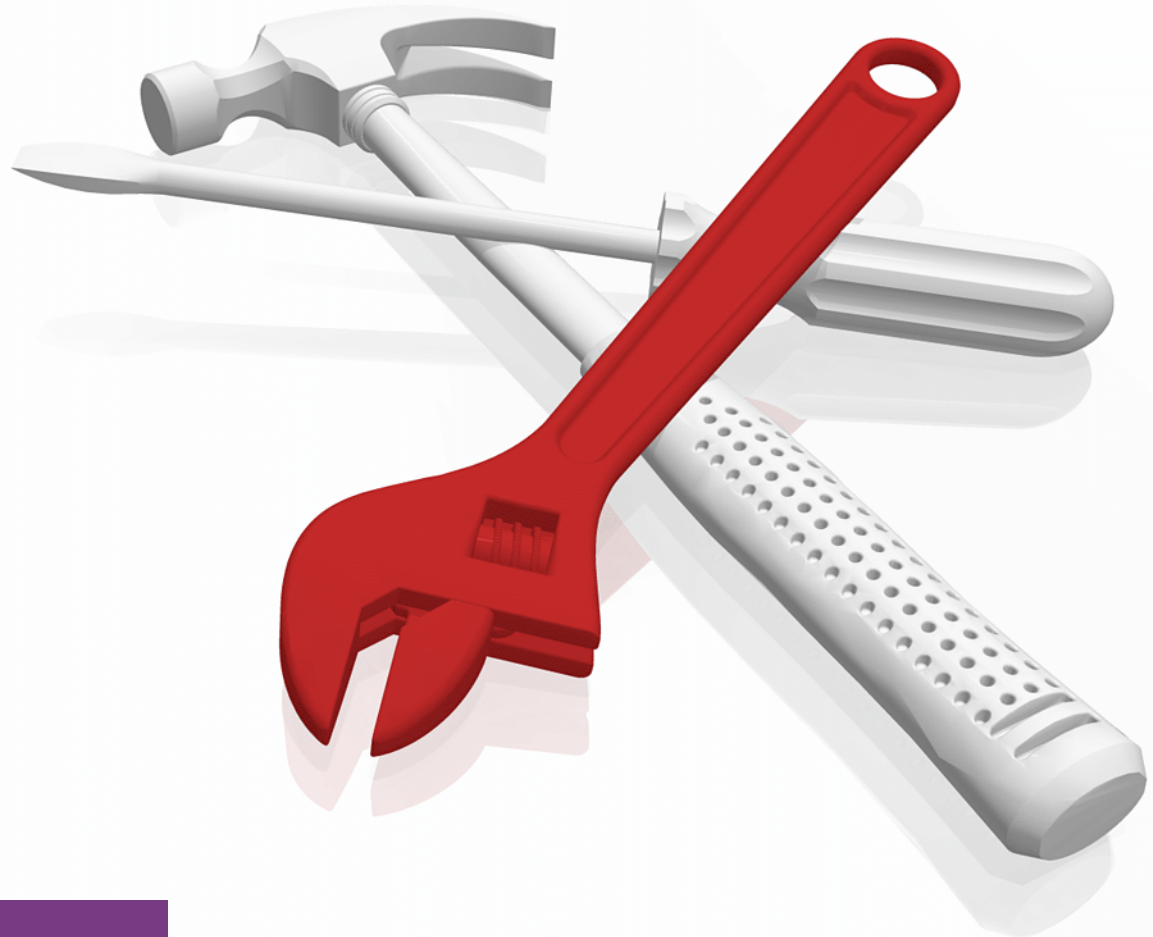


IBM z15 (8561) Technical Guide

Octavian Lascu
Bill White
John Troy
Jannie Houbjerg
Kazuhiro Nakajima
Paul Schouten
Anna Shugol
Frank Packheiser
Hervey Kamga
Bo Xu



IBM Z



IBM Redbooks

IBM z15 (8561) Technical Guide

August 2020

Note: Before using this information and the product it supports, read the information in “Notices” on page xiii.

First Edition (August 2020)

This edition applies to IBM z15 Model T01, Machine Type 8561.

© Copyright International Business Machines Corporation 2020. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

| | |
|---|------|
| Notices | xiii |
| Trademarks | xiv |
| Preface | xv |
| Authors | xv |
| Now you can become a published author, too! | xvii |
| Comments welcome | xvii |
| Stay connected to IBM Redbooks | xvii |
| Chapter 1. Introduction | 1 |
| 1.1 Design considerations for the IBM z15 | 2 |
| 1.1.1 Complementing and augmenting cloud solutions | 2 |
| 1.1.2 Compliance, resiliency, and performance | 2 |
| 1.1.3 Pervasive encryption | 3 |
| 1.1.4 IBM Z Data Privacy Passports | 3 |
| 1.1.5 Blending open source with IBM Z state-of-the-art technologies | 3 |
| 1.2 z15 server highlights | 4 |
| 1.2.1 Processor and memory | 4 |
| 1.2.2 Capacity and performance | 5 |
| 1.2.3 Virtualization | 7 |
| 1.2.4 I/O subsystem and I/O features | 10 |
| 1.2.5 Reliability, availability, and serviceability design | 11 |
| 1.3 z15 server technical overview | 12 |
| 1.3.1 Model and features | 13 |
| 1.3.2 Model upgrade paths | 14 |
| 1.3.3 Frames | 15 |
| 1.3.4 CPC drawer | 15 |
| 1.3.5 I/O connectivity: PCIe+ Generation 3 | 20 |
| 1.3.6 I/O subsystem | 20 |
| 1.3.7 I/O and special purpose features in the PCIe I/O drawer | 21 |
| 1.3.8 Storage connectivity | 22 |
| 1.3.9 Network connectivity | 23 |
| 1.3.10 Coupling and Server Time Protocol connectivity | 25 |
| 1.3.11 Cryptography | 27 |
| 1.4 Reliability, availability, and serviceability | 30 |
| 1.5 Hardware Management Consoles and Support Elements | 31 |
| 1.6 Operating systems | 31 |
| 1.6.1 Supported operating systems | 31 |
| 1.6.2 IBM compilers | 34 |
| Chapter 2. Central processor complex hardware components | 35 |
| 2.1 Frames and configurations | 36 |
| 2.1.1 z15 cover (door) design | 39 |
| 2.1.2 Top exit I/O and cabling | 39 |
| 2.2 CPC drawer | 41 |
| 2.2.1 CPC drawer interconnect topology | 45 |
| 2.2.2 Oscillator | 46 |
| 2.2.3 System control | 47 |
| 2.2.4 CPC drawer power | 48 |

| | | |
|--|---|-----------|
| 2.3 | Single chip modules | 49 |
| 2.3.1 | Processor unit chip | 50 |
| 2.3.2 | Processor unit (core) | 51 |
| 2.3.3 | PU characterization | 52 |
| 2.3.4 | System Controller chip | 53 |
| 2.3.5 | Cache level structure | 54 |
| 2.4 | PCIe+ I/O drawer | 54 |
| 2.5 | Memory | 56 |
| 2.5.1 | Memory subsystem topology | 58 |
| 2.5.2 | Redundant array of independent memory | 58 |
| 2.5.3 | Memory configurations | 59 |
| 2.5.4 | Memory upgrades | 64 |
| 2.5.5 | Drawer replacement and memory | 65 |
| 2.5.6 | Virtual Flash Memory | 65 |
| 2.5.7 | Flexible Memory Option | 65 |
| 2.6 | Reliability, availability, and serviceability | 66 |
| 2.6.1 | RAS in the CPC memory subsystem | 66 |
| 2.6.2 | General z15 T01 RAS features | 67 |
| 2.7 | Connectivity | 70 |
| 2.7.1 | Redundant I/O interconnect | 71 |
| 2.7.2 | Enhanced drawer availability (EDA) | 73 |
| 2.7.3 | CPC drawer upgrade | 73 |
| 2.8 | Model configurations | 73 |
| 2.8.1 | Upgrades | 75 |
| 2.8.2 | Model capacity identifier | 76 |
| 2.9 | Power and cooling | 77 |
| 2.9.1 | PDU-based configurations | 77 |
| 2.9.2 | BPA-based configurations | 79 |
| 2.9.3 | Internal Battery Feature | 81 |
| 2.9.4 | Power estimation tool | 81 |
| 2.9.5 | Cooling | 82 |
| 2.9.6 | Radiator Cooling Unit | 82 |
| 2.9.7 | Water-cooling unit | 85 |
| 2.10 | Summary | 87 |
| Chapter 3. Central processor complex design | | 89 |
| 3.1 | Overview | 90 |
| 3.2 | Design highlights | 90 |
| 3.3 | CPC drawer design | 92 |
| 3.3.1 | Cache levels and memory structure | 93 |
| 3.3.2 | CPC drawer interconnect topology | 96 |
| 3.4 | Processor unit design | 97 |
| 3.4.1 | Simultaneous multithreading | 98 |
| 3.4.2 | Single-instruction multiple-data | 99 |
| 3.4.3 | Out-of-Order execution | 102 |
| 3.4.4 | Superscalar processor | 104 |
| 3.4.5 | Compression and cryptography accelerators on a chip | 104 |
| 3.4.6 | Decimal floating point accelerator | 108 |
| 3.4.7 | IEEE floating point | 109 |
| 3.4.8 | Processor error detection and recovery | 109 |
| 3.4.9 | Branch prediction | 109 |
| 3.4.10 | Wild branch | 110 |
| 3.4.11 | Translation lookaside buffer | 111 |

| | | |
|-------------------|--|------------|
| 3.4.12 | Instruction fetching, decoding, and grouping | 111 |
| 3.4.13 | Extended Translation Facility | 112 |
| 3.4.14 | Instruction set extensions | 112 |
| 3.4.15 | Transactional Execution | 112 |
| 3.4.16 | Runtime Instrumentation | 112 |
| 3.5 | Processor unit functions | 113 |
| 3.5.1 | Overview | 113 |
| 3.5.2 | Central processors | 114 |
| 3.5.3 | Integrated Facility for Linux (FC 1945) | 115 |
| 3.5.4 | Internal Coupling Facility (FC 1946) | 116 |
| 3.5.5 | IBM Z Integrated Information Processor (FC 1947) | 120 |
| 3.5.6 | System assist processors | 124 |
| 3.5.7 | Reserved processors | 125 |
| 3.5.8 | Integrated firmware processor | 125 |
| 3.5.9 | Processor unit assignment | 126 |
| 3.5.10 | Sparing rules | 127 |
| 3.5.11 | CPC drawer numbering | 127 |
| 3.6 | Memory design | 128 |
| 3.6.1 | Overview | 128 |
| 3.6.2 | Main storage | 131 |
| 3.6.3 | Hardware system area | 131 |
| 3.6.4 | Virtual Flash Memory (FC 0643) | 132 |
| 3.7 | Logical partitioning | 132 |
| 3.7.1 | Overview | 132 |
| 3.7.2 | Storage operations | 139 |
| 3.7.3 | Reserved storage | 141 |
| 3.7.4 | Logical partition storage granularity | 142 |
| 3.7.5 | LPAR dynamic storage reconfiguration | 142 |
| 3.8 | Intelligent Resource Director | 142 |
| 3.9 | Clustering technology | 144 |
| 3.9.1 | CF Control Code | 145 |
| 3.9.2 | Coupling Thin Interrupts | 148 |
| 3.9.3 | Dynamic CF dispatching | 148 |
| 3.10 | Virtual Flash Memory | 150 |
| 3.10.1 | IBM Z Virtual Flash Memory overview | 150 |
| 3.10.2 | VFM feature | 150 |
| 3.10.3 | VFM administration | 151 |
| 3.11 | Secure Service Container | 151 |
| Chapter 4. | Central processor complex I/O structure | 153 |
| 4.1 | Introduction to I/O infrastructure | 154 |
| 4.1.1 | I/O infrastructure | 154 |
| 4.1.2 | PCIe Generation 3 | 155 |
| 4.2 | I/O system overview | 156 |
| 4.2.1 | Characteristics | 156 |
| 4.2.2 | Supported I/O features | 157 |
| 4.3 | PCIe+ I/O drawer | 158 |
| 4.3.1 | PCIe+ I/O drawer offerings | 160 |
| 4.4 | CPC drawer fanouts | 161 |
| 4.4.1 | PCIe+ Generation 3 fanout (FC 0175) | 162 |
| 4.4.2 | Integrated Coupling Adapter (FC 0172 and 0176) | 162 |
| 4.4.3 | Fanout considerations | 163 |
| 4.5 | I/O features | 164 |

| | |
|--|------------|
| 4.5.1 I/O feature card ordering information | 165 |
| 4.5.2 Physical channel ID report | 166 |
| 4.6 Connectivity. | 167 |
| 4.6.1 I/O feature support and configuration rules. | 167 |
| 4.6.2 Storage connectivity | 172 |
| 4.6.3 Network connectivity | 181 |
| 4.6.4 Parallel Sysplex connectivity. | 195 |
| 4.7 Cryptographic functions | 200 |
| 4.7.1 CPACF functions (FC 3863) | 200 |
| 4.7.2 Crypto Express7S feature (FC 0898 and FC 0899) | 200 |
| 4.7.3 Crypto Express6S feature (FC 0893) as carry forward only | 201 |
| 4.7.4 Crypto Express5S feature (FC 0890) as carry forward only | 202 |
| 4.8 Integrated Firmware Processor. | 202 |
| Chapter 5. Central processor complex channel subsystem. | 203 |
| 5.1 Channel subsystem. | 204 |
| 5.1.1 Multiple logical channel subsystems. | 205 |
| 5.1.2 Multiple subchannel sets. | 206 |
| 5.1.3 Channel path spanning. | 209 |
| 5.2 I/O configuration management | 212 |
| 5.3 Channel subsystem summary. | 213 |
| Chapter 6. Cryptographic features | 215 |
| 6.1 Cryptography enhancements on IBM z15. | 216 |
| 6.2 Cryptography overview | 217 |
| 6.2.1 Modern cryptography | 217 |
| 6.2.2 Kerckhoffs' principle | 218 |
| 6.2.3 Keys | 218 |
| 6.2.4 Algorithms. | 220 |
| 6.3 Cryptography on IBM z15 | 221 |
| 6.4 CP Assist for Cryptographic Functions | 224 |
| 6.4.1 Cryptographic synchronous functions. | 226 |
| 6.4.2 CPACF protected key | 227 |
| 6.5 Crypto Express7S | 230 |
| 6.5.1 Cryptographic asynchronous functions. | 232 |
| 6.5.2 Crypto Express7S as a CCA coprocessor | 233 |
| 6.5.3 Crypto Express7S as an EP11 coprocessor. | 239 |
| 6.5.4 Crypto Express7S as an accelerator. | 240 |
| 6.5.5 Managing Crypto Express7S | 240 |
| 6.6 Trusted Key Entry workstation | 243 |
| 6.6.1 Logical partition, TKE host, and TKE target | 246 |
| 6.6.2 Optional smart card reader | 246 |
| 6.6.3 TKE hardware support and migration information. | 247 |
| 6.7 Cryptographic functions comparison. | 249 |
| 6.8 Cryptographic operating system support for z15 | 251 |
| 6.8.1 Crypto Express7S Toleration | 251 |
| 6.8.2 Crypto Express7S support of VFPE | 251 |
| 6.8.3 Crypto Express7S support of greater than 16 domains | 252 |
| Chapter 7. Operating system support. | 253 |
| 7.1 Operating systems summary | 254 |
| 7.2 Support by operating system | 254 |
| 7.2.1 z/OS | 255 |
| 7.2.2 z/VM | 255 |

| | | |
|-----------------------------------|---|------------|
| 7.2.3 | z/VSE | 255 |
| 7.2.4 | z/TPF | 255 |
| 7.2.5 | Linux on IBM Z (Linux on Z) | 256 |
| 7.2.6 | KVM hypervisor | 257 |
| 7.3 | z15 features and function support overview | 257 |
| 7.3.1 | Supported CPC functions | 258 |
| 7.3.2 | Coupling and clustering | 261 |
| 7.3.3 | Network connectivity | 266 |
| 7.3.4 | Cryptographic functions | 271 |
| 7.4 | Support by features and functions | 273 |
| 7.4.1 | LPAR Configuration and Management | 273 |
| 7.4.2 | Base CPC features and functions | 277 |
| 7.4.3 | Coupling and clustering features and functions | 288 |
| 7.4.4 | Storage connectivity-related features and functions | 293 |
| 7.4.5 | Networking features and functions | 305 |
| 7.4.6 | Cryptography Features and Functions Support | 317 |
| 7.5 | z/OS migration considerations | 322 |
| 7.5.1 | General guidelines | 322 |
| 7.5.2 | Hardware Fix Categories (FIXCATs) | 323 |
| 7.5.3 | Coupling links | 324 |
| 7.5.4 | z/OS XL C/C++ considerations | 324 |
| 7.5.5 | z/OS V2.4 | 325 |
| 7.5.6 | z/OS V2.3 | 326 |
| 7.6 | z/VM migration considerations | 326 |
| 7.6.1 | z/VM 7.2 | 326 |
| 7.6.2 | z/VM 7.1 | 327 |
| 7.6.3 | z/VM V6.4 | 327 |
| 7.6.4 | ESA/390-compatibility mode for guests | 327 |
| 7.6.5 | Capacity | 328 |
| 7.7 | z/VSE migration considerations | 328 |
| 7.8 | Software licensing | 328 |
| 7.9 | References | 331 |
| Chapter 8. System upgrades | | 333 |
| 8.1 | Permanent and Temporary Upgrades | 335 |
| 8.1.1 | Overview | 335 |
| 8.1.2 | CoD for z15 systems-related terminology | 336 |
| 8.1.3 | Concurrent and nondisruptive upgrades | 338 |
| 8.1.4 | Permanent upgrades | 338 |
| 8.1.5 | Temporary upgrades | 339 |
| 8.2 | Concurrent upgrades | 340 |
| 8.2.1 | PU Capacity feature upgrades | 340 |
| 8.2.2 | Customer Initiated Upgrade facility | 342 |
| 8.2.3 | Concurrent upgrade functions summary | 347 |
| 8.3 | Miscellaneous equipment specification upgrades | 347 |
| 8.3.1 | MES upgrade for processors | 348 |
| 8.3.2 | MES upgrades for memory | 350 |
| 8.3.3 | MES upgrades for I/O | 351 |
| 8.3.4 | Feature on Demand | 352 |
| 8.3.5 | Summary of plan-ahead feature | 353 |
| 8.4 | Permanent upgrade by using the CIU facility | 353 |
| 8.4.1 | Ordering | 355 |
| 8.4.2 | Retrieval and activation | 356 |

| | | |
|--|--|------------|
| 8.5 | On/Off Capacity on Demand | 357 |
| 8.5.1 | Overview | 357 |
| 8.5.2 | Capacity Provisioning Manager | 358 |
| 8.5.3 | Ordering | 359 |
| 8.5.4 | On/Off CoD testing | 362 |
| 8.5.5 | Activation and deactivation | 364 |
| 8.5.6 | Termination | 364 |
| 8.6 | z/OS Capacity Provisioning | 365 |
| 8.7 | System Recovery Boost Upgrade | 369 |
| 8.8 | Capacity for Planned Event | 370 |
| 8.9 | Capacity Backup | 372 |
| 8.9.1 | Ordering | 372 |
| 8.9.2 | CBU activation and deactivation | 374 |
| 8.9.3 | Automatic CBU enablement for GDPS | 375 |
| 8.10 | Planning for nondisruptive upgrades | 376 |
| 8.10.1 | Components | 376 |
| 8.10.2 | Concurrent upgrade considerations | 377 |
| 8.11 | Summary of Capacity on-Demand offerings | 380 |
| Chapter 9. Reliability, availability, and serviceability | | 383 |
| 9.1 | RAS strategy | 384 |
| 9.2 | Technology | 384 |
| 9.2.1 | Processor Unit chip | 384 |
| 9.2.2 | System Controller and main memory | 386 |
| 9.2.3 | I/O and service | 386 |
| 9.3 | Structure | 387 |
| 9.4 | Reducing complexity | 387 |
| 9.5 | Reducing touches | 388 |
| 9.6 | z15 availability characteristics | 388 |
| 9.7 | z15 RAS functions | 392 |
| 9.7.1 | Scheduled outages | 393 |
| 9.7.2 | Unscheduled outages | 395 |
| 9.8 | z15 enhanced drawer availability | 396 |
| 9.8.1 | EDA planning considerations | 396 |
| 9.8.2 | Enhanced drawer availability processing | 398 |
| 9.9 | z15 Enhanced Driver Maintenance | 404 |
| 9.9.1 | Resource Group and native PCIe features MCLs | 405 |
| 9.10 | RAS capability for the HMC and SE | 408 |
| Chapter 10. Hardware Management Console and Support Element | | 411 |
| 10.1 | HMC and SE introduction | 412 |
| 10.1.1 | Dynamic Partition Manager support | 412 |
| 10.2 | HMC and SE changes and new features | 413 |
| 10.2.1 | Driver Level 41 HMC and SE new features | 413 |
| 10.2.2 | New Rack-mounted HMC and Tower HMC | 416 |
| 10.2.3 | New Support Element | 417 |
| 10.2.4 | New service and functional operations for HMCs and SEs | 418 |
| 10.2.5 | SE driver support with the HMC driver | 419 |
| 10.2.6 | HMC feature codes | 419 |
| 10.2.7 | User interface | 420 |
| 10.2.8 | Customize Product Engineering Access: Best practice | 420 |
| 10.3 | HMC and SE connectivity | 421 |
| 10.3.1 | Standard HMC connectivity | 421 |

| | | |
|--------------------|--|------------|
| 10.3.2 | Hardware Management Appliance | 422 |
| 10.3.3 | Network planning for the HMC and SE | 423 |
| 10.3.4 | Hardware considerations | 426 |
| 10.3.5 | TCP/IP Version 6 on the HMC and SE | 426 |
| 10.3.6 | OSA Support Facility | 426 |
| 10.3.7 | Assigning addresses to the HMC and SE | 427 |
| 10.3.8 | HMC Multi-factor authentication | 428 |
| 10.4 | Remote Support Facility | 429 |
| 10.4.1 | Security characteristics | 429 |
| 10.4.2 | RSF connections to IBM and Enhanced IBM Service Support System | 429 |
| 10.4.3 | HMC and SE remote operations | 430 |
| 10.5 | HMC and SE capabilities | 432 |
| 10.5.1 | Central processor complex management | 432 |
| 10.5.2 | LPAR management | 432 |
| 10.5.3 | Operating system communication | 434 |
| 10.5.4 | HMC and SE microcode | 435 |
| 10.5.5 | Monitoring | 439 |
| 10.5.6 | Capacity on-demand support | 441 |
| 10.5.7 | Server Time Protocol support | 442 |
| 10.5.8 | CTN Split and Merge | 445 |
| 10.5.9 | NTP client and server support on the HMC | 445 |
| 10.5.10 | Security and user ID management | 446 |
| 10.5.11 | System Input/Output Configuration Analyzer on the SE and HMC | 448 |
| 10.5.12 | Automated operations | 449 |
| 10.5.13 | Cryptographic support | 449 |
| 10.5.14 | Installation support for z/VM that uses the HMC | 450 |
| 10.5.15 | Dynamic Partition Manager | 451 |
| Chapter 11. | Environmentals | 453 |
| 11.1 | Power and Cooling | 454 |
| 11.1.1 | Intelligent Power Distribution Unit (iPDU) | 454 |
| 11.1.2 | Bulk Power assembly (BPA) | 456 |
| 11.1.3 | Cooling requirements | 461 |
| 11.1.4 | Internal Battery Feature | 464 |
| 11.2 | Physical specifications | 466 |
| 11.3 | Physical planning | 467 |
| 11.3.1 | Raised floor or non-raised floor | 467 |
| 11.3.2 | Top Exit cabling feature (optional) | 467 |
| 11.3.3 | Top or bottom exit cables | 469 |
| 11.3.4 | Bottom Exit cabling feature | 470 |
| 11.3.5 | Frame Bolt-down kit | 470 |
| 11.3.6 | Service clearance areas | 470 |
| 11.4 | Energy management | 471 |
| 11.4.1 | Environmental monitoring | 472 |
| Chapter 12. | Performance | 475 |
| 12.1 | IBM z15 performance characteristics | 476 |
| 12.1.1 | z15 single-thread capacity | 476 |
| 12.1.2 | z15 SMT capacity | 476 |
| 12.1.3 | IBM Integrated Accelerator for zEnterprise Data Compression | 477 |
| 12.1.4 | Primary performance improvement drivers with z15 | 477 |
| 12.2 | z15 Large System Performance Reference ratio | 478 |
| 12.2.1 | LSPR workload suite | 479 |

| | |
|--|------------|
| 12.3 Fundamental components of workload performance | 479 |
| 12.3.1 Instruction path length. | 480 |
| 12.3.2 Instruction complexity | 480 |
| 12.3.3 Memory hierarchy and memory nest. | 480 |
| 12.4 Relative Nest Intensity | 481 |
| 12.5 LSPR workload categories based on RNI. | 483 |
| 12.6 Relating production workloads to LSPR workloads | 483 |
| 12.7 CPU MF counter data and LSPR workload type. | 484 |
| 12.8 Workload performance variation | 485 |
| 12.9 Capacity planning consideration for z15 | 485 |
| 12.9.1 Collect CPU MF counter data | 485 |
| 12.9.2 Creating EDF file with CP3KEXTR | 486 |
| 12.9.3 Loading EDF file to the capacity planning tool | 486 |
| 12.9.4 Tips to maximize z15 server capacity. | 486 |
| Appendix A. Channel options | 489 |
| Appendix B. System Recovery Boost | 493 |
| B.1 Overview. | 494 |
| B.1.1 Use cases. | 494 |
| B.2 Functions | 494 |
| B.3 Delivering extra capacity. | 496 |
| B.3.1 Subcapacity CPs speed boost | 496 |
| B.3.2 zIIP processor capacity boost (zIIP boost) | 497 |
| B.3.3 Optional System Recovery Boost Upgrade capacity record (z15 T01) | 498 |
| B.3.4 Planned shutdown boost | 499 |
| B.3.5 IBM Geographically Dispersed Parallel Sysplex Actions Performance and Parallelism | 500 |
| B.3.6 Process recovery boosts (short duration). | 501 |
| B.4 Setting up the System Recovery Boost | 503 |
| B.5 Monitoring System Recovery Boost | 504 |
| B.6 Automation | 505 |
| B.7 Pricing. | 506 |
| B.7.1 Base System Recovery Boost feature: No extra charge functions. | 506 |
| B.7.2 Priced feature: System Recovery Boost Upgrade record. | 506 |
| B.7.3 Software pricing | 506 |
| B.8 Software support. | 506 |
| Appendix C. IBM Integrated Accelerator for zEnterprise Data Compression | 509 |
| Client value of Z compression | 510 |
| z15 IBM Integrated Accelerator for zEDC | 510 |
| Eliminating adapter sharing by using Nest Compression Accelerator | 511 |
| Compression modes | 511 |
| z15 migration considerations | 511 |
| All z/OS configuration stay the same | 511 |
| Consider fail-over and DR sizing. | 511 |
| Performance metrics | 512 |
| zEDC to z15 zlib Program Flow for z/OS | 512 |
| Software support | 512 |
| C.0.1 IBM Z Batch Network Analyzer. | 513 |
| Compression acceleration and Linux on Z. | 513 |
| Appendix D. Frame configurations | 515 |
| Power Distribution Unit configurations | 516 |

| | |
|--|-----|
| Bulk Power Assembly configurations | 522 |
| Related publications | 529 |
| IBM Redbooks | 529 |
| Other publications | 529 |
| Online resources | 529 |
| Help from IBM | 529 |

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.


COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at <http://www.ibm.com/legal/copytrade.shtml>

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

| | | |
|---|---|-----------------|
| AIX® | IBM z13® | System z® |
| Bluemix® | IBM z13s® | System z10® |
| CICS® | IBM z14® | System z9® |
| Db2® | IBM z15™ | VIA® |
| DB2® | Interconnect® | VTAM® |
| Distributed Relational Database Architecture™ | Language Environment® | Watson™ |
| DS8000® | MVS™ | WebSphere® |
| FICON® | OMEGAMON® | z Systems® |
| FlashCopy® | Parallel Sysplex® | z/Architecture® |
| GDPS® | Passport Advantage® | z/OS® |
| Global Technology Services® | PowerPC® | z/VM® |
| HyperSwap® | RACF® | z/VSE® |
| IBM® | Redbooks® | z13® |
| IBM Watson® | Redbooks (logo)  ® | z13s® |
| IBM Z® | Resource Link® | z15™ |
| IBM z Systems® | S/390® | z9® |
| | System Storage™ | zEnterprise® |

The following terms are trademarks of other companies:

Evolution, are trademarks or registered trademarks of Kenexa, an IBM Company.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Red Hat, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

VMware, and the VMware logo are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

Other company, product, or service names may be trademarks or service marks of others.

Preface

This IBM® Redbooks® publication describes the features and functions the latest member of the IBM Z® platform, the IBM z15™ (machine type 8561). It includes information about the IBM z15 processor design, I/O innovations, security features, and supported operating systems.

The z15 is a state-of-the-art data and transaction system that delivers advanced capabilities, which are vital to any digital transformation. The z15 is designed for enhanced modularity, which is in an industry standard footprint. This system excels at the following tasks:

- ▶ Making use of multicloud integration services
- ▶ Securing data with pervasive encryption
- ▶ Accelerating digital transformation with agile service delivery
- ▶ Transforming a transactional platform into a data powerhouse
- ▶ Getting more out of the platform with IT Operational Analytics
- ▶ Accelerating digital transformation with agile service delivery
- ▶ Revolutionizing business processes
- ▶ Blending open source and Z technologies

This book explains how this system uses new innovations and traditional Z strengths to satisfy growing demand for cloud, analytics, and open source technologies. With the z15 as the base, applications can run in a trusted, reliable, and secure environment that improves operations and lessens business risk.

Authors

This book was produced by a team of specialists from around the world working at IBM Redbooks, Poughkeepsie Center.

Octavian Lascu is an IBM Redbooks Project Leader and a Senior IT Consultant for IBM Romania with over 25 years of experience. He specializes in designing, implementing, and supporting complex IT infrastructure environments (systems, storage, and networking), including high availability and disaster recovery solutions and high-performance computing deployments. He has developed materials for and taught over 50 workshops for technical audiences around the world. He is the author of several IBM publications.

Bill White is an IBM Redbooks Project Leader and Senior Networking and Connectivity Specialist at IBM Redbooks, Poughkeepsie Center.

John Troy is an IBM Z and storage hardware National Top Gun in the northeast area of the United States. He has 40 years of experience in the service field. His areas of expertise include IBM Z servers and high-end storage systems technical and customer support and services. John has also been an IBM Z hardware technical support course designer, developer, and instructor for the last eight generations of IBM high-end servers.

Jannie Houlbjerg is a Systems Programmer working at JN Data in Denmark. She has more than 20 years of experience in the IBM Z field. Her areas of expertise include IBM Z hardware and infrastructure, IBM Parallel Sysplex®, connectivity, performance, IBM GDPS®, and technical project management and documentation.

Kazuhiro Nakajima is a Senior IT Specialist in IBM Japan. He has almost 30 years career in IBM Japan and he has been active as an advanced Subject Matter Expert of IBM Z products for over 20 years. His areas of expertise include IBM Z hardware, performance, z/OS®, and IBM Z connectivity. He has been a co-author of several IBM Z configuration set up IBM Redbooks publications from the IBM zEC12 to the IBM z14®.

Paul Schouten is an IBM Z Client Technical Specialist based in Sydney, Australia. During his 40 years supporting mainframe systems, he has performed many roles, including Certified IT Architect, Systems Software developer, and Systems Programming. He has extensive experience developing and documenting high availability solutions for IBM's enterprise customers.

Anna Shugol is a mainframe technical specialist with IBM UK. She has over 8 years of Mainframe experience and has worked with clients in various geographies. Large system performance is one of her key areas of expertise. Anna holds a Computer Science degree from Bauman Moscow State Technical University.

Frank Packheiser is a Senior zIT Specialist at the Field Technical Sales Support office in Germany. He has over 25 years of experience in IBM Z. Frank has worked for 10 years in the IBM Education Center in Germany, developing and providing professional training. He also provides professional services to IBM Z and mainframe clients. In 2008 and 2009, Frank supported clients in Middle East/North Africa (MENA) as a zIT Architect. In addition to co-authoring several IBM Redbooks publications since 1999, he has been an official ITSO presenter at ITSO workshops since 2010.

Hervey Kamga is an IBM Z Product Engineer with the EMEA I/O Connectivity Team in Montpellier, France. Before serving in his current role, he was a Support Engineer and Engineer On Site for 13 years with Sun Microsystems and Oracle in EMEA. Hervey's areas of expertise include Oracle Solaris (operating system and hardware products), virtualization (VMware and virtualBox), Linux, and IBM Z I/O features and protocols (IBM FICON® and OSA) while co-authoring several IBM Redbooks publications.

Bo Xu is a consulting product Service SSR in China. He has more than 20 years of experience with IBM Z platform maintenance support. He has been working in IBM's Global Technology Services® department to provide IBM Z platform support to clients as a local Top Gun, and as second-level support for IBM DS8000® as one of the country's Top Gun & Skill Owners. His areas of expertise include IBM Z platform hardware, channel connectivity, and IBM DS8000 storage.

Thanks to the following people for their contributions to this project:

Robert Haimowitz
IBM Redbooks, Poughkeepsie Center

Dave Surman, Darelle Gent, David Hutton, Tom Dewkett, Frank Bosco, Patty Driever, Anthony Sofia, Brian Valentine, Purvi Patel, Les Geer III, Bill Bitner, Christine Smith, Barbara Weiler, Dean St Piere, Tom Morris, Ellen Carbarnes, Gregory Hutchison, Riaz Ahmad, Franco Pinto, Walter Niklaus, Martin Recktenwald, Roan Dawkins

IBM

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an IBM Redbooks residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at: ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- ▶ Use the online **Contact us** review Redbooks form found at:
ibm.com/redbooks
- ▶ Send your comments in an email to:
redbooks@us.ibm.com
- ▶ Mail your comments to:
IBM Corporation, IBM Redbooks
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- ▶ Find us on Facebook:
<http://www.facebook.com/IBMRedbooks>
- ▶ Follow us on Twitter:
<http://twitter.com/ibmredbooks>
- ▶ Look for us on LinkedIn:
<http://www.linkedin.com/groups?home=&gid=2130806>
- ▶ Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:
<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>
- ▶ Stay current on recent Redbooks publications with RSS Feeds:
<http://www.redbooks.ibm.com/rss.html>



Introduction

This chapter describes the basic concepts and design considerations around IBM z15™ servers and includes the following topics:

- ▶ 1.1, “Design considerations for the IBM z15” on page 2
- ▶ 1.2, “z15 server highlights” on page 4
- ▶ 1.3, “z15 server technical overview” on page 12
- ▶ 1.4, “Reliability, availability, and serviceability” on page 30
- ▶ 1.5, “Hardware Management Consoles and Support Elements” on page 31
- ▶ 1.6, “Operating systems” on page 31

Naming: The IBM z15 server generation is available as the following machine types and models:

- ▶ Machine Type 8561 (M/T 8561), Model T01, Features Max34, Max71, Max108, Max145, and Max190, which is further identified as *IBM z15 Model T01*, or *z15 T01*, unless otherwise specified.
- ▶ Machine Type 8562 (M/T 8562), Model T02, Features Max4, Max13, Max21, Max32, and Max65, which is further identified as *IBM z15 Model T02*, or *z15 T02*, unless otherwise specified.

In the remainder of this chapter, IBM z15 (z15) refers to both machine types.

1.1 Design considerations for the IBM z15

Delivering new services efficiently and effectively with speed and at scale is crucial to any business, large or small. Managing changes and potential disruptions to the IT infrastructure while maintaining availability of services is a must. Ensuring data privacy and protection to mitigate impacts of security breaches is what every business expects from its IT infrastructure.

There is no doubt that the technology you choose determines business success and differentiates you from the competition. Technology choices must help meet the expectations of rapid dynamic development cycles and give the confidence that new and existing services are resilient and can be delivered quickly and securely. Therefore, the correct balance of open source technologies and the IT platform on which they run is also key.

1.1.1 Complementing and augmenting cloud solutions

Today, many business use cloud technologies to deliver cloud native services at scale and at lower cost, but often the risk of vendor lock-in and escalating costs exists. Also, approximately 80% of enterprise applications and services have not yet moved to the cloud because of struggles with flexibility, connectivity, security, and management across multicloud environments.

1.1.2 Compliance, resiliency, and performance

The latest member of the IBM Z family, the z15, features a tried-and-true architecture to satisfy today's demands. It can help you create an open, secure, and resilient infrastructure that streamlines your ability to integrate disparate cloud environments and create a single, cohesive IT infrastructure that provides high availability, scalability, and performance. The z15 can also help you fully protect your data (in-flight and at-rest), providing the strongest workload isolation, while facilitating regulatory compliance.

The IBM Integrated Accelerator for zEnterprise® Data Compression (zEDC) provides on-chip compression (processor nest compression accelerator), which supports DEFLATE-compliant compression and decompression and GZIP CRC/ZLIB Adler. Compared to the zEDC Express (PCIe feature), the z15 on-chip compression provides low latency and high bandwidth while eliminating the need for the virtualization layer. The nest accelerator is now accessible as a designed instruction, which is available for problem state programs.

New with z15, the System Recovery Boost feature offers the customer more CP capacity during system recovery operations, such as shutdown, IPL, and stand-alone dumps, to “speed up” the recovery. The feature is embedded in the z15 firmware and is available to operating systems (opt-in for supported OS-es) to accelerate recovery from maintenance tasks by providing more CP capacity to opt in LPARs.

For more information about System Recovery Boost, see Chapter 7, “Operating system support” on page 253 and Appendix B, “System Recovery Boost” on page 493.

1.1.3 Pervasive encryption

Cryptography is in the DNA of IBM Z family. The IBM z15 continues that tradition with pervasive encryption to defend and protect your critical assets with unrivaled encryption and intelligent data monitoring without compromising transactional throughput or response times. Most importantly, pervasive encryption requires no application changes.

Pervasive encryption can dramatically simplify data protection and reduce the costs that are associated with regulatory compliance. By using simple policy controls, z15 pervasive computing streamlines data protection for mission critical IBM Db2® for z/OS, IBM IMS, and Virtual Storage Access Method (VSAM) datasets.

The Central Processor Assist for Cryptographic Function (CPACF), which is standard on every core, supports pervasive encryption and provides hardware acceleration for encryption operations. The new Crypto-Express7S gets a performance boost on z15. Combined, these enhancements perform encryption more efficiently on the z15 than on earlier IBM Z servers.

1.1.4 IBM Z Data Privacy Passports

The new IBM Z Data Privacy Passports with IBM z15 is designed to enforce security and privacy protections to data not only on Z, but across platforms during the extract, transform, and load (ETL) process. It provides a data-centric security solution that complements Pervasive Encryption available on IBM Z while enabling data to play an active role in its own protection.

1.1.5 Blending open source with IBM Z state-of-the-art technologies

The IBM z15 T01 was designed specifically to meet the demand for new services and customer experiences, while securing the growing amounts of data and complying with increasingly intricate regulations. With up to 190 configurable cores, z15 T01 has performance and scaling advantage over prior generation and 25% more capacity than the 170-way z14 M05.

Optimized SMT on z15 delivers improved virtualization performance to benefit Linux. High-speed connectivity out to the data is critical in achieving exceptional levels of transaction throughput. The IBM zHyperLink Express introduces disk I/O technology for accessing the IBM DS8880 storage system with low latency, which enables shorter batch windows and a more resilient I/O infrastructure with predictable and repeatable I/O performance.

With up to 40 TB of memory, z15 T01 can open opportunities, such as in-memory data marts and in-memory analytics, while giving you the necessary room to tune applications for optimal performance. By using the Vector Packed Decimal Facility that allows packed decimal operations to be performed in registers rather than memory, and new fast mathematical computations, compilers (such as Enterprise COBOL for z/OS, V6.2, Enterprise PL/I for z/OS, V5.2, z/OS V2.4 XL C/C++), the COBOL optimizer, Automatic Binary Optimizer for z/OS, V1.3, and Java, are optimized on z15. These compilers and optimizers are designed to improve application performance and reduce CPU usage and operating costs.

Java improvements and the use of crypto-acceleration deliver more improvements in throughput per core, which gives a natural boost to z/OS Connect EE, IBM WebSphere® Liberty in IBM CICS®, Spark for z/OS, and IBM Java for Linux on Z.

Smoothly handling the data tsunami requires robust infrastructure that is designed specifically for high-volume data transactions. To take advantage of new unstructured data,¹ businesses on IBM Z can use application programming interfaces (APIs) that can help with creating and delivering innovative services.

Linux on IBM Z, which is optimized for open source software, brings more value to the platform. Linux on IBM Z supports a wealth of new products that are familiar to application developers, such as Python, Scala, Spark, MongoDB, PostgreSQL, and MariaDB. By accessing core business data directly on platform (without the need for ETL², and hence no data offload off Z platform), you can develop new intelligent applications and business processes.

As your business technology needs evolve to compete in today's digital economy, IBM stands ready to help with intelligent, robust, and comprehensive technology solutions. The IBM approach integrates server, software, and storage solutions to ensure that each member of the stack is designed and optimized to work together. The new IBM z15™ leads that approach by delivering the power and speed users demand, the security users and regulators require, and the operational efficiency that maximizes your bottom line.

Terminology note: The remainder of this book uses the designation *CPC* to refer to the *central processor complex*.

1.2 z15 server highlights

This section reviews some of the following most important features and functions of z15 (Driver 41) servers:

- ▶ Processor and memory
- ▶ Capacity and performance
- ▶ Virtualization
- ▶ I/O subsystem and I/O features
- ▶ Reliability, availability, and serviceability design

1.2.1 Processor and memory

The z15 T01 system is packaged in 19-inch format frames. With 1 - 4 frames, and processor and I/O modular architecture, it provides flexible configuration and fit for purpose systems.

IBM continues its technology leadership with the z15 server. The z15 T01 server is built by using the IBM modular multi-processor drawer design that supports 1 - 5 processor drawers per CPC. Each processor drawer contains four Processor Unit (PU) single-chip modules (SCMs) and one Storage Controller (SC) SCM.

Both SCMs are redesigned by using 14 nm FINFET SOI technology.³ Each PU SCM has 12 processor units (PUs, or cores). In addition to SCMs, CPC drawers host memory DIMMs, connectors for I/O, redundant power supplies, combined Flexible Service Processors and Oscillators (FSP/OSC), and cooling manifolds.

¹ This data accounts for 80% of all data that is generated today and is expected to grow to over 93% by 2020.

² Extract, Transform, Load (ETL) database operations that extract data from one or more databases, and transform and load it into another database for analyzing data in a different context.

³ FINFET is the industry solution; SOI is the IBM solution for SER.

The superscalar processor implements third-generation Simultaneous Multi-Threading (SMT)⁴. It also implements redesigned OoO, augmented caches and translation lookaside buffer (TLB), optimized pipeline, and better branch prediction.

Also featured is an expanded instruction set with Vector Packed Decimal Facility, Guarded Storage Facility, Vector Facility enhancements, Semaphore Assist Facility, Order Preserving Compression, Entropy Encoding for Co-processor Compression for better performance in several different areas, and the IBM Integrated Accelerator for zEnterprise Data Compression (on-chip compression accelerator).

Depending on the model, the z15 T01 server can support 512 GB - 40 TB of usable memory, with up to 8 TB of usable memory per CPC drawer. In addition, a fixed amount of 256 GB is reserved for the hardware system area (HSA) and is not part of customer-purchased memory. Memory is implemented as a redundant array of independent memory (RAIM) and uses extra physical memory as spare memory. The RAIM function accounts for 20% of the physical installed memory in each CPC drawer.

New with z15 T01, Virtual Flash Memory (VFM) feature is offered from the main memory capacity in 0.5 TB units (versus 1.5 TB per feature on z14 M0x) increasing granularity for the feature. VFM provides much simpler management and better performance by eliminating the I/O to the adapters in the PCIe+ I/O drawers. VFM does not require any application changes when moving from IBM zFlash Express (previously available on z13® and zEC12).

1.2.2 Capacity and performance

The z15 T01 server provides increased processing and enhanced I/O capabilities over its predecessor, the z14 M0x system. This capacity is achieved by increasing the performance of the individual PUs and the number of PUs per system, redesigning the system cache, increasing the amount of memory, and introducing new I/O technologies.

The increased performance and the total system capacity available (with potential energy savings) allows consolidation of diverse applications on a single platform with significant financial savings. The introduction of new technologies and an expanded and enhanced instruction set ensure that the z15 server is a high-performance, reliable, and rich-security platform. The z15 server is designed to maximize the use of resources and allows you to integrate and consolidate applications and data across the enterprise IT infrastructure.

z15 Model T01 server is offered with five maximum processor features, with 1 - 190 configurable PUs. The processor features Max34, Max71, Max108, and Max145 have one, two, three, respective four CPC drawers with 41 active PUs per CPC drawer. The high-capacity feature (Max190) has five processor (CPC) drawers with 43 PUs per drawer.

The z15 T01 feature Max190 is estimated to provide up to 25% more total system capacity than the z14 Model M05, with the same amount of memory and power requirements. With up to 40 TB of main storage and enhanced SMT, the performance of the z15 T01 processors deliver considerable improvement. Uniprocessor performance also increased significantly. A z15 Model 701 offers average performance improvements of up to 14%⁵ over the z14 Model 701.

The Integrated Facility for Linux (IFL) and IBM Z Integrated Information Processor (zIIP) processor units on the z15 server can be configured to run two simultaneous threads per clock cycle in a single processor (SMT). This feature increases the capacity of these

⁴ Simultaneous multithreading is two threads per core.

⁵ Observed performance increases vary depending on the workload types.

processors with 25% in average⁵ over processors that are running single thread. SMT is also enabled by default on System Assist Processors (SAPs).

The z15 T01 server expands the subcapacity settings, offering three subcapacity levels (in models 4xx, 5xx and 6xx) for up to 34 processors that are characterized as CPs (compared to up to 33 processors for z14). This configuration gives a total of 292 distinct capacity settings. The z15 servers deliver scalability and granularity to meet the needs of medium-sized enterprises, while also satisfying the requirements of large enterprises that have demanding, mission-critical transaction and data processing requirements.

This comparison is based on the Large System Performance Reference (LSPR) mixed workload analysis. For more information about performance and workload variation on z15 servers, see Chapter 12, “Performance” on page 475.

z15 servers continue to offer all specialty engine types that are available on z14.

Workload variability

Consult the LSPR when considering performance on z15 servers. The range of performance ratings across the individual LSPR workloads is likely to have a large spread. More performance variation of individual logical partitions (LPARs) is available when an increased number of partitions and more PUs are available. For more information, see Chapter 12, “Performance” on page 475.

For more information about performance, [see the LSPR website](#).

For more information about millions of service units (MSUs) ratings, see the [IBM Z Software Contracts](#) website.

Capacity on demand

Capacity on demand (CoD) enhancements enable clients to have more flexibility in managing and administering their temporary capacity requirements. The z15 server supports the same architectural approach for CoD offerings as the z14 (temporary or permanent). Within the z15 server, one or more flexible configuration definitions can be available to solve multiple temporary situations, and multiple capacity configurations can be active simultaneously.

Prepaid OoCoD tokens^a: Beginning with IBM z15, new prepaid OoCoD tokens that are purchased do not carry forward to future systems.

- a. IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

Up to 200 staged records can be created to handle many scenarios. Up to eight of these records can be installed on the server at any time. After the records are installed, the activation of the records can be done manually, or the z/OS Capacity Provisioning Manager can automatically start the activation when Workload Manager (WLM) policy thresholds are reached. Tokens are available that can be purchased for On/Off CoD before or after workload execution (pre- or post-paid).

LPAR capping

IBM Processor Resource/Systems Manager (IBM PR/SM) offers different options to limit the amount of capacity that is assigned to and used by an LPAR or a group of LPARs. By using the Hardware Management Console (HMC), a user can define an absolute or a relative capping value for LPARs that are running on the system.

1.2.3 Virtualization

This section describes built-in virtualization capabilities of z15 supporting operating systems, hypervisors, and available virtual appliances.

z15 servers support IBM z/Architecture® mode only, which can be initialized in LPAR mode (also known as PR/SM) or Dynamic Partition Manager (DPM) mode.

PR/SM mode

PR/SM is Licensed Internal Code (LIC) that manages and virtualizes all the installed and enabled system resources as a single large symmetric multiprocessor (SMP) system. This virtualization enables full sharing of the installed resources with high security and efficiency.

On z15 T01, the PR/SM supports configuring up to 85 LPARs, each of which includes logical processors, memory, and I/O resources. Resources of these LPARs are assigned from the installed CPC drawers and features. For more information about PR/SM functions, see 3.7, “Logical partitioning” on page 132.

LPAR configurations can be dynamically adjusted to optimize the virtual servers’ workloads. z15 servers provide improvements to the PR/SM HiperDispatch function. HiperDispatch provides alignment of logical processors to physical processors that ultimately improves cache utilization, minimizes inter-CPC drawer communication, and optimizes operating system work dispatching, which combined results in increased throughput. For more information, see “HiperDispatch” on page 96.

z15 PR/SM implements an Improved memory affinity algorithm and improved logical partition placement algorithms based on z14 experience. PR/SM reoptimization is the process of identifying “homes” for the partitions. PR/SM on z15 tries to assign all memory in one drawer (single SC SCM, shared L4 cache) and attempts to consolidate storage onto drawers with the most processor entitlement.

PR/SM also tries to assign all logical processors to one CPC drawer and packed into chips of that drawer, in cooperation with operating system use of HiperDispatch. In z15, all processor types can be dynamically reassigned, except IFPs.

Dynamic Partition Manager mode

DPM is an administrative mode (front end to PR/SM) that was introduced for Linux only systems for IBM z15, IBM z14, IBM z13®, IBM z13s®, and IBM LinuxONE servers. A system can be configured in DPM mode or in PR/SM mode (POR is required to switch modes). DPM supports the following functions:

- ▶ Create, provision, and manage partitions (processor, memory, and adapters)
- ▶ Monitor and troubleshoot the environment

HiperSockets

z15 servers support defining up to 32 IBM HiperSockets. *HiperSockets* provide for memory-to-memory communication across LPARs without the need for any I/O adapters and have virtual LAN (VLAN) capability.

LPAR modes on z15

The following PR/SM LPAR modes with corresponding operating systems and firmware appliances are supported:

- ▶ General:
 - z/OS

- IBM z/VM®
- IBM z/VSE®
- z/TPF
- Linux on IBM Z
- ▶ Coupling Facility: Coupling Facility Control Code (CFCC)
- ▶ Linux only:
 - Linux on IBM Z
 - z/VM
- ▶ z/VM
- ▶ Secure Service Container:
 - VNA (z/VSE Network Appliance)
 - IBM High Security Business Network (HSBN)⁶

The following LPAR modes are available for DPM:

- ▶ z/VM
- ▶ Linux on IBM Z (also used for KVM deployments)
- ▶ Secure Service Container

IBM Z servers also offer other virtual appliance-based solutions and support other the following hypervisors and containerization:

- ▶ IBM GDPS Virtual Appliance
- ▶ KVM for IBM Z
- ▶ Docker Enterprise Edition for Linux on IBM Systems⁷

Coupling Facility mode logical partition

Parallel Sysplex is a synergy between hardware and software, which is a highly advanced technology for clustering that is designed to enable the aggregate capacity of multiple z/OS systems to be applied against common workloads. To use this technology, a special LIC is used, which is called CFCC. To activate the CFCC, a special logical partition must be defined. Only PUs that are characterized as CPs or Internal Coupling Facilities (ICFs) can be used for Coupling Facility (CF) partitions. For a production CF workload, it is recommended to use dedicated ICFs.

The z/VM-mode LPAR

z15 servers support an LPAR mode, called *z/VM-mode*, that is exclusively for running z/VM as the first-level operating system. The z/VM-mode requires z/VM V6R4 or later, and allows z/VM to use a wider variety of specialty processors in a single LPAR, which increases flexibility and simplifying system management.

For example, in a z/VM-mode LPAR, z/VM can manage Linux on IBM Z guests that are running on IFL processors while also managing z/VSE and z/OS guests on CPs. It also allows z/OS to fully use zIIPs.

Secure Service Container

Secure Service Container (SSC) is an integrated IBM Z appliance and was designed to host most sensitive client workloads and applications, acting as a highly protected and secured digital vault, enforcing security by encrypting the whole stack: memory, network and data (in-flight and at-rest). An application that is running inside SSC (software appliance) is isolated and protected from outsider and insider threats.

⁶ IBM HSBN is a cloud service plan that is available on IBM Bluemix® for Blockchain.

⁷ For more information, see the [Running Docker Containers on IBM Z](#) topic of IBM Knowledge Center.

SSC combines hardware, software, and middleware and is unique to IBM Z platform. Although it is called a Container, it should not be confused with purely software Open Source containers (such as Kubernetes or Docker).

SSC is a part of the Pervasive Encryption concept that was introduced with IBM z14, which is aimed at delivering best IBM Security hardware and software enhancements, services, and practices for 360-degree infrastructure protection.

LPAR is defined as SSC by using Hardware Management Console (HMC).

The SSC solution offers the following advantages:

- ▶ Existing applications require zero changes to use SSC; software developers do not need to write any SSC-specific programming code.
- ▶ End-to-end encryption of data in-flight and at-rest:
 - Automatic Network Encryption (TLS, IPSEC); data in-flight.
 - Automatic File System Encryption (LUKS); data at-rest.
 - Linux Unified Key Setup (LUKS) is the standard way in Linux to provide disk encryption. SSC encrypt all data with a key that is stored within the appliance
 - Protected memory: Up to 16 TB can be defined per SSC LPAR.
- ▶ Encrypted Diagnostic Data
All diagnostic information (debug dump data, and logs) are encrypted and do not contain any user or application data.
- ▶ No operating system access
After the SSC appliance is built, Secure Shell (SSH) and command line-interface (CLI) are disabled, which guarantees that even system administrators cannot access the contents of SSC and do not know what application is running there.
- ▶ Applications that run inside SSC are accessed externally by REST APIs only, in a transparent to user way.
- ▶ Tamper-proof SSC Secure Boot
SSC- eligible applications are booted into SSC by using verified booting sequence, where only trusted and digitally signed and verified by IBM software code is uploaded into the SSC.
- ▶ Vertical workload isolation, which is certified by EAL5+ Common Criteria Standard, which is the highest level that ensures workload separation and isolation.
- ▶ Horizontal workload isolation, which is a separation from the rest of the host environment.

SSC is a powerful IBM technology for providing the extra protection of the most sensitive workloads. The integration with other applications is transparent; all services can be called externally by standard REST APIs.

GDPS Virtual Appliance

The GDPS Virtual Appliance solution implements GDPS/PPRC Multiplatform Resilience for IBM Z (xDR). xDR coordinates near-continuous availability and a disaster recovery (DR) solution through the following features:

- ▶ Disk error detection
- ▶ Heartbeat for smoke tests
- ▶ Re-IPL in place
- ▶ Coordinated site takeover

- ▶ Coordinated IBM HyperSwap®
- ▶ Single point of control

1.2.4 I/O subsystem and I/O features

The z15 server supports PCIe I/O infrastructure. PCIe features are installed in PCIe+ I/O drawers. Up to 12 PCIe+ I/O drawers per z15 T01 server are supported, which provides space for up to 192 PCIe I/O features. PCIe I/O drawers (32 slots, four I/O domains, available on z14 and z13) are not supported on z15 servers and cannot be carried forward during an upgrade.

For a five CPC drawer system, up to 60 PCIe+ fanout slots can be configured for data communications between the CPC drawers and the I/O infrastructure, and for coupling. The multiple channel subsystem (CSS) architecture allows up to six CSSs, each with 256 channels.

For I/O constraint relief, four subchannel sets are available per CSS, which allows access to many logical volumes. The fourth subchannel set allows extending the amount of addressable external storage for Parallel Access Volumes (PAVs), Peer-to-Peer Remote Copy (PPRC) secondary devices, and IBM FlashCopy® devices. z15 T01 supports Initial Program Load (IPL) from subchannel set 1 (SS1), subchannel set 2 (SS2), or subchannel set 3 (SS3), and subchannel set 0 (SS0). For more information, see “Initial program load from an alternative subchannel set” on page 208.

The system I/O buses use the Peripheral Component Interconnect® Express (PCIe) technology, which also is used in coupling links.

z15 T01 connectivity supports the following I/O or special purpose features:

- ▶ Storage connectivity:
 - Fibre Channel connection (IBM FICON):
 - FICON Express16SA 10 KM long wavelength (LX) and short wavelength (SX)
 - FICON Express16S+ 10 KM LX and SX (carry forward only)
 - FICON Express16S 10 KM LX and SX (carry forward only)
 - FICON Express8S 10 KM LX and SX (carry forward only)
 - IBM zHyperLink Express1.1 (new build)
 - IBM zHyperLink Express (carry forward)
- ▶ Network connectivity:
 - Open Systems Adapter (OSA):
 - OSA-Express7S 25 GbE Short Reach1.1 (new build)
 - OSA-Express7S 10 GbE long reach (LR) and short reach (SR) (new build)
 - OSA-Express7S GbE LX and SX (new build)
 - OSA-Express7S 1000BASE-T Ethernet (new build)
 - OSA-Express7S 25GbE SR (carry forward)
 - OSA-Express6S 10 GbE LR and SR (carry forward)
 - OSA-Express6S GbE LX and SX (carry forward)
 - OSA-Express6S 1000BASE-T Ethernet (carry forward)
 - OSA-Express5S 10 GbE LR and SR (carry forward)
 - OSA-Express5S GbE LX and SX (carry forward)
 - OSA-Express5S 1000BASE-T Ethernet (carry forward)
 - IBM HiperSockets
 - Shared Memory Communication - Remote Direct Memory Access (SMC-R):

- 25GbE RoCE (RDMA over Converged Ethernet) Express2.1 (new build)
- 25GbE RoCE (RDMA over Converged Ethernet) Express2 (carry forward)
- 10GbE RoCE Express2.1 (new build)
- 10GbE RoCE Express2 (carry forward)
- 10GbE RoCE Express (carry forward)
- Shared Memory Communication - Direct Memory Access (SMC-D) through Internal Shared Memory (ISM)
- ▶ Coupling and Server Time Protocol connectivity:
 - Internal Coupling (IC) links
 - Integrated Coupling Adapter Short Reach1.1 (ICA SR1.1 - new build)
 - Integrated Coupling Adapter Short Reach (ICA SR - carry forward)
 - CE LR (new build and carry forward)
- ▶ Cryptography:
 - Crypto-Express7S (new build)
 - Crypto-Express6S (carry forward)
 - Crypto-Express5S (carry forward)

1.2.5 Reliability, availability, and serviceability design

System reliability, availability, and serviceability (RAS) is an area of continuous IBM focus and a defining IBM Z platform characteristic. The RAS objective is to reduce, or eliminate if possible, all sources of planned and unplanned outages while providing adequate service information if an issue occurs. Adequate service information is required to determine the cause of an issue without the need to reproduce the context of an event.

IBM Z servers are designed to enable highest availability and lowest downtime. These facts are recognized by various IT analysts, such as ITIC⁸ and IDC⁹. Comprehensive, multi-layered strategy includes the following features:

- ▶ Error Prevention
- ▶ Error Detection and Correction
- ▶ Error Recovery
- ▶ System Recovery Boost

With a properly configured z15 server, further reduction of outages can be attained through First Failure Data Capture (FFDC), which is designed to reduce service times and avoid subsequent errors. It also improves nondisruptive replace, repair, and upgrade functions for memory, drawers, and I/O adapters. z15 servers support the nondisruptive download and installation of LIC updates.

IBM z15™ RAS features provide unique high-availability and nondisruptive operational capabilities that differentiate the Z servers in the marketplace. z15 RAS enhancements are made on many components of the CPC (processor chip, memory subsystem, I/O, and service) in areas, such as error checking, error protection, failure handling, error checking, faster repair capabilities, sparing, and cooling.

The ability to cluster multiple systems in a Parallel Sysplex takes the commercial strengths of the z/OS platform to higher levels of system management, scalable growth, and continuous availability.

⁸ For more information, see [ITIC Global Server Hardware, Server OS Reliability Report](#).

⁹ For more information, see [Quantifying the Business Value of IBM Z](#).

The z15 processor builds upon the RAS of the z14 family with the following RAS improvements:

▶ **System Recovery Boost**

System Recovery Boost was introduced with IBM z15. It offers customers more Central Processor (CP) capacity during system recovery operations to accelerate the system startup (IPL), shutdown or stand-alone dump operations. System Recovery Boost requires operating system support. No other IBM software charges are made during the boost period.

System Recovery Boost might be used during LPAR IPL or LPAR shutdown to make the running operating system and services available in a shorter period.

The System Recovery Boost provides the following options for the capacity increase:

- Subcapacity CP speed boost: During the boost period, subcapacity engines are transparently activated at their full capacity (CP engines).
- zIIP Capacity Boost: During the boost period, all active zIIPs that are assigned to an LPAR are used to extend the CP capacity (CP workload is dispatched to zIIP processors during the boost period).

At the time of this writing, the main System Recovery Boost users are z/OS (running in an LPAR), z/VM, and z/TPF.

z/VM uses the System Recovery Boost if it runs on subcapacity CP processors only (IFLs are always at their full clock speed). Second-level z/VM guest operating systems¹⁰ can inherit the boost if they are running on CPs.

- ▶ Level 3 and Level 4 cache enhancements use symbol ECC to extend the reach of prior z14 cache and memory improvements for improved availability. The level 3 and level 4 cache powerful symbol ECC is designed to make it resistant to more failure mechanisms. Preemptive DRAM marking is added to the main memory to isolate and recover failures more quickly.
- ▶ On-chip compression accelerating compression operations on core level, for all LPARs, which eliminates the virtualization layer that was needed for zEDC Express. The technology replaces zEDC Express PCIe cards, which improves reliability and availability.

1.3 z15 server technical overview

This section briefly reviews the following major elements of z15 servers:

- ▶ Model and features
- ▶ Model upgrade paths
- ▶ Frames
- ▶ CPC drawer
- ▶ I/O connectivity: PCIe+ Generation 3
- ▶ I/O subsystem
- ▶ Coupling and Server Time Protocol connectivity
- ▶ Cryptography
- ▶ IBM Integrated Accelerator for zEnterprise Data Compression

¹⁰ z/OS configured as a guest system under z/VM does not use the boost.

- ▶ Reliability, availability, and serviceability

1.3.1 Model and features

The IBM z15 Model T01 has a machine type of 8561 and is offered with five CPC features: Max34, Max71, Max108, Max145, and Max190. The feature name indicates the number of CPC drawers and available PUs, from one (Max34) to five (Max190).

Systems with up to four drawers have 41 active PUs per CPC drawer, while feature Max190 (five CPC drawers) has 43 active PUs per drawer. A PU is the generic term for the IBM z/Architecture processor unit (processor core) on the CP SCM.

On z15 servers, some PUs are part of the system base; that is, they are not part of the PUs that can be purchased by clients. They include the following characteristics:

- ▶ System assist processor (SAP) that is used by the channel subsystem. The number of predefined SAPs depends on the z15 model.
- ▶ One integrated firmware processor (IFP). The IFP is used in support of select features, such as and RoCE Express.
- ▶ Two spare PUs that can transparently assume any characterization during a permanent failure of another PU.

The PUs that clients can purchase can assume any of the following characteristics:

- ▶ CP for general-purpose use.
- ▶ Integrated Facility for Linux (IFL) for the use of Linux on Z.
- ▶ IBM Z Integrated Information Processor (zIIP) is designed to help free-up general computing capacity and lower overall total cost of computing for select data and transaction processing workloads.

zIIPs: At least one CP must be purchased with, or before, a zIIP can be purchased. Clients can purchase up to two zIIPs for each purchased CP (assigned or unassigned) on the system (2:1). However, during System Recovery Boost periods, the zIIP to CP ratio can be greater than 2:1.

- ▶ Internal Coupling Facility (ICF) is used by the CFCC.
- ▶ Extra (optional) SAPs are used by the channel subsystem.

A PU that is not characterized cannot be used, but is available as a spare. The following rules apply:

- ▶ In the five-feature structure, at least one CP, ICF, or IFL must be purchased and activated for any model.
- ▶ PUs can be purchased in single PU increments and are orderable by feature code.
- ▶ The total number of PUs purchased cannot exceed the total number that are available for that model.
- ▶ The number of installed zIIPs cannot exceed twice the number of installed CPs.

The multi-CPC drawer system design provides the capability to concurrently increase the capacity of the system in the following ways:

- ▶ Add capacity by concurrently activating more CPs, IFLs, ICFs, or zIIPs on a CPC drawer.
- ▶ Add a CPC drawer concurrently and activate more CPs, IFLs, ICFs, or zIIPs.

- ▶ Add a CPC drawer to provide more memory, or one or more adapters to support a larger number of I/O features.

1.3.2 Model upgrade paths

Within z15 Model T01, upgrades from Max34 to Max71 to Max108 are concurrent. Any z14 M0x or z13 model can be upgraded to any z15 T01 feature disruptively. Upgrades from z15 Model T01 features Max34, Max71 and Max108 to features Max145 and Max190 are not allowed as these features are factory build only. Figure 1-1 on page 14 shows the supported upgrade paths.

Note: Consider the following points:

- ▶ An air-cooled z15 T01 system *cannot* be converted to a water-cooled z15 T01 system, and vice versa.
- ▶ The z15 server *cannot* be part of an Ensemble that is managed by the Unified Resource Manager (zManager).

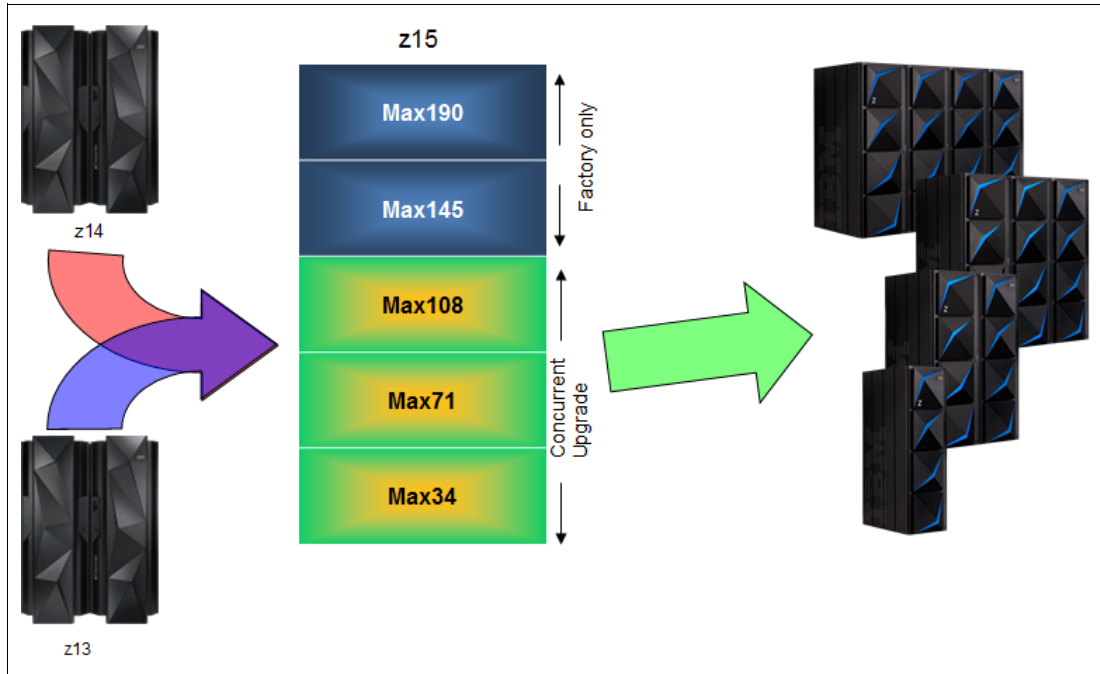


Figure 1-1 z15 T01 upgrades

z14 M0x upgrade to z15 T01

When a z14 M0x (M/T 3906) is upgraded to a z15, the z14 driver level must be at least 36. Upgrading from z14 to z15 system is disruptive (frame roll).

z13 upgrade to z15 T01

When a z13 (M/T 2964) is upgraded to a z15 T01, the z13 must be at least at Driver level 27. Upgrading from z13 to z15 T01 system is disruptive (frame roll).

The following processes are *not* supported:

- ▶ Downgrades within the z15 models
- ▶ Upgrade from a z13s or z14 ZR1 to z15 T01 systems

- ▶ Upgrades from zEC12 or earlier systems

1.3.3 Frames

The z15 Model T01 system is designed in a new format that provides configuration flexibility to fit customer requirements. The z15 T01 is build in a 19-inch form factor and can be configured with one, two, three, or four frames, depending on processor and I/O requirements.

Attention: The z15 system cannot be installed in a customer supplied frame. The system comes configured and build in an 19-inch IBM frame.

The z15 T01 also offers two power choices: Bulk Power Assembly (BPA) and Intelligent Power Distribution Units (iPDU, or PDU). The frames are A, B, C, and Z, bolted together, and feature the following components:

- ▶ Up to three CPC drawers in Frame A
- ▶ Up to two CPC drawers in Frame B (CPC drawers in frame B are factory-installed only)
- ▶ Up to 12 PCIe+ I/O drawers that hold I/O features and special purpose features
- ▶ All CPC drawers and PCIe+ I/O drawers have redundant power supplies
- ▶ For BPA systems only: Bulk Power Assemblies in Frames A and B with Optional Internal Battery Feature (IBF)
- ▶ For PDU systems only: Power Distribution Units in Frames A, B, and C (configuration dependant)
- ▶ CPC Drawer cooling units for either air or water cooling in Frames A and B
- ▶ Two Ethernet switches in Frame A and two in Frame B (configuration dependent) to interconnect the CPC components through Ethernet
- ▶ Two 1U rack-mounted Support Elements (mounted in Frame A). The Support Elements have a new service console (stored in Frame A), which can be connected in the front or rear of a system.

1.3.4 CPC drawer

Up to three CPC drawers are installed in frame A and up to two in Frame B of a z15 T01 server. Each CPC drawer houses the SCMs (PU and SC), memory, and I/O interconnects.

Single Chip Module technology

z15 T01 servers are built on the superscalar microprocessor architecture of its predecessor and provide various enhancements over the z14. Each CPC drawer has four PU (Processor Unit) SCMs arranged in two logical CP clusters, and one SC (System Controller) SCM.

Two CPC drawer sizes are available in z15 T01, depending on the number of active PU (cores). The z15 model T01 Max190 has 43 active cores (PUs) per CPC drawer. All other z15 models have 41 active cores. The PU SCM has 12 cores by design, with 9, 10, or 11 active cores, which can be characterized as CPs, IFLs, ICFs, zIIPs, SAPs, or IFPs.

The SCM provides a significant increase in system scalability and an extra opportunity for server consolidation. All CPC drawers are fully interconnected by using high-speed communication links through the L4 cache (in the SC SCM). This configuration allows the z15

server to be controlled by the PR/SM facility as a memory-coherent and cache-coherent SMP system.

The PU configuration includes two designated spare PUs per system and a variable number of SAPs. The SAPs scale with the number of CPC drawers that are installed in the server. For example, four standard SAPs are available with one CPC drawer that is installed, and up to 22 standard SAPs for five CPC drawers installed. In addition, one PU is used as an IFP and is not available for client use. The remaining PUs can be characterized as CPs, IFL processors, zIIPs, ICF processors, or extra SAPs. For z15, the SAPs operate in Simultaneous Multi-Threading (enabled by default, cannot be disabled).

The PU SCMs of the z15 T01 are cooled by a cold plate that is connected to an internal water cooling loop. In an air-cooled system, the radiator units (RUs) exchange the heat from the internal water loop with air. The RU has redundant pumps and blowers. The SC SCM is air-cooled.

The z15 T01 server offers also a water-cooling option (using data center chilled water supply) for increased system and data center energy efficiency. Water-cooling option is only available with Bulk Power Assembly (BPA)-based systems. The water cooling units (WCUs) are fully redundant in an N+1 arrangement.

Processor features

The processor core of the z15 T01 operates at 5.2 GHz. Depending on the z15 T01 feature, 41 - 215 active PUs are available on 1 - 5 CPC drawers.

Each core on the PU SCM includes an enhanced dedicated coprocessor for data compression and cryptographic functions, which are known as the Central Processor Assist for Cryptographic Functions (CPACF)¹¹. Having standard clear key cryptographic coprocessors that are integrated with the processor provides high-speed cryptography for protecting data.

The z15 supports 64-bit addressing mode and uses Complex Instruction Set Computer (CISC), including highly capable and thus complex instructions. Most of the instructions are implemented at the hardware or firmware level for most optimal and effective execution.

Each PU is a superscalar processor, which can decode up to six complex instructions per clock cycle, running instructions out-of-order. The PU uses a high-frequency, low-latency pipeline that provides robust performance across a wide range of workloads.

z/Architecture addressing modes: The z/Architecture simultaneously supports 24-bit, 31-bit, and 64-bit addressing modes. This feature provides compatibility with earlier versions and with that compatibility, investment protection.

Compared to its predecessor, the z15 processor design includes the following improvements and architectural extensions:

- ▶ Better performance and throughput:
 - Fast processor units with enhanced microarchitecture
 - Larger L2, L3 caches
 - Up to 12 cores per processor chip
 - More capacity (up to 190 characterizable processor units versus 170 on the z14)
 - Larger cache (and shorter path to cache) means faster uniprocessor performance

¹¹ Feature code (FC) 3863 must be ordered to enable CPACF. This feature code is available for no extra fee.

- Innovative core-cache design (L1 and L2 private to the processor core), processor chip-cache design (L3), and processor node design (L4), with focus on keeping more data closer to the processor, increasing the cache sizes, and decreasing the latency to access the next levels of cache.

This on-chip cache implementation optimizes system performance for high-frequency processors, with cache improvements, new Translation/TLB2 design, pipeline optimizations, better branch prediction, new accelerators, and architecture support.

- ▶ Reoptimized design for power and performance:
 - Improved instruction delivery
 - Improved branch prediction
 - Reduced execution latency
 - Optimized third-generation SMT
 - Enhanced out-of-order execution
 - New and enhanced vector instructions
- ▶ Dedicated co-processor for each processor unit (PU):
 - The Central Processor Assist for Cryptographic Function (CPACF) is well-suited for encrypting large amounts of data in real time because of its proximity to the processor unit.

CPACF supports DES, TDES, AES-128, AES-256, SHA-1, SHA-2, SHA-3, and True Random Number Generator. With the z15, CPACF supports Elliptic Curve Cryptography clear key, improving the performance of Elliptic Curve algorithms. The following algorithms are supported: EdDSA (Ed448, Ed25519), ECDSA (P-256, P-384, P-521), ECDH(P-256, P-384, P521, X25519, X448). Protected key signature creation is also supported.

- IBM Integrated Accelerator for zEDC (On-chip Compression): The z15 is enhancing the compression by taking it from the I/O device level (zEDC Express feature) and moving it to the Nest Accelerator Unit on the processor chip, adding the Deflate compliant (lossless data compression algorithm), and GZIP (GNU zip - UNIX compression utility) compression and decompression support as hardware instructions.

The IBM Integrated Accelerator for zEnterprise Data Compression (zEDC) provides on-chip compression (DEFLATE/gzip/zlib) services for all LPARs, whereas the zEDC Express PCIe feature was assignable to up to 15 LPARs only.

This innovation results in improved compression performance and simplified management (no need to manage the zEDC Express PCIe features), on a processor chip level, without any delays associated with I/O requests, and with minimal CPU costs. The enhancement preserves the compatibility with an earlier version with the data that is compressed by zEDC Express features; data, which is compressed and written by zEDC Express features, is read and decompressed on the z15 and vice versa. This simplifies the migration to the z15 (on-chip compression removes the need to acquire zEDC Express PCIe features).

Transactional Execution Facility

The Transactional Execution Facility, which is known in the industry as *hardware transactional memory*, allows instructions to be issued automatically. Therefore, *all results* of the instructions in the group are committed or *no results* are committed, in a truly transactional manner. The execution is optimistic.

The instructions are issued, but previous state values are saved in transactional memory. If the transaction succeeds, the saved values are discarded. If it fails, they are used to restore

the original values. Software can test the success of execution and rederive the code (if needed) by using the same or a different path.

The Transactional Execution Facility provides several instructions, including instructions to declare the start and end of a transaction and to cancel the transaction. This capability can provide performance benefits and scalability to workloads by helping to avoid most of the locks on data. This ability is especially important for heavily threaded applications, such as Java.

Guarded Storage Facility

Also known as less-pausing garbage collection, Guarded Storage Facility is a function that was introduced with the z14 to enable enterprise scale Java applications to run without periodic pause for garbage collection on larger heaps. This facility improves Java performance by reducing program pauses during Java Garbage Collection.

Simultaneous multithreading

Simultaneous multithreading (SMT) is built into the z15 IFLs, zIIPs, and SAPs, which allows more than one thread to simultaneously run in the same core and share all of its resources. This function improves the use of the cores and increases processing capacity.

When a program accesses a memory location that is not in the cache, it is called a *cache miss*. Because the processor must then wait for the data to be fetched before it can continue to run, cache misses affect the performance and capacity of the core to run instructions. By using SMT, when one thread in the core is waiting (such as for data to be fetched from the next cache levels or from main memory), the second thread in the core can use the shared resources rather than remain idle.

Adjusted with the growth in the core cache and TLB2, third-generation SMT on z15 improves thread balancing, supports multiple outstanding translations, optimizes hang avoidance mechanisms, and delivers improved virtualization performance to benefit Linux. z15 provides economies of scale with next generation multithreading (SMT) for Linux and zIIP-eligible workloads while adding support for the I/O System Assist Processor (SAP).

Hardware decimal floating point function

The hardware decimal floating point (HDFP) function is designed to speed up calculations and provide the precision that is demanded by financial institutions and others. The HDFP fully implements the IEEE 754r standard.

Vector Packed Decimal Facility

Vector Packed Decimal Facility allows packed decimal operations to be performed in registers rather than memory by using new fast mathematical computations. Compilers, such as Enterprise COBOL for z/OS, V6.2, Enterprise PL/I for z/OS, V5.2, z/OS V2.4 XL C/C++, the COBOL optimizer, Automatic Binary Optimizer for z/OS, V1.3, and Java, are optimized on z15.

Single instruction, multiple data

The z15 includes a set of instructions called single instruction, which is multiple data (SIMD) that can improve the performance of complex mathematical models and analytics workloads. This improvement is accomplished through vector processing and complex instructions that can process a large volume of data with a single instruction.

SIMD is designed for parallel computing and can accelerate code that contains integer, string, character, and floating point data types. This system enables better consolidation of analytics workloads and business transactions on the Z platform.

Runtime Instrumentation Facility

The Runtime Instrumentation Facility provides managed run times and just-in-time compilers with enhanced feedback about application behavior. This capability allows dynamic optimization of code generation as it is being run.

IBM Integrated Accelerator for zEnterprise Data Compression

With z15, a new on-chip compression accelerator (one per PU chip) was added to improve performance DEFLATE/gzip/zlib operations. The new accelerator replaces the zEDC Express feature and complements the functionality of the coprocessor (CPACF). Their functions are not interchangeable.

Processor RAS

In the unlikely case that a permanent core failure occurs, cores can be individually replaced by one of the available spares. Core sparing is transparent to the operating system and applications.

Concurrent processor unit conversions

The z15 supports concurrent conversion between various PU types, which provides the flexibility to meet the requirements of changing business environments. CPs, IFLs, zIIPs, ICFs, and optional SAPs can be converted to CPs, IFLs, zIIPs, ICFs, and optional SAPs.

Memory subsystem and topology

The z15 T01 systems use double data rate fourth-generation (DDR4) dual inline memory module (DIMM) technology. For this purpose, IBM developed a chip that controls communication with the PU, which is an SC chip, and drives address and control from DIMM-to-DIMM. The DIMM is available in 32 GB, 64 GB, 128 GB, 256 GB, and 512 GB capacities.

Memory topology provides the following benefits:

- ▶ A RAIM for protection at the dynamic random access memory (DRAM), DIMM, and memory channel levels.
- ▶ A maximum of 40 TB of user configurable memory with a maximum of 50 TB of physical memory (with a maximum of 16 TB configurable to a single LPAR).
- ▶ One memory port for each PU SCM, and up to five independent memory ports per CPC drawer.
- ▶ Increased bandwidth between memory and I/O.
- ▶ Asymmetrical memory size and DRAM technology across CPC drawers.
- ▶ Large memory pages (1 MB and 2 GB).
- ▶ Key storage.
- ▶ Storage protection key array that is kept in physical memory.
- ▶ Storage protection (memory) key that is also kept in every L2 and L3 cache directory entry.
- ▶ A larger (256 GB) fixed-size HSA that eliminates planning for HSA.

PCIe fanout hot-plug

The *PCIe fanout* slots provides the path for data between memory and the PCIe features through the PCIe+ Generation 3 feature dual 16 Gbps buses and cables. The PCIe+ fanout supports hot-pluggable features.

During an outage, a redundant I/O interconnect allows a PCIe+ Gen3 fanout feature to be concurrently repaired without loss of access to its associated I/O domains. Up to 12 PCIe+

fanout slots are available per CPC drawer. The PCIe fanout slots can also be used for the ICA SR. If redundancy in coupling link connectivity is ensured, the PCIe+ fanout features can be concurrently repaired.

1.3.5 I/O connectivity: PCIe+ Generation 3

The z15 T01 server offers new and improved I/O features and uses PCIe+ Gen3 for I/O connectivity. This section briefly reviews the most relevant I/O capabilities.

The z15 uses PCIe+ Gen3 to implement the following features:

- ▶ PCIe+ Gen3 fanouts implements dual 16 GBps connections to the PCIe I/O features in the PCIe+ I/O drawers
- ▶ PCIe fanouts that provide dual 8 GBps coupling link connections through the Integrated Coupling Adapter Short Reach1.1 (ICA SR1.1).

1.3.6 I/O subsystem

The z15 PU SCM I/O implements PCIe Generation 4, which is used to connect the PCIe+ Gen3 dual port fanout features in the CPC drawers. The I/O infrastructure is designed to reduce processor usage and I/O latency, and provide increased throughput and availability.

z15 servers offer PCIe+ I/O drawers that host PCIe features. PCIe I/O drawers and I/O drawers that were used in previous IBM Z servers are not supported on z15.

PCIe+ I/O drawer

The *PCIe+ I/O drawer*, together with the PCIe features, offers finer granularity and capacity over previous I/O infrastructures. It can be concurrently added and removed in the field, which eases planning. Only PCIe cards (features) are supported, in any combination. Up to 12 PCIe+ I/O drawers can be installed on a z15 T01 server.

Native PCIe and Integrated Firmware Processor

Native PCIe support was introduced in support of RoCE Express features, which are managed differently from the traditional PCIe I/O (FICON Express and OSA-Express) features. The device drivers for the native features are provided in the operating system. The diagnostic tests for the adapter layer functions of the native PCIe features are managed by LIC that is running as a resource group, which runs on the Integrated Firmware Processor (IFP).

With z15 (as with the z14), the number of PCIe resource groups is four, which helps mitigate the effect of the disruptive Resource Group Microcode Change Level (MCL) installations. This firmware management approach contributes to the RAS of the server.

During the ordering process of the native PCIe features, features of the same type are evenly spread across the four resource groups (RG1, RG2, RG3, and RG4) for availability and serviceability reasons. Resource groups are automatically activated when these features are present in the CPC.

In addition to the 10GbE RoCE Express features, the following native PCIe I/O features are also managed by the resource groups:

- ▶ Coupling Express Long Reach (CE LR)
- ▶ zHyperLink Express1.1 and zHyperLink Express
- ▶ RoCE Express2.1, RoCE Express2, and RoCE Express features.

1.3.7 I/O and special purpose features in the PCIe I/O drawer

The z15 T01 server (new build) supports the following PCIe features that are installed in the PCIe+ I/O drawers:

- ▶ Storage connectivity:
 - FICON Express16SA Short Wave (SX)
 - FICON Express16SA Long Wave (LX) 10 km (6.2 miles)
 - zHyperLink Express1.1
- ▶ Network connectivity:
 - OSA-Express7S 25GbE Short Reach (SR) 1.1
 - OSA-Express7S 10GbE Long Reach (LR)
 - OSA-Express7S 10GbE Short Reach (SR)
 - OSA-Express7S GbE LX
 - OSA-Express7S GbE SX
 - OSA-Express7S 1000BASE-T
 - 25GbE RoCE Express2.1
 - 10GbE RoCE Express2.1
- ▶ Coupling and Server Time Protocol connectivity: Coupling Express LR
- ▶ Cryptography:
 - Crypto-Express7S (2 port)
 - Crypto-Express7S (1 port)

When carried forward on an upgrade, the z15 T01 server also supports the following features in the PCIe+ I/O drawers:

- ▶ Storage connectivity:
 - FICON Express16S+ Short Wave (SX)
 - FICON Express16S+ Long Wave (LX) 10 km (6.2 miles)
 - zHyperLink Express
 - FICON Express 16S SX
 - FICON Express 16S LX
 - FICON Express 8S SX
 - FICON Express 8S LX
- ▶ Network connectivity:
 - OSA-Express7S 25GbE Short Reach (SR)
 - OSA-Express6S 10GbE Long Reach (LR)
 - OSA-Express6S 10GbE Short Reach (SR)
 - OSA-Express6S GbE LX
 - OSA-Express6S GbE SX
 - OSA-Express6S 1000BASE-T
 - 25GbE RoCE Express2
 - 10GbE RoCE Express2
 - OSA-Express5S 10 GbE Long Reach (LR)
 - OSA-Express5S 10 GbE Short Reach (SR)
 - OSA-Express5S GbE LX
 - OSA-Express5S GbE SX
 - OSA-Express5S 1000BASE-T
 - 10GbE RoCE Express
- ▶ Coupling and Server Time Protocol connectivity: Coupling Express LR
- ▶ Cryptography:

- Crypto-Express6S
- Crypto-Express5S

Although they are used for coupling connectivity, the IBM Integrated Coupling Adapter (ICA SR) features are not listed here because they are attached directly to the CPC drawer.

1.3.8 Storage connectivity

z15 T01 supports the IBM zHyperLink Express and FICON channels for storage connectivity.

Note: The IBM zHyperLink Express1.1 (FC 0451, new build for z15) and IBM zHyperLink Express (FC 0431, carry forward from z14) are functionally equivalent. Unless specified, these features are referred to as *zHyperLink Express* or *zHyperLink*.

IBM zHyperLink Express

zHyperLink Express is a short-distance mainframe attach link that is designed to increase the scalability of IBM Z transaction processing and lower I/O latency than High-Performance FICON (HPF) by for bringing data close to processing power.

zHyperLink Express feature directly connects the z15 central processor complex (CPC) to the I/O Bay of the DS8000 series (R8.5 and newer). This short distance of up to 150 m (492 feet) direct connection is intended to reduce I/O latency and improve storage I/O throughput.

The improved performance of zHyperLink Express allows the z15 PU to make a synchronous request for the data that is in the DS8880 cache. This feature eliminates the undispatch of the running request, the queuing delays to resume the request, and the PU cache disruption.

The IBM zHyperLink Express is a two-port feature in the PCIe+ I/O drawer. Up to 16 features with up to 32 zHyperLink Express ports are supported in a z15 CPC. The zHyperLink Express feature uses PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes for storage connectivity. It is designed to support a link data rate of 8 GigaBytes per second (GBps)¹².

The point-to-point link is established by 24x fiber optic cable with Multi-fiber Termination Push-on (MTP) connectors. For more information, see “zHyperLink Express1.1 (FC 0451)” on page 179.

FICON channels

Up to 160 features with up to 320 FICON Express16SA channels are supported on a new build z15. FICON Express 16SA supports 8 or 16 Gbps data link rate (NO 4 Gbps support). FICON Exptess16S+ and FICON Express 16S (both carry forward only) support link data rates of 4, 8, or 16 Gbps. FICON Express8S features (carry forward only) support link data rates of 2, 4, or 8 Gbps.

FICON Express16SA offers the same performance as FICON Express16S+ with its IBM I/O ASIC that supports up to 3x the I/O start rate of previous FICON/FCP solutions. As with the FICON Express16S+, both ports of a feature must be defined as the same CHPID type (no mix of FC and FCP CHPID for the same feature).

¹² The link data rates do not represent the performance of the links. The actual performance is dependent upon many factors, including latency through the adapters, cable lengths, and the type of workload.

The FICON features on z15 support the following protocols:

- ▶ FICON (CHPID type FC) and High-Performance FICON for Z (zHPF). zHPF offers improved performance for data access, which is important to online transaction processing (OLTP) applications.
- ▶ FICON channel-to-channel (CHPID type CTC).
- ▶ Fibre Channel Protocol (CHPID type FCP).

FICON also offers the following capabilities:

- ▶ Modified Indirect Data Address Word (MIDAW) facility: Provides more capacity over native FICON channels for programs that process data sets that use striping and compression, such as Db2, VSAM, partitioned data set extended (PDSE), hierarchical file system (HFS), and z/OS file system (zFS). It does so by reducing channel, director, and control unit processor usage.
- ▶ Enhanced problem determination, analysis, and manageability of the storage area network (SAN) by providing registration information to the fabric name server for FICON and FCP.
- ▶ An Extended Link Service command, Read Diagnostic Parameters (RDP) is used to obtain extra diagnostic data from the Small Form Factor Pluggable optics that are throughout the SAN fabric to improve the accuracy of identifying a failing component.

1.3.9 Network connectivity

The IBM z15 T01 supports the following technologies for network connectivity:

- ▶ Open Systems Adapter (OSA)
- ▶ HiperSockets
- ▶ Shared Memory Communication - Remote Direct Memory Access over Converged Ethernet (SMC-R)
- ▶ Shared Memory Communications - Direct Memory Access over Internal Shared Memory (SMC-D)

Open Systems Adapter

z15 T01 allows any mix of the supported OSA Ethernet features that are listed in 1.3.7, “I/O and special purpose features in the PCIe I/O drawer” on page 21. OSA-Express7S features are a technology refresh of the OSA-Express6S features. Up to 48 OSA-Express7S features, with a maximum of 96 ports, are supported. The maximum number of combined OSA-Express features cannot exceed 48.

OSA-Express features provide important benefits for TCP/IP traffic by reducing latency and improving throughput for standard and jumbo frames. Data router function that is present in all OSA-Express features enables performance enhancements.

With OSA-Express7S, OSA-Express6S, and OSA-Express5S, the functions that were performed in firmware are performed in the hardware. Extra logic in the IBM application-specific integrated circuit (ASIC) that is included with these features handle packet construction, inspection, and routing, which allows packets to flow between host memory and the LAN at line speed without firmware intervention.

On z15, an OSA feature that is configured as an integrated console controller CHPID type (OSC) supports the configuration and enablement of secure connections by using the Transport Layer Security (TLS) protocol versions 1.0, 1.1, and 1.2.

For more information about the OSA features, see 4.6, “Connectivity” on page 167.

HiperSockets

The HiperSockets function (also known as *internal queued direct input/output* or internal QDIO or iQDIO) is an integrated function of the z15 server that provides users with attachments to up to 32 high-speed virtual LANs with minimal system and network processor usage.

For communications between LPARs in the same z15 server, HiperSockets eliminate the need to use I/O subsystem features to traverse an external network. Connection to HiperSockets offers significant value in server consolidation by connecting many virtual servers.

HiperSockets can be customized to accommodate varying traffic sizes. Because the HiperSockets function does not use an external network, it can free system and network resources, which eliminates attachment costs while improving availability and performance.

HiperSockets can also be used for Dynamic cross-system coupling, which is a z/OS Communications Server feature that creates trusted, internal links to other stacks within a Parallel Sysplex.

Shared Memory Communication - Remote Direct Memory Access

zEC12 GA2 was the first IBM Z server generation to support Remote Direct Memory Access over Converged Ethernet (RoCE) technology. This technology is designed to provide fast, reduced CPU consumption and memory-to-memory communications between two IBM Z CPCs.

RoCE Express features reduce CPU consumption for applications that use the TCP/IP stack (sockets communication), such as IBM WebSphere Application Server that accesses a Db2 database. It is transparent to applications and also might help to reduce network latency with memory-to-memory transfers that use SMC-R in supported z/OS releases and Linux on Z.

IBM Z server generations continue to enhance the RoCE architecture. The 10GbE RoCE Express feature (carry forward only) supports sharing among 31 LPARs running z/OS or Linux on Z, while the RoCE Express2 and RoCE Express2.1 (10 GbE and 25 GbE) support 4x the number of LPARs and performance improvements. RoCE Express2 and RoCE Express2.1 support 63 Virtual Functions (VFs) per port for up to 126 VFs per PCHID (physical channel ID).

The 10GbE RoCE Express2, 10GbE RoCE Express2.1, and 10GbE RoCE Express features use SR optics and support the use of a multimode fiber optic cable that ends with an LC Duplex connector. Both support point-to-point and switched connections with an enterprise-class 10 GbE switch. A maximum of eight RoCE Express features can be installed in PCIe+ I/O drawers of z15.

The 25GbE RoCE Express2 and 25GbE RoCE Express2.1 also feature SR optics and supports the use of 50-micron multimode fiber optic that ends with an LC duplex connector. These features support point-to-point and switched connections with 25GbE capable switch (support only for 25 Gbps, no down negotiation to 10 Gbps).

Shared Memory Communications - Direct Memory Access

SMC-D enables low processor usage and low latency communications within a CPC that uses a Direct Memory Access connection over ISM. SMC-D implementation is similar to SMC-R over RoCE; SMC-D over ISM extends the benefits of SMC-R to operating system

instances that are running on the same CPC without requiring physical resources (RoCE adapters, PCI bandwidth, ports, I/O slots, network resources, and 25/10 GbE switches).

Introduced with z13 GA2 and z13s, SMC-D enables high-bandwidth LPAR-to-LPAR TCP/IP traffic (sockets communication) by using the direct memory access software protocols over virtual Internal Shared Memory PCIe devices (vPCIe). SMC-D maintains the socket-API transparency aspect of SMC-R so that applications that use TCP/IP communications can benefit immediately without requiring any application software or IP topology changes.

z15 continues to support SMC-D with its lightweight design that improves throughput, latency, and CPU consumption and complements HiperSockets, OSA, or RoCE without sacrificing quality of service.

SMC-D requires an OSA or a HiperSockets connection to establish the initial TCP communications and can coexist with them. SMC-D uses a virtual PCIe adapter and is configured as a physical PCIe device. Up to 32 ISM adapters are available, each with a unique Physical Network ID per CPC.

Notes: SMC-R and SMC-D do not currently support multiple IP subnets.

1.3.10 Coupling and Server Time Protocol connectivity

IBM z15 support for Parallel Sysplex includes the Coupling Facility (running the CFCC¹³) and coupling links.

Coupling links support

Coupling connectivity in support of Parallel Sysplex environments is provided on the z15 server by the following features:

- ▶ Internal Coupling (IC) links that are operating at memory speed.
- ▶ Integrated Coupling Adapter Short Reach1.11.1
- ▶ Integrated Coupling Adapter Short Reach1.1
- ▶ Coupling Express Long Reach

All physical coupling link types can be used to carry STP messages.

Integrated Coupling Adapter Short Reach1.1

The Integrated Coupling Adapter Short Reach1.1 (ICA SR1.1) feature was refreshed for z15. It is a two-port fanout that is used for short distance coupling connectivity. It uses PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes for coupling.

The ICA SR1.1 is designed to drive distances up to 150 m and support a link data rate of 8 GBps. The ICA SR1.1 fanout takes one PCIe fanout slot in the z15 CPC drawer. It is used for coupling connectivity between z15, z14, z13, and z13s CPCs, and cannot be connected to HCA3-O or HCA3-O LR coupling fanouts. The ICA SR1.1 is compatible with another ICA SR1.1 or ICA SR only.

Integrated Coupling Adapter Short Reach

The Integrated Coupling Adapter Short Reach (ICA SR) feature (introduced with IBM z13) is a two-port fanout that is used for short distance coupling connectivity and can be carried forward to z15. It uses PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes for coupling.

¹³ CFCC - Coupling Facility Control Code

The ICA SR is designed to drive distances up to 150 m (492 feet) and support a link data rate of 8 GBps. The ICA SR fanout takes one PCIe I/O fanout slot in the z15 CPC drawer. It is used for coupling connectivity between z15, z14, z13, and z13s CPCs, and cannot be connected to HCA3-O or HCA3-O LR coupling fanouts. The ICA SR is compatible with another ICA SR or ICA SR 1.1 only.

Coupling Express Long Reach

Coupling Express Long Range (CE LR), which is a two-port feature is used for point-to-point long-distance coupling connectivity and defined as coupling channel type, CL5. The CE LR link is plugged in a PCIe+ I/O drawer slot, which uses industry standard I/O technology. It is used for long-distance coupling connectivity between z15, z14, z13, and z13s CPCs. It is not compatible with 1x InfiniBand (HCA3O-LR or HCA2O-LR) features.

The CE LR link allows for more granularity when scaling up or completing maintenance and uses Single Mode fiber (similar to InfiniBand 1x coupling links). The CE LR link provides point-to-point coupling connectivity at distances of 10 km (6.2 miles) unrepeated and 100 km (62.1 miles) with a qualified dense wavelength division multiplexing (DWDM) device.

CFCC Level 24

CFCC level 24 is delivered on the z15 with driver level 41. CFCC Level 24 introduces the following enhancements:

- ▶ CFCC Fair Latch Manager2
- ▶ Message Path SYID Resiliency Enhancement
- ▶ Shared-Engine CF Default is changed to “DYNDISP=THIN”
- ▶ Coupling Facility monopolization avoidance

z15 servers with CFCC Level 24 require z/OS V2R1 or later, and z/VM V6R4 or later for virtual guest coupling. For more information, see “**Coupling Facility Enhancements with CFCC level 24**” on page 118.

Although the CF LPARs are running on different server generations, different levels of CFCC can coexist in the same sysplex, which enables upgrade from one CFCC level to the next. CF LPARs that are running on the same server share a single CFCC level.

A CF running on a z15 server (CFCC level 24) can coexist in a sysplex with CFCC levels 23, 22, 21 and 20. For more information about determining the CF LPAR size by using the CFSizer tool, see the [System z Coupling Facility Structure Sizer Tool web page](#).

Server Time Protocol facility

Time synchronization for Parallel Sysplex Server Time Protocol (STP) is designed to ensure events that occur in different servers are properly sequenced in time. STP is designed for servers that are configured in a Parallel Sysplex or a basic sysplex (without a CF), and servers that are not in a sysplex but need time synchronization.

STP is a server-wide facility that is implemented in the LIC, which presents a single view of time to PR/SM. Any IBM Z CPC (including CPCs that are running as stand-alone CFs) can be enabled for STP by installing the STP feature.

STP uses a message-based protocol in which timekeeping information is passed over externally defined coupling links between servers. The STP design introduced a concept that is called Coordinated Timing Network (CTN), which is a collection of servers and CFs that are time-synchronized to a time value called Coordinated Server Time (CST).

Network Time Protocol as External Time Source

Network Time Protocol (NTP) client support is available to the STP code on the z15, z14, z13, and z13s servers. By using this function, these servers can be configured to use an NTP server as an External Time Source (ETS). This implementation fulfills the need for a single time source across the heterogeneous platforms in the enterprise, including IBM Z servers and others systems that are running Linux, UNIX, and Microsoft Windows operating systems.

HMC can be configured as an NTP client or an NTP server. To ensure secure connectivity, HMC NTP broadband authentication can be enabled on z15, z14, and z13 servers.

The time accuracy of an STP-only CTN can be improved by using an NTP server with the pulse per second (PPS) output signal as ETS. This type of ETS is available from various vendors that offer network timing solutions.

Precision Time Protocol as External Time Source

For IBM z15 Model T01 (and available on z15 T02 as well), Precision Time Protocol (IEEE 1588) can also be used as External Time Source for STP (replacing NTP). Pulse per second (PPS) is required at this time for providing the highest timing accuracy.

Attention: A z15 server can coexist in the same (STP-only) CTN with z15, z14, and z13 servers. No older servers are supported in the same CTN with z15.

The z15 supports coupling and timing connectivity by using Integrated Coupling Adapter Short Reach (ICA SR) and Coupling Express Long Reach (CE LR). No InfiniBand connectivity is supported on the z15. For STP role playing servers, planning must consider only ICA SR and CE LR coupling or timing links.

1.3.11 Cryptography

A strong synergy exists between cryptography and security. Cryptography provides the primitives to support security functions. Similarly, security functions help to ensure authorized use of key material and cryptographic functions.

Cryptography on IBM Z is built on the platform with integrity. IBM Z platform offers hardware-based cryptography features that are used by the following environments and functions:

- ▶ Java
- ▶ Db2/IMS encryption tool
- ▶ Db2 built in encryption z/OS Communication Server
- ▶ IPsec/IKE/AT-TLS
- ▶ z/OS System SSL
- ▶ z/OS
- ▶ z/OS Encryption Facility
- ▶ Linux on Z
- ▶ CP Assist for Cryptographic Functions
- ▶ Crypto Express7S
- ▶ Crypto-Express6S
- ▶ Trusted Key Entry workstation

CP Assist for Cryptographic Functions

Supporting clear and protected key encryption, CP Assist for Cryptographic Function (CPACF) offers the full complement of the Advanced Encryption Standard (AES) algorithm and Secure Hash Algorithm (SHA) with the Data Encryption Standard (DES) algorithm.

Support for CPACF is available through a group of instructions that are known as the Message-Security Assist (MSA).

z/OS Integrated Cryptographic Service Facility (ICSF) callable services and the z90crypt device driver that is running on Linux on Z also start CPACF functions. ICSF is a base element of z/OS. It uses the available cryptographic functions, CPACF, or PCIe cryptographic features to balance the workload and help address the bandwidth requirements of your applications.

With z15, a new Elliptic Curve Cryptography-supporting Modulo Arithmetic unit was implemented on each PU (core) along with a new message security assist extension 9 and Elliptic Curve Signature Authentication (ECSA) instruction. This provides hardware support for verification and signing using NIST P256, P384, P521 curves, Ed25519, Ed448-Goldilocks curves. The expected use cases include SSL libraries (authentication on the web) and Blockchain cryptography.

With z14 and carried on to z15, CPACF is enhanced to support pervasive encryption to provide faster encryption and decryption than previous servers. For every Processor Unit that is defined as a CP or an IFL, the following benefits are realized:

- ▶ Reduced overhead on short data (hashing and encryption)
- ▶ Up to 4x throughput for AES compared to z13
- ▶ Special instructions for elliptic curve cryptography (ECC)/RSA
- ▶ New hashing algorithms (for example, SHA-3)
- ▶ Support for authenticated encryption (combined encryption and hashing; for example, AES-GCM)
- ▶ True random number generator (for example, for session keys)

The z13 CPACF provides (supported by z15 and z14 also) the following features:

- ▶ For data privacy and confidentiality: DES, Triple Data Encryption Standard (TDES), and AES for 128-bit, 192-bit, and 256-bit keys.
- ▶ For data integrity: Secure Hash Algorithm-1 (SHA-1) 160-bit, and SHA-2 for 224-, 256-, 384-, and 512-bit support. SHA-1 and SHA-2 are included as enabled on all z14s and do not require the no-charge enablement feature.
- ▶ For key generation: Pseudo Random Number Generation (PRNG), Random Number Generation Long (RNGL) (1 - 8192 bytes), and Random Number Generation Long (RNG) with up to 4096-bit key RSA support for message authentication.

CPACF must be explicitly enabled by using a no-charge enablement feature (FC 3863). This requirement excludes the SHAs, which are enabled by default with each server.

The enhancements to CPACF are exclusive to the IBM Z servers and are supported by z/OS, z/VM, z/VSE, z/TPF, and Linux on Z.

Crypto Express7S

Crypto Express7S represents the newest generation of cryptographic features. Cryptographic performance improvements with new Crypto Express7S, which is available with two features, with one (FC 0899) or two (FC 0898) cryptographic co-processors¹⁴ that allow more data to be securely transferred across the internet. Crypto Express7S is designed to complement the cryptographic capabilities of the CPACF. It is an optional feature of the z15 server generation.

¹⁴ The IBM PCIe Cryptographic Co-processor (PCIeCC) is implemented as a Hardware Security Module (HSM).

The Crypto Express7S feature is designed to provide granularity for increased flexibility with one or two PCIe adapters per feature. Although installed in the PCIe+ I/O drawer, Crypto Express7S features do not perform I/O operations. That is, no data is moved between the CPC and any externally attached devices. For availability reasons, a minimum of two features is required.

z15 T01 servers allow sharing of a cryptographic coprocessor across 85 domains (the maximum number of LPARs on the system for z15 T01 is 85).

Crypto Express7S provides higher density of the PCIe cryptographic features with enhanced PCIe Cryptographic Coprocessor (IBM 4769) and carries forward the functionality of the Crypto Express6S, described in the following paragraph.

Crypto-Express6S

Crypto-Express6S (FC #0893) introduced with z14 server generation can be carried forward to z15. For availability reasons, a minimum of two features is required.

z15 servers allow sharing of a cryptographic coprocessor across 85 domains (the maximum number of LPARs on the system for z15 is 85).

The Crypto-Express6S is a state-of-the-art, tamper-sensing, and tamper-responding programmable cryptographic feature that provides a secure cryptographic environment. Each adapter contains a tamper-resistant hardware security module (HSM). The HSM can be configured as a Secure IBM CCA coprocessor, as a Secure IBM Enterprise PKCS #11 (EP11) coprocessor, or as an accelerator. Consider the following points:

- ▶ A Secure IBM CCA coprocessor is for secure key encrypted transactions that use CCA callable services (default).
- ▶ A Secure IBM Enterprise PKCS #11 (EP11) coprocessor implements an industry standardized set of services that adhere to the PKCS #11 specification v2.20 and more recent amendments. This new cryptographic coprocessor mode introduced the PKCS #11 secure key function.
- ▶ An accelerator for public key and private key cryptographic operations is used with Secure Sockets Layer/Transport Layer Security (SSL/TLS) acceleration.

The Crypto-Express6S is designed to meet the following cryptographic standards, among others:

- ▶ FIPS 140-2 Level 4¹⁵
- ▶ Common Criteria EP11 EAL4
- ▶ ANSI 9.97
- ▶ Payment Card Industry (PCI) HSM
- ▶ German Banking Industry Commission (GBIC), (formerly DK, Deutsche Kreditwirtschaft)

Federal Information Processing Standard (FIPS) 140-2 certification is supported only when Crypto-Express6S is configured as a CCA or an EP11 coprocessor.

Crypto-Express6S supports several ciphers and standards that are described next. For more information about cryptographic algorithms and standards, see Chapter 6, “Cryptographic features” on page 215.

¹⁵ Federal Information Processing Standard (FIPS) 140-2 Security Requirements for Cryptographic Modules

Trusted Key Entry workstation

The Trusted Key Entry (TKE) feature is an integrated solution that is composed of workstation firmware, hardware, and software to manage cryptographic keys in a secure environment. The TKE is network-connected or isolated, in which case smart cards are used.

The TKE workstation offers a security-rich solution for basic local and remote key management. It provides authorized personnel with a method for key identification, exchange, separation, update, and backup, and a secure hardware-based key loading mechanism for operational and master keys. TKE also provides secure management of host cryptographic module and host capabilities.

Support for an optional smart card reader that is attached to the TKE workstation allows the use of smart cards that contain an embedded microprocessor and associated memory for data storage. Access to and the use of confidential data on the smart cards are protected by a user-defined personal identification number (PIN).

TKE workstation and the most recent TKE 9.2 LIC are optional features on the z15. TKE workstation is offered in two types: TKE Tower (FC 0088) and TKE Rack Mount (FC 0087).

TKE 9.x¹⁶ requires the crypto-adapter FC 4769. You can use an older TKE version to collect data from previous generations of cryptographic modules and apply the data to Crypto-Express7S and Crypto-Express6S coprocessors.

TKE 9.x is required if you choose to use the TKE to manage a Crypto-Express7S. TKE 9.1 and later is also required to manage the new CCA mode PCI-HSM settings that are available on the Crypto-Express6S and Crypto-Express7S. A TKE is required to manage any Crypto-Express feature that is running in IBM Enterprise PKCS #11 (EP11) mode. If EP11 is to be defined, smart cards that are used require FIPS certification.

For more information about the cryptographic features, see Chapter 6, “Cryptographic features” on page 215.

For more information about the most current ICSF updates that are available, see [the Web Deliverables download web page](#).

1.4 Reliability, availability, and serviceability

The z15 RAS strategy uses a building block approach, which is developed to meet the client’s stringent requirements for achieving continuous reliable operation. Those building blocks are error prevention, error detection, recovery, problem determination, service structure, change management, measurement, and analysis.

The initial focus is on preventing failures from occurring. This goal is accomplished by using Hi-Rel (highest reliability) components that use screening, sorting, burn-in, and run-in, and by taking advantage of technology integration.

For LIC and hardware design, failures are reduced through rigorous design rules; design walk-through; peer reviews; element, subsystem, and system simulation; and extensive engineering and manufacturing testing.

The RAS strategy is focused on a recovery design to mask errors and make them transparent to client operations. An extensive hardware recovery design is implemented to detect and

¹⁶ TKE 9.0 LIC, TKE 9.1 LIC, and TKE 9.2 LIC have the same hardware requirements. TKE 9.0 LIC and 9.1 LIC can be upgraded to TKE 9.2 LIC.

correct memory array faults. When transparency cannot be achieved, you can restart the server with the maximum capacity possible.

System Recovery Boost (see also Appendix B, “System Recovery Boost” on page 493) is a new function that is implemented in the z15 firmware that brings a new dimension to the overall RAS approach. System Recovery Boost is designed to provide more CP capacity to LPARs running on a z15 CPC, capacity that is used for boosting (speeding up) operations during maintenance periods, such as system IPL, system shutdown, and stand-alone dumps.

For more information, see Chapter 9, “Reliability, availability, and serviceability” on page 383.

1.5 Hardware Management Consoles and Support Elements

The HMCs and SEs are appliances that together provide platform management for IBM Z. The HMC is a workstation that is designed to provide a single point of control for managing local or remote hardware elements.

HMC is offered as a Tower (FC 0062) and a Rack Mount (FC 0063) feature. Rack Mount HMC can be placed in a customer-supplied 19-inch rack and occupies 1U rack space. z15 includes Driver level 41 and HMC application Version 2.15.0.

IBM z15 also introduces a new management feature, the IBM Hardware Management Appliance (FC 0100). With this feature, the need for a stand-alone HMC is eliminated, both Hardware Management Console appliance and Support Element Appliance running virtualized on the Support Element hardware servers integrated with the z15 CPC.

For more information, see Chapter 10, “Hardware Management Console and Support Element” on page 411.

1.6 Operating systems

The IBM z15 server is supported by a large set of software products and programs, including independent software vendor (ISV) applications. (This section lists only the supported operating systems.) Use of various features might require the latest releases. For more information, see Chapter 7, “Operating system support” on page 253.

1.6.1 Supported operating systems

The following operating systems with required maintenance applied are supported for z15 servers:

- ▶ z/OS:
 - Version 2 Release 4
 - Version 2 Release 3
 - Version 2 Release 2
 - Version 2 Release 1 (compatibility support only, with extended support agreement)
- ▶ z/VM:
 - Version 7 Release 2¹⁷
 - Version 7 Release 1
 - Version 6 Release 4

¹⁷ Announced on April 14, 2020.

- ▶ z/VSE: Version 6 Release 2
- ▶ z/TPF Version 1 Release 1
- ▶ Linux on IBM Z distributions¹⁸:
 - SUSE SLES 15 SP1 with service, SUSE SLES 12 SP4 with service, and SUSE SLES 11 SP4 with service.
 - Red Hat RHEL 8.0 with service, Red Hat RHEL 7.7 with service, and Red Hat RHEL 6.10 with service.
 - Ubuntu 18.04.1 LTS with service and Ubuntu 16.04.5 LTS with service.
- ▶ KVM: Supported by Linux on IBM Z distributions

For more information about supported Linux on Z distribution levels, see the [Tested platforms for Linux page](#) of the IBM Z website.

For more information about features and functions that are supported on z14 by operating system, see Chapter 7, “Operating system support” on page 253.

z/VM support

z/VM 7.2 (Announced April 14, 2020, planned availability Sept. 2020) brings, in addition to features and functions included with z/VM 7.1 and subsequent 7.1 PTFs, the following new functionality:

- ▶ Centralized Service Management for non-SSI environments to deploy service to multiple systems, regardless of geographic location, from a centralized primary location.
- ▶ Multiple Subchannel Set (MSS) Multi-Target Peer-To-Peer Remote Copy (MT-PPRC) z/VM support for the GDPS environment, allowing a device to be the primary to up to three secondary devices, each defined in a separate alternate subchannel set (supporting up to 3 alternate subchannel sets). Also provides the CP updates necessary for VM/HCD support of alternate subchannel sets.
- ▶ New Architecture Level Set of z13, z13s, or newer processor families

z/VM 7.1 (available as of September 2018) increases the level of engagement with the z/VM user community. z/VM 7.1 includes the following new features:

- ▶ Single System Image and Live Guest Relocation included in the base (no extra charge).
- ▶ Enhances the dump process to reduce the time that is required to create and process dumps.
- ▶ Upgrades to a new Architecture Level Set (requires an IBM zEnterprise EC12 or BC12, or later).
- ▶ Provides the base for more functionality to be delivered as service after general availability.
- ▶ Enhances the dynamic configuration capabilities of a running z/VM system with Dynamic Memory Downgrade* support. For more information, [see this web page](#).
- ▶ Includes SPE¹⁹s shipped for z/VM 6.4, including Virtual Switch Enhanced Load Balancing, DS8000 z-Thin Provisioning, and Encrypted Paging.

To support new functionality that was announced October 2018, z/VM requires fixes for the following APARs:

- ▶ PI99085
- ▶ VM66130

¹⁸ Customers should monitor for new distribution releases supported.

¹⁹ Small Program Enhancements, part of the continuous delivery model, see <http://www.vm.ibm.com/newfunction/>

- ▶ VM65598
- ▶ VM66179
- ▶ VM66180

Support for z15 is provided in z/VM 7.1 and 6.4 with fixes for the following APARs:

- ▶ VM66248: z15 processor compatibility support (including Crypto-Express7S)
- ▶ PI99085: TCP/IP Stack support for OSA-Express7S
- ▶ VM66239: VMHCD support
- ▶ VM65598: VMHCM support
- ▶ PH00902: HLASM support
- ▶ VM66240: IOCP support.

For more information about the features and functions that are supported on z14 by operating system, see Chapter 7, “Operating system support” on page 253.

z/OS support

z/OS uses many of the following new functions and features of z14 (depending on version and release; PTFs might be required to support new functions):

- ▶ Up to 190 processors per LPAR or up to 128 physical processors per LPAR in SMT mode (SMT for zIIP)
- ▶ Up to 16 TB of real memory per LPAR (4 TB maximum for z/OS)
- ▶ Two-way simultaneous multithreading (SMT) optimization and support of SAPs (SAP SMT enabled by default) in addition to zIIP engines
- ▶ XL C/C++ ARCH(13) and TUNE(13) compiler options
- ▶ Use of faster CPACF
- ▶ Pervasive Encryption:
 - Coupling Facility Encryption
 - Dataset and network encryption
- ▶ HiperDispatch Enhancements
- ▶ On-chip compression accelerator (transparent zEDC replacement)
- ▶ z15 Hardware Instrumentation Services (HIS)
- ▶ Entropy-Encoding Compression Enhancements
- ▶ Guarded Storage Facility (GSF)
- ▶ Instruction Execution Protection (IEP)
- ▶ IBM Virtual Flash Memory (VFM)
- ▶ Improved memory management in Real Storage Manager (RSM)
- ▶ CF use of VFM: CFCC Level 24
- ▶ Coupling Express Long Reach (CE LR) CHPID type CL5
- ▶ zHyperLink Express1.1
- ▶ FICON Express16SA; OSA Express7S
- ▶ RoCE-Express2.1 (25GbE and 10GbE)
- ▶ Cryptography:
 - Crypto-Express7S:
 - Next Generation Coprocessor support
 - Support for Coprocessor in PCI-HSM Compliance Mode

- Designed for up to 85 domains
- TKE 9.2 workstation

For more information about the features and functions that are supported on z15 by operating system, see Chapter 7, “Operating system support” on page 253.

1.6.2 IBM compilers

The following IBM compilers for Z servers can use z15 servers:

- ▶ Enterprise COBOL for z/OS
- ▶ Enterprise PL/I for z/OS
- ▶ Automatic Binary Optimizer
- ▶ z/OS XL C/C++
- ▶ XL C/C++ for Linux on Z

The compilers increase the return on your investment in IBM Z hardware by maximizing application performance by using the compilers’ advanced optimization technology for z/Architecture. Through their support of web services, XML, and Java, they allow for the modernization of assets in web-based applications. They also support the latest IBM middleware products (CICS, Db2, and IMS), which allows applications to use their latest capabilities.

To fully use the capabilities of z15 servers, you must compile it by using the minimum level of each compiler. To obtain the best performance, you must specify an architecture level of 13 by using the **ARCH(13)** option.

For more information, see 7.5.4, “z/OS XL C/C++ considerations” on page 324.



Central processor complex hardware components

This chapter provides information about the new IBM z15™ and its hardware building blocks, and how these components physically interconnect. This information is useful for planning purposes and can help in defining configurations that fit your requirements.

Naming: The IBM z15 server generation is available as the following machine types and models:

- ▶ Machine Type 8561 (M/T 8561), Model T01, Features Max34, Max71, Max108, Max145, and Max190, which is further identified as *IBM z15 Model T01*, or *z15 T01*, unless otherwise specified.
- ▶ Machine Type 8562 (M/T 8562), Model T02, Features Max4, Max13, Max21, Max32, and Max65, which is further identified as *IBM z15 Model T02*, or *z15 T02*, unless otherwise specified.

In the remainder of this chapter, IBM z15 (z15) refers to both machine types.

This chapter includes the following topics:

- ▶ 2.1, “Frames and configurations” on page 36
- ▶ 2.2, “CPC drawer” on page 41
- ▶ 2.3, “Single chip modules” on page 49
- ▶ 2.4, “PCIe+ I/O drawer” on page 54
- ▶ 2.5, “Memory” on page 56
- ▶ 2.6, “Reliability, availability, and serviceability” on page 66
- ▶ 2.7, “Connectivity” on page 70
- ▶ 2.8, “Model configurations” on page 73
- ▶ 2.9, “Power and cooling” on page 77
- ▶ 2.10, “Summary” on page 87

2.1 Frames and configurations

The z15 Model T01 system is designed in a 19-inch form factor with configuration of 1 - 4 frames that can be easily installed in any data center. The z15 Machine Type 8561 can include 1 - 4 42U EIA (19-inch) frames, which are bolted together. The configurations can include up to five central processor complex (CPC) drawers and up to 12 Peripheral Component Interconnect Express+ (PCIe+) I/O drawers.

The redesigned CPC drawer and I/O infrastructure also lowers power consumption, reduces the footprint, and allows installation in virtually any data center. The z15 server is rated for ASHRAE class A3¹ data center operating environment.

The z15 server differentiates itself from previous Z server generations through the following significant changes to the modular hardware:

- ▶ All external cabling (power, I/O, and management) is performed at the rear of the system
- ▶ Flexible configurations: Frame quantity is determined by the system configuration (1 - 4 frames)
- ▶ Choice of power Intelligent Power Distribution Unit (iPDU or PDU) or Bulk Power Assembly (BPA)
- ▶ Feature codes that reserve slots for plan-ahead CPC drawers
- ▶ Added internal water cooling plumbing for systems with more than three CPC drawers
- ▶ New PCIe+ Gen3 I/O drawers (19-inch format) supporting 16 PCIe adapters

The power options include PDU-based power or BPA-based power. The z15 T01 server can be configured as a radiator (air) cooled or water cooled (that uses data center chilled water supply) system. Only BPA-based power system can be (optionally) configured for water cooling and with Internal Battery Feature (IBF), while a radiator (air) cooled system has PDU-based power (no Internal Battery Feature [IBF] available for PDU-based systems).

The z15 T01 includes the following basic hardware building blocks:

- ▶ 19-inch 42u frame (1 - 4)
- ▶ CPC (Processor) drawers (1 - 5)
- ▶ PCIe+ Gen3 I/O drawers (up to 12)
- ▶ CPC drawer Cooling Units: Radiator cooling assembly (RCA) or Water Cooling Unit (WCU)
- ▶ Power, with choice of:
 - Intelligent Power Distribution Units (iPDU) pairs (2 - 4 per frame, depending on the configuration).
 - Bulk Power Regulators (1 - 6 pairs, depending on the configuration)
- ▶ Support Elements (two):
 - Single KMM² device (USB-C connection)
 - Optional extra hardware for IBM Hardware Management Appliance feature
- ▶ 24-port 1GbE Switches (two or four, depending on the system configuration)
- ▶ Hardware for cable management at the rear of the system

¹ For more information, see Chapter 2, Environmental specifications in *IBM Z 8561 Installation Manual for Physical Planning*, GC28-7002.

² KMM - Keyboard, Mouse, Monitor

An example of a fully configured system with PDU-based power, five CPC drawers, and maximum 12 PCIe+ I/O drawers is shown in Figure 2-1.

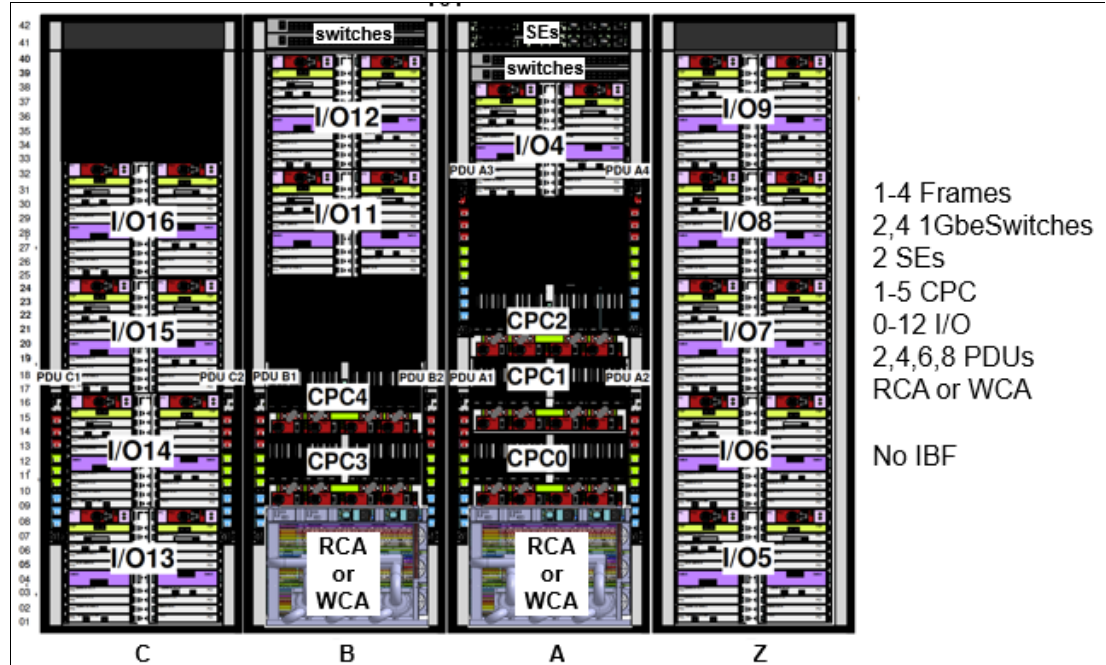


Figure 2-1 Maximum configuration, PDU-based powered system, rear view

An example of a BPA-based powered system with IBF, and a maximum of five CPC drawers and 11 PCIe+ I/O drawers is shown in Figure 2-2.

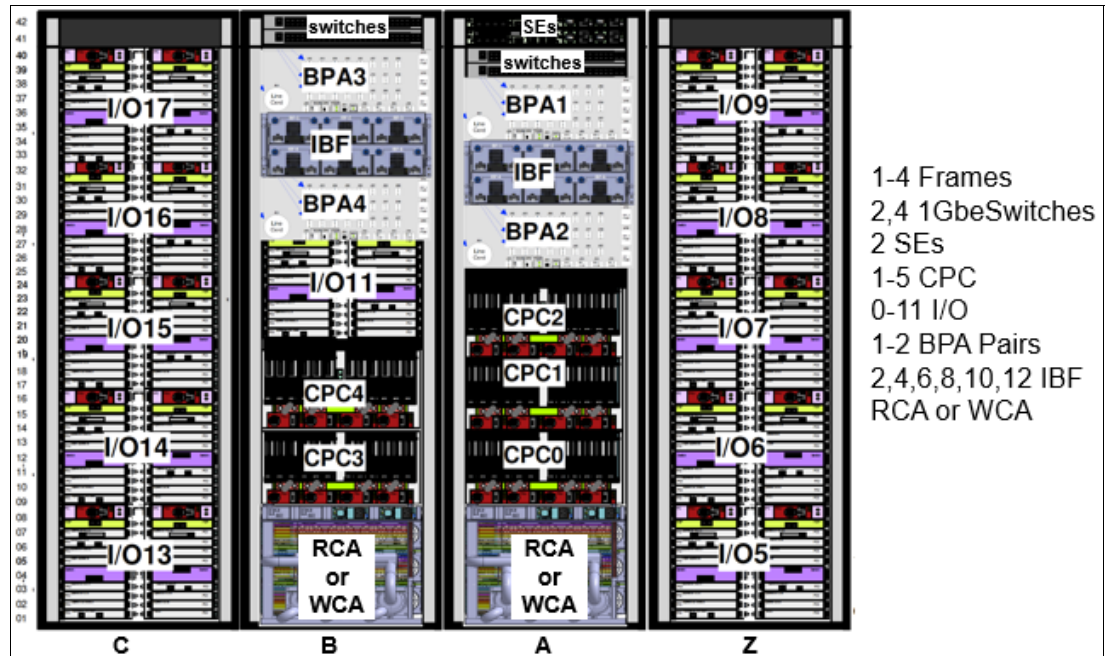


Figure 2-2 Maximum configuration, BPA-based powered system, rear view

The key features that are used to build the system are listed in Table 2-1 on page 38. For more information about the various configurations, see Appendix D, “Frame configurations” on page 515.

Table 2-1 Key features that influence the system configurations

| Feature Code | Description | Comments |
|---------------------------|------------------------------------|--|
| 0503 | Model T01 | Supports CPs and specialty engines |
| 0655 | One CPC Drawer | Feature Max34 |
| 0656 | Two CPC Drawers | Feature Max71 |
| 0657 | Three CPC Drawers | Feature Max108 |
| 0658 | Four CPC Drawers | Feature Max145 |
| 0659 | Five CPC Drawers | Feature Max190 |
| 2271 | CPC1 reserve | Reserve A15 location for future add CPC1 (Max34 to Max71 upgrade) |
| 2272 | CPC2 reserve | Reserve A20 location for future add CPC2 (Max71 to Max108 upgrade) |
| Frames and cooling | | |
| 4033 | A Frame | Radiator (air cooled) |
| 4034 | A Frame | Water cooled |
| 4035 | B Frame | Radiator (air cooled) |
| 4036 | B Frame | Water cooled |
| 4037 | Z Frame | I/O drawers only |
| 4038 | C Frame | I/O drawers only |
| PDU power | | |
| 0629 | 200-208V 60A 3 Phase (Delta - "Δ") | North America and Japan |
| 0630 | 380-415V 60A 3 Phase (Wye- "Y") | Worldwide (except North America and Japan) |
| BPA power | | |
| 0640 | Bulk Power Assembly (BPA) | Quantity 1 = 2 Bulk Power Enclosures Quantity 2 = 4 Bulk Power Enclosures |
| 3003 | Balanced Power Plan ahead | Only available with BPA |
| 3016 | Bulk Power Regulator (BPR) | Quantity per BPA feature: Min. 2, max 6 (in pairs) |
| 3217 | Internal Battery Feature (IBF) | Quantity per BPA feature: Min. 2, max 6 (in pairs) |
| I/O | | |
| 4021 | PCIe+ I/O drawer | Max. 12 (PDU) or max. 11 (BPA) |
| 7917 | Top Exit Cabling | Includes cable management top hat |
| 7919 | Bottom Exit Cabling | Includes rear tailgate hardware at bottom of frame |
| 7928 | Top Exit Cabling without Tophat | Uses rear slide plates at top of frame |

Considerations

Consider the following points:

- ▶ A-Frame is always present in every configuration
- ▶ 1u Support Elements (x2) are always in A-Frame at locations A41 and A42
- ▶ 1u 24-port internal Ethernet switches (x2) are always at locations A39 and A40
More Ethernet switches (x2) are available when necessary in Frame C or B
- ▶ I/O PCHID numbering starts with 0100 and increments depending on the number of features that is ordered. There is no PCHID number affinity to a fixed PCIe+ I/O drawer location as with previous systems.

2.1.1 z15 cover (door) design

The standard cover set for z15 model T01 is shown in Figure 2-3. Depending on the number of frames for the configuration, a Z and IBM accent top panel is installed on the outer frames. The single frame configuration combines the accents.



Figure 2-3 z15 Frames

The front doors of the z15 T01 for systems with 1 - 4 frames also is shown in Figure 2-3.

2.1.2 Top exit I/O and cabling

For the z15 Model T01 server, the top exit of all cables for I/O or power is always an option with no feature codes required. Adjustable cover plates are available for the openings at the top rear of each frame.

The Top Exit feature code (FC 7917) provides an optional Top Exit *cover enclosure*. The optional Top Exit cover enclosure provides cable retention hardware and mounting locations to secure Fiber Quick Connector MPO³ brackets on the top of the frames.

All external cabling enters the system at the rear of the frames for all I/O adapters, management LAN, and power connections.

Feature code 7917 provides a top hat assembly to be installed at the top of each frame in the configuration. This assembly is designed to assist with fiber trunking management.

³ MPO - Multi-fiber Push On connector

Overhead I/O cabling is contained within the frames. Extension “chimneys” that were featured with previous Z systems are no longer used.

A view of the top rear of the frame and the openings for top exit cables and power is shown in Figure 2-4. When FC 7917 is installed, the plated adjustable shields are removed and the top exit enclosure is installed.

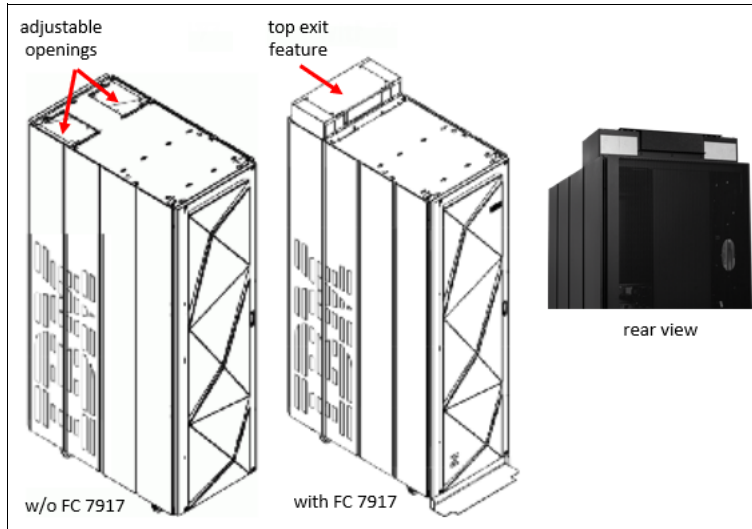


Figure 2-4 Top exit without and with FC7917

A newly designed vertical cable management guide (“spine”) can assist with proper cable management for fiber, copper, and coupling cables. Depending on the configuration, a spine is present from manufacturing with cable organizer clips installed.

The cable retention clips can be relocated for best usage. All external cabling to the system (from top or bottom) can use the spines to minimize interference with the PDUs that are mounted on the sides of the rack.

The rack with the spine mounted is shown in Figure 2-5. If necessary, the spine easily can be relocated for service procedures.

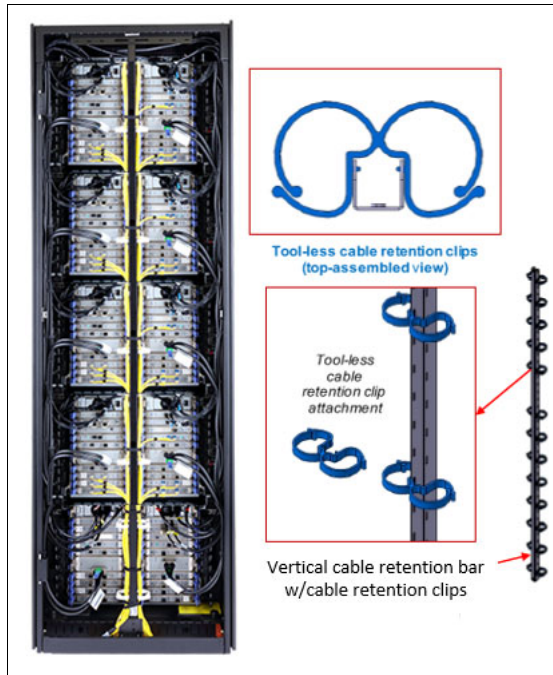


Figure 2-5 I/O cable management spine (Frame Z, rear view)

2.2 CPC drawer

The z15 Model T01 (machine type 8561) server continues the design of z14 by packaging processors in drawers. A z15 T01 CPC drawer includes the following features:

- ▶ Five single chip modules (SCMs)
- ▶ Up to 20 Memory DIMMs
- ▶ Symmetric multiprocessor (SMP) connectivity
- ▶ Connectors to support PCIe+ Gen3 fanout cards for PCIe+ I/O drawers or coupling fanouts for coupling links to other CPCs

The z15 T01 can be configured with 1 - 5 CPC drawers (three in the A frame and two in the B frame). A CPC drawer and its components are shown in Figure 2-6 on page 42.

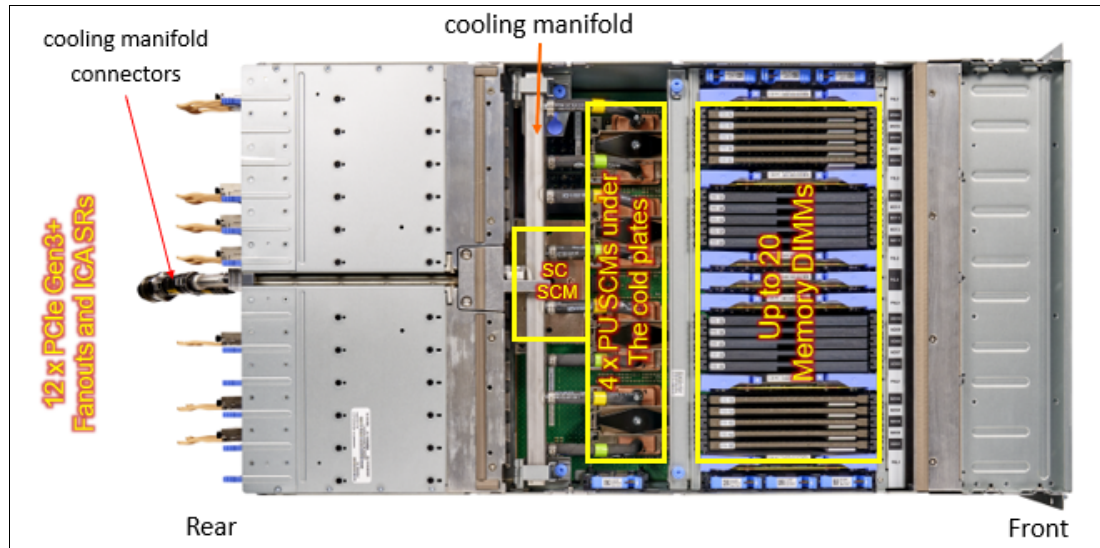


Figure 2-6 CPC drawer components (top view)

The z15 Model T01 5u CPC drawer always contains four Processor Unit (PU) SCMs, one System Controller (SC) SCM, and up to 20 memory DIMMs.

Depending on the feature, the z15 T01 contains the following CPC components:

- ▶ The number of CPC drawers installed is driven by the following feature codes:
 - FC 0655: One CPC drawer, Max34, up to 34 characterizable PUs
 - FC 0656: Two CPC drawers, Max71, up to 71 characterizable PUs
 - FC 0657: Three CPC drawers, Max108, up to 108 characterizable PUs
 - FC 0658: Four CPC drawers, Max145, up to 145 characterizable PUs
 - FC 0659: Five CPC drawers, Max190, up to 190 characterizable PUs
- ▶ The following SCMs are used:
 - PU SCM uses 14nm SOI technology, 17 layers of metal, 9.2 billion transistors, core running at 5.2GHz: (with 12 cores design per PU SCM).
 - SC SCM, 17 layers of metal, 12.2 billion transistors, 960 MB shared eDRAM L4 cache.
- ▶ Memory plugging:
 - Four memory controllers per drawer (one per PU SCM)
 - Each memory controller supports five DIMM slots
 - Four or three memory controllers per drawer are populated (up to 20 DIMMs)
 - Different memory controllers can have different size DIMMs
- ▶ Up to 12 PCIe+ Gen3 fanout slots that can host:
 - 2-Port PCIe+ Gen3 I/O fanout for PCIe+ I/O drawers (ordered and used in pairs for availability)
 - ICA SR and ICA SR1.1 PCIe fanout for coupling (two ports per feature)
- ▶ Management elements: Two dual function flexible service processor (FSP) and oscillator cards (OSC) for system control and to provide system clock (N+1 redundancy).

- ▶ CPC drawer power infrastructure consists of the following components:
 - Three or four Power Supply Units (PSUs) that provide power to the CPC drawer. The loss of one power supply leaves enough power to satisfy the drawer's power requirements (N+1 redundancy). The power supplies can be concurrently removed and replaced (one at a time)
 - 7x 12v distribution point-of-load (POL) that plug in slots that divide the memory banks
 - 7x Voltage Regulator Modules that plug outside of the memory DIMMs
 - Two Power Control cards to control the five CPC fans at the front of the CPC drawer
- ▶ Four SMP connectors that provide the CPC drawer to CPC drawer communication (NUMA).

The front view of the CPC drawer, which includes the cooling fans, FSP/OSC and bulk (power) distribution cards (BDC), is shown in Figure 2-7.

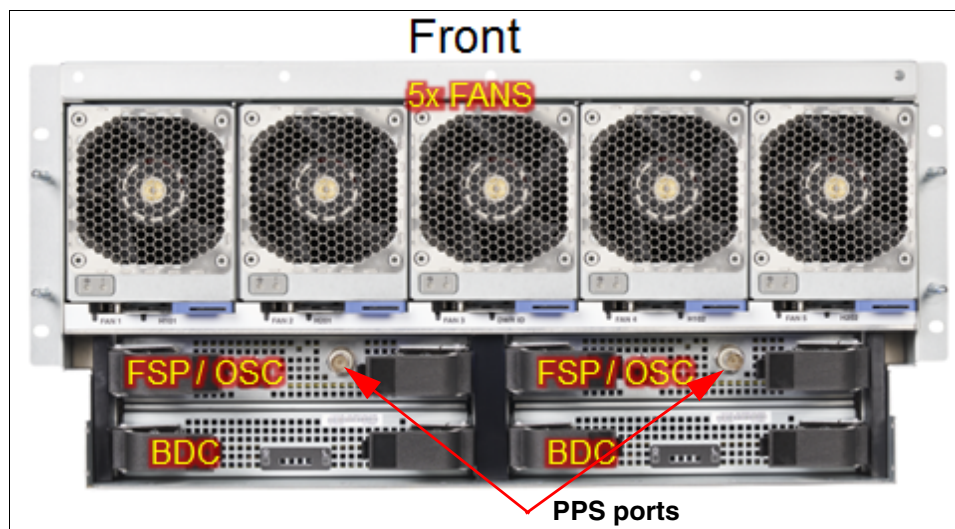


Figure 2-7 Front view of the CPC drawer

The rear view of a fully populated CPC Drawer is shown in Figure 2-8 on page 44. Dual port I/O fanouts and ICA SR adapters are plugged in specific slots for best performance and availability. Redundant power supplies and four SMP ports also are shown.

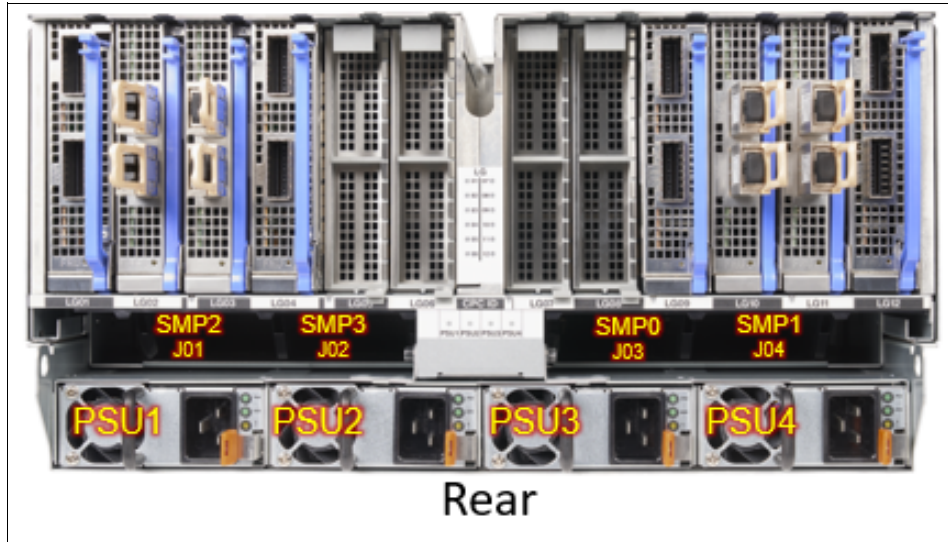


Figure 2-8 Rear view of the CPC drawer

The CPC drawer logical structure, component connections (including the PU SCMs), and the storage control SCMs are shown in Figure 2-9.

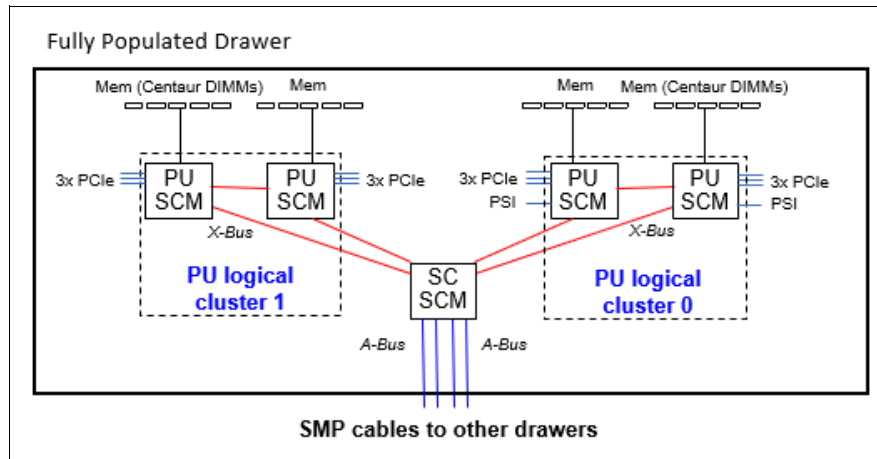


Figure 2-9 CPC drawer logical structure

Memory is connected to the SCMs through memory control units (MCUs). Up to four MCUs are available in a CPC drawer (one per PU SCM) and provide the interface to the DIMM controller. A memory controller uses five DIMM slots.

The buses are organized in the following configurations:

- ▶ The PCIe I/O buses provide connectivity for PCIe fanouts and can sustain up to 16 GBps data traffic per port.
- ▶ The X-bus provides interconnects between SC chip and PUs chips to each other, in the same logical cluster.
- ▶ The A-bus provides interconnects between SC chips (L4 cache) in different drawers by using SMP cables.
- ▶ Processor support interfaces (PSIs) are used to communicate with FSP cards for system control.

2.2.1 CPC drawer interconnect topology

The point-to-point SMP connection topology for CPC drawers is shown in Figure 2-10. Each CPC drawer communicates directly to all of the other CPC drawers SC SCM (L4 cache) by using point-to-point links.

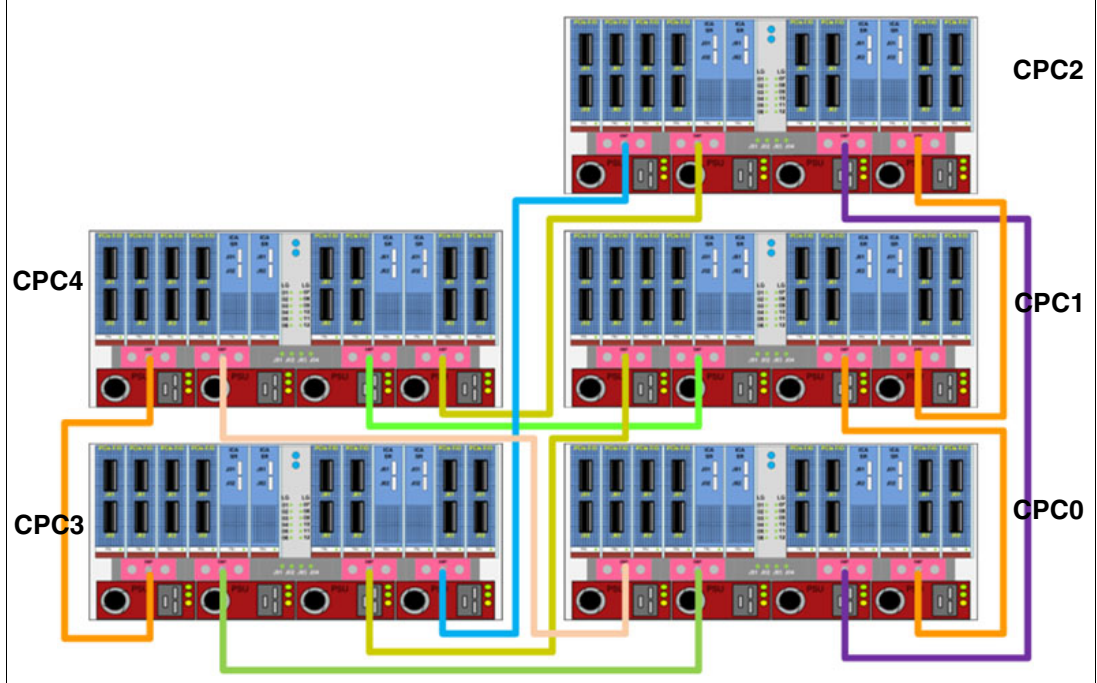


Figure 2-10 Maximum CPC drawer SMP connections (rear view)

The CPC drawers that are installed in Frame A and Frame B are populated from bottom to top.

The order of CPC drawer installation is listed in Table 2-2.

Table 2-2 CPC drawer installation order and position

| CPC drawer ^a | CPC0 | CPC1 | CPC2 | CPC3 | CPC4 |
|-------------------------|-------|--------|-------|--------|-------|
| Installation order | First | Second | Third | Fourth | Fifth |
| Position in Frame A | A10B | A15B | A20B | B10B | B15B |

a. CPC3 and CPC4 are factory installed only (no field MES available)

CPC drawer installation in the A frame is concurrent. Non-disruptive addition of CPC1 or CPC2 drawers is possible in the field (MES upgrade) if the reserve features (FC 2271 or FC 2272) are included with the initial system order. Concurrent drawer repair requires a minimum of two drawers.

2.2.2 Oscillator⁴

With z15 Model T01, the oscillator card design and signal distribution scheme is new; however, the RAS strategy for redundant clock signal and dynamic switchover is unchanged. One primary OSC card and one backup are used. If the primary OSC card fails, the secondary detects the failure, takes over transparently, and continues to provide the clock signal to the CPC.

Manage System Time

On z14, HMC 2.14.1 provided a significant user experience enhancement for timing controls with the new Manage System Time task.

For simplification, the z15 (2.15.0) removes the Support Element “Sysplex/System Timer” task panels. HMC level 2.15.0 (Driver 41) is required to manage system time for z15.

Network Time Protocol

The SEs provide the Simple Network Time Protocol (SNTP) client function. When Server Time Protocol (STP) is used, the time of an STP-only Coordinated Timing Network (CTN) can be synchronized to the time that is provided by a Network Time Protocol (NTP) server. This configuration allows time-of-day (TOD) synchronization in a heterogeneous platform environment and throughout the LPARs running on the CPC.

Precision Time Protocol

New for z15, Precision Time Protocol (PTP, IEEE 1588) can also be used as an external time source for IBM Z Server Time Protocol (STP) for an IBM Z Coordinated Timing Network (CTN). The initial implementation for PTP connectivity is provided by using the IBM Z Support Element (SE).

The accuracy of an STP-only CTN is improved by using an NTP or PTP server with the PPS output signal as the External Time Source (ETS). Devices with PPS output are available from several vendors that offer network timing solutions.

Consider the following points:

- ▶ A new card combines the FSP and OSC was implemented with z15. The internal physical cards (FSP and OSC) are separate, but combined as a single FRU because of a packaging design.
- ▶ Two local redundant oscillator cards are available per CPC drawer each with one PPS port.
- ▶ Current design requires Pulse Per Second use for providing maximum time accuracy for both NTP and PTP.
- ▶ An enhanced precision oscillator (20 PPM⁵ versus 50 PPM on previous systems) is used.
- ▶ The following PPS plugging rules apply (see Figure 2-11):
 - Single CPC drawer plug left and right OSC PPS coaxial connectors.
 - Multi-drawer plug CPC0 left OSC PPS and CPC1 left OSC PPS coaxial connectors.
 - Cables are routed from rear to front by using a pass-through hole in the frame, and under the CPC bezel by using a right-angle Bayonet Neill-Concelman (BNC) connector that provides the pulse per second (PPS) input for synchronization to an external time source with PPS output.

Cables are supplied by the customer.

⁴ Oscillator card (OSC) is combined (single FRU P/N) with the Flexible Support Processor (FSP); installed in pairs for each CPC Drawer
⁵ PPM - Parts Per Million

- Connected PPS ports must be assigned in the Manage System Time menus on the HMC.

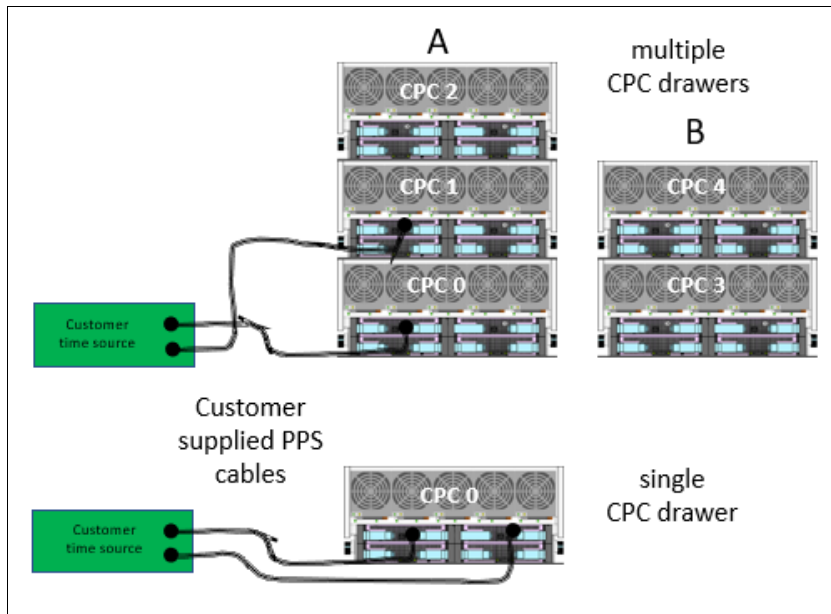


Figure 2-11 Recommended PPS cabling

Tip: STP is available as FC 1021. It is implemented in the Licensed Internal Code (LIC), and allows multiple servers to maintain time synchronization with each other and synchronization to an ETS.

For more information, see the Redbook: *IBM Server Time Protocol Guide*, SG24-8480.

2.2.3 System control

The various system elements are managed through the FSPs. An FSP is based on the IBM PowerPC® microprocessor technology.

With z15, the CPC drawer FSP card is combined with the Oscillator card in a single Field Replaceable Unit (FRU). Two combined FSP/OSC cards are used per CPC drawer.

Also, the PCIe+ I/O drawer has a new FSP. Each FSP card has one Ethernet port that connects to the internal Ethernet LANs through the internal network switches (SW1, SW2, and SW3, SW4, if configured). The FSPs communicate with the SEs and provide a subsystem interface (SSI) for controlling components.

An overview of the system control design is shown in Figure 2-12

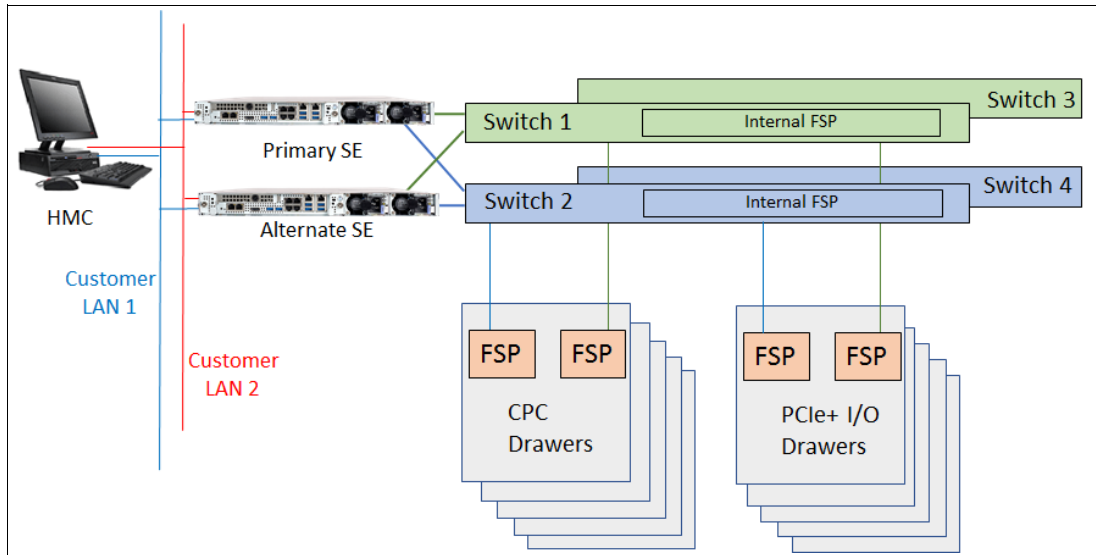


Figure 2-12 Conceptual overview of system control element

Note: The maximum z15 T01 system configuration features four GbE switches, five CPC drawers, and up to 12 PCIe I/O drawers.

A typical FSP operation is to control a power supply. An SE sends a command to the FSP to start the power supply. The FSP cycles the various components of the power supply, monitors the success of each step and the resulting voltages, and reports this status to the SE.

Most SEs are duplexed ($N+1$), and each element has at least one FSP. Two internal Ethernet LANs and two SEs, for redundancy, and crossover capability between the LANs, are available so that both SEs can operate on both LANs.

The Hardware Management Consoles (HMCs) and SEs are connected directly to one or two Ethernet Customer LANs. One or more HMCs can be used.

2.2.4 CPC drawer power

The power for the CPC drawer is a new design. It uses the following combinations of PSUs, POL⁶s, VRMs, and Bulk Distribution Cards:

- ▶ PSUs: Provide AC to 12V DC bulk/standby power and are installed at the rear of the CPC. The quantity that is installed depends on the following configurations:
 - Three PSUs for configurations that use BPA power
 - Four PSUs for configurations that use PDU power
- ▶ POLs: Seven Point of Load $N+2$ Redundant cards are installed next to the Memory DIMMs.
- ▶ VRMs: seven Voltage Regulator Modules ($N+2$ redundancy).
- ▶ Bulk distribution card (BDC): Redundant processor power and control cards connect to the CPC trail board. The control function is powered from 12V standby that is provided by the PSU. The BDC card also includes pressure, temperature, and humidity sensors.

⁶ POL - Point of Load, VRM - Voltage Regulator Module.

2.3 Single chip modules

The SCM is a multi-layer metal substrate module that holds one PU chip or an SC chip. Both PU and SC chip size is 696 mm² (25.3 mm x 27.5 mm). Each CPC drawer has four PU SCMs (9.2 billion transistors each), and one SC SCM (12.2 billion transistors).

The two types of SCMs (PU and SC) are shown in Figure 2-13. For both SCMs, a thermal cap is placed over the chip. Each PU SCM is water cooled by way of a cold plate manifold assembly. The SC SCM is air cooled by using CPC drawer fans and heat sink.

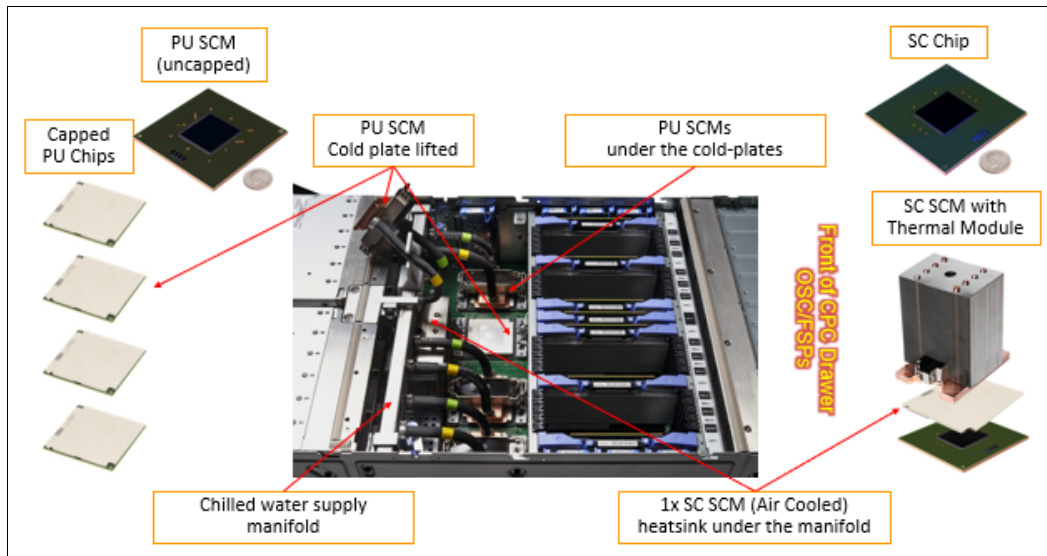


Figure 2-13 Single chip modules (PU SCM and SC SCM)

PU and SC chips use CMOS 14 nm process, 17 layers of metal, and state-of-the-art Silicon-On-Insulator (SOI) technology.

The SCMs are plugged into a socket that is part of the CPC drawer packaging. The interconnectivity between the CPC drawers is accomplished through SMP connectors and cables. Four inter-drawer connections are available on each CPC drawer. This configuration allows a multidrawer system to act as an SMP system.

2.3.1 Processor unit chip

A schematic representation of the PU chip is shown in Figure 2-14.

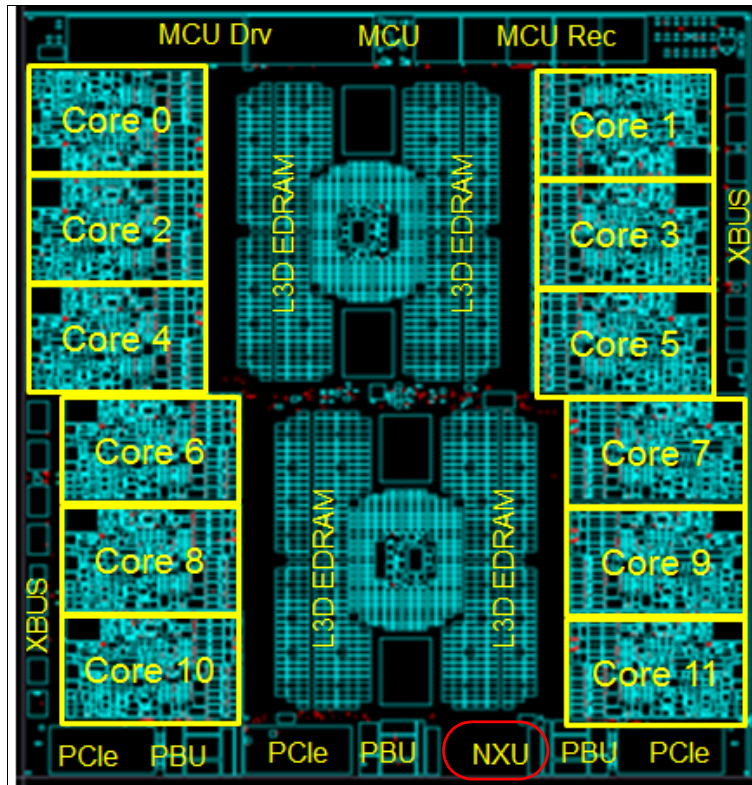


Figure 2-14 PU SCM floor plan

The z15 PU chip (installed as a PU SCM) is an evolution of the z14 chip design. It includes the following features and improvements:

- ▶ CMOS 14nm SOI technology
- ▶ 12 core design (versus 10 for z14) with increased on-chip cache sizes
- ▶ Three PCIe Gen4 interfaces (GX bus was dropped)
- ▶ DDR4 memory controller
- ▶ Two X-buses support cluster connectivity (PU SCM-to-PU SCM and PU SCM-to-SC SCM connectivity by way of X bus).
- ▶ New EDRAM macro design with 2x macro density. Compared to z14 PU:
 - L3 was increased from 128 MB to 256 MB per chip
 - L2-I was increased from 2 MB to 4 MB per core
 - L2-L3 protocol was changed to reduce latency
- ▶ On-chip compression accelerator (Nest Acceleration Unit - NXU)
- ▶ Further optimization of the nest-core staging

- ▶ Dedicated Co-Processor (CoP): The dedicated coprocessor is responsible for data compression and encryption functions for each core.
- ▶ Core pervasive unit (PC) for instrumentation and error collection.
- ▶ Modulo arithmetic (MA) unit: Support for Elliptic Curve Cryptography:
 - Vector and Floating point Units (VFU):
 - BFU: Binary floating point unit
 - DFU: Decimal floating point unit
 - DFx: Decimal fixed-point unit
 - FPd: Floating point divide unit
 - VXx: Vector fixed-point unit
 - VXs: Vector string unit
 - VXp: Vector permute unit
 - VXm: Vector multiply unit
 - L2I/L2D – Level 2 instruction/data cache

2.3.3 PU characterization

The PUs are characterized for client use. The characterized PUs can be used in general to run supported operating systems, such as z/OS, z/VM, and Linux on Z. They also can run specific workloads, such as Java, XML services, IPsec, and some Db2 workloads, or clustering functions, such as the Coupling Facility Control Code (CFCC).

The maximum number of characterizable PUs depends on the z15 CPC drawer feature code. Some PUs are characterized for system use; some are characterized for client workload use.

By default, one spare PU is available to assume the function of a failed PU. The maximum number of PUs that can be characterized for client use are listed in Table 2-3.

Table 2-3 PU characterization

| Feature | CPs | IFLs | Unassigned IFLs | zIIPs | ICFs | IFPs | Std SAPs | Add'l SAPs | Spare PUs |
|---------|-------|-------|-----------------|-------|-------|------|----------|------------|-----------|
| Max34 | 0-34 | 0-34 | 0-33 | 0-22 | 0-34 | 1 | 4 | 0-8 | 2 |
| Max71 | 0-71 | 0-71 | 0-70 | 0-46 | 0-71 | 1 | 8 | 0-8 | 2 |
| Max108 | 0-108 | 0-108 | 0-107 | 0-70 | 0-108 | 1 | 12 | 0-8 | 2 |
| Max145 | 0-145 | 0-145 | 0-144 | 0-96 | 0-145 | 1 | 16 | 0-8 | 2 |
| Max190 | 0-190 | 0-190 | 0-189 | 0-126 | 0-190 | 1 | 22 | 0-8 | 2 |

The rule for the CP to zIIP purchase ratio is that for every CP purchased, up to two zIIPs can be purchased. Java and XML workloads can run on zIIPs.

However, an LPAR definition can go beyond the 1:2 ratio. For example, a maximum of four physical zIIPs can be installed on a system with two physical CPs.

Converting a PU from one type to any other type is possible by using the Dynamic Processor Unit Reassignment process. These conversions occur concurrently with the system operation.

Note: The addition of ICFs, IFLs, zIIPs, and SAP to the z15 does not change the system capacity setting or its million service units (MSU) rating.

2.3.4 System Controller chip

The System Controller (SC) chip uses the CMOS 14nm SOI technology, with 17 layers of metal. It measures 25.3 x 27.5 mm, and has 12.2 billion transistors. Each CPC drawer of the system has one SC chip.

A schematic representation of the SC chip is shown in Figure 2-16. Consider the following points:

- ▶ A Bus (SC-SC off drawer): Minor changes to reflect protocol improvements and new system topology
- ▶ 960 MB shared eDRAM L4 Cache
- ▶ L4 Directory is built with eDRAM
- ▶ New L4 Cache Management: Ratio of L3 to L4 cache capacity is increasing

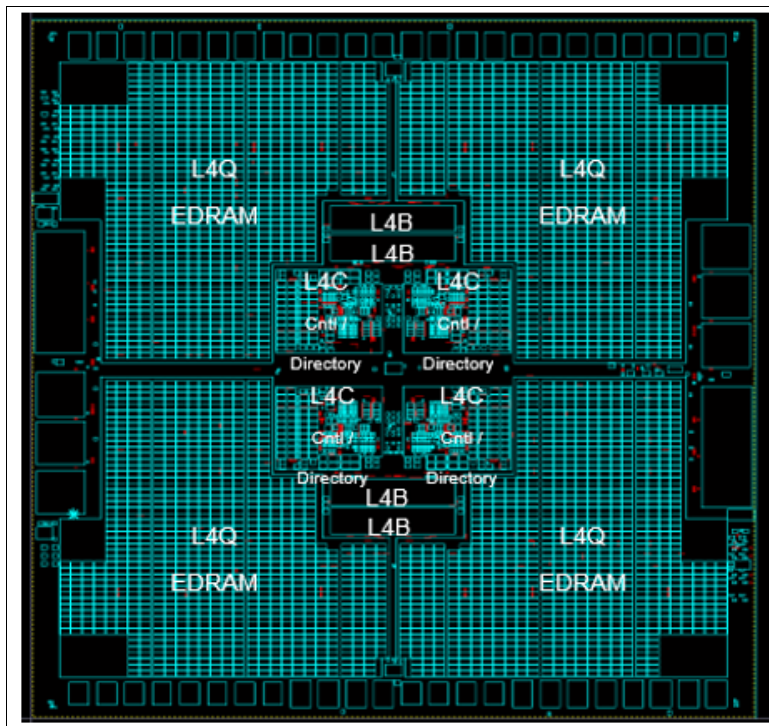


Figure 2-16 SC chip floor plan

2.3.5 Cache level structure

The cache structure comparison between CPC drawers on z14 M0x and z15 T01 is shown in Figure 2-17.

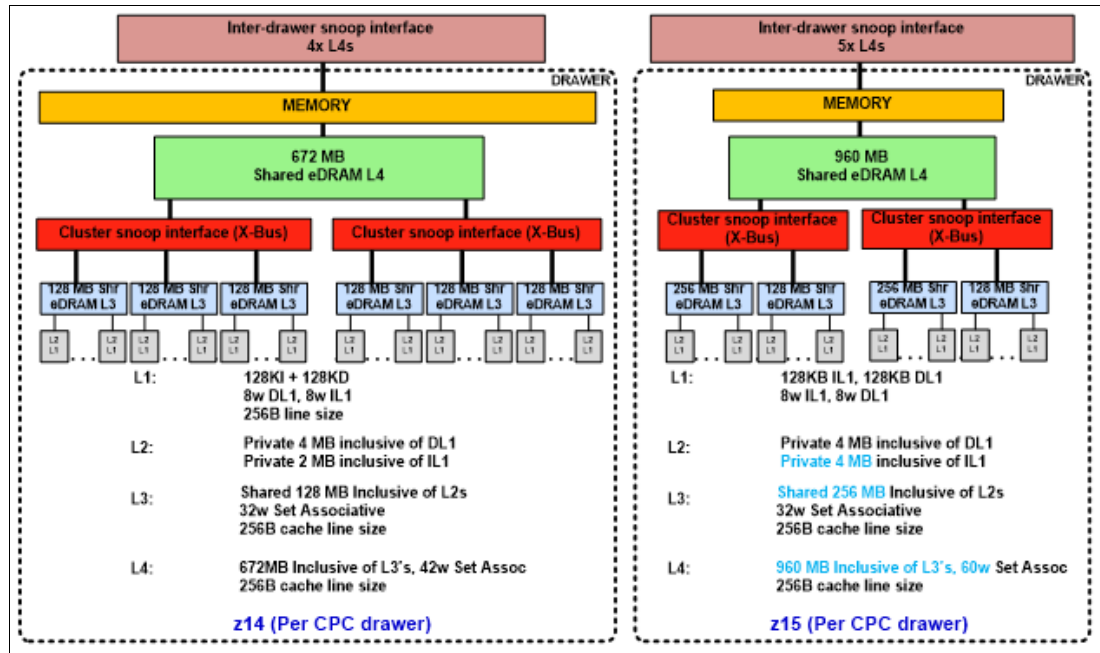


Figure 2-17 Cache structure comparison: z14 versus z15

2.4 PCIe+ I/O drawer

As shown in Figure 2-18 on page 55, each PCIe+ I/O drawer has 16 slots each to support the PCIe I/O infrastructure with a bandwidth of 16 GBps and includes the following features:

- ▶ A total of 16 I/O cards are spread over two I/O domains (0 and 1):
 - Each I/O slot reserves four PCHIDs.
 - Left side slots are numbered LG01-LG10 and right side slots are numbered LG11-LG20 from the rear of the rack. A location and LED identifier panel is at the center of the drawer.
 - New with z15 Model T01, the numbering of the PCHIDs is not related to a fixed location in a frame as with previous generations of Z systems. Instead, the first configured I/O location starts with PCHID 100 and continues the incremental sequence to the next configured PCIe I/O drawer. For more information about examples of the various configurations, see Appendix D, “Frame configurations” on page 515.
- ▶ Two PCIe+ switch cards provide connectivity to the PCIe+ Gen3 fanouts that are installed in the CPC drawers.
- ▶ Each I/O drawer domain has four dedicated support partitions (two per domain) to manage the native PCIe cards.
- ▶ Two Flexible Support Processor (FSP) cards are used to control the drawer function.
- ▶ Redundant N+1 power supplies (two) are mounted on the rear and redundant blowers (six) are mounted on the front.

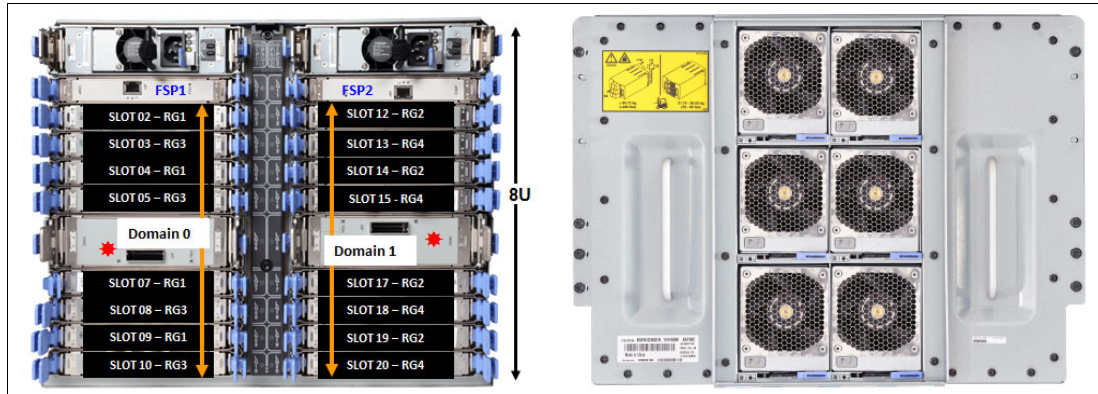


Figure 2-18 PCIe I/O drawer front and rear view

The following configuration examples and how the configurations are different with the power selection, number of CPC drawers and I/O features ordered, the layout of the PCIe+ I/O drawers, and PCHID numbering are shown in Figure 2-19:

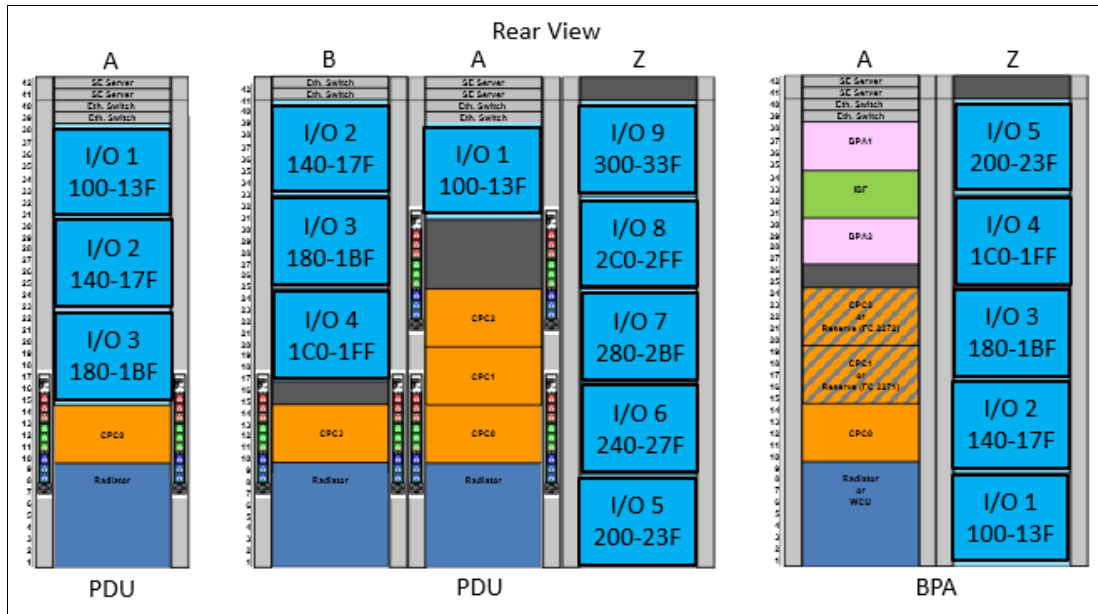


Figure 2-19 Configuration examples with PCHID numbering

- ▶ The first, a single frame system, ordered with PDU power, radiator cooling, one CPC drawer, and greater than 32 I/O features to drive three I/O drawers. PCHID numbering is consecutive from top to bottom.
- ▶ The second, a three frame system, ordered with PDU power, radiator cooling, four CPC drawers, and greater than 128 I/O features to drive nine I/O drawers. PCHID numbering starts in the A-frame, resumes in the B-frame from top down, and continues to the Z-frame working from the bottom up.
- ▶ The third, a two frame system, ordered with BPA, IBF power options, radiator cooling, one CPC drawer, two reserved CPC drawer slots for future CPC drawer add MES, and greater than 64 I/O features to drive five I/O drawers. PCHID numbering starts in the Z-frame, from the bottom and working up.

Consideration for PCHID identification:

In previous PCIe I/O drawers (introduced with zEC12), the orientation of the I/O features were vertical. For z15, the orientation of the PCIe features is horizontal, and the top of the card is now closest to the center of the drawer for the left and right side of the drawer.

The vertical card collapsed horizontal and the awareness of the port and PCHID layout where the top of the adapter (port D1) is closest to the location panel on both sides of the drawer are shown in Figure 2-20.

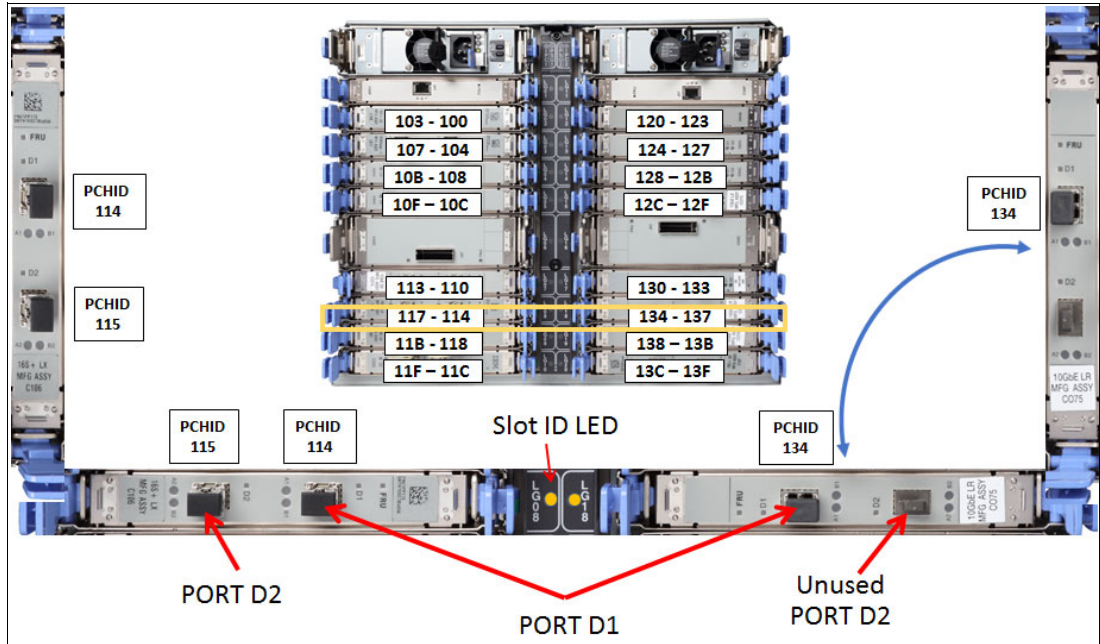


Figure 2-20 I/O feature orientation in PCIe I/O drawer (rear view)

Note: The CHPID Mapping Tool (available on ResourceLink) can be used to print a CHPID Report that displays the drawer and PCHID/CHPID layout.

2.5 Memory

The maximum physical memory size is directly related to the number of CPC drawers in the system. Each CPC drawer can contain up to 8 TB of customer memory, for a total of 40 TB of memory per system.

The minimum and maximum memory sizes that you can order for each z15 feature are listed in Table 2-4.

Table 2-4 Purchased Memory (Memory available for assignment to LPARs)

| Feature | # of CPC drawers | Customer memory GB | Flexible memory GB |
|---------|------------------|--------------------|--------------------|
| Max34 | 1 | 512 - 7936 | NA |
| Max71 | 2 | 512 - 16128 | 512 - 7936 |
| Max108 | 3 | 512 - 24320 | 512 - 16128 |
| Max145 | 4 | 512 - 32512 | 512 - 24320 |

| Feature | # of CPC drawers | Customer memory GB | Flexible memory GB |
|---------|------------------|--------------------|--------------------|
| Max190 | 5 | 512 - 40704 | 512 - 32512 |

The following memory types are available:

- ▶ Purchased: Memory that is available for assignment to LPARs.
- ▶ Hardware System Area (HSA): Standard 256 GB of addressable memory for system use outside of customer memory.
- ▶ Standard: Provides minimum physical memory that is required to hold customer purchase memory plus 256 GB HSA.
- ▶ Flexible: Provides more physical memory that is needed to support that activation of base customer memory and HSA on a multiple CPC drawer z15 with one drawer out of service (concurrent drawer replacement; not available on Max34 feature).

Note: The Plan Ahead Memory feature is not offered with a new order z15 system. The Plan Ahead Memory feature that is available on z13 or z14 can be carried forward to z15.

The memory granularity, which is based on the installed customer memory, is listed in Table 2-5.

Table 2-5 Customer offering memory increments

| Memory increment (GB) | Offered memory sizes (GB) |
|-----------------------|---------------------------|
| 64 | 512 - 768 |
| 128 | 896 - 2048 |
| 256 | 2304 - 3840 |
| 512 | 4352 - 17152 |
| 1024 | 18176 - 32512 |
| 2048 | 34560 - 40704 |

2.5.1 Memory subsystem topology

The z15 memory subsystem uses high-speed, differential-ended communications memory channels to link a host memory to the main memory storage devices.

The CPC drawer memory topology of a z15 server is shown in Figure 2-21.

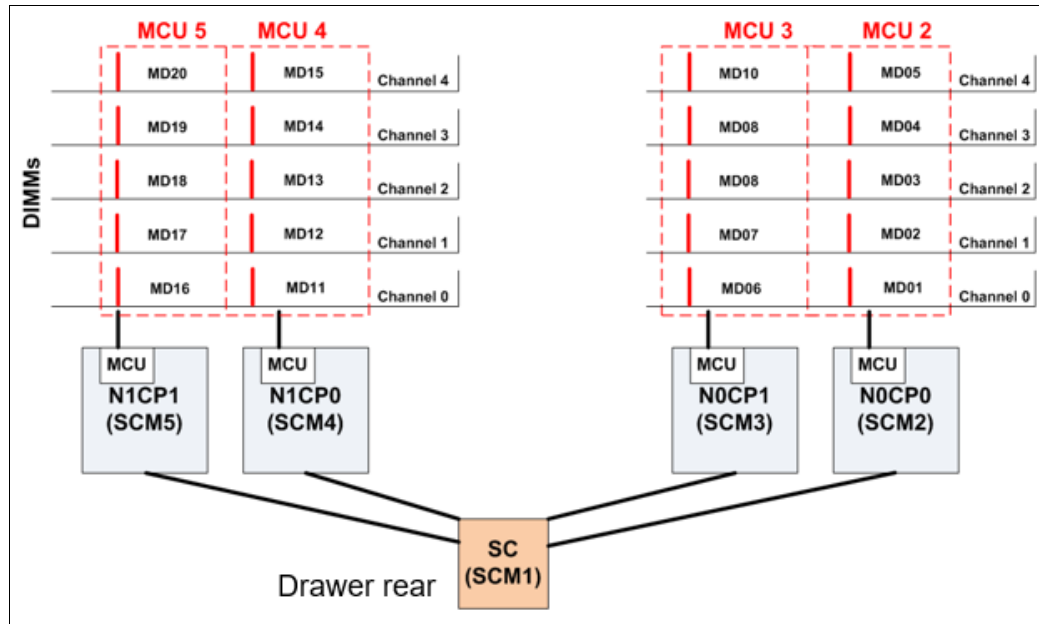


Figure 2-21 CPC drawer memory topology at maximum configuration

Consider the following points regarding the topology:

- ▶ One MCU per processor chip with five memory channels, one DIMM per channel (no DIMM cascading) is used.
- ▶ The fifth channel in each MCU enables memory to be implemented as a Redundant Array of Independent Memory (RAIM). This technology features significant error detection and correction capabilities. Bit, lane, DRAM, DIMM, socket, and complete memory channel failures can be detected and corrected, including many types of multiple failures. Therefore, RAIM takes 20% of DIMM capacity. (No non-RAIM option is available.)
- ▶ DIMM sizes used are 32, 64, 128, 256 and 512 GB with five DIMMs of the same size included in a memory feature (160, 320, 640, 1280 and 2560 GB RAIM array size, respectively).
- ▶ Three or four features (15 or 20 DIMMs) are plugged in each drawer.
- ▶ Features with different DIMMs sizes can be mixed in the same drawer.
- ▶ The five DIMMs per MCU must be the same size.
- ▶ Addressable memory is required for partitions and HSA.

2.5.2 Redundant array of independent memory

The z15 server uses the RAIM technology. The RAIM design detects and recovers from failures of dynamic random access memory (DRAM), sockets, memory channels, or DIMMs.

The RAIM design requires the addition of one memory channel that is dedicated for reliability, availability, and serviceability (RAS).

The five channel RAIM Memory Controller overview is shown in Figure 2-22.

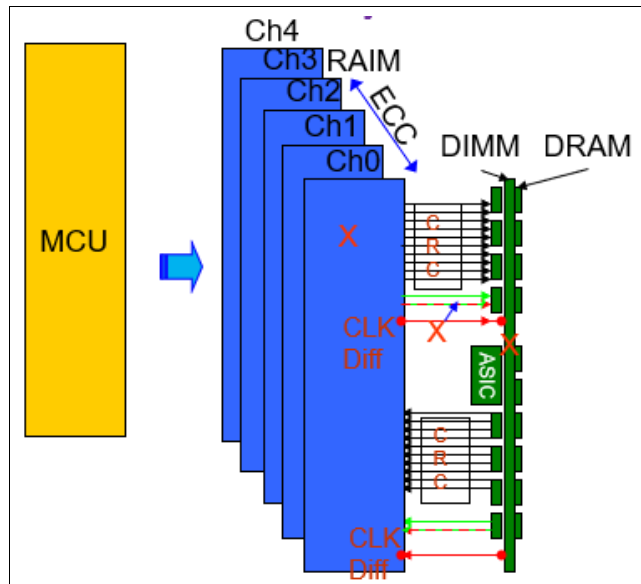


Figure 2-22 Five channel RAIM Memory Controller Overview

The fifth channel in each MCU enables memory to be implemented as a RAIM. This technology features significant error detection and correction capabilities. Bit, lane, DRAM, DIMM, socket, and complete memory channel failures can be detected and corrected, including many types of multiple failures. Therefore, RAIM takes 20% of DIMM capacity (a non-RAIM option is not available).

The RAIM design provides the following layers of memory recovery:

- ▶ ECC with 90B/64B Reed Solomon code.
- ▶ DRAM failure, with marking technology in which two DRAMs can be marked and no half sparing is needed. A call for replacement occurs on the third DRAM failure.
- ▶ Lane failure with CRC retry, data-lane sparing, and clock-RAIM with lane sparing.
- ▶ DIMM failure (discrete components and VTT Reg) with CRC retry, data-lane sparing, and clock-RAIM with lane sparing.
- ▶ DIMM controller ASIC failure.
- ▶ Channel failure started RAIM recovery.

2.5.3 Memory configurations

Memory sizes in each CPC drawer do not have to be similar. Different CPC drawers can contain different amounts of memory. The 10 (10-19) drawer memory configurations that are supported are listed in Table 2-6. Each CPC drawer is included from manufacturing with one of these memory configurations. Total physical memory includes RAIM (20%).

Table 2-6 Drawer memory plugging configurations

| CFG # | Physical memory GB | 32 GB #DIMMs | 64 GB #DIMMs | 128 GB #DIMMs | 256 GB #DIMMs | 512 GB #DIMMs | -RAIM GB | -HSA GB |
|-------|--------------------|--------------|--------------|---------------|---------------|---------------|----------|---------|
| 10 | 640 | 20 | 0 | 0 | 0 | 0 | 512 | 256 |

| CFG # | Physical memory GB | 32 GB #DIMMs | 64 GB #DIMMs | 128 GB #DIMMs | 256 GB #DIMMs | 512 GB #DIMMs | -RAIM GB | -HSA GB |
|-------|--------------------|--------------|--------------|---------------|---------------|---------------|----------|---------|
| 11 | 960 | 10 | 10 | 0 | 0 | 0 | 768 | 512 |
| 12 | 1280 | 0 | 20 | 0 | 0 | 0 | 1024 | 768 |
| 13 | 1600 | 0 | 10 | 10 | 0 | 0 | 1536 | 1280 |
| 14 | 1920 | 0 | 0 | 20 | 0 | 0 | 2048 | 1792 |
| 15 | 3840 | 0 | 0 | 10 | 10 | 0 | 3072 | 2816 |
| 16 | 5120 | 0 | 0 | 0 | 20 | 0 | 4096 | 3840 |
| 17 | 7680 | 0 | 0 | 0 | 10 | 10 | 6144 | 5888 |
| 18 | 10240 | 0 | 0 | 0 | 0 | 20 | 8192 | 7936 |
| 19 | 480 | 15 | 0 | 0 | 0 | 0 | 384 | 128 |

Consider the following points:

- ▶ A CPC drawer contains a minimum of 15 32GB DIMMs as listed in drawer configuration number 19 in Table 2-6 on page 59.
- ▶ A CPC drawer can have more memory installed than what is actually enabled for client use. The amount of memory that can be enabled by the client is the total physically installed memory minus the RAIM amount (20%) and minus the 256 GB of HSA memory.
- ▶ A CPC drawer can have available unused memory, which can be ordered as a memory upgrade and enabled by LIC-CC without DIMM changes.
- ▶ DIMM changes require a disruptive power-on reset (POR) on z15 T01 with a single CPC drawer. DIMM changes can be done concurrently on z15 models with multiple CPC drawers using Enhanced Drawer Availability (EDA).

DIMM plugging for the configurations in each CPC drawer do not have to be similar. Each memory 5 slot DIMM bank must have the same DIMM size; however, a drawer can have a mix of DIMM banks. Table 2-7 lists the memory population by DIMM bank for the 15 configurations that are listed in Table 2-6 on page 59.

As an example, for configuration #14, memory positions MD06-MD10 are populated with five 128 GB DIMMS.

Table 2-7 Memory Population by DIMM Bank

| CFG # | MD01-MD05 | MD06-MD10 | MD11-MD15 | MD16-MD20 | Physical | Total -RAIM | Total-RAIM+HSA |
|-------|-----------|-----------|-----------|-----------|----------|-------------|----------------|
| 10 | 32 | 32 | 32 | 32 | 640 | 512 | 256 |
| 11 | 64 | 32 | 32 | 64 | 960 | 768 | 512 |
| 12 | 64 | 64 | 64 | 64 | 1280 | 1024 | 768 |
| 13 | 128 | 64 | 64 | 128 | 1920 | 1536 | 1280 |
| 14 | 128 | 128 | 128 | 128 | 2560 | 2048 | 1792 |
| 15 | 256 | 128 | 128 | 256 | 3840 | 3072 | 2816 |
| 16 | 256 | 256 | 256 | 256 | 5120 | 4096 | 3840 |

| CFG # | MD01-MD05 | MD06-MD10 | MD11-MD15 | MD16-MD20 | Physical | Total -RAIM | Total-RAIM+HSA |
|-------|-----------|-----------|-----------|-----------|----------|-------------|----------------|
| 17 | 512 | 256 | 256 | 512 | 7680 | 6144 | 5888 |
| 18 | 512 | 512 | 512 | 512 | 10240 | 8192 | 7936 |
| 19 | 32 | 32 | 32 | | 480 | 384 | 128 |

The support element View Hardware Configuration task can be used to determine the size and quantity of the memory plugged in each drawer. Figure 2-23 shows an example of configuration number 16 from the previous tables, and displays the location and description of the installed memory modules.

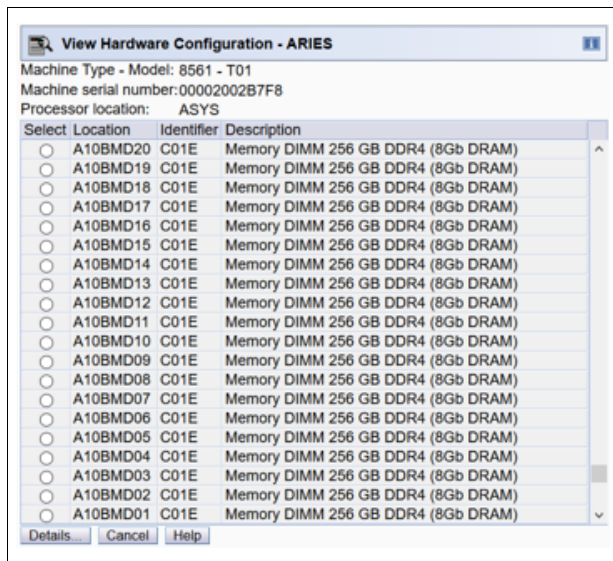


Figure 2-23 View Hardware Configuration task on the Support Element

Figure 2-24 shows the CPC drawer and DIMM locations for a z15.

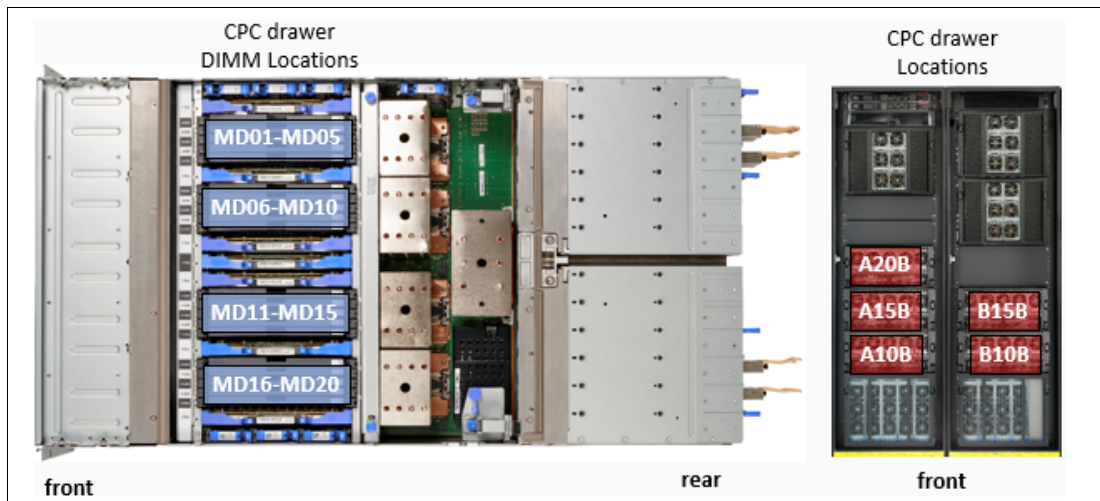


Figure 2-24 CPC drawer and DIMM locations for a z15

Table 2-8 lists the physical memory plugging configurations by feature code from manufacturing when the system is ordered. Consider the following points:

- ▶ The CPC drawer columns for the specific feature contain the Memory Plug Drawer Configuration number that is referenced in Table 2-6 on page 59 and the Population by DIMM Bank that is listed in Table 2-7 on page 60.
- ▶ Dial Max indicates the maximum memory that can be enabled by way of the LICC concurrent upgrade.

If more storage is ordered by using other feature codes, such as Virtual Flash Memory, or Flexible Memory, the extra storage is installed and plugged as necessary.

For example, a customer orders FC 1528 that features 1920 GB memory and Max145 (4 CPC drawers). The drawer configurations include the following components:

- ▶ CPC0 (768GB), CPC3 (768GB) - Configuration #11 (from Table 2-6 on page 59)
- ▶ CPC1 (384GB), CPC2 (384GB) - Configuration #19
- ▶ Total 768 + 768 + 384 + 384 - 256 HSA= 2048GB (Dial Max)

Table 2-8 Memory features and physical plugging

| Feature Code | Increments | Customer Memory Increments | Max34 | | Max71 | | | Max108 | | | | Max145 | | | | | Max190 | | | | | |
|--------------|------------|----------------------------|-------------|----------|-------------|-------------|----------|-------------|-------------|-------------|----------|-------------|-------------|-------------|-------------|----------|-------------|-------------|-------------|-------------|-------------|----------|
| | | | Drawer CPC0 | Dial Max | Drawer CPC0 | Drawer CPC1 | Dial Max | Drawer CPC0 | Drawer CPC1 | Drawer CPC2 | Dial Max | Drawer CPC0 | Drawer CPC1 | Drawer CPC2 | Drawer CPC3 | Dial Max | Drawer CPC0 | Drawer CPC1 | Drawer CPC2 | Drawer CPC3 | Drawer CPC4 | Dial Max |
| 1515 | 64 | 512 | 12 | 768 | 11 | 19 | 896 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1516 | | 576 | 12 | 768 | 11 | 19 | 896 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1517 | | 640 | 12 | 768 | 11 | 19 | 896 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1518 | | 704 | 12 | 768 | 11 | 19 | 896 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1519 | | 768 | 12 | 768 | 11 | 19 | 896 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1520 | 128 | 896 | 13 | 1280 | 11 | 19 | 896 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1521 | | 1024 | 13 | 1280 | 12 | 19 | 1152 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1522 | | 1152 | 13 | 1280 | 12 | 19 | 1152 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1523 | | 1280 | 13 | 1280 | 12 | 11 | 1536 | 11 | 19 | 19 | 1280 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1524 | | 1408 | 14 | 1792 | 12 | 11 | 1536 | 11 | 19 | 11 | 1664 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1525 | | 1536 | 14 | 1792 | 12 | 11 | 1536 | 11 | 19 | 11 | 1664 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1526 | | 1664 | 14 | 1792 | 12 | 12 | 1792 | 11 | 19 | 11 | 1664 | 11 | 19 | 19 | 19 | 1664 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1527 | | 1792 | 14 | 1792 | 12 | 12 | 1792 | 11 | 11 | 11 | 2048 | 11 | 19 | 19 | 11 | 2048 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1528 | | 1920 | 15 | 2816 | 13 | 13 | 2816 | 11 | 11 | 11 | 2048 | 11 | 19 | 19 | 11 | 2048 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1529 | | 2048 | 15 | 2816 | 13 | 13 | 2816 | 11 | 11 | 11 | 2048 | 11 | 19 | 19 | 11 | 2048 | 11 | 19 | 19 | 19 | 19 | 2048 |
| 1530 | 256 | 2304 | 15 | 2816 | 13 | 13 | 2816 | 12 | 11 | 12 | 2560 | 11 | 11 | 11 | 11 | 2816 | 11 | 11 | 19 | 19 | 11 | 2816 |
| 1531 | | 2560 | 15 | 2816 | 13 | 13 | 2816 | 12 | 11 | 12 | 2560 | 11 | 11 | 11 | 11 | 2816 | 11 | 11 | 19 | 19 | 11 | 2816 |
| 1532 | | 2816 | 15 | 2816 | 13 | 13 | 2816 | 12 | 12 | 12 | 2816 | 11 | 11 | 11 | 11 | 2816 | 11 | 11 | 19 | 19 | 11 | 2816 |
| 1533 | | 3072 | 16 | 3840 | 14 | 14 | 3840 | 13 | 12 | 13 | 3840 | 12 | 11 | 11 | 12 | 3328 | 11 | 11 | 11 | 11 | 11 | 3584 |
| 1534 | | 3328 | 16 | 3840 | 14 | 14 | 3840 | 13 | 12 | 13 | 3840 | 12 | 11 | 11 | 12 | 3328 | 11 | 11 | 11 | 11 | 11 | 3584 |
| 1535 | | 3584 | 16 | 3840 | 14 | 14 | 3840 | 13 | 12 | 13 | 3840 | 12 | 12 | 12 | 12 | 3840 | 11 | 11 | 11 | 11 | 11 | 3584 |
| 1536 | | 3840 | 16 | 3840 | 14 | 14 | 3840 | 13 | 12 | 13 | 3840 | 12 | 12 | 12 | 12 | 3840 | 12 | 12 | 11 | 11 | 12 | 4352 |

| Feature Code | Increments | Customer Memory Increments | Max34 | | Max71 | | | Max108 | | | | Max145 | | | | | Max190 | | | | | |
|--------------|------------|----------------------------|-------------|----------|-------------|-------------|----------|-------------|-------------|-------------|----------|-------------|-------------|-------------|-------------|----------|-------------|-------------|-------------|-------------|-------------|----------|
| | | | Drawer OPC0 | Dial Max | Drawer OPC0 | Drawer OPC1 | Dial Max | Drawer OPC0 | Drawer OPC1 | Drawer OPC2 | Dial Max | Drawer OPC0 | Drawer OPC1 | Drawer OPC2 | Drawer OPC3 | Dial Max | Drawer OPC0 | Drawer OPC1 | Drawer OPC2 | Drawer OPC3 | Drawer OPC4 | Dial Max |
| 1537 | 512 | 4352 | 17 | 5888 | 15 | 15 | 5888 | 13 | 13 | 13 | 4352 | 13 | 12 | 12 | 13 | 4864 | 12 | 12 | 11 | 11 | 12 | 4352 |
| 1538 | | 4864 | 17 | 5888 | 15 | 15 | 5888 | 14 | 13 | 14 | 5376 | 13 | 12 | 12 | 13 | 4864 | 12 | 12 | 12 | 12 | 12 | 4864 |
| 1539 | | 5376 | 17 | 5888 | 15 | 15 | 5888 | 14 | 13 | 14 | 5376 | 13 | 13 | 13 | 13 | 5888 | 13 | 12 | 12 | 12 | 13 | 5888 |
| 1540 | | 5888 | 17 | 5888 | 15 | 15 | 5888 | 14 | 14 | 14 | 5888 | 13 | 13 | 13 | 13 | 5888 | 13 | 12 | 12 | 12 | 13 | 5888 |
| 1541 | | 6400 | 18 | 7936 | 16 | 16 | 7936 | 15 | 14 | 15 | 7936 | 14 | 13 | 13 | 14 | 6912 | 13 | 13 | 12 | 13 | 13 | 6912 |
| 1542 | | 6912 | 18 | 7936 | 16 | 16 | 7936 | 15 | 14 | 15 | 7936 | 14 | 13 | 13 | 14 | 6912 | 13 | 13 | 12 | 13 | 13 | 6912 |
| 1543 | | 7424 | 18 | 7936 | 16 | 16 | 7936 | 15 | 14 | 15 | 7936 | 14 | 14 | 14 | 14 | 7936 | 13 | 13 | 13 | 13 | 13 | 7424 |
| 1544 | | 7936 | 18 | 7936 | 16 | 16 | 7936 | 15 | 14 | 15 | 7936 | 14 | 14 | 14 | 14 | 7936 | 14 | 13 | 13 | 13 | 14 | 8448 |
| 1545 | | 8448 | | | 17 | 17 | 12032 | 15 | 15 | 15 | 8960 | 15 | 14 | 14 | 15 | 9984 | 14 | 13 | 13 | 13 | 14 | 8448 |
| 1546 | | 8960 | | | 17 | 17 | 12032 | 15 | 15 | 15 | 8960 | 15 | 14 | 14 | 15 | 9984 | 14 | 14 | 13 | 14 | 14 | 9472 |
| 1547 | | 9472 | | | 17 | 17 | 12032 | 16 | 15 | 16 | 11008 | 15 | 14 | 14 | 15 | 9984 | 14 | 14 | 13 | 14 | 14 | 9472 |
| 1548 | | 9984 | | | 17 | 17 | 12032 | 16 | 15 | 16 | 11008 | 15 | 14 | 14 | 15 | 9984 | 14 | 14 | 14 | 14 | 14 | 9984 |
| 1549 | | 10496 | | | 17 | 17 | 12032 | 16 | 15 | 16 | 11008 | 15 | 15 | 15 | 15 | 12032 | 15 | 14 | 14 | 14 | 15 | 12032 |
| 1550 | | 11008 | | | 17 | 17 | 12032 | 16 | 15 | 16 | 11008 | 15 | 15 | 15 | 15 | 12032 | 15 | 14 | 14 | 14 | 15 | 12032 |
| 1551 | | 11520 | | | 17 | 17 | 12032 | 16 | 16 | 16 | 12032 | 15 | 15 | 15 | 15 | 12032 | 15 | 14 | 14 | 14 | 15 | 12032 |
| 1552 | | 12032 | | | 17 | 17 | 12032 | 16 | 16 | 16 | 12032 | 15 | 15 | 15 | 15 | 12032 | 15 | 14 | 14 | 14 | 15 | 12032 |
| 1553 | | 12544 | | | 18 | 18 | 16128 | 17 | 16 | 17 | 16128 | 16 | 15 | 15 | 16 | 14080 | 15 | 14 | 14 | 15 | 15 | 13056 |
| 1554 | | 13056 | | | 18 | 18 | 16128 | 17 | 16 | 17 | 16128 | 16 | 15 | 15 | 16 | 14080 | 15 | 14 | 14 | 15 | 15 | 13056 |
| 1555 | | 13568 | | | 18 | 18 | 16128 | 17 | 16 | 17 | 16128 | 16 | 15 | 15 | 16 | 14080 | 15 | 15 | 15 | 15 | 15 | 15104 |
| 1556 | | 14080 | | | 18 | 18 | 16128 | 17 | 16 | 17 | 16128 | 16 | 15 | 15 | 16 | 14080 | 15 | 15 | 15 | 15 | 15 | 15104 |
| 1557 | | 14592 | | | 18 | 18 | 16128 | 17 | 16 | 17 | 16128 | 16 | 16 | 16 | 16 | 16128 | 15 | 15 | 15 | 15 | 15 | 15104 |
| 1558 | | 15104 | | | 18 | 18 | 16128 | 17 | 16 | 17 | 16128 | 16 | 16 | 16 | 16 | 16128 | 15 | 15 | 15 | 15 | 15 | 15104 |
| 1559 | | 15616 | | | 18 | 18 | 16128 | 17 | 16 | 17 | 16128 | 16 | 16 | 16 | 16 | 16128 | 16 | 15 | 15 | 15 | 16 | 17152 |
| 1560 | | 16128 | | | 18 | 18 | 16128 | 17 | 16 | 17 | 16128 | 16 | 16 | 16 | 16 | 16128 | 16 | 15 | 15 | 15 | 16 | 17152 |
| 1561 | | 16640 | | | | | | 17 | 17 | 17 | 18176 | 17 | 16 | 16 | 17 | 20224 | 16 | 15 | 15 | 15 | 16 | 17152 |
| 1562 | | 17152 | | | | | | 17 | 17 | 17 | 18176 | 17 | 16 | 16 | 17 | 20224 | 16 | 15 | 15 | 15 | 16 | 17152 |

| Feature Code | Increments | Customer Memory Increments | Max34 | | Max71 | | | Max108 | | | | Max145 | | | | Max190 | | | | | |
|--------------|------------|----------------------------|-------------|----------|-------------|-------------|----------|-------------|-------------|-------------|----------|-------------|-------------|-------------|-------------|----------|-------------|-------------|-------------|-------------|-------------|
| | | | Drawer CPC0 | Dial Max | Drawer CPC0 | Drawer CPC1 | Dial Max | Drawer CPC0 | Drawer CPC1 | Drawer CPC2 | Dial Max | Drawer CPC0 | Drawer CPC1 | Drawer CPC2 | Drawer CPC3 | Dial Max | Drawer CPC0 | Drawer CPC1 | Drawer CPC2 | Drawer CPC3 | Drawer CPC4 |
| 1563 | 1024 | 18176 | | | | | 17 | 17 | 17 | 18176 | 17 | 16 | 16 | 17 | 20224 | 16 | 16 | 16 | 15 | 16 | 19200 |
| 1564 | | 19200 | | | | | 18 | 17 | 18 | 22272 | 17 | 16 | 16 | 17 | 20224 | 16 | 16 | 16 | 15 | 16 | 19200 |
| 1565 | | 20224 | | | | | 18 | 17 | 18 | 22272 | 17 | 16 | 16 | 17 | 20224 | 16 | 16 | 16 | 16 | 16 | 20224 |
| 1566 | | 21248 | | | | | 18 | 17 | 18 | 22272 | 17 | 17 | 17 | 17 | 24320 | 17 | 16 | 16 | 16 | 17 | 24320 |
| 1567 | | 22272 | | | | | 18 | 17 | 18 | 22272 | 17 | 17 | 17 | 17 | 24320 | 17 | 16 | 16 | 16 | 17 | 24320 |
| 1568 | | 23296 | | | | | 18 | 18 | 18 | 24320 | 17 | 17 | 17 | 17 | 24320 | 17 | 16 | 16 | 16 | 17 | 24320 |
| 1569 | | 24320 | | | | | 18 | 18 | 18 | 24320 | 17 | 17 | 17 | 17 | 24320 | 17 | 16 | 16 | 16 | 17 | 24320 |
| 1570 | | 25344 | | | | | | | | | 18 | 17 | 17 | 18 | 28416 | 17 | 17 | 16 | 16 | 17 | 26368 |
| 1571 | | 26368 | | | | | | | | | 18 | 17 | 17 | 18 | 28416 | 17 | 17 | 16 | 16 | 17 | 26368 |
| 1572 | | 27392 | | | | | | | | | 18 | 17 | 17 | 18 | 28416 | 17 | 17 | 17 | 16 | 17 | 28416 |
| 1573 | | 28416 | | | | | | | | | 18 | 17 | 17 | 18 | 28416 | 17 | 17 | 17 | 16 | 17 | 28416 |
| 1574 | | 29440 | | | | | | | | | 18 | 18 | 18 | 18 | 32512 | 17 | 17 | 17 | 17 | 17 | 30464 |
| 1575 | | 30464 | | | | | | | | | 18 | 18 | 18 | 18 | 32512 | 17 | 17 | 17 | 17 | 17 | 30464 |
| 1576 | | 31488 | | | | | | | | | 18 | 18 | 18 | 18 | 32512 | 18 | 17 | 17 | 17 | 18 | 34560 |
| 1577 | | 32512 | | | | | | | | | 18 | 18 | 18 | 18 | 32512 | 18 | 17 | 17 | 17 | 18 | 34560 |
| 1578 | 2048 | 34560 | | | | | | | | | | | | | 18 | 17 | 17 | 17 | 18 | 34560 | |
| 1579 | | 36608 | | | | | | | | | | | | | 18 | 18 | 17 | 17 | 18 | 36608 | |
| 1580 | | 38656 | | | | | | | | | | | | | 18 | 18 | 18 | 17 | 18 | 38656 | |
| 1581 | | 40704 | | | | | | | | | | | | | 18 | 18 | 18 | 18 | 18 | 40704 | |

2.5.4 Memory upgrades

Memory upgrades can be ordered and enabled by LIC, upgrading the DIMM cards, adding DIMM cards, or adding a CPC drawer.

For a model upgrade that results in the addition of a CPC drawer, the minimum memory increment is added to the system. Each CPC drawer has a minimum physical memory size of 480 GB.

During a model upgrade, adding a CPC drawer is a concurrent operation. Adding physical memory to the added drawer is also concurrent. If all or part of the added memory is enabled for use, it might become available to an active LPAR if the partition includes defined reserved storage. (For more information, see 3.7.3, “Reserved storage” on page 141.) Alternatively, the added memory can be used by a defined LPAR that is activated after the memory is added.

Note: Memory downgrades within a z15 are not supported. Feature downgrades (removal of a CPC quantity feature) are not supported.

2.5.5 Drawer replacement and memory

With Enhanced Drawer Availability (EDA), which is supported for z15 T01, sufficient resources must be available to accommodate resources that are rendered unavailable when a CPC drawer is removed for upgrade or repair. For more information, see 2.7.1, “Redundant I/O interconnect” on page 71.

Removing a CPC drawer often results in removing active memory. With the flexible memory option, removing the affected memory and reallocating its use elsewhere in the system is possible. For more information, see 2.5.7, “Flexible Memory Option”. This process requires more available memory to compensate for the memory that is lost with the removal of the drawer.

2.5.6 Virtual Flash Memory

IBM Virtual Flash Memory (VFM) FC 0643 replaces the Flash Express features (0402 and 0403) that were available on the IBM z13s. It offers up to 6.0 TB of virtual flash memory in 512 GB (0.5 TB) increments for improved application availability and to handle paging workload spikes.

No application changes are required to change from IBM Flash Express to VFM. Consider the following points:

- ▶ Dialed memory + VFM = total hardware plugged
- ▶ Dialed memory + VFM + Flex memory option = total hardware plugged
- ▶ VFM is offered in 0.5 TB increment size; VFM for z15 is FC 0643 - Min=0, Max=12

VFM is designed to help improve availability and handling of paging workload spikes when z/OS V2.1, V2.2, V2.3, or V2.4 is run. With this support, z/OS is designed to help improve system availability and responsiveness by using VFM across transitional workload events, such as market openings and diagnostic data collection. z/OS is also designed to help improve processor performance by supporting middleware use of pageable large (1 MB) pages.

VFM can also be used by coupling facility images to provide extended capacity and availability for workloads that use IBM WebSphere MQ Shared Queues structures. The use of VFM can improve availability by reducing latency from paging delays that can occur at the start of the workday or during other transitional periods. It is also designed to help eliminate delays that can occur when collecting diagnostic data.

VFM can help organizations meet their most demanding service level agreements and compete more effectively. VFM is easy to configure in the LPAR Image Profile and provides rapid time to value.

2.5.7 Flexible Memory Option

With the Flexible Memory Option, more physical memory is supplied to support the activation of the actual purchased memory entitlement in a single CPC drawer that is out of service during activation (POR), or in a scheduled concurrent drawer upgrade (memory add) or drawer maintenance (n+1 repair) with the use of enhanced drawer availability.

When you order memory, you can request extra flexible memory. The extra physical memory, if required, is calculated by the configuration and priced accordingly.

The hardware required is pre-plugged based on a target capacity specified by the customer. This pre-plugged hardware is enabled by using an LICCC order that is placed by the customer when they determine more memory capacity is needed.

The flexible memory sizes that are available for the z15 T01 are listed in Table 2-9.

Table 2-9 Flexible memory offering

| Feature Code | Increment | Minimum | Maximum |
|--------------|---------------------------------------|---------|---------|
| 1951 | 32GB Flex Memory | 0 | 250 |
| 1952 | 64GB Flex Memory | 0 | 250 |
| 1953 | 256GB Flex Memory | 0 | 250 |
| 1954 | 64GB Virtual Flash Memory Flex Memory | 0 | 1 |

2.6 Reliability, availability, and serviceability

IBM Z servers continue to deliver enterprise class RAS with IBM z15. The main philosophy behind RAS is about preventing or tolerating (masking) outages. It is also about providing the necessary instrumentation (in hardware, LIC and microcode, and software) to capture or collect the relevant failure information to help identify an issue without requiring a re-creation of the event. These outages can be planned or unplanned. Planned and unplanned outages can include the following situations (examples are not related to the RAS features of IBM Z servers):

- ▶ A planned outage because of the addition of physical processor capacity or memory
- ▶ A planned outage because of the addition of I/O cards
- ▶ An unplanned outage because of a failure of a power supply
- ▶ An unplanned outage because of a memory failure

The IBM Z hardware has decades of intense engineering behind it, which results in a robust and reliable platform. The hardware has many RAS features that are built into it. For more information, see Chapter 9, “Reliability, availability, and serviceability” on page 383.

2.6.1 RAS in the CPC memory subsystem

Patented error correction technology in the memory subsystem continues to provide the most robust error correction from IBM to date. Two full DRAM failures per rank can be spared and a third full DRAM failure can be corrected.

DIMM level failures, including components, such as the memory controller application-specific integrated circuit (ASIC), power regulators, clocks, and system board, can be corrected. Memory channel failures, such as signal lines, control lines, and drivers and receivers on the MCM, can be corrected.

Upstream and downstream data signals can be spared by using two spare wires on the upstream and downstream paths. One of these signals can be used to spare a clock signal line (one upstream and one downstream). The following improvements were also implemented in the z15 server:

- ▶ No cascading of memory DIMMs
- ▶ Independent channel recovery
- ▶ Double tabs for clock lanes
- ▶ Separate replay buffer per channel

- ▶ Hardware driven lane soft error rate (SER) and sparing

2.6.2 General z15 T01 RAS features

The z15 T01 server includes the following RAS features:

- ▶ The z15 T01 server provides a true N+2 pumps and N+1 fans for the cooling function for the radiator-cooled (air-cooled) model and N+1 (fully redundant) cooling function for the water-cooled model.
- ▶ The Power/Thermal Subsystem is new for z15 T01. It uses switchable, intelligent Power Distribution Units (PDUs) or Bulk Power Assemblies.
- ▶ Redundant (N+1), number of PDUs (up to four pairs, configuration dependent) or BPAs (also configuration-dependent - 2 or 4)
- ▶ CPC drawer is packaged to fit in the 19-inch frame. CPC drawer power and cooling includes the following components:
 - PSUs: AC to 12V bulk/standby (N+1 redundant). PSU fans are not separate FRUs CPC drawers have 3 PSUs for BPA-based configurations or 4 PSUs for PDU-based configurations.
 - POLs: N+2 Phase and Master Redundant and are available in quantities of up to 6.
 - VRM sticks⁷: Derivative of z13s design (N+2 Phase and Master redundancy) and are available in quantities of 7.
 - Power Control Card: New power control card to control CPC fans (N+1 redundant) and are available in quantities of 2.
 - Fans: Drawer has five fans and are N+1 redundant.
 - FSP/OSCs: Redundant (N+1).
- ▶ PU SCMs are all water-cooled. The SC SCM is air-cooled and uses a heat sink (same as z14 M0x).
- ▶ The PCIe+ I/O drawer power supplies for the z15 server are also based on the N+1 design. The second power supply can maintain operations and avoid an unplanned outage of the system.
- ▶ N+2 redundant environmental sensors (ambient temperature, relative humidity, and air density⁸).

The internal intelligent Power Distribution Unit (iPDU, or PDU) provide the following capabilities:

- ▶ Individual outlet control by way of Ethernet:
 - Provide a System Reset capability
 - Power cycle an SE if a hang occurs
 - Verify a power cable at installation
- ▶ System Reset Function:
 - No EPO switch is available on the z15. This function provides a means to put a server into a known state similar to past total power reset.

⁷ Voltage Regulator Module stick converts the DC bulk power that is delivered by the PSUs (12V) into localized low voltage that is used by the installed components (for example, PU SCMs, SC SCM, memory DIMMs, and other circuitry).

⁸ The air density sensor measures air pressure and is used to control blower speed.

- This function does not provide the option to power down and keep the power down to the system. The power must be unplugged or the customer-supplied power is turned off at the panel.
- ▶ Other characteristics:
 - PDU Firmware can be concurrently updated
 - Concurrently repairable
 - Power redundancy check
- ▶ Cable verification test by way of PDU:
 - By power cycling individual iPDU outlets, the system can verify proper cable connectivity
 - Power cable test runs during system Power On
 - Runs at system Installation and at every system Power On until the test passes
- ▶ PCIe service enhancements:
 - Mandatory end-to-end cyclic redundancy check (ECRC)
 - Customer operation code separate from maintenance code
 - Native PCIe firmware stack that is running on the integrated firmware processor (IFP) to manage isolation and recovery

The power service and control network (PSCN) is used to control and monitor the elements in the system and include the following components:

- ▶ Ethernet Top of Rack (TOR) switches provide the internal PSCN connectivity:
 - Switches are redundant (N+1)
 - Concurrently maintainable
 - Each switch has one integrated power supply
 - FSPs are cross wired to the Ethernet switches
- ▶ Redundant SEs

Each SE has two power supplies (N+1) and input power is cross-coupled from the PDUs.
- ▶ Concurrent CPC upgrades

CPC1 to (CPC1 + CPC2) and (CPC1 + CPC2) to (CPC1+CPC2+CPC3) - provided CPC1 Reserve and/or CPC2 Reserve features are part of the initial system order (FC 2271 and/or FC 2272)
- ▶ All PCIe+ I/O drawer MESs are concurrent
- ▶ All LICC model changes are concurrent

System Recovery Boost

With z15, a new feature was built into the system to enable restoration of service from, and catch up after, planned and unplanned outages faster than on any previous Z systems with no extra software cost. It also provides faster changeover between GDPS and non-GDPS systems (in a GDPS configured environment). Restoration of service involved speeding up IPL and shutdown operations of an image (LPAR).

Important: The base System Recovery Boost capability is built into z15 firmware and does not require ordering of extra features. For Z15 T01, System Recovery Boost Upgrade (consisting of FC 9930 and FC 6802) is an optional, orderable feature that provides more temporary zIIP capacity for use during boost periods. Consider the following points:

- ▶ FC 9930 is *not* required to use the base System Recovery Boost capability.
- ▶ FC 9930 AND FC 6802 are needed *only* if more zIIP temporary capacity is required.

Functionality for system recovery boost includes:

- ▶ General Purpose Processor (CP) capacity boost using zIIPs, which provides a boost in processor capacity and parallelism and makes all available processor capacity in the boosting images (LPARs) available for processing any kind of work (zIIPs run CP workload) during the boost period⁹ (operating system-dependent) by:
 - Using the client’s already-entitled CPs and zIIPs
 - System Recovery Boost Upgrade requires optional contract feature (FC 9930) and the (priced) feature (FC 6802) providing on-demand zIIP processor capacity. System Recovery Boost Upgrade draws upon some of the unused but available processor resources on the machine (inactive PUs) to provide more zIIPs to be used for general-purpose workload dispatch during boost period.
- ▶ Speed boost: On subcapacity machine models, provides a boost in processor capacity by running the CPs at full-capacity speed for the boosting images (LPARs) during the boost period (operating system dependent).
- ▶ Expedited GDPS reconfiguration: Expediting and parallelizing GDPS reconfiguration actions that can be part of the client’s restart, reconfiguration, and recovery process (requires GDPS code update and z15 CPCs).

IBM z15 servers continue to deliver robust server designs through new technologies, hardening both new and classic redundancy. For more information, see Chapter 9, “Reliability, availability, and serviceability” on page 383.

⁹ For z/OS, boost period is 60 minutes for IPL and 30 minutes for shutdown

2.7 Connectivity

Connections to PCIe+ I/O drawers and Integrated Coupling Adapters are driven from the CPC drawer fanout cards. These fanouts are installed in the rear of the CPC drawer.

Figure 2-25 shows the location of the fanout slots. Each slot is identified with a location code (label) of LGxx.

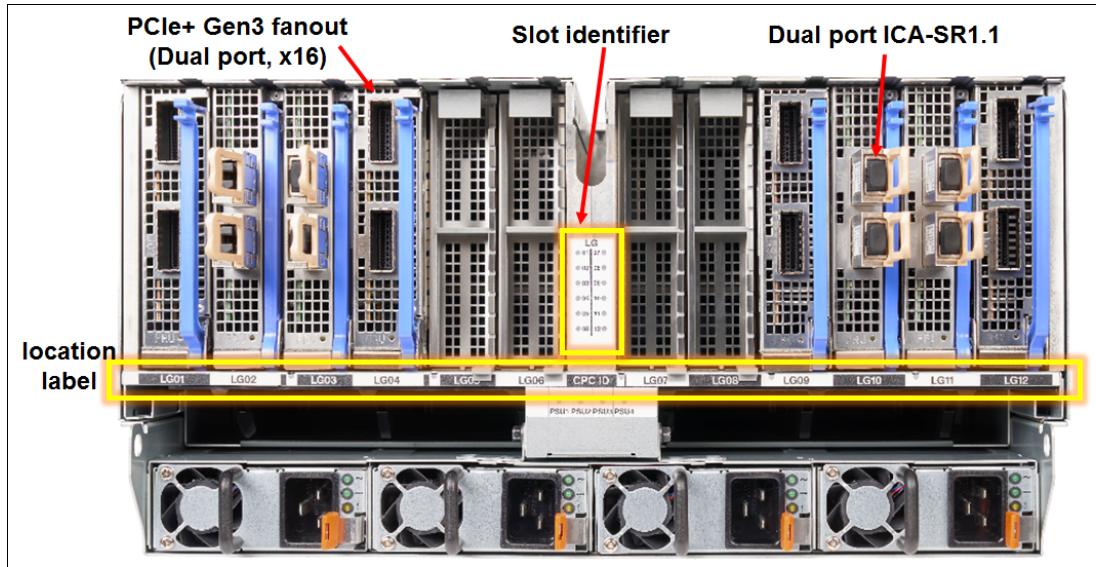


Figure 2-25 Fanout locations in the CPC drawer

Up to 12 PCIe fanouts (LG01 - LG12) can be installed in each CPC drawer.

A fanout can be repaired concurrently with the use of redundant I/O interconnect. For more information, see 2.7.1, “Redundant I/O interconnect” on page 71.

The following types of fanouts are available:

- ▶ A new PCIe+ Generation 3 dual port fanout card: This fanout provides connectivity to the PCIe switch cards in the PCIe+ I/O drawer.
- ▶ A new Integrated Coupling Adapter (ICA SR1.1): This adapter provides coupling connectivity to z15, z14, and z13 servers.

When configured for availability, the channels and coupling links are balanced across CPC drawers. In a system that is configured for maximum availability, alternative paths maintain access to critical I/O devices, such as disks and networks. The CHPID Mapping Tool can be used to assist with configuring a system for high availability.

Enhanced (CPC) drawer availability (EDA) allows a single CPC drawer in a multidrawer CPC to be removed and reinstalled (serviced) concurrently for an upgrade or a repair. Removing a CPC drawer means that the connectivity to the I/O devices that are connected to that CPC drawer is lost. To prevent connectivity loss, the redundant I/O interconnect feature allows you to maintain connection to critical I/O devices (except for ICA SR1.1) when a CPC drawer is removed.

2.7.1 Redundant I/O interconnect

Redundancy is provided for PCIe I/O interconnects.

The PCIe+ I/O drawer supports up to 16 PCIe features, which are organized in two hardware domains (for each drawer). The infrastructure for the fanout to I/O drawers and external coupling is shown in Figure 2-26.

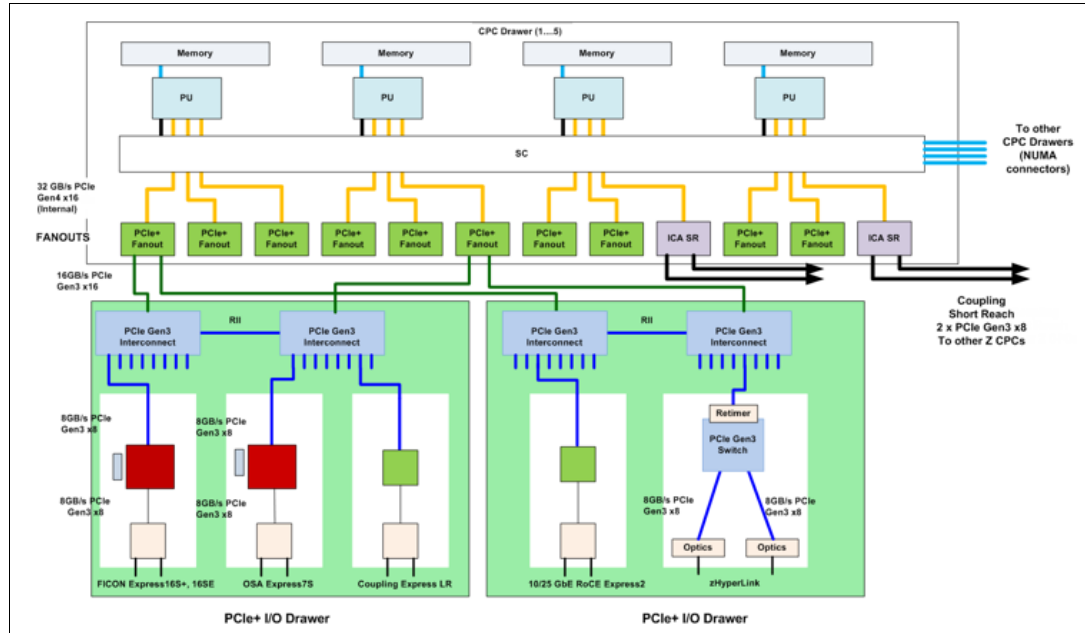


Figure 2-26 Infrastructure for PCIe+ I/O drawer (system with two PCIe+ I/O drawers)

The new PCIe+ Gen3 fanout cards are used to provide the connection from the PU SCM PCIe Bridge Unit (PBU), which uses split the PCIe Gen4 (@32GBps) processor busses into two PCIe Gen3 x16 (@16 GBps) interfaces to the PCIe switch card in the PCIe+ I/O drawer.

The PCIe switch card spreads the x16 PCIe bus to the PCIe I/O slots in the domain.

In the PCIe+ I/O drawer, the two PCIe switch cards (LG06 and LG16, see Figure 2-27) provide a backup path (Redundant I/O Interconnect - RII) for each other through the passive connection in the PCIe+ I/O drawer backplane.

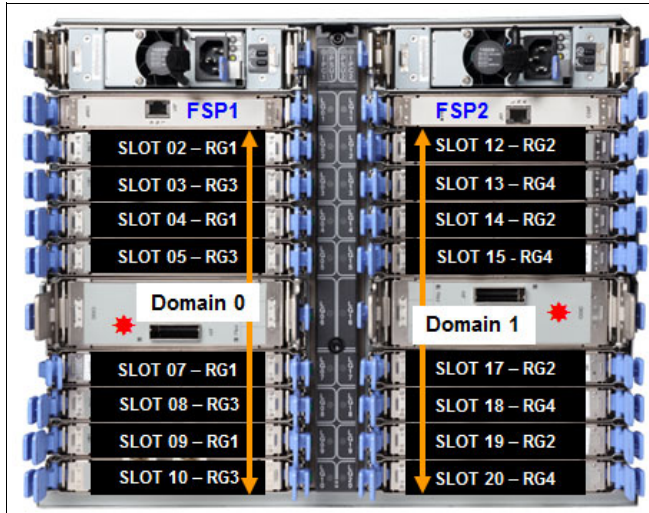


Figure 2-27 PCIe+ I/O drawer locations

During a PCIe fanout or cable failure, all 16 PCIe cards in the two domains can be driven through a single PCIe switch card (see Figure 2-28).

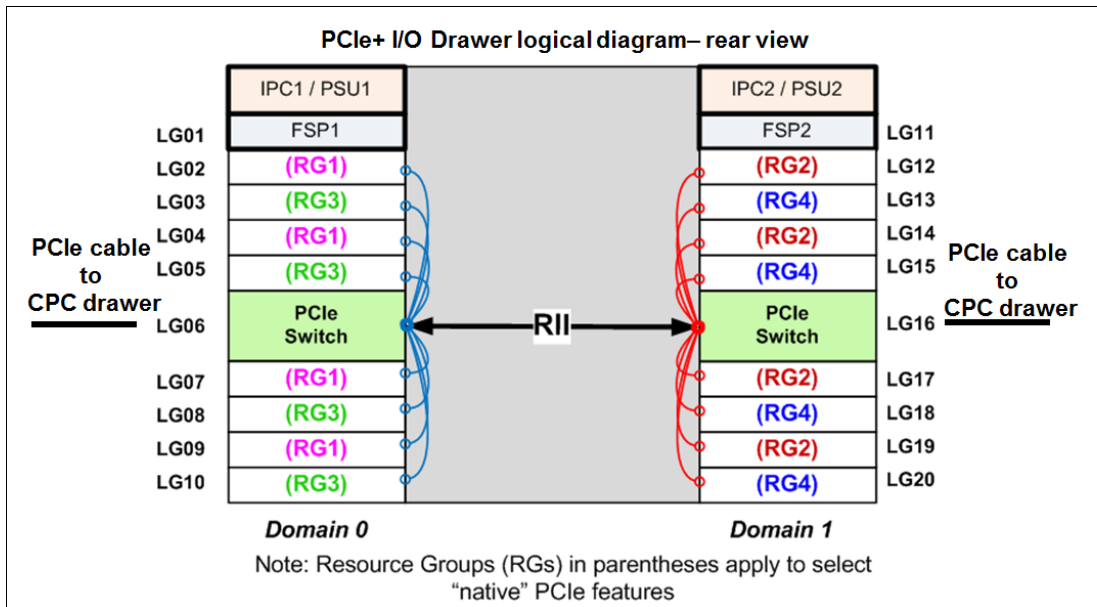


Figure 2-28 Redundant I/O Interconnect

To support Redundant I/O Interconnect (RII) between domain pair 0 and 1, the two interconnects to each pair must be driven from two different PCIe fanouts. Normally, each PCIe interconnect in a pair supports the eight features in its domain. In backup operation mode, one PCIe interconnect supports all 16 features in the domain pair.

Note: The PCIe Interconnect (switch) adapter *must* be installed in the PCIe+ I/O drawer to maintain the interconnect across I/O domains. If the adapter is removed (for a service operation), the I/O cards in that domain (up to eight) become unavailable.

2.7.2 Enhanced drawer availability (EDA)

With EDA, the effect of CPC drawer replacement is minimized. In a multiple CPC drawer system, a single CPC drawer can be concurrently removed and reinstalled for an upgrade or repair. Removing a CPC drawer without affecting the workload requires sufficient resources in the remaining CPC drawer.

Before removing the CPC drawer, the contents of the PUs and memory of the drawer must be relocated. PUs must be available on the remaining CPC drawers to replace the deactivated drawer. Also, sufficient redundant memory must be available if no degradation of applications is allowed. To ensure that the CPC configuration supports removal of a CPC drawer with minimal effect on the workload, consider the flexible memory option. Any CPC drawer can be replaced, including the first CPC drawer that initially contains the HSA.

Removal of a CPC drawer also removes the CPC drawer connectivity to the I/O drawers, PCIe I/O drawers, and coupling links. The effect of the removal of the CPC drawer on the system is limited by the use of redundant I/O interconnect. (For more information, see 2.7.1, “Redundant I/O interconnect” on page 71.) However, all ICA SR1.1 links that are installed in the removed CPC drawer must be configured offline.

If the enhanced drawer availability and flexible memory options are *not* used when a CPC drawer must be replaced, the memory in the failing drawer is also removed. This process might be necessary during an upgrade or a repair action. Until the removed CPC drawer is replaced, a power-on reset of the system with the remaining CPC drawers is supported. The CPC drawer can then be replaced and added back into the configuration concurrently.

2.7.3 CPC drawer upgrade

All fanouts that are used for I/O and coupling links are rebalanced concurrently as part of a CPC drawer addition to support better RAS characteristics.

2.8 Model configurations

When a z15 is ordered, the PUs are characterized according to their intended usage. The PUs can be ordered as any of the following items:

- | | |
|------------|---|
| CP | The processor is purchased and activated. PU supports running the z/OS, z/VSE, z/VM, z/TPF, and Linux on Z ¹⁰ operating systems. It can also run Coupling Facility Control Code. |
| IFL | The Integrated Facility for Linux (IFL) is a processor that is purchased and activated for use by z/VM for Linux guests and Linux on Z ¹⁰ operating systems. |

¹⁰ The KVM hypervisor is part of supported Linux on Z distributions.

- Unassigned IFL** A processor that is purchased for future use as an IFL. It is offline and cannot be used until an upgrade for the IFL is installed. It does not affect software licenses or maintenance charges.
- ICF** An internal coupling facility (ICF) processor that is purchased and activated for use by the Coupling Facility Control Code.
- zIIP** An “Off Load Processor” for workload that supports applications such as Db2 and z/OS Container Extensions. It can also be used for “System Recovery Boost” on page 68
- Additional SAP** An optional processor that is purchased and activated for use as SAP (System Assist Processor).

A minimum of one PU that is characterized as a CP, IFL, or ICF is required per system. The maximum number of characterizable PUs is 190. The maximum number of zIIPs is up to twice the number of PUs that are characterized as CPs.

The following components are present in the z15 server, but they are not part of the PUs that clients purchase and require no characterization:

- ▶ SAP to be used by the channel subsystem. The number of predefined SAPs depends on the z15 model.
- ▶ One IFP, which is used in the support of designated features and functions, such as RoCE (all features), Coupling Express LR, Internal Shared Memory (ISM) SMC-D, and other management functions.
- ▶ Two spare PUs, which can transparently assume any characterization during a permanent failure of another PU.

The z15 uses features to define the number of PUs that are available for client use in each configuration. The models are listed in Table 2-10.

Table 2-10 z15 Processor Configurations

| Feature | CPC Drawers | PUs per drawer | Active PUs | | | | zIIP | IFP | Opt SAPs | Base SAPs | Spares |
|---------|-------------|----------------|------------|-------|-------|-------|-------|-----|----------|-----------|--------|
| | | | CPs | IFLs | ICFs | uIFLs | | | | | |
| Max34 | 1 | 41 | 0-34 | 0-34 | 0-34 | 0-33 | 0-22 | 1 | 0-8 | 4 | 2 |
| Max71 | 2 | 41 | 0-71 | 0-71 | 0-71 | 0-70 | 0-46 | 1 | 0-8 | 8 | 2 |
| Max108 | 3 | 41 | 0-108 | 0-108 | 0-108 | 0-107 | 0-70 | 1 | 0-8 | 12 | 2 |
| Max145 | 4 | 41 | 0-145 | 0-145 | 0-145 | 0-144 | 0-96 | 1 | 0-8 | 16 | 2 |
| Max190 | 5 | 43 | 0-190 | 0-190 | 0-190 | 0-189 | 0-126 | 1 | 0-8 | 22 | 2 |

- ▶ Not all PUs available on a model are required to be characterized with a feature code. Only the PUs purchased by a customer are identified with a feature code.
- ▶ zIIP maximum quantity for new build systems follows the 2:1 ratio. It might be greater if present during MES upgrades.
- ▶ All PU conversions can be performed concurrently.

A *capacity marker* identifies the number of CPs that were purchased. This number of purchased CPs is higher than or equal to the number of CPs that is actively used. The capacity marker marks the availability of purchased but unused capacity that is intended to be used as CPs in the future. This status often is present for software-charging reasons.

Unused CPs are not a factor when establishing the millions of service units (MSU) value that is used for charging monthly license charge (MLC) software, or when charged on a per-processor basis.

2.8.1 Upgrades

Concurrent upgrades of CPs, IFLs, ICFs, zIIPs, or SAPs are available for the z15 server. However, concurrent PU upgrades require that more PUs are installed but not activated.

Spare PUs are used to replace defective PUs. Two spare PUs always are on a z15 T01 server. In the rare event of a PU failure, a spare PU is activated concurrently and transparently and is assigned the characteristics of the failing PU.

If an upgrade request cannot be accomplished within the configuration, a hardware upgrade is required.

The following upgrade paths for the z15 are shown in Figure 2-29:

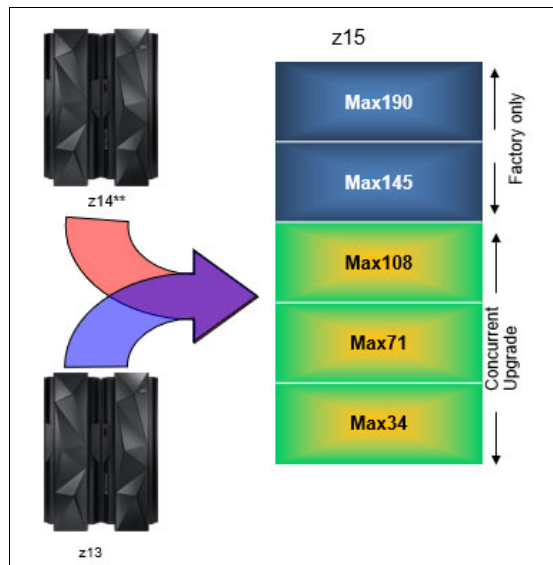


Figure 2-29 z15 T01 system upgrade paths

- ▶ z15 T01 to z15 T01 upgrades:
 - Max34 to Max71, Max108 are concurrent
 - No upgrade to Max154 or Max190 (these features are Factory built only)
 - More I/O drawers can be added based on available space in current frames or I/O expansion frames
- ▶ Any z13 (M/T 2964) to z15 T01:
 - Feature conversion of installed zAAPs to zIIPs (default) or another processor type
 - For installed OnDemand Records, change temporary zAAPs to zIIPs. Stage the record
- ▶ Any z14 (M/T 3906) to z15 T01

Note: The z15 CPC cannot be member in an Ensemble managed by Unified Resource Manager.

2.8.2 Model capacity identifier

To recognize how many PUs are characterized as CPs, the Store System Information (STSI) instruction returns a Model Capacity Identifier (MCI). The MCI determines the number and speed of characterized CPs. Characterization of a PU as an IFL, ICF, or zIIP is not reflected in the output of the STSI instruction because characterization has no effect on software charging. For more information about STSI output, see “Processor identification” on page 377.

The following distinct model capacity identifier ranges are recognized (one for full capacity and three for granular capacity):

- ▶ For full-capacity engines, model capacity identifiers 701 - 7J0 are used. They express capacity settings for 1 - 190 characterized CPs.
- ▶ Three model capacity identifier ranges offer a unique level of granular sub-capacity engines at the low end. They are available when no more than 34 CPs are characterized. These three subcapacity settings are applied to up to 34 CPs, which combined offer 102 more capacity settings. For more information, see “Granular capacity”.

Granular capacity

The z15 T01 server offers 102 capacity settings (granular capacity) for up to 34 CPs.. When subcapacity settings are used, other PUs beyond 34 can be characterized only as specialty engines. For models with more that 34 CPs, all CPs are running at full capacity (7xx).

The three defined ranges of subcapacity settings have model capacity identifiers numbered 401- 434, 501 - 534, and 601 - 634.

Consideration: All CPs have the same capacity identifier. Specialty engines (IFLs, zIIPs, and ICFs) operate at full speed.

List of model capacity identifiers

Regardless of the number of CPC drawers, a configuration with one characterized CP is possible, as listed in Table 2-11.

Table 2-11 Model capacity identifiers

| Feature | Model capacity identifier |
|---------|--|
| Max34 | 701 - 734, 601 - 634, 501 - 534, and 401 - 434 |
| Max71 | 701 - 771, 601 - 634, 501 - 534, and 401 - 434 |
| Max108 | 701 - 7A8, 601 - 634, 501 - 534, and 401 - 434 |
| Max145 | 701 - 7E5, 601 - 634, 501 - 534, and 401 - 434 |
| Max190 | 701 - 7J0, 601 - 634, 501 - 534, and 401 - 434 |

For more information about temporary capacity increases, see Chapter 8, “System upgrades” on page 333.

2.9 Power and cooling

The z15 T01 power and cooling system is a change from previous systems because the system is packaged in an industry standard 19-inch form factor frame for all the internal system elements. The configuration can be 1 - 4 frames. Consider the following points:

- ▶ The power subsystem is based on the following offerings:
 - Power Distribution Units (PDUs) that are mounted at the rear of the system in pairs
 - Bulk Power Assembly (BPA), as with previous systems
- ▶ The system uses 3-phase power:
 - Low voltage 4 wire “Delta”
 - High voltage 5 wire “Wye”
- ▶ No EPO (emergency power off) switch is used.
z15 has a support element task to simulate the EPO function (only used when necessary to do a System Reset Function).
- ▶ No DC input feature is available.
- ▶ The air-cooled z15 T01 server now has a radiator unit (RU) N+2 design for the pumps and blowers.
- ▶ The water-cooled system option for the z15 T01 server is only available with BPA-based systems.
- ▶ No separate Top Exit Power feature is available because the 19-inch frame is capable of top or bottom exit of power. All line cords are 4.26 meters (14 feet). Combined with the Top Exit I/O Cabling feature, more options are available when you are planning your computer room cabling.
- ▶ The new PSCN structure uses industry standard Ethernet switches (up to four) that replace the previous IBM System Control Hubs (SCHs).

2.9.1 PDU-based configurations

The IBM Z systems operate with redundant power infrastructure. The z15 T01 is designed with a new power infrastructure that is based on intelligent (PDUs that are mounted vertically on the rear side of the 19-inch racks and Power Supply Units for the internal components. The PDU configuration supports radiator-cooled systems only.

The PDUs are controlled by using an Ethernet port and support the following input:

- ▶ 3-phase 200 - 240 V AC (wired as “Delta”)
- ▶ 3-phase 380 - 415 V AC (wired as “Wye”)

The power supply units convert the AC power to DC power that is used as input for the Points of Load (POLs) in the CPC drawer and the PCIe+ I/O drawers.

The power requirements depend on the number of CPC drawers (1 - 5), number of PCIe I/O drawers (0 - 12) and I/O features that are installed in the PCIe I/O drawers.

PDUs are installed and serviced from the rear of the frame. Unused power ports should never be used by any external device.

A schematic view of a maximum configured system with PDU-based power is shown in Figure 2-30.

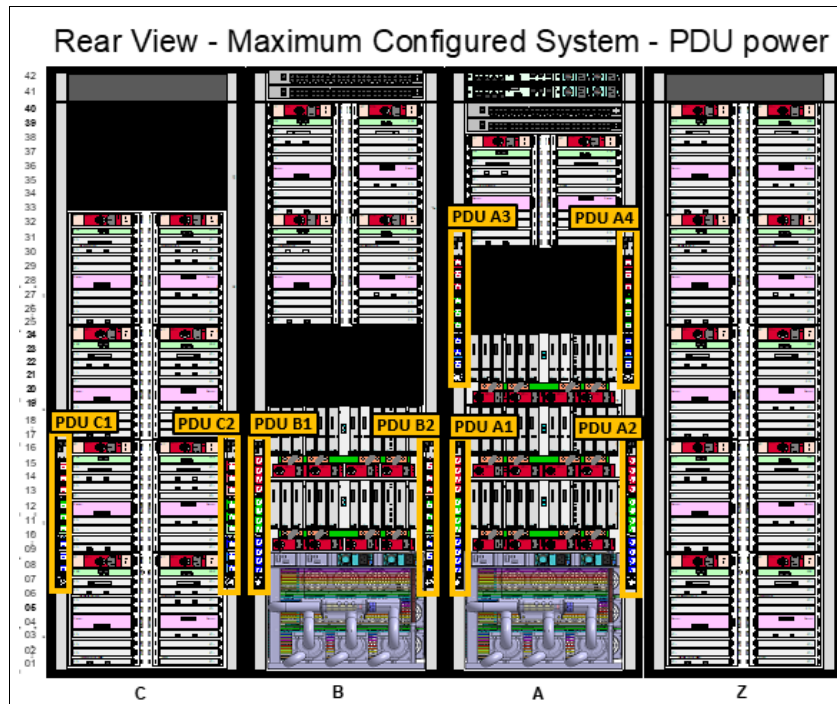


Figure 2-30 Rear view- maximum configured system PDU power

Each PDU installed requires a customer supplied power feed. The number of power cords that are required depends on the system configuration.

Note: For initial installation, all power sources are required to run the system checkout diagnostics successfully.

PDUs are installed in pairs. A system can have 2, 4, 6, or 8 PDUs, depending on the configuration. Consider the following points:

- ▶ Paired PDUs are A1/A2, A3/A4, B1/B2, and C1/C2.
- ▶ From the rear of the system, the odd-numbered PDUs are on the left side of the rack, and the even-numbered PDUs are on the right side of the rack.
- ▶ The total loss of one PDU in a pair has no effect on the system operation.

Components that plug into the PDUs for redundancy (using two power cords) include the following features:

- ▶ CPC Drawers, PCIe+ I/O drawers, Radiators, and Support Elements
- ▶ The redundancy for each component is achieved by plugging the power cables into the paired PDUs.

For example, the top Support Element (1), has one power supply plugged into PDU A1 and the second power supply plugged into the paired PDU A2 for redundancy.

Note: Customer power sources should always maintain redundancy across PDU pairs; that is, one power source or distribution panel supplies power for PDU A1 and the separate power source or distribution panel supplies power for PDU A2.

As a best practice, connect the odd-numbered PDUs (A1, B1, C1, and D1) to one power source or distribution panel, and the even-numbered PDUs (A2, B2, C2, and D2) to a separate power source or distribution panel.

The frame count rules (number of frames) for z15 T01 are listed in Table 2-12.

Table 2-12 Frame count rules for z15 T01

| Frame Count | I/O drawers | | | | | | | | | | | | |
|-------------|-------------|---|---|---|---|---|---|---|---|---|----|----|----|
| CPC drawers | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 3 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | NA |
| 4 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 |
| 5 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 |

The number of CPC drawers and I/O drawers determines the number of racks in the system and the number of PDUs in the system.

The PDU/line cord rules (number of PDU/Cord pairs) for z15 T01 are listed in Table 2-13

Table 2-13 PDU/line cord rules (# PDU/Cord pairs) for z15 T01

| PDU/Linecord | I/O drawers | | | | | | | | | | | | |
|--------------|-------------|---|---|---|---|---|---|---|---|---|----|----|----|
| CPC drawers | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | NA |
| 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 |
| 5 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 |

2.9.2 BPA-based configurations

The Bulk Power Assembly configuration is the feature that allows customers to have truly redundant line cords with phase loss protection and IBF (Internal Battery Feature) support for frame level UPS power backup. Characteristics are similar to previous IBM Z systems, but BPA was redesigned to fit into the 19-inch frame.

No PDUs are installed in this configuration; instead, 1 or 2 pairs of BPAs are installed that house the bulk power distribution to all the components, the Bulk Power Regulators (BPR), and the optional Internal Battery Feature.

With BPA-based configurations, 3-phase balanced power feature is supported. BPA-based systems support radiator-cooled systems and water-cooled systems.

The BPAs support 2- or 4-line cords with a single universal type of 3-phase 200 - 480 V AC.

The power supply units convert the regulated high-voltage DC power (supplied by BPA) to DC power that is used as input for the Points of Load (POLs) in the CPC drawer and the PCIe+ I/O drawers.

The power requirements depend on the number of CPC drawers (1 - 5), number of PCIe I/O drawers (0 - 11), and I/O features that are installed in the PCIe+ I/O drawers.

A schematic view of a maximum configured system with BPA power is shown in Figure 2-31.

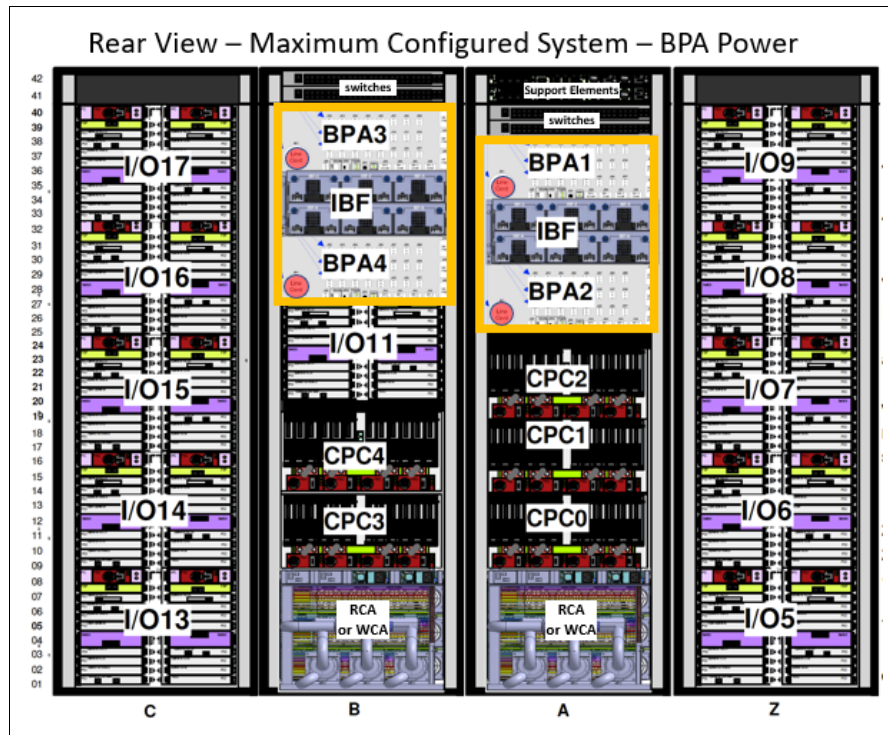


Figure 2-31 Rear view- maximum configured system BPA power

Each BPA installed requires a customer supplied power feed. The number of power cords that are required depends on the system configuration. The total loss of one BPA has no effect on system operation.

Note: For initial installation, all power sources are required to run the system checkout diagnostics successfully.

BPAs are installed in pairs (2 or 4), depending on the configuration. Consider the following points:

- ▶ BPA1 and BPA2 are always installed initially in Frame A
- ▶ BPA3 and BPA4 are installed to support more components for larger configurations (CPC and I/O drawers)
- ▶ BPA1/BPA2 and BPA3/BPA4 for paired redundancy
- ▶ From the rear of the system, the odd-numbered BPAs are above the even-numbered BPAs in both frames
- ▶ The total loss of one BPA in a pair has no effect on system operation

The following components plug into the BPAs X 2 for redundancy:

- ▶ CPC Drawers, I/O Drawers, Switches, Radiators, WCUs, and Support Elements
- ▶ The redundancy for each component plugs into the paired BPAs

For example, for the top Support Element 1, one power supply plugs into BPA1 and the second power supply plugs into the paired BPA2 for redundancy.

Note: Customer power sources should always maintain redundancy across BPA pairs; that is, one power source or distribution panel supplies power for BPA1 and the separate power source or distribution panel supplies power for BPA2.

As a best practice, connect the odd-numbered BPAs (BPA1 and BPA3) to one power source or distribution panel, and the even-numbered BPAs (BPA2 and BPA4) to a separate power source or distribution panel.

2.9.3 Internal Battery Feature

The Internal Battery Feature (IBF) is an option on the z15 T01 server. It is shown in Figure 2-31 on page 80 for air-cooled or water-cooled systems. The IBF provides a temporary local power source for maintaining system power (hold-up times depend on system configuration) in case of an external power source failure (“black-out” or “brown-out”).

Removal of IBF support^a: IBM z15 is planned to be the last IBM Z server to offer an Internal Battery Feature (IBF). As client data centers continue to improve power stability and uninterruptible power supply (UPS) coordination, IBM Z continues to innovate to help clients take advantage of common power efficiency and monitoring across their ecosystems. Additional support for data center power planning can be requested through your IBM Sales contact.

- a. Statements by IBM regarding its plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM.

The IBF further enhances the robustness of the power design, which increases power line disturbance immunity. It provides battery power to preserve processor data during a loss of power on all power feeds from the computer room. The IBF can hold power briefly during a brownout, or for orderly shutdown for a longer outage. For information about the hold times, which depend on the I/O configuration and amount of CPC drawers, see Chapter 11, “Environmentals” on page 453.

2.9.4 Power estimation tool

The power estimation tool for the z15 server allows you to enter your precise server configuration to obtain an *estimate* of power consumption. Log in to the Resource link with your user ID. Click **Planning** → **Tools** → **Power Estimation Tools**. Specify the quantity for the features that are installed in your system.

This tool estimates the power consumption for the specified configuration. The tool does *not* verify that the specified configuration can be physically built.

Tip: The exact power consumption for your system varies. The object of the tool is to estimate the power requirements to aid you in planning for your system installation. Actual power consumption after installation can be confirmed by using the HMC Monitors Dashboard task.

2.9.5 Cooling

The PU SCMs for z15 T01 are cooled by a cold plate that is connected to the internal water-cooling loop. The SC SCMs are air-cooled. In an air-cooled system, the radiator unit dissipates the heat from the internal water loop with air. The radiator unit provides improved availability with N+ 2 pumps and blowers. The WCUs are fully redundant in an N+1 arrangement for both the A-frame and the B-frame when present.

For all z15 T01 servers, the CPC drawer components (except for PU SCMs) and the PCIe+ I/O drawers are air cooled by redundant fans. Airflow of the system is directed from front (cool air) to the back of the system (hot air).

Radiator-cooled (air-cooled) models (FC 4033, FC 4035)

The z15 T01 PU SCMs in the CPC drawers are cooled by water. The internal closed water loop removes heat from PU SCMs by circulating water between the radiator heat exchanger and the cold plate that is mounted on the PU SCMs. For more information, see 2.9.6, “Radiator Cooling Unit”.

Although the PU SCMs are cooled by water, the heat is exhausted into the room from the radiator heat exchanger by forced air with blowers. At the system level, these z15 T01 are still air-cooled systems.

Water-cooled models (FC 4034, FC 4036)

z15 T01 servers are available as water-cooled systems. With WCU technology, z15 T01 servers can transfer most of the heat that they generate into the building’s chilled water, which effectively reduces the heat output to the computer room.

Unlike the radiator in air-cooled models, a WCU has two water loops: An internal closed water loop and an external (chilled) water loop. The external water loop connects to the client-supplied building’s chilled water. The internal water loop circulates between the WCU heat exchanger and the PU SCMs cold plates. The loop takes heat away from the PU SCMs and transfers it to the external water loop in the WCU’s heat exchanger. For more information, see 2.9.7, “Water-cooling unit” on page 85.

In addition to the PU SCMs, the internal water loop circulates through two heat exchangers that are in the path of the exhaust air in the rear of the frames. These heat exchangers remove approximately 60% - 65% of the residual heat from the I/O drawers, PCIe I/O drawers, the air-cooled logic in the CPC drawers, and the power enclosures. Almost two-thirds of the total heat that is generated can be removed from the room by the chilled water.

Air-cooled models or water-cooled models are chosen when ordering, and the corresponding equipment is factory-installed. An MES (conversion) from an air-cooled model to a water-cooled model and vice versa is not possible.

2.9.6 Radiator Cooling Unit

There are 1 - 2 Radiator Cooling Units (RCU) in the system: One in Frame A and one in Frame B in support of water cooling of the PU SCMs within the CPC drawers. The unit includes n+2 pumps and N+1 fans. Water loops to each drawer are directly delivered by way of hoses to each drawer from manifolds.

The RCU discharges heat from the internal frame water loop to the customer’s data center.

Each RCU provides cooling to PU SCMs with closed loop water within the respective frame. No connection to an external chilled water supply is required. For the z15 T01 server, the internal circulating water is conditioned water (BTA) that is added to the radiator unit during system installation with the Fill and Drain Tool (FC 3393).

Fill and Drain Tool

The Fill and Drain Tool (FDT) is a new design and is included with new z15 servers. The new FDT (FC 3393) is *not* compatible with previous systems. The FDT is used to provide the internal water at the installation and for maintenance, and to remove it at discontinuance.

The new design uses a pump for filling the system and FRUs, and a compressor for draining FRUs and discontinuance. The process is faster and more effective. The FDT is shown in Figure 2-32.

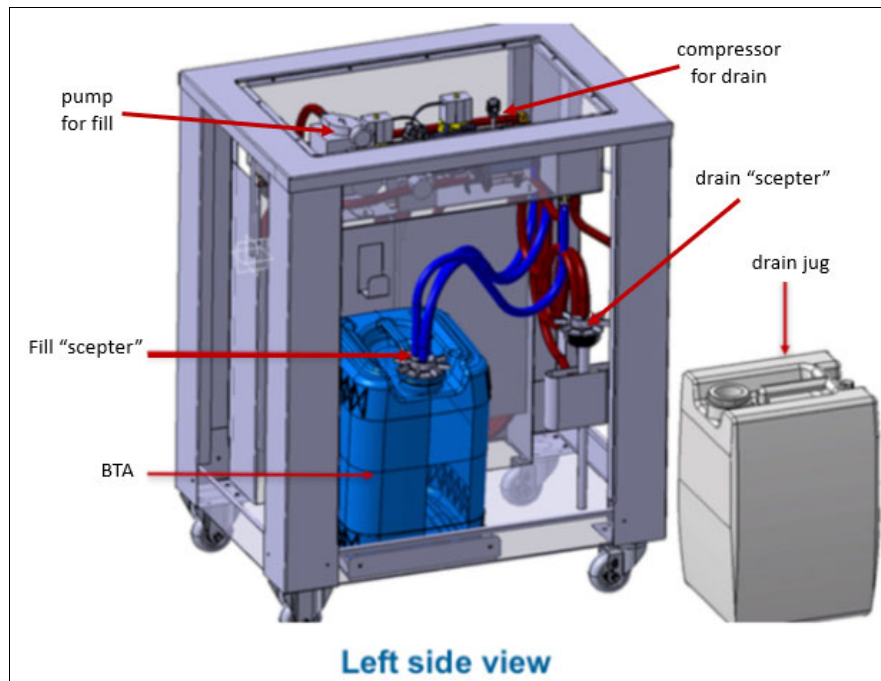


Figure 2-32 Fill Drain Tool (FDT)

The radiator cooling unit (RCU) contains three independent pump FRUs. The cooling capability is a redundant N+2 design, so a single working pump and blower can support the entire load. The replacement of one pump or blower can be done concurrently and does not affect performance.

Each radiator cooling unit contains up to five independent fan assemblies that can be concurrently serviced. The number of fans present depends on the number of CPC drawers installed in the frame.

The water pumps, manifold assembly, radiator assembly (which includes the heat exchanger), and fan assemblies are the main components of the z15 RCU, as shown in Figure 2-33.

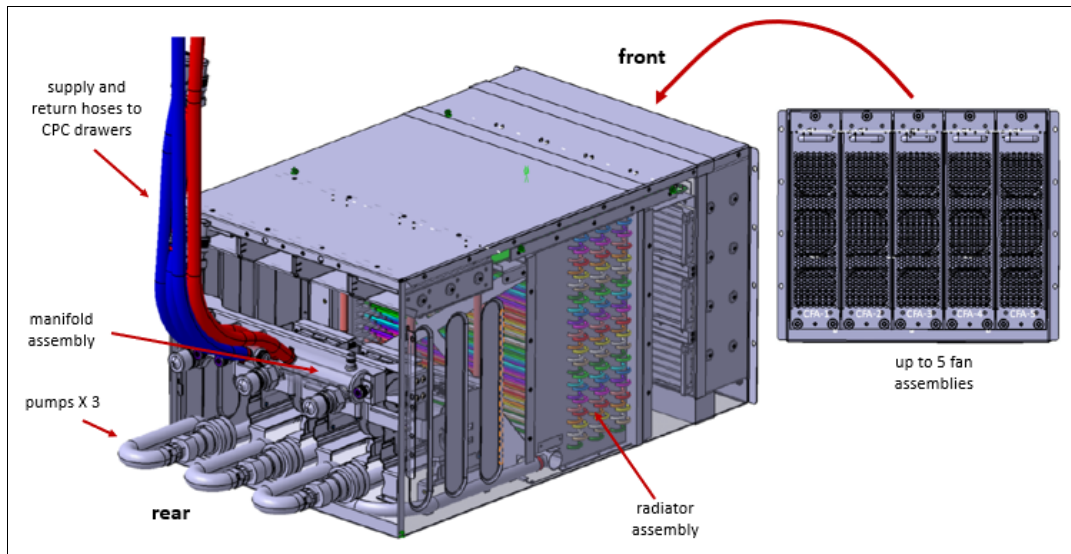


Figure 2-33 Radiator Cooling Unit (RCU)

The closed water loop in the radiator unit is shown in Figure 2-34. The warm water that is exiting from the PU SCMs cold plates enters pumps through a common manifold and is pumped through a heat exchanger where heat is extracted by the air flowing across the heat exchanger fins. The cooled water is then recirculated back into the PU SCMs cold plates.

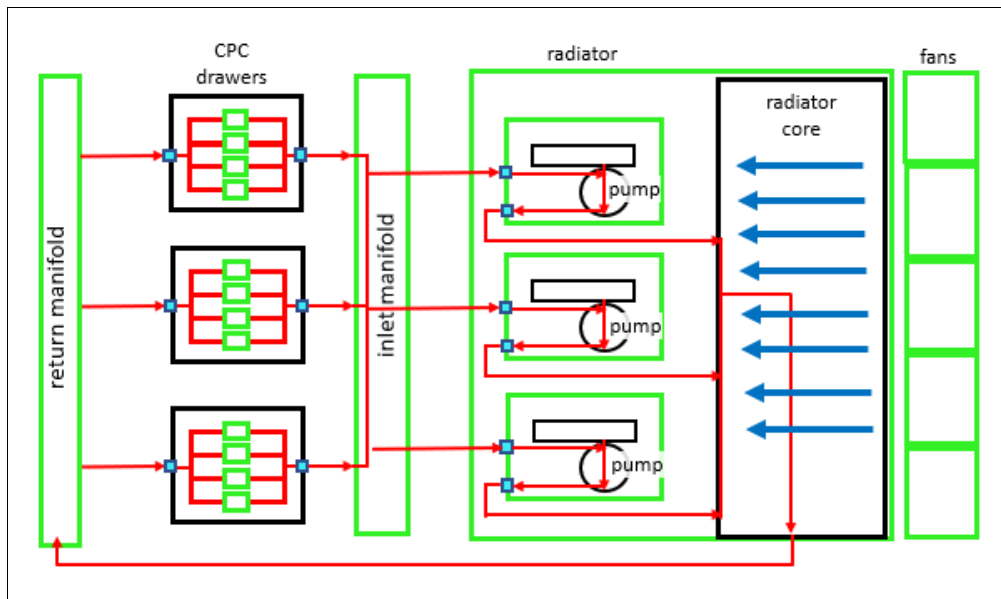


Figure 2-34 Radiator cooling system

2.9.7 Water-cooling unit

z15 T01 servers continue to provide the ability to cool systems with building-provided chilled water by using the WCU technology. The PU SCMs in the CPC drawers are cooled by internal closed loop water. The internal closed loop water exchanges heat with building-provided chilled water in the WCU heat exchanger. The source of the building's chilled water is provided by the client.

A WCU is shown in Figure 2-35.

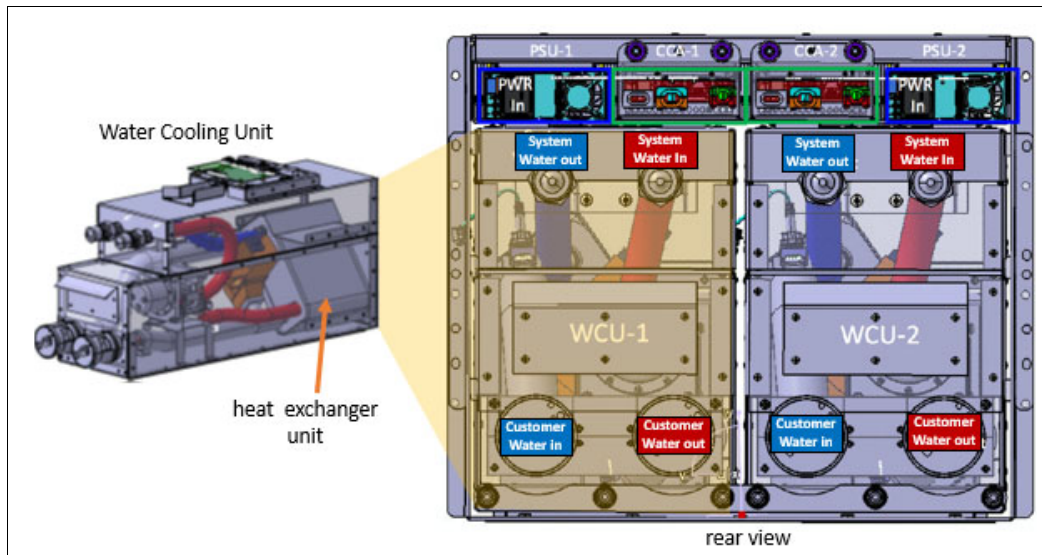


Figure 2-35 Water cooling unit

The water in the closed loop within the system exchanges heat with the continuous supply of building-provided chilled water. This water circulates between the PU SCMs cold plates and a heat exchanger within the WCU. Heat from the PU SCMs is transferred to the cold plates, where it is in turn transferred to the circulating system water (closed loop). The system water then dissipates its heat to the building-provided chilled water within the WCU's heat exchanger. The PU SCMs are cooled efficiently in this manner.

z15 T01 servers operate with two fully redundant WCUs in the A-frame and two fully redundant WCUs in the B-frame (when present), each on separate loops. These water-cooling units each have their own facility feed and return water connections. If water is interrupted to one of the units, the other unit picks up the entire load, and the server continues to operate without interruption. You must provide independent redundant water loops to the water-cooling units to obtain full redundancy.

The internal circulating water is conditioned water that is added to the radiator unit during system installation with the Fill and Drain Tool (FC 3393). The FDT is included with new z15 servers. The FDT is used to provide the internal water at the installation and for maintenance, and to remove it at discontinuance.

Indoor Heat Exchanger (IDX)

In z15 T01 servers, all water-cooled models have up to four IDX Indoor Heat Exchanger units that are installed on the rear of each frame that is present, as shown in Figure 2-36 on page 86. These units remove heat from the internal system water loop and internal air exits the server into the hot air exhaust aisle.

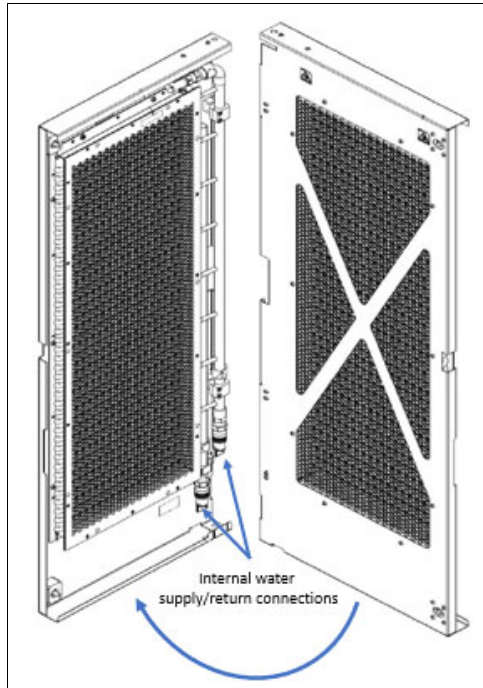


Figure 2-36 IDX Indoor Heat Exchanger

Consider the following points:

- ▶ IDX is an air-to-water heat exchange inside the frame
- ▶ Used to remove heat from the frame air stream and reject it to system water
- ▶ System water rejects heat to building chilled water
- ▶ Water turbulators are used to improve HX efficiency

In addition to the PU SCMs cold plates, the internal water loop circulates through heat exchangers that are mounted on each frame in the configuration. These exchangers are in the path of the exhaust air in the rear of the frames.

Depending on the configuration, these heat exchangers remove up to 93% of the residual heat from the I/O drawers, PCIe I/O drawer, the air-cooled logic in the CPC drawer, and the power enclosures. The goal is for two-thirds of the total heat that is generated to be removed from the room by the chilled water.

Figure 2-37 shows an example of various frame configurations. Every frame includes a heat exchanger when the water-cooling feature is ordered.

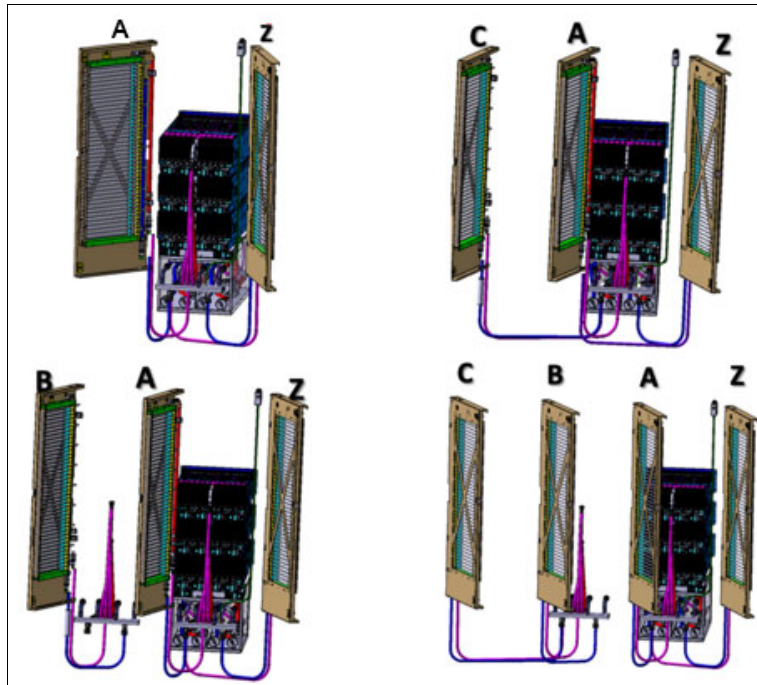


Figure 2-37 Rear door heat exchanger plumbing schematic

If one client water supply or one WCU fails, the remaining feed maintains PU SCM cooling for the frame. The WCUs and the associated drive card are concurrently replaceable. In addition, the heat exchangers can be disconnected and removed from the system concurrently.

2.10 Summary

All aspects of the z15 T01 structure are listed in Table 2-14

Table 2-14 System structure summary

| Description | Max34 | Max71 | Max108 | Max145 | Max190 |
|-------------------------------------|--------|--------|---------|---------|---------|
| Maximum number of characterized PUs | 34 | 71 | 108 | 145 | 190 |
| Number of CPs | 0 - 34 | 0 - 71 | 0 - 108 | 0 - 145 | 0 - 190 |
| Number of IFLs | 0 - 34 | 0 - 71 | 0 - 108 | 0 - 145 | 0 - 190 |
| Number of Unassigned IFLs | 0 - 33 | 0 - 70 | 0 - 107 | 0 - 144 | 0 - 189 |
| Number of ICFs | 0 - 34 | 0 - 71 | 0 - 108 | 0 - 145 | 0 - 190 |
| Number of zIIPs | 0 - 22 | 0 - 46 | 0 - 70 | 0 - 96 | 0 - 126 |
| Standard SAPs | 4 | 8 | 12 | 16 | 22 |
| Additional SAPs | 0 - 8 | 0 - 8 | 0 - 8 | 0 - 8 | 0 - 8 |

| Description | Max34 | Max71 | Max108 | Max145 | Max190 |
|-----------------------------------|----------------|----------------|----------------|----------------|----------------|
| Number of IFP | 1 | 1 | 1 | 1 | 1 |
| Standard spare PUs | 2 | 2 | 2 | 2 | 2 |
| Enabled Memory sizes GB | 512 - 7936 | 512 - 16128 | 512 - 24320 | 512 - 32512 | 512 - 40704 |
| Flexible memory sizes GB | N/A | 512 - 7936 | 512 - 16128 | 512 - 24320 | 512 - 32512 |
| L1 cache per PU (I/D) | 128/128 KB | 128/128 KB | 128/128 KB | 128/128 KB | 128/128 KB |
| L2 cache per PU (I/D) | 48/48 MB (I/D) | 48/48 MB (I/D) | 48/48 MB (I/D) | 48/48 MB (I/D) | 48/48 MB (I/D) |
| L3 shared cache per PU chip | 256 MB | 256 MB | 256 MB | 256 MB | 256 MB |
| L4 shared cache per drawer | 960 MB | 960 MB | 960 MB | 960 MB | 960 MB |
| Cycle time (ns) | 0.192 | 0.192 | 0.192 | 0.192 | 0.192 |
| Clock frequency | 5.2 GHz | 5.2 GHz | 5.2 GHz | 5.2 GHz | 5.2 GHz |
| Maximum number of PCIe fanouts | 12 | 24 | 36 | 48 | 60 |
| PCIe Bandwidth | 16 GBps | 16 GBps | 16 GBps | 16 GBps | 16 GBps |
| Number of support elements | 2 | 2 | 2 | 2 | 2 |
| External AC power | 3-phase | 3-phase | 3-phase | 3-phase | 3-phase |
| Internal Battery Feature BPA only | Optional | Optional | Optional | Optional | Optional |



Central processor complex design

This chapter describes the design of the IBM z15 processor unit. By understanding this design, users become familiar with the functions that make the z15 server a system that accommodates a broad mix of workloads for large enterprises.

Naming: The IBM z15 server generation is available as the following machine types and models:

- ▶ Machine Type 8561 (M/T 8561), Model T01, Features Max34, Max71, Max108, Max145, and Max190, which is further identified as *IBM z15 Model T01*, or *z15 T01*, unless otherwise specified.
- ▶ Machine Type 8562 (M/T 8562), Model T02, Features Max4, Max13, Max21, Max32, and Max65, which is further identified as *IBM z15 Model T02*, or *z15 T02*, unless otherwise specified.

In the remainder of this chapter, IBM z15 (z15) refers to both machine types.

For more information about the processor unit, see [z/Architecture Principles of Operation, SA22-7832](#).

This chapter includes the following topics:

- ▶ 3.1, “Overview” on page 90
- ▶ 3.2, “Design highlights” on page 90
- ▶ 3.3, “CPC drawer design” on page 92
- ▶ 3.4, “Processor unit design” on page 97
- ▶ 3.5, “Processor unit functions” on page 113
- ▶ 3.6, “Memory design” on page 128
- ▶ 3.7, “Logical partitioning” on page 132
- ▶ 3.8, “Intelligent Resource Director” on page 142
- ▶ 3.9, “Clustering technology” on page 144
- ▶ 3.10, “Virtual Flash Memory” on page 150
- ▶ 3.11, “Secure Service Container” on page 151

3.1 Overview

The z15 symmetric multiprocessor (SMP) system is the next step in an evolutionary trajectory that began with the introduction of the IBM System/360 in 1964. Over time, the design was adapted to the changing requirements that were dictated by the shift toward new types of applications on which clients depend.

z15 servers offer high levels of reliability, availability, serviceability (RAS), resilience, and security. It fits into the IBM strategy in which mainframes play a central role in creating an infrastructure for cloud, analytics, and mobile, underpinned by security. The z15 server is designed so that everything around it, such as operating systems, middleware, storage, security, and network technologies that support open standards, helps you achieve your business goals.

The modular CPC drawer design aims to reduce, or in some cases even eliminate, planned and unplanned outages. The design does so by offering concurrent repair, replace, and upgrade functions for processors, memory, and I/O. For more information about the z15 RAS features, see Chapter 9, “Reliability, availability, and serviceability” on page 383.

z15 T01 servers include the following features:

- ▶ Ultra-high frequency, large, high-speed buffers (caches) and memory
- ▶ Superscalar processor design
- ▶ Improved Out-of-order core execution
- ▶ Simultaneous multithreading (SMT)
- ▶ Enhanced Single-instruction multiple-data (SIMD)
- ▶ On-chip integrated accelerator for z Enterprise Data Compression (zEDC) (one per PU chip)
- ▶ Flexible configuration options

It is the next implementation of IBM Z servers to address the ever-changing IT environment.

For more information about frames and configurations, see Chapter 2, “Central processor complex hardware components” on page 35, and Appendix D, “Frame configurations” on page 515.

3.2 Design highlights

The physical packaging of z15 servers CPC drawer is a continuation and evolution of the z14 systems. Its modular CPC drawer and single chip module (SCM) design address the augmenting costs that are related to building systems with ever-increasing capacities. The modular CPC drawer design is flexible and expandable, offering unprecedented capacity to meet consolidation needs, and might contain even larger capacities in the future.

z15 servers CPC continues the line of mainframe processors that are compatible with an earlier version. Evolution® brings the following processor design enhancements:

- ▶ Twelve cores per PU chip design
- ▶ Each PU chip has 3 PCIe Generation 4 ports (x16@32GBps)
- ▶ Pipeline optimization
- ▶ Improved SMT and SIMD

- ▶ Improved branch prediction
- ▶ Improved co-processor functionality
- ▶ IBM Integrated Accelerator for zEnterprise Data Compression (zEDC) (on-chip compression accelerator)

It uses 24-, 31-, and 64-bit addressing modes, multiple arithmetic formats, and multiple address spaces for robust inter-process security.

The z15 system design has the following main objectives:

- ▶ Offer a data-centric approach to information (data) security that is simple, transparent, and consumable (extensive data encryption from inception to archive, in-flight and at-rest).
- ▶ Offer a *flexible infrastructure* to concurrently accommodate a wide range of operating systems and applications, from the traditional systems (for example, z/OS and z/VM) to the world of Linux, cloud, analytics, and mobile computing.
- ▶ Offer state-of-the-art *integration* capability for server consolidation by using virtualization capabilities in a highly *secure environment*:
 - Logical partitioning, which allows 85 independent logical servers.
 - z/VM, which can virtualize hundreds to thousands of servers as independently running virtual machines (guests).
 - Hipersockets, which implement virtual LANs between logical partitions (LPARs) within the system.
 - Efficient data transfer that uses direct memory access (SMC-D), Remote Direct Memory Access (SMC-R), and reduced storage access latency for transactional environments - zHyperLink Express.
 - The IBM Z Processor Resource/System Manager (PR/SM) is designed for Common Criteria Evaluation Assurance Level 5+ (EAL 5+) certification for security, so an application that is running on one partition (LPAR) cannot access another application on a different partition, which provides essentially the same security as an air-gapped system.
 - A new feature was introduced: Secure Execution, which not only securely separates second-level guest operating systems running under KVM for Z from each other but securely separates access to second-level guests from the hypervisor.

This configuration allows for a logical and virtual server coexistence and maximizes system utilization and efficiency by sharing hardware resources.

- ▶ Offer *high-performance computing* to achieve the outstanding response times that are required by new workload-type applications. This performance is achieved by high-frequency, enhanced superscalar processor technology, out-of-order core execution, large high-speed buffers (cache) and memory, an architecture with multiple complex instructions, and high-bandwidth channels.
- ▶ Offer the *high capacity* and *scalability* that are required by the most demanding applications, from the single-system and clustered-systems points of view.
- ▶ Offer the capability of *concurrent upgrades* for processors, memory, and I/O connectivity, which prevents system outages in planned situations.
- ▶ Implement a system with *high availability* and *reliability*. These goals are achieved with redundancy of critical elements and sparing components of a single system, and the clustering technology of the Parallel Sysplex environment.
- ▶ Have internal and external *connectivity* offerings, supporting open standards, such as Gigabit Ethernet (GbE) and Fibre Channel Protocol (FCP).

- ▶ Provide leading *cryptographic* performance. Every processor unit (PU) includes a dedicated and optimized CP Assist for Cryptographic Function (CPACF). Optional Crypto Express features with cryptographic coprocessors provide the highest standardized security certification.¹ These optional features can also be configured as Cryptographic Accelerators to enhance the performance of Secure Sockets Layer/Transport Layer Security (SSL/TLS) transactions.
- ▶ Provide on-chip compression. Every PU chip design incorporates a compression unit, which is the IBM Integrated Accelerator for z Enterprise Data Compression (zEDC) (which is different from the CMPSC implemented in each core).
- ▶ Be *self-managing* and *self-optimizing*, adjusting itself when the workload changes to achieve the best system throughput. This process can be done by using the Intelligent Resource Director or the Workload Manager functions, which are assisted by HiperDispatch.
- ▶ Have a *balanced system* design with pervasive encryption, which provides large data rate bandwidths for high-performance connectivity along with processor and system capacity, while protecting every byte that enters and exits the z15.

The remaining sections in this chapter describe the z15 system structure, showing a logical representation of the data flow from PUs, caches, memory cards, and various interconnect capabilities.

3.3 CPC drawer design

A z15 T01 system can have up to five CPC drawers in a full configuration, with up to 190 PUs that can be characterized for customer use, and up to 40 TB of customer usable memory. Each CPC drawer is logically divided in two clusters to improve the processor and memory affinity and availability.

The following types of CPC drawer configurations are available for z15 T01:

- ▶ One drawer: Max34
- ▶ Two drawers: Max71
- ▶ Three drawers: Max108
- ▶ Four drawers: Max145
- ▶ Five drawers: Max190

Note: Max145 and Max190 are factory build only. It is not possible to upgrade in the field to Max145 or Max190.

The z15 T01 has up to 20 memory controller units (MCUs) for a Max190 feature (four MCUs per CPC drawer). The MCU configuration uses five-channel redundant array of independent memory (RAIM) protection, with dual inline memory modules (DIMM) bus cyclic redundancy check (CRC) error retry.

The cache hierarchy (L1, L2, L3, and L4) is implemented with embedded dynamic random access memory (eDRAM). Until recently, eDRAM was considered to be too slow for this use. However, a breakthrough in technology that was made by IBM eliminated that limitation. In addition, eDRAM offers higher density, less power utilization, fewer soft errors, and better performance. Concurrent maintenance allows dynamic central processing complex (CPAC) drawer add and repair.²

¹ Federal Information Processing Standard (FIPS)140-2 Security Requirements for Cryptographic Modules

² For configurations with two or more CPC drawers installed

z15 servers use CMOS Silicon-on-Insulator (SOI) 14 nm chip technology with advanced low latency pipeline design, which creates high-speed yet power-efficient circuit designs. The PU SCM has a dense packaging, which allows closed water loop cooling. The heat from the closed loop is air-cooled by a radiator unit (RU) or optionally, by a water-cooling unit (WCU). The water-cooling option can lower the total power consumption of the system. This benefit is significant for larger configurations. For more information, see 2.9, “Power and cooling” on page 77.

3.3.1 Cache levels and memory structure

The z15 memory subsystem focuses on keeping data “closer” to the PU core. With the current processor configuration, all on chip cache levels increased.

The z15 T01 cache levels and memory hierarchy are shown in Figure 3-1.

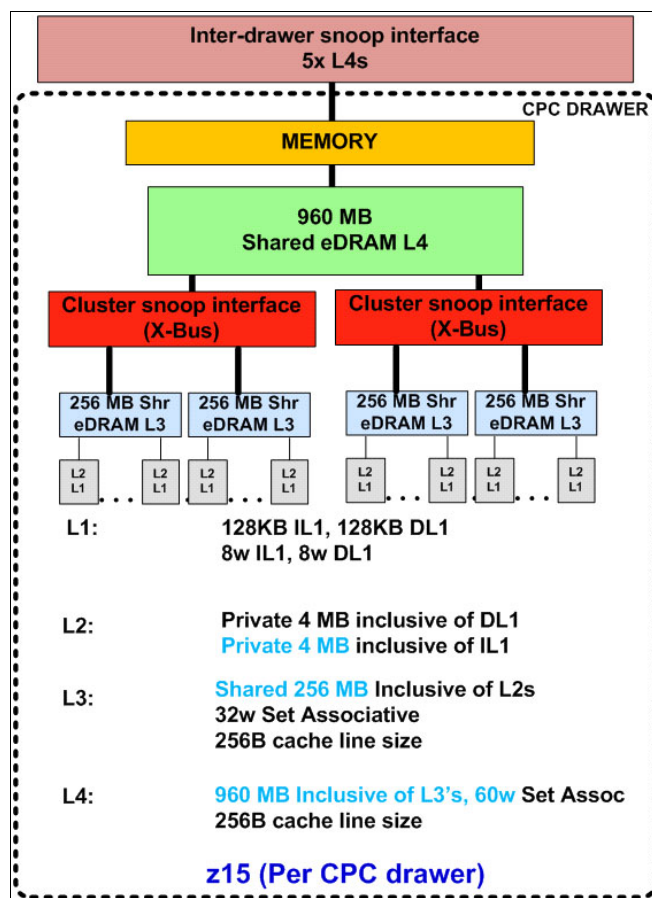


Figure 3-1 z15 cache levels and memory hierarchy

Although L1, L2, and L3 caches are implemented on the PU SCM, the fourth cache level (L4) is implemented within the system controller (SC) SCM. One L4 cache is present in each CPC drawer, which is shared by all PU SCMs. The cache structure of the z15 has the following characteristics:

- ▶ Larger L1, L2, and L3 caches (more data closer to the core).
- ▶ L1 and L2 caches use eDRAM, and are private for each PU core.
- ▶ L2-L3 interface has a new Fetch cancel protocol, a revised L2 Least Recent Used (LRU) Demote handling.

- ▶ L3 cache also uses eDRAM and is shared by all 12 cores within the PU chip. For availability and reliability, L3 cache implements symbol ECC.
- ▶ L4 cache also uses eDRAM, and is shared by all PU chips on the CPC drawer. Each L4 cache has 960 MB inclusive of L3's, 60w Set Associative, and 256-byte cache line size. A five-CPC drawer system has 4800 MB (5 x 960 MB) of shared L4 cache.

In most real-world situations, a fair number of cache lines exist in multiple L3s underneath a specific L4. The L4 does not contain the same line multiple times, but rather once with an indication of all the cores that have a copy of that line. As such, 960 MB of inclusive L4 can easily cover 1024 MB of underlying L3 caches (4 x 256 MB per CPC drawer).

- ▶ Main storage has up to 8 TB addressable memory per CPC drawer, which uses 20 DIMMs. A five-CPC drawer system can have up to 40 TB of main storage.

Considerations

Cache sizes are being limited by ever-diminishing cycle times because they must respond quickly without creating bottlenecks. Access to large caches costs more cycles. Instruction and data cache (L1) sizes must be limited because larger distances must be traveled to reach long cache lines. This L1 access time generally occurs in one cycle, which prevents increased latency.

Also, the distance to remote caches as seen from the microprocessor becomes a significant factor. An example is an L4 cache that is not on the microprocessor (and might not even be in the same CPC drawer). Although the L4 cache is rather large, several cycles are needed to travel the distance to the cache. The node-cache topology of z15 servers is shown in Figure 3-2.

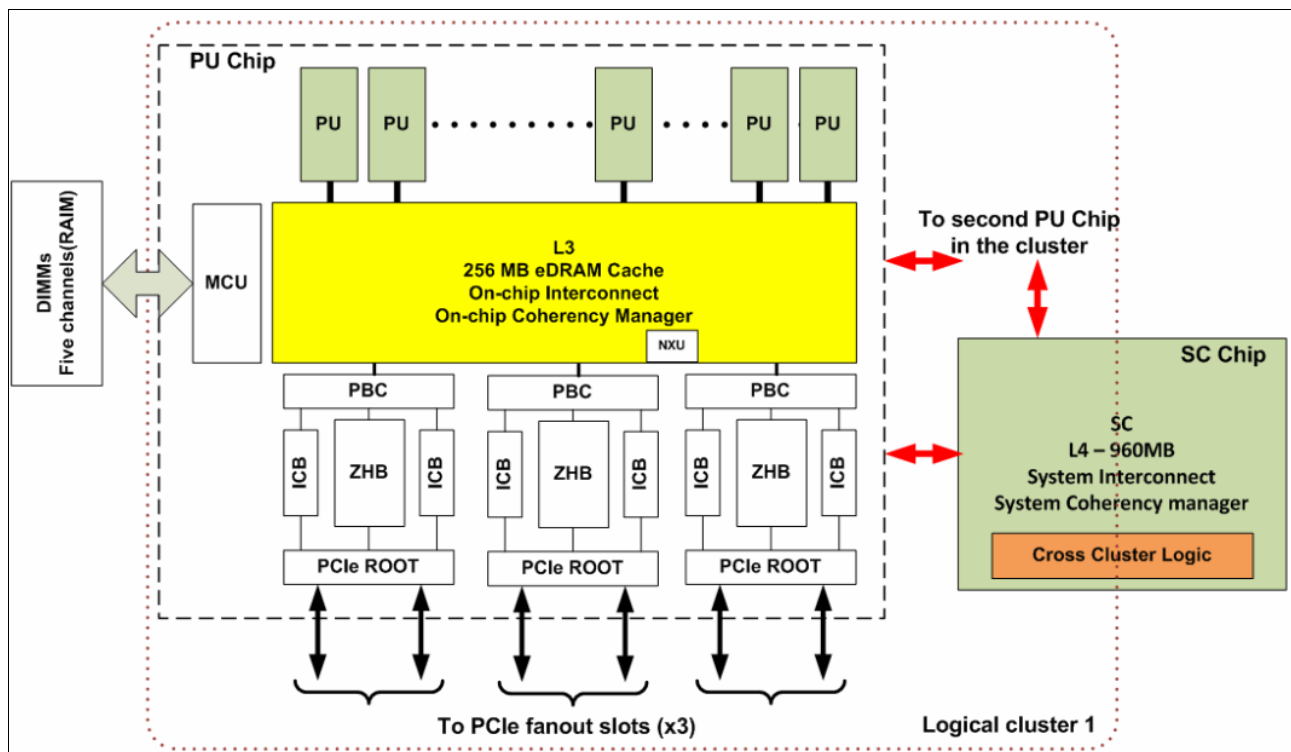


Figure 3-2 z15 cache topology

Although large caches mean increased access latency, the new technology of CMOS 14S0 (14 nm chip lithography) and the lower cycle time allows z15 servers to increase the size of cache levels (L4, L3, and L2) within the PU chip by using denser packaging. This design reduces traffic to and from the shared L4 cache, which is on another chip (SC chip). Only when a cache miss occurs in L1, L2, or L3 is a request sent to L4. L4 is the coherence manager, which means that all memory fetches must be in the L4 cache before that data can be used by the processor. However, in the z15 cache design, some lines of the L3 cache are not included in the L4 cache.

Another approach is available for avoiding L4 cache access delays (latency). The L4 cache straddles up to five CPC drawers. This configuration means that relatively long distances exist between the higher-level caches in the processors and the L4 cache content.

To overcome the delays that are inherent in the SMP CPC drawer design and save cycles to access the remote L4 content, the system keeps instructions and data as close to the processors as possible. This configuration can be managed by directing as much work of a particular LPAR workload to the processors in the same CPC drawer as the L4 cache. This configuration is achieved by having the IBM Processor Resource/Systems Manager (PR/SM) scheduler and the z/OS WLM and dispatcher work together. Have them keep as much work as possible within the boundaries of as few processors and L4 cache space (which is best within a CPC drawer boundary) without affecting throughput and response times.

The cache structures of z15 T01 systems are compared with the previous generation of IBM Z (z14 M0x) in Figure 3-3.

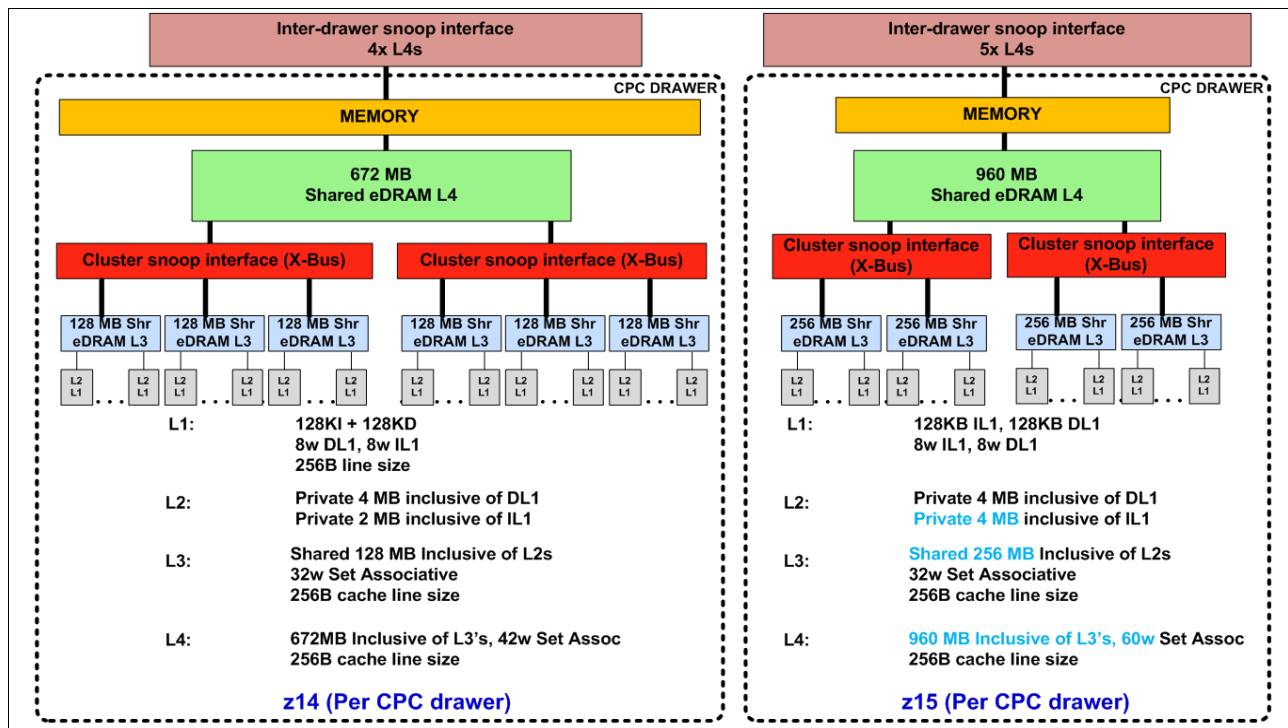


Figure 3-3 z15 and z14 cache level comparison

Compared to z14, the z15 cache design has larger L4, L3, and L2 cache sizes. As in z14 servers, in z15 servers, more affinity exists between the memory of a partition, the L4 cache in the SC, accessed by the two logical clusters in the same CPC drawer, and the cores in the PU. The access time of the private cache usually occurs in one cycle.

As in z14, the z15 cache level structure is focused on keeping more data closer to the PU. This design can improve system performance on many production workloads.

HiperDispatch

To help avoid latency in a high-frequency processor design, PR/SM and the dispatcher must be prevented from scheduling and dispatching a workload on *any* processor available, which keeps the workload in as small a portion of the system as possible. The cooperation between z/OS and PR/SM is bundled in a function called *HiperDispatch*. HiperDispatch uses the z15 cache topology, which features reduced cross-cluster “help” and better locality for multi-task address spaces.

PR/SM can use dynamic PU reassignment to move processors (CPs, ZIIPs, IFLs, ICFs, SAPs, and spares) to a different chip, node, and drawer to improve the reuse of shared caches by processors of the same partition. It can use dynamic memory relocation (DMR) to move a running partition’s memory to different physical memory to improve the affinity and reduce the distance between the memory of a partition and the processors of the partition. For more information about HiperDispatch, see 3.7, “Logical partitioning” on page 132.

3.3.2 CPC drawer interconnect topology

CPC drawers are interconnected in a point-to-point topology at SC level, which allows a CPC drawer to communicate with every CPC drawer. Data transfer does not always have to go through another CPC drawer (L4 cache) to address the requested data or control information.

The z15 T01 intra-CPC drawer communication structure is shown in Figure 3-4.

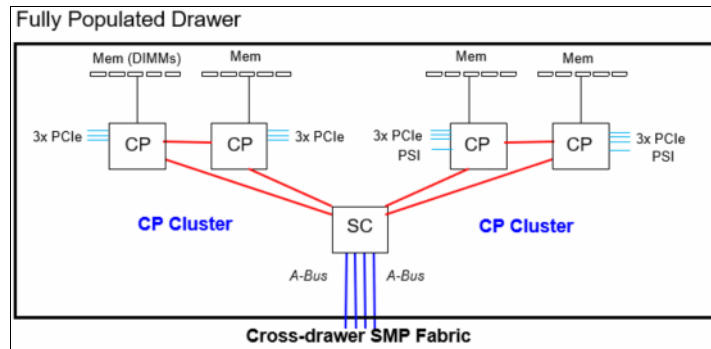


Figure 3-4 z15 CPC drawer communication topology.

A simplified topology of a five-CPC drawer system is shown in Figure 3-5.

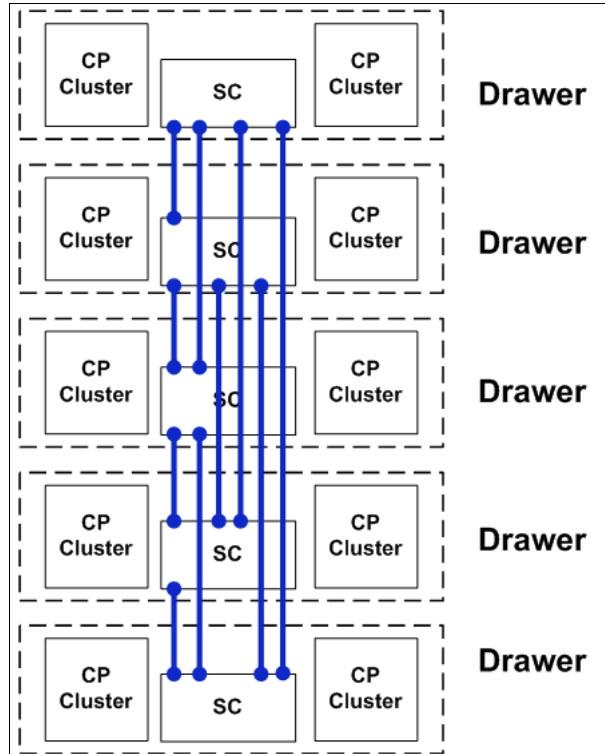


Figure 3-5 Point-to-point topology with five CPC drawers

Inter-CPC drawer communication occurs at the L4 cache level, which is implemented on the SC chips in each drawer. The SC function regulates coherent drawer-to-drawer traffic.

3.4 Processor unit design

Processor cycle time is especially important for processor-intensive applications. Current systems design is driven by processor cycle time, although improved cycle time does not automatically mean that the performance characteristics of the system improve.

z15 T01 core frequency is 5.2 GHz (same as z14 M0x), but with increased number of processors that share larger caches to have shorter access times and improved capacity and performance.

Through innovative processor design (pipeline and cache management redesigns), the IBM Z processor performance continues to evolve. Enhancements were made on the processor unit design, such as on out-of-order execution, branch prediction mechanism, Floating point and vector unit, Divide engine scheduler, Load/Store Unit and Operand Store Compare (OSC), simultaneous multi-threading, and the relative nest intensity (RNI) redesigns. For more information about RNI, see 12.4, “Relative Nest Intensity” on page 481.

The processing performance was enhanced by bringing the following changes on the z15 processor design:

- ▶ Core optimization to enable performance and capacity growth.
- ▶ New EDRAM macro design with 2x macro density (cache growths and L2-L3 Protocol changes to reduce latency).

- ▶ On-chip IBM Integrated Accelerator for zEnterprise Data Compression (Nest compression accelerator, or NXU. For more information, see Figure 2-14 on page 50.)
- ▶ Enhancement of nest-core staging.

Because of those enhancements, the z15 processor full speed z/OS single-thread performance is on average 1.12 - 1.14 times faster than the z14 at equal N-way. For more information about performance, see Chapter 12, “Performance” on page 475.

z13 servers introduced architectural extensions with instructions that reduce processor quiesce effects, cache misses, and pipeline disruption, and increase parallelism with instructions that process several operands in a single instruction (SIMD). The processor architecture was further developed for z14 and z15. z15 includes the following enhancements:

- ▶ Optimized second-generation SMT
- ▶ Enhanced SIMD instructions set
- ▶ Improved Out-of-Order core execution
- ▶ Improvements in branch prediction and handling
- ▶ Pipeline optimization
- ▶ Secure Execution³
- ▶ Co-processor compression enhancement

The z15 enhanced Instruction Set Architecture (ISA) includes a set of instructions that are added to improve compiled code efficiency. These instructions optimize PUs to meet the demands of various business and analytics workload types without compromising the performance characteristics of traditional workloads.

3.4.1 Simultaneous multithreading

Aligned with industry directions, z15 servers can process up to two simultaneous threads in a single core while sharing certain resources of the processor, such as execution units, translation lookaside buffers (TLBs), and caches. When one thread in the core is waiting for other hardware resources, the second thread in the core can use the shared resources rather than remaining idle. This capability is known as *simultaneous multithreading* (SMT).

An operating system with SMT support can be configured to dispatch work to a thread on a zIIP (for eligible workloads in z/OS) or an IFL (for z/VM and Linux on Z) core in single thread or SMT mode so that HiperDispatch cache optimization can be considered. For more information about operating system support, see Chapter 7, “Operating system support” on page 253.

SMT technology allows instructions from more than one thread to run in any pipeline stage at a time. SMT can handle up to four pending translations.

Each thread has its own unique state information, such as Program Status Word - S/360 Architecture (PSW) and registers. The simultaneous threads cannot necessarily run instructions instantly and must at times compete to use certain core resources that are shared between the threads. In some cases, threads can use shared resources that are not experiencing competition.

³ Secure execution requires OS support.

Two threads (A and B) running on the same processor core on different pipeline stages and sharing the core resources is shown in Figure 3-6.

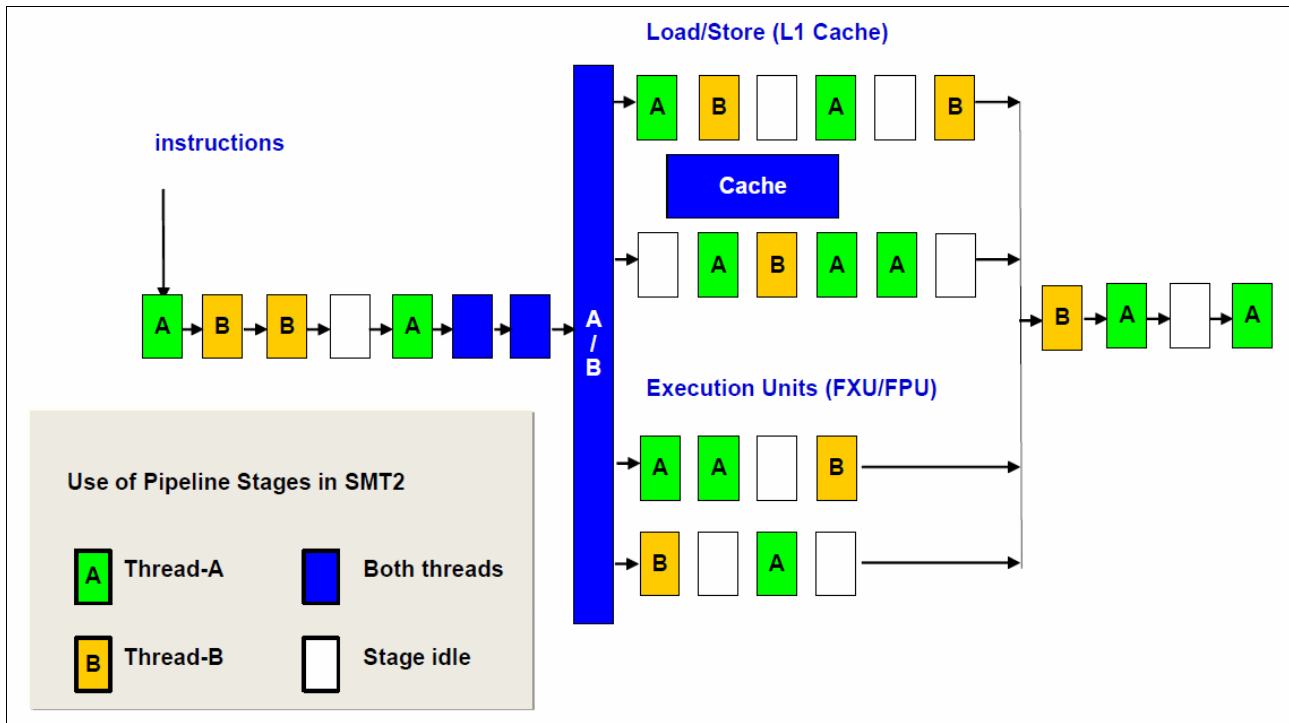


Figure 3-6 Two threads running simultaneously on the same processor core

The use of SMT provides more efficient use of the processors' resources and helps address memory latency, which results in overall throughput gains. The active thread shares core resources in space, such as data and instruction caches, TLBs, branch history tables, and, in time, pipeline slots, execution units, and address translators.

Although SMT increases the processing capacity, the performance in some cases might be superior if a single thread is used. Enhanced hardware monitoring supports measurement through CPUMF for thread usage and capacity.

For workloads that need maximum thread speed, the partition's SMT mode can be turned off. For workloads that need more throughput to decrease the dispatch queue size, the partition's SMT mode can be turned on.

SMT use is functionally transparent to middleware and applications, and no changes are required to run them in an SMT-enabled partition.

3.4.2 Single-instruction multiple-data

The z15 superscalar processor has 32 vector registers and an instruction set architecture that includes a subset of instructions (known as SIMD) that were added to improve the efficiency of complex mathematical models and vector processing. These new instructions allow a larger number of operands to be processed with a single instruction. The SIMD instructions use the superscalar core to process operands in parallel.

SIMD provides the next phase of enhancements of IBM Z analytics capability. The set of SIMD instructions are a type of data parallel computing and vector processing that can decrease the amount of code and accelerate code that handles integer, string, character, and floating point data types. The SIMD instructions improve performance of complex mathematical models and allow integration of business transactions and analytic workloads on IBM Z servers.

The 32 vector registers feature 128 bits. The instructions include string operations, vector integer, and vector floating point operations. Each register contains multiple data elements of a fixed size. The following instructions code specifies which data format to use and the size of the elements:

- ▶ Byte (16 8-bit operands)
- ▶ Halfword (eight 16-bit operands)
- ▶ Word (four 32-bit operands)
- ▶ Doubleword (two 64-bit operands)
- ▶ Quadword (one 128-bit operand)

In addition to the instructions that were introduced in z14, the SIMD provides the following enhancements for z15:

- ▶ Double-bandwidth vector loads
- ▶ Multiply/Divide speed ups
- ▶ Conversion speed ups
- ▶ New and enhanced vector instructions
- ▶ Load/store reversed (to help with little endian conversion)
- ▶ More vector shift operations
- ▶ VECTOR STRING SEARCH, for fast string search, supporting different encodings
- ▶ New vector FP converts
- ▶ New and enhanced Vector Packed Decimal instructions

The collection of elements in a register is called a *vector*. A single instruction operates on all of the elements in the register. Instructions include a non-destructive operand encoding that allows the addition of the register vector A and register vector B and stores the result in the register vector A ($A = A + B$).

A schematic representation of a SIMD instruction with 16-byte size elements in each vector operand is shown in Figure 3-7.

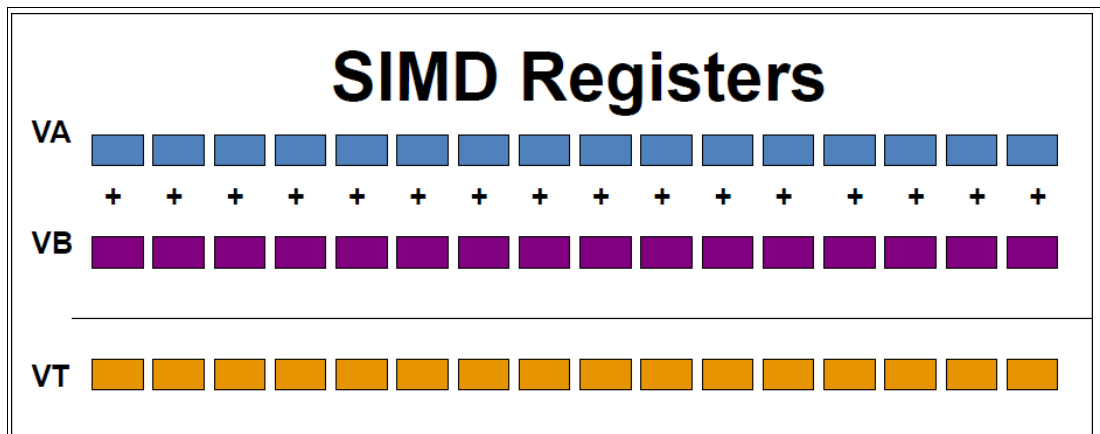


Figure 3-7 SIMD operation logic

The vector register file overlays the floating-point registers (FPRs), as shown in Figure 3-8. The FPRs use the first 64 bits of the first 16 vector registers, which saves hardware area and power, and makes it easier to mix scalar and SIMD codes. Effectively, the core gets 64 FPRs, which can further improve FP code efficiency.

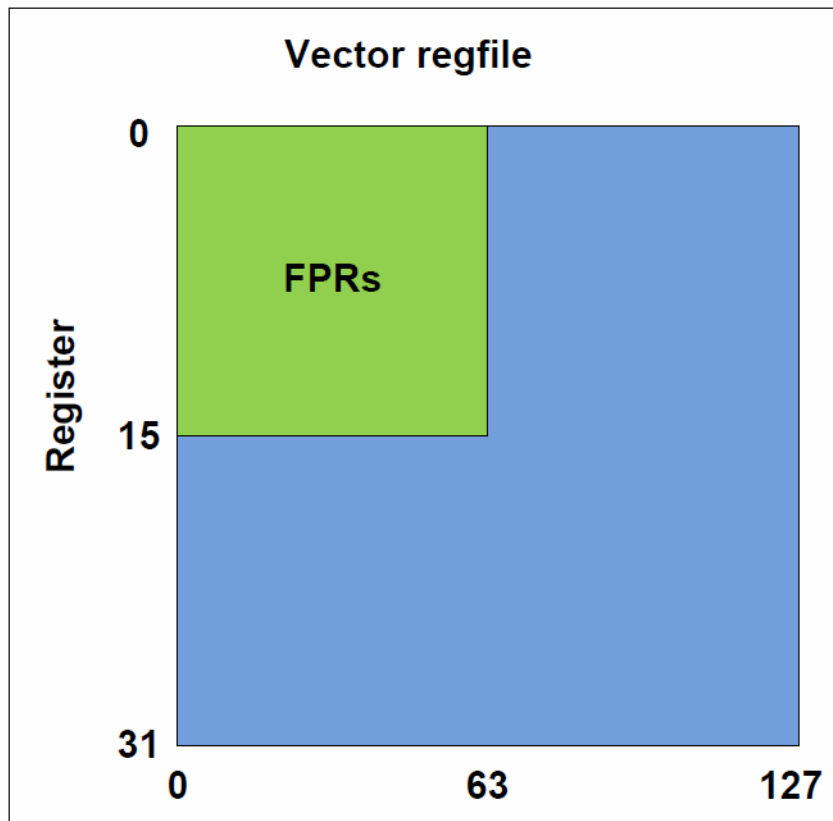


Figure 3-8 Floating point registers overlaid by vector registers

SIMD instructions include the following examples:

- ▶ Integer byte to quadword add, sub, and compare
- ▶ Integer byte to doubleword min, max, and average
- ▶ Integer byte to word multiply
- ▶ String find 8-bit, 16-bit, and 32-bit
- ▶ String range compare
- ▶ String find any equal
- ▶ String load to block boundaries and load/store with length

For most operations, the condition code is not set. A summary condition code is used only for a few instructions.

z15 SIMD features the following enhancements:

- ▶ Doubled vector double precision Binary Floating Point (BFP) operations throughput (2x 64b)
- ▶ Added vector single precision BFP (4x 32b)
- ▶ Added vector quad precision BFP (128b)
- ▶ Added binary Fixed Multiply Add (FMA) operations to speed up code
- ▶ Vector Single Precision/ Double Precision/ Quad Precision (SP/DP/QP) compare/min/max with programming language support

- ▶ Enhanced to Storage-to-Storage Binary Coded Decimal (BCD)
- ▶ Vector load/store right-most with length

3.4.3 Out-of-Order execution

z15 servers have an Out-of-Order core, much like the z14 and z13. This optimized Out-of-Order feature yields significant performance benefits for compute-intensive applications. It does so by reordering instruction execution, which allows later (younger) instructions to be run ahead of a stalled instruction, and reordering storage accesses and parallel storage accesses. Out-of-Order maintains good performance growth for traditional applications. Out-of-Order execution can improve performance in the following ways:

- ▶ Reordering instruction execution

Instructions stall in a pipeline because they are waiting for results from a previous instruction or the execution resource that they require is busy. In an in-order core, this stalled instruction stalls all later instructions in the code stream. In an out-of-order core, later instructions are allowed to run ahead of the stalled instruction.

- ▶ Reordering storage accesses

Instructions that access storage can stall because they are waiting on results that are needed to compute the storage address. In an in-order core, later instructions are stalled. In an out-of-order core, later storage-accessing instructions that can compute their storage address are allowed to run.

- ▶ Hiding storage access latency

Many instructions access data from storage. Storage accesses can miss the L1 and require 7 - 50 more clock cycles to retrieve the storage data. In an in-order core, later instructions in the code stream are stalled. In an out-of-order core, later instructions that are not dependent on this storage data are allowed to run.

The z15 processor includes pipeline enhancements that benefit Out-of-Order execution. The IBM Z processor design features advanced micro-architectural innovations that provide the following benefits:

- ▶ Maximized instruction-level parallelism (ILP) for a better cycles per instruction (CPI) design.
- ▶ Maximized performance per watt. Two cores are added (as compared to the z14 chip) at slightly higher chip power.
- ▶ Enhanced instruction dispatch and grouping efficiency.
- ▶ Increased OoO resources (Global Completion Table entries, physical GPR entries, and physical FPR entries).
- ▶ Improved completion rate.
- ▶ Reduced cache/TLB miss penalty.
- ▶ Improved execution of D-Cache store and reload and new Fixed-point divide.
- ▶ New Operand Store Compare (OSC) (load-hit-store conflict) avoidance scheme.
- ▶ Enhanced branch prediction structure and sequential instruction fetching.

Program results

The Out-of-Order execution does not change any program results. Execution can occur out of (program) order, but all program dependencies are accepted, ending up with the same results as in-order (program) execution. It design was optimized by increasing the Global Completion Table (GCT) from 48x3 to 60x3, increasing Issue queue size from 2x30 to 2x36 and designing a new Mapper.

This implementation requires special circuitry to make execution and memory accesses display in order to the software. The logical diagram of a z15 core is shown in Figure 3-9.

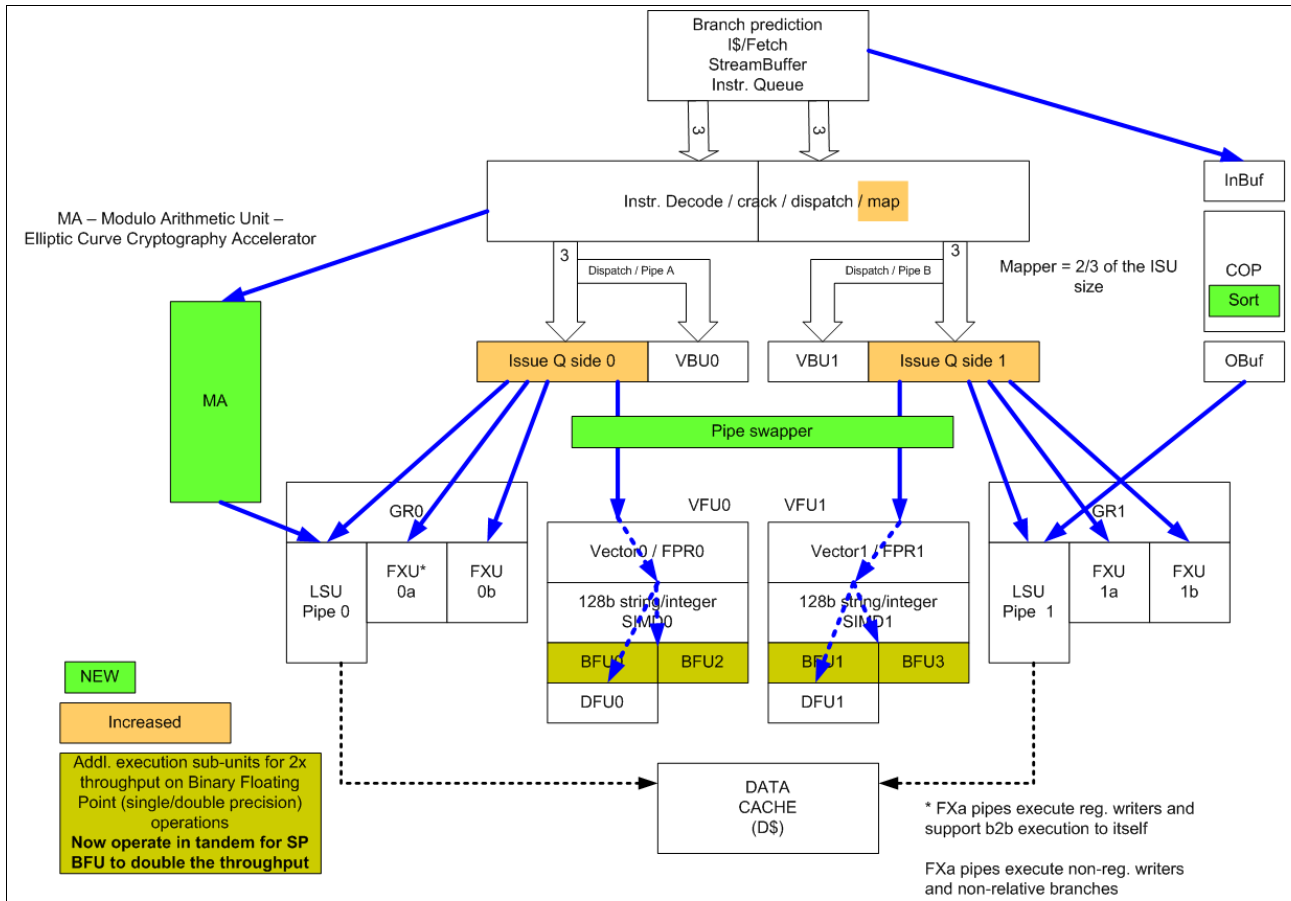


Figure 3-9 z15 PU core logical diagram

Memory address generation and memory accesses can occur out of (program) order. This capability can provide a greater use of the z15 superscalar core, and improve system performance.

The z15 T01 processor unit core is a superscalar, out-of-order, SMT processor with 12 execution units. Up to six instructions can be decoded per cycle, and up to 12 instructions or operations can be started to run per clock cycle (0.192 ns). The execution of the instructions can occur out of program order, and memory address generation and memory accesses can also occur out of program order. Each core has special circuitry to display execution and memory accesses in order to the software.

The z15 superscalar PU core can have up to 12 instructions or operations that are running per cycle. This technology results in shorter workload runtime.

Branch prediction

If the branch prediction logic of the microprocessor makes the wrong prediction, all instructions in the parallel pipelines are removed. The wrong branch prediction is expensive in a high-frequency processor design. Therefore, the branch prediction techniques that are used are important to prevent as many wrong branches as possible.

For this reason, various history-based branch prediction mechanisms are used, as shown on the in-order part of the z15 PU core logical diagram in Figure 3-9 on page 103. The branch target buffer (BTB) runs ahead of instruction cache pre-fetches to prevent branch misses in an early stage. Furthermore, a branch history table (BHT), in combination with a pattern history table (PHT) and the use of tagged multi-target prediction technology branch prediction, offers a high branch prediction success rate.

The z15 microprocessor improves the branch prediction throughput by using a simplified and larger BTB.

3.4.4 Superscalar processor

A *scalar processor* is a processor that is based on a single-issue architecture, which means that only a single instruction is run at a time. A *superscalar processor* allows concurrent (parallel) execution of instructions by adding more resources to the microprocessor in multiple pipelines, each working on its own set of instructions to create parallelism.

A superscalar processor is based on a multi-issue architecture. However, when multiple instructions can be run during each cycle, the level of complexity is increased because an operation in one pipeline stage might depend on data in another pipeline stage. Therefore, a superscalar design demands careful consideration of which instruction sequences can successfully operate in a long pipeline environment.

Many challenges exist in creating an efficient superscalar processor. The superscalar design of the PU made significant strides in avoiding address generation interlock (AGI) situations. Instructions that require information from memory locations can suffer multi-cycle delays to get the needed memory content. Because high-frequency processors wait “faster” (spend processor cycles more quickly while idle), the cost of getting the information might become prohibitive.

3.4.5 Compression and cryptography accelerators on a chip

This section introduces the CPACF enhancements for z15 and the new on-chip IBM Integrated Accelerator for zEnterprise Data Compression (zEDC).

IBM integrated Accelerator for zEDC (on-chip)

z15 On-Chip Compression (Nest Accelerator Unit - NXU, see Figure 3-10 on page 106) provides value for existing and new compression users.

z15 Compression/Decompression is implemented in the Nest Accelerator Unit (NXU) on each processor chip. z15 On-Chip Compression delivers industry-leading throughput and replaces the zEDC Express PCIe adapter available on the IBM z14 and earlier servers.

One Nest Accelerator Unit (NXU) is used per processor chip, which is shared by all cores on the chip and features the following benefits:

- ▶ Brand new concept of sharing and operating an accelerator function in the nest
- ▶ Supports DEFLATE compliant compression/decompression and GZIP CRC/ZLIB Adler
- ▶ Low latency

- ▶ High bandwidth
- ▶ Problem state execution
- ▶ Hardware/Firmware interlocks to ensure system responsiveness
- ▶ Architected instruction
- ▶ Executed in millicode

Based on IBM benchmarks, the largest IBM z15 T01 with the Integrated Accelerator for zEDC provides up to 17 times the total compression throughput (compress up to 260 GBps with the Integrated Accelerator for zEDC) of a z14 (3906) configured with the maximum number of zEDC Express cards. Also, the Integrated Accelerator for zEDC on z15 improves the compression ratio by 5% over z14 zEDC Express.⁴

Sharing of zEDC cards is limited to 15 LPARs per adaptor. The On-Chip Compression Accelerator removes this virtualization constraint because it is shared by all PUs on the processors chip and is therefore available to all LPARs and guests.

Moving the compression function from the IO drawer to the processor chip means that compression can operate directly on L3 cache and data does not need to be passed by using I/O.

Compression is run in one of two execution modes: Synchronous or Asynchronous.

Synchronous execution occurs in problem state where the user application starts the instruction in its virtual address space.

Asynchronous execution is optimized for Large Operations under z/OS for authorized applications (for example, BSAM) and issues I/O using EADMF for asynchronous execution.

Asynchronous execution maintains the current user experience and provides a transparent implementation for existing authorized users of zEDC.

The On-Chip Compression implements compression as defined by RFC1951 (DEFLATE).

⁴ Measurements were collected in a controlled environment running an IBM developed workload under z/OS. Individual results can vary. Results are workload dependent.

Figure 3-10 shows the nest compression accelerator (NXU) for On-Chip Compression acceleration.

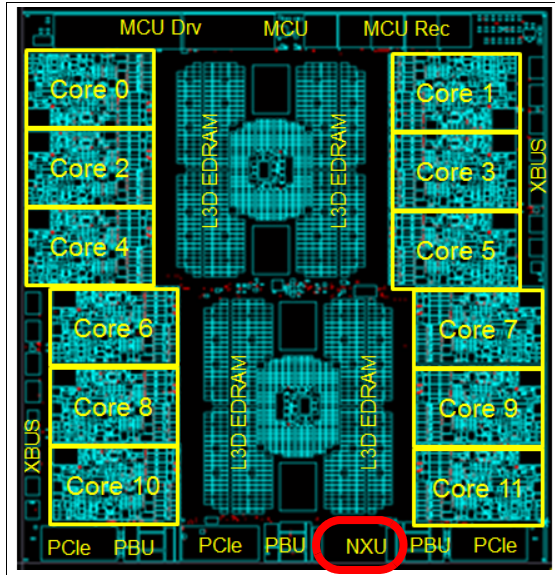


Figure 3-10 z15 PU Chip

For information about sizing, migration considerations, and software support, see Appendix C, “IBM Integrated Accelerator for zEnterprise Data Compression” on page 509.

Coprocessor units

One coprocessor unit is available for compression and cryptography on *each core* in the chip. The compression engine uses static dictionary compression and expansion. The compression dictionary uses the L1-cache (instruction cache).

The cryptography engine is used for the CPACF, which offers a set of symmetric cryptographic functions for encrypting and decrypting of clear key operations.

The coprocessors feature the following characteristics:

- ▶ Each core has an independent compression and cryptographic engine.
- ▶ The coprocessor was redesigned to support SMT operation and for throughput increase.
- ▶ It is available to any processor type (regardless of the processor characterization).
- ▶ The owning processor is busy when its coprocessor is busy.

The location of the coprocessor on the chip is shown in Figure 3-11.

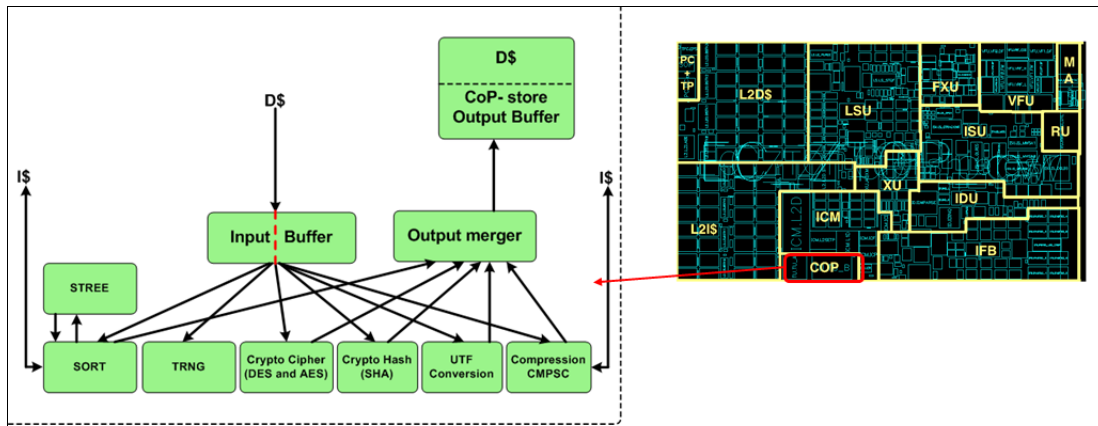


Figure 3-11 Core co-processor

Compression enhancements

The compression features the following enhancements:

- ▶ Huffman compression on top of CMPSC compression (embedded in dictionary, reuse of generators)
- ▶ Order Preserving compression in B-Trees and other index structures
- ▶ Faster expansion algorithms
- ▶ Reduced overhead on short data

CPACF

CPACF accelerates the encrypting and decrypting of SSL/TLS transactions, virtual private network (VPN)-encrypted data transfers, and data-storing applications that do not require FIPS 140-2 level 4 security. The assist function uses a special instruction set for symmetrical clear key cryptographic encryption and decryption, and for hash operations. This group of instructions is known as the *Message-Security Assist (MSA)*. For more information about these instructions, see *z/Architecture Principles of Operation, SA22-7832*.

Crypto functions enhancements

The crypto functions enhancements include Modulo Arithmetic unit in support of Elliptic Curve Cryptography. Expected use cases include the following examples:

- ▶ SSL libraries (authentication on the web)
- ▶ Blockchain

For more information about cryptographic functions on z15 servers, see Chapter 6, “Cryptographic features” on page 215.

IBM Integrated Accelerator for Z SORT

Sorting data is a significant part of IBM Z workloads including batch workloads, database query processing, and utility processing. The amount of data stored and processed on Z continues to grow at a high rate driving an ever-increasing sort workload.

IBM z15 processor adds special hardware to significantly accelerate frequently used functions. The IBM Integrated Accelerator for Z sort has been implemented on each core and it provides new architected instructions for sorting data to speed up sorting operations. It supports multiple active sorts in parallel and it is designed to:

- ▶ Optimize elapsed time and CPU time
- ▶ Shorten the batch window (primarily targeting existing batch type sort workloads)
- ▶ Improve select database functions, such as reorganization to help reorganize data more frequently as sorted data is faster to access in interactive use

3.4.6 Decimal floating point accelerator

The decimal floating point (DFP) accelerator function is present on each of the microprocessors (cores) on the 12-core chip. Its implementation meets business application requirements for better performance, precision, and function.

Base 10 arithmetic is used for most business and financial computation. Floating point computation that is used for work that is typically done in decimal arithmetic involves frequent data conversions and approximation to represent decimal numbers. This process makes floating point arithmetic complex and error-prone for programmers who use it for applications in which the data is typically decimal.

Hardware DFP computational instructions provide the following features:

- ▶ Data formats of 4, 8, and 16 bytes
- ▶ An encoded decimal (base 10) representation for data
- ▶ Instructions for running decimal floating point computations
- ▶ An instruction that runs data conversions to and from the decimal floating point representation

Benefits of the DFP accelerator

The DFP accelerator offers the following benefits:

- ▶ Avoids rounding issues, such as those issues that occur with binary-to-decimal conversions.
- ▶ Controls existing binary-coded decimal (BCD) operations better.
- ▶ Follows the standardization of the dominant decimal data and decimal operations in commercial computing, supporting the industry standardization (IEEE 745R) of decimal floating point operations. Instructions are added in support of the Draft Standard for Floating-Point Arithmetic - IEEE 754-2008, which is intended to supersede the ANSI/IEEE Standard 754-1985.
- ▶ Allows COBOL programs that use zoned-decimal operations to take advantage of the z/Architecture DFP instructions.

z15 servers have two DFP accelerator units per core, which improve the decimal floating point execution bandwidth. The floating point instructions operate on newly designed vector registers (32 new 128-bit registers).

z15 servers include new decimal floating point packed conversion facility support with the following benefits:

- ▶ Reduces code path length because extra instructions to format conversion are no longer needed.
- ▶ Packed data is operated in memory by all decimal instructions without general-purpose registers, which were required only to prepare for decimal floating point packed conversion instruction.
- ▶ Converting from packed can now force the input packed value to positive instead of requiring a separate OI, OILL, or load positive instruction.

- ▶ Converting to packed can now force a positive zero result instead of requiring ZAP instruction.

Software support

DFP is supported in the following programming languages and products:

- ▶ Release 4 and later of the High Level Assembler
- ▶ C/C++, which requires supported z/OS version
- ▶ Enterprise PL/I Release 3.7 and Debug Tool Release 8.1 or later
- ▶ Java Applications that use the BigDecimal Class Library
- ▶ SQL support as of Db2 Version 9 and later

3.4.7 IEEE floating point

Binary and hexadecimal floating-point instructions are implemented in z15 servers. They incorporate IEEE standards into the system.

The z15 core implements two other execution subunits for 2x throughput on BFP (single/double precision) operations (see Figure 3-9 on page 103).

The key point is that Java and C/C++ applications tend to use IEEE BFP operations more frequently than earlier applications. Therefore, the better the hardware implementation of this set of instructions, the better the performance of applications.

3.4.8 Processor error detection and recovery

The PU uses a process called *transient recovery* as an error recovery mechanism. When an error is detected, the instruction unit tries the instruction again and attempts to recover the error. If the second attempt is unsuccessful (that is, a permanent fault exists), a relocation process is started that restores the full capacity by moving work to another PU.

Relocation under hardware control is possible because the R-unit has the full designed state in its buffer. PU error detection and recovery are shown in Figure 3-12 on page 109.

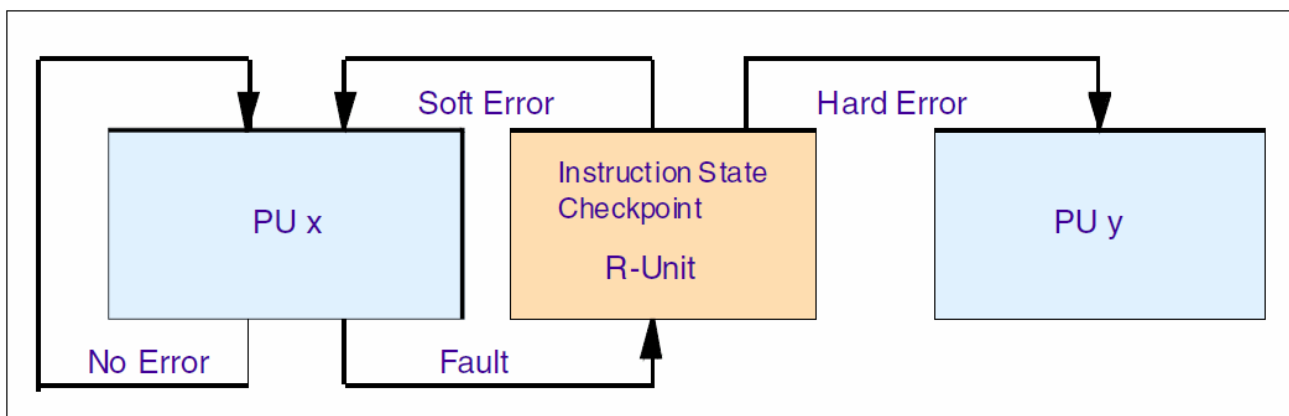


Figure 3-12 PU error detection and recovery

3.4.9 Branch prediction

Because of the ultra-high frequency of the PUs, the penalty for a wrongly predicted branch is high. Therefore, a multi-pronged strategy for branch prediction, based on gathered branch

history that is combined with other prediction mechanisms, is implemented on each microprocessor.

The BHT implementation on processors provides a large performance improvement. Originally introduced on the IBM ES/9000 9021 in 1990, the BHT is continuously improved.

The BHT offers significant branch performance benefits. The BHT allows each PU to take instruction branches that are based on a stored BHT, which improves processing times for calculation routines. In addition to the BHT, z15 servers use the following techniques to improve the prediction of the correct branch to be run:

- ▶ BTB
- ▶ PHT
- ▶ BTB data compression

The success rate of branch prediction contributes significantly to the superscalar aspects of z15 servers. This success is because the architecture rules prescribe that, for successful parallel execution of an instruction stream, the correctly predicted result of the branch is essential.

The z15 branch prediction includes the following enhancements over z14:

- ▶ Branch prediction search pipeline extended from five to six cycles to accommodate new predictors for increased accuracy/performance.
- ▶ Predictors enhancements:
 - SKOOT enhanced BTB search (SKip Over OffsetT entries remember where the branch search should continue). This enhancement saves BTB1 access cycles while searching for the next branch.
 - SSCRS (hardware-based super simple call-return stack). The major change is that it does not necessarily have to return to the next sequential instruction (NSIA). This feature supports return branches up to 8 bytes past the NSIA.
 - New TAGE (TAGged GEometric) history length branch predictor.
- ▶ Branch prediction is simplified by removing the BTBp “BTB1 victim cache”, which is replaced by a write buffer. The two independent read ports were replaced by one double-bandwidth port, which simplifies physical design significantly:
 - Level 1 Branch Target Buffer (BTB1): 2K rows x 4sets → 2 K rows x 8sets
 - Level 2 Branch Target Buffer (BTB2) size remains unchanged
- ▶ Better power efficiency: Several structures were redesigned to maintain their accuracy while less power is used through smart access algorithms.
- ▶ New static IBM IA regions expanded from four to eight. To conserve space, prediction structures do not store full target addresses. Instead, they use the locality and limited ranges of “4gig regions” of virtual instruction addresses - IA(0:31).

3.4.10 Wild branch

When a bad pointer is used or when code overlays a data area that contains a pointer to code, a random branch is the result. This process causes a 0C1 or 0C4 abend. Random branches are difficult to diagnose because clues about how the system got there are not evident.

With the wild branch hardware facility, the last address from which a successful branch instruction was run is kept. z/OS uses this information with debugging aids, such as the **SLIP** command, to determine from where a wild branch came. It can also collect data from that

storage location. This approach decreases the number of debugging steps that are necessary when you want to know from where the branch came.

3.4.11 Translation lookaside buffer

The TLB in the instruction and data L1 caches use a secondary TLB to enhance performance.

The size of the TLB is kept as small as possible because of its short access time requirements and hardware space limitations. Because memory sizes recently increased significantly as a result of the introduction of 64-bit addressing, a smaller working set is represented by the TLB.

To increase the working set representation in the TLB without enlarging the TLB, large (1 MB) page and giant page (2 GB) support is available and can be used when appropriate. For more information, see “Large page support” on page 130.

With the enhanced DAT-2 (EDAT-2) improvements, the IBM Z servers support 2 GB page frames.

z15 TLB enhancements

IBM z15 switches to a logical-tagged L1 directory and inline TLB2. Each L1 cache directory entry contains the virtual address and Address Space Control Element (ASCE) because it no longer must access TLB for L1 cache hit. TLB2 is accessed in parallel to L2, which saves significant latency compared to TLB1-miss.

The new translation engine allows up to four translations pending concurrently. Each translation step is ~2x faster, which helps second level guests.

3.4.12 Instruction fetching, decoding, and grouping

The superscalar design of the microprocessor allows for the decoding of up to six instructions per cycle and the execution of up to 12 instructions per cycle. Both execution and storage accesses for instruction and operand fetching can occur out of sequence.

Instruction fetching

Instruction fetching normally tries to get as far ahead of instruction decoding and execution as possible because of the relatively large instruction buffers available. In the microprocessor, smaller instruction buffers are used. The operation code is fetched from the I-cache and put in instruction buffers that hold prefetched data that is awaiting decoding.

Instruction decoding

The processor can decode up to six instructions per cycle. The result of the decoding process is queued and later used to form a group.

Instruction grouping

From the instruction queue, up to 12 instructions can be completed on every cycle. A complete description of the rules is beyond the scope of this publication.

The compilers and JVMs are responsible for selecting instructions that best fit with the superscalar microprocessor. They abide by the rules to create code that best uses the superscalar implementation. All IBM Z compilers and JVMs are constantly updated to benefit from new instructions and advances in microprocessor designs.

3.4.13 Extended Translation Facility

The z/Architecture instruction set has instructions in support of the Extended Translation Facility. They are used in data conversion operations for Unicode data, which causes applications that are enabled for Unicode or globalization to be more efficient. These data-encoding formats are used in web services, grid, and on-demand environments in which XML and SOAP technologies are used. The High Level Assembler supports the Extended Translation Facility instructions.

3.4.14 Instruction set extensions

The processor supports the following instructions to support functions:

- ▶ Hexadecimal floating point instructions for various unnormalized multiply and multiply add instructions.
- ▶ Divide engine scheduler.
- ▶ Second generation of BCD-RR architecture.
- ▶ Modulo arithmetic.
- ▶ Immediate instructions, including various add, compare, OR, exclusive-OR, subtract, load, and insert formats. The use of these instructions improves performance.
- ▶ Load instructions for handling unsigned halfwords, such as those used for Unicode.
- ▶ Cryptographic instructions, which are known as the MSA, offer the full complement of the AES, SHA-1, SHA-2, and DES algorithms. They also include functions for random number generation.
- ▶ Extended Translate Facility-3 instructions, which are enhanced to conform with the current Unicode 4.0 standard.
- ▶ Assist instructions that help eliminate hypervisor processor usage.
- ▶ SIMD instructions, which allow the parallel processing of multiple elements in a single instruction.

3.4.15 Transactional Execution

The Transactional Execution (TX) capability, which is known in the industry as *hardware transactional memory*, runs a group of instructions atomically; that is, all of their results are committed or no result is committed. The execution is optimistic. The instructions are run, but previous state values are saved in a transactional memory. If the transaction succeeds, the saved values are discarded; otherwise, they are used to restore the original values.

The Transaction Execution Facility provides instructions, including declaring the beginning and end of a transaction, and canceling the transaction. TX is expected to provide significant performance benefits and scalability by avoiding most locks. This benefit is especially important for heavily threaded applications, such as Java.

3.4.16 Runtime Instrumentation

Runtime Instrumentation (RI) is a hardware facility for managed run times, such as the Java Runtime Environment (JRE). RI allows dynamic optimization of code generation as it is being run. It requires fewer system resources than the current software-only profiling, and provides information about hardware and program characteristics. RI also enhances JRE in making the correct decision by providing real-time feedback.

3.5 Processor unit functions

The PU functions are described in this section.

3.5.1 Overview

All PUs on a z15 server are physically identical. When the system is initialized, one integrated firmware processor (IFP) is allocated from the pool of PUs that is available for the entire system. The other PUs can be characterized to specific functions (CP, IFL, ICF, zIIP, or SAP).

The function that is assigned to a PU is set by the Licensed Internal Code (LIC). The LIC is loaded when the system is initialized at power-on reset (POR) and the PUs are *characterized*.

Only characterized PUs include a designated function. Non-characterized PUs are considered spares. Order at least one CP, IFL, or ICF on a z15 server.

This design brings outstanding flexibility to z15 servers because any PU can assume any available characterization. The design also plays an essential role in system availability because PU characterization can be done dynamically, with no system outage.

For more information about software level support of functions and features, see Chapter 7, “Operating system support” on page 253.

Concurrent upgrades

For all z15 T01 features that have more processor units (PUs) installed (non-characterized) than activated, concurrent upgrades can be done by LIC activation, which assigns a PU function to a previously non-characterized PU. No hardware changes are required. The upgrade can be done concurrently through the following facilities:

- ▶ Customer Initiated Upgrade (CIU) for permanent upgrades
- ▶ On/Off Capacity on Demand (On/Off CoD) for temporary upgrades
- ▶ Capacity BackUp (CBU) for temporary upgrades
- ▶ Capacity for Planned Event (CPE) for temporary upgrades

If the PU chips in the installed CPC drawers have no available remaining PUs, an upgrade results in a feature upgrade and the installation of an extra CPC drawer. Field add (MES) of a CPC drawer is possible for z15 Model T01 features Max34 and Max71 only. These features can be upgraded to a Max108 provided initial order for the CPC Reserve features FC 2271 or FC 2272. CPC drawer installation is nondisruptive, but takes more time than a simple LIC upgrade. Features Max145 and Max190 are factory build only.

For more information about Capacity on Demand, see Chapter 8, “System upgrades” on page 333.

PU sparing

If a PU failure occurs, the failed PU’s characterization is dynamically and transparently reassigned to a spare PU. z15 servers have two spare PUs. PUs that are not characterized on a CPC configuration can also be used as extra spare PUs. For more information about PU sparing, see 3.5.10, “Sparing rules” on page 127.

PU pools

PU pools are defined as CPs, IFLs, ICFs, and zIIPs are grouped in their own pools from where they can be managed separately. This configuration significantly simplifies capacity planning and management for LPARs. The separation also affects weight management because CP and zIIP weights can be managed separately. For more information, see “[PU weighting](#)” on page 114.

All assigned PUs are grouped in the PU pool. These PUs are dispatched to online logical PUs. As an example, consider a z15 server with 10 CPs, 2 IFLs, 5 zIIPs, and 1 ICF. This system has a PU pool of 18 PUs, called the *pool width*. Subdivision defines the following pools:

- ▶ A CP pool of 10 CPs
- ▶ An ICF pool of one ICF
- ▶ An IFL pool of two IFLs
- ▶ A zIIP pool of five zIIPs

PUs are placed in the pools in the following circumstances:

- ▶ When the system is POREd
- ▶ At the time of a concurrent upgrade
- ▶ As a result of adding PUs during a CBU
- ▶ Following a capacity on-demand upgrade through On/Off CoD or CIU

PUs are removed from their pools when a concurrent downgrade occurs as the result of the removal of a CBU. They are also removed through the On/Off CoD process and the conversion of a PU. When a dedicated LPAR is activated, its PUs are taken from the correct pools. This process is also the case when an LPAR logically configures a PU as on, if the width of the pool allows for it.

For an LPAR, logical PUs are dispatched from the supporting pool only. The logical CPs are dispatched from the CP pool, logical zIIPs from the zIIP pool, logical IFLs from the IFL pool, and the logical ICFs from the ICF pool.

PU weighting

Because CPs, zIIPs, IFLs, and ICFs have their own pools from where they are dispatched, they can be given their own weights. For more information about PU pools and processing weights, see the *IBM Z Processor Resource/Systems Manager Planning Guide*, SB10-7175.

3.5.2 Central processors

A central processor (CP) is a PU that uses the full z/Architecture instruction set. It can run z/Architecture-based operating systems (z/OS, z/VM, TPF, z/TPF, z/VSE, and Linux on Z) and the Coupling Facility Control Code (CFCC). Up to 190 PUs can be characterized as CPs, depending on the configuration.

The z15 server can be initialized in LPAR (PR/SM) mode or in Dynamic Partition Manager (DPM) mode.

CPs are defined as dedicated or shared. Reserved CPs can be defined to an LPAR to allow for nondisruptive image upgrades. If the operating system in the LPAR supports the logical processor add function, reserved processors are no longer needed. Regardless of the installed model, an LPAR can have up to 190 logical CPs that are defined (the sum of active and reserved logical CPs). In practice, define no more CPs than the operating system supports.

All PUs that are characterized as CPs within a configuration are grouped into the CP pool. The CP pool can be seen on the Hardware Management Console (HMC) workplace. Any z/Architecture operating systems and CFCCs can run on CPs that are assigned from the CP pool.

The z15 T01 server recognizes four distinct capacity settings for CPs. Full-capacity CPs are identified as CP7. In addition to full-capacity CPs, three subcapacity settings (CP6, CP5, and CP4), each for up to 34 PUs, are offered.

The following capacity settings appear in hardware descriptions:

- ▶ CP4 Feature Code 1941 (up to 34 PUs)
- ▶ CP5 Feature Code 1942 (up to 34 PUs)
- ▶ CP6 Feature Code 1943 (up to 34 PUs)
- ▶ CP7 Feature Code 1944 (up to 34 PUs)

Granular capacity adds 102 subcapacity settings to the 190 capacity settings that are available with full capacity CPs (CP7). Each of the 102 subcapacity settings applies to up to 34 CPs only, independent of the model installed.

Information about CPs in the remainder of this chapter applies to all CP capacity settings, unless indicated otherwise. For more information about granular capacity, see Chapter 2.3.3, “PU characterization” on page 52.

3.5.3 Integrated Facility for Linux (FC 1945)

An IFL is a PU that can be used to run Linux, Linux guests on z/VM operating systems, and Secure Service Container (SSC). Up to 190 PUs can be characterized as IFLs, depending on the configuration.

Note: IFLs can be dedicated to a Linux, a z/VM, or LPAR, or can be shared by multiple Linux guests, z/VM LPARs, or SSC that are running on the same z15 server. Only z/VM, Linux on Z operating systems, SSC, and designated software products can run on IFLs. IFLs are orderable by using FC 1945.

IFL pool

All PUs that are characterized as IFLs within a configuration are grouped into the IFL pool. The IFL pool can be seen on the HMC workplace.

IFLs do not change the model capacity identifier of the z15 server. Software product license charges that are based on the model capacity identifier are not affected by the addition of IFLs.

Unassigned IFLs

An IFL that is purchased but not activated is registered as an unassigned IFL (FC 1948). When the system is later upgraded with another IFL, the system recognizes that an IFL was purchased and is present.

The allowable number of IFLs and Unassigned IFLs numbers per feature is listed in Table 3-1.

Table 3-1 IFLs and Unassigned IFLs per feature

| Features | Max34 | Max71 | Max108 | Max145 | Max190 |
|---------------------------------------|-------|-------|--------|--------|--------|
| Maximum of IFLs FC 1933 | 34 | 71 | 108 | 145 | 190 |
| Maximum of Unassigned IFLs FC 1937 | 33 | 70 | 107 | 144 | 189 |

3.5.4 Internal Coupling Facility (FC 1946)

An Internal Coupling Facility (ICF) is a PU that is used to run the CFCC for Parallel Sysplex environments. Within the sum of all unassigned PUs in up to five CPC drawers, up to 190 ICFs can be characterized, depending on the model. However, the maximum number of ICFs that can be defined on a coupling facility LPAR is limited to 16. ICFs are orderable by using FC 1946.

The allowable number of zIIPs for each model is listed in Table 3-2.

Table 3-2 ICFs per feature

| Features | Max34 | Max71 | Max108 | Max145 | Max190 |
|-----------------|-------|-------|--------|--------|--------|
| Maximum of ICFs | 34 | 71 | 108 | 145 | 190 |

ICFs exclusively run CFCC. ICFs do not change the model capacity identifier of the z15 system. Software product license charges that are based on the model capacity identifier are not affected by the addition of ICFs.

All ICFs within a configuration are grouped into the ICF pool. The ICF pool can be seen on the HMC workplace.

The ICFs can be used by coupling facility LPARs only. ICFs are dedicated or shared. ICFs can be dedicated to a CF LPAR, or shared by multiple CF LPARs that run on the same system. However, having an LPAR with dedicated and shared ICFs at the same time is not possible.

Coupling Thin Interrupts (default CF LPAR setting with z15)

With the introduction of Driver 15F (zEC12 and zBC12), the IBM z/Architecture provides a new thin interrupt class called *Coupling Thin Interrupts*. The capabilities that are provided by hardware, firmware, and software support the generation of coupling-related “thin interrupts” when the following situations occur:

- ▶ On the coupling facility (CF) side:
 - A CF command or a CF signal (arrival of a CF-to-CF duplexing signal) is received by a shared-engine CF image.
 - The completion of a CF signal that was previously sent by the CF occurs (completion of a CF-to-CF duplexing signal).
- ▶ On the z/OS side:
 - CF signal is received by a shared-engine z/OS image (arrival of a List Notification signal).
 - An asynchronous CF operation completes.

The interrupt causes the receiving partition to be dispatched by an LPAR if it is not dispatched. This process allows the request, signal, or request completion to be recognized and processed in a more timely manner.

After the image is dispatched, “poll for work” logic in CFCC and z/OS can be used largely as-is to locate and process the work. The new interrupt expedites the redispaching of the partition.

LPAR presents these Coupling Thin Interrupts to the guest partition, so CFCC and z/OS both require interrupt handler support that can deal with them. CFCC also changes to relinquish control of the processor when all available pending work is exhausted, or when the LPAR undispaches it off the shared processor, whichever comes first.

CF processor combinations

A CF image can have one of the following combinations that are defined in the image profile:

- ▶ Dedicated ICFs
- ▶ Shared ICFs
- ▶ Dedicated CPs
- ▶ Shared CPs

Shared ICFs add flexibility. However, running only with shared coupling facility PUs (ICFs or CPs) is not a preferable production configuration. It is preferable for a production CF to operate by using dedicated ICFs.

In Figure 3-13, the CPC on the left has two environments that are defined (production and test), and each has one z/OS and one coupling facility image. The coupling facility images share an ICF.

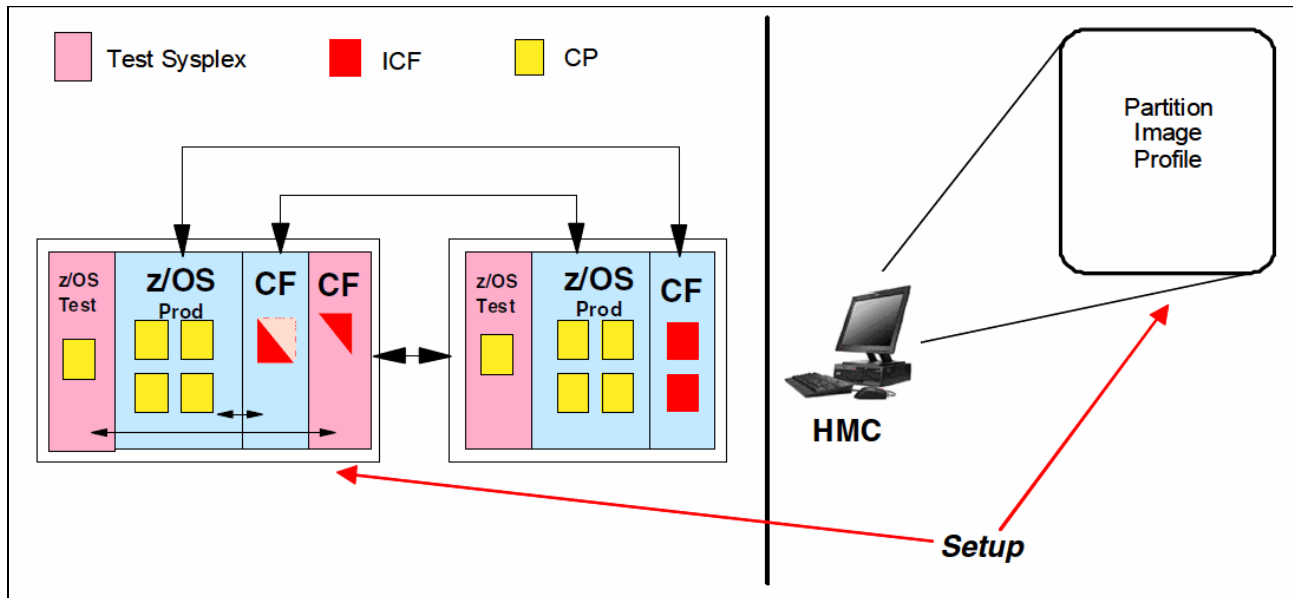


Figure 3-13 ICF options - shared ICFs

The LPAR processing weights are used to define how much processor capacity each CF image can include. The capped option can also be set for a test CF image to protect the production environment.

Connections between these z/OS and CF images can use internal coupling links to avoid the use of real (external) coupling links, and get the best link bandwidth available.

Dynamic CF dispatching

The *dynamic coupling facility dispatching* function has a dispatching algorithm that you can use to define a backup CF in an LPAR on the system. When this LPAR is in backup mode, it uses few processor resources. When the backup CF becomes active, only the resources that are necessary to provide coupling are allocated.

DYNDISP allows more environments with multiple CF images to coexist in a server, and to share CF engines with reasonable performance. For more information, see 3.9.3, “Dynamic CF dispatching” on page 148.

Coupling Facility Processor scalability

CF work management and dispatcher changed to improve efficiency as processors are added to scale up the capacity of a CF image.

CF images support up to 16 processors. To obtain sufficient CF capacity, customers might be forced to split the CF workload across more CF images. However, this change brings more configuration complexity and granularity (more, smaller CF images, more coupling links, and logical CHPIDs to define and manage for connectivity, and so on).

To improve CF processor scaling for the customer’s CF images and to make effective use of more processors as the sysplex workload increases, CF work management and dispatcher provide the following improvements (z15):

- ▶ Non-prioritized (FIFO-based) work queues, which avoids overhead of maintaining ordered queues in the CF.
- ▶ Streamlined system-managed duplexing protocol, which avoids costly latching deadlocks that can occur between primary and secondary structure.
- ▶ “Functionally specialized” ICF processors that operate for CF images with dedicated processors defined under certain conditions that realizes the following benefits:
 - One “functionally specialized” processor for inspecting suspended commands
 - One “functionally specialized” processor for pulling in new commands
 - The remaining processors are non-specialized for general CF request processing
 - Avoids many inter-processor contentions that were associated with CF dispatching

Coupling Facility Enhancements with CFCC level 24

z15 supports 384 coupling CHPIDs per CPC (50% increase over z14). In addition, the CFCC level 24 adds the following features:

- ▶ CFCC Fair Latch Manager2

This feature is an enhancement to the internals of the Coupling Facility (CFCC) dispatcher to provide CF work management efficiency and processor scalability improvements, and improve the “fairness” of arbitration for internal CF resource latches across tasks.

The tasks that are waiting for CF latches are not placed on the global suspend queue at all; instead they are placed on latch-specific waiter queues for the exact instance of the latch they are requesting, and in the exact order in which they requested the latch. As a result, the global suspend queue is much less heavily used and thus is much less a source of global contention or cache misses in the CF.

Also, when a latch is released, the specific latch's latch waiter queue is used to transfer ownership of the latch directly to the next request in line (or multiple requests, in the case of a shared latch), and make that task (or tasks) ready to run, with the transferred latch already held. No possibility of any unfairness or "cutters" in line between the time the latch is released versus when it is reobtained.

For managing latches properly for structures that are System-Managed (SM) synchronous duplexing, it is now important for the CF to understand which of the duplexed pair of requests operates as the "master" versus "slave" from a latching perspective, requiring more SM duplexing setup information from z/OS

z/OS XCF/XES toleration APAR support for z15 is required to provide this enhancement.

► Message Path SYID Resiliency Enhancement

When a z/OS system IPLs, message paths are supposed to be deactivated by using system reset, and their SYIDs are supposed to be cleared in the process. During the IPL, z/OS then reactivates the message paths with a new SYID that represents the new instance of z/OS that is using the paths.

On rare occasions, a message path might not be deactivated during system reset or IPL processing, which leaves the message path active with the z/OS image's OLD, now-obsolete SYID. Because the path erroneously remained active, z/OS does not see any need to reactivate it with a new, correct SYID.

From the CF's perspective, the incorrect SYID persists, and prevents delivery of signals to the z/OS image that is using that message path.

With z15, CFCC provides a new resiliency mechanism that transparently recovers for this "missing" message path deactivate (if and when that situation ever occurs).

The CF provides more information to z/OS about every message path that appears active; namely, the current SYID with which the message path is registered in the CF. Whenever z/OS interrogates the state of the message paths to the CF, z/OS checks this SYID information for currency and correctness. If an obsolete or incorrect SYID exists in the message path for any reason, z/OS performs the following steps:

- i. Requests non-disruptive gathering of diagnostic information for the affected message paths and CF image
- ii. Reactivates the message path with the correct SYID for the current z/OS image to seamlessly correct the problem

This enhancement requires z/OS XCF/XES exploitation APAR support for z15.

► Shared-Engine CF Default is changed to "DYNDISP=THIN"

The CF operates with the following Dynamic Dispatching (DYNDISP) models:

- DYNDISP=OFF: LPAR time-slicing completely controls the CF processor; the processor polls the entire time it is dispatched by LPAR, and it is idle (not dispatched) when undispached by LPAR. The result is least efficient sharing, worst shared-engine performance.
- DYNDISP=ON: An optimization over pure LPAR timeslicing, in which the CFCC code judiciously sets timer interrupts to give LPAR initiative to redispach it, and the CF sometimes voluntarily gives up control of the shared processor when it runs out of work to do. The result is more efficient sharing, better shared-engine performance. This setting is the default setting for CF LPAR running on z15.
- DYNDISP=THIN: An interrupt-driven model in which the CF processor is dispatched in response to a set of events that generate Thin Interrupts and runs until it runs out of things to do, then gives up control voluntarily (until the next interrupt causes it to get dispatched again). This model is the most efficient sharing, best shared-engine performance.

Thin Interrupt support is available since zEC12/zBC12, and proved to be efficient and performant in numerous different test and customer shared-engine coupling facility configurations.

For CFCC running on z15, DYNDISP=THIN is now the default mode of operation for coupling facility images that use shared processors.

- ▶ CF monopolization avoidance
 - With z15 T01/T02, the CF dispatcher will monitor in real-time the number of CF tasks that have a command assigned to them for a given structure, on a structure by structure basis.
 - When the number of CF tasks being used by any given structure exceeds a model-dependent CF threshold, and a global threshold on the number of active tasks is also exceeded, the structure will be considered to be “monopolizing” the CF, and z/OS will be informed of this monopolization.
 - New support in z/OS will observe the monopolization state for a structure, and start to selectively queue and throttle incoming requests to the CF, on a structure-specific basis – while other requests, for other “non-monopolizing” structures and workloads, are completely unaffected.
 - z/OS will dynamically manage the queue of requests for the “monopolizing” structures to limit the number of active CF requests (parallelism) to them, and will monitor the CF’s monopolization state information so as to observe the structure becoming “non-monopolized” again, so that request processing can eventually revert back to a non-throttled mode of operation.
 - The overall goal of z/OS anti-monopolization support is to protect the ability of ALL well-behaved structures and workloads to access the CF, and get their requests processed in the CF in a timely fashion – while implementing queueing and throttling mechanisms in z/OS to hold back the specific abusive workloads that are causing problems for other workloads.

z/OS XCF/XES exploitation APAR support is required to provide this functionality.

3.5.5 IBM Z Integrated Information Processor (FC 1947)

A zIIP⁵ reduces the standard processor (CP) capacity requirements for z/OS Java, XML system services applications, and a portion of work of z/OS Communications Server and Db2 UDB for z/OS Version 8 or later, which frees up capacity for other workload requirements.

A zIIP enables eligible z/OS workloads to have a portion of them directed for execution to a processor that is characterized as a zIIP. The zIIPs do not increase the MSU value of the processor and so do not affect the IBM software license changes.

z15 is the third generation of IBM Z processors to support SMT. z15 servers implement two threads per core on IFLs and zIIPs. SMT must be enabled at the LPAR level and supported by the z/OS operating system. SMT was enhanced for z15 and it is enabled for SAPs by default (no customer intervention required).

New with z/OS 2.4, the z/OS Container Extensions⁶ allows deployment of Linux on Z software components, such as Docker Containers in a z/OS system, in direct support of z/OS workloads without requiring a separately provisioned Linux server, while maintaining overall solution operational control within z/OS and with z/OS qualities of service. Workload deployed in z/OS Container Extensions is zIIP eligible.

⁵ IBM z Systems® Application Assist Processors (zAAPs) are not available since z14 servers. A zAAP workload is dispatched to available zIIPs (zAAP on zIIP capability).

⁶ z/OS Container Extensions require ordering the Container Hosting Foundation feature with Z systems (FC 0104).

How zIIPs work

zIIPs are designed for supporting designated z/OS workloads. One of the workloads is Java code execution. When Java code must be run (for example, under control of IBM WebSphere), the z/OS JVM calls the function of the zIIP. The z/OS dispatcher then suspends the JVM task on the CP that it is running on and dispatches it on an available zIIP. After the Java application code execution is finished, z/OS redispaches the JVM task on an available CP. After this process occurs, normal processing is resumed.

This process reduces the CP time that is needed to run Java WebSphere applications, which frees that capacity for other workloads.

The logical flow of Java code that is running on a z15 server that has a zIIP available is shown in Figure 3-14. When JVM starts the execution of a Java program, it passes control to the z/OS dispatcher that verifies the availability of a zIIP.

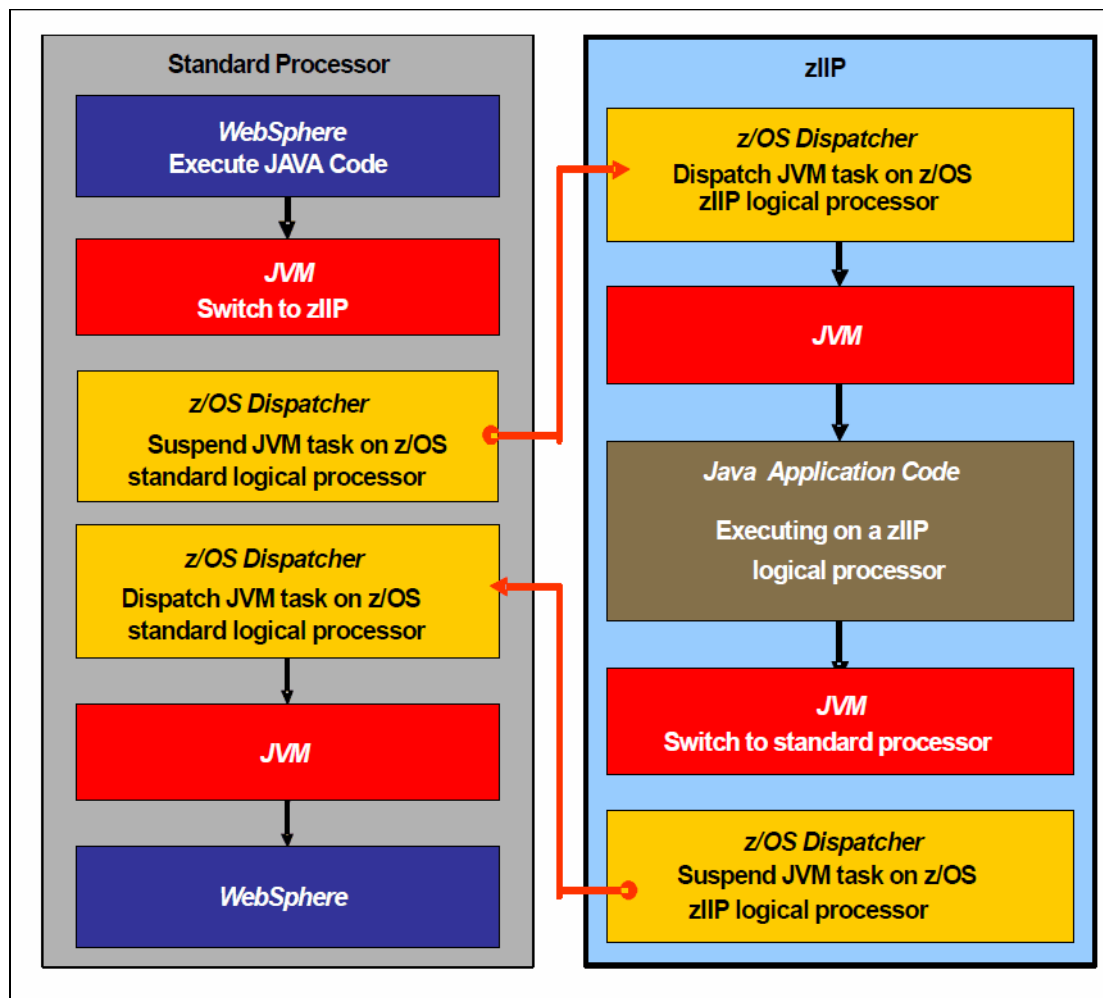


Figure 3-14 Logical flow of Java code execution on a zIIP

The availability is treated in the following manner:

- ▶ If a zIIP is available (not busy), the dispatcher suspends the JVM task on the CP and assigns the Java task to the zIIP. When the task returns control to the JVM, it passes control back to the dispatcher. The dispatcher then reassigns the JVM code execution to a CP.
- ▶ If no zIIP is available (all busy), the z/OS dispatcher allows the Java task to run on a standard CP. This process depends on the option that is used in the OPT statement in the IEAOPTxx member of SYS1.PARMLIB.

A zIIP runs only IBM authorized code. This IBM authorized code includes the z/OS JVM in association with parts of system code, such as the z/OS dispatcher and supervisor services. A zIIP cannot process I/O or clock comparator interruptions, and it does not support operator controls, such as IPL.

Java application code can run on a CP or a zIIP. The installation can manage the use of CPs so that Java application code runs only on CPs, only on zIIPs, or on both.

Two execution options for zIIP-eligible code execution are available. These options are user-specified in IEAOPTxx and can be dynamically altered by using the **SET OPT** command. The following options are supported for z/OS⁷:

- ▶ Option 1: Java dispatching by priority (IIPHONORPRIORITY=YES)

This option is the default option and specifies that CPs must not automatically consider zIIP-eligible work for dispatching on them. The zIIP-eligible work is dispatched on the zIIP engines until Workload Manager (WLM) determines that the zIIPs are overcommitted. WLM then requests help from the CPs. When help is requested, the CPs consider dispatching zIIP-eligible work on the CPs themselves based on the dispatching priority relative to other workloads. When the zIIP engines are no longer overcommitted, the CPs stop considering zIIP-eligible work for dispatch.

This option runs as much zIIP-eligible work on zIIPs as possible. It also allows it to spill over onto the CPs only when the zIIPs are overcommitted.

- ▶ Option 2: Java dispatching by priority (IIPHONORPRIORITY=NO)

zIIP-eligible work runs on zIIPs only while at least one zIIP engine is online. zIIP-eligible work is not normally dispatched on a CP, even if the zIIPs are overcommitted and CPs are unused. The exception is that zIIP-eligible work can sometimes run on a CP to resolve resource conflicts.

Therefore, zIIP-eligible work does not affect the CP utilization that is used for reporting through the subcapacity reporting tool (SCRT), no matter how busy the zIIPs are.

If zIIPs are defined to the LPAR but are not online, the zIIP-eligible work units are processed by CPs in order of priority. The system ignores the IIPHONORPRIORITY parameter in this case and handles the work as though it had no eligibility to zIIPs.

zIIPs provide the following benefits:

- ▶ Potential software cost savings.
- ▶ Simplification of infrastructure as a result of the colocation and integration of new applications with their associated database systems and transaction middleware, such as Db2, IMS, or CICS. Simplification can happen, for example, by introducing a uniform security environment, and by reducing the number of TCP/IP programming stacks and system interconnect links.

⁷ z/OS V2R1 and later (older z/OS versions are out of support)

- ▶ Prevention of processing latencies that occur if Java application servers and their database servers are deployed on separate server platforms.

The following Db2 UDB for z/OS V8 or later workloads are eligible to run in Service Request Block (SRB) mode:

- ▶ Query processing of network-connected applications that access the Db2 database over a TCP/IP connection by using IBM Distributed Relational Database Architecture™ (DRDA). DRDA enables relational data to be distributed among multiple systems. It is native to Db2 for z/OS, which reduces the need for more gateway products that can affect performance and availability. The application uses the DRDA requester or server to access a remote database. IBM Db2 Connect is an example of a DRDA application requester.
- ▶ Star schema query processing, which is mostly used in business intelligence work. A *star schema* is a relational database schema for representing multidimensional data. It stores data in a central fact table and is surrounded by more dimension tables that hold information about each perspective of the data. For example, a star schema query joins various dimensions of a star schema data set.
- ▶ Db2 utilities that are used for index maintenance, such as LOAD, REORG, and REBUILD. Indexes allow quick access to table rows, but over time, the databases become less efficient and must be maintained as data in large databases is manipulated.

The zIIP runs portions of eligible database workloads, which helps to free computer capacity and lower software costs. Not all Db2 workloads are eligible for zIIP processing. Db2 UDB for z/OS V8 and later gives z/OS the information to direct portions of the work to the zIIP. The result is that in every user situation, different variables determine how much work is redirected to the zIIP.

On a z15 server, the following workloads can also benefit from zIIPs:

- ▶ z/OS Communications Server uses the zIIP for eligible Internet Protocol Security (IPSec) network encryption workloads. This configuration requires z/OS V1R10 or later. Portions of IPSec processing take advantage of the zIIPs, specifically end-to-end encryption with IPSec. The IPSec function moves a portion of the processing from the general-purpose processors to the zIIPs. In addition, to run the encryption processing, the zIIP also handles the cryptographic validation of message integrity and IPSec header processing.
- ▶ z/OS Global Mirror, formerly known as Extended Remote Copy (XRC), also uses the zIIP. Most z/OS Data Facility Storage Management Subsystem (DFSMS) system data mover (SDM) processing that is associated with z/OS Global Mirror can run on the zIIP. This configuration requires z/OS V1R10 or later releases.
- ▶ The first IBM user of z/OS XML system services is Db2 V9. For Db2 V9 before the z/OS XML System Services enhancement, z/OS XML System Services non-validating parsing was partially directed to zIIPs when used as part of a distributed Db2 request through DRDA. This enhancement benefits Db2 V9 by making all z/OS XML System Services non-validating parsing eligible to zIIPs. This configuration is possible when processing is used as part of any workload that is running in enclave SRB mode.
- ▶ z/OS Communications Server also allows the HiperSockets Multiple Write operation for outbound large messages (originating from z/OS) to be run by a zIIP. Application workloads that are based on XML, HTTP, SOAP, and Java, and traditional file transfer can benefit.
- ▶ For business intelligence, IBM Scalable Architecture for Financial Reporting provides a high-volume, high-performance reporting solution by running many diverse queries in z/OS batch. It can also be eligible for zIIP.

For more information about zIIP and eligible workloads, see [the IBM zIIP website](#).

zIIP installation

One CP must be installed with or before any zIIP is installed. In z15 T01, the zIIP-to-CP ratio is 2:1⁸, which means that up to 126 zIIPs on feature Max190 can be characterized. The allowable number of zIIPs for each model is listed in Table 3-3.

Table 3-3 Number of zIIPs per feature

| Features | Max34 | Max71 | Max108 | Max145 | Max190 |
|---------------|--------|--------|--------|--------|---------|
| Maximum zIIPs | 0 - 22 | 0 - 46 | 0 - 70 | 0 - 96 | 0 - 126 |

zIIPs are orderable by using FC 1947. Up to two zIIPs can be ordered for each CP or marked CP configured in the system. If the installed CPC drawer has no remaining unassigned PUs, the assignment of the next zIIP might require the installation of another CPC drawer.

PUs that are characterized as zIIPs within a configuration are grouped into the zIIP pool. This configuration allows zIIPs to have their own processing weights, independent of the weight of parent CPs. The zIIP pool can be seen on the hardware console.

The number of permanent zIIPs plus temporary zIIPs cannot exceed twice the number of purchased CPs plus temporary CPs. Also, the number of temporary zIIPs cannot exceed the number of permanent zIIPs.

zIIPs and logical partition definitions

zIIPs are dedicated or shared, depending on whether they are part of an LPAR with dedicated or shared CPs. In an LPAR, at least one CP must be defined before zIIPs for that partition can be defined. The number of zIIPs that are available in the system is the number of zIIPs that can be defined to an LPAR.

LPAR: In an LPAR, as many zIIPs as are available can be defined together with at least one CP.

3.5.6 System assist processors

A system assist processor (SAP) is a PU that runs the channel subsystem LIC to control I/O operations. All SAPs run I/O operations for all LPARs. As with z14 server, in z15 SMT is enabled⁹ for SAPs. All features include standard SAPs configured. The number of standard SAPs depends on the z15 feature, as listed in Table 3-4.

Table 3-4 SAPs per feature

| Features | Max34 FC 0655 | Max71 FC 0656 | Max108 FC 0657 | Max145FC FC 0658 | Max190 FC 0659 |
|---------------|------------------|------------------|-------------------|---------------------|-------------------|
| Standard SAPs | 4 | 8 | 12 | 16 | 22 |

SAP configuration

A standard SAP configuration provides a well-balanced system for most environments. However, some application environments feature high I/O rates, typically Transaction Processing Facility (TPF) environments. In this case, more SAPs can be ordered. Assigning more SAPs can increase the capability of the channel subsystem to run I/O operations. In z15

⁸ 2:1 ratio can be exceeded (during boost periods) if System Recovery Boost Upgrade (FC 9930 and FC 6802) is used for activating temporary zIIP capacity.

⁹ Enabled by default, cannot be changed or altered by user

systems, the number of SAPs plus the number of optional SAPs cannot exceed firmware limit of 128 threads.

Optional other orderable SAPs (FC 1949)

The option to order more SAPs is available on all models (FC 1949). These extra SAPs increase the capacity of the channel subsystem to run I/O operations, which is suggested for TPF environments. The maximum number of optional extra orderable SAPs depends on the configuration and the number of available uncharacterized PUs. The number of SAPs is listed in Table 3-5.

Table 3-5 Optional SAPs per feature

| Features | Max34 | Max71 | Max108 | Max145 | Max190 |
|---------------|-------|-------|--------|--------|--------|
| Optional SAPs | 0 - 8 | 0 - 8 | 0 - 8 | 0 - 8 | 0 - 8 |

3.5.7 Reserved processors

Reserved processors are defined by PR/SM to allow for a nondisruptive capacity upgrade. Reserved processors are similar to spare logical processors and can be shared or dedicated. Reserved CPs can be defined to an LPAR dynamically to allow for nondisruptive image upgrades.

Reserved processors can be dynamically configured online by an operating system that supports this function if enough unassigned PUs are available to satisfy the request. The PR/SM rules that govern logical processor activation remain unchanged.

By using reserved processors, you can define more logical processors than the number of available CPs, IFLs, ICFs, and zIIPs in the configuration to an LPAR. This process makes it possible to nondisruptively configure online more logical processors after more CPs, IFLs, ICFs, and zIIPs are made available concurrently. They can be made available with one of the capacity on-demand options.

The maximum number of reserved processors that can be defined to an LPAR depends on the number of logical processors that are defined. The maximum number of logical processors plus reserved processors is 190. If the operating system in the LPAR supports the logical processor add function, reserved processors are no longer needed.

Do not define more active and reserved processors than the operating system for the LPAR can support. For more information about logical processors and reserved processors and their definitions, see 3.7, “Logical partitioning” on page 132.

3.5.8 Integrated firmware processor

An IFP is allocated from the pool of PUs and is available for the entire system. Unlike other characterized PUs, the IFP is standard and not defined by the client. It is a single PU that is dedicated solely to supporting the following *native* Peripheral Component Interconnect Express (PCIe) features:

- ▶ 10 GbE Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) Express
- ▶ 25GbE and 10GbE RoCE Express2.1
- ▶ 25GbE and 10GbE RoCE Express2
- ▶ Coupling Express Long Reach

The IFP is also initialized at POR. The IFP supports Resource Group (RG) LIC¹⁰ to provide native PCIe I/O feature management and virtualization functions.

3.5.9 Processor unit assignment

The processor unit assignment of characterized PUs is done at POR time, when the system is initialized. The initial assignment rules keep PUs of the same characterization type grouped as much as possible in relation to PU chips and CPC drawer boundaries to optimize shared cache usage.

The z15 T01 PU assignment is based on CPC drawer plug order (*not* “ordering”). Feature upgrade provides more processor (CPC) drawers. Max108 cannot be upgraded because the supposed targeted features (Max145 and Max190) are factory built only.

The CPC drawers are populated from the bottom up. This process defines following the low-order and the high-order CPC drawers:

- ▶ CPC drawer 1 (CPC 0 at position A10): Plug order 1 (low-order CPC drawer)
- ▶ CPC drawer 2 (CPC 1 at position A15): Plug order 2
- ▶ CPC drawer 3 (CPC 2 at position A20): Plug order 3 (high-order CPC drawer)

The assignment rules comply with the following order:

- ▶ Spare: CPC drawers 0 and 1 are assigned one spare each on the high PU chip. In the feature Max34, both spares are assigned to CPC drawer 0.
- ▶ IFP: One IFP is assigned to CPC drawer 0.
- ▶ SAPs: Spread across CPC drawers and high PU chips. Each CPC drawer has at least four standard SAPs. Start with the highest PU chip high core, then the next highest PU chip high core. This process prevents all the SAPs from being assigned on one PU chip.
- ▶ IFLs and ICFs: Assign IFLs and ICFs to cores on chips in higher CPC drawers working downward.
- ▶ CPs and zIIPs: Assign CPs and zIIPs to cores on chips in lower CPC drawers working upward.

These rules are intended to isolate, as much as possible, on different CPC drawers and even on different PU chips, processors that are used by different operating systems. This configuration ensures that different operating systems do not use the same shared caches. For example, CPs and zIIPs are all used by z/OS, and can benefit by using the same shared caches. However, IFLs are used by z/VM and Linux, and ICFs are used by CFCC. Therefore, for performance reasons, the assignment rules prevent them from sharing L3 and L4 caches with z/OS processors.

This initial PU assignment, which is done at POR, can be dynamically rearranged by an LPAR by swapping an active core to a core in a different PU chip in a different CPC drawer or cluster to improve system performance. For more information, see “LPAR dynamic PU reassignment” on page 138.

When a CPC drawer is added concurrently after POR and new LPARs are activated, or processor capacity for active partitions is dynamically expanded, the extra PU capacity can be assigned from the new CPC drawer. The processor unit assignment rules consider the newly installed CPC drawer dynamically.

¹⁰ IBM zHyperLink Express1.1 and IBM zHyperLink Express are not managed by Resource Groups LIC

3.5.10 Sparing rules

On a z15 T01 system, two PUs are reserved as spares. The spare PUs are available to replace any two characterized PUs, whether they are CP, IFL, ICF, zIIP, SAP, or IFP.

Systems with a failed PU for which no spare is available *call home* for a replacement. A system with a failed PU that is spared and requires an SCM to be replaced (referred to as a *pending repair*) can still be upgraded when sufficient PUs are available.

Transparent CP, IFL, ICF, zIIP, SAP, and IFP sparing

Depending on the model, sparing of CP, IFL, ICF, zIIP, SAP, and IFP is transparent and does not require operating system or operator intervention.

With *transparent sparing*, the status of the application that was running on the failed processor is preserved. The application continues processing on a newly assigned CP, IFL, ICF, zIIP, SAP, or IFP (allocated to one of the spare PUs) without client intervention.

Application preservation

If no spare PU is available, *application preservation* (z/OS only) is started. The state of the failing processor is passed to another active processor that is used by the operating system. Through operating system recovery services, the task is resumed successfully (in most cases, without client intervention).

Dynamic SAP and IFP sparing and reassignment

Dynamic recovery is provided if a failure of the SAP or IFP occurs. If the SAP or IFP fails, and if a spare PU is available, the spare PU is dynamically assigned as a new SAP or IFP. If no spare PU is available, and more than one CP is characterized, a characterized CP is reassigned as an SAP or IFP. In either case, client intervention is not required. This capability eliminates an unplanned outage and allows a service action to be deferred to a more convenient time.

3.5.11 CPC drawer numbering

z15 T01 CPC drawer numbering starts with CPC 0, the first installed CPC drawer. It is in frame A at A10. The second one, in the same frame, at A15. The third one, in the same frame at A20. The fourth and the fifth are both in frame B (locations B10 and B15). Figure 3-15 shows CPC drawer numbering.

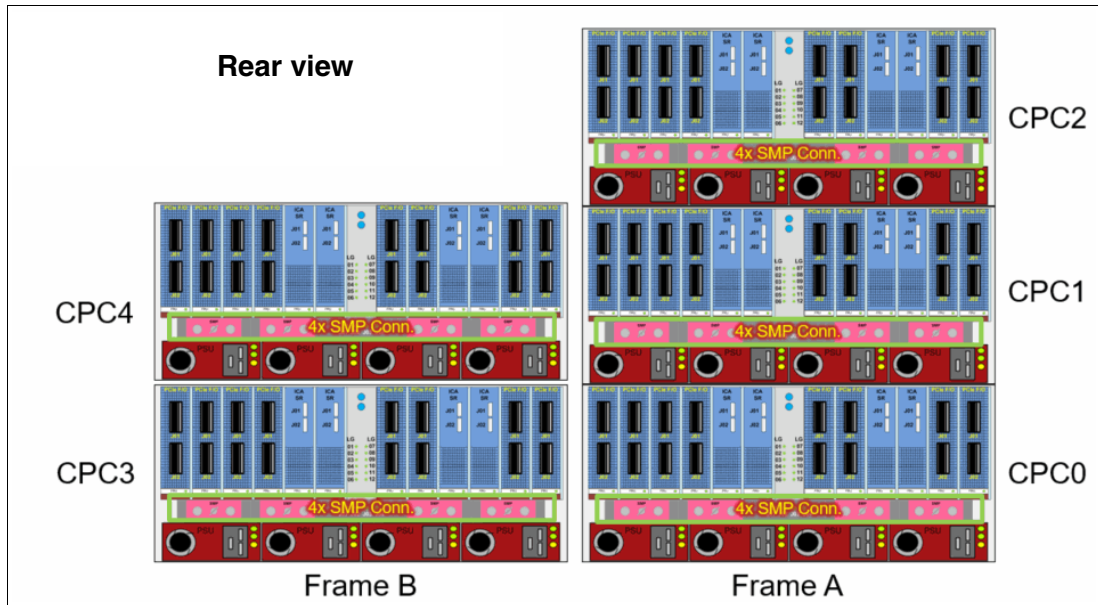


Figure 3-15 CPC drawer numbering

3.6 Memory design

Various considerations of the z15 memory design are described in this section.

3.6.1 Overview

The z15 T01 memory design also provides flexibility, high availability, and the following upgrades:

- ▶ Concurrent memory upgrades if the physically installed capacity is not yet reached
z15 servers can have more physically installed memory than the initial available capacity. Memory upgrades within the physically installed capacity can be done concurrently by LIC, and no hardware changes are required. However, memory upgrades *cannot* be done through CBU or On/Off CoD.
- ▶ Concurrent memory upgrades if the physically installed capacity is reached
Physical memory upgrades require a processor drawer to be removed and reinstalled after replacing the memory cards in the processor drawer. Except for the feature Max34, the combination of enhanced drawer availability and the flexible memory option allows you to concurrently add memory to the system. For more information, see 2.5.5, “Drawer replacement and memory” on page 65, and 2.5.7, “Flexible Memory Option” on page 65.

When the total capacity that is installed has more usable memory than required for a configuration, the LIC Configuration Control (LICCC) determines how much memory is used from each processor drawer. The sum of the LICCC provided memory from each CPC drawer is the amount that is available for use in the system.

Memory allocation

When the system is activated by using a POR, PR/SM determines the total installed memory and the customer enabled memory. Later in the process, during LPAR activation, PR/SM assigns and allocates each partition memory according to their image profile.

PR/SM controls all physical memory, and can make physical memory available to the configuration when a CPC drawer is added.

In older IBM Z processors, memory allocation was striped across the available CPC drawers because relatively fast¹¹ connectivity existed between the drawers. Splitting the work between all of the memory controllers allowed a smooth performance variability.

The memory allocation algorithm changed starting with IBM z13. For z15, PR/SM tries to allocate memory into a single CPC drawer. If memory does not fit into a single drawer, PR/SM tries to allocate the memory into the CPC drawer with the most processor entitlement.

The PR/SM memory and logical processor resources allocation goal is to place all partition resources on a single CPC drawer, if possible. The resources, such as memory and logical processors, are assigned to the logical partitions at the time of their activation. Later on, when all partitions are activated, PR/SM can move memory between CPC drawers to benefit the performance of each LPAR, without operating system knowledge. This process was done on the previous families of IBM Z servers only for PUs that use PR/SM dynamic PU reallocation.

With z15 servers, this process occurs whenever the configuration changes, such as in the following circumstances:

- ▶ Activating or deactivating an LPAR
- ▶ Changing the LPARs processing weights
- ▶ Upgrading the system through a temporary or permanent record
- ▶ Downgrading the system through deactivation of a temporary record

PR/SM schedules a global reoptimization of the resources in use. It does so by looking at all the partitions that are active and prioritizing them based on their processing entitlement and weights, which creates a high and low priority rank. Then, the resources, such as logical processors and memory, can be moved from one CPC drawer to another to address the priority ranks that were created.

When partitions are activated, PR/SM tries to find a home assignment CPC drawer, home assignment node, and home assignment chip for the logical processors that are defined to them. The PR/SM goal is to allocate all the partition logical processors and memory to a single CPC drawer (the home drawer for that partition).

If all logical processors can be assigned to a home drawer and the partition-defined memory is greater than what is available in that drawer, the exceeding memory amount is allocated on another CPC drawer. If all the logical processors cannot fit in one CPC drawer, the remaining logical processors spill to another CPC drawer. When that overlap occurs, PR/SM stripes the memory (if possible) across the CPC drawers where the logical processors are assigned.

The process of reallocating memory is based on the *memory copy/reassign* function, which is used to allow enhanced drawer availability (EDA) and concurrent drawer replacement (CDR)¹². This process was enhanced starting with z13 and z13s to provide more efficiency and speed to the process without affecting system performance.

z15 T01 implements a faster dynamic memory reallocation mechanism, which is especially useful during service operations (EDA and CDR). PR/SM controls the reassignment of the content of a specific physical memory array in one CPC drawer to a physical memory array in another CPC drawer. To do accomplish this task, PR/SM uses all the available physical memory in the system. This memory includes the memory that is not in use by the system that is available but not purchased by the client, and the planned memory options, if installed.

¹¹ Relatively fast to the processor clock frequency

¹² In previous IBM Z generations (before z13), these service operations were known as enhanced book availability (EBA) and concurrent book repair (CBR).

Because of the memory allocation algorithm, systems that undergo many miscellaneous equipment specification (MES) upgrades for memory can have different memory mixes and quantities in all processor drawers of the system. If the memory fails, it is technically feasible to run a POR of the system with the remaining working memory resources. After the POR completes, the memory distribution across the processor drawers is different, as is the total amount of available memory.

Large page support

By default, page frames are allocated with a 4 KB size. z15 servers also support large page sizes of 1 MB or 2 GB. The first z/OS release that supports 1 MB pages is z/OS V1R9. Linux on Z 1 MB pages support is available in SUSE Linux Enterprise Server 10 SP2 and Red Hat Enterprise Linux (RHEL) 5.2 and later.

The TLB reduces the amount of time that is required to translate a virtual address to a real address. This translation is done by dynamic address translation (DAT) when it must find the correct page for the correct address space. Each TLB entry represents one page. As with other buffers or caches, lines are discarded from the TLB on a least recently used (LRU) basis.

The worst-case translation time occurs when a TLB miss occurs and the segment table (which is needed to find the page table) and the page table (which is needed to find the entry for the particular page in question) are not in cache. This case involves two complete real memory access delays plus the address translation delay. The duration of a processor cycle is much shorter than the duration of a memory cycle, so a TLB miss is relatively costly.

It is preferable to have addresses in the TLB. With 4 K pages, holding all of the addresses for 1 MB of storage takes 256 TLB lines. When 1 MB pages are used, it takes only one TLB line. Therefore, large page size users have a much smaller TLB footprint.

Large pages allow the TLB to better represent a large working set and suffer fewer TLB misses by allowing a single TLB entry to cover more address translations.

Users of large pages are better represented in the TLB and are expected to see performance improvements in elapsed time and processor usage. These improvements are because DAT and memory operations are part of processor busy time, even though the processor waits for memory operations to complete without processing anything else in the meantime.

To overcome the processor usage that is associated with creating a 1 MB page, a process must run for some time. It also must maintain frequent memory access to keep the pertinent addresses in the TLB.

Short-running work does not overcome the processor usage. Short processes with small working sets are expected to receive little or no improvement. Long-running work with high memory-access frequency is the best candidate to benefit from large pages.

Long-running work with low memory-access frequency is less likely to maintain its entries in the TLB. However, when it does run, few address translations are required to resolve all of the memory it needs. Therefore, a long-running process can benefit even without frequent memory access.

Weigh the benefits of whether something in this category must use large pages as a result of the system-level costs of tying up real storage. A balance exists between the performance of a process that uses large pages and the performance of the remaining work on the system.

On z15 servers, 1 MB large pages become pageable if Virtual Flash Memory¹³ is available and enabled. They are available only for 64-bit virtual private storage, such as virtual memory that is above 2 GB.

It is easy to assume that increasing the TLB size is a feasible option to deal with TLB-miss situations. However, this process is not as straightforward as it seems. As the size of the TLB increases, so does the processor usage that is involved in managing the TLB's contents. Correct sizing of the TLB is subject to complex statistical modeling to find the optimal tradeoff between size and performance.

3.6.2 Main storage

Main storage consist of memory space addressable by programs and storage that is not directly addressable by programs. Non-addressable storage includes the hardware system area (HSA).

Main storage provides the following functions:

- ▶ Data storage and retrieval for PUs and I/O
- ▶ Communication with PUs and I/O
- ▶ Communication with and control of optional expanded storage
- ▶ Error checking and correction

Main storage can be accessed by all processors, but cannot be shared between LPARs. Any system image (LPAR) must include a defined main storage size. This defined main storage is allocated exclusively to the LPAR during partition activation.

3.6.3 Hardware system area

The HSA is a non-addressable storage area that contains system LIC and configuration-dependent control blocks. On z15 T01 servers, the HSA has a fixed size of 256 GB and is not part of the purchased memory that you order and install.

The fixed size of the HSA eliminates planning for future expansion of the HSA because the hardware configuration definition (HCD)/input/output configuration program (IOCP) always reserves space for the following items:

- ▶ Six channel subsystems (CSSs)
- ▶ A total of 15 LPARs in CSSs 1 through 5, and 10 LPARs for the sixth CSS for a total of 85 LPARs
- ▶ Subchannel set 0 with 63.75-K devices in each CSS
- ▶ Subchannel set 1 with 64-K devices in each CSS
- ▶ Subchannel set 2 with 64-K devices in each CSS
- ▶ Subchannel set 3 with 64-K devices in each CSS

The HSA features sufficient reserved space to allow for dynamic I/O reconfiguration changes to the maximum capability of the processor.

¹³ Virtual Flash Memory has replaced IBM zFlash Express. No carry forward of zFlash Express exists.

3.6.4 Virtual Flash Memory (FC 0643)

IBM Virtual Flash Memory (VFM, FC 0643) is the replacement for the Flash Express features that were available on the IBM zEC12 and IBM z13. No application changes are required to change from IBM Flash Express to VFM.

For z15 T01, IBM VFM provides up to 6.0 TB of virtual flash memory in 512 GB increments. The minimum is 0, while the maximum is 12 features. The number of VFM features ordered reduces the maximum orderable memory for the z15 server.

3.7 Logical partitioning

The logical partitioning features are described in this section.

3.7.1 Overview

Logical partitioning is a function that is implemented by the PR/SM on z15. z15 can run in LPAR mode, or in DPM mode. DPM provides a GUI for PR/SM to manage I/O resources dynamically.

PR/SM is aware of the processor drawer structure on z15 servers. However, LPARs do not feature this awareness. LPARs feature resources that are allocated to them from various physical resources. From a systems standpoint, LPARs have no control over these physical resources, but the PR/SM functions do have this control.

PR/SM manages and optimizes allocation and the dispatching of work on the physical topology. Most physical topology that was handled by the operating systems is the responsibility of PR/SM.

As described in 3.5.9, “Processor unit assignment” on page 126, the initial PU assignment is done during POR by using rules to optimize cache usage. This step is the “physical” step, where CPs, zIIPs, IFLs, ICFs, and SAPs are allocated on the processor drawers.

When an LPAR is activated, PR/SM builds logical processors and allocates memory for the LPAR.

PR/SM assigns all logical processors to one CPC drawer that are packed into chips of that drawer and cooperates with operating system use of HiperDispatch.

New for z15, all processor types can be dynamically reassigned except IFPs.

Memory allocation changed from the previous IBM Z servers. Partition memory is now allocated based on processor drawer affinity and striped across processor clusters. For more information, see “Memory allocation” on page 128.

Logical processors are dispatched by PR/SM on physical processors. The assignment topology that is used by PR/SM to dispatch logical processors on physical PUs is also based on cache usage optimization.

Processor drawers and logical node level assignments are more important because they optimize L4 cache usage. Therefore, logical processors from a specific LPAR are packed into a processor drawer as much as possible.

PR/SM optimizes chip assignments within the assigned processor drawers (or drawers) to maximize L3 cache efficiency. Logical processors from an LPAR are dispatched on physical processors on the same PU chip as much as possible. The number of processors per chip (up to 12) matches the number of z/OS processor affinity queues that is used by HiperDispatch, which achieves optimal cache usage within an affinity logical node.

PR/SM also tries to redispach a logical processor on the same physical processor to optimize private cache (L1 and L2) usage.

HiperDispatch

PR/SM and z/OS work in tandem to use processor resources more efficiently. HiperDispatch is a function that combines the dispatcher actions and the knowledge that PR/SM has about the topology of the system.

Performance can be optimized by redispershing units of work to the same processor group, which keeps processes running near their cached instructions and data, and minimizes transfers of data ownership among processors and processor drawers.

The nested topology is returned to z/OS by the Store System Information (STSI) instruction. HiperDispatch uses the information to concentrate logical processors around shared caches (L3 at PU chip level, and L4 at drawer level), and dynamically optimizes the assignment of logical processors and units of work.

z/OS dispatcher manages multiple queues, called *affinity queues*, with a target number of eight processors per queue, which fits well onto a single PU chip. These queues are used to assign work to as few logical processors as are needed for an LPAR workload. Therefore, even if the LPAR is defined with many logical processors, HiperDispatch optimizes this number of processors to be near the required capacity. The optimal number of processors to be used is kept within a processor drawer boundary, when possible.

Tip: z/VM V6.4 and later also support HiperDispatch.

Logical partitions

PR/SM enables z15 T01 systems to be initialized for a logically partitioned operation, supporting up to 85 LPARs. Each LPAR can run its own operating system image in any image mode, independently from the other LPARs.

An LPAR can be added, removed, activated, or deactivated at any time. Changing the number of LPARs is not disruptive and does not require a POR. Certain facilities might not be available to all operating systems because the facilities might have software corequisites.

Each LPAR has the following resources that are the same as a real CPC:

- ▶ Processors

Called *logical processors*, they can be defined as CPs, IFLs, ICFs, or zIIPs. They can be dedicated to an LPAR or shared among LPARs. When shared, a processor weight can be defined to provide the required level of processor resources to an LPAR. Also, the capping option can be turned on, which prevents an LPAR from acquiring more than its defined weight and limits its processor consumption.

LPARs for z/OS can have CP and zIIP logical processors. The logical processor types can be defined as all dedicated or all shared. The zIIP support is available in z/OS.

The weight and number of online logical processors of an LPAR can be dynamically managed by the LPAR CPU Management function of the Intelligent Resource Director (IRD). These functions can be used to achieve the defined goals of this specific partition and of the overall system. The provisioning architecture of z15 systems, as described in Chapter 8, “System upgrades” on page 333, adds a dimension to the dynamic management of LPARs.

PR/SM is enhanced to support an option to limit the amount of physical processor capacity that is used by an individual LPAR when a PU is defined as a general-purpose processor (CP) or an IFL that is shared across a set of LPARs.

This enhancement is designed to provide a physical capacity limit that is enforced as an absolute (versus relative) limit. It is not affected by changes to the logical or physical configuration of the system. This physical capacity limit can be specified in units of CPs or IFLs. The Change LPAR Controls and Customize Activation Profiles tasks on the HMC were enhanced to support this new function.

For the z/OS Workload License Charges (WLC) pricing metric and metrics that are based on it, such as Advanced Workload License Charges (AWLC), an LPAR *defined capacity* can be set. This defined capacity enables the soft capping function. Workload charging introduces the capability to pay software license fees that are based on the processor utilization of the LPAR on which the product is running, rather than on the total capacity of the system. Consider the following points:

- In support of WLC, the user can specify a defined capacity in millions of service units (MSUs) per hour. The defined capacity sets the capacity of an individual LPAR when soft capping is selected.

The defined capacity value is specified on the Options tab in the Customize Image Profiles window.

- WLM keeps a four-hour rolling average of the processor usage of the LPAR. When the four-hour average processor consumption exceeds the defined capacity limit, WLM dynamically activates LPAR capping (soft capping). When the rolling four-hour average returns below the defined capacity, the soft cap is removed.

For more information about WLM, see *System Programmer's Guide to: Workload Manager*, SG24-6472.

For more information about software licensing, see 7.8, “Software licensing” on page 328.

Weight settings: When defined capacity is used to define an uncapped LPAR's capacity, carefully consider the weight settings of that LPAR. If the weight is much smaller than the defined capacity, PR/SM uses a discontinuous cap pattern to achieve the defined capacity setting. This configuration means PR/SM alternates between capping the LPAR at the MSU value that corresponds to the relative weight settings, and no capping at all. It is best to avoid this scenario and instead attempt to establish a defined capacity that is equal or close to the relative weight.

► Memory

Memory (main storage) must be dedicated to an LPAR. The defined storage must be available during the LPAR activation; otherwise, the LPAR activation fails.

Reserved storage can be defined to an LPAR, which enables nondisruptive memory addition to and removal from an LPAR by using the LPAR dynamic storage reconfiguration (z/OS and z/VM). For more information, see 3.7.5, “LPAR dynamic storage reconfiguration” on page 142.

► Channels

Channels can be shared between LPARs by including the partition name in the partition list of a channel-path identifier (CHPID). I/O configurations are defined by the IOCP or the HCD with the CHPID mapping tool (CMT). The CMT is an optional tool that is used to map CHPIDs onto physical channel IDs (PCHIDs). PCHIDs represent the physical location of a port on a card in an I/O cage, I/O drawer, or PCIe I/O drawer.

IOCP is available on the z/OS, z/VM, and z/VSE operating systems, and as a stand-alone program on the hardware console. For more information, see *IBM Z Input/Output Configuration Program User's Guide for ICP IOCP*, SB10-7172. HCD is available on the z/OS and z/VM operating systems. Consult the appropriate 8561DEVICE Preventive Service Planning (PSP) buckets before implementation.

Fibre Channel connection (FICON) channels can be managed by the Dynamic CHPID Management (DCM) function of the Intelligent Resource Director. DCM enables the system to respond to ever-changing channel requirements by moving channels from lesser-used control units to more heavily used control units, as needed.

Modes of operation

The modes of operation are listed in Table 3-6. All available mode combinations, including their operating modes and processor types, operating systems, and addressing modes, also are listed. Only the currently supported versions of operating systems are considered.

Table 3-6 z15 modes of operation

| Image mode | PU type | Operating system | Addressing mode |
|----------------------|-----------------------|--------------------------------------|-----------------|
| General ^a | CP and zIIP | z/OS z/VM | 64-bit |
| | CP | z/VSE Linux on IBM Z z/TPF | 64-bit |
| Coupling facility | ICF or CP | CFCC | 64-bit |
| Linux only | IFL or CP | Linux on Z (64-bit) | 64-bit |
| | | z/VM | |
| | | Linux on Z (31-bit) | 31-bit |
| z/VM | CP, IFL, zIIP, or ICF | z/VM | 64-bit |
| SSC ^b | IFL or CP | z/VSE Network Appliance ^c | 64 bit |

a. General mode uses 64-bit z/Architecture

b. Secure Service Container

c. More appliances to be announced and supported in the future

The 64-bit z/Architecture mode has no special operating mode because the architecture mode is not an attribute of the definable images operating mode. The 64-bit operating systems are in 31-bit mode at IPL and change to 64-bit mode during their initialization. The operating system is responsible for taking advantage of the addressing capabilities that are provided by the architectural mode.

For information about operating system support, see Chapter 7, “Operating system support” on page 253.

Logically partitioned mode

If the z15 T01 system runs in LPAR mode, each of the 85 LPARs can be defined to operate in one of the following image modes:

- ▶ General mode to run the following systems:
 - A z/Architecture operating system, on dedicated or shared CPs
 - A Linux on Z operating system, on dedicated or shared CPs
 - z/OS, on any of the following processor units:
 - Dedicated or shared CPs
 - Dedicated CPs *and* dedicated zIIPs
 - Shared CPs *and* shared zIIPs

zIIP usage: zIIPs can be defined to General mode or z/VM mode image, as listed in Table 3-6 on page 135. However, zIIPs are used only by z/OS. Other operating systems cannot use zIIPs, even if they are defined to the LPAR. z/VM V6R4 and later support real and virtual zIIPs to guest z/OS systems.

- ▶ General mode is also used to run the z/TPF operating system on dedicated or shared CPs
- ▶ CF mode, by loading the CFCC code into the LPAR that is defined as one of the following types:
 - Dedicated or shared CPs
 - Dedicated or shared ICFs
- ▶ Linux only mode to run the following systems:
 - A Linux on Z operating system, on either of the following types:
 - Dedicated or shared IFLs
 - Dedicated or shared CPs
 - A z/VM operating system, on either of the following types:
 - Dedicated or shared IFLs
 - Dedicated or shared CPs
- ▶ z/VM mode to run z/VM on dedicated or shared CPs or IFLs, plus zIIPs and ICFs
- ▶ Secure Service Container (SSC) mode LPAR can run on dedicated or shared:
 - CPs
 - IFLs

All LPAR modes, required characterized PUs, operating systems, and the PU characterizations that can be configured to an LPAR image are listed in Table 3-7. The available combinations of dedicated (DED) and shared (SHR) processors are also included. For all combinations, an LPAR also can include reserved processors that are defined, which allows for nondisruptive LPAR upgrades.

Table 3-7 LPAR mode and PU usage

| LPAR mode | PU type | Operating systems | PUs usage |
|-----------|-------------------------|--|--|
| General | CPs | z/Architecture operating systems Linux on Z | CPs DED or CPs SHR |
| | CPs <i>and</i> zIIPs | z/OS z/VM (guest exploitation) | CPs DED or zIIPs DED or CPs SHR or zIIPs SHR |
| General | CPs | z/TPF | CPs DED or CPs SHR |

| LPAR mode | PU type | Operating systems | PUs usage |
|-------------------|---------------------------------|-------------------------|--|
| Coupling facility | ICFs <i>or</i> CPs | CFCC | ICFs DED or ICFs SHR or CPs DED or CPs SHR |
| Linux only | IFLs <i>or</i> CPs | Linux on Z z/VM | IFLs DED or IFLs SHR or CPs DED or CPs SHR |
| z/VM | CPs, IFLs, zIIPs, or ICFs | z/VM (V6R4 and later) | All PUs must be SHR or DED |
| SSC ^a | IFLs, <i>or</i> CPs | z/VSE Network Appliance | IFLs DED or IFLs SHR or CPs DED or CPs SHR |

a. Secure Service Container

Dynamically adding or deleting a logical partition name

Dynamically adding or deleting an LPAR name is the ability to add or delete LPARs and their associated I/O resources to or from the configuration without a POR.

The extra channel subsystem and multiple image facility (MIF) image ID pairs (CSSID/MIFID) can be later assigned to an LPAR for use (or later removed). This process can be done through dynamic I/O commands by using the HCD. At the same time, required channels must be defined for the new LPAR.

Partition profile: Cryptographic coprocessors are not tied to partition numbers or MIF IDs. They are set up with Adjunct Processor (AP) numbers and domain indexes. These numbers are assigned to a partition profile of a given name. The client assigns these AP numbers and domains to the partitions and continues to have the responsibility to clear them out when their profiles change.

Adding logical processors to a logical partition

Logical processors can be concurrently added to an LPAR by defining them as reserved in the image profile and later configuring them online to the operating system by using the appropriate console commands. Logical processors also can be concurrently added to a logical partition dynamically by using the Support Element (SE) “Logical Processor Add” function under the CPC Operational Customization task. This SE function allows the initial and reserved processor values to be dynamically changed. The operating system must support the dynamic addition¹⁴ of these resources.

Adding a crypto feature to a logical partition

You can plan the addition of Crypto Express7S features to an LPAR on the crypto page in the image profile by defining the Cryptographic Candidate List, and the Usage and Control Domain indexes, in the partition profile. By using the Change LPAR Cryptographic Controls task, you can add crypto adapters dynamically to an LPAR without an outage of the LPAR. Also, dynamic deletion or moving of these features does not require pre-planning. Support is provided in z/OS, z/VM, z/VSE, Secure Service Container (based on appliance requirements), and Linux on Z.

¹⁴ In z/OS, this support is available since Version 1 Release 10 (z/OS V1.10), while z/VM supports this addition since z/VM V5.4, and z/VSE since V4.3. However, z15 supports z/OS 2.1 and later, z/VSE 6.2 and z/VM 6.4 and later.

LPAR dynamic PU reassignment

The system configuration is enhanced to optimize the PU-to-CPC drawer assignment of physical processors dynamically. The initial assignment of client-usable physical processors to physical processor drawers can change dynamically to better suit the LPAR configurations that are in use. For more information, see 3.5.9, “Processor unit assignment” on page 126.

Swapping of specialty engines and general processors with each other, with spare PUs, or with both, can occur as the system attempts to compact LPAR configurations into physical configurations that span the least number of processor drawers.

LPAR dynamic PU reassignment can swap client processors of different types between processor drawers. For example, reassignment can swap an IFL on processor drawer 1 with a CP on processor drawer 2. Swaps can also occur between PU chips within a processor drawer or a node and can include spare PUs. The goals are to pack the LPAR on fewer processor drawers and also on fewer PU chips, based on the z15 processor drawers' topology. The effect of this process is evident in dedicated and shared LPARs that use HiperDispatch.

LPAR dynamic PU reassignment is transparent to operating systems.

LPAR group capacity limit (LPAR group absolute capping)

The group capacity limit feature allows the definition of a group of LPARs on a z15 system, and limits the combined capacity usage by those LPARs. This process allows the system to manage the group so that the group capacity limits in MSUs per hour are not exceeded. To take advantage of this feature, you must be running z/OS V2.1 or later in the all LPARs in the group.

PR/SM and WLM work together to enforce the capacity that is defined for the group and the capacity that is optionally defined for each individual LPAR.

LPAR absolute capping

Absolute capping is a logical partition control that was made available with zEC12 and is supported on z15 systems. With this support, PR/SM and the HMC are enhanced to support a new option to limit the amount of physical processor capacity that is used by an individual LPAR when a PU is defined as a general-purpose processor (CP), zIIP, or an IFL processor that is shared across a set of LPARs.

Unlike traditional LPAR capping, absolute capping is designed to provide a physical capacity limit that is enforced as an absolute (versus relative) value that is not affected by changes to the virtual or physical configuration of the system.

Absolute capping provides an optional maximum capacity setting for logical partitions that is specified in the absolute processors capacity (for example, 5.00 CPs or 2.75 IFLs). This setting is specified independently by processor type (namely CPs, zIIPs, and IFLs) and provides an enforceable upper limit on the amount of the specified processor type that can be used in a partition.

Absolute capping is ideal for processor types and operating systems that the z/OS WLM cannot control. Absolute capping is not intended as a replacement for defined capacity or group capacity for z/OS, which are managed by WLM.

Absolute capping can be used with any z/OS, z/VM, or Linux on Z LPAR (that is running on an IBM Z server). If specified for a z/OS LPAR, absolute capping can be used concurrently with defined capacity or group capacity management for z/OS. When used concurrently, the absolute capacity limit becomes effective before other capping controls.

Dynamic Partition Manager mode

DPM is an IBM Z server operation mode that provides a simplified approach to create and manage virtualized environments, which reduces the barriers of its adoption for new and existing customers.

The implementation provides built-in integrated capabilities that allow advanced virtualization management on IBM Z servers. With DPM, you can use your Linux and virtualization skills while taking advantage of the full value of IBM Z hardware, robustness, and security in a workload optimized environment.

DPM provides facilities to define and run virtualized computing systems by using a firmware-managed environment that coordinate the physical system resources that are shared by the partitions. The partitions' resources include processors, memory, network, storage, crypto, and accelerators.

DPM provides a new mode of operation for IBM Z servers that provide the following services:

- ▶ Facilitates defining, configuring, and operating PR/SM LPARs in a similar way to how someone performs these tasks on another platform.
- ▶ Lays the foundation for a general IBM Z new user experience.

DPM is not another hypervisor for IBM Z servers. DPM uses the PR/SM hypervisor infrastructure and provides an intelligent interface on top of it that allows customers to define, use, and operate the platform virtualization without IBM Z experience or skills.

3.7.2 Storage operations

In z15 T01 systems, memory can be assigned as main storage, supporting up to 85 LPARs. Before you activate an LPAR, main storage must be defined to the LPAR. All installed storage can be configured as main storage. Each z/OS individual LPAR can be defined with a maximum of 4 TB of main storage. z/VM V6R4, z/VM V7R1, and z/VM V7R2 support 2 TB of main storage.

Memory *cannot* be shared between system images (LPARs). It is possible to dynamically reallocate storage resources for z/Architecture LPARs that run operating systems that support dynamic storage reconfiguration (DSR). This process is supported by z/OS, and z/VM. z/VM, in turn, virtualizes this support to its guests. For more information, see 3.7.5, “LPAR dynamic storage reconfiguration” on page 142.

Operating systems that run as guests of z/VM can use the z/VM capability of implementing virtual memory to guest virtual machines. The z/VM dedicated real storage can be shared between guest operating systems.

LPAR main storage allocation and usage

The IBM z15 T01 storage allocation and usage capabilities depend on the image mode and the operating system deployed in the LPAR.

Important: The memory allocation and usage depends on the operating system architecture and the tested (as per specific operating system documentation) limits.

While the maximum supported memory per LPAR for IBM z15 T01 is 16TB, each operating system has its own support specifications.

For general guidelines, check the *PR/SM Planning Guide*, SB10-7175 available on [IBM Resource Link](#) (IBM ID required to access Resource Link).

The following modes are provided:

► z/Architecture mode

In z/Architecture (General) mode, storage addressing is 64-bit, which allows for virtual addresses up to 16 exabytes (16 EB). The 64-bit architecture theoretically allows a maximum of 16 EB to be used as main storage. However, the current main storage limit for LPARs on a z15 T01 is 16 TB of main storage. The operating system that runs in z/Architecture mode must support the real storage. Currently, z/OS V2R4 supports up to 4 TB¹⁵ of real storage.

► CF mode

In CF mode, storage addressing is 64 bit for a CF image that runs at CFCC Level 12 or later. This configuration allows for an addressing range up to 16 EB. However, the current z15 T01 definition limit for CF LPARs is 16 TB of storage. The following CFCC levels are supported in a Sysplex with IBM z15:

- CFCC Level 24, available on z15 (Driver level 41)
- CFCC Level 23, available on z14 (Driver level 36)
- CFCC Level 22, available on z14 (Driver level 32)
- CFCC Level 21, available on z13 and z13s (Driver Level 27)
- CFCC Level 20, available for z13 servers with Driver Level 22

For more information, see [3.9.1, “CF Control Code” on page 145](#).

Expanded storage cannot be defined for a CF image. Only IBM CFCC can run in CF mode.

► Linux only mode

In Linux only mode, storage addressing can be 31 bit or 64 bit, depending on the operating system architecture and the operating system configuration.

Only Linux and z/VM operating systems can run in Linux only mode. Linux on Z 64-bit distributions (SUSE Linux Enterprise Server 11 and later, Red Hat RHEL 6 and later, and Ubuntu 16.04.5 LTS and later) use 64-bit addressing and operate in z/Architecture mode. z/VM also uses 64-bit addressing and operates in z/Architecture mode.

Note: For information about supported amount of memory check the Linux Distribution specific documentation.

► z/VM mode

In z/VM mode, certain types of processor units can be defined within one LPAR. This feature increases flexibility and simplifies systems management by allowing z/VM to run the following tasks in the same z/VM LPAR:

- Manage guests to operate Linux on Z on IFLs
- Operate z/VSE and z/OS on CPs
- Offload z/OS system software processor usage, such as Db2 workloads on zIIPs
- Provide an economical Java execution environment under z/OS on zIIPs

► Secure Service Container (SSC) mode

In SSC mode, storage addressing is 64 bit for an embedded product. The amount of main storage usable by the appliance code deployed in the SSC LPAR is documented by the appliance code supplier.

¹⁵ 1 TB if an I/O drawer is installed in the z13 system (carry forward only). z15, z14 do not support I/O drawer. z/OS V2.1 or later are supported at the time of this writing.

3.7.3 Reserved storage

Reserved storage can be optionally defined to an LPAR, allowing a nondisruptive image memory upgrade for this partition. Reserved storage can be defined to both central and expanded storage, and to any image mode except CF mode.

An LPAR must define an amount of main storage and optionally (if not a CF image), an amount of expanded storage. Both main storage and expanded storage can feature the following storage sizes defined:

- ▶ The *initial value* is the storage size that is allocated to the partition when it is activated.
- ▶ The *reserved value* is another storage capacity beyond its initial storage size that an LPAR can acquire dynamically. The reserved storage sizes that are defined to an LPAR do not have to be available when the partition is activated. They are predefined storage sizes to allow a storage increase, from an LPAR point of view.

Without the reserved storage definition, an LPAR storage upgrade is a disruptive process that requires the following steps:

1. Partition deactivation.
2. An initial storage size definition change.
3. Partition activation.

The extra storage capacity for an LPAR upgrade can come from the following sources:

- ▶ Any unused available storage
- ▶ Another partition that features released storage
- ▶ A memory upgrade

A concurrent LPAR storage upgrade uses DSR. z/OS uses the reconfigurable storage unit (RSU) definition to add or remove storage units in a nondisruptive way.

z/VM V6R4 and later releases support the dynamic addition of memory to a running LPAR by using reserved storage. It also virtualizes this support to its guests. Removing storage from the z/VM LPAR is disruptive. Removing memory from a z/VM guest is not disruptive to the z/VM LPAR.

SLES 11 supports concurrent add and remove.

3.7.4 Logical partition storage granularity

Granularity of main storage for an LPAR depends on the largest main storage amount that is defined for initial or reserved main storage, as listed in Table 3-8¹⁶.

Table 3-8 Logical partition main storage granularity (z15)

| Logical partition: Largest main storage amount | Logical partition: Main storage granularity |
|---|--|
| Main storage amount <= 512 GB | 1 GB |
| 512 GB < main storage amount <= 1 TB | 2 GB |
| 1 TB < main storage amount <= 2 TB | 4 GB |
| 2 TB < main storage amount <= 4 TB | 8 GB |
| 4 TB < main storage amount <= 8 TB | 16 GB |
| 8 TB < main storage amount <= 16 TB | 32 GB |

LPAR storage granularity information is required for LPAR image setup and for z/OS RSU definition. On z15 T01 LPARs are limited to a maximum size of 16 TB of main storage. However, the maximum amount of memory that is supported by z/OS V2.3 and z/OS V2.4 is 4 TB. For z/VM V6R4, V7R1, and V7R2 the limit is 2 TB.

3.7.5 LPAR dynamic storage reconfiguration

Dynamic storage reconfiguration on z15 systems allows an operating system that is running on an LPAR to add (nondisruptively) its reserved storage amount to its configuration. This process can occur only if unused storage exists. This unused storage can be obtained when another LPAR releases storage, or when a concurrent memory upgrade occurs.

With dynamic storage reconfiguration, the unused storage does not have to be continuous.

When an operating system running on an LPAR assigns a storage increment to its configuration, PR/SM determines whether any free storage increments are available. PR/SM then dynamically brings the storage online.

PR/SM dynamically takes offline a storage increment and makes it available to other partitions when an operating system running on an LPAR releases a storage increment.

3.8 Intelligent Resource Director

Intelligent Resource Director (IRD) is a z15 and IBM Z capability that is used only by z/OS. IRD is a function that optimizes processor and channel resource utilization across LPARs within a single IBM Z server.

This feature extends the concept of goal-oriented resource management. It does so by grouping system images that are on the same z15 or Z servers that are running in LPAR mode (and in the same Parallel Sysplex) into an *LPAR cluster*. This configuration allows WLM to manage resources (processor and I/O) across the entire cluster of system images and not only in one single image.

¹⁶ When defining an LPAR on the HMC, the 2G boundary should still be followed in PR/SM.

An LPAR cluster is shown in Figure 3-16. It contains three z/OS images and one Linux image that is managed by the cluster. Included as part of the entire Parallel Sysplex is another z/OS image and a CF image. In this example, the scope over which IRD has control is the defined LPAR cluster.

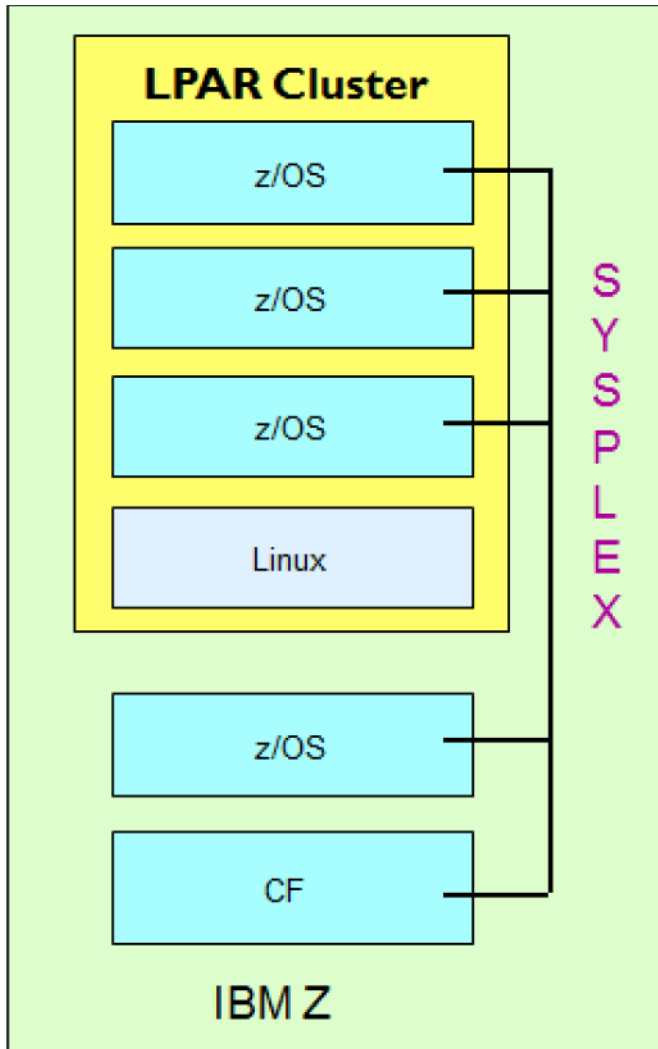


Figure 3-16 IRD LPAR cluster example

IRD features the following characteristics:

- ▶ IRD processor management

WLM dynamically adjusts the number of logical processors within an LPAR and the processor weight based on the WLM policy. The ability to move the processor weights across an LPAR cluster provides processing power where it is most needed, based on WLM goal mode policy.

The processor management function is automatically deactivated when HiperDispatch is active. However, the LPAR weight management function remains active with IRD with HiperDispatch. For more information about HiperDispatch, see 3.7, “Logical partitioning” on page 132.

HiperDispatch manages the number of logical CPs in use. It adjusts the number of logical processors within an LPAR to achieve the optimal balance between CP resources and the requirements of the workload.

HiperDispatch also adjusts the number of logical processors. The goal is to map the logical processor to as few physical processors as possible. This configuration uses the processor resources more efficiently by trying to stay within the local cache structure. Doing so makes efficient use of the advantages of the high-frequency microprocessors, and improves throughput and response times.

- ▶ Dynamic channel path management (DCM)

DCM moves FICON channel bandwidth between disk control units to address current processing needs. z15 systems support DCM within a channel subsystem.

- ▶ Channel subsystem priority queuing

This function on z15 and Z systems allows the priority queuing of I/O requests in the channel subsystem and the specification of relative priority among LPARs. When running in goal mode, WLM sets the priority for an LPAR and coordinates this activity among clustered LPARs.

For more information about implementing LPAR processor management under IRD, see *z/OS Intelligent Resource Director*, SG24-5952.

3.9 Clustering technology

Parallel Sysplex is the clustering technology that is used with IBM Z servers. The components of a Parallel Sysplex as implemented within the z/Architecture are shown in Figure 3-17. The example in Figure 3-17 shows one of many possible Parallel Sysplex configurations.

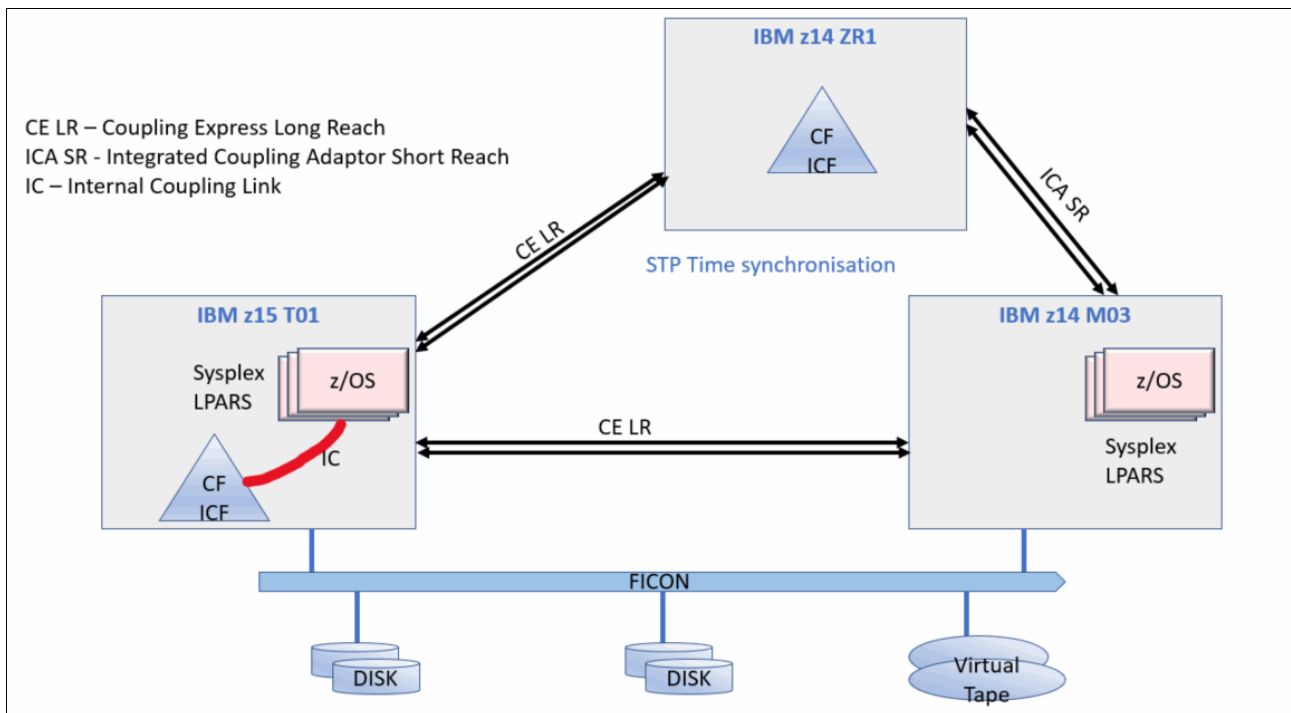


Figure 3-17 Sysplex hardware overview

Figure 3-17 shows a z15 Model T01 system that contains multiple z/OS sysplex partitions. It contains an internal CF (CF02), a z14 ZR1 system that contains a stand-alone CF (CF01), and a z14 M03 that contains multiple z/OS sysplex partitions.

STP over coupling links provides time synchronization to all systems. Selection of the appropriate CF link technology Coupling Express Long Reach (CE LR) or Integrate Coupling Adaptor Short Reach (ICA SR and SR1.1) depends on the system configuration and how distant they are physically. For more information about link technologies, see “Coupling links” on page 195.

Parallel Sysplex is an enabling technology that allows highly reliable, redundant, and robust IBM Z technology to achieve near-continuous availability. A Parallel Sysplex consists of one or more (z/OS) operating system images that are coupled through one or more Coupling Facilities.

A correctly configured Parallel Sysplex cluster maximizes availability in the following ways:

- ▶ **Continuous availability:** Changes can be introduced, such as software upgrades, one image at a time, while the remaining images continue to process work. For more information, see *Parallel Sysplex Application Considerations*, SG24-6523.
- ▶ **High capacity:** 1- 32 z/OS images in a Parallel Sysplex operating as a single system.
- ▶ **Dynamic workload balancing:** Because it is viewed as a single logical resource, work can be directed to any operating system image in a Parallel Sysplex cluster that has available capacity.
- ▶ **Systems management:** The architecture defines the infrastructure to satisfy client requirements for continuous availability. It also provides techniques for achieving simplified systems management consistent with this requirement.
- ▶ **Resource sharing:** Several base z/OS components use CF shared storage. This configuration enables sharing of physical resources with significant improvements in cost, performance, and simplified systems management.
- ▶ **Single logical system:** The collection of system images in the Parallel Sysplex is displayed as a single entity to the operator, user, and database administrator. A single system view means reduced complexity from operational and definition perspectives.
- ▶ **N-2 support:** Multiple hardware generations (normally three) are supported in the same Parallel Sysplex. This configuration provides for a gradual evolution of the systems in the Parallel Sysplex without changing all of them simultaneously. Software support for multiple releases or versions is also supported.

Note: Parallel sysplex coupling and timing links connectivity for z15 (M/T 8561) is supported to N-2 generation CPCs (z15, z14, z13/z13s).

Through state-of-the-art cluster technology, the power of multiple images can be harnessed to work in concert on common workloads. The IBM Z Parallel Sysplex cluster takes the commercial strengths of the platform to improved levels of system management, competitive price performance, scalable growth, and continuous availability.

3.9.1 CF Control Code

The LPAR that is running the CFCC can be on z15, z14/z14 ZR1, z13, and z13s systems. For more information about CFCC requirements for supported systems, see “Coupling facility and CFCC considerations” on page 288.

Consideration: z15, z14, z14 ZR1, z13, and z13s servers cannot coexist in the same sysplex with System zEC12 or earlier generation systems.

CFCC Level 24

CFCC level 24 is delivered on the z15 with driver level 41.

For CFCC Level 24 enhancements see “**Coupling Facility Enhancements with CFCC level 24**” on page 118.

CFCC Level 23

CFCC level 23 is delivered on the z14 with driver level 36. CFCC Level 23 introduces the following enhancements:

- ▶ Asynchronous cross-invalidate (XI) of CF cache structures. Requires PTF support for z/OS and explicit data manager support (Db2 V12 with PTFs). Consider the following points:
 - Instead of performing XI signals synchronously on every cache update request that causes them, data managers can “opt in” for the CF to perform these XIs asynchronously (and then sync them up with the CF at or before transaction completion). Data integrity is maintained if all XI signals complete by the time transaction locks are released.
 - Results in faster completion of cache update CF requests, especially with cross-site distance involved.
 - Provides improved cache structure service times and coupling efficiency.
- ▶ Coupling Facility hang detect enhancements provide a significant reduction in failure scope and client disruption (CF-level to structure-level), with no loss of FFDC collection capability:
 - When a hang is detected, in most cases the CF confines the scope of the failure to “structure damage” for the single CF structure the hung command was processing against, capture diagnostics with a non-disruptive CF dump, and continue operating without ending or restarting the CF image.
 - Provides a significant reduction in failure scope and client disruption (CF-level to structure-level), with no loss of FFDC collection capability.
- ▶ Coupling Facility ECR granular latching:
 - With this support, most CF list and lock structure ECR processing no longer use structure-wide latching. It serializes its execution by using the normal structure object latches that all mainline commands use.
 - Eliminates the performance degradation that is caused by structure-wide latching.
 - A few “edge conditions” in ECR processing still require structure-wide latching to be used to serialize them.
 - Cache structure ECR processing continues to require and use structure-wide latches for its serialization.

z14 servers with CFCC Level 23 require z/OS V1R13 or later, and z/VM V6R4 or later for virtual guest coupling.

CFCC Level 22

CFCC level 22 is delivered on the z14 servers with driver level D32. CFCC Level 22 introduces the following enhancements:

- ▶ CF Enhancements:
 - CF structure encryption

CF Structure encryption is transparent to CF-using middleware and applications, while CF users are unaware of and not involved in the encryption. All data and adjunct data that flows between z/OS and the CF is encrypted. The intent is to encrypt all data that might be sensitive.

Internal control information and related request metadata is not encrypted, including locks and lock structures.

z/OS generates the required structure-related encryption keys and does much of the key management automatically by using CFRM that uses secure, protected keys (never clear keys). Secure keys maintained in CFRM couple dataset.

- ▶ CF Asynchronous Duplexing for Lock Structures:
 - New asynchronous duplexing protocol for lock structures:
 - z/OS sends command to primary CF only
 - Primary CF processes command and returns result
 - Primary CF forwards description of required updates to secondary CF
 - Secondary CF updates secondary structure instance asynchronously
 - Provided for lock structures only:
 - z/OS V2.2 SPE with PTFs for APAR OA47796
 - Db2 V12 with PTFs
 - Most performance-sensitive structures for duplexing
 - Benefit/Value:
 - Db2 locking receives performance similar to simplex operations
 - Reduces CPU and CF link overhead
 - Avoids the overhead of synchronous protocol handshakes on every update
 - Duplexing failover much faster than log-based recovery
 - Targeted at multi-site clients who run split workloads at distance to make duplexing lock structures at distance practical.
- ▶ CF Processor Scalability:
 - CF work management and dispatcher changes to allow improved efficiency as processors are added to scale up the capacity of a CF image.
 - Functionally specialized ICF processors that operate for CF images having more than a threshold number of dedicated processors defined for them:
 - One functionally specialized processor for inspecting suspended commands.
 - One functionally specialized processor for pulling in new commands.
 - The remaining processors are non-specialized for general CF request processing.
 - Avoids many inter-processor contentions that were associated with CF dispatching
- ▶ Enable systems management applications to collect valid CF LPAR information through z/OS BCPIi:
 - System Type (CFCC)
 - System Level (CFCC LEVEL)
 - Dynamic Dispatch settings to indicate CF state (dedicated, shared, and Thin Interrupt), which are useful when investigating functional performance problems

To support an upgrade from one CFCC level to the next, different levels of CFCC can be run concurrently while the CF LPARs are running on different servers. CF LPARs that run on the same server share the CFCC level.

z15 servers (CFCC level 24) can coexist in a sysplex with CFCC levels 23, 22, 21, and 20.

The CFCC is implemented by using the active wait technique. This technique means that the CFCC is always running (processing or searching for service) and never enters a wait state. Therefore, the CF Control Code uses all the processor capacity that are available for the CF LPAR.

If the LPAR that is running the CFCC includes only dedicated processors (CPs or ICFs), the use of all processor capacity (cycles) is not an issue. However, this configuration can be an issue if the LPAR that is running the CFCC includes shared processors. We suggest that you enable thin interrupts on the CF LPAR (Default for z15).

Performance consideration: Dedicated processor CF still provides the best CF image performance for production environments.

CF structure sizing changes are expected when moving to CFCC Level 24. Always review the CF structure size by using the [CFSizer tool](#) when changing CFCC levels.

For more information about the recommended CFCC levels, see the [current exception letter that is published on IBM Resource Link®](#).

3.9.2 Coupling Thin Interrupts

CFCC Level 19 introduced Coupling Thin Interrupts to improve performance in environments that share CF engines. Although dedicated engines are preferable to obtain the best CF performance, Coupling Thin Interrupts can help facilitate the use of a shared pool of engines, which helps to lower hardware acquisition costs.

The interrupt causes a shared logical processor CF partition to be dispatched by PR/SM (if it is not already dispatched), which allows the request or signal to be processed in a more timely manner. The CF relinquishes control when work is exhausted or when PR/SM takes the physical processor away from the logical processor.

The use of Coupling Thin Interrupts is controlled by the DYNDISP specification.

You can experience CF response time improvements or more consistent CF response time when using CFs with shared engines. This improvement can allow more environments with multiple CF images to coexist in a server, and share CF engines with reasonable performance.

The response time for asynchronous CF requests can also be improved as a result of the use of Coupling Thin Interrupts on the z/OS host system, regardless of whether the CF is using shared or dedicated engines.

3.9.3 Dynamic CF dispatching

Dynamic CF dispatching provides the following process on a CF:

1. If no work exists, CF enters a wait state (by time).
2. After an elapsed time, CF wakes up to see whether any new work is available (that is, if any requests are in the CF Receiver buffer).
3. If no work exists, CF sleeps again for a longer period.
4. If new work is available, CF enters the normal active wait until no other work is available. After all work is complete, the process starts again.

With the introduction of the Coupling Thin Interrupt support, which is used only when the CF partition is using shared engines and the new **DYNDISP=THIN** parameter, the CFCC code is changed to handle these interrupts correctly. CFCC was also changed to relinquish voluntarily control of the processor whenever it runs out of work to do. It relies on Coupling Thin Interrupts to dispatch the image again in a timely fashion when new work (or new signals) arrives at the CF to be processed. With z15, **DYNDISP=THIN** is defined as the default mode of operation for coupling facility images that use shared processors.

This capability allows ICF engines to be shared by several CF images. In this environment, it provides faster and far more consistent CF service times. It can also provide performance that is reasonably close to dedicated-engine CF performance if the CF engines are not CF Control Code Thin Interrupts.

The introduction of Thin Interrupts allows a CF to run by using a shared processor while maintaining good performance. The shared engine is allowed to be undispached when no more work exists, as in the past. The new Thin Interrupt now gets the shared processor that is dispatched when a command or duplexing signal is presented to the shared engine.

This function saves processor cycles and is an excellent option to be used by a production backup CF or a testing environment CF. This function is activated by using the CFCC command **DYNDISP ON**.

Note: CFCC Change Shared-Engine CF Default to **DYNDISP=THIN**

The CPs can run z/OS operating system images and CF images. For software charging reasons, generally use only ICF processors to run CF images.

Dynamic CF dispatching is shown in Figure 3-18.

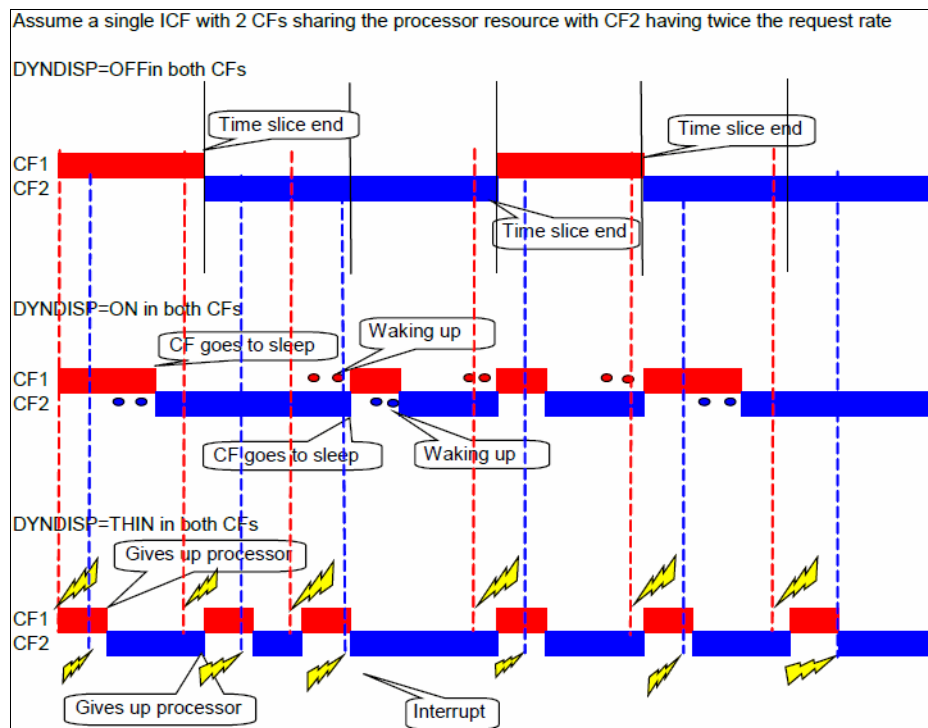


Figure 3-18 Dynamic CF dispatching options

For more information about CF configurations, see *Coupling Facility Configuration Options*, GF22-5042, and WP102400¹⁷ Coupling Interrupt and Coupling Facility Performance.

3.10 Virtual Flash Memory

Flash Express is not supported on z15. This feature was replaced by IBM Z Virtual Flash Memory (VFM), from z14. The Virtual Flash Memory feature code is 0643 on z15 T01.

3.10.1 IBM Z Virtual Flash Memory overview

Virtual Flash Memory (VFM) is an IBM solution to replace the external zFlash Express feature with support that is based on main memory.

The “storage class memory” that is provided by Flash Express adapters is replaced with memory allocated from main memory (VFM).

VFM is designed to help improve availability and handling of paging workload spikes when running z/OS. With this support, z/OS is designed to help improve system availability and responsiveness by using VFM across transitional workload events, such as market openings and diagnostic data collection. z/OS is also designed to help improve processor performance by supporting middleware use of pageable large (1 MB) pages.

VFM can also be used in CF images to provide extended capacity and availability for workloads that use IBM WebSphere MQ Shared Queues structures. The use of VFM can help availability by reducing latency from paging delays that can occur at the start of the workday or during other transitional periods. It is also designed to eliminate delays that can occur when collecting diagnostic data during failures.

3.10.2 VFM feature

A VFM feature (FC 0643) is 512 GB of memory on z15 T01. The maximum number of VFM features is 12 per z15 T01 system.

Ordered VFM memory reduces the maximum orderable memory for the model.

Simplification in its management is of great value because no hardware adapter is needed to manage. It also has no hardware repair and verify. It has a better performance because no I/O to attached adapter occurs. Finally, because this feature is part of memory, it is protected by RAIM and ECC.

VFM provides physical memory DIMMs that are needed to support activation of all customer purchased memory and HSA on a multiple drawer z15 T01 with one drawer done for the following features:

- ▶ Scheduled concurrent drawer upgrade, such as memory add
- ▶ Scheduled concurrent drawer maintenance, such N+1 repair
- ▶ Concurrent repair of an out of service CPC drawer “fenced” during Activation (POR)

Note: All of these features can be done without VFM. However, all customer purchased memory is not available for use in most cases. Some work might need to be shut down or not restarted.

¹⁷ For more information about WP102400, see this web page:
<http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102400>

3.10.3 VFM administration

The allocation and definition information of VFM for all partitions is viewed through the Storage Information panel that is under the Operational Customization panel.

The information is relocated during CDR in a manner that is identical to the process that was used for expanded storage. VFM is much simpler to manage (HMC task) and no hardware repair and verify (no cables and no adapters) are needed. Also, because this feature is part of internal memory, VFM is protected by RAIM and ECC and can provide better performance because no I/O to an attached adapter occurs.

Note: Use cases for Flash did not change (for example, z/OS paging and CF shared queue overflow). Instead, they transparently benefit from the changes in the hardware implementation.

No option is available for VFM plan ahead. The only option is to always include zVFM plan ahead when Flexible Memory option is selected.

3.11 Secure Service Container

Client applications are subject to a number of security risks in a production environment. These risks might include external risks (cyber hacker attacks) or internal risks (malicious software, system administrators using their privileged rights for unauthorized access and many others).

Secure Service Container (SSC) is an integrated IBM Z appliance and is designed to host most sensitive client workloads and applications, acting as a highly protected and secured digital vault, enforcing security by encrypting the whole stack: memory, network and data (both in-flight and at-rest). Applications running inside SSC are isolated and protected from outsider and insider threats.

SSC combines hardware, software, and middleware and is unique to IBM Z platform. Though it is called a container, it should not be confused with purely software open source containers (such as Kubernetes or Docker).

SSC is a part of the Pervasive Encryption concept that was introduced with IBM z14, which is aimed at delivering best IBM Security hardware and software enhancements, services, and practices for 360 degree infrastructure protection.

LPAR is defined as SSC by using the Hardware Management Console (HMC).

The SSC solution includes the following key advantages:

- ▶ Existing applications require zero changes to use SSC; software developers do not need to write any SSC specific programming code.
- ▶ End-to-end encryption, both of in-flight and at-rest data:
 - Automatic Network Encryption (TLS, IPSEC): Data-in-flight.
 - Automatic File System Encryption (LUKS): Data-at-rest.
 - Linux Unified Key Setup (LUKS) is the standard way in Linux to provide disk encryption. SSC encrypts all data with a key that is stored within the appliance.
 - Protected memory: Up to 16 TB can be defined per SSC LPAR.
- ▶ Encrypted Diagnostic Data

All diagnostic information (debug dump data, logs) are encrypted and do not contain any user or application data.

▶ No operating system access:

After the SSC appliance is built, Secure Shell (SSH) and command line-interface (CLI) are disabled, which ensures that even system administrators do not have access to the contents of SSC and do not know what application is running there.

▶ Applications that run inside SSC are being accessed externally by REST APIs only, in a transparent to user way.

▶ Tamper-proof SSC Secure Boot:

– SSC- eligible applications are being booted into SSC by using verified booting sequence, where only trusted and digitally signed and verified by IBM software code is uploaded into the SSC.

– Vertical workload isolation, certified by EAL5+ Common Criteria Standard, which is the highest level that ensures workload separation and isolation.

– Horizontal workload isolation: Separation from the rest of the host environment.

SSC is a powerful IBM technology for providing the extra protection of the most sensitive workloads.

IBM Hyper Protect Crypto Solutions family uses the SSC technology as a core layer to provide hyper protected services. For more information, see the [IBM Cloud Hyper Protect Crypto Services website](#).

IBM Blockchain Platform can be deployed on IBM Z by using SSC to host the IBM Blockchain network. For more information, see [this web page](#).



Central processor complex I/O structure

This chapter describes the I/O system structure and connectivity options that are available on the IBM z15 servers.

Naming: The IBM z15 server generation is available as the following machine types and models:

- ▶ Machine Type 8561 (M/T 8561), Model T01, Features Max34, Max71, Max108, Max145, and Max190, which is further identified as *IBM z15 Model T01*, or *z15 T01*, unless otherwise specified.
- ▶ Machine Type 8562 (M/T 8562), Model T02, Features Max4, Max13, Max21, Max32, and Max65, which is further identified as *IBM z15 Model T02*, or *z15 T02*, unless otherwise specified.

In the remainder of this chapter, IBM z15 (z15) refers to both machine types.

This chapter includes the following topics:

- ▶ 4.1, “Introduction to I/O infrastructure” on page 154
- ▶ 4.2, “I/O system overview” on page 156
- ▶ 4.3, “PCIe+ I/O drawer” on page 158
- ▶ 4.4, “CPC drawer fanouts” on page 161
- ▶ 4.5, “I/O features” on page 164
- ▶ 4.6, “Connectivity” on page 167
- ▶ 4.7, “Cryptographic functions” on page 200
- ▶ 4.8, “Integrated Firmware Processor” on page 202

4.1 Introduction to I/O infrastructure

This section describes the I/O features that are available on the IBM z15 system. The z15 systems support PCIe+ I/O drawers only.

I/O cage, I/O drawer, and PCIe I/O drawer are not supported.

Note: Throughout this chapter, the terms *adapter* and *card* refer to a PCIe I/O feature that is installed in a PCIe+ I/O drawer.

4.1.1 I/O infrastructure

IBM extends the use of industry standards on the IBM Z platform by offering a Peripheral Component Interconnect Express Generation 3 (PCIe Gen3) I/O infrastructure. The PCIe I/O infrastructure that is provided by the central processor complex (CPC) improves I/O capability and flexibility, while allowing for the future integration of PCIe adapters and accelerators.

The PCIe I/O infrastructure in z15 T01 consists of the following components:

- ▶ PCIe+ Gen3 dual port fanouts that support 16 GBps I/O bus for CPC drawer connectivity to the PCIe+ I/O drawers. It connects to the PCIe Interconnect Gen3 in the PCIe+ I/O drawers.
- ▶ PCIe Gen3 feature that support coupling links, Integrated Coupling Adapter Short Reach (ICA SR and ICA SR1.1). The ICA SR and ICA SR1.1 features have two ports, each port supporting 8 GBps.
- ▶ The 8U, 16-slot, and 2-domain PCIe+ I/O drawer for PCIe I/O features.

The z15 T01 I/O infrastructure provides the following benefits:

- ▶ The bus connecting the CPC drawer to the I/O domain in the PCIe+ I/O drawer bandwidth is 16 GBps.
- ▶ Up to 32 channels (16 PCIe I/O cards) are supported in the PCIe+ I/O drawer.
- ▶ Granularity for the storage area network (SAN) and the local area network (LAN):
 - The FICON Express16SA features two channels per feature for Fibre Channel connection (FICON), High-Performance FICON on Z (zHPF), and Fibre Channel Protocol (FCP) storage area networks. The FICON Express16S+ features two channels per feature for Fibre Channel connection (FICON), High-Performance FICON on Z (zHPF), and Fibre Channel Protocol (FCP) storage area networks. The FICON Express16S and FICON Express8S feature two channels per feature for Fibre Channel connection (FICON), High-Performance FICON on Z (zHPF), and Fibre Channel Protocol (FCP) storage area networks.
 - The Open Systems Adapter (OSA)-Express7S GbE, OSA-Express7S 1000BASE-T, OSA-Express6S GbE, and the OSA-Express6S 1000BASE-T features include two ports each (LAN connectivity); the OSA-Express7S 25GbE SR1.1, OSA-Express7S 25GbE, OSA-Express7S 10GbE, and the OSA-Express6S 10 GbE features have one port each (LAN connectivity).
- ▶ Native PCIe features (plugged into the PCIe+ I/O drawer):
 - IBM zHyperLink Express 1.1
 - IBM zHyperlink Express

- 25GbE and 10GbE Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) Express2.1
- 25GbE and 10GbE RoCE Express2
- Coupling Express Long Reach (CE LR) (available also on z14 M0x, z14 ZR1, z13, and z13s)
- 10 GbE RoCE Express
- Crypto Express7S (single/dual ports)
- Crypto Express6s
- Crypto Express5s

4.1.2 PCIe Generation 3

The PCIe Generation 3 uses 128b/130b encoding for data transmission. This configuration reduces the encoding overhead to about 1.54% versus the PCIe Generation 2 overhead of 20% that uses 8b/10b encoding.

The PCIe standard uses a low-voltage differential serial bus. Two wires are used for signal transmission, and a total of four wires (two for transmit and two for receive) form a lane of a PCIe link, which is full-duplex. Multiple lanes can be aggregated into a larger link width. PCIe supports link widths of 1, 2, 4, 8, 12, 16, and 32 lanes (x1, x2, x4, x8, x12, x16, and x32).

The data transmission rate of a PCIe link is determined by the link width (numbers of lanes), the signaling rate of each lane, and the signal encoding rule. The signaling rate of one PCIe Generation 3 lane is eight gigatransfers per second (GTps), which means that nearly 8 gigabits are transmitted per second (Gbps).

Note: I/O infrastructure for z15 is implemented as PCIe Generation 3. The PU chip PCIe interface is PCIe Generation 4 (x16 @32 GBps), but the CPC I/O Fanout infrastructure uses externally PCIe Generation 3 connectivity.

A PCIe Gen3 x16 link features the following data transmission rates:

- ▶ The maximum theoretical data transmission rate per lane:
8 Gbps * 128/130 bit (encoding) = 7.87 Gbps=984.6 MBps
- ▶ The maximum theoretical data transmission rate per link:
984.6 MBps * 16 (lanes) = 15.75 GBps

Considering that the PCIe link is full-duplex mode, the data throughput rate of a PCIe Gen3 x16 link is 31.5 GBps (15.75 GBps in both directions).

Link performance: The link speeds do not represent the performance of the link. The performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.

PCIe Gen3 x16 links are used in z15 servers for driving the PCIe+ I/O drawers, and for coupling links for CPC to CPC communications.

Note: Unless specified otherwise, *PCIe* refers to PCIe Generation 3 in remaining sections of this chapter.

4.2 I/O system overview

The z15 I/O characteristics and supported features are described in this section.

4.2.1 Characteristics

The z15 T01 I/O subsystem is designed to provide great flexibility, high availability, and the following excellent performance characteristics:

- ▶ High bandwidth

Link performance: The link speeds do not represent the performance of the link. The performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.

z15 servers use PCIe Gen3 protocol to drive PCIe+ I/O drawers and CPC to CPC (coupling) connections. The I/O bus infrastructure data rate of up to 128 GBps per system (12 PCIe+ Gen3 fanout slots). For more information about coupling link connectivity, see 4.6.4, “Parallel Sysplex connectivity” on page 195.

- ▶ Connectivity options:
 - z15 servers can be connected to an extensive range of interfaces, such as FICON/FCP for SAN connectivity, OSA features for LAN connectivity and zHyperLink Express for storage connectivity (low latency compared to FICON).
 - For CPC to CPC connections, z15 servers use Integrated Coupling Adapter (ICA SR) and the Coupling Express Long Reach (CE LR). The Parallel Sysplex InfiniBand is not supported.
 - The 25GbE RoCE Express2.1, 10GbE RoCE Express2.1, 25GbE RoCE Express2, 10GbE RoCE Express2, and 10GbE RoCE Express features provide high-speed memory-to-memory data exchange to a remote CPC by using the Shared Memory Communications over RDMA (SMC-R) protocol for TCP (socket-based) communications.
- ▶ Concurrent I/O upgrade

You can concurrently add I/O features to z15 servers if unused I/O slot positions are available.
- ▶ Concurrent PCIe+ I/O drawer upgrade

Extra PCIe+ I/O drawers can be installed concurrently if free frame slots for the PCIe+ I/O drawers and PCIe fanouts in the CPC drawer are available.
- ▶ Dynamic I/O configuration

Dynamic I/O configuration supports the dynamic addition, removal, or modification of the channel path, control units, and I/O devices without a planned outage.
- ▶ Pluggable optics:
 - The FICON Express16SA, FICON Express16S+ FICON Express16S and FICON Express8S, OSA Express7S, OSA Express6S, OSA Express5S, RoCE Express2.1, RoCE Express2, and RoCE Express features include Small Form-Factor Pluggable (SFP) optics.¹ These optics allow each channel to be individually serviced in a fiber optic module failure. The traffic on the other channels on the same feature can continue to flow if a channel requires servicing.

¹ OSA-Express5S and 6S 1000BASE-T features do not have optics (copper only, RJ45 connectors).

- The zHyperLink Express feature uses fiber optics cable with MTP² connector and the cable uses a CXP connection to the adapter. The CXP³ optics are provided with the adapter.
- ▶ Concurrent I/O card maintenance
 - Every I/O card that is plugged in PCIe+ I/O drawer supports concurrent card replacement during a repair action.

4.2.2 Supported I/O features

The following I/O features are supported on a z15 T01 system (max. for each individual adapter type):

- ▶ Up to 384 FICON Express16SA channels
- ▶ Up to 384 FICON Express16S+ channels
- ▶ Up to 384 FICON Express16S channels
- ▶ Up to 384 FICON Express8S channels
- ▶ Up to 48 OSA-Express7S 25GbE SR1.1 ports
- ▶ Up to 48 OSA-Express7S 25GbE SR ports
- ▶ Up to 48 OSA-Express7S 10GbE ports
- ▶ Up to 96 OSA-Express7S GbE ports
- ▶ Up to 96 OSA-Express7S 1000BASE-T ports
- ▶ Up to 48 OSA-Express6S 10GbE ports
- ▶ Up to 96 OSA-Express6S GbE ports
- ▶ Up to 96 OSA-Express6S 1000BASE-T ports
- ▶ Up to 48 OSA-Express5S 10GbE ports
- ▶ Up to 96 OSA-Express5S GbE ports
- ▶ Up to 96 OSA-Express5S 1000BASE-T ports
- ▶ Up to eight 25GbE RoCE Express2.1 features
- ▶ Up to eight 25GbE RoCE Express2 features
- ▶ Up to eight 10GbE RoCE Express2.1 features
- ▶ Up to eight 10GbE RoCE Express2 features
- ▶ Up to eight 10GbE RoCE Express features
- ▶ Up to 16 zHyperLink Express1.1 features
- ▶ Up to 16 zHyperLink Express features
- ▶ Up to 48 ICA SR1.1 features with up to 120 coupling links
- ▶ Up to 48 ICA SR features with up to 120 coupling links
- ▶ Up to 32 CE LR features with up to 32 coupling links

² Multifiber Termination Push-On.

³ For more information, see this web page: <https://cw.infinibandta.org/document/dl/7157>

Notes: Consider the following points:

- ▶ The number of I/O features that are supported might be affected by the machine power infrastructure (PDU versus BPA). z15 T01 PDU models support a maximum of 12 PCIe+ I/O drawers, BPA models support a maximum of 11 PCIe+ I/O drawers.
- ▶ The maximum number of coupling CHPIDs on a z15 server was increased to 384, which is a combination of the following ports (not all combinations are possible; subject to I/O configuration options):
 - Up to 96 ICA SR ports (48 ICA SR features)
 - Up to 64 CE LR ports (32 CE LR features)
- ▶ IBM Virtual Flash Memory replaces IBM zFlash Express feature on z15 servers.
- ▶ zEDC features are not supported.
- ▶ The maximum combined number of RoCE features that can be installed is 8; that is, any combination of 25GbE RoCE Express2.1, 25GbE RoCE Express2, 10GbE RoCE Express2.1, 10GbE RoCE Express2, and 10GbE RoCE Express (carry forward only) features.
- ▶ 25GbE RoCE Express2 (or 2.1) features should not be configured in the same SMC-R link group with 10GbE RoCE Express2 (or 2.1) or RoCE Express features. However, 10GbE RoCE Express2 (or 2.1) can be mixed with 10 GbE RoCE Express.

4.3 PCIe+ I/O drawer

The PCIe+ I/O drawers (see Figure 4-1) are attached to the CPC drawer through a PCIe cable and use PCIe Gen3 as the infrastructure bus within the drawer. The PCIe Gen3 I/O bus infrastructure data rate is up to 16 GBps.

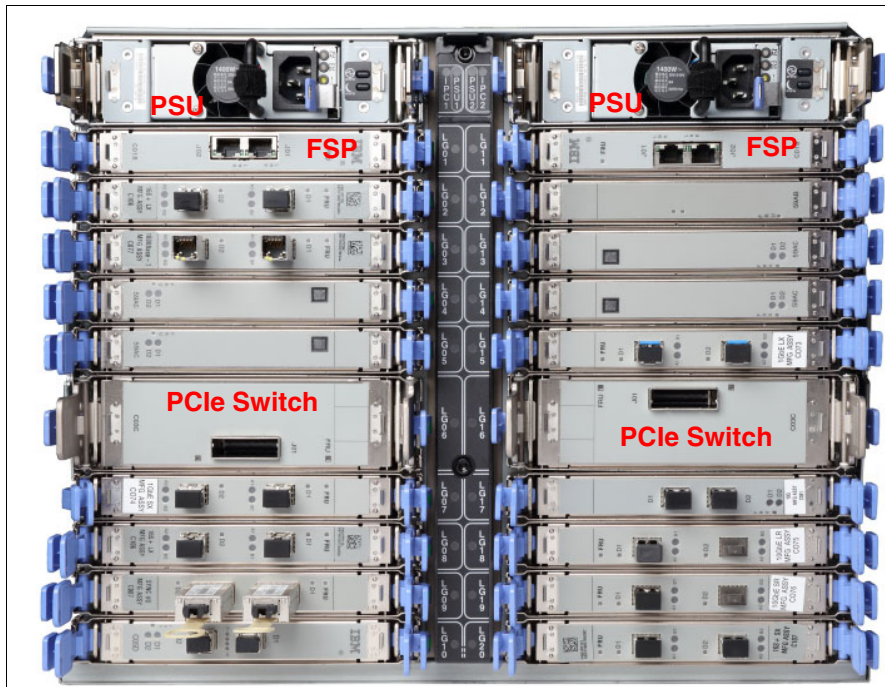


Figure 4-1 Rear view of PCIe+ I/O drawer

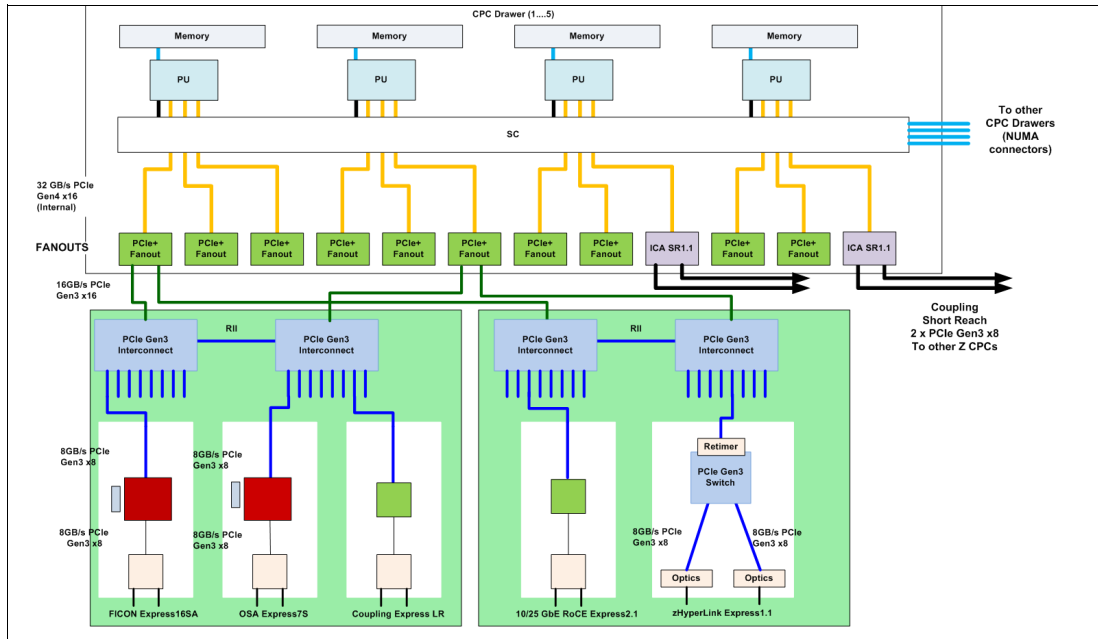


Figure 4-3 z15 I/O connectivity (Max34 feature with two PCIe+ I/O drawers represented)

The PCIe slots in a PCIe+ I/O drawer are organized into two I/O domains. Each I/O domain supports up to eight features and is driven through a PCIe switch card. Two PCIe switch cards always provide a backup path for each other through the passive connection in the PCIe+ I/O drawer backplane. During a PCIe fanout card or cable failure, 16 I/O cards in two domains can be driven through a single PCIe switch card. It is not possible to drive 16 I/O cards after one of the PCIe switch cards is removed.

The two switch cards are interconnected through the PCIe+ I/O drawer board (Redundant I/O Interconnect, or R11). In addition, switch cards in same PCIe+ I/O drawer are connected to PCIe fanouts across clusters in CPC drawer for higher availability.

The R11 design provides a failover capability during a PCIe fanout card failure. Both domains in one of these PCIe+ I/O drawers are activated with two fanouts. The flexible service processors (FSPs) are used for system control.

The domains and their related I/O slots are shown in Figure 4-2 on page 159.

Each I/O domain supports up to eight features (FICON, OSA, Crypto, and so on.) All I/O cards connect to the PCIe switch card through the backplane board. The I/O domains and slots are listed in Table 4-1.

Table 4-1 I/O domains of PCIe+ I/O drawer

| Domain | I/O slot in domain |
|--------|--|
| 0 | LG02, LG03, LG04, LG05, LG07, LG08, LG09, and LG10 |
| 1 | LG12, LG13, LG14, LG15, LG17, LG18, LG19, and LG20 |

4.3.1 PCIe+ I/O drawer offerings

A maximum of 12 PCIe+ I/O drawers, depending on system power choice (PDU or BPA), can be installed for supporting up to 192 PCIe I/O features.

For an upgrade to z15 T01 servers, only the following PCIe features can be carried forward:

- ▶ FICON Express16S+
- ▶ FICON Express16S
- ▶ FICON Express8S
- ▶ zHyperLink Express
- ▶ OSA-Express7S 25GbE SR
- ▶ OSA-Express6S (all features)
- ▶ OSA-Express5S (all features)
- ▶ 25GbE RoCE Express2
- ▶ 10GbE RoCE Express2
- ▶ 10GbE RoCE Express
- ▶ Crypto Express6S
- ▶ Crypto Express5S
- ▶ Coupling Express Long Reach (CE LR)

Consideration: On a z15 server, only PCIe+ I/O drawers are supported. No older generation drawers can be carried forward.

z15 T01 server supports the following PCIe I/O new features that are hosted in the PCIe+ I/O drawers:

- ▶ FICON Express16SA
- ▶ OSA-Express7S 25GbE SR1.1
- ▶ OSA-Express7S 10GbE
- ▶ OSA-Express7S GbE
- ▶ OSA-Express7S 1000BASE-T
- ▶ 25GbE RoCE Express2.1
- ▶ 10GbE RoCE Express2.1
- ▶ Crypto Express7S (one or two ports)
- ▶ Coupling Express Long Reach (CE LR)
- ▶ zHyperLink Express1.1

4.4 CPC drawer fanouts

The z15 server uses fanout cards to connect the I/O subsystem in the CPC drawer to the PCIe+ I/O drawers. The fanout cards also provide the ICA SR (ICA SR and ICA SR1.1) coupling links for Parallel Sysplex. All fanout cards support concurrent add, delete, and move.

The z15 CPC drawer I/O infrastructure consists of the following features:

- ▶ The PCIe+ Generation 3 fanout cards: Two ports per card (feature) that connect to PCIe+ I/O drawers.
- ▶ ICA SR (ICA SR and ICA SR1.1) fanout cards: Two ports per card (feature) that connect to other (external) CPCs.

Note: IBM z15 does not support Parallel Sysplex InfiniBand (PSIFB) links.

Unless otherwise noted, ICA SR is used for ICA SR and ICA SR1.1 for the rest of the chapter.

The PCIe fanouts cards are installed in the rear of the CPC drawers. Each CPC drawer features 12 PCIe+ Gen3 fanout slots.

The PCIe fanouts and ICA SR fanouts are installed in locations LG01 - LG12 at the rear in the CPC drawers (see Figure 2-25 on page 70). The oscillator card (OSC) is combined with the Flexible Support Processor (FSP) and is in the front of the drawer. Two combined FSP/OSC cards are used per CPC drawer.

The following types of fanout cards are supported by z15 servers. Each CPC drawer fanout slot can hold one of the following fanouts:

- ▶ PCIe+ Gen3 fanout card: This dual-port copper fanout provides connectivity to the PCIe switch card in the PCIe I/O drawer.
- ▶ Integrated Coupling Adapter (ICA SR): This two-port adapter provides coupling connectivity between z14 ZR1, z14, z13, and z13s servers, up to 150 meters (492 feet), @8 GBps link rate.

An I/O connection diagram is shown in Figure 4-3 on page 160.

4.4.1 PCIe+ Generation 3 fanout (FC 0175)

The PCIe+ Gen3 fanout card provides connectivity to a PCIe+ I/O drawer by using a copper cable. One port on the fanout card is dedicated for PCIe I/O. This PCIe fanout card supports a link rate of 16 GBps (with two links per card).

A 16x PCIe copper cable of 1.5 meters (4.92 feet) to 4.0 meters (13.1 feet) is used for connection to the PCIe switch card in the PCIe+ I/O drawer. PCIe fanout cards are always plugged in pairs and provide redundancy for I/O domains within the PCIe+ I/O drawer.

PCIe fanout: The PCIe fanout is used exclusively for I/O and cannot be shared for any other purpose.

4.4.2 Integrated Coupling Adapter (FC 0172 and 0176)

Introduced with IBM z13, the IBM ICA SR (FC 0172) is a two-port fanout feature that is used for short distance coupling connectivity and uses channel type CS5. For z15, the new build feature is ICA SR1.1 (FC 0176).

The ICA SR (FC 0172) and ICA SR1.1 (FC 0176) use PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes for coupling. No performance degradation is expected compared to the coupling over InfiniBand 12x IFB3 protocol.

Both cards are designed to drive distances up to 150 meters (492 feet) with a link data rate of 8 GBps. ICA SR supports up to four channel-path identifiers (CHPIDs) per port and eight subchannels (devices) per CHPID.

The coupling links can be defined as shared between images (z/OS) within a CSS. They also can be spanned across multiple CSSs in a CPC. For ICA SR features, a maximum four CHPIDs per port can be defined.

When STP (FC 1021) is available, ICA SR coupling links can be defined as timing-only links to other z15, z14 ZR1, z14, and z13/z13s CPCs.

These two fanouts features are housed in the PCIe+ Gen3 I/O fanout slot on the z15 CPC drawers. Up to 48 ICA SR and ICA SR1.1 features (up to 96 ports) are supported on a z15 T01 system. This configuration enables greater connectivity for short distance coupling on a single processor node compared to previous Z generations.

The ICA SR can be used for coupling connectivity between z15, z14/z14 ZR1, and z13/z13s servers. It cannot be connected to HCA3-O or HCA3-O LR coupling fanouts.

OM3 fiber optic can be used for distances up to 100 meters (328 feet). OM4 fiber optic cables can be used for distances up to 150 meters (492 feet). For more information, see the following manuals:

- ▶ *Planning for Fiber Optic Links, GA23-1408*
- ▶ *IBM Z 8561 Installation Manual for Physical Planning, GC28-7002*

4.4.3 Fanout considerations

Fanout slots in the CPC drawer can be used to plug different fanouts. On z15 T01, the CPC drawers can hold up to 60 PCIe fanout cards (five-CPC drawers configuration).

Adapter ID number assignment

PCIe fanouts and ports are identified by an Adapter ID (AID) that is initially dependent on their physical locations, which is unlike channels that are installed in a PCIe+ I/O drawer. Those channels are identified by a physical channel ID (PCHID) number that is related to their physical location. This AID must be used to assign a CHPID to the fanout in the IOCDs definition. The CHPID assignment is done by associating the CHPID to an AID port (see Table 4-2).

Table 4-2 Fanout locations and their AIDs for the CPC drawer (z15 T01)

| Fanout locations | CPC0 Location A10 AID(Hex) | CPC1 Location A15 AID(Hex) | CPC2 Location A20 AID(Hex) | CPC3 Location B10 AID(Hex) | CPC4 Location B15 AID(Hex) |
|------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| LG01 | 00 | 0C | 18 | 24 | 30 |
| LG02 | 01 | 0D | 19 | 25 | 31 |
| LG03 | 02 | 0E | 1A | 26 | 32 |
| LG04 | 03 | 0F | 1B | 27 | 33 |
| LG05 | 04 | 10 | 1C | 28 | 34 |
| LG06 | 05 | 11 | 1D | 29 | 35 |
| LG07 | 06 | 12 | 1E | 2A | 36 |
| LG08 | 07 | 13 | 1F | 2B | 37 |
| LG09 | 08 | 14 | 20 | 2C | 38 |
| LG10 | 09 | 15 | 21 | 2D | 39 |
| LG11 | 0A | 16 | 22 | 2E | 3A |
| LG12 | 0B | 17 | 23 | 2F | 3B |

Fanout slots

The fanout slots are numbered LG01 - LG12, from left to right, as listed in Table 4-2. All fanout locations and their AIDs for the CPC drawer are shown for reference only.

Important: The AID numbers that are listed in Table 4-2 are valid only for a new build system. If a fanout is moved, the AID follows the fanout to its new physical location.

The AID assignment is listed in the PCHID REPORT that is provided for each new server or for an MES upgrade on existing servers. Part of a PCHID REPORT for a z15 T01 is shown in Example 4-1. In this example, four fanout cards are installed at in CPC drawer at location A10A, in slots LG01, LG02, LG04, and LG11 with AIDs 00, 01, 03, and 0A.

Example 4-1 AID assignment in PCHID REPORT

| CHPIDSTART | | PCHID REPORT | | | Aug 25, 2019 |
|------------------------|------|--------------|------|--------------------|--------------|
| 23754978 | | | | | |
| Machine: 8561-T01 NEW1 | | | | | |
| ----- | | | | | |
| Source | Drwr | Slot | F/C | PCHID/Ports or AID | Comment |
| A10/LG01 | A10A | LG01 | 0176 | AID=00 | |
| A10/LG02 | A10A | LG02 | 0176 | AID=01 | |
| A10/LG04 | A10A | LG04 | 0176 | AID=03 | |
| A10/LG11 | A10A | LG11 | 0176 | AID=0A | |

Fanout features that are supported by the z15 server are listed in Table 4-3, which includes the feature type, feature code, and information about the link supported by the fanout feature.

Table 4-3 Fanout summary

| Fanout feature | Feature code | Use | Cable type | Connector type | Maximum distance | Link data rate ^a |
|-------------------|--------------|----------------------------|------------|----------------|------------------|-----------------------------|
| PCIe+ Gen3 fanout | 0173 | Connect to PCIe I/O drawer | Copper | N/A | 4 m (13.1 ft) | 16 GBps |
| ICA SR | 0172 | Coupling link | OM4 | MTP | 150 m (492 ft) | 8 GBps |
| | | | OM3 | MTP | 100 m (328 ft) | 8 GBps |
| ICA SR1.1 | 0176 | Coupling link | OM4 | MTP | 150 m (492 ft) | 8 GBps |
| | | | OM3 | MTP | 100 m (328 ft) | 8 GBps |

a. The link data rates do not represent the performance of the link. The performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.

4.5 I/O features

I/O features (adapters) include ports⁴ to connect the z15 T01 system to external devices, networks, or other servers. I/O features are plugged into the PCIe+ I/O drawers, based on the configuration rules for the server. Different types of I/O cards are available, one for each channel or link type. I/O cards can be installed or replaced concurrently.

⁴ Certain I/O features do not have external ports, such as Crypto Express.

4.5.1 I/O feature card ordering information

The I/O features that are supported by z15 T01 systems and the ordering information for them are listed in Table 4-4.

Table 4-4 I/O features and ordering information

| Channel feature | Feature code | New build | Carry-forward |
|---|--------------|-----------|---------------|
| FICON Express16SA LX | 0436 | Y | N/A |
| FICON Express16SA SX | 0437 | Y | N/A |
| FICON Express16S+ LX | 0427 | N | Y |
| FICON Express16S+ SX | 0428 | N | Y |
| FICON Express16S LX | 0418 | N | Y |
| FICON Express16S SX | 0419 | N | Y |
| FICON Express8S LX | 0409 | N | Y |
| FICON Express8S SX | 0410 | N | Y |
| OSA-Express7S 25GbE SR1.1 | 0449 | Y | N/A |
| OSA-Express7S 25GbE SR | 0429 | N | Y |
| OSA-Express7S 10GbE LR | 0444 | Y | N/A |
| OSA-Express7S 10GbE SR | 0445 | Y | N/A |
| OSA-Express7S GbE LX | 0442 | Y | N/A |
| OSA-Express7S GbE SX | 0443 | Y | N/A |
| OSA-Express7S 1000BASE-T Ethernet | 0446 | | N/A |
| OSA-Express6S 10GbE LR | 0424 | N | Y |
| OSA-Express6S 10GbE SR | 0425 | N | Y |
| OSA-Express6S GbE LX | 0422 | N | Y |
| OSA-Express6S GbE SX | 0423 | N | Y |
| OSA-Express6S 1000BASE-T Ethernet | 0426 | N | Y |
| OSA-Express5S 10GbE LR | 0415 | N | Y |
| OSA-Express5S 10GbE SR | 0416 | N | Y |
| OSA-Express5S GbE LX | 0413 | N | Y |
| OSA-Express5S GbE SX | 0414 | N | Y |
| OSA-Express5S 1000BASE-T Ethernet | 0417 | N | Y |
| PCIe+ Gen3 fanout | 0175 | Y | N/A |
| Integrated Coupling Adapter (ICA SR1.1) | 0176 | Y | N/A |
| Integrated Coupling Adapter (ICA SR) | 0172 | Y | Y |
| Coupling Express LR | 0433 | Y | Y |
| Crypto Express7S (2 ports) | 0898 | Y | N/A |

| Channel feature | Feature code | New build | Carry-forward |
|---------------------------|--------------|-----------|---------------|
| Crypto Express7S (1 port) | 0899 | Y | N/A |
| Crypto Express6S | 0893 | N | Y |
| Crypto Express5S | 0890 | N | Y |
| 25GbE RoCE Express2.1 | 0450 | Y | N/A |
| 25GbE RoCE Express2 | 0430 | Y | Y |
| 10GbE RoCE Express2.1 | 0432 | Y | N/A |
| 10GbE RoCE Express2 | 0412 | N | Y |
| 10GbE RoCE Express | 0411 | N | Y |
| zHyperLink Express1.1 | 0451 | Y | N/A |
| zHyperLink Express | 0431 | N | Y |

Coupling links connectivity support: zEC12 and zBC12 are *not* supported in same Parallel Sysplex or STP CTN with z15.

4.5.2 Physical channel ID report

A physical channel ID (PCHID) reflects the physical location of a channel-type interface. A PCHID number is based on the following factors:

- ▶ PCIe+ I/O drawer location
- ▶ Channel feature slot number
- ▶ Port number of the channel feature

A CHPID does not directly correspond to a hardware channel port. Instead, it is assigned to a PCHID in the hardware configuration definition (HCD) or IOCP.

A PCHID REPORT is created for each new build server and for upgrades on servers. The report lists all I/O features that are installed, the physical slot location, and the assigned PCHID. A portion of a sample PCHID REPORT is shown in Example 4-2.

Example 4-2 PCHID REPORT

```

CHPIDSTART
23754978                                PCHID REPORT                                Aug 25,2019
Machine: 8561-T01  NEW1
-----
Source          Drwr  Slot  F/C   PCHID/Ports or AID          Comment
A10/LG01        A10A  LG01  0176  AID=00
A10/LG02        A10A  LG02  0176  AID=01
A10/LG04        A10A  LG04  0176  AID=03
A10/LG11        A10A  LG11  0176  AID=0A
A10/LG12/J02    Z01B  02    0437  100/D1 101/D2
A10/LG12/J02    Z01B  03    0437  104/D1 105/D2
A10/LG12/J02    Z01B  04    0443  108/D1D2
.....<< snippet >>.....

```

The PCHID REPORT that is shown in Example 4-2 on page 166 includes the following components (among others):

- ▶ Feature code 0176 (Integrated Coupling Adapters (ICA SR1.1) is installed in the CPC drawer (location A10A, slots LG01, LG02, LG04. and LG11), and have AIDs 00, 01, 03, and 0A assigned.
- ▶ Feature codes 0437 (FICON Express16SA SX) are installed in PCIe+ I/O drawer 1:
 - Location Z01B, slot 02 with PCHIDs 100/D1 and 101/D2 assigned
 - Location Z01B, slot 03 with PCHIDs 104/D1 and 105/D2 assigned
- ▶ Feature code 0443 (OSA-Express7S GbE SX) installed in PCIe+ I/O drawer 1 in location Z01B, slots 04 with PCHID108/D1D2

A resource group (RG) parameter is also shown in the PCHID REPORT for native PCIe features. A balanced plugging of native PCIe features exists between four resource groups (RG1, RG2, RG3, and RG4).

The preassigned PCHID number of each I/O port relates directly to its physical location (jack location in a specific slot).

4.6 Connectivity

I/O channels are part of the CSS. They provide connectivity for data exchange between servers, between servers and external control units (CUs) and devices, or between networks.

For more information about connectivity to external I/O subsystems (for example, disks), see “Storage connectivity” on page 172.

For more information about communication to LANs, see “Network connectivity” on page 181.

Communication between servers is implemented by using CE LR, ICA SR, or channel-to-channel (CTC) connections. For more information, see “Parallel Sysplex connectivity” on page 195.

4.6.1 I/O feature support and configuration rules

The supported I/O features for a z15 T01 system are listed in Table 4-5. Also listed in Table 4-5 are the number of ports per card, port increments, the maximum number of feature cards, and the maximum number of channels for each feature type. The CHPID definitions that are used in the IOCDs also are listed.

Table 4-5 z15 T01 supported I/O features

| I/O feature | Ports per card | Port increments | Max. ports | Max. I/O slots | PCHID | CHPID definition |
|-------------------------|----------------|-----------------|------------|----------------|-------|----------------------|
| Storage access | | | | | | |
| FICON Express16SA LX/SX | 2 | 2 | 384 | 192 | Yes | FC, FCP ^a |
| FICON Express16S+ LX/SX | 2 | 2 | 128 | 64 | Yes | FC, FCP ^a |
| FICON Express16S LX/SX | 2 | 2 | 128 | 64 | Yes | FC, FCP ^b |
| FICON Express8S LX/SX | 2 | 2 | 128 | 64 | Yes | FC, FCP ^b |

| I/O feature | Ports per card | Port increments | Max. ports | Max. I/O slots | PCHID | CHPID definition |
|---|----------------|-----------------|------------|----------------|-------|------------------|
| zHyperLink Express 1.1 | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| zHyperLink Express | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| OSA-Express features^c | | | | | | |
| OSA-Express7S 25GbE SR1.1 | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express7S 25GbE SR | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express7S 10GbE LR/SR | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express7S GbE LR/SR | 1 | 1 | 48 | 48 | Yes | OSD, OSC |
| OSA-Express7S 1000BASE-T ^d | 2 | 2 | 96 | 48 | Yes | OSC, OSD, OSE |
| OSA-Express6S 10 GbE LR/SR | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express6S GbE LX/SX | 2 | 2 | 96 | 48 | Yes | OSD |
| OSA-Express6S 1000BASE-T ^d | 2 | 2 | 96 | 48 | Yes | OSC, OSD, OSE |
| OSA-Express5S 10 GbE LR/SR | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express5S GbE LX/SX | 2 | 2 | 96 | 48 | Yes | OSD |
| OSA-Express5S 1000BASE-T ^d | 2 | 2 | 96 | 48 | Yes | OSC, OSD, OSE |
| RoCE Express features | | | | | | |
| 25GbE RoCE Express2.1 | 2 | 2 | 32 | 16 | Yes | N/A |
| 10GbE RoCE Express2.1 | 2 | 2 | 32 | 16 | Yes | N/A |
| 25GbE RoCE Express2 | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| 10GbE RoCE Express2 | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| 10GbE RoCE Express | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| Coupling features | | | | | | |
| Coupling Express LR | 2 | 2 | 32 | 16 | Yes | CL5 |
| Integrated Coupling Adapter (ICA SR1.1) | 2 | 2 | 96 | 48 | N/A | CS5 |
| Integrated Coupling Adapter (ICA SR) | 2 | 2 | 96 | 48 | N/A | CS5 |
| Cryptographic features^f | | | | | | |

| I/O feature | Ports per card | Port increments | Max. ports | Max. I/O slots | PCHID | CHPID definition |
|---|----------------|-----------------|------------|----------------|-------|------------------|
| zHyperLink Express 1.1 | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| zHyperLink Express | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| OSA-Express features^c | | | | | | |
| OSA-Express7S 25GbE SR1.1 | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express7S 25GbE SR | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express7S 10GbE LR/SR | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express7S GbE LR/SR | 1 | 1 | 48 | 48 | Yes | OSD, OSC |
| OSA-Express7S 1000BASE-T ^d | 2 | 2 | 96 | 48 | Yes | OSC, OSD, OSE |
| OSA-Express6S 10 GbE LR/SR | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express6S GbE LX/SX | 2 | 2 | 96 | 48 | Yes | OSD |
| OSA-Express6S 1000BASE-T ^d | 2 | 2 | 96 | 48 | Yes | OSC, OSD, OSE |
| OSA-Express5S 10 GbE LR/SR | 1 | 1 | 48 | 48 | Yes | OSD |
| OSA-Express5S GbE LX/SX | 2 | 2 | 96 | 48 | Yes | OSD |
| OSA-Express5S 1000BASE-T ^d | 2 | 2 | 96 | 48 | Yes | OSC, OSD, OSE |
| RoCE Express features | | | | | | |
| 25GbE RoCE Express2.1 | 2 | 2 | 32 | 16 | Yes | N/A |
| 10GbE RoCE Express2.1 | 2 | 2 | 32 | 16 | Yes | N/A |
| 25GbE RoCE Express2 | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| 10GbE RoCE Express2 | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| 10GbE RoCE Express | 2 | 2 | 32 | 16 | Yes | N/A ^e |
| Coupling features | | | | | | |
| Coupling Express LR | 2 | 2 | 32 | 16 | Yes | CL5 |
| Integrated Coupling Adapter (ICA SR1.1) | 2 | 2 | 96 | 48 | N/A | CS5 |
| Integrated Coupling Adapter (ICA SR) | 2 | 2 | 96 | 48 | N/A | CS5 |
| Cryptographic features^f | | | | | | |

| I/O feature | Ports per card | Port increments | Max. ports | Max. I/O slots | PCHID | CHPID definition |
|---|----------------|-----------------|-----------------|----------------|-------|------------------|
| Crypto Express7S (2 ports) ^g | 2 | 2 | 60 ^h | 30 | n/a | n/a |
| Crypto Express7S (1 port) ^g | 1 | 1 | 16 | 16 | n/a | n/a |
| Crypto Express6S | 1 | 1 | 16 | 16 | n/a | n/a |
| Crypto Express5S | 1 | 1 | 16 | 16 | n/a | n/a |

- a. Both ports must be defined with the same CHPID type.
- b. CHPID type mixture is allowed. The keyword is MIXTYPE.
- c. On z15, the OSX type CHPID cannot be defined. z15 cannot be part of an ensemble managed by zManager.
- d. On z15, the OSM type CHPID cannot be defined for user configurations in PR/SM mode. It is used in DPM mode for internal management only.
- e. These features are defined by using Virtual Functions IDs (FIDs).
- f. Crypto Express features are defined through the HMC.
- g. Crypto Express7S can be ordered with one or two IBM 4769 PCIeCC (cryptographic coprocessor) adapters. Each adapter supports 85 domains.
- h. One port equals one IBM PCIeCC. The PCIeCC is a Hardware Security Module (HSM). z15 T01 supports up to 16 single-port (HSM) Crypto Express features, or, if Crypto Express7S (2 port) is used, up to 60 HSMs in any combination.

At least one I/O feature (FICON) or one coupling link feature (ICA SR or CE LR) must be present in the minimum configuration.

The following features can be shared and spanned:

- ▶ FICON channels that are defined as FC or FCP
- ▶ OSA-Express features that are defined as OSC, OSD, OSE or OSM
- ▶ Coupling links that are defined as CS5 or CL5
- ▶ HiperSockets that are defined as IQD

The following features are plugged into a PCIe+ I/O drawer and do not require the definition of a CHPID and CHPID type:

- ▶ Each Crypto Express (7S/6S/5S) feature occupies one I/O slot, but does not include a PCHID type. However, LPARs in all CSSs can access the features. Each Crypto Express adapter can support up to 85 domains.
- ▶ Each 25GbE RoCE Express2.1 feature occupies one I/O slot but does not include a CHPID type. However, LPARs in all CSSs can access the feature. The 25GbE RoCE Express2.1 can be defined to up to 126 virtual functions (VFs) per feature (port is defined in z/OS Communications Server). The 25GbE RoCE Express2.1 features support up to 63 VFs per port (up to 126 VFs per feature).
- ▶ Each 10 RoCE Express2.1 feature occupies one I/O slot but does not include a CHPID type. The 10GbE RoCE Express2.1 can be defined to up to 126 virtual functions (VFs) per feature (port is defined in z/OS Communications Server). The 10GbE RoCE Express2.1 features support up to 63 VFs per port (up to 126 VFs per feature).
- ▶ Each RoCE Express/Express2 feature occupies one I/O slot but does not include a CHPID type. However, LPARs in all CSSs can access the feature. The 10GbE RoCE Express can be defined to up to 31 LPARs per feature (port is defined in z/OS Communications Server). The 25GbE RoCE Express2 and the 10GbE RoCE Express2 features support up to 63 LPARs per port (up to 126 LPARs per feature).
- ▶ Each zHyperLink Express/zHyperlink Express2.1 feature occupies one I/O slot but does not include a CHPID type. However, LPARs in all CSSs can access the feature. The

zHyperLink Express adapter works as native PCIe adapter and can be shared by multiple LPARs.

Each port supports up to 127 Virtual Functions (VFs), with one or more VFs/PFIDs being assigned to each LPAR. This support gives a maximum of 254 VFs per adapter.

I/O feature cables and connectors

The IBM Facilities Cabling Services fiber transport system offers a total cable solution service to help with cable ordering requirements. These services can include the requirements for all of the protocols and media types that are supported (for example, FICON, Coupling Links, and OSA). The services can help whether the focus is the data center, SAN, LAN, or the end-to-end enterprise.

Cables: All fiber optic cables, cable planning, labeling, and installation are client responsibilities for new z15 installations and upgrades. Fiber optic conversion kits and mode conditioning patch cables are not orderable as features on z15 servers. All other cables must be sourced separately.

The Enterprise Fiber Cabling Services use a proven modular cabling system, the fiber transport system (FTS), which includes trunk cables, zone cabinets, and panels for servers, directors, and storage devices. FTS supports Fiber Quick Connect (FQC). FQC feature code is 7960. The FC 7961 is made of FQC feature, bracket and mounting hardware. The FC 7924 is made of FQC, bracket and mounting hardware and LC Duplex 2 meter (6.6 feet) harness. They are optional in z15.

Whether you choose a packaged service or a custom service, high-quality components are used to facilitate moves, additions, and changes in the enterprise to prevent the need to extend the maintenance window.

The required connector and cable type for each I/O feature on z15 T01 servers are listed in Table 4-6.

Table 4-6 I/O feature connector and cable types

| Feature code | Feature name | Connector type | Cable type |
|--------------|----------------------------|----------------|---------------------------|
| 0436 | FICON Express16SA LX | LC Duplex | 9 µm SM |
| 0437 | FICON Express16SA SX | LC Duplex | 50, 62.5 µm MM |
| 0427 | FICON Express16S+ LX 10 km | LC Duplex | 9 µm SM |
| 0428 | FICON Express16S+ SX | LC Duplex | 50, 62.5 µm MM |
| 0418 | FICON Express16S LX 10 km | LC Duplex | 9 µm SM |
| 0419 | FICON Express16S SX | LC Duplex | 50, 62.5 µm MM |
| 0409 | FICON Express8S LX 10 km | LC Duplex | 9 µm SM |
| 0410 | FICON Express8S SX | LC Duplex | 50, 62.5 µm MM |
| 0449 | OSA-Express7S 25GbE SR1.1 | LC Duplex | 50 µm MM OM4 ^d |
| 0429 | OSA-Express7S 25GbE SR | LC Duplex | 50 µm MM OM4 ^d |
| 0444 | OSA-Express7S 10GbE LR | LC Duplex | 9 µm SM |
| 0445 | OSA-Express7S 10GbE SR | LC Duplex | 50, 62.5 µm MM |
| 0442 | OSA-Express7S GbE LX | LC Duplex | 9 µm SM |

| Feature code | Feature name | Connector type | Cable type |
|--------------|---|----------------|-----------------------------|
| 0443 | OSA-Express7S GbE SX | LC Duplex | 50, 62.5 µm MM |
| 0446 | OSA-Express7S 1000BASE-T | RJ-45 | Category 5 UTP ^a |
| 0424 | OSA-Express6S 10GbE LR | LC Duplex | 9 µm SM |
| 0425 | OSA-Express6S 10 GbE SR | LC Duplex | 50, 62.5 µm MM |
| 0422 | OSA-Express6S GbE LX | LC Duplex | 9 µm SM |
| 0423 | OSA-Express6S GbE SX | LC Duplex | 50, 62.5 µm MM |
| 0426 | OSA-Express6S 1000BASE-T | RJ-45 | Category 5 UTP ^b |
| 0415 | OSA-Express5S 10 GbE LR | LC Duplex | 9 µm SM |
| 0416 | OSA-Express5S 10 GbE SR | LC Duplex | 50, 62.5 µm MM |
| 0413 | OSA-Express5S GbE LX | LC Duplex | 9 µm SM |
| 0414 | OSA-Express5S GbE SX | LC Duplex | 50, 62.5 µm MM |
| 0417 | OSA-Express5S 1000BASE-T | RJ-45 | Category 5 UTP |
| 0450 | 25GbE RoCE Express2.1 | LC Duplex | 50 µm MM |
| 0405 | 10GbE RoCE Express2.1 | LC Duplex | 50, 62.5 µm MM |
| 0430 | 25GbE RoCE Express2 | LC Duplex | 50 µm MM OM4 ^d |
| 0412 | 10GbE RoCE Express2 | LC Duplex | 50, 62.5 µm MM |
| 0411 | 10GbE RoCE Express | LC Duplex | 50, 62.5 µm MM |
| 0433 | CE LR | LC Duplex | 9 µm SM |
| 0176 | Integrated Coupling Adapter SR1.1 (ICA SR1.1) | MTP | 50 µm MM OM4 ^c |
| 0172 | Integrated Coupling Adapter (ICA SR) | MTP | 50 µm MM OM4 ^d |
| 0451 | zHyperLink Express1.1 | MPO | 50 µm MM OM4 ^d |
| 0431 | zHyperLink Express | MPO | 50 µm MM OM4 ^d |

a. UTP is unshielded twisted pair. Consider the use of category 6 UTP for 1000 Mbps connections.

b. UTP is unshielded twisted pair. Consider the use of category 6 UTP for 1000 Mbps connections.

c. Or 50 µm MM OM3, but OM4/OM5 is highly recommended.

MM = Multi-Mode
SM = Single-Mode

d. Or 50 µm MM OM3, but OM4 is highly recommended.

MM = Multi-Mode
SM = Single-Mode

4.6.2 Storage connectivity

Connectivity to external storage I/O subsystems (for example, disks) is provided by FICON channels and zHyperLink⁵.

⁵ zHyperLink feature operates with a FICON channel.

FICON channels

z15 T01 supports the following FICON features:

- ▶ FICON Express16SA (FC 0436/0437)
- ▶ FICON Express16S+ (FC 0427/0428)
- ▶ FICON Express16S (FC 0418/0419)
- ▶ FICON Express8S (FC 0409/0410)

The FICON Express16SA, FICON Express16S+, FICON Express16S, and FICON Express8S features conform to the following architectures:

- ▶ Fibre Connection (FICON)
- ▶ High-Performance FICON on Z (zHPF)
- ▶ Fibre Channel Protocol (FCP)

The FICON features provide connectivity between any combination of servers, directors, switches, and devices (control units, disks, tapes, and printers) in a SAN.

Each FICON Express feature occupies one I/O slot in the PCIe+ I/O drawer. Each feature includes two ports, each supporting an LC Duplex connector, with one PCHID and one CHPID that is associated with each port.

Each FICON Express feature uses SFP (SFP+ for FICON Express16SA) optics that allow for concurrent repairing or replacement for each SFP. The data flow on the unaffected channels on the same feature can continue. A problem with one FICON Express port does not require replacement of a complete feature.

Each FICON Express feature also supports cascading, which is the connection of two FICON Directors in succession. This configuration minimizes the number of cross-site connections and helps reduce implementation costs for disaster recovery applications, IBM Geographically Dispersed Parallel Sysplex (GDPS), and remote copy.

z15 servers support 32 K devices per FICON channel for all FICON features.

Each FICON Express channel can be defined independently for connectivity to servers, switches, directors, disks, tapes, and printers, by using the following CHPID types:

- ▶ CHPID type FC: The FICON, zHPF, and FCTC protocols are supported simultaneously.
- ▶ CHPID type FCP: Fibre Channel Protocol that supports attachment to SCSI devices directly or through Fibre Channel switches or directors.

FICON channels (CHPID type FC or FCP) can be shared among LPARs and defined as spanned. All ports on a FICON feature must be of the same type (LX or SX). The features are connected to a FICON capable control unit (point-to-point or switched point-to-point) through a Fibre Channel switch.

FICON Express16SA

The FICON Express16SA feature is installed in the PCIe+ I/O drawer. Each of the two independent ports is capable of 8 Gbps or 16 Gbps. The link speed depends on the capability of the attached switch or device. The link speed is auto-negotiated, point-to-point, and is transparent to users and applications.

The following types of FICON Express16SA optical transceivers are supported (no mix on same card):

- ▶ FICON Express16SA LX feature, FC 0436, with two ports per feature, supporting LC Duplex connectors
- ▶ FICON Express16SA SX feature, FC 0437, with two ports per feature, supporting LC Duplex connectors

For supported distances, see Table 4-6 on page 171.

Consideration: FICON Express16SA features do not support auto-negotiation to a data link rate of 2 or 4 Gbps (only 8, or 16 Gbps) for point-to-point connections except with through a switch with 8 or 16 Gb optics.

IBM Fibre Channel Endpoint Security (z15 T01 only)

IBM z15 Model T01 supports IBM Fibre Channel Endpoint Security feature (FC 1146). FC 1146 provides FC/FCP link encryption and endpoint authentication. This is an optional priced feature which requires the following:

- ▶ FICON Express16SA for both link encryption and endpoint authentication
 - FICON Express16S+ for endpoint authentication only.
- ▶ Select DS8000 storage
- ▶ Supporting infrastructure - IBM Security Key Lifecycle Manager 3.01
- ▶ CPACF enablement (FC 3863)

See the following announcement letter:

<https://www.ibm.com/downloads/cas/US-ENUS120-013-CA/name/US-ENUS120-013-CA.PDF>

FICON Express16S+

The FICON Express16S+ feature is installed in the PCIe+ I/O drawer. Each of the two independent ports is capable of 4 Gbps, 8 Gbps, or 16 Gbps. The link speed depends on the capability of the attached switch or device. The link speed is auto-negotiated, point-to-point, and is transparent to users and applications.

The following types of FICON Express16S+ optical transceivers are supported (no mix on same card):

- ▶ FICON Express16S+ LX feature, FC 0427, with two ports per feature, supporting LC Duplex connectors
- ▶ FICON Express16S+ SX feature, FC 0428, with two ports per feature, supporting LC Duplex connectors

For more information about supported distances, see Table 4-6 on page 171.

For more information, see FICON Express chapter in *IBM Z Connectivity Handbook*, SG24-5444.

Consideration: FICON Express16S+ features do not support auto-negotiation to a data link rate of 2 Gbps (only 4, 8, or 16 Gbps).

FICON Express16S

The FICON Express16S feature is installed in the PCIe+ I/O drawer. Each of the two independent ports is capable of 4 Gbps, 8 Gbps, or 16 Gbps. The link speed depends on the capability of the attached switch or device. The link speed is auto-negotiated, point-to-point, and is transparent to users and applications.

The following types of FICON Express16S optical transceivers are supported:

- ▶ FICON Express16S LX feature, FC 0418, with two ports per feature, supporting LC Duplex connectors
- ▶ FICON Express16S SX feature, FC 0419, with two ports per feature, supporting LC Duplex connectors

For more information about supported distances, see Table 4-6 on page 171.

Consideration: FICON Express16S features do not support auto-negotiation to a data link rate of 2 Gbps (only 4, 8, or 16 Gbps).

FICON Express8S

The FICON Express8S feature is installed in the PCIe I/O drawer. Each of the two independent ports is capable of 2 Gbps, 4 Gbps, or 8 Gbps. The link speed depends on the capability of the attached switch or device. The link speed is auto-negotiated, point-to-point, and is transparent to users and applications.

The following types of FICON Express8S optical transceivers are supported:

- ▶ FICON Express8S LX feature, FC 0409, with two ports per feature, supporting LC Duplex connectors
- ▶ FICON Express8S SX feature, FC 0410, with two ports per feature, supporting LC Duplex connectors

For more information about supported distances, see Table 4-6 on page 171.

FICON enhancements

Together with the FICON Express16SA and FICON Express16S+, z15 servers provide enhancements for FICON in functional and performance aspects with IBM Endpoint Security solution.

Forward Error Correction

Forward Error Correction (FEC) is a technique that is used for reducing data errors when transmitting over unreliable or noisy communication channels (improving signal to noise ratio). By adding redundancy error-correction code (ECC) to the transmitted information, the receiver can detect and correct several errors without requiring retransmission. This process feature improves signal reliability and bandwidth use by reducing retransmissions because of bit errors, especially for connections across long distance, such as an inter-switch link (ISL) in a GDPS Metro Mirror environment.

The FICON Express16SA, FICON Express16S+, and FICON Express16S are designed to support FEC coding on top of its 64b/66b data encoding for 16 Gbps connections. This design can correct up to 11 bit errors per 2112 bits transmitted. Therefore, while connected to devices that support FEC at 16 Gbps connections, the FEC design allows FICON Express16SA, FICON Express16S+, and FICON Express16S channels to operate at higher speeds over longer distances with reduced power and higher throughput while retaining the same reliability and robustness for which FICON channels are traditionally known.

With the IBM DS8870 or newer, z15 servers can extend the use of FEC to the fabric N_Ports for a completed end-to-end coverage of 16 Gbps FC links. For more information, see the *IBM DS8884 and z13s: A new cost optimized solution*, REDP-5327.

FICON dynamic routing

With the IBM z15, IBM z14 ZR1, IBM z14 M0x, IBM z13, and IBM z13s servers, FICON channels are no longer restricted to the use of static SAN routing policies for ISLs for cascaded FICON directors. The Z servers now support dynamic routing in the SAN with the FICON Dynamic Routing (FIDR) feature. It is designed to support the dynamic routing policies that are provided by the FICON director manufacturers; for example, Brocade's exchange-based routing (EBR) and Cisco's originator exchange ID (OxID)⁶ routing.

A static SAN routing policy normally assigns the ISL routes according to the incoming port and its destination domain (port-based routing), or the source and destination ports pairing (device-based routing).

The port-based routing (PBR) assigns the ISL routes statically that is based on "first-come, first-served" when a port starts a fabric login (FLOGI) to a destination domain. The ISL is round-robin that is selected for assignment. Therefore, I/O flow from same incoming port to same destination domain always is assigned the same ISL route, regardless of the destination port of each I/O. This setup can result in some ISLs overloaded while some are under-used. The ISL routing table is changed whenever Z server undergoes a power-on-reset (POR), so the ISL assignment is unpredictable.

Device-based routing (DBR) assigns the ISL routes statically that is based on a hash of the source and destination port. That I/O flow from same incoming port to same destination is assigned to same ISL route. Compared to PBR, the DBR is more capable of spreading the load across ISLs for I/O flow from the same incoming port to different destination ports within a destination domain.

When a static SAN routing policy is used, the FICON director features limited capability to assign ISL routes based on workload. This limitation can result in unbalanced use of ISLs (some might be overloaded, while others are under-used).

The dynamic routing ISL routes are dynamically changed based on the Fibre Channel exchange ID, which is unique for each I/O operation. ISL is assigned at I/O request time, so different I/Os from same incoming port to same destination port are assigned different ISLs.

With FIDR, z15 servers feature the following advantages for performance and management in configurations with ISL and cascaded FICON directors:

- ▶ Support sharing of ISLs between FICON and FCP (PPRC or distributed)
- ▶ I/O traffic is better balanced between all available ISLs
- ▶ Improved use of FICON director and ISL
- ▶ Easier to manage with a predictable and repeatable I/O performance

⁶ Check with the switch provider for their support statement.

FICON dynamic routing can be enabled by defining dynamic routing-capable switches and control units in HCD. Also, z/OS implemented a health check function for FICON dynamic routing.

Improved zHPF I/O execution at distance

By introducing the concept of pre-deposit writes, zHPF reduces the number of round trips of standard FCP I/Os to a single round trip. Originally, this benefit is limited to writes that are less than 64 KB. zHPF on z15, z14 ZR1, z14 M0x, z13s, and z13 servers were enhanced to allow all large write operations (greater than 64 KB) at distances up to 100 km (62.1 miles) to be run in a single round trip to the control unit. This improvement avoids elongating the I/O service time for these write operations at extended distances.

Read Diagnostic Parameter Extended Link Service support

To improve the accuracy of identifying a failed component without unnecessarily replacing components in a SAN fabric, a new Extended Link Service (ELS) command called Read Diagnostic Parameters (RDP) was added to the Fibre Channel T11 standard to allow Z servers to obtain extra diagnostic data from the SFP optics that are throughout the SAN fabric.

z15, z14 ZR1, z14 M0x, z13s, and z13 servers now can read this extra diagnostic data for all the ports that are accessed in the I/O configuration and make the data available to an LPAR. For z/OS LPARs that use FICON channels, z/OS displays the data with a new message and display command. For Linux on Z, z/VM, and z/VSE, and LPARs that use FCP channels, this diagnostic data is available in a new window in the SAN Explorer tool.

N_Port ID Virtualization enhancement

N_Port ID Virtualization (NPIV) allows multiple system images (in LPARs or z/VM guests) to use a single FCP channel as though each were the sole user of the channel. First introduced with IBM z9® EC, this feature can be used with earlier FICON features that were carried forward from earlier servers.

By using the FICON Express16S (or newer) as an FCP channel with NPIV enabled, the maximum numbers of the following aspects for one FCP physical channel are doubled:

- ▶ Maximum number of NPIV hosts defined: 64
- ▶ Maximum number of remote N_Ports communicated: 1024
- ▶ Maximum number of addressable LUNs: 8192
- ▶ Concurrent I/O operations: 1528

For more information about operating systems that support NPIV, see [“N_Port ID Virtualization” on page 304](#).

Export/import physical port WWPNs for FCP Channels

IBM Z automatically assigns worldwide port names (WWPNs) to the physical ports of an FCP channel that is based on the PCHID. This WWPN assignment changes when an FCP channel is moved to a different physical slot position.

z15, z14 ZR1, z14 M0x, z13, and z13s servers allow for the modification of these default assignments, which also allows FCP channels to keep previously assigned WWPNs, even after being moved to a different slot position. This capability can eliminate the need for reconfiguration of the SAN in many situations, and is especially helpful during a system upgrade (FC 0099 - WWPN Persistence).

Note: For more information about the FICON enhancement, see *Get More Out of Your IT Infrastructure with IBM z13 I/O Enhancements*, REDP-5134.

FICON support for multiple-hop cascaded SAN configurations

Before the introduction of z13 and z13s servers, IBM Z FICON SAN configurations supported a single ISL (a single hop) in a cascaded FICON SAN environment only. The z15, z14 ZR1, z14 M0x, z13, and z13s servers now support up to three hops in a cascaded FICON SAN environment. This support allows clients to more easily configure a three- or four-site disaster recovery solution.

For more information about the FICON multi-hop, see the [FICON Multihop: Requirements and Configurations white paper](#) at the IBM Techdocs Library website.

FICON feature summary

The FICON feature codes, cable type, maximum unrepeated distance, and the link data rate on a z15 T01 server are listed in Table 4-7. All FICON features use LC Duplex connectors.

Table 4-7 FICON Features

| Channel feature | Feature codes | Bit rate | Cable type | Maximum unrepeated distance ^a (MHz -km) |
|-----------------------------------|---------------|----------------------------|------------------------|---|
| FICON Express16SA 10KM LX | 0436 | 8, or 16 Gbps ^b | SM 9 μm | 10 km |
| FICON Express16SA SX ^c | 0437 | 16 Gbps | MM 50 μm | 35 m (500) 100 m (2000) 125 m (4700) |
| | | 8 Gbps | MM 62.5 μm MM 50 μm | 21 m (200) 50 m (500) 150 m (2000) 190 m (4700) |
| FICON Express16S+ 10KM LX | 0427 | 4, 8, or 16 Gbps | SM 9 μm | 10 km |
| FICON Express16S+ SX | 0428 | 16 Gbps | MM 50 μm | 35 m (500) 100 m (2000) 125 m (4700) |
| | | 8 Gbps | MM 62.5 μm MM 50 μm | 21 m (200) 50 m (500) 150 m (2000) 190 m (4700) |
| | | 4 Gbps | MM 62.5 μm MM 50 μm | 70 m (200) 150 m (500) 380 m (2000) 400 m (4700) |
| FICON Express16S 10KM LX | 0418 | 4, 8, or 16 Gbps | SM 9 μm | 10 km |
| FICON Express16S SX | 0419 | 16 Gbps | MM 50 μm | 35 m (500) 100 m (2000) 125 m (4700) |
| | | 8 Gbps | MM 62.5 μm MM 50 μm | 21 m (200) 50 m (500) 150 m (2000) 190 m (4700) |
| | | 4 Gbps | MM 62.5 μm MM 50 μm | 70 m (200) 150 m (500) 380 m (2000) 400 m (4700) |

| Channel feature | Feature codes | Bit rate | Cable type | Maximum unrepeated distance ^a (MHz -km) |
|-------------------------|---------------|-----------------|------------------------|---|
| FICON Express8S 10KM LX | 0409 | 2, 4, or 8 Gbps | SM 9 μm | 10 km |
| FICON Express8S SX | 0410 | 8 Gbps | MM 62.5 μm MM 50 μm | 21 m (200) 50 m (500) 150 m (2000) 190 m (4700) |
| | | 4 Gbps | MM 62.5 μm MM 50 μm | 70 m (200) 150 m (500) 380 m (2000) 400 m (4700) |
| | | 2 Gbps | MM 62.5 μm MM 50 μm | 150 m (200) 300 m (500) 500 m (2000) N/A (4700) |

- a. Minimum fiber bandwidths in MHz/km for multimode fiber optic links are included in parentheses, where applicable.
- b. 2 and 4 Gbps connectivity is not supported for point-to-point connections
- c. 2 and 4 Gbps connectivity is supported through a switch with 8 or 16 Gb optics.

zHyperLink Express1.1 (FC 0451)

zHyperLink is a new technology that provides up to 5x reduction in I/O latency times for Db2 read requests with the qualities of service IBM Z clients expect from I/O infrastructure for Db2 v12 with z/OS. The z/OS supported versions for Reads are:

- ▶ z/OS V2.4
- ▶ z/OS V2.3 with PTFs
- ▶ z/OS V2.2 with PTFs
- ▶ z/OS V2.1 with PTFs

The z/OS supported versions for Writes support are:

- ▶ z/OS V2.4
- ▶ z/OS V2.3 with PTFs
- ▶ z/OS V2.2 with PTFs

The zHyperLink Express1.1 feature (FC 0451) provides a low latency direct connection between z15 and DS8k storage system.

The zHyperLink Express1.1 is the result of new business requirements that demand fast and consistent application response times. It dramatically reduces latency by interconnecting the z15 directly to I/O Bay of the DS8k by using PCIe Gen3 x 8 physical link (up to 150-meter [492-foot] distance). A new transport protocol is defined for reading and writing IBM CKD data records⁷, as documented in the zHyperLink interface specification.

On z15, zHyperLink Express1.1 card is a new PCIe Gen3 adapter, which installed in the PCIe+ I/O drawer. HCD definition support was added for new PCIe function type with PORT attributes.

Requirements of zHyperLink Express1.1

The zHyperLink Express feature is available on z15 servers, and includes the following requirements:

⁷ CKD data records are handled by using IBM Enhanced Count Key Data (ECKD) command set.

- ▶ z/OS 2.1 or later
- ▶ 150 m maximum distance in a point-to-point configuration
- ▶ DS8k with I/O Bay Planar board and firmware level 8.4 or later
- ▶ z15 with zHyperLink Express1.1 adapter (FC 0451) installed
- ▶ FICON channel as a driver
- ▶ Only ECKD supported
- ▶ z/VM is not supported

Up to 16 zHyperLink Express adapters can be installed in a z15 (up to 32 links).

The zHyperLink Express1.1 is virtualized as a native PCIe adapter and can be shared by multiple LPARs. Each port can support up to 127 Virtual Functions (VFs), with one or more VFs/PFIDs being assigned to each LPAR. This configuration gives a maximum of 254 VFs per adapter. The zHyperLink Express requires the following components:

- ▶ zHyperLink connector on DS8k I/O Bay
 - For DS8880 firmware R8.3 or newer, the I/O Bay planar is updated to support the zHyperLink interface. This update includes the update of the PEX 8732 switch to PEX8733 that includes a DMA engine for the zHyperLink transfers, and the upgrade from a copper to optical interface by a CXP connector (provided).
- ▶ Cable
 - The zHyperLink Express1.1 uses optical cable with MTP connector. Maximum supported cable length is 150 meters (492 feet).

zHyperLink Express (FC 0431)

zHyperLink is a new technology that provides up to 5x reduction in I/O latency times for Db2 read requests with the qualities of service IBM Z clients expect from I/O infrastructure for Db2 v12 with z/OS. The z/OS supported versions for Reads are:

- ▶ z/OS V2.4
- ▶ z/OS V2.3 with PTFs
- ▶ z/OS V2.2 with PTFs
- ▶ z/OS V2.1 with PTFs

The z/OS supported versions for Writes support are:

- ▶ z/OS V2.4
- ▶ z/OS V2.3 with PTFs
- ▶ z/OS V2.2 with PTFs

The zHyperLink Express feature (FC 0431) provides a low latency direct connection between z15 and DS8k I/O Port.

The zHyperLink Express is the result of new business requirements that demand fast and consistent application response times. It dramatically reduces latency by interconnecting the z15 directly to I/O Bay of the DS8880 by using PCIe Gen3 x 8 physical link (up to 150-meter [492-foot] distance). A new transport protocol is defined for reading and writing IBM CKD data records⁸, as documented in the zHyperLink interface specification.

On z15, zHyperLink Express card is a new PCIe adapter, which installed in the PCIe+ I/O drawer. HCD definition support was added for new PCIe function type with PORT attributes.

Requirements of zHyperLink

The zHyperLink Express feature is available on z15 servers, and includes the following requirements:

⁸ CKD data records are handled by using IBM Enhanced Count Key Data (ECKD) command set.

- ▶ z/OS 2.1 or later
- ▶ DS888x with I/O Bay Planar board and firmware level 8.4 or later
- ▶ z15 with zHyperLink Express adapter (FC 0431) installed
- ▶ FICON channel as a driver
- ▶ Only ECKD supported
- ▶ z/VM is not supported

Up to 16 zHyperLink Express adapters can be installed in a z14 ZR1 (up to 32 links).

The zHyperLink Express is virtualized as a native PCIe adapter and can be shared by multiple LPARs. Each port can support up to 127 Virtual Functions (VFs), with one or more VFs/PFIDs being assigned to each LPAR. This configuration gives a maximum of 254 VFs per adapter. The zHyperLink Express requires the following components:

- ▶ zHyperLink connector on DS8880 I/O Bay

For DS8880 firmware R8.4 or newer, the I/O Bay planar is updated to support the zHyperLink interface. This update includes the update of the PEX 8732 switch to PEX8733 that includes a DMA engine for the zHyperLink transfers, and the upgrade from a copper to optical interface by a CXP connector (provided).

- ▶ Cable

The zHyperLink Express uses optical cable with MTP connector. Maximum supported cable length is 150 meters (492 feet).

4.6.3 Network connectivity

Communication for LANs is provided by the OSA-Express7S, OSA-Express6S, OSA-Express5S, 25GbE RoCE Express2.1, 10GbE RoCE Express2.1, 25GbE RoCE Express2, 10GbE RoCE Express2, and 10GbE RoCE Express features.

OSA-Express7S 25GbE SR1.1 (FC 0449)

OSA-Express7S 25 Gigabit Ethernet SR1.1 (FC 0449) is installed in the PCIe+ I/O drawer.

OSA-Express7S 25 Gigabit Ethernet Short Reach1.1 (SR1.1) feature includes one PCIe Gen3 adapter and one port per feature. The port supports CHPID types OSD.

The OSA-Express7S 25GbE SR1.1 feature is designed to support attachment to a multimode fiber 25 Gbps Ethernet LAN or Ethernet switch that is capable of 25 Gbps. The port can be defined as a spanned channel and shared among LPARs within and across logical channel subsystems.

The OSA-Express7S 25GbE SR1.1 feature supports the use of an industry standard small form factor (SFP+) LC Duplex connector. Ensure that the attaching or downstream device has an SR transceiver. The sending and receiving transceivers must be the same (SR to SR).

The OSA-Express7S 25GbE SR1.1 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 50 μ m multimode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

OSA-Express7S 25GbE SR (FC 0429)

OSA-Express7S 25 Gigabit Ethernet SR (FC 0429) is installed in the PCIe+ I/O drawer.

The OSA-Express7S 25GbE Short Reach (SR) feature includes one PCIe adapter and one port per feature. The port supports CHPID types OSD. The OSA-Express7S 25GbE feature is designed to support attachment to a multimode fiber 25 Gbps Ethernet LAN or Ethernet switch that is capable of 25 Gbps. The port can be defined as a spanned channel and shared among LPARs within and across logical channel subsystems.

The OSA-Express7S 25GbE SR feature supports the use of an industry standard small form factor (SFP+) LC Duplex connector. Ensure that the attaching or downstream device has an SR transceiver. The sending and receiving transceivers must be the same (SR-to-SR).

The OSA-Express7S 25GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 50 μ m multimode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

The following other OSA-Express7S features can be installed on z15 servers:

- ▶ OSA-Express7S 10 Gigabit Ethernet LR, FC 0444
- ▶ OSA-Express7S 10 Gigabit Ethernet SR, FC 0445
- ▶ OSA-Express7S Gigabit Ethernet LX, FC 0442
- ▶ OSA-Express7S Gigabit Ethernet SX, FC 0443
- ▶ OSA-Express7S 1000BASE-T Ethernet, FC 0446

The supported OSA-Express7S features are listed in Table 4-5 on page 167.

OSA-Express7S 10 Gigabit Ethernet LR (FC 0444)

The OSA-Express7S 10 Gigabit Ethernet (GbE) Long Reach (LR) feature includes one PCIe Gen3 adapter and one port per feature. The port supports CHPID types OSD. The 10 GbE feature is designed to support attachment to a single-mode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 10 Gbps. The port can be defined as a spanned channel and can be shared among LPARs within and across logical channel subsystems.

The OSA-Express7S 10 GbE LR feature supports the use of an industry standard small form factor (SFP+) LC Duplex connector. Ensure that the attaching or downstream device includes an LR transceiver. The transceivers at both ends must be the same (LR-to-LR).

The OSA-Express7S 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 9 μ m single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting this feature to the selected device.

OSA-Express7S 10 GbE SR

The OSA-Express7S 10 GbE Short Reach (SR) feature (FC 0445) includes one PCIe Gen3 adapter and one port per feature. The port supports CHPID types OSD. The 10 GbE feature is designed to support attachment to a multimode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 10 Gbps. The port can be defined as a spanned channel and shared among LPARs within and across logical channel subsystems.

The OSA-Express7S 10 GbE SR feature supports the use of an industry standard small form factor (SFP+) LC Duplex connector. Ensure that the attaching or downstream device has an SR transceiver. The sending and receiving transceivers must be the same (SR to SR).

The OSA-Express7S 10 GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 50 or a 62.5 μm multimode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

OSA-Express7S Gigabit Ethernet LX (FC 0442)

The OSA-Express7S GbE LX feature includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID types OSD or OSC). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and can be shared among LPARs and across logical channel subsystems.

The OSA-Express7S GbE LX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an LX transceiver. The sending and receiving transceivers must be the same (LX to LX).

A 9 μm single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device. If multimode fiber optic cables are being reused, a pair of Mode Conditioning Patch cables is required, with one cable for each end of the link.

OSA-Express7S GbE SX

The OSA-Express7S GbE SX feature (FC 0443) includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID types OSD or OSC). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and shared among LPARs and across logical channel subsystems.

The OSA-Express7S GbE SX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an SX transceiver. The sending and receiving transceivers must be the same (SX-to-SX).

A multi-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

OSA-Express7S 1000BASE-T Ethernet, FC 0446

Feature code 0446 occupies one slot in the PCIe+ I/O drawer. It features two ports that connect to a 1000 Mbps (1 Gbps) Ethernet LAN. Each port has an SFP+ with an RJ-45 receptacle for cabling to an Ethernet switch. The RJ-45 receptacle is required to be attached by using an EIA/TIA Category 5 or Category 6 UTP cable with a maximum length of 100 meters (328 feet). The SFP allows a concurrent repair or replace action.

The OSA-Express7S 1000BASE-T Ethernet feature does not support auto-negotiation. It supports links at 1000 Mbps in full duplex mode only.

The OSA-Express7S 1000BASE-T Ethernet feature can be configured as CHPID type OSC, OSD, or OSE. Non-QDIO operation mode requires CHPID type OSE.

Notes: Consider the following points:

- ▶ CHPID type OSM is not supported on z15 for user configurations. It is used only in DPM mode for internal management.
- ▶ CHPID types OSN and OSX are not supported on z15.

OSA-Express6S

The OSA-Express6S feature is installed in the PCIe+ I/O drawer. The following OSA-Express6S features can be installed on z15 servers (carry forward only):

- ▶ OSA-Express6S 10 Gigabit Ethernet LR (FC 0424)

- ▶ OSA-Express6S 10 Gigabit Ethernet SR (FC 0425)
- ▶ OSA-Express6S Gigabit Ethernet LX (FC 0422)
- ▶ OSA-Express6S Gigabit Ethernet SX (FC 0423)
- ▶ OSA-Express6S 1000BASE-T Ethernet (FC 0426)

The supported OSA-Express6S features are listed in Table 4-5 on page 167.

OSA-Express6S 10 GbE LR

The OSA-Express6S 10 GbE LR feature (FC 0424) includes one PCIe adapter and one port per feature. On z15, the port supports CHPID type OSD. The 10 GbE feature is designed to support attachment to a single-mode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 10 Gbps. The port can be defined as a spanned channel and can be shared among LPARs within and across logical channel subsystems.

The OSA-Express6S 10 GbE LR feature supports the use of an industry standard small form factor LC Duplex connector. Ensure that the attaching or downstream device includes an LR transceiver. The transceivers at both ends must be the same (LR-to-LR).

The OSA-Express6S 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 9 μm single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting this feature to the selected device.

For supported distances, see Table 4-8 on page 188.

OSA-Express6S 10 GbE SR

The OSA-Express6S 10 GbE SR feature (FC 0416) includes one PCIe adapter and one port per feature. On z15, the port supports CHPID type OSD. The 10 GbE feature is designed to support attachment to a multimode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 10 Gbps. The port can be defined as a spanned channel and shared among LPARs within and across logical channel subsystems.

The OSA-Express6S 10 GbE SR feature supports the use of an industry-standard small form factor LC Duplex connector. Ensure that the attaching or downstream device has an SR transceiver. The sending and receiving transceivers must be the same (SR-to-SR).

The OSA-Express6S 10 GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 50 or a 62.5 μm multimode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

For supported distances, see Table 4-8 on page 188.

OSA-Express6S GbE LX

The OSA-Express6S GbE LX feature (FC 0422) includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID type OSD). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and can be shared among LPARs and across logical channel subsystems.

The OSA-Express6S GbE LX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an LX transceiver. The sending and receiving transceivers must be the same (LX-to-LX).

A 9 µm single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device. If multimode fiber optic cables are being reused, a pair of Mode Conditioning Patch cables is required, with one cable for each end of the link.

For supported distances, see Table 4-8 on page 188.

OSA-Express6S GbE SX

The OSA-Express6S GbE SX feature (FC 0423) includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID type OSD). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and shared among LPARs and across logical channel subsystems.

The OSA-Express6S GbE SX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an SX transceiver. The sending and receiving transceivers must be the same (SX-to-SX).

A multi-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

For supported distances, see Table 4-8 on page 188.

OSA-Express6S 1000BASE-T Ethernet feature

This feature (FC 0426) occupies one slot in the PCIe+ I/O drawer. It features two ports that connect to a 1000 Mbps (1 Gbps) or 100 Mbps Ethernet LAN. Each port has an SFP with an RJ-45 receptacle for cabling to an Ethernet switch. The RJ-45 receptacle is required to be attached by using an EIA/TIA Category 5 or Category 6 UTP cable with a maximum length of 100 meters (328 feet). The SFP allows a concurrent repair or replace action.

The OSA-Express6S 1000BASE-T Ethernet feature supports auto-negotiation when attached to an Ethernet router or switch. If you allow the LAN speed and duplex mode to default to auto-negotiation, the OSA-Express port and the attached router or switch auto-negotiate the LAN speed and duplex mode settings between them. They then connect at the highest common performance speed and duplex mode of interoperation. If the attached Ethernet router or switch does not support auto-negotiation, the OSA-Express port examines the signal that it is receiving and connects at the speed and duplex mode of the device at the other end of the cable.

The OSA-Express6S 1000BASE-T Ethernet feature can be configured as CHPID type OSC, OSD, or OSE. Non-QDIO operation mode requires CHPID type OSE.

Notes: Consider the following points:

- ▶ CHPID type OSM is not supported on z15 for user configurations. It is used only in DPM mode for internal management.
- ▶ CHPID types OSN and OSX are not supported on z15.

The following settings are supported on the OSA-Express6S 1000BASE-T Ethernet feature port:

- ▶ Auto-negotiate
- ▶ 100 Mbps half-duplex or full-duplex
- ▶ 1000 Mbps full-duplex

If auto-negotiate is not used, the OSA-Express port attempts to join the LAN at the specified speed and duplex mode. If this specified speed and duplex mode do not match the speed and duplex mode of the signal on the cable, the OSA-Express port does not connect.

For more information about supported distances, see Table 4-8 on page 188.

OSA-Express5S

The OSA-Express5S feature is installed in the PCIe I/O drawer. The following OSA-Express5S features can be installed on z14 servers (carry forward only):

- ▶ OSA-Express5S 10 Gigabit Ethernet LR, FC 0415
- ▶ OSA-Express5S 10 Gigabit Ethernet SR, FC 0416
- ▶ OSA-Express5S Gigabit Ethernet LX, FC 0413
- ▶ OSA-Express5S Gigabit Ethernet SX, FC 0414
- ▶ OSA-Express5S 1000BASE-T Ethernet, FC 0417

The OSA-Express5S features are listed in Table 4-5 on page 167.

OSA-Express5S 10 GbE LR

The OSA-Express5S 10 GbE LR feature (FC 0415) includes one PCIe adapter and one port per feature. On z15, the port supports CHPID type OSD.

The 10 GbE feature is designed to support attachment to a single-mode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 10 Gbps. The port can be defined as a spanned channel and shared among LPARs within and across logical channel subsystems.

The OSA-Express5S 10 GbE LR feature supports the use of an industry-standard small form factor LC Duplex connector. Ensure that the attaching or downstream device includes an LR transceiver. The transceivers at both ends must be the same (LR-to-LR).

The OSA-Express5S 10 GbE LR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 9 μm single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting this feature to the selected device.

For supported distances, see Table 4-8 on page 188.

OSA-Express5S 10 Gigabit Ethernet SR

The OSA-Express5S 10 GbE SR feature (FC 0416) includes one PCIe adapter and one port per feature. On z15, the port supports CHPID type OSD.

The 10 GbE feature is designed to support attachment to a multimode fiber 10 Gbps Ethernet LAN or Ethernet switch that is capable of 10 Gbps. The port can be defined as a spanned channel and shared among LPARs within and across logical channel subsystems.

The OSA-Express5S 10 GbE SR feature supports the use of an industry standard small form factor LC Duplex connector. Ensure that the attaching or downstream device includes an SR transceiver. The sending and receiving transceivers must be the same (SR-to-SR).

The OSA-Express5S 10 GbE SR feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

A 50 or a 62.5 μm multimode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

For more information about supported distances, see Table 4-8 on page 188.

OSA-Express5S Gigabit Ethernet LX (FC 0413)

The OSA-Express5S GbE LX feature includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID type OSD exclusively). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and shared among LPARs and across logical channel subsystems.

The OSA-Express5S GbE LX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an LX transceiver. The sending and receiving transceivers must be the same (LX-to-LX).

A 9 μ m single-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device. If multimode fiber optic cables are being reused, a pair of Mode Conditioning Patch cables is required, with one cable for each end of the link.

For more information about supported distances, see Table 4-8 on page 188.

OSA-Express5S Gigabit Ethernet SX (FC 0414)

The OSA-Express5S GbE SX feature includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID type OSD exclusively). The ports support attachment to a 1 Gbps Ethernet LAN. Each port can be defined as a spanned channel and can be shared among LPARs and across logical channel subsystems.

The OSA-Express5S GbE SX feature supports the use of an LC Duplex connector. Ensure that the attaching or downstream device has an SX transceiver. The sending and receiving transceivers must be the same (SX-to-SX).

A multi-mode fiber optic cable that ends with an LC Duplex connector is required for connecting each port on this feature to the selected device.

For more information about supported distances, see Table 4-8 on page 188.

OSA-Express5S 1000BASE-T Ethernet feature

This feature (FC 0417) occupies one slot in the PCIe I/O drawer. It has two ports that connect to a 1000 Mbps (1 Gbps) or 100 Mbps Ethernet LAN. Each port has an SFP with an RJ-45 receptacle for cabling to an Ethernet switch. The RJ-45 receptacle is required to be attached by using an EIA/TIA Category 5 or Category 6 UTP cable with a maximum length of 100 meters (328 feet). The SFP allows a concurrent repair or replace action.

The OSA-Express5S 1000BASE-T Ethernet feature supports auto-negotiation when attached to an Ethernet router or switch. If you allow the LAN speed and duplex mode to default to auto-negotiation, the OSA-Express port and the attached router or switch auto-negotiate the LAN speed and duplex mode settings between them. They then connect at the highest common performance speed and duplex mode of interoperation. If the attached Ethernet router or switch does not support auto-negotiation, the OSA-Express port examines the signal that it is receiving and connects at the speed and duplex mode of the device at the other end of the cable.

The OSA-Express5S 1000BASE-T Ethernet feature can be configured as CHPID type OSC, OSD, or OSE. Non-QDIO operation mode requires CHPID type OSE.

Notes: Consider the following points:

- ▶ CHPID type OSM is not supported on z15 for user configurations. It is used only in DPM mode for internal management.
- ▶ CHPID types OSN and OSX are not supported on z15.

The following settings are supported on the OSA-Express5S 1000BASE-T Ethernet feature port:

- ▶ Auto-negotiate
- ▶ 100 Mbps half-duplex or full-duplex
- ▶ 1000 Mbps full-duplex

If auto-negotiate is not used, the OSA-Express port attempts to join the LAN at the specified speed and duplex mode. If this specified speed and duplex mode do not match the speed and duplex mode of the signal on the cable, the OSA-Express port does not connect.

For more information about supported distances, see Table 4-8 on page 188.

OSA-Express features summary

The OSA-Express feature codes, cable type, maximum unrepeated distance, and the link rate on a z15 T01 system are listed in Table 4-8.

Table 4-8 OSA features

| Channel feature | Feature code | Bit rate in Gbps | Cable type | Maximum unrepeated distance ^a (MHz - km) |
|---------------------------|--------------|------------------|--|---|
| OSA-Express7S 25GbE SR1.1 | 0449 | 25 | MM 50 µm | 70 m (2000) |
| OSA-Express7S 25GbE SR | 0429 | | | 100 m (4700) |
| OSA-Express7S 10GbE LR | 0444 | 10 | SM 9 µm | 10 km (6.8 miles) |
| OSA-Express7S 10GbE SR | 0445 | 10 | MM 62.5 µm MM 50 µm | 33 m (200) 82 m (500) 300 m (2000) |
| OSA-Express7S GbE LX | 0442 | 1.25 | SM 9 µm | 5 km (3.1 miles) |
| OSA-Express7S GbE SX | 0443 | 1.25 | MM 62.5 µm MM 50 µm | 275 m (200) 550 m (500) |
| OSA-Express7S 1000BASE-T | 0446 | 1000 Mbps | Cat 5, Cat 6 unshielded twisted pair (UTP) | 100 m |
| OSA-Express6S 10GbE LR | 0424 | 10 | SM 9 µm | 10 km (6.8 miles) |
| OSA-Express6S 10GbE SR | 0425 | 10 | MM 62.5 µm MM 50 µm | 33 m (200) 82 m (500) 300 m (2000) |
| OSA-Express6S GbE LX | 0422 | 1.25 | SM 9 µm | 5 km (3.1 miles) |
| OSA-Express6S GbE SX | 0423 | 1.25 | MM 62.5 µm MM 50 µm | 275 m (200) 550 m (500) |
| OSA-Express6S 1000BASE-T | 0426 | 100 or 1000 Mbps | Cat 5, Cat 6 unshielded twisted pair (UTP) | 100 m |

| Channel feature | Feature code | Bit rate in Gbps | Cable type | Maximum unrepeated distance ^a (MHz - km) |
|---------------------------------------|--------------|------------------|--|---|
| OSA-Express5S 10GbE LR | 0415 | 10 | SM 9 μm | 10 km (6.8 miles) |
| OSA-Express5S 10GbE SR | 0416 | 10 | MM 62.5 μm MM 50 μm | 33 m (200) 82 m (500) 300 m (2000) |
| OSA-Express5S GbE LX | 0413 | 1.25 | SM 9 μm | 5 km (3.1 miles) |
| OSA-Express5S GbE SX | 0414 | 1.25 | MM 62.5 μm MM 50 μm | 275 m (200) 550 m (500) |
| OSA-Express5S 1000BASE-T ^b | 0417 | 100 or 1000 Mbps | Cat 5, Cat 6 unshielded twisted pair (UTP) | 100 m |

a. Minimum fiber bandwidths in MHz/km for multimode fiber optic links are included in parentheses, where applicable.

b. With OSA-Express5S, the only link rate supported is 1000 Mbps.

25GbE RoCE Express2.1

25GbE RoCE Express2.1 (FC 0450) is installed in the PCIe+ I/O drawer and is supported only on IBM z15 servers. The 25GbE RoCE Express2.1 is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

Switch configuration for RoCE Express2.1: If the IBM 25GbE RoCE Express2.1 features are connected to 25GbE switches, the switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled
- ▶ No firewalls, no routing, and no IEDN

The 25GbE RoCE Express2.1 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

10GbE and 25GbE RoCE features should not be mixed in a z/OS SMC-R Link Group. Mixing same speed RoCE features in the same z/OS SMC-R link group is allowed.

The maximum supported unrepeated distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 25GbE switch or to the 25GbE RoCE Express2.1 feature on the attached server:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet)
- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet)

On IBM z15 servers, both ports are supported by z/OS and can be shared by up to 126 partitions (LPARs) per PCHID. The 25GbE RoCE Express2.1 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector. Both point-to-point connections and switched connections with an enterprise-class 25GbE switch are supported.

On z15, RoCE Express2 and 2.1 features support 63 Virtual Functions per port (126 VFs per feature).

10GbE RoCE Express2.1

10GbE RoCE Express2.1 (FC 0432) is installed in the PCIe+ I/O drawer and is supported on IBM z15 servers. The 10GbE RoCE Express2.1 is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

Switch configuration for RoCE Express2.1: If the IBM 10GbE RoCE Express2.1 features are connected to 10GbE switches, the switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled
- ▶ No firewalls, no routing, and no IEDN

The 10GbE RoCE Express2.1 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

10GbE and 25GbE RoCE features should not be mixed in a z/OS SMC-R Link Group. Mixing same speed RoCE features in the same z/OS SMC-R link group is allowed.

The maximum supported unrepeatable distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 10GbE switch or to the 10GbE RoCE Express2 feature on the attached server:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet)
- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet)

The 10GbE RoCE Express2.1 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector. Both point-to-point connections and switched connections with an enterprise-class 10GbE switch are supported.

On z15, RoCE Express2 and 2.1 support 63 Virtual Functions per port (126 VFs per feature).

25GbE RoCE Express2

25GbE RoCE Express2 (FC 0430) is installed in the PCIe I/O drawer and is supported on IBM z15 servers. The 25GbE RoCE Express2 is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

The 25GbE RoCE Express2 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector. Both point-to-point connections and switched connections with an enterprise-class 25GbE switch are supported.

On z15, RoCE Express2 and 2.1 features support 63 Virtual Functions per port (126 VFs per feature).

Switch configuration for RoCE Express2: If the IBM 25GbE RoCE Express2 features are connected to 25GbE switches, the switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled
- ▶ No firewalls, no routing, and no IEDN

The 25GbE RoCE Express2 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

10GbE and 25GbE RoCE features should not be mixed in a z/OS SMC-R Link Group. Mixing same speed RoCE features in the same z/OS SMC-R link group is allowed.

The maximum supported unrepeatable distance, point-to-point, is 100 meters (328 feet). A client-supplied cable is required. Two types of cables can be used for connecting the port to the selected 25GbE switch or to the 25GbE RoCE Express2 feature on the attached server:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector, which supports 70 meters (229 feet)
- ▶ OM4 50-micron multimode fiber optic cable that is rated at 4700 MHz-km that ends with an LC Duplex connector, which supports 100 meters (328 feet)

10GbE RoCE Express2

RoCE Express2 (FC 0412) is installed in the PCIe+ I/O drawer and is supported on z15 servers. The 10GbE RoCE Express2 is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

On z15 servers, both ports are supported by z/OS and can be shared by up to 126 partitions (LPARs) per PCHID. The 10GbE RoCE Express2 feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector. Both point-to-point connections and switched connections with an enterprise-class 10 GbE switch are supported.

On z15, RoCE Express2 and 2.1 features support 63 Virtual Functions per port (126 VFs per feature). The RAS was improved and ECC double bit correction added starting with FC 0412.

Switch configuration for RoCE Express2: If the IBM 10GbE RoCE Express2 features are connected to 10GbE switches, the switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled
- ▶ No firewalls, no routing, and no IEDN

The 10GbE RoCE Express2 feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

10GbE and 25GbE RoCE features should not be mixed in a z/OS SMC-R Link Group. Mixing same speed RoCE features in the same z/OS SMC-R link group is allowed.

The maximum supported unrepeated distance, point-to-point, is 300 meters (984 feet). A client-supplied cable is required. The following types of cables can be used for connecting the port to the selected 10 GbE switch or to the 10GbE RoCE Express2 feature on the attached server:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector; supports 300 meters (984 feet)
- ▶ OM2 50-micron multimode fiber optic cable that is rated at 500 MHz-km that ends with an LC Duplex connector; supports 82 meters (269 feet)
- ▶ OM1 62.5-micron multimode fiber optic cable that is rated at 200 MHz-km that ends with an LC Duplex connector; supports 33 meters (108 feet)

10GbE RoCE Express

The 10GbE RoCE Express feature (FC 0411) is installed in the PCIe+ I/O drawer. This feature is supported on z14, z14 ZR1, z13, z13s servers and can be carried forward during an MES upgrade to a z15.

The 10GbE RoCE Express is a native PCIe feature. It does not use a CHPID and is defined by using the IOCP **FUNCTION** statement or in the hardware configuration definition (HCD).

Both ports are supported by z/OS and can be shared by up to 31 partitions (LPARs) per PCHID on z15, z14 ZR1, z14 M0x, z13s, and z13.

The 10GbE RoCE Express feature uses SR optics and supports the use of a multimode fiber optic cable that ends with an LC Duplex connector. Point-to-point connections and switched connections with an enterprise-class 10 GbE switch are supported.

Switch configuration for RoCE: If the IBM 10GbE RoCE Express features are connected to 10 GbE switches, the switches must meet the following requirements:

- ▶ Global Pause function enabled
- ▶ Priority flow control (PFC) disabled
- ▶ No firewalls, no routing, and no IEDN

The 10GbE RoCE Express feature does not support auto-negotiation to any other speed and runs in full duplex mode only.

10GbE and 25GbE RoCE features should not be mixed in a z/OS SMC-R Link Group. Mixing same speed RoCE features in the same z/OS SMC-R link group is allowed.

The maximum supported unrepeated distance, point-to-point, is 300 meters (984 feet). A client-supplied cable is required. The following types of cables can be used for connecting the port to the selected 10 GbE switch or to the 10GbE RoCE Express feature on the attached server:

- ▶ OM3 50-micron multimode fiber optic cable that is rated at 2000 MHz-km that ends with an LC Duplex connector; supports 300 meters (984 feet)
- ▶ OM2 50-micron multimode fiber optic cable that is rated at 500 MHz-km that ends with an LC Duplex connector; supports 82 meters (269 feet)
- ▶ OM1 62.5-micron multimode fiber optic cable that is rated at 200 MHz-km that ends with an LC Duplex connector; supports 33 meters (108 feet)

Shared Memory Communications functions

The Shared Memory Communication (SMC) capabilities of the z15 help optimize the communications between applications for server-to-server (SMC-R) or LPAR-to-LPAR (SMC-D) connectivity.

SMC-R

SMC-R provides application transparent use of the RoCE-Express feature. This feature reduces the network overhead and latency of data transfers, which effectively offers the benefits of optimized network performance across processors.

SMC-D

SMC-D was used with the introduction of the Internal Shared Memory (ISM) virtual PCI function. ISM is a virtual PCI network adapter that enables direct access to shared virtual memory, which provides a highly optimized network interconnect for IBM Z intra-CPC communications.

SMC-D maintains the socket-API transparency aspect of SMC-R so that applications that use TCP/IP communications can benefit immediately without requiring any application software or IP topology changes. SMC-D completes the overall SMC solution, which provides synergy with SMC-R.

SMC-R and SMC-D use shared memory architectural concept, which eliminates the TCP/IP processing in the data path, yet preserves TCP/IP Qualities of Service for connection management purposes.

Internal Shared Memory (ISM)

ISM is a function that is supported by z15, z14 ZR1, z14 M0x, z13, and z13s machines. It is the firmware that provides connectivity by using shared memory access between multiple operating system images within the same CPC. ISM creates virtual adapters with shared memory that is allocated for each operating system image.

ISM is defined by the FUNCTION statement with a virtual CHPID (VCHID) in hardware configuration definition (HCD)/IOCDS. Identified by the PNETID parameter, each ISM VCHID defines an isolated, internal virtual network for SMC-D communication, without any hardware component required. Virtual adapters are defined by virtual function (VF) statements. Multiple LPARs can access the same virtual network for SMC-D data exchange by associating their VF with same VCHID.

Applications that use HiperSockets can realize network latency and CPU reduction benefits and performance improvement by using the SMC-D over ISM.

z15 servers support up to 32 ISM VCHIDs per CPC. Each VCHID supports up to 255 VFs, with a total maximum of 8,000 VFs.

HiperSockets

The HiperSockets function of z15 servers provides up to 32 high-speed virtual LAN attachments.

HiperSockets can be customized to accommodate varying traffic sizes. Because HiperSockets does not use an external network, it can free up system and network resources. This advantage can help eliminate attachment costs and improve availability and performance.

HiperSockets eliminates the need to use I/O subsystem operations and traverse an external network connection to communicate between LPARs in the same z15 server. HiperSockets offers significant value in server consolidation when connecting many virtual servers. It can be used instead of certain coupling link configurations in a Parallel Sysplex.

HiperSockets internal networks support the following transport modes:

- ▶ Layer 2 (link layer)
- ▶ Layer 3 (network or IP layer)

Traffic can be IPv4 or IPv6, or non-IP, such as AppleTalk, DECnet, IPX, NetBIOS, or SNA.

HiperSockets devices are protocol-independent and Layer 3-independent. Each HiperSockets device (Layer 2 and Layer 3 mode) features its own Media Access Control (MAC) address. This address allows the use of applications that depend on the existence of Layer 2 addresses, such as Dynamic Host Configuration Protocol (DHCP) servers and firewalls.

Layer 2 support helps facilitate server consolidation and can reduce complexity and simplify network configuration. It also allows LAN administrators to maintain the mainframe network environment similarly to non-mainframe environments.

Packet forwarding decisions are based on Layer 2 information instead of Layer 3. The HiperSockets device can run automatic MAC address generation to create uniqueness within and across LPARs and servers. The use of Group MAC addresses for multicast is supported, and broadcasts to all other Layer 2 devices on the same HiperSockets networks.

Datagrams are delivered only between HiperSockets devices that use the same transport mode. A Layer 2 device cannot communicate directly to a Layer 3 device in another LPAR network. A HiperSockets device can filter inbound datagrams by VLAN identification, the destination MAC address, or both.

Analogous to the Layer 3 functions, HiperSockets Layer 2 devices can be configured as primary or secondary connectors, or multicast routers. This configuration enables the creation of high-performance and high-availability link layer switches between the internal HiperSockets network and an external Ethernet network. It also can be used to connect to the HiperSockets Layer 2 networks of different servers.

HiperSockets Layer 2 is supported by Linux on Z, and by z/VM for Linux guest use.

z15 supports the HiperSockets Completion Queue function that is designed to allow HiperSockets to transfer data synchronously (if possible) and asynchronously, if necessary. This feature combines ultra-low latency with more tolerance for traffic peaks.

With the asynchronous support, data can be temporarily held until the receiver has buffers that are available in its inbound queue during high volume situations. The HiperSockets Completion Queue function requires the following minimum applications⁹:

- ▶ z/OS V2.2 with PTFs
- ▶ Linux on Z distributions:
 - Red Hat Enterprise Linux (RHEL) 6.2
 - SUSE Linux Enterprise Server (SLES) 11 SP2
 - Ubuntu server 16.04 LTS
- ▶ z/VSE V6.2

⁹ Minimum OS support for z15 can differ. For more information, see Chapter 7, “Operating system support” on page 253.

- ▶ z/VM V6.4 with maintenance

In z/VM V6.4 and newer, the virtual switch function transparently bridges a guest virtual machine network connection on a HiperSockets LAN segment. This bridge allows a single HiperSockets guest virtual machine network connection to communicate directly with the following systems:

- ▶ Other guest virtual machines on the virtual switch
- ▶ External network hosts through the virtual switch OSA UPLINK port

RoCE Express features summary

The RoCE Express feature codes, cable type, maximum unrepeated distance, and the link rate on a z15 server are listed in Table 4-9.

Table 4-9 RoCE Express features summary

| Channel feature | Feature code | Bit rate in Gbps | Cable type | Maximum unrepeated distance ^a (MHz - km) |
|-----------------------|--------------|------------------|------------------------|---|
| 25GbE RoCE Express2.1 | 0450 | 25 | MM 50 μm | 70 m (2000) 100 m (4700) |
| 10GbE RoCE Express2.1 | 0432 | 10 | MM 62.5 μm MM 50 μm | 33 m (200) 82 m (500) 300 m (2000) |
| 25GbE RoCE Express2 | 0430 | 25 | MM 50 μm | 70 m (2000) 100 m (4700) |
| 10GbE RoCE Express2 | 0412 | 10 | MM 62.5 μm MM 50 μm | 33 m (200) 82 m (500) 300 m (2000) |
| 10GbE RoCE Express | 0411 | 10 | MM 62.5 μm MM 50 μm | 33 m (200) 82 m (500) 300 m (2000) |

a. Minimum fiber bandwidths in MHz/km for multimode fiber optic links are included in parentheses, where applicable.

4.6.4 Parallel Sysplex connectivity

Coupling links are required in a Parallel Sysplex configuration to provide connectivity from the z/OS images to the coupling facility (CF). A properly configured Parallel Sysplex provides a highly reliable, redundant, and robust IBM Z technology solution to achieve near-continuous availability. A Parallel Sysplex is composed of one or more z/OS operating system images that are coupled through one or more CFs.

This section describes coupling link features supported in a Parallel Sysplex in which a z15 can participate.

Coupling links

The type of coupling link that is used to connect a CF to an operating system LPAR is important. The link performance significantly affects response times and coupling processor usage. For configurations that extend over large distances, the time that is spent on the link can be the largest part of the response time.

IBM z15 supports three coupling link types:

- ▶ Integrated Coupling Adapter Short Reach (ICA SR) links connect directly to the CPC drawer and are intended for short distances between CPCs of up to 150 meters (492.1 feet).
- ▶ Coupling Express Long Reach (CE LR) adapters are in the PCIe+ drawer and support unrepeated distances of up to 10 km (6.21 miles) or up to 100 km (62.1 miles) over qualified WDM services.
- ▶ Internal Coupling (IC) links are for internal links within a CPC.

Attention: Parallel Sysplex supports connectivity between systems that differ by up to two generations (n-2). For example, an IBM z15 can participate in an IBM Parallel Sysplex cluster with z14, z14 ZR1, z13, and z13s systems.

However, the IBM z15 and IBM z14 ZR1 do not support InfiniBand connectivity so these servers support connectivity by using only Integrated Coupling Adapter Short Reach (ICA SR) and Coupling Express Long Reach (CE LR) features. z15 can connect to z13 and z13s only if these servers have ICA SR or CE LR coupling features.

Figure 4-4 shows the following supported Coupling Link connections for the z15:

- ▶ InfiniBand links are supported between z13, z13s and z14 machines
- ▶ Only ICA SR and CE LR links are supported on z15 and z14 ZR1 machines

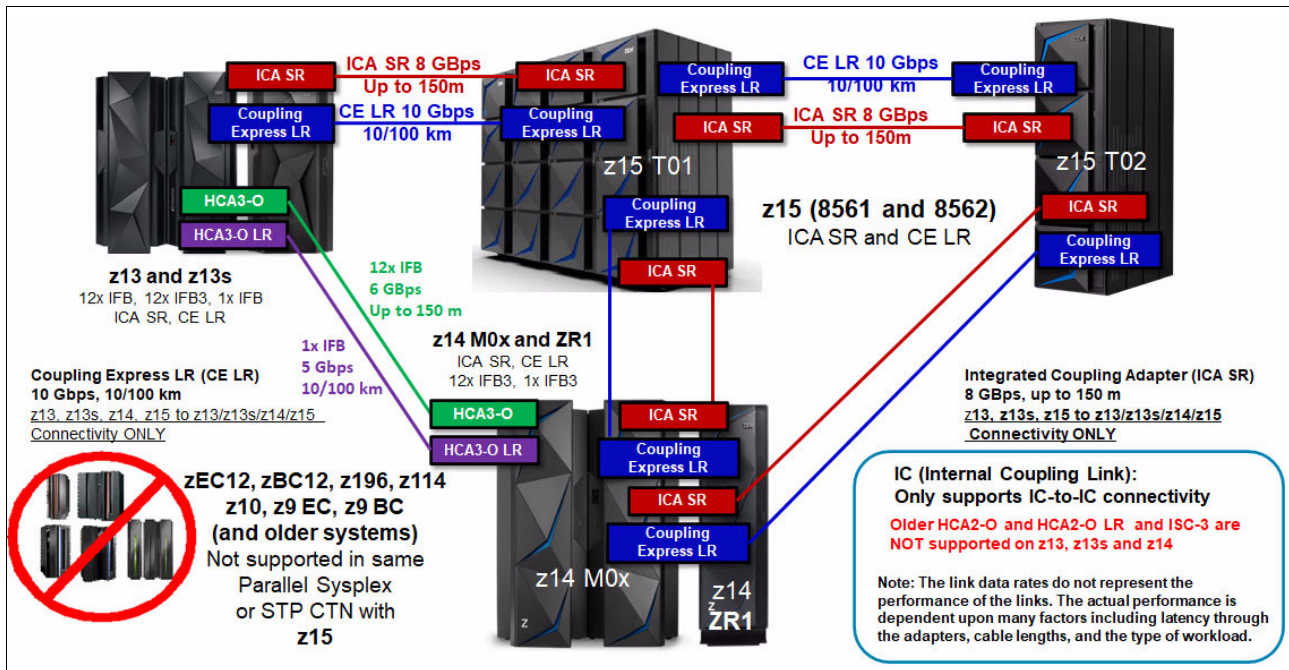


Figure 4-4 Parallel Sysplex connectivity options

The coupling link options that are listed in Table 4-10. Also listed are the coupling link support for each IBM Z platform. Restrictions on the maximum numbers can apply, depending on the configuration. Always check with your IBM support team for more information.

Table 4-10 Coupling link options that are supported on z15

| Type | Description | Feature Code | Link rate | Max unrepeat distance | Maximum number of supported links | | | | |
|-----------|---------------------------------|--------------|-----------------|-----------------------|-----------------------------------|---------|-----|------|-----|
| | | | | | z15 | z14 ZR1 | z14 | z13s | z13 |
| CE LR | Coupling Express LR | 0433 | 10 Gbps | 10 kms (6.2 miles) | 64 | 32 | 64 | 32 | 64 |
| ICA SR1.1 | Integrated Coupling Adapter | 0176 | 8 GBps | 150 meters (492 feet) | 96 | N/A | N/A | N/A | N/A |
| ICA SR | Integrated Coupling Adapter | 0172 | 8 GBps | 150 meters (492 feet) | 96 | 16 | 80 | 16 | 40 |
| IC | Integrated Coupling Adapter | N/A | Internal speeds | N/A | 64 | 32 | 32 | 32 | 32 |
| HCA3-O LR | InfiniBand Long Reach (1 x IFB) | 0170 | | 10 kms (6.2 miles) | N/A | N/A | 64 | 32 | 64 |
| HCA3-O | InfiniBand (12 x IFB) | 0171 | | 150 meters (492 feet) | N/A | N/A | 32 | 16 | 32 |

The maximum number of combined external coupling links (active CE LR, ICA SR links) is 160 per z15 T01 system. z15 systems support up to 384 coupling CHPIDs per CPC. A z15 coupling link support summary is shown in Figure 4-4 on page 196. Consider the following points:

- ▶ The maximum supported links depends on the IBM Z model or capacity feature code and the numbers are marked with an asterisk (*).
- ▶ z15 ICA SR maximum depends on the number of CPU drawers. A total of 12 PCIe+ fanouts are used per CPU drawer, which gives a maximum of 24 ICA SR ports. The z15 machine maximum ICA SR1.1 and ICA SR ports combined is 96.

For more information about distance support for coupling links, see *System z End-to-End Extended Distance Guide*, SG24-8047.

Internal Coupling link

IC links are Licensed Internal Code-defined links to connect a CF to a z/OS logical partition in the same CPC. These links are available on all IBM Z platforms. The IC link is an IBM Z coupling connectivity option that enables high-speed, efficient communication between a CF partition and one or more z/OS logical partitions that are running on the same CPC. The IC is a linkless connection (implemented in LIC) and does not require any hardware or cabling.

An IC link is a fast coupling link that uses memory-to-memory data transfers. IC links do not have PCHID numbers, but do require CHPIDs.

IC links have the following attributes:

- ▶ They provide the fastest connectivity that is significantly faster than external link alternatives.

- ▶ They result in better coupling efficiency than with external links, effectively reducing the CPU cost that is associated with Parallel Sysplex.
- ▶ They can be used in test or production configurations, reduce the cost of moving into Parallel Sysplex technology, and enhance performance and reliability.
- ▶ They can be defined as spanned channels across multiple channel subsystems.
- ▶ They are available at no extra hardware cost (no feature code). Employing ICFs with IC links results in considerable cost savings when configuring a cluster.

IC links are enabled by defining CHPID type ICP. A maximum of 64 IC links can be defined on an IBM z15 CPC.

Integrated Coupling Adapter Short Range

The ICA SR (FC 0172) was introduced with the IBM z13. z15 introduces ICA SR1.1 (FC 0176). ICA SR and ICA SR1.1 are two-port, short-distance coupling features that allow the supported IBM Z systems to connect to each other. ICA SR and ICA SR1.1 use coupling channel type CS5. The ICA SR uses PCIe Gen3 technology, with x16 lanes that are bifurcated into x8 lanes for coupling. ICA SR1.1 uses PCIe Gen4 technology, with x16 lanes that are bifurcated into x8 lanes for coupling.

The ICA SR & SR1.1 are designed to drive distances up to 150 m and supports a link data rate of 8 GBps. It is designed to support up to four CHPIDs per port and eight subchannels (devices) per CHPID.

For more information, see *IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, and Open System Adapters)*, GA23-1407 as [this web page](#).

Coupling Express Long Reach

The Coupling Express LR occupies one slot in a PCIe I/O drawer or PCIe+ I/O drawer¹⁰. It allows the supported IBM Z systems to connect to each other over extended distance. The Coupling Express LR (FC 0433) is a two-port that uses coupling channel type CL5.

The Coupling Express LR uses 10GbE RoCE technology and is designed to drive distances up to 10 km (6.21 miles) unrepeated and support a link data rate of 10 Gigabits per second (Gbps). For distance requirements greater than 10 km (6.21 miles), clients must use a Wavelength Division Multiplexer (WDM). The WDM vendor must be qualified by IBM Z.

Coupling Express LR is designed to support up to four CHPIDs per port, 32 buffers (that is, 32 subchannels) per CHPID. The Coupling Express LR feature is in the PCIe+ I/O drawer on IBM z15.

For more information, see *IBM Z Planning for Fiber Optic Links (FICON/FCP, Coupling Links, Open Systems Adapters, and zHyperLink Express)*, GA23-1408, which is available at [this web page](#).

Extended distance support

For more information about extended distance support, see *System z End-to-End Extended Distance Guide*, SG24-8047.

¹⁰ PCIe+ I/O drawer (FC 4021) is introduced with z15 as-is and built in a 19-inch format. FC 4021 contains 16 I/O slots. FC 4021 can host up to 16 PCIe I/O features (adapters). The PCIe I/O drawer (4032 on z14) cannot be carried forward during and MES upgrade to z15. z15 support only PCIe+ I/O drawers.

Migration considerations

Upgrading from previous generations of IBM Z systems in a Parallel Sysplex to z15 servers in that same Parallel Sysplex requires proper planning for coupling connectivity. Planning is important because of the change in the supported type of coupling link adapters and the number of available fanout slots of the z15 CPC drawers.

The ICA SR fanout features provide short-distance connectivity to another z15, z14 ZR1, z14, z13s, or z13 server.

The CE LR adapter provides long-distance connectivity to another z15, z14 ZR1, z14, z13s, or z13 server.

The z15 server fanout slots in the CPC drawer provide coupling link connectivity through the ICA SR fanout cards. In addition to coupling links for Parallel Sysplex, the fanout cards provide connectivity for the PCIe+ I/O drawer (PCIe+ Gen3 fanout).

Up to 12 PCIe fanout cards can be installed in a z15 CPC drawer.

To migrate from an older generation machine to a z15 without disruption in a Parallel Sysplex environment requires that the older machines are no more than n-2 generation (namely, at least z13) and that they carry enough coupling links to connect to the existing systems while also connecting to the new machine. N-2 generations rule and enough coupling links are necessary to allow individual components (z/OS LPARs, CFs) to be shut down and moved to the target machine and continue connect to the remaining systems.

It is beyond the scope of this book to describe all possible migration scenarios. Always consult with subject matter experts to help you to develop your migration strategy.

Coupling links and Server Time Protocol

All external coupling links can be used to pass time synchronization signals by using Server Time Protocol (STP). STP is a message-based protocol in which timing messages are passed over data links between servers. The same coupling links can be used to exchange time and CF messages in a Parallel Sysplex.

The use of the coupling links to exchange STP messages has the following advantages:

- ▶ STP can scale with distance by using the same links to exchange STP messages and CF messages in a Parallel Sysplex. Servers that are exchanging messages over short distances, such as IFB or ICA SR links, can meet more stringent synchronization requirements than servers that exchange messages over long IFB LR links, with distances up to 100 kilometers (62 miles). This advantage is an enhancement over the IBM Sysplex Timer implementation, which does not scale with distance.
- ▶ Coupling links also provide the connectivity that is necessary in a Parallel Sysplex. Therefore, a potential benefit can be realized of minimizing the number of cross-site links that is required in a multi-site Parallel Sysplex.

Between any two servers that are intended to exchange STP messages, configure each server so that at least two coupling links exist for communication between the servers. This configuration prevents the loss of one link from causing the loss of STP communication between the servers. If a server does not have a CF LPAR, timing-only links can be used to provide STP connectivity.

4.7 Cryptographic functions

Cryptographic functions are provided by the CP Assist for Cryptographic Function (CPACF) and the PCI Express cryptographic adapters. z15 servers support the Crypto Express6S feature.

4.7.1 CPACF functions (FC 3863)

FC 3863¹¹ is required to enable CPACF functions.

4.7.2 Crypto Express7S feature (FC 0898 and FC 0899)

The Crypto Express7S represents the newest generation of the Peripheral Component Interconnect Express (PCIe) cryptographic coprocessors, which are an optional feature that is available on the z15. These coprocessors are Hardware Security Modules (HSMs) that provide high-security cryptographic processing as required by banking and other industries.

This feature provides a secure programming and hardware environment wherein crypto processes are performed. Each cryptographic coprocessor includes general-purpose processors, non-volatile storage, and specialized cryptographic electronics, which are all contained within a tamper-sensing and tamper-responsive enclosure that eliminates all keys and sensitive data on any attempt to tamper with the device. The security features of the HSM are designed to meet the requirements of FIPS 140-2, Level 4, which is the highest security level defined.

The Crypto Express7S (2 port), FC 0898 includes two IBM PCIe Cryptographic Coprocessors (PCIeCC) per feature. The IBM PCIeCC is a hardware security module (HSM). The Crypto Express7S (1 port), FC 0899 includes one IBM PCIe Cryptographic Coprocessors (PCIeCC) per feature. For availability reasons, a minimum of two features is required for the one port feature. Up to 30 Crypto Express7S (2 port) features are supported on z15 T01. The maximum number of the one-port features is 16. The total number of HSMs supported on z15 T01 is 60 in a combination of Crypto Express7S (2 port), Crypto Express7S (1 port), Crypto Express6S, or Crypto Express5S.

The Crypto Express7S feature occupies one I/O slot in a PCIe+ I/O drawer.

Each adapter can be configured as a Secure IBM CCA coprocessor, Secure IBM Enterprise PKCS #11 (EP11) coprocessor, or accelerator.

Crypto Express7S provides domain support for up to 85 logical partitions.

The accelerator function is designed for maximum-speed Secure Sockets Layer and Transport Layer Security (SSL/TLS) acceleration, rather than for specialized financial applications for secure, long-term storage of keys or secrets. The Crypto Express7S can also be configured as one of the following configurations:

- ▶ The Secure IBM CCA coprocessor includes secure key functions with emphasis on the specialized functions that are required for banking and payment card systems. It is optionally programmable to add custom functions and algorithms by using User Defined Extensions (UDX).

A new mode, called Payment Card Industry (PCI) PIN Transaction Security (PTS) Hardware Security Module (HSM) (PCI-HSM), is available exclusively for Crypto

¹¹ Subject to export regulations.

Express6S in CCA mode. PCI-HSM mode simplifies compliance with PCI requirements for hardware security modules.

- ▶ The Secure IBM Enterprise PKCS #11 (EP11) coprocessor implements an industry-standardized set of services that adheres to the PKCS #11 specification v2.20 and more recent amendments. It was designed for extended FIPS and Common Criteria evaluations to meet industry requirements.

This cryptographic coprocessor mode introduced the PKCS #11 secure key function.

TKE feature: The Trusted Key Entry (TKE) Workstation feature is required for supporting the administration of the Crypto Express6S when configured as an Enterprise PKCS #11 coprocessor or managing the CCA mode PCI-HSM.

When the Crypto Express7S PCI Express adapter is configured as a secure IBM CCA co-processor, it still provides accelerator functions. However, up to 3x better performance for those functions can be achieved if the Crypto Express7S PCI Express adapter is configured as an accelerator.

CCA enhancements include the ability to use triple-length (192-bit) Triple-DES (TDES) keys for operations, such as data encryption, PIN processing, and key wrapping to strengthen security. CCA also extended the support for the cryptographic requirements of the German Banking Industry Committee Deutsche Kreditwirtschaft (DK).

Several features that support the use of the AES algorithm in banking applications also were added to CCA. These features include the addition of AES-related key management features and the AES ISO Format 4 (ISO-4) PIN blocks as defined in the ISO 9564-1 standard. PIN block translation and the use of AES PIN blocks in other CCA callable services are supported. IBM continues to add enhancements as AES finance industry standards are released.

4.7.3 Crypto Express6S feature (FC 0893) as carry forward only

Crypto Express5S was introduced from z14 servers. On the initial configuration, a minimum of two features are installed. The number of features then increases one at a time up to a maximum of 16 features.

Each Crypto Express6S feature holds one PCI Express cryptographic adapter. Each adapter can be configured by the installation as a Secure IBM Common Cryptographic Architecture (CCA) coprocessor, as a Secure IBM Enterprise Public Key Cryptography Standards (PKCS) #11 (EP11) coprocessor, or as an accelerator.

The tamper-resistant hardware security module, which is contained on the Crypto Express6S feature, conforms to the Federal Information Processing Standard (FIPS) 140-2 Level 4 Certification. It supports User Defined Extension (UDX) services to implement cryptographic functions and algorithms (when defined as an IBM CCA coprocessor).

The following CCA compliance levels are available:

- ▶ Non-compliant (default)
- ▶ PCI-HSM 2016
- ▶ PCI-HSM 2016 (migration, key tokens while migrating to compliant)

The following EP11 compliance levels are available (Crypto Express6S and Crypto Express5S):

- ▶ FIPS 2009 (default)
- ▶ FIPS 2011

- ▶ BSI 2009
- ▶ BSI 2011

Each Crypto Express6S feature occupies one I/O slot in the PCIe I/O drawer, and features no CHPID assigned. However, it includes one PCHID.

4.7.4 Crypto Express5S feature (FC 0890) as carry forward only

Crypto Express5S was introduced from z13 servers. On the initial configuration, a minimum of two features are installed. The number of features then increases individually to a maximum of 16 features.

Each Crypto Express5S feature holds one PCI Express cryptographic adapter. Each adapter can be configured by the installation as a Secure IBM CCA coprocessor, as a Secure IBM Enterprise Public Key Cryptography Standards (PKCS) #11 (EP11) coprocessor, or as an accelerator.

Each Crypto Express5S feature occupies one I/O slot in the PCIe I/O drawer, and features no CHPID assigned. However, it includes one PCHID.

4.8 Integrated Firmware Processor

The Integrated Firmware Processor (IFP) was introduced with the zEC12 and zBC12 servers. The IFP is dedicated for managing a new generation of PCIe features. The following features are installed in the PCIe+ I/O drawer:

- ▶ 25GbE RoCE Express2.1
- ▶ 10GbE RoCE Express2.1
- ▶ 25GbE RoCE Express2
- ▶ 10GbE RoCE Express2
- ▶ 10GbE RoCE Express
- ▶ Coupling Express Long Reach (CE LR)

All native PCIe features should be ordered in pairs for redundancy. The features are assigned to one of the four resource groups (RGs) that are running on the IFP according to their physical location in the PCIe+ I/O drawer, which provides management functions and virtualization functions.

If two features of the same type are installed, one always is managed by resource group 1 (RG 1) or resource group 3 (RG 3) while the other feature is managed by resource group 2 (RG 2) or resource group 4 (RG 4). This configuration provides redundancy if one of the features or resource groups needs maintenance or fails.

The IFP and RGs support the following infrastructure management functions:

- ▶ Firmware update of adapters and resource groups
- ▶ Error recovery and failure data collection
- ▶ Diagnostic and maintenance tasks



Central processor complex channel subsystem

This chapter describes the concepts of the z15 T01 channel subsystem, including multiple channel subsystems and multiple subchannel sets. It also describes the technology, terminology, and implementation aspects of the channel subsystem.

This chapter includes the following topics:

- ▶ 5.1, “Channel subsystem” on page 204
- ▶ 5.2, “I/O configuration management” on page 212
- ▶ 5.3, “Channel subsystem summary” on page 213

5.1 Channel subsystem

Channel subsystem (CSS) is a collective name of facilities that Z servers use to control I/O operations.

The channel subsystem directs the flow of information between I/O devices and main storage. It allows data processing to proceed concurrently with I/O processing, which relieves data processors (central processor (CP) and Integrated Facility for Linux [IFL]) of the task of communicating directly with I/O devices.

The channel subsystem includes subchannels, I/O devices that are attached through control units, and channel paths between the subsystem and control units. For more information about the channel subsystem, see 5.1.1, “Multiple logical channel subsystems”.

The design of IBM Z servers offers considerable processing power, memory size, and I/O connectivity. In support of the larger I/O capability, the CSS structure is scaled up by introducing the multiple logical channel subsystem (LCSS) since z990, and multiple subchannel sets (MSS) since z9.

An overview of the channel subsystem for z15 servers is shown in Figure 5-1. z15 T01 systems are designed to support up to six logical channel subsystems, each with four subchannel sets and up to 256 channels.

| z15 Model T01 | | | | | |
|---|---|---|---|---|---|
| HSA = 256 GB | | | | | |
| LCSS 0 | LCSS 1 | LCSS 2 | LCSS 3 | LCSS 4 | LCSS 5 |
| Up to 15 Logical Partitions | Up to 15 Logical Partitions | Up to 15 Logical Partitions | Up to 15 Logical Partitions | Up to 15 Logical Partitions | Up to 10 Logical Partitions |
| Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k | Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k | Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k | Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k | Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k | Subchannel Sets: SS 0 – 63.75 k SS 1 – 64 k SS 2 – 64 k SS 3 – 64 k |
| Up to 256 Channels | Up to 256 Channels | Up to 256 Channels | Up to 256 Channels | Up to 256 Channels | Up to 256 Channels |

Figure 5-1 Multiple channel subsystems and multiple subchannel sets

All channel subsystems are defined within a single configuration, which is called I/O configuration data set (IOCDs). The IOCDs is loaded into the hardware system area (HSA) during a central processor complex (CPC) power-on reset (POR) to start all of the channel subsystems.

On z15 T01 systems, the HSA is pre-allocated in memory with a fixed size of 256 GB, which is in addition to the customer purchased memory. This fixed size memory for HSA eliminates the requirement for more planning of the initial I/O configuration and pre-planning for future I/O expansions.

CPC drawer repair: The HSA can be moved from one CPC drawer to a different drawer in an enhanced availability configuration as part of a concurrent CPC drawer repair (CDR) action.

The following objects are always reserved in the z15 T01 HSA during POR, whether they are defined in the IOCDs for use:

- ▶ Six CSSs
- ▶ A total of 15 LPARs in each CSS0 to CSS4
- ▶ A total of 10 LPARs in CSS5
- ▶ Subchannel set 0 with 63.75 K devices in each CSS
- ▶ Subchannel set 1 with 64 K minus one device in each CSS
- ▶ Subchannel set 2 with 64 K minus one device in each CSS
- ▶ Subchannel set 3 with 64 K minus one device in each CSS

5.1.1 Multiple logical channel subsystems

In the z/Architecture, a *single channel subsystem* can have up to 256 channel paths that are defined, which limited the total numbers of I/O connectivity on older Z servers to 256.

The introduction of *multiple LCSSs* enabled an IBM Z server to have more than one channel subsystems logically, while each logical channel subsystem maintains the same manner of I/O processing. Also, a logical partition (LPAR) is now attached to a specific logical channel subsystem, which makes the extension of multiple logical channel subsystems not apparent to the operating systems and applications. The multiple image facility (MIF) in the structure enables resource sharing across LPARs within a single LCSS or across the LCSSs.

The multiple LCSS structure extended the Z servers' total number of I/O connectivity to support a balanced configuration for the growth of processor and I/O capabilities.

A one-digit number ID starting from 0 (CSSID) is assigned to an LCSS, and a one-digit hexadecimal ID (MIF ID) starting from 0 is assigned to an LPAR within the LCSS.

Note: The phrase *channel subsystem* has same meaning as *logical channel subsystem* in this section, unless otherwise stated.

Subchannels

A *subchannel* provides the logical appearance of a device to the program and contains the information that is required for sustaining a single I/O operation. Each device is accessible by using one subchannel in a channel subsystem to which it is assigned according to the active IOCDs of the Z server.

A subchannel set (SS) is a collection of subchannels within a channel subsystem. The maximum number of subchannels of a subchannel set determines how many devices are accessible to a channel subsystem.

In z/Architecture, the first subchannel set of an LCSS can have 63.75 K subchannels (with 0.25 K reserved), with a subchannel set ID (SSID) of 0. By enabling the multiple subchannel sets, extra subchannel sets are available to increase the device addressability of a channel subsystem. For more information about multiple subchannel sets, see 5.1.2, "Multiple subchannel sets" on page 206.

Channel paths

A *channel path* provides a connection between the channel subsystem and control units that allows the channel subsystem to communicate with I/O devices. Depending on the type of connections, a channel path might be a physical connection to a control unit with I/O devices, such as FICON, or an internal logical control unit, such as HiperSockets.

Each channel path in a channel subsystem features a unique 2-digit hexadecimal identifier that is known as a *channel-path identifier* (CHPID), which ranges 00 - FF. Therefore, a total of 256 CHPIDs are supported by a CSS, and a maximum of 1536 CHPIDs are available on a z15 server with six logical channel subsystems.

By assigning a CHPID to a physical port of an I/O feature adapter, such as FICON Express16SA, or a fanout adapter (ICA SR) port, the channel subsystem connects to the I/O devices through these physical ports.

A port on an I/O feature card features a unique physical channel identifier (PCHID) according to the physical location of this I/O feature adapter, and the sequence of this port on the adapter.

In addition, a port on a fanout adapter has a unique adapter identifier (AID), according to the physical location of this fanout adapter, and the sequence of this port on the adapter.

A CHPID is assigned to a physical port by defining the corresponding PCHID or AID in the I/O configuration definitions.

Control units

A *control unit* provides the logical capabilities that are necessary to operate and control an I/O device. It adapts the characteristics of each device so that it can respond to the standard form of control that is provided by the CSS.

A control unit can be housed separately or can be physically and logically integrated with the I/O device, channel subsystem, or within the Z server.

I/O devices

An *I/O device* provides external storage, a means of communication between data-processing systems, or a means of communication between a system and its environment. In the simplest case, an I/O device is attached to one control unit and is accessible through one or more channel paths that are connected to the control unit.

5.1.2 Multiple subchannel sets

A subchannel set is a collection of subchannels within a channel subsystem. The maximum number of subchannels of a subchannel set determines how many I/O devices that a channel subsystem can access. This number also determines the number of addressable devices to the program (for example, an operating system) that is running in the LPAR.

Each subchannel has a unique four-digit hexadecimal number 0x0000 - 0xFFFF. Therefore, a single subchannel set can address and access up to 64 K I/O devices.

As with the z13 server, the z15 T01 systems support four subchannel sets for each logical channel subsystem. It can access a maximum of 255.74 K devices for a logical channel subsystem and a logical partition and the programs that are running on it.

Note: Do not confuse the multiple subchannel sets function with multiple channel subsystems.

Subchannel number

The subchannel number is a four-digit hexadecimal number 0x0000 - 0xFFFF, which is assigned to a subchannel within a subchannel set of a channel subsystem. Subchannels in each subchannel set are always assigned subchannel numbers within a single range of contiguous numbers.

The lowest-numbered subchannel is subchannel 0, and the highest-numbered subchannel includes a subchannel number equal to one less than the maximum numbers of subchannels that are supported by the subchannel set. Therefore, a subchannel number is always unique within a subchannel set of a channel subsystem and depends on the sequence of assigning.

With the subchannel numbers, a program that is running on an LPAR (for example, an operating system) can specify all I/O functions relative to a specific I/O device by designating a subchannel that is assigned to the I/O devices.

Normally, subchannel numbers are used only in communication between the programs and the channel subsystem.

Subchannel set identifier

While introducing the MSS, the channel subsystem is extended to assign a value 0 - 3 for each subchannel set, which is the SSID. A subchannel can be identified by its SSID and subchannel number.

Device number

A device number is an arbitrary number 0x0000 - 0xFFFF, which is defined by a system programmer in an I/O configuration for naming an I/O device. The device number must be unique within a subchannel set of a channel subsystem. It is assigned to the corresponding subchannel by channel subsystem when an I/O configuration is activated. Therefore, a subchannel in a subchannel set of a channel subsystem includes a device number together with a subchannel number for designating an I/O operation.

The device number provide a means to identify a device, independent of any limitations that are imposed by the system model, configuration, or channel-path protocols.

A device number also can be used to designate an I/O function to a specific I/O device. Because it is an arbitrary number, it can easily be fit into any configuration management and operating management scenarios. For example, a system administrator can set all of the z/OS systems in an environment to device number 1000 for their system RES volumes.

With multiple subchannel sets, a subchannel is assigned to a specific I/O device by the channel subsystem with an automatically assigned subchannel number and a device number that is defined by user. An I/O device can always be identified by an SSID with a subchannel number or a device number. For example, a device with device number AB00 of subchannel set 1 can be designated as 1AB00.

Normally, the subchannel number is used by the programs to communicate with the channel subsystem and I/O device, whereas the device number is used by a system programmer, operator, and administrator.

Device in subchannel set 0 and extra subchannel sets

An LCSS always includes the first subchannel set (SSID 0), which can have up to 63.75 K subchannels with 256 subchannels that are reserved by the channel subsystem. Users can always define their I/O devices in this subchannel set for general use.

For the extra subchannel sets enabled by the MSS facility, each has 65535 subchannels (64 K minus one) for specific types of devices. These extra subchannel sets are referred as *alternative subchannel sets* in z/OS. Also, a device that is defined in an alternative subchannel set is considered a *special device*, which normally features a special device type in the I/O configuration.

Currently, a z15 T01 system that is running z/OS defines the following types of devices in another subchannel set, with proper APAP or PTF installed:

- ▶ Alias devices of the parallel access volumes (PAV).
- ▶ Secondary devices of GDPS Metro Mirror Copy Service (formerly Peer-to-Peer Remote Copy [PPRC]).
- ▶ FlashCopy SOURCE and TARGET devices with program temporary fix (PTF) OA46900.
- ▶ Db2 data backup volumes with PTF OA24142.

The use of another subchannel set for these special devices helps reduce the number of devices in the subchannel set 0, which increases the growth capability for accessing more devices.

Initial program load from an alternative subchannel set

z15 T01 systems support initial program load (IPL) from alternative subchannel sets in addition to subchannel set 0. Devices that are used early during IPL processing now can be accessed by using subchannel set 1, subchannel set 2, or subchannel set 3 on a z15 server. This configuration allows the users of Metro Mirror (formerly PPRC) secondary devices that are defined by using the same device number and a new device type in an alternative subchannel set to be used for IPL, an I/O definition file (IODF), and stand-alone memory dump volumes, when needed.

The display ios,config command

The z/OS `display ios,config(a11)` command that is shown in Figure 5-2 includes information about the MSSs.

```
D IOS,CONFIG(ALL)
IOS506I 11.32.19 I/O CONFIG DATA 340
ACTIVE IODF DATA SET = SYS6.IODF39
CONFIGURATION ID = L06RMVS1      EDT ID = 01
TOKEN:  PROCESSOR DATE      TIME      DESCRIPTION
SOURCE: SCZP501 14-10-31 08:51:47 SYS6      IODF39
ACTIVE CSS: 0      SUBCHANNEL SETS CONFIGURED: 0, 1, 2, 3
CHANNEL MEASUREMENT BLOCK FACILITY IS ACTIVE
LOCAL SYSTEM NAME (LSYSTEM): SCZP501
HARDWARE SYSTEM AREA AVAILABLE FOR CONFIGURATION CHANGES
PHYSICAL CONTROL UNITS          8099
CSS 0 - LOGICAL CONTROL UNITS   3996
  SS 0  SUBCHANNELS             54689
  SS 1  SUBCHANNELS             58862
  SS 2  SUBCHANNELS             65535
  SS 3  SUBCHANNELS             65535
CSS 1 - LOGICAL CONTROL UNITS   4088
  SS 0  SUBCHANNELS             65280
  SS 1  SUBCHANNELS             65535
  SS 2  SUBCHANNELS             65535
  SS 3  SUBCHANNELS             65535
CSS 2 - LOGICAL CONTROL UNITS   4088
  SS 0  SUBCHANNELS             65280
  SS 1  SUBCHANNELS             65535
  SS 2  SUBCHANNELS             65535
  SS 3  SUBCHANNELS             65535
CSS 3 - LOGICAL CONTROL UNITS   4088
  SS 0  SUBCHANNELS             65280
  SS 1  SUBCHANNELS             65535
  SS 2  SUBCHANNELS             65535
  SS 3  SUBCHANNELS             65535
CSS 4 - LOGICAL CONTROL UNITS   4088
  SS 0  SUBCHANNELS             65280
  SS 1  SUBCHANNELS             65535
  SS 2  SUBCHANNELS             65535
  SS 3  SUBCHANNELS             65535
CSS 5 - LOGICAL CONTROL UNITS   4088
  SS 0  SUBCHANNELS             65280
  SS 1  SUBCHANNELS             65535
  SS 2  SUBCHANNELS             65535
  SS 3  SUBCHANNELS             65535
```

Figure 5-2 Output for `display ios,config(all)` command with MSS

5.1.3 Channel path spanning

With the implementation of multiple LCSSs, a channel path can be available to LPARs as dedicated, shared, and spanned.

While a shared channel path can be shared by LPARs within a same LCSS, a spanned channel path can be shared by LPARs within and across LCSSs.

By assigning the same CHPID from different LCSSs to the same channel path (for example, a PCHID), the channel path can be accessed by any LPARs from these LCSSs at the same time. The CHPID is spanned across those LCSSs. The use of spanned channels paths decreases the number of channels that are needed in an installation of Z servers.

A sample of channel paths that are defined as dedicated, shared, and spanned is shown in Figure 5-3.

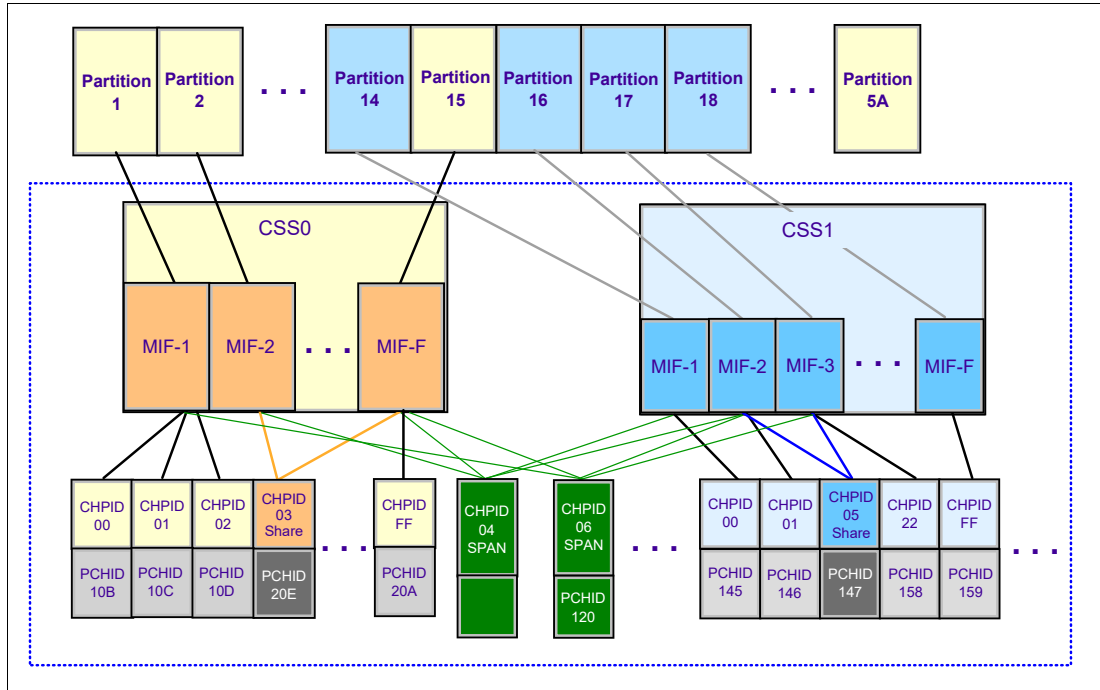


Figure 5-3 IBM Z CSS: Channel subsystems with channel spanning

In the sample, the following definitions of a channel path are shown:

- ▶ CHPID FF, assigned to PCHID 20A, is dedicated access for partition 15 of LCSS0. The same applies to CHPID 00,01,02 of LCSS0, and CHPID 00,01,FF of LCSS1.
- ▶ CHPID 03, assigned to PCHID 20E, is shared access for partition 2, and 15 of LCSS0. The same applies to CHPID 05 of LCSS1.
- ▶ CHPID 06, assigned to PCHID 120 is spanned access for partition 1, 15 of LCSS0, and partition 16, 17 of LCSS1. The same applies to CHPID 04.

Channel spanning is supported for internal links (HiperSockets and IC links) and for certain types of external links. External links that are supported on z15 T01 systems include FICON Express16SA, FICON Express16S+, FICON Express16S, FICON Express8S, OSA-Express7S, OSA-Express6S, OSA-Express5S, and Coupling Links.

The definition of LPAR name, MIF image ID, and LPAR ID are used to identify an LPAR by the channel subsystem to identify I/O functions from different LPARs of multiple LCSSs, which support the implementation of these dedicated, shared, and spanned paths.

An example of definition of these LPAR-related identifications is shown in Figure 5-4.

| CSS0 | CSS1 | CSS2 | CSS3 | CSS4 | CSS5 | Specified in HCD / IOCP |
|--|--|--------------------------|----------------------------|----------------------------|---------------------------|-------------------------------|
| Logical Partition Name TST1 PROD1 PROD2 | Logical Partition Name TST2 PROD3 PROD4 | LPAR Name TST3 TST4 | LPAR Name PROD5 PROD6 | LPAR Name TST55 PROD7 | LPAR Name PROD8 TST6 | Specified in HCD / IOCP |
| Logical Partition ID 02 04 0A | Logical Partition ID 14 16 1D | LPAR ID 22 26 | LPAR ID 35 3A | LPAR ID 44 47 | LPAR ID 56 5A | Specified in Image Profile |
| MIF ID 2 4 A | MIF ID 4 6 D | MIF ID 2 6 | MIF ID 5 A | MIF ID 4 7 | MIF ID 6 A | Specified in HCD / IOCP |

Figure 5-4 CSS, LPAR, and identifier example

LPAR name

The LPAR name is defined as partition name parameter in the **RESOURCE** statement of an I/O configuration. The LPAR name must be unique across the server.

MIF image ID

The MIF image ID is defined as a parameter for each LPAR in the **RESOURCE** statement of an I/O configuration. It ranges 1 - F, and must be unique within an LCSS. However, duplicates are allowed in different LCSSs.

If a MIF image ID is not defined, an arbitrary ID is assigned when the I/O configuration activated. The z15 server supports a maximum of six LCSSs, with a total of 85 LPARs that can be defined.

Each LCSS of a z15 T01 system can support the following numbers of LPARs:

- ▶ LCSS0 to LCSS4 support 15 LPARs each, and the MIF image ID is 1 - F.
- ▶ LCSS5 supports 10 LPARs, and the MIF image IDs are 1 - A.

LPAR ID

The LPAR ID is defined by a user in an image activation profile for each LPAR. It is a 2-digit hexadecimal number 00 - 7F. The LPAR ID must be unique across the server. Although it is arbitrarily defined by the user, an LPAR ID often is the CSS ID concatenated to its MIF image ID, which makes the value more meaningful for the system administrator. For example, an LPAR with LPAR ID 1A defined in that manner means that the LPAR is defined in LCSS1, with the MIF image ID A.

5.2 I/O configuration management

The following tools are available to help maintain and optimize the I/O configuration:

- ▶ IBM Configurator for e-business (eConfig)

The eConfig tool is used by your IBM representative. It is used to create configurations or upgrades of a configuration, and maintains tracking to the installed features of those configurations. eConfig produces reports that help you understand the changes that are being made for a new system, or a system upgrade, and what the target configuration looks like.

- ▶ Hardware configuration definition (HCD)

HCD supplies an interactive dialog to generate the IODF, and later the IOCDs. Generally, use HCD or Hardware Configuration Manager (HCM) to generate the I/O configuration rather than writing I/O configuration program (IOCP) statements. The validation checking that HCD runs against a IODF source file helps minimize the risk of errors before an I/O configuration is activated.

HCD support for multiple channel subsystems is available with z/VM and z/OS. HCD provides the capability to make dynamic hardware and software I/O configuration changes.

Note: Certain functions might require specific levels of an operating system, PTFs, or both.

- ▶ Consult the appropriate fix categories:

- z15 T01: IBM.Device.Server.z15-8561
- z15 T02: IBM.Device.Server.z15-8562
- z14 M0x: IBM.Device.Server.z14-3906
- z14 ZR1: IBM.Device.Server.z14ZR1-3907
- z13: IBM.Device.Server.z13-2964
- z13s: IBM.Device.Server.z13s-2965

- ▶ HCM

HCM is a priced optional feature that supplies a graphical interface of HCD. It is installed on a PC and allows you to manage the physical and logical aspects of a mainframe's hardware configuration.

- ▶ CHPID Mapping Tool (CMT)

The CMT helps to map CHPIDs onto PCHIDs that are based on an IODF source file and the eConfig configuration file of a mainframe. It provides a CHPID to PCHID mapping with high availability for the targeted I/O configuration. It also features built-in mechanisms to generate a mapping according to customized I/O performance groups. More enhancements are implemented in CMT to support z15 servers.

The CMT is available for download from the [IBM Resource Link website](#).

The configuration file for a new machine or upgrade is also available from IBM Resource Link. Ask your IBM technical sales representative for the name of the file to download.

5.3 Channel subsystem summary

z15 T01 systems support the channel subsystem features of multiple LCSS, MSS, and the channel spanning that is described in this chapter. The channel subsystem capabilities of z15 T01 systems are listed in Table 5-1.

Table 5-1 z15 T01 CSS overview

| | |
|--|--|
| Maximum number of CSSs | 6 |
| Maximum number of LPARs per CSS | CSS0 - CSS4: 15 CSS5: 10 |
| Maximum number of LPARs per system | 85 |
| Maximum number of subchannel sets per CSS | 4 |
| Maximum number of subchannels per CSS | 255.74 K SS0: 65280 SS1 - SS3: 65535 |
| Maximum number of CHPIDs per CSS | 256 |



Cryptographic features

This chapter describes the hardware cryptographic functions that are available on IBM z15. The CP Assist for Cryptographic Function (CPACF), together with the Peripheral Component Interconnect Express (PCIe) cryptographic coprocessors (PCIeCC), offer a balanced use of processing resources and unmatched scalability for fulfilling pervasive encryption demands.

The z15 is designed for delivering a transparent and consumable approach that enables extensive (pervasive) encryption of data in flight and at rest, with the goal of substantially simplifying data security and reducing the costs that are associated with protecting data while achieving compliance mandates.

Naming: The IBM z15 server generation is available as the following machine types and models:

- ▶ Machine Type 8561 (M/T 8561), Model T01, which is further identified as *IBM z15 Model T01*, or *z15 T01*, unless otherwise specified.
- ▶ Machine Type 8562 (M/T 8562), Model T02, which is further identified as *IBM z15 Model T02*, or *z15 T02*, unless otherwise specified.

In the remainder of this chapter, IBM z15 (z15) refers to both machine types.

This chapter also introduces the principles of cryptography and describes the implementation of cryptography in the hardware and software architecture of IBM Z. It also describes the features that IBM z15 offers. Finally, the chapter summarizes the cryptographic features and required software.

This chapter includes the following topics:

- ▶ 6.1, “Cryptography enhancements on IBM z15” on page 216
- ▶ 6.2, “Cryptography overview” on page 217
- ▶ 6.3, “Cryptography on IBM z15” on page 221
- ▶ 6.4, “CP Assist for Cryptographic Functions” on page 224
- ▶ 6.5, “Crypto Express7S” on page 230
- ▶ 6.6, “Trusted Key Entry workstation” on page 243
- ▶ 6.7, “Cryptographic functions comparison” on page 249
- ▶ 6.8, “Cryptographic operating system support for z15” on page 251

6.1 Cryptography enhancements on IBM z15

IBM z15 introduced the new PCI Crypto Express7S feature, together with a further improved CPACF Coprocessor, that can be managed by a new Trusted Key Entry (TKE) workstation. In addition, the IBM Common Cryptographic Architecture (CCA) and the IBM Enterprise PKCS #11 (EP11) Licensed Internal Code (LIC) were enhanced.

The functions support new standards and are designed to meet the following compliance requirements:

- ▶ Payment Card Industry (PCI) Hardware Security Module (HSM) certification to strengthen the cryptographic standards for attack resistance in the payment card systems area.
PCI HSM certification is exclusive for Crypto Express7S and Crypto Express6S.
- ▶ National Institute of Standards and Technology (NIST) through the Federal Information Processing Standard (FIPS) standard to implement guidance requirements.
- ▶ Common Criteria EP11 EAL4.
- ▶ German Banking Industry Commission (GBIC).
- ▶ Visa Format Preserving Encryption (VFPE) for credit card numbers.
- ▶ Enhanced public key Elliptic Curve Cryptography (ECC) for users such as Chrome, Firefox, and Apple's iMessage.
- ▶ Accredited Standards Committee X9 Inc Technical Report-34 (ASC X9 TR-34)

These enhancements are described in this chapter.

IBM z15 includes standard cryptographic hardware and optional cryptographic features for flexibility and growth capability. IBM has a long history of providing hardware cryptographic solutions. This history stretches from the development of the Data Encryption Standard (DES) in the 1970s to the Crypto Express tamper-sensing and tamper-responding programmable features.

Crypto Express is designed to meet the US Government's highest security rating of FIPS 140-2 Level 4¹. It also meets several other security ratings, such as the Common Criteria for Information Technology Security Evaluation, the PCI HSM criteria, and the criteria for German Banking Industry Commission (formerly known as Deutsche Kreditwirtschaft evaluation).

The cryptographic functions include the full range of cryptographic operations that are necessary for local and global business and financial institution applications. User Defined Extensions (UDX) allow you to add custom cryptographic functions to the functions that z15 systems offer.

¹ FIPS 140-2 Security Requirements for Cryptographic Modules.

6.2 Cryptography overview

Throughout history, a need existed for secret communication between people that cannot be understood by outside parties.

Also, it is necessary to ensure that a message cannot be corrupted (message integrity), while ensuring that the sender and the receiver really are the persons who they claim to be. Over time, several methods were used to achieve these objectives, with more or less success. Many procedures and algorithms for encrypting and decrypting data were developed that are increasingly complicated and time-consuming.

6.2.1 Modern cryptography

With the development of computing technology, the encryption and decryption algorithms can be performed by computers, which enables the use of complicated mathematical algorithms. Most of these algorithms are based on the prime factorization of large numbers.

Cryptography is used to meet the following requirements:

- ▶ Data protection

The protection of data usually is the main concept that is associated with cryptography. Only authorized persons should be able to read the message or get information about it. Data is encrypted by using a known algorithm and secret keys, such that the intended party can de-scramble the data, but an interloper cannot. This concept is also referred to as *confidentiality*.

- ▶ Authentication (identity validation)

This process decides whether the communication partners are who they claim to be, which can be done by using certificates and signatures. It must be possible to clearly identify the owner of the data or the sender and the receiver of the message.

- ▶ Message (data) Integrity

The verification of data ensures that what was received is identical to what was sent. It must be proven that the data is complete and was not altered during the moment it was transmitted (by the sender) and the moment it was received (by the receiver).

- ▶ Non-repudiation

It must be impossible for the owner of the data or the sender of the message to deny authorship. Non-repudiation ensures that both sides of a communication know that the other side agreed to what was exchanged, and not someone else. This specification implies a legal liability and contractual obligation, which is the same as a signature on a contract.

These goals should all be possible without unacceptable overhead to the communication. The goal is to keep the system secure, manageable, and productive.

The basic data protection method is achieved by encrypting and decrypting the data, while hash algorithms, message authentication codes (MACs), digital signatures, and certificates are used for authentication, data integrity, and non-repudiation.

When encrypting a message, the sender transforms the clear text into a secret text. Doing so requires the following main elements:

- ▶ The *algorithm* is the mathematical or logical formula that is applied to the key and the clear text to deliver a ciphered result, or to take a ciphered text and deliver the original clear text.
- ▶ The *key* ensures that the result of the encrypting data transformation by the algorithm is only the same when the same key is used. That decryption of a ciphered message results only in the original clear message when the correct key is used. Therefore, the receiver of a ciphered message must know which algorithm and key must be used to decrypt the message.

6.2.2 Kerckhoffs' principle

In modern cryptography, the algorithm is published and known to everyone, whereas the keys are kept secret. This configuration corresponds to Kerckhoffs' principle, which is named after Auguste Kerckhoffs, a Dutch cryptographer, who formulated it in 1883:

“A system should not depend on secrecy, and it should be able to fall into the enemy's hands without disadvantage.”

In other words, the security of a cryptographic system should depend on the security of the key, so the key must be kept secret. Therefore, the secure management of keys is the primal task of modern cryptographic systems.

Adhering to Kerckhoffs' Principle is done for the following reasons:

- ▶ It is much more difficult to keep an algorithm secret than a key.
- ▶ It is harder to exchange a compromised algorithm than to exchange a compromised key.
- ▶ Secret algorithms can be reconstructed by reverse engineering software or hardware implementations.
- ▶ Errors in public algorithms can generally be found more easily, when many experts examine it.
- ▶ In history, most secret encryption methods proved to be weak and inadequate.
- ▶ When a secret encryption method is used, it is possible that a back door was built in.
- ▶ If an algorithm is public, many experts can form an opinion about it. Also, the method can be more thoroughly investigated for potential weaknesses and vulnerabilities.

6.2.3 Keys

The keys that are used for the cryptographic algorithms often are sequences of numbers and characters, but can also be any other sequence of bits. The length of a key influences the security (strength) of the cryptographic method. The longer the used key, the more difficult it is to compromise a cryptographic algorithm.

For example, the DES (symmetric key) algorithm uses keys with a length of 56 bits, Triple-DES (TDES) uses keys with a length of 112 bits, and Advanced Encryption Standard (AES) uses keys of 128, 192, 256, or 512 bits. The asymmetric key RSA algorithm (named after its inventors Rivest, Shamir, and Adleman) uses keys with a length of 1024 - 4096 bits.

In modern cryptography, keys must be kept secret. Depending on the effort that is made to protect the key, keys are classified into the following levels:

- ▶ A *clear key* is a key that is transferred from the application in clear text to the cryptographic function. The key value is stored in the clear (at least briefly) somewhere in unprotected memory areas. Therefore, the key can be made available to someone under certain circumstances who is accessing this memory area.

This risk must be considered when clear keys are used. However, many applications exist where this risk can be accepted. For example, the transaction security for the (widely used) encryption methods Secure Sockets Layer (SSL) and Transport Layer Security (TLS) is based on clear keys.

- ▶ The value of a *protected key* is stored only in clear in memory areas that cannot be read by applications or users. The key value does not exist outside of the physical hardware, although the hardware might not be tamper-resistant. The principle of protected keys is unique to IBM Z. For more information, see 6.4.2, “CPACF protected key” on page 227.
- ▶ For a *secure key*, the key value does not exist in clear format outside of a special hardware device (HSM), which must be secured and tamper-resistant. A secure key is protected from disclosure and misuse, and can be used for the trusted execution of cryptographic algorithms on highly sensitive data. If used and stored outside of the HSM, a secure key must be encrypted with a *master key*, which is created within the HSM and never leaves the HSM.

Because a secure key must be handled in a special hardware device, the use of secure keys usually is far slower than the use of clear keys, as shown in Figure 6-1.

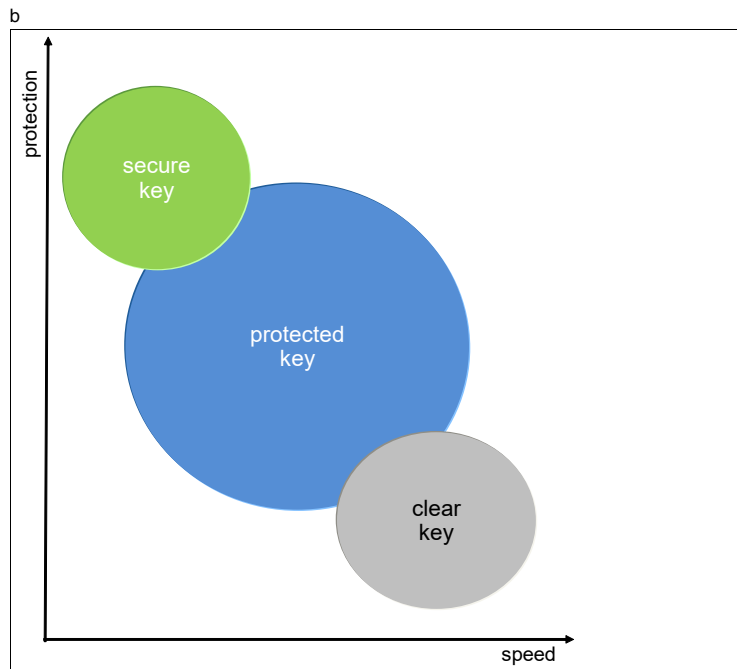


Figure 6-1 Three levels of protection with three levels of speed

6.2.4 Algorithms

The following algorithms of modern cryptography are differentiated based on whether they use the same key for the encryption of the message as for the decryption:

- ▶ *Symmetric algorithms* use the same key to encrypt and to decrypt data. The function that is used to decrypt the data is the opposite of the function that is used to encrypt the data. Because the same key is used on both sides of an operation, it must be negotiated between both parties and kept secret. Therefore, symmetric algorithms are also known as *secret key algorithms*.

The main advantage of symmetric algorithms is that they are fast and therefore can be used for large amounts of data, even if they are not run on specialized hardware. The disadvantage is that the key must be known by both sender and receiver of the messages, which implies that the key must be exchanged between them. This key exchange is a weak point that can be attacked.

Prominent examples for symmetric algorithms are DES, TDES, and AES.

- ▶ *Asymmetric algorithms* use two distinct but related keys: the *public key* and the *private key*. As the names imply, the private key must be kept secret, whereas the public key is shown to everyone. However, with asymmetric cryptography, it is not important who sees or knows the public key. Whatever is done with one key can be undone by the other key only.

For example, data that is encrypted by the public key can be decrypted by the associated private key only, and vice versa. Unlike symmetric algorithms, which use distinct functions for encryption and decryption, only one function is used in asymmetric algorithms.

Depending on the values that are passed to this function, it encrypts or decrypts the data. Asymmetric algorithms are also known as *public key algorithms*.

Asymmetric algorithms use complex calculations and are relatively slow (about 100 - 1000 times slower than symmetric algorithms). Therefore, such algorithms are not used for the encryption of bulk data.

Because the private key is never exchanged between the parties in communication, they are less vulnerable than symmetric algorithms. Asymmetric algorithms mainly are used for authentication, digital signatures, and for the encryption and exchange of secret keys, which in turn are used to encrypt bulk data with a symmetric algorithm.

Examples for asymmetric algorithms are RSA and the elliptic curve algorithms.

- ▶ *One-way algorithms* are not cryptographic functions. They do not use keys, and they can scramble data only, not de-scramble it. These algorithms are used extensively within cryptographic procedures for digital signing and tend to be developed and governed by using the same principles as cryptographic algorithms. One-way algorithms are also known as *hash algorithms*.

The most prominent one-way algorithms are the Secure Hash Algorithms (SHA).

6.3 Cryptography on IBM z15

In principle, cryptographic algorithms can run on processor hardware. However, these workloads are compute-intensive, and the handling of secure keys also requires special hardware protection. Therefore, IBM Z offer several cryptographic hardware features, which are specialized to meet the requirements for cryptographic workload.

The cryptographic hardware that is supported on IBM z15 is shown in Figure 6-2. These features are described in this chapter.

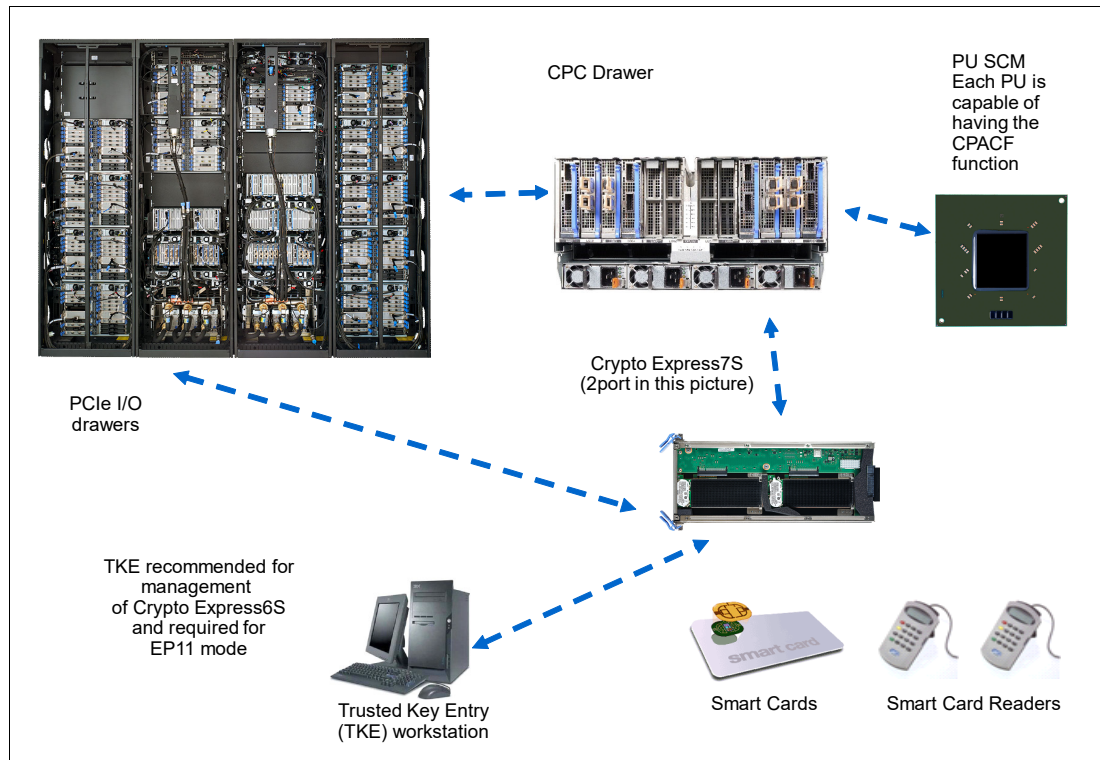


Figure 6-2 Cryptographic hardware that is supported in IBM z15

Implemented in every processor unit (PU) or core in a central processor complex (CPC) is a cryptographic coprocessor that can be used² for cryptographic algorithms that uses clear keys or protected keys. For more information, see 6.4, “CP Assist for Cryptographic Functions” on page 224.

The Crypto Express7S adapter is an HSM that is placed in the PCIe+ I/O drawer of z15. It also supports cryptographic algorithms by using secret keys. For more information, see 6.5, “Crypto Express7S” on page 230.

Finally, a TKE workstation is required for entering keys in a secure way into the Crypto Express7S HSM, which often also is equipped with smart card readers. For more information, see 6.6, “Trusted Key Entry workstation” on page 243.

² CPACF enablement feature must be ordered (FC 3863).

The feature codes and purpose of the cryptographic hardware features that are available for IBM z15 are listed in Table 6-1.

Table 6-1 Cryptographic features for IBM z15 T01

| Feature code | Description |
|--------------|---|
| 3863 | CP Assist for Cryptographic Function (CPACF) enablement This feature is a prerequisite to use CPACF (except for SHA-1, SHA-224, SHA-256, SHA-384, and SHA-512) and the PCIe Crypto Express features. |
| 0898 | Crypto Express7S feature (2-port) ^a These features are optional. The 2-port feature contains two IBM 4769 PCIe Cryptographic Coprocessors (HSMs), which can be independently defined as Coprocessor or Accelerator. New feature. Not supported on previous generations Z systems. |
| 0899 | Crypto Express7S feature (1-port) ^a These features are optional. The 1-port feature contains one IBM 4769 PCIe Cryptographic Coprocessor (HSM), which can be defined as Coprocessor or Accelerator. New feature. Not supported on previous generations Z systems |
| 0893 | Crypto Express6S adapter ^a This feature is available as a carry forward MES from z14. This feature is optional. Each feature one IBM 4768 PCIe Cryptographic Coprocessor (HSM). This feature is supported in z15 and z14. |
| 0890 | Crypto Express5S adapter ^a This feature is available as a carry forward MES from z14 or z13. This feature is optional and each feature of which contains one IBM 4767 PCIe Cryptographic Coprocessor (HSM). This feature is supported in z15, z14, z13, and z13s systems. |
| 0088 | TKE tower workstation A TKE provides basic key management (key identification, exchange, separation, update, and backup) and security administration. It is optional for running a Crypto Express feature in CCA mode in non PCI-compliant environment. It is required for running in EP11 mode and CCA mode with full PCI compliance. The TKE workstation has one 1000BASE-T Ethernet port, and supports connectivity to an Ethernet local area network (LAN). Up to 10 features combined with 0087 per z15 can be ordered. |
| 0087 | TKE rack-mounted workstation The rack-mounted version of the TKE, which needs a customer-provided standard 19-inch rack. It features a 1u TKE unit and an (optional) 1u console tray (screen, keyboard, and pointing device). When smart card readers are used, another customer-provided tray is needed. Up to 10 features combined with 0088 per z15 can be ordered. |
| 0881 | TKE 9.2 Licensed Internal Code (LIC) Included with the TKE tower workstation FC 0088 and the TKE rack-mounted workstation FC 0087 for z15. Earlier versions of TKE features (feature codes: 0080, 0081, 0085, and 0086) can also be upgraded to TKE 9.2 LIC, if the TKE is assigned to a z14 or later. |

| Feature code | Description |
|--------------|---|
| 0891 | TKE Smart Card Reader Access to information in the smart card is protected by a PIN. One feature code includes two smart card readers, two cables to connect to the TKE workstation, and 20 smart cards. |
| 0900 | New TKE smart cards This card allows the TKE to support zones with EC 521 key strength (EC 521 strength for Logon Keys, Authority Signature Keys, and EP11 signature keys). |
| 0892 | More TKE smart cards When one feature code is ordered, 10 smart cards are included. The order increment is 1 - 99 (990 blank smart cards). |

- a. The maximum number of combined features of all types cannot exceed 60 HSMs on a z15 T01. This means the maximum number for feature code 0898 is 30, for all other (single HSM) types is 16 when installed exclusively.

A TKE includes support for the AES encryption algorithm with 256-bit master keys and key management functions to load or generate master keys to the cryptographic coprocessor.

If the TKE workstation is chosen to operate the Crypto Express7S adapter in a z15, TKE workstation with the TKE 9.2 LIC is required. For more information, see 6.6, “Trusted Key Entry workstation” on page 243.

Important: Products that include any of the cryptographic feature codes contain cryptographic functions that are subject to special export licensing requirements by the United States Department of Commerce. It is your responsibility to understand and adhere to these regulations when you are moving, selling, or transferring these products.

To access and use the cryptographic hardware devices that are provided by z15, the application must use an application programming interface (API) that is provided by the operating system. In z/OS, the Integrated Cryptographic Service Facility (ICSF) provides the APIs and is managing the access to the cryptographic devices, as shown in Figure 6-3 on page 224.

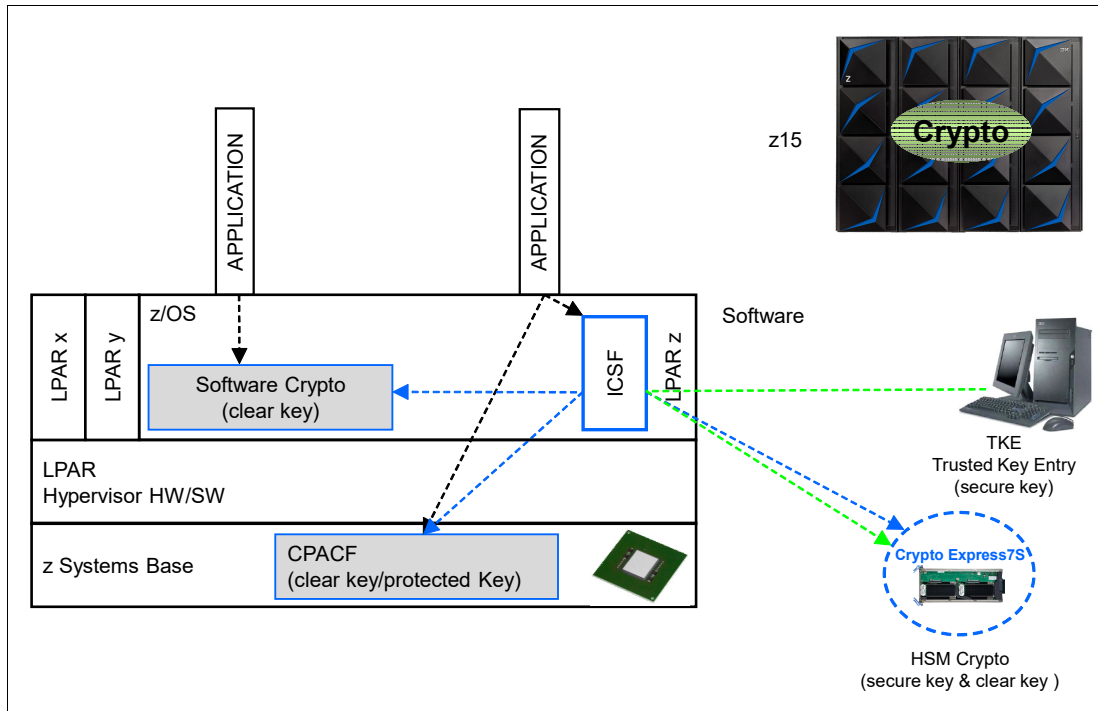


Figure 6-3 z15 Cryptographic Support in z/OS

ICSF is a software component of z/OS. ICSF works with the hardware cryptographic features and the Security Server (IBM Resource Access Control Facility [IBM RACF®] element) to provide secure, high-speed cryptographic services in the z/OS environment. ICSF provides the APIs by which applications request the cryptographic services, and from the CPACF and the Crypto Express features.

ICSF transparently routes application requests for cryptographic services to one of the integrated cryptographic engines (CPACF or a Crypto Express feature), depending on performance or requested cryptographic function. ICSF is also the means by which the secure Crypto Express features are loaded with master key values, which allows the hardware features to be used by applications.

The cryptographic hardware that is installed in z15 determines the cryptographic features and services that are available to the applications.

The users of the cryptographic services call the ICSF API. Some functions are performed by the ICSF software without starting the cryptographic hardware features. Other functions result in ICSF going into routines that contain proprietary IBM Z crypto instructions. These instructions are run by a CPU engine and result in a work request that is generated for a cryptographic hardware feature.

6.4 CP Assist for Cryptographic Functions

Attached to every PU (core) of a z15 system are two independent engines, one for compression and one for cryptographic functions, as shown in Figure 6-4 on page 225.

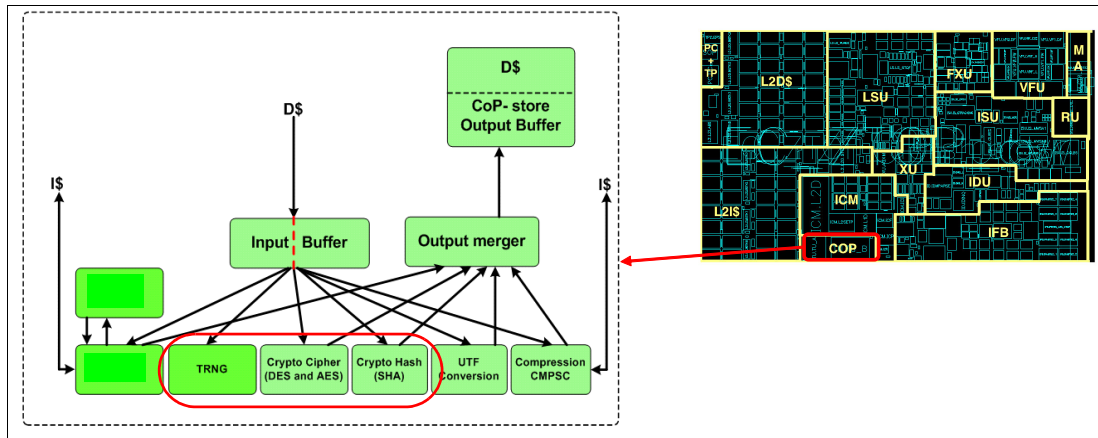


Figure 6-4 The cryptographic coprocessor CPACF

This cryptographic coprocessor, which is known as the CPACF, is not qualified as an HSM; therefore, it is not suitable for handling algorithms that use secret keys. However, the coprocessor can be used for cryptographic algorithms that use clear keys or protected keys. The CPACF works synchronously with the PU, which means that the owning processor is busy when its coprocessor is busy. This setup provides a fast device for cryptographic services.

CPACF supports pervasive encryption. Simple policy controls allow businesses to enable encryption to protect data in mission-critical databases without the need to stop the database or re-create database objects. Pervasive encryption includes z/OS Dataset Encryption, z/OS Coupling Facility Encryption, z/VM encrypted hypervisor paging, and z/TPF transparent database encryption, which use performance enhancements in the hardware.

The CPACF offers a set of symmetric cryptographic functions that enhances the encryption and decryption performance of clear key operations. These functions are for SSL, virtual private network (VPN), and data-storing applications that do not require FIPS 140-2 Level 4 security.

CPACF is designed to facilitate the privacy of cryptographic key material when used for data encryption through key wrapping implementation. It ensures that key material is not visible to applications or operating systems during encryption operations. For more information, see 6.4.2, “CPACF protected key” on page 227.

The CPACF feature provides hardware acceleration for the following cryptographic services:

- ▶ DES
- ▶ Triple-DES
- ▶ AES-128
- ▶ AES-192
- ▶ AES-256 (all for clear and protected keys)
- ▶ SHA-1
- ▶ SHA-256 (SHA-2 or SHA-3 standard)
- ▶ SHA-384 (SHA-2 or SHA-3 standard)
- ▶ SHA-512 (SHA-2 or SHA-3 standard)
- ▶ SHAKE-128
- ▶ SHAKE-256
- ▶ PRNG
- ▶ DRNG
- ▶ TRNG

It provides high-performance hardware encryption, decryption, hashing, and random number generation support. The following instructions support the cryptographic assist function:

- ▶ KMAC: Compute Message Authentic Code
- ▶ KM: Cipher Message
- ▶ KMC: Cipher Message with Chaining
- ▶ KMF: Cipher Message with CFB
- ▶ KMCTR: Cipher Message with Counter
- ▶ KMO: Cipher Message with OFB
- ▶ KIMD: Compute Intermediate Message Digest
- ▶ KLMD: Compute Last Message Digest
- ▶ PCKMO: Provide Cryptographic Key Management Operation

These functions are provided as problem-state *z/Architecture* instructions that are directly available to application programs. These instructions are known as Message-Security Assist (MSA). When enabled, the CPACF runs at processor speed for every CP, IFL, and zIIP. For more information about MSA instructions, see *z/Architecture Principles of Operation*, SA22-7832.

For activating these functions, the CPACF must be enabled by using feature code (FC) 3863, which is available for no extra charge. Support for hashing algorithms SHA-1, SHA-256, SHA-384, and SHA-512 is always enabled.

6.4.1 Cryptographic synchronous functions

Because the CPACF works synchronously with the PU, it provides cryptographic synchronous functions. For IBM and client-written programs, CPACF functions can be started by using the MSA instructions. *z/OS ICSF callable services on z/OS*, in-kernel crypto APIs, and a *libica* cryptographic functions library that is running on Linux on Z can also start CPACF synchronous functions.

The CPACF coprocessor in z14 was redesigned for improved performance compared to the z13 and was further improved in z15, depending on the function that is being used. The following tools might benefit from the throughput improvements:

- ▶ Db2/IMS encryption tool
- ▶ Db2 built-in encryption
- ▶ *z/OS Communication Server: IPsec/IKE/AT-TLS*
- ▶ *z/OS System SSL*
- ▶ *z/OS Network Authentication Service (Kerberos)*
- ▶ DFDSS Volume encryption
- ▶ *z/OS Java SDK*
- ▶ *z/OS Encryption Facility*
- ▶ Linux on Z: Kernel, openSSL, openCryptoki, and GSKIT

The z15 hardware includes the implementation of algorithms as hardware synchronous operations. This configuration holds the PU processing of the instruction flow until the operation completes.

z15 offers the following synchronous functions:

- ▶ Data encryption and decryption algorithms for data privacy and confidentiality:
 - Data Encryption Standard (DES):
 - Single-length key DES
 - Double-length key DES
 - Triple-length key DES (also known as Triple-DES)

- Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys
- ▶ Hashing algorithms for data integrity, such as SHA-1 and SHA-2. New for z14 ZR1 is SHA-3 support for SHA-224, SHA-256, SHA-384, and SHA-512 and the two extendable output functions as described by the standard SHAKE-128 and SHAKE-256.
- ▶ Message authentication code (MAC):
 - Single-length key MAC
 - Double-length key MAC
- ▶ Pseudo-Random Number Generator (PRNG), Deterministic Random Number Generation (DRNG), and True Random Number Generation (TRNG) for cryptographic key generation.
- ▶ Galois Counter Mode (GCM) encryption, which is enabled by a single hardware instruction.

For the SHA hashing algorithms and the random number generation algorithms, only clear keys are used. For the symmetric encryption and decryption DES and AES algorithms and clear keys, protected keys can also be used. Protected keys require a Crypto Express7S, Crypto Express6S, or a Crypto Express5S adapter that is running in CCA mode. For more information, see 6.5.2, “Crypto Express7S as a CCA coprocessor” on page 233.

The hashing algorithms SHA-1, SHA-2, and SHA-3 support for SHA-224, SHA-256, SHA-384, and SHA-512, are enabled on all systems and do not require the CPACF enablement feature. For all other algorithms, the no-charge CPACF enablement feature (FC 3863) is required.

The CPACF functions are implemented as processor instructions and require operating system support for use. Operating systems that use the CPACF instructions include z/OS, z/VM, z/VSE, z/TPF, and Linux on Z.

6.4.2 CPACF protected key

z15 supports the protected key implementation. Since PCIXCC³ deployment, secure keys are processed on the PCI-X and PCIe adapters. This process requires an asynchronous operation to move the data and keys from the general-purpose central processor (CP) to the crypto adapters.

Clear keys process faster than secure keys because the process is done synchronously on the CPACF. Protected keys blend the security of Crypto Express7S, Crypto Express6S, or Crypto Express5S coprocessors and the performance characteristics of the CPACF. This process allows it to run closer to the speed of clear keys.

CPACF facilitates the continued privacy of cryptographic key material when used for data encryption. In Crypto Express7S, Crypto Express6S, or Express5S coprocessors, a secure key is encrypted under a master key. However, a protected key is encrypted under a wrapping key that is unique to each LPAR.

Because the wrapping key is unique to each LPAR, a protected key cannot be shared with another LPAR. By using key wrapping, CPACF ensures that key material is not visible to applications or operating systems during encryption operations.

³ IBM 4764 PCI-X cryptographic coprocessor.

CPACF code generates the wrapping key and stores it in the protected area of the hardware system area (HSA). The wrapping key is accessible only by firmware. It cannot be accessed by operating systems or applications. DES/T-DES and AES algorithms are implemented in CPACF code with the support of hardware assist functions. Two variations of wrapping keys are generated: one for DES/T-DES keys and another for AES keys.

Wrapping keys are generated during the clear reset each time an LPAR is activated or reset. No customizable option is available at Support Element (SE) or Hardware Management Console (HMC) that permits or avoids the wrapping key generation. This function flow for the Crypto Express7S and Crypto Express6S adapters is shown in Figure 6-5.

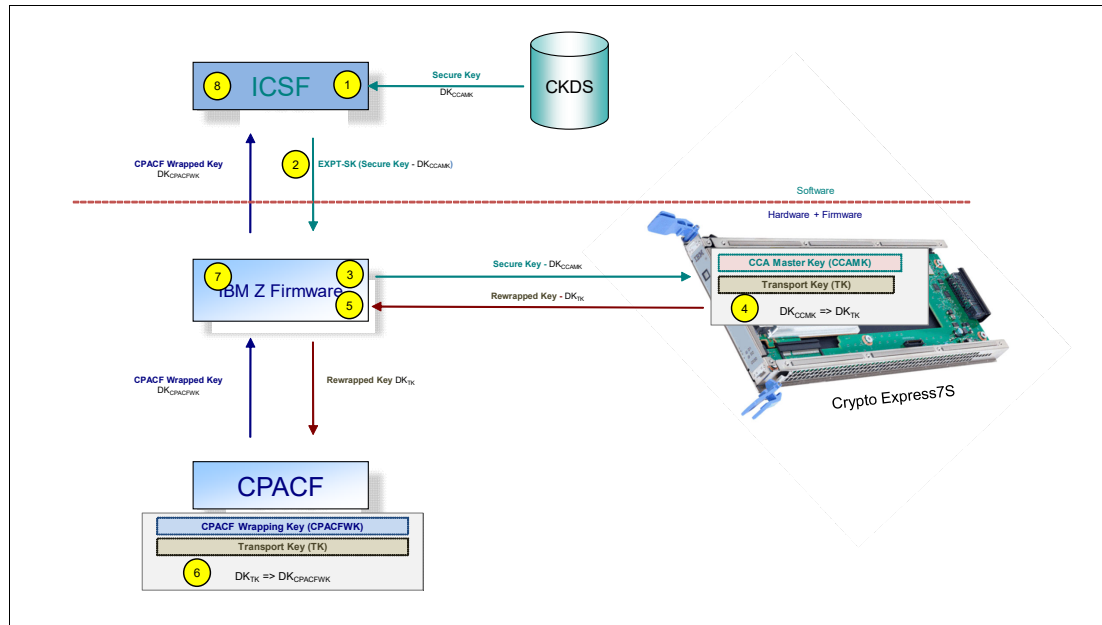


Figure 6-5 CPACF key wrapping for Crypto Express7S and Crypto Express6S

The key wrapping for Crypto Express5S is similar to Crypto Express7S or Crypto Express6S; however, the Data Key that is exchanged between the Crypto Express5S and the CPACF is not wrapped by way of a Transport Key.

The CPACF Wrapping Key and the Transport Key for use with Crypto Express7S or Crypto Express6S are in a protected area of the HSA that is not visible to operating systems or applications.

The function flow for Crypto Express5S is shown in Figure 6-6.

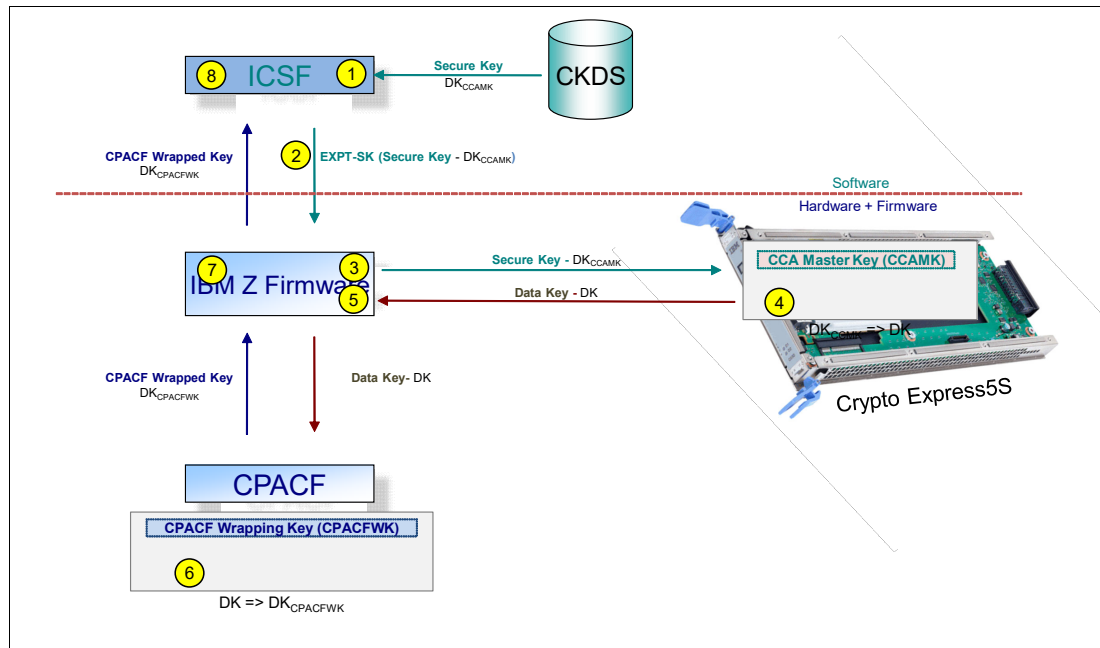


Figure 6-6 CPACF key wrapping for Crypto Express5S

If a Crypto ExpressxS coprocessor (CEX7C, CEX6C, or CEX5C) is available, a protected key can begin its life as a secure key. Otherwise, an application is responsible for creating or loading a clear key value, and then uses the PCKMO instruction to wrap the key. ICSF is not called by the application if the CEXxC is not available.

A new segment in the profiles of the CSFKEYS class in IBM RACF restricts which secure keys can be used as protected keys. By default, all secure keys are considered not eligible to be used as protected keys. The process that is shown in Figure 6-5 on page 228 considers a secure key as the source of a protected key.

The source key in this case is stored in the ICSF Cryptographic Key Data Set (CKDS) as a secure key, which was encrypted under the master key. This secure key is sent to CEX7C, CEX6C, or CEX5C to be deciphered and then, sent to the CPACF in clear text. At the CPACF, the key is wrapped under the LPAR wrapping key, and is then returned to ICSF. After the key is wrapped, ICSF can keep the protected value in memory. It then passes it to the CPACF, where the key is unwrapped for each encryption or decryption operation.

The protected key is designed to provide substantial throughput improvements for a large volume of data encryption and low latency for encryption of small blocks of data. A high-performance secure key solution, also known as a *protected key solution*, requires the ICSF HCR7770 as a minimum release.

6.5 Crypto Express7S

The Crypto Express7S feature (FC 0898 or FC 0899) is an optional feature that is exclusive to z15 systems. Each feature FC 0898 has one IBM 4769 PCIe cryptographic adapter (hardware security module - HSM), whereas FC 0899 has two IBM 4768 PCIe cryptographic adapters (two HSMs). The Crypto Express7S (CEX7S) feature occupies one I/O slot in a z15 PCIe+ I/O drawer. This feature provides one or two HSMs and for a secure programming and hardware environment on which crypto processes are run.

Each cryptographic coprocessor includes a general-purpose processor, non-volatile storage, and specialized cryptographic electronics. The Crypto Express7S feature provides tamper-sensing and tamper-responding, high-performance cryptographic operations.

Each Crypto Express7S PCI Express adapter (HSM) is available in one of the following configurations:

- ▶ Secure IBM CCA coprocessor (CEX7C) for FIPS 140-2 Level 4 certification. This configuration includes secure key functions. It is optionally programmable to deploy more functions and algorithms by using UDX. For more information, see 6.5.2, “Crypto Express7S as a CCA coprocessor” on page 233.

A TKE workstation is required to support the administration of the Crypto Express7S when it is configured in CCA mode when in full PCI-compliant mode for the necessary certificate management in this mode. The TKE is optional in all other use cases for CCA.

- ▶ Secure IBM Enterprise PKCS #11 (EP11) coprocessor (CEX7P) implements an industry-standardized set of services that adheres to the PKCS #11 specification V2.20 and more recent amendments. It was designed for extended FIPS and Common Criteria evaluations to meet public sector requirements. This new cryptographic coprocessor mode introduced the PKCS #11 secure key function. For more information, see 6.5.3, “Crypto Express7S as an EP11 coprocessor” on page 239.

A TKE workstation is always required to support the administration of the Crypto Express7S when it is configured in EP11 mode.

- ▶ Accelerator (CEX7A) for acceleration of public key and private key cryptographic operations that are used with SSL/TLS processing. For more information, see 6.5.4, “Crypto Express7S as an accelerator” on page 240.

These modes can be configured by using the SE. The PCIe adapter must be configured offline to change the mode.

Attention: Switching between configuration modes erases all adapter secrets. The exception is when you are switching from Secure CCA to accelerator, and vice versa.

The Crypto Express7S feature is released for enhanced cryptographic performance. Clients who migrated to variable-length AES key tokens cannot take advantage of faster encryption speeds by using CPACF. Support is being added to translate a secure variable-length AES CIPHER token to a protected key token (protected by the system wrapping key). This support allows for faster AES encryption speeds when variable-length tokens are used while maintaining strong levels of security.

The Crypto Express7S feature does not include external ports and does not use optical fiber or other cables. It does not use channel path identifiers (CHPIDs), but requires one slot in the PCIe I/O drawer and one physical channel ID (PCHID) for each PCIe cryptographic adapter. Removal of the feature or adapter *zeroizes* its content. Access to the PCIe cryptographic adapter is controlled through the setup in the image profiles on the SE.

Adapter: Although PCIe cryptographic adapters include no CHPID type and are not identified as external channels, all logical partitions (LPARs) in all channel subsystems can access to the adapter. In z15 systems, up to 85 LPARs are available per adapter (HSM). Accessing the adapter requires a setup in the image profile for each partition. The adapter must be in the candidate list.

Each z15 T01 supports up to 60 HSMs in total (combination of Crypto Express7S (2port) Crypto Express7S (1 port), Crypto Express6S, and Crypto Express5S). Crypto Express6S and Crypto Express5S features have one HSM (port) and *are not orderable* for a new build z15 T01 system, but can be carried forward from a z14 or z13 by using an MES. Configuration information for Crypto Express7S is listed in Table 6-2.

Table 6-2 Crypto Express7S features

| Feature | Quantity |
|--|-----------------|
| Minimum number of orderable features 0898 for z15 T01 | 2 |
| Minimum number of orderable features 0899 for z15 ^a T01 | 2 |
| Order increment (above two features for features 0898 and 0899) | 1 |
| Maximum number of HSMs for z15 (combining all CEX7S,CEX6S, and CEX5S) | 60 ^b |
| Number of PCIe cryptographic adapters for each feature 0898 (coprocessor or accelerator) | 2 |
| Number of PCIe cryptographic adapters for each feature 0899 (coprocessor or accelerator) | 1 |
| Number of cryptographic domains at z15 for each PCIe adapter ^c | 85 |

- a. The minimum initial order of Crypto Express7S feature 0899 is two. After the initial order, more Crypto Express7S features 0899 can be ordered one feature individually.
- b. Crypto Express7S (2 port) has two hardware security modules (HSMs) per feature. The HSM is one IBM 4769 PCIe Cryptographic Coprocessor (PCIeCC). The max. number of HSMs per T01 system, combining all cryptographic features is 60, while the max. number of single HSM (port) cryptographic features is 16 (CEX7S (1 port), CEX6S, and CEX5S)
- c. More than one partition, which is defined to the same channel subsystem (CSS) or to different CSSs, can use the same domain number when assigned to different PCIe cryptographic adapters.

The concept of *dedicated processor* does not apply to the PCIe cryptographic adapter. Whether configured as a coprocessor or an accelerator, the PCIe cryptographic adapter is made available to an LPAR. It is made available as directed by the domain assignment and the candidate list in the LPAR image profile. This availability is not changed by the shared or dedicated status that is given to the PUs in the partition.

When installed non-concurrently, Crypto Express7S features are assigned PCIe cryptographic adapter numbers sequentially during the power-on reset (POR) that follows the installation. When a Crypto Express7S feature is installed concurrently, the installation can select an out-of-sequence number from the unused range. When a Crypto Express7S (Crypto Express6S or Crypto Express5S) feature is removed concurrently, the PCIe adapter numbers are automatically freed.

The definition of domain indexes and PCIe cryptographic adapter numbers in the candidate list for each LPAR must be planned to allow for nondisruptive changes. Consider the following points:

- Operational changes can be made by using the Change LPAR Cryptographic Controls task from the SE, which reflects the cryptographic definitions in the image profile for the

partition. With this function, adding and removing the cryptographic feature without stopping a running operating system can be done dynamically.

- ▶ The same usage domain index can be defined more than once across multiple LPARs. However, the PCIe cryptographic adapter number that is coupled with the usage domain index that is specified must be unique across all active LPARs.

The same PCIe cryptographic adapter number and usage domain index combination can be defined for more than one LPAR (up to 85 for z15). For example, you might define a configuration for backup situations. However, only one of the LPARs can be active at a time.

For more information, see 6.5.5, “Managing Crypto Express7S” on page 240.

6.5.1 Cryptographic asynchronous functions

The optional PCIe cryptographic coprocessors Crypto Express7S provides asynchronous cryptographic functions to z15. Over 300 Cryptographic algorithms and modes are supported, including the following algorithms and modes:

- ▶ DES/TDES w DES/TDES MAC/CMAC: The Data Encryption Standard is a widespread symmetrical encryption algorithm. DES, along with its double-length and triple length variations, TDES today are considered to be not sufficient secure for many applications. They were replaced by the AES as the official US standard, but it is still used in the industry with the MAC and the Cipher-based Message Authentication Code (CMAC) for verifying the integrity of messages.
- ▶ AES, AESKW, AES GMAC, AES GCM, AES XTS, AES CIPHER mode, and CMAC: AES replaced DES as the official US standard in October 2000. The enhanced standards for AES Key Warp (AESKW), the AES Galois Message Authentication Code (AES GMAC) and Galois/Counter Mode (AES GCM), the XEX-based tweaked-codebook mode with ciphertext stealing (AES XTS), and CMAC are supported.
- ▶ MD5, SHA-1, SHA-2, or SHA-3⁴ (224, 256, 384, and 512), and HMAC: The Secure Hash Algorithm (SHA-1 and the enhanced SHA-2 or SHA-3 for different block sizes), the older message-digest (MD5) algorithm, and the advanced keyed-hash method authentication code (HMAC) are used for verifying the data integrity and the authentication of a message.
- ▶ VFPE: A method of encryption in which the resulting cipher text features the same form as the input clear text, which is developed for use with credit cards.
- ▶ RSA (512, 1024, 2048, and 4096): RSA was published in 1977. It is widely used asymmetric public-key algorithm, which means that the encryption key is public whereas the decryption key is kept secret. It is based on the difficulty of factoring the product of two large prime numbers. The number describes the length of the keys.
- ▶ ECDSA (192, 224, 256, 384, and 521 Prime/NIST): ECC is a family of asymmetric cryptographic algorithms that are based on the algebraic structure of elliptic curves. ECC can be used for encryption, pseudo-random number generation, and digital certificates. The Elliptic Curve Digital Signature Algorithm (ECDSA) Prime/NIST method is used for ECC digital signatures, which are recommended for government use by NIST.
- ▶ ECDSA (160, 192, 224, 256, 320, 384, and 512 BrainPool): ECC Brainpool is a workgroup of companies and institutions that collaborate on developing ECC algorithms. The ECDSA algorithms that are recommended by this group are supported.
- ▶ ECDH (192, 224, 256, 384, and 521 Prime/NIST): Elliptic Curve Diffie-Hellman (ECDH) is an asymmetric protocol that is used for key agreement between two parties by using ECC-based private keys. The recommendations by NIST are supported.

⁴ SHA-3 was standardized by NIST in 2015. SHA-2 is still acceptable and no indication exists that SHA-2 is vulnerable or that SHA-3 is more or less vulnerable than SHA-2.

- ▶ ECDH (160, 192, 224, 256, 320, 384, and 512 BrainPool): ECDH according to the Brainpool recommendations.
- ▶ Montgomery Modular Math Engine: The Montgomery Modular Math Engine is a method for fast modular multiplication. Many crypto systems, such as RSA and Diffie-Hellman key Exchange, can use this method.
- ▶ Random Number Generator (RNG): The generation of random numbers for cryptographic key generation is supported.
- ▶ Prime Number Generator (PNG): The generation of prime numbers is also supported.
- ▶ Clear Key Fast Path (Symmetric and Asymmetric): This mode of operation gives a direct hardware path to the cryptographic engine and provides high performance for public-key cryptographic functions.

Several of these algorithms require a secure key and must run on an HSM. Some of these algorithms can also run with a clear key on the CPACF. Many standards are supported only when Crypto Express7S is running in CCA mode. Others are supported only when the adapter is running in EP11 mode.

The three modes for Crypto Express6S are described next. For more information, see 6.7, “Cryptographic functions comparison” on page 249.

6.5.2 Crypto Express7S as a CCA coprocessor

A Crypto Express7S adapter that is running in CCA mode supports IBM CCA. CCA is an architecture and a set of APIs. It provides cryptographic algorithms, secure key management, and many special functions that are required for banking. Over 129 APIs with more than 600 options are provided, with new functions and algorithms always being added.

The IBM CCA provides functions for the following tasks:

- ▶ Encryption of data (DES/TDES/AES)
- ▶ Key management:
 - Using TDES or AES keys
 - Using RSA or Elliptic Curve keys
- ▶ Message authentication for MAC/HMAC/AES-CMAC
- ▶ Key generation
- ▶ Digital signatures
- ▶ Random number generation
- ▶ Hashing (SHA, MD5, and others)
- ▶ ATM PIN generation and processing
- ▶ Credit card transaction processing
- ▶ Visa Data Secure Platform (DSP) Point to Point Encryption (P2PE)
- ▶ Europay, MasterCard, and Visa (EMV) card transaction processing
- ▶ Card personalization
- ▶ Other financial transaction processing
- ▶ Integrated role-based access control system
- ▶ Compliance support for:
 - All DES services

- AES services
 - RSA services, including full use of X.509 certificates
- ▶ TR-34 Remote Key Load

User-defined extensions support

User-defined extension (UDX) allows a developer to add customized operations to IBM's CCA Support Program. UDXs to the CCA support customized operations that run within the Crypto Express features when defined as a coprocessor.

UDX is supported under a special contract through an IBM or approved third-party service offering. The Crypto Cards website directs your request to an IBM Global Services location for your geographic location. A special contract is negotiated between IBM Global Services and you for the development of the UDX code by IBM Global Services according to your specifications and an agreed-upon level of the UDX.

A UDX toolkit for IBM Z is tied to specific versions of the CCA code and the related host code. UDX is available for the Crypto Express7S and Crypto Express6S (Secure IBM CCA coprocessor mode only) features. An UDX migration is no more disruptive than a normal Microcode Change Level (MCL) or ICSF release migration.

In z15, up to four UDX files can be imported. These files can be imported from a USB media stick or an FTP server. The UDX configuration window is updated to include a Reset to IBM Default button.

Consideration: CCA features a new code level starting with z13 systems, and the UDX clients require a new UDX.

On z15 T01, Crypto Express7S is delivered with CCA Level 6.3 firmware. A new set of cryptographic functions and callable services is provided by the IBM CCA LIC to enhance the functions that secure financial transactions and keys. The Crypto Express7S includes the following features:

- ▶ Greater than 16 domains support up to 85 LPARs on z15 T01.
- ▶ Payment Card Industry (PCI) PIN Transaction Security (PTS) HSM Certification that is exclusive to z15 in combination with CEX7S or CEX6S features, and z14 and CEX6S features.
- ▶ VFPE support, which was introduced with z13/z13s systems.
- ▶ AES PIN support for the German banking industry.
- ▶ PKA Translate UDX function into CCA.
- ▶ Verb Algorithm Currency.

CCA Version 7.1 improvements

- ▶ Supported curves:
 - NIST Prime Curves: P192, P224, P256, P384, P521
 - Brainpool Curves: 160, 192, 224, 256, 320, 384, 512
- ▶ Support in the CCA coprocessor for these Edwards curves:
 - ED25519 (128-bit security strength), ED448 (224-bit security strength)
 - ED25519 is faster but ED448 is more secure. Practically though, 128-bit security strength is very secure.
- ▶ Edwards curves are used for digitally signing documents and verifying those signatures
- ▶ Edwards curves are less susceptible to side channel attacks when compared to Prime and Brainpool curves

- ▶ **ECC Protected Keys**
 Crypto Express7S provides support in CCA coprocessors to take advantage of fast DES, AES data encryption speeds in CPACF while maintaining high levels of security for the secure key material. The key remains encrypted and the key encrypting key never appears in host storage.
 When using CCA ECC services, ICSF can now take advantage of ECC support in CPACF (protected key support) for these curves:
 - Prime: P256, P384, P521
 - Edwards: ED25519, ED448
 CPACF can achieve much faster crypto speeds compared to the coprocessor
 The translation to protected key happens automatically once the attribute is set in the key token. No application change is required.
- ▶ **New signatures**
 Support for the Cryptographic Suite for Algebraic Lattices signatures algorithm with the largest key sizes (MODE=3)
 - Public Key size: 1760 bytes
 - Private Key Size: 3856 bytes
 - Signature Size: 3366 bytes
 Lattice-based cryptographic keys will be protected by the 256-bit AES MK. The lattice-based key has a security strength of 128 bits.
- ▶ **TR-31 for Hash-based Message Authentication Code (HMAC)**
 HMAC keys are used to verify the integrity and authenticity of a message. This support provides a standard method of exchanging HMAC keys with a partner using symmetric key techniques. The key is exchanged in the standard TR-31 key block format which can be consumed by any crypto system supporting the standard

CCA Version 6.3 improvements⁵

- ▶ Compliance support for:
 - All DES services
 - AES services
 - RSA services, including full use of X.509 certificates
- ▶ TR-34 Remote Key Load

Greater than 16 domains support

z15 T01 supports up to 85 LPARs. The IBM Z crypto architecture was designed to support 16 domains, which matched the LPAR maximum at the time. Before z13 systems, crypto workload separation can be complex in customer environments where the number of LPARs was larger than 16. These customers mapped a large set of LPARs to a small set of crypto domains.

Starting with z14, the IBM Z crypto architecture can support up to 256 domains in an adjunct processor (AP) with the AP extended addressing (APXA) facility that is installed. As such, the Crypto Express adapters are enhanced to handle 256 domains. The IBM Z firmware provides up to 85 domains for z15 to customers (to match the current LPAR maximum). Customers can map individual LPARs to unique crypto domains or continue to share crypto domains across LPARs.

The following requirements must be met to support 85 domains:

- ▶ Hardware: z15 T01 and Crypto Express7S (or Crypto Express6S)

⁵ A TKE is required to manage a PCI-compliant coprocessor and for certificate management

- ▶ Operating systems:
 - z/OS all functions require ICSF WD17 (HCR77C1), unless otherwise noted. WD17 supports z/OS V2R1, V2R2, and V2R3.
 - z/VM Version 6.4 with PTFs or newer for guest use.

Payment Card Industry-HSM certification

Payment Card Industry (PCI) standards are developed to help ensure security in the PCI. PCI defines their standards as a set of security standards that is designed to ensure that all companies that accept, process, store, or transmit credit card information that is maintained a secure environment.

Compliance with the PCI-HSM standard is valuable for customers, particularly those customers who are in the banking and finance industry. This certification is important to clients for the following fundamental reasons:

- ▶ Compliance is increasingly becoming mandatory.
- ▶ The requirements in PCI-HSM make the system more secure.

Industry requirements for PCI-HSM compliance

The PCI organization cannot require compliance with its standards. Compliance with PCI standards is enforced by the payment card brands, such as Visa, MasterCard, American Express, JCB International, and Discover.

If you are a bank, acquirer, processor, or other participant in the payment card systems, the card brands can impose requirements on you if you want to process their cards. One set of requirements they are increasingly enforcing is the PCI standards.

The card brands work with PCI in developing these standards, and they focused first on the standards they considered most important, particularly the PCI Data Security Standard (PCI-DSS). Some of the other standards were written or required later, and PCI-HSM is one of the last standards to be developed. In addition, the standards themselves were increasing the strength of their requirements over time. Some requirements that were optional in earlier versions of the standards are now mandatory.

In general, the trend is for the card brands to enforce more of the PCI standards and to enforce them more rigorously. The trend in the standards is to impose more and stricter requirements in each successive version. The net result is that companies subject to these requirements can expect that they eventually must comply with all of the requirements.

Improved security through use of PCI-HSM

PCI-HSM was developed primarily to improve security in payment card systems. It imposes requirements in key management, HSM API functions, and device physical security. It also controls during manufacturing and delivery, device administration, and several other areas. It prohibits many things that were in common use for many years, but are no longer considered secure.

The result of these requirements is that applications and procedures often must be updated because they used some of the things that are now prohibited. Although this issue is inconvenient and imposes some costs, it does increase the resistance of the systems to attacks of various kinds. Updating a system to use PCI-HSM compliant HSMs is expected to reduce the risk of loss for the institution and its clients.

The following requirements must be met to use PCI-HSM:

- ▶ Hardware: z15⁶ systems and Crypto Express7S (or Crypto Express6S)

- ▶ Operating systems:
 - z/OS - ICSF WD18 (HCR77D0), unless otherwise noted. WD18 supports z/OS V2R1, V2R2, and V2R3; WD19 (HCR77D1 is required for z/OS V2R4.
 - z/VM Version 6.4 with PTFs or newer for guest use

Visa Format Preserving Encryption

VFPE refers to a method of encryption in which the resulting cipher text features the same form as the input clear text. The form of the text can vary according to use and application. One of the classic examples is a 16-digit credit card number. After VFPE is used to encrypt a credit card number, the resulting cipher text is another 16-digit number. This process helps older databases contain encrypted data of sensitive fields without having to restructure the database or applications.

VFPE allows customers to add encryption to their applications in such a way that the encrypted data can flow through their systems without requiring a massive redesign of their application. In our example, if the credit card number is VFPE-encrypted at the point of entry, the cipher text still behaves as a credit card number. It can flow through business logic until it meets a back-end transaction server that can VFPE-decrypt it to get the original credit card number to process the transaction.

Note: VFPE technology forms part of Visa, Inc.'s, Data Secure Platform (DSP). The use of this function requires a service agreement with Visa. You must maintain a valid service agreement with Visa when you use DSP/FPE.

The FPE features the following requirements:

- ▶ Hardware: z15 systems and Crypto Express 7S (or Crypto Express6S or Crypto Express5S with CCA V5.2 firmware).
- ▶ Operating systems:
 - z/OS: All functions require ICSF WD18 (HCR77D0), unless otherwise noted. WD18 supports z/OS V2R1 and later; WD19 (HCR77D1 is required for z/OS V2R4.
 - z/OS V2.1 and z/OS V1.13 with the Cryptographic Support for z/OS V1R13-z/OS V2R1 web deliverable (FMID HCR77B0).
 - z/VM Version 6.4 with PTFs or newer for guest use.

AES PIN support for the German banking industry

The German banking industry organization, DK, defined a new set of PIN processing functions to be used on the internal systems of banks and their servers. CCA is designed to support the functions that are essential to those parts of the German banking industry that are governed by DK requirements. The functions include key management support for new AES key types, AES key derivation support, and several DK-specific PIN and administrative functions.

This support includes PIN method APIs, PIN administration APIs, new key management verbs, and new access control points support that is needed for DK-defined functions.

The following requirements must be met to use AES PIN support:

- ▶ Hardware:
 - z15 systems and Crypto Express7S (Crypto Express6S with CCA V6.0 firmware or Crypto Express5S with CCA V5.2 firmware are also supported)

⁶ Always check the latest information about security certification status for your specific model.

- z14 systems and Crypto Express6S with CCA V6.0 firmware or Crypto Express5S with CCA V5.2
- z13s systems and Crypto Express5S with CCA V5.2 firmware
- z13 systems and Crypto Express5S with CCA V5.0 or later firmware
- ▶ Operating systems requirements for z15:
 - z/OS: All functions require ICSF WD18 (HCR77D0), unless otherwise noted. WD18 supports z/OS V2R1, V2R2, and V2R3; WD19 (HCR77D1 is required for z/OS V2R4.
 - z/VM Version 6.4 with PTFs for guest use.

Support for the updated German Banking standard (DK)

Update support requires ICSF WD18 (HCR77D0) for z/OS V2R2 and V2R3 for:

- ▶ CCA 5.4 & 6.1⁷:
 - ISO-4 PIN Blocks (ISO-9564-1).
 - Directed keys: A key can encrypt or decrypt data, but not both.
 - Allow AES transport keys to be used to export/import *DES* keys in a standard ISO 20038 key block. This feature helps with interoperability between CCA and non-CCA systems.
 - Allow AES transport keys to be used to export/import a small subset of *AES* keys in a standard ISO 20038 key block. This feature helps with interoperability between CCA and non-CCA systems.
 - Triple-length TDES keys with Control Vectors for increased data confidentiality.
- ▶ CCA 5.5: DK Random PIN Generate2 and DK PRW Card Number Update2.
- ▶ CCA 6.2: PCI HSM 3K DES: Support for triple length DES keys (standards compliance).

PKA Translate UDX function into CCA

UDX is custom code that allows the client to add unique operations or extensions to the CCA firmware. Certain UDX functions are integrated into the base CCA code over time to accomplish the following tasks:

- ▶ Remove headaches and challenges that are associated with UDX management and currency.
- ▶ Make available popular UDX functions to a wider audience to encourage adoption.

UDX is integrated into the base CCA code to support translating an external RSA CRT key into new formats. These formats use tags to identify key components. Depending on which new rule array keyword is used with the PKA Key Translate callable service, the service TDES encrypts those components in CBC or ECB mode. In addition, AES CMAC support is delivered.

The following requirements must be met to use this function:

- ▶ Hardware:
 - z15 systems and Crypto Express7s (or Crypto Express6S with CCA 6.0 or Crypto Express5S with CCA V5.2 firmware)
 - z14 systems and Crypto Express6S (or Crypto Express5S with CCA V5.2 firmware)
 - z13s systems and Crypto Express5S with CCA V5.2 firmware

⁷ CCA 5.4 and 6.1 enhancements are also supported for z/OS V2R1 with ICSF HCR77C1 (WD17) with Small Program Enhancements (SPEs) (z/OS continuous delivery model).

- z13 systems and Crypto Express5S with CCA V5.0 or later firmware
- ▶ Operating systems requirements:
 - z/OS: All functions require ICSF WD18 (HCR77D0), unless otherwise noted. WD18 supports z/OS V2R1, V2R2, and V2R3; WD19 (HCR77D1 is required for z/OS V2R4.
 - z/VM Version 6.4 with PTFs for guest use.

Note: Although older IBM Z and operating systems also are supported, they are beyond the scope of this IBM Redbooks publication.

Verb Algorithm Currency

Verb Algorithm Currency is a collection of CCA verb enhancements that are related to customer requirements, with the intent of maintaining currency with cryptographic algorithms and standards. It is also intended for customers who want to maintain the following latest cryptographic capabilities:

- ▶ Secure key support AES GCM encryption
- ▶ Key Check Value (KCV) algorithm for service CSNBKYT2 Key Test 2
- ▶ Key derivation options for CSNDEDH EC Diffie-Hellman service

The following requirements must be met to use this function:

- ▶ Hardware:
 - z15 systems and Crypto Express7s (or Crypto Express6S with CCA 6.0 or Crypto Express5S with CCA V5.2 firmware)
 - z14 systems and Crypto Express6S (or Crypto Express5S with CCA V5.2 firmware)
 - z13s or z13 systems and Crypto Express5S with CCA V5.2 firmware
- ▶ Software:
 - z/OS: All functions require ICSF WD18 (HCR77D0), unless otherwise noted. WD18 supports z/OS V2R1, V2R2, and V2R3; WD19 (HCR77D1 is required for z/OS V2R4.
 - z/VM Version 6.4 with PTFs for guest use.

6.5.3 Crypto Express7S as an EP11 coprocessor

A Crypto Express7S adapter that is configured in Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode provides PKCS #11 secure key support for public sector requirements. Before EP11, the ICSF PKCS #11 implementation supported only clear keys. In EP11, keys can now be generated and securely wrapped under the EP11 Master Key. The secure keys never leave the secure coprocessor boundary decrypted.

The secure IBM Enterprise PKCS #11 (EP11) coprocessor runs the following tasks:

- ▶ Encrypt and decrypt (AES, DES, TDES, and RSA)
- ▶ Sign and verify (DSA, RSA, and ECDSA)
- ▶ Generate keys and key pairs (DES, AES, DSA, ECC, and RSA)
- ▶ HMAC (SHA1, SHA2 or SHA3 [SHA224, SHA256, SHA384, and SHA512])
- ▶ Digest (SHA1, SHS2 or SHA3 [SHA224, SHA256, SHA384, and SHA512])
- ▶ Wrap and unwrap keys
- ▶ Random number generation
- ▶ Get mechanism list and information
- ▶ Attribute values
- ▶ Key Agreement (Diffie-Hellman)

The function extension capability through UDX is not available to the EP11.

When defined in EP11 mode, the TKE workstation is required to manage the Crypto Express6S feature.

z/OS V2.2 and V2.3 require ICSF Web Deliverable WD18 (HCR77D0) to support the following new features:

- ▶ EP11 Stage 4:
 - New elliptic curve algorithms for PKCS#11 signature, key derivation operations:
 - Ed448 elliptic curve
 - EC25519 elliptic curve
 - EP11 Concurrent Patch Apply: Allows service to be applied to the EP11 coprocessor dynamically without taking the crypto adapter offline (already available for CCA coprocessors).
 - eIDAS compliance: eIDAS: Cross-border EU regulation for portable recognition of electronic identification.

6.5.4 Crypto Express7S as an accelerator

A Crypto Express7S adapter that is running in accelerator mode supports only RSA clear key and SSL Acceleration. A request is processed fully in hardware. The Crypto Express accelerator is a coprocessor that is reconfigured by the installation process so that it uses only a subset of the coprocessor functions at a higher speed. Reconfiguration is disruptive to coprocessor and accelerator operations. The coprocessor or accelerator must be deactivated before you begin the reconfiguration.

FIPS 140-2 certification is not relevant to the accelerator because it operates with clear keys only. The function extension capability through UDX is not available to the accelerator.

The functions that remain available when the Crypto Express6S feature is configured as an accelerator are used for the acceleration of modular arithmetic operations. That is, the RSA cryptographic operations are used with the SSL/TLS protocol. The following operations are accelerated:

- ▶ PKA Decrypt (CSNDPKD) with PKCS-1.2 formatting
- ▶ PKA Encrypt (CSNDPKE) with zero-pad formatting
- ▶ Digital Signature Verify

The RSA encryption and decryption functions support key lengths of 512 - 4,096 bits in the Modulus-Exponent (ME) and Chinese Remainder Theorem (CRT) formats.

6.5.5 Managing Crypto Express7S

Each cryptographic coprocessor has 85 physical sets of registers or queue registers, which corresponds to the maximum number of LPARs that are running on a z15 T01, which is also 85. Each of these sets belongs to the following domains:

- ▶ A cryptographic domain index, in the range of 0 - 84 for z15 T01, is allocated to a logical partition in its image profile. The same domain must also be allocated to the ICSF instance that is running in the logical partition that uses the Options data set.
- ▶ Each ICSF instance accesses only the Master Keys or queue registers that correspond to the domain number that is specified in the logical partition image profile at the SE and in its Options data set. Each ICSF instance sees a logical cryptographic coprocessor that

consists of the physical cryptographic engine and the unique set of registers (the domain) that is allocated to this logical partition.

The installation of CP Assist for Cryptographic Functions (CPACF) DES/TDES enablement (FC 3863) is required to use the Crypto Express7S feature.

Each Crypto Express7S FC 0898 includes two IBM 4769 PCIe Cryptographic Coprocessors (PCIeCC - which is a hardware security module - HSM); FC 0899 includes one IBM 4769 PCIeCC. The adapters are available in the following configurations:

- ▶ IBM Enterprise Common Cryptographic Architecture (CCA) Coprocessor (CEX7C)
- ▶ IBM Enterprise Public Key Cryptography Standards #11 (PKCS) Coprocessor (CEX7P)
- ▶ IBM Crypto Express7S Accelerator (CEX7A)

During the feature installation, the PCI-X adapter is configured by default as the CCA coprocessor.

The configuration of the Crypto Express7S adapter as EP11coprocessor requires a TKE tower workstation (FC 0088) or a TKE rack-mounted workstation (FC 0087) with TKE 9.2 (FC 0881) LIC. The same requirement applies to CCA mode for a full PCI-compliant environment.

The Crypto Express7S feature does not use CHPIDs from the channel subsystem pool. However, the Crypto Express7S feature requires one slot in a PCIe I/O drawer, and one PCHID for each PCIe cryptographic adapter.

For enabling an LPAR to use a Crypto Express7S adapter, the following cryptographic resources in the image profile must be defined for each partition:

- ▶ Usage domain index
- ▶ Control domain index
- ▶ PCI Cryptographic Coprocessor Candidate List
- ▶ PCI Cryptographic Coprocessor Online List

This task is accomplished by using the Customize/Delete Activation Profile task, which is in the Operational Customization Group, from the HMC or from the SE. Modify the cryptographic initial definition from the Crypto option in the image profile, as shown in Figure 6-7 on page 242.

Important: After this definition is modified, any change to the image profile requires a DEACTIVATE and ACTIVATE of the logical partition for the change to take effect. Therefore, this cryptographic definition is disruptive to a running system.

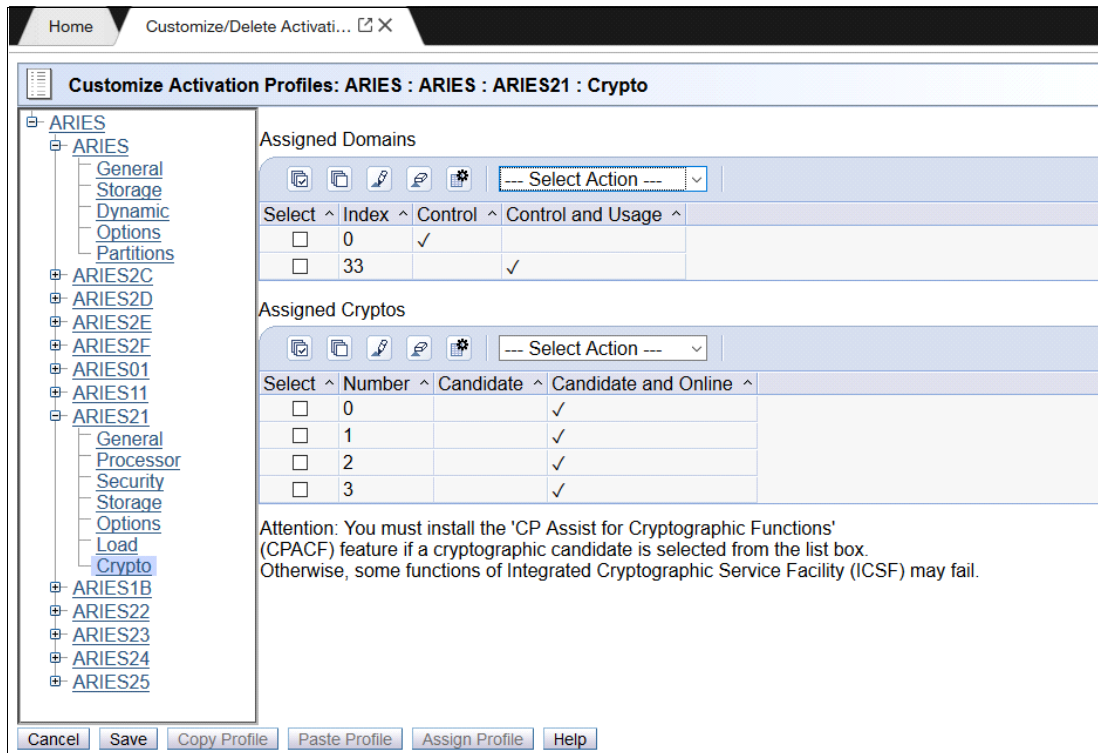


Figure 6-7 Customize Image Profiles: Crypto

The following cryptographic resource definitions are used:

► Control Domain

Identifies the cryptographic coprocessor domains that can be administered from this logical partition if it is set up as the TCP/IP host for the TKE.

If you are setting up the host TCP/IP in this logical partition to communicate with the TKE, the partition is used as a path to other domains' Master Keys. Indicate all the control domains that you want to access (including this partition's own control domain) from this partition.

► Control and Usage Domain

Identifies the cryptographic coprocessor domains that are assigned to the partition for all cryptographic coprocessors that are configured on the partition. The usage domains cannot be removed if they are online. The numbers that are selected must match the domain numbers that are entered in the Options data set when you start this partition instance of ICSF.

The same usage domain index can be used by multiple partitions, regardless to which CSS they are defined. However, the combination of PCIe adapter number and usage domain index number must be unique across all active partitions.

► Cryptographic Candidate list

Identifies the cryptographic coprocessor numbers that can be accessed by this logical partition. From the list, select the coprocessor numbers (in the range 0 - 15) that identify the PCIe adapters to be accessed by this partition.

► Cryptographic Online list

Identifies the cryptographic coprocessor numbers that are automatically brought online during logical partition activation. The numbers that are selected in the online list must also be part of the candidate list.

After they are activated, the active partition cryptographic definitions can be viewed from the HMC. Select the CPC, and click **View LPAR Cryptographic Controls** in the CPC Operational Customization window. The resulting window displays the definition of Usage and Control domain indexes, and PCI Cryptographic candidate and online lists, as shown in Figure 6-8. Information is provided for active logical partitions only.

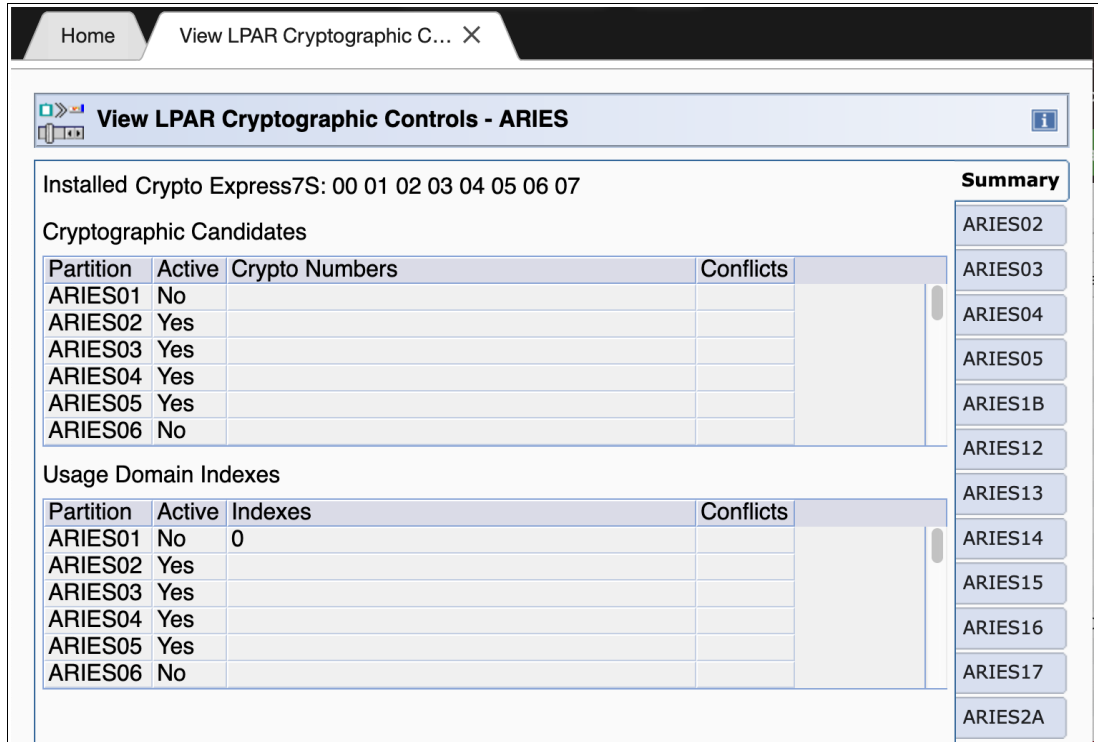


Figure 6-8 View LPAR Cryptographic Controls

Operational changes can be made by using the Change LPAR Cryptographic Controls task, which reflects the cryptographic definitions in the image profile for the partition. With this function, the cryptographic feature can be added and removed dynamically, without stopping a running operating system.

For more information about the management of Crypto Express7S, see IBM z15 Configuration Setup, SG24-8860.

6.6 Trusted Key Entry workstation

The TKE workstation is an optional feature that offers key management functions. It can be a TKE tower workstation (FC 0088) or TKE rack-mounted workstation (FC 0087) for z15 systems to manage Crypto Express7S, Crypto Express6S, or Crypto Express5S.

The TKE contains a combination of hardware and software. A mouse, keyboard, flat panel display, PCIe adapter, and a writable USB media to install the TKE Licensed Internal Code (LIC) are included with the system unit. The TKE workstation requires an IBM 4768 crypto adapter.

A TKE workstation is part of a customized solution for the use of the Integrated Cryptographic Service Facility for z/OS (ICSF for z/OS) or Linux for IBM Z. This program provides a basic key management system for the cryptographic keys of a z15 system that has Crypto Express features installed.

The TKE provides a secure, remote, and flexible method of providing Master Key Part Entry, and to remotely manage PCIe cryptographic coprocessors. The cryptographic functions on the TKE run by one PCIe cryptographic coprocessor. The TKE workstation communicates with the IBM Z system through a TCP/IP connection. The TKE workstation is available with Ethernet LAN connectivity only. Up to 10 TKE workstations can be ordered.

TKE FCs 0087 and 0088 can be used to control any supported Crypto Express feature supported on z15. They also can be used to control the Crypto Express6S, Crypto Express5S on z14, Crypto Express5S on z13 and z13s systems, and the Crypto adapters on older, still supported systems.

The TKE 9.2 LIC (FC 0881) features the following enhancements over the described functions of LIC 9.1 and 9.0:

- ▶ TKE 9.2 LIC is enhanced with functions to support the management of the Crypto Express7S 4769 host crypto module.
- ▶ TKE 9.2 allow Host Transaction Programs to run over a TLS connection for z/OS LPARs. The TKE is checking whether TLS is configured on the Host Transaction Program port and automatically uses TLS when communicating with the host.

TKE 9.2 can handle a combination of AT-TLS configured hosts with those hosts that are not TLS capable.
- ▶ The TKE 9.2. supports AES Operational Key parts to be tagged as PCI-compliant. With CCA 6.3, support for marking some AES operational key parts as being PCI-compliant is introduced. The TKE is mandatory to support the PCI-compliant environment for CCA-mode.
- ▶ A new Stronger encryption is used when negotiating the session key to an EP11 Domain. When loading key parts into an EP11 host domain, a session key is derived by the smart card and the target domain. The BLUE TKE smart cards (00RY790) includes 521-bit EC capability. The 521-bit EC strength is used during the EP11 session key derivation process. For CCA-mode, this stronger encryption was made available with TKE 9.1 LIC.
- ▶ TKE 9.2 supports 32 character user IDs when the TKE Host Transaction Program is on a LINUX system.

The following enhancements were introduced with TKE 9.1 LIC (FC 0880):

- ▶ TKE 9.1 License Internal Code enhancements for support EC521 strength TKE and Migration zones. An EC521 Migration zone is required if you want to use the migration wizard to collect and apply PCI-compliant domain information.
- ▶ TKE 9.1 also has a new family of wizards that makes it easy to create EC521 zones on all of its smart cards. This feature simplifies the process of deploying a TKE for the first time or moving data from a weaker TKE zone to a new EC521 zone.
- ▶ A new smart card for the TKE allows stronger Elliptic Curve Cryptography (ECC) levels. Other TKE Smart Cards (FC 0900, packs of 10, FIPS certified blanks) require TKE 9.1 LIC.

The TKE 9.0 LIC (FC 0879) includes the following features:

- ▶ Key material copy to alternative zone
By using TKE 9.0, key material can be copied from smart cards in one TKE zone to smart cards in another zone. You might have old 1024-bit strength TKE zones, and might want to move or copy the key material in those zones into a new, stronger TKE zone. To use this new feature, you create TKE or EP11 smart cards on your TKE 9.0 system. You then enroll the new TKE or EP11 smart cards in an alternative zone. This process allows you to copy smart card content from a smart card that is enrolled in the alternative zone.
- ▶ Save TKE data directory structure with files to USB
TKE data can be saved to, or restored from, removable media in the same directory structure they were found on the TKE.
- ▶ Create key parts without opening a host
Administrators can now use the TKE application to create key parts without opening a host. This ability allows the key administrator to create key parts while being offline or before any hosts are defined. This feature can be found in the TKE application under the **Utilities** → **Create CCA key parts** pull-down menu.
- ▶ New TKE Audit Log application
A new TKE Audit Log application is available for the Privileged Mode Access ID of AUDITOR. This application provides an easy-to-use interface to view the TKE workstation security audit records from the TKE workstation.
- ▶ Heartbeat audit record
TKE workstations cut an audit record when the TKE starts or when no audit events occurred during a client-configured duration. The record shows the serial number of the TKE local crypto adapter and indicates whether the local crypto adapter was changed since the last check.
- ▶ Performance improvements for domain groups
Depending on the size of a domain group, you might experience performance improvements with CCA version 5.3 when a Load, Set, or Clear operation is performed from inside a domain group. For example, if you group all 85 domains on a Host Crypto Express and issue a Clear New Master Key register operation, the number of commands that is issued to the module drops from 85 to 1.
- ▶ Secure key entry on EP11
TKE 9.0 EP11 smart card applet now supports secure key entry of EP11 master key parts.
- ▶ New certificate manager for domains
Every domain now can manage a set of parent X.509 certificates for validating operating X.509 certificates that are used by applications that are running in the domain.

The following features are related to support for the Crypto Express6S with CCA 6.0. The Crypto Express6S with CCA 6.0 is designed to meet the PCI-HSM PIN Transaction Security v3.0, 2016 standard:

- ▶ Domain mode management
With CCA 6.0, individual domains are in one of the following modes:
 - Normal Mode
 - Imprint Mode
 - Compliant Mode

Imprint and compliant mode were added to indirectly and directly meet the PCI-HSM PIN Transaction Security v3.0, 2016 requirement. TKE is required to manage Host Crypto Module domains in imprint and compliant mode.

- ▶ Set clock

With TKE 9.0, the host crypto module's clock can be set. The clock must be set before a domain can be placed in imprint mode.

- ▶ Domain-specific Host Crypto Module Audit Log management

Domains in imprint mode or compliant mode on a Crypto Express6S maintain a domain-specific module audit log. The TKE provides a feature for downloading the audit records so they can be viewed.

- ▶ Domain-specific roles and authorities

Domains in imprint mode or compliant mode on a Crypto Express6S must be managed by using domain-specific roles and authorities. The TKE provides new management features for the domain-specific roles and authorities. The roles are subject to forced dual control policies that prevent roles from issuing and co-signing a command. For information about how to manage imprint and compliant mode domains, see the TKE User's Guide.

- ▶ Setup PCI Environment Wizard

To simplify the management of a compliant domain, the TKE provides a setup wizard that creates a minimum set of forced dual control roles and authorities that are needed to manage a compliant domain. For more information about how to manage imprint and compliant mode domains, see the TKE User's Guide.

Tip: For more information about handling a TKE, see the [TKE Introduction video](#).

6.6.1 Logical partition, TKE host, and TKE target

If one or more LPARs are configured to use Crypto Express coprocessors, the TKE workstation can be used to manage DES, AES, ECC, and PKA master keys. This management can be done for all cryptographic domains of each Crypto Express coprocessor feature that is assigned to the LPARs that are defined to the TKE workstation.

Each LPAR in the same system that uses a domain that is managed through a TKE workstation connection is a TKE host or TKE target. An LPAR with a TCP/IP connection to the TKE is referred to as the *TKE host*; all other partitions are *TKE targets*.

The cryptographic controls that are set for an LPAR through the SE determine whether the workstation is a TKE host or a TKE target.

6.6.2 Optional smart card reader

An optional smart card reader (FC 0895) can be added to the TKE workstation. One FC 0895 includes two smart card readers, two cables to connect them to the TKE workstation, and 20 smart cards. The reader supports the use of smart cards that contain an embedded microprocessor and associated memory for data storage. The memory can contain the keys to be loaded into the Crypto Express features. These readers can be used with smart cards only that have applets that were loaded from a TKE 8.1 or later. These cards are FIPS certified.

Smart card readers from FC 0885 or FC 0891 can be carried forward. Smart cards can be used on TKE 9.0 with these readers. Access to and use of confidential data on the smart card are protected by a user-defined PIN. Up to 990 other smart cards can be ordered for backup. (The extra smart card feature code is FC 0892.) When one feature code is ordered, 10 smart cards are included. The order increment is 1 - 99 (10 - 990 blank smart cards).

If smart cards with applets that are not supported by the new smart card reader are reused, new smart cards on TKE 8.1 or later must be created and the content from the old smart cards to the new smart cards must be copied. The new smart cards can be created and copied on a TKE 8.1 system. If the copies are done on TKE 9.0, the source smart card must be placed in an older smart card reader from feature code 0885 or 0891.

A new smart card for the Trusted Key Entry (TKE) allows stronger Elliptic Curve Cryptography (ECC) levels. More TKE Smart Cards (FC 0900, packs of 10, FIPS certified blanks) require TKE 9.1 LIC or up.

6.6.3 TKE hardware support and migration information

The new TKE 9.2 LIC (FC 0881) is originally shipped with a new z15 server. The following TKE workstations can be ordered with a new z15:

- ▶ TKE 9.2 tower workstation (FC 0088)
- ▶ TKE 9.2 rack-mounted workstation (FC 0087)

Note: Several options for ordering the TKE with or without ordering Keyboard, Mouse, and Display are available. Ask your IBM Representative for more information about which option is the best option for you.

The TKE 9.2 LIC requires the 4768 crypto adapter. The TKE 8.x and TKE 7.3 workstations can be upgraded to the TKE 9.2 tower workstation by purchasing a 4768 crypto adapter.

The Omnikey Cardman 3821 smart card readers can be carried forward to any TKE 9.2 workstation. Smart cards 45D3398, 74Y0551, and 00JA710 can be used on TKE 9.2.

When performing a MES upgrade from TKE 7.3, TKE 8.0, or TKE 8.1 to a TKE 9.2 installation, the following steps must be completed:

1. Save Upgrade Data on an old TKE to USB memory to save client data.
2. Replace the 4767 crypto adapter with the 4768 crypto adapter.
3. Upgrade the firmware to TKE 9.2.
4. Install the Frame Roll to apply Save Upgrade Data (client data) to the TKE 9.2 system.
5. Run the TKE Workstation Setup wizard.

TKE upgrade considerations

If you are migrating your configuration with Crypto Express5S and TKE Release 8.x to a z15, you do not need to upgrade the TKE LIC.

Note: A workstation that was upgraded to TKE V8.x includes the 4767 cryptographic adapter that is required to manage Crypto Express5S; however, it cannot be used to manage the Crypto Express7S or Crypto Express6S.

If your z15 includes Crypto Express7S or Crypto Express6S, you must upgrade to TKE V9.2, which requires the 4768 cryptographic adapter.

Upgrading to TKE V9.2 requires that your TKE hardware is compatible with the 4768 cryptographic adapter. The following older TKE hardware features are compatible 4768 cryptographic adapters:

- ▶ FC 0842
- ▶ FC 0847
- ▶ FC 0092
- ▶ FC 0098

Important: TKE workstations that are at FC 0841 or older do not support the 4767 or 4768 cryptographic adapters.

For more information about TKE hardware support, see Table 6-3. For some functionality, requirements must be considered; for example, the characterization of a Crypto Express adapter in EP 11 mode always requires the use of a TKE.

Table 6-3 TKE Compatibility Matrix

| TKE workstation | TKE Release LIC | 7.3 ^a | 8.0 ^a | 8.1 ^a | 9.0 | 9.1 | 9.2 |
|----------------------------------|-------------------|------------------|------------------|------------------|--------------|--------------|--------------|
| | HW Feature Code | 0842 | 0847 | 0847 or 0097 | 0085 or 0086 | 0085 or 0086 | 0087 or 0088 |
| | LICC | 0872 | 0877 | 0878 | 0879 | 0880 | 0881 |
| | Smart Card Reader | 0885 | 0891 | 0891 | 0895 | 0895 | 0895 |
| | Smart Card | 0884 | 0892 | 0892 | 0892 | 0892 | 0892 |
| Manage Host Crypto Module | CEX7C (CCA) | No | No | No | No | No | Yes |
| | CEX7P (EP11) | No | No | No | No | No | Yes |
| | CEX6C (CCA) | No | No | No | Yes | Yes | Yes |
| | CEX6P (EP11) | No | No | No | Yes | Yes | Yes |
| | CEX5C (CCA) | Yes | Yes | Yes | Yes | Yes | Yes |
| | CEX5P (EP11) | Yes | Yes | Yes | Yes | Yes | Yes |

a. The TKE workstation FC 0842 that is running LIC 7.3, or 8.x can be upgraded to TKE LIC V9.2 by adding a 4786 cryptographic adapter.

Attention: The TKE is unaware of the CPC type where the host crypto module is installed. That is, the TKE does not consider whether a Crypto Express is running on z15, 14, or z13, or z13s system. Therefore, the LIC can support any CPC where the coprocessor is supported, but the TKE LIC must support the specific crypto module.

6.7 Cryptographic functions comparison

The functions or attributes on z15 for the two cryptographic hardware features are listed in Table 6-4, where “X” indicates that the function or attribute is supported.

Table 6-4 Cryptographic functions on z15

| Functions or attributes | CPACF | CEX7C | CEX7P | CEX7A |
|--|-------|-------|-------|----------------|
| Supports z/OS applications that use CSF | X | X | X | X |
| Supports Linux on Z CCA applications | X | X | - | X |
| Encryption and decryption by using secret-key algorithm | - | X | X | - |
| Provides the highest SSL/TLS handshake performance | - | - | - | X |
| Supports SSL/TLS functions | X | X | X | X |
| Provides the highest symmetric (clear key) encryption performance | X | - | - | - |
| Provides the highest asymmetric (clear key) encryption performance | - | - | - | X |
| Provides the highest asymmetric (encrypted key) encryption performance | - | X | X | - |
| Nondisruptive process to enable ^a | - | X | X | X |
| Requires IOCDs definition | - | - | - | - |
| Uses CHPID numbers | - | - | - | - |
| Uses PCHIDs (one PCHID) | - | X | X | X |
| Requires CPACF enablement (FC 3863) ^b | X | X | X | X |
| Requires ICSF to be active | - | X | X | X |
| Offers UDX | - | X | - | - |
| Usable for data privacy: Encryption and decryption processing | X | X | X | - |
| Usable for data integrity: Hashing and message authentication | X | X | X | - |
| Usable for financial processes and key management operations | - | X | X | - |
| Crypto performance IBM RMF monitoring | - | X | X | X |
| Requires system master keys to be loaded | - | X | X | - |
| System (master) key storage | - | X | X | - |
| Retained key storage | - | X | - | - |
| Tamper-resistant hardware packaging | - | X | X | X ^c |
| Designed for FIPS 140-2 Level 4 certification | - | X | X | X |

| Functions or attributes | CPACF | CEX7C | CEX7P | CEX7A |
|--|-------|-------|-------|-------|
| Supports Linux applications that perform SSL handshakes | - | - | - | X |
| RSA functions | - | X | X | X |
| High-performance SHA-1, SHA-2, and SHA-3 | X | X | X | - |
| Clear key DES or triple DES | X | - | - | - |
| Advanced Encryption Standard (AES) for 128-bit, 192-bit, and 256-bit keys | X | X | X | - |
| True random number generator (TRNG) | X | X | X | - |
| Deterministic random number generator (DRNG) | X | X | X | - |
| Pseudo random number generator (PRNG) | X | X | X | - |
| Clear key RSA | - | - | - | X |
| Payment Card Industry (PCI) PIN Transaction (PTS) Hardware Security Module (HSM) PCI-HSM | | X | X | |
| Europay, MasterCard, and Visa (EMV) support | - | X | - | - |
| Public Key Decrypt (PKD) support for Zero-Pad option for clear RSA private keys | - | X | - | - |
| Public Key Encrypt (PKE) support for Mod_Raised_to Power (MRP) function | - | X | X | - |
| Remote loading of initial keys in ATM | - | X | - | - |
| Improved key exchange with non-CCA systems | - | X | - | - |
| ISO 16609 CBC mode triple DES message authentication code (MAC) support | - | X | - | - |
| AES GMAC, AES GCM, AES XTS mode, CMAC | - | X | - | - |
| SHA-2, SHA-3 (384,512), HMAC | - | X | - | - |
| Visa Format Preserving Encryption | - | X | - | - |
| AES PIN support for the German banking industry | - | X | - | - |
| ECDSA (192, 224, 256, 384, 521 Prime/NIST) | - | X | - | - |
| ECDSA (160, 192, 224, 256, 320, 384, 512 BrainPool) | - | X | - | - |
| ECDH (192, 224, 256, 384, 521 Prime/NIST) | - | X | - | - |
| ECDH (160, 192, 224, 256, 320, 384, 512 BrainPool) | - | X | - | - |
| PNG (Prime Number Generator) | - | X | - | - |

- a. To make adding the Crypto Express features nondisruptive, the logical partition must be predefined with the appropriate PCI Express cryptographic adapter number. This number must be selected from its candidate list in the partition image profile.
- b. This feature is not required for Linux if only RSA clear key operations are used. DES or triple DES encryption requires CPACF to be enabled.
- c. This feature is physically present, but is not used when configured as an accelerator (clear key only).

6.8 Cryptographic operating system support for z15

The following section gives an overview of the operating systems requirements in relation to cryptographic elements.

6.8.1 Crypto Express7S Toleration

Crypto Express7S (0898/0899) Toleration treats Crypto Express7S cryptographic coprocessors and accelerators as Crypto Express5 coprocessors and accelerators. The following minimum prerequisites must be met:

- ▶ z/OS V2.4; certain functions require WD19 (HCR77D1)
- ▶ z/OS V2.3 with PTFs
- ▶ z/OS V2.2 with PTFs
- ▶ z/OS V2.1 with PTFs
- ▶ z/VM V7.1 and V7.2 for guest use
- ▶ z/VM V6.4 with PTFs for guest use
- ▶ z/VSE V6.2 with PTFs
- ▶ z/TPF V1.1 with PTFs
- ▶ Linux on Z: IBM is working with its Linux distribution partners to provide support by way of maintenance or future releases for the following distributions:
 - SUSE Linux Enterprise Server 12 and SLES 11
 - Red Hat Enterprise Linux (RHEL) 7 and Red Hat Enterprise Linux 6
 - Ubuntu 16.04 LTS (or higher)
- ▶ The KVM hypervisor, which is offered with SLES 12 SP2 with service, RHEL 7.5 with kernel-alt package (kernel 4.14) and Ubuntu 16.04 LTS with service, and Ubuntu 18.04 LTS with service Linux distributions. For more information about minimal and recommended distribution levels, see the [Tested platforms for Linux web page](#) of the IBM IT infrastructure website.

6.8.2 Crypto Express7S support of VFPE

The following minimum prerequisites must be met to use this element:

- ▶ z/OS V2.4 with PTFs
- ▶ z/OS V2.3 with PTFs
- ▶ z/OS V2.2 with the Enhanced Cryptographic Support for z/OS V1.13-V2.2 web deliverable installed
- ▶ z/VM V7.2 for guest use
- ▶ z/VM V7.1 with PTFs for guest use
- ▶ z/VM V6.4 with PTFs for guest use
- ▶ Linux on IBM Z:
 - SUSE SLES 15 SP1 with service, SUSE SLES 12 SP4 with service, and SUSE SLES 11 SP4 with service.
 - Red Hat RHEL 8.0 with service, Red Hat RHEL 7.7 with service, and Red Hat RHEL 6.10 with service.

- Ubuntu 18.04.1 LTS with service and Ubuntu 16.04.6 LTS with service.
- The support statements for z15 also cover the KVM hypervisor on distribution levels that have KVM support.

For more information about the minimum required and recommended distribution levels, see the [IBM Z website](#).

6.8.3 Crypto Express7S support of greater than 16 domains

The following prerequisites must be met to support more than 16 domains:

- ▶ z/OS V2.4 with PTFs
- ▶ z/OS V2.3 with PTFs
- ▶ z/OS V2.2 with the Enhanced Cryptographic Support for z/OS V1.13-V2.2 Web deliverable installed
- ▶ z/VM V7.1 and V7.2 for guest use
- ▶ z/VM V6.4 with PTFs for guest use
- ▶ z/VSE V6.2 with PTFs
- ▶ Linux on IBM Z:
 - SUSE SLES 15 SP1 with service, SUSE SLES 12 SP4 with service, and SUSE SLES 11 SP4 with service.
 - Red Hat RHEL 8.0 with service, Red Hat RHEL 7.7 with service, and Red Hat RHEL 6.10 with service.
 - Ubuntu 18.04.1 LTS with service and Ubuntu 16.04.6 LTS with service.
 - The support statements for z15 also cover the KVM hypervisor on distribution levels that have KVM support.

For more information about the minimum required and recommended distribution levels, see the [IBM Z website](#).

For more information about the software support levels for cryptographic functions, see Chapter 7, “Operating system support” on page 253.



Operating system support

This chapter describes the minimum operating system requirements and support considerations for the IBM z15™ servers and their features. It addresses z/OS, z/VM, z/VSE, z/TPF, Linux on Z, and the KVM hypervisor.

Naming: The IBM z15 server generation is available as the following machine types and models:

- ▶ Machine Type 8561 (M/T 8561), Model T01, Features Max34, Max71, Max108, Max145, and Max190, which is further identified as *IBM z15 Model T01*, or *z15 T01*, unless otherwise specified.
- ▶ Machine Type 8562 (M/T 8562), Model T02, Features Max4, Max13, Max21, Max32, and Max65, which is further identified as *IBM z15 Model T02*, or *z15 T02*, unless otherwise specified.

In the remainder of this chapter, IBM z15 (z15) refers to both machine types.

Because this information is subject to change, see the following hardware fix categories for most current information:

- ▶ IBM.Device.Server.z15-8561.* for z15 T01, and
- ▶ IBM.Device.Server.z15-8562.* for z15 T02.

Support of z15 functions depends on the operating system, its version, and release.

This chapter includes the following topics:

- ▶ 7.1, “Operating systems summary” on page 254
- ▶ 7.2, “Support by operating system” on page 254
- ▶ 7.3, “z15 features and function support overview” on page 257
- ▶ 7.4, “Support by features and functions” on page 273
- ▶ 7.5, “z/OS migration considerations” on page 322
- ▶ 7.6, “z/VM migration considerations” on page 326
- ▶ 7.7, “z/VSE migration considerations” on page 328
- ▶ 7.8, “Software licensing” on page 328
- ▶ 7.9, “References” on page 331

7.1 Operating systems summary

The minimum operating system levels that are required on z15 servers are listed in Table 7-1.

End of service operating systems: Operating system levels that are no longer in service are not covered in this publication. These older levels might support some features.

Table 7-1 z15 minimum operating systems requirements

| Operating systems ^a | Supported Version and release on z15 ^b |
|--------------------------------|---|
| z/OS | V2R1 ^c |
| z/VM | V6R4 |
| z/VSE | V6.2 ^{d,e} |
| z/TPF | V1R1 |
| Linux on Z ^f | See Table 7-2 on page 256 |

a. Only z/Architecture mode is supported.

b. Service is required.

c. z/OS V2R1 - Compatibility only. The IBM Software Support Services for z/OS V2R1 offered as October 1st, 2018, provides the ability for customers to purchase extended defect support service for z/OS V2.1.

d. End of service date for z/VSE V5R2 was October 31, 2018.

e. End of service date for z/VSE V6R1 is Sept. 30, 2019.

f. KVM hypervisor is supported by Linux distribution partners.

The use of certain features depends on the operating system. In all cases, program temporary fixes (PTFs) might be required with the operating system level that is indicated. Check the z/OS fix categories, or the subsets of the 8561DEVICE (z15 T01) and 8652DEVICE (z15 T02) PSP buckets for z/VM and z/VSE. The fix categories and the PSP buckets are continuously updated, and contain the latest information about maintenance:

- ▶ Hardware and software buckets contain installation information, hardware and software service levels, service guidelines, and cross-product dependencies.
- ▶ For more information about Linux on Z distributions and KVM hypervisor, see the distributor's support information.

7.2 Support by operating system

z15 systems introduce several new functions. This section describes the support of those functions by the current operating systems. Also included are some of the functions that were introduced in previous IBM Z servers and carried forward or enhanced in z15 servers. Features and functions that are available on previous servers, but no longer supported by z15 servers were removed.

For more information about supported functions that are based on operating systems, see 7.3, "z15 features and function support overview" on page 257. Tables are built by function and feature classification to help you determine, by a quick scan, what is supported and the minimum operating system level that is required.

7.2.1 z/OS

z/OS Version 2 Release 2 is the earliest in-service release that supports z15 servers. Consider the following points:

- ▶ Service support for z/OS Version 2 Release 1 ended in September of 2018; however, a fee-based extension for defect support (for up to three years) can be obtained by ordering IBM Software Support Services - Service Extension for z/OS 2.1.
- ▶ z15 capabilities differ depending on the z/OS release. Toleration support is provided on z/OS V2R1. Exploitation support is provided on z/OS V2R2 and later only¹.

For more information about supported functions and their minimum required support levels, see 7.3, “z15 features and function support overview” on page 257.

7.2.2 z/VM

z/VM V6R4, z/VM V7R1, and z/VM V7R2 provide support that enables guests to use the following features that are supported by z/VM on IBM z15™:

- ▶ z/Architecture support
- ▶ New hardware facilities
- ▶ ESA/390-compatibility mode for guests
- ▶ Crypto Clear Key ECC operations
- ▶ RoCE Express2 support
- ▶ Dynamic I/O support

Provided for managing the configuration of OSA-Express7S and OSA-Express6S OSD CHPIDs, FICON Express16SA² and FICON Express16S+ (FC and FCP CHPIDs), and RoCE Express2 features.

- ▶ Improved memory management

For more information about supported functions and their minimum required support levels, see 7.3, “z15 features and function support overview” on page 257.

7.2.3 z/VSE

z15 support is provided by z/VSE V6R2 and later, with the following considerations:

- ▶ z/VSE runs in z/Architecture mode only.
- ▶ z/VSE supports 64-bit real and virtual addressing.

For more information about supported functions and their minimum required support levels, see 7.3, “z15 features and function support overview” on page 257.

7.2.4 z/TPF

z15 support is provided by z/TPF V1R1 with PTFs. For more information about supported functions and their minimum required support levels, see 7.3, “z15 features and function support overview” on page 257.

¹ Use support for select features by way of PTFs. Toleration support for new hardware might also require PTFs.

² FICON Express16SA supported on z15 T01 only.

7.2.5 Linux on IBM Z (Linux on Z)

Generally, a new machine is not apparent to Linux on Z. For z15, toleration support is required for the following functions and features:

- ▶ IPL in “z/Architecture” mode
- ▶ Crypto Express7S cards
- ▶ RoCE Express cards
- ▶ 8-byte LPAR offset

The service levels of SUSE, Red Hat, and Ubuntu releases that are supported at the time of this writing are listed in Table 7-2.

Table 7-2 Linux on Z distributions

| Linux on Z distribution ^a | Supported Version and Release on z15 ^b |
|--------------------------------------|---|
| SUSE Linux Enterprise Server | 15 SP1 |
| SUSE Linux Enterprise Server | 12 SP4 ^c |
| SUSE Linux Enterprise Server | 11 SP4 ^c |
| Red Hat RHEL | 8.0 with service |
| Red Hat RHEL | 7.7 ^c with service |
| Red Hat RHEL | 6.10 ^c with service |
| Ubuntu | 18.04.1 LTS ^c |
| Ubuntu | 16.04.5 LTS ^c |
| KVM Hypervisor ^d | Offered with the supported Linux distributions. |

a. Only z/Architecture (64-bit mode) is supported. IBM testing identifies the minimum required level and the recommended levels of the tested distributions.

b. Fix installation is required for toleration.

c. Maintenance is required.

d. For more information about minimal and recommended distribution levels, see [the Linux on Z website](#).

For more information about supported Linux distributions on IBM Z servers, see the [Tested platforms for Linux page](#) of the IBM IT infrastructure website.

IBM is working with Linux distribution Business Partners to provide further use of selected z15 functions in future Linux on Z distribution releases.

Consider the following guidelines:

- ▶ Use SUSE Linux Enterprise Server 15, Red Hat RHEL 8, or Ubuntu 18.04 LTS or newer in any new projects for z15 servers.
- ▶ Update any Linux distribution to the latest service level before migrating to z15 servers.
- ▶ Adjust the capacity of any Linux on Z and z/VM guests, in terms of the number of IFLs and CPs, real or virtual, according to the PU capacity of the z15 servers.

7.2.6 KVM hypervisor

KVM is offered through our Linux distribution partners to help simplify delivery and installation. Linux and KVM is provided from a single source. With KVM being included in the Linux distribution, ordering and installing KVM is easier.

For KVM support information, see [the IBM Z website](#).

7.3 z15 features and function support overview

The following list the z15 features and functions and their minimum required operating system support levels:

- ▶ Table 7-3, “Supported Base CPC Functions or z/OS and z/VM” on page 258
- ▶ Table 7-4, “Supported base CPC functions for z/VSE, z/TPF, and Linux on Z” on page 259
- ▶ Table 7-5, “Supported coupling and clustering functions for z/OS and z/VM” on page 261
- ▶ Table 7-6, “Supported storage connectivity functions for z/OS and z/VM” on page 262
- ▶ Table 7-7, “Supported storage connectivity functions for z/VSE, z/TPF, and Linux on Z” on page 264
- ▶ Table 7-8, “Supported network connectivity functions for z/OS and z/VM” on page 266
- ▶ Table 7-9, “Supported network connectivity functions for z/VSE, z/TPF, and Linux on Z” on page 268
- ▶ Table 7-10, “Supported cryptography functions for z/OS and z/VM” on page 271
- ▶ Table 7-11, “Supported cryptography functions for z/VSE, z/TPF, and Linux on Z” on page 272

Information about Linux on Z refers exclusively to the appropriate distributions of SUSE, Red Hat, and Ubuntu.

All tables use the following conventions:

- ▶ Y: The function is supported.
- ▶ N: The function is not supported.
- ▶ -: The function is not applicable to that specific operating system.

Note: The tables in this section list but do not explicitly mark all the features that require fixes that are required by the corresponding operating system for toleration or exploitation. For more information, see the PSP buckets for their respective devices and subsets:

- ▶ z15 T01 - PSP bucket 8561DEVICE
- ▶ z15 T02 - PSP bucket 8562DEVICE

7.3.1 Supported CPC functions

The supported Base CPC Functions or z/OS and z/VM are listed in Table 7-3.

Note: z/OS V2R1 support has ended on as of September 2018. No new function is provided for exploiting the new HW features (toleration support only). Although extended (fee-based) support for z/OS 2.1 can be obtained, support for z/OS 2.1 is not covered extensively in this document.

Table 7-3 Supported Base CPC Functions or z/OS and z/VM

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|---|----------------|------------------|------------------|-----------------|-----------------|-----------------|
| z15 servers | Y | Y | Y | Y | Y | Y |
| Maximum processor unit (PUs) per system image | 190 | 190 ^b | 190 ^b | 80 ^c | 80 ^c | 64 ^d |
| Maximum main storage size ^e | 4 TB | 4 TB | 4 TB | 2 TB | 2 TB | 2 TB |
| z15 T01 - 85 LPARs | Y | Y | Y | Y | Y | Y |
| z15 T02 - 40 LPARs ^f | Y | Y | Y | Y | Y | Y |
| Separate LPAR management of PUs | Y | Y | Y | Y | Y | Y |
| Dynamic PU add | Y | Y | Y | Y | Y | Y |
| Dynamic LPAR memory upgrade | Y | Y | Y | Y | Y | Y |
| LPAR group absolute capping | Y | Y | Y | Y | Y | Y |
| Capacity Provisioning Manager | Y | Y | Y | N | N | N |
| Program-directed re-IPL | - | - | - | Y | Y | Y |
| HiperDispatch | Y | Y | Y | Y | Y | Y |
| IBM Z Integrated Information Processors (zIIPs) | Y | Y | Y | Y | Y | Y |
| Transactional Execution | Y | Y | Y | Y ^{gh} | Y ^{gh} | Y ^{gh} |
| Java Exploitation of Transactional Execution | Y | Y | Y | Y ^h | Y ^h | Y ^h |
| Simultaneous multithreading (SMT) | Y | Y | Y | Y ⁱ | Y ⁱ | Y ⁱ |
| Single Instruction Multiple Data (SIMD) | Y | Y | Y | Y ^{hj} | Y ^{hj} | Y ^{hj} |
| Hardware decimal floating point ^k | Y | Y | Y | Y | Y | Y |
| 2 GB large page support | Y | Y | Y | N | N | N |
| Large page (1 MB) support | Y | Y | Y | Y ^h | Y ^h | Y ^h |
| Out-of-order execution | Y ^l | Y | Y | Y | Y | Y |
| CPUMF (CPU measurement facility) for z15 | Y | Y | Y | Y | Y | Y |
| Enhanced flexibility for Capacity on Demand (CoD) | Y | Y | Y | Y | Y | Y |
| IBM Virtual Flash Memory (VFM) | Y | Y | Y | N | N | N |
| 1 MB pageable large pages ^m | Y | Y | Y | N | N | N |
| Guarded Storage Facility (GSF) | Y | Y | Y | Y ^h | Y ^h | Y ^h |

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|--|----------------|----------------|----------------|----------------|----------------|----------------|
| Instruction Execution Protection (IEP) | Y | Y | Y | Y ^h | Y ^h | Y ^h |
| Co-processor Compression Enhancements ⁿ (CMPSC) | Y | Y | Y | N | N | N |
| System Recovery Boost | Y ^o | Y ^o | N | Y ^p | Y ^p | N |
| IBM Integrated Accelerator for zEDC ^q (on-chip compression) | Y | Y | Y ^r | N ^s | N ^s | N ^s |
| IBM Integrated Accelerator for Z SORT | Y ^o | Y ^o | N | Y ^h | Y ^h | Y ^h |

- a. PTFs might be required for toleration support or exploitation of z15 features and functions.
- b. 190-way without multithreading; 128-way with multithreading enabled.
- c. 80-way without multithreading; 40-way with multithreading enabled.
- d. 64-way without multithreading; 32-way with multithreading enabled
- e. A total of 40 TB of real storage is supported per z15 T01 server while z15 T02 supports up to 16 TB per server.
- f. z15 T02 configuration supports up to 6 CPs and up to 65 IFIs or ICFs. It is recommended that the number of logical processors assigned to an LPAR to be less or equal to the number of physical processors available on the system.
- g. Guests are informed that TX facility is available for use.
- h. Guest exploitation support.
- i. Dynamic SMT with z15.
- j. Guests are informed that SIMD is available for use.
- k. Packed decimal conversion support.
- l. Enhanced OoO execution for z15 - see 3.4.3, "Out-of-Order execution" on page 102.
- m. With IBM Virtual Flash Memory for middleware exploitation.
- n. With PTFs for exploitation.
- o. With PTFs.
- p. Support for subcapacity CPs allocated to the z/VM LPAR.
- q. IBM zEnterprise Data Compression Express replacement
- r. Compatibility (read only) - with PTFs
- s. Transparent for Guest support use of the gzip acceleration; guest support for z/OS Storage Compression

The supported base CPC functions for z/VSE, z/TPF, and Linux on Z are listed in Table 7-4.

Table 7-4 Supported base CPC functions for z/VSE, z/TPF, and Linux on Z

| Function ^a | z/VSE V6R2 | z/TPF V1R1 | Linux on Z ^b |
|---|------------|------------|-------------------------|
| z15 servers | Y | Y | Y |
| z15 T01 Maximum processor unit (PUs) per system image | 10 | 86 | 190 ^c |
| z15 T02 Maximum processor unit (PUs) per system image | 6 | 6 | 65 |
| z15 T01 Maximum main storage size ^d | 32 GB | 4 TB | 16 TB ^e |
| z15 T01 85 LPARs | Y | Y | Y |
| z15 T02 Maximum main storage size ^f | 32 GB | 4 TB | 8 TB ^g |
| z15 T02 40 LPARs | Y | Y | Y |
| Separate LPAR management of PUs | Y | Y | Y |
| Dynamic PU add | Y | N | Y |
| Dynamic LPAR memory upgrade | N | N | Y |
| LPAR group absolute capping | Y | N | N |

| Function ^a | z/VSE V6R2 | z/TPF V1R1 | Linux on Z ^b |
|--|----------------|----------------|-------------------------|
| Program-directed re-IPL | Y | N | Y |
| HiperDispatch | N | N | Y |
| IBM Z Integrated Information Processors (zIIPs) | N | N | N |
| Transactional Execution | N | N | Y |
| Java Exploitation of Transactional Execution | N | N | Y |
| Simultaneous multithreading (SMT) | N | N | Y |
| Single Instruction Multiple Data (SIMD) | Y | N | Y |
| Hardware decimal floating point ^h | N | N | Y |
| 2 GB large page support | N | Y | Y |
| Large page (1 MB) support | Y | Y | Y |
| Out-of-order execution | Y | Y | Y |
| CPUMF (CPU measurement facility) for z15 | N | Y | N ⁱ |
| Enhanced flexibility for CoD | N | N | N |
| IBM Virtual Flash Memory (VFM) | N | N | Y |
| Guarded Storage Facility (GSF) | N | N | Y |
| Instruction Execution Protection (IEP) | N | N | Y |
| Co-processor Compression Enhancements | N | N | N |
| System Recovery Boost | Y ^j | Y ^j | N |
| Secure Boot (code integrity check) | N/A | N/A | Y ^k |
| Secure Execution Support for Linux | N/A | N/A | Y ^l |
| IBM Integrated Accelerator for zEDC ^m (on-chip compression) | N | N | Y ⁿ |
| IBM Integrated Accelerator for Z SORT | N | N | N |

a. PTFs might be required for toleration support or exploitation of z15 features and functions.

b. Support statement varies based on Linux on Z distribution and release.

c. For SLES12/RHEL7/Ubuntu 16.04 and later, Linux kernel supports 256 cores without SMT and 128 cores with SMT (= 256 threads).

d. A total of 40 TB of real storage is supported per z15 T01 server.

e. z15 T01 supports 16 TB per LPAR. Linux on Z releases can support up to 64 TB of memory.

f. A total of 16 TB of real storage is supported per z15 T02 server.

g. z15 T02 supports up to 8 TB per LPAR.

h. Packed decimal conversion support.

i. IBM is working with its Linux distribution Business Partners to provide this feature.

j. Subcapacity CP speed boost (no zIIP boost)

k. For SCSI IPL

l. For second level guests running under KVM

m. IBM zEnterprise Data Compression Express replacement.

n. Requires Linux kernel exploitation support for gzip/zlib compression.

7.3.2 Coupling and clustering

The supported coupling and clustering functions for z/OS and z/VM are listed in Table 7-5.

Note: z/OS V2R1 support has ended on as of September 2018. No new function is provided for exploiting the new HW features (toleration support only). Although extended (fee-based) support for z/OS 2.1 can be obtained, support for z/OS 2.1 is not covered extensively in this document.

Table 7-5 Supported coupling and clustering functions for z/OS and z/VM

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|--|----------------|----------------|----------------|----------------|----------------|----------------|
| Server Time Protocol (STP) | Y | Y | Y | Y | Y | Y |
| CFCC Level 24 ^b | Y | Y | Y | Y ^g | Y ^g | Y ^g |
| CFCC Level 24 Fair Latch Manager | Y | Y ^c | Y ^c | Y ^g | Y ^g | Y ^g |
| Message Path System ID (SYID) resiliency | Y | Y ^c | Y ^c | Y ^g | Y ^g | Y ^g |
| CF Monopolization Avoidance | Y ^c | Y ^c | Y ^c | Y ^g | Y ^g | Y ^g |
| CFCC Level 23 ^d | Y | Y | Y | Y ^g | Y ^g | Y ^g |
| CFCC Level 22 ^e | Y | Y | Y | Y ^f | Y ^g | Y ^g |
| CFCC Level 22 Coupling Thin Interrupts | Y | Y | Y | N | N | N |
| CFCC Level 22 Large Memory support | Y | Y | Y | N | N | N |
| CFCC Level 22 Support for 256 Coupling CHPIDs per CPC | Y | Y | Y | Y ^g | Y ^g | Y ^g |
| CFCC Level 22 Coupling Facility Processor Scalability | Y | Y | Y | Y ^g | Y ^g | Y ^g |
| CFCC Level 22 List Notification Enhancements | Y | Y | Y | Y ^g | Y ^g | Y ^g |
| CFCC Level 22 Encryption Support | Y | Y | N ^h | Y ^g | Y ^g | Y ^g |
| CFCC Level 22 Exploitation of VFM (Virtual Flash Memory) | Y | Y | Y | N | N | N |
| RMF coupling channel reporting | Y | Y | Y | N | N | N |
| Coupling over InfiniBand CHPID type CIB ⁱ | Y | Y | Y | N | N | N |
| InfiniBand coupling links 12x at a distance of 150 m (492 ft.) ⁱ | Y | Y | Y | N | N | N |
| InfiniBand coupling links 1x at an unrepeated distance of 10 km (6.2 miles) ⁱ | Y | Y | Y | N | N | N |
| Integrated Coupling Adapter (ICA-SR) links CHPID CS5 | Y | Y | Y | Y ^j | Y ^j | Y ^j |
| Coupling Express LR (CE LR) CHPID CL5 | Y | Y | Y | Y ^j | Y ^j | Y ^j |
| z/VM Dynamic I/O support for InfiniBand CHPIDs ⁱ | - | - | - | Y ^j | Y ^j | Y ^j |
| z/VM Dynamic I/O support for ICA SR CHPIDs | - | - | - | Y ^j | Y ^j | Y ^j |
| Asynchronous CF Duplexing for lock structures | Y | Y | Y | Y ^g | Y ^g | Y ^g |
| Asynchronous cross-invalidate (XI) for CF cache structures ^k | Y | Y ^l | Y ^l | Y ^g | Y ^g | Y ^g |
| Dynamic I/O activation for stand-alone CF CPCs ^m | Y ⁿ | Y ⁿ | Y ⁿ | Y ⁿ | Y ⁿ | Y ⁿ |

a. PTFs might be required for toleration support or exploitation of z15 features and functions.

- b. CFCC Level 24 with Driver 41 (z15).
- c. Requires z/OS XCF/XES toleration APAR.
- d. CFCC Level 23 with Driver 36 (z14).
- e. CFCC Level 22 with Driver 32 (z13/z13s).
- f. Virtual guest coupling.
- g. Virtual guest coupling.
- h. Toleration support ("locking out" down level systems that cannot use encrypted structure) provided for z/OS 2.1 and later.
- i. InfiniBand Coupling Links are *not* supported on z15 and z14 ZR1.
- j. To define, modify, and delete CHPID type CS5 when z/VM is the controlling LPAR for dynamic I/O.
- k. Requires data manager support (Db2 fixes).
- l. Requires fixes for APAR OA54688 for exploitation.
- m. Managing dynamic I/O activation for stand-alone CF CPCs on a z15 requires HMC 2.15.0 (Driver 41)
- n. Requires HMC 2.14.1(Driver 36) or newer and various OS fixes (HCD, HCM, IOS, IOCP).

In addition to operating system support that is listed in Table 7-5 on page 261, Server Time Protocol is supported on z/TPF V1R1 and Linux on Z. Also, CFCC Level 22, Level 23, and Level 24 are supported for z/TPF V1R1.

Storage connectivity

The supported storage connectivity functions for z/OS and z/VM are listed Table 7-6.

Table 7-6 Supported storage connectivity functions for z/OS and z/VM

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|---|----------------|----------------|----------------|----------------|----------------|----------------|
| zHyperLink Express Read Support | Y | Y ^b | Y ^b | N | N | N |
| zHyperLink Express Write Support | Y | Y ^b | Y ^b | N | N | N |
| The 63.75-K subchannels | Y | Y | Y | Y | Y | Y |
| Six logical channel subsystems (LCSSs) | Y | Y | Y | Y | Y | Y |
| Four subchannel set per LCSS | Y | Y | Y | Y ^c | Y ^c | Y ^c |
| Health Check for FICON Dynamic routing | Y | Y | Y | N | N | N |
| z/VM Dynamic I/O support for FICON Express16SA ^e FC and FCP CHPIDs | - | - | - | Y | Y | Y |
| z/VM Dynamic I/O support for FICON Express16S+ FC and FCP CHPIDs | - | - | - | Y | Y | Y |
| CHPID (Channel-Path Identifier) type FC | | | | | | |
| Extended distance FICON ^d | Y | Y | Y | Y | Y | Y |
| FICON Express16SA ^e for support of zHPF (IBM Z High-Performance FICON) | Y | Y | Y | Y | Y | Y |
| IBM Fibre Channel Endpoint Security ^f | Y ^g | Y ^g | Y ^g | Y ^g | Y ^g | Y ^g |
| FICON Express16S+ for support of zHPF (IBM Z High-Performance FICON) | Y | Y | Y | Y ^h | Y ^h | Y ^h |
| FICON Express16S for support of zHPF | Y | Y | Y | Y ^h | Y ^h | Y ^h |
| FICON Express8S for support of zHPF | Y | Y | Y | Y ^h | Y ^h | Y ^h |
| MIDAW (Modified Indirect Data Address Word) | Y | Y | Y | Y ^h | Y ^h | Y ^h |
| zDAC (z/OS Discovery and Auto-Configuration) | Y | Y | Y | N | N | N |

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|---|-----------|-----------|-----------|----------------|----------------|----------------|
| FICON Express16SA ^e when using FICON or CTC (channel-to-channel) | Y | Y | Y | Y ⁱ | Y ⁱ | Y ⁱ |
| FICON Express16S+ when using FICON or CTC (channel-to-channel) | Y | Y | Y | Y ⁱ | Y ⁱ | Y ⁱ |
| FICON Express16S when using FICON or CTC | Y | Y | Y | Y ⁱ | Y ⁱ | Y ⁱ |
| FICON Express8S when using FICON or CTC | Y | Y | Y | Y ⁱ | Y ⁱ | Y ⁱ |
| Global resource serialization (GRS) FICON CTC toleration | Y | Y | Y | N | N | N |
| IPL from an alternative subchannel set | Y | Y | Y | N | N | N |
| 32 K subchannels for the FICON Express16SA ^e | Y | Y | Y | Y | Y | Y |
| 32 K subchannels for the FICON Express16S+ | Y | Y | Y | Y | Y | Y |
| 32 K subchannels for the FICON Express16S | Y | Y | Y | Y | Y | Y |
| Request node identification data | Y | Y | Y | N | N | N |
| FICON link incident reporting | Y | Y | Y | N | N | Y |
| CHPID (Channel-Path Identifier) type FCP | | | | | | |
| FICON Express16SA ^e for support of SCSI devices | - | - | - | Y | Y | Y |
| FICON Express16S+ for support of SCSI devices | - | - | - | Y | Y | Y |
| FICON Express16S for support of SCSI devices | - | - | - | Y | Y | Y |
| FICON Express8S for support of SCSI devices | - | - | - | Y | Y | Y |
| FICON Express16SA ^e support of hardware data router | - | - | - | Y ^h | Y ^h | Y ^h |
| FICON Express16S+ support of hardware data router | - | - | - | Y ^h | Y ^h | Y ^h |
| FICON Express16S support of hardware data router | - | - | - | Y ^h | Y ^h | Y ^h |
| FICON Express8S support of hardware data router | - | - | - | Y ^h | Y ^h | Y ^h |
| FICON Express16SA ^e T10-DIF support | - | - | - | Y ^h | Y ^h | Y ^h |
| FICON Express16S+ T10-DIF support | - | - | - | Y ^h | Y ^h | Y ^h |
| FICON Express16S T10-DIF support | - | - | - | Y ^h | Y ^h | Y ^h |
| FICON Express8S T10-DIF support | - | - | - | Y ^h | Y ^h | Y ^h |
| Increased performance for the FCP protocol | - | - | - | Y | Y | Y |
| N_Port ID Virtualization (NPIV) | - | - | - | Y | Y | Y |
| Worldwide port name tool | - | - | - | Y | Y | Y |

a. PTFs might be required for toleration support or exploitation of z15 features and functions.

b. With PTFs

c. For specific Geographically Dispersed Parallel Sysplex (GDPS) usage only.

d. Transparent to operating systems.

e. FICON Express16SA is *not* supported on z15 T02.

f. FC 1146 (optional) is available for z15 T01 only. Requires select DS8000 storage, CPACF enablement and FICON Express16SA (endpoint authentication and link encryption) or FICON Express16S+ (endpoint authentication only).

- g. Feature is OS independent, but requires OS support for displaying configuration and monitoring (PTFs required).
- h. For guest use.
- i. CTC channel type not supported when CPC is managed in DPM mode.

The supported storage connectivity functions for z/VSE, z/TPF, and Linux on Z are listed in Table 7-7.

Table 7-7 Supported storage connectivity functions for z/VSE, z/TPF, and Linux on Z

| Function ^a | z/VSE V6R2 | z/TPF V1R1 | Linux on Z ^b |
|---|----------------|----------------|-------------------------|
| zHyperLink Express | - | - | - |
| The 63.75-K subchannels | N | N | Y |
| Six logical channel subsystems (LCSSs) | Y | N | Y |
| Four subchannel set per LCSS | Y | N | Y |
| Health Check for FICON Dynamic routing | N | N | N |
| CHPID (Channel-Path Identifier) type FC | | | |
| Extended distance FICON ^c | Y | Y | Y |
| FICON Express16SA ^d for support of zHPF (IBM Z High-Performance FICON) | Y ^e | Y | Y |
| IBM Fibre Channel Endpoint Security ^f | Y ^g | Y ^g | Y ^g |
| FICON Express16S+ for support of zHPF (IBM Z High-Performance FICON) | Y ^e | Y | Y |
| FICON Express16S for support of zHPF | Y | Y | Y |
| FICON Express8S for support of zHPF | Y | Y | Y |
| MIDAW (Modified Indirect Data Address Word) | N | N | N |
| FICON Express16SA ^d when using FICON or CTC (channel-to-channel) | Y | Y | Y ^h |
| FICON Express16S+ when using FICON or CTC | Y | Y | Y ^h |
| FICON Express16S when using FICON or CTC | Y | Y | Y ^h |
| FICON Express8S when using FICON or CTC | Y | Y | Y ^h |
| IPL from an alternative subchannel set | N | N | N |
| 32 K subchannels for the FICON Express16SA ^d | N | N | Y |
| 32 K subchannels for the FICON Express16S+ | N | N | Y |
| 32 K subchannels for the FICON Express16S | N | N | Y |
| Request node identification data | N | N | N |
| FICON link incident reporting | N | N | N |
| CHPID (Channel-Path Identifier) type FCP | | | |
| FICON Express16SA ^d for support of SCSI devices | Y | - | Y |
| FICON Express16SA ^d for support of SCSI devices | Y | - | Y |
| FICON Express16S+ for support of SCSI devices | Y | - | Y |

| Function ^a | z/VSE V6R2 | z/TPF V1R1 | Linux on Z ^b |
|--|------------|------------|-------------------------|
| FICON Express16S for support of SCSI devices | Y | - | Y |
| FICON Express8S for support of SCSI devices | Y | - | Y |
| FICON Express16SA ^d support of hardware data router | N | N | Y |
| FICON Express16S+ support of hardware data router | N | N | Y |
| FICON Express16S support of hardware data router | N | N | Y |
| FICON Express8S support of hardware data router | N | N | Y |
| FICON Express16SA ^d T10-DIF support | N | N | Y |
| FICON Express16S+ T10-DIF support | N | N | Y |
| FICON Express16S T10-DIF support | N | N | Y |
| FICON Express8S T10-DIF support | N | N | Y |
| Increased performance for the FCP protocol | Y | - | Y |
| N_Port ID Virtualization (NPIV) | Y | N | Y |
| Worldwide port name tool | - | - | Y |

a. PTFs might be required for toleration support or exploitation of z15 features and functions.

b. Support statement varies based on Linux on Z distribution and release.

c. Transparent to operating systems.

d. FICON Express16SA is *not* supported on z15 T02.

e. Supported on z/VSE V6.2 with PTFs.

f. FC 1146 (optional) is available for z15 T01 only. Requires select DS8000 storage, CPACF enablement and FICON Express16SA (both endpoint authentication and link encryption) or FICON Express16S+ (endpoint authentication only).

g. Feature is OS independent (transparent to OS); OS support is only needed for displaying configuration and monitoring (fixes may be required).

h. CTC channel type not supported when CPC is managed in DPM mode.

7.3.3 Network connectivity

The supported network connectivity functions for z/OS and z/VM are listed in Table 7-8.

Note: z/OS V2R1 is End of Support as of September 2018. No new function is provided for exploiting HW features and functions introduced with IBM z15 (toleration support only). Although extended (fee-based) support for z/OS 2.1 can be obtained, support for z/OS 2.1 is not covered extensively in this document.

Table 7-8 Supported network connectivity functions for z/OS and z/VM

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|--|-----------|-----------|-----------|----------------|----------------|----------------|
| Checksum offload for IPV6 packets | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| Checksum offload for LPAR-to-LPAR traffic with IPv4 and IPv6 | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| Querying and displaying an OSA configuration | Y | Y | Y | Y | Y | Y |
| QDIO data connection isolation for z/VM | - | - | - | Y | Y | Y |
| QDIO interface isolation for z/OS | Y | Y | Y | - | - | - |
| QDIO OLM (Optimized Latency Mode) | Y | Y | Y | - | - | - |
| Adapter interruptions for QDIO | Y | N | N | Y | Y | Y |
| QDIO Diagnostic Synchronization | Y | Y | Y | N | N | N |
| IWQ (Inbound Workload Queuing) for OSA | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| VLAN management enhancements | Y | Y | Y | Y ^c | Y ^c | Y ^c |
| GARP VLAN Registration Protocol | Y | Y | Y | Y | Y | Y |
| Link aggregation support for z/VM | - | - | - | Y | Y | Y |
| Multi-vSwitch Link Aggregation | - | - | - | Y | Y | Y |
| Large send for IPV6 packets | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| z/VM Dynamic I/O Support for OSA-Express6S OSD CHPIDs | - | - | - | Y | Y | Y |
| z/VM Dynamic I/O Support for OSA-Express7S ^d OSD CHPIDs | - | - | - | Y | Y | Y |
| OSA Dynamic LAN idle | Y | Y | Y | N | N | N |
| OSA Layer 3 virtual MAC for z/OS environments | Y | Y | Y | - | - | - |
| Network Traffic Analyzer | Y | Y | Y | N | N | N |
| Hipersockets | | | | | | |
| HiperSockets ^e | Y | Y | Y | Y | Y | Y |
| 32 HiperSockets | Y | Y | Y | Y | Y | Y |
| HiperSockets Completion Queue | Y | Y | Y | Y | Y | Y |
| HiperSockets Virtual Switch Bridge | - | - | - | Y | Y | Y |
| HiperSockets Multiple Write Facility | Y | Y | Y | N | N | N |
| HiperSockets support of IPV6 | Y | Y | Y | Y | Y | Y |

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|--|-----------|----------------|----------------|----------------|----------------|----------------|
| HiperSockets Layer 2 support | Y | Y | Y | Y | Y | Y |
| SMC-D and SMC-R | | | | | | |
| SMC-D ^f over ISM (Internal Shared Memory) | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| 10GbE RoCE ^g Express | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| 25GbE and 10GbE RoCE Express2 for SMC-R | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| 25GbE and 10GbE RoCE Express2 and 2.1 for Ethernet communications ^h including Single Root I/O Virtualization (SR-IOV) | N | N | N | Y ^b | Y ^b | Y ^b |
| z/VM Dynamic I/O support for RoCE Express2 and 2.1 | - | - | - | Y | Y | Y |
| Shared RoCE environment | Y | Y | Y | Y | Y | Y |
| Open Systems Adapter (OSA)ⁱ | | | | | | |
| OSA-Express7S ^d 1000BASE-T Ethernet CHPID type OSC | Y | Y | Y | Y | Y | Y |
| OSA-Express6S 1000BASE-T Ethernet CHPID type OSC | Y | Y | Y | Y | Y | Y |
| OSA-Express5S 1000BASE-T Ethernet CHPID type OSC | Y | Y | Y | Y | Y | Y |
| OSA-Express7S 25-Gigabit Ethernet Short Reach (SR and SR1.1 ^d) CHPID type OSD | Y | Y ^j | Y ^j | Y | Y ^k | Y ^k |
| OSA-Express7S ^d 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD | Y | Y | Y | Y | Y | Y |
| OSA-Express6S 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD | Y | Y | Y | Y | Y | Y |
| OSA-Express5S 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD | Y | Y | Y | Y | Y | Y |
| OSA-Express7S ^d Gigabit Ethernet LX and SX CHPID type OSD | Y | Y | Y | Y | Y | Y |
| OSA-Express7S ^d Gigabit Ethernet LX and SX CHPID type OSC | Y | Y | Y | Y | Y | Y |
| OSA-Express6S Gigabit Ethernet LX and SX CHPID type OSD | Y | Y | Y | Y | Y | Y |
| OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD | Y | Y | Y | Y | Y | Y |
| OSA-Express6S 1000BASE-T Ethernet CHPID type OSD | Y | Y | Y | Y | Y | Y |
| OSA-Express5S 1000BASE-T Ethernet CHPID type OSD | Y | Y | Y | Y | Y | Y |
| OSA-Express7S ^d 1000BASE-T Ethernet CHPID type OSE | Y | Y | Y | Y | Y | Y |

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|--|-----------|-----------|-----------|-----------|-----------|-----------|
| OSA-Express6S 1000BASE-T Ethernet CHPID type OSE | Y | Y | Y | Y | Y | Y |
| OSA-Express5S 1000BASE-T Ethernet CHPID type OSE | Y | Y | Y | Y | Y | Y |

- a. PTFs might be required for toleration support or exploitation of z15 features and functions.
- b. For guest use or exploitation.
- c. Support of guests is transparent to z/VM if the device is directly connected to the guest (pass through).
- d. z15 T02 only supports OSA-Express7S 25GbE SR. All other OSA-Express7S features are not supported on z15 T02. z15 T01 supports all OSA-Express7S features.
- e. On z15, the CHPID statement of HiperSockets devices requires the keyword VCHID. Therefore, the z15 IOCP definitions must be migrated to support the HiperSockets definitions (CHPID type IQD). VCHID specifies the virtual channel identification number that is associated with the channel path (valid range is 7C0 - 7FF). VCHID is not valid on IBM Z servers before z13.
- f. Shared Memory Communications - Direct Memory Access.
- g. Remote Direct Memory Access (RDMA) over Converged Ethernet.
- h. Does not require a peer OSA.
- i. Supported CHPID types: OSC, OSD, OSE, and OSM.
- j. Require PTFs for APARs OA55256 (IBM VTAM®) and PI95703 (TCP/IP).
- k. Require PTF for APAR PI99085.

The supported network connectivity functions for z/VSE, z/TPF, and Linux on Z are listed in Table 7-9.

Table 7-9 Supported network connectivity functions for z/VSE, z/TPF, and Linux on Z

| Function ^a | z/VSE V6R2 | z/TPF V1R1 | Linux on Z ^b |
|--|------------|------------|-------------------------|
| Checksum offload for IPV6 packets | N | N | Y |
| Checksum offload for LPAR-to-LPAR traffic with IPv4 and IPv6 | N | N | Y |
| Adapter interruptions for QDIO | Y | N | Y |
| QDIO Diagnostic Synchronization | N | N | N |
| IWQ (Inbound Workload Queuing) for OSA | N | N | N |
| VLAN management enhancements | N | N | N |
| GARP VLAN Registration Protocol | N | N | Y ^c |
| Link aggregation support for z/VM | N | N | N |
| Multi-vSwitch Link Aggregation | N | N | N |
| Large send for IPV6 packets | N | N | Y |
| z/VM Dynamic I/O Support for OSA-Express6S and OSA-Express 7S OSD CHPIDs | N | N | N |
| OSA Dynamic LAN idle | N | N | N |
| Hipersockets | | | |
| HiperSockets ^d | Y | N | Y |
| 32 HiperSockets | Y | N | Y |
| HiperSockets Completion Queue | Y | N | Y |

| Function ^a | z/VSE V6R2 | z/TPF V1R1 | Linux on Z ^b |
|--|------------|------------|-----------------------------|
| HiperSockets Virtual Switch Bridge | - | - | Y ^e |
| HiperSockets Multiple Write Facility | N | N | N |
| HiperSockets support of IPV6 | Y | N | Y |
| HiperSockets Layer 2 support | Y | N | Y |
| HiperSockets Network Traffic Analyzer for Linux on Z | N | N | Y |
| SMC-D and SMC-R | | | |
| SMC-D ^f over ISM (Internal Shared Memory) | N | N | Y ^g |
| 10GbE RoCE ^h Express | N | N | Y ^g ⁱ |
| 25GbE and 10GbE RoCE Express2 for SMC-R | N | N | Y ^g ⁱ |
| 25GbE and 10GbE RoCE Express2 for Ethernet communications ^j including Single Root I/O Virtualization (SR-IOV) | N | N | Y ^g ⁱ |
| Shared RoCE environment | N | N | Y |
| Open Systems Adapter (OSA) | | | |
| OSA-Express7S ^k 1000BASE-T Ethernet CHPID type OSC | Y | Y | - |
| OSA-Express6S 1000BASE-T Ethernet CHPID type OSC | Y | Y | - |
| OSA-Express5S 1000BASE-T Ethernet CHPID type OSC | Y | Y | - |
| OSA-Express7S ^k Gigabit Ethernet (GbE) LX and SX CHPID type OSC | Y | Y | - |
| OSA-Express7S 25-Gigabit Ethernet Short Reach (SR and SR1.1 ^k) CHPID type OSD | Y | Y | Y |
| OSA-Express7S ^k 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD | Y | Y | Y |
| OSA-Express6S 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD | Y | Y | Y |
| OSA-Express5S 10-Gigabit Ethernet Long Reach (LR) and Short Reach (SR) CHPID type OSD | Y | Y | Y |
| OSA-Express7S ^k Gigabit Ethernet (GbE) LX and SX CHPID type OSD | Y | Y | Y |
| OSA-Express6S Gigabit Ethernet LX and SX CHPID type OSD | Y | Y | Y |
| OSA-Express5S Gigabit Ethernet LX and SX CHPID type OSD | Y | Y | Y |
| OSA-Express7S ^k 1000BASE-T Ethernet CHPID type OSD | Y | Y | Y |
| OSA-Express6S 1000BASE-T Ethernet CHPID type OSD | Y | Y | Y |

| Function ^a | z/VSE V6R2 | z/TPF V1R1 | Linux on Z ^b |
|--|---------------|---------------|----------------------------|
| OSA-Express5S 1000BASE-T Ethernet CHPID type OSD | Y | Y | Y |
| OSA-Express7S ^k 1000BASE-T Ethernet CHPID type OSE | Y | N | N |
| OSA-Express6S 1000BASE-T Ethernet CHPID type OSE | Y | N | N |
| OSA-Express5S 1000BASE-T Ethernet CHPID type OSE | Y | N | N |

a. PTFs might be required for toleration support or exploitation of z15 features and functions.

b. Support statement varies based on Linux on Z distribution and release.

c. By using VLANs.

d. On z15, the CHPID statement of HiperSockets devices requires the keyword VCHID.

Therefore, the z15 IOCP definitions must be migrated to support the HiperSockets definitions (CHPID type IQD). VCHID specifies the virtual channel identification number that is associated with the channel path (valid range is 7C0 - 7FF). VCHID is not valid on IBM Z servers before z13.

e. Applicable to guest operating systems.

f. Shared Memory Communications - Direct Memory Access.

g. SMC-R and SMC-D are supported on Linux kernel; see:

<https://linux-on-z.blogspot.com/p/smc-for-linux-on-ibm-z.html>

h. Remote Direct Memory Access (RDMA) over Converged Ethernet.

i. Linux can also use RocE Express as a standard NIC (Network Interface Card) for Ethernet.

j. Does not require a peer OSA.

k. z15 T02 only supports OSA-Express7S 25GbE SR. All other OSA-Express7S features are not supported on z15 T02. z15 T01 supports all OSA-Express7S features.

7.3.4 Cryptographic functions

The z15 supported cryptography functions for z/OS and z/VM are listed in Table 7-10.

Note: z/OS V2R1 is End of Support as of September 2018. No new function is provided for exploiting HW features and functions introduced with IBM z15 (toleration support only). Although extended (fee-based) support for z/OS 2.1 can be obtained, support for z/OS 2.1 is not covered extensively in this document.

Table 7-10 Supported cryptography functions for z/OS and z/VM

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|--|-----------|-----------|----------------|----------------|----------------|----------------|
| CP Assist for Cryptographic Function (CPACF) | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| CPACF greater than 16 Domain Support | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| CPACF AES-128, AES-192, and AES-256 | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| CPACF SHA-1, SHA-224, SHA-256, SHA-384, and SHA-512 | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| CPACF protected key | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| Crypto Express7S | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| Crypto Express7S Support for Visa Format Preserving Encryption | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| Crypto Express7S Support for Coprocessor in PCI-HSM Compliance Mode ^d | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| z15 T01 Crypto Express7S supporting up to 85 domains | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| z15 T02 Crypto Express7S supporting up to 40 domains | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| Crypto Express6S | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| Crypto Express6S Support for Visa Format Preserving Encryption | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| Crypto Express6S Support for Coprocessor in PCI-HSM Compliance Mode ^e | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| Crypto Express6S supporting up to 85 domains | Y | Y | Y ^c | Y ^b | Y ^b | Y ^b |
| Crypto Express5S | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| Crypto Express5S spouting up to 85 domains | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| Elliptic Curve Cryptography (ECC) | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode | Y | Y | Y | Y ^b | Y ^b | Y ^b |
| z/OS Data Set Encryption | Y | Y | Y | - | - | - |
| z/VM Encrypted paging support | - | - | - | Y | Y | Y |
| RMF Support for Crypto Express7 | Y | Y | Y | - | - | - |
| RMF Support for Crypto Express6 | Y | Y | Y | - | - | - |

| Function ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 |
|---|-----------|-----------|-----------|-----------|-----------|-----------|
| z/OS encryption readiness technology (zERT) | Y | Y | N | - | - | - |

- a. PTFs might be required for toleration support or exploitation of z15 features and functions.
- b. For guest use or exploitation.
- c. A web deliverable is required. For more information and to download the deliverable, see [the z/OS downloads page](#) of the IBM IT infrastructure website.
- d. Requires TKE 9.1 or newer.
- e. Requires TKE 9.1 or newer.

The z15 supported cryptography functions for z/VSE, z/TPF, and Linux on Z are listed in Table 7-11.

Table 7-11 Supported cryptography functions for z/VSE, z/TPF, and Linux on Z

| Function ^a | z/VSE V6R2 | z/TPF V1R1 | Linux on Z ^b |
|--|------------|----------------|-------------------------|
| CP Assist for Cryptographic Function (CPACF) | Y | Y | Y |
| CPACF greater than 16 Domain Support | Y | N | Y |
| CPACF AES-128, AES-192, and AES-256 | Y | Y ^c | Y |
| CPACF SHA-1, SHA-224, SHA-256, SHA-384, and SHA-512 | Y | Y ^d | Y |
| CPACF protected key | N | N | N |
| Crypto Express7S | Y | Y | Y |
| Crypto Express6S Support for Visa Format Preserving Encryption | N | N | N |
| Crypto Express7S Support for Coprocessor in PCI-HSM Compliance Mode ^e | N | N | N |
| Crypto Express7S supporting up to 85 domains | Y | N | Y |
| Crypto Express6S | Y | Y | Y |
| Crypto Express6S Support for Visa Format Preserving Encryption | N | N | N |
| Crypto Express6S Support for Coprocessor in PCI-HSM Compliance Mode ^e | N | N | N |
| Crypto Express6S supporting up to 85 domains | Y | N | Y |
| Crypto Express5S | Y | Y | Y |
| Crypto Express5S spouting up to 85 domains | Y | N | Y |
| Elliptic Curve Cryptography (ECC) | Y | N | Y |
| Secure IBM Enterprise PKCS #11 (EP11) coprocessor mode | N | N | Y |
| z/VM Encrypted paging support | N | - | N |
| z/TPF transparent database encryption | - | Y | - |

- a. PTFs might be required for toleration support or exploitation of z15 features and functions.
- b. Support statement varies based on Linux on Z distribution and release.
- c. z/TPF supports only AES-128 and AES-256.
- d. z/TPF supports only SHA-1 and SHA-256.
- e. Requires TKE 9.1 or newer.

7.4 Support by features and functions

This section addresses operating system support by function. Only the currently in-support releases are covered.

Tables in this section use the following convention:

- ▶ N/A: Not applicable
- ▶ NA: Not available

7.4.1 LPAR Configuration and Management

A single system image can control several processor units, such as CPs, zIIPs, or IFLs.

Note: z/OS V2R1 is End of Support as of September 2018. No new function is provided for exploiting HW features and functions introduced with IBM z15 (toleration support only). Although extended (fee-based) support for z/OS 2.1 can be obtained, support for z/OS 2.1 is not covered extensively in this document.

Maximum number of PUs per system image

The maximum number of PUs that is supported by each operating system image and by special-purpose LPARs are listed in Table 7-12.

Table 7-12 Maximum number of PUs per system image

| Operating system | Maximum number of PUs per system image |
|--------------------------|--|
| z/OS V2R4 | 256 ^{abc} |
| z/OS V2R3 | 256 ^{abc} |
| z/OS V2R2 | 256 ^{abc} |
| z/VM V7R2 | 80 ^d |
| z/VM V7R1 | 80 ^e |
| z/VM V6R4 | 64 ^f |
| z/VSE V6.2 and later | z/VSE Turbo Dispatcher can use up to 4 CPs, and tolerates up to 10-way LPARs |
| z/TPF V1R1 | 86 CPs |
| CFCC Level 24 | 16 CPs or ICFs CPs and ICFs cannot be mixed. |
| Linux on Z | SUSE Linux Enterprise Server 12 and later: 256 CPs or IFLs. SUSE Linux Enterprise Server 11: 64 CPs or IFLs. Red Hat RHEL 7 and later: 256 CPs or IFLs. Red Hat RHEL 6: 64 CPs or IFLs. Ubuntu 16.04 LTS and 18.04 LTS: 256 CPs or IFLs. |
| KVM Hypervisor | The KVM hypervisor is offered with the following Linux distributions -- 256CPs or IFLs--: SLES 12 SP4 and later. RHEL 7.5 with kernel-alt package (kernel 4.14). Ubuntu 16.04 LTS and Ubuntu 18.04 LTS. |
| Secure Service Container | 80 |

| Operating system | Maximum number of PUs per system image |
|------------------------|--|
| GDPS Virtual Appliance | 80 |

- a. z15 T01 LPARs support 190-way without multithreading; 128-way with multithreading (SMT).
- b. Total characterizable PUs, including zIIPs and CPs.
- c. z15 T02 supports up to 6 CPs and up to 65 IFLs. It is not recommended to define more logical processors to an image that the actual number of physical PUs available on the system.
- d. 80-way without multithreading and 40-way with multithreading enabled
- e. An 80-way without multithreading and 40-way with multithreading enabled. Requires PTF for APAR VM66265. Supported on z14 and z15.
- f. A 64-way without multithreading and 32-way with multithreading enabled.

Maximum main storage size

The maximum amount of main storage that is supported by current operating systems is listed in Table 7-13. A maximum of 16 TB of main storage can be defined for an LPAR on a z15 server.

Table 7-13 Maximum memory that is supported by the operating system

| Operating system | Maximum supported main storage ^a |
|---------------------------|---|
| z/OS | z/OS V2R1 and later support 4 TB. |
| z/VM | z/VM V6R4 and V7R1 support 2 TB |
| z/VSE | z/VSE V6R2 supports 32 GB. |
| z/TPF | z/TPF supports 4 TB. |
| CFCC | Level 22, 23, and 24 supports up to 3 TB. |
| Secure Service Containers | Supports up to 16TB ^a . |
| Linux on Z (64-bit) | 16TB ^{ab} |

a. On z15 T01 LPAR storage definition supports 16TB (z15 T01 supports up to 40 TB of usable memory), while on z15 T02 LPAR storage definition supports 8 TB (z15 T02 server supports up to 16 TB of usable memory).

b. Support may vary by distribution. Check with your distribution provider.

IBM z15 Model T01 - Up to 85 LPARs

This feature was first made available on z13 servers and allows the z15 T01 system to be configured with up to 85 LPARs. Because channel subsystems can be shared by up to 15 LPARs, it is necessary to configure six channel subsystems to reach the 85 LPARs limit.

The supported operating systems are listed in Table 7-3 and Table 7-4 on page 259.

Remember: A virtual appliance that is deployed in a Secure Service Container runs in a dedicated LPAR. When activated, it reduces the maximum number of available LPARs by one.

IBM z15 Model T02 - Up to 40 LPARs

This feature was first made available on z13s servers and allows the system to be configured with up to 40 LPARs. Because channel subsystems can be shared by up to 15 LPARs, it is necessary to configure three channel subsystems to reach the 40 LPARs limit.

The supported operating systems are listed in Table 7-3 and Table 7-4 on page 259.

Remember: A virtual appliance that is deployed in a Secure Service Container runs in a dedicated LPAR. When activated, it reduces the maximum number of available LPARs by one.

Separate LPAR management of PUs

z15 servers use separate PU pools for each optional PU type. The separate management of PU types enhances and simplifies capacity planning and management of the configured LPARs and their associated processor resources.

The supported operating systems are listed in Table 7-3 and Table 7-4 on page 259.

Dynamic PU add

Planning an LPAR configuration includes defining reserved PUs that can be brought online when extra capacity is needed. Operating system support is required to use this capability without an IPL; that is, nondisruptively. This support is available in z/OS for some time.

The dynamic PU add function enhances this support by allowing you to dynamically define and change the number and type of reserved PUs in an LPAR profile, which removes any planning requirements. The new resources are immediately made available to the operating system and in the case of z/VM, to its guests.

The supported operating systems are listed in Table 7-3 and Table 7-4 on page 259.

Dynamic LPAR memory upgrade

An LPAR can be defined with an initial and a reserved amount of memory. At activation time, the initial amount is made available to the partition and the reserved amount can be added later, partially or totally. Although these two memory zones do not have to be contiguous in real memory, they appear as logically contiguous to the operating system that runs in the LPAR.

z/OS can take advantage of this support and nondisruptively acquire and release memory from the reserved area. z/VM V6R4 and later can acquire memory nondisruptively and immediately make it available to guests. z/VM virtualizes this support to its guests, which now also can increase their memory nondisruptively if supported by the guest operating system. Currently, releasing memory from z/VM is not supported³. Releasing memory from the z/VM guest depends on the guest's operating system support.

Linux on Z also supports acquiring and releasing memory nondisruptively. This feature is enabled for SUSE Linux Enterprise Server 11 and RHEL 6 and later releases.

LPAR group absolute capping

On z13 servers, PR/SM was enhanced to support an option to limit the amount of physical processor capacity that is used by an individual LPAR when a PU that is defined as a CP or an IFL is shared across a set of LPARs. This enhancement is designed to provide a physical capacity limit that is enforced as an absolute (versus a relative) limit. It is not affected by changes to the logical or physical configuration of the system. This physical capacity limit can be specified in units of CPs or IFLs.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

Capacity Provisioning Manager

The provisioning architecture enables clients to better control the configuration and activation of the On/Off CoD. For more information, see Chapter 8., "System upgrades" on page 333. The new process is inherently more flexible and can be automated. This capability can result in easier, faster, and more reliable management of the processing capacity.

The Capacity Provisioning Manager, which is a feature that is first available with z/OS V1R9, interfaces with z/OS Workload Manager (WLM) and implements capacity provisioning policies. Several implementation options are available, from an analysis mode that issues only guidelines, to an autonomic mode that provides fully automated operations.

³ z/VM Dynamic Memory Downgrade (releasing memory from z/VM LPAR) will be made available in the future with PTFs for APAR VM66271. For more information, see: <http://www.vm.ibm.com/newfunction/#dmd>

Replacing manual monitoring with autonomic management or supporting manual operation with guidelines can help ensure that sufficient processing power is available with the least possible delay. The supported operating systems are listed in Table 7-3 on page 258.

Program-directed re-IPL

First available on System z9®, program directed re-IPL allows an operating system on a z15 to IPL again without operator intervention. This function is supported for SCSI and IBM extended count key data (IBM ECKD) devices.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

IOCP

All IBM Z servers require a description of their I/O configuration. This description is stored in I/O configuration data set (IOCDS) files. The I/O configuration program (IOCP) allows for the creation of the IOCDS file from a source file that is known as the I/O configuration source (IOCS).

The IOCS file contains definitions of LPARs and channel subsystems. It also includes detailed information for each channel and path assignment, control unit, and device in the configuration.

IOCP for z15 provides support for the following features:

- ▶ z15 Base machine definition
- ▶ New PCI function adapter for zHyperLink (HYL)
- ▶ New PCI function adapter for RoCE Express2 (CX4)
- ▶ New IOCP Keyword MIXTYPE required for prior FICON⁴ cards
- ▶ New hardware (announced with Driver 41)
- ▶ IOCP support for Dynamic I/O for stand-alone CF (Driver 36 and later)

IOCP required level for z15 servers: The required level of IOCP for the z15 is IOCP 5.5.0 with PTFs. For more information, see the following publications:

- ▶ *IBM Z Stand-Alone Input/Output Configuration Program User's Guide*, SB10-7173-01
- ▶ *IBM Z Input/Output Configuration Program User's Guide for ICP IOCP*, SB10-7172-04

Dynamic Partition Manager V4.0: Dynamic Partition Manager V4.0 is available for managing IBM Z servers that are running Linux. DPM 4.0 is available with HMC Driver Level 41 (HMC Version 2.15.0). IOCP does not need to configure a server that is running in DPM mode. For more information, see *IBM Dynamic Partition Manager (DPM) Guide*, SB10-7176-02.

7.4.2 Base CPC features and functions

In this section, we describe the features and functions of Base CPC.

HiperDispatch

The **HIPERDISPATCH=YES/NO** parameter in the IEAOPTxx member of SYS1.PARMLIB and on the **SET OPT=xx** command controls whether HiperDispatch is enabled or disabled for a z/OS image. It can be changed dynamically, without an IPL or any outage.

⁴ FICON Express16SA and FICON Express16S+ do not allow a mixture of CHPID types on the same card.

In z/OS, the IEAOPTxx keyword **HIPERDISPATCH** defaults to YES when it is running on a z15, z14, z13, and z13s system.

The use of SMT on z15 systems requires that HiperDispatch is enabled on the operating system. For more information, see “Simultaneous multithreading” on page 282.

Additionally, any LPAR that is running with more than 64 logical processors is required to operate in HiperDispatch Management Mode.

The following rules control this environment:

- ▶ If an LPAR is defined at IPL with more than 64 logical processors, the LPAR automatically operates in HiperDispatch Management Mode, regardless of the HIPERDISPATCH= specification.
- ▶ If logical processors are added to an LPAR that has 64 or fewer logical processors and the extra logical processors raise the number of logical processors to more than 64, the LPAR automatically operates in HiperDispatch Management Mode, regardless of the HIPERDISPATCH=YES/NO specification. That is, even if the LPAR has the HIPERDISPATCH=NO specification, that LPAR is converted to operate in HiperDispatch Management Mode.
- ▶ An LPAR with more than 64 logical processors that are running in HiperDispatch Management Mode cannot be reverted to run in non-HiperDispatch Management Mode.

HiperDispatch on z15 systems uses a new chip and CPC drawer configuration to improve the access cache performance. Beginning with z/OS V2R1, HiperDispatch was changed to use the PU Chip/cluster/drawer cache structure of z15 servers. The base support is provided by PTFs that are identified by:

- ▶ For z15 T01 - IBM.device.server.z15-8561.requiredservice
- ▶ For z15 T02 - IBM.device.server.z15-8562.requiredservice

PR/SM on z15 servers seeks to assign all memory in one CPC drawer that is striped across the clusters of that drawer to take advantage of the lower latency memory access in a drawer. Also, PR/SM tries to consolidate storage onto drawers with the most processor entitlement.

PR/SM on z15 servers seeks to assign all logical processors of a partition to one CPC drawer, packed into PU chips of that CPC drawer in cooperation with operating system HiperDispatch optimize shared cache usage.

PR/SM automatically keeps a partition's memory and logical processors on the same CPC drawer. This arrangement looks simple for a partition, but it is a complex optimization for multiple logical partitions because some must be split among processors drawers.

In z15, all processor types can be dynamically reassigned except IFPs.

To use HiperDispatch effectively, WLM goal adjustment might be required. Review the WLM policies and goals and update them as necessary. WLM policies can be changed without turning off HiperDispatch. A health check is provided to verify whether HiperDispatch is enabled on a system image.

z/VM V7R2, V7R1 and V6R4

z/VM also uses the HiperDispatch facility for improved processor efficiency by better use of the processor cache to take advantage of the cache-rich processor, node, and drawer design of the z15 system. The supported processor limit was increased to 80 for z/VM 7.1 and z/VM 7.2 (40 with SMT and up to 80 threads running simultaneously), whereas it remains at 64 for z/VM 6.4 (32 with SMT and supports up to 64 threads that are running simultaneously).

CPU polarization support in Linux on Z

You can optimize the operation of a vertical SMP environment by adjusting the SMP factor based on the workload demands. For more information about CPU polarization support in Linux on Z, see the [CPU polarization page](#) of IBM Knowledge Center.

z/TPF

z/TPF on z15 can use more processors immediately without reactivating the LPAR or IPLing the z/TPF system.

In installations older than z14, z/TPF workload is evenly distributed across all available processors, even in low-utilization situations. This configuration causes cache and core contention with other LPARs. When z/TPF is running in a shared processor configuration, the achieved MIPS is higher when z/TPF is using a minimum set of processors.

In low-utilization periods, z/TPF now minimizes the processor footprint by compressing TPF workload onto a minimal set of I-streams (engines), which reduces the effect on other LPARs and allows the entire CPC to operate more efficiently.

As a consequence, z/OS and z/VM experience less contention from the z/TPF system when the z/TPF system is operating at periods of low demand.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

zIIP support

zIIPs do not change the model capacity identifier of z15 servers. IBM software product license charges that are based on the model capacity identifier are not affected by the addition of zIIPs. On a z15 server, z/OS Version 2 Release 1 is the minimum level for supporting zIIPs.

No changes to applications are required to use zIIPs. They can be used by the following applications:

- ▶ Db2 V8 and later for z/OS data serving for applications that use data Distributed Relational Database Architecture (DRDA) over TCP/IP, such as data serving, data warehousing, and selected utilities.
- ▶ z/OS XML services.
- ▶ z/OS CIM Server.
- ▶ z/OS Communications Server for network encryption (Internet Protocol Security [IPSec]) and for large messages that are sent by HiperSockets.
- ▶ IBM GBS Scalable Architecture for Financial Reporting.
- ▶ IBM z/OS Global Mirror (formerly XRC) and System Data Mover.
- ▶ IBM z/OS Container Extensions.
- ▶ IBM OMEGAMON® XE on z/OS, OMEGAMON XE on Db2 Performance Expert, and Db2 Performance Monitor.
- ▶ Any Java application that uses the current IBM SDK.
- ▶ WebSphere Application Server V5R1 and later, and products that are based on it, such as WebSphere Portal, WebSphere Enterprise Service Bus (WebSphere ESB), and WebSphere Business Integration (WBI) for z/OS.
- ▶ CICS/TS V2R3 and later.
- ▶ Db2 UDB for z/OS Version 8 and later.
- ▶ IMS Version 8 and later.

- ▶ zIIP Assisted HiperSockets for large messages.
- ▶ z/OSMF (z/OS Management Facility).
- ▶ IBM z/OS Platform for Apache Spark.
- ▶ IBM Watson® Machine Learning for z/OS.
- ▶ z/OS System Recovery Boost.

The functioning of a zIIP is transparent to application programs. The supported operating systems are listed in Table 7-3 on page 258.

On z15 servers, the zIIP processor is designed to run in SMT mode, with up to two threads per processor. This function is designed to help improve throughput for zIIP workloads and provide appropriate performance measurement, capacity planning, and SMF accounting data. This support is available for z/OS V2.1 with PTFs and higher.

Use the **PROJECTCPU** option of the IEAOPTxx parmlib member to help determine whether zIIPs can be beneficial to the installation. Setting PROJECTCPU=YES directs z/OS to record the amount of eligible work for zIIPs in SMF record type 72 subtype 3. The field APPL% IIPCP of the Workload Activity Report listing by WLM service class indicates the percentage of a processor that is zIIP eligible. Because of the zIIP's lower price as compared to a CP, even a utilization as low as 10% can provide cost benefits.

Transactional Execution

The IBM zEnterprise EC12 introduced an architectural feature called Transactional Execution (TX). This capability is known in academia and industry as *hardware transactional memory*. Transactional execution is also implemented on subsequent IBM Z servers.

This feature enables software to indicate to the hardware the beginning and end of a group of instructions that must be treated in an atomic way. All of their results occur or none occur, in true transactional style. The execution is optimistic.

The hardware provides a memory area to record the original contents of affected registers and memory as the instruction's execution occurs. If the transactional execution group is canceled or must be rolled back, the hardware transactional memory is used to reset the values. Software can implement a fallback capability.

This capability increases the software's efficiency by providing a way to avoid locks (lock elision). This advantage is of special importance for speculative code generation and highly parallelized applications.

TX is used by IBM Java virtual machine (JVM) and might be used by other software. The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

System Recovery Boost

System Recovery Boost is a new feature that was implemented on the IBM z15 system. This feature provides higher temporary processing capacity to LPARs for speeding up shutdown and IPL operations, without increasing software costs. For more information, Appendix B, "System Recovery Boost" on page 493.

The temporary processing capacity boost is intended to help workloads catch-up after an IPL (planned or unplanned) to work faster through a backlog after a downtime, which improves overall system availability by reducing the elapsed time required to recover service.

The capacity boost is available on an LPAR basis and is provided for general-purpose processors (CPs) and the using operating systems. The following types of temporary boost capacity (see Table 7-14 on page 281 for operating system support) are available:

- ▶ Subcapacity Boost (for z15 T01 systems with CPs running at a subcapacity index 4xx, 5xx, or 6xx and z15 T02 systems with CPs running at a subcapacity index CP-A ... CP-Y)). For LPARS running on subcapacity processors, during the boost period the allocated CPs are boosted to full capacity (7xx) without increasing software cost.
- ▶ CP capacity boost using zIIP conversion during the boost period. If the customer has zIIP processors, these processors can be converted for the temporary boost interval to general-purpose processors (CPs) for the selected LPARs. After the boost period ends, the zIIPs resume their characterization. No other software cost is associated with the zIIPs during the conversion (boost) period.

For z15 T01 only, System Recovery Boost Upgrade (FC 9930 and FC 6802) provides more temporary zIIP boost records which can be obtained by way of eBOD⁵ for supplementing the existing available customer zIIPs. The temporary zIIP boost records must be activated before planned operations and deactivated at the end of the boost period (automation software can be used for this purpose).

Table 7-14 Operating system support for System Recovery Boost

| Boost type ^a | z/OS V2R4 | z/OS V2R3 | z/OS V2R2 | z/VM V7R2 | z/VM V7R1 | z/VM V6R4 | z/TPF V1R1 | z/VSE V6R2 | Linux on Z |
|---|----------------|----------------|-----------|-----------|-----------------|-----------|----------------|----------------|------------|
| Subcapacity Boost for IPL, Shutdown and dumps | Y | Y ^b | N | Y | Y ^{bc} | N | Y ^b | Y ^b | N |
| zIIP to CP capacity boost ^d for IPL, shutdown and dump events | Y | Y ^b | N | N | N | N | N | N | N |
| Subcapacity and zIIP Boost for process recovery boost in a sysplex ^e . | Y ^f | Y ^f | N | N | N | N | N | N | N |

a. Boost must be enabled for LPARs to opt in.

b. With Fixes.

c. Subcapacity boost might be available during the boost period to guest operating systems except for z/OS.

d. zIIP processor capacity boost is only available if customer has at least one active processor characterized as zIIP. For z15 T01 only, more zIIPs can be used if obtained through eBOD (temporary zIIP boost records).

e. Process recovery boosts support subcapacity CPs speed boost and entitled (purchased) customer zIIPs only - up to two (zIIPs provided by FC 9930 and FC 6802 cannot be used for process recovery boosts).

f. Requires fixes for APAR OA59813

Automation

The client's automation product can be used to automate and control the following System Recovery Boost activities:

- ▶ To activate and deactivate the eBod temporary capacity record to provide more physical zIIPs for an IPL or Shutdown Boost.
- ▶ To dynamically modify LPAR weights, as might be needed to modify the sharing of physical zIIP capacity during a Boost period.
- ▶ To drive the invocation of the PROC that indicates the beginning of a shutdown process (and the start of the shut-down Boost).
- ▶ To take advantage of new composite HW API reconfiguration actions.

⁵ System Recovery Boost Upgrade using temporary zIIP capacity is *NOT available* for z15 T02. System Recovery boost is available with existing hardware only (subcapacity CP speed boost and zIIP boost with existing zIIPs).

- ▶ To control the level of parallelism that is present in the workload at startup (for example, starting middleware regions) and shutdown (for example, performing an orderly shutdown of middleware).

Simultaneous multithreading

SMT is the hardware capability to process up to two simultaneous threads in a single core, sharing the resources of the superscalar core. This capability improves the system capacity and efficiency in the usage of the processor, which increases the overall throughput of the system.

The z15 can run up two threads simultaneously in the same processor, which dynamically shares resources of the core, such as cache, translation lookaside buffer (TLB), and execution resources. It provides better utilization of the cores and more processing capacity.

SMT⁶ is supported for zIIPs and IFLs.

Note: For zIIPs and IFLs, SMT must be enabled on z/OS, z/VM, or Linux on Z instances. An operating system with SMT support can be configured to dispatch work to a thread on a zIIP (for eligible workloads in z/OS) or an IFL (for z/VM) core in single-thread or SMT mode.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

An operating system that uses SMT controls each core and is responsible for maximizing their throughput and meeting workload goals with the smallest number of cores. In z/OS, consider HiperDispatch cache optimization when you must choose the two threads to be dispatched in the same processor.

HiperDispatch attempts to dispatch guest virtual CPUs on the same logical processor on which they ran. PR/SM attempts to dispatch a vertical low logical processor in the same physical processor. If that process is not possible, it attempts to dispatch it in the same node, or then the same CPC drawer where it was dispatched before to maximize cache reuse.

From the point of view of an application, SMT is transparent and no changes are required in the application for it to run in an SMT environment, as shown in Figure 7-1.

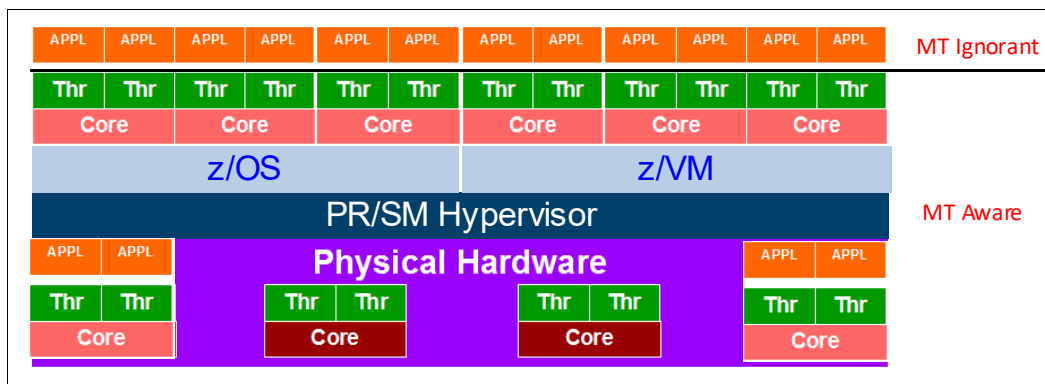


Figure 7-1 Simultaneous multithreading

z/OS

The following APARs must be applied to z/OS V2R1 to use SMT:

⁶ SMT is also enabled (not user configurable) by default for SAPs.

- ▶ OA43366 (BCP)
- ▶ OA43622 (WLM)
- ▶ OA44439 (XCF)

The use of SMT on z/OS V2R1 requires enabling HiperDispatch, and defining the processor view (**PROCVIEW**) control statement in the LOADxx parmlib member and the **MT_ZIIP_MODE** parameter in the IEAOPTxx parmlib member.

The **PROCVIEW** statement is defined for the life of IPL, and can have the following values:

- ▶ **CORE**: This value specifies that z/OS should configure a processor view of core, in which a core can include one or more threads. The number of threads is limited by z15 to two threads. If the underlying hardware does not support SMT, a core is limited to one thread.
- ▶ **CPU**: This value is the default. It specifies that z/OS should configure a traditional processor view of CPU and not use SMT.
- ▶ **CORE,CPU_OK**: This value specifies that z/OS should configure a processor view of core (as with the **CORE** value) but the CPU parameter is accepted as an alias for applicable commands.

When **PROCVIEW CORE** or **CORE,CPU_OK** are specified in z/OS that is running in z15, HiperDispatch is forced to run as enabled, and you cannot disable HiperDispatch. The **PROCVIEW** statement cannot be changed dynamically; therefore, you must run an IPL after changing it to make the new setting effective.

The **MT_ZIIP_MODE** parameter in the IEAOPTxx controls zIIP SMT mode. It can be 1 (the default), where only one thread can be running in a core, or 2, where up two threads can be running in a core. If **PROCVIEW CPU** is specified, the **MT_ZIIP_MODE** is always 1. Otherwise, the use of SMT to dispatch two threads in a single zIIP logical processor (**MT_ZIIP_MODE=2**) can be changed dynamically by using the **SET OPT=xx** setting in the IEAOPTxx parmlib. Changing the MT mode for all cores can take some time to complete.

PROCVIEW CORE requires **DISPLAY M=CORE** and **CONFIG CORE** to display the core states and configure an entire core.

With the introduction of Multi-Threading support for SAPs, a maximum of 88 logical SAPs can be used. RMF is updated to support this change by implementing page break support in the I/O Queuing Activity report that is generated by the RMF Post processor.

z/VM V7R2, V7R1, and V6R4

The use of SMT in z/VM is enabled by using the **MULTITHREADING** statement in the system configuration file. Multithreading is enabled only if z/VM is configured to run with the HiperDispatch vertical polarization mode enabled and with the dispatcher work distribution mode set to reshuffle.

The default in z/VM is multithreading disabled. With the addition of dynamic SMT capability to z/VM V6R4 through an SPE, the number of active threads per core can be changed without a system outage and potential capacity gains going from SMT-1 to SMT-2 (one to two threads per core) can now be achieved dynamically. Dynamic SMT requires applying PTFs that are running in SMT enabled mode and enables dynamically varying the active threads per core.

z/VM V7R2 and V7R2 support up to 40 multithreads cores (80 threads), while V6R4 supports up to 32 multithreaded cores (64 threads) for IFLs, and each thread is treated as an independent processor. z/VM dispatches virtual IFLs on the IFL logical processor so that the same or different guests can share a core. Each core has a single dispatch vector, and z/VM attempts to place virtual sibling IFLs on the same dispatch vector to maximize cache reuses.

The guests have no awareness of SMT, and cannot use it. z/VM SMT exploitation does not include guest support for multithreading. The value of this support for guests is that the first-level z/VM hosts under the guests can achieve higher throughput from the multi-threaded IFL cores.

Linux on Z and the KVM hypervisor

The upstream kernel 4.0 features SMT functionality that was developed by the Linux on Z development team. SMT is supported on LPAR only (not as a second-level guest). For more information, see the [Kernel 4.0 page of the developerWorks website](#).

The following *minimum* releases of Linux on Z distributions are supported on z15 (native SMT support):

- ▶ SUSE:
 - SLES 15 SP1 with service
 - SUSE SLES 12 SP4 with service
 - SUSE SLES 11 SP4 with service
- ▶ Red Hat:
 - Red Hat RHEL 8.0 with service
 - Red Hat RHEL 7.7 with service
 - Red Hat RHEL 6.10 with service
- ▶ Ubuntu:
 - Ubuntu 18.04.1 LTS with service
 - Ubuntu 16.04.5 LTS with service

The KVM hypervisor is supported on the same Linux on Z distributions in this list.

For most current support, see the [Linux on IBM Z Tested platforms website](#).

Single-instruction multiple-data

The SIMD feature introduces a new set of instructions to enable parallel computing that can accelerate code with string, character, integer, and floating point data types. The SIMD instructions allow a larger number of operands to be processed with a single complex instruction.

z15 is equipped with new set of instructions to improve the performance of complex mathematical models and analytic workloads through vector processing and new complex instructions, which can process numerous data with a single instruction. This new set of instructions, which is known as SIMD, enables more consolidation of analytic workloads and business transactions on Z servers.

SIMD on z15 has support for enhanced math libraries that provide performance improvements for analytical workloads by processing more information with a single CPU instruction.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259. Operating System support includes the following features⁷:

- ▶ Enablement of vector registers.
- ▶ Use of vector registers that use XL C/C++ ARCH(11) and TUNE(11).

⁷ The features that are listed here might not be available on all operating systems that are listed in the tables.

- ▶ A math library with an optimized and tuned math function (Mathematical Acceleration Subsystem [MASS]) that can be used in place of some of the C standard math functions. It includes a SIMD vectorized and non-vectorized version.
- ▶ A specialized math library, which is known as Automatically Tuned Linear Algebra Software (ATLAS), that is optimized for the hardware.
- ▶ IBM Language Environment® for C runtime function enablement for ATLAS.
- ▶ DBX to support the disassembly of the new vector instructions, and to display and set vector registers.
- ▶ XML SS exploitation to use new vector processing instructions to improve performance.

MASS and ATLAS can reduce the time and effort for middleware and application developers. IBM provides compiler built-in functions for SIMD that software applications can use as needed, such as for using string instructions.

The use of new hardware instructions require the z/OS V2R4 XL C/C++ compiler with ARCH(13) and TUNE(13) options for targeting z15 instructions. The ARCH(13) compiler option allows the compiler to use any new z15 instructions where appropriate. The TUNE(13) compiler option allows the compiler to tune for any z15 micro-architecture.

Vector programming support is extended for z15 to provide access to the new instructions that were introduced by the VEF 2⁸ specification.

Older levels of z/OS XL C/C++ compilers do not provide z15 exploitation; however, the z/OS V2R4 XL C/C++ compiler can be used to generate code for the older levels of z/OS running on z15.

The followings compilers include built-in functions for SIMD:

- ▶ IBM Java
- ▶ XL C/C++
- ▶ Enterprise COBOL
- ▶ Enterprise PL/I

Code must be developed to take advantage of the SIMD functions. Applications with SIMD instructions abend if they run on a lower hardware level system. Some mathematical function replacement can be done without code changes by including the scalar MASS library before the standard math library.

The MASS and standard math library include different accuracies, so assess the accuracy of the functions in the context of the user application before deciding whether to use the MASS and ATLAS libraries.

The SIMD functions can be disabled in z/OS partitions at IPL time by using the **MACHMIG** parameter in the LOADxx member. To disable SIMD code, use the MACHMIG VEF hardware-based vector facility. If you do not specify a **MACHMIG** statement, which is the default, the system is unlimited in its use of the Vector Facility for z/Architecture (SIMD).

Hardware decimal floating point

Industry support for decimal floating point is growing, with IBM leading the open standard definition. Examples of support for the draft standard IEEE 754r include Java BigDecimal, C#, XML, C/C++, GCC, COBOL, and other key software vendors, such as Microsoft and SAP.

Decimal floating point support was introduced with z9 EC. z15 servers inherited the decimal floating point accelerator feature that was introduced with z10 EC.

⁸ Hardware-based Vector-extension facility 2.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259. For more information, see 7.5.4, “z/OS XL C/C++ considerations” on page 324.

Out-of-order execution

Out-of-order (OOO) execution yields significant performance benefits for compute-intensive applications by reordering instruction execution, which allows later (newer) instructions to be run ahead of a stalled instruction, and reordering storage accesses and parallel storage accesses. OOO maintains good performance growth for traditional applications.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259. For more information, see “3.4.3, “Out-of-Order execution” on page 102.

CPU Measurement Facility

Also known as Hardware Instrumentation Services (HIS), CPU Measurement Facility (CPUMF) was initially introduced with z10 EC to gain insight into the interaction of workload and hardware it runs on. CPU MF data can be collected by z/OS System Measurement Facility on SMF 113 records. The supported operating systems are listed in Table 7-3 on page 258.

For more information about this function, see [The Load-Program-Parameter and the CPU-Measurement Facilities](#).

For more information about the CPU Measurement Facility, see the [CPU MF - Update and WSC Experiences page](#) of the IBM Techdocs Library website.

For more information, see “12.2, “z15 Large System Performance Reference ratio” on page 478.

Large page support

In addition to the existing 1-MB large pages, 4-KB pages, and page frames, z15 servers support pageable 1-MB large pages, large pages that are 2 GB, and large page frames. The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

Virtual Flash Memory

IBM Virtual Flash Memory (VFM) is the replacement for the PCIe based Flash Express features, which were available on the IBM zEC12 and IBM z13. No application changes are required to change from IBM Flash Express to VFM for it implements EADM Architecture using HSA-like memory instead of Flash card pairs.

IBM Virtual Flash Memory (FC 0643) offers up to 6.0 TB of memory for z15 T01, and up to 2.0 TB for z15 T02 in 0.5 TB Increments. VFM is provided for improved application availability and to handle paging workload spikes.

IBM Virtual Flash Memory is designed to help improve availability and handling of paging workload spikes when running z/OS V2.1, V2.2, V2.3, or V2.4. With this support, z/OS is designed to help improve system availability and responsiveness by using VFM across transitional workload events, such as market openings, and diagnostic data collection. z/OS is also designed to help improve processor performance by supporting middleware exploitation of pageable large (1 MB) pages.

Therefore, VFM can help organizations meet their most demanding service level agreements and compete more effectively. VFM is easily configurable, and to provide rapid time to value.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

Guarded Storage Facility

Also known as *less-pausing garbage collection*, Guarded Storage Facility (GSF) is a new architecture that was introduced with z14 to enable enterprise scale Java applications to run without periodic pause for garbage collection on larger heaps.

z/OS

GSF support allows an area of storage to be identified such that an Exit routine assumes control if a reference is made to that storage. GSF is managed by new instructions that define Guarded Storage Controls and system code to maintain that control information across undispach and redispach.

Enabling a less-pausing approach improves Java garbage collection. Function is provided on z14 and subsequent servers that are running z/OS 2.2 and later with APAR OA51643 installed. **MACHMIG** statement in **LOADxx** of **SYS1.PARMLIB** provides ability to disable the function.

z/VM

With the PTF for APAR VM65987, z/VM V6.4 provides support for guest exploitation of the guarded storage facility. This facility is designed to improve the performance of garbage-collection processing by various languages, in particular Java.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

Instruction Execution Protection

Instruction Execution Protection (IEP) is a new hardware function that was introduced with z14 that enables software, such as Language Environment, to mark certain memory regions (for example, a heap or stack), as non-executable to improve the security of programs running on IBM Z servers against stack-overflow or similar attacks.

Through enhanced hardware features (based on DAT table entry bit) and explicit software requests to obtain memory areas as non-executable, areas of memory can be protected from unauthorized execution. A Protection Exception occurs if an attempt is made to fetch an instruction from an address in such an element or if an address in such an element is the target of an execute-type instruction.

z/OS

To use IEP, Real Storage Manager (RSM) is enhanced to request non-executable memory allocation. Use new keyword **EXECUTABLE=YES|NO** on **STORAGE OBTAIN** or **IARV64** to indicate whether memory to be used contains executable code. Recovery Termination Manager (RTM) writes LOGREC record of any program-check that results from IEP.

IEP support is for z/OS 2.2 and later running on z15 with APARs OA51030 and OA51643 installed.

z/VM

Guest exploitation support for the Instruction Execution Protection Facility is provided with APAR VM65986.

The supported operating systems are listed in Table 7-3 on page 258 and Table 7-4 on page 259.

IBM Integrated Accelerator for zEnterprise Data Compression

The IBM Integrated Accelerator for zEnterprise Data Compression (zEDC) is implemented as on-chip data compression accelerator; that is, Nest Compression Accelerator (NXU) and designed to support Deflate/gzip/zlib algorithms. For more information, see Figure 3-10 on page 106).

Each PU chip has one on-chip compression unit, which is designed to replace the zEnterprise Data Compression (zEDC) Express PCIe feature.

The zEDC Express feature available on older systems is NOT carried forward to z15.

The IBM Integrated Accelerator for zEDC maintains software compatibility with existing zEDC Express use cases. For more information, see [Integrated Accelerator for zEnterprise Data Compression](#).

The z/OS zEDC capability is a software-priced feature that is designed to support compression capable hardware. With z15, the zEDC Express (hardware) PCIe feature is replaced by the on-chip compression accelerator unit, but the software (z/OS) component is required to maintain same functionality without increasing CPU costs.

All data interchange with existing (zEDC) compressed data remains compatible as z15 and zEDC capable machines coexist (accessing same data). Data that is compressed and written with zEDC will be read and decompressed by z15 well into the future.

The on-chip compression unit has the following operating modes:

- ▶ Synchronous execution in Problem State, where user application starts instruction in its virtual address space, which provides low latency and high-bandwidth compression/decompression operations). This mod does not require any special hypervisor support, which removes the virtualization layer (sharing the zEDC Express PCIe adapter among LPARs requires virtualization support).
- ▶ Asynchronous optimization for Large Operations under z/OS. The authorized application (for example, BSAM/QSAM) issues I/O for asynchronous execution and SAP (PU) starts instruction (synchronously as described in the previous paragraph) on behalf of application. The on-chip accelerator enables load balancing of high compression loads and low latency and high bandwidth compared to zEDC Express, while maintaining current user experience on compression.

Functionality support for the IBM Integrated Accelerator for zEDC is listed in Table 7-3 on page 258 and Table 7-4 on page 259.

For more information, see Appendix C, “IBM Integrated Accelerator for zEnterprise Data Compression” on page 509.

7.4.3 Coupling and clustering features and functions

In this section, we describe the coupling and cluster features.

Coupling facility and CFCC considerations

Coupling facility (CF) connectivity to a z15 is supported on the z14, z13, z13s, or another z15. The CFCC levels that are supported on Z systems are listed in Table 7-15.

Table 7-15 IBM Z CFCC code-levels

| IBM Z server | Code level |
|---------------------|--------------------------------|
| z15 | CFCC Level 24 |
| z14 M0x and z14 ZR1 | CFCC Level 22 or CFCC Level 23 |
| z13 | CFCC Level 20 or CFCC Level 21 |
| z13s | CFCC Level 21 |

Consideration: Because coupling link connectivity with z14 does not support InfiniBand, introducing z15 into an installation requires extra planning. Consider the level of CFCC. For more information, see “Migration considerations” on page 199.

CFCC Level 24

CFCC Level 24 is delivered on z15 servers with driver level 41. CFCC Level 24 introduces the following enhancements:

- ▶ CFCC Fair Latch Manager

This enhancement to the internals of the Coupling Facility (CFCC) dispatcher provides CF work management efficiency and processor scalability improvements, and improve the “fairness” of arbitration for internal CF resource latches across tasks

- ▶ CFCC Message Path Resiliency enhancement

CF Message Paths use a z/OS-provided system identifier (SYID) to uniquely identify which z/OS system image, and instance of that system image, is sending requests over a message path to the CF. With z15, we are providing a new resiliency mechanism that transparently recovers for this “missing” message path deactivate (if and when that deactivation ever occurs).

During path initialization, the CF provides more information to z/OS about every message path that appears active, including the SYID for the path. Whenever z/OS interrogates the state of the message paths to the CF, z/OS checks this SYID information for currency and correctness, and if incorrect, gather diagnostic information and reactivates the path to correct the problem.

- ▶ CF monopolization avoidance

z/OS will take advantage of current CF support in CFLEVEL 24 (z15 T01/T02) to deliver improved z/OS support for handling CF monopolization.

With z15 T01/T02, the CF dispatcher will monitor in real-time the number of CF tasks that have a command assigned to them for a given structure, on a structure-by-structure basis.

When the number of CF tasks being used by any given structure exceeds a model-dependent CF threshold, and a global threshold on the number of active tasks is also exceeded, the structure will be considered to be “monopolizing” the CF, and z/OS will be informed of this monopolization.

New support in z/OS will observe the monopolization state for a structure, and start to selectively queue and throttle incoming requests to the CF, on a structure-specific basis – while other requests, for other “non-monopolizing” structures and workloads, are completely unaffected.

z/OS will dynamically manage the queue of requests for the “monopolizing” structures to limit the number of active CF requests (parallelism) to them, and will monitor the CF’s monopolization state information so as to observe the structure becoming

“non-monopolized” again, so that request processing can eventually revert back to a non-throttled mode of operation.

The overall goal of z/OS anti-monopolization support is to protect the ability of ALL well-behaved structures and workloads to access the CF, and get their requests processed in the CF in a timely fashion – while implementing queueing and throttling mechanisms in z/OS to hold back the specific abusive workloads that are causing problems for other workloads.

z/OS XCF/XES exploitation APAR support is required to provide this functionality.

► **CFCC Change Shared-Engine *CF Default* to **DYNDISP=THIN****

Coupling Facility images can run with shared or dedicated processors. Shared processor CFs can operate with different Dynamic Dispatching (DYNDISP) models:

- **DYNDISP=OFF**: LPAR timeslicing completely controls the CF processor.
- **DYNDISP=ON**: an optimization over pure LPAR timeslicing, in which the CFCC code manages timer interrupts to share processors more efficiently.
- **DYNDISP=THIN**: An interrupt-driven model in which the CF processor is dispatched in response to a set of events that generate Thin Interrupts.

Thin Interrupt support was available since zEC12/zBC12, and is proven to be efficient and well-performing in numerous different test and customer shared-engine coupling facility configurations.

Therefore, z15 is making **DYNDISP=THIN** the *default mode* of operation for coupling facility images that use shared processors.

CFCC Level 23

CFCC Level 23 is delivered on z14 servers with driver level 36. In addition to CFCC Level 22 enhancements, it introduces the following enhancements:

► **Asynchronous cross invalidation (XI) for CF cache structures**

This enhancement requires z/OS fixes for APARs OA54688 (exploitation) and OA54985 (toleration). It also requires explicit data manager support (Db2 V12 with PTFs).

► **Coupling Facility hang detection**

These enhancements provide a significant reduction in failure scope and client disruption (CF-level to structure-level), with no loss of FFDC collection capability. With this support, the CFCC dispatcher significantly reduces the CF hang detection interval to only 2 seconds, which allows more timely detection and recovery from such events.

When a hang is detected, in most cases the CF confines the scope of the failure to “structure damage” for the single CF structure the hung command was processing against, capture diagnostics with a nondisruptive CF dump, and continue operating without stopping or rebooting the CF image.

► **Coupling Facility granular latching**

This enhancement eliminates the performance degradation that is caused by structure-wide latching. With this support, most CF list and lock structure ECR processing no longer uses structure-wide latching. It serializes its execution by using the normal structure object latches that all mainline commands use. However, a few “edge conditions” in ECR processing still require structure-wide latching.

Before you begin the migration process, install the compatibility and coexistence PTFs. A planned outage is required when you upgrade the CF or CF LPAR to CFCC Level 23.

CFCC Level 22

CFCC Level 22 is delivered on z14 servers with driver level 32. CFCC Level 22 introduced the following enhancements:

- ▶ **Coupling Express Long Range (CE LR):** A new link type that was introduced with z14 for long-distance coupling connectivity.
- ▶ **Coupling Facility (CF) Processor Scalability:** CF work management and dispatching changes for IBM z14™ allow improved efficiency and scalability for coupling facility images.

First, ordered work queues were eliminated from the CF in favor of first-in/first-out queues, which avoids the overhead of maintaining ordered queues.

Second, protocols for system-managed duplexing were simplified to avoid the potential for latching deadlocks between duplexed structures.

Third, the CF image can now use its processors to perform specific work management functions when the number of processors in the CF image exceeds a threshold. Together, these changes improve the processor scalability and throughput for a CF image.

- ▶ **CF List Notification Enhancements:**

Significant enhancements were made to CF notifications that inform users about the status of shared objects within in a Coupling Facility.

First, structure notifications can use a round-robin scheme for delivering immediate and deferred notifications that avoids excessive “shotgun” notifications, which reduces notification overhead.

Second, an option is now available for delivering “aggressive” notifications, which can drive a notification when new elements are added to a queue. This feature provides initiative to get new work processed in a timely manner.

Third, notifications can now be driven when a queue transitions between full and not-full, which allows users to redrive messages that could not be written previously to a “full” queue. The combination of these notification enhancements provides flexibility to accommodate notification preferences among various CF users and yields more consistent, timely notifications.

- ▶ **Coupling Link Constraint Relief:**

IBM z14™ provides more physical and logical coupling link connectivity compared to z13. Consider the following points:

- The maximum number of physical ICA SR coupling links (ports) is increased from 40 per CPC to 80 per CPC. These higher limits on z14 support concurrent use of InfiniBand coupling, ICA SR, and CE LR links, for coupling link technology migration purposes.
- Maximum number of coupling CHPIDs (of all types) is 256 per CPC (same as z13).

- ▶ **CF Encryption:**

z/OS 2.3 provides support for end-to-end encryption for CF data in flight and data at rest in CF structures (as a part of the Pervasive Encryption solution). Host-based CPACF encryption is used for high performance and low latency. IBM z14™ CF images are not required, but are recommended to simplify some sysplex recovery and reconciliation scenarios involving encrypted CF structures. (The CF image never decrypts or encrypts any data). IBM z14™ z/OS images are not required, but are recommended for the improved AES CBC encrypt/decrypt performance that z14 provides.

The supported operating systems are listed in Table 7-5 on page 261.

For more information about CFCC code levels, see [the Parallel Sysplex page](#) of the IBM IT infrastructure website.

For more information about the latest CFCC code levels, see [the current exception letter](#) that is published on Resource Link website (login is required).

CF structure sizing changes are expected when upgrading from a previous CFCC Level to CFCC Level 21. Review the CF LPAR size by using the available CFSizer tool, which is available for [download at the IBM Systems support website](#).

The Sizer Utility, which is an authorized z/OS program download, is useful when you are upgrading a CF. The tool is available [for download at the IBM Systems support website](#).

Before you begin the migration process, install the compatibility and coexistence PTFs. A planned outage is required when you upgrade the CF or CF LPAR to CFCC Level 22.

Coupling links support

Integrated Coupling Adapter (ICA) Short Reach and Coupling Express Long Reach (CE LR) coupling link options provide high-speed connectivity at short and longer distances over fiber optic interconnections. For more information, see 4.6.4, “Parallel Sysplex connectivity” on page 195.

Integrated Coupling Adapter

PCIe Gen3 coupling fanout, which is also known as Integrated Coupling Adapter Short Range (ICA SR, ICA SR1.1), supports a maximum distance of 150 meters (492 feet) and is defined as CHPID type CS5 in IOCP.

Coupling Express Long Reach

The CE LR link provides point-to-point coupling connectivity at distances of 10 km (6.21 miles) unrepeated and defined as CHPID type CL5 in IOCP. The supported operating systems are listed in Table 7-5 on page 261.

Virtual Flash Memory use by CFCC

VFM can be used in coupling facility images to provide extended capacity and availability for workloads that use WebSphere MQ Shared Queues structures. The use of VFM can help availability by reducing latency from paging delays that can occur at the start of the workday or during other transitional periods. It is also designed to help eliminate delays that can occur when diagnostic data during failures is collected.

CFCC Coupling Thin Interrupts

The Coupling Thin Interrupts enhancement is delivered with CFCC 19. It improves the performance of a CF partition and the dispatching of z/OS LPARs that are awaiting the arrival of returned asynchronous CF requests when used in a shared engine environment.

For more information, see “**Coupling Thin Interrupts (default CF LPAR setting with z15)**” on page 116. The supported operating systems are listed in Table 7-5 on page 261.

Asynchronous CF Duplexing for lock structures

Asynchronous CF Duplexing enhancement is a general-purpose interface for any CF Lock structure user. It enables secondary structure updates to be performed asynchronously regarding primary updates. Initially delivered with CFCC 21 on z13 as an enhanced continuous availability solution, it offers performance advantages for duplexing lock structures and avoids the need for synchronous communication delays during the processing of every duplexed update operation.

Asynchronous CF Duplexing for lock structures requires the following software support:

- ▶ z/OS V2R4
- ▶ V2R3, z/OS V2.2 SPE with PTFs for APAR OA47796 and OA49148
- ▶ z/VM V7R2 and V7R1, z/VM V6.4 with PTFs for z/OS exploitation of guest coupling environment
- ▶ Db2 V12 with PTFs for APAR PI66689
- ▶ IRLM V2.3 with PTFs for APAR PI68378

The supported operating systems are listed in Table 7-5 on page 261.

Asynchronous cross-invalidate for CF cache structures

Asynchronous cross-invalidate (XI) for CF cache structures enables improved efficiency in CF data sharing by adopting a more transactional behavior for cross-invalidate (XI) processing, which is used to maintain coherency and consistency of data managers' local buffer pools across the sysplex.

Instead of performing XI signals synchronously on every cache update request that causes them, data managers can "opt in" for the CF to perform these XIs asynchronously (and then sync them up with the CF at or before transaction completion). Data integrity is maintained if all XI signals complete by the time transaction locks are released.

The feature enables faster completion of cache update CF requests, especially with cross-site distance involved and provides improved cache structure service times and coupling efficiency. It requires explicit data manager exploitation/participation, which is not transparent to the data manager. No SMF data changes were made for CF monitoring and reporting.

The following requirements must be met:

- ▶ CFCC Level 23 support, plus
- ▶ z/OS 2.4
- ▶ PTFs on every exploiting system in the sysplex:
 - Fixes for APAR OA54688 - Exploitation support z/OS 2.2 and 2.3
 - Fixes for APAR OA54985 - Toleration support for z/O 2.1
- ▶ Db2 V12 with PTFs for exploitation

z/VM Dynamic I/O support for InfiniBand⁹ and ICA CHPIDs

z/VM dynamic I/O configuration support allows you to add, delete, and modify the definitions of channel paths, control units, and I/O devices to the server and z/VM without shutting down the system.

This function refers exclusively to the z/VM dynamic I/O support of InfiniBand and ICA coupling links. Support is available for the CIB and CS5 CHPID type in the z/VM dynamic commands, including the **change channel path** dynamic I/O command.

Specifying and changing the system name when entering and leaving configuration mode are also supported. z/VM does not use InfiniBand or ICA, and does not support the use of InfiniBand or ICA coupling links by guests. The supported operating systems are listed in Table 7-5 on page 261.

7.4.4 Storage connectivity-related features and functions

In this section, we describe the storage connectivity-related features and functions.

⁹ InfiniBand coupling is *not* supported on z15.

zHyperlink Express

z14 introduced IBM zHyperLink Express as a brand new IBM Z input/output (I/O) channel link technology since FICON. zHyperLink Express 1.1 is available with new z15 system and is designed to help bring data close to processing power, increase the scalability of Z transaction processing, and lower I/O latency.

zHyperLink Express is designed for up to 5x lower latency than High-Performance FICON for Z (zHPF) by directly connecting the Z central processor complex (CPC) to the I/O Bay of the DS8000 (DS8880 or later). This short distance (up to 150 m [492.1 feet]), direct connection is intended to speed Db2 for z/OS transaction processing and improve active log throughput.

The improved performance of zHyperLink Express allows the Processing Unit (PU) to make a synchronous request for the data that is in the DS8000 cache. This feature eliminates the undispatch of the running request, the queuing delays to resume the request, and the PU cache disruption.

Support for zHyperLink Writes can accelerate Db2 log writes to help deliver superior service levels by processing high-volume Db2 transactions at speed. IBM zHyperLink Express requires compatible levels of DS8000/F hardware, firmware R8.5.1 or later, and Db2 12 with PTFs.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

FICON Express16SA

Important: FICON Express16SA is *only available on z15 T01* (new build system). z15 T02 supports FICON Express16S+ for new build and as carry forward.

FICON Express16SA supports a link data rate of 16 gigabits per second (Gbps) and autonegotiation to 8 Gbps for synergy with switches, directors, and storage devices. With support for native FICON, High-Performance FICON for Z (zHPF), and Fibre Channel Protocol (FCP), the IBM z15™ server enables you to position your SAN for even higher performance, which helps you to prepare for an end-to-end 16 Gbps infrastructure to meet the lower latency and increased bandwidth demands of your applications.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

IBM Fibre Channel Endpoint Security (z15 T01 only)

IBM z15 Model T01 supports IBM Fibre Channel Endpoint Security feature (FC 1146). FC 1146 provides FC/FCP link encryption and endpoint authentication. This is an optional priced feature which requires the following:

- ▶ FICON Express16SA for both link encryption and endpoint authentication
 - FICON Express16S+ for endpoint authentication only.
- ▶ Select DS8000 storage
- ▶ Supporting infrastructure - IBM Security Key Lifecycle Manager 3.01
- ▶ CPACF enablement (FC 3863)

See the following announcement letter:

<https://www.ibm.com/downloads/cas/US-ENUS120-013-CA/name/US-ENUS120-013-CA.PDF>

FICON Express16S+

FICON Express16S+ supports a link data rate of 16 Gbps and autonegotiation to 4 or 8 Gbps for synergy with switches, directors, and storage devices. With support for native FICON, High-Performance FICON for Z (zHPF), and Fibre Channel Protocol (FCP), the IBM Z systems enable you to position your SAN for even higher performance, which helps you to prepare for an end-to-end 16 Gbps infrastructure to meet the lower latency and increased bandwidth demands of your applications.

The new FICON Express16S+ channel works with your fiber optic cabling environment (single mode and multimode optical cables). The FICON Express16S+ feature running at end-to-end 16 Gbps link speeds provides reduced latency for large read/write operations and increased bandwidth compared to the FICON Express8S feature.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

FICON Express16S

FICON Express16S supports a link data rate of 16 Gbps and autonegotiation to 4 or 8 Gbps for synergy with existing switches, directors, and storage devices. With support for native FICON, zHPF, and FCP, the z14 server enables SAN for even higher performance, which helps to prepare for an end-to-end 16 Gbps infrastructure to meet the increased bandwidth demands of your applications.

The new features for the multimode and single mode fiber optic cabling environments reduce latency for large read/write operations and increase bandwidth compared to the FICON Express8S features.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

FICON Express8S

The FICON Express8S provides a link rate of 8 Gbps, with auto negotiation to 4 or 2 Gbps for compatibility with previous devices and investment protection. Both 10 km (6.2 miles) LX and SX connections are offered (in a feature, all connections must include the same type).

FICON Express8S introduced a hardware data router for more efficient zHPF data transfers. It is the first channel with hardware that is designed to support zHPF, as compared to FICON Express8, FICON Express4, and FICON Express2, which include a firmware-only zHPF implementation.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

Extended distance FICON

An enhancement to the industry-standard FICON architecture (FC-SB-3) helps avoid degradation of performance at extended distances by implementing a new protocol for persistent IU pacing. Extended distance FICON is transparent to operating systems and applies to all FICON Express16S+ FICON Express16S, and FICON Express8S features that carry native FICON traffic (CHPID type FC).

To use this enhancement, the control unit must support the new IU pacing protocol. IBM System Storage™ DS8000 series supports extended distance FICON for IBM Z environments. The channel defaults to current pacing values when it operates with control units that cannot use extended distance FICON.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

High-performance FICON

High-performance FICON (zHPF) was first provided on System z10®, and is a FICON architecture for protocol simplification and efficiency. It reduces the number of information units (IUs) that are processed. Enhancements were made to the z/Architecture and the FICON interface architecture to provide optimizations for online transaction processing (OLTP) workloads.

Important: FICON Express16SA is *only available on z15 T01* (new build system). z15 T02 supports FICON Express16S+ for new build and as carry forward.

zHPF is available on z15, z14, z13, z13s, zEC12, and zBC12 servers. The FICON Express16SA, FICON Express16S+ FICON Express16S, and FICON Express8S (CHPID type FC) concurrently support the existing FICON protocol and the zHPF protocol in the server LIC.

When used by the FICON channel, the z/OS operating system, and the DS8000 control unit or other subsystems, the FICON channel processor usage can be reduced and performance improved. Appropriate levels of Licensed Internal Code (LIC) are required.

Also, the changes to the architectures provide end-to-end system enhancements to improve reliability, availability, and serviceability (RAS).

zHPF is compatible with the following standards:

- ▶ Fibre Channel Framing and Signaling standard (FC-FS)
- ▶ Fibre Channel Switch Fabric and Switch Control Requirements (FC-SW)
- ▶ Fibre Channel Single-Byte-4 (FC-SB-4) standards

For example, the zHPF channel programs can be used by the z/OS OLTP I/O workloads, Db2, VSAM, the partitioned data set extended (PDSE), and the z/OS file system (zFS).

At the zHPF announcement, zHPF supported the transfer of small blocks of fixed size data (4 K) from a single track. This capability was extended, first to 64 KB, and then to multitrack operations. The 64 KB data transfer limit on multitrack operations was removed by z196. This improvement allows the channel to fully use the bandwidth of FICON channels, which results in higher throughputs and lower response times.

The multitrack operations extension applies to the FICON Express16SA, FICON Express16S+ FICON Express16S, and FICON Express8S, when configured as CHPID type FC and connecting to z/OS. zHPF requires matching support by the DS8000 series. Otherwise, the extended multitrack support is transparent to the control unit.

zHPF is enhanced to allow all large write operations (greater than 64 KB) at distances up to 100 km (62.13 miles) to be run in a single round trip to the control unit. This process does not elongate the I/O service time for these write operations at extended distances. This enhancement to zHPF removes a key inhibitor for clients adopting zHPF over extended distances, especially when the IBM HyperSwap capability of z/OS is used.

From the z/OS perspective, the FICON architecture is called *command mode* and the zHPF architecture is called *transport mode*. During link initialization, the channel node and the control unit node indicate whether they support zHPF.

Requirement: All FICON channel path identifiers (CHPIDs) that are defined to the same LCU must support zHPF. The inclusion of any non-compliant zHPF features in the path group causes the entire path group to support command mode only.

The mode that is used for an I/O operation depends on the control unit that supports zHPF and its settings in the z/OS operating system. For z/OS use, a parameter is available in the IECIOSxx member of SYS1.PARMLIB (**ZHPF=YES** or **NO**) and in the **SETIOS** system command to control whether zHPF is enabled or disabled. The default is ZHPF=NO.

Support is also added for the **D IOS,ZHPF** system command to indicate whether zHPF is enabled, disabled, or not supported on the server.

Similar to the existing FICON channel architecture, the application or access method provides the channel program (CCWs). How zHPF (transport mode) manages channel program operations is different from the CCW operation for the existing FICON architecture (command mode). While in command mode, each CCW is sent to the control unit for execution. In transport mode, multiple channel commands are packaged together and sent over the link to the control unit in a single control block. Fewer processors are used compared to the existing FICON architecture. Certain complex CCW chains are not supported by zHPF.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

For more information about FICON channel performance, see the performance technical papers that are available [at the IBM Z I/O connectivity page](#) of the IBM IT infrastructure website.

Modified Indirect Data Address Word facility

The Modified Indirect Data Address Word (MIDAW) facility improves FICON performance. It provides a more efficient channel command word (CCW)/indirect data address word (IDAW) structure for certain categories of data-chaining I/O operations.

The MIDAW facility is a system architecture and software feature that is designed to improve FICON performance. This facility was first made available on System z9 servers, and is used by the Media Manager in z/OS.

The MIDAW facility provides a more efficient CCW/IDAW structure for certain categories of data-chaining I/O operations.

MIDAW can improve FICON performance for extended format data sets. Non-extended data sets can also benefit from MIDAW.

MIDAW can improve channel utilization and I/O response time. It also reduces FICON channel connect time, director ports, and control unit processor usage.

IBM laboratory tests indicate that applications that use EF data sets, such as Db2, or long chains of small blocks can gain significant performance benefits by using the MIDAW facility.

MIDAW is supported on FICON channels that are configured as CHPID type FC. The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

MIDAW technical description

An IDAW is used to specify data addresses for I/O operations in a virtual environment.¹⁰ The IDAW design allows the first IDAW in a list to point to any address within a page. Subsequent

IDAWs in the same list must point to the first byte in a page. Also, IDAWs (except the first and last IDAW) in a list must manage complete 2 K or 4 K units of data.

Figure 7-2 shows a single CCW that controls the transfer of data that spans non-contiguous 4 K frames in main storage. When the IDAW flag is set, the data address in the CCW points to a list of words (IDAWs). Each IDAW contains an address that designates a data area within real storage.

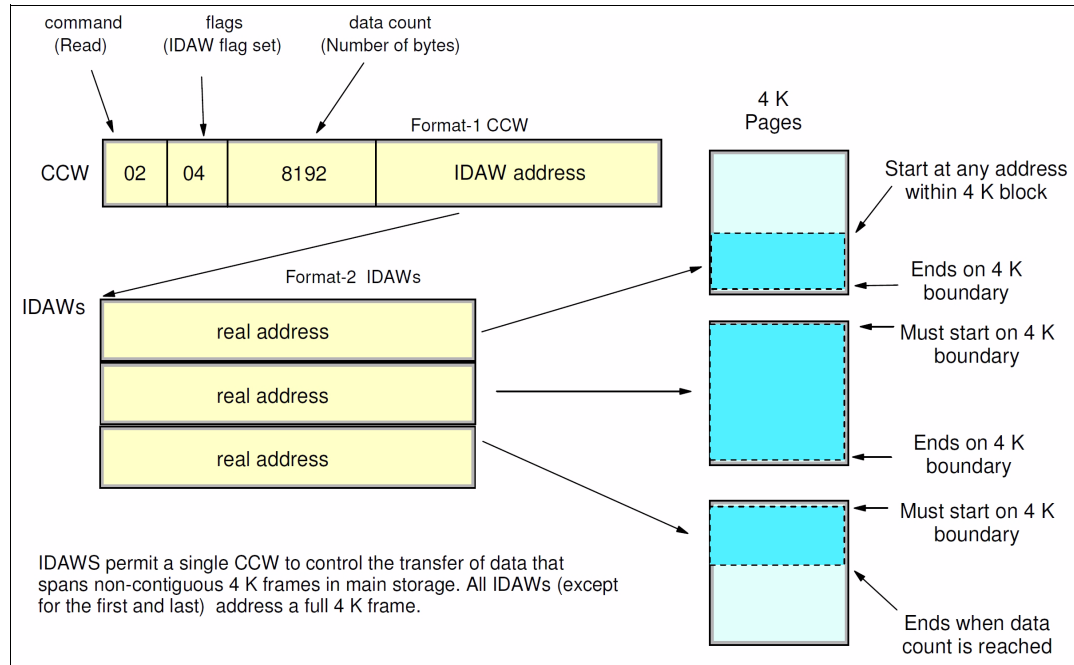


Figure 7-2 IDAW usage

The number of required IDAWs for a CCW is determined by the following factors:

- ▶ IDAW format as specified in the operation request block (ORB)
- ▶ Count field of the CCW
- ▶ Data address in the initial IDAW

For example, three IDAWs are required when the following events occur:

- ▶ The ORB specifies format-2 IDAWs with 4 KB blocks.
- ▶ The CCW count field specifies 8 KB.
- ▶ The first IDAW designates a location in the middle of a 4 KB block.

CCWs with data chaining can be used to process I/O data blocks that have a more complex internal structure, in which portions of the data block are directed into separate buffer areas. This process is sometimes known as *scatter-read* or *scatter-write*. However, as technology evolves and link speed increases, data chaining techniques become less efficient because of switch fabrics, control unit processing and exchanges, and other issues.

¹⁰ Exceptions are made to this statement, and many details are omitted in this description. In this section, we assume that you can merge this brief description with an understanding of I/O operations in a virtual memory environment.

The MIDAW facility is a method of gathering and scattering data from and into discontinuous storage locations during an I/O operation. The MIDAW format is shown in Figure 7-3. It is 16 bytes long and aligned on a quadword.

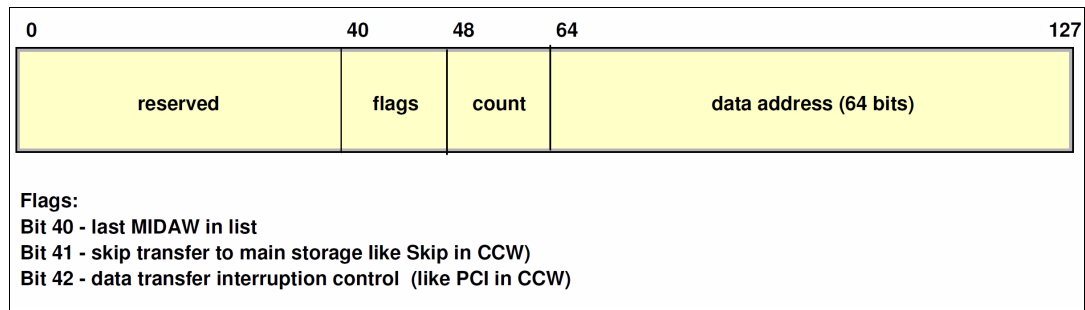


Figure 7-3 MIDAW format

An example of MIDAW usage is shown in Figure 7-4.

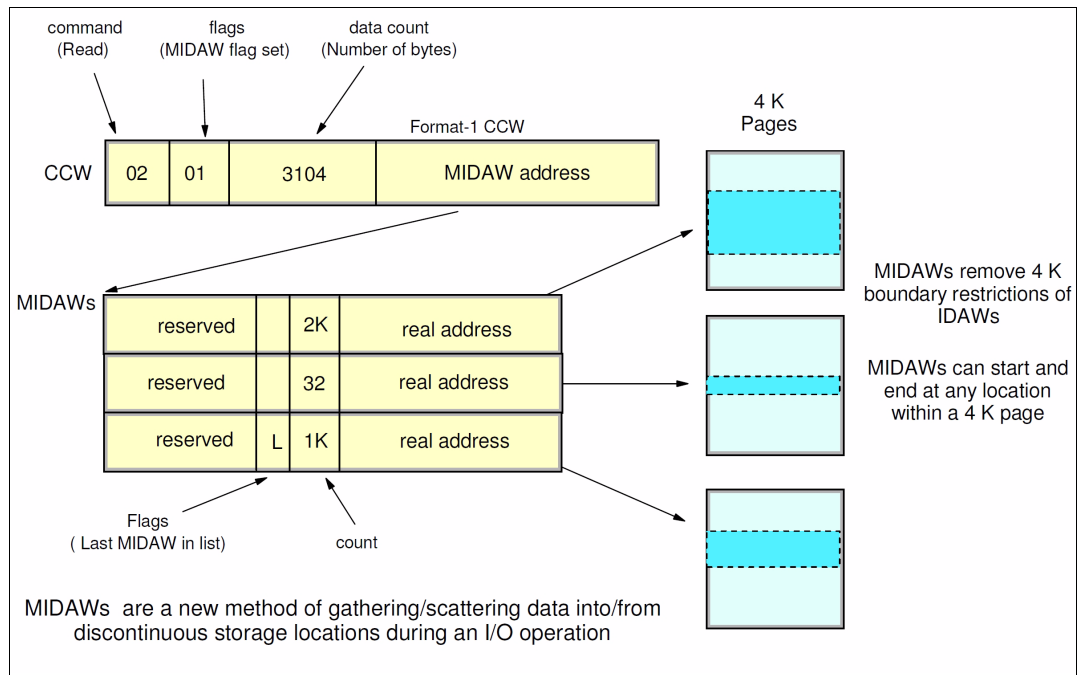


Figure 7-4 MIDAW usage

The use of MIDAWs is indicated by the MIDAW bit in the CCW. If this bit is set, the skip flag cannot be set in the CCW. The skip flag in the MIDAW can be used instead. The data count in the CCW must equal the sum of the data counts in the MIDAWs. The CCW operation ends when the CCW count goes to zero or the last MIDAW (with the last flag) ends.

The combination of the address and count in a MIDAW cannot cross a page boundary. Therefore, the largest possible count is 4 K. The maximum data count of all the MIDAWs in a list cannot exceed 64 K, which is the maximum count of the associated CCW.

The scatter-read or scatter-write effect of the MIDAWs makes it possible to efficiently send small control blocks that are embedded in a disk record to separate buffers from those that are used for larger data areas within the record. MIDAW operations are on a single I/O block, in the manner of data chaining. Do not confuse this operation with CCW command chaining.

Extended format data sets

z/OS extended format (EF) data sets use internal structures (often not visible to the application program) that require a scatter-read (or scatter-write) operation. Therefore, CCW data chaining is required, which produces less than optimal I/O performance. Because the most significant performance benefit of MIDAWs is achieved with EF data sets, a brief review of the EF data sets is included here.

VSAM and non-VSAM (DSORG=PS) sets can be defined as EF data sets. For non-VSAM data sets, a 32-byte suffix is appended to the end of every physical record (that is, block) on disk. VSAM appends the suffix to the end of every control interval (CI), which normally corresponds to a physical record.

A 32 K CI is split into two records to span tracks. This suffix is used to improve data reliability, and facilitates other functions that are described next. Therefore, for example, if the DCB BLKSIZE or VSAM CI size is equal to 8192, the actual block on storage consists of 8224 bytes. The control unit does not distinguish between suffixes and user data. The suffix is transparent to the access method and database.

In addition to reliability, EF data sets enable the following functions:

- ▶ DFSMS striping
- ▶ Access method compression
- ▶ Extended addressability (EA)

EA is useful for creating large Db2 partitions (larger than 4 GB). Striping can be used to increase sequential throughput, or to spread random I/Os across multiple logical volumes. DFSMS striping is useful for using multiple channels in parallel for one data set. The Db2 logs are often striped to optimize the performance of Db2 sequential inserts.

Processing an I/O operation to an EF data set normally requires at least two CCWs with data chaining. One CCW is used for the 32-byte suffix of the EF data set. With MIDAW, the additional CCW for the EF data set suffix is eliminated.

MIDAWs benefit EF and non-EF data sets. For example, to read 12 4 K records from a non-EF data set on a 3390 track, Media Manager chains together 12 CCWs by using data chaining. To read 12 4 K records from an EF data set, 24 CCWs are chained (two CCWs per 4 K record). By using Media Manager track-level command operations and MIDAWs, an entire track can be transferred by using a single CCW.

Performance benefits

z/OS Media Manager includes I/O channel program support for implementing EF data sets, and automatically uses MIDAWs when appropriate. Most disk I/Os in the system are generated by using Media Manager.

Users of the Executing Fixed Channel Programs in Real Storage (EXCPVR) instruction can construct channel programs that contain MIDAWs. However, doing so requires that they construct an IOBE with the IOBEMIDA bit set. Users of the EXCP instruction cannot construct channel programs that contain MIDAWs.

The MIDAW facility removes the 4 K boundary restrictions of IDAWs and for EF data sets, reduces the number of CCWs. Decreasing the number of CCWs helps to reduce the FICON channel processor utilization. Media Manager and MIDAWs do not cause the bits to move any faster across the FICON link. However, they reduce the number of frames and sequences that flow across the link, and therefore use the channel resources more efficiently.

The performance of a specific workload can vary based on the conditions and hardware configuration of the environment. IBM laboratory tests found that Db2 gains significant performance benefits by using the MIDAW facility in the following areas:

- ▶ Table scans
- ▶ Logging
- ▶ Utilities
- ▶ Use of DFSMS striping for Db2 data sets

Media Manager with the MIDAW facility can provide significant performance benefits when used in combination applications that use EF data sets (such as Db2) or long chains of small blocks.

For more information about FICON and MIDAW, see the following resources:

- ▶ The [I/O Connectivity page](#) of the IBM IT infrastructure website includes information about FICON channel performance
- ▶ *DS8000 Performance Monitoring and Tuning, SG24-7146*

ICKDSF

Device Support Facilities, ICKDSF, Release 17 is required on all systems that share disk subsystems with a z15 processor.

ICKDSF supports a modified format of the CPU information field that contains a two-digit LPAR identifier. ICKDSF uses the CPU information field instead of CCW reserve/release for concurrent media maintenance. It prevents multiple systems from running ICKDSF on the same volume, and at the same time allows user applications to run while ICKDSF is processing. To prevent data corruption, ICKDSF must determine all sharing systems that might run ICKDSF. Therefore, this support is required for z15.

Remember: The need for ICKDSF Release 17 also applies to systems that are not part of the same sysplex, or are running an operating system other than z/OS, such as z/VM.

z/OS Discovery and Auto-Configuration

z/OS Discovery and Auto Configuration (zDAC) is designed to automatically run several I/O configuration definition tasks for new and changed disk and tape controllers that are connected to a switch or director, when attached to a FICON channel.

The zDAC function is integrated into the hardware configuration definition (HCD). Clients can define a policy that can include preferences for availability and bandwidth that include parallel access volume (PAV) definitions, control unit numbers, and device number ranges. When new controllers are added to an I/O configuration or changes are made to existing controllers, the system discovers them and proposes configuration changes that are based on that policy.

zDAC provides real-time discovery for the FICON fabric, subsystem, and I/O device resource changes from z/OS. By exploring the discovered control units for defined logical control units (LCUs) and devices, zDAC compares the discovered controller information with the current system configuration. It then determines delta changes to the configuration for a proposed configuration.

All added or changed logical control units and devices are added into the proposed configuration. They are assigned proposed control unit and device numbers, and channel paths that are based on the defined policy. zDAC uses channel path chosen algorithms to minimize single points of failure. The zDAC proposed configurations are created as work I/O definition files (IODFs) that can be converted to production IODFs and activated.

zDAC is designed to run discovery for all systems in a sysplex that support the function. Therefore, zDAC helps to simplify I/O configuration on z15 systems that run z/OS, and reduces complexity and setup time.

zDAC applies to all FICON features that are supported on z15 when configured as CHPID type FC. The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

Platform and name server registration in FICON channel

The FICON Express16SA, FICON Express16S+, FICON Express16S, and FICON Express8S features support platform and name server registration to the fabric for CHPID types FC and FCP.

Important: FICON Express16SA is *only available on z15 T01* (new build system). z15 T02 supports FICON Express16S+ for new build and as carry forward.

Information about the channels that are connected to a fabric (if registered) allows other nodes or storage area network (SAN) managers to query the name server to determine what is connected to the fabric.

The following attributes are registered for the z15 systems:

- ▶ Platform information
- ▶ Channel information
- ▶ Worldwide port name (WWPN)
- ▶ Port type (N_Port_ID)
- ▶ FC-4 types that are supported
- ▶ Classes of service that are supported by the channel

The platform and name server registration service are defined in the Fibre Channel Generic Services 4 (FC-GS-4) standard.

The 63.75-K subchannels

Servers before z9 EC reserved 1024 subchannels for internal system use, out of a maximum of 64 K subchannels. Starting with z9 EC, the number of reserved subchannels was reduced to 256, which increased the number of subchannels that are available. Reserved subchannels exist in subchannel set 0 only. One subchannel is reserved in each of subchannel sets 1, 2, and 3.

The informal name, 63.75-K subchannels, represents 65280 subchannels, as shown in the following equation:

$$63 \times 1024 + 0.75 \times 1024 = 65280$$

This equation is applicable for subchannel set 0. For subchannel sets 1, 2 and 3, the available subchannels are derived by using the following equation:

$$(64 \times 1024) - 1 = 65535$$

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

Multiple subchannel sets

First introduced in z9 EC, multiple subchannel sets (MSS) provide a mechanism for addressing more than 63.75-K I/O devices and aliases for FICON (CHPID types FC) on the z15, z14, z13, z13s, zEC12, and zBC12. z196 introduced the third subchannel set (SS2).

With z13, one more subchannel set (SS3) was introduced, which expands the alias addressing by 64-K more I/O devices.

z/VM V6R3 MSS support for mirrored direct access storage device (DASD) provides a subset of host support for the MSS facility to allow the use of an alternative subchannel set for Peer-to-Peer Remote Copy (PPRC) secondary volumes.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264. For more information about channel subsystem, see Chapter 5, “Central processor complex channel subsystem” on page 203.

Chapter 5, “Central processor complex channel subsystem” on page 203.

Subchannel sets

z15 T01 supports four subchannel sets (SS0, SS1, SS2, SS3), while z15 T02 support three subchannel sets (SS0, SS1, SS2).

Subchannel sets SS1, SS2, and SS3 (SS3 for z15 T01 only) can be used for disk alias devices of primary and secondary devices, and as Metro Mirror secondary devices. This set helps facilitate storage growth and complements other functions, such as extended address volume (EAV) and Hyper Parallel Access Volumes (HyperPAV).

See Table 7-6 on page 262 and Table 7-7 on page 264 for list of supported operating systems.

IPL from an alternative subchannel set

z15 supports IPL from subchannel set 1 (SS1), subchannel set 2 (SS2), or subchannel set 3 (SS3), in addition to subchannel set 0.

See Table 7-6 on page 262 and Table 7-7 on page 264 for list of supported operating systems. For more information, refer to “IPL from an alternative subchannel set” on page 303.

32 K subchannels

To help facilitate growth and continue to enable server consolidation, the z15 supports up to 32 K subchannels per FICON ExpressSE, FICON Express16S+ and FICON Express16S channels (CHPID). More devices can be defined per FICON channel, which includes primary, secondary, and alias devices. The maximum number of subchannels across all device types that are addressable within an LPAR remains at 63.75 K for subchannel set 0 and 64 K (64 X 1024)-1 for subchannel sets 1, 2, and 3.

Important: FICON Express16SA is *only available on z15 T01* (new build system). z15 T02 supports FICON Express16S+ for new build and as carry forward.

This support is available to the z15, z14, z13, and z13s servers and applies to the FICON ExpressSE, FICON Express16S+, and FICON Express16S features (defined as CHPID type FC). FICON Express8S remains at 24 subchannel support when defined as CHPID type FC.

The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

Request node identification data

The request node identification data (RNID) function for native FICON CHPID type FC allows isolation of cabling-detected errors. The supported operating systems are listed in Table 7-6 on page 262.

FICON link incident reporting

FICON link incident reporting allows an operating system image (without operator intervention) to register link incident reports. The supported operating systems are listed in Table 7-6 on page 262.

Health Check for FICON Dynamic routing

Starting with z13, the channel microcode was changed to support FICON dynamic routing. Although change is required in z/OS to support dynamic routing, I/O errors can occur if the FICON switches are configured for dynamic routing despite the missing support in the processor or storage controllers. Therefore, a health check is provided that interrogates the switch to determine whether dynamic routing is enabled in the switch fabric.

No action is required on z/OS to enable the health check; it is automatically enabled at IPL and reacts to changes that might cause problems. The health check can be disabled by using the **PARMLIB** or **SDSF** modify commands.

The supported operating systems are listed in Table 7-6 on page 262. For more information about FICON Dynamic Routing (FIDR), see Chapter 4, “Central processor complex I/O structure” on page 153.

Global resource serialization FICON CTC toleration

For some configurations that depend on ESCON CTC definitions, global resource serialization (GRS) FICON CTC toleration that is provided with APAR OA38230 is essential, especially after ESCON channel support was removed from IBM Z starting with zEC12.

The supported operating systems are listed in Table 7-6 on page 262.

Increased performance for the FCP protocol

The FCP LIC is modified to help increase I/O operations per second for small and large block sizes, and to support 16-Gbps link speeds.

For more information about FCP channel performance, see [the performance technical papers that are available](#) at the IBM Z I/O connectivity page of the IBM IT infrastructure website.

The FCP protocol is supported by z/VM, z/VSE, and Linux on Z. The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

T10-DIF support

American National Standards Institute (ANSI) T10 Data Integrity Field (DIF) standard is supported on IBM Z for SCSI end-to-end data protection on fixed block (FB) LUN volumes. IBM Z provides added end-to-end data protection between the operating system and the DS8870 unit. This support adds protection information that consists of Cyclic Redundancy Checking (CRC), Logical Block Address (LBA), and host application tags to each sector of FB data on a logical volume.

IBM Z support applies to FCP channels only. The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) allows multiple system images (in LPARs or z/VM guests) to use a single FCP channel as though each were the sole user of the channel. First introduced with z9 EC, this feature can be used with supported FICON features on z14 servers. The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

Worldwide port name tool

Part of the z15 system installation is the pre-planning of the SAN environment. IBM includes a stand-alone tool to assist with this planning before the installation.

The capabilities of the WWPN are extended to calculate and show WWPNs for virtual and physical ports ahead of system installation.

The tool assigns WWPNs to each virtual FCP channel or port by using the same WWPN assignment algorithms that a system uses when assigning WWPNs for channels that use NPIV. Therefore, the SAN can be set up in advance, which allows operations to proceed much faster after the server is installed. In addition, the SAN configuration can be retained instead of altered by assigning the WWPN to physical FCP ports when a FICON feature is replaced.

The WWPN tool takes a .csv file that contains the FCP-specific I/O device definitions and creates the WWPN assignments that are required to set up the SAN. A binary configuration file that can be imported later by the system is also created. The .csv file can be created manually or exported from the HCD/HCM. The supported operating systems are listed in Table 7-6 on page 262 and Table 7-7 on page 264.

The WWPN tool is applicable to all FICON channels that are defined as CHPID type FCP (for communication with SCSI devices) on z15. It is available [for download at the Resource Link](#) at the following website (log in is required).

Note: An optional feature can be ordered for WWPN persistency before shipment to keep the same I/O serial number on the new CPC. Current information must be provided during the ordering process.

7.4.5 Networking features and functions

In this section, we describe the networking features and functions.

25GbE RoCE Express2.1 and 25GbE RoCE Express2

Based on the RoCE Express2 generation hardware, the 25GbE RoCE Express2 (FC 0430 and 0450) provides two 25GbE physical ports and requires 25GbE optics and Ethernet switch 25GbE support. The switch port must support 25GbE (negotiation down to 10GbE is not supported).

The 25GbE RoCE Express2 has one PCHID and the same virtualization characteristics and the 10GbE RoCE Express2 (FC 0412 and FC 0432); that is, 126 Virtual Functions per PCHID.

z/OS requires fixes for APAR OA55686. RMF 2.2 and later is also enhanced to recognize the CX4 card type and properly display CX4 cards in the PCIe Activity reports.

25GbE RoCE Express2 feature also are used by Linux on Z for applications that are coded to the native RoCE verb interface or use Ethernet (such as TCP/IP). This native exploitation does not require a peer OSA.

Support for select Linux on Z distributions is now provided for Shared Memory Communications over Remote Direct Memory Access (SMC-R) by using RoCE Express features. For more information, see [this Linux on Z Blogspot web page](#).

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

10GbE RoCE Express2.1 and 10GbE RoCE Express2

IBM 10GbE RoCE Express2 provides a natively attached PCIe I/O Drawer-based Ethernet feature that supports 10 Gbps Converged Enhanced Ethernet (CEE) and RDMA over CEE (RoCE). The RoCE feature, with an OSA feature, enables shared memory communications between two CPCs by using a shared switch.

RoCE Express2 provides increased virtualization (sharing capability) by supporting 63 Virtual Functions (VFs) per physical port for a total of 126 VFs per PCHID. This configuration allows RoCE to be extended to more workloads.

z/OS Communications Server (CS) provides a new software device driver ConnectX4 (CX4) for RoCE Express2. The device driver is not apparent to both upper layers of the CS (the SMC-R and TCP/IP stack) and application software (using TCP sockets). RoCE Express2 introduces a minor change in how the physical port is configured.

RMF 2.2 and later is also enhanced to recognize the new CX4 card type and properly display CX4 cards in the PCIE Activity reports.

Support in select Linux on Z distributions is now provided for Shared Memory Communications over Remote Direct Memory Access (SMC-R) using the supported RoCE Express features. For more information, see [this Linux on Z Blogspot web page](#).

The 10GbE RoCE Express2 feature also is used by Linux on Z for applications that are coded to the native RoCE verb interface or use Ethernet (such as TCP/IP). This native use does not require a peer OSA.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

10GbE RoCE Express

z14 servers support carrying forward the 10GbE RoCE Express feature. This feature provides support to the second port on the adapter and sharing the ports to up to 31 partitions (per adapter) by using both ports.

The 10GbE RoCE Express feature reduces consumption of CPU resources for applications that use the TCP/IP stack (such as WebSphere accessing a Db2 database). Use of the 10GbE RoCE Express feature also can help reduce network latency with memory-to-memory transfers by using Shared Memory Communications over Remote Direct Memory Access (SMC-R) in z/OS V2R1 or later.

It is transparent to applications and can be used for LPAR-to-LPAR communication on a single z14 server or for server-to-server communication in a multiple CPC environment.

Support in select Linux on Z distributions is now provided for Shared Memory Communications over Remote Direct Memory Access (SMC-R) by using RoCE Express features. For more information, see [this Linux on Z Blogspot web page](#).

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

Shared Memory Communication - Direct Memory Access

First introduced with z13 servers, the Shared Memory Communication - Direct Memory Access (SMC-D) feature maintains the socket-API transparency aspect of SMC-R so that applications that use TCPI/IP communications can benefit immediately without requiring application software to undergo IP topology changes.

Similar to SMC-R, this protocol uses shared memory architectural concepts that eliminate TCP/IP processing in the data path, yet preserve TCP/IP Qualities of Service for connection management purposes.

Support in select Linux on Z distributions is now provided for Shared Memory Communications over Direct Memory Access (SMC-D). For more information, see [this Linux on Z Blogspot web page](#).

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

HiperSockets Completion Queue

The HiperSockets Completion Queue function is implemented on z15, z14, z13, and z13s. This function is designed to allow HiperSockets to transfer data synchronously (if possible) and asynchronously, if necessary. Therefore, it combines ultra-low latency with more tolerance for traffic peaks. HiperSockets Completion Queue can be especially helpful in burst situations. The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

HiperSockets Virtual Switch Bridge

The HiperSockets Virtual Switch Bridge is implemented on z15, z14, z13, and z13s. With the HiperSockets Virtual Switch Bridge, z/VM virtual switch is enhanced to transparently bridge a guest virtual machine network connection on a HiperSockets LAN segment. This bridge allows a single HiperSockets guest virtual machine network connection to also directly communicate with the following components:

- ▶ Other guest virtual machines on the virtual switch
- ▶ External network hosts through the virtual switch OSA UPLINK port

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

HiperSockets Multiple Write Facility

The HiperSockets Multiple Write Facility allows the streaming of bulk data over a HiperSockets link between two LPARs. Multiple output buffers are supported on a single Signal Adapter (SIGA) write instruction. The key advantage of this enhancement is that it allows the receiving LPAR to process a much larger amount of data per I/O interrupt. This process is transparent to the operating system in the receiving partition. HiperSockets Multiple Write Facility with fewer I/O interrupts is designed to reduce processor utilization of the sending and receiving partitions.

Support for this function is required by the sending operating system. For more information, see “HiperSockets” on page 193. The supported operating systems are listed in Table 7-8 on page 266.

HiperSockets support of IPV6

IPv6 is expected to be a key element in the future of networking. The IPv6 support for HiperSockets allows compatible implementations between external networks and internal HiperSockets networks. The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

HiperSockets Layer 2 support

For flexible and efficient data transfer for IP and non-IP workloads, the HiperSockets internal networks on z14 can support two transport modes: Layer 2 (Link Layer) and the current Layer 3 (Network or IP Layer). Traffic can be Internet Protocol (IP) Version 4 or Version 6 (IPv4, IPv6) or non-IP (AppleTalk, DECnet, IPX, NetBIOS, or SNA).

HiperSockets devices are protocol-independent and Layer 3-independent. Each HiperSockets device features its own Layer 2 Media Access Control (MAC) address. This MAC address allows the use of applications that depend on the existence of Layer 2 addresses, such as Dynamic Host Configuration Protocol (DHCP) servers and firewalls.

Layer 2 support can help facilitate server consolidation. Complexity can be reduced, network configuration is simplified and intuitive, and LAN administrators can configure and maintain the mainframe environment the same way as they do a non-mainframe environment.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

HiperSockets network traffic analyzer for Linux on Z

Introduced with IBM System z10, HiperSockets network traffic analyzer (HS NTA) provides support for tracing Layer2 and Layer3 HiperSockets network traffic in Linux on Z. This support allows Linux on Z to control the trace for the internal virtual LAN to capture the records into host memory and storage (file systems).

Linux on Z tools can be used to format, edit, and process the trace records for analysis by system programmers and network administrators.

OSA-Express7S 25 Gigabit Ethernet SR

OSA-Express7S 25 GbE SR1.1 (FC 0449) and OSA-Express7S 25GbE (FC 0429) are installed in the PCIe I/O drawer and have one 25GbE physical port and requires 25GbE optics and Ethernet switch 25GbE support (negotiation down to 10GbE is not supported).

Note: OSA Express7S features are available on z15 T01 only, except for OSA-Express7S 25GbE SR (FC 0429).

New build z15 T02 supports OSA-Express6S (all features) and OSA-Express7S 25GbE SR (FC 0429).

Consider the following points regarding operating system support:

- ▶ z/OS V2R1, V2R2, and V2R3 require fixes for the following APARs: OA55256 (VTAM) and PI95703 (TCP/IP).
- ▶ z/VM V6R4 and V7R1 require PTF for APAR PI99085.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express7S 10-Gigabit Ethernet LR and SR

OSA-Express6S 10-Gigabit Ethernet features are installed in the PCIe I/O drawer, which is supported by the 16 GBps PCIe Gen3 host bus. The performance characteristics are comparable to the OSA-Express6S features and they also retain the same form factor and port granularity.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express6S 10-Gigabit Ethernet LR and SR

OSA-Express6S 10-Gigabit Ethernet features are installed in the PCIe I/O drawer, which is supported by the 16 GBps PCIe Gen3 host bus. The performance characteristics are comparable to the OSA-Express5S features and they also retain the same form factor and port granularity. OSA-Express6S features have been introduced with z14 and can be carried forward to a z15 (T01 and T02), as well as ordered with a new z15 T02

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express5S 10-Gigabit Ethernet LR and SR

Introduced with the zEC12 and zBC12, the OSA-Express5S 10-Gigabit Ethernet feature is installed exclusively in the PCIe I/O drawer. Each feature includes one port, which is defined as CHPID type OSD that supports the queued direct input/output (QDIO) architecture for high-speed TCP/IP communication.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express7S Gigabit Ethernet LX and SX

z15 introduces an Ethernet technology refresh with OSA-Express7S Gigabit Ethernet features to be installed in the PCIe I/O drawer, which is supported by the 16 GBps PCIe Gen3 host bus. The performance characteristics are comparable to the OSA-Express6S features and they also retain the same form factor and port granularity.

Note: OSA Express7S features are available on z15 T01 only, except for OSA-Express7S 25GbE SR (FC 0429).

New build z15 T02 supports OSA-Express6S (all features) and OSA-Express7S 25GbE SR (FC 0429).

Each adapter can be configured in the following modes:

- ▶ QDIO mode, with CHPID types OSD
- ▶ Local 3270 emulation mode, including OSA-ICC, with CHPID type OSC

Note: Operating system support is required to recognize and use the second port on the OSA-Express6S 1000BASE-T Ethernet feature.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express6S Gigabit Ethernet LX and SX

z14 introduces an Ethernet technology refresh with OSA-Express6S Gigabit Ethernet features to be installed in the PCIe I/O drawer, which is supported by the 16 Gbps PCIe Gen3 host bus. The performance characteristics are comparable to the OSA-Express5S features and they also retain the same form factor and port granularity. OSA-Express6S features have been introduced with z14 and can be carried forward to a z15 (T01 and T02), as well as ordered with a new z15 T02.

Note: Operating system support is required to recognize and use the second port on the OSA-Express6S Gigabit Ethernet feature.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express5S Gigabit Ethernet LX and SX

The OSA-Express5S Gigabit Ethernet feature is installed exclusively in the PCIe I/O drawer. Each feature includes one PCIe adapter and two ports. The two ports share a channel path identifier (CHPID type OSD exclusively). Each port supports attachment to a 1 Gigabit per second (Gbps) Ethernet LAN. The ports can be defined as a spanned channel, and can be shared among LPARs and across logical channel subsystems.

Note: Operating system support is required to recognize and use the second port on the OSA-Express5S Gigabit Ethernet feature.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express7S 1000BASE-T Ethernet

z15 introduces an Ethernet technology refresh with OSA-Express7S 1000BASE-T Ethernet features to be installed in the PCIe I/O drawer, which is supported by the 16 Gbps PCIe Gen3 host bus. The performance characteristics are comparable to the OSA-Express6S features and they also retain the same form factor and port granularity.

Each adapter can be configured in the following modes:

- ▶ QDIO mode, with CHPID types OSD
- ▶ Non-QDIO mode, with CHPID type OSE
- ▶ Local 3270 emulation mode, including OSA-ICC, with CHPID type OSC

Notes: Consider the following points:

- ▶ Operating system support is required to recognize and use the second port on the OSA-Express7S 1000BASE-T Ethernet feature.
- ▶ OSA-Express7S 1000BASE-T Ethernet feature supports only 1000 Mbps duplex mode (no auto-negotiation to 100 or 10 Mbps)

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express6S 1000BASE-T Ethernet

z14 introduces an Ethernet technology refresh with OSA-Express6S 1000BASE-T Ethernet features to be installed in the PCIe I/O drawer, which is supported by the 16 Gbps PCIe Gen3 host bus. The performance characteristics are comparable to the OSA-Express5S features

and they also retain the same form factor and port granularity. OSA-Express6S features have been introduced with z14 and can be carried forward to a z15 (T01 and T02), as well as ordered with a new z15 T02.

Each adapter can be configured in the following modes:

- ▶ QDIO mode, with CHPID types OSD
- ▶ Non-QDIO mode, with CHPID type OSE
- ▶ Local 3270 emulation mode, including OSA-ICC, with CHPID type OSC

Note: Operating system support is required to recognize and use the second port on the OSA-Express6S 1000BASE-T Ethernet feature.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Express5S 1000BASE-T Ethernet

The OSA-Express5S 1000BASE-T Ethernet feature is installed exclusively in the PCIe I/O drawer. Each feature includes one PCIe adapter and two ports. The two ports share a CHPID, which can be defined as OSC, OSD or OSE. The ports can be defined as a spanned channel, and can be shared among LPARs and across logical channel subsystems.

Each adapter can be configured in the following modes:

- ▶ QDIO mode, with CHPID types OSD
- ▶ Non-QDIO mode, with CHPID type OSE
- ▶ Local 3270 emulation mode, including OSA-ICC, with CHPID type OSC

Note: Operating system support is required to recognize and use the second port on the OSA-Express5S 1000BASE-T Ethernet feature.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA-Integrated Console Controller

The OSA-Express 1000BASE-T Ethernet features provide the Integrated Console Controller (OSA-ICC) function, which supports TN3270E (RFC 2355) and non-SNA DFT 3270 emulation. The OSA-ICC function is defined as CHPID type OSC and console controller, and includes multiple LPAR support as shared or spanned channels.

With the OSA-ICC function, 3270 emulation for console session connections is integrated in the z15 through a port on the OSA-Express7S 1000BASE-T, OSA-Express7S GbE, OSA-Express6S 1000BASE-T, or OSA-Express5S 1000BASE-T.

OSA-ICC can be configured on a PCHID-by-PCHID basis, and is supported at any of the feature settings. Each port can support up to 120 console session connections.

To improve security of console operations and to provide a secure, validated connectivity, OSA-ICC supports Transport Layer Security/Secure Sockets Layer (TLS/SSL) with Certificate Authentication starting with z13 GA2 (Driver level 27).

Note: OSA-ICC supports up to 48 *secure* sessions per CHPID (the overall maximum of 120 connections is unchanged).

OSA-ICC Enhancements

With HMC 2.14.1 and newer the following enhancements are available:

- ▶ The IPv6 communications protocol is supported by OSA-ICC 3270 so that clients can comply with regulations that require all computer purchases to support IPv6.
- ▶ TLS negotiation levels (the supported TLS protocol levels) for the OSA-ICC 3270 client connection can now be specified:
 - TLS 1.0 OSA-ICC 3270 server permits TLS 1.0, TLS 1.1, and TLS 1.2 client connections.
 - TLS 1.1 OSA-ICC 3270 server permits TLS 1.1 and TLS 1.2 client connections.
 - TLS 1.2 OSA-ICC 3270 server permits only TLS 1.2 client connections.
- ▶ Separate and unique OSA-ICC 3270 certificates are supported (for each PCHID), for the benefit of customers who host workloads across multiple business units or data centers, where cross-site coordination is required. Customers can avoid interruption of all the TLS connections at the same time when having to renew expired certificates. OSA-ICC continues to also support a single certificate for all OSA-ICC PCHIDs in the system.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

Checksum offload for in QDIO mode (CHPID type OSD)

Checksum offload provides the capability of calculating the Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and IP header checksum. Checksum verifies the accuracy of files. By moving the checksum calculations to a Gigabit or 1000BASE-T Ethernet feature, host processor cycles are reduced and performance is improved.

Checksum offload provides checksum offload for several types of traffic and is supported by the following features when configured as CHPID type OSD (QDIO mode only):

- ▶ OSA-Express7S GbE
- ▶ OSA-Express7S 1000BASE-T Ethernet
- ▶ OSA-Express6S GbE
- ▶ OSA-Express6S 1000BASE-T Ethernet
- ▶ OSA-Express5S GbE
- ▶ OSA-Express5S 1000BASE-T Ethernet

When checksum is offloaded, the OSA-Express feature runs the checksum calculations for Internet Protocol version 4 (IPv4) and Internet Protocol version 6 (IPv6) packets. The checksum offload function applies to packets that go to or come from the LAN.

When multiple IP stacks share an OSA-Express, and an IP stack sends a packet to a next hop address that is owned by another IP stack that is sharing the OSA-Express, OSA-Express sends the IP packet directly to the other IP stack. The packet does not have to be placed out on the LAN, which is termed LPAR-to-LPAR traffic. Checksum offload is enhanced to support the LPAR-to-LPAR traffic, which was not originally available.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

Querying and displaying an OSA configuration

OSA-Express3 introduced the capability for the operating system to query and display directly the current OSA configuration information (similar to OSA/SF). z/OS uses this OSA capability

by introducing the TCP/IP operator command `display OSAINFO`. z/VM provides this function with the `NETSTAT OSAINFO TCP/IP` command.

The use of `display OSAINFO` (z/OS) or `NETSTAT OSAINFO` (z/VM) allows the operator to monitor and verify the current OSA configuration and helps improve the overall management, serviceability, and usability of OSA-Express cards.

These commands apply to CHPID type OSD. The supported operating systems are listed in Table 7-8 on page 266.

QDIO data connection isolation for z/VM

The QDIO data connection isolation function provides a higher level of security when sharing an OSA connection in z/VM environments that use VSWITCH. The VSWITCH is a virtual network device that provides switching between OSA connections and the connected guest systems.

QDIO data connection isolation allows disabling internal routing for each QDIO connected. It also provides a means for creating security zones and preventing network traffic between the zones.

QDIO data connection isolation is supported by all OSA-Express features on z15. The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

QDIO interface isolation for z/OS

Some environments require strict controls for routing data traffic between servers or nodes. In certain cases, the LPAR-to-LPAR capability of a shared OSA connection can prevent such controls from being enforced. With interface isolation, internal routing can be controlled on an LPAR basis. When interface isolation is enabled, the OSA discards any packets that are destined for a z/OS LPAR that is registered in the OSA Address Table (OAT) as isolated.

QDIO interface isolation is supported on all OSA-Express features on z15. The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

QDIO optimized latency mode

QDIO optimized latency mode (OLM) can help improve performance for applications that feature a critical requirement to minimize response times for inbound and outbound data.

OLM optimizes the interrupt processing in the following manner:

- ▶ For inbound processing, the TCP/IP stack looks more frequently for available data to process. This process ensures that any new data is read from the OSA-Express features without needing more program controlled interrupts (PCIs).
- ▶ For outbound processing, the OSA-Express cards also look more frequently for available data to process from the TCP/IP stack. Therefore, the process does not require a Signal Adapter (SIGA) instruction to determine whether more data is available.

The supported operating systems are listed in Table 7-8 on page 266.

QDIO Diagnostic Synchronization

QDIO Diagnostic Synchronization enables system programmers and network administrators to coordinate and simultaneously capture software and hardware traces. It allows z/OS to signal OSA-Express features (by using a diagnostic assist function) to stop traces and capture the current trace records.

QDIO Diagnostic Synchronization is supported by the OSA-Express features on z15 when in QDIO mode (CHPID type OSD). The supported operating systems are listed in Table 7-8 on page 266.

Adapter interruptions for QDIO

Linux on Z and z/VM work together to provide performance improvements by using extensions to the QDIO architecture. First added to z/Architecture with HiperSockets, adapter interruptions provide an efficient, high-performance technique for I/O interruptions to reduce path lengths and processor usage. These reductions are in the host operating system and the adapter (supported OSA-Express cards when CHPID type OSD is used).

In extending the use of adapter interruptions to OSD (QDIO) channels, the processor utilization to handle a traditional I/O interruption is reduced. This configuration benefits OSA-Express TCP/IP support in z/VM, z/VSE, and Linux on Z. The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

Inbound workload queuing (IWQ) for OSA

OSA-Express3 introduced inbound workload queuing (IWQ), which creates multiple input queues and allows OSA to differentiate workloads “off the wire.” It then assigns work to a specific input queue (per device) to z/OS.

Each input queue is a unique type of workload, and has unique service and processing requirements. The IWQ function allows z/OS to preassign the appropriate processing resources for each input queue. This approach allows multiple concurrent z/OS processing threads to process each unique input queue (workload), which avoids traditional resource contention.

IWQ reduces the conventional z/OS processing that is required to identify and separate unique workloads. This advantage results in improved overall system performance and scalability.

A primary objective of IWQ is to provide improved performance for business-critical interactive workloads by reducing contention that is created by other types of workloads. In a heavily mixed workload environment, this “off the wire” network traffic separation is provided by OSA-Express7S, OSA-Express6S, or OSA-Express5S¹¹ features that are defined as CHPID type OSD. OSA IWQ is shown in Figure 7-5.

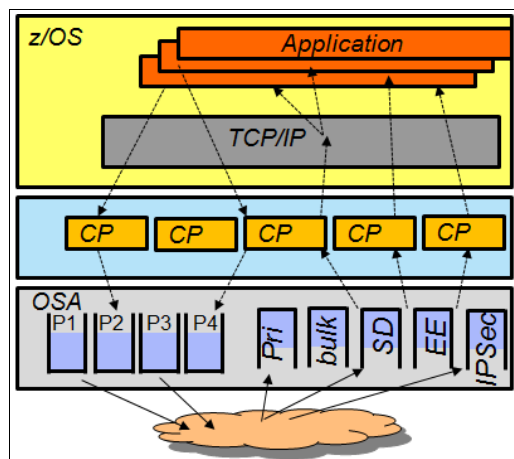


Figure 7-5 OSA inbound workload queuing

¹¹ Only OSA-Express6S and OSA-Express5S cards are supported on z15 as carry forward.

The following types of z/OS workloads are identified and assigned to unique input queues:

- ▶ z/OS Sysplex Distributor traffic
Network traffic that is associated with a distributed virtual Internet Protocol address (VIPA) is assigned to a unique input queue. This configuration allows the Sysplex Distributor traffic to be immediately distributed to the target host.
- ▶ z/OS bulk data traffic
Network traffic that is dynamically associated with a streaming (bulk data) TCP connection is assigned to a unique input queue. This configuration allows the bulk data processing to be assigned the appropriate resources and isolated from critical interactive workloads.
- ▶ EE (Enterprise Extender / SNA traffic)
IWQ for the OSA-Express features is enhanced to differentiate and separate inbound Enterprise Extender traffic to a dedicated input queue.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

VLAN management enhancements

VLAN management enhancements are valid for supported OSA-Express features on z15 defines as CHPID type OSD. The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

GARP VLAN Registration Protocol

All OSA-Express features support VLAN prioritization, which is a component of the IEEE 802.1 standard. GARP VLAN Registration Protocol (GVRP) support allows an OSA-Express port to register or unregister its VLAN IDs with a GVRP-capable switch and dynamically update its table as the VLANs change. This process simplifies the network administration and management of VLANs because manually entering VLAN IDs at the switch is no longer necessary. The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

Link aggregation support for z/VM

Link aggregation (IEEE 802.3ad) that is controlled by the z/VM Virtual Switch (VSWITCH) allows the dedication of an OSA-Express port to the z/VM operating system. The port must be participating in an aggregated group that is configured in Layer 2 mode. Link aggregation (trunking) combines multiple physical OSA-Express ports into a single logical link. This configuration increases throughput, and provides nondisruptive failover if a port becomes unavailable. The target links for aggregation must be of the same type.

Link aggregation is applicable to CHPID type OSD (QDIO). The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

Multi-VSwitch Link Aggregation

Multi-VSwitch Link Aggregation support allows a port group of OSA-Express features to span multiple virtual switches within a single z/VM system or between multiple z/VM systems. Sharing a Link Aggregation Port Group (LAG) with multiple virtual switches increases optimization and utilization of the OSA-Express features when handling larger traffic loads.

Higher adapter utilization protects customer investments, which is increasingly important as 10 GbE deployments become more prevalent. The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

Large send for IPv6 packets

Large send for IPv6 packets improves performance by offloading outbound TCP segmentation processing from the host to an OSA-Express feature by using a more efficient memory transfer into it.

Large send support for IPv6 packets applies to the OSA-Express7S, OSA-Express6S, and OSA-Express5S¹¹ features (CHPID type OSD) on z15, z14, z13, and z13s.

z13 added support of large send for IPv6 packets (segmentation offloading) for LPAR-to-LPAR traffic. OSA-Express6S added TCP checksum on large send, which reduces the cost (CPU time) of error detection for large send.

The supported operating systems are listed in Table 7-8 on page 266 and Table 7-9 on page 268.

OSA Dynamic LAN idle

The OSA Dynamic LAN idle parameter change helps reduce latency and improve performance by dynamically adjusting the inbound blocking algorithm. System administrators can authorize the TCP/IP stack to enable a dynamic setting that previously was static.

The blocking algorithm is modified based on the following application requirements:

- ▶ For latency-sensitive applications, the blocking algorithm is modified considering latency.
- ▶ For streaming (throughput-sensitive) applications, the blocking algorithm is adjusted to maximize throughput.

In all cases, the TCP/IP stack determines the best setting based on the current system and environmental conditions, such as inbound workload volume, processor utilization, and traffic patterns. It can then dynamically update the settings.

Supported OSA-Express features adapt to the changes, which avoids thrashing and frequent updates to the OAT. Based on the TCP/IP settings, OSA holds the packets before presenting them to the host. A dynamic setting is designed to avoid or minimize host interrupts.

OSA Dynamic LAN idle is supported by the OSA-Express7S, OSA-Express6S, and OSA-Express5S features on z15 when in QDIO mode (CHPID type OSD). The supported operating systems are listed in Table 7-8 on page 266.

OSA Layer 3 virtual MAC for z/OS environments

To help simplify the infrastructure and facilitate load balancing when an LPAR is sharing an OSA MAC address with another LPAR, each operating system instance can have its own unique logical or virtual MAC (VMAC) address. All IP addresses that are associated with a TCP/IP stack are accessible by using their own VMAC address instead of sharing the MAC address of an OSA port. This situation also applies to Layer 3 mode and to an OSA port spanned among channel subsystems.

OSA Layer 3 VMAC is supported by the OSA-Express7S, OSA-Express6S, and OSA-Express5S features on z15 when in QDIO mode (CHPID type OSD). The supported operating systems are listed in Table 7-8 on page 266.

Network Traffic Analyzer

The z15 offers systems programmers and network administrators the ability to more easily solve network problems despite high traffic. With the OSA-Express Network Traffic Analyzer and QDIO Diagnostic Synchronization on the server, you can capture trace and trap data.

This data can then be forwarded to z/OS tools for easier problem determination and resolution.

The Network Traffic Analyzer is supported by the OSA-Express7S, OSA-Express6S, and OSA-Express5S features on z15 when in QDIO mode (CHPID type OSD). The supported operating systems are listed in Table 7-8 on page 266.

7.4.6 Cryptography Features and Functions Support

IBM z15™ provides the following major groups of cryptographic functions:

- ▶ Synchronous cryptographic functions, which are provided by CPACF
- ▶ Asynchronous cryptographic functions, which are provided by the Crypto Express7S feature

The minimum software support levels are described in the following sections. Review the current PSP buckets to ensure that the latest support levels are known and included as part of the implementation plan.

CP Assist for Cryptographic Function

Central Processor Assist for Cryptographic Function (CPACF), which is standard¹² on every z15 core, now supports pervasive encryption. Simple policy controls allow business to enable encryption to protect data in mission-critical databases without needing to stop the database or re-create database objects. Database administrators can use z/OS Dataset Encryption, z/OS Coupling Facility Encryption, z/VM encrypted hypervisor paging, and z/TPF transparent database encryption, which use the performance enhancements in the hardware.

CPACF supports the following features in z15:

- ▶ Advanced Encryption Standard (AES, symmetric encryption)
- ▶ Data Encryption Standard (DES, symmetric encryption)
- ▶ Secure Hash Algorithm (SHA, hashing)
- ▶ SHAKE Algorithms
- ▶ True Random Number Generation (TRNG)
- ▶ Improved GCM (Galois Counter Mode) encryption (enabled by a single hardware instruction)

In addition, the z15 core implements a Modulo Arithmetic unit in support of Elliptic Curve Cryptography.

CPACF also is used by several IBM software product offerings for z/OS, such as IBM WebSphere Application Server for z/OS. For more information, see 6.4, “CP Assist for Cryptographic Functions” on page 224.

The supported operating systems are listed in Table 7-10 on page 271 and Table 7-11 on page 272.

¹² CPACF hardware is implemented on each z15 core. CPACF functionality is enabled with FC 3863.

Crypto Express7S

Introduced with z15, Crypto Express7S includes a single- or dual-port adapter (single or dual IBM 4769 PCIe Cryptographic Co-processor [PCIeCC]) and complies with the following Physical Security Standards:

- ▶ FIPS 140-2 level 4
- ▶ Common Criteria EP11 EAL4
- ▶ Payment Card Industry (PCI) HSM
- ▶ German Banking Industry Commission (GBIC, formerly DK)

Support of Crypto Express6S functions varies by operating system and release and by the way the card is configured as a coprocessor or an accelerator. For more information, see 6.5, “Crypto Express7S” on page 230. The supported operating systems are listed in Table 7-10 on page 271 and Table 7-11 on page 272.

Crypto Express6S (carry forward on z15)

Introduced with z14, Crypto Express6S includes one IBM 4768 PCIe Cryptographic Co-processor (PCIeCC) and complies with the following Physical Security Standards:

- ▶ FIPS 140-2 level 4
- ▶ Common Criteria EP11 EAL4
- ▶ Payment Card Industry (PCI) HSM
- ▶ German Banking Industry Commission (GBIC, formerly DK)

Support of Crypto Express6S functions varies by operating system and release and by the way the card is configured as a coprocessor or an accelerator. For more information, see 6.5, “Crypto Express7S” on page 230. The supported operating systems are listed in Table 7-10 on page 271 and Table 7-11 on page 272.

Crypto Express5S (carry forward on z15)

Support of Crypto Express5S functions varies by operating system and release and by the way the card is configured as a coprocessor or an accelerator. The supported operating systems are listed in Table 7-10 on page 271 and Table 7-11 on page 272.

Web deliverables

For more information about web-deliverable code on z/OS, see [the z/OS downloads website](#).

For Linux on Z, support is delivered through IBM and the distribution partners. For more information, see [Linux on Z on the IBM developerWorks website](#).

z/OS Integrated Cryptographic Service Facility

Integrated Cryptographic Service Facility (ICSF) is a base component of z/OS. It is designed to transparently use the available cryptographic functions, whether CPACF or Crypto Express, to balance the workload and help address the bandwidth requirements of the applications.

Despite being a z/OS base component, ICSF functions are generally made available through web deliverable support a few months after a new z/OS release. Therefore, new functions are related to an ICSF function modification identifier (FMID) instead of a z/OS version.

ICSF HCR77D1 - Cryptographic Support for z/OS V2R2, V2R3, and V2R4

z/OS V2.2, V2.3, and V2.4 require ICSF Web Deliverable WD19 (HCR77D1) to support the following features:

- ▶ Support for CCA 5.5 and CCA 6.3:
 - PCI HSM Phase 4 (AES and RSA) and ANSI TR-34

- ICSF Support with an SPE for HCR77D0
- ▶ Support for Crypto Express7S
- ▶ Support for more than 16 Adapters
- ▶ Support for carry forward for Crypto Express5S and Crypto Express6S
- ▶ Support for:
 - EP11 and ECC Protected Key
 - CPACF ECC Enablement MSA-9
 - EP11 and CCA Support for new ECC Curves
 - FPE Voltage Algorithms
 - Post Quantum Crypto PoC

ICSF HCR77D0 - Cryptographic Support for z/OS V2R2 and z/OS V2R3

z/OS V2.2 and V2.3 require ICSF Web Deliverable WD18 (HCR77D0) to support the following features:

- ▶ Support for the updated German Banking standard (DK):
 - CCA 5.4 & 6.1¹³:
 - ISO-4 PIN Blocks (ISO-9564-1)
 - Directed keys: A key can either encrypt or decrypt data, but not both.
 - Allow AES transport keys to be used to export/import *DES* keys in a standard ISO 20038 key block. This feature helps with interoperability between CCA and non-CCA systems.
 - Allow AES transport keys to be used to export/import a small subset of *AES* keys in a standard ISO 20038 key block. This feature helps with interoperability between CCA and non-CCA systems.
 - Triple-length TDES keys with Control Vectors for increased data confidentiality.
 - CCA 6.2: PCI HSM 3K DES - Support for triple length DES keys (standards compliance)
- ▶ EP11 Stage 4:
 - New elliptic curve algorithms for PKCS#11 signature, key derivation operations:
 - Ed448 elliptic curve
 - EC25519 elliptic curve
 - EP11 Concurrent Patch Apply: Allows service to be applied to the EP11 coprocessor dynamically without taking the crypto adapter offline (already available for CCA coprocessors).
 - eIDAS compliance: eIDAS cross-border EU regulation for portable recognition of electronic identification.

ICSF HCR77C1 - Cryptographic Support for z/OS V2R1 - z/OS V2R3

ICSF Web Deliverable HCR77C1 provides support for the following features:

- ▶ Usage and administration of Crypto Express6S

This feature might be configured as an accelerator (CEX6A), a CCA coprocessor (CEX6C), or an EP-11 coprocessor (CEX6P).
- ▶ Coprocessor in PCI-HSM Compliance Mode (enablement requires TKE 9.0 or newer).

¹³ CCA 5.4 and 6.1 enhancements are also supported for z/OS V2R1 with ICSF HCR77C1 (WD17) with SPEs (Small Program Enhancements (z/OS continuous delivery model)).

- ▶ z14 CPACF support. For more information, see “CP Assist for Cryptographic Function” on page 317.

The following software enhancements are available in ICSF Web Deliverable HCR77C1:

- ▶ Crypto Usage Statistics: When enabled, ICSF aggregates statistics that are related to crypto workloads and logs to an SMF record.
- ▶ Panel-based CKDS Administration: ICSF added an ISPF, panel-driven interface that allows interactive administration (View, Create, Modify, and Delete) of CKDS keys.
- ▶ CICS End User Auditing: When enabled, ICSF retrieves the CICS user identity and includes it as a log string in the SAF resource check. The user identity is not checked for access to the resource. Instead, it is included in the resource check (SMF Type 80) records that are logged for any of the ICSF SAF classes protecting crypto keys and services (CSFKEYS, XCSFKEY, CRYPTOZ, and CSFSERV).

For more information about ICSF versions and FMID cross-references, see the [z/OS: ICSF Version and FMID Cross Reference](#), TD103782, abstract that is available at the IBM Techdoc website.

For PTFs that allow previous levels of ICSF to coexist with the Cryptographic Support for z/OS 2.1 - z/OS V2R3 (HCR77C1) web deliverable, check below FIXCAT, as shown in the following example:

```
IBM.Coexistence.ICSF.z/OS_V2R1-V2R3-HCR77C1
```

RMF Support for Crypto Express7S and Crypto Express6S

RMF enhances the Monitor I Crypto Activity data gatherer to recognize and use performance data for the new Crypto Express7S (CEX7) and CryptoExpress6S (CEX6) cards. RMF supports all valid card configurations on z15 and provides CEX7 and CEX6 crypto activity data in the SMF type 70 subtype 2 records and RMF Postprocessor Crypto Activity Report.

Reporting can be done at an LPAR/domain level to provide more granular reports for capacity planning and diagnosing problems. This feature requires fix for APAR OA54952.

The supported operating systems are listed in Table 7-10 on page 271.

z/OS Data Set Encryption

Aligned with IBM Z Pervasive Encryption initiative, IBM provides application-transparent, policy-controlled dataset encryption in IBM z/OS.

Policy-driven z/OS Data Set Encryption enables users to perform the following tasks:

- ▶ De-couple encryption from data classification; encrypt data automatically independent of labor-intensive data classification work.
- ▶ Encrypt data immediately and efficiently at the time it is written.
- ▶ Reduce risks that are associated with mis-classified or undiscovered sensitive data.
- ▶ Help protect digital assets automatically.
- ▶ Achieve application transparent encryption.

IBM Db2® for z/OS and IBM Information Management System (IMS) intend to use z/OS Data Set Encryption.

With z/OS, Data Set Encryption DFSMS enhances data security with support for data set level encryption by using DFSMS access methods. This function is designed to give users the ability to encrypt their data sets without changing their application programs.

DFSMS users can identify which data sets require encryption by using JCL, Data Class, or the RACF data set profile. Data set level encryption can allow the data to remain encrypted during functions, such as backup and restore, migration and recall, and replication.

z/OS Data Set Encryption requires CP Assist for Cryptographic Functions (CPACF).

Considering the significant enhancements that were introduced with z14, the encryption mode of XTS is used by access method encryption to obtain the best performance possible. It is not recommended to enable z/OS data set encryption until all sharing systems, fallback, backup, and DR systems support encryption.

In addition to applying PTFs enabling the support, ICSF configuration is required. The supported operating systems are listed in Table 7-10 on page 271.

Crypto Analytics Tool for Z

The IBM Crypto Analytics Tool (CAT) for Z is an analytics solution that collects data on your z/OS cryptographic infrastructure, presents reports, and analyzes if any vulnerabilities exist. CAT collects cryptographic information from across the enterprise and provides reports to help users better manage the crypto infrastructure and ensure it follows best practices. The use of CAT can help you deal with managing complex cryptography resources across your organization.

z/VM encrypted hypervisor paging (encrypted paging support)

With the PTF for APAR VM65993, z/VM V6.4 provides support for encrypted paging in support of the z15 pervasive encryption philosophy of encrypting all data in flight and at rest. Ciphering occurs as data moves between active memory and a paging volume that is owned by z/VM.

Included in this support is the ability to dynamically control whether a running z/VM system is encrypting this data. This support protects guest paging data from administrators or users with access to volumes. Enabled with AES encryption, z/VM Encrypted Paging includes low overhead by using CPACF.

The supported operating systems are listed in Table 7-10 on page 271.

z/TPF transparent database encryption

Shipped in August 2016, z/TPF at-rest Data Encryption provides following features and benefits:

- ▶ Automatic encryption of at-rest data by using AES CBC (128 or 256).
- ▶ No application changes required.
- ▶ Database level encryption by using highly efficient CPACF.
- ▶ Inclusion of data on disk and cached in memory.
- ▶ Ability to include data integrity checking (optionally by using SHA-256) to detect accidental or malicious data corruption.
- ▶ Tools to migrate a database from unencrypted to encrypted state or change the encryption key/algorithm for a specific DB while transactions are flowing (no database downtime).

Pervasive encryption for Linux on Z

Pervasive encryption for Linux on Z combines the full power of Linux with z15 capabilities by using the support of the following features:

- ▶ Kernel Crypto: z15 CPACF

- ▶ LUKS dm-crypt Protected-Key CPACF
- ▶ Libica and openssl: z15 CPACF and acceleration of RSA handshakes by using SIMD
- ▶ Secure Service Container: High security virtual appliance deployment infrastructure

Protection of data at-rest

By using the integration of industry-unique hardware accelerated CPACF encryption into the standard Linux components, users can achieve optimized encryption transparently to prevent raw key material from being visible to operating systems and applications.

Because of the potential costs and overheads, most of the organizations avoid the use of host-based network encryption today. By using enhanced CPACF and SIMD on z15, TLS and IPsec can use hardware performance gains while benefitting from transparent enablement. Reduced cost of encryption enables broad use of network encryption.

7.5 z/OS migration considerations

Except for base processor support, z/OS releases do not require any of the functions that are introduced with the z15. Minimal toleration support that is needed depends on z/OS release.

Although z15 servers *do not* require any “functional” software, it is recommended to install all z15 service before upgrading to the new server. The support matrix for z/OS releases and the Z servers that support them are listed in Table 7-16.

Table 7-16 z/OS support summary

| z/OS Release | z10 EC z10 BC WDFM ^a | z196 z114 WDFM ^a | zEC12 zBC12 WDFM ^a | z13 z13s WDFM ^a | z14 | z15 | End of Service | Extended Defect Support ^b |
|--------------|---------------------------------|-----------------------------|-------------------------------|----------------------------|-----|----------------|----------------------|--------------------------------------|
| V2R1 | X | X | X | X | X | X ^b | 09/2018 | 09/2021 |
| V2R2 | X | X | X | X | X | X | 09/2020 ^b | 09/2023 ^b |
| V2R3 | | | X | X | X | X | 09/2022 ^b | 09/2025 ^b |
| V2R4 | | | | X | X | X | - | - |

a. Server was withdrawn from marketing.

b. The IBM Software Support Services provides the ability for customers to purchase extended defect support service for z/OS.

7.5.1 General guidelines

The IBM z15™ introduces the latest IBM Z technology. Although support is provided by z/OS starting with z/OS 2.1, the capabilities and use of z15 depends on the z/OS release. Also, web deliverables¹⁴ are needed for some functions on some releases. In general, consider the following guidelines:

- ▶ Do not change software releases and hardware at the same time.
- ▶ Keep members of the sysplex at the same software level, except during brief migration periods.
- ▶ Migrate to an STP-only network before introducing a z15 into a sysplex.
- ▶ Review any restrictions and migration considerations before creating an upgrade plan.

¹⁴ For example, the use of Crypto Express7S requires the Cryptographic Support for z/OS V2R1 - z/OS V2R3 web deliverable.

- ▶ Acknowledge that some hardware features cannot be ordered or carried forward for an upgrade from an earlier server to z15 and plan accordingly.
- ▶ Determine the changes in IOCP, HCD, and HCM to support defining z15 configuration and the new features and functions it introduces.
- ▶ Ensure that none of the new z/Architecture Machine Instructions (mnemonics) that were introduced with z15 are colliding with the names of Assembler macro instructions you use¹⁵.
- ▶ Check the use of **MACHMIG** statements in **LOADxx** PARMLIB commands.

7.5.2 Hardware Fix Categories (FIXCATs)

Base support includes fixes that are required to run z/OS on the IBM z15™ server. They are identified by:

```
IBM.Device.Server.z15-8561.RequiredService
IBM.Device.Server.z15-8562.RequiredService
```

The use of many functions covers fixes that are required to use the capabilities of the IBM z15™ servers. They are identified by:

```
IBM.Device.Server.z15-8561.Exploitation
IBM.Device.Server.z15-8562.Exploitation
```

Recommended service is identified by:

```
IBM.Device.Server.z15-8561.RecommendedService
IBM.Device.Server.z15-8562.RecommendedService
```

Support for z15 is provided by using a combination of web deliverables and PTFs, which are documented in:

- For z15 T01: PSP Bucket Upgrade = 8561DEVICE, Subset = 8561/ZOS.
- For z15 T02: PSP Bucket Upgrade = 8562DEVICE, Subset = 8562/ZOS.

Consider the following other Fix Categories of Interest:

- ▶ Fixes that are required to use Parallel Sysplex InfiniBand Coupling links:


```
IBM.Function.ParallelSysplexInfiniBandCoupling
```
- ▶ Fixes that are required to use the Server Time Protocol function:


```
IBM.Function.ServerTimeProtocol
```
- ▶ Fixes that are required to use the High-Performance FICON function:


```
IBM.Function.zHighPerformanceFICON
```
- ▶ PTFs that allow previous levels of ICSF to coexist with the latest Cryptographic Support for z/OS V2R2 - z/OS V2R4 (HCR77D1) web deliverable:


```
IBM.Coexistence.ICSF.z/OS_V2R2-V2R4-HCR77D1
```
- ▶ PTFs that allow previous levels of ICSF to coexist with the latest Cryptographic Support for z/OS V2R2 - z/OS V2R3 (HCR77D0) web deliverable:


```
IBM.Coexistence.ICSF.z/OS_V2R2-V2R3-HCR77D0
```
- ▶ PTFs that allow previous levels of ICSF to coexist with the Cryptographic Support for z/OS V2R1 - z/OS V2R3 (HCR77C1) web deliverable:

¹⁵ For more information, see the [Tool to Compare IBM z15 Instruction Mnemonics with Macro Libraries](#) IBM Technote.

Use the SMP/E **REPORT MISSINGFIX** command to determine whether any FIXCAT APARs exist that are applicable and are not yet installed, and whether any SYSMODs are available to satisfy the missing FIXCAT APARs.

For more information about IBM Fix Category Values and Descriptions, see the [IBM Fix Category Values and Descriptions page](#) of the IBM IT infrastructure website.

7.5.3 Coupling links¹⁶

z15 servers support only active participation in the same Parallel Sysplex with z14, z13, and z13s. Configurations with z/OS on one of these servers can add a z15 Server to their Sysplex for a z/OS or a Coupling Facility image.

Configurations with a Coupling Facility on one of these servers can add a z15 Server to their Sysplex for a z/OS or a Coupling Facility image. z15 does not support participating in a Parallel Sysplex with System zEC12/zBC12 and earlier systems.

Each system can use, or not use, internal coupling links, InfiniBand coupling links, or ICA coupling links independently of what other systems are using.

Coupling connectivity is available only when other systems also support the same type of coupling. For more information about supported coupling link technologies on z15, see 4.6.4, “Parallel Sysplex connectivity” on page 195, and the [Coupling Facility Configuration Options](#) white paper.

7.5.4 z/OS XL C/C++ considerations

z/OS V2R4 is required to use the latest level (13) of the following C/C++ compiler options:

- ▶ **ARCHITECTURE**: This option selects the minimum level of system architecture on which the program can run. Certain features that are provided by the compiler require a minimum architecture level. **ARCH(13)** uses instructions that are available on the z15.
- ▶ **TUNE**: This option allows optimization of the application for a specific system architecture within the constraints that are imposed by the **ARCHITECTURE** option. The **TUNE** level must not be lower than the setting in the **ARCHITECTURE** option.

The following new functions provide performance improvements for applications by using new z15 instructions:

- ▶ Vector Programming Enhancements
- ▶ New z15 hardware instruction support
- ▶ Packed Decimal support using vector registers
- ▶ Auto-SIMD enhancements to make use of new data types

To enable the use of new functions, specify **ARCH(13)** and **VECTOR** for compilation. The binaries that are produced by the compiler on z15 can be run only on z15 and above because it uses the vector facility on z15 for new functions. The use of older versions of the compiler on z15 does not enable new functions.

For more information about the **ARCHITECTURE**, **TUNE**, and **VECTOR** compiler options, see *z/OS V2R2.0 XL C/C++ User's Guide*, SC09-4767.

¹⁶ IBM z15 does not support InfiniBand coupling links. More planning might be required to integrate the z15 in a Parallel Sysplex with z14 and z13/z13s servers.

Important: Use the previous **Z ARCHITECTURE** or **TUNE** options for C/C++ programs if the same applications run on the previous IBM Z servers. However, if C/C++ applications run on z15 servers only, use the latest **ARCHITECTURE** and **TUNE** options to ensure that the best performance possible is delivered through the latest instruction set additions.

For more information, see *Migration from z/OS V2R1 to z/OS V2R2*, GA32-0889.

7.5.5 z/OS V2.4

IBM z/OS, Version 2 Release 4, was announced on July 23, 2019 and was made generally available on September 30, 2019. This release delivers innovation through an agile, optimized, and resilient platform that helps companies build applications and services based on a highly scalable and secure infrastructure that provides the performance and availability for on-premise or provisioned as-a-service workloads.

z/OS V2.4 delivers the following capabilities (list is not exclusive):

- ▶ IBM z/OS Container Extensions (zCX), which enables the ability to run almost any Linux on IBM Z Docker container in z/OS alongside existing z/OS applications and data without a separate provisioned Linux server
- ▶ Easier integration of z/OS into private and multi-cloud environments with improvements that deliver a more robust, easy to use, and highly available implementation using IBM Cloud™ Provisioning and Management for z/OS, IBM z/OS Cloud Broker and IBM Cloud Storage Access for z/OS Data,
- ▶ Enhancements that continue to simplify and modernize the z/OS environment for a better user experience and improved productivity by reducing the level of IBM Z specific skills that are required to maintain z/OS,
- ▶ Ongoing industry-wide simplification improvements to help companies install and configure software using a common and modern method. These installation improvements range from the packaging of software through the configuration so that faster time to value can be realized throughout the enterprise,
- ▶ IBM Open Data Analytics for z/OS provides enhancements to simplify data analysis by combining open source runtimes and libraries with analysis of z/OS data at its source,
- ▶ Enhancements to security and data protection on the system with support for new industry cryptography and continued enhancements driving pervasive encryption through the ability to encrypt data without application changes. A new RACF capability improves management of access and privileges
- ▶ Leveraging z15 capabilities - System Recovery Boost which reduces the time that z/OS is offline when the operating system is offline for any reason. The use of IBM System Recovery Boost expedites planned operating system shutdown processing, operating system IPL (Initial Program Load), middleware/workload restart and recovery, and the client workload execution that follows. It will let businesses return their systems to work faster, not just from catastrophes, but after all kinds of disruptions, both planned and unplanned. Another aspect of System Recovery Boost is to expedite and streamline the execution of GDPS recovery scripts which perform reconfiguration actions during various planned and unplanned operational scenarios.
- ▶ Dynamic activation of I/O configurations for stand-alone Coupling Facilities Coupling Facilities (CFs) provide locking, caching, and list services between coupling-capable z/OS processors. They are a significant component of highly available Parallel Sysplex configurations. Stand-alone CFs (Coupling Facility images that reside on a server without a co-resident z/OS image), are now able to participate in dynamic I/O configuration changes that affect the stand-alone CF and no longer require the server to be restarted to activate such changes

7.5.6 z/OS V2.3

IBM announced z/OS Version 2 Release 3 - Engine for digital transformation through Announcement letter 217-246 on July 17, 2017. Focusing on three critical areas (Security, Simplification, and Cloud), z/OS V2.3 provides a simple and transparent approach to enable extensive encryption of data and to simplify the overall management of the z/OS system to increase productivity. Focus is also given to providing a simple approach for self-service provisioning and rapid delivery of software as a service, while enabling for the API economy.

Consider the following points before migrating z/OS 2.3 to IBM z15™:

- ▶ IBM z/OS V2.3 with z15 requires a minimum of 8 GB of memory. When running as a z/VM guest or on an IBM System z® Personal Development Tool, a minimum of 2 GB is required for z/OS V2.3. If the minimum is not met, a warning WTOR is issued at IPL.

Continuing with less than the minimum memory might affect availability. A migration health check will be introduced at z/OS V2.1 and z/OS V2.2 to warn if the system is configured with less than 8 GB.

- ▶ Dynamic splitting and merging of Coordinated Timing Network (CTN) is available with z15.
- ▶ The z/OS V2.3 real storage manager (RSM) is planned to support a new asynchronous memory clear operation to clear the data from 1M page frames by using I/O processors (SAPs). The new asynchronous memory clear operation eliminates the CPU cost for this operation and help improve performance of RSM first reference page fault processing and system services, such as IARV64 and STORAGE OBTAIN.
- ▶ RMF support is provided to collect SMC-D related performance measurements in SMF 73 Channel Path Activity and SMF 74 subtype 9 PCIE Activity records. It also provides these measurements in the RMF Postprocessor and Monitor III PCIE and Channel Activity reports. This support is also available on z/OS V2.2 with PTF UA80445 for APAR OA49113.
- ▶ HyperSwap support is enhanced to allow RESERVE processing. When a system runs a request to swap to secondary devices that are managed by HyperSwap, z/OS detects when RESERVEs are held and ensures that the devices that are swapped also hold the RESERVE. This enhancement is provided with collaboration from z/OS, GDPS HyperSwap, and CSM HyperSwap.

7.6 z/VM migration considerations

IBM z15 supports z/VM 7.2, z/VM 7.1, and z/VM 6.4. z/VM is moving to continuous delivery model. For more information, see [this web page](#).

7.6.1 z/VM 7.2

z/VM 7.2 has been announced April 14, 2020 with planned general availability September, 2020. It features the following (excerpt):

- ▶ Centralized Service Management for non-SSI environments to deploy service to multiple systems, regardless of geographic location, from a centralized primary location.
- ▶ Multiple Subchannel Set (MSS) Multi-Target Peer-To-Peer Remote Copy (MT-PPRC) z/VM support for the GDPS environment, allowing a device to be the primary to up to three secondary devices, each defined in a separate alternate subchannel set (supporting up to 3 alternate subchannel sets). Also provides the CP updates necessary for VM/HCD support of alternate subchannel sets.

- ▶ New Architecture Level Set of z13, z13s (LinuxONE Emperor / Rockhopper), or newer processor families
- ▶ z/VM 7.2 includes New Function APARs shipped for z/VM 7.1, such as:
VSwitch Priority Queuing, EAV Paging, 80 Logical Processors, Dynamic Crypto, System Recovery Boost support (subcapacity CPs speed boost only), and so on.

7.6.2 z/VM 7.1

z/VM 7.1 can be installed directly on IBM z15. z/VM V7R1 includes the following new features:

- ▶ Single System Image and Live Guest Relocation included in the base. In z/VM 6.4, this feature was the VMSSI-priced feature.
- ▶ Enhances the dump process to reduce the time that is required to create and process dumps.
- ▶ Upgrades to a new Architecture Level Set. This feature requires an IBM zEnterprise EC12 or BC12, or later.
- ▶ Provides the base for more functionality to be delivered as service Small Program Enhancements (SPEs) after general availability.

z/VM 7.1 includes SPEs shipped for z/VM 6.4, including Virtual Switch Enhanced Load Balancing, DS8000 z-Thin Provisioning, and Encrypted Paging.

7.6.3 z/VM V6.4

z/VM V6.4 can be installed directly on a z15 server with an image that is obtained from IBM after Sept. 23, 2019. The PTF for APAR VM65942 must be applied immediately after installing z/VM V6.4 and before configuring any part of the new z/VM system.

A z/VM Release Status Summary for supported z/VM versions is listed in Table 7-17.

Table 7-17 z/VM Release Status Summary

| z/VM Level ^a | General Availability | End of Marketing | End of Service | Minimum Processor Level | Maximum Processor Level |
|-------------------------|-----------------------------|------------------|----------------|-------------------------|-------------------------|
| 7.2 | September 2020 ^b | Not announced | Not announced | z13 and z13s | - |
| 7.1 | September 2018 | Not announced | Not announced | zEC12 and zBC12 | - |
| 6.4 | November 2016 | Not announced | Not announced | z196 and z114 | - |

a. Older z/VM versions (6.3, 6.2, 5.4 are End Of Support)

b. Planned GA, Announced April 14, 2020.

7.6.4 ESA/390-compatibility mode for guests

IBM z15™ no longer supports the full ESA/390 architectural mode. However, IBM z15™ does provide ESA/390-compatibility mode, which is an environment that supports a subset of DAT-off ESA/390 applications in a hybrid architectural mode.

z/VM provides the support that is necessary for DAT-off guests to run in this new compatibility mode. This support allows guests, such as CMS, GCS, and those guests that start in ESA/390 mode briefly before switching to z/Architecture mode to continue to run on IBM z15™.

The available PTF for APAR VM65976 provides infrastructure support for ESA/390 compatibility mode within z/VM V6.4. It must be installed on all members of an SSI cluster before any z/VM V6.4 member of the cluster is run on an IBM z15™ server.

In addition to operating system support, all the stand-alone utilities a client uses must be at a minimum level or need a PTF.

7.6.5 Capacity

For the capacity of any z/VM logical partition (LPAR) and any z/VM guest, in terms of the number of Integrated Facility for Linux (IFL) processors and central processors (CPs), real or virtual, you might want to adjust the number to accommodate the PU capacity of z15 servers.

7.7 z/VSE migration considerations

As described in “z/VSE” on page 255, IBM z15 supports z/VSE 6.2.

Consider the following general guidelines when you are migrating z/VSE environment to z15 servers:

- ▶ Collect reference information before migration

This information includes baseline data that reflects the status of, for example, performance data, CPU utilization of reference workload, I/O activity, and elapsed times.

This information is required to size z15 and is the only way to compare workload characteristics after migration.

For more information, see the *z/VSE Release and Hardware Upgrade* document.

- ▶ Apply required maintenance for z15

Review the Preventive Service Planning (PSP) bucket 8561DEVICE for z15 and apply the required PTFs for IBM and independent software vendor (ISV) products.

Note: IBM z15™ supports z/Architecture mode only.

7.8 Software licensing

The IBM z15™ software portfolio includes operating system software (that is, z/OS, z/VM, z/VSE, and z/TPF) and middleware that runs on these operating systems. The portfolio also includes middleware for Linux on Z environments.

For the z15, the following metric groups for software licensing are available from IBM, depending on the software product:

- ▶ Monthly license charge (MLC)

MLC pricing metrics feature a recurring charge that applies monthly. In addition to the permission to use the product, the charge includes access to IBM product support during

the support period. MLC pricing applies to z/OS, z/VSE, and z/TPF operating systems. Charges are based on processor capacity, which is measured in millions of service units (MSU) per hour.

▶ IPLA

IPLA metrics have a single, up-front charge for an entitlement to use the product. An optional and separate annual charge (called *subscription and support*) entitles clients to access IBM product support during the support period. With this option, you can also receive future releases and versions at no extra charge.

Software Licensing References

For more information about software licensing, see the following websites:

- ▶ [Learn about Software licensing](#)
- ▶ [Base license agreements](#)
- ▶ [IBM Z Software Pricing reference guide](#)
- ▶ [IBM Z Software Pricing](#)
- ▶ [The IBM International Passport Advantage® Agreement](#) can be downloaded from the [Learn about Software licensing website](#).

Subcapacity license charges

For eligible programs, subcapacity licensing allows software charges that are based on the measured utilization by logical partitions instead of the total number of MSUs of the CPC. Subcapacity licensing removes the dependency between the software charges and CPC (hardware) installed capacity.

The subcapacity licensed products are charged monthly based on the highest observed 4-hour rolling average utilization of the logical partitions in which the product runs. The exception is products that are licensed by using the Select Application License Charge (SALC) pricing metric. This type of charge requires measuring the utilization and reporting it to IBM.

The 4-hour rolling average utilization of the logical partition can be limited by a defined capacity value on the image profile of the partition. This value activates the soft capping function of the PR/SM, which limits the 4-hour rolling average partition utilization to the defined capacity value. Soft capping controls the maximum 4-hour rolling average usage (the last 4-hour average value at every 5-minute interval), but does not control the maximum instantaneous partition use.

You can also use an LPAR group capacity limit, which sets soft capping by PR/SM for a group of logical partitions that are running z/OS.

Even by using the soft capping option, the use of the partition can reach up to its maximum share based on the number of logical processors and weights in the image profile. Only the 4-hour rolling average utilization is tracked, which allows utilization peaks above the defined capacity value.

Some pricing metrics apply to stand-alone Z servers. Others apply to the aggregation of multiple Z server workloads within the same Parallel Sysplex.

For more information about WLC and how to combine logical partition utilization, see *z/OS Planning for Sub-Capacity Pricing*, SA23-2301.

Key MLC Metrics and Offerings

MLC metrics include various offerings. The following metrics and pricing schemes are available. Offerings often are tied to or made available to only on certain Z servers:

- ▶ Key MLC Metrics:
 - WLC (Workload License Charges)
 - AWLC (Advanced Workload License Charges)
 - CMLC (Country Multiplex License Charges)
 - VWLC (Variable Workload License Charges)
 - FWLC (Flat Workload License Charges)
 - AEWLC (Advanced Entry Workload License Charges)
 - EWLC (Entry Workload License Charges)
 - TWLC (Tiered Workload License Charges)
 - zNALC (System z New Application License Charges)
 - PSLC (Parallel Sysplex License Charges)
 - MWLC (Midrange Workload License Charges)
 - zELC (zSeries Entry License Charges)
 - GOLC (Growth Opportunity License Charges)
 - SALC (Select Application License Charges)
- ▶ Pricing:
 - GSSP (Getting Started Sub-Capacity Pricing)
 - IWP (Integrated Workload Pricing)
 - MWP (Mobile Workload Pricing)
 - zCAP (Z Collocated Application Pricing)
 - Parallel Sysplex Aggregated Pricing
 - CMP (Country Multiplex Pricing)
 - ULC (IBM S/390® Usage Pricing)

One of the recent changes in software licensing for z/OS and z/VSE is Multi-Version Measurement (MVM), which replaced Single Version Charging (SVC), Migration Pricing Option (MPO), and the IPLA Migration Grace Period.

MVM for z/OS and z/VSE removes time limits for running multiple eligible versions of a software program. Clients can run different versions of a program simultaneously for an unlimited duration during a program version upgrade.

Clients can also choose to run multiple different versions of a program simultaneously for an unlimited duration in a production environment. MVM allows clients to selectively deploy new software versions, which provides more flexible control over their program upgrade cycles. For more information, see *Software Announcement 217-093*, dated February 14, 2017.

Technology Transition Offerings with z15

Complementing the announcement of the z15 server, IBM introduced the following Technology Transition Offerings (TTOs):

- ▶ Technology Update Pricing for the IBM z15™.
- ▶ New and revised Transition Charges for Sysplexes or Multiplexes TTOs for actively coupled Parallel Sysplexes (z/OS), Loosely Coupled Complexes (z/TPF), and Multiplexes (z/OS and z/TPF).

Technology Update Pricing for the IBM z15™ extends the software price and performance that is provided by AWLC and CMLC for z15 servers. The new and revised Transition Charges for Sysplexes or Multiplexes offerings provide a transition to Technology Update Pricing for the IBM z15™ for customers who did not yet fully migrate to z15 servers. This

transition ensures that aggregation benefits are maintained and also phases in the benefits of Technology Update Pricing for the IBM z15™ pricing as customers migrate.

When a z15 server is in an actively coupled Parallel Sysplex or a Loosely Coupled Complex, you might choose aggregated Advanced Workload License Charges (AWLC) pricing or aggregated Parallel Sysplex License Charges (PSLC) pricing (subject to all applicable terms and conditions).

When a z15 server is part of a Multiplex under Country Multiplex Pricing (CMP) terms, Country Multiplex License Charges (CMLC), Multiplex zNALC (MzNALC), and Flat Workload License Charges (FWLC) are the only pricing metrics available (subject to all applicable terms and conditions).

When a z15 server is running z/VSE, you can choose Mid-Range Workload License Charges (MWLC), which are subject to all applicable terms and conditions.

For more information about AWLC, CMLC, MzNALC, PSLC, MWLC, or the Technology Update Pricing and Transition Charges for Sysplexes or Multiplexes TTO offerings, see the [IBM z Software Pricing page](#) of the IBM IT infrastructure website.

7.9 References

For more information about planning, see the home pages for the following operating systems:

- ▶ [z/OS](#)
- ▶ [z/VM](#)
- ▶ [z/VSE](#)
- ▶ [z/TPF](#)
- ▶ [Linux on Z](#)
- ▶ [KVM for IBM Z](#)



System upgrades

This chapter provides an overview of the IBM Z server upgrade process and how, in many cases, customers can manage capacity upgrades by using online tools and automation. The chapter also includes a detailed description of capacity on demand (CoD) offerings available on the z15.

IBM z15 servers support many dynamic provisioning features to give clients exceptional flexibility and control over system capacity and costs.

A key resource for managing client IBM Z servers is the [IBM Resource Link website](#). Once registered, a client can view product information by clicking **Resource Link** → **Client Initiated Upgrade Information**, and selecting **Education**. Select your particular product from the list of available systems.

The scalability of z15 servers includes the following benefits:

- ▶ Enabling new business opportunities
- ▶ Support for dynamic capacity growth and cloud environments
- ▶ Risk management of volatile, high-growth, and high-volume applications
- ▶ Enabling 24 x 7 application availability
- ▶ Enabling capacity growth during lockdown periods
- ▶ Enabling planned-downtime without availability impacts

This chapter includes the following topics:

- ▶ 8.1, “Permanent and Temporary Upgrades” on page 335
- ▶ 8.2, “Concurrent upgrades” on page 340
- ▶ 8.3, “Miscellaneous equipment specification upgrades” on page 347
- ▶ 8.4, “Permanent upgrade by using the CIU facility” on page 353
- ▶ 8.5, “On/Off Capacity on Demand” on page 357
- ▶ 8.6, “z/OS Capacity Provisioning” on page 365
- ▶ 8.7, “System Recovery Boost Upgrade” on page 369
- ▶ 8.8, “Capacity for Planned Event” on page 370
- ▶ 8.9, “Capacity Backup” on page 372
- ▶ 8.10, “Planning for nondisruptive upgrades” on page 376
- ▶ 8.11, “Summary of Capacity on-Demand offerings” on page 380

Note: Throughout this chapter, *z15* refers to IBM z15 Model T01 (Machine Type 8651), unless otherwise specified.

8.1 Permanent and Temporary Upgrades

The terminology for CoD and the types of upgrades for a z15 server are described in this section.

8.1.1 Overview

Upgrades can be categorized as described in this section.

Permanent versus temporary upgrades

Deciding whether to perform a Permanent or a Temporary upgrade depends on the situation. For example, a growing workload might require more memory, I/O cards, or processor capacity. However, to handle a peak workload, or to temporarily replace a system that is down during a disaster or data center maintenance, might require only a temporary upgrade. z15 servers offer the following solutions:

► Permanent upgrades

- Miscellaneous equipment specification (MES)

An MES upgrade might involve the addition of physical hardware or the installation of Licensed Internal Code Configuration Control (LICCC). In both cases, the installation is performed by IBM personnel.

- Customer Initiated Upgrade (CIU)

The use of the CIU facility for a system requires that the online CoD buying feature (FC 9900) is installed on the system and for the relevant CIU contract agreements to be in place. The CIU facility supports only LICCC upgrades.

For more information, see 8.1.4, “Permanent upgrades” on page 338.

Tip: An MES provides system upgrades that can result in more enabled processors, a different central processor (CP) capacity level, more processor drawers, memory, PCIe+ I/O drawers, and I/O features (physical upgrade). Extra planning tasks are required for nondisruptive logical upgrades. An MES is ordered through your IBM representative and installed by IBM service support representatives (IBM SSRs).

► Temporary

All temporary upgrades are LICCC-based. The one billable capacity offering is On/Off Capacity on Demand (On/Off CoD), which can be used for short-term capacity requirements and are pre-paid or post-paid.

The two replacement capacity offerings available are Capacity Backup (CBU) and Capacity for Planned Event (CPE).

System Recovery Boost zIIP capacity is a new pre-paid offering for z15 and is intended to provide temporary zIIP capacity to be used to speed up IPL, shutdown, and stand-alone dump events.

8.1.2 CoD for z15 systems-related terminology

The most frequently used terms that are related to CoD for z15 systems are listed in Table 8-1.

Table 8-1 CoD terminology

| Term | Description |
|-------------------------------------|--|
| Activated capacity | Capacity that is purchased and activated. Purchased capacity can be greater than the activated capacity. |
| Billable capacity | Capacity that helps handle workload peaks (expected or unexpected). The one billable offering that is available is On/Off Capacity on Demand (OOCoD). |
| Capacity | Hardware resources (processor and memory) that can process the workload can be added to the system through various capacity offerings. |
| Capacity Backup (CBU) | Capacity Backup allows you to place model capacity or specialty engines in a backup system. CBU is used in an unforeseen loss of system capacity because of an emergency or for Disaster Recovery testing. |
| Capacity for Planned Event (CPE) | Used when temporary replacement capacity is needed for a short-term event. CPE activates processor capacity temporarily to facilitate moving systems between data centers, upgrades, and other routine management tasks. CPE is an offering of CoD. |
| Capacity levels | Can be full capacity or subcapacity. For the z15 system, capacity levels for the CP engine are 7, 6, 5, and 4. |
| Capacity setting | Derived from the capacity level and the number of processors. For the z15 system, the capacity levels are 7nn, 6yy, 5yy, and 4xx, where xx, yy, or nn indicates the number of active CPs. The number of processors can have the following ranges: <ul style="list-style-type: none"> ▶ 0 - 34 for capacity levels 4xx. An all IFL or an all ICF system has a capacity level of 400. ▶ 1 - 34 for capacity levels 5yy and 6yy. ▶ 1 - 99 in decimal and A0 - J0, where A0 represents 100 and J0 represents 190, for capacity level 7nn. |
| Customer Initiated Upgrade (CIU) | A web-based facility where you can request processor and memory upgrades by using the IBM Resource Link and the system's Remote Support Facility (RSF) connection. |
| Capacity on Demand (CoD) | The ability of a system to increase or decrease its performance capacity as needed to meet fluctuations in demand. |
| Capacity Provisioning Manager (CPM) | As a component of z/OS Capacity Provisioning, CPM monitors business-critical workloads that are running z/OS on z15 systems. |
| Customer profile | This information is on Resource Link and contains client and system information. A customer profile can contain information about systems that are related to their IBM customer numbers. |
| Full capacity CP feature | For z15 servers, feature (CP7) provides full capacity. Capacity settings 7nn are full capacity settings with the ranges of 1 - 99 in decimal and A0 - J0, where A0 represents 100 and J0 represents 190, for capacity level 7nn. |
| High-water mark | Capacity that is purchased and owned by the client. |
| Installed record | The LICCC record is downloaded, staged to the Support Element (SE), and is installed on the central processor complex (CPC). A maximum of eight different records can be concurrently installed. |

| Term | Description |
|--|---|
| Model capacity identifier (MCI) | Shows the current active capacity on the system, including all replacement and billable capacity. For z15 servers, the model capacity identifier is in the form of 4xx, 5yy, 6yy, or 7nn, where xx, yy, or nn indicates the number of active CPs: <ul style="list-style-type: none"> ▶ xx can have a range of 00 - 34. An all IFL or an all ICF system has a capacity level of 400. ▶ yy can have a range of 01 - 34. ▶ 1 - 99 in decimal and A0 - J0, where A0 represents 100 and J0 represents 190, for capacity level 7nn. |
| Model Permanent Capacity Identifier (MPCI) | Keeps information about the capacity settings that are active before any temporary capacity is activated. |
| Model Temporary Capacity Identifier (MTCI) | Reflects the permanent capacity with billable capacity only, without replacement capacity. If no billable temporary capacity is active, MTCI equals the MPCI. |
| On/Off Capacity on Demand (CoD) | Represents a function that allows spare capacity in a CPC to be made available to increase the total capacity of a CPC. For example, On/Off CoD can be used to acquire more capacity for handling a workload peak. |
| Features on Demand (FoD) | FoD is a centralized way to flexibly entitle features and functions on the system. |
| Permanent capacity | The capacity that a client purchases and activates. This amount might be less capacity than the total capacity purchased. |
| Permanent upgrade | LICC that is licensed by IBM to enable the activation of applicable computing resources, such as processors or memory, for a specific CIU-eligible system on a permanent basis. |
| Purchased capacity | Capacity that is delivered to and owned by the client. It can be higher than the permanent capacity. |
| Permanent/Temporary entitlement record | The internal representation of a temporary (TER) or permanent (PER) capacity upgrade that is processed by the CIU facility. An <i>entitlement record</i> contains the encrypted representation of the upgrade configuration with the associated time limit conditions. |
| Replacement capacity | A temporary capacity that is used for situations in which processing capacity in other parts of the enterprise is lost. This loss can be a planned event or an unexpected disaster. The two replacement offerings available are Capacity for Planned Events and Capacity Backup. |
| Resource Link | The IBM Resource Link is a technical support website that provides a comprehensive set of tools and resources (log in required). |
| Secondary approval | An option that is selected by the client that requires second approver control for each CoD order. When a secondary approval is required, the request is sent for approval or cancellation to the Resource Link secondary user ID. |
| Staged record | The point when a record that represents a temporary or permanent capacity upgrade is retrieved and loaded on the SE disk. |
| Subcapacity | For z15 servers, CP features (CP4, CP5, and CP6) provide reduced capacity relative to the full capacity CP feature (CP7). |
| System Recovery Boost Record | Available on z15 servers, the optional System Recovery Boost Record is an orderable feature that provides more capacity for a limited time to enable speeding up shutdown, restart, and catchup processing for a limited event duration. |
| Temporary capacity | An optional capacity that is added to the current system capacity for a limited amount of time. It can be capacity that is owned or not owned by the client. |

| Term | Description |
|--------------------------|--|
| Vital product data (VPD) | Information that uniquely defines system, hardware, software, and microcode elements of a processing system. |

8.1.3 Concurrent and nondisruptive upgrades

Depending on the effect on the system and application availability, upgrades can be classified in the following manner:

▶ Concurrent

In general, *concurrency* addresses the continuity of operations of the *hardware* during an upgrade; for example, whether a system (hardware) must be turned off during the upgrade. For more information, see 8.2, “Concurrent upgrades” on page 340.

▶ Non-concurrent

This type of upgrade requires turning off the hardware that is being upgraded. Examples include memory upgrades to a z15 T01 max 34.

▶ Nondisruptive

Nondisruptive upgrades do not require the software or operating system to be restarted for the upgrade to take effect.

▶ Disruptive

An upgrade is considered *disruptive* when resources that are modified or added to an operating system image require that the operating system be restarted to configure the newly added resources.

A Concurrent upgrade might be disruptive to operating systems or programs that do not support the upgrades while being nondisruptive to others. For more information, see 8.10, “Planning for nondisruptive upgrades” on page 376.

8.1.4 Permanent upgrades

Permanent upgrades can be obtained by using the following processes:

- ▶ Ordered through an IBM marketing representative
- ▶ Initiated by the client with the CIU on the IBM Resource Link

Tip: The use of the CIU facility for a system requires that the online CoD buying feature (FC 9900) is installed on the system. The CIU facility is enabled through the permanent upgrade authorization feature code (FC 9898).

Permanent upgrades that are ordered through an IBM representative

Through a permanent upgrade, you can accomplish the following tasks:

- ▶ Add processor drawers
- ▶ Add Peripheral Component Interconnect Express (PCIe) drawers and features
- ▶ Add model capacity
- ▶ Add specialty engines
- ▶ Add memory
- ▶ Activate unassigned model capacity or IFLs
- ▶ Deactivate activated model capacity or IFLs
- ▶ Activate channels
- ▶ Activate cryptographic engines

- ▶ Change specialty engines (recharacterization)

Considerations: Most of the MESs can be concurrently applied without disrupting the workload. For more information, see 8.2, “Concurrent upgrades” on page 340. However, certain MES changes can be disruptive, such as adding PCIE IO drawers.

Memory upgrades that require dual inline memory module (DIMM) changes can be made nondisruptively if multiple CPC drawers are available and the flexible memory option is used.

Permanent upgrades by using CIU on the IBM Resource Link

Ordering the following permanent upgrades by using the CIU application through Resource Link allows you to add capacity to fit within your hardware:

- ▶ Add model capacity
- ▶ Add specialty engines
- ▶ Add memory
- ▶ Activate unassigned model capacity or IFLs
- ▶ Deactivate activated model capacity or IFLs

8.1.5 Temporary upgrades

z15 servers offer the following types of temporary upgrades:

- ▶ On/Off Capacity on Demand (On/Off CoD)

This offering allows you to temporarily add capacity or specialty engines to cover seasonal activities, period-end requirements, peaks in workload, or application testing. This temporary upgrade can be ordered by using the CIU application through Resource Link only.

Prepaid OoCoD tokens^a: Beginning with IBM z15, new prepaid OoCoD tokens that are purchased do not carry forward to future systems.

- a. Statements by IBM regarding its plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM.

- ▶ CBU

This offering allows you to replace model capacity or specialty engines in a backup system that is used in an unforeseen loss of system capacity because of a disaster.

- ▶ CPE

This offering allows you to replace model capacity or specialty engines because of a relocation of workload during system migrations or a data center move.

- ▶ System Recovery Boost Record

This offering allows you to add up to 20 zIIPs for use with the System Recovery Boost facility. System Recovery Boost provides temporary extra capacity for CP workloads to allow rapid shutdown, restart, and recovery of eligible systems. System Recovery Boost records are prepaid, licensed for a one-year period, and can be renewed at any time.

CBU, CPE, and System Recovery Boost Records can be ordered by using the CIU application through Resource Link or by contacting your IBM marketing representative.

Temporary upgrade capacity changes might be billable or a replacement.

Billable capacity

To handle a peak workload, you can activate up to double the purchased capacity of any processor unit (PU) type temporarily. You are charged daily.

This billable capacity offering is On/Off Capacity on Demand (On/Off CoD).

Replacement capacity

When processing capacity is lost in part of an enterprise, replacement capacity can be activated. It allows you to activate any PU type up to your authorized limit.

The following replacement capacity offerings are available:

- ▶ Capacity Backup
- ▶ Capacity for Planned Event

8.2 Concurrent upgrades

Concurrent upgrades on z15 servers can provide more capacity with no system outage. In most cases, a concurrent upgrade can be nondisruptive to the operating system with planning and operating system support.

This capability is based on the flexibility of the design and structure, which allows concurrent hardware installation and Licensed Internal Control Code (LICC) configuration changes.

The subcapacity models allow more configuration granularity within the family. The added granularity is available for models that are configured with up to 34 CPs, and provides 102 extra capacity settings. Subcapacity models provide for CP capacity increase in two dimensions that can be used together to deliver configuration granularity. The first dimension is adding CPs to the configuration. The second is changing the capacity setting of the CPs currently installed to a higher model capacity identifier.

z15 servers allow the concurrent and nondisruptive addition of processors to a running logical partition (LPAR). As a result, you can have a flexible infrastructure to which you can add capacity. This function is supported by z/OS, z/VM, and z/VSE. This addition is made by using one of the following methods:

- ▶ With planning ahead for the future need of extra processors. Reserved processors can be specified in the LPAR's profile. When the extra processors are installed, the number of active processors for that LPAR can be increased without the need for a partition reactivation and initial program load (IPL).
- ▶ Another (easier) way is to enable the dynamic addition of processors through the z/OS LOADxx member. Set the **DYNCPADD** parameter in member LOADxx to ENABLE.

Another function concerns the *system assist processor* (SAP). When more SAPs are concurrently added to the configuration, the SAP-to-channel affinity is dynamically remapped on all SAPs on the system to rebalance the I/O configuration.

8.2.1 PU Capacity feature upgrades

z15 servers feature machine type and model 8561-T01 capacity identifier.

The 8561-T01 is available in the following CPC drawer configurations:

- ▶ Feature Max34 (one CPC Drawer installed) can have a maximum of 34PUs for client characterization.

- ▶ Feature Max71 (two CPC Drawers) can have a maximum of 71 client PUs
- ▶ Feature Max108 (three CPC Drawers) can have a maximum of 108 client PUs
- ▶ Feature Max145 (four CPC Drawers) can have a maximum of 145 client PUs
- ▶ Feature Max190 (five CPC Drawers) can have a maximum of 190 client PUs
- ▶ Model capacity identifiers 4xx, 5yy, 6yy, or 7nn

The xx is a range of 00 - 34¹, yy is a range of 01 - 34, and nn is a range of 01 - 99, A0 - J0, where A0 represents the decimal number 100, which combines the character A with decimal 0 and where J0 represents the decimal number 190. It is obtained by continuing the hexadecimal counting to F that equals 15, G equals 16, H equals 17, I equals 18 and J equals 19 and adding the decimal digit 0 to make 190. A z15 server with 190 client usable processors is a z15 7J0. The model capacity identifier describes how many CPs are characterized (xx, yy, or nn) and the capacity setting (4, 5, 6, or 7) of the CPs.

A hardware configuration upgrade always requires more physical hardware (processor drawers, PCIe+ I/O drawers, or both). A system upgrade can change the system model or the MCI.

Consider the following points regarding model upgrades:

- ▶ LICCC upgrade:
 - Can add memory or Virtual Flash Memory (VFM) up to the amount that is physically installed
 - Can change the model capacity identifier, the capacity setting, or both
- ▶ Hardware installation upgrade:
 - Can change the CPC drawer feature by adding one or more drawers
 - Can change the model capacity identifier, the capacity setting, or both
 - Can add physical memory, PCIe+ I/O drawers, and other hardware features

The model capacity identifier can be concurrently changed. Concurrent upgrades can be performed for permanent and temporary upgrades.

Tip: A drawer feature upgrade can be performed concurrently only for a max 34 or a max 71 machine if feature codes 2271 or 2272 were ordered with the base machine.

Licensed Internal Code upgrades (MES ordered)

The LICCC provides for system upgrades without hardware changes by activating extra (physically installed) unused capacity. Concurrent upgrades through LICCC can be performed for the following resources:

- ▶ Processors, such as CPs, ICFs, z Integrated Information Processors (zIIPs), IFLs, and SAPs, if unused PUs are available on the installed processor drawers, or if the model capacity identifier for the CPs can be increased.
- ▶ Memory, when unused capacity is available on the installed memory cards. The Flexible memory option is available to give you better control over future memory upgrades. For more information, see 2.5.7, “Flexible Memory Option” on page 65.

Note: Plan ahead memory is not offered on new z15 orders. It can be carried forward only from z13 and z14 machines.

¹ The z15 zero CP MCI is 400. This setting applies to an all-IFL or all-ICF system.

Concurrent hardware installation upgrades (MES ordered)

Configuration upgrades can be concurrent when installing the following resources:

- ▶ Processor drawers (which contain processors, memory, and fanouts). Up to two processor drawers can be added concurrently on a z15 T01 max 34 if feature codes 2271 and 2272 were ordered with the initial configuration.
- ▶ PCIe+ Gen3 fanouts.
- ▶ I/O cards, when slots are still available on the installed PCIe+ I/O drawers.
- ▶ PCIe+ I/O drawers.

The concurrent I/O upgrade capability can be better used if a future target configuration is considered during the initial configuration.

Concurrent PU conversions (MES ordered)

z15 servers support concurrent conversion between all PU types, which includes SAPs, to provide flexibility and meet changing business requirements.

Important: The LICCC-based PU conversions require that at least one PU (CP, ICF, or IFL), remains unchanged. Otherwise, the conversion is disruptive. The PU conversion generates a LICCC that can be installed concurrently in two steps:

1. Remove the assigned PU from the configuration.
2. Activate the newly available PU as the new PU type.

LPARs also might have to free the PUs to be converted. The operating systems must include support to configure processors offline or online so that the PU conversion can be done nondisruptively.

Considerations: Client planning and operator action are required to use concurrent PU conversion. Consider the following points about PU conversion:

- ▶ It is disruptive if *all* current PUs are converted to different types.
- ▶ It might require individual LPAR outages if dedicated PUs are converted.

Unassigned CP capacity is recorded by a model capacity identifier. CP feature conversions change (increase or decrease) the model capacity identifier.

8.2.2 Customer Initiated Upgrade facility

The CIU facility is an IBM online system through which you can order, download, and install permanent and temporary upgrades for IBM Z servers. Access to and use of the CIU facility requires a contract between the client and IBM through which the terms and conditions for use of the CIU facility are accepted.

The CIU facility is controlled through the permanent upgrade authorization FC 9898. A prerequisite to FC 9898 is the online CoD buying feature code (FC 9900). Although FC 9898 can be installed on your z15 servers at any time, often it is added when ordering a z15 server.

After you place an order through the CIU facility, you receive a notice that the order is ready for download. You can then download and apply the upgrade by using functions that are available through the Hardware Management Console (HMC), along with the RSF. After all of the prerequisites are met, the entire process (from ordering to activation of the upgrade) is performed by the client and does not require any onsite presence of IBM SSRs.

CIU prerequisites

The CIU facility supports LICCC upgrades only. It does not support I/O upgrades. All other capacity that is required for an upgrade must be previously installed. Extra processor drawers or I/O cards cannot be installed as part of an order that is placed through the CIU facility. The sum of CPs, unassigned CPs, ICFs, zIIPs, IFLs, and unassigned IFLs cannot exceed the client PU count of the installed processor drawers. The total number of zIIPs can be twice the number of purchased CPs.

CIU registration and contract for CIU

To use the CIU facility, a client must be registered and the system must be set up. After you complete the CIU registration, access to the CIU application is available through the [IBM Resource Link website](#).

As part of the setup, provide one resource link ID for configuring and placing CIU orders and, if required, a second ID as an approver. The IDs are then set up for access to the CIU support. The CIU facility allows upgrades to be ordered and delivered much faster than through the regular MES process.

To order and activate the upgrade, log on to the [IBM Resource Link website](#) and start the CIU application to upgrade a system for processors or memory. You can request a client order approval to conform to your operational policies. You also can allow the definition of more IDs to be authorized to access the CIU. More IDs can be authorized to enter or approve CIU orders, or only view orders.

Permanent upgrades

Permanent upgrades can be ordered by using the CIU facility. Through the CIU facility, you can generate online permanent upgrade orders to concurrently add processors (CPs, ICFs, zIIPs, IFLs, and SAPs) and memory, or change the model capacity identifier. You can do so up to the limits of the installed processor drawers on a system.

Temporary upgrades

The base model z15 server describes permanent and dormant capacity by using the capacity marker and the number of PU features that are installed on the system. Up to eight temporary offerings can be present. Each offering includes its own policies and controls, and each can be activated or deactivated independently in any sequence and combination. Although multiple offerings can be active at any time, *only one On/Off CoD offering can be active at any time* if enough resources are available to fulfill the offering specifications.

Temporary upgrades are represented in the system by a *record*. All temporary upgrade records are on the SE hard disk drive (HDD). The records can be downloaded from the RSF or installed from portable media. At the time of activation, you can control everything locally.

The provisioning architecture is shown in Figure 8-1.

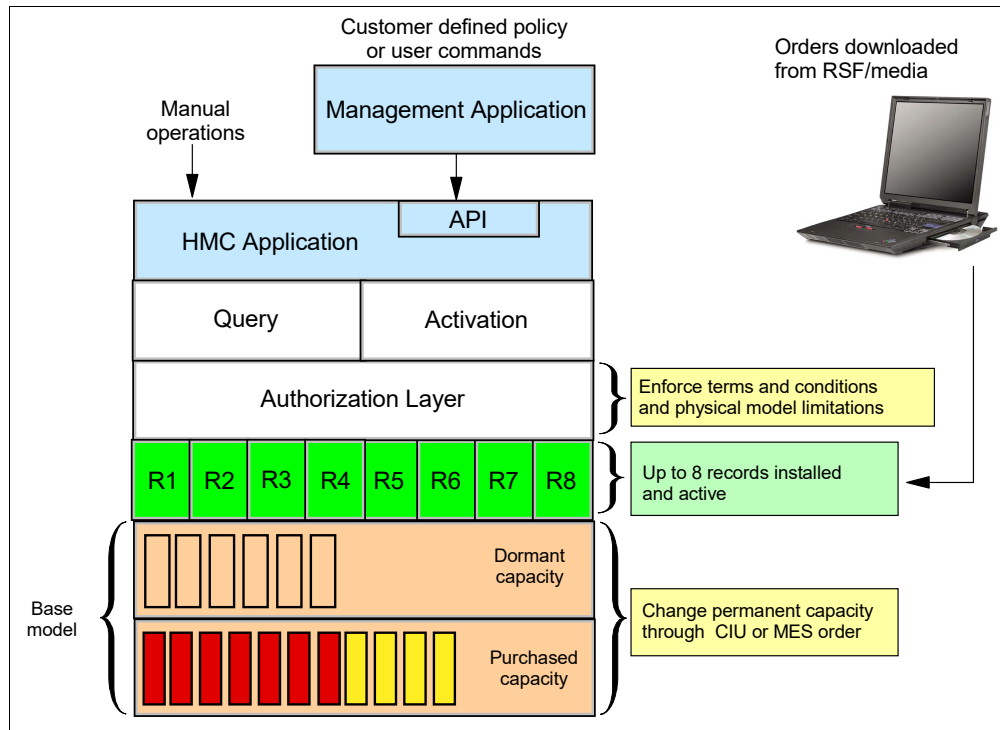


Figure 8-1 Provisioning architecture

The authorization layer enables administrative control over the temporary offerings. The activation and deactivation can be driven manually or under the control of an application through a documented application programming interface (API).

By using the API approach, you can customize at activation time the resources that are necessary to respond to the current situation up to the maximum that is specified in the order record. If the situation changes, you can add or remove resources without having to go back to the base configuration. This process eliminates the need for temporary upgrade specifications for all possible scenarios.

For a CPE record, only the ordered configuration can be activated.

This approach also enables you to update and replenish temporary upgrades, even in situations where the upgrades are active. Likewise, depending on the configuration, permanent upgrades can be performed while temporary upgrades are active. Examples of the activation sequence of multiple temporary upgrades are shown in Figure 8-2.

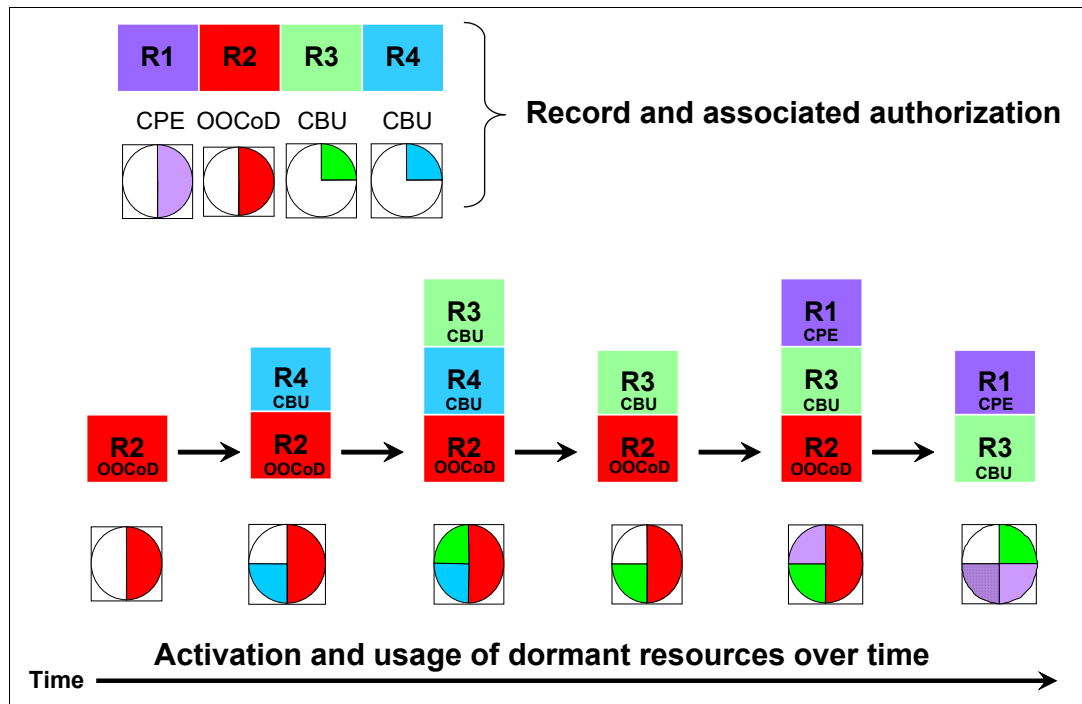


Figure 8-2 Example of temporary upgrade activation sequence

As shown in Figure 8-2, if R2, R3, and R1 are active at the same time, only parts of R1 can be activated because not enough resources are available to fulfill all of R1. When R2 is deactivated, the remaining parts of R1 can be activated as shown.

Temporary capacity can be billable as On/Off CoD, or replacement capacity as CBU, CPE, or System Recovery Boost. Consider the following points:

- ▶ On/Off CoD is a function that enables *concurrent* and *temporary* capacity growth of the system.

On/Off CoD can be used for client peak workload requirements, for any length of time, and includes a daily hardware and maintenance charge. The software charges can vary according to the license agreement for the individual products. For more information, contact your IBM Software Group representative.

On/Off CoD can concurrently add processors (CPs, ICFs, zIIPs, IFLs, and SAPs), increase the model capacity identifier, or both. It can do so up to the limit of the installed processor drawers of a system. It is restricted to twice the installed capacity. On/Off CoD requires a contractual agreement between you and IBM.

You decide whether to pre-pay or post-pay On/Off CoD. Capacity tokens that are inside the records are used to control activation time and resources.
- ▶ CBU is a concurrent and temporary activation of more CPs, ICFs, zIIPs, IFLs, and SAPs; or an increase of the model capacity identifier; or both.

Note: CBU cannot be used for peak workload management in any form.

On/Off CoD is the correct method to use for workload management. A CBU activation can last up to 90 days when a disaster or recovery situation occurs.

CBU features are optional, and require unused capacity to be available on installed processor drawers of the backup system. They can be available as unused PUs, an increase in the model capacity identifier, or both.

A CBU contract must be in place before the special code that enables this capability can be loaded on the system. The standard CBU contract provides for five 10-day tests (the *CBU test activation*) and one 90-day activation over a five-year period. For more information, contact your IBM representative.

You can run production workload on a CBU upgrade during a CBU test. At least an *equivalent amount* of production capacity must be shut down during the CBU test. If you signed CBU contracts, you also must sign an Amendment (US form #Z125-8145) with IBM to allow you to run production workload on a CBU upgrade during your CBU tests. More 10-day tests can be purchased with the CBU record.

- ▶ CPE is a concurrent and temporary activation of extra CPs, ICFs, zIIPs, IFLs, and SAPs; or an increase of the model capacity identifier; or both.

The CPE offering is used to replace temporary lost capacity within a client's enterprise for planned downtime events, such as data center changes.

Note: CPE cannot be used for peak load management of client workload or for a disaster situation.

The CPE feature requires unused capacity to be available on installed processor drawers of the backup system. The capacity must be available as unused PUs, as a possibility to increase the model capacity identifier on a subcapacity system, or as both.

A CPE contract must be in place before the special code that enables this capability can be loaded on the system. The standard CPE contract provides for one 3-day planned activation at a specific date. For more information, contact your IBM representative.

- ▶ The System Recovery Boost Record allows a concurrent activation of extra zIIPs.

The System Recovery Boost Record offering can be used to provide extra zIIP capacity that can be used by the System Recovery Boost facility. You might want to consider the use of this offering if your server is a full capacity model (7nn) and can benefit from more CPs during system shutdown and restart. The capacity is delivered as zIIPs that can perform CP work during the boost periods for an LPAR.

A System Recovery Boost Record contract must be in place before the special code that enables this capability can be loaded on the system. The standard contract provides for one 6-hour activation for the specific purpose of System Recovery Boost only. For more information, contact your IBM representative.

Activation of System Recovery Boost Record does not change the MCI of your system.

8.2.3 Concurrent upgrade functions summary

The possible concurrent upgrades combinations are listed in Table 8-2.

Table 8-2 Concurrent upgrade summary

| Type | Name | Upgrade | Process |
|-----------|------------------------------|--|--------------------------------------|
| Permanent | MES | CPs, ICFs, zIIPs, IFLs, SAPs, processor drawer, memory, and I/Os | Installed by IBM SSRs |
| | Online permanent upgrade | CPs, ICFs, zIIPs, IFLs, SAPs, and memory | Performed through the CIU facility |
| Temporary | On/Off CoD | CPs, ICFs, zIIPs, IFLs, and SAPs | Performed through the OOCOD facility |
| | CBU | CPs, ICFs, zIIPs, IFLs, and SAPs | Activated through model conversion |
| | CPE | CPs, ICFs, zIIPs, IFLs, and SAPs | Activated through model conversion |
| | System Recovery Boost Record | zIIPs | Activated through model conversion |

8.3 Miscellaneous equipment specification upgrades

MES upgrades enable concurrent and permanent capacity growth. MES upgrades allow the concurrent adding of processors (CPs, ICFs, zIIPs, IFLs, and SAPs), memory capacity, and I/O ports. For subcapacity models, MES upgrades allow the concurrent adjustment of both the number of processors and the capacity level. The MES upgrade can be performed by using LICCC only, installing more processor drawers, adding PCIe+ I/O drawers, adding I/O² features, or using the following combinations:

- ▶ MES upgrades for processors are done by any of the following methods:
 - LICCC assigning and activating unassigned PUs up to the limit of the installed processor drawers.
 - LICCC to adjust the number and types of PUs to change the capacity setting, or both.
 - Installing more processor drawers and LICCC assigning and activating unassigned PUs on the installed processor drawers.
- ▶ MES upgrades for memory are done by one of the following methods:
 - By using LICCC to activate more memory capacity up to the limit of the memory cards on the currently installed processor drawers. Flexible memory features enable you to implement better control over future memory upgrades. For more information about the memory features, see 2.5.7, “Flexible Memory Option” on page 65.
 - Installing more processor drawers and the use of LICCC to activate more memory capacity on installed processor drawers.
 - By using the CPC Enhanced Drawer Availability (EDA), where possible, on multi-drawer systems to add or change the memory cards.
- ▶ MES upgrades for I/O are done by installing more I/O features and supporting infrastructure (if required) on PCIe drawers that are installed, or installing more PCIe drawers to hold the new cards.

² Other adapter types, such as zHyperlink, Coupling Express LR, and Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE), also can be added to the PCIe+ I/O drawers through an MES.

An MES upgrade requires IBM SSRs for the installation. In most cases, the time that is required for installing the LICCC and completing the upgrade is short.

To better use the MES upgrade function, carefully plan the initial configuration to allow a concurrent upgrade to a target configuration. The availability of PCIe+ I/O drawers improves the flexibility to perform unplanned I/O configuration changes concurrently.

The Store System Information (STSI) instruction gives more useful and detailed information about the base configuration and temporary upgrades.

The model and model capacity identifiers that are returned by the STSI instruction are updated to coincide with the upgrade. For more information, see “Store System Information instruction” on page 378.

Upgrades: An MES provides the physical upgrade, which results in more enabled processors, different capacity settings for the CPs, and more memory, I/O ports, I/O adapters, and I/O drawers. Extra planning tasks are required for nondisruptive logical upgrades. For more information, see “Guidelines to avoid disruptive upgrades” on page 380.

8.3.1 MES upgrade for processors

An MES upgrade for processors can concurrently add CPs, ICFs, zIIPs, IFLs, and SAPs to a z15 server by assigning available PUs on the processor drawers through LICCC. Depending on the quantity of the extra processors in the upgrade, more processor drawers might be required, and can be concurrently installed before the LICCC is enabled if plan-ahead features are available. With the subcapacity models, more capacity can be provided by adding CPs, changing the capacity identifier on the current CPs, or both.

Limits: The sum of CPs, inactive CPs, ICFs, zIIPs, IFLs, unassigned IFLs, and SAPs cannot exceed the maximum limit of PUs available for client use. The number of zIIPs cannot exceed twice the number of purchased CPs.

An example of an MES upgrade for processors (with two upgrade steps) is shown in Figure 8-3.

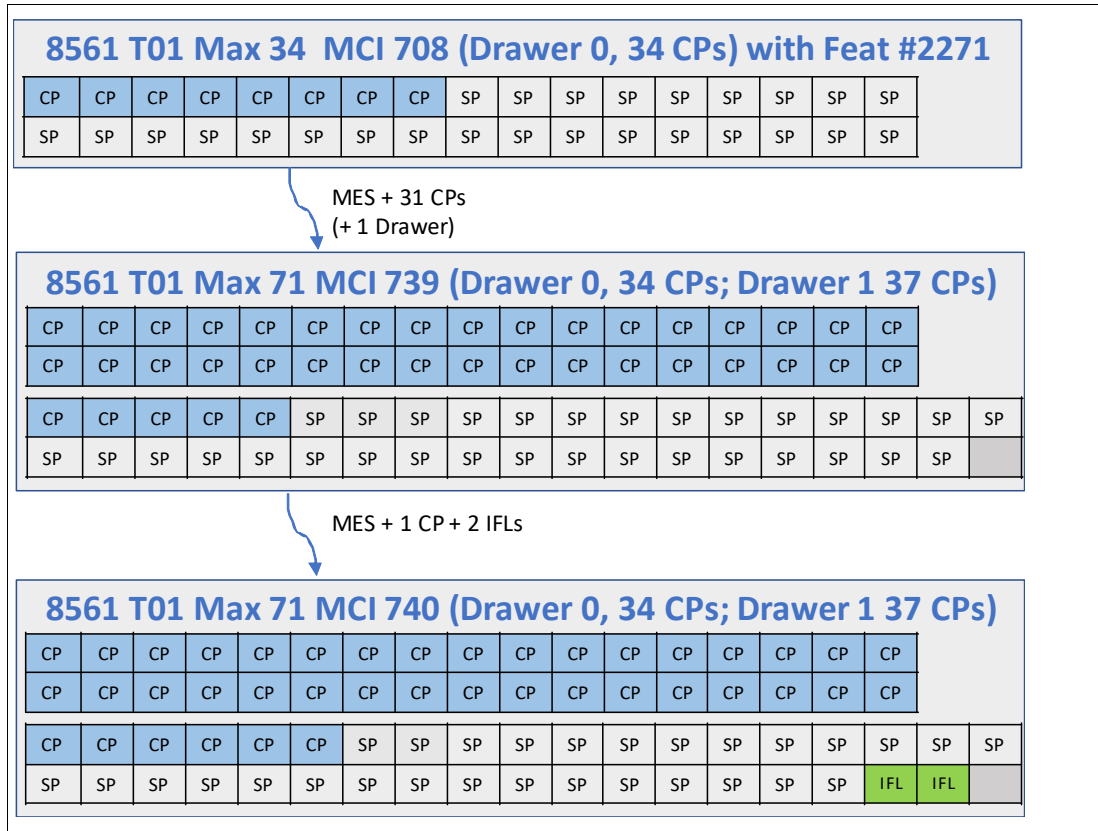


Figure 8-3 MES for processor example

A model T01 max 34 (one processor drawer), model capacity identifier 708 (eight CPs), is concurrently upgraded to a model T01 max 71 (two processor drawers), with MCI 739 (39 CPs). The model upgrade requires adding a processor drawer and assigning and activating 39 PUs as CPs. Then, model max 71, MCI 739, is concurrently upgraded to a capacity identifier 740 (40 CPs) with two IFLs. This process is done by assigning and activating three more unassigned PUs (one as CP and two as IFLs). If needed, more LPARs can be created concurrently to use the newly added processors.

The example that is shown in Figure 8-3 was used to show how the addition of PUs as CPs and IFLs and the addition of a processor drawer works. The addition of a processor drawer to a z15 Max 34 upgrades the machine to Max 71.

After the second CPC drawer addition, CPC drawer 0 has 34 configurable PUs and CPC drawer 1 has 37 configurable PUs, which allows 71 PUs to be characterized on the new Max 71 configuration.

Consideration: Up to 190 logical processors (including reserved processors) can be defined to an LPAR. However, do not define more processors to an LPAR than the target operating system supports.

The number of processors that are supported by various z/OS and z/VM releases are listed in Table 8-3.

Table 8-3 Number of processors that are supported by the operating system

| Operating system | Number of processors that are supported |
|-------------------------|--|
| z/OS V2R2 | 190 PUs per z/OS LPAR in non-SMT mode and 128 PUs per z/OS LPAR in SMT mode. For both, the PU total is the sum of CPs and zIIPs. |
| z/OS V2R3 | 190 PUs per z/OS LPAR in non-SMT mode and 128 PUs per z/OS LPAR in SMT mode. For both, the PU total is the sum of CPs and zIIPs. |
| z/OS V2R4 | 190 PUs per z/OS LPAR in non-SMT mode and 128 PUs per z/OS LPAR in SMT mode. For both, the PU total is the sum of CPs and zIIPs. |
| z/VM V7R1 | 64 (or 32 in SMT mode). |
| z/VM V6R4 | 64 (or 32 in SMT mode). |
| z/VSE | z/VSE Turbo Dispatcher can use up to 4 CPs, and tolerates up to 10-way LPARs. |
| z/TPF | 86 CPs. |
| Linux on IBM Z -190 CPs | Linux ^a supports 256 cores without SMT and 128 cores with SMT (256 threads). |

a. Supported Linux on Z distributions (for more information, see Chapter 7, “Operating system support” on page 253).

Software charges, which are based on the total capacity of the system on which the software is installed, are adjusted to the new capacity after the MES upgrade.

Software products that use Workload License Charges (WLC) or Taylor Fit Pricing (TFP) might not be affected by the system upgrade. Their charges are based on partition usage, not on the system total capacity. For more information about WLC, see 7.8, “Software licensing” on page 328.

8.3.2 MES upgrades for memory

MES upgrades for memory can concurrently add more memory in the following ways:

- ▶ Through LICCC, which enables more capacity up to the limit of the currently installed DIMM memory cards
- ▶ Concurrently installing more CPC drawers and LICCC-enabling memory capacity on the new CPC drawers.

The Flexible Memory Feature is available to allow better control over future memory upgrades. For more information about flexible memory features, see 2.5.7, “Flexible Memory Option” on page 65.

If the z15 server is a multiple processor drawer configuration, you can use the EDA feature to remove a processor drawer and add DIMM memory cards. It can also be used to upgrade the installed memory cards to a larger capacity size. You can then use LICCC to enable the extra memory.

With proper planning, memory can be added nondisruptively to z/OS partitions and z/VM partitions. If necessary, new LPARs can be created nondisruptively to use the newly added memory.

Concurrency: Upgrades that require DIMM changes can be concurrent by using the EDA feature. Planning is required to see whether this option is a viable for your configuration. The use of the flexible memory option ensures that EDA can work with the least disruption.

The one-processor drawer feature Max34 requires a minimum of 768 GB addressable memory. The client addressable storage in this case is 512 GB. If you require more memory, an extra memory upgrade can install up to 8 TB of memory. It does so by changing the DIMM sizes and adding DIMMs in all available slots in the processor drawer.

You can also add memory by *concurrently* adding a second processor drawer with sufficient memory into the configuration and then using LICCC to enable that memory. Changing DIMMs in a single CPC drawer system is disruptive.

An LPAR can dynamically take advantage of a memory upgrade if reserved storage is defined to that LPAR. The reserved storage is defined to the LPAR as part of the image profile.

Reserved memory can be configured online to the LPAR by using the LPAR dynamic storage reconfiguration (DSR) function. DSR allows a z/OS operating system image and z/VM partitions to add reserved storage to their configuration if any unused storage exists.

The nondisruptive addition of storage to a z/OS and z/VM partition requires the correct operating system parameters to be set. If reserved storage is not defined to the LPAR, the LPAR must be deactivated, the image profile changed, and the LPAR reactivated. This process allows the extra storage resources to be available to the operating system image.

8.3.3 MES upgrades for I/O

MES upgrades for I/O can concurrently add more I/O features by using one of the following methods:

- ▶ Installing more I/O features on an installed PCIe+ I/O drawer.
- ▶ Adding a PCIe+ I/O drawer to hold the new I/O features.

For more information about PCIe+ I/O drawers, see 4.2, “I/O system overview” on page 156.

The number of PCIe+ I/O drawers that can be present in a z15 server depends on how many CPC drawers are present and on whether the configuration includes the Bulk Power Assembly (BPA) offering. It also depends on whether the CPC drawer reserve features are present.

The number of drawers for specific configuration options are listed in Table 8-4 on page 352. It is based on no CPC drawer reserve options being configured.

Note: The maximum number of IO drawers in the table is reduced by 1 for each CPC drawer reserve feature present.

Table 8-4 PCIe+ I/O drawers summary

| Description | Frame A only | Frames Z,A or A,B | Frames Z,A,B | Frames Z,A,B,C |
|-------------|--------------|-------------------|--------------|----------------|
| PDU Max 34 | 0 - 3 | 4 - 8 | 9 - 12 | 13 - 16 |
| BPA Max 34 | 0 - 1 | 2 - 6 | 7 - 9 | 7 - 11 |
| PDU Max 71 | 0 - 2 | 3 - 7 | 8 - 11 | 12 - 15 |
| BPA Max 71 | 0 | 1 - 5 | 6 - 8 | 6 - 11 |
| PDU Max 108 | 0 - 1 | 2 - 6 | 7 - 10 | 11 - 15 |
| BPA Max 108 | 0 | 0 - 3 | 4 - 8 | 9 - 12 |
| PDU Max 145 | N/A | 0 - 4 | 5 - 9 | 10 - 12 |
| BPA Max 145 | N/A | 0 - 1 | 2 - 6 | 7 - 11 |
| PDU Max 190 | N/A | 0 - 3 | 4 - 8 | 9 - 12 |
| BPA Max 190 | N/A | 0 - 1 | 2 - 6 | 7 - 11 |

Depending on the number of I/O features, the configurator determines the number of PCIe+ I/O drawers required.

To better use the MES for I/O capability, carefully plan the initial configuration to allow concurrent upgrades up to the target configuration.

If a PCIe+ I/O drawer is added to a z15 server and original features must be physically moved to another PCIe+ I/O drawer, original card moves are disruptive.

z/VSE, z/TPF, and Linux on Z do *not* provide dynamic I/O configuration support. Although installing the new hardware is done concurrently, defining the new hardware to these operating systems requires an IPL.

Tip: z15 servers feature a hardware system area (HSA) of 256 GB. z14 servers have a 192 GB HSA. HSA is *not* part of the client-purchased memory.

8.3.4 Feature on Demand

Only one FoD LICCC record is installed or staged at any time in the system and its contents can be viewed in the Manage window, as shown in Figure 8-4 on page 353.

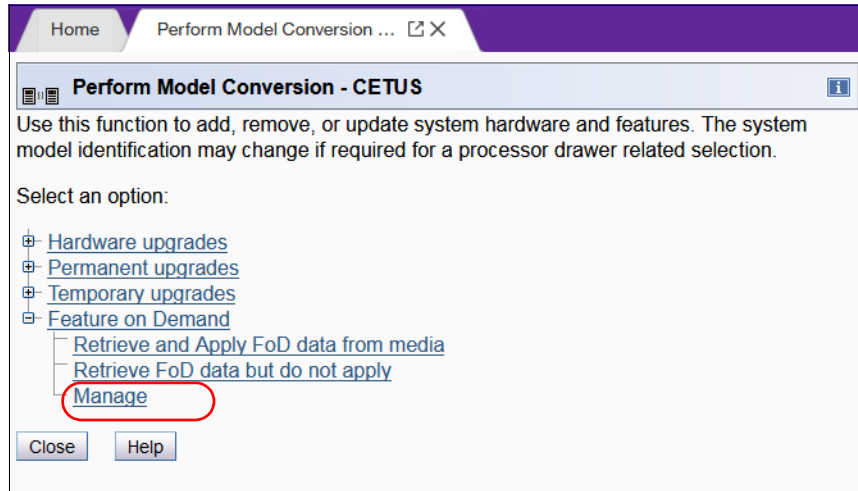


Figure 8-4 Features on-Demand

A staged record can be removed without installing it. An FoD record can be installed only completely; no selective feature or partial record installation is available. The features that are installed are merged with the CPC LICCC after activation.

An FoD record can be installed only once. If it is removed, a new FoD record is needed to reinstall. A remove action cannot be undone.

8.3.5 Summary of plan-ahead feature

The flexible memory plan-ahead feature is available for z15 servers. No feature code is associated with flexible memory. The purpose of flexible memory is to enable enhanced processor drawer availability. If a processor drawer must be serviced, the flexible memory is activated to accommodate storing the CPC drawer that is taken offline. After the repair action, the memory is taken offline again and is made unavailable for use.

Tip: Accurate planning and the definition of the target configuration allows you to maximize the value of these plan-ahead features.

8.4 Permanent upgrade by using the CIU facility

By using the CIU facility (through [the IBM Resource Link](#)), you can start a permanent upgrade for CPs, ICFs, zIIPs, IFLs, SAPs, or memory. When performed through the CIU facility, you add the resources without IBM personnel present at your location. You can also unassign previously purchased CPs and IFL processors through the CIU facility.

Adding permanent upgrades to a system through the CIU facility requires that the permanent upgrade enablement feature (FC 9898) is installed on the system. A permanent upgrade might change the system model capacity identifier (4xx, 5yy, 6yy, or 7nn) if more CPs are requested, or if the capacity identifier is changed as part of the permanent upgrade. If necessary, more LPARs can be created concurrently to use the newly added processors.

Consideration: A permanent upgrade of processors can provide a concurrent upgrade, which results in more enabled processors that are available to a system configuration. More planning and tasks are required for *nondisruptive* logical upgrades. For more information, see “Guidelines to avoid disruptive upgrades” on page 380.

Maintenance charges are automatically adjusted as a result of a permanent upgrade.

Software charges that are based on the total capacity of the system on which the software is installed are adjusted to the new capacity after the permanent upgrade is installed. Software products that use WLC or customers with TFP might not be affected by the system upgrade because their charges are based on LPAR usage rather than system total capacity. For more information about WLC, see 7.8, “Software licensing” on page 328.

The CIU facility process on IBM Resource Link is shown in Figure 8-5.

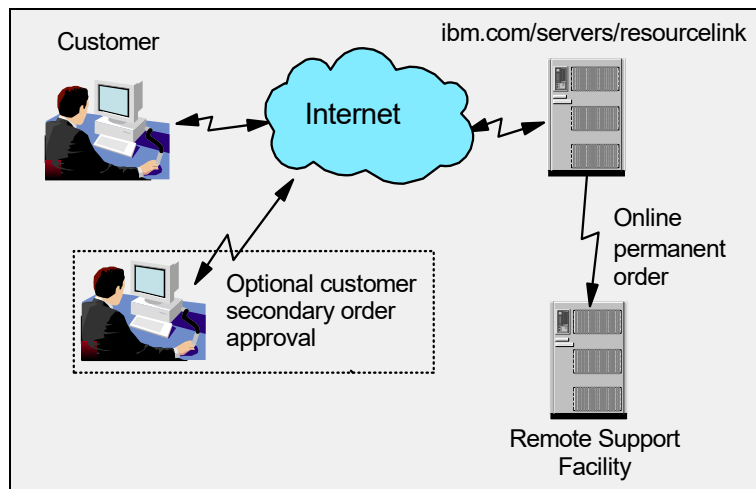


Figure 8-5 Permanent upgrade order example

The following sample sequence shows how to start an order on the IBM Resource Link:

1. Sign on to Resource Link.
2. Select **Customer Initiated Upgrade** from the main Resource Link page. Client and system information that is associated with the user ID are displayed.
3. Select the system to receive the upgrade. The current configuration (PU allocation and memory) is shown for the selected system.
4. Select **Order Permanent Upgrade**. The Resource Link limits the options to those options that are valid or possible for the selected configuration (system).
5. After the target configuration is verified by the system, accept or cancel the order. An order is created and verified against the pre-established agreement.
6. Accept or reject the price that is quoted. A secondary order approval is optional. Upon confirmation, the order is processed. The LICCC for the upgrade is available within hours.

The order activation process for a permanent upgrade is shown in Figure 8-6. When the LICCC is passed to the Remote Support Facility, you are notified through an email that the upgrade is ready to be downloaded.

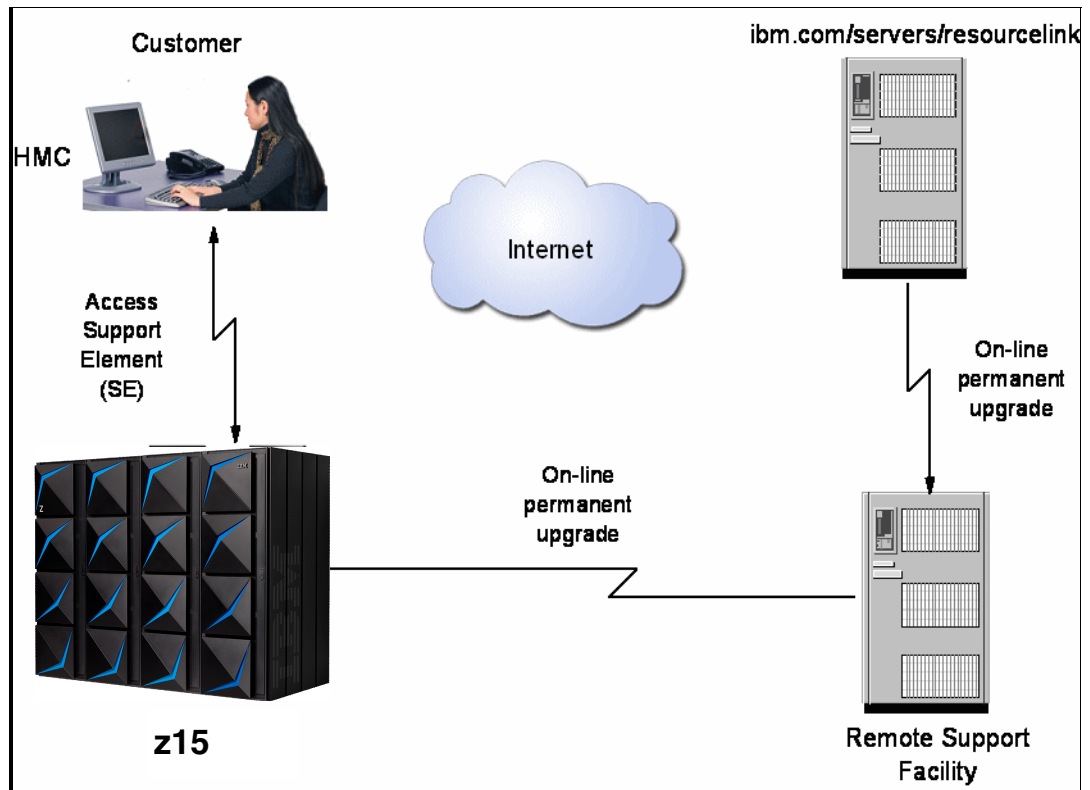


Figure 8-6 CIU-eligible order activation example

8.4.1 Ordering

IBM Resource Link provides the interface that enables you to order a concurrent upgrade for a system. You can create, cancel, or view the order, and view the history of orders that were placed through this interface.

Configuration rules enforce that only valid configurations are generated within the limits of the individual system. Warning messages are issued if you select invalid upgrade options. The process allows only one permanent CIU-eligible order for each system to be placed at a time.

For more information, see the [IBM Resource Link website](#) (log in required).

The initial view of the Machine profile on Resource Link is shown in Figure 8-7.

IBM Systems > z Systems > Resource Link > Customer Initiated Upgrade >

Machine profile

2964 - 8DA87 - 4604724

| Current configuration | |
|--|--------------|
| Model Capacity: | 735 (35 CPs) |
| ICF: | 8 |
| zIIP: | 12 |
| IFL: | 8 |
| SAP: | 12 |
| Memory: | 1952 |
| Unassigned IFLs: | 0 |
| Management enablement level: 2. Automate | |
| Current configuration as of 24 Mar 2015 09:38:00 | |

Machine summary

Type, model, serial:
2964 - N63 - 8DA87

System name:
SCZP501

Customer summary

Company name:
IBM CORP

Customer number:
4604724

GEO, country:
Americas - zDutchy of Merwyn

Ordering options

- [→ Order permanent upgrade](#)
- [→ Order On/Off CoD record](#)
- [→ Order On/Off CoD test record](#)
- [→ Order On/Off CoD record with prepaid upgrades](#)
- [→ Order On/Off CoD record with spending limits](#)
- [→ Order administrative On/Off CoD test record](#)
- [→ Order Capacity Backup \(CBU\) record](#)
- [→ Order Capacity for Planned Events \(CPE\) record](#)
- [→ Display upgrade matrix](#)

To update profile

- [→ Upload VPD](#)
- [→ Upload upgrade billing XML data](#)
- [→ Disable machine profile...](#)

For more information

- [→ View machine's On/Off CoD order billing history](#)

About ordering

Authorization to create orders
User ID: haimo@us.ibm.com

Name: Robert Haimowitz

Authorization to approve orders
Not required

Notes:

Ordering options

CIU Permanent: Enabled
On/Off CoD: Enabled
Auto Renewal: Enabled
CBU: Enabled
CPE: Enabled

Figure 8-7 Machine profile window

The number of CPs, ICFs, zIIPs, IFLs, SAPs, memory size, and unassigned IFLs on the current configuration are displayed on the left side of the page.

Resource Link retrieves and stores relevant data that is associated with the processor configuration, such as the number of CPs and installed memory cards. It allows you to select only those upgrade options that are deemed valid by the order process. It also allows upgrades only within the bounds of the currently installed hardware.

8.4.2 Retrieval and activation

After an order is placed and processed, the appropriate upgrade record is passed to the IBM support system for download.

When the order is available for download, you receive an email that contains an activation number. You can then retrieve the order by using the Perform Model Conversion task from the SE, or through the Single Object Operation to the SE from an HMC.

In the Perform Model Conversion window, select **Permanent upgrades** to start the process, as shown in Figure 8-8.

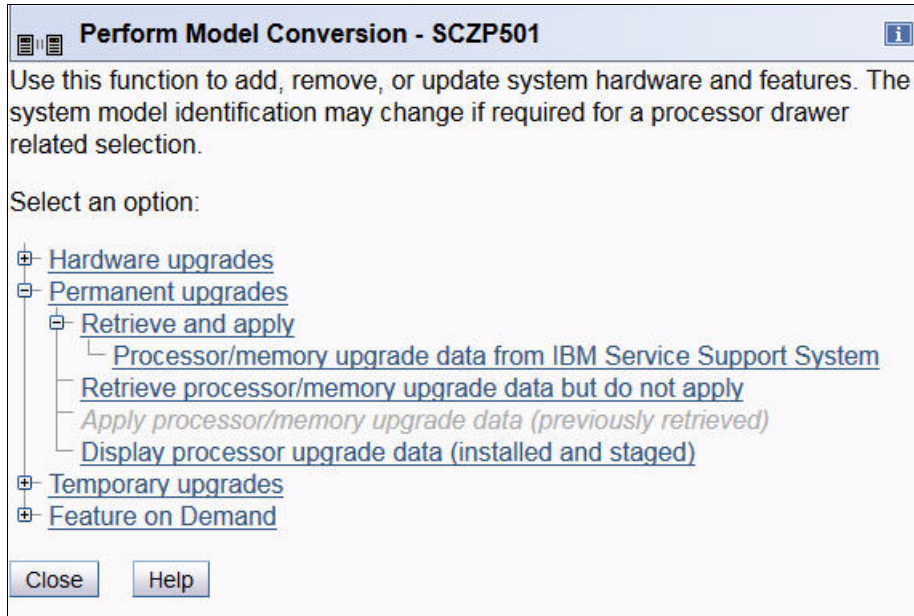


Figure 8-8 z15 Perform Model Conversion window

The window provides several possible options. If you select the **Retrieve and apply** data option, you are prompted to enter the order activation number to start the permanent upgrade, as shown in Figure 8-9.

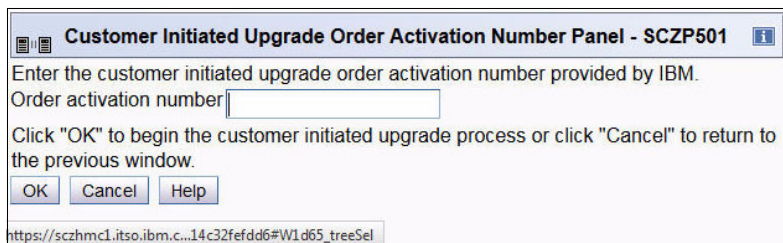


Figure 8-9 Customer Initiated Upgrade Order Activation Number window

8.5 On/Off Capacity on Demand

On/Off CoD allows you to temporarily enable PUs and unassigned IFLs that are available within the current hardware model. You can also use it to change capacity settings for CPs to help meet your peak workload requirements.

8.5.1 Overview

The capacity for CPs is expressed in millions of service units (MSUs). Capacity for speciality engines is expressed in number of speciality engines. *Capacity tokens* are used to limit the resource consumption for all types of processor capacity.

Capacity tokens are introduced to provide better control over resource consumption when On/Off CoD offerings are activated. Tokens represent the following resource consumptions:

- ▶ For CP capacity, each token represents the amount of CP capacity that results in one MSU of software cost for one day (*an MSU-day token*).
- ▶ For speciality engines, each token is equivalent to one speciality engine capacity for one day (*an engine-day token*).

Each speciality engine type features its own tokens, and each On/Off CoD record includes separate token pools for each capacity type. During the ordering sessions on Resource Link, select how many tokens of each type to create for an offering record. Each engine type must include tokens for that engine type to be activated. Capacity that has no tokens cannot be activated.

When resources from an On/Off CoD offering record that contains capacity tokens are activated, a *billing window* is started. A billing window is always 24 hours. Billing occurs at the end of each billing window.

The resources that are billed are the highest resource usage inside each billing window for each capacity type. An activation period is one or more complete billing windows. The activation period is the time from the first activation of resources in a record until the end of the billing window in which the last resource in a record is deactivated.

At the end of each billing window, the tokens are decremented by the highest usage of each resource during the billing window. If any resource in a record does not have enough tokens to cover usage for the next billing window, the entire record is deactivated.

Note: On/Off CoD requires that the Online CoD Buying feature (FC 9900) is installed on the system that you want to upgrade.

The On/Off CoD to Permanent Upgrade Option gives customers a window of opportunity to assess capacity additions to your permanent configurations by using On/Off CoD. If a purchase is made, the hardware On/Off CoD charges during this window (three days or less) are waived. If no purchase is made, you are charged for the temporary use.

The resources eligible for temporary use are CPs, ICFs, zIIPs, IFLs, and SAPs. The temporary addition of memory and I/O ports or adapters is not supported.

Unassigned PUs that are on the installed processor drawers can be temporarily and concurrently activated as CPs, ICFs, zIIPs, IFLs, and SAPs through LICCC. You can assign PUs up to twice the currently installed CP capacity, and up to twice the number of ICFs, zIIPs, or IFLs. An On/Off CoD upgrade cannot change the system capacity feature. The addition of new processor drawers is not supported. However, the activation of an On/Off CoD upgrade can increase the model capacity identifier (4xx, 5yy, 6yy, or 7nn).

8.5.2 Capacity Provisioning Manager

The installation of the capacity provision function on z/OS requires the following prerequisites:

- ▶ Setting up and customizing z/OS RMF, including the Distributed Data Server (DDS).
- ▶ Setting up the z/OS CIM Server (included in z/OS base).
- ▶ Performing capacity provisioning customization. For more information, see *z/OS MVS Capacity Provisioning User's Guide*, SC34-2661.

Using the capacity provisioning function requires the following prerequisites:

- ▶ TCP/IP connectivity to observed systems.
- ▶ RMF Distributed Data Server must be active.
- ▶ CIM server must be active.
- ▶ Security and CIM customization.
- ▶ Capacity Provisioning Manager customization.

The Capacity Provisioning Manager Console is provided as part of z/OS MF, which provides a browser interface for managing z/OS systems.

Customizing the capacity provisioning function is required on the following systems:

- ▶ Observed z/OS systems

These systems are in one or multiple sysplexes that are to be monitored. For more information about the capacity provisioning domain, see 8.10, “Planning for nondisruptive upgrades” on page 376.

- ▶ Runtime systems

These systems are the systems where the Capacity Provisioning Manager is running, or to which the server can fail over after server or system failures.

8.5.3 Ordering

Concurrently installing temporary capacity by ordering On/Off CoD is possible in the following manner:

- ▶ CP features equal to the MSU capacity of installed CPs
- ▶ IFL features up to the number of installed IFLs
- ▶ ICF features up to the number of installed ICFs
- ▶ zIIP features up to the number of installed zIIPs
- ▶ SAPs - up to 8

On/Off CoD can provide CP temporary capacity in two ways:

- ▶ By increasing the number of CPs.
- ▶ For subcapacity models, capacity can be added by increasing the number of CPs, changing the capacity setting of the CPs, or both. The capacity setting for all CPs must be the same. If the On/Off CoD is adding CP resources that have a capacity setting different from the installed CPs, the base capacity settings are changed to match.

On/Off CoD includes the following limits that are associated with its use:

- The number of CPs cannot be reduced.
- The target configuration capacity is limited to these amounts:
 - Twice the currently installed capacity, expressed in MSUs for CPs.
 - Twice the number of installed IFLs, ICFs, and zIIPs. Up to 8 SAPs can be activated. For more information, see 8.2.1, “PU Capacity feature upgrades” on page 340.

On/Off CoD can be ordered as prepaid or postpaid. A prepaid On/Off CoD offering record contains resource descriptions, MSUs, speciality engines, and tokens that describe the total capacity that can be used. For CP capacity, the token contains MSU-days. For speciality engines, the token contains speciality engine-days.

When resources on a prepaid offering are activated, they must have enough capacity tokens to allow the activation for an entire billing window, which is 24 hours. The resources remain active until you deactivate them or until one resource uses all of its capacity tokens. Then, all activated resources from the record are deactivated.

A postpaid On/Off CoD offering record contains resource descriptions, MSUs, speciality engines, and can contain capacity tokens that denote MSU-days and speciality engine-days.

When resources in a postpaid offering record *without* capacity tokens are activated, those resources remain active until they are deactivated, or until the offering record expires. The record often expires 180 days after its installation.

When resources in a postpaid offering record *with* capacity tokens are activated, those resources must include enough capacity tokens to allow the activation for an entire billing window (24 hours). The resources remain active until they are deactivated, until all of the resource tokens are used, or until the record expires. The record usually expires 180 days after its installation. If one capacity token type is used, resources from the entire record are deactivated.

For example, for a z15 server with capacity identifier 502 (two CPs), a capacity upgrade through On/Off CoD can be delivered in the following ways:

- ▶ Add CPs of the same capacity setting. With this option, the model capacity identifier can be changed to a 503, which adds another CP to make it a three-way CP. It can also be changed to a 504, which adds two CPs, making it a four-way CP.
- ▶ Change to a different capacity level of the current CPs and change the model capacity identifier to a 602 or 702. The capacity level of the CPs is increased, but no other CPs are added. The 502 also can be temporarily upgraded to a 603, which increases the capacity level and adds another processor. The capacity setting 434 does not have an upgrade path through On/Off CoD because you cannot reduce the number of CPs and a 534 is more than twice the capacity.

Use the Large System Performance Reference (LSPR) information to evaluate the capacity requirements according to your workload type. For more information about LSPR data for current IBM processors, see the [Large Systems Performance Reference for IBM Z page](#) of the IBM Systems website.

The On/Off CoD hardware capacity is charged on a 24-hour basis. A grace period is granted at the end of the On/Off CoD day. This grace period allows up to an hour after the 24-hour billing period to change the On/Off CoD configuration for the next 24-hour billing period or deactivate the current On/Off CoD configuration. The times when the capacity is activated and deactivated are maintained in the z15 server and sent back to the support systems.

If On/Off capacity is active, On/Off capacity can be added without having to return the system to its original capacity. If the capacity is increased multiple times within a 24-hour period, the charges apply to the highest amount of capacity active in that period.

If more capacity is added from an active record that contains capacity tokens, the systems checks whether the resource has enough capacity to be active for an entire billing window (24 hours). If that criteria is not met, no extra resources are activated from the record.

If necessary, more LPARs can be activated concurrently to use the newly added processor resources.

Consideration: On/Off CoD provides a concurrent hardware upgrade that results in more capacity being made available to a system configuration. Extra planning tasks are required for nondisruptive upgrades. For more information, see “Guidelines to avoid disruptive upgrades” on page 380.

To participate in this offering, you must accept contractual terms for purchasing capacity through the Resource Link, establish a profile, and install an On/Off CoD enablement feature on the system. Later, you can concurrently install temporary capacity up to the limits in On/Off CoD and use it for up to 180 days.

Monitoring occurs through the system call-home facility. An invoice is generated if the capacity is enabled during the calendar month. You are billed for the use of temporary capacity until the system is returned to the original configuration. Remove the enablement code if the On/Off CoD support is no longer needed.

On/Off CoD orders can be pre-staged in Resource Link to allow multiple optional configurations. The pricing of the orders is done at the time that you order them, and the pricing can vary from quarter to quarter. Staged orders can have different pricing.

When the order is downloaded and activated, the daily costs are based on the pricing at the time of the order. The staged orders do not have to be installed in the order sequence. If a staged order is installed out of sequence and later a higher-priced order is staged, the daily cost is based on the lower price.

Another possibility is to store multiple On/Off CoD LICCC records on the SE with the same or different capacities, which gives you greater flexibility to enable quickly needed temporary capacity. Each record is easily identified with descriptive names, and you can select from a list of records that can be activated.

Resource Link provides the interface to order a dynamic upgrade for a specific system. You can create, cancel, and view the order. Configuration rules are enforced, and only valid configurations are generated based on the configuration of the individual system. After you complete the prerequisites, orders for the On/Off CoD can be placed. The order process uses the CIU facility on Resource Link.

Memory and channels are not supported on On/Off CoD.

An individual record can be activated only once. Subsequent sessions require a new order to be generated, which produces a new LICCC record for that specific order. Alternatively, you can use an *auto-renewal* feature to eliminate the need for a manual replenishment of the On/Off CoD order. This feature is implemented in Resource Link, and you must also select this feature in the machine profile, as shown in Figure 8-10.

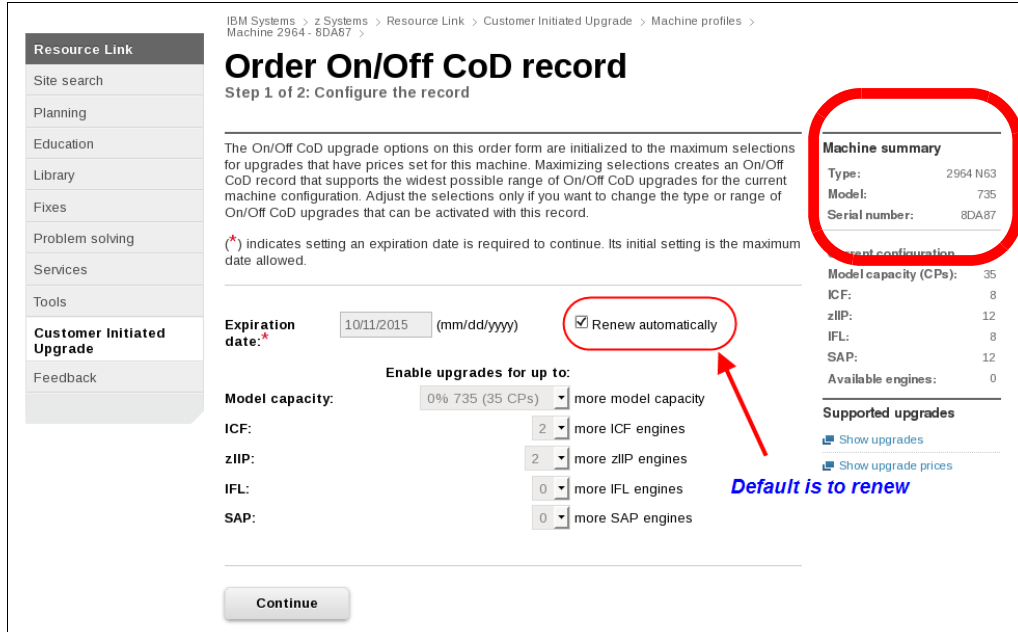


Figure 8-10 Order On/Off CoD record window

8.5.4 On/Off CoD testing

Each On/Off CoD-enabled system is entitled to one no-charge 24-hour test. No IBM charges are assessed for the test, including charges that are associated with temporary hardware capacity, IBM software, and IBM maintenance. The test can be used to validate the processes to download, stage, install, activate, and deactivate On/Off CoD capacity.

This test can have a maximum duration of 24 hours, which commences upon the activation of any capacity resource that is contained in the On/Off CoD record. Activation levels of capacity can change during the 24-hour test period. The On/Off CoD test automatically stops at the end of the 24-hour period.

You also can perform administrative testing. No capacity is added to the system, but you can test all the procedures and automation for the management of the On/Off CoD facility.

An example of an On/Off CoD order on the Resource Link web page is shown in Figure 8-11.

IBM Systems > z Systems > Resource Link > Customer Initiated Upgrade > Machine profiles > Machine 2964 - 8DA87 >

Order On/Off CoD record

Step 2 of 2: Review and submit your order

Review the range of upgrades you selected on the previous page. The On/Off CoD record you are about to order will be configured to support activating any configurations within the range.

(*) indicates accepting the [Terms and Conditions of this order](#) is required to submit it. Mark the check box to indicate acceptance.

Machine summary

Type: 2964 N63
 Model: 735
 Serial number: 8DA87

| | Enable upgrades up to | Daily hardware prices | Daily maintenance prices (estimated) ¹ |
|------------------------|------------------------|-----------------------|---|
| Model capacity: | 0% more model capacity | | |
| ICF: | 2 more ICF engines | \$0.00 | ≈12.00 |
| zIIP: | 2 more zIIP engines | \$0.00 | ≈12.00 |
| IFL: | 0 more IFL engines | | |
| SAP: | 0 more SAP engines | | |

Description: +0% model capacity, +2 ICF, +2 zIIP, +0 IFL, +0 SAP

Notes:

- Reflects current established prices for the selected machine. Prices are subject to change; the actual prices in effect at the time of use will apply.
- Daily prices for ICF, zIIP, IFL, and SAP upgrades are **per engine**.
- The IFL upgrade daily hardware price includes per IFL for the management enablement level in effect for this machine.

Figure 8-11 On/Off CoD order example

The example order that is shown in Figure 8-11 is an On/Off CoD order for 0% more CP capacity (system is at capacity level 7), and for two more ICFs and two more zIIPs. The maximum number of CPs, ICFs, zIIPs, and IFLs is limited by the current number of available unused PUs of the installed processor drawers. The maximum number of SAPs is determined by the model number and the number of available PUs on the already installed processor drawers.

To finalize the order, you must accept Terms and Conditions for the order, as shown in Figure 8-12.

Terms of Order

You have requested an On/Off Capacity on Demand, or Temporary Capacity upgrade. Your enterprise has previously accepted the Temporary Capacity terms, restated here. In the event there is a conflict between the terms shown on this website and the terms specified in your contract with IBM, the terms of such contract prevail:

1) upon download and installation of this Temporary Capacity Upgrade, IBM grants you only a temporary license to use the LIC enabling such Temporary Capacity Upgrade. You may use such Temporary Capacity Upgrade only on the TC Eligible Machine for which such LIC is provided, and only to the extent of the authorization identified via the CIU Facility.

I accept the Terms and Conditions of this order*

Submit

Figure 8-12 CIU order Terms and Conditions

8.5.5 Activation and deactivation

When a previously ordered On/Off CoD is retrieved from Resource Link, it is downloaded and stored on the SE HDD. You can activate the order manually or through automation when the capacity is needed.

If the On/Off CoD offering record does not contain resource tokens, you must deactivate the temporary capacity manually. Deactivation is done from the SE and is nondisruptive. Depending on how the capacity was added to the LPARs, you might be required to perform tasks at the LPAR level to remove it. For example, you might have to configure offline any CPs that were added to the partition, deactivate LPARs that were created to use the temporary capacity, or both.

On/Off CoD orders can be staged in Resource Link so that multiple orders are available. An order can be downloaded and activated only once. If a different On/Off CoD order is required or a permanent upgrade is needed, it can be downloaded and activated without having to restore the system to its original purchased capacity.

In support of automation, an API is if allows the activation of the On/Off CoD records. The activation is performed from the HMC, and requires specifying the order number. With this API, automation code can be used to send an activation command along with the order number to the HMC to enable the order.

8.5.6 Termination

A client is contractually obligated to end the On/Off CoD right-to-use feature when a transfer in asset ownership occurs. A client also can choose to end the On/Off CoD right-to-use feature without transferring ownership.

Removing FC 9898 ends the right to use the On/Off CoD. This feature cannot be ordered if a temporary session is active. Similarly, the CIU enablement feature cannot be removed if a temporary session is active. When the CIU enablement feature is removed, the On/Off CoD right-to-use feature is simultaneously removed. Reactivating the right-to-use feature subjects the client to the terms and fees that apply then.

Upgrade capability during On/Off CoD

Upgrades that involve physical hardware are supported while an On/Off CoD upgrade is active on a particular z15 server. LICCC-only upgrades can be ordered and retrieved from Resource Link, and can be applied while an On/Off CoD upgrade is active. LICCC-only memory upgrades can be retrieved and applied while an On/Off CoD upgrade is active.

Repair capability during On/Off CoD

If the z15 server requires service while an On/Off CoD upgrade is active, the repair can take place without affecting the temporary capacity.

Monitoring

When you activate an On/Off CoD upgrade, an indicator is set in vital product data. This indicator is part of the call-home data transmission, which is sent on a scheduled basis. A time stamp is placed into the call-home data when the facility is deactivated. At the end of each calendar month, the data is used to generate an invoice for the On/Off CoD that was used during that month.

Maintenance

The maintenance price is adjusted as a result of an On/Off CoD activation.

Software

Software Parallel Sysplex license charge (PSLC) clients are billed at the MSU level that is represented by the combined permanent and temporary capacity. All PSLC products are billed at the peak MSUs that are enabled during the month, regardless of usage. Clients with WLC licenses are billed by product at the highest four-hour rolling average for the month. In this instance, temporary capacity does not increase the software bill until that capacity is allocated to LPARs and used.

Results from the STSI instruction reflect the current permanent and temporary CPs. For more information, see “Store System Information instruction” on page 378.

8.6 z/OS Capacity Provisioning

This section describes how z/OS Capacity Provisioning can help you manage the addition of capacity to a server to handle workload peaks.

z/OS Capacity Provisioning is delivered as part of the z/OS MVS™ Base Control Program (BCP).

Capacity Provisioning includes the following components:

- ▶ Capacity Provisioning Manager (Provisioning Manager)
- ▶ Capacity Provisioning Management Console, available in the IBM z/OS Management Facility
- ▶ Sample data sets and files

The Provisioning Manager monitors the workload on a set of z/OS systems and organizes the provisioning of extra capacity to these systems when required. You define the systems to be observed in a domain configuration file.

The details of extra capacity and the rules for its provisioning are stored in a policy file. These two files are created and maintained through the Capacity Provisioning Management Console (CPMC). The operational flow of Capacity Provisioning is shown in Figure 8-13 on page 366.

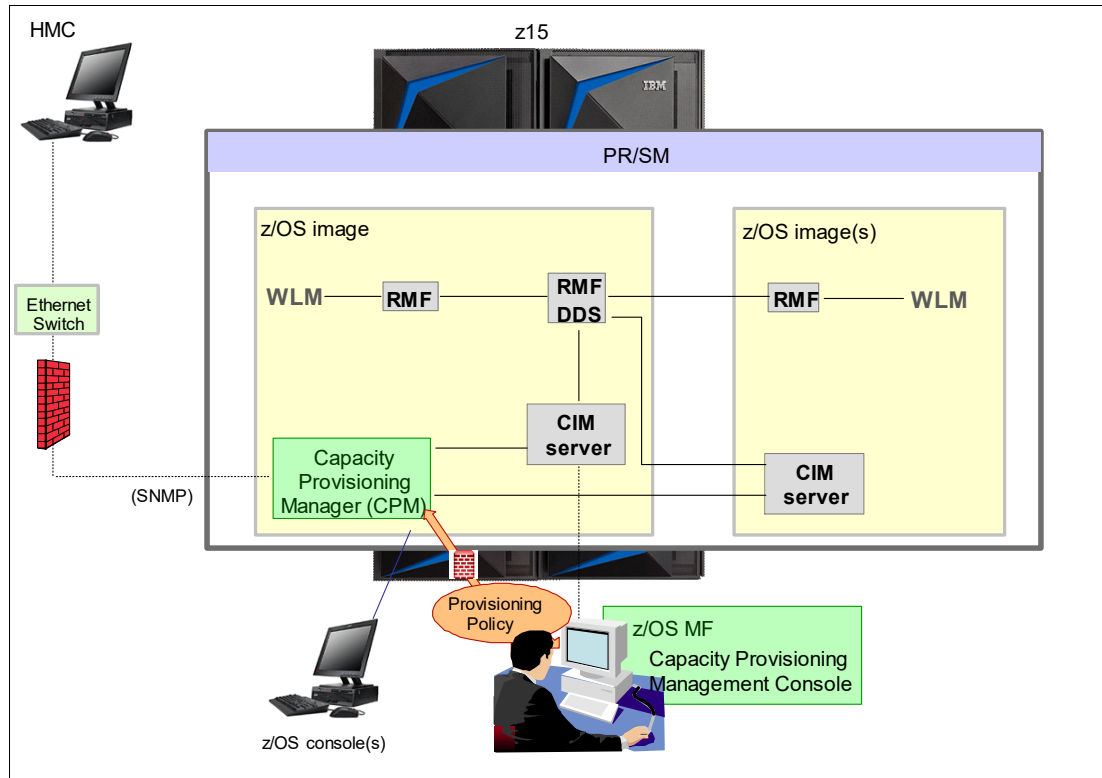


Figure 8-13 The capacity provisioning process and infrastructure

The z/OS WLM manages the workload by goals and business importance on each z/OS system. WLM metrics are available through existing interfaces, and are reported through IBM Resource Measurement Facility (RMF) Monitor III, with one RMF gatherer for each z/OS system.

Sysplex-wide data aggregation and propagation occur in the RMF Distributed Data Server (DDS). The RMF Common Information Model (CIM) providers and associated CIM models publish the RMF Monitor III data.

CPM retrieves critical metrics from one or more z/OS systems' CIM structures and protocols. CPM communicates to local and remote SEs and HMCs by using the Simple Network Management Protocol (SNMP).

CPM can see the resources in the individual offering records and the capacity tokens. When CPM activates resources, a check is run to determine whether enough capacity tokens remain for the specified resource to be activated for at least 24 hours. If insufficient tokens remain, no resource from the On/Off CoD record is activated.

If a capacity token is used during an activation that is driven by the CPM, the corresponding On/Off CoD record is deactivated prematurely by the system. This process occurs even if the CPM activates this record, or parts of it. However, you do receive warning messages if capacity tokens are close to being fully used.

You receive the messages five days before a capacity token is fully used. The five days are based on the assumption that the consumption is constant for the five days. You must put operational procedures in place to handle these situations. You can deactivate the record manually, allow it occur automatically, or replenish the specified capacity token by using the Resource Link application.

The Capacity Provisioning Management Console (CPMC) is a console that administrators use to work with provisioning policies and domain configurations and to monitor the status of a Provisioning Manager. The management console is implemented by the Capacity Provisioning task in the IBM z/OS Management Facility (z/OSMF). z/OSMF provides a framework for managing various aspects of a z/OS system through a web browser interface

Capacity Provisioning Domain

The provisioning infrastructure is managed by the CPM through the Capacity Provisioning Domain (CPD), which is controlled by the Capacity Provisioning Policy (CPP). The CPD is shown in Figure 8-14.

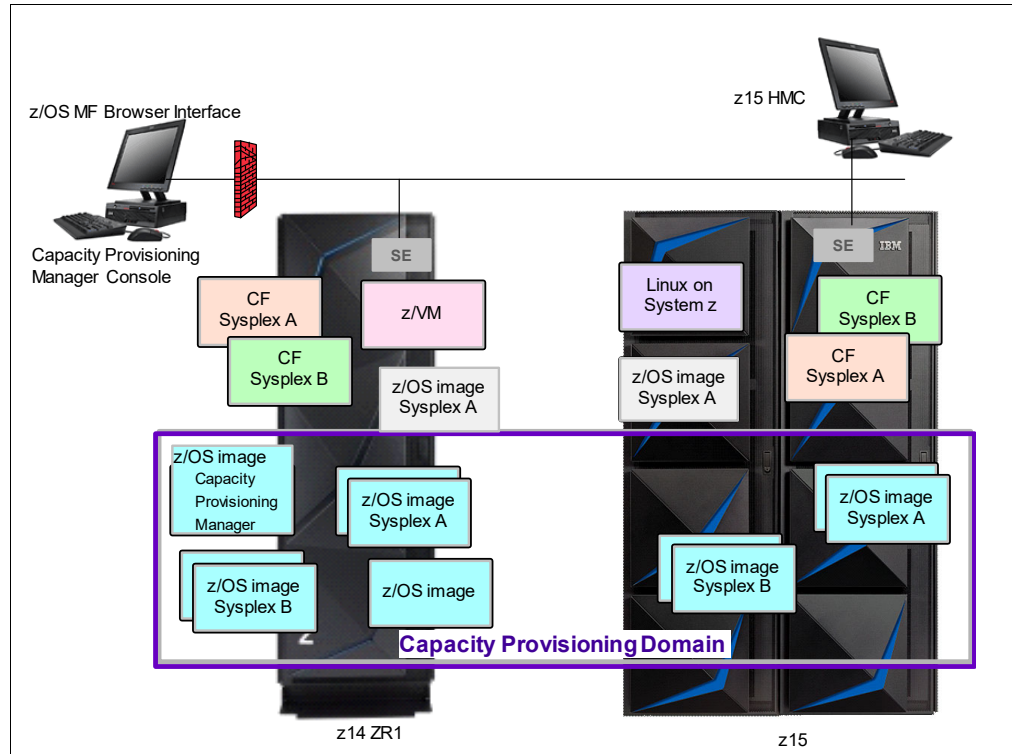


Figure 8-14 Capacity Provisioning Domain

The CPD configuration defines the CPCs and z/OS systems that are controlled by an instance of the CPM. One or more CPCs, sysplexes, and z/OS systems can be defined into a domain. Although sysplexes and CPCs do not have to be contained in a domain, they must not belong to more than one domain.

Each domain has one active capacity provisioning policy.

CPM operates in the following modes, which allows four different levels of automation:

- ▶ Manual mode

Use this command-driven mode when no CPM policy is active.

- ▶ Analysis mode

In analysis mode, CPM processes capacity-provisioning policies and informs the operator when a provisioning or deprovisioning action is required according to policy criteria.

Also, the operator determines whether to ignore the information or to manually upgrade or downgrade the system by using the HMC, SE, or available CPM commands.

► Confirmation mode

In this mode, CPM processes capacity provisioning policies and interrogates the installed temporary offering records. Every action that is proposed by the CPM must be confirmed by the operator.

► Autonomic mode

This mode is similar to the confirmation mode, but no operator confirmation is required.

Several reports are available in all modes that contain information about the workload, provisioning status, and the rationale for provisioning guidelines. User interfaces are provided through the z/OS console and the CPMC application.

The provisioning policy defines the circumstances under which more capacity can be provisioned (when, which, and how). The criteria features the following elements:

- A time condition is when provisioning is allowed:
 - Start time indicates when provisioning can begin.
 - Deadline indicates that provisioning of more capacity is no longer allowed.
 - End time indicates that deactivation of capacity must begin.
- A workload condition is which work qualifies for provisioning. It can have the following parameters:
 - The z/OS systems that can run eligible work.
 - The importance filter indicates eligible service class periods, which are identified by WLM importance.
 - Performance Index (PI) criteria:
 - Activation threshold: PI of service class periods must exceed the activation threshold for a specified duration before the work is considered to be suffering.
 - Deactivation threshold: PI of service class periods must fall below the deactivation threshold for a specified duration before the work is considered to no longer be suffering.
 - Included service classes are eligible service class periods.
 - Excluded service classes are service class periods that must not be considered.

Tip: If no workload condition is specified, the full capacity that is described in the policy is activated and deactivated at the start and end times that are specified in the policy.

- Provisioning scope is how much more capacity can be activated and is expressed in MSUs.

The number of zIIPs must be one specification per CPC that is part of the CPD and are specified in MSUs.

The maximum provisioning scope is the maximum extra capacity that can be activated for all the rules in the CPD.

In the specified time interval, the provisioning rule is that up to the defined extra capacity can be activated if the specified workload is behind its objective.

The rules and conditions are named and stored in the Capacity Provisioning Policy.

For more information about z/OS Capacity Provisioning functions, see *z/OS MVS Capacity Provisioning User's Guide*, SC34-2661.

Planning considerations for using automatic provisioning

Although only one On/Off CoD offering can be active at any one time, several On/Off CoD offerings can be present on the system. Changing from one to another requires stopping the active one before the inactive one can be activated. This operation decreases the current capacity during the change.

The provisioning management routines can interrogate the installed offerings, their content, and the status of the content of the offering. To avoid the decrease in capacity, create only one On/Off CoD offering on the system by specifying the maximum allowable capacity. The CPM can then, when an activation is needed, activate a subset of the contents of the offering sufficient to satisfy the demand. If more capacity is needed later, the Provisioning Manager can activate more capacity up to the maximum allowed increase.

Multiple offering records can be pre-staged on the SE hard disk. Changing the content of the offerings (if necessary) is also possible.

Remember: CPM controls capacity tokens for the On/Off CoD records. In a situation where a capacity token is used, the system deactivates the corresponding offering record. Therefore, you must prepare routines for catching the warning messages about capacity tokens being used, and have administrative procedures in place for such a situation.

The messages from the system begin five days before a capacity token is fully used. To avoid capacity records being deactivated in this situation, replenish the necessary capacity tokens before they are used.

The Capacity Provisioning Manager operates based on Workload Manager (WLM) indications, and the construct that is used is the Performance Index (PI) of a service class period. It is important to select service class periods that are appropriate for the business application that needs more capacity. For example, the application in question might be running through several service class periods, where the first period is the important one. The application might be defined as importance level 2 or 3, but might depend on other work that is running with importance level 1. Therefore, it is important to consider which workloads to control and which service class periods to specify.

8.7 System Recovery Boost Upgrade

Important: The base System Recovery Boost capability is BUILT INTO z15 firmware and does not require ordering more features. System Recovery Boost Upgrade (consisting of FC 9930 and FC 6802) is an optional, orderable feature that provides more temporary zIIP capacity for use during boost periods. Consider the following points:

- ▶ FC 9930 is *not* required to use the base System Recovery Boost capability.
- ▶ FC 9930 is *only* needed if more zIIP temporary capacity is required.

The System Recovery Boost Upgrade optional feature is offered with z15 servers to provide more capacity for System Recovery Boost processing. For example, if a z/OS operating system change requires a sequence of system shutdowns and restarts, and these procedures can benefit from extra CPU capacity, the System Recovery Boost record can be used to activate more zIIPs on the server at the commencement of the change window. These zIIPs are used by the z/OS systems to run general CP work during the boost periods.

The System Recovery Boost Upgrade requires the following feature codes:

- ▶ FC 6802: System Recovery Boost Record

This feature code provides extra zIIP capacity (20 zIIP records for one year; must be renewed after one year) for use in System Recovery Boost events (shutdown, restart/IPL, and stand-alone dumps).

Note: At least one permanent zIIP record must be present for ordering System Recovery Boost Upgrade (FC 6802).

- ▶ FC 9930: Boost Authorization contract

Enables the ordering of On/Off CoD for System Recovery Boost through Resource Link

Important: The System Recovery Boost Upgrade record is for System Recovery Boost capacity only, and cannot be used for peak workload management. The customer must deactivate the boost capacity at the end of the system restart procedure.

The zIIP processors that can be activated by System Recovery Boost record come from the “dark core” capacity on the server. They can be added to a z15 server nondisruptively.

The base system configuration must have sufficient memory and channels to accommodate the potential requirements of the larger capacity system.

Note: The System Recovery Boost configuration is activated temporarily and provides up to a maximum of 20 extra zIIPs to the system’s original, permanent configuration and can violate the 2:1 zIIP rule. The number of zIIPs that can be activated is limited by the unused capacity that is available on the system.

When activating the System Recovery Boost record, the extra zIIPs are added to the zIIP pool when they are activated. Review the LPAR zIIP assignments and weights in the image profiles to ensure that the LPAR can use the extra capacity when it becomes available.

Configure a quantity for the initial and reserved zIIPs in the image profile so that extra zIIPs can be brought online dynamically when the boost record is activated. Also consider adjusting the LPAR zIIP weight.

When the system recovery event ends, the system must be returned to its original configuration. The System Recovery Boost Upgrade record can be used only once and must be replenished before it can be used again.

A System Recovery Boost Upgrade contract (through FC 9930) must be in place before the special code that enables this capability can be installed on the system.

8.8 Capacity for Planned Event

CPE is offered with z15 servers to provide replacement backup capacity for planned downtime events. For example, if a server room requires an extension or repair work, replacement capacity can be installed temporarily on another z15 server in the client’s environment.

Important: CPE is for planned replacement capacity only, and cannot be used for peak workload management.

CPE includes the following feature codes:

- ▶ FC 6833: Capacity for Planned Event enablement
- ▶ FC 0116: 1 CPE Capacity Unit
- ▶ FC 0117: 100 CPE Capacity Unit
- ▶ FC 0118: 10000 CPE Capacity Unit
- ▶ FC 0119: 1 CPE Capacity Unit - IFL
- ▶ FC 0120: 100 CPE Capacity Unit - IFL
- ▶ FC 0121: 1 CPE Capacity Unit - ICF
- ▶ FC 0122: 100 CPE Capacity Unit - ICF
- ▶ FC 0125: 1 CPE Capacity Unit - zIIP
- ▶ FC 0126: 100 CPE Capacity Unit - zIIP
- ▶ FC 0127: 1 CPE Capacity Unit - SAP
- ▶ FC 0128: 100 CPE Capacity Unit - SAP

The feature codes are calculated automatically when the CPE offering is configured. Whether the eConfig tool or the Resource Link is used, a target configuration must be ordered. The configuration consists of a model identifier, several speciality engines, or both. Based on the target configuration, several feature codes from the list are calculated automatically, and a CPE offering record is constructed.

CPE is intended to replace capacity that is lost within the enterprise because of a planned event, such as a facility upgrade or system relocation.

Note: CPE is intended for short duration events that last a maximum of three days.

After each CPE record is activated, you can access dormant PUs on the system for which you have a contract, as described by the feature codes. Processor units can be configured in any combination of CP or specialty engine types (zIIP, SAP, IFL, and ICF). At the time of CPE activation, the contracted configuration is activated. The general rule of two zIIPs for each configured CP is enforced for the contracted configuration.

The processors that can be activated by CPE come from the available unassigned PUs on any installed processor drawer. CPE features can be added to a z15 server nondisruptively. A one-time fee is applied for each CPE event. This fee depends on the contracted configuration and its resulting feature codes. Only one CPE record can be ordered at a time.

The base system configuration must include sufficient memory and channels to accommodate the potential requirements of the large CPE-configured system. Ensure that all required functions and resources are available on the system where CPE is activated. These functions and resources include CF LEVELs for coupling facility partitions, memory, and cryptographic functions, and include connectivity capabilities.

The CPE configuration is activated temporarily and provides more PUs in addition to the system's original, permanent configuration. The number of extra PUs is predetermined by the number and type of feature codes that are configured, as described by the feature codes. The number of PUs that can be activated is limited by the unused capacity that is available on the system; for example:

- ▶ A z15 Max 71 with 26 CPs, and no IFLs or ICFs has 45 unassigned PUs available.
- ▶ A z15 Max 145 with 38 CPs, 1 IFL, and 1 ICF has 105 unassigned PUs available.

When the planned event ends, the system must be returned to its original configuration. You can deactivate the CPE features at any time before the expiration date.

A CPE contract must be in place before the special code that enables this capability can be installed on the system. CPE features can be added to a z15 server nondisruptively.

8.9 Capacity Backup

CBU provides reserved emergency backup processor capacity for unplanned situations in which capacity is lost in another part of your enterprise. It allows you to recover by adding the reserved capacity on a designated z15 server.

CBU is the quick, temporary activation of PUs:

- ▶ For up to 90 contiguous days, for a loss of processing capacity as a result of an emergency or disaster recovery situation.
- ▶ For 10 days, for testing your disaster recovery procedures or running the production workload. This option requires that IBM Z workload capacity that is equivalent to the CBU upgrade capacity is shut down or otherwise made unusable during the CBU test.³

Important: CBU is for disaster and recovery purposes only. It *cannot* be used for peak workload management or for a planned event.

8.9.1 Ordering

The CBU process allows for CBU to activate CPs, ICFs, zIIPs, IFLs, and SAPs. To use the CBU process, a CBU enablement feature (FC 9910) must be ordered and installed. You must order the quantity and type of PU that you require by using the following feature codes:

- ▶ FC 6805: More CBU test activations
- ▶ FC 6817: Total CBU years ordered
- ▶ FC 6818: CBU records that are ordered
- ▶ FC 6820: Single CBU CP-year
- ▶ FC 6821: 25 CBU CP-year
- ▶ FC 6822: Single CBU IFL-year
- ▶ FC 6823: 25 CBU IFL-year
- ▶ FC 6824: Single CBU ICF-year
- ▶ FC 6825: 25 CBU ICF-year
- ▶ FC 6828: Single CBU zIIP-year
- ▶ FC 6829: 25 CBU zIIP-year
- ▶ FC 6830: Single CBU SAP-year
- ▶ FC 6831: 25 CBU SAP-year
- ▶ FC 6832: CBU replenishment

The CBU entitlement record (FC 6818) contains an expiration date that is established at the time of the order. This date depends on the quantity of CBU years (FC 6817). You can extend your CBU entitlements through the purchase of more CBU years.

The number of FC 6817 per instance of FC 6818 remains limited to five. Fractional years are rounded up to the nearest whole integer when calculating this limit.

³ All new CBU contract documents contain new CBU test terms to allow execution of production workload during CBU test. CBU clients must sign the IBM client Agreement Amendment for IBM Z Capacity Backup Upgrade Tests (US form #Z125-8145).

If two years and eight months exist before the expiration date at the time of the order, the expiration date can be extended by no more than two years. One test activation is provided for each CBU year that is added to the CBU entitlement record.

FC 6805 allows for ordering more tests in increments of one. The maximum number of tests that is allowed is 15 for each FC 6818.

The processors that can be activated by CBU come from the available unassigned PUs on any installed processor drawer. The maximum number of CBU features that can be *ordered* is 190. The number of features that can be *activated* is limited by the number of unused PUs on the system; for example:

- ▶ A z15 Max 34 with Capacity Model Identifier 401 can activate up to 34 CBU features. These CBU features can be used to change the capacity setting of the CPs, and to activate unused PUs.
- ▶ A z15 Max 71 with 15 CPs, 4 IFLs, and 1 ICF has 51 unused PUs available. It can *activate* up to 51 CBU features.

The ordering system allows for over-configuration in the order. You can order up to 190 CBU features, regardless of the current configuration. However, at activation, only the capacity that is installed can be activated. At activation, you can decide to activate only a subset of the CBU features that are ordered for the system.

Subcapacity makes a difference in the way that the CBU features are completed. On the full-capacity models, the CBU features indicate the amount of extra capacity that is needed. If the amount of necessary CBU capacity is equal to four CPs, the CBU configuration is four CBU CPs.

The subcapacity models feature multiple capacity settings of 4xx, 5yy, or 6yy. The standard models use the capacity setting 7nn. To change the capacity setting, the number of CBU CPs must be equal to or greater than the number of CPs in the base configuration.

For example, if the base configuration is a two-way 402, providing a CBU configuration of a four-way of the same capacity setting requires two CBU feature codes. If the required CBU capacity changes the capacity setting of the CPs, going from model capacity identifier 402 to a CBU configuration of a four-way 504 requires four CBU feature codes with a capacity setting of 5yy.

If the capacity setting of the CPs is changed, more CBU features are required, not more physical PUs. Therefore, your CBU contract requires more CBU features when the capacity setting of the CPs is changed.

CBU can add CPs through LICCC only, and the z15 server must have the correct number of installed processor drawers to allow the required upgrade. CBU can change the model capacity identifier to a *higher* value than the base setting (4xx, 5yy, or 6yy), but the CBU feature cannot *decrease* the capacity setting.

A CBU contract must be in place before the special code that enables this capability can be installed on the system. CBU features can be added to a z15 server nondisruptively. For each system enabled for CBU, the authorization to use CBU is available for 1 - 5 years.

The alternative configuration is activated *temporarily*, and provides more capacity than the system's original, *permanent* configuration. At activation time, determine the capacity that you require for that situation. You can decide to activate only a subset of the capacity that is specified in the CBU contract.

The base system configuration must have sufficient memory and channels to accommodate the potential requirements of the large CBU target system. Ensure that all required functions and resources are available on the backup systems. These functions include CF LEVELs for coupling facility partitions, memory, and cryptographic functions, and connectivity capabilities.

When the emergency is over (or the CBU test is complete), the system must be returned to its original configuration. The CBU features can be deactivated at any time before the expiration date. Failure to deactivate the CBU feature before the expiration date can cause the system to downgrade resources gracefully to the original configuration. The system does not deactivate dedicated engines, or the last of in-use shared engines.

Planning: CBU for processors provides a concurrent upgrade. This upgrade can result in more enabled processors, changed capacity settings that are available to a system configuration, or both. You can activate a subset of the CBU features that are ordered for the system. Therefore, more planning and tasks are required for *nondisruptive* logical upgrades. For more information, see “Guidelines to avoid disruptive upgrades” on page 380.

For more information, see the *IBM Z Capacity on Demand User's Guide*, SC28-6846.

8.9.2 CBU activation and deactivation

The activation and deactivation of the CBU function is your responsibility and does not require the onsite presence of IBM SSRs. The CBU function is activated or deactivated concurrently from the HMC by using the API. On the SE, CBU is activated by using the Perform Model Conversion task or through the API. The API enables task automation.

CBU activation

CBU is activated from the SE by using the HMC and SSO to the SE, by using the Perform Model Conversion task, or through automation by using the API on the SE or the HMC. During a real disaster, use the Activate CBU option to activate the 90-day period.

Image upgrades

After CBU activation, the z15 server can have more capacity, more active PUs, or both. The extra resources go into the resource pools and are available to the LPARs. If the LPARs must increase their share of the resources, the LPAR weight can be changed or the number of logical processors can be concurrently increased by configuring reserved processors online. The operating system must concurrently configure more processors online. If necessary, more LPARs can be created to use the newly added capacity.

CBU deactivation

To deactivate the CBU, the extra resources must be released from the LPARs by the operating systems. In some cases, this process involves varying the resources offline. In other cases, it can mean shutting down operating systems or deactivating LPARs. After the resources are released, the same facility on the HMC/SE is used to turn off CBU. To deactivate CBU, select the **Undo temporary upgrade** option from the Perform Model Conversion task on the SE.

CBU testing

Test CBUs are provided as part of the CBU contract. CBU is activated from the SE by using the Perform Model Conversion task. Select the test option to start a 10-day test period. A standard contract allows one test per CBU year. However, you can order more tests in increments of one up to a maximum of 15 for each CBU order.

Tip: The CBU test activation is done the same way as the real activation; that is, by using the same SE Perform a Model Conversion window and selecting the **Temporary upgrades** option. The HMC windows were changed to avoid accidental real CBU activations by setting the test activation as the default option.

The test CBU must be deactivated in the same way as the regular CBU. Failure to deactivate the CBU feature before the expiration date can cause the system to degrade gracefully back to its original configuration. The system does not deactivate dedicated engines or the last in-use shared engine.

CBU example

An example of a CBU operation is shown in Figure 8-15. The permanent configuration is a 504, and a record contains seven CP CBU features. During an activation, multiple target configurations are available. With 7 CP CBU features, you can add up to 7CPs within the same MCI, which allows the activation of a 506, 507, through to a 511 (the blue path).

Alternatively, 4 CP CBU features can be used to change the MCI (in the example from a 504 to a 704) and then add the remaining 3 CP CBU features to upgrade to a 707 (the red path).

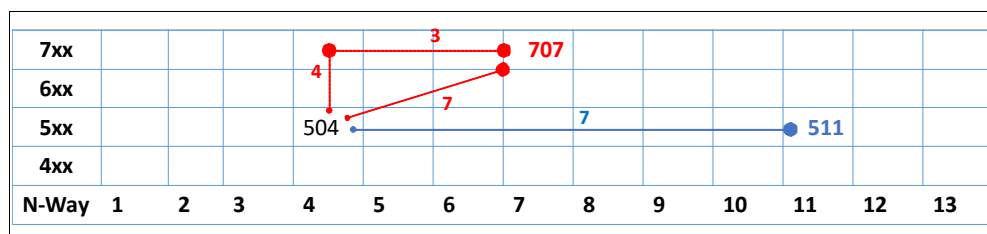


Figure 8-15 CBU example

Note: System Recovery Boost record does *not* affect model capacity identifier.

8.9.3 Automatic CBU enablement for GDPS

The IBM Geographically Dispersed Parallel Sysplex (GDPS) enables automatic management of the PUs that are provided by the CBU feature during a system or site failure. Upon detection of a site failure or planned disaster test, GDPS concurrently adds CPs to the systems in the take-over site to restore processing power for mission-critical production workloads. GDPS automation runs the following tasks:

- ▶ The analysis that is required to determine the scope of the failure. This process minimizes operator intervention and the potential for errors.
- ▶ Automates authentication and activation of the reserved CPs.
- ▶ Automatically restarts the critical applications after reserved CP activation.
- ▶ Reduces the outage time to restart critical workloads from several hours to minutes.

The GDPS service is for z/OS only, or for z/OS in combination with Linux on Z.

8.10 Planning for nondisruptive upgrades

Continuous availability is an important requirement for clients, and planned outages are no longer acceptable. Although Parallel Sysplex clustering technology is the best continuous availability solution for z/OS environments, nondisruptive upgrades within a single system can avoid system outages and cover non-z/OS operating systems.

z15 servers allow *concurrent* upgrades, which means that dynamically adding capacity to the system is possible. If the operating system images that run on the upgraded system do not require disruptive tasks to use the new capacity, the upgrade is also *nondisruptive*. This process avoids power-on resets (POR), LPAR deactivation, and IPLs.

If the concurrent upgrade is intended to satisfy an *image* upgrade to an LPAR, the operating system that is running in this partition must concurrently configure more capacity online. z/OS operating systems include this capability. z/VM can concurrently configure new processors and I/O devices online, and memory can be dynamically added to z/VM partitions.

If the concurrent upgrade is intended to satisfy the need for more operating system images, more LPARs can be created *concurrently* on the z15 system. These LPARs include all resources that are needed. These extra LPARs can be activated concurrently.

These enhanced configuration options are available through the HSA, which is an IBM reserved area in system memory.

Linux operating systems, in general, cannot add more resources concurrently. However, Linux, and other types of virtual machines that run under z/VM, can benefit from the z/VM capability to nondisruptively configure more resources online (processors and I/O).

With z/VM, Linux guests can manipulate their logical processors by using the Linux CPU hotplug daemon. The daemon can start and stop logical processors that are based on the Linux *load average* value. The daemon is available in Linux SLES 10 SP2 and later, and in Red Hat Enterprise Linux (RHEL) V5R4 and up.

8.10.1 Components

The following components can be added, depending on the considerations as described in this section:

- ▶ PUs
- ▶ Memory
- ▶ I/O
- ▶ Cryptographic adapters
- ▶ Special features

PU

CPs, ICFs, zIIPs, IFLs, and SAPs can be added concurrently to a z15 server if unassigned PUs are available on any installed processor drawer. The number of zIIPs cannot exceed twice the number of CPs. The z15 allows the concurrent addition of a second and third processor drawer if the CPC reserve features are installed.

If necessary, more LPARs can be created concurrently to use the newly added processors.

The Coupling Facility Control Code (CFCC) can also configure more processors online to coupling facility LPARs by using the CFCC image operations window.

Memory

Memory can be added concurrently up to the physical installed memory limit. More processor drawers can be installed concurrently, which allows further memory upgrades by LICCC, and enables memory capacity on the new processor drawers.

By using the previously defined reserved memory, z/OS operating system images, and z/VM partitions, you can dynamically configure more memory online. This process allows nondisruptive memory upgrades. Linux on Z supports Dynamic Storage Reconfiguration.

I/O

I/O features can be added concurrently if all the required infrastructure (I/O slots and PCIe Fanouts) is present in the configuration. PCIe+ I/O drawers can be added concurrently without planning if free space is available in one of the frames and the configuration permits.

Dynamic I/O configurations are supported by certain operating systems (z/OS and z/VM), which allows nondisruptive I/O upgrades. Dynamic I/O reconfiguration on a stand-alone coupling facility system is also possible using the Dynamic I/O activation for stand-alone CF CPCs features

Cryptographic adapters

Crypto Express7S features can be added concurrently if all the required infrastructure is in the configuration.

Special features

Special features such as zHyperlink, Coupling Express LR, and RoCE features can be added concurrently if all infrastructure is available in the configuration.

8.10.2 Concurrent upgrade considerations

By using an MES upgrade, On/Off CoD, CBU, or CPE, a z15 server can be upgraded concurrently from one model to another (temporarily or permanently).

Enabling and using the extra processor capacity is not apparent to most applications. However, certain programs depend on processor model-related information, such as ISV products. Consider the effect on the software that is running on a z15 server when you perform any of these configuration upgrades.

Processor identification

The following instructions are used to obtain processor information:

- ▶ Store System Information (STSI) instruction

The STSI instruction can be used to obtain information about the current execution environment and any processing level below the current environment. It can be used to obtain processor model and model capacity identifier information from the basic machine configuration form of the system information block (SYSIB). It supports concurrent upgrades and is the recommended way to request processor information.

- ▶ Store CPU ID (STIDP) instruction

STIDP returns information that identifies the execution environment, system serial number, and machine type.

Note: To ensure unique identification of the configuration of the issuing CPU, use the STSI instruction specifying basic machine configuration (SYSIB 1.1.1).

Store System Information instruction

The format of the basic machine configuration SYSIB that is returned by the STSI instruction is shown in Figure 8-16. The STSI instruction returns the model capacity identifier for the permanent configuration and the model capacity identifier for any temporary capacity. This data is key to the functioning of CoD offerings.

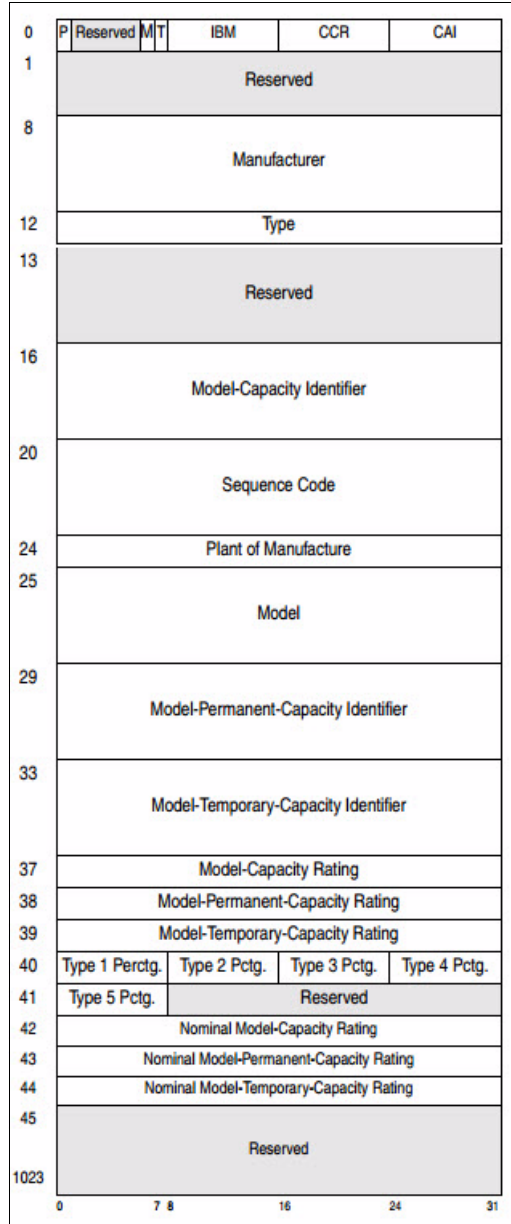


Figure 8-16 Format of system-information block (SYSIB)

The model capacity identifier contains the base capacity, On/Off CoD, and CBU. The Model Permanent Capacity Identifier and the Model Permanent Capacity Rating contain the base capacity of the system. The Model Temporary Capacity Identifier and Model Temporary Capacity Rating contain the base capacity and On/Off CoD.

For more information about the STSI instruction, see *z/Architecture Principles of Operation*, SA22-7832.

Store CPU ID (STIDP) instruction

The STIDP instruction returns information about the processor type, serial number, and LPAR identifier, as shown in Figure 8-17.

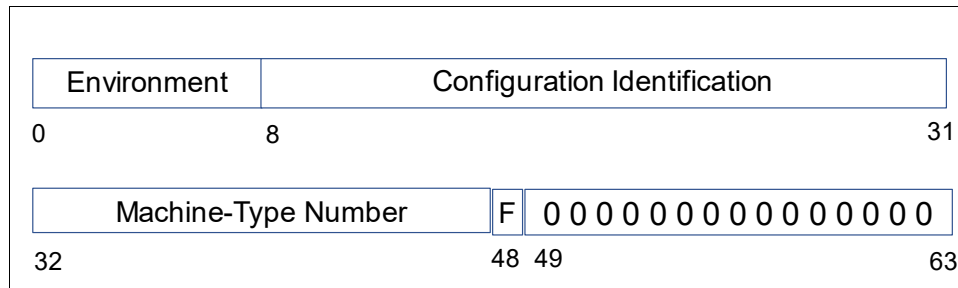


Figure 8-17 STIDP Information

Consider the following points:

- ▶ Bits 0 - 7:
 - For a program that is run by an IBM machine in a level-1 configuration (basic machine mode), or for a program being run by a level-2 configuration (in a logical partition), the environment field contains 00 hex.
 - For a program that is run natively by the System z Personal-Development Tool, the environment field contains C1 hex or D3 hex.
 - For a program that is run by a level-3 configuration (a virtual machine, such as a z/VM guest), the environment field contains FF hex.
- ▶ Bit positions 8 - 31
Contains six hexadecimal digits. The right-most of these digits can represent the machine's serial number.
- ▶ Bit positions 32 - 47
Contains an unsigned packed-decimal number that identifies the machine type of the CPU.
- ▶ Bit position 48
Specifies the format of the first two hexadecimal digits of the configuration-identification field.
- ▶ Bit positions 49 - 63 are reserved and stored as zeros.

For more information about the STIDP instruction, see *z/Architecture Principles of Operation*, SA22-7832.

Planning for nondisruptive upgrades

Online permanent upgrades, On/Off CoD, CBU, and CPE can be used to upgrade a z15 server concurrently. However, certain situations require a disruptive task to enable capacity that was recently added to the system. Some of these situations can be avoided if planning is done in advance. Planning ahead is a key factor for nondisruptive upgrades.

Disruptive upgrades are performed for the following reasons:

- ▶ LPAR memory upgrades when reserved storage was not previously defined are disruptive to image upgrades. z/OS and z/VM support this function.

- ▶ An I/O upgrade when the operating system cannot use the dynamic I/O configuration function is disruptive to that partition. Linux, z/VSE, and z/TPF do not support dynamic I/O configuration.

You can minimize the need for these outages by carefully planning and reviewing “Guidelines to avoid disruptive upgrades” on page 380.

Guidelines to avoid disruptive upgrades

Based on the reasons for disruptive upgrades (see “Planning for nondisruptive upgrades” on page 379), you can use the following guidelines to avoid or at least minimize these situations, which increases the chances for nondisruptive upgrades:

- ▶ By using an SE function that is called Logical Processor add, which is under Operational Customization tasks, CPs and zIIPs can be added concurrently to a running partition. The CP and zIIP and initial or reserved number of processors can be changed dynamically.
- ▶ The operating system that runs in the targeted LPAR must support the dynamic addition of resources and to configure processors online. The total number of defined and reserved CPs cannot exceed the number of CPs that are supported by the operating system. z/OS V2.R4, V2.R3, and V2.R2 support 190 PUs per z/OS LPAR in non-SMT mode and 128 PUs per z/OS LPAR in SMT mode. For both, the PU total is the sum of CPs and zIIPs. z/VM supports up to 64 processors.

- ▶ Configure reserved storage to LPARs.

Configuring reserved storage for all LPARs before their activation enables them to be nondisruptively upgraded. The operating system that is running in the LPAR must configure memory online. The amount of reserved storage can be greater than the CPC drawer threshold limit, even if no other CPC drawer is installed. With z15 servers, the current partition storage limit is 4 TB for z/OS V2.R1 and later. z/VM V7.R1 and V6.R4 support 2 TB memory partitions.

- ▶ Consider the flexible memory options.

Use a convenient entry point for memory capacity, and select memory options that allow future upgrades within the memory cards that are installed on the CPC drawers. For more information about the offerings, see 2.5.7, “Flexible Memory Option” on page 65.

Considerations when installing CPC drawers

During an upgrade, a second and third processor drawer can be installed concurrently if they are pre-planned. Depending on the number of processor drawers in the upgrade and your I/O configuration, a fanout rebalancing might be needed for availability reasons. A fourth or fifth processor drawer can be installed at the IBM Manufacturing plant only.

8.11 Summary of Capacity on-Demand offerings

The CoD infrastructure and its offerings are based on client requirements for more flexibility, granularity, and better business control over the IBM Z infrastructure, operationally, and financially.

One major client requirement was to eliminate the need for a client authorization connection to the IBM Resource Link system when activating an offering. This requirement is met by the z13, z14, and z15 servers.

After the offerings are installed on the z15 SE, they can be activated at any time at the client’s discretion. No intervention by IBM or IBM personnel is necessary. In addition, the activation of CBU does not require a password.

The z15 server can have up to eight offerings installed at the same time, with the limitation that only *one* of them can be an On/Off CoD offering. The others can be any combination. The installed offerings can be activated fully or partially, and in any sequence and any combination. The offerings can be controlled manually through command interfaces on the HMC, or programmatically through a number of APIs. IBM applications, ISV programs, and client-written applications can control the use of the offerings.

Resource usage (and therefore, financial exposure) can be controlled by using capacity tokens in the On/Off CoD offering records.

The CPM is an example of an application that uses the CoD APIs to provision On/Off CoD capacity that is based on the requirements of the workload. The CPM cannot control other offerings.

For more information about any of the topics in this chapter, see *IBM Z Capacity on Demand User's Guide*, SC28-6943.



Reliability, availability, and serviceability

From the Quality perspective, the z15 reliability, availability, and serviceability (RAS) design is driven by a set of high-level program RAS objectives. The IBM Z platform continues to drive toward Continuous Reliable Operation (CRO) at the single footprint level.

Note: Throughout this chapter, *z15* refers to IBM z15 Model T01 (Machine Type 8561), unless otherwise specified.

The key objectives, in order of priority, are to ensure data integrity, computational integrity, reduce or eliminate unscheduled outages, reduce scheduled outages, reduce planned outages, and reduce the number of Repair Actions.

RAS can be accomplished with improved concurrent replace, repair, and upgrade functions for processors, memory, drawers, and I/O. RAS also extends to the nondisruptive capability for installing Licensed Internal Code (LIC) updates. In most cases, a capacity upgrade can be concurrent without a system outage. As an extension to the RAS capabilities, environmental controls are implemented in the system to help reduce power consumption and meet cooling requirements.

This chapter includes the following topics:

- ▶ 9.1, “RAS strategy” on page 384
- ▶ 9.2, “Technology” on page 384
- ▶ 9.3, “Structure” on page 387
- ▶ 9.4, “Reducing complexity” on page 387
- ▶ 9.5, “Reducing touches” on page 388
- ▶ 9.6, “z15 availability characteristics” on page 388
- ▶ 9.7, “z15 RAS functions” on page 392
- ▶ 9.8, “z15 enhanced drawer availability” on page 396
- ▶ 9.9, “z15 Enhanced Driver Maintenance” on page 404
- ▶ 9.10, “RAS capability for the HMC and SE” on page 408

9.1 RAS strategy

The RAS strategy is to manage change by learning from previous generations and investing in new RAS function to eliminate or minimize all sources of outages. Enhancements to z14 RAS designs are implemented on the z15 system through the introduction of new technology, structure, and requirements. Continuous improvements in RAS are associated with new features and functions to ensure that IBM Z servers deliver exceptional value to clients.

The following overriding RAS requirements are principles as shown in Figure 9-1:

- ▶ Inclusion of existing (or equivalent) RAS characteristics from previous generations.
- ▶ Learn from current field issues and addressing the deficiencies.
- ▶ Understand the trend in technology reliability (hard and soft) and ensure that the RAS design points are sufficiently robust.
- ▶ Invest in RAS design enhancements (hardware and firmware) that provide IBM Z and Customer valued differentiation.

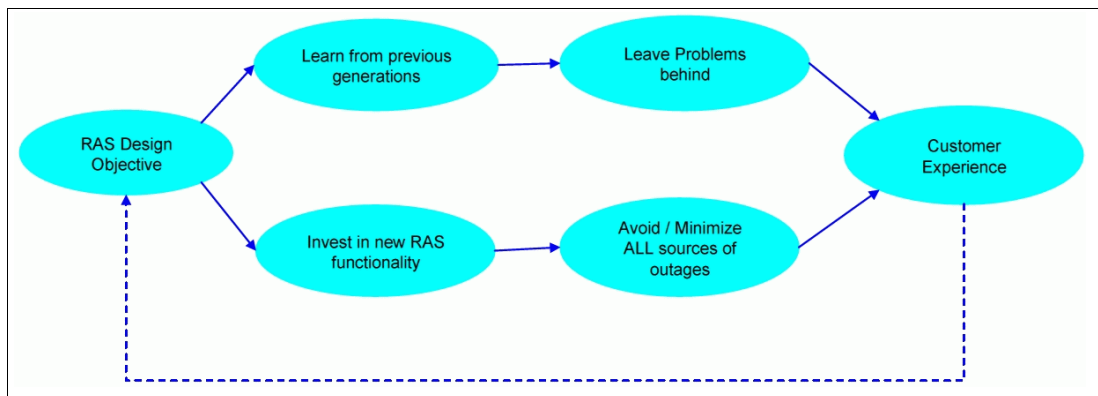


Figure 9-1 Overriding RAS requirements

9.2 Technology

This section introduces some of the RAS features that are incorporated in the z15 design.

9.2.1 Processor Unit chip

The Processor Unit (PU) chip includes the following features:

- ▶ A Single Chip Module (SCM) uses 14nm SOI technology and consists of 17 layers of metal, over 9.1 billion transistors, cores running at 5.2 GHz with 12 cores per PU SCM, all of which enhances the thermal conductivity and improves the reliability.
- ▶ L3:
 - Symbol ECC on L3 data cache
 - 256MB (double the size of L3 in z14)
 - Ability to monitor (dynamically) fenced macros
 - Dynamic cache monitor (“stepper”) to find and demote HSA lines in the cache
 - Wordline span reduced (less impact)
 - Dynamic uMasking for subarrays
- ▶ L2: L2-I (Instruction) is now 4MB (double compared to z14 L2-I)

- ▶ Ability to spare the PU core upon non-L2 cache/DIR Core Array delete.
- ▶ Improved error thresholding on PU cores, which avoids continuous recovery.
- ▶ Memory Control Unit (MCU) Cache Symbol ECC.
- ▶ L1 and L1+; L2 protected by PU sparing.
- ▶ PU Core mandatory address checking.
- ▶ Redundant parity on error in RU bit to protect wordline (WL).
- ▶ On-Chip Compression

The On-Chip Compression, which is new to z15 is a major improvement from the zEDC cards. The on-chip compression offers an industry-leading, hardware-based acceleration for data compression with faster single thread performance.

This On-Chip Compression capability replaces the zEDC Express adapter on the IBM z14 and earlier servers, whereby all data interchange remains compatible.

With zEDC cards the throughput was 1GBps per feature with a maximum of 16 features.

With On-Chip Compression the throughput is improved to 12 GBps per PU, which equates to 48 GBps per drawer and 240 GBps for a fully populated z15.

The On-Chip Compression can be virtualized between all LPARs on the z15, whereas the zEDC feature was limited to 15 LPARs.

The On-Chip Compression module implements DEFLATE/gzip/lzip algorithms and works in a synchronous mode in problem state and an asynchronous mode for larger operations under z/OS.

For more information about IBM Integrated Accelerator for z Enterprise Data Compression (zEDC - On-Chip Compression) on z15 servers, see Appendix C, “IBM Integrated Accelerator for zEnterprise Data Compression” on page 509.

z15 processor memory and cache structure are shown in Figure 9-2.

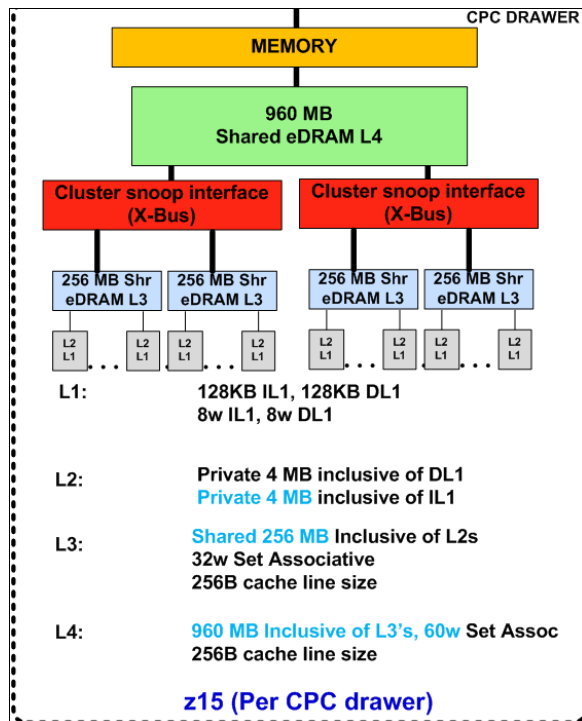


Figure 9-2 Memory and Cache Structure

9.2.2 System Controller and main memory

The System Controller (SC) and main memory consist of the following features:

- ▶ A Single Chip Module (SCM) uses 14nm SOI technology and consists of 17 layers of metal, 9.7 billion transistors, all which enhances the thermal conductivity and improves the reliability.
- ▶ Reduced chip count (single SC chip) improves reliability with fewer components involved.
- ▶ Reduced SMP cables (fewer SC chips) improves reliability and results in faster repair actions and hardware upgrade times because of fewer components to quiesce, test, and resume.
- ▶ L4:
 - Symbol ECC on the cache data, directory, configuration array and on the store protects key cache data.
 - Ability to monitor (dynamically) fenced macros and allow integrated sparing.
 - Array recovery as dynamic array sparing, line deleting, and dynamic uMasking of subarrays.
- ▶ Preemptive memory channel marking:
 - Analysis of uncorrectable errors considers pattern of prior correctable errors
 - More robust uncorrectable error handling
 - Simplified repair action
- ▶ Improved resilience in KUE¹ repair capability
- ▶ Virtual Flash Memory (Flash Express replacement) solution is moved to DIMM:
 - Solution is moved to more robust storage Redundant Array of Independent Memory (RAIM) protected (same function that main memory uses)
 - Concurrent Drawer Repair (CDR) and concurrent drawer addition (CDA)

9.2.3 I/O and service

I/O and service consist of the following features:

- ▶ PCIe+ Fanout Gen3

z15 has a new and improved PCIe+ Fanout Gen3 card with dual 16x ports to provide better availability
- ▶ The number of PSP, support partitions, for managing native PCIe I/O:
 - Four partitions
 - Reduced effect on MCL updates
 - Better availability
- ▶ Faster Dynamic Memory Relocation engine:
 - Enables faster reallocation of memory that is used for LPAR activations, CDR, and concurrent upgrade
 - Provides faster, more robust service actions
- ▶ Dynamic Time Domain Reflectometry (TDR):
 - Hardware facility that is used to isolate failures on wires provides better FRU isolation and improved service actions.

¹ Key in storage error uncorrected: Indicates that the hardware cannot repair a storage key that was in error.

- ▶ Universal spare for PU SCMs and SC SCMs and processor drawer.
- ▶ z15 has a new and improved Fill and Drain tool.

9.3 Structure

The z15 server is built in a new form factor of 1- 4 19-inch frames. The z15 server can be delivered as an air-cooled and water-cooled system and fulfills the requirements for an ASHRAE A3 environment.

The z15 server can have up to 11 PCIe+ I/O drawers when delivered with Bulk Power Assembly (BPA) and 12 PCIe+ I/O drawers when delivered with Power Distribution Unit (PDU). The structure changes to the z15 server are done with the following goals:

- ▶ Enhanced system modularity
- ▶ Standardization to enable rapid integration
- ▶ Platform simplification

Cables are keyed to ensure that correct lengths are plugged. Plug detection ensures correct location, and custom latches ensure retention. Further improvements to the fabric bus include symmetric multiprocessing (SMP) cables that connect the drawers.

To improve field-replaceable unit (FRU) isolation, TDR techniques are applied to the SMP cables, between chips (PU-PU, and PU-SC), and between the PU chips and dual inline memory modules (DIMMs).

Enhancements to thermal RAS also were introduced, such as a field-replaceable water manifold for PU cooling. The z15 has the following characteristics:

- ▶ Processing infrastructure is designed by using drawer technology.
- ▶ Keyed cables and plugging detection.
- ▶ SMP cables that are used for fabric bus connections.
- ▶ Water manifold is a FRU.
- ▶ Master-master redundant oscillator design in the main memory.
- ▶ Processor and nest chips are separate FRUs.
- ▶ Point of load cards are separate FRUs.
- ▶ Two combined Flexible Service Processor (FSP) and Oscillator Cards (OSC) are provided per CPU draw.
- ▶ Built in time domain reflectometer for FRU isolation in interface errors.
- ▶ Redundant N+1 Power Supply Units (PSU) to CPC drawer and PCIe+ drawer.

9.4 Reducing complexity

z15 servers continue the z14 enhancements that reduced system RAS complexity. Specifically, simplifications were made in RAIM recovery in the memory subsystem design. Memory DIMMs are no longer cascaded, which eliminates the double FRU call for DIMM errors.

Independent channel recovery with replay buffers on all interfaces allows recovery of a single DIMM channel, while other channels remain active. Further redundancies are incorporated in I/O pins for clock lines to main memory, which eliminates the loss of memory clocks because of connector (pin) failure. The following RAS enhancements reduce service complexity:

- ▶ Continued use of RAIM ECC.
- ▶ No cascading of memory DIMM to simplify the recovery design.
- ▶ Replay buffer for hardware retry on soft errors on the main memory interface.
- ▶ Redundant I/O pins for clock lines to main memory.

9.5 Reducing touches

IBM Z RAS efforts focus on the reduction of unscheduled, scheduled, planned, and unplanned outages. IBM Z technology has a long history of demonstrated RAS improvements, and this effort continues with changes that reduce service *touches* on the system.

Firmware was updated to improve filtering and resolution of errors that do not require action. Enhanced integrated sparing in processor cores, cache relocates, N+1 SEEPROM and POL N+2 redundancies, and DRAM marking also are incorporated to reduce touches. The following RAS enhancements reduce service touches:

- ▶ Improved error resolution to enable filtering
- ▶ Enhanced integrated sparing in processor cores
- ▶ Cache relocates
- ▶ N+1 SEEPROM
- ▶ N+2 POL
- ▶ DRAM marking
- ▶ (Dynamic) Spare lanes for PU-SC, PU-PU, PU-mem, and SC-SMP fabric
- ▶ N+1 radiator pumps, controllers, blowers, and sensors
- ▶ N+1 Ethernet switches
- ▶ N+1 Support Element (SE) (with N+1 SE power supplies)
- ▶ Redundant SEEPROM on memory DIMM
- ▶ Redundant temperature sensor (one SEEPROM and one temperature sensor per I2C bus)
- ▶ FICON forward error correction

9.6 z15 availability characteristics

The following functions include availability characteristics on z15 servers:

- ▶ Enhanced drawer availability (EDA)
EDA is a *procedure* under which a CPC drawer in a multidrawer system can be removed and reinstalled during an upgrade or repair action with no effect on the workload.
- ▶ Concurrent memory upgrade or replacement
Memory can be upgraded concurrently by using Licensed Internal Code Configuration Control (LICCC) if physical memory is available on the drawers.

The EDA function can be useful if the physical memory cards must be changed in a multidrawer configuration (requiring the drawer to be removed).

It requires the availability of more memory resources on other drawers or reducing the need for memory resources during this action. Select the flexible memory option to help ensure that the appropriate level of memory is available in a multiple-drawer configuration. This option provides more resources to use EDA when repairing a drawer or memory on a drawer. They are also available when upgrading memory when larger memory cards might be required.

- ▶ Enhanced driver maintenance (EDM)

One of the greatest contributors to downtime during planned outages is LIC driver updates that are performed in support of new features and functions. z15 servers are designed to support the concurrent activation of a selected new driver level.

- ▶ Plan Ahead for Balanced Power (FC 3003)

This feature allows you to order the maximum number of bulk power regulators (BPRs) on any server configuration. This feature helps to ensure that your configuration is in a balanced power environment if you intend to add CPC drawers and I/O drawers to your server in the future. The feature is available with Bulk Power Adapters (BPA) only.

- ▶ Concurrent fanout addition or replacement

A PCIe+ fanout card provides the path for data between memory and I/O through PCIe cables. With z15 servers, a hot-pluggable and concurrently upgradeable fanout card is available. Up to 12 PCIe fanout cards per CPC drawer are available for z15 servers. A z15 Model T01 feature Max190 holds five CPC drawers and can have 60 PCIe fan out slots.

Internal I/O paths from the CPC drawer fanout ports to a PCIe drawer or an I/O drawer are spread across multiple CPC drawers (for feature Max71, Max108, Max145, and Max190) and across different nodes within a single CPC drawer Feature Max34. During an outage, a fanout card that is used for I/O can be repaired concurrently while redundant I/O interconnect ensures that no I/O connectivity is lost.

- ▶ Redundant I/O interconnect

Redundant I/O interconnect helps maintain critical connections to devices. z15 servers allow a single drawer, in a multidrawer system, to be removed and reinstalled concurrently during an upgrade or repair. Connectivity to the system I/O resources is maintained through a second path from a different drawer.

- ▶ Flexible Service Processor (FSP) / Oscillator Cards (OSC).

z15 servers have two combined Flexible Service Processor (FSP) and Oscillator Cards (OSC) per CPU draw. The strategy of redundant clock and switchover stays the same. One primary and one backup is available. If the primary OSC fails, the backup detects the failure, takes over transparently, and continues to provide the clock signal to the CPC.

- ▶ Processor unit (PU) sparing

z15 servers have two spare PUs per CPU drawer to maintain performance levels if an active PU, Internal Coupling Facility (ICF), Integrated Facility for Linux (IFL), IBM Z Integrated Information Processor (zIIP), integrated firmware processor (IFP), or system assist processor (SAP) fails. Transparent sparing for failed processors is supported and sparing is supported across the drawers in the unlikely event that the drawer with the failure does not have spares available.

- ▶ Application preservation

This function is used when a PU fails and no spares are available. The state of the failing PU is passed to another active PU, where the operating system uses it to successfully resume the task, in most cases without client intervention.

► Cooling improvements

The z15 air-cooled configuration includes a newly designed front to rear radiator cooling system. The radiator pumps, blowers, controls, and sensors are N+2 redundant. In normal operation, one active pump supports the system. A second pump is turned on and the original pump is turned off periodically, which improves reliability of the pumps. The replacement of pumps or blowers is concurrent with no affect on performance.

A water-cooling system also is an option in z15 servers, with water-cooling unit (WCU) technology. Two redundant WCUs run with two independent chilled water feeds. One WCU and one water feed can support the entire system load. The water-cooled configuration is backed up by the rear door heat exchangers in the rare event of a problem with the chilled water facilities of the customer.

► FICON Express16SA / FICON Express16S+ with Forward Error Correction (FEC)

FICON Express16SA and FICON Express16S+ features continue to provide a new standard for transmitting data over 16 Gbps links by using 64b/66b encoding. The new standard that is defined by T11.org FC-FS-3 is more efficient than the current 8b/10b encoding.

FICON Express16SA and FICON Express16S+ channels that are running at 16 Gbps can take advantage of FEC capabilities when connected to devices that support FEC.

FEC allows FICON Express16SA and FICON Express16S+ channels to operate at higher speeds, over longer distances, with reduced power and higher throughput. They also retain the same reliability and robustness for which FICON channels are traditionally known.

FEC is a technique that is used for controlling errors in data transmission over unreliable or noisy communication channels. When running at 16 Gbps link speeds, clients often see fewer I/O errors, which reduces the potential effect to production workloads from those I/O errors.

Read Diagnostic Parameters (RDP) improve Fault Isolation. After a link error is detected (for example, IFCC, CC3, reset event, or a link incident report), link data that is returned from Read Diagnostic Parameters is used to differentiate between errors that result from failures in the optics versus failures because of dirty or faulty links.

Key metrics can be displayed on the operator console. The results of a display matrix command with the LINKINFO=FIRST parameter, which collects information from each device in the path from the channel to the I/O device (see Figure 9-3 on page 391):

- Transmit (Tx) and Receive (Rx) optic power levels from the PCHID, Switch Input and Output, and I/O device
- Capable and Operating speed between the devices
- Error counts
- Operating System requires new function APAR OA49089

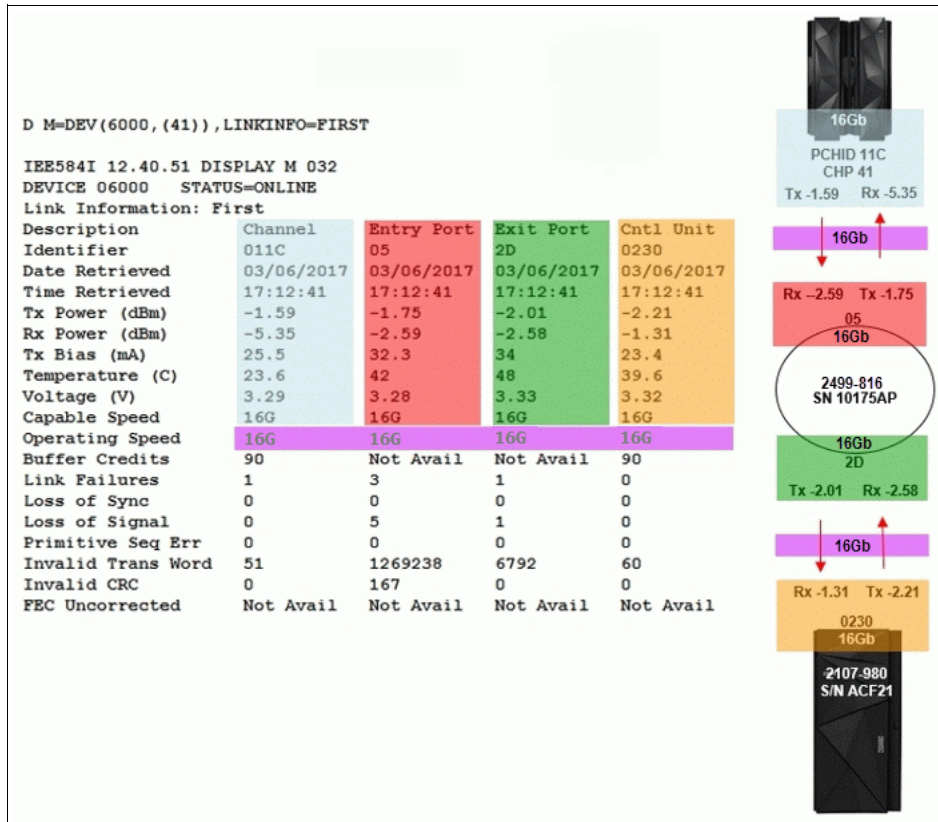


Figure 9-3 Read Diagnostic Parameters function

The new IBM Z Channel Subsystem Function performs periodic polling from the channel to the end points for the logical paths that are established and reduces the number of useless Repair Actions (RAs).

The RDP data history is used to validate Predictive Failure Algorithms and identify Fibre Channel Links with degrading signal strength before errors start to occur. The new Fibre Channel Extended Link Service (ELS) retrieves signal strength.

► FICON Dynamic Routing

FICON Dynamic Routing (FIDR) enables the use of storage area network (SAN) dynamic routing policies in the fabric. With the z15 server, FICON channels are no longer restricted to the use of static routing policies for inter-switch links (ISLs) for cascaded FICON directors.

FICON Dynamic Routing dynamically changes the routing between the channel and control unit based on the Fibre Channel Exchange ID. Each I/O operation has a unique exchange ID. FIDR is designed to support static SAN routing policies and dynamic routing policies.

FICON Dynamic Routing can help clients reduce costs by providing the following features:

- Share SANs between their FICON and FCP traffic.
- Improve performance because of SAN dynamic routing policies that better use all the available ISL bandwidth through higher use of the ISLs,
- Simplify management of their SAN fabrics by using static routing policies that assign different ISL routes with each power-on-reset (POR), which makes the SAN fabric performance difficult to predict.

Clients must ensure that all devices in their FICON SAN support FICON Dynamic Routing before they implement this feature.

9.7 z15 RAS functions

Hardware RAS function improvements focus on addressing all sources of outages. Sources of outages feature the following classifications:

- ▶ **Unscheduled**

This outage occurs because of an unrecoverable malfunction in a hardware component of the system.

- ▶ **Scheduled**

This outage is caused by changes or updates that must be done to the system in a timely fashion. A scheduled outage can be caused by a disruptive patch that must be installed, or other changes that must be made to the system.

- ▶ **Planned**

This outage is caused by changes or updates that must be done to the system. A planned outage can be caused by a capacity upgrade or a driver upgrade. A planned outage usually is requested by the client, and often requires pre-planning. The z15 design phase focuses on enhancing planning to simplify or eliminate planned outages.

The difference between scheduled outages and planned outages might not be obvious. The general consensus is that scheduled outages occur sometime soon. The time frame is approximately two weeks.

Planned outages are outages that are planned well in advance and go beyond this approximate two-week time frame. This chapter does not distinguish between scheduled and planned outages.

Preventing unscheduled, scheduled, and planned outages was addressed by the IBM Z system design for many years.

z15 servers have a fixed size HSA of 256 GB (up from 192 GB on z14). This size helps eliminate pre-planning requirements for HSA and provides the flexibility to update dynamically the configuration. You can perform the following tasks dynamically:²

- ▶ Add a logical partition (LPAR)
- ▶ Add a logical channel subsystem (LCSS)
- ▶ Add a subchannel set
- ▶ Add a logical PU to an LPAR
- ▶ Add a cryptographic coprocessor
- ▶ Remove a cryptographic coprocessor
- ▶ Enable I/O connections
- ▶ Swap processor types
- ▶ Add memory
- ▶ Add a physical processor

By addressing the elimination of planned outages, the following tasks also are possible:

- ▶ Concurrent driver upgrades
- ▶ Concurrent and flexible customer-initiated upgrades

² Some pre-planning considerations might exist. For more information, see Chapter 8, "System upgrades" on page 333.

For more information about the flexible upgrades that are started by clients, see 8.2.2, “Customer Initiated Upgrade facility” on page 342.

- ▶ STP management of concurrent CTN Split and Merge
- ▶ Dynamic I/O for stand-alone CF CPCs

Dynamic I/O configuration changes can be made to a stand-alone CF without requiring a disruptive power on reset. An LPAR with a firmware-based appliance version of an HCD instance is used to apply the new I/O configuration changes. The firmware-based LPAR is driven by updates from an HCD instance that is running in a z/OS LPAR on a different CPC that is connected to the same z15 HMC.

- ▶ System Recovery Boost

System Recovery Boost, which is new on z15, introduces the possibility of reducing the downtime from an operating systems perspective in both scheduled and unscheduled events on a partition basis. This reduction in downtime is achieved by delivering more CP capacity for a boost period before a scheduled shutdown and following a restart.

For more information about System Recovery Boost, see Appendix B, “System Recovery Boost” on page 493.

9.7.1 Scheduled outages

Concurrent hardware upgrades, parts replacement, driver upgrades, and firmware fixes that are available with z15 servers all address the elimination of scheduled outages. Also, the following indicators and functions that address scheduled outages are included:

- ▶ Double memory data bus lane sparing.
This feature reduces the number of repair actions for memory.
- ▶ Single memory clock sparing.
- ▶ Double DRAM chipkill tolerance.
- ▶ Field repair of the cache fabric bus.
- ▶ Processor drawer power distribution $N+2$ design.

The CPC Drawer uses point of load (POL) cards in a highly redundant $N+2$ configuration. POL regulators are daughter cards that contain the voltage regulators for the principle logic voltage boundaries in the z15 CPC drawer. They plug onto the CPC drawer system board and are nonconcurrent FRUs for the affected drawer, similar to the memory DIMMs. If you can use EDA, the replacement of POL cards is concurrent for the whole Z server.

- ▶ Redundant ($N+1$) Ethernet switches.
- ▶ Redundant ($N+2$) humidity sensors.
- ▶ Redundant ($N+2$) altimeter sensors.
- ▶ Redundant ($N+2$) ambient temperature sensors.
- ▶ Dual inline memory module (DIMM) field-replaceable unit (FRU) indicators.

These indicators imply that a memory module is not error-free and might fail sometime in the future. This indicator gives IBM a warning and provides time to concurrently repair the storage module if the z15 is a multidrawer system.

The process to repair the storage module is to isolate or “fence off” the drawer, remove the drawer, replace the failing storage module, and then add the drawer. The flexible memory option might be necessary to maintain sufficient capacity while repairing the storage module.

- ▶ Single processor core checkstop and sparing.
This indicator shows that a processor core malfunctioned and is *spared*. IBM determines what to do based on the system and the history of that system.
- ▶ Point-to-point fabric for symmetric multiprocessing (SMP).
Having fewer components that can fail is an advantage. In a multidrawer system, all of the drawers are connected by point-to-point connections. A drawer can always be added concurrently.
- ▶ Air-cooled system: radiator with redundant ($N+2$) pumps.
z15 servers implement true $N+2$ redundancy on pumps and blowers for the radiator.
One radiator unit in Frame A and one radiator unit in frame B (configuration-dependant).
The radiator cooling system can support up to three CPC drawers simultaneously with a redundant design that consists of two pumps and two blowers.
The replacement of a pump or blower causes no performance effect.
- ▶ Water-cooled system: $N+1$ Water-Cooling Units (WCUs).
A water-cooling system is an option in z15 servers with WCU technology. Two redundant WCUs run with two independent chilled water feeds for each frame A and B (if B is installed). One WCU and one water feed can support the entire system load. The water-cooled configuration is backed up by the rear door heat exchangers in the rare event of a problem with the chilled water facilities of the customer.
- ▶ The PCIe+ I/O drawer is available for z15 servers. It and all of the PCIe+ I/O drawer-supported features can be installed concurrently.
- ▶ Memory interface logic to maintain channel synchronization when one channel goes into replay. z15 servers can isolate recovery to only the failing channel.
- ▶ Out-of-band access to DIMM (for background maintenance functions).
Out-of-band access (by using an I2C interface) allows maintenance (such as logging) without disrupting customer memory accesses.
- ▶ Lane shadowing function to each lane that periodically is taken offline (for recalibration).
The (logical) spare bit lane is rotated through the (physical) lanes. This configuration allows the lane to be tested and recalibrated transparently to customer operations.
- ▶ Automatic lane recalibration on offline lanes on the main memory interface. Hardware support for transparent recalibration is included.
- ▶ Automatic dynamic lane sparing based on pre-programmed CRC thresholds on the main memory interface. Hardware support to detect a defective lane and spare it out is included.
- ▶ Improved DIMM exerciser for testing memory during IML.
- ▶ PCIe redrive hub cards plug straight in (no blind mating of connector). Simplified plugging that is more reliable is included.
- ▶ ICA (short distance) coupling cards plug straight in (no blind mating of connector). Simplified plugging that is more reliable is included.
- ▶ Coupling Express LR (CE LR) coupling cards plug into the PCIe+ I/O drawer, which allows more connections with a faster bandwidth.
- ▶ Hardware-driven dynamic lane sparing on fabric (SMP) buses. Increased bit lane sparing is featured.

9.7.2 Unscheduled outages

An *unscheduled outage* occurs because of an unrecoverable malfunction in a hardware component of the system.

The following improvements can minimize unscheduled outages:

- ▶ Continued focus on firmware quality

For LIC and hardware design, failures are eliminated through rigorous design rules; design walk-through; peer reviews; element, subsystem, and system simulation; and extensive engineering and manufacturing testing.

- ▶ Memory subsystem

RAIM on Z servers is a concept similar to the concept of Redundant Array of Independent Disks (RAID). The RAIM design detects and recovers from dynamic random access memory (DRAM), socket, memory channel, or DIMM failures. The RAIM design requires the adding one memory channel that is dedicated for RAS.

The parity of the four data DIMMs is stored in the DIMMs that are attached to the fifth memory channel. Any failure in a memory component can be detected and corrected dynamically. z15 servers inherited this memory architecture.

The memory system on z15 servers is implemented with an enhanced version of the Reed-Solomon ECC that is known as 90B/64B. It provides protection against memory channel and DIMM failures.

A precise marking of faulty chips helps ensure timely DIMM replacements. The design of the z15 server further improved this chip marking technology. Graduated DRAM marking is available, and channel marking and scrubbing calls for replacement on the third DRAM failure is available. For more information about the memory system on z15 servers, see 2.5, “Memory” on page 56.

- ▶ Soft-switch firmware

z15 servers are equipped with the capabilities of soft-switching firmware. Enhanced logic in this function ensures that every affected circuit is powered off during the soft-switching of firmware components. For example, when you are upgrading the microcode of a FICON feature, enhancements are implemented to avoid any unwanted side effects that were detected on previous systems.

- ▶ Server Time Protocol (STP) recovery enhancement

When PCIe-based integrated communication adapter (ICA) Short Reach (SR) links are used, an unambiguous “going away signal” is sent when the server on which the coupling link is running is about to enter a failed (check stopped) state.

When the “going away signal” that is sent by the Current Time Server (CTS) in an STP-only Coordinated Timing Network (CTN) is received by the Backup Time Server (BTS), the BTS can safely take over as the CTS without relying on the previous Offline Signal (OLS) in a two-server CTN, or as the Arbiter in a CTN with three or more servers.

Enhanced Console Assisted Recovery (ECAR) was new with z13s and z13 GA2 and carried forward to z14 and z15. It contains better recovery algorithms during a failing Primary Time Server (PTS) and uses communication over the HMC/SE network to assist with BTS takeover. For more information, see Chapter 10, “Hardware Management Console and Support Element” on page 411.

Coupling Express LR does not support the “going away signal”; however, ECAR can be used to assist with recovery in the following configurations:

- ▶ Design of pervasive infrastructure controls in processor chips in memory ASICs.
- ▶ Improved error checking in the processor recovery unit (RU) to better protect against word line failures in the RU arrays.

9.8 z15 enhanced drawer availability

Enhanced drawer availability (EDA) is a procedure in which a drawer in a multidrawer system can be removed and reinstalled during an upgrade or repair action. This procedure has no effect on the running workload.

The EDA procedure and careful planning help ensure that all the resources are still available to run critical applications in an ($n-1$) drawer configuration. This process allows you to avoid planned outages. Consider the flexible memory option to provide more memory resources when you are replacing a drawer. For more information about flexible memory, see 2.5.7, “Flexible Memory Option” on page 65.

To minimize the effect on current workloads, ensure that sufficient inactive physical resources exist on the remaining drawers to complete a drawer removal. Also, consider deactivating non-critical system images, such as test or development LPARs. After you stop these non-critical LPARs and free their resources, you might find sufficient inactive resources to contain critical workloads while completing a drawer replacement.

9.8.1 EDA planning considerations

To use the EDA function, configure enough physical memory and engines so that the loss of a single drawer does not result in any degradation to critical workloads during the following occurrences:

- ▶ A degraded restart in the rare event of a drawer failure
- ▶ A drawer replacement for repair or a physical memory upgrade

The following configurations especially enable the use of the EDA function. These z15 features need enough spare capacity so that they can cover the resources of a fenced or isolated drawer. This configuration imposes limits on the following number of the client-owned PUs that can be activated when one drawer within a model is fenced:

- ▶ A maximum of 34 client PUs are configured on the Max34.
- ▶ A maximum of 71 client PUs are configured on the Max71.
- ▶ A maximum of 108 client PUs are configured on the Max108.
- ▶ A maximum of 145 client PUs are configured on the Max145.
- ▶ A maximum of 190 client PUs are configured on the Max190.
- ▶ No special feature codes are required for PU and model configuration.
- ▶ Feature Max34 to Max145 each have 4 SAPs in each drawer. Max190 has in total 22 standard SAPs.
- ▶ The flexible memory option delivers physical memory so that 100% of the purchased memory increment can be activated even when one drawer is fenced.

The system configuration must have sufficient dormant resources on the remaining drawers in the system for the *evacuation* of the drawer that is to be replaced or upgraded. Dormant resources include the following possibilities:

- ▶ Unused PUs or memory that is not enabled by LICCC
- ▶ Inactive resources that are enabled by LICCC (memory that is not being used by any activated LPARs)
- ▶ Memory that is purchased with the flexible memory option
- ▶ Extra drawers

The I/O connectivity must also support drawer removal. Most of the paths to the I/O feature redundant I/O interconnect support in the I/O infrastructure (drawers) that enable connections through multiple fanout cards.

If sufficient resources are not present on the remaining drawers, certain non-critical LPARs might need to be deactivated. One or more PUs or storage might need to be configured offline to reach the required level of available resources. Plan to address these possibilities to help reduce operational errors.

Exception: Single-drawer systems cannot use the EDA procedure.

Include the planning as part of the initial installation and any follow-on upgrade that modifies the operating environment. A client can use the Resource Link machine information report to determine the number of drawers, active PUs, memory configuration, and channel layout.

If the z15 server is installed, click **Prepare for Enhanced Drawer Availability** in the Perform Model Conversion window of the EDA process on the Hardware Management Console (HMC). This task helps you determine the resources that are required to support the removal of a drawer with acceptable degradation to the operating system images.

The EDA process determines which resources, including memory, PUs, and I/O paths, are free to allow for the removal of a drawer. You can run this preparation on each drawer to determine which resource changes are necessary. Use the results as input in the planning stage to help identify critical resources.

With this planning information, you can examine the LPAR configuration and workload priorities to determine how resources might be reduced and still allow the drawer to be concurrently removed.

Include the following tasks in the planning process:

- ▶ Review of the z15 configuration to determine the following values:
 - Number of drawers that are installed and the number of PUs enabled. Consider the following points:
 - Use the Resource Link machine information or the HMC to determine the model, number, and types of PUs (CPs, IFLs, ICFs, and zIIPs).
 - Determine the amount of memory (physically installed and LICCC-enabled).
 - Work with your IBM Service Support Representative (IBM SSR) to determine the memory card size in each drawer. The memory card sizes and the number of cards that are installed for each drawer can be viewed from the SE under the CPC configuration task list. Use the View Hardware Configuration option.

- ICA SR fanout layouts and ICA to ICA connections.

Use the Resource Link machine information to review the channel configuration. This process is a normal part of the I/O connectivity planning. The alternative paths must be separated as far into the system as possible.

- ▶ Review the system image configurations to determine the resources for each image.
- ▶ Determine the importance and relative priority of each LPAR.
- ▶ Identify the LPAR or workloads and the actions to be taken:
 - Deactivate the entire LPAR.
 - Configure PUs.
 - Reconfigure memory, which might require the use of reconfigurable storage unit (RSU) values.
 - Vary off the channels.
- ▶ Review the channel layout and determine whether any changes are necessary to address single paths.
- ▶ Develop a plan to address the requirements.

When you perform the review, document the resources that can be made available if the EDA is used. The resources on the drawers are allocated during a POR of the system and can change after that process. Perform a review when changes are made to z15 servers, such as adding drawers, PUs, memory, or channels. Also, perform a review when workloads are added or removed, or if the HiperDispatch feature was enabled and disabled since the last time you performed a POR.

9.8.2 Enhanced drawer availability processing

To use the EDA, first ensure that the following conditions are met:

- ▶ Free the used processors (PUs) on the drawer that is removed.
- ▶ Free the used memory on the drawer.
- ▶ For all I/O domains that are connected to the drawer and ensure that alternative paths exist. Otherwise, place the I/O paths offline.

For the EDA process, this phase is the preparation phase. It is started from the SE, directly or on the HMC, by using the Single object operation option on the Perform Model Conversion window from the CPC configuration task list, as shown in Figure 9-4.

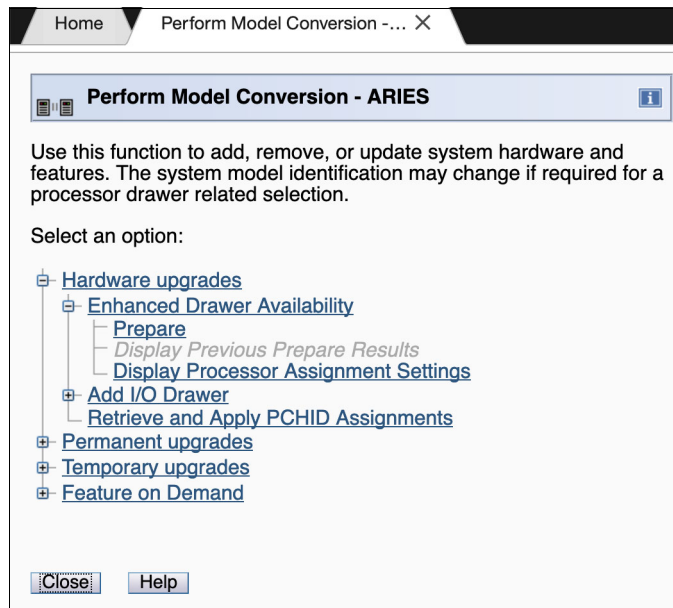


Figure 9-4 Clicking Prepare for Enhanced Drawer Availability option

Processor availability

Processor resource availability for reallocation or deactivation is affected by the type and quantity of the resources in use, such as:

- ▶ Total number of PUs that are enabled through LICCC
- ▶ PU definitions in the profiles that can be dedicated and dedicated reserved or shared
- ▶ Active LPARs with dedicated resources at the time of the drawer repair or replacement

To maximize the PU availability option, ensure that sufficient inactive physical resources are on the remaining drawers to complete a drawer removal.

Memory availability

Memory resource availability for reallocation or deactivation depends on the following factors:

- ▶ Physically installed memory
- ▶ Image profile memory allocations
- ▶ Amount of memory that is enabled through LICCC
- ▶ Flexible memory option
- ▶ Virtual Flash Memory if enabled and configured

For more information, see 2.7.2, “Enhanced drawer availability (EDA)” on page 73.

Fan out card to I/O connectivity requirements

The optimum approach is to maintain maximum I/O connectivity during drawer removal. The redundant I/O interconnect (RII) function provides for redundant connectivity to all installed I/O domains in the PCIe+ I/O drawers.

Preparing for enhanced drawer availability

The Prepare Concurrent Drawer replacement option validates that enough dormant resources are available for this operation. If enough resources are not available on the remaining drawers to complete the EDA process, the process identifies those resources. It then guides you through a series of steps to select and free up those resources. The preparation process does not complete until all processors, memory, and I/O conditions are successfully resolved.

Preparation: The preparation step does not reallocate any resources. It is used only to record client choices and produce a configuration file on the SE that is used to run the concurrent drawer replacement operation.

The preparation step can be done in advance. However, if any changes to the configuration occur between the preparation and the physical removal of the drawer, you must rerun the preparation phase.

The process can be run multiple times because it does not move any resources. To view the results of the last preparation operation, click **Display Previous Prepare Enhanced Drawer Availability Results** from the Perform Model Conversion window in the SE.

The preparation step can be run without performing a drawer replacement. You can use it to dynamically adjust the operational configuration for drawer repair or replacement before IBM SSR activity. The Perform Model Conversion window in you click **Prepare for Enhanced Drawer Availability** is shown in Figure 9-4 on page 399.

After you click **Prepare for Enhanced Drawer Availability**, the Enhanced Drawer Availability window opens. Select the drawer that is to be repaired or upgraded; then, select **OK**, as shown in Figure 9-5. Only one target drawer can be selected at a time.

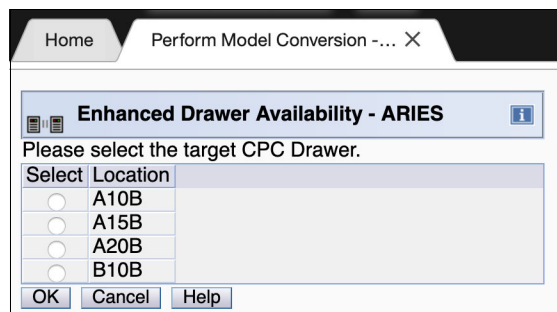


Figure 9-5 Selecting the target drawer

The system verifies the resources that are required for the removal, determines the required actions, and presents the results for review. Depending on the configuration, the task can take from a few seconds to several minutes.

The preparation step determines the readiness of the system for the removal of the targeted drawer. The configured processors and the memory in the selected drawer are evaluated against unused resources that are available across the remaining drawers. The system also analyzes I/O connections that are associated with the removal of the targeted drawer for any single path I/O connectivity.

If insufficient resources are available, the system identifies the conflicts so that you can free other resources.

The following states can result from the preparation step:

- ▶ The system is ready to run the EDA for the targeted drawer with the original configuration.
- ▶ The system is not ready to run the EDA because of conditions that are indicated by the preparation step.
- ▶ The system is ready to run the EDA for the targeted drawer. However, to continue with the process, processors are reassigned from the original configuration.

Review the results of this reassignment relative to your operation and business requirements. The reassignments can be changed on the final window that is presented. However, before making any changes or approving reassignments, ensure that the changes are reviewed and approved by the correct level of support based on your organization's business requirements.

Preparation tabs

The results of the preparation are presented for review in a tabbed format. Each tab indicates conditions that prevent the EDA option from being run. The following tab selections are available:

- ▶ Processors
- ▶ Memory
- ▶ Single I/O
- ▶ Single Domain I/O
- ▶ Single Alternate Path I/O

Only the tabs that feature conditions that prevent the drawer from being removed are displayed. Each tab indicates the specific conditions and possible options to correct them.

For example, the preparation identifies single I/O paths that are associated with the removal of the selected drawer. These paths must be varied offline to perform the drawer removal. After you address the condition, rerun the preparation step to ensure that all the required conditions are met.

Preparing the system to perform enhanced drawer availability

During the preparation, the system determines the PU configuration that is required to remove the drawer. The results and the option to change the assignment on non-dedicated processors are shown in Figure 9-6.

The screenshot shows a window titled "Processor Assignments - ARIES" with a tabbed interface. The active tab is "Drawer_A20B". Below the tabs is a table with the following data:

| Processor Type | Dedicated Count | Non-Dedicated Count | Processor Totals | LICCC Count |
|-----------------------------------|-----------------|---------------------|------------------|-------------|
| CPU | 8 | 8 | 16 | 16 |
| ICF | 0 | 12 | 12 | 12 |
| IFL | 8 | 40 | 48 | 48 |
| zIIP | 1 | 15 | 16 | 16 |
| CBP | 0 | 0 | 0 | 0 |
| SAP | 16 | 0 | 16 | 16 |
| Available to use | | 0 | 0 | |
| Remaining processor drawer Totals | 33 | 90 | 123 | |

At the bottom of the window are "Cancel" and "Help" buttons.

Figure 9-6 Reassign Non-Dedicated Processors results

Important: Consider the results of these changes relative to the operational environment. Understand the potential effect of making such operational changes. Changes to the PU assignment, although technically correct, can result in constraints for critical system images. In certain cases, the solution might be to defer the reassignments to another time that has less effect on the production system images.

After you review the reassignment results and make any necessary adjustments, click **OK**.

The final results of the reassignment, which include the changes that are made as a result of the review, are displayed (see Figure 9-7). These results are the assignments when the drawer removal phase of the EDA is completed.

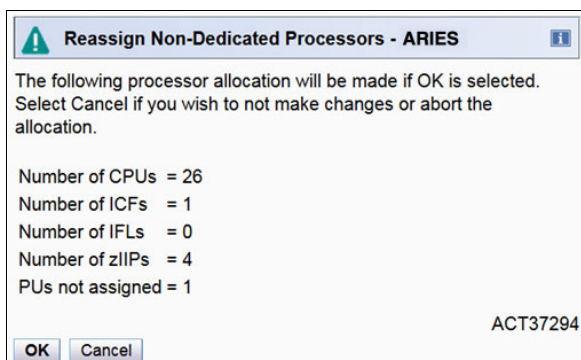


Figure 9-7 Reassign Non-Dedicated Processors, message ACT37294

Summary of the drawer removal process steps

To remove a drawer, the following resources must be moved to the remaining active drawers:

- ▶ PUs: Enough PUs must be available on the remaining active drawers, including all types of PUs that can be characterized (CPs, IFLs, ICFs, zIIPs, SAPs, and IFP).
- ▶ Memory: Enough installed memory must be available on the remaining active drawers.
- ▶ I/O connectivity: Alternative paths to other drawers must be available on the remaining active drawers, or the I/O path must be taken offline.

By understanding the system configuration and the LPAR allocation for memory, PUs, and I/O, you can make the best decision about how to free the necessary resources to allow for drawer removal.

Complete the following steps to concurrently replace a drawer:

1. Run the preparation task to determine the necessary resources.
2. Review the results.
3. Determine the actions to perform to meet the required conditions for EDA.
4. When you are ready to remove the drawer, free the resources that are indicated in the preparation steps.
5. Repeat the step that is shown in Figure 9-4 on page 399 to ensure that the required conditions are all satisfied.
6. Upon successful completion, the system is ready for the removal of the drawer.

The preparation process can be run multiple times to ensure that all conditions are met. It does not reallocate any resources; instead, it produces only a report. The resources are not reallocated until the Perform Drawer Removal process is started.

Rules during EDA

During EDA, the following rules are enforced:

- ▶ Processor rules

All processors in any remaining drawers are available to be used during EDA. This requirement includes the two spare PUs or any available PU that is non-LICCC.

The EDA process also allows conversion of one PU type to another PU type. One example is converting a zIIP to a CP during the EDA function. The preparation for the concurrent drawer replacement task indicates whether any SAPs must be moved to the remaining drawers.

- ▶ Memory rules

All physical memory that is installed in the system, including flexible memory, is available during the EDA function. Any physical installed memory, whether purchased or not, is available to be used by the EDA function.

- ▶ Single I/O rules

Alternative paths to other drawers must be available, or the I/O path must be taken offline.

Review the results. The result of the preparation task is a list of resources that must be made available before the drawer replacement can occur.

Freeing any resources

At this stage, create a plan to free these resources. The following resources and actions are necessary to free them:

- ▶ Freeing any PUs:

- Vary off the PUs by using the Perform a Model Conversion window, which reduces the number of PUs in the shared PU pool.
- Deactivate the LPARs.

- ▶ Freeing memory:

- Deactivate an LPAR.
- Vary offline a portion of the reserved (online) memory. For example, in z/OS, run the following command:

```
CONFIG_STOR(E=1),<OFFLINE/ONLINE>
```

This command enables a storage element to be taken offline. The size of the storage element depends on the RSU value. In z/OS, the following command configures offline smaller amounts of storage than the amount that was set for the storage element:

```
CONFIG_STOR(nnM),<OFFLINE/ONLINE>
```

- A combination of both LPAR deactivation and varying memory offline.

Reserved storage: If you plan to use the EDA function with z/OS LPARs, set up reserved storage and an RSU value. Use the RSU value to specify the number of storage units that are to be kept free of long-term fixed storage allocations. This configuration allows for storage elements to be varied offline.

9.9 z15 Enhanced Driver Maintenance

EDM is one more step toward reducing the necessity for and the duration of a scheduled outage. One of the components to planned outages is LIC Driver updates that are run in support of new features and functions.

When correctly configured, z15 servers support concurrently activating a selected new LIC Driver level. Concurrent activation of the selected new LIC Driver level is supported only at specific released sync points. Concurrently activating a selected new LIC Driver level anywhere in the maintenance stream is not possible. Certain LIC updates do not allow a concurrent update or upgrade.

Consider the following key points about EDM:

- ▶ The HMC can query whether a system is ready for a concurrent driver upgrade.
- ▶ Previous firmware updates, which require an initial machine load (IML) of the z15 system to be activated, can block the ability to run a concurrent driver upgrade.
- ▶ An icon on the SE allows you or your IBM SSR to define the concurrent driver upgrade sync point to be used for an EDM.
- ▶ The ability to concurrently install and activate a driver can eliminate or reduce a planned outage.
- ▶ z15 servers introduce Concurrent Driver Upgrade (CDU) cloning support to other CPCs for CDU preinstallation and activation.
- ▶ Concurrent crossover from Driver level N to Driver level $N+1$, then to Driver level $N+2$, must be done serially. No composite moves are allowed.
- ▶ Disruptive upgrades are permitted at any time, and allow for a composite upgrade (Driver N to Driver $N+2$).
- ▶ Concurrently backing up to the previous driver level is not possible. The driver level must move forward to driver level $N+1$ after EDM is started. Unrecoverable errors during an update might require a scheduled outage to recover.

The EDM function does not eliminate the need for planned outages for driver-level upgrades. Upgrades might require a system level or a functional element scheduled outage to activate the new LIC. The following circumstances require a scheduled outage:

- ▶ Specific complex code changes might dictate a disruptive driver upgrade. You are alerted in advance so that you can plan for the following changes:
 - Design data or hardware initialization data fixes
 - CFCC release level change
- ▶ OSA CHPID code changes might require PCHID Vary OFF/ON to activate new code.
- ▶ Crypto code changes might require PCHID Vary OFF/ON to activate new code.

Note: zUDX clients should contact their User Defined Extensions (UDX) provider before installing Microcode Change Levels (MCLs). Any changes to Segments 2 and 3 from a previous MCL level might require a change to the client's UDX. Attempting to install an incompatible UDX at this level results in a Crypto checkstop.

9.9.1 Resource Group and native PCIe features MCLs

Microcode fixes, referred to as *individual MCLs* or *packaged in Bundles*, might be required to update the Resource Group code and the native PCIe features. Although the goal is to minimize changes or make the update process concurrent, the maintenance updates at times can require the Resource Group or the affected native PCIe to be toggled offline and online to implement the updates. The native PCIe features (managed by Resource Group code) are listed Table 9-1.

Table 9-1 Native PCIe cards for z15

| Native PCIe adapter type | Feature code | Resource required to be offline |
|--------------------------|--------------|---------------------------------|
| 25 GbE RoCE Express2.1 | 0450 | FIDs/PCHID |
| 25 GbE RoCE Express2 | 0430 | FIDs/PCHID |
| 10 GbE RoCE Express2.1 | 0432 | FIDs/PCHID |
| 10 GbE RoCE Express2 | 0412 | FIDs/PCHID |
| zHyperLink Express1.1 | 0451 | FIDs/PCHID |
| zHyperLink Express | 0431 | FIDs/PCHID |
| Coupling Express LR | 0433 | CHPIDs/PCHID |

Consider the following points for managing native PCIe adapters microcode levels:

- ▶ Updates to the Resource Group require all native PCIe adapters that are installed in that RG to be offline.
- ▶ Updates to the native PCIe adapter require the adapter to be offline. If the adapter is not defined, the MCL session automatically installs the maintenance that is related to the adapter.

The PCIe native adapters are configured with Function IDs (FIDs) and might need to be configured offline when changes to code are needed. To help alleviate the number of adapters (and FIDs) that are affected by the Resource Group code update, z15 have four Resource Groups per system (CPC).

Note: Other adapter types, such as FICON Express, OSA Express, and Crypto Express that are installed in the PCIe+ I/O drawer are not effected because they are not managed by the Resource Groups.

The front, rear, and top view of the PCIe+ I/O drawer and the Resource Group assignment by card slot are shown in Figure 9-8. All PCIe+ I/O drawers that are installed in the system feature the same Resource Group assignment.

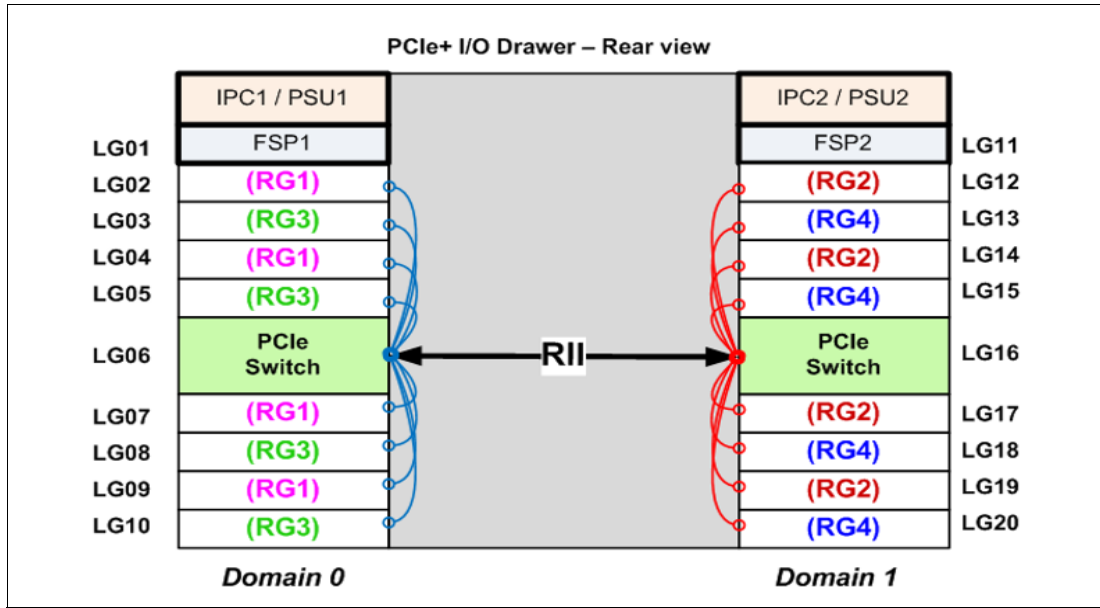


Figure 9-8 Resource Group slot assignment

The adapter locations and PCHIDs for the four Resource Groups are listed in Table 9-2.

Table 9-2 Resource Group affinity to native PCIe adapter locations

| RG | Adapter locations | PCIe+ I/O drawer | PCHIDs |
|-------------|-------------------------------|------------------|---------------------------------|
| RG1 Left | LG02, LG04, LG07, and LG09 | 1 | 100-101,108-109,110-111,118-119 |
| | | 2 | 140-141,148-149,150-151,158-159 |
| | | 3 | 180-181,188-189,190-191,198-199 |
| | | 4 | 1C0-1C1,1C8-1C9,1D0-1D1,1D8-1D9 |
| | | 5 | 200-201,208-209,210-211,218-219 |
| | | 6 | 240-241,248-249,250-251,258-259 |
| | | 7 | 280-281,288-289,290-291,298-299 |
| | | 8 | 2C0-2C1,2C8-2C9,2D0-2D1,2D8-2D9 |
| | | 9 | 300-301,308-309,310-311,318-319 |
| | | 10 | 340-341,348-349,350-351,358-359 |
| | | 11 | 380-381,388-389,390-391,398-399 |
| | | 12 | 3C0-3C1,3C8-3C9,3D0-3D1,3D8-3D9 |

| RG | Adapter locations | PCIe+ I/O drawer | PCHIDs |
|--------------|-------------------------------|------------------|---------------------------------|
| RG2 Right | LG12, LG14, LG17, and LG19 | 1 | 120-121,128-129,130-131,138-139 |
| | | 2 | 160-161,168-169,170-171,178-179 |
| | | 3 | 1A0-1A1,1A8-1A9,1B0-1B1,1B8-1B9 |
| | | 4 | 1E0-1E1,1E8-1E9,1F0-1F1,1F8-1F9 |
| | | 5 | 220-221,228-229,230-231,238-239 |
| | | 6 | 260-261,268-269,270-271,278-279 |
| | | 7 | 2A0-2A1,2A8-2A9,2B0-2B1,2B8-2B9 |
| | | 8 | 2E0-2E1,2E8-2E9,2F0-2F1,2F8-2F9 |
| | | 9 | 320-321,328-329,330-331,338-339 |
| | | 10 | 360-361,368-369,370-371,378-379 |
| | | 11 | 3A0-3A1,3A8-3A9,3B0-3B1,3B8-3B9 |
| | | 12 | 3E0-3E1,3E8-3E9,3F0-3F1,3F8-3F9 |
| RG3 Left | LG03, LG05, LG08, and LG10 | 1 | 104-105,10C-10D,114-115,11C-11D |
| | | 2 | 144-145,14C-14D,154-155,15C-15D |
| | | 3 | 184-185,18C-18D,194-195,19C-19D |
| | | 4 | 1C4-1C5,1CC-1CD,1D4-1D5,1DC-1DD |
| | | 5 | 204-205,20C-20D,214-215,21C-21D |
| | | 6 | 244-245,24C-24D,254-255,25C-25D |
| | | 7 | 284-285,28C-28D,294-295,29C-29D |
| | | 8 | 2C4-2C5,2CC-2CD,2D4-2D5,2DC-2DD |
| | | 9 | 304-305,30C-30D,314-315,31C-31D |
| | | 10 | 344-345,34C-34D,354-355,35C-35D |
| | | 11 | 384-385,38C-38D,394-395,39C-39D |
| | | 12 | 3C4-3C5,3CC-3CD,3D4-3D5,3DC-3DD |

| RG | Adapter locations | PCIe+ I/O drawer | PCHIDs |
|--------------|-------------------------------|------------------|---------------------------------|
| RG4 Right | LG13, LG15, LG18, and LG20 | 1 | 124-125,12C-12D,134-135,13C-13D |
| | | 2 | 164-165,16C-16D,174-175,17C-17D |
| | | 3 | 1A4-1A5,1AC-1AD,1B4-1B5,1BC-1BD |
| | | 4 | 1E4-1E5,1EC-1ED,1F4-1F5,1FC-1FD |
| | | 5 | 224-225,22C-22D,234-235,23C-23D |
| | | 6 | 264-265,26C-26D,274-275,27C-27D |
| | | 7 | 2A4-2A5,2AC-2AD,2B4-2B5,2BC-2BD |
| | | 8 | 2E4-2E5,2EC-2ED,2F4-2F5,2FC-2FD |
| | | 9 | 324-325,32C-32D,334-335,33C-33D |
| | | 10 | 364-365,36C-36D,374-375,37C-37D |
| | | 11 | 3A4-3A5,3AC-3AD,3B4-3B5,3BC-3BD |
| | | 12 | 3E4-3E5,3EC-3ED,3F4-3F5,3FC-3FD |

9.10 RAS capability for the HMC and SE

The HMC and the SE include the following RAS capabilities:

- ▶ Back up from HMC and SE

For the customers who do not have an FTP server that is defined for backups, the HMC can be configured as an FTP server.

On a scheduled basis, the HMC hard disk drive (HDD) is backed up to the USB flash memory drive (UFD), a defined FTP server, or both.

SE HDDs are backed up on to the primary SE HDD and an alternative SE HDD. In addition, you can save the backup to a defined FTP server.

For more information, see 10.2, “HMC and SE changes and new features” on page 413.

- ▶ Remote Support Facility (RSF)

The HMC RSF provides the important communication to a centralized IBM support network for hardware problem reporting and service. For more information, see 10.4, “Remote Support Facility” on page 429.

- ▶ Microcode Change Level (MCL)

Regular installation of MCLs is key for RAS, optimal performance, and new functions. Generally, plan to install MCLs quarterly at a minimum. Review hiper MCLs continuously. You must decide whether to wait for the next scheduled apply session, or schedule one earlier if your risk assessment of the new hiper MCLs warrants.

For more information, see 10.5.4, “HMC and SE microcode” on page 435.

► SE

z15 servers are provided with two 1U trusted servers inside the IBM Z server A frame: One is always the primary SE and the other is the alternative SE. The primary SE is the active SE. The alternative acts as the backup. Information is mirrored once per day. The SE servers include N+1 redundant power supplies.

For more information, see 10.2.3, “New Support Element” on page 417.



Hardware Management Console and Support Element

The Hardware Management Console (HMC) supports the functions and tasks that are required to manage the IBM Z CPCs. When tasks are performed on the HMC, the commands are sent to the primary Support Element (SE) of the targeted system, which then issues commands to their respective central processor complex (CPC).

This chapter describes the newest elements for the HMC and SE.

Note: The Help function is a good starting point to get more information about all of the functions that can be used by the HMC and SE. The Help feature is available by clicking **Help** from a drop-down menu that appears when you click your user ID.

For more information, see [IBM Knowledge Center](#).

This chapter includes the following topics:

- ▶ 10.1, “HMC and SE introduction” on page 412
- ▶ 10.2, “HMC and SE changes and new features” on page 413
- ▶ 10.3, “HMC and SE connectivity” on page 421
- ▶ 10.4, “Remote Support Facility” on page 429
- ▶ 10.5, “HMC and SE capabilities” on page 432

Note: Throughout this chapter, *z15* refers to IBM z15 Model T01 (Machine Type 8651), unless otherwise specified.

10.1 HMC and SE introduction

The HMC is a closed system (appliance), which means that no other applications can be installed on it.

The HMC runs a set of management applications. On z15, the HMC can be a stand-alone computer (mini-tower or rack mounted), or (new with z15) can run (as a Hardware Management Appliance) on the SEs hardware (1U rack-mounted servers that are integrated in z15 A frame).

The SEs are two 1U servers integral to the z15 frame. One SE is the primary SE (active) and the other is the alternative SE (backup). As with the HMCs, the SEs are closed systems, and no other applications can be installed on them.

The HMC is used to set up, manage, monitor, and operate one or more CPCs. It manages IBM Z hardware, its logical partitions (LPARs), and provides support applications. At least one HMC is required to operate an IBM Z. An HMC can manage multiple Z CPCs, and can be at a local or a remote site.

When tasks are performed at the HMC, the commands are routed to the active SE of the z15. The SE then issues those commands to their CPC. One HMC can control up to 100 SEs and 1 SE can be controlled by up to 32 HMCs.

New with z15, a number of “traditional” SE-only functions moved to HMC tasks. On z15, these functions appear as native HMC tasks, but run on the SE. These HMC functions run in parallel with Single Object Operations (SOOs), which simplifies and streamlines system management. For more information about SOOs, see “Single Object Operations” on page 431.

10.1.1 Dynamic Partition Manager support

With Driver 27 (Version 2.13.1), the IBM Dynamic Partition Manager (DPM) was introduced for CPCs that are running Linux only with Fibre Channel Protocol (FCP) attached storage. HMC Driver 32 (Version 2.14.0) with MCLs added support for ECKD FICON disks to the DPM (Release 3.1).

HMC 2.14.1 includes DPM 3.2, with enhanced storage management capabilities. HMC driver 41 (Version 2.15.0) with MCLs added support for FCP and FICON support to the DPM Release 4.0. DPM is a mode of operation that enables customers with little or no knowledge of IBM Z technology to set up the system efficiently and with ease.

For more information, see [IBM Knowledge Center](#), click the search engine window, and enter DPM.

The HMC Remote Support Facility (RSF) provides an important communication to a centralized IBM support network for hardware problem reporting and service. For more information, see 10.4, “Remote Support Facility” on page 429.

10.2 HMC and SE changes and new features

The initial release that is included with z15 is HMC application Version 2.15.0. Use the “What’s New” task to examine the new features that are available for each release. For more information about HMC and SE functions, use the HMC and SE (Version 2.15.0) console help system or see [IBM Knowledge Center](#).

At IBM Knowledge Center, search for “z15 HMC”.

10.2.1 Driver Level 41 HMC and SE new features

The following support was added with Driver 41:

- ▶ Hardware Management Appliance

Before z15, in addition to the two integrated rack-mounted SEs, at least one external (stand-alone) HMC is needed (two HMCs are recommended for redundancy).

Starting with z15, the two 1U rack-mounted SEs increased hardware capacity (processor, memory), which allows virtual instances of both HMC and SE to run collocated on the same physical appliance (SE server). The SE application (appliance code) runs as guest of the Hardware Management Appliance, and can be managed by the Virtual Support Element Management task, as shown in Figure 10-1. The SE interface can still be accessed by using the Single Object Operation as usual (by way of HMC web interface).

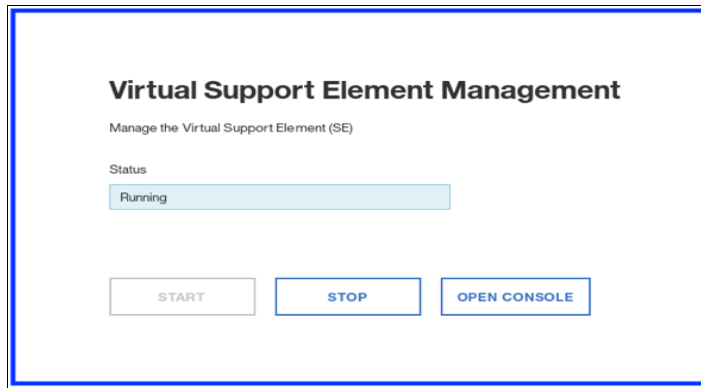


Figure 10-1 Virtual Support Element Management

The Hardware Management Appliance (HMC application) is accessible by using a remote web browser (the user experience for HMC interaction is free of charge) and can manage N-2 generations systems (z15, z14 ZR1, z14, z13s, and z13).

Important: With IBM Hardware Management appliance, shutdown or restart is disruptive to the SE appliance (the SE appliance runs as a guest). An application restart of the HMC appliance is not disruptive to the SE.

Updating the driver or applying HMC MCLs requires planning to ensure that these operations are not disruptive to the CPC by ensuring availability of the primary or alternative SE appliance.

Stand-alone (physical) HMCs are also available as rack-mounted or mini-tower.

► Integration of z/OS MFA support for RSA SecurID

HMC Version 2.15.0 provides RSA SecurID authentication by way of centralized support from IBM MFA for z/OS. The MFA policy is defined in RACF and assigned to RACF user IDs, and the RSA SecurID passcode is verified by RSA authentication server.

User Management task is changed on User Definition and User Template Definition to define and select the MFA Server, map HMC user ID to RACF user ID, and select the RACF policy.

► System Recovery Boost status (observe)

System Recovery Boost helps clients improve SLAs at lower risk over service disruptions, both planned and unplanned.

System Recovery boost delivers increased processor capacity for the boost period at the beginning of a planned shutdown or following an IPL so client workloads start can be accelerated and catch up with work through a backlog after the downtime.

Boost capacity can be provided by using one of the following options:

- On a subcapacity machine, the CPs that are allocated to the opt-in LPAR are converted to full-speed CPs during the boost period.
- Dispatching general processor (CP) workloads to zIIPs during the boost period.
- Turbo feature for System Recovery Boost, which is a temporary capacity record that adds physical zIIP capacity to the machine by activating uncharacterized PUs on the system. These added zIIPs can perform CP work as in the second option. This option requires purchase of a prepaid (priced) feature (FC 6802) and is covered by contract terms and conditions (FC 9930 - Boost Authorization).

Processor boost status (usage) is shown on the partition image details (from HMC and SE), as shown in Figure 10-2. System Recovery Boost can be “On” during the boost period (30 minutes for shutdown 60 minutes for IPL), or “Off” during normal LPAR operation. The software image that is running in the LPAR can opt in or opt out the boost facility.

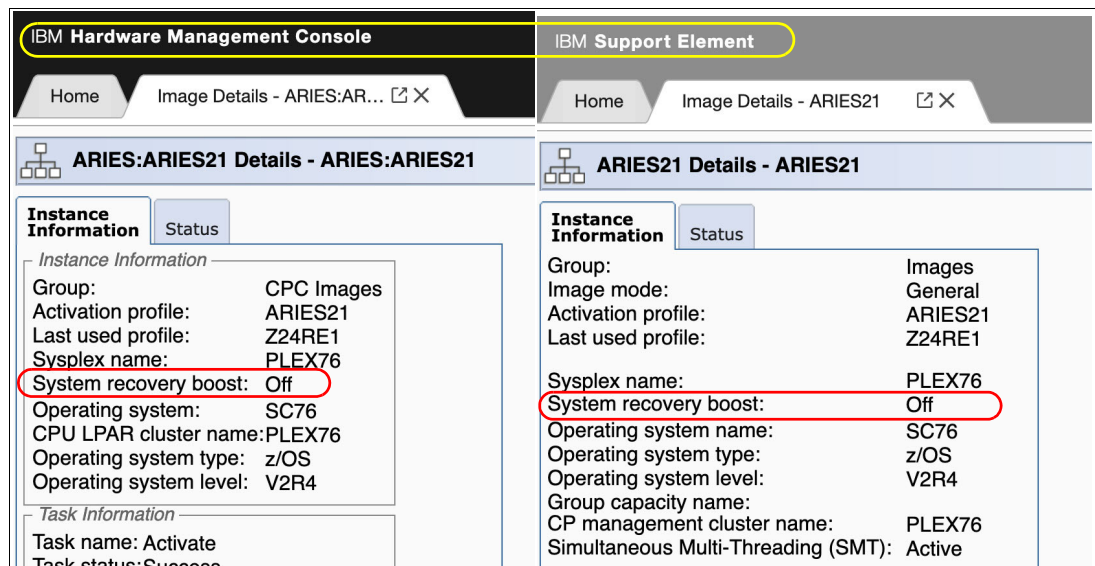


Figure 10-2 Boost status observe from SE

► Change LPAR Group Controls scheduled operation

HMC/SE Version 2.15.0 added a Change LPAR Group Controls scheduled operation (see Figure 10-3) to allow scheduling the change of the CPs/zIIPs/ICFs/IFLs absolute capping status for an LPAR Group.

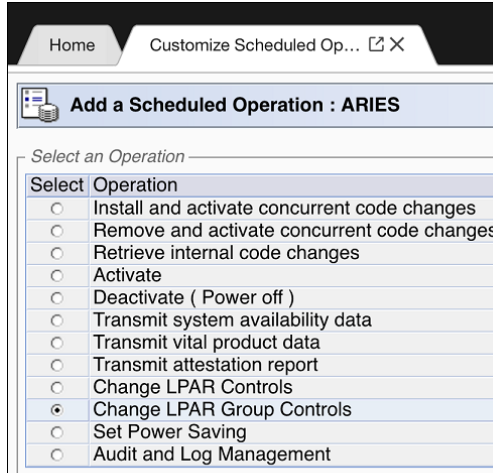


Figure 10-3 Scheduled Operations for Change LPAR Group Controls

► Linux Secure IPL is available for LPARs running on z15 CPCs

This new feature checks the software signature before loading the code into the LPAR storage (memory). When Linux Secure IPL is enabled, the signature of the operating system being loaded is compared with the signature from the Linux on Z distribution provider. The load fails if the signatures do not match.

The feature can be enabled by the selecting the **Verify software signature with distributor** option in the image profile load Load task, as shown in Figure 10-4. This feature supports Linux on Z running in an LPAR on a z15 CPC.

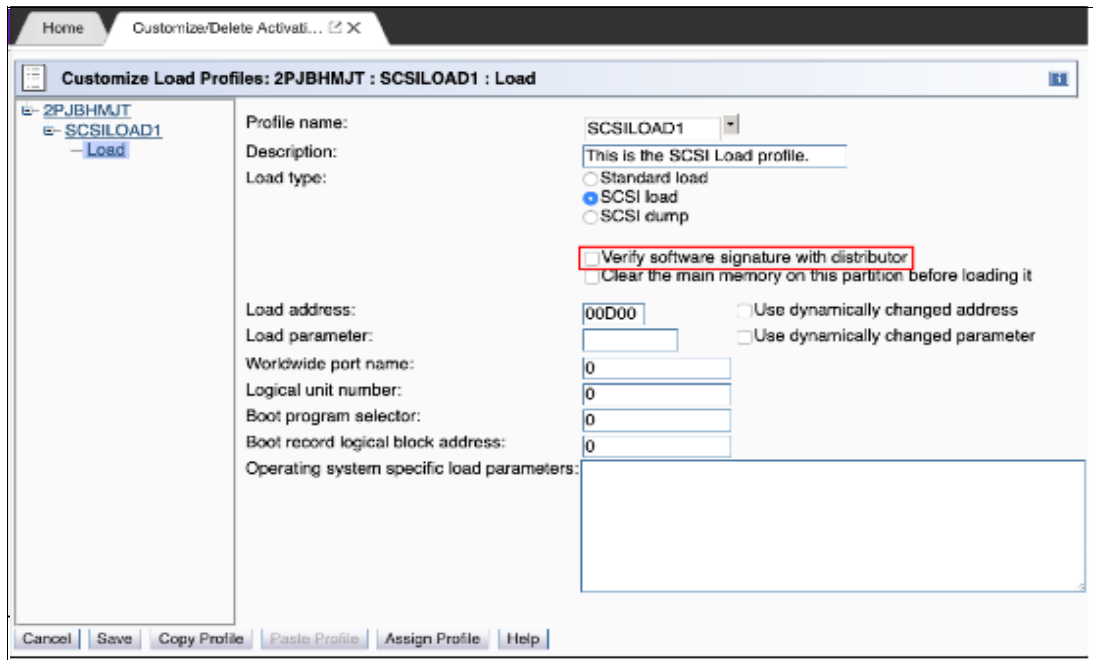


Figure 10-4 Customize load profile

- ▶ Integrated 3270 console security and performance enhancements

On HMC 2.15.0, the internal framework was reworked; therefore, performance for integrated 3270 console is improved. Also, RACF security checking is required based on user logon.

Note: Consider the following points:

- ▶ Starting with HMC Version 2.15.0, zBX is no longer supported.
- ▶ With HMC/SE Driver 41, the Sysplex Time task was removed from the SE; as such, an HMC version 2.15.0 is required to manage STP on a z15 CPC.
- ▶ HMC 2.15.0 supports managing N-2 generation CPCs only (z15, z14 ZR1, z14, z13s, and z13).
- ▶ The DVD drives for the z15 hardware are no longer available on the new build HMC or the SE.

10.2.2 New Rack-mounted HMC and Tower HMC

Feature code (FC) 0063 provides rack-mounted HMC, and FC 0062 provides the tower version of the stand-alone HMC.

The HMC FC0063 (2461-SE3) is a 1U server that can be ordered with an optional IBM 1U rack-mounted tray that features a monitor and a keyboard/pointing device (KMM FC 0154). The HMC system unit and the KMM tray must be rack-mounted in two adjacent 1U locations in the “ergonomic zone” between 21U and 26U in a standard 19-inch rack.

The rack-mounted HMC can be installed in a customer-provided rack (it cannot be mounted in the z15 CPC frames). Three C13 power receptacles are required: two for the system unit and one for the display and keyboard, as shown in Figure 10-5.

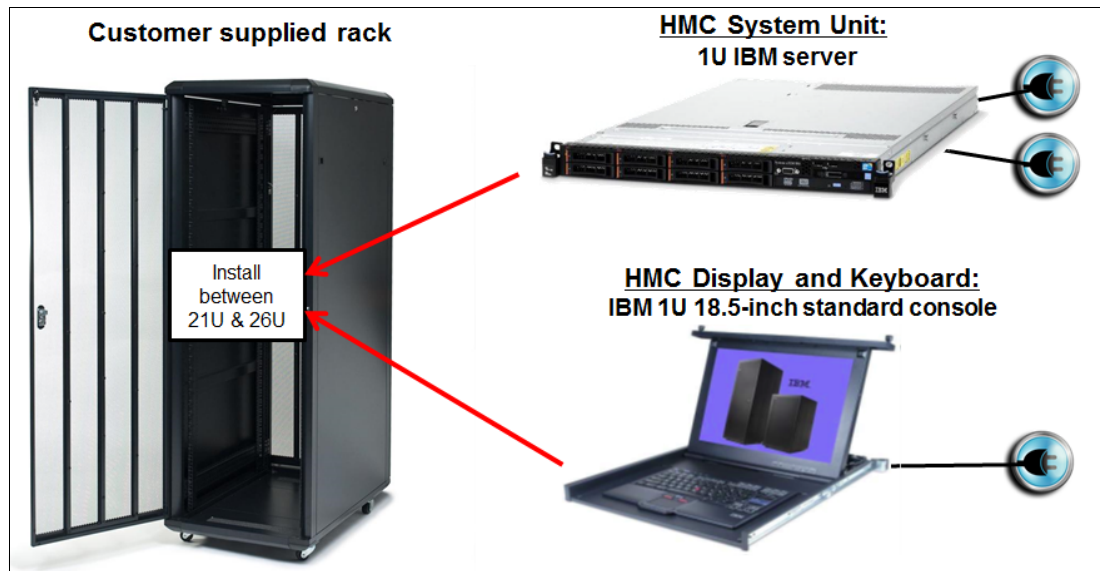


Figure 10-5 Rack-mounted HMC installed in an extra rack

10.2.3 New Support Element

In z15, the SEs are two 1U rack-mounted hardware appliances that are installed at the top of the z15 A-frame, as shown in Figure 10-6.



Figure 10-6 SEs location

Support Element Keyboard Mouse Monitor

New with z15, a new Keyboard Mouse Monitor (KMM) device replaces the previous KMM assembly that was mounted on a swing gate. Consider the following points:

- ▶ The device is intended to be used by service personnel only.
- ▶ One KMM is used.
- ▶ KMM is stored in a cubby at the front of the A-frame, just below the SEs at EIA Rack Unit 39. A USB-C cable and mounting bracket are stored together with the KMM.
- ▶ A cable is used to plug the device into a KVM switch at the front or the rear of the rack when servicing the system.
- ▶ Switching between SEs is done by using a button that is on the KVM. It also indicates which SE is selected (see Figure 10-7 on page 418).
- ▶ The KMM mounting bracket can be used to mount the device to any frame in the system (front or rear sides).
- ▶ The KMM can be used on any z15 system (no affinity to system with which it is shipped).

The SE KMM device is shown in Figure 10-7.

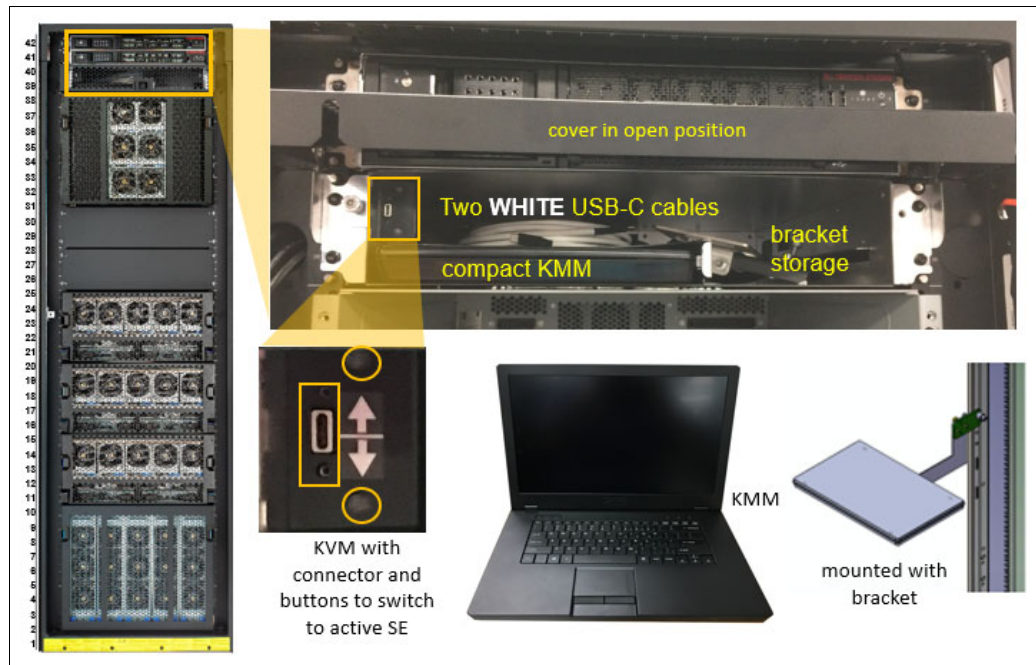


Figure 10-7 Support Element KMM device

For more information about the mini-KMM and how to attach it to the A frame, see *IBM Z 8561 Installation Manual for Physical Planning, GC28-7002*.

10.2.4 New service and functional operations for HMCs and SEs

Because a DVD drive is not available on the HMC or the SE, this section describes some service and functional operations for HMC Version 2.15.0.

Firmware load

Firmware can be loaded by using the following modes:

- ▶ USB

If the HMC and SE firmware is shipped on a USB drive when a new system is ordered, the load procedure is similar that was used with a DVD load.

- ▶ Electronic

If USB load is not allowed, or when FC 0846 is ordered, an ISO image is used for a firmware load over a local area network (LAN). New build HMCs include HMC/SE ISO images and the HMC provides the server function for loading the code. ISO images can also be downloaded through zRSF or FTP from IBM.

Important: The ISO image server (HMC) *must* be on the same subnet with the target system; that is, the system to be loaded with HMC or SE code from ISO images.

Operating system load from removable media or server

z/OS, z/VM, z/VSE, and Linux on Z are planned for USB/network distribution. z/TPF does not use the HMC for code load.

10.2.5 SE driver support with the HMC driver

The driver of the HMC and SE is equivalent to a specific HMC and SE version:

- ▶ Driver 22 is equivalent to HMC Version 2.13.0
- ▶ Driver 27 is equivalent to HMC Version 2.13.1
- ▶ Driver 32 is equivalent to HMC Version 2.14.0
- ▶ Driver 36 is equivalent to HMC Version 2.14.1
- ▶ Driver 41 is equivalent to HMC Version 2.15.0

An HMC with Version 2.15.0 can support N-2 IBM Z server generations. Some functions that are available on Version 2.15.0 and later are supported only when the HMC is connected to an IBM Z with Driver 41 (z15).

The SE drivers and versions that are supported by the z15 HMC Version 2.15.0 (Driver 41) and earlier versions are listed in Table 10-1.

Table 10-1 Summary of SE drivers

| IBM Z family name | Machine type | SE driver | HMC/SE version | Ensemble node potential |
|-------------------|--------------|-----------|---------------------|-------------------------|
| z15 | 8561 | 41 | 2.15.0 ^a | No |
| z14 ZR1 | 3907 | 32, 36 | 2.14.0, 2.14.1 | Yes ^b |
| z14 | 3906 | 32, 36 | 2.14.0, 2.14.1 | Yes ^b |
| z13s | 2965 | 27 | 2.13.1 | Yes ^b |
| z13 | 2964 | 22, 27 | 2.13.0, 2.13.1 | Yes ^a |

a. HMC 2.15.0 cannot be used to manage ensembles.

b. A CPC in DPM mode cannot be a member of an ensemble; however, the CPC can still be managed by the ensemble HMC (z14 and earlier systems only).

10.2.6 HMC feature codes

HMCs that are earlier than FC 0095 are not supported for z15 at Driver 41.

The following HMC feature codes are available for a new order:

- ▶ FC 0062: M/T 2461-TW3

This feature is the new tower HMC that supports z15, z14 ZR1, z14, z13, and z13s systems.

- ▶ FC 0063: M/T 2461-SE3

This feature is the new rack-mounted HMC that supports z15, z14 ZR1, z14, z13, and z13s systems.

The following older HMCs can be carried forward (the carry forward HMCs do not provide all enhancements that are available with FC 0062 and FC 0063):

- ▶ Tower FC 0082
- ▶ Tower FC 0095
- ▶ 1U Rack FC 0083
- ▶ 1U Rack FC 0096

10.2.7 User interface

Starting with HMC Version 2.15.0, HMC Dashboard status was enhanced for accessibility.

In z14, the status bar is in Home tab, and the HMC tasks are started in tabs. While working in a task tab, the console status is not visible. As such, the user cannot be notified of hardware or operating system messages, and other unacceptable status messages until the current task is closed and user returns to the Home tab.

In z15, the status bar was moved to the masthead of the HMC interface, which is always visible (even when working in a task tab), as shown in Figure 10-8.



Figure 10-8 Dashboard status enhancement

10.2.8 Customize Product Engineering Access: Best practice

At times, the HMC or the SE must be accessed in a support role to perform problem determination tasks.

The task to authorize IBM Product Engineering access to the console is shown in Figure 10-9. When access is authorized, an IBM product engineer can use an exclusive user ID and reserved password to log on to the console for problem determination actions.

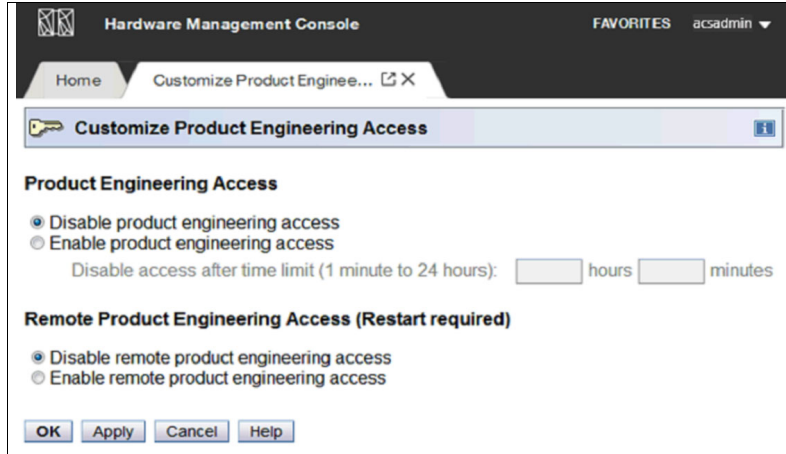


Figure 10-9 Customize Product Engineering Access tab

As shown in Figure 10-9, the task is available only to users with ACSADMIN authority. Consider the following points:

- ▶ Customers must ensure that redundant administrator users (ACSADMIN role) are available for each console.
- ▶ Customers must document contact information and procedures.
- ▶ The “Welcome Text” task can be used to identify contact information so that IBM Service personnel are informed about how to engage customer administrators if HMC or SE access is needed.
- ▶ The Product Engineering access options are disabled by default.

10.3 HMC and SE connectivity

The *standard* (stand-alone) HMC feature s two Ethernet adapters that enable connectivity to two distinct Ethernet LANs: one for communicating to the support elements, and one for client access through a web browser.

The for CPC management, the SEs on z15 are connected to the Ethernet switches that are installed at the top of the z15 rack, under the SEs. In previous IBM Z systems, the customer network was connected directly to the bulk power hub (BPH). Now, the SEs are directly connected to the customer network by using distinct (separate) Ethernet ports than the SE Ethernet ports that are used for internal CPC management.

10.3.1 Standard HMC connectivity

The HMC communicates with the SE through a customer-supplied Ethernet switch (two switches are recommended for redundancy) that is connected to the J03 or J04 ports on the SEs. Other IBM Z systems and HMCs also can be connected to the same switch (set of switches). Standard HMC connectivity is shown in Figure 10-10.

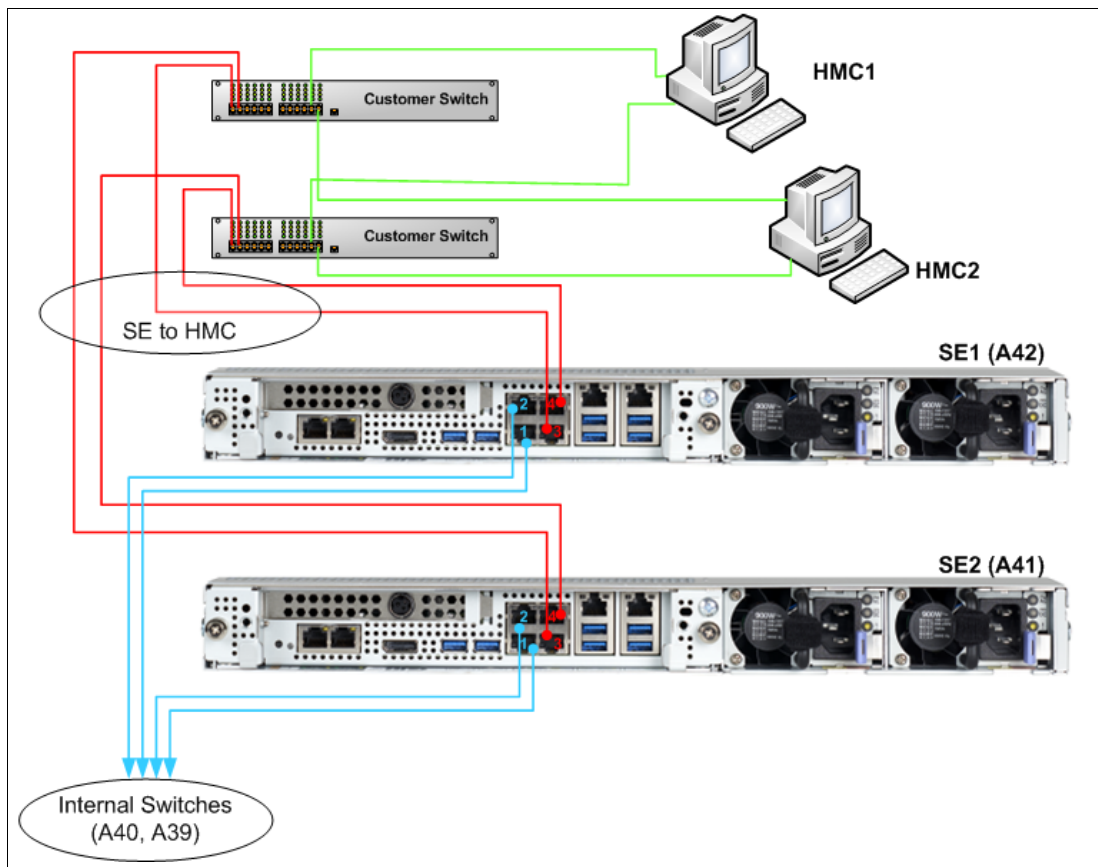


Figure 10-10 Standard HMC connectivity

Note: The HMC *must* be connected to the SEs by using a switch. Direct connection between the HMC and the SEs is *not* supported.

10.3.2 Hardware Management Appliance

Starting with z15, customers have a choice of the standard (physical) HMC or the new Hardware Management Appliance (FC 0100). Both options can be used to manage the CPCs.

With FC 0100, HMCs and SEs are packaged redundantly inside the Z CPC frame, which eliminates the need for managing separate HMC hardware appliances outside of the z15 CPC.

SE redundant physical appliances use increased capacity devices (1U Rack Mounted). The SE physical appliances run virtual instances of HMC and SE on each physical appliance. This configuration provides redundancy for the HMC and SE.

Note: Consider the following points:

- ▶ Although Hardware Management Appliance does not require external (stand-alone HMC) hardware, it can be used in parallel with a stand-alone HMC appliance.
- ▶ The HMC code runs SE appliance code as a guest of the HMC. This setup requires more planning when updates to the HMC or SE code are required.
- ▶ HMC implemented as Hardware Management Appliance can be used to manage N-2 systems (z13/z13s and z14), not only z15.

With FC 0100, the SE hardware (2461-SE3) provides the required processing resources and networking connectivity. With Hardware Management Appliance, client access to the HMC is performed by using a browser (remote access because no HMC KMM console is required). Hardware Management Appliance connectivity is shown in Figure 10-11.

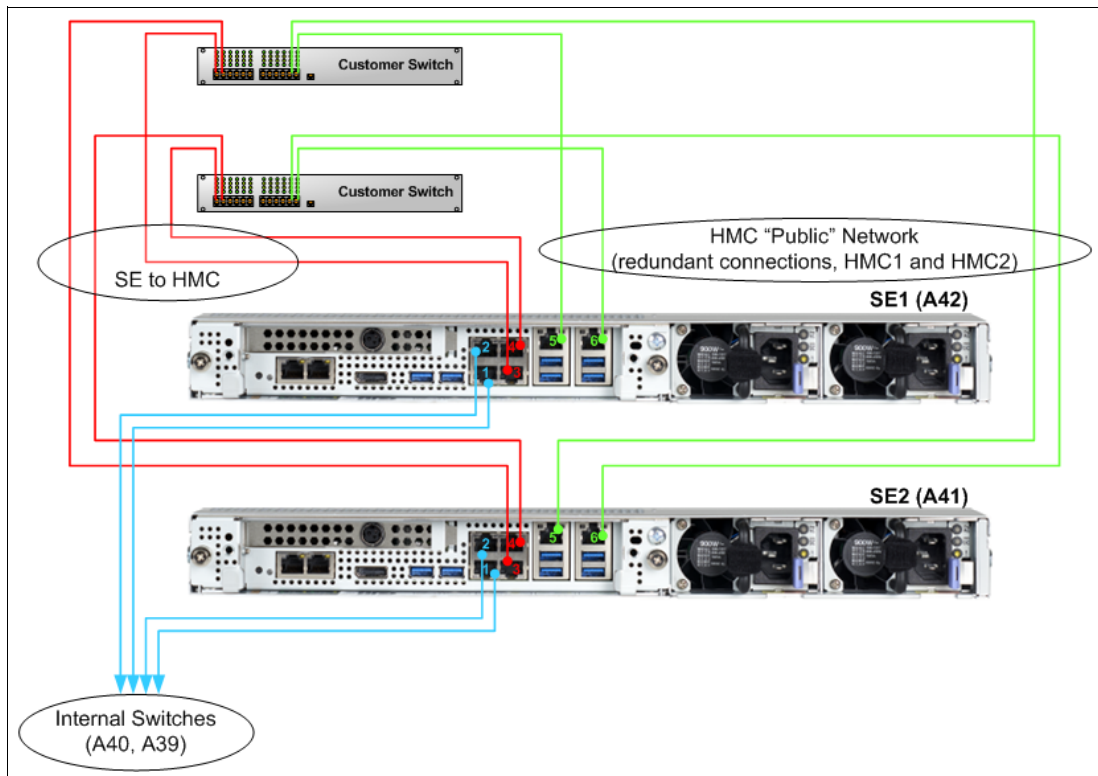


Figure 10-11 Hardware Management Appliance connectivity

The Hardware Management Appliance connectivity for multiple CPC environments (z15 N-2 only) is shown in Figure 10-12.

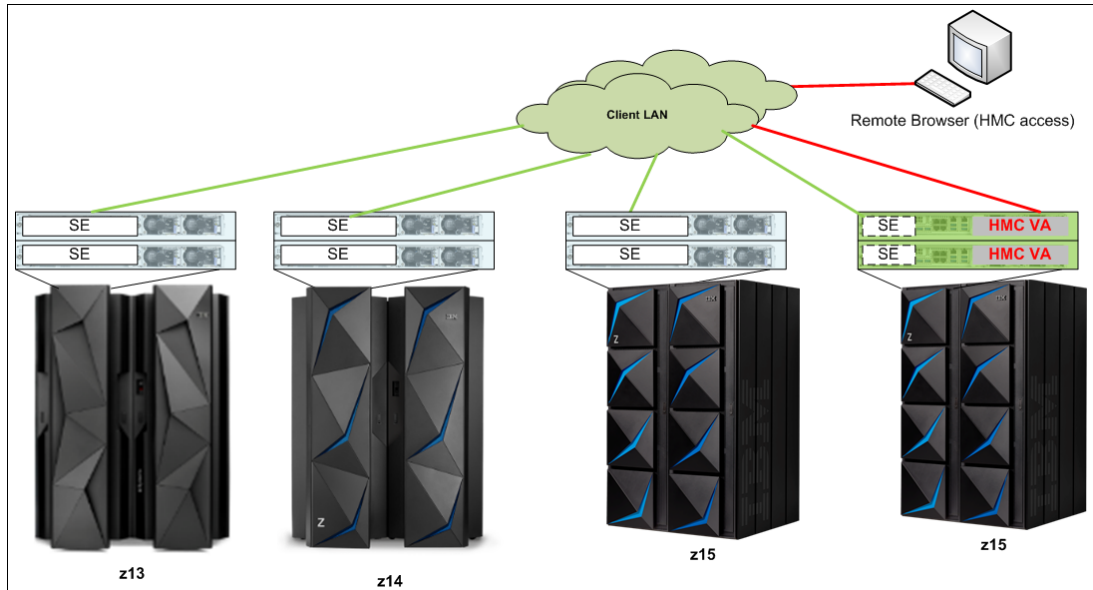


Figure 10-12 HMC virtual appliance managing multiple CPCs

Various methods are available for setting up the network. Designing and planning the HMC and SE connectivity is the clients' responsibility, based on the environment's connectivity and security requirements.

Security: The configuration of network components, such as routers or firewalls, is beyond the scope of this document. Whenever the networks are interconnected, security exposures can exist. For more information about HMC security, see *Integrating the Hardware Management Console's Broadband Remote Support Facility into your Enterprise*, SC28-6951.

For more information about the HMC settings that are related to access and security, see the HMC and SE console help function or [IBM Knowledge Center](#).

10.3.3 Network planning for the HMC and SE

Plan the HMC and SE network connectivity carefully to allow for current and future use. Many of the IBM Z capabilities benefit from the various network connectivity options that are available. The following functions, which depend on the HMC connectivity, are available to the HMC:

- ▶ Lightweight Directory Access Protocol (LDAP) support, which can be used for HMC user authentication
- ▶ Network Time Protocol (NTP) support
- ▶ RSF through broadband
- ▶ HMC remote access and HMC Mobile
- ▶ RSA SecurID support
- ▶ Enablement of the SNMP and CIM¹ APIs to support automation or management applications, such as IBM System Director Active Energy Manager (AEM).

HMC File Transfer support

FTP, FTPS, and SFTP protocols are now supported on the HMC and SE. All three file transfer protocols (applications) require login ID and password (credentials).

FTPS is based on Secure Sockets Layer cryptographic protocol (SSL) and requires certificates to authenticate the servers. SFTP is based on Secure Shell protocol (SSH) and requires SSH keys to authenticate the servers. Certificates and key pairs are hosted on the z15 HMC Console.

The recommended network topology for HMC, SE, and FTP server is shown in Figure 10-13.

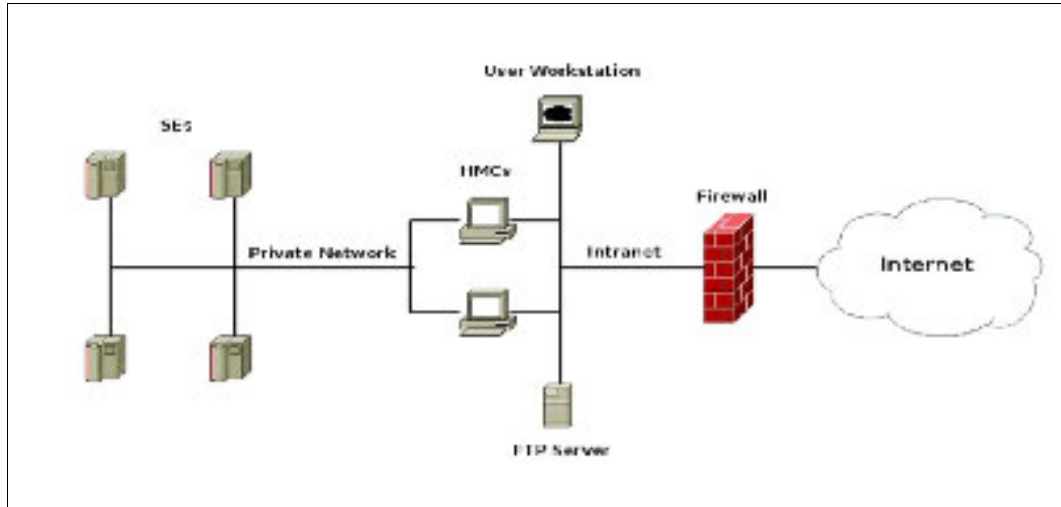


Figure 10-13 Recommended Network Topology for HMC, SE, and FTP server

The following FTP server requirements must be met:

- ▶ Support “passive” data connections
- ▶ A server configuration that allows the client to connect on an ephemeral port

The following FTPS server requirements must be met:

- ▶ Operate in “explicit” mode
- ▶ Allows a server to offer secure and unsecured connections
- ▶ Must support “passive” data connections
- ▶ Must support secure data connections

The SFTP server must support password-based authentication.

¹ CIM support was removed from the HMC with Version 2.14.0.

The file transfer server choices for HMC are shown in Figure 10-14.

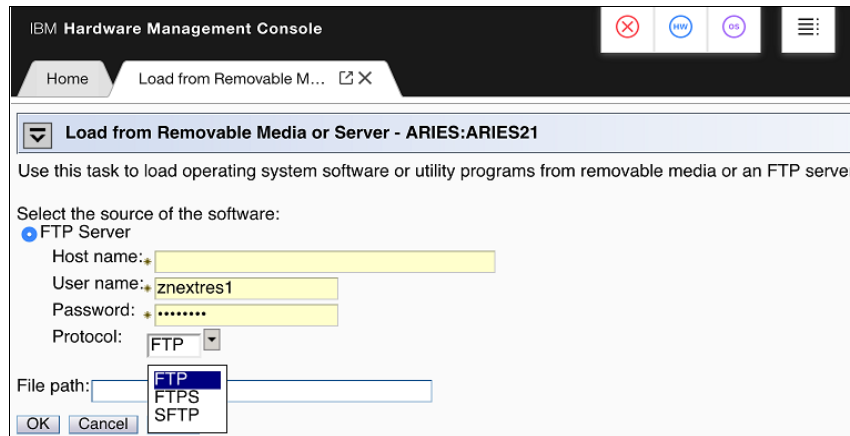


Figure 10-14 FTP protocols drop-down list

FTP through HMC

It is highly recommended to keep IBM Z systems, HMC consoles, and SEs on an isolated network. This approach prevents SEs initiating FTP connections with outside networks and applies to all supported file transfer protocols (FTP, FTPS, and SFTP).

With z15 HMC, all FTP connections that originate from the SEs are taken to HMC consoles. Secure FTP server credentials must be imported to one or more managing HMC consoles.

After the HMC console completes all FTP operations, the HMC console performs the FTP operation on SE's behalf and returns the results. The IBM Z platform must be managed by at least one HMC to allow FTP operations.

Secure console-to-console communications

Before the z14 server generation, the HMC consoles used anonymous cipher suites to establish console-to-console communication. Anonymous cipher suite is a part of SSL/TLS protocol and it can be used to create point-to-point connections. Anonymous cipher suite does not exchange certificates, which can be a security exposure.

Similar to z14, the z15 HMC consoles abandon anonymous cipher suite and implement an industry standard-based, password-driven cryptography system. The Domain Security Settings are used to provide authentication and high-quality encryption. Because of these changes, we now recommend that clients use unique Domain Security settings to provide maximum security. The new system provides greater security than anonymous cipher suites, even if the default settings are used.

To allow greater flexibility in password selection, the password limit was increased to 64 characters and special characters are allowed for z15 and z14 installations. If communication with older systems before z14 is needed, the previous password limits must be followed (6 - 8 characters, only uppercase and number characters allowed).

For more information about HMC networks, see the following resources:

- ▶ The HMC and SE (Version 2.15.0) console help system, or [IBM Knowledge Center](#)
- ▶ *IBM Z 8561 Installation Manual for Physical Planning*, GC28-7002

10.3.4 Hardware considerations

The following HMC changes are important for z15:

- ▶ No DVD/CD drive with HMC
- ▶ IBM does not provide Ethernet switches with the system
- ▶ RSF is broadband-only

DVD/CD Drive

Starting with z15, a DVD/CD drive is not available with the HMC (nor with the SE).

Ethernet switches

Ethernet switches for HMC and SE connectivity are provided by the client. Existing supported switches can still be used.

Ethernet switches and hubs often include the following characteristics:

- ▶ A total of 16 auto-negotiation ports
- ▶ 100/1000 Mbps data rate
- ▶ Full or half duplex operation
- ▶ Auto medium-dependent interface crossover (MDIX) on all ports
- ▶ Port status LEDs

Note: The recommendation is to use a switch with 1000 Mbps/Full duplex support.

RSF is broadband-only

RSF through a modem is *not* supported the z15 HMC. Broadband is needed for hardware problem reporting and service. For more information, see 10.4, “Remote Support Facility” on page 429.

10.3.5 TCP/IP Version 6 on the HMC and SE

The HMC and SE can communicate by using IPv4, IPv6, or both. Assigning a static IP address to an SE is unnecessary if the SE communicates only with the HMCs on the same subnet. The HMC and SE can use IPv6 link-local addresses to communicate with each other.

IPv6 link-local addresses feature the following characteristics:

- ▶ Every IPv6 network interface is assigned a link-local IP address.
- ▶ A link-local address is used on a single link (subnet) only and is never routed.
- ▶ Two IPv6-capable hosts on a subnet can communicate by using link-local addresses, without having any other IP addresses assigned.

10.3.6 OSA Support Facility

Since OSA/SF was moved from z/OS to HMC/SE environment, it was noted that it is no longer easy to obtain a global view of all OSA PCHIDs and the monitoring and diagnostic information that was available in the Query Host command.

To address this issue, the following changes were made:

- ▶ If a CPC is targeted, the initial window provides a global view of all OSA PCHIDs.
- ▶ The user can browse to various OSA Advanced Facilities subtasks from the initial window, which makes the process of getting to them less cumbersome.

- ▶ Today's View Port Parameters and Display OAT entries support exporting data of one OSA PCHID. Also, the data for all OSA PCHIDs can be exported to USB or FTP from the View Port Parameters menu.
- ▶ The initial window was changed to display status information of all OSA PCHIDs (see Figure 10-15).

| Select | PCHID | Hardware Type | Status | CHPID Type | Code Level | Port 0 Status | Port 0 MAC Address | Port 1 Status | Port 1 MAC Address |
|----------------------------------|-------|-----------------------------------|-----------|------------|------------|---------------|--------------------|---------------|--------------------|
| <input checked="" type="radio"/> | 0110 | OSA-Express7S 10Gb SR Ethernet | Operating | OSD | 0142 | Enabled | AC620D090B36 | | |
| <input type="radio"/> | 0114 | OSA-Express7S 1000Base-T Ethernet | Operating | OSC | 0083 | Enabled | AC620D0909E8 | Enabled | AC620D0909E9 |
| <input type="radio"/> | 011C | OSA-Express6S 1000Base-T Ethernet | Operating | OSC | 0089 | Enabled | 98BE9479D6E4 | Enabled | 98BE9479D6E5 |
| <input type="radio"/> | 0128 | OSA-Express7S 10Gb SR Ethernet | Operating | OSD | 0142 | Enabled | AC620D090B18 | | |
| <input type="radio"/> | 0130 | OSA-Express7S 10Gb SR Ethernet | Operating | OSD | 0142 | Enabled | AC620D09093C | | |
| <input type="radio"/> | 0134 | OSA-Express7S 1000Base-T Ethernet | Operating | OSD | 0142 | Enabled | AC620D090AF0 | Enabled | AC620D090AF1 |
| <input type="radio"/> | 0138 | OSA-Express6S 10Gb SR Ethernet | Operating | OSD | 0142 | Enabled | 98BE9479FBF0 | | |
| <input type="radio"/> | 0150 | OSA-Express7S 10Gb SR Ethernet | Operating | OSD | 0142 | Enabled | AC620D090AB6 | | |
| <input type="radio"/> | 0154 | OSA-Express7S 10Gb SR Ethernet | Operating | OSD | 0142 | Enabled | AC620D090B2C | | |
| <input type="radio"/> | 015C | OSA-Express7S 1000Base-T Ethernet | Operating | OSD | 0142 | Enabled | AC620D090966 | Enabled | AC620D090967 |
| <input type="radio"/> | 016C | OSA-Express7S 10Gb SR Ethernet | Operating | OSD | 0142 | Enabled | AC620D090A02 | | |
| <input type="radio"/> | 0174 | OSA-Express7S 10Gb SR Ethernet | Operating | OSD | 0142 | Enabled | AC620D090B40 | | |

Total: 28 Filtered: 28 Selected: 1

Figure 10-15 OSA Advanced Facilities window

10.3.7 Assigning addresses to the HMC and SE

An HMC can have the following IP configurations:

- ▶ Statically assigned IPv4 or statically assigned IPv6 addresses
- ▶ Dynamic Host Configuration Protocol (DHCP)-assigned IPv4 or DHCP-assigned IPv6 addresses
- ▶ Auto-configured IPv6:
 - Link-local is assigned to every network interface.
 - Router-advertised, which is broadcast from the router, can be combined with a Media Access Control (MAC) address to create a unique address.
 - Privacy extensions can be enabled for these addresses as a way to avoid the use of the MAC address as part of the address to ensure uniqueness.

An SE can have the following IP addresses:

- ▶ Statically assigned IPv4 or statically assigned IPv6
- ▶ Auto-configured IPv6 as link-local or router-advertised

IP addresses on the SE cannot be dynamically assigned through DHCP to ensure repeatable address assignments. DHCP privacy extensions are not used on the SE.

The HMC uses IPv4 and IPv6 multicasting² to automatically discover the SEs. The HMC Network Diagnostic Information task can be used to identify the IP addresses (IPv4 and IPv6) which are used by the HMC to communicate to the SEs (of a CPC).

² For a customer-supplied switch, multicast must be enabled at the switch level.

IPv6 addresses are easily identified. A fully qualified IPV6 address features 16 bytes. It is written as eight 16-bit hexadecimal blocks that are separated by colons, as shown in the following example:

```
2001:0db8:0000:0000:0202:b3ff:fe1e:8329
```

Because many IPv6 addresses are not fully qualified, shorthand notation can be used. In shorthand notation, the leading zeros can be omitted, and a series of consecutive zeros can be replaced with a double colon. The address in the previous example also can be written in the following manner:

```
2001:db8::202:b3ff:fe1e:8329
```

If an IPv6 address is assigned to the HMC for remote operations that use a web browser, browse to it by specifying that address. The address must be surrounded with square brackets in the browser's address field, as shown in the following example:

```
https://[fdab:1b89:fc07:1:201:6cff:fe72:ba7c]
```

The use of link-local addresses must be supported by your browser.

10.3.8 HMC Multi-factor authentication

Multi-factor authentication is an optional and configurable feature on per-user, per-template basis. It enhances security by requiring not only what you know (which is first factor) but also what you have available, which means that only person who owns a specific phone number can log in.

Multi-factor authentication first factor is login and password; the second factor is TOTP (Time-based One-Time Password) that is sent to your smartphone, desktop, or app (for example, Google Authenticator). This TOTP is defined in RFC 6238 standard and uses a cryptographic hash function that combines a secret key with the current time to generate a one-time password.

The secret key is generated by HMC/SE/TKE while the user is performing first factor logon. The secret key is known only to HMC/SE/TKE and to the user's smartphone. For that reason, it must be protected as much as your first factor password.

Multi-factor authentication code (MFA code) that was generated as a second factor is time-sensitive. Therefore, it is important to remember that it should be used soon after it is generated.

The algorithm within the HMC that is responsible for MFA code generation changes the code every 30 seconds. However, to make things easier, the HMC and SE console accepts current, previous, and next MFA codes. It is also important to have HMC, SE, and smartphone clocks synchronized. If the clocks are not synchronized, the MFA logon attempt fails. Time zone differences are irrelevant because the MFA code algorithm uses UTC.

On z15, HMC Version 2.15.0 provides integration of HMC authentication and z/OS MFA support, which means RSA SecurID authentication is achieved by way of centralized support from IBM MFA for z/OS, with the MFA policy defined in RACF and the HMC IDs assigned to RACF user IDs. The RSA SecurID passcode (from an RSA SecurID Token) is verified by the RSA authentication server. This authentication is supported on HMC only, *not* on the SE.

User Management task is changed on User Definition and User Template Definition to define and select the MFA Server, and for mapping the HMC user ID to the RACF user ID and selecting the RACF policy.

10.4 Remote Support Facility

The HMC Remote Support Facility (RSF) provides important communication to a centralized IBM support network for hardware problem reporting and service. The following types of communication are provided:

- ▶ Problem reporting and repair data
- ▶ Microcode Change Level (MCL) delivery
- ▶ Hardware inventory data, which is also known as vital product data (VPD)
- ▶ On-demand enablement

Consideration: RSF through a modem is *not* supported on the z15 HMC. Broadband connectivity is needed for hardware problem reporting and service.

10.4.1 Security characteristics

The following security characteristics are in effect:

- ▶ RSF requests always are started from the HMC to IBM. An inbound connection is never started from the IBM Service Support System.
- ▶ All data that is transferred between the HMC and the IBM Service Support System is encrypted with high-grade SSL/Transport Layer Security (TLS) encryption.
- ▶ When starting the SSL/TLS-encrypted connection, the HMC validates the trusted host with the digital signature that is issued for the IBM Service Support System.
- ▶ Data that is sent to the IBM Service Support System consists of hardware problems and configuration data.

More information: For more information about the benefits of Broadband RSF and the SSL/TLS-secured protocol, and a sample configuration for the Broadband RSF connection, see *Integrating the HMC Broadband Remote Support Facility into Your Enterprise*, SC28-6986.

10.4.2 RSF connections to IBM and Enhanced IBM Service Support System

If the HMC and SE are at Driver 22 or later, the driver uses a new remote infrastructure at IBM when the HMC connects through RSF for certain tasks. Check your network infrastructure settings to ensure that this new infrastructure works.

At the time of this writing, RSF still uses the “traditional” RETAIN connection. You must add access to the new Enhanced IBM Service Support System to your current RSF infrastructure (proxy, firewall, and so on).

To have the best availability and redundancy and to be prepared for the future, the HMC must access IBM by using the internet through RSF in the following manner: Transmission to the enhanced IBM Support System requires a domain name server (DNS). The DNS must be configured on the HMC if a proxy for RSF is not used. If a proxy for RSF is used, the proxy must provide the DNS.

The following host names and IP addresses are used and your network infrastructure must allow the HMC to access the following host names:

- ▶ www-945.ibm.com on port 443
- ▶ esupport.ibm.com on port 443

The following IP addresses (IPv4, IPv6, or both) can be used:

- ▶ IBM Enhanced support facility:
 - IPV4:
 - 129.42.54.189
 - 129.42.56.189
 - 129.42.60.189
 - IPV6:
 - 2620:0:6c0:200:129:42:54:189
 - 2620:0:6c2:200:129:42:56:189
 - 2620:0:6c4:200:129:42:60:189
- ▶ Legacy IBM support Facility:
 - IPV4:
 - 129.42.26.224
 - 129.42.42.224
 - 129.42.50.224
 - IPV6:
 - 2620:0:6c0:1::1000
 - 2620:0:6c2:1::1000
 - 2620:0:6c4:1::1000

Note: All other previous IP addresses are no longer supported.

10.4.3 HMC and SE remote operations

You can use the following methods to perform remote manual operations on the HMC:

- ▶ Use of a remote HMC

A remote HMC is a physical HMC that is on a different subnet from the SE. This configuration prevents the SE from being automatically discovered with IP multicast.

A remote HMC requires TCP/IP connectivity to each SE to be managed. Therefore, any customer-installed firewalls between the remote HMC and its managed objects must permit communication between the HMC and the SE. For service and support, the remote HMC also requires connectivity to IBM, or to another HMC with connectivity to IBM through RSF. For more information, see 10.4, “Remote Support Facility” on page 429.

- ▶ Use of a web browser to connect to an HMC

The z15 HMC application simultaneously supports one local user and any number of remote users. The user interface in the web browser is the same as the local HMC and has the same functions. Some functions are not available.

Note: Remote browser access is the default for the Hardware Management Appliance.

Access by using the UFD requires physical access to the HMC. Logon security for a web browser is provided by the local HMC user logon procedures. Certificates for secure communications are provided, and can be changed by the user. A remote browser session to the primary HMC that is managing an ensemble allows a user to perform ensemble-related actions, such as limiting remote web browser access.

You can now limit remote web browser access by specifying an IP address from the Customize Console Services task. To enable or disable the Remote operation service, click **Change...** in the Customize Console Services window, as shown in Figure 10-16.

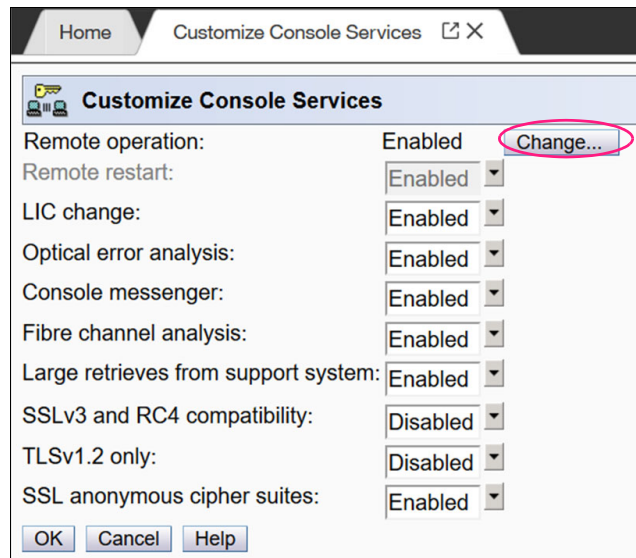


Figure 10-16 Customizing HMC remote operation

Microsoft Internet Explorer, Mozilla Firefox, and Goggle Chrome were tested as remote browsers. For more information about web browser requirements, see the HMC and SE console help system or [IBM Knowledge Center](#).

Single Object Operations

It is not necessary to be physically close to a SE to use it. The HMC can be used to access the SE remotely by using the SOO task. The interface is the same as the interface that is used on the SE. For more information, see the HMC and SE console help system or [IBM Knowledge Center](#).

Note: With HMC 2.15.0, certain tasks that required in the past to access to the SE in SOO mode were implemented as HMC tasks. With this enhancement, the HMC runs the tasks on the SE directly, without the need to lock the SE in SOO mode.

HMC mobile interface

The new mobile application interface allows HMC users to securely monitor and manage systems from anywhere. iOS and Android HMC applications are available to provide system and partition views, monitor the status and Hardware and Operating System Messages, and receive mobile push notifications from the HMC by using the IBM Z Remote Support Facility (zRSF) connection.

A full set of granular security controls are provided from the HMC console, to the user, monitor only, and mobile app password, including multi-factor authentication. This mobile interface is optional and is disabled by default.

10.5 HMC and SE capabilities

The HMC and SE feature many capabilities. This section describes the key areas. For more information about these capabilities, see the HMC and SE (Version 2.15.0) console help system or [IBM Knowledge Center](#).

With the introduction of the DPM mode for Linux on Z, only CPCs the user interface and user interaction with the HMC changed dramatically; the capabilities underneath are still the same. The figures and command examples that are shown in this section were taken in PR/SM mode.

10.5.1 Central processor complex management

The HMC is the primary place for CPC control. For example, the input/output configuration data set (IOCDs) includes definitions of LPARs, channel subsystems, control units, and devices, and their accessibility from LPARs. IOCDs can be created and put into production from the HMC.

The HMC is used to start the power-on reset (POR) of the system. During the POR, processor units (PUs) are characterized and placed into their respective pools, memory is put into a single storage pool, and the IOCDs is loaded and started into the hardware system area (HSA).

The hardware messages task displays hardware-related messages at the CPC, LPAR, or SE level. It also displays hardware messages that relate to the HMC.

10.5.2 LPAR management

Use the HMC to define LPAR properties, such as the number of processors of each type, how many are reserved, and how much memory is assigned to it. These parameters are defined in LPAR profiles and stored on the SE.

Because Processor Resource/Systems Manager (PR/SM) must manage LPAR access to processors and the initial weights of each partition, weights are used to prioritize partition access to processors.

You can use the Load task on the HMC to perform an IPL of an operating system. This task causes a program to be read from a designated device, and starts that program. You can perform the IPL of the operating system from storage, the USB flash memory drive (UFD), or an FTP server.

When an LPAR is active and an operating system is running in it, you can use the HMC to dynamically change certain LPAR parameters. The HMC provides an interface to change partition weights, add logical processors to partitions, and add memory.

LPAR weights can also be changed through a scheduled operation. Use the Customize Scheduled Operations task to define the weights that are set to LPARs at the scheduled time.

Channel paths can be dynamically configured on and off (as needed for each partition) from an HMC.

The Change LPAR Controls task for z15 can export the Change LPAR Controls table data to a comma-separated value (.csv)-formatted file. This support is available to a user when they are connected to the HMC remotely by a web browser.

Partition capping values can be scheduled and are specified on the Change LPAR Controls scheduled operation support. Viewing more information about a Change LPAR Controls scheduled operation is available on the SE.

One example of managing the LPAR settings is the absolute physical hardware LPAR capacity setting. Driver 15 (zEC12/zBC12) introduced the capability to define (in the image profile for shared processors) the absolute processor capacity that the image is allowed to use (independent of the image weight or other cappings).

To indicate that the LPAR can use the non-dedicated processors absolute capping, select **Absolute capping** on the Image Profile Processor settings to specify an absolute number of processors at which to cap the LPAR's activity. The absolute capping value can be "None" or a value for the number of processors (0.01 - 255.0).

The LPAR group absolute capping was the next step in partition capping options that are available on z15, z14 M0x, z14 ZR1, z13s, and z13 CPCs at Driver level 27 and greater. Following on to LPAR absolute capping, LPAR group absolute capping uses a similar methodology to enforce the following components:

- ▶ Customer licensing
- ▶ Non-z/OS partitions where group soft capping is not an option
- ▶ z/OS partitions where ISV does not support software capping

A group name, processor capping value, and partition membership are specified at the hardware console, along with the following properties:

- ▶ Set an absolute capacity cap by CPU type on a group of LPARs.
- ▶ Allow each of the partitions to use capacity up to their individual limits if the group's aggregate consumption does not exceed the group absolute capacity limit.
- ▶ Include updated SysEvent QVS support (used by vendors who implement software pricing).
- ▶ Only shared partitions are managed in these groups.
- ▶ Specify caps for one or more processor types in the group.
- ▶ Specify in absolute processor capacity (for example, 2.5 processors).

- Use Change LPAR Group Controls (as with windows that are used for software group-defined capacity), as shown in Figure 10-17 (snapshot on a z15).

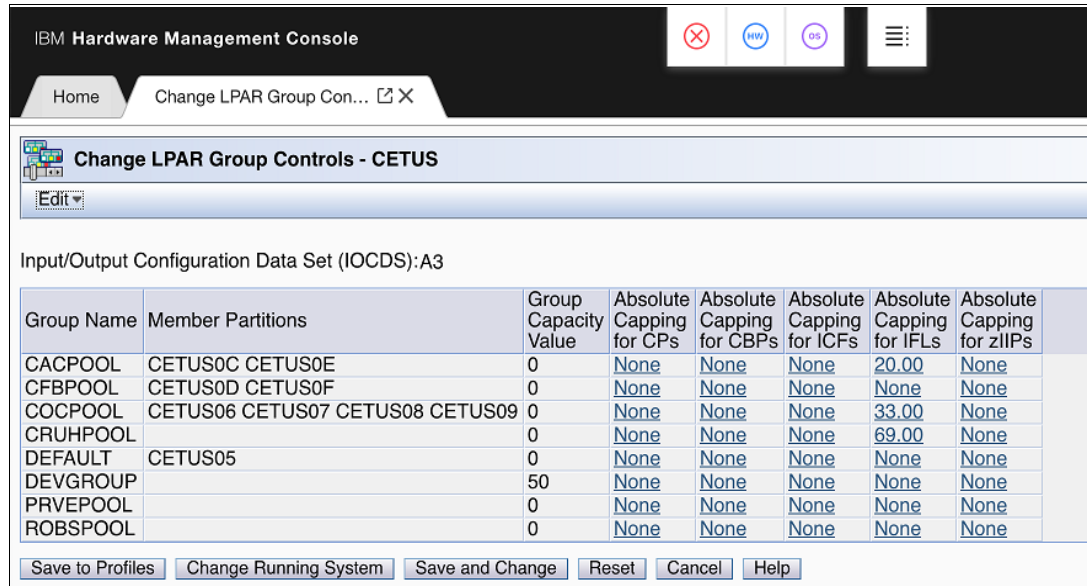


Figure 10-17 Change LPAR Group Controls: Group absolute capping

Absolute capping is specified as an absolute number of processors to which the group's activity is capped. The value is specified to hundredths of a processor (for example, 4.56 processors) worth of capacity.

The value is not tied to the Licensed Internal Code (LIC) configuration code (LICCC). Any value 0.01 - 255.00 can be specified. This configuration makes the profiles more portable, which means that you do not have issues in the future when profiles are migrated to new machines.

Although the absolute cap can be specified to hundredths of a processor, the exact amount might not be that precise. The same factors that influence the “machine capacity” also influence the precision with which the absolute capping works.

LPAR absolute capping can be changed through scheduled operations start with HMC version 2.15.0.

10.5.3 Operating system communication

The Operating System Messages task displays messages from an LPAR. You can also enter operating system commands and interact with the system. This task is especially valuable for entering Coupling Facility Control Code (CFCC) commands.

The HMC also provides integrated 3270 and ASCII consoles. These consoles allow an operating system to be accessed without requiring other network or network devices, such as TCP/IP or control units.

Updates to x3270 support

The Configure 3270 Emulators task on the HMC and TKE consoles was enhanced with Driver 15 to verify the authenticity of the certificate that is returned by the 3270 server when a secure and encrypted SSL connection is established to an IBM host. This 3270 Emulator with encrypted connection is also known as *Secure 3270*.

Use the Certificate Management task if the certificates that are returned by the 3270 server are not signed by a well-known trusted certificate authority (CA) certificate, such as VeriSign or Geotrust. An advanced action within the Certificate Management task, Manage Trusted Signing Certificates, is used to add trusted signing certificates.

For example, if the certificate that is associated with the 3270 server on the IBM host is signed and issued by a corporate certificate, it must be imported, as shown in Figure 10-18.

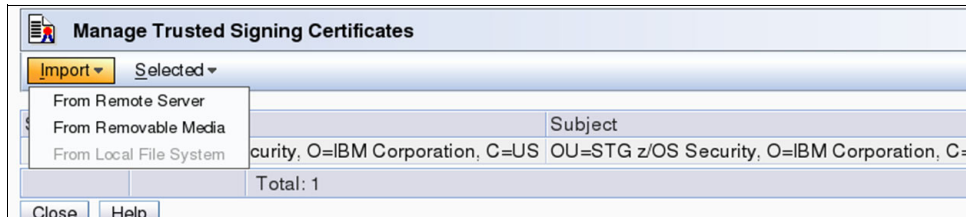


Figure 10-18 Manage Trusted Signing Certificates

The import from the remote server option can be used if the connection between the console and the IBM host can be trusted when the certificate is imported, as shown in Figure 10-19. Otherwise, import the certificate by using removable media.

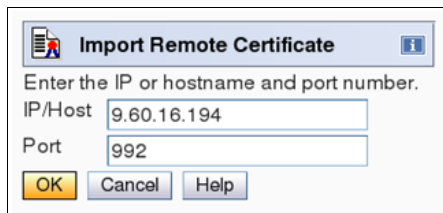


Figure 10-19 Import Remote Certificate example

A secure Telnet connection is established by adding the prefix L: to the IP address:port of the IBM host, as shown in Figure 10-20.

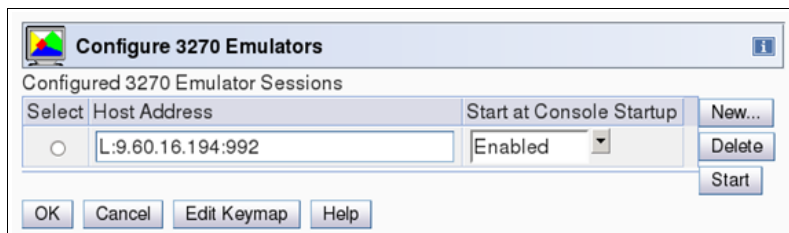


Figure 10-20 Configure 3270 Emulators

10.5.4 HMC and SE microcode

The microcode for the HMC, SE, and CPC is included in the driver or version. The HMC provides the management of the driver upgrade through Enhanced Driver Maintenance (EDM). EDM also provides the installation of the latest functions and the patches (MCLs) of the new driver.

When you perform a driver upgrade, always check the Driver (41) Customer Exception Letter option in the Fixes section at the IBM Resource Link.

For more information, see 9.9, “z15 Enhanced Driver Maintenance” on page 404.

Microcode Change Level

Regular installation of Microcode Change Levels (MCLs) is key for reliability, availability, and serviceability (RAS), optimal performance, and the following new functions:

- ▶ Install MCLs on a quarterly basis at a minimum.
- ▶ Review hiper MCLs continuously to decide whether to wait for the next scheduled fix application session or to schedule one earlier if the risk assessment warrants.
- ▶ Sign On the “IBM Z Security Portal” website and review for security alerts and related MCL fixes.

Tip: The IBM Resource Link provides access to the system information for your IBM Z system according to the system availability data that is sent on a scheduled basis. It provides more information about the MCL status of your z15 systems.

For more information about accessing the Resource Link, see [the IBM Resource Link website](#) (login required).

At the Resource Link website, click **Tools** → **Machine Information**, choose your IBM Z system, and then, click **EC/MCL**.

Microcode terms

The microcode features the following characteristics:

- ▶ The driver contains engineering change (EC) streams.
- ▶ Each EC stream covers the code for a specific component of z15. It includes a specific name and an ascending number.
- ▶ The EC stream name and a specific number are one MCL.
- ▶ MCLs from the same EC stream must be installed in sequence.
- ▶ MCLs can include installation dependencies on other MCLs.
- ▶ Combined MCLs from one or more EC streams are in one bundle.
- ▶ An MCL contains one or more Microcode Fixes (MCFs).

How the driver, bundle, EC stream, MCL, and MCFs interact with each other is shown in Figure 10-21.

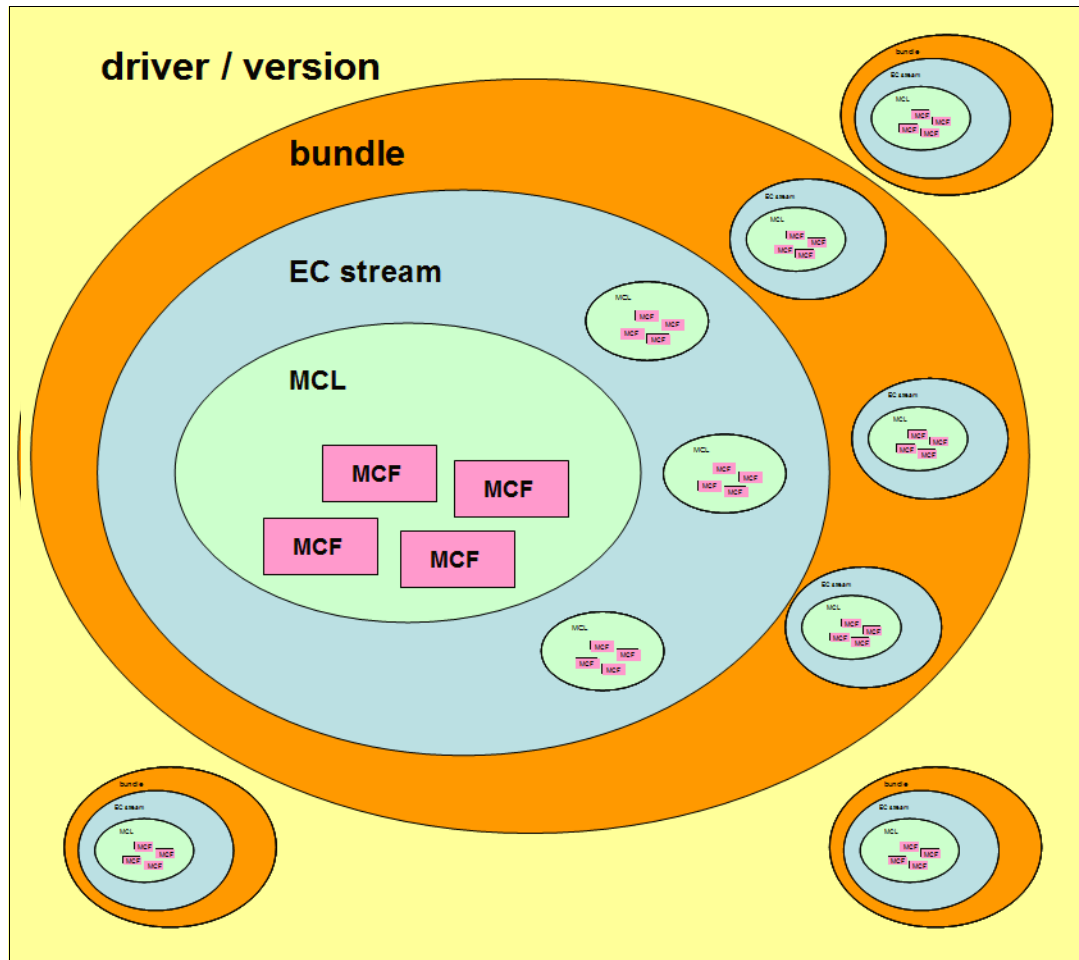


Figure 10-21 Microcode terms and interaction

MCL Application

By design, all MCLs can be applied concurrently, including:

- ▶ During MCL Bundle application.
- ▶ During Enhanced Driver Maintenance (Concurrent Driver Upgrade). For example, for z14 at Driver 32, the original GA level, upgrade from Driver 32 to Driver 36 (also applicable to z14 ZR1).

MCL Activation

By design and with planning, MCLs can be activated concurrently. Consider the following points:

- ▶ Most MCLs activate concurrently when applied.
- ▶ A few MCLs are “Pended” for scheduled activation because activation is disruptive in some way (Recent History: Most commonly seen for traditional OSA-Express features or Crypto-Express features).
- ▶ Activate traditional I/O Feature Pended MCL – LIC on the hardware feature:
 - Display Pending MCLs using HMC function or Resource Link Machine Information Reports

- Activate using HMC function on a feature basis by PCHID one at a time – disruptive: CONFIG the CHPID OFF to all sharing LPARs, activate, and then CONFIG ON to all
- ▶ Activate Native PCIe Pended MCL – LIC on a hardware feature OR Resource Group (RG) LIC:
 - Display Pending MCLs using HMC function or Resource Link Machine Information Reports
 - Feature LIC: Activate using HMC function on a one feature (PCHID) at a time basis - disruptive: CONFIG FUNCTIONS mapped to the feature OFF to all LPARs, activate, and then CONFIG ON
 - RG LIC: Activate using HMC function to each RG in turn – disruptive to all PCHIDs in the RG: CONFIG all FUNCTIONS mapped to all PCHIDs in RG1 OFF, activate, then CONFIG ON. Repeat for all PCHIDs in RG2, RG3, RG4

Note: For hardware that does not need CHPID or a FUNCTION definition (for example, Crypto Express), a different method that is specific to the feature is used.

- ▶ Alternative: Apply and activate all Pended MCLs disruptively with a scheduled Power On Reset (POR)

To discover this “Pended” situation, the following actions are done whenever an MCL is applied:

- ▶ Logon HMC and select CPC under “System Management”
- ▶ Change Management
- ▶ System Information
- ▶ Query Additional Actions

Or:

- ▶ Logon HMC and select CPC under “System Management”
- ▶ Change Management
- ▶ Query Channel/Crypto Configure Off/On Pending

Microcode installation by MCL bundle target

A *bundle* is a set of MCLs that are grouped during testing and released as a group on the same date. You can install an MCL to a specific target bundle level. The System Information window is enhanced to show a summary bundle level for the activated level, as shown in Figure 10-22 on page 439.

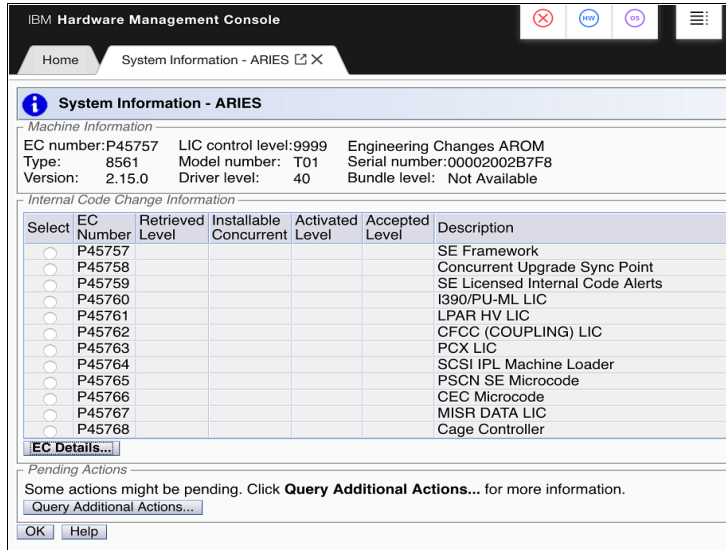


Figure 10-22 System Information: Bundle level

10.5.5 Monitoring

This section describes monitoring considerations.

Monitor task group

The Monitor task group on the HMC and SE includes monitoring-related tasks for IBM Z CPCs, as shown in Figure 10-23.

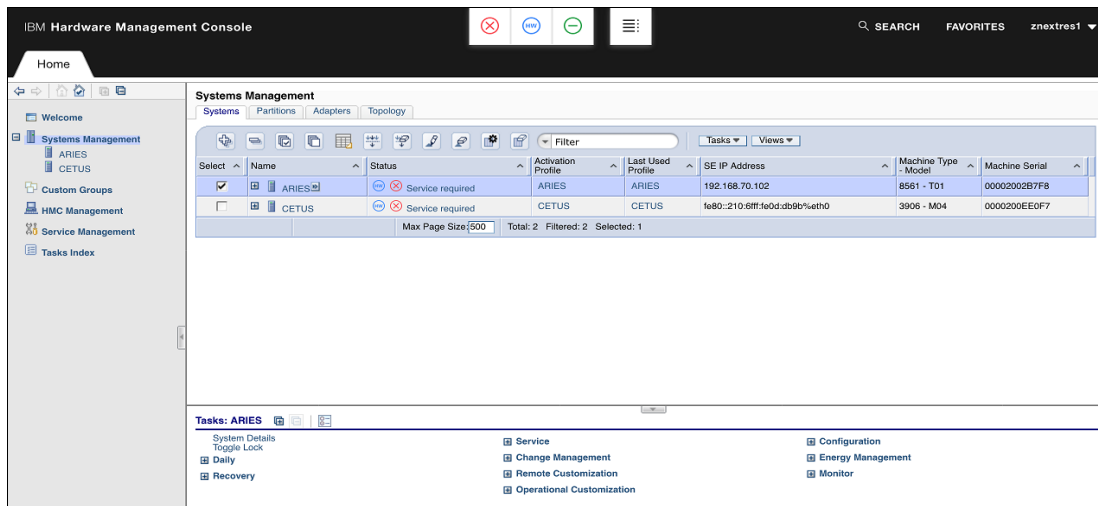


Figure 10-23 HMC Monitor Task Group

Monitors Dashboard task

The Monitors Dashboard task supersedes the System Activity Display (SAD). In the z15, the Monitors Dashboard task in the Monitor task group provides a tree-based view of resources.

Multiple graphical views are available for displaying data, including history charts. The Monitors Dashboard monitors processor and channel usage. It produces data that includes power monitoring information, power consumption, and the air input temperature for the system.

An example of the Monitors Dashboard task is shown in Figure 10-24.

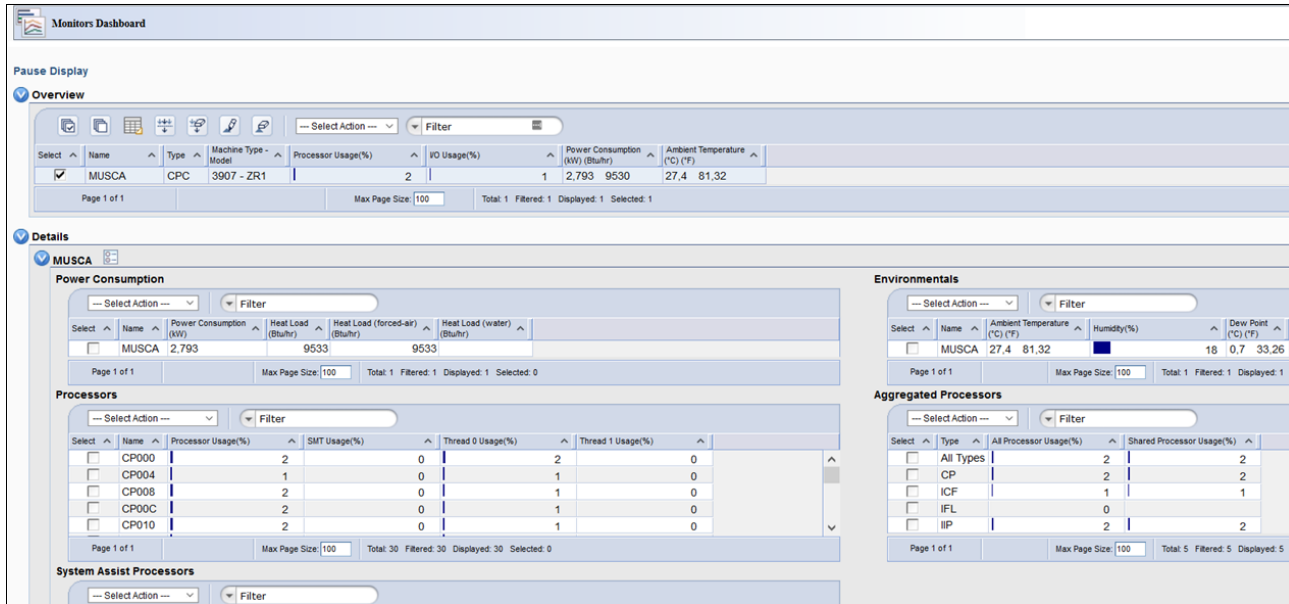


Figure 10-24 Monitors Dashboard task

You can display more information for the following components (see Figure 10-25 on page 441):

- ▶ Power consumption
- ▶ Environmental
- ▶ Aggregated processors
- ▶ Processors (with SMT information)
- ▶ System Assist Processors
- ▶ Logical Partitions
- ▶ Channels
- ▶ Adapters: Crypto use percentage is displayed according to the physical channel ID (PCHID number)

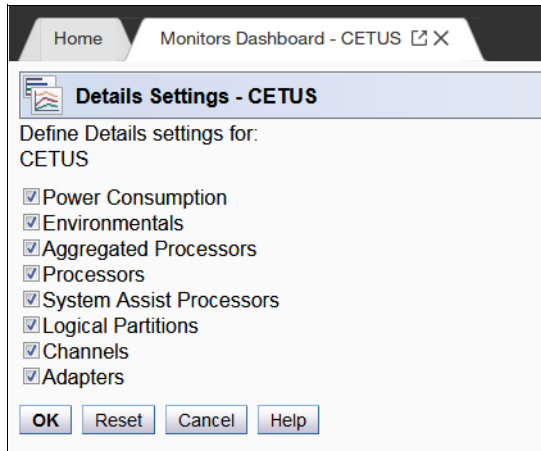


Figure 10-25 Monitors dashboard Detailed settings

Environmental Efficiency Statistics task

The Environmental Efficiency Statistics task is part of the Monitor task group. It provides historical power consumption and thermal information for the IBM Z CPC, and is available on the HMC.

The data is presented in table format and graphical “histogram” format. The data also can be exported to a .csv-formatted file so that the data can be imported into a spreadsheet. For this task, you must use a web browser to connect to an HMC.

10.5.6 Capacity on-demand support

All capacity on demand (CoD) upgrades are performed by using the SE Perform a Model Conversion task. Use the task to retrieve and activate a permanent upgrade, and to retrieve, install, activate, and deactivate a temporary upgrade. The task shows a list of all installed or staged LICCC records to help you manage them. It also shows a history of recorded activities.

The HMC for IBM z15 features the following CoD capabilities:

- ▶ SNMP API support:
 - API interfaces for granular activation and deactivation
 - API interfaces for enhanced CoD query information
 - API event notification for any CoD change activity on the system
 - CoD API interfaces, such as On/Off CoD and Capacity Back Up (CBU)
- ▶ SE window features (accessed through HMC Single Object Operations):
 - Window controls for granular activation and deactivation
 - History window for all CoD actions
 - Description editing of CoD records
- ▶ HMC/SE provides the following CoD information:
 - Millions of service units (MSU) and processor tokens
 - Last activation time
 - Pending resources that are shown by processor type instead of only a total count
 - Option to show more information about installed and staged permanent records
 - More information for the Attention state by providing seven more flags

HMC and SE are a part of the z/OS Capacity Provisioning environment. The Capacity Provisioning Manager (CPM) communicates with the HMC through IBM Z APIs, and enters CoD requests. For this reason, SNMP must be configured and enabled by using the Customize API Settings task on the HMC.

For more information about using and setting up CPM, see [IBM Knowledge Center](#) or the following publications:

- ▶ *z/OS MVS Capacity Provisioning User's Guide*, SC33-8299
- ▶ *IBM Z System Capacity on-Demand User's Guide*, SC28-6985

10.5.7 Server Time Protocol support

Important: The Sysplex Time task on the SE was discontinued for z15. Therefore, an HMC at Version 2.15.0 is required to manage system time for z15 CPCs.

With the Server Time Protocol (STP) functions, the role of the HMC is extended to provide the user interface for managing the Coordinated Timing Network (CTN). Consider the following points:

- ▶ IBM Z CPCs rely on STP for time synchronization, and continue to provide support of a pulse per second (PPS) port. Consider the following points:
 - **New for z15** - STP can be configured to use Precision Time Protocol (PTP, IEEE 1588) as external time source. Current implementation requires that the support element is connected to a PTP capable network infrastructure and has access to a PTP server.
 - An STP that uses Network Time Protocol (NTP) or PTP as External Time Source (ETS) server *with* PPS maintains an accuracy of 10 ms
 - An STP that uses ETS *without* PPS maintains accuracy of 100 ms
- ▶ The z15 cannot be in the same CTN with zEC12, zBC12, or earlier systems and cannot become member of a mixed CTN.

An STP-only CTN can be managed by using different HMCs. However, the HMC must be at the same driver level (or later) than any SE that is to be managed. Also, all SEs to be managed must be known (defined) to that HMC. In a STP-only CTN, the HMC can be used to perform the following tasks:

- ▶ Start or modify the CTN ID.
- ▶ Start the time (manually or by contacting an NTP server).
- ▶ Start the time zone offset, Daylight Saving Time offset, and leap second offset.
- ▶ Assign the roles of preferred, backup, and current time servers, and arbiter.
- ▶ Adjust time by up to plus or minus 60 seconds.
- ▶ Schedule changes to the offsets listed. STP can automatically schedule Daylight Saving Time, based on the selected time zone.
- ▶ Monitor the status of the CTN.
- ▶ Monitor the status of the coupling links that are started for STP message exchanges.
- ▶ For diagnostic purposes, the PPS port state on a z15 can be displayed and fenced ports can be reset individually.

STP changes and enhancements

Important: The Sysplex Time task on the SE was discontinued for z15. Therefore, an HMC at Version 2.15.0 is required to manage system time for z15 CPCs.

Detailed instructions and guidelines are provided within task workflow. z15 HMC provides a visual representation of the CTN topology. A preview of any configuration action is also shown in topological display. An example of the topology view is shown in Figure 10-26.

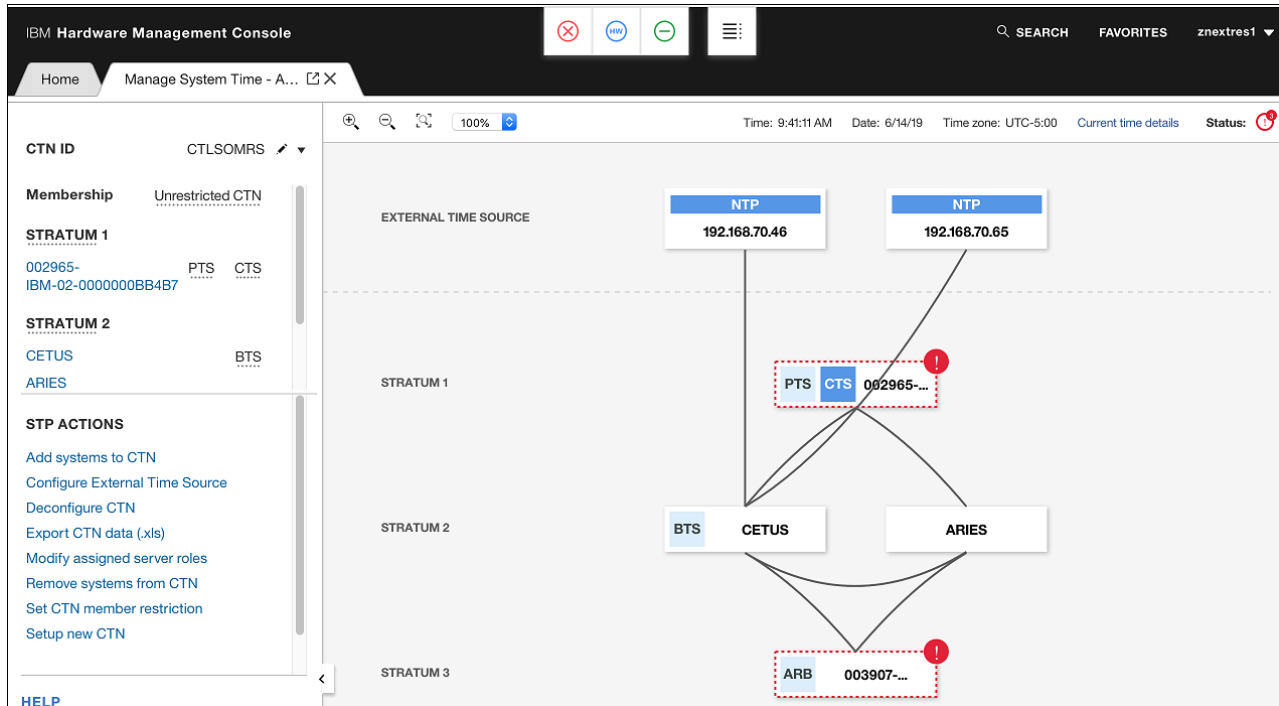


Figure 10-26 CTN topology visible on HMC Manage System Time window

Click **Current time details** for more information and available options (for example, adjusting time zone and leap second offset), as shown in Figure 10-27.

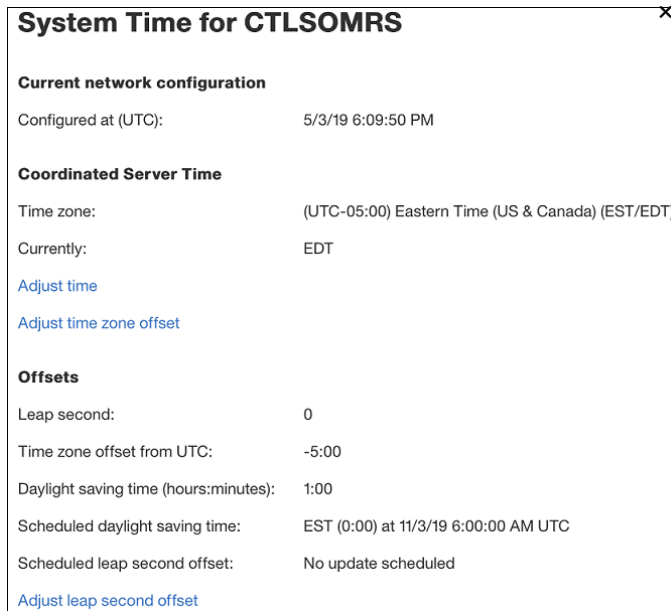


Figure 10-27 Current time details

Enhanced Console Assisted Recovery

Enhanced Console Assisted Recovery (ECAR) speeds up the process of BTS takeover by performing the following steps:

1. When the Primary Time Server (PTS/CTS) detects a checkstop condition, the CEC informs its SE and HMC.
2. The PTS SE recognizes the checkstop pending condition, and calls the PTS SE STP code.
3. The PTS SE sends an ECAR request through HMC to the Backup Time Server (BTS) SE.
4. The BTS SE communicates with the BTS to start the takeover.

ECAR support is faster than the original CAR support because the console path changes from a 2-way path to a 1-way path. Also, almost no lag time is incurred between the system checkstop and the start of CAR processing. Because the request is generated from the PTS before system logging, it avoids the potential of recovery being held up.

Requirements

ECAR is available on z15, z14 M0x, z14 ZR1, and z13/z13s systems on Driver 27 and later.

Attention: z15 and z14 ZR1 do not support InfiniBand connectivity; therefore, these servers cannot be connected by using IFB coupling/timing links to a z14 (3906), z13, or z13s. As such, in a CTN with servers that still use InfiniBand coupling, CTN roles (PTS, CTS, or Arbiter) must be assigned carefully in such a way that a failure of a CTN role playing server does not affect CTN functionality (loss of synchronization or transition to lower STP Stratum levels).

For more information about planning and setup, see the following publications:

- ▶ *Server Time Protocol Planning Guide*, SG24-7280
- ▶ *Server Time Protocol Implementation Guide*, SG24-7281
- ▶ *Server Time Protocol Recovery Guide*, SG24-7380

10.5.8 CTN Split and Merge

z15, z14 ZR1, and z14 support CTN split and CTN merge.

CTN Split

The HMC menus for Server Time Protocol (STP) provide support when one or more systems must be split in to a separate CTN without interruption in the clock source.

The task is available under the Advanced Actions menu in the Manage System Time task. Several checks are performed to avoid potential disruptive actions. If targeted CTN includes only members with the roles, task start fails with error message. If targeted CTN includes at least one system without any roles, the task starts. An informational warning is presented to the user to acknowledge that sysplex workloads are divided appropriately.

Merging two CTNs

When two separate CTNs must be merged in to the single CTN without interruption in the clock source, the system administrator must perform the Join existing CTN action, which is available in the Advanced Actions menu.

Note: After joining the selected CTN, all systems within the current CTN are synchronized with the Current Time Server of the selected CTN. A coupling link must be in place that connects the CTS of the selected CTN and the CTS of the current CTN.

During the transition state, most of the STP actions for the two affected CTNs are disabled. After the merge is completed, STP actions are enabled again.

For more information about planning and understanding STP server roles, see the following publications:

- ▶ *Server Time Protocol Planning Guide*, SG24-7280
- ▶ *Server Time Protocol Implementation Guide*, SG24-7281
- ▶ *Server Time Protocol Recovery Guide*, SG24-7380

10.5.9 NTP client and server support on the HMC

The NTP client support allows a STP-only CTN to use an NTP server as an ETS. This capability addresses the following requirements:

- ▶ Clients who want time synchronization for the servers members of the STP-only CTN
- ▶ Clients who use a common time reference across heterogeneous systems

The NTP server becomes the single time source (the ETS) for STP and other servers that are not IBM Z systems (such as AIX®, and Microsoft Windows) that include NTP clients.

The HMC can act as an NTP server. With this support, the z15 can receive the time from the HMC without accessing a LAN other than the HMC and SE network. When the HMC is used as an NTP server, it can be configured to receive the NTP source from the internet. For this type of configuration, a LAN that is separate from the HMC/SE LAN can be used.

HMC NTP broadband authentication support

HMC NTP authentication can be used since HMC Driver 15 (zEC12/zBC12). The SE NTP support is unchanged. To use this option on the SE, configure the HMC with this option as an NTP server for the SE.

Authentication support with a proxy

Some client configurations use a proxy for external access outside the corporate data center. NTP requests are User Datagram Protocol (UDP) socket packets and cannot pass through the proxy. The proxy must be configured as an NTP server to get to target servers on the web. Authentication can be set up on the client's proxy to communicate with the target time sources.

Authentication support with a firewall

If you use a firewall, HMC NTP requests can pass through it. Use HMC authentication to ensure untampered time stamps.

NTP symmetric key and autokey authentication

With symmetric key and autokey authentication, the highest level of NTP security is available. HMC Level 2.12.0 and later provide windows that accept and generate key information to be configured into the HMC NTP configuration. They can also issue NTP commands.

The HMC offers the following symmetric key and autokey authentication and NTP commands:

- ▶ Symmetric key (NTP V3-V4) authentication

Symmetric key authentication is described in RFC 1305, which was made available in NTP Version 3. Symmetric key encryption uses the same key for encryption and decryption. Users that are exchanging data keep this key to themselves. Messages encrypted with a secret key can be decrypted only with the same secret key. Symmetric key authentication supports network address translation (NAT).

- ▶ Symmetric key autokey (NTP V4) authentication

This autokey uses public key cryptography, as described in RFC 5906, which was made available in NTP Version 4. You can generate keys for the HMC NTP by clicking **Generate Local Host Key** in the Autokey Configuration window. This option issues the **ntp-keygen** command to generate the specific key and certificate for this system. Autokey authentication is not available with the NAT firewall.

- ▶ Issue NTP commands

NTP command support is added to display the status of remote NTP servers and the current NTP server (HMC).

For more information about planning and setup for STP and NTP, see the following publications:

- ▶ *Server Time Protocol Planning Guide*, SG24-7280
- ▶ *Server Time Protocol Implementation Guide*, SG24-7281
- ▶ *Server Time Protocol Recovery Guide*, SG24-7380

10.5.10 Security and user ID management

This section addresses security and user ID management considerations.

HMC and SE HD encryption

On z15 (continued emphasis on encryption):

- ▶ Password is never stored in clear (one-way hash)
- ▶ HMC/SE is a closed appliance; no means exist to get HDD
- ▶ All network traffic is TLS encrypted
- ▶ HMC/SE features embedded firewall
- ▶ Firmware is digitally signed and validated for delivery
- ▶ Firmware Integrity Monitoring is used for any attempted tempering post delivery

For HMC/SE Version 2.15.0, HDD encryption uses Trusted Platform Module (TPM) and Linux Unified Key Setup (LUKS) technology.

HMC and SE security audit improvements

With the Audit and Log Management task, audit reports can be generated, viewed, saved, and offloaded. The Customize Scheduled Operations task allows you to schedule audit report generation, saving, and offloading. The Monitor System Events task allows Security Logs to send email notifications by using the same type of filters and rules that are used for hardware and operating system messages.

With z15, you can offload the following HMC and SE log files for customer audit:

- ▶ Console event log
- ▶ Console service history
- ▶ Tasks performed log
- ▶ Security logs
- ▶ System log

Full log offload and delta log offload (since the last offload request) are provided. Offloading to removable media and to remote locations by FTP is available. The offloading can be manually started by the new Audit and Log Management task or scheduled by the Customize Scheduled Operations task. The data can be offloaded in the HTML and XML formats.

HMC user ID templates and LDAP user authentication

Lightweight Directory Access Protocol (LDAP) user authentication and HMC user ID templates enable the addition and removal of HMC users according to your own corporate security environment. These processes use an LDAP server as the central authority.

Each HMC user ID template defines the specific authorization levels for the tasks and objects for the user who is mapped to that template. The HMC user is mapped to a specific user ID template by user ID pattern matching. The system then obtains the name of the user ID template from content in the LDAP server schema data.

Default HMC user IDs

It is no longer possible to change the Managed Resource or Task Roles of the default user ID's operator, advanced, sysprog, acsadmin, and service.

If you want to change the roles for a default user ID, create your own version by copying a default user ID.

View-only user IDs and view-only access for HMC and SE

On z15 HMC version 2.15.0, with HMC and SE user ID support, users can be created that have "view-only" access to selected tasks. Support for "view-only" user IDs is available for the following purposes:

- ▶ Hardware messages
- ▶ Operating system messages
- ▶ View activation profiles
- ▶ Manage system time
- ▶ Manage Coupling Facility Port
- ▶ OSA Advanced Facilities
- ▶ Advance Facilities
- ▶ Configure Channel Path On/Off
- ▶ Configure on and off
- ▶ Cryptographic Configuration

HMC and SE secure FTP support

You can use a secure FTP connection from a HMC/SE FTP client to a customer FTP server location. This configuration is implemented by using the Secure Shell (SSH) File Transfer Protocol, which is an extension of SSH. You can use the Manage SSH Keys console action, which is available to the HMC and SE, to import public keys that are associated with a host address.

The Secure FTP infrastructure allows HMC and SE applications to query whether a public key is associated with a host address and to use the Secure FTP interface with the appropriate public key for a host. Tasks that use FTP now provide a selection for the secure host connection.

When selected, the task verifies that a public key is associated with the specified host name. If a public key is not provided, a message window opens that points to the Manage SSH Keys task to enter a public key. The following tasks provide this support:

- ▶ Import/Export IOCDs
- ▶ Advanced Facilities FTP IBM Content Collector Load
- ▶ Audit and Log Management (Scheduled Operations only)
- ▶ FCP Configuration Import/Export
- ▶ OSA view Port Parameter Export
- ▶ OSA-Integrated Console Configuration Import/Export

10.5.11 System Input/Output Configuration Analyzer on the SE and HMC

The System Input/Output Configuration Analyzer task supports the system I/O configuration function.

The information that is needed to manage a system's I/O configuration must be obtained from many separate sources. The System Input/Output Configuration Analyzer task enables the system hardware administrator to access, from one location, the information from those sources. Managing I/O configurations then becomes easier, particularly across multiple systems.

The System Input/Output Configuration Analyzer task runs the following functions:

- ▶ Analyzes the current active IOCDs on the SE.
- ▶ Extracts information about the defined channel, partitions, link addresses, and control units.
- ▶ Requests the channels' node ID information. The Fibre Channel connection (FICON) channels support remote node ID information, which is also collected.

The System Input/Output Configuration Analyzer is a view-only tool. It does not offer any options other than viewing. By using the tool, data is formatted and displayed in five different views. The tool provides various sort options, and data can be exported to a UFD for later viewing.

The following views are available:

- ▶ PCHID Control Unit View shows PCHIDs, channel subsystems (CSS), CHPIDs, and their control units.
- ▶ PCHID Partition View shows PCHIDs, CSS, CHPIDs, and the partitions in which they exist.
- ▶ Control Unit View shows the control units, their PCHIDs, and their link addresses in each CSS.

- ▶ Link Load View shows the Link address and the PCHIDs that use it.
- ▶ Node ID View shows the Node ID data under the PCHIDs.

10.5.12 Automated operations

As an alternative to manual operations, an application can interact with the HMC and SE through an API. The interface allows a program to monitor and control the hardware components of the system in the same way a user performs these tasks. On z15, the HMC APIs provide monitoring and control functions through SNMP. The API can get and set a managed object's attributes, issue commands, receive asynchronous notifications, and generate SNMP traps.

The older system, such as z13's HMC, supports the CIM as an extra systems management API. Starting with z14, the CIM support is removed.

For more information about APIs, see *IBM Z Application Programming Interfaces*, SB10-7164.

10.5.13 Cryptographic support

This section describes the cryptographic management and control functions that are available in the HMC and SE.

Cryptographic hardware

z15 systems include standard cryptographic hardware and optional cryptographic features for flexibility and growth capability.

The HMC/SE interface provides the following capabilities:

- ▶ Defining the cryptographic controls
- ▶ Dynamically adding a Crypto feature to a partition for the first time
- ▶ Dynamically adding a Crypto feature to a partition that already uses Crypto
- ▶ Dynamically removing a Crypto feature from a partition

The Crypto Express7S, which is a new Peripheral Component Interconnect Express (PCIe) cryptographic coprocessor, is an optional z15 exclusive feature. Crypto Express7S provides a secure programming and hardware environment on which crypto processes are run. Each Crypto Express7S adapter can be configured by the installation as a Secure IBM CCA coprocessor, a Secure IBM Enterprise Public Key Cryptography Standards (PKCS) #11 (EP11) coprocessor, or an accelerator.

When EP11 mode is selected, a unique Enterprise PKCS #11 firmware is loaded into the cryptographic coprocessor. It is separate from the Common Cryptographic Architecture (CCA) firmware that is loaded when a CCA coprocessor is selected. CCA firmware and PKCS #11 firmware cannot coexist in a card.

The Trusted Key Entry (TKE) Workstation with smart card reader feature is required to support the administration of the Crypto Express7S when configured as an Enterprise PKCS #11 coprocessor.

To support the new Crypto Express7S card, the TKE9.2 is needed. An example of the Cryptographic Configuration window is shown in Figure 10-28 on page 450.

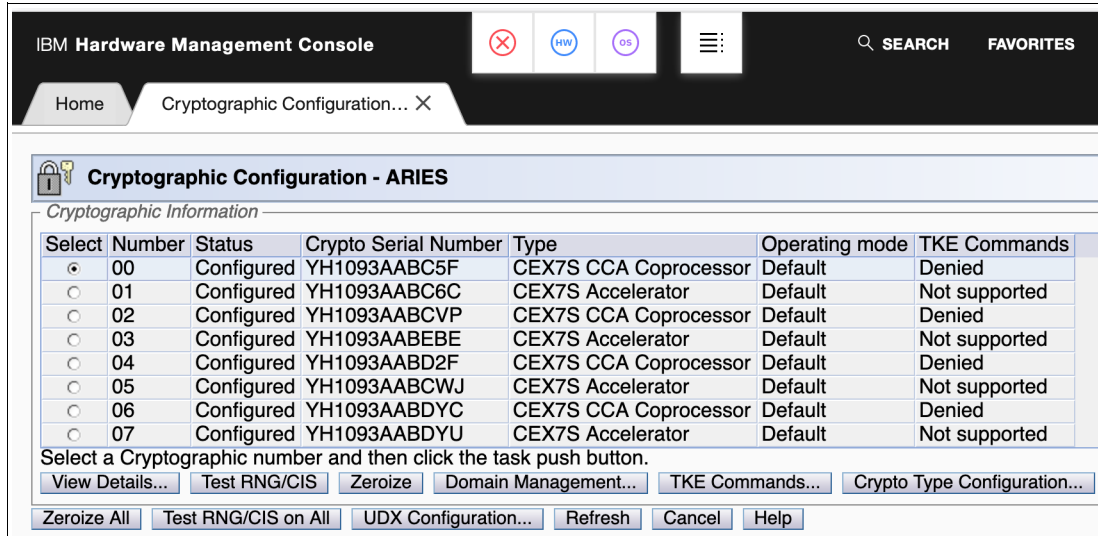


Figure 10-28 Cryptographic Configuration window

The Usage Domain Zeroize task is provided to clear the appropriate partition crypto keys for a usage domain when you remove a crypto card from a partition. Crypto Express7/6/5S in EP11 mode is configured to the standby state after the zeroize process.

For more information, see *IBM z15 (8561) Configuration Setup*, SG24-8860.

Digitally signed firmware

Security and data integrity are critical issues with firmware upgrades. Procedures are in place to use a process to digitally sign the firmware update files that are sent to the HMC, SE, and TKE. By using a hash algorithm, a message digest is generated that is then encrypted with a private key to produce a digital signature.

This operation ensures that any changes that are made to the data are detected during the upgrade process by verifying the digital signature. It helps ensure that no malware can be installed on IBM Z products during firmware updates. It also enables the z15 Central Processor Assist for Cryptographic Function (CPACF) functions to comply with Federal Information Processing Standard (FIPS) 140-2 Level 1 for Cryptographic LIC changes. The enhancement follows the IBM Z focus of security for the HMC and the SE.

The Crypto Express7S (CEX7S) is compliant with CCA PCI HSM. TKE workstation is optional when used to manage a Crypto Express7S feature that is defined as a CCA coprocessor in normal mode. However, it is mandatory when it is used to manage a Crypto Express7S feature that is defined as a CCA coprocessor in PCI-HSM mode or is defined as an EP11 coprocessor (CCA in PCI-HSM mode and EP11 also require a smart card reader plus smart cards with FIPS certification).

10.5.14 Installation support for z/VM that uses the HMC

Starting with z/VM V5R4 and z10, Linux on Z can be installed in a z/VM virtual machine from HMC workstation media. This Linux on Z installation can use the communication path between the HMC and the SE. No external network or extra network setup is necessary for the installation.

10.5.15 Dynamic Partition Manager

DPM is an IBM Z mode of operation that provides a simplified approach to create and manage virtualized environments, which reduces the barriers of its adoption for new and existing customers.

Setting up is a disruptive action. The selection of the DPM mode of operation is done by using the Enable Dynamic Partition Manager function, which is available in the SE CPC Configuration menu. Enabling DPM is performed on the SE and requires a system POR.

Attention: The Enabling Dynamic Partition Manager task is run on the SE and is performed by your IBM system service representative (SSR).

After the CPC is restarted and you log on to the HMC in which this CPC is defined, the HMC shows the Welcome window that is shown in Figure 10-29.

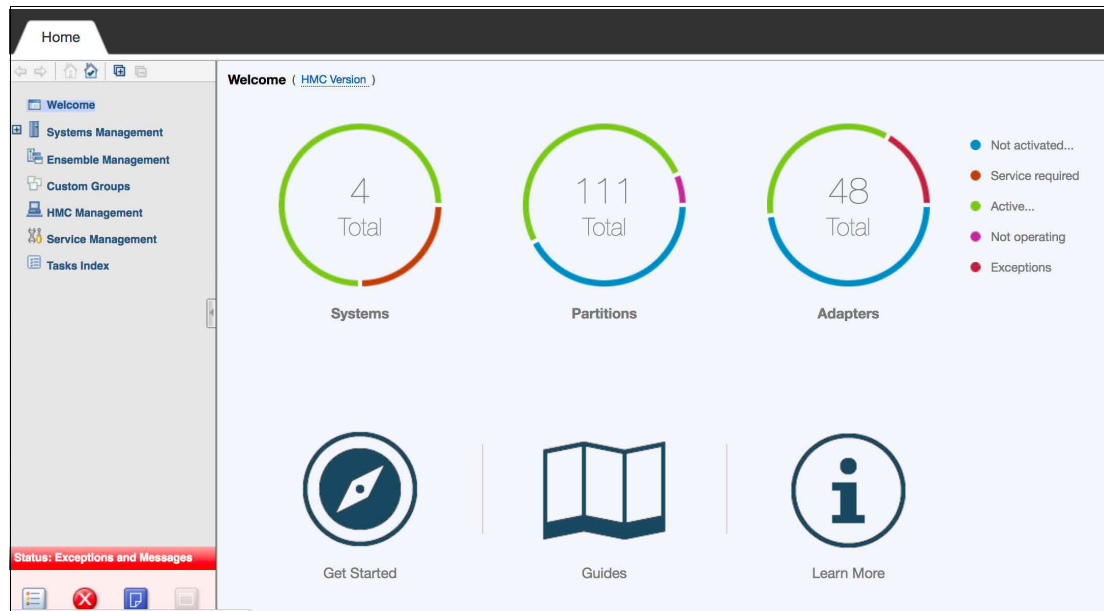


Figure 10-29 HMC welcome window

New partitions can be added by selecting **Get Started**. For more information, see [IBM Knowledge Center](#).



Environmentals

This chapter describes the environmental requirements for IBM z15™ servers. It also lists the dimensions, weights, power, and cooling requirements that are needed to plan for the installation of an z15 server.

Naming: Throughout this chapter, *z15* refers to IBM z15 Model T01 (Machine Type 8561), unless otherwise specified.

The following options are available for physically installing the server:

- ▶ Air or water cooling
- ▶ Power Distribution Unit (PDU) or Bulk Power Assembly (BPA) for power
- ▶ Installation on a raised floor or non-raised floor
- ▶ I/O and power cables can exit under the raised floor or off the top of the server frames

For more information about physical planning, see *IBM Z 8561 Installation Manual for Physical Planning*, GC28-7002.

This chapter includes the following topics:

- ▶ 11.1, “Power and Cooling” on page 454
- ▶ 11.2, “Physical specifications” on page 466
- ▶ 11.3, “Physical planning” on page 467
- ▶ 11.4, “Energy management” on page 471

11.1 Power and Cooling

The z15 server can be a 1 - 4 19-inch rack system, depending on the configuration. Frames are shipped separately in Arbo crates, along with separate front and rear cover sets for each frame boxed on a shipping pallet. The frames are bolted together during the installation procedure. z15 servers support installation on a raised floor or non-raised floor.

The z15 servers are available in the following power and cooling options:

- ▶ Intelligent Power Distribution Unit-based power (iPDU) - or PDU
Radiator-based cooling (air cooled system)
- ▶ Bulk Power Assembly-based power (BPA):
 - Radiator-based cooling (air cooled system)
 - Water Cooling Unit that uses customer-provided chilled water supply

11.1.1 Intelligent Power Distribution Unit (iPDU)

The iPDU can be ordered as the following feature codes, per client datacenter power infrastructure requirements:

- ▶ 60A / 3 Phase “Delta” PDU feature code (FC 0629)
- ▶ 60A / 3 Phase “Wye” PDU feature code (FC 0630)

A PDU-based system can have 2 - 8 power cords, depending on the configuration. The use of iPDU on z15 might enable fewer frames, which allows for more I/O slots to be available and improves power efficiency to lower overall energy costs. It also offers some standardization and ease of data center installation planning, which allows the z15 to easily coexist with other platforms within the data center.

Power requirements

The z15 is designed with a fully redundant power system. To make full use of the redundancy that is built into the server, the PDUs within one pair must be powered from different power distribution panels. In that case, if one PDU in a pair fails, the second PDU ensures continued operation of the server without interruption.

The second, third, and fourth PDU pairs are installed dependent on other CPC or PCIe+ I/O drawers installed. The locations of the PDU pairs and frames are listed in Table 11-1.

Table 11-1 PDU pairs and frames location

| Frame C | Frame B | Frame A |
|-------------|-------------|-------------|
| N/A | N/A | PDU A3 / A4 |
| PDU C1 / C2 | PDU B1 / B2 | PDU A1 / A2 |

Power cords for the PDUs are attached to the options that are listed in Table 11-2.

Table 11-2 Power cords for PDUs

| Supply type | Input voltage | Input frequency | Input current rating |
|------------------------------------|----------------|--------------------------------------|----------------------|
| 2, 4, 6, or 8, 3-phase power cords | 200 - 240 V AC | 50/60 Hz (47 - 63 Hz with tolerance) | 48 A |
| 2, 4, 6, or 8, 3-phase power cords | 380 - 415 V AC | 50/60 Hz (47 - 63 Hz with tolerance) | 24 A |

A rear view of a maximum configured PDU-powered system with five CPC drawers and 12 PCIe+ I/O drawers is shown in Figure 11-1.

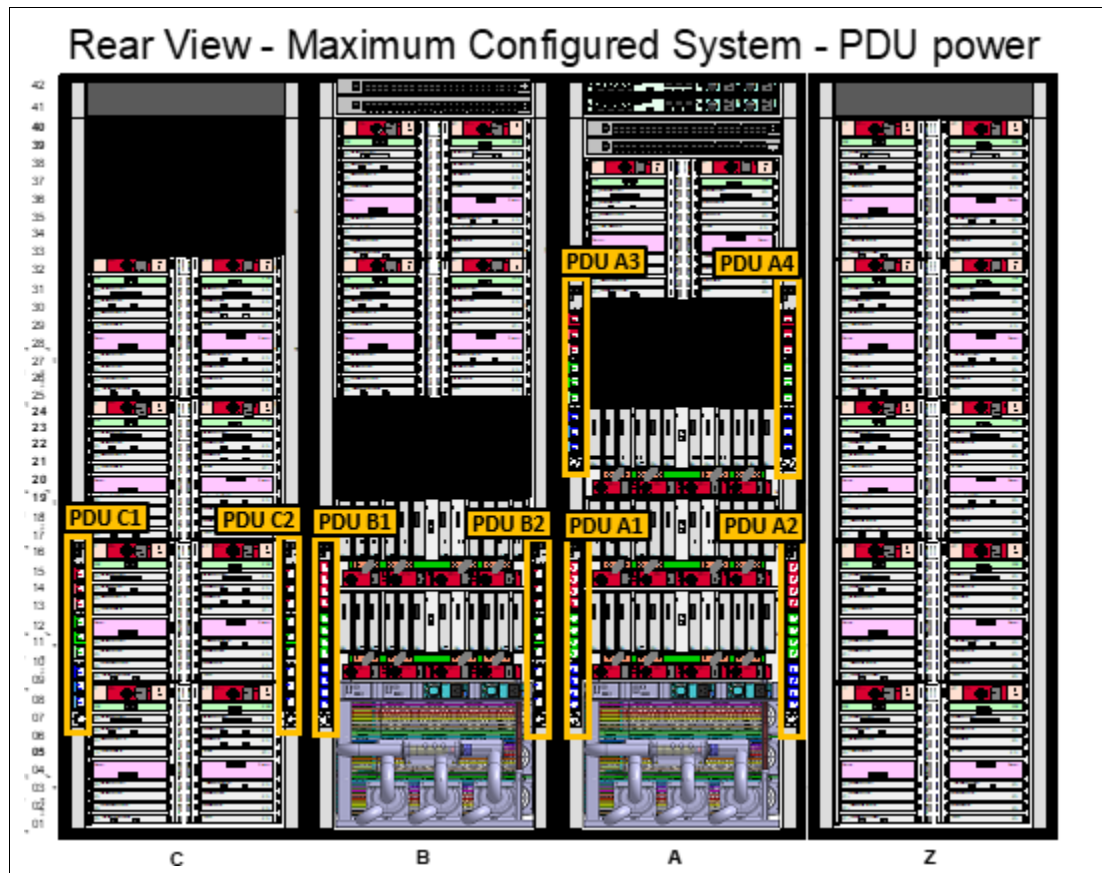


Figure 11-1 Rear View of maximum configured PDU system

The number of PDUs and power cords that are required based on the number of CPC drawers and PCIe+ I/O drawers are in Table 11-3.

Table 11-3 Number of PDUs installed

| Number of CPCs | Number of PCIe+ I/O drawers | | | | | | | | | | | | |
|----------------|-----------------------------|---|---|---|---|---|---|---|---|---|----|----|-----|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 2 | 2 | 2 | 2 | | | | | | 6 | 6 | 6 | 6 |
| 2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 6 | 6 | 6 | 6 | 6 |
| 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 6 | 6 | 6 | 6 | 6 | N/A |

| Number of CPCs | Number of PCIe+ I/O drawers | | | | | | | | | | | | |
|----------------|-----------------------------|---|---|---|---|---|---|---|---|---|----|----|----|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 4 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 |
| 5 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 8 |

Power consumption

The utility power consumption for the z15 for PDU option is listed in Table 11-4.

Table 11-4 z15 utility power consumption for PDU

| FC CPC # | Number of PCIe+ I/O drawers | | | | | | | | | | | | |
|------------------|-----------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| max34 FC0655 | 3.8 kw | 4.8 kw | 5.9 kw | 6.9 kw | 8.0 kw | 9.0 kw | 10.0 kw | 11.1 kw | 12.1 kw | 13.2 kw | 14.2 kw | 15.3 kw | 16.1 kw |
| max71 FC0656 | 6.9 kw | 7.9 kw | 9.0 kw | 10.0 kw | 11.1 kw | 12.1 kw | 13.2 kw | 14.2 kw | 15.3 kw | 16.3 kw | 17.4 kw | 18.4 kw | 19.2 kw |
| max108 FC0657 | 10.1 kw | 11.1 kw | 12.2 kw | 13.2 kw | 14.3 kw | 15.3 kw | 16.4 kw | 17.4 kw | 18.5 kw | 19.5 kw | 20.6 kw | 21.6 kw | N/A |
| max145 FC0658 | 13.5 kw | 14.5 kw | 15.6 kw | 16.6 kw | 17.7 kw | 18.7 kw | 19.8 kw | 20.8 kw | 21.9 kw | 22.9 kw | 24.0 kw | 25.0 kw | 25.8 kw |
| max190 FC0659 | 16.6 kw | 17.6 kw | 18.7 kw | 19.7 kw | 20.8 kw | 21.8 kw | 22.9 kw | 23.9 kw | 25.0 kw | 26.0 kw | 27.1 kw | 28.1 kw | 28.9 kw |

Note: Consider the following points:

- ▶ The power values that are listed in this table assume the CPC process drawer and PCIe+ I/O drawers are plugged to the maximum with highest power features (that is, memory and I/O adapters and fanouts). Also assumed is that maximum ambient temperature is used.
- ▶ Typical configurations and data center conditions result in lower power. A calculator available on Resource Link calculates power and weight for specific configurations and environmental conditions.

Considerations: Power consumption is lower in a normal ambient temperature room, and for configurations that feature a lesser number of I/O slots, smaller amount of memory, and fewer PUs.

Power estimation for any configuration, power source, and room condition can be obtained by using the power estimation tool that is available at the [IBM Resource Link website](#) (login required).

On the Resource Link page, click **Tools** → **Power and weight estimation**.

11.1.2 Bulk Power assembly (BPA)

The Bulk Power Assembly consists of the following features:

- ▶ BPA (FC 0640)
- ▶ Balanced Power Plan ahead (FC 3003)
- ▶ Bulk Power Regulator (BPR) (FC 3016)
- ▶ Internal Battery Feature (FC 3217)

Removal of IBF support^a: IBM z15 is planned to be the last IBM Z server to offer an Internal Battery Feature (IBF). As client data centers continue to improve power stability and uninterruptible power supply (UPS) coordination, IBM Z continues to innovate to help clients take advantage of common power efficiency and monitoring across their ecosystems. Additional support for data center power planning can be requested through your IBM Sales contact.

- a. Statements by IBM regarding its plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM.

The BPA option is required for clients who order an Internal Battery Feature (IBF), Water Cooling Unit (WCU), or Balanced Power. The BPA requires two or four power cords. All BPAs are concurrently repairable in the field.

Power requirements

The 8561 operates from 2 or 4 fully redundant power supplies. These redundant power supplies each have their own power cords, or pair of power cords, which allows the system to survive the loss of customer power to either power cord or power cord pair.

If power is interrupted to one of the power supplies, the other power supply assumes the entire load and the system continues to operate without interruption. Therefore, the power cords for each power supply must be wired to support the entire power load of the system.

For the most reliable availability, the power cords in the rear of the frame should be powered from different PDUs. All power cords exit through the rear of the frame. The utility current distribution across the phase conductors (phase current balance) depends on the system configuration. Each front end power supply is provided with phase switching redundancy.

The loss of an input phase is detected and the total input current is switched to the remaining phase pair without any power interruption. Depending on the configuration input power draw, the system can run from several minutes to indefinitely in this condition. Because most single phase losses are transients that recover in seconds, this redundancy provides protection against virtually all single phase outages.

Power cords for the BPAs are attached to the options that are listed in Table 11-5.

Table 11-5 Power cords for the BPAs

| Supply type | Nominal voltage range | Voltage tolerance | Frequency range |
|---|-----------------------|-------------------|-----------------|
| Two or four redundant 3-phase power cords | 200 - 480 V AC | 180 - 508 V AC | 50 / 60 Hz |

The source power cords ratings for BPA systems are listed in Table 11-6.

Table 11-6 Source power cords for BPA

| Source type | Frequency | Input voltage range | Rated input current |
|---------------------------|------------|----------------------------|---------------------|
| Three-phase (60A plug) | 50 / 60 Hz | 200 V ^a | 50 A |
| Three-phase (60A plug) | 50 / 60 Hz | 208 - 240 V ^b | 48 A |
| Three-phase (30A plug) | 50 / 60 Hz | 380 V - 415 V ^c | 24 A |
| Three-phase (30A plug) | 60 Hz | 480 V ^d | 20 A |
| Three-phase (63A no plug) | 50 / 60 Hz | 220 V - 480 V | 48 A |

| Source type | Frequency | Input voltage range | Rated input current |
|---------------------------|------------|---------------------|---------------------|
| Three-phase (32A no plug) | 50 / 60 Hz | 380 V - 415 V | 25 A |

- a. Japan (same physical cord as Note b)
- b. US, Canada (same physical cord as Note a)
- c. US, Canada only
- d. US only

A rear view of a maximum configured BPA powered system with two BPA pairs, IBF, five CPC drawers, and 11 PCIe+ I/O drawers is shown in Figure 11-2.

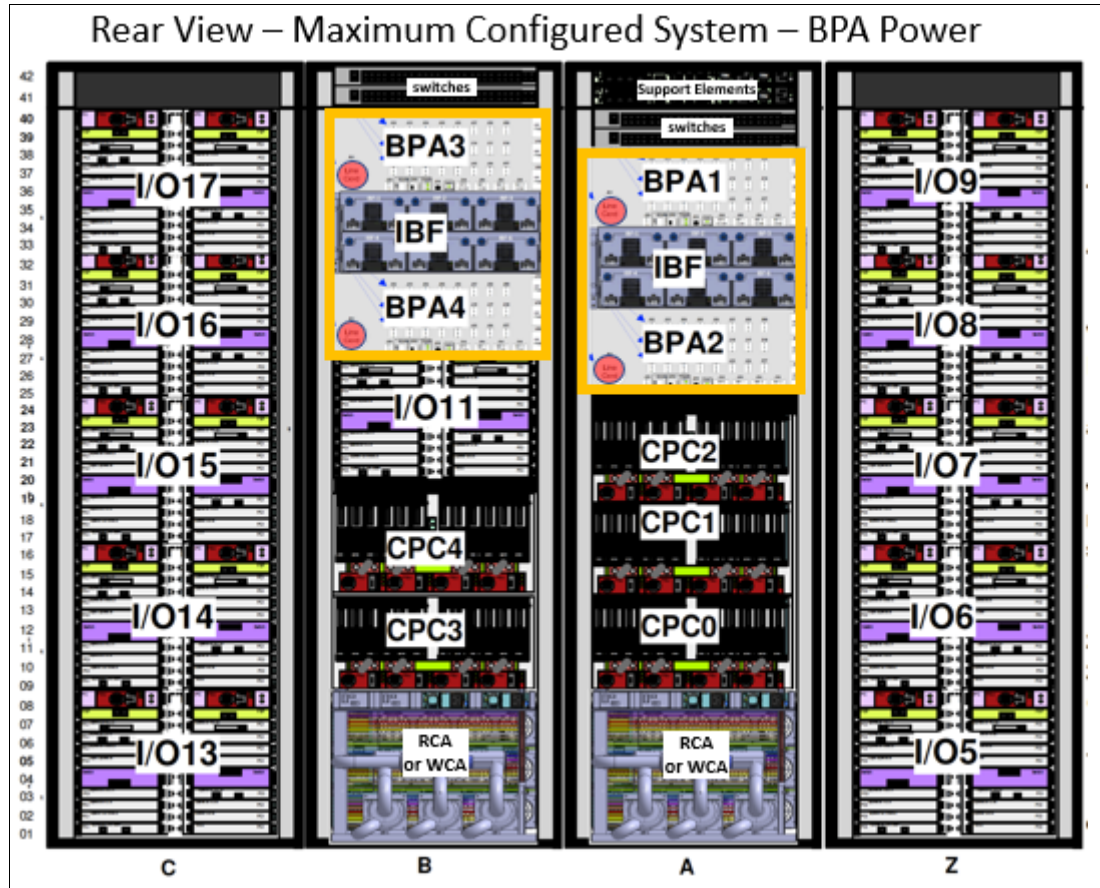


Figure 11-2 Rear view of maximum configured BPA system

The number of power cords that are required based on the number of CPC drawers and PCIe+ I/O drawers are listed in Table 11-7.

Table 11-7 Number of power cords installed

| Number of CPCs | Number of PCIe+ I/O drawers | | | | | | | | | | | |
|----------------|-----------------------------|---|---|---|---|---|---|---|---|---|----|----|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 |
| 2 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 4 |
| 3 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 4 |
| 4 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 |

| Number of CPCs | Number of PCIe+ I/O drawers | | | | | | | | | | | |
|----------------|-----------------------------|---|---|---|---|---|---|---|---|---|----|----|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 5 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 |

The number of power cords and number of Bulk Power Regulators required based on the number of CPC processor drawers and number of PCIe+ I/O drawers in the configuration (radiator or WCU) is shown in Table 11-8 on page 459.

Table 11-8 Number of BPRs installed per BPA

| | Number of PCIe+ I/O drawers | | | | | | |
|-------|-----------------------------|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 CPC | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| 2 CPC | 2 | 2 | 2 | 2 | 2 | 3 | 3 |
| 3 CPC | 2 | 2 | 3 | 3 | 3 | 3 | - |

Note: Balanced power feature includes all BPRs that are plugged in all BPAs in the system.

Power consumption

The utility power consumption for the z15 for the BPA radiator cooled system option is listed in Table 11-9.

Table 11-9 Utility Power consumption (BPA) for radiator-cooled systems

| CPC Feature Code | Number of PCIe+ I/O drawers | | | | | | | | | | | |
|-------------------------|-----------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| max34 FC0655 (1) | 4.2 kw | 5.3 kw | 6.4 kw | 7.5 kw | 8.6 kw | 9.7 kw | 10.9 kw | 12.0 kw | 13.1 kw | 14.2 kw | 15.3 kw | 16.4 kw |
| max71 FC0656 (2) | 7.5 kw | 8.6 kw | 9.7 kw | 10.8 kw | 11.9 kw | 13.0 kw | 14.2 kw | 15.3 kw | 16.4 kw | 17.5 kw | 18.6 kw | 18.6 kw |
| max108 FC0657 (3) | 10.9 kw | 12.0 kw | 13.1 kw | 14.2 kw | 15.3 kw | 16.4 kw | 17.6 kw | 18.7 kw | 19.8 kw | 20.9 kw | 22.0 kw | 22.0 kw |
| max145 FC0658 (4) | 14.8 kw | 15.9 kw | 17.0 kw | 18.1 kw | 19.2 kw | 20.3 kw | 21.4 kw | 22.5 kw | 23.6 kw | 24.8 kw | 25.9 kw | 25.9 kw |
| max190 FC0659 (5) | 18.1 kw | 19.2 kw | 20.3 kw | 21.4 kw | 22.5 kw | 23.6 kw | 24.7 kw | 25.8 kw | 26.9 kw | 28.0 kw | 29.2 kw | 29.2 kw |

Note: Consider the following points:

- ▶ The power values that are listed in this table assume the CPC process drawer and PCIe+ I/O drawers are plugged to the maximum with highest power features (that is, memory and I/O adapters and fanouts). Also assumed is that maximum ambient temperature is used.
- ▶ Typical configurations and data center conditions result in lower power. A calculator that is available on Resource Link calculates power and weight for specific configurations and environmental conditions.

The utility power consumption for the z15 for the BPA water-cooled system option is listed in Table 11-10.

Table 11-10 Utility Power consumption (BPA) for water-cooled systems

| CPC Feature Code | Number of PCIe+ I/O drawers | | | | | | | | | | | |
|-------------------------|-----------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| max34 FC0655 (1) | 4.0 kw | 5.1 kw | 6.2 kw | 7.3 kw | 8.4 kw | 9.5 kw | 10.6 kw | 11.8k w | 12.9 kw | 14.0 kw | 15.1 kw | 16.2 kw |
| max71 FC0656 (2) | 7.2 kw | 8.3 kw | 9.4 kw | 10.5 kw | 11.6 kw | 12.7 kw | 13.8 kw | 14.9 kw | 16.1 kw | 17.2 kw | 18.3 kw | 19.4 kw |
| max108 FC0657 (3) | 10.4 kw | 11.5 kw | 12.6 kw | 13.7 kw | 14.8 kw | 15.9 kw | 17.0 kw | 18.1 kw | 19.2 kw | 20.4 kw | 21.5 kw | 22.6 kw |
| max145 FC0658 (4) | 14.0 kw | 15.1 kw | 16.2 kw | 17.3 kw | 18.5 kw | 19.6 kw | 20.7 kw | 21.8 kw | 22.9 kw | 24.0 kw | 25.1 kw | 26.2 kw |
| max190 FC0659 (5) | 17.2 kw | 18.3 kw | 19.4 kw | 20.5 kw | 21.6 kw | 22.8 kw | 23.9 kw | 25.0 kw | 26.1 kw | 27.2 kw | 28.3 kw | 29.4 kw |

Notes: Consider the following points:

- ▶ The power values that are listed in this table assume the CPC process drawer and PCIe+ I/O drawers are plugged to the maximum with highest power features (that is, memory and I/O adapters and fanouts). Also assumed is that maximum ambient temperature is used.
- ▶ Typical configurations and data center conditions result in lower power. A calculator that is available on Resource Link calculates power and weight for specific configurations and environmental conditions.

Balanced Power Plan Ahead feature

Phase currents are minimized when they are balanced among the three input phases. Balanced Power Plan Ahead (FC 3003) is designed to allow you to order the full complement of bulk power regulators (BPRs) on any configuration to help ensure that the configuration is in a balanced power environment

If the z15 server is configured with the Internal Battery Feature (IBF), Balanced Power Plan Ahead automatically supplies the maximum number of batteries (IBFs) with the system.

Consideration: Power consumption is lower when in a normal ambient temperature room, and for configurations that feature a lesser number of I/O slots, smaller amount of memory, and fewer processors. Power consumption is also slightly lower for DC input voltage. The numbers that are listed in this section assume that batteries are present and charging.

Power estimation for any configuration, power source, and room condition can be obtained by using the power estimation tool at [IBM Resource Link website](#) (authentication required).

On the Resource Link page, click **Tools** → **Power and weight estimation**.

11.1.3 Cooling requirements

The z15 cooling system includes with two options: Radiator (air) cooled or water-cooled system. Single chip modules (SCMs) are always cooled with an internal water loop, no matter which z15 cooling option is chosen. The liquid in the internal water circuit can be cooled by using a radiator (for air-cooling option) or customer-supplied chilled water supply (for water-cooling option). I/O drawers, PCIe I/O drawers, power enclosures, and CPC drawers are cooled by chilled air with blowers.

The z15 servers include a recommended (long-term) ambient temperature range of 18°C (64.4°F) - 27°C (80.6°F). The minimum allowed ambient temperature is 5°C (41°F) and the maximum allowed temperature is 40°C (104°F).

For more information about cooling requirements, see *IBM Z 8561 Installation Manual for Physical Planning*, GC28-7002.

Radiator (air) cooling

The following radiator (air) cooling options are available:

- ▶ A-Frame radiator air-cooled feature code (FC 4033)
- ▶ B-Frame radiator air-cooled feature code (FC 4035)

The radiator cooling system requires chilled air to fulfill the air-cooling requirements. z15 system airflow is from the front (intake, chilled air) to the rear (exhausts, warm air) of the frames. The chilled air is provided through perforated floor panels in front of the system

The hot and cold airflow and the arrangement of server aisles are shown in Figure 11-3.

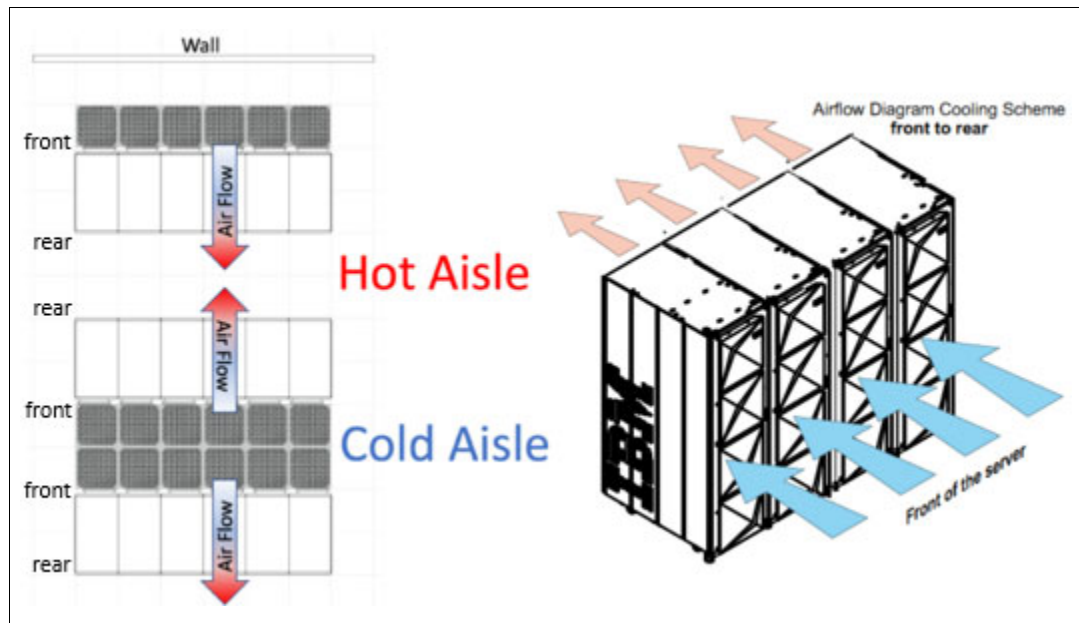


Figure 11-3 Hot and cold aisles

As shown in Figure 11-3, rows of servers must be placed front-to-front. Chilled air is provided through perforated floor panels that are placed in rows between the fronts of servers (the cold aisles). Perforated tiles generally are not placed in the hot aisles. If your computer room causes the temperature in the hot aisles to exceed a comfortable temperature, add as many perforated tiles as necessary to create a satisfactory comfort level. Heated exhaust air exits the computer room above the computing equipment.

For more information about the requirements for air-cooling options, see *IBM Z 8561 Installation Manual for Physical Planning*, GC28-7002.

Water-cooled system requirements

The following water-cooled options are available:

- ▶ A-Frame water-cooled feature code (FC 4034)
- ▶ B-Frame water-cooled feature code (FC 4036)

The water cooled 8561 system requires four or eight connections to the facility water, depending on the system configuration. Consider the following points:

- ▶ Systems with 1 - 3 CPC drawers require two feeds and two returns in the rear of the A frame.
- ▶ Systems with more than three CPC drawers require two feeds and two returns in the rear for both the A and the B frames.

These connections are made by using hoses that are fixed to the facility plumbing. They are routed up through the rear tailgate of the machine and terminated by using quick connect couplings.

Hoses and a means to fasten them to the facility plumbing are provided with the server.

The water cooled 8561 systems with one to three CPC drawers feature one water cooling assembly (WCA). The water cooled 8561 systems with four or five CPC drawers include two WCAs. Each WCA contains two fully redundant water control units (WCUs). These water control units each have their own facility feed and return water connections. If water is interrupted to one of the WCUs, the other water control unit assumes the entire load and the server continues to operate without interruption.

Therefore, each water connection to the facility plumbing must support the entire flow requirement for the WCA. If water is lost to both WCUs, the system attempts to reject heat by using the inner door heat exchangers in each frame and increasing system blower speeds. The server can also run in a degraded mode during this event.

Raised floor: The minimum raised floor height for a water-cooled system is 22.86 cm (8.6 in.).

These connections are made by using hoses that are fixed to the facility plumbing and are routed up through the rear tailgate of the system. They end with quick connect couplings.

Before you install z15 servers with water-cooled option, your facility must meet following requirements:

- ▶ Total water hardness cannot exceed 200 mg/L of calcium carbonate.
- ▶ The facility water pH is 7 - 9.
- ▶ Turbidity is less than 10 Nephelometric Turbidity Units (NTUs).
- ▶ Bacteria is less than 1000 colony-forming units (CFUs)/ml.
- ▶ The water is as free of particulate matter as feasible.
- ▶ The allowable system inlet water temperature range is 6°C - 20°C (43°F - 68°F) by using standard building chilled water. A special water system is typically not required.

- ▶ The flow rate to the frame is 3.7 - 79.4 lpm (1 - 21 gpm), depending on the inlet water temperature and the number of processor drawers in the z13 server. Colder inlet water temperatures require less flow than warmer water temperatures. Fewer processor drawers require less flow than a maximum populated z13 server.
- ▶ The minimum water pressure across the IBM hose ends is 0.34 - 2.32 BAR (5 - 33.7 psi), depending on the minimum flow required.

The maximum water pressure that is supplied at the IBM hose connections to the client's water supply cannot exceed 6.89 BAR (100 psi).

Supply hoses for the water-cooled systems

The z15 water-cooled system includes a customer "kit" that contains the hoses, clamps, and necessary fittings to install the supply and return hoses that connect to the WCUs in the system. Included are 4.2 m (13.7 ft) water hoses. The WCU water supply connections are shown Figure 11-4.

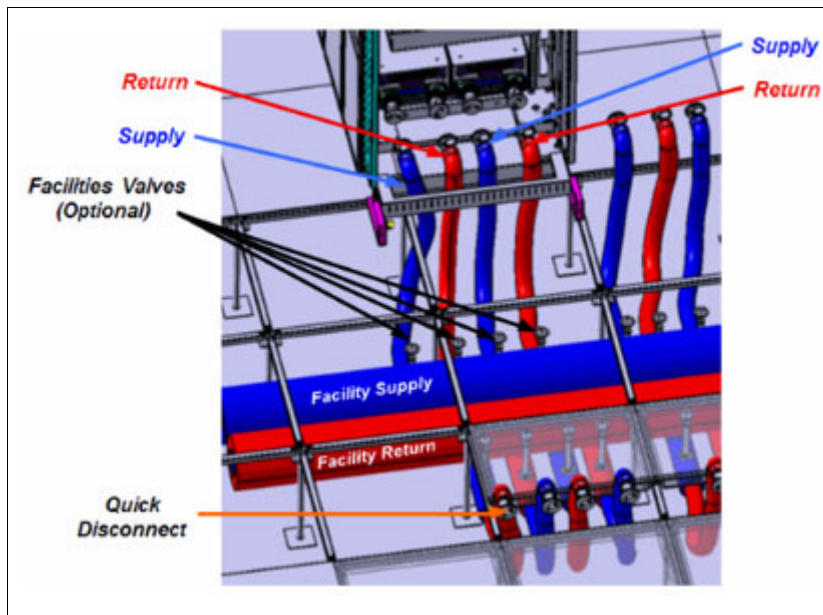


Figure 11-4 WCU water supply connections

The client's ends of the hoses are left open, which allows you to cut the hose to a custom length. An insulation clamp is provided to secure the insulation and protective sleeving after you cut the hose to the correct length and install it onto your plumbing.

Important: Use shut-off valves in front of the hoses. This configuration allows for the removal of the hoses for a service procedure or relocation. Valves are not included in the order. A stainless steel fitting is available for ordering. The fitting is barbed on one side and has a 2.54 cm (1 in.) male national pipe thread (NPT).

For more information about the tools that are needed for the water supply connections, see *IBM Z 8561 Installation Manual for Physical Planning*, GC28-7002.

11.1.4 Internal Battery Feature

Removal of IBF support^a: IBM z15 is planned to be the last IBM Z server to offer an Internal Battery Feature (IBF). As client data centers continue to improve power stability and uninterruptible power supply (UPS) coordination, IBM Z continues to innovate to help clients take advantage of common power efficiency and monitoring across their ecosystems. Additional support for data center power planning can be requested through your IBM Sales contact.

- a. Statements by IBM regarding its plans, directions, and intent are subject to change or withdrawal without notice at the sole discretion of IBM.

If a power shutdown occurs, the optional Internal Battery Feature (IBF) provides sustained system operations for a relatively short time, which allows a proper z15 server's shutdown. In addition, an external uninterrupted power supply system can be connected, which allows for longer periods of sustained operation.

Attention: The optional Internal Battery Feature (FC 3217) contains lithium ion batteries, which are packaged separately from the frame. A pair of batteries weighs approximately 43.5 kg (96 lb), as each individual battery weighs approximately 21.8 kg (48 lb).

Because of the hazardous nature of these Lithium ion batteries, only a certified Dangerous Goods Transportation employee is authorized to package and ship the IBFs. In the event of a system relocation, this restriction must be considered and is the responsibility of the client.

For more information, see *IBM Z 8561 Installation Manual for Physical Planning*, GC28-7002.

The IBF (when installed) can provide emergency power for the estimated time that is listed in Table 11-11. The number of IBFs depends on the number of BPRs. For the number of BPRs that are installed in relation to I/O units and the number of CPC drawers, see Table 11-8 on page 459. They are installed in pairs. You can have two, four, or six batteries (odd numbers are not allowed) per BPA.

Table 11-11 Battery hold-up times (minutes) for IBM z15 Model T01

| CPC Feature | number of PCIe+ I/O drawers | | | | | | | | | | | |
|-------------------|-----------------------------|---|----|----|---|---|---|---|---|---|----|----|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| | IBF hold-up time (minutes) | | | | | | | | | | | |
| Max34 FC 0655 | 10 | 7 | 12 | 10 | 9 | 8 | 7 | 7 | 7 | 7 | 7 | 7 |
| Max71 FC 0656 | 11 | 9 | 8 | 7 | 6 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| Max108 FC 0657 | 7 | 6 | 9 | 8 | 8 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| Max145 FC 0658 | 7 | 7 | 6 | 8 | 8 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| Max190 FC 0659 | 7 | 7 | 6 | 9 | 8 | 8 | 7 | 7 | 7 | 7 | 7 | 7 |

| CPC Feature | number of PCIe+ I/O drawers | | | | | | | | | | | |
|--|-----------------------------|---|---|---|---|---|---|---|---|---|----|----|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| | IBF hold-up time (minutes) | | | | | | | | | | | |
| <p>Notes: Consider the following points:</p> <ul style="list-style-type: none"> ▶ The numbers that are shown assume both BPAs feeding power with the batteries charged at full capacity ▶ Hold-up times are influenced by temperature, battery age, and fault conditions within the system. | | | | | | | | | | | | |

Consideration: The system holdup times that are listed in Table 11-11 on page 464 assume that both sides are functional and have fresh batteries under normal room ambient conditions.

Holdup times are greater for configurations that do not have every I/O slot plugged, the maximum installed memory, and are not using the maximum processors.

These holdup times are estimates. Your particular battery holdup time for any specific circumstance might be different.

Holdup times vary depending on the number of BPRs that are installed. As the number of BPRs increases, the holdup time also increases until the maximum number of BPRs is reached. After six BPRs (three per side) are installed, no other batteries are added; therefore, the time decreases from that point.

Holdup times for actual configurations are provided in the power estimation tool at [IBM Resource Link website](#).

On the Resource Link page, click **Tools** → **Machine information**. Then, select your IBM Z system and click **Power Estimation Tool**.

The BPR plugging order in the Bulk Power Enclosure is shown in Figure 11-5. Only the A frame is shown, but the same order is used in the B or C racks if present. Consider the following points:

- ▶ BPRs (and IBF units when the feature is installed) always are plugged in pairs.
- ▶ The number of IBFs (when installed) always equals the number of BPRs.

The plug order of BPRs is shown Figure 11-5.

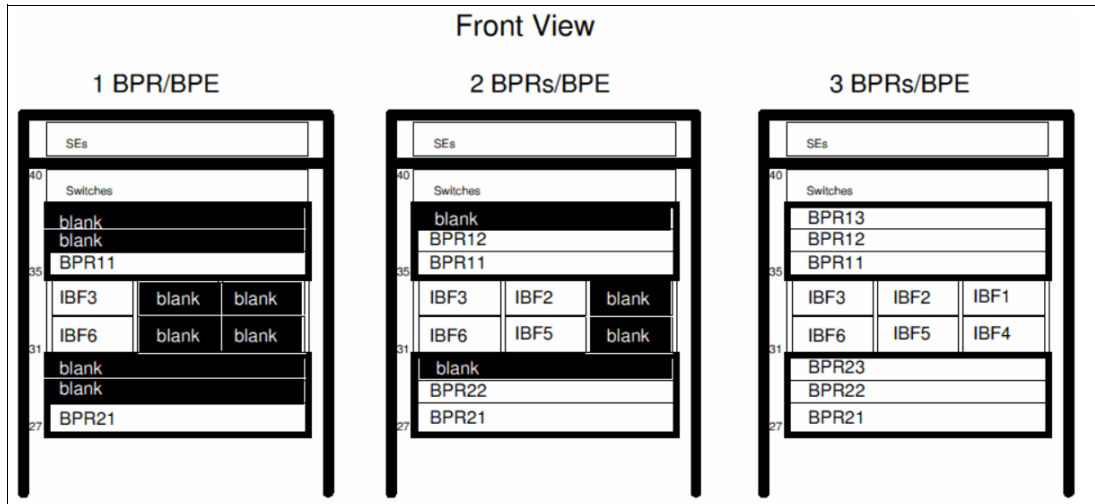


Figure 11-5 BPR / IBF (when present) plugging order

11.2 Physical specifications

This section describes the weights and dimensions of z15 server. The z15 is the first IBM Z enterprise class server that is based on a 19-inch rack; therefore, dimensions and weights are significantly different from previous generations of IBM Z servers.

The z15 can be installed on a raised or non-raised floor. For more information about weight distribution and floor loading tables, see *IBM 8561 Installation Manual for Physical Planning*, GC28-7002. This data is used with the maximum frame weight, frame width, and frame depth to calculate the floor loading.

Weight estimates for the maximum system configurations on the 8561 are listed in Table 11-12.

Table 11-12 Weights for maximum machine configurations (PDU)

| Frame configurations | Individual frame weights kg (lbs) | | | | Total weight kg (lbs) |
|----------------------|--------------------------------------|----------------------|----------------------|----------------------|--------------------------|
| | Z Frame | A Frame | B Frame | C Frame | |
| A | - | 795 kg (1753 lbs) | - | - | 795 kg (1753 lbs) |
| ZA | 711 kg (1568 lbs) | 775 kg (1709 lbs) | - | - | 1486 kg (3277 lbs) |
| ZAC | 711 kg (1568 lbs) | 756 kg (1667 lbs) | - | 740 kg (1632 lbs) | 2207 kg (4866 lbs) |
| AB | - | 775 kg (1709 lbs) | 725 kg (1599 lbs) | - | 1500 kg (3308 lbs) |
| ZAB | 711 kg (1568 lbs) | 756 kg (1667 lbs) | 725 kg (1599 lbs) | - | 2192 kg (4833 lbs) |
| ZABC | 711 kg (1568 lbs) | 756 kg (1667 lbs) | 705 kg (1555 lbs) | 641 kg (1413 lbs) | 2813 kg (6203 lbs) |

| Frame configurations | Individual frame weights kg (lbs) | | | | Total weight kg (lbs) |
|---|--------------------------------------|---------|---------|---------|--------------------------|
| | Z Frame | A Frame | B Frame | C Frame | |
| Notes: Consider the following points: <ul style="list-style-type: none"> ▶ Weight is based on the maximum system configuration. ▶ All weights are approximate and do not include Earthquake Kit hardware. ▶ Be certain that the raised floor on which you are installing the server can support the weight. | | | | | |

The power and weight estimation tool for Z servers on Resource Link covers the estimated weight for your designated configuration. It is available on [IBM Resource Link website](#).

On the Resource Link page, click **Tools** → **Power and weight estimation**.

11.3 Physical planning

This section describes the floor mounting, power, and I/O cabling options. For more information, see *IBM 8561 Installation Manual for Physical Planning*, GC28-7002.

11.3.1 Raised floor or non-raised floor

z15 servers can be installed on a raised or non-raised floor. The water-cooled models require a raised floor.

Note: On the z15, all I/O cabling and power cords come from the rear of the machine; therefore, all related features for Bottom and Top Exit cabling are in the rear of the frame.

Raised floor

If the z15 server is installed in a raised floor environment, air-cooled and water-cooled models are supported. You can select top exit features to manage I/O cables from the top frame of the z15 server.

The following top exit options are available for z15 servers:

- ▶ Top Exit I/O cabling feature code (FC 7917)
- ▶ Bottom Exit cabling feature code (FC 7919)
- ▶ Top Exit Cabling without Tophat feature code (FC 7928)

11.3.2 Top Exit cabling feature (optional)

The optional Top Exit cabling feature (FC 7917) allows for I/O cabling and power cords to exit the top of the frame. This feature adds cable management options, such as trunking and retainer brackets, as shown in Figure 11-6. The Top Exit cabling feature can be placed as shown in Figure 11-6, with the exit area towards the front of the frame, or with the exit area towards the rear of the frame.

Note: The feature provides the same extra hardware for every frame in the configuration.

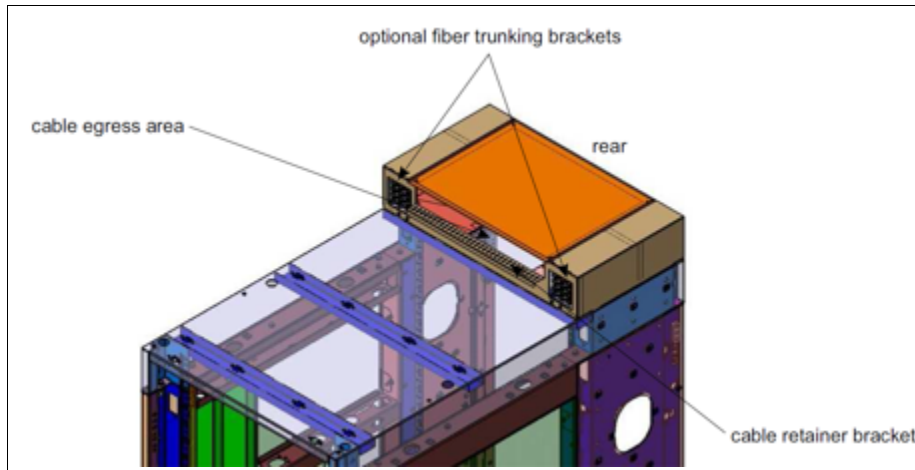


Figure 11-6 Top Exit cabling feature

The Top Exit cabling feature adds 117.5 mm (4.63 in.) to the height of the frame and approximately 5.4 kg (12 lbs) to the weight.

If the Top Exit cabling feature is not ordered, two sliding plates are available on the top of the frame (one on each side of the rear of the frame) that can be partially opened. By opening these plates, I/O cabling and power cords can exit the frame. The plates should be removed to install the Top Exit cabling feature as shown in Figure 11-7 on page 468.

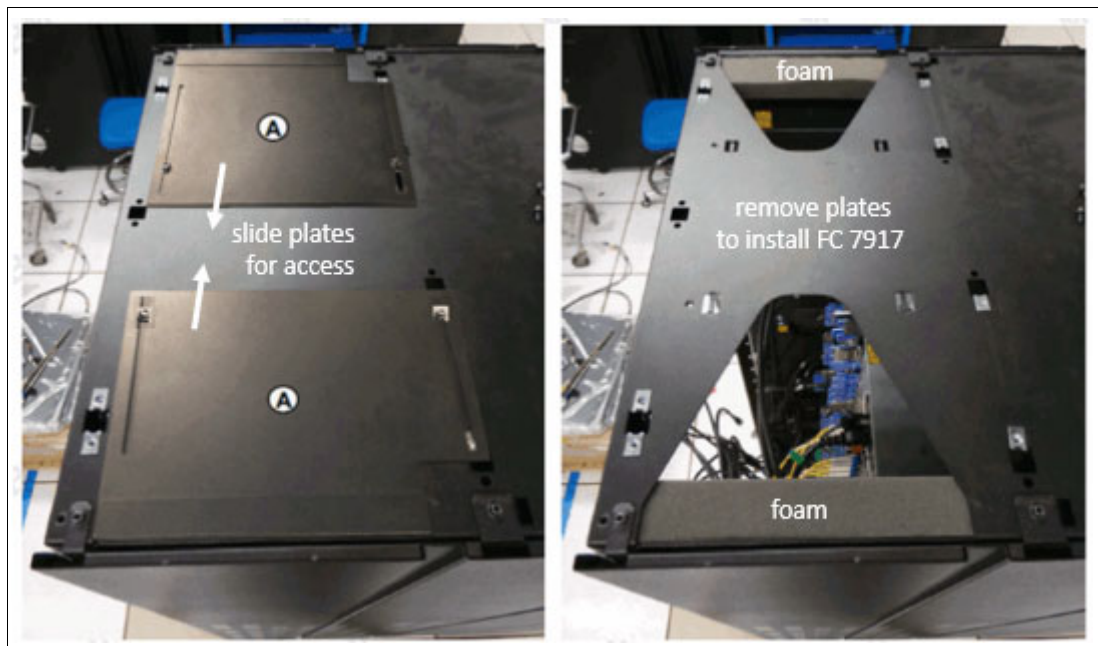


Figure 11-7 Top Exit access plates

11.3.3 Top or bottom exit cables

Features allow for Top Exit Cabling (FC 7917) or Bottom Exit Cabling (FC 7919) cabling, or a combination of both. These features are independent of raised floor or non-raised floor installations and offer flexible possibilities for the data center.

All external cabling enters the rear of the rack from under floor or from above the rack. Different from previous Z Systems, no cabling access or cable plugging is available at the front of the rack. The top view of the rack with and without FC 7917 is shown in Figure 11-8.

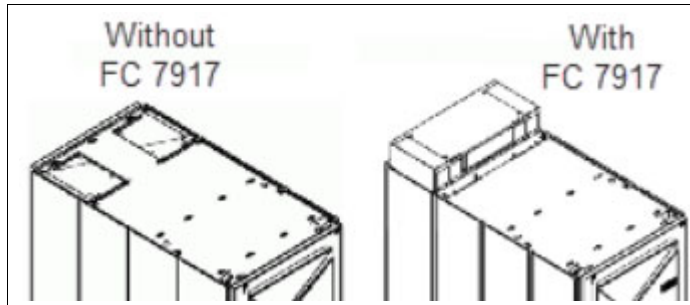


Figure 11-8 With and without FC 7917

The Top Exit Cabling feature provides new hardware. The new hardware resembles a rectangular box with an open side that faces the front or rear of the rack. It includes other hardware to organize and fasten cables.

The Top Exit Cabling option can be used for routing power cables and IO cables out the top of the machine.

Without the Top Exit Cabling feature, power and cables still can be run out the top of the rack through two adjustable openings at the top rear of the rack, as shown on the left side of Figure 11-8.

The Bottom Exit Cabling feature provides tailgate hardware for routing power cables or IO cables out the bottom of the machine.

For more information, see *IBM 8561 Installation Manual for Physical Planning*, GC28-7002, and 11.3, “Physical planning” on page 467

11.3.4 Bottom Exit cabling feature

The Bottom Exit cabling feature (FC 7919) is required for raised floor environments, where I/O cabling or power cords must exit from the bottom of the frame. This feature includes the hardware to allow bottom exit, and other components for cable management and filler plates to preserve the recommended air circulation, as shown in Figure 11-9.

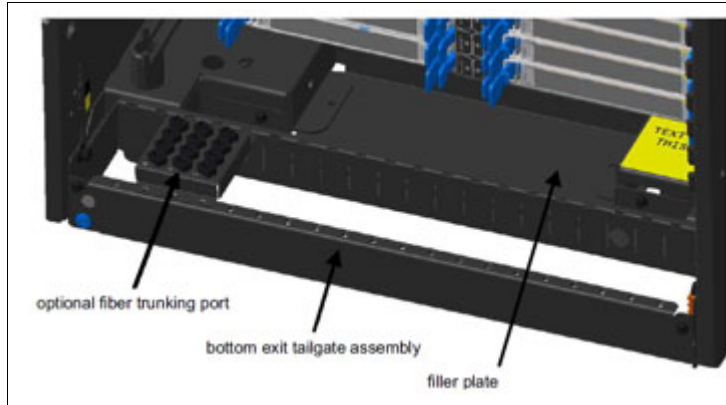


Figure 11-9 Bottom exit cabling feature

11.3.5 Frame Bolt-down kit

An Earthquake Kit, RF (FC 8010) is available for the z14 ZR1 servers. The kit provides hardware to enhance the ruggedness of the frame, the frame stiffener, and to tie down the frame to a concrete floor.

The frame tie-down kit can be used on a non-raised floor where the frame is secured directly to a concrete floor, or on a raised floor where the frame is secured to the concrete floor underneath the raised floor. Raised floors 241.3 mm (9.5 inches) - 1270 mm (50 inches) are supported.

The kits help secure the frames and their contents from damage when they are exposed to shocks and vibrations, such as in a seismic event. The frame tie-downs are intended for securing a frame that weighs up to 1308 kg (2885 lbs).

For more information, see *IBM 8561 Installation Manual for Physical Planning*, GC28-7002.

11.3.6 Service clearance areas

z15 servers require specific service clearance (see Figure 11-10 on page 471) to ensure the fastest possible repair in the unlikely event that a part must be replaced. Failure to provide enough clearance to open the front and rear covers results in extended service times or outages.

For more information, see *IBM 8561 Installation Manual for Physical Planning*, GC28-7002.

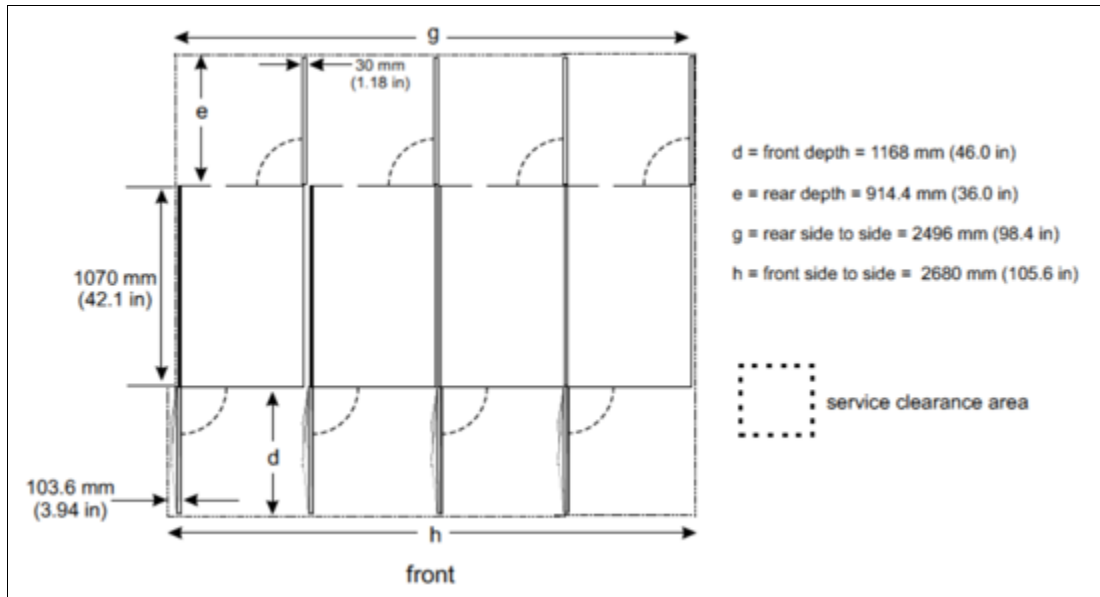


Figure 11-10 Maximum service clearance area (four frames)

11.4 Energy management

This section describes the elements of energy management to help you understand the requirements for power and cooling, monitoring and trending, and reducing power consumption. The energy management structure for the server is shown in Figure 11-11.

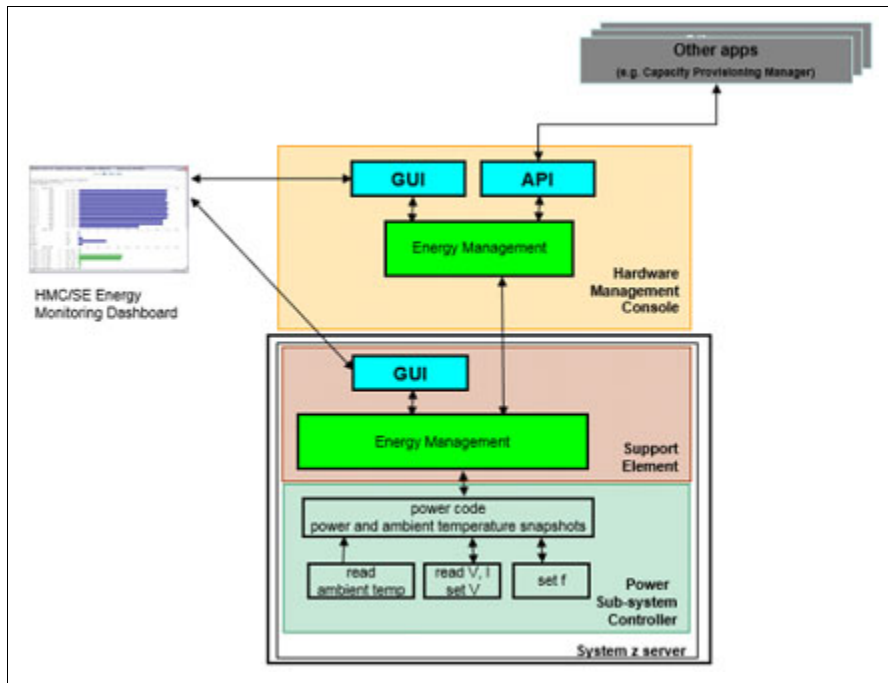


Figure 11-11 z15 Energy Management

The hardware components in the z15 server are monitored and managed by the energy management component in the Support Element (SE) and HMC. The graphical user interfaces (GUIs) of the SE and HMC provide views, such as the Monitors Dashboard, Environmental Efficiency Statistics, and Energy Optimization Advisor.

The following tools are available to plan and monitor the energy consumption of z15 servers:

- ▶ Power estimation tool on Resource Link
- ▶ Energy Optimization Advisor task for maximum potential power on HMC and SE
- ▶ Monitors Dashboard and Environmental Efficiency Statistics tasks on HMC and SE

11.4.1 Environmental monitoring

This section describes energy monitoring HMC and SE tasks.

Energy Optimization Advisor

This window is started from the HMC targeting the system and task under Energy Management. The window displays the following recommendations:

- ▶ Thermal Advice
- ▶ Processor Utilization Advice

Select the advice hyperlinks to provide specific recommendations for your system, as shown in Figure 11-12.



Figure 11-12 Energy Optimization Advisor

Monitors Dashboard task

In z15 servers, the Monitors Dashboard task in the Monitor task group provides a tree-based view of resources. Multiple graphical views display data, including history charts. This task monitors processor and channel usage. It produces data that includes power monitoring information, power consumption, and the air input temperature for the server.

An example of the Monitors Dashboard task is shown in Figure 11-13.

IBM Hardware Management Console

Home Monitors Dashboard - AR... X

Monitors Dashboard

Last refresh time: 03:28:45 AM Date: 09/01/19 Time zone: UTC-04:00 [Pause Refresh](#)

Overview

| Select | Name | Status | Type | Machine Type Model | Processor Usage(%) | I/O Usage(%) | Power Consumption (kW) (Btu/hr) | Ambient Temperature (°C) (°F) |
|-------------------------------------|-------|-----------|------|--------------------|--------------------|--------------|---------------------------------|-------------------------------|
| <input checked="" type="checkbox"/> | ARIES | Operating | CPC | 8561 - T01 | 4 | 1 | 9.842 33582 | |

Page 1 of 1 Max Page Size: 100 Total: 1 Filtered: 1 Displayed: 1 Selected: 1

Details

ARIES

Expand for CPC details

Close Help

Figure 11-13 Monitors Dashboard task

Environmental Efficiency Statistics task

The Environmental Efficiency Statistics task (see Figure 11-14) is part of the Monitor task group. It provides historical power consumption and thermal information for the CPC.

The data is presented in table format and graphical “histogram” format. The data can also be exported to a .csv-formatted file so that the data can be imported into a spreadsheet. For this task, you must use a web browser to connect to an HMC.

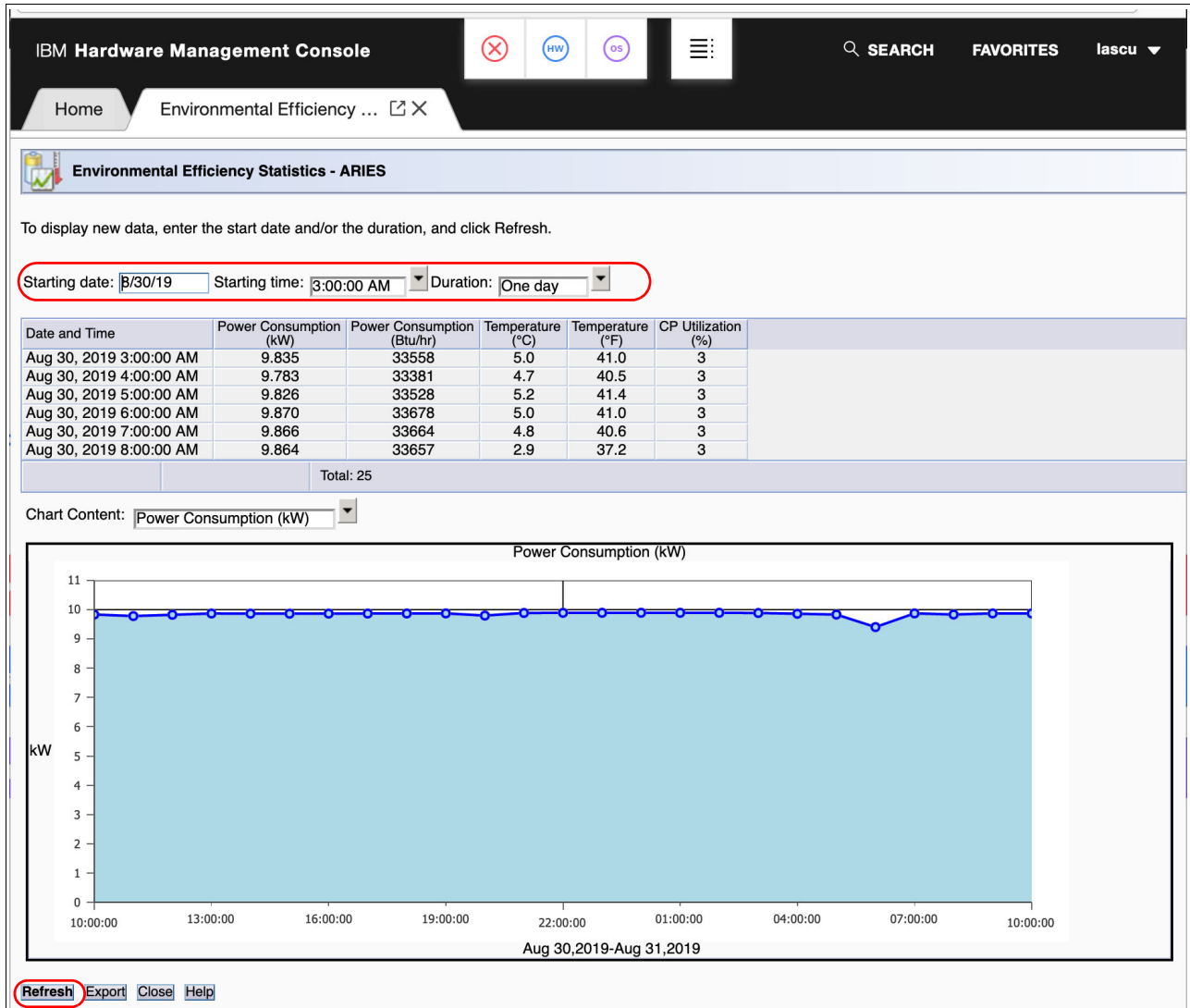


Figure 11-14 Environmental Efficiency Statistics task



Performance

This chapter describes the performance and capacity planning of z15.

Note: Throughout this chapter, *z15* refers to IBM z15 Model T01 (Machine Type 8561) unless otherwise specified.

This chapter includes the following topics:

- ▶ 12.1, “IBM z15 performance characteristics” on page 476
- ▶ 12.2, “z15 Large System Performance Reference ratio” on page 478
- ▶ 12.3, “Fundamental components of workload performance” on page 479
- ▶ 12.4, “Relative Nest Intensity” on page 481
- ▶ 12.5, “LSPR workload categories based on RNI” on page 483
- ▶ 12.6, “Relating production workloads to LSPR workloads” on page 483
- ▶ 12.7, “CPU MF counter data and LSPR workload type” on page 484
- ▶ 12.8, “Workload performance variation” on page 485
- ▶ 12.9, “Capacity planning consideration for z15” on page 485

12.1 IBM z15 performance characteristics

The IBM z15 Model T01 Feature Max190 (7J0) is designed to offer up to 25% more capacity and a 25% increase in the amount of memory that is compared to an IBM z14 Model M05 (7H0) system.

Uniprocessor performance also increased. On average, a z15 Model 701 offers performance improvements of more than 12% over the z14 Model 701. Figure 12-1 shows a system performance comparison of successive IBM Z servers.

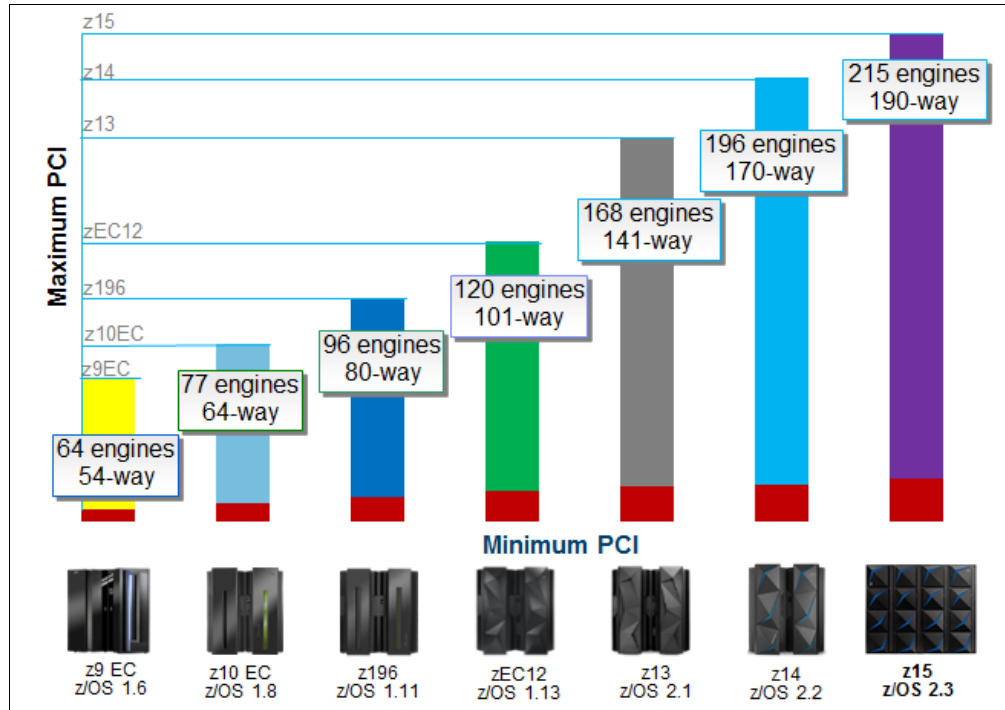


Figure 12-1 System performance comparison of successive IBM Z servers

Note: PCI = Processor Capacity Index.

Operating system support varies for the number of “engines” that are supported.

12.1.1 z15 single-thread capacity

The z15 processor chip runs at 5.2 GHz clock speed, which is the same as the z14 processor chip, but the performance is increased. For a uniprocessor, it increases 10 - 13% on average. For N-way processors model, it increases 12 - 14% on average at equal N-way configuration. These numbers differ depending on the workload type and LPAR configuration.

12.1.2 z15 SMT capacity

As with z13 and z14, customers can choose to run two threads on IFL and zIIP cores by using SMT mode. SMT increases throughput by 10 - 40% (average 25%), depending on workload.

12.1.3 IBM Integrated Accelerator for zEnterprise Data Compression

Starting with z13, IBM introduced the zEnterprise Data Compression (zEDC) Express PCIe feature, bringing efficiency and economies for data storing and data transfers.

The zEDC Express feature was adopted by enterprises because it helps with software costs for compression/decompression operations (by offloading these operations), and increases data encryption (compression before encryption) efficiency.

With z15, the zEDC Express functionality was moved off from the PCIe infrastructure into the processor nest. By moving the compression and decompression into the processor nest (on-chip), IBM z15 processor provides a new level of performance for these tasks and eliminates the need for the zEDC Express feature virtualization. It also brings new use cases to the platform.

For more information, see Appendix C, “IBM Integrated Accelerator for zEnterprise Data Compression” on page 509.

12.1.4 Primary performance improvement drivers with z15

The attributes and design points of z15 contribute to overall performance and throughput improvements as compared to the z14. The following major items contribute to z15 performance improvements:

- ▶ z15 microprocessor architecture:
 - Larger GCT (60x3 entries versus 48x3 in z14, 1.25x entries)
 - New Mapper Design (2x entries)
 - Larger Issue Queue (2x36 entries versus 2x30 in z14)
 - 4x sized 2-gig TLB2 (256 entries versus 64 in z14)
 - Doubled # of core FARs (12 versus 6 in z14)
 - Double sized BTB1 (16 K entries versus 8 K in z14, Improved prediction)
 - New TAGE-based PHT branch predictor design
 - Pipeline optimization
 - Third-generation SMT processing for zIIPs and IFLs
- ▶ Cache:
 - L2 I-Cache increased from 2 MB to 4 MB per Core, or 2x
 - L3 Cache increased from 128 MB to 256 MB per CP chip, or 2x
 - L4 Cache increased from 672 MB to 960 MB, or +43%
 - New power efficient logical directory design
- ▶ Storage hierarchy:
 - Reduced cache latencies (L2 → L3, L3 → L4)
 - Improved system protocols (Reduced contention points)
 - Core stores target L3 cache compartment directly
 - Improved on-chip data by using (shared by 3 versus 5 on z14)
 - On-cluster memory fetches only cached in L3 (LRU write-back to L4)
 - Bus speeds and feeds Improvements (meso-sync buses)
 - Nest Early Eviction fetch hints (non-mru caching of temporal data)
 - Improved hot cache line handling (contention affinity, single SC)
- ▶ Software and hardware:
 - Drawer-based memory affinity
 - z/OS HiperDispatch Optimizations
 - WiseOps, z/OS MicroTrend Analysis
 - PR/SM Algorithm Improvements (placement, ICF relocation)

- Hot Cache line handling improvements
- Post Quantum Encryption
- Speed Boost:
- ▶ z/Architecture implementation:
 - DEFLATE-Conversion Facility (on chip compression)

Replaces zEDC Express PCIe feature. DEFLATE CONVERSION CALL (DFLTCC) instruction can compress and decompress data by using the DEFLATE standard.
 - Move-Page-and-Set-Key Facility

Enables setting the storage key that is associated with the page to which is being stored.
 - PER Storage-Key-Alteration Facility

Identify changes to ACC (access key) and F (fetch protect) bits of storage keys.
 - Message-Security-Assist Extension 9

Add support to CPACF for performing digital signatures.
 - Vector-Enhancements Facility 2:

Provides eight new instructions to help deal with endian conversions.
 - Adapter CPU Directed Interrupts

New architecture to allow native PCI devices to present interrupts that are directed to a specific CPU.
 - PCI Mapped I/O (MIO) Address Space

Allows for an operating system to enable a problem state program to access PCI memory.
 - Miscellaneous-Instruction-Extensions Facility 3

12.2 z15 Large System Performance Reference ratio

The Large System Performance Reference (LSPR) provides capacity ratios among various processor families that are based on various measured workloads. It is a common practice to assign a capacity scaling value to processors as a high-level approximation of their capacities.

For z/OS V2R3 studies, the capacity scaling factor that is commonly associated with the reference processor is set to a 2094-701 with a Processor Capacity Index (PCI) value of 593. This value is unchanged since z/OS V1R11 LSPR. The use of the same scaling factor across LSPR releases minimizes the changes in capacity results for an older study and provides more accurate capacity view for a new study.

Performance data for z15 servers were obtained with z/OS V2R3 (running Db2 for z/OS V12, CICS TS V5R3, IMS V14, Enterprise COBOL V6R2, and WebSphere Application Server for z/OS V9.0.0.8). All IBM Z server generations are measured in the same environment with the same workloads at high usage.

Note: If your software configuration is different from what is described here, the performance results might vary.

On average, z15 servers can deliver up to 25% more performance in a 190-way configuration than an z14 170-way. However, the observed performance increase varies depending on the workload type.

Consult the LSPR when you consider performance on the z15. The range of performance ratings across the individual LSPR workloads is likely to include a large spread. Performance of the individual logical partitions (LPARs) varies depending on the fluctuating resource requirements of other partitions and the availability of processor units (PUs). Therefore, it is important to know which LSPR workload type suite your production environment. For more information, see 12.8, “Workload performance variation” on page 485.

For more information about performance, see the [Large Systems Performance Reference for IBM Z page](#) of the Resource Link website.

For more information about millions of service units (MSU) ratings, see the [IBM z Systems Software Contracts page](#) of the IBM IT infrastructure website.

12.2.1 LSPR workload suite

Historically, LSPR capacity tables, including pure workloads and mixes, were identified with application names or a *software* characteristic; for example, CICS, IMS, OLTP-T,¹ CB-L,² LoIO-mix,³ and TI-mix.⁴ However, capacity performance is more closely associated with how a workload uses and interacts with a particular processor *hardware* design.

The CPU Measurement Facility (CPU MF) data that was introduced on the z10 provides insight into the interaction of workload and *hardware design* in production workloads. CPU MF data helps LSPR to adjust workload capacity curves that are based on the underlying hardware sensitivities; in particular, the processor access to caches and memory. This processor access to caches and memory is called *nest*. By using this data, LSPR introduces three workload capacity categories that replace all older primitives and mixes.

LSPR contains the internal throughput rate ratios (ITRRs) for the z15 and the previous generation processor families. These ratios are based on measurements and projections that use standard IBM benchmarks in a controlled environment.

The throughput that any user experiences can vary depending on the amount of multiprogramming in the user’s job stream, the I/O configuration, and the workload processed. Therefore, no assurance can be given that an individual user can achieve throughput improvements that are equivalent to the performance ratios that are stated.

12.3 Fundamental components of workload performance

Workload performance is sensitive to the following major factors:

- ▶ Instruction path length
- ▶ Instruction complexity
- ▶ Memory hierarchy and memory nest

These factors are described next.

¹ Traditional online transaction processing workload (formerly known as IMS).

² Commercial batch with long-running jobs.

³ Low I/O Content Mix Workload.

⁴ Transaction Intensive Mix Workload.

12.3.1 Instruction path length

A transaction or job runs a set of instructions to complete its task. These instructions are composed of various paths through the operating system, subsystems, and application. The total count of instructions that are run across these software components is referred to as the *transaction or job path length*.

The path length varies for each transaction or job, and depends on the complexity of the tasks that must be run. For a particular transaction or job, the application path length tends to stay the same, assuming that the transaction or job is asked to run the same task each time.

However, the path length that is associated with the operating system or subsystem can vary based on the following factors:

- ▶ Competition with other tasks in the system for shared resources. As the total number of tasks grows, more instructions are needed to manage the resources.
- ▶ The number of logical processors (*n-way*) of the image or LPAR. As the number of logical processors grows, more instructions are needed to manage resources that are serialized by latches and locks.

12.3.2 Instruction complexity

The type of instructions and the sequence in which they are run interacts with the design of a microprocessor to affect a performance component. This factor is defined as *instruction complexity*. The following design alternatives affect this component:

- ▶ Cycle time (GHz)
- ▶ Instruction architecture
- ▶ Pipeline
- ▶ Superscalar
- ▶ Out-of-order execution
- ▶ Branch prediction
- ▶ Transaction Lookaside Buffer (TLB)
- ▶ Transactional Execution (TX)
- ▶ Single instruction multiple data instruction set (SIMD)
- ▶ Simultaneous multithreading (SMT)⁵

As workloads are moved between microprocessors with various designs, performance varies. However, when on a processor, this component tends to be similar across all models of that processor.

12.3.3 Memory hierarchy and memory nest

The *memory hierarchy* of a processor generally refers to the caches, data buses, and memory arrays that stage the instructions and data that must be run on the microprocessor to complete a transaction or job.

The following design choices affect this component:

- ▶ Cache size
- ▶ Latencies (sensitive to distance from the microprocessor)
- ▶ Number of levels, the Modified, Exclusive, Shared, Invalid (MESI) protocol, controllers, switches, the number and bandwidth of data buses, and so on.

⁵ Only available for IFL, zIIP, and SAP processors

Certain caches are *private* to the microprocessor core, which means that only that microprocessor core can access them. Other caches are shared by multiple microprocessor cores. The term *memory nest* for an IBM Z processor refers to the shared caches and memory along with the data buses that interconnect them.

A memory nest in a z15 CPC drawer is shown in Figure 12-2.

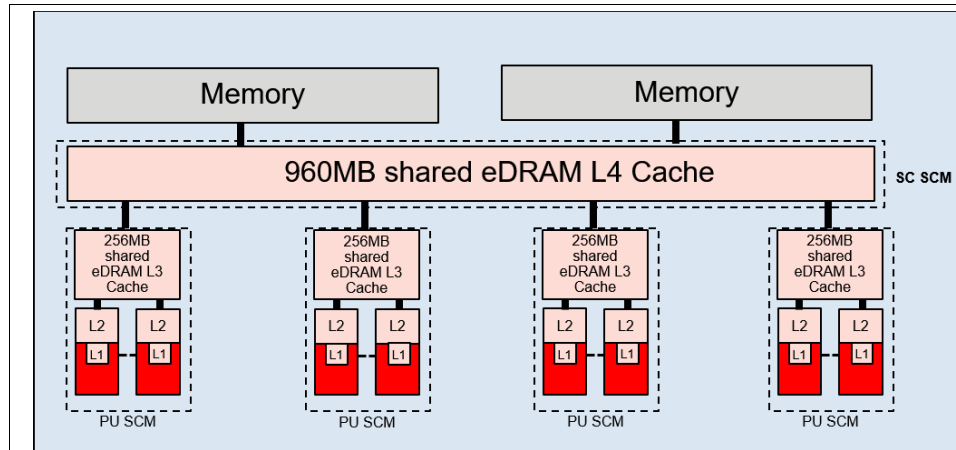


Figure 12-2 Memory hierarchy in a z15 CPC drawer

Workload performance is sensitive to how deep into the memory hierarchy the processor must go to retrieve the workload instructions and data for running. The best performance occurs when the instructions and data are in the caches nearest the processor because little time is spent waiting before running. If the instructions and data must be retrieved from farther out in the hierarchy, the processor spends more time waiting for their arrival.

As workloads are moved between processors with various memory hierarchy designs, performance varies because the average time to retrieve instructions and data from within the memory hierarchy varies. Also, when on a processor, this component continues to vary because the location of a workload's instructions and data within the memory hierarchy is affected by several factors that include, but are not limited to, the following factors:

- ▶ Locality of reference
- ▶ I/O rate
- ▶ Competition from other applications and LPARs

12.4 Relative Nest Intensity

The most performance-sensitive area of the memory hierarchy is the activity to the memory nest. This area is the distribution of activity to the shared caches and memory.

The term *Relative Nest Intensity* (RNI) indicates the level of activity to this part of the memory hierarchy. By using data from CPU MF, the RNI of the workload that is running in an LPAR can be calculated. The higher the RNI, the deeper into the memory hierarchy the processor must go to retrieve the instructions and data for that workload.

RNI reflects the distribution and latency of sourcing data from shared caches and memory, as shown in Figure 12-3.

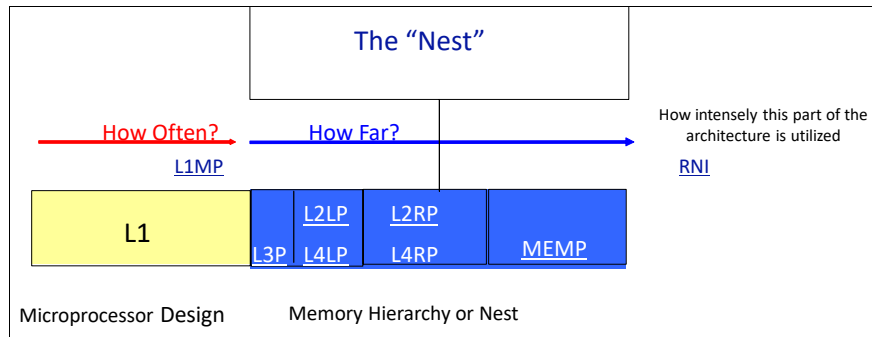


Figure 12-3 Relative Nest Intensity

Many factors influence the performance of a workload. However, these factors often are influencing the RNI of the workload. The interaction of all these factors results in a net RNI for the workload, which in turn directly relates to the performance of the workload.

These factors are tendencies, not absolutes. For example, a workload might have a low I/O rate, intensive processor use, and a high locality of reference, which all suggest a low RNI. However, it might be competing with many other applications within the same LPAR and many other LPARs on the processor, which tends to create a higher RNI. It is the net effect of the interaction of all these factors that determines the RNI.

The traditional factors that were used to categorize workloads in the past are shown with their RNI tendency in Figure 12-4.

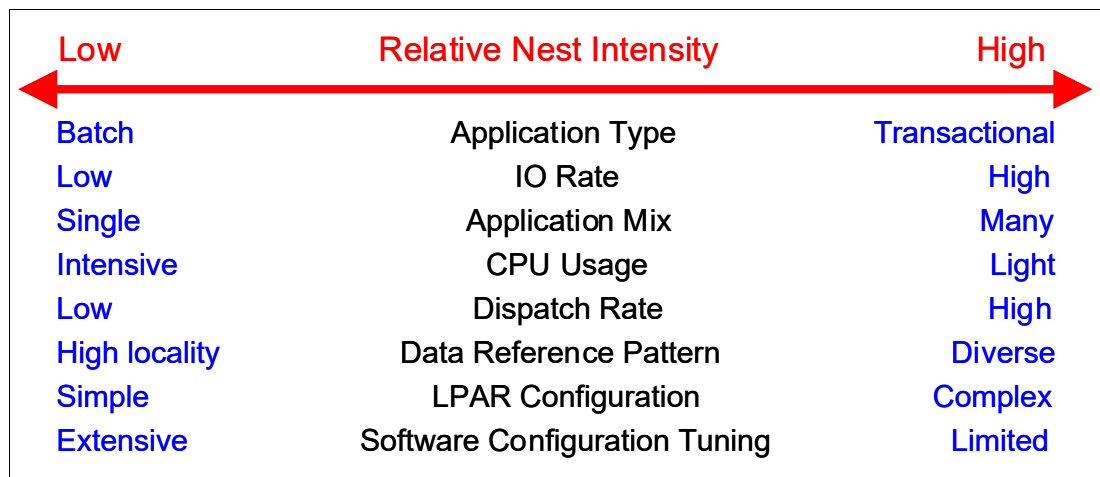


Figure 12-4 Traditional factors that were used to categorize workloads

Little can be done to affect most of these factors. An application type is whatever is necessary to do the job. The data reference pattern and processor usage tend to be inherent to the nature of the application. The LPAR configuration and application mix are mostly a function of what must be supported on a system. The I/O rate can be influenced somewhat through buffer pool tuning.

However, one factor, *software configuration tuning*, is often overlooked but can have a direct effect on RNI. This term refers to the number of address spaces (such as CICS application-owning regions (AORs) or batch initiators) that are needed to support a workload.

This factor always existed, but its sensitivity is higher with the current high frequency microprocessors. Spreading the same workload over more address spaces than necessary can raise a workload's RNI. This increase occurs because the working set of instructions and data from each address space increases the competition for the processor caches.

Tuning to reduce the number of simultaneously active address spaces to the optimum number that is needed to support a workload can reduce RNI and improve performance. In the LSPR, the number of address spaces for each processor type and *n-way* configuration is tuned to be consistent with what is needed to support the workload. Therefore, the LSPR workload capacity ratios reflect a presumed level of software configuration tuning. Retuning the software configuration of a production workload as it moves to a larger or faster processor might be needed to achieve the published LSPR ratios.

12.5 LSPR workload categories based on RNI

A workload's RNI is the most influential factor in determining workload performance. Other more traditional factors, such as application type or I/O rate, have RNI tendencies. However, it is the net RNI of the workload that is the underlying factor in determining the workload's performance. The LSPR now runs various combinations of former workload primitives, such as CICS, Db2, IMS, OSAM, VSAM, WebSphere, COBOL, and utilities, to produce capacity curves that span the typical range of RNI.

The following workload categories are represented in the LSPR tables:

- ▶ **LOW** (relative nest intensity)
A workload category that represents light use of the memory hierarchy.
- ▶ **AVERAGE** (relative nest intensity)
A workload category that represents average use of the memory hierarchy. This category is expected to represent most production workloads.
- ▶ **HIGH** (relative nest intensity)
A workload category that represents a heavy use of the memory hierarchy.

These categories are based on the RNI. The RNI is influenced by many variables, such as application type, I/O rate, application mix, processor usage, data reference patterns, LPAR configuration, and the software configuration that is running. CPU MF data can be collected by z/OS System Measurement Facility on SMF 113 records or z/VM Monitor starting with z/VM V5R4.

12.6 Relating production workloads to LSPR workloads

Historically, the following techniques were used to match production workloads to LSPR workloads:

- ▶ Application name (a client that is running CICS can use the CICS LSPR workload)
- ▶ Application type (create a mix of the LSPR online and batch workloads)
- ▶ I/O rate (the low I/O rates that are used a mix of low I/O rate LSPR workloads)

The IBM Processor Capacity Reference for IBM Z (zPCR) tool supports the following workload categories:

- ▶ Low
- ▶ Low-Average

- ▶ Average
- ▶ Average-high
- ▶ High

For more information about the no-charge IBM zPCR tool (which reflects the latest IBM LSPR measurements), see the [Getting Started with zPCR \(IBM's Processor Capacity Reference\)](#) page of the IBM Techdoc Library website.

As described in 12.5, “LSPR workload categories based on RNI” on page 483, the underlying performance sensitive factor is how a workload interacts with the processor hardware.

12.7 CPU MF counter data and LSPR workload type

Beginning with the z10 processor, the hardware characteristics can be measured by using CPU MF (SMF 113) counters data. A production workload can be matched to an LSPR workload category through these hardware characteristics.

For more information about RNI, see 12.5, “LSPR workload categories based on RNI” on page 483.

The AVERAGE RNI LSPR workload is intended to match most client workloads. When no other data is available, use the AVERAGE RNI LSPR workload for capacity analysis.

Low-Average and Average-High categories allow better granularity for workload characterization but these categories can apply on zPCR only.

The CPU MF data can be used determine workload type. When available, this data allows the RNI for a production workload to be calculated.

By using the RNI and another factor from CPU MF, the L1MP (percentage of data and instruction references that miss the L1 cache), a workload can be classified as LOW, AVERAGE, or HIGH RNI. This classification and resulting hit are automated in the zPCR tool. It is preferable to use zPCR for capacity sizing.

Starting with z/OS V2R1 with APAR OA43366, zFS file is not required any more for CPU MF and Hardware Instrumentation Services (HIS). HIS is a z/OS function that collects hardware event data for processors in SMF records type 113, and a z/OS UNIX System Services output files.

Only SMF 113 record is required to know proper workload type by using CPU MF counter data. CPU overhead of CPUMF is minimal. Also, the amount of SMF 113 record is 1% of typical SMF 70 and 72 which RMF writes.

CPU MF and HIS can use not only for deciding workload type but also use another purpose. For example, starting with z/OS V2R1, you can record Instruction Counts in SMF type 30 record when you activate CPU MF. Therefore, we strongly recommend that you *always* activate CPU MF.

For more information about getting CPUMF counter data, see the CPU MF - 2017 Update and WSC Experiences of the IBM Techdoc Library website.

12.8 Workload performance variation

As the size of transistors approaches the size of atoms that stand as a fundamental physical barrier, a processor chip's performance can no longer double every two years (Moore's Law⁶ does not apply).

A holistic performance approach is required when the performance gains are reduced because of frequency. Therefore, hardware and software synergy becomes an absolute requirement.

Starting with z13, Instructions Per Cycle (IPC) improvements in core and cache became the driving factor for performance gains. As these microarchitectural features increase (which contributes to instruction parallelism), overall workload performance variability also increases because not all workloads react the same way to these enhancements.

Because of the nature of the z15 multi-CPC drawer system and resource management across those drawers, performance variability from application to application is expected.

Also, the memory and cache designs affect various workloads in many ways. All workloads are improved, with cache-intensive loads benefiting the most. For example, having more PUs per CPC drawer, each with higher capacity than z14, more workload can fit on a z15 CPC drawer. This configuration can result in better performance. For example, z14 two drawer system model M02 can populate maximum 69 PUs.

In contrast, z15 two drawer system Max71 can populate maximum 71 PUs. Therefore, two more PUs can share caches and memories within the drawer, so the performance improvements is expected.

The workload variability for moving from z13 and z14 to z15 expected to be stable. Workloads that are migrating from z10 EC, z196, and zEC12 to z15 can expect to see similar results with slightly less variability than the migration from z13 and z14.

Experience demonstrates that IBM Z servers can be run at up to 100% utilization levels, sustained. However, most clients prefer to leave some room and run at 90% or slightly under.

12.9 Capacity planning consideration for z15

In this section, we describe recommended ways conduct capacity planning for z15.

Do not use MIPs or MSUs for capacity planning: Do *not* use “one number” capacity comparisons, such as MIPs or MSUs. IBM does not officially announce the processor performance as “MIPs”. MSU is only a number for software license charge and it does *not* represent for performance for the processor.

12.9.1 Collect CPU MF counter data

It is important to recognize the LSPR workload type of your production system. As described in 12.7, “CPU MF counter data and LSPR workload type” on page 484, the capacity of the processor is different from the LSPR workload type. By collecting the CPU MF SMF 113 record, you can recognize the workload type in a specific IBM-provided capacity planning tool. Therefore, collecting CPU MF counter data is a first step to begin the capacity planning.

⁶ For more information, see the [Moore's Law website](#).

12.9.2 Creating EDF file with CP3KEXTR

EDF file is an input file of the IBM Z capacity planning tool. You can create this file by using the CP3KEXTR program. The CP3KEXTR program reads SMF records and extracts needed data as input to IBM's Processor Capacity Reference (zPCR) and z Systems Batch Network Analyzer (zBNA) tools.

CP3KEXTR is offered as a “no-charge” application. It can also create the EDF file for ZCP3000. ZCP3000 is an IBM internal tool, but you can create the EDF file for it on your system. For more information about CP3KEXTR, see the IBM Techdoc *z/OS Data Extraction Program (CP3KEXTR) for zPCR and zBNA*.

12.9.3 Loading EDF file to the capacity planning tool

By loading EDF file to IBM capacity planning tool, you can see the LSPR workload type based on CPU MF counter data. Figure 12-5 shows a sample zPCR window of a workload type. In this example, the workload type displays in the “Assigned Workload” column. When you load the EDF file to zPCR, it automatically sets your LPAR configuration. It also makes easy to define the LPAR configuration to the zPCR.

The screenshot shows the 'Create LPAR Configuration from EDF' window. At the top, it displays the LPAR configuration from EDF, including the z/OS SMF Data Set Name (SAMPLE2827SMF), Extract Version (CP3KEXTR04/01/17), EDF File Name (C:\PSTOOL\KzPCREDF\Files\EDF\Sample zEC12.edf), Interval #1 (Date=2016-04-11 Time=00:00:00 Length=01:00:00), CPC ID (CPC0002, GP Processor Model = 2827-600), and zEC12 Host (2827-H43/600 with 8 CPs: GP=3 zIIP=1 IFL=2 ICF=2). Below this, it shows the 'Create LPAR Configuration' section with Configuration #1 and LPAR Host as specified above. The main part of the window is a table with columns for Copy LP, LP is Active, LP from EDF, Partition Identification (No., Type, Name, SCP), Assigned Workload, Partition Configuration (Mode, Total LCPs, Weight, Weight %), Capping (checked, ABS), HiperDispatch (Is Active, Parked LCPs, RNI), CPU-MF (Workload Assignment), and Method Used. The table lists 8 partitions with various types (GP, zIIP, IFL, ICF) and workloads (Average, Average/LV, CFCC). At the bottom, there are controls for Default SCP for GP Partitions (radio buttons for z/OS, z/VM, z/VM), IFL Partitions (radio buttons for z/VM), and checkboxes for 'Estimate parked LCPs where unknown for: GP partitions' and 'IFL partitions'. There are also buttons for 'Select All', 'Select Active', 'Remove All', and 'Choose Another EDF Interval', and a checkbox for 'Remove Parked LCPs from the LCP Count when copying partitions into zPCR'.

| Copy LP | LP is Active | LP from EDF | Partition Identification | | | Assigned Workload | Partition Configuration | | | Capping | HiperDispatch | | | CPU-MF | Method Used | |
|-------------------------------------|-------------------------------------|-------------------------------------|--------------------------|------|---------|-------------------|-------------------------|------|------------|---------|---------------|-------------------------------------|-----------|--------|-------------|-------------|
| | | | No. | Type | Name | | SCP | Mode | Total LCPs | | Weight | Weight % | Is Active | | | Parked LCPs |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 1 | GP | GP-001 | z/OS-2.2 | Average | SHR | 3.0 | 800 | 80.0% | <input checked="" type="checkbox"/> | ABS | 0.0 | 0.0 | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 2 | GP | GP-002 | z/OS-2.2 | Average | SHR | 1.0 | 10 | 33.3% | <input checked="" type="checkbox"/> | ABS | 0.0 | 0.0 | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 2 | GP | GP-002 | z/OS-2.2 | Average | SHR | 2.0 | 180 | 18.0% | <input checked="" type="checkbox"/> | ABS | | | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 3 | GP | GP-003 | z/OS-2.2 | Average | SHR | 1.0 | 10 | 33.3% | <input checked="" type="checkbox"/> | ABS | | | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 3 | GP | GP-003 | z/OS-2.2 | Average | SHR | 1.0 | 20 | 2.0% | <input checked="" type="checkbox"/> | ABS | | | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 4 | zIIP | GP-003 | z/OS-2.2 | Average | SHR | 1.0 | 10 | 33.3% | <input checked="" type="checkbox"/> | ABS | | | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 4 | IFL | IFL-001 | z/VM-6.4 | Average/LV | SHR | 2.0 | 90 | 90.0% | <input checked="" type="checkbox"/> | ABS | | | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 5 | IFL | IFL-002 | z/VM-6.4 | Average/LV | SHR | 2.0 | 10 | 10.0% | <input checked="" type="checkbox"/> | ABS | | | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 6 | ICF | ICF-001 | CFCC | CFCC | DED | 1.0 | n/s | | <input checked="" type="checkbox"/> | ABS | | | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 7 | ICF | ICF-002 | CFCC | CFCC | SHR | 1.0 | 90 | 90.0% | <input checked="" type="checkbox"/> | ABS | | | Default |
| <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 8 | ICF | ICF-003 | CFCC | CFCC | SHR | 1.0 | 10 | 10.0% | <input checked="" type="checkbox"/> | ABS | | | Default |

Figure 12-5 zPCR LPAR Configuration from EDF window

12.9.4 Tips to maximize z15 server capacity

The server capacity of the z15 can be maximized by using the following tips:

- ▶ Turn on HiperDispatch in every LPARs. Hiperdispatch optimizes processor cache usage by creating an affinity between a PU and the workload.
- ▶ Assign an appropriate number of logical CPs. If you assign too many logical CPs to the LPAR, unnecessary LPAR management cost is exhausted. This issue reduces the efficiency of the cache.

The server capacity declines relative to the LCP:RCP ratio (sum of logical CPs defined in all LPARs: the number of physical CPs on your configuration). Therefore, assigning the correct number of CPU to LPAR is important.

If your LPARs configuration LCP:RCP ratio reaches its limit, zPCR warns your configuration. Figure 12-6 shows a sample zPCR error message window when the practical LCP:RCP ratio is exceeded.

Partition Detail Report
 Based on LSPR Data for IBM Z Processors
 Study ID: Not specified
 #1 Configuration #1
 Description: Created from EDF EDFsample zEC12.edf for CPO0002 interval #1: Date=2016-04-11 Time=00:00:00
zEC12 Host = 2827-H43/600 with 8 CPs: GP=3 zIIP=1 IFL=2 ICF=2
98 Active Partitions: GP=47 zIIP=46 IFL=2 ICF=3
 Capacity basis: 2094-701 @ 1.000 ITRR for a shared single-partition configuration
 Capacity for z/OS on z10 and later processors is represented with HiperDispatch turned ON

| Include | Partition Identification | | | | Partition Configuration | | | | | | | | |
|-------------------------------------|--------------------------|------|--------|----------|-------------------------|------|-------------|--------|----------------|--------------------------|--------------------------|----------|---------|
| | No. | Type | Name | SCP | Assigned Workload | Mode | Active LCPs | Weight | Weight Percent | Capping | | Capacity | |
| <input checked="" type="checkbox"/> | 1 | GP | GP-001 | z/OS-2.2 | Average | SHR | 3 | 800 | | <input type="checkbox"/> | <input type="checkbox"/> | Minimum | Maximum |
| <input checked="" type="checkbox"/> | | zIIP | GP-001 | z/OS-2.2 | Average | SHR | 1 | 10 | | <input type="checkbox"/> | <input type="checkbox"/> | | |
| <input checked="" type="checkbox"/> | 2 | GP | GP-002 | z/OS-2.2 | Average | SHR | 3 | 180 | | <input type="checkbox"/> | <input type="checkbox"/> | | |
| <input checked="" type="checkbox"/> | | zIIP | GP-002 | z/OS-2.2 | Average | SHR | 1 | 10 | | <input type="checkbox"/> | <input type="checkbox"/> | | |

Table View Controls
 Display zAAP/zIIP/IFL Partitions
 With Associated GP Separate by Pool
 Show: GP Pool GP zAAP zIIP
 Includes Only IFL ICF

Capacity Summary by Pool

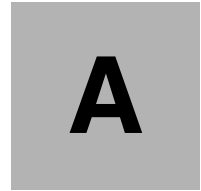
| GP Pool | Real CPs | LPs | DED LCPs | SHR | | Sum of Weights | Capacity Totals |
|---------|----------|-----|----------|------|---------|----------------|-----------------|
| | | | | LCPs | LCP:RCP | | |
| GP | 3 | 47 | | 141 | 47.000 | 1,880 | |
| zAAP | | | | | | | |
| zIIP | 1 | 46 | | 46 | 46.000 | 460 | |
| IFL | 2 | 2 | | 4 | 2.000 | 100 | |
| ICF | 2 | 3 | 1 | 2 | 2.000 | 100 | |
| Totals | 8 | 98 | 1 | 193 | | | |

Host Summary | LCP Alternatives | zAAP/zIIP Loading | Calibrate Capacity

For significant configuration changes such as upgrading the processor family, consider capacity comparisons to have a +/-5% margin-of-error

Error: GP shared LCP:RCP ratio exceeds cutoff value of 16.81 for 3 shared RCPs; No results will be generated
Error: zIIP shared LCP:RCP ratio exceeds cutoff value of 30.00 for 1 shared RCPs; No results will be generated
Error: zIIP shared LCP:RCP ratio exceeds cutoff value of %v for %cps shared RCPs; No results will be generated
 Note: 13 defined partitions are excluded from consideration in the results

Figure 12-6 zPCR message window



Channel options

This appendix describes all channel attributes, the required cable types, the maximum unrepeated distance, and the bit rate for IBM z15. The features that are hosted in the PCIe drawer for Cryptography are also listed.

For all optical links, the connector type is LC Duplex (except for the zHyperLink) and the ICA SR connections, which are established with multifiber push-on (MPO) connectors.

The MPO connector of the zHyperLink, and the ICA connection features two rows of 12 fibers and are interchangeable.

The electrical Ethernet cable for the Open Systems Adapter (OSA) connectivity is connected through an RJ45 jack.

The attributes of the channel options that are supported on z15 are listed in Table A-1.

Table A-1 z15 channel feature support

| Channel feature | Feature codes | Bit rate ^a in Gbps (or stated) | Cable type | Maximum unrepeated distance ^b | Ordering information |
|--|---------------|---|---------------|--|----------------------|
| zHyperLink and Fiber Connection (FICON) | | | | | |
| zHyperlink Express1.1 | 0451 | 8 Gbps | OM4, OM3 | See Table A-2 on page 491 | New build |
| zHyperlink Express | 0431 | | | | Carry forward |
| FICON Express16SA LX | 0436 | 8, or 16 | SM 9 μm | 10 km (6.2 miles) | New build |
| FICON Express16SA SX | 0437 | 8, or 16 | OM2, OM3, OM4 | See Table on page 491. | New build |
| FICON Express16S+ LX | 0427 | 4, 8, or 16 | SM 9 μm | 10 km (6.2 miles) | Carry forward |
| FICON Express16S+ SX | 0428 | 4, 8, or 16 | OM2, OM3, OM4 | See Table on page 491. | Carry forward |

| Channel feature | Feature codes | Bit rate ^a in Gbps (or stated) | Cable type | Maximum unrepeated distance ^b | Ordering information |
|---|---------------|---|--|--|-----------------------------|
| FICON Express16S LX | 0418 | 4, 8, or 16 | SM 9 µm | 10 km (6.2 miles) | Carry forward |
| FICON Express16S SX | 0419 | 4, 8, or 16 | OM2, OM3, OM4 | See Table on page 491. | Carry forward |
| FICON Express8S LX | 0409 | 2, 4, or 8 | SM 9 µm | 10 km (6.2 miles) | Carry forward |
| FICON Express8S SX | 0410 | 2, 4, or 8 | OM2, OM3, OM4 | See Table on page 491. | Carry forward |
| Open Systems Adapter (OSA) and Remote Direct Memory over Converged Ethernet (RoCE) | | | | | |
| OSA-Express7S 25GbE SR1.1 | 0449 | 25 | MM 50 µm | 70 m (2000) 100 m (4700) | New build and Carry forward |
| OSA-Express7S 25GbE SR | 0429 | | | | Carry forward |
| OSA-Express7S 10GbE LR | 0444 | 10 | SM 9 µm | 10 km (6.2 miles) | New build |
| OSA-Express6S 10GbE LR | 0424 | | | | Carry forward |
| OSA-Express5S 10GbE LR | 0415 | | | | |
| OSA-Express7S 10GbE SR | 0445 | 10 | MM 62.5 µm | 33 m (200) 82 m (500) | New build |
| OSA-Express6S 10GbE SR | 0425 | | | | Carry forward |
| OSA-Express5S 10GbE SR | 0416 | | MM 50 µm | 300 m (2000) | |
| OSA-Express7S GbE LX | 0442 | 1.25 | SM 9 µm | 5 km (3.1 miles) | New build |
| OSA-Express6S GbE LX | 0422 | | | | Carry forward |
| OSA-Express5S GbE LX | 0413 | | | | |
| OSA-Express7S GbE SX | 0443 | 1.25 | MM 62.5 µm | 275 m (200) | New build |
| OSA-Express6S GbE SX | 0423 | | | | Carry forward |
| OSA-Express5S GbE SX | 0414 | | MM 50 µm | 550 m (500) | |
| OSA-Express7S 1000BASE-T | 0446 | 1000 Mbps | Cat 5, Cat 6 unshielded twisted pair (UTP) | 100 m | New build |
| OSA-Express6S 1000BASE-T | 0426 | 100 or 1000 Mbps | | | Carry forward |
| OSA-Express5S 1000BASE-T | 0417 | | | | |
| 25GbE RoCE Express2.1 | 0450 | 25 | OM4 | 70 m (2000) 100 m (4700) | New build |
| 25GbE RoCE Express2 | 0430 | | | | Carry forward |
| 10GbE RoCE Express2.1 | 0432 | 10 | OM4 | 33 m (200) 82 m (500) | New build |
| 10GbE RoCE Express2 | 0412 | | | | Carry forward |
| 10GbE RoCE Express | 0411 | | OM3 | 300m (2000) ^c | |
| Parallel Sysplex | | | | | |

| Channel feature | Feature codes | Bit rate ^a in Gbps (or stated) | Cable type | Maximum unrepeated distance ^b | Ordering information |
|-----------------|---------------|---|------------|--|---------------------------|
| CE LR | 0433 | 10 Gbps | SM 9 µm | 10 km (6.2 miles) | New build & Carry forward |
| ICA SR1.1 | 0176 | 8 Gbps | OM4 OM3 | 150 m 100 m | New build |
| ICA SR | 0172 | | | | Carry forward |

a. The link data rate does not represent the performance of the link. The performance depends on many factors, including latency through the adapters, cable lengths, and the type of workload.

b. Where applicable, the minimum fiber bandwidth distance in MHz·km for multi-mode fiber optic links is included in parentheses.

c. A 600 meters maximum when sharing the switch across two RoCE Express features.

The unrepeated distances for different multimode fiber optic types for zHyperlink Express are listed in Table A-2.

Table A-2 Unrepeated distances

| Cable type (modal bandwidth) | 8 Gbps |
|-------------------------------|------------|
| OM3 (50 µm at 2000 MHz·km) | 100 meters |
| | 328 feet |
| OM4 (50 µm at 4700 MHz·km) | 150 meters |
| | 492 feet |

The maximum unrepeated distances for FICON SX features are listed in Table A-3.

Table A-3 Maximum unrepeated distance for FICON SX features

| Cable type/bit rate | 1 Gbps | 2 Gbps | 4 Gbps | 8 Gbps | 16 Gbps |
|--|------------|------------|------------|------------|------------|
| OM1 (62.5 µm at 200 MHz·km) | 300 meters | 150 meters | 70 meters | 21 meters | N/A |
| | 984 feet | 492 feet | 230 feet | 69 feet | N/A |
| OM2 (50 µm at 500 MHz·km) | 500 meters | 300 meters | 150 meters | 50 meters | 35 meters |
| | 1640 feet | 984 feet | 492 feet | 164 feet | 115 feet |
| OM3 (50 µm at 2000 MHz·km) | 860 meters | 500 meters | 380 meters | 150 meters | 100 meters |
| | 2822 feet | 1640 feet | 1247 feet | 492 feet | 328 feet |
| OM4 ^a (50 µm at 4700 MHz·km) | N/A | 500 meters | 400 meters | 190 meters | 125 meters |
| | N/A | 1640 feet | 1312 feet | 693 feet | 410 feet |

a. Fibre Channel Standard (not certified for Ethernet)



System Recovery Boost

In this appendix, we introduce System Recovery Boost, which is a new function of the IBM z15. System Recovery Boost delivers substantially faster system shutdown and restart, as well as short duration process recovery boosts for sysplex events.

Note: System Recovery Boost is a new firmware feature and is available on z15 CPC. It is *not* available on older systems.

Naming: The IBM z15 server generation is available as the following machine types and models:

- ▶ Machine Type 8561 (M/T 8561), Model T01, Features Max34, Max71, Max108, Max145, and Max190, which is further identified as *IBM z15 Model T01*, or *z15 T01*, unless otherwise specified.
- ▶ Machine Type 8562 (M/T 8562), Model T02, Features Max4, Max13, Max21, Max32, and Max65, which is further identified as *IBM z15 Model T02*, or *z15 T02*, unless otherwise specified.

In the remainder of this appendix, IBM z15 (z15) refers to both machine types.

This appendix includes the following topics:

- ▶ B.1, “Overview” on page 494
- ▶ B.2, “Functions” on page 494
- ▶ B.3, “Delivering extra capacity” on page 496
- ▶ B.4, “Setting up the System Recovery Boost” on page 503
- ▶ B.5, “Monitoring System Recovery Boost” on page 504
- ▶ B.6, “Automation” on page 505
- ▶ B.7, “Pricing” on page 506
- ▶ B.8, “Software support” on page 506

B.1 Overview

System Recovery Boost is a new feature that is implemented in z15 CPC firmware. It delivers improved overall system and application availability by minimizing the downtime that results from system shutdown and restart operations, and short duration process recovery boost for sysplex environments.

System Recovery Boost realizes the following benefits:

- ▶ During a planned or unplanned system restart,
 - Shuts down the system substantially faster than any prior Z machine.
 - Helps restart and recover the middleware environment and client workloads substantially faster than on any prior Z machine.
 - Delivers higher processor capacity for a limited time following an IPL during a boost period so that the operating system can start faster and client workloads can catch up and work through a backlog after a downtime.
- ▶ During events in a parallel sysplex it delivers additional CP capacity for short duration process recovery boosts.

B.1.1 Use cases

System Recovery Boost provides value for many use cases, including the following examples:

- ▶ Single-system IPL (planned and unplanned):
 - Planned or rolling IPLs (for example, to install software maintenance and disruptive system maintenance)
 - Unplanned IPLs to recover after an operating system failure, crash or “sick but not dead” occurrence that required a system shutdown or restart
- ▶ Multi-system IPL (planned and unplanned)
 - Restart all images on a CPC after planned CPC IML/POR (CPC non-concurrent upgrade)
 - Restart all images on a CPC after unplanned CPC failure following a CPC IML/POR
 - Bring back up sysplex after sysplex-wide (or sysplex multi-system) failure or crash or “sick but not dead” occurrence that required a sysplex shutdown or restart
- ▶ DR/Site Switch:
 - Planned DR test, bringing up test systems at DR site
 - Planned or unplanned site switch, bringing up systems at DR site
- ▶ Process recovery boosts (short duration) for helping with sysplex recovery
 - Providing boosted processor capacity to mitigate the impact on workload processing following short-term recovery events in a sysplex, and restore normal sysplex steady-state sysplex operation as quickly as possible.

B.2 Functions

IBM Z Recovery boost is a new feature that was introduced with IBM z15 firmware (Driver 41), which is designed to provide higher temporary processing capacity to LPARs for speeding up shutdown, IPL, and stand-alone memory dump operations, as well as for short duration

process recovery boosts (following an event in a parallel sysplex), without increasing IBM software costs.

By default, System Recovery Boost capacity is provided in the following ways:

- ▶ By converting subcapacity CPs to full-capacity CPs, also known as *speed boost* for the opted-in images¹ during the boost period.
- ▶ By dispatching general-purpose workloads to zIIPs (for z/OS LPARs enabled for boost and with allocated zIIPs in the LPAR profile) during the boost period (zIIP boost).
- ▶ By way of firmware enhancements that support greater parallelism and performance improvements in the hardware API services. These enhancements are used by IBM GDPS to speed up the orchestration of shutdown and restart activities.

This support is available in GDPS V4.2 for workloads running on z15 CPCs.

- ▶ By providing short duration boosted processor capacity for boosting process recovery in a parallel sysplex to mitigate the impact on workload processing following short-term recovery events in a sysplex, and restore normal sysplex steady-state sysplex operation as quickly as possible.

The boost capacity does not contribute to other IBM software license charges.

Figure B-1 shows a typical System Recovery Boost timeline for z/OS.

Note: Timelines might differ for other operating systems.

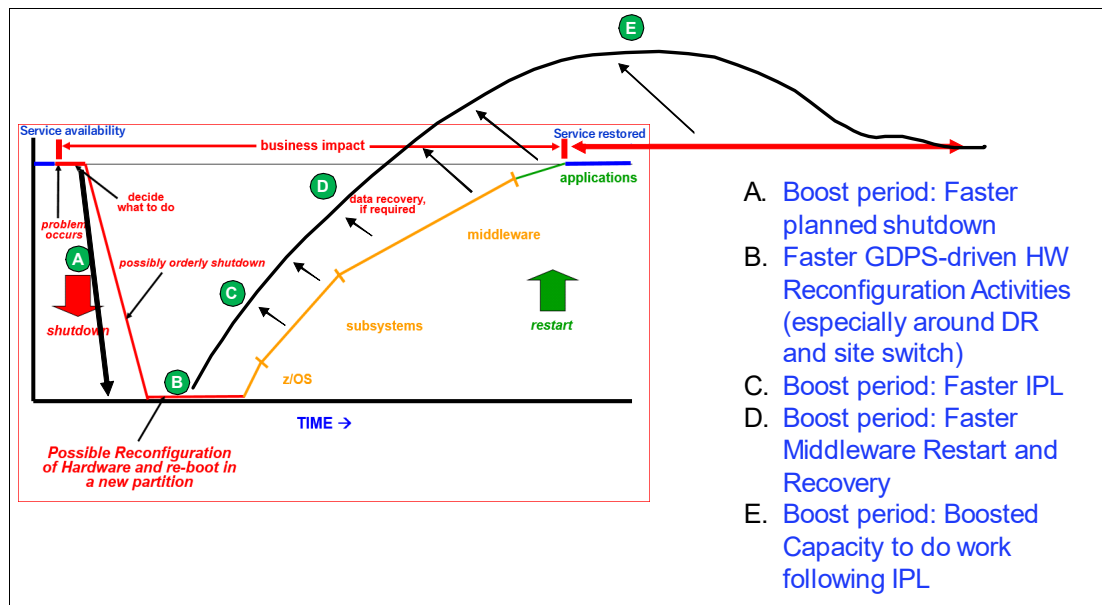


Figure B-1 z/OS typical System Recovery Boost timeline

Boost timeline

For z/OS, current implementation allows 60 minutes boost period for IPL and 30 minutes for shutdown events². The 60 minutes boost period for IPL-ing the z/OS system also allow catching up with the backlog work.

¹ Supported operating images are enabled for boost that is running in an LPAR.

² Boost periods are operating system-specific. Different operating systems can feature different boost periods.

For stand-alone memory dumps, the boost period extends to the duration of the event and it uses subcapacity CP speed boost only (no zIIPs).

For process recovery boosts, boost is provided up to 30 minutes (sum of boost times for multiple processes) period over a period of 24 hours per system image (z/OS LPAR).

B.3 Delivering extra capacity

This section describes the ways in which extra capacity is delivered for System Recovery Boost.

B.3.1 Subcapacity CPs speed boost

When the z15 is configured as a subcapacity model (4xx, 5xx and 6xx), LPARs that are running in a boost period can access the speed boost. This feature requires operating system opt-in and support.

At the time of this writing, IBM z/OS (2.3 and 2.4 with PTFs), IBM z/VM 7.1 with PTFs, z/VSE V6.2, and z/TPF 1.1 with PTFs can use subcapacity boost.

Note: Speed boost applies to general-purpose processors (CPs) only. All other engines run at full capacity (IFLs, zIIPs, and ICFs).

Speed boost example

Figure B-2 shows an example of subcapacity CP speed boost.

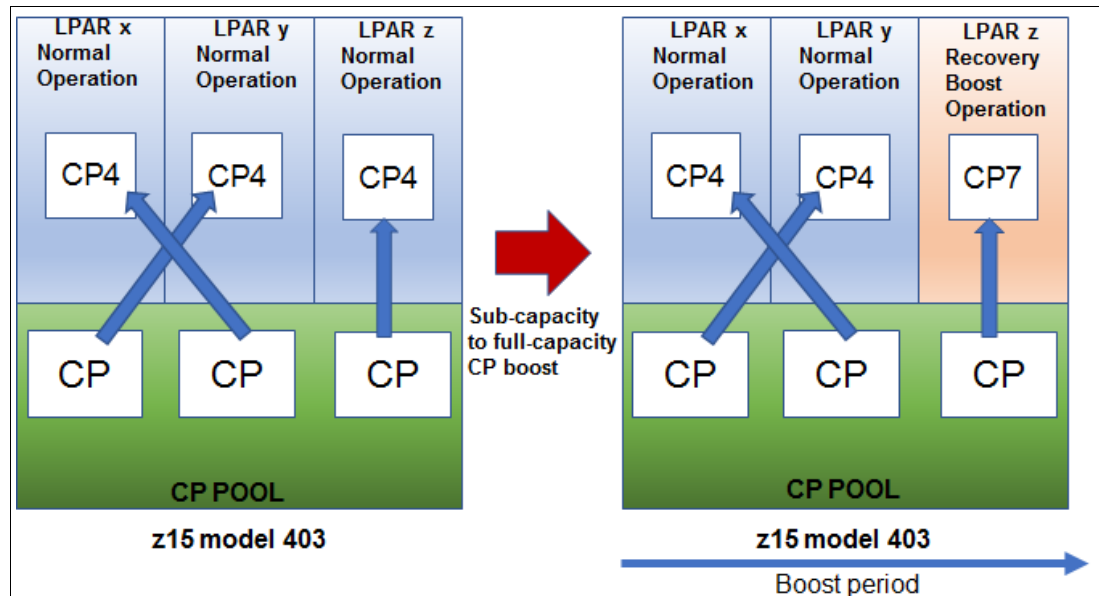


Figure B-2 Subcapacity to full-capacity boost example

In this example, three LPARs are defined in the z15 model 403. In the normal operation, all work is dispatched on subcapacity CPs.

When LPARs enter a boost period, work that is dispatched from LPARz runs at CP7 (full capacity). Other LPARs continue to be dispatched at CP4 (subcapacity). One boost period

started at LPARz shutdown and a new boost period started at IPL (of LPARz). At the end of the IPL boost period, LPARz returns to normal operation at CP4 (subcapacity).

B.3.2 zIIP processor capacity boost (zIIP boost)

Normally, only zIIP eligible work (such as DRDA and IBM Db2 Utilities) is dispatched to zIIPs. During System Recovery Boost period, both zIIP eligible and general CP work is dispatched to available zIIPs for the boost opt-in z/OS images (running in an LPAR).

Notes: Consider the following points:

- ▶ Currently, z/OS uses the zIIP boost feature.
- ▶ At least one zIIP entitlement must be available to use zIIP boost.

In this period, the system can use following processors to run CP workload:

- ▶ Entitled purchased CPs
- ▶ Entitled purchased zIIPs
- ▶ Extra zIIPs, which can be added by using the temporary boost capacity records (FC 9930 and FC 6802 on z15 T01³).

If more logical zIIPs are available and configured in the LPAR profile (whether added by way of the temporary boost capacity record or other temporary capacity activation) while in the boost period, images bring more logical zIIP processors online to use the extra physical zIIP capacity.

After the boost period ends, z/OS dispatching of work on CPs versus zIIPs returns to normal.

Important: These extra logical processors should be taken offline at the end of the boost period (either manually or by using automation), or they are taken offline automatically when the temporary boost record expires.

The start and end of the boost period is signaled by way of a console message, ENF signal (84), and cutting an SMF record. Also, the start and end of boost period starts new SMF interval. A system command or PROC (IEABE) is provided to allow for early opt-out of the boost, if wanted.

³ System Recovery Boost upgrade (FC 9930 and 6802) is not available for z15 T02. Only base System Recovery Boost is available on z15 T02 (built into the system).

System Recovery Boost using zIIP boost example

Figure B-3 shows an example of recovery boost using zIIP capacity boost.

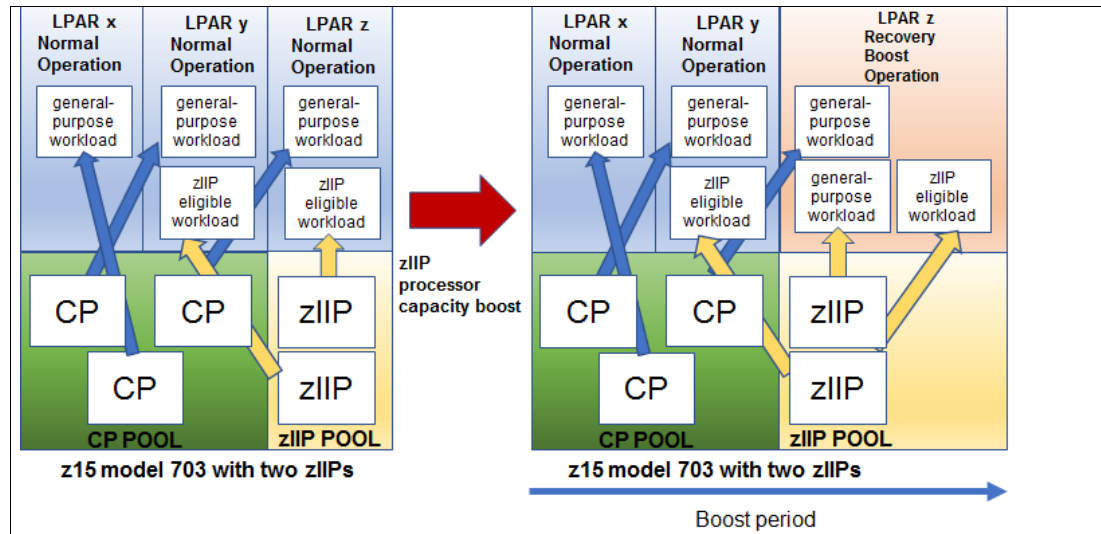


Figure B-3 Example of zIIP recovery boost (z/OS LPAR)

In this example, three LPARs are defined on the z15 model 703 (M/T 8561) with two zIIPs. Two zIIPs are shared between LPARy and LPARz.

During normal operation, only zIIP eligible work is dispatched to the zIIPs. When LPARz enters a boost period, general-purpose work and zIIP eligible work might be dispatched to the zIIPs.

When the boost period ends, only zIIPs-eligible work is dispatched to the zIIPs.

B.3.3 Optional System Recovery Boost Upgrade capacity record (z15 T01)

You can temporarily increase the number of physical zIIPs to use for System Recovery Boost. This feature is the new System Recovery Boost Upgrade record that you can activate from the HMC/SE Perform Model Conversion menu, or by using automation that calls the hardware API services.

After it is activated, zIIPs are added to the zIIP pool and LPAR shares of the extra physical zIIP capacity follows normal LPAR weight rules. You can set up the LPAR image profiles in advance with extra logical zIIP processors to enable the effective use of the extra physical zIIP processor capacity.

Deactivate the temporary boost capacity record at the end of the IPL or shutdown actions or change windows for which they intend to provide a boost (the record auto-deactivates after 6 hours).

For systems that ordered the new System Recovery Boost Upgrade record, the zIIP:CP ratio of 2:1 can be exceeded during the boost periods for the boost opt-in images (LPARs). The boost record activates the zIIPs for up to 6 hours. The boost record has an expiration date of one year.

You must activate the boost record *before* a boost event.

Consider the following points regarding the optional System Recovery Boost Upgrade record:

- ▶ It is a priced feature.
- ▶ The subscription feature is valid for one year (must be renewed after one year).
- ▶ Each activation includes an entitled number of zIIPs, which can be up to 20 and might violate the 2:1 ratio rule between zIIPs and CPs.
- ▶ Activation of this record cannot cause the activation of the first or only zIIP on the machine; therefore, at least one entitled zIIP must be present.

Boost temporary capacity record example

Figure B-4 shows an example of the use the optional boost temporary capacity record.

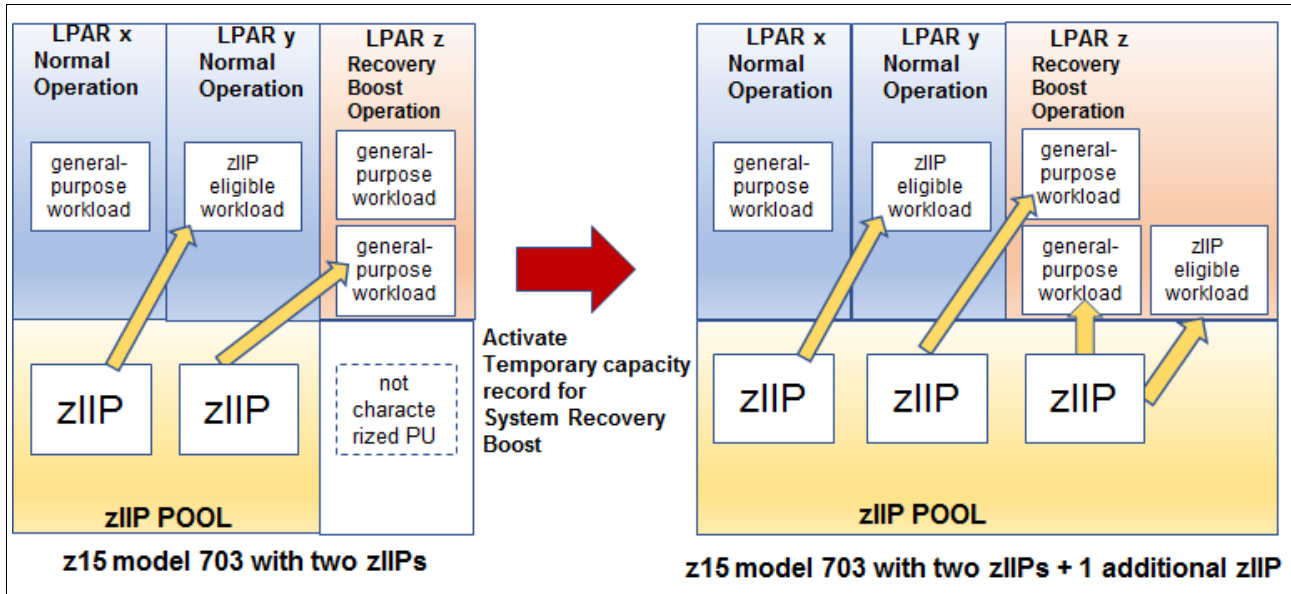


Figure B-4 Example of boost using temporary capacity record

In this example, three LPARs are defined in the z15 model 703 with two zIIPs. Two zIIPs are assigned to both LPARy and LPARz. LPARy is in normal operation.

When LPARz is in a boost period; therefore, both general-purpose work and zIIP eligible work can be dispatched to the zIIPs. LPARz includes a reserved zIIP specified in its image profile.

By activating temporary boost capacity record, one zIIP is added to the zIIP pool and automatically allocated to the LPARz and brought online. In this boost period, general CP work for LPARz is dispatched to zIIPs and CPs.

B.3.4 Planned shutdown boost

A z/OS system can signal that it wants to enter a boost for a planned shutdown by starting the IEASDBS PROC. Consider the following points:

- ▶ In response to starting the PROC, which is driven manually or by way of automation, z/OS opts in to the allowed Boosts permitted by using parmlib.
- ▶ The start and end of the Boost period is signaled by way of a console message, ENF signal (84), and cutting an SMF record. The beginning and end of the boost period triggers new SMF interval.

In a sysplex, WLM sysplex routing starts to route work away from a system after the shutdown PROC is started to further accelerate shutdown.

All z/OS and middleware processing during the shutdown boost period benefits from higher capacity CP processors or extra parallelism that is provided by zIIPs and allows CP work to run on zIIPs.

Shutdown boost example

Figure B-5 shows an example of shutdown boost using subcapacity CP speed boost and zIIP capacity boost.

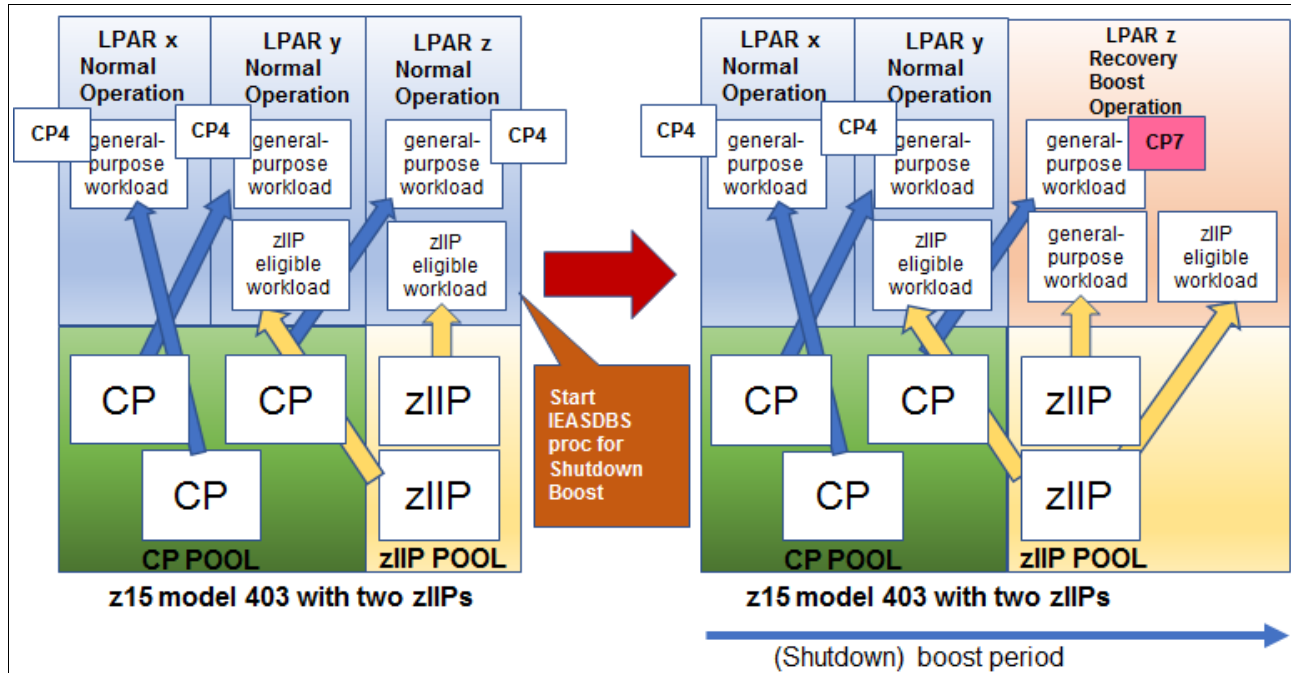


Figure B-5 Example of shutdown boost in a subcapacity model

In this example, three LPARs are defined in the z15 model 403 (M/T 8561) with two zIIPs. Two zIIPs are assigned to LPARy and LPARz. During normal operation, all CP work is dispatched at subcapacity (CP4). Only zIIP eligible work is dispatched to zIIPs.

Note: Stand-alone memory dump does not use zIIP for boost purposes.

Before the planned shutdown of LPARz, the IEASDBS proc is started by an operator or automation. This process starts the shutdown boost. CP work that is dispatched by LPARz is run at full-capacity (CP7) and also general-purpose workload is dispatched to zIIPs. LPARx and LPARy continue in the normal operation at subcapacity (CP4) and only zIIP-eligible workload is dispatched to zIIP.

B.3.5 IBM Geographically Dispersed Parallel Sysplex Actions Performance and Parallelism

IBM Geographically Dispersed Parallel Sysplex (GDPS) drives BCPii HW APIs for orchestrating CBU capacity activations, image activations, resets, and IPLs for multiple images in many planned and unplanned DR site-switch scenarios.⁴

Firmware changes on the z15 HMC and SEs support greater parallelism and performance improvements in the HW API services.

⁴ z/OS use requires z/OS 2.4 with rollback to 2.3 with PTFs (for subcapacity and zIIP boost, on z15 CPC).

GDPS uses these changes to take advantage of available parallelism in the following underlying hardware services:

- ▶ GDPS usage that requires GDPS 4.2
- ▶ Firmware support that is delivered on the z15 machine

B.3.6 Process recovery boosts (short duration)

Important: The capabilities described in this section are available for both IBM z15 T01 and IBM z15 T02 running z/OS images member of a parallel Sysplex. These require:

- ▶ Specific (concurrently installable) LPAR MCL for z15 Driver 41 or higher (check IBM ResourceLink).
- ▶ z/OS 2.4 with PTFs for APAR OA59813
- ▶ z/OS 2.3 with PTFs for APAR OA59813

Note that z/OS APARs will be associated with new FIXCAT for System Recovery Boost (Keyword *SRB/K* and Fixcat Name: *IBM.Function.SystemRecoveryBoost*).

With the enhanced System Recovery Boost support, IBM is extending the boost technologies to provide short-term *recovery process boost* acceleration for other specific sysplex recovery events in z/OS.

Currently, these sysplex recovery events often cause short-duration workload impacts or workload spikes while the sysplex is busy recovering for a recovery event. Recovery affects the normal execution of the client workload in the sysplex, until the recovery processing completes.

The process recovery boost is designed to provide boosted processor capacity to mitigate short-term recovery impacts, and restore normal sysplex steady-state sysplex operation as quickly as possible following specific recovery events, as well as to provide boosted processor capacity for a short period of time following restoration of steady-state operation, to help with workload “catch-up” from the recovery event.

Each z/OS image can receive boosts as follows:

- ▶ One long-duration boost for image startup (60 minutes), of each type (CP speed boost, zIIP boost), and one long-duration boost for image shutdown (30 minutes), of each type.
- ▶ Several short-duration recovery process boosts, of each type, each of less than five (5) minutes duration, with a total usage of no more than 30 minutes of recovery process boost time in any given consecutive 24-hour period.
 - LPAR support times the use of recovery process boosts and limits the total usage per-day for each image

z/OS manages the recovery process boosts internally, with the operating system initiating the boosts as these recovery events take place, and only on the images that are affected by these events. If recovery process boosts happen to “overlap” – a second recovery process boost occurs before a first one has used its entire boost period – then the overlapping boosts are merged and the boost period may be extended to allow the full boost period time for the second recovery process

Process recovery boost candidates

System Recovery Boost process recovery boost provides boosted processor capacity and parallelism to accelerate the following recovery events:

- ▶ **Sysplex Partitioning Recovery:** Boost all surviving systems in the sysplex as they take on the additional workload of sysplex partitioning related recovery, after planned or unplanned removal of a system from the sysplex

When a system in the sysplex is removed, the surviving systems have to do a large amount of recovery processing to clean up after the failed system, free up resources that were held on the failed system, etc.

- ▶ **CF Structure Recovery:** Boost all systems participating in CF structure recovery processing – CF structure rebuild, duplexing failover, re-duplexing.

Recovering failed CF structures and their data can be a laborious process that requires the participation of all systems that were using those CF structures, and can apply to many structures in cases like loss of a CF image

- ▶ **CF Datasharing Member Recovery:** Boost all systems participating in recovery from termination of a CF datasharing member.

When a datasharing member (e.g. a Db2 instance) fails, the other surviving members have to do a lot of recovery/cleanup processing to free up locks and other datasharing resources held by the failed member

- ▶ **Hyperswap Recovery:** Boost all systems participating in a Hyperswap recovery process

Hyperswap processing is a coordinated, sysplex-wide recovery process that restores access to DASD devices following the failure of a storage controller. Its recovery time is sometimes limited by slow processing on one or more participating systems.

Operational considerations

- ▶ During a Recovery Process boost period, WLM neither routes work away from the system (as it does during shutdown boost) nor towards the system (as it does during startup boost). WLM ignores short-duration recovery boosts for workload routing purposes.
- ▶ When bringing reserved logical zIIP processors online and offline at the start and end of a recovery process boost period, z/OS limits the number of “transient” zIIPs brought online/offline automatically to at most two (more transient zIIPs during IPL and shutdown boost periods may be configured).
- ▶ System Recovery Boost Upgrade record activation (used to activate additional temporary zIIP capacity on the z15 T01) is NOT supported for use with recovery process boost periods.
- ▶ z/OS starts and ends a new SMF interval during a recovery process boost period, but when two or more recovery process boosts “overlap” they are merged into a single boost period and a single SMF interval
- ▶ z/OS issues ENF signals and console messages as appropriate when starting, extending, or stopping a recovery process boost
- ▶ z/OS does not permit overlap between the use of recovery process boosts, and the longer image-level startup/shutdown boosts:
 - Recovery process boosts are not initiated while an image-level startup boost is still in progress – the system is already boosted
 - If a recovery process boost is in progress when a system image-level shutdown is initiated, then z/OS “cancels” the in-progress recovery process boost, and initiates the shutdown boost period for system shutdown
 - If additional transient zIIPs were already online during the recovery process boost, z/OS would potentially have to bring additional ones online for the shutdown boost, up to the full quota of reserved logical zIIPs.

B.4 Setting up the System Recovery Boost

System Recovery Boost is a firmware feature of the z15 CPC for operating systems that are running in an LPAR, which requires operating system support.

Important: The base System Recovery Boost capability is built into z15 firmware and does not require ordering extra features. System Recovery Boost Upgrade (consisting of FC 9930 and FC 6802) is an optional, orderable feature that provides more temporary zIIP capacity for use during boost periods, and is available on z15 T01 (M/T 8561) only. Consider the following points:

- ▶ FC 9930 is *not* required to use the base System Recovery Boost capability.
- ▶ FC 9930 is *only* needed if more zIIP temporary capacity is required.
- ▶ By default, System Recovery Boost is enabled for z/OS (opt-in). z/OS must run on z15 CPC.
- ▶ For extra zIIP temporary boost capacity, FC 9930 and FC 6802 (System Recovery Boost Upgrade) must be ordered with the z15 T01 CPC.

You can configure a z/OS system-level parameter (IEASYSxx) to control whether a specific z/OS image opts in for the zIIP processor Boost, as shown in the following example:

```
BOOST=SYSTEM | ZIIP | SPEED | NONE
```

No hardware configuration setup is required.

If you want to use offline zIIPs or extra zIIPs that are provided by the System Recovery Boost record, you must define reserved zIIPs in the image profile, as shown in Figure B-6 on page 503.

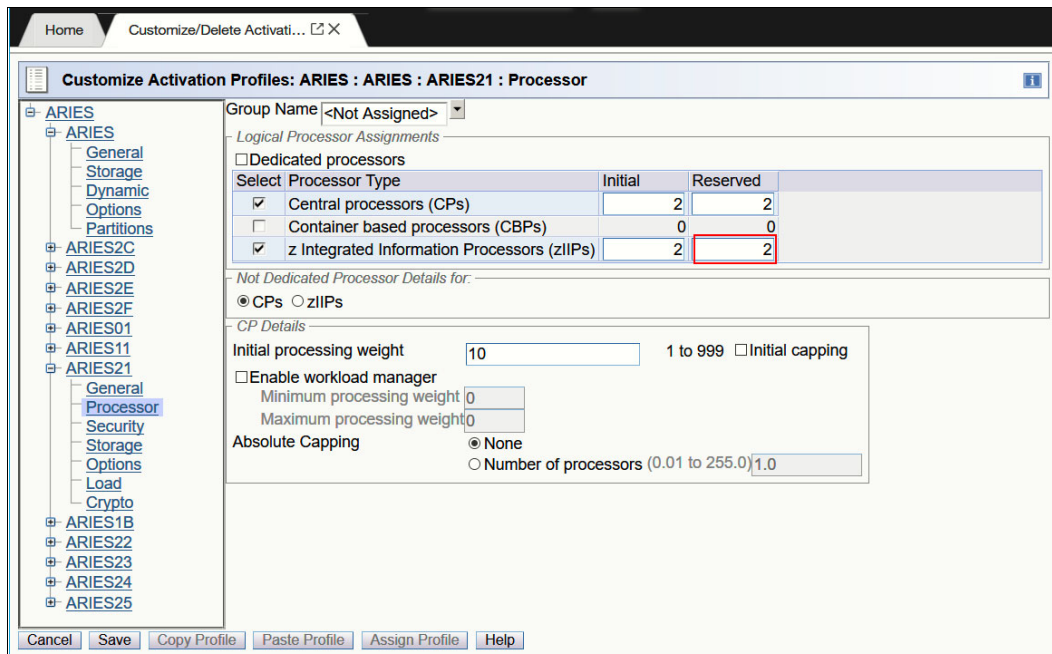


Figure B-6 Reserved zIIPs definition window in the image profile

You also should review LPAR weights and storage allocation to ensure that they meet your system requirements.

Capping, accounting and measurement, capacity reporting, workload management and workload routing, capacity management, operating system messages, ENF signals, control block APIs, SMF data during the boost period, and so on are identical for recovery process boosts as they are for image startup and shutdown boosts.

A new “boost class” is available for Recovery Process boosts, which appears in various system messages, ENF signals, SMF fields, and other z/OS APIs.

B.5 Monitoring System Recovery Boost

The **D IPLINFO, BOOST, STATE** command has been enhanced to show an image’s current boost state, both for startup/shutdown boosts, and the new recovery process boosts, as shown in Example B-1.

Example B-1 z/OS display for boost status

Example 1:

```
IEE257I Boost State
Boost class: IPL
zIIP boost: active with 5 transient zIIP cores
Speed boost: active
```

Example 2:

```
IEE257I Boost State
Boost class: Recovery Process
Requestor: Hyperswap
zIIP boost: active with 2 transient zIIP cores
Speed boost: active
```

In addition, the **DISPLAY M=CPU** has also been enhanced (see Example B-2):

- “I” indicates zIIPs
- “B” indicates (transient) boost zIIPs. This CPU was configured online at the start of the boost period, and will be configured offline when the boost ends

Example B-2 CPU information for transient zIIPs

```
SY1 IEE174I 09.58.10 DISPLAY M 328
PROCESSOR STATUS
ID CPU SERIAL
00 + 0449D74381
01 + 1449D74381
02 +I 2449D74381
03 +B 3449D74381
04 +I 4449D74381
```

When an LPAR is in the Boost period, you can confirm the status of System Recovery Boost in the HMC/SE Partition Image Details window, as shown in Figure B-7. During the boost period, Processor Boost is shown as ON.

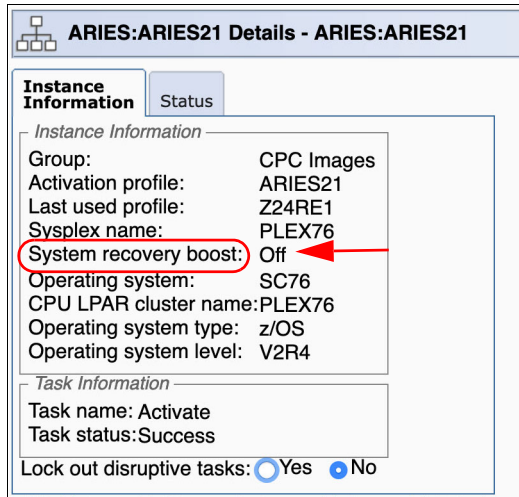


Figure B-7 HMC Partition Image Details window

Also, the processor boost status is shown in HMC Monitors Dashboard.

B.6 Automation

Your automation product can be used in the following ways to automate and control System Recovery Boost activities:

- ▶ To activate and deactivate the eBod temporary capacity record to provide extra physical zIIPs for an IPL or Shutdown Boost.
- ▶ To dynamically modify LPAR weights, as might be needed to modify or “skew” the sharing of physical zIIP capacity during a Boost period.
- ▶ To drive the invocation of the PROC that indicates the beginning of a shutdown process (and the start of the shutdown Boost).
- ▶ To take advantage of new composite hardware API reconfiguration actions.
- ▶ To control the level of parallelism present in the workload at start (for example, starting middleware regions) and shutdown (for example, perform an orderly shutdown of middleware).

Automation can pace or throttle these activities to varying degrees; with Boost, less pacing or more parallelism might be wanted.

- ▶ To automate on the new z/OS messages that are issued at start or end of boost periods to take whatever actions are needed.

B.7 Pricing

In this section, the available pricing options are described.

B.7.1 Base System Recovery Boost feature: No extra charge functions

The following standard, *no-charge* z15 hardware facilities are available:

- ▶ Subcapacity to full-capacity boost for CPs
- ▶ zIIP boost that uses a client's entitled zIIPs
- ▶ GDPS scripting and firmware enhancements
- ▶ Process recovery boosts

B.7.2 Priced feature: System Recovery Boost Upgrade record

A charge function is a *priced activation* of extra zIIP capacity for boost usage. An annual subscription model for entitlement to renewals for zIIP Boost with activation of unassigned processor units (PUs) by way of temporary record also is available.

System Recovery Boost records on z15 T01 require Boost Enablement contract (FC 9930) and temporary pre-paid zIIP boost records (FC 6802).

B.7.3 Software pricing

Boost should *not* increase a customers' IBM software costs, regardless of whether the client is using 4HRA Pricing, Solution Pricing, or Consumption-based Pricing.

B.8 Software support

The following software is supported to use System Recovery Boost:

z/OS z/OS V2R4 with rollback to V2R3 with PTFs.

GDPS V4R2.

Stand-alone Dump Stand-alone memory dump (SADMP) uses subcapacity to full-capacity boost for CPs during memory dump processing (zIIP capacity boost is *not* used for SADMP).

z/VM When running on CP processors, z/VM LPAR uses subcapacity to full-capacity boost for CPs for startup and shutdown (IFLs always run at full capacity, no boost is available for IFLs).

Second-level guests of z/VM I “inherit” the subcapacity boost from VM during these periods, which accelerates the start and shutdown of hosted second-level guests (except for z/OS as a second-level guest). Available in z/VM 7.1 (with fixes) and 7.2 (enabled by default).

z/TPF z/TPF uses Speed Boost for CPs for start and shutdown boost. Existing support for a function (called TPF Dynamic CPU) can be used to provide more CP capacity when needed.

IBM z/VSE

General purpose processors (CPs) running at subcapacity can be boosted to full capacity for a limited time. z/VSE support System Recovery Boost support is available as of September 2020 for both z15 T01 and T02. z/VSE requires PTF for APAR DY47832. During IPL and standalone dump processing, these processors will be boosted for up to 60 minutes, while during system shutdown, these processors will be boosted for up to 30 minutes. The z/VSE system automatically enables the boost during IPL and during standalone dump processing. To enable the boost for system shutdown, use the `SYSDEF SYSTEM` command.



IBM Integrated Accelerator for zEnterprise Data Compression

This appendix describes the new IBM Integrated Accelerator for z Enterprise Data Compression (zEDC) that is implemented in IBM z15 hardware.

The appendix includes the following topics:

- ▶ “Client value of Z compression” on page 510
- ▶ “z15 IBM Integrated Accelerator for zEDC” on page 510
- ▶ “z15 migration considerations” on page 511
- ▶ “Software support” on page 512
- ▶ “Compression acceleration and Linux on Z” on page 513

Client value of Z compression

The amount of data that is captured, transferred, and stored continues to grow. Software-based compression algorithms can be costly in terms of processor resources, storage costs, and network bandwidth.

An optional feature that is available for z14, z13, and z13s servers, zEDC Express addressed customer requirements by providing hardware-based acceleration for data compression and decompression. zEDC provided data compression with lower CPU consumption than compression technology that was available on the IBM Z server.

Existing clients deployed zEDC compression to deliver the following types of compression:

- ▶ Storage
- ▶ Data transfer
- ▶ Database
- ▶ In-application

Data compression delivers the following benefits:

- ▶ Disk space savings
- ▶ Improved elapse times
- ▶ Reduced CPU consumption
- ▶ Reduced network bandwidth requirements and transfer times

Many clients are increasing their zEDC footprint to 8GBps with up to 16 features per z14 system at 1 GBps throughput per feature (redundancy reduces total throughput to 8 GBps).

The z15 further addresses the growth of data compression requirements with the integrated on-chip compression unit (implemented in processor Nest, one per PU chip) that significantly increases compression throughput and speed compared to zEDC deployments.

z15 IBM Integrated Accelerator for zEDC

z15 on-chip compression provides value for existing and new compression users by bringing the compression facility into the PU chip, which is tied in L3 cache.

The z15 Integrated Accelerator for zEDC delivers industry-leading throughput and replaces the zEDC Express PCIe adapter that is available on the IBM z14 and earlier servers.

z15 compression/decompression is implemented in the Nest Accelerator Unit (NXU, see Figure 3-10 on page 106) on each processor chip and replaces the existing zEDC Express adapter in the PCIe+ I/O drawer.

One Nest Accelerator Unit is available per processor chip, which is shared by all cores on the chip and features the following benefits:

- ▶ New concept of sharing and operating an accelerator function in the nest
- ▶ Supports DEFLATE compliant compression/decompression and GZIP CRC/ZLIB Adler
- ▶ Low latency
- ▶ High bandwidth
- ▶ Problem state execution
- ▶ Hardware and firmware interlocks to ensure system responsiveness
- ▶ Designed instruction
- ▶ Run in millicode

Based on IBM benchmarks, the throughput for each On-Chip Compression unit is 12 GBps, which equates to 48 GBps per drawer or 240 GBps for a fully populated five-drawer z15.

On-Chip Compression provides an up to 5% improvement in compression ratios for BSAM/VSAM datasets over zEDC while maintaining full compatibility.

Eliminating adapter sharing by using Nest Compression Accelerator

Sharing of zEDC cards is limited to 15 LPAR guests per adaptor. The Nest Compression Accelerator removes this virtualization constraint because it is shared by all PUs on the processor chip and therefore is available to all LPARs and guests.

Moving the compression function from the (PCIe) I/O drawer to the processor chip means that compression can operate directly in L3 cache and data does not need to be passed by using I/O operations.

Compression modes

Compression is run in one of the following modes:

- ▶ Synchronous

Execution occurs in problem state where the user application starts the instruction in its virtual address space.

- ▶ Asynchronous

Execution is optimized for Large Operations under z/OS for authorized applications (for example, BSAM) and issues I/O by using EADMF for asynchronous execution.

This type of execution maintains the current user experience and provides a transparent implementation for authorized users of zEDC.

Note: The zEDC Express feature does *not* carry forward to z15.

z15 migration considerations

The IBM Integrated Accelerator for zEDC is fully compatible with zEDC. Data compressed by zEDC can be read by z15 (the on-chip) nest accelerator unit and vice versa.

All z/OS configuration stay the same

No changes are required when moving from earlier systems using zEDC to z15.

The IFAPRDxx feature is still required for authorized services. For problem state services, such as zlib usage of Java, it is not required.

Consider fail-over and DR sizing

The order of magnitude throughput increase on z15 means that the throughput requirements need to be considered whether failing over to earlier systems with zEDC.

Performance metrics

On-chip compression introduces the following system reporting changes:

- ▶ No RMF PCIE reporting for zEDC
- ▶ Synchronous executions are not recorded (just an instruction invocation)
- ▶ Asynchronous executions are recorded:
 - SMF30 information is captured for asynchronous usage
 - RMF EADM reporting is enhanced (RMF 74.10)
 - SAP utilization is updated to include time spent compressing and decompressing

zEDC to z15 zlib Program Flow for z/OS

The z/OS provided zlib library is statically linked into many IBM and ISV products and remains functional. However, to get the best optimization for z15, some minor changes are made to zlib.

The current zlib and the new zlib function on z14 and earlier servers and z15 hardware. It functions with or without the z15 z/OS PTFs on z14 and earlier servers.

Software support

Support of the On-Chip Compression function is compatible with zEDC support and is available in z/OS V2R1 or later for data compression and decompression. Support for data recovery (decompression) in the case that zEDC or On-Chip Compression not available; however, it is provided through software in z/OS V2R2, and V2R1 with the appropriate program temporary fixes (PTFs).

Software decompression is slow and can involve considerable processor resources. Therefore, it is not recommended for production environments.

A specific fix category that is named `IBM.Function.zEDC` identifies the fixes that enable or use the zEDC and On-Chip Compression function.

z/OS guests that run under z/VM V6.4 with PTFs and later can use the zEDC Express feature and z15 On-Chip Compression.

For more information, see the [Enhancements to z/VM 6.4 page](#) of the IBM Systems website.

IBM 31-bit and 64-bit SDK for z/OS Java Technology Edition, Version 7 Release 1 (5655-W43 and 5655-W44) (IBM SDK 7 for z/OS Java) now provides use of the zEDC Express feature and Shared Memory Communications-Remote Direct Memory Access (SMC-R), which is used by the 10GbE RoCE Express feature.

For more information about how to implement and use the IBM Z compression features, see *Reduce Storage Occupancy and Increase Operations Efficiency with IBM zEnterprise Data Compression*, SG24-8259.

C.0.1 IBM Z Batch Network Analyzer

IBM Z Batch Network Analyzer (zBNA) is a no-charge, “as is” tool. It is available to clients, IBM Business Partners, and IBM employees.

zBNA is based on Microsoft Windows, and provides graphical and text reports, including Gantt charts, and support for alternative processors.

zBNA can be used to analyze client-provided System Management Facilities (SMF) records to identify jobs and data sets that are candidates for zEDC and z15 On-Chip Compression across a specified time window (often a batch window).

zBNA can generate lists of data sets by the following jobs:

- ▶ Jobs that perform hardware compression and might be candidates for On-Chip Compression.
- ▶ Jobs that might be On-Chip Compression candidates, but are not in extended format.

Therefore, zBNA can help you estimate the use of On-Chip Compression features and help identify savings. The following resources are available:

- ▶ IBM Employees can obtain zBNA and other CPS tools at the [IBM Z Batch Network Analyzer \(zBNA\) Tool page](#) of the IBM Techdoc website.
- ▶ IBM Business Partners can obtain zBNA and other CPS tools at the [IBM PartnerWorld website](#) (log in required).
- ▶ IBM clients can obtain zBNA and other CPS tools at the [IBM Z Batch Network Analyzer \(zBNA\) Tool page](#) of the IBM Techdoc Library website.

Compression acceleration and Linux on Z

The zEDC I/O adapter use is limited in many Linux on Z environments because SR-IOV does not provide a high degree of virtualization; therefore, the user must pick and choose which guests are granted access to the accelerator.

The z15 On-Chip Compression accelerator solves these virtualization limitations because the function is no longer an I/O device and is available as a problem state instruction to all Linux on Z guests without constraints.

This feature enables pervasive usage in highly virtualized environments.

z15 On-Chip Compression is available to open source applications by way of zlib.



D

Frame configurations

This appendix describes the various frame configurations for Power Distribution Units (PDUs) and Bulk Power Assembly (BPA)-based systems. All the diagrams are views from the rear of the system.

The common building blocks are displayed and range from 1 - 4 frames, with various numbers of CPC drawers, PCIe+ I/O drawers.

This chapter includes the following topics:

- ▶ “Power Distribution Unit configurations” on page 516
- ▶ “Bulk Power Assembly configurations” on page 522

Power Distribution Unit configurations

The various PDU-based system configurations are shown in Figure D-1 - Figure D-12 on page 521.

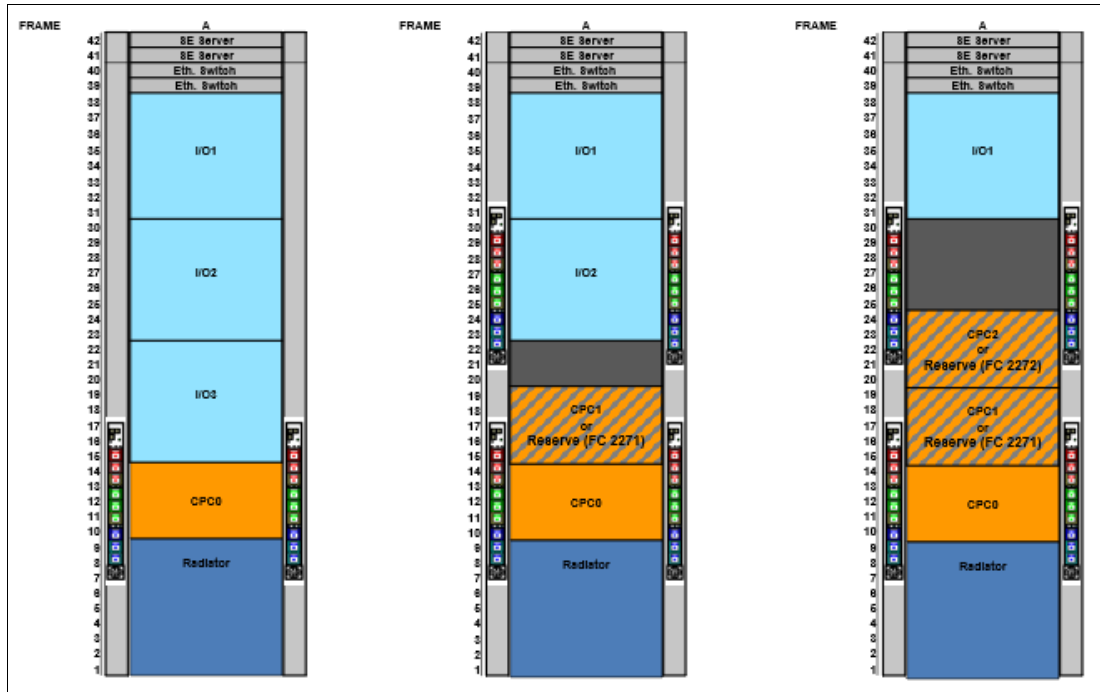


Figure D-1 Single frame

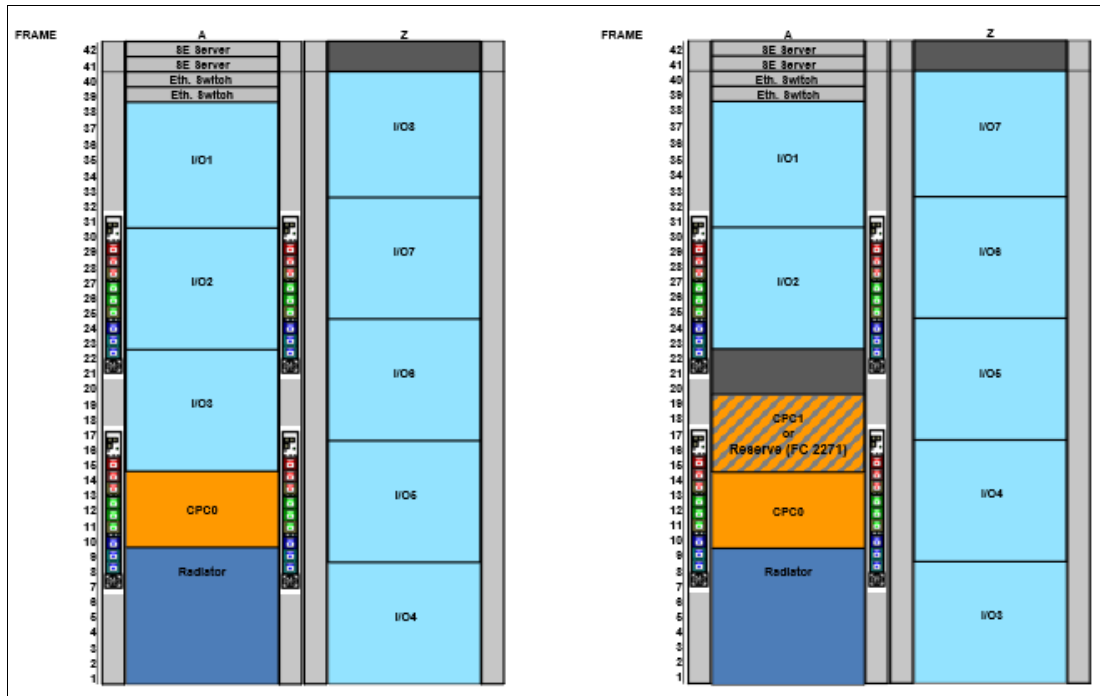


Figure D-2 Two frames AZ

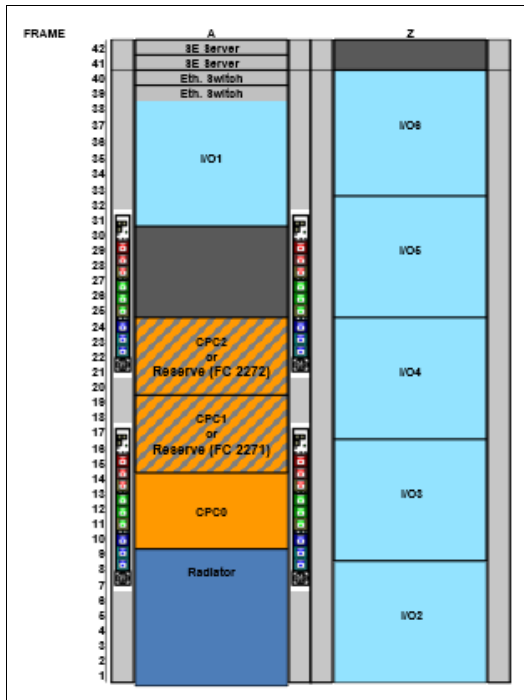


Figure D-3 Two frames AZ three CPC

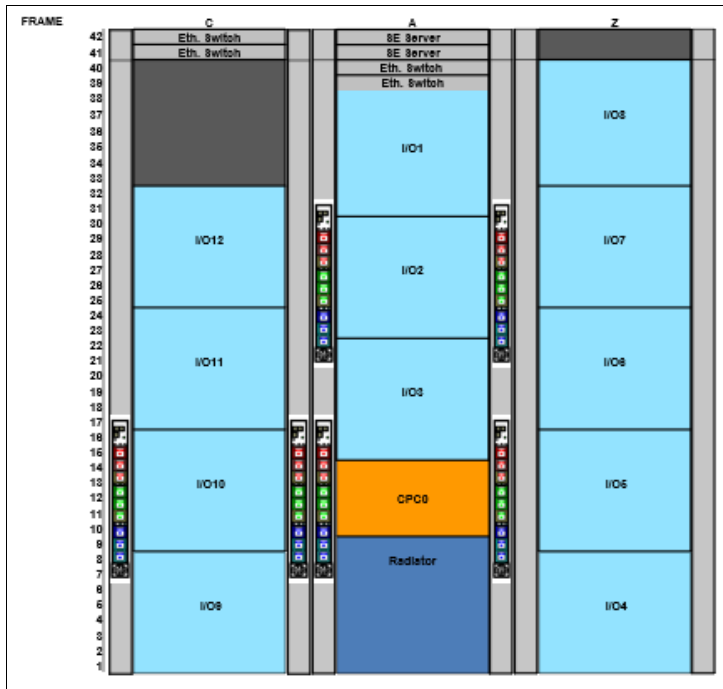


Figure D-4 Three frames CAZ single CPC



Figure D-5 Three frames CAZ two CPC

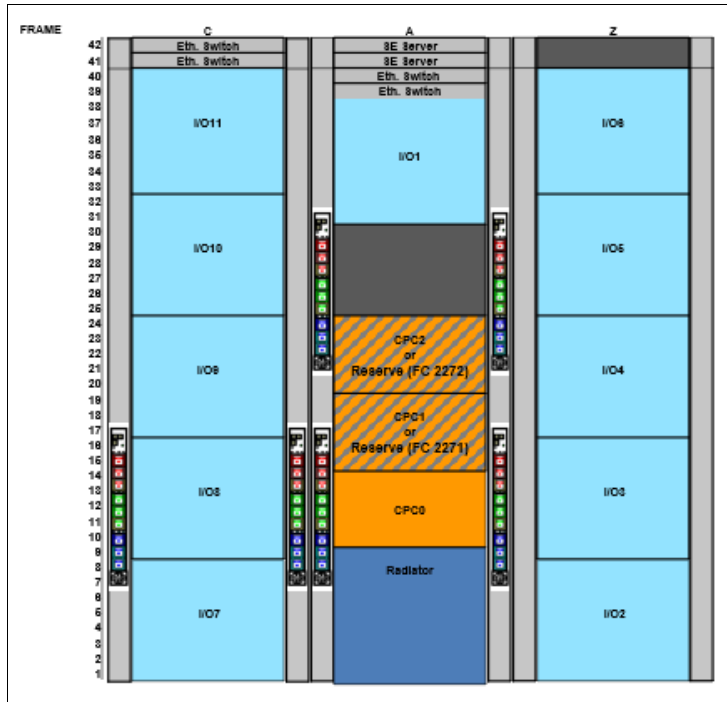


Figure D-6 Three frames CAZ three CPC

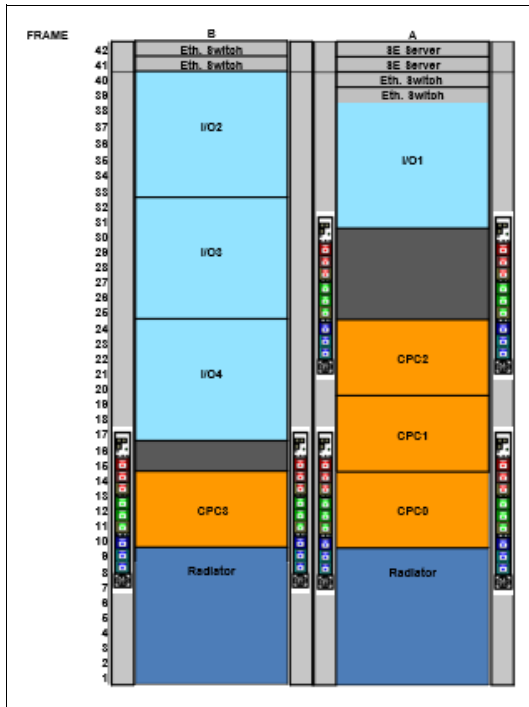


Figure D-7 Two frames BA four CPC

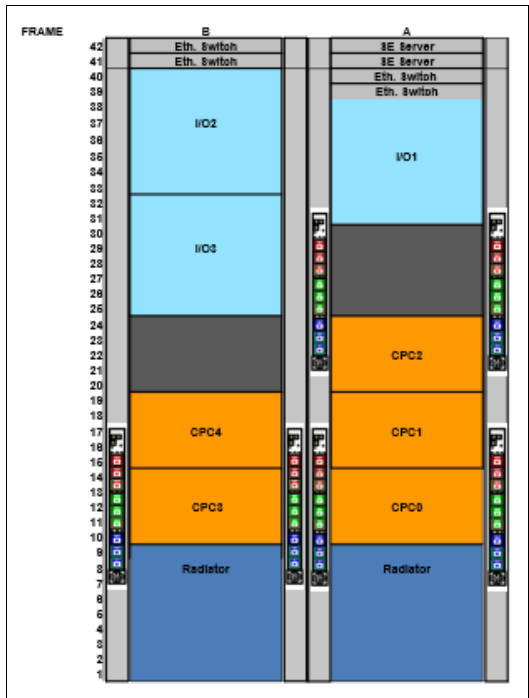


Figure D-8 Two frames BA five CPC

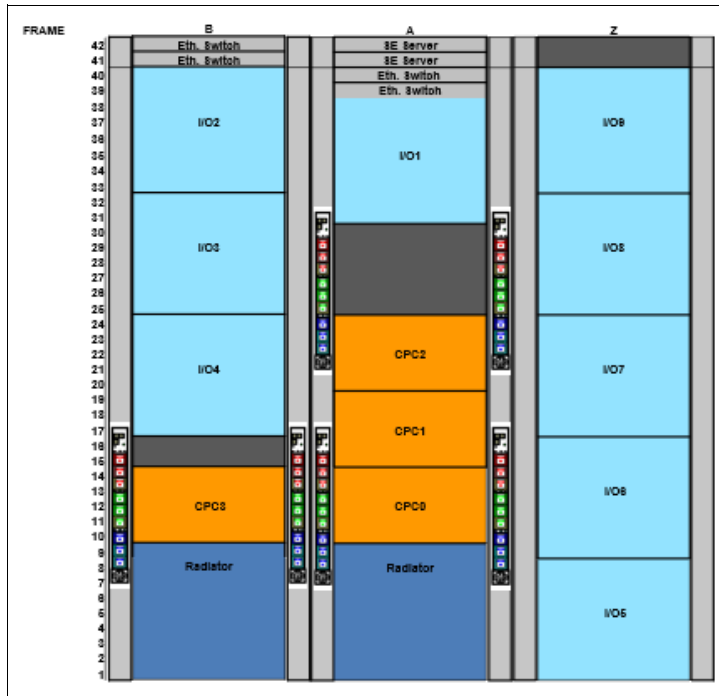


Figure D-9 Three frames BAZ four CPC

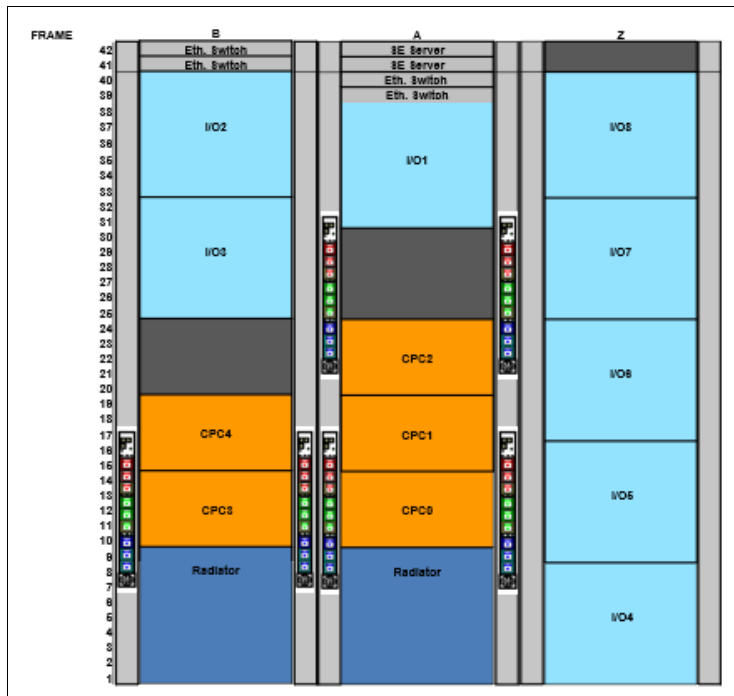


Figure D-10 Three frames BAZ five CPC

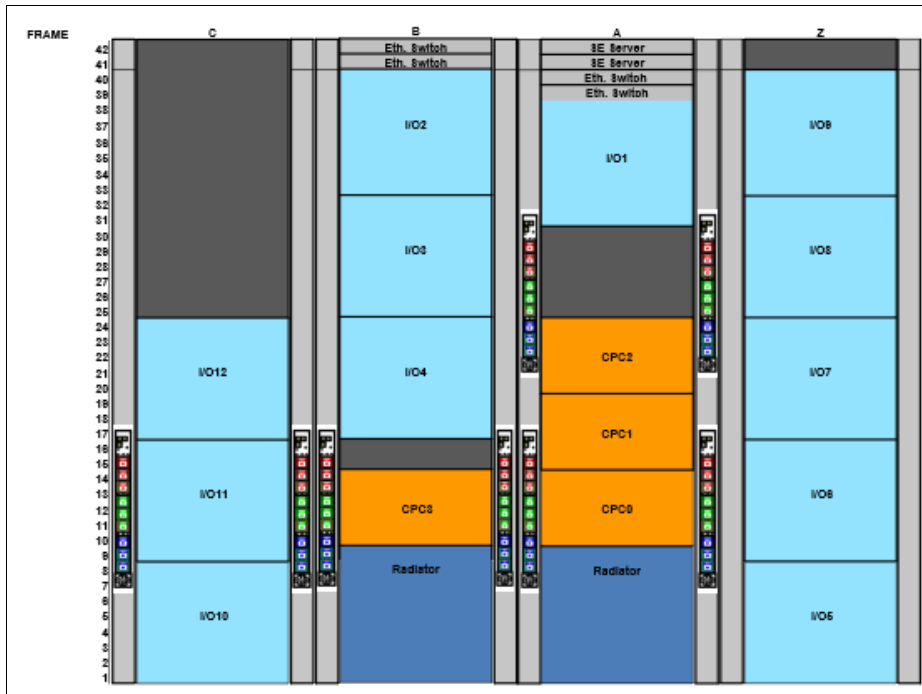


Figure D-11 Four frames CBAZ four CPC

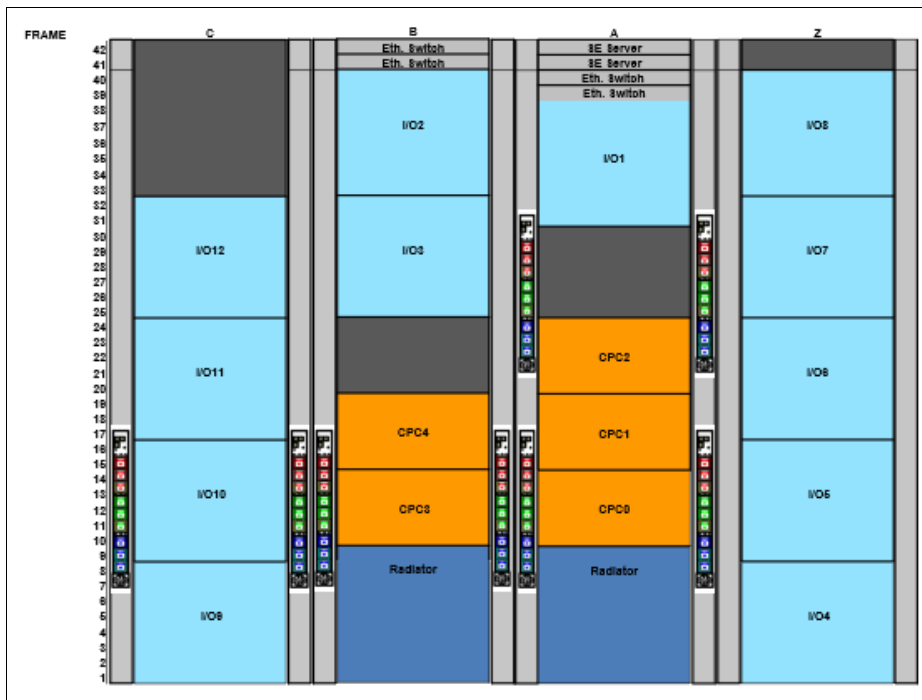


Figure D-12 Four frames CBAZ five CPC

Bulk Power Assembly configurations

The BPA-based system configurations are shown in Figure D-13 - Figure D-26 on page 528.

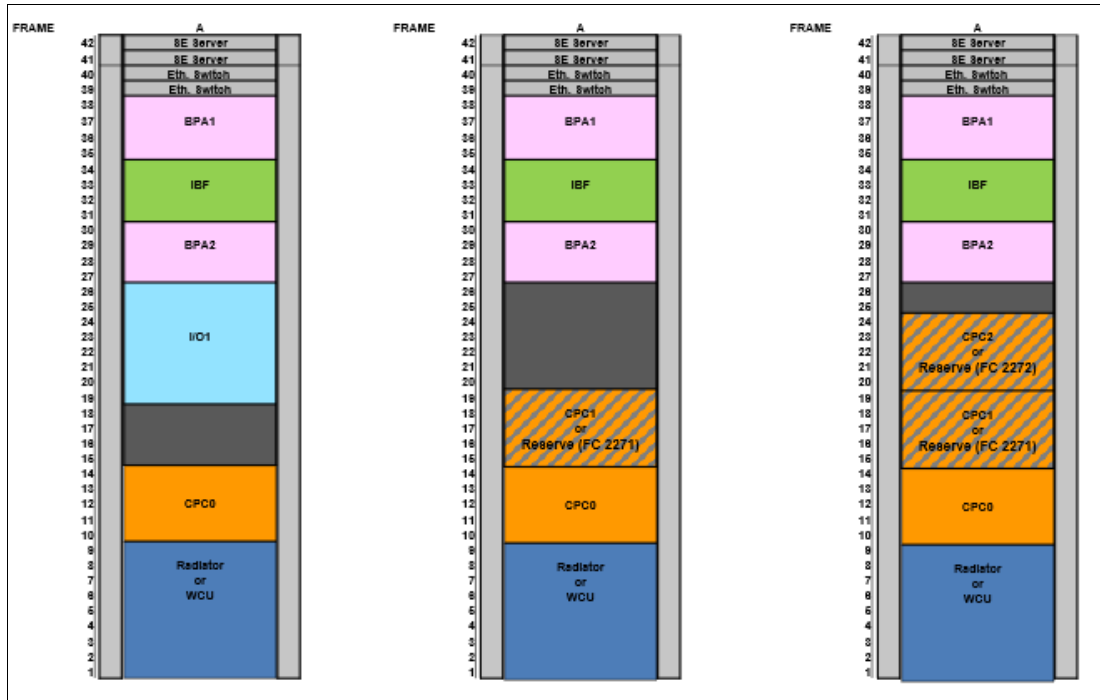


Figure D-13 Single frame

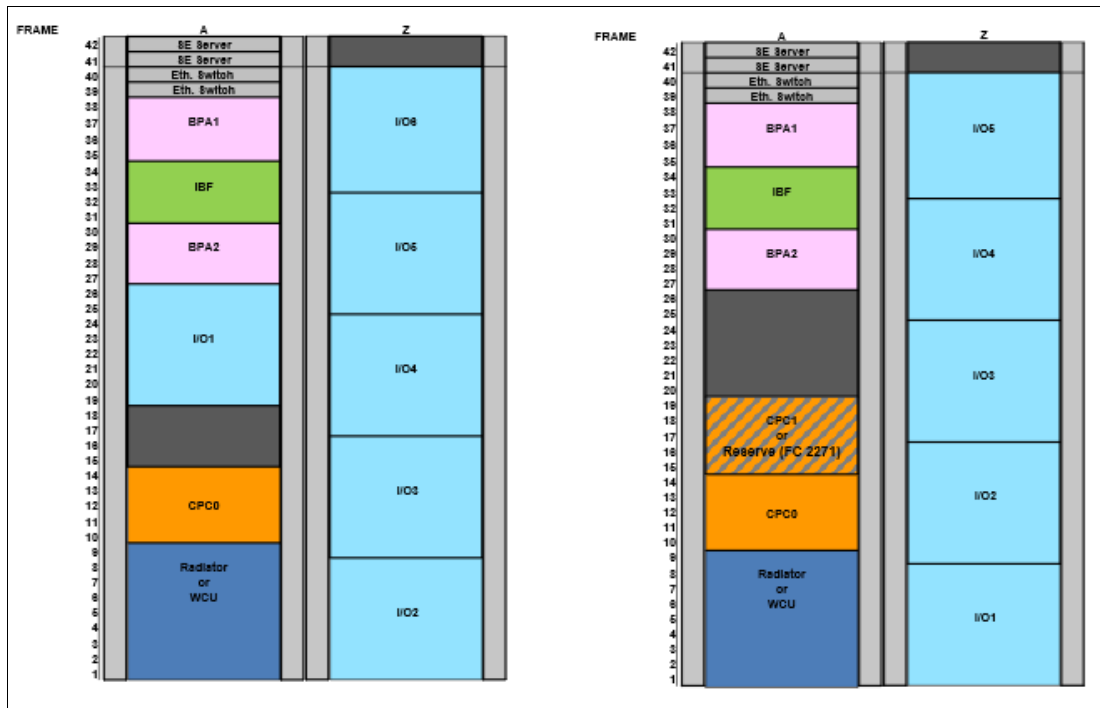


Figure D-14 Two frames AZ

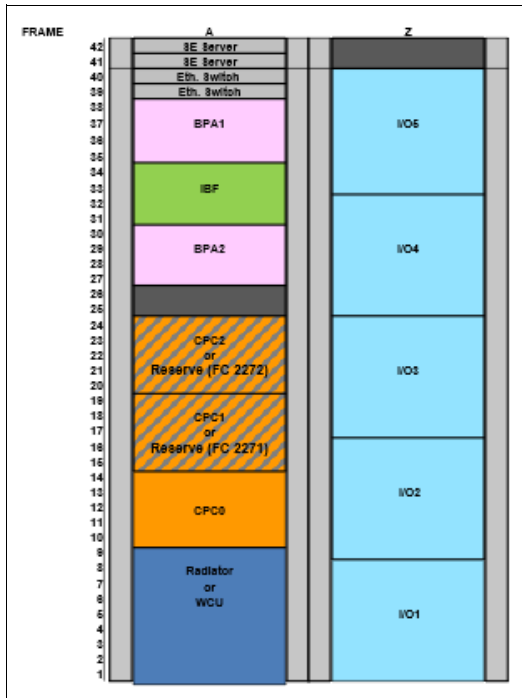


Figure D-15 Two frames AZ three CPC

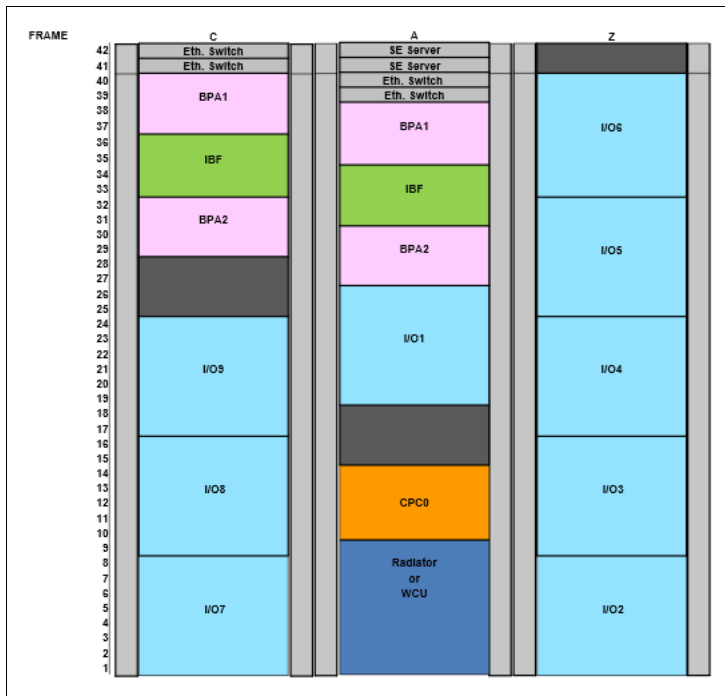


Figure D-16 Three frames CAZ single CPC

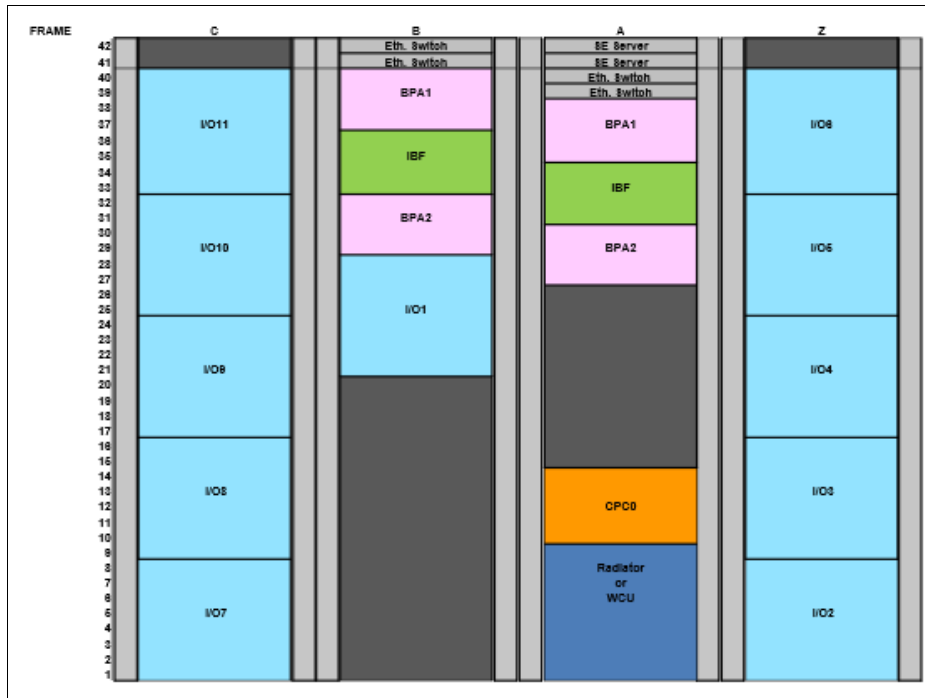


Figure D-17 Four frames CBAZ single CPC

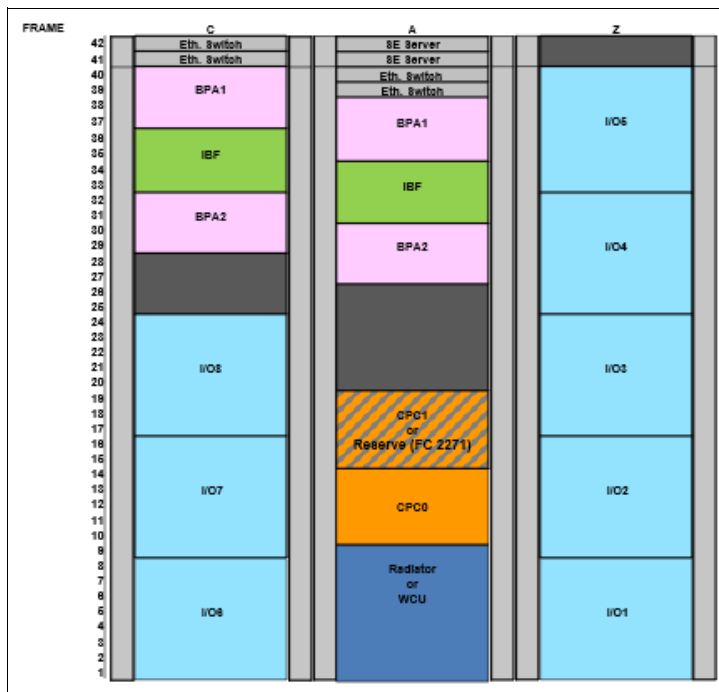


Figure D-18 Three frames CAZ two CPC

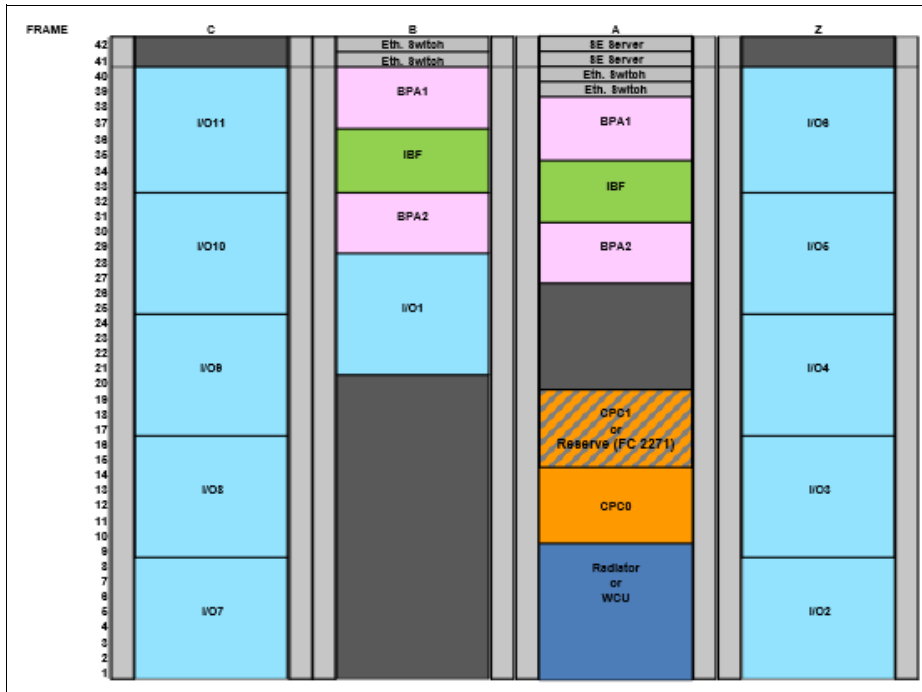


Figure D-19 Four frames CBAZ two CPC

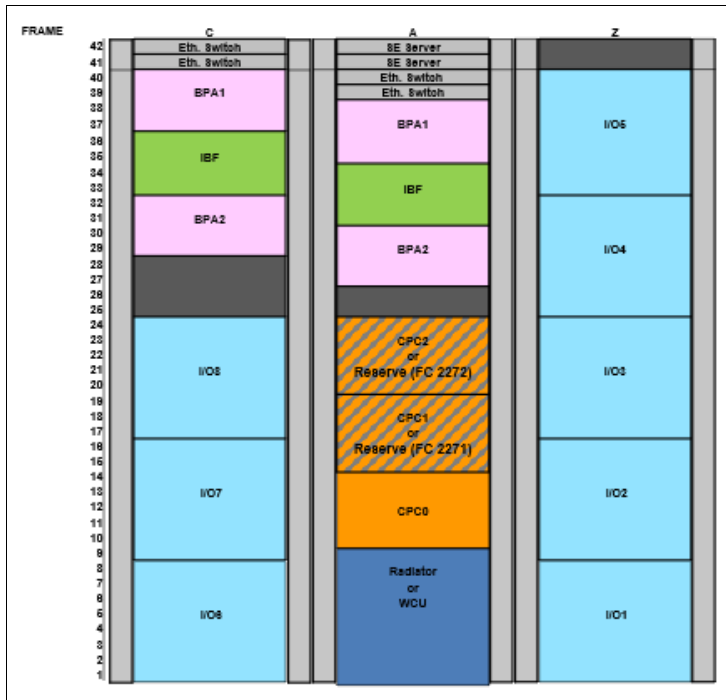


Figure D-20 Three frames CAZ three CPC

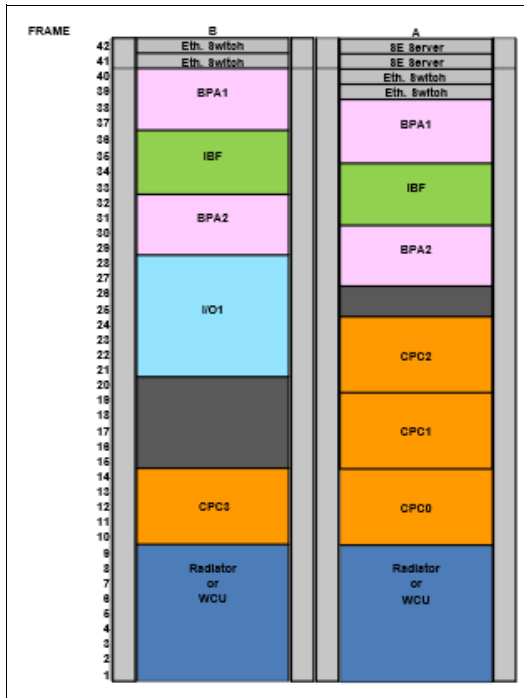


Figure D-21 Two frames BA four CPC

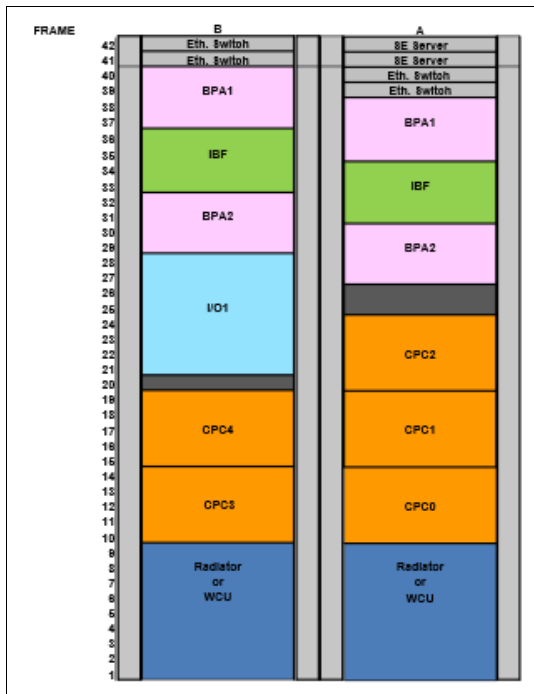


Figure D-22 Two frames BA five CPC

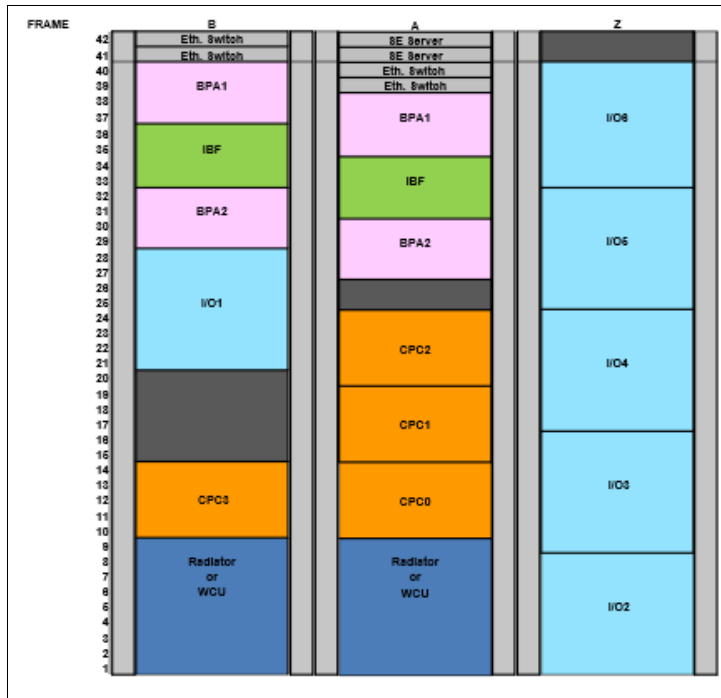


Figure D-23 Three frames BAZ four CPC

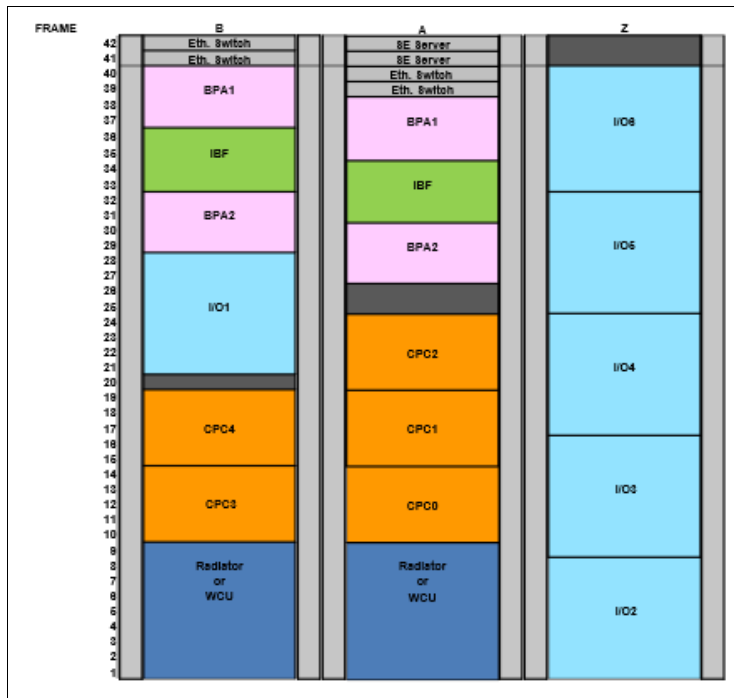


Figure D-24 Three frames BAZ five CPC

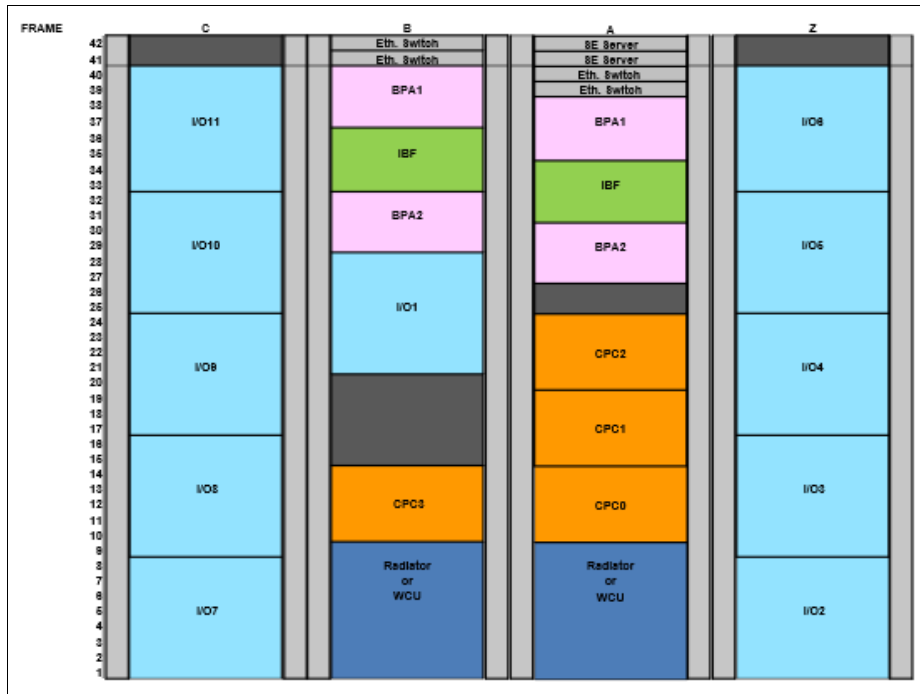


Figure D-25 Four frames CBAZ four CPC

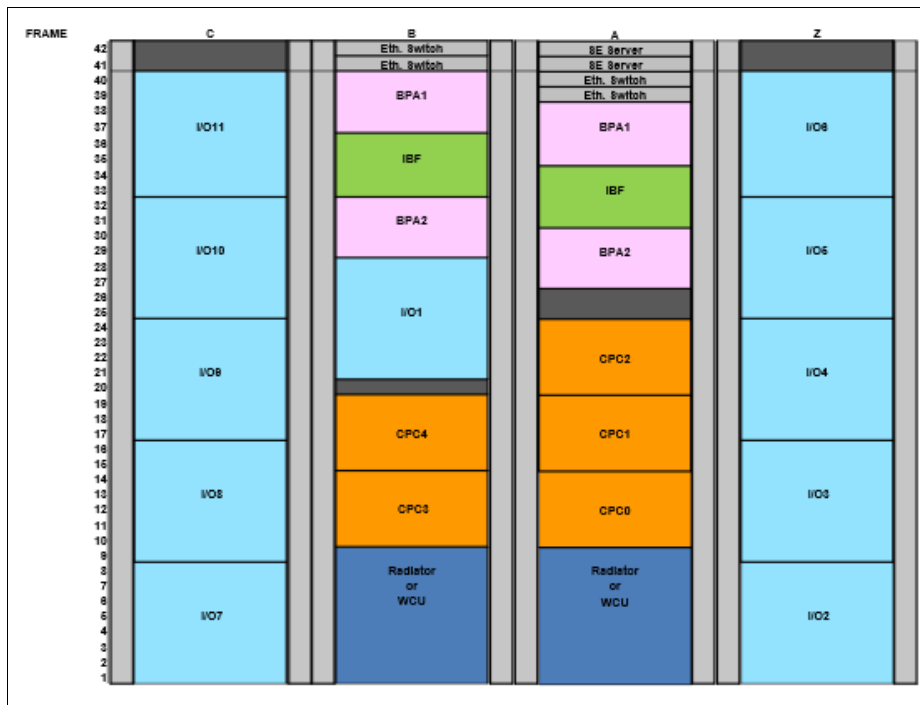


Figure D-26 Four frames CBAZ five CPC

Related publications

The publications that are listed in this section are considered particularly suitable for a more detailed discussion of the topics that are covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide more information about the topic in this document. Note that some publications that are referenced in this list might be available in softcopy only:

- ▶ *IBM z15 Technical Introduction*, SG24-8850
- ▶ *IBM Z Connectivity Handbook*, SG24-5444
- ▶ *IBM z14 (M/T 3907) Technical Guide*, SG24-8451
- ▶ *IBM z14 ZR1 Technical Guide*, SG24-8651
- ▶ *IBM z14 (M/T 3906) Technical Introduction*, SG24-8450
- ▶ *IBM z14 ZR1 Technical Introduction*, SG24-8550

You can search for, view, download, or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other publications

The publication *IBM Z 8561 Installation Manual for Physical Planning*, GC28-7002 also is relevant as another information source.

Online resources

The IBM Resource Link for documentation and tools website is also relevant as another information source:

<http://www.ibm.com/servers/resourceLink>

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services



Redbooks

IBM z15 (8561) Technical Guide

SG24-8851-00

ISBN 0738458120



(1.0" spine)

0.875" x 1.498"

460 <-> 788 pages



SG24-8851-00

ISBN 0738458120

Printed in U.S.A.

Get connected

