Second-Generation Intel Xeon Scalable Processor Selection Guide for Virtual Client Computing on Cisco UCS with Citrix Virtual Apps and Desktops 7

Last Updated: July 8, 2019

Contents

What you will learn

Intel recently announced the availability of the second-generation Intel® Xeon® Scalable processors. Intel sees across an evolving digital world disruptive and emerging technology trends in business, industry, science, and entertainment. According to Intel, by 2020, the success of half of the world's Global 2000 companies will depend on their ability to create digitally enhanced products and services.

Intel suggests that this global transformation is rapidly scaling the demand for flexible computing, storage, and networking resources. Future workloads will require infrastructure that can scale to support rapid response to changing demands and scale requirements.

The Intel Xeon 2nd Gen Scalable platform provides a strong foundation for that evolution. This processor line enables new platform capabilities across computing, memory, storage, network, and security resources.

The new processor line introduces a unique persistent memory class, called Intel® Optane™ DC persistent memory. This new class of memory and storage innovation is designed for data-centric environments. Individual memory modules can be as large as 512 GB, allowing up to 36 TB of system-level memory when combined with traditional DRAM.

Cisco has evaluated the new processor line for its use in virtual client computing to provide our customers with guidance as to which processors provide the best starting price-to-performance ratio for three key benchmark workloads. These workloads represent three user personas and two of the most commonly used delivery techniques for each persona.

The main benefits of the second-generation Intel Xeon Scalable processor Gold line for virtual client computing, sometimes referred to as virtual desktop infrastructure (VDI), include the following:

- Intel Xeon Gold 6200 and 5200 series processors support higher memory speeds (2933 MHz for the 6200 series), enhanced memory capacity, and four-socket scalability.
- Intel Xeon Gold 6200 and 5200 series processors support advanced reliability and hardware-enhanced security.
- With higher core counts and processor frequencies, higher user densities are possible.
- Lower-core-count high-frequency processors facilitate enhanced performance for high-performance professional graphics applications when coupled with server graphics processing units (GPUs).

This document provides an overview of the fifth-generation Cisco Unified Computing System™ (Cisco UCS®) product line, an update on the latest VMware ESXi and Citrix products that support the second-generation Intel Xeon Scalable processors, and an overview of our selection process and test methodology.

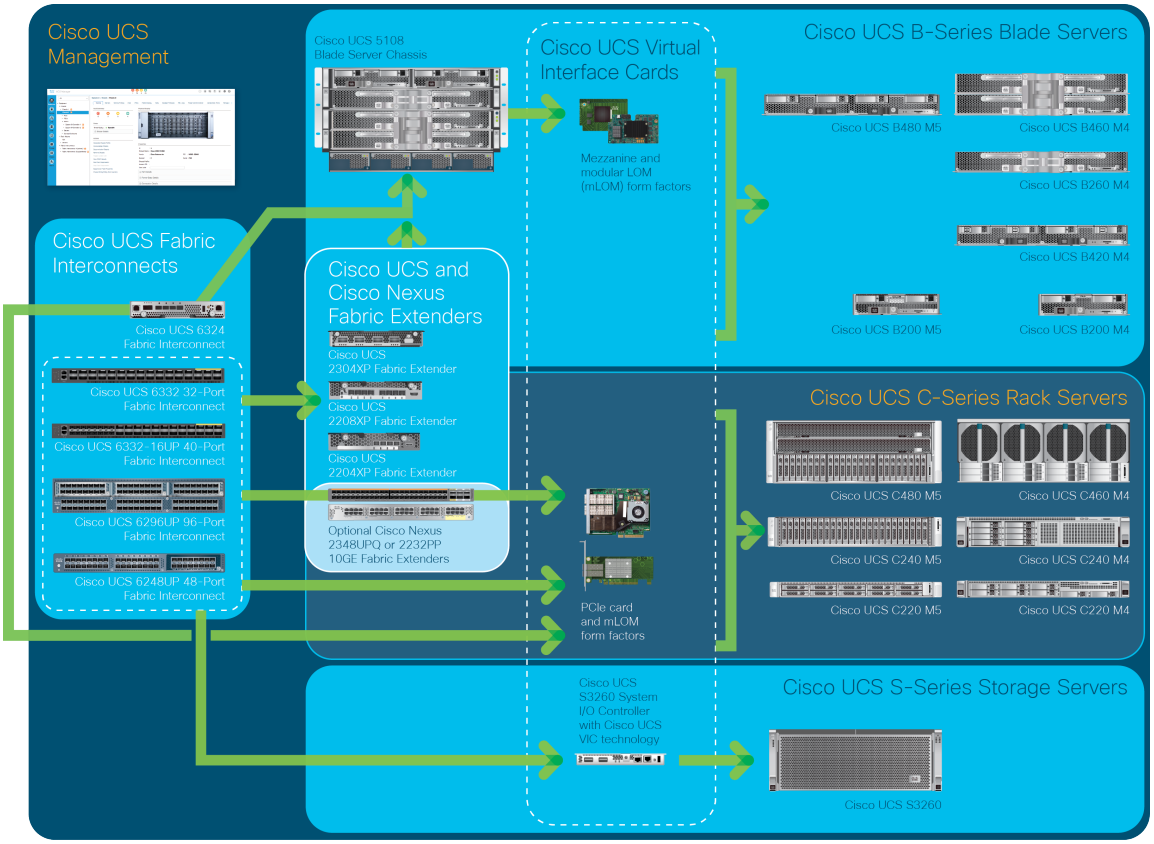
The document concludes with guidance for the starting configuration of second-generation Intel Xeon Scalable processors by user persona and delivery type.

First, though, this document introduces Cisco UCS, VMware ESXi, and Citrix Virtual Apps and Desktops.

Cisco Unified Computing System

Cisco UCS is a next-generation data center platform that unites computing, networking, and storage access. The platform, optimized for virtual environments, is designed using open industry-standard technologies and aims to reduce total cost of ownership (TCO) and increase business agility. The system integrates a low-latency lossless 40 Gigabit Ethernet unified network fabric with enterprise-class x86-architecture servers. It is an integrated, scalable, multichassis platform in which all resources participate in a unified management domain (Figure 1).

Figure 1. Cisco UCS components



The main components of Cisco UCS are:

- Computing: The system is based on an entirely new class of computing system that incorporates blade servers and modular servers based on Intel processors.
- Network: The system is integrated onto a low-latency, lossless, 40-Gbps unified network fabric. This network foundation consolidates LANs, SANs, and high-performance computing (HPC) networks, which are separate networks today. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables and by decreasing power and cooling requirements.
- Virtualization: The system unleashes the full potential of virtualization by enhancing the scalability, performance, and operational control of virtual environments. Cisco security, policy enforcement, and diagnostic features are now extended into virtualized environments to better support changing business and IT requirements.
- Storage access: The system provides consolidated access to local storage, SAN storage, and network-attached storage (NAS) over the unified fabric. With storage access unified, Cisco UCS can access storage over Ethernet, Fibre Channel, Fibre Channel over Ethernet (FCoE), and Small Computer System Interface over IP (iSCSI) protocols. This capability provides customers with a choice for storage access and investment protection. In addition, server administrators can preassign storage-access policies for system connectivity to storage resources, simplifying storage connectivity and management and helping increase productivity.

- Management: Cisco UCS uniquely integrates all system components, enabling the entire solution to be managed as a single entity by Cisco UCS Manager. The manager has an intuitive GUI, a command-line interface (CLI), and a robust API for managing all system configuration processes and operations.

Cisco UCS is designed to deliver:

- Reduced TCO and increased business agility
- Increased IT staff productivity through just-in-time provisioning and mobility support
- A cohesive, integrated system that unifies the technology in the data center; the system is managed, serviced, and tested as a whole
- Scalability through design for hundreds of discrete servers and thousands of virtual machines and the capability to scale I/O bandwidth to match demand
- Industry standards supported by a partner ecosystem of industry leaders

Cisco UCS Manager

Cisco UCS Manager provides unified, embedded management of all software and hardware components of Cisco UCS through an intuitive GUI, a CLI, and an XML API. The manager provides a unified management domain with centralized management capabilities and can control multiple chassis and thousands of virtual machines. Tightly integrated Cisco UCS manager and NVIDIA GPU cards provide better management of firmware and graphics card configuration.

Cisco UCS 6332 Fabric Interconnect

The Cisco UCS 6332 Fabric Interconnect (Figure 2) is the management and communication backbone for Cisco UCS B-Series Blade Servers, C-Series Rack Servers, and 5100 Series Blade Server Chassis. All servers attached to 6332 Fabric Interconnects become part of one highly available management domain.

Because they support unified fabric, Cisco UCS 6300 Series Fabric Interconnects provide both LAN and SAN connectivity for all servers within their domains. For more details see, https://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-b-series-blade-servers/6332-specsheet.pdf.

The 6332 Fabric Interconnect provides these features and capabilities:

- Bandwidth of up to 2.56 Tbps of full-duplex throughput
- Thirty-two 40-Gbps Enhanced Quad Small Form-Factor Pluggable (QSFP+) ports in a 1-rack-unit (1RU) form factor
- Support for four 10-Gbps breakout cables
- Ports capable of line-rate, low-latency, lossless 40 Gigabit Ethernet and FCoE
- Centralized unified management with Cisco UCS Manager
- Efficient cooling and serviceability

Figure 2. Cisco UCS 6332 Fabric Interconnect

Cisco UCS B-Series Blade Servers

Cisco UCS B-Series Blade Servers are based on Intel Xeon processors. They work with virtualized and nonvirtualized applications to increase performance, energy efficiency, flexibility, and administrator productivity. Cisco UCS servers offer a balanced approach between the need for servers with smaller form factors and more server density per rack. Cisco UCS blade servers deliver a high-performance virtual client computing end-user experience.

Cisco UCS B200 M5 Blade Server

Delivering performance, versatility, and density without compromise, the Cisco UCS B200 M5 Blade Server (Figure 3) addresses a broad set of workloads, from IT and web infrastructure to distributed databases. The enterprise-class B200 M5 Blade Server extends the capabilities of the Cisco UCS portfolio in a half-width blade form factor. The B200 M5 harnesses the power of the latest Intel Xeon Scalable CPUs with up to 3072 GB of RAM (using 128-GB DIMMs), two solid-state disks (SSDs) or hard-disk drives (HDDs), and connectivity with throughput of up to 80 Gbps.

The Cisco UCS B200 M5 server mounts in a Cisco UCS 5100 Series Blade Server Chassis or Cisco UCS Mini blade server chassis. It has 24 total slots for error-correcting code (ECC) registered DIMMs (RDIMMs) or load-reduced DIMMs (LR DIMMs). It supports one connector for the Cisco UCS Virtual Interface Card (VIC) 1340 adapter, which provides Ethernet and FCoE.

The B200 M5 has one rear mezzanine adapter slot, which can be configured with a Cisco UCS port expander card for additional connectivity bandwidth or with an NVIDIA P6 GPU. These hardware options enable an additional four ports of the VIC 1340, bringing the total capability of the VIC 1340 to a dual native 40-Gbps interface or a dual 4 x 10 Gigabit Ethernet port-channel interface. Alternatively, the same rear mezzanine adapter slot can be configured with an NVIDIA P6 GPU.

The B200 M5 also has one front mezzanine slot. The B200 M5 can be ordered with or without a front mezzanine card. The front mezzanine card can accommodate a storage controller or NVIDIA P6 GPU.

For more information, see https://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-b-series-blade-servers/b200m5-specsheet.pdf.

Figure 3. Cisco UCS B200 M5 Blade Server front view

Cisco UCS B480 M5 Blade Server

The enterprise-class Cisco UCS B480 M5 Blade Server (Figure 4) delivers market-leading performance, versatility, and density without compromise for memory-intensive mission-critical enterprise applications and virtualized workloads, among others.

With the B480 M5, you can quickly deploy stateless physical and virtual workloads with the programmability that Cisco UCS Manager and Cisco® SingleConnect technology enable. Customers gain support for Intel Xeon Scalable processors, higher-density 256-GB DDR4 DIMMs, new second-generation Intel Xeon Scalable processors, and the new Intel Optane DC persistent memory. Additionally, these B480 M5 servers offer up to 12 TB of DDR4 memory or 18 TB using 24 x 256-GB DDR4 DIMMs and 24 x 512-GB Intel Optane DC persistent memory; four SAS, SATA, and Non-Volatile Memory Express (NVMe) drives; M.2 storage; up to four GPUs; and 160 Gigabit Ethernet connectivity for I/O throughput, all leading to exceptional performance, flexibility, and I/O throughput to run your most demanding applications.

The B480 M5 is a full-width blade server supported by the Cisco UCS 5108 Blade Server Chassis. The 5108 chassis and the Cisco UCS B-Series Blade Servers provide inherent architectural advantages:

- Through Cisco UCS, gives you the architectural advantage of not having to power, cool, manage, and purchase excess switches (management, storage, and networking), host bus adapters (HBAs), and network interface cards (NICs) in each blade chassis
- Reduces TCO by removing management modules from the chassis, making the chassis stateless
- Provides a single, highly available Cisco UCS management domain for all system chassis and rack servers, reducing administrative tasks

The Cisco UCS B480 M5 Blade Server delivers flexibility, density, and expandability in a 4-socket, full-width form factor for enterprise and mission-critical applications. It offers:

- Four new second-generation Intel Xeon Scalable CPUs (up to 28 cores per socket)
- Four existing Intel Xeon Scalable CPUs (up to 28 cores per socket)
- Support for higher-density DDR4 memory: from 6 TB (128-GB DDR4 DIMMs) to 12 TB (256-GB DDR4 DIMMs)
- Increased memory speeds: from 2666 MHz to 2933 MHz
- Intel Optane DC persistent memory modules (DCPMMs): 128, 256, and 512 GB
- Up to 18 TB using 24 x 256-GB DDR4 DIMMs and 24 x 512-GB Intel Optane DCPMMs
- Cisco FlexStorage storage subsystem

- Five mezzanine adapters and support for up to four NVIDIA GPUs
- Cisco UCS VIC 1340 modular LAN on motherboard (mLOM) and upcoming fourth-generation VIC mLOM
- Internal Secure Digital (SD) and M.2 boot options

Note: Each Intel Optane DCPMM requires a DDR4 DIMM for deployment (for example, 12 Intel Optane DCPMMs require 12 DDR4 DIMMs). The sizes of the persistent memory modules must be the same, and the DDR4 DIMMs must be the same. But between the persistent memory modules and DIMMs, the sizes can vary.

Figure 4. Cisco UCS B480 M5 Blade Server (front view)

Cisco UCS C-Series Rack Servers

Cisco UCS C-Series Rack Servers keep pace with Intel Xeon processor innovation by offering the latest processors with an increase in processor frequency and improved security and availability features. With the increased performance provided by the Intel Xeon Scalable processors, Cisco UCS C-Series servers offer an improved price-to-performance ratio. They also extend Cisco UCS innovations to an industry-standard rack-mount form factor, including a standards-based unified network fabric, Cisco VN-Link virtualization support, and Cisco Extended Memory Technology.

Designed to operate both in standalone environments and as part of a Cisco UCS managed configuration, these servers enable organizations to deploy systems incrementally—using as many or as few servers as needed—on a schedule that best meets the organization's timing and budget. Cisco UCS C-Series servers offer investment protection through the capability to deploy them either as standalone servers or as part of Cisco UCS.

One compelling reason that many organizations prefer rack-mount servers is the wide range of I/O options available in the form of PCI Express (PCIe) adapters. Cisco UCS C-Series servers support a broad range of I/O options, including interfaces supported by Cisco as well as adapters from third parties.

Cisco UCS C220 M5 Rack Server

The Cisco UCS C220 M5 Rack Server (Figures 5 and 6) is among the most versatile general-purpose enterprise infrastructure and application server in the industry. It is a high-density 2-socket rack server that delivers industry-leading performance and efficiency for a wide range of workloads, including virtualization, collaboration, and bare-metal applications. The Cisco UCS C-Series Rack Servers can be deployed as standalone servers or as part of Cisco UCS to take advantage of Cisco's standards-based unified computing innovations that help reduce customers' TCO and increase their business agility.

The Cisco UCS C220 M5 server extends the capabilities of the Cisco UCS portfolio in a 1RU form factor. It incorporates the Intel Xeon Scalable processors, supporting up to 20 percent more cores per socket, twice the memory capacity, 20 percent greater storage density, and five times more PCIe NVMe SSDs compared to the previous generation of servers. These

improvements deliver significant performance and efficiency gains that will improve your application performance. The C220 M5 delivers outstanding levels of expandability and performance in a compact package, with:

- The latest second-generation Intel Xeon Scalable CPUs, with up to 28 cores per socket
- Support for first-generation Intel Xeon Scalable CPUs, with up to 28 cores per socket
- Up to 24 DDR4 DIMMs for improved performance
- Support for the Intel Optane DC persistent memory (128, 256, and 512 GB)
- Up to 10 small-form-factor (SFF) 2.5-inch drives or 4 large-form-factor (LFF) 3.5-inch drives (77 TB of storage capacity with all NVMe PCIe SSDs)
- Support for a 12-Gbps SAS modular RAID controller in a dedicated slot, leaving the remaining PCIe Generation 3.0 slots available for other expansion cards
- mLOM slot that can be used to install a Cisco UCS VIC without consuming a PCIe slot
- Dual embedded Intel x550 10GBASE-T LAN-on-motherboard (LOM) ports

Note: Each Intel Optane DCPMM requires a DDR4 DIMM for deployment (for example, 12 Intel Optane DCPMMs require 12 DDR4 DIMMs). The sizes of the persistent memory modules must be the same, and the DDR4 DIMMs must be the same. But between the persistent memory modules and DIMMs, the sizes can vary.

Figure 5. Cisco UCS C220 M5 Rack Server

Figure 6. Cisco UCS C220 M5 Rack Server Rear View

The server includes two PCIe 3.0 slots plus one dedicated 12-Gbps RAID controller slot and one dedicated mLOM slot for the Cisco VIC converged network adapter (CNA).

Cisco UCS C240 M5 Rack Server

The Cisco UCS C240 M5 Rack Server (Figures 7 and 8 and Table 1) is designed for both performance and expandability over a wide range of storage-intensive infrastructure workloads, from big data to collaboration.

The C240 M5 SFF server extends the capabilities of the Cisco UCS portfolio in a 2RU form factor with the addition of the Intel Xeon Scalable processor family, 24 DIMM slots for 2666-MHz DDR4 DIMMs, and up to 128-GB capacity points, up to 6 PCIe 3.0 slots, and up to 26 internal SFF drives. The C240 M5 SFF server also includes one dedicated internal slot for a 12-GB SAS storage controller card. The C240 M5 server includes a dedicated internal mLOM slot for installation of a Cisco VIC or third-party NIC without consuming a PCI slot, in addition to two 10GBASE-T Intel x550 embedded (on the motherboard) LOM ports.

In addition, the C240 M5 offers outstanding levels of internal memory and storage expandability with exceptional performance. It delivers:

- Up to 24 DDR4 DIMMs at speeds up to 2666 MHz for improved performance and lower power consumption
- One or two Intel Xeon Scalable CPUs
- Up to six PCIe 3.0 slots (four full-height, full-length for GPU)
- Six hot-swappable fans for front-to-rear cooling
- Twenty-four SFF front-facing SAS/SATA HDDs or SAS/SATA SSDs
- Optionally, up to two front-facing SFF NVMe PCIe SSDs (replacing SAS/SATA drives). These drives must be placed in front drive bays 1 and 2 only and are controlled from Riser 2, Option C.
- Optionally, up to two SFF, rear-facing SAS/SATA HDDs and SSDs or up to two rear-facing SFF NVMe PCIe SSDs
  - Rear-facing SFF NVMe drives connected from Riser 2, Option B or C
  - Support for 12-Gbps SAS drives
- Dedicated mLOM slot on the motherboard, which can flexibly accommodate the following cards:
  - Cisco VICs
  - Quad-port Intel i350 1 Gigabit Ethernet RJ-45 mLOM NIC
- Two 1 Gigabit Ethernet embedded LOM ports
- Support for up to two double-wide NVIDIA GPUs, providing a graphics-rich experience to more virtual users
- Excellent reliability, availability, and serviceability (RAS) features with tool-free CPU insertion, easy-to-use latching lid, and hot-swappable and hot-pluggable components
- One slot for a MicroSD card on PCIe Riser 1 (Options 1 and 1B)
  - The MicroSD card serves as a dedicated local resource for utilities such as the Cisco Host Upgrade Utility (HUU).
  - Images can be pulled from a file share (Network File System [NFS] or Common Internet File System [CIFS]) and uploaded to the cards for future use.
- A mini-storage module connector on the motherboard that supports either:
  - SD card module with two SD card slots; mixing different-capacity SD cards is not supported
  - M.2 module with two SATA M.2 SSD slots; mixing different-capacity M.2 modules is not supported

Note: SD cards and M.2 modules cannot be mixed. M.2 modules do not support RAID 1 with VMware. Only Microsoft Windows and Linux are supported.

The C240 M5 also increases performance and customer choice over many types of storage-intensive applications such as:

- Collaboration
- Small and medium-sized business (SMB) databases
- Big data infrastructure
- Virtualization and consolidation
- Storage servers
- High-performance appliances

The C240 M5 can be deployed as a standalone server or as part of a Cisco UCS managed domain. Cisco UCS unifies computing, networking, management, virtualization, and storage access into a single integrated architecture that enables end-to-end server visibility, management, and control in both bare-metal and virtualized environments. Within a Cisco UCS deployment, the C240 M5 takes advantage of Cisco's standards-based unified computing innovations, which significantly reduce customers' TCO and increase business agility.

For more information about the Cisco UCS C240 M5 Rack Server, see https://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-c-series-rack-servers/c240m5-sff-specsheet.pdf.

Figure 7. Cisco UCS C240 M5 Rack Server

Figure 8. Cisco UCS C240 M5Rack Server rear view

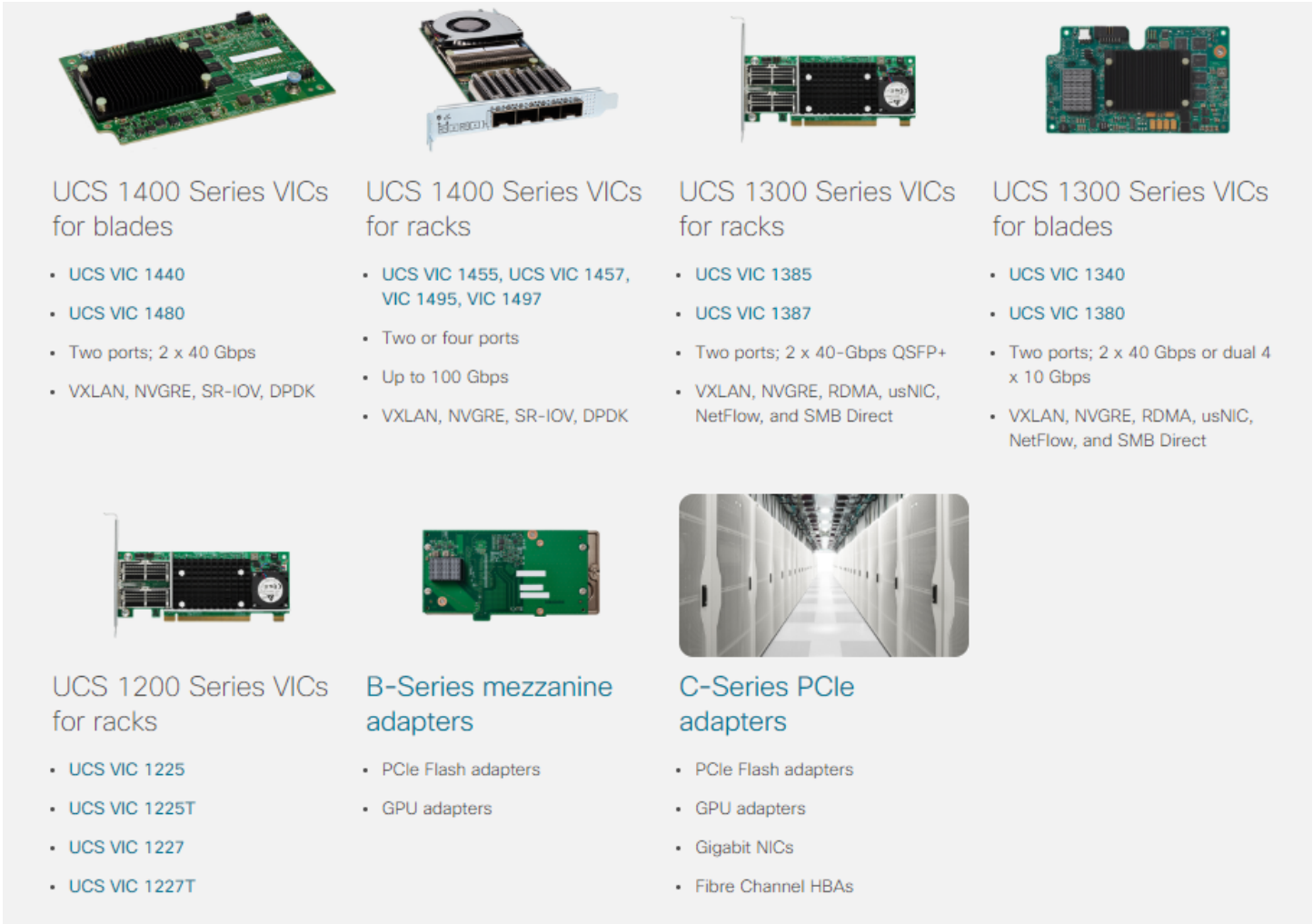
Table 1. Cisco UCS C240 M5 PCIe slots

PCIe slot	Length	Lane
1	Half	x8
2	Full	x16
3	Half	x8
4	Half	x8
5	Full	x16
6	Full	x8

Cisco UCS virtual interface cards

Whether you are using blade, rack, or storage servers from Cisco, the Cisco UCS VIC (Figure 9) provides optimal connectivity.

Figure 9. Cisco VICs for Cisco UCS blade and rack servers



Cisco UCS VIC 1340

The Cisco UCS VIC 1340 (Figure 10) is a 2-port 40-Gbps Ethernet or dual 4 x 10-Gbps Ethernet, FCoE-capable mLOM designed exclusively for the M5 generation of Cisco UCS B-Series Blade Servers. When used in combination with an optional port expander, the VIC 1340 is enabled for two ports of 40-Gbps Ethernet. The VIC 1340 enables a policy-based, stateless, agile server infrastructure that can present more than 256 PCIe standards-compliant interfaces to the host that can be dynamically configured as either NICs or HBAs. In addition, the VIC 1340 supports Cisco Virtual Machine Fabric Extender (VM-FEX) technology, which extends the Cisco UCS fabric interconnect ports to virtual machines, simplifying server virtualization deployment and management.

For more information, see https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/ucs-virtual-interface-card-1340/datasheet-c78-732517.html.

Figure 10. Cisco UCS VIC 1340

Cisco UCS VIC 1387

The Cisco UCS VIC 1387 (Figure 11) is a dual-port Enhanced Small Form-Factor Pluggable (SFP+) 40-Gbps Ethernet and FCoE-capable PCIe mLOM adapter installed in Cisco UCS C-Series Rack Servers. The mLOM slot can be used to install a Cisco VIC without consuming a PCIe slot, which provides greater I/O expandability. It incorporates next-generation CNA technology from Cisco, providing investment protection for future feature releases. The card enables a policy-based, stateless, agile server infrastructure that can present more than 256 PCIe standards-compliant interfaces to the host that can be dynamically configured as either NICs or HBAs. The personality of the card is determined dynamically at boot time using the service profile associated with the server. The number, type (NIC or HBA), identity (MAC address and World Wide Name [WWN]), failover policy, bandwidth, and quality-of-service (QoS) policies of the PCIe interfaces are all determined using the service profile.

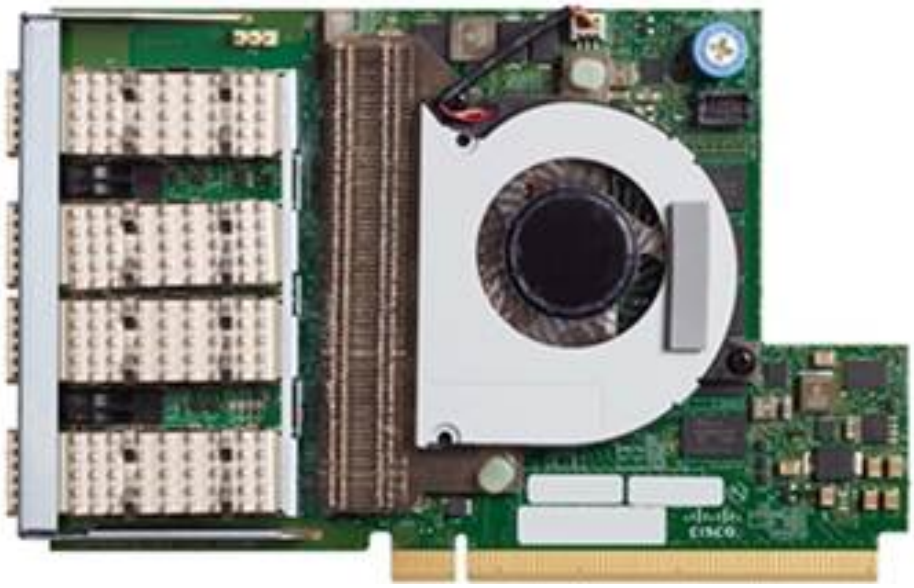
For more information about the VIC, see https://www.cisco.com/c/en/us/products/interfaces-modules/ucs-virtual-interface-card-1387/index.html.

Figure 11. Cisco UCS VIC 1387

Cisco UCS VIC 1457

The Cisco UCS VIC 1457 (Figure 12) is a quad-port SFP28 mLOM card designed for the M5 generation of Cisco UCS C-Series Rack Servers. The card supports 10 and 25 Gigabit Ethernet and FCoE. The card can present PCIe standards-compliant interfaces to the host, and these can be dynamically configured as either NICs or HBAs.

Figure 12. Cisco UCS VIC 1457



Cisco UCS VIC 1497

The Cisco VIC 1497 (Figure 13) is a dual-port quad SFP28 mLOM card designed for the M5 generation of Cisco UCS C-Series Rack Servers. The card supports 40 and 100 Gigabit Ethernet and FCoE. The card can present PCIe standards-compliant interfaces to the host, and these can be dynamically configured as NICs or HBAs.

Figure 13. Cisco UCS VIC 1497

VMware vSphere 6.7 U2

VMware provides virtualization software. VMware's enterprise software hypervisors for servers—VMware vSphere ESX, vSphere ESXi, and vSphere—are bare-metal hypervisors that run directly on server hardware without requiring an additional underlying operating system. VMware vCenter Server for vSphere provides central management and complete control and visibility into clusters, hosts, virtual machines, storage, networking, and other critical elements of your virtual infrastructure.

VMware vSphere 6.7 introduces many enhancements to vSphere Hypervisor, VMware virtual machines, vCenter Server, virtual storage, and virtual networking, further extending the core capabilities of the vSphere platform.

Now VMware has announced vSphere 6.7, which has one of the most comprehensive feature sets of any of the recent releases of vSphere. The vCenter Server Appliance is taking charge in this release, with several new features that are discussed in this document. For starters, the installer has been overhauled with a new modern look and feel. Users of both Linux and Mac OS will note that the installer is now supported on those platforms along with Microsoft Windows. The vCenter Server Appliance also now has several exclusive features such as the following:

- Migration capabilities
- Improved appliance management
- VMware Update Manager
- Native high availability
- Built-in backup and restore capabilities

VMware vSphere Client

VMware vSphere 6.7 provides a fully supported version of the HTML-5-based vSphere Client that will run alongside the vSphere Web Client. The vSphere Client is built in to vCenter Server 6.7 (both Windows and Appliance versions) and is enabled by default. Although the HTML-5 based vSphere Client does not have full feature parity, the upgrade prioritized many of the day-to-day tasks of administrators, and VMware continues to seek feedback about items that will enable customers to use it full time. The vSphere Web Client continues to be accessible through http://<vcenter_fqdn>/vsphere-client, and the vSphere Client can be accessed through http://<vcenter_fqdn>/ui.

VMware is periodically updating the vSphere Client outside the normal vCenter Server release cycle. To make sure it is easy and simple for customers to stay up-to-date, the vSphere Client can be updated without affecting the rest of vCenter Server.

These are some of the benefits of the new vSphere Client:

- Clean, consistent user interface built on VMware's new Clarity user interface standards (to be adopted across the portfolio)
- Built on HTML-5 so that it is truly a cross-browser and cross-platform application
- No browser plug-ins to install or manage
- Integrated into vCenter Server 6.7 and fully supported
- Full support for enhanced linked mode

VMware ESXi 6.7 hypervisor

VMware vSphere 6.7 introduces the following new features in the hypervisor:

- Scalability improvements
  - ESXi 6.7 dramatically increases the scalability of the platform. With vSphere Hypervisor 6.0, clusters can scale to as many as 64 hosts: up from 32 in previous releases. With 64 hosts in a cluster, vSphere 6.0 can support 8000 virtual machines in a single cluster. This capability enables greater consolidation ratios, more efficient use of VMware vSphere Distributed Resource Scheduler (DRS), and fewer clusters that must be separately managed. Each vSphere Hypervisor 6.7 instance can support up to 480 logical CPUs, 12 TB of RAM, and 1024 virtual machines. By using the newest hardware advances, ESXi 6.7 enables the virtualization of applications that previously had been thought to be nonvirtualizable.
- ESXI 6.7 security enhancements
  - Account management: ESXi 6.7 enables management of local accounts on the ESXi server using new ESXi CLI commands. The capability to add, list, remove, and modify accounts across all hosts in a cluster can be centrally managed using a vCenter Server system. Previously, the account and permission management functions for ESXi hosts were available only for direct host connections. The setup, removal, and listing of local permissions on ESXi servers can also be centrally managed.
  - Account lockout: ESXi Host Advanced System Settings have two new options for the management of failed local account login attempts and account lockout duration. These parameters affect Secure Shell (SSH) and vSphere Web Services connections, but not ESXi direct console user interface (DCUI) or console shell access.
  - Password complexity rules: In previous versions of ESXi, password complexity changes had to be made by manually editing the /etc/pam.d/passwd file on each ESXi host. In vSphere 6.0, an entry in Host Advanced System Settings enables changes to be centrally managed for all hosts in a cluster.
  - Improved auditability of ESXi administrator actions: Prior to vSphere 6.0, actions at the vCenter Server level by a named user appeared in ESXi logs with the vpxuser user name: for example, [user=vpxuser]. In vSphere 6.7, all actions at the

vCenter Server level for an ESXi server appear in the ESXi logs with the vCenter Server user name: for example, [user=vpxuser: DOMAIN\User]. This approach provides a better audit trail for actions run on a vCenter Server instance that conducted corresponding tasks on the ESXi hosts.

- Flexible lockdown modes: Prior to vSphere 6.7, only one lockdown mode was available. Feedback from customers indicated that this lockdown mode was inflexible in some use cases. With vSphere 6.7, two lockdown modes are available:
  - In normal lockdown mode, DCUI access is not stopped, and users on the DCUI access list can access the DCUI.
  - In strict lockdown mode, the DCUI is stopped.
- Exception users: vSphere 6.0 offers a new function called exception users. Exception users are local accounts or Microsoft Active Directory accounts with permissions defined locally on the host to which these users have host access. These exception users are not recommended for general user accounts, but they are recommended for use by third-party applications—for service accounts, for example—that need host access when either normal or strict lockdown mode is enabled. Permissions on these ac-counts should be set to the bare minimum required for the application to perform its task and with an account that needs only read-only permissions on the ESXi host.
- Smart card authentication to DCUI: This function is for U.S. federal customers only. It enables DCUI login access using a Common Access Card (CAC) and Personal Identity Verification (PIV). The ESXi host must be part of an Active Directory domain.

Citrix Virtual Apps and Desktops 7 1808

Citrix Virtual Apps and Desktops 7 1808 provides the following features and enhancements.

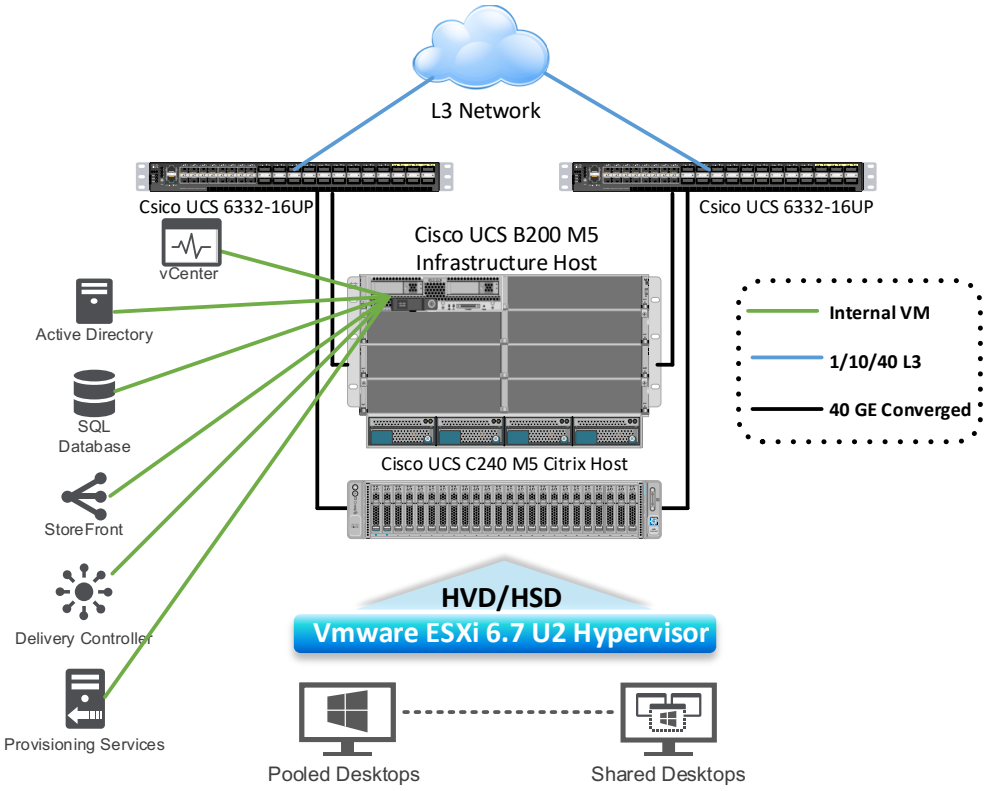
- Citrix Virtual Apps and Desktops: Citrix Virtual Apps and Desktops offers a virtual app and desktop solution, provided as a cloud service and as an on-premises product, giving employees the freedom to work from anywhere on any device while reducing IT costs. Deliver Windows, Linux, web, and software-as-a-service SaaS applications or full virtual desktops from any cloud: public, on premises, or hybrid. Virtual Apps and Desktops was formerly called XenApp and XenDesktop.
- Citrix Workspace app: The Citrix Workspace app incorporates existing Citrix Receiver technology as well as the other Citrix Workspace client technologies. It has been enhanced to deliver additional capabilities to provide end users with a unified, contextual experience in which they can interact with all the work apps, files, and devices they need to do their best work. For more information, see this blog post.
- Citrix SD-WAN: NetScaler SD-WAN, a crucial technology for Citrix customers and partners transforming their branch networks and WANs with cloud technology, is now Citrix SD-WAN.
- Citrix Secure Web Gateway: Citrix Secure Web Gateway service was previously known as NetScaler Secure Web Gateway.
- Citrix Gateway: Robust NetScaler Unified Gateway, which allows secure, contextual access to the apps and data you need to do your best work, is now Citrix Gateway.
- Citrix Content Collaboration and Citrix Files for Windows: The advanced access, collaboration, workflows, rights management, and integration features of ShareFile are now available in the Citrix Content Collaboration component set in the secure, contextual, integrated Citrix Workspace. Citrix Files for Windows allows you to access your Content Collaboration files directly through a mapped drive, providing a native Windows Explorer experience.
- Citrix Hypervisor: The technology from XenServer for virtualization infrastructure, based on the XenProject hypervisor, is now Citrix Hypervisor.

For additional details, refer to the Citrix documentation here.

Citrix Virtual Apps and Desktops 7 test platform for second-generation Intel Xeon Scalable processors

Figure 14 provides an overview of the test platform used to evaluate and select processors.

Figure 14. Reference architecture



The solution includes the following hardware components:

- Cisco UCS C240 M5 Rack Server for Horizon payloads (two second-generation Intel Xeon Scalable Gold processors with 1.5 TB of memory (64 GB x 24 DIMMs at 2666 to 2933 MHz) with:
  - Intel Xeon Scalable processor Gold series 5220 (18 cores at 2.2 GHz with 125 watts (W) and 2666-MHz for memory)
  - Intel Xeon Scalable processor Gold series 6230 (20 cores at 2.1 GHz with 125W and 2933 MHz for memory)
  - Intel Xeon Scalable processor Gold series 6248 (20 cores at 2.5GHz with 150W and 2933 MHz for memory)
- Cisco UCS B200 M5 Blade Server for Horizon infrastructure (two Intel Xeon Gold 5120 CPUs at 2.20 GHz) with 768 GB of memory (64 GB x 12 DIMMs at 2666 MHz)
- Cisco UCS VIC 1387 mLOM (Cisco UCS C240 M5 Rack Server)
- Cisco UCS VIC 1340 mLOM (Cisco UCS B200 M5 Blade Server)
- Two Cisco UCS 6332 Fabric Interconnects (third-generation fabric interconnects)
- Two Cisco Nexus® 93180YC-FX Switches (optional access switches)

The software components of the solution are as follows:

- Cisco UCS Firmware Release 4.0(3.104)
- VMware ESXi 6.7 U2 for VDI hosts
- Citrix Virtual Apps and Desktops 7 1808
- Microsoft Windows 10 64-bit (1607)
- Microsoft Server 2016 (1607)
- Microsoft Office 2016

Test strategy

To evaluate the second-generation Intel Xeon Scalable processors, we created a strategy to test for the optimal price-to-performance ratio for the three mainstream user personas for virtual client computing, also referred to as end-user computing by Citrix. They are:

- Task workers
- Knowledge workers
- Power users

In addition, we identified the two most common delivery mechanisms for each user type:

- Windows 10 hosted virtual desktops using HDX
- Windows Server 2016 hosted virtual server desktop sessions using Remote Desktop Session Host (RDSH)

We evaluated each processor tested against the combinations of Login VSI test workloads shown in Table 2, in benchmark mode.

Table 2. User persona and delivery mechanism combinations tested

Delivery Mechanism	Task Worker	Knowledge worker	Power user
Windows 10 (PVS) HVD	Tested	Tested	Tested
Windows Server 2016 (PVS) HSD session	Tested	Tested	

We started with our knowledge of the performance of the first-generation Intel Scalable Gold processors and compared their benchmark performance to that of the second-generation processors to guide our initial selections for evaluation. Our plan was to identify a processor for each user type that delivered the best price-to-performance ratio for both the VDI and RDSH delivery modalities for Citrix Virtual Apps and Desktops 7.

Test methodology

For each user type and processor combination tested, we created a Citrix Virtual Apps and Desktops 7 virtual machine with specifications as shown in Table 3.

Table 3. Citrix Virtual Apps and Desktops 7 virtual machine

Combination	Virtual CPU	Memory	Virtual NIC
Task worker: Windows 10	1 vCPU	2 GB of memory	1 x 10-GB vNIC
Knowledge worker: Windows 10	2 vCPUs	4 GB of memory	1 x 10-GB vNIC
Power user: Windows 10	4 vCPUs	4 GB of memory	1 x 10-GB vNIC
Windows Server 2016 RDSH: Virtual machine (all)	9 vCPUs	24 GB of memory	1 x 10-GB vNIC

For each user type, we selected one or more second-generation Intel Xeon Scalable processors for evaluation. We used SPECrate2017*_int_base and SPECrate2017*_int_fp data generated by Cisco for first- and second-generation processors to identify candidates.

We installed the chosen processor candidates in a Cisco UCS C240 M5 server and ran Login VSI benchmark mode tests at calculated maximum user densities to determine actual maximum user density per server. The maximum recommended user density is some number of users that complete the Login VSI workload with all attempted users active and logged off without triggering Login VSImax. In addition, CPU utilization on the host should not exceed 90 percent during the test.

We used the maximum recommended user density achieved to determine server loading in a server maintenance or failure scenario: typically N-1. We expect that customers would run their environment only at this load in those cases.

We compared performance and price per user at the maximum recommended user density to determine the best processor for the user type. We included Windows 10 and RDSH metrics where appropriate.

Test data

This section presents the data from the test runs for the processor selected for each user type.

Task workers

Task workers are individuals in an organization who use a limited number of applications to perform their duties. Examples of task workers are customer service agents, medical transcriptionists, accounts receivable and payable workers, and some enterprise resource planning (ERP) workers.

In many cases, these workers are well served by RDSH server sessions or published applications. In some cases, organizations provide a light-duty Windows 10 virtual desktop to these users.

We tested both use cases using the Login VSI Task Worker workload in benchmark mode. You can find additional information about Login VSI and all of the workloads we tested for this document here.

In addition to using the Login VSI test suite, we measured host utilization by gathering data from ESXtop. We also captured perfmon data from sample virtual machines during the full server load tests.

Windows 10 and Citrix Virtual Apps and Desktops 7 single-server synopsis: Intel Xeon Scalable Gold processor 5220

The test results are summarized here and in Figures 15, 16, and 17.

- Operating system: Windows 10 64-bit (1607) with Citrix optimizations
- 1 vCPU; 2 GB of RAM
- Number of users: 260 users running Login VSI Task Worker workload with Windows 10
- No VSImax; Login VSI baseline = 812 milliseconds (ms)

Figure 15. Login VSI end-user experience summary

Summary | Settings | VSImax v4 | VSImax v4 Detailed | VSImax v4 Detailed Weighted | VSImax v4 Scatter | UMEM | IO | CPU | ZLC | ZHC | NFP | NFO

0609-0821-5220- 260HVDtask

Successfully completed Login VSI test with **260** **taskworker** sessions. VSImax (system saturation) was not reached. All Login VSI users completed the test.

Test result review

260 sessions were configured to be launched in **2880** seconds.

In total **0** sessions failed during the test:

- **0** sessions was/were not successfully launched
- **0** launched sessions failed to become active
- **260** sessions were active during the test
- **0** sessions got stuck during the test (before VSImax threshold)

With **260** sessions the maximum capacity VSImax (v4.1) **taskworker** was not reached with a Login VSI baseline performance score of **812**

Login VSI index average score is **714** lower than threshold. It might be possible to launch more sessions in this configuration.

Baseline performance of **812** is: **Good**

Figure 16. Login VSI end-user experience performance chart

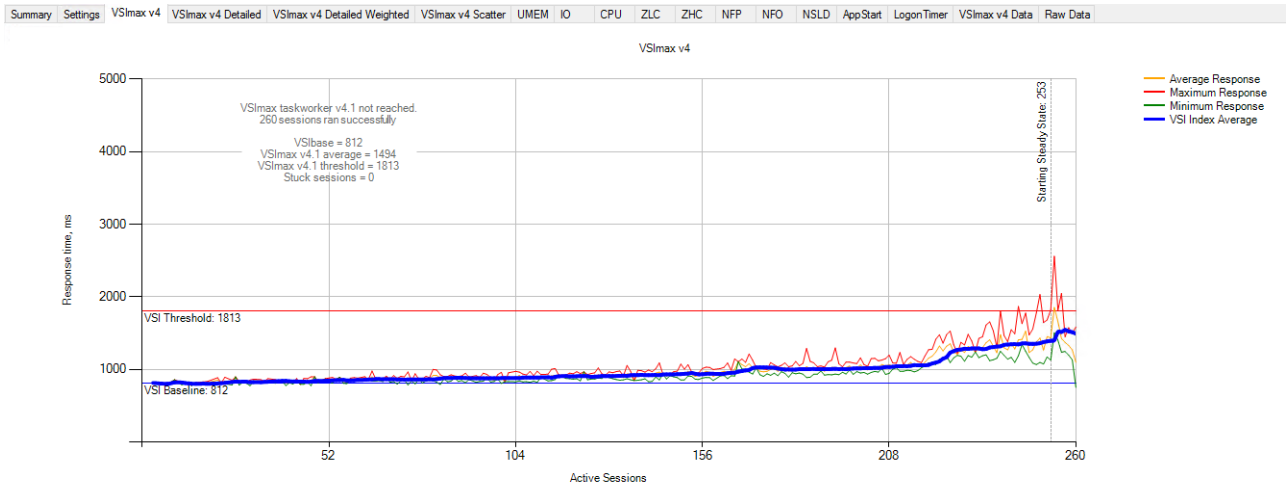
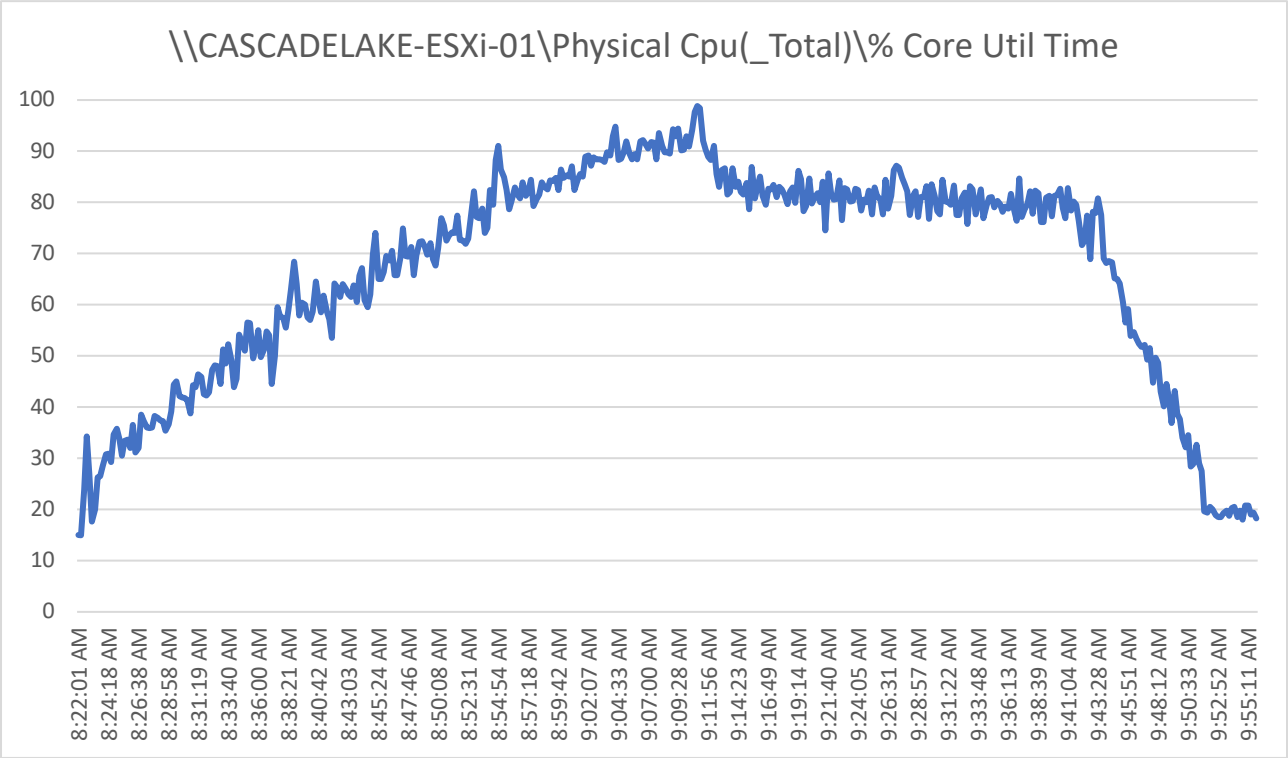


Figure 17. VMware ESXi host CPU Util %during testing



Windows Server 2016 and Citrix Virtual Apps and Desktops 7 RDSH synopsis: Intel Xeon Scalable Gold processor 5220

The test results are summarized here and in Figures 18, 19, 20, and 21.

- Operating system: Windows Server 2016 (1607) with Citrix optimizations
- 9 vCPUs; 24 GB of RAM per virtual machine
- 11 virtual machines per host
- Number of users: 360 users running Login VSI Task Worker workload RDSH sessions
- No VSImax; Login VSI baseline = 541 ms

Figure 18. Login VSI end-user experience summary

Summary | Settings | VSImax v4 | VSImax v4 Detailed | VSImax v4 Detailed Weighted | VSImax v4 Scatter | UMEM | IO | CPU | ZLC | ZHC | NFP | NFO

0610-1701-360HSDtask

Successfully completed Login VSI test with **360** **taskworker** sessions. VSImax (system saturation) was not reached. All Login VSI users completed the test.

Test result review

360 sessions were configured to be launched in **2880** seconds.

In total **0** sessions failed during the test:

- **0** sessions was/were not successfully launched
- **0** launched sessions failed to become active
- **360** sessions were active during the test
- **0** sessions got stuck during the test (before VSImax threshold)

With **360** sessions the maximum capacity VSImax (v4.1) **taskworker** was not reached with a Login VSI baseline performance score of **675**

Login VSI index average score is **541** lower than threshold. It might be possible to launch more sessions in this configuration.

Baseline performance of **675** is: **Very good**

Figure 19. Login VSI end-user experience performance chart

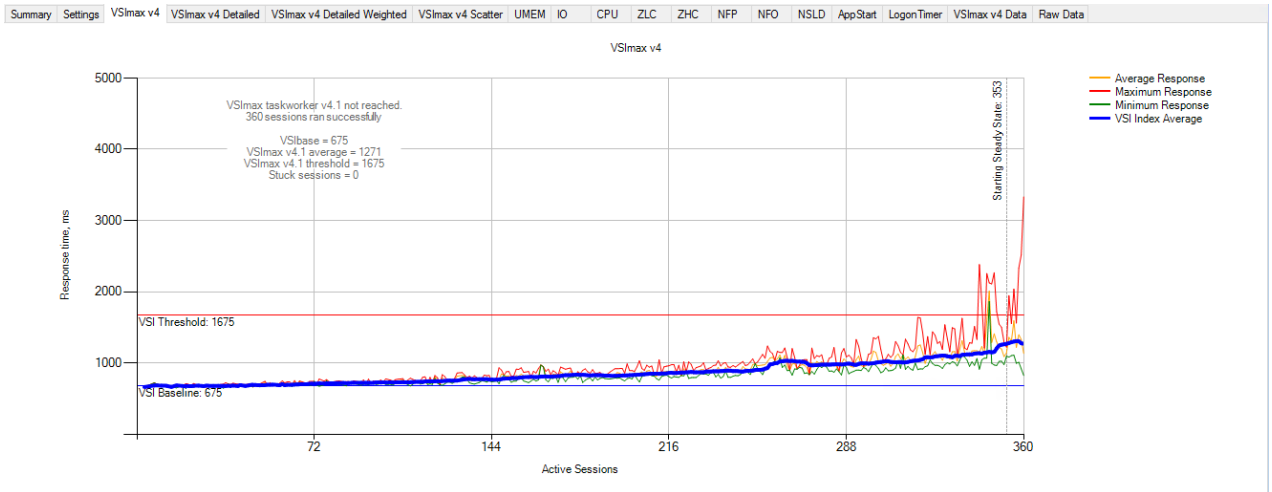


Figure 20. VMware ESXi host CPU Util%during testing

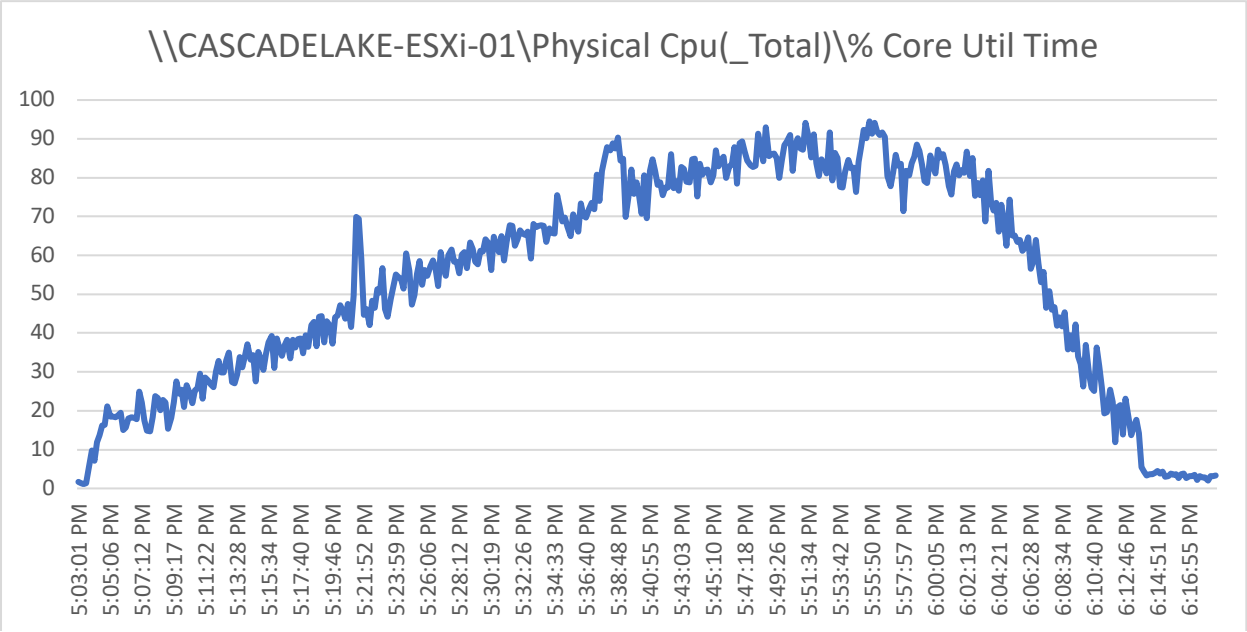
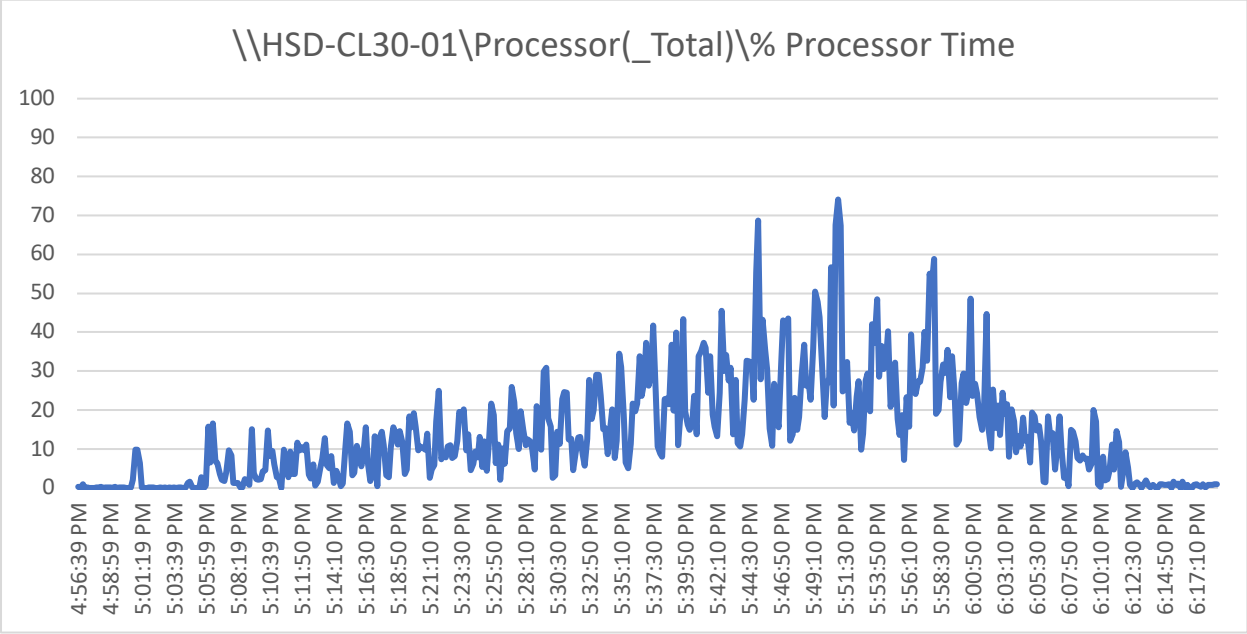


Figure 21. Processor time Util% on Microsoft Server 2016 virtual machine during testing



Knowledge Workers

Knowledge workers are individuals in an organization who use a large number of applications to perform their duties. Examples of knowledge workers are sales and marketing professionals, business development managers, healthcare clinicians, and project managers.

In some cases, these workers can be served by RDSH server sessions or published applications. In most cases, organizations provide a medium-capability Windows 10 virtual desktop to these users.

We tested both use cases using the Login VSI Knowledge Worker workload in benchmark mode. You can find additional information about Login VSI and all the workloads we tested for this document here.

In addition to the Login VSI test suite, we measured host utilization by gathering data from ESXtop. We also captured perfmon data from sample RDSH server virtual machines during the full server load tests.

Windows 10 and Citrix Virtual Apps and Desktops 7 single-server synopsis: Intel Xeon Scalable Gold processor 6230

The test results are summarized here and in Figures 22, 23, and 24.

- Operating system: Windows 10 64-bit (1607) with VMware optimizations
- 2 vCPUs; 4 GB of RAM
- Number of users: 225 users running Login VSI Knowledge Worker workload with Windows 10
- No VSImax; Login VSI baseline = 778 ms

Figure 22. Login VSI end-user experience summary

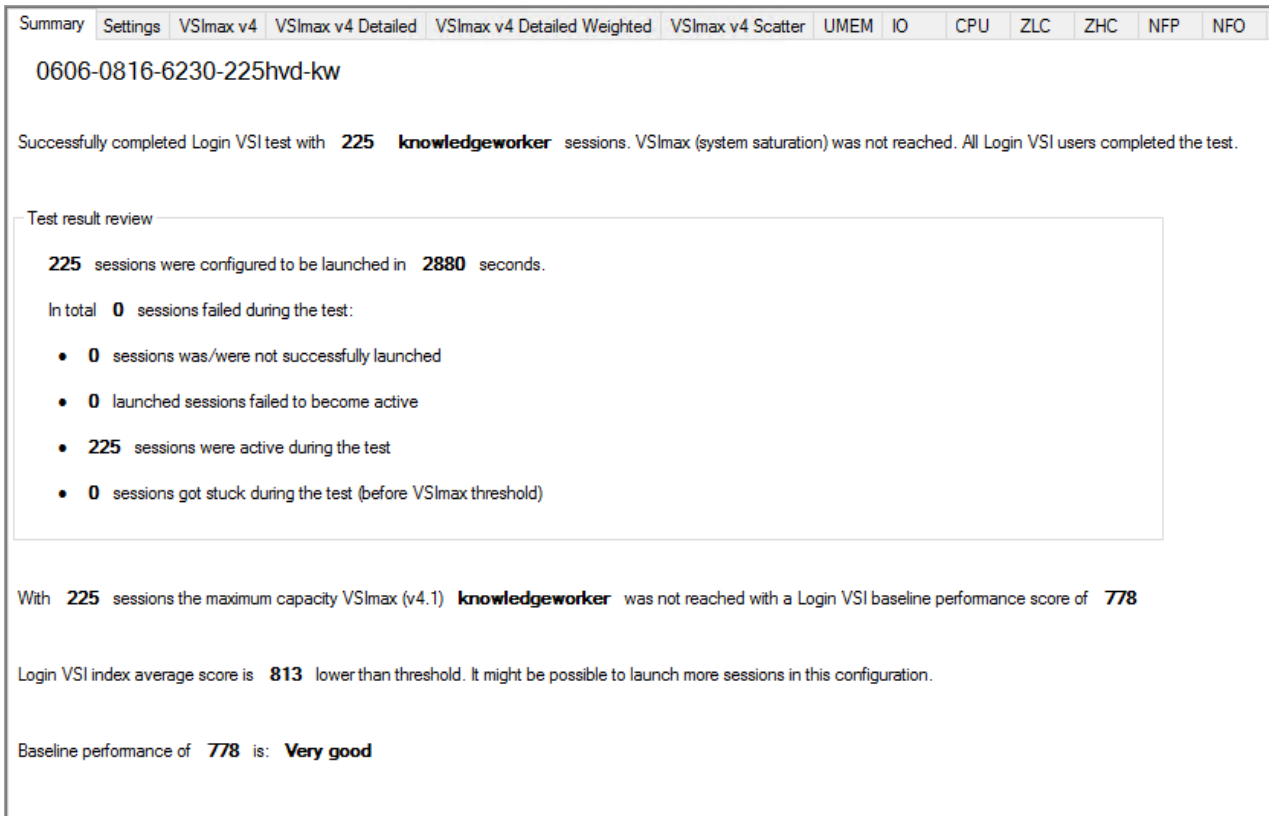


Figure 23. Login VSI end-user experience performance chart

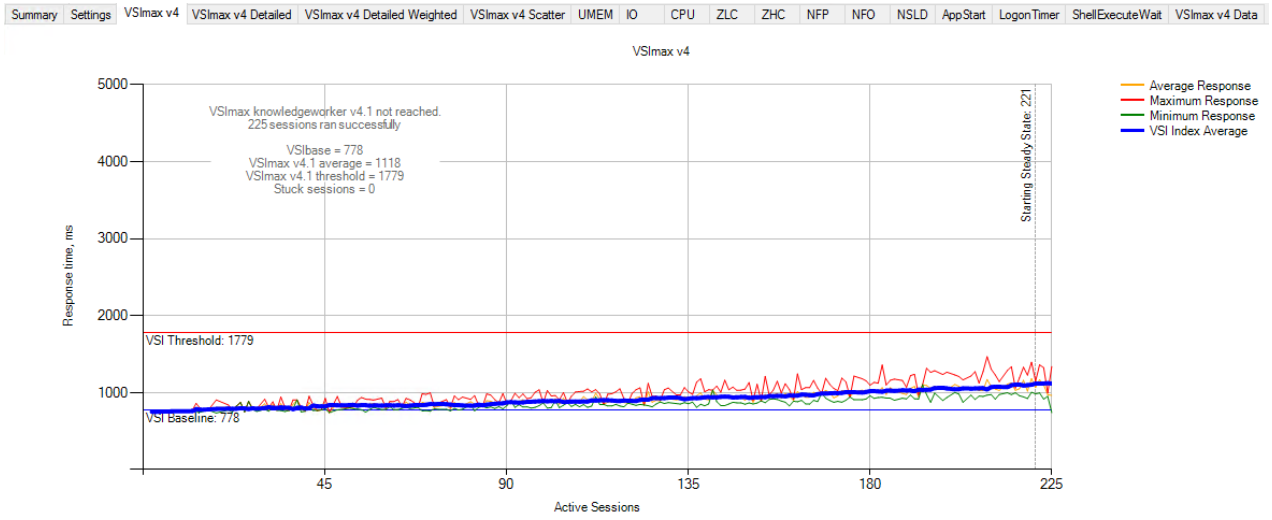
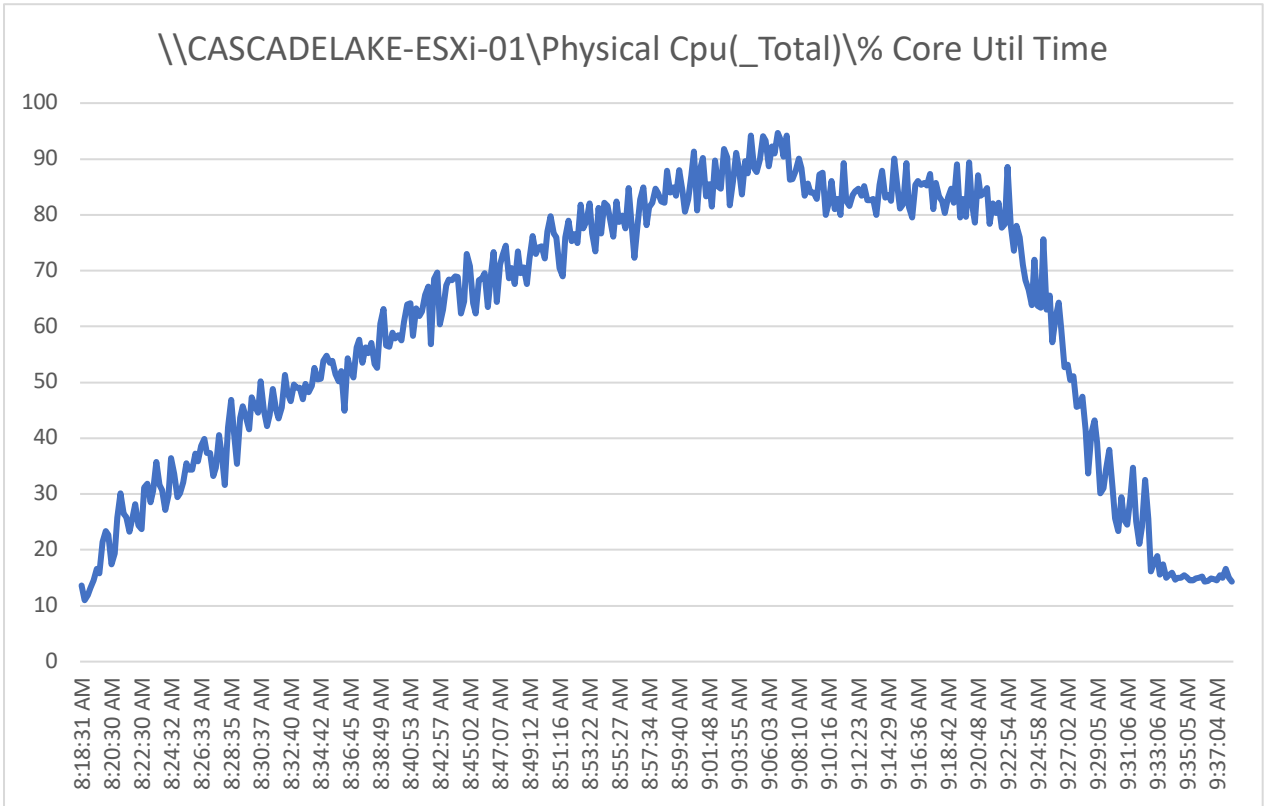


Figure 24. VMware ESXi host CPU Util% during testing



Windows Server 2016 and Citrix Virtual Apps and Desktops 7 RDSH synopsis: Intel Xeon Scalable Gold processor 6230

The test results are summarized here and in Figures 25, 26, 27, and 28.

- Operating system: Windows Server 2016 (1607) with Citrix optimizations
- 9 vCPUs; 24 GB of RAM per virtual machine
- 11 virtual machines per host
- Number of users: 330 users running Login VSI Knowledge Worker workload RDSH sessions
- No VSImax; Login VSI baseline = 644 ms

Figure 25. Login VSI end-user experience summary

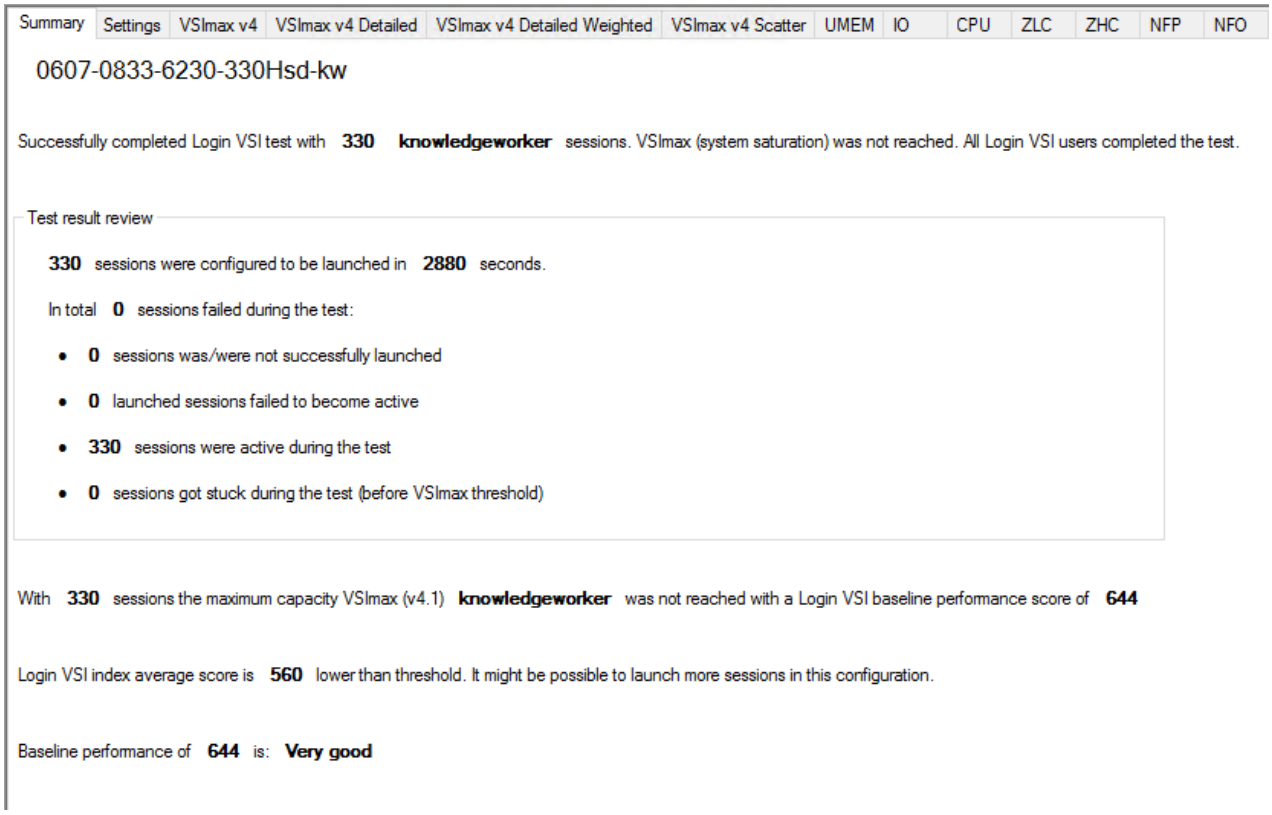


Figure 26. Login VSI end-user experience performance chart

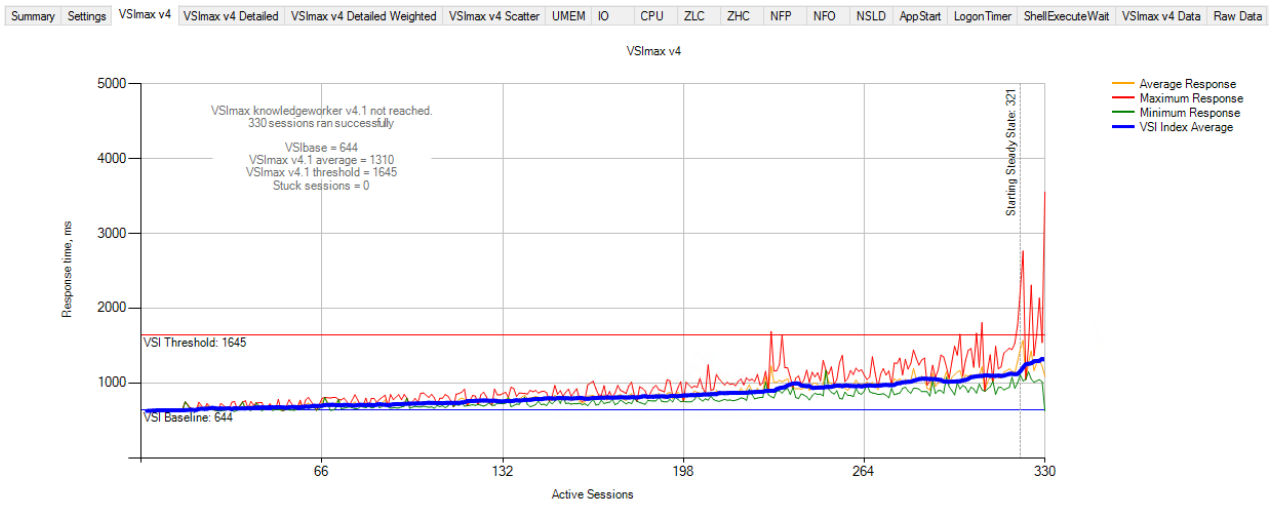


Figure 27. VMware ESXi host CPU Util% during testing

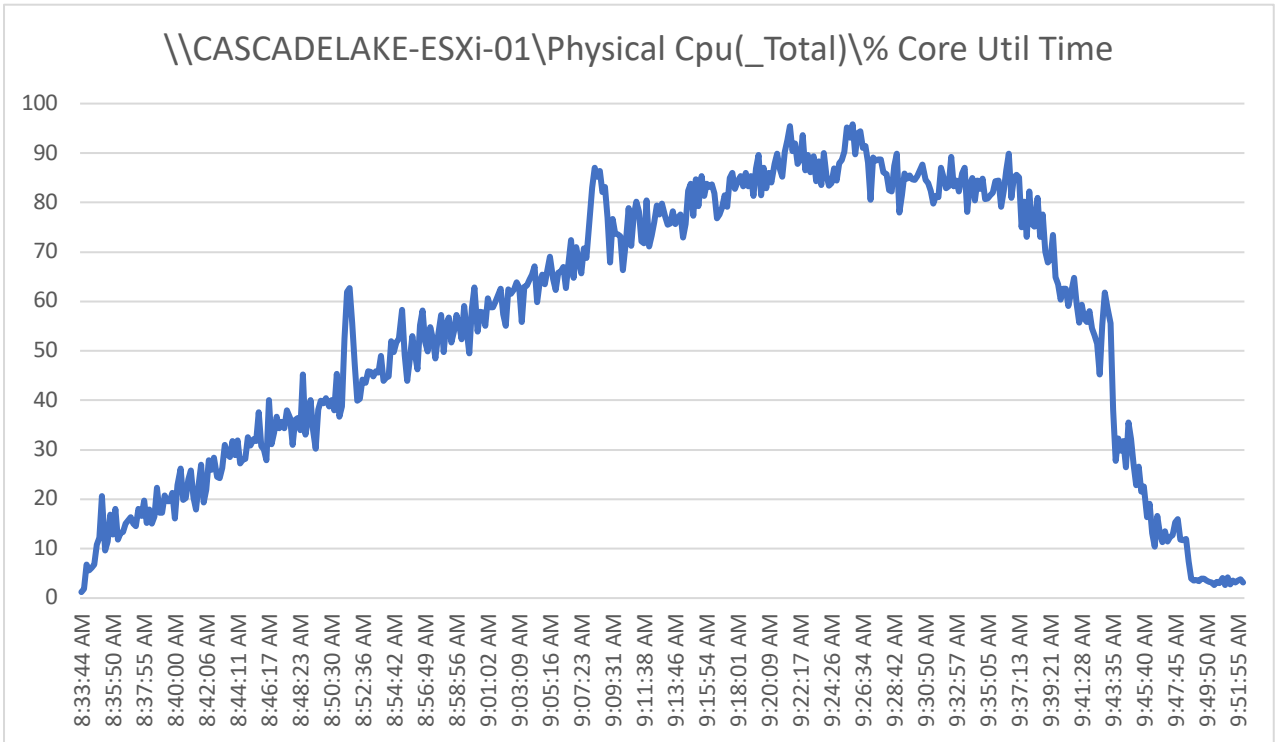
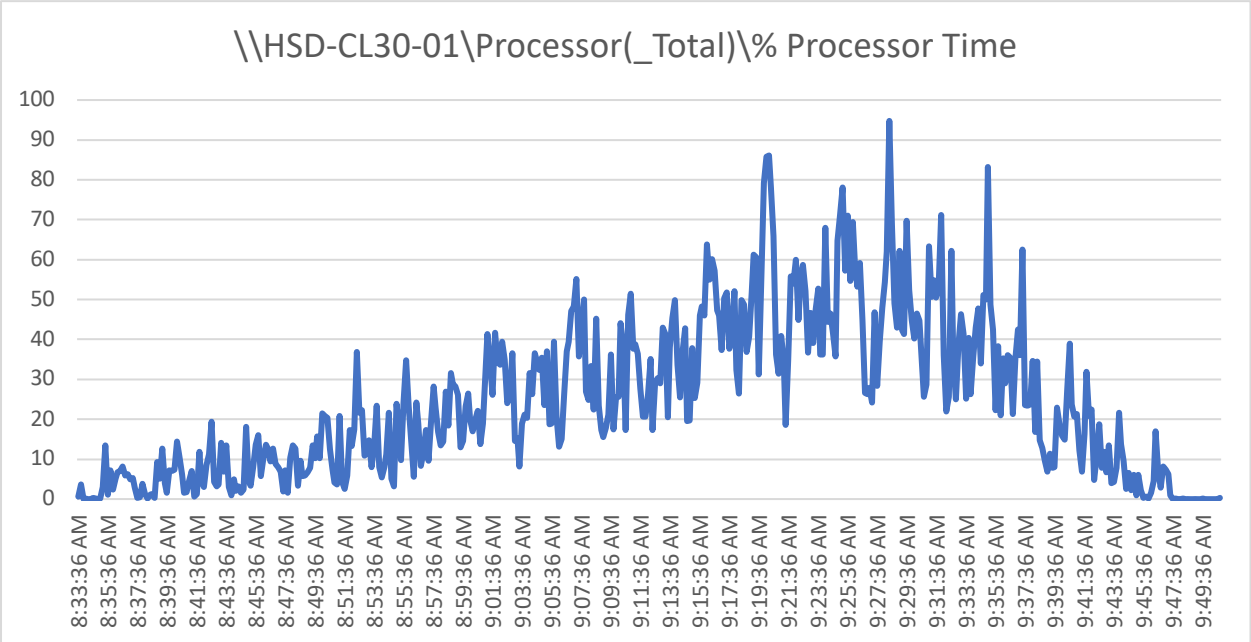


Figure 28. Processor time Util% on Microsoft Windows Server 2016 virtual machine during testing



Power Users

Power users are individuals in an organization who use a large number of installed and web applications to perform their duties. These users typically have many applications open concurrently and perform more complex operations than other workers. Examples of power users are business analysts, strategic and tactical planners, manufacturing planners, operations planners, and financial analysts.

These workers cannot typically be served by RDSH server sessions or published applications. In almost all cases, organizations provide a highly capable Windows 10 virtual desktop to these users.

For that reason, we tested only the Windows 10 use case using the Login VSI Power User workload in benchmark mode. You can find additional information about Login VSI and all the workloads we tested for this document here.

In addition to the Login VSI test suite, we measured host utilization by gathering data from ESXtop.

Windows 10 and Citrix Virtual Apps and Desktops 7 single-server synopsis: Intel Xeon Scalable Gold processor 6248

The test results are summarized here and in Figures 29, 30, and 31.

- Operating system: Windows 10 64-bit (1607) with VMware optimizations
- 4 vCPUs; 8 GB of RAM
- Number of users: 220 users running Login VSI Power User workload with Windows 10
- No VSImax; Login VSI baseline = 699 ms

Figure 29. Login VSI end-user experience summary

Summary | Settings | VSImax v4 | VSImax v4 Detailed | VSImax v4 Detailed Weighted | VSImax v4 Scatter | UMEM | IO | CPU | ZLC | ZHC | NFP | NFO

0612-1729-220HVDpower

Successfully completed Login VSI test with **220** **powerworker** sessions. VSImax (system saturation) was not reached. All Login VSI users completed the test.

Test result review

220 sessions were configured to be launched in **2880** seconds.

In total **0** sessions failed during the test:

- **0** sessions was/were not successfully launched
- **0** launched sessions failed to become active
- **220** sessions were active during the test
- **0** sessions got stuck during the test (before VSImax threshold)

With **220** sessions the maximum capacity VSImax (v4.1) **powerworker** was not reached with a Login VSI baseline performance score of **699**

Login VSI index average score is **792** lower than threshold. It might be possible to launch more sessions in this configuration.

Baseline performance of **699** is: **Very good**

Figure 30. Login VSI end-user experience performance chart

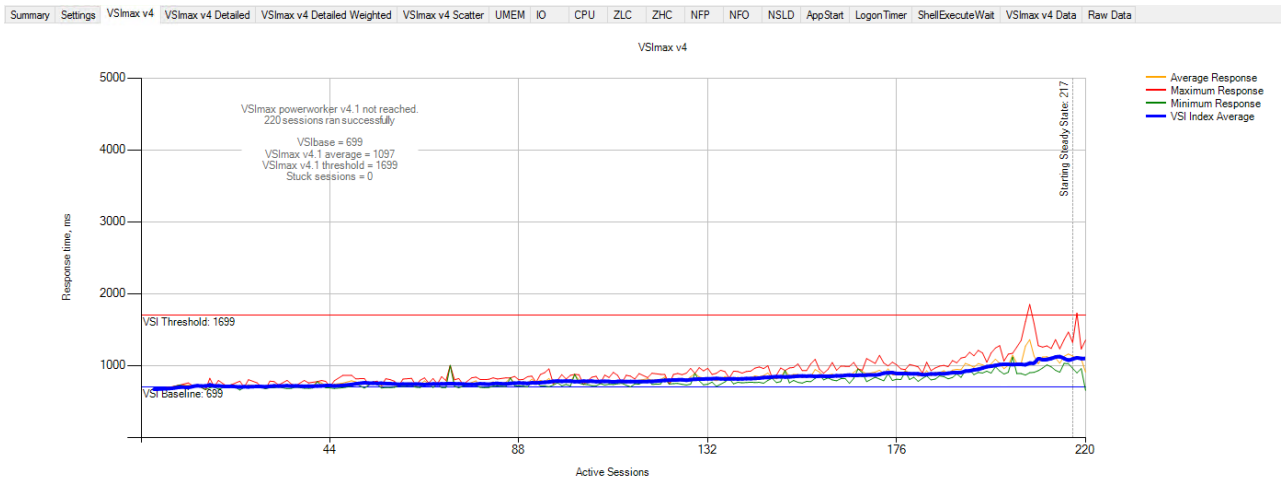
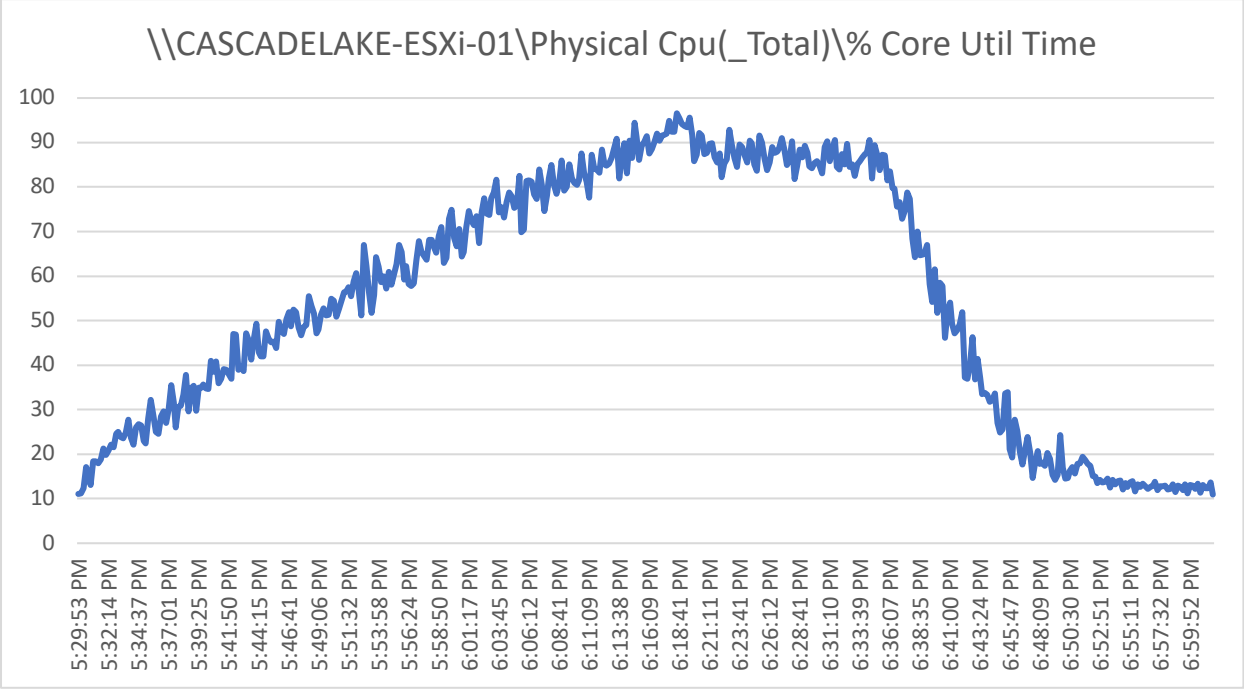


Figure 31. VMware ESX host CPU Util % during testing



Conclusion

By carefully evaluating key processors in the second-generation Intel Xeon Scalable processor families, we identified three processors that have optimal price-to-performance characteristics for the three main user types—task workers, knowledge workers, and power users—for Citrix Virtual Apps and Desktops 7 and Microsoft Windows 10 virtual desktops and RDSH hosted server sessions. The user type configuration details are detailed in Table 4.

Table 4. Configurations by user type

Microsoft Windows 10	Task Worker	Knowledge Worker	Power User
vCPU per VM	1	2	4
Memory per VM	2 GB	4 -8 GB	8 -16 GB

We identified the maximum recommended workload for each processor and user type pair for Windows 10 virtual machines and for RDSH sessions for task workers and knowledge workers. The maximum recommended workload is used to plan for maintenance and failure scenarios. During normal operations, fewer virtual machines would run on the clusters supporting your users. Normal operation densities are shown in Tables 5 and 6.

Because of the nature of power users' workloads, we focused solely on the Windows 10 virtual machines and identified the maximum recommended workload.

Tables 5 and 6 summarize our starting-point recommendations (not the maximum recommended workloads) for each user type and delivery method.

Table 5. Microsoft Windows 10 (Build 1607) and Citrix Virtual Apps and Desktops 7 virtual desktops

Processor PID x Qty	Cores	Memory (GB)	Task Worker	Knowledge Worker	Power User
UCS-CPU-I5220 x 2	36	768 (12 x 64 GB)	220 -250		
UCS-CPU-I6230 x 2	40	768 (12 x 64 GB)		95 -190	
UCS-CPU-I6248 x 2	40	1.5 TB (24 x 64 GB)			95 -190

Table 6. Microsoft Server 2016 (Build 14393.2485) and Citrix Virtual Apps and Desktops 7 RDSH sessions

Processor PID x Qty	Cores	Memory (GB)	Task Worker RDSH Sessions	Knowledge Worker RDSH Sessions	Power User RDSH Sessions
UCS-CPU-I5220 x 2	36	384 (12 x 32 GB)	220 -240		
UCS-CPU-I6230 x 2	40	384 (12 x 32 GB)		220 - 240	
UCS-CPU-I6248 x 2	40	384 (12 x 32 GB)			N/A

Each customer's environment and workloads are different. The recommended ranges shown in the tables here are starting points for your unique environment. They are not intended to be performance guarantees.

For graphics-intensive workloads and for enhanced-experience Windows 10 workloads, you can use GPUs with additional processors that are suited for that purpose.

For more information

For additional information, see the following:

- Cisco UCS C-Series Rack Servers and B-Series Blade Servers:
  - http://www.cisco.com/en/US/products/ps10265/
- Cisco HyperFlex™ hyperconverged servers:
  - https://www.cisco.com/c/en/us/products/hyperconverged-infrastructure/hyperflex-hx-series/index.html
- Citrix Virtual Apps and Desktops 7:
  - https://docs.citrix.com/en-us/citrix-virtual-apps-desktops.html
- VMware vSphere 6.7 Update 2:
  - https://docs.vmware.com/en/VMware-vSphere/6.7/rn/vsphere-esxi-67u2-release-notes.html
  - https://blogs.vmware.com/vsphere/2019/04/vcenter-server-6-7-update-2-whats-new.html
- Citrix Optimizer:
  - https://support.citrix.com/article/CTX224676
- Login VSI
  - https://www.loginvsi.com/
  - https://www.loginvsi.com/products/login-vsi

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-742527-00 7/19