Lenovo™

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series

Last update: 26 March 2024
Version 2.9

Provides a technical overview of Lenovo ThinkAgile HX Series appliances and certified nodes

---

Contains performance data and sizing recommendations

Shows variety of workloads and software stack that can be used in a Nutanix hyper-converged platform

---

Explains reliability and performance features of hyper-converged appliances

Chandrakandh Mouleeswaran

Dan Ionut Ghervase

Patrick Hartman

Vinay Kulkarni

Xiaotong Jiang

Markesha Parker

Cristian Ghetau

David West

LENOVO PRESS

Table of Contents

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

1 Introduction

The intended audience for this document is technical IT architects, system administrators, and managers who are interested in executing workloads on the Lenovo ThinkAgile HX Series appliances and certified nodes.

ThinkAgile HX Series provides a hyper-converged infrastructure. Hyper-converged means incorporating multiple components like compute and storage into a single entity through software. A hyper-converged infrastructure seamlessly pools compute and storage to deliver high performance for the virtual workloads and provides flexibility to combine the local storage using a distributed file system to eliminate shared storage such as SAN or NAS. These factors make the solution cost effective without compromising the performance.

Chapter 2 provides a technical overview of ThinkAgile HX Series and explains why the combination of Lenovo servers and Nutanix software provides best of breed system performance and reliability. Chapter 3 provides some deployment models.

Each of the subsequent chapters in the document describes a particular virtualized workload and provides recommendations on what appliance model to use and how to size the appliance to that workload. Some best practice recommendations are also listed. ThinkAgile HX Series appliances and certified nodes are not limited to just the workloads described in this reference architecture and can execute any virtualized workload on the supported hypervisors.

This Reference Architecture describes seven workloads:

- Citrix Virtual Apps and Desktops
- Microsoft SQL Server
- Red Hat OpenShift Container Platform
- VMware Horizon
- VMware vCloud Suite
- Veeam Backup with ThinkSystem DE/DM/DG
- Nutanix Cloud Platform
- Hybrid Backup with AWS
- Enterprise AI with NVIDIA

For SAP workloads on ThinkAgile HX systems see https://lenovopress.lenovo.com/lp1413.pdf

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

2 Technical overview of appliances

This chapter provides an overview of the ThinkAgile HX Series appliances and certified nodes including the associated software, systems management, and networking. The last section provides an overview of the performance and reliability features.

2.1 ThinkAgile HX series

Lenovo ThinkAgile HX Series appliances and certified nodes are designed to help you simplify IT infrastructure, reduce costs, and accelerate time to value. These hyper-converged appliances from Lenovo combine industry-leading hyper-convergence software from Nutanix with Lenovo enterprise platforms. Several common uses are:

- Enterprise workloads
- Private and hybrid clouds
- Remote office and branch office (ROBO)
- Server virtualization
- Virtual desktop infrastructure (VDI)
- Data science, Artificial Intelligence(AI) and Machine Learning(ML)
- Small-medium business (SMB) workloads

Starting with as few as three nodes to keep your acquisition costs down, the Lenovo ThinkAgile HX Series appliances and certified nodes are capable of immense scalability as your needs grow.

Lenovo ThinkAgile HX Series appliances and certified nodes are available in five families that can be tailored to your needs:

- Lenovo ThinkAgile HX1000 Series: optimized for ROBO environments
- Lenovo ThinkAgile HX2000 Series: optimized for SMB environments
- Lenovo ThinkAgile HX3000 Series: optimized for compute-heavy environments
- Lenovo ThinkAgile HX5000 Series: optimized for storage-heavy workloads
- Lenovo ThinkAgile HX7000 Series: optimized for high-performance workloads

For Lenovo ThinkAgile HX 1U Appliances & Certified Nodes are 2-socket 1U systems, see
https://lenovopress.lenovo.com/lp1521-thinkagile-hx1320-hx1321-hx2320-hx2321-hx3320-hx3321-1u
For Lenovo ThinkAgile HX 2U Appliances & Certified Nodes are 2-socket 2U systems, see
https://lenovopress.lenovo.com/lp1522-thinkagile-hx3520g-hx3521g-hx5520-hx5521-hx7520-hx7521-2u

Table 1 shows the similarities and differences between ThinkAgile HX Series appliances and certified nodes.

Table 1: Comparison of ThinkAgile HX Series appliances and certified nodes

Feature	HX Series Appliances	HX Series certified nodes
Validated and integrated hardware and firmware	Yes	Yes
Certified and preloaded with Nutanix software	Yes	Yes
Includes Nutanix licenses	Yes	No

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

ThinkAgile Advantage Single Point of Support for quick 24/7 problem reporting and resolution	Yes	Yes
Includes deployment services	Optional	Optional
Supports ThinkAgile HX2000 Series	Yes	No

For more information about the system specifications and supported configurations, refer to the product guides for the Lenovo ThinkAgile HX Series appliances and certified nodes based on the Intel Xeon Scalable processors and AMD EPYC processors.

For Integrated systems see   https://lenovopress.lenovo.com/datasheet/ds0019-lenovo-thinkagile-hx-series

For certified nodes see https://lenovopress.lenovo.com/datasheet/ds0046-lenovo-thinkagile-hx-certified-nodes

Table *2* shows ThinkAgile HX V3 systems with Intel Xeon 4th Generation Scalable Processors.

Table 2: ThinkAgile HX V3 Models with Intel Xeon 4th Gen Scalable Processors

Model	HX630 V3	HX650 V3	HX650 V3
Configuration	All Flash	All Flash	Hybrid Storage
Form Factor	1U 2S	2U 2S	2U 2S
Memory	TruDDR5	TruDDR5	TruDDR5
	32 DIMMs	32 DIMMs	32 DIMMs
	(8 TB Max)	(8 TB Max)	(8 TB Max)
GPU	2xSW 75W	7xSW 4xDW	
Drives NVMe/SAS/SATA	10x2.5" + 2x2.5"	16x3.5" + 8x2.5"	12x3.5" + 4x2.5"
PCIe 5.0	2 Slots 1/10/25/100 GbE	7 Slots 1/10/25/100 GbE	7 Slots 1/10/25 GbE
OCP 3.0	1x1Gb, 10Gb, 25Gb	1x1Gb, 10Gb, 25Gb	1x1Gb, 10Gb, 25Gb
HX Integrated System	7D6MCTO1WW	7D6NCTO1WW	7D6NCTO2WW
HX Certified Node	7D6MCTO3WW	7D6NCTO3WW	7D6NCTO4WW

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Max Possible Capacity	24x7.68 TB	24x7.68 TB	16x7.68 TB

Table 3 shows ThinkAgile HX V3 systems with AMD EPYC 4th Generation processors.

Table 3: ThinkAgile HX V3 Models with 4th Gen AMD EPYC processors

Model	**HX645 V3**	**HX665 V3**	**HX665 V3 Storage**
Configuration	All Flash	All FlashHybrid	All FlashHybrid
Form Factor	1U 2S	2U 2S	2U 2S
Memory	6 TB Maximum	6 TB Maximum	6 TB Maximum
GPU	Up to 2x single-width GPUs	No	No
Drives NVMe/SAS/SATA	12 Maximum4 NVMe	40 Maximum8 NVMe	16 Maximum
PCIe 5.0	3 Slots	10 Slots	10 Slots
OCP 3.0	1x 1Gb, 10Gb, 25Gb	1x 1Gb, 10Gb, 25Gb	1x 1Gb, 10Gb, 25Gb
Hypervisor	Nutanix AHV, VMware ESXi	Nutanix AHV, VMware ESXi	 .
HX Integrated System	7D9MCTO1WW	7D9NCTO1WW	7D9NCTO2WW
HX Certified Node	7D9MCTO2WW	7D9NCTO3WW	7D9NCTO4WW

2.2 Software components

This section gives an overview of the software components used in the solution.

2.2.1 Hypervisor

The ThinkAgile HX Series appliances and certified nodes (generally) support the following hypervisors:

- Nutanix Acropolis Hypervisor based on KVM (AHV)
- VMware ESXi 6.7
- VMware ESXi 7.0

The HX1520-R, HX5520-C, HX7820, and all SAP HANA models support only the following hypervisor: The HX Series appliances come standard with the hypervisor preloaded in the factory. This software is optional for the ThinkAgile HX Series certified nodes.

2.2.2 Lenovo XClarity Administrator

Lenovo XClarity Administrator is a centralized systems management solution that helps administrators deliver infrastructure faster. This solution integrates easily with Lenovo servers, ThinkAgile HX Series appliances and certified nodes, and Flex System, providing automated agent-less discovery, monitoring, firmware updates, and configuration management.

Lenovo XClarity Pro goes one step further and provides entitlement to additional functions such as XClarity Integrators for Microsoft System Center and VMware vCenter, XClarity Administrator Configuration Patterns and Service and Support.

Lenovo XClarity Administrator is an optional software component and can be used to manage firmware upgrades outside of the Nutanix Prism web console. Note that XClarity should not be used to install hypervisors and Nutanix Foundation should be used instead.

Lenovo XClarity Administrator is provided as a virtual appliance that can be quickly imported into a virtualized environment. XClarity can either be installed on a separate server or a server within a Nutanix cluster providing that the hardware management network with the server IMMs is routable from the server hosting the XClarity VM.

Figure 1 shows the Lenovo XClarity administrator interface.
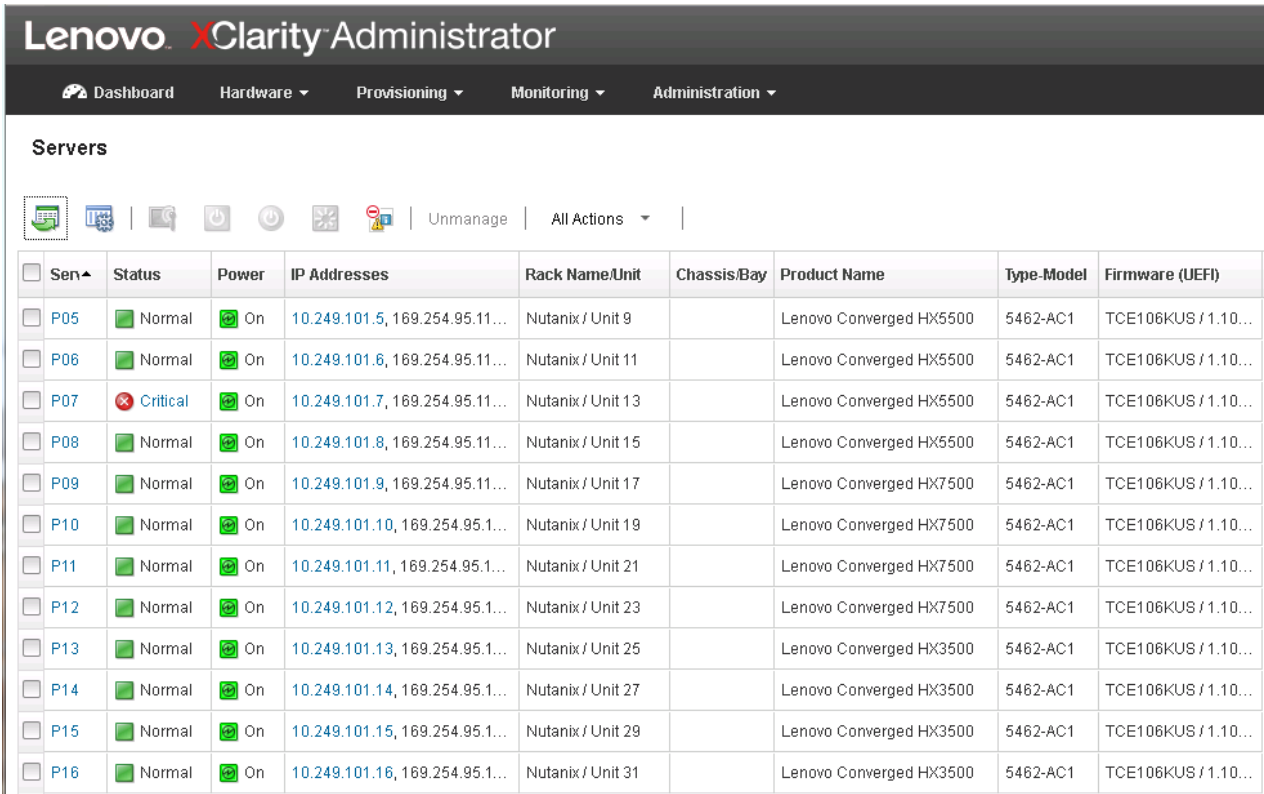
Figure 1: XClarity Administrator interface

2.2.3 Nutanix Prism

Nutanix Prism gives administrators a simple and elegant way to manage virtual environments. Powered by advanced data analytics and heuristics, Prism simplifies and streamlines common workflows within a data center.

Nutanix Prism is a part of the Nutanix software preloaded on the appliances and offers the following features:

- Single point of control

    o Accelerates enterprise-wide deployment

    o Manages capacity centrally

    o Adds nodes in minutes

    o Supports non-disruptive software upgrades with zero downtime

    o Integrates with REST APIs and PowerShell

- Monitoring and alerting

    o Tracks infrastructure utilization (storage, processor, memory)

    o Centrally monitors multiple clusters across multiple sites

    o Monitors per virtual machine (VM) performance and resource usage

    o Checks system health

    o Generates alerts and notifications

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

- Integrated data protection

    o Offers customizable RPO/RTO and retention policies

    o Supports configurable per-VM replication (1:1, 1:many and many:1)

    o Provides efficient VM recovery

    o Deploys affordable data recovery (DR) and backup to the cloud

- Diagnostics and troubleshooting

    o Provides time-based historical views of VM activity

    o Performs proactive alert analysis

    o Correlates alerts and events to quickly diagnose issues

    o Generates actionable alerts and reduces resolution times

    o Analyzes trending patterns for accurate capacity planning

2.2.4 Nutanix Foundation

Nutanix Foundation is a separate utility that you use to orchestrate the installation of hypervisors and Nutanix software on one or more nodes. The maximum number of nodes that can be deployed at one time is 20.

Foundation is available both as a stand-alone VM and also integrated into the CVM. Because CVM is pre-installed in the factory, the CVM integration of Foundation simplifies the deployment and cluster creation of new servers delivered from the factory.

The dual M.2 boot drives must be configured as a RAID 1 mirrored array for installation to be successful.

2.2.5 Nutanix Controller VM

The Nutanix Controller VM (CVM) is the key to hyper-converged capability and each node in a cluster has its own instance. Figure 2 shows the main components of the CVM.

Stargate - Data I/O Manager
- Its responsible for all data management and I/O operations.
- This service runs on each node in the cluster to serve localized I/O.

Cassandra - Distributed Metadata Store
- It stored and manages the cluster metadata.
- Cassandra runs on each node in the cluster.

Curator – Map reduce cluster management and cleanup
- Its responsible for managing and distributing tasks like disk balancing etc. across the cluster.
- It runs on each node and controlled by an elected Curator Master.

Zookeeper – Cluster Configuration Manager
- It stores all the cluster configuration e.g. Hosts, state, IP addresses etc.
- This service runs on three nodes in the cluster and one among these nodes elected as a leader.

Prism – User Interface
- Its an user interface to configure and monitor Nutanix cluster.
- Its runs on each node in the cluster.

Figure 2: Controller VM components

The CVM works as interface between the storage and hypervisor to manage all I/O operations for the hypervisor and user VMs running on the nodes as shown in Figure 3.
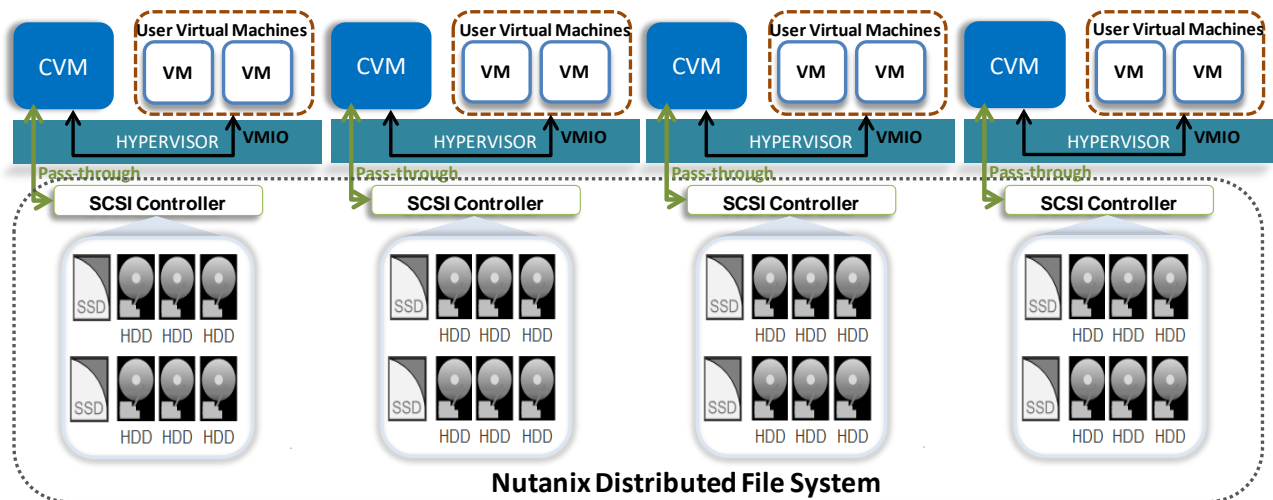
Figure 3: CVM interaction with Hypervisor and User VMs

CVM virtualizes all the local storage attached to each node in a cluster and presents it as centralized storage array using Nutanix Distributed File System (NDFS). All I/O operations are handled locally to provide the highest performance. See section 2.5 for more details on the performance features of NDFS.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

2.3 Data network components

The data network is the fabric that carries all inter-node storage I/O traffic for the shared Lenovo HX distributed file system, in addition to the user data traffic via the virtual Network Interface Cards (NICs) exposed through the hypervisor to the virtual machines. The

ThinkAgile HX supports 10GbE/25GbE OCP and PCIe ethernet network adapters from Broadcom and Mellanox, but mixing vendors is not supported. HX configurations only support network adapters from one vendor. For example, if you select a Broadcom OCP adapter, you cannot select a Mellanox PCIe network adapter. All HX systems supports one OCP adapter and 0-5 Broadcom PCIe adapter and 0-7 Mellanox PCIe adapters.

The hypervisors are configured by the Nutanix software so that the fastest network ports on the appliance are pooled for the data network. The hypervisor VM management network should use the same network. Because all of the network ports are pooled, each appliance only needs two network IP addresses; one for the hypervisor and one for the Nutanix CVM. These IP addresses should be all on the same subnet.

All storage I/O for virtual machines (VMs) running on a HX Series appliance node is handled by the hypervisor on a dedicated private network. The I/O request is handled by the hypervisor, which then forwards the request to the private IP on the local controller VM (CVM). The CVM then performs the remote data replication with other nodes in the cluster using its external IP address. In most cases, read request traffic is served locally and does not enter the data network. This means that the only traffic in the public data network is remote replication traffic and VM network I/O (i.e. user data). In some cases, the CVM will forward requests to other CVMs in the cluster, such as if a CVM is down or data is remote. Also, cluster-wide tasks, such as disk balancing, temporarily generate I/O traffic on the data network.

For more information on the network architecture see nutanixbible.com.

2.3.1 Data network switches

ThinkAgile HX series are interoperable with many TOR and aggregation switches available in the industry. A resilient network design is important to ensure connectivity between HX CVMs, for virtual machine traffic, and for management functions. The basic design overview of the switches for interconnecting HX Series nodes is shown in Figure 4

For the basic interconnect, two TOR switches are placed in an LAG configuration (called VLAG, similar to Cisco Nexus vPC) which enables this switch pair to act as a single logical switch over which single link aggregations can be formed between ports on both hardware switches. The HX Series are connected redundantly to each of the VLAG peer switches and rely on vSwitch features to spread the traffic over the two VLAG peers. The VLAG presents a flexible basis for interconnecting to the uplink/core network, ensures the active usage of all available links, and provides high availability in case of a switch failure or a required maintenance outage.
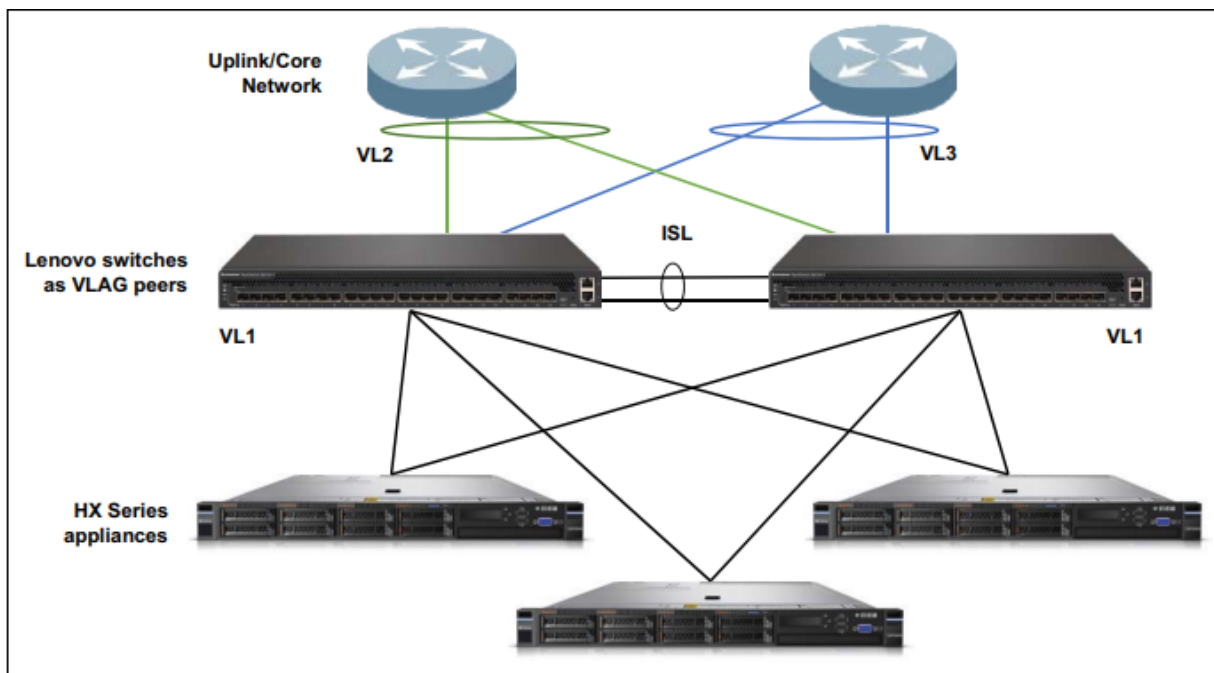
Figure 4: Basic Network Design for HX Cluster

2.3.2 VLANs

It is a networking best practice to use VLANs to logically separate different kinds of network traffic. The following standard VLANs are recommended:

- Management   Used for all management traffic for the hypervisor

- Storage network   Used for NDFS storage traffic

The following ESXi specific VLANs are recommended:

- vSphere vMotion   Used to move VMs from one server to another.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

- **Fault Tolerance**   Used to support the fault tolerance (FT) feature of vSphere.

In addition, each workload application might require one or more VLANs for its logical networks. For larger networks with many workloads, it is easy to run out of unique VLANs. In this case, VXLANs could be used.

The procedure for configuring VLANs for HX Series appliances is outside of the scope of this document.

2.3.3 Redundancy

It is recommended that two top of rack (TOR) switches are used for redundancy in the data network. It is recommended to use two dual-port 10/25Gbps network adapters for workloads that require high throughput on the network or scale-out cluster deployments. This will effectively provide two redundant, bonded links per host for 20Gbps of bandwidth per logical link. Note that by default, the bonding configuration for ThinkAgile HX Series is active/passive, but this can be changed to active/active with the proper configuration on the hypervisor host and switch side.

In order to support the logical pairing of the network adapter ports and to provide automatic failover of the switches, it is recommended to use virtual link aggregation groups (VLAGs). When VLAG is enabled over the inter-switch link (ISL) trunk, it enables logical grouping of these switches. When one of the switches is lost, or the uplink from the host to the switch is lost, the connectivity is automatically maintained over the other switch.

2.4 Hardware management network components

The hardware management network is used for out-of-band access to ThinkAgile HX Series appliances and certified nodes via the optional Lenovo XClarity Administrator. It may also be needed to re-image an appliance. All systems management is handled in-band via Intelligent Platform Management Interface (IPMI) commands. The dedicated Integrated Management Module (IMM) port on all of the Lenovo ThinkAgile HX series appliances and certified nodes needs to be connected to a 1GbE TOR switch.

2.5 Reliability and performance features

Reliability and excellent performance are important for any workload but particularly for hyper-converged infrastructures like ThinkAgile HX Series. These requirements are met through the following design features of Nutanix software combined with Lenovo Servers.

Hardware reliability

Lenovo uses the highest quality hardware components combined with firmware that is thoroughly tested. As a consequence Lenovo servers have been rated #1 in hardware reliability for the last 3 years. This is important as it lowers the frequency of a server failure which in turn lowers OPEX.

A HX appliance has redundant hardware components by including two power supplies, multiple chassis fans, two Intel CPUs, multiple memory DIMMs, multiple SSDs and HDDs, and optionally up to two dual-port network interface cards.

Hardware performance

The HX Series appliances have been carefully designed for performance. In addition to all of the usual attributes like processors and memory, the 24 drive HX7520 uses three HBA controllers instead of the one. As a consequence the latency is halved for some workloads that heavily utilize the cold tier. This allows a higher throughput and improved transaction rates.

Distributed file system

The Nutanix Distributed file system (NDFS) is an intelligent file system which virtualizes the local attached storage (SSD/HDD) on all the nodes in a cluster and presents it as single storage entity to cluster. Figure 5 shows the high level structure of NDFS:

Storage Pool
- Group of physical devices for the cluster.
- It can span multiple nodes and is expanded as the cluster scales

Container
- It's a logical segmentation of storage pool.
- It contains virtual machines or files (vDisks).
- Its typically have 1:1 mapping with a datastore (in case of NFS/SMB).

vDisk
- It's a file above 512KB size on NDFS including .vmdks and virtual machine hard disks.

Figure 5: Nutanix Distributed File System

Data protection via replication

The Nutanix platform replication factor (RF) and checksum is used to ensure data redundancy and accessibility in the event of a node or disk failure or corruption. It uses an OpLog which acts as a staging area for incoming writes on low latency SSDs which are then replicated to the OpLogs for one or two other Controller VMs before acknowledging a successful write. This approach ensures that data available in at least

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

two to three different locations and is fault tolerant. While the data is being written a checksum is calculated and stored as part of its metadata.

In the case of a drive or node failure, that data is replicated out to more nodes to maintain the replication factor. A checksum is computed every time the data is read to ensure the data validity. If the checksum and data mismatch, then the data replica is read to replace the invalid copy.

Performance with data tiering

Nutanix uses a disk tiering concept in which disk resources (SSD and HDD) are pooled together to form a cluster wide storage tier. This tier can be accessed by any node within the cluster for data placement and can leverage the full tier capacity. The following data tiering functions are provided:

- The SSD on a local node always has the highest tier priority for write I/O.
- If the local node's SSD is full then the other SSDs in the cluster are used for I/O.
- The NDFS Information Lifecycle Management (ILM) component migrates cold data from the local SSD to HDD to free up SSD space. It also moves heavily accessed data to the local SSD to provide high performance.

Performance by data locality

Data locality is a crucial factor for cluster and VM performance. In order to minimize latency the CVM will work to ensure that all I/O happens locally. This ensures optimal performance and provides very low latencies and high data transfer speeds that cannot be achieved easily with shared storage arrays, even if all-flash.

The following occurs in case of a VM migration or high availability event that moves a VM from Node-A to Node-B:

- The VM's data is provided by the CVM running on Node-B.
- All write I/O requests occur locally i.e. to the local storage of Node-B.
- When a request comes for reading old data, the I/O request is forwarded by Node-B to Node-A. NDFS detects that the I/O request originated from different node and migrates the data locally in the background i.e. from Node-A to Node-B so that all subsequent read I/O operations are served locally. This approach (migration only on a read) helps to avoid network flooding.

Performance of snapshots and clones

NDFS provides support for offloaded snapshots and clones using a redirect-on-write algorithm. When a snapshot or clone is created, the base vDisk is marked as read only and another vDisk is created with read/write permissions as shown in Figure 6 and Figure 7 below.

At this point both vDisks have the same block map - a metadata mapping of the vDisk to its corresponding extents. This approach reduces the overhead of creating snapshots and allows snapshots to be taken very quickly with little performance impact.
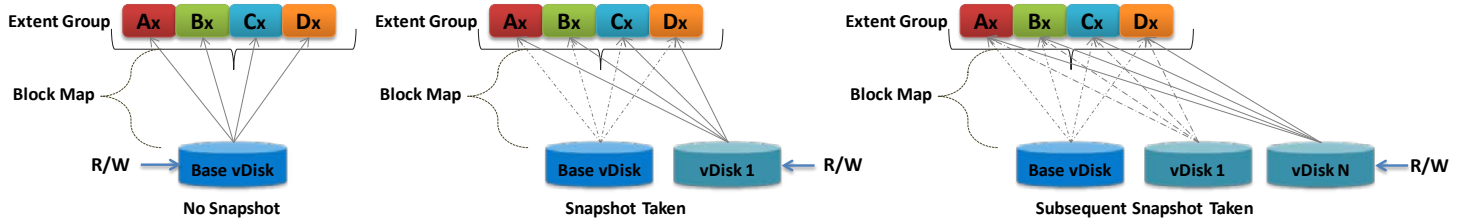
Figure 6: VM snapshots

When a VM is cloned the current block map is locked and then clones are created. These updates are metadata only so again no actual I/O takes place. The logic applies for clones of clones as well where a previously cloned VM acts as a base vDisk. All the clones inherit the prior block map and any new writes take place on the individual block maps.
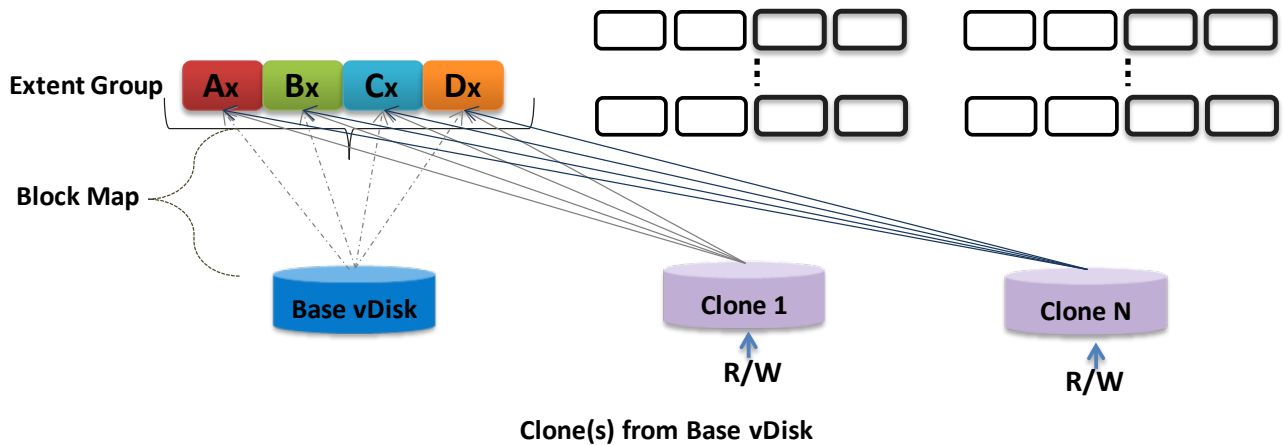

Clone(s) from Base vDisk

Figure 7: VM clones

Storage reduction via De-duplication and Compression

The Nutanix elastic de-duplication engine increases the effective capacity of a disk, as well as the RAM and cache of the system by removing duplicate data. It's an intelligent technology which performs following actions to increase storage efficiency:

- Sequential streams of data fingerprinted at 4K granularity
- Single instance of the shared VM data is loaded into the cache upon read
- Each node in a cluster performs its own fingerprinting and deduplication

The Nutanix capacity optimization engine is responsible for performing data transformations and compression to achieve data optimization. NDFS provides following compression methods:

- In-line compression sequential streams of data or large I/O sizes are compressed in memory before written to the disk
- Post-process compression whereby data is written in an uncompressed state and the curator framework is used to compress the data in a cluster wide manner

The Nutanix capacity optimization engine uses the Google snappy compression library to deliver good compression ratios with minimal compute overhead and very fast compression or decompression rates.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Elimination of "split-brain" errors

In a distributed system it is possible for one participant to become disconnected which will cause differences in the stored data. NDFS uses the proven "Paxos" algorithm to eliminate these "split-brain" issues by reaching a consensus (quorum) among the participants in a distributed system before the writes are made.

Drive reliability via active monitoring

The CVM actively monitors the performance of every drive in a node. The deterioration of a drive's performance may indicate that the drive is about to fail. The CVM proactively moves data off the drive before it fails and marks the drive offline and in need to replacement. The idea is to avoid the expensive data transfers to maintain data redundancy and possible loss of data.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

3 Deployment models

This chapter provides recommended deployment models for different examples of using a HX Series cluster.

3.1 SMB deployment model

There are two specific models of ThinkAgile HX Series appliances targeted for the small-medium business (SMB) environment. These are the HX2320-E in a 1U form factor and HX2720-E in a 2U4N form factor.

As described in the product guide for the HX2000 Series appliances, the cluster size is limited to 4 nodes. This example deployment for a SMB customer with low performance requirements includes 4 HX2720-E nodes in a 2U chassis and a 1GbE switch. Figure 8 shows the front view.

Figure 8: Front view of SMB Deployment

Figure 9 shows the rear view and the cabling for both the data (blue, red) and hardware management (green) networks into the same switch. It is also typical in these environments that the data network and management network IP addresses are in the same subnet.
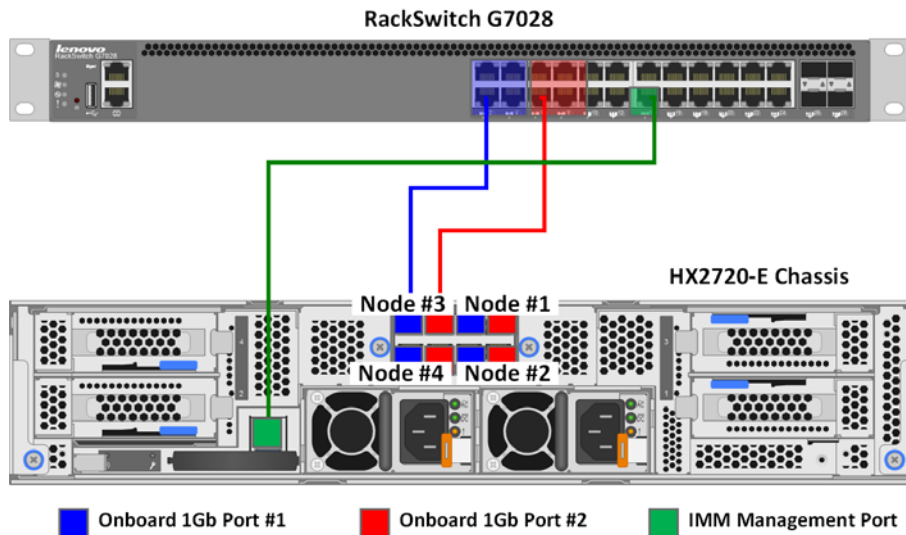

Figure 9: Rear view of SMB Deployment

3.2 Rack-scale deployment models

This section gives some example deployment models.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

3.2.1 VDI deployment model

This deployment model applies equally for either Citrix Virtual Apps and Desktops(see chapter 4) or VMware Horizon (see chapter 7). In this example the requirement is to support 5000 stateless virtual desktops. Each Windows 10 desktop virtual machine requires 4GB of RAM and 2-3 vCPU (similar to Knowledgeworker profile).

In order to promote reusability, the management nodes are configured the same as the compute nodes running the virtual desktops. The ThinkAgile HX650 V3 Integrated System are used and configured as follows:

- Intel Xeon Gold 6454S 32C 270W 2.2GHz Processor
- 16 x 64 GB ThinkSystem 64GB TruDDR5 4800MHz
- 4 x ThinkSystem 2.5" S4620 3.84TB Mixed Use SATA 6Gb HS SSD

Using the sizing tables presented in chapter 4 and chapter 7, it is recommended to have on average 226 virtual desktops per node with 75% CPU utilization per node. This translates to twenty-two HX650 V3 nodes. The recommended design is to split the cluster into multiple clusters of size 8-12 nodes for easy management and deployment. Each cluster can be sized with N+1 redundancy to support failover scenarios. For 5000 desktops, two clusters with 12 nodes are sufficient for knowledge worker and 24 nodes are needed in total. In addition three nodes are needed for the VDI management VMs and to provide adequate failover characteristics. A fourth node could be used as a quick deploy spare.

Figure 10 shows the deployment for 5000 VDI users with 28 x   HX650 V3 nodes(8 chassis) and the TOR switches. Each of the compute nodes is numbered C1 to C24 and each of the management nodes is numbered M1 to M3.
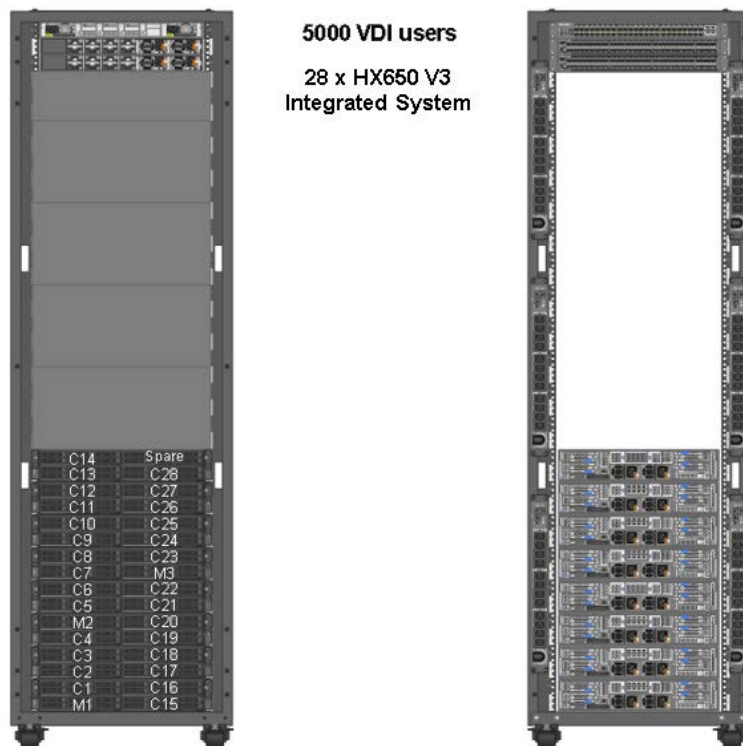

Figure 10: Example deployment model for 5000 VDI users

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

3.2.2 Microsoft SQL Server deployment model

In this example deployment model, the requirement is to support 24 generic applications and 8 analytic applications that use Microsoft SQL Server for the database. The solution uses ThinkAgile HX Flash servers powered by 4th Gen Intel Xeon Scalable Processors and 4th Gen AMD EPYC processors.

The 8 analytic workloads are supported using 8 x HX650 V3 appliances hosting 16 Microsoft SQL server VMs. An additional 4 x HX665 V3 appliances are used for sixteen Microsoft SQL Server VMs for 24 generic workloads and   uses Microsoft AlwaysOn availability groups (AAGs) for database redundancy. The VMs for the web front-ends and backend microservices are redundantly deployed on 4 x HX650 V3 appliances.

The HX650 V3 Integrated System is configured as follows:

- 2x Intel Xeon Platinum 8450H 28C 250W 2.0GHz Processor
- 24 x ThinkSystem 64GB TruDDR5 4800MHz
- 12 x ThinkSystem S4620 3.84TB Mixed Use SATA 6Gb HS SSD

The HX665 V3 Integrated System is configured as follows:

- ThinkSystem AMD EPYC 9354 32C 280W 3.25GHz Processor
- 16 x 64 GB ThinkSystem 64GB TruDDR5 4800MHz
- 6 x ThinkSystem P5620 3.2TB Mixed Use NVMe PCIe 4.0 x4 HS SSD

Figure 11 shows the deployment of the 12 x HX650 V3 appliances, 4 x HX665 V3 appliances, and the TOR switches.
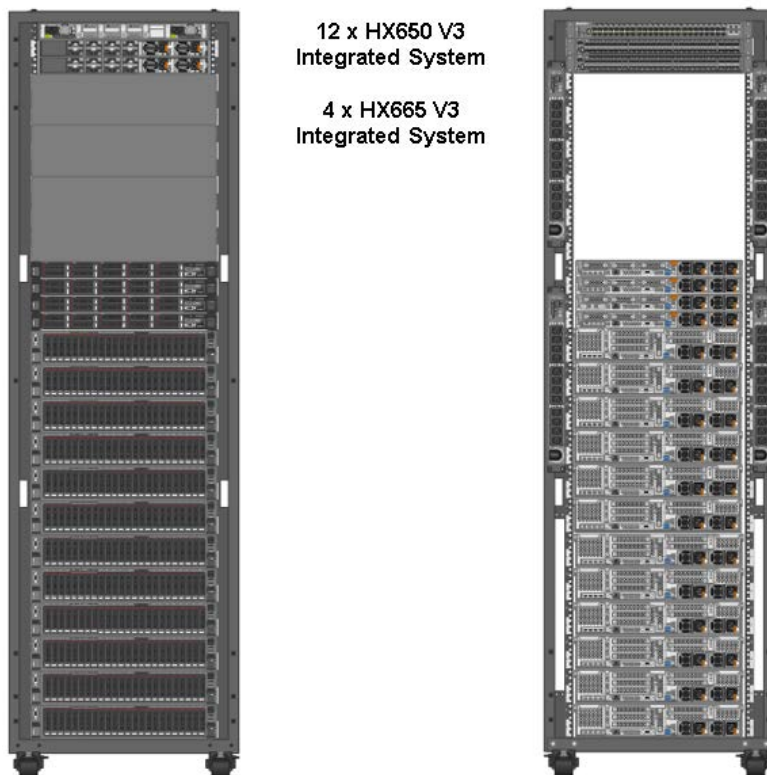
Figure 11: Example deployment model for Microsoft SQL Server

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

3.2.3 VMware vCloud Suite deployment model

In this example deployment model, the requirement is to support generic workloads using a VMware vCloud Suite. For more details see section 8.7 on page 96.

The vCloud edge-compute cluster uses 12 x HX650 V3 appliances configured as follows:

- Intel Xeon Gold 6454S 32C 270W 2.2GHz Processor
- 16 x 64 GB ThinkSystem 64GB TruDDR5 4800MHz
- 6 x ThinkSystem S4620 3.84TB Mixed Use SATA 6Gb HS SSD

The vCloud management cluster uses 4 x HX630 V3 appliances configured as follows:

- Intel Xeon Gold 6438Y+ 32C 205W 2.0GHz Processor
- 16 x 32 GB ThinkSystem 32GB TruDDR5 4800MHz
- 4 x ThinkSystem S4620 1.92TB Mixed Use SATA 6Gb HS SSD

Figure 12 shows the deployment of the 12x HX650 V3 appliances, 4x HX630 V3 appliances, and the TOR switches.

4 x HX630 V3
Integrated System

12 x HX650 V3
Integrated System
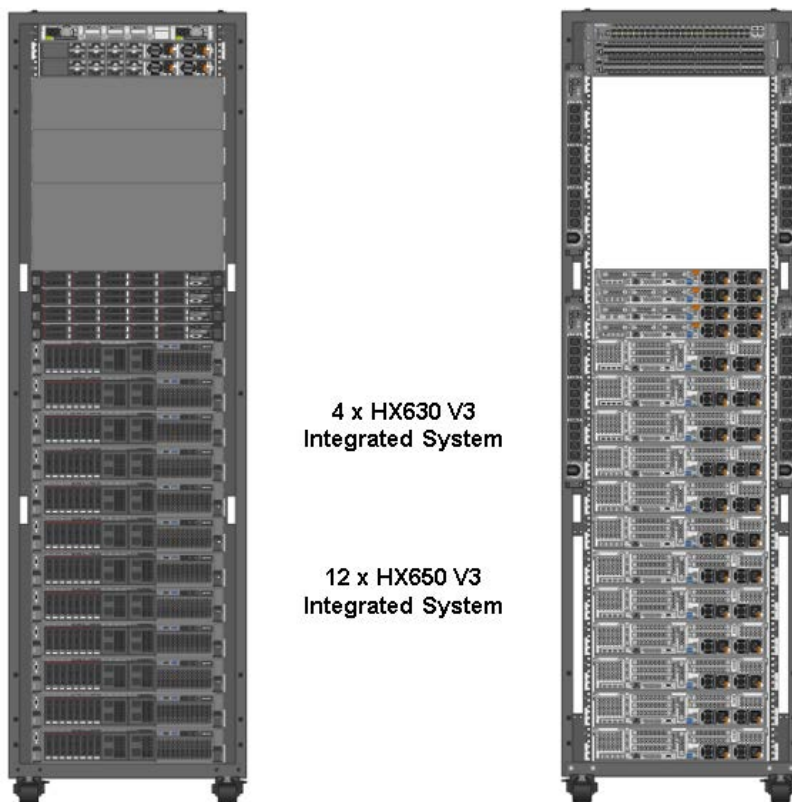
Figure 12: Example deployment model for VMware vCloud

3.3 TruScale Infrastructure as a Service deployment models

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Lenovo TruScale is an Infrastructure-as-a-service solution which gives flexible solutions and support to accelerate business needs and demands. From the data center to the edge and everything in between. Combine the convenience and flexibility of a cloud-like infrastructure that can be scaled according to your demand, with all the security and control of a traditional on-prem solution. Many ThinkAgile HX solutions are provided as TruScale offering with subscription models to suit business requirements from edge to cloud https://techtoday.lenovo.com/us/en/iaas

4 Citrix Virtual Apps and Desktops

Citrix Virtual Apps and Desktopsis a suite of virtualization software which delivers Windows virtual desktops as well as virtual applications to meet the demands of any use case. It is based on the unified FlexCast Management Architecture (FMA) platform. See this website for more details: citrix.com/products/citrix-virtual-apps-and-desktops.

4.1 Solution overview

Figure 13 provides an architecture overview of Lenovo hyper-converged Nutanix solution's reference architecture with Citrix Virtual Apps and Desktops7.15 on VMware ESXi 7.0 U2 and Nutanix AHV 20.2 hypervisors. This chapter does not address the general issues of multi-site deployment and network management and limits the description to the components that are inside the customer's intranet.
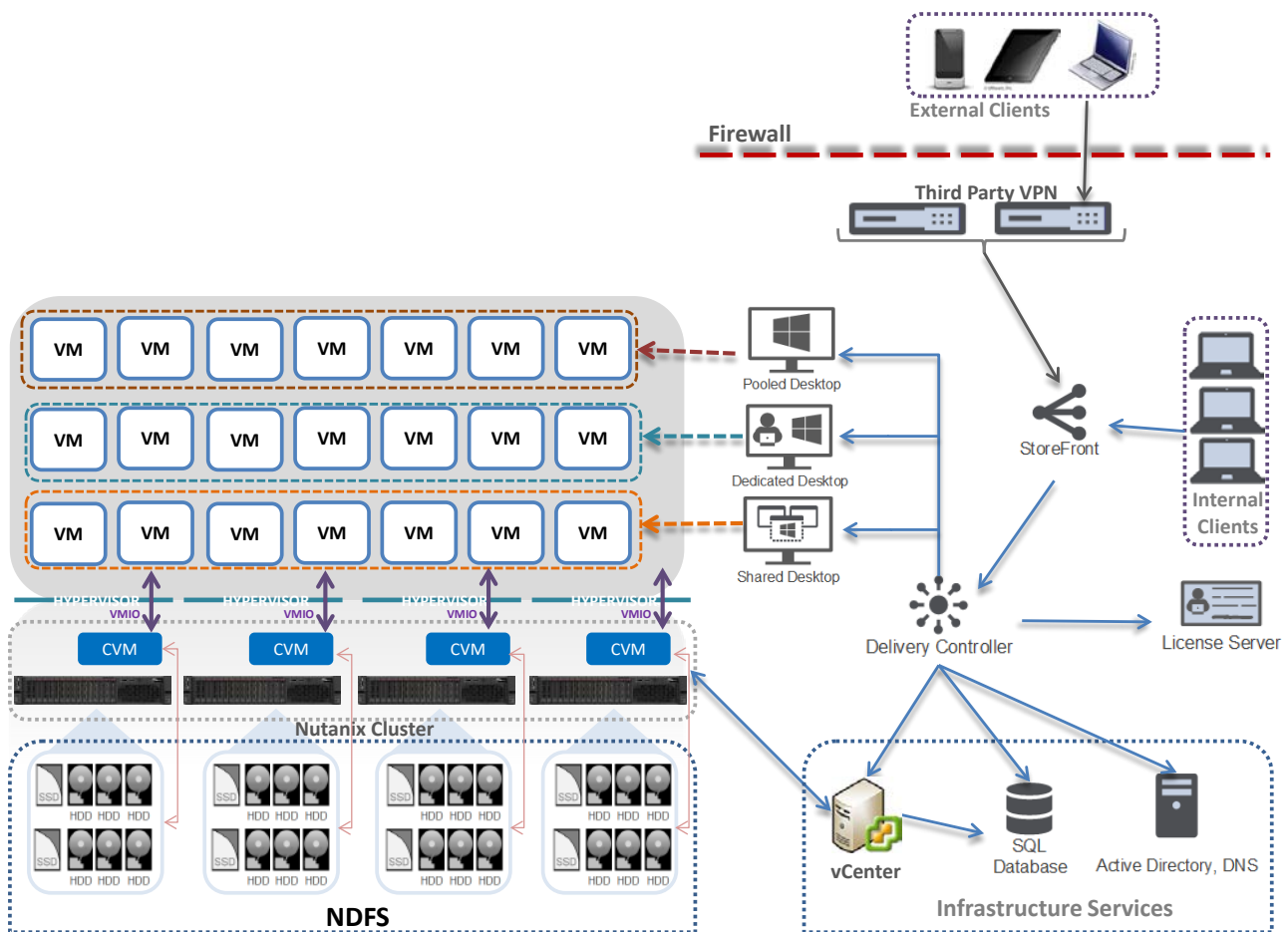

Figure 13: Lenovo ThinkAgile HX Series solution with Citrix Virtual Apps and Desktops

Pooled desktops are stateless (non-persistent) virtual desktops and dedicated desktops are persistent. Shared desktops are used for hosted shared desktops or hosted shared applications.

4.2 Component model

Figure 14 is a layered component view for the Citrix Virtual Apps and Desktops virtualization infrastructure.
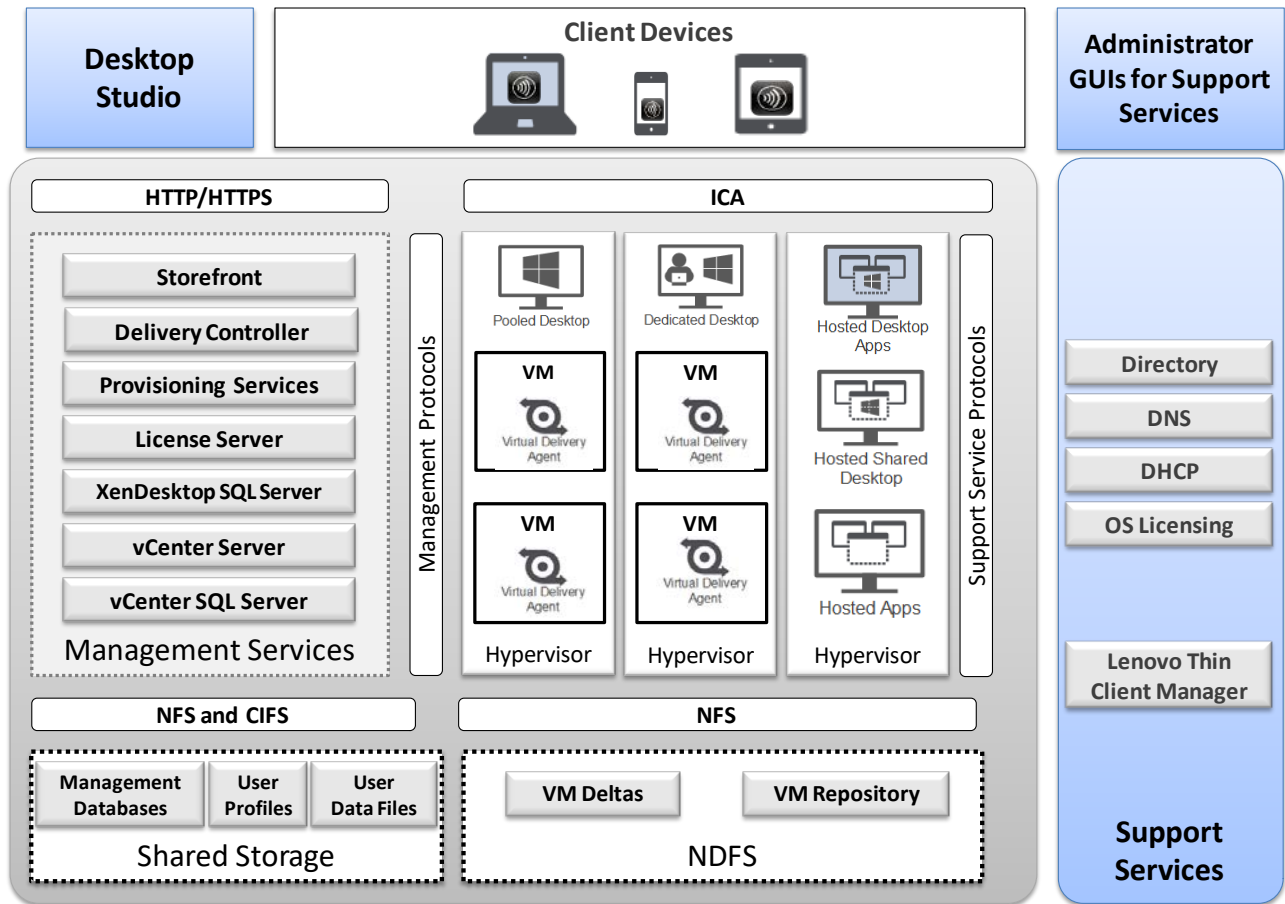
Figure 14: Component model with Citrix Virtual Apps and Desktops

Citrix Virtual Apps and Desktopsfeatures the following main components:

Desktop Studio   Desktop Studio is the main administrator GUI for Citrix Virtual Apps and Desktops. It is used to configure and manage all of the main entities, including servers, desktop pools and provisioning, policy, and licensing.

Storefront   Storefront provides the user interface to the virtual desktops environment. The Web Interface brokers user authentication, enumerates the available desktops and, upon start, delivers a .ica file to the Citrix Receiver on the user's local device to start a connection. The Independent Computing Architecture (ICA) file contains configuration information for the Citrix receiver to communicate with the virtual desktop. Because the Web Interface is a critical component, redundant servers must be available to provide fault tolerance.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Delivery controller   The Delivery controller is responsible for maintaining the proper level of idle desktops to allow for instantaneous connections, monitoring the state of online and connected desktops, and shutting down desktops as needed.

A farm is a larger grouping of virtual machine servers. Each delivery controller in the Virtual Apps and Desktops acts as an XML server that is responsible for brokering user authentication, resource enumeration, and desktop starting. Because a failure in the XML service results in users being unable to start their desktops, it is recommended that you configure multiple controllers per farm.

PVS and MCS   Provisioning Services (PVS) is used to provision stateless desktops at a large scale. Machine Creation Services (MCS) is used to provision dedicated or stateless desktops in a quick and integrated manner. For more information, see "Citrix Virtual Apps and Desktops provisioning" section on page 24.

License Server   The Citrix License Server is responsible for managing the licenses for all Virtual Apps and Desktops components. Citrix Virtual Apps and Desktops has a 30-day grace period that allows the system to function normally for 30 days if the license server becomes unavailable. This grace period offsets the complexity of otherwise building redundancy into the license server.

SQL Server   Each Citrix Virtual Apps and Desktopssite requires an SQL Server database that is called the data store, which used to centralize farm configuration information and transaction logs. The data store maintains all static and dynamic information about the virtual desktop environment. Because the SQL server is a critical component, redundant servers must be available to provide fault tolerance.

vCenter Server   By using a single console, vCenter Server provides centralized management of the virtual machines (VMs) for the VMware ESXi hypervisor. VMware vCenter can be used to perform live migration (called VMware vMotion), which allows a running VM to be moved from one physical server to another without downtime.

Redundancy for vCenter Server is achieved through VMware high availability (HA). The vCenter Server also contains a licensing server for VMware ESXi.

vCenter SQL Server   vCenter Server for VMware ESXi hypervisor requires an SQL database. The vCenter SQL server might be Microsoft® Data Engine (MSDE), Oracle, or SQL Server. Because the vCenter SQL server is a critical component, redundant servers must be available to provide fault tolerance. Customer SQL databases (including respective redundancy) can be used.

Client devices	Citrix Virtual Apps and Desktopssupports a broad set of devices and all major device operating platforms, including Apple iOS, Google Android, and Google ChromeOS. Citrix Virtuall Apps and Desktops enables a rich, native experience on each device, including support for gestures and multi-touch features, which customizes the experience based on the type of device. Each client device has a Citrix Receiver, which acts as the agent to communicate with the virtual desktop by using the ICA/HDX protocol.
Thin-client Manager	The Lenovo Thin-client Manager (LTM) is used to manage and support Lenovo thin-client devices individually or in groups.
VDA	Each VM needs a Citrix Virtual Desktop Agent (VDA) to capture desktop data and send it to the Citrix Receiver in the client device. The VDA also emulates keyboard and gestures sent from the receiver. ICA is the Citrix remote display protocol for VDI.
Citrix Workspace Client	Citrix Workspace is the client software that provides access to applications, desktops and data easily and securely from any device, including smartphones, tablets, PCs and Macs

For more information, see the Lenovo Client Virtualization base reference architecture document that is available at this website: lenovopress.com/lp0756.

4.3 Citrix Virtual Apps and Desktops provisioning

Citrix Virtual Apps and Desktopsfeatures the following primary provisioning components for desktops and applications:

- Provisioning Services (PVS)
- Machine Creation Services (MCS)

4.3.1 Provisioning services

Hosted VDI desktops can be deployed with or without Citrix PVS. The advantage of PVS is that you can stream a single desktop image to create multiple virtual desktops on one or more servers in a data center. Figure 15 shows the sequence of operations that are executed by Citrix Virtual Apps and Desktops to deliver a hosted VDI virtual desktop.
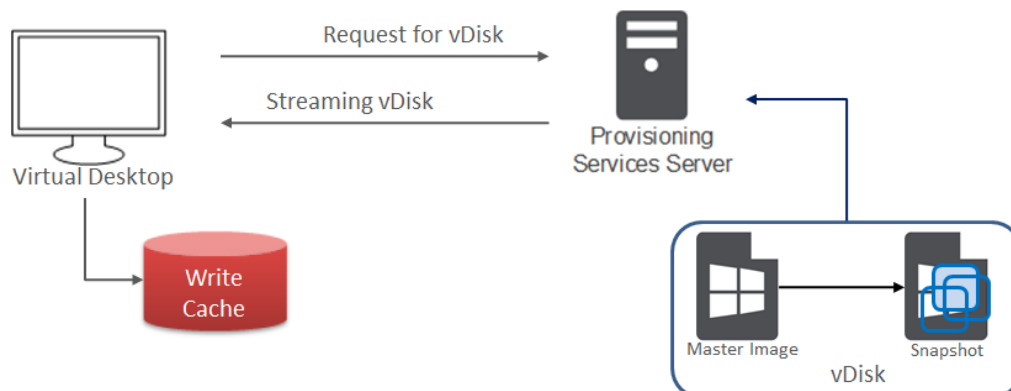

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Figure 15: Using PVS for a stateless model

When the virtual disk (vDisk) master image is available from the network, the VM on a target device no longer needs its local hard disk drive (HDD) to operate; it boots directly from the network and behaves as if it were running from a local drive on the target device, which is why PVS is recommended for stateless virtual desktops. PVS often is not used for dedicated virtual desktops because the write cache is not stored on shared storage.

PVS is also used with Microsoft Roaming Profiles (MSRPs) so that the user's profile information can be separated out and reused. Profile data is available from CIFS based shared storage.

It is a best practice to use snapshots for changes to the master VM images and also keep copies as a backup.

4.3.2 Machine creation services

Unlike PVS, MCS does not require more servers. Instead, it uses integrated functionality that is built into the hypervisor and communicates through the APIs. Each desktop has one difference disk and one identity disk (as shown in Figure 16). The difference disk is used to capture any changes that are made to the master image. The identity disk is used to store information, such as device name and password.
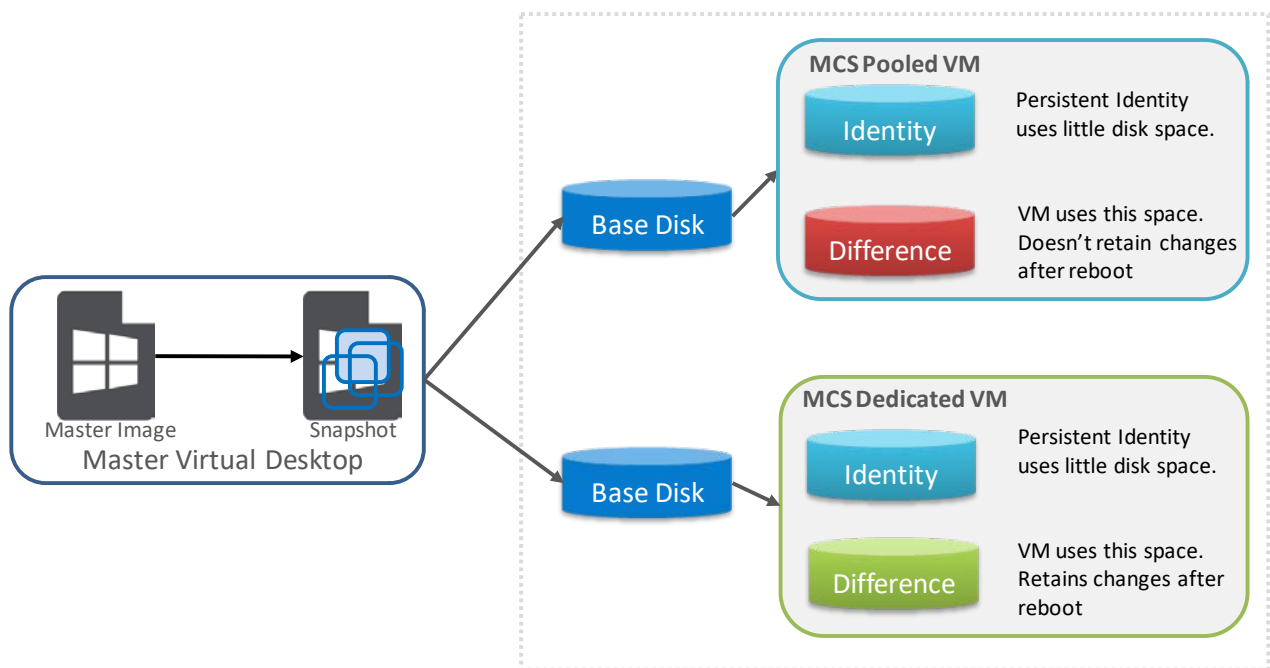
Figure 16: MCS image and difference/identity disk storage model

The following types of image assignment models for MCS are available:

- Pooled-random: Desktops are assigned randomly. When they log off, the desktop is free for another user. When rebooted, any changes that were made are destroyed.
- Pooled-static: Desktops are permanently assigned to a single user. When a user logs off, only that user can use the desktop, regardless if the desktop is rebooted. During reboots, any changes that are made are destroyed.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

- Dedicated: Desktops are permanently assigned to a single user. When a user logs off, only that user can use the desktop, regardless if the desktop is rebooted. During reboots, any changes that are made persist across subsequent restarts.

MCS thin provisions each desktop from a master image by using built-in technology to provide each desktop with a unique identity. Only changes that are made to the desktop use more disk space.

There is a new caching option in Citrix Virtual Apps and Desktops 7.9 for Pooled and Hosted Shared desktops. Figure 17 shows a screenshot of the option.
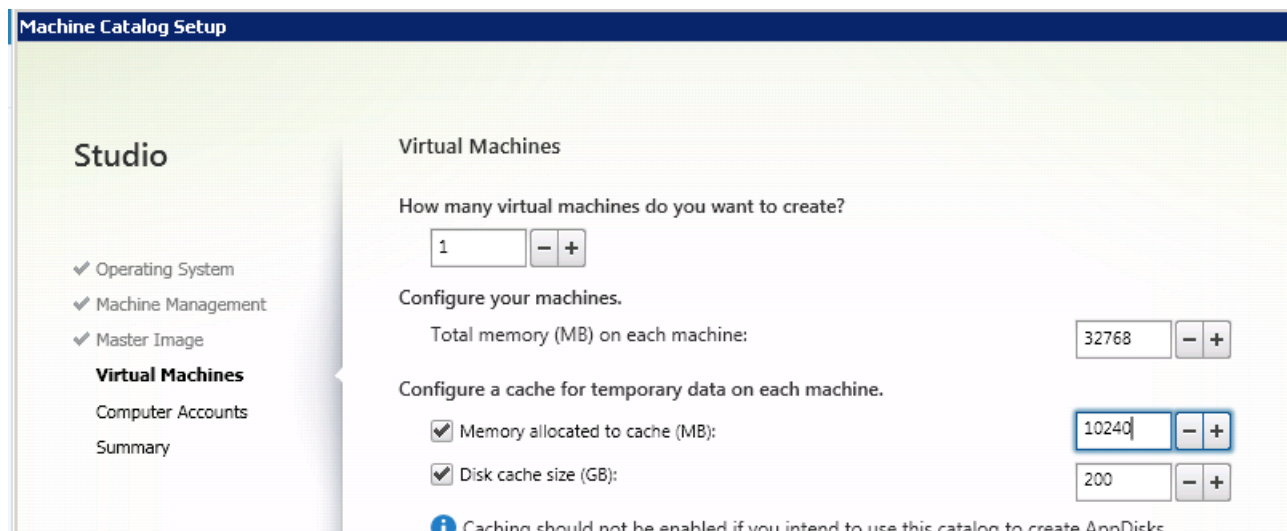
Figure 17: Caching option for Pooled and Hosted Shared desktops

4.4 Management VMs

A key part of the Citrix Virtual Apps and Desktopsenvironment is the various management VMs used to manage the VDI infrastructure and user VMs. Table 4 lists the VM requirements and performance characteristics of each management service for Citrix Virtual Apps and Desktops.

Table 4: Characteristics of Citrix Virtual Apps and Desktops and ESXi management services

Management service VM	Virtual processors	System memory	Storage	Windows OS	HA needed	Performance characteristic
Delivery controller	4	8 GB	60 GB	2012 R2	Yes	5000 user connections
Storefront	4	4 GB	60 GB	2012 R2	Yes	30,000 connections per hour
Citrix licensing server	2	4 GB	60 GB	2012 R2	No	170 licenses per second
SQL server	2	8 GB	60 GB	2012 R2	Yes	5000 users

Management service VM	Virtual processors	System memory	Storage	Windows OS	HA needed	Performance characteristic
PVS servers	4	32 GB	60 GB (depends on number of images)	2012 R2	Yes	Up to 1000 desktops, memory should be a minimum of 2 GB plus 1.5 GB per image served
vCenter server	8	16 GB	60 GB	2012 R2	No	Up to 2000 desktops
vCenter SQL server	4	8 GB	200 GB	2012 R2	Yes	Double the virtual processors and memory for more than 2500 users

Table 5 lists the number of management VMs for each size of users following the recommendations for high availability and performance. The number of vCenter servers is half of the number of vCenter clusters because each vCenter server can handle two clusters of up to 1000 desktops.

Table 5: Management VMs needed

Management VMs	300 users	600 users	1200 users	3000 users
Delivery Controllers Includes Citrix Licensing server Includes Web server	2 (1+1) Y Y	2 (1+1) Y Y	2 (1+1) N N	2 (1+1) N N
Web Interface	N/A	N/A	2 (1+1)	2 (1+1)
Citrix licensing servers	N/A	N/A	1	1
SQL servers	2 (1+1)	2 (1+1)	2 (1+1)	2 (1+1)
PVS servers for stateless case only	2 (1+1)	2 (1+1)	2 (1+1)	2 (1+1)
ESXi management service VM	**300 users**	**600 users**	**1200 users**	**3000 users**
vCenter servers	1	1	2	2
vCenter SQL servers	2 (1+1)	2 (1+1)	2 (1+1)	2 (1+1)

It is assumed that common services, such as Microsoft Active Directory, Dynamic Host Configuration Protocol (DHCP), domain name server (DNS), and Microsoft licensing servers exist in the customer environment.

There are 3 ways to deploy the management VMs for Citrix Virtual Apps and Desktops:

- Separate servers outside HX cluster
- Servers within the HX cluster
- In the Citrix Cloud using the Citrix Workspace Appliance (see section 4.4.1)

These management VMs can be run on separate servers from the HX series cluster or within the cluster itself. Separating out the VMs means that the management VMs can be separately managed and sized to the

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

requirements and dedicated servers used for the user VMs. Putting all of the VMs together in one cluster means that the compute servers will execute less user VMs and additional resources are needed for the much larger and more granular management VMs. Lenovo recommends that the management and user VMs are separated for all but the smallest deployments (i.e. less than 600 users).

4.4.1 Citrix Workspace Appliance

Citrix Workspace Appliance (CWA) hosts the management VMs for virtual apps and desktops in the Citrix Cloud; for a service fee. This significantly reduces the skills, effort, and hardware needed to manage a VDI environment. The customer only needs to configure a cluster of servers to host user VMs for dedicated or stateless desktops.

Citrix installs and manages the following components in Citrix Cloud:

Delivery Controller

The Delivery Controller is the central control layer component in a deployment. The Controller's services communicate through the Cloud Connectors in each resource location to:

- Distribute applications and desktops.

- Authenticate and manage user access.

- Broker connections between users and their virtual desktops and applications.

- Optimize use connections, and load-balance these connections.

- Track which users are logged on and where, which session resources the users have, and if users need to reconnect to existing applications. This includes managing the state of desktops, starting and stopping them based on demand and administrative configuration.

Data from the Controller services is stored in a Microsoft SQL Server site database. A deployment also uses a Configuration Logging database, plus a monitoring database used by Director.

When the Citrix Cloud is hosted on Microsoft Azure, SQL Server on Azure VMs (IaaS) is supported. Azure SQL (PaaS) Database is not supported.

Citrix license management

License management functionality communicates with the Controller to manage licensing for each user's session and allocate license files. The customer administrator does not need to configure or manage anything with licensing. All that work is done automatically in Citrix Cloud.

Citrix Studio

Studio is the management console you use to configure and manage connections, machine catalogs, and Delivery Groups. Studio launches when you select **Manage** in the Citrix Cloud console.

Citrix Director

Director enables IT support and help desk teams to monitor an environment, troubleshoot issues before they become system-critical, and perform support tasks for end users. Displays include:

- Real-time session data from the Broker Service in the Controller, which includes data from the broker agent in the VDA.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

- Historical data from the Monitor Service in the Controller.

- Data about HDX traffic (also known as ICA traffic).

Director provides the data when you select **Monitor** in the Citrix Cloud console.

Workspace configuration

From Workspace configuration in Citrix Cloud, you can:

- Specify which services are integrated with Citrix Workspace.

- Customize the URL that your subscribers use to access their workspace.

- Customize the appearance of subscribers' workspaces, such as logos, color, and preferences.

- Specify how subscribers authenticate to their workspace, such as using Active Directory or Azure Active Directory.

- Specify external connectivity for resource locations used by your subscribers.

Citrix StoreFront

Citrix StoreFront is the predecessor to Citrix Workspace, and is used in on-premises deployments. It offers several similar features to Citrix Workspace, including subscriber access to apps and desktops, authentication, and other subscriber data that ensures consistent experience across devices.

In a service deployment, you can optionally install StoreFront servers in resource locations. Having local stores can help deliver apps and desktops during network outages.

Citrix Gateway

When users connect from outside the corporate firewall, Citrix Virtual Apps and Desktops can use Citrix Gateway technology to secure these connections with TLS. The Citrix Gateway or VPX virtual appliance is an SSL VPN appliance deployed in the DMZ. It provides a single secure point of access through the corporate firewall.

Although Citrix installs and manages Citrix Gateway in Citrix Cloud, you can also optionally install Citrix Gateway in resource locations.

Components in resource locations (on-premises):

A resource location contains resources required to deliver services to your subscribers (users). You manage these resources from Citrix Cloud. Resource locations contain different resources depending on which Citrix Cloud services you're using and the services you want to provide to your users.

To interact with Citrix Cloud, each resource location needs Cloud Connectors and access to a Microsoft Active Directory domain.

In a Citrix Virtual Apps and Desktops service deployment, a resource location contains items from the access layer and resource layer:

- Cloud Connectors

- Active Directory domain controller

- Virtual Delivery Agents (VDAs)

- Hypervisors that provision VDAs and store their data, if used

- Citrix Gateway (optional)

- StoreFront servers (optional)

Cloud Connector

Every resource location contains at least one Cloud Connector (two or more are recommended for redundancy). A Cloud Connector is the communications channel between the components in the Citrix Cloud and components in the resource location. In the resource location, the Cloud Connector acts as a proxy for the Delivery Controller in Citrix Cloud.

You install Cloud Connectors from the Citrix Cloud console. Citrix then manages and updates the Cloud Connectors automatically.

Virtual Delivery Agents (VDAs)

Each physical or virtual machine that delivers applications and desktops must have a VDA. The VDA registers with a Cloud Connector. After registration, connections are brokered from those resources to users. VDAs establish and manage the connection between the machine and the user device, and apply policies that are configured for the session.

The VDA communicates session information to the Cloud Connector through a broker agent in the VDA. The broker agent hosts multiple plugins and collects real-time data.

VDAs are available for Windows server and desktop operating systems. VDAs for Windows server operating systems allow multiple users to connect to the server at one time. VDAs for Windows desktop operating systems allow only one user to connect to the desktop at a time. Linux VDAs are also available.

Throughout this documentation, "VDA" refers to the agent and the machine on which it is installed.

Hypervisors and cloud services

A hypervisor or cloud service contains the VDAs that host applications and desktops.

To provision virtual machines that deliver applications and desktops, you can use:

- Machine Creation Services: The MCS technology is built into Studio and is accessed automatically through the Citrix Cloud Console. MCS creates copies of a master image to create and provision VMs.

- Citrix Provisioning (formerly Provisioning Services): The Citrix Provisioning technology streams a master image to user devices. Citrix Provisioning doesn't require a hypervisor, so you can provision physical machines.

- Another provisioning tool of your choice.

Although many deployments use hypervisors, you don't need one if:

- Your applications and desktops are hosted on physical machines.

- You use Citrix Provisioning to provision VMs.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

- You want to deploy Remote PC Access, which enables employees to remotely access their physical PCs.

Active Directory

Although not a Citrix component, Microsoft Active Directory is required for authentication and authorization in any deployment. The Kerberos infrastructure in Active Directory is used to guarantee the authenticity and confidentiality of communications with Citrix Cloud.

Items that help deliver desktops and applications

As part of delivering applications and desktops to users in a production environment, you configure the following items.

Host connection

A host connection enables communication between components in the control plane (Citrix Cloud) and VDAs in a hypervisor or cloud service. Connection specifications include:

- The address and credentials to access the host

- Which tool you use to create VMs

- The storage method to use, and the machines to use for storage

- Which network the VMs will use

Machine catalog

A machine catalog is a collection of virtual or physical machines that have the same operating system type: server or desktop.

If you use VMs, you can create a master image (also known as template) on the hypervisor or cloud service, and install a VDA on the master image. You can also install applications on the master image, if you want them to appear on all machines created from that image and don't want to virtualize them. Then, you create a catalog using a Citrix tool (MCS or Citrix Provisioning) or your own tools. With Citrix tools, the catalog creation process provisions identical VMs from that image.

If you use your own tools to provision VMs, or if you use physical machines, the catalog creation process adds those machines to the catalog.

For technical details about the Citrix provisioning tools, see [Citrix Virtual Apps and Desktops Image Management](#).

Delivery Group

A Delivery Group specifies:

- One or more machines from a machine catalog.

- Optionally, users who are allowed to access those machines. Alternatively, you can specify users through the Citrix Cloud console.

- Optionally, which applications and desktops users can access. Alternatively, you can specify applications through the Citrix Cloud console.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Delivering applications and desktops

[Delivery methods](#) describes the choices available to deliver applications and desktops to users.

Service Level Agreement

The Citrix Virtual Apps and Desktops service (the Service) is designed using industry best practices to achieve cloud scale and a high degree of service availability.

For complete details about Citrix's commitment for availability of Citrix Cloud services, see the [Service Level Agreement](#).

Performance against this goal can be monitored on an ongoing basis at https://status.cloud.com.

For more information on CWA and the Citrix Cloud, see docs.citrix.com/en-us/citrix-cloud.html.

4.5 Graphics acceleration

This section is specific to the Lenovo ThinkAgile HX3520-G and ThinkAgile HX3521-G that support GPU acceleration. The VMware ESXi hypervisor supports the following options for graphics acceleration:

- Dedicated GPU with one GPU per user, which is called virtual dedicated graphics acceleration (vDGA) mode.
- GPU hardware virtualization (vGPU) that partitions each GPU.
- Shared GPU with users sharing a GPU, which is called virtual shared graphics acceleration (vSGA) mode and is not recommended because of user contention for shared use of the GPU.

The vDGA (or pass-through) option has a low user density as it restricts a single user to access each very powerful GPU. This option is not flexible and is no longer cost effective even for high-end power users. Therefore vDGA is no longer recommended especially given that the performance of the equivalent vGPU mode is similar.

When using the vGPU option with ESXi 6.5 and the latest drivers from NVidia, it is necessary to change the default GPU mode from "Shared" (vSGA) to "Shared Direct" (vGPU) for each GPU using VMware vCenter. This enables the correct GPU support for the VMs which would otherwise result in the VM not powering on correctly and the standard "graphics resources not available" error message. The host needs to be rebooted for the changes to take effect.

The performance of graphics acceleration was tested using the Lenovo ThinkSystem SR650 servers. Each server supports up to two GPU adapters. The Heaven benchmark is used to measure the per user frame rate for different GPUs, resolutions, and image quality. This benchmark is graphics-heavy and is fairly realistic for designers and engineers. Power users or knowledge workers usually have less intense graphics workloads and can achieve higher frame rates. Table 6 lists the results of the Heaven benchmark as FPS that are available to each user with the GRID 2,0 M60 adapter by using vGPU mode with DirectX 11.

Table 6: Performance of GRID 2.0 M60 vGPU modes with DirectX 11

Quality	Tessellation	Anti-Aliasing	Resolution	M60-8Q	M60-4Q	M60-2Q	M60-4A	M60-2A
High	Normal	0	1280x1024	Untested	Untested	32.07	59.94	33.12
High	Normal	0	1680x1050	Untested	50.43	25.68	N/A	N/A
High	Normal	0	1920x1200	Untested	41.82	21.72	N/A	N/A
Ultra	Extreme	8	1280x1024	Untested	Untested	18.24	36.91	18.76
Ultra	Extreme	8	1680x1050	57.61	29.79	14.40	N/A	N/A
Ultra	Extreme	8	1920x1080	50.50	26.18	12.63	N/A	N/A
Ultra	Extreme	8	1920x1200	46.17	23.01	Untested	N/A	N/A
Ultra	Extreme	8hai	2560x1600	27.46	14.19	Untested	N/A	N/A

Lenovo recommends that a medium to high powered CPU, such as the Xeon Scalable 6130, is used for accelerated graphics applications tend to also require extra load on the processor. For vGPU mode, Lenovo recommends at least 384GB of server memory. Because there are many variables when graphics acceleration is used, Lenovo recommends that testing is done in the customer environment to verify the performance for the required user workloads.

4.6 Performance testing

This section describes the performance benchmarking tool and the results obtained for different configurations of a cluster of 4 Lenovo ThinkAgile HX3000 appliances and certified nodes.

4.6.1 Login Enterprise benchmarking

Login Enterprise is tool for benchmarking virtual desktops and applications. It provides VSIMax metric to measure the density of virtual desktops and EUX score metric to measure application response time and end user experience. Login Enterprise comes with knowledge worker workload which uses standard office applications such as MicroSoft office applications and Edge/Chrome browser.   It also provides response time for user login, application open and closing time. Also note that the previous LoginVSI benchmark is withdrawn from the market and Login Enterprise will be a replacement tool for virtual desktop benchmark. The Login Enterprise workload is 60-70% heavier than previous LoginVSI and it matches real world requirements.

This section shows the virtual desktop performance results for Lenovo ThinkAgile HX all flash appliances and certified nodes.

The recommended configuration of the Nutanix CVM is as follows:
- vCPU   12
- CPU Reservation   10000 MHz
- Memory   48 GB
- Memory Reservation   48 GB
- NUMA   No Affinity
- Advance CPU – Scheduling Affinity   No Affinity

4.6.2 Login Enterprise Results on ESXi with 3rd Generation Intel Xeon Processors

Table 7 shows the Login Enterprise 4.9.10 benchmark results for knowledge worker workload on Windows 10. Lenovo modified scripts to support office 2007 and PDF print operation to match with previous LoginVSI benchmark runs. The test was done on 3 node HX7530 vSphere 7.0u3 cluster with 800 desktops with 2x1.6TB NVMe cache drives and 4x1.92TB SATA SSD capacity drives. The Login Enterprise VSIMax is 764 which provides 255 virtual desktops per node. The results shows Login Enterprise workload is ~50% heavier than previous version.

Table 7: Login Enterprise **Performance with VMware ESXi 7.0 U3 on ThinkAgile HX7530**

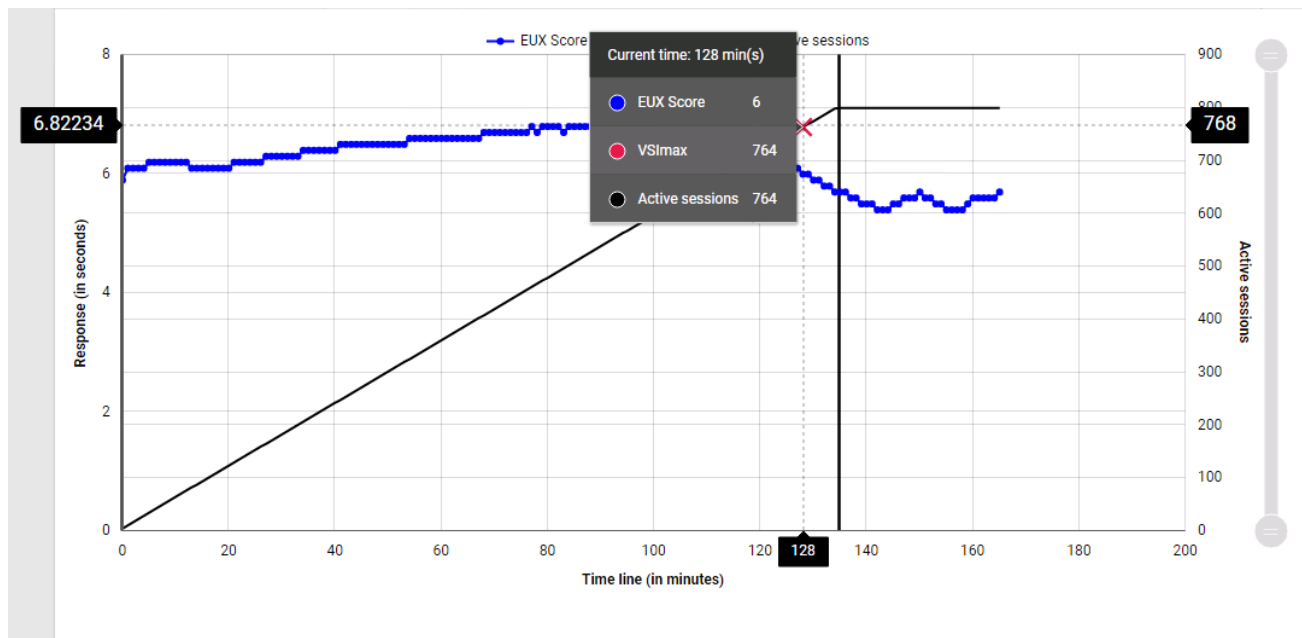| Server	Processor	Workload	Total Desktops	VSIMax	EUX Score
HX7530	2x Gold 6338N @ 2.20GHz, 32C	Login Enterprise Knowledge worker (3 vcpu+2GB+40GB disk)	800	764	6.0

Figure 18: Login Entreprise EUX score for knowledge worker on ESXi

4.6.3 Login Enterprise Results on AHV with 3rd Gen Intel Xeon Scalable Processors

Table 8 shows the Login Enterprise 4.9.10 benchmark results for knowledge worker workload on Windows 10.. The test was done on 3 node HX7530 appliance AHV cluster with 720 desktops with 2x1.6TB NVMe cache drives and 4x1.92TB SATA SSD capacity drives. The VSIMax is 669 which provides 225 virtual desktops per node.

Table 8: Login Enterprise **Performance with AHV on ThinkAgile HX7530**

Server	Processor	Workload	Total Desktops	VSIMax	EUX Score

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

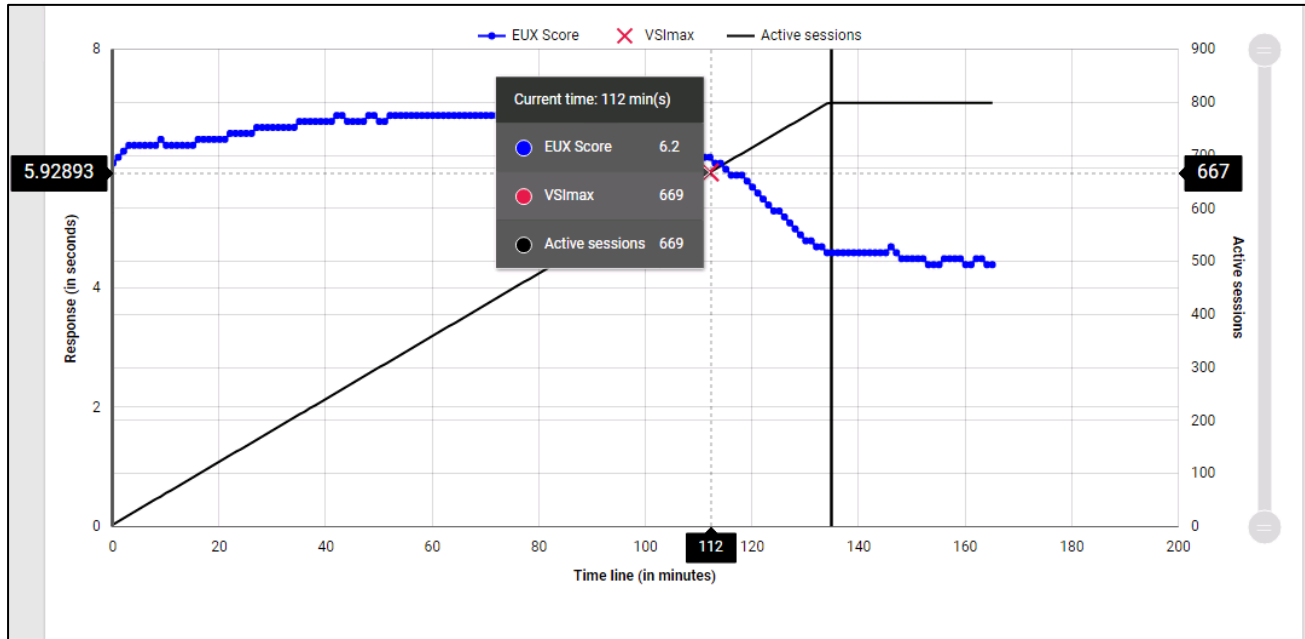|					
HX7530	2 x Intel Xeon Gold 6338N @ 2.20GHz, 32C	Login Enterprise Knowledge worker (3 vcpu+2GB+40GB disk)	720	669	6.2

Figure 19: Login Entreprise EUX score for knowledge worker on AHV

4.6.4 Login Enterprise Results on AHV with 4th Gen Intel Xeon Scalable Processors

Table 9 shows the Login Enterprise 5.1.2 benchmark results for knowledge worker workload on Windows 10. The test was done on 3 node AHV cluster with ThinkAgile HX630 V3 nodes with 1000 desktops 4x ThinkSystem P5620 3.2TB Mixed Use NVMe PCIe 4.0 x4 HS SSD. The VSIMax is 970 which provides 323 virtual desktops per node. The 4th Gen Intel Xeon Scalable processors provide better performance and high density than previous generations.

Table 9: Login Enterprise **Performance with AOS 6.5.4 on ThinkAgile HX630 V3**

Server	Processor	Workload	Total Desktops	VSIMax	EUX Score
HX7530	2x Intel Xeon Gold 6454S @ 2.20GHz, 32C	Login Enterprise Knowledge worker (3 vcpu+2GB+40GB disk)	1000	970	5.9

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**
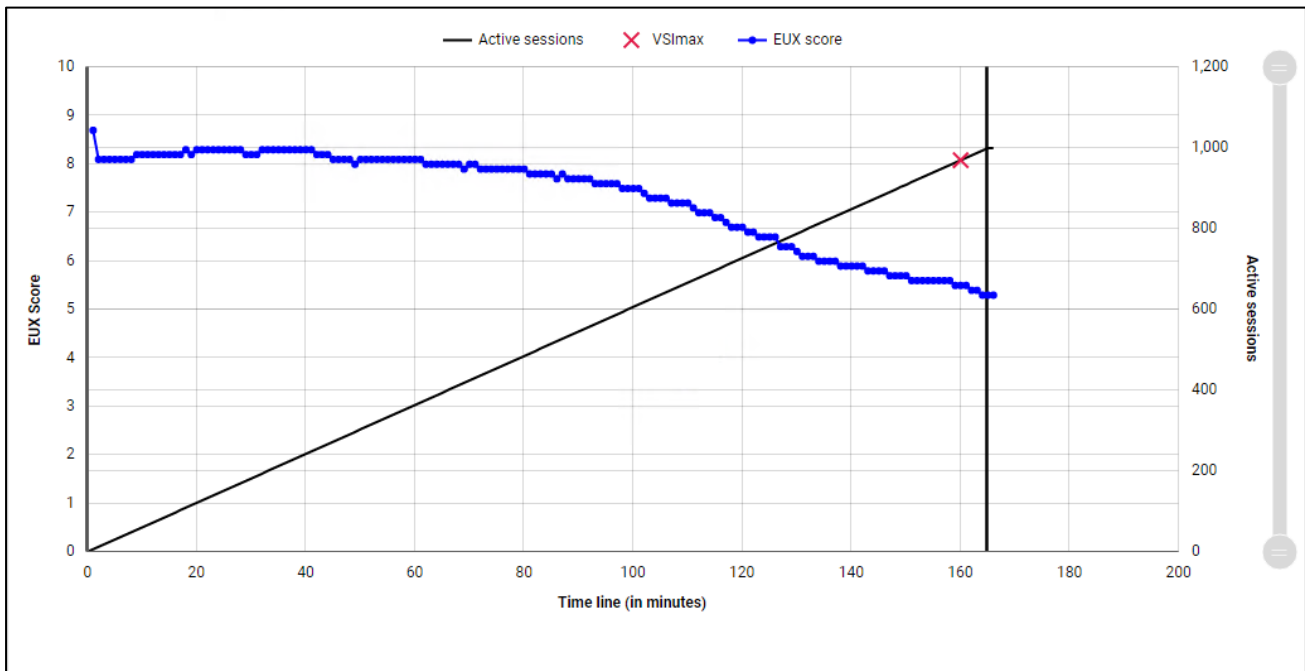
Figure 20: Login Entreprise EUX score for knowledge worker on AHV

4.7 Performance recommendations

This section provides sizing performance recommendations and best practices.

4.7.1 Sizing recommendations for virtual desktops

The default recommendation is two Xeon Scalable 6338N processors and 768 GB of system memory because this configuration provides the best coverage and density for a range of users. Assuming there is enough storage configured for the VMs, this configuration is recommended for any of the HX 3000 Series appliances.

For a knowledge worker, Lenovo testing shows that 250 users per server is a good baseline and has an average of 75% usage of the processors in the server. For the degraded failover case, Lenovo testing shows that 330 users per server have an average of 87% usage of the processors. By using a target of 250 users per server, the maximum number of knowledge workers is 4000 in a 16 node cluster.

For a power worker, Lenovo testing shows that 220 users per server is a good baseline and has an average of 75% usage of the processors in the server, For the degraded failover case, Lenovo testing shows that 275 users per server have an average of 84% usage of the processors. By using target of 220 users per server, the maximum number of power workers is 3500 in a 16 node cluster.

Table _10_ summarizes the processor usage with ESXi for the recommended user counts for normal mode and failover mode.

Table 10: Processor usage

Processor	Workload	Users per Server	CPU Utilization
Two 6338N	Knowledge worker	250 users – Normal Mode	75%
Two 6338N	Knowledge worker	330 users – Failover Mode	85%
Two 6338N	Power worker	220 users – Normal Mode	75%
Two 6338N	Power worker	275 users – Failover Mode	84%

Table *11* lists the recommended number of virtual desktops per server for different workload types and VM memory sizes. The number of users is reduced in some cases to fit within the available memory and still maintain a reasonably balanced system of compute and memory.

Table 11: Recommended number of virtual desktops per server

Workload	Knowledge worker	Power worker
Processor	Two 6338N	Two 6338N
VM memory size	4 GB	5 GB
System memory	1000 GB	1000 GB
Memory overhead of CVM	24 GB	24 GB
Desktops per server (normal mode)	180	150
Desktops per server (failover mode)	240	200

Table *12* lists the approximate number of compute servers that are needed for different numbers of users and Knowledge worker workloads.

Table 12: Compute servers needed for Knowledge workers and different numbers of users

Knowledge workers	500 users	1000 users	1500 users	3000 users
Compute servers @180 users (normal)	3	7	9	16
Compute servers @240 users (failover)	2	6	8	15

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

4.7.2 Best practices

The number of desktops that can be run on a specific server depends upon the available system memory, compute power of the processors, and number of logons per second during a logon storm. For a cost-effective solution, the maximum number of users should be put on each server to balance processor, memory, storage I/O, and networking. Lenovo recommends using all flash appliances for situations where the user logon rate is high or time to reboot all the VMs on a node must be less than 10 minutes.

Another important consideration for compute servers is system memory. For stateless users, the typical range of memory that is required for each desktop is 2 GB - 4 GB. For dedicated users, the range of memory for each desktop is 2 GB - 6 GB. In general, power users that require larger memory sizes also require more virtual processors. This reference architecture standardizes on 2 GB per desktop as the minimum requirement of a Windows 10 desktop. The virtual desktop memory should be large enough so that swapping is not needed and vSwap can be disabled.

It is a best practice not to overcommit on memory as swapping to disk can have a severe effect on performance; a better strategy is to give each desktop more memory. Alternatively, a monitoring tool can be run to gather information about existing desktops. The desktop memory size that is required does not necessarily have to match the memory supplied in a desktop machine; it can be larger or smaller.
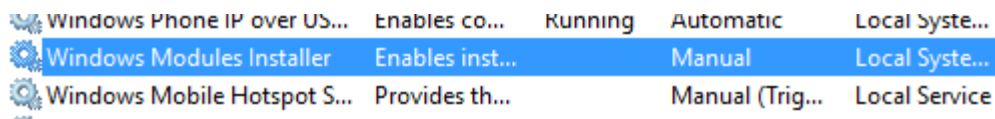
Lenovo recommends the use of VLANs to partition the network traffic. The following VLANs should be used:

- User (for web protocols, display protocols, and support service protocols)
- Management (for management protocols)
- Storage (for NDFS)

Lenovo recommends to always perform user virtualization, even if users have dedicated desktops. This separation of user-specific data makes it much easier to manage and perform upgrades.

Windows 10 was used for all of the performance testing. In general Windows 10 requires 10% to 20% more compute power than Windows 7. The following optimizations are recommended for the Windows 10 base image:

- Applied #VDILIKEAPRO Tuning Template(developed by loginVSI) – see the following for more details:
  loginvsi.com/blog/520-the-ultimate-windows-10-tuning-template-for-any-vdi-environment

- Set Adobe acrobat as a default app for PDF files using steps in following webpage:
  adobe.com/devnet-docs/acrobatetk/tools/AdminGuide/pdfviewer.html

- Disabled **Windows Modules installer** service on the base image because the CPU utilization can remain high after rebooting all the VMs. By default this service is set to manual rather than disabled.

Please refer to the websites below for best practices and optimizations recommended by Citrix:

- Windows 10 Optimization Guide:
  https://flings.vmware.com/vmware-os-optimization-tool

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

4.8 Deployment ready solutions

There are 3 deployment ready solutions for Citrix Virtual Apps and Desktops in the Lenovo Data Center Solution Configurator (DCSC): https://dcsc.lenovo.com/#/reference-solution.

Solution Identifier	SID0000443	SID0000444	SID0000445
ThinkAgile HX node	HX630 V3 Integrated System	HX630 V3 Integrated System	HX650 V3 Integrated System
Configuration	Flash	Flash	Flash
Processor	Intel Xeon Gold 5416S 16C 150W 2.0GHz	Intel Xeon Gold 6230 20C 125W 2.1GHz	Intel Xeon Gold 6448Y 32C 225W 2.1GHz
Office Workers (3GB)	200 users/node	150 users/node	N/A
Knowledge workers (4GB)	N/A	50 users/node	N/A
Power users (5GB)	N/A	N/A	200 users/node
Power user with GPU (6GB)	N/A	N/A	N/A
Light Cad with GPU (8GB)	N/A	N/A	N/A
Heavy Cad with GPU (16GB)	N/A	N/A	N/A

There are 4 TruScale deployment ready solutions are available in Lenovo Data Center Solution Configurator (DCSC) for Citrix Virtual Apps and Desktops Hosted Desktops solutions.
https://dcsc.lenovo.com/#/reference-solution/categories/b641ce2abc584a40b65c09889fe2d3ae/solutions

Solution Identifier	SID0000289	SID0000290	SID0000291	SID0000292
ThinkAgile HX node	3xHX2320	7xHX2320	8xHX2320	10xHX2320
Configuration	Flash	Flash	Flash	
Processor	Intel Xeon Silver 4214R 12C 100W 2.4GHz	Intel Xeon Gold 6246 12C 165W 3.3GHz	Intel Xeon Gold 5218R 20C 125W 2.1GHz	Intel Xeon Gold 6238R 28C 165W 2.2GHz

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

| Mixed users | 100 users/node | 300 users per node | 500 users per node | 800 users per node |

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

5 Microsoft SQL Server

Microsoft SQL Server is a database platform for large-scale online transaction processing (OLTP), data warehousing, and a business intelligence platform for data integration, analysis, and reporting solutions. It uses a common set of tools to deploy and manage databases for in-house and cloud environments.

5.1 Solution overview

Figure 21 shows high level architecture of Microsoft SQL Server on Lenovo ThinkAgile HX7520 appliances. This chapter applies equally to the ThinkAgile HX7521 certified nodes.
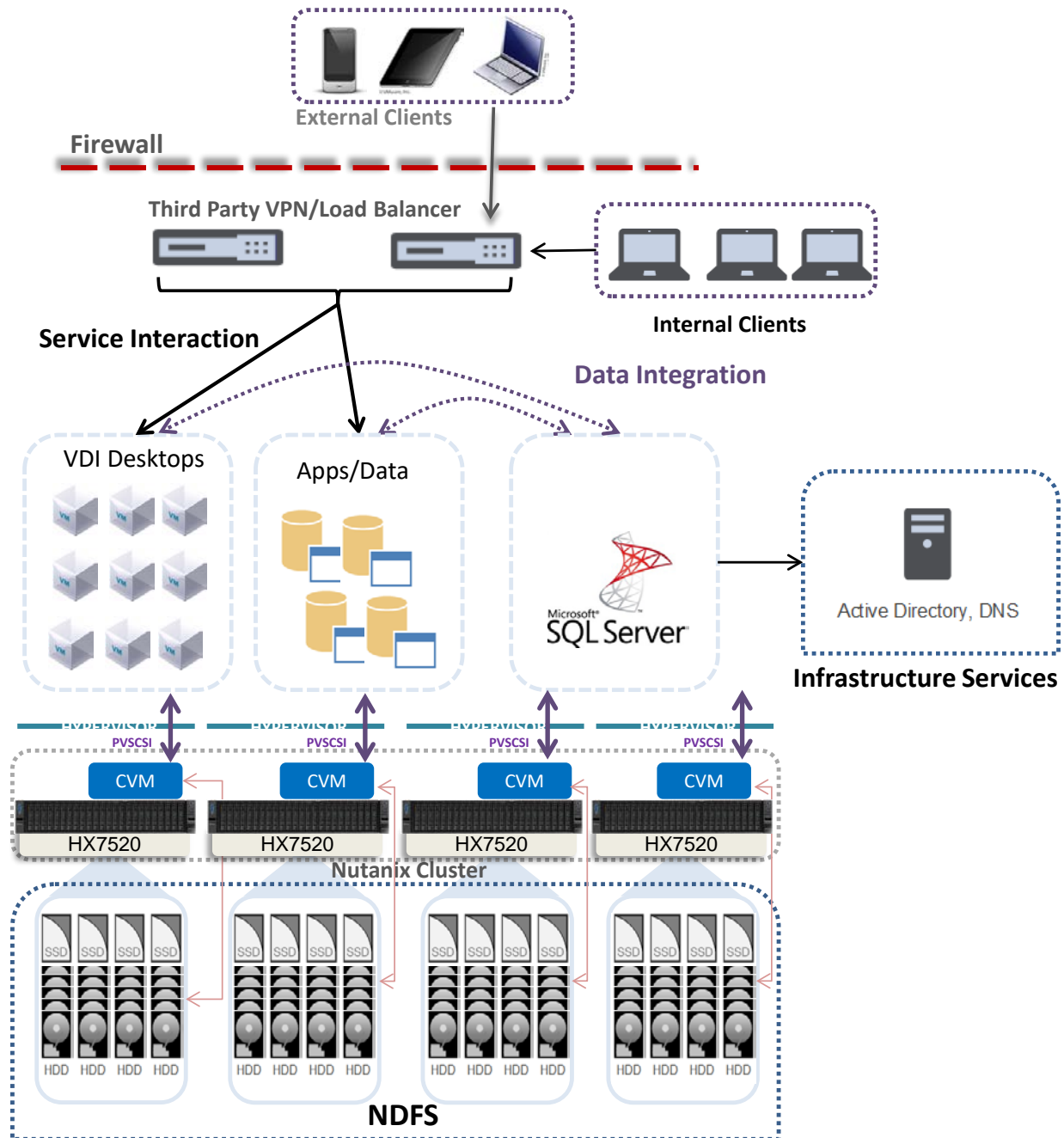

Figure 21: Lenovo ThinkAgile HX Series solution with Microsoft SQL Server

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Microsoft SQL Server can be deployed and can operate in combination with other hosted applications and provides a single scalable platform for all deployments.

5.2 Component model

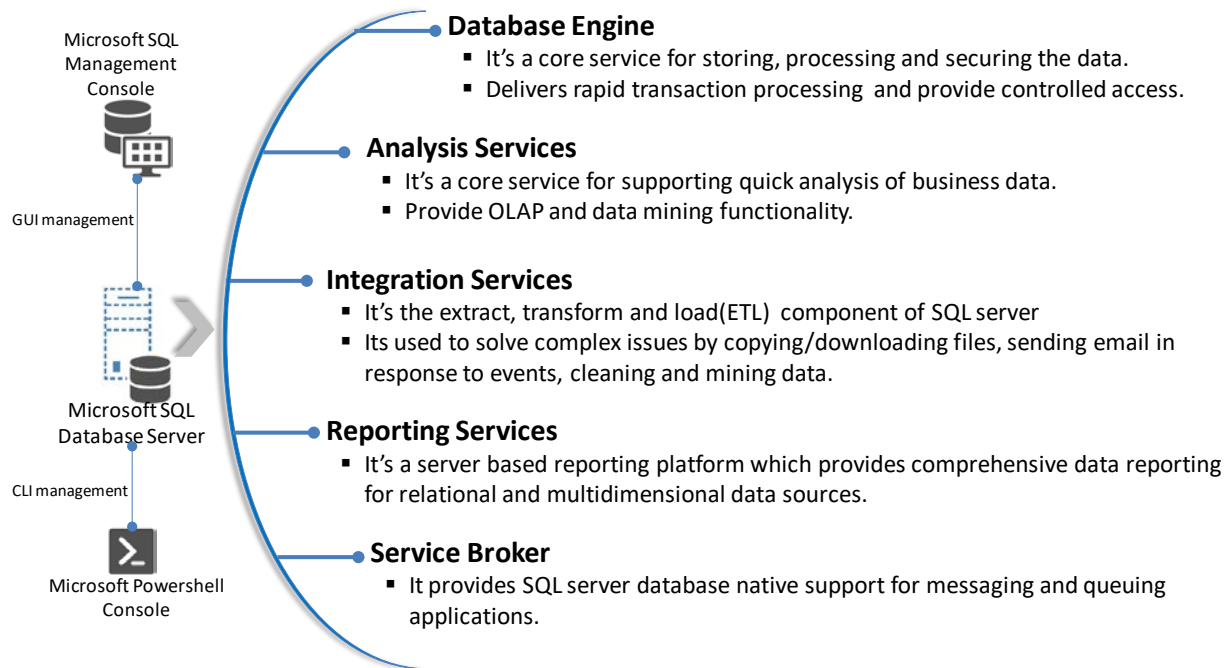Figure 22 is a layered component view for Microsoft SQL Server.

Figure 22: Component model with Microsoft SQL Server

Microsoft SQL Server features the following main components:

Database Engine	This part of SQL Server actually creates and drives relational databases.
Analysis Services	SQL Server Analysis Services (SSAS) is the data analysis component of SQL Server. It can create OLAP (OnLine Analytical Processing) cubes — sophisticated programming objects for organizing data inside a relational database — and do data mining (pulling relevant data out of a database in response to an ad-hoc question).
Integration Services	SQL Server Integration Services (SSIS) performs the extract-transform-load (ETL) process that cleans up and formats raw data from source systems for inclusion in the database as ready-to-use information.
Reporting Services	SQL Server Reporting Services (SSRS) provides reporting regardless of a database's operating system.
Service Broker	SQL Server Service Broker provides native support for messaging and queuing applications which makes it easier to build distributed and reliable applications that use the Database Engine components.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

5.3 SQL Server deployment best practices

This section describes recommended best practices to provide data optimization and high availability of Microsoft SQL Server. See also this website for Nutanix Best Practices: Microsoft SQL server: go.nutanix.com/microsoft-sql-server-converged-infrastructure.html.

5.3.1 Data optimization

By default all Nutanix storage containers are thin provisioned which reduces unused capacity and automatically provisions additional storage capacity when needed. It is also very easy to add additional storage capacity for databases by simply adding nodes to the cluster. It is also possible to set a storage reservation amount to guarantee a minimum amount of storage capacity.

Data compression can be used to further increase data capacity especially for data that is less frequently accessed. Lenovo recommends enabling compression with a delay of 1440 minutes (1 day) which minimizes the performance impact on I/O writes.

Data de-duplication is not recommended and should be disabled for SQL Server because of the frequency of changes. Note that de-duplication may be beneficial for backup volumes which are not changed very often.

A resiliency factor of 2 is the default. This provides a minimum level of data redundancy but a resiliency factor of 3 might be important in some environments. Using erasure coding saves significant storage capacity but it only recommended for archive data.

5.3.2 Cluster high availability

The minimum number of nodes in each cluster is 3 and should be at least 4 to provide failover. The following high availability features are recommended for an ESXi-based cluster:

- VMware vSphere high availability (HA) for failover
- VMware vSphere distributed resource scheduler (DRS) for load balancing
- Microsoft AlwaysOn availability groups (AAGs) for data redundancy

VMware vSphere HA pools VMs into a cluster to increase data resiliency. If a host fails, VMware HA moves the VMs to other hosts with spare capacity. Lenovo recommends enabling the "Admission Control Setting" and using the "Admission Control Policy" to set the percentage of cluster resources reserved as failover spare capacity.

VMware vSphere DRS can be used to group ESXi hosts into resource clusters to provide highly available resources and balance workloads. In order to keep the active working set for each SQL Server VM local to the node, Lenovo recommends creating a host group for each node and a "should" rule that keeps each SQL Server VM on a 1 to 1 ratio with the ESXi host. The hosts should be configured with a minimum of N+1 availability.

Because of the Nutanix architecture, Lenovo recommends spreading SQL Server instances across multiple VMs, each with adequate vdisks, rather than using several SQL Server instances on a single VM. Deploying a greater number of smaller VMs also allows advanced hypervisor features such as the distributed resource scheduler (DRS) to better manage workloads. It is easier to place small VMs than very large VMS. The distribution of SQL instance ultimately depends on performance requirements and database size.

The Microsoft AlwaysOn availability groups (AAGs) is a recommended high availability mechanism. It uses a shared-nothing approach where transactions are replicated to other nodes so each node has a full copy of the database and transaction logs. This provides a very fast failover mechanism. The DRS anti-affinity rules need to be used to ensure that the SQL Server VMs are placed on different physical hosts.

DR across datacenters can also be done using AAGs assuming there is sufficient band-width between the sites. The scenarios for active-active and active-passive DR sites using AAGs are outside the scope of this document.

5.3.3 Virtual Disk Configuration

Multiple virtual disks should be used to keep SQL binaries, database, and database logs files and achieve optimal SQL performance. All four SCSI controllers should be utilized and virtual disks should be distributed evenly across controllers as shown in Table 13.

Table 13: Mapping of virtual disks for ESXi and cluster size

Virtual Disk	vSCSI Controller Type	Controller #	Cluster size
Operating System	LSI Logic SAS	0	4 KB
SQL Binaries	LSI Logic SAS	0	4 KB
Backup/Restore	LSI Logic SAS	0	64 KB
SQL Database - 1	VMware Paravirtual	1	64 KB
SQL Database - 2	VMware Paravirtual	1	64 KB
TempDB - 1	VMware Paravirtual	2	64 KB
TempDB - 2	VMware Paravirtual	2	64 KB
TempDB log files	VMware Paravirtual	3	64 KB
Database log files	VMware Paravirtual	3	64 KB

All SQL database and log drives should be formatted with 64KB NTFS cluster size as it enhances the I/O performance without adding any overhead. The OS and SQL binary drives should be formatted with the standard 4KB NTFS cluster size. Drives space utilization should not be above 80% to achieve optimal performance.

To maximize the storage performance of SQL Server VMs, Lenovo recommends using the ESXi Paravirtual SCSI (PVSCSI) adapters. Each PVSCSI adapter can support up to 15 VMDKs.

Lenovo recommends at least eight vdisk per SQL Server VM, spread intelligently across the four available SCSI controllers. Figure 23 shows a mixture of LSI Logic SAS and PVSCSI controllers.
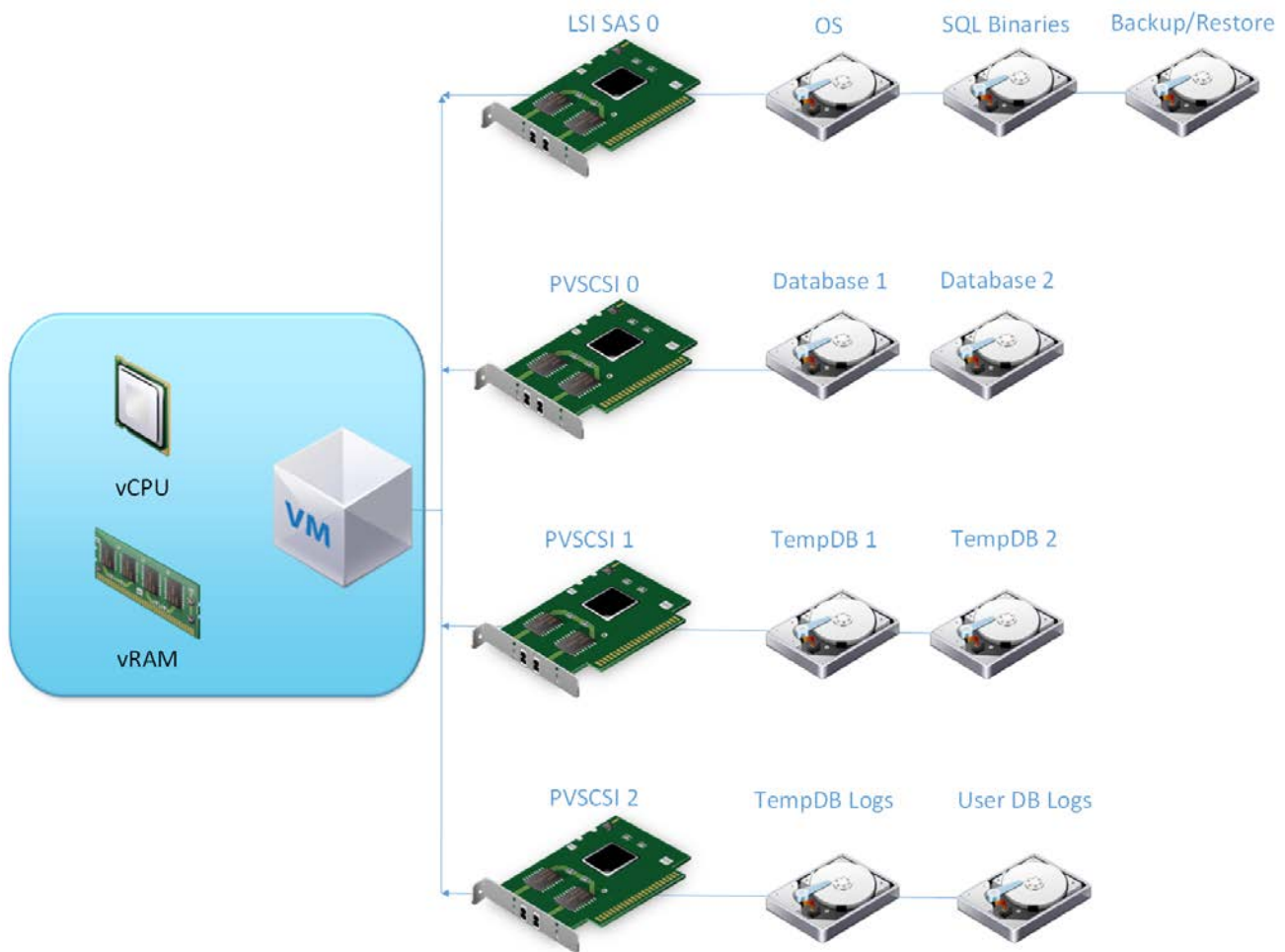
Figure 23: Example controller configuration

5.3.4 SQL Server Files

To achieve high performance, the database should be split into multiple files across multiple virtual disks. In general, one database file per vCPU is ideal. For example a VM with 4 vCPUs hosting 400GB database could be split into four 100GB database files and should spread evenly across the two virtual disks.

For write intensive databases, it is recommended to distribute the database files over four or more virtual disks as it improves the write performance on the back-end and delivers consistent performance. To avoid complexity, the PAGEIOLATCH_XX values should be monitored for contention.

SQL log files (DB and TempDB) are written sequentially, so using multiple log files wouldn't improve the performance. Using a single log file per database is recommended.

TempDB is used as scratch space by the applications and is one of the most important factors of SQL performance. The number of TempDB files to be used is based on the vCPU count. If the vCPUs are less than 8, then configure the same number of TempDB files. If the number of vCPUs is higher than 8, then start with 8 TempDB files and monitor the contention for in-memory allocation (PAGELATCH_XX). The number of TempDB file should be increased in increments of four until contention is eliminated.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

It is recommended to create all TempDBs with the same size and not allow for autogrowth. The TempDB file sizing should be done based on required application and is usually 1-10% of the database size.

5.4 Performance test configuration

This section describes the test configuration for Microsoft SQL Server using a cluster of 4 Lenovo ThinkAgile HX752x appliances. Each appliance has the following configuration:

- 2 x Intel Xeon Scalable 8276 (28 cores @ 2.2 GHz) processors
- 768GB RAM
- 96 x 1.92TB SATA SSDs
- ESXi
- Set UEFI to "Performance Bias" – for more information, see lenovopress.com/lp0780

From Microsoft SQL licensing guide we see that from SQL Server 2012 onwards, there are two enterprise licenses – Core Based and SERVER/CAL Based. As per the Microsoft license document Compute capacity limits by edition of SQL Server, we found that CAL based license limits SQL Enterprise to use only 20 physical processors. Hence, we will perform two tests based on purchased license, one with core based license using 28 vCPUs per SQL Server, and another test with SERVER/CAL license using only 12 vCPUs per SQL Server.

Each host has ESXi and is configured with HA and DRS. The cluster has 8 SQL Server VMs, each configured as follows:

- 28 vCPUs
- 192 GB RAM
- Windows Server Standard Edition
- Microsoft SQ Enterprise Edition

The Nutanix CVM is configured as follows:

- 12 vCPUs
- 96 GB RAM
- CPU affinity 0-55

See also the optimization recommendation in the HammerDB optimization document: hammerdb.com/hammerdb_mssql_oltp_best_practice.pdf.

5.5 Performance test results

This section provides a high-level summary of the results of executing the HammerDB test suite. HammerDB is a graphical open source database load testing and benchmarking tool for Linux and Windows to test databases running on any operating system. HammerDB is automated, multi-threaded and extensible with dynamic scripting support. See this website for more details: hammerdb.com.

This test concentrated on executing 8 VMs on a four node cluster of Lenovo ThinkAgileX752x appliances with two VMs per node. Each cluster has a raw storage of 92TB of SSDs.

A 5,000 scale OLTP database was used for each VM, which equates to about 500GB of database records on the cluster per VM and a total storage of 4TB in the cluster.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Separate load servers were used to simulate the user load. Each OLTP database workload was simulated by running 400 users simultaneously for a total of 3,200 users across the 8 VMs. Each test run had a 5 minute ramp up phase and executed for 15 minutes to simulate 320,000 transactions per user.

Three different scenarios were tested using the ThinkAgile HX752x all flash configuration:

- 12 vCPUs per SQL Server VM
- 28 vCPUs per SQL Server VM
- 44 vCPUs per SQL Server VM

5.5.1 Testing with 12 vCPUs

The first scenario uses 12 vCPUs per VM from the available 56 cores (28 per socket). In this situation a single SQL SERVER/CAL license can be used. It is limited to 20 vCPUs and in this case the physical CPU resources are not overcommitted. Each SQL VM is forced to stay on one vNUMA.

Figure 24 shows the results from this test. The two SQL Server VMs on a single host, one per socket, have a flat profile with a similar transaction rate of 1.8 to 1.9 million transactions per minute.
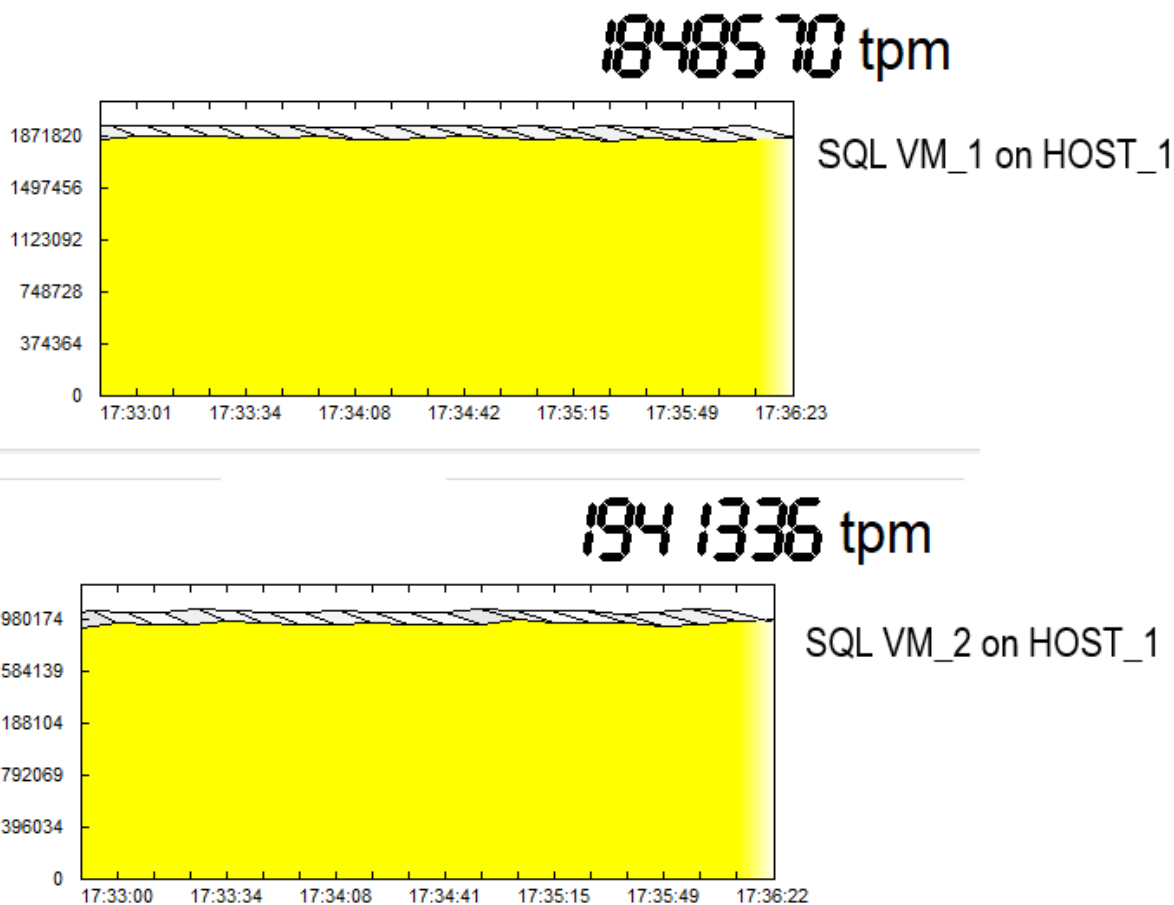

Figure 24: TPM profile obtained during 12 vCPUs tests on two VMs which are situated on same host

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Table 14 shows the transactions per minute and new operations per minute results of the HammerDB test using 2 VMs each with 12vCPUs.

Table 14: Microsoft SQL Server results with cluster of four HX752x appliances and 12 vCPUs

Node	VM Name	HammerDB TPM	HammerDB NOPM
Node 1	SQL Instance - 1	1,634,813	355,652
	SQL Instance - 2	1,730,826	376,263
Node 2	SQL Instance - 3	1,668,535	362,940
	SQL Instance - 4	1,701,626	369,669
Node 3	SQL Instance - 5	1,615,699	351,488
	SQL Instance - 6	1,697,651	369,355
Node 4	SQL Instance - 7	1,603,921	348,847
	SQL Instance - 8	1,703,324	370,336
	Total	13,356,395	2,904,550

The 4 node cluster can execute over 13 million transactions per minute and almost 3 million new orders per minute. The CPU utilization is 90-95%.

5.5.2 Testing with 28 vCPUs

The second scenario uses 28 vCPUs per VM from the available 56 cores (28 per socket). This also requires using a CPU based SQL license as the SQL SERVER/CAL license is limited to 20 vCPUs.

In this scenario one socket has a SQL Server VM and the second socket has both the SQL Server VM and the Nutanix CVM completing for resources. The total number of vCPUs is 28 + 28 + 12 = 68, which exceeds the physical cores. In this case hyper-threading can be used to provide the necessary resources but it is less efficient. Benchmarks and actual user experiences have demonstrated that at most, a 30 percent gain can be obtained when transactions are extremely small, e.g., they execute in under 200 milliseconds. Hence, there are small turbulences in the graph for the second socket as there are some transactions waiting for free threads in the CPU.

Figure 25 shows the results from this test. The transaction rates on the second socket is less than the first one because of the CVM. Nevertheless, the transaction rate is twice that achieved with 12 vCPUs per VM because of the higher number of vCPUs.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**
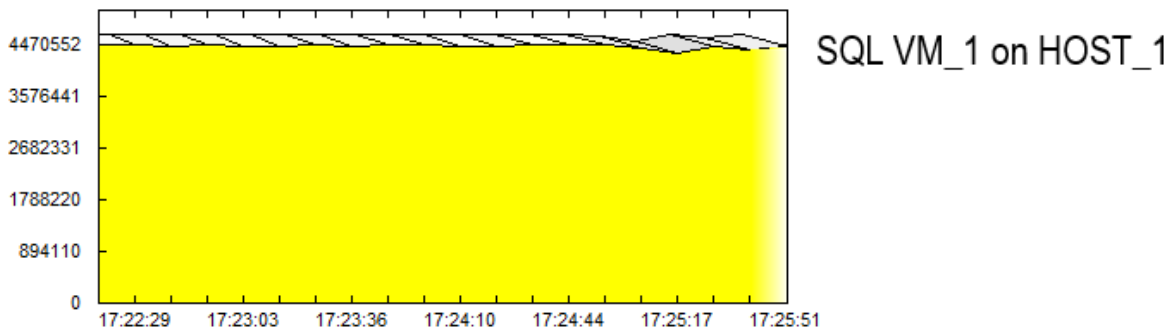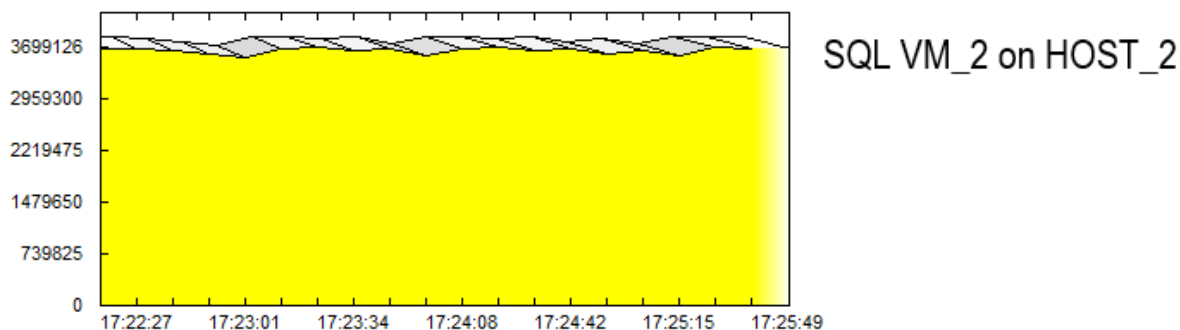
Figure 25: TPM profile obtained during 28 vCPUs tests on two VMs which are situated on same host

Table 15 shows the transactions per minute and new operations per minute results of the HammerDB test using 2 VMs each with 28vCPUs.

Table 15: Microsoft SQL Server results with cluster of four HX752x appliances and 28 vCPUs

Node	VM Name	HammerDB TPM	HammerDB NOPM
Node 1	SQL Instance - 1	3,928,848	854,444
	SQL Instance - 2	3,260,314	708,869
Node 2	SQL Instance - 3	3,863,147	839,716
	SQL Instance - 4	3,091,193	672,483
Node 3	SQL Instance - 5	3,807,745	828,074
	SQL Instance - 6	3,222,624	701,694
Node 4	SQL Instance - 7	3,850,576	837,641
	SQL Instance - 8	3,377,472	734,849
	Total	28,401,919	6,177,770

The 4 node cluster can execute over 28 million transactions per minute and over 6 million new orders per minute. The CPU utilization is 95-99%.

5.5.3 Testing with 44 vCPUs

The third scenario uses 44 vCPUs per VM from the available 56 cores (28 per socket). When hyper-threading is activated, one socket is using 44 vCPUs for the SQL Server VM. The second socket is using 56 vCPUs (44 for the SQL Server VM and 12 for the CVM). However, in this case the threading is not keeping up with transactions.

Figure 26 shows a significant decrease in transaction per minute compared with 28 vCPUs per VM (see Figure 25). The profile curve is also not very flat and consistent in the steady state condition.
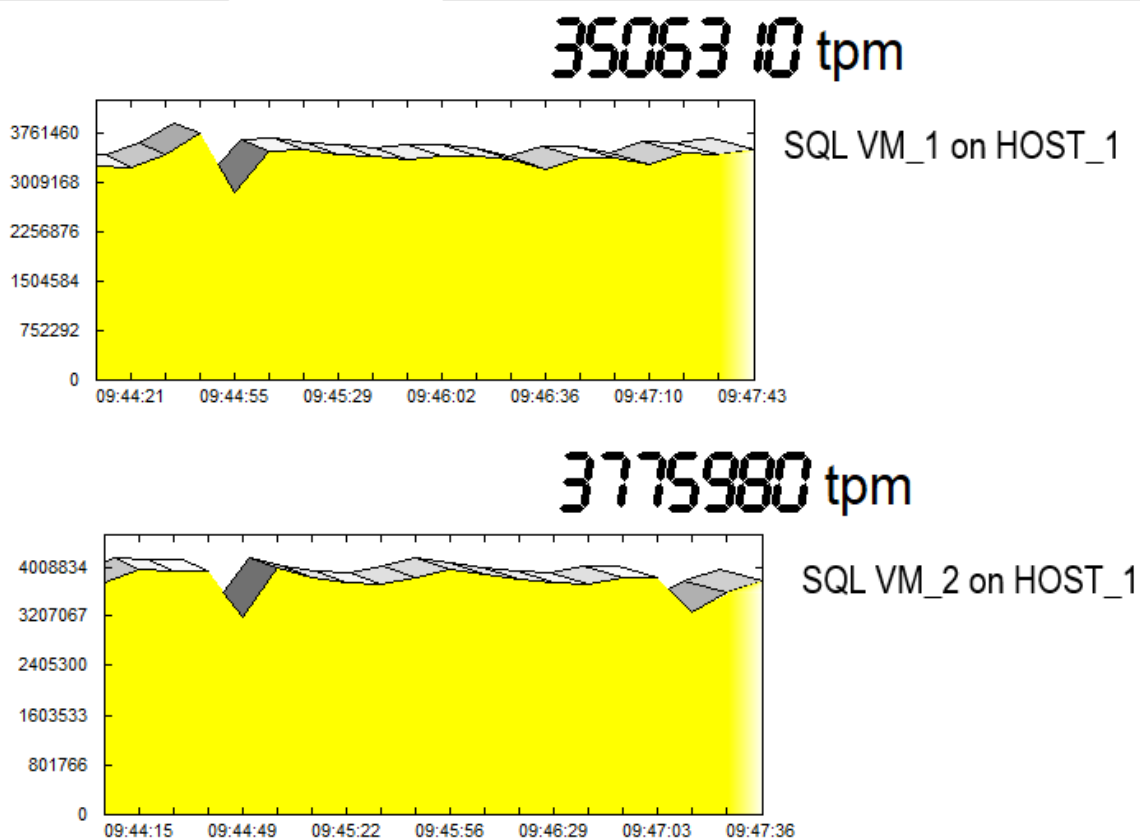

Figure 26: TPM profile obtained during 44 vCPUs tests on two VMs which are situated on same host

5.5.4 Conclusions

The best choice of processor partly depends on which SQL license is purchased. The SQL SERVER/CAL license limits the number of cores to that available from the license. The CPU license does not have this restriction. In either case Lenovo recommends adding 12 more cores per socket for the CVM and ESXi.

Lenovo recommends tying the number of vCPUs to the processor core count. When the number of vCPUs exceeds the core count, hyper-threading is used. Depending on the workload, hyper-threading can increase system performance by up to 30%.However in some cases and when the number of vCPUs is too large, it can increase contention and thereby cause a drop in overall performance.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

6 Red Hat OpenShift Container Platform

The Red Hat OpenShift Container Platform is a complete container application platform that provides all aspects of the application development process in one consistent solution across multiple infrastructure footprints. OpenShift integrates all of the architecture, processes, platforms, and services needed to help development and operations teams traverse traditional siloed structures and produce applications that help businesses succeed.

6.1 Solution Overview

Figure 27 below shows the high level architecture of the Red Hat OpenShift Container Platform and the core building blocks based on Lenovo ThinkAgile HX cluster. OpenShift is a platform designed to orchestrate containerized workloads across a cluster of nodes. The system uses the Kubernetes as the core container orchestration engine, which manages the Docker container images and their lifecycle. Lenovo ThinkAgile HX cluster provides NFS storage and VM running environment. CVM virtualizes all the local storage attached to each node in a cluster and presents it as centralized storage array using Nutanix Distributed File System (NDFS)
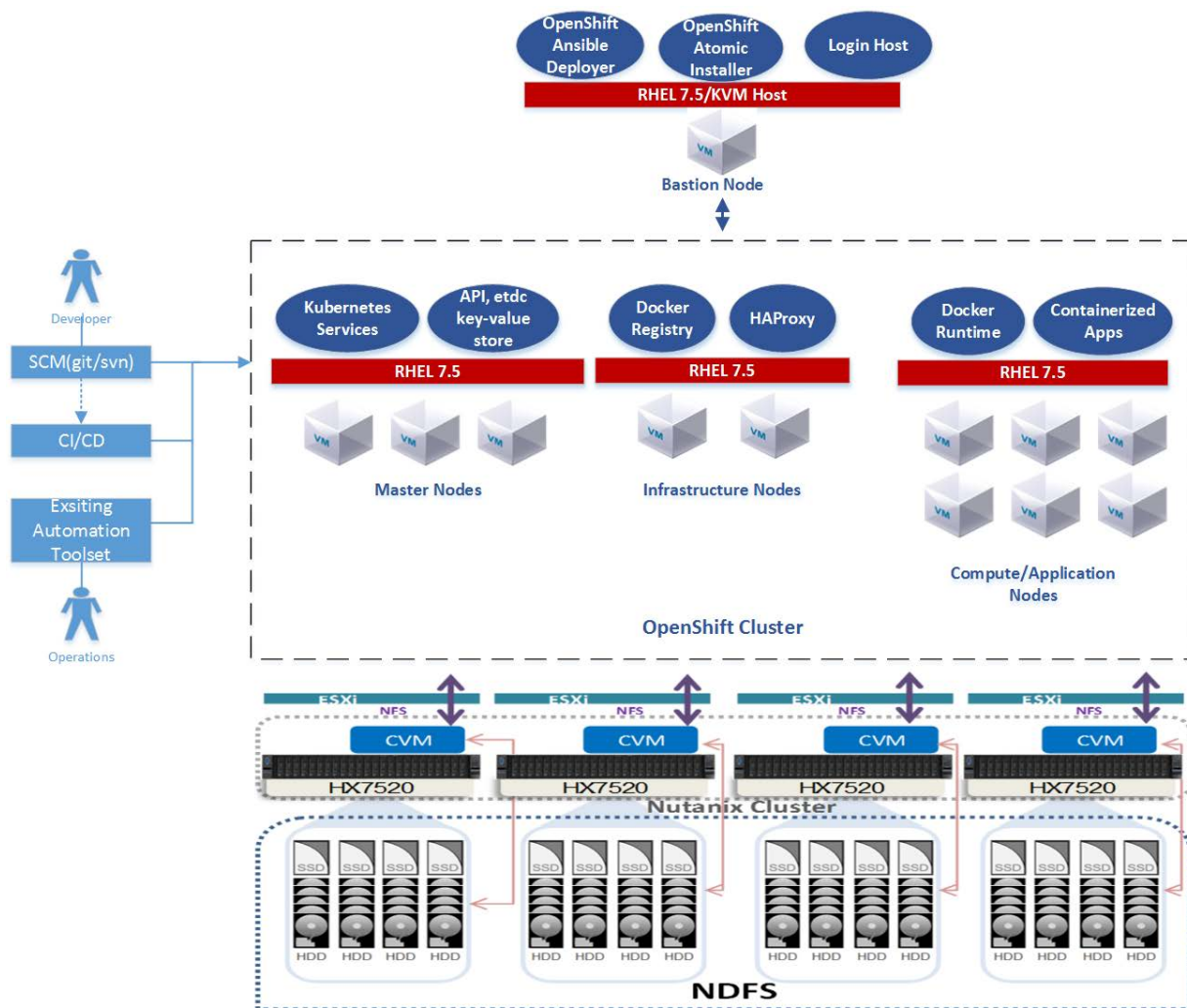

Figure 27. Red Hat OpenShift Container Platform Architecture

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

The VM node configuration of the OpenShift platform is based on the Kubernetes cluster architecture. The master node is the primary node on which the Kubernetes scheduler, along with the distributed cluster data store (etcd), the REST API services, and other associated management services run. In a product environment, you need to ensure high availability of the master services through replicating the services to multiple VM nodes and implementing monitoring and load-balancing services such as Keepalived and HAproxy. The infrastructure nodes can be used in a product setting to implement such services.

Application nodes (or just shown as Node in the diagram) run the users containerized applications on top of the Docker container environment.

6.2 Component Model

As shown in Figure 28, this chapter describes the components and logical architecture of the Red Hat OpenShift solution.
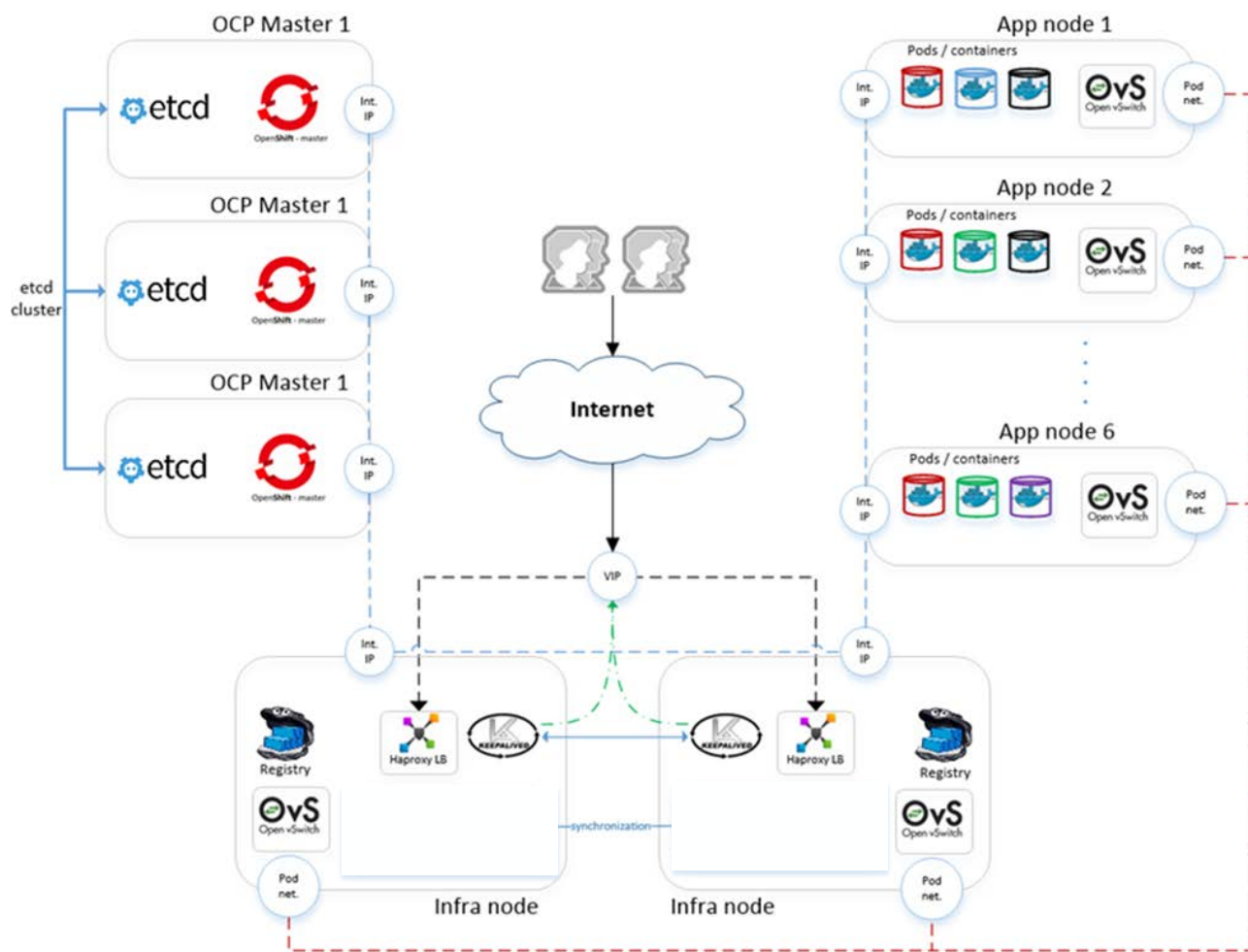
Figure 28. Red Hat OpenShift Container Platform logical architecture

All the OpenShift nodes are connected via the internal network, where they can communicate with each other. Furthermore, Open vSwitch creates its own network for OpenShift pod-to-pod communication. Because of the multi-tenant plugin, Open vSwitch pods can communicate to each other only if they share the same project namespace. There is a virtual IP address managed by Keepalived on two *infrastructure* hosts for external access to the OpenShift web console and applications. Lastly, there is a Red Hat OpenShift Container

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Storage server that shares disk space with Docker Registry for Docker image storage. This storage is backed up by Red Hat OpenShift Container Storage, so Docker Registry storage can be easily switched in case of a node failure.

6.2.1 OpenShift infrastructure components

Figure 29 shows the four types of OpenShift nodes: *bastion*, *master*, *infrastructure,* and *application* in ThinkAgile HX cluster. In this solution, OpenShift platform use NFS storage in ThinkAgile HX cluster to store images and deploy applications. A lightweight DNS server, dnsmasq, is running in OpenShift platform. User can also use Active Directory and DNS server in ThinkAgile HX cluster to provide AD and DNS services when a more stable and large-scale OpenShift is required.
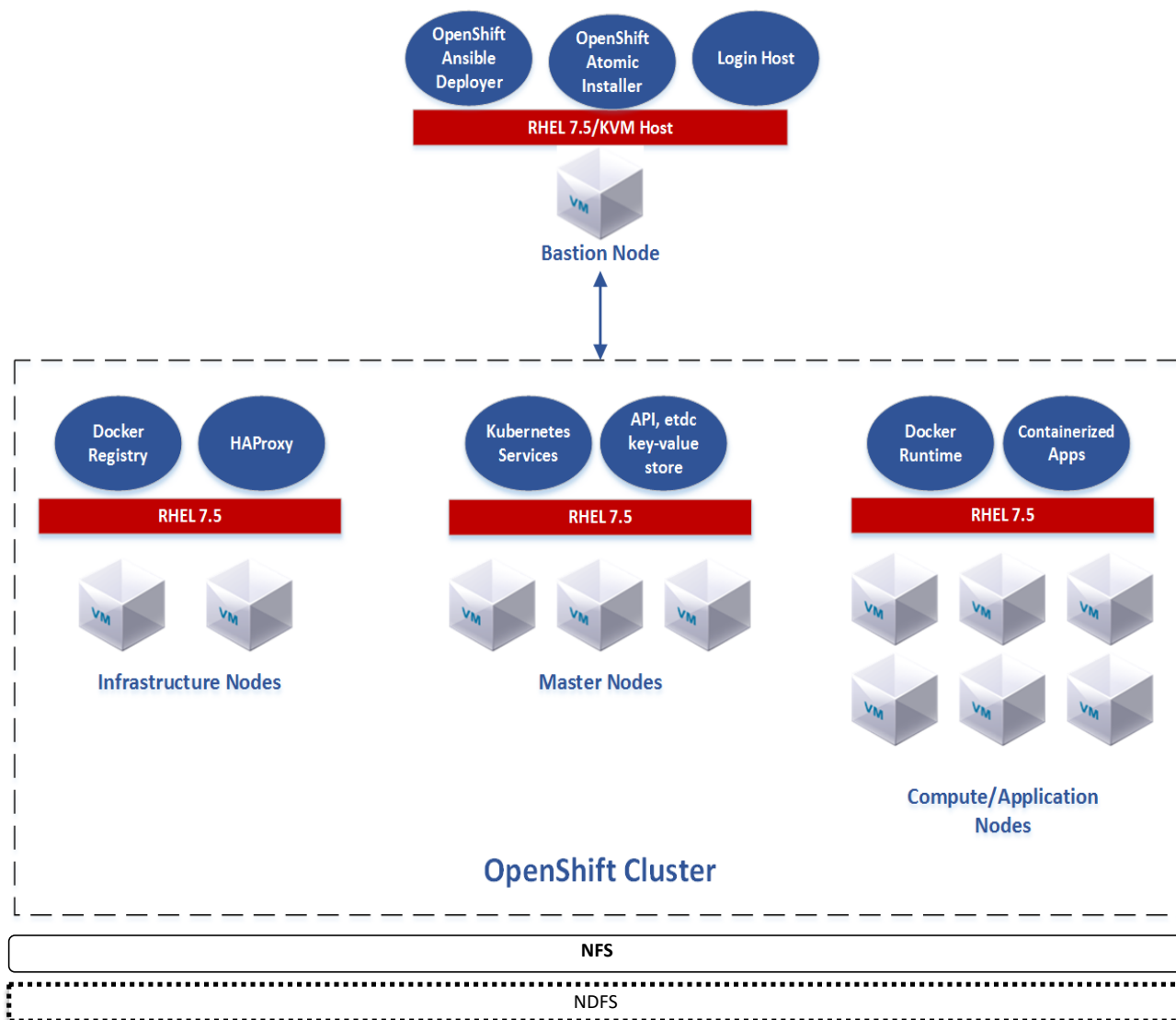

Figure 29. OpenShift Nodes

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Bastion node

This is a dedicated node that serves as the main deployment and management server for the OpenShift cluster. This is used as the logon node for the cluster administrators to perform the system deployment and management operations, such as running the Ansible OpenShift deployment playbooks. The *Bastion* node runs RHEL 7.5 Server with the Linux KVM packages installed.

Master node

The OpenShift Container Platform *master* is a server that performs control functions for the whole cluster environment. It is responsible for the creation, scheduling, and management of all objects specific to OpenShift. It includes API, controller manager, and scheduler capabilities in one OpenShift binary. It is also a common practice to install an etcd key-value store on OpenShift *masters* to achieve a low-latency link between etcd and OpenShift *masters*. It is recommended that you run both OpenShift *masters* and etcd in highly available environments. This can be achieved by running multiple OpenShift *masters* in conjunction with an external active-passive load balancer and the clustering functions of etcd. The OpenShift *master* node runs RHEL 7.5 Server.

Infrastructure node

The OpenShift *infrastructure* node runs infrastructure-specific services such as the Docker Registry and the HAProxy router. The Docker Registry stores application images in the form of containers. The HAProxy router provides routing functions for OpenShift applications. It currently supports HTTP(S) traffic and TLS-enabled traffic via Server Name Indication (SNI). Additional applications and services can be deployed on OpenShift *infrastructure* nodes. The OpenShift *infrastructure* node runs RHEL Server 7.5.

Application node

The OpenShift *application* nodes run containerized applications created and deployed by developers. An OpenShift *application* node contains the OpenShift node components combined into a single binary, which can be used by OpenShift *masters* to schedule and control containers. An OpenShift *application* node runs RHEL Server 7.5.

6.2.2 OpenShift architecture

Kubernetes is an open source project developed by Google. The project gained popularity via its open and flexible architecture for managing containerized workloads at large scale. It provides APIs that can be easily integrated into other tools such as the Red Hat OpenShift Container platform. Kubernetes provides the orchestration capabilities for containers, including scheduling the container images to nodes in a cluster, managing the container life cycle, availability, replication, persistent and non-persistent storage for containers, policy, multi-tenancy, network virtualization, routing, hierarchical clusters via federation APIs, and so forth.

A detailed software description of the Kubernetes components is described on this website: docs.openshift.com/enterprise/3.0/architecture/infrastructure_components/kubernetes_infrastructure.html.

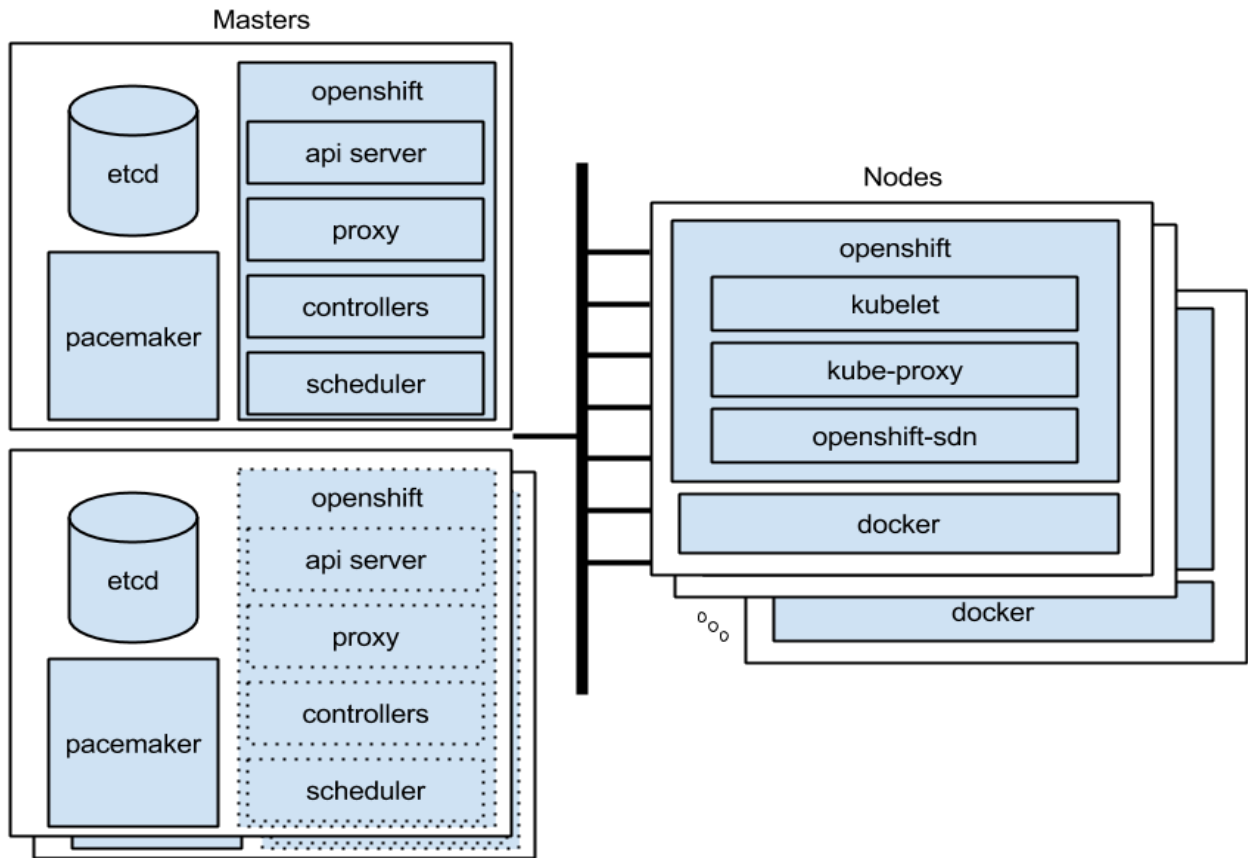Figure 30 shows the OpenShift high-level architecture and components.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Figure 30. OpenShift component architecture

The master nodes, as described previously, are responsible for core services such as API interface, authentication/authorization, container scheduling, controller management, and configuration database. The master manages the state of the cluster and the lifecycle of the user container images. For redundancy and high availability, you can have multiple master nodes with frontend load-balancers such as HAproxy. The command line interface to the master nodes is implemented via the "oc" command.

The (worker) nodes are where users' container images are executed. In OpenShift terminology the worker nodes run "pods", each of which manages one or more running containers. Each node implements a "kubelet", which is the node level controller that manages the pods and interacts with the OpenShift master.

In addition to the core OpenShift services, the Red Hat OpenShift platform also includes other features such as the Web based user self-service console, monitoring, an integrated container registry, storage management, authentication/authorization, automation via built-in Ansible playbooks, and other administrative tools for managing the container platform.

6.3 Operational model

The OpenShift Container Platform can be implemented in development/test, staging, and production settings. Each node role has its own dedicated servers or VMs for performance and availability. However, in a non-production environment, a minimal environment can be provided to test applications before moving them to a staging or production environment.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

For a test/development environment, you can use minishift or 'oc cluster up' to deploy all on one server, or implement OpenShift with five servers for a more formal platform as shown in table below:

Node type	Quantity	Node role
Deployment	1	Deployment of the environment, Ansible playbooks, hardware management, etc.
Master	1	OpenShift API master, Kubernetes scheduler, etcd, other core services
Compute	3	Runs the application containers

For a production OpenShift deployment, all of the core services such as the API servers, Kubernetes scheduler, etcd, etc., need to be highly available. The table below shows the recommended configuration for a production deployment.

Node type	Quantity	Node role
Bastion	1	Deployment of the environment, Ansible playbooks, hardware management, etc.
Infrastructure	2	OpenShift HAProxy, container registry, routing, etcd.
Master	3	OpenShift API master, Kubernetes scheduler
Compute	3+	Runs the application containers
Storage	0	Using NFS shared storage in Nutanix cluster

There are performance and availability implications of running the Red Hat OpenShift Container Storage alongside the workload containers in a hyperconverged environment. For production environments, it is recommended to separate hyper-converged compute servers from storage-only servers, or to ensure that the servers have sufficient CPU, memory, and storage resources to avoid any performance bottlenecks.

6.3.1 Compute servers

The OpenShift Container Platform can be implemented on a small footprint of VMs on Lenovo ThinkAgile HX HX752x cluster. The VMs could be clustered together and scaled as the user workloads grow. Each HX server has the following configuration:

- 2 x Intel Xeon Scalable 8170 (26 cores @ 2.1 GHz) processors
- 768GB RAM
- 4 x 1.92TB SATA SSDs
- 20 x 2TB SATA HDDs
- ESXi 6.5 U1
- Set UEFI to "Performance Bias" – for more information, see lenovopress.com/lp0780

The Nutanix CVM is configured as follows:

- 12 vCPUs
- 96 GB RAM

6.3.2 Persistent storage for containerized workloads

There are two types of storage consumed by containerized applications – ephemeral (non-persistent) and persistent. As the names suggest, non-persistent storage is created and destroyed along with the container and is only used by applications during their lifetime as a container. Hence, non-persistent storage is used for temporary data. When implementing the OpenShift Container Platform, local disk space on the application nodes can be configured and used for the non-persistent storage volumes.

Persistent storage, on the other hand, is used for data that needs to be persisted across container instantiations. An example is a 2 or 3-tier application that has separate containers for the web and business logic tier and the database tier. The web and business logic tier can be scaled out using multiple containers for high availability. The database that is used in the database tier requires persistent storage that is not destroyed.

OpenShift uses a persistent volume framework that operates on two concepts – persistent storage and persistent volume claim. Persistent storage is the physical storage volumes that are created and managed by the OpenShift cluster administrator. When an application container requires persistent storage, it would create a persistent volume claim (PVC). The PVC is a unique pointer/handle to a persistent volume on the physical storage, except that PVC is not bound to a physical volume. When a container makes a PVC request, OpenShift would allocate the physical disk and binds it to the PVC. When the container image is destroyed, the volume bound to the PVC is not destroyed unless you explicitly destroy that volume. In addition, during the lifecycle of the container if it relocates to another physical server in the cluster, the PVC binding will still be maintained. After the container image is destroyed, the PVC is released, but the persisted storage volume is not deleted. The specific persistent storage policy for the volume will determine when the volume gets deleted.

For more detailed conceptual information on persistent volumes see: access.redhat.com/documentation/en-us/openshift_container_platform/3.9/html-single/architecture/#architecture-additional-concepts-storage

A variety of persistent storage options are available for OpenShift, choices including NFS, OpenStack Cinder, Ceph RBD, iSCSI, fiber channel SAN, hyperconverged storage using Red Hat OpenShift Container Storage, AWS elastic block storage (EBS), and others. For a complete list of these choices and the corresponding requirements, see the link below: access.redhat.com/documentation/en-us/openshift_container_platform/3.9/html-single/installation_and_configuration/#configuring-persistent-storage

In the OpenShift platform, persistent storage use NFS storage in the ThinkAgile HX cluster as backend. Lenovo recommends configuration the following persistent storage:

- 500 GB NFS storage for Docker registry
- 5 TB NFS storage for other applications

6.3.3 Networking

For OpenShift Container Platform deployment in Lenovo ThinkAgile HX752x cluster, VMware virtual infrastructure networking is recommended as the choice for all cluster-wide communication for the core OpenShift services, application container workloads, storage services access with NFS, as well as all east-west traffic and north-south traffic across the container workloads.

There are three logical networks:

- **External Network**: The external network is used for the public API, the OpenShift web interface, and exposed applications (services and routes).
- **Internal Network**: This is the primary, non-routable network used for cluster management and inter-node communication. The same network acts as the layer for server provisioning using PXE and HTTP. Domain Name Servers (DNS) and Dynamic Host Configuration Protocol (DHCP) services also reside on this network to provide the functionality necessary for the deployment process and the cluster to work. Communication with the Internet is provided by NAT configured on the *bastion* node.
- **Storage Network**: This is isolated network used for connection between OpenShift cluster and shared NFS storage created in Lenovo ThinkAgile HX752x cluster.

Figure 31 shows the Red Hat OpenShift servers and the recommended network architecture.
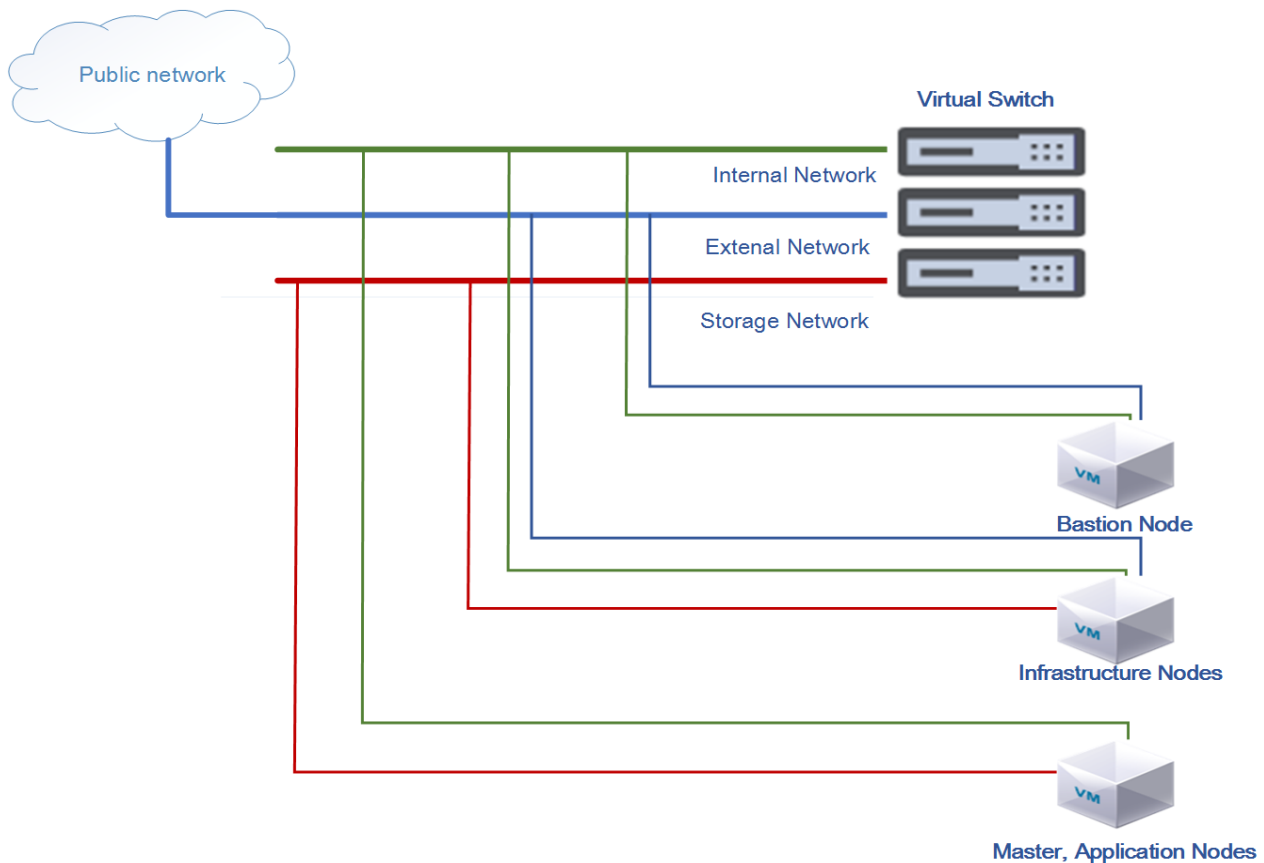
Figure 31. OpenShift Network Connectivity

All OpenShift nodes are connected via the internal network, where they can communicate with each other. Furthermore, Open vSwitch creates its own network for OpenShift pod-to-pod communication. Because of the multi-tenant plugin, Open vSwitch pods can communicate to each other only if they share the same project namespace. There is a virtual IP address managed by Keepalived on two *infrastructure* hosts for external access to the OpenShift web console and applications.

6.4 Best practices

This section describes recommended the best practices for implementing Red Hat OpenShift Container platform in the ThinkAgile HX Nutanix environment.

6.4.1 Shared NFS storage as persistent storage

A Nutanix storage container is a defined subset of available storage within a storage pool. Storage containers allow users to apply rules or transformations such as compression to a data set.

By default all Nutanix storage containers are thin provisioned which reduces unused capacity and automatically provisions additional storage capacity when needed. It is also very easy to add additional storage capacity for databases by simply adding nodes to the cluster. It is also possible to set a storage reservation amount to guarantee a minimum amount of storage capacity.

Data compression can be used to further increase data capacity especially for data that is less frequently accessed. Lenovo recommends disabling compression for the Red Hat OpenShift Container platform when supporting multiple types of applications.

A resiliency factor of 2 is the default. This provides a minimum level of data redundancy but a resiliency factor of 3 might be important in some environments. Using erasure coding saves significant storage capacity but it only recommended for archive data.

Nutanix virtual storage using shared NFS storage should be configured on each OpenShift VM node. Users can create volume manually on shared NFS and configure it in container configuration files. Users can also use third-party tools to utilize shared NFS by PVC (persistent volume claim).

Nutanix shared NFS storage can support most types of applications services. However, for some no-SQL databases such as MongoDB and Elastic Search, it is recommended to use other types of storage such as fiber channel SAN, Red Hat OpenShift Container Storage for the persistent storage backend.

6.4.2 Cluster high availability

The minimum number of nodes in each cluster is 3 and should be at least 4 to provide failover. The following high availability features are recommended for an ESXi-based ThinkAgile HX cluster:

- VMware vSphere high availability (HA) for failover
- VMware vSphere distributed resource scheduler (DRS) for load balancing

VMware vSphere HA pools VMs into a cluster to increase data resiliency. If a host fails, VMware HA moves the VMs to other hosts with spare capacity. Lenovo recommends enabling the "Admission Control Setting" and using the "Admission Control Policy" to set the percentage of cluster resources reserved as failover spare capacity.

VMware vSphere DRS can be used to group ESXi hosts into resource clusters to provide highly available resources and balance workloads. In order to keep the active working set for each OpenShift application VM to the ThinkAgile HX node, Lenovo recommends creating a host group for each node and a "should" rule that keeps each OpenShift application VM on a 1 to 1 ratio with the ESXi host. The hosts should be configured with a minimum of N+1 availability. Other VM nodes (bastion node, master node, infrastructure node) are recommended to run on non-DRS hosts to allow OpenShift to manage the load balancing for these nodes.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

6.5 Deployment example

This deployment example was fully tested and verified by Lenovo. All of the deployment files are available from GitHub at the following location: https://github.com/lenovo/deployment-scripts-for-OpenShift-Platform/tree/master/OCP_HX.

The example uses 10 nodes as follows:

- 1 Bastion node
- 3 Master nodes
- 2 Infrastructure nodes
- 4 Application nodes with Red Hat OpenShift Container Storage to provide hyper-converged compute

This configuration represents a production grade OpenShift implementation that meets high-availability, redundancy, and scale requirements for enterprises. Additional Application nodes can be added to increase the available compute and storage capacity.

Table 16 provides a VM configuration summary for this example deployment on Lenovo ThinkAgile HX752x cluster.

Table 16. VM Configuration for each OpenShift Node

Node Type	CPU	Memory	Hard Disk	Network Adapter
Bastion	2 vCPU(s)	20 GB	100 GB	2 VMXNET3 NIC
Master	4 vCPU(s)	40 GB	200 GB	2 VMXNET3 NIC
Infrastructure	4 vCPU(s)	100 GB	300 GB	3 VMXNET3 NIC
Application	4 vCPU(s)	100 GB	300 GB	2 VMXNET3 NIC

6.5.1 Ansible inventory file

In order to install the OpenShift Container Platform cluster, an Ansible inventory file has to be created with the environment's description. The rest of this section explains the inventory file and provides the additional variables that are used for the automatic prerequisites and a Keepalived deployment.

The Ansible section below specifies the types of nodes that are used in an OpenShift Container Platform environment. Required groups are nodes, masters, and etcd. Optional groups are lb (for load balancing in multi-master clusters), local (which specifies the *Bastion* node).

```
[OSEv3:children]
masters
nodes
etcd
lb
local
```

The Ansible section below describes global cluster parameters. Parameter openshift_master_cluster_method specifies the load balancing method in a multi-master environment. With the native value, there will be a separated HAProxy load balancer installed on the specified host and configured for the whole environment.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

The hostname for users and cluster components to access the cluster load balancer from external and internal networks is set in the **openshift_master_cluster_hostname** and **openshift_master_cluster_public_hostname** parameters. The parameter **openshift_master_identity_providers** configures the way for authentication of OpenShift users. In this example, this parameter is based on htpasswd files stored in the OpenShift configuration directory. However, many other authentication methods can be used such as LDAP, Keystone*, or GitHub* accounts. **os_sdn_network_plugin_name** specifies the SDN Open vSwitch plugin used in environment. For this example, **redhat/openshift-ovs-multitenant** provides isolation between OpenShift projects on the network level. In this solution, Docker Registry uses Nutanix shared NFS storage for Docker image storage.

```
[OSEv3:vars]
ansible_ssh_user=root
ansible_become=true
containerized=true
openshift_master_cluster_method=native
openshift_master_cluster_hostname=lb.ocp.example.local
openshift_master_cluster_public_hostname=openshift.ocp.example.com
openshift_master_default_subdomain=apps.ocp.example.com
openshift_master_cluster_ip=172.29.7.30
openshift_master_cluster_public_ip=10.240.39.98
openshift_master_portal_net=10.0.0.0/16
openshift_deployment_type=openshift-enterprise
openshift_release=v3.9
os_sdn_network_plugin_name='redhat/openshift-ovs-multitenant'

# Allow all auth
openshift_master_identity_providers=[{'name': 'allow_all', 'login': 'true',
'challenge': 'true', 'kind': 'AllowAllPasswordIdentityProvider'}]

rhel_subscription_user=??????
rhel_subscription_pass=******

openshift_enable_unsupported_configurations=True
openshift_hosted_registry_storage_kind=nfs
openshift_hosted_registry_storage_access_modes=['ReadWriteMany']
openshift_hosted_registry_storage_host=192.168.5.2
openshift_hosted_registry_storage_nfs_directory=/RedHat
openshift_hosted_registry_storage_volume_name=registry
openshift_hosted_registry_storage_volume_size=200Gi

external_interface=ens192
external_netmask=255.255.252.0
external_gateway=10.240.36.1
external_dns=10.240.0.10
```

```
internal_netmask=255.255.0.0
bastion_ip=172.29.7.1
internal_gateway={{ bastion_ip }}
internal_interface=ens224
local_dns={{ bastion_ip }}
dhcp_first_ip=172.29.7.100
dhcp_last_ip=172.29.7.150
root_password=Passw0rd


openshift_install_examples=true


openshift_master_api_port=8443
openshift_master_console_port=8443


openshift_disable_check=docker_image_availability
```

The Ansible section below describes which servers act as OpenShift masters. For this example, three OpenShift masters are implemented for control plane HA purposes. OpenShift master components can be installed with two methods: rpm-based or container-based. In this RA, all OpenShift components are implemented as containers, which is determined by the **containerized=True** parameter.

```
[masters]
master1.ocp.example.local containerized=true openshift_ip=172.29.7.2
openshift_hostname=master1.ocp.example.local
master2.ocp.example.local containerized=true openshift_ip=172.29.7.3
openshift_hostname=master2.ocp.example.local
master3.ocp.example.local containerized=true openshift_ip=172.29.7.4
openshift_hostname=master3.ocp.example.local
```

The Ansible section below describes which servers act as OpenShift nodes. For this example, seven OpenShift nodes are implemented. Two of them perform infrastructure functions, which is determined by the openshift_node_labels="{'region': 'infra'}" parameter. OpenShift node components are also installed on OpenShift master servers. User application could be deployed on these servers when configuring parameter openshift_schedulable=true. In this RA, all node components are implemented as containers, which is determined by the **containerized=true** parameter.

```
[nodes]
master1.ocp.example.local  containerized=true openshift_ip=172.29.7.2
openshift_hostname=master1.ocp.example.local
master2.ocp.example.local  containerized=true openshift_ip=172.29.7.3
openshift_hostname=master2.ocp.example.local
master3.ocp.example.local  containerized=true openshift_ip=172.29.7.4
openshift_hostname=master3.ocp.example.local
infra2.ocp.example.local openshift_node_labels="{'region': 'infra'}"
containerized=true  openshift_ip=172.29.7.6
openshift_hostname=infra2.ocp.example.local
```

```
infra1.ocp.example.local openshift_node_labels="{'region': 'infra'}"
containerized=true  openshift_ip=172.29.7.5
openshift_hostname=infra1.ocp.example.local
app1.ocp.example.local openshift_schedulable=true containerized=true
openshift_ip=172.29.7.7 openshift_hostname=app1.ocp.example.local
openshift_hostname_check=false
app2.ocp.example.local openshift_schedulable=true containerized=true
openshift_ip=172.29.7.8 openshift_hostname=app2.ocp.example.local
app3.ocp.example.local openshift_schedulable=true containerized=true
openshift_ip=172.29.7.9 openshift_hostname=app3.ocp.example.local
app4.ocp.example.local openshift_schedulable=true containerized=true
openshift_ip=172.29.7.10 openshift_hostname=app4.ocp.example.local
```

The Ansible section below describes hosts that will run etcd instances. For this example, three *etcd* instances are installed on three *master* servers to achieve low-latency traffic between them. When many etcd instances are specified in an inventory file, they are automatically clustered in order to provide a highly available key-value etcd store. An etcd cluster that consists of three etcd instances resists a failure of one etcd instance. It is also recommended to have an odd number of etcd instances in a cluster.

```
[etcd]
master1.ocp.example.local containerized=true openshift_ip=172.29.7.2
openshift_hostname=master1.ocp.example.local
master2.ocp.example.local containerized=true openshift_ip=172.29.7.3
openshift_hostname=master2.ocp.example.local
master3.ocp.example.local containerized=true openshift_ip=172.29.7.4
openshift_hostname=master3.ocp.example.local
```

When **openshift_master_cluster_method** is set to **native**, then the Ansible section below specifies a host on which HAProxy load balancer will be installed and configured. For this example, two HAProxy load balancers are installed on two infrastructure servers. They use one common virtual IP address that is managed by Keepalived software to achieve a highly available OpenShift Container Platform cluster.

```
[lb]
infra1.ocp.example.local openshift_hostname=infra1.ocp.example.local
openshift_hostname_check=false openshift_ip=172.29.7.5
openshift_public_ip=10.240.39.91
infra2.ocp.example.local openshift_hostname=infra2.ocp.example.local
openshift_hostname_check=false openshift_ip=172.29.7.6
openshift_public_ip=10.240.39.92
```

6.5.2 Software

For this example, the following software is needed:

- **OpenShift Container Platform**, which adds developer- and operation-centric tools to enable rapid application development, easy deployment, scaling, and long-term lifecycle maintenance for small and large teams and applications

In addition, the OpenShift Container Platform requires the following software packages:

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

- **Docker** to build, ship, and run containerized applications
- **Kubernetes** to orchestrate and manage containerized applications
- **Etcd***, which is a key-value store for the OpenShift Container Platform cluster
- **Open vSwitch*** to provide software-defined networking (SDN)-specific functions in the OpenShift Container Platform environment
- **Ansible®** for installation and management of the OpenShift Container Platform deployment
- **HAProxy*** for routing and load-balancing purposes
- **Keepalived*** for virtual IP management for HAProxy instances

Table 17 lists the software versions used for this example deployment

Table 17. Software versions

Component	Version
Red Hat Enterprise Linux	7.5
OpenShift Container Platform	3.9
Docker	1.13.1
Ansible	2.4.6.0
rhel7/etcd	latest
openshift3/openvswitch	v3.9.43
openshift3/ose	v3.9.43
openshift3/ose-pod	v3.9.43
openshift3/node	v3.9.43
openshift3/ose-haproxy-router	v3.9.43
openshift3/keepalived	1.0_ra

Each VM node is installed with RHEL 7.5 as the base operating system and the following resources:

- Minimum 40 GB hard disk space for the file system containing `/var/`.
- Minimum 1 GB hard disk space for the file system containing `/usr/local/bin/`.
- Minimum 1 GB hard disk space for the file system containing the system's temporary directory

6.5.3 Networking

The Lenovo ThinkAgile HX cluster provides virtual networks for user VMs. Table 18 shows the network configuration used in OpenShift cluster (as shown in Figure 31).

Table 18. Network configuration for OpenShift Cluster

Node Type	External Network	Internal Network	Storage Network
Bastion	VMXNET 3	VMXNET 3	
Master		VMXNET 3	VMXNET 3
Infrastructure	VMXNET 3	VMXNET 3	VMXNET 3
Application		VMXNET 3	VMXNET 3

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

6.5.4 Automatic prerequisites installation

After the operating system has been installed and configured, the nodes need to be prepared for OpenShift installation. Perform the following preliminary steps: prepare an *openshift* account and exchange SSH keys across all nodes, attach software licenses, install and configure the DNS service, install additional packages, and configure Docker Engine.

All the remaining tasks can be executed automatically using the Ansible playbooks available at https://github.com/lenovo/deployment-scripts-for-OpenShift-Platform/tree/master/OCP_HX.

Based on information from the operating system deployment, prepare an Ansible inventory file and place it in the location /etc/ansible/hosts on the *Bastion* node. The hosts file for this deployment example is named hosts.example and is provided in the GitHub repository.

After that, clone the git repository onto the *Bastion* node:

```
$ git clone https://github.com/lenovo/deployment-scripts-for-OpenShift-Platform.git
$ cd deployment-scripts-for-OpenShift-Platform/OCP_HX/src/prerequisites
```

In the inventory file, set up the following additional variables:

rhel_subscription_user:	Name of the user who will be used for registration
rhel_subscription_pass:	Password of the user who will be used for registration
ansible_ssh_user:	Insert root or other user with root privileges
ansible_become:	Set to True to run commands with sudo privileges
local_dns:	Type a proper IP address for your bastion node that runs the DNS service

Finally start the Ansible playbook by entering the following command:

```
$ ansible-playbook nodes_setup.yaml -k
```

6.5.5 Automatic Keepalived deployment

The OpenShift Container Platform delivers two flavors of HAProxy load balancing software. The first flavor, which is spawned as a daemon, distributes API calls between *master* servers. The second flavor, spawned as a Docker container, provides the *router* mechanism for exposing applications inside a cluster.

To achieve high availability (HA), maximum fault tolerance, and performance, this deployment example includes an additional package called Keepalived. It is open-source software distributed under the GPL license and is recognized by Red Hat as their recommended solution. Please see the following web site for more information: Red Hat Enterprise Linux documentation.

This example deployment uses both flavors of HAProxy instances, which are installed on both Infrastructure nodes. A single point of failure is eliminated is eliminated when used in conjunction with floating IP addresses provided by Keepalived,

Installation and configuration of this HA Solution can be performed manually or through a single command using an Ansible playbook. First the following variables must be defined in the Ansible inventory:

```
external_interface=enp6s0f0
external_netmask=255.255.250.128
external_gateway=10.240.202.129
```

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

```
external_vlan= 302
external_dns=114.114.114.114
internal_interface=bond0
internal_netmask=255.255.0.0
openshift_master_cluster_ip=172.30.4.30
openshift_master_cluster_public_ip=10.240.202.188
```

To deploy the Keepalived daemons using an Ansible playbook on Infrastructure nodes, enter following command inside the cloned GitHub repository:

```
$ su openshift
$ ansible-playbook \
    openshift-container-architecture/OCP_HX/src/keepalived-multimaster/keepalived.yaml
```

6.5.6 OpenShift container platform installation

When the inventory file with the environment description is prepared and all prerequisites are configured, the OpenShift Container Platform install can be performed from the *Bastion* node.

For this deployment example, a containerized version of the OpenShift Container Platform is installed on servers. This installer image provides the same functionality as the RPM-based installer, but it runs in a containerized environment that provides all of its dependencies rather than being installed directly on the node. The only requirement to use it is the ability to run a container, and atomic packages are installed. The installer image can be used as a system container. System containers are stored and run outside of the traditional Docker service. This enables running the installer image from one of the target hosts without concern for the install restarting Docker on the host.

This install process is straightforward, and requires three steps:

First, Create shared storage:

```
Create a NFS storage container (named Redhat and granted ReadWriteMany access modes)
in Nutanix cluster. And Create a directory (named registry and granted ReadWriteMany
access modes) in storage container.
```

Second, execute the prerequisites:

```
$ atomic install --system --set INVENTORY_FILE=/etc/ansible/hosts --storage=ostree
--set PLAYBOOK_FILE=/usr/share/ansible/openshift-ansible/playbooks/prerequisites.yml
--set OPTS="-v" registry.access.redhat.com/openshift3/ose-ansible:v3.9
```

Third, deploy OpenShift platform：

```
$ atomic install --system --storage=ostree  --set INVENTORY_FILE=/etc/ansible/hosts
--set PLAYBOOK_FILE=/usr/share/ansible/openshift-ansible/playbooks/deploy_cluster.yml
--set OPTS="-v" registry.access.redhat.com/openshift3/ose-ansible:v3.9
```

After the installation process, the Ansible playbook should report no errors and the OpenShift Container Platform environment is set up. If needed, you can easily uninstall the environment with the following command:

```
$ ansible-playbook -i /etc/ansible/hosts /usr/share/ansible/openshift-
ansible/playbooks/adhoc/uninstall.yml
```

Next user credentials should be created using the following commands:

```
$ sudo yum install httpd-tools
$ touch users.htpasswd
$ htpasswd -n <user_name> >> users.htpasswd
```

Use the **htpasswd** command for each user account and propagate the users.htpasswd file to every OpenShift *master* node, into the **/etc/origin/master/** directory.

Lastly restart the API services on each *master* node using the command:

```
$ sudo systemctl restart atomic-openshift-master-api
```

6.5.7 Deployment validation

The deployment should be validated before it is used.

First, log on to one of the OpenShift *master* nodes and check that all nodes are connected to the cluster using the commands:

```
$ ssh master1.ocp.example.local
$ oc get nodes
```

Here is some example output from the command:

```
NAME                         STATUS  ROLES    AGE    VERSION
app1.ocp.example.local       Ready   compute  26d    v1.9.1+a0ce1bc657
app2.ocp.example.local       Ready   compute  26d    v1.9.1+a0ce1bc657
app3.ocp.example.local       Ready   compute  26d    v1.9.1+a0ce1bc657
app4.ocp.example.local       Ready   compute  26d    v1.9.1+a0ce1bc657
infra1.ocp.example.local     Ready   <none>   26d    v1.9.1+a0ce1bc657
infra2.ocp.example.local     Ready   <none>   26d    v1.9.1+a0ce1bc657
master1.ocp.example.local    Ready   master   26d    v1.9.1+a0ce1bc657
master2.ocp.example.local    Ready   master   26d    v1.9.1+a0ce1bc657
master3.ocp.example.local    Ready   master   26d    v1.9.1+a0ce1bc657
```

All cluster nodes should be listed and marked as **Ready**. If any node is in a **NotReady** state then it is not properly assigned to a cluster and should be inspected using the following command to verify the etcd state:

```
$ sudo etcdctl -C https://etcd1.ocp.example.local:2379 --ca-file=/etc/etcd/ca.crt --
cert-file=/etc/etcd/peer.crt --key-file=/etc/etcd/peer.key cluster-health
```

Here is some example output from the command:

```
member 5f0aab880290ddeb is healthy: got healthy result from
https://etcd1.ocp.example.local:2379
member c305190f3c57613c is healthy: got healthy result from
https://etcd2.ocp.example.local:2379
```

```
member c434590bbf158f3d is healthy: got healthy result from
https://etcd3.ocp.example.local:2379
```

All etcd members should be listed and marked as **healthy**. If any etcd member is in an **unhealthy** state then it is not properly assigned to an etcd cluster.

To further inspect and verify all the components of the OpenShift Container Platform cluster, use the command:

```
$ oc get all
```

All the listed items should have a **Running** status.

At a final verification step, log on to the OpenShift Container Platform web console using the following URL address: https://openshift.ocp.example.com:8443 and display the OpenShift container catalog. Figure 32 shows an example.
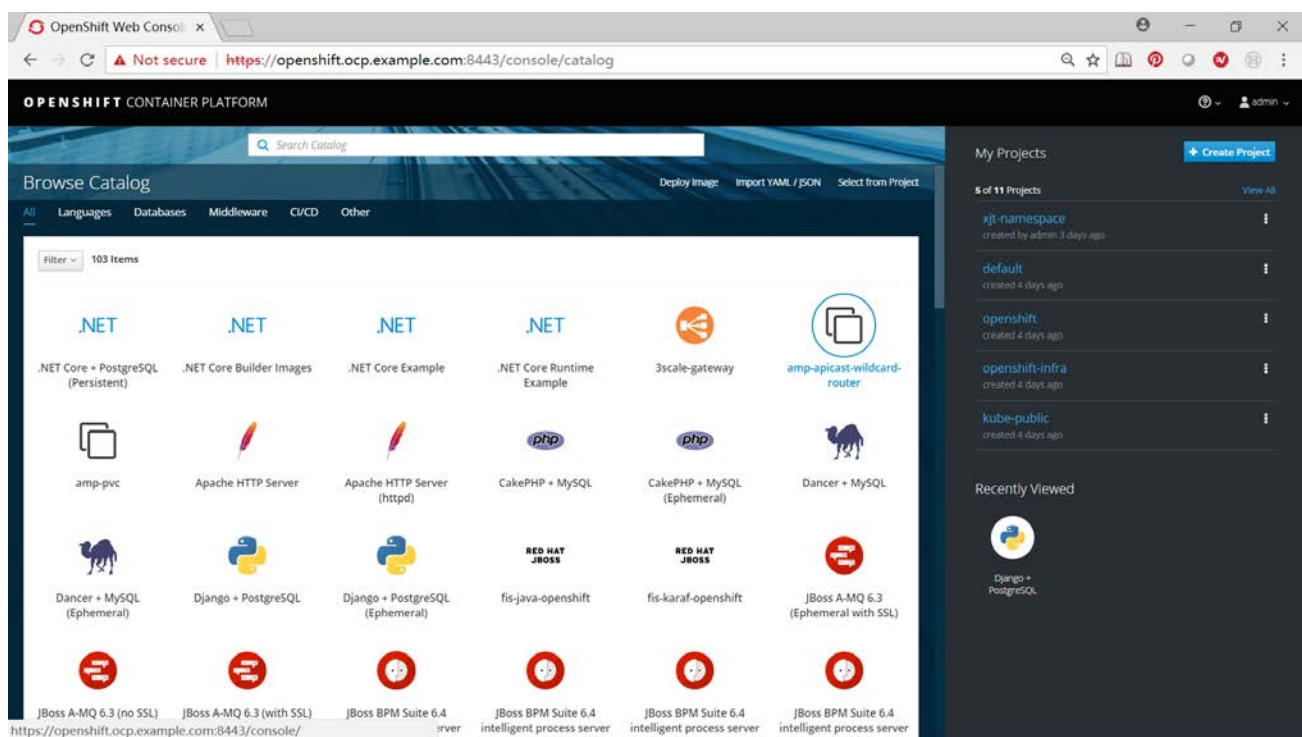
Figure 32 Example Catalog for OpenShift Container Platform

Figure 33 shows other OpenShift application components.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**
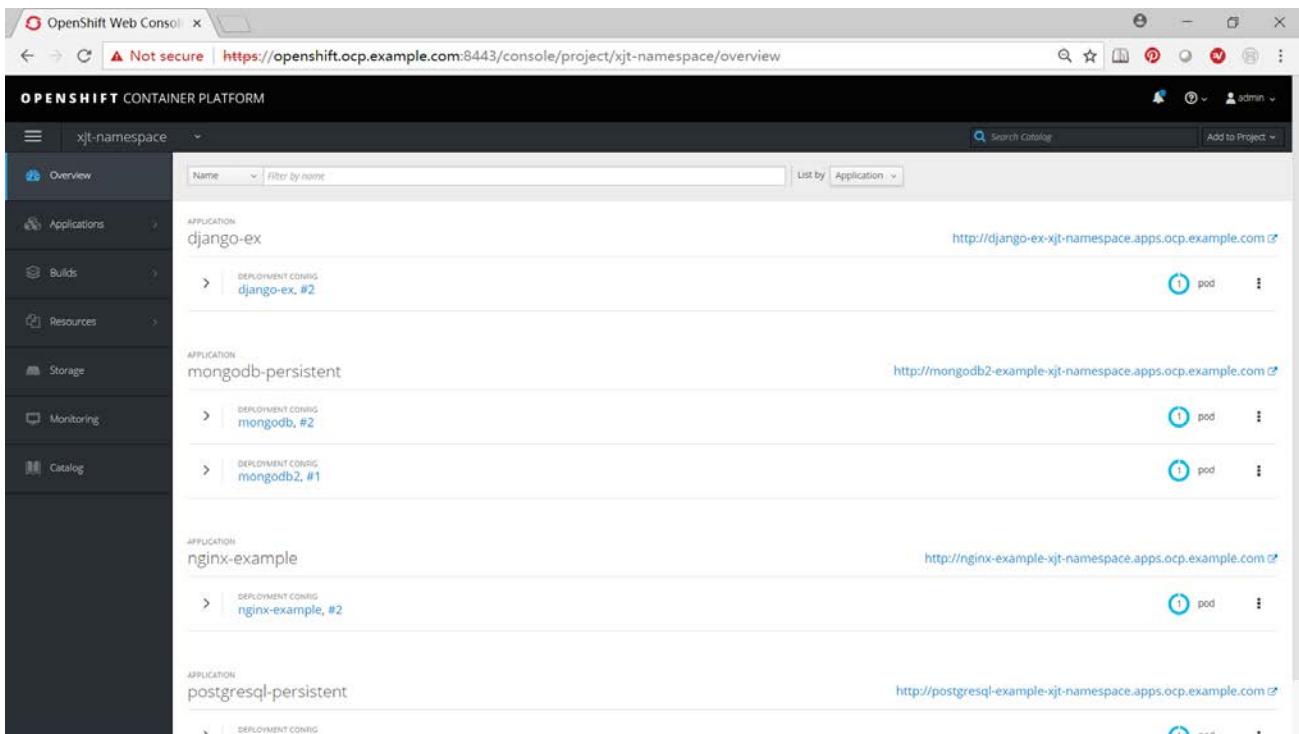
Figure 33 OpenShift Application Components

Figure 34 shows storage status.

Figure 34 OpenShift Storage Status

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**
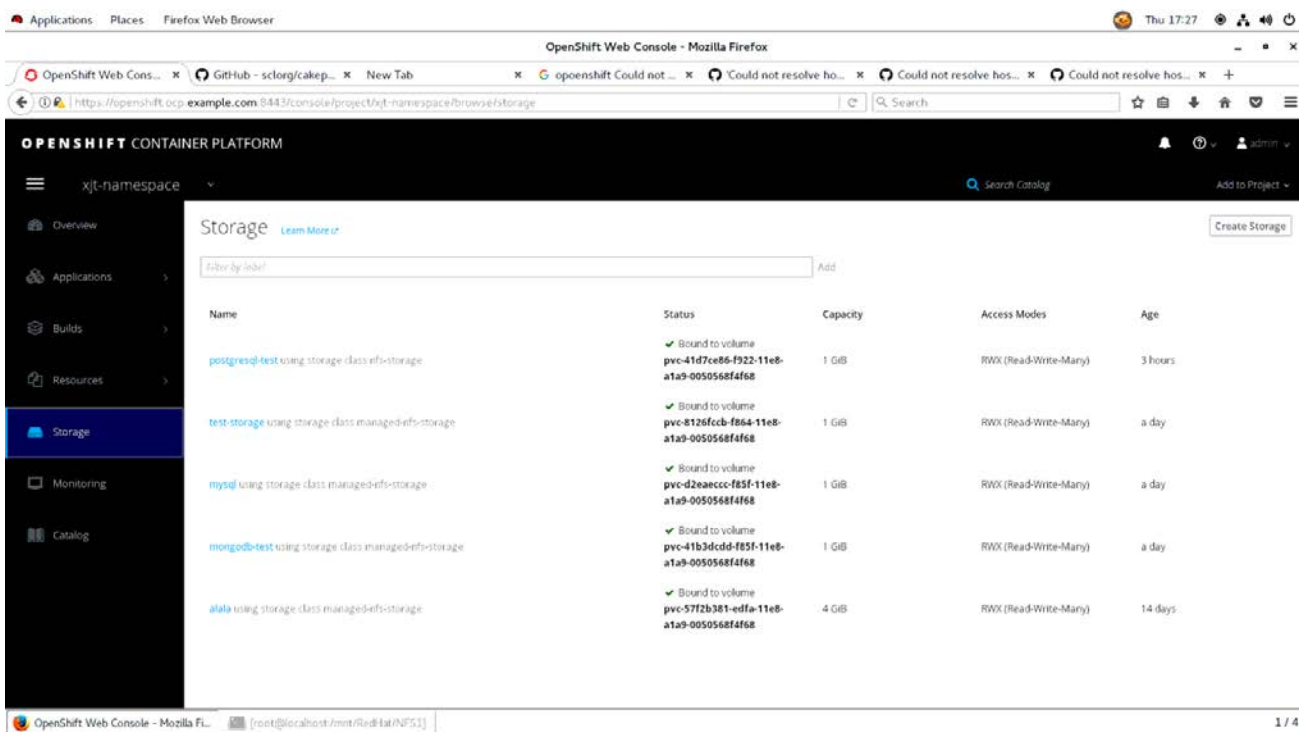
7 VMware Horizon

Horizon View is a desktop virtualization product developed by VMware Inc. It provides remote desktop capabilities by using VMware virtualization technology and can deliver a consistent user experience across devices and locations while keeping corporate data secure and compliant. See this website for more details: vmware.com/products/horizon-view.

7.1 Solution overview

Figure 35 shows all of the main features of the Lenovo Hyper-converged Nutanix solution's reference architecture with VMware Horizon 7.2 on VMware ESXi 7.2 U2 hypervisor. This chapter does not address the general issues of multi-site deployment and network management and limits the description to the components that are inside the customer's intranet.
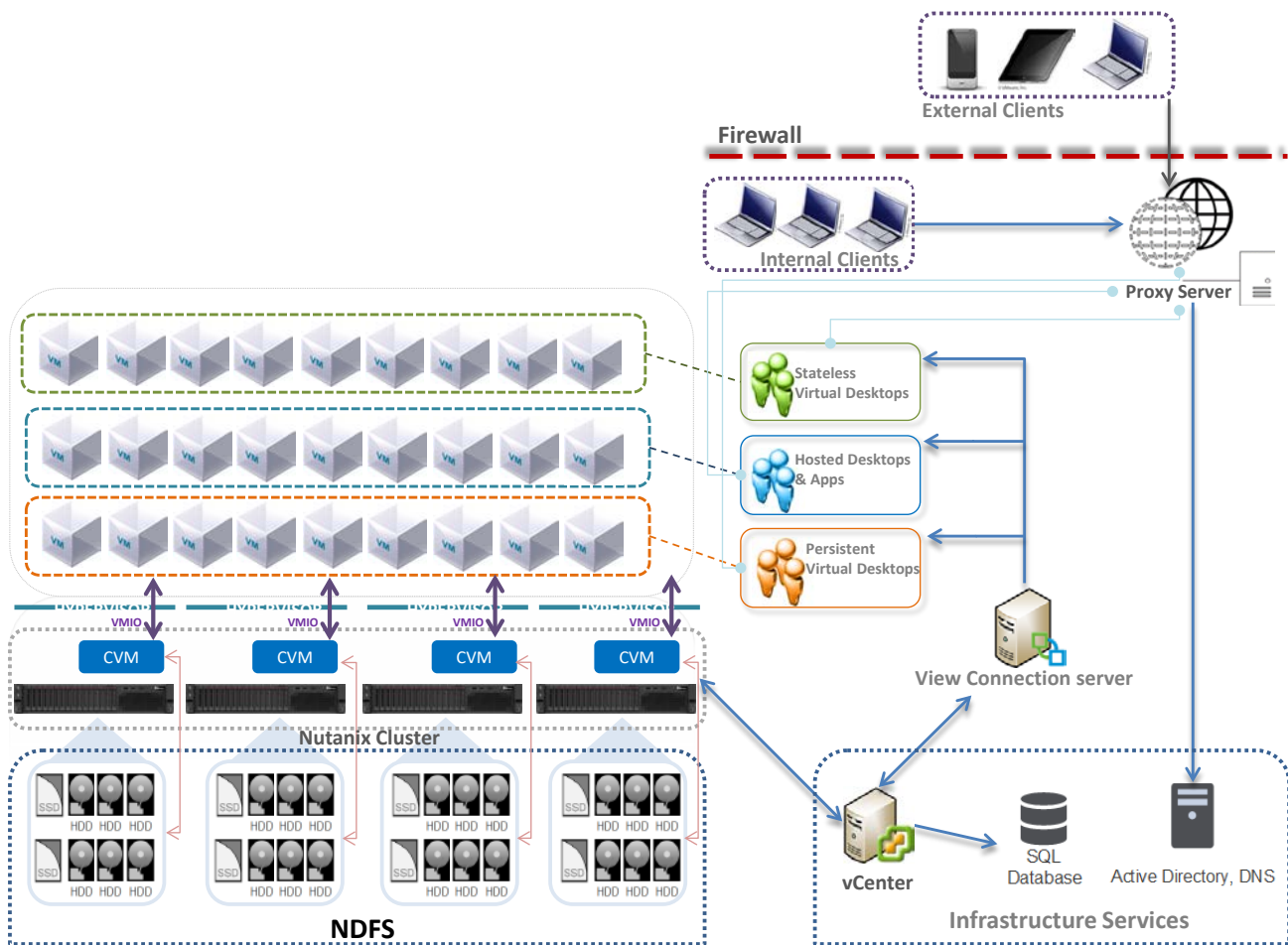

Figure 35: Lenovo ThinkAgile HX Series solution with VMware Horizon

7.2 Component model

Figure 36 is a layered component view for the VMware Horizon virtualization infrastructure.
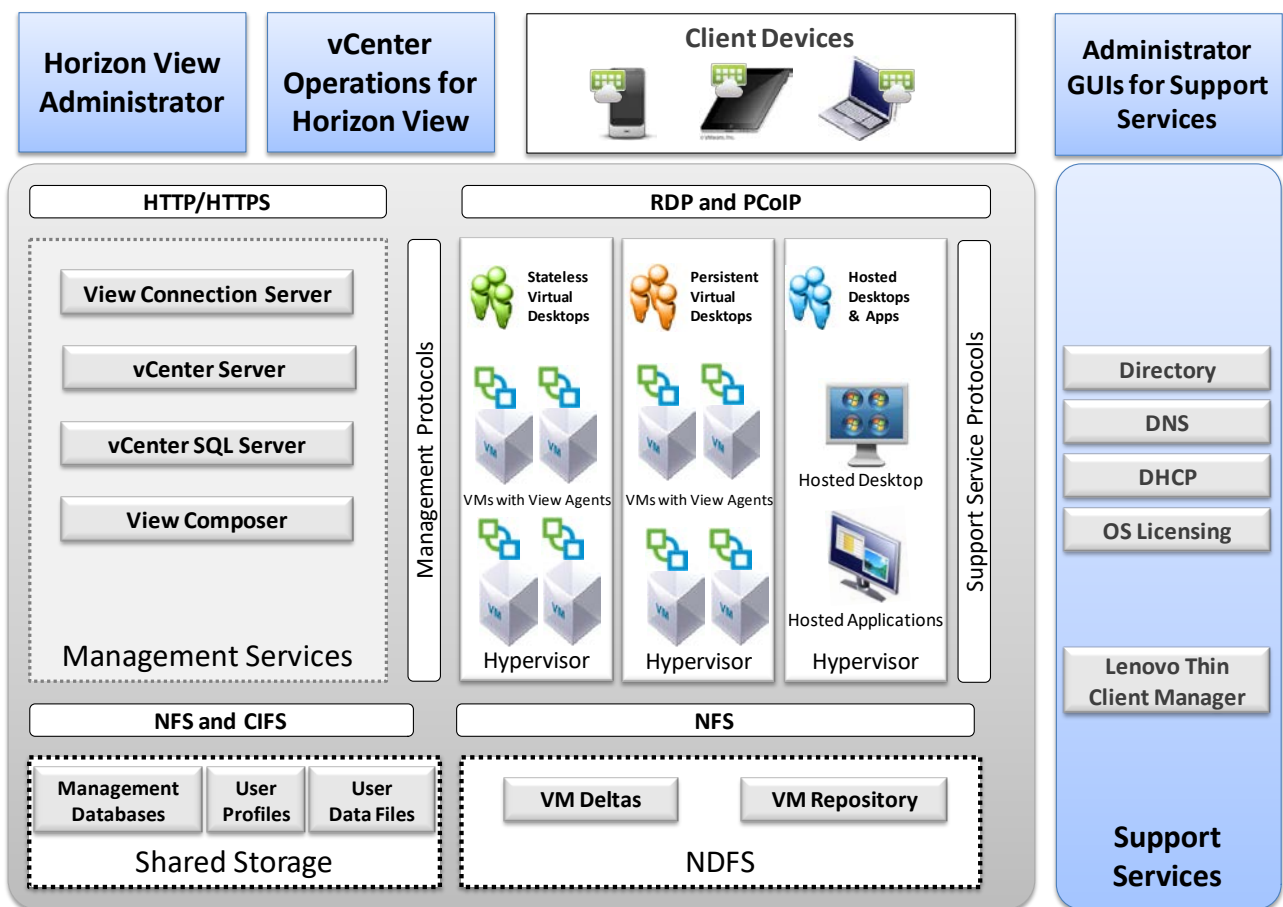
Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Figure 36: Component model with VMware Horizon

VMware Horizon with the VMware ESXi hypervisor features the following main components:

Horizon View Administrator	By using this web-based application, administrators can configure ViewConnection Server, deploy and manage View desktops, control user authentication, and troubleshoot user issues. It is installed during the installation of ViewConnection Server instances and is not required to be installed on local (administrator) devices.
vCenter Operations for Horizon View	This tool provides end-to-end visibility into the health, performance, and efficiency of the virtual desktop infrastructure (VDI) configuration. It enables administrators to proactively ensure the best user experience possible, avert incidents, and eliminate bottlenecks before they become larger issues.
View Connection Server	The VMware Horizon Connection Server is the point of contact for client devices that are requesting virtual desktops. It authenticates users and directs the virtual desktop request to the appropriate virtual machine (VM) or desktop, which ensures that only valid users are allowed access. After the authentication is complete, users are directed to their assigned VM or desktop.
	If a virtual desktop is unavailable, the View Connection Server works with the management and the provisioning layer to have the VM ready and available.

View Composer	In a VMware vCenter Server instance, View Composer is installed. View Composer is required when linked clones are created from a parent VM.
vCenter Server	By using a single console, vCenter Server provides centralized management of the virtual machines (VMs) for the VMware ESXi hypervisor. VMware vCenter can be used to perform live migration (called VMware vMotion), which allows a running VM to be moved from one physical server to another without downtime.
	Redundancy for vCenter Server is achieved through VMware high availability (HA). The vCenter Server also contains a licensing server for VMware ESXi.
vCenter SQL Server	vCenter Server for VMware ESXi hypervisor requires an SQL database. The vCenter SQL server might be Microsoft® Data Engine (MSDE), Oracle, or SQL Server. Because the vCenter SQL server is a critical component, redundant servers must be available to provide fault tolerance. Customer SQL databases (including respective redundancy) can be used.
View Event database	VMware Horizon can be configured to record events and their details into a Microsoft SQL Server or Oracle database. Business intelligence (BI) reporting engines can be used to analyse this database.
Clients	VMware Horizon supports a broad set of devices and all major device operating platforms, including Apple iOS, Google Android, and Google ChromeOS. Each client device has a VMware View Client, which acts as the agent to communicate with the virtual desktop.
Thin-client Manager	The Lenovo Thin-client Manager (LTM) is used to manage and support Lenovo thin-client devices individually or in groups.
RDP, PCoIP	The virtual desktop image is streamed to the user access device by using the display protocol. Depending on the solution, the choice of protocols available are Remote Desktop Protocol (RDP) and PC over IP (PCoIP).

For more information, see the Lenovo Client Virtualization base reference architecture document that is available at this website: lenovopress.com/lp0756.

7.3 VMware Horizon provisioning

VMware Horizon supports stateless and persistent virtual desktop models. Provisioning for VMware Horizon is a function of vCenter server and View Composer for linked clones.

vCenter Server allows for manually created pools and automatic pools. It allows for provisioning full clones and linked clones of a parent image for dedicated and stateless virtual desktops.

Because persistent virtual desktops use large amounts of storage, linked clones can be used to reduce the storage requirements. Linked clones are created from a snapshot (replica) that is taken from a golden master image. One pool can contain up to 2000 linked clones.

This document describes the use of automated pools (with linked clones) for dedicated and stateless virtual desktops. The deployment requirements for full clones using Nutanix de-duplication functionality is beyond the scope of this document.

7.4 Management VMs

A key part of the VMware Horizon environment is the various management VMs used to manage the VDI infrastructure and user VMs. Table 19 lists the VM requirements and performance characteristics of each management service.

Table 19: Characteristics of VMware Horizon management services

Management service VM	Virtual processors	System memory	Storage	Windows OS	HA needed	Performance characteristic
vCenter Server	8	12 GB	60 GB	2012 R2	No	Up to 2000 VMs.
vCenter SQL Server	4	8 GB	200 GB	2012 R2	Yes	Double the virtual processors and memory for more than 2500 users.
View Connection Server	4	16 GB	60 GB	2012 R2	Yes	Up to 2000 connections.

These management VMs can be run on separate servers from the HX series cluster or within the cluster itself. Separating out the VMs means that the management VMs can be separately managed and sized to the requirements and dedicated servers used for the user VMs. Putting all of the VMs together in one cluster means that the compute servers will execute less user VMs and need to leave enough resources for the much larger and more granular management VMs. Lenovo recommends that the management and user VMs are separated for all but the smallest deployments (i.e. less than 600 users).

Table 20 lists the number of management VMs for each size of users following the requirements for high-availability and performance. The number of vCenter servers is half of the number of vCenter clusters because each vCenter server can handle two clusters of up to 1000 desktops.

Table 20: Management VMs needed

Horizon management service VM	300 users	600 users	1200 users	3000 users
vCenter servers	1	1	2	2
vCenter SQL servers	2 (1+1)	2 (1+1)	2 (1+1)	2 (1+1)
View Connection Server	2 (1+1)	2 (1+1)	2 (1+1)	2 (1+1)

It is assumed that common services, such as Microsoft Active Directory, Dynamic Host Configuration Protocol (DHCP), domain name server (DNS), and Microsoft licensing servers exist in the customer environment.

7.5 Graphics acceleration

This section is specific to the Lenovo ThinkAgile HX3520-G and ThinkAgile HX3521-G that support GPU acceleration. The VMware ESXi hypervisor supports the following options for graphics acceleration:

- Dedicated GPU with one GPU per user, which is called virtual dedicated graphics acceleration (vDGA) mode.
- GPU hardware virtualization (vGPU) that partitions each GPU for 1 - 8 users.
- Shared GPU with users sharing a GPU, which is called virtual shared graphics acceleration (vSGA) mode and is not recommended because of user contention for shared use of the GPU.

The vDGA option has a low user density as it restricts a single user to access each very powerful GPU. This option is not flexible and is no longer cost effective even for high-end power users. Therefore vDGA is no longer recommended especially given that the performance of the equivalent vGPU mode is similar.

When using the vGPU option with ESXi 6.5 and the latest drivers from NVidia, it is necessary to change the default GPU mode from "Shared" (vSGA) to "Shared Direct" (vGPU) for each GPU using VMware vCenter. This enables the correct GPU support for the VMs which would otherwise result in the VM not powering on correctly and the standard "graphics resources not available" error message. The host needs to be rebooted for the changes to take effect.

The performance of graphics acceleration was tested using the Lenovo ThinkSystem SR650 servers. Each server supports up to two GPU adapters. The Heaven benchmark is used to measure the per user frame rate for different GPUs, resolutions, and image quality. This benchmark is graphics-heavy and is fairly realistic for designers and engineers. Power users or knowledge workers usually have less intense graphics workloads and can achieve higher frame rates. *Table 21* lists the results of the Heaven benchmark as FPS that are available to each user with the GRID 2,0 M60 adapter by using vGPU mode with DirectX 11.

Table 21: Performance of GRID 2.0 M60 vGPU modes with DirectX 11

Quality	Tessellation	Anti-Aliasing	Resolution	M60-8Q	M60-4Q	M60-2Q	M60-4A	M60-2A
High	Normal	0	1280x1024	Untested	Untested	32.03	59.81	32.98
High	Normal	0	1680x1050	Untested	49.97	25.41	N/A	N/A
High	Normal	0	1920x1200	Untested	41.36	21.03	N/A	N/A
Ultra	Extreme	8	1280x1024	Untested	Untested	18.02	37.01	18.68
Ultra	Extreme	8	1680x1050	56.01	29.67	14.18	N/A	N/A
Ultra	Extreme	8	1920x1080	50.38	25.69	12.76	N/A	N/A
Ultra	Extreme	8	1920x1200	46.01	22.79	Untested	N/A	N/A
Ultra	Extreme	8hai	2560x1600	27.42	14.16	Untested	N/A	N/A

Lenovo recommends that a medium to high powered CPU, such as the Xeon Scalable 6130, is used for accelerated graphics applications tend to also require extra load on the processor. For vGPU mode, Lenovo recommends at least 384GB of server memory. Because there are many variables when graphics acceleration is used, Lenovo recommends that testing is done in the customer environment to verify the performance for the required user workloads.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

7.6 Performance testing

This section describes the performance benchmarking tool and the results obtained for different configurations of a cluster of 4 Lenovo ThinkAgile HX3320 appliances. This chapter applies equally to the ThinkAgile HX3321 certified nodes.

7.6.1 Login VSI benchmarking tool

Login VSI is a vendor-independent benchmarking tool that is used to objectively test and measure the performance and scalability of server-based Windows desktop environments. Leading IT analysts recognize and recommend Login VSI as an industry-standard benchmarking tool for client virtualization and can be used by user organizations, system integrators, hosting providers, and testing companies.

Login VSI provides multiple workloads to simulate real user work and suitable in performing load test, benchmarking and capacity planning for VDI environments. Table 22 lists the characteristics of the Login VSI 4.1 workloads that are used in the Lenovo testing.

Table 22. Login VSI Workload Comparison

Workload Name	Login VSI Version	Apps Open	CPU Usage	Disk Reads	Disk Writes	IOPS	Memory	vCPU
Office worker	4.1	5-8	82%	90%	101%	8.1	2GB	1vCPU
Knowledge worker	4.1	5-9	100%	100%	100%	8.5	2GB	2vCPU
Power worker	4.1	8-12	119%	133%	123%	10.8	3GB	3vCPU

The VSImax score parameter (the number indicates user density) is used to determine the performance of a particular system configuration. The following parameters and rules are used for Login VSI tests:

- User login interval: 30 seconds per node
- Workload: Office Worker, Knowledge Worker, or Power User
- All virtual desktops were pre-booted before the tests
- The number of powered-on VMs was adjusted to stay within a 10% margin of VSImax to avoid unreasonable overhead by "idling" virtual machines
- VSImax score is derived using the "classic model" calculation

7.6.2 Performance results for virtual desktops

This section shows the virtual desktop performance results for Lenovo ThinkAgile HX all flash appliances.

The recommended configuration of the Nutanix CVM is as follows:
- vCPU   12
- CPU Reservation   10000 MHz
- Memory   40GB
- Memory Reservation   40GB
- NUMA   No Affinity
- Advance CPU – Scheduling Affinity   No Affinity

Table *23* lists the Login VSI performance results of a HX3330 Series appliance 4 node cluster with 2TB memory and 2 x1.6 TB NVMe cache drives and 4 x 1.92TB SATA SSD capacity drives. The benchmark

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

perfomed on VMware ESXi 7.0 U2 and Windows 10 with the latest security fixes and the new Xeon Scalable Processors generation 3.

Table 23: Login VSI Performance with VMware ESXI 7.0 U2 on ThinkAgile HX3330

Processor	Workload	Stateless	Dedicated
2x Intel Xeon 6338N 2.20 GHz, 32C 185W	Knowledge worker		350 users

Table *24* lists the Login VSI performance results of a HX3376 Series Certified 4 node cluster with on AMD EPYC 7003 series processors, 2TB memory and 2x 800 GB SAS SSD cache drives and 6 x 7.68 TB SAS SSD capacity drives using ESXi 6.7 U2 and Windows 10.

Table 24: Login VSI Performance with WMware ESXi 6.7 U2 on ThinkAgile HX3376

Processor	Workload	Stateless	Dedicated
2xAMD EPYC 73F3,16 Core, 3.5 Ghz 240W	Knowledge worker		214 users
2xAMD EPYC 7543, 32 Core, 2.8 GHz 225W	Power worker		315 users

7.6.3 Performance results from boot storm testing

A boot storm occurs when a substantial number of VMs are all booted within a short period of time. Booting a large number of VMs simultaneously requires large IOPS otherwise the VMs become slow and unresponsive.

Different numbers of VMs were booted on a cluster of 4 HX3320 hybrid appliances. The VMs were unpowered in vCenter and the boot storm created by powering on all of the VMs simultaneously. The time for all of the VMs to become visible in VMware Horizon was measured.

Figure 37 shows a comparison of the boot times for a variety of VMs. With an even spread of VMs on each node, the boot time for the VMs on each node was similar to the overall cluster boot time.
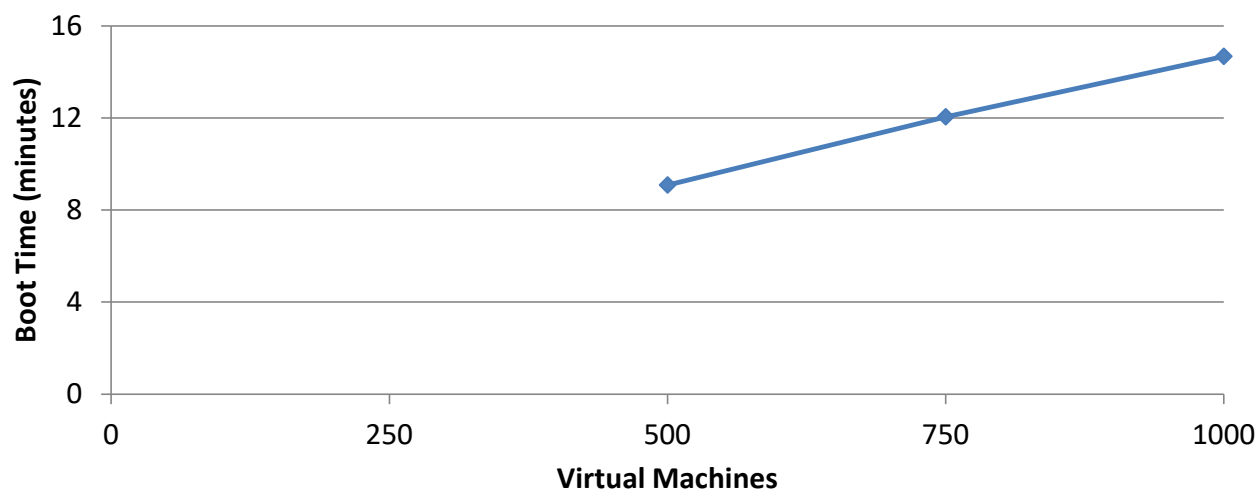

Figure 37: Boot storm comparison

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

7.7 Performance recommendations

This section provides sizing performance recommendations and best practices.

7.7.1 Sizing recommendations for virtual desktops

The default recommendation is two Xeon Scalable 6338N processors and 768 GB of system memory because this configuration provides the best coverage and density for a range of users. Assuming there is enough storage configured for the VMs, this configuration is recommended for any of the HX 3000 Series appliances.

For a knowledge worker, Lenovo testing shows that 240 users per server is a good baseline and has an average of 75% usage of the processors in the server. For the degraded failover case, Lenovo testing shows that 300 users per server have an average of 87% usage of the processors. By using a target of 240 users per server, the maximum number of knowledge workers is 4000 in a 16 node cluster.

Table *25* summarizes the processor usage with ESXi for the recommended user counts for normal mode and failover mode.

Table 25: Processor usage

Processor	Workload	Users per Server	CPU Utilization
Two 6338N	Knowledge worker	240 users – Normal Mode	75%
Two 6338N	Knowledge worker	300 users – Failover Mode	87%

Table *26* lists the recommended number of virtual desktops per server for different workload types and VM memory sizes. The number of users is reduced in some cases to fit within the available memory and still maintain a reasonably balanced system of compute and memory.

Table 26: Recommended number of virtual desktops per server

Workload	Knowledge worker
Processor	Two 6338N
VM memory size	4 GB
System memory	1000 GB
Memory overhead of CVM	24 GB
Desktops per server (normal mode)	180
Desktops per server (failover mode)	240

Table *27* lists the approximate number of compute servers that are needed for different numbers of users and Knowledge worker workloads.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Table 27: Compute servers needed for Knowledge workers and different numbers of users

Knowledge workers	500 users	1000 users	1500 users	3000 users
Compute servers @180 users (normal)	3	7	9	16
Compute servers @240 users (failover)	2	6	8	15

7.7.2 Best practices

The number of desktops that can be run on a specific server depends upon the available system memory, compute power of the processors, and number of logons per second during a logon storm. For a cost-effective solution, the maximum number of users should be put on each server to balance processor, memory, storage I/O, and networking. Lenovo recommends using all flash appliances for situations where the user logon rate is high or time to reboot all the VMs on a node must be less than 10 minutes.

Another important consideration for compute servers is system memory. For stateless users, the typical range of memory that is required for each desktop is 2 GB - 4 GB. For dedicated users, the range of memory for each desktop is 2 GB - 6 GB. In general, power users that require larger memory sizes also require more virtual processors. This reference architecture standardizes on 2 GB per desktop as the minimum requirement of a Windows 10 desktop. The virtual desktop memory should be large enough so that swapping is not needed and vSwap can be disabled.

It is a best practice not to overcommit on memory as swapping to disk can have a severe effect on performance; a better strategy is to give each desktop more memory. Alternatively, a monitoring tool can be run to gather information about existing desktops. The desktop memory size that is required does not necessarily have to match the memory supplied in a desktop machine; it can be larger or smaller.

Lenovo recommends the use of VLANs to partition the network traffic. The following VLANs should be used:

- User (for web protocols, display protocols, and support service protocols)
- Management (for management protocols)
- Storage (for NDFS)

Lenovo recommends to always perform user virtualization, even if users have dedicated desktops. This separation of user-specific data makes it much easier to manage and perform upgrades.

Windows 10 was used for all of the performance testing. In general Windows 10 requires 10% to 20% more compute power than Windows 7. The following optimizations are recommended for the Windows 10 base image:

- Applied #VDILIKEAPRO Tuning Template(developed by loginVSI) – see the following for more details:
  loginvsi.com/blog/520-the-ultimate-windows-10-tuning-template-for-any-vdi-environment

- Set Adobe acrobat as a default app for PDF files using steps in following webpage:
  adobe.com/devnet-docs/acrobatetk/tools/AdminGuide/pdfviewer.html

- Disabled **Windows Modules installer** service on the base image because the CPU utilization can remain high after rebooting all the VMs. By default this service is set to manual rather than disabled.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Please refer below links for best practices and optimizations recommended by VMware:

- View Architecture Planning – VMware Horizon 7.0:
  pubs.vmware.com/horizon-view-60/topic/com.vmware.ICbase/PDF/horizon-view-60-architecture-planning.pdf

7.8 Deployment ready solutions

There are 3 deployment ready solutions can be leveraged for VMware Horizon in the Lenovo Data Center Solution Configurator (DCSC): The hypervisor needs to be changed to ESXi on the following solution indentifiers to deploy VMware Horizon instead of Citrix Virtual Apps and Desktops https://dcsc.lenovo.com/#/reference-solution.

Solution Identifier	SID0000443	SID0000444	SID0000445
ThinkAgile HX node	HX630 V3 Integrated System	HX630 V3 Integrated System	HX650 V3 Integrated System
Configuration	Flash	Flash	Flash
Processor	Intel Xeon Gold 5416S 16C 150W 2.0GHz	Intel Xeon Gold 6230 20C 125W 2.1GHz	Intel Xeon Gold 6448Y 32C 225W 2.1GHz
Office Workers (3GB)	200 users/node	150 users/node	N/A
Knowledge workers (4GB)	N/A	50 users/node	N/A
Power users (5GB)	N/A	N/A	200 users/node
Power user with GPU (6GB)	N/A	N/A	N/A
Light Cad with GPU (8GB)	N/A	N/A	N/A
Heavy Cad with GPU (16GB)	N/A	N/A	N/A

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

8 VMware vCloud Suite

VMware vCloud Suite is an integrated offering that brings together VMware's industry-leading vSphere hypervisor and VMware vRealize Suite multi-vendor hybrid cloud management platform. This chapter covers the following VMware products:

- vSphere 6.7 (and vSphere 6.5 U2), which provides compute virtualization
- vCloud Suite 7.4, which provides a VMware vSphere-based private cloud using vRealize Suite products and additional products to support vSphere Data Protection and Availability
- vRealize Suite 7.4, which provides cloud management capabilities for private, public and hybrid clouds with support for multiple hypervisors
- AWS Server Migration Service Connector 1.0.12.50 which supports migration of virtual machines from on-premise vSphere cloud to AWS pubic cloud.

VMware NSX 6.4.1 provides network virtualization by using software defined networking (SDN) and supports integration with hardware layer 2 gateways. The use of NSX with Lenovo ThinkAgile HX Series appliances and certified nodes was verified but a full description is outside the scope of this Reference Architecture. For more details see: www.nutanix.com/go/vmware-nsx-for-vsphere.php.

8.1 Solution Overview

This section gives an architectural overview of vCloud Suite products. Figure 38 gives an overview of how those products are deployed into shared edge and compute, management, and additional compute clusters.

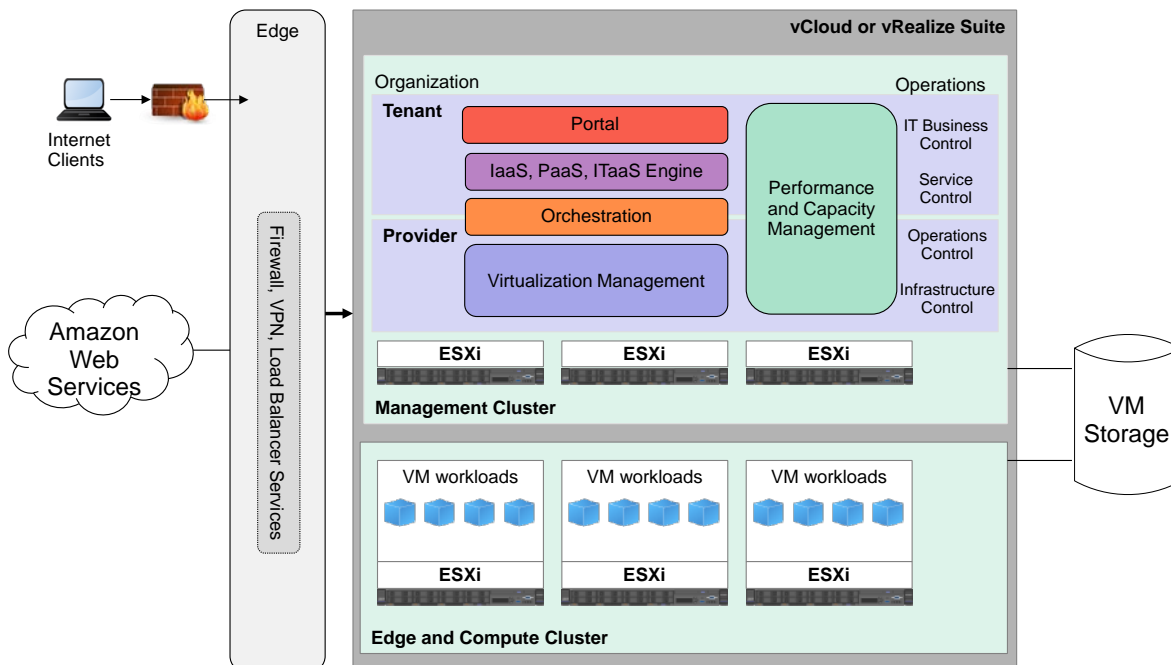This separation of function into these clusters allows for scaling in larger environments.

Figure 38: Conceptual design of vCloud Suite

The management cluster runs the components required to support vCloud Suite and is used for management, monitoring, and infrastructure services. A management cluster provides resource isolation which helps these services to operate at their best possible performance level. A separate cluster can satisfy an organization's

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

policy to have physical isolation between management and production hardware and a single management cluster is required for each physical location.

The shared edge and compute cluster supports virtualized infrastructure services as well as network devices that provide interconnectivity between environments. It provides protected capacity by which internal data center networks connect via gateways to external networks. Networking edge services and network traffic management occur in this cluster and all external facing network connectivity ends in this cluster. The shared edge and compute cluster also supports the delivery of all other (non-edge) customer workloads and there can be one or more compute clusters, depending on the customer environment. Multiple compute clusters can be for different organizations or tenants, different workload types, or to spread the load in a large enterprise.

8.2 Component model

This section describes the component model for VMware vCloud Suite and optionally extending it into public clouds with hybrid cloud connections.

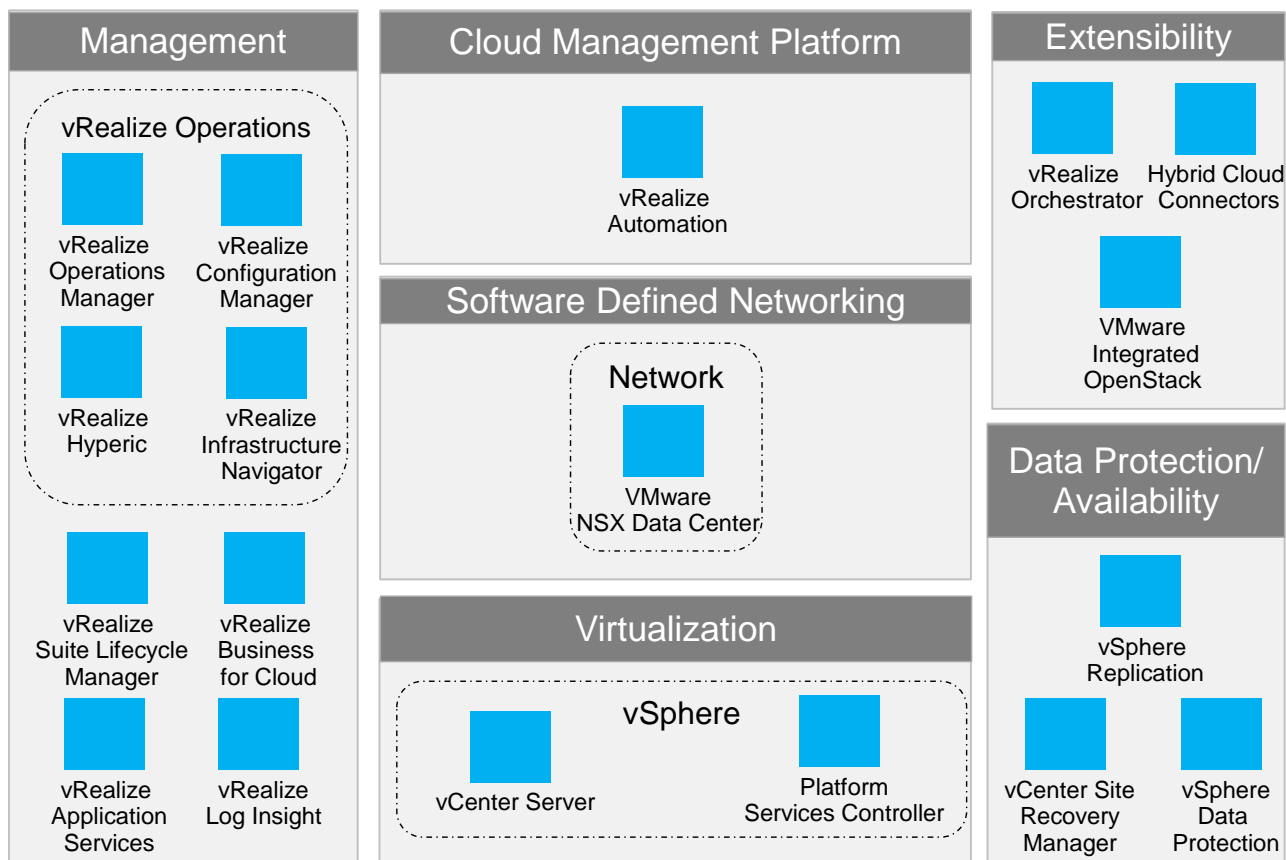Figure 39 shows an overview of the major components of the VMware vCloud Suite.

Figure 39: vCloud Suite components

The VMware vCloud Suite features the following components:

ESXi hypervisor   Provides bare-metal virtualization of servers so you can
                                   consolidate your applications on less hardware.

vCenter Server	Provides a centralized platform for managing vSphere environments and includes vSphere replication and vSphere data protection.
Platform Services Controller (PSC)	Provides a set of common infrastructure services that encompasses single sign-on (SSO), licensing, and a certificate authority (CA).
vRealize Suite Lifecycle Manager	Provides deployment options such as install, configure, import, and upgrade vRealize Suite environments and perform drift analysis and view the health of those environments
vRealize Automation	Provides a self-service, policy-enabled IT and application services catalog for deploying and provisioning of business-relevant cloud services across private and public clouds, physical infrastructure, hypervisors, and public cloud providers.
vRealize Operations	Provides a set of components for automation of operations including infrastructure health, configurations and compliance, application discovery, and monitoring of hardware and software.
• vRealize Operations Manager	Provides comprehensive visibility and insights into the performance, capacity and health of your infrastructure.
• vRealize Configuration Manager	Provides automation of configuration and compliance management across your virtual, physical, and cloud environments, which assesses them for operational and security compliance.
• vRealize Infrastructure Navigator	Provides automated discovery of application services, visualizes relationships, and maps dependencies of applications on virtualized compute, storage, and network resources.
• vRealize Hyperic	Provides monitoring of operating systems, middleware, and applications that are running in physical, virtual, and cloud environments.
vRealize Business for Cloud	Provides transparency and control over the costs and quality of IT services that are critical for private (vCloud Suite) or hybrid cloud (vRealize Suite) success.
vRealize Log Insight	Provides analytics capabilities to unstructured data and log management, which gives operational intelligence and deep, enterprise-wide visibility across all tiers of the IT infrastructure and applications. Standard for vRealize Suite.
vCenter Site Recovery Manager (SRM)	Provides disaster recovery capability with which you can perform automated orchestration and non-disruptive testing for virtualized applications by using ESXi hypervisor only. SRM is standard for vCloud Suite and optional for vRealize Suite.

vRealize Orchestrator	Provides the capability to create workflows that automate activities, such as provisioning VM, performing scheduled maintenance, and starting backups.
NSX	NSX provides virtualization of networking and is optional for vCloud Suite deployments.
VMware Integrated OpenStack (VIO)	Provides a VMware-supported OpenStack distribution (distro) that makes it easier for IT to run a production-grade, OpenStack-based deployment on top of their VMware infrastructure. For more information, see this website: vmware.com/products/openstack.
Hybrid Cloud Connectors	Allows an administrator to provide hybridization using public cloud providers such as Amazon AWS. See the next section for more information.

The vCloud Suite products also have dependencies on the following external components:

Identity source	Identity sources (Active Directory, OpenLDAP, or Local OS) or similar is required to implement and operate the vCloud Suite or vRealize Suite infrastructure.
DNS	DNS must be configured for connectivity between vCenter Server, Active Directory, ESXi hosts, and the VMs
DHCP/TFTP	PXE boot is required for vSphere Auto Deploy functionality.
Time synchronization	Accurate time keeping and time synchronization is critical for a healthy infrastructure. All components (including ESXi hosts, vCenter Server, the SAN, physical network infrastructure, and VM guest operating systems) must have accurate time keeping.
Microsoft SQL Server or Oracle database	Many of the vCloud Suite components come with embedded databases or they can use external databases such as Microsoft SQL Server or Oracle, depending on the component and the intended environment.

Other software components such as Lenovo XClarity Administrator are not shown. As well as providing management of Lenovo hardware, XClarity Administrator also has plugins for VMware vCenter, VMware vRealize Orchestrator, and VMware vRealize Log Insight which are further described in "Systems management" on page 90.

8.2.1 Hybrid Clouds

On-premise VMware vCloud suite connects to public clouds such as Amazon Web Services (AWS), IBM Public Cloud and OVH Public Cloud. This document only discusses AWS support.

The Amazon Elastic Compute Cloud (EC2) provides scalable computing capacity in the Amazon Web Services (AWS) public cloud by offering compute, storage, networking, software, and development tools. AWS provides Virtual Private Cloud and Dedicated Hosts for compute and different services. It supports a hybrid

architecture by integrating networking, security and access control, automated workload migrations and controlling AWS from an on-premise infrastructure management tool.

AWS Server Migration Service (AWS SMS) is an agentless service to migrate on-premise workloads from vCenter to AWS. It supports incremental replications and migration can be performed faster while minimizing network bandwidth consumption and reducing server downtime. Each server volume replicated is saved as a new Amazon Machine Image (AMI) which can be launched as an EC2 instance (VM) in the AWS cloud. The AWS Server Migration Service replicates server volumes from on-premises environment to S3 temporarily and purges them from S3 immediately after creating the Elastic Block Store (EBS) snapshots.

Table 28 describes the features supported by the AWS SMS connector. Some best practices for deployment are described in "Hybrid networking to public clouds" on page 44.

Table 28: Hybrid Cloud Features

Feature	AWS SMS Connector
Bi-Directional Migration	No
Integration with vCenter	Yes
vCenter→Connector Association	Many-Many
Connector→Public Cloud Association	1-1
vCenter linked Mode support	No
Integration with vSphere client	Use AWS Management Console
Integration with vSphere web client	Use AWS Management Console
Integration with vSphere Replication	No
Integration with vRealize Automation	No
Multi-tenant support	No
VM Management public cloud	Use AWS Management Console
VM Management on-premise	No
Migration to all public cloud regions	Limited Currently
Copy/Migrate Templates to public cloud	Yes (AMI)
Deploy VM from Template to public cloud	Yes (AMI)
Live Migration to public cloud	Yes
Cold Migration to public cloud	Yes
Bulk Migration to public cloud	Yes
Layer 2 Extension	No
Migrate VM to on-premise	No (OVA download)
Offline Data Transfer	No
Common Content Library	No
Number of Concurrent Migration	50 per account
License	90 days Free
vSphere Standard Switch Support	Yes
vSphere Distributed Switch Support	Yes
Network(NSX) Policy Migration	No

8.2.2 VMware Licensing

The licensing for vSphere is based on a CPU metric and licensing for other products is based on the number of OS instances. The vCloud Suite license is a single perpetual license that includes vSphere Enterprise Plus and vRealize Suite. For vSphere environments, the license can be purchased separately for vSphere Enterprise Plus and vRealize Suite. Other components have their own separate licenses and are optional add-ons.

Table 29 lists the standard and optional components that are provided with a vCloud Suite License or vRealize Suite License.

Table 29: VMware Licensing

License	Component	vCloud Standard	vCloud Advanced	vCloud Enterprise	vRealize Standard	vRealize Advanced	vRealize Enterprise
Base	vSphere	Enterprise Plus	Enterprise Plus	Enterprise Plus			
	vRealize Suite Lifecycle Mgr	Included	Included	Included	Included	Included	Included
	vRealize Automation	N/A	Advanced	Enterprise	N/A	Advanced	Enterprise
	vRealize Operations • vRealize Operations Mgr • vRealize Configuration Mgr • vRealize Infrastructure Nav • vRealize Hyperic	Advanced	Advanced	Enterprise	Advanced	Advanced	Enterprise
	vRealize Business for Cloud	Standard	Advanced	Advanced	Standard	Advanced	Advanced
	vSphere Replication	Included	Included	Included	N/A	N/A	N/A
	vSphere Data Protection	Included	Included	Included	N/A	N/A	N/A
	vSphere Big Data Extensions	Included	Included	Included	N/A	N/A	N/A
	vRealize Log Insight	Included	Included	Included	Included	Included	Included
	vCenter Site Recovery Mgr	Add On	Add On	Add On	Add On	Add On	Add On
Other licenses	vRealize Orchestrator	Included	Included	Included	Add On	Add On	Add On
	vRealize Automation Public Cloud Extension	N/A	Add On	Add On	Add On	Add On	Add On
	vRealize Operations Public Cloud Extension	N/A	Add On	Add On	Add On	Add On	Add On
	NSX Data Center for vSphere	Included	Included	Included	Add On	Add On	Add On
	VMware Integrated OpenStack	Add On	Add On	Add On	Add On	Add On	Add On

8.3 Shared edge and compute cluster

The shared edge and compute cluster uses its own dedicated vCenter server.

8.3.1 Edge and Infrastructure Services VMs

The VMs used for infrastructure services such as Active Directory, DNS/DHCP, firewalls, proxy and anti-virus are deployed in the shared edge and compute cluster. Table 30 lists each infrastructure service VM with the recommended sizes in terms of virtual CPUs, RAM, storage, and networking.

Table 30: Infrastructure services VMs

VM description	CPU (vCPUs)	Memory (GB)	Storage (GB)	Network bandwidth	High availability
AD, DHCP, DNS server	2	4	70	1 GbE	clustered
http proxy server	2	4	30	1 GbE	clustered

8.3.2 Hybrid cloud VMs

Table 31 lists the cloud connectivity VMs with the recommended sizes in terms of virtual CPUs, RAM, storage, networking, and location. Note that these VMs do not have options for high availability.

Table 31: Cloud connectivity VMs

VM description	CPU (vCPUs)	Memory (GB)	Storage (GB)	Network bandwidth	Location
AWS SMS Connector for vCenter	2	4	300	1 GbE	On-Premise

8.3.3 Server configuration

Since the shared cluster hosts compute workloads and edge services, the servers need to be sized appropriately. See other chapters in this Reference Architecture for specific workloads.

8.3.4 Load balancing and protection

An essential part of the infrastructure is load balancing of the server VMs and recognizing when a server is down and failing over to a second server.

For the shared edge and compute cluster connected to the Internet, it is also important to provide a firewall and protection against external threats. There are many ways to solve these problems such as using a F5 Big-IP edge gateway device or virtual machine. Using F5 protection and load balancing is outside the scope of this document.

8.4 Management cluster

The number of VMware vCloud Suite components in the management cluster increases as capabilities are added. This section addresses the management components that could be used. Third party add-ons must be sized separately.

8.4.1 Management cluster VMs

There are several considerations that contribute to an end-to-end sizing of an entire VMware vCloud environment including Lenovo software for systems management. This section is intended to provide some high-level guidance for management cluster configuration sizing. The recommended number of virtual CPUs, memory size, storage size, and network bandwidth is given for each VM and the VMs are grouped by each major component or appliance.

An essential part of the infrastructure is load balancing of the server VMs and recognizing when a server is down and failing over to another server. The following cases are available for VMs in the management cluster:

- vSphere HA: vCenter automatically restarts the VM on another server, but there is some downtime while the VM starts up.
- Microsoft SQL server clustering: The SQL server cluster automatically handles failover.
- Clustering within component to provide built-in high availability.
- Load balancing: An external load balancer such as a Big-IP switch from F5

Table 32 lists each management cluster VM for vSphere with its recommended size in terms of virtual CPUs, RAM, storage, and networking.

Table 32: Management cluster VMs for vSphere

VM description	CPU (vCPUs)	Memory (GB)	Storage (GB)	Network bandwidth	High availability
vCenter Server(1) Management Cluster	8	24	50	1 GbE	load balancer
vCenter Server(2) Edge and Compute Cluster	8	24	50	1 GbE	load balancer
vCenter Server Database (MS SQL)	4	8	200	1 GbE	SQL AlwaysOn Availability Group
Platform Service Controller (1) Management Cluster	2	4	50	1 GbE	load balancer
Platform Service Controller (2) Edge and Compute Cluster	2	4	50	1 GbE	load balancer
vSphere Replication	2	4	20	1 GbE	not required
vSphere Data Protection	4	4	1600	1 GbE	not required
vRealize Orchestrator Appliance	2	3	12	1 GbE	Clustered

Table 33 lists each management cluster VM for vRealize Automation with its size in terms of virtual CPUs, RAM, storage, and networking.

Table 33: Management cluster VMs for vRealize Automation

VM description	CPU (vCPUs)	Memory (GB)	Storage (GB)	Network bandwidth	High availability
vRealize Suite Lifecycle Manager	4	16	135	1 GbE	N/A
vRealize Automation Appliance	4	16	30	1 GbE	load balancer
IaaS Database (MS SQL)	8	16	100	1 GbE	SQL AlwaysOn Availability Group
Infrastructure Web Server	2	4	40	1 GbE	load balancer
Infrastructure Manager Server	2	4	40	1 GbE	load balancer
Distributed Execution Manager (DEM)	2	6	40	1 GbE	load balancer
vSphere Proxy Agent	2	4	40	1 GbE	load balancer
vRealize Application Services	8	16	50	1 GbE	vSphere HA

Table 34 lists each management cluster VM for vRealize Operations Manager with its size in terms of virtual CPUs, RAM, storage, and networking.

Table 34: Management cluster VMs for vRealize Operations Manager

VM description	CPU (vCPUs)	Memory (GB)	Storage (GB)	Network bandwidth	High availability
vRealize Operations Manager – Master	4	16	500	1 GbE	clustered
vRealize Operations Manager – Data	4	16	500	1 GbE	not required
vRealize Configuration Manager – Collector	4	16	72	1 GbE	not required
vRealize Configuration Manager Database (MS SQL)	4	16	1000	1 GbE	SQL AlwaysOn Availability Group
vRealize Hyperic Server	8	12	16	1 GbE	load balancer
vRealize Hyperic Server - Postgres DB	8	12	75	1 GbE	load balancer
vRealize Infrastructure Navigator	2	4	24	1 GbE	not required

Table 35 lists each of the remaining management cluster VMs.

Table 35: Other Management cluster VMs

VM description	CPU (vCPUs)	Memory (GB)	Storage (GB)	Network bandwidth	High Availability
vRealize Business Standard	2	4	50	1 GbE	vSphere HA
Site Recovery Manager	4	4	20	1 GbE	not required
Site Recovery Manager Database (MS SQL)	2	4	100	1 GbE	SQL AlwaysOn Availability Group
vRealize Log Insight	8	16	100	1 GbE	Cluster of 3 nodes

Table 36 lists the VMs that are needed for Lenovo software for systems management.

Table 36: Lenovo System Management VMs

VM description	CPU (vCPUs)	Memory (GB)	Storage (GB)	Network bandwidth	High availability
Lenovo XClarity Administrator	2	4	64	1 GbE	not required
Lenovo XClarity Integrator (Windows OS)	1	2	30	1 GbE	not required

8.4.2 Server configuration

The management cluster should have a minimum of four hosts for high availability. Because of the large number of management VMs that can be used in the management cluster, the following configuration is recommended for each server:

- Lenovo ThinkAgile HX332x or HX372x
- 2 x Intel® Xeon® Gold 6130 Processor (2.10 GHz 16 cores)
- 384 GB of system memory
- Dual M.2 boot drives with ESXi 6.5 U1

8.5 Hybrid networking to public clouds

This section contains deployment considerations for hybrid networking from an on-premise cloud to public clouds such as AWS.

8.5.1 AWS networking

Table 37 shows the required connectivity for each component used for AWS. . For more details, see the Shared Edge and Compute cluster underlay for the deployment example on page 59.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Table 37: AWS component connectivity

Virtual Machine	DvSwitch	http proxy support
AWS SMS Connector for vCenter	Edge-Compute	Yes

8.5.2 Best Practices

Amazon Web Services (AWS) provide internet and direct connect accessibility options to connect from the vSphere private cloud environments. The AWS connectors leverage an http proxy to establish connectivity to the AWS cloud.

The server hardware for an on-premise vSphere cloud may not use the same processor family. In this case Lenovo recommends enabling Enhanced vMotion Compatibility (EVC) on the cluster to avoid compatibility errors during vMotion.

AWS Connectors are deployed in an on-premise vSphere environment and they do not provide capabilities for bi-directional migration. The on-premise virtual machines can be migrated to AWS cloud but there is no option in the connectors to migrate a virtual machine back to vSphere environment. Instead the VMs need to be exported to OVA format and stored in S3. Then the image can be imported or downloaded to the on-premise vSphere environment.

8.6 Systems management

Lenovo XClarity™ Administrator is a centralized resource management solution that reduces complexity, speeds up response, and enhances the availability of Lenovo® server systems and solutions. See section 2.2.2 on page 5 for more details.

In addition Lenovo provides a number of plugins for VMware vCloud Suite components which are described below.

8.6.1 Lenovo XClarity integration

Lenovo also provides XClarity integration modules for VMware vCenter, VMware vRealize Orchestrator and VMware vRealize Log Insight. For more information, see this website: http://www3.lenovo.com/us/en/data-center/software/systems-management/c/systems-management.

By using the Lenovo XClarity Integrator for VMware vCenter, administrators can consolidate physical resource management in VMware vCenter, which reduces the time that is required for routine system administration.
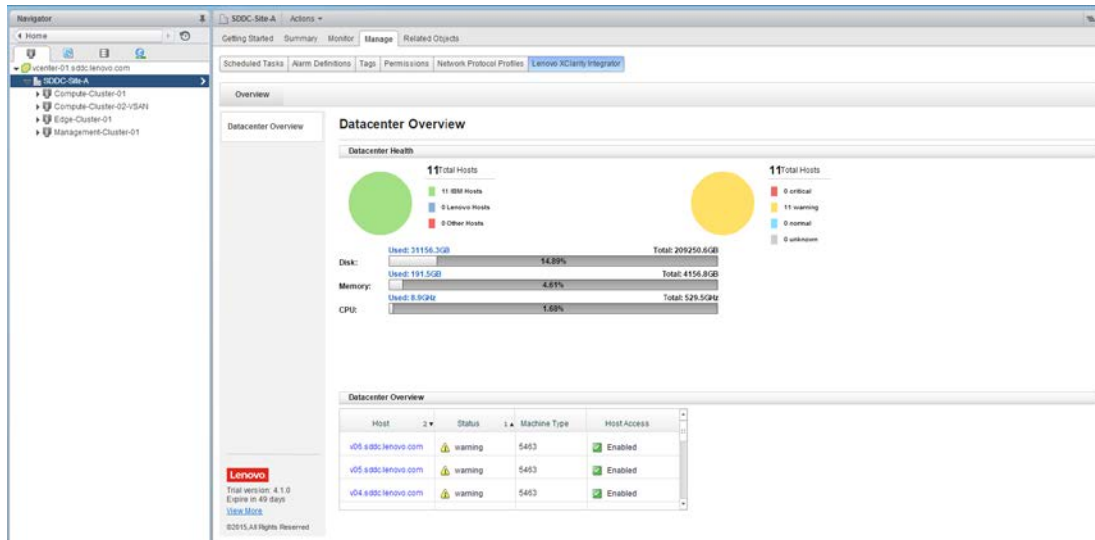
Figure 40: Lenovo XClarity Integrator for VMware vCenter

The Lenovo XClarity Integrator for VMware vCenter provides the following features and benefits:

- Extends Lenovo XClarity Administrator features to the virtualization management console
- Enables management of legacy infrastructure from the virtualization management console
- Reduces workload downtime by dynamically triggering workload migration in clustered environments during rolling server reboots or firmware updates, and predicted hardware failures

The Lenovo XClarity Integrator for VMware vRealize Orchestrator provides IT administrators with the ability to coordinate physical server provisioning features of Lenovo XClarity Pro with broader vRealize Orchestrator workflows. Lenovo XClarity Integrator for VMware vRealize Orchestrator provides a library of simple yet robust and customizable workflow routines and actions designed to automate complex, repetitive IT infrastructure tasks such as system discovery and configuration, hypervisor installation, and addition of new hosts to vCenter.
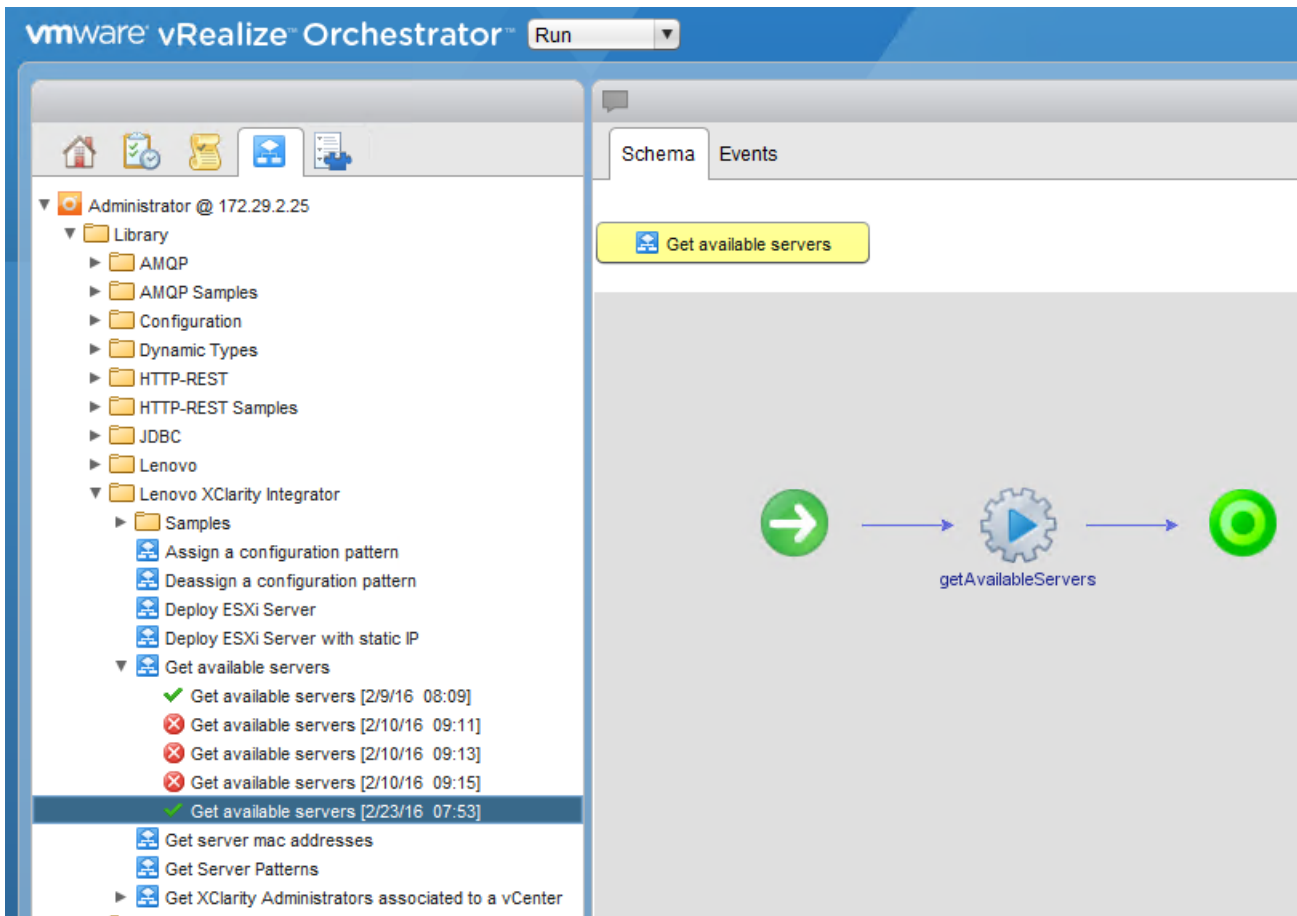
Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Figure 41: Lenovo XClarity Integrator for VMware vRealize Orchestrator interface

The Lenovo XClarity Administrator Content Pack for VMware vRealize Log Insight simplifies the collection and forwarding of Lenovo XClarity Administrator logs to VMware vRealize Log Insight for powerful processing and analytics, and displaying insightful information in an intuitive format.

The VMs for VMware vCenter, vRealize Orchestrator, Lenovo XClarity Administrator and Lenovo XClarity Administrator Integrator should have access to the management network used for managing servers, storage and networking.

Lenovo XClarity Integrator for vRealize Automation provides a set of blueprints to provision infrastructure services based on Lenovo servers, network switches and vSphere. This eases provisioning a new Lenovo server with vSphere installed, apply vSphere distributed switch configurations and adding the server to the existing or new vSphere Cluster. These services leverage the workflows defined in the Lenovo vRealize SoftBundle for vRealize Orchestrator, Lenovo XClarity Integrator for vCenter, Lenovo XClarity Integrator for vRealize Orchestrator, and Lenovo Networking Integration plugin for vRealize Orchestrator.

The Lenovo vRealize SoftBundle package for vRealize Automation needs to imported into vRealize Orchestrator and then the Blueprints package is imported using the vRealize Cloud Client command line utility by Tenant Administrators and it creates catalog items automatically. The catalog items are created under Lenovo Servers, Lenovo Network, and Lenovo Virtualization services.

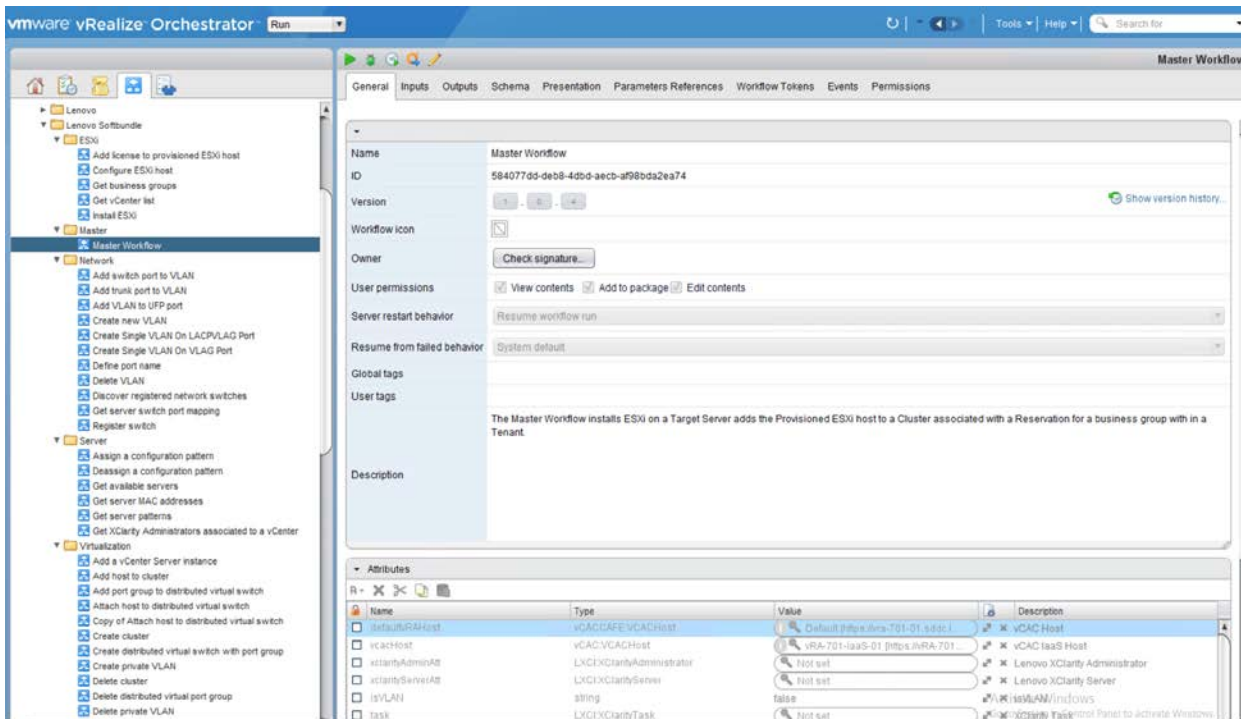Figure 42 shows the workflows available in the Lenovo SoftBundle for vRealize Orchestrator.

Figure 42: Lenovo SoftBundle workflows for vRealize Orchestrator

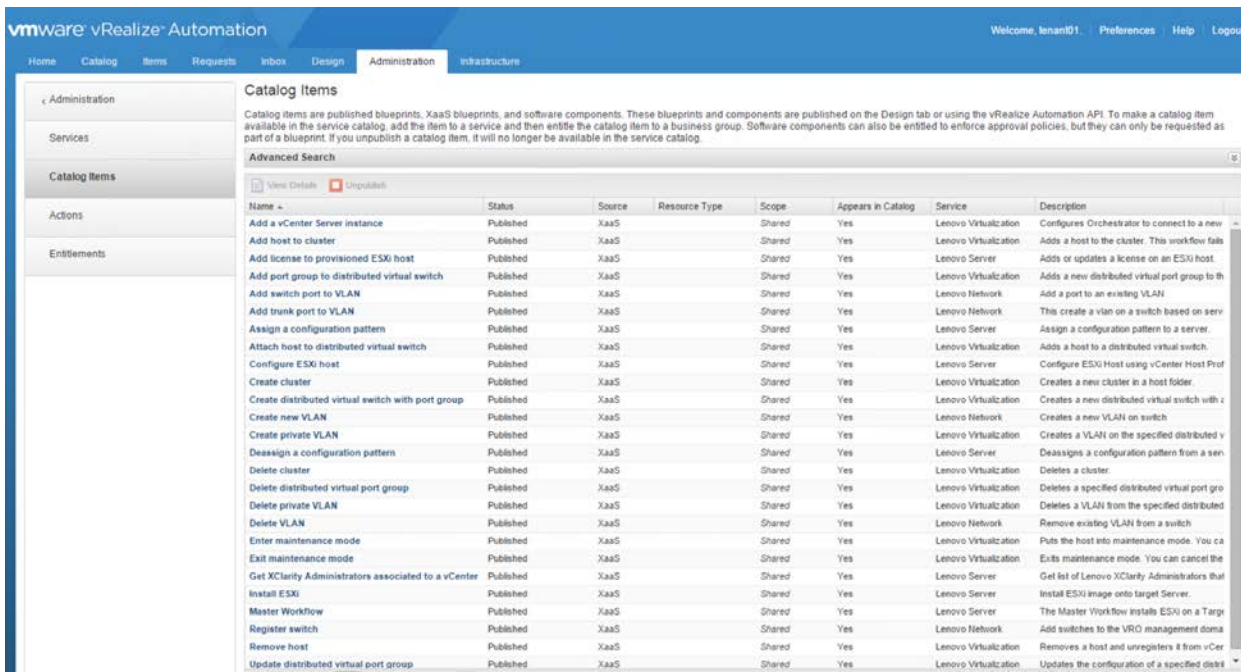Figure 43 shows Lenovo XClarity Integrator catalog items for vRealize Automation.

Figure 43: Lenovo XClarity Integrator for vRealize Automation Catalog Items

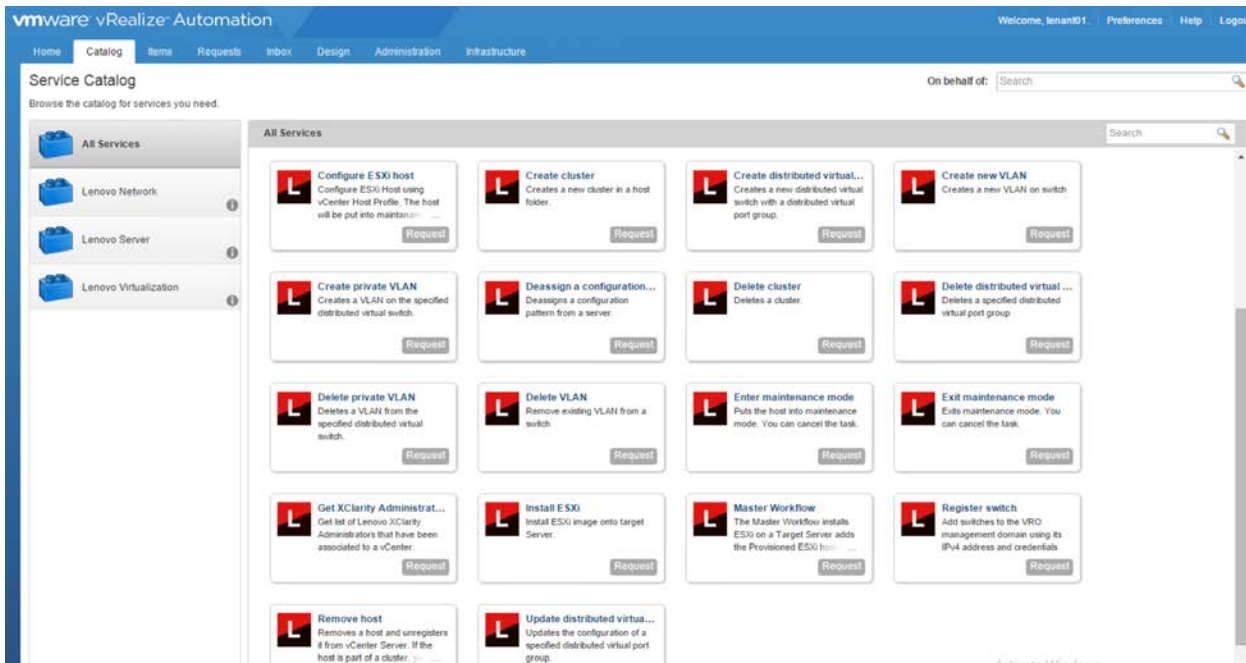Figure 44 shows Lenovo XClarity Integrator services available for vRealize Automation.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Figure 44: Lenovo XClarity Integrator for vRealize Automation Services Catalog

8.6.2 Lenovo plug-ins compatibility

Table 38 below lists current versions of Lenovo integration plugins and the required or supported VMware vCenter and vRealize Suite products.

Table 38: Plug-in compatibility

Component Name	Version	Supported Product Versions
Lenovo XClarity Administrator(LXCA)	2.1.0	VMware vCenter 6.0U2/6.5/6.7, ESXi 6.0U2/6.5 U1/6.7
Lenovo XClarity Integrator(LXCI) for vCenter	5.4.0	Lenovo XClarity Administrator 1.4.x, 2.x VMware vCenter 5.x U1/U2/U3, 6.0 U1/U2/U3, 6.5 U1/U2,6.7
Lenovo XClarity Administrator content pack for VMWare vRealize Log Insight	2.0	Lenovo XClarity Administrator 1.1 or higher VMware vRealize Log Insight 3.0 or higher
Lenovo XClarity Integrator for VMware vRealize Automation	1.0.1	VMware vRealize Automation 7.x
Lenovo XClarity Integrator for VMware vRealize Orchestrator	1.1.0	VMware vRealize Automation 7.0 VMware vRealize Orchestrator 6.0/7.0
Lenovo Network Plugin for VMware vRealize Orchestrator	1.4.0	VMware vRealize Orchestrator 7.4.x

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Lenovo Networking Content Pack for VMware vRealize Log Insight	1.2	VMware vRealize Log Insight 4.3.x, 4.5.x

8.7 Deployment example

There are 3 strategies to deploy vCloud Suite:

- Use VLANs for both user VMs and the Nutanix CVM
- Use NSX and VXLANs for user VMs and the Nutanix CVM in a VLAN
- Use NSX and VXLANs for both user VMs and the Nutanix CVM

This section describes an example deployment of vRealize Suite 7.2 using VLANs for all VMs. The term vRA is used as a short-hand to denote all of the vRealize Automation VMs.

The Lenovo ESXi 6.5 U1 image is used for this example deployment. To use ESXi 6.5 U1, download the Lenovo ThinkSystem custom image from the following website: my.vmware.com/web/vmware/info/slug/datacenter_cloud_infrastructure/vmware_vsphere/6_5#custom_iso.

8.7.1 Physical Model

Four physical servers are used for each of the shared edge and compute, management, and additional compute clusters. Lenovo ThinkSystem HX3320 appliances are used for the shared edge and compute cluster and management cluster. The servers for the additional compute cluster are workload- dependent and Lenovo ThinkSystem HX3520-G appliances are used in this case. See also the "VMware vCloud Suite deployment model" on page 19.

Figure 45 shows a view of the physical 10 GbE network and connections to the external internet.
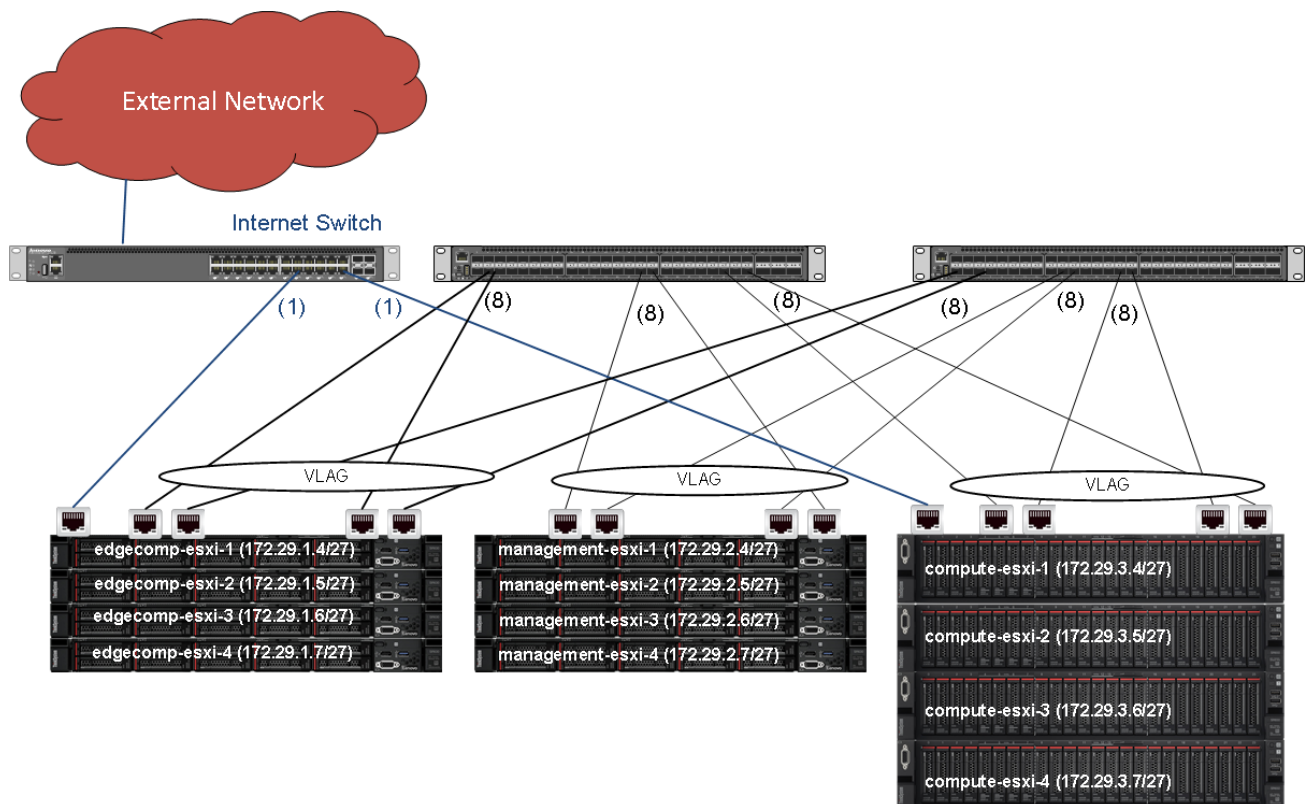

Figure 45: Networking Overview

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

For the shared edge and compute, management and additional compute clusters, the nodes use VLAG technology and as such are using a LAG configuration within the vSphere Distributed Switches. It is recommended to use VLAG for all the clusters connected to the same set of switches.

The servers in the shared edge and compute cluster and the additional compute cluster are connected to a 1G switch. This switch in turn is connected to the internet via a gateway and firewall (not shown).

8.7.2   IP/VLAN mapping

It is recommended to use different VLANs in each of the clusters (shared edge and compute, management, and compute). These different VLANs can be connected together by leveraging Layer 3 routing either in the TOR switch or in upstream switches.

This example deployment uses the following five VLANs:

- Management (includes Nutanix CVM)
- vMotion
- FT
- vRA
- Comp (for customer specific computation data traffic)

It is a Nutanix best practice to include the CVM as part of the management VLAN. Alternatively the Nutanix CVM could be executed in its own Storage VLAN. Local storage traffic between VMs and the local CVM is done with a Nutanix specific vSwitch (vSwitchNutanix) using the svm-iscsi-pg port group. Do not modify vSwitchNutanix, as this provides critical CVM-to-hypervisor storage communication on CVM eth1.

Table 39 lists example IP address ranges for the VLANs in each cluster where RID means Rack ID.

Table 39: Network Segments

Cluster	Shared Edge and Compute (RID 1)		Management (RID 2)		Compute (RID 3)	
	Subnet	VLAN	Subnet	VLAN	Subnet	VLAN
Manage	172.29.1.0/27	101	172.29.2.0/27	201	172.29.3.0/27	301
vMotion	172.29.1.32/27	102	172.29.2.32/27	202	172.29.3.32/27	302
FT	172.29.1.64/27	103	172.29.2.64/27	203	172.29.3.64/27	303
vRA	N/A		172.29.2.192/27	207	N/A	
Comp	172.29.1.192/27	109	N/A		172.29.1.192/27	109

In this example, each cluster needs a minimum of five network segments within the 172.29.RID.x address range. Each segment does not require more than 30 IP addresses; therefore, a 255.255.255.224 (/27) netmask provides enough addresses. The same VLAN IDs can be used across racks with different IP segments as shown by VLAN 109.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

8.7.3　Layer 3 Routing with VRRP

Virtual Routing Redundancy Protocol (VRRP) should be enabled in the switches for layer 3 routing. Each switch defaults to IP routing. VMs can use the respective routing IP to reach the switches. Routing occurs through either one of the switches and this causes intermittent routing failures when used with VLAG.

Layer 3 routing with VRRP removes this limitation by using a single routing IP address for both the switches. Each subnet requires 2 IPs reserved for normal layer 3 routers and one IP reserved for VRRP router.

Table 40 lists the example layer 3 routing IPs for each of the VLANs and the clusters.

Table 40: Layer 3 Example Routing

Cluster	VLAN	Subnet	L3 Interface IP		VRRP Router IP	
			TOR(1)	TOR(2)	TOR(1) Master	TOR(2) Backup
Shared Edge and Compute	Management	172.29.1.0/27	172.29.1.1	172.29.1.2	172.29.1.3	172.29.1.3
	vMotion	172.29.1.32/27	172.29.1.33	172.29.1.34	172.29.1.35	172.291.35
	FT	172.29.1.64/27	172.29.1.65	172.29.1.66	172.29.1.67	172.29.1.67
	Comp	172.29.1.192/27	172.29.1.193	172.29.1.194	172.29.1.195	172.29.1.195
Management	Management	172.29.2.0/27	172.29.2.1	172.29.2.2	172.29.2.3	172.29.2.3
	vMotion	172.29.2.32/27	172.29.2.33	172.29.2.34	172.29.2.35	172.29.2.35
	FT	172.29.2.64/27	172.29.2.65	172.29.2.66	172.29.2.67	172.29.2.67
	vRA	172.29.2.192/27	172.29.2.193	172.29.2.194	172.29.2.195	172.29.2.195
Additional Compute	Management	172.29.3.0/27	172.29.3.1	172.29.3.2	172.29.3.3	172.29.3.3
	vMotion	172.29.3.32/27	172.29.3.33	172.29.3.34	172.29.3.35	172.29.3.35
	FT	172.29.3.64/27	172.29.3.65	172.29.3.66	172.29.3.67	172.29.3.67
	Comp	172.29.1.192/27	172.29.1.193	172.29.1.194	172.29.1.195	172.29.1.195

8.7.4   Management cluster

Although a single distributed switch can be deployed across various clusters, it is recommended to deploy unique distributed switches per cluster. Figure 46 shows the distributed switch for the management cluster.
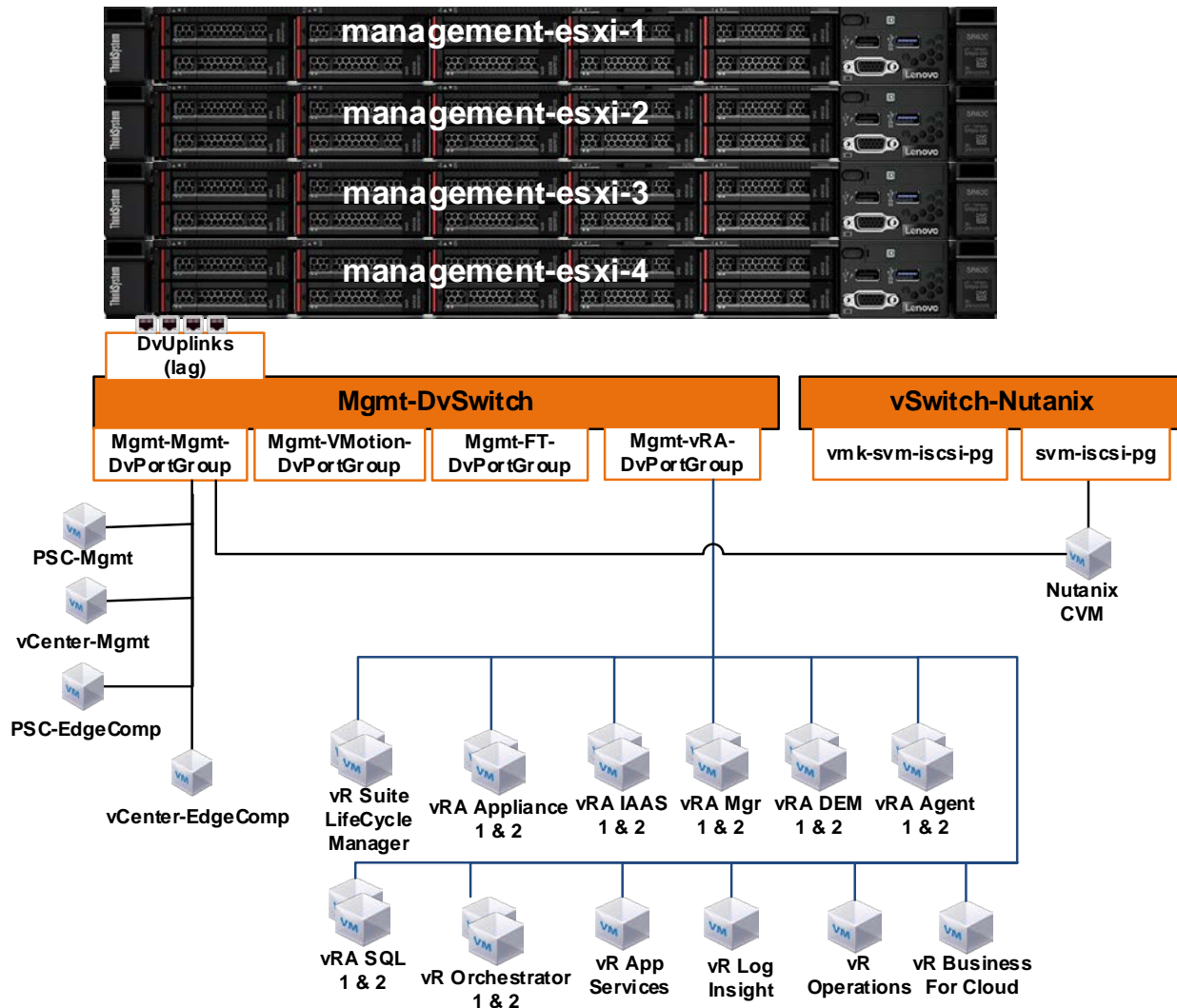

Figure 46: Management Cluster VDS

The infrastructure uses single SSO domain and all platform service controllers and vCenter servers are connected to this domain. All VMs in the management cluster can be configured on the same network segment. Separate vCenter and platform service controller instances are used for management cluster and shared edge and compute cluster. It is recommended that vCenter is deployed in a highly available configuration (depending on each customer's needs) which can result in multiple vCenters, multiple PSCs, and physical load balancers.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

8.7.5   Shared Edge and Compute cluster

Figure 47 shows the distributed switches for the edge and compute cluster. One DVS is used for accessing the Internet via a 1GbE port on the server. The other DVS is used for all of the edge and compute network flows.
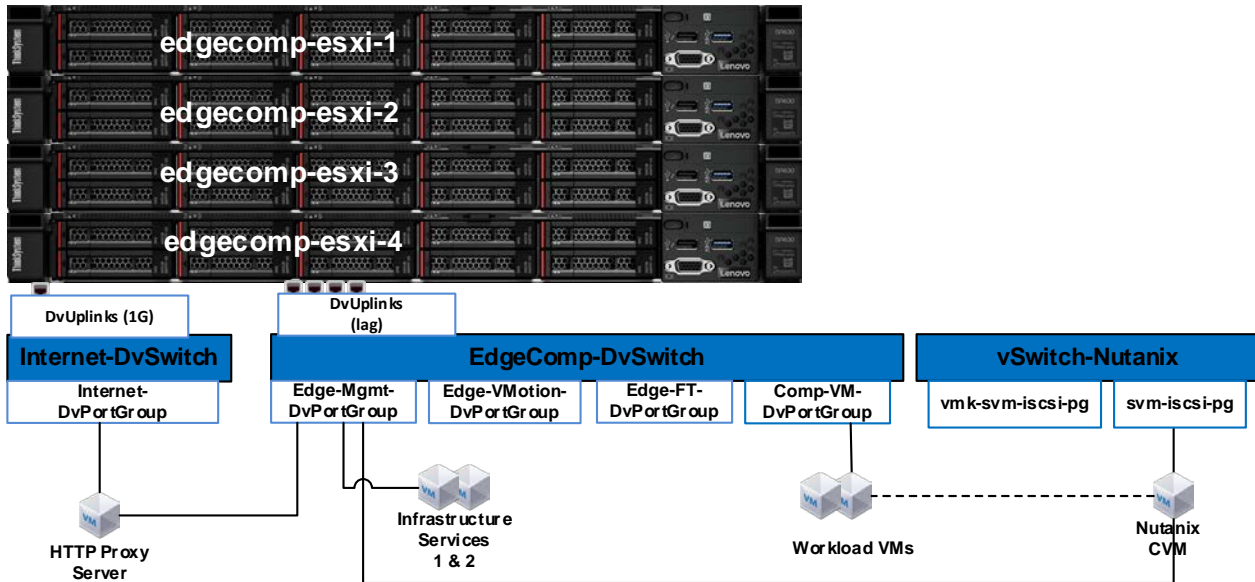

Figure 47: Shared Edge and Compute Cluster VDS

Infrastructure service VMs, such as Active Directory, DHCP, DNS, and NTP might exist in the customers' environment and these services are accessed through a clustered configuration, as shown in this example. However, if there is a requirement to virtualize these services within this environment, then they can be accessed through Edge-Mgmt-DvPortGroup, as shown in Figure 47.

The Comp-VM-DvPortGroup is used for the workload VMs.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

8.7.6   Compute cluster

Figure 48 shows the distributed switch for the compute cluster. Workload-specific DvPortGroups and VMs are not shown because they are highly dependent on the specific workloads. More than one compute cluster can be used to add more workloads.
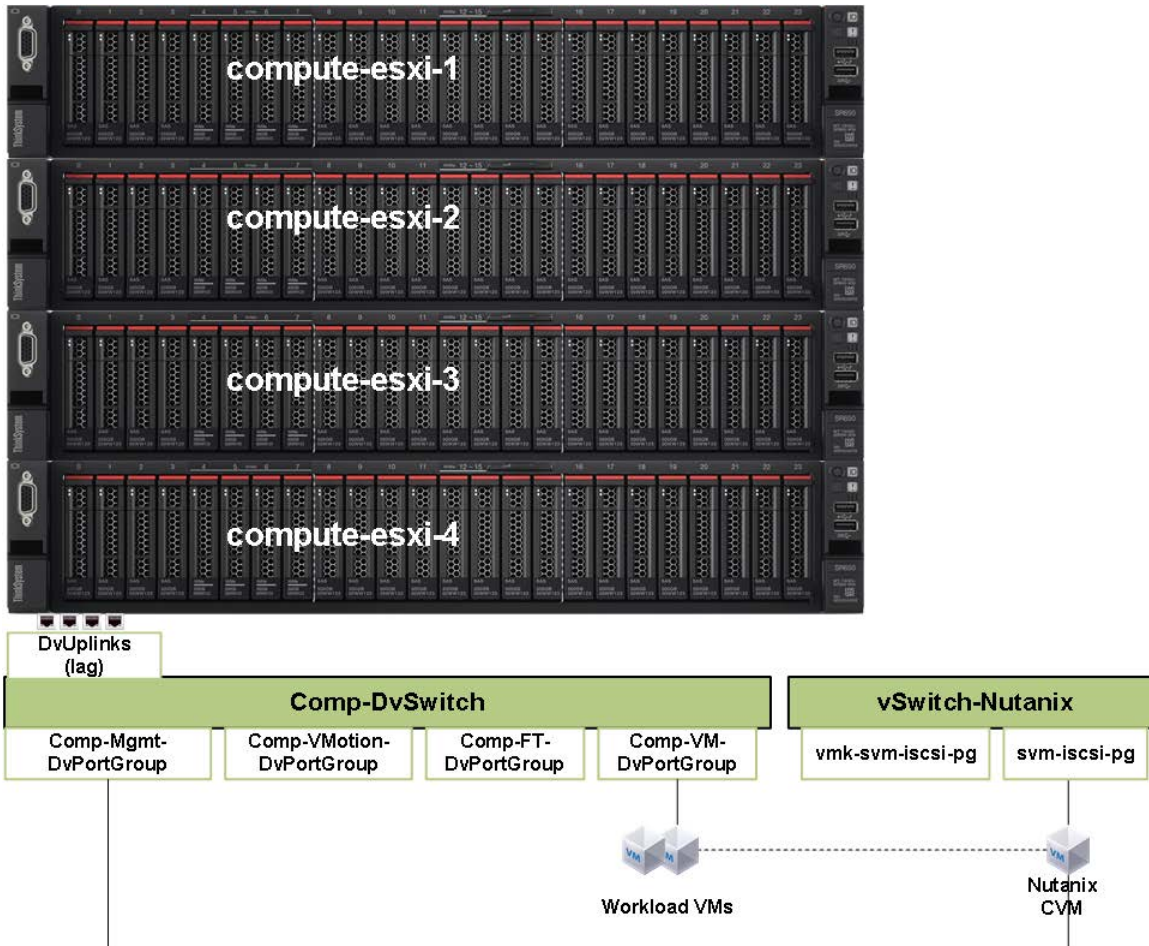

Figure 48: Compute Cluster VDS

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

9 Veeam Backup with ThinkSystem DE/DM/DG

The Veeam Data Platform is a collection of products and services that work together to provide a comprehensive data protection and management solution. Refer this reference architecture for more details https://lenovopress.lenovo.com/lp1769-reference-architecture-backup-solution-with-thinksystem-de-dm-dg-storage-veeam

9.1 Solution Overview

Figure 49 shows common backup and recovery architecture spanning across multiple sites and different type of storage systems with Lenovo DE/DM/DG series and Veeam platform.   The general recommended practice is to have more than one copy of backup to recovery from normal failover and disaster recovery scenarios. The backup storage architecture comprises of different media types all flash, hard drives and tape systems to provide flexible options and operationally cost efficient without impacting recovery point (RPO)and recovery time (RTO) objectives. The backup architecture can leverage either storage array-based replication or host based replication or combination of both for different applications backup performance requirements.
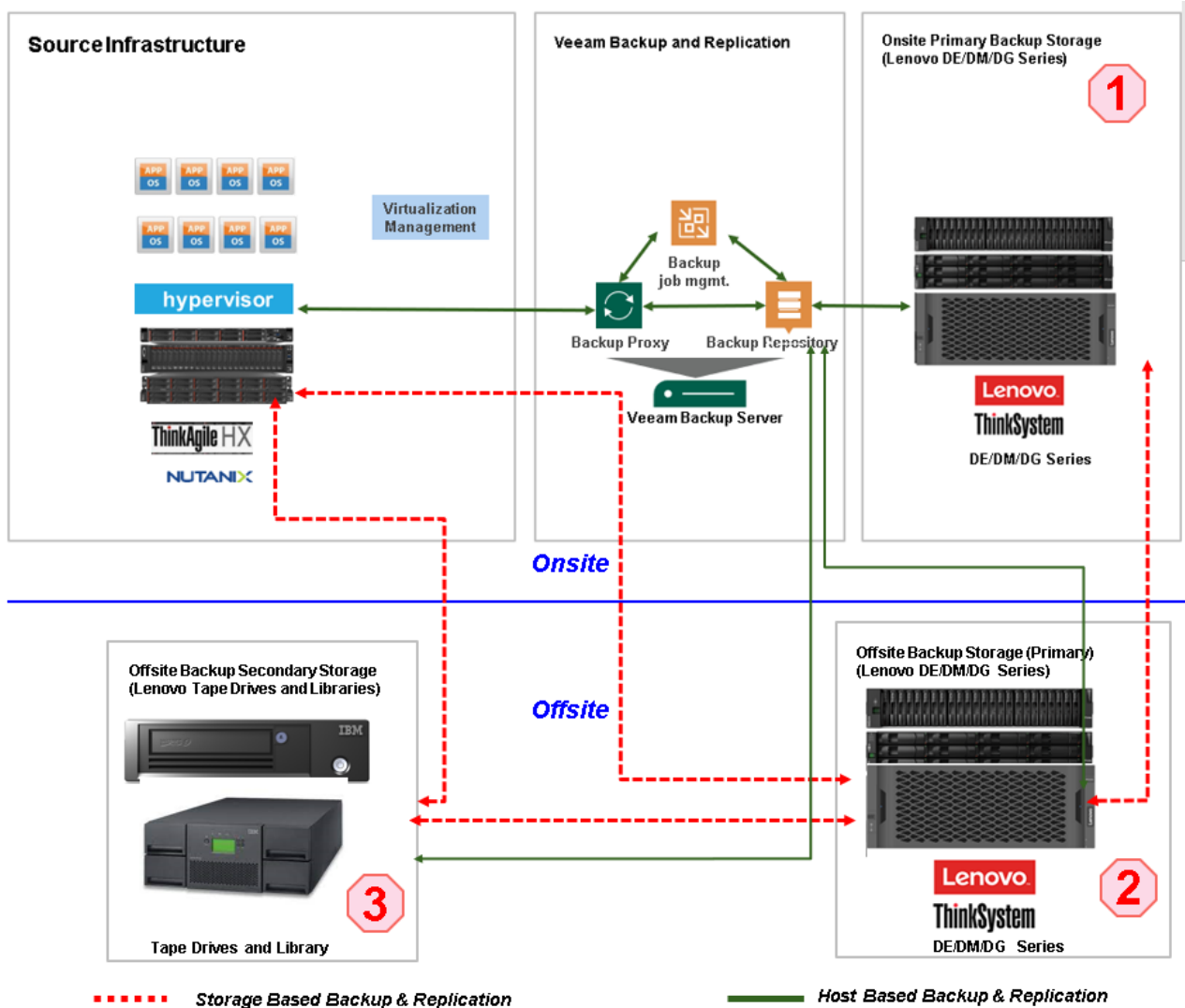

Figure 49. Onsite-Offsite Backup and Replication architecture with Lenovo DE/DM/DG Storage and Veeam

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

9.2 Component Model

Figure 50 is a layered view of the Veeam Backup and Replication solution that is mapped to the Nutanix virtualization infrastructure.
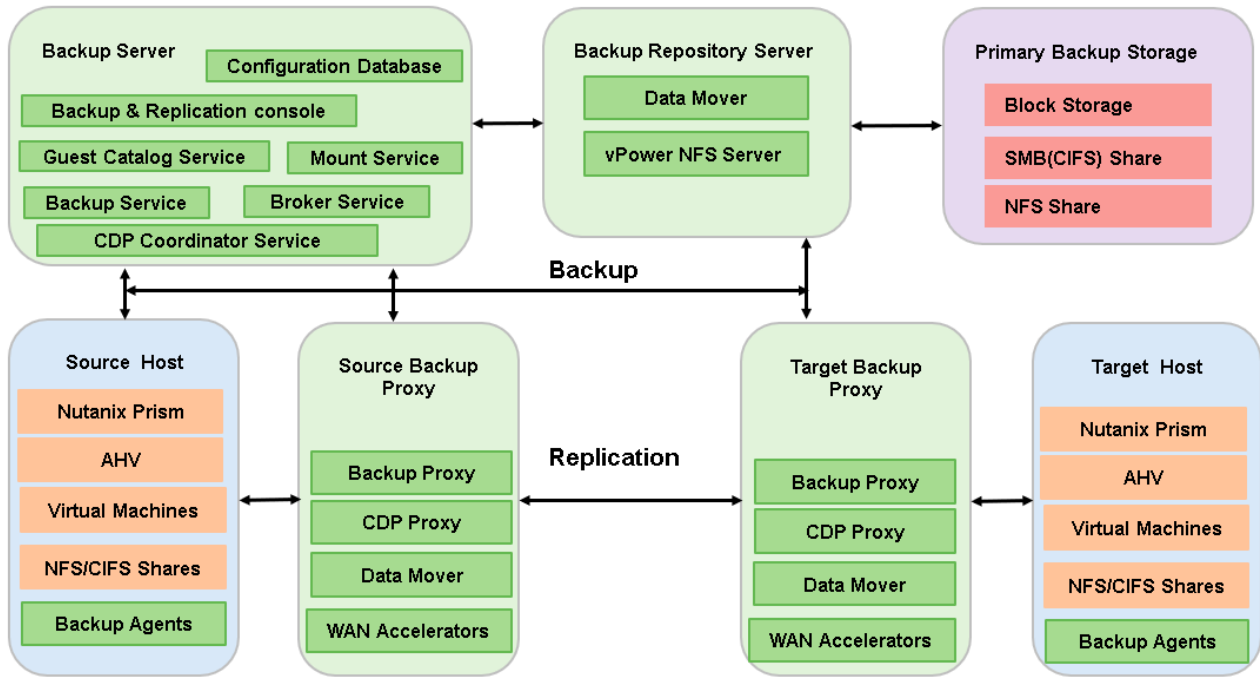

Figure 50. Component model with Veeam Backup and Replication

Veeam Backup and Replication solution infrastructure comprises the following components

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Veeam Backup and Replication Server	Coordinates backup and replication tasks, controls resource allocation and replica job scheduling. It contains the following services and components.
	Backup Service is a Windows service that coordinates all operations performed by Veeam Backup & Replication such as backup, replication, recovery verification and restore tasks
	Broker Service interacts with the virtual infrastructure to collect and cache the virtual infrastructure topology
	Guest Catalog Service manages guest OS file system indexing for VMs and replicates system index data files to enable search through guest OS files.
	Mount Service mounts backups and replicas for file-level access, browsing the VM guest file system and restoring VM guest OS files and application items to the original location
	CDP Coordinator Service communicates with vCenter, assigns continuous data protection (CDP) tasks and manages the infrastructure components involved
	Configuration Database stores data about the backup infrastructure, jobs, sessions and other configuration data. The database instance can be located on a SQL Server
	Backup & Replication Console provides the application user interface and allows user access to the application functionality.
	PowerShell Module is an extension for Microsoft Windows PowerShell that adds a set of cmdlets to allow users to perform backup, replication and recovery tasks through the command-line interface to run custom scripts to fully automate operation
Veeam Backup Proxy	It collects, transforms and transports VM data during the backup and replication process. It can be deployed on physical or virtual machines. It uses different services to perform the end-end operations.
	Data Mover: It copies VM data from the source storage in one of transport modes. It filters out zero data blocks, blocks of swap files and blocks of excluded VM guest OS files, compresses and deduplicates VM data blocks and moves them to the target-side Veeam Data Mover.
Veeam Backup Repository	The backup repository stores replica metadata that contains information on the read data blocks. Also stores backup metadata along with backup data.
Nutanix AHV Backup Appliance	It sits between the backup server and other components of the backup infrastructure. While the backup server administers tasks, the backup appliance processes jobs and delivers backup traffic.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Nutanix AHV Plug-in	It enables integration between the backup server and the Nutanix AHV backup appliance. Nutanix AHV Plug-in also allows the backup server to deploy and manage the Nutanix AHV backup appliance.
Source	The sources are Nutanix Prism, VMware vCenter, Physical servers, virtual machines and databases.
Target	Backup repository for backup data and the host where virtual machine replicas will be created and maintained in the ready-to-start state
Backup Storage	Block, file and object storage systems are used for backup.
WAN Accelerators	WAN accelerators are optional components in the backup infrastructure. You can use WAN accelerators if you replicate VMs over a slow connection or over WAN.

9.2.1 Veeam Backup for Nutanix AHV

Figure 51 shows componenents in Veeam Backup for Nutanix AHV architecture.

Veeam Backup & Replication(VBR) uses image level backup at block level and copies complete virtual machine object and restore can be done at virtual machine level, virtual machine files and file level recovery within virtual machine.
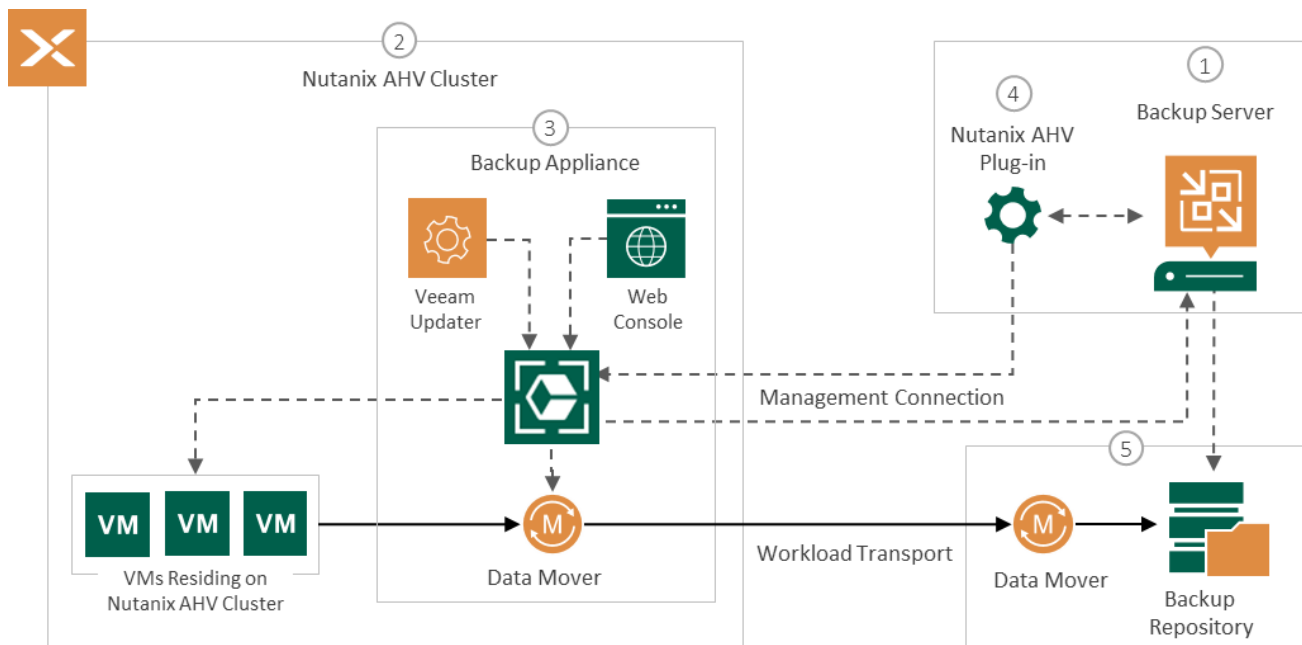

Figure 51. Veeam Backup for Nutanix AHV

Veeam Backup for Nutanix AHV does not install agent software inside VMs to retrieve data. To back up VMs, Veeam Backup for Nutanix AHV uses native Nutanix AHV capabilities. During every backup session, Veeam Backup for Nutanix AHV creates a snapshot for each VM or protected domain that were added to a job. The native snapshot is further used to create an image-level backup of the VM.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Veeam Backup for Nutanix AHV workflow: the Nutanix AHV backup appliance receives the backup job configuration from the web console and it starts a backup job and forwards the backup session data to the backup server. It then connects to the Nutanix AHV cluster using the Nutanix REST API and creates the backup snapshots of the protected VMs or domains that were previously added to the backup job. It creates a volume group on the Nutanix AHV cluster, mounts the VM disks over iSCSI and retrieves the VM data on the block level. The Veeam Data Mover service compresses and deduplicates the VM data to save bandwidth then forwards the backup data to the backup repository in native Veeam format.
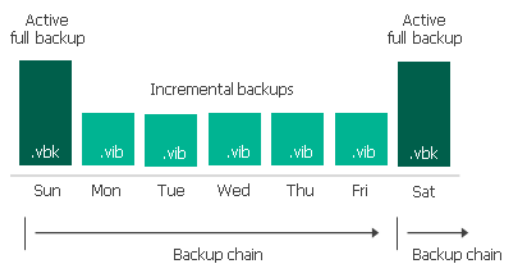
To configure and manage backup for Nutanix AHV infrastructure, Refer Veeam user guide for Nutanix AHV https://helpcenter.veeam.com/docs/vbahv/userguide/overview.html?ver=40

9.2.2 Backup Methods and Retention Policies

Veeam supports different backup methods and retention polices to address all sort of backup requirements by efficiently planning for storage capacity and fast recovery. Table 41 lists different backup types supported by Veeam platform.

Table 41: Veeam supported backup methods

Backup Method	Description
Forever forward incremental (FFI) 	The backup chain consists of the first full backup file (VBK) and a set of forward incremental backup files (VIBs). It stores only one full backup file and removes incremental backup files once the retention period is exceeded. It is the fastest compared to other methods because the first available restore point is always a full backup.
Forward incremental (FI) 	The backup chain consists of multiple full backup files (VBKs) and sets of forward incremental backup files (VIBs) following each full backup file. It is split into shorter series so this lowers the chances of losing the backup chain completely. Restore from backup files created using the FI method is the most optimal in time compared to other methods
Reverse Incremental Backup* 	The backup chain consists of the last full backup file (VBK) and a set of reverse incremental backup files (VRB) preceding it. The most recent restore point in the backup chain is always a full backup and it makes restoring virtual machine immediately without any additional processing.

Active Full Backup	It compresses and deduplicates whole VM and stores it to the full backup file (vbk). The latest full backup file is new starting point for all incremental backups.
Synthetic Full Backup	It is identical to a regular full backup and it synthesizes a full back up from data you already have in the backup repository rather than from source VM

* Reverse Incremental Backup is supported technically in Veeam currently, but it may not be available in future releases.

Short Term Retention Policy

Whenever the number of restore points or days is exceeded, the backup chain makes room for most recent restore point. The full backup also moves one step forward with the next incremental backup in the backup chain. The earliest incremental backup file is deleted immediately when it is outdated, and the full backup chain will be removed only after the last incremental backup file in the chain becomes outdated. All the incremental backup methods apply relevant short term retention policy based on the restore points and full backup requirements.

Long Term Retention Policy (GFS)

The long-term or Grandfather-Father-Son (GFS) retention policy allows to store for longer period and backup can be done weekly, monthly, and yearly. Veeam Backup and Replication follows different algorithms based on different type of GFS policies to take full backup based on the retention period.

9.2.3 Veeam Backup Copy

Backup copy or secondary backup storage helps to recover from disaster recovery scenarios and when any issue with primary backup storage. Veeam Backup copy contains the same data and format as the primary backup, and it is created using jobs and retention settings can be applied with desired number of restore points. All backup jobs can be linked to backup copy jobs to store on secondary backup storage.

9.3 Backup System Design

Designing robust backup solution is based on the Recovery Time Objectives (RPO) and Recovery Time Objectives(RTO) and backup storage infrastructure must be sized to meet these objectives. These objectives are purely depending on applications or workloads and compute and storage need to be assessed appropriately. The following steps will provide guidance to make decision and refer operational model for detailed information.

1. Decide RPO and RTO for applications and virtual machines
2. Decide number of onsite and offsite copies

3. Group applications based on RPO/RTO

4. Decide Backup Method and Backup Window

5. Decide Total storage capacity and Retention Period

6. Choose VM snapshot vs Storage Snapshot

7. Choose storage system and drives to meet RPO/RTO

8. Define volumes for different backup groups and RAID levels

9. Define number of backup jobs

10. Evaluate sharing compute and network from source infrastructure

11. Decide compute for Veeam backup components

9.4 Operational Model

This section describes the options for mapping the logical components of backup and replication solution onto hardware and software. The "Operational model scenarios" section gives an overview of the available mappings and has pointers into the other sections for the related hardware.

9.4.1 Operational model scenarios

Figure 52 shows the following operational models (solutions) in Lenovo Backup solution with Veeam for enterprise and small-medium business (SMB) with dedicated and shared deployment and different performance classes. Backup is necessary for mission critical applications for any business and having more copies on different locations and different media prevents data loss from circumstances ranges from planned maintenance to disaster scenarios. Having more copies are good for large business and it brings additional operational costs for backup storage and compute depends on RPO and RTO objectives. Table 42 shows typical backup requirements for SMB and enterprises.

Table 42: Backup for different business sizes

Type of Business Size	Total Backup Size	Number of Mission Critical Applications	Number of Sites	Backup Model
Small	Up to 100 TB	1-2	1	1 Onsite copy + 1 Archival copy
Medium	101-500TB	2-5	1-2	1 Onsite copy + 1 Offsite copy + 1 Archival Copy
Enterprise	500TB+	5-10 or more	3+	1 Onsite + 1 or more Offsite copies + 1 Archival Copy

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

The lower RTO improves availability by providing faster recovery and it does require more compute, faster connectivity, and all flash storage systems. RTO is a critical factor for mission critical applications irrespective of company size. Dedicated compute for backup server for small deployments is not required for small and medium deployment and the source infrastructure should be able to accommodate and scale to support backup software computing requirements. The performance of the source storage systems also important to achieve lower RTO. Figure 52 shows operation model with different Lenovo servers and storage systems to meet backup RTO objectives.
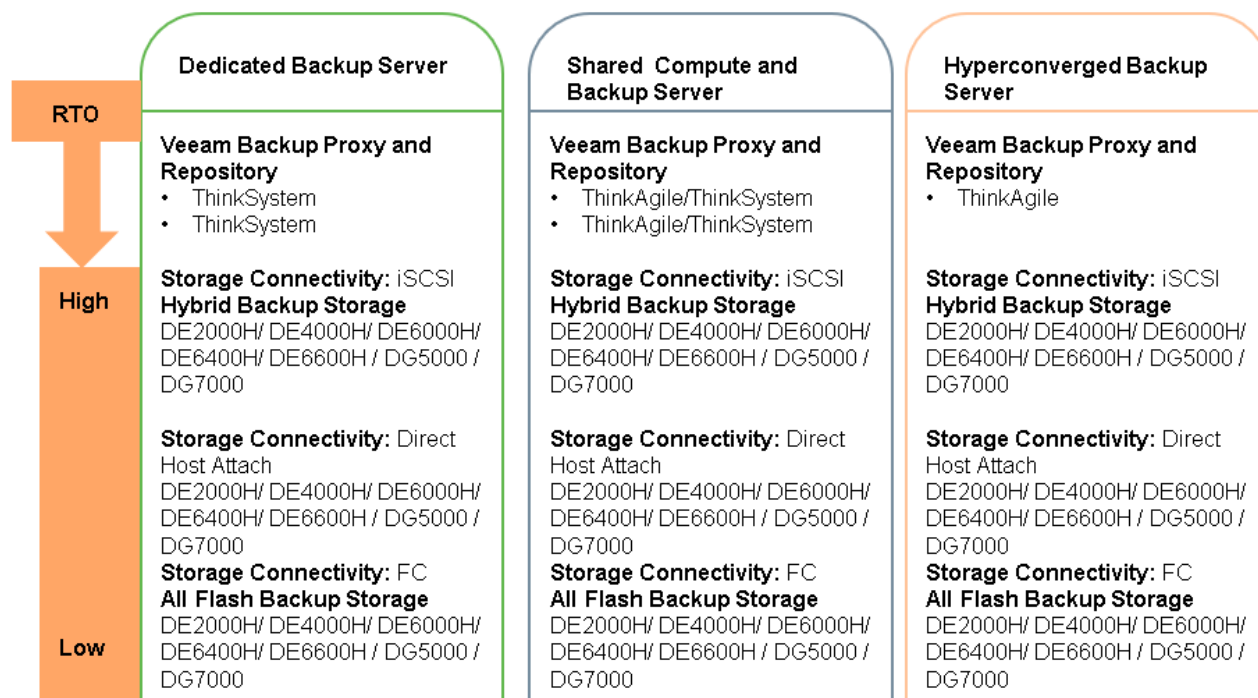

Figure 52: Backup Operational model scenarios

9.4.2 Primary Backup Storage with ThinkSystem DE/DM/DG Series

Lenovo ThinkSystem DE/DM/DG series provide various options to support backup storage requirements. Usually, HDDs are recommended for backup due to the cost factor. However, SSDs and NVMe drives can be used for deployments which require shorter recovery time. The primary backup storage can coexist on the same data center, but it is recommended to keep one backup copy on separate physical location than the source environment. Table 43 below lists Lenovo storage systems support for backup solutions.

Table 43: Lenovo ThinkSystem Storage for SMB Backup

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Backup Storage	All Flash	SnapMirror	Protocol	Expansion	Connectivity	Maximum Capacity
DE2000H 2U 12 LFF	No	No	Block		Fibre Channel, iSCSI	1.47 TB
DE4000H 2U 12 LFF / 4U 60LFF	No	No	Block	DE120S 2U12 LFF, DE600S 4U60	Fibre Channel, iSCSI	3.07 PB
DE6000H 2U 12 LFF / 4U 60 LFF	No	No	Block	DE120S 2U12 LFF, DE600S 4U60	Fibre Channel, iSCSI	4.2 PB
DM3000H 2U 12LFF	No	Yes	Block/NFS		Fibre Channel, iSCSI, Ethernet	2.3 PB Cluster (27.6 PB NAS or 13.8 PB SAN)
DE4000F 2U24 SFF	Yes	No	Block		Fibre Channel, iSCSI, Ethernet	1.440 PB
DE6000F 2U24 SFF	Yes	No	Block		Fibre Channel, iSCSI, Ethernet	1.84 PB
DM5000F 2U24 SFF	Yes	Yes	Block		Fibre Channel, iSCSI, Ethernet	2.2 PB Cluster (26.5 PB NAS or 13.2 PB SAN)
DG5000	QLC	Yes	Block/NFS	DM240N	Fibre Channel, iSCSI, Ethernet	8.8 PB
DG7000	QLC	Yes	Block/NFS	DM240N	Fibre Channel, iSCSI, Ethernet	17.6 TB

9.4.3 Dedicated Backup Server with ThinkSystem Servers

For the enterprise performance and availability requirements, dedicated Lenovo ThinkSystem SR630 V2 nodes are recommended to host Veeam backup proxy and repository server components. In many circumstance the source rack might not have enough space and also source virtual machines and applications need to reserve for compute overhead and deploying Veeam components brings additional management complexities during maintenance and upgrade operations. So, it is recommended to separate Veeam components to dedicated servers to isolate from source and to backup different workloads running across converged and hyperconverged compute clusters in the data center. The ThinkSystem servers require

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

connectivity to external storage for Veeam management components and it can leverage backup storage as well.

9.4.4 Shared Compute and Backup Server

For small and medium size source environments, it is expensive to have dedicated servers to host Veeam backup components. Also, many of the source workloads are being deployed on hyperconverged servers and having separate shared storage for only Veeam management components it not a viable solution. The Veeam backup components can coexists on the source cluster without impacting source applications. Deploying Veeam proxy and backup server closer to source virtual machines provide better performance. This deployment is well suited for storage array-based snapshots where hosts are freed from resources required for virtual machine based snapshots.

9.4.5 Hyperconverged Backup Server

Veeam backup components require considerable amount of compute, memory and storage and it increases when the backup storage and number of backup jobs increase. Veeam provides additional features such as orchestrator, monitoring platform and container backup support and it requires additional resources. For 200TB backup requirements, the complete Veeam management stack mentioned above requires 32 cores, 64 GB and 3TB storage and it can be managed with either 2 node ThinkAgile VX 2000 VMware vSAN systems or ThinkAgile HX 2000 Nutanix series or ThinkAgile MX 1021 AzureStack HCI systems are perfect fit and these system can also host other infrastructure management software such as vCenter, vCloud Foundation and others.

9.4.6 Host Attached Storage

Lenovo ThinkSystem DE/DM series supports connecting storages to local port of the hosts. It reduces additional switches and hops in the network and it is an ideal solution for onsite backup model. This design is perfect fit for dedicated backup server with ThinkSystem servers and hyperconverged systems and DE/DM storage can be placed on the same rack or within reasonable distance within data center if feasible. The support is available for 1GbE/10GbE/100GbE connectivity depends on the storage systems.

9.5 Lenovo ThinkSystem DE/DM/DG Series Storage Arrays

For backup requirements the Lenovo DE/DM/DG series storage provides a wide range of options, supporting high-capacity, high performance and scale-out options as well as rich set of software capabilities such as storage management, block/file level support, encryption, deduplication, and so forth

For information and specifications of the Lenovo ThinkSystem DE,DM and DG series storage arrays, see the following Lenovo Press publications:

Lenovo ThinkSystem DM Series Unified Storage Arrays

Lenovo ThinkSystem DE Series Storage Arrays

Lenovo ThinkSystem DG Series

The IBM TS2270, TS2280, and TS2290 Tape Drives are excellent tape storage solutions for small and midsized businesses that require backup and low-cost archival data storage. These tape drives incorporate the latest generation of industry leading LTO technology that helps organizations to handle the growing data demands for backup and archival.

The TS2270 Tape Drive provides a physical capacity of 6TB of native storage capacity (15TB with 2.5:1 compression) and a transfer rate of up to 300MBps with 6Gbps SAS interface connectivity. The TS2280 Tape Drive claims a physical capacity of 12TB (up to 30TB with 2.5:1 compression).

The TS2290 Tape Drive provides even higher capacity than the previous generation, with a native storage capacity of 18TB (up to 45TB per cartridge with 2.5:1 compression).

Refer Lenovo datasheet for more details IBM TS2270, TS2280 & TS2290 Tape Drives

IBM TS4300 Tape Library is a high-density, highly scalable, easy-to-manage solution designed to keep data securely stored long-term, while helping reduce the costs associated with data center space and utilities. The supported tape drives are IBM LTO Ultrium 9, 8, 7, or 6 Tape Drives with SAS or Fibre Channel host connectivity interfaces and with LTO Ultrium 9 technology, the TS4300 Tape Library provides up to 5.04 PB of native backup storage. Refer Lenovo press paper for more details IBM TS4300 Tap Library

9.6 Backup Management Servers

For medium to large environment, the backup management components should have dedicated compute servers so that it will not impact application workloads performance if it is sharing the compute. However, if there are many clusters to be backed up and backup window is not impacting application performance, then shared compute can be used and it is recommended to keep proxy closer to applications which require back up. Veeam Backup and Replication server components can run on ESXi along with virtualization and cloud management software such as vCenter server and vCloud Foundation. The table below shows minimal configuration requirements for Veeam components. Table 44 shows the minimal system requirement to install Veeam backup and server components.

Table 44: Veeam Management components system requirements

Management VM	Virtual processors	System memory	Storage	OS	HA needed
Veeam Backup and Replication Server	8	8 GB	60 GB	Windows Server	No
Backup Proxy	8	24 GB	60 GB	Windows Server or Linux	Yes
Backup Repository Server	12	24 GB	100 GB	Windows Server or Linux	No

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Management VM	Virtual processors	System memory	Storage	OS	HA needed
Nutanix AHV backup appliance	4	4 GB	100 GB		

9.7 Sizing Guidelines

The backup infrastructure consists of primary and secondary backup storage systems and compute required for running Veeam backup components.

9.7.1 Storage Capacity

The factors below are considered to estimate total back up storage consumption.

Source Data: The source can be virtual machines disks, images, files, and databases and the actual size of each protected component needs to be considered for full backup. In hyperconverged environments, the usable capacity is calculated and varies based on RAID level and it is recommended to estimate based on the virtual machine actual disk requirements instead of considering total storage on the source environment.

Number of Virtual Machines: The size of virtual machines cannot be expected same in large environments. However, it can be grouped based on the same disk size to estimate target backup capacity easily.

Daily Change Rate: Estimation daily change rate for each protected source helps to size storage appropriately. It can be started with 5-10% conservatively for any applications and then it must be monitored and revise sizing appropriately. The more the change rate, it does require frequent backup to avoid data loss.

Retention Scheme: How many copies are retained for how long an import factor. Many backup can be designed to have latest copy but many critical applications require more retention period based on industry and regulatory requirements. It is recommended to have multiple retention points created daily, weekly, monthly, and yearly.

Compression Ratio: The backup storage footprint can be reduced by increasing compression ratio and it requires processing capacity based on the compression level chosen.

Table 45 lists the backup storage needed for different size of source environment. The sizing is based on forward incremental with weekly full backup method with 50% compression ratio.

Table 45: Primary Backup Storage Sizing - Forwards Incremental with Weekly Full

Source Capacity	Daily change rate (%)	Yearly growth (%)	Growth Scope	Retention Points	Reduction(%)	Estimated Backup Storage	Estimated Backup Copy
10 TB	10%	10%	3 years	Daily 12 Weekly 0 Monthly 0	50%	20.39 TB	20.39 TB

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

				Yearly 0			
10 TB	10%	10%	3 years	Daily 12 Weekly 2 Monthly 2 Yearly 2	50%	41.05 TB	41.05 TB
20 TB	10%	10%	3 years	Daily 12 Weekly 2 Monthly 2 Yearly 1	50%	80.14 TB	80.14 TB
50 TB	10%	10%	3 years	Daily 12 Weekly 2 Monthly 2 Yearly 1	50%	192.68 TB	192.68 TB
100 TB	10%	10%	3 years	Daily 12 Weekly 2 Monthly 2 Yearly 1	50%	379.28 TB	379.28 TB
250 TB	10%	10%	3 years	Daily 12 Weekly 2 Monthly 2 Yearly 1	50%	924.8 TB	924.8 TB

9.7.2 Backup Window

Choosing right backup window is critical for any environment as it is tightly coupled with recovery point and time objectives and it can impact running applications if it is not planned well. It has be designed considering RPO/RTO objectives for each application and virtual machines. Each backup method has it own advantages and drawbacks and we need to design purely based on the recovery objectives and tolerance factors. Many businesses can follow 4-8 hours night backup window for normal applications, but critical applications require continuous backup along with multiple backup copies across different locations. Table 46 shows different design factors which should be taken care while planning backup window.

Table 46: Backup Window influencing factors

Factors	Design Considerations
Full Backup	Full back up takes more time, and it depends on number of full backup required Daily/Weekly/Monthly. It will occupy more storage

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

	space as well. Many critical applications require full back up during weekdays. If the source is large, then backup window should be increased to fit.
Incremental Backup	Since full backup takes more time and it is advisable to use combination of full and incremental back up. Incremental back up require less space and consumes less resources and can be taken frequently.
Recovery Point Objective (RPO)	Having frequent backup and restore points help to improve RPO. Restoring from full back is quicker than recovering from incremental backups. During recovery time the RPO would reduce as backup cannot be done when restore is in progress and restore takes priority than backup.
Recovery Time Objective (RTO)	Restoring backup is dependant on the backup size and the storage media and network speed. Storage array-based backups and restore are faster than software-based recovery which involves traverse through backup chain to rebuild required restore point.
Parallel Backup Proxies	Configuring more concurrent backup proxies to run backup for different application would reduce the backup time and hence backup window.
Storage Array based replication	Storage array based replication provide faster backup and recovery and it can reduce backup window.
Storage System Performance	Storage systems maximum throughput capacity needs to be considered. Either we have to scale out storage systems to support within backup window or adjust backup window to fit storage system performance.

9.7.3 Storage System Performance

The selection of storage model, drive, host connectivity and storage controller should be based on RPO and RTO objectives and backup window. If RTO is high, then it is recommended to use SSD and Fibre Channel connectivity and storage array-based snapshots for faster backup and recovery. The backup NAS storage is recommended for applications and workloads require to access large number of files. The storage system performance may degrade after 80% utilization and the storage capacity should be sized considering headroom and performance limitations.

The storage selection should address RPO/RTO objectives, recovery speed and backup window. The storage system should support enough concurrent backup jobs without impacting backup window. The backup operation is sequential and full backup takes longer time based on the disk size of source virtual machine or applications. All Flash is good for mission critical applications which require RTO in few minutes range. Storage array-based snapshot and replication is recommended in these scenarios as it avoids additional CPU

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

cycle for source system and network latencies. Also, onsite backup provides better recovery time than offsite backup due to network latency. Figure 53 shows different DE systems supported throughput.

Operation	DE200H	DE4000H	DE6000H	DE6400H	DE6600H	DE4000F	DE6000F	DE6400F	DE6600F
Sequential Read	3 GB/s	9.2 GB/s	21 GB/s	20 GB/s	44 GB/s	9.2 GB/s	21 GB/s	20 GB/s	44 GB/s
Sequential Write	0.9 GB/s	2.7 GB/s	7 GB/s	7 GB/s	12.5 GB/s	2.7 GB/s	9 GB/s	7 GB/s	12.5 GB/s

Figure 53: Lenovo DE storage system throughput

Lenovo ThinkSystem DG series are with QLC all flash drives and it provides better price, performance and power consumption over HDD based storage systems. DG5000 and DG7000 systems are optimal solution for backup to satisfy RPO and RTO objectives and capacity requirements. Table 47 below shows comparison of Vdbench performance on DM3000H and DG5000 with 512k and 1MB block sizes. The DG5000 provides 4:1 compression ratio and can reduce backup storage footprints with improved throughput and much reduced sub seconds latency.

Table 47: Vdbench performance results for DG5000 and DM3000H

Storage	DG5000		DM3000H	
Number of Drives	24x 15.3TB QLC SSD		12x 4TB HDDs + DM600S 120x 10TB HDDs	
Total Capacity	367.2 TB		1248 TB	
Number of LUNs	10		10	
Host Connectivity	4 x 10 Gb iSCSI		4 x 10 Gb iSCSI	
Compression Ratio	4:1		N/A	
Que Depth	100/LUN		100/LUN	
Block Size	1M Seq Write	512k Seq Write	1M Seq Write	512k Seq Write
Throughput	1.2MB/s	1.2MB/s	1.12MB/s	1.12 MB/s
Latency	1.4ms	1.1ms	98ms	62ms
Total Watts	645	644	2118	2109

9.7.4 Compute

Veeam backup components consumes considerable resources, and these should be sized and scaled appropriately to meet availability and recovery objectives. These components are sized based on the number of backup jobs and backup storage requirements. It is recommended to use dedicated servers to avoid resource contention in the source environment. However if the source is with latest generation high performance processors and it can be run along with workloads.

Backup Proxy

The backup proxy transfer data from source to Repository server and it should have enough cores to process. It is recommended to allocate at least 1 core per 100MB/s transfer and split backup to multiple tasks to increase concurrency and reduce time. Each proxy task require 2GB memory and proxies can be distributed

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

evenly across nodes in the source cluster to balance performance and prevent it from reserving compute for source workloads. More sizing examples can be found here Veeam Backup Proxy for VMware

Backup Repository Server: The backup repository server is sized based on the number of proxy servers and backup copy requirements. For each proxy, the repository server requires one third processing power required for the proxy. The additional processing power required for secondary backup is depends on the total backup storage capacity. Also, it is recommended to keep repository server closer to proxy servers to reduce latency and improve backup and recovery time. Refer design guidelines here Veeam Backup Repositories

9.7.5 Network Bandwidth

The network requirement needs be sized based on recovery speed, storage connectivity options, storage system throughput limit and backup window requirements. It is recommended to use dedicated 10Gbps/25Gbps links for backup traffic. The higher backup and recovery speed provides better RPO/RTO, and it should be at least 10Gbps. If the backup storage is connected through iSCSI or NAS, it is recommended to use more parallel links between proxies and repository servers and between repository server to backup storage systems due to speed constraints.

9.8 Deployment example

9.8.1 Offsite Backup with DE Storage and Dedicate ThinkSystem

In this deployment model the backup storage DE2000H and Veeam Repository server is deployed at the secondary site. All other Veeam backup management components deployed on the source site. The backup repository server uses dedicated ThinkSystem servers.
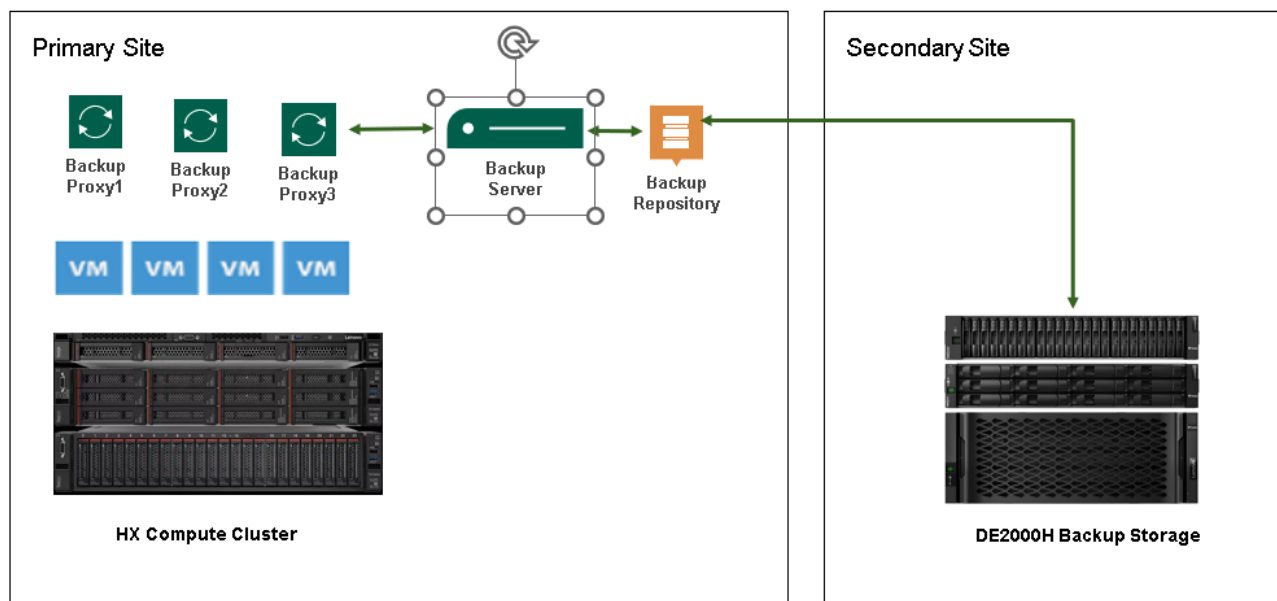

Figure 54: Offsite Backup with DE and ThinkSystem

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

10 Nutanix Cloud Platform

The Nutanix Cloud Platform(NCP) is a set of software packages to build, manage and monitor hybrid and multicloud infrastrutrue and applications across on-premises and hyperscalers. The platform provides diversied options to build private cloud, infrastructure as a services, storage cloud, SaaS services for database and VDI.

10.1 Solution Overview

The Nutanix Cloud Platform(NCP) suite of software provides complete functionality to scale deployments on-premises and public cloud and address infrastructure challenges to deliver highly resilient and secured services for different workloads. The following software are included in the platform.
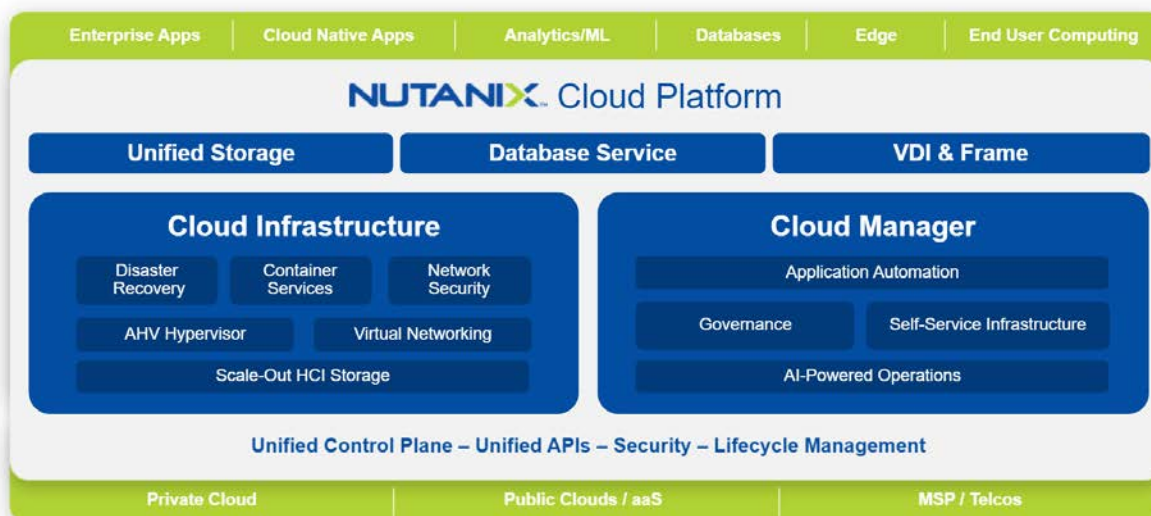
Figure 55: Nutanix Cloud Platform for Hybrid Cloud

Nutanix Cloud Infrastructure (NCI) provides a complete software stack to unify your hybrid cloud infrastructure including compute, storage and network, hypervisors and containers, in public or private clouds; all with built-in resilience, self-healing, disaster recovery capabilities, and security.

Nutanix Cloud Manager (NCM) adds the control plane for provisioning, operating, automating, and governing workloads across clouds. NCM is an end-to-end consumer-grade management, operations, and automation solution for every environment – from a single virtualized datacenter to a distributed hybrid multicloud environment. NCM brings unprecedented simplicity by combining aspects of administration, reporting, intelligent infrastructure, self-service and application automation, application provisioning, with cost and security governance.

Nutanix Unified Storage provides an alternative to traditional storage servers for unstructured files and objects stores. It addresses exponential data growth challenges and simplifies data management with a single platform of shared resources for unstructured data. NUS is a distributed and software defined storage solution that provides the scale organizations need to serve any workload anywhere - from the core to the edge, to the cloud. A single point of management for all storage resources eliminates complexity of multiple interfaces and consumer grade design enables non-storage experts to handle most day-to-day storage and data management tasks. Intelligent analytics integrated into the solution provide data visibility and deep insights for

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

governance and security of data. NUS can be deployed as dedicated storage clusters or as part of AOS clusters based on customer's needs and use case.

Nutanix Database Service (NDB) enables small, nimble teams to easily manage large fleets of SQL Server, Oracle Database,MySQL, MongoDB and PostgreSQL databases using a single API and console, freeing their time and skills to focus on mission critical databases. With Nutanix Database Service, developers can self-service their database needs through a simple database-as-a-service experience.

Nutanix Virtual Desktop Infrastructure (VDI) & Nutanix Frame enable a wide range of end user computing solutions that deliver virtual apps and desktops to users worldwide from public, private and hybrid cloud infrastructure. Nutanix VDI is a per-user licensing option for Nutanix Cloud Infrastructure that simplifies capacity planning by matching the infrastructure cost model to that of the end user computing platform (e.g. Citrix CVAD , VMware Horizon, and Nutanix Frame). This makes it possible to arrive at a total per-user cost top to bottom, removing the need for guesswork in sizing the infrastructure to match individual user needs.

10.2 Component Model

Figure 56 shows layered view of different software and services in the Nutanix Cloud Platform solution.
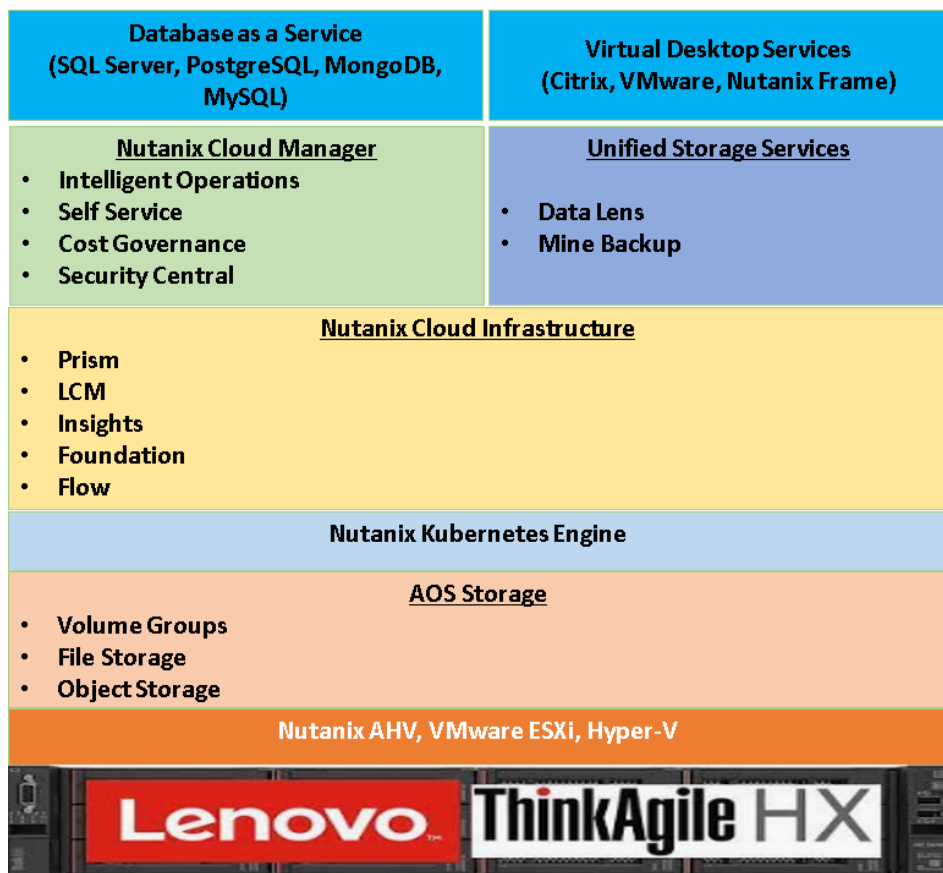

Figure 56: Nutanix Cloud Platform components with ThinkAgile HX

10.2.1 Nutanix Cloud Infrastructure

AOS	AOS is the operating system of the Nutanix Controller VM–the VM that runs in the hypervisor to provide Nutanix-specific functionality. It provides core functionality used by workloads and services running on the platform. It contains several data services and features for data protection, space efficiency, scalability, automated data tiering, and security.
AHV	The Acropolis Hypervisor (AHV), based on the proven Linux KVM hypervisor. It provides enhanced security, self healing capabilities based on SaltStack, and enterprise-grade VM management.
Prism Central Admin Center	For more information see Prism Central Admin Center Guide
Life Cycle Manager (LCM)	Tracks and updates software and firmware versions of all entities in the cluster.
Foundation Central	Create clusters from factory-imaged nodes and reimage existing nodes that are already registered with Foundation Central, remotely from Prism Central.
Nutanix Kubernetes Engine	Enterprise Kubernetes edtion from Nutanix for containers and cloud native applications.
Flow Network security	Provides Application centric visibility and enterprise grade network microsegmentation

10.2.2 Nutanix Cloud Manager

The Nutanix Cloud Manager components are enabled as services in Prism Central to access respective fuctionalites.

Componenet	Description
Self-Service	Provides infrastructure and application deployment and automation and life cycle management and role based governance. For more information see Self Service Admin Operations Guide
Intelligent Operations	Solution for end-to-end operations, monitoring and planning IT operations that combines aspects of administration, reporting, and intelligent automation. For more information see Intelligent Operations Guide
Nutanix Security Central	Provides automate security operations, discover vulnarabilites and recommend remediation, regulatory compliance
Cost Governance	Cost Governance is a SaaS offering by Nutanix that helps cloud-focused organizations gain visibility into their cloud spending across

	multiple cloud environments. Cost Governance provides comprehensive visibility and analytics into cloud consumption patterns and a one-click feature for optimizing cloud costs

10.2.3 Nutanix Unified Storage

Nutanix Volumes	Ssoftware-defined block storage solution that exposes storage resources directly to virtualized guest operating systems or physical hosts using the iSCSI protocol or, in the case of AHV VMs specifically, through the guest's native storage adaptor.
Nutanix Files	Scale-out file storage solution that provides a repository for unstructured data, such as home directories, user profiles, departmental shares, application logs, backups, and archives. Nutanix It provides Server Message Block (SMB) and Network File System (NFS) file services to clients. Nutanix Files instances are composed of a set of file server VMs (FSVMs). Files requires at least three FSVMs running on three nodes to satisfy a quorum for high availability
Nutanix Objects	Object store service S3 compatible REST API
Data Lens	Cloud-hosted analytics and monitoring service that provides data analytics and security solution, including ransomware protection for file servers running on Nutanix Files and object stores running on Nutanix Objects. For more information see Data Lens User Guide
Mine Integrated Backup	Backup solution, where only backup component VMs run on the Mine cluster and the cluster storage are used to store backup workloads. The solution is integrated with Veeam and HYCU backup products.

10.2.4 Nutanix Database Service

Nutanix Database Service (NDB) is a software suite that automates and simplifies database administration, bringing simplicity and invisible operations to database provisioning and life-cycle management. With one-click database provisioning, copy data management (CDM), database protection, and patching, NDB enables database administrators to provision, clone, and refresh their databases to any point in time. NDB's API-first architecture can easily integrate with your preferred self-service tools, and every operation has a unique ID that's fully visible for auditing.
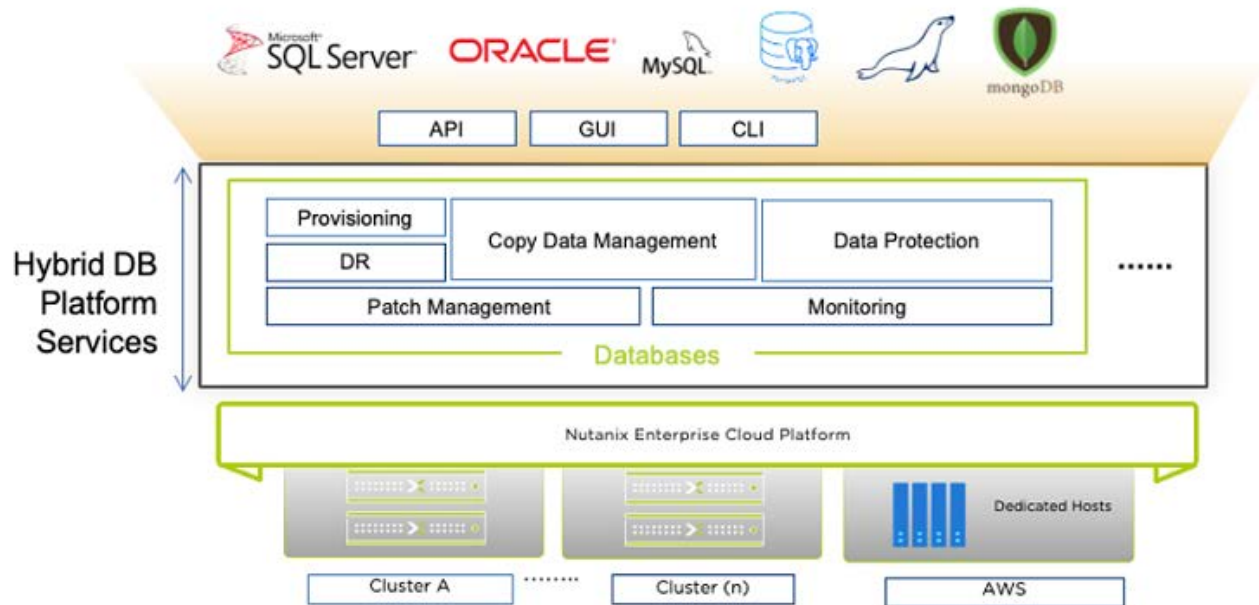
Figure 57: Nutanix Database Service (NDB)

10.3 Licensing Model

This section provides different licensing options and combinations in the Nutanix Cloud Platform.

Table 48: Nutanix Cloud Platfrom components licensing options

Software	Options	Add On
Nutanix Cloud Infrastructure	Core Based	• Nutanix Unified Stroage • Nutanix Database Service
Nutanix Cloud Manager	Core Based	• NCM Cloud • Cost Governance • Security
Nutanix Unified Stroage(NUS)	**Mixed Mode:** Unified Stroage + NCI / NCP **Dedicated Mode:** Unified Storage only (TB Based)	• Advanced Replication • Security • Data Lens •
Nutanix Cloud Infrastructure - VDI	Core Based / Maximum Concurrent Users/Virtual machines (Includes Unified Storage)	• Unified Storegae (TB based) • Advanced Replication • Security

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Nutanix Cloud Infrastructure – Edge	Core Based (Maximum 25 virtual machines, 96GB memory per VM) (Includes Unified Storage)	• Unified Storegae (TB based) • Advanced Replication • Security
Nutanix Database Service (NDB)	Core based	N/A
Nutanix Mine Integrated Backup	TB based (Nutanix Unified Storage)	N/A
Nutanix Data Lens	TB based (Nutanix Unified Storage)	N/A

10.4 Performance recommendations

This section provides sizing performance recommendations and best practices.

System Requirements for Nutanix Cloud Manager Self Service Service on Prism Control

Deployment Size	VMs	vCPU	Memory	Disk
Small	2500	6	30 GB	1 TB
Large	5000	10	52 GB	1 TB

System Requirements for Nutanix Cloud Manager Intelligent Operations service in Prism Central

Deployment Size	VMs	vCPU	Memory	Disk
Small	2500	4	16 GB	200 GB
Large	5000	8	32 GB	500 GB

System Requirements for additional services

Component	vCPU	RAM	Disk	Network
LDAP	2	4 GB	40 GB	1 Gbps
DNS	1	2 GB	40 GB	1 Gbps
Admin VM	2	4 GB	100 GB	1 Gbps
SMTP	1	2 GB	40 GB	1 Gbps
NTP	1	2 GB	40 GB	1 Gbps
Syslog	2	4 GB	100 GB	1 Gbps

10.5   Deployment example

10.5.1 Private Cloud with AHV

This example uses Lenovo ThinkAgile HX V3 nodes with Nutanix Cloud Platform to provide full-stack solution for private cloud deployments that integrates multiple products including AOS, AHV, Prism, Nutanix Cloud Manager (NCM) Self-Service , Flow Network Security, Nutanix Disaster Recovery and Mine backup. The following Nutanix and partner software required to deploy private cloud on Lenovo ThinkAgile HX.

- Nutanix AOS
- Nutanix AHV
- Nutanix Objects
- Object manager
- NCM Self-Service
- Nutanix Life Cycle Manager
- Nutanix Microservices Platform
- Nutanix Flow Network Security
- Nutanix Cluster Check
- Nutanix Foundation
- Nutanix Mine Integrated Backup
- Veeam

The private cloud comprises of one management cluster and multiple workload clusters. The figure below shows management cluster with Nutanix Cloud platform services and one workload cluster for Nutanix Database service and another dedicated workload cluster for VDI with Citrix. The clusters are built with Intel Xeon 4th Generation Sclable Processors and AMD EPYC 4th Generation processors.

Management Cluster Node Configuration

- 4 x Lenovo ThinkAgile HX650 V3 Integrated System
- 2 x Intel Xeon Gold 5416S 16C 150W 2.0GHz Processor
- 12 x ThinkSystem 32GB TruDDR5 4800MHz (1Rx4) 9x4 RDIMM-A
- 4 x ThinkSystem 2.5" S4620 3.84TB Mixed Use SATA 6Gb HS SSD
- 1 x ThinkSystem Broadcom 57414 10/25GbE SFP28 2-Port OCP Ethernet Adapter

Workload Cluster 1(Database as a Service) Node configuration

- 4 x Lenovo ThinkAgile HX645 V3 Integrated System

- 2 x ThinkSystem AMD EPYC 9334 32C 210W 2.7GHz Processor

- 16 x ThinkSystem 32GB TruDDR5 4800MHz (1Rx4) 9x4 RDIMM-A

- 6 x ThinkSystem 2.5" S4620 3.84TB Mixed Use SATA 6Gb HS SSD

- 1 x ThinkSystem Broadcom 57454 10/25GbE SFP28 4-Port OCP Ethernet Adapter

Workload Cluster 1 (VDI as a Service) Node Configuration

- 4 x Lenovo ThinkAgile HX650 V3 Certified Nodes

- 2 x Intel Xeon Gold 6448Y 32C 225W 2.1GHz Processor

- 8 x ThinkSystem 64GB TruDDR5 4800MHz (2Rx4) 10x4 RDIMM

- 4 x ThinkSystem 2.5" PM1653 7.68TB Read Intensive SAS 24Gb HS SSD

- 1 x ThinkSystem Broadcom 57414 10/25GbE SFP28 2-Port OCP Ethernet Adapter
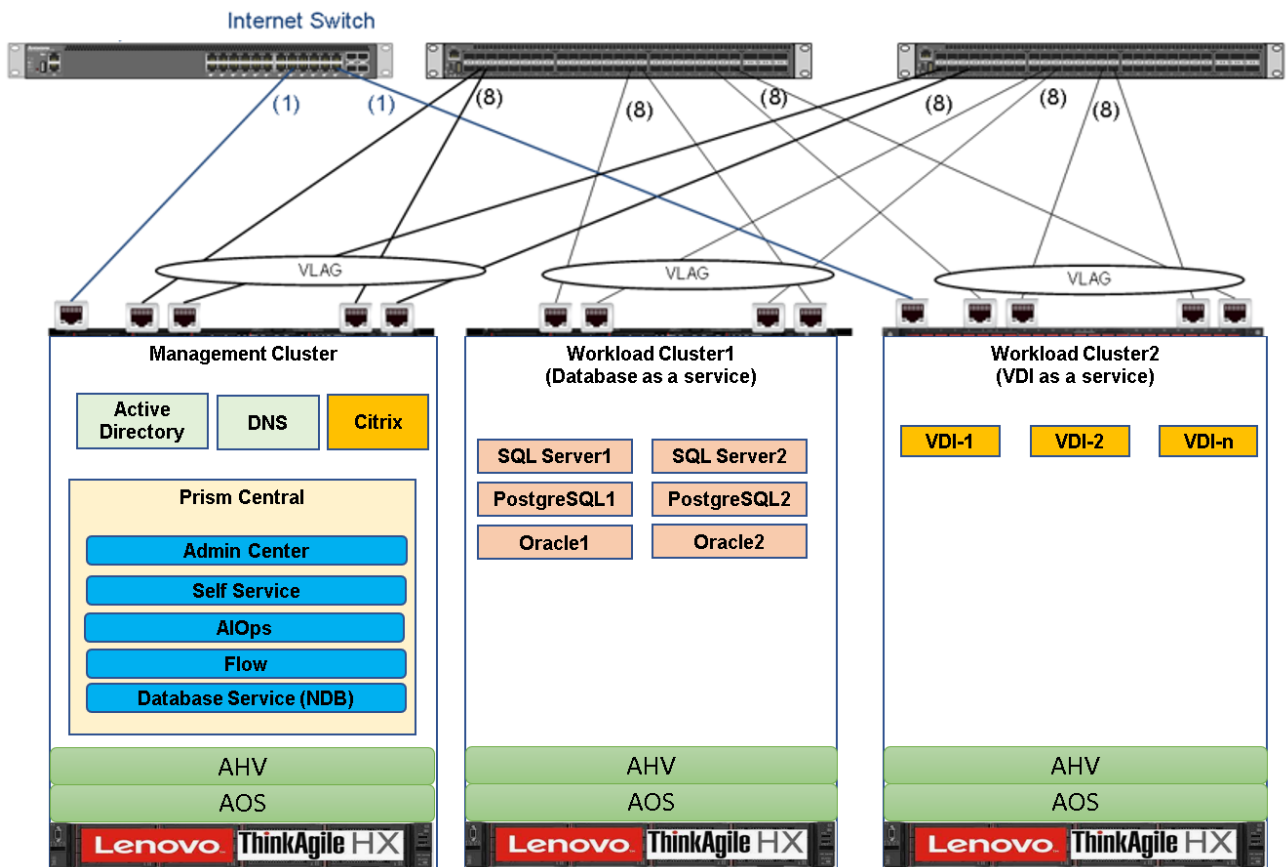
Figure 58: Private Cloud with Lenovo ThinkAgile HX and Nutanix Cloud Platform

The private cloud solution can use Nutanix Mine Integrated Backup with Veeam and backup require separate storage cluster which can be built using Lenovo ThinkAgile HX Storage nodes.The backup can also be done on shared storage with Lenovo ThinkSystem DE/DM/DG storage with Veeam. For more information see Lenovo ThinkSystem DE/DM/DG Series Storage Arrays

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

11 Hybrid backup with AWS

Amazon Web Services (AWS) provides a backup and restore solution that offers on-premise data centers a seamless hybrid cloud backup solution. AWS Backup integrates with the on-premise AWS Storage Gateway to act as a bridge between local storage and Amazon's various cloud storage solutions to provide hybrid storage and backup capabilities. Virtual machines running on the HX platform or physical servers can connect to local volumes on the AWS Storage Gateway via iSCSI. This solution can be used for nearly any application or systems, with the use of industry standard connectivity.

11.1 Solution overview

The figure below shows the AWS hybrid backup and restore architecture. Although file based storage is also supported, this design is focused on the block storage Volume Gateway on the AWS Storage Gateway. Volumes are created on the gateway and connected to local on-premise hosts, while the Elastic Block Storage (EBS) service in the AWS cloud takes snapshots on a defined schedule. AWS Backup can also be used for more centrally managed and automated data protection strategies.

The key advantages of using the AWS Storage Gateway include:

- Integration of hardware and software without any hardware changes

- Use on-premises storage and cloud storage in Amazon - hybrid storage

- Simple migration from physical infrastructure to the AWS cloud

Note that the AWS Storage Gateway is a custom Linux VM image provided by AWS and it is currently not configured for the Nutanix hypervisor; as a result we deployed the storage gateway VM on a Hyper-V host.
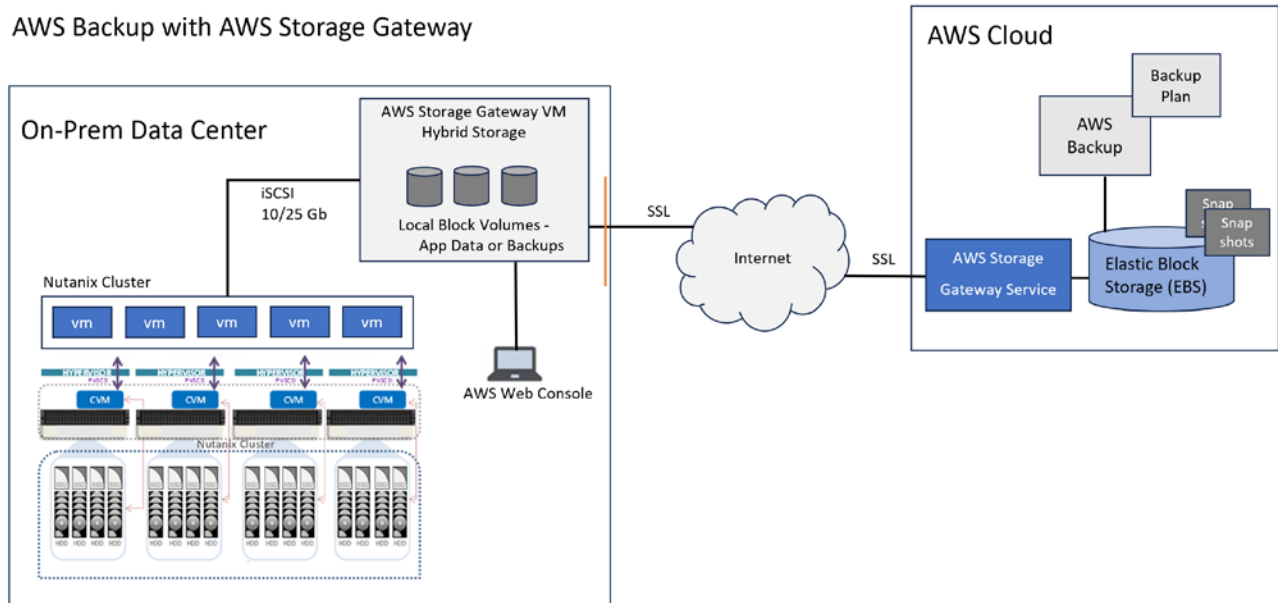

Figure 59 AWS Backup with AWS Storage Gateway Architecture

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

11.2 Component model

The figure below is a layered view of the components making up the AWS hybrid cloud storage and backup solution for the HX systems with Lenovo and Nutanix on the left and AWS components on the right. For more detailed reading, please see https://aws.amazon.com/backup/
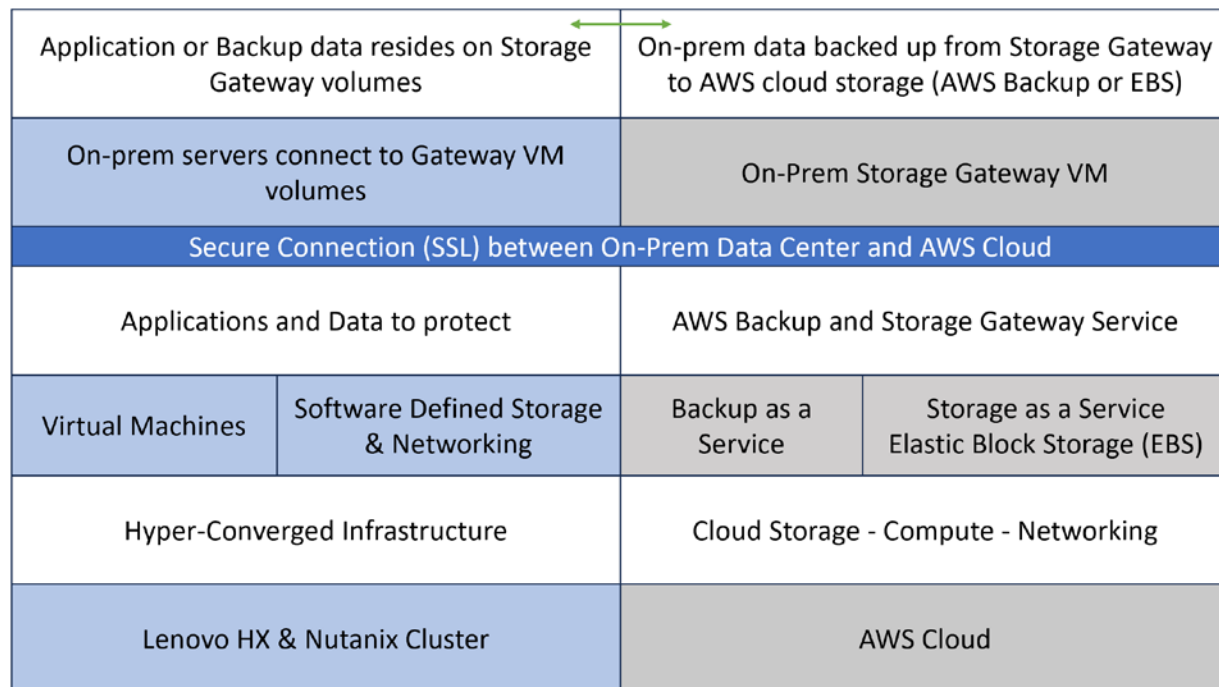

Figure 60 Component model AWS hybrid backup

11.3 AWS Storage Gateway

A key to understanding this hybrid storage and backup solution is to take a closer look at the AWS Storage Gateway. It is the bridge that sits between the on-premise resources and the AWS cloud. Amazon is using iSCSI to connect the on-premise servers to data volumes on the Storage Gateway. Then the AWS Storage Gateway service in the cloud is able to connect to thes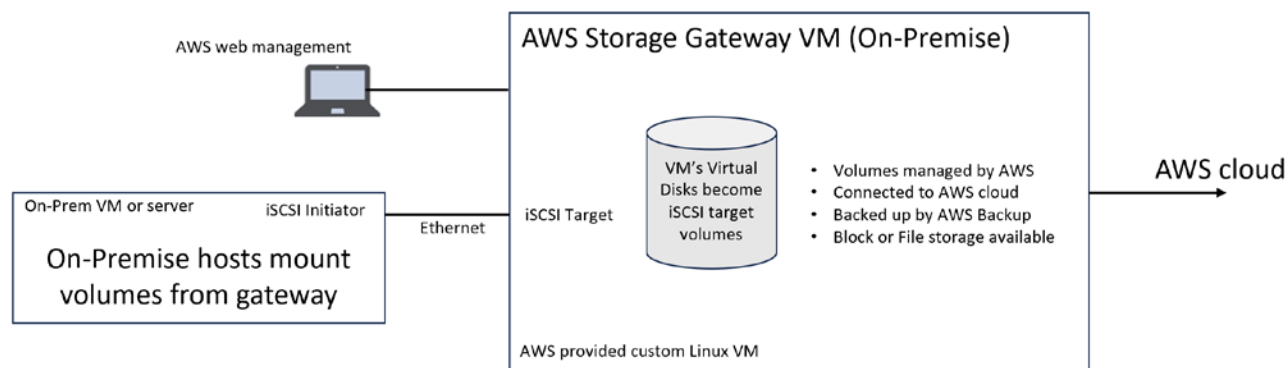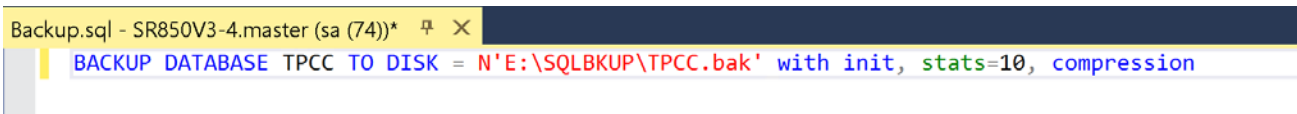e volumes on the gateway, and back them up. https://docs.aws.amazon.com/storagegateway/latest/vgw/WhatIsStorageGateway.html

Figure 61 AWS Storage Gateway detailed view

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

11.4 Application backups – SQL Server

Amazon is not able to reach inside your data center and do native application backups, at least not yet. AWS Backup doesn't even support SQL backups on SQL Servers hosted inside the AWS cloud. So for many applications, an alternative approach is needed to make use of this hybrid infrastructure. Some will only use the AWS Storage Gateway volumes as a backup destination, while others will use it for application data.

Although there are backup solutions that are SQL aware, and can natively backup and restore SQL databases, many organizations still rely on SQL T-SQL backup scripts, launch by some chron method. This is the method we'll use, with the SQL backup destination being an iSCSI attached volume on the Storage Gateway. The figure below shows a typical SQL backup script. There are also similar restore scripts.

Figure 62 Sample SQL backup script

After the SQL .bak file is on the gateway volume, a snap shot can be scheduled from AWS to pick up the new data. This can be done with either AWS Backup or the EBS snap shot scheduler. Both can be configured easily from the AWS web console.

11.5 Deployment details

This section covers the high level steps and links to detailed information necessary to deploy the solution, and connect your on-premise hosts to it. The steps in the links below have been validated in our lab proof of concept testing.

After creating an AWS account, the next steps are to determine which region to align with and which type of AWS Storage Gateway to deploy. The region is in the upper right of the AWS console, you'll want to select the one closest to your data center for best performance.

Regarding the types of gateways, there are three to choose from:

- File Gateway – NAS and NFS
- Volume Gateway – iSCSI and either Stored or Cached volume type
- Tape Gateway – iSCSI, archival level

For this paper, we are only covering the Volume Gateway, and using Stored Volumes option.

Follow the steps and links below to deploy the gateway.

1. Setup a Volume Gateway with the steps at this link:
   https://docs.aws.amazon.com/storagegateway/latest/vgw/create-volume-gateway.html
   Note that you need to create some fixed size virtual disks and make them available on your VM. The
   steps in the link don't mention this. These are the volumes that AWS will configure as iSCSI data
   volumes later in the process.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

2. Create the volumes by following these steps:

https://docs.aws.amazon.com/storagegateway/latest/vgw/GettingStartedCreateVolumes.html

3. Connect to a volume by following these steps:

https://docs.aws.amazon.com/storagegateway/latest/vgw/GettingStarted-use-volumes.html

4. Backup the volumes with AWS Backup here:

https://docs.aws.amazon.com/storagegateway/latest/vgw/backing-up-volumes.html

5. Below is a view of the Storage Gateway management console.
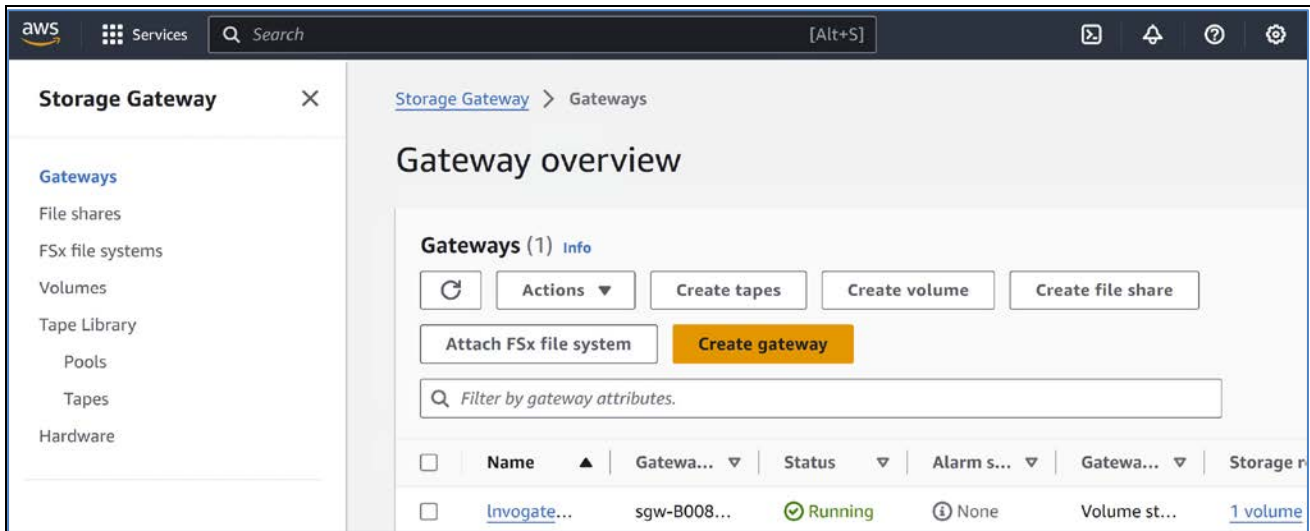
Figure 63 View of Storage Gateway management console

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

12 Enterprise AI with NVIDIA

The value of artificial intelligence technologies and democratization of AI in the enterprise business brings development of relevant applications or software programs to address tasks or use cases across industries and sectors. Lenovo Enterprise AI solutions provide different hardware platforms and software stack for data preparation and analysis, model development, training, inference, application development and integrations with existing systems and workflows. Refer Enterprise AI Reference Architecture for more details. The Nutanix Enterprise AI solution comprises the following software validated to run on Lenovo ThinkAgile HX infrastructure.

- Nutanix AHV Hypervisor
- Nutanix Cloud Platform
- RedHat OpenShift Container Platfom
- NVIDIA AI Enterprise Suite

12.1 Solution Overview

Lenovo ThinkAgile HX systems with Nutanix Cloud Platform and NVIDIA Enterprise AI provides unified solution for managing lifecycle of data science and AI/ML applications span across edge, core and hybrid cloud infrastructure. Nutanix AHV is certified hypervisor for NVIDIA AI Enterprise and Lenovo ThinkAgile platform supports NVIDIA GPUs to run training and inference workloads for variety for AI/ML usecases. Nutanix Prism Central cluster management solution provides a single management plane for virtual environments, including VMs, storage, and networking. Redhat OpenShift running on AHV provides solid foundation to containerize and deploy AI/ML workloads at scale. Lenovo ThinkAgile HX with Nutanix Cloud Platfor addresses data security, latency, deployment speed and scalability challenges in the machine learning life cycle.
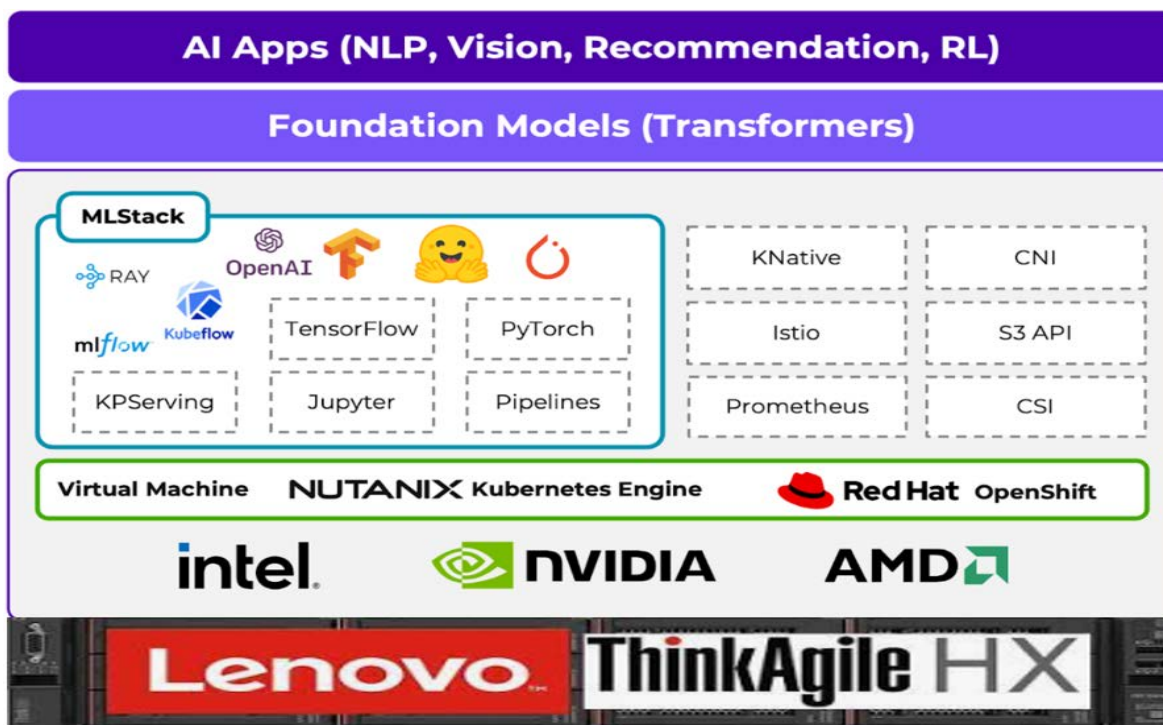

Figure 64 ThinkAgile HX with NVIDIA AI Enterprise Suite

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

12.2 Component model

12.2.1 NVIDIA AI Enterprise

NVIDIA AI Enterprise is an end-to-end, cloud native software platform that accelerates the data science pipeline and streamlines development and deployment of production-grade AI applications, including generative AI, computer vision, speech AI, and more. Enterprises that run their businesses on AI rely on the security, support, and stability provided by NVIDIA AI Enterprise to improve productivity of AI teams, reduce total cost of AI infrastructure, and ensure a smooth transition from pilot to production.
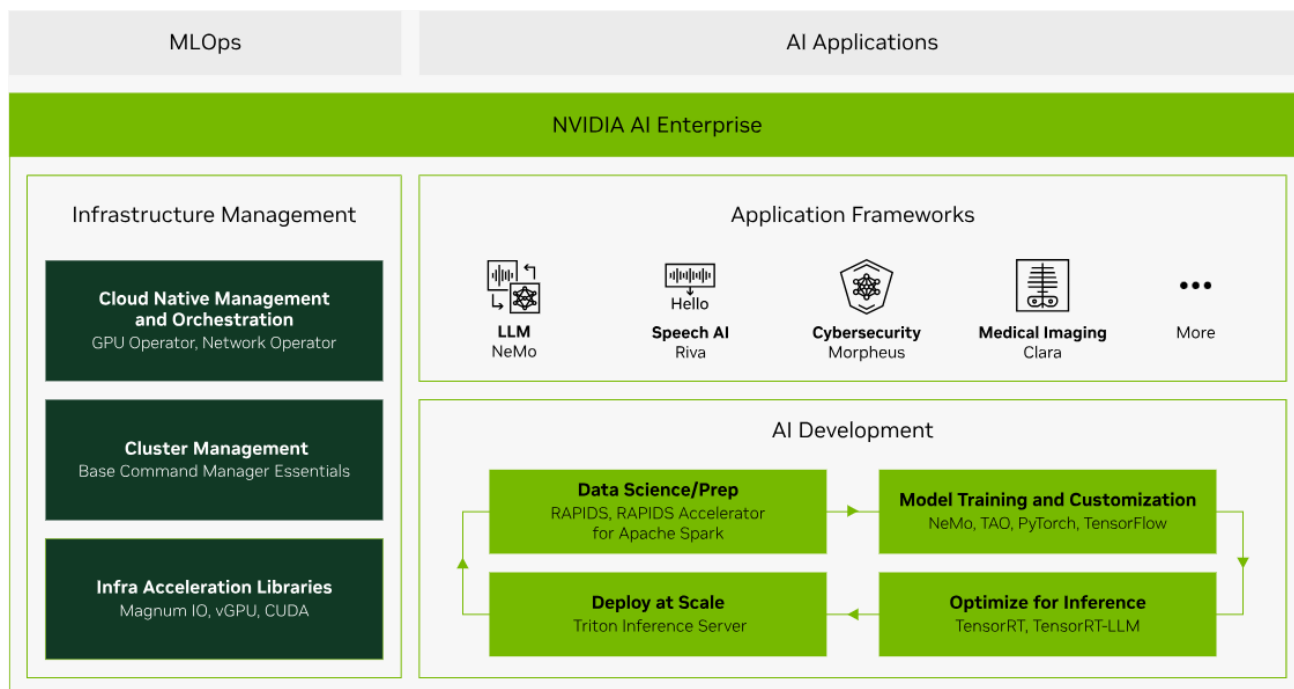

Figure 65 NVIDIA AI Enterprise

NVIDIA AI Enterprise includes the following components.

NVIDIA NeMo an end-to-end framework for building, customizing, and deploying enterprise-grade generative AI models; NeMo lets organizations easily customize pretrained foundation models from NVIDIA and select community models for domain-specific use cases.

NVIDIA RAPIDS is an open-source suite of GPU-accelerated data science and AI libraries with APIs that match the most popular open-source data tools. It accelerates performance by orders of magnitude at scale across data pipelines.

NVIDIA TAO Toolkit simplifies model creation, training, and optimization with TensorFlow and PyTorch and it enables creating custom, production-ready AI models by fine-tuning NVIDIA pretrained models and large training datasets.

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

NVIDIA TensorRT, an SDK for high-performance deep learning inference, includes a deep learning inference optimizer and runtime that delivers low latency and high throughput for inference applications. TensorRT is built on the NVIDIA CUDA parallel programming model, enables you to optimize inference using techniques such as quantization, layer and tensor fusion, kernel tuning, and others on NVIDIA GPUs.

https://developer.nvidia.com/tensorrt-getting-started

NVIDIA TensorRT-LLM is an open-source library that accelerates and optimizes inference performance of the latest large language models (LLMs). TensorRT-LLM wraps TensorRT's deep learning compiler and includes optimized kernels from FasterTransformer, pre- and post-processing, and multi-GPU and multi-node communication.

https://developer.nvidia.com/tensorrt

NVIDIA Triton Inference Server optimizes the deployment of AI models at scale and in production for both neural networks and tree-based models on GPUs.

NVIDIA AI Enterprise is certified on private cloud platforms VMware Cloud Foundation, Nutanix AHV, Red Hat Enterprise Linux and Ubuntu KVM and works seamlessly with Container orchestration VMware Tanzu, Red Hat OpenShift, Google Kubernetes Engine (GKE), Amazon Elastic Kubernetes, Service (EKS), Azure Kubernetes, Service (AKS), and upstream Kubernetes. It can be integrated with MLOps platforms such as ClearML, Domino Data Lab, Run:ai, UbiOps, and Weights & Biases.

12.2.2 Nutanix Cloud Platform

Nutanix Cloud Platform (NCP) is a unified solution built on top of hyperconverged infrastructure to address hybrid cloud requirements and simplify operations for different workload scenarios. Lenovo ThinkAgile HX systems support NCP to provide reliable infrastructure as a foundation for a variety of use cases in hybrid cloud deployments. Refer section 10 Nutanix Cloud Platform for more details.

12.2.3 Redhat OpenShift Container Platform

Red Hat OpenShift Container Platform is Kubernetes solution provides many enterprise-ready features like enhanced security, multi-tenancy, virtualization, data science platform, simplified application deployment, and continuous integration/continuous deployment tools. The AI/ML workloads can be developed and deployed at scale leveraging capabilities of ThinkAgile HX, Nutanix AHV, Nutanix Cloud Platform and NVIDIA AI Enterprise Suite. Refer section 6 Red Hat OpenShift Container Platform for more details.

12.3 Deployment Example

ThinkAgile HX systems with NVIDIA AI Enterprise solution supports requirements for enterprise and small-medium business (SMB). The vertical axis is split into two categories of AI/ML workload compute requirements and many model inference can be performed on CPU and with low-end GPUs and accelerators. Machine learning training requires GPUs and inference for large model are addressed with GPU. CPU is required for

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

both training and inference and the lower half can fit for SMB requirements and upper half for Enterprise requirements.
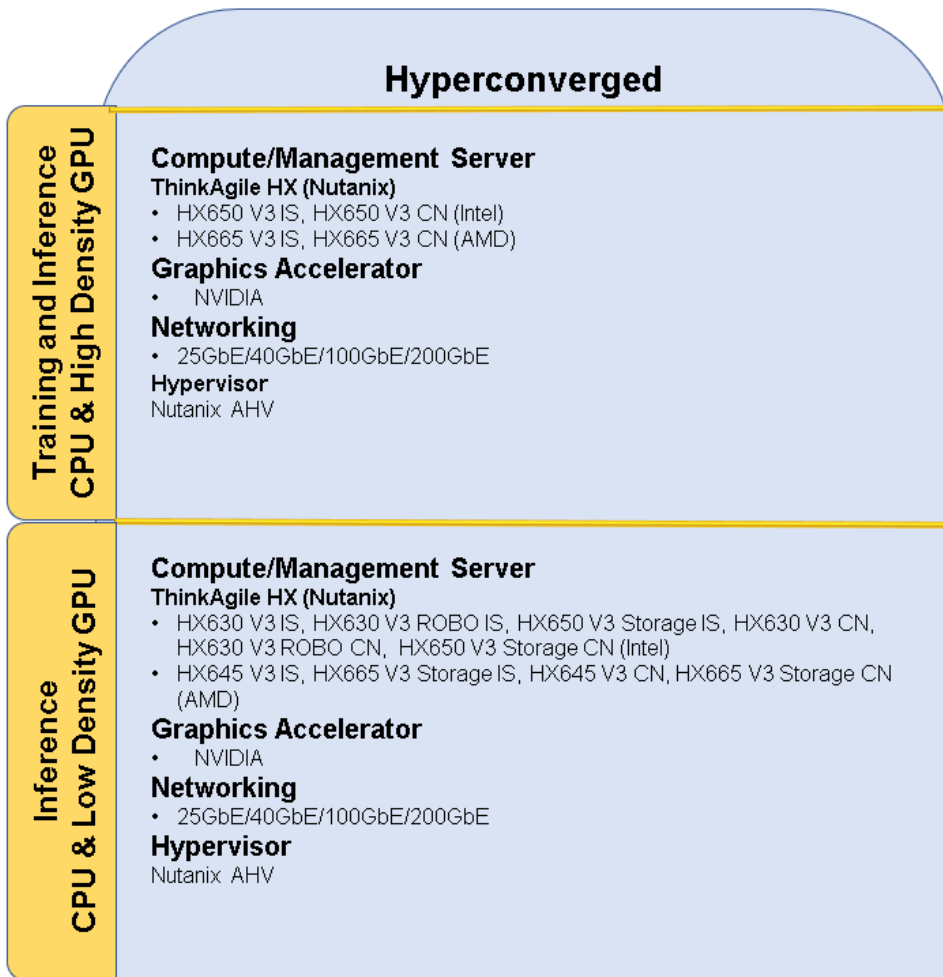

Figure 66 ThinkAgile HX Enterprise AI Operational Models

Table below shows an approximate estimate to size required infrastructure for different business wide need for AI deployment.

Table 49: AI Models for different business sizes

Type of Business Size	Total models	Dataset Size (Total)	Number of Machine Learning models	Number of Deep Learning Models	Number of Generative AI Applications
Small	Up to 10	<100 TB	8	2	2
Medium	10-30	100-500TB	25	6	8
Enterprise	30+	1PB +	30+	10+	5+

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

12.3.1 Example Enterprise AI Infrastructure Solution

This section provides compute sizing for software and applications for a Enterprise AI requirements with 40 models of different sizes. This example is considered as base to provide sizing for different Lenovo ThinkSystem and ThinkAgile platform solutions described in the following sections.

Table 50 provides software components and resource requirements for data science and preparation phase in machine learning life cycle. Data preparation is critical task and the data sources can be any and this example considers streaming to process data coming from different data sources and process it to store features in different databases for different models.

Table 50: Data science and Machine Learning Framework System Configuration

Management service	Virtual processors	System memory	Storage	OS	GPU	HA needed	Workload Characteristics
Kafka Broker Cluster (2)	8 VCPU	32 GB	300 GB	Linux	N/A	Yes	Inventory events 1 million per day
Kafka Connector Cluster (2)	6 CPU	32 GB	200 GB	Linux	N/A	Yes	Inventory and product database
PostgreSQL	16 VCPU	128 GB	2000 GB	Linux	N/A	Yes	100k new product records per day
MongoDB	24 VCPU	128 GB	3000 GB	Linux	N/A	Yes	Training Data
Hopsworks Feature Store	16 VCPU	96 GB	300 GB	Linux	N/A	Yes	Feature store for model training and serving
KubeFlow	8 VCPU	24 GB	300 GB	Linux	N/A	Yes	MLOps

Compute for Machine Learning Training Cluster Workload

The number of AI/ML models in an enterprise varies for different use cases and each model also requires different CPU and GPU requirements based on the model algorithm and dataset complexity. Table 51 provides t-shirt sizing for ML training with NVIDIA multi-instance GPU (MIG) profile which partitions single H100 to 7 instances. The table below shows t-shirt sizing for training environment for 40 models of different sizes. The same cluster can be used for HPC workloads which is not sized in this reference architecture. The ML training can be done in distributed manner by running models across multiple nodes.

Table 51: Machine Learning Training Applications System Configuration

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Model Type (count)	Virtual processors	System memory	Storage	GPU Profile	Workload Characteristics
Small (16)	8 VCPU	32 GB	100 GB	MIG 1g.10gb	Machine Learning
Medium (12)	16 VCPU	GB	200 GB	MIG 2g.20gb	Machine Learning
Large (8)	32 VCPU	128 GB	500 GB	MIG 3g.40gb	Machine Learning
Extra Large (3)	64 VCPU	256 GB	1000 GB	MIG 7g.80gb	Deep Learning
High End (1)	96 VCPU	1000 GB	2000 GB	8 x GPU (Direct)	Deep Learning

Compute for AI Inference Cluster Workload

AI Inference workloads can leverage both CPU and GPU. The number of inference instances can be scaled out to optimally use GPU memory. Inference can be performed with CPU only for many use cases. The inference scaling is based on the number of requests served and it must be scaled appropriately.

Table 52: Machine Learning Inference System Configuration

Model Type (count)	Virtual processors	System memory	Storage	GPU Profile	Workload Characteristics
Small (16)	12 VCPU	32 GB	100 GB	MIG 1g.10gb	Machine Learning
Medium (12)	32 VCPU	64 GB	200 GB	MIG 2g.20gb	Machine Learning
Large (8)	8 VCPU	128 GB	500 GB	MIG 3g.40gb	Machine Learning
Extra Large (3)	64 VCPU	512 GB	1000 GB	MIG 7g.80gb	Deep Learning
High End (1)	96 VCPU	1024 GB	2000 GB	4 x GPU (Direct)	Deep Learning

12.3.2 Compute Cluster with Intel 4th Gen Intel Xeon Scalable processors

The compute requirement considers end-end data science and AI/ML pipeline for model development, training, and production deployment for the sample sizing described in the previous section. The compute requirement considers end-end data science and AI/ML pipeline for model development, training, and production deployment. Data preparation is a critical task in any AI project, and it is recommended to have these applications and services isolated on separate cluster. The sample sizing given below uses Lenovo ThinkSystem HX650 V3 IS server Intel 4th Gen Xeon Scalable Processors and NVIDIA H100 GPU for data science, AI/ML training AI training and inference use cases. ThinkSystem HX650 V3 IS systems support three NVIDIA H100 GPU per node and servers can be scaled out to meet workloads.

Table 53: Compute Clusters Configuration for AI Applications

Cluster Name	Lenovo Platform	CPU	Memory	OS / Hypervisor	GPU
Data Science and AI Apps	4 x ThinkSystem HX650 V3 IS	Intel Xeon Gold 6448H 32C 250W 2.4GHz	2000 GB	RHEL	No
AI Training Nodes	8 x ThinkSystem HX650 V3 IS	Intel Xeon Platinum 8468 48C 350W 2.1GHz	1000 GB	RHEL	3xNVIDIA H100 80GB
AI Inference	4 x ThinkSystem HX650 V3 IS	Intel Xeon Platinum 8468 48C 350W 2.1GHz	1000 GB	RHEL	2xNVIDIA H100 80GB

12.3.3 Compute Cluster with Intel 4th Gen AMD EPYC processors

The sample sizing given below uses Lenovo ThinkSystem HX665 V3 IS server 4th Gen AMD EPYC Processors and NVIDIA H100 GPU for data science, AI/ML training AI training and inference use cases. ThinkSystem HX665 V3 IS systems support three NVIDIA H100 GPU per node and servers can be scaled out to meet workloads.

Table 54: Compute Clusters Configuration for AI Applications

Cluster Name	Lenovo Platform	CPU	Memory	OS / Hypervisor	GPU
Data Science and AI Apps	4 x ThinkSystem HX645 V3 IS	ThinkSystem AMD EPYC 9354 32C 280W 3.25GHz	2000 GB	RHEL	No
AI Training Nodes	8 x ThinkSystem HX665 V3 IS	ThinkSystem AMD EPYC 9454 48C 290W 2.75GHz	1000 GB	RHEL	3xNVIDIA H100 80GB
AI Inference	4 x ThinkSystem HX665 V3 IS	ThinkSystem AMD EPYC 9474F 48C 360W 3.6GHz	1000 GB	RHEL	2xNVIDIA H100 80GB

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Resources

- Nutanix Portal (requires registration)

  portal.nutanix.com

- Nutanix Bible

  nutanixbible.com/

- Nutanix Tech Note: VMware vSphere Networking on Nutanix

  go.nutanix.com/rs/nutanix/images/Nutanix_TechNote-
  VMware_vSphere_Networking_with_Nutanix.pdf

- VMware vSphere

  vmware.com/products/datacenter-virtualization/vsphere

- SAP Support Portal (requires registration)

  support.sap.com

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Document History

Version 1.0	31 October 2017	• Initial version
Version 1.1	15 November 2017	• Added chapter for Microsoft SQL Server workload • Added rack deployment models with more detail on ThinkAgile SXN
Version 1.2	10 December 2017	• Added chapter for Microsoft Exchange 2013 workload • Added deployment model for Microsoft Exchange
Version 1.3	22 January 2018	• Added Nutanix AHV hypervisor and Citrix Workspace Appliance (CWA) for Citrix Virtual Apps and Desktops • Added Microsoft Exchange 2016
Version 1.4	29 January 2018	• Added chapter for VMware vCloud Suite • Added deployment model for VMware vCloud Suite • Added performance results for VDI graphics acceleration • Add Microsoft Exchange 2016 results for all flash appliances
Version 1.5	18 August 2018	• Added certified nodes • Added pre-configured deployment ready solutions for Citrix Virtual Apps and Desktopsand VMware Horizon • Removed ThinkAgile SXN integrated solution
Version 1.6	2 October 2018	• Added HX782x appliances and certified nodes • Added chapter for SAP Business Applications
Version 1.7	12 December 2018	• Added chapter for Red Hat OpenShift Container Platform
Version 1.8	23 April 2019	• Updated chapter on Microsoft SQL Server to use SQL Server 2019 and Intel Xeon Scalable Processor gen 2 CPUs.
Version 1.9	10 June 2019	• Added Citrix Virtual Apps and Desktopsresults for Xeon SP2 processors • Added VMware Horizon results for Xeon SP2 processors
Version 2.0	16 July 2019	• Updated performance data for VDI Power Worker profile from the Login VSI benchmark
Version 2.1	25 September 2019	• Updated chapter SAP HANA to include HX332x and HX752x models based on Intel Xeon SP 1.Gen processors • Added HX7820 SSD Only descriptions to SAP HANA section • Removed requirement for Mellanox ConnectX Adapters with SAP HANA

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Version 2.2	15 November 2019	• Added Citrix Workspace appliance
		• Added SAP HANA models for Intel Xeon SP Gen 2. Processors
Version 2.3	10 December 2020	• Moved the SAP Solutions chapter to a separate Reference Architecture document - lenovopress.com/lp1413
Version 2.4	27 July 2022	• Section 2.1 References for ThinkAgile HX appliances and certified nodes with Intel Xeon Scalable Gen 2 and Gen 3 are added
		• Section 3.2.1 VDI Deployment model is updated with Intel Xeon Scalable Gen 2 processors
		• Section 3.3 Added TruScale deployment models
		• Section 4 Updated Citrix XenDesktop to Citrix Virtual Apps and Desktops
		• Section 4.6.2, 4.7.1 Updated LoginVSI benchmark results on Intel Icelake and AMD EPYC 7003 processors
		• Section 4.8 Updated deployment ready solutions with HX3521-G certified node and HX2320 (for TruScale)
		• Section 8.6.2, 8.7.1 Updated LoginVSI benchmark results on Intel Icelake and AMD EPYC 7003 processors
		• Section 8.8 Updated deployment ready solutions with HX3521-G certified node
Version 2.5	23 August 2022	• Section 2.2.4, 2.3, 2.4 – Lenovo specific switch configurations and references are removed
		• Section 5 – Exchange workload is removed
		• Section 8.6.2 – Lenovo network plugins are removed
		• Section 3.2.2 – Exchange server is removed from the deployment model
		• Chapter 5 – Updated to remove SQL Server version info
Version 2.6	20 December 2022	• Section 4.6.1 – Login Enterprise workload is added
		• Added Section 4.6.3 for Login Enterprise results for Citrix on ESXi with Intel 3rd Generation Intel Xeon Processors
		• Added Section 4.6.4 for Login Enterprise results for Citrix on AHV with Intel 3rd Generation Intel Xeon Processors

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**

Version 2.7	26 September 2023	• Section 9 – Added Backup solution for ThinkAgile HX with Lenovo ThinkSystem DE/DM/DG storage and Veeam
Version 2.8	02 January 2024	• Section 2.1 – Updated with ThinkAgile HX V3 systems • Section 10 – Added Nutanix Cloud Platform for private and hybrid cloud solutions with ThinkAgile HX systems
Version 2.9	26 March 2024	• Section 2.1 – Removed withdrawn processors • Section 3.2.1, 3.2.2, 3.3.4 – Updated with 4th Gen Intel Xeon Scalable Processors and 4th Gen AMD APYC Processors • Section 4.6.1, 4.6.2, 4.6.5 - Removed LoginVSI Pro benchmark results and added Login Enterprise results for 4th Gen Intel Scalable Processors • Section 4.8 & 7.8 – Updated Solution Identifier for HX V3 systems • Section 11 – Added Hybrid backup with AWS solution • Section 12 – Added Enterprise AI with NVIDIA solution

Reference Architecture for Workloads using Lenovo ThinkAgile HX Series **version 2.9**