

# REFERENCE ARCHITECTURE

## Dell EMC Ready System for VDI on XC Series

Integration of Citrix XenDesktop with Dell EMC XC Series Hyper-Converged Appliances

### [Abstract](#)

A Reference Architecture for integrating Dell EMC XC Series Hyper-Converged Appliances and Citrix XenDesktop software to create virtual application and virtual desktop environments on 14th generation Dell EMC PowerEdge Servers.

May 2018

## Revisions

Date	Description
November 2017	Initial release
January 2018	vGPU updates
May 2018	XC Core, Citrix Ready Workspace Appliance, and XenServer updates

## Acknowledgements

This paper was produced by the following members of the Dell EMC Ready Solutions team:

Authors: Peter Fine – Chief Architect

Nicholas Busick – Solution Architect

Geoff Dillon – Sr. Solutions Engineer

Andrew Breedy – Sr. Solutions Engineer

Colin Byrne – Sr. Solutions Engineer

Jonathan Chamberlain – Solutions Engineer

Support: David Hulama – Sr. Technical Marketing Advisor

Other:

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

© 2017-2018 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.

Dell believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

# Table of contents

Revisions.....	2
Acknowledgements.....	2
Executive summary.....	6
<b>1 Introduction.....</b>	<b>7</b>
1.1 Objective.....	7
1.2 What's new .....	7
<b>2 Solution architecture overview .....</b>	<b>8</b>
2.1 Introduction .....	8
2.2 Dell EMC XC Series Hyper-Converged Appliances.....	8
2.3 Dell EMC XC Core.....	8
2.4 Distributed Storage Fabric.....	9
2.5 App Mobility Fabric .....	9
2.5.1 Nutanix architecture.....	9
2.6 Nutanix Hyper-Converged Infrastructure.....	12
2.7 Nutanix all-flash .....	13
2.8 Dell EMC XC Series - VDI solution architecture.....	13
2.8.1 Networking.....	13
2.8.2 XC Series – Enterprise solution pods.....	15
<b>3 Hardware components .....</b>	<b>18</b>
3.1 Network.....	18
3.1.1 Dell Networking S3048 (1Gb ToR switch).....	18
3.1.2 Dell Networking S4048 (10Gb ToR switch).....	19
3.2 Dell EMC XC Series hardware changes between generations.....	20
3.2.1 Dell EMC XC640.....	21
3.2.2 Dell EMC XC740xd.....	24
3.3 NVIDIA Tesla GPUs .....	26
3.3.1 NVIDIA Tesla M10 .....	26
3.3.2 NVIDIA Tesla M60 .....	27
3.4 Dell Wyse Endpoints .....	28
3.4.1 Wyse 3040 Thin Client (ThinOS, ThinLinux) .....	28
3.4.2 Wyse 5040 AIO Thin Client (ThinOS).....	28
3.4.3 Wyse 5060 Thin Client (ThinOS, ThinLinux, WES7P, WIE10) .....	28
3.4.4 Wyse 7020 Thin Client (WES 7/7P/8, WIE10, ThinLinux).....	29
3.4.5 Wyse 7040 Thin Client (WES7P, WIE10) .....	29
<b>4 Software components.....</b>	<b>30</b>

4.1	Citrix.....	30
4.1.1	XenDesktop .....	30
4.1.2	Machine Creation Services (MCS) .....	32
4.1.3	Personal vDisk.....	33
4.1.4	AppDisks.....	33
4.1.5	HDX 3D Pro .....	33
4.1.6	Profile Manager .....	34
4.1.7	XenApp .....	34
4.1.8	NUMA architecture considerations .....	36
4.1.9	Local Host Cache .....	39
4.1.10	NetScaler.....	40
4.1.11	Hypervisor platforms .....	42
4.2	NVIDIA GRID vGPU .....	44
4.2.1	vGPU profiles.....	44
4.3	Citrix Ready Workspace Appliance .....	51
5	<b>Solution architecture for XenDesktop.....</b>	<b>52</b>
5.1	Management role configuration .....	52
5.1.1	Citrix XenDesktop on vSphere management role requirements .....	52
5.1.2	Citrix XenDesktop on Hyper-V management role requirements .....	53
5.1.3	RDSH/XenApp on vSphere .....	53
5.1.4	RDSH/XenApp on Hyper-V .....	54
5.1.5	NVIDIA GRID license server requirements .....	54
5.1.6	SQL databases.....	55
5.1.7	DNS .....	55
5.2	Storage architecture overview .....	56
5.2.1	Nutanix containers .....	56
5.3	Virtual networking .....	56
5.3.1	vSphere .....	57
5.3.2	Hyper-V.....	57
5.4	Scaling guidance .....	59
5.5	Solution high availability .....	61
5.6	Communication flow for XenDesktop .....	62
6	<b>Solution performance and testing.....</b>	<b>63</b>
6.1	Summary .....	63
6.2	Test and performance analysis methodology.....	64
6.2.1	Testing process .....	64

6.2.2 Resource monitoring .....	66
6.2.3 Resource utilization .....	66
6.3 Test configuration details.....	67
6.3.1 Compute VM configurations .....	68
6.4 Standard VDI test results analysis.....	69
6.4.1 XC740xd-C7 .....	71
6.5 vGPU test results and analysis.....	94
6.5.1 XC740xd-C7 with Tesla M60.....	95
<b>A Related resources .....</b>	<b>105</b>

## Executive summary

This document provides the reference architecture for integrating Dell EMC XC Series Hyper-Converged Appliances and the Citrix® software suite (XenDesktop™ and XenApp™) to create virtual desktop infrastructure environments.

The Dell EMC XC Series is a hyper-converged solution that combines storage, compute, networking, and virtualization using industry-proven Dell EMC PowerEdge™ server technology and Nutanix™ software. By combining the hardware resources from each appliance into a shared-everything model for simplified operations, improved agility, and greater flexibility, Dell EMC and Nutanix together deliver simple, cost-effective solutions for enterprise workloads.

Citrix XenDesktop provides a complete end-to-end virtualization solution delivering Microsoft Windows virtual desktops or server-based hosted shared sessions to users on a wide variety of endpoint devices.

# 1 Introduction

This document addresses the architecture design, configuration and implementation considerations for the key components required to deliver virtual desktops or shared sessions via Citrix XenDesktop and XenApp on Microsoft® Windows® Server® Hyper-V® 2012 R2, VMware® vSphere™ 6, or Citrix XenServer 7 running on the Dell EMC XC Series Hyper-Converged Appliances.

For manuals, support info, tools, and videos, please visit: [www.Dell.com/xcseriesmanuals](http://www.Dell.com/xcseriesmanuals).

## 1.1 Objective

Relative to delivering the virtual desktop environment, the objectives of this document are to:

- Define the detailed technical design for the solution.
- Define the hardware requirements to support the design.
- Define the constraints which are relevant to the design.
- Define relevant risks, issues, assumptions and concessions – referencing existing ones where possible.
- Provide a breakdown of the design into key elements such that the reader receives an incremental or modular explanation of the design.
- Provide solution scaling and component selection guidance.

## 1.2 What's new

- XC Core
- One-Click Cloud integration of the Citrix Ready Workspace Appliance on XenServer

## 2 Solution architecture overview

### 2.1 Introduction

Dell EMC customers benefit in leveraging this integrated solution for their primary workload data protection needs. This integrated solution offers Virtual Machine (VM) deployment and lifecycle management for the combined solution offering. Protection for newly deployed and existing VMs. Usage of policies and best practices and the consequent streamlining of the data protection workflow are the primary goals for this solution. This section will provide an overview of the products used to validate the solution.

### 2.2 Dell EMC XC Series Hyper-Converged Appliances

Dell EMC XC Series Hyper-converged Appliances start with the proven Dell EMC PowerEdge 14<sup>th</sup> generation server platform and incorporate many of the advanced software technologies that power leading web-scale and cloud infrastructures. Backed by Dell EMC global service and support, these 1- and 2U appliances are preconfigured for specific virtualized workloads, and are designed to maintain data availability in case of node and disk failure.

The XC Series infrastructure is a scalable cluster of high-performance appliances, or servers, each running a standard hypervisor and containing processors, memory, and local storage (consisting of solid state disk (SSD) flash for high performance and high-capacity disk drives), hybrid or all-flash. Each appliance runs virtual machines just like a standard hypervisor host as displayed below.

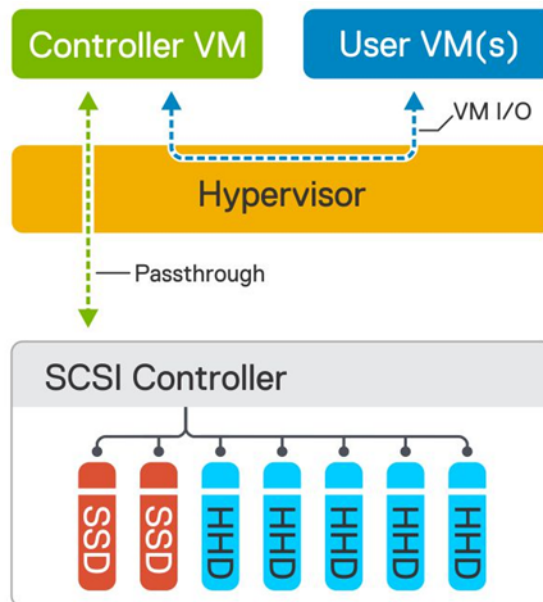


Figure 1 Architecture overview

### 2.3 Dell EMC XC Core

XC Core uses the same PowerEdge hardware and software as the XC Series appliances, however, the HCI software is licensed and supported directly by Nutanix. Support and service for Dell EMC hardware and system integration software are provided through our ProSupport centers and teams located in 167 countries around the world.



This alternative lets customers buy Nutanix software licenses from authorized partners, and then add the licenses to pre-validated XC Core systems that are configured, built and tested by Dell EMC. It enables license portability across infrastructure components and separate management and support of hardware and Nutanix software lifecycles. Customers also can update the Dell EMC hardware and Nutanix software independently to take full advantage of the latest technology enhancements to the XC Family.

For additional information, please visit: [Link](#)

## 2.4 Distributed Storage Fabric

The Distributed Storage Fabric (DSF) delivers enterprise data storage as an on-demand service by employing a highly distributed software architecture. Nutanix eliminates the need for traditional SAN and NAS solutions while delivering a rich set of VM-centric software-defined services. Specifically, the DSF handles the data path of such features as snapshots, clones, high availability, disaster recovery, deduplication, compression, and erasure coding.

The DSF operates via an interconnected network of Controller VMs (CVMs) that form a Nutanix cluster, and every node in the cluster has access to data from shared SSD, HDD, and cloud resources. The hypervisors and the DSF communicate using the industry-standard NFS, iSCSI, and SMB3 protocols, depending on the hypervisor in use.

## 2.5 App Mobility Fabric

The App Mobility Fabric (AMF) collects powerful technologies that give IT professionals the freedom to choose the best environment for their enterprise applications. The AMF encompasses a broad range of capabilities for allowing applications and data to move freely between runtime environments, including between Nutanix systems supporting different hypervisors, and from Nutanix to public clouds. When VMs can migrate between hypervisors, administrators can host production and development or test environments concurrently on different hypervisors and shift workloads between them as needed. AMF is implemented via a distributed, scale-out service that runs inside the CVM on every node within a Nutanix cluster.

### 2.5.1 Nutanix architecture

Nutanix software provides a hyper-converged platform that uses DSF to share and present local storage to server nodes within a cluster while creating a clustered volume namespace accessible to all nodes. The figure below shows an overview of the Nutanix architecture including, user VMs, the Nutanix storage CVM, and its local disk devices. Each CVM connects directly to the local storage controller and its associated disks. Using local storage controllers on each host localizes access to data through the DSF, thereby reducing storage I/O latency. The DSF replicates writes synchronously to at least one other XC Series node in the system, distributing data throughout the cluster for resiliency and availability. Replication factor 2 (RF2) creates two identical data copies in the cluster, and replication factor 3 (RF3) creates three identical data copies.

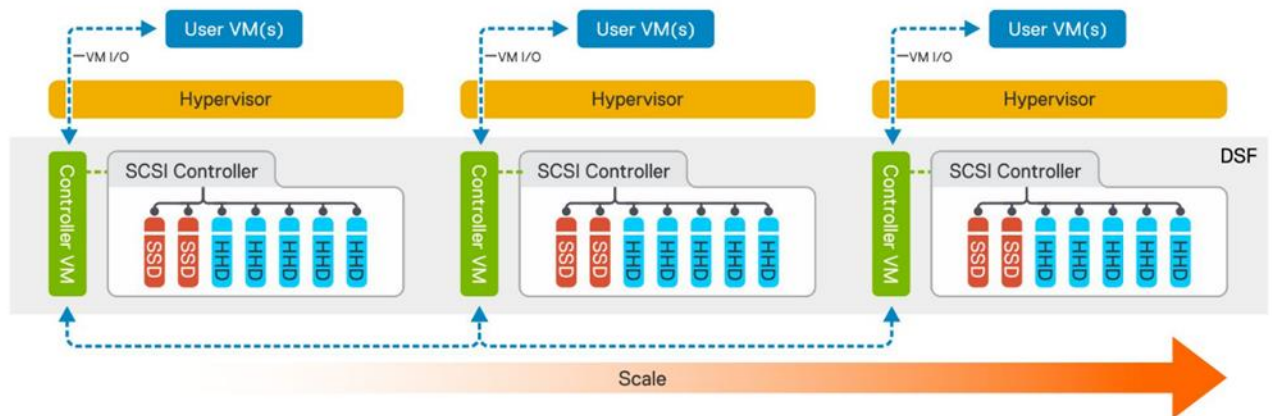


Figure 2 Nutanix architecture overview

DSF virtualizes local storage from all appliances into a unified pool. DSF uses local SSDs and capacity disks from all appliances to store virtual machine data. Virtual machines running on the cluster write data to DSF as if they were writing to local storage. Nutanix data locality ensures that the XC Series node providing CPU and memory to a VM also provides its disk as well, thus minimizing IO that must cross the network. XC Series supports multiple hypervisors and provides choice and flexibility to customer.

XC Series offers customer choice of hypervisors without being locked-in. The hypervisors covered in this reference architecture are:

- VMware ESXi
- Microsoft Hyper-V
- XenServer

In addition, the solution includes the Nutanix Controller VM (CVM), which runs the Nutanix software and serves I/O operations for the hypervisor and all VMs running on that host. Each CVM connects directly to the local storage controller and its associated disks thereby reducing the storage I/O latency. The data locality feature ensures virtual machine I/Os are served by the local CVM on the same hypervisor appliance, improving the VM I/O performance regardless of where it runs.

The Nutanix solution has no LUNs to manage, no RAID groups to configure, and no complicated storage multipathing to set up since there is no reliance on traditional SAN or NAS. All storage management is VM-centric, and the DSF optimizes I/O at the VM virtual disk level. There is one shared pool of storage that includes flash-based SSDs for high performance and low-latency HDDs for affordable capacity. The file system automatically tiers data across different types of storage devices using intelligent data placement algorithms. These algorithms make sure that the most frequently used data is available in memory or in flash for optimal performance. Organizations can also choose flash-only storage for the fastest possible storage performance. The following figure illustrates the data I/O path for a write in a hybrid model with a mix of SSD and HDD disks.

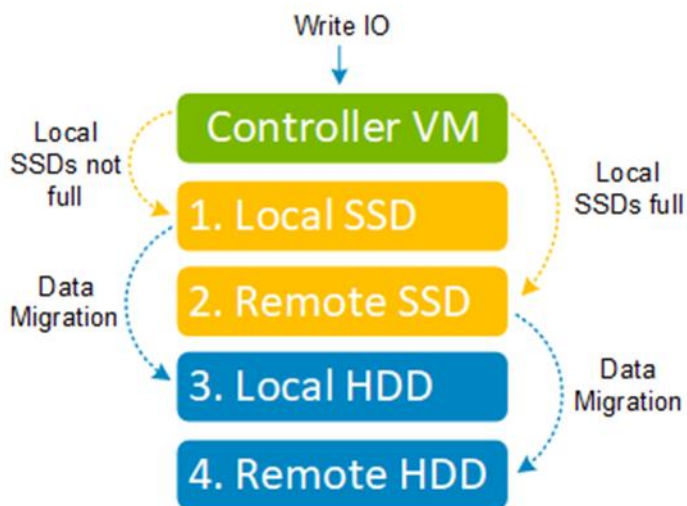


Figure 3 Data I/O path for a write in a hybrid model

Local storage for each XC Series node in the architecture appears to the hypervisor as one large pool of shared storage. This allows the DSF to support all key virtualization features. Data localization maintains performance and quality of service (QoS) on each host, minimizing the effect noisy VMs have on their neighbors' performance. This functionality allows for large, mixed-workload clusters that are more efficient and more resilient to failure when compared to traditional architectures with standalone, shared, and dual-controller storage arrays.

When VMs move from one hypervisor to another, such as during live migration or a high availability (HA) event, the now local CVM serves a newly migrated VMs data. While all write I/O occurs locally, when the local CVM reads old data stored on the now remote CVM, the local CVM forwards the I/O request to the remote CVM. The DSF detects that I/O is occurring from a different node and migrates the data to the local node in the background, ensuring that all read I/O is served locally as well. The next figure shows how data follows the VM as it moves between hypervisor nodes.

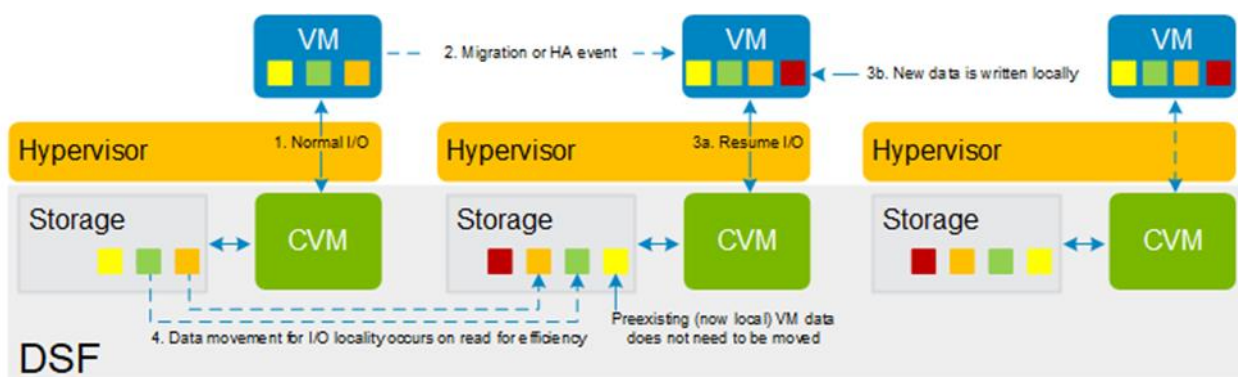


Figure 4 Data follows the VM as it moves between hypervisor nodes

Nutanix Shadow Clones delivers distributed localized caching of virtual disks performance in multi-reader scenarios, such as desktop virtualization using Citrix XenDesktop or XenApp. With Shadow Clones, the CVM actively monitors virtual disk access trends. If there are requests originating from more than two remote CVMs, as well as the local CVM, and all of the requests are read I/O and the virtual disk will be marked as immutable. When the disk is immutable, each CVM then caches it locally, so local storage can now satisfy read operations. Shadow Clones also help to reduce latency when using App Volumes, as you can cache the applications to local memory.

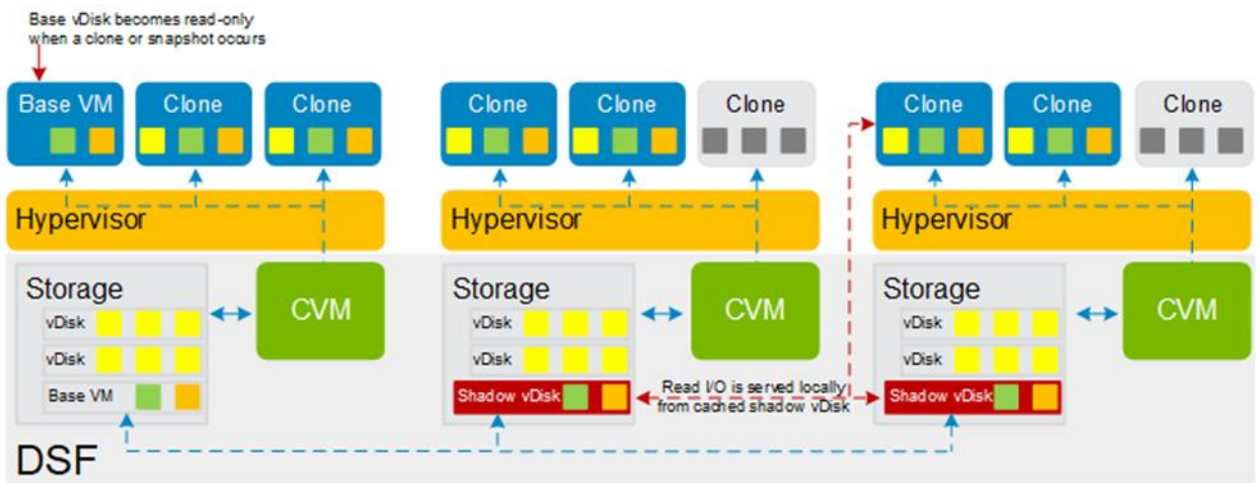


Figure 5 Distributed localized caching of virtual disks

## 2.6 Nutanix Hyper-Converged Infrastructure

The Nutanix hyper-converged infrastructure provides an ideal combination of both high-performance compute with localized storage to meet any demand. True to this capability, this reference architecture has been validated as optimized for the VDI use case.

The next figure shows a high-level example of the relationship between an XC Series node, storage pool, container, pod and relative scale out:

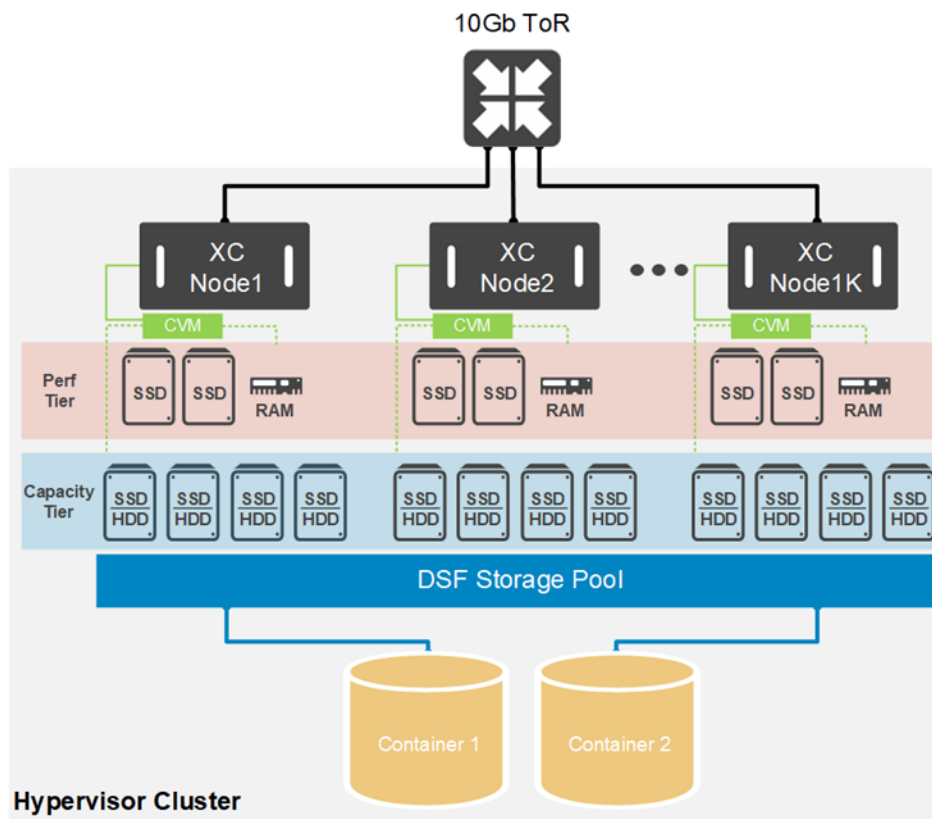


Figure 6 Relationship between an XC Series node and storage pool

This solution allows organizations to deliver virtualized or remote desktops and applications through a single platform and support end users with access to all of their desktops and applications in a single place.

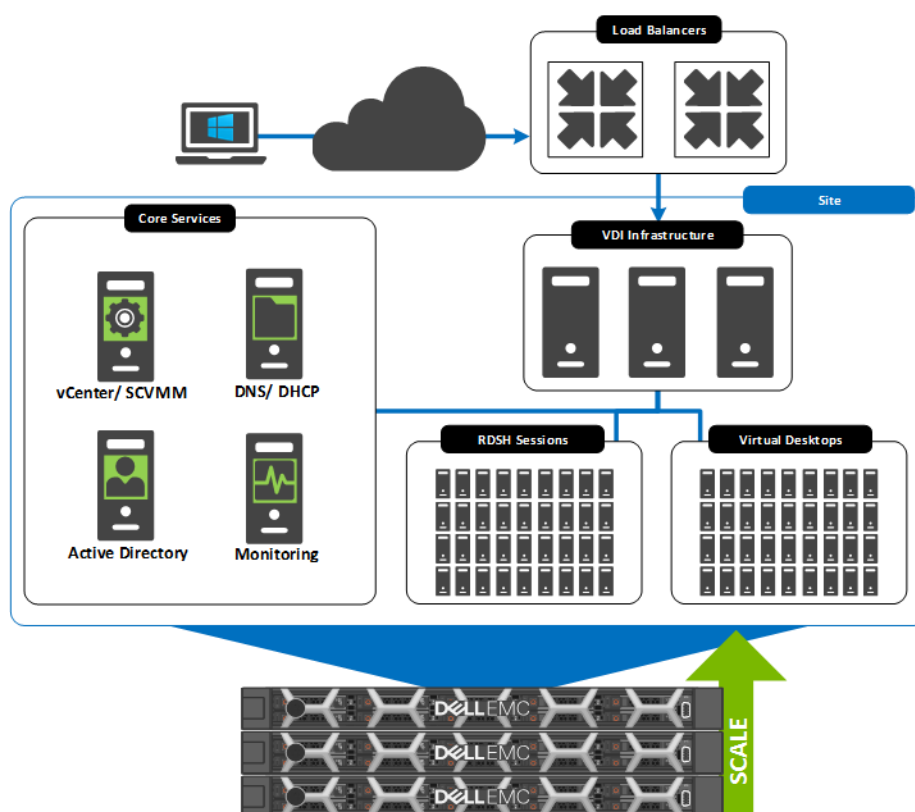


Figure 7 Virtual desktop delivery overview

## 2.7 Nutanix all-flash

Nutanix supports an all-flash configuration where all local disks are SSDs and therefore, the storage pool is fully comprised of SSDs for both capacity and performance. The previously described features and functionality for management, data optimization and protection, and disaster recovery are still present. With all-flash, hot data is stored on SSDs local to each VM. If capacity needs exceed the local SSD storage, capacity on other nodes is automatically and transparently utilized. Compared to traditional all-flash shared storage arrays, XC Series all-flash clusters won't have the typical performance limitations due to network and storage controller bottlenecks. Benefits for VDI include faster provisioning times, low latency, ability to handle extremely high application I/O needs, and accommodating bursts of activity such as boot storms and anti-virus scans.

## 2.8 Dell EMC XC Series - VDI solution architecture

### 2.8.1 Networking

The networking layer consists of the 10 Gb Dell Networking S4048 utilized to build a leaf/spine architecture with robust 1Gb switching in the S3048 for iDRAC connectivity.

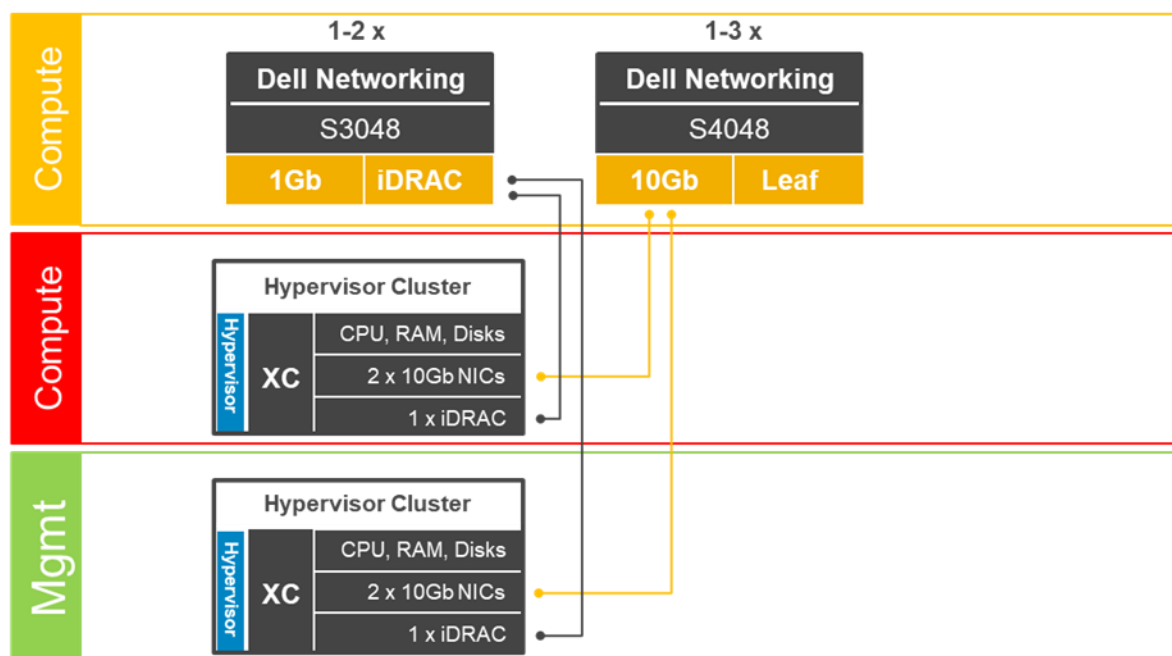


Figure 8 Networking layer

Designed for true linear scaling, XC Series leverages a Leaf-Spine network architecture. A Leaf-Spine architecture consists of two network tiers: a 10Gb layer-2 (L2) Leaf segment and a layer-3 (L3) Spine segment based on 40GbE and non-blocking switches. This architecture maintains consistent performance without any throughput reduction due to a static maximum of three hops from any node in the network.

The following figure shows a design of a scale-out Leaf-Spine network architecture that provides 20Gb active throughput from each node to its Leaf and scalable 80Gb active throughput from each Leaf to Spine switch providing scale from 3 XC Series nodes to thousands without any impact to available bandwidth:



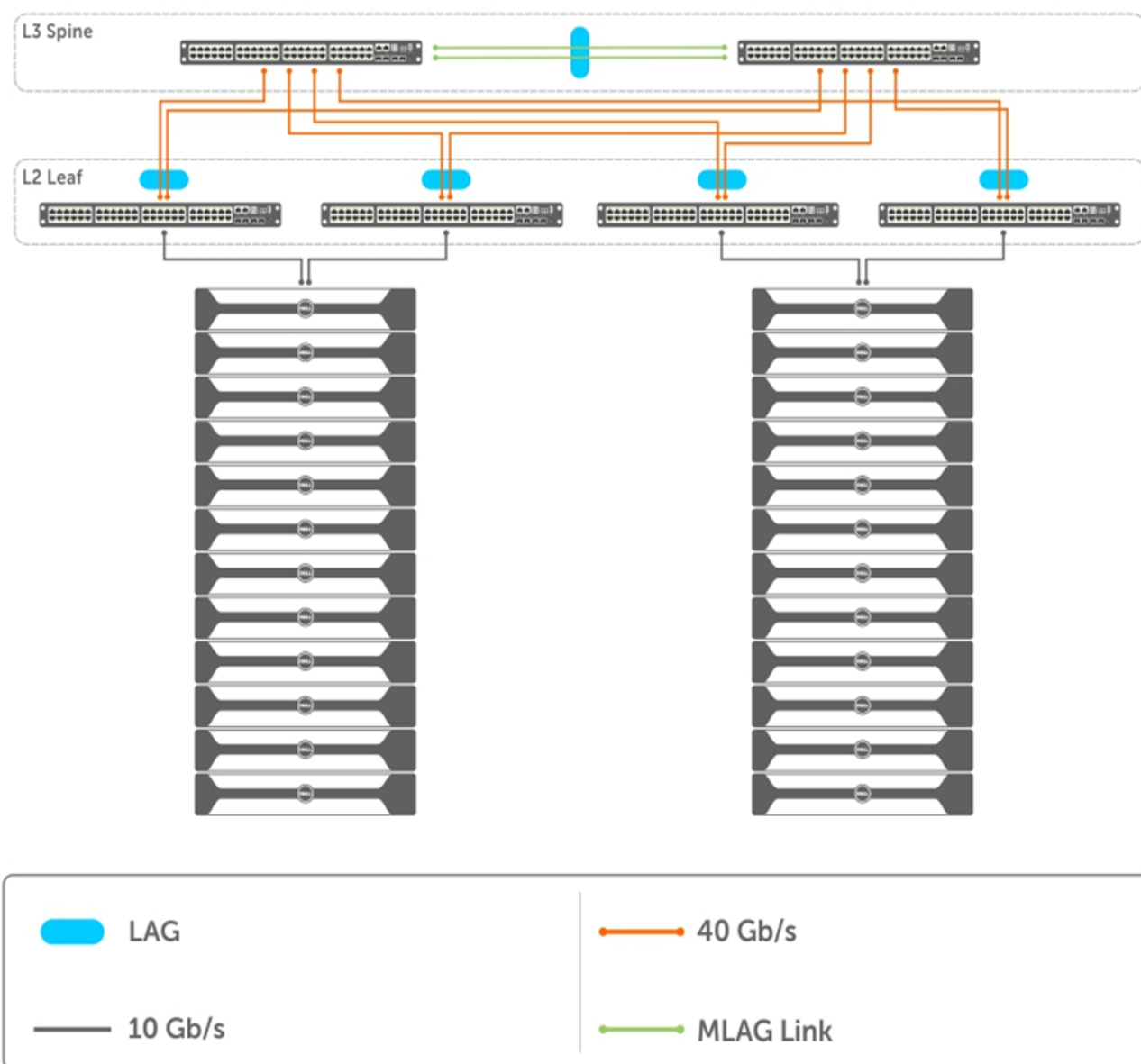


Figure 9 Scale-out Leaf-Spine network architecture

## 2.8.2 XC Series – Enterprise solution pods

The compute, management and storage layers are converged into each XC Series node in the cluster, hosting VMware vSphere, or Microsoft Hyper-V. The recommended boundaries of an individual pod are based on number of nodes supported within a given hypervisor cluster, 64 nodes for vSphere 6 and Hyper-V, although the Nutanix ADFS cluster can scale much larger, well beyond the boundaries of the hypervisor in use.

Dell EMC recommends that the VDI management infrastructure nodes be separated from the compute resources onto their own appliance cluster with a common DSF namespace shared between them based on NFS for vSphere or SMB for Hyper-V. One node for VDI management is required, minimally, and expanded based on size of the pod. The designations `ds_rdsh`, `ds_compute`, `ds_vgpu` and `ds_mgmt` as seen below are logical DSF containers used to group VMs of a particular type.

Using distinct containers allows features and attributes, such as compression and deduplication, to be applied to groups of VMs that share similar characteristics. Compute hosts can be used interchangeably for XenApp or Microsoft Remote Desktop Session Hosts (RDSH) as required. Distinct clusters should be built for management and compute hosts for HA, respectively, to plan predictable failover, scale and load across the pod. The DSF namespace can be shared across multiple hypervisor clusters adding disk capacity and performance for each distinct cluster.

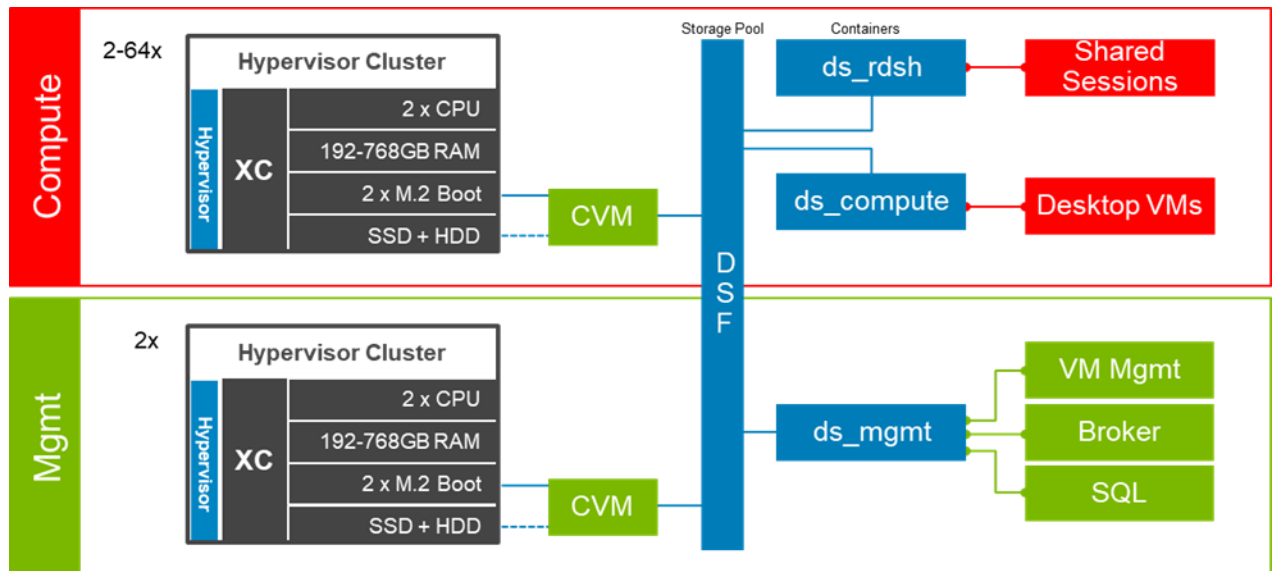


Figure 10 DSF namespace can be shared across multiple hypervisor clusters



High-performance graphics capabilities compliment the solution and can be added at any time to any new or existing XC Series vSphere-based deployment. Simply add the appropriate number of XC740xd appliances to your DSF cluster and provide a superior user experience with vSphere 6 and NVIDIA GRID vGPU technology. Any XC Series appliance can be utilized for the non-graphics compute or management portions of this solution and vSphere will provide HA accordingly based on the type of VM.

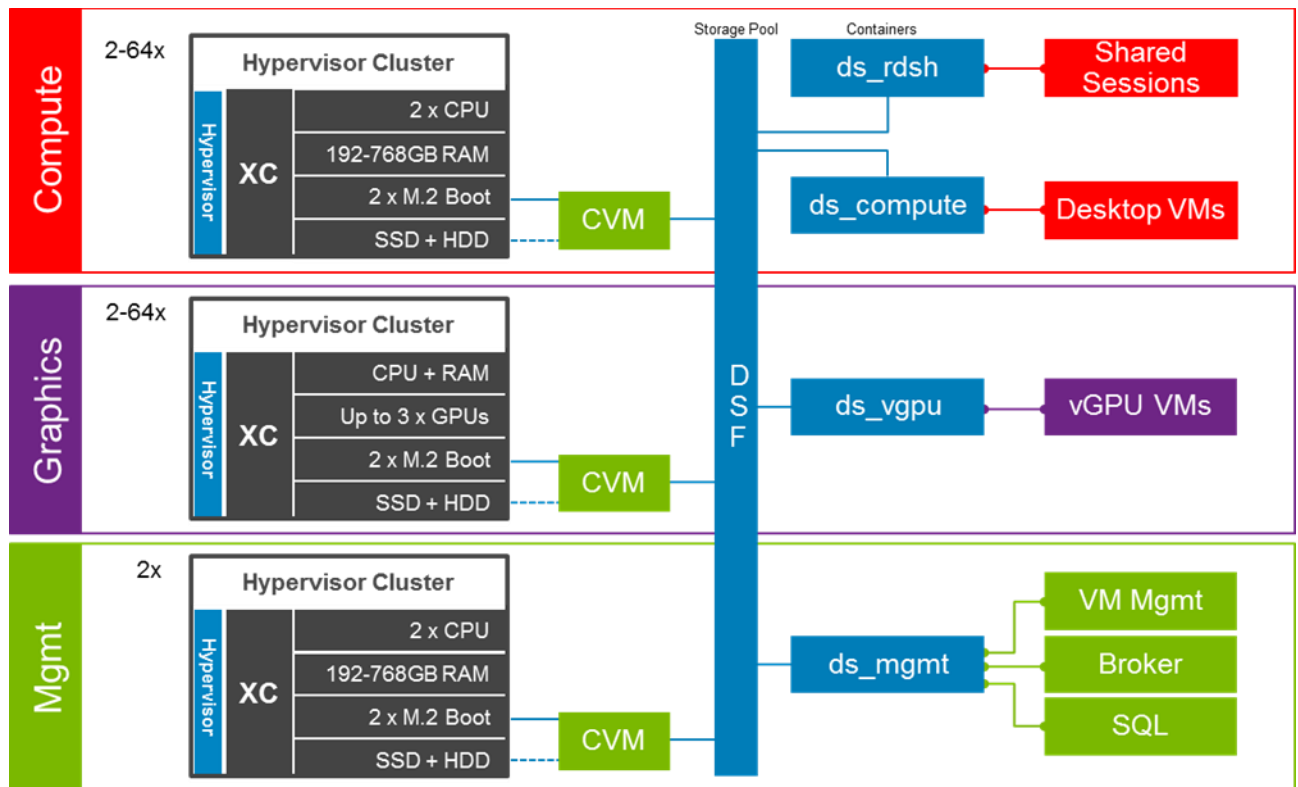


Figure 11 High performance graphics capabilities added to an existing environment

## 3 Hardware components

### 3.1 Network

The following sections contain the core network components for the solution. General uplink cabling guidance to consider in all cases is that TwinAx or CAT6 is very cost effective for short 10Gb runs and for longer runs use fiber with SFPs.

#### 3.1.1 Dell Networking S3048 (1Gb ToR switch)

Accelerate applications in high-performance environments with a low-latency top-of-rack (ToR) switch that features 48 x 1GbE and 4 x 10GbE ports, a dense 1U design and up to 260Gbps performance. The S3048-ON also supports Open Network Installation Environment (ONIE) for zero-touch installation of alternate network operating systems.

Table 1 Dell Networking S3048 features

Model	Features	Options	Uses
• Dell Networking S3048-ON	• 48 x 1000BaseT • 4 x 10Gb SFP+ • Non-blocking, line-rate performance • 260Gbps full-duplex bandwidth • 131 Mpps forwarding rate	• Redundant hot-swap PSUs & fans	• 1Gb connectivity
		• VRF-lite, Routed VLT, VLT Proxy Gateway	
		• User port stacking (up to 6 switches)	
		• Open Networking Install Environment (ONIE)	

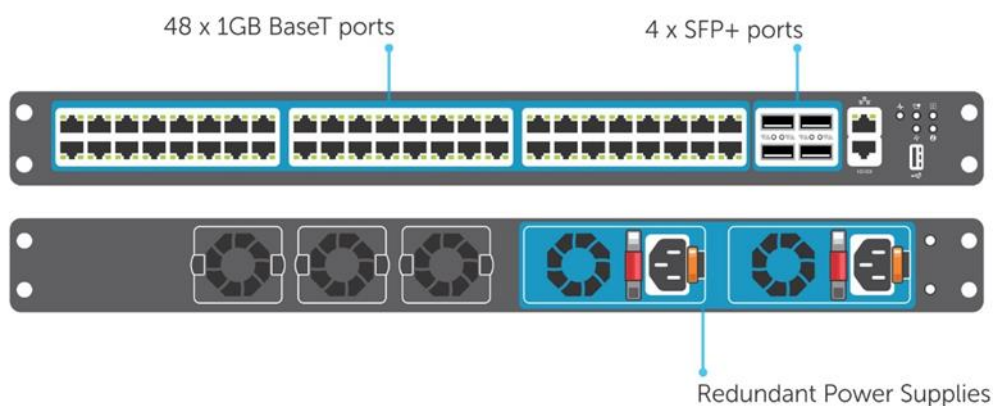


Figure 12 Dell Networking S3048

### 3.1.2 Dell Networking S4048 (10Gb ToR switch)

Optimize your network for virtualization with a high-density, ultra-low-latency ToR switch that features 48 x 10GbE SFP+ and 6 x 40GbE ports (or 72 x 10GbE ports in breakout mode) and up to 720Gbps performance. The S4048-ON also supports ONIE for zero-touch installation of alternate network operating systems.

Table 2 Dell Networking S4048 features

Model	Features	Options	Uses
<ul style="list-style-type: none"> <li>Dell Networking S4048-ON</li> </ul>	<ul style="list-style-type: none"> <li>48 x 10Gb SFP+</li> <li>6 x 40Gb QSFP+</li> <li>Non-blocking, line-rate performance</li> <li>1.44Tbps bandwidth</li> <li>720 Gbps forwarding rate</li> <li>VXLAN gateway support</li> </ul>	<ul style="list-style-type: none"> <li>Redundant hot-swap PSUs &amp; fans</li> </ul>	<ul style="list-style-type: none"> <li>10Gb connectivity</li> </ul>
		<ul style="list-style-type: none"> <li>72 x 10Gb SFP+ ports with breakout cables</li> </ul>	
		<ul style="list-style-type: none"> <li>User port stacking (up to 6 switches)</li> </ul>	
		<ul style="list-style-type: none"> <li>Open Networking Install Environment (ONIE)</li> </ul>	

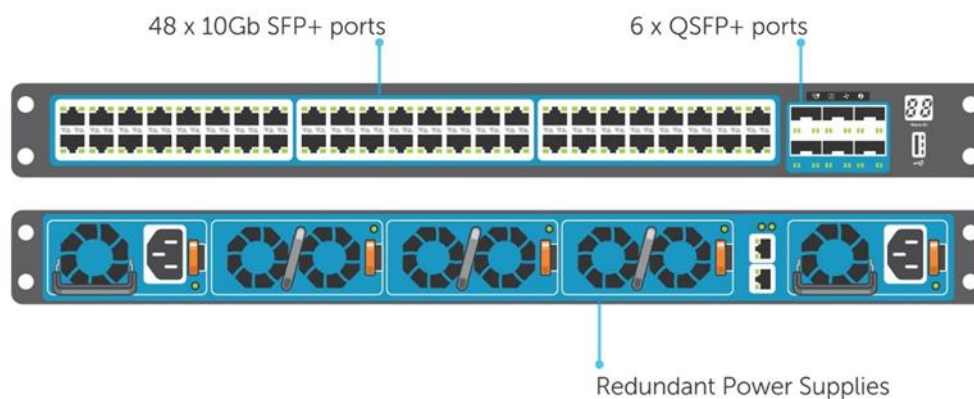


Figure 13 Dell Networking S4048

For more information on the S3048, S4048 switches and Dell Networking, please visit: [Link](#)

## 3.2 Dell EMC XC Series hardware changes between generations

Dell EMC XC Series is based on the award-winning 14th generation of Dell EMC PowerEdge servers which offer a number of performance and feature enhancements. The table below outlines the hardware changes between generations.

Table 3 Comparison of generations

	<b>XC630</b>	<b>XC640</b>	<b>XC730xd</b>	<b>XC740xd</b>	<b>XC730 to XC740 Increase</b>
CPU and chipset	Broadwell-EP	Skylake	Broadwell-EP	Skylake	
Front side bus	Intel QuickPath Interconnect @ 9.6 GT/s	Intel UltraPath Interconnect @ 11.2 GT/s	Intel QuickPath Interconnect @ 9.6 GT/s	Intel UltraPath Interconnect @ 10.4 GT/s	8%
Cores (max)	18 cores	28 cores	22 cores	28 cores	27%
TDP (max)	145 W	205 W	145 W	205 W	
Instruction set	AVX2	AVX2/ AVX-512	AVX2	AVX2/ AVX-512	
Max DP FLOPS / CLK	16 per core (w /AVX2)	32 per core (w / AVX-512)	16 per core (w /AVX2)	32 per core (w / AVX-512)	100%
Memory channels per socket	4 channels, DDR4	6 channels, DDR4	4 channels, DDR4	6 channels, DDR4	50%
Memory (max)	384 GB/ socket (768 GB total)	768 GB/ socket (1.5 TB total)	768 GB/ socket (1.5 TB total)	1.5 TB / socket (3 TB total)	100%
Memory speed (max)	2133 MT/s	2667 MT/s	2400 MT/s	2667 MT/s	11%
PCIe Lanes	40	48	40	48	20%

Consolidate compute and storage into a single chassis with XC Series Hyper-converged appliances, powered by Nutanix software. XC Series appliances install quickly, integrate easily into any data center, and can be deployed for multiple virtualized workloads including desktop virtualization, test and development, and private cloud projects. For general purpose virtual desktop and virtual application solutions, Dell EMC recommends the XC640 and XC740xd. For workloads requiring graphics the XC740xd with NVIDIA GRID vGPU can be integrated into any environment running any other XC Series appliance. For small Remote Office – Branch Office scenarios we offer the XC640. For more information on the Dell EMC XC Series, please visit: [Link](#)

The XC Series portfolio, optimized for VDI, has been designed and arranged in three top-level optimized configurations which apply to the available physical platforms showcased below.

- **A3** configuration is perfect for small scale, POC or low-density cost-conscious environments. Available on all standard hybrid platform configurations.
- **B5** configuration is geared toward larger scale general purpose workloads, balancing performance and cost-effectiveness. Available on all XC Series platforms.
- **C7** is the premium configuration offering an abundance of high performance and tiered capacity where user density is maximized. Available on all XC Series platforms.

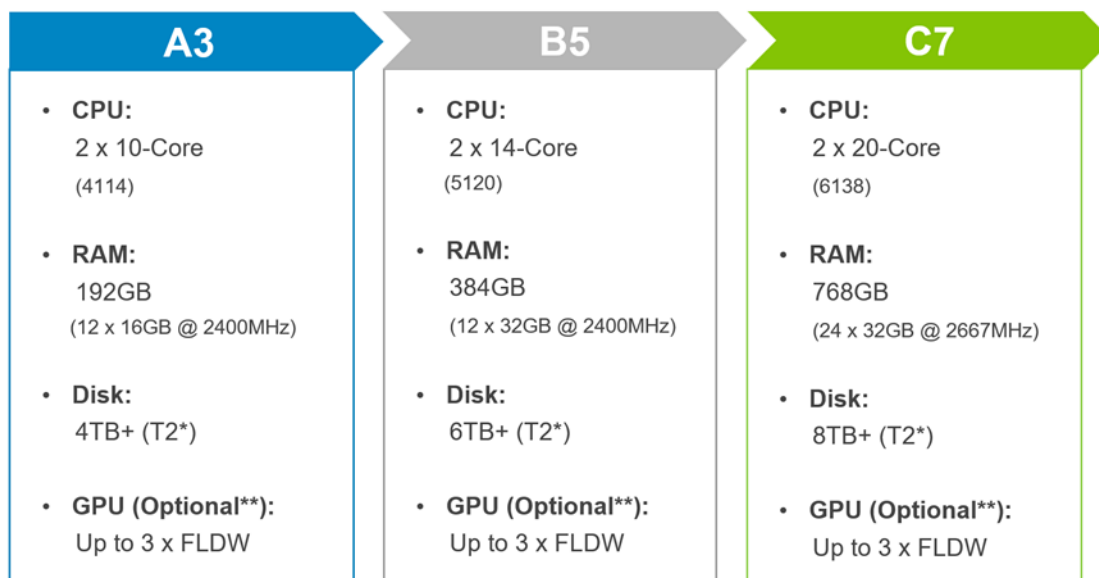


Figure 14 XC Series configurations. \*Raw disk capacity target. \*\*Available on XC740xd

### 3.2.1 Dell EMC XC640

The Dell EMC XC640 is a 10-disk 1U platform with a broad range of configuration options. Each appliance comes equipped with dual CPUs, 10 to 28cores, and up to 1.5TB of high-performance RAM. For the hybrid disk configuration, a minimum of six disks is required in each host, 2 x SSD for the performance tier (Tier1) and 4 x HDD for the capacity tier (Tier2) which can be expanded up to eight HDDs as required. For the all-flash disk configuration, the chassis must be populated with a minimum of 4 x SSDs. The M.2-based BOSS module boots the hypervisor and Nutanix Controller VM while the PERC HBA330 connects the CVM to the SSDs and HDDs. 64GB is consumed on each of the first two SSDs for the Nutanix “home”. All HDD/SSD disks are presented to the Nutanix CVM running locally on each host which contributes to the clustered DSF storage pool. Each platform can be outfitted with SFP+ or BaseT NICs.

### 3.2.1.1 XC640 hybrid disk storage

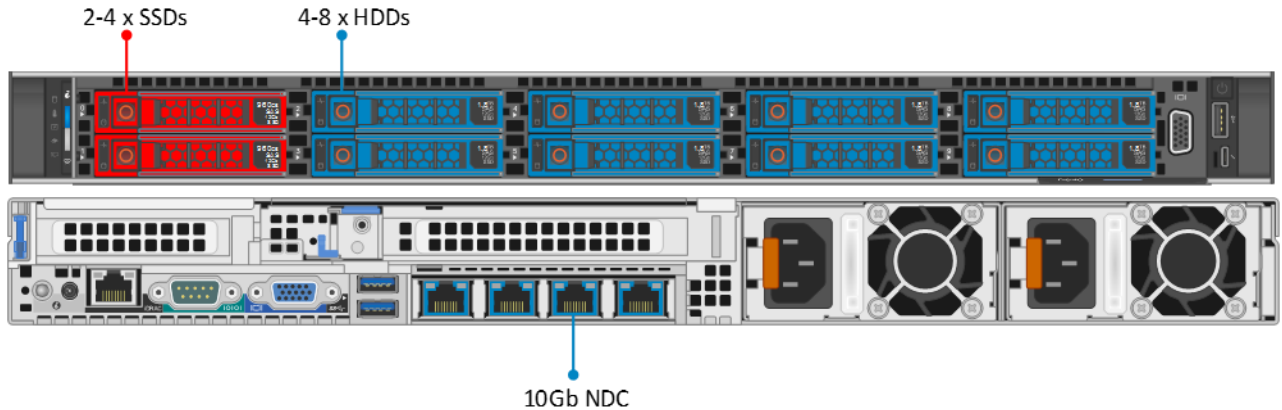


Figure 15 XC640 hybrid disk storage

Table 4 XC640 hybrid configuration details

XC640	A3	B5	C7
<b>Hybrid</b>			
<b>CPU</b>	<ul style="list-style-type: none"> <li>2 x Intel Xeon Silver 4114</li> <li>(10C, 2.2GHz)</li> </ul>	<ul style="list-style-type: none"> <li>2 x Intel Xeon Gold 5120</li> <li>(14C, 2.2GHz)</li> </ul>	<ul style="list-style-type: none"> <li>2 x Intel Gold 6138</li> <li>(20C, 2.0GHz)</li> </ul>
<b>Memory</b>	<ul style="list-style-type: none"> <li>12 x 16GB 2667MT/s RDIMMs</li> <li>Effective speed: 2400MT/s @ 192GB</li> </ul>	<ul style="list-style-type: none"> <li>12 x 32GB 2667MT/s RDIMMs</li> <li>Effective speed: 2400MT/s @ 384GB</li> </ul>	<ul style="list-style-type: none"> <li>24 x 32GB 2667MT/s RDIMMs</li> <li>Effective speed: 2667MT/s @ 768GB</li> </ul>
<b>Storage Ctrl</b>	HBA330 LP		
<b>Storage</b>	<ul style="list-style-type: none"> <li>CVM/ OS: 2 x 120GB M.2 RAID1</li> <li>Tools/ Recovery: 32GB SD</li> <li>T1: 2 x 480GB SSD 2.5"</li> <li>T2: 4 x 1.8TB 2.5"/ 2TB HDD 3.5"</li> </ul>	<ul style="list-style-type: none"> <li>CVM/ OS: 2 x 120GB M.2 RAID1</li> <li>Tools/ Recovery: 32GB SD</li> <li>T1: 2 x 960GB SSD 2.5"</li> <li>T2: 4 x 1.8TB 2.5"/ 2TB HDD 3.5"</li> </ul>	<ul style="list-style-type: none"> <li>CVM/ OS: 2 x 120GB M.2 RAID1</li> <li>Tools/ Recovery: 32GB SD</li> <li>T1: 2 x 960GB SSD 2.5"</li> <li>T2: 6 x 1.8TB 2.5"/ 2TB HDD 3.5"</li> </ul>
<b>Network</b>	2 x 10Gb, 2 x 1Gb SFP+/ BT		
<b>iDRAC</b>	iDRAC9 Enterprise		
<b>Power</b>	2 x 1100W PSUs		

### 3.2.1.2 XC640 all-flash disk storage

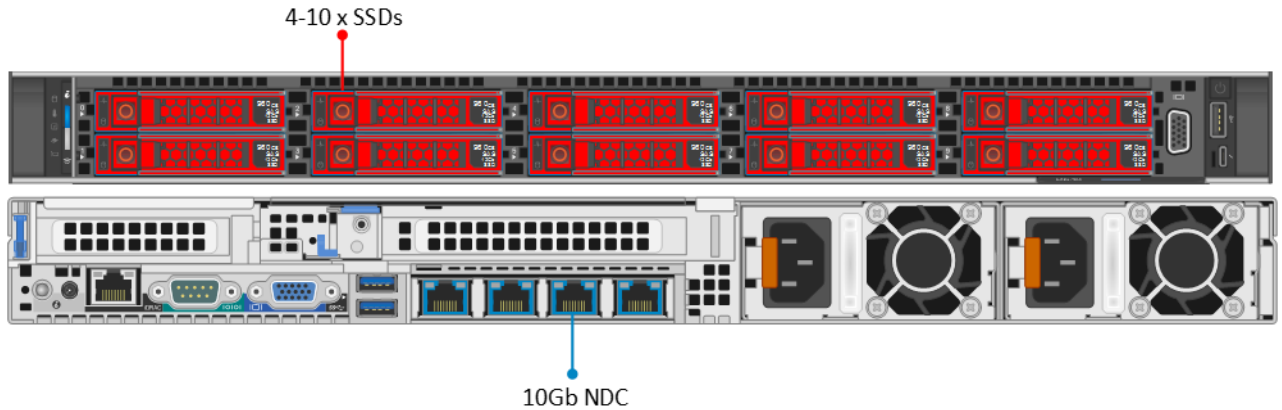


Figure 16 XC640 all-flash disk storage

Table 5 XC640 all-flash configuration details

<b>XC640</b>	<b>B5-AF</b>	<b>C7-AF</b>
<b>All-Flash</b>		
<b>CPU</b>	<ul style="list-style-type: none"> <li>• 2 x Intel Xeon Gold 5120</li> <li>• (14C, 2.2GHz)</li> </ul>	<ul style="list-style-type: none"> <li>• 2 x Intel Gold 6138</li> <li>• (20C, 2.0GHz)</li> </ul>
<b>Memory</b>	<ul style="list-style-type: none"> <li>• 12 x 32GB 2667MT/s RDIMMs</li> <li>• Effective speed: 2400MT/s @ 384GB</li> </ul>	<ul style="list-style-type: none"> <li>• 24 x 32GB 2667MT/s RDIMMs</li> <li>• Effective speed: 2667MT/s @ 768GB</li> </ul>
<b>Storage Ctrl</b>	HBA330 LP	
<b>Storage</b>	<ul style="list-style-type: none"> <li>• <b>CVM/ OS:</b> 2 x 120GB M.2 RAID1</li> <li>• Tools/ Recovery: 32GB SD</li> <li>• <b>T1/ T2:</b> 6 x 960GB SSD 2.5"</li> </ul>	<ul style="list-style-type: none"> <li>• <b>CVM/ OS:</b> 2 x 120GB M.2 RAID1</li> <li>• Tools/ Recovery: 32GB SD</li> <li>• <b>T1/ T2:</b> 10 x 960GB SSD 2.5"</li> </ul>
<b>Network</b>	2 x 10Gb, 2 x 1Gb SFP+ or BaseT	
<b>iDRAC</b>	iDRAC9 Enterprise	
<b>Power</b>	2 x 1100W PSUs	

All-flash configuration requirements:

- 4 x SSD min XC640-4
- 6 x SSD min XC640-10



### 3.2.2 Dell EMC XC740xd

The Dell EMC XC740xd is a 2U platform that can be configured with 24 x 2.5" disks or 12 x 3.5" disks to serve a broad range of capacity requirements. Each appliance comes equipped with dual CPUs, 10 to 28 cores, and up to 1.5TB of high-performance RAM. A minimum of six disks is required in each host, 2 x SSD for the performance tier (Tier1) and 4 x HDD for the capacity tier (Tier2) which can be expanded as required up to a possible 45TB+ per node raw. The M.2-based BOSS module boots the hypervisor and Nutanix Controller VM while the PERC HBA330 connects the CVM to the SSDs and HDDs. 64GB is consumed on each of the first two SSDs for the Nutanix "home". All HDD/SSD disks are presented to the Nutanix CVM running locally on each host which contributes to the clustered DSF pool. Each platform can be outfitted with SFP+ or BaseT NICs. The 24-disk XC740xd can support up to 3 NVIDIA M60 or 2 x M10 GPU cards. Please note that higher wattage power supplies will also be required when GPUs are in use, up to 2000W per PSU.

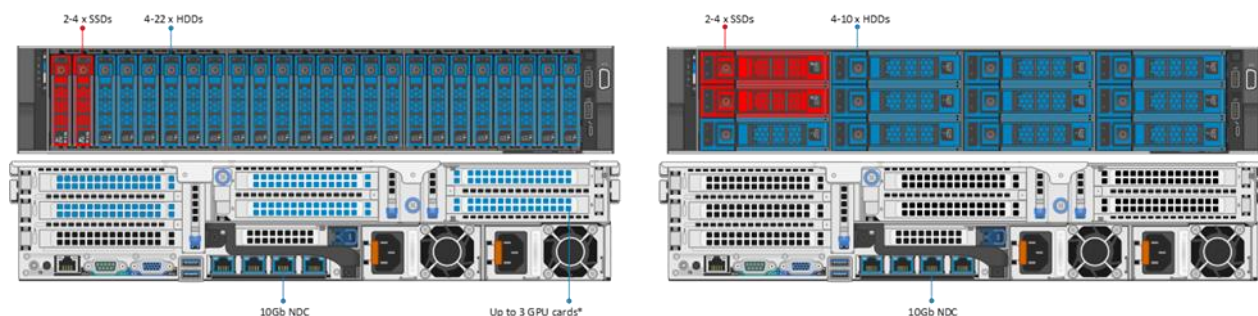


Figure 17 XC740xd

Table 6 XC740xd configuration details

XC740xd	A3	B5	C7
<b>Hybrid</b>			
<b>CPU</b>	<ul style="list-style-type: none"> <li>2 x Intel Xeon Silver 4114</li> <li>(10C, 2.2GHz)</li> </ul>	<ul style="list-style-type: none"> <li>2 x Intel Xeon Gold 5120</li> <li>(14C, 2.2GHz)</li> </ul>	<ul style="list-style-type: none"> <li>2 x Intel Gold 6138</li> <li>(20C, 2.0GHz)</li> </ul>
<b>Memory</b>	<ul style="list-style-type: none"> <li>12 x 16GB 2667MT/s RDIMMs</li> <li>Effective speed: 2400MT/s @ 192GB</li> </ul>	<ul style="list-style-type: none"> <li>12 x 32GB 2667MT/s RDIMMs</li> <li>Effective speed: 2400MT/s @ 384GB</li> </ul>	<ul style="list-style-type: none"> <li>24 x 32GB 2667MT/s RDIMMs</li> <li>Effective speed: 2667MT/s @ 768GB</li> </ul>
<b>Storage Ctrl</b>	HBA330		
<b>Storage</b>	<ul style="list-style-type: none"> <li><b>CVM/ OS:</b> 2 x 120GB M.2 RAID1</li> <li><b>Tools/ Recovery:</b> 32GB SD</li> <li><b>T1:</b> 2 x 480GB SSD 2.5"</li> <li><b>T2:</b> 4 x 1.8TB 2.5"/ 2TB HDD 3.5"</li> </ul>	<ul style="list-style-type: none"> <li><b>CVM/ OS:</b> 2 x 120GB M.2 RAID1</li> <li><b>Tools/ Recovery:</b> 32GB SD</li> <li><b>T1:</b> 2 x 960GB SSD 2.5"</li> <li><b>T2:</b> 4 x 1.8TB 2.5"/ 2TB HDD 3.5"</li> </ul>	<ul style="list-style-type: none"> <li><b>CVM/ OS:</b> 2 x 120GB M.2 RAID1</li> <li><b>Tools/ Recovery:</b> 32GB SD</li> <li><b>T1:</b> 2 x 960GB SSD 2.5"</li> <li><b>T2:</b> 6 x 1.8TB 2.5"/ 2TB HDD 3.5"</li> </ul>
<b>GPU</b>	<ul style="list-style-type: none"> <li>2 x Tesla M10 or</li> </ul>		



	<ul style="list-style-type: none"> <li>• 3 x Tesla M60</li> </ul>
<b>Network</b>	2 x 10Gb, 2 x 1Gb SFP+ or BaseT
<b>iDRAC</b>	iDRAC9 Enterprise
<b>Power</b>	2 x 1100W PSUs (2 x 2000w PSUs for GPU)

All-flash configuration requirements:

- 6 x SSD min XC740, XC740xd-12
- 12 x SSD min XC740xd-24

## 3.3 NVIDIA Tesla GPUs

Accelerate your most demanding enterprise data center workloads with NVIDIA® Tesla® GPU accelerators. Scientists can now crunch through petabytes of data up to 10x faster than with CPUs in applications ranging from energy exploration to deep learning. Plus, Tesla accelerators deliver the horsepower needed to run bigger simulations faster than ever before. For enterprises deploying VDI, Tesla accelerators are perfect for accelerating virtual desktops. GPUs can only be used with the Dell EMC XC730 platform.



### 3.3.1 NVIDIA Tesla M10

The NVIDIA® Tesla® M10 is a dual-slot 10.5 inch PCI Express Gen3 graphics card featuring four mid-range NVIDIA Maxwell™ GPUs and a total of 32GB GDDR5 memory per card (8GB per GPU). The Tesla® M10 doubles the number of H.264 encoders over the NVIDIA® Kepler™ GPUs and improves encoding quality, which enables richer colors, preserves more details after video encoding, and results in a high-quality user experience.

The NVIDIA® Tesla® M10 GPU accelerator works with NVIDIA GRID™ software to deliver the industry's highest user density for virtualized desktops and applications. It supports up to 64 desktops per GPU card using a 1GB framebuffer (up to 128 desktops per server) and gives businesses the power to deliver great graphics experiences to all of their employees at an affordable cost.

Table 7 NVIDIA Tesla M10 specs

Specs	Tesla M10
Number of GPUs/ card	4 x NVIDIA Maxwell™ GPUs
Total CUDA cores	2560 (640 per GPU)
GPU Clock	Idle: 405MHz / Base: 1033MHz
Total memory size	32GB GDDR5 (8GB per GPU)
Max power	225W
Form Factors	Dual slot (4.4" x 10.5")
Aux power	8-pin connector
PCIe	x16 (Gen3)
Cooling solution	Passive

### 3.3.2 NVIDIA Tesla M60

The NVIDIA® Tesla® M60 is a dual-slot 10.5 inch PCI Express Gen3 graphics card featuring two high-end NVIDIA Maxwell™ GPUs and a total of 16GB GDDR5 memory per card. This card utilizes NVIDIA GPU Boost™ technology which dynamically adjusts the GPU clock to achieve maximum performance. Additionally, the Tesla® M60 doubles the number of H.264 encoders over the NVIDIA® Kepler™ GPUs.



The NVIDIA® Tesla® M60 GPU accelerator works with NVIDIA GRID™ software to provide the industry's highest user performance for virtualized workstations, desktops, and applications. It allows enterprises to virtualize almost any application (including professional graphics applications) and deliver them to any device, anywhere. M60 can support 3 cards in the XC740xd providing 48 x Windows10 users assigned a 1GB framebuffer each.

Table 8 NVIDIA Tesla M60 specs

Specs	Tesla M60
Number of GPUs/ card	2 x NVIDIA Maxwell™ GPUs
Total CUDA cores	4096 (2048 per GPU)
Base Clock	899 MHz (Max: 1178 MHz)
Total memory size	16GB GDDR5 (8GB per GPU)
Max power	300W
Form Factors	Dual slot (4.4" x 10.5")
Aux power	8-pin connector
PCIe	x16 (Gen3)
Cooling solution	Passive/ Active

## 3.4 Dell Wyse Endpoints

The following Dell Wyse clients will deliver a superior user experience for Citrix XenDesktop and are the recommended choices for this solution.

### 3.4.1 Wyse 3040 Thin Client (ThinOS, ThinLinux)

The Wyse 3040 is the industry's first entry-level Intel x86 quad-core thin client, powered by a quad-core Intel Atom 1.44GHz processor, delivering robust connectivity options with a choice of Wyse ThinOS or ThinLinux operating systems. The Wyse 3040 is Dell's lightest, smallest and most power-efficient thin client – it consumes 3.3 Watts in idle state – and offers superb performance and manageability for task and basic productivity users. Despite its small size, the 3040 includes all typical interfaces such as four USB ports including USB 3.1, two DisplayPort interfaces and wired and wireless options. It is highly manageable as it can be monitored, maintained, and serviced remotely via Wyse Device Manager (WDM) or Wyse Management Suite. For more information, please visit: [Link](#)



### 3.4.2 Wyse 5040 AIO Thin Client (ThinOS)



The Dell Wyse 5040 AIO all-in-one (AIO) thin client runs ThinOS (with or without PCoIP), has a 21.5" Full HD display and offers versatile connectivity options for use in a wide range of industries. With four USB 2.0 ports, Gigabit Ethernet and integrated dual band Wi-Fi options, users can link to their peripherals and quickly connect to the network while working with processing-intensive, graphics-rich applications. Built-in speakers, a camera and a microphone make video conferencing and desktop communication simple and easy. It even supports a second attached display for those who need a dual monitor configuration. A simple one-cord design and out-of-box automatic setup makes deployment effortless while

remote management from a simple file server, Wyse Device Manager (WDM), or Wyse Management Suite can help lower your total cost of ownership as you grow from just a few thin clients to tens of thousands. For more information, please visit: [Link](#)

### 3.4.3 Wyse 5060 Thin Client (ThinOS, ThinLinux, WES7P, WIE10)

The Wyse 5060 offers high performance and reliability, featuring all the security and management benefits of Dell thin clients. It comes with flexible OS options: ThinOS (with or without PCoIP), ThinLinux, Windows Embedded Standard 7P (WES7P) or Windows 10 IoT Enterprise (WIE10). Designed for knowledge workers demanding powerful virtual desktop performance, and support for unified communications solutions like Skype for Business, the Wyse 5060 thin client delivers the flexibility, efficiency and security organizations require for their cloud environments. It is powered by a quad-core AMD 2.4GHz processor, supports dual 4K (3840x2160) monitors and provides multiple connectivity options with six USB ports, two of which are USB 3.0 for high-speed peripherals, as well as two DisplayPort connectors, wired networking or wireless 802.11 a/b/g/n/ac. The Wyse 5060 can be monitored, maintained, and serviced remotely via Wyse Device Manager (WDM), cloud-based Wyse Management Suite or Microsoft SCCM (5060 with Windows versions). For more information, please visit: [Link](#)



### 3.4.4 Wyse 7020 Thin Client (WES 7/7P/8, WIE10, ThinLinux)

The versatile Dell Wyse 7020 thin client is a powerful endpoint platform for virtual desktop environments. It is available with Windows Embedded Standard 7/7P/8 (WES), Windows 10 IoT Enterprise (WIE10), Wyse ThinLinux operating systems and it supports a broad range of fast, flexible connectivity options so that users can connect their favorite peripherals while working with processing-intensive, graphics-rich applications. This 64-bit thin client delivers a great user experience and support for local applications while ensuring security. Designed to provide a superior user experience, ThinLinux features broad broker support including Citrix Receiver, VMware Horizon and Amazon Workspace, and support for unified communication platforms including Skype for Business, Lync 2013 and Lync 2010. For additional security, ThinLinux also supports single sign-on and VPN. With a powerful quad core AMD G Series APU in a compact chassis with dual-HD monitor support, the Wyse 7020 thin client delivers stunning performance and display capabilities across 2D, 3D and HD video applications. Its silent diskless and fan less design helps reduce power usage to just a fraction (it only consumes about 15 watts) of that used in traditional desktops. Wyse Device Manager (WDM) helps lower the total cost of ownership for large deployments and offers remote enterprise-wide management that scales from just a few to tens of thousands of cloud clients. For more information, please visit [Link](#)



### 3.4.5 Wyse 7040 Thin Client (WES7P, WIE10)

The Wyse 7040 is a high-powered, ultra-secure thin client running Windows Embedded Standard 7P (WES7P) or Windows 10 IoT Enterprise (WIE10) operating systems. Equipped with an Intel i5/i7 processors, it delivers extremely high graphical display performance (up to three displays via display-port daisy-chaining, with 4K resolution available on a single monitor) for seamless access to the most demanding applications. The Wyse 7040 is compatible with both data center hosted and client-side virtual desktop environments and is compliant with all relevant U.S. Federal security certifications including OPAL compliant hard-drive options, VPAT/Section 508, NIST BIOS, Energy-Star and EPEAT. Wyse enhanced WES7P OS provides additional security features such as BitLocker. The Wyse 7040 offers a high level of connectivity including dual NIC, 6 x USB3.0 ports and an optional second network port, with either copper or fiber SFP interface. Wyse 7040 devices are highly manageable through Intel vPRO, Wyse Device Manager (WDM), Microsoft System Center Configuration Manager (SCCM) and Dell Command Configure (DCC). For more information, please visit: [Link](#)



#### Enhanced Security

Note that all the above thin clients running Windows Embedded Standard 7 or Windows 10 IoT can be protected against viruses, ransomware and zero-day threats by installing [Dell Threat Defense](#), a revolutionary anti-malware software solution using artificial intelligence and mathematical modeling and is not signature-based. Threat Defense prevents 99percent of executable malware, far above the average 50percent of threats identified by the top anti-virus solutions. It doesn't need a constant internet connection nor frequent updates (only about twice a year), it only uses 1-3percent CPU and has only a ~40MB memory footprint, making it an ideal choice to protect thin clients without impacting the end user productivity.

If you also want to protect virtual desktops against such malware and threats with a similar success, Dell recommends using [Dell Endpoint Security Suite Enterprise](#), a full suite featuring advanced threat prevention and data-centric encryption using an on-premise management console. This suite can also be used to protect physical PCs, MAC OS X systems and Windows Server.

## 4 Software components

### 4.1 Citrix

#### 4.1.1 XenDesktop

The solution is based on Citrix XenDesktop which provides a complete end-to-end solution delivering Microsoft Windows virtual desktops or server-based hosted shared sessions to users on a wide variety of endpoint devices. Virtual desktops are dynamically assembled on demand, providing users with pristine, yet personalized, desktops each time they log on.

Citrix XenDesktop provides a complete virtual desktop delivery system by integrating several distributed components with advanced configuration tools that simplify the creation and real-time management of the virtual desktop infrastructure.

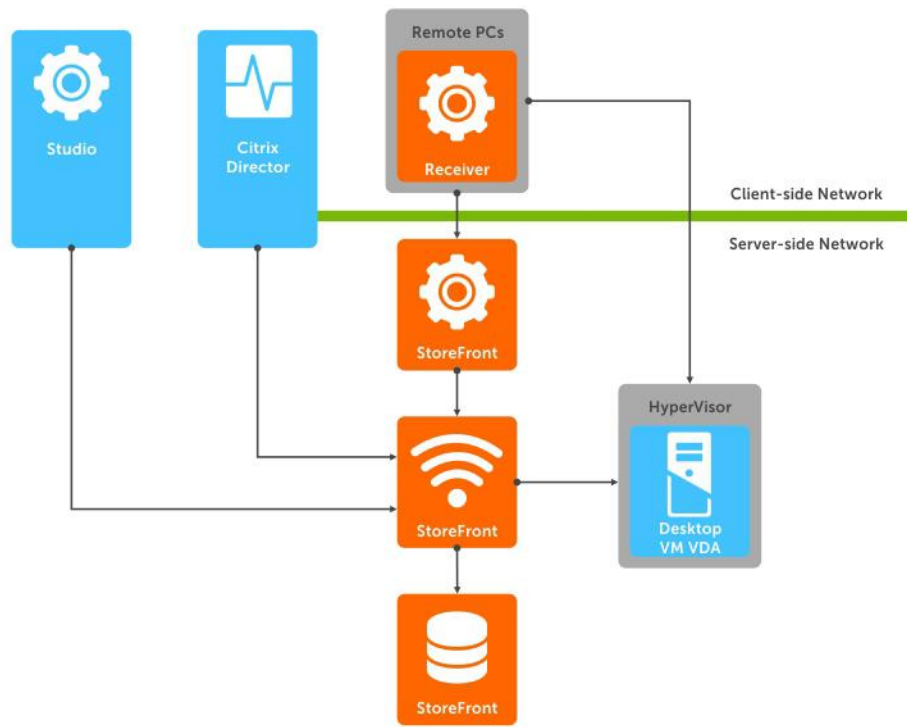


Figure 18 Citrix XenDesktop overview diagram

## **The core XenDesktop components include:**

**Studio** – Studio is the management console that enables you to configure and manage your deployment, eliminating the need for separate management consoles for managing delivery of applications and desktops. Studio provides various wizards to guide you through the process of setting up your environment, creating your workloads to host applications and desktops, and assigning applications and desktops to users.

**Delivery Controller (DC)** – Installed on servers in the data center, the controller authenticates users, manages the assembly of users' virtual desktop environments, and brokers connections between users and their virtual desktops. The Controller also manages the state of desktops, starting and stopping them based on demand and administrative configuration.

**Database** – At least one Microsoft SQL Server database is required for every XenApp or XenDesktop Site to store configuration and session information. The Delivery Controller must have a persistent connection to the database as it stores data collected and managed by the Controller services.

**Director** – Director is a web-based tool that enables IT support teams to monitor an environment, troubleshoot issues before they become system-critical, and perform support tasks for end users. You can also view and interact with a user's sessions using Microsoft Remote Assistance. Starting in version 7.12, Director now includes detailed descriptions for connection and machine failures, one-month historical data (Enterprise edition), custom reporting, and notifications via SNMP traps.

**Receiver** – Installed on user devices, Citrix Receiver provides users with quick, secure, self-service access to documents, applications, and desktops from any of the user's devices including smartphones, tablets, and PCs. Receiver provides on-demand access to Windows, Web, and Software as a Service (SaaS) applications. For devices that cannot install the Receiver software, Citrix Receiver for HTML5 provides connectivity through a HTML5-compatible web browser.

**StoreFront** – StoreFront authenticates users to sites hosting resources and manages stores of desktops and applications that user's access. StoreFront version 3.8 (released with XenDesktop 7.12) and above includes ability to create and use multiple IIS websites each having its own domain name.

**License Server** – The Citrix License Server is an essential component at any Citrix-based solution. Every Citrix product environment must have at least one shared or dedicated license server. License servers are computers that are either partly or completely dedicated to storing and managing licenses. Citrix products request licenses from a license server when users attempt to connect.

**Machine Creation Services (MCS)** – A collection of services that work together to create virtual servers and desktops from a master image on demand; optimizing storage utilization and providing a pristine virtual machine to users every time they log on. Machine Creation Services is fully integrated and administrated in Citrix Studio.

**Provisioning Services (PVS)** – The Provisioning Services infrastructure is based on software-streaming technology. This technology allows computers to be provisioned and re-provisioned in real-time from a single shared-disk image.

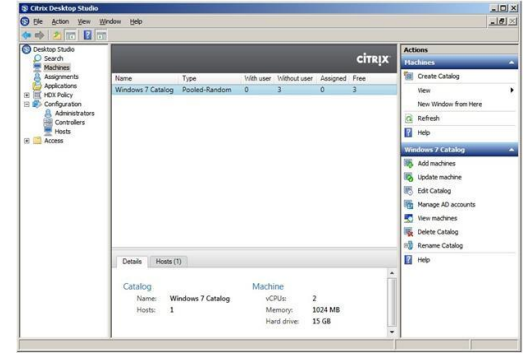
**Virtual Delivery Agent (VDA)** – The Virtual Desktop Agent is a transparent plugin that is installed on every virtual desktop or XenApp host (RDSH) and enables the direct connection between the virtual desktop and users' endpoint devices. Windows and Linux VDAs are available.



## 4.1.2 Machine Creation Services (MCS)

Citrix Machine Creation Services is the native provisioning mechanism within Citrix XenDesktop for virtual desktop image creation and management. Machine Creation Services uses the hypervisor APIs to create, start, stop, and delete virtual desktop images. Desktop images are organized in a Machine Catalog and within that catalog there are a number of options available to create and deploy virtual desktops:

- **Random:** Virtual desktops are assigned randomly as users connect. When they logoff, the desktop is reset to its original state and made free for another user to login and use. Any changes made by the user are discarded at log off.
- **Static:** Virtual desktops are assigned to the same user every time with three options for how to handle changes made to the desktop: Store on local vDisk, Personal vDisk, or discarded on user log off.



All the desktops in a random or static catalog are based off a master desktop template which is selected during the catalog creation process. MCS then takes snapshots of the master template and layers two additional virtual disks on top: an Identity vDisk and a Difference vDisk. The Identity vDisk includes all the specific desktop identity information such as host names and passwords. The Difference vDisk is where all the writes and changes to the desktop are stored. These Identity and Difference vDisks for each desktop are stored on the same data store as their related clone.

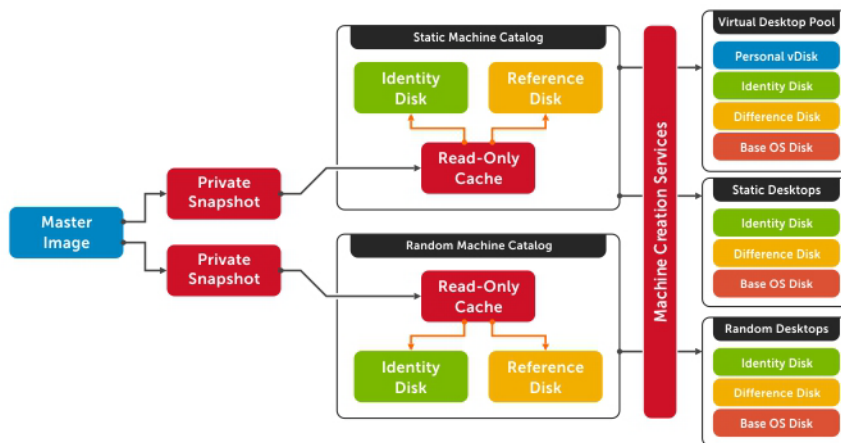


Figure 19 MCS overview diagram

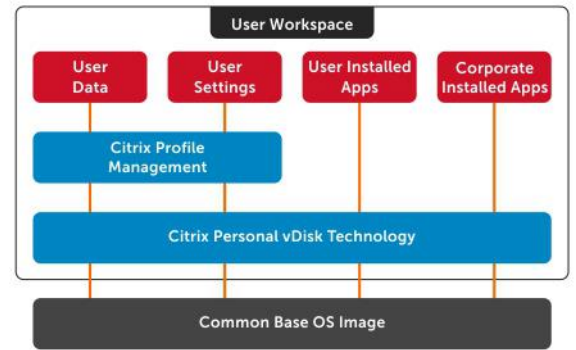
While traditionally used for small to medium sized XenDesktop deployments, MCS can bring along with it some substantial Tier 1 storage cost savings because of the snapshot/identity/difference disk methodology. The Tier 1 disk space requirements of the identity and difference disks when layered on top of a master image snapshot, is far less than that of a dedicated desktop architecture.



### 4.1.3 Personal vDisk

Citrix Personal vDisk is an enterprise workspace virtualization solution that is built into Citrix XenDesktop. Personal vDisk provides the user customization and personalization benefits of a persistent desktop image with the storage savings and performance of a single/shared image.

Used in conjunction with a static desktop experience, Citrix Personal vDisk allows each user to receive personal storage in the form of a layered vDisk (3GB minimum). This personal vDisk enables users to personalize and persist their desktop environment while providing storage for any user or departmental apps.



- Personal vDisk provides the following benefits to XenDesktop:
- Persistent personalization of user profiles, settings and data
- Enables deployment and management of user installed and entitlement based applications
- Fully compatible with Microsoft SCCM and App-V
- 100percent persistence with VDI pooled Storage management
- Near Zero management overhead

### 4.1.4 AppDisks

Citrix AppDisk provides layering technology to manage departmental applications as an independent storage layer. AppDisk eases the management burden of maintaining multiple departmental images by instantly layering applications onto a single golden image that remains separate and pristine. AppDisks can be associated with either published desktops or published applications via XenApp. AppDisk does not replace the functionality provided by Personal vDisk but currently cannot be used within the same golden image either. AppDisks, when integrated with AppDNA, provide the ability to analyze OS and application performance, compatibility, and remediation capabilities.

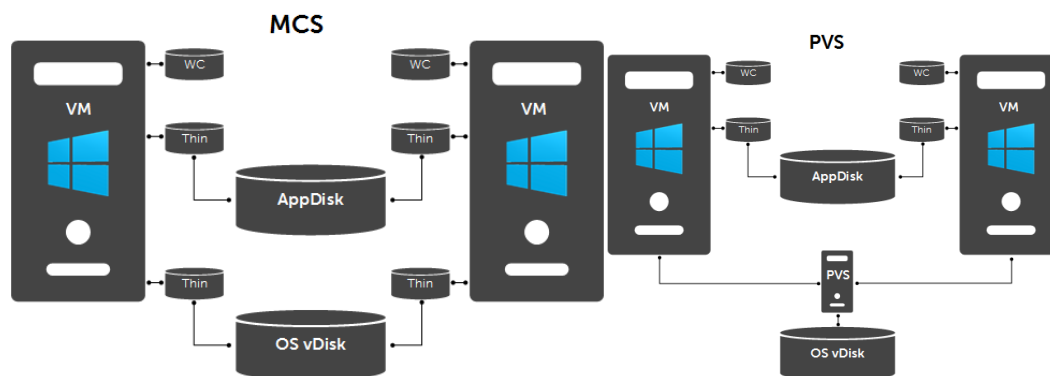


Figure 20 AppDisk overview diagrams

### 4.1.5 HDX 3D Pro

XenDesktop with HDX 3D Pro is a desktop and app virtualization solution that supports high-end designers and engineers of 3D professional graphics applications as well as provides cost-effective support to viewers and editors of 3D data. With XenDesktop, you can deliver a persistent user experience and leverage other virtualization benefits such as single-image management and improved data security.

Use HDX 3D Pro technologies with:

- Computer-aided design, manufacturing, and engineering (CAD/CAM/CAE) applications
- Geographical information system (GIS) software
- Picture Archiving Communication System (PACS) workstations for medical imaging
- Latest OpenGL, DirectX, CUDA and CL versions supported
- Latest NVIDIA Grid cards
- Shared or dedicated GPUs or a mix of both on desktop or server OS VMs

HDX 3D Pro provides the best user experience over any bandwidth using Framehawk integration:

- On wide area network (WAN) connections: Deliver an interactive user experience over WAN connections with bandwidths as low as 1.5 Mbps.
- On local area network (LAN) connections: Deliver a user experience equivalent to that of a local desktop on LAN connections.

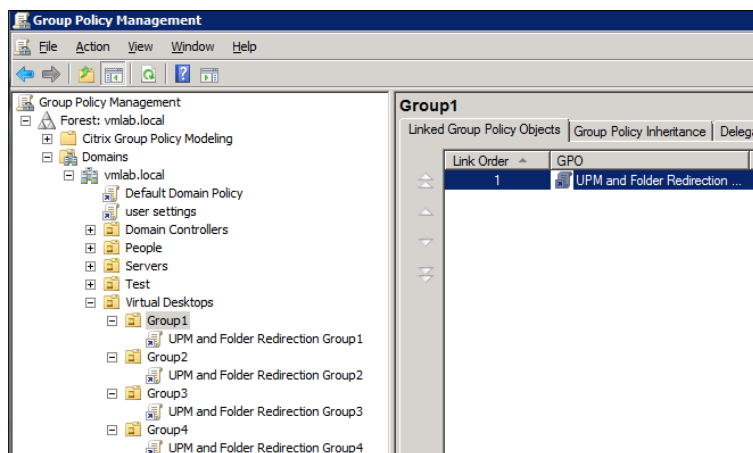
Framehawk is a display remoting technology implemented as an ICA virtual channel that optimizes delivery of virtual desktops and applications to users on broadband wireless connections where high packet loss or congestion occurs.

### 4.1.6 Profile Manager

Citrix Profile Management is a component of the XenDesktop suite which is used to manage user profiles and minimize many of the issues associated with traditional Windows roaming profiles in an environment where users may have their user profile open on multiple devices at the same time. The profile management toolset has two components: the profile management agent, installed on any device where the user profiles is managed, and a Group Policy Administrative Template, which is imported to a group policy.

In order to further optimize, the profile management folders within the user profile is redirected the users' home drive. The folder redirection is managed via group policy objects within Active Directory. The following folders are redirected:

- Contacts
- Downloads
- Favorites
- Links
- My Documents
- Searches
- Start Menu
- Windows
- My Music
- My Pictures
- My Videos
- Desktop



### 4.1.7 XenApp

Citrix XenApp version 7.6 and above includes enhancements in the areas of faster access to virtual apps with higher connection resiliency, improved graphics rendering, and new app-usage reporting and monitoring tools.

Citrix XenApp delivers Windows apps as secure mobile services. With XenApp, IT can mobilize the business - increasing user productivity, while reducing costs by centralizing control and security of intellectual property. XenApp delivers high-performance apps to any PC, Mac, laptop, tablet or smartphone that enable the delivery of a native experience that is optimized for the type of device, as well as the network. XenApp is built on a 3rd generation FlexCast Management Architecture (FMA) and is the only hybrid cloud-ready platform that separates the management plane from the workload to enable IT to securely deliver published apps on-premises, and manage workers and mobile workspaces either on-premises or in the cloud.

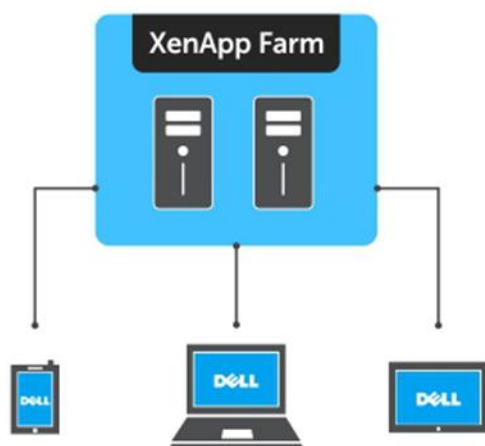


Figure 21 XenApp overview diagram

Benefits of hosted desktop sessions and applications:

- Management of applications (single instance)
- Management of simple desktop images (no applications installed)
- PVS to stream XenApp servers as well as user desktops
- Scalability of XenDesktop compute hosts: CPU and IOPS reduction via application offload
- Shared storage scalability: less IOPS = more room to grow

Citrix XenDesktop with XenApp integration can effectively deliver a desktop/application hybrid solution as well. Specifically where a single or small number of shared VDI desktop images are deployed via XenDesktop, each with common shared applications installed within the golden image. A user-specific application set is then deployed and made accessible via the hosted application compute infrastructure, accessible from within the virtual desktop.

User Environment	XenDesktop	XenApp
User-Specific Applications		✓
Profile and User Data	✓	✓
Shared Applications	✓	
Shared Virtual Desktop Image	✓	

Figure 22 XenDesktop + XenApp features

Alternatively, XenApp provides a platform for delivering Windows server-based sessions to users who may not need a full desktop VM. Hosted desktops increase infrastructure resource utilization while reducing complexity as all applications and sessions are centrally managed.

User Environment	XenDesktop	XenApp
User-Specific Applications		✓
Profile and User Data		✓
Dedicated Virtual Desktop Image		✓

Figure 23 Hosted system XenApp features

#### 4.1.7.1 XenApp integration

The XenApp servers can exist as physical or virtualized instances of Windows Server 2012 R2 or Windows 2016. A minimum of one (1), up to a maximum of ten (10) virtual servers are installed per physical compute host. Since XenApp instances are easily added to an existing XenDesktop stack, the only additional components required are one or more Windows Server OS instances running the Citrix VDA added to the XenDesktop site.

The total number of required virtual XenApp servers is dependent on application type, quantity and user load. Deploying XenApp virtually and in a multi-server farm configuration increases overall farm performance, application load balancing as well as farm redundancy and resiliency.

#### 4.1.7.2 XenDesktop with XenApp and Personal vDisk integration

In a XenDesktop implementation that leverages hosted applications, these execute from a centralized Windows Server and are then accessed via the Citrix Receiver. There are some instances, however, where certain departmental or custom applications cannot run using XenApp. At the same time, for organizational policy or certain storage considerations, delivering these applications as a part of a base image is not possible either. In this case, Citrix Personal vDisk technology is the appropriate solution.

With Citrix Personal vDisk, each user of that single shared virtual desktop image also receives a personal layered vDisk, which enables the user to personalize their desktop and receive native application execution within a Windows client OS and not from a server. When leveraging the integration of XenApp within XenDesktop, all profile and user data is seamlessly accessed within both environments.

User Environment	XenDesktop	XenApp
User-Specific Applications		✓
Profile and User Data	✓	✓
Departmental Applications	✓	
Shared Applications	✓	

Figure 24 XenDesktop + XenApp + Personal vDisk features

#### 4.1.8 NUMA architecture considerations

Best practices and testing has showed that aligning RDSH design to the physical Non-Uniform Memory Access (NUMA) architecture of the server CPUs results in increased and optimal performance. NUMA alignment ensures that a CPU can access its own directly-connected RAM banks faster than those banks of the adjacent processor which are accessed via the Quick Path Interconnect (QPI). The same is true of VMs with large vCPU assignments, best performance will be achieved if your VMs receive their vCPU allotment from a single physical NUMA node. Ensuring that your virtual RDSH servers do not span physical NUMA nodes will ensure the greatest possible performance benefit.

The general guidance for RDSH NUMA-alignment on the Dell EMC XC Series is as follows:

#### 4.1.8.1 A3 NUMA alignment

10 physical cores per CPU in the A3 configuration, 20 logical with Hyper-threading active, gives a total of 40 consumable cores per appliance. The Nutanix CVM will receive its vCPU allotment from the first physical CPU and by configuring the RDSH VMs as shown below will ensure that no NUMA spanning occurs which could lower performance. Per the example below, we have three total RDSH VMs configured with 8 vCPUs each, along with the Nutanix CVM configured with 8 or 10 vCPUs (automatically detects the number of cores on the host and sizes accordingly). Note that the CVM actually reserves 10,000MHz which equates to roughly 5 cores reserved. This leaves both sockets with some additional headroom for workload assignment and burst scheduling. Please note that the A3 and B5 configs have 2 x UPI channels, versus 3 channels on the C7.

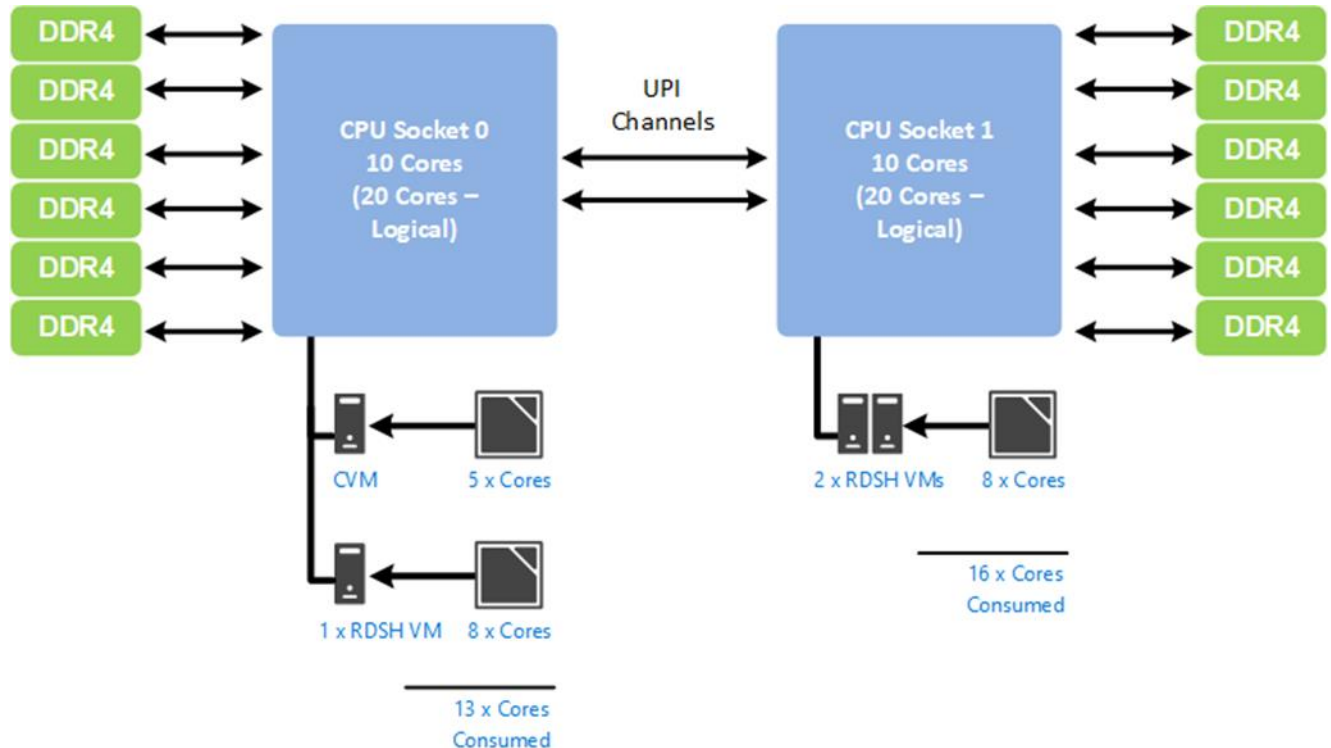


Figure 25 A3 NUMA alignment

#### 4.1.8.2 B5 NUMA alignment

14 physical cores per CPU in the B5 configuration, 28 logical with Hyper-threading active, gives a total of 56 consumable cores per node. The Nutanix CVM will receive its vCPU allotment from the first physical CPU and by configuring the RDSH VMs as shown below will ensure that no NUMA spanning occurs, which could lower performance. Per the example below, we have six RDSH VMs configured with 8 vCPUs each, along with the Nutanix CVM with 12 vCPUs. Note that the CVM actually reserves 10,000MHz which equates to roughly 5 cores reserved. This leaves some additional headroom on socket 0 for workload assignment and burst scheduling. Please note that the A3 and B5 configs have 2 x UPI channels, versus 3 channels on the C7.

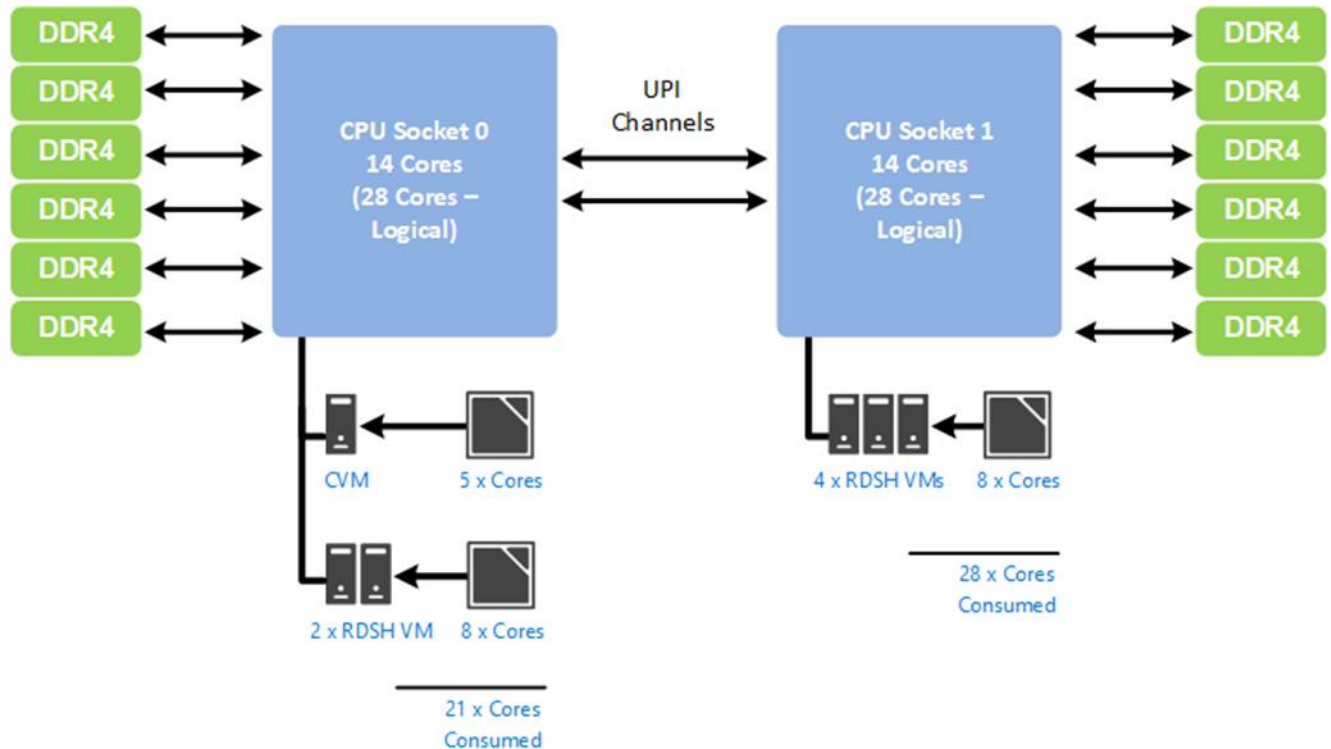


Figure 26 B5 NUMA alignment

#### 4.1.8.3 C7 NUMA alignment

20 physical cores per CPU in the C7 configuration, 40 logical with Hyper-threading active, gives us a total of 80 consumable cores per node. The Nutanix CVM will receive its vCPU allotment from the first physical CPU and by configuring the RDSH VMs as shown below will ensure that no NUMA spanning occurs which could lower performance. Per the example below, we have eight RDSH VMs configured with 8 vCPUs each, along with the Nutanix CVM with 12 vCPUs. Note that the CVM actually reserves 10,000MHz which equates to 5 cores reserved. This leaves some additional headroom on socket 0 for workload assignment and burst scheduling.

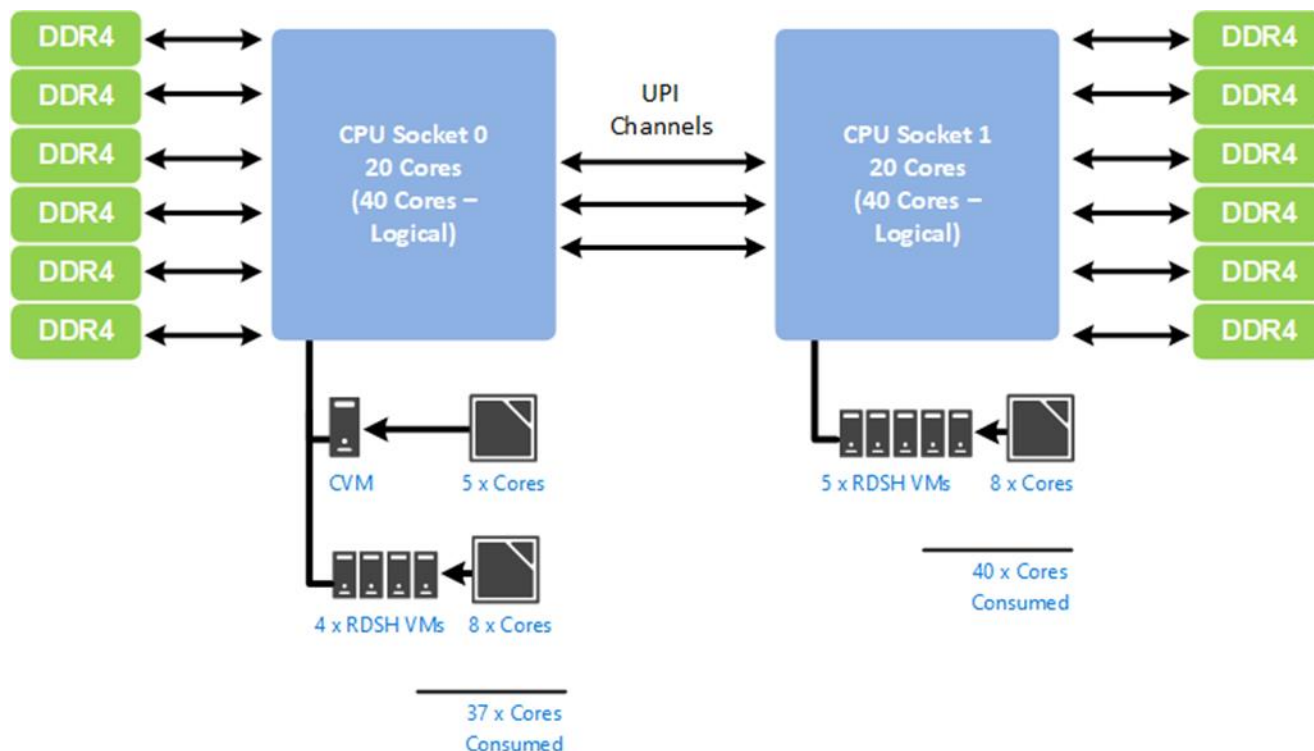


Figure 27 C7 NUMA alignment

#### 4.1.9 Local Host Cache

In XenApp and XenDesktop version 7.12 and above, the Local Host Cache (LHC) feature allows connection brokering operations to continue when connectivity to the Site database has been interrupted. This includes both failures between the Delivery Controller and Site database in on-premises deployments and when the WAN link between the Site and Citrix control plane fails in a Citrix Cloud environment.

LHC replaces the connection leasing feature as the recommended XenApp and XenDesktop high availability solution. During an outage, LHC will support new users and existing users launching new resources, as well as users accessing pooled resources (shared desktops). Earlier versions of XenApp had a feature named Local Host Cache but this is an entirely different implementation that is more robust and immune to corruption.

The principal broker on a delivery controller accepts requests and communicates with the Site database to connect users. A check is made every two minutes to determine if changes have been made to the principal broker's configuration and if so, the information is synchronized with the secondary broker. All configuration data is copied to ensure the LocalDB database matches the site database.



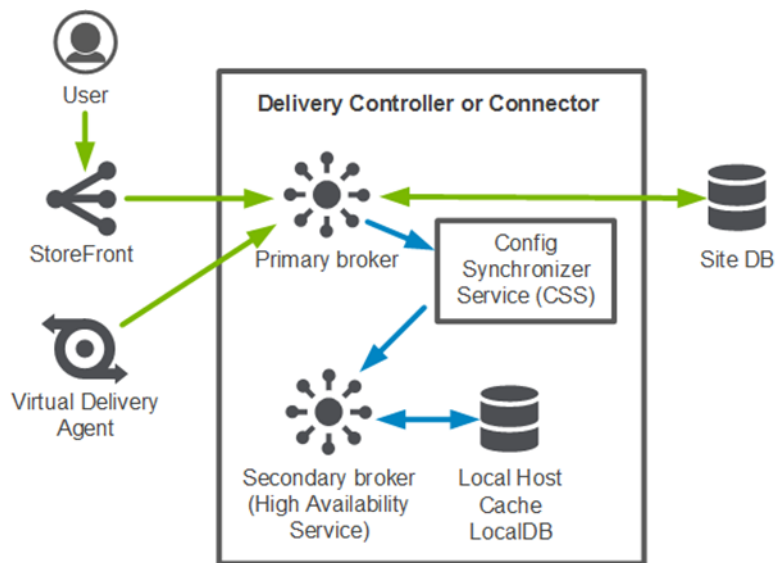


Figure 28 Communication paths during normal operations

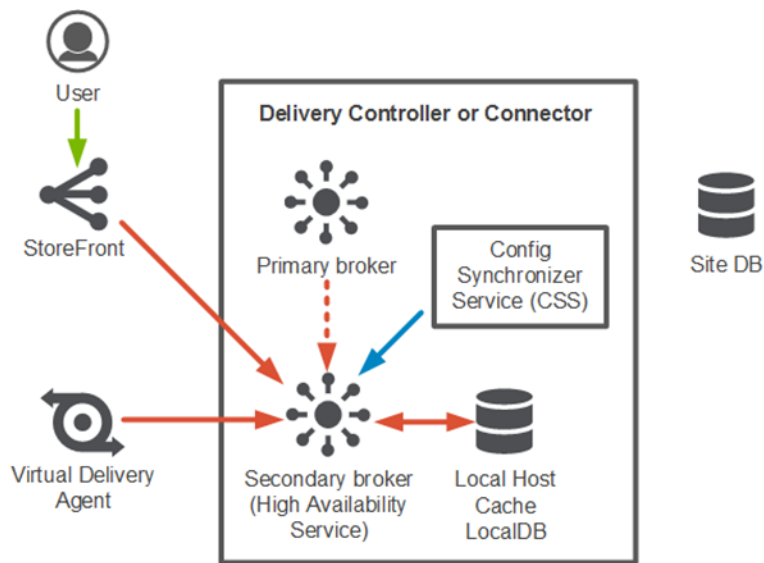


Figure 29 Changes in communication when the principal broker is unable to connect to the Site database

The principal broker stops listening for requests and instructs the secondary broker to begin listening and processing requests. When a VDA communicates with the secondary broker, a re-registration process is triggered during which current session information is delivered. During this time, the principal broker continually monitors the connection to the Site database. Once restored, the principal broker resumes brokering operations and instructs the secondary broker to stop listening for connection information.

#### 4.1.10 NetScaler

Citrix NetScaler is an all-in-one web application delivery controller that makes applications run better, reduces web application ownership costs, optimizes the user experience, and makes sure that applications are always available by using:

- Proven application acceleration such as compression and caching
- High application availability through advanced L4-7 load balancer
- Application security with an integrated application firewall



- Server offloading to significantly reduce costs and consolidate servers

A NetScaler appliance resides between the clients and the servers, so that client requests and server responses pass through it. In a typical installation, virtual servers (vservers) configured on the NetScaler provide connection points that clients use to access the applications behind the NetScaler. In this case, the NetScaler owns public IP addresses that are associated with its vservers, while the real servers are isolated in a private network. It is also possible to operate the NetScaler in a transparent mode as an L2 bridge or L3 router, or even to combine aspects of these and other modes. NetScaler can also be used to host the StoreFront function eliminating complexity from the environment.

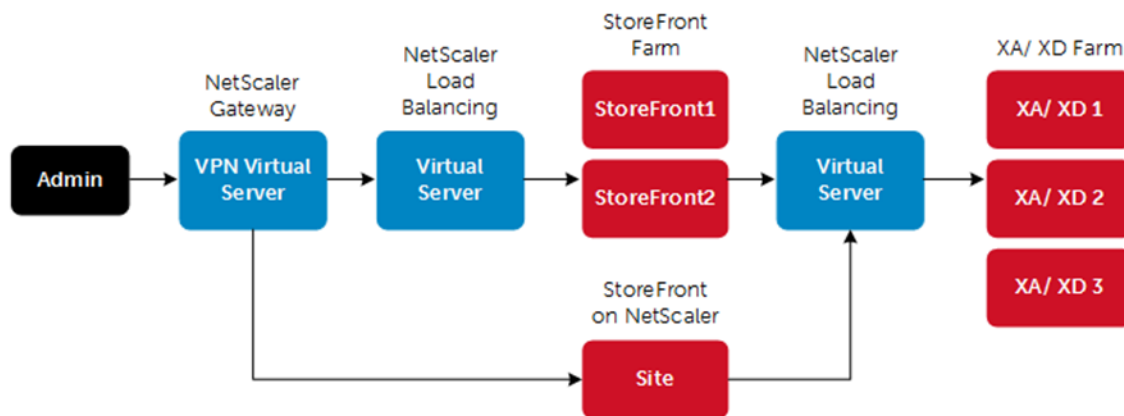


Figure 30 Citrix NetScaler overview diagram

### Global Server Load Balancing

GSLB is an industry standard function. It is in widespread use to provide automatic distribution of user requests to an instance of an application hosted in the appropriate data center where multiple processing facilities exist. The intent is to seamlessly redistribute load on an as required basis, transparent to the user community. These distributions are used on a localized or worldwide basis. Many companies use GSLB in its simplest form.

They use the technology to automatically redirect traffic to Disaster Recovery (DR) sites on an exception basis. That is, GSLB is configured to simply route user load to the DR site on a temporary basis only in the event of a catastrophic failure or only during extended planned data center maintenance. GSLB is also used to distribute load across data centers on a continuous load balancing basis as part of normal processing.

### NetScaler and XenDesktop Deployment Guide: [Link](#)

Several of the management components of the XenDesktop stack are made highly-available using NetScaler to load balance traffic. The following management components require the use of a load balancer to function in a high availability mode:

- StoreFront Servers
- Licensing Server
- XenDesktop XML Service
- XenDesktop Desktop Director
- Provisioning Services TFTP Service
- Framehawk UDP virtual channel (supported on NetScaler Gateway 11.0.62.10 or later and NetScaler Unified Gateway 11.0.64.34 or later)

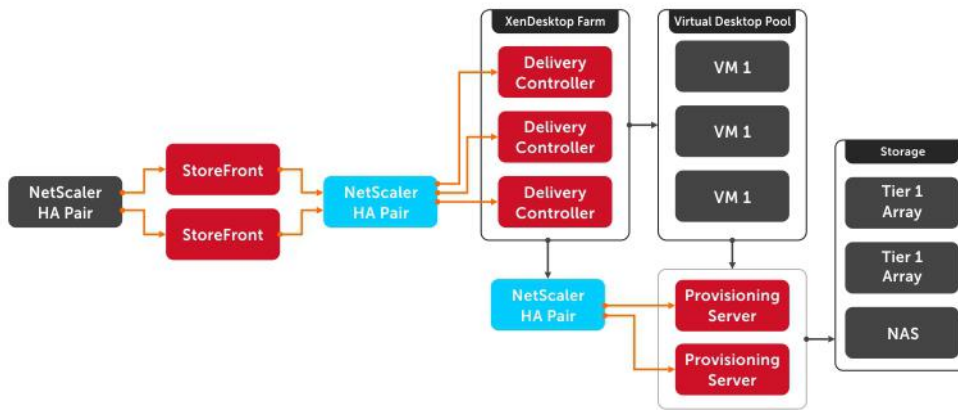


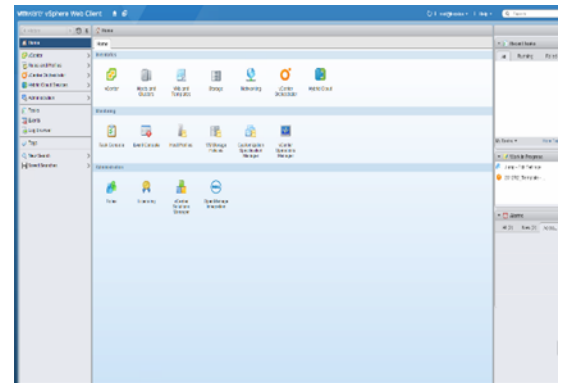
Figure 31 Citrix NetScaler in high availability mode

## 4.1.11 Hypervisor platforms

### 4.1.11.1 VMware vSphere 6

The vSphere hypervisor also known as ESXi is a bare-metal hypervisor that installs directly on top of your physical server and partitions it into multiple virtual machines. Each virtual machine shares the same physical resources as the other virtual machines and they can all run at the same time. Unlike other hypervisors, all management functionality of vSphere is done through remote management tools. There is no underlying operating system, reducing the install footprint to less than 150MB.

VMware vSphere 6 includes three major layers: Virtualization, Management and Interface. The Virtualization layer includes infrastructure and application services. The Management layer is central for configuring, provisioning and managing virtualized environments. The Interface layer includes the vSphere web client.

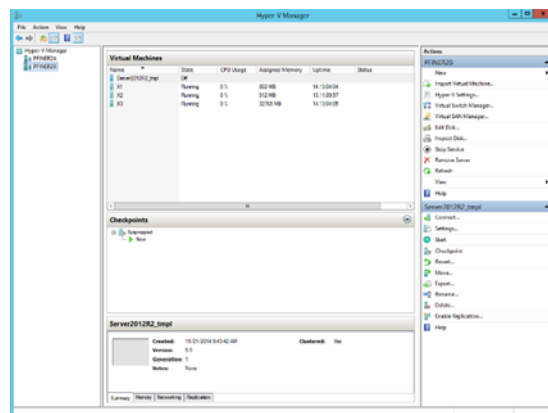


Throughout this Dell EMC solution, all VMware and Microsoft best practices and prerequisites for core services are adhered to (NTP, DNS, Active Directory, etc.). The vCenter 6 VM used in the solution is a single Windows Server 2012 R2 VM or vCenter 6 virtual appliance, residing on a host in the management layer. SQL server is a core component of the Windows version of vCenter and is hosted on another VM also residing in the management layer. It is recommended that all additional XenDesktop components be installed in a distributed architecture, one role per server VM.

### 4.1.11.2 Microsoft Windows Server 2012 R2 Hyper-V

Windows Server 2012 R2 Hyper-V™ is a powerful virtualization technology that enables businesses to leverage the benefits of virtualization. Hyper-V reduces costs, increases hardware utilization, optimizes business infrastructure, and improves server availability. Hyper-V works with virtualization-aware hardware to tightly control the resources available to each virtual machine. The latest generation of Dell EMC servers includes virtualization-aware processors and network adapters.

From a network management standpoint, virtual machines are much easier to manage than physical computers. To this end, Hyper-V includes many management features designed to make managing virtual machines simple and familiar, while enabling easy access to powerful VM-specific management functions. The primary management platform within a Hyper-V based XenDesktop virtualization environment is Microsoft Systems Center Virtual Machine Manager SP1 (SCVMM).

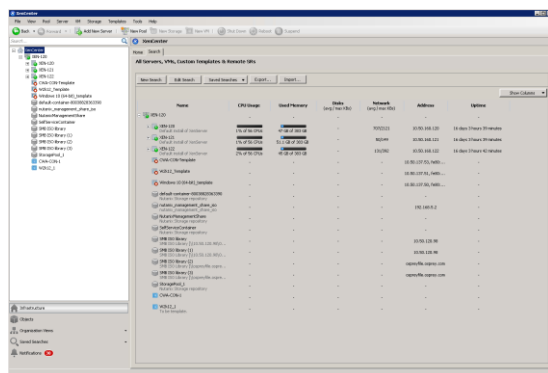


SCVMM provides centralized and powerful management, monitoring, and self-service provisioning for virtual machines. SCVMM host groups are a way to apply policies and to check for problems across several VMs at once. Groups are organized by owner, operating system, or by custom names such as “Development” or “Production”. The interface also incorporates Remote Desktop Protocol (RDP); double-click a VM to bring up the console for that VM—live and accessible from the management console.

### 4.1.11.3 Citrix XenServer 7

XenServer is a comprehensive server virtualization platform with enterprise-class features built in to easily handle different workload types, mixed operating systems and storage or networking configurations. For the most demanding app, and desktop virtualization use cases, its industry-leading scalability and performance under load, can cater to the largest XenApp or XenDesktop deployments.

XenServer is managed by XenCenter to provide VM monitoring, management, and general administration functions through a single, intuitive interface.



XenServer combined with Dell EMC XC Core provides simple One-Click Cloud Integration for deploying the Citrix Workspace Appliance to eliminate complexity and provide built-in automation and orchestration.

As stated above, in the previous paragraph to take advantage of this virtualization platform, Dell EMC XC Series systems should be ordered leveraging the XC Core licensing model.

---

**Note:** Full performance and testing characterization has not been completed on this virtualization platform at this time. User densities and characterization will be available in a subsequent release of this reference architecture.

---

## 4.2 NVIDIA GRID vGPU

NVIDIA GRID™ vGPU™ brings the full benefit of NVIDIA hardware-accelerated graphics to virtualized solutions. This technology provides exceptional graphics performance for virtual desktops equivalent to local PCs when sharing a GPU among multiple users.

GRID vGPU is the industry's most advanced technology for sharing true GPU hardware acceleration between multiple virtual desktops—without compromising the graphics experience. Application features and compatibility are exactly the same as they would be at the user's desk.

With GRID vGPU technology, the graphics commands of each virtual machine are passed directly to the GPU, without translation by the hypervisor. This allows the GPU hardware to be time-sliced to deliver outstanding shared virtualized graphics performance.

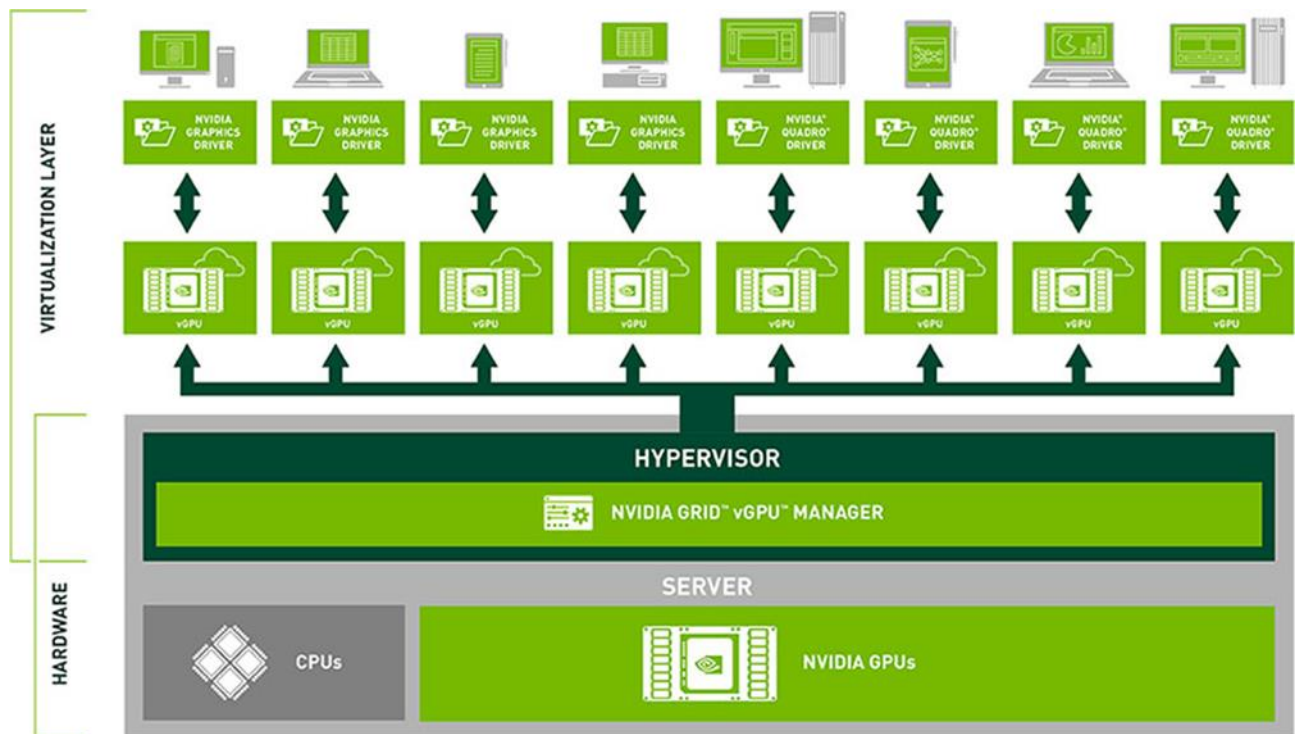


Figure 32 NVIDIA GRID vGPU overview diagram. Image provided courtesy of NVIDIA Corporation, Copyright NVIDIA Corporation.

### 4.2.1 vGPU profiles

Virtual Graphics Processing Unit, or GRID vGPU™, is technology developed by NVIDIA® that enables hardware sharing of graphics processing for virtual desktops. This solution provides a hybrid shared mode allowing the GPU to be virtualized while the virtual machines run the native NVIDIA video drivers for better performance. Thanks to OpenGL support, VMs have access to more graphics applications. When utilizing vGPU, the graphics commands from virtual machines are passed directly to the GPU without any hypervisor translation. Every virtual desktop has dedicated graphics memory so they always have the resources they need to launch and run their applications at full performance. All this is done without sacrificing server performance and so is truly cutting edge.

The combination of Dell EMC servers, NVIDIA GRID vGPU™ technology and NVIDIA Tesla™ cards enable high-end graphics users to experience high fidelity graphics quality and performance, for their favorite applications at a reasonable cost.

For more information about NVIDIA GRID vGPU, please visit: [Link](#)

The number of users per appliance is determined by the number of GPU cards in the system (max 2 x M10 or 3 x M60), vGPU profiles used for each GPU in a card, and GRID license type. The same profile must be used on a single GPU but profiles can differ across GPUs within a single card.

Table 9 NVIDIA® Tesla® M10 GRID vGPU Profiles:

Card	vGPU Profile	Graphics Memory (Frame Buffer)	Virtual Display Heads	Maximum Resolution	Maximum Graphics-Enabled VMs		
					Per GPU	Per Card	Per Server (2 cards)
Tesla M10	M10-8Q	8GB	4	4096x2160	1	4	8
	M10-4Q	4GB	4	4096x2160	2	8	16
	M10-2Q	2GB	4	4096x2160	4	16	32
	M10-1Q	1GB	2	4096x2160	8	32	64
	M10-0Q	512MB	2	2560x1600	16	64	128
	M10-1B	1GB	4	2560x1600	8	32	64
	M10-0B	512MB	2	2560x1600	16	64	128
	M10-8A	8GB	1	1280x1024	1	4	8
	M10-4A	4GB			2	8	16
	M10-2A	2GB			4	16	32
	M10-1A	1GB			8	32	64

Card	vGPU Profile	Guest VM OS Supported*		License Required
		Win	64bit Linux	
Tesla M10	M10-8Q	•	•	NVIDIA® Quadro® Virtual Data Center Workstation
	M10-4Q	•	•	
	M10-2Q	•	•	
	M10-1Q	•	•	
	M10-0Q	•	•	
	M10-1B	•		GRID Virtual PC
	M10-0B	•		
	M10-8A	•		GRID Virtual Application
	M10-4A	•		
	M10-2A	•		
	M10-1A	•		

Supported Guest VM Operating Systems*	
Windows	Linux
Windows 7 (32/64-bit)	RHEL 6.6 & 7
Windows 8.x (32/64-bit)	CentOS 6.6 & 7
Windows 10 (32/64-bit)	Ubuntu 12.04 & 14.04 LTS
Windows Server 2008 R2	
Windows Server 2012 R2	
Windows Server 2016	

---

\*NOTE: Supported guest operating systems listed as of the time of this writing. Please refer to NVIDIA's documentation for latest supported operating systems.

---

Table 10 NVIDIA® Tesla® M60 GRID vGPU Profiles:

Card	vGPU Profile	Graphics Memory (Frame Buffer)	Virtual Display Heads	Maximum Resolution	Maximum Graphics-Enabled VMs		
					Per GPU	Per Card	Per Server (3 cards)
<b>Tesla M60</b>	M60-8Q	8GB	4	4096x2160	1	2	6
	M60-4Q	4GB	4	4096x2160	2	4	12
	M60-2Q	2GB	4	4096x2160	4	8	24
	M60-1Q	1GB	2	4096x2160	8	16	48
	M60-0Q	512MB	2	2560x1600	16	32	96
	M60-1B	1GB	4	2560x1600	8	16	48
	M60-0B	512MB	2	2560x1600	16	32	96
	M60-8A	8GB	1	1280x1024	1	2	6
	M60-4A	4GB			2	4	12
	M60-2A	2GB			4	8	24
	M60-1A	1GB			8	16	48



Card	vGPU Profile	Guest VM OS Supported*		License Required
		Win	64bit Linux	
Tesla M60	M60-8Q	•	•	NVIDIA® Quadro® Virtual Data Center Workstation
	M60-4Q	•	•	
	M60-2Q	•	•	
	M60-1Q	•	•	
	M60-0Q	•	•	
	M60-1B	•		GRID Virtual PC
	M60-0B	•		
	M60-8A	•		GRID Virtual Application
	M60-4A	•		
	M60-2A	•		
	M60-1A	•		

Supported Guest VM Operating Systems*	
Windows	Linux
Windows 7 (32/64-bit)	RHEL 6.6 & 7
Windows 8.x (32/64-bit)	CentOS 6.6 & 7
Windows 10 (32/64-bit)	Ubuntu 12.04 & 14.04 LTS
Windows Server 2008 R2	
Windows Server 2012 R2	
Windows Server 2016	

\*NOTE: Supported guest operating systems listed as of the time of this writing. Please refer to NVIDIA's documentation for latest supported operating systems.

#### 4.2.1.1 GRID vGPU licensing and architecture

NVIDIA® GRID vGPU™ is offered as a licensable feature on Tesla® GPUs. vGPU can be licensed and entitled using one of the three following software editions.

Table 11 NVIDIA vGPU licensing editions



<b>NVIDIA® GRID® Virtual Applications</b>	<b>NVIDIA® GRID® Virtual PC</b>	<b>NVIDIA® Quadro® Virtual Data Center Workstation</b>
For organizations deploying or other RDSH solutions. Designed to deliver Windows applications at full performance.	For users who want a virtual desktop, but also need a great user experience leveraging PC applications, browsers, and high-definition video.	For users who need to use professional graphics applications with full performance on any device, anywhere.
Up to 2 displays @ 1280x1024 resolution supporting virtualized Windows applications	Up to 4 displays @ 2560x1600 resolution supporting Windows desktops, and NVIDIA Quadro features	Up to 4 displays @ 4096x2160* resolution supporting Windows or Linux desktops, NVIDIA Quadro, CUDA**, OpenCL** & GPU pass-through

\*0Q profiles only support up to 2560x1600 resolution

\*\*CUDA and OpenCL only supported with M10-8Q, M10-8A, M60-8Q, or M60-8A profiles

The GRID vGPU Manager, running on the hypervisor installed via the VIB, controls the vGPUs that can be assigned to guest VMs. A properly configured VM obtains a license from the GRID license server during the boot operation for a specified license level. The NVIDIA graphics driver running on the guest VM provides direct access to the assigned GPU. When the VM is shut down, it releases the license back to the server. If a vGPU enabled VM is unable to obtain a license, it will run at full capability without the license but users will be warned each time it tries and fails to obtain a license.

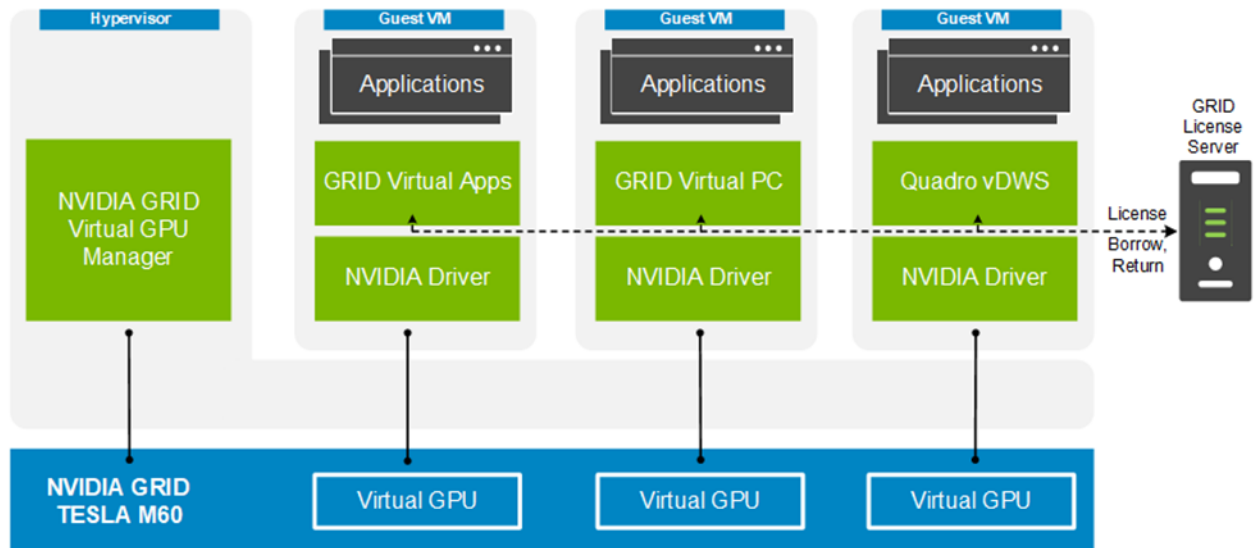


Figure 33 NVIDIA GRID licensing architecture overview

## 4.3 Citrix Ready Workspace Appliance

The Citrix Ready Workspace Appliance enables customers to leverage the flexibility of a hybrid cloud architecture to enable additional cost and resource consumption models. With Dell EMC XC Core and Citrix XenServer, organizations can quickly deploy new appliances and quickly connect them to the cloud to address organizational growth while managing these appliances in a central location. This solution provides a fast and flexible way to provide secure access to virtual Windows, Linux, and applications.

The overall solution consists of two main components:

1. A hosted Citrix Cloud XenApp/XenDesktop Service (running in the Citrix Cloud)
2. A Dell EMC XC Core cluster (hosted in an on-premises resource location)

In addition to the above components, XC Core licensing and Citrix XenServer must be used to enable and take advantage of this solution and features.

To manage and monitor on the on-premises Dell EMC XC Series cluster, the hosted Citrix Cloud XenApp/XenDesktop Service must deploy a Cloud Connector at the on-premise side to act as a proxy and agent for communication between the two components. Nutanix has automated the deployment of this connector within the Prism management interface with a feature called "Connect to Citrix Cloud". This feature provides single-click integration of the on-premise Dell EMC XC Series clusters as a resource location within the Citrix Cloud environment.

For additional information, please visit: [Link](#)

## 5 Solution architecture for XenDesktop

### 5.1 Management role configuration

The Management role recommendations for the base solution are summarized below. Use data disks for role-specific application files such as data, logs and IIS web files in the Management volume.

#### 5.1.1 Citrix XenDesktop on vSphere management role requirements

Table 12 Citrix XenDesktop on vSphere management role requirements

Role	vCPU	vRAM (GB)	NIC	OS vDisk	
				Size (GB)	Location
Nutanix CVM	12*	32	2	-	BOSS
DDC + Lic	4	8	1	40	DSF: ds_mgmt
Storefront	2	4	1	40	DSF: ds_mgmt
Primary SQL	4	8	1	250	DSF: ds_mgmt
VCSA	2	8	1	125	DSF: ds_mgmt
<b>Total</b>	<b>24</b>	<b>44</b>	<b>6</b>	<b>455</b>	-

## 5.1.2 Citrix XenDesktop on Hyper-V management role requirements

Table 13 Citrix XenDesktop on Hyper-V management role requirements

Role	vCPU	Startup RAM (GB)	Dynamic Memory			vNIC	OS vDisk	
			Min Max	Buffer	Weight		Size (GB)	Location
Nutanix CVM	12*	32	Dynamic Memory Disabled		High	2	-	BOSS
DDC + Lic	4	8	512MB   10GB	20%	Med	1	40	DSF: ds_mgmt
Storefront	2	4	512MB   6GB	20%	Med	1	40	DSF: ds_mgmt
Primary SQL	4	8	512MB   10GB	20%	Med	1	40 + 200	DSF: ds_mgmt
SCVMM	2	8	512MB   10GB	20%	Med	1	40 + 50	DSF: ds_mgmt
<b>Total</b>	<b>20</b>	<b>60</b>	<b>2GB   36GB</b>	-	-	<b>6</b>	<b>445</b>	-

\*NOTE: The Nutanix CVM utilizes physical core reservations that may vary based on platform configuration.

## 5.1.3 RDSH/XenApp on vSphere

The recommended number of RDSH VMs and their configurations on vSphere are summarized below based on applicable hardware platform.

Table 14 RDSH/XenApp on vSphere VM configurations

Role	HW Config	VMs per host	vCPUs per VM	RAM (GB)	vNIC	OS vDisk	
						Size (GB)	Location
RDSH VM	A3	3	8	32	1	80	DSF: ds_rdsh
RDSH VM	B5	6	8	32	1	80	DSF: ds_rdsh
RDSH VM	C7	8	8	32	1	80	DSF: ds_rdsh

## 5.1.4 RDSH/XenApp on Hyper-V

Table 15 RDSH/XenApp on Hyper-V VM configurations

Role	HW Config	VMs per Host	vCPUs per VM	Startup RAM (GB)	Dynamic Memory			vNIC	OS vDisk	
					Min Max	Buffer	Weight		Size (GB)	Location
RDSH VM	A3	3	8	16	8GB   32GB	20%	Med	1	80	DSF: ds_rdsh
RDSH VM	B5	6	8	16	8GB   32GB	20%	Med	1	80	DSF: ds_rdsh
RDSH VM	C7	8	8	16	8GB   32GB	20%	Med	1	80	DSF: ds_rdsh

## 5.1.5 NVIDIA GRID license server requirements

When using NVIDIA Tesla cards, graphics enabled VMs must obtain a license from a GRID License server on your network to be entitled for vGPU. To configure, a virtual machine with the following specifications must be added to a management host in addition to the management role VMs.

Role	vCPU	vRAM (GB)	NIC	OS vDisk	
				Size (GB)	Location
NVIDIA GRID License Srv	2	4	1	40 + 5	DSF: ds_mgmt

GRID License server software can be installed on a system running the following operating systems:

- Windows 7 (x32/x64)
- Windows 8.x (x32/x64)
- Windows 10 x64
- Windows Server 2008 R2
- Windows Server 2012 R2
- Red Hat Enterprise 7.1 x64
- CentOS 7.1 x64

Additional license server requirements:

- A fixed (unchanging) IP address. The IP address may be assigned dynamically via DHCP or statically configured, but must be constant.
- At least one unchanging Ethernet MAC address, to be used as a unique identifier when registering the server and generating licenses in NVIDIA's licensing portal.
- The date/time must be set accurately (all hosts on the same network should be time synchronized).

## 5.1.6 SQL databases

The Citrix, Microsoft and VMware databases are hosted by a single dedicated SQL 2012 R2 Server VM in the Management layer. Use caution during database setup to ensure that SQL data, logs, and TempDB are properly separated onto their respective volumes. Create all Databases that are required for:

- Citrix XenDesktop
- vCenter (if using Windows version) or SCVMM

Initial placement of all databases into a single SQL instance is fine unless performance becomes an issue, in which case database need to be separated into separate named instances. Enable auto-growth for each DB.

Best practices defined by Citrix, Microsoft and VMware are to be adhered to, to ensure optimal database performance.

Align all disks to be used by SQL Server with a 1024K offset and then formatted with a 64K file allocation unit size (data, logs, and TempDB).

## 5.1.7 DNS

DNS plays a crucial role in the environment not only as the basis for Active Directory but is used to control access to the various Citrix and Microsoft software components. All hosts, VMs, and consumable software components need to have a presence in DNS, preferably via a dynamic and AD-integrated namespace. Microsoft best practices and organizational requirements are to be adhered to.

Pay consideration for eventual scaling, access to components that may live on one or more servers (SQL databases, Citrix services) during the initial deployment. Use CNAMEs and the round robin DNS mechanism to provide a front-end “mask” to the back-end server actually hosting the service or data source.

### 5.1.7.1 DNS for SQL

To access the SQL data sources, either directly or via ODBC, a connection to the server name\instance name must be used. To simplify this process, as well as protect for future scaling (HA), instead of connecting to server names directly, alias these connections in the form of DNS CNAMEs. So instead of connecting to SQLServer1\<instance name> for every device that needs access to SQL, the preferred approach is to connect to <CNAME>\<instance name>.

For example, the CNAME “VDISQL” is created to point to SQLServer1. If a failure scenario was to occur and SQLServer2 would need to start serving data, we would simply change the CNAME in DNS to point to SQLServer2. No infrastructure SQL client connections would need to be touched.




 SQLServer1	Host (A)	10.1.1.28
 SQLServer2	Host (A)	10.1.1.29
 SQLVDI	Alias (CNAME)	SQLServer1.fcs.local

Figure 34 SQL DNS CNAMEs

## 5.2 Storage architecture overview

All Dell EMC XC Series appliances come with two tiers of storage by default, SSD for performance and HDD for capacity. Additionally, all-flash configurations are available utilizing only SSD disks. A single common Software Defined Storage namespace is created across the Nutanix cluster and presented as either NFS or SMB to the hypervisor of each host. This constitutes a storage pool and one should be sufficient per cluster. Within this common namespace, logical containers are created to group VM files as well as control the specific storage-related features that are desired to be enabled such as deduplication and compression.

### 5.2.1 Nutanix containers

The following table outlines the recommended containers, their purpose and settings given the use case. Best practices suggest using as few features as possible, only enable what is absolutely required. For example, if you are not experiencing disk capacity pressure then there is no need to enable Capacity Tier Deduplication. Enabling unnecessary services increases the resource demands of the Controller VMs. Capacity tier deduplication requires that CVMs be configured with 32GB RAM. Erasure Coding (EC-X) is recommended to increase usable capacity of the cluster.

Table 16 Nutanix container details

Container	Purpose	Replication Factor	EC-X	Perf Tier Deduplication	Capacity Tier Deduplication	Compression
Ds_compute	Desktop VMs	2	Enabled	Enabled	Disabled	Disabled
Ds_mgmt	Mgmt Infra VMs	2	Enabled	Enabled	Disabled	Disabled
Ds_rdsh	RDSH VMs	2	Enabled	Enabled	Disabled	Disabled
Ds_vgpu	vGPU VMs	2	Enabled	Enabled	Disabled	Disabled

## 5.3 Virtual networking

The network configuration for the Dell EMC XC Series appliances utilizes a 10Gb converged infrastructure model. All required VLANs will traverse 2 x 10Gb NICs configured in an active/active team. For larger scaling it is recommended to separate the infrastructure management VMs from the compute VMs to aid in predictable compute host scaling. The following outlines the suggested VLAN requirements for the Compute and Management hosts in this solution model:

- Compute hosts
  - Management VLAN: Configured for hypervisor infrastructure traffic – L3 routed via spine layer
  - Live Migration VLAN: Configured for Live Migration traffic – L2 switched via leaf layer
  - VDI VLAN: Configured for VDI session traffic – L3 routed via spine layer
- Management hosts
  - Management VLAN: Configured for hypervisor Management traffic – L3 routed via spine layer
  - Live Migration VLAN: Configured for Live Migration traffic – L2 switched via leaf layer
  - VDI Management VLAN: Configured for VDI infrastructure traffic – L3 routed via spine layer
- An iDRAC VLAN is configured for all hardware management traffic – L3 routed via spine layer



### 5.3.1 vSphere

Both the compute and management host network configuration consists of a standard vSwitch teamed with 2 x 10Gb physical adapters assigned to VMNICS. The CVM connects to a private internal vSwitch to communicate directly with the hypervisor as well as the standard external vSwitch to communicate with other CVMs in the cluster. All VDI infrastructure VMs connect through the primary port group on the external vSwitch.

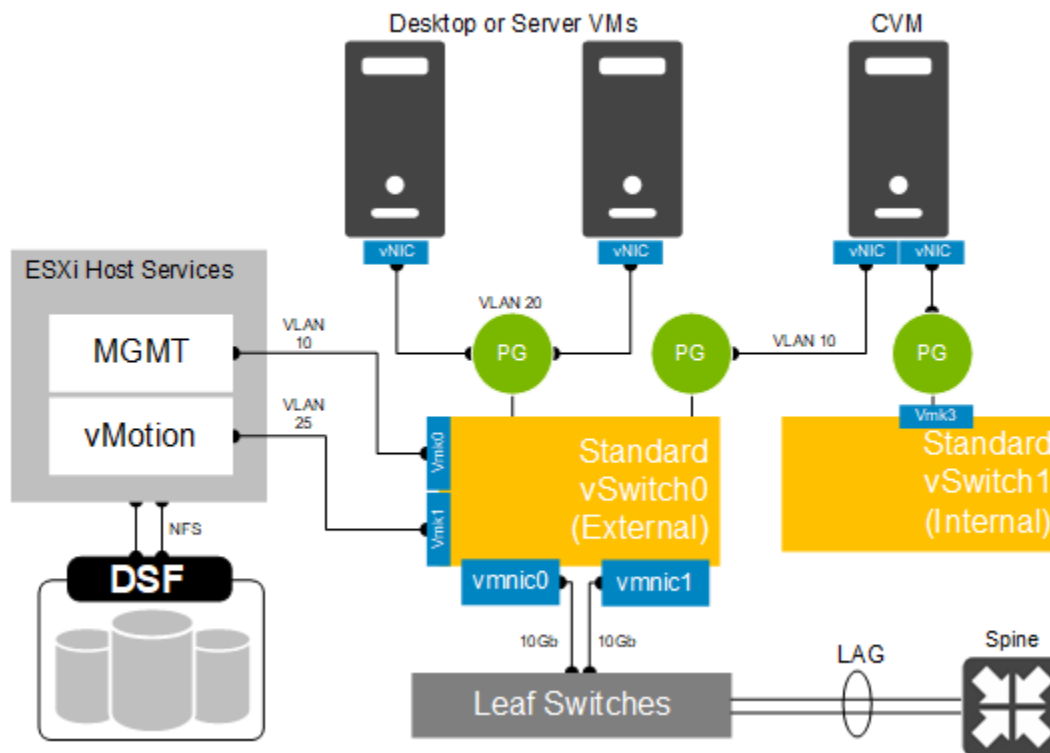


Figure 35 vSphere virtual networking overview

### 5.3.2 Hyper-V

The Hyper-V compute and management configuration, while identical in core requirements and hardware, is executed differently due to how Hyper-V and Windows Server 2012 R2 implement networking and virtual switches. As shown in the diagram below, native Windows Server 2012 R2 NIC Teaming is utilized to load balance and provide resiliency for network connections. For the management host in this scenario, a single LBFO NIC team is configured to connect to a Hyper-V switch for external traffic and one internal Hyper-V switch is used for the Nutanix CVM alone. All vNICs associated with the Management OS connect directly to the external Hyper-V switch.

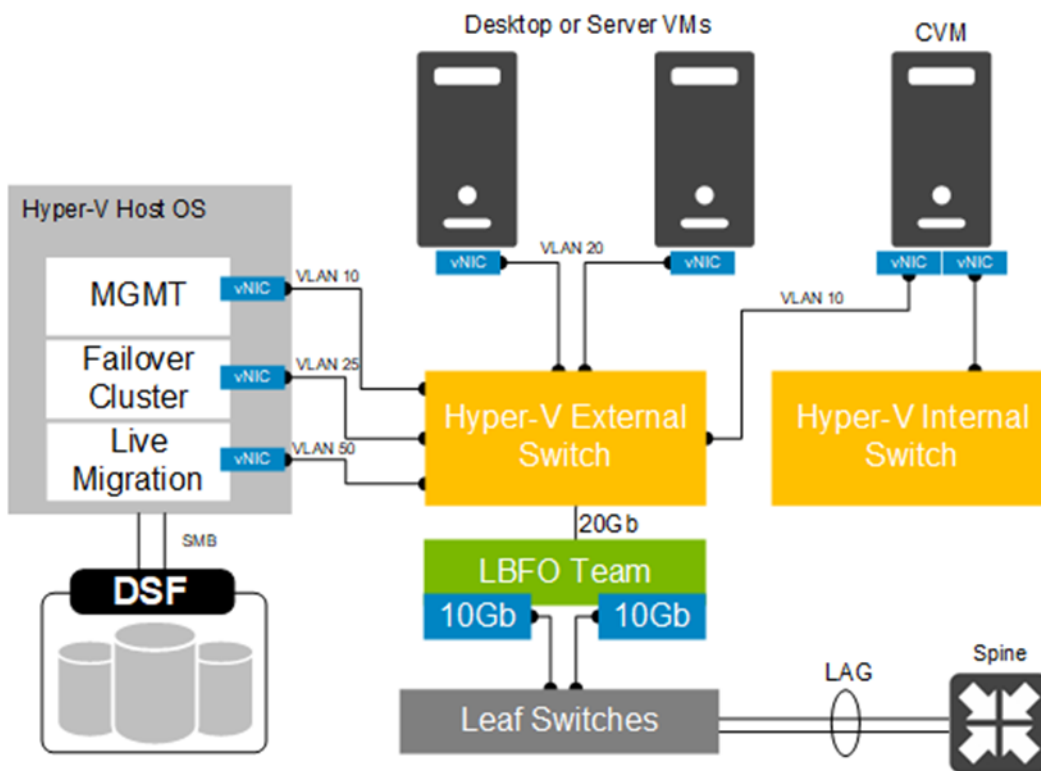


Figure 36 Hyper-V virtual networking overview

The NIC team for the Hyper-V switch is configured as switch independent, dynamic for the load balancing mode with all adapters set to active. This team is used exclusively by Hyper-V.

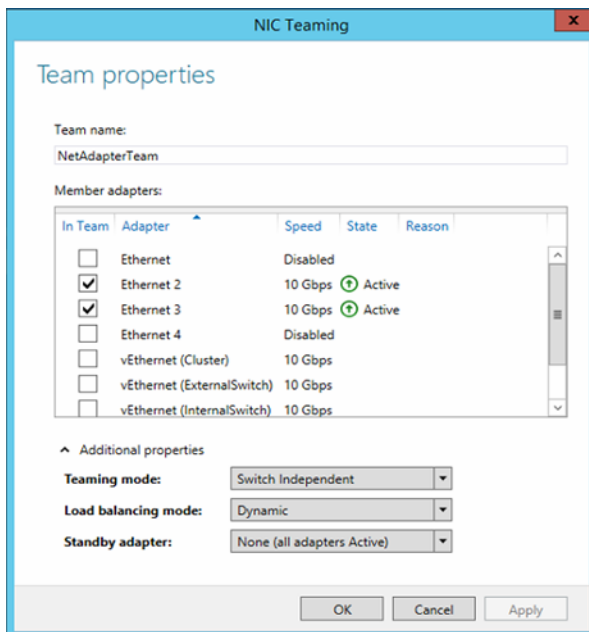


Figure 37 Hyper-V NIC teaming

## 5.4 Scaling guidance

Each component of the solution architecture scales independently according to the desired number of supported users. Additional appliance nodes can be added at any time to expand the Nutanix SDS pool in a modular fashion. While there is no scaling limit of the Nutanix architecture itself, practicality might suggest scaling pods based on the limits of hypervisor clusters (64 nodes for vSphere or Hyper-V). Isolating management and compute to their own HA clusters provides more flexibility with regard to scaling and functional layer protection while stretching the DSF cluster namespace between them.

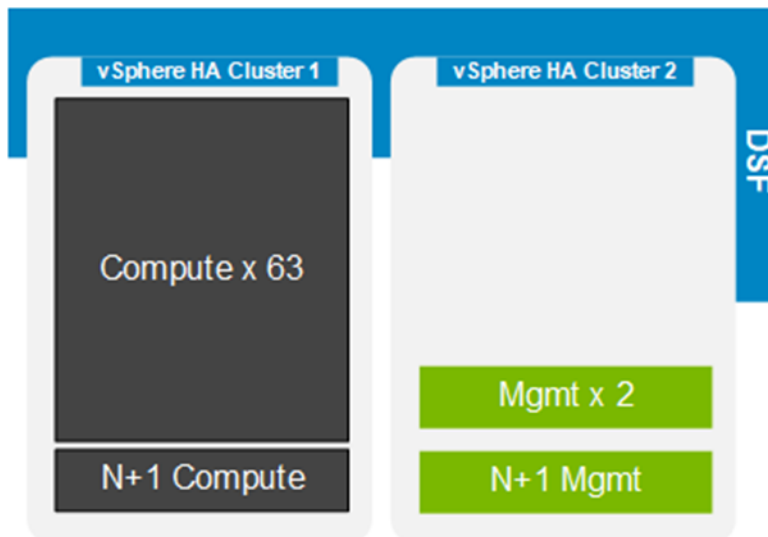


Figure 38 Scaling guidance

Another option is to design a large single contiguous NDFS namespace with multiple hypervisor clusters within to provide single pane of glass management. For example, portrayed below is a large-scale user environment segmented by vSphere HA cluster and broker farm. Each farm compute instance is segmented into an HA cluster with a hot standby node providing N+1, served by a dedicated pair of management nodes per compute cluster in a separate HA cluster dedicated to mgmt. This provides multiple broker farms with separated HA protection while maintaining a single NDFS cluster across all nodes. Some hypervisor specific features like Cluster Aware Updating (CUA) in Hyper-V might not work as expected using this architecture. If CAU is to be implemented, the recommended guidance is to deploy a single CAU for the DSF cluster at a time.

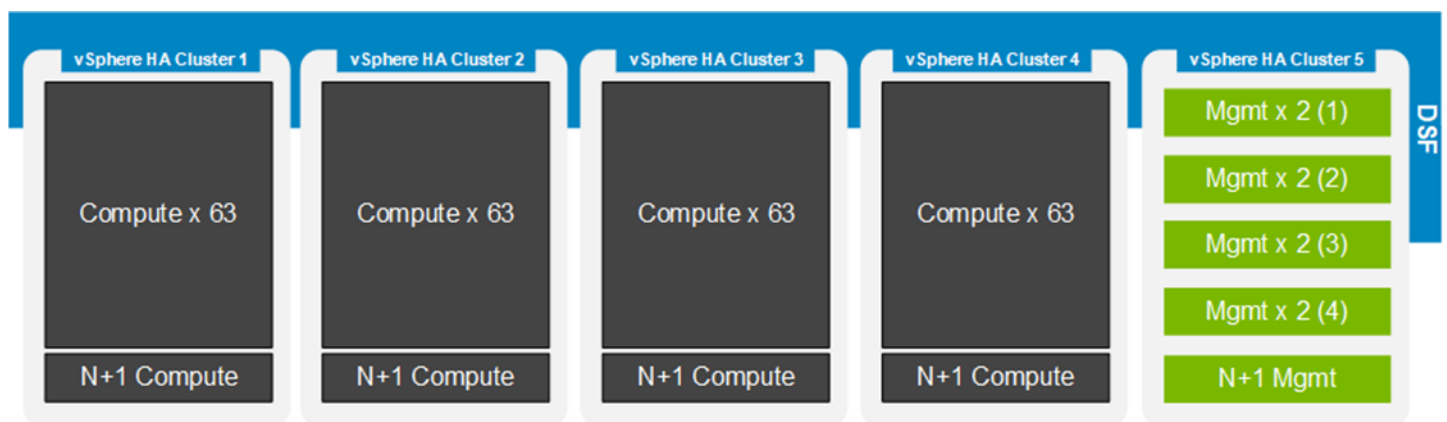


Figure 39 Additional scaling guidance

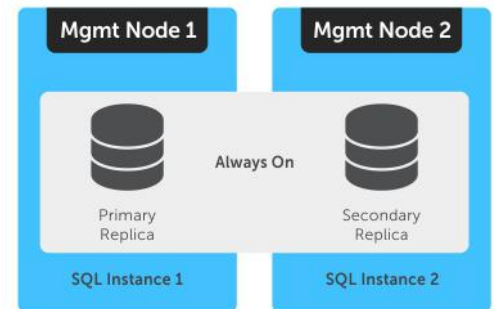
- The components are scaled either horizontally (by adding additional physical and virtual servers to the server pools) or vertically (by adding virtual resources to the infrastructure)
- Eliminate bandwidth and performance bottlenecks as much as possible
- Allow future horizontal and vertical scaling with the objective of reducing the future cost of ownership of the infrastructure.

Table 17 Components, metrics and scalability

Component	Metric	Horizontal scalability	Vertical scalability
Compute Servers	Desktop VMs per physical host based on available CPU	Additional hosts and clusters added as necessary	Additional RAM or CPU compute power
Mgmt Servers	Number of server VMs per host	Add additional hosts	Add RAM or network adapters
Provisioning Servers	Desktops per instance	Additional servers added to the Provisioning Server farm	Additional network and I/O capacity added to the servers
Desktop Delivery Servers	Desktops per instance (dependent on SQL performance as well)	Additional servers added to the XenDesktop Site	Additional virtual machine resources (RAM and CPU)
XenApp Servers	Desktops per instance	Additional virtual servers added to the XenDesktop Site	Additional physical servers to host virtual XenApp servers.
Storefront Servers	Logons/ minute	Additional servers added to the Storefront environment	Additional virtual machine resources (RAM and CPU)
Database Services	Concurrent connections, responsiveness of reads/ writes	Migrate databases to a dedicated SQL server and increase the number of management nodes	Additional RAM and CPU for the management nodes
File Services	Concurrent connections, responsiveness of reads/ writes	Split user profiles and home directories between multiple file servers in the cluster. File services can also be migrated to the optional NAS device to provide high availability.	Additional RAM and CPU for the management nodes

## 5.5 Solution high availability

High availability (HA) is offered to protect each architecture solution layer, individually if desired. Following the N+1 model, additional ToR switches are added to the Network layer and stacked to provide redundancy as required, additional compute and management hosts are added to their respective layers, vSphere or Hyper-V clustering is introduced in both the management and compute layers, SQL is configured for AlwaysOn or clustered and NetScaler is leveraged for load balancing.



The HA options provide redundancy for all critical components in the stack while improving the performance and efficiency of the solution as a whole.

- Additional switches added to the existing thereby equally spreading each host's network connections across multiple switches.
- Additional ESXi or Hyper-V hosts added in the compute or management layers to provide N+1 protection.
- Applicable Citrix infrastructure server roles are duplicated and spread amongst management host instances where connections to each are load balanced via the addition of virtual NetScaler appliances.
- SQL Server databases also are protected through the addition and configuration of an "AlwaysOn" Failover Cluster Instance or Availability Group.

Please refer to these links for more information: [SQL Server AlwaysOn Availability Groups](#) and [Windows Server Failover Clustering with SQL Server](#)

## 5.6 Communication flow for XenDesktop

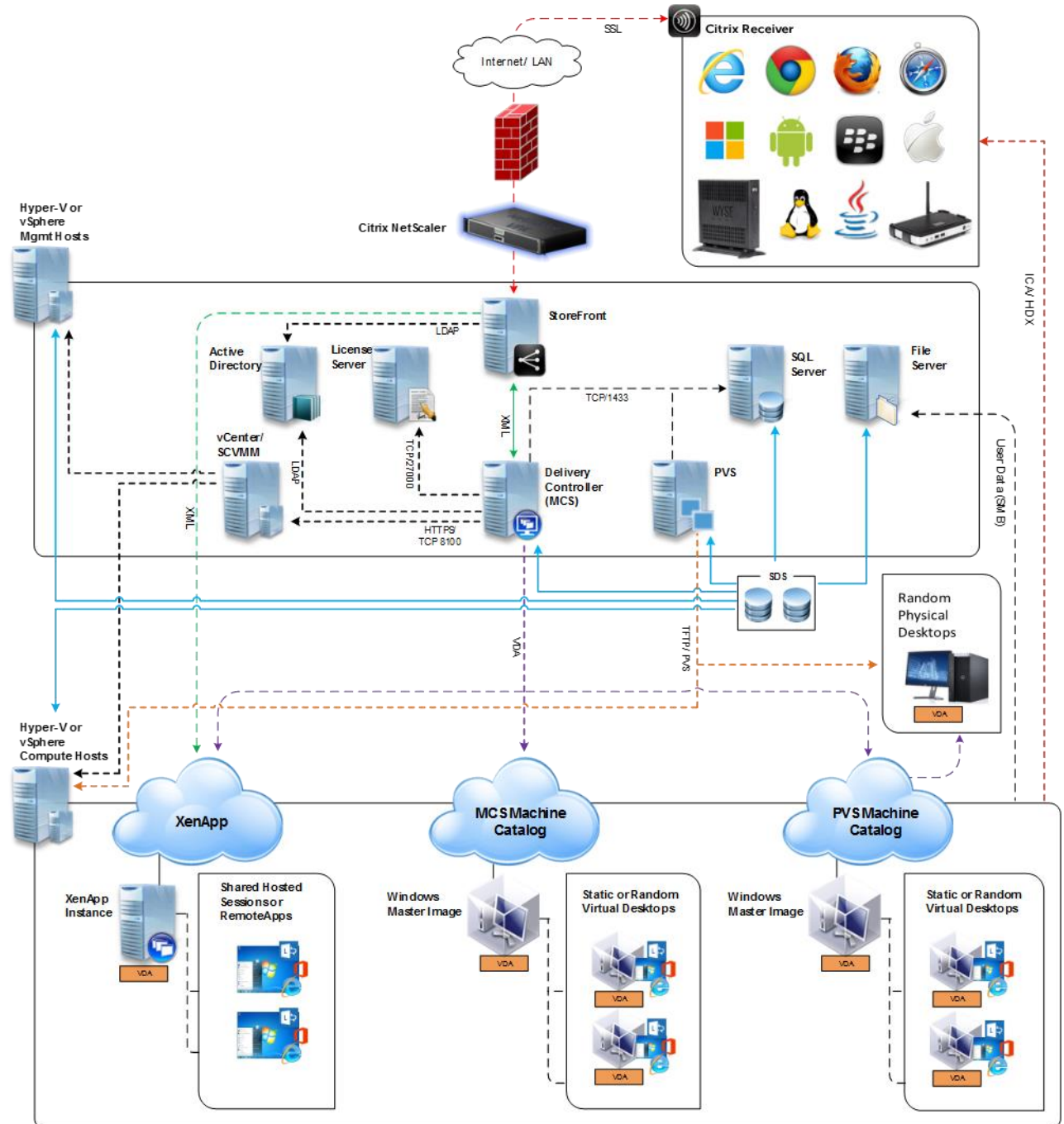


Figure 40 Communication flow for XenDesktop

## 6 Solution performance and testing

### 6.1 Summary

At the time of publication, these are the available density recommendations per appliance/node. Please refer to the Platform Configurations section for hardware specifications.

NOTE: All-flash configurations yield the same user densities with our test methodology since processor and memory resources are exhausted before storage resources are impacted.

Table 18 Standard user density summary

Host Config*	Hypervisor	Broker & Provisioning	Workload	Template	User Density*
XC Series C7	ESXi 6.5 U1	XenDesktop 7.15 MCS	Task Worker	Windows 10 & Office 2016	190/200**
XC Series C7	ESXi 6.5 U1	XenDesktop 7.15 MCS	Knowledge Worker	Windows 10 & Office 2016	150/160
XC Series C7	ESXi 6.5 U1	XenDesktop 7.15 MCS	Power Worker	Windows 10 & Office 2016	128/130
XC Series C7	ESXi 6.5 U1	XenDesktop 7.15 MCS	Power Worker	Windows 10 & Office 2016	48 + 120***
XC Series C7	Hyper-V 2012R2	XenDesktop 7.15 MCS	Task Worker	Windows 10 & Office 2016	140/210
XC Series C7	Hyper-V 2012R2	XenDesktop 7.15 MCS	Knowledge Worker	Windows 10 & Office 2016	120/185
XC Series C7	Hyper-V 2012R2	XenDesktop 7.15 MCS	Power Worker	Windows 10 & Office 2016	120/145

Dell EMC is aware of the new side-channel analysis vulnerabilities, known as Meltdown and Spectre, affecting many modern microprocessors that were discovered and published by a team of security researchers on January 3, 2018. Further information is available at the following locations:

- [https://emcservice--c.na55.visual.force.com/apex/KB\\_Security\\_KB?id=kA6f1000000FD0g](https://emcservice--c.na55.visual.force.com/apex/KB_Security_KB?id=kA6f1000000FD0g)
- <http://www.dell.com/support/article/SLN308588>
- <http://www.dell.com/support/article/SLN308587>

XenDesktop and vSphere management roles were also deployed on the cluster on a single host that also hosted desktops. This optional configuration is good for POCs or small deployments looking to maximize user density. The first value in the User Density column shows the Management Host's density if desktop VMs were to be deployed in addition to the mgmt. VMs, the second value represents the Compute Hosts density only. (ex. 140/210). Proper HA design should be part of any production deployment!

The Nutanix CVM uses 12 cores with high priority on each host, 10,000MHz reserved on ESXi (roughly 5 physical cores reserved). This was a factor in reducing density to a level well below the usual target CPU and Memory usage for a vSphere or Hyper-V cluster. The detailed validation results and analysis of these reference designs are in the following sections.



\*User density values are reported here as management + desktop VMs/ desktop VMs only. In smaller 3-4 node clusters, this shows the impact of running desktop VMs in addition to mgmt. infra VMs on the mgmt. node to maximize user density. For example, the first result of 140/210 indicates 140 desktop VMs on the mgmt. host and 210 desktop VMs per node on the compute hosts, totaling 560 users on this 3-node cluster.

\*\*The Task Worker User Density was projected based on the results of a successful test run with 560 users with 180/190 user densities that appeared to be below the full loading for the system. An additional test run to confirm the 590 user density was not possible given schedule limitations.

\*\*\*For graphics acceleration testing we performed assigned 1GB framebuffers both with vGPU VMs only as well as vGPU VMs + standard non-accelerated VMs. Here 48 represents the vGPU VMs, 120 represents the non-accelerated VMs also running on the same server node.

## 6.2 Test and performance analysis methodology

### 6.2.1 Testing process

In order to ensure the optimal combination of end-user experience (EUE) and cost-per-user, performance analysis and characterization (PAAC) on Dell EMC VDI solutions is carried out using a carefully designed, holistic methodology that monitors both hardware resource utilization parameters and EUE during load-testing.

Login VSI is currently the load-generation tool used during PAAC of Dell EMC solutions. Each user load is tested against multiple runs. First, a pilot run to validate that the infrastructure is functioning and valid data can be captured, and then, subsequent runs allowing correlation of data.

At different times during testing, the testing team will complete some manual “User Experience” Testing while the environment is under load. This will involve a team member logging into a session during the run and completing tasks similar to the User Workload description. While this experience will be subjective, it will help provide a better understanding of the end user experience of the desktop sessions, particularly under high load, and ensure that the data gathered is reliable.

#### 6.2.1.1 Load generation

Login VSI by Login Consultants is the de-facto industry standard tool for testing VDI environments and server-based computing (RDSH environments). It installs a standard collection of desktop application software (e.g. Microsoft Office, Adobe Acrobat Reader) on each VDI desktop; it then uses launcher systems to connect a specified number of users to available desktops within the environment. Once the user is connected, the workload is started via a logon script which starts the test script once the user environment is configured by the login script. Each launcher system can launch connections to a number of ‘target’ machines (i.e. VDI desktops). The launchers and Login VSI environment are configured and managed by a centralized management console.

Additionally, the following login and boot paradigm is used:

- Users are logged in within a login timeframe of 1 hour. Exception to this login timeframe occurs when testing low density solutions such as GPU/graphics based configurations. With those configurations, users are logged on every 10-15 seconds.
- All desktops are pre-booted in advance of logins commencing.
- All desktops run an industry-standard anti-virus solution. Windows Defender is used for Windows 10 due to issues implementing McAfee.



### 6.2.1.2 Profiles and workloads

It's important to understand user workloads and profiles when designing a desktop virtualization solution in order to understand the density numbers that the solution can support. At Dell EMC, we use five workload / profile levels, each of which is bound by specific metrics and capabilities with two targeted at graphics-intensive use cases. We will present more detailed information in relation to these workloads and profiles below but first it is useful to define the terms “profile” and “workload” as they are used in this document.

- **Profile:** This is the configuration of the virtual desktop - number of vCPUs and amount of RAM configured on the desktop (i.e. available to the user).
- **Workload:** This is the set of applications used by performance analysis and characterization (PAAC) of Dell EMC VDI solutions (e.g. Microsoft Office applications, PDF Reader, Internet Explorer etc.)

Load-testing on each profile is carried out using an appropriate workload that is representative of the relevant use case and summarized in the table below:

Table 19 Profile to workload mapping

Profile Name	Workload
Task Worker	Login VSI Task worker
Knowledge Worker	Login VSI Knowledge worker
Power Worker	Login VSI Power worker

Login VSI workloads are summarized in the sections below. Further information for each workload can be found on Login VSI's [website](#).

#### Login VSI Task Worker Workload

The Task Worker workload runs fewer applications than the other workloads (mainly Excel and Internet Explorer with some minimal Word activity, Outlook, Adobe, copy and zip actions) and starts/stops the applications less frequently. This results in lower CPU, memory and disk IO usage.

#### Login VSI Knowledge Worker Workload

The Knowledge Worker workload is designed for virtual machines with 2vCPUs. This workload contains the following activities:

- Outlook, browse messages.
- Internet Explorer, browse different webpages and a YouTube style video (480p movie trailer) is opened three times in every loop.
- Word, one instance to measure response time, one instance to review and edit a document.
- Doro PDF Printer & Acrobat Reader, the Word document is printed and exported to PDF.
- Excel, a very large randomized sheet is opened.
- PowerPoint, a presentation is reviewed and edited.
- FreeMind, a Java based Mind Mapping application.
- Various copy and zip actions.

#### Login VSI Power Worker Workload

The Power Worker workload is the most intensive of the standard workloads. The following activities are performed with this workload:

- Begins by opening four instances of Internet Explorer which remain open throughout the workload.
- Begins by opening two instances of Adobe Reader which remain open throughout the workload.
- There are more PDF printer actions in the workload as compared to the other workloads.
- Instead of 480p videos a 720p and a 1080p video are watched.
- The idle time is reduced to two minutes.
- Various copy and zip actions.

## 6.2.2 Resource monitoring

The following sections explain respective component monitoring used across all Dell EMC solutions where applicable.

### 6.2.2.1 GPU resources

#### ESXi hosts

For gathering of GPU related resource usage, a script is executed on the ESXi host before starting the test run and stopped when the test is completed. The script contains NVIDIA System Management Interface commands to query each GPU and log GPU utilization and GPU memory utilization into a .csv file.

ESXi 6.5 and above includes the collection of this data in the vSphere Client/Monitor section. GPU processor utilization, GPU temperature, and GPU memory utilization can be collected the same was as host CPU, host memory, host Network, etc.

### 6.2.2.2 Microsoft Performance Monitor

Microsoft Performance Monitor is used for Hyper-V based solutions to gather key data (CPU, Memory, Disk and Network usage) from each of the compute hosts during each test run. This data is exported to .csv files for single hosts and then consolidated to show data from all hosts (when multiple are tested). While the report does not include specific performance metrics for the Management host servers, these servers are monitored during testing to ensure they are performing at an expected performance level with no bottlenecks.

### 6.2.2.3 VMware vCenter

VMware vCenter is used for VMware vSphere-based solutions to gather key data (CPU, Memory, Disk and Network usage) from each of the compute hosts during each test run. This data is exported to .csv files for single hosts and then consolidated to show data from all hosts (when multiple are tested). While the report does not include specific performance metrics for the Management host servers, these servers are monitored during testing to ensure they are performing at an expected performance level with no bottlenecks.

## 6.2.3 Resource utilization

Poor end-user experience is one of the main risk factors when implementing desktop virtualization but a root cause for poor end-user experience is resource contention: hardware resources at some point in the solution have been exhausted, thus causing the poor end-user experience. In order to ensure that this does not happen, PAAC on Dell EMC solutions monitors the relevant resource utilization parameters and applies relatively conservative thresholds as shown in the table below. Thresholds are carefully selected to deliver an optimal combination of good end-user experience and cost-per-user, while also providing burst capacity for seasonal / intermittent spikes in usage. Utilization within these thresholds is used to determine the number of virtual applications or desktops (density) that are hosted by a specific hardware environment (i.e. combination of server, storage and networking) that forms the basis for a Dell EMC RA.

Table 20 Resource utilization thresholds

Parameter	Pass/Fail Threshold
Physical Host CPU Utilization (ESXi hypervisor)*	100%
Physical Host Memory Utilization	85%
Network Throughput	85%
Storage IO Latency	20ms

\*Turbo mode is enabled; therefore, the CPU threshold is increased as it will be reported as over 100percent utilization when running with turbo.

## 6.3 Test configuration details

The following components were used to complete the validation testing for the solution:

Table 21 XC Series hardware and software test components

Component	Description/Version
Hardware platform(s)	Dell EMC XC740xd C7
Hypervisor(s)	VMware vSphere ESXi 6.5 U1
Broker technology	XenDesktop 7.15
Broker database	Microsoft SQL 2014
Management VM OS	Microsoft Windows Server 2012 R2
Virtual desktop OS	Microsoft Windows 10 Enterprise 64-bit
Office application suite	Microsoft Office 2016 Professional Plus
Login VSI test suite	4.1.25

### 6.3.1 Compute VM configurations

The following table summarizes the compute VM configurations for the various profiles/workloads tested.

Table 22 ESXi desktop VM specifications

User Profile	vCPUs	ESXi Memory Configured	ESXi Memory Reservation	Screen Resolution	Operating System
Task Worker	2	2GB	1GB	1280 X 720	Windows 10 Enterprise 64-bit
Knowledge Worker	2	3GB	1.5GB	1920 X 1080	Windows 10 Enterprise 64-bit
Power Worker	2	4GB	2GB	1920 X 1080	Windows 10 Enterprise 64-bit
Graphics Density Configuration	2	4 GB	4GB	4GB	Windows 10 Enterprise 64-bit
Graphics Performance Configuration	4	8GB	8GB	8GB	Windows 10 Enterprise 64-bit

Table 23 Hyper-V desktop VM specifications

User Profile	vCPUs	Hyper-V Startup Memory	Hyper-V Min   Max Dynamic	Screen Resolution	Operating System
Task Worker	2	1GB	1GB   2GB	1280 X 720	Windows 10 Enterprise 64-bit
Knowledge Worker	2	1.5GB	1.5GB   3GB	1920 X 1080	Windows 10 Enterprise 64-bit
Power Worker	2	2GB	2GB   4GB	1920 X 1080	Windows 10 Enterprise 64-bit

Table 24 Profile to workload mapping

Profile Name	Workload	OS Image
Standard	Login VSI Task worker	Shared
Enhanced	Login VSI Knowledge worker	Shared
Professional	Login VSI Power worker	Shared + Profile Virtualization
Graphics Density Configuration	Login VSI Power worker with ProLibrary	Persistent
Graphics Performance Configuration	Custom specific	Persistent

## 6.4 Standard VDI test results analysis

The following table summarizes the test results for the cluster using the various workloads and configurations.

In typical VDI density tests we attempt to approach “full loading” of the compute hosts at some predefined threshold of CPU or Memory. For Hyper-V hypervisors the threshold for CPU is defined as 85 percent, while the threshold for memory is defined as 90 percent. In this case it was not feasible to achieve full loading at those thresholds due to the resource reservations imposed on each host by the Nutanix CVM. The CVM is assigned 12 cores with high priority according to the recommended default settings from Nutanix. We found through testing that when the Host CPU usage by VMs approached or exceeded a certain threshold there were often problems with sessions failing or failing to logon. These were often evident in the VSI results in the chart named “VSIMax per computer” where we could see that a small subset of user sessions were experiencing extreme latency on the order of 24000ms in many cases. While the overall performance of the system appeared to be well within constraints, the number of VMs contending for CPU time saturated the CPUs such that many desktop VMs had to wait behind the CVM and other VMs.

Due to these issues with core-scheduling, it was necessary to respect a much lower CPU threshold to avoid a higher-than-desired CPU “readiness” on the desktop and management VMs, which caused excessive application latency and very bad user experience despite not “fully loading” the host.

In practice, we found that even with the host CPU usage near or below 50 percent desktop latency issues would appear, such as failed logons, sessions failing to become active, and in worst case VSI Errors. Our new criteria for a successful test run was to reduce these occurrences to below 2percent of all sessions. We also found that the Management host exhibited these problems at a much lower number of desktops, so the density on that node was reduced substantially. The “Density Per Host” column shows the relative numbers of desktops on the management host vs the compute hosts that were achieved using these criteria. Despite a relatively low average CPU reporting, the overall per node density using Hyper-V still exceeds that of ESXi with much higher rates of CPU utilization.

Table 25 Test results summary

Platform Config	Hypervisor	Broker & Provisioning	Login VSI Workload	Density Per Host	Avg CPU	Avg Mem Consumed	Avg IOPS / User	Avg Net Mbps / User
XC Series C7	ESXi 6.5U1	XenDesktop 7.15	Task Worker	180/190	77%	427 GB	3.3	3.57
XC Series C7	ESXi 6.5U1	XenDesktop 7.15	Knowledge Worker	155/160	83%	527 GB	4.2	6.21
XC Series C7	ESXi 6.5U1	XenDesktop 7.15	Power Worker	128/130	87%	561 GB	3.5	9.27
XC Series C7	Hyper-V 2012R2	XenDesktop 7.15	Task Worker	140/210	37%	370 GB	2.31	3.17
XC Series C7	Hyper-V 2012R2	XenDesktop 7.15	Knowledge Worker	120/185	44%	400 GB	1.89	5.29
XC Series C7	Hyper-V 2012R2	XenDesktop 7.15	Power Worker	120/145	44%	354 GB	1.68	7.47

**Density per Host:** Density reflects number of users per compute host that successfully completed the workload test within the acceptable resource limits for the host. For clusters, this reflects the average of the density achieved for all compute hosts in the cluster.

**Avg CPU:** This is the average CPU usage over the steady state period. For clusters, this represents the combined average CPU usage of all compute hosts.

NOTE: On the latest Intel series processors, the ESXi host CPU metrics will exceed the reported 100percent utilization for the host if Turbo Boost is enabled on the CPU (by default). Up to an additional 35percent of CPU is available from the Turbo Boost feature but this additional CPU headroom is not reflected in the metrics where the performance data is gathered. Therefore, CPU usage for ESXi hosts is adjusted and a line indicating the potential performance headroom provided by Turbo boost is included in each CPU graph. Hyper-V CPU percentages take into account this additional capability.

**Avg Consumed Memory:** ESXi consumed memory is the amount of host physical memory granted within a host. For clusters, this is the average consumed memory across all compute hosts over the steady state period.

**Avg Mem Active:** For ESXi hosts, active memory is the amount of memory that is actively used, as estimated by VMkernel based on recently touched memory pages. For clusters, this is the average amount of guest “physical” memory actively used across all compute hosts over the steady state period.

**Avg IOPS/User:** IOPS calculated from the average Disk IOPS figure over the steady state period divided by the number of users.

**Avg Net Mbps/User:** Amount of network usage over the steady state period divided by the number of users. For clusters, this is the combined average of all compute hosts over the steady state period divided by the number of users on a host. The unit is Mega-bits per second, not Mega-Bytes per second (MBps).

On both Hyper-V and ESXi the CVMs were configured with 12 x vCPUs and 32GB RAM. In past testing comparisons between VMware ESXi and Windows 2012 R2 Hyper-V we have found that the desktop density with Windows Server is higher, and that the limiting resource on the hosts is often Memory rather than CPU Usage.

One difference we have seen in the Hyper-V results is that we have seen significant intermittent spikes in the Disk IO Latency. These often correlate to spikes in the Network Bandwidth Usage. We believe these are related to cache deduplication activity between the CVMs on separate hosts. These latency spikes seemed to have little impact on the user experience.

## 6.4.1 XC740xd-C7

Refer to section 3 for hardware configuration details. Please note that the performance results reported below can be expected to apply to any platform with a C7 configuration.

The hardware configurations that were tested are summarized in the table(s) below.

Table 26 XC Series hardware configuration for standard VDI

Enterprise Platform	Platform Config	CPU	Memory	RAID Ctlr	Drive Config	Network
XC740XD-24	C7	6138 Gold (20-Core 2.0 GHz)	768GB @2666 MT/s	HBA 330	<ul style="list-style-type: none"> <li>2 x 120GB M.2</li> <li>2x 960 GB SSD 4x 1.8TB HDD</li> </ul>	4x Intel X710 rNDC

Compute and Management resources were split out with the following configuration across a three node Nutanix cluster and all test runs were completed with this configuration.

- Node 1 – Management + compute: vCenter Appliance (ESXi only), SQL Server, XenDesktop Delivery Controller, XenDesktop Storefront, Nutanix CVM, desktop VMs using remaining resources
- Node 2 – Dedicated Compute, Nutanix CVM and User VMs only.
- Node 3 – Dedicated Compute, Nutanix CVM and User VMs only.

1GB networking was used for the deployment of the XC Series appliances only while 10GB networking is required for standard cluster operation.

Instead of dedicated nodes, Nutanix CVMs and VDI management roles were deployed on the cluster with desktop VMs, which reduced maximum density on that node. Each compute node was loaded with desktops to its maximum density; no failover capacity was reserved.

### 6.4.1.1 Task Worker, 560 Total Users, ESXi 6.5, XenDesktop 7.15 MCS

The Task Worker test run provisioned 190 VMs on each Compute Host and 180 VMs on the Management Host along with the XenDesktop management roles. The Peak CPU usage was 94percent on one node at the end of the Logon phase, while the steady state average CPU usage was 77percent.

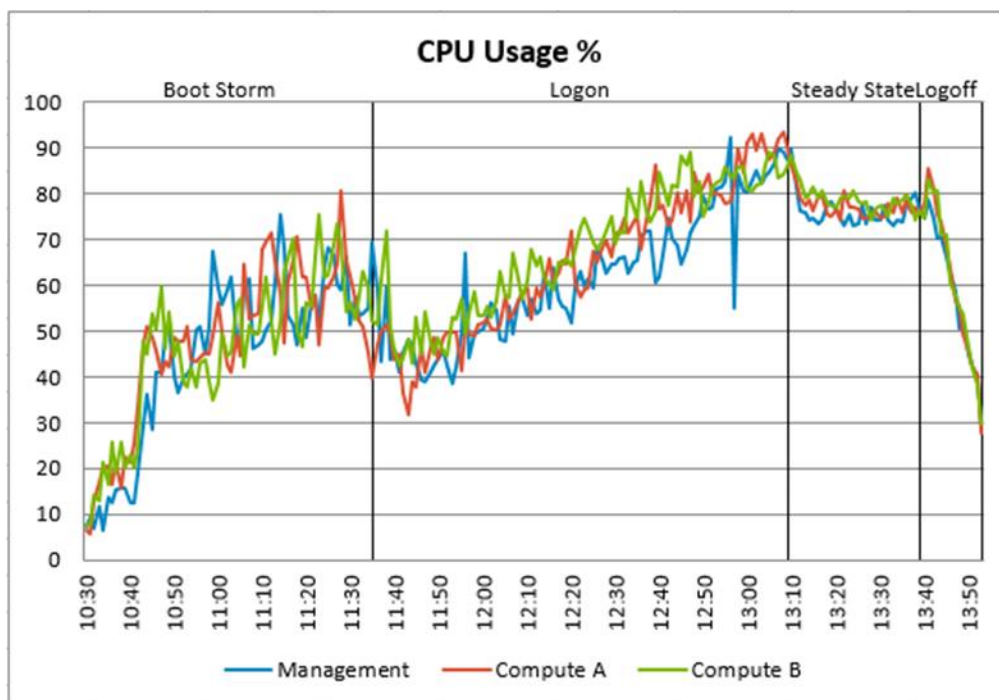


Figure 41 CPU usage

The memory consumed per host during the test run averaged 427 GB in Steady State Phase, and the peak memory usage was 430 GB on one host during Steady State. There was no Ballooning or Swapping of VM memory during the test run. Each user session consumed 2.07 GB of memory after accounting for CVM and management memory usage.

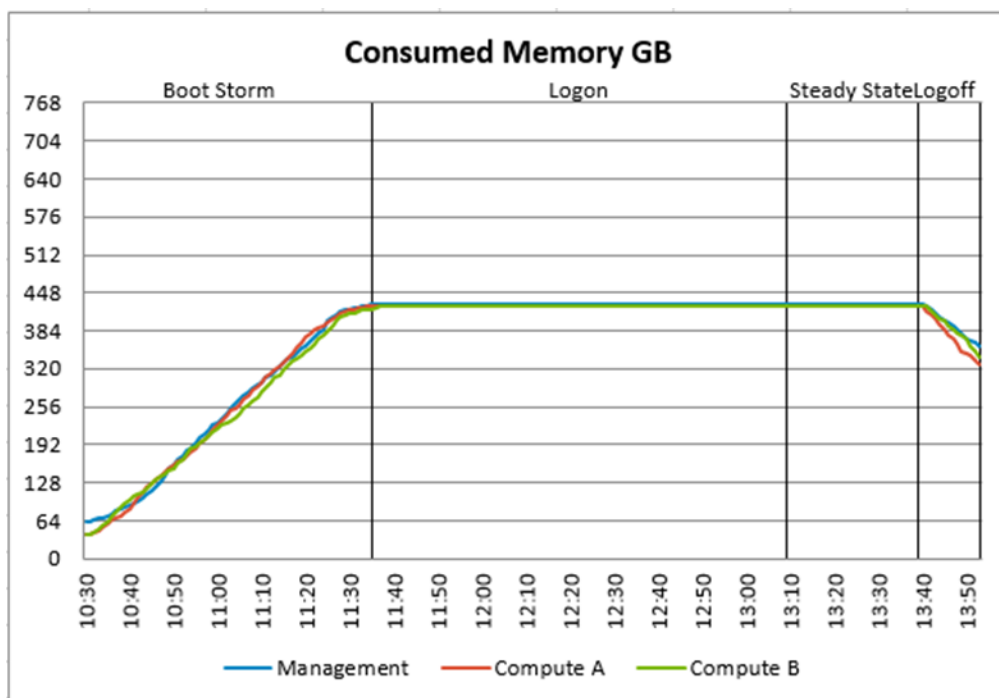


Figure 42 Consumed memory



The Steady State average active memory use was 180 GB, while the peak was 195 GB on one host during Logoff phase. Each user session used 0.78 GB of active memory during Steady State, after accounting for VM and management memory usage.

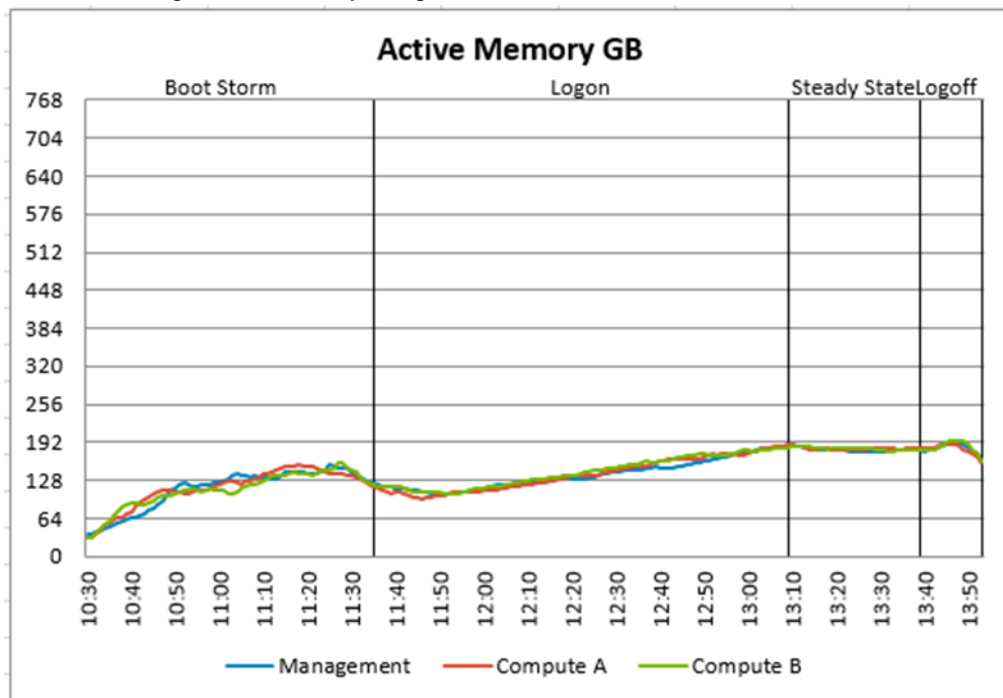
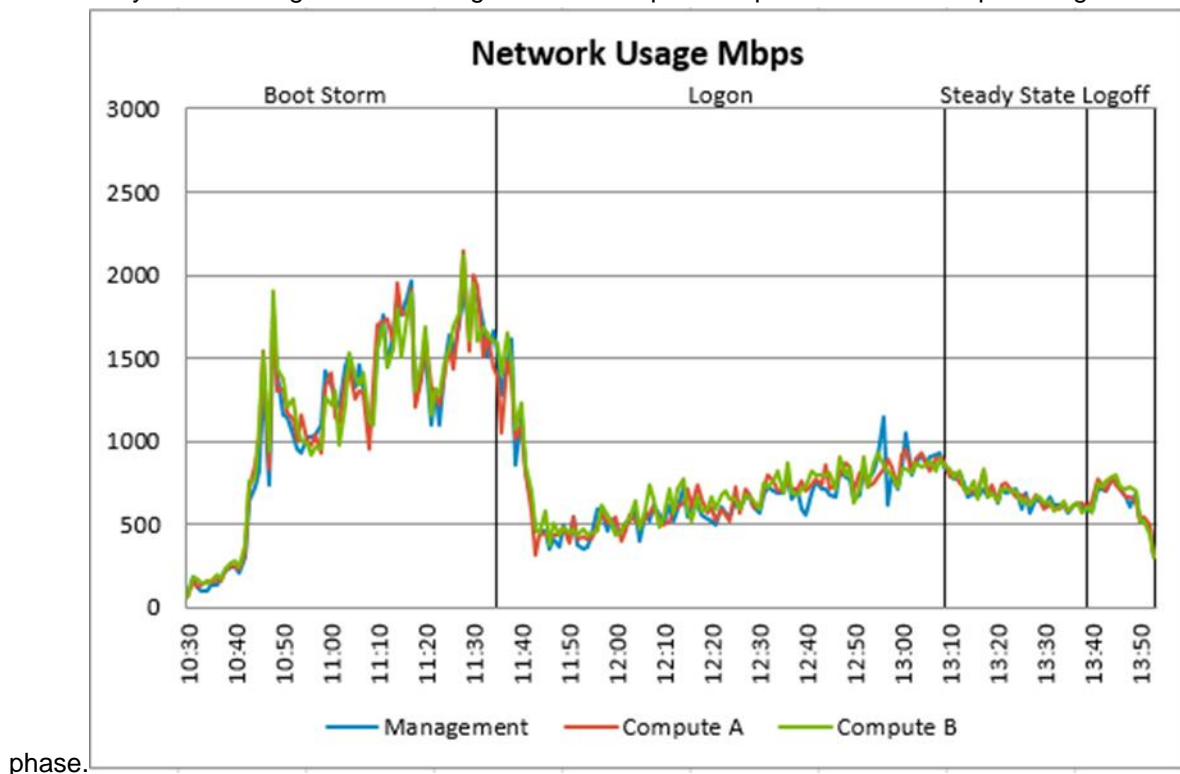


Figure 43 Active memory

The Steady State average network usage was 667 Mbps. The peak was 2140 Mbps during Boot Storm



phase.

Figure 44 Network usage

The peak Cluster IOPS for the test run was 6850 IOPS during the Boot Storm phase, while the average in Steady State was 1850 IOPS. Based on these numbers each user session generated 3.3 IOPS in Steady State.

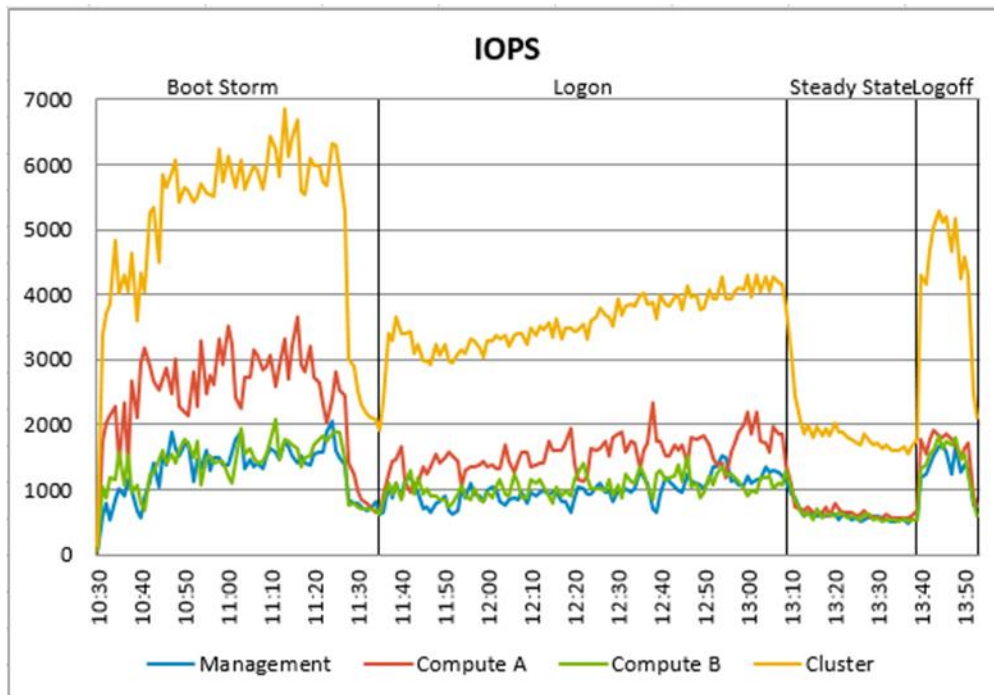


Figure 45 IOPS

The peak IO Latency was 1.1 ms during the Boot Storm. The average IO latency during Steady State was 0.4 ms. The chart clearly shows a very steady and very low level of IO Latency throughout the test run.

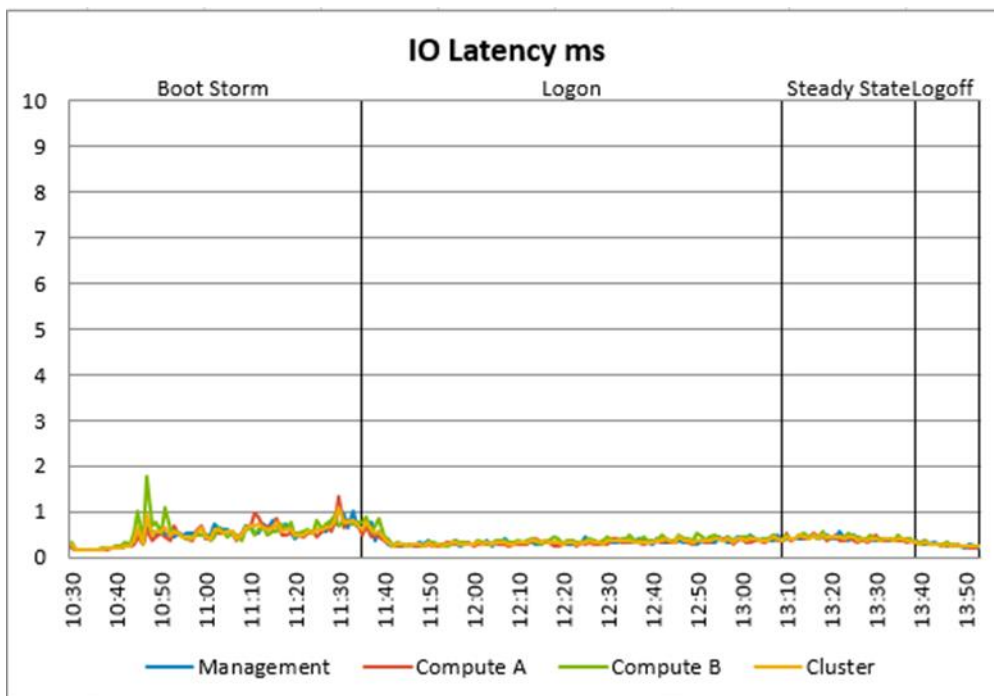


Figure 46 IO latency

The baseline performance of 842 indicated that the user experience for this test run was Good. The Index average reached 982 well below the threshold of 1843.

Table 27 Performance metrics

Login VSI Baseline	VSI Index Average	VSIMax Reached	VSI Threshold
842	982	NO	1843

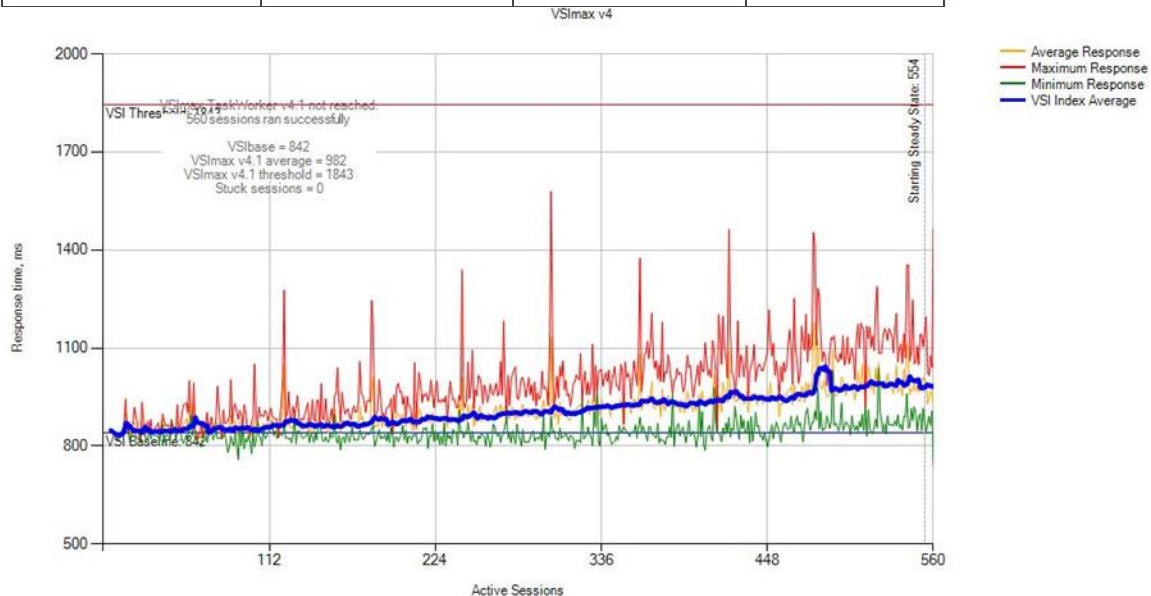


Figure 47 VSImax

#### 6.4.1.2 Knowledge Worker, 475 Total Users, ESXi 6.5, XenDesktop 7.15 MCS

In this workload test run the compute hosts each had 160 user sessions, while the management host had 155 sessions in addition to the XenDesktop management VMs. The peak CPU Usage was 99percent on one host during logon phase, while the Steady State average was 83percent across all hosts. The relatively low steady state CPU usage was essential to avoid CPU scheduling problems that would have placed many of the desktops into the Ready-waiting state. The average CVM CPU usage during steady state was 7.8percent, while the management VMs altogether used 1.2percent.

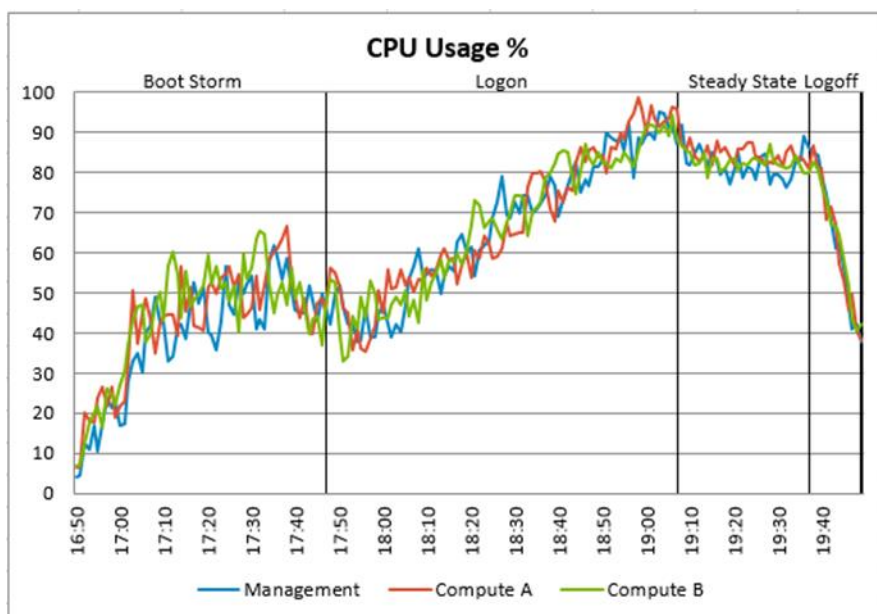


Figure 48 CPU usage

The memory consumption averaged 527 GB in steady state across all hosts, and the peak usage on any host was 532 GB during the steady state phase. The peak usage was 69percent, well below the 85percent threshold, while the Steady State average usage for all hosts was 68.6percent. There was no swapping or ballooning during the test run and each desktop consumed 3.06 GB, slightly more than the granted 3.0 GB, after accounting for CVM and management VM memory consumption. The CVM on each host consumed its full 32GB of reserved memory throughout the test run.

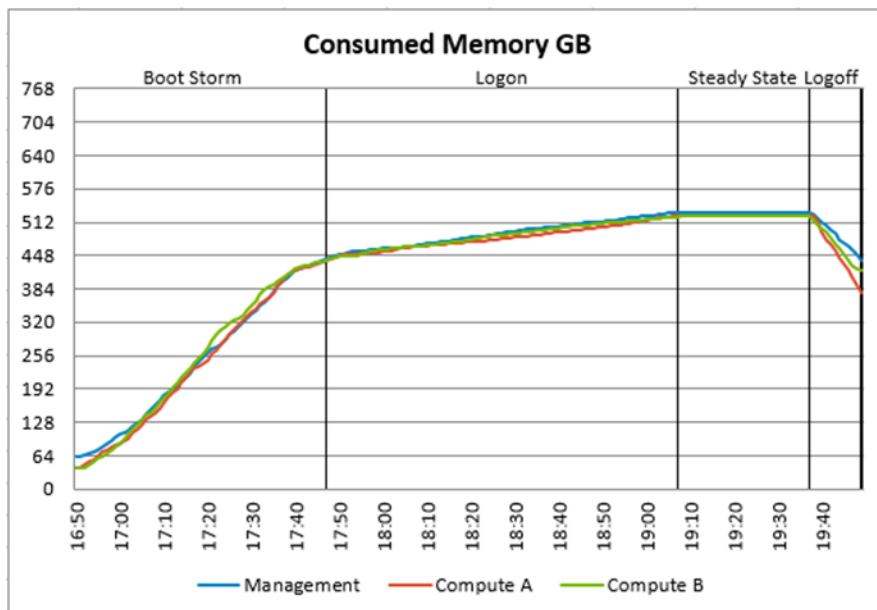


Figure 49 Consumed memory

Active memory during the test run peaked at 226 GB on the management host, and averaged 204 GB during Steady state. Each desktop accounted for 1.06 GB of active memory usage after deducting CVM and management VM active memory. The CVM on each host used a full 32 GB of active memory throughout the test run.

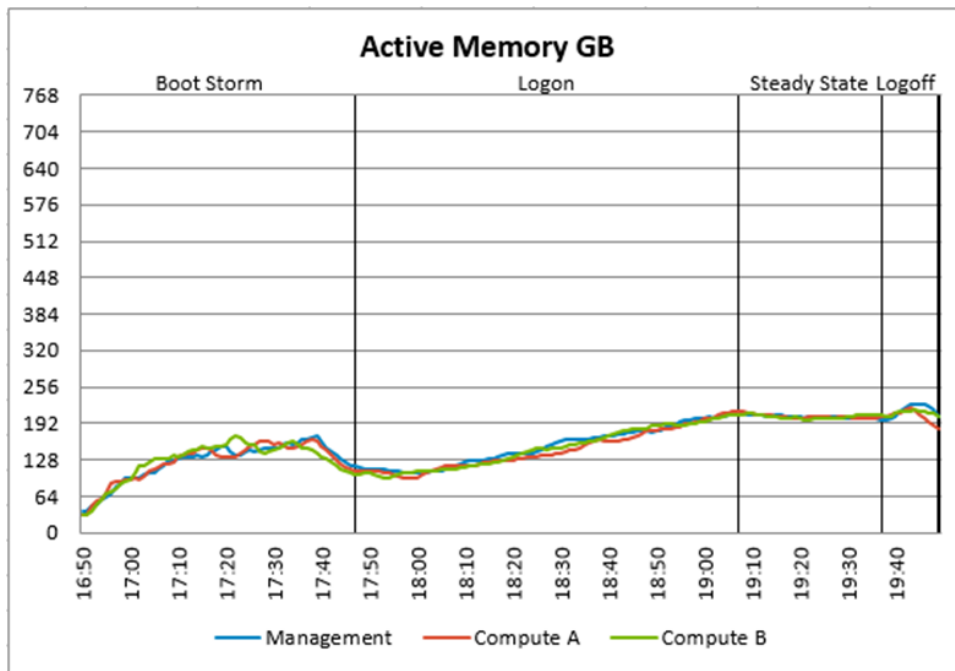


Figure 50 Active memory

Network usage peaked at 1709 Mbit per sec during Boot storm on one host, and the average network usage for all hosts was 984 Mbit per sec during Steady State. Each desktop produced network throughput of 6.32 Mbit per sec in steady state.

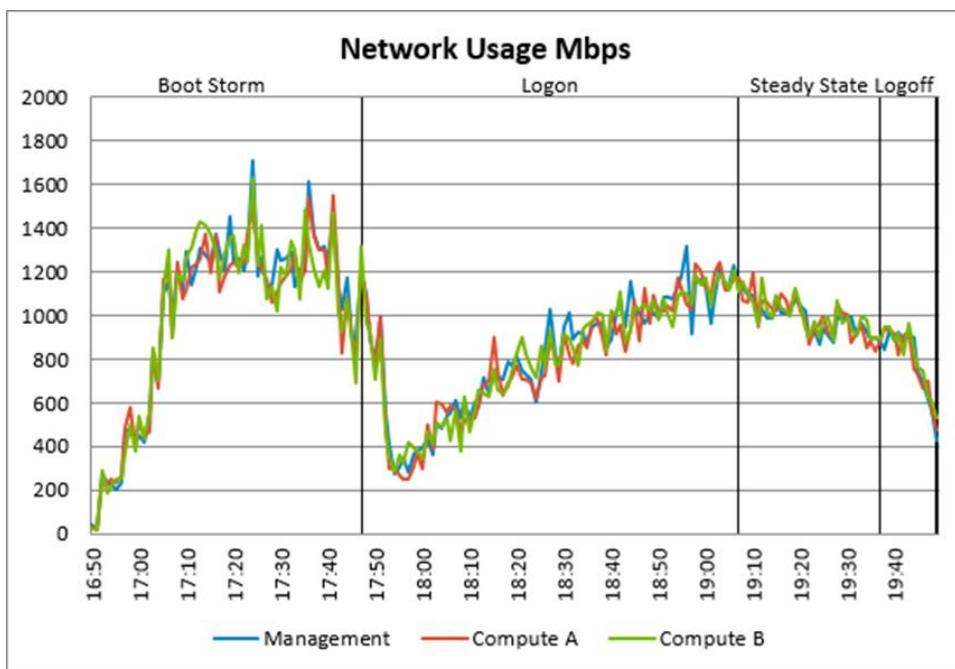


Figure 51 Network usage

The peak Cluster IOPS for the test run was 7589 IOPS at the end of Logon phase, while the average in Steady State was 1997 IOPS. Based on these numbers each user session generated 4.20 IOPS in steady state.

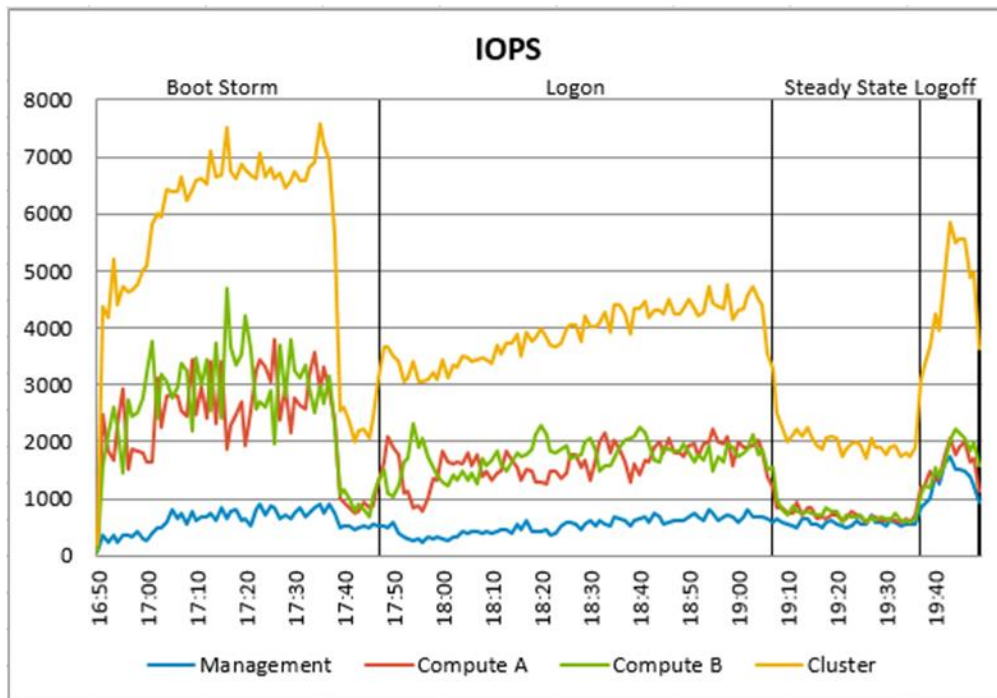


Figure 52 IOPS

The peak Cluster IO Latency was 0.9 ms during the Boot Storm. The average Cluster IO latency during steady state was 0.4 ms. The highest IO Latency on any host was 1.4 ms during the Boot Storm. The chart clearly shows a very steady and very low level of IO Latency throughout the test run.

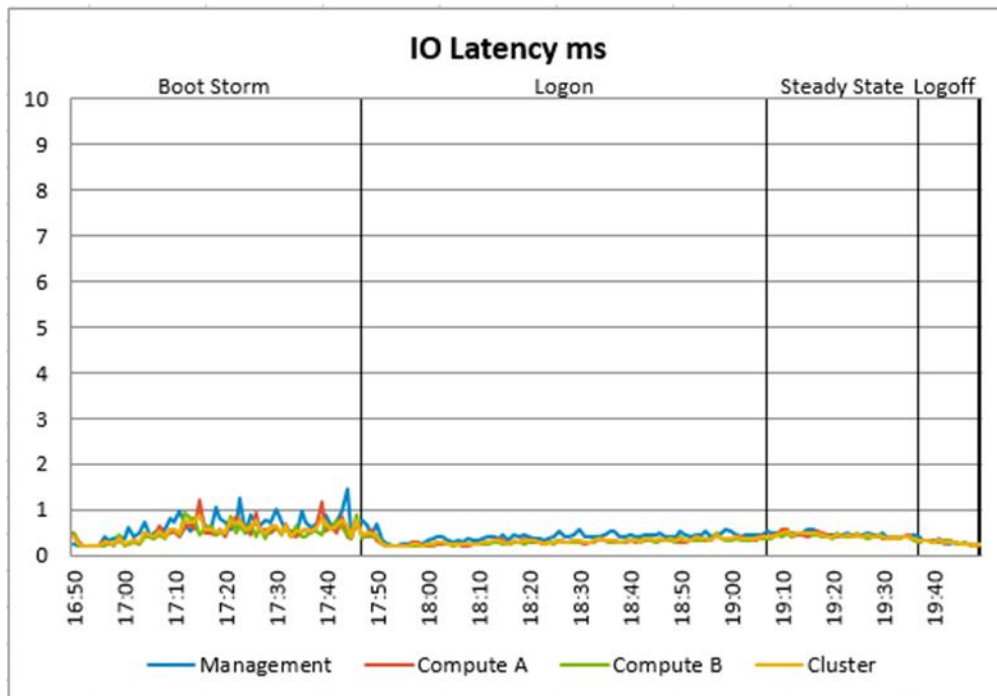


Figure 53 IO latency

The baseline performance of 851 indicated that the user experience for this test run was good. The Index average reached 1524, well below the VSIMax threshold of 1851.

Table 28 Performance metrics

Login VSI Baseline	VSI Index Average	VSIMax Reached	VSI Threshold
851	1524	NO	1851

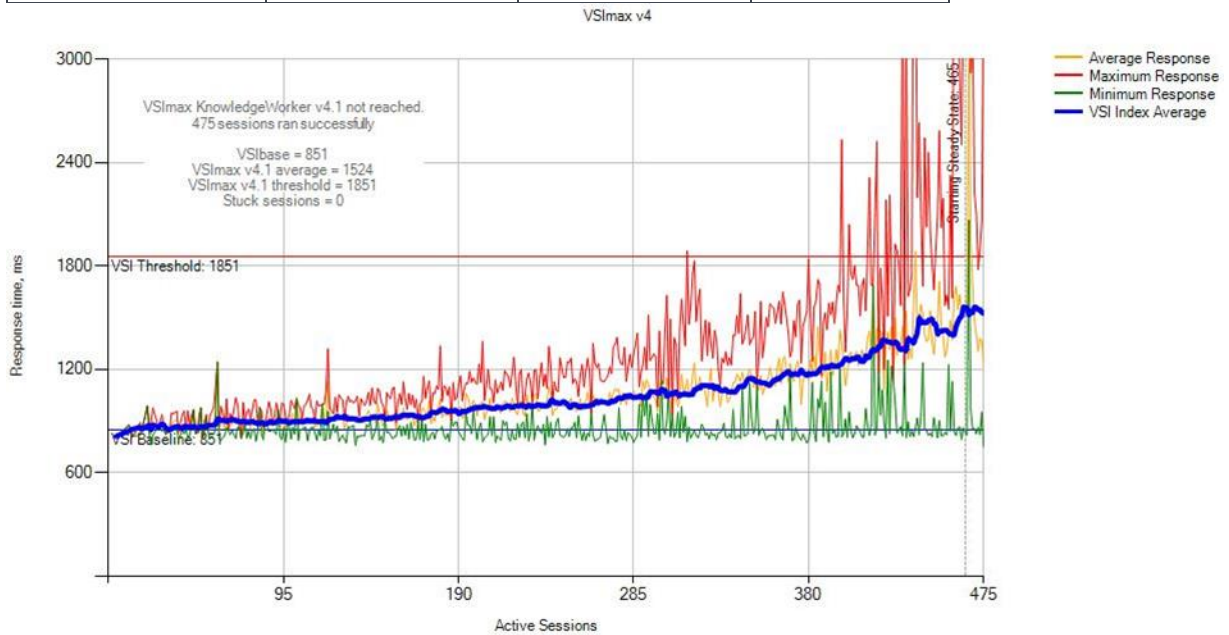


Figure 54 VSIMax



### 6.4.1.3 Power Worker, 390 Total Users, ESXi 6.5, XenDesktop 7.15 MCS

In this workload test run each compute host had 130 user sessions while the designated management host had the full set of management VMs plus 130 desktops. The peak CPU Usage was 99percent on one host during logon phase, while the Steady State average was 87percent across all hosts. The relatively low steady state CPU usage was essential to avoid CPU scheduling problems that would have placed many of the desktops into the Ready-waiting state.

The CVMs on each host averaged 7.5percent CPU usage during steady state. The management VMs used only 1.4percent CPU on the management host in steady state.

Provisioning 390 power worker desktops with XenDesktop MCS Fast Cloning on vSphere 6.5 U1 took only 11min 23 secs, not counting time required to power on the entire complement of desktops, which was 40 minutes once the desktops were added to a desktop group. Of the total provisioning time, 5 min 37 secs was required to prepare the base image, and the remaining time 5 min 46 secs was for creating the cloned VMs.

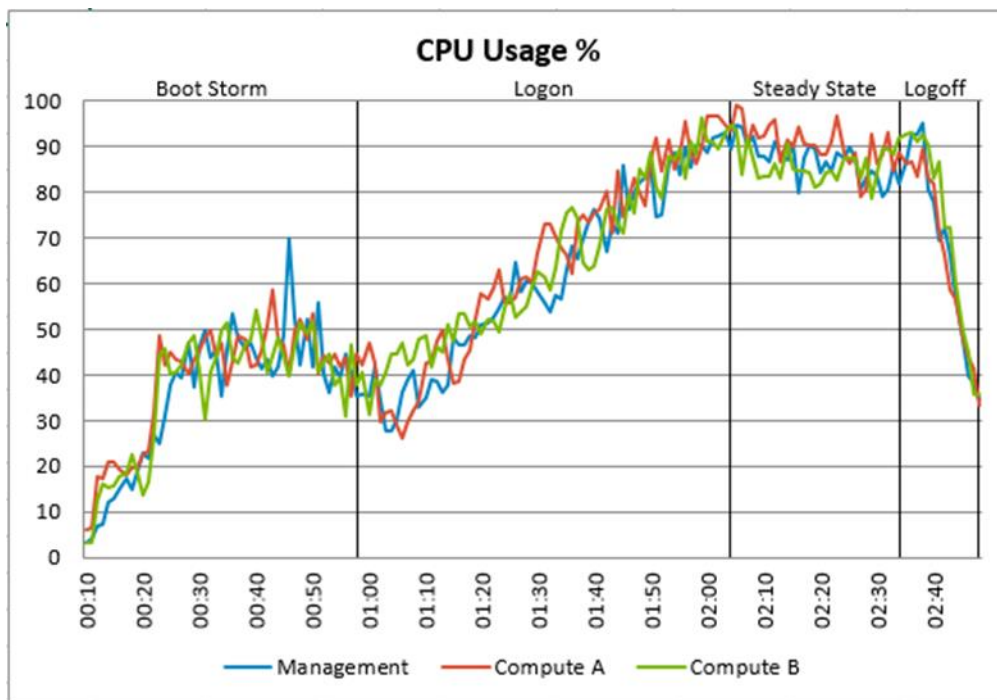


Figure 55 CPU usage



The memory consumption averaged 561 GB in steady state across all hosts, and the peak usage on any host was 584 GB during the steady state phase. The peak usage was 76percent, well below the 85percent threshold, while the Steady State average usage for all hosts was 73percent. There was no swapping or ballooning during the test run and each desktop consumed 4.01 GB, slightly more than the granted 4.0 GB, after accounting for CVM and management VM memory consumption.

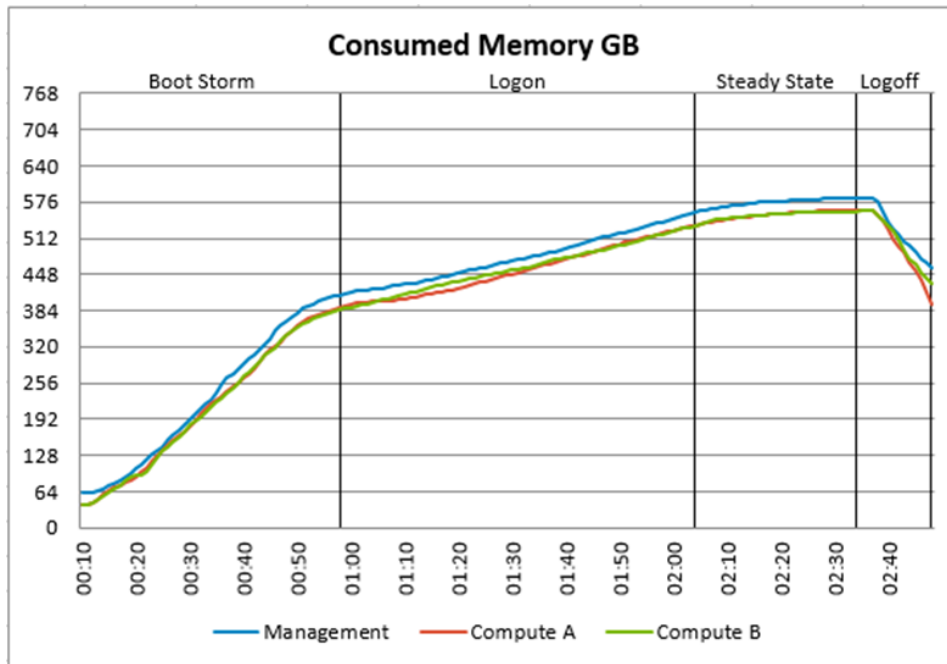


Figure 56 Consumed memory

Active memory usage reached a maximum of 196 GB on the management host during steady state, and the average steady state usage was 184 GB. Each desktop used 1.15 GB of active memory after deducting for CVM and management VM usage. The CVM used its full 32GB of active memory throughout the test run.

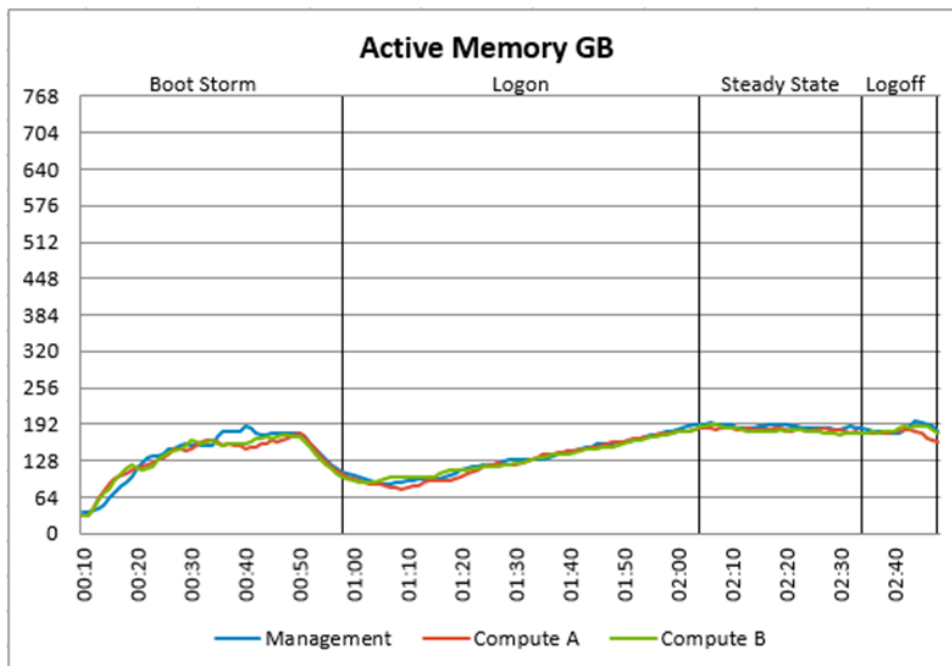


Figure 57 Active memory

Network usage peaked at 1495 Mbit per sec on one host during logon phase, and the average network usage for all hosts during Steady State was 1205 Mbit per sec. Each desktop accounted for 9.27 Mbit per sec in steady state.

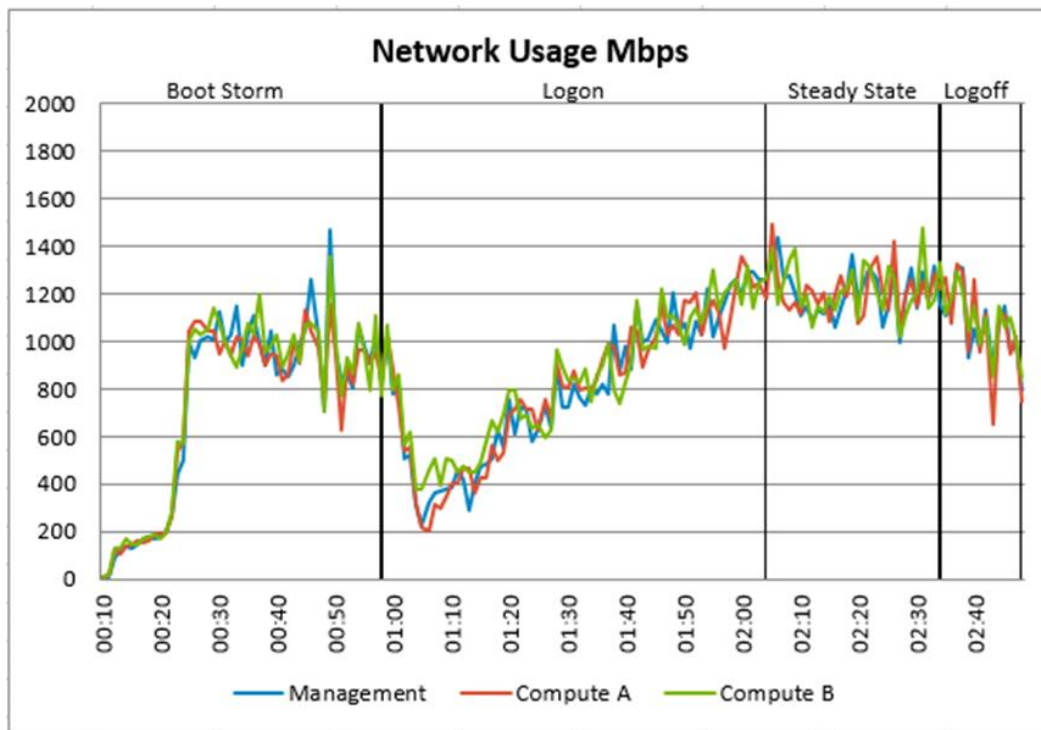


Figure 58 Network usage

The peak Cluster IOPS for the test run was 6885 IOPS during Logon, while the average in Steady State was 1381 IOPS. Based on these numbers each user session generated 3.54 IOPS during Steady State.

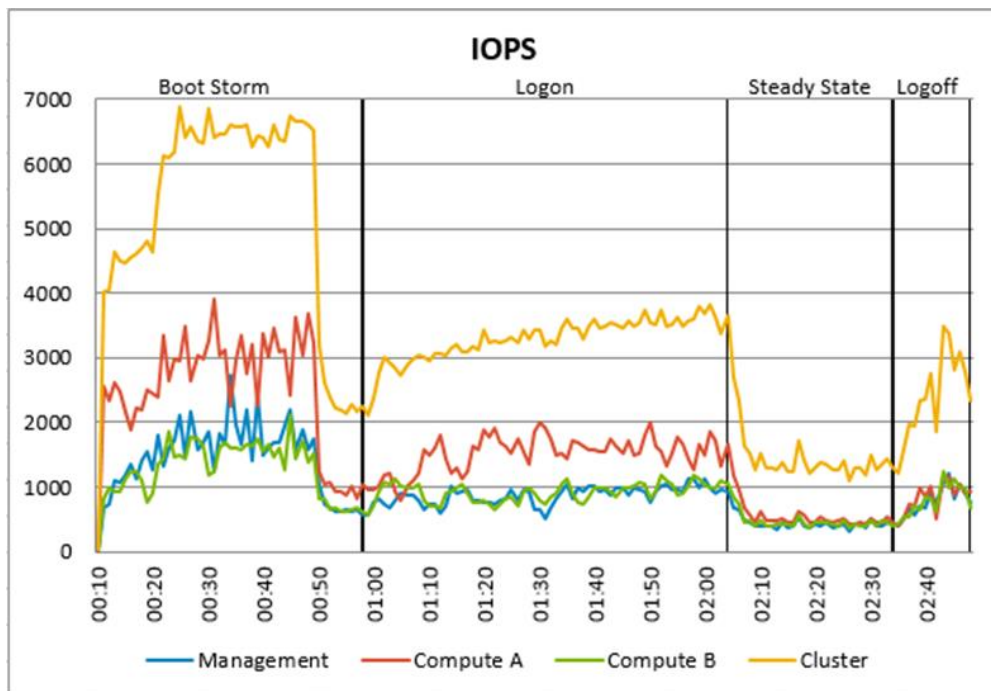


Figure 59 IOPS

The peak Cluster IO Latency was 0.8 ms during the Boot Storm, while the peak on any host was 1.0 ms. The average Cluster IO latency during steady state was 0.6 ms. The chart clearly shows a very steady and very low level of IO Latency throughout the test run.

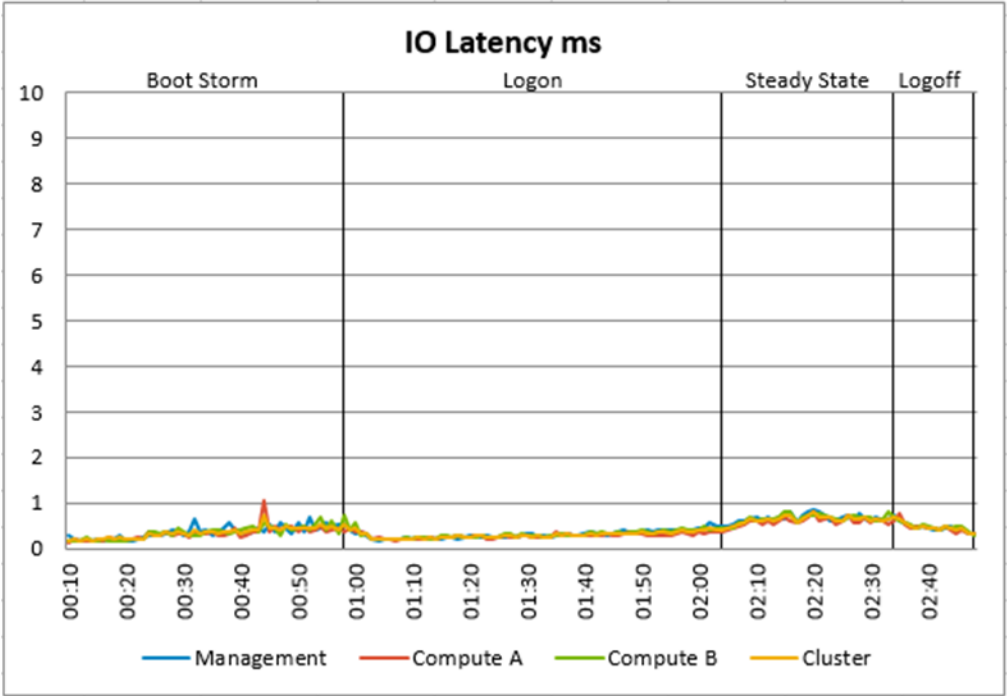


Figure 60 IO latency

The baseline performance of 834 indicated that the user experience for this test run was good. The Index average reached 1175, well below the threshold of 1834. Although the difference between the VSIMax and the average would seem to indicate that more desktops could be used, the CPU limitations described above would have reduced the user experience dramatically.

Table 29 Performance metrics

Login VSI Baseline	VSI Index Average	VSIMax Reached	VSI Threshold
834	1175	NO	1834

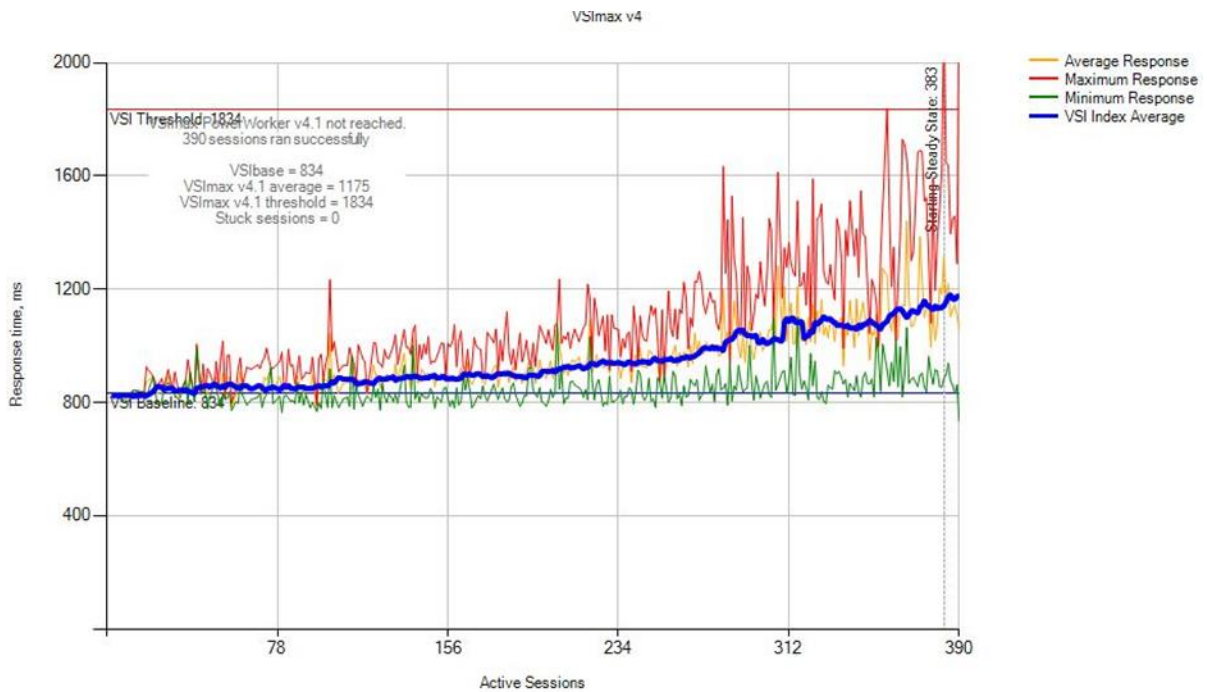


Figure 61 VSImax

#### 6.4.1.4 Task Worker, 560 Total Users, Hyper-V 2012R2, XenDesktop 7.15 MCS

The Task Worker test run provisioned 210 VMs on each compute host and 140 on the management host, along with all XenDesktop VM roles. The total provisioning time was 55 minutes and 45 secs, not counting the time required to power on and register the desktops in the Delivery Group.

The Peak CPU usage was 45percent on one node at the end of the Logon phase, while the steady state average CPU usage was 37percent. During steady state the CVM roles averaged 6.0percent on each host while the management roles averaged 0.9percent of the Management Host CPU.

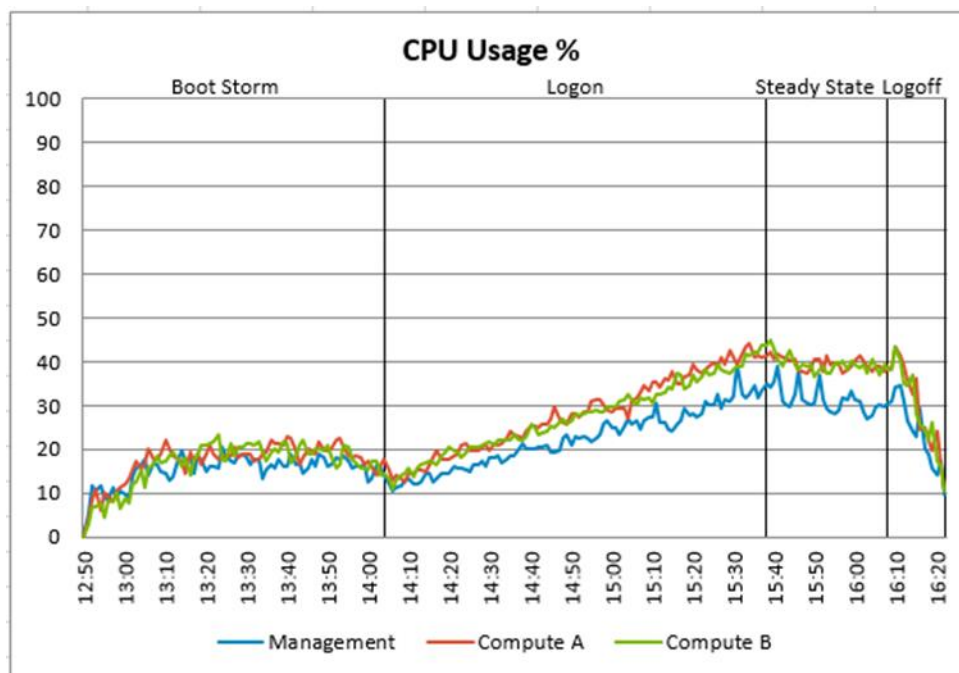


Figure 62 CPU usage

The memory used during the steady state averaged 370 GB, and the peak memory usage was 416 GB on one node during Steady State. The management host averaged only 308 GB in steady state.

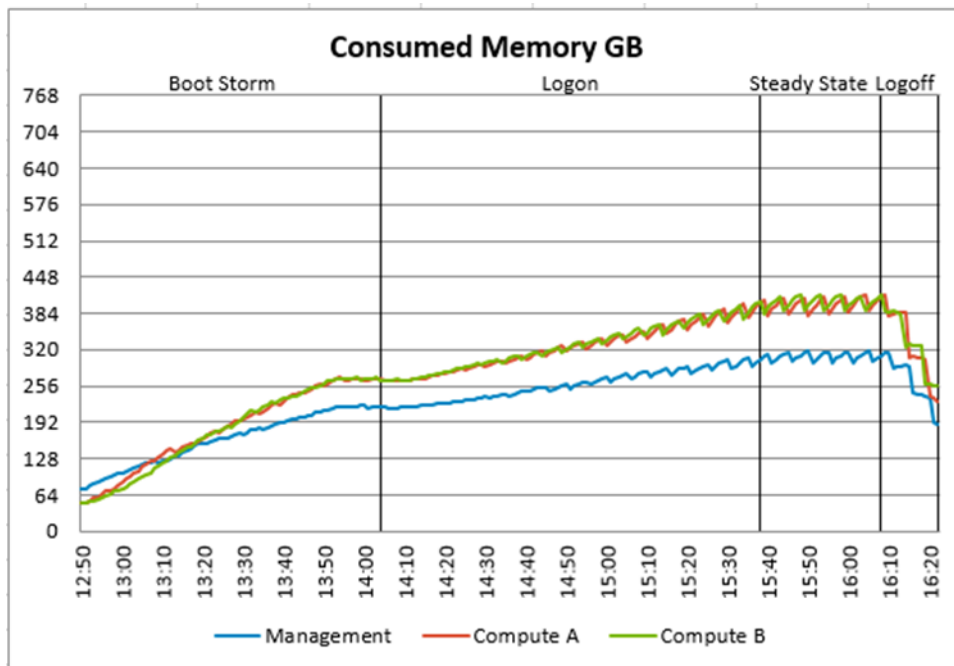


Figure 63 Consumed memory

The Steady State average network usage was 618 Mbit per sec on the compute hosts. The peak was 1720 during Boot Storm phase on node Compute B. The Management host averaged 508 Mbit per sec in steady state.

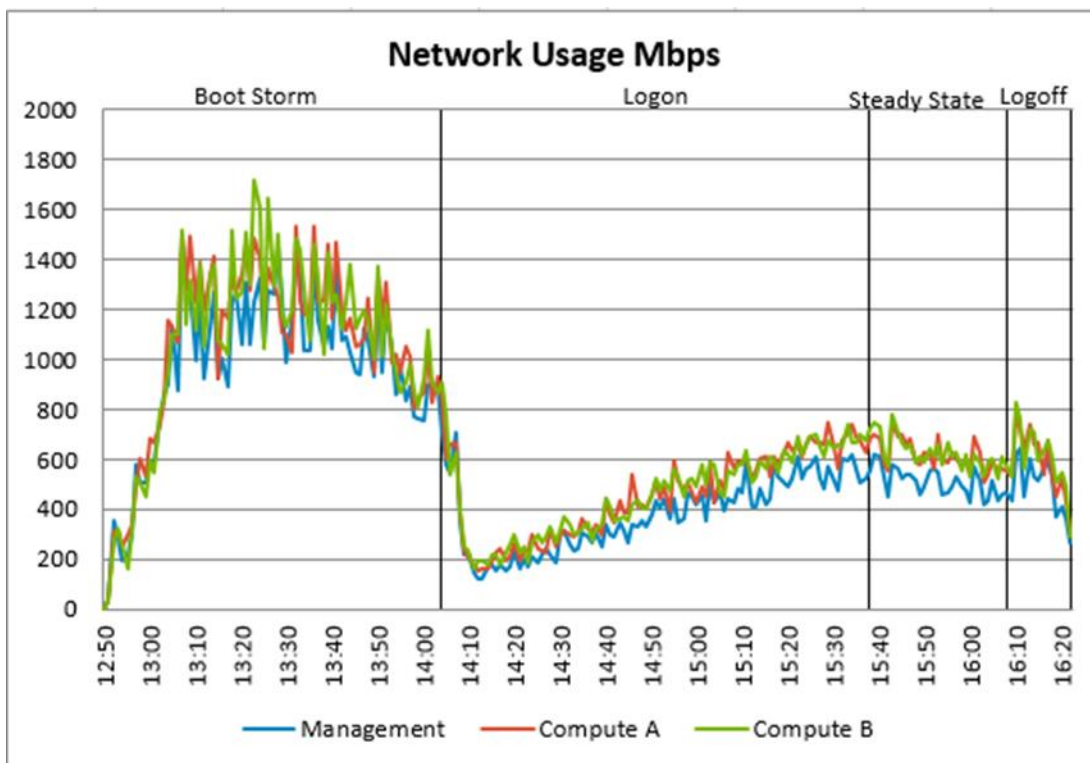


Figure 64 Network usage

The peak Cluster IOPS for the test run was 4140 IOPS during Boot Storm phase, while the average in Steady State was 1295 IOPS. The peak IOPS for any single node was 1848 IOPS on the Management host during the Boot Storm.

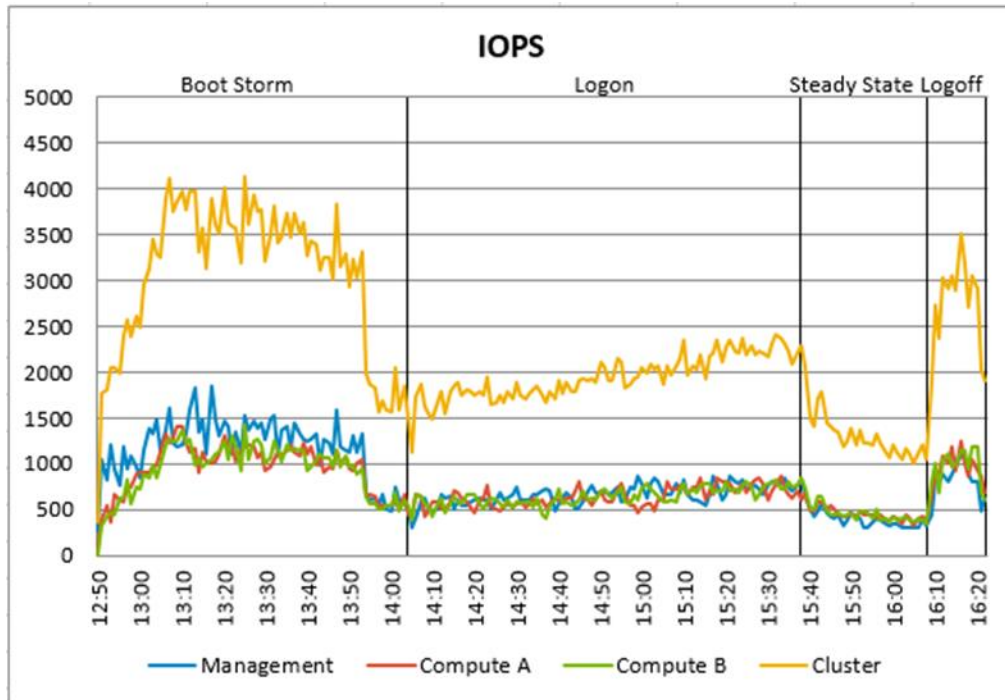


Figure 65 IOPS

The peak IO Latency was 1.4 ms during the Boot Storm. The average IO latency during steady state was 0.9 ms. The chart clearly shows a very steady and very low level of IO Latency throughout the test run.

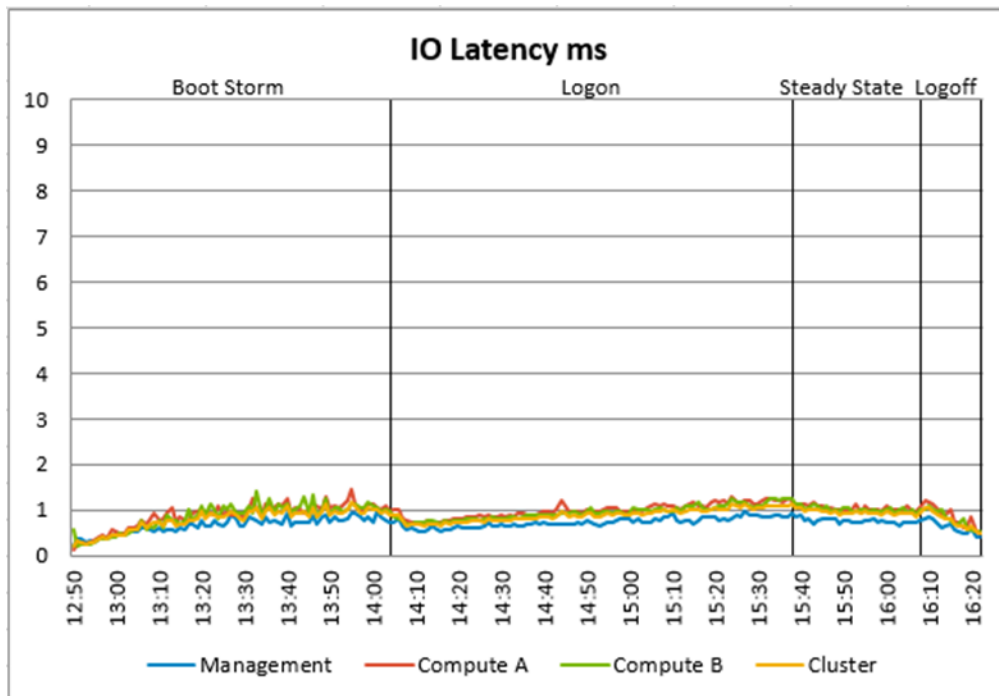


Figure 66 IO latency

The baseline performance of 776 indicated that the user experience for this test run was very good. The Index average reached 897, well below the threshold of 1777. A total of 9 sessions failed to login or failed to become active.

Table 30 Performance metrics

Login VSI Baseline	VSI Index Average	VSIMax Reached	VSI Threshold
776	897	NO	1777

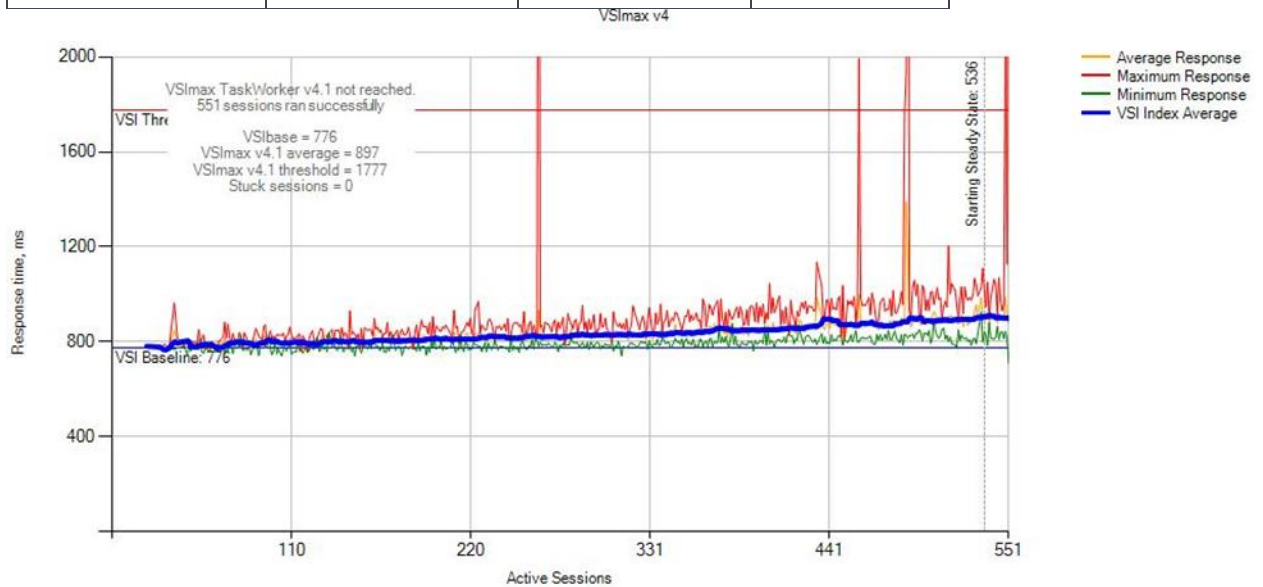


Figure 67 VSImax



#### 6.4.1.5 Knowledge Worker, 490 Total Users, Hyper-V 2012R2, XenDesktop 7.15 MCS

In this workload test run the compute hosts each had 185 user sessions, while the management host had 120 sessions in addition to the XenDesktop management VMs. The total provisioning time was 52 minutes and 14 secs, not counting the time required to power on and register the desktops in the Delivery Group.

The peak CPU Usage was 53% on one host at the end of Logon phase, while the Steady State average was 46% on the compute hosts and 38% on the management host. The relatively low steady state CPU usage was essential to avoid CPU scheduling problems that caused desktops to timeout during the workload. The average CVM CPU usage during steady state was 6.4%, while the management VMs altogether used 1.1%.

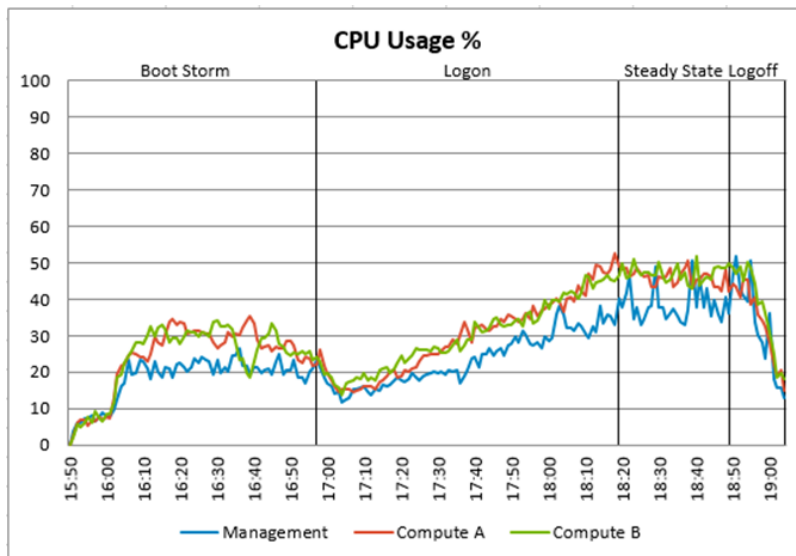


Figure 68 CPU usage

The memory consumption averaged 437 GB in steady state on the compute hosts, and 324 GB on the Management host. The management roles consumed 23.6 GB in steady state on the Management host. The peak usage on any host was 453 GB during the steady state phase. Each desktop consumed 2.21 GB, after accounting for CVM and management VM memory consumption. The CVM on each host consumed its full 32GB of reserved memory throughout the test run.

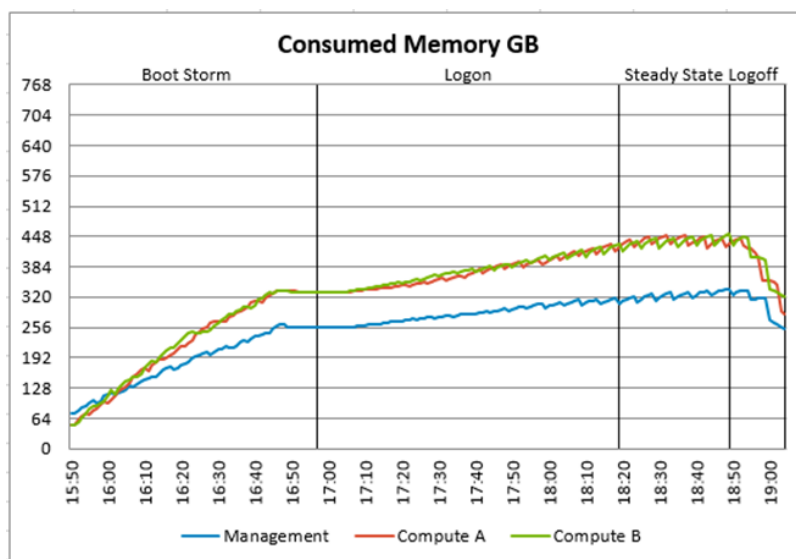


Figure 69 Consumed memory



Network usage peaked at 1516 Mbit per sec during Boot storm on one host, and the average network usage for all hosts was 849 Mbit per sec during Steady State. Each desktop produced network throughput of 5.29 Mbit per sec in steady state.

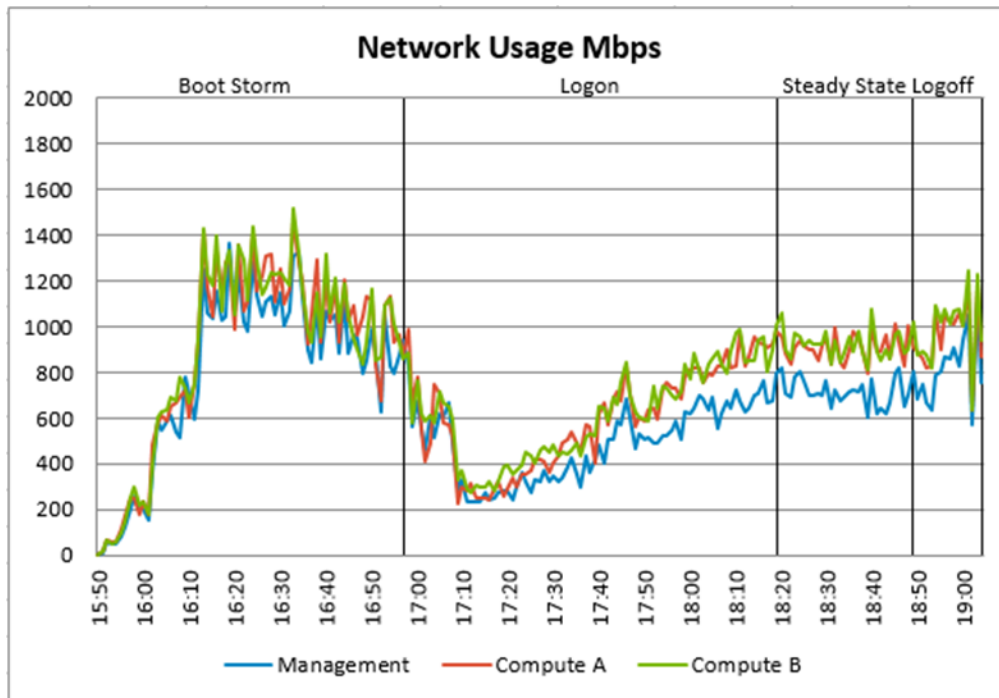


Figure 70 Network usage

The peak Cluster IOPS for the test run was 4703 IOPS during Boot Storm phase, while the average in Steady State was 924 IOPS. Based on these numbers each user session generated 1.89 IOPS in steady state.

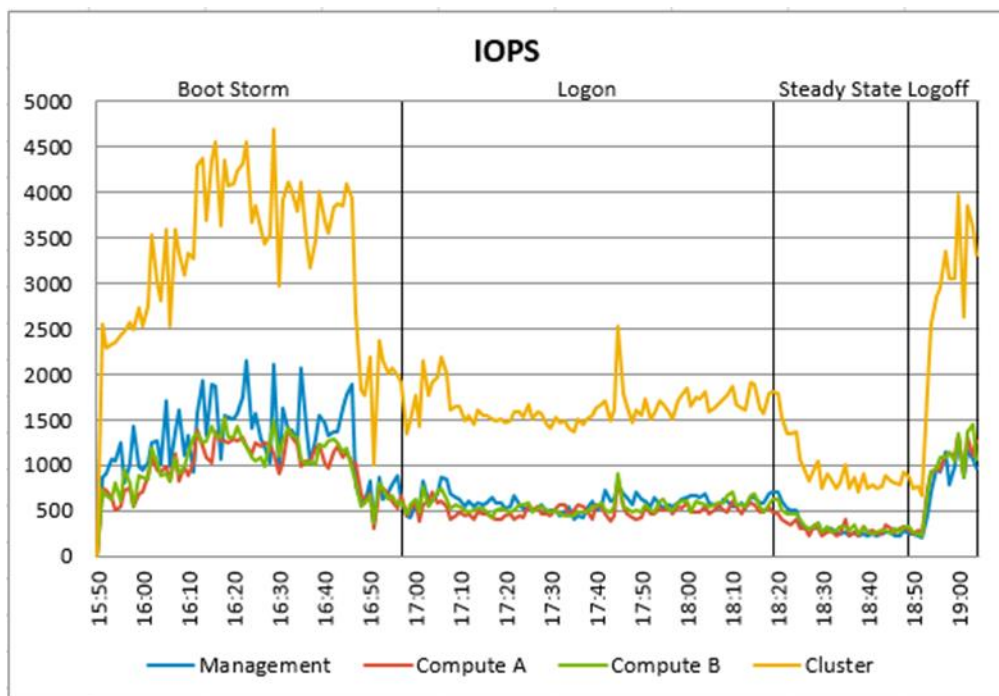


Figure 71 IOPS

The peak Cluster IO Latency was 1.2 ms during the Boot Storm. The average Cluster IO latency during steady state was 1.1 ms. The highest IO Latency on any host was 1.5 ms during the Boot Storm. The chart clearly shows a very steady and very low level of IO Latency throughout the test run.

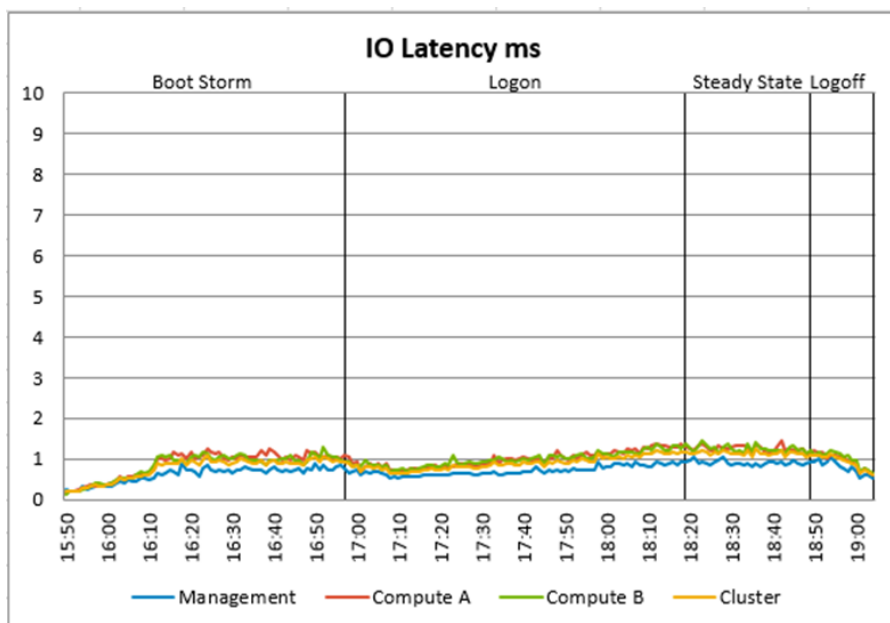


Figure 72 IO latency

The baseline performance of 771 indicated that the user experience for this test run was very good. The Index average reached 991, well below the VSIMax threshold of 1771. A total of 9 sessions failed to logon or become active, and there were no VSI errors.

Table 31 Performance metrics

Login VSI Baseline	VSI Index Average	VSIMax Reached	VSI Threshold
771	991	NO	1771

VSIMax v4

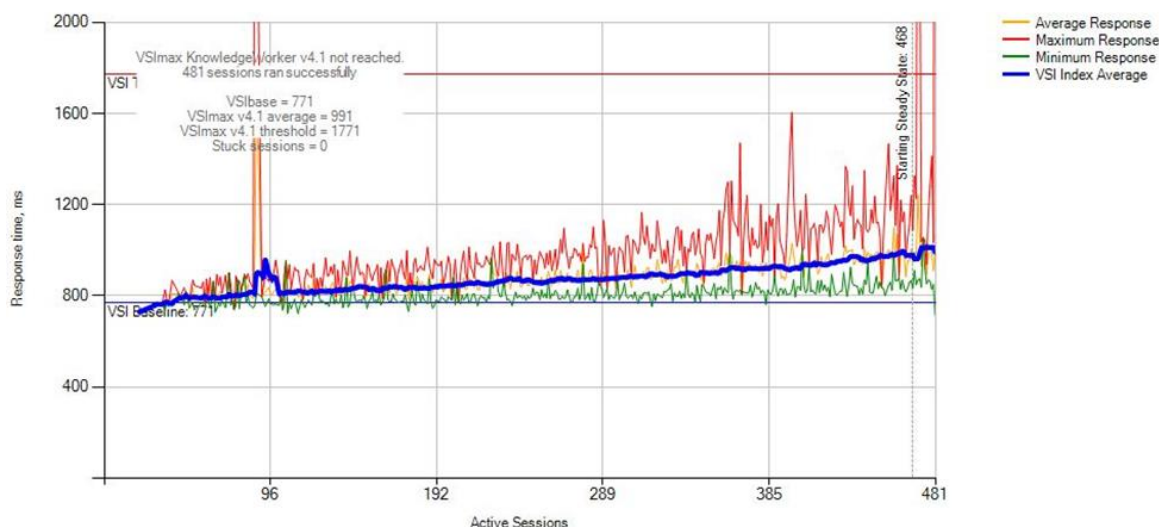


Figure 73 VSIMax

#### 6.4.1.6 Power Worker, 410 Total Users, Hyper-V 2012R2, XenDesktop 7.x MCS

In this workload test run each compute host had 145 user sessions while the designated management host had the full set of management VMs plus 120 desktops. The provisioning time was 37 minutes 44 seconds, not counting the time required to power on and register desktops with the Delivery Group.

The peak CPU Usage was 57percent on one host during Steady State phase, while the Steady State average was 44percent across all hosts. The relatively low steady state CPU usage was essential to avoid CPU scheduling problems that caused sessions to exceed their timeouts and fail.

The CVMs on each host averaged 5.9percent CPU usage during steady state. The management VMs used only 1.3percent CPU on the management host in steady state.

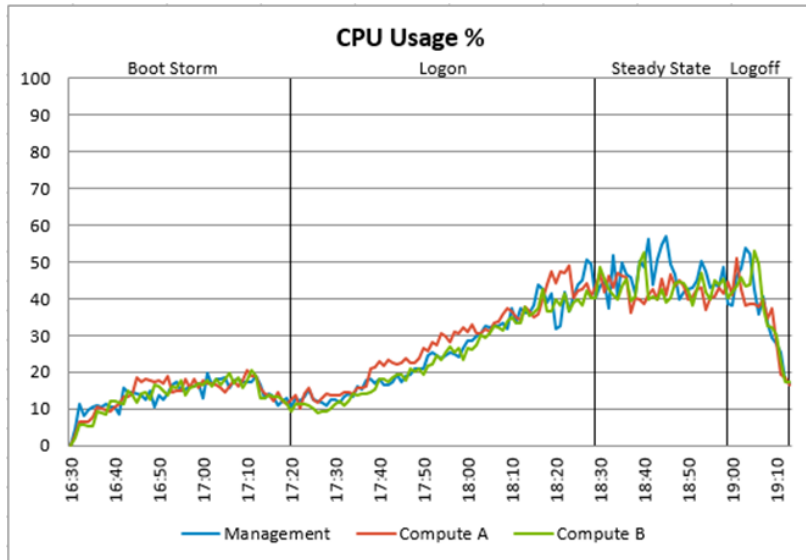


Figure 74 CPU usage

The memory consumption averaged 365 GB in steady state across on the Compute hosts, and 333 GB on the Management host. The peak usage on any host was 376 GB during the Steady State phase. Each desktop consumed 2.30 GB, after accounting for CVM and management VM memory consumption.

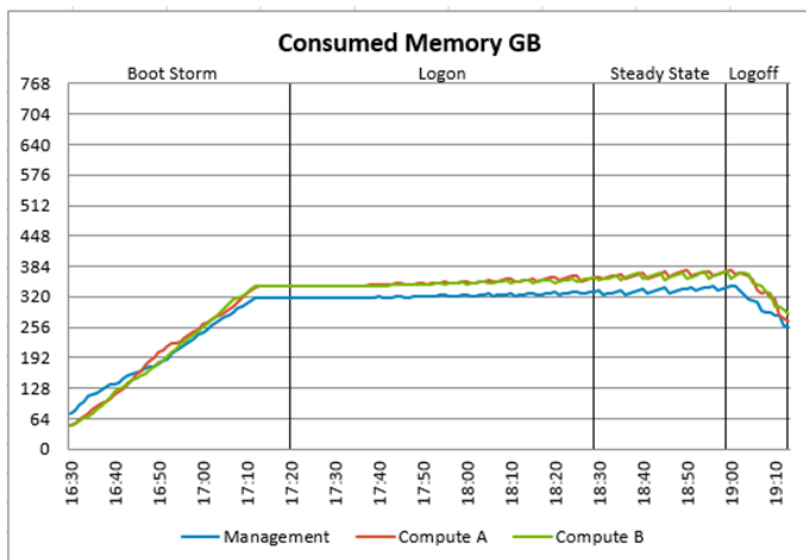


Figure 75 Consumed memory

Network usage peaked at 1571 Mbit per sec on one host during Logoff phase, and the average network usage for all hosts during Steady State was 1020 Mbit per sec. Each desktop accounted for 7.47 Mbit per sec in steady state.

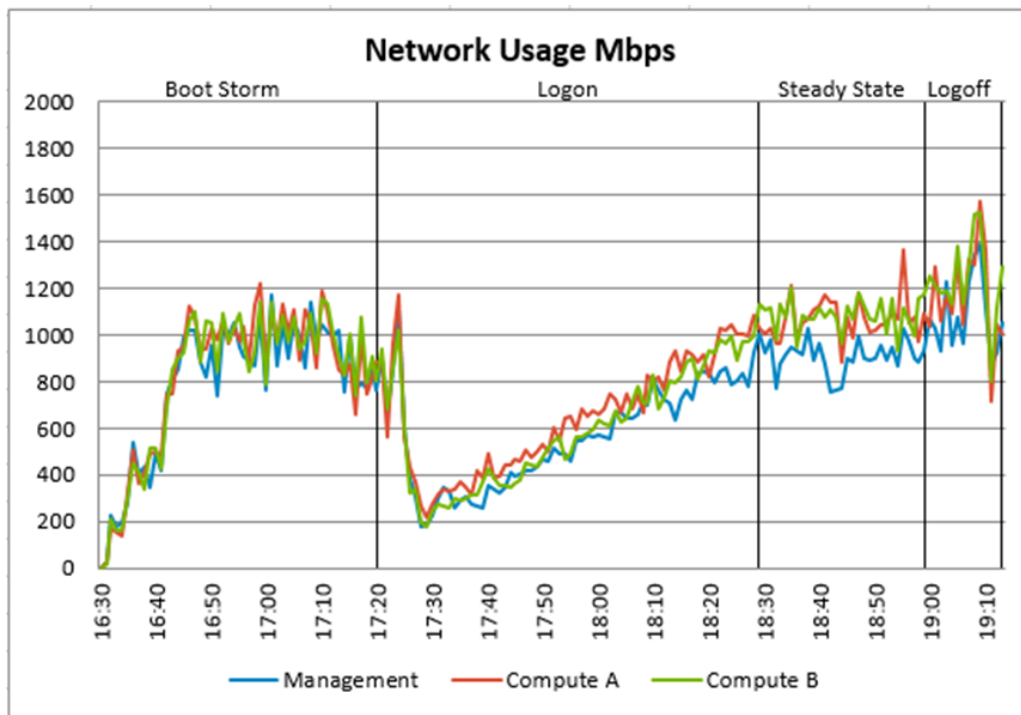


Figure 76 Network usage

The peak Cluster IOPS for the test run was 4066 IOPS during Logon phase, while the average in Steady State was 687 IOPS. Based on these numbers each user session generated 1.68 IOPS during Steady State.

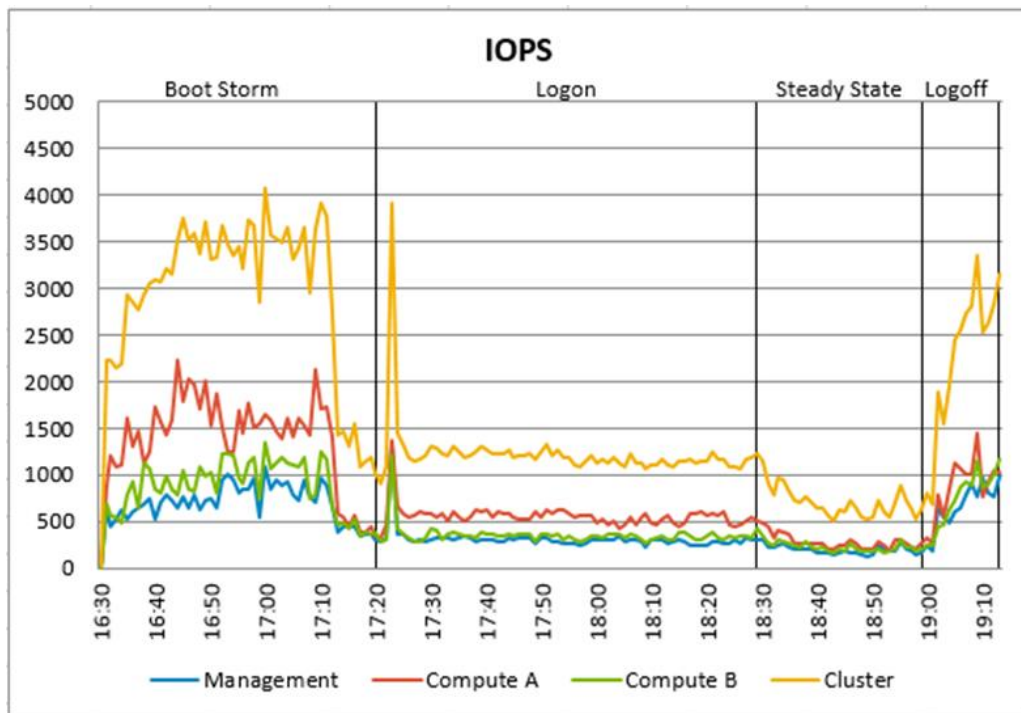


Figure 77 IOPS

The peak Cluster IO Latency was 1.2 ms at the end of Steady State phase, while the peak on any host was 1.3 ms. The average Cluster IO latency during steady state was 1.1 ms. The chart clearly shows a very steady and very low level of IO Latency throughout the test run.

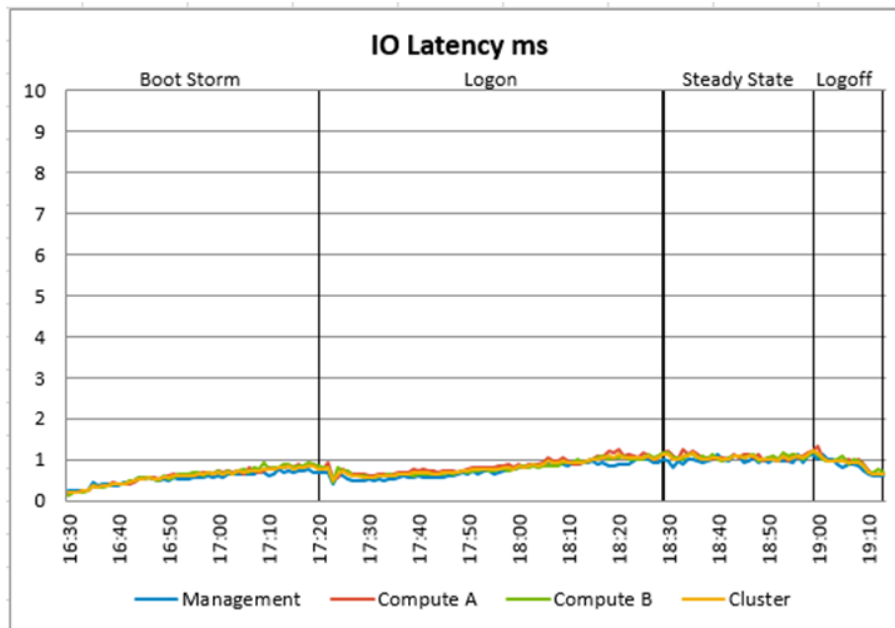


Figure 78 IO latency

The baseline performance of 773 indicated that the user experience for this test run was very good. The Index average reached 975, well below the threshold of 1773. Only 2 sessions failed to logon or become active and there were no VSI Errors.

Table 32 Performance metrics

Login VSI Baseline	VSI Index Average	VSIMax Reached	VSI Threshold
773	975	NO	1773

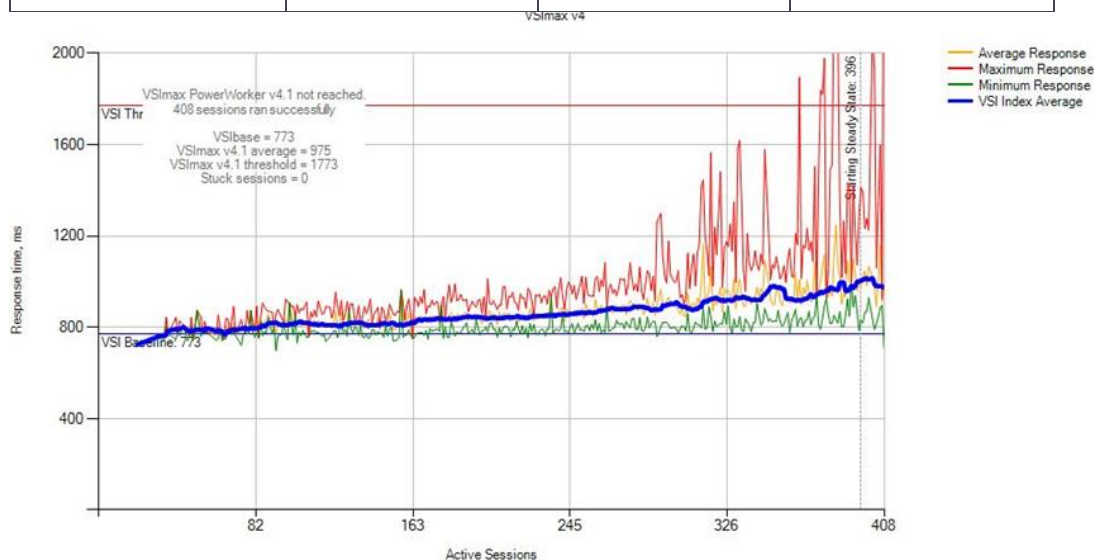


Figure 79 VSImax

## 6.5 vGPU test results and analysis

All test results graphs include the performance of the platform during the deletion and recreation of the pooled virtual machines after all users log off when the test run has completed. The different phases of the test cycle are displayed in the test results graphs later in this document as 'Boot Storm', 'Logon', 'Steady State' and 'Logoff'.

We tested three scenarios for graphics acceleration on Windows10 VMs: vGPU only using M60-1Q profile, and 48 standard non-vGPU VMs to compare system performance against the accelerated variants. Please note that all scenarios consist of the minimum three-node cluster with one mgmt. node and only one compute node active for these tests. Since GPUs were only added to a single host, we performed all relevant testing against this one host. Hence you will see the 2nd compute node marked at "unused" in some of the graphs that follow.

Table 33 Test results for workloads and configurations.

Hyper-visor	Provisioning	Login VSI Workload	Density Per Host	Avg' CPU %	Avg' GPU %	Avg' Memory Consumed GB	Avg' Memory Active GB	Avg' IOPS/User	Avg' Net Mbps/User
ESXi	Citrix MCS	Power Worker	48 vGPU VMs	38%	11%	239 GB	234 GB	9.1	4.9 Mbps
ESXi	Citrix MCS	Power Worker	48 Std	32%	-	190 GB	86 GB	7.9	4.7 Mbps

**CPU Usage.** The figure shown in the table, 'Avg' CPU percent', is the combined average CPU usage of all compute hosts over the steady state period.

**GPU Usage.** The figure shown in the table, 'Avg' GPU percent', is the average GPU usage of all hosts containing GPU cards over the steady state period.

**Consumed Memory.** Consumed memory is the amount of host physical memory consumed by a virtual machine, host, or cluster. The figure 'Avg' Memory Consumed GB' in the table is the average consumed memory across all compute hosts over the steady state period.

**Active Memory.** Active Memory is the amount of memory that is actively used, as estimated by VMkernel based on recently touched memory pages. The figure 'Avg' Memory Active GB' in the table is the average amount of guest "physical" memory actively used across the compute (and or management) hosts over the steady state period.

**Disk IOPS.** Disk IOPS are calculated from the Cluster Disk IOPS steady state average divided by the number of users to produce the 'IOPS / User' figure.

**Network Usage.** Network Usage per User. The figure shown in the table 'Avg' Net Mbps/User' is the Network Usage average of the hosts over the steady state period divided by the number of users per host in Megabits per second.

CPU usage for ESXi hosts is adjusted to account for the fact that on the latest Intel series processors, the ESXi host CPU metrics will exceed the rated 100percent for the host if Turbo Boost is enabled (by default). An additional 35percent of CPU is available from the Turbo Boost feature when all cores are active, but this



additional CPU headroom is not reflected in the VMware vSphere metrics where the performance data is gathered from. As a result, a line indicating the potential performance headroom provided by Turbo boost is included in each CPU graph.

Without the inclusion of the turbo there is a total of 80,000 MHz available for Desktops, with Turbo boost the total available MHz value is 108,000 MHz.

The user virtual machines were created using Citrix XenDesktop Machine Creation Services (MCS). The virtual machine desktops used local logon profiles and each user was assigned the same desktop for all logons. The vGPU enabled VMs used Windows 10 Enterprise and had the NVIDIA GRID drivers for Windows 10 installed and aligned with the Login VSI 4.1 virtual machine configuration. Office 2016 was used as the office suite and each virtual machine's virtual disk sized at 60 GB.

Table 34 Workload configuration of the user virtual machines

User Workload Profile	vCPUs	Memory GB	Reserved Memory	vGPU Profile	OS Bit Level	HD Size GB	Screen Resolution
Power Worker	2	4GB	4 GB	M60-1Q	Windows 10 64 Bit	60 GB	1920 x 1080

### 6.5.1 XC740xd-C7 with Tesla M60

Refer to the section 3 for hardware configuration details. GPUs can only be added to the 24-disk variant of the XC740xd.

Table 35 XC Series hardware configuration for vGPU

Enterprise Platform	CPU	Memory	RAID Ctlr	Drive Config	Network	GPU
XC740xd-24	6138 Gold (20- Core 2.0 GHz)	768GB @2666 MT/s	Dell HBA 330 Adapter	<ul style="list-style-type: none"> <li>2 x 120GB M.2</li> <li>2x 960 GB SSD 4x 1.8TB HDD</li> </ul>	Intel 10GbE 4P X710 rNDC	3 X NVIDIA M60 GPU cards installed in one host.

Compute and Management resources were split out with the following configuration across a three-node cluster and all test runs were completed with this configuration.

- Node 1 –XC740xd-24 – Dedicated Management.
- Node 2 – XC740xd-24– Dedicated Compute (Unused for hosting VMs for this testing).
- Node 3 – XC740xd-24– Dedicated Compute with 3 X M60 GPU Cards installed.

10GB networking was used for all PAAC testing. The host containing the M60 GPU cards had the appropriate ESXi drivers for this card installed.

#### 6.5.1.1 Power Worker, 48 vGPU users, ESXi 6.5, XenDesktop 7.15 MCS

The GPU enabled Compute Host was populated with 48 vGPU enabled virtual machines and used the NVIDIA M60-1Q profile. With all user virtual machines powered on and before starting test, the CPU usage was approximately 11percent.

The below graph shows the performance data for 48 user sessions on the Management and GPU enabled Compute hosts. The CPU reaches a steady state average of 38percent during the test cycle when all users are logged on to the GPU enabled Compute host.

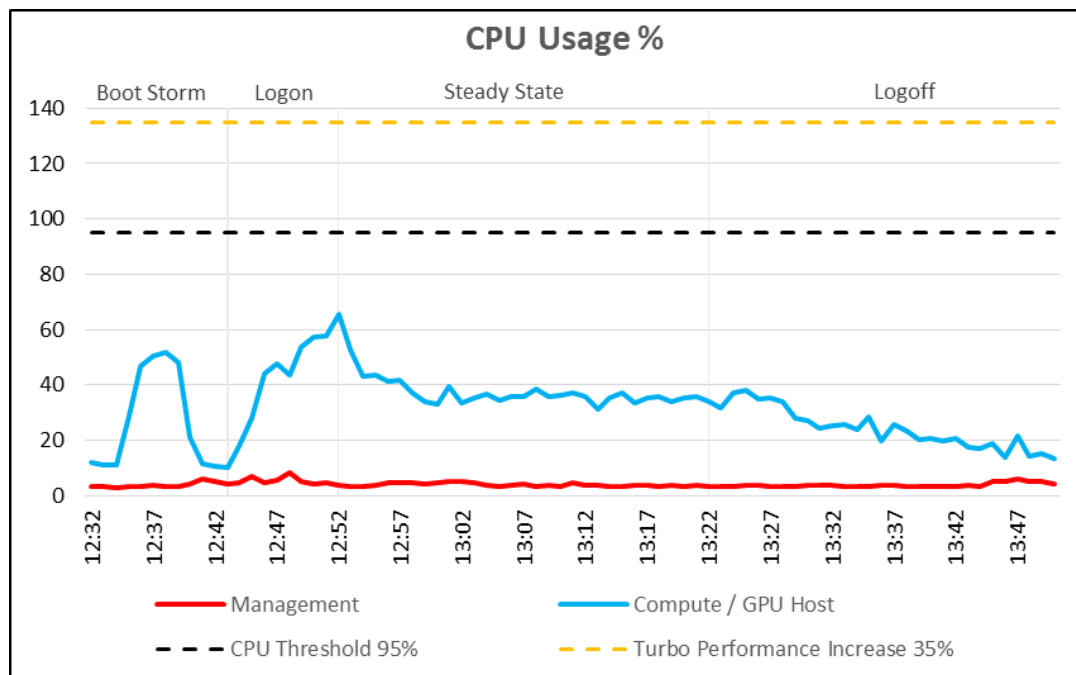


Figure 80 CPU usage

The GPU Metrics were gathered from the vSphere client and the GPU Profiler application were run on a test session to determine the framebuffer and vGPU usage. The GPU usage during the steady state period averaged approximately 11percent and reached a peak usage of 30percent with the Power worker workload. Framebuffer in use during the run averaged ~60% based on the sampling from one of the 48 VMs under load.

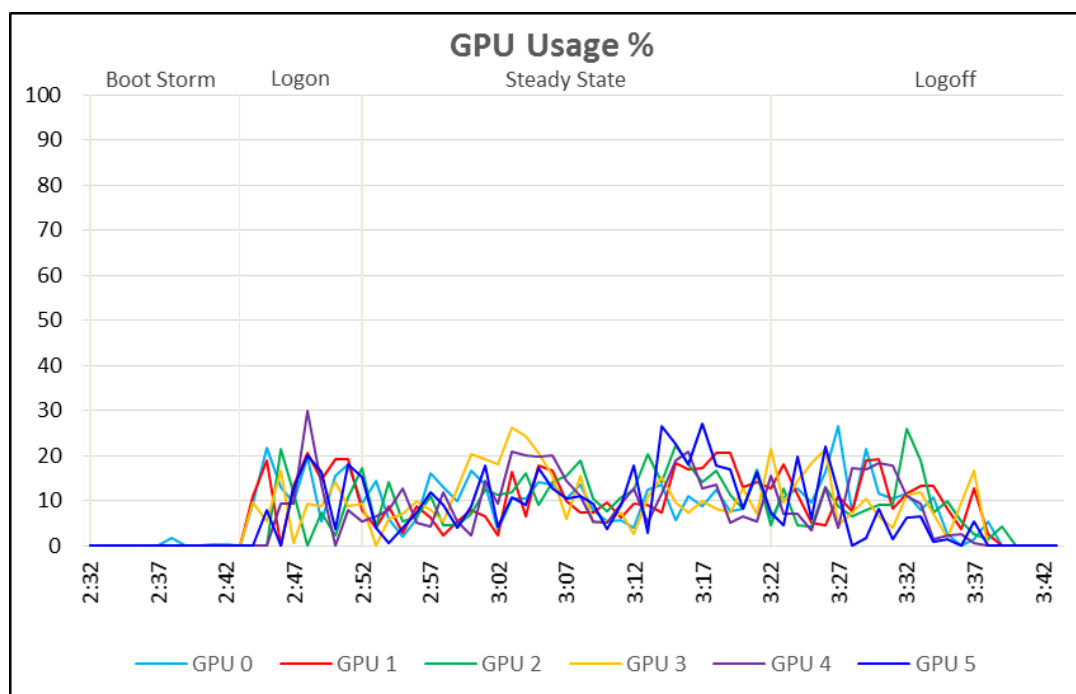


Figure 81 GPU usage



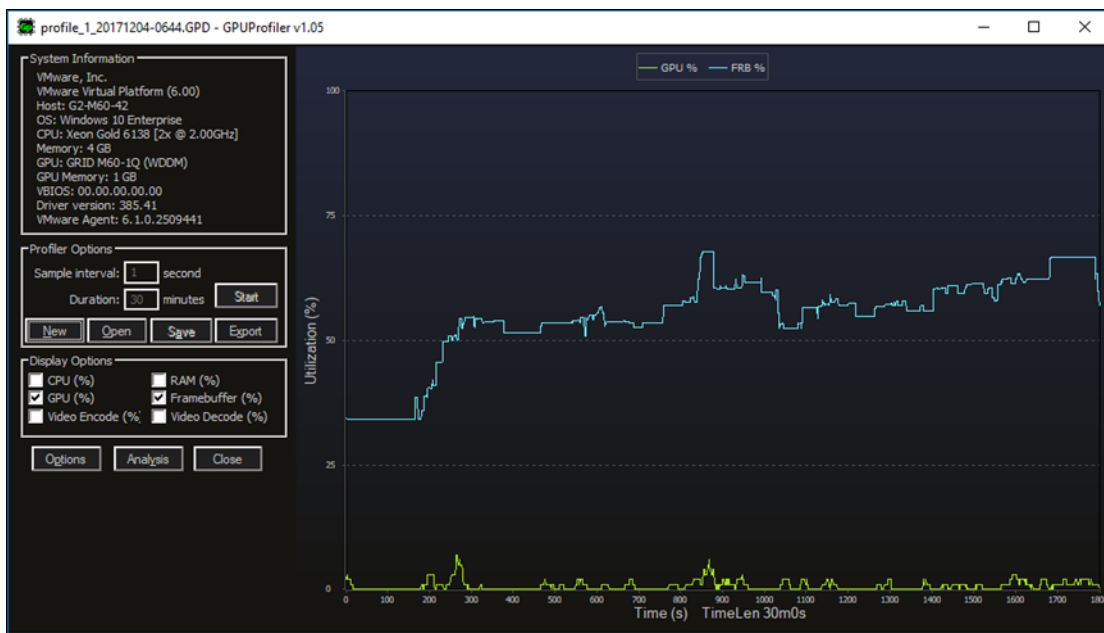


Figure 82 GPU and frame buffer usage

In regards to memory consumption for this test run there were no constraints on the Management or GPU enabled Compute hosts. Of a total of 768 GB available memory per node, the GPU Compute host reached a maximum memory consumption of 239 GB with active memory usage reaching a max of 224 GB. There were no variations in memory usage throughout the test as all vGPU enabled VM memory was reserved. There was no memory ballooning or swapping on either host.

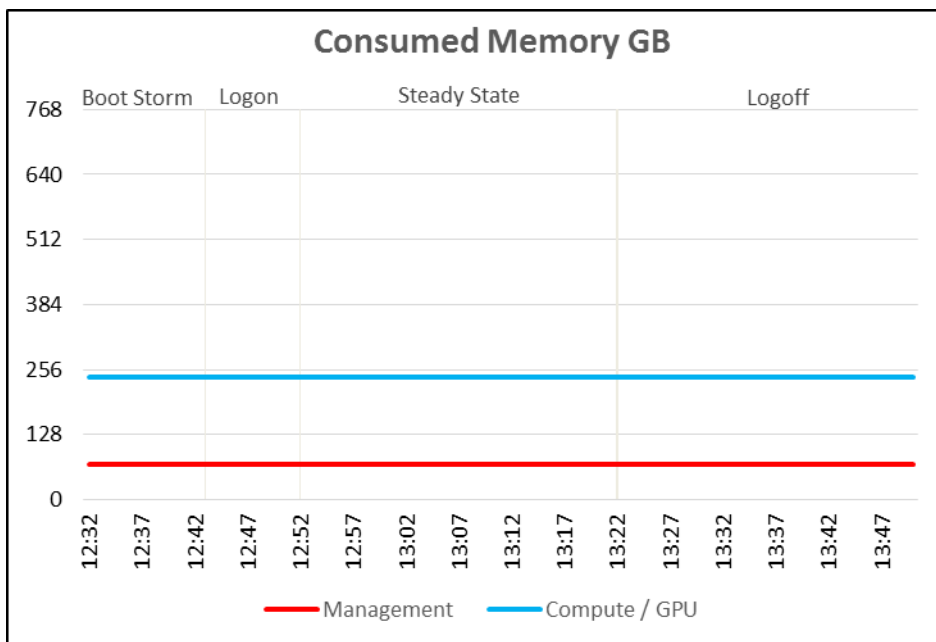


Figure 83 Consumed memory

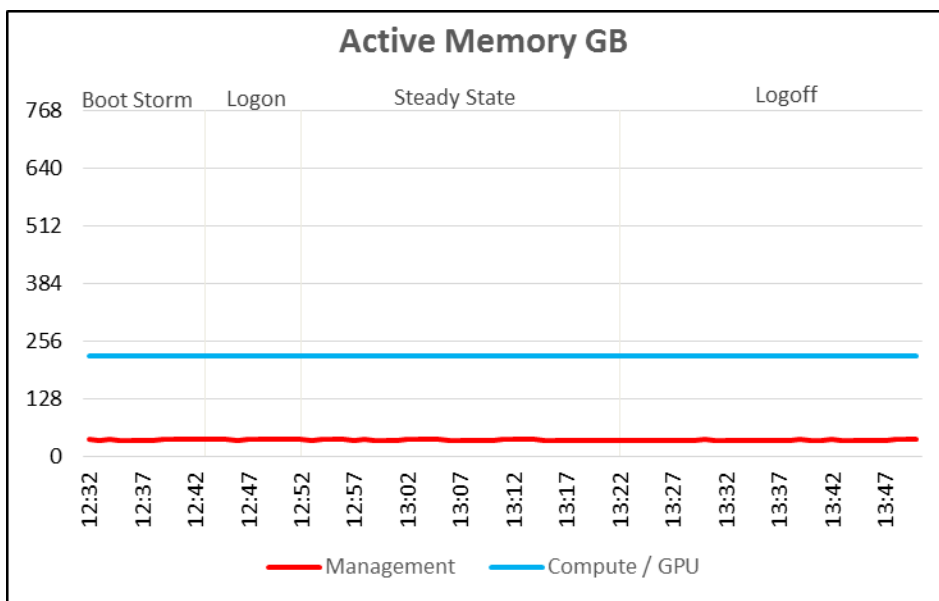


Figure 84 Active memory

Network bandwidth is not an issue on this test run with a steady state peak of approximately 311 Mbps. The busiest period for network traffic was during the Boot Storm phase during the reboot of all the VMs before testing started. The Compute / GPU hosts reached a peak of 936 Mbps during the Boot Storm.

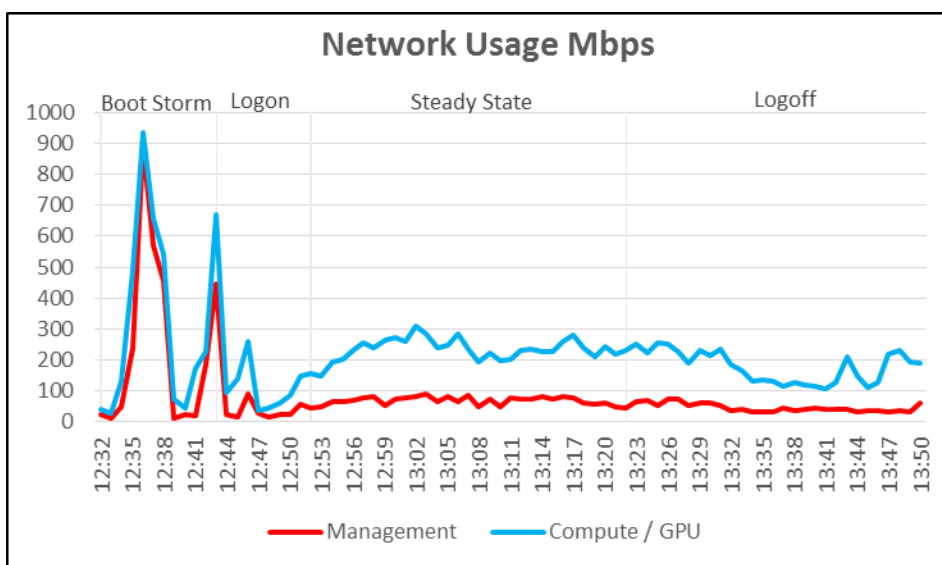


Figure 85 Network usage

The IOPS graphs and IOPS numbers are taken from the Nutanix Prism web console and they clearly display the boot storm, the initial logon of the desktops then the steady state and finally the logoff phase. The graphs show IOPS data for the individual hosts in the cluster and for the cluster as a whole.

The cluster reached a maximum of 15,308 Disk IOPS during the reboot of all the VMs before test start and 1,268 IOPS at the start of steady state. The Compute / GPU host reached a peak of 11,525 Disk IOPS during the reboot of all the VMs and 1,469 at the start of steady state.

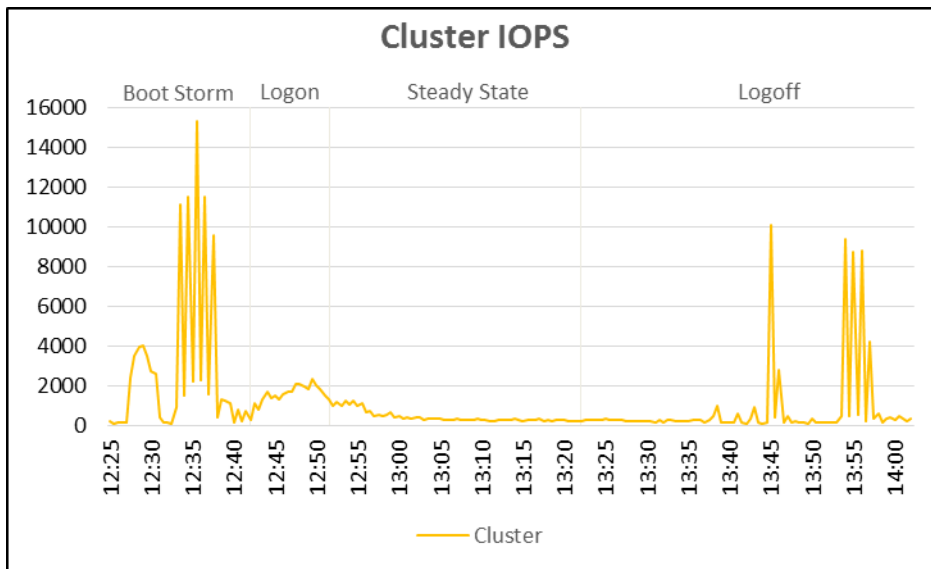


Figure 86 Cluster IOPS

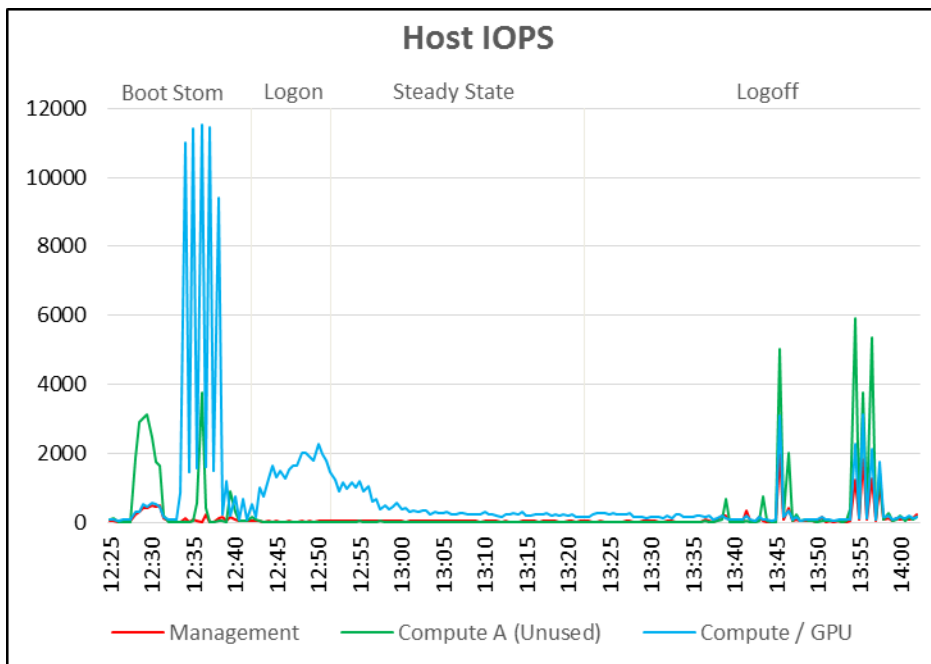


Figure 87 Host IOPS

The Login VSI Max user experience score shown below for this test was not reached. When manually interacting with the sessions during steady state the mouse and window movement was responsive and video playback was good.

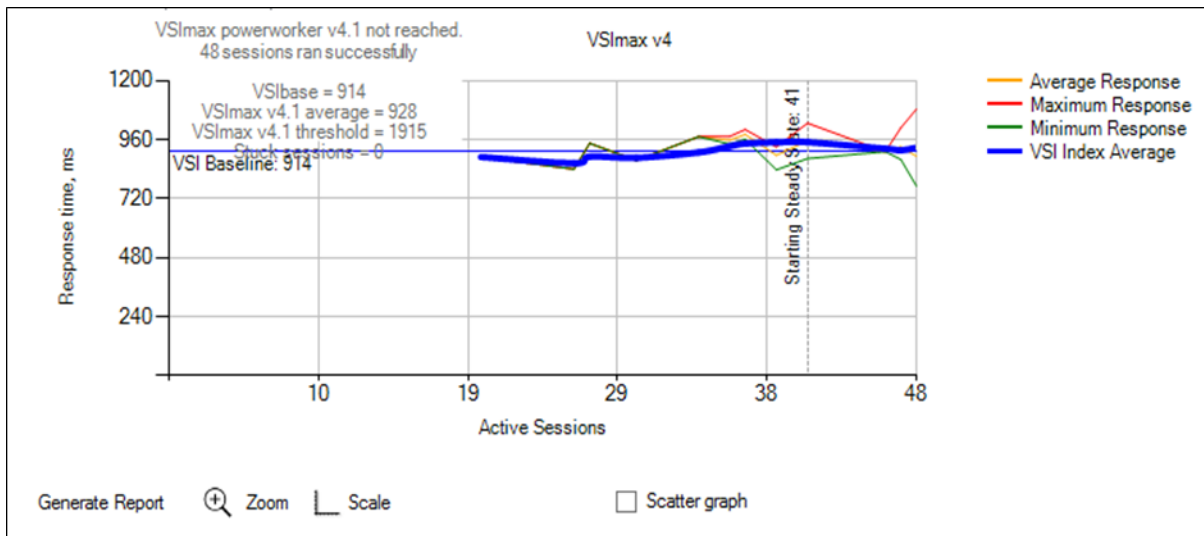


Figure 88 VSImax

Notes:

- As indicated above, the CPU graphs do not take into account the extra 35percent of CPU resources available through the Intel Xenon Gold 6138 processors turbo feature.
- 256 MB of memory was allocated to the temporary disk cache setting ('Memory allocated to cache MB:' was ticked) which is configured during the creation of the machine catalog.
- The Disk Cache Size box was ticked during the creation of the machine catalog. This meant any temporary data is written directly to the temporary data cache disk, using a minimal amount of memory cache.
- There were no disk latency issues during testing.

### 6.5.1.2 Power Worker, 48 standard users (non-vGPU), ESXi 6.5, XenDesktop 7.15 MCS

The GPU enabled Compute Host was populated with 48 standard non-vGPU enabled virtual machines and run to compare the performance of 48 non-vGPU enabled VMs with 48 vGPU enabled VMs on the same hardware. With all user virtual machines powered on and before starting test, the Compute host's CPU usage was approximately 4percent.

The below graph shows the performance data for 48 user sessions on the Management and Compute hosts. The CPU reaches a steady state average of 32percent during the test cycle when all 48 users are logged on to the Compute host.

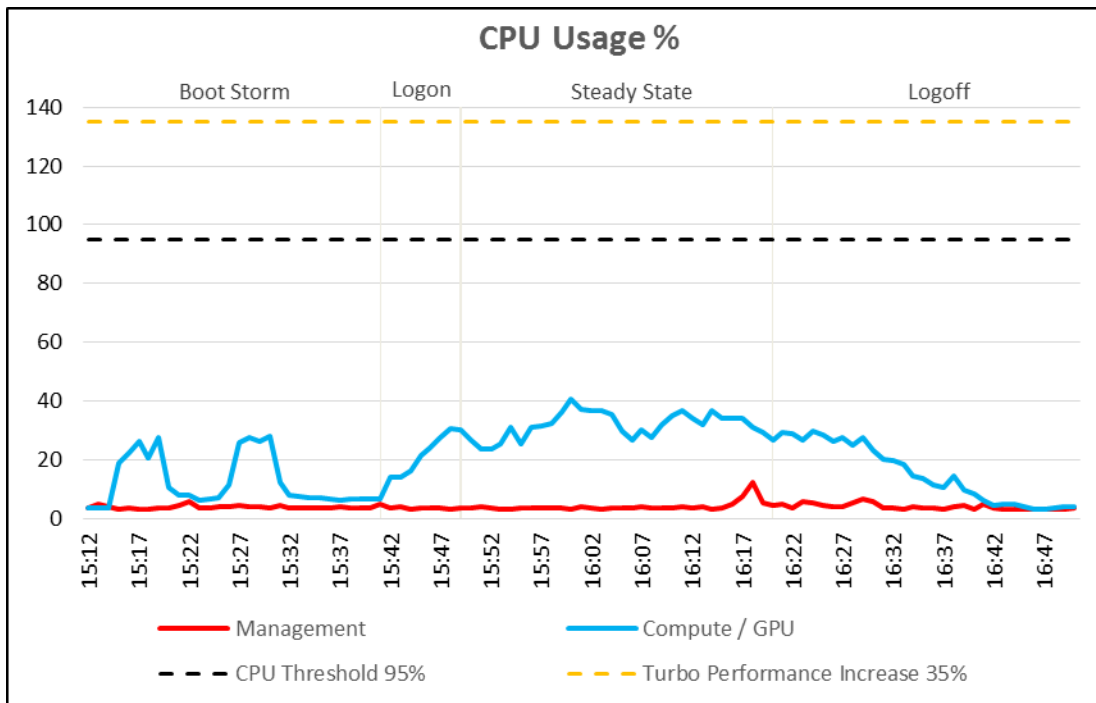


Figure 89 CPU usage

In regards to memory consumption for this test run there were no constraints on the Management or Compute hosts. Of a total of 768 GB available memory per node, the Compute host reached a maximum memory consumption of 211 GB with active memory usage reaching a max of 94 GB. There was no swapping or ballooning on either host.

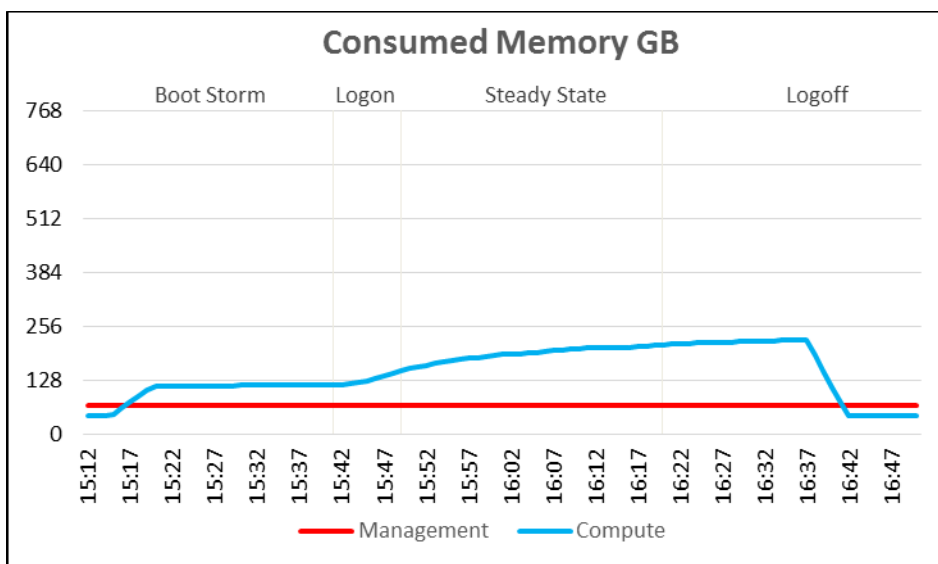


Figure 90 Consumed memory

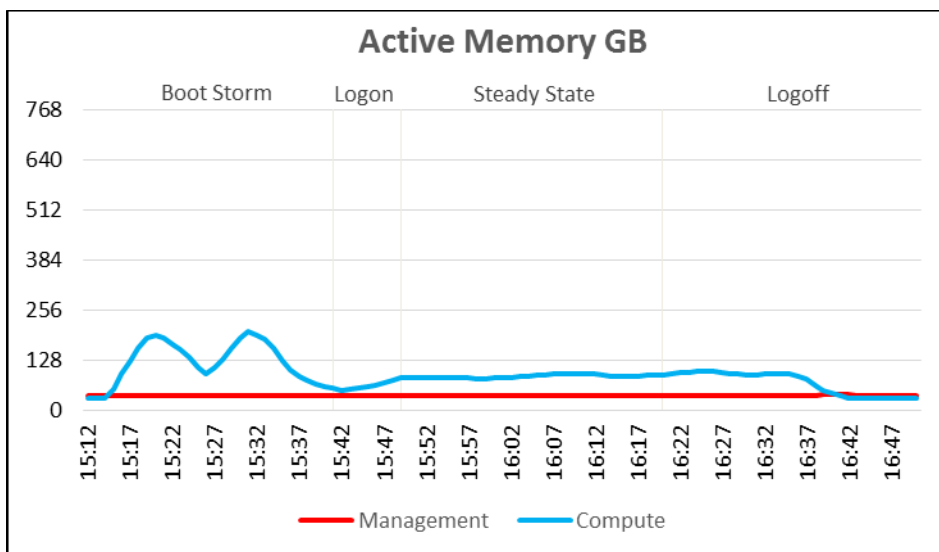


Figure 91 Active memory

Network bandwidth is not an issue on this test run with a steady state peak of approximately 292 Mbps. The busiest period for network traffic was during the Logoff phase. The Compute host reached a peak of 312 Mbps during this period.

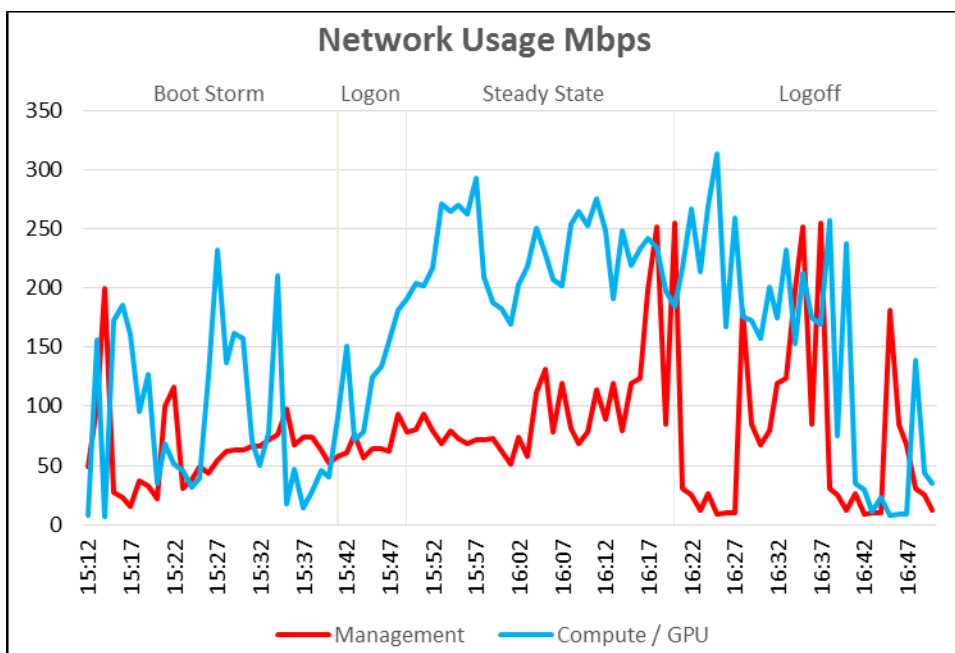


Figure 92 Network usage

The IOPS graphs and IOPS numbers are taken from the Nutanix Prism web console and they clearly display the boot storm, the initial logon of the desktops then the steady state and finally the logoff phase. The graphs show IOPS data for the individual hosts in the cluster and for the cluster as a whole.

The cluster reached a maximum of 7,379 Disk IOPS during the reboot of all the VMs before test start and 1,185 IOPS at the start of steady state. The Compute host reached a peak of 4,519 Disk IOPS during the reboot of all the VMs and 618 IOPS at the start of steady state.

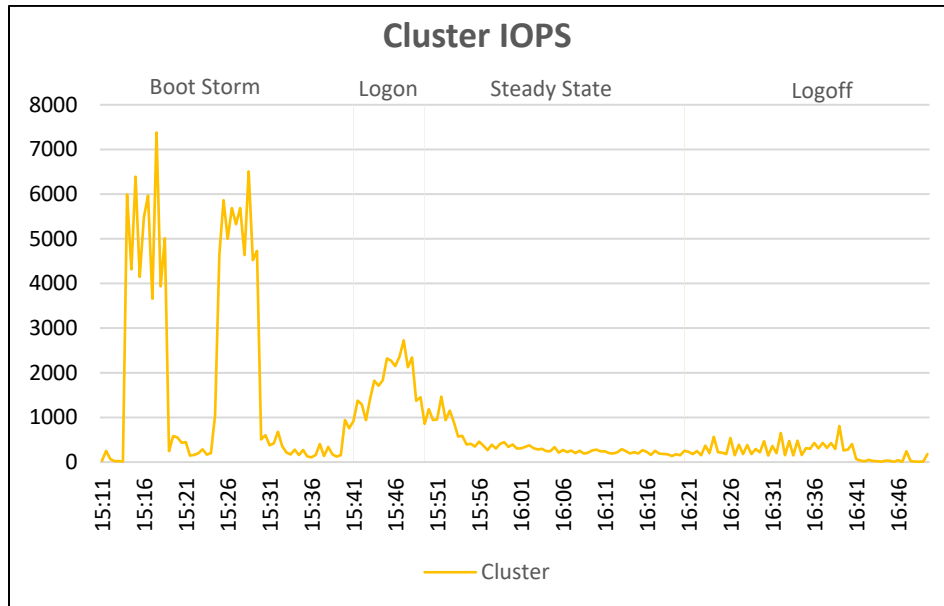


Figure 93 Cluster IOPS

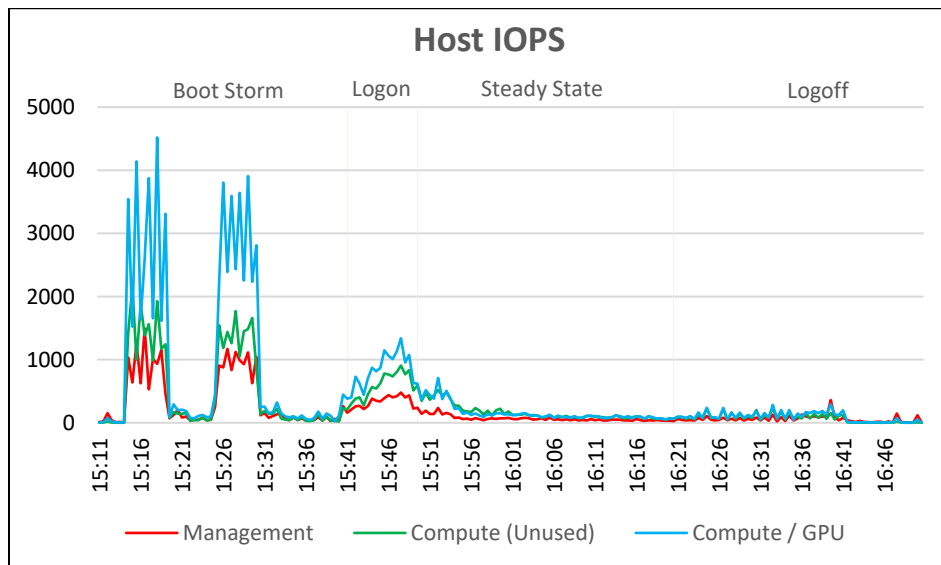


Figure 94 Host IOPS

The Login VSI Max user experience score shown below for this test was not reached. When manually interacting with the sessions during steady state the mouse and window movement was responsive and video playback was good.

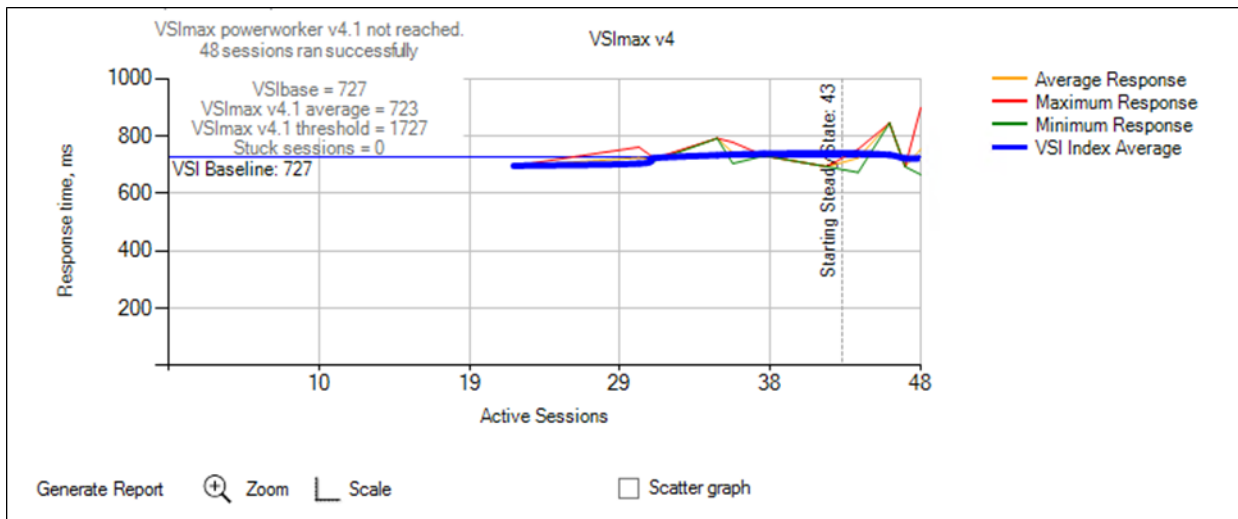


Figure 95 VSImax

Notes:

- As indicated above, the CPU graphs do not take into account the extra 35percent of CPU resources available through the Intel Xenon Gold 6138 processors turbo feature.
- 256 MB of memory was allocated to the temporary disk cache setting ('Memory allocated to cache MB:' was ticked) which is configured during the creation of the machine catalog.
- The Disk Cache Size box was ticked during the creation of the machine catalog. This meant any temporary data is written directly to the temporary data cache disk, using a minimal amount of memory cache.
- There were no disk latency issues during testing.



## A Related resources

See the following referenced or recommended resources:

- The [Dell EMC Cloud-Client Computing Solutions for Citrix Tech Center page](#) which includes this RA and other Citrix XenDesktop based RAs.
- Dell EMC Tech Center for XC Series:  
<http://en.community.dell.com/techcenter/storage/w/wiki/11454.dell-emc-xc-series-hyper-converged-solution>
- <http://www.dell.com/XC-Series-Solutions> for Dell EMC XC Series white papers.
- [www.Dell.com/xcseriesmanuals](http://www.Dell.com/xcseriesmanuals) for deployment guides (XC Xpress only), manuals, support info, tools, and videos.
- [Dell EMC XC Series Family Best Practices for Running Windows Server 2012 R2 with Hyper-V](#) white paper.