

NetApp Verified Architecture

# VMware End-User Computing with NetApp HCI and NVIDIA GPUs

NetApp Verified Architecture Design

Suresh Thoppay, NetApp January 2019 | NVA-1129-DESIGN | Version 1.0

#### Abstract

VMware End-User Computing with NetApp<sup>®</sup> HCI is a prevalidated, best-practice data center architecture for deploying graphics-intensive workloads at an enterprise scale. This document describes the architectural design and best practices for deploying the solution at production scale in a reliable and risk-free manner.



#### TABLE OF CONTENTS

1	Executive Summary	5
2	NetApp HCI and VMware End-User Computing Architecture	5
	2.1 NetApp HCI	6
	2.2 VMware End User Computing	7
	2.3 NetApp HCI and VMware End-User Computing Design Principles	9
3	Solution Overview	9
	3.1 Target Audience	9
	3.2 NetApp HCI Use Cases	10
	3.3 Solution Use Case Summary	10
4	Technology Overview	11
	4.1 NetApp HCI	11
	4.2 VMware vSphere	12
	4.3 Virtual Dedicated Graphics Acceleration (vDGA)	13
	4.4 Virtual Shared Graphics Acceleration	14
	4.5 Virtual Shared Pass-Through Graphics Acceleration with the NVIDIA vGPU	16
	4.6 VMware Horizon 7 Enterprise	20
	4.7 VMware Identity Manager	20
	4.8 VMware Horizon	21
	4.9 VMware ThinApp	25
	4.10 VMware App Volumes	25
	4.11 User Environment Manager	26
	4.12 Just-In-Time Management Platform	27
	4.13 VMware Mirage	28
	4.14 VMware NSX for vSphere	28
	4.15 VMware vRealize Operations Manager	29
	4.16 VMware vRealize Orchestrator Plug-In for Horizon	31
	4.17 VMware vRealize Log Insight	32
	4.18 VMware Workstation	34
	4.19 VMware Fusion	34
5	Solution Design	34
6	Technology Components	36
	6.1 Hardware Components	36
	6.2 Software Components	36

7	Solution Verification	. 37
8	Conclusion	. 37
Wh	ere to Find Additional Information	. 38
	NetApp	. 38
	NVIDIA	. 38
	VMware	. 38
Ve	rsion History	. 39

#### LIST OF TABLES

Table 1) vSphere maximums	13
Table 2: Horizon sizing	24
Table 3) Desktop VM configuration.	35
Table 4) Datastore estimates.	35
Table 5) Hardware components.	36
Table 6) Software requirements.	36

#### LIST OF FIGURES

Figure 1) VMware Horizon Cloud architecture	6
Figure 2) VMware Workspace One and NetApp HCI.	8
Figure 3) VMware Horizon JMP technology.	8
Figure 4) Solution architecture.	11
Figure 5) Common Element OS.	Error! Bookmark not defined.
Figure 6) SolidFire-to ONTAP disaster recovery solution	12
Figure 7) Enable DirectPath.	14
Figure 8) Pass-through PCI device for VMs.	14
Figure 9) Virtual Shared Graphics Acceleration.	15
Figure 10) gpuvm CLI	15
Figure 11) NVIDIA vGPU architecture	16
Figure 12) Host graphics settings	17
Figure 13) vGPU profiles for Tesla M10 (from NVIDIA documentation)	17
Figure 14) Homogenous virtual GPU profile combination	
Figure 15) nvidia-smi vgpu cli	
Figure 16) CUDA-Z screenshot.	19
Figure 17) Web launcher portal (from VMware documentation)	21
Figure 18) Provisioning Instant Clones	
Figure 19) Blast protocol features	Error! Bookmark not defined.
Figure 20) Unified Access Gateway logical diagram.	24
Figure 21) VMware App Volumes	25

Figure 22) AppStacks and Writable Volumes	26
Figure 23) UEM file shares	27
Figure 24) JMP technologies	28
Figure 25) NSX microsegmentation.	29
Figure 26) Five key features of VMware vRealize Operations Manager.	29
Figure 27) NetApp HCI vROPS integration.	30
Figure 28) NVIDIA vROPS integration.	31
Figure 29) GPU metrics	31
Figure 30) Horizon vRealize Orchestrator Plug-In architecture.	32
Figure 31) VMware vRealize Log Insight architecture.	33
Figure 32) Resource block scalability	34

# **1** Executive Summary

The right information at the right time becomes more crucial than ever for business success. Smartphones and tablets can retrieve information easily from the internet, but securely retrieving data from workspaces is always challenging. VMware Horizon virtual desktops and hosted apps along with VMware Workspace ONE have helped address this issue by providing secure access to virtual desktop environments or to specific applications connected to end devices.

Modern operating systems and other applications such as Microsoft Office have now become more graphics intensive, which adds CPU cycles to virtual desktops and applications and reduces user density on host servers. Having a graphics acceleration card on the server offloads graphics processing and rendering work from the CPU, creates a better user experience, and increases user density.

NetApp<sup>®</sup> HCI provides scale-out all-flash storage and guaranteed quality of service (QoS), which makes it easier to run virtual desktops and applications along with other workloads. With the NetApp HCI vCenter Plug-In, administrators can provision datastores, expand clusters, and perform other vital functions directly from vCenter. NetApp HCI makes it the ideal platform for consolidating your workloads.

# 2 NetApp HCI and VMware End-User Computing Architecture

This combination of technologies from NetApp and VMware allows customers to experience the benefits of the End-User Computing ecosystem. This NetApp Verified Architecture details the design decisions made to deploy VMware End-User Computing on NetApp HCI.

NetApp HCI is a hybrid cloud infrastructure solution capable of transforming and empowering organizations to move faster, promote operational efficiencies, and reduce costs. NetApp HCI is the foundation of the End-User Computing strategy, which can run multiple applications with the predictable performance that enterprises and customers demand. NetApp HCI enables the independent scaling of compute and storage resources so that systems are rightsized. With VMware Horizon, NetApp HCI can provision desktops for users in minutes along with the required applications, eliminating the complex management of traditional architectures. Integration into the NetApp Data Fabric means that you can easily integrate your infrastructure with the cloud by using the required data services.

NetApp HCI frees you from the limitations of current infrastructure solutions that are complex, that cannot consolidate all workloads, that force customers to scale in ways that strand resources, and that throttle the performance required by next-generation applications. With VMware Horizon and NetApp HCI, customers can quickly deploy virtual desktops and applications on an infrastructure that can be deployed quickly and scaled as needs change. Figure 1 illustrates how VMware Horizon running on-premises with NetApp HCI can integrate with cloud providers like AWS, Azure, and so on using the VMware Horizon Cloud service.

#### Figure 1) Horizon Cloud architecture.



## 2.1 NetApp HCI

NetApp HCI offers various benefits to consumers seeking a hybrid cloud infrastructure by combining industry best practices and the VMware vSphere Hypervisor. NetApp HCI delivers features and capabilities that first-generation HCI vendors could not. NetApp HCI is predictable, flexible, and scalable; provides simple administration and deployment; and is integrated into the NetApp Data Fabric.

#### **Predictable**

One of the biggest challenges for anyone managing infrastructure is delivering predictable performance, especially in the face of proliferating applications and workloads. Dedicated platforms and massive overprovisioning are not economically viable. However, when multiple applications share infrastructure, one application might interfere with the performance of another. NetApp HCI alleviates this concern with QoS limits available natively with NetApp Element<sup>®</sup> software. Element enables the granular control of each applications and volume, eliminates noisy neighbors, and satisfies all performance SLAs. All applications can be deployed on a shared platform, predictably and with confidence. The multitenancy capabilities of NetApp HCI can help eliminate more than 90% of traditional performance-related problems.

#### **Flexible and Scalable**

Hyper converged infrastructures have required fixed resource ratios, limiting deployments to configurations between four and eight nodes. NetApp HCI, however, scales compute and storage resources independently. Independent scaling avoids costly and inefficient overprovisioning, eliminates the 10% to 30% "HCI tax" from controller VM overhead, and simplifies capacity and performance planning.

With NetApp HCI, licensing costs are reduced. NetApp HCI is available in mixtures of small, medium, and large storage and compute configurations. The architectural design choices enable customers to confidently scale on their terms, making HCI viable for core data center applications and platforms.

No data center scales linearly, because business needs change constantly, and each application has different infrastructure requirements. NetApp HCI enables independent scaling of compute and storage resources, allowing on-demand scaling, avoiding costly and inefficient overprovisioning, and simplifying capacity and performance planning.

NetApp HCI is architected in building blocks at either the chassis or the node level. Each chassis can hold four nodes that are made up of storage nodes, compute nodes, or both. A minimum configuration is two chassis with six nodes, consisting of four storage nodes and two compute nodes. Two more blank spots can be used for expansion. Compute and storage nodes can be mixed if best practices are followed. Resources can be scaled nondisruptively through a simple GUI-driven process.

#### Simple

An imperative within the IT community is to automate all routine tasks, eliminating the risk of user error while freeing up resources to focus on more interesting, higher-value projects. NetApp HCI allows IT departments to become more agile and responsive by simplifying deployment and ongoing management.

The new NetApp Deployment Engine (NDE) eliminates most manual steps needed to deploy infrastructure, such as assigning names, network settings, and IP addresses, and provisioning ESXi hosts and VMware datastores. You can expect the infrastructure to be functional in less than 30 minutes.

The VMware vCenter Plug-In simplifies management in an intuitive way. Additionally, NetApp HCI uses a robust suite of APIs to promote integration into higher-level management, orchestration, backup, and disaster recovery tools.

#### **NetApp Data Fabric**

Traditional HCI platforms often involved the introduction of a silo of resources into an existing data center. Such platforms have little in common with other infrastructure-consumption choices that consumers might have made already or would like to make in the future. This approach is not efficient in the long term.

In contrast, NetApp HCI integrates into the NetApp Data Fabric for enhanced data portability, visibility, and protection of workloads, whether they reside on premises, in near-cloud storage, or in a public cloud. The NetApp Data Fabric removes lock-in and provides you with new choices. It allows the full potential of your data to be unleashed across cloud environments.

## 2.2 VMware End-User Computing

VMware Horizon enables IT organizations to provide virtual desktops and hosted applications to their users. As users start using smartphones, tablets, and other portable devices to access information, managing such devices becomes a challenge. VMware Workspace ONE provides unified management for all endpoints. It also provides single sign-on to hosted apps, desktops, intranet applications, and SaaS-based applications, as is depicted in Figure 2.

Figure 2) VMware Workspace ONE and NetApp HCI.



The VMware Horizon Just-In-Time Management Platform (JMP) allows customers to perform ultrafast desktop or Remote Desktop Session Host (RDSH) provisioning with VMware Instant Clones, real-time application delivery with VMware App Volumes, and contextual policy management with VMware User Environment Manager.

Figure 3) Horizon JMP technology.



Horizon Instant Clones, App Volumes, and User Environment Manager are all part of Horizon 7 Enterprise Edition, which also includes other components such as the following:

- vRealize Operations Manager for performance monitoring of desktop and application pools
- vRealize Log Insight for centralized log collection and analysis
- VMware NSX for vSphere for software-defined networking; securing desktops with microsegmentation; and edge services such as load balancers and the Dynamic Host Configuration Protocol (DHCP).

## 2.3 NetApp HCI and VMware End-User Computing Design Principles

NetApp HCI and VMware End-User Computing products provide an integrated system that offers all the benefits of VMware Horizon and the scalability and granularity of NetApp HCI. The underlying NetApp HCI system allows you to expand or resize a data center according to CPU, memory, storage capacity, and storage IOPS requirements.

NetApp HCI also lets you add and repurpose compute and storage nodes of various capacities to expand or contract any of the compute or storage parameters according to a data center's needs. This scaling is managed through vCenter and the NDE.

The NDE manages hardware configuration and deployment of the NetApp HCI environment, so compute and NetApp HCI storage nodes can be added or deleted easily in any configuration. To add compute nodes to a VMware cloud configuration, you simply add them to the vCenter data center and compute clusters. You add storage nodes to the NetApp HCI cluster and provision the datastores from the vCenter Plug-In. To add capacity and throughput, add more datastores to the desktop pools.

# **3 Solution Overview**

Media content usage has significantly increased recently with users watching training videos, using the latest office applications, or using image-editing tools. An increased media streaming workload is a challenge for a virtualization environment, because it adds load to the environment and degrades performance. If the environment is not virtualized, it can become difficult for administrators to provide security updates for workstations spread across the enterprise, because they can push updates only when mobile devices are connected to the network.

NetApp HCI with a graphics adapter allows users to stream media (even at 4K resolution) with a virtual desktop or in a hosted app environment. With VMware Horizon, customers can securely host their virtual desktop and apps in the data center, so users can use any supported device to access desktops or applications in the enterprise or from remote locations.

As you start designing graphics-intensive workloads on virtual desktops, you face several questions:

- What guest operating systems can I use?
- How can I provide high availability for the solution components?
- What versions of DirectX or OpenGL does it support?
- Does the environment support CUDA?
- How many users can I host with each server?
- How do I build a cost-effective solution to meet customer demands?
- What security features are available?
- How do I scale the infrastructure for workload demands?
- How do I manage the infrastructure?

This solution address all of these questions.

#### 3.1 Target Audience

The target audience for this technical report includes the following groups:

- Field consultants: to help with design decisions with the VMware End-User Computing environment
- Executives and sales engineers: to understand the value of the solution
- Professional services and IT managers: to understand and identify the components of the solution
- Partners: to learn and assist the customers who face similar challenges

## 3.2 NetApp HCI Use Cases

In addition to the previously mentioned benefits, NetApp HCI is ideal for the use cases described in this section. For customers performing the following tasks, NetApp HCI is architected to deliver exceptional value:

- Deploying private clouds
- Designing end-user computing environments that include hardware-accelerated graphics applications
- Considering workload consolidation

#### **Private Cloud**

NetApp HCI is an optimal foundation for an enterprise private cloud model, whether you choose OpenStack, VMware, or a solution developed in-house. This is because NetApp HCI uses native NetApp Element APIs that allow the on-demand provisioning of workloads through storage drivers and management plug-ins.

As an example, NetApp HCI integrates with VMware Virtual Volumes, enabling VMware administrators to achieve the most granular control over storage performance on a per-VM basis. With this functionality, you can set minimum, maximum, and burst IOPS levels, confirming exact amounts of capacity and performance for even the most sensitive virtual machines (VMs). You can change capacity and performance dynamically without migrating data or affecting system performance.

#### **End-User Computing**

NetApp HCI is optimal for end-user computing environments because capacity and performance are allocated independently for every virtual desktop and every application. The allocations can be easily adjusted as workloads shift or needs evolve without complexity. If an application needs more performance, the initial configuration can become a bottleneck, but NetApp HCI eliminates the penalty for underestimating requirements. You can modify the QoS policies to easily change the settings for minimum, maximum, and burst. The new settings take effect immediately.

#### **Workload Consolidation**

NetApp HCI eliminates workload silos, allowing customers to predictably run multiple applications on the same infrastructure. Traditionally, when multiple applications share infrastructure, all performance resources, both IOPS and bandwidth, are freely available to all applications all the time across the shared resources. Without a more precise resource allocation, one application or "noisy neighbor" can easily consume an unfair share of the resources, leaving little available for others. This "first-come, first-served" allocation methodology can have a huge negative effect on all the other applications on the system.

Performance expectations on an application-by-application basis are erratic and unpredictable. One misbehaving application can cripple the entire system. To keep these variances in check, customers must constantly monitor and manage which applications share resources. Often, alleviating resource contention requires migration of either the "noisy neighbor" or the unhappy customer to a new system.

The NetApp HCI QoS settings eliminate resource contention and the variable application performance caused by noisy neighbors. Each volume on the system is assigned its own minimum, maximum, and burst settings. Thus, performance becomes predictable for each application, but you avoid the capacity sprawl and low utilization that are common in today's modern infrastructures.

## 3.3 Solution Use Case Summary

This solution applies to the following use cases:

- Hardware-based graphics acceleration for multimedia workloads
- On-demand desktop and application deployment for end users

- Self-service of user profile management for end users
- Ease of management for security patches
- Integration with an existing private cloud
- Secure multitenant infrastructure for enterprises

## 4 Technology Overview

VMware Horizon 7 Enterprise Edition provides components for deploying and managing virtual desktops and hosted apps. To deploy virtual desktops for users in minutes, VMware introduced the JMP technology. JMP uses Horizon Instant Clones for quick virtual desktop provisioning, application assignments from App Volumes, and the association of virtual desktop policies managed by User Environment Manager.

NetApp HCI H610C includes the NVIDIA Tesla M10 graphics adapter, which helps accelerate media streaming for virtual desktops and apps hosted in VMware Horizon.

Figure 4 shows the logical diagram of the components of the VMware Horizon solution on NetApp HCI.



Figure 4) Solution architecture.

#### 4.1 NetApp HCI

NetApp HCI is scale-out architecture that permits the independent expansion of storage or compute capacity. It can scale up to 40 storage nodes and 64 compute nodes per cluster. The form factor of NetApp HCI varies by model. Mixed nodes (nodes of various NetApp HCI models) are supported in a cluster, thus protecting your investment. NetApp HCI provides APIs that provide deep ecosystem integration with various automation tools. The NetApp HCI H610C is a 2U compute node with two NVIDIA Tesla M10 graphics adapters, which accommodates a greater number of task and knowledge workers who use graphics-intensive media. An NVIDIA Tesla M10 graphics card is a dual PCIe slot device. Each board is equipped with the following components:

- Four mid-tier NVIDIA Maxwell GPUs
- 32GB of GDDR5 memory (8GB per GPU)

- 2560 NVIDIA CUDA cores (640 per GPU)
- 28 H.264 1080p30 streams

NetApp HCI provides the NDE to minimize the number of inputs and steps required to configure the system for a VMware vSphere environment. The NDE creates the storage cluster and the compute cluster, configures networks including iSCSI, and creates an initial datastore. Additional datastores can be provisioned and managed from the vCenter plug-in that NetApp HCI provides. NetApp HCI storage supports hardware acceleration (VAAI, or vSphere Storage APIs—Array Integration), and no further tuning is required for VMFS6 volumes.

Figure 5) NetApp SolidFire to NetApp ONTAP disaster recovery solution.

## SolidFire-to-ONTAP disaster recovery solution

New functionality available with ONTAP 9.4 and Element 10.3

- ONTAP LUNs can now be replicated to Element OS
- Create new LUNs while failed over to ONTAP and restore to HCI on failback
- Migrate volumes to NetApp HCI using SnapMirror
- Increased fan-in ratio of up to 32 HCI

clusters to one ONTAP cluster

- SnapMirror to ONTAP Select now supported
- ONTAP Cloud support planned for Element 11.0



NetApp HCI storage provides inline duplication, QoS, and RESTful APIs to integrate into any automation framework. NetApp HCI can perform synchronous and asynchronous replication between NetApp HCI systems or to NetApp ONTAP<sup>®</sup> systems on premises or in the cloud.

NetApp HCI uses 10/25Gb ports for the iSCSI, virtual machine, and vMotion traffic and 1Gb for the Intelligent Platform Management Interface (IPMI) and management ports. Although the NDE supports both virtual LAN (VLAN) untagged and tagged networks, it is easier to work with tagged networks. Configure any uplink ports used in vSphere distributed switches as trunk ports on the network switches, thus creating appropriate VLANs.

#### 4.2 VMware vSphere

Separate vSphere clusters are recommended for management, desktop pools, and application pools to provide logical separation and fault isolation between the components. NetApp recommends that you enable high availability and the Dynamic Resources Scheduler (DRS) on those vSphere clusters. High availability provides fault tolerance to host failures, which results in only a short outage before VMs are automatically brought back online.

Enable host monitoring and admission control so that at least one host failure or maintenance operation can be tolerated while sufficient resources are still provided to run the entire workload of the cluster. More capacity can be reserved to provide greater headroom for concurrent host failures or maintenance.

DRS automatically balances CPU and memory workloads across the cluster members. vSphere supports up to 64 nodes per cluster. VMware Horizon supports up to 32 nodes for linked clones and up to 64 nodes for the Instant Clones.

vCenter Server provides the control pane for the vSphere environment. Up to 15 vCenter instances can be joined in linked mode. The NDE deploys one vCenter Server, or it can join to any existing vCenter Server. The NetApp HCI vCenter Plug-In supports linked mode and can provision datastores for hosts within that vCenter data center. A vSphere host can exist in only one vCenter data center.

Item	Maximum	Description
VMs per host	1,024	For H610C nodes, with vGPU: 128.
Virtual disks per host	2,048	
VMFS volume size	64TB	
Volumes per host	1,024	
Hosts per volume	64	
Powered-on VMs per VMFS volume	2,048	
Maximum file size on VMFS5/VMFS6	62TB	
Shared GPUs per ESXi host	16	
Hosts per vCenter Server	2,000	
Linked vCenter Servers	15	
Hosts in linked vCenter Servers	5,000	
Powered-on VMs in linked vCenter Servers	50,000	
Latency between vCenter instances in linked mode	100ms	

Table 1) vSphere maximums.

Graphics adapters in vSphere hosts can be used in three ways:

- Virtual Dedicated Graphics Acceleration (vDGA)
- Virtual Shared Graphics Acceleration (vSGA)
- Virtual Shared Pass-Through Graphics Acceleration (with NVIDIA, referred to as "vGPU"; with AMD, referred to as "multiuser GPU" or "MxGPU")

#### 4.3 Virtual Dedicated Graphics Acceleration (vDGA)

VMs access the graphics adapter using a PCI pass-through device. You must enable DirectPath for that device on the vSphere host (see Figure 6). Using the pass-through device, the VMs cannot be suspended, be migrated with vMotion, or use VM snapshot features.

#### Figure 6) Enable DirectPath.

	DirectP	Al PCI De	Notes					-		
Storage	15	(Date)				0.1	Elter		9	- Clinit -
Starage Perdoan	0			( Marca )	This is a strength of the second	Canada Manta	ETTY Manual		Denos Name	
Delectores		· 10			a Intel Corporation	Sity Lake-F PCI				
Understores	10000		La 0000 10.00.0	Not Confourable	PLX Sectimology	PEX 6747 48-La		1		
Rolevel Computation			Law 0000.10.11.0	Not Configurable	PLX Technology	PEX 8747 48-La.				
PTOTOCOI Engpoints	1.0		P 0000 1D	00.0 Available (pend	NVIDIA Corporat	Testa M10				
TO Filters	1 month		18 0000.19.00.0	Not Configuratio	a PLX Technology	PEX 0747 48-La				
Networking			0000 1A	00.0 Unavailable	NVIDIA Corporat.	Testa M10				
Virtual switches	1 Descent		III 0000 19:10.0	Not Configurate	PEX Technology	PEX 0747 40-La.				
VMkernel adapters			2 DE 0000 1C	00.0 Available (pend	NVIDIA Corpora	Tesla M10				
Physical adapters TCP/IP configuration			)(III 0000:19:00 0	Not Configurati	PLX Technology	PEX 6747 40-La.				
Advanced	1.000	0000:10:0	0.00							
Virtual Machines		This devic	e is not available to M	ls. R will become available	after its host is reboote	d.				
VM Startup/Shutdown	I I come	Name	Tesla M1	0	Vendor Name	NVIDIA Corpora	ation			
Agent VM Settings		Device 10	D 1380		Vendor ID	10DE				
Swap file location	1 International	Subdevia	ce ID 1160		Subvendor ID	10DE				
Default VM Compatibility		Class ID	300							
System	1 months	Bus Locat	ton							
Licensing	1	ID:	0000.10	0.0	Slot	0				
Time Configuration		But	10		Function	NAN				
Authentication Services			105.1					_		
Certificate										
Power Management										
Advanced System Settings										
System Resource Reservation							OK Ca	ncel:		
Security Profile										
System Swan										
March Decking										

Figure 7) Pass-through PCI device for VMs.

VDI	DO VDGA01 - Edit Sett	ngs				0	10	0.00 Hz
Con Con	Vitual Hardware VM	Options SDRS Rules	VADD OD	lons				MENORY USADE
	· C CPU	4		0				ET BTORAGE LISAD
0	> Memory	8196		MB				(B) 64.00 GB
150	+ m Hard disk 1	64		08				
	• Q SCSI controller 0	LSI Logic SAS						
· UM Handware	+ Ketwork adapter 1	Default (H610	C VDS)			Connect		
	+ @ CD/D/D drive 1	Clent Device				Connect		
- Tags	+ Floppy drive 1	Clent Device				Connect.		
Assigned Say Collegers This last is proved	+ Mideo card	Specify custo	n settings					
	. ES USB xHCl control	use 3.0			-			
	SATA controller 0							
	+ 🥥 VMCI device							
	Other Devices							
	· New PCI device	0000 1c 00 0	NVIDIA Cor	poration 1				
	Physical POLIPCIe de	NCE 0000 1c 00 0	NVIDIA Cor	poration 1	iesta I	w10		
Notes		0000 1d 00 0	NVIDIA Con	poration 1	esia I	who plable when		
		suspend, r such virtus	rigrate with v ( machines.	Motion, o	r take	or restore snapshots of		
	New device		Device			Add	-	
vApp Details	Concellulary ESY 6.5	and later A.M. version 15						
· Update Manager Compliance			_	_	_	un carce		
Status -		VM Storage Policy Cor	nplance -					
3	Law Detailed Status	Last Checked Date						
				Chec	k Con	nplance		

A native graphics adapter driver must be installed on the VM, but no drivers are required on the hypervisor. The VM has access to the complete GPU and GPU memory even if it is idle or powered off. NetApp doesn't recommend using vDGA, because it doesn't support many vSphere features and it provides low resource utilization.

Partners might need to demonstrate the NVIDIA vGPU vMotion feature on a single server and use vSphere nested virtualization. NetApp hasn't tested this functionality, and you should know its limitations before you consider it. With NVIDIA, it requires a Quadro Virtual Data Center Workstation license or a GRID Virtual Apps license.

#### 4.4 Virtual Shared Graphics Acceleration (vSGA)

vSGA provides hardware 3D acceleration by sharing GPUs across multiple VMs. This solution is attractive for users who require a GPU's full potential during brief periods. The NVIDIA vGPU Manager or the GRID software driver is installed on hypervisor hosts, and the VMware 3D driver, which is part of

VMware Tools, is installed on the guest OS. Graphics API support is limited to vSGA, and only selective versions of DirectX and OpenGL are supported. There is no CUDA support.

Figure 8) vSGA.



To install the NVIDIA driver on multiple hypervisor hosts, VMware vSphere Update Manager can be used with the host extension type. You can monitor VM GPU association by using the vSphere web client or the gpuvm CLI command.

Figure 9) gpuvm CLI.

# VM GPU Info

gbrack Survay Marter       curdque Termisons Vile Datatives Helinets Unders Heinets Un	Using vS	Sphere We	eb Client /	gpuvr	n CLI				pid 108304, VM "H610C-LC088", REMERVed 131072M of CPU MEMORY. pid 108307, VM "H610C-LC022", reserved 131072M of GPU MEMORY. pid 11100, VM "H610C-LC038", remerved 131072M of GPU MEMORY. pid 139444, VM "H610C-LC038", remerved 131072M of GPU MEMORY.
Image: section of the section of th	Getting Started Summary Monitor	Configure Permissions V	Vis Datastores Networks Up	pdate Manager					pid 111259, VM "H610C-LC009", reserved 131072KB of GPU memory. pid 111254, VM "H610C-LC009", reserved 131072KB of GPU memory.
Harrer funder         (-1/2)           Vield Allgriss         (-1/2) <td< th=""><th>Virbai switches</th><th>Host Graphics Graphics</th><th>Devices</th><th></th><th></th><th></th><th></th><th></th><th>pid 111266, VM "H610C-LC040", reserved 131072KB of GPU memory. pid 111268, VM "H610C-LC033", reserved 131072KB of GPU memory.</th></td<>	Virbai switches	Host Graphics Graphics	Devices						pid 111266, VM "H610C-LC040", reserved 131072KB of GPU memory. pid 111268, VM "H610C-LC033", reserved 131072KB of GPU memory.
ymod Adams           Sind Cadgradina	Withornal extension	Grephics Devices							GPU memory left 7819552KB.
yndrafagang yndr	Discrimination and and and and	1						O FIN .	pid 109209 IN UNCLOCATION, Vacantee, GPU Maximum memory 030012041
nd comparison production pr	Physical adapters	Name	Mander	Anti-e Tone		Conference Tree		the second secon	pid 109645, VM "H610C-LC020", reserved 131072KB of GPU memory.
HateCd       NotO-Note be 1000       NotO-Note be	ICPAP conguration	NO (DIATARIA MIS)	NU(D)). Comparison	Ghand		Ghand		7.00.00	pid 109644, VH "H610C-LC027", reserved 131072KB of GPU memory.
thal Mondos thal Mondos thal Mondos thal Mondos that We handling we handling	Advanced	N/DiaTeria M10	N/DIA Conservation	Shared		Shared		7.00.00	pid 109648, VM "H610C-LC036", reserved 131072KB of GPU memory.
4 Setus Statement       1 Setus       1 Se	<ul> <li>Virtual Machines</li> </ul>	NOTDIATASIS MID	NODIA Corporation	Shared		Charled		7.00.00	pid 111120, VH "H610C-LC063", reserved 131072KB of GPU memory.
part M standard make W Chaves B 10 motor Consider a based to part of the standard of the sta	VM Startup/Shutdown	NUCLEARING MAD	Multilla Constantion	Shared		Etward		7.00.00	pid 111269, VH "H610C-LC055", reserved 131072MB of GPO memory.
wm/b         Model         Same         Same <t< td=""><td>Agent VM Settings</td><td>NUTUATION MIL</td><td>NUTLA Corporation</td><td>Strangel</td><td></td><td>Strand</td><td></td><td>7 00 100</td><td>pid 111333, VM "Heloc-LCO28", reserved 131072KB of GPU memory.</td></t<>	Agent VM Settings	NUTUATION MIL	NUTLA Corporation	Strangel		Strand		7 00 100	pid 111333, VM "Heloc-LCO28", reserved 131072KB of GPU memory.
data VM Constants       made       ma	Swap file location	NOTICIA AND	Multiple Corporation	Stand		Stared		7 50 03	GPU mamory loft 7310552KB.
wei         income         income <td>Default VM Competibility</td> <td>NOTIONAL MAD</td> <td>NUCLE Conservation</td> <td>Streeted</td> <td></td> <td>Shand</td> <td></td> <td>7.00.00</td> <td>Xserver unix:2, FCI ID 0:28:0:0, vSGA mode, GPU maximum memory 8368128KB</td>	Default VM Competibility	NOTIONAL MAD	NUCLE Conservation	Streeted		Shand		7.00.00	Xserver unix:2, FCI ID 0:28:0:0, vSGA mode, GPU maximum memory 8368128KB
control       minute	- System	NOTION AND AND	NORTH Constant	Obarad		Chanad		7.00.00	pid 111298, VH "H610C-LC015", reserved 131072KB of GPU memory.
mere menery merery mer	Licensing	INVIDIOUSIA IN ID	NYIDIA COrporation	DIMAD		Director		7.90 00	pid 108309, VM "H610C-LC023", reserved 131072KB of GPU memory.
Mar Comparison Mar Compariso	Electrony and a	ML (o Fint	•					Stirms ExExport - Fo Copy -	pid 112527, VH "H610C-LC041", reserved 131072KB of GPU memory.
MondControlster Hinding Hans control eff the grade dates AND Controls dates AND Controls dates and the And Control Handley and Handley	time computation								pid 112530, VH "H610C-LC052", reserved 131072KB of GPU memory.
Image: Marging: Imarge: Marging: Image: Marging: Image: Marging	Authentication Services	Who associated with the gra	phics device "NVDIATesite M10"						pid 111322, VM "H610C-LC024", reserved 131072KB of GPU memory.
New Management Name Status Name Status Status Status Name Status Name Status Na	Certificate		and and and						pid 121015 VN "Heide-Leose", reserved 151072XB of GPU memory
Name         Total	Power Management	87 / <b>-</b> 63 (A)	Savenous +					12 (Q.PHDF *)	pid 112547, VM "H610C-LC045", reserved 131072KB of GPU memory.
Autore Hancards Harvitan         Value         Value <thvalue< th="">         Value         Valu</thvalue<>	Advanced System Settings	New received and the second se	1. 2667	89.0	Provisioned Space	Used Space	Hest CPU	Hasl Ven	GPU memory left 7319552KB.
Sang Yang Mag Sang Y	System Resource Reservation	City Hemoscians	Pawered Ch	O Normal	81.72 GB	10.20 023	20 /04/12	1,772 MB	Xserver unix:3, PCI ID 0:29:0:0, vSGA mode, GPU maximum memory 8868128K
Anto Sauge Anto Sau	Security Profile	Hetuc-Louzs	Powered On	<ul> <li>Normal</li> </ul>	86.26 GB	18.32 08	20 //842	1,749 MB	pid 100310, VH "H610C-LC001", reserved 131072KB of GPU memory.
sk fvdfer veder veder ververver ve	System Swap	H610G-LC024	Powered On	<ul> <li>Normal</li> </ul>	85.32 GB	18.38 (39	20 ////2	1,783 MB	pid 109647, VH "H610C-LC010", reserved 131072KB of CPU memory.
<pre>share share s</pre>	Host Profile	M0100-L0041	Powered On	o normal	00.70 (38	10.22 08	335 MH2	2,079 WB	pid 111262. VM "H610C-LC029", reserved 181072KB of GPU memory.
Normal State Control State	- Hardware	M0100-L0045	Powered On	<ul> <li>Normal</li> </ul>	86.23 GB	18.29 08	900 MH2	3,389 WB	pid 111258, VM "H610C-LC011", reserved 131072KB of GPU memory.
Alter Construction of the Construction of t	Brocestore	H0100-L0052	Powered On	<ul> <li>Normal</li> </ul>	88.24 GB	18.3 GB	330 MHz	2,861 MB	pid 111271, VH "H610C-LC005", reserved 131072KB of GPU memory.
<pre>minory</pre>	Flocessols	H010C-LC056	Powered On	<ul> <li>Normal</li> </ul>	88.19 GB	18.25 GB	377 MHz	2,834 MB	pid 111265, VH "H610C-LC008", reserved 131072KB of GPU memory.
adem of de financy Lett 733555170. Group our pression and pression an	Memory	He10C-LC080	Powered On	<ul> <li>Normal</li> </ul>	83.28 GB	16.32 GB	20.0842	1,882 MB	pid 111263, VH "H610C-LC057", reserved 131072KB of GPU memory.
Kerrere unikaliji, CLI 10 vijekiji, Odu nakama nominy Sabila vijekiji, Odu nakama nomi nakama nominy Sabila vijekiji, Odu nakama nomi nakama nominy Sabila vijekiji, Odu nakama nomi nakama nominy Sabila vijekiji, Odu nakama nomi nakama nakama nakama nomi nakama nomi nakama nakama nomi nakama nomi nakama nomi nakama nomi nakama nomi nakama na	Grepherse								GPU memory left 7319552KB.
10 Outcos         Entra Fach         Entra F	Power Management								Xaerver unix:4, PCI ID 0:96:0:0, VSGA mode, GPU maximum memory 8568128K
than Faab than Faab the F	PCI Devices								pid 100305, VH "HEIOC-LOUHY", reserved 131072KB of GPU memory.
that Faab Resources appendix the searce and the searce appendix to a searce of the searce appendix to a sea	<ul> <li>Virtual Flash</li> </ul>								pid 11320, VM "M610C-LC050", reserved 131072KB of GPU memory.
nageneet In Fach toot Han Cafe _ 1 In Fach too	Virtual Flash Resource								pid 112532, VH "H610C-LC006", reserved 131072KB of GPU memory.
tau Faab Nest Seep Cade attematikes the see Cade the see Cade t	Management								pid 111267, VH "H610C-LC062", reserved 131072KB of GPU memory.
pid 112339, VN VR10-CC051*, reserved 131072VB of OTO memory left. 748054KB.	Virtual Flash Host Swap Cache	M						8 Objects Deport Copy -	pid 112538, VH "H610C-LC049", reserved 131072KB of GPU memory.
GPU memory left 7450624KB.	Configuration								pid 112539, VH "H610C-LCOS1", reserved 131072KB of GPU memory.
									GPU memory left 7450624KB.

The vSGA mode provides higher user density with little management overhead. This mode is the default option when you install the NVIDIA driver. All the vSphere features—such as vMotion, Suspend, and snapshots—are supported with a wide range of NVDIA driver versions.

When no graphics adapter driver is installed on the hypervisor, it shows the memory size as zero on the vSphere web client.

## 4.5 Virtual Shared Pass-Through Graphics Acceleration with the NVIDIA vGPU

When users demand more graphics performance than the consolidation ratio provides in a given environment, the NVIDIA vGPU provides the balance of dedicated resource features, such as vDGA and higher consolidation ratio like vSGA.

In a manner similar to how vSphere virtualizes hardware resources, the NVIDIA vGPU Manager and the GRID software virtualize the GPU hardware.

Figure 10) NVIDIA vGPU architecture (source).

# NVIDIA vGPU Architecture

From their user guide



To use this mode after installing the NVDIA driver on the hypervisor host, switch the graphics adapter mode, as Figure 11 shows.

Figure 11) Host graphics settings.

#### 172.18.214.9 - Edit Host Graphics Settings

Settings will take effect after restarting the host or "xorg" service.

Shared
 VMware shared virtual graphics

Shared Direct

Vendor shared passthrough graphics

#### Shared passthrough GPU assignment policy:

Spread VMs across GPUs (best performance)

Group VMs on GPU until full (GPU consolidation)



A shared PCI device is added to the VM, and the appropriate GPU profile is selected. The GPU profile list is provided in Figure 12.

Physical GPUs per board: 4 Required License Edition Maximum vGPUs per GPU Virtual GPU Type Intended Use Case Frame Buffer (Mbytes) Virtual Display Heads Maximum Resolution per Display Head Maximum vGPUs per Board M10-80 Designer 8192 4096×2160 Quadro vDWS 4 4 M10-4Q Designer 4096 4 4096×2160 2 Quadro vDWS 8 M10-2Q Designer 2048 4 4096×2160 4 16 Quadro vDWS 32 M10-1Q Power User, Designer 1024 2 4096×2160 8 Quadro vDWS M10-0Q Power User, Designer 512 2 2560×1600 16 64 Quadro vDWS M10-2B Power User 2048 2 4096×2160 GRID Virtual PC or Quadro vDWS 4 16 M10-2B4 Power User 2048 4 2560×1600 4 16 GRID Virtual PC or Quadro vDWS M10-1B Power User 1024 4 2560×1600 8 32 GRID Virtual PC or Ouadro vDWS M10-0B Power User 512 2 2560×1600 16 64 GRID Virtual PC or Quadro vDWS M10-8A Virtual Application User 8192 1 1 4 GRID Virtual 1280×1024<sup>1</sup> Application M10-4A Virtual Application User 4096 1 2 8 GRID Virtual 1280×1024<sup>1</sup> Application M10-2A Virtual Application User 2048 1 4 16 GRID Virtual Application 1280×1024 M10-1A Virtual Application User 1024 1 32 GRID Virtual 8 1280×1024<sup>1</sup> Application

Figure 12) vGPU profiles for Tesla M10 (source).

?)

A GRID Virtual PC (profiles typically end with B) is used for virtual desktops, and a GRID Virtual App (profiles end with A) is used for hosted apps. Most of the Q profiles (Quadro Virtual Data Center Workstation) support the 4K resolution, which provides an enhanced user experience for image-editing tools.

According to the policy set, the vGPU Manager schedules a VM to the GPU to either spread the VMs (best performance) or to group VMs (GPU consolidation), as Figure 13 shows.

This release of NVIDIA vGPU supports only homogeneous virtual GPUs. At any moment, the virtual GPUs resident on a single physical GPU must all be of the same type. However, this restriction doesn't extend across physical GPUs on the same card. Different physical GPUs on the same card can host different types of virtual GPUs at the same time, as long as the vGPU types on any one physical GPU are the same.

Figure 13) Homogenous virtual GPU profile combination.

	M10	)-4B		M10	)-2B	M10	)-2B
Valid Pr	rofiles						
M10- 1B							
					_		_
			GP	U 0			

To monitor this relationship, either the vSphere web client or the nvidia-smi vgpu CLI can be used.

Figure 14) nvidia-smi vgpu CLI.

# VM GPU Info

Invalid Profile Combo

With vSphere Web Client/ nvidia-smi vgpu cli

	Hist Oraphics Gr	raphics Devices						
• Storage	Genetica Devices							
Storage Adapters	Configura Dentine							
Storage Devices	/					Q,	Filter	
Detastores	Name	(menter		Anne Type	Configured Type		Marrary	
Host Cache Configuration	NADIATesia M10	NVIDIA 0	orporation	Shared Drect	Shared Direct		7.99 06	
Protocol Endpoints	NVIDIATesia M10	NVIDIA C	brooration	Shared Direct	Shared Direct		7.99.08	
10 Fillers	NV1DIATesia M10	NVIDIA C	orporation	Shared Drect	Shared Direct		7.99 0.0	
Networking	NV1DIATesia M10	NVIDIA C	orporation	Shared Direct	Shared Direct		7.99.08	
Manual Association	NVIDIATesta M10	NVIDIA C	lorporation	Shared Direct	Shared Direct		7.99.08	
Ve Gaar energines	NMDIATesia M10	NVIDIA 0	lorporation	Shared Direct	Shared Direct		7.99 GB	
Velkarmet adaptors	NVIDIATesia M10	NVIDIA 0	orporation	Shared Direct	Shared Direct		7.99-08	
Physical adapters	NVIDH/Tesia M10	NVIDIA 0	lorporation	Shared Drivit	Shared Direct		7.99-08	
TCPIP configuration								
Advanced								
<ul> <li>Virtual Machines</li> </ul>								
VM Startup/Shutdown	M Q.SHI	•				2 Rents	Probeit - 19	Pcebh .
Agent VM Settings	We associated with	. Pas manhing da	in Walter					
Swap file location	100 2510,000 00	o and firsthours and	ALE MELINIES	a aro				
Default VM Compatibility	👻 🕨 🖬 🕤	Acto	ts +			16 Q	Filter.	•
- System	Name 1	. Sinta	. 2014	Provincent Space	Shed Taxes	Print CP	of the state	
Licensing	th HEROCHCOOK	Powered On	O Normal	83.99 08	14.03 08	63 MH	#,222	VB .
Time Configuration	H6100-40012	Powered On	<ul> <li>Nomai</li> </ul>	83.99 08	14.00 OB	83 104	4,222	NB
Authentication Associate	H5100-80018	Powered On	<ul> <li>Nomal</li> </ul>	83.99 OB	14.03 OB	63 544	4,222	MB
Continue of the second	A H010C-IC013	Powered On	Nomal	83.99 08	14.03 GB	03 MH	4,222	NB
Ceroncate	H6100-IC030	Powered On	Nomai	13.99 08	14.03 08	03 MH	a 4,222	10
Power Management	H5100-R0042	Powered On	<ul> <li>Normal</li> </ul>	83.99 08	14.03 08	62 MH	a 4,222	MB .
Advanced System Settings	A H610C-IC059	Powered On	Normal	£3.99 GB	14.03 OB	628 M	Hz 4,2161	149
System Resource Reservation								
Security Proble								
System Swip								
Host Profile	100	.i						
							the second se	

ILA-SMI 396.				
HARR HERE ID		Bus-1d VNC 1D		0/0-4411 +0/0-041
Tesla M10		 0000000	0114100.0	 14
	THID WID-18			
	GB15 M18-18			
	12810 815-18			
	CRID MID/LB			
	0875 M55-18			
	2810 915-18		84100-50018	
	G#15 H10+18			
Tesla Mil		\$555555	0-13-00.6	45%
	URIO M10-18			
	URID NO.6-18		84100-10033	
	GRID N00-19	54494T	84100-50022	
	9810 M10-18		84100-50010	
	GB10 M15-18			
	C0410 H30+18			
	483D M10-18			
Twels MIC		6000000	0+10+00.0	 119
	BE15 M18-18			
	2810 MS.0-18	544458		
	GRID W10-18			
	0835 W10+18			
	8010 NO.0-18			
	0830 M26+18			
Tesla Mid		0000000	0:10:00.6	
	3810 W15-15			
	GRID M34-18		84100-50132	
	GRID MI0-18			
	C810 X10-18			
	G#10 H30+18			
	GRID N10+15			
	G830 NO.0-18			
Tesla M10		0000000	0)40+00_0	
	G#30 H50+18			
	G810 N20-18			
	G810 M10-18			
	ORID N10-18			
	BRID NO.6-LB			
	GR32 N2.0-13			
	CB10 W10-18			
Tesia MIG		0000000	0163108.8	

The vGPU 8Q profile supports CUDA applications. Figure 15 shows a screenshot of CUDA-Z for the 8Q profile.

Figure 15) CUDA-Z screenshot.

CUDA-Z 0.10.251 64	bit	-		×
Core Memory	Performance About			
Name	GRID M10-8Q			
Compute Capability	5.0	- 1	6	
Clock Rate	1032.5 MHz		DVIDI	
PCI Location	0:2:2		TIVIDI	Α.
Multiprocessors	5 (640 Cores)			
Threads Per Multiproc.	2048	14		
Warp Size	32		2	
Regs Per Block	65536	- Contraction		
Threads Per Block	1024	To long		
Threads Dimensions	1024 x 1024 x 64			
Grid Dimensions	2147483647 x 65535 x 65535	Driver	ersion	
Watchdog Enabled	Yes		39	91.81
Integrated GPU	No	Driver [	Oll Version	
Concurrent Kernels	Yes	9.10 (	23.21.13.9	9181)
Compute Mode	Default	Runtim	e Dll Versio	n
Stream Priorities	Yes			6.50
0: GRID M10-80	-			NK -
0. 0100 1110 00				

The combination of VMware vSphere 6.7 Update 1 and NVIDIA GRID software 7.0 supports vMotion, making it easier to manage the host.

Using the frame buffer for the NVIDIA hardware-based H.264/HEVC video encoder (NVENC) can cause memory exhaustion with vGPU profiles that have 512MB or less of frame buffer. To reduce the risk of memory exhaustion, NVENC is disabled on profiles that have 512MB or less of frame buffer.

Application GPU acceleration remains fully supported and available for all profiles, including profiles with 512MB or less of frame buffer. NVENC support from both Citrix and VMware is a recent feature, so you should experience no change in functionality if you are using an older version.

The following vGPU profiles have 512MB or less of frame buffer:

- Tesla M6-0B, M6-0Q
- Tesla M10-0B, M10-0Q
- Tesla M60-0B, M60-0Q

If you require NVENC, use a profile that has at least 1GB of frame buffer.

To use vGPU, the NVIDIA driver must be installed on the guest OS. You must also have other ways to access the VM, such as VNC or Horizon View Agent Direct-Connection. The vSphere web console can't be used to access the VM console session after driver installation and reboot.

Use of vGPU requires an NVIDIA license for the appropriate profile.

#### 4.6 VMware Horizon 7 Enterprise Edition

With VMware, you can pick a subscription-based license or a perpetual license for desktop pools or application pools. Subscription pricing options are available for Horizon 7, Horizon Cloud Service with Hosted Infrastructure, and Horizon Cloud on Microsoft Azure. Horizon 7 subscription licenses can be used on premises or with VMware Cloud on AWS to burst to the cloud. Horizon 7 provides IT with a new streamlined approach to deliver, protect, and manage Windows and Linux desktops and applications while containing costs and making sure that end users can work anytime, anywhere, and on any device.

VMware Horizon 7 Enterprise Edition includes the following components:

- VMware Identity Manager
- VMware Horizon (Connection Server, Agent, Client, and Unified Access Gateway [UAG])
- VMware ThinApp
- VMware App Volumes
- User Environment Manager
- Just-in-Time Management Platform (JMP)
- VMware Mirage
- VMware vSphere Hypervisor
- VMware vCenter Server
- VMware NSX for vSphere
- VMware vRealize Operations Manager (including for vRealize Operations for Published Applications)
- VMware vRealize Orchestrator Plug-In for Horizon
- VMware vRealize Log Insight
- VMware Workstation
- VMware Fusion

#### 4.7 VMware Identity Manager

VMware Identity Manager provides single sign-on (SSO) for virtual desktops and applications, which includes SaaS, web, ThinApp, and mobile applications. With SSO, users do not need to remember multiple user names and passwords. It provides a central location to instantly disable user access to all resources, which protects systems from data leakage. It also acts as an app store for the company, and it provides a self-service catalog that can be customized to add branding.



Figure 16) Web launcher portal (source).

Figure 16 shows the web launcher portal with apps from various resources. This component is optional for this solution. Notably, the Horizon client can show the links to virtual desktops and published apps. VMware Identity Manager is part of the private cloud solution to provide SSO. If you want to implement SSO, see the <u>Reviewer's Guide for On-Premises VMware Identity Manager</u>.

#### 4.8 VMware Horizon

VMware Horizon deploys and manages desktop pools and application pools, manages the connection broker and user entitlements, and so on. For virtual desktops, it can automatically provision full clones, linked clones, or Instant Clones in VMware vSphere environments. It can also broker connections to preexisting physical machines or vCenter VMs, including Remote Desktop Session Hosts (RDSHs), to provide session-based desktops (like terminal services).

The connection server can provide dedicated or floating assignments to desktop pools. With dedicated assignments, a one-to-one relationship is maintained between users and computers. Therefore, when a user logs in, the user always gets the same desktop. The assignments can be made manually or automatically when users first log in. If you do not need a one-to-one user-to-machine relationship—for example, if users are working multiple shifts, sharing the same computer—you should use a floating pool.

A full clone is cloned from an existing vSphere template, whereas a linked clone or an Instant Clone is provisioned from VMware snapshots.

If you are not using thin provisioning from vCenter or from the storage vendor, the storage requirements for virtual desktops are huge. The storage I/O requirements for virtual desktops with spinning disks were challenging, and many companies have kept a separate silo infrastructure for virtual desktops as a result. With NetApp HCI, you can consolidate these workloads along with other infrastructure workloads on an all-flash storage system with guaranteed service levels.

Linked clones help organizations save on storage space. However, the provisioning time is longer, and linked clones require additional components such as a SQL database and a composer service. An Instant Clone doesn't require any external databases, and it uses fewer vCenter operations than a linked clone.

Figure 17) Provisioning Instant Clones.



Priming or image push operations include the creation of an internal template that is a linked clone of the master image snapshot. The internal template joins the computer to the Windows domain and reboots the computer, and then the replica is created for every datastore that is chosen for the provision of pools. The replica is a thin-provisioned, full clone of the internal template. The parent VM is booted from the replica on every host that is part of the resource pool selected for the desktop pool. The booted-up parent VM is quiesced and "hot-cloned" to rapidly produce derivative (child) VMs, taking advantage of the same disk and memory as the parent VM. The clone starts in a booted-up state.

An Instant Clone, even though the initial priming process takes some time, can expand a pool in seconds. An Instant Clone uses vSphere VMFork technology to provision desktops, and it joins the domain by using cloneprep without any reboots required.

The master image should have Horizon Agent installed so that it can communicate with Horizon connection servers, and virtual desktops can be managed from the connection server. The default option is to install the Horizon View Composer Agent. Make sure to disable the VMware Horizon View Composer Agent and use the VMware Horizon Instant Clone Agent.

Horizon 7 supports both Windows and Linux-based desktops, including the Red Hat Enterprise Linux, Ubuntu, CentOS, and NeoKylin operating systems. For application pools, it needs RDSH servers deployed in a farm. On supported Windows Servers, enable the Remote Desktop Services (RDS) role and select the RDSH role services. If you plan to host graphics applications, install Horizon Agent with the 3D RDSH option enabled.

A connection server can create a farm automatically from the VM snapshot, or it can manage manually deployed physical machines or vCenter VMs. The manual option can also be used for the VMs provisioned from other tools, such as VMware vRealize Automation as part of a private cloud.

While deploying an automated farm, remember to set the maximum sessions per RDS host to 60.

In one farm, up to 200 RDSHs can be added. After the farms are defined, RDS-based desktop pools or application pools can be created from the farm. While the application pool is being created, the pool can

automatically sense the installed applications from the farm and pick the applications to publish. If an app is not listed, you can add it manually by providing its name and the executable path.

Within a connection server, you can entitle desktop pools or application pools to users or groups. After entitlement, users can launch a Horizon client to connect to the desktop or application.

If an HTML access agent is installed and enabled in the pool, users can launch the desktop or application from a web browser, which makes access from any device easy. VMware Horizon 7 provides the Blast communication protocol along with PCoIP. Blast Extreme is a new display technology built on the H.264 protocol. Blast Extreme Adaptive Transport (BEAT) maintains a great user experience across a wide variety of network types, ranging from a corporate LAN to public Wi-Fi and mobile networks.



Figure 18) Blast protocol features.

Blast Multimedia provides high-performance multimedia streaming for a rich user experience. Blast 3D provides rich virtualized graphics that deliver workstation-class performance. Blast Local Access provides access to local devices, USB, and device peripherals.

The Session Collaboration feature of Horizon Agent allows multiple users to view and modify the same desktop. This feature can be useful in healthcare, design, engineering, and education organizations for peer reviews, design iterations, and training. The desktop owner can invite users to collaborate in real time on their desktops.

One connection server can handle up to 4,000 sessions (2,000 sessions is recommended), and connection servers (up to a total of seven) can be added for high availability with N+1 and load balancing. The Horizon clients can access the connection servers through a load balancer to distribute the load. If NSX is deployed, the load balancer is available as part of NSX Edge services.

When users try to connect remotely, they can access the desktop pools, hosted applications, intranet resources, or SaaS-based applications through the UAG without VPN access. One UAG can support 2,000 connections.

Figure 19) Unified Access Gateway logical diagram.



NetApp recommends having the Horizon 7 Enterprise components in the management block and keeping the desktop pool and application pool in the resource blocks, each with its own vCenter and NSX server, as shown in Figure 4. A Horizon pod—a combination of management blocks and resource blocks—can support up to 10,000 connections. If more connections are required, you must enable the Cloud Pod Architecture (CPA) feature, which is a federated pod.

Table 2) Horizon sizing.

Item	Maximum
Number of active sessions in the CPA	200,000
Number of pods in the CPA	25
Number of sites in CPA	10
Active connections per pod	20,000 (10,000 recommended)
Number of active connection server instances per pod	7
Active sessions per connection server with direct connection, RDP, tunnel connection, or PCoIP	4,000 (2,000 recommended)
Sessions per RDSH	150 (60 recommended)
RDSHs per farm	200
VMs per vCenter instance	10,000 (5,000 for Instant Clones)
VMs per pool	4,000 (2,000 recommended)

Item	Maximum
VMs per LUN	500
Hosts per vSphere cluster	64 (Instant Clones) 32 (linked clones)
vGPU-enabled VMs per vSphere host	128
Connection servers per pod	7
Connections per UAG	2000

The CPA allows customers to dynamically move and locate Horizon desktop pools and application pools across multiple data centers for efficient management of end users across distributed locations.

#### 4.9 VMware ThinApp

VMware ThinApp is an agentless architecture for application virtualization and packaging. This product is primarily provided for backward compatibility. If you are already using VMware ThinApp, you can continue to use it in the Horizon 7 environment. For new deployments, VMware App Volumes is preferred.

#### 4.10 VMware App Volumes

VMware App Volumes is a Ruby-based application running on NGINX to provide applications on demand to users and computers. The App Volumes Manager runs this application and requires an external SQL database to store metadata. An AppStack is a collection of applications packaged as a virtual disk. An App Volumes agent is installed on virtual desktops or RDSHs that are used for capturing applications and also on virtual desktops or RDSHs that must consume AppStacks.



Figure 20) App Volumes.

**Real-Time App Model with App Volumes** 

**Traditional App Model** 

When an AppStack is assigned to users and computers, it only has read-only access to the AppStack volumes. User-writable volumes are used to allow users to install their own applications and to create the profiles.

A major benefit of using App Volumes is to reduce the number of master images in virtual desktop environments by removing the application dependencies on the base image.

**Note:** An AppStack can only be used with the same OS that is used for capture.

AppStack capture is an easy process. Start AppStack creation by using the App Volumes Manager and pick an AppStack template that was provisioned as part of the deployment process. Also pick one of the machines in your desktop pool or RDSH farm. You must have an App Volumes agent installed on that machine, and the agent must be registered with the App Volumes Manager. Deploy applications manually on that machine, and, after you have installed all the required applications, notify the agent that you have completed the task. The machine reboots and an AppStack is created.

Figure 21) AppStacks and writable volumes.



A storage group in an App Volume is a collection of datastores that are used to serve AppStack volumes or user-writable volumes. Storage groups used for AppStack volumes are primarily used for replicating the AppStack volumes across multiple LUNs for availability and load balancing.

VMware recommends that you not assign more than 8 to 10 AppStacks for each user or device. You must mount the writable volume before mounting AppStack volumes to avoid the reboot prompt. If you are using writable volumes, assign AppStack volumes to users rather than to machines. To reduce the performance impact, set the number of AppStack volumes for each user or machine to a low number.

#### 4.11 User Environment Manager

A user profile contains many personalization options for operating systems in addition to application settings. User Environment Manager captures these details to a network share and imports them back according to defined conditions. It doesn't require the user to have a roaming profile to work. User Environment Manager can set smart policies for the Horizon client—for example, enabling printing when accessed in the office and disabling it when accessed from outside.



Figure 22) User Environment Manager file shares.

Because this configuration is CIFS-share based, each environment (like testing, production, HR, sales, and engineering) can have its own network configuration shares.

#### 4.12 Just-In-Time Management Platform (JMP)

JMP is a new administrative interface for Horizon connection servers that allows you to pick desktops from the pool and AppStacks from App Volumes. You can also use a single workflow to manage User Environment Manager settings such as ADMX, the display language, drive mappings, environment variables, file and folders, login tasks, application personalization, registry settings, shortcuts, triggered tasks, and Horizon Client smart policies. JMP reduces the number of management interfaces needed for onboarding users.

Figure 23) JMP technologies.



## 4.13 VMware Mirage

VMware Mirage provides streamlined image management for physical desktops, full-clone virtual machines, and point-of-sale devices. VMware Mirage categorizes a PC into logical layers owned by either IT or the end user. It also sends a complete copy of the system to the Mirage server in the data center and keeps it synchronized. With the layering technology in Mirage, IT has three options for desktop recovery:

- Restore the entire device (OS, applications, user data, and profile)
- Restore just the applications, user data, and profile
- Restore just the user data and profile

VMware Mirage can manage Horizon FLEX (an optional component not included with Horizon 7 Enterprise) virtual images. Horizon FLEX provides virtual desktop access in offline or disconnected mode with VMware Workstation and Fusion products.

If you plan to use VMware Mirage, see Deployment and Design Considerations for VMware Mirage.

## 4.14 VMware NSX for vSphere

VMware NSX for vSphere brings speed and simplicity to virtual desktop infrastructure (VDI) networking, with policies that dynamically follow virtual desktops. You can create, change, and manage security policies across all virtual desktops with a few clicks. Map these policies to user groups to accelerate virtual desktop onboarding. The security policies can be mapped to users according to role, logical grouping, desktop operating system, and more. VMware NSX Edge services include load balancing (which can be used for connection servers, UAG, and so on), DHCP/DHCP relay (for desktop or application pools), firewall, network address translation (NAT), routing, VPN, and so on.

Figure 24) NSX microsegmentation.



# Micro-segmentation enables security that follows the desktop

#### 4.15 VMware vRealize Operations Manager

VMware vRealize Operations Manager provides the five key features listed in Figure 25.

Figure 25) Five key features of vRealize Operations Manager.



To monitor NetApp HCI with vRealize Operations Manager, use Blue Medora's management pack for NetApp HCI.

vm vRealize Operations Manager	Home Dashboards	Alerts E	nvironment.	Administration				d G	₽ 8-		
K K	NetApp HCI Cluster	Overview	Actions 🖌 All Das	hboards 🛩					Shared		
<ul> <li>Dashboards</li> <li>Getting Started</li> </ul>	Cluster Overview										
NetApp HCI Cluster Overview NetApp HCI VVOL Overview NetApp HCI Health Investigation	Compression Factor	Deduplicat	lon Factor	2.08	Overall Efficiency Factor 47	.32 • Thi	32  Thin Provisioning Factor 13.99				
NetApp HCI Volume Overview Views Reports	Space Provisioned	7.24 18	Space Use	d	332.7 GB	Capacity Utilization 0.00	22 • Act	live Block Space	318.9 <sub>GB</sub>		
	Cluster Selector					Relationships					
	Name	Collection State	Collection Status	BadgelHealth (%)							
	bluemedora-cluster		0	75			tastas iC at				
							tilvermoote-chutter metzes-rig-9t 15.77 A		o. Harrenze ag Dr		
				1.	of 1 items	Viscous	- 0 est	uernedoriș etg. Gz			
	Cluster IOPS			Cluster Throughp	ut	Oust	er Latency				
	3.84	Read Write		mmm	53,934.1 1,649,16	72 Read		4 Read 153 Write			
	132.23	Total			1,634,73	17.5 Total		153 Total			
	Storage Node Selector					Storage Node Overview					
	Name	Collection State	Collection Status	BadgelHealth (%)							
	bluemedora-stg-03		•	75							
	bluemedora-stg-01	-	0	100		CPU Usage	0 • Me	mory Used	39.38		
	bluemedora-stg-04	^	0	75	5 S			GB			
	bluemedora-stg-02	-	0	300							
	0			1-4	of 4 items						
	Storage Node IOPS			Storage Node Th	oughput	Stora	ge Node Latend	τy			
	101 Lon	Read			2,551,03	14.5 Read	ullus	1,649.42 Read			
	33.06	Write Total			95.75 Write 65,454.52 Write 59.75 Total 71,432.62 Total						
	Host System Selector (Com	pute Node)				Host System Overview					
	Name	Collection State	Collection Status	BadgelHealth (%)							
	10.77.8.152	•	0	100		CPU Usage 1	1.74 • Me	amory Usage	32.99		

Figure 26) NetApp HCI integration with vRealize Operations Manager.

NVIDIA provides a GPU management pack for vRealize Operations Manager to monitor metrics from the graphics adapter.

#### Figure 27) NVIDIA integration with vRealize Operations Manager.

Inclusion infrastructure Marcon Infrastructure M	the         50           Object Type         Object Type           Adapt         GPU           J A	Petry           Default Policy           Default Policy           Default Policy           Default Policy           Default Policy           Default Policy           Default Policy	Collection State	Y Filter
Name     Adapter Tage     Option Tage     Name     Adapter Tage       Victorio User Session Status     Calaschold demo     vCenter Adapter     Host System     Default Policy     A     G     1218 274 105 Tesis MI00-00000     NUDIA VG       Vistorio DVS Possion Status     1218 274 105 Tesis MI00-0000     VCINTer Adapter     Host System     Default Policy     A     G     1218 274 105 Tesis MI00-0000     NUDIA VG       Vistorio DVS Policio     1218 274 105< Tesis MI00-0000     VCINTer Adapter     Host System     Default Policy     A     G     1218 274 105 Tesis MI00-0000     NUDIA VG       Vistorio DVS Policio     1218 274 105< Tesis MI00-0001     VCINTer Adapter     Host System     Default Policy     A     G       Vistorio Adapter Ser Heath     Vistorio Hist System     Default Policy     A     G     1218 274 105 Tesis MI00-0000     NUDIA VG       Vistorio Adapter Ser Heath     Vistorio Hist System     Default Policy     A     G     1218 274 105 Tesis MI00-0000     NUDIA VG       Vistorio Adapter Ser Heath     Vistorio Hist System     Default Policy     A     G     1218 274 105 Tesis MI00-0000     NUDIA VG       Vistorio Adapter Ser Heath     Vistorio Hist System     Default Policy     A     G     1218 274 105 Tesis MI00-0000     NUDIA VG       Vistori MSC Vistorio Hist Ser Heath <t< th=""><th>이용(et Type ) Adapt_ GPU ) CPU (CPU) Adapt_ GPU ) CPU ) CPU )</th><th>Petry Default Policy Default Policy Default Policy Default Policy Default Policy Default Policy</th><th>Collection State</th><th>Collection State</th></t<>	이용(et Type ) Adapt_ GPU ) CPU (CPU) Adapt_ GPU ) CPU )	Petry Default Policy Default Policy Default Policy Default Policy Default Policy Default Policy	Collection State	Collection State
Construction     Charla holding deemo     VCenter Adapter     Host System     Default Pairsy     ▲     ●       Construction     VCEnter Adapter     Host System     Default Pairsy     ▲     ●       Construction     VCEnter Adapter     Host System     Default Pairsy     ▲     ●       VCENDER     VCEnter Adapter     Host System     Default Pairsy     ▲     ●       VCENDER     VCEnter Adapter     Host System     Default Pairsy     ▲     ●       VCENDER     VCEnter Adapter     Host System     Default Pairsy     ●     ●       VCENDER     VCEnter Adapter     Host System     Default Pairsy     ●     ●       VCENDER     VCENter Adapter     Host System     Default Pairsy     ●     ●       VCENTER     VCENter Adapter     Host System     Default Pairsy     ●     ●       VCENTER     VCENter Adapter     Host System     Default Pairsy     ●     ●       VCENTER     VCENter Adapter     Host System     Default Pairsy     ●     ●       VCENTER     VCENter Adapter     Host System     Default Pairsy     ●     ●       VCENTER     VCENter Adapter     Host System     Default Pairsy     ●     ●       VCENTER     VCENter Adapter     Host	J Adapt. GPU J Adapt. GPU J Adapt. GPU J Adapt. GPU J Adapt. GPU J Adapt. GPU St St Created On Status	Default Policy Default Policy Default Policy Default Policy Default Policy Default Policy Atent Type Alert	n n n n n n n n n n n n n n n n n n n	© © © 1-8of8i ¥ Filter
n DS Pools n PolS Pool 1218 214 0 Verter Adapter Host System n Adapter Seit Neuman n Pol Verter Pools n PolS Pool 1218 214 10 Verter Adapter Host System n Adapter Seit Neuman n PolS Pool 1218 214 10 Verter Adapter Host System n Adapter Seit Neuman n PolS Pool 1218 214 10 Verter Adapter Host System n Adapter Seit Neuman n PolS Verter Adapter Host System n Adapter Seit Neuman n PolS Pool 1218 214 10 Verter Adapter Host System 1218 214 10 Verter Adapter Host System 1100 Verter Adapter Host System	J Adapt. GPU J Adapt. GPU J Adapt. GPU J Adapt. GPU J Adapt. GPU Status Created On Status	Default Policy Default Policy Default Policy Default Policy Atent Type Atent	m m 1 Subtype	● ● ● ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■
A Applications n Applications n Destance Databas n Destance Da	J Adapt. GPU J Adapt. GPU J Adapt. GPU DAdapt. GPU Created On Status	Default Policy Default Policy Default Policy Alert Type Alert	n 11 Subtype	● ● 1-8of8
Lissuage Li	J Adapt. GPU J Adapt. GPU Ost Created On Status	Default Policy Default Policy Alert Type Alert	rt Subtype	● ● 1-8of8 ▼Filter
RROS Hould Details Academ Ser Hould Rob Loge Reb Concert New Concert Ser Mark Heter VMs ore Utilization New Concert August and the selected Host Heter VMs ore Utilization New Concert New Concert Ne	DAdapt GPU	Default Policy Alert Type Alert	rt Subtype	● 1-8 of 8 ▼ Filter
Adapter Freider Bord Urster Steringer Root Cauter Adarytes Mort With Logis Kert Mar Kert Mar Kert Mar Kert Mar Chicatity – Alert Store Chicatity – Ale	ost 👶 🍮 Created On Status	Alert Type Alert	rt Subtype	1-8of8 ⊤Fiter
Adverts on VSPUs running on the selected Host	ost a a Created On Status	Alert Type Alert	rt Subtype	
tel Visit Difficultation figuration Talization Bio Original States 112 1823410-1est. 10/6/18 7.20. Hardware ( Capacity Randware ( Capacity	Created On Status	Alert Type Alert	rt Subtype	11100
ohfguration k Configuration k Configuration Utilization Utilization Utilization Utilization				
Environmet Overview Hots Summary GPU Summary	No Results Fo	Found		
VGPU Summary 1-1 of Items 1-1 of Items				No
HCI VVOL Overview HCI Volume Overview Top 25 vGPUs By 3D/Compute Utilization (%) (select any vGPU and navigate to see details) Top 25 vGPUs By Memory Utilization (%)	elect any vGPU and navig	rigate to see details)		
p HC Cluster Overview 🛃 🗸 😹 👔 😴 🔤 🖓				Top 25 H
p HCl Health gation Utilization Index Objects Utilization Index Utilization Index	01	Objects		

#### Figure 28) GPU metrics.

vm vRealize Operations Man	ager Home	Dashboards	Alerts	Environmen	t Administration											Ľ,	
BACK ~ «	Horizon User S	Session Details	Actions 🛩	All Dashboard	ls •											3	Shared
<ul> <li>Dashboards</li> </ul>	Horizon Remote Se	ssions													*	0 ?	•
Horizon Help Desk	User	Pool Name		Session State	e Object Type	Protocol	Session Duration	(hours)	Logon Duration (sec)	App Stacks							
Horizon Infrastructure	hcivdl/loginvsi003	?		Disconnect	ed Application Session	?	403.43	1	?	?							-
Horizon User Sessions Horizon VDI Pools	hcivdl\loginvsi002	?		Disconnect	ed Application Session	2	14.73	1	?	?							11
Horizon RDS Pools	hcivdi\administrator	Windows 10		Disconnect	ed VDI Desktop Session	7	778.8	3	2	7							
Horizon Applications Horizon Desktop Usage					VDI Desktop Session												
Horizon User Session Details					VDI Desktop Session												
Horizon RDS Host Details Horizon Adapter Self Health													1 - 50 of 218 it	terns	¢ <u>1</u> 2	3 4 1	2
Horizon End User Experience	Session Indicator M	letrics															
Horizon Root Cause Analysis Troubleshoot with Loos	Health			100	Workload		?	Total Logon	Time		?	Round Trip Latency				1	2 ^
Getting Started				5			5		No Data		sec		No Data				a
Heavy Hitter VMs					•				TWO DUIL				140 0010				- 11
Host Configuration	Encode Frame	e Rate		?	TX Bandwidth		?	BLAST Estim	ate Bandwidth		?	PCoIP RX Bandwidth				mbit/se	2
Network Configuration		No Data			No E	lata			No Data				No Data				
Cluster Utilization	PCoIP TX Pac	ket Loss		?	PCoIP RX Packet Loss		?										
NVIDIA Environment Overview																	
NVIDIA Host Summary	Session Logon Brea	akdown															
NVIDIA GPU Summary NVIDIA Application Summary	Profile Load T	ime		0.11 sec	Shell Load Time		3.13	Interactive 5	Session Time		1.09						Î.
NetApp HCI VVOL Overview NetApp HCI Volume Overview	Session Processes																
NetApp HCI Cluster Overview																	
NetApp HCI Health Investigation																	
Views																	
Reports																	

The vRealize for Horizon adapter is required to monitor the virtual desktops, application pools, App Volumes, and so on. vRealize Operations Manager Agent must be installed on the VMware Horizon connection servers, and agents must be installed on virtual desktop templates.

For more information, see the vRealize Operations Manager sizing guidelines.

#### 4.16 VMware vRealize Orchestrator Plug-In for Horizon

VMware Horizon provides vRealize Orchestrator Plug-In to extend vCenter or vRealize Automation with custom workflows. This plug-in allows users to choose which tool to deploy to manage the desktop and

application pools. With Horizon Agent Direct Connect, users can connect to desktops directly without going through the connection broker. VMware Horizon also supports HTML Direct Connection. To identify which features are provided and which features can be created with custom code, see the <u>plug-in</u> <u>documentation</u>.



Figure 29) Architecture of vRealize Orchestrator Plug-In for Horizon.

#### 4.17 VMware vRealize Log Insight

VMware vRealize Log Insight delivers heterogeneous and highly scalable log management with intuitive, actionable dashboards, sophisticated analytics, and broad third-party extensibility, providing deep operational visibility and faster troubleshooting.

To collect logs from desktops, you need to install vRealize Log Insight Agent on the desktops. For hosts such as vSphere Hypervisor or NetApp HCI storage, you can configure it to forward the logs to vRealize Log Insight as a syslog server.

Figure 30) vRealize Log Insight architecture.



Figure 25) vRealize Log Insight UI.

vm Log Insight	Dashboards Interactive Analytics	🛦 admin 😑
Custom Dashboards > My Dashboards > Shared Dashboards Content Pack Dashboards > General	Latest 48 hours of data v c Displey legand on all widgets () + ADD FAITER Active User Count	
✓VMware - Horizon View General - Resource Systems Cluster	No results	
Uses Authenocation Desktop Sessions Security Geterey PCoIP - General PCoIP Server - Bandwidth Utilization BCoIP Server - Bandwidth	Initial Status Codes over Time         Initial Status Codes over Time           1         1           1         55	agent_ent_ok
Poor Comerce automatic Utilization PCoIP Server - Packet Counts PCoIP Server - Image PCoIP Server - Image	0         11/00         21/00         Nui-28         orize         orize         12/00         12/00         12/00         Nui-27         01/00         orize         orize         12/00         15/00         12/00         10/00         Active Agent Versions         Active Virtual Desktop Operating System         100	
PCoIP Server - Vernings PCoIP Server - Errors PCoIP Client - Connection Quality PCoIP Client - Packet Counts PCoIP Client - Decoder	No results 100 100 100 100 100 100 100 100 100 10	
PCoIP Client - Warnings PCoIP Client - Errors Errors > VMware - VSAN > VMware - vRops 6.7+		

For sizing information, see Sizing the vRealize Log Insight Virtual Appliance.

#### 4.18 VMware Workstation

VMware Workstation is a desktop hypervisor tool that can run on Windows and Linux desktops or laptops to create VMs running multiple Windows or Linux instances at the same time. When used with the optional product Horizon Flex, a virtual desktop from the desktop pool can be used in offline or disconnected mode.

#### 4.19 VMware Fusion

VMware Fusion allows Apple macOS machines to create Windows, Linux, or macOS VMs. When used with the optional product Horizon Flex, a virtual desktop from the desktop pool can be used in offline or disconnected mode.

# 5 Solution Design

As Figure 4 shows, the End-User Computing on NetApp HCI solution keeps infrastructure VMs separate from workload VMs in the management block and resource block. Having a resource block keeps the workloads securely contained and isolated. With NetApp HCI, each resource block corresponds to an account and a separate VLAN for storage traffic. Every block requires at least four compute nodes when used with NSX. Keep the three controllers on separate hosts and keep an additional one for fault tolerance.

In linked mode, 15 vCenter Servers are allowed. Therefore, 14 resource blocks and one management block are possible for each Horizon pod. The number of compute nodes in vCenter linked mode is 5,000. If you exceed the maximum storage space available in each cluster, you can deploy more clusters for the resource blocks and manage them centrally from the vCenter plug-in.



Figure 31) Resource block scalability.

Based on VMware guidelines, keeping 10,000 sessions for each Horizon pod and 2,000 sessions for each resource block provides five resource blocks for each Horizon pod. If you require more sessions, use the CPA feature to create global entitlement for users across the Horizon pods within the same site, across the sites, or to the cloud.

We have seen user density for NetApp HCI GPU compute nodes in the range of 130 to 150 virtual desktops per node for vSGA mode. To be conservative, we picked 128 users per node. That number is also valid if we use vGPU mode.

The configuration of desktop VMs is described in Table 3.

Table 3) Desktop VM configuration.

Item	Value
vCPU	2
RAM	4GB
NIC (VMXNET3 adapter)	1
VMDK	<ul> <li>70GB (OS)</li> <li>10GB (UWV + profile)</li> <li>20GB (AppStack)</li> </ul>

AppStacks support only Windows OS. Supported Linux VMs with Horizon can use hosted applications on Windows RDSH served by AppStacks.

An Instant Clone provides a delta disk for every virtual desktop. The size of the delta disk depends on the change rate. We assumed a 10% change rate for the sizing calculations, and we also assumed no backup volumes configured on App Volumes and considered 500GB for AppStacks. Table 4 provides a rough estimate for the number of datastores required for 1.5TB volumes.

Table 4) Datastore estimates.

Component	800 Users	1,200 Users	2,000 Users		
NetApp HCI compute nodes (resource block)	8	11	17		
NetApp HCI compute nodes (management block)	4	4	4		
NetApp HCI storage nodes (based on the model used)	> or = 4	> or = 4	> or = 4		
1.5TB datastore for hosting virtual desktops	4	6	10		
1.5TB datastore for user writable volumes	8	12	20		

The datastores that hold AppStack volumes are mostly for read-only access, and the datastores that hold user-writable volumes provide read/write access. With NetApp HCI, there is no need to separate those workloads, and both can coexist on the same datastore.

Hosted applications serve task-oriented users such as hospital workers who require access to specific applications and don't have time to wait for desktops to start up. The application pool is hosted on a farm of RDSHs. If the RDSH reboots during application access, the user gets an error. When the application is relaunched, it is served from the other surviving hots.

For desktop pools serving Instant Clones, Horizon refreshes the desktop on another host if one of the vSphere hosts crashes. Typically, the desktop is protected by vSphere HA.

The UAG, App Volumes Manager, and connection servers are accessed through a load balancer, and the traffic is redirected to the server in service. AppStack volumes are replicated to multiple datastores by using the storage group, and writable volumes are distributed based on the following two policy sets:

- **Spread.** Distributes files evenly across all the storage locations. When a file is created, the storage with the most available space is selected.
- **Round-robin.** Distributes files by sequentially using the storage locations. When a file is created, the storage with the oldest used time is selected.

The App Volumes datastores that hold the virtual disks for AppStacks and user-writable volumes can be replicated from NetApp HCI to any ONTAP devices on premises or in the cloud. If the User Environment Manager file shares are also hosted on NetApp ONTAP, they too can be replicated to other ONTAP devices on premises or in the cloud. This configuration enables users to burst into the cloud for a short duration and then to bring data back on premises.

## 6 Technology Components

This section covers the technology components that we used for validating the End-User Computing on NetApp HCI solution.

#### 6.1 Hardware Components

Table 5 lists the hardware components that we used for the solution. These components might vary according to customer requirements.

Component	Count	Description
NetApp HCI compute nodes (resource block)	2	H610C
NetApp HCI compute nodes (management block)	2	H500E
NetApp HCI storage nodes	4	H500S
Cisco Nexus 9000 switch	2	iSCSI/VM/vMotion
1Gb switch	1	IPMI/Management

Table 5) Hardware components.

#### 6.2 Software Components

Table 6) lists the software components that are required to implement the solution. These components might vary according to customer requirements.

Table 6) Software requirements.

Software	Version
NetApp HCI NDE	1.4
VMware Horizon 7 Enterprise	
VMware Identity Manager	3.3.0
VMware Horizon	7.6
VMware App Volumes	2.14.2

Software	Version
VMware NSX for vSphere	6.4.3
VMware vCenter	6.7 Update 1
VMware vSphere	6.7 Update 1
VMware vRealize Operations Manager	6.6.1
VMware vRealize Operations for Published Applications	6.5.1
VMware vRealize Orchestrator Plug-in for Horizon	1.4.0
VMware User Environment Manager	9.5.0
VMware vRealize Log Insight for vCenter	4.6.1
NVIDIA GRID	7.0

# 7 Solution Verification

This solution architecture is based on VMware vSphere 6.7 Update 1, VMware Horizon 7.6, and NVIDIA GRID 7.0 so that it can support vMotion for the VM running graphics workloads, including vGPU mode. In our testing, our management block was hosted on NetApp HCI 500E/410C nodes, and our resource block was hosted on NetApp HCI H610C nodes. We used Cisco Nexus 9K switches for the 10Gb network, and we used a 1Gb switch for the management network.

We verified that we were able to deploy desktop pools and application pools utilizing the GPU with VMware Horizon. We performed on-demand desktop assignment and application provisioning for a test user, and we confirmed that their settings were preserved with logout and login.

We determined that we could put a vSphere host in maintenance mode, and we migrated a VM running a graphics workload with a vGPU profile to the other host. We were able to perform session sharing while running multiple workloads with Horizon clients and by using HTML with a web browser.

We validated the solution by simulating user-like workloads using the Login Virtual Session Indexer (Login VSI). Login VSI is the industry-standard load-testing tool for testing the performance and scalability of centralized Windows desktop environments, such as server-based computing and VDI.

The testing included the following scenarios:

- Get VSIMax for a single server with no GPU used for a knowledge worker.
- Get VSIMax for a single server with vSGA mode for a knowledge worker.
- Get VSIMax for a single server with vGPU mode for a knowledge worker.
- Get VSIMax for a single server with vSGA mode for a multimedia workload.
- Get VSIMax for a single server with vGPU mode for a multimedia workload.

All the above testing was performed using JMP to entitle the desktop pool, an AppStack volume for Login VSI workloads, and the User Environment Manager policies.

## 8 Conclusion

NetApp HCI can scale or shrink based on your business needs so that you can provide flexible options for your customers. The QoS feature makes NetApp HCI simple to integrate with existing workloads. NetApp HCI GPU compute nodes provide an enhanced user experience for knowledge-based workers who can

benefit from hardware 3D acceleration. With Horizon 7, users can securely connect to desktop pools or application pools from any device, anywhere.

# Where to Find Additional Information

To learn more about the information described in this document, refer to the following documents and websites:

#### NetApp

- NetApp Product Documentation <u>https://www.netapp.com/us/documentation/index.aspx</u>
- NetApp HCI Documentation Center <u>http://docs.netapp.com/hci/index.jsp</u>
- NetApp Data Fabric <u>https://www.netapp.com/us/info/what-is-data-fabric.aspx</u>
- NetApp HCI Datasheet <u>https://www.netapp.com/us/media/ds-3881.pdf</u>
- NetApp HCI Technical docs <u>http://docs.netapp.com/hci/index.jsp</u>
- NetApp HCI Deployment Guide <u>https://library.netapp.com/ecm/ecm\_download\_file/ECMLP2844053</u>
- NetApp HCI Network Setup Guide <u>https://www.netapp.com/us/media/tr-4679.pdf</u>
- VMware vRealize Operations Management Pack for NetApp HCI and SolidFire <u>https://bluemedora.com/resource/vmware-vrealize-operations-management-pack-for-netapp-hci-solidfire/</u>
- NetApp SolidFire<sup>®</sup> vRealize Orchestrator Plug-in <u>https://github.com/solidfire/vrealize-orchestrator-plugin</u>
- NetApp HCI Theory of Operations
   <u>https://www.netapp.com/us/media/wp-7261.pdf</u>
- NetApp Element software <u>https://www.netapp.com/us/products/data-management-software/element-os.aspx</u>
- NetApp ONTAP Select <u>https://www.netapp.com/us/products/data-management-software/ontap-select-sds.aspx</u>
- NetApp Interoperability Matrix Tool <u>https://mysupport.netapp.com/matrix/#welcome</u>

#### **NVIDIA**

 NVIDIA Virtual GPU Software Documentation <u>https://docs.nvidia.com/grid/</u>

#### VMware

- VMware Tech Zone
   <u>https://techzone.vmware.com/</u>
- VMware Workspace ONE and VMware Horizon 7 Enterprise Edition On-Premises Reference Architecture <u>https://techzone.vmware.com/resource/vmware-workspace-one-and-vmware-horizon-7-enterprise-</u> edition-premises-reference

- Deploying Hardware-Accelerated Graphics with VMware Horizon 7
   <u>https://techzone.vmware.com/resource/deploying-hardware-accelerated-graphics-vmware-horizon-7</u>
- VMware Horizon 7 Sizing Limits and Recommendations <u>https://kb.vmware.com/s/article/2150348</u>
- VMware Configuration Maximums <u>https://configmax.vmware.com/</u>
- Reviewer's Guide for On-Premises VMware Identity Manager
   <u>https://techzone.vmware.com/resource/reviewers-guide-premises-vmware-identity-manager</u>
- Deployment and Design Considerations for VMware Mirage
   <u>https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/vmware-horizon-mirage-deployment-design-considerations.pdf</u>
- Network Ports in VMware Horizon 7
   <u>https://techzone.vmware.com/resource/network-ports-vmware-horizon-7</u>
- Horizon for Linux FAQ
   <u>https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/horizon/vmware-horizon-for-linux-faq.pdf</u>
- vRealize Operations Manager 6.6.1 Sizing Guidelines <u>https://kb.vmware.com/s/article/2150421</u>
- Sizing the vRealize Log Insight Virtual Appliance <u>https://docs.vmware.com/en/vRealize-Log-Insight/4.6/com.vmware.log-insight.getting-</u> started.doc/GUID-284FC5F4-B832-47A7-912E-D407A760CAE4.html

# **Version History**

Version	Date	Document Version History
Version 1.0	January 2019	Initial version

Refer to the <u>Interoperability Matrix Tool (IMT)</u> on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

#### **Copyright Information**

Copyright © 2018–2019 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

#### **Trademark Information**

NETAPP, the NETAPP logo, and the marks listed at <u>http://www.netapp.com/TM</u> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.

