



NVIDIA GRID DEPLOYMENT GUIDE FOR CITRIX XENDESKTOP 7.12 ON VMWARE VSPHERE 6

SP-00000-001_v01.0/. | May 2017

NVIDIA TechPub



DOCUMENT CHANGE HISTORY

SP-00000-001_v01.0/.

Version	Date	Description of Change
01	05/2015	Initial release
02	05/2017	Update to Citrix 7.12

TABLE OF CONTENTS

Chapter 1. About this Document	8
1.1 About the Technologies	8
1.1.1 NVIDIA GRID.....	8
1.1.2 VMware vSphere.....	9
1.1.3 Citrix XenDesktop.....	9
1.2 Organization	11
1.3 Formatting Conventions.....	12
1.4 Related Documentation	13
1.5 The Deployment Guide Lab	13
Chapter 2. Getting Started	15
2.1 Why NVIDIA GRID?.....	15
2.2 NVIDIA GRID Architecture.....	16
2.3 Supported GPUs	19
2.4 Tesla M60/M6 Specifications	20
2.5 Tesla M10 Specifications	21
2.6 GRID K1 / K2 Specifications	22
2.7 GRID vGPU Profiles	22
2.8 Supported Graphics Protocols.....	25
2.9 Homogenous GRID vGPUs.....	25
2.10 GRID Licensing.....	26
2.11 Graphics vs. Compute Mode	29
2.12 Distribution of GRID vGPUs	30
2.13 The Before You Begin.....	31
2.13.1 General Prerequisites.....	31
2.13.2 Build an Evaluation Plan	32
2.13.3 Size Your Environment	32
2.13.4 Choosing Your Hardware	32
2.13.5 Choose Virtualization Components	33
2.13.6 Pre-Installation Preparation	34
2.13.7 Known Limitations	34
2.13.8 Server BIOS Settings	34
Chapter 3. Installing VMware ESXi 6.0 U2	36
3.1 Installing ESXi 6.0 U2	36
3.2 Initial Host Configuration	41
Chapter 4. Installing VMware vCenter Server 6.0 U2	48
4.1 Installing vCenter 6.0 U2 Server Appliance	48
4.1.1 vCenter Server Client Integration Plug-in Installation	49
4.1.2 vCenter Server Appliance (VCSA) Installation	54

4.2 Installing vCenter 6.0 U2 Server on Windows Server 2012	64
4.2.1 Initial Installation	64
4.2.2 Adding Licenses to your vCenter Server	76
4.3 Setting the NTP Service on a Host	81
4.4 Setting vCenter to Auto-Start	81
4.5 Mounting an NFS ISO Data Store	84
4.6 Creating vSphere clusters for GRID servers	87
4.6.1 Creating a Datacenter in vCenter	87
4.6.2 Creating a Cluster in the Datacenter	89
4.6.3 Adding a Host	90
Chapter 5. GRID software preparation	96
5.1 Obtain an evaluation license PAK and download software	96
5.2 Create the License Manager Server	107
5.2.1 Prepare the License Manager VM	107
5.2.2 Install the License Server	108
5.3 Obtain and install your license	113
5.4 Exchanging your evaluation license for a permanent one	117
5.5 Support, update, and maintenance (SUMS)	122
Chapter 6. NVIDIA GRID vGPU Manager Installation	128
6.1 Prepare and Install the VIB Files	128
6.1.1 Uploading VIB in vSphere	128
6.1.2 Place the Host in Maintenance Mode	133
6.1.3 Installing the VIB	134
6.1.4 Confirm the Driver VIB Installation	134
6.2 Uninstalling the VIB	135
Chapter 7. Building Citrix XenDesktop 7.12	137
7.1 Registering and Allocating the Trial License	137
7.2 Installing XenDesktop 7.12	144
7.3 Importing vSphere and vSphere Web Services SDK Certificates	151
7.3.1 Importing the vSphere Certificate	151
7.3.2 Importing vSphere Web Services SDK Certificates	154
7.4 Using Citrix Studio to Perform Site Setup	164
7.5 Configuring Citrix Licenses Server	171
7.6 Verify Citrix StoreFront Configuration	175
Chapter 8. Selecting GRID vGPU profiles	176
8.1 Profiles explained	176
8.2 The role of the GRID vGPU Manager	177
8.3 Power User & Designer profiles	177
8.4 Knowledge Worker profiles	178
8.5 Application delivery vs full desktop	178
8.6 The full list of GRID vGPU profiles	179
8.7 Matching Profiles to User Needs	180
8.8 The 8Q and 8A PROFILES vs vDGA/Pass-Through	181

Chapter 9. Application Based Host Sizing Example – Esri	182
9.1 Application Based Sizing	182
9.2 Typical Physical Workstation	182
9.2.1 Recommended System Requirements	183
9.3 UX – The VDI User Experience	184
9.3.1 Esri Benchmark Metrics.....	184
9.3.2 Real Life Experience vs Benchmarking.....	185
9.3.3 The Importance of Eyes On	185
9.4 Testing Methodology	185
9.5 Results Overview.....	186
9.6 Conclusion.....	187
Chapter 10. Creating your first GRID Virtual workstation.....	188
10.1 Creating a Virtual Machine	188
10.2 Installing Windows 10	194
10.3 Installing VMware Tools on the Virtual Machine.....	201
10.4 Adding the Golden Master to the Domain	206
10.5 Installing Citrix Virtual Delivery Agent	209
10.6 Enabling the NVIDIA GRID vGPU	218
10.7 Installing NVIDIA Driver in the VM	222
10.8 License the desktop	226
10.9 Removing the GRID vGPU Device from the Master Image	228
Chapter 11. Creating XenDesktop Machine Catalogs and Delivery Groups ..	231
11.1 Creating a Snapshot.....	232
11.2 Creating a Machine Catalog for the Master Image (Optional).....	233
11.3 Creating a XenDesktop Random Machine Catalog From a Snapshot.....	241
11.4 Configuring Catalog VMs to Enable GRID vGPU Profiles	247
11.5 Creating a XenDesktop Delivery Group for the Master Image (Optional)	252
11.6 Creating a XenDesktop Delivery Group for the Pooled Machine Catalog	260
11.7 Connecting to the Citrix Delivered Desktop for the First Time	266
Chapter 12. Testing, Monitoring, & Benchmarking	271
12.1 Testing Virtual Desktops vs Physical Ones, Important Differences.....	271
12.2 NVIDIA System Management Interface.....	272
12.2.1 Querying nvidia-smi	272
12.2.2 Monitoring	273
12.3 Testing/Simulating Apps and Tools.....	275
12.3.1 GPUProfiler	275
12.3.2 Browsers.....	275
12.3.3 Microsoft.....	275
12.3.4 Dassault Systèmes	275
12.3.5 PTC.....	275
12.3.6 Siemens	275
12.3.7 OpenGL	276
12.3.8 Lego Designer	276

12.3.9 Adobe	276
12.3.10 BitSquid Stone Giant	276
12.3.11 Google Earth	276
12.3.12 Hulu/YouTube videos	276
12.3.13 Autodesk	276
12.3.14 Slicer3D	277
12.3.15 FutureMark	277
Chapter 13. Known Issues	278
13.1 Uninstalling Existing vSGA VIB to Install the NVIDIA GRID vGPU VIB	278
13.2 NVIDIA WDDM Issues	279
13.3 vGPU-Enabled VMs and nvidia-smi Fail to Start	279
13.4 VMotion and DRS Not Supported	279
Chapter 14. Troubleshooting	280
14.1 Troubleshooting Procedure	280
14.1.1 Confirm the GPU Installation	280
14.1.2 Setting the GPU Graphics Mode	281
14.1.3 Verify the VIB Installation	285
14.1.4 Confirm VIB Loading	285
14.1.5 Confirm PCI Bus Slot Order	285
14.1.6 Check for X.Org Server Issues	286
14.1.7 Check nvidia-smi	289
14.1.8 Resolving Black Screen Errors	290
14.1.9 Multi-monitor Display Issues	290
14.2 Additional Remediating Steps	291
14.2.1 Enabling VM Console VNC Access	291
14.2.2 Joining an Endpoint to the Domain	291
14.2.3 Filing a Bug Report	293
Chapter 15. Using WINSXP	295
Appendix A. Acronyms	299

LIST OF FIGURES

Figure 1-1 Citrix XenDesktop components diagram	10
Figure 2-1 NVIDIA GRID System Architecture	17
Figure 2-2 NVIDIA GRID vGPU Internal Architecture	18
Figure 2-3 NVENC overview	19
Figure 2-4 Example of GRID vGPU Configurations on Tesla M60	26

Figure 2-5	GRID Licensing management system	28
Figure 2-6	Example of GRID vGPU Distribution on Tesla M60 Card	31
Figure 7-1	Citrix vSphere certificate error	168
Figure 12-1	Graphed Log Output.....	275

LIST OF TABLES

Table 1-1	Lab Servers	13
Table 2-1	Tesla M60 and Tesla M6 Hardware Specifications.....	20
Table 2-2	Tesla M10 Hardware Specifications	21
Table 2-3	GRID K1/K2 Hardware Specifications.....	22
Table 2-4	GRID vGPU Types	23
Table 2-5	GRID License Editions.....	27
Table 2-6	Compute Mode Settings	29
Table 2-7	Graphics Mode Settings.....	29
Table 8-1	NVIDIA GRID Graphics Memory Options.....	179
Table 8-2	User types and appropriate profiles	180

CHAPTER 1. ABOUT THIS DOCUMENT

This guide makes two assumptions about you: that you are interested in hosting graphics accelerated applications in a virtual environment, and that you want to get a suitable environment running quickly, even though you may have little or no experience deploying a virtualized environment. This deployment guide is written to give you enough information to successfully design, build, and execute a proof of concept (POC) project using NVIDIA GRID™, VMware vSphere® 6.0 U2, and Citrix® XenDesktop® 7.12. Once completed, this POC system will allow you to deliver virtual workstations (vWS) and virtual PCs (vPC) with high resolution, graphics accelerated applications, remotely to end users.

This section contains a brief overview of the technologies presented, an overview of the chapters, and the formatting conventions. We also describe the lab environment used to test the contents of this guide and capture the screenshots used.

1.1 ABOUT THE TECHNOLOGIES

Let's start off with a brief description of the technologies involved.

1.1.1 NVIDIA GRID

NVIDIA GRID brings the power of NVIDIA GPU to virtual desktops, applications, and workstations, with an immersive user experience for everyone, from office workers to mobile professionals to designers. GRID technology meets the rising performance demands of modern applications and Windows 10, bringing graphics acceleration to the data center, enabling IT to centralize applications and data, and deliver virtual workspaces that offer improved management security, and productivity. Business can deliver cost-effective VDI performance that scales, transforming workflow and

liberating users from the confines of PC's and offices, collaborating in real-time, from anywhere, using any device. IT can now virtualize 100% its users with a native experience, achieving dramatically lower cost per desktop, with manageability that ensures a consistently great user experience. NVIDIA GRID improves security, ensuring data and IP are centrally stored and protected, while enabling business partners to securely collaborate without risk.

The following NVIDIA GRID technologies are used in this guide:

- ▶ NVIDIA GRID vGPU Manager – This software runs in the hypervisor layer and allocates physical GPU resources to VMs.
- ▶ NVIDIA GRID Monitoring – which provides features like:
 - ▶ vGPU Discovery, which shows the currently active vGPU as well as supported and creatable types of vGPU.
 - ▶ vGPU Properties such as number of displays, maximum resolution, frame buffer size, current license status and more.
 - ▶ Utilization Reports, that provide metrics on encode and decode engine utilization, and other features.
- ▶ NVIDIA® Quadro® Driver – Vendors of high-performance graphics applications certify their software to work with NVIDIA Quadro drivers, and NVIDIA GRID provides those same drivers for use in Virtual Workstations.
- ▶ NVIDIA GRID License Manager – This server software helps administrators deploy, manage, and audit GRID license usage among virtual workstations, desktops, and applications.
- ▶ NVIDIA® Tesla® multi-GPU boards – These provide the physical GPU and frame buffer resources for the VMs to use.

1.1.2 VMware vSphere

VMware vSphere 6.0 U2 provides virtualization technology. vSphere consists of VMware® ESXi™, which installs on bare-metal servers and provides the virtualization necessary, and VMware® vCenter Server™ (simply referred to as vCenter) which manages the ESXi servers and provides a consolidated, centralized user interface. This guide uses vSphere to supply the hypervisor that will create the virtual workstations and desktops.

1.1.3 Citrix XenDesktop

The following Citrix XenDesktop technologies are included in this guide:

- ▶ Citrix® StoreFront™ - A web based, self-service portal for users to access their virtual desktops and other resources.
- ▶ Citrix Studio - a centralized management console for administrators to create deployment groups of virtual workstations and desktops to be consumed.
- ▶ Citrix License server - A server to audit and provision licenses for the XenDesktop environment.
- ▶ Citrix Virtual Desktop Agent (VDA) - Software that resides within the virtual workstation or desktop and provides access to the end user.
- ▶ Citrix Receiver - Client software enabling the end user to connect to virtual desktops and other resources.
- ▶ Citrix Director - A web based, daily management console for administrators and help desk staff to view and take action on user sessions as well as environment reports.
- ▶ Citrix Delivery Controller - A server to broker and manage user connections and environment configuration details.

Figure 1-1 shows the key components of a XenDesktop deployment.

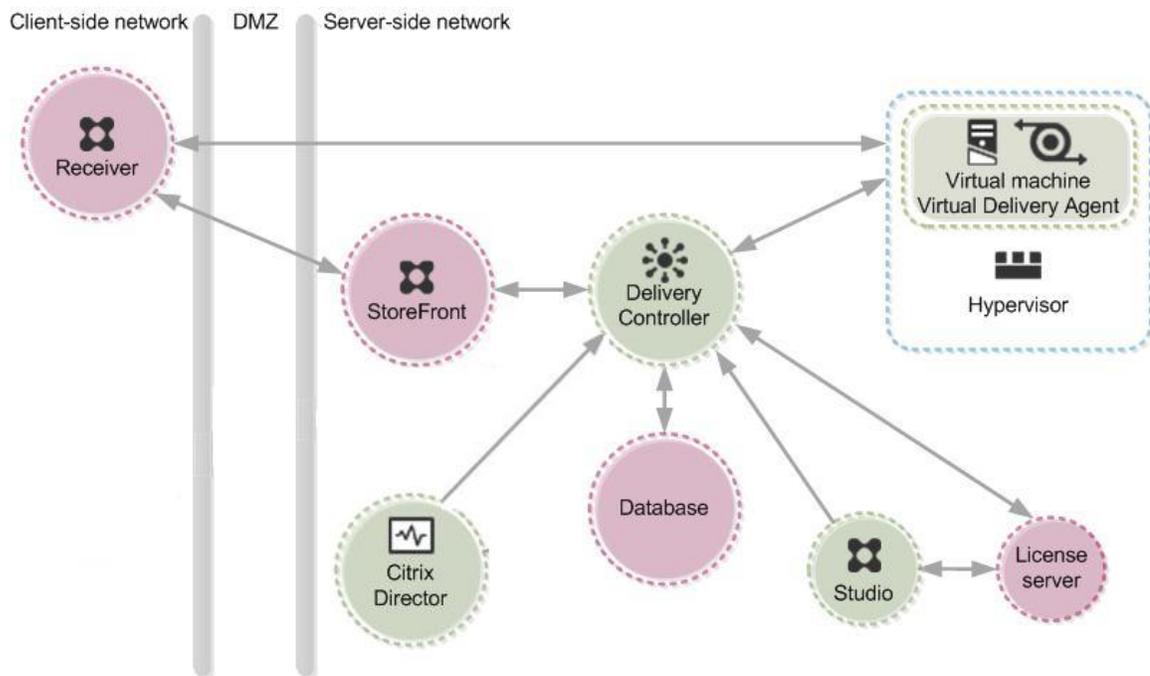


Figure 1-1 Citrix XenDesktop components diagram

1.2 ORGANIZATION

This manual contains the following chapters.

- ▶ Chapter 1 – *About this Document*: Explains the conventions used in this document, related documentation, and a list of acronyms.
- ▶ Chapter 2– *Getting Started*: Introduces you to NVIDIA GRID technology and guides you through preparing to deploy GRID with Citrix XenDesktop.
- ▶ Chapter 3 - *Installing VMware ESXi 6.0 U2*: Guides you through preparing a bootable USB drive and installing vSphere 6 on the selected host.
- ▶ Chapter 4 – *Installing VMware vCenter Server 6.0 U2*: Instructions for installing and configuring VMware vCenter 6 Server, either Windows or OVA Linux virtual appliance.
- ▶ Chapter 5– *GRID software preparation*: Explains how to build a license server, obtain licenses, and register for NVIDIA Support, Update, and Maintenance Subscription (SUMS)
- ▶ Chapter 6 – *NVIDIA GRID vGPU Manager Installation*: Step-by-step guide to downloading, preparing, uploading, installing, or uninstalling an NVIDIA GRID vGPU Manager VIB file.
- ▶ Chapter 7 – *Building Citrix XenDesktop 7.12*: Instructions for installing and configuring Citrix XenDesktop 7.12.
- ▶ Chapter 8 – *Selecting GRID vGPU profiles*: Understanding the range of available profiles and how to select the correct ones for your user groups.
- ▶ Chapter 9 – *Application Based Host Sizing Example – Esri*: give basic guidance on how to size host resources to accommodate a typical GPU dependent application.
- ▶ Chapter 10 – *Creating your first GRID Virtual workstation* : Guides you through configuring a GRID vGPU for a single guest desktop to create a template.
- ▶ Chapter 11 – *Creating XenDesktop Machine Catalogs and Delivery Groups*: Using Citrix Studio to create a random pool of GRID vGPU powered desktops using a single GRID vGPU profile.
- ▶ Chapter 12 – *Testing, Monitoring, & Benchmarking*: Describes procedures for testing and benchmarking your NVIDIA GRID deployment.
- ▶ Chapter 13 – *Known Issues*: Lists known issues related to vCenter, NVIDIA, and hardware.
- ▶ Chapter 14 – *Troubleshooting*: Provides a procedure for resolving common issues.
- ▶ Chapter 15 – *Using WINSXP*: Describes how to use WINSXP to transfer files.

1.3 FORMATTING CONVENTIONS

This document uses several formatting conventions to present information of special importance.

Lists of items, points to consider, or procedures that do not need to be performed in a specific order appear in bullet format:

- ▶ Item 1
- ▶ Item 2

Procedures that must be followed in a specific order appear in numbered steps:

1. Perform this step first.
2. Perform this step second.

Specific keyboard keys are bolded and capitalized, for example: **ESC**. If more than one key should be pressed simultaneously, it is displayed as **KEY1+KEY 2**, for example **ALT+F4**.

Interface elements such as document titles, fields, windows, tabs, buttons, commands, options, and icons appear in bold or italics text.

Menus and submenus have the notation **Menu→Submenu**. For example, “Select **File→Save**” means that you should first open the **File** menu, and then select the **Save** option.

Specific commands appear in standard Courier font. Sequences of commands appear in the order in which you should execute them and include horizontal or vertical spaces between commands. The following additional formatting also applies when discussing command-line commands:

Plain-text responses from the system appear in bold **Courier** font.

This manual also contains important safety information and instructions in specially formatted callouts with accompanying graphic symbols. These callouts and their symbols appear as follows throughout the manual:



CAUTION: Cautions alert you to the possibility of a serious error, data loss, or other adverse condition



Notes provide helpful information

1.4 RELATED DOCUMENTATION

If you're just getting started with NVIDIA GRID, see [NVIDIA GRID Technology](#) for basic information.

Refer to [Resources for NVIDIA GRID](#) for additional information about NVIDIA GRID technology, including:

- ▶ Datasheets
- ▶ Guides and Solution Overviews
- ▶ Case Studies
- ▶ White Papers
- ▶ Videos

For a list of GRID certified servers, see [Certified Servers](#).

For information on VMware vSphere and ESXi installation check the documentation at [About vSphere Installation and Setup](#).

1.5 THE DEPLOYMENT GUIDE LAB

To write this guide, test the steps necessary, gather the screenshots, and research what would become the contents, we used the following servers:

Table 1-1 Lab Servers

Role	Manufacturer	GPU	CPU	Memory
Management host	Supermicro	None	Intel Xeon E5-1650 v3 @ 3.50 GHz	64 GB
vGPU host	HPE	2 X NVIDIA Tesla M60	Intel Xeon CPU E5-2697 v4 @ 2.30GHz	768 GB

The Management server hosted the following services in virtual machines (VM).

- ▶ VMware vCenter Server Appliance (VCSA), Linux based
- ▶ Microsoft Windows Active Directory server
- ▶ VMware vCenter Server on Windows
- ▶ NVIDIA GRID License server
- ▶ Citrix XenDesktop components

The production server was used to host the virtual workstations and desktops, which were based on Microsoft Windows 10 Professional x64. The management server hosted the infrastructure components (Active Directory, Citrix XenDesktop, GRID License Server) which were based on Windows Server 2012 R2, excepting the VMware vCenter Server Appliance, a Linux based system.

Centralized storage provides a host of benefits to a virtualized environment, but they can frequently be slower than local disk. Disk latency on a SAN can be problematic as well, causing overall application delays. We used a Dell EMC XtremeIO Storage device connected with 10Gb Ethernet. We found this provided adequate speed and low latency while also giving us a centralized storage environment.

It is not necessary to have all the listed equipment building a POC environment; in fact, it can be done on a single physical server. This would not, however, be representative of a recommended enterprise deployment. To give your POC the best chance of wide adoption and general favorability, build the POC as close as possible to production design. Here are some guidelines we've found useful for planning your POC:

- ▶ Given that CPU is the most common bottleneck in many visualization applications, it is strongly recommended to avoid putting non-GPU workloads on a GPU equipped host. This allows those resources to be dedicated to provisioning guests with the performance of NVIDIA GRID.
- ▶ By separating the management and production (or virtual desktop) clusters, administrators can retain foundational services and exported virtual machines while reinstalling new versions of code.

CHAPTER 2. GETTING STARTED

NVIDIA GRID software enables multiple virtual machines (VMs) to have simultaneous, direct access to pools of physical, data center proven NVIDIA Tesla GPUs, using a microhypervisor installed within the host hypervisor stack. It leverages a feature rich SDK to provide monitoring and management, and the same proven NVIDIA graphics drivers that are deployed on non-virtualized operating systems. This gives VMs unparalleled graphics performance and application compatibility, together with cost-effectiveness and scalability. This is brought about by sharing a GPU among multiple VMs.

This chapter covers how NVIDIA GRID fundamentally alters the landscape of desktop virtualization. It describes the NVIDIA GRID architecture, the types of virtual GPUs supported, and key standards supported by GRID.

This guide focuses on the deployment of full virtual desktops as opposed to streamed applications or other solutions such as Citrix XenApp®. Citrix XenApp is supported on NVIDIA GRID, but is outside the scope of this guide.

2.1 WHY NVIDIA GRID?

As IT seeks to fully realize the power of virtualization, end user computing has remained a stumbling block. Organizations want the benefits of greater elasticity of scale, data security, and simplified manageability. However, adoption of desktop virtualization solutions has been limited by their failure to deliver a consistently great user experience that's equivalent to physical PC's and workstations.

Increasingly, today's operating systems like Microsoft Windows 10, and modern applications including office productivity tools, browsers, and multimedia, demand

levels of graphics performance and user experience not traditionally supported by desktop and application virtualization. NVIDIA GRID addresses the deficiencies of VDI by bringing graphics acceleration to the virtualized data center, enabling IT to extend the reach of virtualization to every user, from task workers to designers, delivering unparalleled NVIDIA graphics performance and a great user experience that rivals physical PC's.

This new “graphics-accelerated data center” is the realization of a fully-virtualized end user computing environment, where NVIDIA GPU technology powers every desktop, every app, and every workstation, making virtualized workspaces accessible to an enterprise's entire workforce. NVIDIA GRID brings the power of NVIDIA GPU to virtual desktops, applications, and workstations, with an engaging user experience for everyone, from office workers to mobile professionals to designers. GRID technology meets the rising performance demands of modern apps and Windows 10, bringing graphics acceleration to the data center, enabling IT to centralize apps and data, and deliver virtual workspaces that offer improved management security, and productivity. Business can deliver cost-effective VDI performance that scales, transforming workflows and liberating users from the confines of PC's and offices, collaborating in real-time, from anywhere, using any device. IT can now virtualize 100% of its users with a native experience, achieving dramatically lower cost per workstation, with manageability that ensures a consistently great user experience. NVIDIA GRID improves security, ensuring data and intellectual property are centrally stored and protected, while enabling business partners to securely collaborate remotely.

2.2 NVIDIA GRID ARCHITECTURE

Using GRID software, NVIDIA Tesla GPUs can be virtualized, making them capable of supporting multiple *virtual* GPU devices (vGPUs) that can be assigned directly to VMs. The VMs present the vGPU to its guest OS as an instance of the shared physical GPU. The guest OS inside the VM sees the vGPU as if it were a physical GPU, down to the model number. The Guest uses the vGPU in the same way it would a physical GPU that has been passed through by the hypervisor. An NVIDIA driver loaded into the VM's guest OS leverages direct access to the GPU for performance and critical fast paths. Non-critical performance management operations use a para-virtualized interface to the GRID vGPU Manager. Figure 2-1 shows the NVIDIA GRID system architecture.

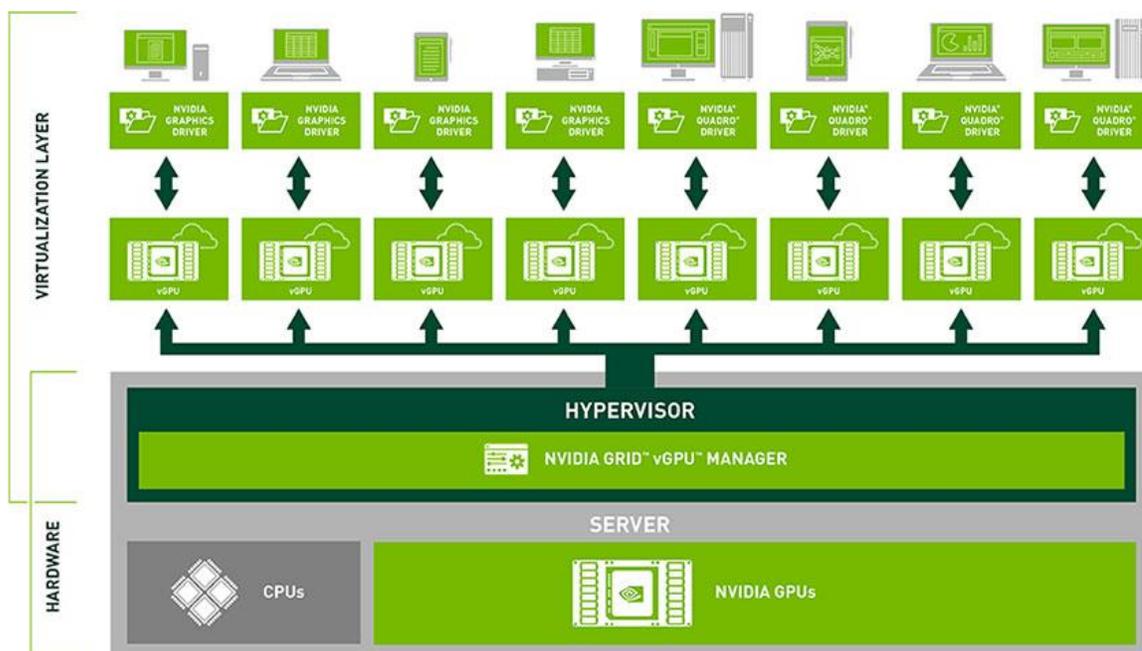


Figure 2-1 NVIDIA GRID System Architecture

GRID vGPUs are comparable to conventional GPUs in that they have a fixed amount of GPU frame buffer and one or more virtual display outputs or *heads*. Multiple heads support multiple displays. Managed by the NVIDIA GRID vGPU Manager software installed in the hypervisor, the vGPU frame buffer is allocated out of the physical GPU frame buffer at the time the vGPU is created. The vGPU retains exclusive use of that frame buffer until it is destroyed.

All vGPUs resident on a physical GPU share access to the GPU's engines, including the graphics (3D), and video decode and encode engines. The physical GPU has an internal scheduler that assigns all the GPU engine resources to a particular VM's graphics requests while that VM is scheduled. The input and output data for these requests are stored in the frame buffer assigned to that VM, so that one VM doesn't overwrite another's data. Figure 2-2 shows the vGPU internal architecture.

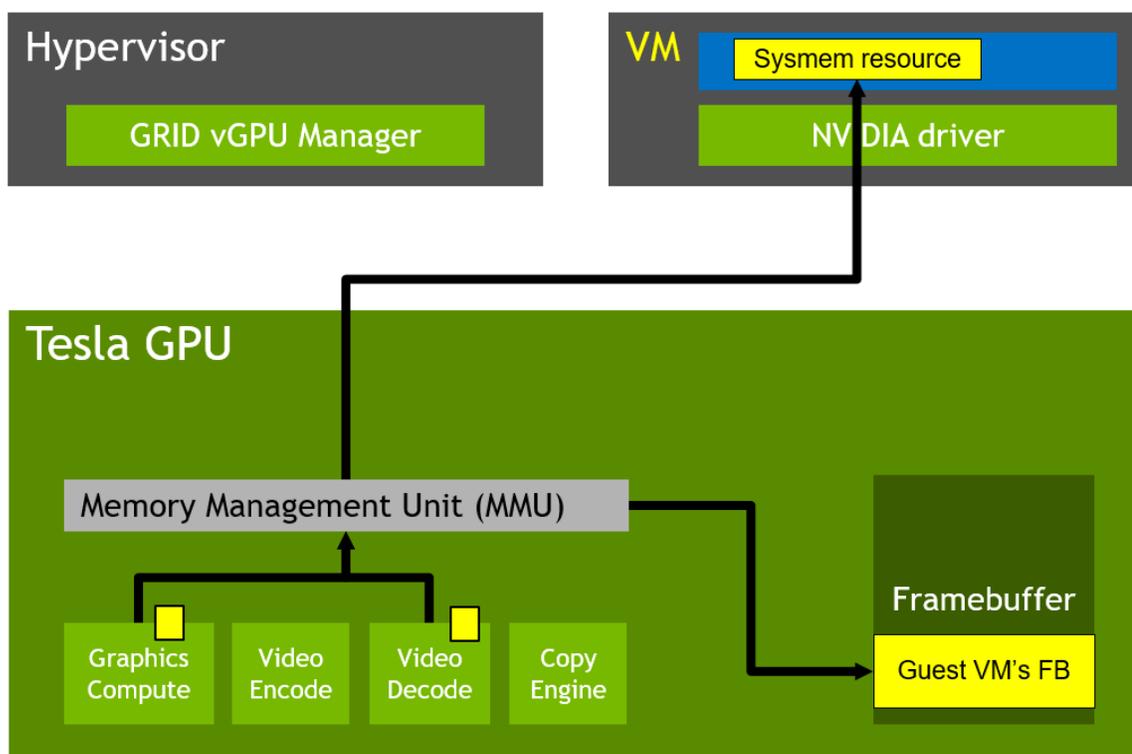


Figure 2-2 NVIDIA GRID vGPU Internal Architecture

Another innovation of the GRID architecture is support for NVIDIA Encode (NVENC) video encode acceleration with Citrix HDX 3D Pro. NVENC allows the hypervisor to use the physical GPU when encoding H.264 video output for transmission. Without NVENC, the server would use extra CPU cycles in the process of encoding the video stream. With NVENC, remote desktop software can more quickly encode frames, leading to increased performance for hosted applications. Figure 2-3 gives an overview of NVENC operation.

To find out more about NVENC, please refer to the following website:

[NVIDIA VIDEO CODEC SDK](#)

And for Citrix documentation:

[GPU acceleration for Windows Desktop OS](#)

NVENC OVERVIEW

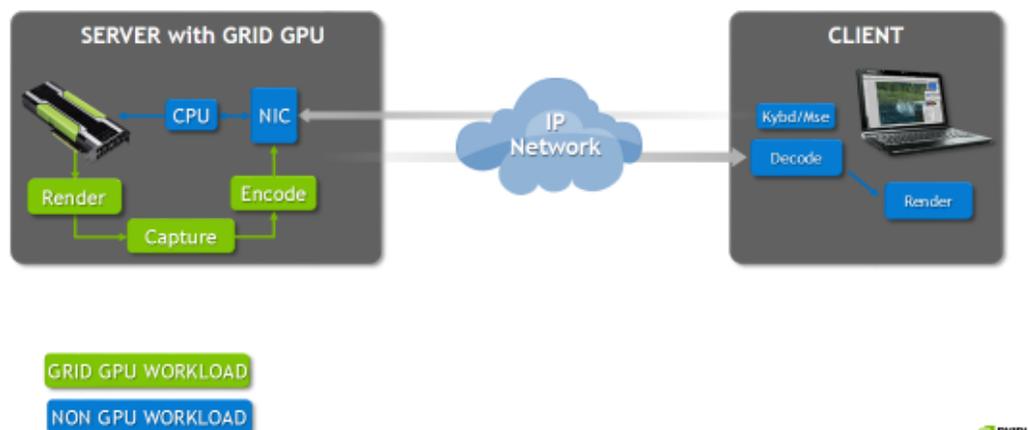


Figure 2-3 NVENC overview

2.3 SUPPORTED GPUS

NVIDIA GRID software is supported on Tesla M60 and Tesla M10 boards, both of which implement multiple physical GPUs, as well as the Tesla M6 single GPU MXM for blades:

- ▶ Tesla M60 is a PCI Express board with two high-end Maxwell GPUs, optimized for high performance
- ▶ Tesla M10 is a PCI Express board with four mid-tier Maxwell GPUs, optimized for high density
- ▶ Tesla M6 is an MXM board with a single high-end Maxwell GPU optimized for use in blade servers.

The Tesla Maxwell series succeeds the original GRID K1 and K2 cards, and provides several new features:

- ▶ Maximum screen resolutions up to 4096 X 2160
- ▶ Increased user density
- ▶ Support for NVIDIA® CUDA® and *OpenCL™ software*.
- ▶ Blade server MXM support (M6)
- ▶ H.264 encoding on the GPU
- ▶ Monitoring and management

► Linux support

The Tesla M6 is a single GPU blade optimized MXM module. The Tesla M60 and Tesla M10 are full height, full length, double width PCI-e cards, and are typically passively cooled (no onboard fan) requiring certified servers designed to properly power and cool them. Refer to [NVIDIA Certified Servers](#) to see all supported server platforms and cross-reference the NVIDIA list against the VMware vSphere HCL for servers. This list continues to grow rapidly, check back often for more certified platforms.

Each card requires auxiliary power cables connected to it, see the specification charts for specific pin counts. Most servers require an enablement kit for proper mounting of the Tesla and GRID GPU cards. Check with your server OEM of choice for more specific requirements.

GRID can also be run on the first-generation GRID K1 and K2 GPU cards, however they don't benefit from GRID software features.

For more information, see the [NVIDIA GRID datasheet](#).

2.4 TESLA M60/M6 SPECIFICATIONS

Table 2-1 lists the Tesla M60 and M6 hardware specifications.

Table 2-1 Tesla M60 and Tesla M6 Hardware Specifications

	Tesla M60	Tesla M6
Number of GPUs	2	1
Total NVIDIA CUDA Cores	2048 per GPU (4096 total)	1536
Total Memory Size	8GB GDDR5 per GPU (16GB total)	8 GB GDDR5
Max Power	300 W	100 W
Board Length	10.5	4.134
Board Height	4.4	3.228
Board Width	Dual slot	MXM 3.1 type B form factor
Display IO	None	None

Aux Power	8-pin connector	8-pin connector
PCIe	x16	x16
PCIe Generation	Gen3 (Gen2 compatible)	Gen3 (Gen1 compatible)
Cooling Solution	Passive	Passive
Complete Technical Specifications	See https://images.nvidia.com/content/pdf/tesla/tesla-m60-product-brief.pdf	See http://images.nvidia.com/content/pdf/tesla/tesla-m6-product-brief.pdf
Note: Refer to http://www.nvidia.com/object/grid-technology.html for more specifications.		

2.5 TESLA M10 SPECIFICATIONS

Table 2-2 lists the Tesla M10 hardware specifications.

Table 2-2 Tesla M10 Hardware Specifications

	Tesla M10
Number of GPUs	4
Total NVIDIA CUDA Cores	640 per GPU (2560 total)
Total Memory Size	8GB GDDR5 per GPU (32GB total)
Max Power	225 W
Board Length	10.5
Board Height	4.4
Board Width	Dual slot
Display IO	None
Aux Power	8-pin connector
PCIe	x16
PCIe Generation	Gen3 (Gen2 compatible)
Cooling Solution	Passive
Complete Technical Specifications	See http://images.nvidia.com/content/pdf/tesla/Tesla-M10-Product-Brief.pdf
Note: Refer to http://www.nvidia.com/object/grid-technology.html for more specifications.	

2.6 GRID K1 / K2 SPECIFICATIONS

Table 2-3 lists the GRID K1 and K2 hardware specifications

Table 2-3 GRID K1/K2 Hardware Specifications

	GRID K1	GRID K2
Number of GPUs	4	2
Total NVIDIA CUDA Cores	192 per GPU (768 total)	1536 per GPU (3072 total)
Total Memory Size	4GB DDR3 per GPU (16GB total)	4 GB GDDR5 per GPU (8GB total)
Max Power	130 W	225 W
Board Length	10.5	10.5
Board Height	4.4	4.4
Board Width	Dual slot	Dual slot
Display IO	None	None
Aux Power	6-pin connector	8-pin connector
PCIe	x16	x16
PCIe Generation	Gen3 (Gen2 compatible)	Gen3 (Gen2 compatible)
Cooling Solution	Passive	Passive
Complete Technical Specifications	See http://www.nvidia.com/content/grid/pdf/GRID_K1_BD-06633-001_v02.pdf	See http://www.nvidia.com/content/grid/pdf/GRID_K2_BD-06580-001_v02.pdf
Note: Refer to http://www.nvidia.com/object/grid-technology.html for more specifications.		

2.7 GRID vGPU PROFILES

Allocation of physical GPU frame buffer to virtual GPUs is controlled by vGPU profiles. Selection of a profile defines the amount of graphics frame buffer memory assigned to the vGPU, and subsequently how many virtual desktops can share that physical GPU. For example, a Tesla M60 physical GPU can support up to eight Tesla M60-1q vGPUs on each of its two physical GPUs (for a total of 16 vGPUs), but only two Tesla M60-4q vGPUs (for a total of eight vGPUs). You cannot oversubscribe frame buffer and it must be shared equally for each physical GPU, but you can select different profiles for each GPU. Like a GRID K2 card, the Tesla M60 with its two physical GPUs can support up to two different profiles, while the Tesla M10 can support up to four. See *Homogenous GRID vGPUs* on page 25 for more detail.

Each physical GPU can support several different types of vGPU, each having a fixed amount of frame buffer, number of supported display heads and maximum resolutions, and targeted at different classes of workload.

Table 2-4 lists the GRID vGPU types supported on Tesla M60, M6, M10, with GRID software, and the GRID K1 and K2 cards. It also shows the GRID software versions for each vGPU profile: Virtual Workstation (vWS), Virtual PC (vPC), or Virtual Applications (vApps).

Table 2-4 GRID vGPU Types

Card	Physical GPUs	GRID Software Version (See section 2.10 GRID Licensing)	vGPU Profile	Frame Buffer (MB)	Virtual Display Heads	Max Resolution	Maximum vGPUs	
							Per GPU	Per Board
Maxwell Architecture								
Tesla M60 (PCI-e)	Two	vWS	GRID M60-8Q	8192	4	4096x2160	1	2
			GRID M60-4Q	4096	4	4096x2160	2	4
			GRID M60-2Q	2048	4	4096x2160	4	8
			GRID M60-1Q	1024	2	4096x2160	8	16
			GRID M60-0Q	512	1	2560x1600	16	32
		vPC	GRID M60-1B	1024	4	2560x1600	8	16
			GRID M60-0B	512	1	2560x1600	16	32
		vApps	GRID M60-8A	8192	1	1280x1024	1	2
			GRID M60-4A	4096	1	1280x1024	2	4
			GRID M60-2A	2048	1	1280x1024	4	8
			GRID M60-1A	1024	1	1280x1024	8	16
		Tesla M6 (MXM)	One	vWS	GRID M6-8Q	8192	4	4096x2160
GRID M6-4Q	4096				4	4096x2160	2	2

			GRID M6-2Q	2048	4	4096x2160	4	4		
			GRID M6-1Q	1024	2	4096x2160	8	8		
			GRID M6-0Q	512	1	2560x1600	16	16		
		vPC	GRID M6-1B	1024	4	2560x1600	8	8		
			GRID M6-0B	512	1	2560x1600	16	16		
		vApps	GRID M6-8A	8192	1	1280x1024	1	1		
			GRID M6-4A	4096	1	1280x1024	2	2		
			GRID M6-2A	2048	1	1280x1024	4	4		
			GRID M6-1A	1024	1	1280x1024	8	8		
Tesla M10 (PCI-e)	Four	vWS	GRID M10-8Q	8192	4	4096x2160	1	4		
			GRID M10-4Q	4096	4	4096x2160	2	8		
			GRID M10-2Q	2048	4	4096x2160	4	16		
			GRID M10-1Q	1024	2	4096x2160	8	32		
			GRID M10-0Q	512	1	2560x1600	16	64		
		vPC	GRID M10-1B	1024	4	2560x1600	8	32		
			GRID M10-0B	512	1	2560x1600	16	64		
		vApps	GRID M10-8A	8192	1	1280x1024	1	4		
			GRID M10-4A	4096	1	1280x1024	2	8		
			GRID M10-2A	2048	1	1280x1024	4	16		
			GRID M10-1A	1024	1	1280x1024	8	32		
		Kepler architecture								
		GRID K2	Two	GRID 2.0 features	GRID K280Q	4096	4	2560x1600	1	2

(PCI-e)		not available	GRID K260Q	2048	4	2560x1600	2	4
			GRID K240Q	1024	2	2560x1600	4	8
			GRID K220Q	512	2	2560x1600	8	16
GRID K1 (PCI-e)	Four	GRID 2.0 features not available	GRID K180Q	4096	4	2560x1600	1	4
			GRID K160Q	2048	4	2560x1600	2	8
			GRID K140Q	1024	2	2560x1600	4	16
			GRID K120Q	512	2	2560x1600	8	32

2.8 SUPPORTED GRAPHICS PROTOCOLS

This release of GRID includes support for:

- ▶ DirectX 12, Direct2D, and DirectX Video Acceleration (DXVA)
- ▶ OpenGL 4.5
- ▶ NVIDIA Capture SDK 5.0 (remote graphics acceleration)

CUDA and OpenCL are supported on these vGPUs:

- ▶ GRID M60-8Q
- ▶ GRID M10-8Q
- ▶ GRID M6-8Q
- ▶ GRID M60-8A
- ▶ GRID M10-8A
- ▶ GRID M6-8A

2.9 HOMOGENOUS GRID VGPUS

To maintain predictable and stable performance, GRID software supports homogeneous vGPUs on each specific physical GPU. At any given time, the vGPU profiles running on a single physical GPU must be identical. However, this restriction does not extend to physical GPUs on the same card. Like the GRID K1 and K2 before, each physical GPU on

a Tesla M60 or Tesla M10 may host different types of vGPUs at the same time. For example, a Tesla M10 card has four physical GPUs and can support up to four different types of vGPU, one pre-physical GPU, allowing for great flexibility. Figure 2-4 shows an example of vGPU distribution on the Tesla M60.



In the case of the blade optimized Tesla M6 there is only one GPU per MXM, so one profile type can be assigned, but with multiple MXM's in a blade one additional profile can be chosen for each.

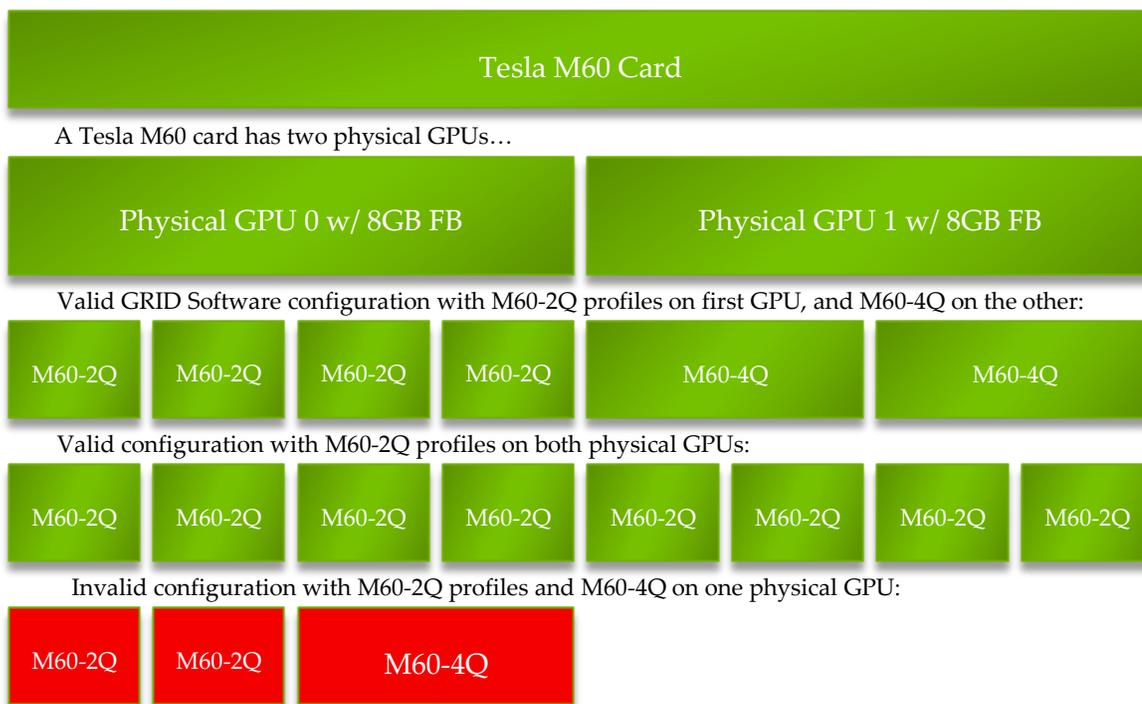


Figure 2-4 Example of GRID vGPU Configurations on Tesla M60

2.10 GRID LICENSING

These NVIDIA GRID products are available as licensed products on NVIDIA Tesla GPUs:

- ▶ Virtual Workstation (vWS)

- ▶ Virtual PC (vPC)
- ▶ Virtual Applications (vApps)

GRID Virtual PC

This product is ideal for users who want a virtual desktop but need great user experience leveraging Windows applications, browsers, and high definition video. NVIDIA GRID Virtual PC delivers a native experience to users in a virtual environment, allowing them to run all their PC applications at full performance.

GRID Virtual Workstation

This edition is ideal for mainstream and high-end designers who use powerful 3D content creation applications such as Dassault's CATIA and SOLIDWORKS, 3DExcite, Siemens NX, PTC Creo, Schlumberger Petrel, and Autodesk Maya. NVIDIA GRID Virtual Workstation allows users to access their professional graphics applications with full features and performance, anywhere, on any device, with all the features and performance delivered by the fully tested and supported NVIDIA Quadro driver.

GRID Virtual Applications

For organizations deploying XenApp or other RDSH solutions. It delivers Windows applications at full performance. NVIDIA GRID Virtual Applications allows users to access any Windows application at full performance on any device, anywhere. This edition is suited for users who would like to virtualize applications using XenApp or other RDSH solutions.

Table 2-5 summarizes the GRID License editions

Table 2-5 GRID License Editions

GRID License Edition	GRID Features
GRID Virtual Application	<input type="checkbox"/> Virtual GPUs for virtual application computing
GRID Virtual PC	<input type="checkbox"/> Virtual GPUs for business desktop computing
GRID Virtual Workstation	<input type="checkbox"/> Virtual GPUs for midrange and high-end workstation computing <input type="checkbox"/> Workstation graphics on GPU pass-through

Licenses can be purchased from NVIDIA (<http://www.nvidia.com/buygrid>) or customers may evaluate the licensing for a trial period. Customers who purchase licenses will receive an email with direction on how to download and install their new

licenses. Evaluation customers will immediately see a link connecting them to their evaluation keys.

Once you've obtained your license, the GRID License Manager will help you manage distribution and compliance. The GRID License Manager provides monitoring and reporting on license usage for capacity planning and is included with the GRID software packages. This License Manager can be installed on either a physical server or, more likely, a dedicated virtual machine.

The following Figure 2-5 provides an overview of the GRID License management system:

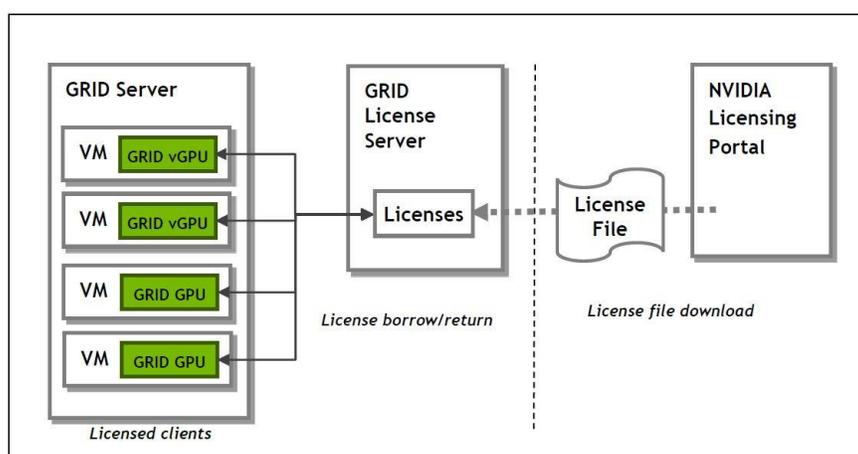


Figure 2-5 GRID Licensing management system

When enabled on Tesla GPUs, licensed editions of GRID products are activated by obtaining a license over the network from an NVIDIA GRID License Server. The license is “checked out” or “borrowed” at the time the Virtual Machine (VM) is booted, and returned when the VM is shut down.

GRID licenses come in three editions that enable different classes of GRID features. The GRID software automatically selects the right license edition based on the features being used.

For more information on NVIDIA GRID Licensing, see the [NVIDIA GRID Packaging and Licensing](#).



Note: NVIDIA GRID and vGPUs require licensing for EULA compliance. To improve end user experience, the GRID software will run with or without a valid license server connection. NVIDIA GRID K1 and K2 GPUs do not require a license to run vGPU.

2.11 GRAPHICS VS. COMPUTE MODE

NVIDIA Tesla GPUs are shipped in a configuration optimized for graphics applications. However, they can also be configured for Compute mode, as listed in Table 2-6.

Table 2-6 Compute Mode Settings

Setting	Value	Notes
Classcode	3D Controller	This classcode indicates to operating systems (OS) that the GPU is not intended for use as a primary display device.
Memory BAR	8 gigabytes	Tesla GPUs expose a large memory base address register (BAR) for direct access to the frame buffer from the CPU, and other PCI Express devices.
I/O base BAR	Disabled	The GPU need not consume any legacy I/O resources when used as a non-display device.
ECC protection	Enabled	Error Correcting Code (ECC) is enabled on the GPU frame buffer to protect against single- and multi-bit memory errors.

Collectively this GPU configuration is referred to as **compute mode**. While compute mode is optimal for single-precision compute usage, it can cause compatibility problems with OS and hypervisors when the GPU is used primarily as a graphics device:

- ▶ Some OS require that the GPU advertise a VGA display controller classcode for the GPU to be used as a primary graphics device.
- ▶ Some hypervisors cannot support vDGA/pass-through of GPUs with large memory BARs to guest virtual machines.

To address these problems, certain NVIDIA Tesla GPUs support **graphics mode**. The settings for graphics mode are listed in Table 2-7.

Table 2-7 Graphics Mode Settings

Setting	Value	Notes
Classcode	VGA Controller	This classcode indicates to OS that the GPU can function as a primary display device.
Memory BAR	256 megabytes	The GPUs a smaller memory BAR for direct access to the frame buffer.
I/O base BAR	Enabled	The GPU exposes an I/O BAR to claim the resources required to operate as a VGA controller.

Setting	Value	Notes
ECC protection	Disabled	ECC protection is disabled by default, though it can still be enabled by use of the nvidia-smi management tool

The mode of the GPU is established directly at power-on, from settings stored in the GPU's non-volatile memory. **gpumodeswitch** is the software used to program the mode of the GPU by updating the GPU's non-volatile memory settings.

To learn more about graphics vs. compute mode, see the [gpumodeswitch user guide](#).



Tesla M10 GPU Accelerators do not enable Compute mode. These are Graphics mode only cards.



For the purposes of this guide, all GPU will be set to graphics mode.

2.12 DISTRIBUTION OF GRID VGPU

On each host the GRID vGPU guests should be distributed across all available physical GPUs, ensuring the guests get the maximum available performance, versus loading up the first physical GPU before moving on to the next to ensure maximum density. This is referred to as breadth-first over depth-first distribution.

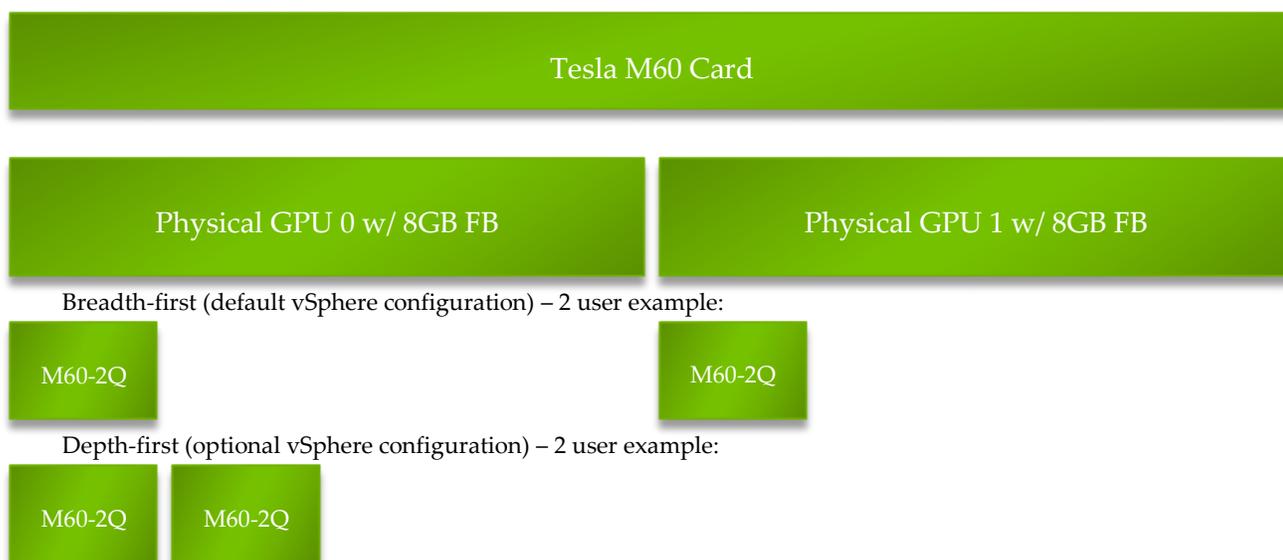


Figure 2-6 Example of GRID vGPU Distribution on Tesla M60 Card

Architecturally the challenge comes with guests of different profiles but for the same card type, for example the M60-1Q and M60-8Q are both M60 profiles. In this case, it would be possible for the pool of M60-1Q guests to be spread across all available GPUs on a host, leaving no free GPUs for the M60-8Q pool. If there is another M60 host available, then the second pool can start placing guests there. Careful calculation of available hosts and GPUs, or assigning hosts to profile specific clusters, is necessary to ensure all guests have the necessary resources to power up.

2.13 THE BEFORE YOU BEGIN

This section describes the general prerequisites and some general preparatory steps that must be addressed before proceeding with the deployment.

2.13.1 General Prerequisites

The following elements are required to install and configure HDX 3D Pro with vGPU on VMware vSphere 6:

- ▶ Evaluation plan consisting of all the following:
- ▶ Listed business drivers and goals
- ▶ List of all user groups, their workloads, and applications

- ▶ Current end-user experience measurements and analysis
- ▶ ROI/density goals
- ▶ Review Citrix XenDesktop documentation

2.13.2 Build an Evaluation Plan

To be successful, we highly recommend you start with evaluation plan consisting of the following:

- ▶ Listed business drivers and goals with cost benefit analysis
- ▶ List of all user groups, their workloads, and applications
- ▶ Current end-user experience measurements and analysis
- ▶ ROI/density goals

2.13.3 Size Your Environment

Based on your evaluation plan, we recommend sizing an appropriate environment for each user group you are trying to reach with your evaluation. Use the system requirements for the most intensive graphics apps, as well as existing physical workstations as reference points. Often server performance hits the CPU bottleneck much before it consumes all available GPU power. Based on all this, make sure to include enough RAM, CPUs, and GPUs to meet your sizing needs.

2.13.4 Choosing Your Hardware

The following elements are required install and configure vGPU on VMware vSphere 6:

- ▶ VMware and NVIDIA certified servers with NVIDIA Tesla M60, Tesla M6 or Tesla M10 cards (see [the NVIDIA Certified Server page](#) for a list of certified NVIDIA servers and cross check that against the VMware HCL) that meet the following recommended specifications:
 - ▶ 2.5GHz CPU or faster (Intel “Haswell” or later)
 - ▶ High-speed RAM
 - ▶ Fast networking
 - ▶ If using local storage IOPS plays a major role in performance.
- ▶ Select the appropriate NVIDIA GRID card for your use case, either Tesla M60 for higher performance solutions, Tesla M10 for density, or Tesla M6 for blade servers (refer to <http://www.nvidia.com/object/grid-boards.html> for details). For additional guidance, please contact your NVIDIA and Citrix reps.

- ▶ High performance storage with high IOPS SAN or similar
- ▶ Higher-performance end points for testing.

2.13.5 Choose Virtualization Components

- ▶ VMware vSphere 6.0 U2
- ▶ You may deploy vCenter Server 6 on a Windows server or as an OVA Appliance.
- ▶ Citrix XenDesktop 7.12

If needed, you may register for a free 90-day, 99-user trial to obtain the license keys, software, and quick-guides to deploy and manage XenDesktop at:

<http://www.citrix.com/tryxendesktop>

- ▶ NVIDIA GRID software and licenses:
- ▶ NVIDIA GRID vGPU Manager VIB
- ▶ NVIDIA WDDM guest driver (32-bit and 64-bit versions)



The vGPU Manager VIB is loaded like a driver in the vSphere hypervisor, and is then managed by the vCenter Server.

- ▶ Windows Server 2012 R2 (recommended) or Server 2008 R2 (minimum)
- ▶ Windows 10 Enterprise (for use as the guest operating system)
- ▶ Your choice of a CLI/SSH/SCP tools installed on your Windows-based toolbox PC, for example:
 - ▶ Putty (SSH) and WinSCP (SCP), available from <http://www.putty.org> and <http://www.winscp.net>, respectively. See Chapter 15, *Using WINSCP*, on page 295 for information on using WinSCP.
- ▶ Licenses:
 - ▶ NVIDIA GRID licenses (or [evaluation](#))
 - ▶ VMware vSphere 6.0 U2 (or evaluation)
 - ▶ Citrix XenDesktop 7.12 (or evaluation)
 - ▶ Microsoft OS (volume licenses recommended)
- ▶ Testing and Benchmarking:
 - ▶ NVIDIA System Management Interface
 - ▶ OPTIONAL: VMware vRealize Operations Manager
 - ▶ OPTIONAL: Citrix Director
 - ▶ OPTIONAL: Lakeside SysTrack 7 with GPU monitoring
- ▶ Server where you will install Citrix XenDesktop 7.12 Delivery Controller:

- ▶ Windows Server 2012 R2
- ▶ Server must be joined to a domain
- ▶ Server must have a static IP address assigned

2.13.6 Pre-Installation Preparation

Perform the following procedure prior to installation:

1. Determine how vSphere will run on the physical hosts. Consider booting from a thumb drive or SAN.
2. Download and install a SSH and SCP tool, examples are:
 - ▶ PuTTY (SSH)
 - ▶ WinSCP (SCP)
3. Set the BIOS on the physical hosts as appropriate. See 2.13.8, *Server BIOS Settings* on page 34.
4. Install Virtual Clone Drive or similar to easily mount and un-mount ISO images within the virtual machines.

2.13.7 Known Limitations

See Chapter 13, *Known Issues*, on page 278 for a list of known limitations that is current as of the publication date of this manual.

2.13.8 Server BIOS Settings

Configure the BIOS as appropriate for your physical hosts, as per direction from your OEM for a VMware vSphere environment with NVIDIA Tesla products.

Older versions of many hypervisors were often limited to 32-bit addressing (i.e. a 4GB limit) and as such required server BIOSs to be configured such that MMIO (Memory Mapped I/O) above 4GB is disabled. This advice applies to VMware ESXi 5.1 and earlier. Information on addressing limitation can be found in http://nvidia.custhelp.com/app/answers/detail/a_id/4119/~/incorrect-bios-settings-on-a-server-when-used-with-a-hypervisor-can-cause-mmio.

MMIO above 4GB should be enabled and may be necessary on many servers particularly when using NVIDIA GPUs with higher numbers of physical GPUs i.e. this is more likely to be necessary when using Tesla M10 cards (4 GPUs per card) than Tesla M60 cards (2 physical GPUs per card).

Some newer versions of VMware ESXi have raised the addressing limit to 44-bits but full 64-bit addressing is not supported in many versions of VMware ESXi. Information on the current MMIO access limits for specific versions of VMware ESXi can be found here: https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2087943, where it is documented for example that the Maximum Physical Address (MAXPA) is limited to 16TB in versions 6.0, 16TB (44-bit, 2 to the power of 44). However, care may need to be taken to ensure other BIOS settings prevent the use of MMIO address space above the 44-bit limit. An example is given in http://nvidia.custhelp.com/app/answers/detail/a_id/4119/~/incorrect-bios-settings-on-a-server-when-used-with-a-hypervisor-can-cause-mmio, where by a Supermicro server requires the MMIOHBase BIOS setting to be set to 2T to ensure memory indexed remains below the 16TB (44-bit) limit of VMware ESXi version 6.0. Customers are advised to consult their OEM server vendor for advice on appropriate BIOS settings.

CHAPTER 3. INSTALLING VMWARE ESXI 6.0 U2

This chapter covers the following vSphere 6 installation topics:

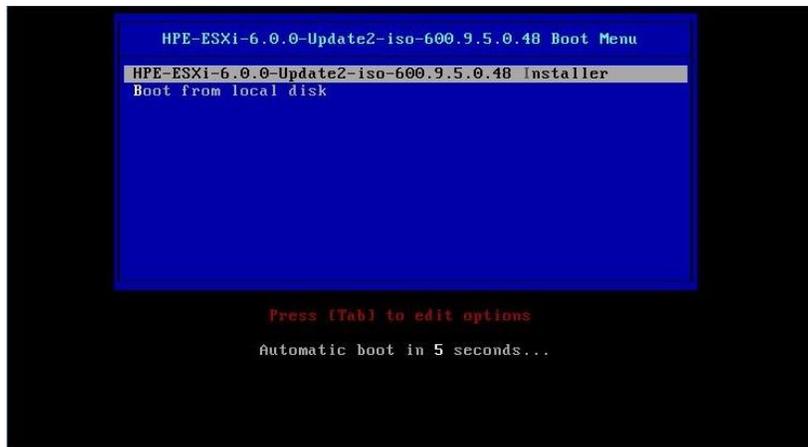
- ▶ Installing vSphere 6.0 U2
- ▶ Initial host configuration



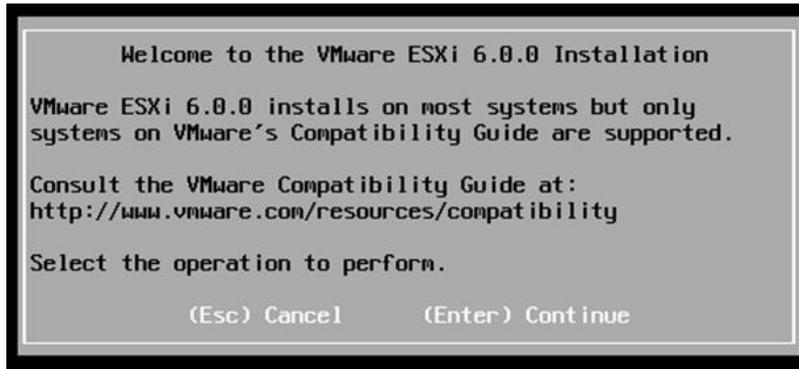
This deployment guide assumes you are building an environment as a proof of concept and is not meant to be a production deployment, thus choices made are meant to speed up and ease the process. For production use, consult the various documentation resources for each technology.

3.1 INSTALLING ESXI 6.0 U2

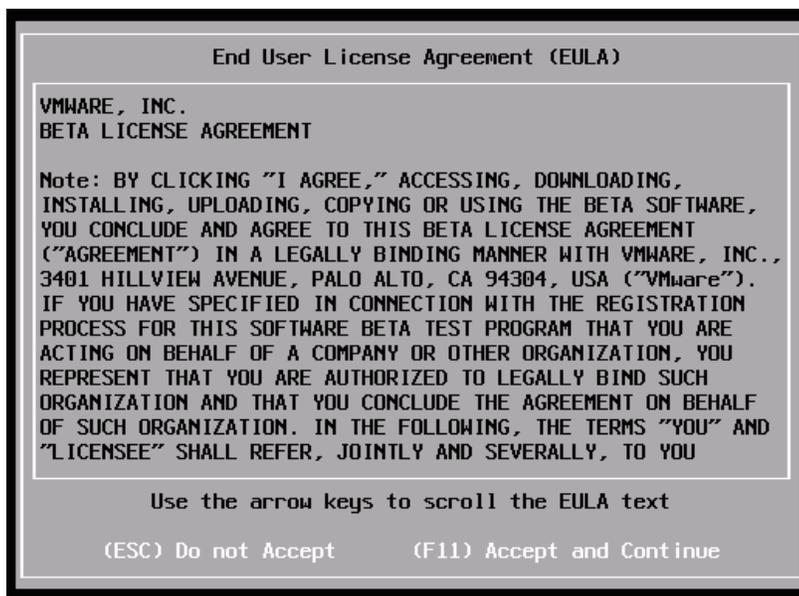
Use the following procedure to install the vSphere 6.0 U2. Select the boot media with the vSphere ISO on your host's boot menu.



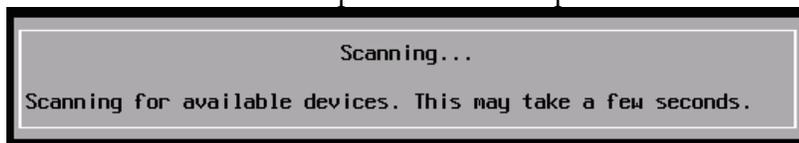
1. Apply power to start the host. The **boot menu** displays when the host starts up.



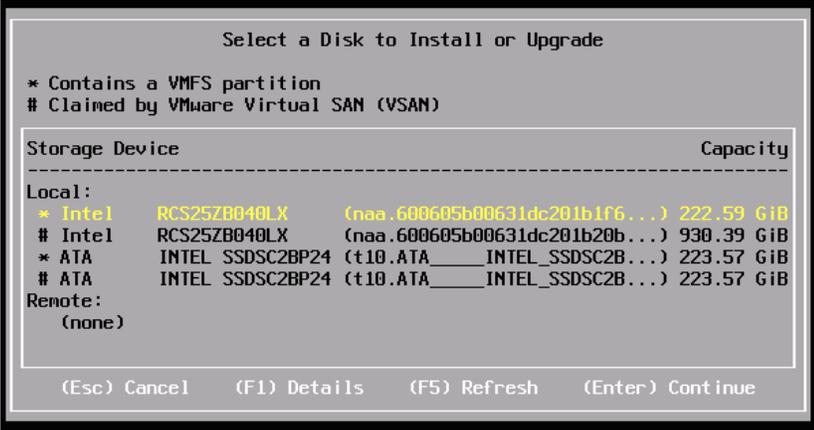
2. Once the installation process boots, the **Welcome** screen will be displayed. press **ENTER** to continue.



3. Read the EULA and then press **F11** to accept it and continue the installation.

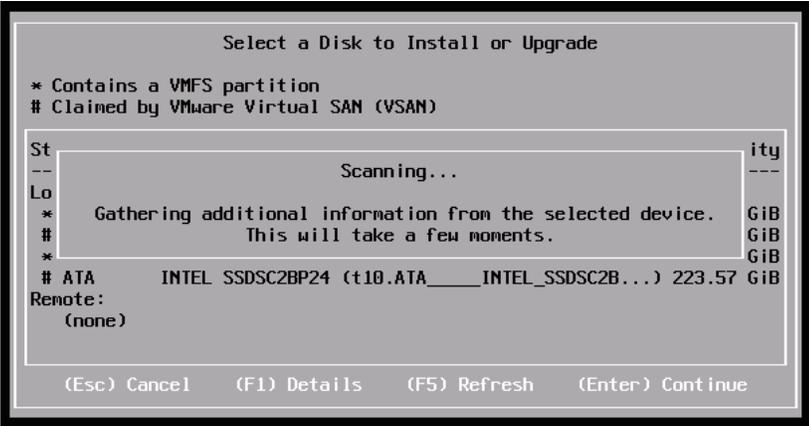


The installer scans the host to locate a suitable installation drive.

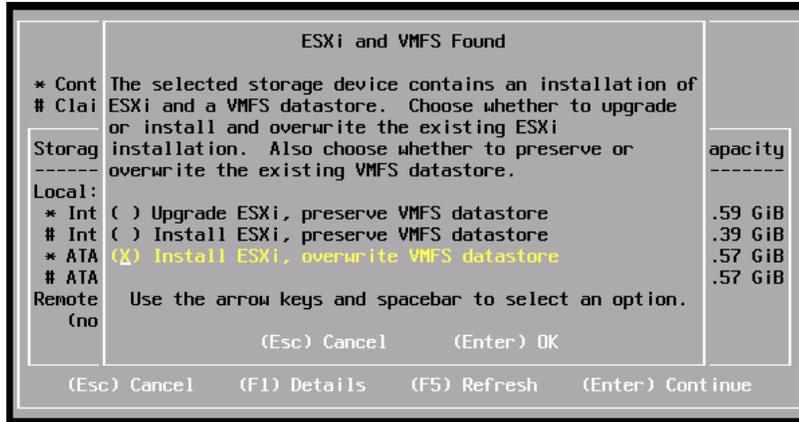


- Use the arrow keys to select the drive on which you will install vSphere 6 and then press **ENTER** to continue.

You can install vSphere 6 to a USB drive and then boot and run the system from that USB drive. This sample installation shows vSphere being installed on a local hard drive.



The installer scans the chosen drive to determine suitability for install:



5. The **Confirm Disk Selection** window may display, if a VMFS datastore has been found on the install device. Press **ENTER** to accept your selection and continue.



6. Select your desired keyboard layout using the arrow keys and then press **ENTER**.

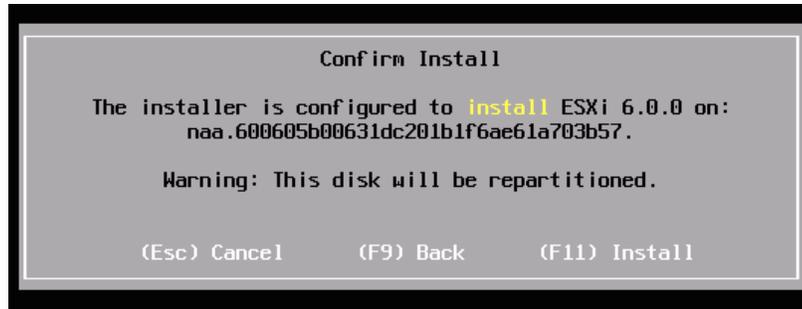


7. Enter a root password in the **Root password** field.



CAUTION: TO PREVENT UNAUTHORIZED ACCESS, YOUR SELECTED ROOT PASSWORD SHOULD CONTAIN AT LEAST EIGHT (8) CHARACTERS AND CONSIST OF A MIX OF LOWERCASE AND CAPITAL LETTERS, DIGITS, AND SPECIAL CHARACTERS.

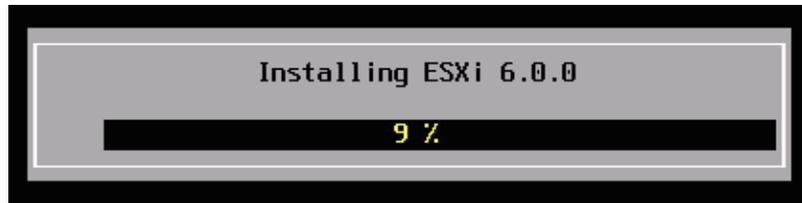
8. Confirm the password in the **Confirm password** field and then press **ENTER** to proceed.



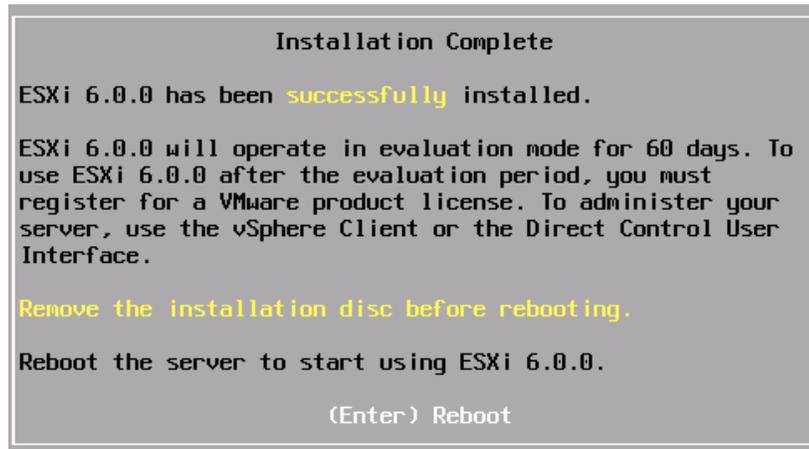
- In the **Confirm Install** window press **F11**.



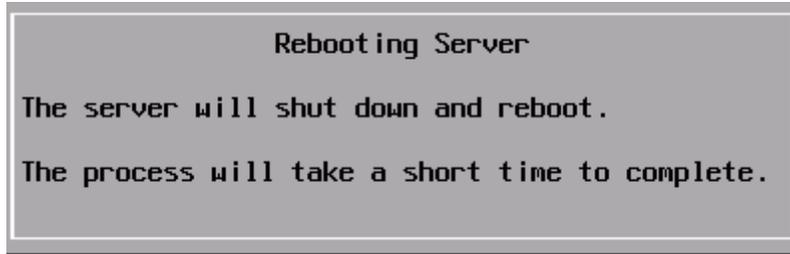
CAUTION: THE INSTALLER WILL REPARTITION THE SELECTED DISK. ALL DATA ON THE SELECTED DISK WILL BE DESTROYED



- The vSphere 6 installation proceeds.



- The **Installation Complete** window displays when the installation process is completed. Press **ENTER** to reboot the system. (Make sure installation media has been ejected and bios set to boot from the boot disk.)

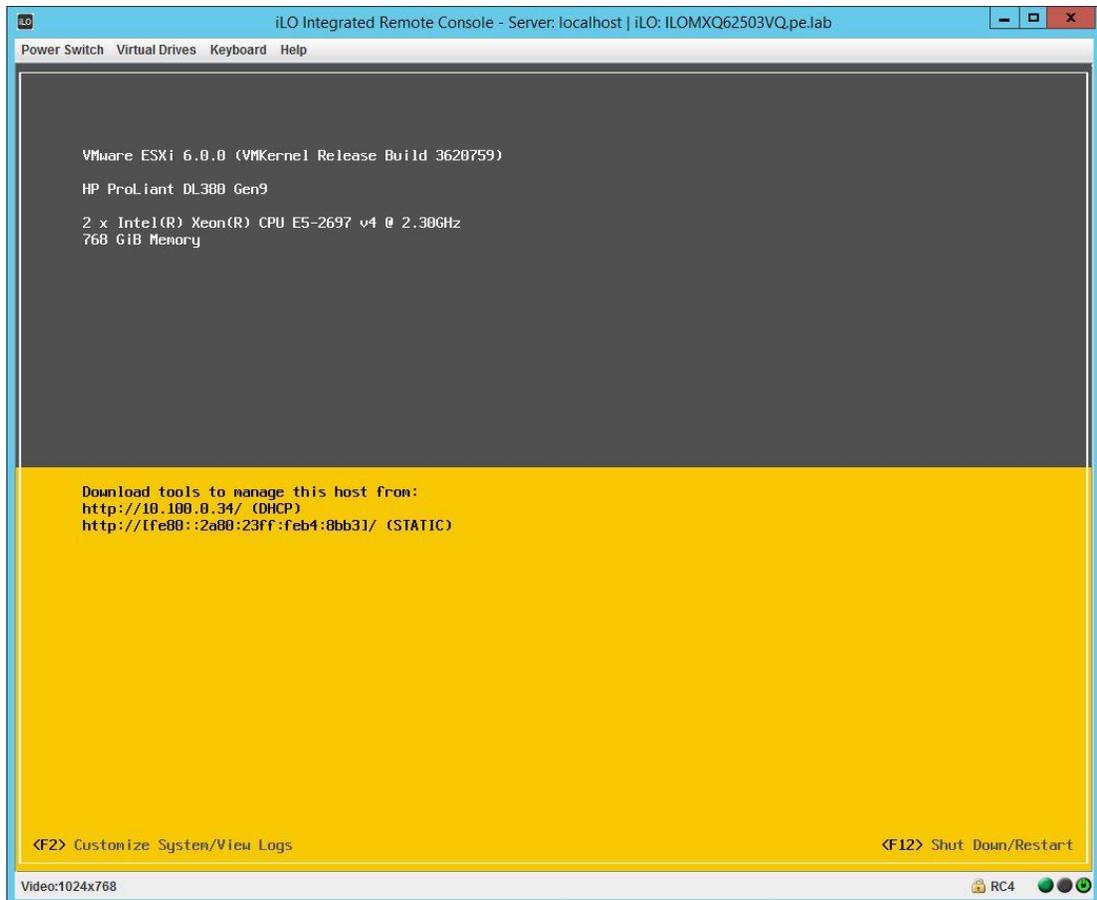


12. The installation is now complete.

3.2 INITIAL HOST CONFIGURATION

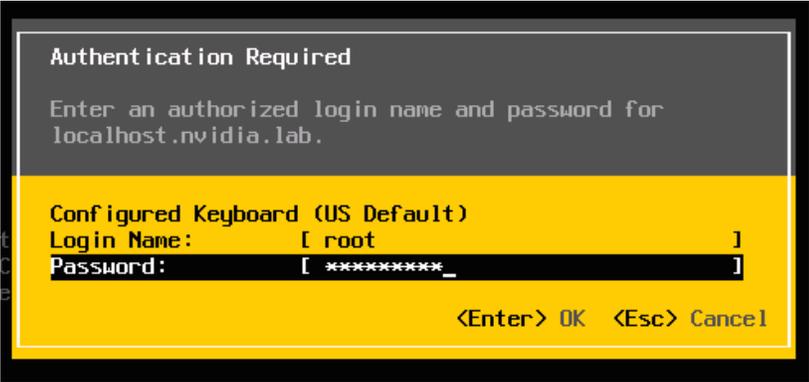
A countdown timer displays when you first boot ESXi 6. You can wait for the countdown to expire or press **ENTER** to proceed with booting. A series of notifications display during boot.

The VMware ESXi 6 screen displays once the boot completes:



Use the following procedure to configure the host:

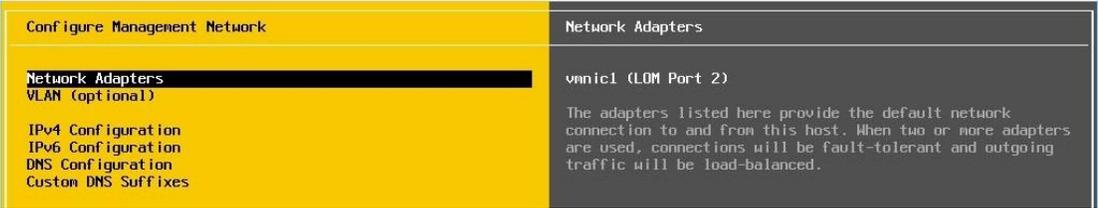
- 1. Press **F2**.



- 2. In the Authentication Required window, enter the root account credentials that you created during the installation process and then press **ENTER**.



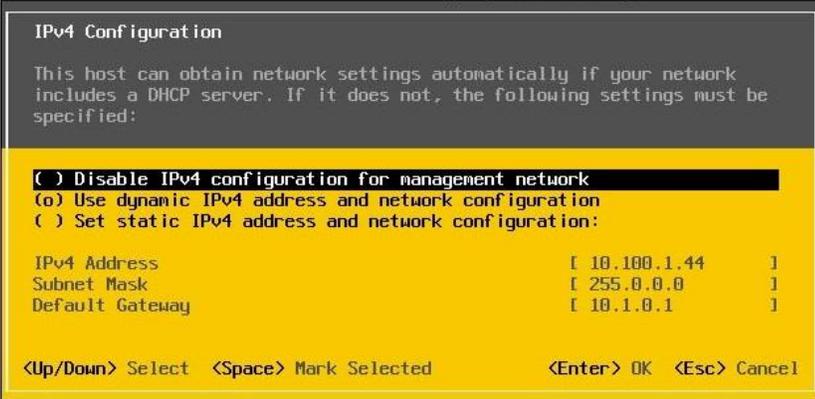
- 3. In the **System Customization** window, scroll down to select **Configure Management Network** and then press **ENTER**.



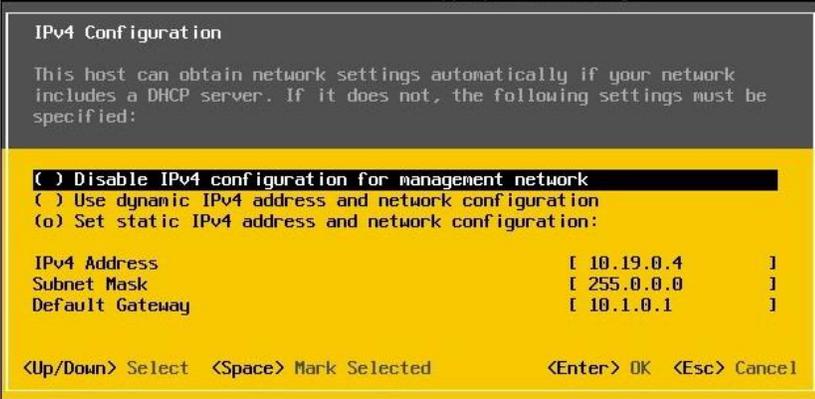
- 4. The **Network Adapters** window appears. Press **ENTER**.



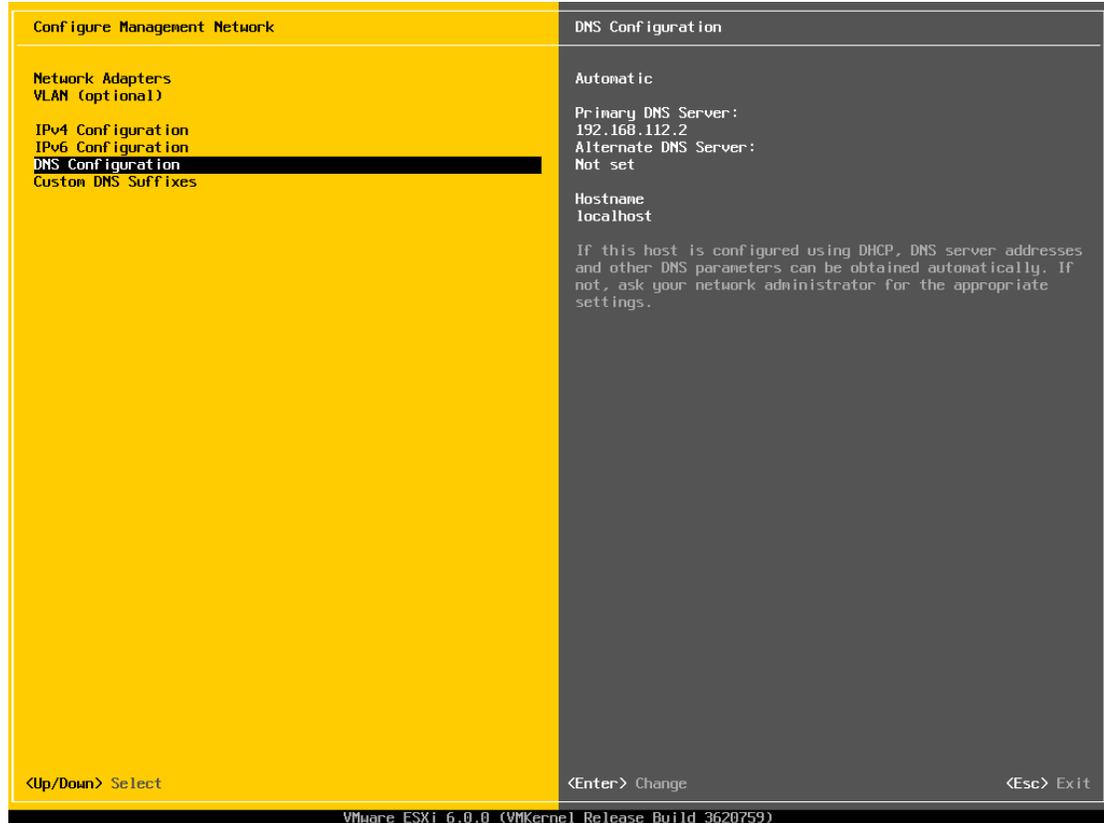
- 5. Use the arrow keys to select the adapter to use as the default management network, press **SPACE** to select, and then press **ENTER**.



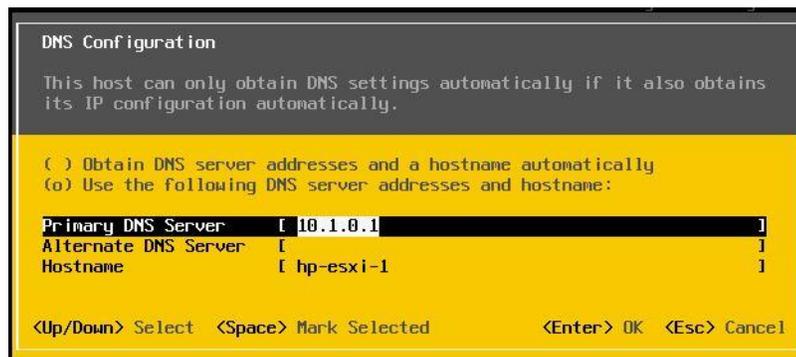
- 6. Scroll down to IPv4 Configuration and press **ENTER**. The IPv4 Configuration window displays.



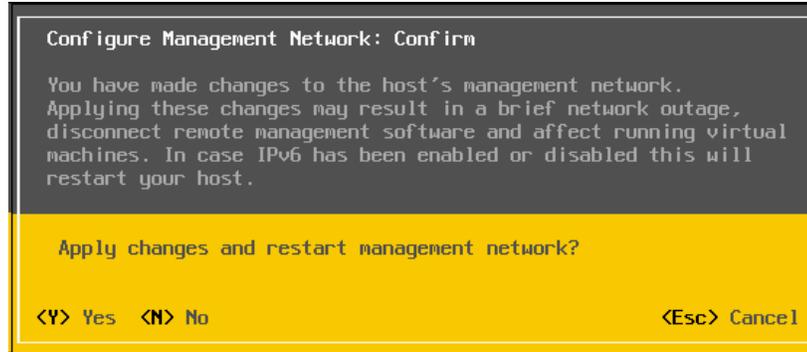
- Use the arrow keys to select **Set static IPv4 address and network configuration** and then enter the IPv4 address, subnet mask, and default gateway in the respective fields. Press **Enter** to continue.



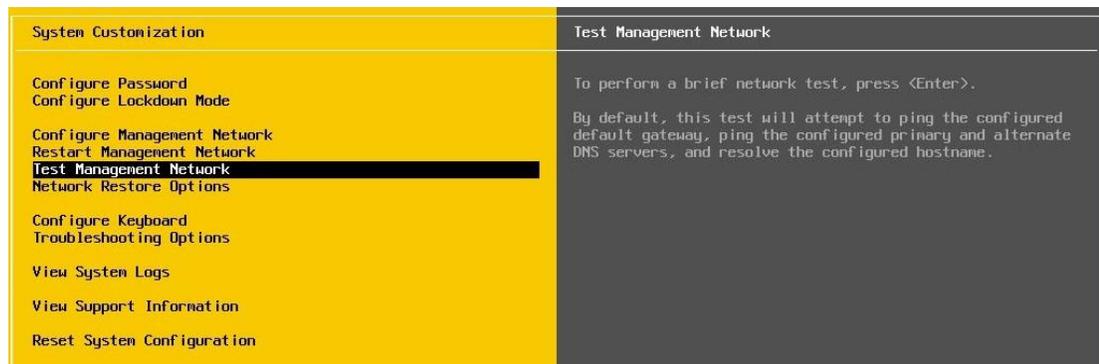
- Scroll down to **DNS Configuration** and press **ENTER**.



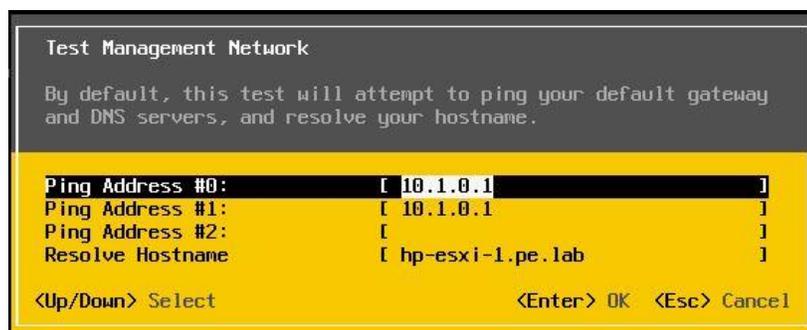
9. Add the primary and (if available) secondary DNS server address(es) in the respective fields. Set the host name for this vSphere host in the **Hostname** field. Press **ENTER** when finished.



10. Press **ESC** to save network changes. Press **Y** to confirm changes.

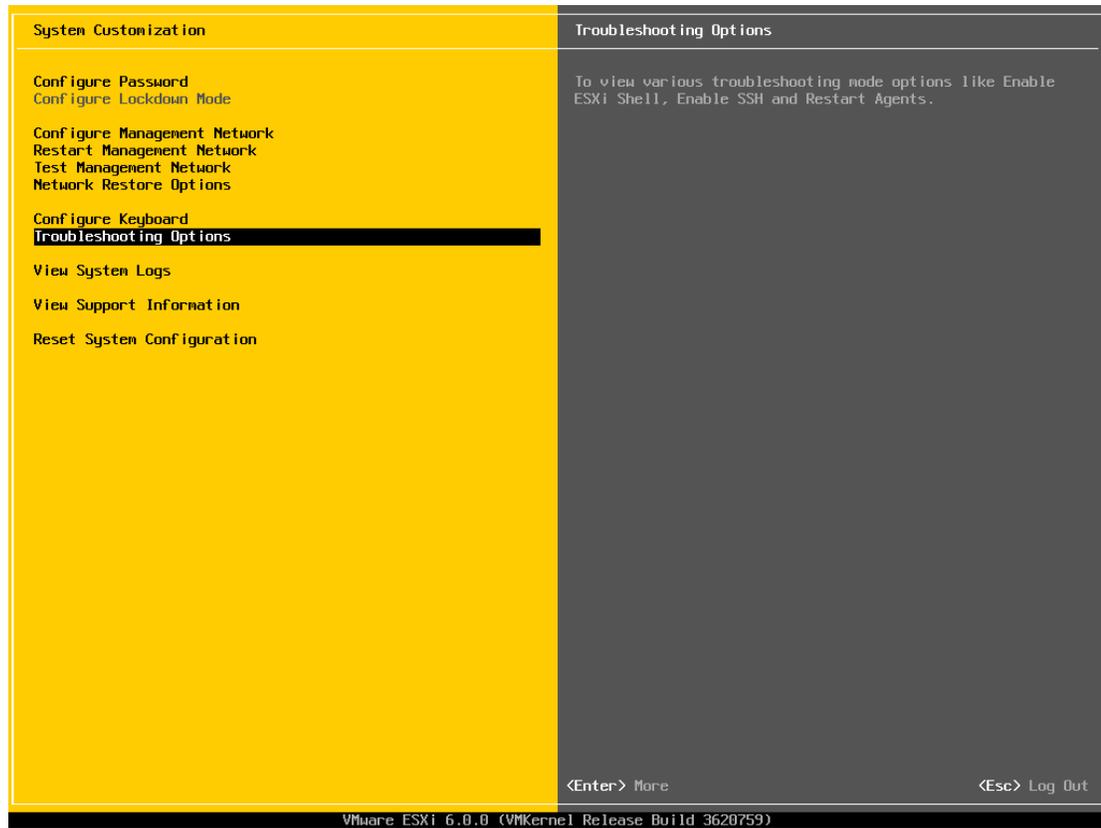


11. Select **Test Management Network** on the main vSphere 6 screen to open the **Test Management Network** window.



12. Perform the following tests:
 - Ping the default gateway.

- ▶ Ping the DNS server.
- ▶ Resolve a known address.



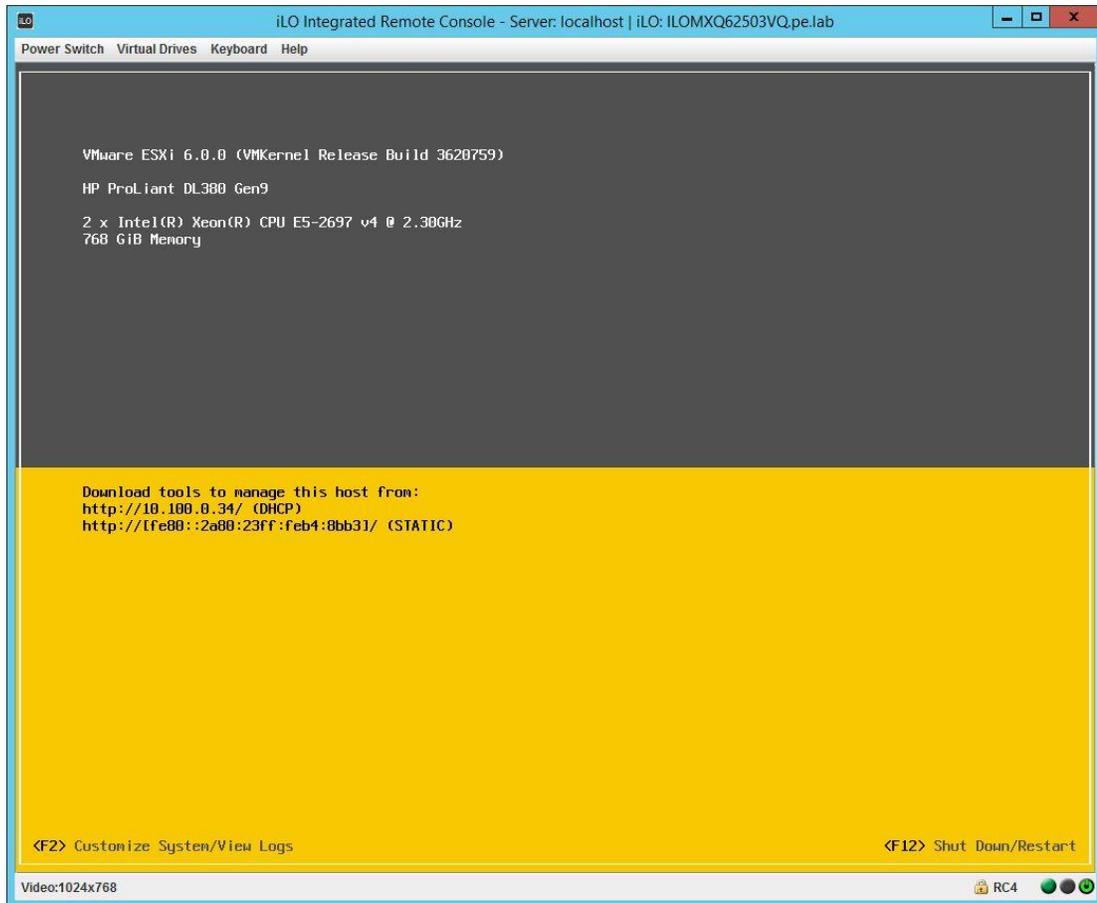
13. Return to the main vSphere 6 screen when you have completed testing, and then select **Troubleshooting Options**. The **Troubleshooting Mode Options** window displays.



14. To install the NVIDIA VIB in a later step, you will need to enable the ESXi Shell. This can be accomplished by selecting **Enable ESXi Shell** and pressing **ENTER** to toggle **Enable ESXi Shell** on.
The window on the right displays the status: **ESXi Shell is Enabled**



15. Enable SSH by selecting **Enable SSH** and pressing **ENTER** to toggle this option on. The window on the right displays the status **SSH is Enabled**.



16. Press **ESC** twice to log out of the ESXi UI.

VMware ESXi is now installed on the server.

CHAPTER 4. INSTALLING VMWARE VCENTER SERVER 6.0 U2

This chapter covers installing VMware vCenter Server 6.0 U2, including:

- ▶ Installing a vCenter Server Appliance (VCSA)
- ▶ Installing vCenter in a Windows Server 2012 VM
- ▶ The initial vCenter Server configuration
- ▶ Setting a vCenter appliance to auto-start
- ▶ Mounting an NFS ISO data store
- ▶ Creating vSphere clusters for GRID servers and management servers

Review the prerequisites in section 2.13.5, *Choose Virtualization Components* on page 31 before proceeding with these installations.



This deployment guide assumes you are building an environment for a proof of concept. Refer to VMware best practice guides before building your production environment.

4.1 INSTALLING VCENTER 6.0 U2 SERVER APPLIANCE

Use the following procedure to install a vCenter 6 Server Appliance.

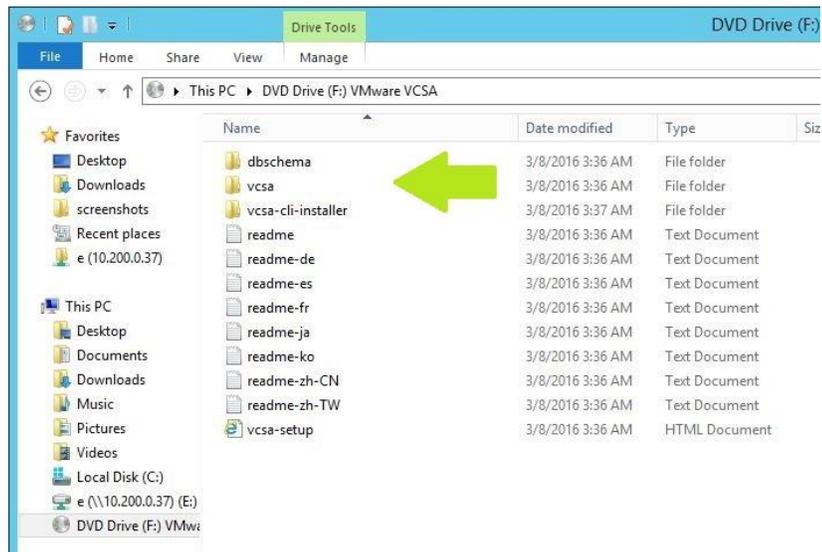
4.1.1 vCenter Server Client Integration Plug-in Installation

The vCenter Server Client Integration Plug-in needs to be installed to enable the vCenter Server Appliance install (as well as uploading files from your desktop to data stores). Use the following procedure to install the plug-in:

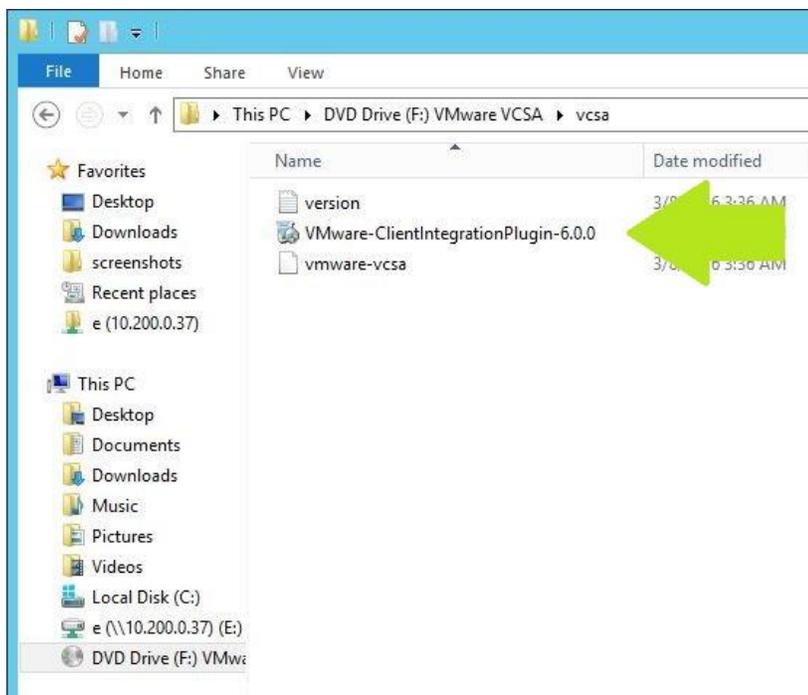
1. From any desktop client that can connect to the ESXi server previously created, mount the **VMware-VCSA-all-6.0.0-3634788.iso** in the CDROM drive.



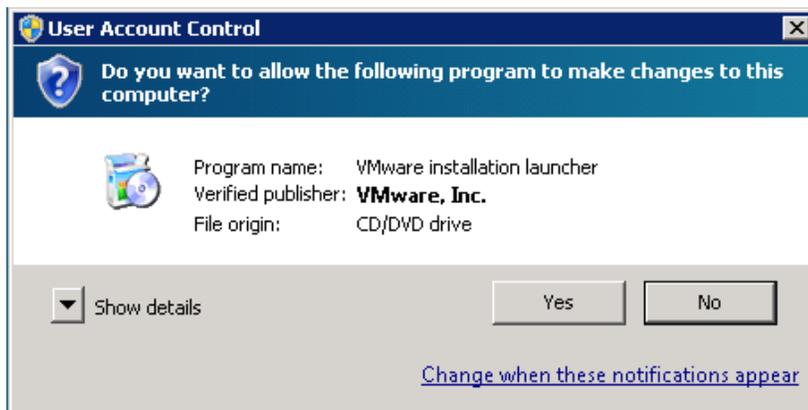
Windows 10 can connect this ISO file simply by double clicking on the file in Explorer. Other operating systems may need special software for mounting ISO files to CD drives.



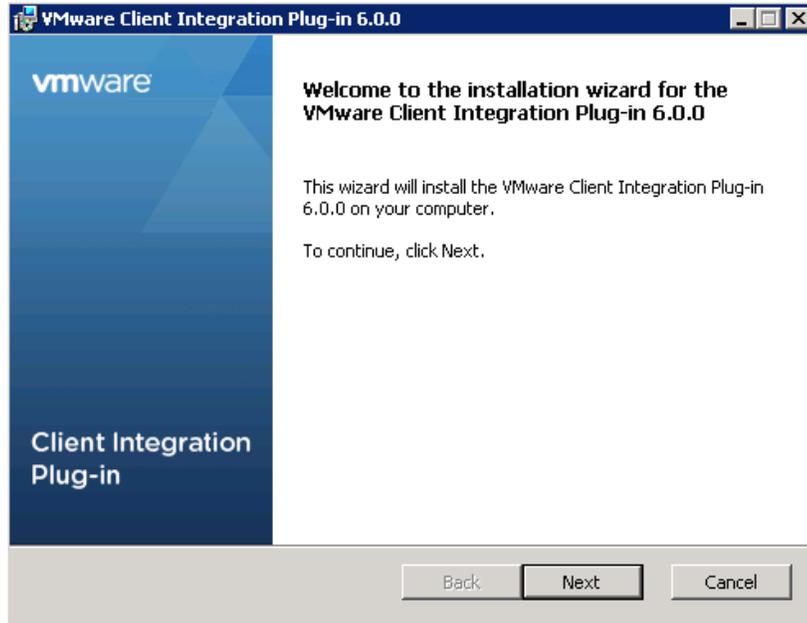
2. Browse to the **VCSA** folder:



3. In the VCSA folder, double click the **VMware-ClientIntegrationPlugin-6.0.0** installer.



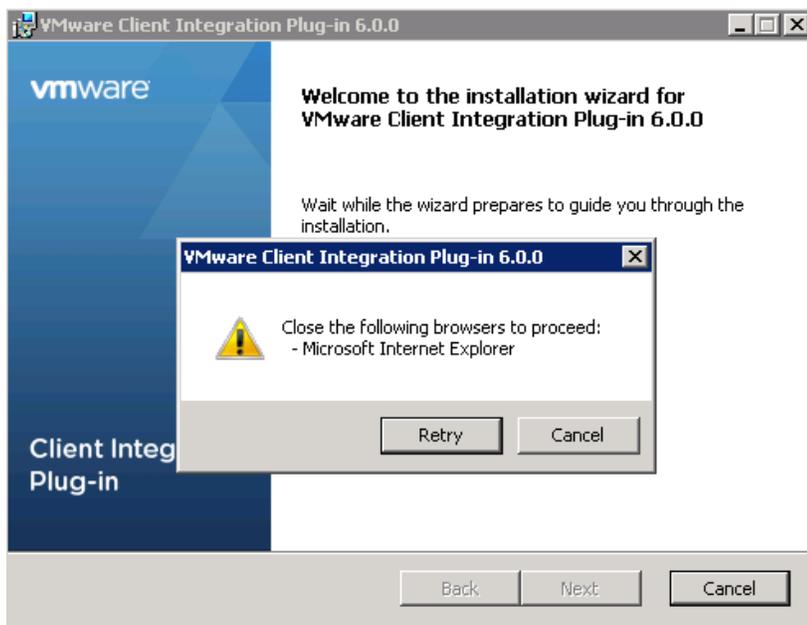
4. If the **Microsoft User Account Control** dialog prompts, click **Yes** to allow the install to begin.

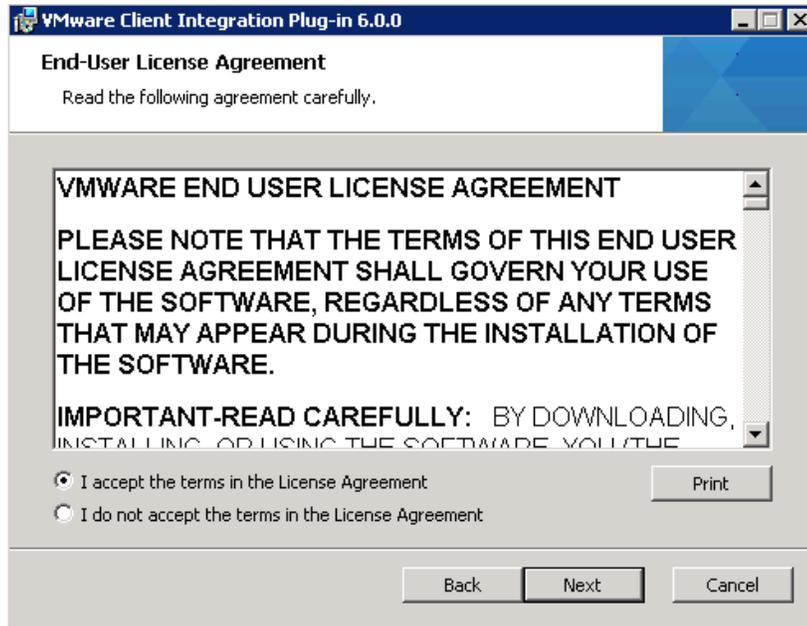


5. The installer starts, click **Next** to continue:

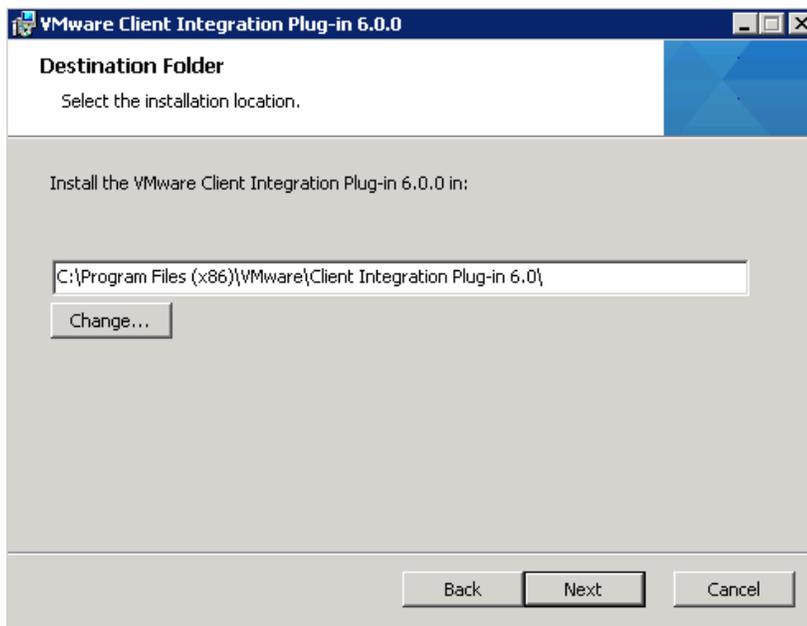
 **NOTE:** All browsers must be closed or you will be prompted to close them.

If this window pops up, close all listed browsers, and click **Retry** to continue:

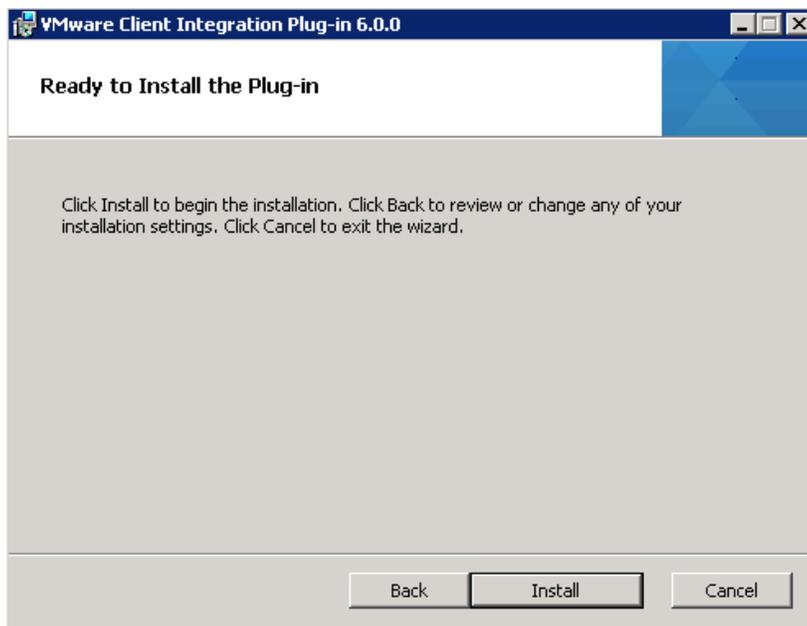




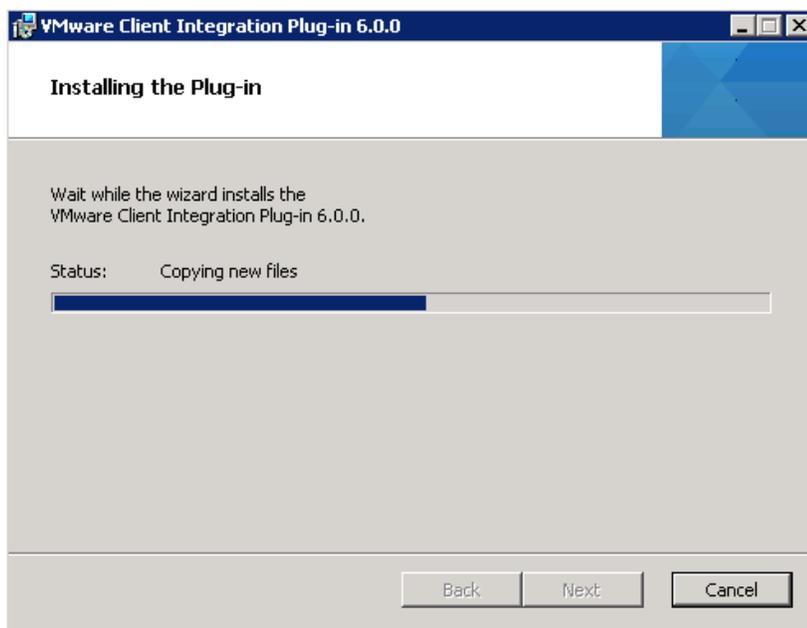
6. Accept the EULA and click **Next** to continue.



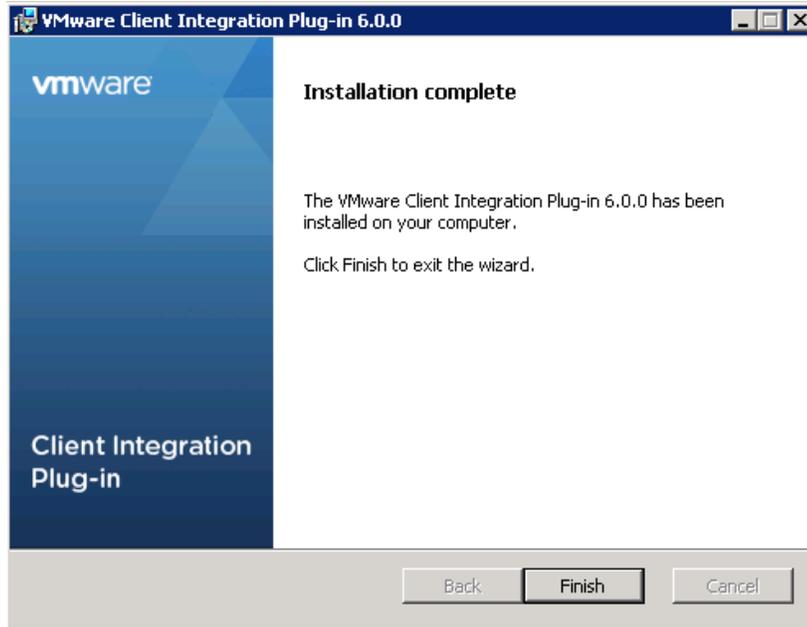
7. Select the install destination folder and click **Next**.



8. Click **Install** to begin the installation.



9. Installation will progress...



10. When install completes, click **Finish**:

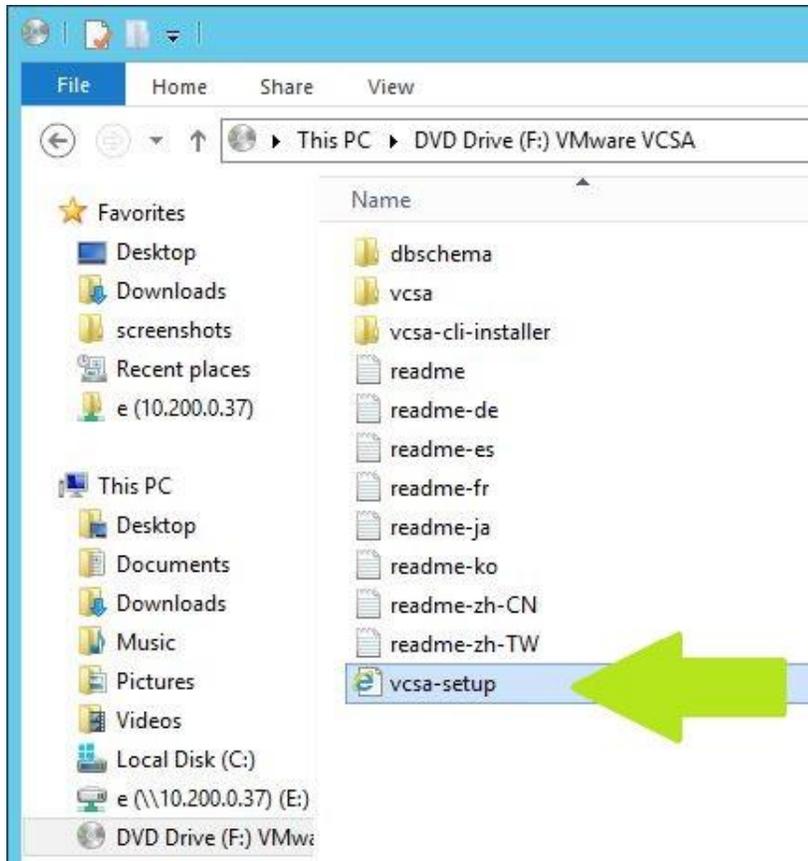
4.1.2 vCenter Server Appliance (VCSA) Installation

Follow this procedure to install the VCSA.

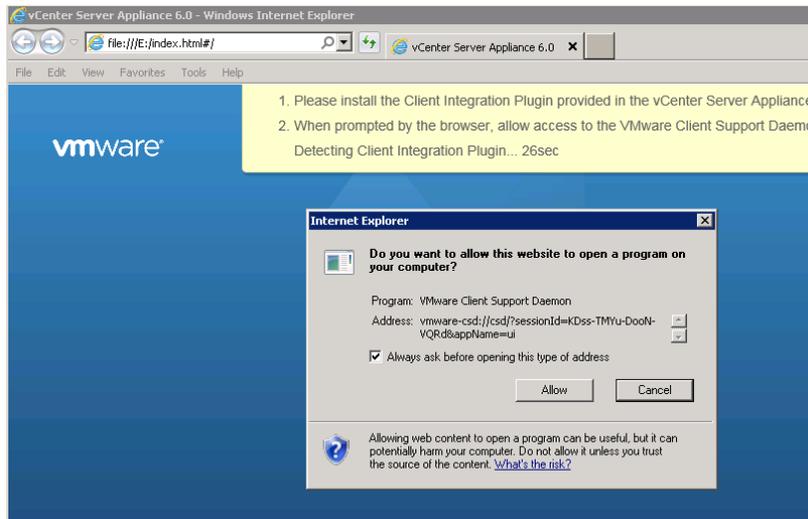
1. From any desktop client that can connect to the ESXi server previously created, mount the **VMware-VCSA-all-6.0.0-3634788.iso** in the CDROM drive.



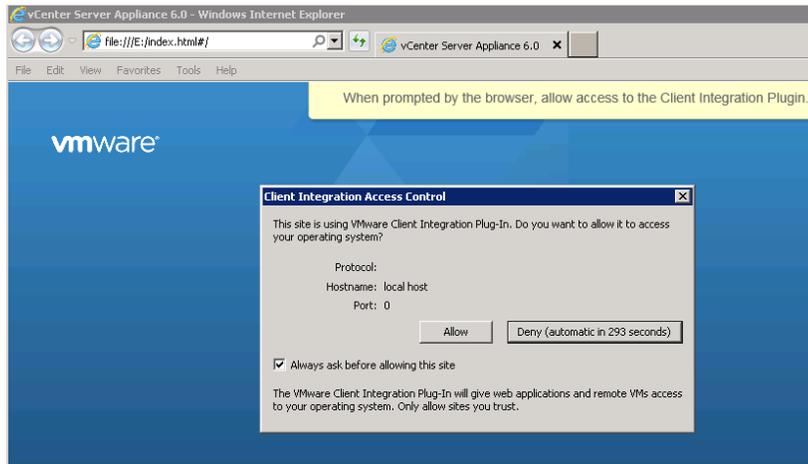
Windows 10 can connect this ISO file simply by double clicking on the file in Explorer. Other operating systems may need special software for mounting ISO files to CD drives.



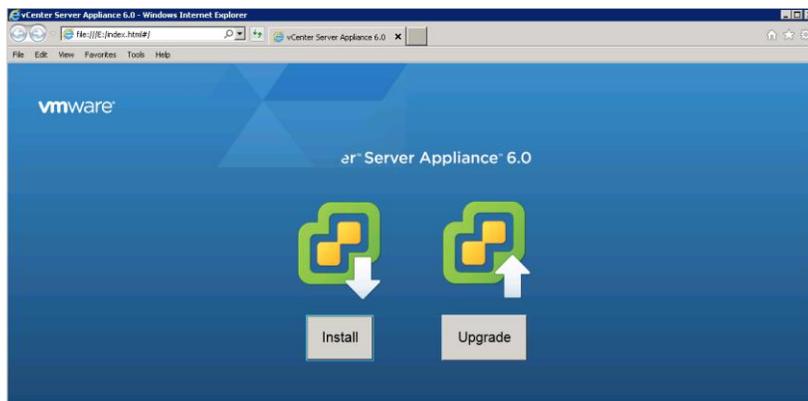
2. Open the html file **vcsa-setup** in Internet Explorer.



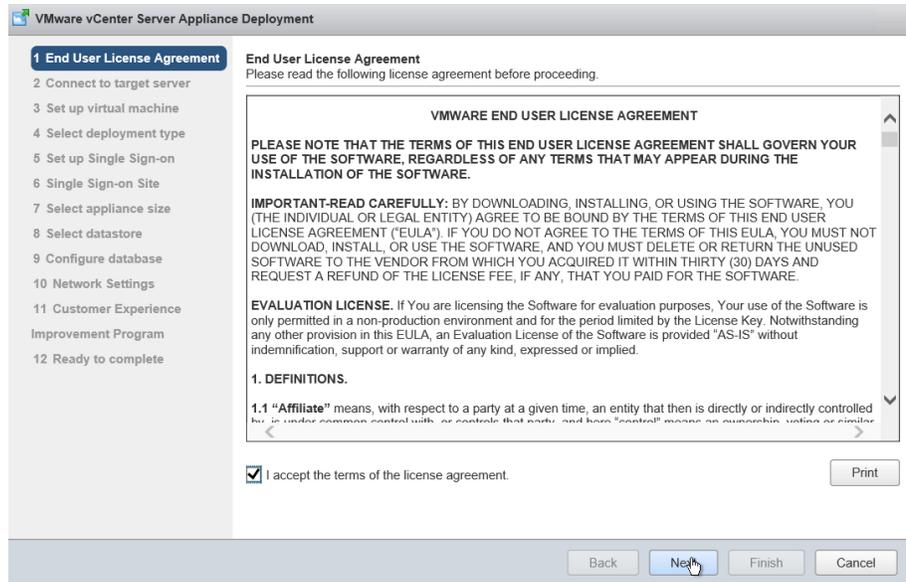
3. This alert is displayed if the VMware Client Integration Plug-in is not installed. Click **Allow**.



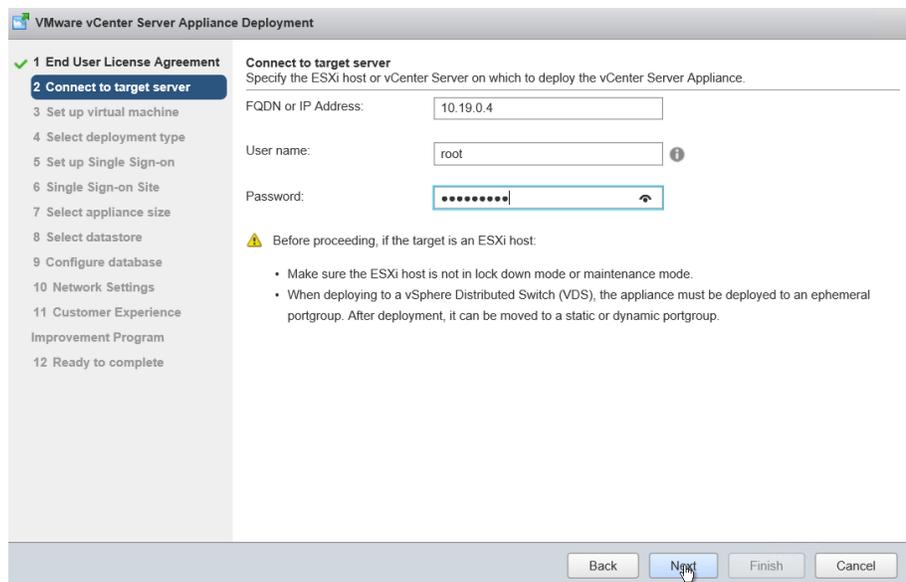
4. If the plug-in is active, you are prompted to allow it access to your operating system, click **Allow**.



5. To begin the install of the VCSA on a host click **Install**.



6. Read and then accept the EULA, click **Next** to continue.



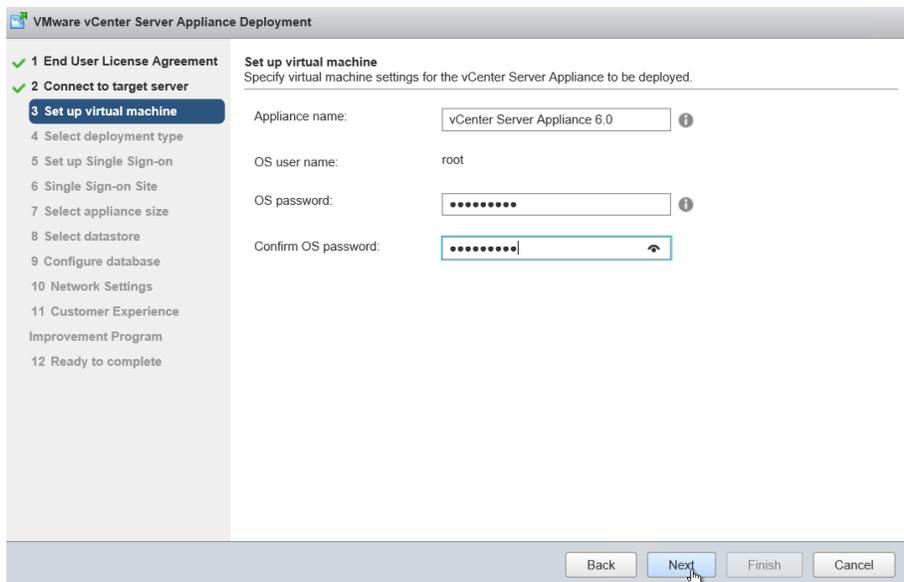
7. In this step select the vSphere host on which to install the VCSA as a guest. This can be a host running ESXi 5.5 or later. It is recommended that the vCenter server (Windows or appliance based) run on a separate management cluster from the one designated for VDI workloads. Enter the IP address or FQDN of the chosen host, then its root user name and password and click **Next**.



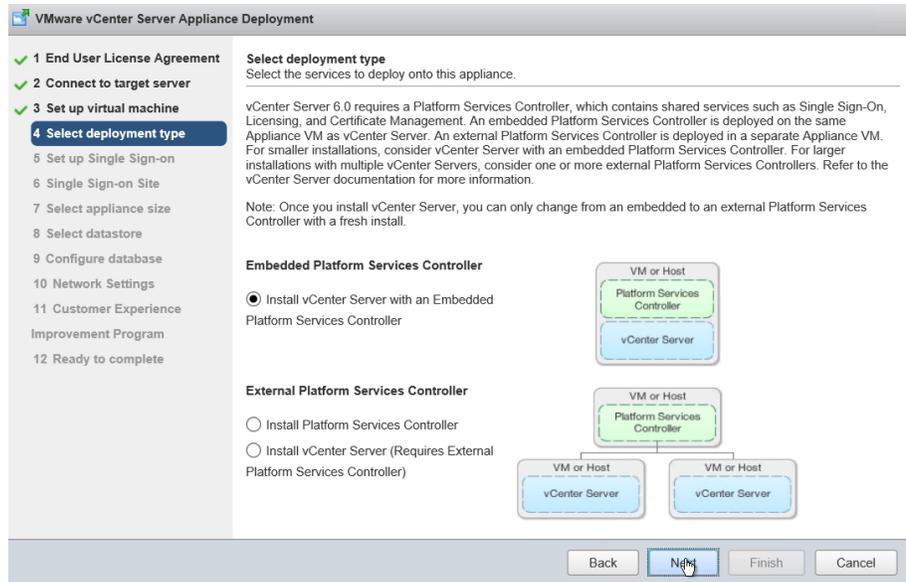
- If the web installer can reach the host then a certificate warning will occur. Click **Yes** to continue.



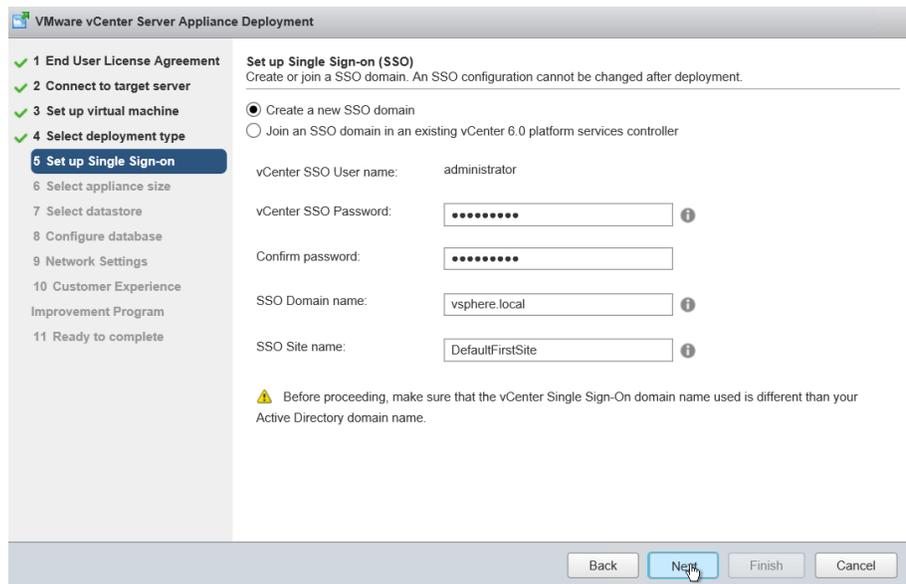
- The credentials provided are validated.



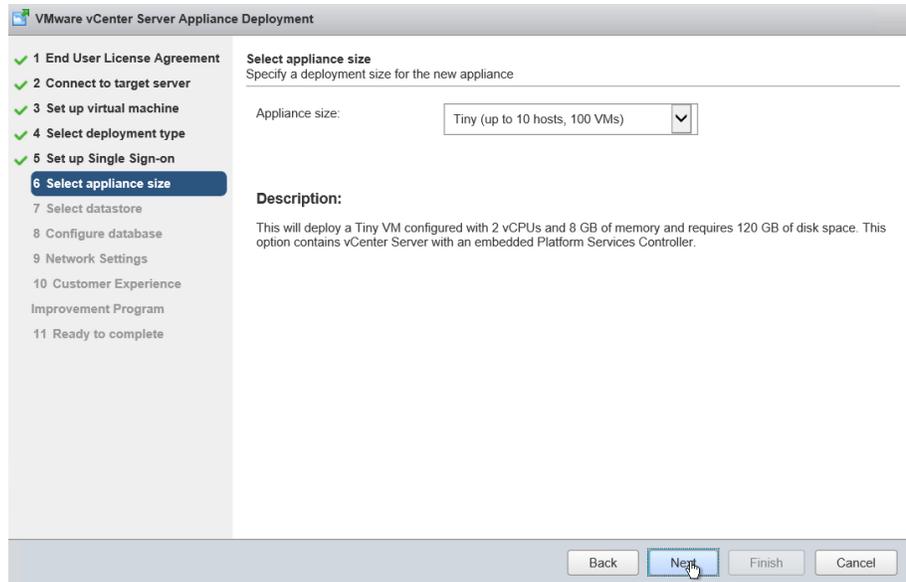
- Enter a name for the appliance, then enter a root password for the appliance, enter it again, and click **Next**.



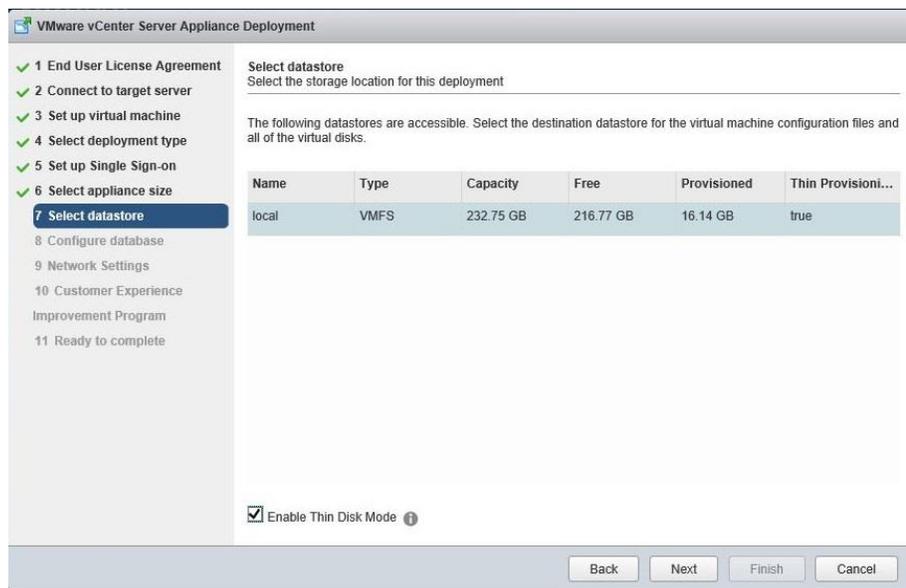
11. Select **Install vCenter Server with an Embedded Platform Services Controller** and click **Next**.



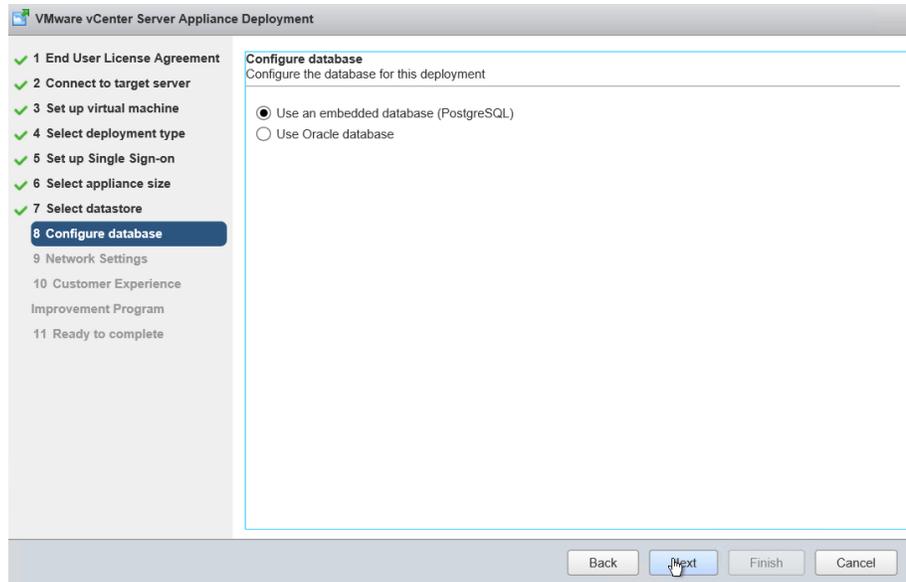
12. Select **Create a new SSO domain**. For the purposes of a proof of concept, enter (and confirm) a password, a domain name (we chose the default of **vsphere.local**), and a site name, then click **Next**.



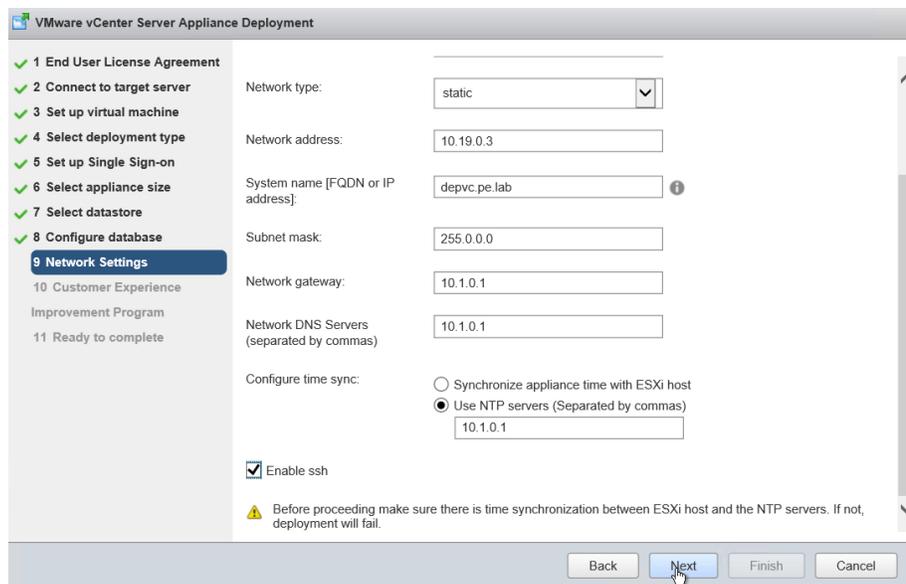
13. Select an appliance size, appropriate for the potential scale of your proof of concept or trial. Here we have selected **Tiny**. Click **Next**.



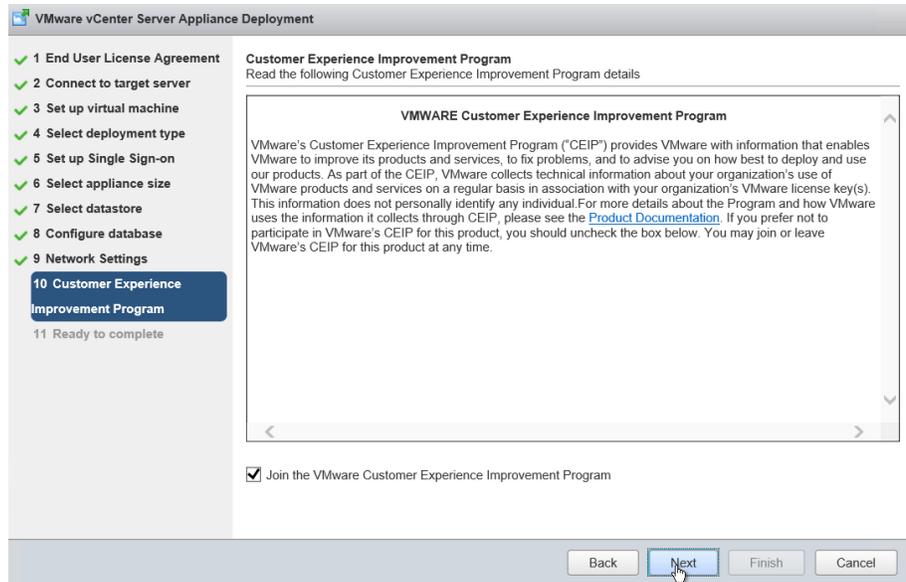
14. Select a datastore from those available on the host you have connected with, select **Enable Thin Disk Mode**, and then click **Next**.



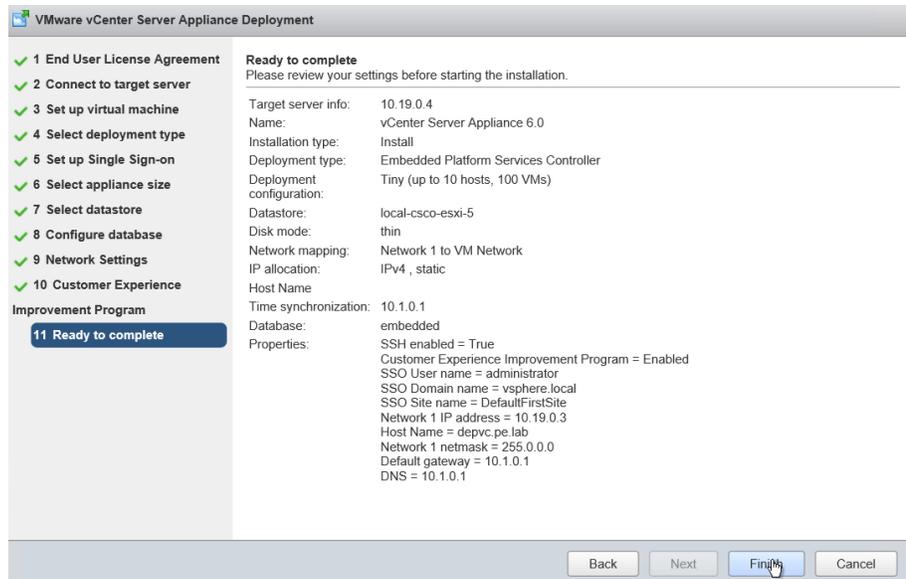
15. Choose **Use an embedded database (PostgreSQL)**, then click **Next**.



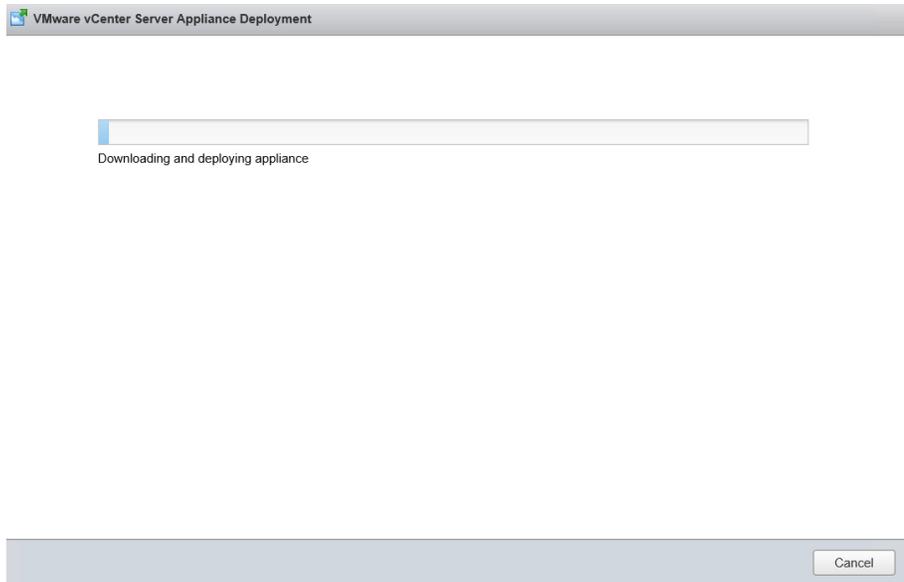
16. Before configuring network settings, choose a static IP address, and enter that into local DNS (e.g. on the Domain Controller). Once you can resolve the address, enter that IP address, host name, and then scroll down for remaining entries. This is a long page and will require scrolling down to see all settings. Make sure the configuration leverages an NTP service. The lower half of the **Network Settings** pane adds NTP, enter the address of the local time server or a reachable external one. Select **Enable SSH**. Click **Next**.



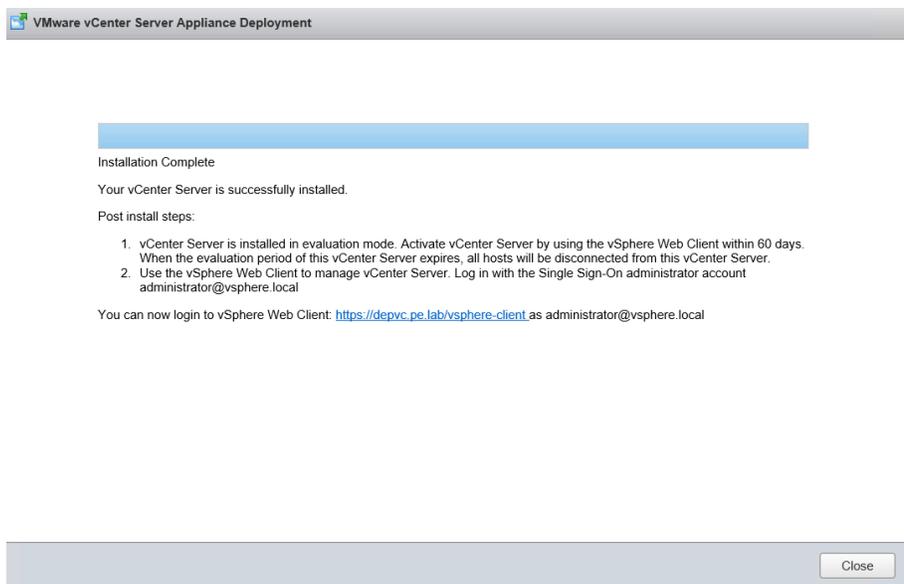
17. The VMware Customer Experience Program sends information about usage and problems directly to VMware for collection. Using this information VMware can more easily identify bugs and other discrepancies. No confidential information is ever sent. Join the Customer Experience Program or not, based on internal security requirements and desire, and click **Next**.



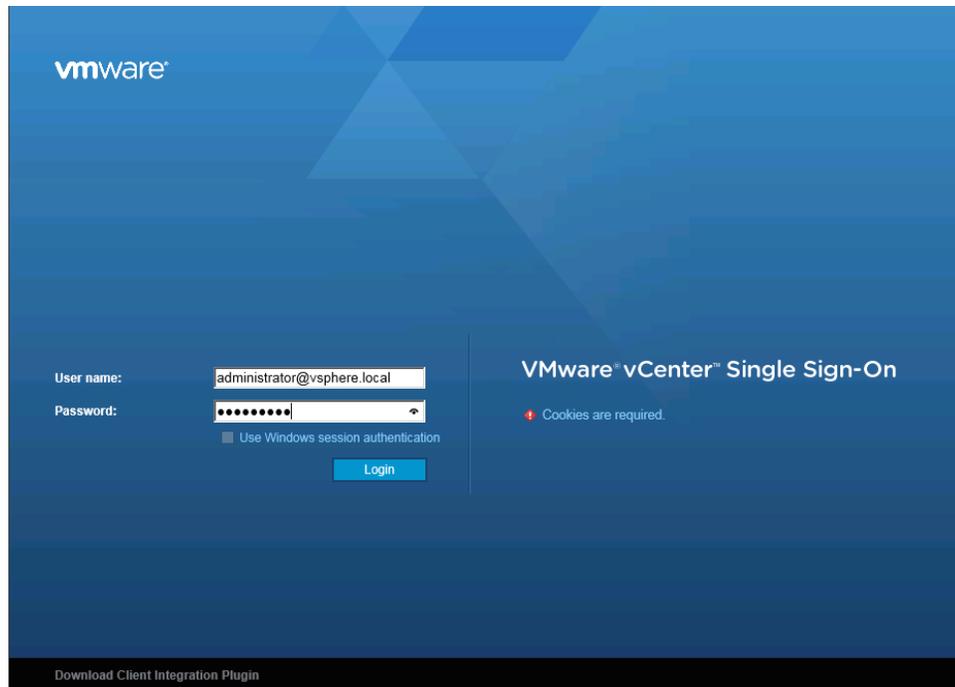
18. Confirm your settings and, if correct, click **Finish** to deploy your VCSA.



19. If all settings are correct then a progress bar will appear.



20. When complete, a window appears with a link to the VCSA. Click the link to reach the login page. You can now close this page.



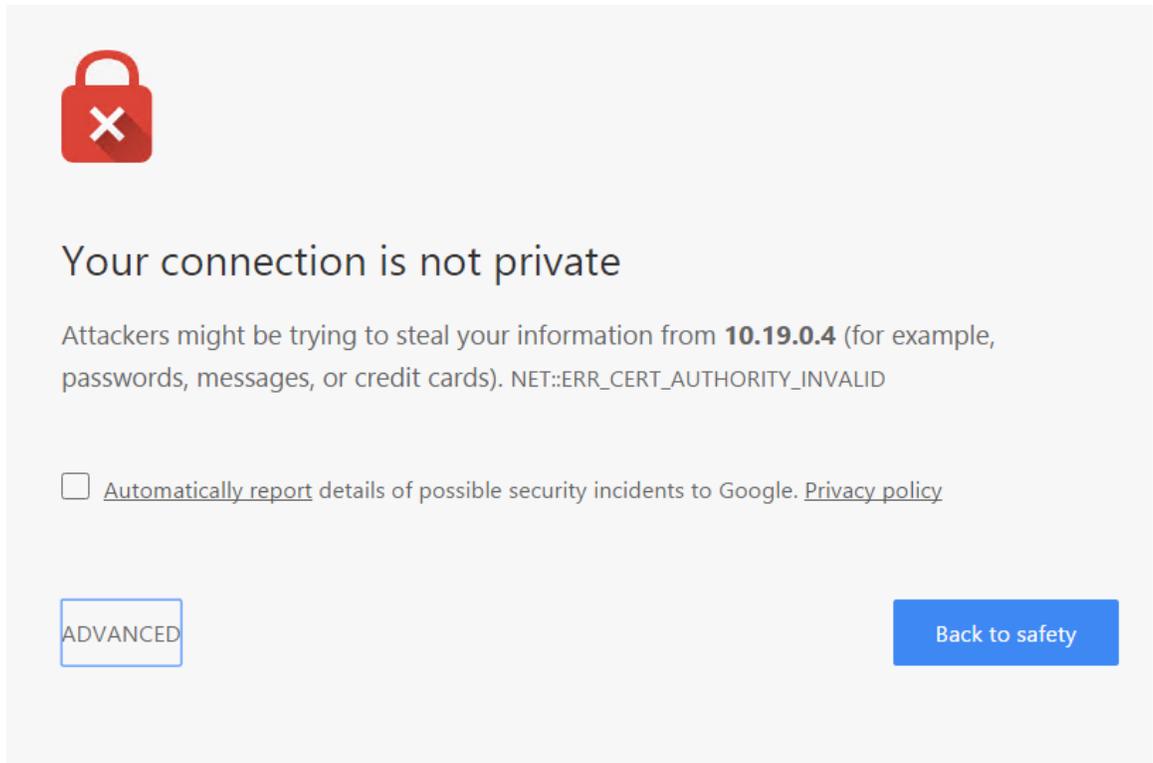
21. The VCSA is now installed and ready for use. From your browser, the login page should look like this (be sure to enable cookies if you have not).

4.2 INSTALLING VCENTER 6.0 U2 SERVER ON WINDOWS SERVER 2012

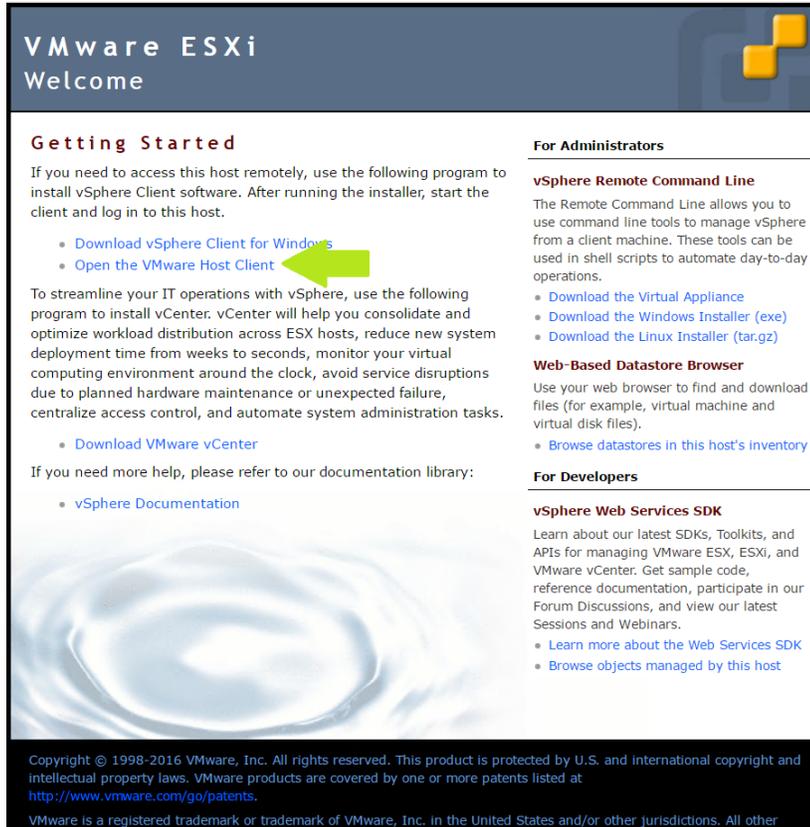
This section describes how to install and configure vCenter Server 6.0 U2 in a Windows Server VM.

4.2.1 Initial Installation

Use the following procedure to install vCenter Server.



1. Open a browser and navigate to the URL for your ESXi server. A certificate error will occur. This is expected, click **ADVANCED**, then **Proceed to <your ESXi server address>**.



VMware ESXi
Welcome

Getting Started

If you need to access this host remotely, use the following program to install vSphere Client software. After running the installer, start the client and log in to this host.

- [Download vSphere Client for Windows](#)
- [Open the VMware Host Client](#)

To streamline your IT operations with vSphere, use the following program to install vCenter. vCenter will help you consolidate and optimize workload distribution across ESX hosts, reduce new system deployment time from weeks to seconds, monitor your virtual computing environment around the clock, avoid service disruptions due to planned hardware maintenance or unexpected failure, centralize access control, and automate system administration tasks.

- [Download VMware vCenter](#)

If you need more help, please refer to our documentation library:

- [vSphere Documentation](#)

For Administrators

vSphere Remote Command Line

The Remote Command Line allows you to use command line tools to manage vSphere from a client machine. These tools can be used in shell scripts to automate day-to-day operations.

- [Download the Virtual Appliance](#)
- [Download the Windows Installer \(exe\)](#)
- [Download the Linux Installer \(tar.gz\)](#)

Web-Based Datastore Browser

Use your web browser to find and download files (for example, virtual machine and virtual disk files).

- [Browse datastores in this host's inventory](#)

For Developers

vSphere Web Services SDK

Learn about our latest SDKs, Toolkits, and APIs for managing VMware ESX, ESXi, and VMware vCenter. Get sample code, reference documentation, participate in our Forum Discussions, and view our latest Sessions and Webinars.

- [Learn more about the Web Services SDK](#)
- [Browse objects managed by this host](#)

Copyright © 1998-2016 VMware, Inc. All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. VMware products are covered by one or more patents listed at <http://www.vmware.com/go/patents>.

VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other

2. On the VMware ESXi Welcome screen, click **Open the VMware Host Client**.



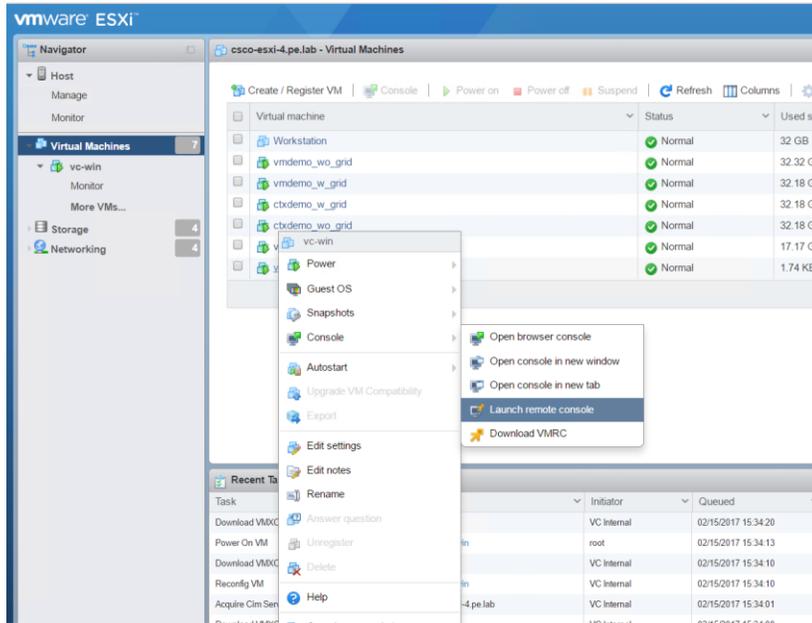
vmware®

User name:

Password:

vmware® ESXi™

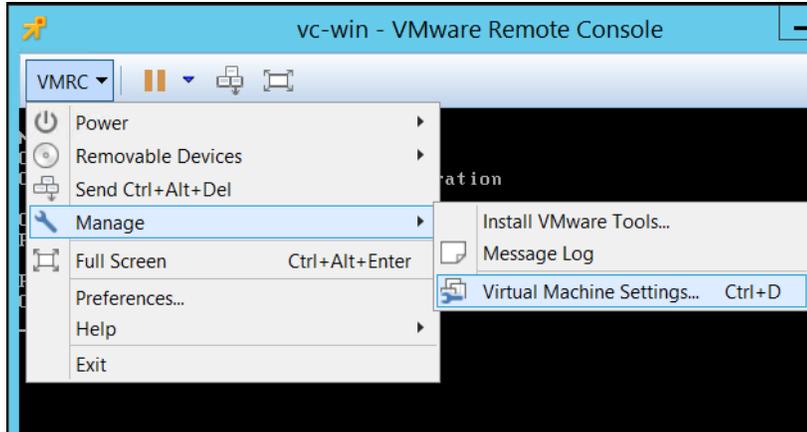
3. Log in to the ESXi web client using User name **root** and the root password.



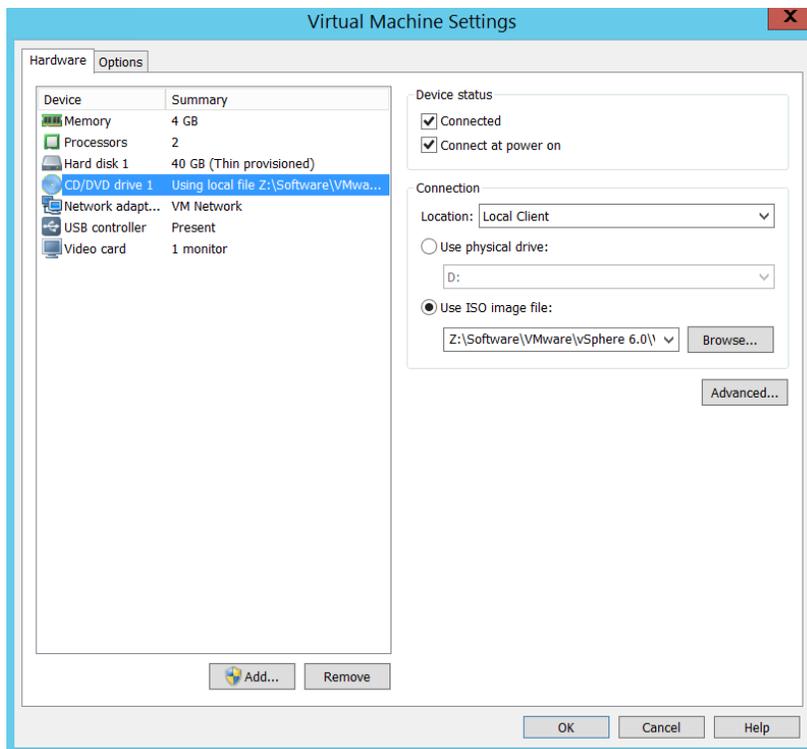
- In the left-hand navigation pane, click on Virtual Machines. In the right-hand pane, right click on the VM you will be installing vCenter in and select **Console** → **Launch Remote Console**.



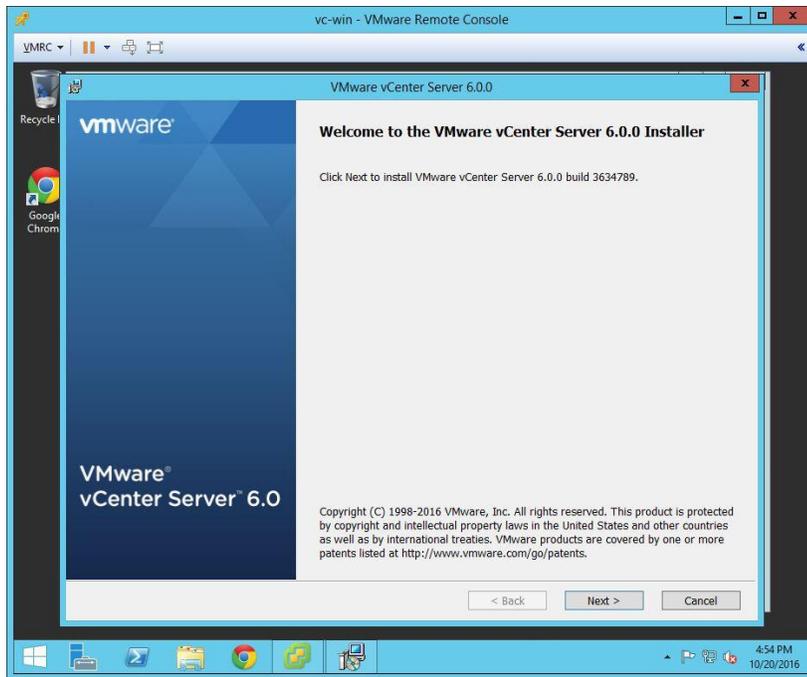
- On the **Invalid Security Certificate** dialog, click **Connect Anyway**.



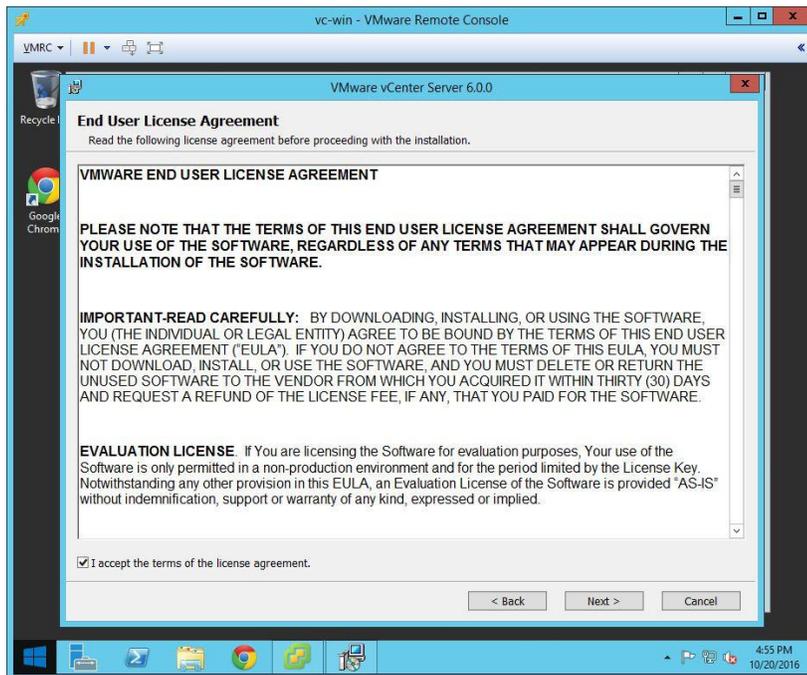
6. In the VMware Remote Console window, click **VMRC** → **Manage** → **Virtual Machine Settings...**



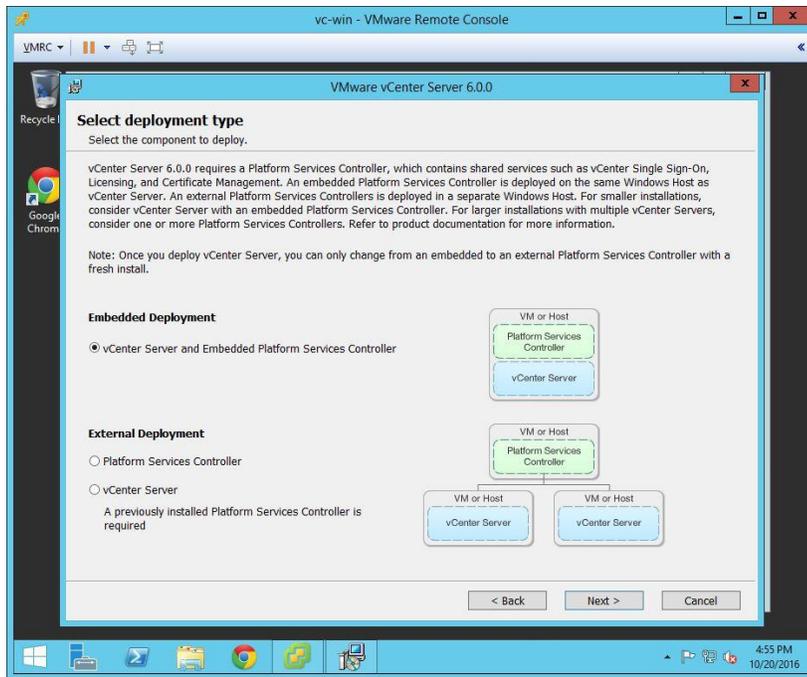
7. Mount the **VMware-VIMSetup-all-6.0.0-3634788** ISO to the CD drive by clicking **CD/DVD drive 1**, clicking **Connected**, **Connected at power on**, then **Use ISO image file**. Use the **Browse** button to find the correct ISO. Click **OK**. The vCenter Server software is now mounted inside the VM as the CD drive.
8. Inside the VM console, right-click the CD drive and select **Install or run program** from your media.



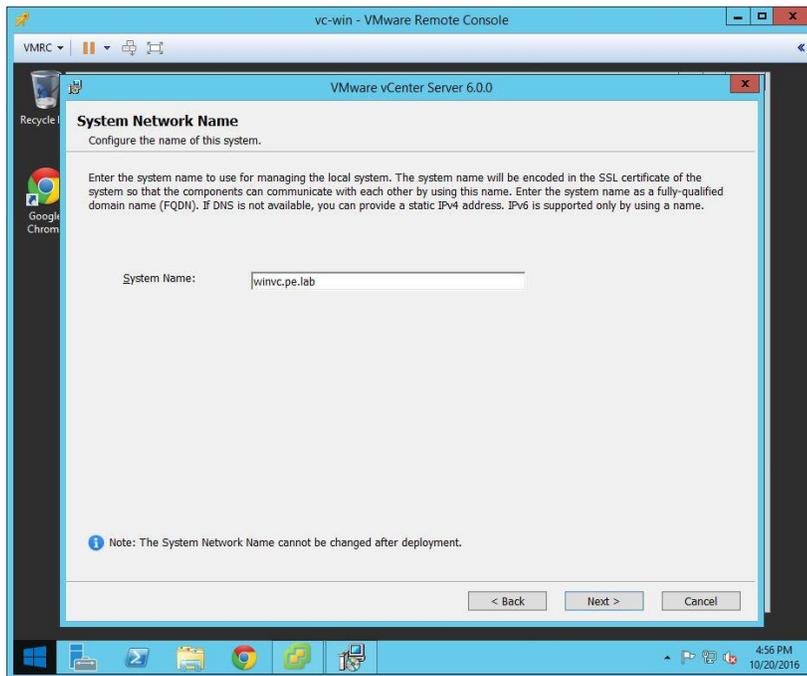
9. Click **Next** to proceed with the installation.



10. Check **I accept the terms of the license agreement** checkbox, click **Next** to proceed.

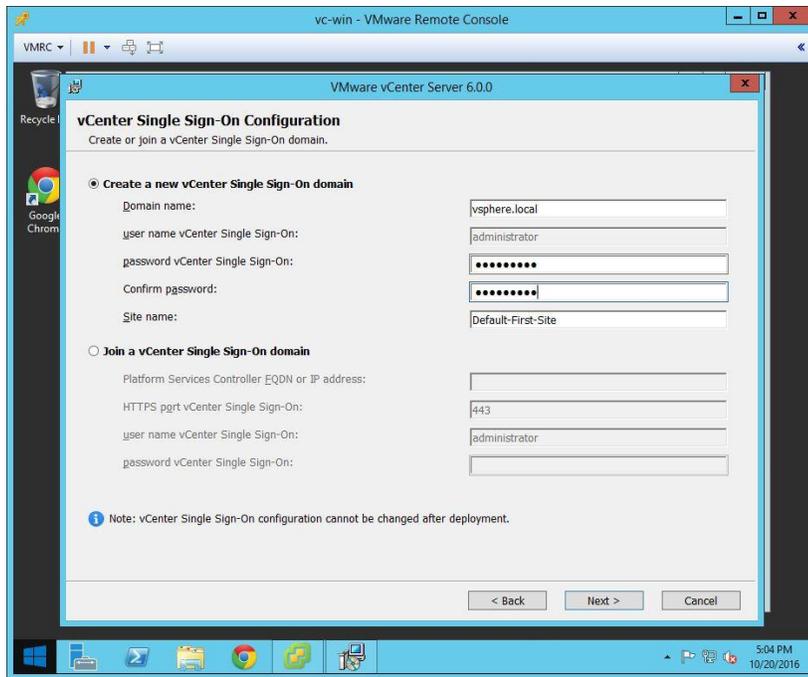


11. Select **vCenter Server with an embedded infrastructure controller**. Click **Next** to proceed.



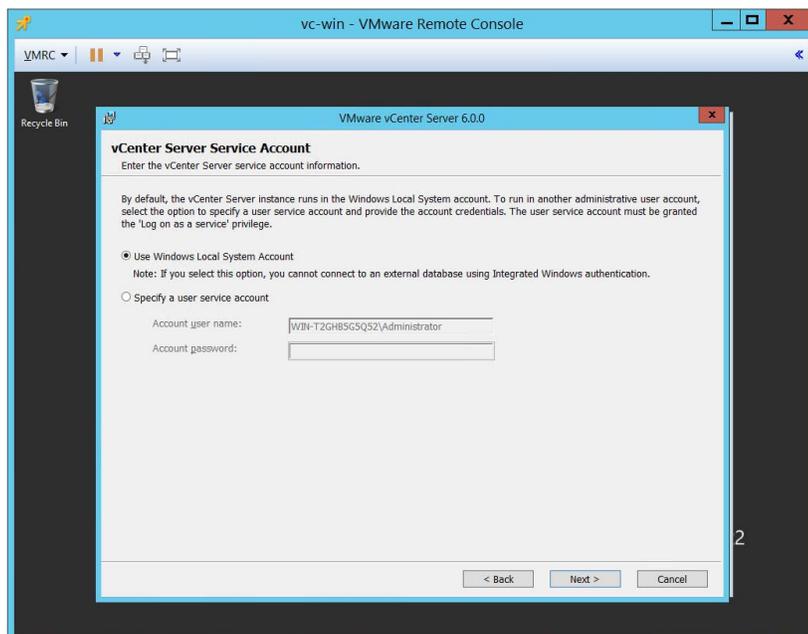
12. Enter the FQDN or IP address of the host in the **System Name** field, click **Next** to proceed.

 A warning dialog displays if you enter an IP address; click **OK** to proceed.

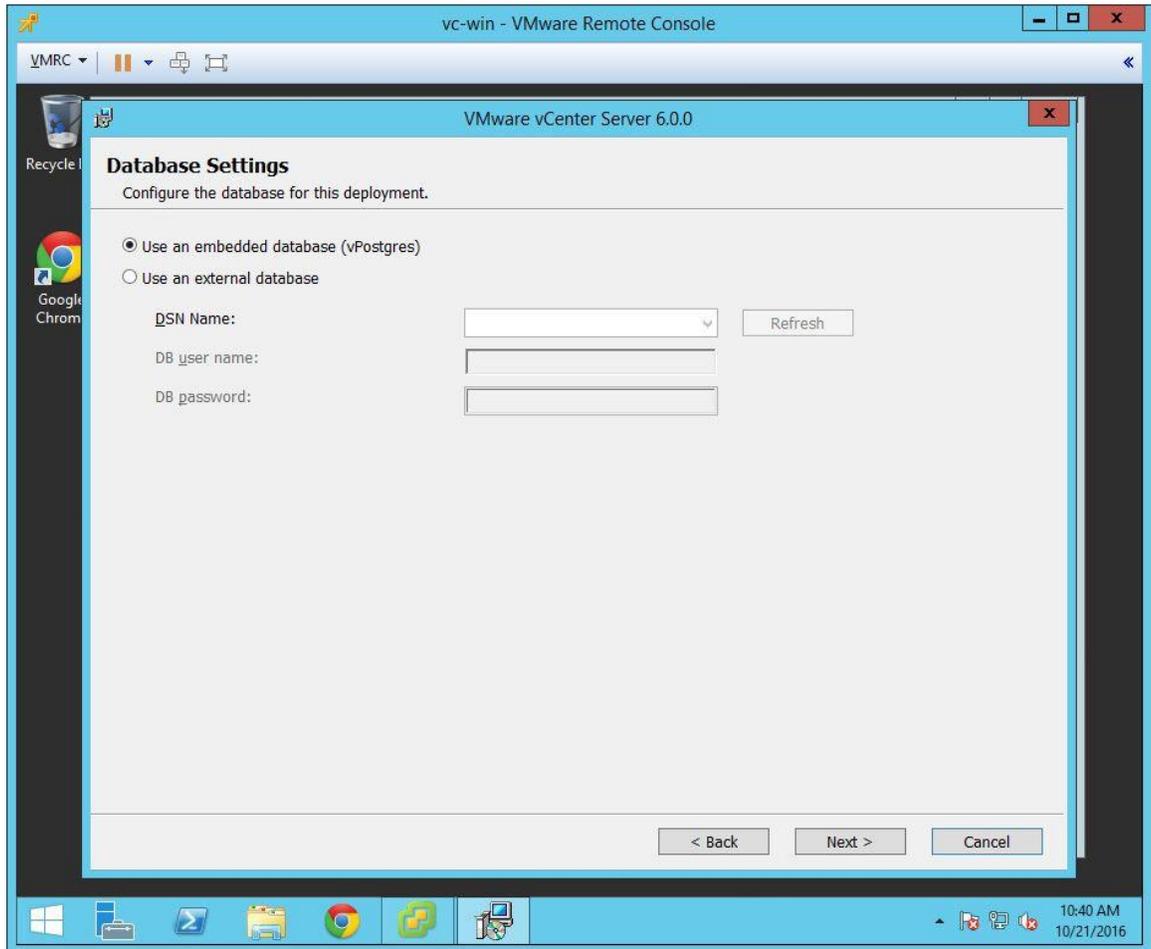


- The Single Sign-On Configuration window displays. Enter the administrator password for the new Single Sign-On domain in the **Password** and **Confirm password** fields. Click **Next** to proceed.

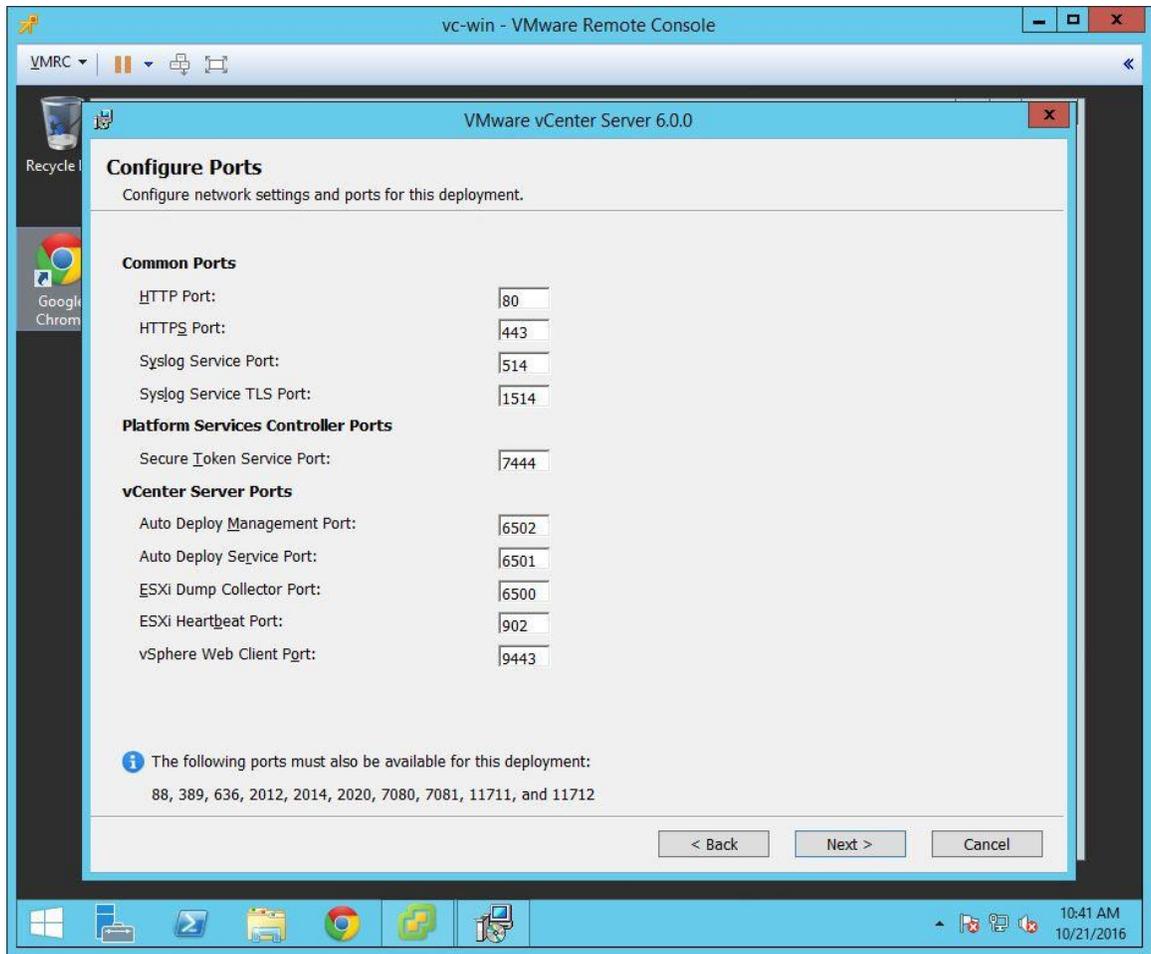
CAUTION: TO PREVENT UNAUTHORIZED ACCESS, YOUR SELECTED ROOT PASSWORD SHOULD CONTAIN AT LEAST EIGHT (8) CHARACTERS AND CONSIST OF A MIX OF LOWERCASE AND CAPITAL LETTERS, DIGITS, AND SPECIAL CHARACTERS



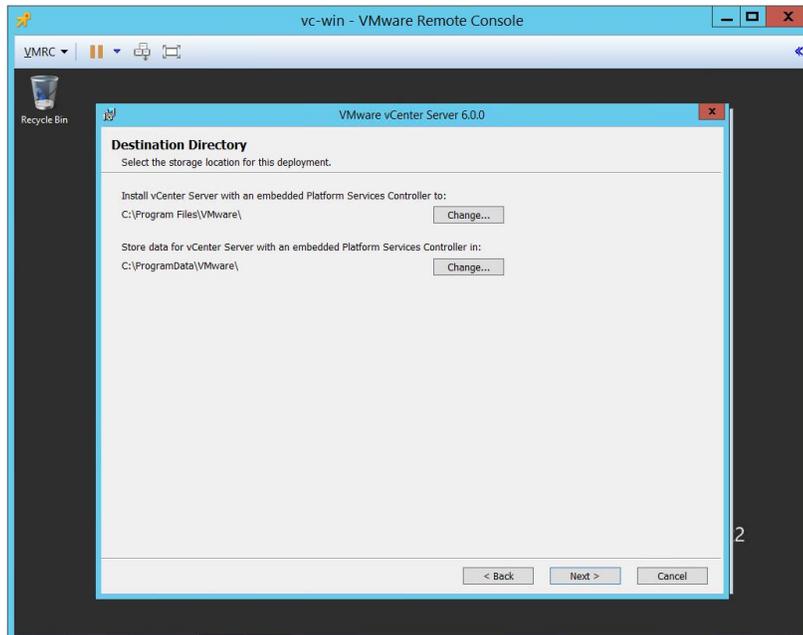
14. Select **Use Windows Local Account**. Click **Next** to proceed.



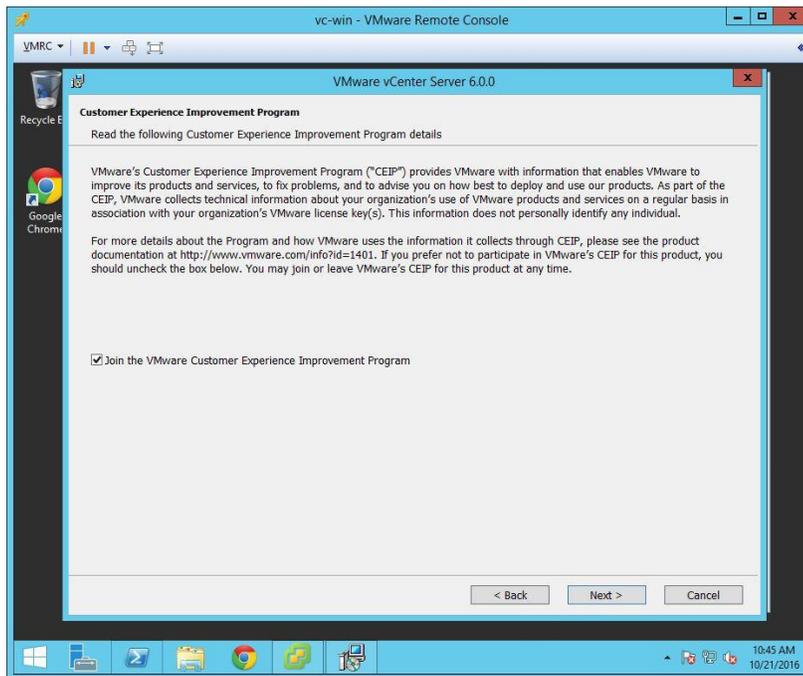
15. Select **Use an embedded database (vPostgres)**, Click **Next** to proceed.



16. Ensure that the ports are configured as shown in the **Configure Ports** window and click **Next** to proceed. The *Destination Folder* window displays.

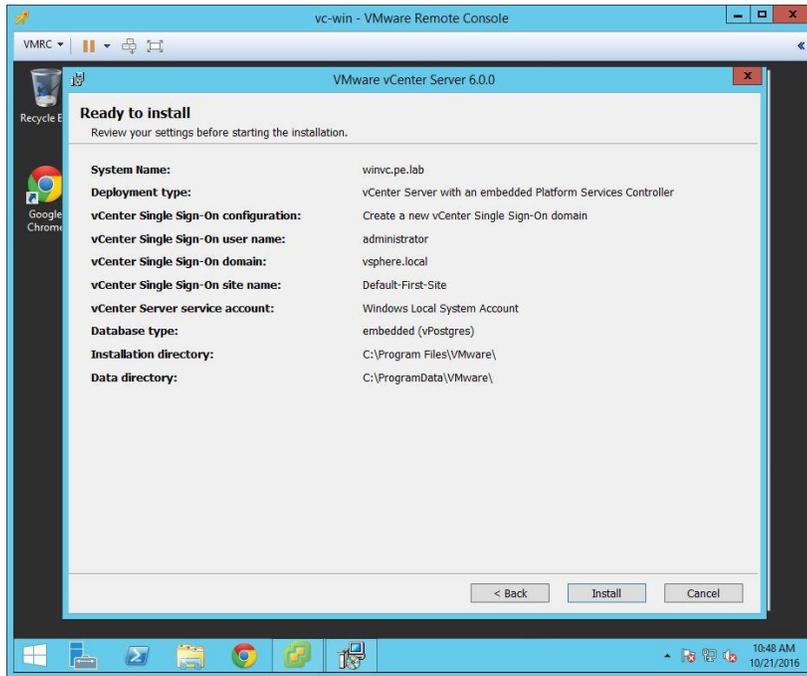


17. In the **Destination Directory** window accept the suggested path. Click **Next** to proceed.

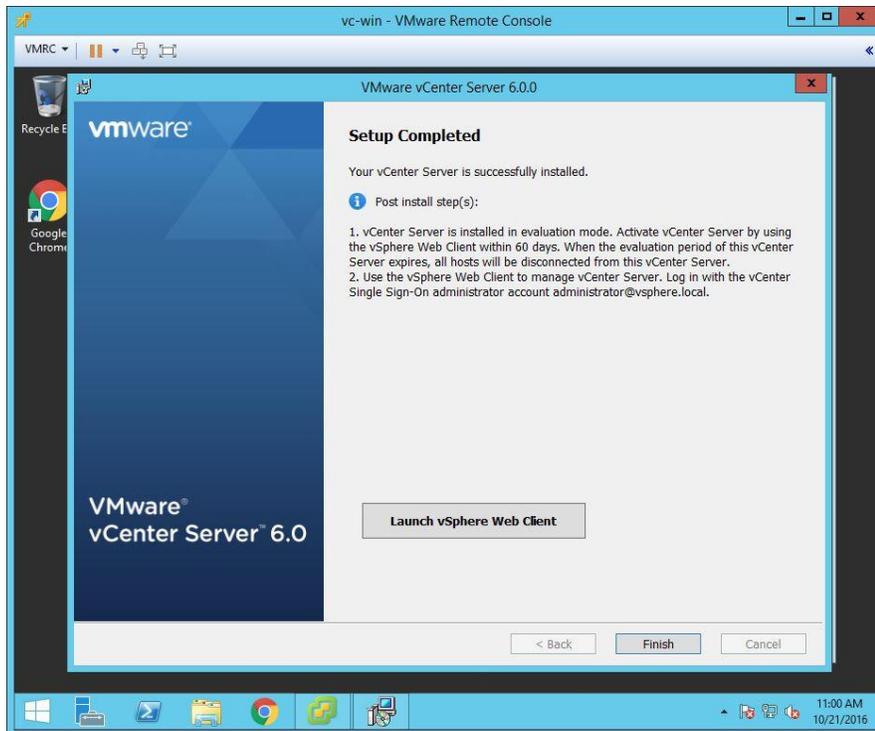


18. The Customer Experience Improvement program sends benign usage data to VMware for the purposes of investigating possible defects and general enhancement of the customer experience. Enroll in the program or not, at the discretion of the installer. Click **Next** to proceed.

The Ready to install window appears.



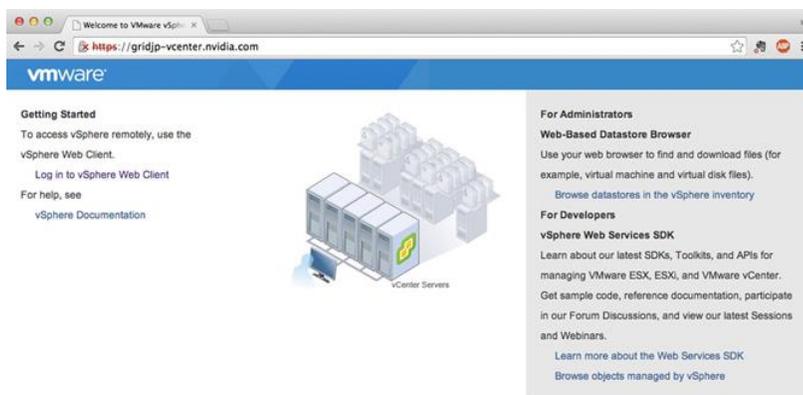
19. Verify that all the selections are correct and click **Install** to proceed. The installation takes place.



20. Click **Finish** to complete the installation process

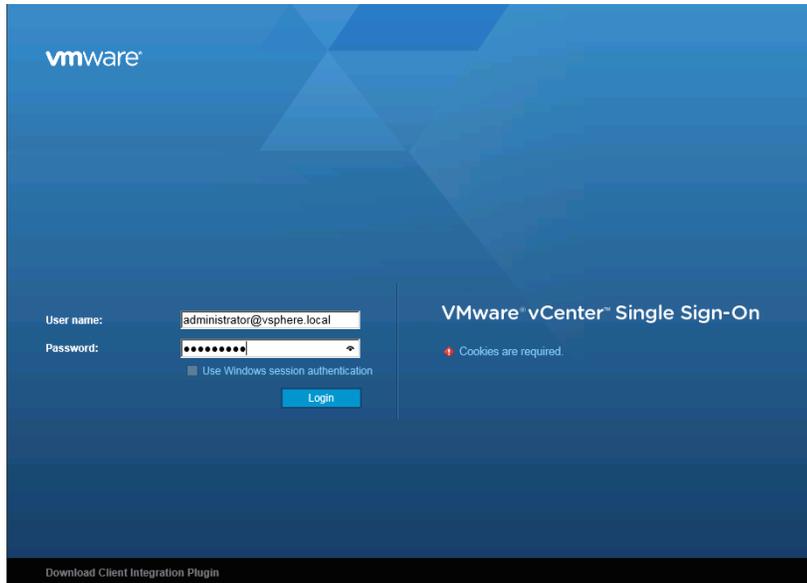
4.2.2 Adding Licenses to your vCenter Server

Use the following procedure to add licenses to your vCenter Server.

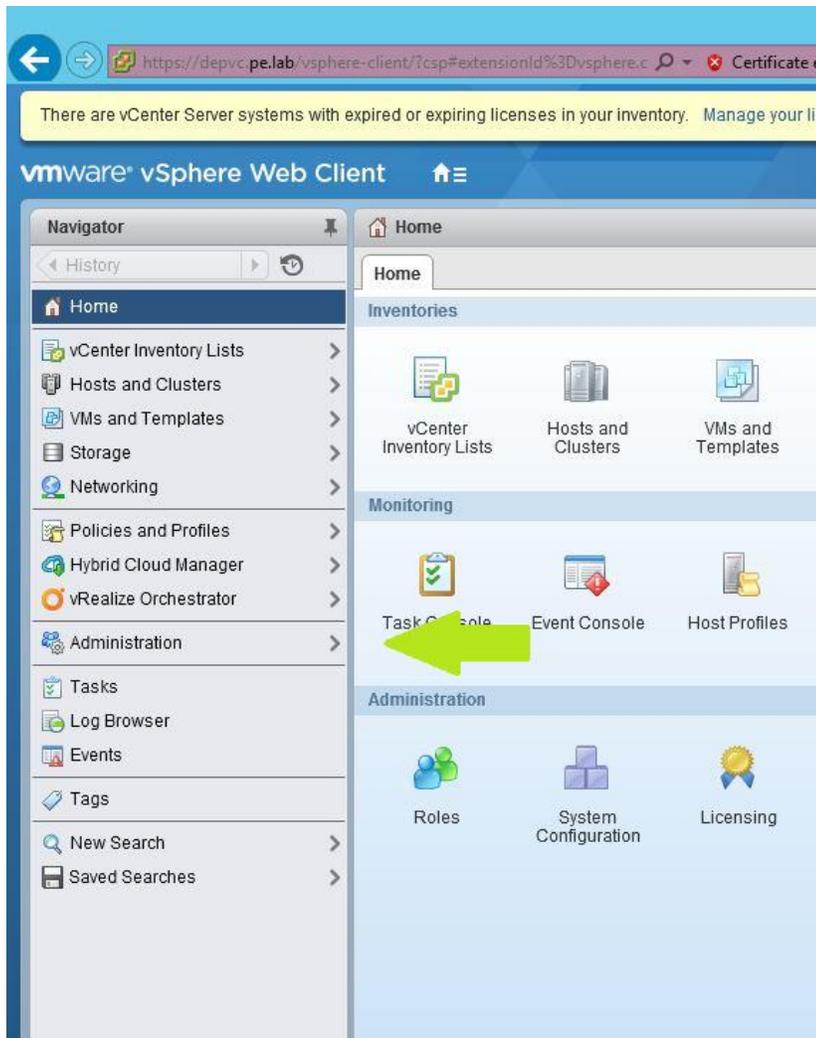


1. Open a web browser and navigate to the vCenter host at `https://<host>`, where `<host>` is either the hostname or IP address. Ignore the untrusted certificate warning if it displays. Select **Log in to vSphere Web client**.

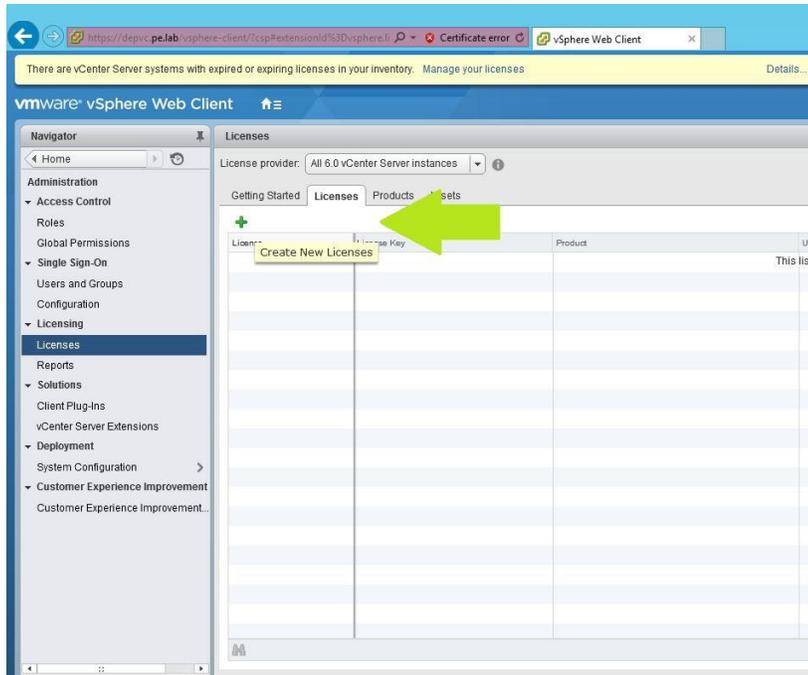
2. Install the VMware Client Support Plug-in by selecting **Always run on this site**, if prompted.



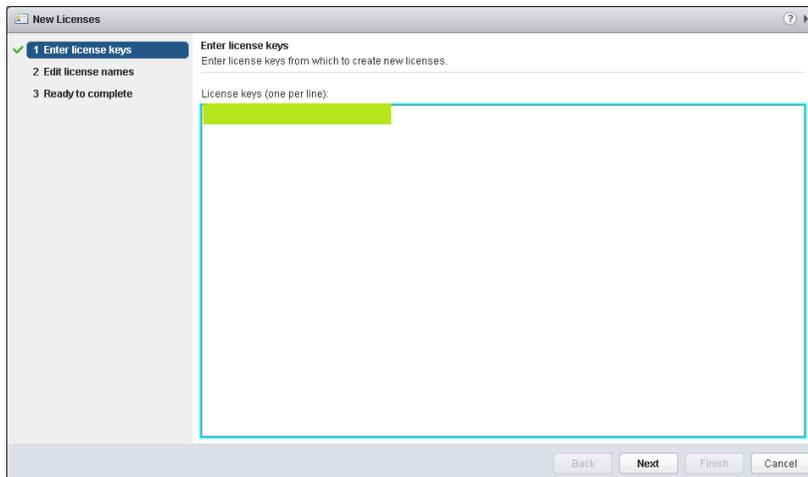
3. Enter User name **administrator@vsphere.local** and password that you specified during installation, and then click the **Login** button.



4. Click the **Home** icon (house) and then select **Administration**.



5. Select **Licenses** from the left-hand menu and then select the **Licenses** tab. Click the green **New License** icon (plus sign) to open the **New Licenses** popup.



6. Enter the vCenter Server license key

The screenshot shows the 'New Licenses' wizard in the VMware vSphere Web Client. The left sidebar indicates the current step is '2 Edit license names'. The main area is titled 'Edit license names' and contains the following information:

- License name: VMware vCenter Server 6 Standard EA
- License key: [Redacted]
- Expires: [Redacted]
- Product: VMware vCenter Server 6 Standard (Instances)
- Capacity: Unlimited L...

Buttons at the bottom include Back, Next, Finish, and Cancel.

7. Enter a unique name for the license in the **License Name** field and then click **Next**.

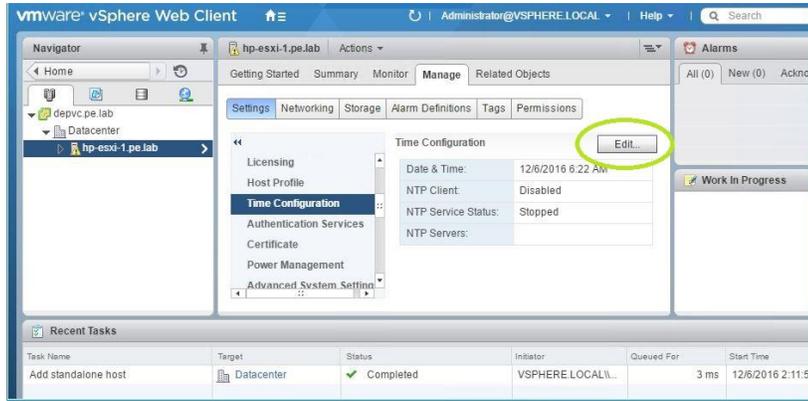
The screenshot shows the 'New Licenses' wizard in the VMware vSphere Web Client. The left sidebar indicates the current step is '3 Ready to complete'. The main area is titled 'Ready to complete' and contains the following information:

- Number of licenses: 1
- License name: VMware vCenter Server 6 Standard EA
- License key: [Redacted]

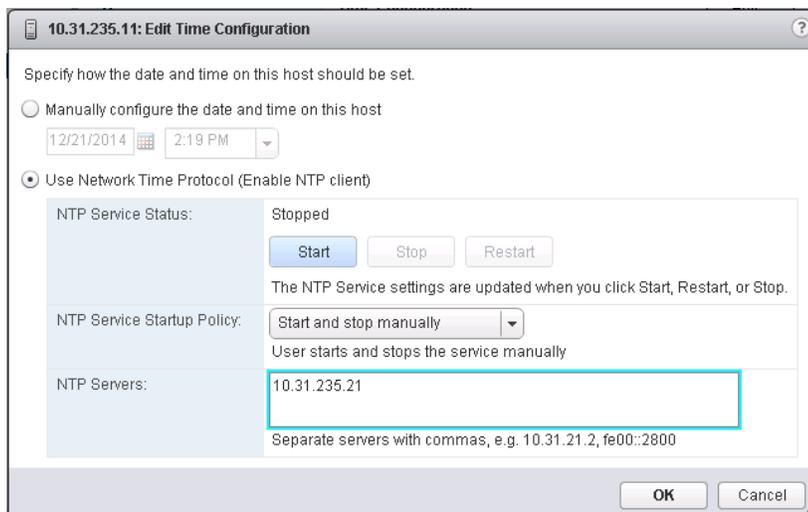
Buttons at the bottom include Back, Next, Finish, and Cancel.

8. Review your selections and then click **Finish** to close the **Enter New License** wizard and return to the VMware vSphere Web Client page.

4.3 SETTING THE NTP SERVICE ON A HOST



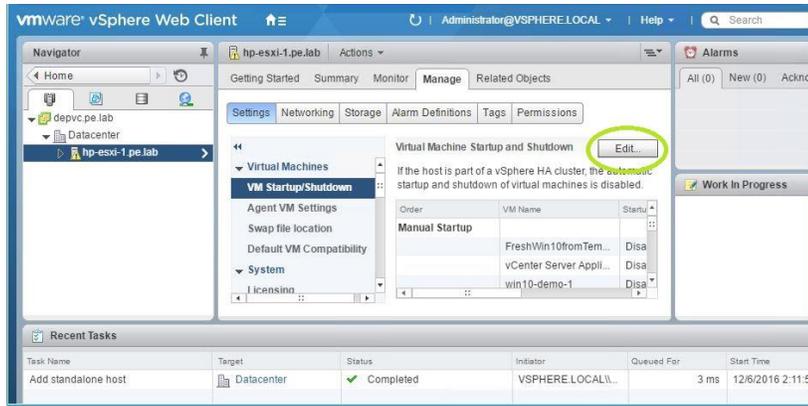
1. Click a host object in the menu on the left, click **Manage** → **Settings** → **Time Configuration** → **Edit**



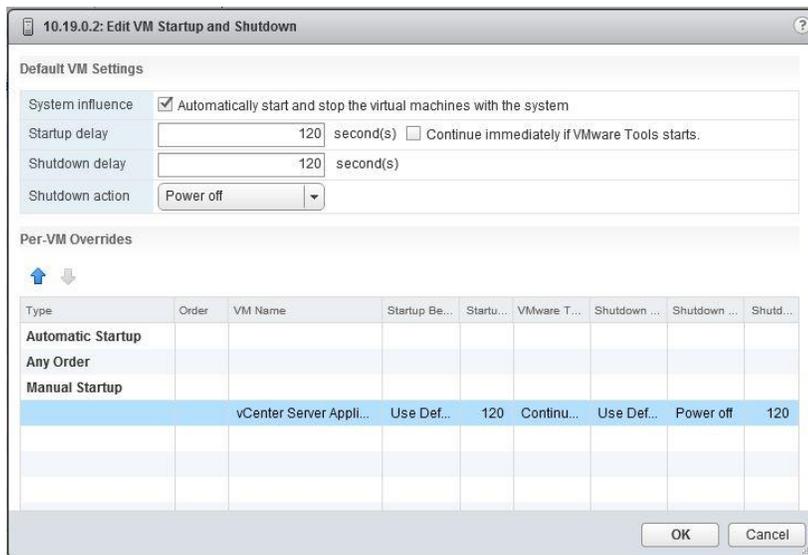
2. Enter a valid time server and click **OK**.
3. Do this for each host to ensure time is accurate for all guests.

4.4 SETTING VCENTER TO AUTO-START

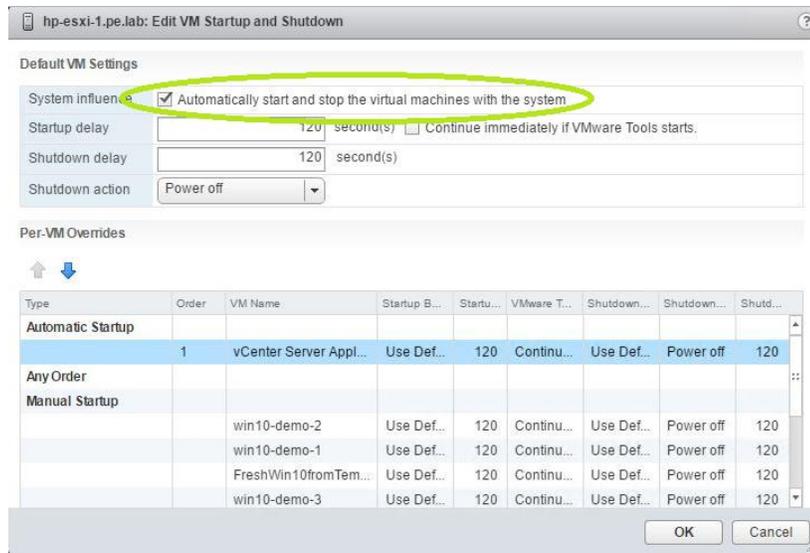
Use the following procedure to set vCenter to start automatically with the management server:



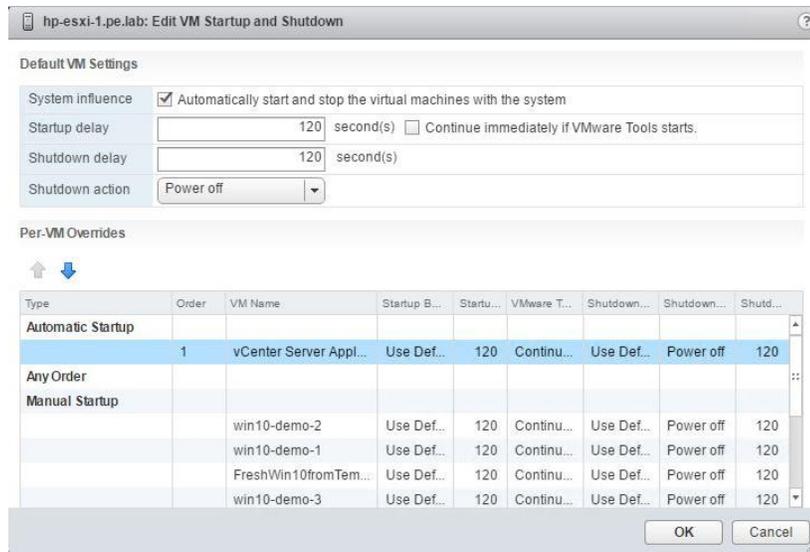
1. In the vSphere Web Client, select the host then select **Manage** → **Settings** → **VM Startup/Shutdown**.



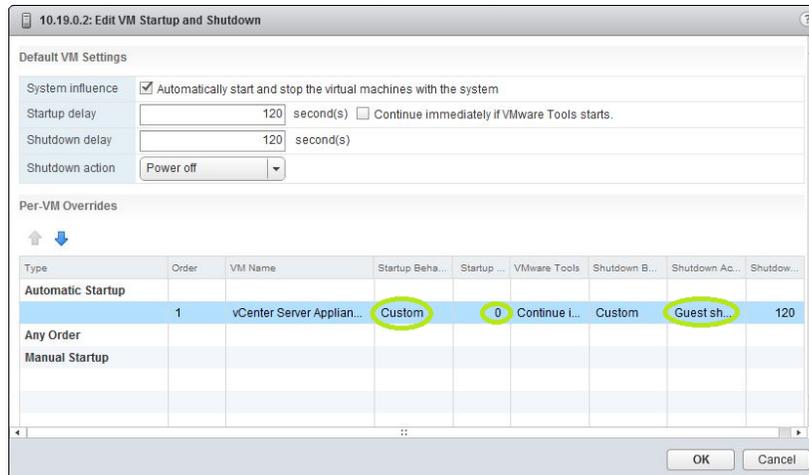
2. Click the **Edit** button.



3. Check the box marked **Automatically start and stop the virtual machines with the system**.



4. Select the **vCenter Appliance** and click the **Up** arrow to move that virtual machine up to the **Automatic Startup** section.



5. Select the following options:

- ▶ Set **Startup Behavior** to **Custom** and **select Continue immediately if VMware Tools starts**
- ▶ Set **Startup Delay** to **0**
- ▶ Set **Shutdown Behavior** to **Custom**
- ▶ Select **Guest Shutdown**

Click **OK** to apply the configuration.



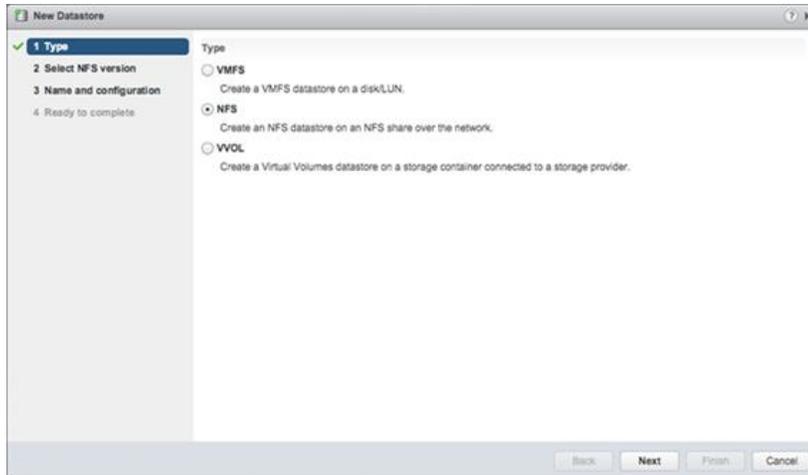
The vCenter Web Client may not reflect these configuration changes immediately. Either click the Refresh icon or a different configuration group and return to the current setting.

4.5 MOUNTING AN NFS ISO DATA STORE

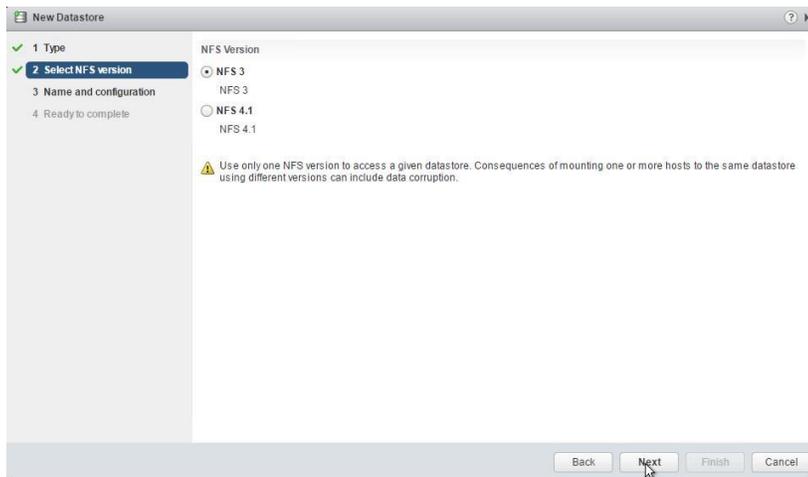
An NFS datastore can come in handy, since so many OS can access it. It becomes a storage point for files used by the ESXi server. We recommend using it to store ISO files for operating system installations, like Microsoft Windows.

Use the following procedure to mount an NFS ISO data store:

1. In the main vSphere Web Client window, select **Hosts and Clusters** and select the host.
2. Select **Storage** → **New Datastore** from the **Actions** drop-down menu.



3. Select **NFS** and click **Next** to proceed.



4. Select the correct NFS version and click **Next** to proceed.

New Datastore

1 Type
2 Select NFS version
3 **Name and configuration**
4 Ready to complete

Datastore name: FreeNAS

NFS Share Details

Folder: /mnt/pool0
E.g. /vol1/vol0/datastore-001

Server: freenas.pe.lab
E.g. nas, nas.it.com or 192.168.0.1

Info If you plan to configure an existing datastore on new hosts in the datacenter, it is recommended to use the "Mount to additional hosts" action instead.

Access Mode

Mount NFS as read-only

Back Next Finish Cancel

5. Enter the NFS exported folder path and the NFS server address in the **Folder** and **Server** fields, respectively. Click **Next** to proceed.

Since this an ISO data store, consider mounting it as read-only by checking the **Mount NFS as read-only** checkbox.

New Datastore

1 Type
2 Select NFS version
3 Name and configuration
4 **Ready to complete**

General:

Name: FreeNAS
Type: NFS
Version: NFS 3

NFS share and access mode:

Server: freenas.pe.lab
Folder: /mnt/pool0
Access Mode: Read-write

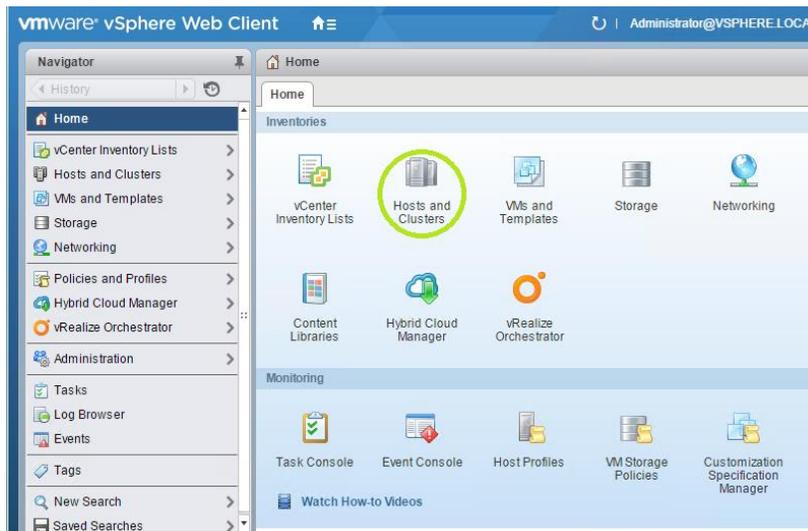
Back Next Finish Cancel

6. Review the settings. Click **Finish** to complete adding the NFS ISO data store. This data store is now accessible as an installation source for virtual machine CD drives.

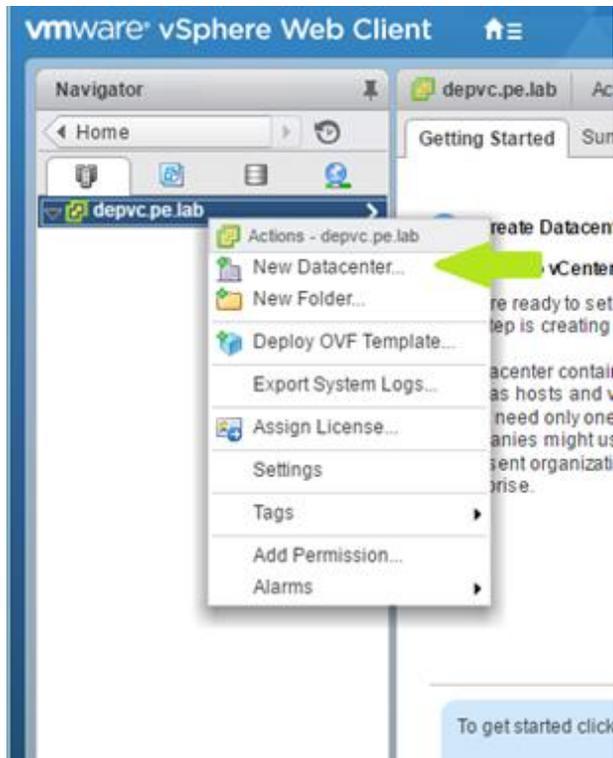
4.6 CREATING VSPHERE CLUSTERS FOR GRID SERVERS

This section covers adding logical organizational units in vCenter to cluster hosts and manage them under a data center object. This is especially important for distributing GRID vGPU-enabled desktops across clustered vSphere hosts with NVIDIA Tesla GPUs.

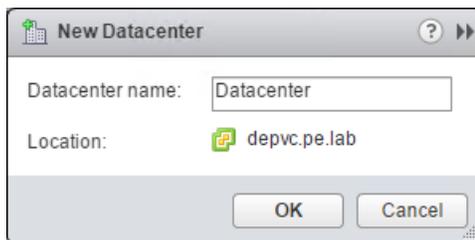
4.6.1 Creating a Datacenter in vCenter



1. Log in to vCenter Web Client and click on **Hosts and Clusters**.

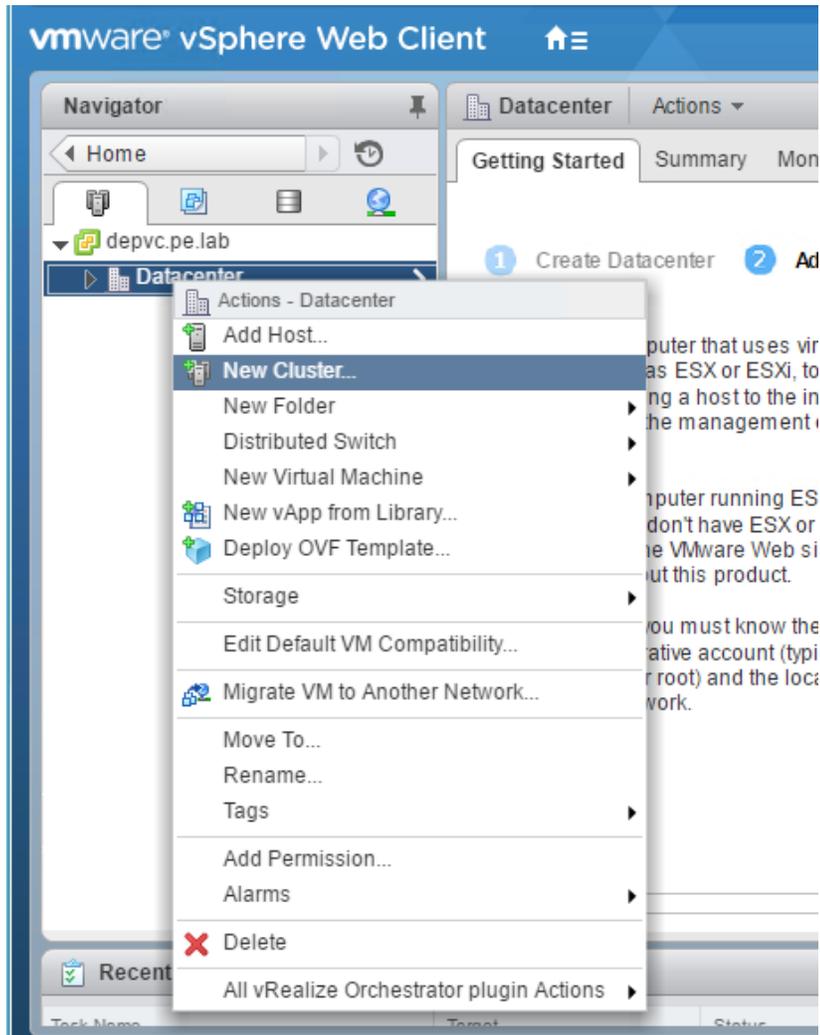


2. In the left-hand **Navigator** pane, right click on the vCenter and select **New Datacenter** from the pop-up menu.

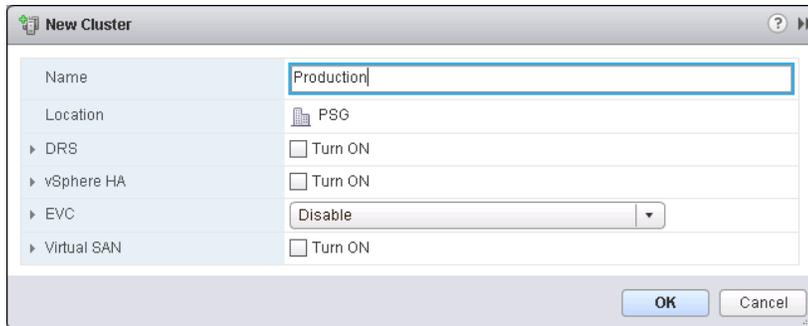


3. On the resulting pop-up, enter a name for the new Datacenter, and click **OK**.

4.6.2 Creating a Cluster in the Datacenter



1. In the left-hand **Navigator** pane, right click on the **Datacenter** object, then, on the menu that pops up, click **New Cluster...**

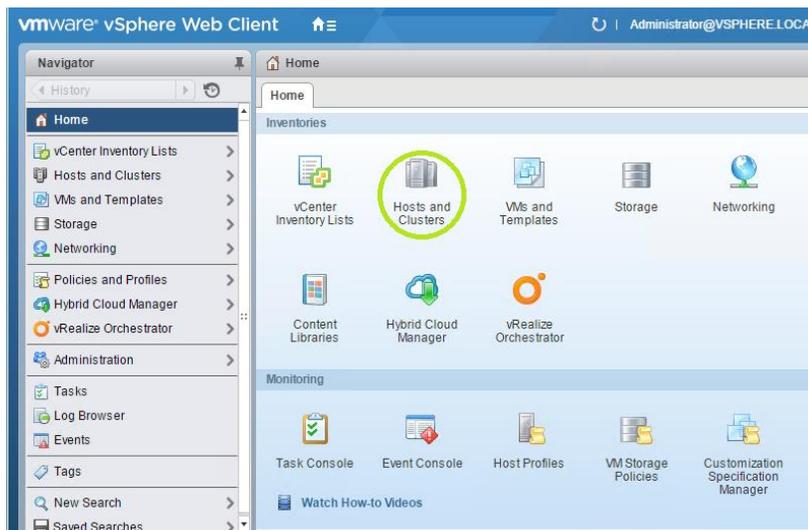


2. In the **New Cluster** dialog box, enter **Production** in the **Name** area. Click **OK**.
3. Repeat this procedure to create a cluster named **Management**.

You should now have a logical organizational structure of clusters in the datacenter: one for management, the other for production VDI.

4.6.3 Adding a Host

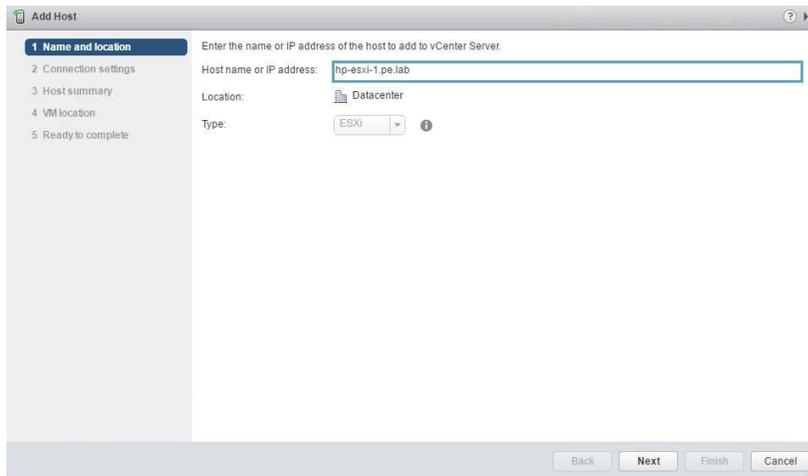
Use the following procedure to add a host in the vCenter cluster.



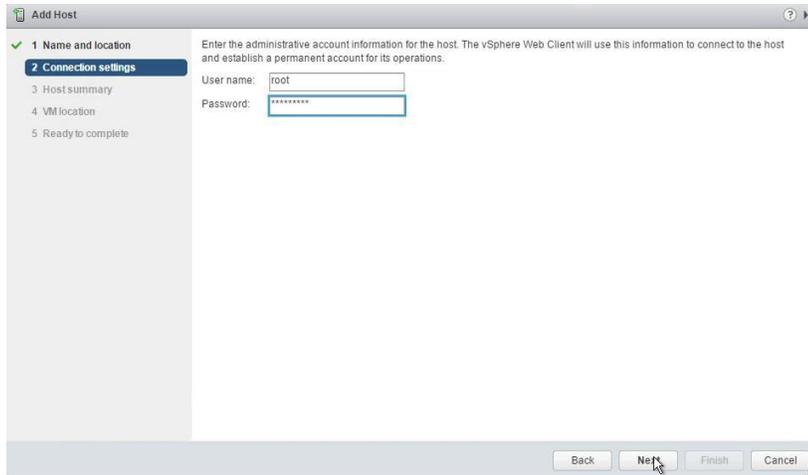
1. Select **Hosts and Clusters**.



2. In the left-hand **Navigation** pane, select the vCenter. Select the **Getting Started** tab and select **Add a Host**.



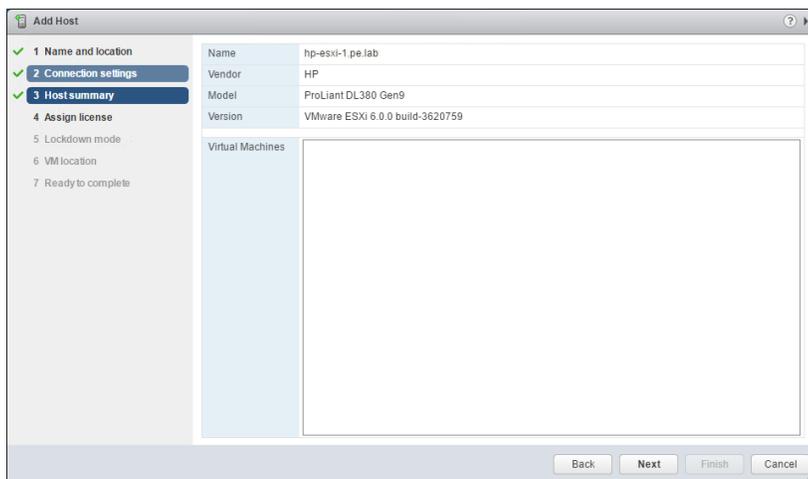
3. Enter the host name or IP address of the vSphere host and click **Next**.



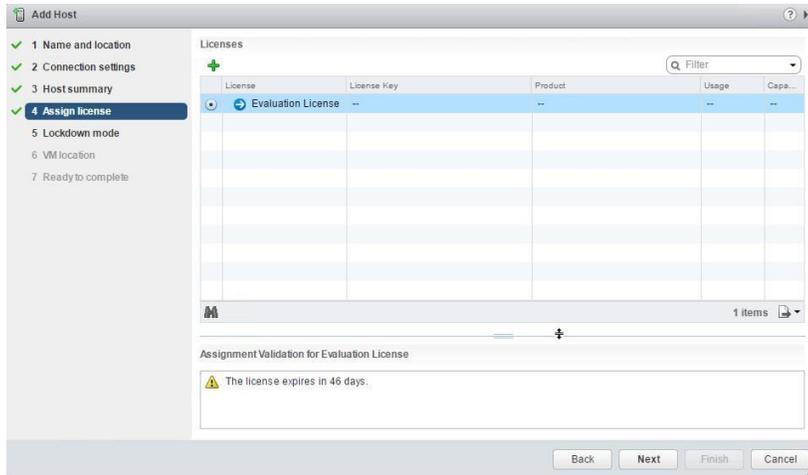
4. Enter the administrator account credentials in the **User name** and **Password** fields and click **Next**.



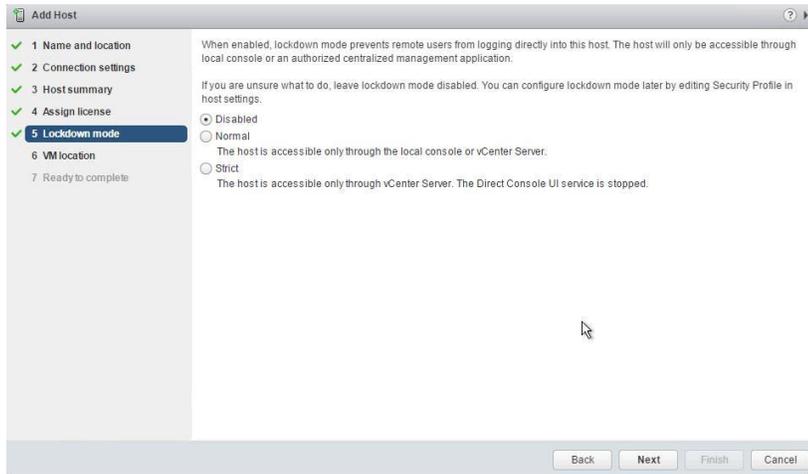
5. Click **Yes** to replace the host certificate.



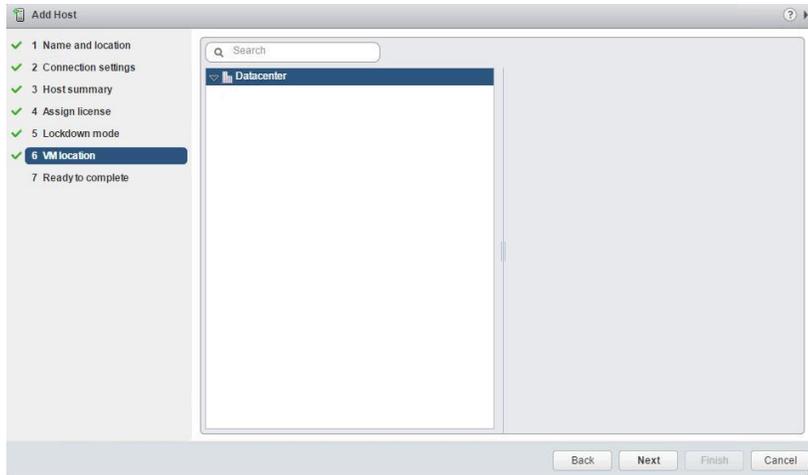
6. Review the settings and click **Next** to proceed.



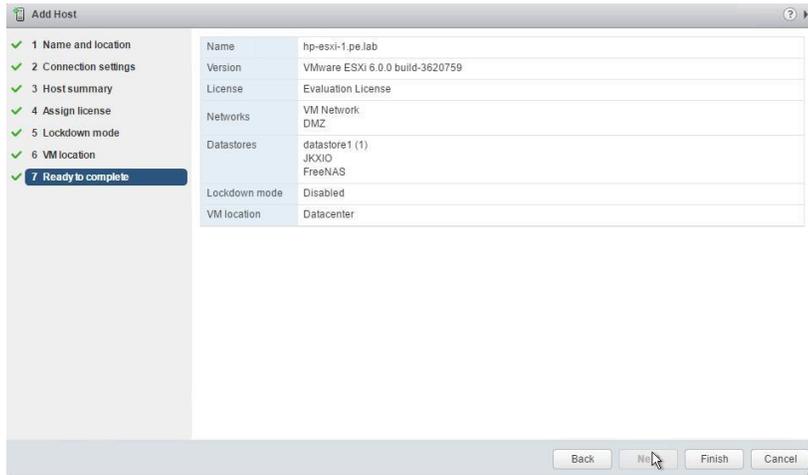
7. Confirm the license selection and click **Next**



8. Accept the default setting (**Disabled**) and click **Next**.



9. Select a cluster or accept the default option and click **Next** to proceed.



10. Click **Finish** to complete adding the new host.



11. The new host is now visible in the left panel when you click the datacenter name.

CHAPTER 5. GRID SOFTWARE PREPARATION

The following steps provide needed software and licenses for your GRID installation.

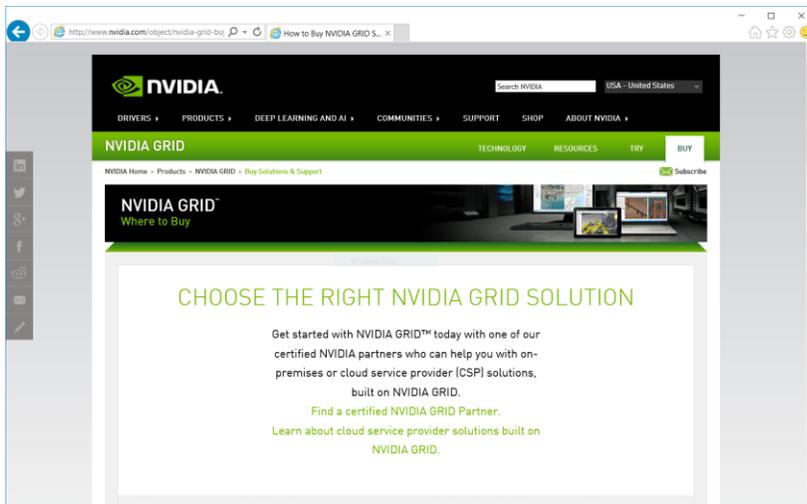
They include:

- ▶ Obtaining software and evaluation licenses
- ▶ Creating the License Manager server
- ▶ Registering the License Manager server
- ▶ Installing license files in the License Manager server
- ▶ Replacing evaluation licenses with permanent licenses
- ▶ How to obtain Support, Update, and Maintenance (SUMS)

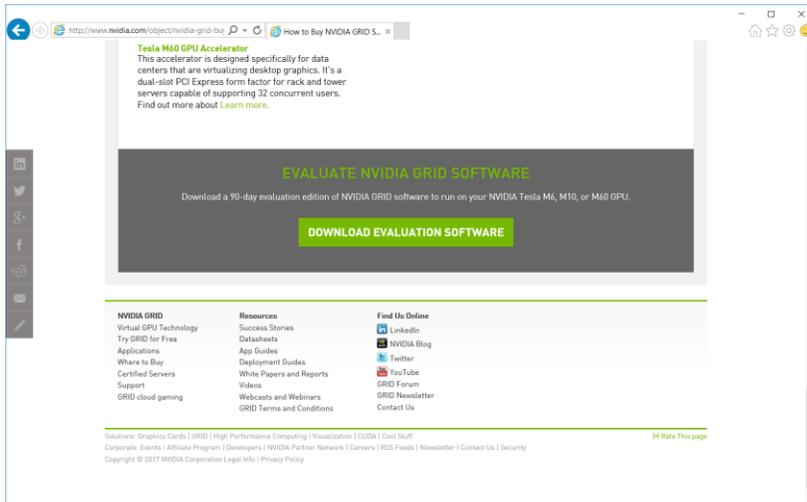
After we have created the first desktop VM, we will configure it to receive licenses from the License Manager server.

5.1 OBTAIN AN EVALUATION LICENSE PAK AND DOWNLOAD SOFTWARE

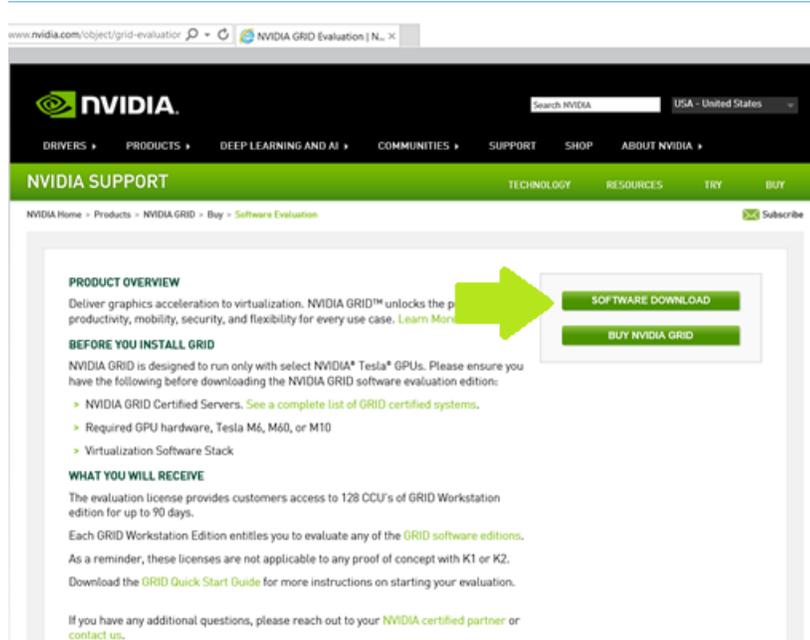
Use the following directions to obtain the GRID software and evaluation licenses.



1. In a browser navigate to <http://www.nvidia.com/buygrid>.



2. Scroll to the bottom of the page. Click on **Download Evaluation Software**.



3. Click on **Software Download**. The **NVIDIA Enterprise** window opens.

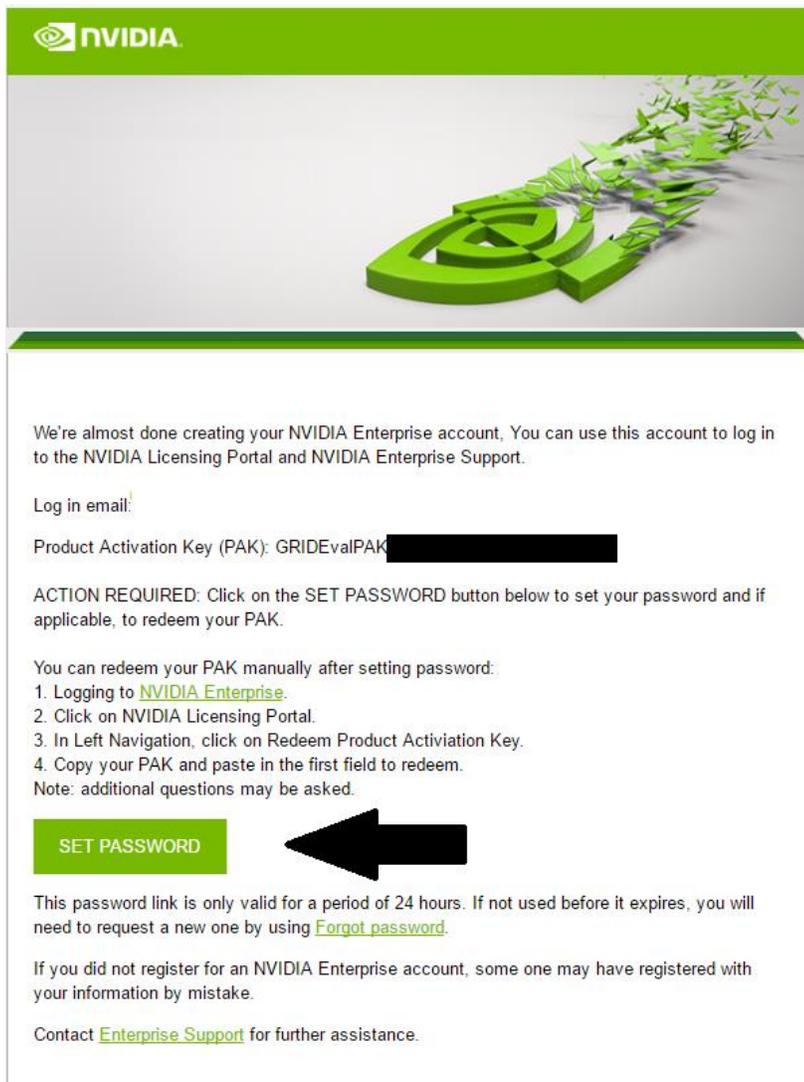
The screenshot shows the 'CREATE YOUR NVIDIA ENTERPRISE ACCOUNT' form. The form is titled 'CREATE YOUR NVIDIA ENTERPRISE ACCOUNT' and includes a sub-header 'NVIDIA. Enterprise'. Below the title, there is a note: 'If you need assistance with registration, please email enterprisesupport@nvidia.com'. The form is labeled '*Required Fields' and contains the following input fields:

- *Email address:
- *Company:
- *First name:
- *Last name:
- *Job role:
- *Industry:
- *Phone:
- *Street 1:
- Street 2:

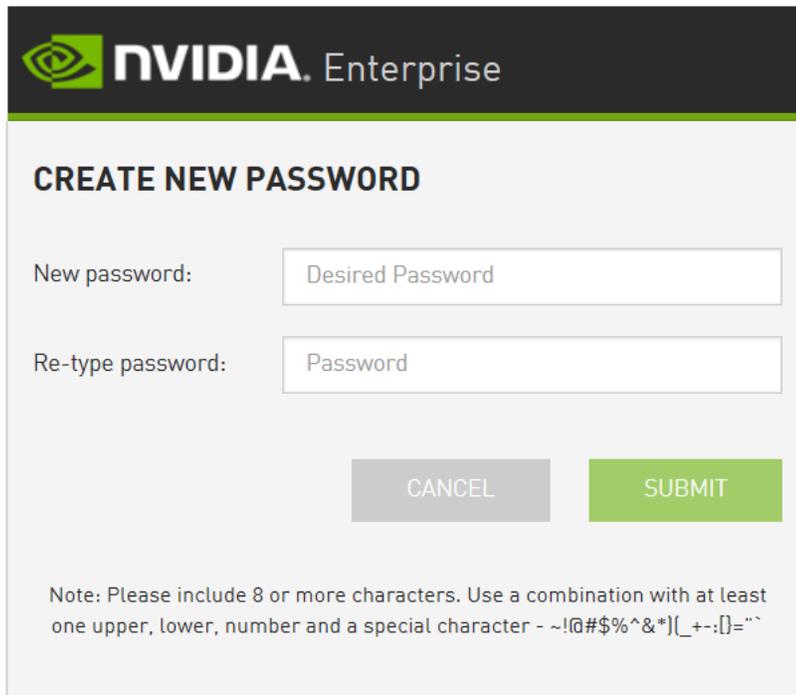
4. Fill out the form completely and click **Register**. Once your account has been created, you will receive an email with further instructions.



5. Check your email for licensing instructions.

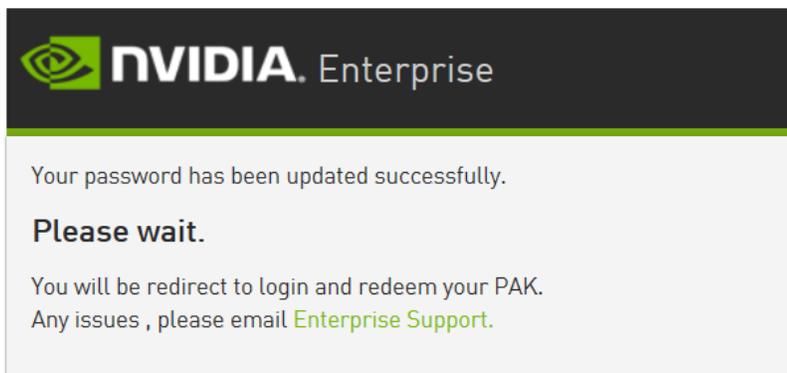


6. In the email you received with your Product Activation Key, click on the **Set Password** icon. This will open a browser window.



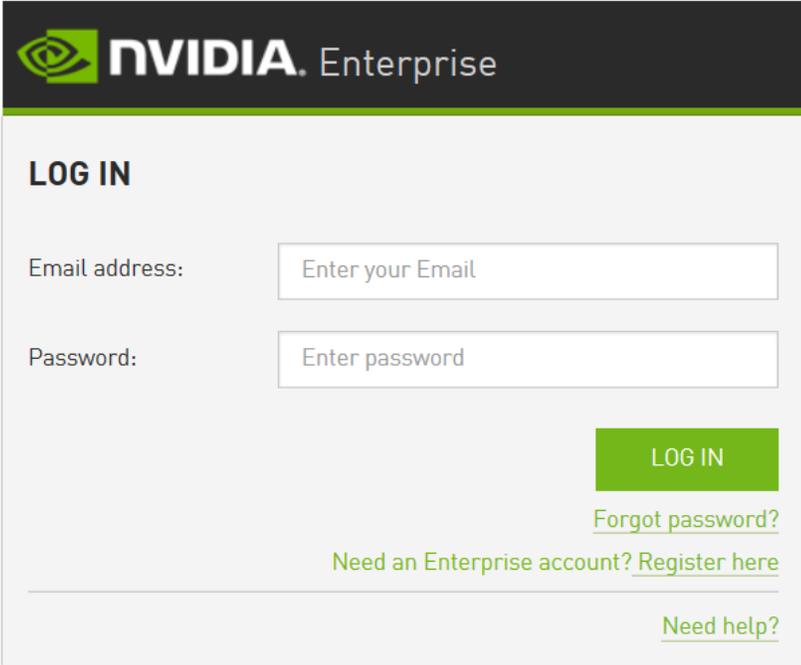
The screenshot shows the NVIDIA Enterprise password creation interface. At the top is the NVIDIA logo and the word 'Enterprise'. Below this is the heading 'CREATE NEW PASSWORD'. There are two input fields: 'New password:' with the placeholder text 'Desired Password' and 'Re-type password:' with the placeholder text 'Password'. Below the fields are two buttons: a grey 'CANCEL' button and a green 'SUBMIT' button. At the bottom, there is a note: 'Note: Please include 8 or more characters. Use a combination with at least one upper, lower, number and a special character - ~!@#\$%^&*][_+~:]=``'.

7. In the Set Password window, enter a new password, re-type, and click **Submit**.

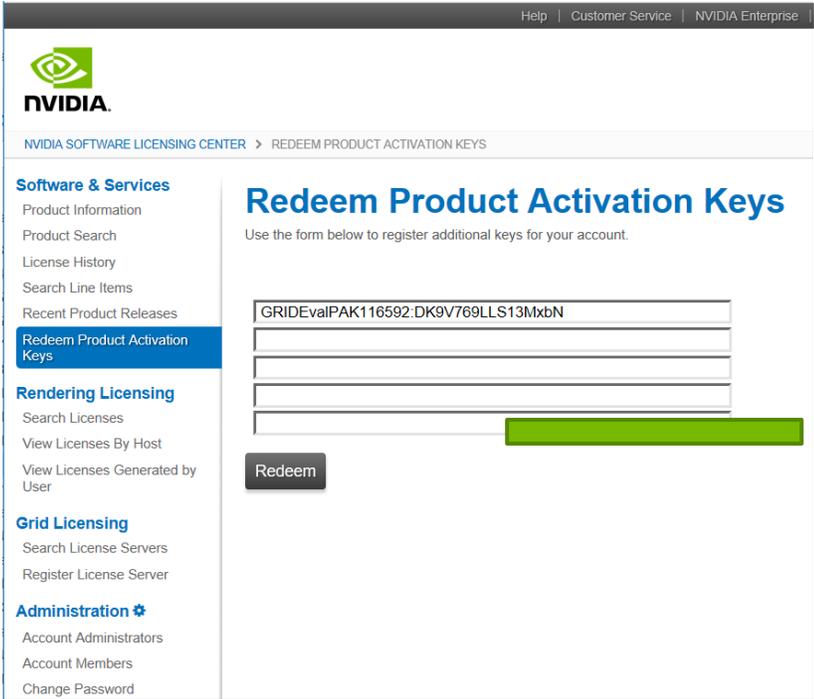


The screenshot shows a success message from NVIDIA Enterprise. At the top is the NVIDIA logo and the word 'Enterprise'. Below this is the text 'Your password has been updated successfully.' followed by the bold heading 'Please wait.' and the text 'You will be redirect to login and redeem your PAK. Any issues , please email [Enterprise Support](#).'

8. If successful, you will be advised, and redirected to the Login page.



9. Login with your email address and new password.



10. On the **Redeem Product Activation Keys** page, your activation key will be listed in the box. Click **Redeem**.

The screenshot shows the NVIDIA Software Licensing Center interface. The main heading is "Register Product Activation Key". The form includes the following fields and options:

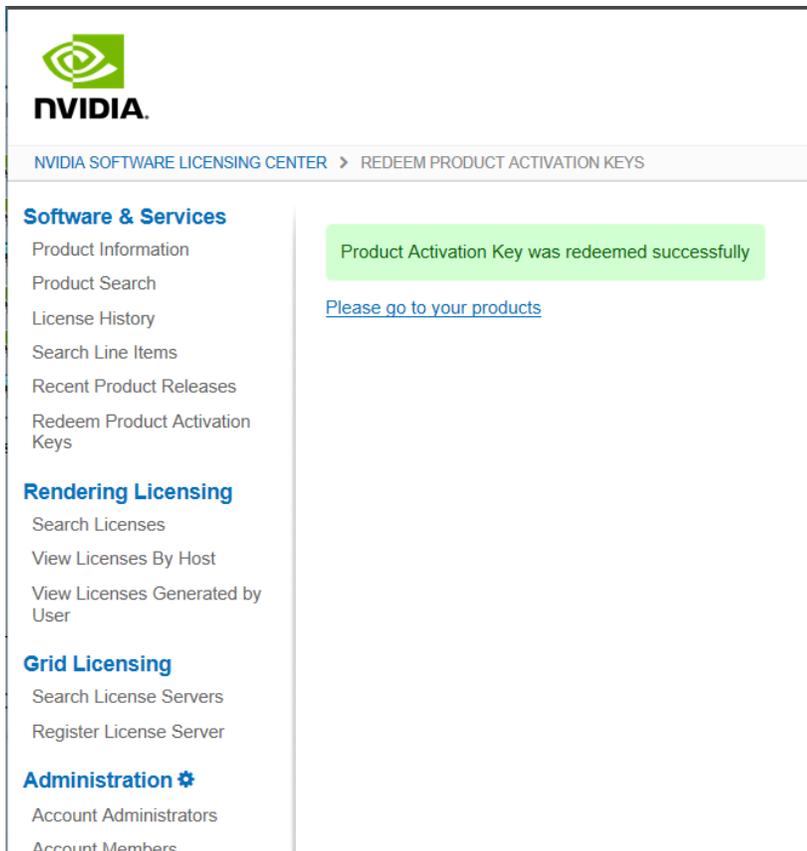
- Department: [Text Input]
- Number of VDI Seats: * [Dropdown Menu: -- Please Choose One --]
- Timeframe for Deployment: * [Dropdown Menu: -- Please Choose One --]
- Who is your Preferred Partner? * [Text Input]
- Design services or applications to be enabled with GRID: * [Text Input]
- Register for Nvidia Newsletters:

At the bottom of the form are "Submit" and "Cancel" buttons.

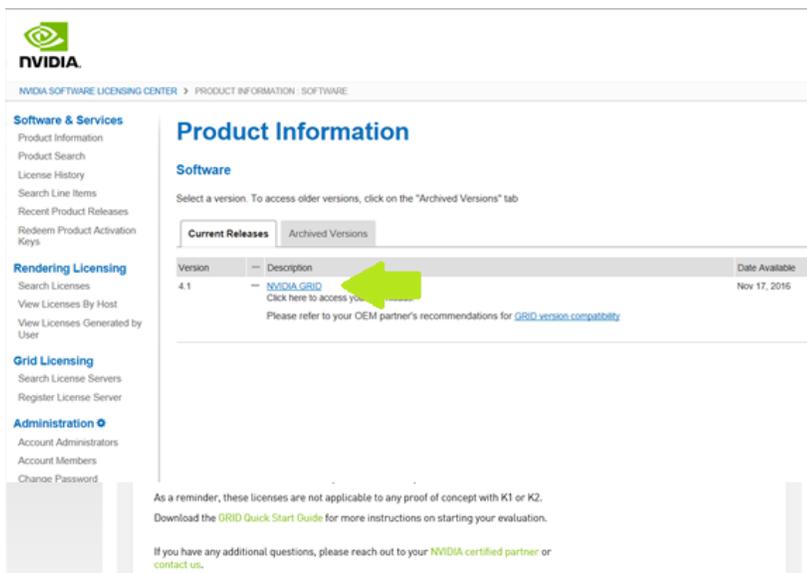
11. Fill out the **Register Product Activation Key** form and click **Submit**.

This screenshot shows the same "Register Product Activation Key" page, but the form fields are empty, indicating that the form has been submitted. The "Submit" and "Cancel" buttons are still visible at the bottom of the form area.

12. Click **Submit**.



13. Click **Please go to your products.**



14. On the **Product Information** page, click the **NVIDIA GRID** hyperlink.

Software Terms and Conditions

NVIDIA GRID

Please read the following agreement and select [ACCEPT_EULA] at the bottom before downloading your software.

END USER LICENSE AGREEMENT
Release Date: April 13, 2016

NVIDIA SOFTWARE END-USER LICENSE AGREEMENT

IMPORTANT – READ BEFORE DOWNLOADING, INSTALLING, COPYING OR USING THE LICENSED SOFTWARE.

READ CAREFULLY: This Enterprise End User License Agreement ("EULA"), made and entered into as of the time and date of click through action ("Effective Date"), is a legal agreement between you and NVIDIA Corporation ("NVIDIA") and governs the use of the NVIDIA computer software and the documentation made available for use with such NVIDIA software. By downloading, installing, copying, or otherwise using the NVIDIA software and/or documentation, you agree to be bound by the terms of this EULA. If you do not agree to the terms of this EULA, do not download, install, copy or use the

15. Read the End User License Agreement. Click **I Agree**.

Product Download

NVIDIA GRID

- Download appropriate version of the NVIDIA GRID software below. You can refer to the [Product Support Matrix](#) to determine which software package is right for you.
 - [GRID for vSphere 6.5](#) / [GRID for vSphere 6.0](#) / [GRID for XenServer 7.0](#) / [GRID for XenServer 6.5](#) / [GRID for Windows](#) / [GRID for RHEL KVM](#) / [GRID for LVP](#)
- Download the GPU mode change utility here: [Mode Change Utility](#)
- Download the NVIDIA License Manager here: [License Manager for Windows](#) / [License Manager for Linux](#)
- After you install the license server, read the MAC ID displayed on it. Use this MAC ID to map required number of licenses and download license file by clicking on "Create License Server" in the left side panel.

Thank you for your NVIDIA Enterprise software product purchase.

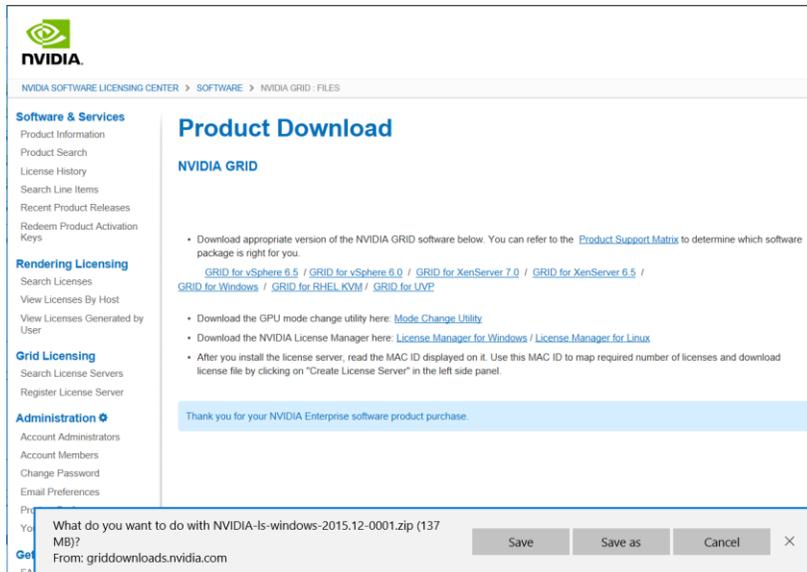
16. On the **Product Download** page you will find links to all available versions of GRID software as well as the GPU Mode Change Utility and the License Manager. Click on **GRID for vSphere 6.0**.

The screenshot shows the NVIDIA Software Licensing Center interface. The main content area is titled 'Product Download' and 'NVIDIA GRID'. It contains several bullet points and links for downloading software. A modal dialog box is overlaid at the bottom, with the text: 'What do you want to do with NVIDIA-GRID-vSphere-6.0-367.64-369.71.zip (0.97 GB)?'. The dialog has three buttons: 'Save', 'Save as', and 'Cancel', along with a close icon (X).

17. Click **Save**.

The screenshot shows the NVIDIA Software Licensing Center interface. The main content area is titled 'Product Download' and 'NVIDIA GRID'. It contains several bullet points and links for downloading software. A modal dialog box is overlaid at the bottom, with the text: 'What do you want to do with NVIDIA-gpumodeswitch-2016-04.zip (96.6 MB)?'. The dialog has three buttons: 'Save', 'Save as', and 'Cancel', along with a close icon (X).

18. Click on **Mode Change Utility** and save the resulting file.



19. Click on **License Manager for Windows** and save the resulting file.

5.2 CREATE THE LICENSE MANAGER SERVER

When enabled on Tesla GPUs, licensed editions of GRID products are activated by obtaining a license over the network from an NVIDIA GRID License Server. The license is “checked out” or “borrowed” at the time the VM is booted, and returned when the VM is shut down.

Use the following instructions to create an NVIDIA License Manager server.

5.2.1 Prepare the License Manager VM

The NVIDIA GRID License Manager can be installed on stand-alone hardware on in a VM. For the purposes of this procedure a VM will be used.

The license server is supported on Windows and Linux hosts. Check the GRID License Server Release Notes for supported OS.

Use the following procedure to create a Windows based License Manager server.

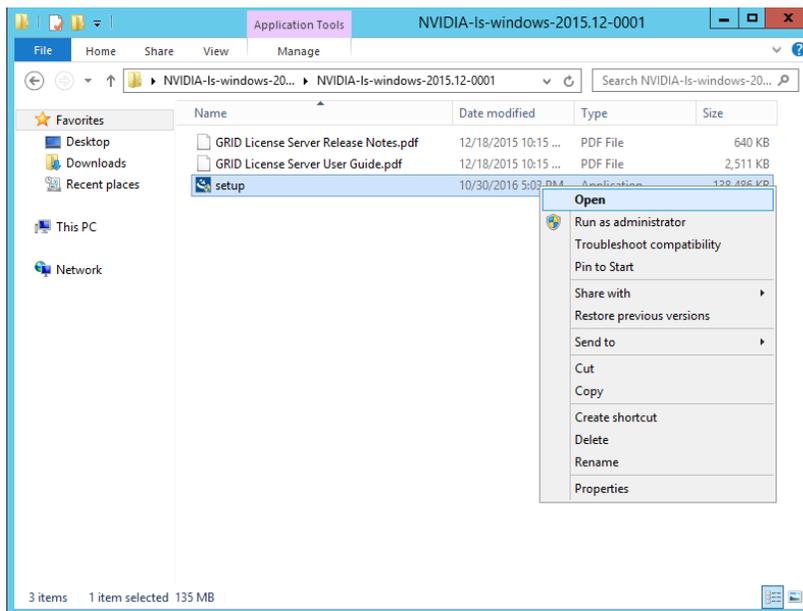
1. Create a VM with a supported version of Windows installed. Refer to sections 10.1 - *Creating a Virtual Machine* through 10.3 - *Installing VMware Tools on the Virtual Machine* for instructions.

- ▶ The recommended minimum configuration is 2 CPU cores, 4 gigabytes of RAM. A high-end configuration of 4+ CPU cores with 16 gigabytes of RAM is suitable for handling up to 150k licensed clients.
- ▶ The hosting platform must run a supported Windows OS.
- ▶ The platform must have a fixed (unchanging) IP address. The IP address may be assigned dynamically via DHCP or statically configured, but must be constant.
- ▶ The platform must have 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 platform's date/time must be set accurately.

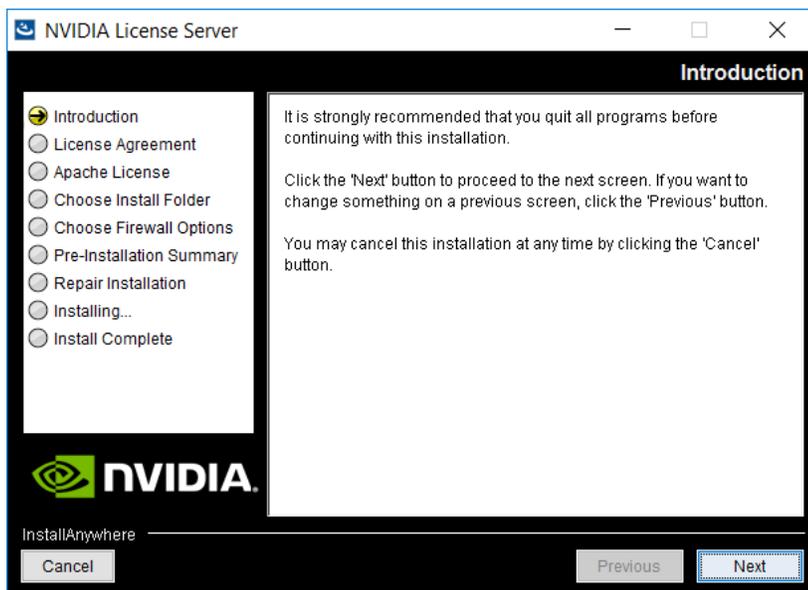
Prior to running the NVIDIA license server installer, download the current Java 32-bit runtime environment from www.oracle.com and install it.

Install the 32-bit Java runtime, regardless of whether your platform is Windows 32-bit or 64-bit.

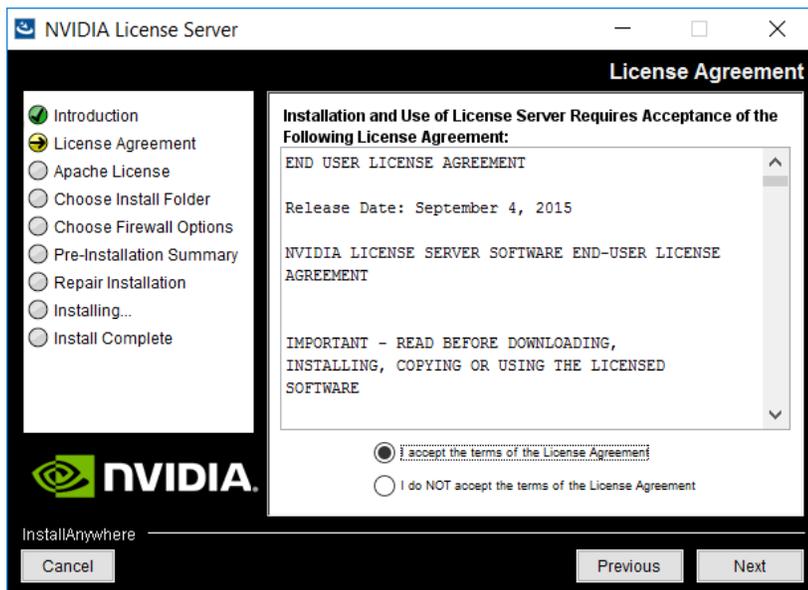
5.2.2 Install the License Server



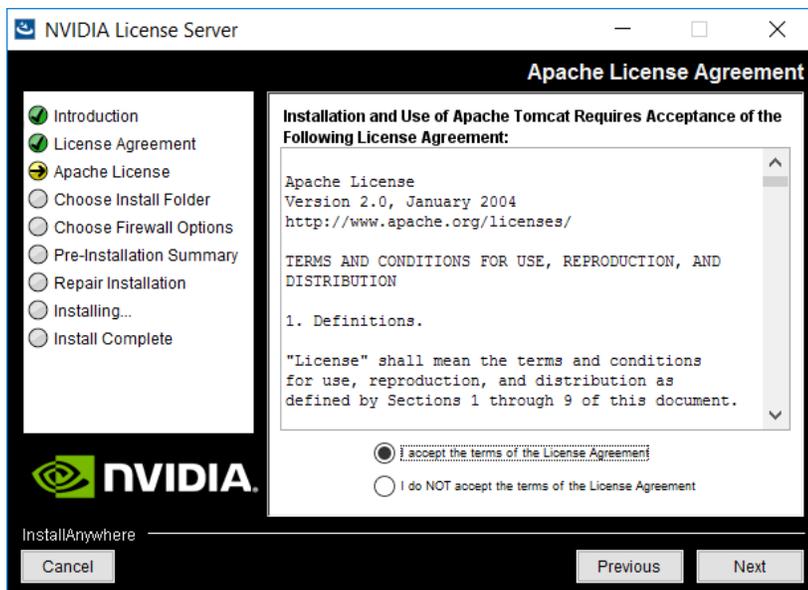
1. Unzip the license server installer and run **setup.exe**.



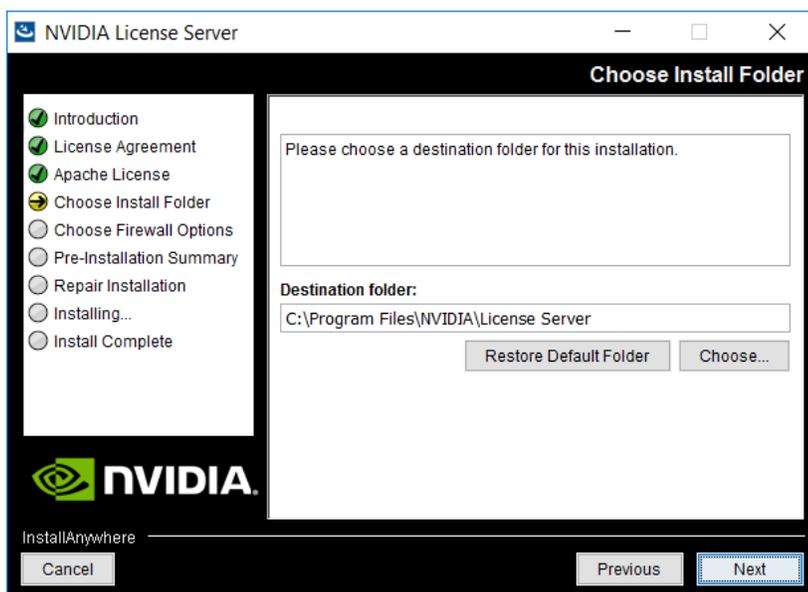
2. On the introduction screen click **Next**.



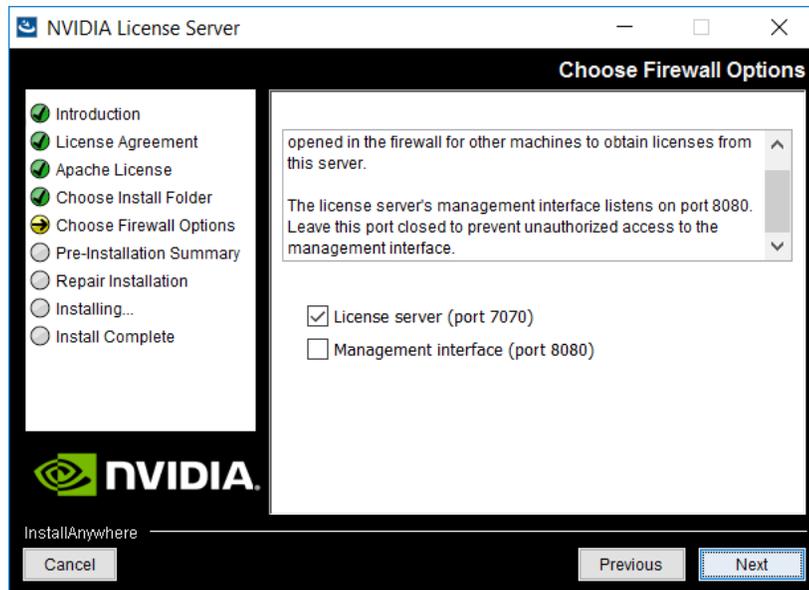
3. Accept the EULA for the license server software and click **Next**.



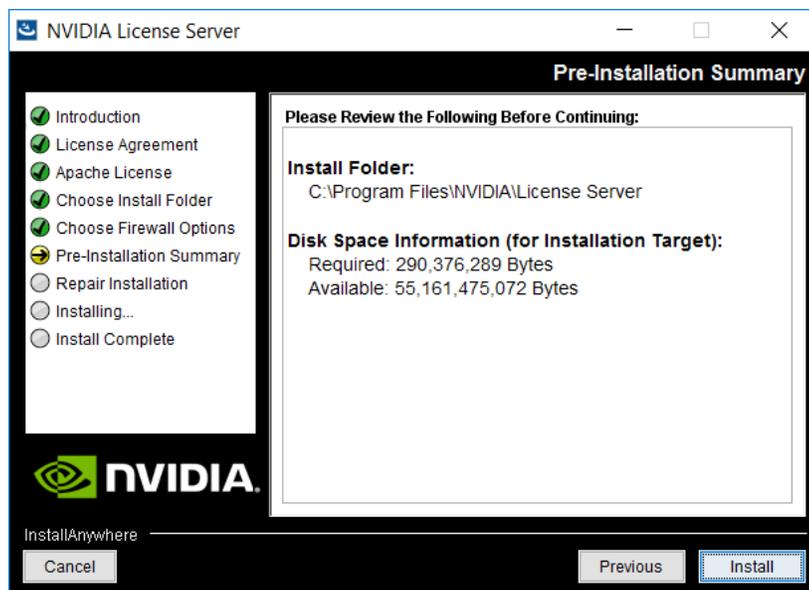
4. Accept the EULA for the Apache Tomcat software used to support the license server's management interface. Click **Next**.



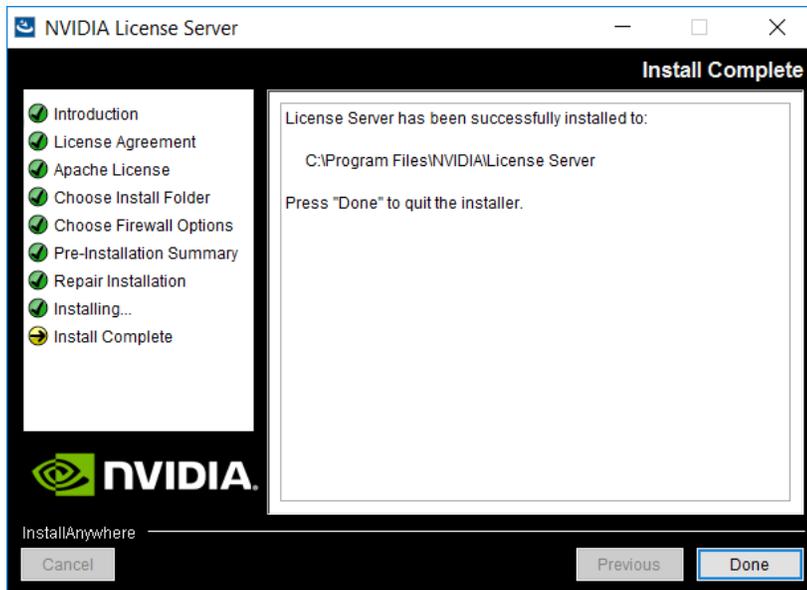
5. Accept the default destination folder and click **Next**.



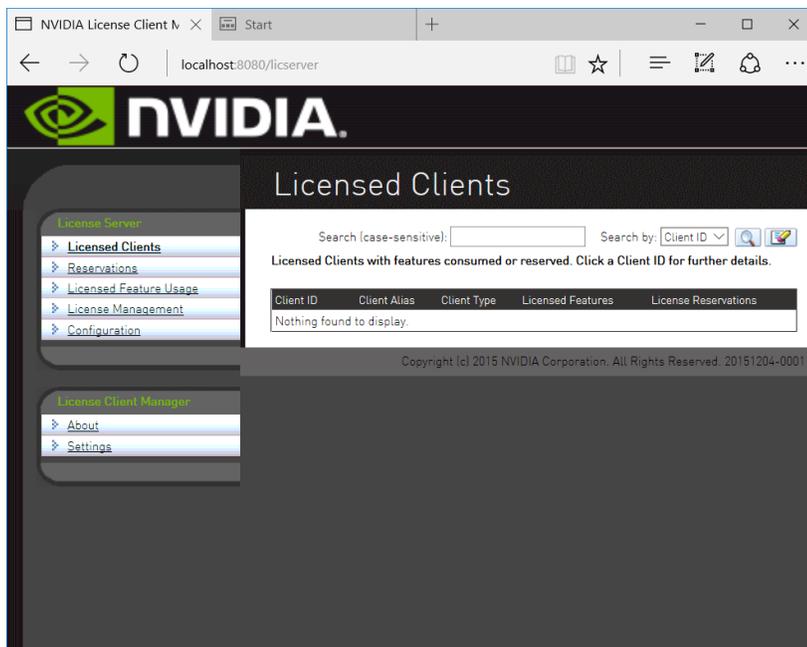
- On the Firewall Options dialog, select the ports to be opened in the firewall. We recommend you use the default setting, which opens port 7070 so that remote clients can access licenses from the server, but leaves port 8080 closed, so that the management interface is only available via web browser running locally on the license server host. Click **Next**.



- Click **Install** to commence installation.



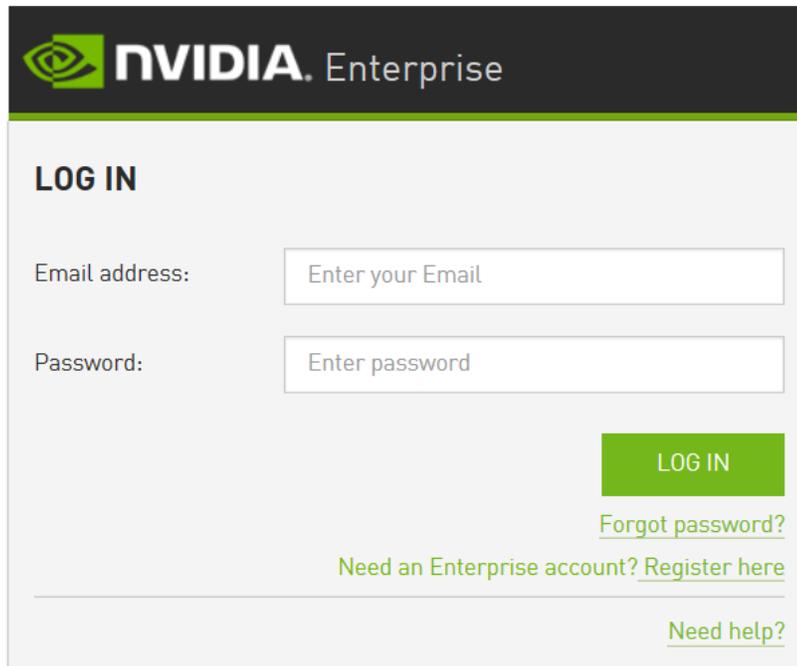
8. Click **Done**.



9. To verify that the license server is operating correctly, open a web browser on the license server host and connect to URL <http://localhost:8080/licserver>. The browser should display the management interface home page, as shown.

5.3 OBTAIN AND INSTALL YOUR LICENSE

Licenses can be purchased from NVIDIA (<http://www.nvidia.com/buygrid>) or customers may evaluate the licensing for a trial period. Customers who purchase licenses should receive an email with direction on how to implement licensing. Evaluation customers will immediately see a link connecting them to their evaluation keys.



The screenshot shows the NVIDIA Enterprise login interface. At the top, there is a dark header with the NVIDIA logo and the word "Enterprise". Below the header, the page is titled "LOG IN". There are two input fields: "Email address:" with a placeholder "Enter your Email" and "Password:" with a placeholder "Enter password". To the right of the password field is a green "LOG IN" button. Below the button are three links: "Forgot password?", "Need an Enterprise account? Register here", and "Need help?".

1. In a browser navigate to <https://nvidia.flexnetoperations.com>. Log in with your previously created credentials. If you have not created credentials, go to Section 5.1 Obtain an evaluation license PAK and download software on page 96

The screenshot shows the NVIDIA Software Licensing Center interface. The browser address bar displays <https://nvidia.flexnetoperations.com/control/nvda/home>. The page title is "Product Information". On the left sidebar, under "Grid Licensing", the "Register License Server" link is highlighted with a green circle. The main content area shows "Current Releases" for NVIDIA GRID 4.0, with a link to "Click here to access your downloads" and a note to refer to OEM partner recommendations for [GRID version compatibility](#).

2. Click on **Register License Server**.

The screenshot shows the "Register License Server" page. The browser address bar displays <https://nvidia.flexnetoperations.com/control/nvda/createServer.ifs>. The page title is "Register License Server". The left sidebar shows "Register License Server" as the active page. The main content area contains a form with the following fields: "MAC address*", "Alias", and "Site Name". Below the form is a "Create" button. A note above the form states: "Note: Please do not use colons or spaces in the MAC Address."

- Enter the MAC address of the network card associated with the static IP address on your license server, **without any dashes or other non-number characters**. Adding alias and site information can be beneficial for management but is not strictly required. Click **Create**.



The MAC address of the nic associated with the VM can be found in the VM settings on vCenter

View Server

MAC address 0050569F1328
 ID Type ETHERNET
 Alias
 Site Name pe.lab

[Map Add-Ons](#) [View History](#) [View Served Clients](#) [Download License File](#)

Add-Ons

Add-On Name	Status	Entitlement
No add-ons are currently mapped.		

- In the **View Server** windows, click **Map Add-Ons**.

Map Add-Ons

Search Add-Ons for Server 0050569c3c2

Activation Code

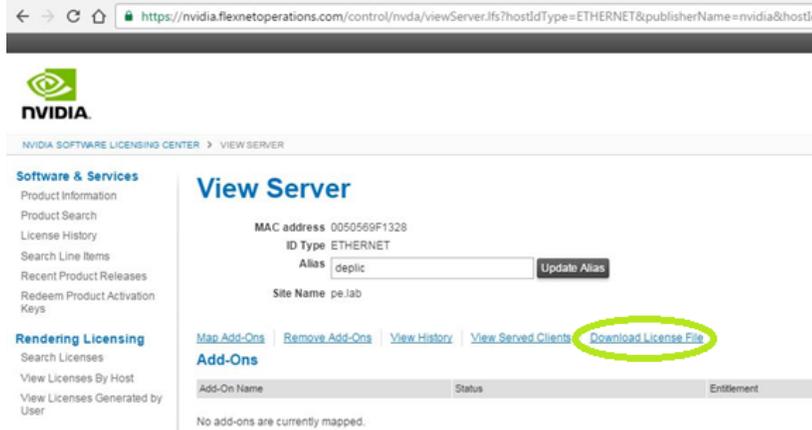
Entitlement ID

Add-On Name

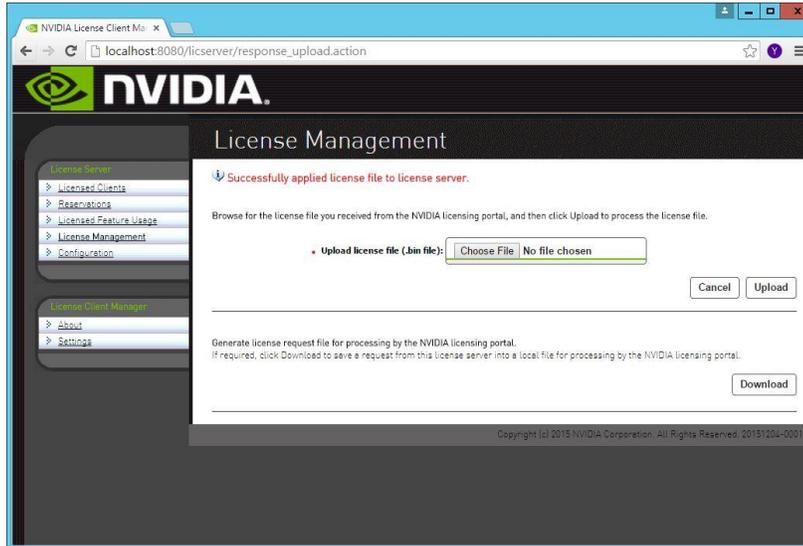
Feature Name

Add-On Name	Activation Code	Entitlement	Expiration	Available Units in Line Item	Total Units in Line Item	Qty to Add
GRID Evaluation Edition	GRIDEvalPK117322_XJ6R558JUX 6TQ2BY	GRIDEvalPK117322_XJ6R558JUX UJQ2BY (20000000)	May 3, 2017	128	128	<input type="text" value="128"/>

- In the **Quantity to Add** box, enter the number of licenses to be assigned to this server. Other licenses can be assigned to other NVIDIA License Servers in future, if desired. For the purposes of this procedure, enter the entire available amount.



- On the **View Server** window, click on **Download License File**. Save the file to a location accessible to the License Server.



- Once your license file has been obtained from NVIDIA, Login to your License Server, navigate to <http://localhost:8080/licserver>. Click on **License Management** and upload your license file.

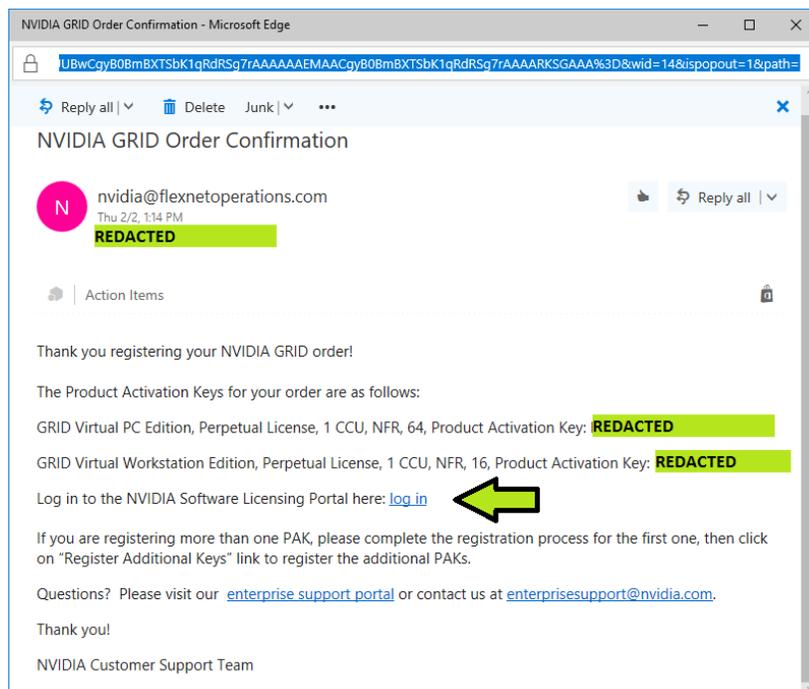
At this point, new clients will be able to obtain licenses automatically from the License Server.

5.4 EXCHANGING YOUR EVALUATION LICENSE FOR A PERMANENT ONE

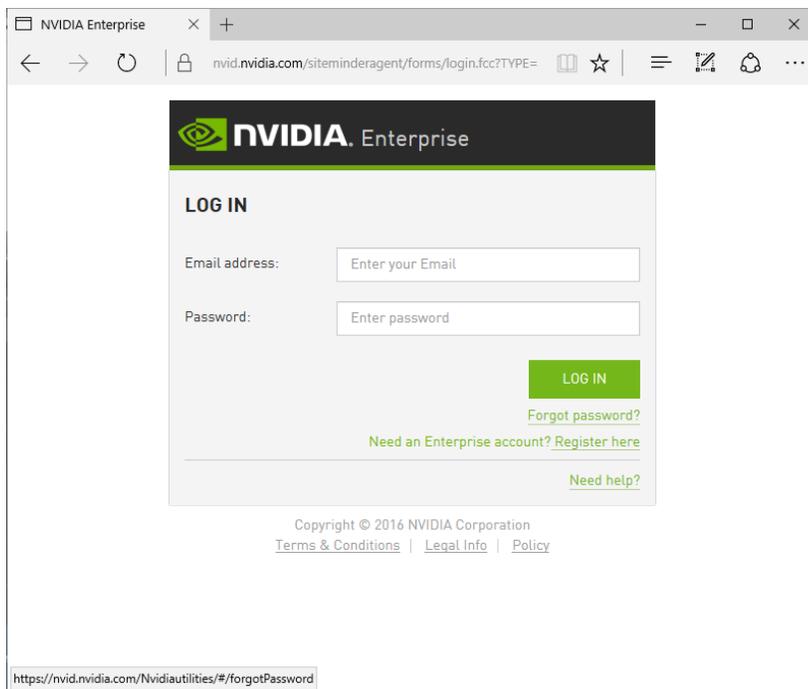
Evaluation licenses will eventually expire, and need to be replaced with permanent licenses. Follow this procedure to replace your evaluation licenses with permanent ones.

1. Purchase NVIDIA GRID licenses

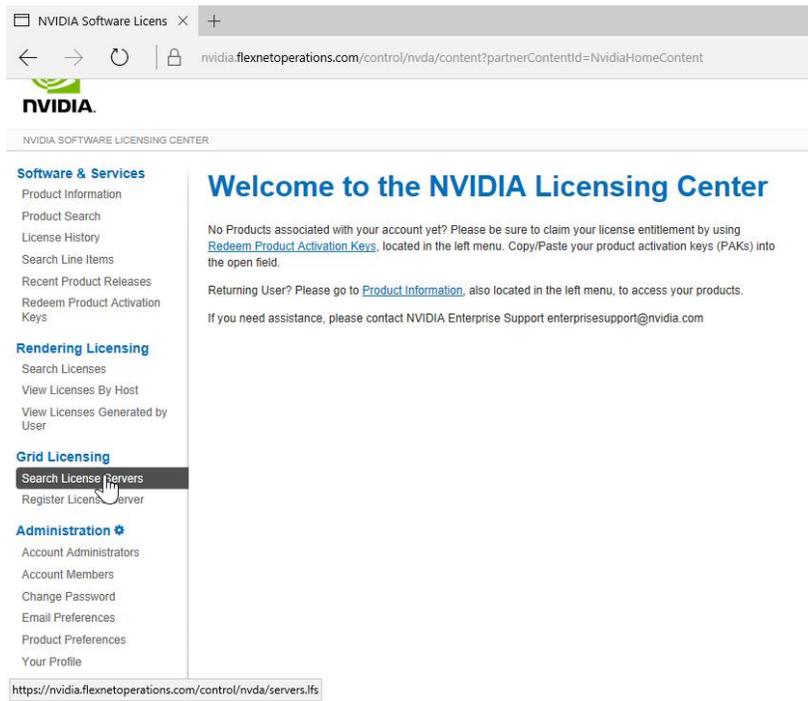
These can be obtained in many ways, see your local reseller. For help finding a reseller, open <http://www.nvidia.com/buygrid> in a web browser.



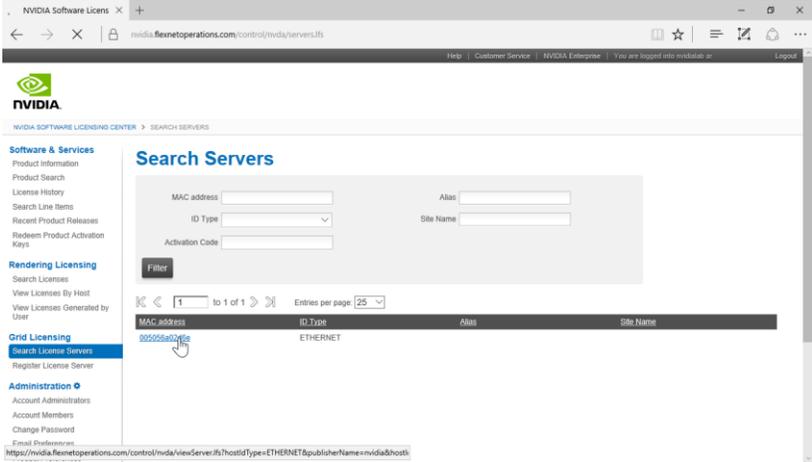
2. Open the email that came from nvidia@flexnetoperations.com. Click on the log in hyperlink.



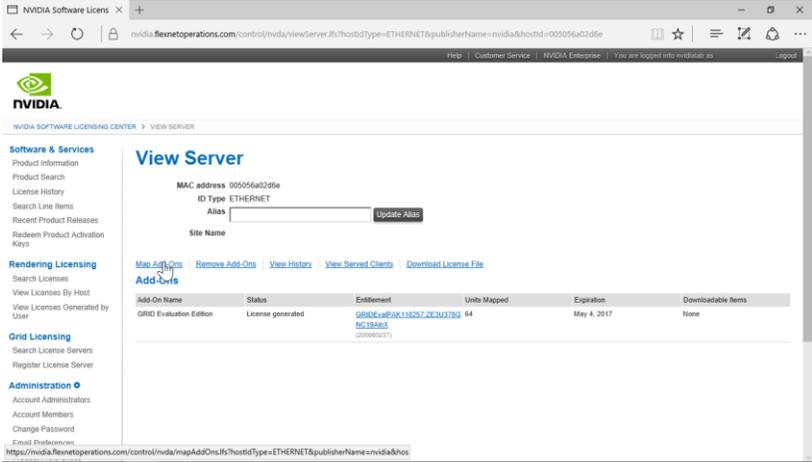
3. Log in with your email address and password.



4. On the **Welcome to the NVIDIA Licensing Center** screen, click on **Search License Servers** in the left-hand navigation area.



5. In the **Search Servers** screen, click on the hyperlink for your license server.



6. In the **View Servers** screen, click on **Map Add-Ons**.

Map Add-Ons

Search Add-Ons for Server: 005056a02d6e

Activation Code: Entitlement ID:

Add-On Name: Feature Name:

Add-On Name	Activation Code	Entitlement	Expiration	Available Units in Line Item	Total Units in Line Item	Qty to Add
GRID Evaluation Edition	GRDE-94P1X118257-2E3U 3780NC19AAX	GRDE-94P1X118257-2E3U 3780NC19AAX	May 4, 2017	64	128	<input type="text"/>
GRID Virtual PC Edition, Perpetual License, 1 CCU, NFR	4185-C88B-3E64-BAEF	NFR1235-2a513N (200891027)	Permanent	64	64	<input type="text" value="20"/>
GRID Virtual Workstation Edition, Perpetual License, 1 CCU, NFR	ASFG-EB63-FBDA-0029	NFR1235-2a513N (200891027)	Permanent	16	16	<input type="text" value="16"/>

- In the **Map Add-Ons** screen, find your permanent licenses. In the **Qty to Add** box for each license type, enter the number of licenses to add to this license server. Click **Map Add-Ons**.

View Server

The add-ons were successfully mapped.

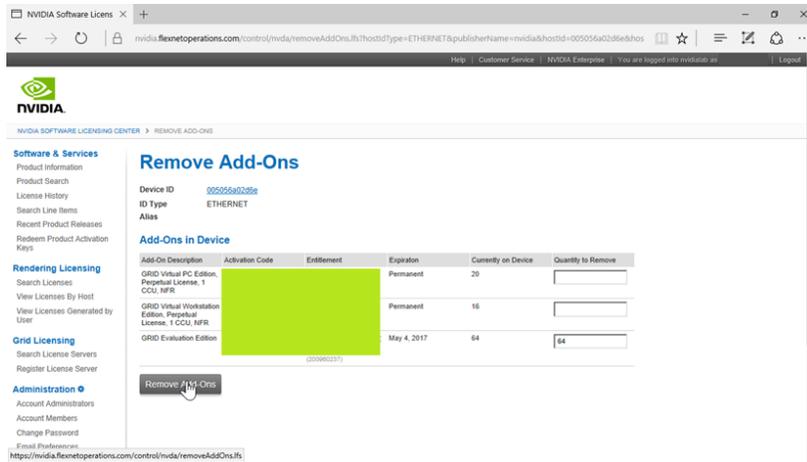
MAC address: 005056a02d6e
ID Type: ETHERNET
Alias:
Site Name:

[Map Add-Ons](#) | [Remove Add-Ons](#) | [View History](#) | [View Served Clients](#) | [Download License File](#)

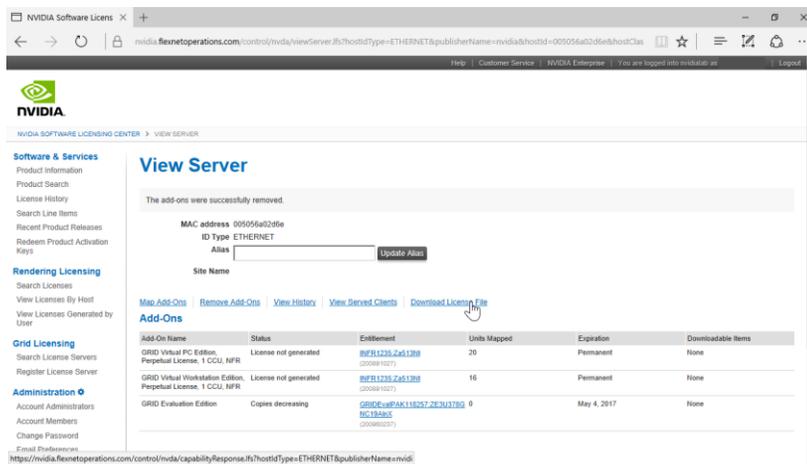
Add-Ons

Add-On Name	Status	Entitlement	Units Mapped
GRID Virtual PC Edition, Perpetual License, 1 CCU, NFR	License not generated	NFR1235-2a513N (200891027)	20
GRID Virtual Workstation Edition, Perpetual License, 1 CCU, NFR	License not generated	NFR1235-2a513N (200891027)	16

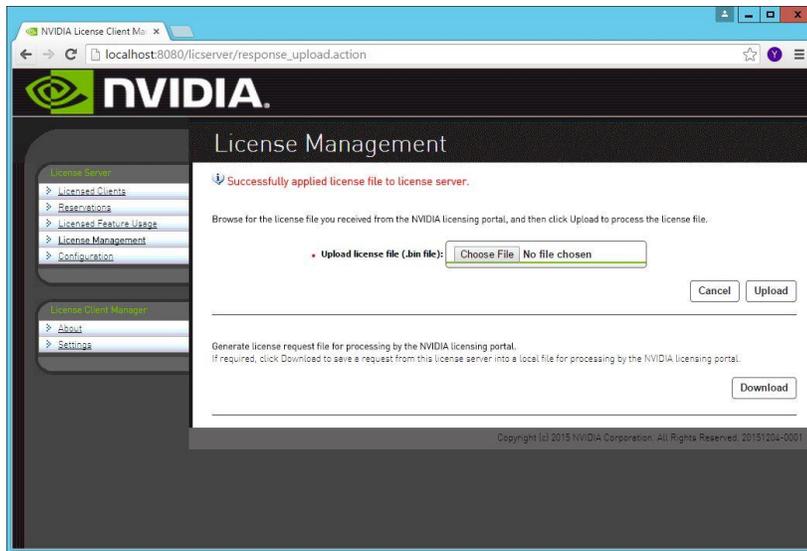
- In the **View Server** window, click **Remove Add-Ons**.



9. In the **Remove Add-Ons** window, find your evaluation licenses and, in the **Quantity to Remove** box, enter all the licenses currently on the device. You will find this number in the **Currently on Device** column on the same page. Click **Remove Add-Ons**.



10. In the **View Server** window, observe that the **Units Mapped** of evaluation licenses is now 0. Click on **Download License File** and save the file in a place accessible by your license server.



11. Once your license file has been obtained from NVIDIA, Login to your License Server, navigate to <http://localhost:8080/licserver>. Click on **License Management** and upload your license file.

At this point, any GRID clients that reboot will obtain new licenses from the permanent group.

5.5 SUPPORT, UPDATE, AND MAINTENANCE (SUMS)

NVIDIA Support, Update, and Maintenance Subscription (SUMS) for NVIDIA GRID gives you comprehensive software patches, updates, and upgrades, plus technical support from the experts. You now have an easy, reliable way to improve productivity and reduce downtime for your production systems.

NVIDIA GRID SUMS delivers a comprehensive solution for NVIDIA GRID software versions 2.0 and beyond on your NVIDIA Tesla Certified Systems, including:

- ▶ Access to the latest updates and upgrades
- ▶ Maintenance for NVIDIA GRID software
- ▶ Direct communication with NVIDIA technical experts
- ▶ Searchable knowledge base of specific documentation, application notes and articles
- ▶ Rapid response & timely issue resolution
- ▶ Active issue prioritization

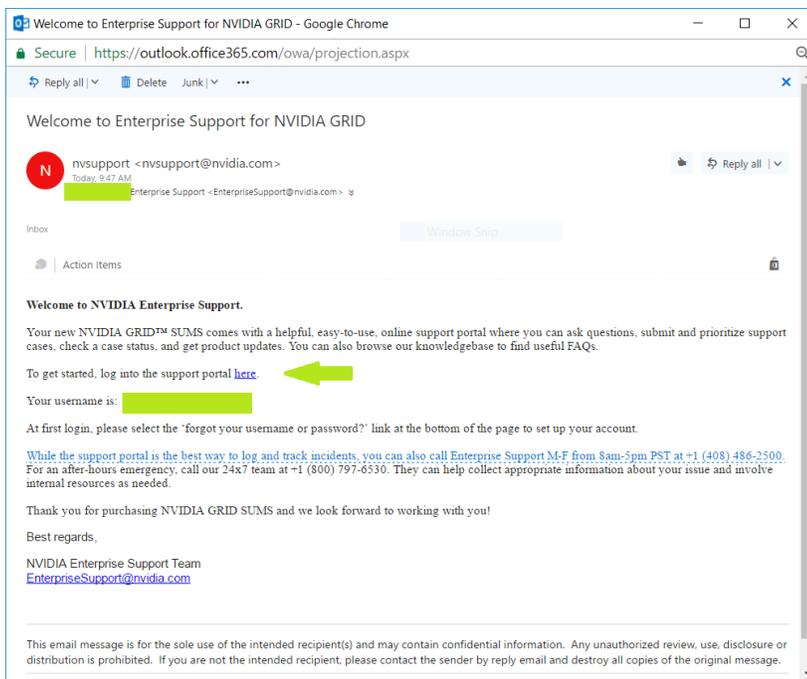
SUMS comes bundled as part of an annual subscription, and can be purchased as a yearly subscription alongside a perpetual license.

	Production
Maintenance releases	X
Updates	X
Upgrades	X
Maintenance, Updates, and Upgrades Access Period	3 Years from General Availability
Response Time	4 Hours
Online Access	NVIDIA GRID E-Mail Support Alias and NVIDIA GRID Support Portal
Support Hours	8am-5pm PST, Monday-Friday
Knowledgebase	

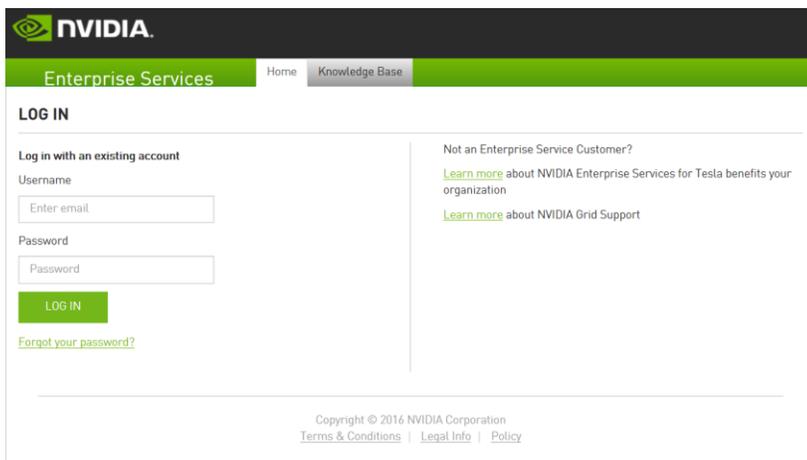
Support currently available only in English.

For additional service details, please review the [NVIDIA GRID End User License Agreement](#) and the [NVIDIA GRID Software Service Supplemental Terms](#).

Use the following procedure to obtain support for your GRID products.



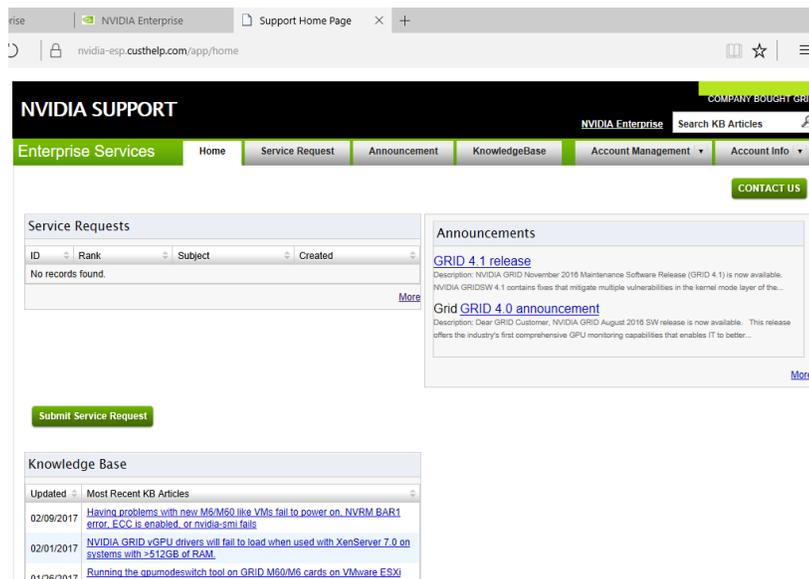
1. Once your SUMS entitlement has been assigned, you will receive and email from NVIDIA Enterprise Support. Click the **here** hyperlink.



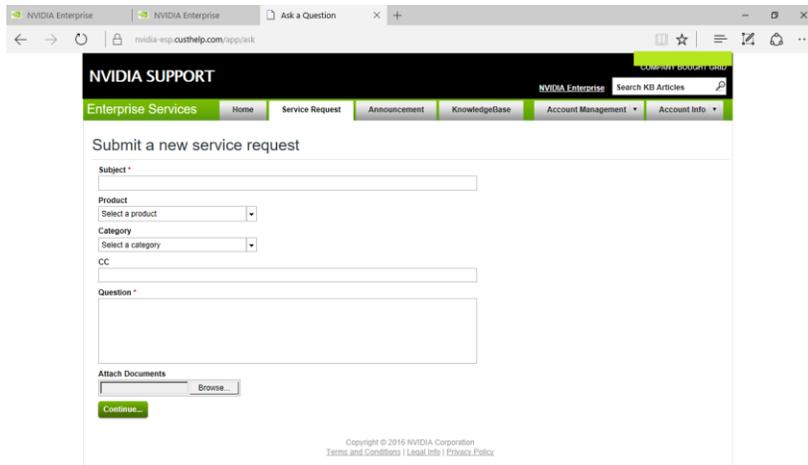
2. Log in with your **Username** (your email address) and **Password**.



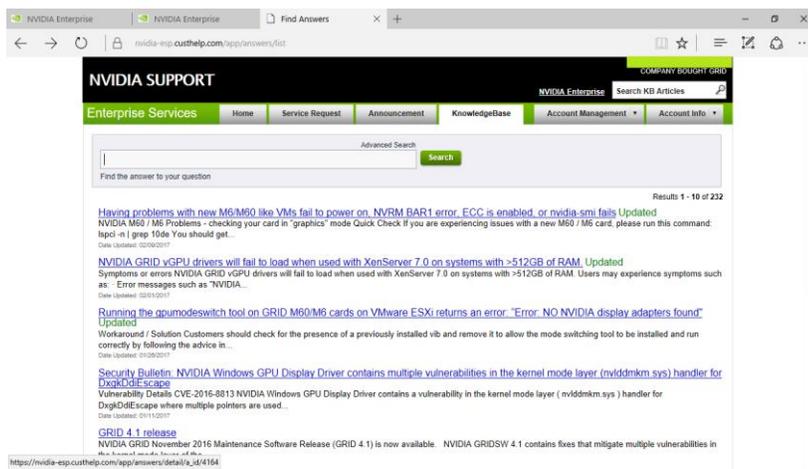
3. On the **NVIDIA Enterprise Customer Service** page, click the **NVIDIA ENTERPRISE SUPPORT** icon.



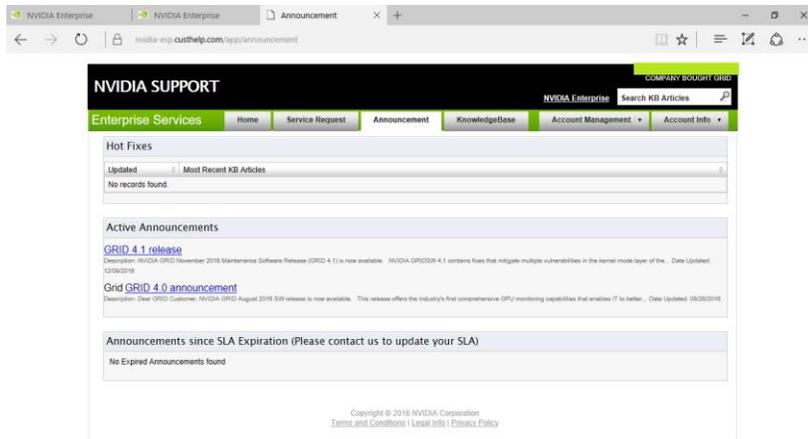
4. You will arrive at the **Home** tab. Here you will see a list of your service requests, Announcements from NVIDIA pertaining to your product, links to the Knowledge Base, and a button for submitting a service request.



5. When you wish to submit a service request, click the **Service Request** button.



6. The Knowledge Base has up-to-date articles and tips for maximizing your NVIDIA GRID experience.



7. The Announcements page shows the latest versions of GRID software, and any late breaking updates or patches.

Be sure to check back at the SUMS portal frequently. This will help you stay up to date on your subscription, find patches, and keep your GRID software running smoothly.

CHAPTER 6. NVIDIA GRID vGPU MANAGER INSTALLATION

This chapter discusses the following NVIDIA GRID vGPU Manager installation topics:

- ▶ Installing the VIB.
- ▶ Uninstalling the VIB.
- ▶ Upgrading to Permanent GRID licenses.



Upgrading from previous versions of the VIB is not supported.

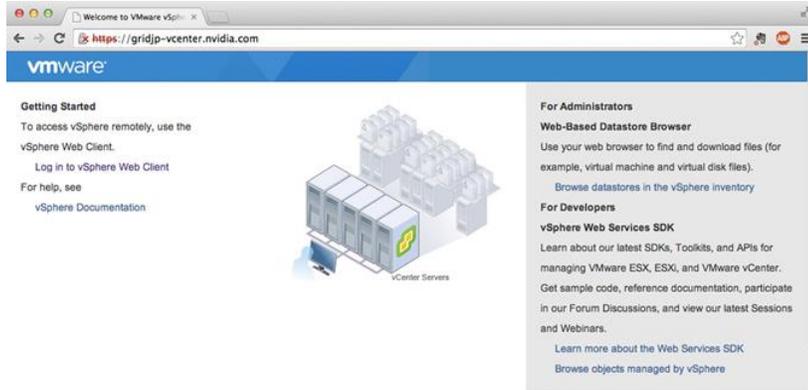
6.1 PREPARE AND INSTALL THE VIB FILES

This section guides you through installing, confirming, and uninstalling the VIB files.

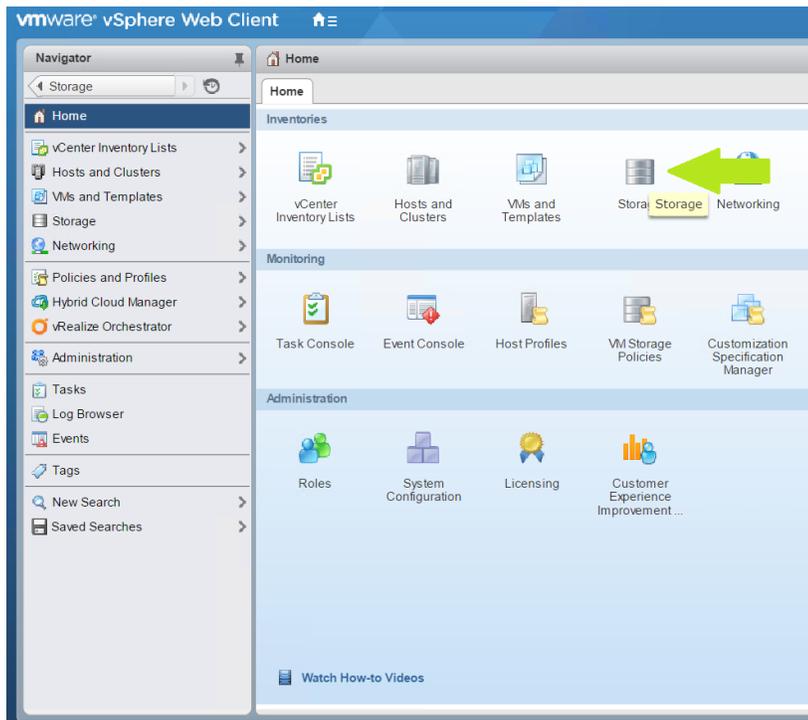
6.1.1 Uploading VIB in vSphere

This section describes how to upload a VIB file to the host using the vSphere Web Client. See Chapter 15, *Using WINSCP*, on page 295 if you are using WinSCP.

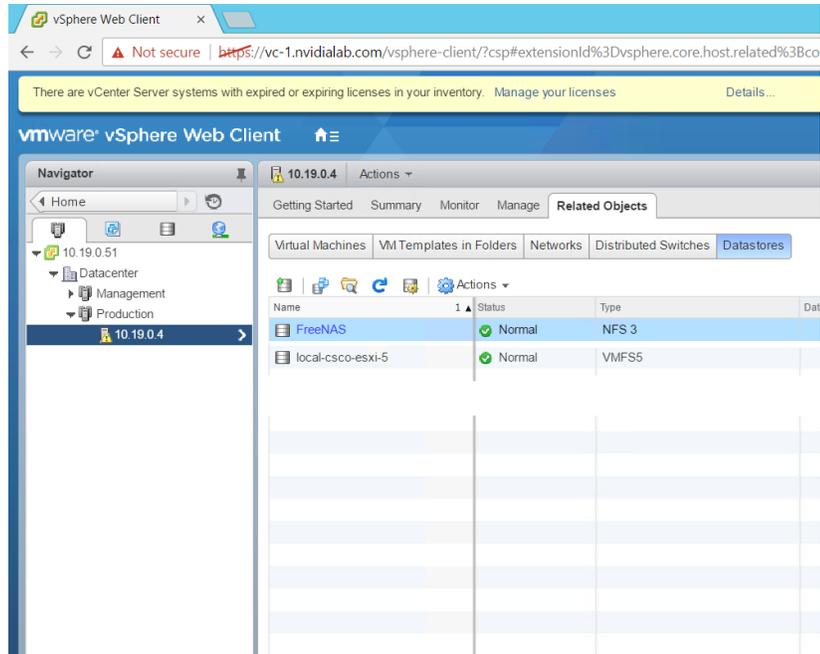
To upload the file to the data store using vSphere Web Client, do the following:



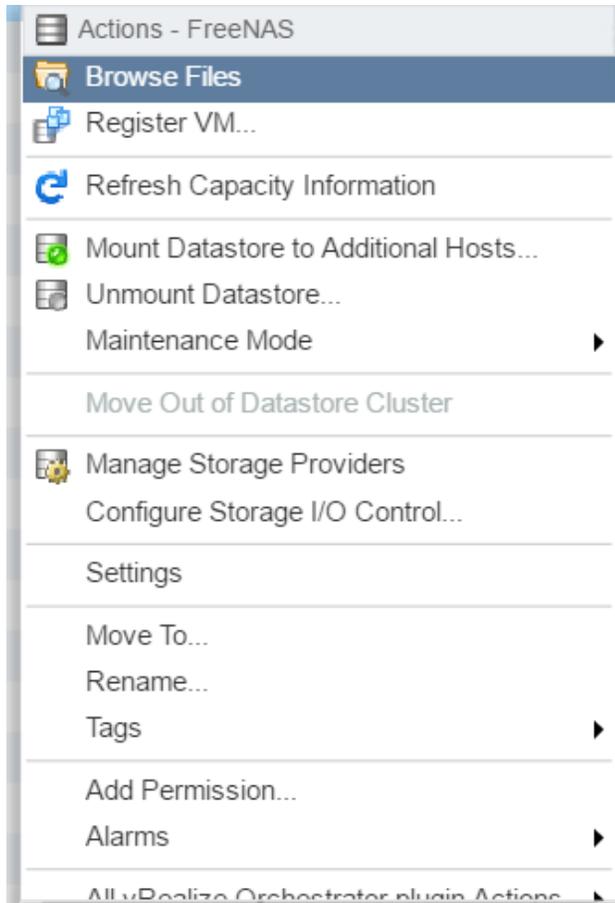
1. Open a web browser and navigate to the vCenter host at `https://<host>`, where `<host>` is either the hostname or IP address. Ignore the untrusted certificate warning if it displays. Select **Log in to vSphere Web client**.



2. On the Home screen, click **Hosts and Clusters**.



3. Click on the host in the left-hand Navigator pane, then click **Related Objects** → **Datastores**.



4. Right click on the datastore that will accept the upload, and click **Browse Files**.

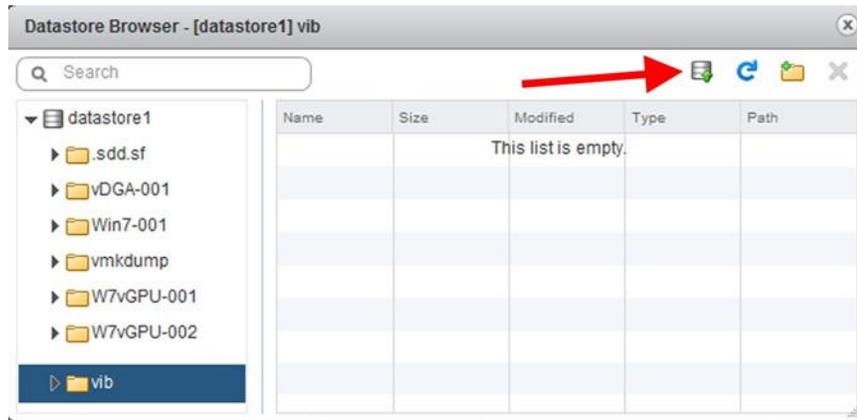


5. Click the **New Folder** icon.

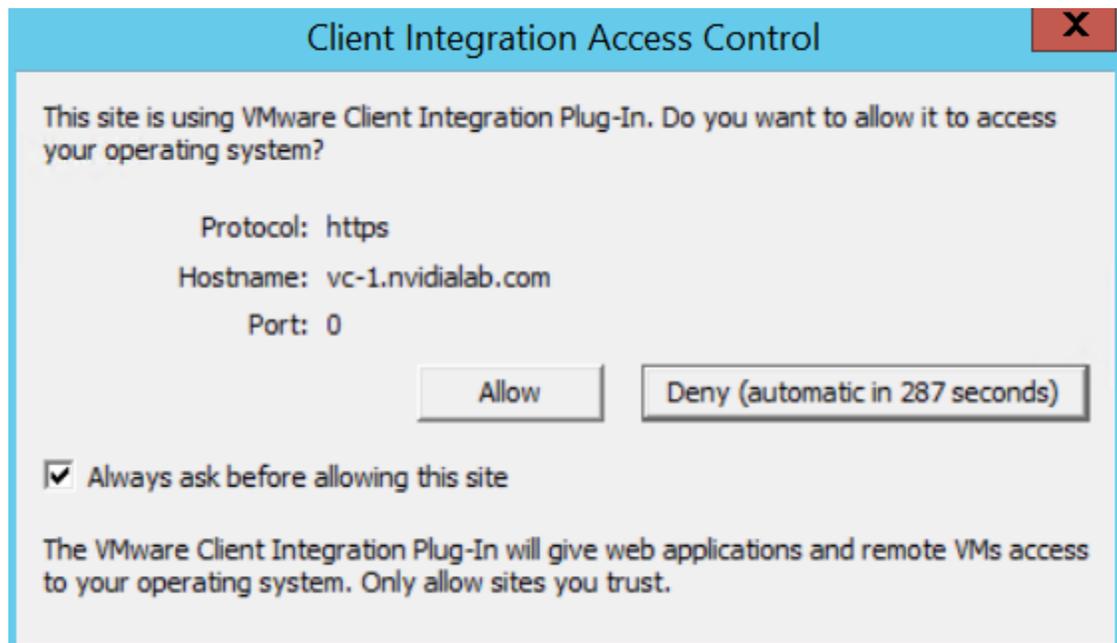


6. Name the new folder **vib** and then click **Create**.

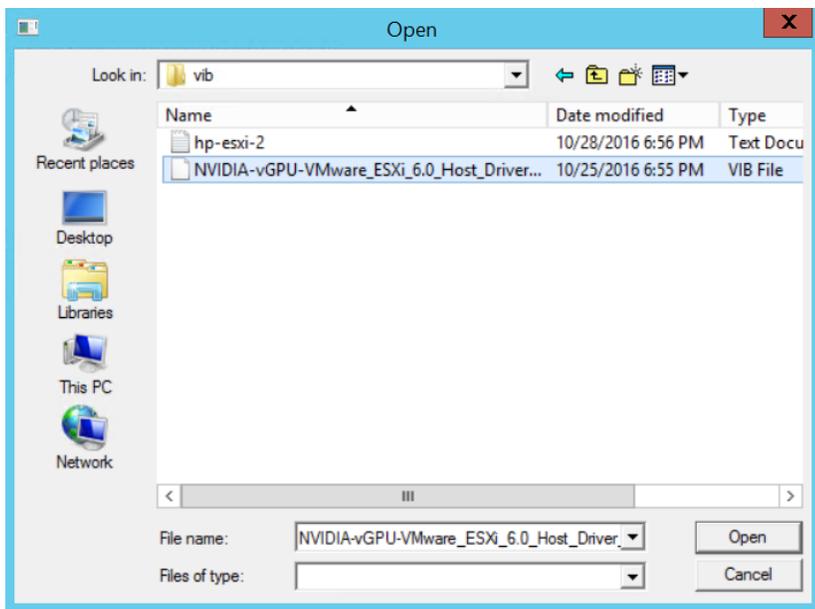
7. Select the **vib** folder in the **Datastore Browser** window.



8. Click the **Upload** icon.



9. Select **Allow**.

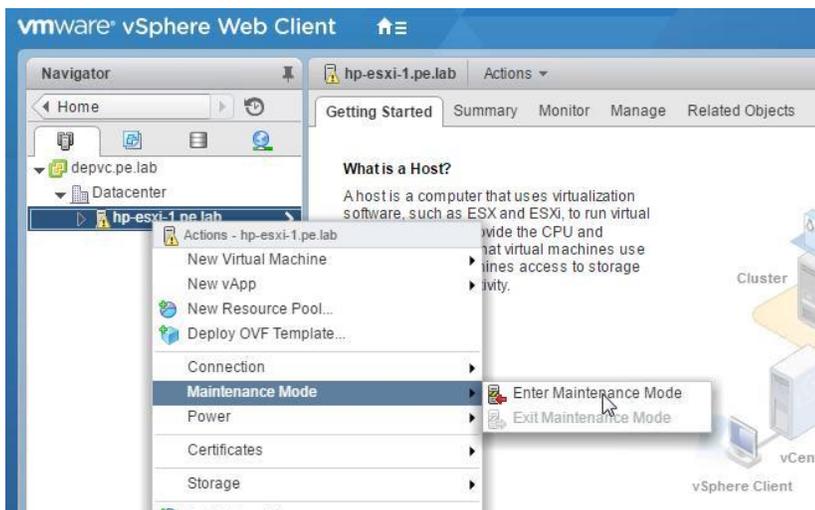


10. Select the VIB file to be uploaded. Click **Open**.

The VIB file is uploaded to the data store on the host.

 If you do not click Allow before the timer runs out, then further attempts to upload a file will silently fail. If this happens, exit and restart vSphere Web Client, repeat this procedure and be sure to click Allow before the timer runs out.

6.1.2 Place the Host in Maintenance Mode



1. Place the host into Maintenance mode by right-clicking it and then selecting **Maintenance Mode** → **Enter Maintenance Mode**.

Alternatively, you can place the host into Maintenance mode using the command prompt by entering

```
$ esxcli system maintenanceMode set --enable=true
```

This command will not return a response. Making this change using the command prompt will not refresh the vSphere Web Client UI. Click the Refresh icon in the upper right corner of the vSphere Web Client window.



CAUTION: PLACING THE HOST INTO MAINTENANCE MODE DISABLES ANY VCENTER APPLIANCE RUNNING ON THIS HOST UNTIL YOU EXIT MAINTENANCE MODE AND THEN RESTART THAT VCENTER APPLIANCE.

2. Click **OK** to confirm your selection.

6.1.3 Installing the VIB

Use the following procedure to install the VIB file, **NVIDIA-vGPU-VMware_ESXi_6.0_Host_Driver_367.43-1OEM.600.0.0.2494585.vib**.

1. Issue the following command to install the vGPU VIB, where **<path>** is the full path to the VIB file:

```
$ esxcli software vib install -v /<path>/NVIDIA-vGPU-VMware_ESXi_6.0_Host_Driver_367.43-1OEM.600.0.0.2494585.vib
```

The following error displays if you *do not* use the full path:
[VibDownloadError].

When the installation is successful, the following displays:

```
Installation Result:
Message: Operation finished successfully. Reboot Required: false
VIBs Installed: NVIDIA-vGPU-VMware_ESXi_6.0_Host_Driver_367.43-1OEM.600.0.0.2494585
```

```
VIBs Removed: VIBs Skipped:
```



Although the display states “Reboot Required: false”, a reboot is necessary for the vib to load and xorg to start.

6.1.4 Confirm the Driver VIB Installation

1. Use the following command to determine the installation of vGPU driver in vSphere:

```
$ esxcli software vib list | grep -i nvidia
```

This command returns output similar to the following:

```
NVIDIA-vGPU-VMware_ESXi_6.0_Host_Driver_367.43-1OEM.600.0.0.2494585
NVIDIA VMwareAccepted 2016-10-26
```

2. Confirm GPU detection using the following command:

```
$ nvidia-smi
```

This command returns output similar to the following (output shown below is for a Tesla M60 card in a single host):

```
+-----+
| nvidia-smi 367.43                               Driver Version: 367.43 |
+-----+-----+-----+-----+-----+
| GPU  Name          Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf   Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+-----+-----+-----+
|   0   Tesla M60           On      | 0000:05:00.0   Off  |    0%      Default  |
| N/A   44C    P8     24W / 150W | 19MiB / 8191MiB |          |
+-----+-----+-----+-----+-----+
|   1   Tesla M60           On      | 0000:06:00.0   Off  |    0%      Default  |
| N/A   41C    P8     23W / 150W | 19MiB / 8191MiB |          |
+-----+-----+-----+-----+-----+

+-----+
| Processes:                                     GPU Memory |
|  GPU           PID  Type  Process name                               Usage      |
+-----+-----+-----+-----+-----+
| No running processes found                    |
+-----+
```

If the **nvidia-smi** command does not return any output, see Chapter 14, vGPU-Enabled VMs and nvidia-smi Fail to Start, on page 280)



SMI also allows GPU monitoring using the following command:

```
$ nvidia-smi -l
```

This command switch adds a loop, auto refreshing the display.

6.2 UNINSTALLING THE VIB

This section describes how to uninstall the VIB



Upgrading is not supported. The previous vib must be removed completely before installing the new vib.

Use the following procedure to uninstall the VIB:

1. Determine the name of the vGPU driver bundle as described in section 6.1.4 Confirm the Driver VIB Installation on page 134
2. Run the following command to uninstall the driver package:

```
$ esxcli software vib remove -n NVIDIA-vGPU-  
VMware_ESXi_6.0_Host_Driver --maintenance-mode
```

The following message displays when uninstallation is successful:

```
Removal Result  
  Message: Operation finished successfully.  
  Reboot Required: false  
  VIBs Installed:  
  VIBs Removed: NVIDIA_bootbank_NVIDIA-vGPU-  
VMware_ESXi_6.0_Host_Driver_367.43-1OEM.600.0.0.2494585  
  VIBs Skipped:
```

3. Reboot the host to complete the uninstallation process.

CHAPTER 7.

BUILDING CITRIX XENDESKTOP 7.12

This chapter covers installing Citrix XenDesktop 7.12, including:

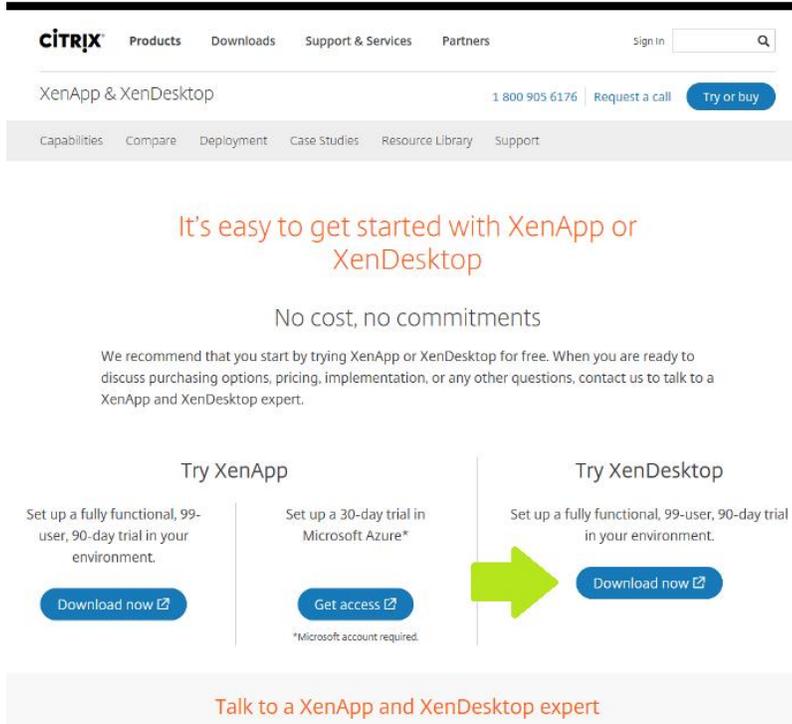
- ▶ Registering and allocating the trial license
- ▶ Installing XenDesktop
- ▶ Importing vSphere and vSphere SDK Certificates
- ▶ Using Citrix Studio to performing site setup
- ▶ Configuring Citrix license server
- ▶ Verify Citrix StoreFront Configuration

7.1 REGISTERING AND ALLOCATING THE TRIAL LICENSE

Use the following procedure to obtain and allocate a Citrix XenDesktop 7.12 license:



Testing was done on XenDesktop 7.12 though Citrix always has the latest version available for trials so please keep that in mind and note that some screenshots and steps may vary slightly.



1. Using an internet browser. Navigate to the <http://www.citrix.com/tryxendesktop> website. Under **Try XenDesktop** click **Download Now**.



Try XenDesktop for free

Securely deliver Windows, Linux, web, and SaaS apps plus full virtual desktops to any device, anywhere.

Increase mobile productivity

Increase employee productivity and business mobility in a single solution, without compromising security.

Simple unified platform

Reduce IT complexity by centralizing app and desktop management and automating common tasks.

Secure remote access without risk

Deliver remote access while addressing privacy, compliance, and risk management mandates.

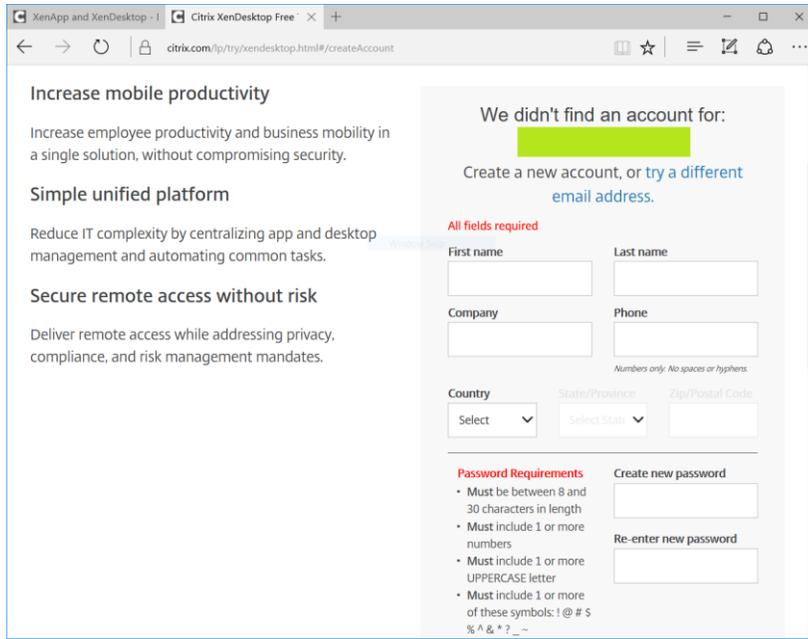


Enter your email address.
We will send you a trial license.

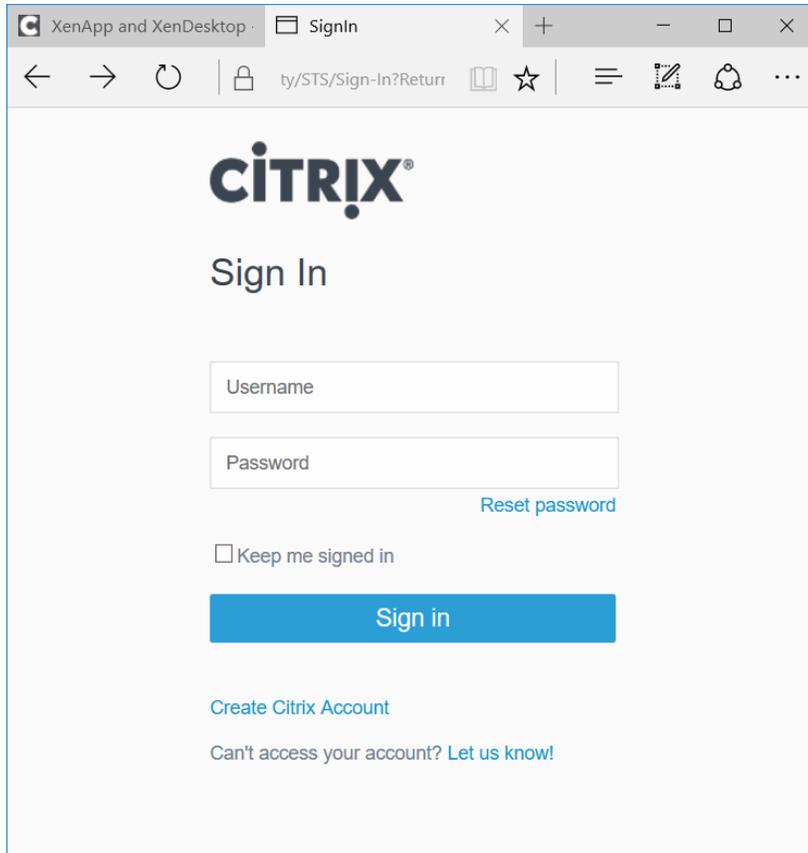
Email

Next

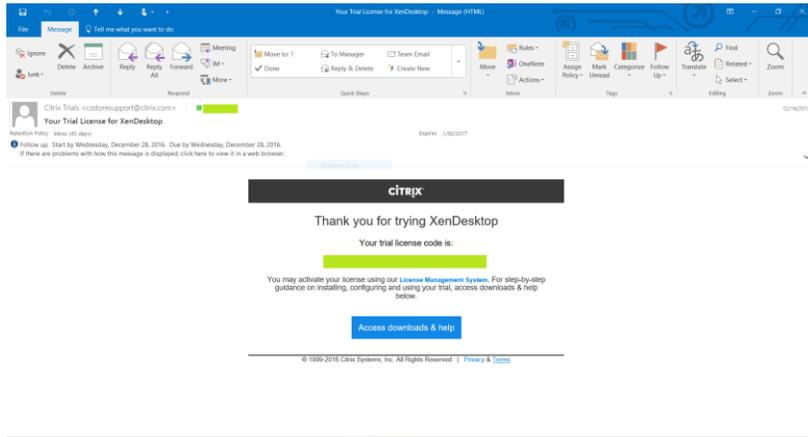
2. Enter your email address and click **Next**.



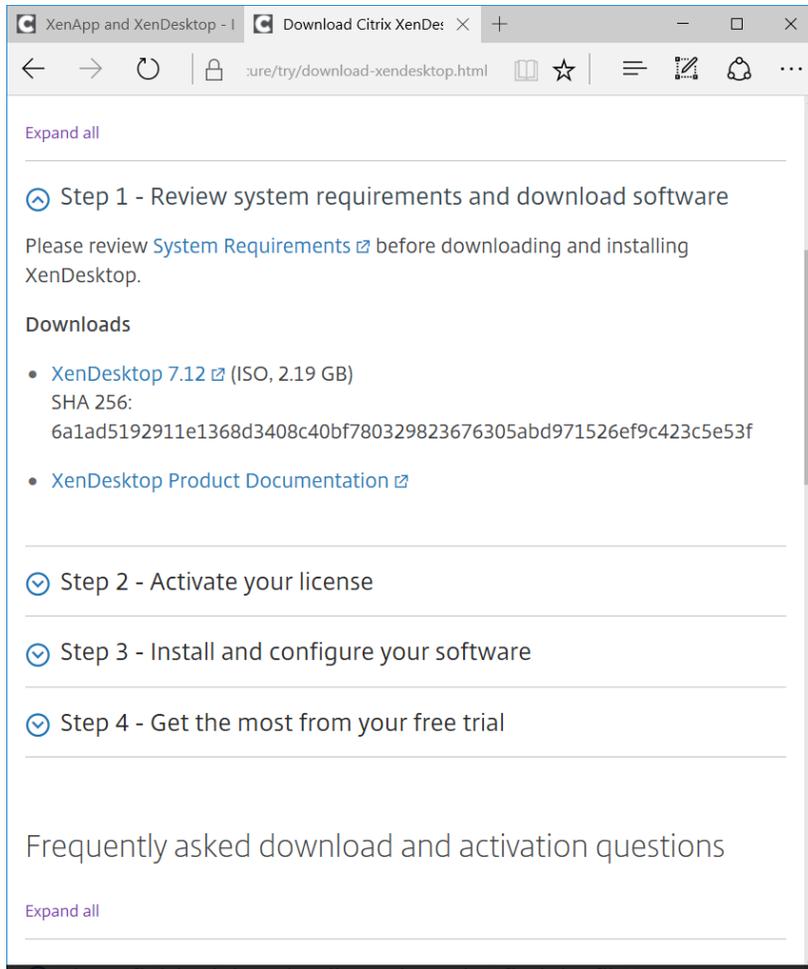
3. Create a new account by adding your account info. Click **Create account and send license**.



4. Once your account has been created sign in to your account.

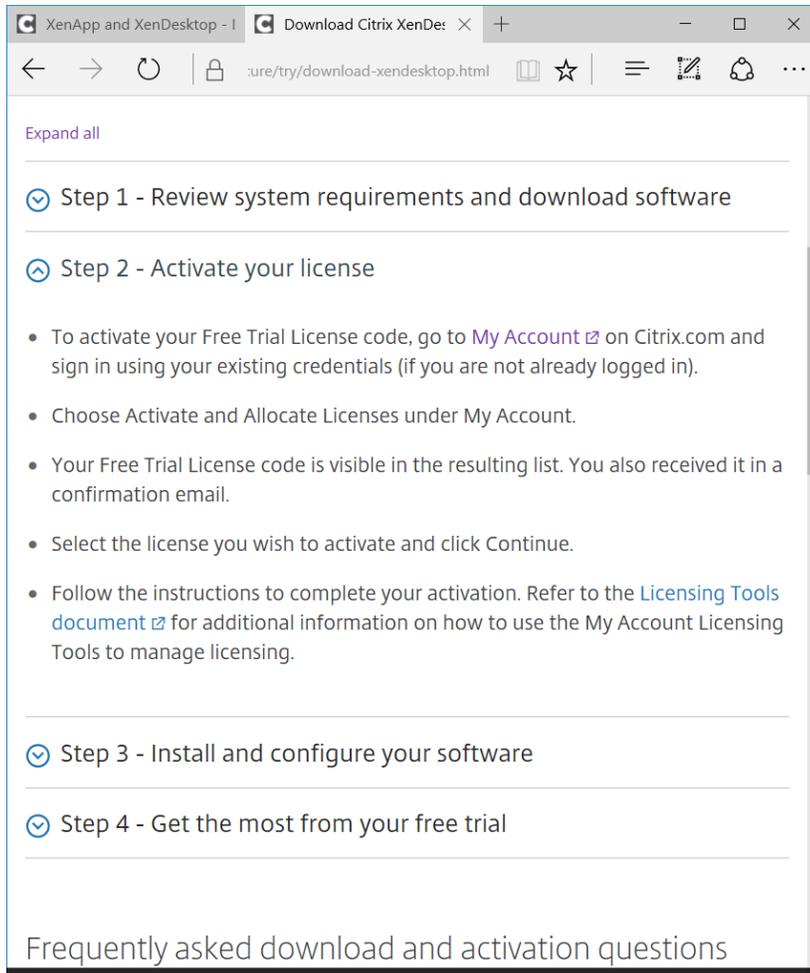


5. Open the email that you received from **citrix.cs@digitalriver.com** as it contains the **License Code** needed to allocate your trial license. This code should match the one displayed in your Citrix account.

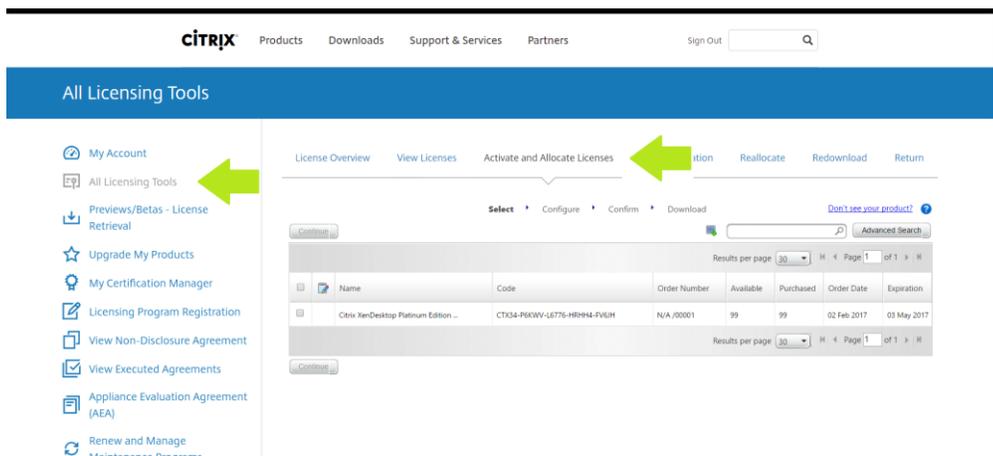


 You may choose to enter the license key later. Even without a license key, all features are enabled and the evaluation will run successfully for up to 30 days

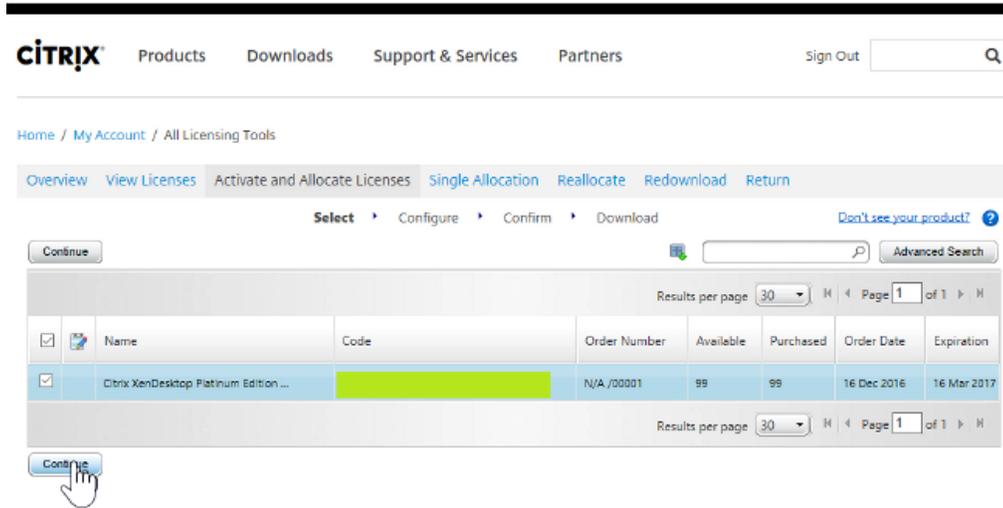
6. On the Citrix License page, Click on **Review system requirements and download software**. Click the **XenDesktop 7.12** hyperlink and download the software to a storage location accessible to the ESXi Management server.



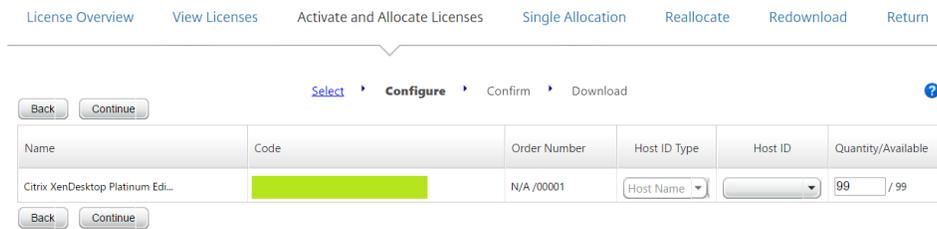
7. Once the software has downloaded, click on **Step 2 - Activate your license**. Click on **My Account** and sign in if necessary.



8. Click on the **All Licensing Tools** link. Then click **Activate and Allocate Licenses**.



9. Click the checkbox next to your license, then click **Continue**.

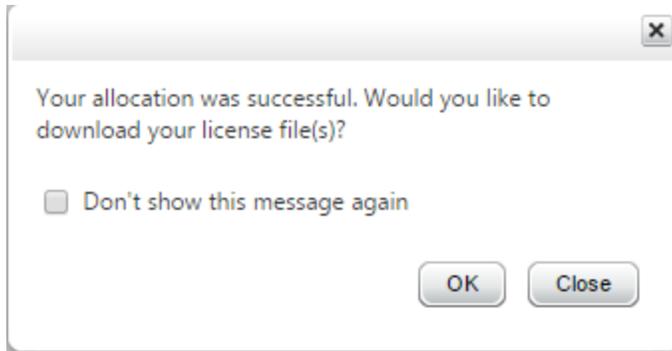


10. For **Host ID Type**, ensure **Host Name** is selected. Under **Host ID**, enter the NetBios name of the host on which you will install Citrix XenDesktop. Click **Continue**.

 The Host ID field is case sensitive. If your Host ID is not exactly the same as your NETBIOS Hostname it will not work properly. Do not enter the FQDN name of your server.



11. Confirm your selection and Host ID and click **Confirm**. Select the **Download** button to proceed to download your license file.



12. A confirmation dialog will be displayed, select **OK** to proceed to download your license file.
13. Once the download completes, copy the license file to the server where you will install XenDesktop 7.12.

7.2 INSTALLING XENDESKTOP 7.12

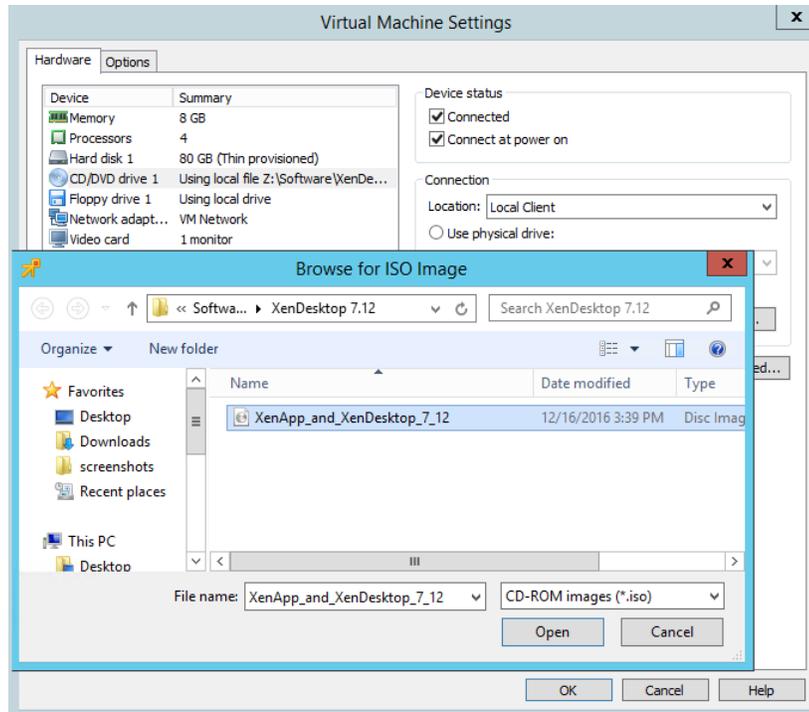
The server on which you are installing Citrix XenDesktop 7.12 must meet the requirements listed in *General Prerequisites* on page 33. As of the publication date of this manual the current version of the Citrix XenDesktop is:

Version: 7.12.0.81

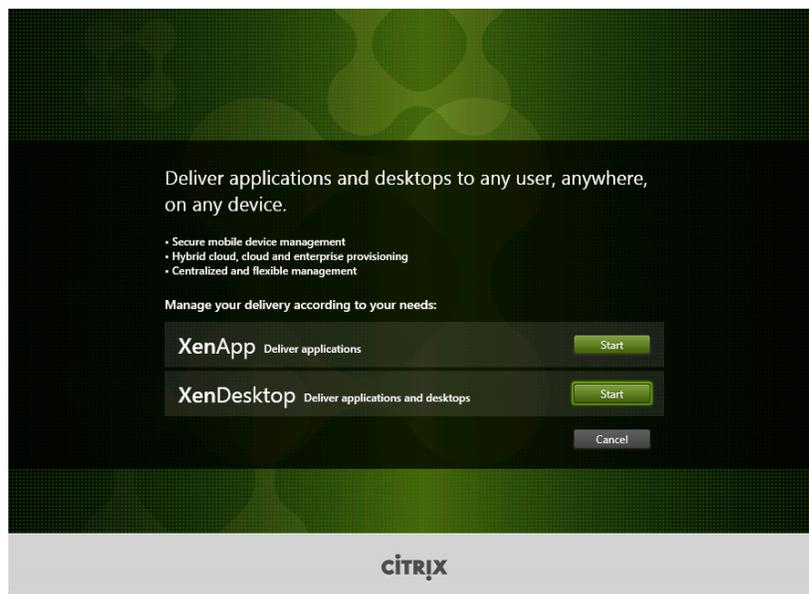
XenApp_and_XenDesktop7_12.iso

Use the following procedure to install XenDesktop 7.12:

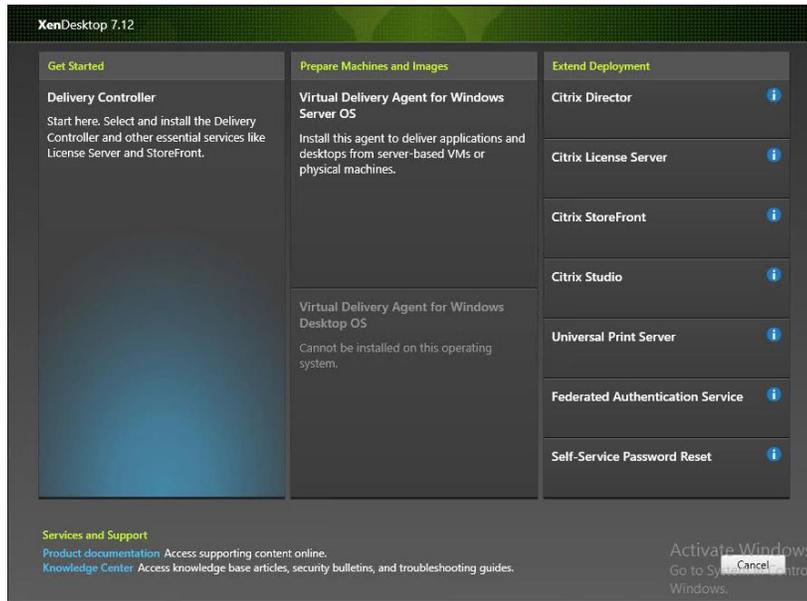
1. Log in to the vCenter Web Client.



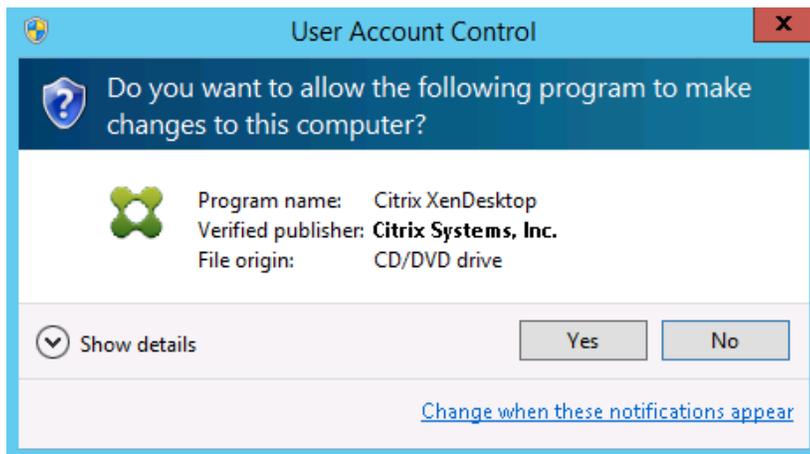
2. Use the vCenter Client to mount the .iso from an ISO data store to the virtual machine CD drives. Ensure the **Connected** and **Connected at power on** check boxes are selected.



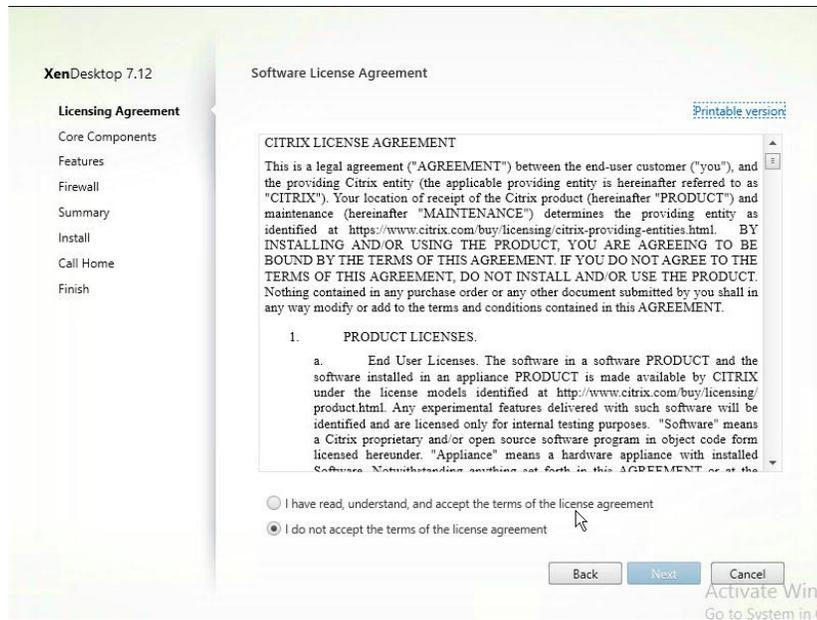
3. Launch the CD with AutoPlay or select the **AutoSelect.exe** application to start the Citrix XenApp and XenDesktop installer. Select the **Start** button next to **XenDesktop Deliver applications and desktops**.



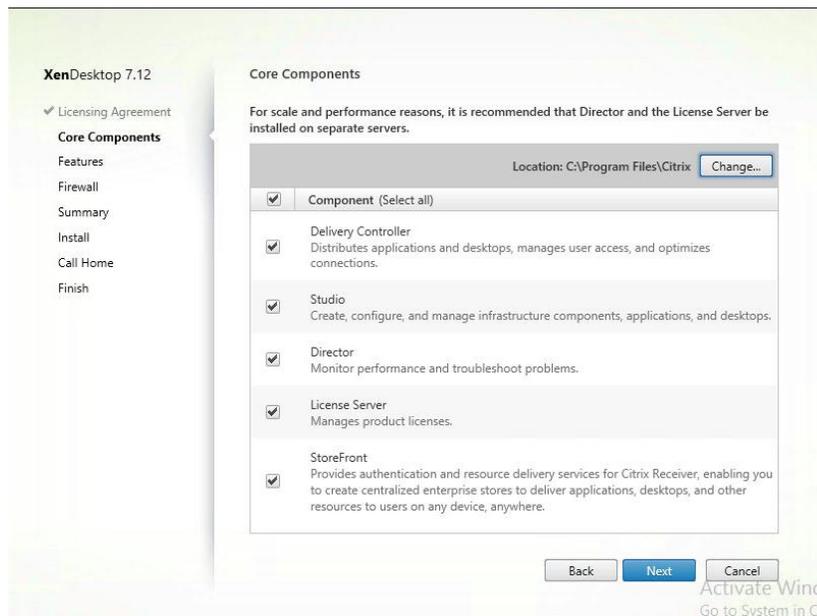
4. Select **Delivery Controller** under **Get Started** on the **XenDesktop 7.12** window.



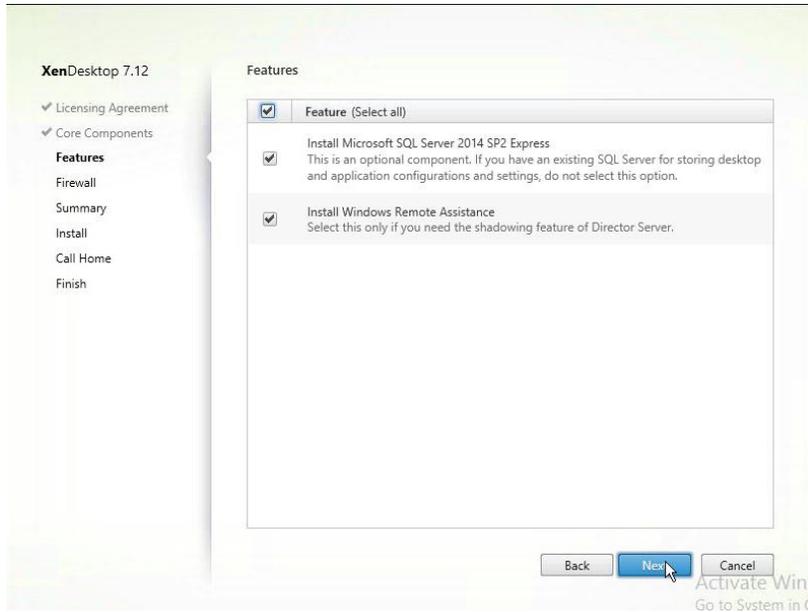
5. Select **Yes** to agree to the User Account Control prompt if displayed.



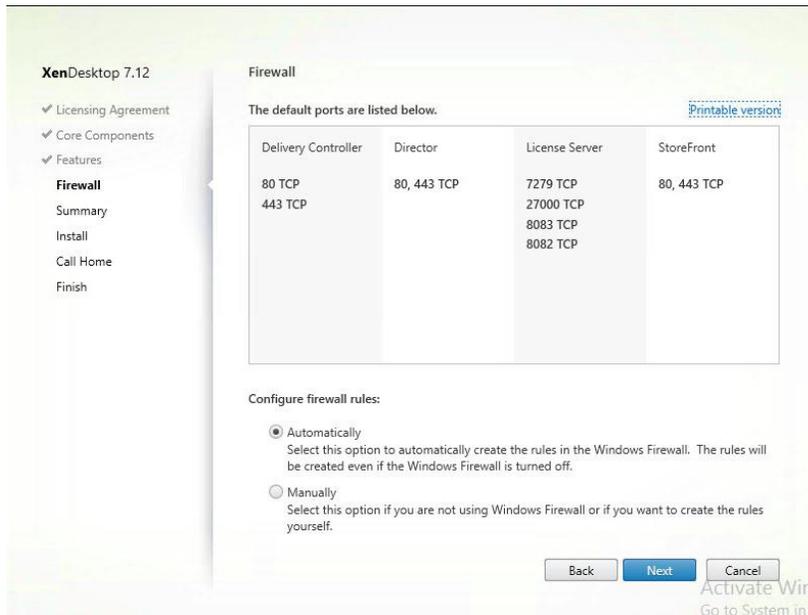
6. Check **I have read, understand, and accept the terms of the license agreement** radio button to accept the agreement. Select **Next** to continue.



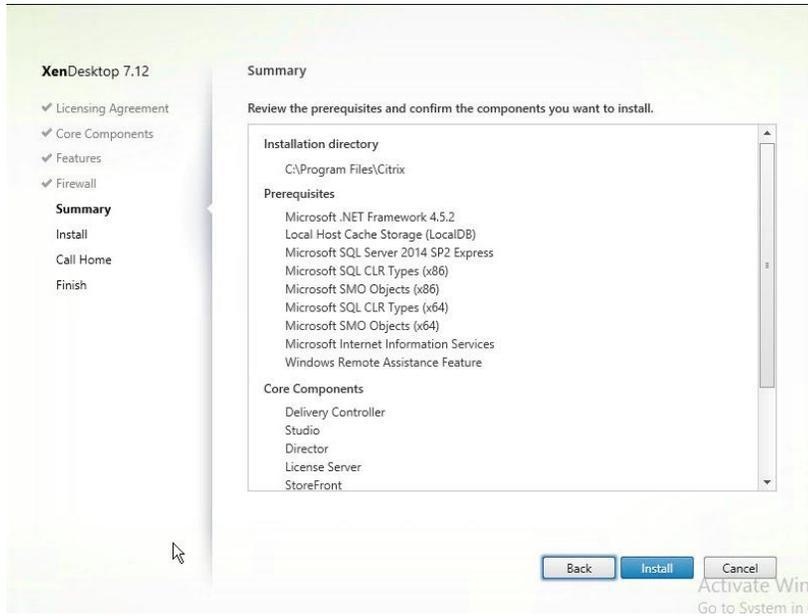
7. Select the Core Components that you wish to install, by default all components are selected. For a single server installation select **Next** with all Core Components checked to continue.



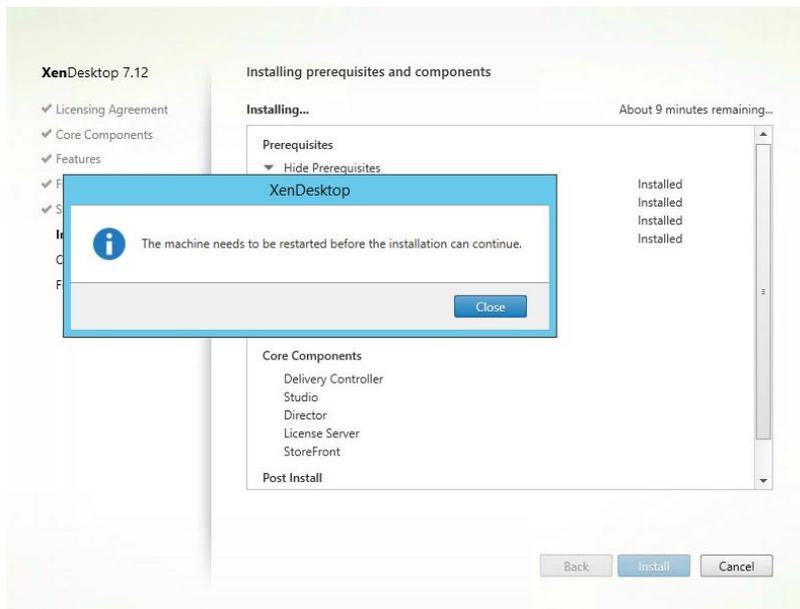
8. Select **Features** needed for this installation. Citrix best practice is to use a SQL Server separate from the XenDesktop server. For a single server installation accept the default selections and select **Next** to continue.



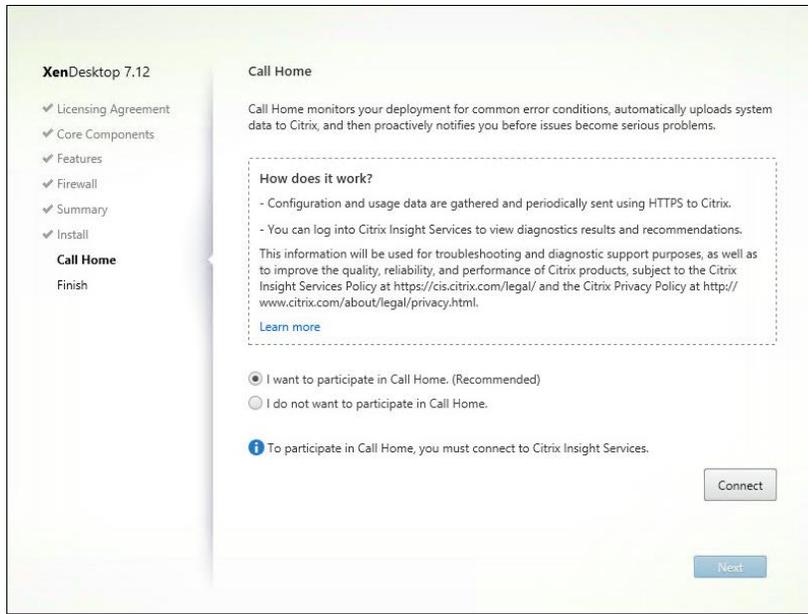
9. Under **Configure firewall rules:** select **Automatically**, unless you are not using Windows Firewall. Select **Next** to continue.



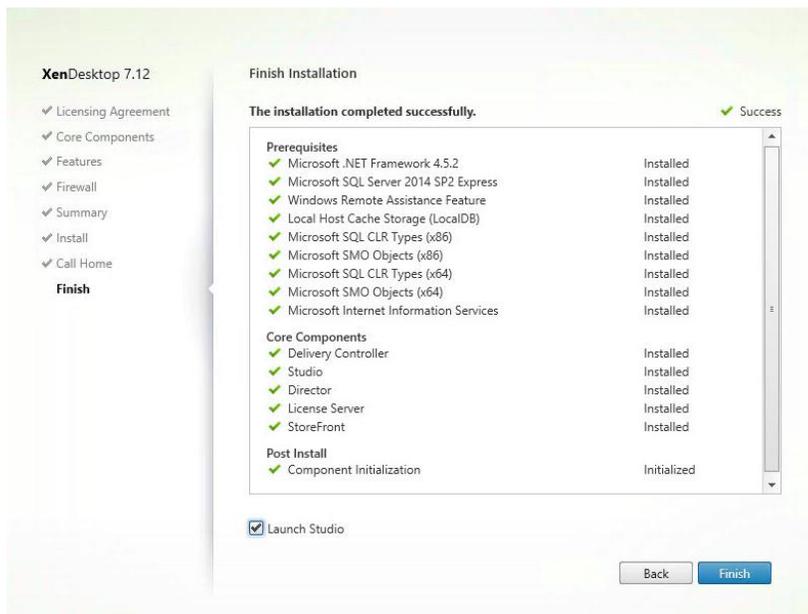
10. Review Summary and Select **Install** to begin installation.



11. Restart the OS as necessary. Once the OS restarts log back in to continue.



12. After install, the **Call Home** window appears. Select whether to participate in Citrix Call Home and continue.

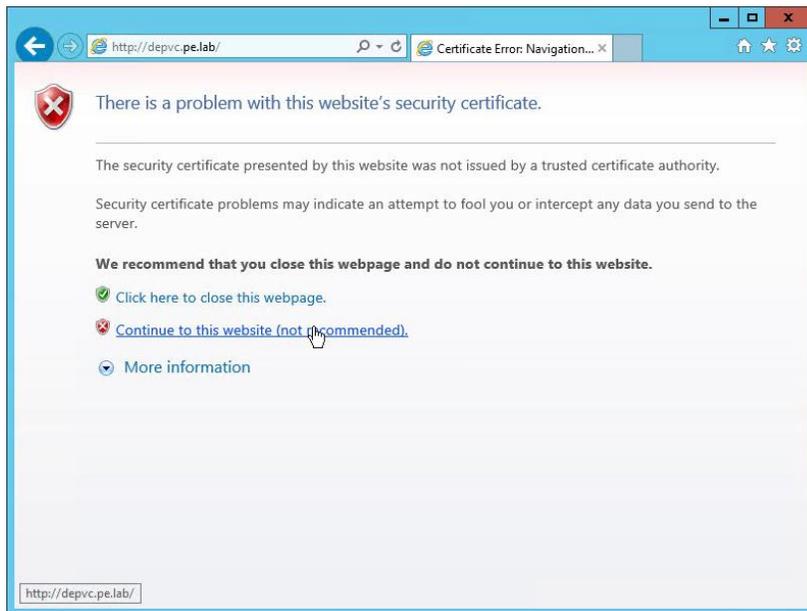


13. Select **Finish** when the **Finish Installation** dialog is displayed.

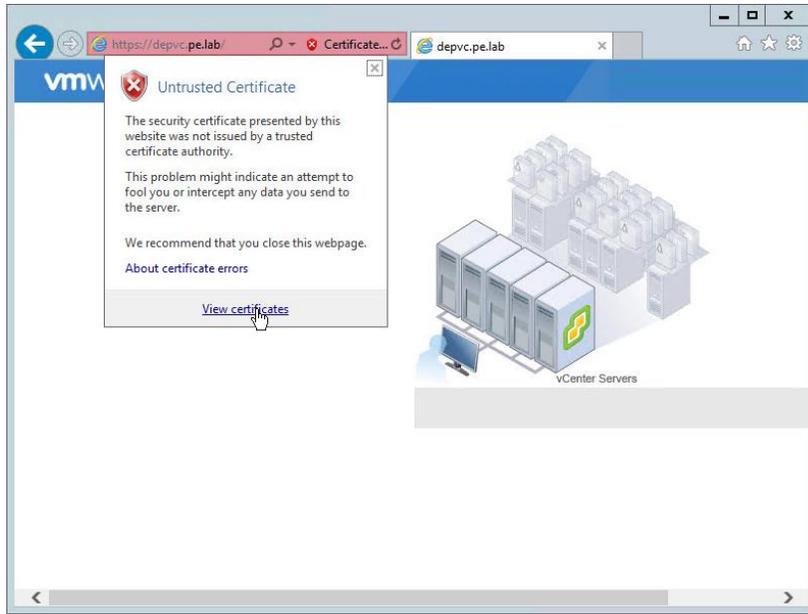
7.3 IMPORTING VSPHERE AND VSPHERE WEB SERVICES SDK CERTIFICATES

Before you can continue with the XenDesktop configuration you must import the vSphere 6.0 server certificate and vSphere Web Service SDK Certificates into the VM that contains the XenDesktop instance.

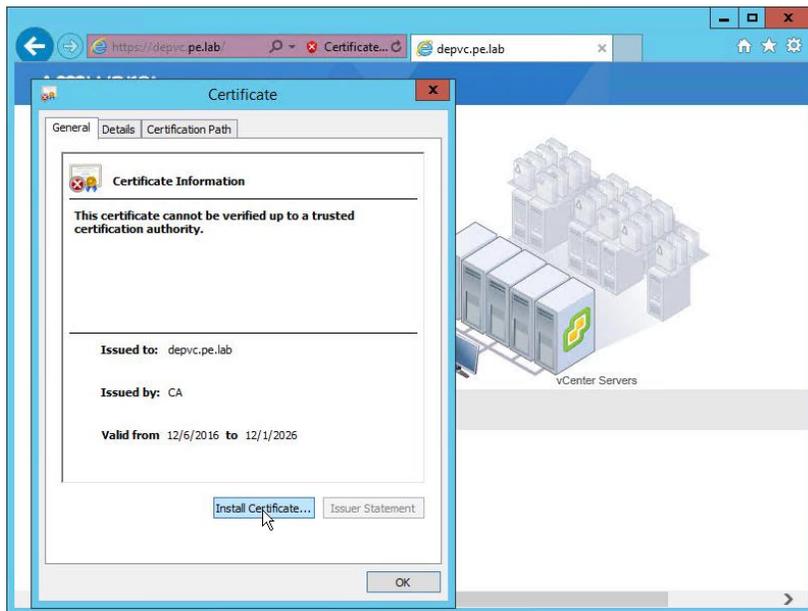
7.3.1 Importing the vSphere Certificate



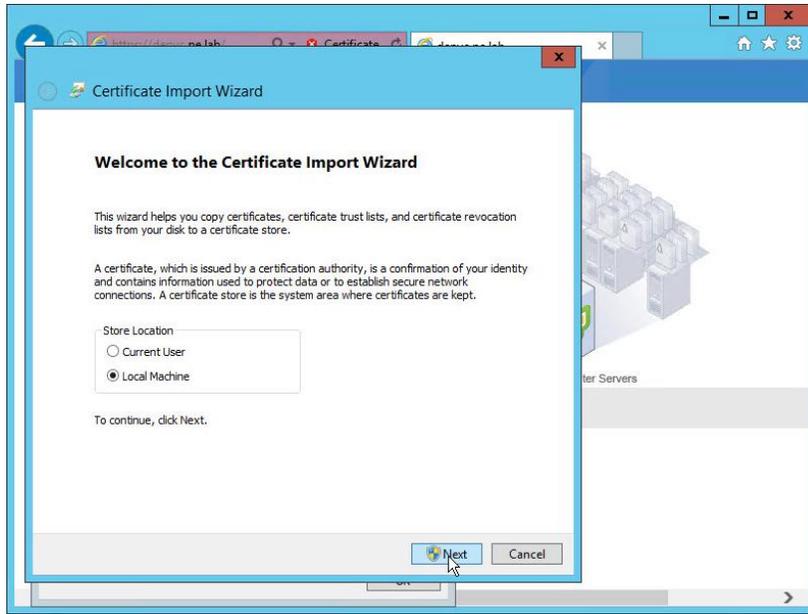
1. Navigate to the website for the VMware vCenter Server UI. Select **Continue to this website (not recommended)**.



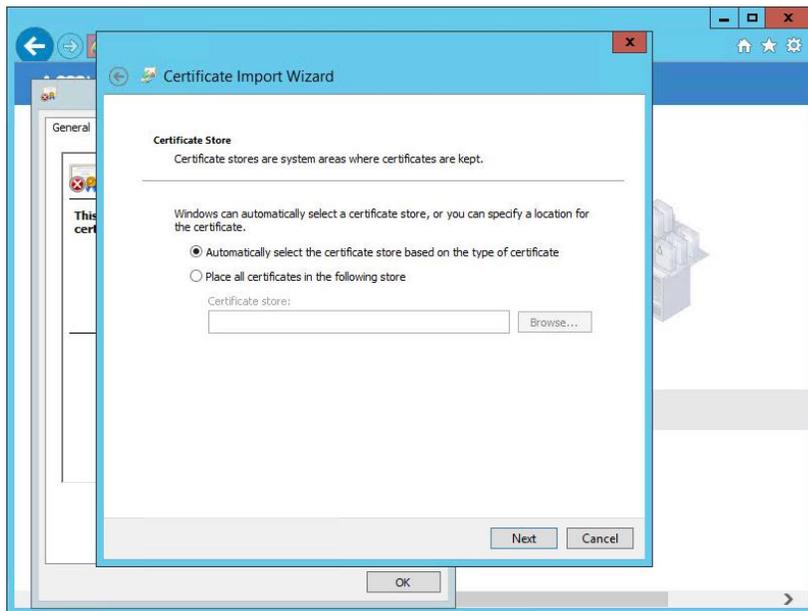
2. Click on the **certificate warning** next to the URL in the address bar and select **View certificates**.



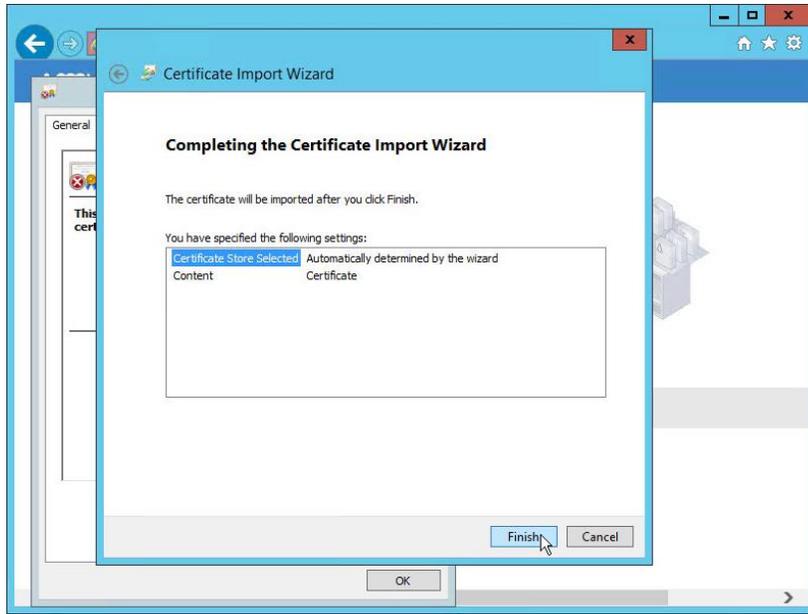
3. Select the **Install Certificate** button to import the certificate into the local certificate store.



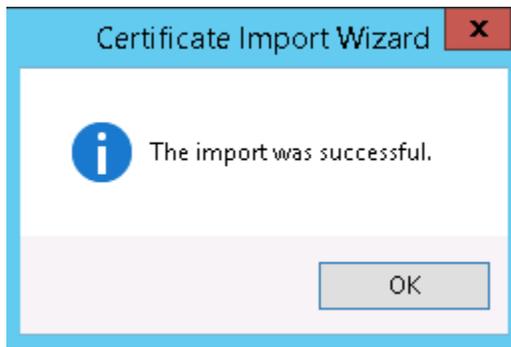
4. Select **Local Machine** radio button on the **Certificate Import Wizard** and then select **Next** to import the server certificate into the local machine certificate store.



5. Select **Next** to accept the default to automatically select the certificate store based on the certificate type.

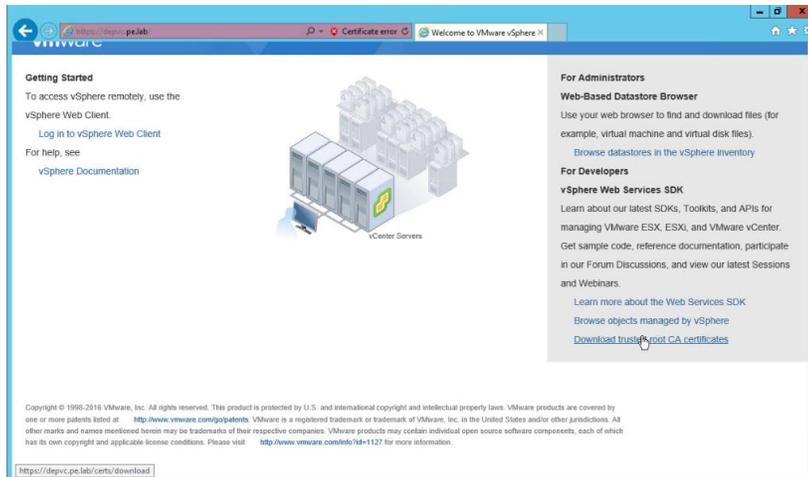


6. Select **Finish** to continue.

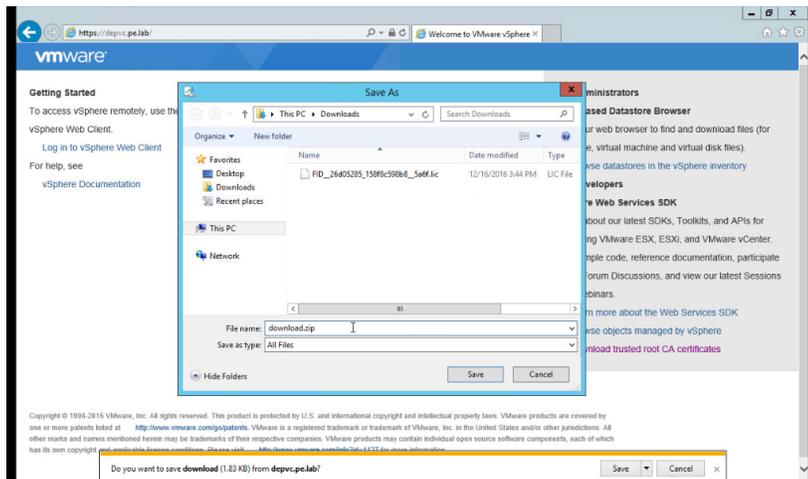


7. Select **OK** to close the notice that the certificate was imported successfully.

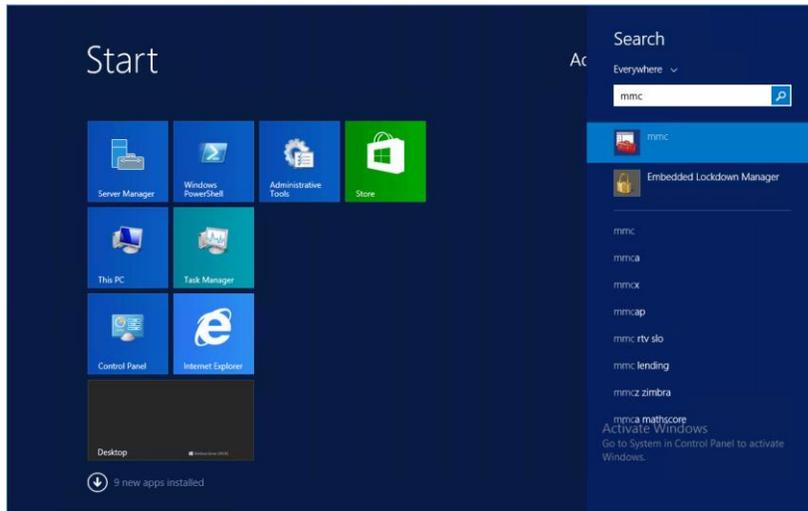
7.3.2 Importing vSphere Web Services SDK Certificates



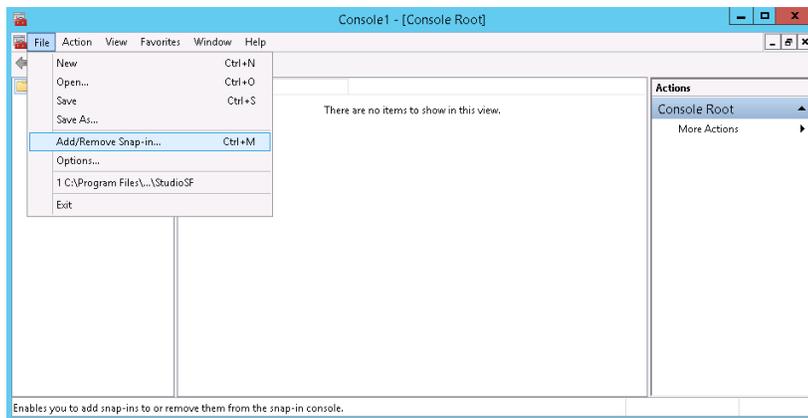
1. Select **Download the trusted root CA certificates** from the vSphere Web Services SDK section.



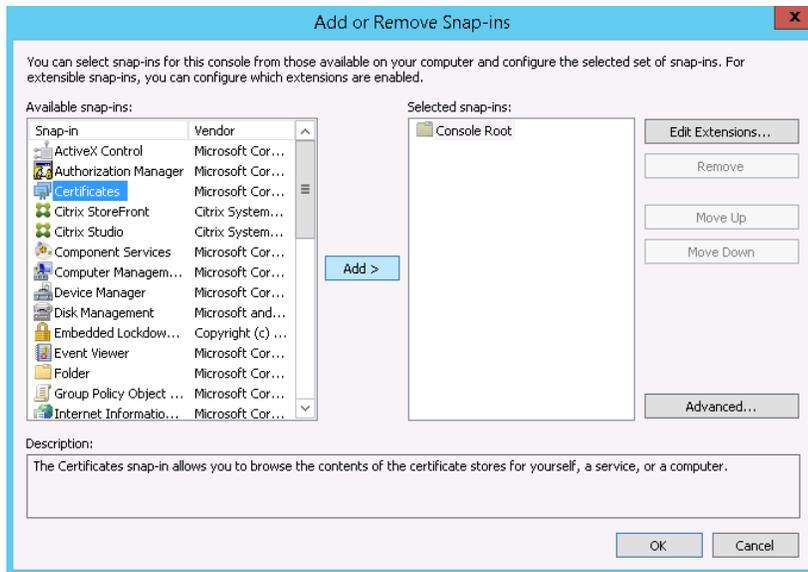
2. Save the file, adding **“.zip”** to the end. Extract the certs from the download.zip file.



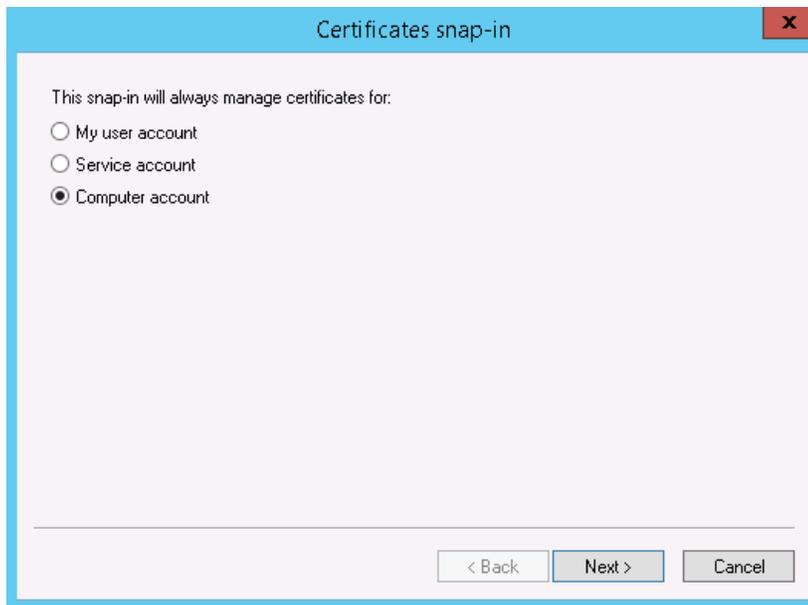
3. Launch MMC control panel from the **Start** → **Search** menu by entering **MMC** in the search window clicking the MMC icon.



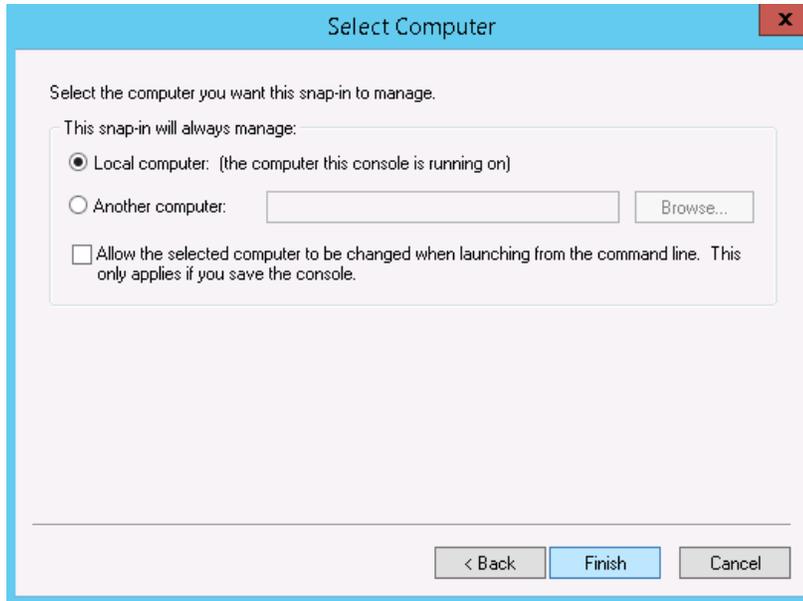
4. Select **Add/Remove Snap-in** from the **File** menu.



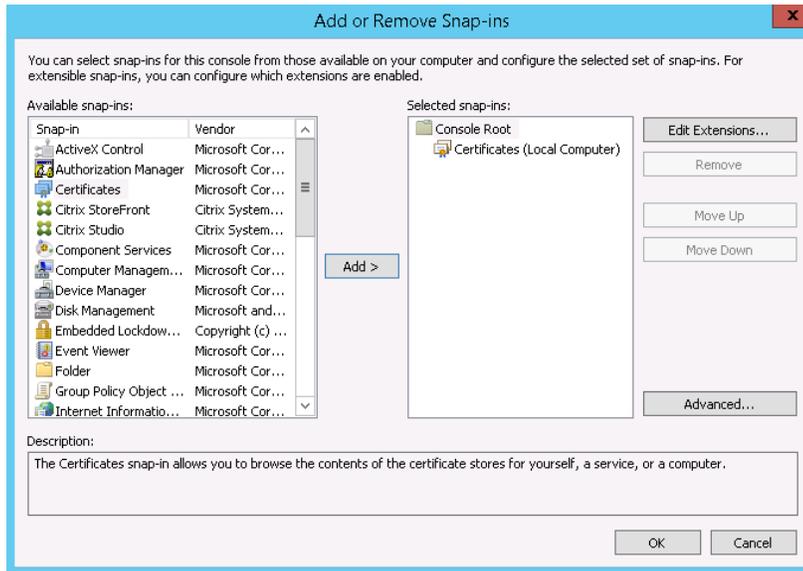
5. Select **Certificates** from the **Available snap-ins** and then click **Add** button.



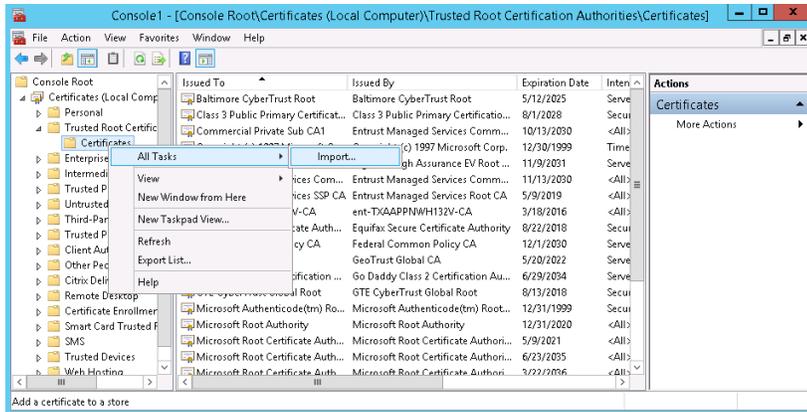
6. Select **Computer Account** radio button and then select **Next** to continue.



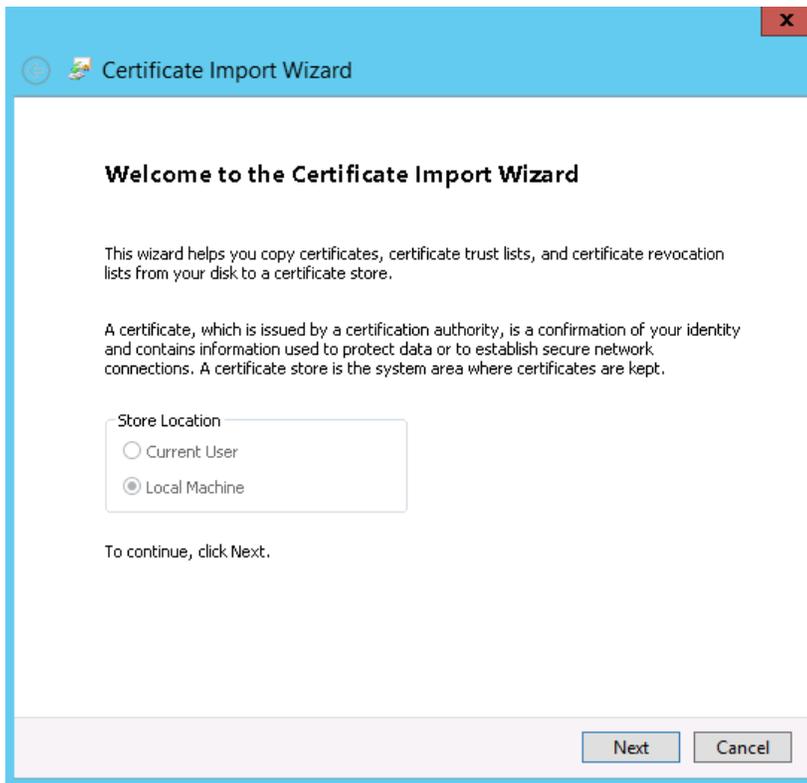
7. Accept the default of **Local computer** as the computer you want to manage, and then select **Finish** to continue.



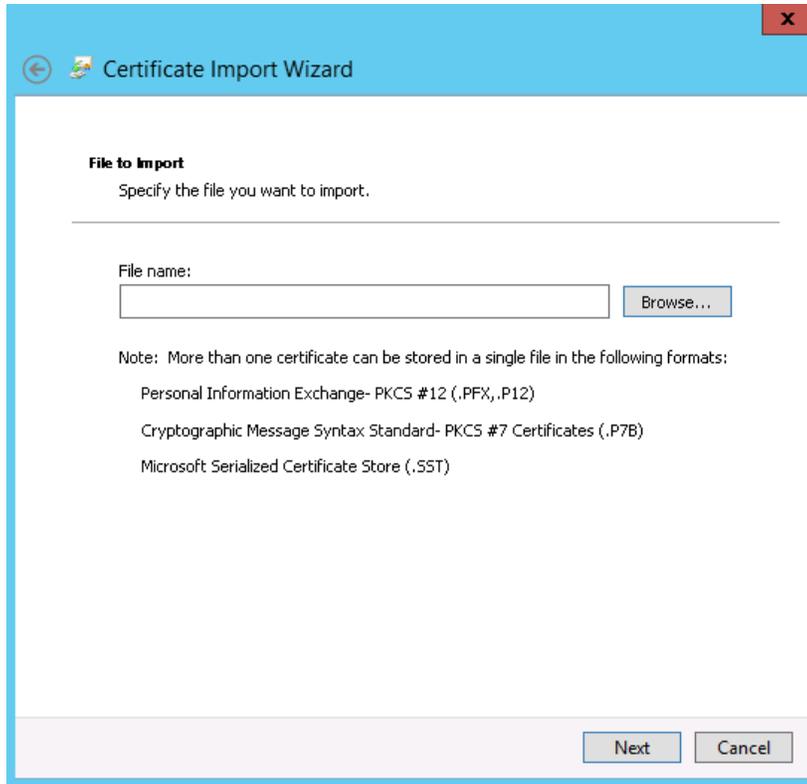
8. Select **OK** to load the snap-in.



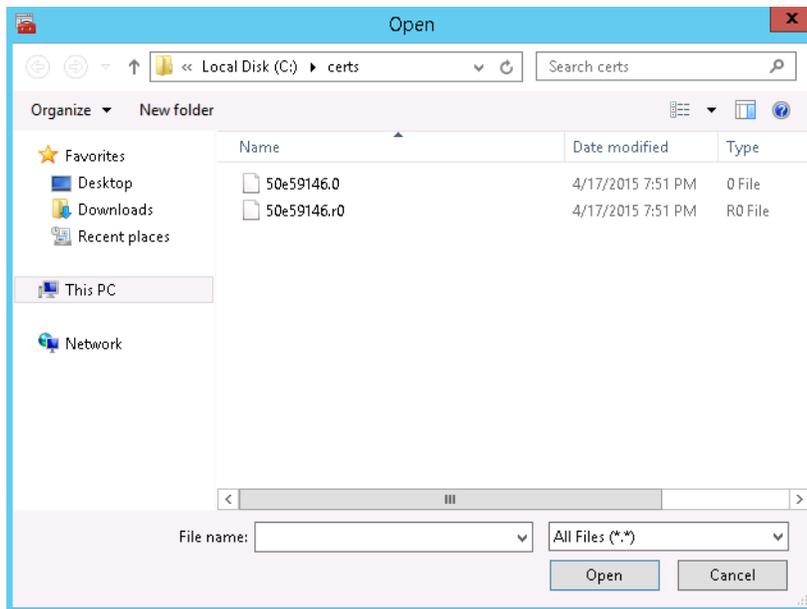
- Expand **Certificates**, and then expand **Trusted Root Certificates**. Right click on **Certificates** under **Trusted Root Certificates** and select **All Tasks**→**Import**.



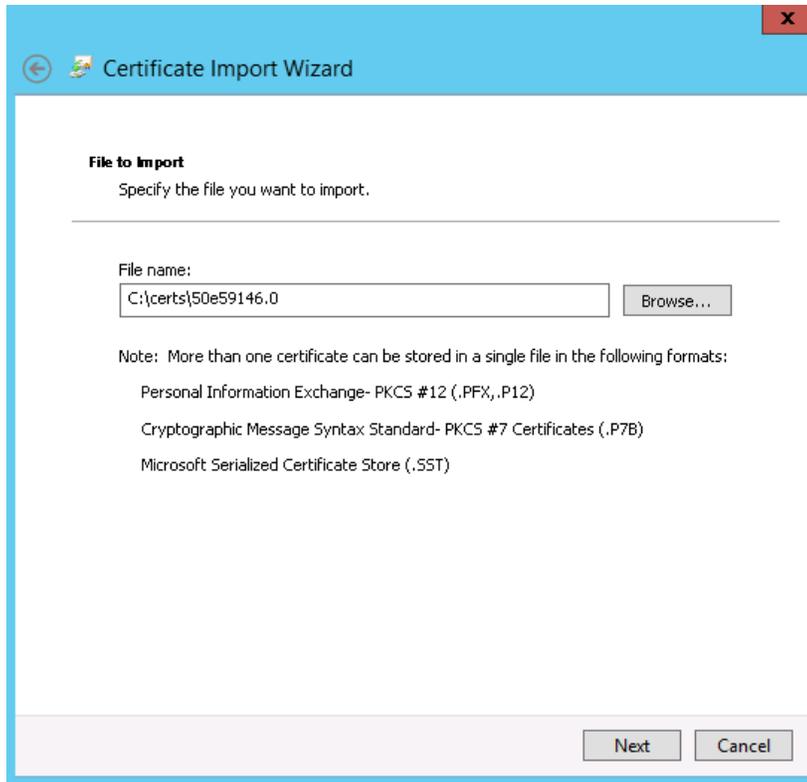
- Select **Next** to continue to import certificates using the **Certificate Import Wizard**.



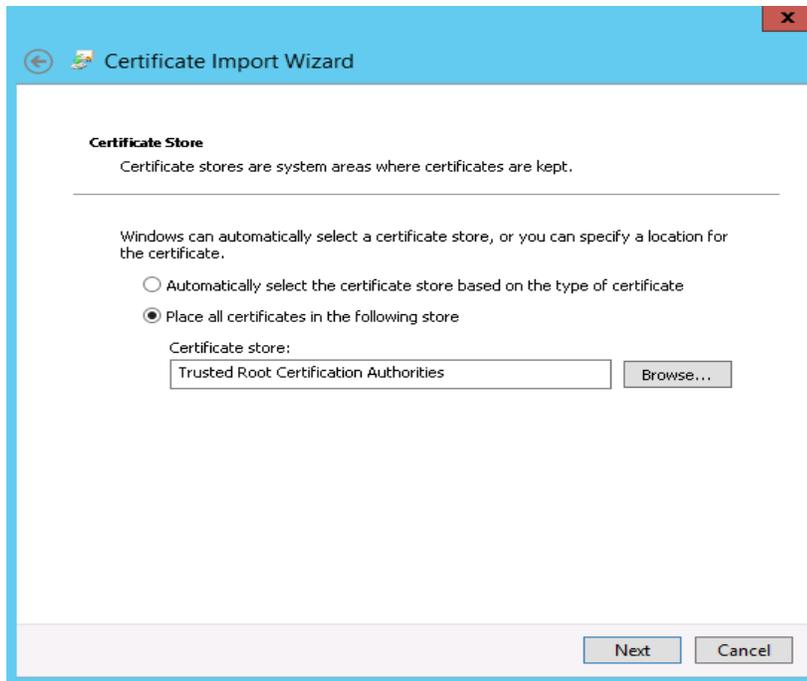
11. Click on the **Browse** button to locate the recently downloaded certificates.



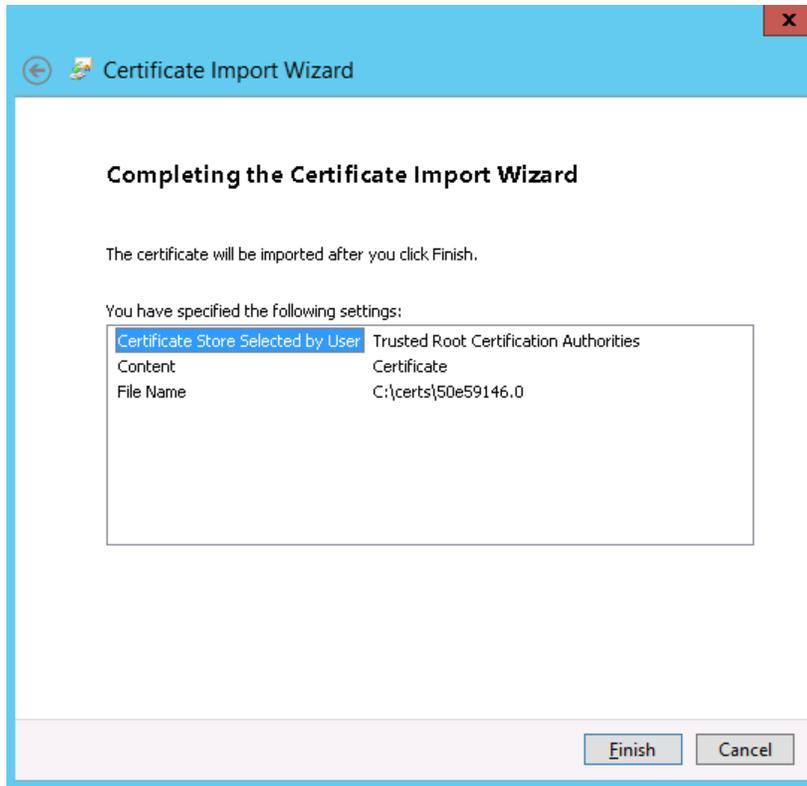
12. Locate the certs folder where you recently extracted it from the download.zip. Select All Files (*.*) file type and then select the **first certificate .0 file**. Click **Open** to choose this certificate file.



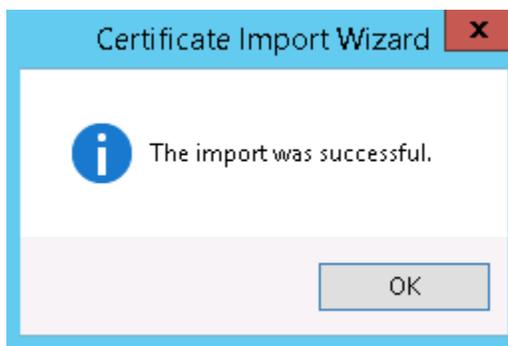
13. Select **Next** to continue.



14. Select **Next** to continue with importing the certificate into the **Trusted Root Certification Authorities**.



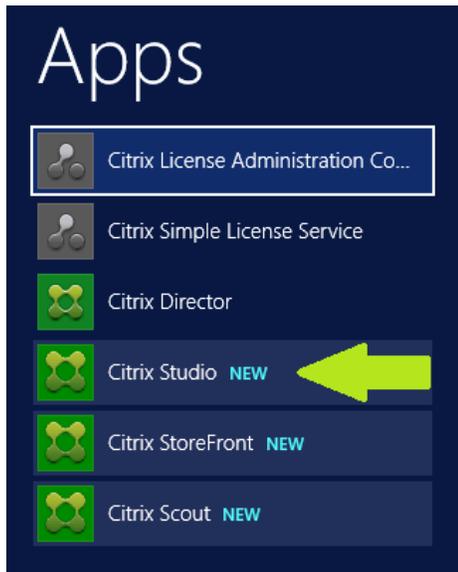
15. Select **Finish** to import the certificate.



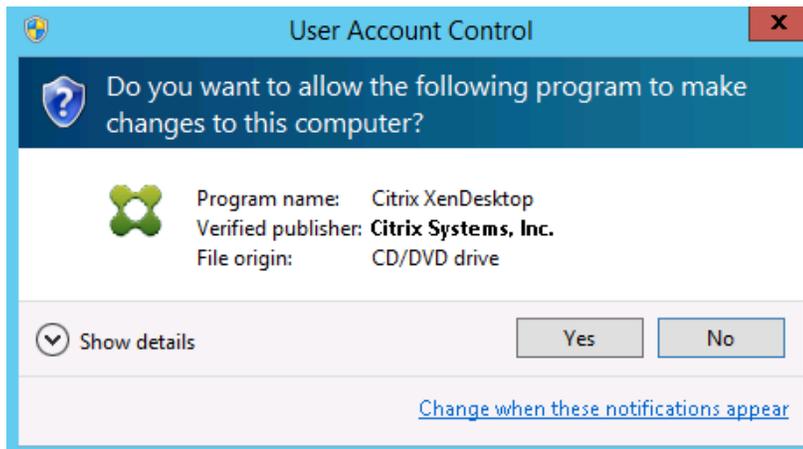
16. Select **OK** to close the notice that the certificate was imported successfully.
17. Repeat the import process for the vSphere Web Services SDK certificate on the second certificate .r0 file extracted from download.zip.
18. Close the MMC window.
19. Select **No**, to not save console settings.

7.4 USING CITRIX STUDIO TO PERFORM SITE SETUP

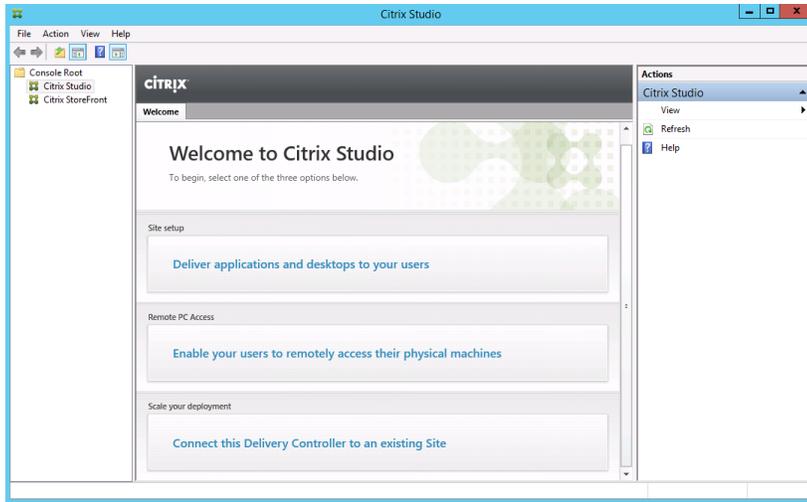
Use the following procedure to configure Citrix Studio, which will be used to manage the XenDesktop 7.12 deployment.



1. On the **Start Menu** search for and select **Citrix Studio**.



2. Select **Yes** to agree to the User Account Control prompt if displayed.



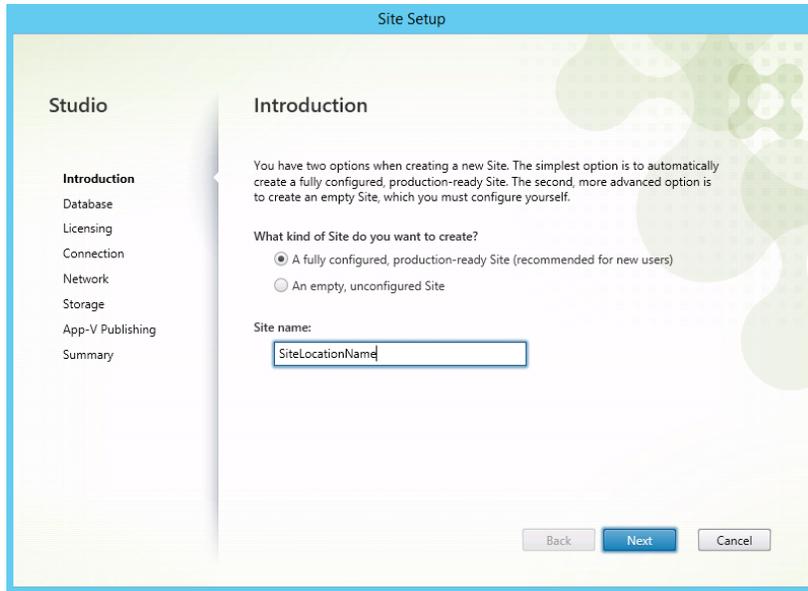
3. Under **Site Setup**, Select **Deliver applications and desktops to your users**.

 If you have not joined the server to the domain you will not be able to run the Site Setup process. The following message will be displayed. If you have received this error, uninstall XenDesktop, join the server to the domain and reinstall XenDesktop before continuing.

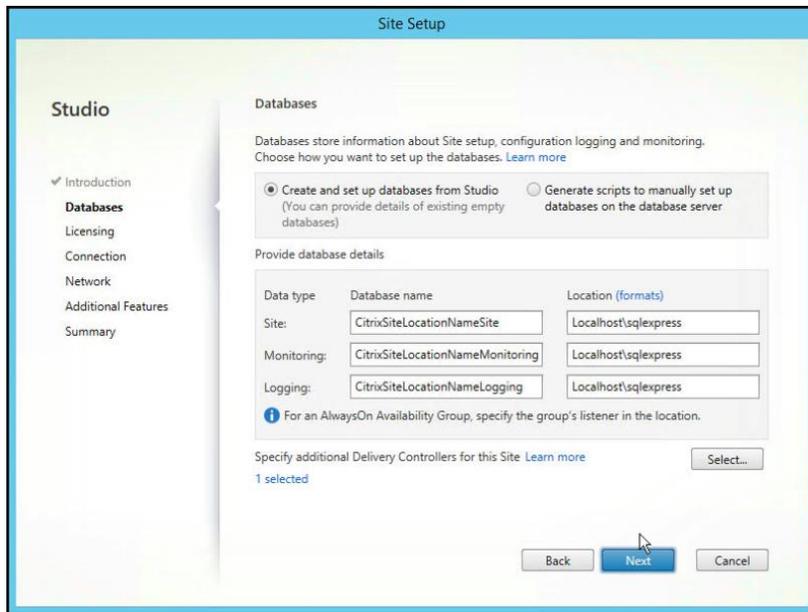
 This machine is currently not joined to an Active Directory domain. To use Studio on this machine, join the machine to a domain.

If you are logged in as a local user rather than a domain user you will not be able to run the site setup process. The following message will be displayed.

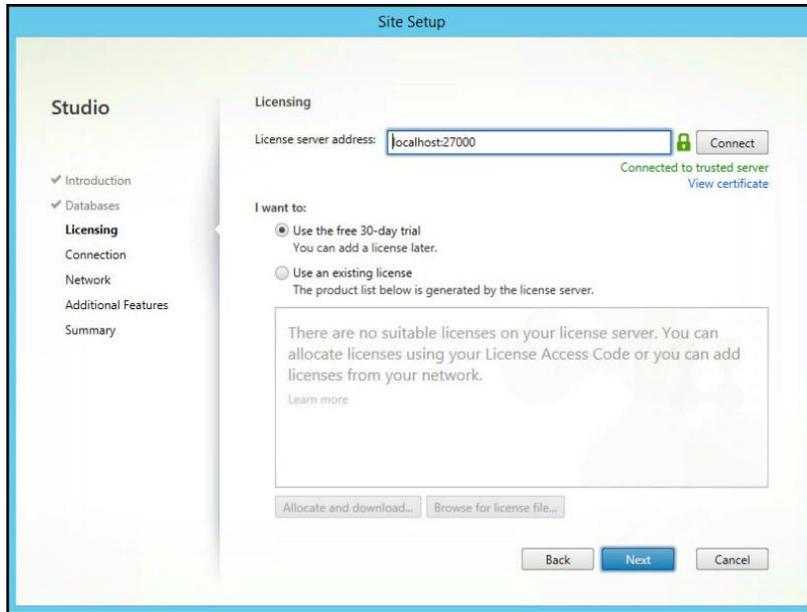
 The current user does not appear to be the member of an Active Directory domain. Studio cannot be run by a local user.



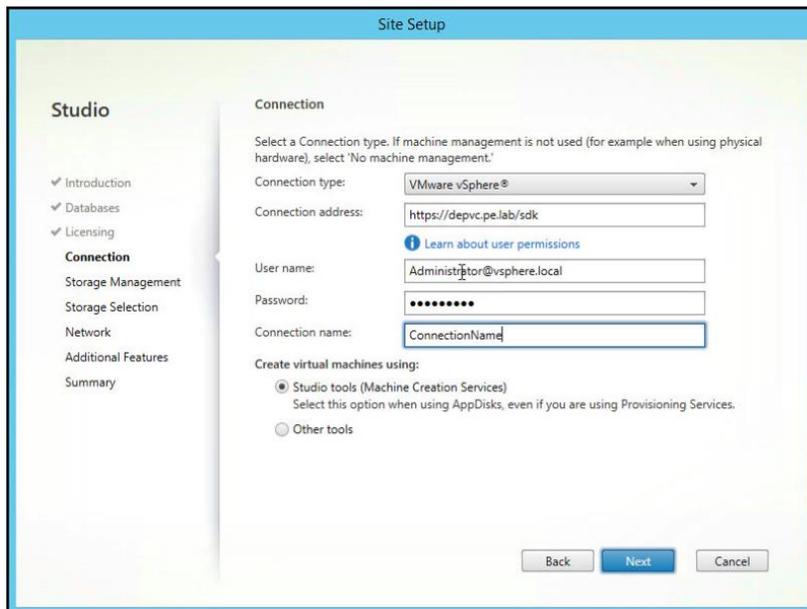
4. On the **Site Setup: Introduction** screen. Enter a site name, and select **Next** to continue.



5. For a single server deployment accept the default settings. For a multi-server deployment with SQL Server 2008 R2 or newer enter the appropriate Database server location and database name. Select **Test Connection** to verify the database server connection. Click **Next** to continue.



6. On the **Licensing** dialog, accept the default settings which will **Use the free 30-day trial**. Select **Next** to continue.



7. On the **Connection** dialog, enter the following:
- ▶ **Connection type:** VMware vSphere
 - ▶ **Connection address:** The vCenter Web Services SDK url
 - ▶ **User name:** administrator@vsphere.local
 - ▶ **Password:** the vCenter administrator password.
 - ▶ **Connection Name:** ConnectionName

Accept the default for **Studio Tools (Machine Creation Services)**. Select **Next** to continue.

If you receive a certificate error, ensure that you have imported the vSphere Certificate, and vSphere Web Services SDK certificates into the Local Machines Trusted Root Certificate Authorities.

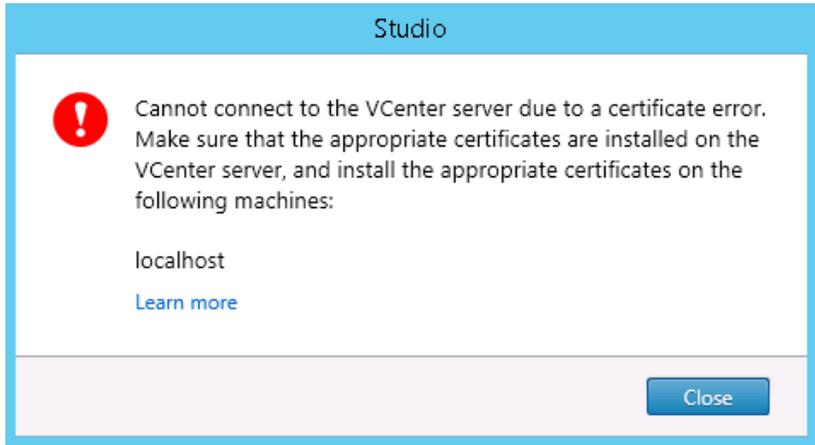
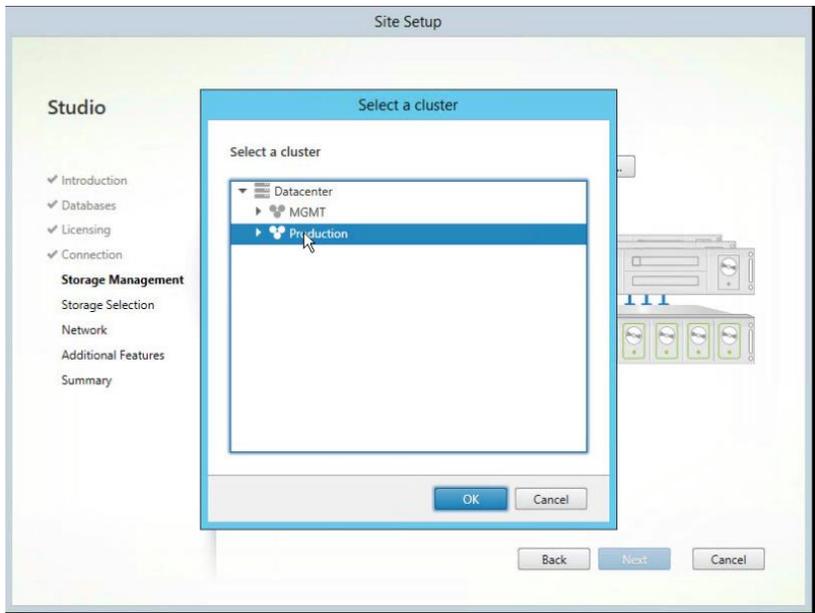
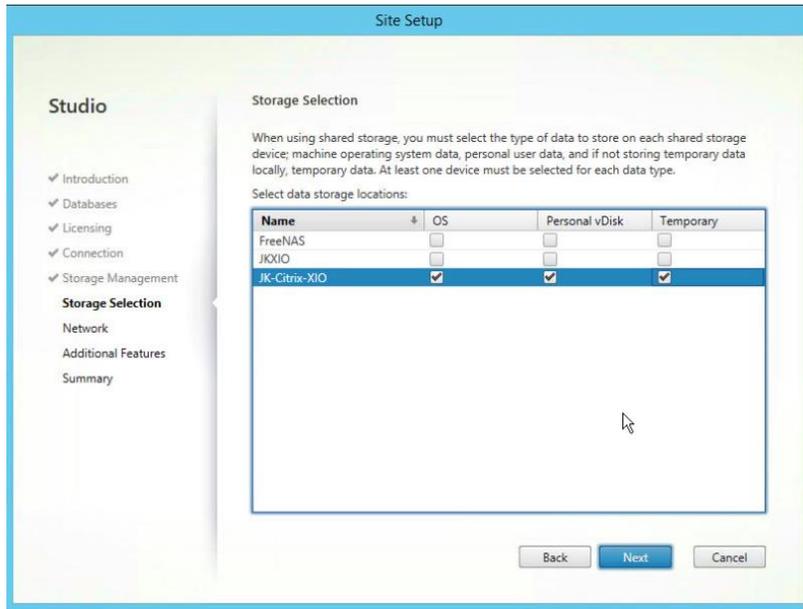


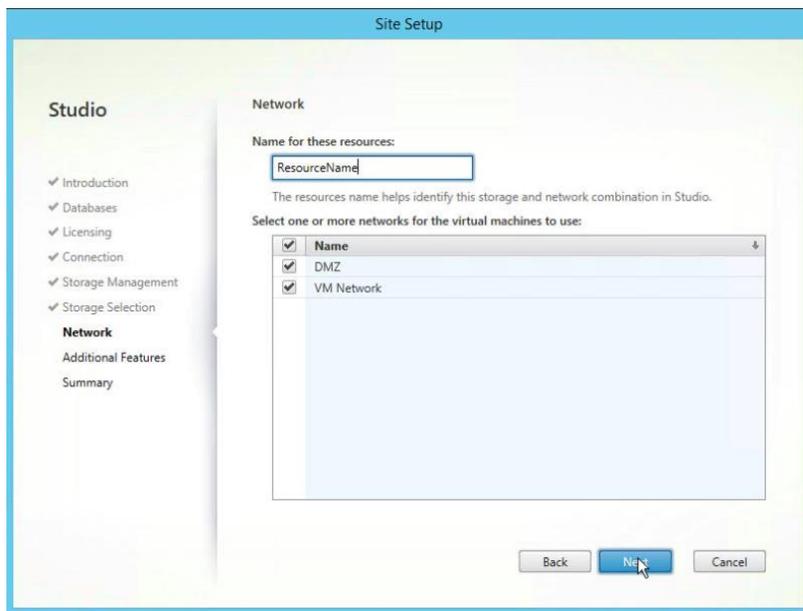
Figure 7-1 Citrix vSphere certificate error



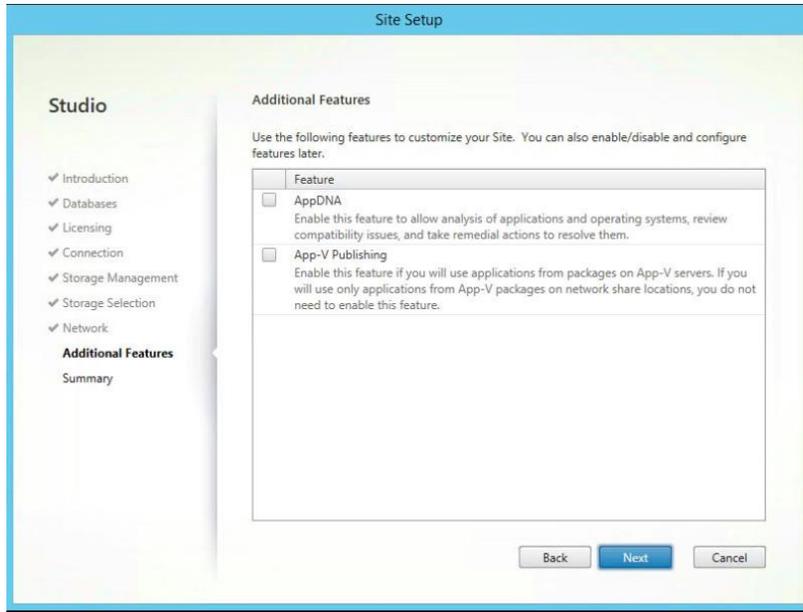
8. In the **Storage Management** section, select a cluster on which to run the virtual desktops by clicking **Browse**. Choose the **Production** cluster and click **Next**.



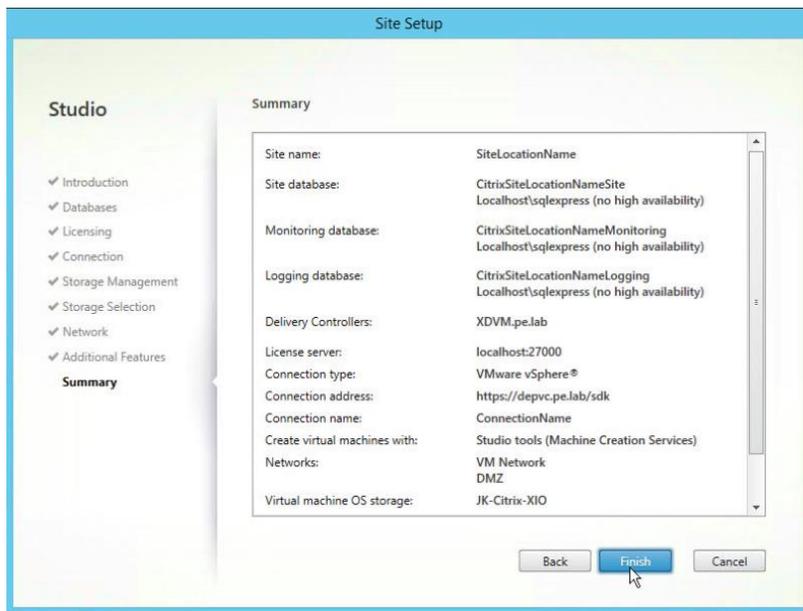
9. On the **Storage Selection** window, select a datastore for each of the storage categories. Then click **Next**.



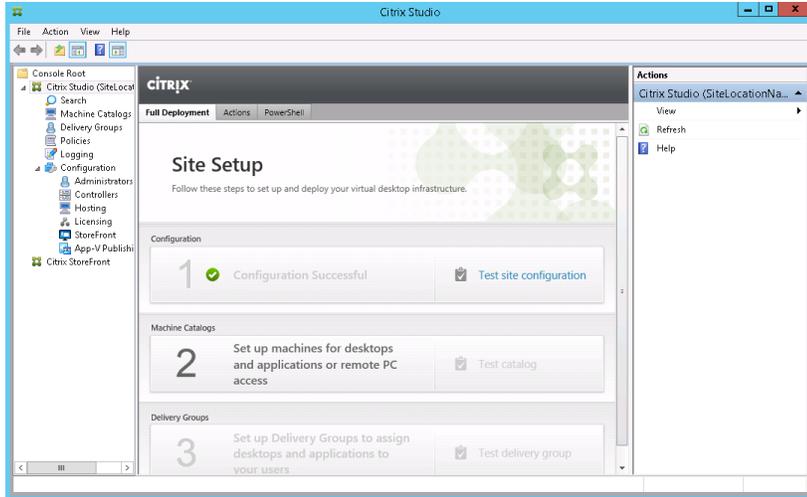
10. On the **Network** dialog box, under **Name for these resources**, enter a resource name for the vSphere networks. Select the Network(s) that you wish the virtual machines to use. Select **Next** to continue.



11. Do not select additional features. Click **Next**.



12. Select **Finish** to complete the Site Setup.



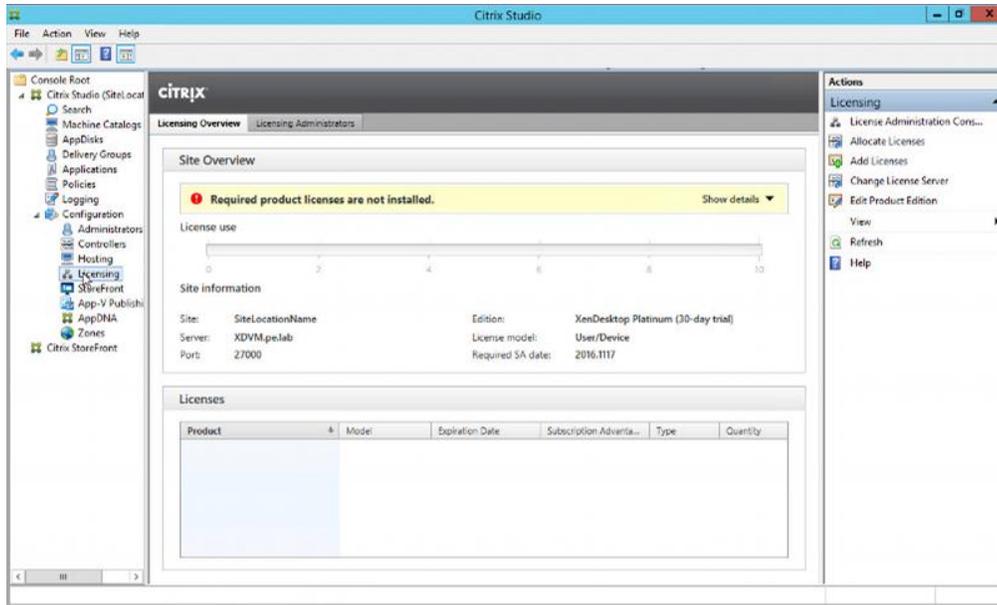
13. Select **Test Site Configuration** to verify everything is configured properly.



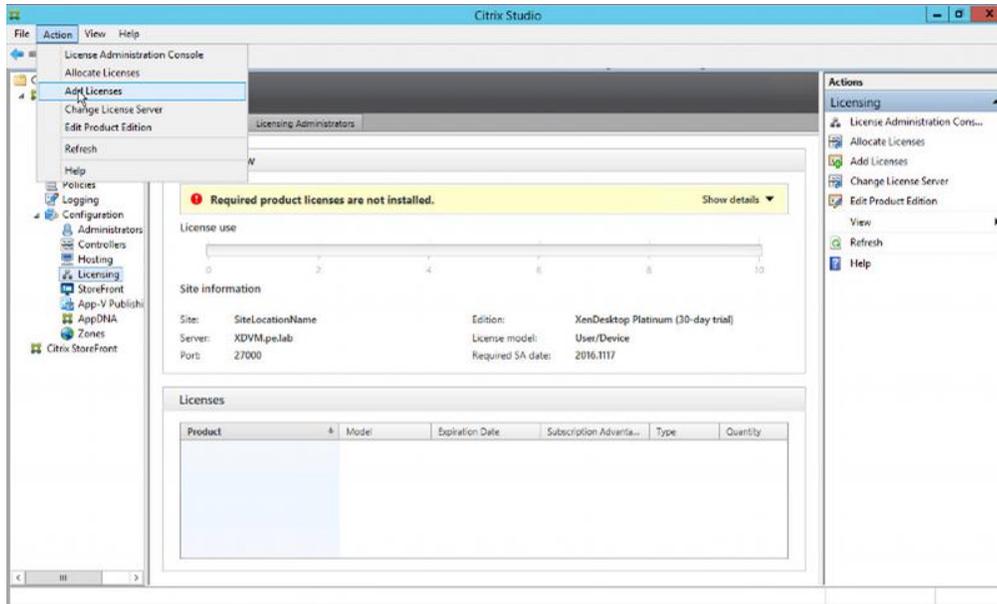
14. Select **Show Report** to view any warnings or failed tests. Select **Close** to exit the **Site Configuration Testing** dialog once all errors are corrected.

7.5 CONFIGURING CITRIX LICENSES SERVER

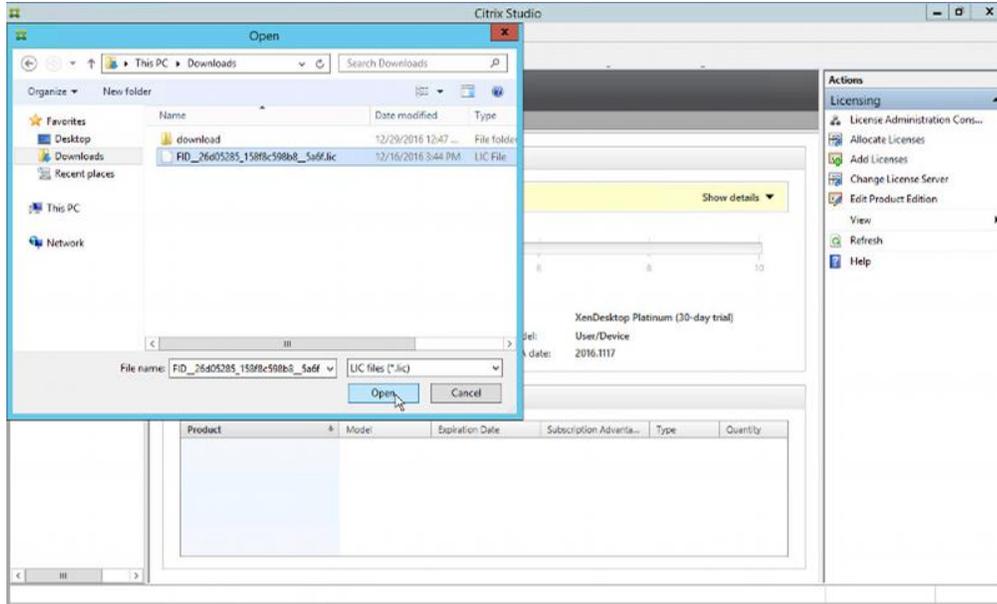
Use the following procedure to configure the Citrix XenDesktop 7.12 Enterprise License server.



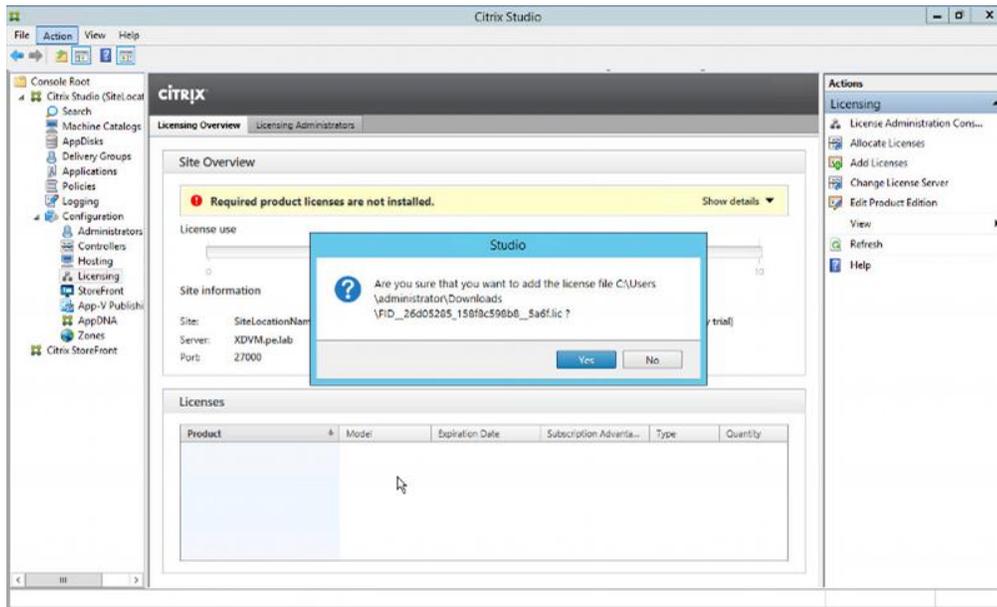
1. Open Citrix Studio on the XenDesktop server. Expand Citrix Studio and then expand Configuration on the left-hand pane. Select Licensing to display the Actions menu on the right pane. The currently allocated licensing will be displayed in the center pane.



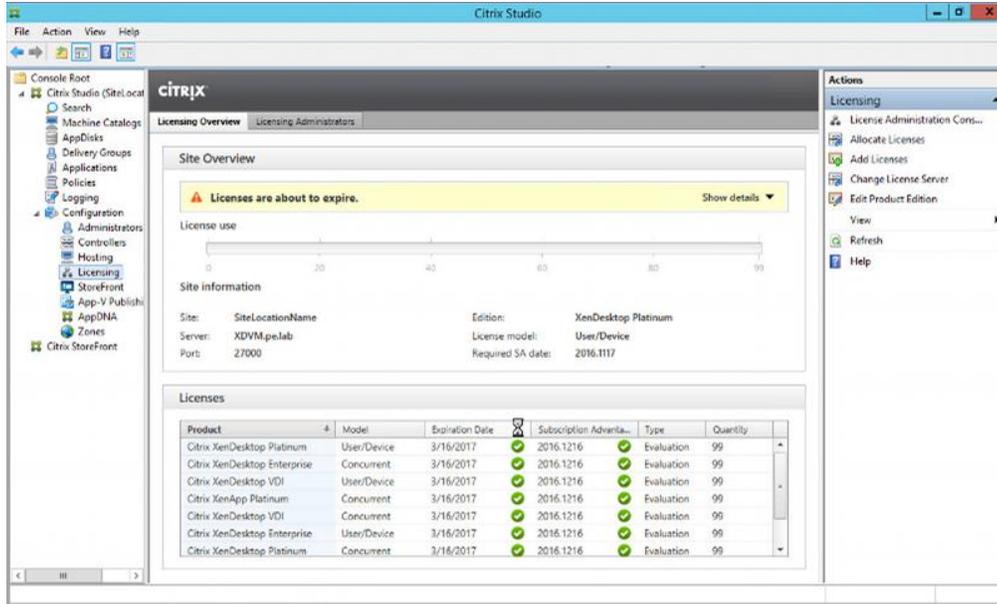
2. Select Add Licenses from the Action pane.



3. Browse the file system on the XenDesktop server to where you previously saved the .lic file downloaded from your Citrix My Account. Select the file, and then click **Open** button to continue.



4. Select **Yes** to confirm to add the license file selected.



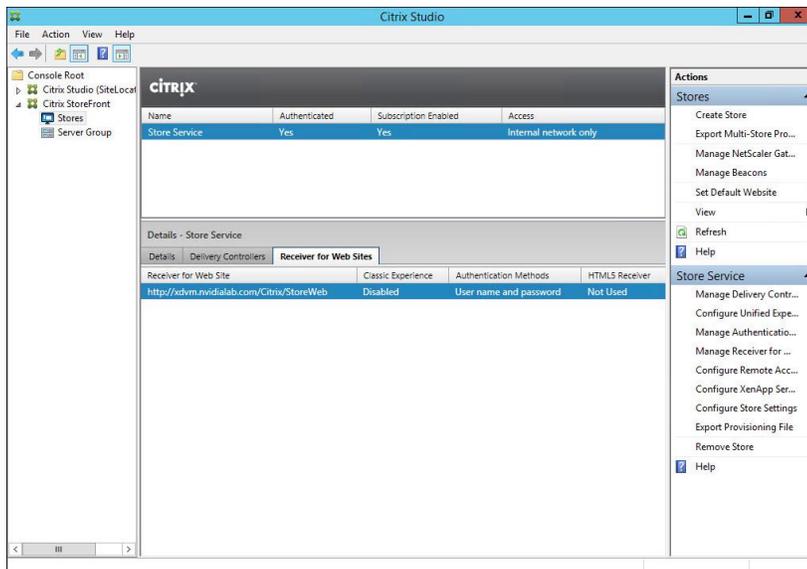
5. The license will be displayed in the center pane of the **License Overview** panel.

7.6 VERIFY CITRIX STOREFRONT CONFIGURATION

Use the following procedure to verify the Citrix StoreFront configuration.



1. Open **Citrix Studio** on the XenDesktop server. Select **Citrix StoreFront** from the left pane. Select **View or Change Stores** to continue to the **Store** dialog.



2. Select the **Receiver for Web Sites** tab in the middle pane. Make note of the **Website URL** that is displayed on the center pane. This is the **Citrix Web Receiver** url and will be used to log in to XenDesktop from a web browser.

CHAPTER 8. SELECTING GRID VGPU PROFILES

This chapter covers GRID vGPU profiles and considerations in choosing them for your user's needs, including:

- ▶ An explanation of the profiles and what they do
- ▶ How to match profiles to user needs

8.1 PROFILES EXPLAINED

As mentioned in the NVIDIA GRID Architecture section, the GPU can be viewed as having two sections: the engines, which do the mathematics of rendering, shading, and deep learning, and the frame buffer, which is used to store the input and output results for access by the VM. To maximize GPU performance for each VM, access to the GPU engines is time sliced between VMs. When a VM has graphics processing to do, it has all the engine resources, like graphics compute engine and video encode/decode, available to it. This ensures that critical graphics functions, which must execute as fast as possible to provide a smooth, usable screen experience, have all the resources they need to finish quickly. This type of sharing means that the individual data for each VMs operations must be partitioned from each other.

GPU frame buffer memory is not, therefore, universally available. In the interest of keeping one VM from interfering with another, the frame buffer is allocated specifically to each VM. Each VM needs its own frame buffer area for the GPU to store the results of graphics operations. This frame buffer memory is allocated using vGPU profiles. GRID vGPU profiles assign custom amounts of dedicated graphics frame buffer to each VM. This way, administrators can allocate the right amount of frame buffer for the expected

role that the VM will play (workstation, desktop, or virtual application). Every virtual machine gets the resources needed to handle the expected graphics load.

Profiles also determine other graphics characteristics, such as maximum number of monitors and maximum display resolutions.

8.2 THE ROLE OF THE GRID VGPU MANAGER

GRID vGPU Manager software, running in the hypervisor layer, implements the requested profiles, and performs GPU scheduling. It allows up to 16 users to share each physical Tesla GPU. It assigns the graphics resources of the available GPUs to virtual machines using a balanced approach.

8.3 POWER USER & DESIGNER PROFILES

NVIDIA GRID was designed to provide excellent performance for all parts of the enterprise. Prior to NVIDIA GRID it was impossible to move your most demanding end users into the data center. They were essentially denied mobility, ease of management, data centrality and security, disaster recovery protection, and the rest of the benefits of virtualization. These users were chained to their workstation, which was chained to a desk. vDGA/Pass-through and other solutions were developed that allowed remote workstation access, but with a 1:1 ratio the costs were high and no resource pooling efficiency was gained. Now with GRID multiple workstation users can be placed on a single host, even combined with other less demanding users to increase density.

Start your profile selection by looking to your primary application's requirements. The software vendors work hard certifying that their application will perform well if the right hardware and software is used. Using those requirements, select the appropriate vGPU profile to meet your end user's needs. To understand more about the graphics requirements of your applications, consult your application vendor.

The key to picking a profile for these users is compatibility and performance. Leveraging the NVIDIA platform, built from proven software and hardware, means you can expect the same high quality experience you, and your key design application, expect from certifying with and using NVIDIA Quadro. The vGPU profiles ending in "Q" undergo the same rigorous application certification process as those Quadro-class processors for professional graphics stand-alone workstations. We refer to them as virtual workstation (vWS) profiles. As a result, you can expect 100% compatibility and

performance with your applications, and ISVs requiring Quadro can certify against the GRID vGPU profiles as well.

8.4 KNOWLEDGE WORKER PROFILES

No user wants a flat, boring experience from their apps and desktop. Application developers know this and continue to increase the rich visual experience when interacting with their applications. While obvious with 3D applications for design or drawing, even the apps used by the clear majority of end users now leverage graphics to one degree or another. These include the underlying operating system's desktop such as Windows 10. Users expect rich web content such as Adobe Flash or HTML 5, better video playback, or the full interactive experience of Microsoft Office (PowerPoint, Excel, Word) with embedded graphics, charts, animation in presentations, etc.

The Knowledge Worker application workload is filled with subtle graphics we take for granted and miss when not present. Adobe PDF adds shading effects to the text in a document, giving it a rich, high quality appearance. When you scroll in a document, you expect the text to smoothly move up or down so you can easily scan the content. The same goes for web pages. Take these effects away from an end user and their adoption of your solution is impacted.

The letter "B" in the profile indicates a Knowledge Worker; someone who uses a virtual desktop but isn't utilizing high end professional visualization applications. The B profiles (vPC) provide more display heads at lower resolution than their Q profile counterparts. This is designed to provide more screen space for a Knowledge Worker, who doesn't need the high resolution that a Designer does.

8.5 APPLICATION DELIVERY VS FULL DESKTOP

There are times when an entire desktop isn't needed, but only one or more applications. The letter "A" indicates a profile designed to support an application versus an entire virtual workstation (vApp). XenApp and other RDSH application delivery software use this profile. These profiles provide only one display, since the application is being rendered instead of the entire desktop. They leverage the power of the NVIDIA Quadro

driver, for a single application. Also, they provide lower resolutions, increasing the responsiveness of the application.

The

8.6 THE FULL LIST OF GRID VGPU PROFILES

The profiles represent a lineup of GRID vGPUs, varying in size of frame buffer and number of heads. The division of frame buffer is what defines the number of users possible per physical GPU with that specific profile, while the number of heads defines the number of displays supported. Max resolution is consistent across all the profiles.

Table 8-1 NVIDIA GRID Graphics Memory Options

Card	vGPU Profile	Frame Buffer (MB)	Virtual Display Heads	Max Resolution	Maximum vGPUs	
					Per GPU	Per Board
Tesla M60 (PCI-e)	M60-8Q	8192	4	4096x2160	1	2
	M60-4Q	4096	4	4096x2160	2	4
	M60-2Q	2048	4	4096x2160	4	8
	M60-1Q	1024	2	4096x2160	8	16
	M60-0Q	512	2	2560x1600	16	32
	M60-1B	1024	4	2560x1600	8	8
	M60-0B	512	2	2560x1600	16	16
	M60-8A	8192	1	1280x1024	1	2
	M60-4A	4096	1	1280x1024	2	4
	M60-2A	2048	1	1280x1024	4	8
M60-1A	1024	1	1280x1024	8	16	
Tesla M6 (MXM)	M6-8Q	8192	4	4096x2160	1	1
	M6-4Q	4096	4	4096x2160	2	2
	M6-2Q	2048	4	4096x2160	4	4
	M6-1Q	1024	2	4096x2160	8	8
	M6-0Q	512	2	2560x1600	16	16
	M6-1B	1024	4	2560x1600	8	8
	M6-0B	512	2	2560x1600	16	16

	M6-8A	8192	1	1280x1024	1	1
	M6-4A	4096	1	1280x1024	2	2
	M6-2A	2048	1	1280x1024	4	4
	M6-1A	1024	1	1280x1024	8	8
Tesla M10 (PCI-e)	M10-8Q	8192	4	4096x2160	1	4
	M10-4Q	4096	4	4096x2160	2	8
	M10-2Q	2048	4	4096x2160	4	16
	M10-1Q	1024	2	4096x2160	8	32
	M10-0Q	512	2	2560x1600	16	64
	M10-1B	1024	4	2560x1600	8	32
	M10-0B	512	2	2560x1600	16	64
	M10-8A	8192	1	1280x1024	1	4
	M10-4A	4096	1	1280x1024	2	8
	M10-2A	2048	1	1280x1024	4	16
	M10-1A	1024	1	1280x1024	8	32

8.7 MATCHING PROFILES TO USER NEEDS

To obtain the best user experience, you need to define your user needs and match them to available NVIDIA GRID vGPU profiles.

If we take three example Revit 2016 user use cases and their computing needs, we can now match the video RAM and display needs to desired GRID vGPU profiles. Your settings may vary, adjust to best meet your user needs.

Table 8-2 User types and appropriate profiles

Configuration	RAM	vCPU	Frame buffer	Displays/resolution	Profile
Entry Level Engineer	4 GB	4 (2.5 GHz)	1 GB	2/1920x1080	M10-1Q or M60-1Q
Mid-Level Engineer	8 GB	4-8 (2.5 GHz)	2 GB	4/4096x2160	M10-2Q or M60-2Q

Configuration	RAM	vCPU	Frame buffer	Displays/resolution	Profile
Advanced Engineer	16 GB	8 (2.5 GHz)	4 GB	4/4096x2160	M10-4Q or M60-4Q

8.8 THE 8Q AND 8A PROFILES VS VDGA/PASS-THROUGH

With the addition of these profiles, each assigned the full 8GB of Frame Buffer memory available in a GPU, you no longer must pull valuable GPUs out and manage them separately, as you would with vDGA/pass-through. Each of these profiles gives the full power of a GPU, without the drawbacks associated with forcing a VM to remain tethered to a specific GPU. When that VM isn't powered on, other VMs can use the GPU for their work, and when the VM is powered on, it can be assigned to another GPU not in use.

Now you can manage all your user groups, from high density Knowledge Workers needing the rich and responsive experience of 1GB of frame buffer and access to a shared Quadro K2200 equivalent GPU (M10-1B), to the most demanding of workstation end users needing 8GB of frame buffer and access to a full M5000 equivalent GPU (M60-8Q). When needs change, just shut down your virtual desktop and assign a new profile, which is far simpler than having to shut down a physical machine, change the PCI card (which inevitably requires a change in driver software), and reboot.

The 8Q and 8A profiles also support NVIDIA CUDA® (<https://developer.nvidia.com/cuda-zone>) and OpenCL (<https://www.khronos.org/opencl/>), required for such programs as XGen, part of the Autodesk Maya product, and the newest frontier in computing: [Deep Learning](#). These technologies play a vital role in the future of Artificial Intelligence, and NVIDIA GRID supports them today in the 8Q and 8A profiles.

CHAPTER 9. APPLICATION BASED HOST SIZING EXAMPLE - ESRI

9.1 APPLICATION BASED SIZING

This section contains an example of the testing, the methodology, and the results that we use to determine server sizing for a given application i.e. what resources do we need to run multiple copies of an application on one server.

First, we look at the specifications of a physical workstation build based on the recommendations of the ISV.

9.2 TYPICAL PHYSICAL WORKSTATION

The ISV's website traditionally details a recommended hardware specification, and while a physical workstation is a dedicated solution very different from the shared, flexible solution of VDI, it provides a good starting point to architect your virtual desktops.

9.2.1 Recommended System Requirements

From [Esri's ArcGIS product page](#):

CPU speed	Minimum: Hyperthreaded dual core Recommended: Quad core Optimal: 2x hyperthreaded hexa core
Platform	x64 with SSE2 extensions
Memory/RAM	Minimum: 4 GB Recommended: 8 GB Optimal: 16+ GB
Display properties	24-bit color depth
Screen resolution	1024x768 recommended or higher at normal size (96 dpi)
Visualization cache	The temporary visualization cache for ArcGIS Pro can consume up to 32 GB of space, if available, in your user profile location.
Disk space	Minimum: 4 GB Recommended: 6 GB or higher
Video/Graphics adapter	Minimum: DirectX 11 feature level 10.1, Shader Model 4.1, OpenGL 3.3, and the EXT_texture_filter_anisotropic and EXT_texture_compression_s3tc extensions. Recommended: DirectX 11 feature level 11.0; Shader Model 5.0; OpenGL 4.5; 2 GB RAM; and the EXT_texture_filter_anisotropic, EXT_texture_compression_s3tc, EXT_swap_control, and ARB_shader_draw_parameters extensions. Note: DirectX 9 and OpenGL2 are not supported in ArcGIS Pro 1.4. Be sure to use the latest available driver.

For testing, we key on recommended specifications when feasible. The goal is to test both performance and scalability, to maintain the flexibility and manageability advantages of virtualization without sacrificing the performance end users expect from NVIDIA powered graphics.

9.3 UX - THE VDI USER EXPERIENCE

To define user experience (UX) requires defining elements of application and user interaction. These can be obvious, like the rendering time for an image to appear, or to smoothly pan across that image, or they may be subtler, like the ability to smoothly scroll down a page, or the “snappy” reaction for a menu to appear after a right click. While elements such as these can be measured, the user’s perception of them is subjective, the measured values harder to interpret.

Users don’t interact with the applications the way that benchmarking software does. They add variables like “think time”, the time they spend looking at their display before interacting again with the application. This time offers an advantage to the underlying resources, such as CPU, as it allows tasks to finish, processes to complete, and is even more beneficial in a shared resource environment such as VDI where one user “thinking” frees up resources for another user who chose that moment to interact with their application. Now factor in other “time” away from the application (meetings, lunch, etc.) and one could expect to get more such benefit to shared resources. The benefit equates to more resources for the user’s session, typically a more responsive application, and thus a better-perceived experience by the end user.

9.3.1 Esri Benchmark Metrics

Esri provides an API that when loaded allows the gathering of user experience (UX) metrics during benchmarking. Esri, as the ISV, knows their product best and defines great user experience as the combination of the following metrics:

- ▶ Draw Time Sum: Consists of the sum of time taken for all of the benchmarks to fully draw, this was defined by Esri to be acceptable up to 45 seconds, less time would be a better UX, but more would be a worsening UX.
- ▶ Frames Per Second (FPS): Esri stated that 30FPS is what most users perceive as a good UX, 60 is optimal but most users do not see a significant difference.
- ▶ FPS Minimum: Esri stated that a drop below 5-10FPS would appear to an end user that the drawing had stopped or “frozen”.
- ▶ Standard Deviation: This would represent the number of tests that were outside the average of the others, typically representing a faulty test. Values should be <2 for 2D and <4 for 3D workloads.

After initial testing it was clear that Draw Time Sum would be inside its acceptability threshold if the other three metrics were also inside their respective thresholds. This

gave us a single value to track. Once that single value was validated, we could then validate that the rest of the results were within acceptable ranges as well.

9.3.2 Real Life Experience vs Benchmarking

Our goal is to find the most accurate possible proxy for testing, but this is still not the same as real users doing real work on real data. The NVIDIA GRID Performance Engineering Labs is committed to working with customers to find more and better models, and field confirmation of findings.

9.3.3 The Importance of Eyes On

It's important to view the tests to be sure the experience is in fact something a user would enjoy. We discovered that the initial tests did not contain panning, or "Navigation", something that users do on a common basis. In adding this functionality to the "Heavy 3D + Increased Think Time" tests resulted in a drop from the maximum 16 users down to 12.

9.4 TESTING METHODOLOGY

To ensure you will be able to reproduce our results, we have deliberately chosen a peak workload and run simultaneous tests. Meaning all tested virtual desktops are doing the same activities at the same time, a "Peak Workload" that should be unrealistic of real user interaction but shows the number of users per host when extreme demand is put on the shared resources. This gives us one end of the spectrum.

In the case of ArcGIS, we focused on rendering pipeline (DirectX 11 based) to determine the impact of GPU on performance on scalability. OpenGL testing will be covered in future guides.

These tests did not focus on analytics, as this is more impactful on networking (assuming remote data) and CPU. We intend to test analytics operations in the future. Given ArcGIS Pro is synchronous you can render while running analytics in the background. The impact of this will need to be tested. The following information details the test methodology used with the Esri testing API:

- ▶ Sample workload: Esri provided their “Philly 3D” workload for us to test with. This test is described by Esri as representative of a “heavy 3D” map.
- ▶ Scripting: Esri provided scripting designed to run the application through several benchmarks. In later passes we added panning (“Navigation”) to the script to better imitate end user behavior when viewing a map.
- ▶ Think Time: The Esri script allowed for “think time” adjustment, we started with the default of 2 seconds, then adjusted to 5 seconds, then tried 10, this to imitate time a user might think before selecting another bookmark in the test.
- ▶ Scalability: Scripting also allowed for test runs of as many virtual desktops as required. We ran 1, then 8, then 16, to get a baseline of results and accompanying logs (CPU, GPU, RAM, networking, storage IOPS, etc.).

9.5 RESULTS OVERVIEW

▶ vCPU:

Esri stated that ArcGIS is CPU bound, meaning it relies heavily on the CPU to perform and that this shared resource will be most likely to bottleneck first. Based on the preceding physical spec we chose 4 vCPUs/virtual workstation to start, then tested with 6 and 8 vCPUs respectively.

Result: Our testing found 4 vCPUs performed the best based on the benchmarks Esri provided.

▶ vRAM:

Esri stated they are not RAM intensive, based on recommended specifications we started with 6GB/virtual desktop. During testing we also tested with 4GB of RAM, but this caused CPU spikes, we determined this was caused by the application requiring 5+GB of RAM to load the “Philly 3D” dataset. Tests with 8GB of RAM per virtual desktop did not produce improved performance. You should base the amount of RAM on the needs of your actual workloads.

Result: For this workload ≥ 6 GB of RAM is required.

▶ GPU

Esri states that a GPU is required for 3D workloads. A test with the full GPU and 4GB of frame buffer, offered high performance but lacked scalability. We then tested with 2GB of frame buffer, up to its maximum of 8 users it was still within the metrics provided by Esri. This meant there was room for more users if we lowered frame buffer to 1GB. We also tried 512MB of frame buffer, but this was too little VRAM and caused CPU swapping impacting performance and scalability, and proved GPU is necessary.

Result: For ArcGIS Pro performance AND scalability, 1GB profile was best.

► **Storage**

We used iSCSI attached SAN over 10G non-trunked networking. Our tests never exceed ~25,000 IOPS.

Result: Clearly local spindle bound storage would have been a bottleneck and impacted performance, the fast cache SAN handled the IOPS load.

► **Networking**

We used 10G and 1GB distribution networking At no time was networking a bottleneck and results were unremarkable.

9.6 CONCLUSION

After thorough testing, we have concluded that a profile providing 1GB of frame buffer was sufficient to provide a robust, immersive and productive experience and remains within the thresholds set by Esri. **Therefore, GRID vGPU profiles M10-1Q, M60-1Q, or M6-1Q, should be sufficient to allocate to Esri ArcGIS virtual workstations.**

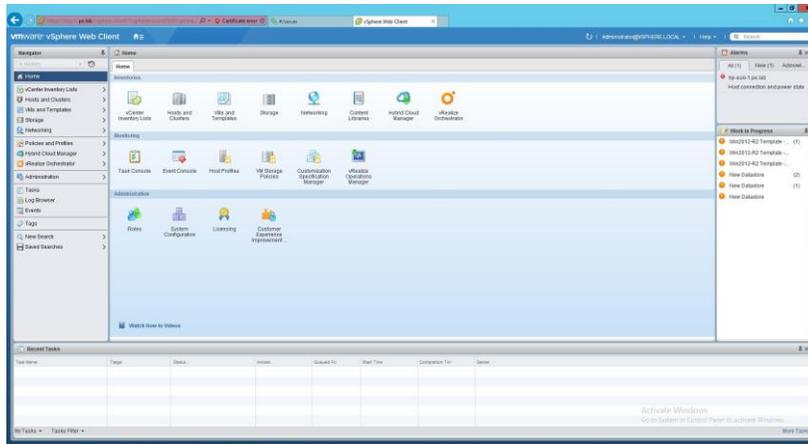
CHAPTER 10. CREATING YOUR FIRST GRID VIRTUAL WORKSTATION

This chapter describes how to:

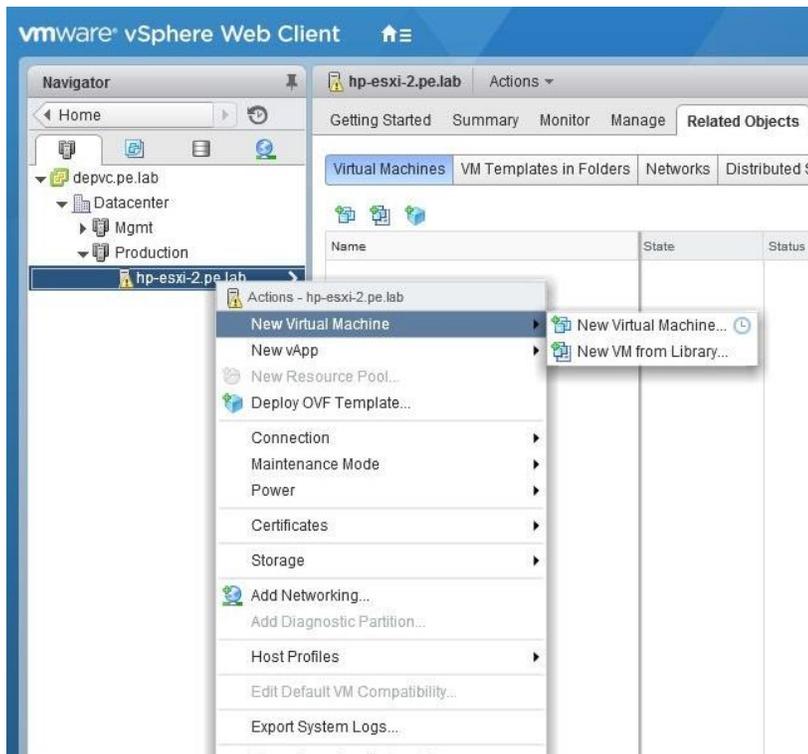
- ▶ Create and configure a virtual machine in vSphere
- ▶ Install Windows 10 and VMware Tools in the VM
- ▶ Adding the Gold Master image to the domain
- ▶ Install Citrix Virtual Desktop Agent on the VM
- ▶ Enable the vGPU in the VM
- ▶ Installing the NVIDIA Driver
- ▶ Licensing the vGPU in the VM
- ▶ Removing the vGPU in preparation for creating XenDesktop machine catalogs

10.1 CREATING A VIRTUAL MACHINE

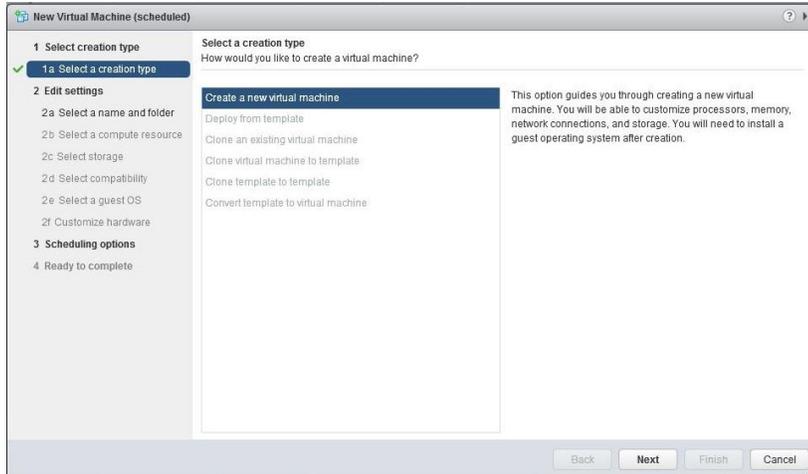
These instructions detail making a VM from scratch that will support NVIDIA GRID. Later, the VM will be used as a “golden” master image for subsequent VMs. Use the following procedure to create the VM.



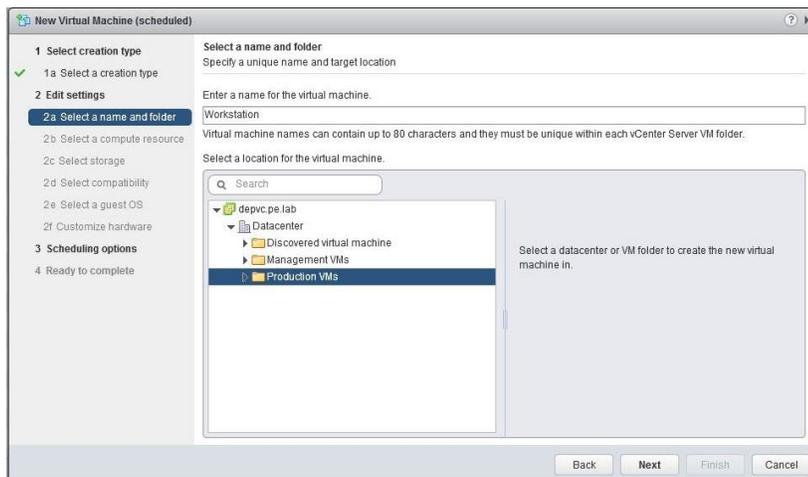
1. Browse to the host or cluster using the vSphere Web Client.



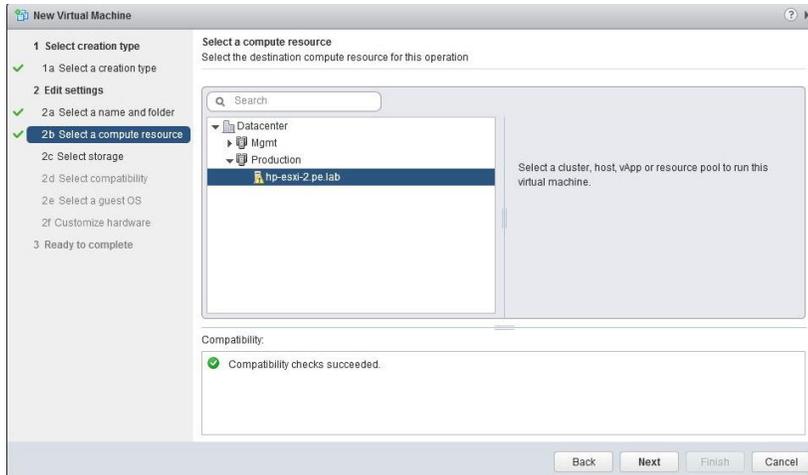
2. Right-click the desired host or cluster and select **New Virtual Machine**.



3. Select **Create a new virtual machine** and click **Next**.

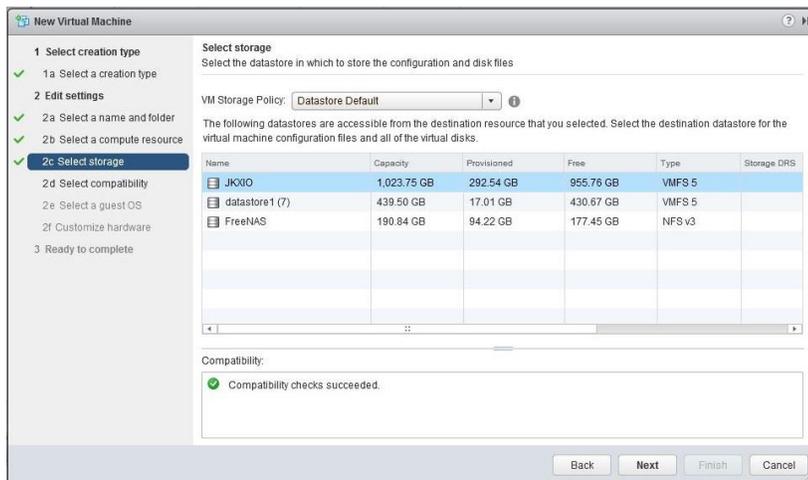


4. Enter a name for the virtual machine. Choose the location to host the virtual machine using the **Select a location for the virtual machine** section. Click **Next** to continue.

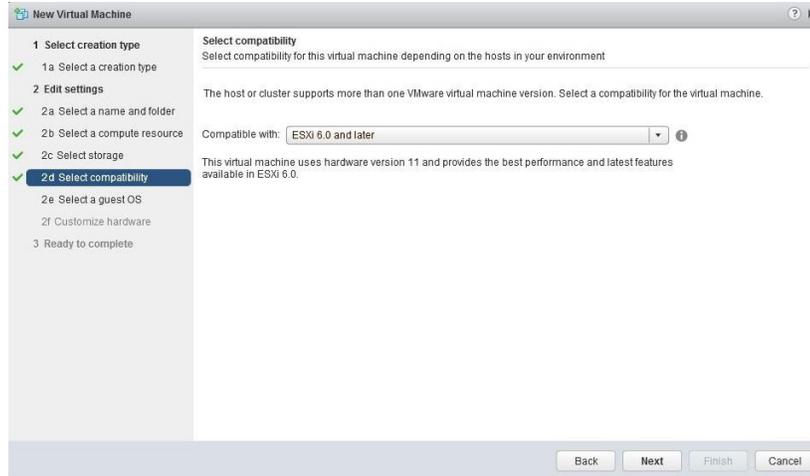


5. Select a compute resource to run the VM. Click **Next** to continue.

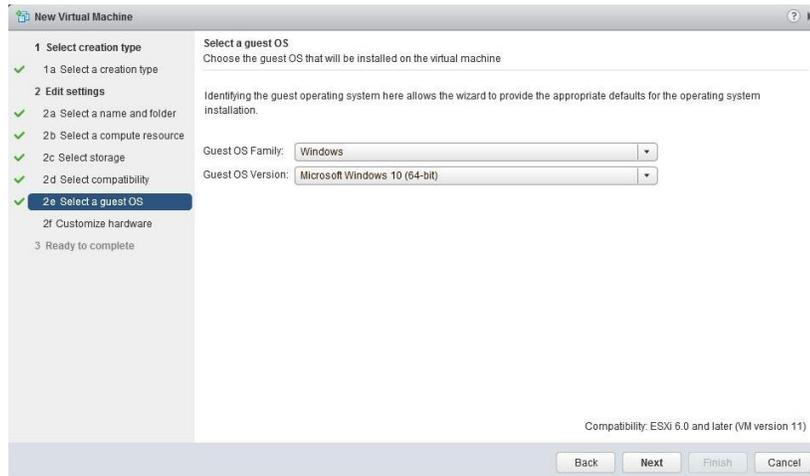
 This compute resource should include an NVIDIA Tesla GPU installed and configured correctly.



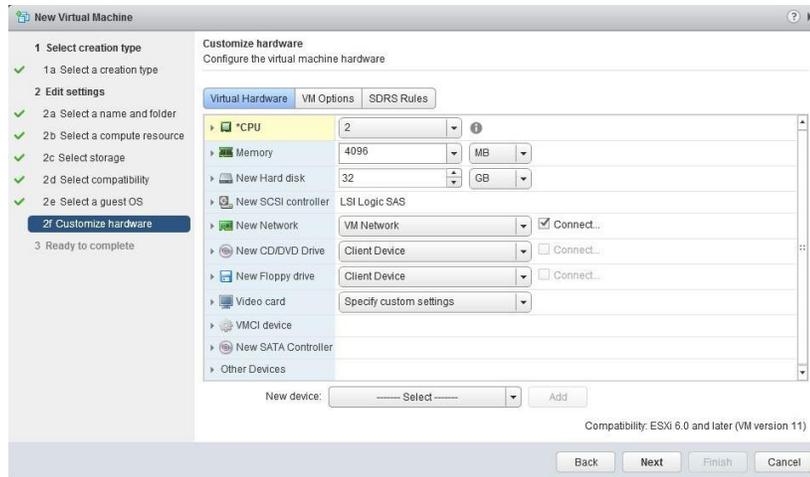
6. Select the datastore to host the virtual machine. Click **Next** to continue.



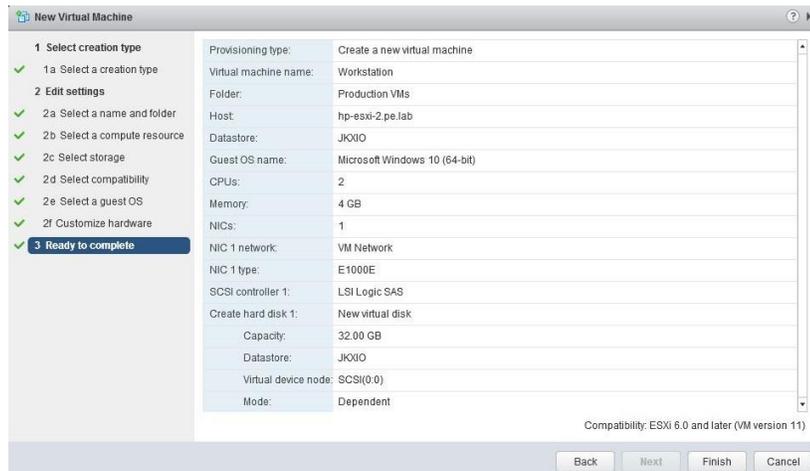
7. Select **ESXi 6.0 and later**. Click **Next** to continue.



8. Select the appropriate Windows OS from the **Guest OS Family** and **Guest OS Version** pull-down menus. Click **Next** to continue.



9. Select 2 vCPU and 4096 MB memory. Click **Next** to continue.

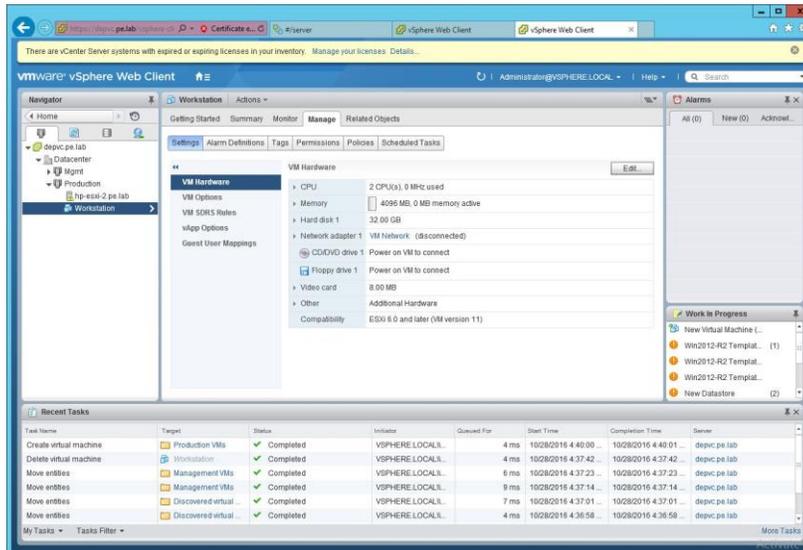


10. Review the **New Virtual Machine** configuration prior to completion. Click **Finish** when ready.

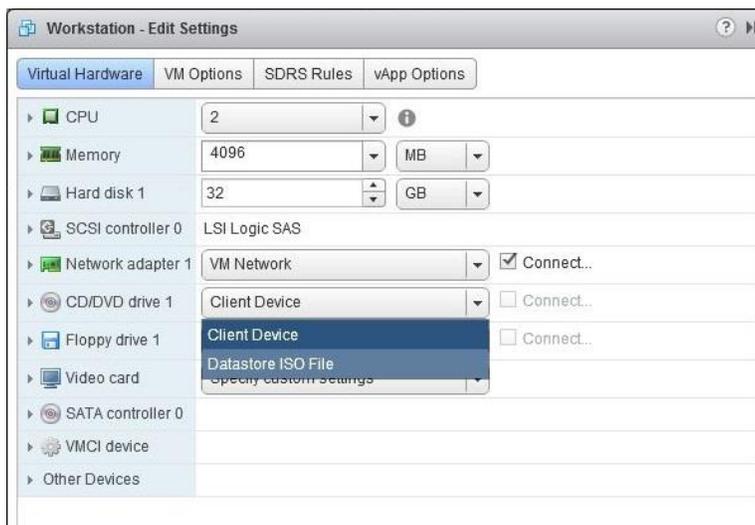
The new virtual machine has now been created.

10.2 INSTALLING WINDOWS 10

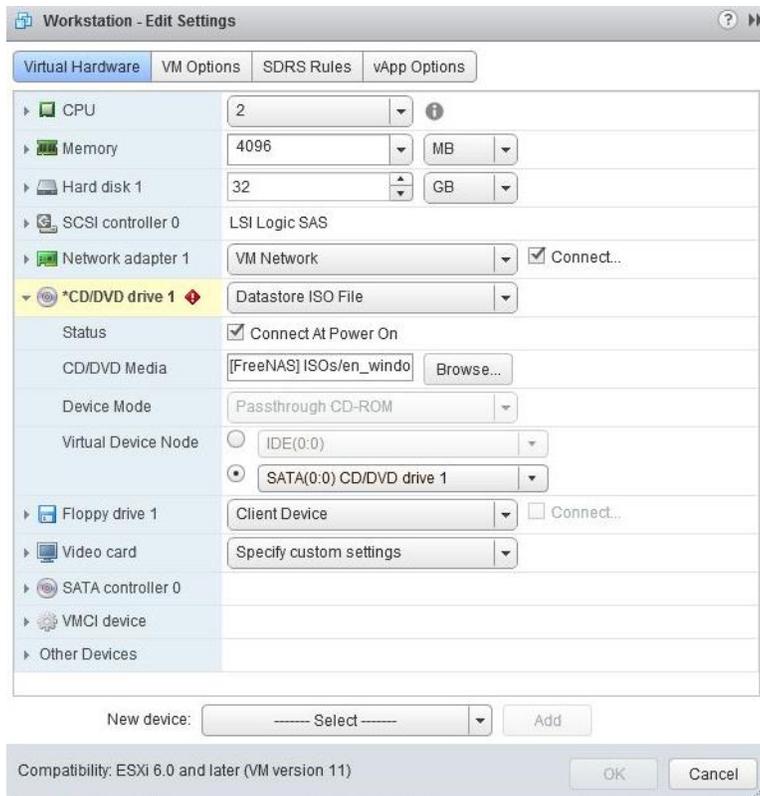
Use the following procedure to install Windows on the virtual machine.



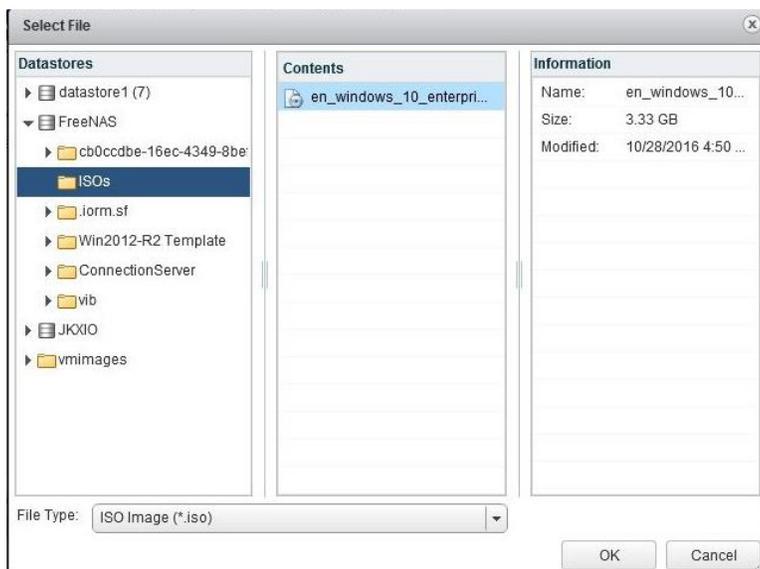
1. Select the virtual machine, click on the **Manage** tab and select **Settings**. Click on the **Edit** button.



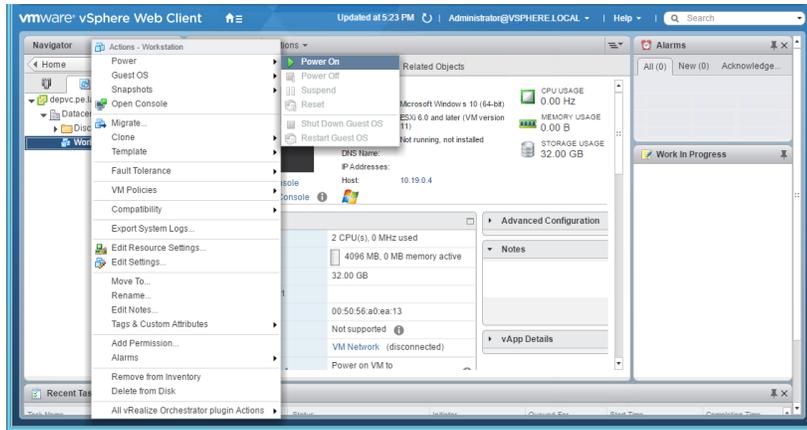
2. Locate the **CD/DVD drive 1** entry under the **Virtual Hardware** tab. Select the arrow drop down to reveal data sources for the CD/DVD media. (In this example a datastore ISO file will be used.) Check the **Connect** checkbox for **CD/DVD drive 1**.



3. Toggle the carrot next to the **CD/DVD drive 1** icon to reveal the details of the virtual device. For Status check the **Connect At Power On** checkbox. This will connect the ISO file to the VM's virtual CD/DVD drive during boot up.

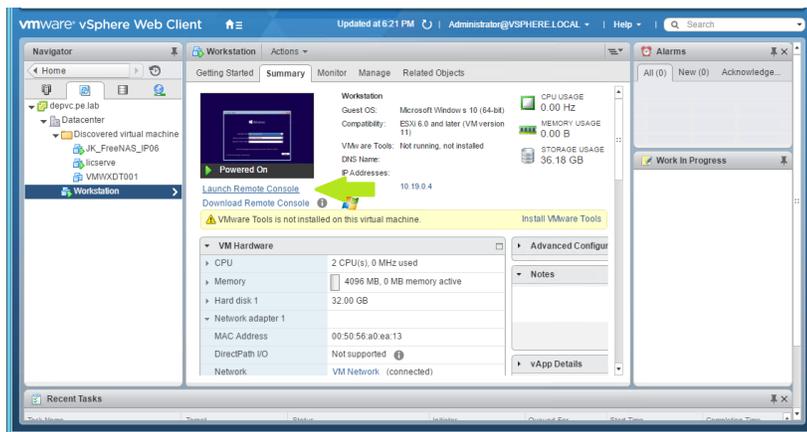


- Click on the **Browse** button for the CD/DVD Media. Navigate to and select the OS ISO file to be used. Click **OK** to select the ISO.

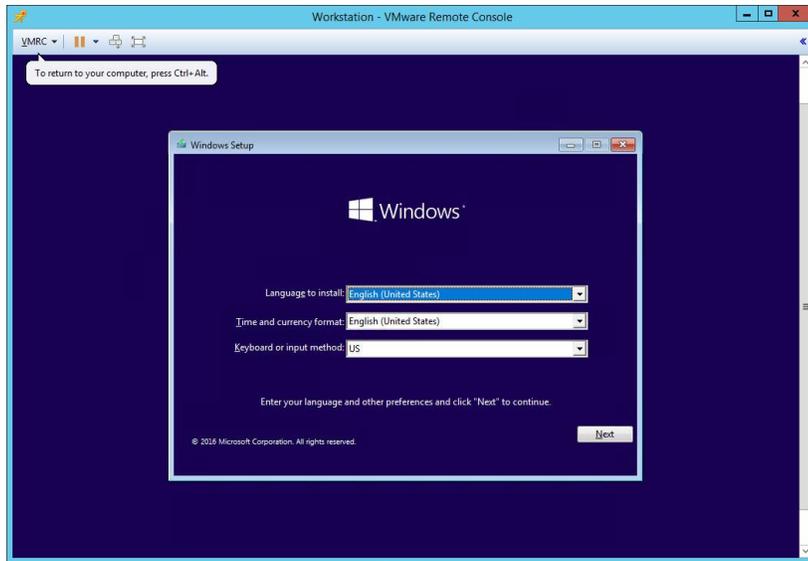


- Right-click the virtual machine, and then select **Power>Power On** to start the virtual machine and boot from the .ISO to install the operating system.

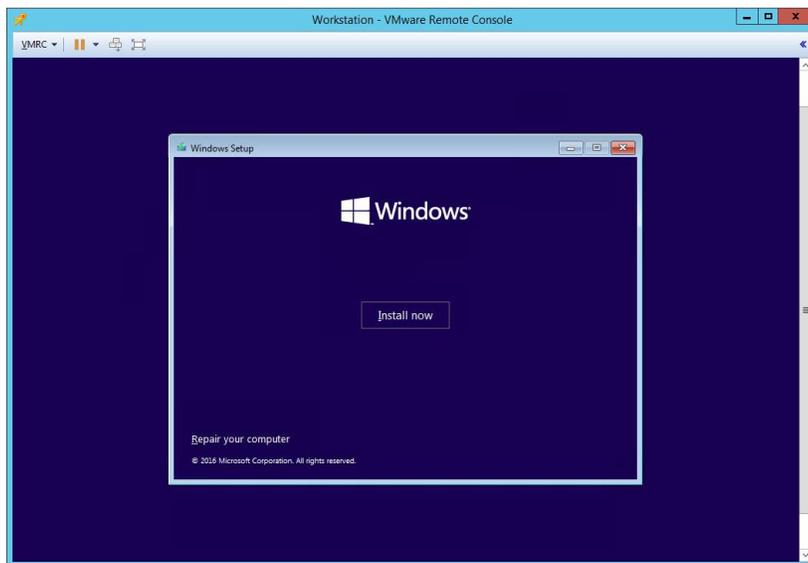
If you are creating a new virtual machine and using the vSphere Web Client's VM console functionality, then the mouse may not be usable within the virtual machine until after both the operating system and VMware tools have been installed.



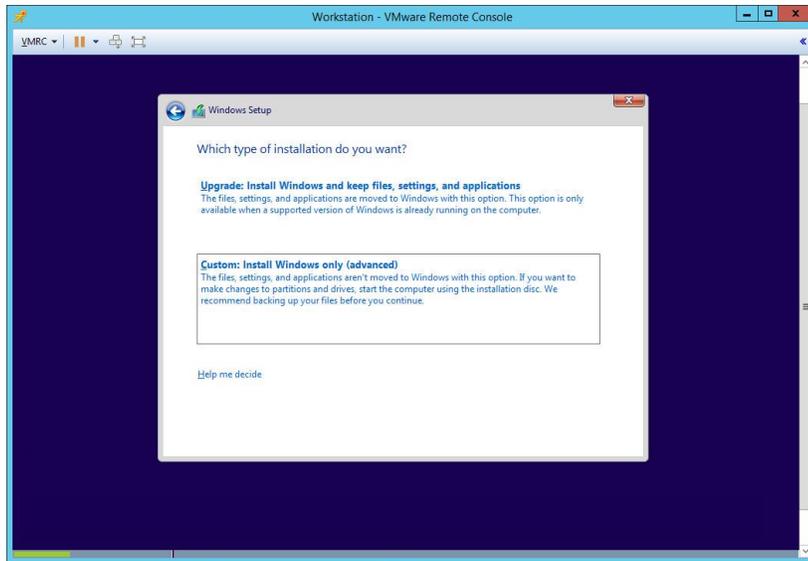
- Click on **Launch Remote Console**.



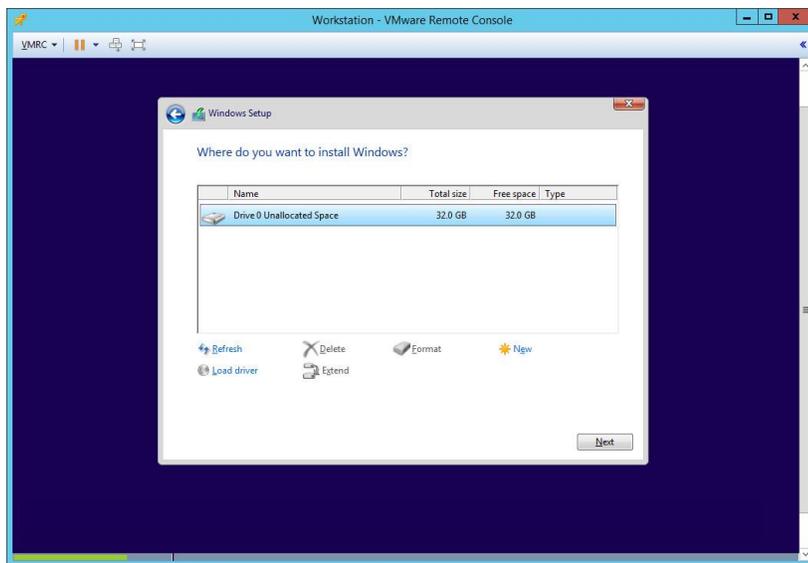
7. At the Windows Setup screen, enter **Alt+n**



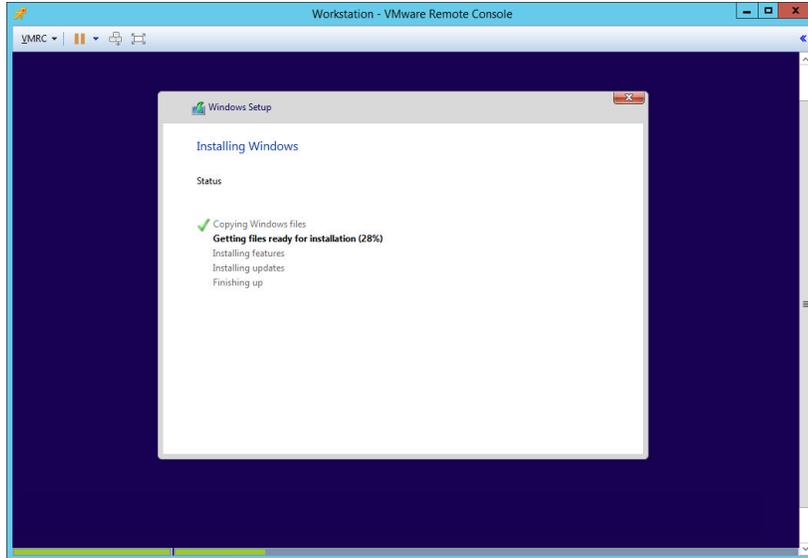
8. Enter **Atl+i** to install now.



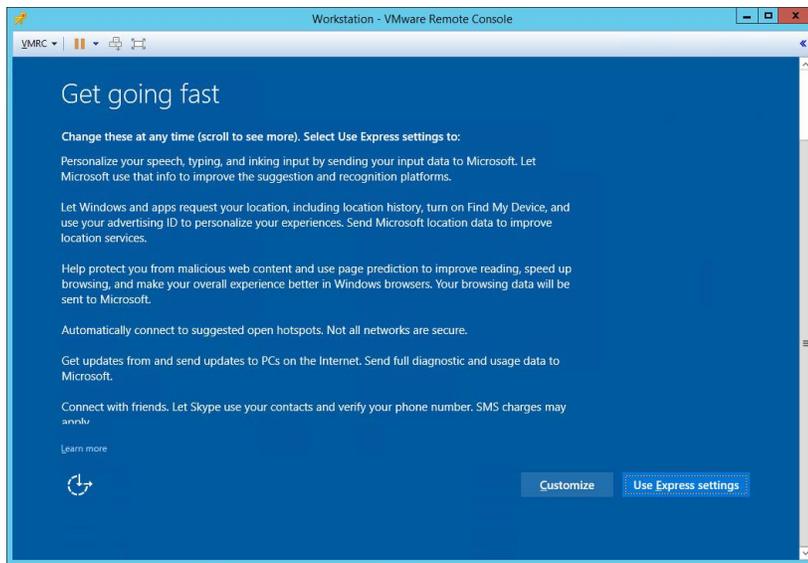
9. Use the arrow keys to select **Custom: Install Windows only (advanced)**. Press **Enter**.



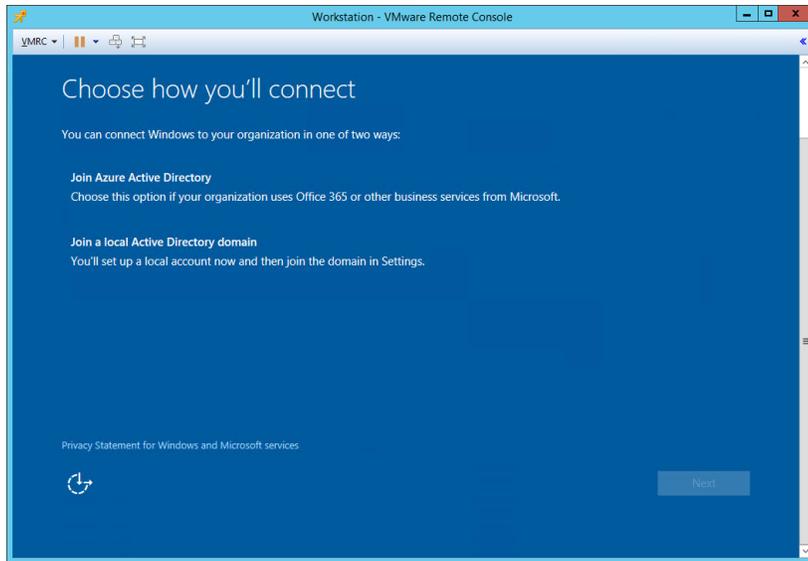
10. Select Drive 0 and press **Alt+n** to continue.



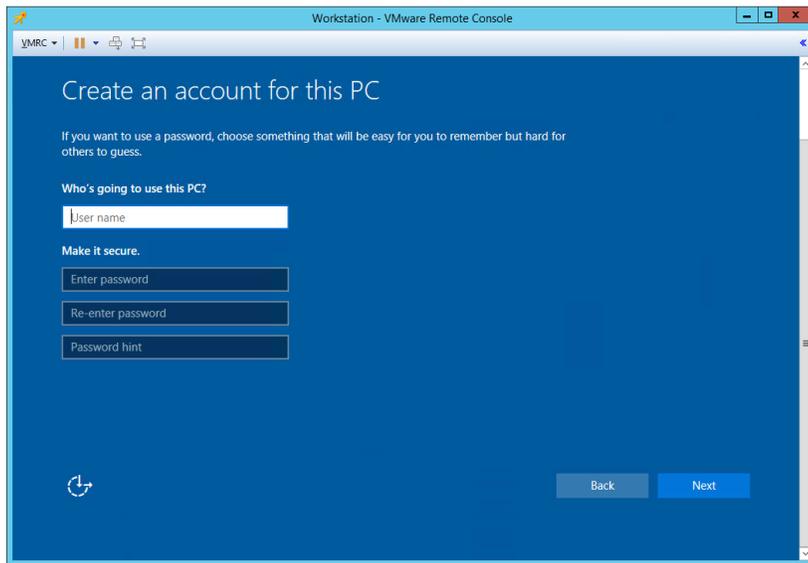
11. During installation, Windows reboots the VM several times.



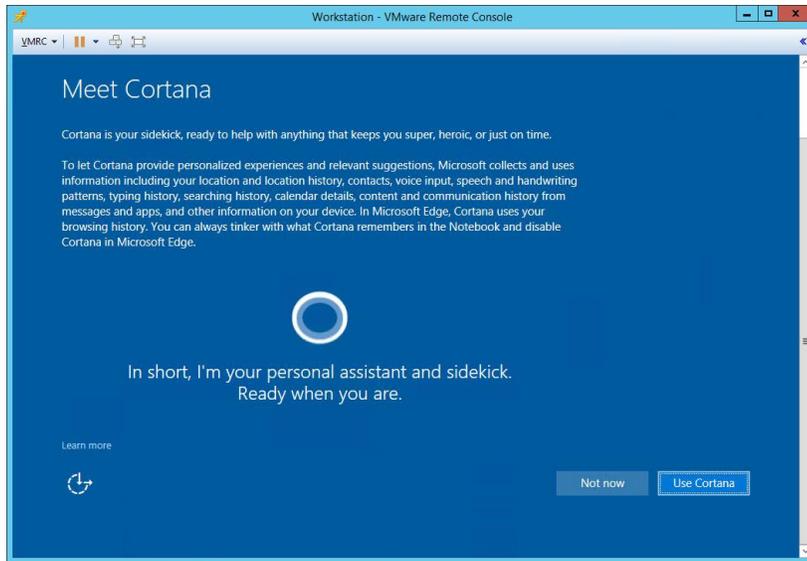
12. At the **Get Going Fast** window, press **Alt+e** to use express settings.



13. At the **Choose how you'll connect** window, use the **TAB** and **DOWN ARROW** keys to maneuver to **Join a local Active Directory domain**. Press **ENTER**. Tab to the **Next** button and press **Enter**.



14. At the **Create an account for this PC** window, use the **TAB** key to enter a user name, password, re-enter password, and password hint. Tab over to **Next** and press **Enter**.



15. At the **Meet Cortana** window, press **Enter** to use Cortana.



16. Disconnect the .ISO from the VM when Windows is done installing.

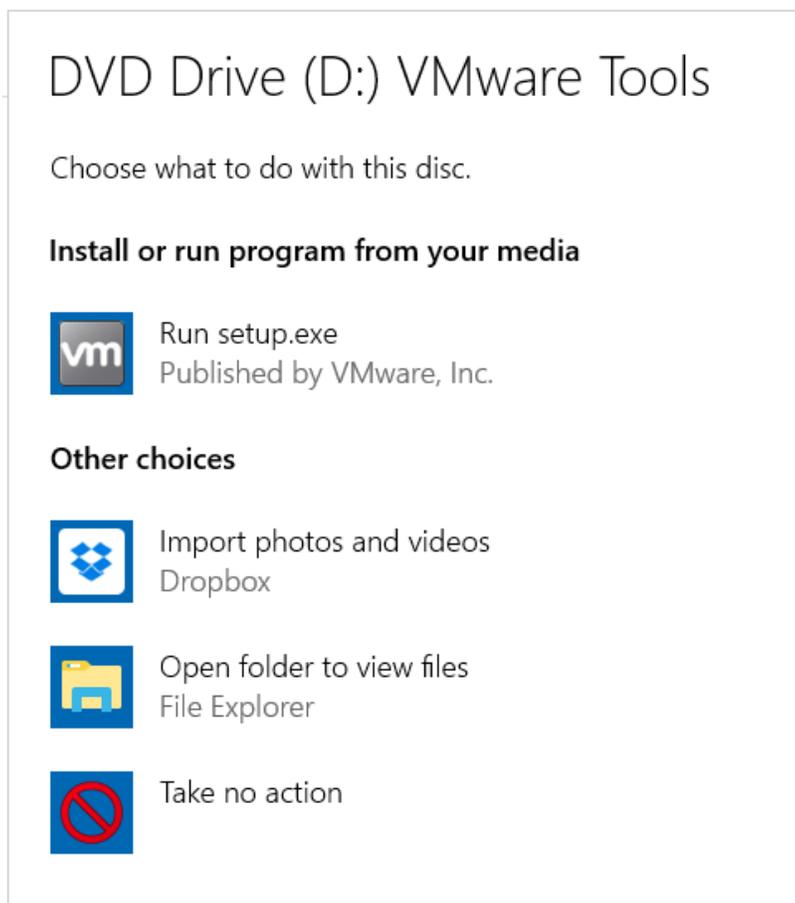
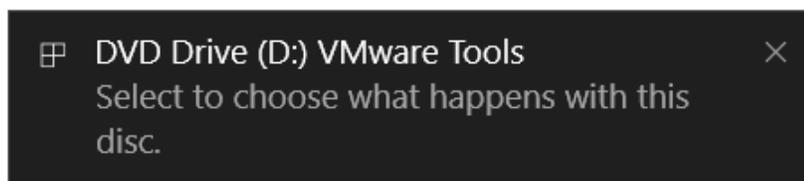
Windows 10 is now installed in the virtual machine.

10.3 INSTALLING VMWARE TOOLS ON THE VIRTUAL MACHINE

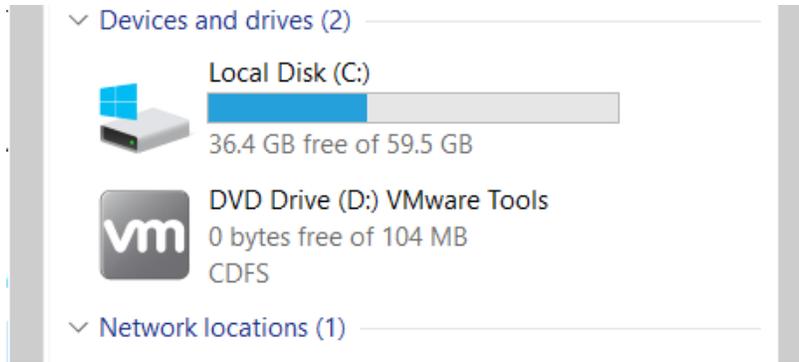
After Windows completes the initial installation and configuration process, the next step is to install VMware Tools on the virtual machine.



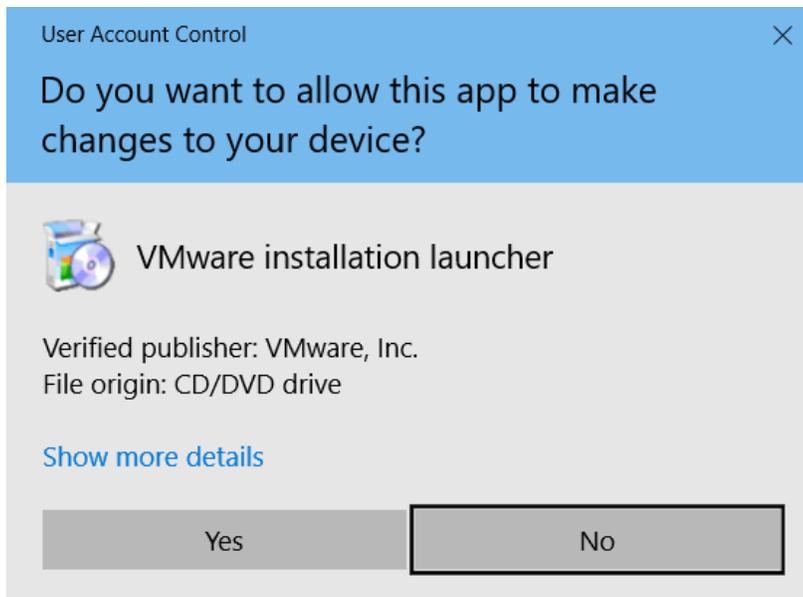
17. Select the **Summary** tab from the virtual machine console. Click the **Install VMware Tools** link in the yellow bar.



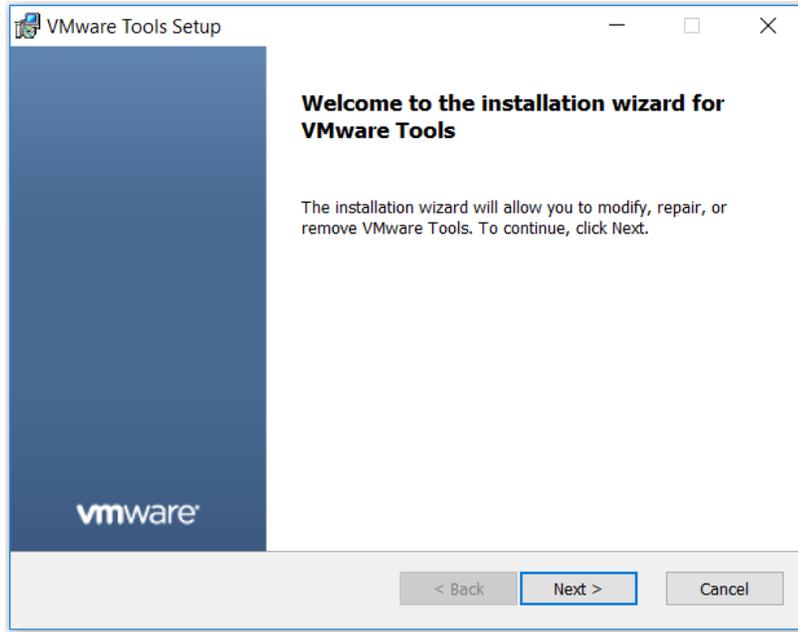
18. Back in the virtual machine console, Windows 10 detects the CD image and notifies the user. Click on the notification window. In the popup window, **click Run setup.exe**.



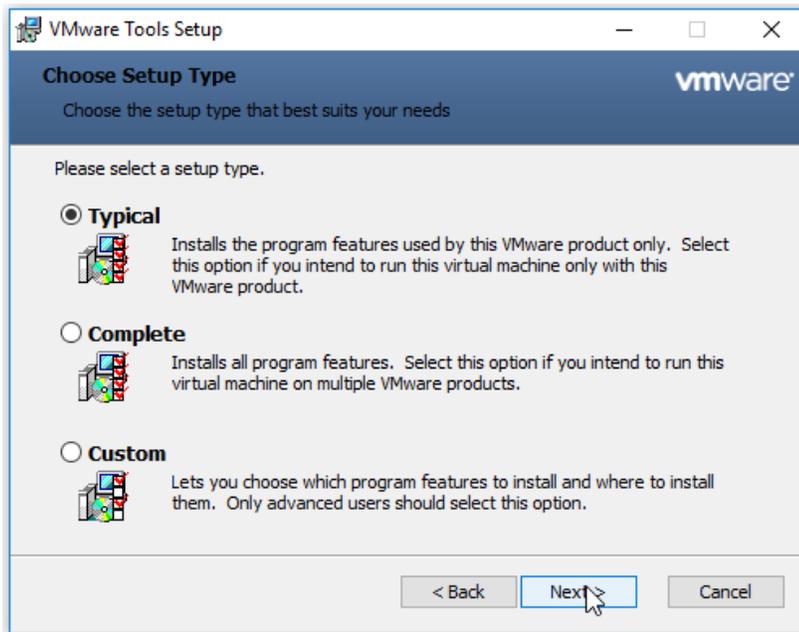
19. If the window does not display (or you miss it) then open Explorer, locate the CDROM, and double-click it.



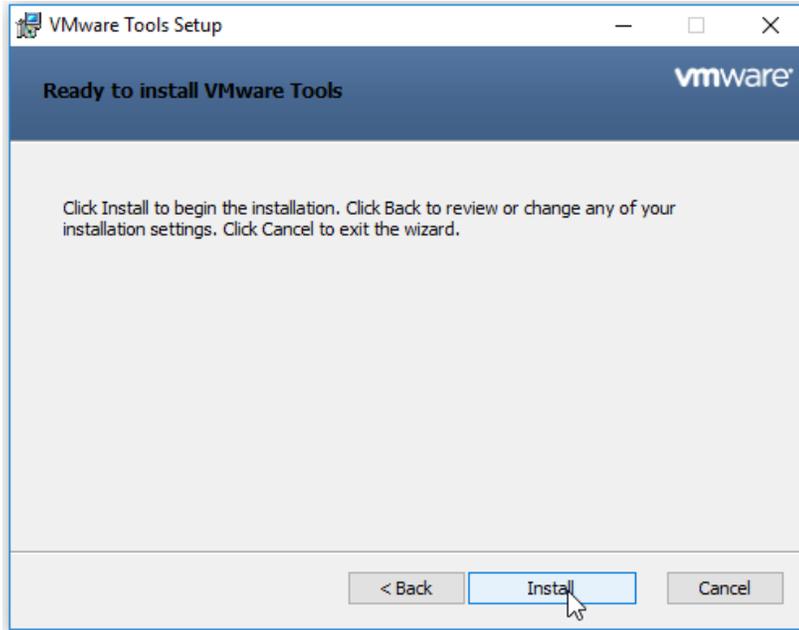
20. The **User Account Control** popup may display. If so, click **Yes**.



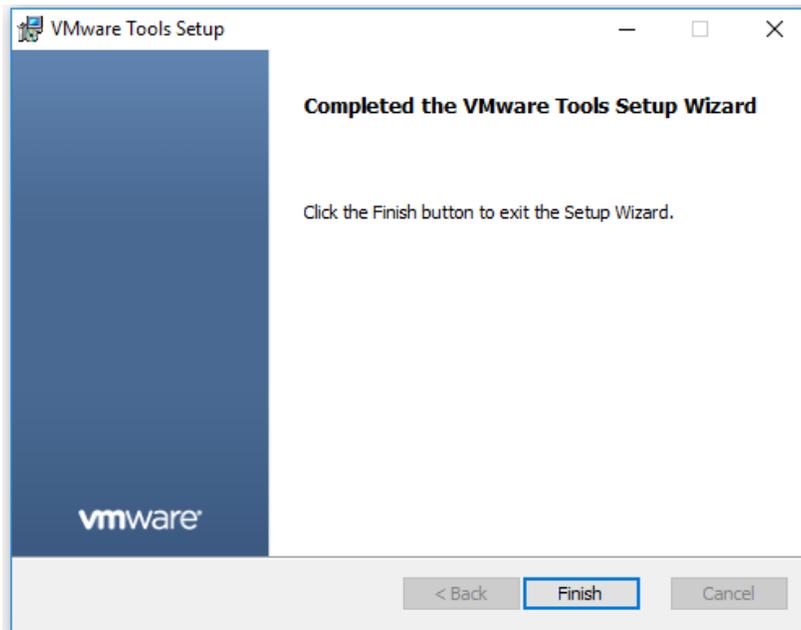
21. The installer begins, click **Next**.



22. When prompted, select **Typical** installation, click **Next**, and accept all defaults.



23. When prompted, select **Install** to begin installation.



24. Click **Finish** in the VMware Tools installer and reboot the virtual machine when prompted.

! **CAUTION:** This reboot is critical to ensure the tools and drivers are loaded.

VMware Tools are now installed in the virtual machine

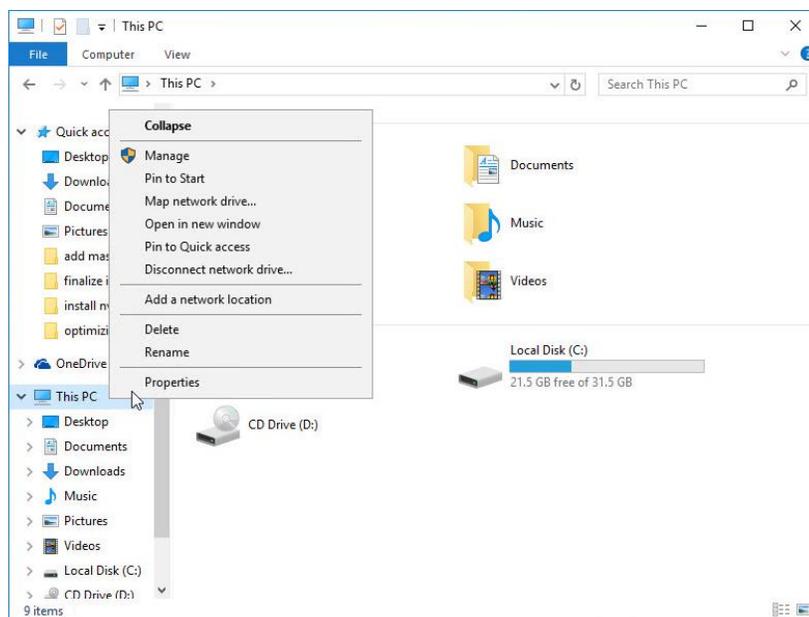
10.4 ADDING THE GOLDEN MASTER TO THE DOMAIN

By joining the VM to the Windows Active Directory domain you are then able to manage it as you would any physical desktop in the domain.

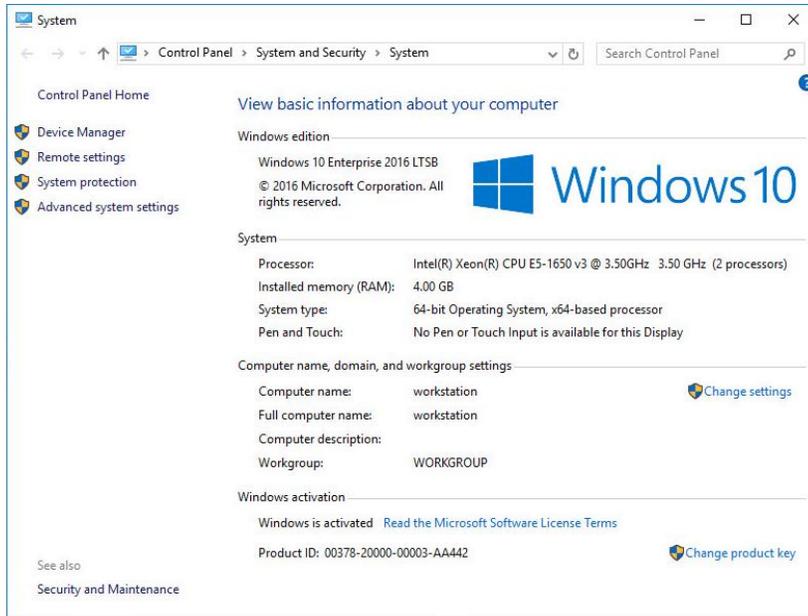
Customize Windows on the virtual machine as follows:

- ▶ Join the domain
- ▶ Add appropriate Domain groups to Local Administrators

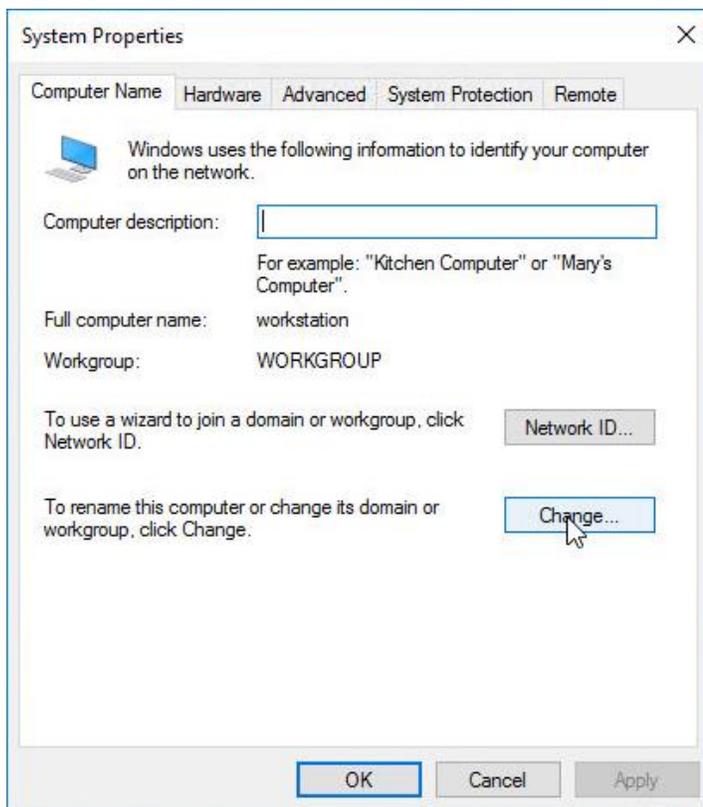
Use the following procedure to add the VM to the domain.



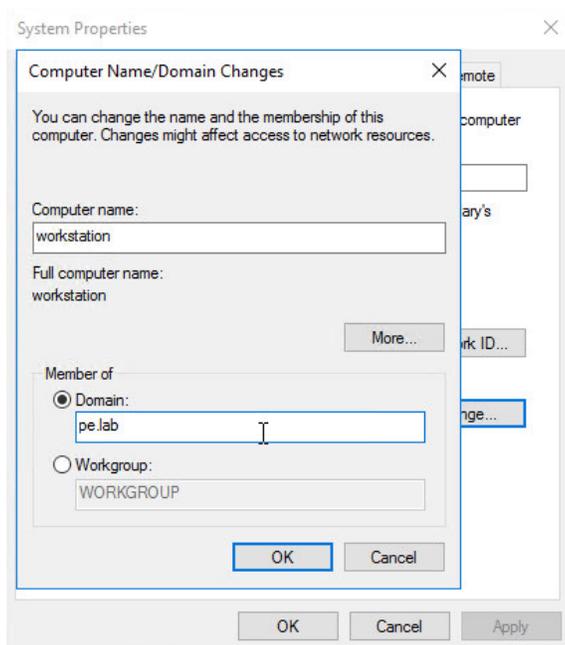
1. In the VM, click **Start**, then right click on **Computer** to bring up its menu, then click **Properties**.



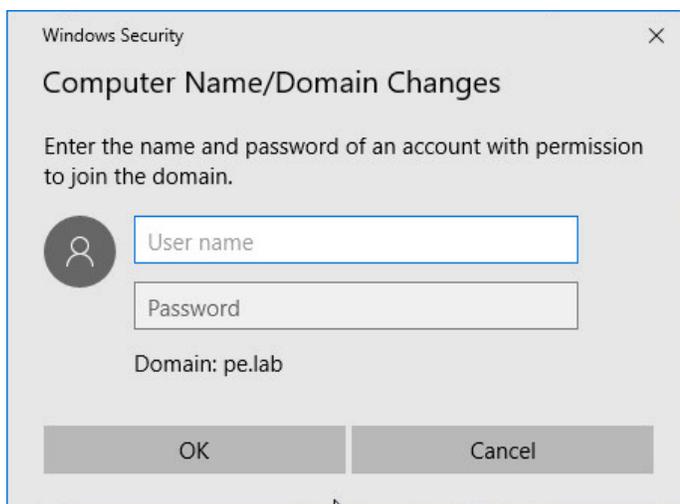
2. On the **Properties** window, click **Change settings**.



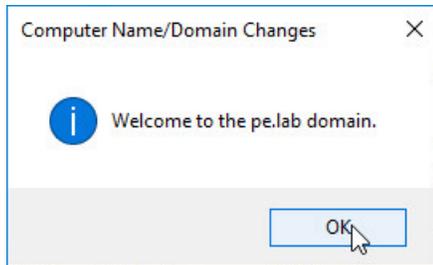
3. This brings up the **System Properties** window, on the **Computer Name** tab click **Change**.



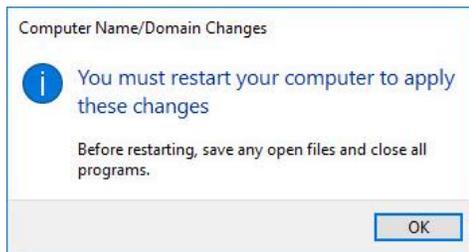
4. On the **Computer Name/Domain Changes** window, enter an appropriate computer name, then domain name, and click **OK**. Our chosen naming is shown below, use what is appropriate for your POC.



5. A security window pops up, fill in your specific domain administrator credentials and click **OK**.



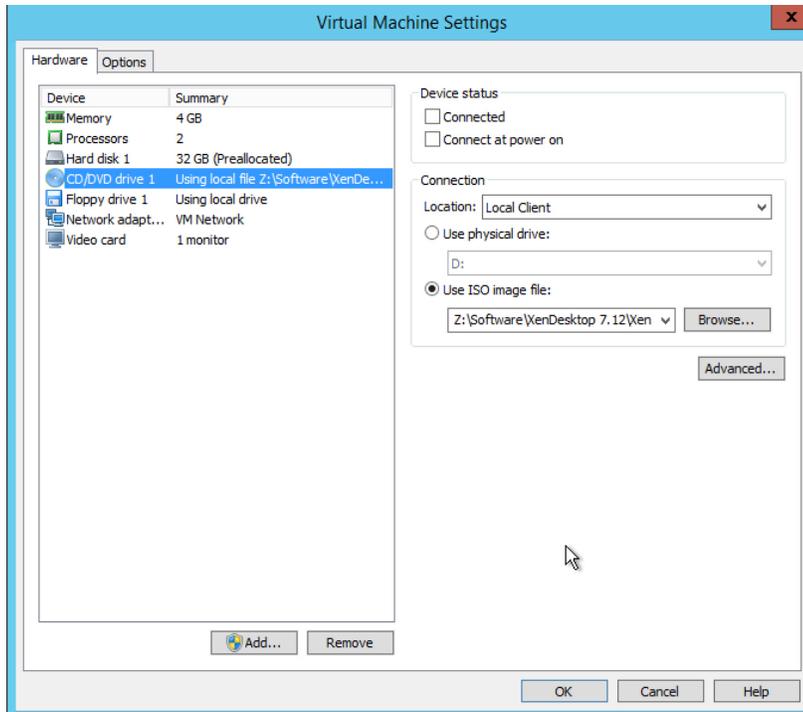
6. On successful authentication, you will see the following welcome pop-up showing your VM is now in the domain (the domain name should reflect your domain information). Click **OK**.



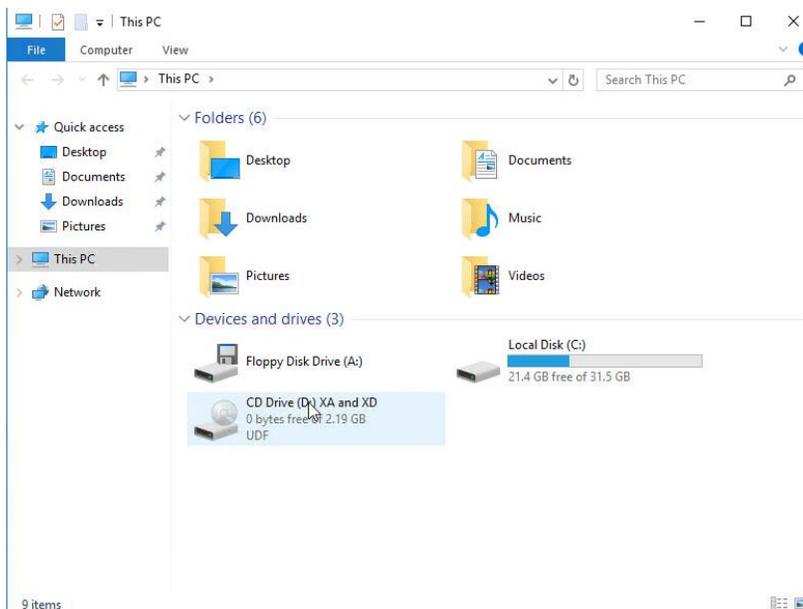
7. The VM needs to reboot to complete the process, click **OK** again and the VM reboots immediately.

10.5 INSTALLING CITRIX VIRTUAL DELIVERY AGENT

For the virtual machines to be managed by the XenDesktop delivery controller, a small piece of code called Virtual Delivery Agent needs to run on each desktop machine. Use the following procedure to install Citrix Virtual Desktop Agent on the master image.



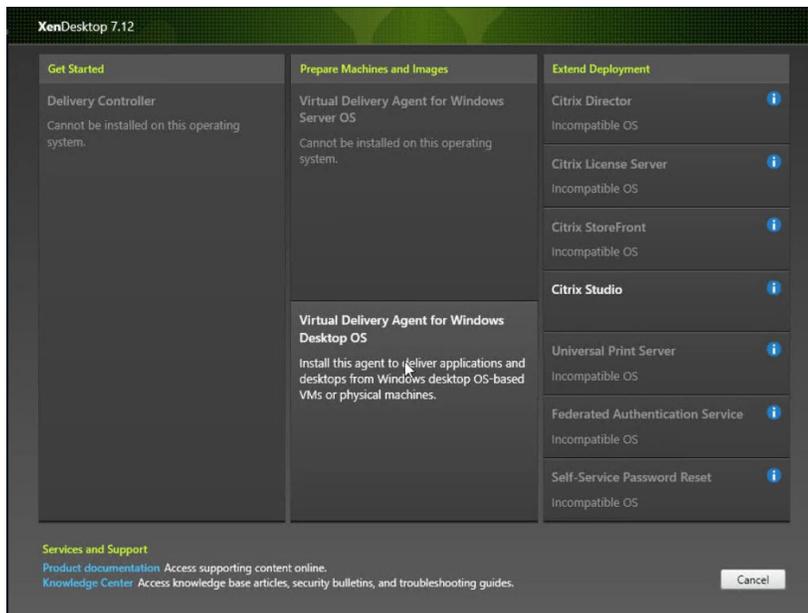
1. Mount the .iso image using Virtual Clone Drive or similar. Alternately you can use vCenter Client to mount the .iso from an ISO datastore to the virtual machine CD drives.



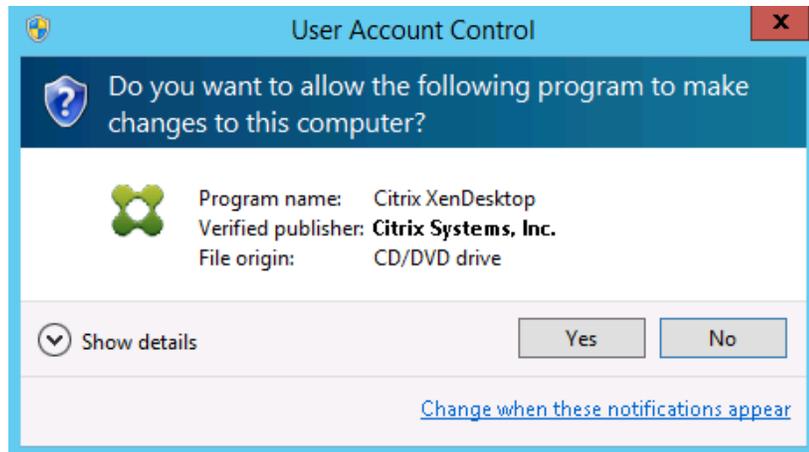
2. Launch the CD with AutoPlay or select the **AutoSelect.exe** application to start the Citrix XenApp and XenDesktop installer.



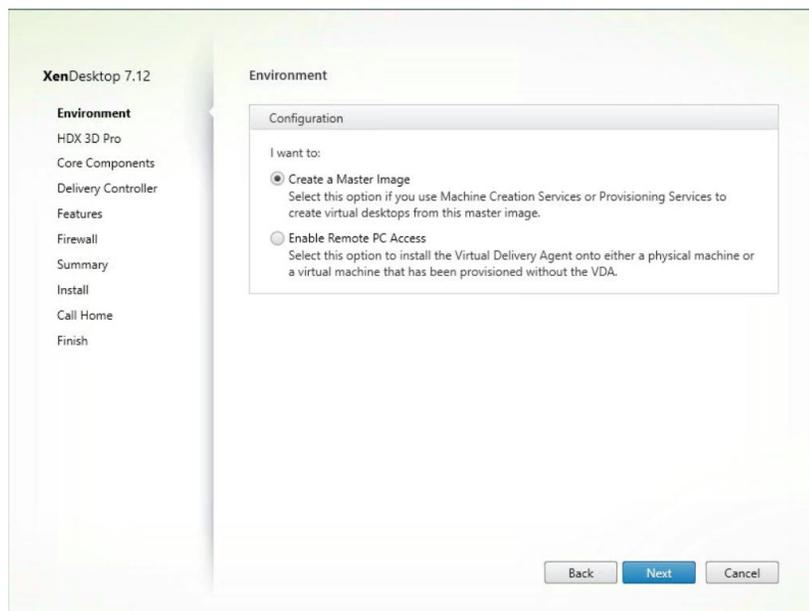
3. Select **Start** button next to XenDesktop Deliver applications and desktops.



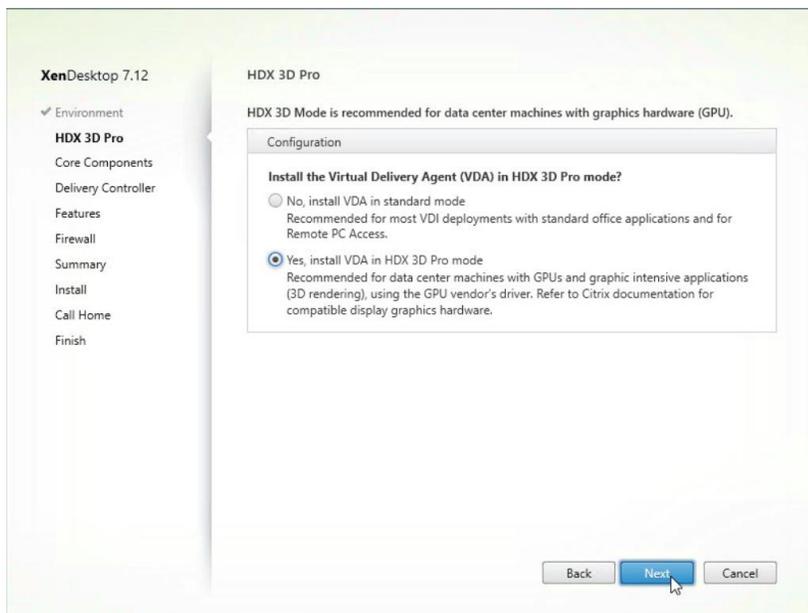
4. Select **Virtual Delivery Agent for Windows Desktop OS** under **Prepare Machines and Images** on the XenDesktop 7.12 window.



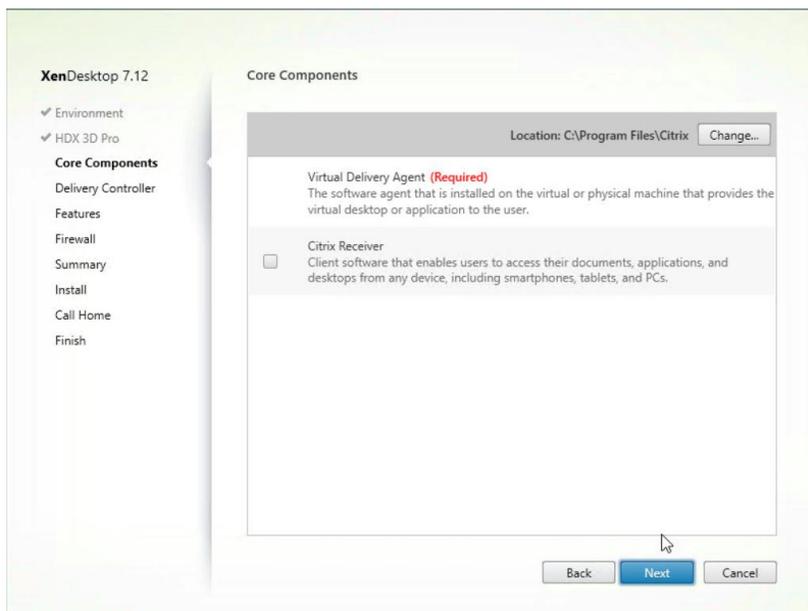
5. Select **Yes** to agree to the User Account Control prompt if displayed.



6. Accept the default **Environment** value of **Create a Master Image** radio button, and select **Next** to continue.

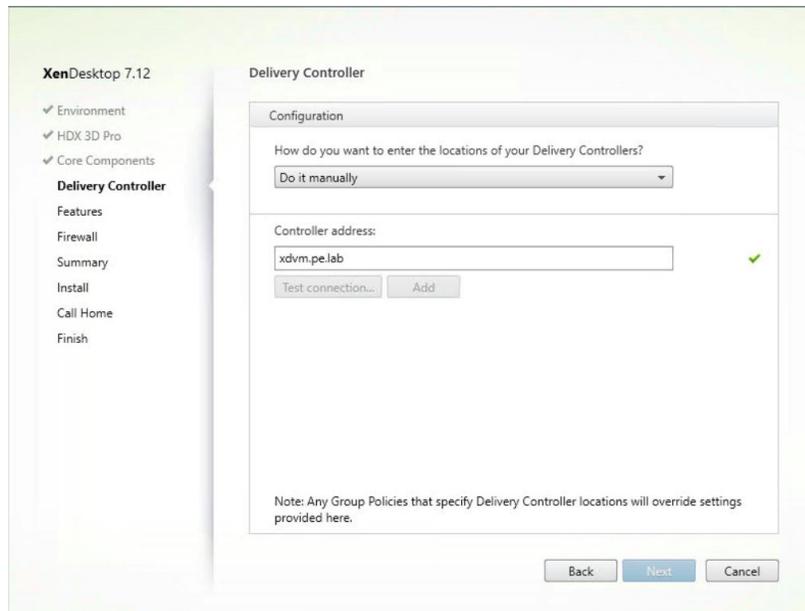


7. Select the **Yes, install the VDA for HDX 3D Pro** radio button on the HDX 3D Pro Dialog, and select **Next** to continue.



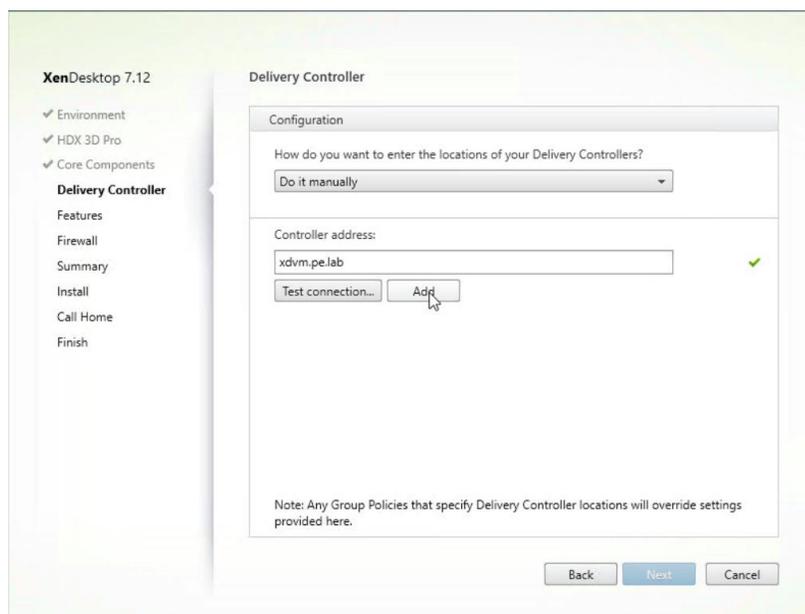
8. Deselect the **Citrix Receiver** check box on the **Core Components** dialog, and select **Next** to continue.

 Citrix Receiver can be installed separately later and would only be used if connecting to another XenApp or XenDesktop Store.

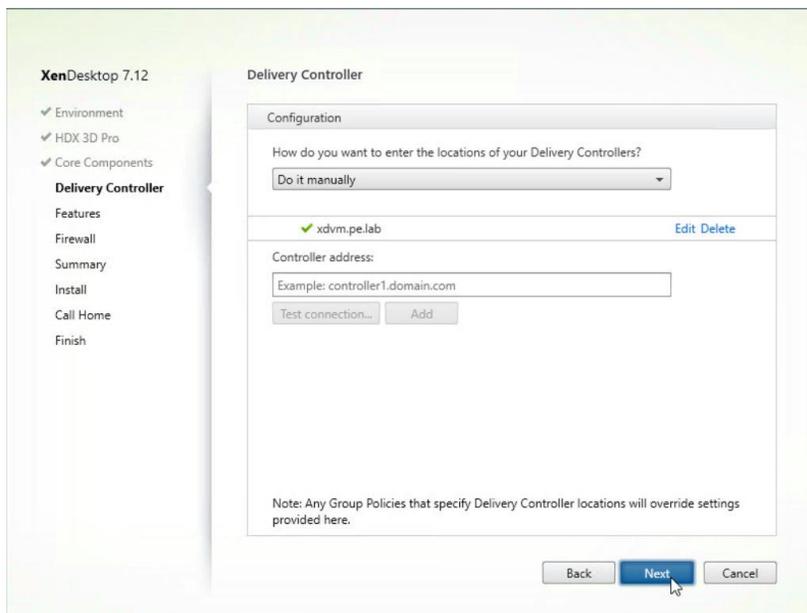


9. Enter the XenDesktop delivery controller address. This is the FQDN of the XenDesktop server. Click on the **Test connection...** button to validate the address.

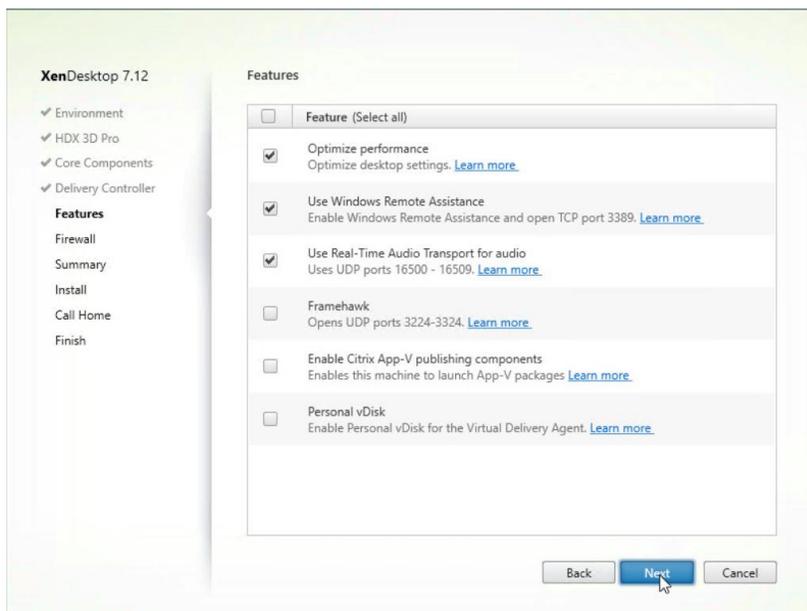
IP Addresses are not supported as a controller address, only Fully Qualified Domain Names are allowed.



10. A green checkmark will be displayed on the right of the controller address input box if the FQDN address of the delivery controller is verified. Click the **Add** button to confirm the delivery controller address and add this address as the Virtual Delivery Agent target.

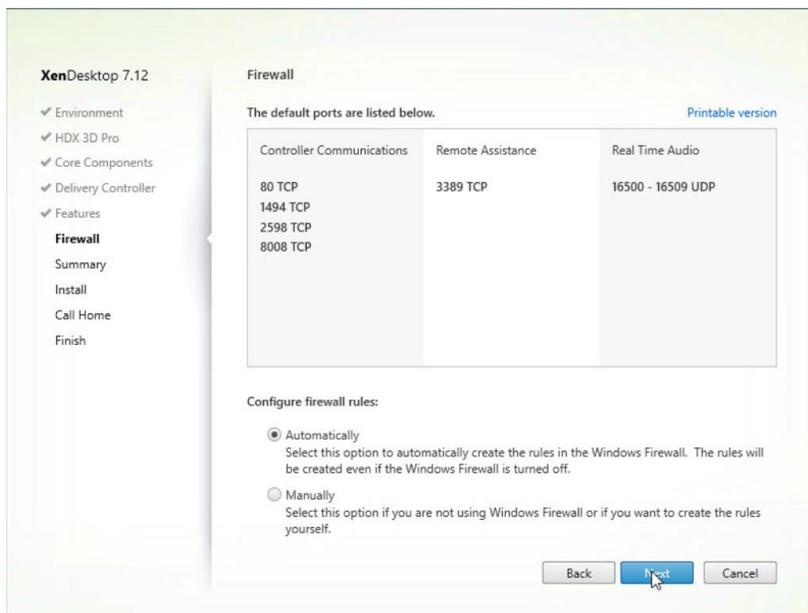


11. The delivery controller FQDN will be displayed above the delivery controller address entry box once it has been added. Select **Next** to continue.

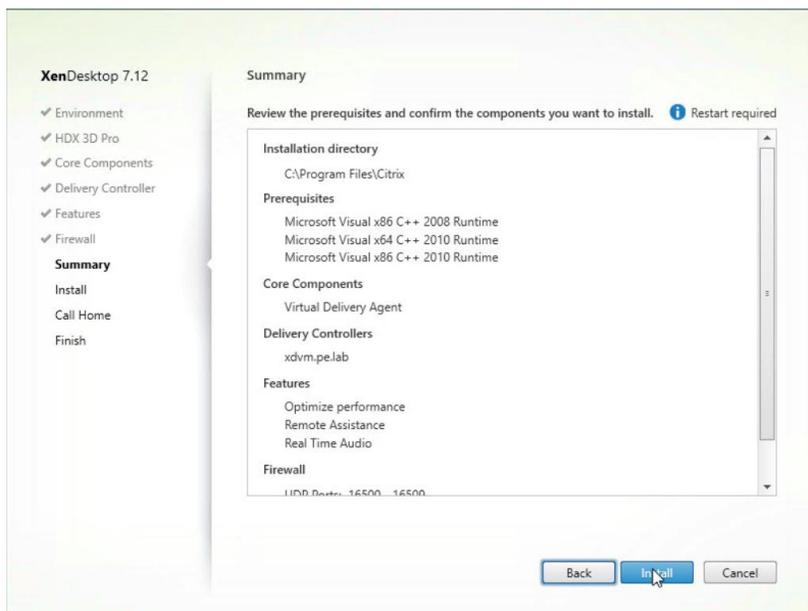


12. **Optimize performance** should be selected by default. Select **Use Windows Remote Assistance** and **Use Real-Time Audio Transport for audio** on the **Features** dialog and select **Next** to continue.

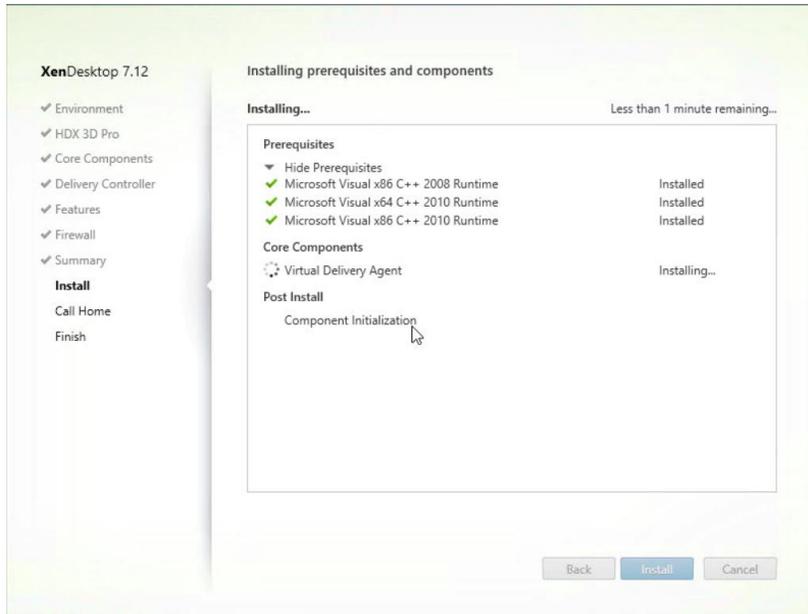
 Personal vDisk can be enabled at a later date if desired



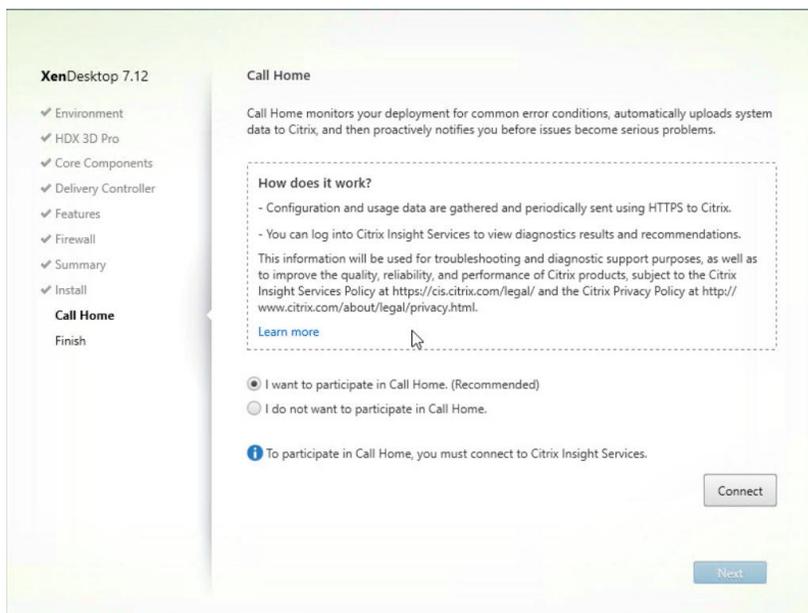
13. Accept the default value to automatically create the Firewall entries. Select **Next** to continue.



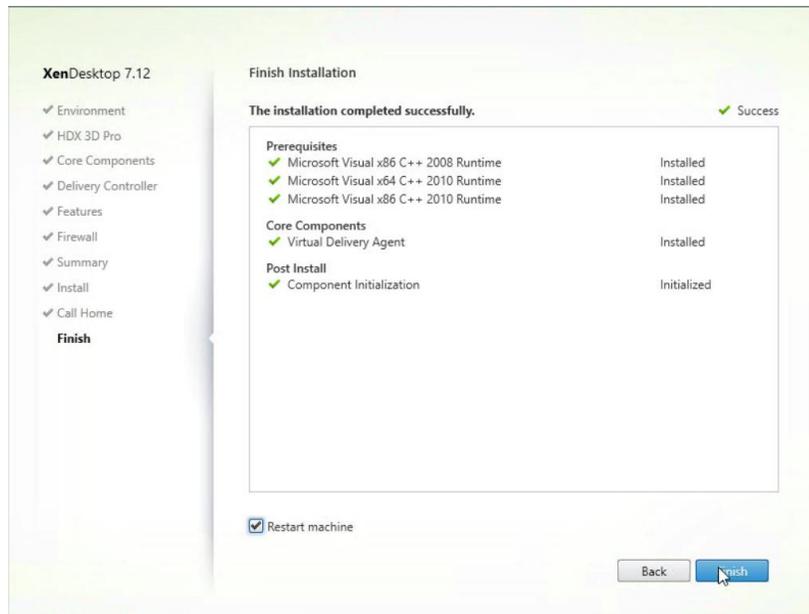
14. Review the summary and select **Install** to continue.



15. Installation will progress.



16. At the **Call Home** dialog box, choose whether to participate in Citrix Call Home. Click **Next**.



17. When the installation has completed successfully accept the default to restart the machine and select **Finish**.

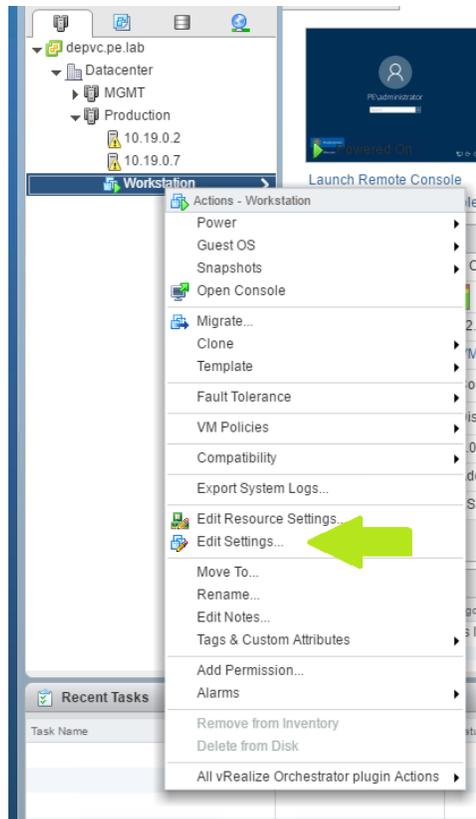
10.6 ENABLING THE NVIDIA GRID vGPU

Use the following procedure to enable GRID vGPU for your virtual machine. This will require you to edit the virtual machine settings.

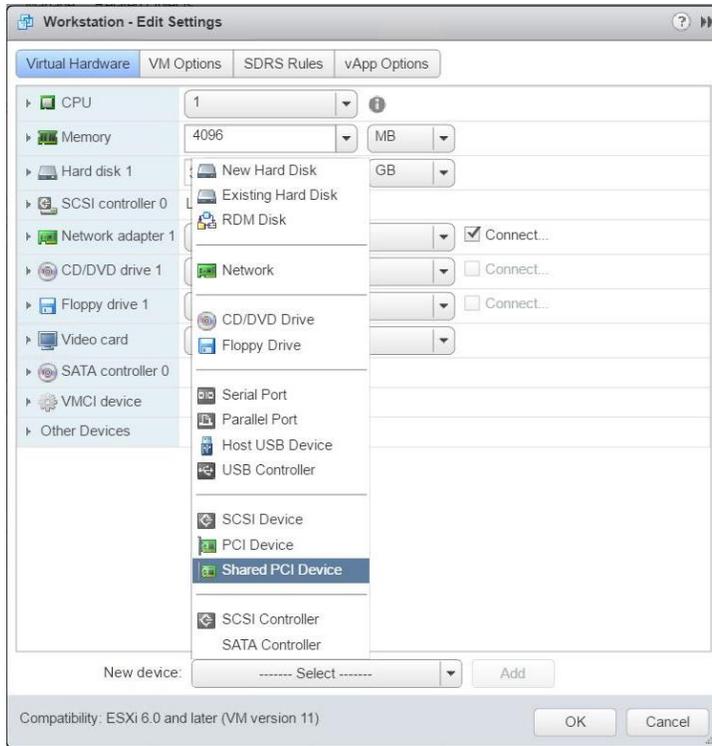
1. Close the remote console; this will not be functional when GRID vGPU is configured.

 Take a snapshot of the virtual machine to preserve your work. Label this snapshot pre-vGPU, and revert back to it if you encounter any problems going forward, such as driver issues

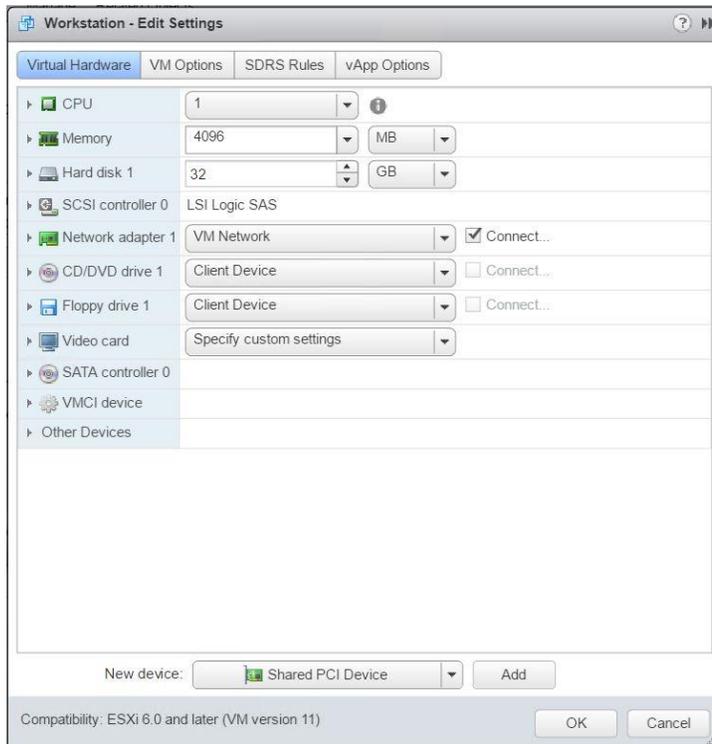
2. Power down the virtual machine.



3. Right click on the VM in the Navigator window. Select **Edit Settings**. The **Edit Settings** dialog appears.



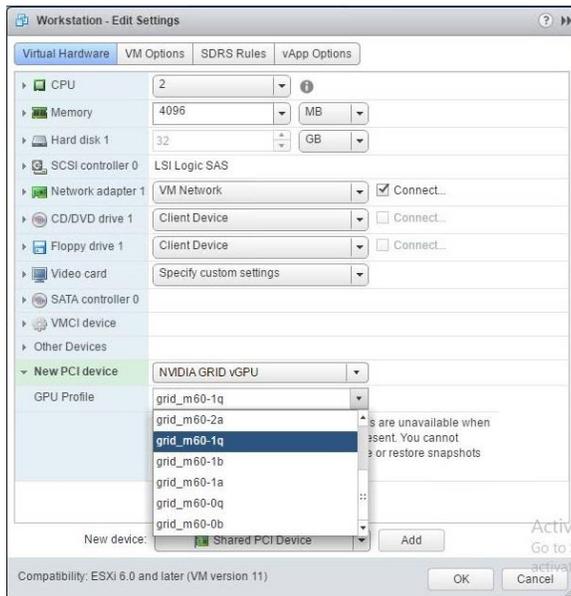
4. Click on the **New Device** bar and select **Shared PCI device**.



5. Click on **Add** to continue.



- The new PCI device will show the **NVIDIA GRID vGPU** device has been added.



- Click the down triangle on the **GPU Profile** selection bar to drop down the list of available vGPU profiles. Select the appropriate profile for this VM.

For information about what profile to choose, see *Selecting GRID vGPU profiles* on page 176

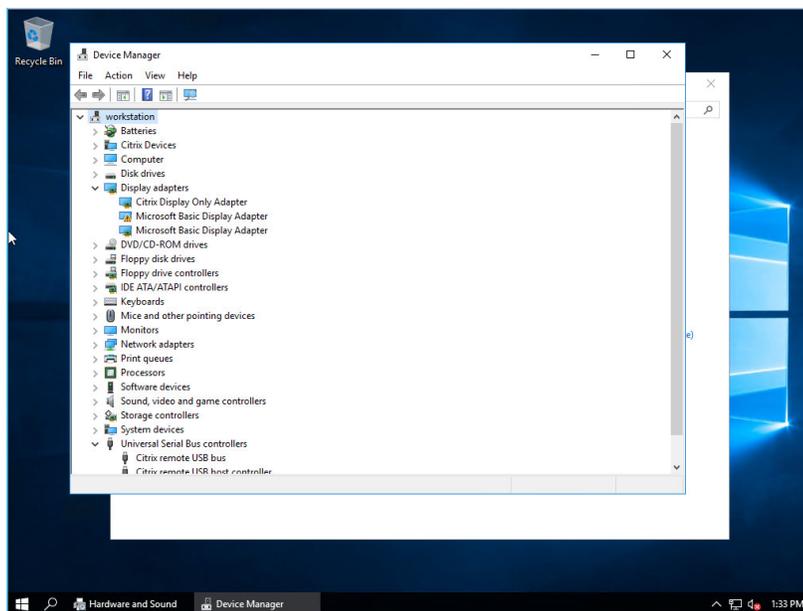
- Click **OK** to complete the configuration.

10.7 INSTALLING NVIDIA DRIVER IN THE VM

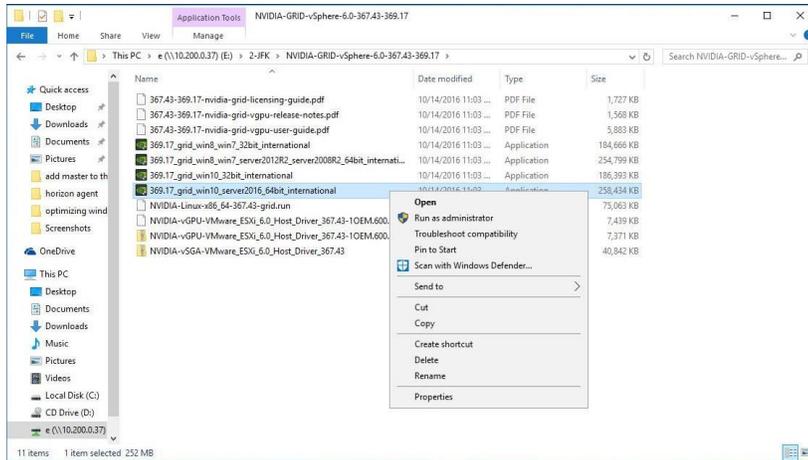
1. Start the virtual machine, and then connect to it using VMware Remote Console through the vSphere Web Client (*When connected, a popup warning requesting that you restart the computer to apply changes will display the first time it is booted after enabling the NVIDIA GRID vGPU.*)
2. Click **Restart Later** to continue booting the virtual machine.



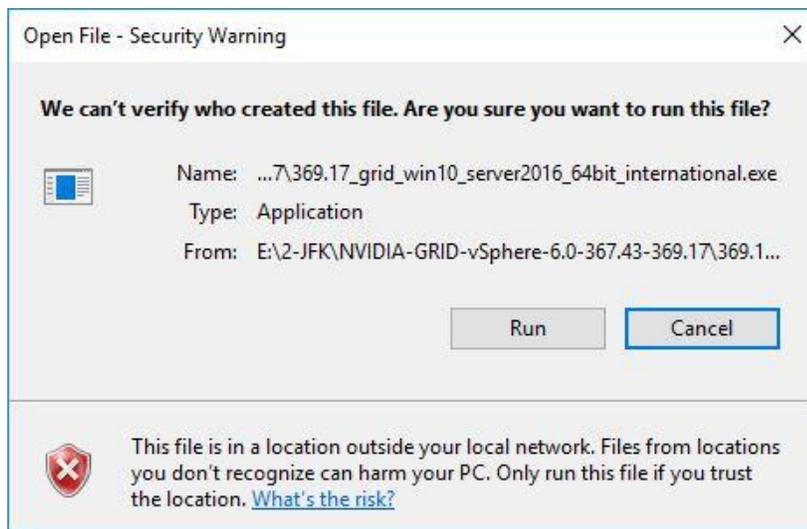
DO NOT REBOOT THE VIRTUAL MACHINE IF YOU HAVE OLDER NVIDIA DRIVERS INSTALLED. DOING THIS WILL CAUSE A BLUE SCREEN ERROR.



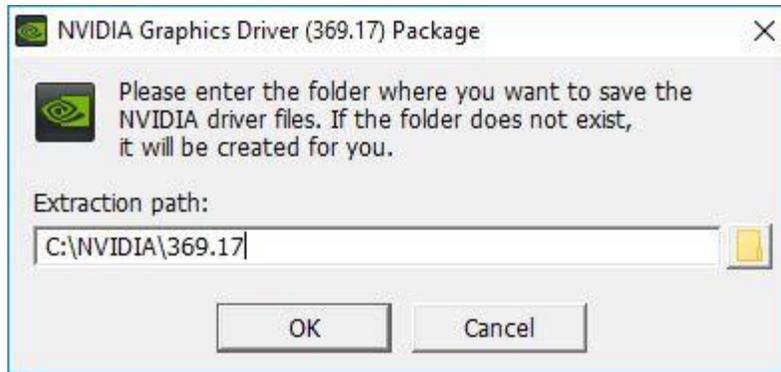
3. Log into Windows and open the Device Manager. The **Microsoft Basic Display Adapter** displays in the **Display adapters** section of the Device Manager with an exclamation point by it to indicate a driver problem. This is normal.



4. Locate the NVIDIA driver and double-click the **Setup** icon to launch it. (Recommendation: Have the software on a shared volume that can be mounted by the VM for quick access).



5. Accept the Security Warning to continue install.



6. Click **OK** to install the files to the default location.



7. The NVIDIA software license agreement window displays. Click the **AGREE AND CONTINUE** button to proceed.



8. The **Installation Options** window displays. Check the **Custom (Advanced)** radio button, then click **Next**. The **Custom installation options** window appears.



9. Check the **Perform a clean installation** checkbox, and then click **Next**.



10. A window displays when the NVIDIA Graphics Driver installation is complete. Click **RESTART NOW** to reboot the VM and complete the installation.

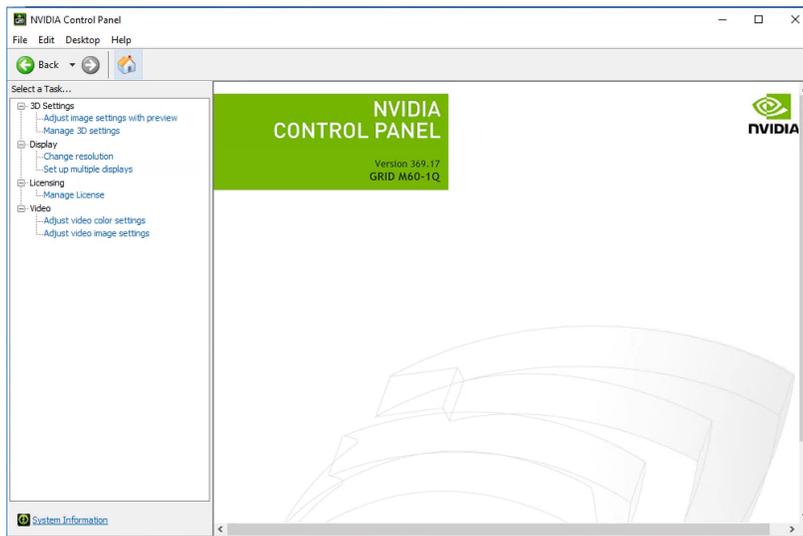
10.8 LICENSE THE DESKTOP

To manage licenses for each desktop, the FQDN or IP address of the GRID License Manager must be entered in the VM. When the VM boots, it will obtain a short-term lease for a license from the License Manager. Upon expiration, the NVIDIA driver software will automatically renew the lease, if the VM is running. If the VM shuts down, the lease will be renewed on VM startup.

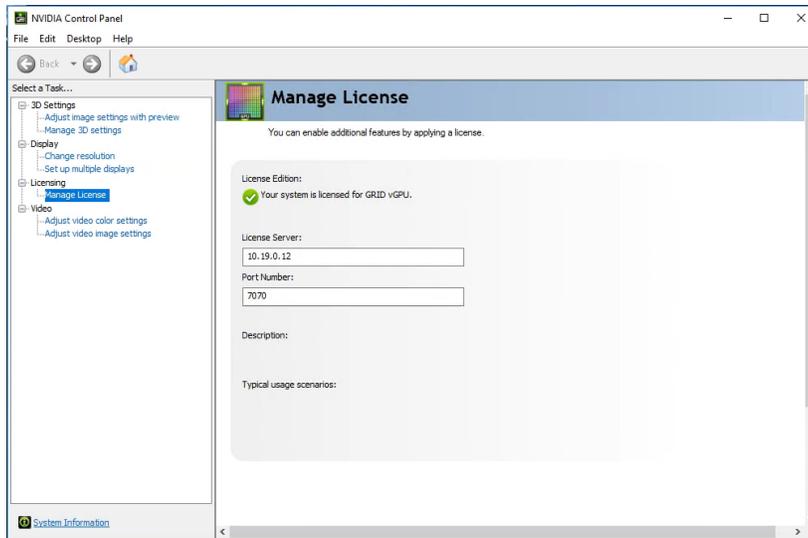
Use the following procedure to license the VM.



1. Right-click on the screen and select the **NVIDIA Control Panel** menu option.



2. The **NVIDIA Control Panel** will indicate which GRID Profile is loaded.



3. Click on **Manage License** in the left-hand navigation pane. Enter the IP address of your License Server, and default port 7070. Click **Apply**.

The VM will now contact the License Manager for updated licenses upon reboot.

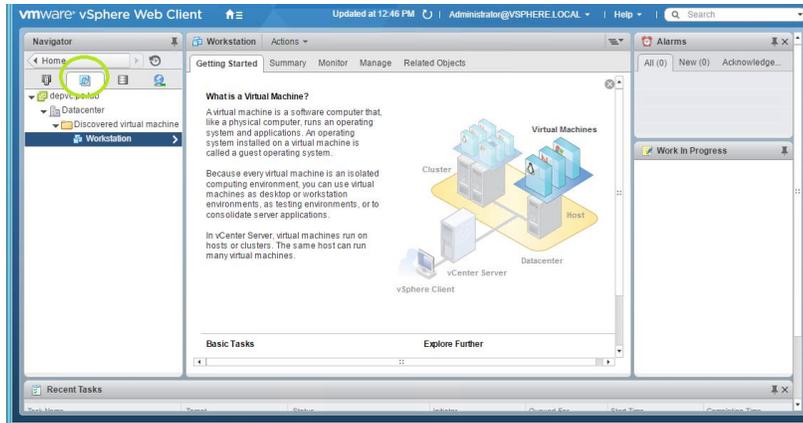
10.9 REMOVING THE GRID vGPU DEVICE FROM THE MASTER IMAGE

Use the following procedure to remove the vGPU shared PCI device from the master image before provisioning the XenDesktop machine catalog.

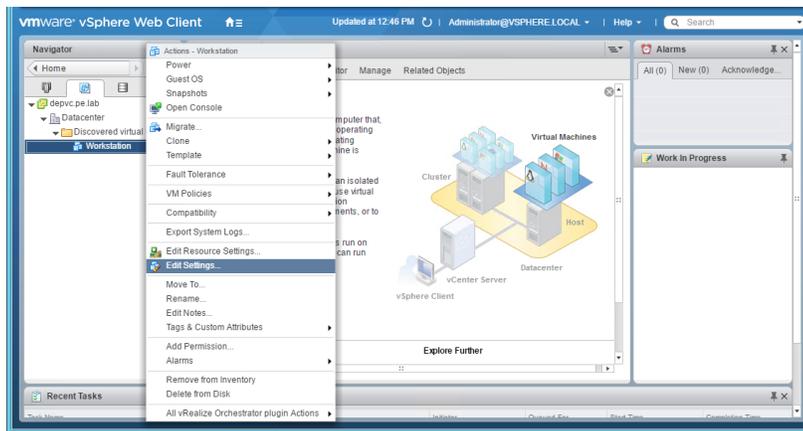


CAUTION: Not performing this procedure will cause device conflicts between vSphere and XenDesktop.

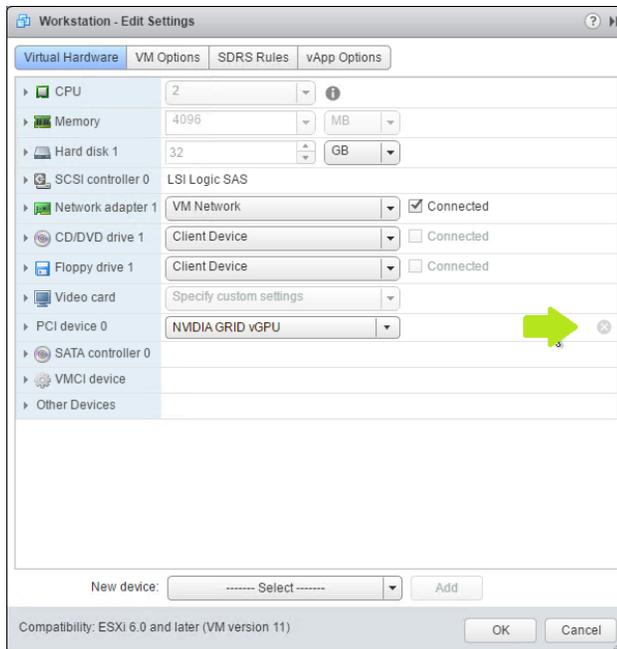
Creating your first GRID Virtual workstation



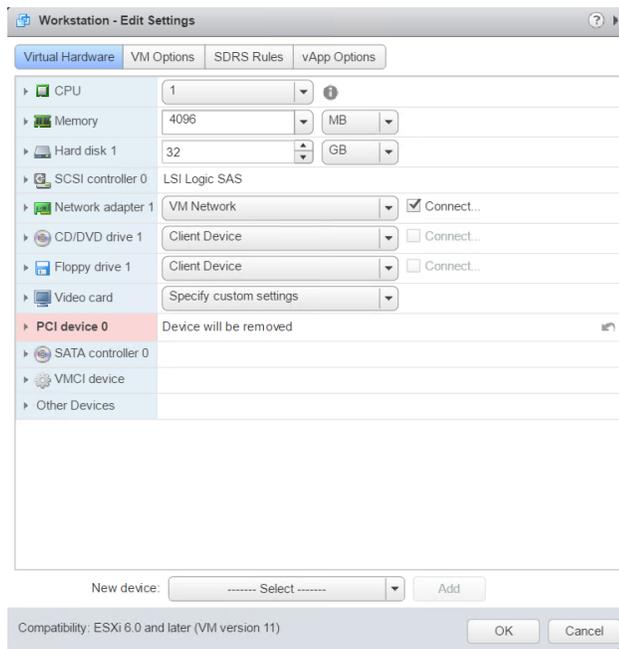
1. In the vSphere Web Client **Navigator** left-hand pane, select the **VMs and Templates** tab.



2. Right-click on the master image VM and select the **Edit Settings** menu item.



3. Select the gray “x” on the far right on the **PCI Device** item that is connected to the **NVIDIA GRID vGPU** to select the device for removal.



4. Select the **OK** button at the bottom of the **Edit Settings** dialog to save the settings.

The Recent Tasks pane at the bottom will display the status of Complete once the device has been removed. A XenDesktop machine catalog can be safely configured now.

CHAPTER 11. CREATING XENDESKTOP MACHINE CATALOGS AND DELIVERY GROUPS

This chapter describes the following:

- ▶ Creating a VM snapshot
- ▶ Creating a machine catalog for the master image
- ▶ Creating a XenDesktop random machine catalog from a snapshot
- ▶ Configuring VM to enable vGPU profiles
- ▶ Creating a XenDesktop delivery group for the master image
- ▶ Creating a XenDesktop delivery group for the catalog
- ▶ Connecting to the Citrix delivered desktop for the first time

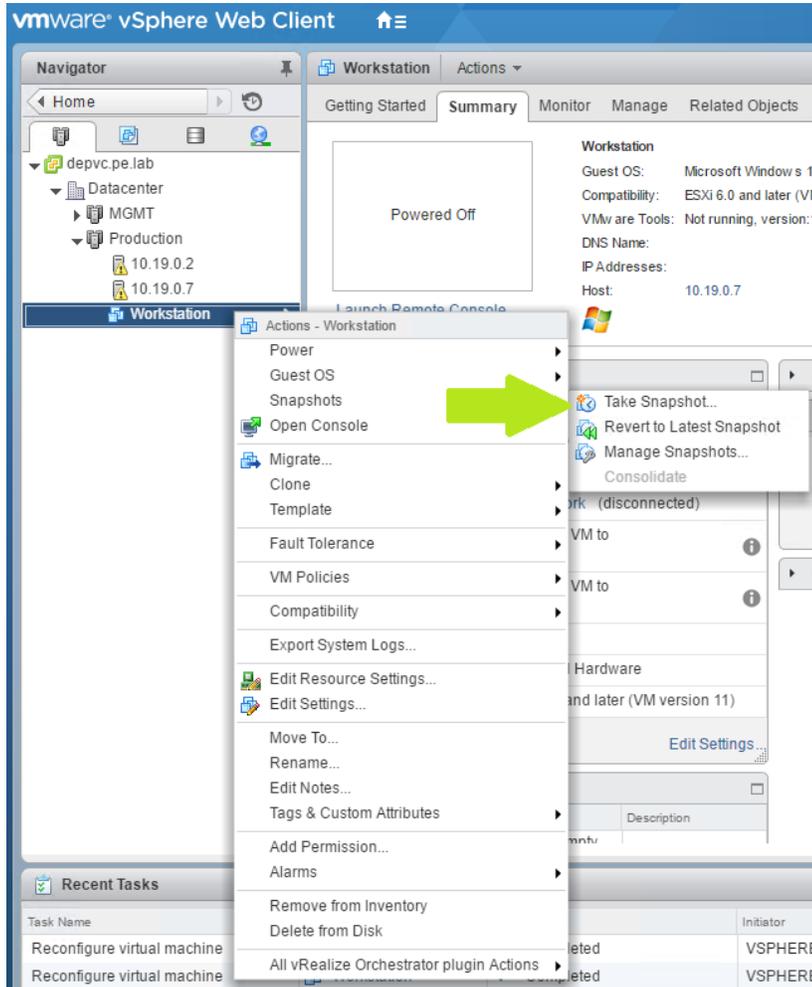
To create a pool of virtual machines, you must first take a snapshot of an existing virtual machine, which then allows you to either create a single virtual machine or use it to dynamically create virtual machines on demand.



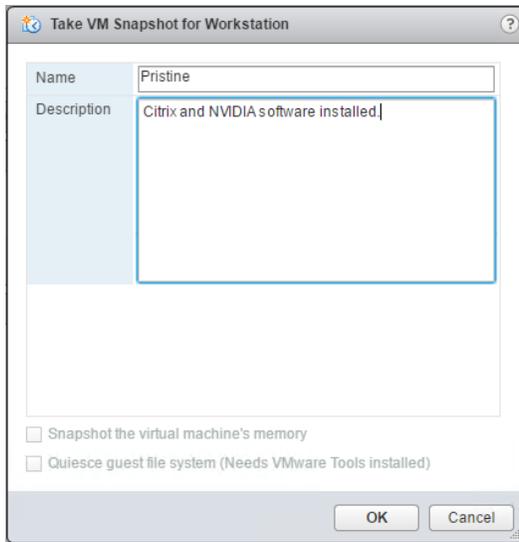
A complete demonstration of these features is beyond the scope of this document; however, taking a virtual machine snapshot and then deploying virtual machines from that snapshot is a fundamental operation that will reduce evaluation time.

11.1 CREATING A SNAPSHOT

Use the following procedure to create a snapshot for an existing virtual machine.



1. Right-click the virtual machine and then select **Snapshot** → **Take Snapshot...**



2. Enter a name for the snapshot. Optionally enter a description for the snapshot. Select **OK** to take the snapshot.

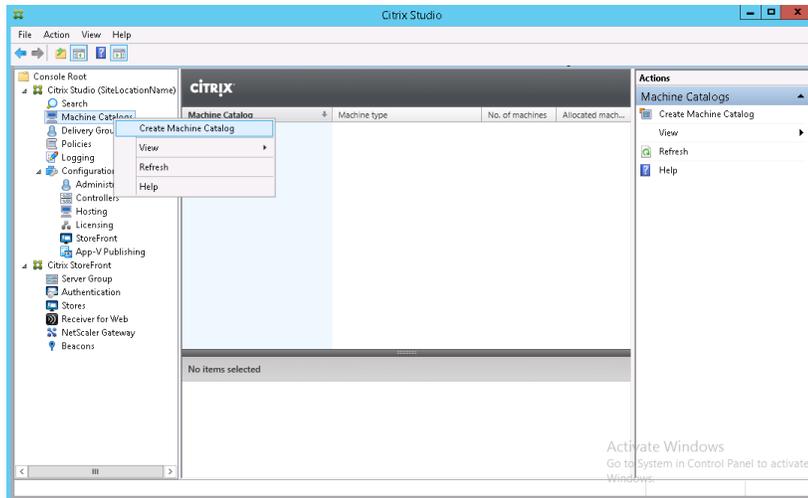
11.2 CREATING A MACHINE CATALOG FOR THE MASTER IMAGE (OPTIONAL)



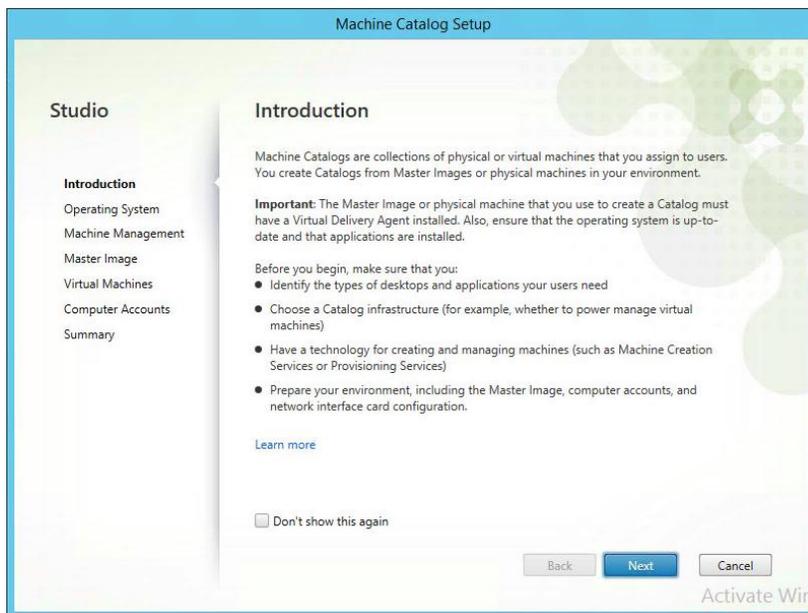
Although this is not a required step, the following procedure can be used to validate the communication paths between the end user and the target VMs that will be created from this master before they are created. This can reduce troubleshooting time and rework.

Use the following procedure to provision a single vGPU-enabled machine catalog:

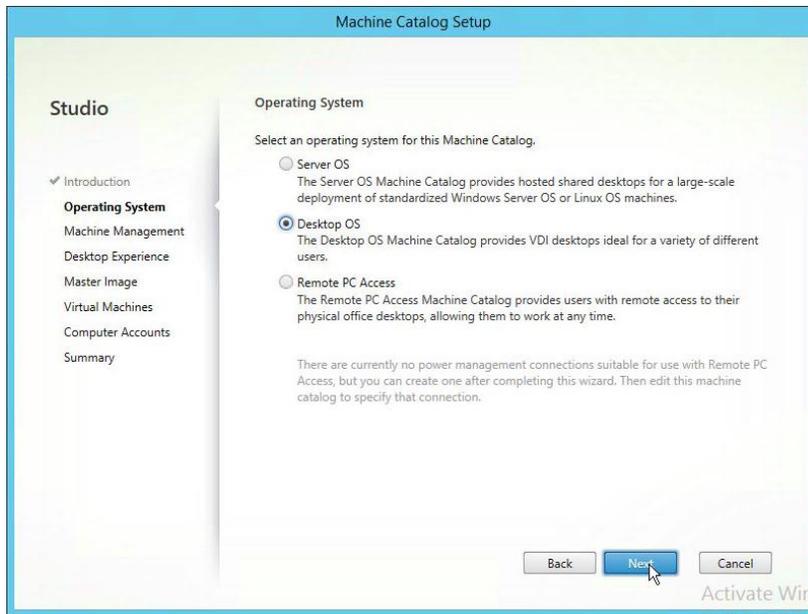
1. Logon to the XenDesktop server and open **Citrix Studio** from the start menu.



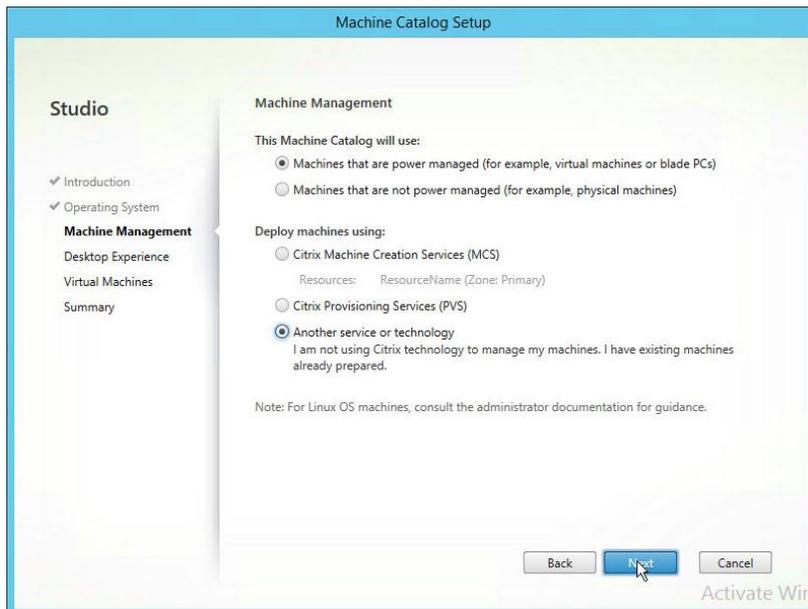
2. Expand **Citrix Studio** on the left-hand pane if needed. Right-Click on **Machine Catalogs** menu item and select **Create Machine Catalog** menu item.



3. Select **Next** button on the **Machine Catalog Setup** dialog to begin.

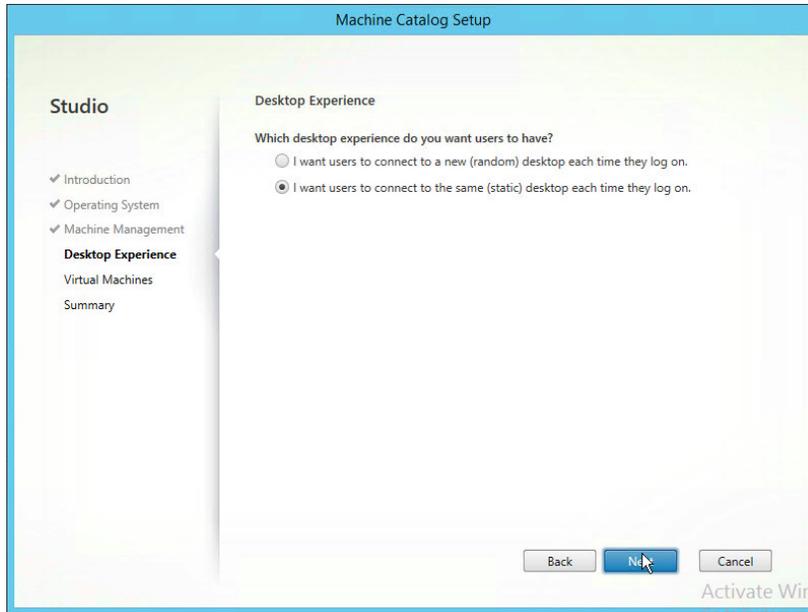


4. On the **Operating System** dialog, select the **Desktop OS** radio button and then select **Next** to continue.

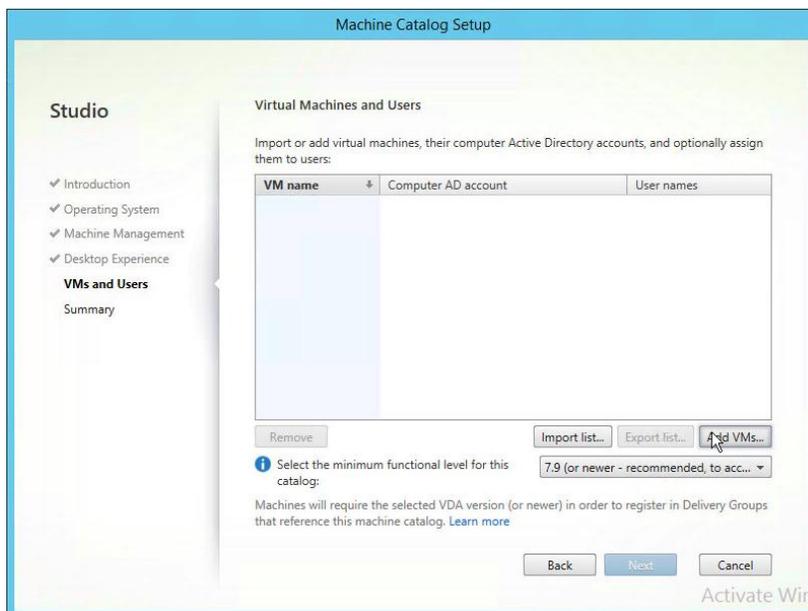


5. On the **Machine Management** dialog select the **Another service or technology** radio button for the **Deploy machines using:** section. Select **Next** to continue.

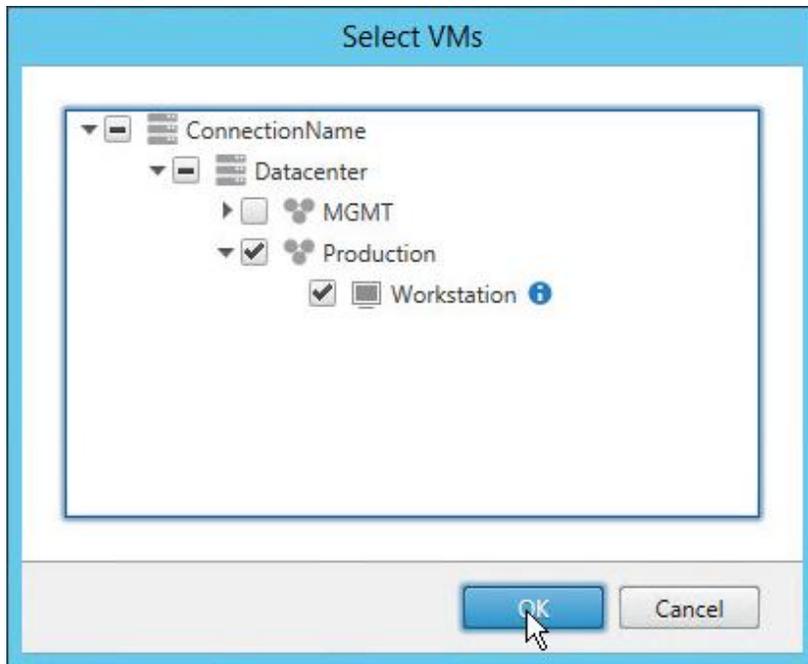
 This is only used for the master image, the random pool will use Citrix Machine Creation Services (MCS)



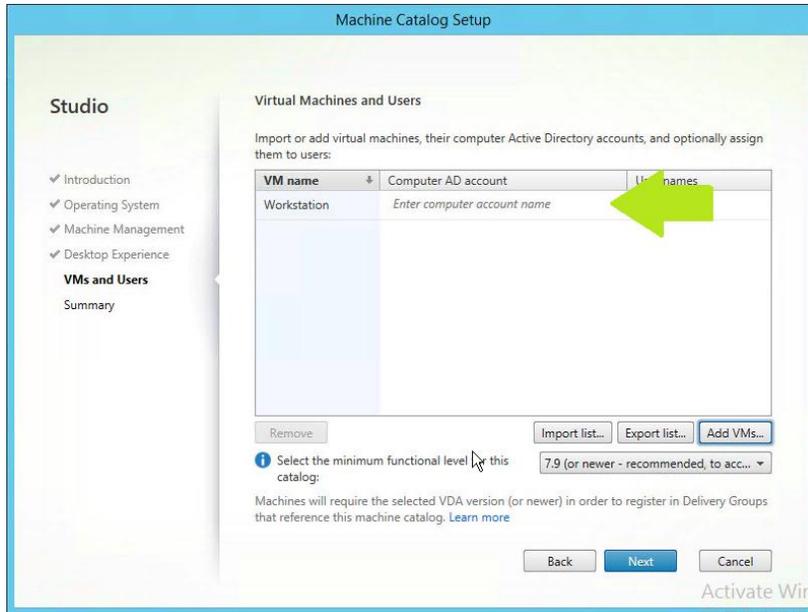
6. On the **Desktop Experience** dialog, accept the default and select **Next** to continue.



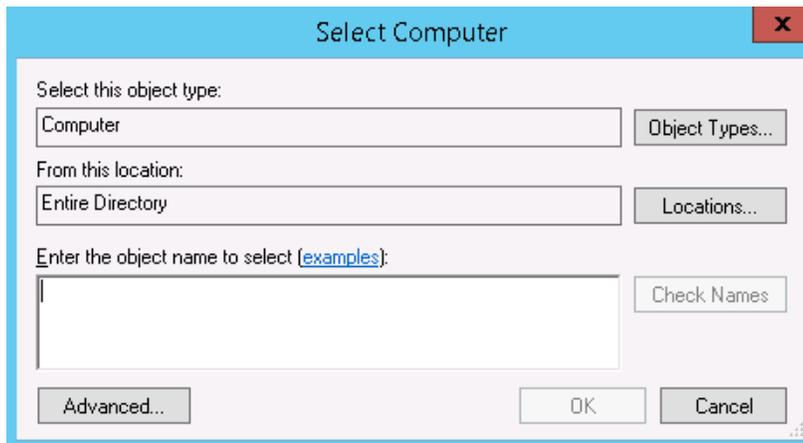
7. On the **Virtual Machines and Users** dialog, select the **Add VMs...** button to select the master image to the machine catalog.



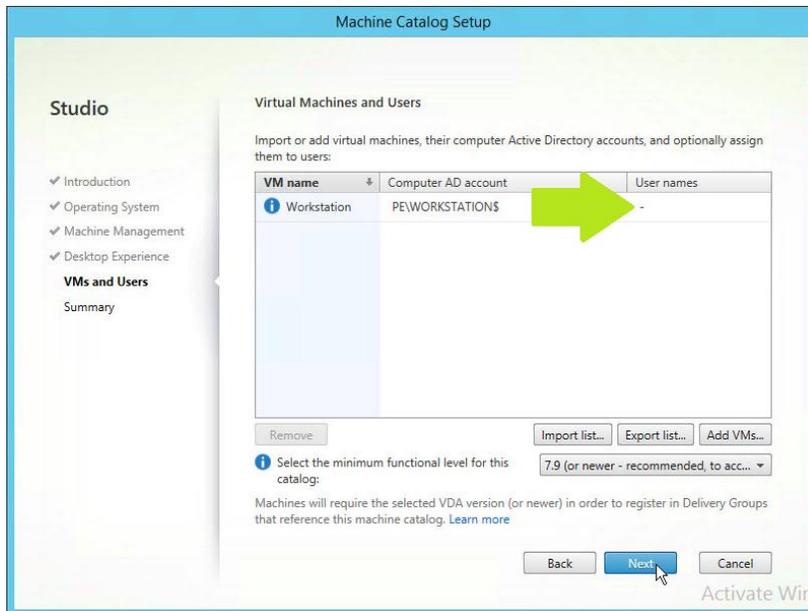
8. On the **Select VMs** dialog, navigate through the vCenter VMs site tree and select the check box next to the VM you created to be the master image. Select the **OK** button to continue.



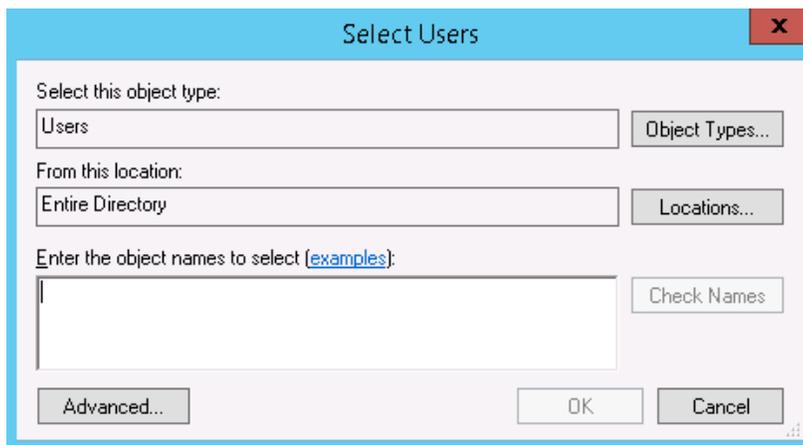
9. Select the ellipsis button next to the **Computer AD Account** field to search Active Directory for the machine account to add.



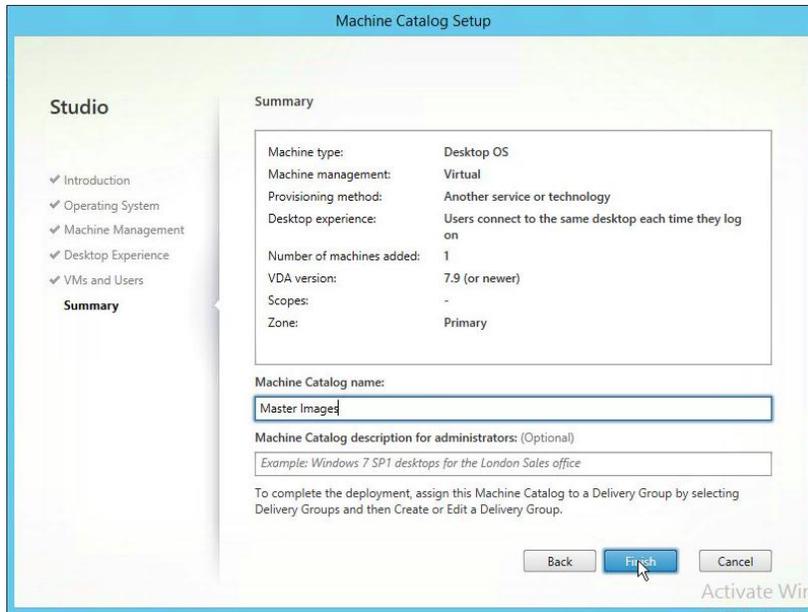
10. Enter the computer name into the Active Directory **Select Computer** dialog. Click the **Check Names** button to verify the computer account exists and then select **OK** button to continue.



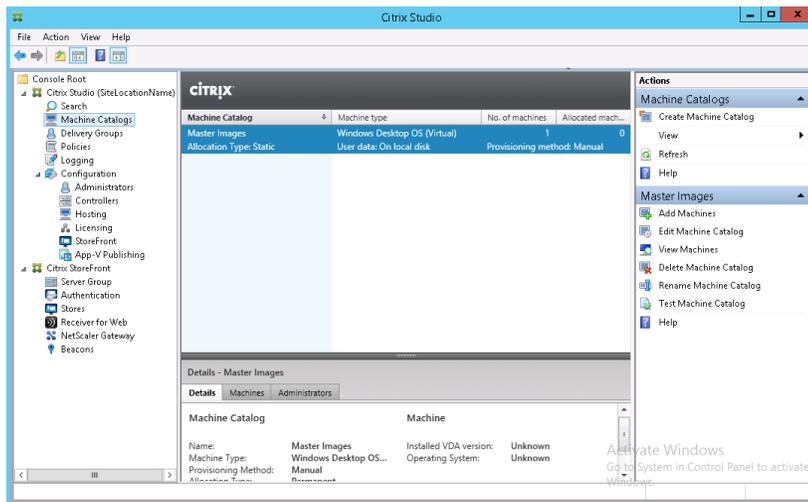
11. Select the **ellipsis** next to **User Names** field to search Active Directory for the user account to add.



12. Enter the users name into the Active Directory **Select Users** dialog. Select the **Check Names** button to verify the users account then select the **OK** button to continue.
13. Once you have completed adding computer and user account references for all machines in the catalog select **Next** to continue.



14. On the **Summary** dialog enter a Machine Catalog name and then select **Finish** button to continue.



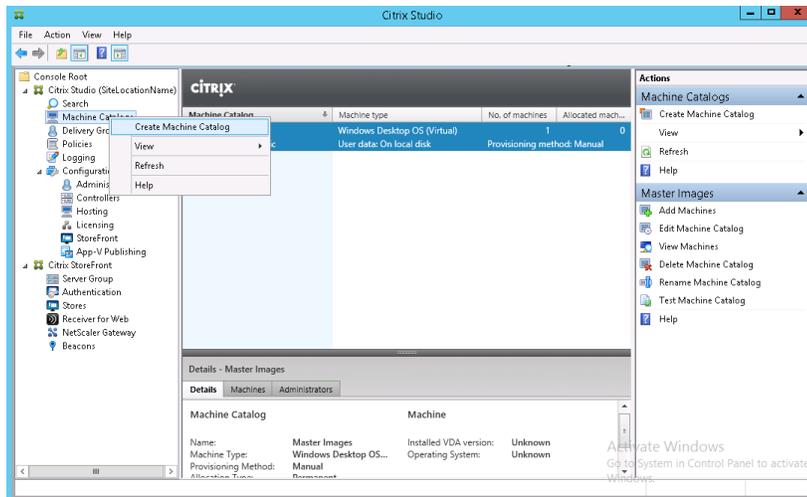
15. Once complete, the machine catalog will be displayed in the center pane.

11.3 CREATING A XENDESKTOP RANDOM MACHINE CATALOG FROM A SNAPSHOT

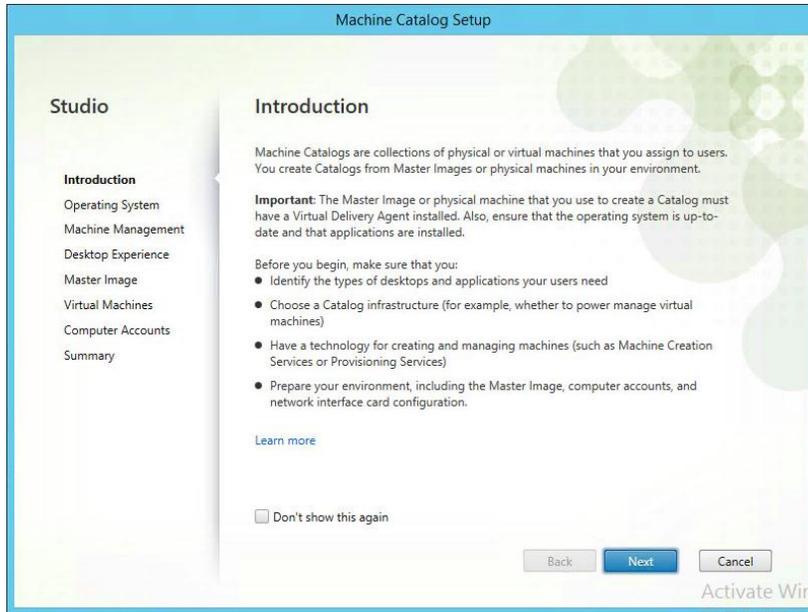
Use the following procedure to provision a random machine catalog.

MCS Random pool deployments require DHCP scopes to be available for the VM's. You cannot assign a static IP address to a MCS pool random VM. A DHCP IP reservation can be assigned to the VM MAC Address.

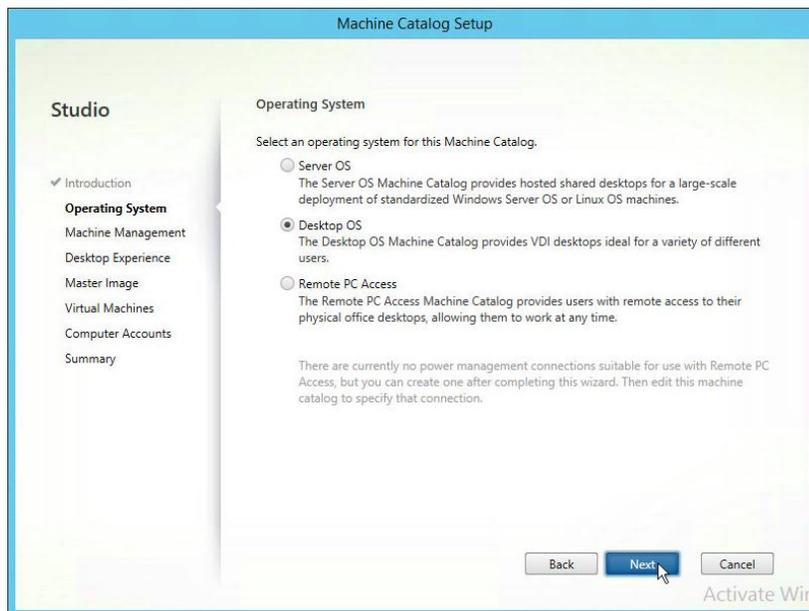
MCS Machine Catalogs require a snapshot of the Master image. Select the reference VM in the Navigator and right click. Ensure you have a snapshot of the master image before proceeding.



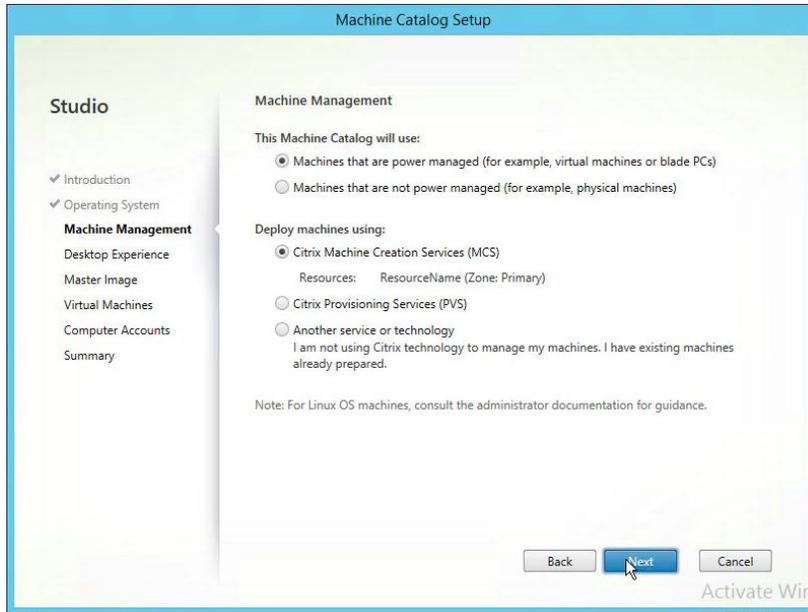
1. Right-Click on **Machine Catalog** from the left-hand pane and select **Create Machine Catalog** menu item.



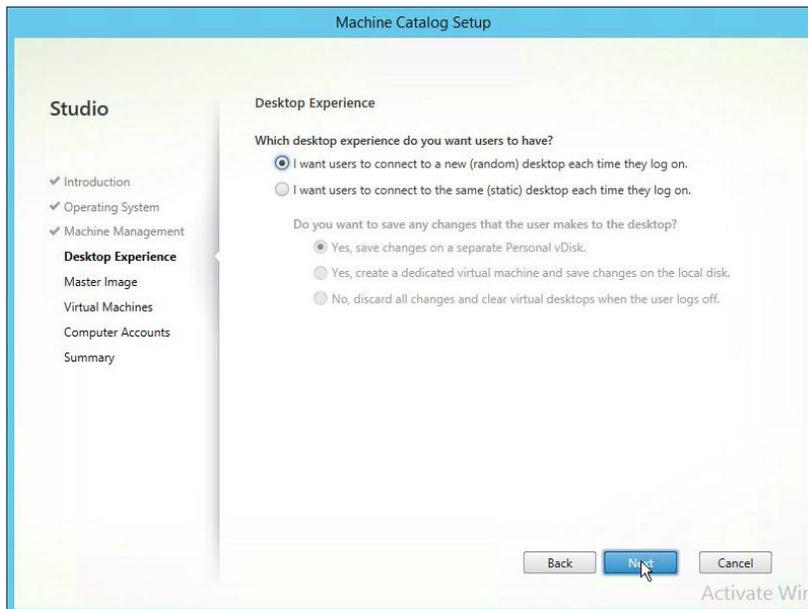
2. Select **Next** to continue.



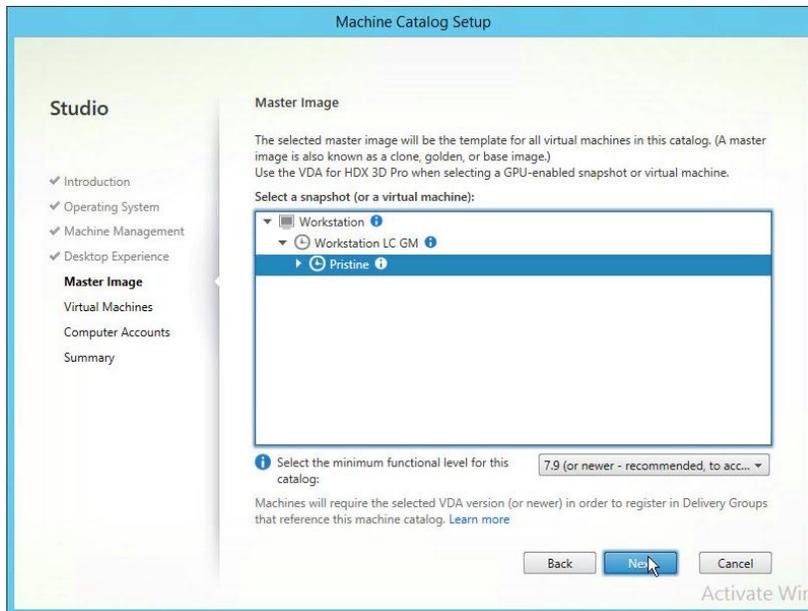
3. On **Operating System** dialog select **Desktop OS** radio button then select **Next** to continue.



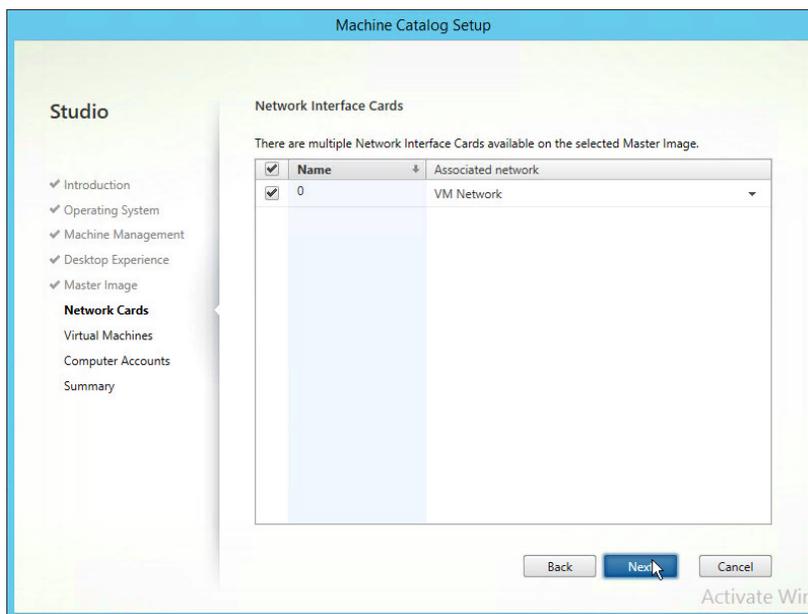
4. On **Machine Management** dialog accept defaults for **Citrix Machine Creation Services (MCS)** and then select **Next to continue...**



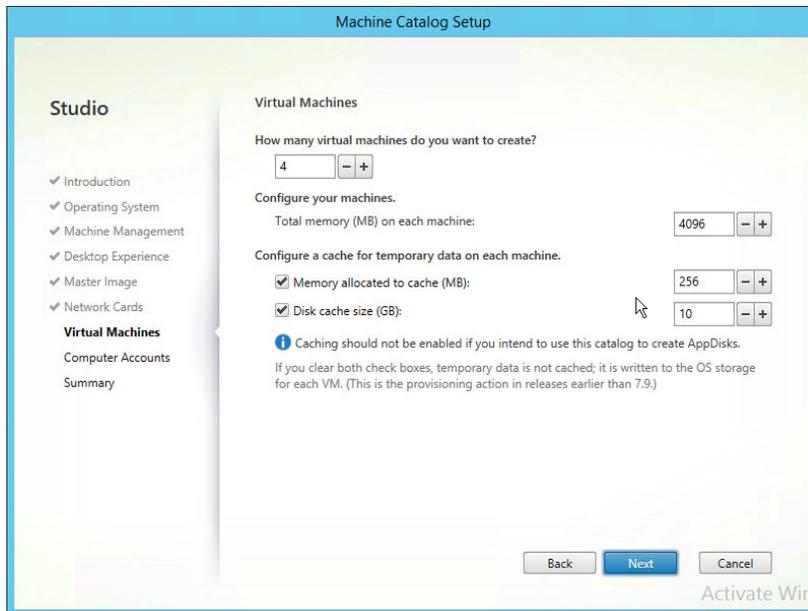
5. On the **Desktop Experience** dialog select the **I want users to connect to a new (random) desktop each time they logon** radio button. Click **Next**.



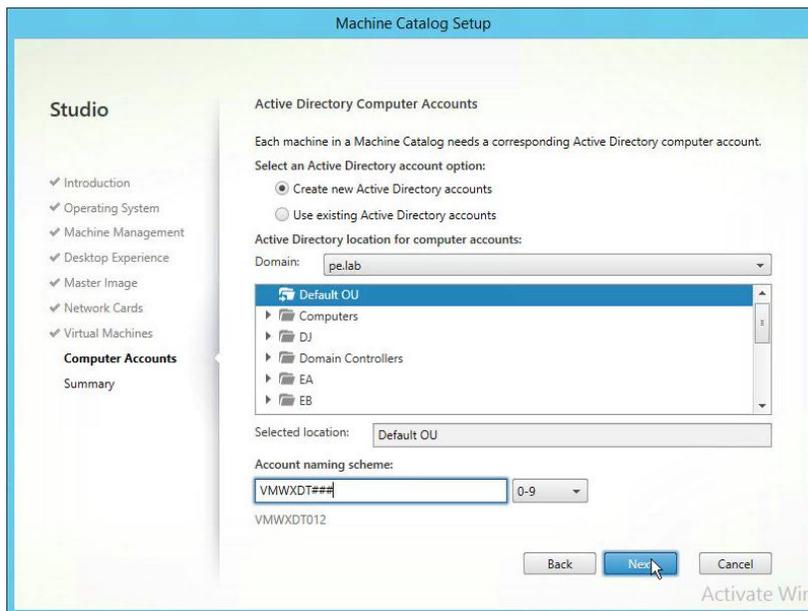
6. On the **Master Image** dialog expand the virtual machine that you wish to deploy from and browse down until you can select the Snapshot you wish to use for the random pool. Select **Next** to continue.



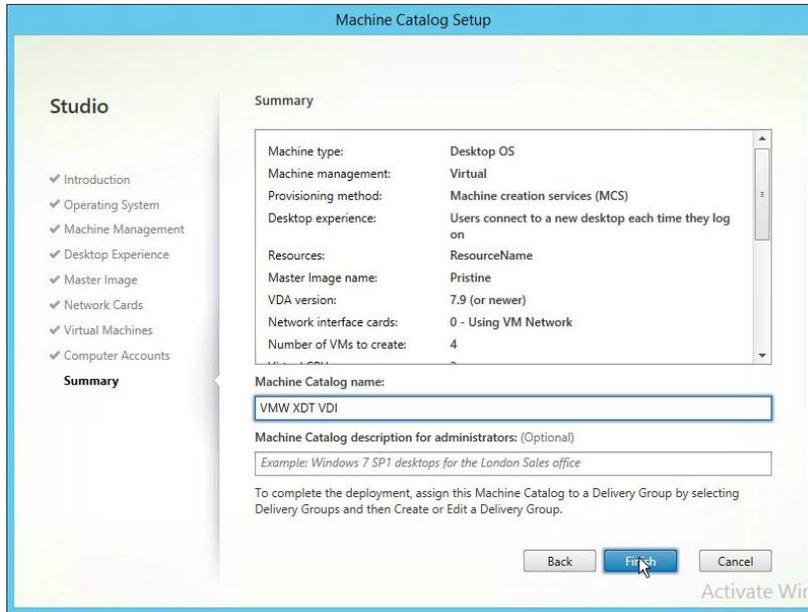
7. On the **Network Cards** dialog specify the network cards and networks you wish to associate. Click **Next**.



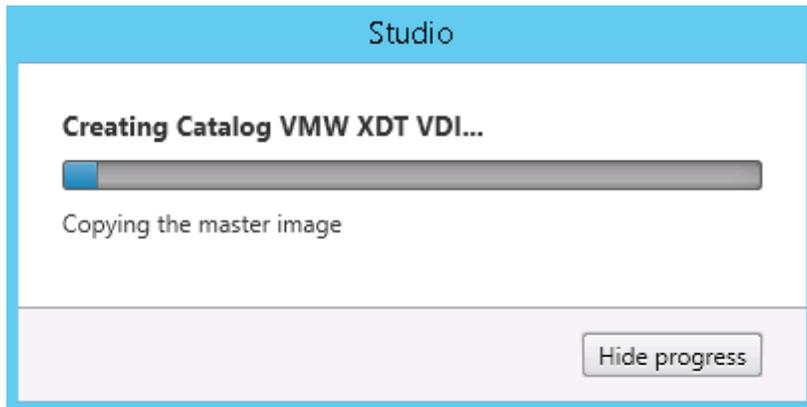
8. On the **Virtual Machines** dialog specify the **Number of virtual machines** needed and then select **Next** to continue.



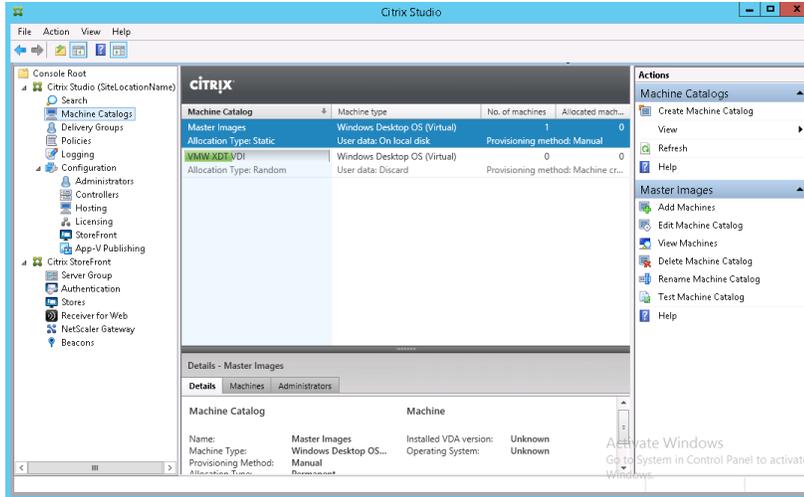
9. On the **Active Directory Computer Accounts** select the Domain to create machines in, browse the **OU tree** and select the OU to create machine accounts in, the OU should be displayed in the **Selected location** display field. Specify an **Account naming scheme** (append ## to the name for automatic numeric increments) then select **Next** to continue.



10. On the **Summary** dialog enter the **Machine Catalog** name and then select **Finish** button to continue.



11. A progress bar will be displayed on the **Studio** dialog while the machines are being provisioned. Click the **Hide progress** button to allow the process to complete in the background.



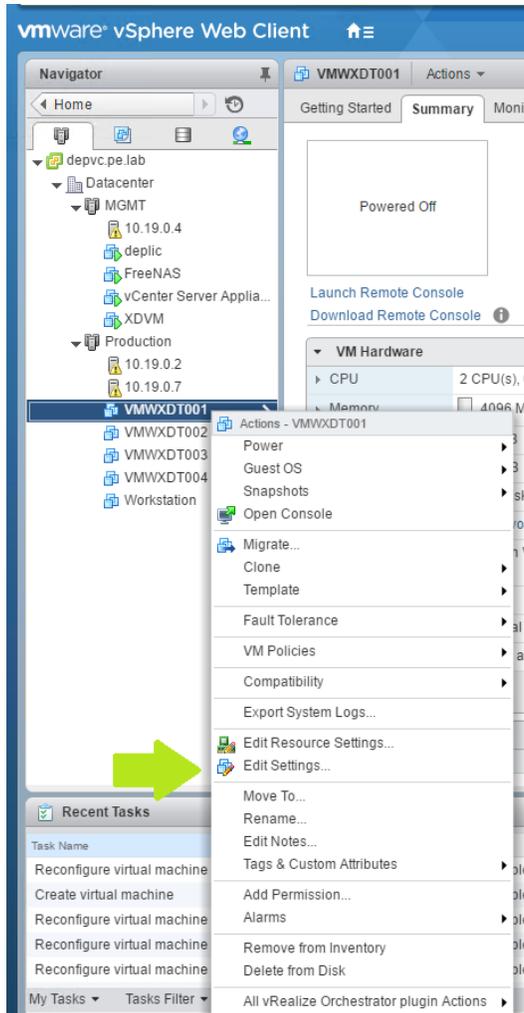
12. A progress bar continues to be displayed on the center pane of the **Machine Catalogs** screen in Citrix Studio.

Once the machine catalog is created, re-enable the vGPU profiles in the next step.

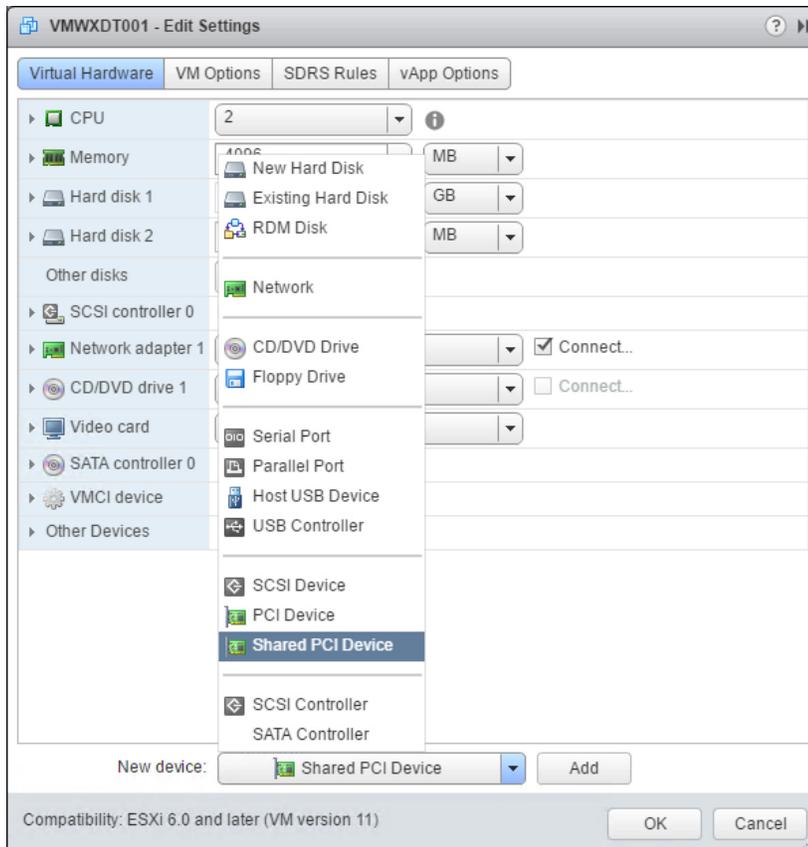
11.4 CONFIGURING CATALOG VMS TO ENABLE GRID VGPU PROFILES

Perform the following procedure to each VM provisioned in both the Random and Static Machine Catalogs. Use a Static Machine Catalog if DHCP is not available.

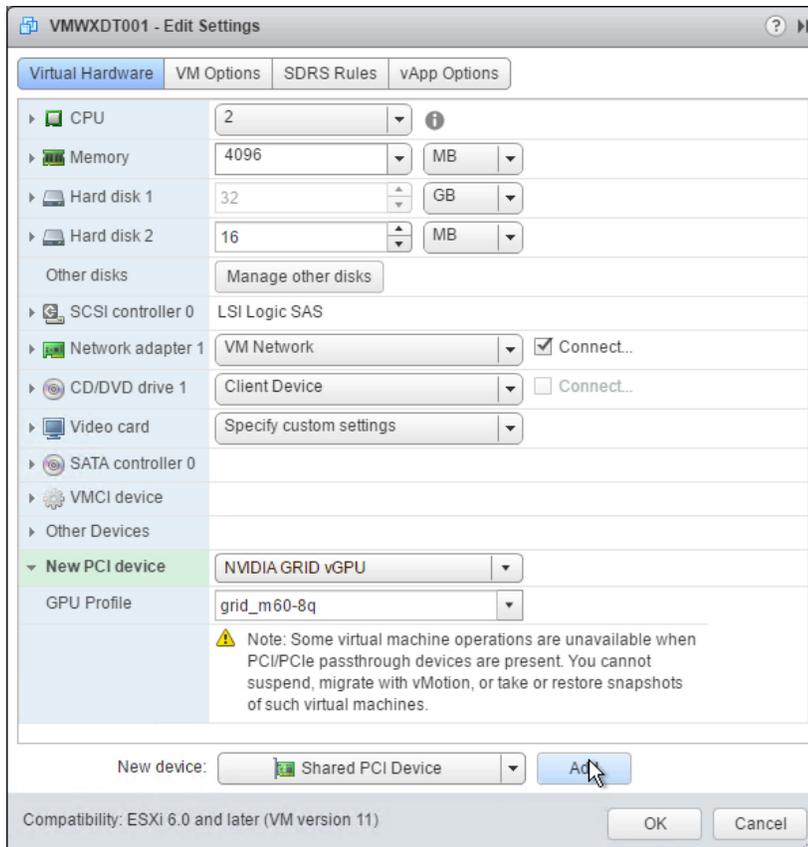
! CAUTION: If using static assigned IP addresses boot machines up and assign IP addresses before adding Shared PCI device for vGPU. VMware console is not usable with vGPU enabled.



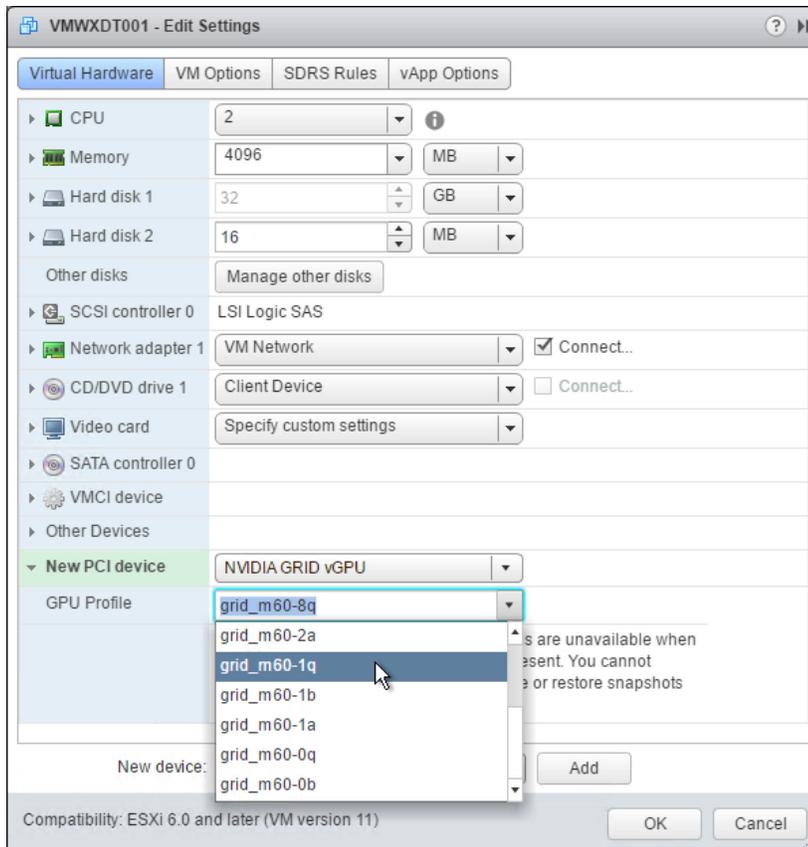
1. In the vSphere Web Client right click on a VM in the Catalog. Select **Edit Settings** to open the VM Settings dialog.



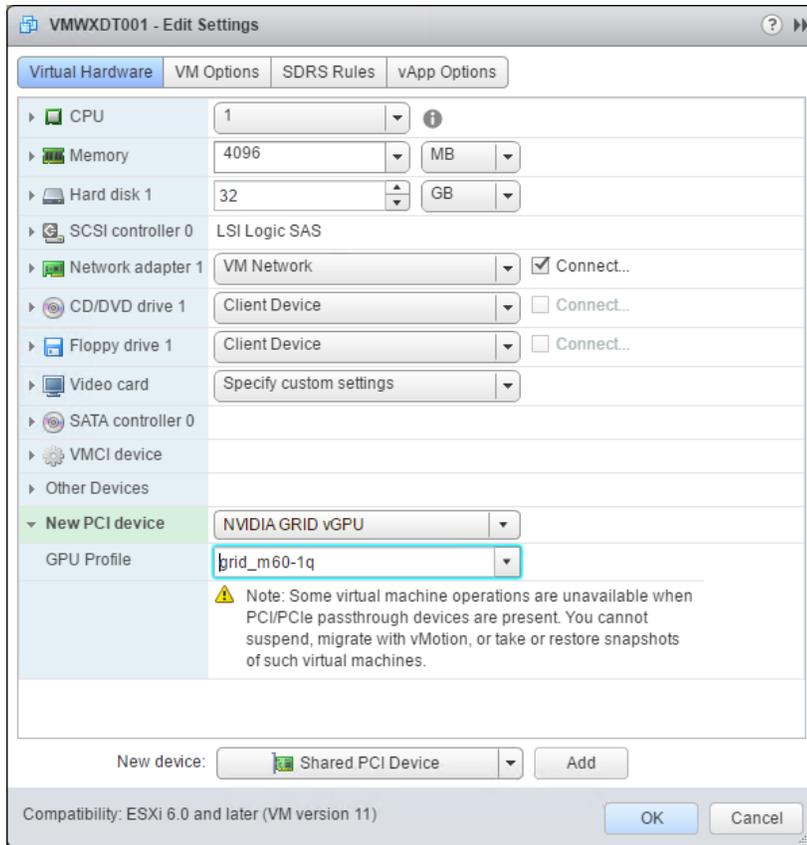
2. On the VM **Edit Settings** dialog click on the **New device** drop down menu at the bottom and select **Shared PCI device**.



3. Select **Add** button to add the **Shared PCI Device** to the VM.



4. Select the **GPU Profile** for this VM and then select **OK** to close the dialog and save the VM settings.

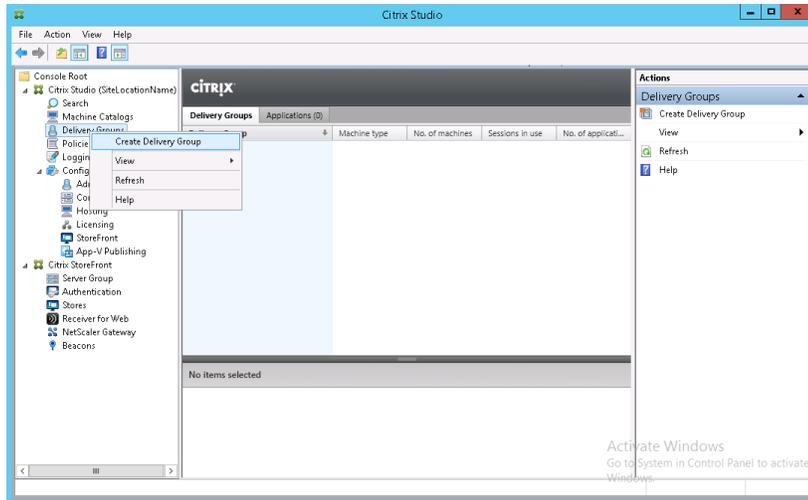


5. Repeat for all VM's provisioned on the Machine Catalog.

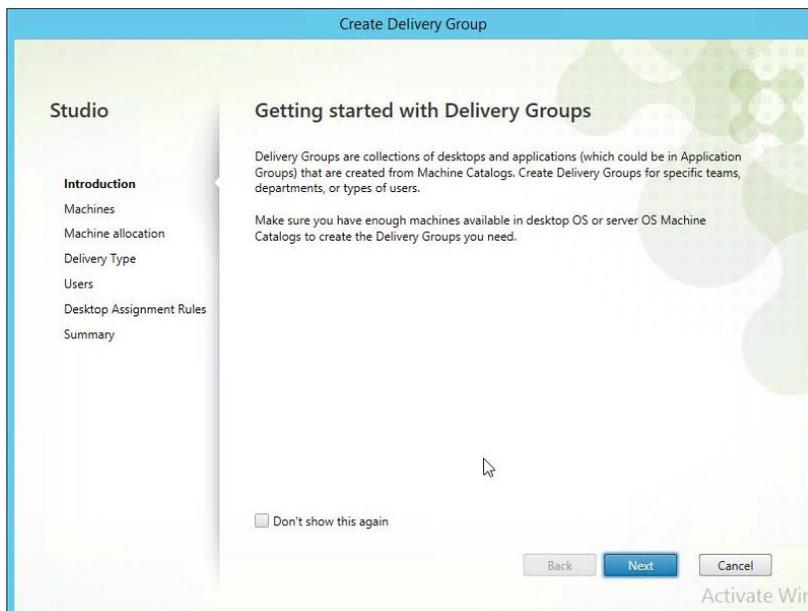
11.5 CREATING A XENDESKTOP DELIVERY GROUP FOR THE MASTER IMAGE (OPTIONAL)

Although this is not a required step, the following procedure can be used to validate the communication paths between the end user and the target VMs that will be created from this master before they are created. This can reduce troubleshooting time and rework.

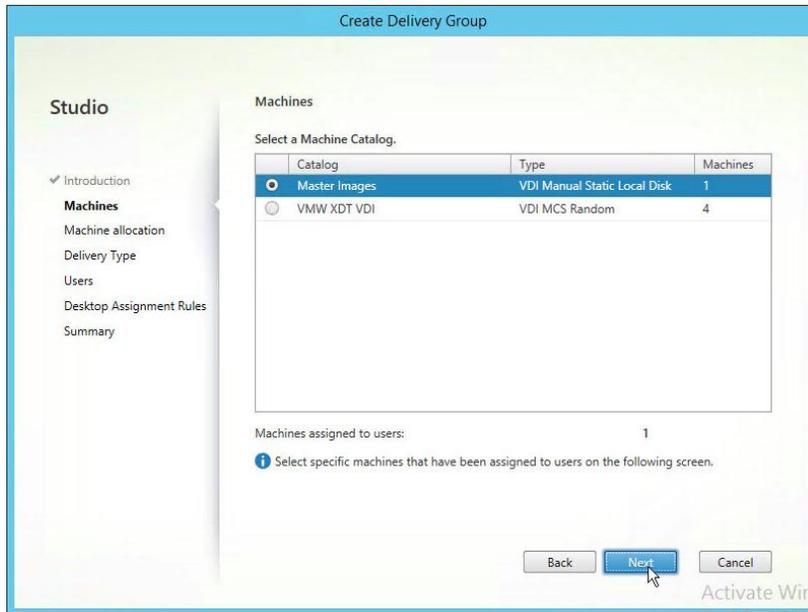
Perform the following procedure to configure the master image access through Citrix Receiver.



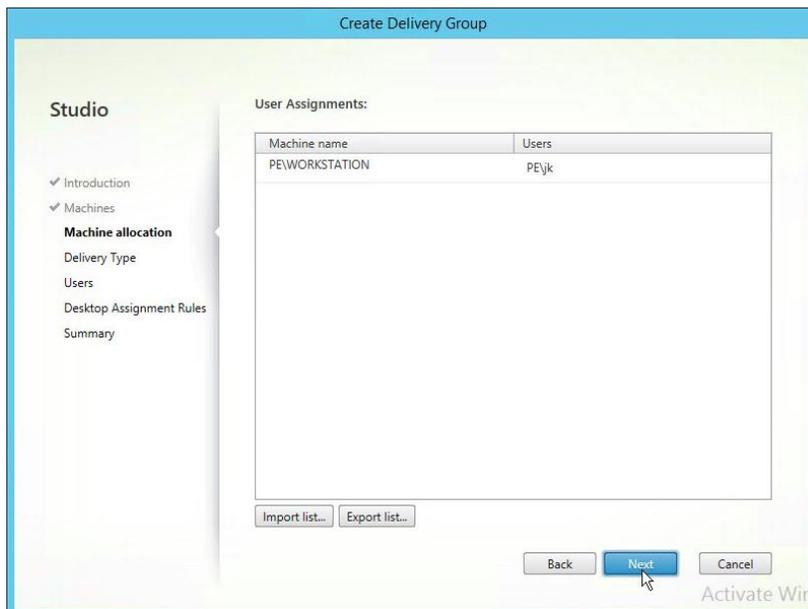
1. Login to the XenDesktop server and open **Citrix Studio**. Right-click on **Delivery Groups** from the left-hand pane and select the **Create Delivery Group** menu item.



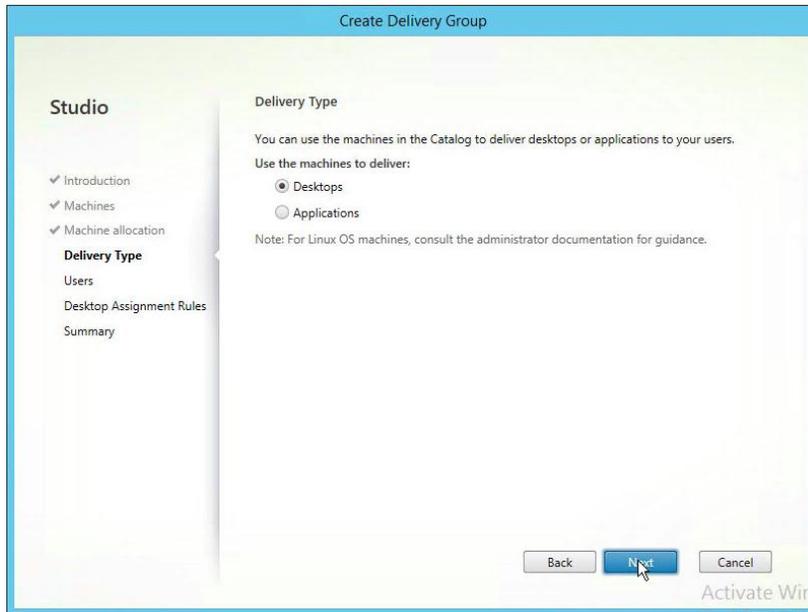
2. On the **Create Delivery Group** dialog select the **Next** button to continue.



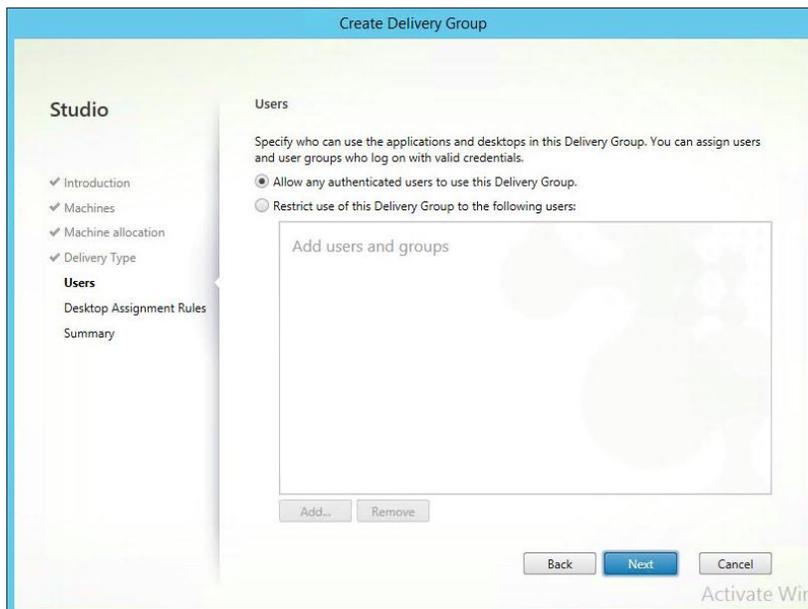
3. On the **Machines** dialog select the master image catalog that we created earlier. Select the **Next** button to continue.



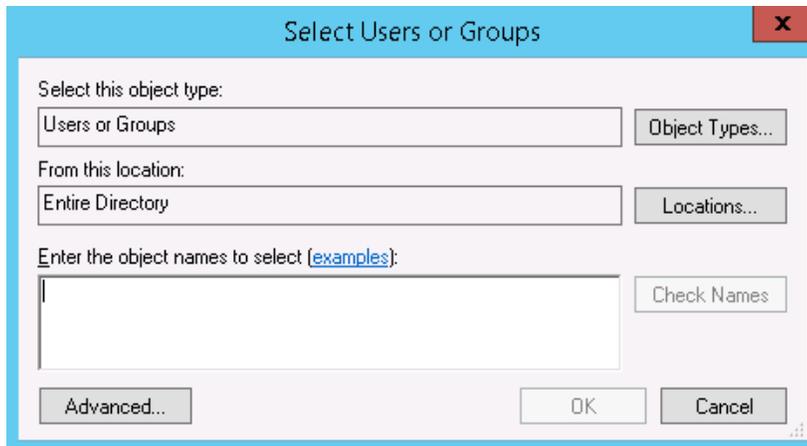
4. On the **User Assignments** dialog optionally assign users to machines and then select the **Next** button to continue.



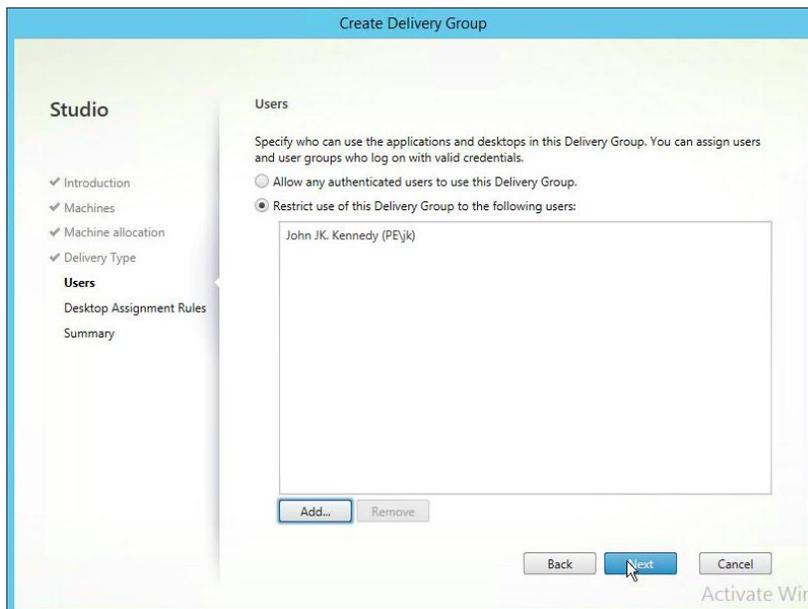
5. On the **Delivery Type** dialog accept the default selection of **Desktops** radio button and then select the **Next** button to continue.



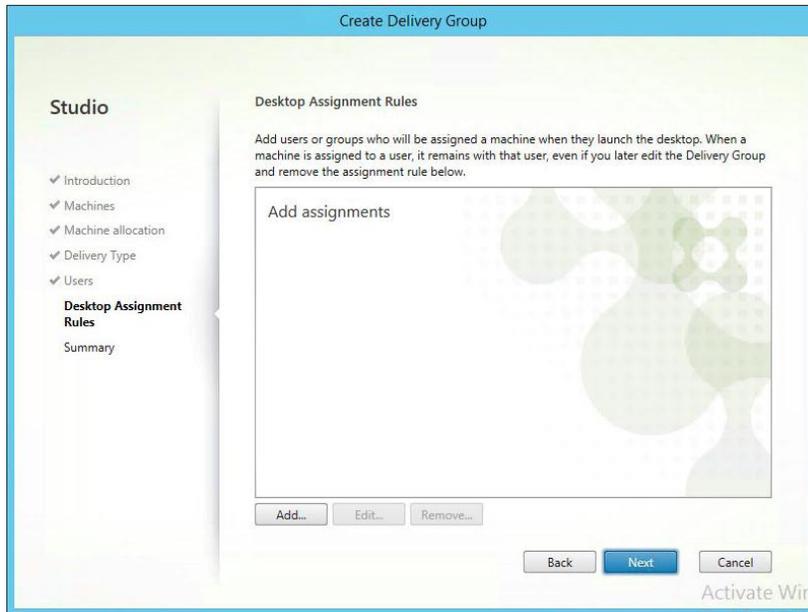
6. On the **Users** dialog, select **Restrict use of this delivery group to the following users**, then select the **Add** button to use the **Active Directory Select Users and Groups** to grant users or groups access to the delivery group.



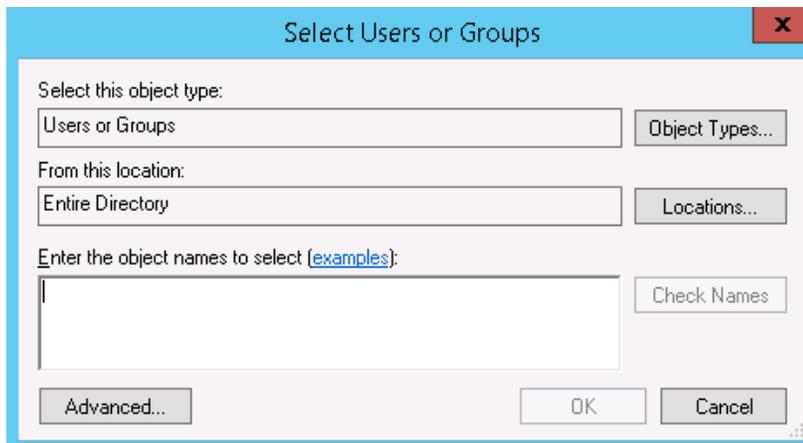
7. On the **Select Users and Groups** dialog search for and add the users and groups that will be granted access to the delivery group. Select the **OK** button to continue.



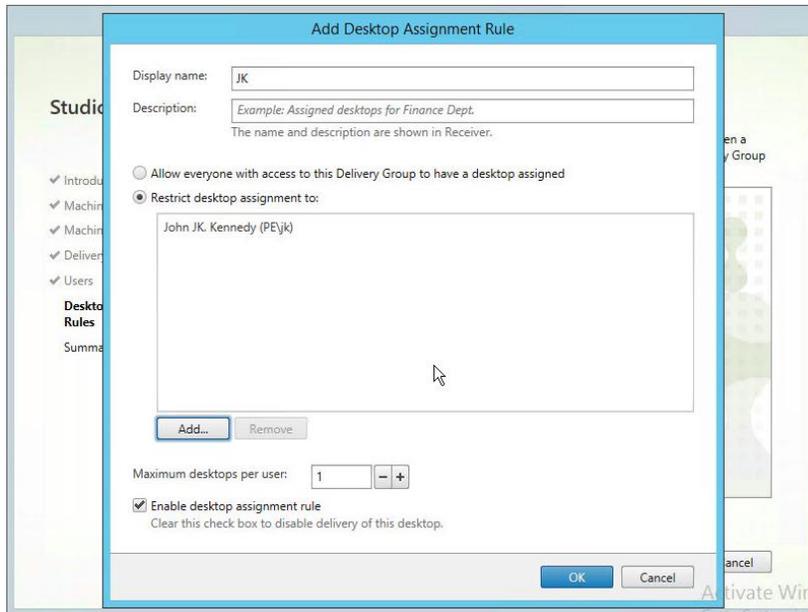
8. On the **Users** dialog select the **Next** button to continue.



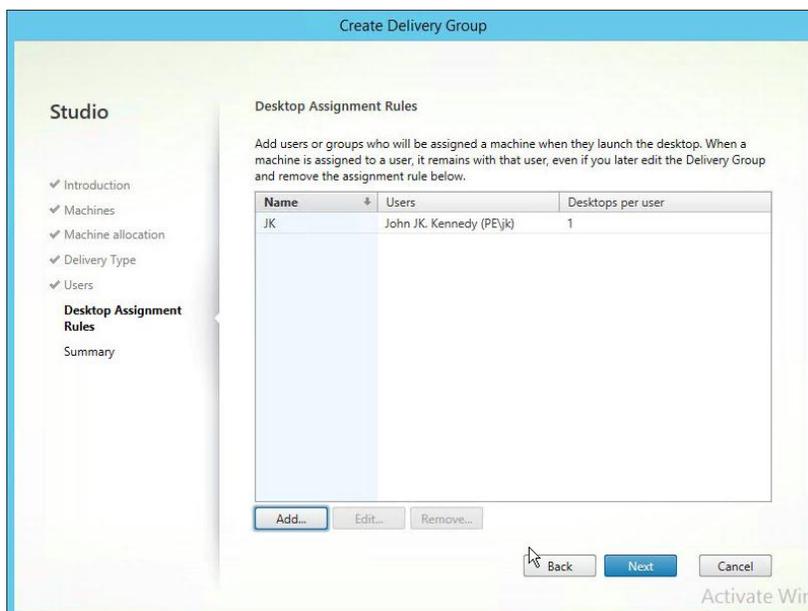
9. On the **Desktop Assignment Rules** dialog, click **Add**. Give the new rule a display name. Select **Restrict use of this delivery group to the following users**, then select the **Add** button to use the **Active Directory Select Users and Groups** to grant users or groups access to the delivery group.



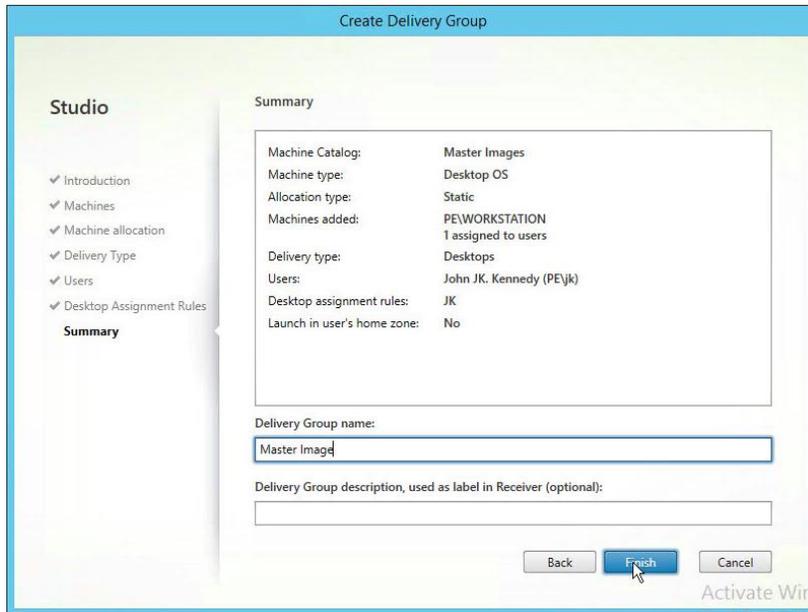
10. On the **Select Users and Groups** dialog search for and add the users and groups that will be granted access to the delivery group. Select the **OK** button to continue.



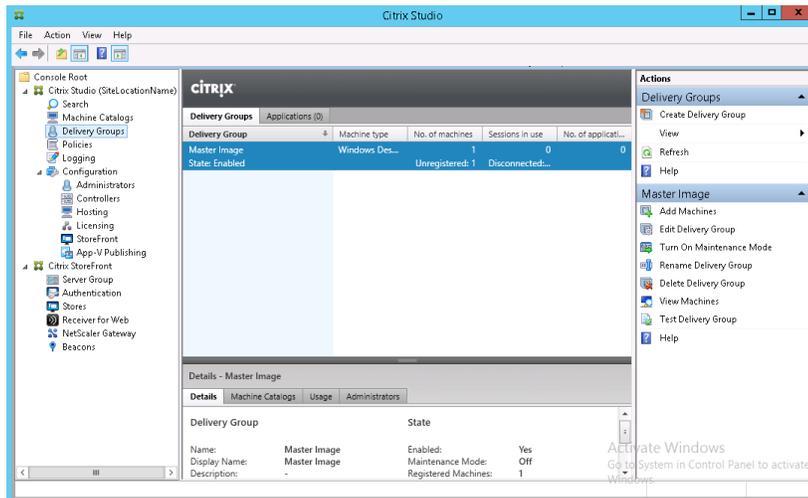
11. On the **Add Desktop Assignment Rule** dialog, click **OK**.



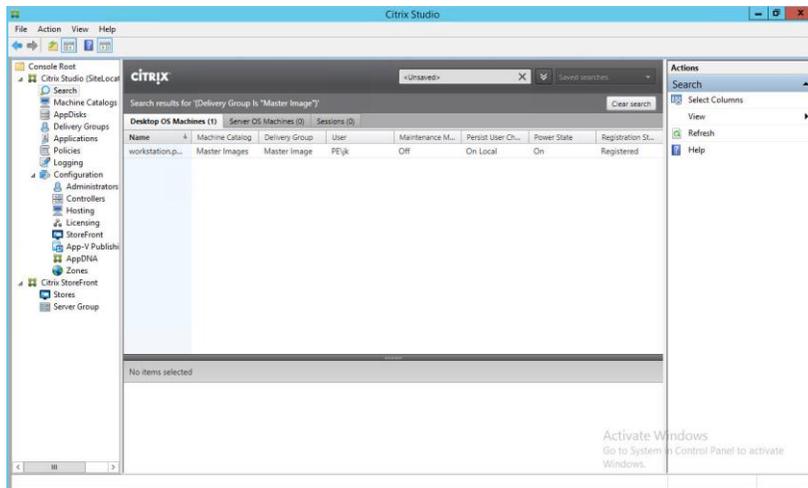
12. On the **Desktop Assignment Rules** dialog select the **Next** button to continue.



- On the **Summary** dialog enter a unique **Delivery Group** name, and **Display Name** for the delivery group. Select the **Finish** button to complete the creation of the delivery group.



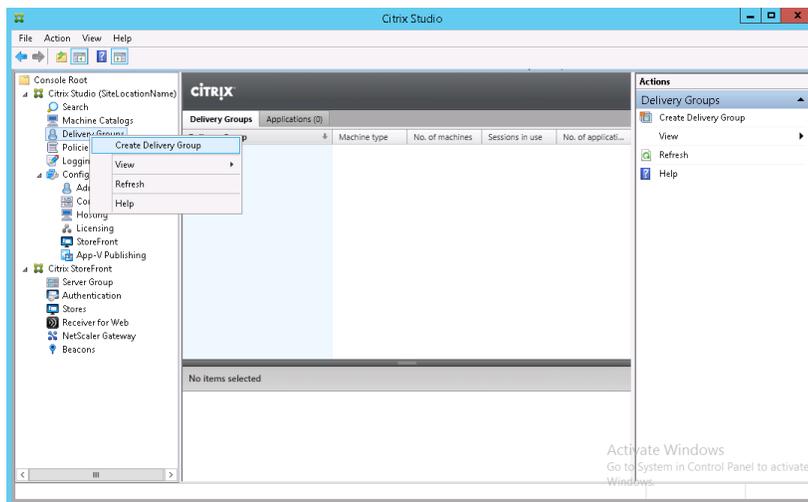
- On the **Delivery Groups** center pane, the newly created delivery group should now be available. Double-click on the **Delivery Group** row on the center pane to view the machines that are available for delivery.



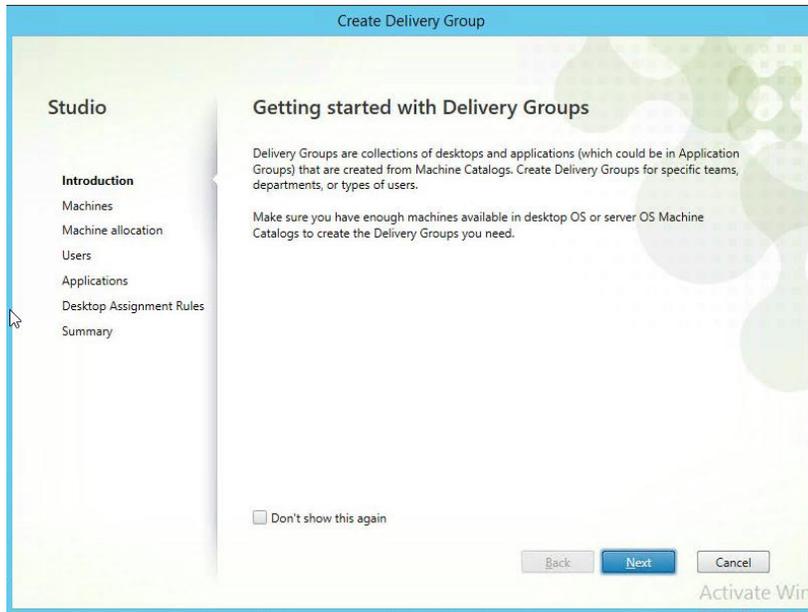
15. The machines available for this **Delivery Group** will be listed in the center pane.

11.6 CREATING A XENDESKTOP DELIVERY GROUP FOR THE POOLED MACHINE CATALOG

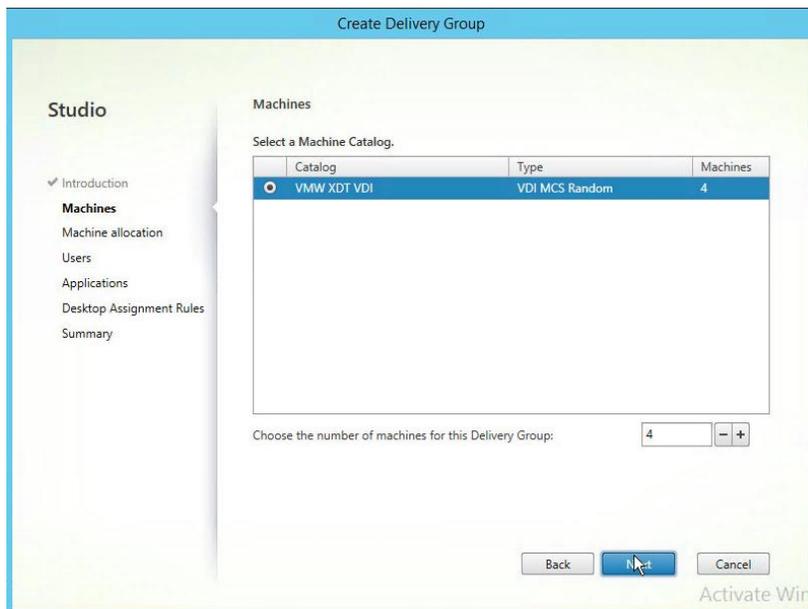
Use the following procedure to create a delivery group for the pooled machine catalog.



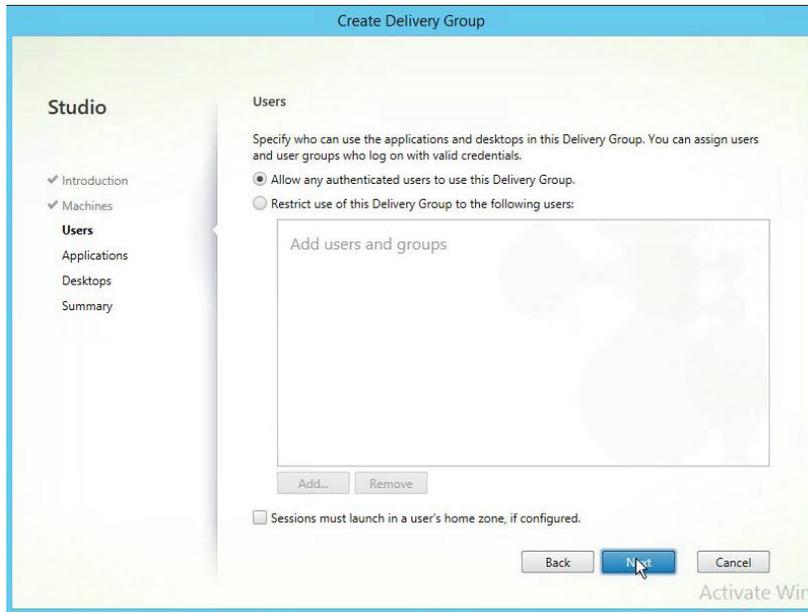
1. Right-click on **Delivery Groups** from the left-hand pane and select the **Create Delivery Group** menu item.



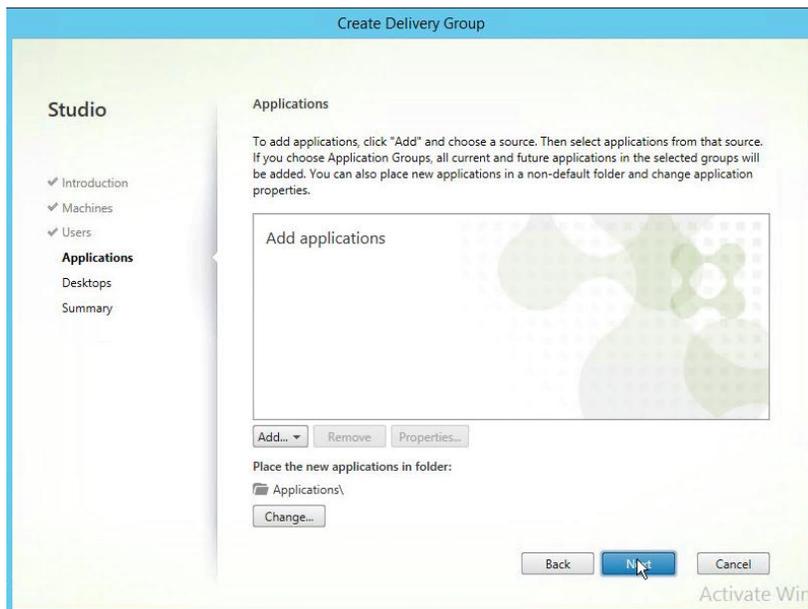
2. On the **Create Delivery Group** dialog select the **Next** button to continue.



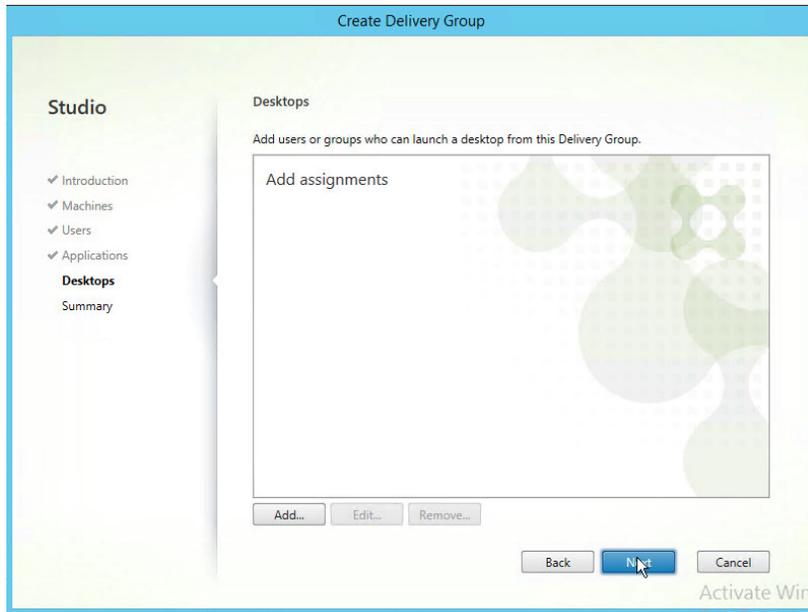
3. On the **Machines** dialog select the random pool catalog that we create earlier. Increase the counter for the number of machines in this group to 4. Select the **Next** button to continue.



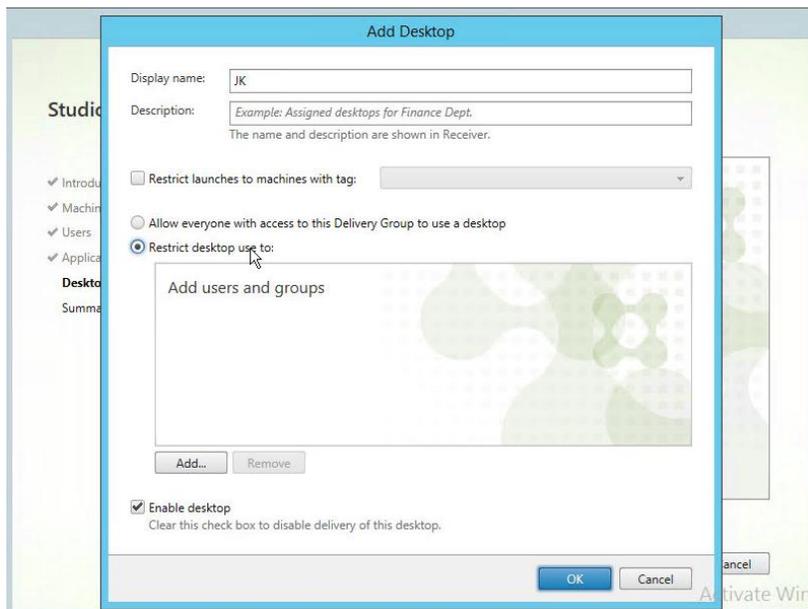
4. On the **Users** dialog select the **Allow any authenticated users to use this Delivery Group**. Click **Next**.



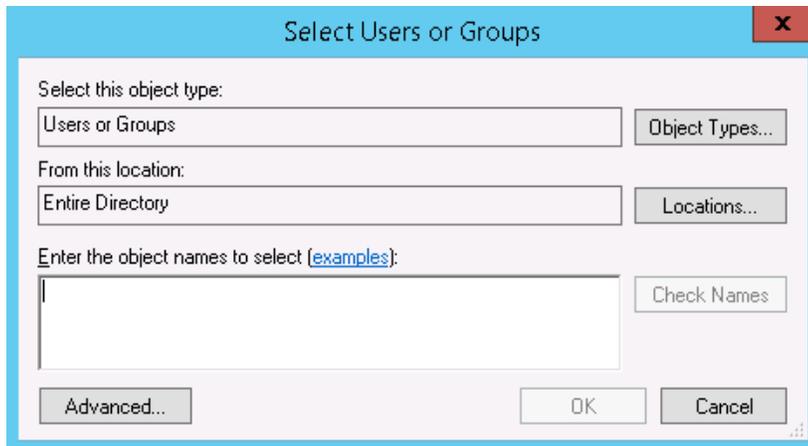
5. On the **Applications** dialog click **Next**.



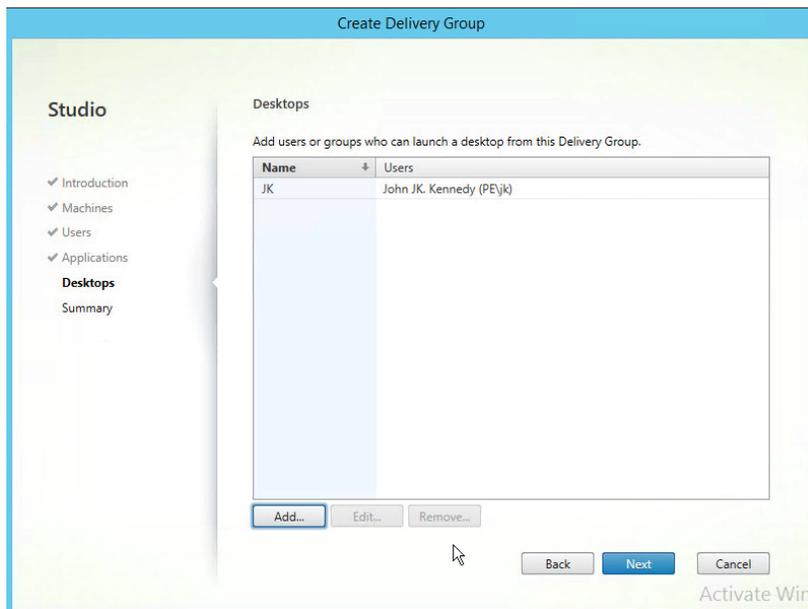
6. On the **Desktops** dialog box click **Add** to permit a user or group to launch desktops from this delivery group.



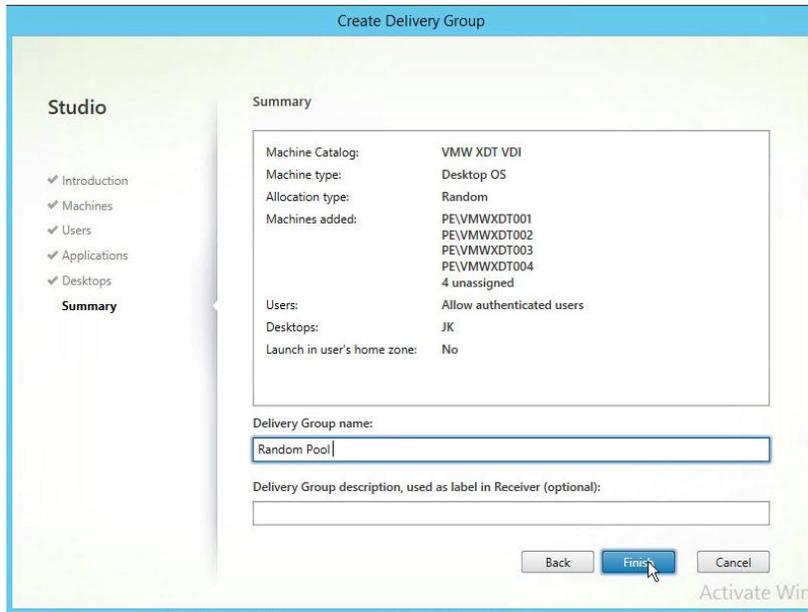
7. On the **Add Desktop** dialog box, give this person or group a display name, and select the **Restrict desktop use to** radio button. Then click **Add**.



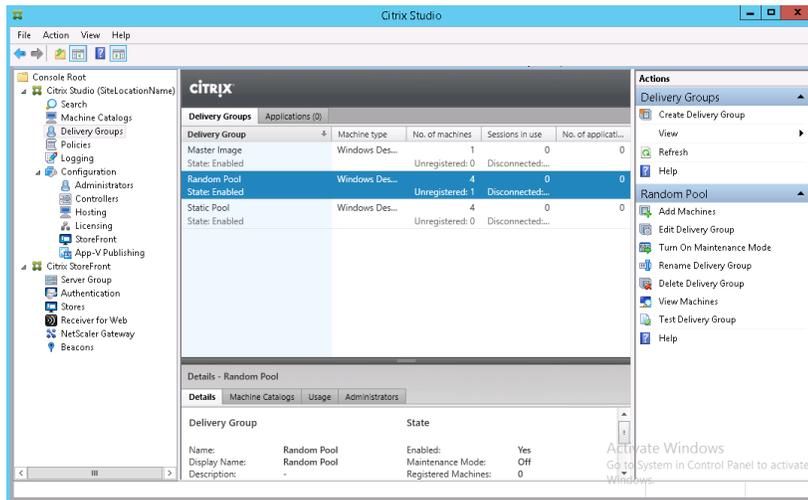
8. On the **Select Users and Groups** dialog search for and add the users and groups that will be granted access to the delivery group. Select the **OK** button to continue.



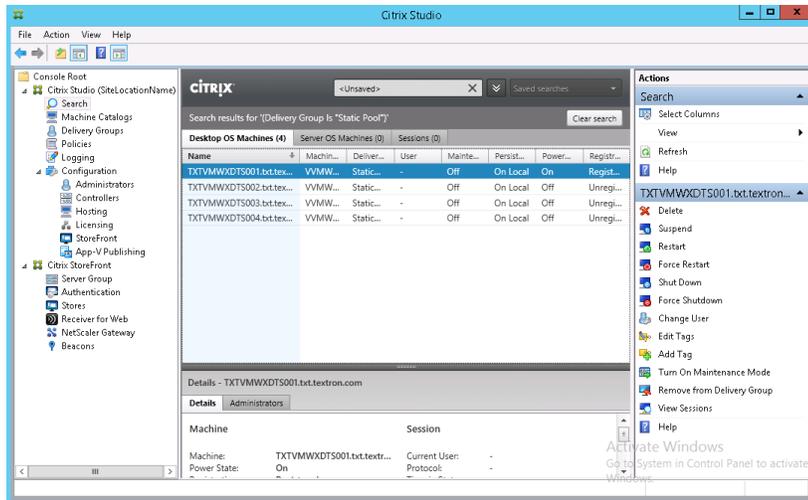
9. On the **Desktop** dialog select the **Next** button to continue.



- On the **Summary** dialog enter a unique **Delivery Group** name, and **Display Name** for the delivery group. Select the **Finish** button to complete the creation of the delivery group.



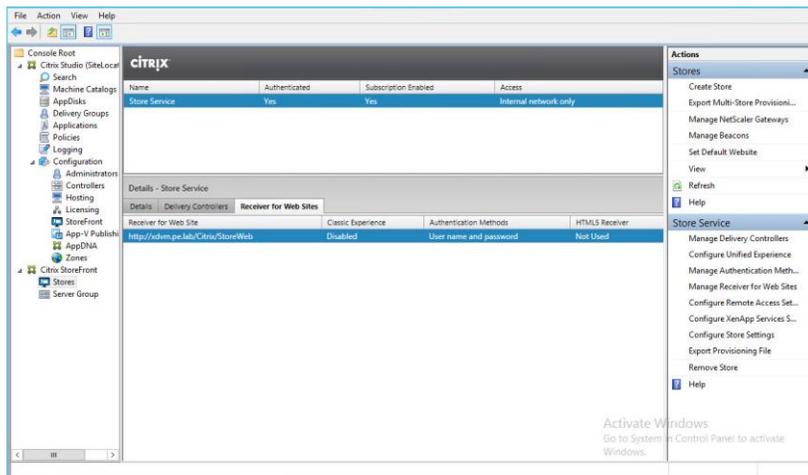
- On the **Delivery Groups** center pane, the newly created delivery group should now be available. Double-click on the **Delivery Group** row on the center pane to view the machines that are available for delivery.



12. The machines available for this **Delivery Group** will be listed in the center pane.

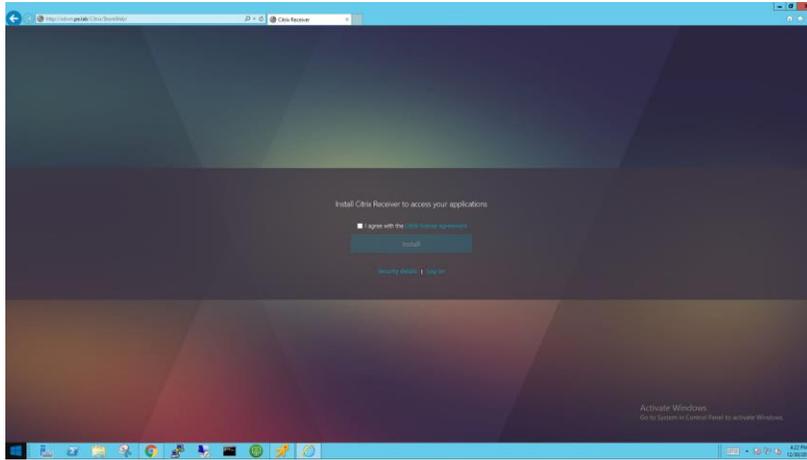
11.7 CONNECTING TO THE CITRIX DELIVERED DESKTOP FOR THE FIRST TIME

Perform the following procedure to access the Citrix delivered virtual machines for the first time.

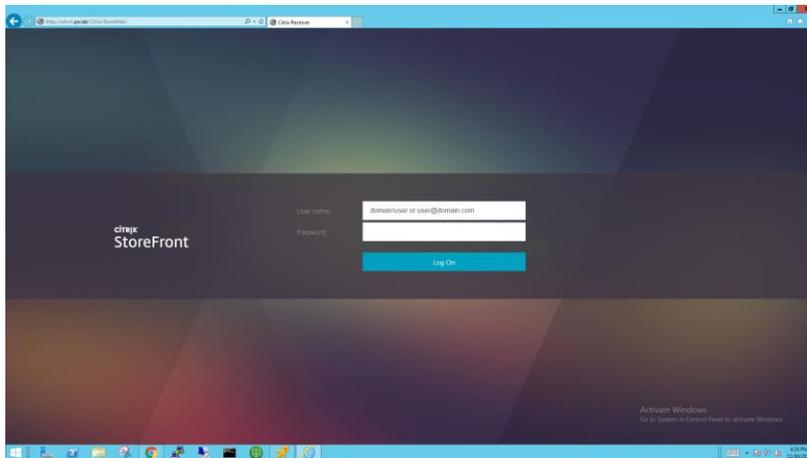


1. Using Internet Explorer or your favorite web browser navigate to the XenDesktop servers StoreFront URL. If you are not sure what this URL is, log in to the XenDesktop server and open **Citrix Studio**. On the left-hand pane, select the **Stores**

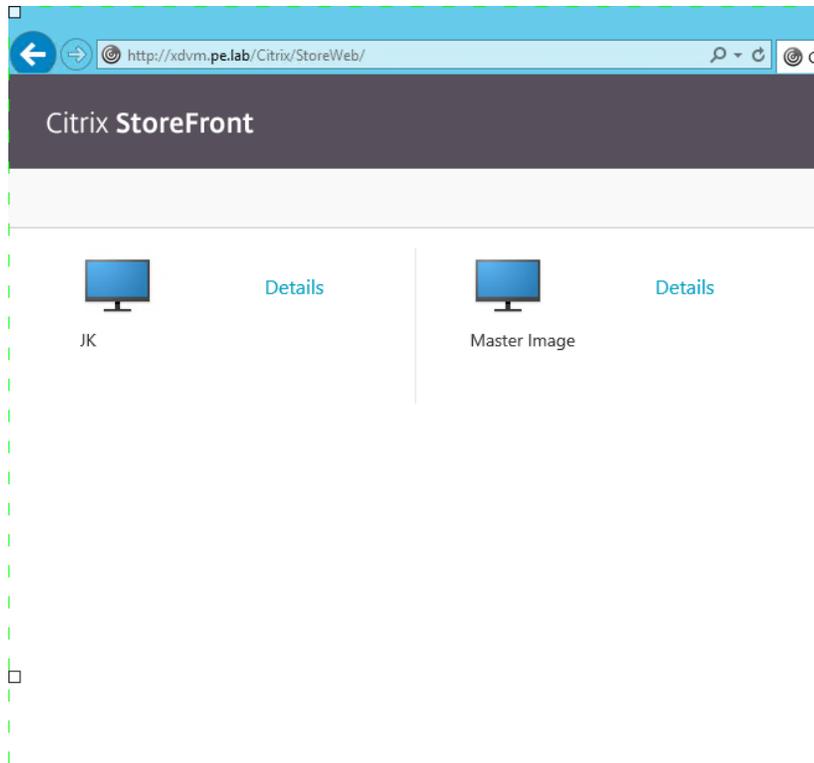
menu option. Under **Details** select **Receiver for Web Sites**. Note the **Website URL** displayed on the center pane.



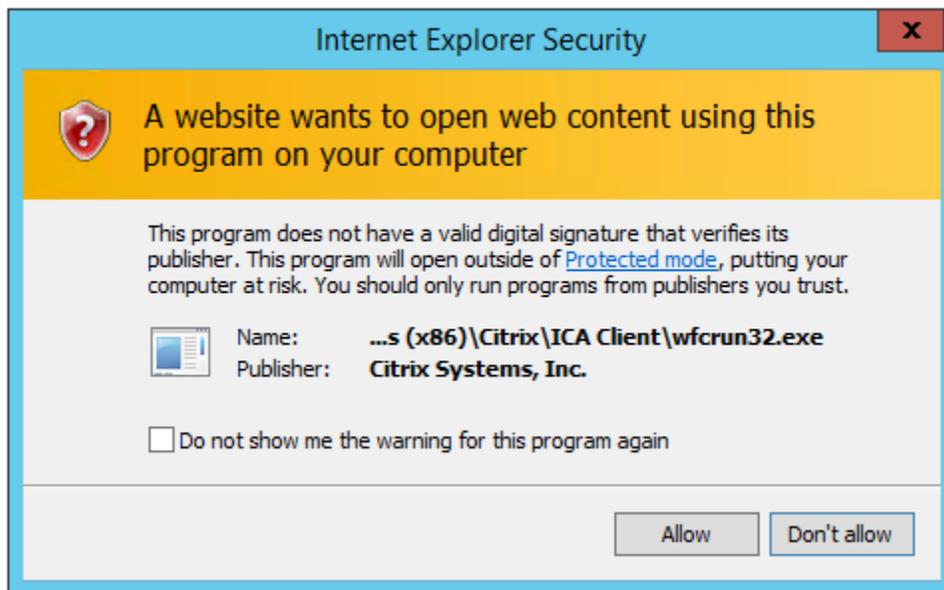
2. Open the Receiver for Web URL in a web browser. If Citrix Receiver client is not installed a prompt will be displayed to **Install** the client. Select the **I agree with the Citrix license agreement** check box. Select **Install** to begin installation of the Citrix Receiver client. Alternately select the **Log on** link to skip the installation and continue to **Receiver for Web**.



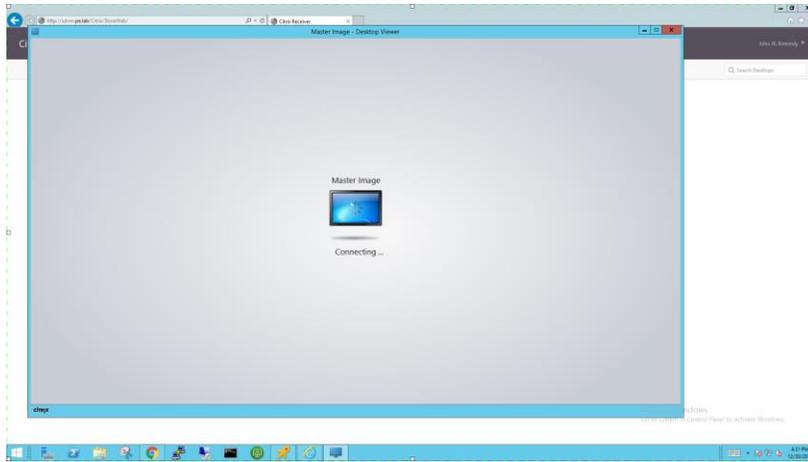
3. Enter **User name** (including domain account) and **password**, then select **Log On** to continue.



4. The Delivery Group resources that the logged in user has been granted access to will be displayed on the **Desktop** catalog page. Select the **Master Image** icon, to test that XenDesktop functionality is configured properly.



5. If an Internet Explorer Security dialog displays, select **Allow**.



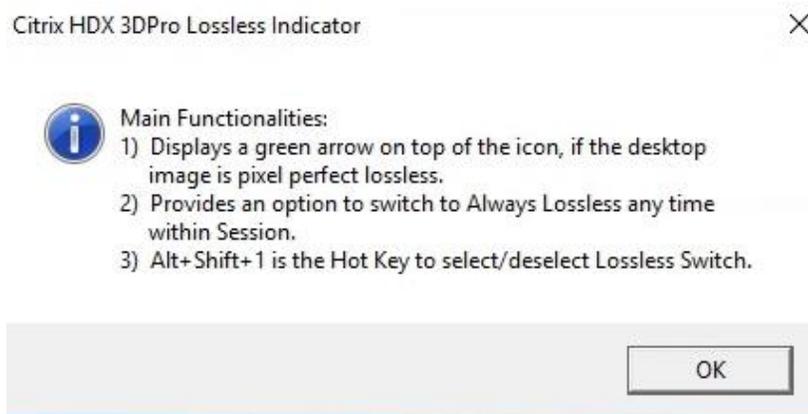
6. The **Desktop Viewer** will display with a connecting message.



7. When the image is fully loaded, the **Windows Desktop** will be displayed.



8. The drop down tab at the top can be accessed to enter **Full Screen** mode.



9. The Citrix HDX 3DPro Lossless Indicator notice is displayed when you change screen attributes such as entering full screen mode. This is indicating that the ICA session is being graphically accelerated with the HDX 3D Pro technology.

CHAPTER 12. TESTING, MONITORING, & BENCHMARKING

This chapter describes various means of testing and benchmarking your NVIDIA GRID installation, including:

- ▶ NVIDIA System Management Interface (nvidia-smi)
- ▶ Testing and simulation applications and tools

12.1 TESTING VIRTUAL DESKTOPS VS PHYSICAL ONES, IMPORTANT DIFFERENCES

Virtualization is a shared environment. It provides efficiencies and administrative ease that are impossible with bare metal installs. Where a physical workstation or PC is one solution to the problem of making professional visualization applications available, virtualizing those same workloads is a very different solution. The testing and benchmarking must be viewed accordingly. In a physical workstation, the connection to the user is immediate; screens are delivered directly to a monitor, and keystrokes and mouse operations are conducted directly to the system. In a virtualized system, the video must travel across networks to devices at the user's location. Where the goal on bare metal is to maximize performance at that desktop, with virtual desktops the goal is to maximize performance all the way across the network, to whatever device, over bandwidth and latency constraints that the end user is experiencing.

12.2 NVIDIA SYSTEM MANAGEMENT INTERFACE

The NVIDIA System Management Interface (`nvidia-smi`) is a command line utility based on the NVIDIA Management Library (NVML). The NVML helps you manage and monitor NVIDIA GPU devices by allowing you to query and (if you have the appropriate privileges) modify the GPU device state. Information can be reported in either XML or human-readable plain text using either standard or file output. Refer to the `nvidia-smi` documentation for more information.



`nvidia-smi` is designed for Tesla and Fermi-based Quadro devices; however, it does offer limited support for other NVIDIA GPUs as well.

`Nvidia-smi` ships with NVIDIA GRID vGPU Manager and with the graphics drivers for each OS.

12.2.1 Querying `nvidia-smi`

To obtain a system-wide overview of the NVIDIA GPUs, call the `nvidia-smi` command without any arguments:

```
$ nvidia-smi
```

In the following example system output, a eight virtual machines are utilizing a GRID Tesla M60 with the **M60-1Q vGPU** profile, and the GPU utilization for the GPU that are hosting the vGPU profiles is **11%** and **0%** respectively.



The `nvidia-smi` documentation is available at <http://developer.download.nvidia.com/compute/DCGM/docs/nvidia-smi-367.38.pdf>.

```
Tue Nov  1 01:00:56 2016
+-----+
| nvidia-smi 367.43                Driver Version: 367.43                |
+-----+-----+-----+-----+-----+-----+
| GPU  Name          Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+-----+-----+-----+-----+
|   0   Tesla M60           On      | 0000:05:00.0    Off  |           Off       |
| N/A   44C    P8         24W / 150W | 4520MiB / 8191MiB |    11%    Default  |
+-----+-----+-----+-----+-----+-----+
|   1   Tesla M60           On      | 0000:06:00.0    Off  |           Off       |
| N/A   40C    P8         23W / 150W | 3619MiB / 8191MiB |     0%    Default  |
+-----+-----+-----+-----+-----+-----+

+-----+
| Processes:                                     GPU Memory |
|  GPU       PID    Type   Process name                               Usage      |
+-----+-----+-----+-----+-----+-----+
|    0      108782  C+G    CLonevGPU001                               896MiB    |
+-----+-----+-----+-----+-----+-----+
```

	0	113970	C+G	LinkedClone002	896MiB	
	0	114062	C+G	LinkedClone004	896MiB	
	0	114313	C+G	LinkedClone006	896MiB	
	0	114314	C+G	LinkedClone007	896MiB	
	1	114059	C+G	LinkedClone001	896MiB	
	1	114060	C+G	LinkedClone005	896MiB	
	1	114061	C+G	LinkedClone003	896MiB	
	1	114827	C+G	CLonevGPU002	896MiB	
+-----+						

12.2.2 Monitoring

Adding command switches allows you to use `nvidia-smi` to monitor GPU performance metrics. To monitor the GPU utilization for the GPU hosting the vGPU (not the vGPU utilization) in the section 12.2.1, Querying `nvidia-smi` example on page 272, use the following command switches:

- ▶ `--id` is the PCI address of the device being queried. You can determine the address from the `nvidia-smi` output or by using the `gpudev` command.
- ▶ `--format` is the desired output format (CSV format, in this example).
- ▶ `--query` is the types of data you wish to query and output. Be sure to separate data using commas, and refrain from using spaces.
- ▶ `-l` is the number of seconds to loop the query (3 seconds in this example)
- ▶ `> Win7-c01-GPU.log` redirects the console output into a file.

Press **CTRL+c** to stop logging.

This following example logs GPU performance to the file `Win7-c01- GPU.log`:

```
$ nvidia-smi --id=0000:86:00.0 --format=csv --query-gpu=utilization.gpu,utilization.memory,memory.total,memory.used,memory.free -l 3 > Win7-c01-GPU.log
```

Begin collecting GPU metrics by launching an application. When finished, press **CTRL+c** to stop logging.

Use the following command to display the logging results:

```
$ cat Win7-c01-GPU.log
```

The system returns output similar to the following:

```
utilization.gpu [%], utilization.memory [%], memory.total [MiB],
memory.used [MiB], memory.free [MiB]
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
```

```

0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
0 %, 0 %, 8191 MiB, 919 MiB, 7272 MiB
42 %, 31 %, 8191 MiB, 920 MiB, 7271 MiB
39 %, 30 %, 8191 MiB, 920 MiB, 7271 MiB
5 %, 1 %, 8191 MiB, 920 MiB, 7271 MiB
4 %, 1 %, 8191 MiB, 920 MiB, 7271 MiB
21 %, 12 %, 8191 MiB, 920 MiB, 7271 MiB
23 %, 12 %, 8191 MiB, 920 MiB, 7271 MiB
18 %, 10 %, 8191 MiB, 920 MiB, 7271 MiB
81 %, 56 %, 8191 MiB, 920 MiB, 7271 MiB
90 %, 60 %, 8191 MiB, 920 MiB, 7271 MiB
92 %, 61 %, 8191 MiB, 920 MiB, 7271 MiB
82 %, 55 %, 8191 MiB, 920 MiB, 7271 MiB
92 %, 61 %, 8191 MiB, 920 MiB, 7271 MiB
89 %, 59 %, 8191 MiB, 920 MiB, 7271 MiB
90 %, 58 %, 8191 MiB, 920 MiB, 7271 MiB
95 %, 62 %, 8191 MiB, 920 MiB, 7271 MiB
92 %, 61 %, 8191 MiB, 920 MiB, 7271 MiB
82 %, 54 %, 8191 MiB, 920 MiB, 7271 MiB
94 %, 66 %, 8191 MiB, 920 MiB, 7271 MiB
5 %, 1 %, 8191 MiB, 921 MiB, 7270 MiB
2 %, 0 %, 8191 MiB, 921 MiB, 7270 MiB
    
```

You can use a graphing application to display the overall GPU utilization and GPU memory utilization.

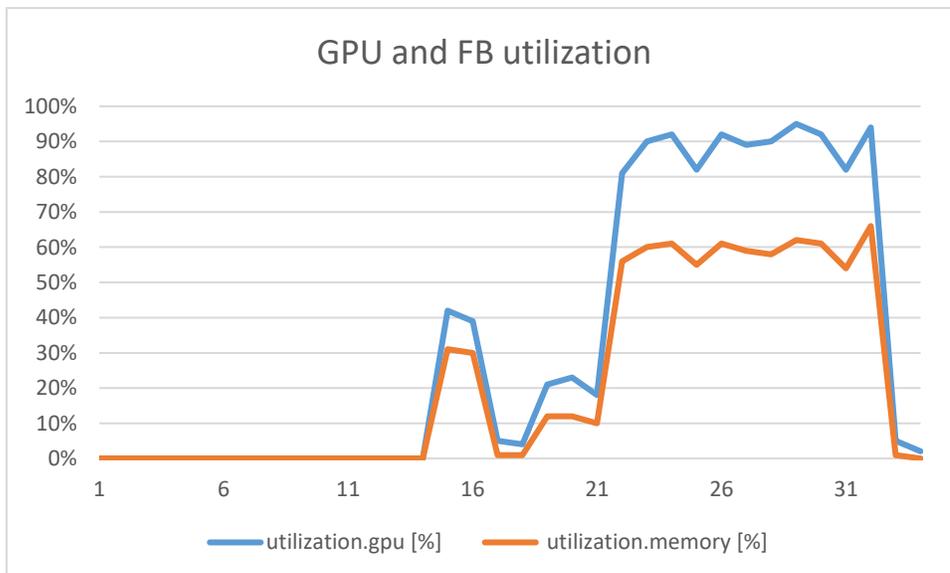


Figure 12-1 Graphed Log Output

12.3 TESTING/SIMULATING APPS AND TOOLS

This section contains numerous examples of tools and applications used for testing performance, benchmarking, and other verification purposes.

12.3.1 GPUProfiler

- ▶ <https://github.com/JeremyMain/GPUProfiler/releases>

12.3.2 Browsers

- ▶ HTML 5 (Fishbowl): <http://ie.microsoft.com/testdrive>
- ▶ Internet Explorer: <http://windows.microsoft.com/en-us/internet-explorer/download-ie>
- ▶ Chrome: <http://chrome.google.com>
- ▶ Firefox: www.mozilla.org/firefox
- ▶ Facebook: www.facebook.com
- ▶ NVIDIA: www.nvidia.com
- ▶ ESPN: www.espn.com

12.3.3 Microsoft

- ▶ Office 2013 (turn on and off GPU Acceleration): <http://www.microsoft.com/en-us/evalcenter/evaluate-office-professional-plus-2013>
- ▶ Office 2016: <https://products.office.com>

12.3.4 Dassault Systèmes

- ▶ SOLIDWORKS: <http://www.solidworks.com/sw/purchase/solidworks-trial.htm>
- ▶ CATIA: <http://www.3ds.com/products-services/catia>

12.3.5 PTC

- ▶ Creo: <http://www.ptc.com/product/creo>

12.3.6 Siemens

- ▶ Siemens NX: http://www.plm.automation.siemens.com/en_us/products/nx/

12.3.7 OpenGL

- ▶ FurMark: <http://www.ozone3d.net/benchmarks/fur>
- ▶ Unigine: <http://unigine.com/products/benchmarks/>
- ▶ Heaven (newest): <http://unigine.com/products/heaven/>
- ▶ Valley: <http://unigine.com/products/valley/>
- ▶ Tropics: <http://unigine.com/en/products/benchmarks/tropics>
- ▶ Sanctuary: <http://unigine.com/en/products/benchmarks/sanctuary>

12.3.8 Lego Designer

- ▶ Lego Designer: <http://ldd.lego.com/en-us/>

12.3.9 Adobe

- ▶ Reader: <http://get.adobe.com/reader>
- ▶ Acrobat: http://www.adobe.com/Acrobat_XI
- ▶ Creative Cloud Suite: <http://www.adobe.com/CreativeCloud>
- ▶ Photoshop (with high-resolution images): <http://www.adobe.com/Photoshop>
- ▶ Illustrator: <http://www.adobe.com/Illustrator>

12.3.10 BitSquid Stone Giant

- ▶ Bitsquid Stone Giant: http://international.download.nvidia.com/geforce-com/international/techdemos/stone_giant_setup.exe

12.3.11 Google Earth

- ▶ Locally installed: <https://www.google.com/earth/explore/products/desktop.html>
- ▶ Chrome delivered: <https://www.google.com/earth/explore/products/plugin.html>

12.3.12 Hulu/YouTube videos

- ▶ Try an assortment of videos with dialog, as this is critical for the video playback test (making sure lips are moving in time with the audio).

12.3.13 Autodesk

- ▶ AutoCAD: <http://www.autodesk.com/products/autocad/free-trial>
- ▶ Revit: <http://www.autodesk.com/products/revit-family/free-trial>
- ▶ Revit samples: <http://grabcad.com/library/software/autodesk-revit>

12.3.14 Slicer3D

- ▶ 3D Slicer: <http://download.slicer.org/>

12.3.15 FutureMark

- ▶ 3DMark: <http://www.futuremark.com/benchmarks/3dmark>
- ▶ PCMark: <http://www.futuremark.com/benchmarks/pcmark8>

CHAPTER 13.

KNOWN ISSUES

This section lists some of the known issues that are current as of the publication date of this manual.

13.1 UNINSTALLING EXISTING VSGA VIB TO INSTALL THE NVIDIA GRID VGPU VIB

You must uninstall an existing NVIDIA VIB file *before* installing a new vSGA or vGPU VIB. As of this publication, the NVIDIA VIBs for vSGA and vGPU are mutually exclusive.

To uninstall the NVIDIA VIB file for the purposes of upgrading the vib, issue the following command:

```
esxcli software vib remove -n NVIDIA-vgx-VMware_vSphere_6_Host_Driver
```

The output should be:

```
Removal Result
  Message: Operation finished successfully.
  Reboot Required: false
  VIBs Installed:
  VIBs Removed: NVIDIA_bootbank_NVIDIA-vGPU-
VMware_ESXi_6.0_Host_Driver_367.43-1OEM.600.0.0.2494585
  VIBs Skipped:
```

To verify the installation status of the NVIDIA VIB, issue the following command in a vSphere shell:

```
esxcli software vib list | grep NVIDIA
```

- ▶ If an NVIDIA VIB is not installed, the command will not return any output.
- ▶ If an NVIDIA VIB is installed, the output will be:

```
NVIDIA_bootbank_NVIDIA-vGPU-VMware_vSphere_6_Host_Driver_367.43-
1OEM.600.0.0.2494585
NVIDIA_VMwareAccepted 2016-11-22
```

13.2 NVIDIA WDDM ISSUES

After installing the NVIDIA WDDM driver in a VM and connecting with Horizon Client, the desktop will not be visible from the Host Client console.

- ▶ Without a console, the vSphere Web Client can still provide the IP address of a VM.
- ▶ With Horizon Agent Direct Connection, a Horizon Client can connect directly to the IP address of a VM.

13.3 vGPU-ENABLED VMS AND NVIDIA-SMI FAIL TO START

If vGPU-enabled VMs are assigned too high a proportion of the server's total memory, one or more of the VMs may fail to start with the error "**The available memory resources in the parent resource pool are insufficient for the operation**", and nvidia-smi run in the host shell returns the error "**-sh: can't fork**".

For example, on a server configured with 256GB of memory, these errors may occur if vGPU-enabled VMs are assigned more than 243GB of memory. Reduce the total amount of system memory assigned to the VMs.

13.4 VMOTION AND DRS NOT SUPPORTED

Virtual machines utilizing vGPU, or vDGA/pass-through, or SR-IOV, do not support vMotion or similar live motion functionality due to the dependence on a physical hardware resource. Cold motion, shutting down the virtual machine and moving it to another identically GPU equipped host, works well.

CHAPTER 14.

TROUBLESHOOTING

This section contains procedures for resolving common issues.



Before troubleshooting or filing a bug report, review the release notes for information about known issues with the current release, and potential workarounds.

14.1 TROUBLESHOOTING PROCEDURE

If a vGPU-enabled virtual machine either fails to start or does not display any output when it does start, follow these procedures to narrow down the probable cause.

14.1.1 Confirm the GPU Installation

Run the following command on the vSphere host to confirm that the graphics adapter has been physically installed correctly, and that vSphere is aware of the device:

```
$ lspci | grep -i display
```

Output should look similar to the following:

```
0000:0f:00.0 Display controller: Matrox Electronics Systems Ltd. MGA  
G200e [Pilot] ServerEngines (SEP1)  
0000:86:00.0 Display controller: NVIDIA Corporation NVIDIATesla M60  
[vmgfx0]  
0000:87:00.0 Display controller: NVIDIA Corporation NVIDIATesla M60  
[vmgfx1]
```

In this example, the bottom two devices `86:00.0` and `87:00.0` are associated with the Tesla M60 card.

Run the following command to list details about the GPUs listed above:

```
$ esxcli hardware pci list -c 0x0300 -m 0xff
```

You should see output like the following for each of the device class `0x300` (**Display**):

```
0000:86:00.0
  Address: 0000:86:00.0
  Segment: 0x0000
  Bus: 0x86
  Slot: 0x00
  Function: 0x0
  VMkernel Name: vmgfx0
  Vendor Name: NVIDIA Corporation
  Device Name: NVIDIATesla M60
  Configured Owner: Unknown
  Current Owner: VMkernel
  Vendor ID: 0x10de
  Device ID: 0x13f2
  SubVendor ID: 0x10de
  SubDevice ID: 0x115e
  Device Class: 0x0300
  Device Class Name: VGA compatible controller
  Programming Interface: 0x00
  Revision ID: 0xa1
  Interrupt Line: 0x0b
  IRQ: 255
  Interrupt Vector: 0x37
  PCI Pin: 0x00
  Spawned Bus: 0x00
  Flags: 0x0201
  Module ID: 21
  Module Name: nvidia
  Chassis: 0
  Physical Slot: 4294967295
  Slot Description: SlotID:5; relative bdf 01:00.0
  Passthru Capable: true
  Parent Device: PCI 0:133:8:0
  Dependent Device: PCI 0:134:0:0
  Reset Method: Bridge reset
  FPT Sharable: true
```

14.1.2 Setting the GPU Graphics Mode

To ensure that the GPUs are set to the proper GPU mode, complete the following procedure.



CAUTION: The gpumodeswitch VIB and the GPU Driver VIB cannot be simultaneously installed. Doing so prevents the gpumodeswitch driver from operating properly.

1. If the GPU Driver VIB is currently installed, uninstall it. Refer to Section 6.2 *Uninstalling the VIB* on page 135
2. Upload the gpumodeswitch software VIB to the ESXi server. Refer to Section 6.1.1 *Uploading VIB in vSphere* on page 128
3. Place the server in Maintenance Mode. See Section 6.1.2 *Place the Host in Maintenance Mode* on page 133
4. Issue the following command to install the gpumodeswitch VIB, where **<path>** is the full path to the VIB file:

```
$ esxcli software vib install --no-sig-check
-v /<path>/NVIDIA-GpuModeSwitch-1OEM.600.0.0.2494585.x86_64.vib
```

The following error displays if you *do not* use the full path:
[VibDownloadError] .

When the installation is successful, the following displays:

```
Installation Result:
Message: Operation finished successfully. Reboot Required: false
VIBs Installed: NVIDIA_bootbank_NVIDIA-
VMware_ESXi_6.0_GpuModeSwitch_Driver_1.0-1OEM.600.0.0.2494585
VIBs Removed:
VIBs Skipped:
```



Although the display states “Reboot Required: false”, a reboot is necessary for the vib to load.

5. Reboot the host by typing **Reboot** and pressing **ENTER**.
6. Once the server has rebooted, log in to the console and execute the following command:

```
$ gpumodeswitch --listgpumodes
```

The following output (or similar) will be displayed.

```
NVIDIA GPU Mode Switch Utility Version 1.23.0
```

```
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.
```

```
Tesla M60 (10DE,13F2,10DE,115E) H:--:NRM
S:00,B:05,PCI,D:00,F:00
Adapter: Tesla M60 (10DE,13F2,10DE,115E) H:--:NRM
S:00,B:05,PCI,D:00, F:00
```

```
Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
GPU Mode: Graphics
```

```
Tesla M60 (10DE,13F2,10DE,115E) H:--:NRM
S:00,B:06,PCI,D:00,F:00
Adapter: Tesla M60 (10DE,13F2,10DE,115E) H:--:NRM
S:00,B:06,PCI,D:00, F:00
```

```
Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
GPU Mode: Graphics
```

7. If the GPU is not in Graphics mode, set all available GPU to graphics mode by entering the following command:

```
$ gpumodeswitch --gpumode graphics
```

8. When prompted, type **y** to confirm the mode switch.

```
NVIDIA GPU Mode Switch Utility Version 1.02 Copyright (C) 2015, NVIDIA
Corporation. All Rights Reserved.
```

```
Update GPU Mode of all adapters to "graphics"?
```

```
Press 'y' to confirm or 'n' to choose adapters or any other key to
abort: y
```

You should see output like the following:

```
Updating GPU Mode of all eligible adapters to "graphics"
```

```
Tesla M60 (10DE,13F2,10DE,115E) H:--:NRM
S:00,B:05,PCI,D:00,F:00
Adapter: Tesla M60 (10DE,13F2,10DE,115E) H:--:NRM
S:00,B:05,PCI,D:00,F:00
```

```
Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
```

```
Programming UPR setting for requested mode..
License image updated successfully.
```

```
Programming ECC setting for requested mode..
```

The display may go *BLANK* on and off for up to 10 seconds or more during the update process depending on your display adapter and output device.

```
Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
.....
Verifying update...
Update successful.
```

Firmware image has been updated from version 84.04.85.00.03 to 84.04.85.00.03.

A reboot is required for the update to take effect.

InfoROM image updated successfully.

```
Tesla M60 (10DE,13F2,10DE,115E) H:--:NRM
S:00,B:06,PCI,D:00,F:00
Adapter: Tesla M60 (10DE,13F2,10DE,115E) H:--:NRM
S:00,B:06,PCI,D:00,F:00
```

```
Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
```

```
Programming UPR setting for requested mode..
License image updated successfully.
```

```
Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more
during the update process depending on your display adapter and output
device.
```

```
Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
.....
Verifying update...
Update successful.
```

Firmware image has been updated from version 84.04.85.00.04 to 84.04.85.00.04.

A reboot is required for the update to take effect.

```
InfoROM image updated successfully.
```

9. Uninstall the gpumodeswitch VIB by executing the following command:

```
$ esxcli software vib remove -n NVIDIA-  
VMware_ESXi_6.0_GpuModeSwitch_Driver
```

10. Reinstall the vGPU Manager VIB. Refer to section 6.1 *Prepare and Install the VIB Files* on page 128

14.1.3 Verify the VIB Installation

Run the following command using a shell on the host server to confirm that the VIB is installed:

```
$ esxcli software vib list | grep NVIDIA
```

If the vib is installed, then the output should look something like this:

```
NVIDIA-vGPU-VMware_ESXi_6.0_Host_Driver 367.43-1OEM.600.0.0.2494585  
NVIDIA VMwareAccepted 2016-11-23
```

14.1.4 Confirm VIB Loading

To confirm that the NVIDIA module has been loaded, run the following command:

```
$ esxcfg-module -l | grep nvidia
```

The result should be:

```
nvidia 2 11312
```

If there is no output, manually load the module by issuing the following command:

```
$ esxcli system module load -m nvidia
```

You can verify that the module loaded using the following command:

```
$ cat /var/log/vmkernel.log | grep NVRM
```

Include any NVRM specific output from `/var/log/vmkernel.log` in any support requests regarding module loading errors.



The `vmkernel.log` should be one of the first places you look when troubleshooting vGPU issues.

14.1.5 Confirm PCI Bus Slot Order

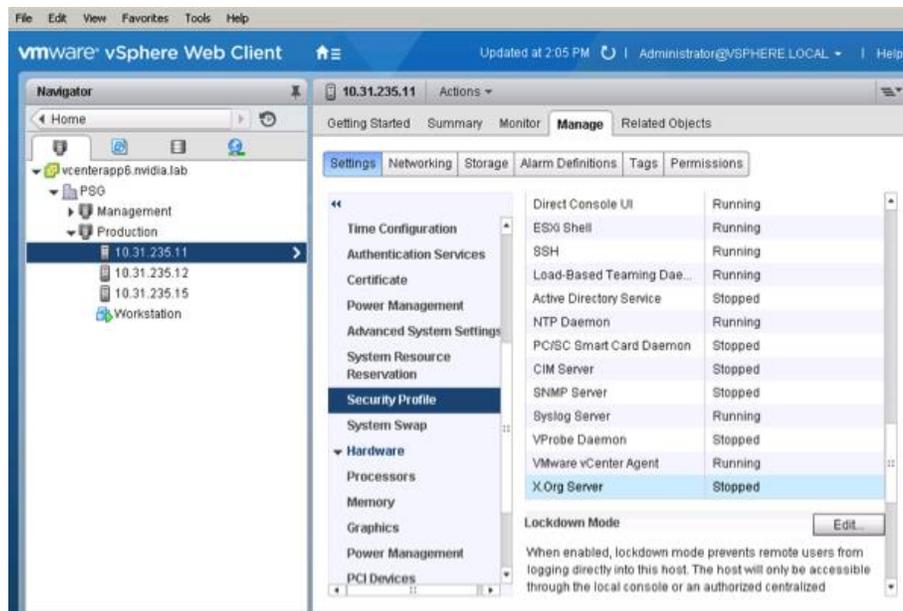
If you installed a second lower-end GPU in the server, it is possible that the order of the cards in the PCIe slots will choose the higher-end card for the vSphere console session. If

this occurs, either swap the two GPUs between PCIe slots or change the Primary GPU settings in the server BIOS.

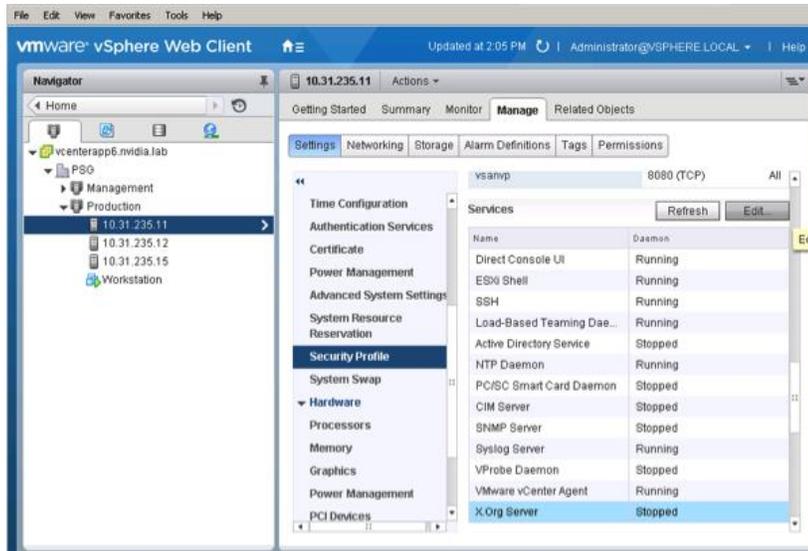
14.1.6 Check for X.Org Server Issues

If vGPU-enabled virtual machines fail to start, confirm that the X.Org server is functioning, as this is a fundamental service and necessary for vGPU. You can do this via the web UI of the vSphere Web Client, or via SSH to that vSphere host.

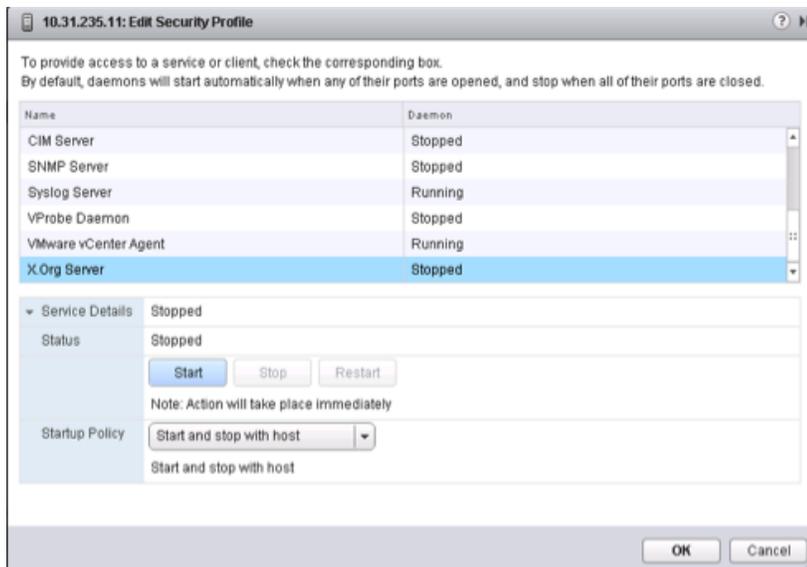
14.1.6.1 Managing X.Org Server via the Web Client



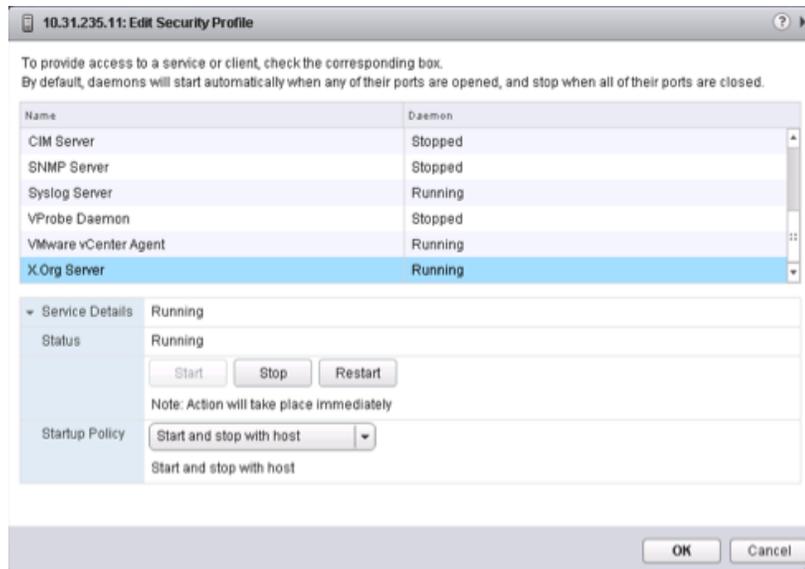
1. Using the VMware vSphere Web Client, browse to the host in question via the menu on the left. From **Home**, click **Hosts**, then drill down and single click on the specific host. Using the view on the main window to the right, click the **Manage** tab, then the **Settings** sub-tab, then scroll down then menu to the Security Profile section as shown below. On the Security Profile list, scroll down to the **Services** section, look for X.Org Server.



2. In the Services section you can see the current reported status of the X.Org Server service. To change or restart it, single click on **X.Org Server**, then click **Edit**.



3. The service edit window appears, showing current reported status and allowing you access to starting, stopping, restarting, and managing service behavior on server boot. For a vGPU host, you want the service to start and stop with the host. In the example below we show the service as stopped, click **Start**.



4. The window will refresh and the service should start.
5. Check that your vGPU guests can now start.

14.1.6.2 Managing X.Org via CLI

1. Using SSH, connect to the host running vGPU and log in.

 PuTTY is a useful Windows based tool to gain SSH access to a host.

2. Check the status of X.Org Server:

```
$ /etc/init.d/xorg status
```

If the X.Org server is operating correctly, the system displays the following output:

```
Xorg is running
```

If the X.Org server is not running, the system displays the following output:

```
Xorg is not running
```

3. Start the X server using the following command, if necessary:

```
$ /etc/init.d/xorg start
```

The system should return the following messages:

```
Xorg0 started
Xorg1 started
```

4. To troubleshoot an X.Org startup issue, examine the log file for NVIDIA- and/or GPU-related message using the following command:

```
$ cat /var/log/Xorg.log | grep -E "GPU|nv"
```

14.1.7 Check nvidia-smi

The NVIDIA System Management Interface (`nvidia-smi`) is the primary tool to see how much of each GPU is in use when using the NVIDIA driver. To see data about the physical card, run `nvidia-smi` on the host by issuing the following command in an SSH session:

```
$ nvidia-smi
```

This shows various details of GPU status and usage at the time the command is issued. (The data display is not dynamically updated).

The system returns output similar to the following:

```
Wed Nov 23 18:44:31 2016
+-----+
+-----+
| nvidia-smi 367.43                Driver Version: 367.43
|
|-----+-----+-----+-----+
+-----+
| GPU  Name           Persistence-M| Bus-Id           Disp.A | Volatile
Uncorr. ECC |
| Fan  Temp  Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util
Compute M. |
|====+=====+=====+=====+
=====|
|  0  Tesla M60             On      | 0000:86:00.0     Off   |
Off |
| N/A   43C    P8     24W / 150W |    19MiB /  8191MiB |    0%
Default |
+-----+-----+-----+-----+
+-----+
|  1  Tesla M60             On      | 0000:87:00.0     Off   |
Off |
| N/A   36C    P8     24W / 150W |    19MiB /  8191MiB |    0%
Default |
+-----+-----+-----+-----+
+-----+
+-----+
+-----+
+-----+
```

```

| Processes: GPU
Memory |
| GPU      PID  Type  Process name
Usage    |
|=====|
| No running processes found
|
+-----+
-----+ +

```

 The Volatile GPU-Util metric is the most meaningful for troubleshooting, which shows the percentage of the cores on each GPU that are busy at that point in time. This metric can help troubleshoot performance problems, such as by helping you determine whether a GPU is being overloaded.

nvidia-smi requires CPU and other system resources. Repeated loading and running of this command can cause a reduction in performance. Instead use the loop command by adding a -l to the command, such as # `nvidia-smi -l`

If you do not specify a delay value, then nvidia-smi will refresh every five seconds. See the nvidia-smi documentation for further usage information.

14.1.8 Resolving Black Screen Errors

This issue may occur with high-resolution displays or multi-monitor configurations. Use the following procedure to correct this issue:

1. Open vCenter.
2. Select the **Settings** tab for the virtual machine.
3. Select the video card.
4. Select **Specify custom settings**.
5. Adjust the video memory value higher (such as 32 MB, 64 MB, or 128 MB), as needed; see the Video Memory Calculator for recommended settings.

14.1.9 Multi-monitor Display Issues

Some older GRID K1 and GRID K2 BIOS do not support multi-monitor mode. Contact the vendor for updated BIOS and updating instructions, or upgrade to Tesla M series cards.

14.2 ADDITIONAL REMEDIATING STEPS

14.2.1 Enabling VM Console VNC Access

Enabling VNC support for the virtual machine console will negatively impact performance and should only be used for troubleshooting driver or Horizon connection issues when using vGPU within the virtual machine.

To enable VNC access for the virtual machine console, add the following to the virtual machine's .VMX file:

```
RemoteDisplay.vnc.enabled = TRUE
RemoteDisplay.vnc.port = 5901
RemoteDisplay.vnc.password = your-password
```

You will need to either manually open the port(s) in the virtual machine's firewall settings or disable it completely.



CAUTION: THESE INSTRUCTIONS ASSUME THAT THE VM IS BEING USED AS A PROOF-OF-CONCEPT ONLY AND THAT DISABLING THE FIREWALL WILL THEREFORE POSE ONLY A MINIMAL SECURITY BREACH. ALWAYS FOLLOW YOUR ESTABLISHED SECURITY PROCEDURES AND BEST PRACTICES WHEN SETTING UP SECURITY FOR A PRODUCTION MACHINE OR ANY ENVIRONMENT THAT CAN BE ACCESSED FROM OUTSIDE YOUR NETWORK.

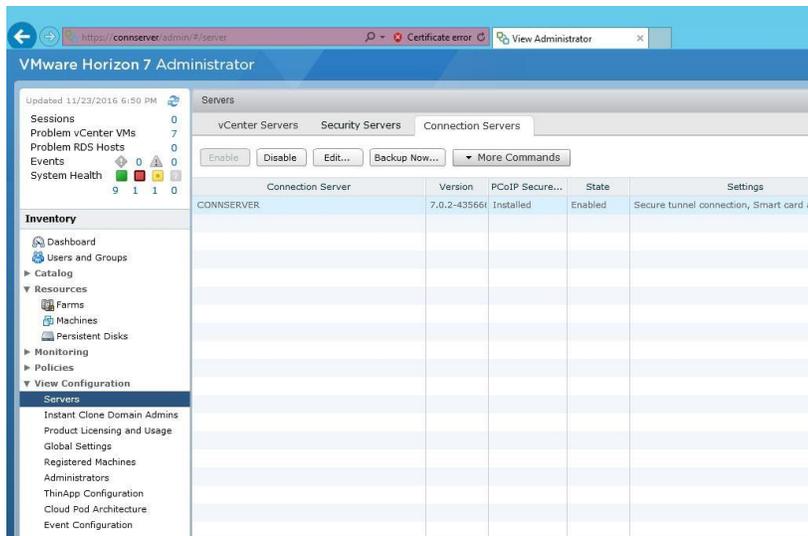
14.2.2 Joining an Endpoint to the Domain

If the endpoint is not joined to the domain, change the Horizon connection server settings to the IP address of the server (as opposed to the DNS name of the server), as follows:

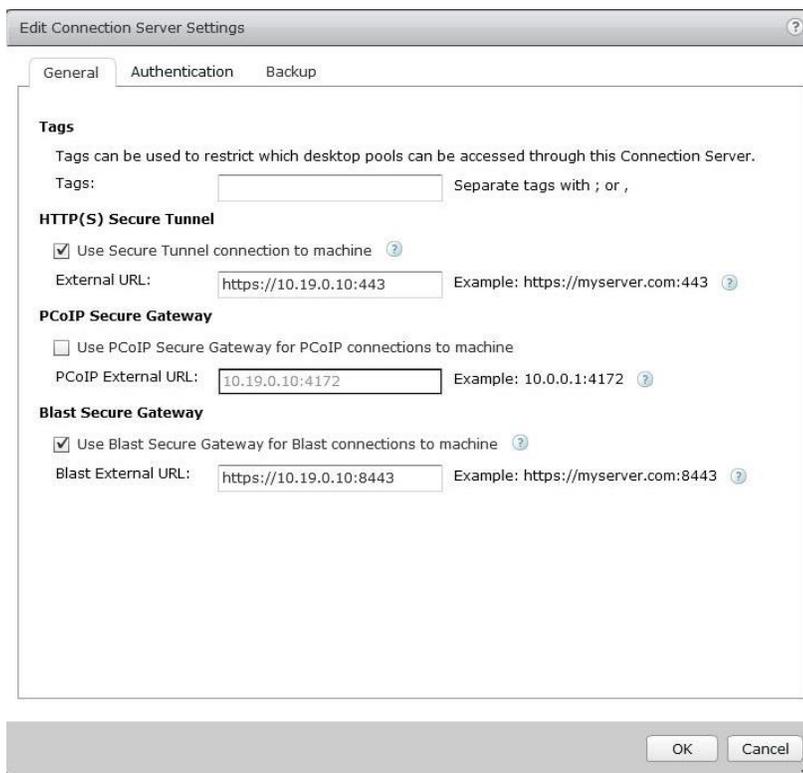
1. Select the server to edit in the VMware Horizon Client.
2. Enter your username and password in the displayed pop-up.
3. Select **Continue** to display an alert popup with the following message:

```
Error: Unable to resolve server address: nodename nor servname
provided, or not known.
```

4. Select **OK** to continue to the Horizon interface.
5. Select **Servers** in the Horizon interface,
6. Select the server to modify.



7. Select **Edit** to display the **Edit Connection Server Settings** window.



8. Change the hostnames to IP addresses in the **External URL** and **Blast External URL** fields. Select **OK** to return to the Horizon interface.

14.2.3 Filing a Bug Report

When filing a bug report or requesting support assistance from NVIDIA, it is critical to include information about the environment, so that the technical staff that can help you resolve the issue. Include in this information details on the host and virtual machine builds, software versions for host and virtual machines, the hardware, the impacted applications (with versions, patch levels, etc.) with screenshots or similar, and any service ticket numbers or similar help requests with other vendors.

NVIDIA includes the **nvidia-bug-report.sh** script within the VIB installation package to help collect and package some of this critical information. The script collects the following information:

- ▶ VMware version
- ▶ X.Org log and configuration
- ▶ PCI information
- ▶ CPU information
- ▶ GPU information
- ▶ **esxcfg** information for PLX devices
- ▶ **esxcfg** information for GPU devices
- ▶ VIB information
- ▶ NVRM messages from vmkernel.log
- ▶ System **dmesg** output
- ▶ Which virtual machines have vGPU or vSGA configured
- ▶ **nvidia-smi** output

When running this script:

- ▶ You may specify the output location for the bug report using either the **-o** or **-output** switch followed by the output file name. If you do not specify an output directory, the script will write the bug report to the current directory.
- ▶ If you do not specify a file name, the script will use the default name **nvidia-bug-report.log.gz**.
- ▶ If the selected directory already contains a bug report file, then the script will change the name of that existing report file to **nvidia-bug-report.log.old.gz** before generating a new **nvidia-bug-report.log.gz** file.

To collect a bug report, issue the command:

```
$ nvidia-bug-report.sh
```

The system displays the following message during the collection process:

```
nvidia-bug-report.sh will now collect information about your system and
create the file 'nvidia-bug-report.log.gz' in the current directory. It
may take several seconds to run. In some cases, it may hang trying to
capture data generated dynamically by the vSphere kernel and/or the
NVIDIA kernel module. While the bug report log file will be incomplete
if this happens, it may still contain enough data to diagnose your
problem.
```

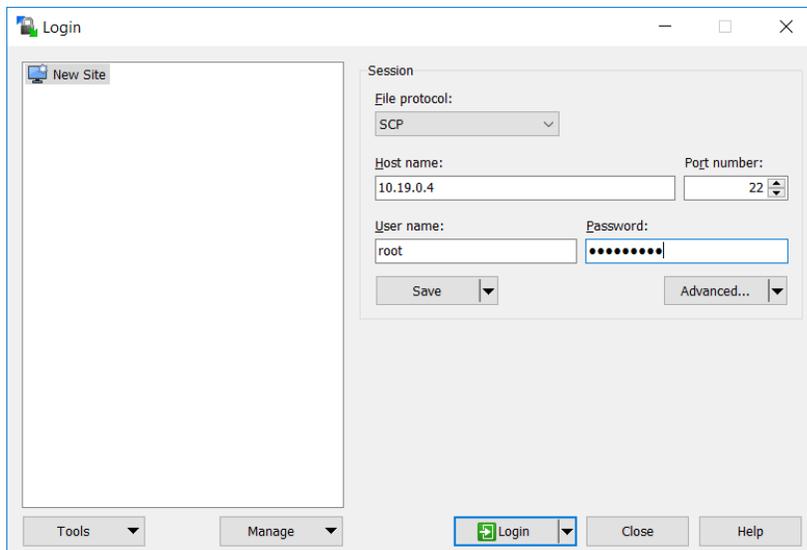
Be sure to include the **nvidia-bug-report.log.gz** file when reporting problems to NVIDIA.

CHAPTER 15. USING WINSCP

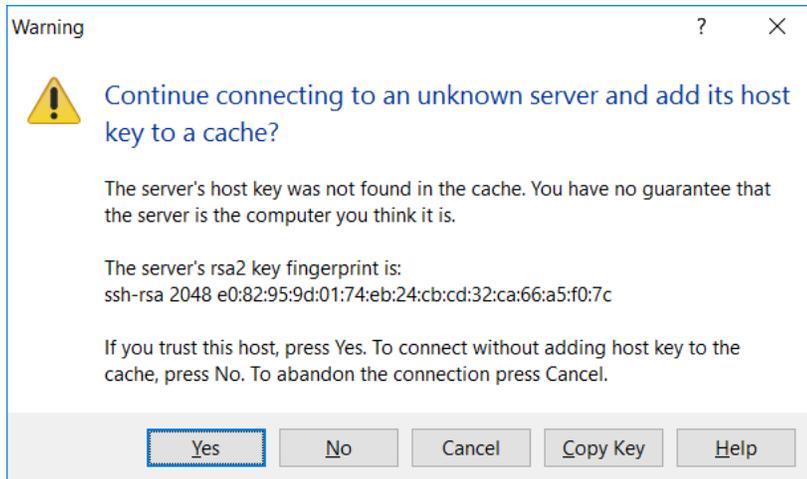
This section describes how to upload a VIB file using WinSCP. Before doing this:

- ▶ Ensure that SSH is enabled on vSphere.
- ▶ Download and install WinSCP on a Windows PC that has network connectivity to your vSphere host.

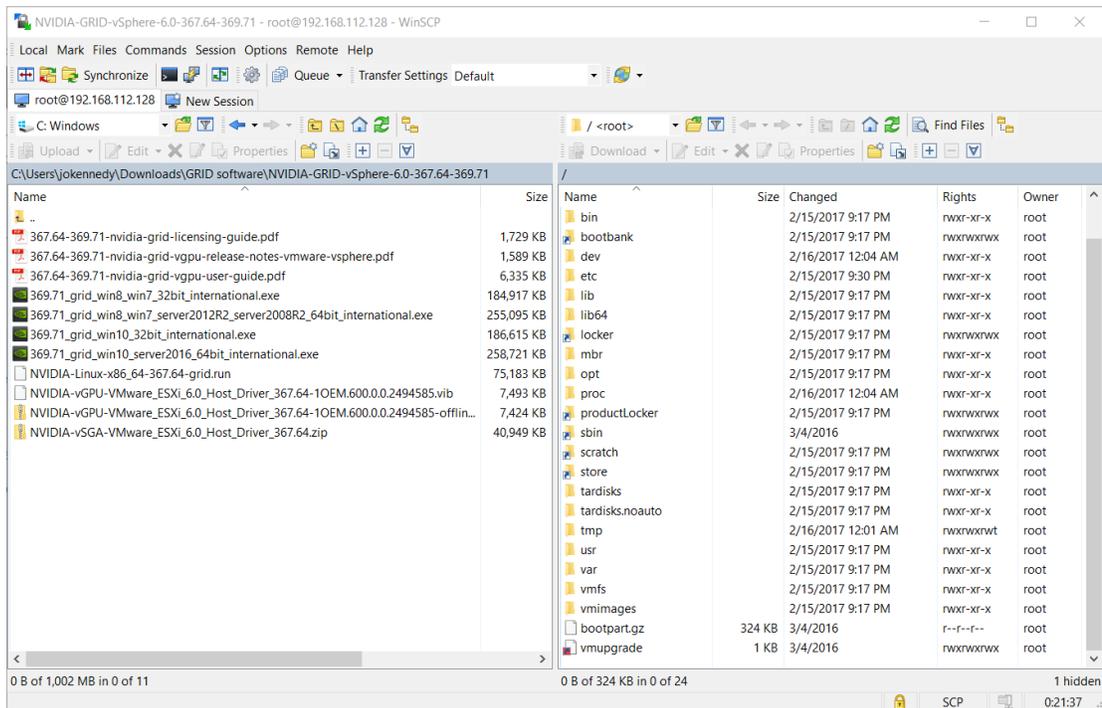
Use the following procedure to upload a VIB file using WinSCP:



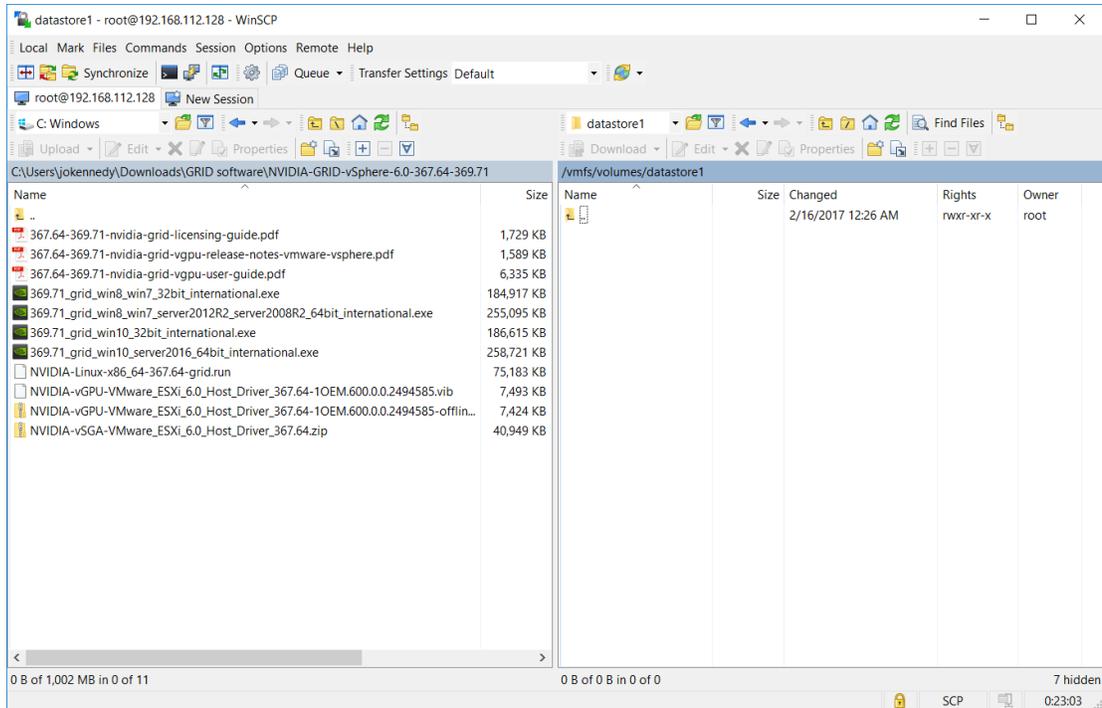
1. Start WinSCP to display the Login screen. Enter the connection details required to connect to the vSphere host.



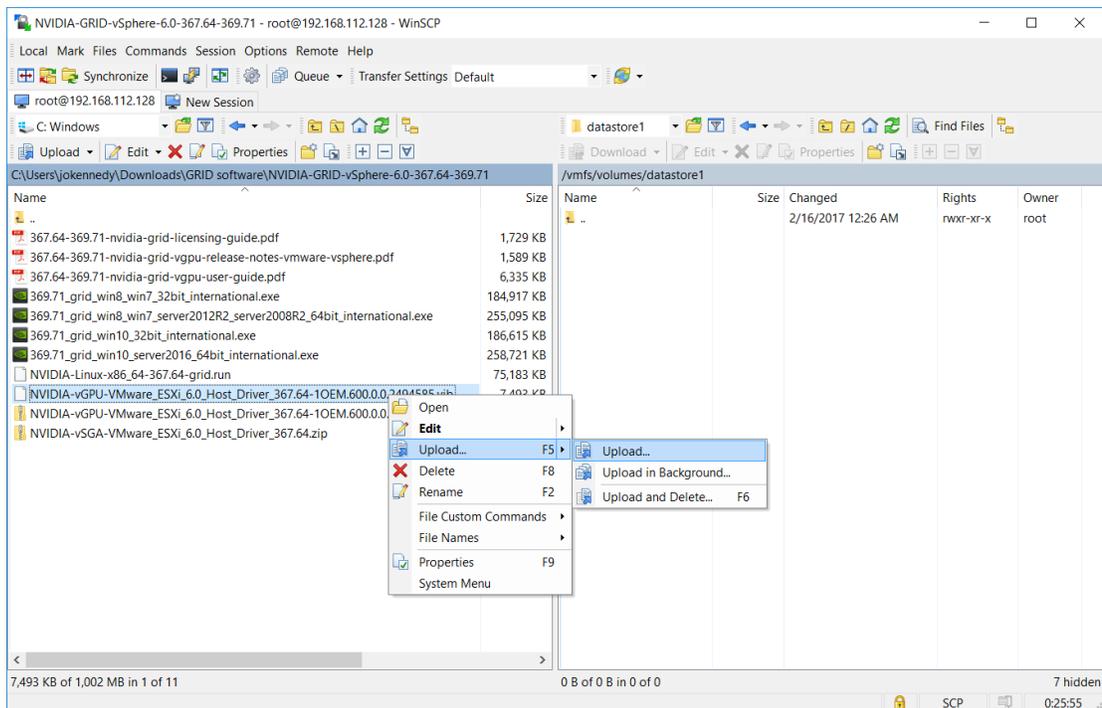
2. Select **Login** to display the **Commander** window. If this is your first time connecting to this server, a warning dialog will appear asking you to confirm the connection.



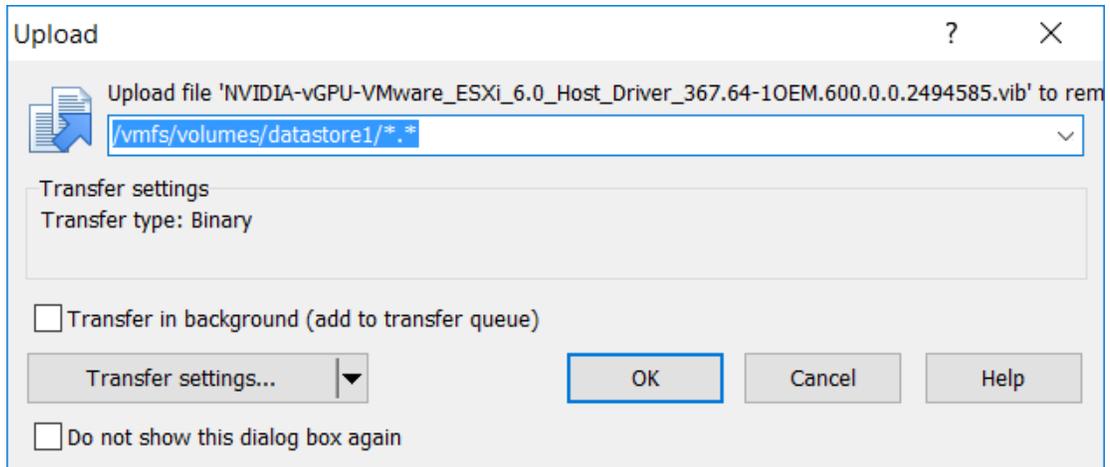
3. Once the connection is established, use the left pane of the WinSCP interface to navigate to the local folder containing the VIB file that you want to upload to the vSphere host.



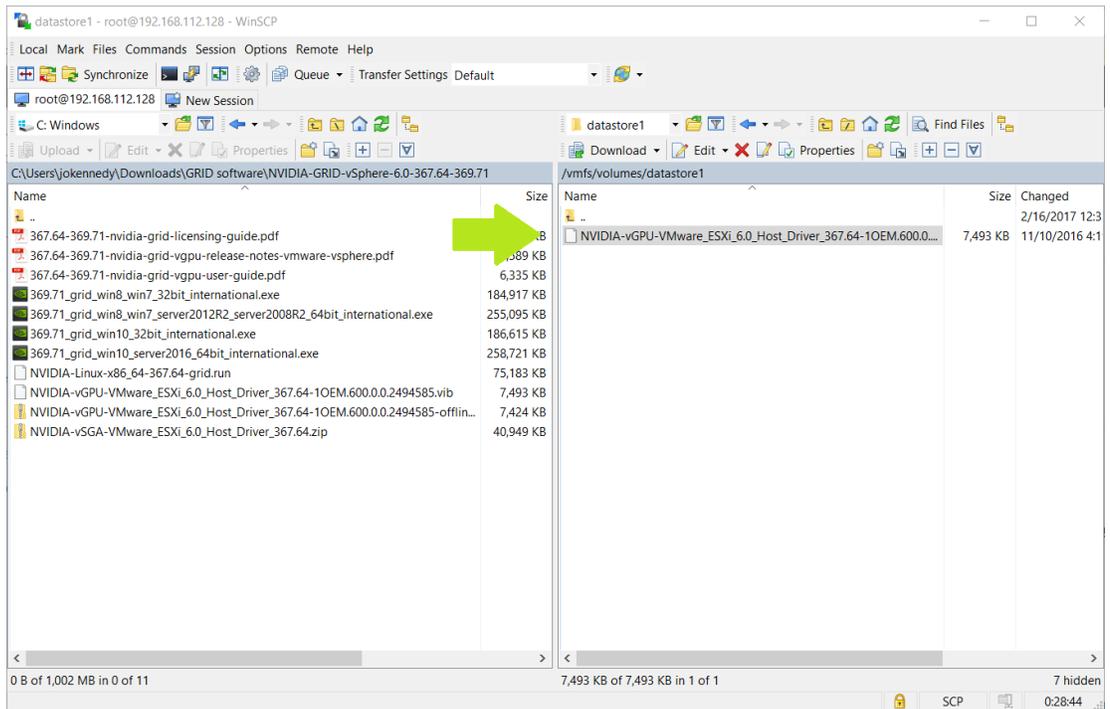
4. Navigate to the datastore using the right pane of the WinSCP interface.



5. Right-click the VIB file and select **Upload** to upload the file to the host.



6. If the Upload dialog box appears, click **OK**.



7. Verify that the VIB file has uploaded.

APPENDIX A. ACRONYMS

CLI	“Command Line Interface” - as opposed to using a something graphical, this is ASCII text based interaction with an app or OS
GPU	“Graphics Processing Unit” - the physical chip that accelerates the creation of graphical imagery in frame buffer for output on a display.
GUI	“Graphical User Interface” - a visual, windowed based means of interacting with an app or OS
OS	“Operating System” - a base level installation such as Windows that controls the underlying hardware (or virtual hardware)
OVA	“Open Virtualization Appliance” - an OVF bundled into a discrete and portable appliance
OVF	“Open Virtualization Format” - portable files representing discrete virtual machines, these can be bundled into OVA packages as a discrete appliance as well
POC	“Proof of concept”
UI	“User Interface” - refers to a tool used to enter data or otherwise interact with an application or OS
VCSA	“vCenter Server Appliance” - The management server for a VMware vSphere instance.
VDA	“Virtual Desktop Agent” - software that allows the VM to connect to remote Citrix Receivers.
VDI	“Virtual Desktop Infrastructure” - a general term for desktop operating systems running as a guest on a host server
vGPU	“Virtual Graphics Processing Unit” - a virtual GPU provided by the hypervisor and managed by the GRID vGPU Manager to provide VM’s access to the power of a physical GPU.
VIB	“vSphere Installation Bundle”
VM	“Virtual Machine” - an operating system running as a guest on a host server
vSGA	“Virtual Shared Graphics Acceleration”
WebUI	“Web User Interface” - a visual, windowed web based means of interacting with an app or OS

Notice

The information provided in this specification is believed to be accurate and reliable as of the date provided. However, NVIDIA Corporation (“NVIDIA”) does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This publication supersedes and replaces all other specifications for the product that may have been previously supplied.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and other changes to this specification, at any time and/or to discontinue any product or service without notice. Customer should obtain the latest relevant specification before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer. NVIDIA hereby expressly objects to applying any customer general terms and conditions with regard to the purchase of the NVIDIA product referenced in this specification.

NVIDIA products are not designed, authorized or warranted to be suitable for use in medical, military, aircraft, space or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer’s own risk.

NVIDIA makes no representation or warranty that products based on these specifications will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer’s product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this specification. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this specification, or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this specification. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA. Reproduction of information in this specification is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, “MATERIALS”) ARE BEING PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

VESA DisplayPort

DisplayPort and DisplayPort Compliance Logo, DisplayPort Compliance Logo for Dual-mode Sources, and DisplayPort Compliance Logo for Active Cables are trademarks owned by the Video Electronics Standards Association in the United States and other countries.

HDMI

HDMI, the HDMI logo, and High-Definition Multimedia Interface are trademarks or registered trademarks of HDMI Licensing LLC.

ROVI Compliance Statement

NVIDIA Products that support Rovi Corporation’s Revision 7.1.L1 Anti-Copy Process (ACP) encoding technology can only be sold or distributed to buyers with a valid and existing authorization from ROVI to purchase and incorporate the device into buyer’s products.

This device is protected by U.S. patent numbers 6,516,132; 5,583,936; 6,836,549; 7,050,698; and 7,492,896 and other intellectual property rights. The use of ROVI Corporation’s copy protection technology in the device must be authorized by ROVI Corporation and is intended for home and other limited pay-per-view uses only, unless otherwise authorized in writing by ROVI Corporation. Reverse engineering or disassembly is prohibited.

OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.

Trademarks

NVIDIA, the NVIDIA logo, NVIDIA GRID, and NVIDIA GRID vGPU™ are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2015 NVIDIA Corporation. All rights reserved.