NVIDIA GRID™ VGPU™ DEPLOYMENT GUIDE FOR CITRIX XENDESKTOP 7.6 ON VMWARE VSPHERE 6

DOCUMENT CHANGE HISTORY

Version	Date	Authors	Description of Change
01	05/2015	Jeff Weiss, Luke Wignall, & Fred Devoir	Initial release

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1.
ABOUT THIS DOCUMENT

This section contains an overview of the chapters, the formatting conventions and information contained in this deployment guide. There is also a quick list of acronyms used. Lastly, we detail the lab environment used to test the contents of this guide and capture the screenshots used.

1.1 ORGANIZATION

This manual contains the following chapters:

▶ Chapter 1 – About this Document: Explains the conventions used in this document, related documentation, contact information, and a list of acronyms.

▶ Chapter 2 – Getting Started: Introduces you to NVIDIA GRID™ technology and guides you through preparing to deploy NVIDIA® GRID™ vGPU™ with Citrix XenDesktop.

▶ Chapter 3 – vSphere 6 Installation: Guides you through preparing a bootable USB drive and installing vSphere 6 on the selected host.

▶ Chapter 4 – Installing VMware vCenter 6 Server: Instructions for installing and configuring VMware vCenter 6 Server, either Windows or OVA Linux virtual appliance.

▶ Chapter 5 – NVIDIA vGPU Manager VIB Installation: Step-by-step guide to downloading, preparing, uploading, installing, or uninstalling an NVIDIA vGPU Manager VIB file.

▶ Chapter 6 – Building Citrix XenDesktop 7.6: Instructions for installing and configuring Citrix XenDesktop 7.6 Broker.

- Chapter 7 – Building a vGPU vSphere Host Cluster: Adding vGPU enabled hosts to create a vGPU Cluster.

- Chapter 8 – Choosing a NVIDIA vGPU Profile: Understanding the range of available profiles and how to select the correct ones for your user groups.

- Chapter 9 – Application Based VDI Host sizing Example: ESRI ARCGIS

- Chapter 10 – Creating your first vGPU Virtual Desktop: Guides you through configuring an NVIDIA vGPU for a single guest desktop to create a template.

- Chapter 11 – Create a XenDesktop vGPU Pool: Using Citrix Studio to create a random pool of vGPU powered desktops using a single 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 processes for resolving common issues.

- Chapter 15 – Using WinSCP: Describes how to perform relevant functions of this SSF/SCP client.

1.2 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

1.3 RELATED DOCUMENTATION

Refer to http://www.nvidia.com/gridresources for additional information about NVIDIA GRID technology, including:

- ▶ NVIDIA GRID vGPU
- ▶ Lists of GRID-certified servers
- ▶ Datasheet
- ▶ Guides and Solution Overviews
- ▶ Case Studies
- ▶ White Papers
- ▶ Videos

1.4 ACRONYMS

CLI	"Command Line Interface" - as opposed to using a something graphical, this is ASCI 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 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
vDGA	"Virtual Direct Graphics Acceleration"
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 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

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 equipment:

NVIDIA Lab:

- Management Cluster
  - Hosts:
  - Supermicro 1U servers, local storage, 1GB NICs
  - Guests:
    - Microsoft Windows 2012 Domain Controller
    - Microsoft SQL Server 2008

- vCenter Server Appliance (VCSA), Linux based
- Microsoft Windows 2012 Servers for running:
  - View Composer
  - vCenter Server on Windows
- Production Cluster
  - Hosts:
    - Supermicro 2U servers, 2x NVIDIA GRID K2s, vSAN (SSD & spindle HDD), 1GB NICs, 10G NICs for iSCSI to Pure Storage SAN
  - Guests:
    - Microsoft Windows 7 Professional x64
    - Microsoft Windows 8.1
- SAN Storage
  - Pure Storage FA-420 (10G iSCSI)
- Networking
  - Arista 10G core switch
    - No LAG or similar
  - Juniper & Cisco 1GB distribution switches
- Load Generation
  - VMware View Planner 3.5
  - Login VSI
- Monitoring
  - Lakeside Systrack v7 (with GPU monitoring)

It is not necessary to have all of the equipment we used for building out your POC/trial environment; in fact it can be done on a single physical server. This is not, however, representative of a recommended enterprise deployment and can impact performance. To insure the most accurate results and performance you intend to deliver to your end users, build your POC/trial as close as possible to what it will look like once in production. Further, by separating the management and production (or virtual desktop) clusters, we were able to reinstall new versions of code while retaining foundational services and exported virtual machines.

Special Thanks:

Having access to fast cache storage allowed us to scale out and work with density to test the ability to handle greater quantities of users. We would like to thank Pure Storage for providing a SAN for us to build against, and the support of their Solution Architects who quickly brought it up and configured it. For more information see: http://www.purestorage.com/flash-array/

Chapter 2.
GETTING STARTED

NVIDIA GRID™ vGPU™ enables multiple virtual machines (VMs) to have simultaneous, direct access to a single physical GPU, using the same 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 brought about by sharing a GPU among multiple workloads.

This chapter covers how NVIDIA GRID vGPU fundamentally alters the landscape of desktop virtualization and enables users and applications of all levels of complexity and graphics requirements. It also describes the NVIDIA GRID vGPU architecture, the types of virtual GPUs supported, and key standards supported by GRID vGPU.

This guide focuses on the deployment of full virtual desktops as opposed to streamed apps or other solutions such as XenApp. XenApp is a key part of the XenDesktop solution but is not covered in this guide. A separate document will be published to ensure the proper detail is available.

2.1 WHY NVIDIA VGPU GRID?

The promise of desktop virtualization, realized for server workloads years ago, is flexibility and manageability. Initially, desktop virtualization was used where the flexibility and security were the primary drivers due to cost considerations. Over the last five years, the cost of desktop virtualization has been coming down quickly due to advances in storage and multi-core processors.

The big remaining challenge for desktop virtualization is providing a cost effective yet rich user experience. There have been attempts to solve this problem with software graphics or shared GPU technologies, but those technologies don't support the rich application support needed to be successful and ensure end user adoption. This compared to dedicated GPU pass-through that does provide 100% application compatibility, but only for the highest end use cases due to the high cost with limited density of virtual machines per host server.

Due to the lack of scalable, sharable, and cost effective per user GPUs that provide 100% application compatibility, providing a cost effective rich user experience has been challenging for broad use cases in desktop virtualization. Meanwhile, high-end 3D applications simply did not work in a virtualized environment, or were so expensive to implement with pass-thru it was reserved for only the most limited of circumstances.

Today, this is no longer true to thanks to NVIDIA GRID vGPU combined with Citrix XenDesktop running on VMware vSphere. NVIDIA GRID vGPU gives you the best of both worlds where multiple virtual desktops share a single physical GPU, and multiple GPUs on a single physical PCI card, all providing the 100% application compatibility of pass-through graphics, but with lower cost of multiple desktops sharing a single graphics card to provide a rich, yet more cost effective user experience. With XenDesktop you are able to centralize, pool, and more easily manage traditionally complex and expensive, distributed workstations and desktops. Now all your user groups can take advantage of the promise of virtualization.

2.2 NVIDIA GRID VGPU ARCHITECTURE

Physical NVIDIA GRID GPUs are capable of supporting multiple *virtual* GPU devices (vGPUs) that can be assigned directly to guest VMs under the control of NVIDIA's GRID Virtual GPU Manager running in VMware vSphere 6. Guest VMs use the GRID vGPUs in the same manner as 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 Virtual GPU Manager. Figure 2-1 shows the GRID vGPU system architecture.
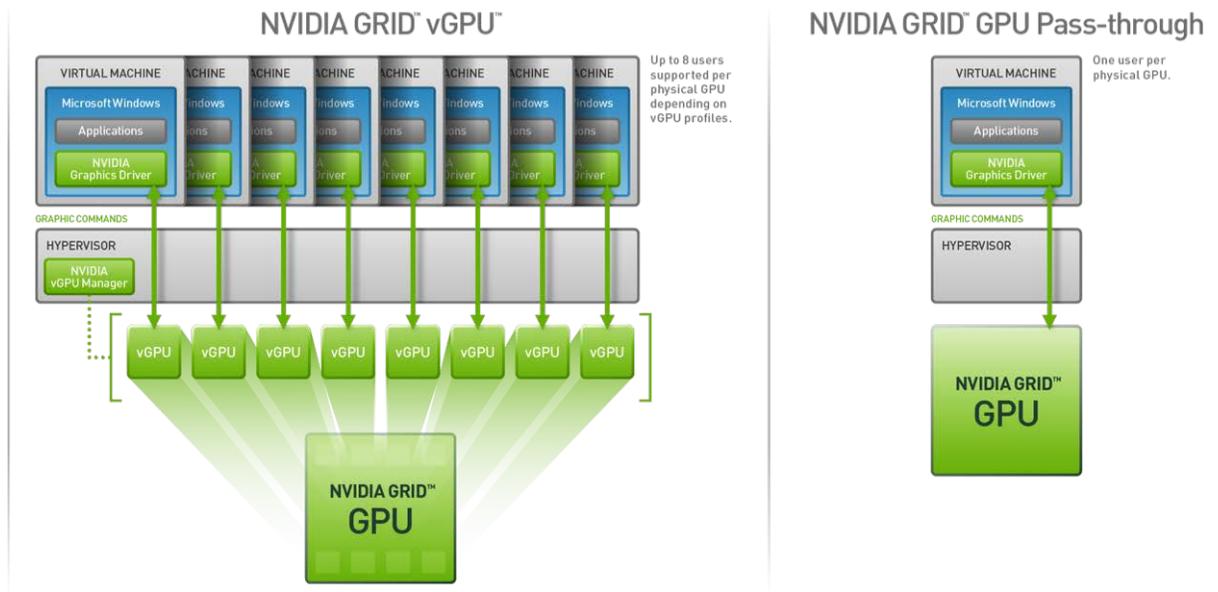
Figure 2-1.   GRID vGPU System Architecture

NVIDIA 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 vGPU Manager 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. Figure 2-2 shows the GRID vGPU internal architecture.
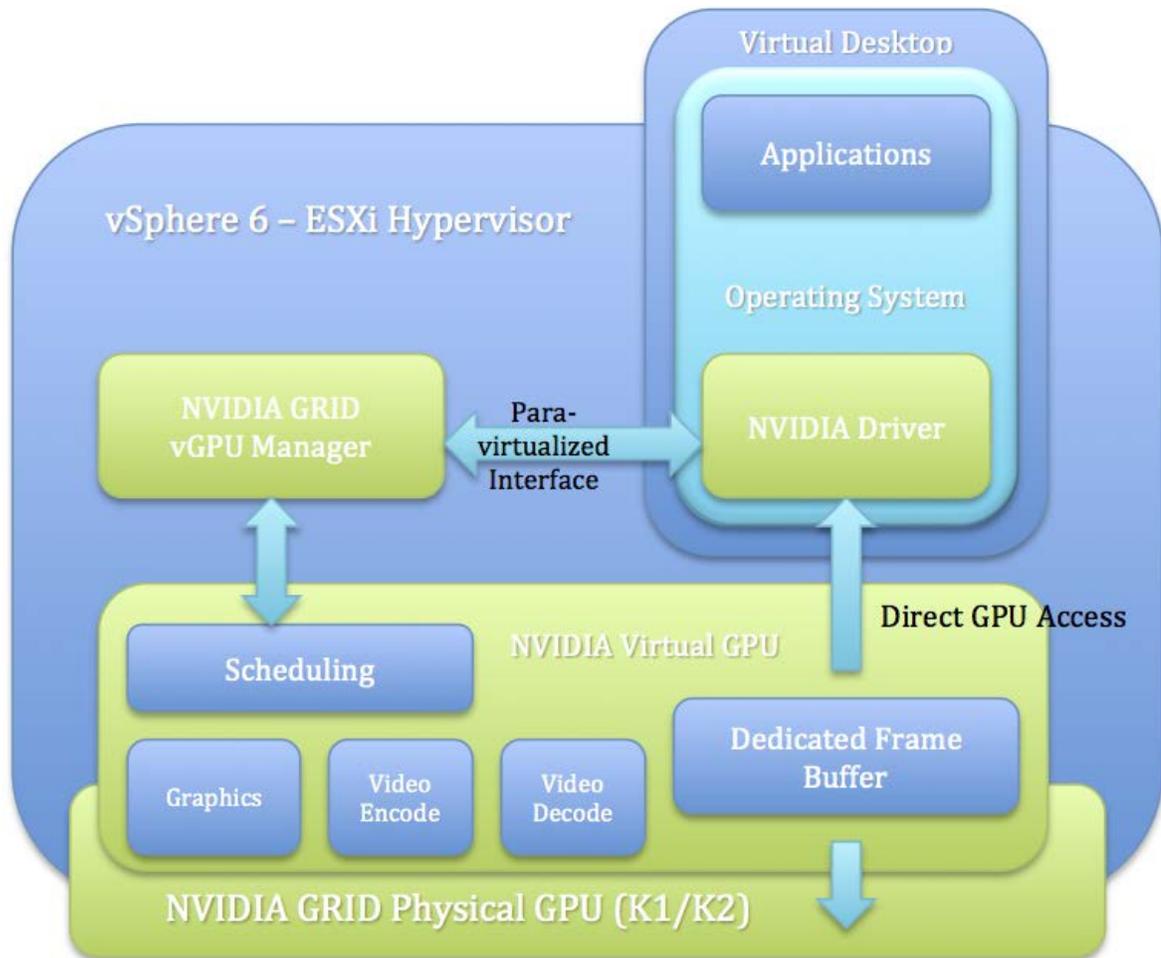
Figure 2-2.   GRID vGPU Internal Architecture

2.3 SUPPORTED GPUS

NVIDIA GRID vGPU is supported on NVIDIA GRID K2 and GRID K1 cards. GRID K2 and K1 each implement multiple physical GPUs:

▶ GRID K2 has two GPUs on board, best for performance over density.

▶ GRID K1 has four GPUs on board, best for density over performance.

Each physical GPU can support several different types of virtual GPU (vGPU). Each of the vGPU types has a fixed amount of frame buffer, number of supported display heads and maximum resolutions, and targeted at different classes of workload. Table 2-1 lists the virtual GPU types supported by GRID K1 and K2:

Table 2-1.   Virtual GPU Types

Card	Physical GPUs	Virtual GPU	Intended User(s)	Frame Buffer (MB)	Virtual Display Heads	Max Resolution	Maximum vGPUs	
							Per GPU	Per Board
GRID K2	Two	GRID K280Q	Designer	4096	4	2560x1600	1	2
		GRID K260Q	Designer	2048	4	2560x1600	2	4
		GRID K240Q	Power User	1024	2	2560x1600	4	8
		GRID K220Q	Power User	512	2	2560x1600	8	16
GRID K1	Four	GRID K180Q	Power User	4096	4	2560x1600	1	4
		GRID K160Q	Power User	2048	4	2560x1600	2	8
		GRID K140Q	Knowledge Worker	1024	2	2560x1600	4	16
		GRID K120Q	Knowledge Worker	512	2	2560x1600	8	32

The maximum number of vGPUs that can be created simultaneously on a physical GPU is eight, and each GPU has 4GB of frame buffer memory. Your selection of a profile defines the amount of graphics frame buffer assigned per guest virtual desktop and thus how many virtual desktops can share that physical GPU. For example, a GRID K2 physical GPU can support up to four K240Q vGPUs on each of its two physical GPUs (for a total of eight vGPUs), but only two K260Qs vGPUs (for a total of four 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. As a result a K2 card, with its two physical GPUs can support up to two different profiles, each K1 can support up to four. See *Homogeneous Virtual GPUs* for more detail.

The K2 and K1 are full height, full length, double width PCI-e cards, and are passively cooled (no onboard fan) requiring certified servers designed to properly power and cool them. Refer to http://www.nvidia.com/buygrid for a list of recommended server platforms to use with GRID K1 and K2 and cross-reference the NVIDIA list against the VMware vSphere HCL for servers and find servers on both lists. This list continues to grow rapidly, check back often for more certified platforms. Each card requires auxiliary power cables connected to it, see chart below for specific pin counts, and most servers require an enablement kit for proper mounting the of the GRID cards. Check with your server OEM of choice for more specific requirements.

2.4 GRID K1 /K2 SPECIFICATIONS

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

Table 2-2.   GRID K1 and GRID K2 Hardware Specifications

	GRID K1	GRID K2
Number of GPUs	4 x K600 equivalent Kepler GPUs	2 x K5000 equivalent Kepler GPUs
Total NVIDIA CUDA Cores	768	3072
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.

2.5 SUPPORTED GRAPHICS PROTOCOLS

This version of GRID vGPU includes support for:

▶ Full DirectX 9/10/11, Direct2D, and DirectX Video Acceleration (DXVA)

▶ OpenGL 4.4

▶ NVIDIA GRID SDK (remote graphics acceleration)

The following are NOT supported in this release of vGPU:

▶ CUDA

▶ OpenCL

NOTE: For CUDA and OpenCL we recommend using vDGA (not covered in this guide).

2.6 HOMOGENOUS VIRTUAL GPUS

GRID vGPU supports homogeneous virtual GPUs. At any given time, the virtual GPUs residing on a single physical GPU must be all of the same type. However, this restriction does not extend across physical GPUs on the same card. Each physical GPU on a K1 or K2 may host different types of virtual GPUs at the same time. For example, a GRID K2 card has two physical GPUs and can support four types of virtual GPU: GRID K220Q, GRID K240Q, GRID K260Q, and GRID K280Q.

Graphic showing possible vGPU profile distribution:

GRID K2 Card							

A K2 card has two physical GPUs…

Physical GPU 0 w/ 4GB FB	Physical GPU 1 w/ 4GB FB

Valid configuration with K240Q profiles on one physical GPU, and K260Q on the other:

K240Q	K240Q	K240Q	K240Q	K260Q	K260Q

Valid configuration with K240Q profiles on both physical GPUs:

K240Q	K240Q	K240Q	K240Q	K240Q	K240Q	K240Q	K240Q

Invalid configuration with K240Q profiles and K260Q on one physical GPU:

K240Q	K240Q	K260Q

Figure 2-3.   Example of a vGPU Configurations on GRID K2

2.7 BEFORE YOU BEGIN

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

2.7.1   General Prerequisites

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

▶ Evaluation plan consisting of all of 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's XenDesktop documentation

2.7.2   Build An Evaluation Plan

To be successful, we highly recommend you start with evaluation plan consisting of all 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

2.7.3   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.

2.7.4   Choosing Your Hardware

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

▶ VMware and NVIDIA certified servers with NVIDIA GRID K1 or K2 cards (see http://www.nvidia.com/buygrid 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 (Ivy Bridge, Haswell)
- High-speed RAM
- Fast networking
- If using local storage IOPS plays a major role in performance. If using VMware for Virtual SAN, see the VMware Virtual SAN requirements website for more details.

▶ High performance storage with Virtual SAN or high IOPS storage system

▶ Higher-performance end points for testing access

2.7.5 Choose Other Virtualization Components

Select the appropriate NVIDIA GRID card for your use case, either K2 for higher performance solutions or K1 for density (refer to http://www.nvidia.com/object/grid-boards.html for details). For additional guidance, please contact your NVIDIA and Citrix reps.

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

▶ Citrix XenDesktop 7.6

  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 vGPU software:
  - NVIDIA vGPU manager VIB
  - NVIDIA WDDM guest driver (32-bit and 64-bit versions)

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

▶ Microsoft software:
  - Windows Server 2012 R2 (recommended) or Server 2008 R2 (minimum)
  - Windows 7 SP1 Professional or Enterprise VL (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 WinSPC (SPC), available from www.putty.org and www.winscp.net, respectively. See Chapter 15, *Using WINSCP*, on page 206 for information on using WinSCP.

▶ Licenses:

- VMware vSphere 6
- Citrix XenDesktop 7.6 Enterprise (or free evaluation)
- Microsoft (volume licenses recommended)

▶ Testing and Benchmarking:
- NVIDIA System Management interface (NSMI)
- OPTIONAL: VMware vRealize Operations Manager
- OPTIONAL: Lakeside Systrack 7 with GPU monitoring

▶ Server where you will install Citrix XenDesktop 7.6 Desktop Delivery Controller:
- Windows Server 2012 R2
- Server must be joined to a domain
- Server must have a static IP address assigned

2.7.6   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 2.8.8, *Server BIOS* Settings on page 15.

   4. Install Virtual Clone Drive or similar to easily mount and un-mount ISO images within the virtual machines.

2.7.7   Known Limitations

See Chapter 13, *Known Issues*, on page 190 for a list of known limitations that is current as of the publication date of this manual.

2.7.8   Server BIOS Settings

Configure the BIOS as appropriate for your physical hosts, as described below:

2.7.8.1   Cisco

▶ CPU Performance: Set this parameter to either Enterprise or High Throughput.

▶ Energy Performance: Set this parameter to **Performance**.

▶ PCI Configuration: Set the following parameters:
- MMCFG BASE: Change this value from **Auto** to **2 GB**.

> **!** CAUTION: FAILURE TO CHANGE THIS VALUE MAY CAUSE THE VSPHERE INSTALLAITON TO FAIL WHILE BOOTING THE ISO.

- Memory Mapped I/O above 4-GB: Set this parameter to **Disabled**.

▶ VGA Priority: Set this parameter to **Onboard VGA Primary**.

▶ Refer to the *Cisco Unified Computing System BIOS Settings* page at the Cisco website for additional Cisco-specific BIOS settings.

2.7.8.2   Dell

▶ System Profile: Set this parameter to Performance.

▶ Memory Mapped I/O Above 4 GB: Set this parameter to Disabled

> 💬 If NVIDIA card detection does not include all of the installed GPUs, set this option to Enabled

▶ Embedded Video Controller: Set this parameter to Enabled.

▶ Refer to the Dell BIOS Settings page at the Dell website for additional Dell-specific BIOS settings.

2.7.8.3   HP

▶ HP Power Profile: Set this parameter to Advanced.

▶ HP Power Regulator: Set this parameter to HP Static High Performance Mode.

▶ Energy/Performance Bias: Set this parameter to Maximum Performance.

▶ Memory Power Savings Mode: Set this parameter to Maximum Performance.

▶ Video Options: Set this parameter to Embedded Video Primary, Optional Video Secondary.

2.7.8.4   Supermicro

▶ Power Technology: Set this parameter to either Maximum Performance or Energy Efficient.

▶ Energy/Performance Bias: Set this parameter to either Performance or Balanced Performance.

▶ Above 4G Decoding: Set this parameter to Disabled.

▶ MMCFG base: Set this parameter to 0xC0000000.

2.7.8.5   Others

Memory Mapped I/O above 4-GB: When available, set this parameter to Disabled, unless vSphere installation fails or returns an error; in this case, enable the 4-GB memory mapping until the base vSphere installation is complete, and then disable this setting again before proceeding to vGPU VIB install.

Chapter 3.
INSTALLING VSPHERE 6

This chapter covers the following vSphere 6 installation topics:

▶ Choosing an install method

▶ Preparing the USB boot media

▶ Installing vSphere 6 from the USB media

▶ Initial host configuration

▶ Assigning a host license

NOTE: This deployment guide assumes you are building an environment as a proof of concept and is not meant to be a production deployment, as a result choices made are meant to speed up and ease the process. See the corresponding guides for each technology, and make choices appropriate for your needs, before building your production environment.

3.1 CHOOSING INSTALL METHOD

With the ability to install from and onto a SD card or USB memory stick, vSphere offers flexibility verses local hard drive install. Please see vSphere documentation regarding best practices for logs when booting from USB or similar. In our main lab we used Supermicro's IPMI and virtual media to boot from ISO file and install on local storage. In home labs USB was used to quickly move from one version to another.

3.2 PREPARING USB BOOT MEDIA

Reference VMware KB: "Installing ESXi 5.x on a supported USB flash drive or SD flash card (2004784)"

Booting vSphere 6 from a USB drive is useful if your host has an existing vSphere Version 5.5 or earlier installation that you want to retain.

Use the following procedure to prepare a USB drive for booting:

1. Download **UNetbootin** from http://unetbootin.sourceforge.net/.
    The Windows version of the application does not include an installer; however, the OSX version is packaged in a **.DMG** file that you must mount. You must also copy the application to the **Applications** folder before launching. Alternatively, you can use YUMI, which allows booting multiple installation images on one USB device plus the option to load the entire installation into RAM. The download link is http://www.pendrivelinux.com/yumi-multiboot-usb-creator/.

2. Start the application, select **Diskimage**, and then click the … icon to browse for the installation **.ISO** file.
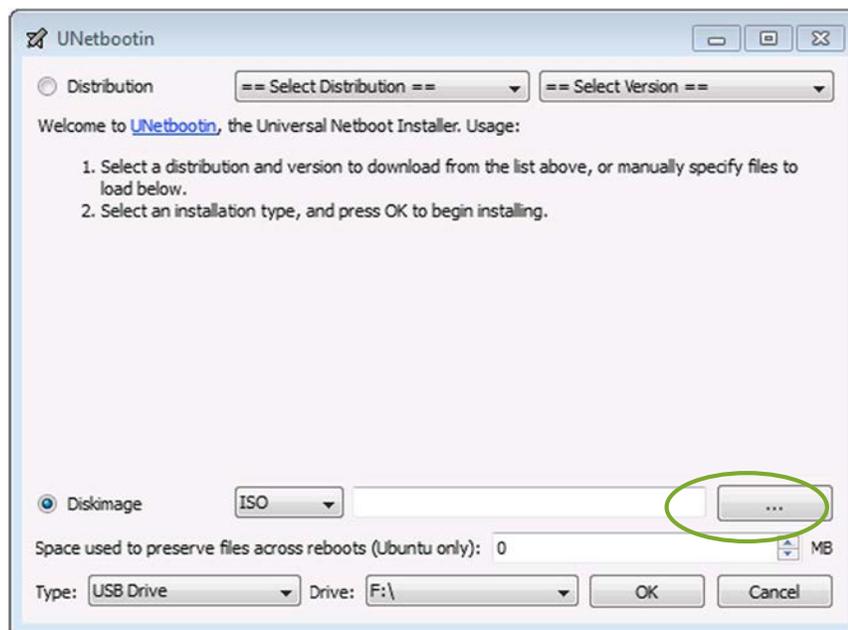
3. Navigate to the location that contains the installation **.ISO** file and then select **Open**.

4. Select the mounted USB drive on which to perform the installation and then select **OK**. The copying process begins and a series of progress bars are displayed.
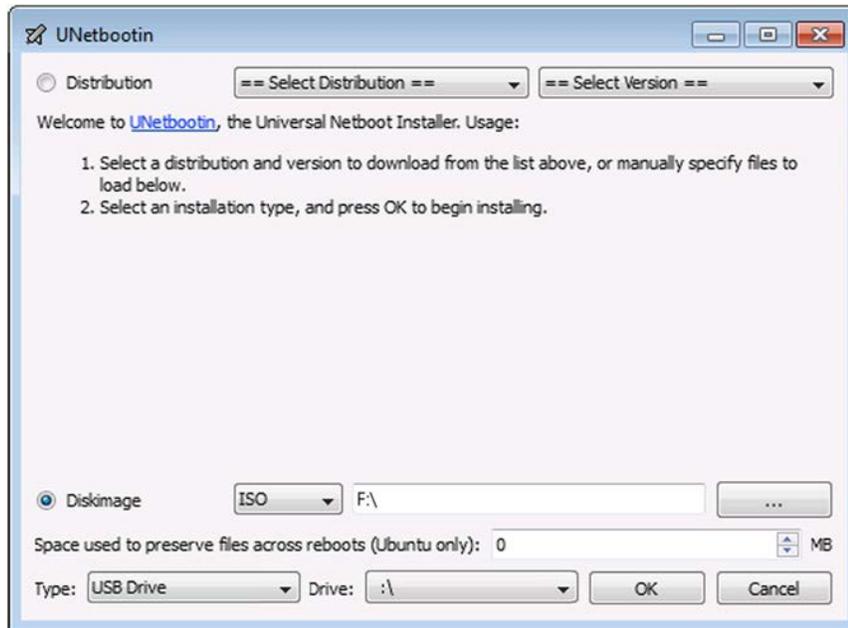
5. When the copying process is complete, click Exit and then remove the USB drive.

6. To install from this USB drive, insert into the host using either an internal or on motherboard USB port, then set that as the primary boot source or select from the boot menu on power up.

3.3 INSTALLING VSPHERE 6

Use the following procedure to install the vSphere 6 regardless of boot source. Select the boot media with the vSphere ISO on your host's boot menu.

1. Apply power to start the host.
   The following menu displays when the host starts up:
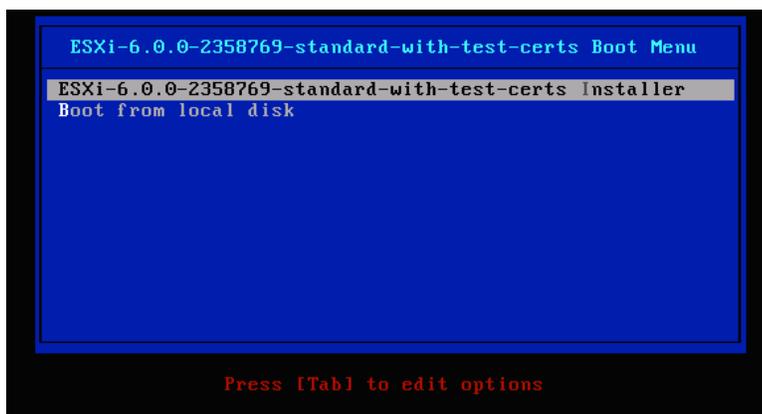
2. Select the installer using the arrow keys and then press **[ENTER]** to begin booting the vSphere 6 installer.

3. A compatibility warning displays:

```
        Welcome to the VMware ESXi 6.0.0 Installation

VMware ESXi 6.0.0 installs on most systems but only
systems on VMware's Compatibility Guide are supported.

Consult the VMware Compatibility Guide at:
http://www.vmware.com/resources/compatibility

Select the operation to perform.

        (Esc) Cancel        (Enter) Continue
```
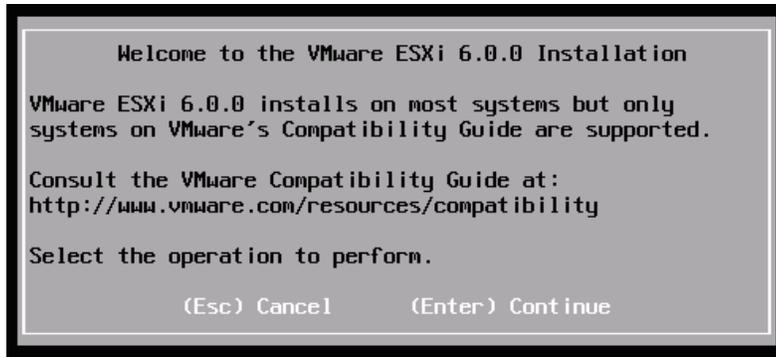
4. Press **[ENTER]** to proceed.
   The End User License Agreement (EULA) displays:

```
            End User License Agreement (EULA)

VMWARE, INC.
BETA LICENSE AGREEMENT

Note: BY CLICKING "I AGREE," ACCESSING, DOWNLOADING,
INSTALLING, UPLOADING, COPYING OR USING THE BETA SOFTWARE,
YOU CONCLUDE AND AGREE TO THIS BETA LICENSE AGREEMENT
("AGREEMENT") IN A LEGALLY BINDING MANNER WITH VMWARE, INC.,
3401 HILLVIEW AVENUE, PALO ALTO, CA 94304, USA ("VMware").
IF YOU HAVE SPECIFIED IN CONNECTION WITH THE REGISTRATION
PROCESS FOR THIS SOFTWARE BETA TEST PROGRAM THAT YOU ARE
ACTING ON BEHALF OF A COMPANY OR OTHER ORGANIZATION, YOU
REPRESENT THAT YOU ARE AUTHORIZED TO LEGALLY BIND SUCH
ORGANIZATION AND THAT YOU CONCLUDE THE AGREEMENT ON BEHALF
OF SUCH ORGANIZATION. IN THE FOLLOWING, THE TERMS "YOU" AND
"LICENSEE" SHALL REFER, JOINTLY AND SEVERALLY, TO YOU

        Use the arrow keys to scroll the EULA text

    (ESC) Do not Accept        (F11) Accept and Continue
```

5. 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:

```
                    Scanning...
Scanning for available devices. This may take a few seconds.
```
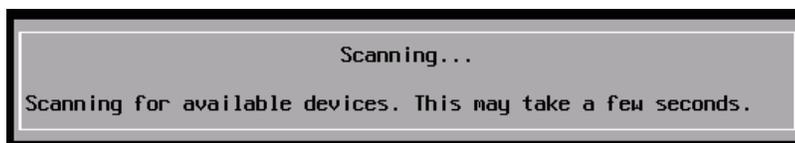
6. It should display all drives available for install:

```
                 Select a Disk to Install or Upgrade

  * Contains a VMFS partition
  # Claimed by VMware Virtual SAN (VSAN)

  Storage Device                                           Capacity
  ----------------------------------------------------------------------
  Local:
   * Intel    RCS25ZB040LX    (naa.600605b00631dc201b1f6...) 222.59 GiB
   # Intel    RCS25ZB040LX    (naa.600605b00631dc201b20b...) 930.39 GiB
   * ATA      INTEL SSDSC2BP24 (t10.ATA_____INTEL_SSDSC2B...) 223.57 GiB
   # ATA      INTEL SSDSC2BP24 (t10.ATA_____INTEL_SSDSC2B...) 223.57 GiB
  Remote:
     (none)


    (Esc) Cancel    (F1) Details    (F5) Refresh    (Enter) Continue
```

7. Use the arrow keys to select the drive you want to 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.

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

```
                 Select a Disk to Install or Upgrade

  * Contains a VMFS partition
  # Claimed by VMware Virtual SAN (VSAN)

  St                                                           ity
  --                                                           ---
  Lo               Scanning...
   *    Gathering additional information from the selected device.   GiB
   #               This will take a few moments.                     GiB
   *                                                                 GiB
   # ATA      INTEL SSDSC2BP24 (t10.ATA_____INTEL_SSDSC2B...) 223.57 GiB
  Remote:
     (none)


    (Esc) Cancel    (F1) Details    (F5) Refresh    (Enter) Continue
```
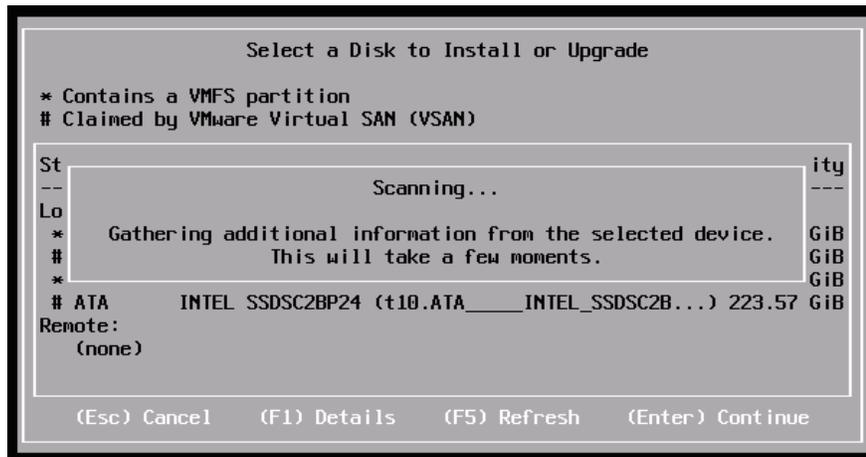
9. The Confirm Disk Selection window displays:

```
                            ESXi and VMFS Found
 * Cont  The selected storage device contains an installation of
 # Clai  ESXi and a VMFS datastore.  Choose whether to upgrade
         or install and overwrite the existing ESXi
 Storag  installation.  Also choose whether to preserve or          apacity
 ------  overwrite the existing VMFS datastore.                     -------
 Local:
 * Int   ( ) Upgrade ESXi, preserve VMFS datastore                  .59 GiB
 # Int   ( ) Install ESXi, preserve VMFS datastore                  .39 GiB
 * ATA   (X) Install ESXi, overwrite VMFS datastore                 .57 GiB
 # ATA                                                              .57 GiB
 Remote     Use the arrow keys and spacebar to select an option.
   (no
                     (Esc) Cancel        (Enter) OK


      (Esc) Cancel     (F1) Details     (F5) Refresh     (Enter) Continue
```

10. Press **[ENTER]** to accept your selection and continue

    The **Please select a keyboard layout** window displays:

```
              Please select a keyboard layout

    Swiss French
    Swiss German
    Turkish
    US Default
    US Dvorak
    Ukrainian
    United Kingdom

              Use the arrow keys to scroll.

    (Esc) Cancel     (F9) Back     (Enter) Continue
```

11. Select your desired keyboard layout using the arrow keys and then press **[ENTER]**.

    The **Enter a root password** window displays.

```
                Enter a root password

      Root password: **********
   Confirm password: **********

                 Passwords match.


       (Esc) Cancel     (F9) Back     (Enter) Continue
```

12. 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.

13. Confirm the password in the **Confirm password** field and then press **[ENTER]** to proceed. The installer rescans the system:

```
                        Scanning system...

Gathering additional system information. This may take a few moments.
```
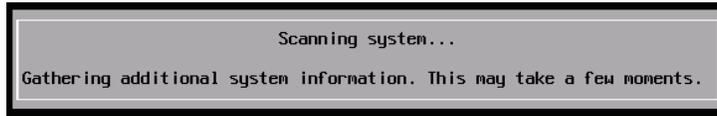
14. Then displays the **Confirm Install** window:

```
                       Confirm Install

     The installer is configured to install ESXi 6.0.0 on:
             naa.600605b00631dc201b1f6ae61a703b57.

        Warning: This disk will be repartitioned.


     (Esc) Cancel        (F9) Back        (F11) Install
```
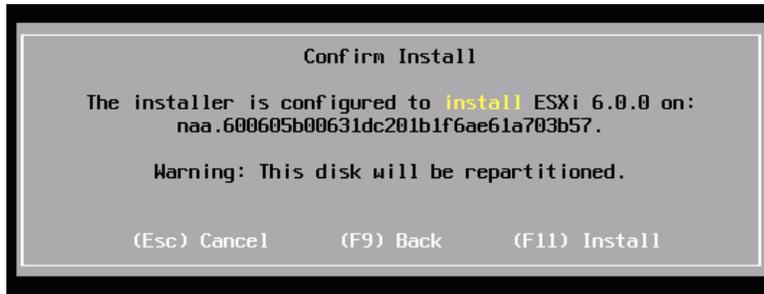
15. Press **[F11]** to proceed with the installation.

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

16. The vSphere 6 installation proceeds:

```
            Installing ESXi 6.0.0
                   9 %
```

17. The **Installation Complete** window displays when the installation process is completed:

```
                  Installation Complete

ESXi 6.0.0 has been successfully installed.

ESXi 6.0.0 will operate in evaluation mode for 60 days. To
use ESXi 6.0.0 after the evaluation period, you must
register for a VMware product license. To administer your
server, use the vSphere Client or the Direct Control User
Interface.

Remove the installation disc before rebooting.

Reboot the server to start using ESXi 6.0.0.

                    (Enter) Reboot
```

18. Press **[ENTER]** to reboot the system. (Make sure your installation media has been ejected and you bios set to the boot disk.)

```
                    Rebooting Server

The server will shut down and reboot.

The process will take a short time to complete.
```
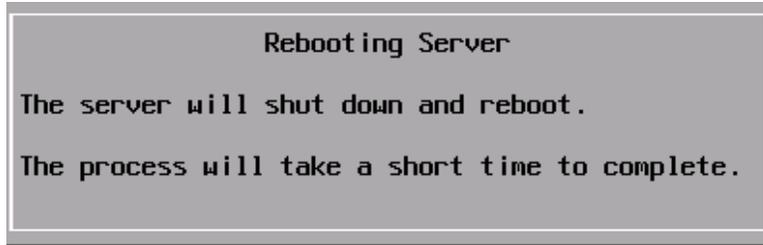
The installation is now complete.

3.4 INITIAL HOST CONFIGURATION

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

The VMware vSphere 6 screen displays once the boot completes:

```
VMware ESXi 6.0.0 (VMKernel Release Build 2358769)

Supermicro SYS-2027GR-TRFH

2 x Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00GHz
128 GiB Memory




Download tools to manage this host from:
http://10.31.235.109/ (DHCP)
http://[fe80::225:90ff:fee3:1296]/ (STATIC)




<F2> Customize System/View Logs                    <F12> Shut Down/Restart
```
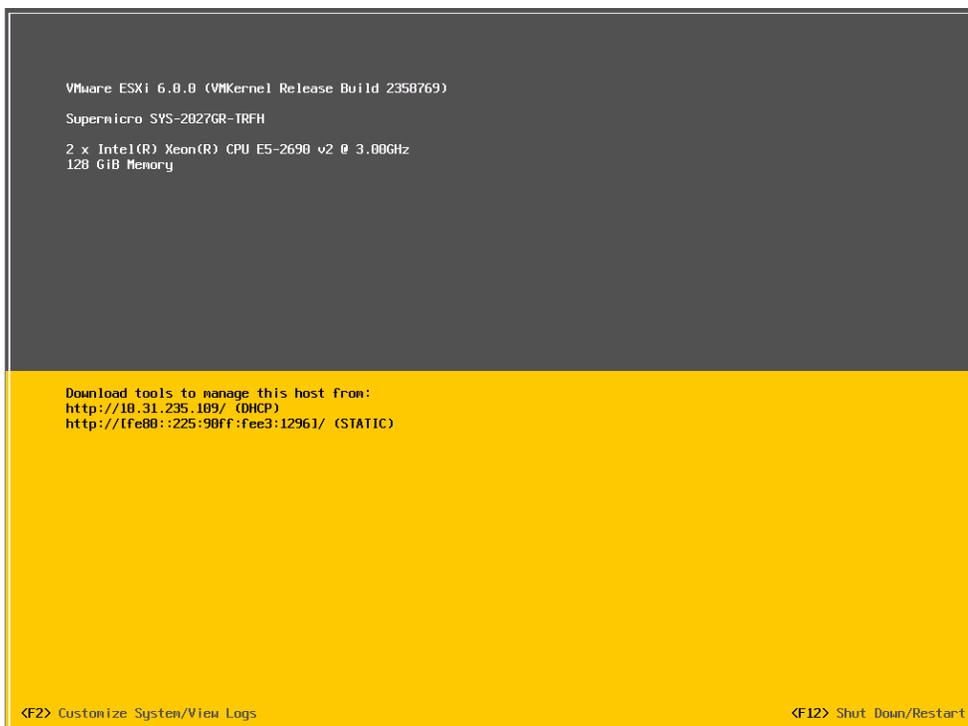
Figure 3-1.   VMware vSphere 6 Screen

Use the following procedure to configure the host:

1. Press [F2].

The **Authentication Required** window displays:

```
Authentication Required

Enter an authorized login name and password for
localhost.nvidia.lab.

Configured Keyboard (US Default)
Login Name:        [ root                         ]
Password:          [ *********_                    ]

                            <Enter> OK   <Esc> Cancel
```
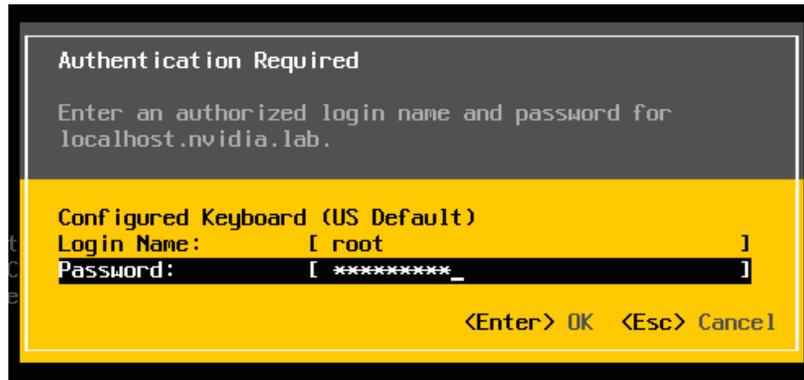
2. Enter the root account credentials that you created during the installation process and then press [ENTER]. The **System Customization** screen displays.

```
System Customization                     Configure Password

Configure Password                       Set
Configure Lockdown Mode
                                         To prevent unauthorized access to this system, set the
Configure Management Network             password for the user.
Restart Management Network
Test Management Network
Network Restore Options

Configure Keyboard
Troubleshooting Options

View System Logs

View Support Information

Reset System Configuration
```

3. Scroll down to select **Configure Management Network** and then press [ENTER].

The **Network Adapters** window appears:

```
System Customization                     Configure Management Network

Configure Password                       Hostname:
Configure Lockdown Mode                  localhost

Configure Management Network             IPv4 Address:
Restart Management Network               10.31.235.109
Test Management Network
Network Restore Options                  Network identity acquired from DHCP server 10.31.235.21

Configure Keyboard                       IPv6 Addresses:
Troubleshooting Options                  fe80::225:90ff:fee3:1296/64

View System Logs                         To view or modify this host's management network settings in
                                         detail, press <Enter>.
View Support Information

Reset System Configuration
```

4. Use the arrow keys to select the adapter to use as the default management network and then press [ENTER]. The **IPv4 Configuration** window displays:

```
appropriate settings.

IPv4 Configuration

This host can obtain network settings automatically if your network
includes a DHCP server. If it does not, the following settings must be
specified:

( ) Disable IPv4 configuration for management network
( ) Use dynamic IPv4 address and network configuration
(o) Set static IPv4 address and network configuration:

IPv4 Address                                         [ 10.31.235.109    ]
Subnet Mask                                          [ 255.255.255.0    ]
Default Gateway                                      [ 10.31.235.1      ]

<Up/Down> Select   <Space> Mark Selected          <Enter> OK   <Esc> Cancel
```

5. 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.
6. Press **[ENTER]** when finished to apply the new management network settings.
   The **Confirm Management Network** popup displays.
7. Press **[Y]** to confirm your selection.
   The **DNS Configuration** window displays.
8. Add the primary and (if available) secondary DNS server address(es) in the respective fields.
9. Set the host name for this vSphere host in the **Hostname** field.
10. Press **[ENTER]** when finished.
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.


```
Troubleshooting Mode Options              ESXi Shell

Enable ESXi Shell                         ESXi Shell is Disabled
Enable SSH
Modify ESXi Shell and SSH timeouts        Change current state of the ESXi Shell
Modify DCUI idle timeout
Restart Management Agents
```

14. In order to install the NVIDIA VIB in a later step, you will need to enable the vSphere Shell. This can be accomplished by selecting **Enable vSphere Shell**.

15. Press **[ENTER]** to toggle **Enable vSphere Shell** on.
    The window on the right displays the status: **Enable vSphere Shell Disabled**

```
Troubleshooting Mode Options          ESXi Shell

Enable ESXi Shell                     ESXi Shell is Disabled
Enable SSH
Modify ESXi Shell and SSH timeouts    Change current state of the ESXi Shell
Modify DCUI idle timeout
Restart Management Agents
```

16. Enable SSH by selecting **Enable SSH** and press **[ENTER]** to toggle this option on.
    The window on the right displays the status: **SSH is Enabled**

Chapter 4.
INSTALLING VMWARE VCENTER 6 SERVER

This chapter covers installing VMware vCenter 6 Server, including:

▶ An installation walk through

▶ The initial vCenter Server Configuration

▶ Adding a host

▶ Setting a vCenter appliance to auto-start

▶ Mounting an NFO ISO data store

Review the prerequisites in section 2.8.5, *General Prerequisites* on page 13 before proceeding with these installations.

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

4.1 INSTALLING VCENTER 6 SERVER APPLIANCE

4.1.1   Initial Installation

Use the following procedure to install vCenter 6 Server Linux Appliance, starting with an Internet Explorer plug-in that allows vCenter Server install from your current desktop.

4.1.1.1   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. Mount the VMware ISO with the installation media in the drive

2. Browse to the VCSA folder:
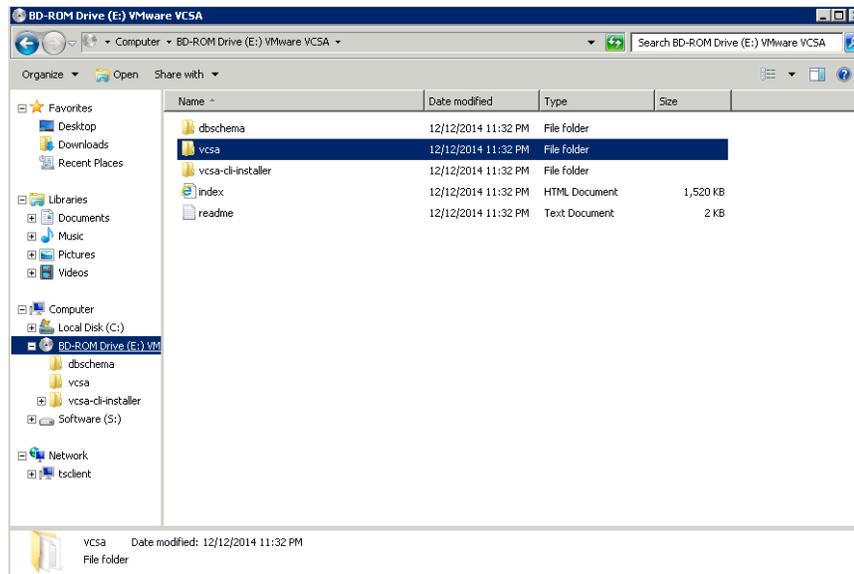
3. In the VCSA folder, double click the VMware-ClientIntegrationPlugin-6.0.0 installer:
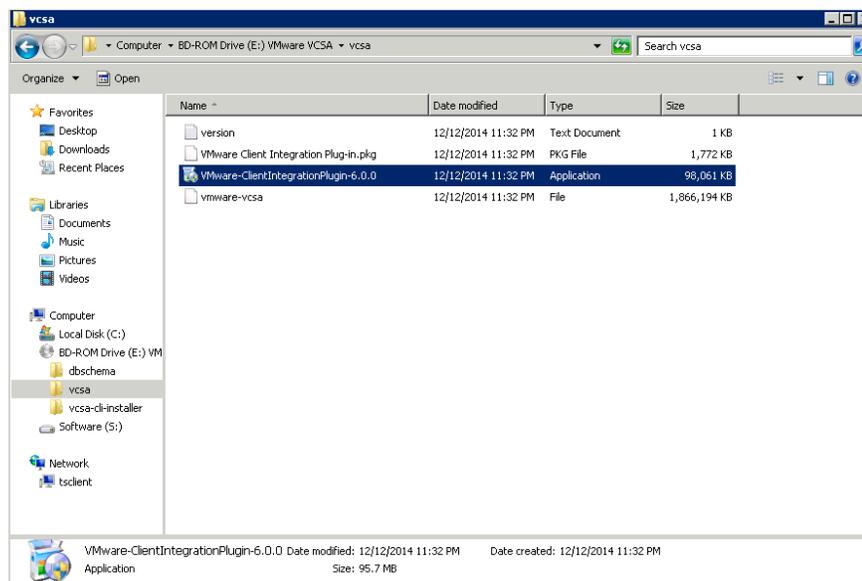
4. If the Microsoft UAC prompts, click [YES] to allow the install to begin:
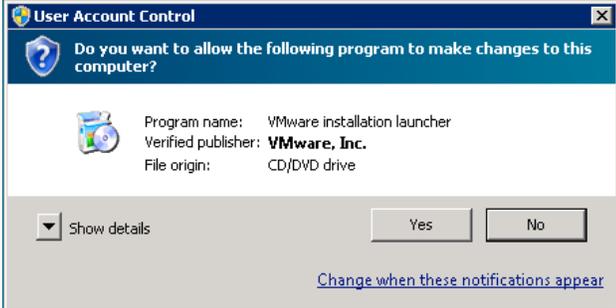
5.Installer starts, click **[Next]** to continue:
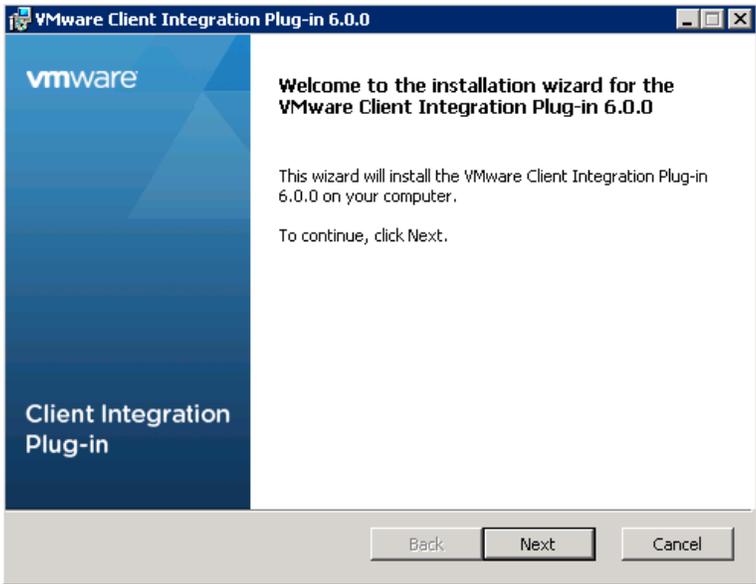
NOTE: All browsers must be closed or you will be prompted to close them, if this window pops up then close all listed browsers and click **[Retry]** to continue:
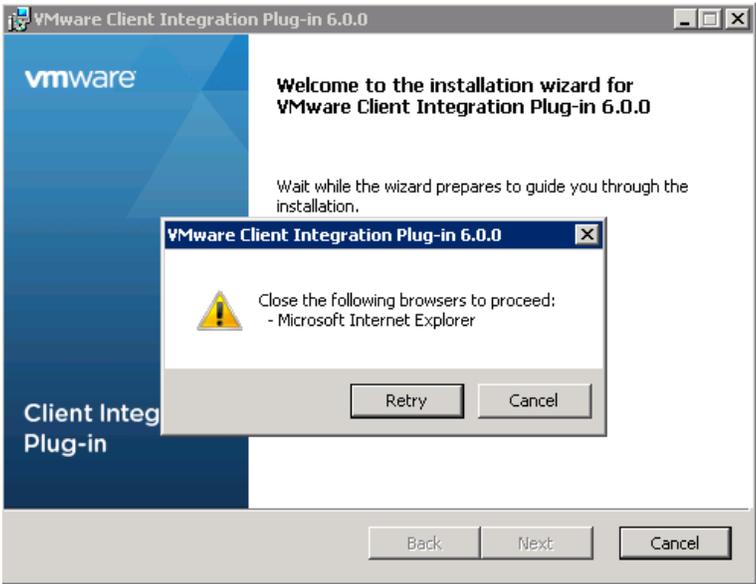
6. Accept the EULA and click [Next] to continue:
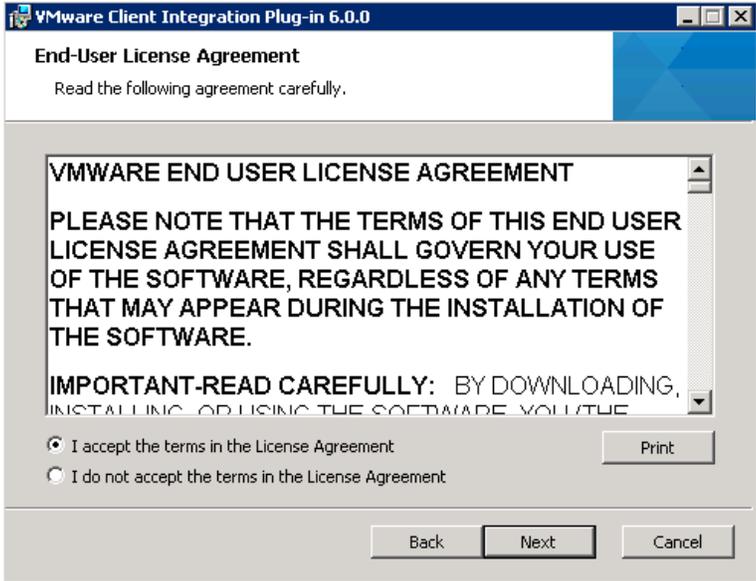
7. Select the install destination folder and click [Next]:
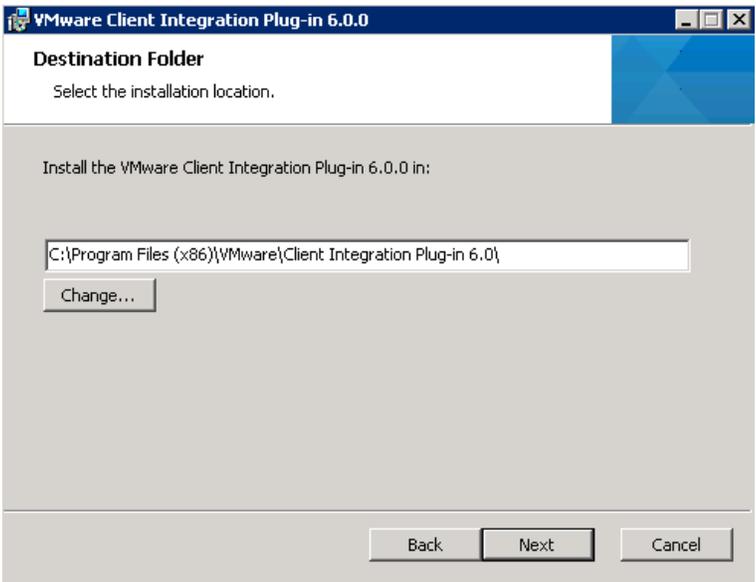
8. Click [Install] to begin the installation:
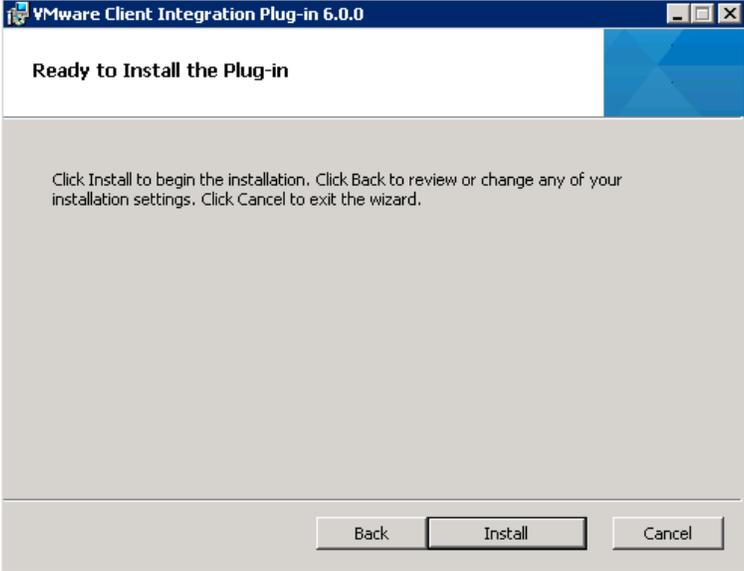
9. Install in progress…
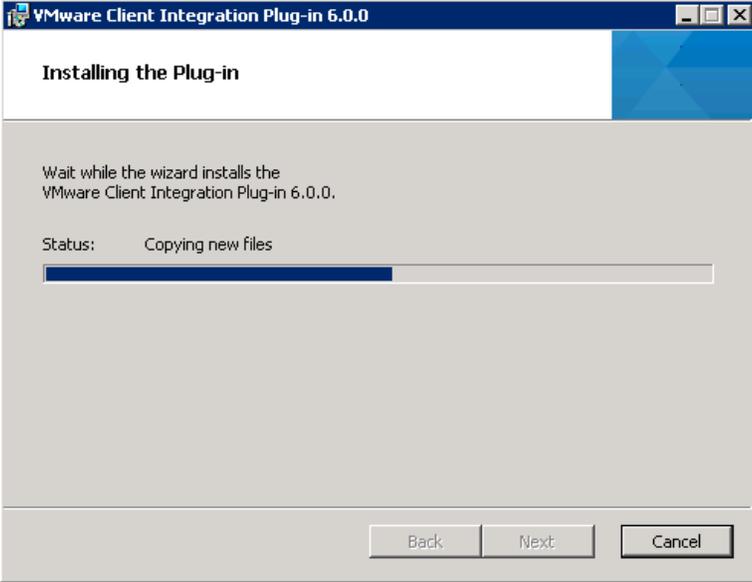
10. When install completes, click **[Finish]**:
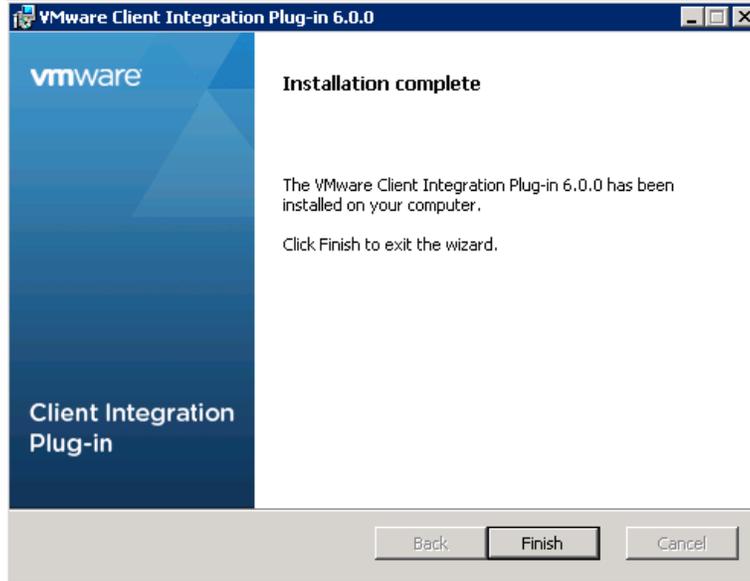
4.1.1.2   vCenter Server Appliance (VCSA) installation

Using a web installer and the client integration plug-in from above, you now install the VCSA.

11. Mount the VMware VCSA ISO with the installation media in the drive.

12. Browse to the VCSA folder:
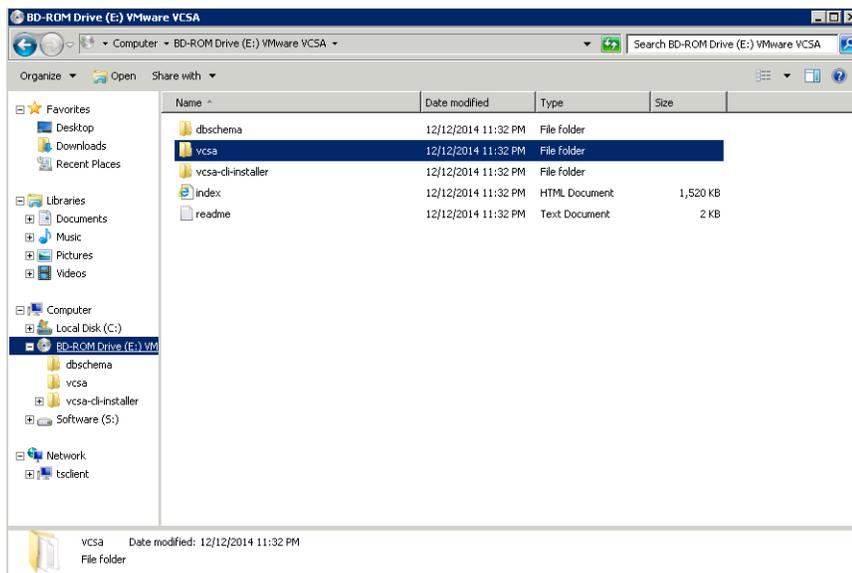
13. Open the html file "index" in Internet Explorer:
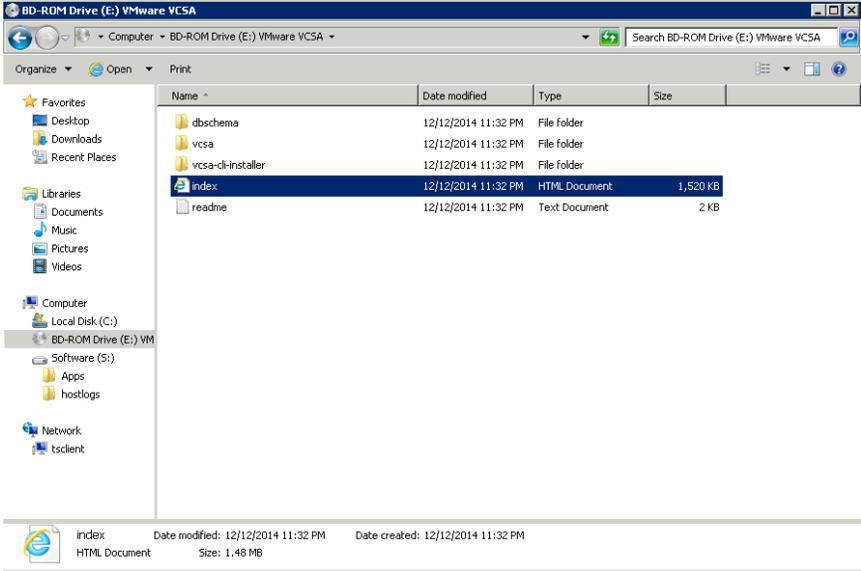
14. Below is the alert you will receive if the VMware Client Integration Plug-in is not installed.
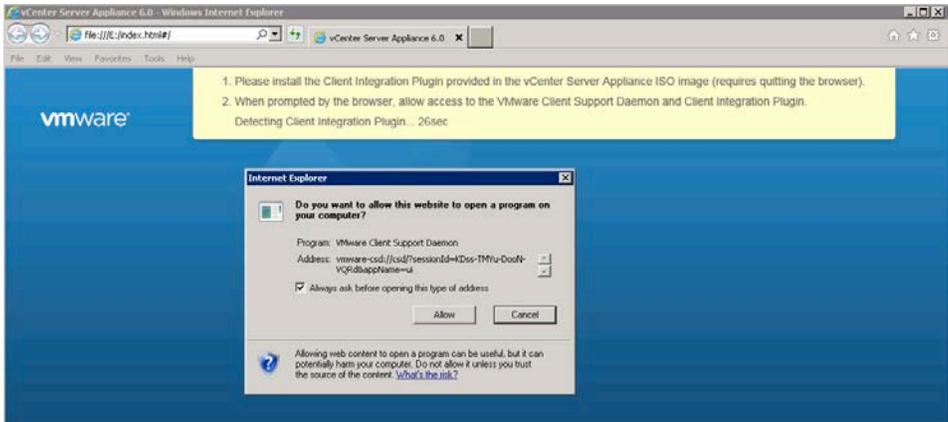
15. If the plug-in is active, you are prompted to allow it access to your operating system, click **[Allow]**:
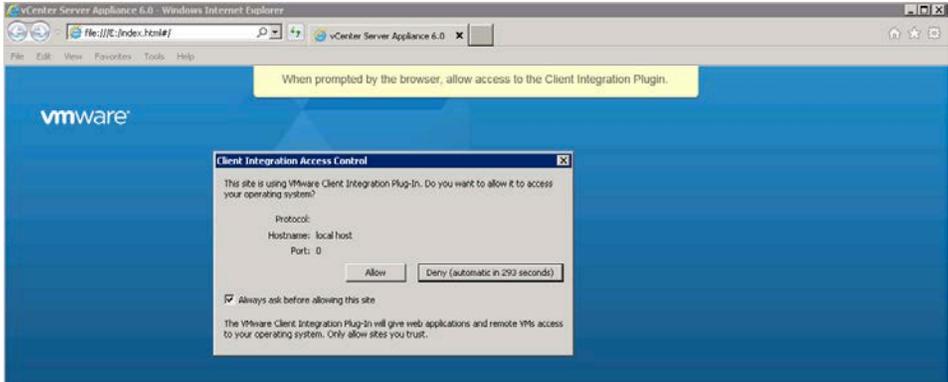
16. To begin the install of the VCSA on a host click **[Install]**:
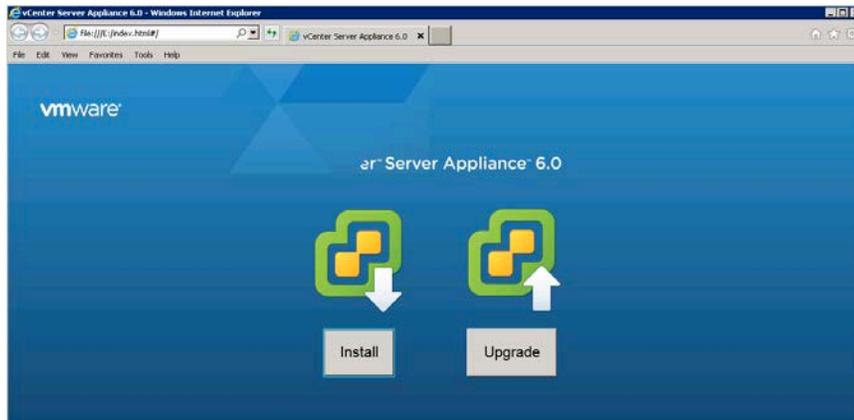
17. Read and then accept the EULA, click **[Next]** to continue:
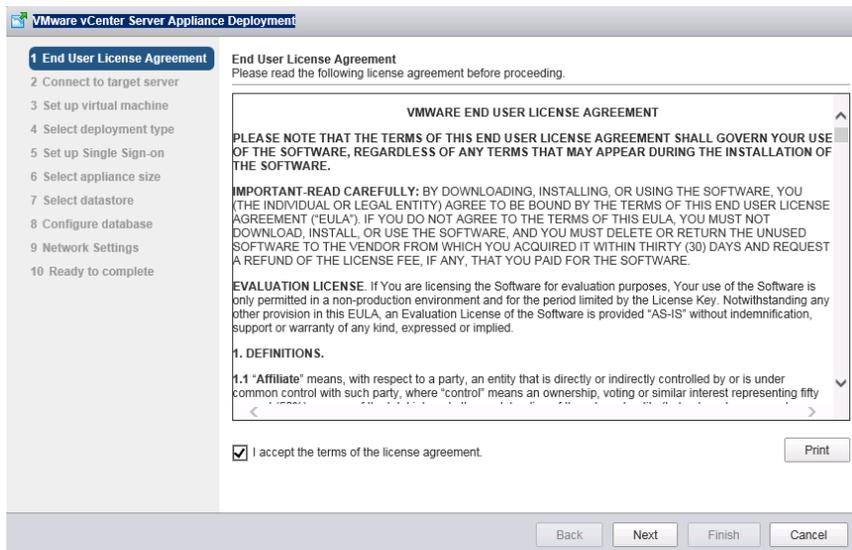
18. In this step you are selecting the vSphere host to install the VCSA on as a guest, this can be on a host running ESXi 5.5 or later. It is recommended that the vCenter server (Windows or appliance based) run on a separate management cluster from the one designated for VDI workloads. Enter the IP address or FQDN of the chosen host, then its root user name and password and click **[Next]**:
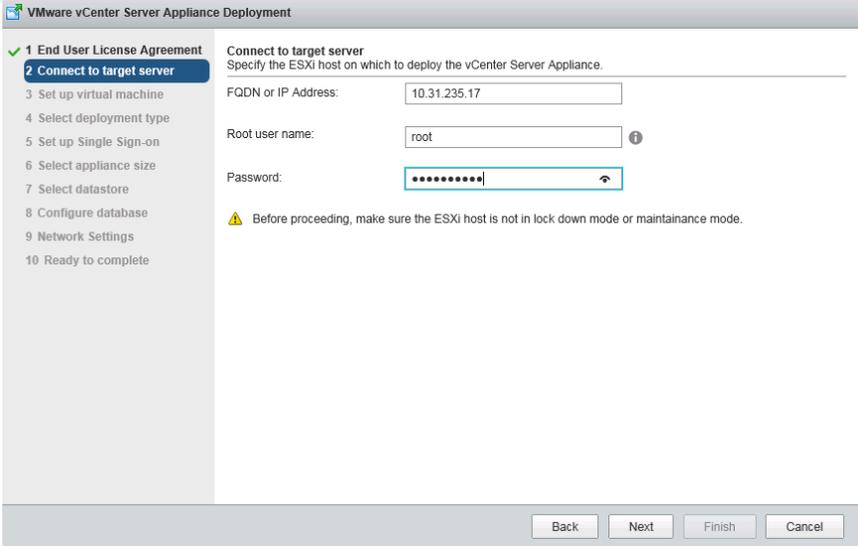
19. If your desktop, and the web installer, can reach the host then you should see a certificate warning as it connects, this is due to the use of a self-signed cert. If you are using signed certificates then you will not see this warning. Click [Yes] and continue:

20. The credentials you provided are validated:
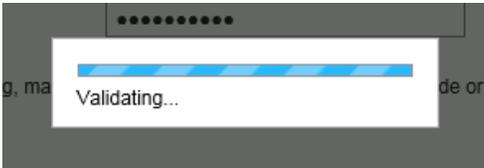
21. With a successful connection you are now prompted to name the appliance, then enter a root password for the appliance, enter it twice and click next:

VMware vCenter Server Appliance Deployment

✔ 1 End User License Agreement
✔ 2 Connect to target server
3 Set up virtual machine
4 Select deployment type
5 Set up Single Sign-on
6 Select appliance size
7 Select datastore
8 Configure database
9 Network Settings
10 Ready to complete

Set up virtual machine
Specify virtual machine settings for the vCenter Server Appliance to be deployed.

Appliance name: vCenter Server Appliance 6.0

OS user name: root

OS password: ••••••••• ⓘ

Confirm OS password: ••••••••• ⚏

Back   Next   Finish   Cancel

22. Select a deployment type for your appliance. Since you are creating an instance for building your proof of concept, this should be a separate install of vCenter Server with an Embedded Platform Services Controller, choose that option and click [Next]:

VMware vCenter Server Appliance Deployment

✔ 1 End User License Agreement
✔ 2 Connect to target server
✔ 3 Set up virtual machine
4 Select deployment type
5 Set up Single Sign-on
6 Select appliance size
7 Select datastore
8 Configure database
9 Network Settings
10 Ready to complete

Select deployment type
Select the services to deploy onto this appliance.

vCenter Server 6.0 requires a Platform Services Controller, which contains shared services such as Single Sign-On, Licensing, and Certificate Management. An embedded Platform Services Controller is deployed on the same Appliance VM as vCenter Server. An external Paltform Services Controller is deployed in a seperate Appliance VM. For smaller installations, consider vCenter Server with an embedded Platform Services Controller. For larger installations with multiple vCenter Servers, consider one or more external Platform Services Controllers. Refer to product documentation for more information.

Note: Once you install vCenter Server, you can only change from an embedded to an external Platform Services Controller with a fresh install.

Embedded Platform Services Controller

◉ Install vCenter Server with an Embedded Platform Services Controller

VM or Host
Platform Services Controller
vCenter Server

External Platform Services Controller

○ Install Platform Services Controller
○ Install vCenter Server

VM or Host
Platform Services Controller

VM or Host   VM or Host
vCenter Server   vCenter Server

Back   Next   Finish   Cancel

23. Create a Single Sign-On instance, 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]:

VMware vCenter Server Appliance Deployment

✓ 1 End User License Agreement
✓ 2 Connect to target server
✓ 3 Set up virtual machine
✓ 4 Select deployment type
5 **Set up Single Sign-on**
6 Select appliance size
7 Select datastore
8 Configure database
9 Network Settings
10 Ready to complete

Set up Single Sign-on
Create or join a Single Sign-On domain. Single Sign-On configuration cannot be changed after deployment.

● Create Single Sign-On
○ Join a Single Sign-On domain in an existing vCenter 6.0 platform services controller

vCenter Single Sign-On User name: administrator

vCenter Single Sign-On Password: [•••••••••] ⓘ

Confirm password: [•••••••••]

Domain name: [vsphere.local] ⓘ

Site name: [PSGSC] ⓘ

Back | Next | Finish | Cancel

24. Select an appliance size, appropriate for the potential scale of your proof of concept or trial, here we selected small, then click **[Next]**:

VMware vCenter Server Appliance Deployment

✓ 1 End User License Agreement
✓ 2 Connect to target server
✓ 3 Set up virtual machine
✓ 4 Select deployment type
✓ 5 Set up Single Sign-on
6 **Select appliance size**
7 Select datastore
8 Configure database
9 Network Settings
10 Ready to complete

Select appliance size
Specify a deployment size for the new appliance

Appliance size: [Small (up to 150 hosts, 3,000 VMs) ▼]

Description

This will deploy a Small VM configured with 4 vCPUs and 16 GB of memory. Please note this option contains vCenter Server with an embedded Platform Services Controller.

Back | Next | Finish | Cancel

25. Select a datastore from those available on the host you have connected with, also consider selecting Enable Thin Disk Mode, and then click **[Next]**:

26. For a proof of concept environment, choose the embedded database, then click [Next]:

27. Configure network settings, this is a long page and will require scrolling down to see all settings. Before configuring these settings, choose an appropriate 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:
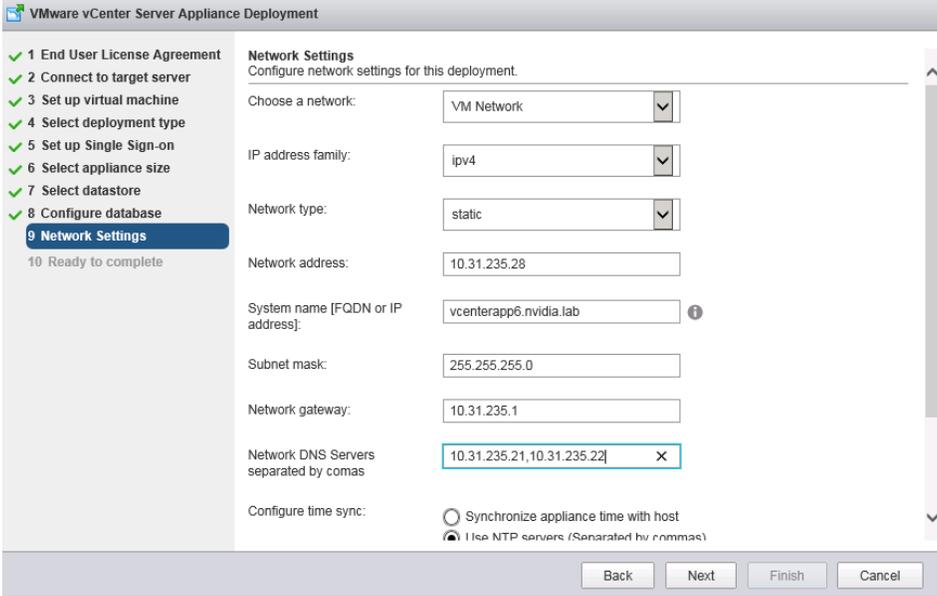
28. Make sure your entire configuration leverages an NTP service, this includes the vCenter Server, ESXi hosts, VMs and Management components. The lower half of the Network Settings pane adds NTP, enter your local time server or a reachable external one, and click **[Next]**:
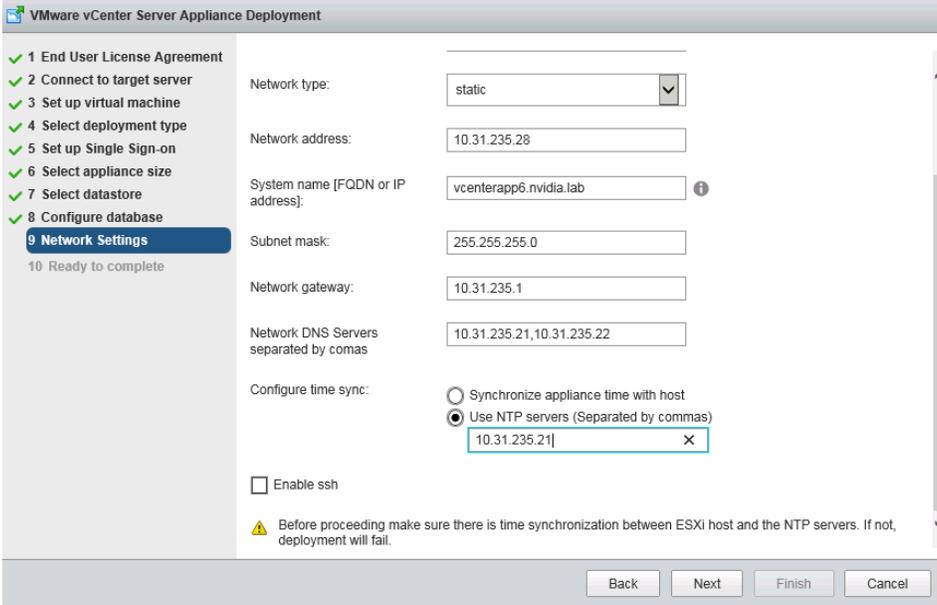
29. Confirm your settings and if correct click **[Finish]** to deploy your vCenter server:
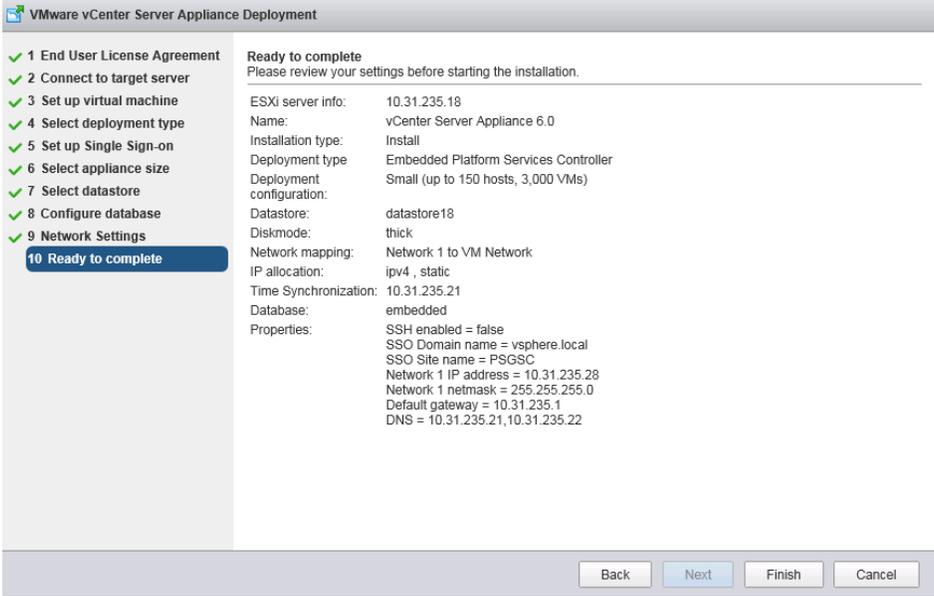
30. If all settings are correct then you should see a progress bar:
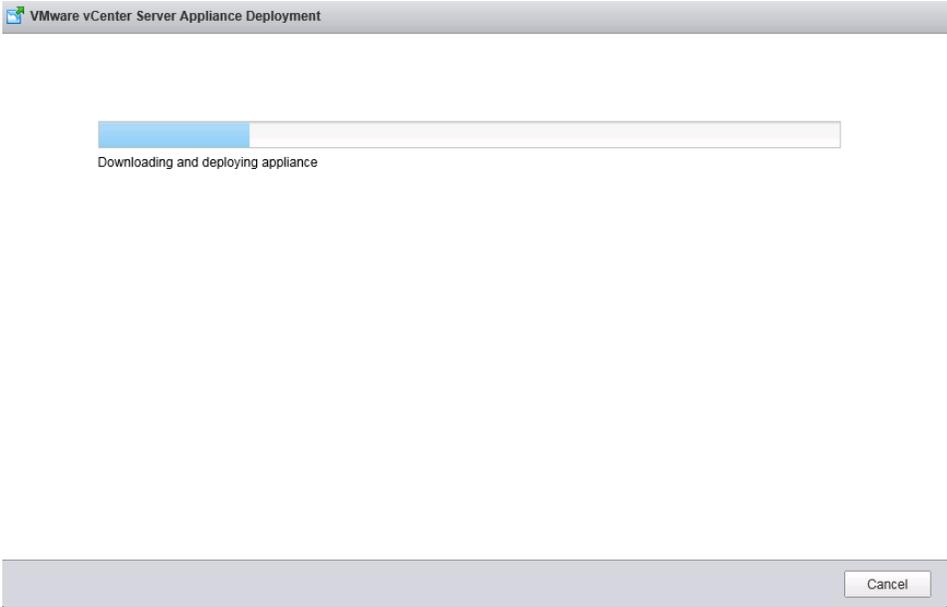
31. When complete the window below appears with a link to your VCSA, click the link to reach the login page. You can now close this page:

VMware vCenter Server Appliance Deployment

Installation Complete

Your vCenter Server is successfully installed.

Post install steps:

1. vCenter Server is installed in evaluation mode. Activate vCenter Server by using the vSphere Web Client within 60 days. When the evaluation period of this vCenter Server expires, all hosts will be disconnected from this vCenter Server.
2. Use the vSphere Web Client to manage vCenter Server. Log in with the Single Sign-On administrator account administrator@vsphere.local

You can now login to vSphere Web Client: https://vcenterapp6.nvidia.lab/vsphere-client as administrator@vsphere.local

Close

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

vmware

User name: administrator@vsphere.local

Password: •••••••••

☐ Use Windows session authentication

Login

VMware® vCenter™ Single Sign-On

❗ Cookies are required.

Download Client Integration Plugin
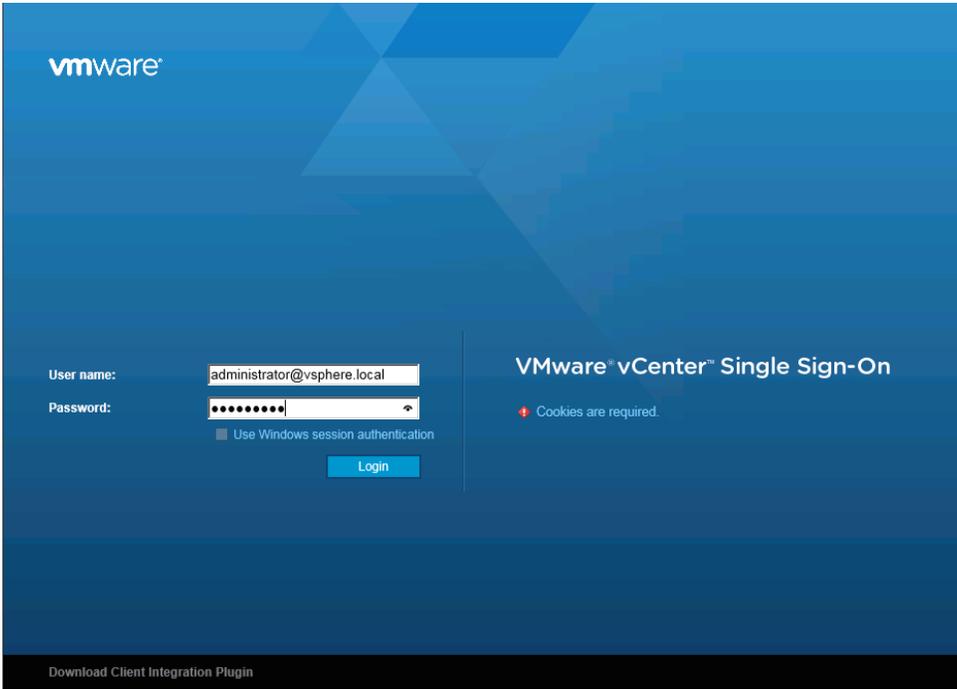
4.2 INSTALLING VCENTER 6 SERVER ON WINDOWS SERVER 2012

This section describes how to install and configure vCenter 6 Server.

4.2.1 Initial Installation

Use the following procedure to install vCenter 6 Server:

1. Mount the VMware ISO with the installation media in the drive

2. Right-click the drive and select **Install or run program** from your media.
    The *VMware vCenter Installer* window displays.

3. Select **vCenter Server for Windows** and click **Install**.
    The Welcome window displays.

4. Click **Next** to proceed with the installation.
    The *End User License Agreement* window displays.

5. Check **I accept the terms of the license agreement** checkbox,

6. Click Next to proceed.
    The *Deployment Type* window displays.

7. Select **vCenter Server with an embedded infrastructure controller**.

8. Click **Next** to proceed.
    The *System Network Name* window displays.

9. Enter the name or IP address of the host in the **System Name** field,

10. Click **Next** to proceed.
    The *Single Sign-On Configuration* window displays.

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

11. Enter the administrator password for the new Single Sign- On domain in the **Password** and **Confirm password** fields

12. Click Next to proceed.
    The *vCenter Server Service Account* window displays.

    > ❗ **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

13. Select **Use Windows Local Account**.

14. Click **Next** to proceed.
    The *Database Settings* window displays.

15. Select **Use an embedded database (vPostgres)**,

16. Click **Next** to proceed.
    The *Configure Ports* window displays. The values shown are the default values; you should not have to modify any of them.

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

18. Accept the suggested path or enter a custom path.

19. Click Next to proceed.
    The *Ready to Install* window displays.

20. Verify that all of the selections are correct and click **Next** to proceed.
    The installation takes place. The *Setup Completed* window displays once the installation is completed.
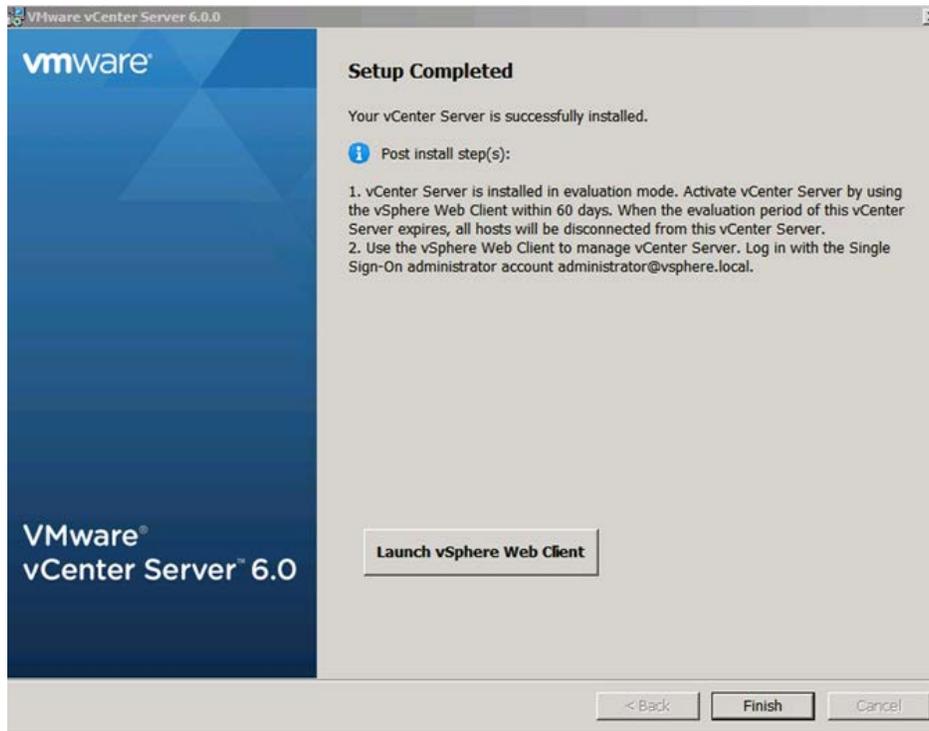
21. Click **Finish** to complete the installation process

22. Connect to the host using the Web client, using your Single Sign-On credentials.

4.2.2   Adding Licenses to your vCenter Server

Use the following procedure to configure vCenter 6:

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. The *Getting Started* page displays, select **Log in to vSphere Web client**:
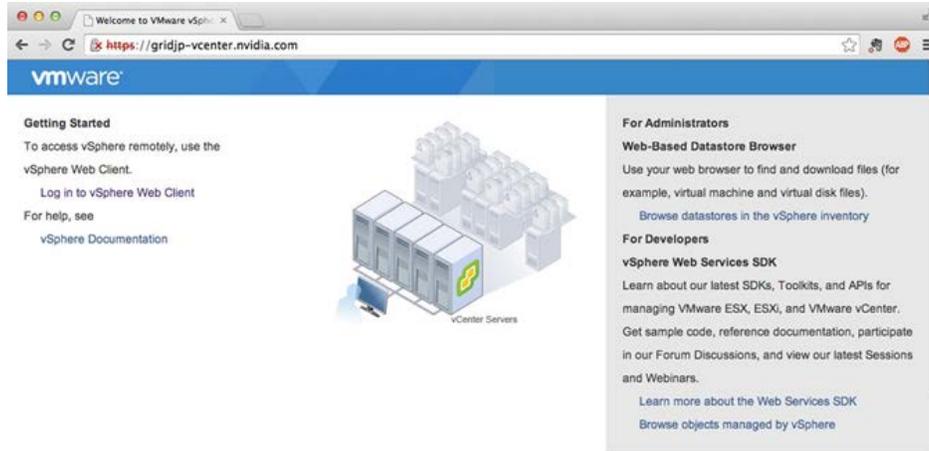
2. Install the VMware Client Support Plug-in by selecting **Always run on this site**, if prompted. The *VMware Single Single-On* page displays:
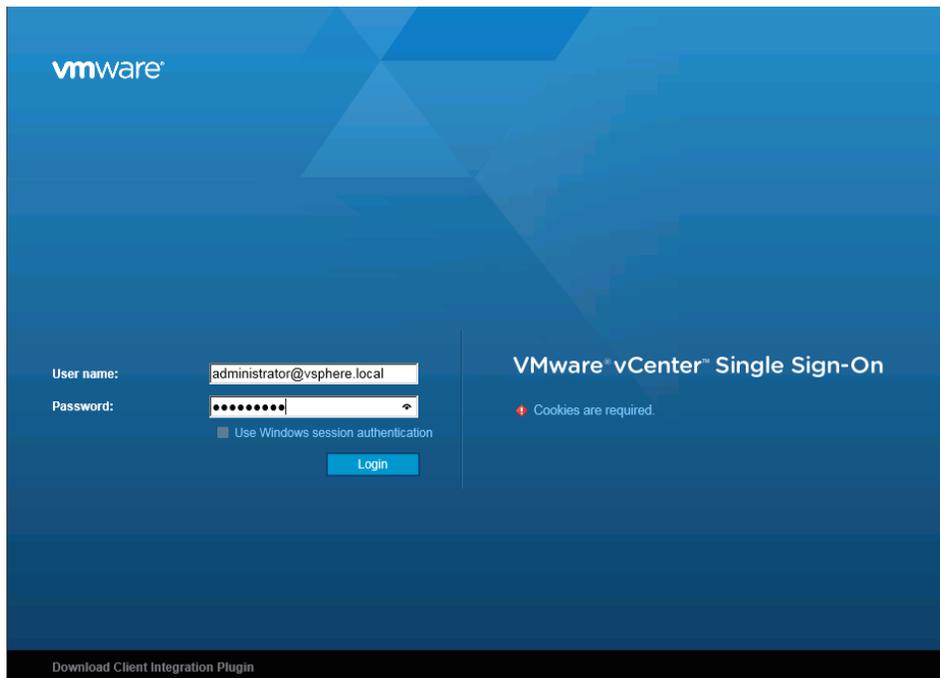
3. Enter the username and password that you specified during installation, and then click the **Login** button. The *VMware vSphere Web Client* page displays.
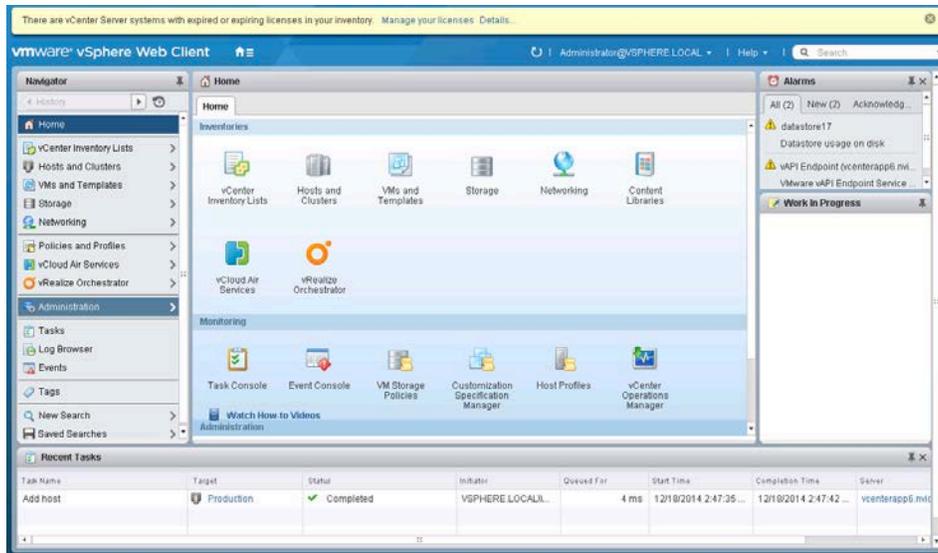
4. Click the Home icon (house) and then select **Administration**.

The *Administration* page displays.
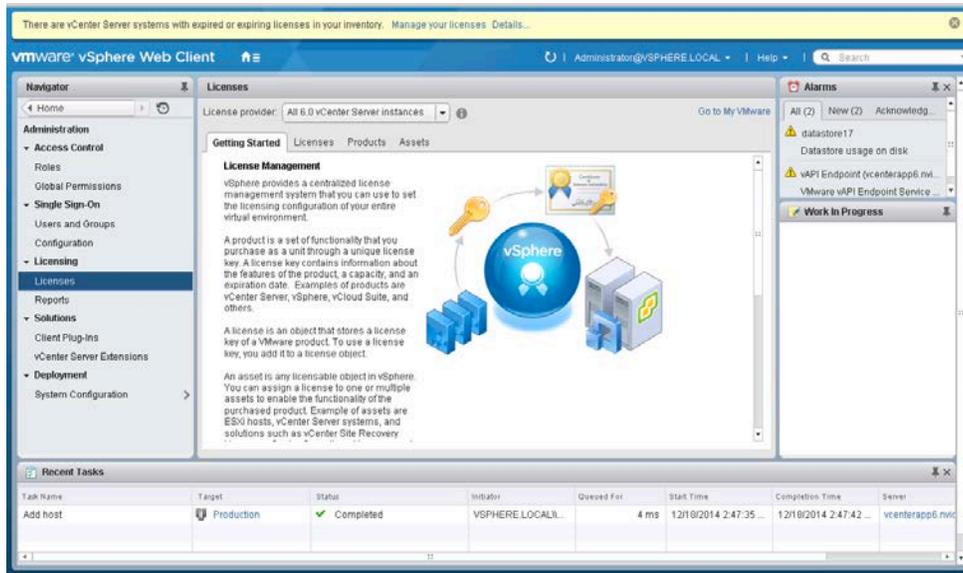
5. Select **Licenses** from the left-hand menu and then select the **Licenses** tab to open the **Licenses** tab. Click the green **New License** icon (plus sign) to open the *New Licenses* popup.
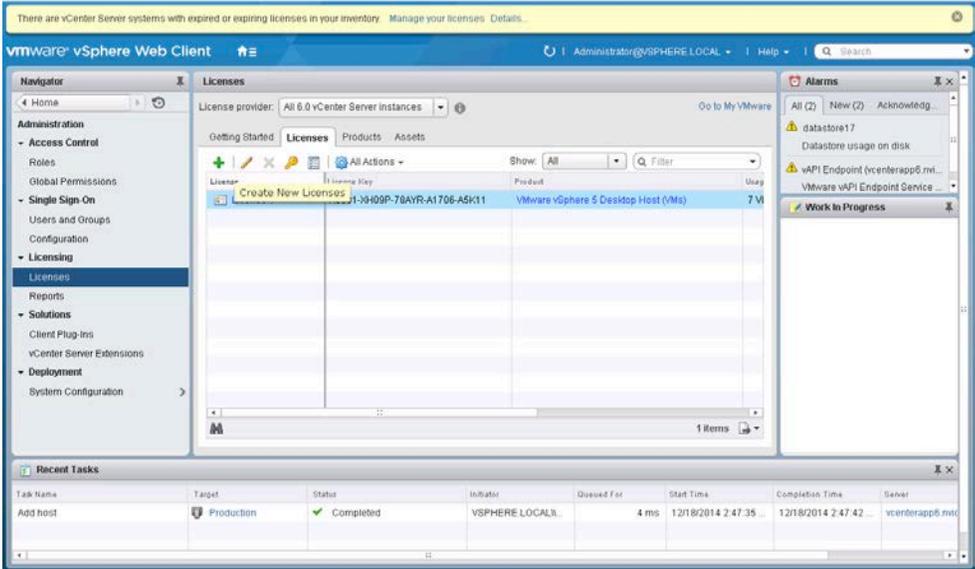
6. Enter the Center Server Standard license key
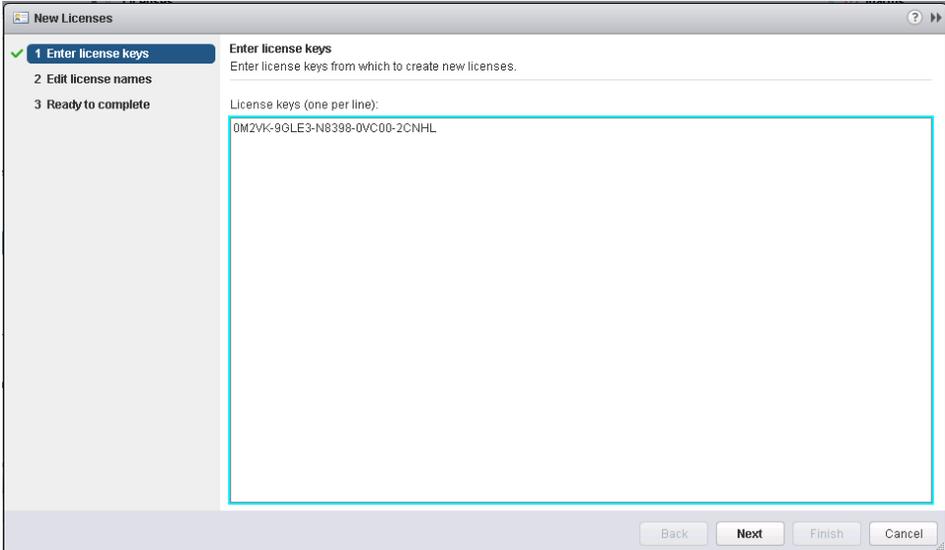
7. Enter a unique name for the license in the **License Name** field and then click **Next**.
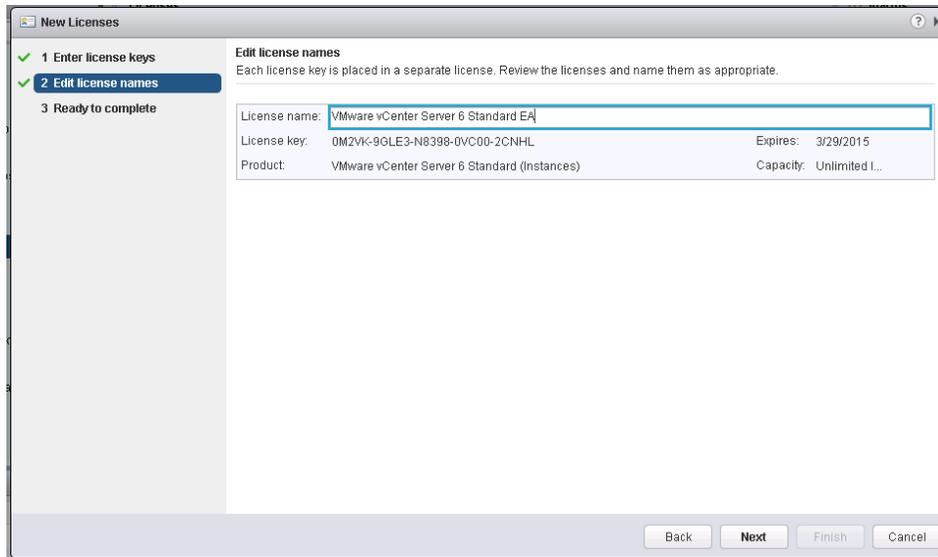
8. Review your selections and then click **Finish** to close the *Enter New License* popup and return to the VMware vSphere Web Client page.
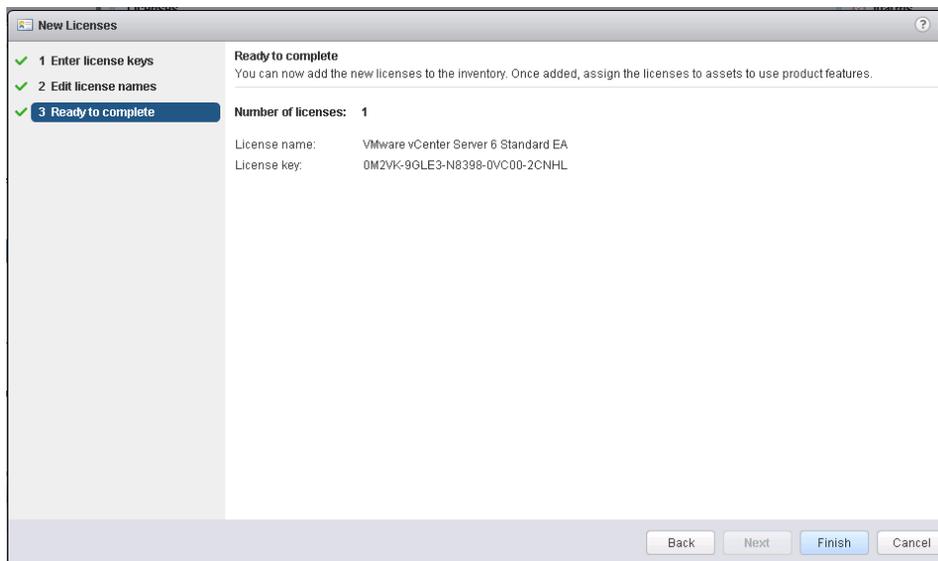
4.2.3   Adding a Host

Use the following procedure to add a host in vCenter 6:

1. Select the **Home** icon (house) on the *VMware vSphere Web CIient* page

2. Select **Hosts** and **Clusters**.

3. Select the desired host.

4. Go to the **Getting Started** tab and select **Create Datacenter**.
   The New Datacenter popup Displays
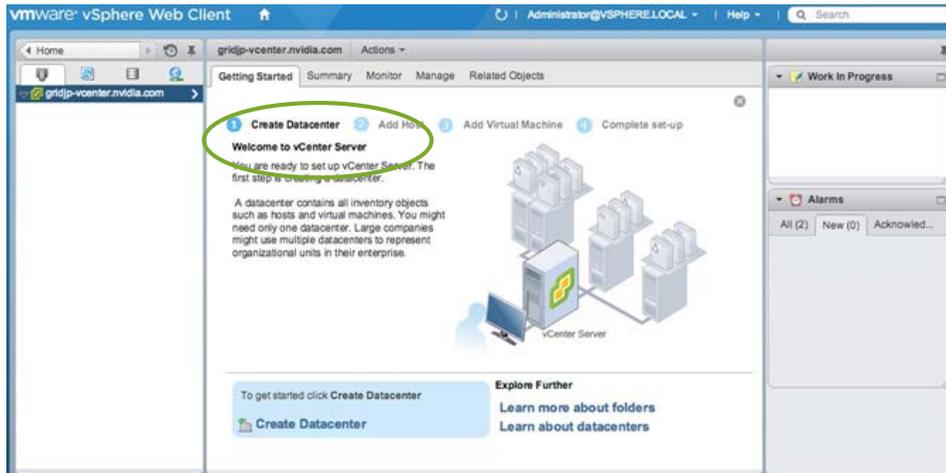
5. Enter a name for the datacenter in the **Datacenter name** field and click **OK**.
   The new datacenter is visible in the left panel of the *vSphere Web Client*.
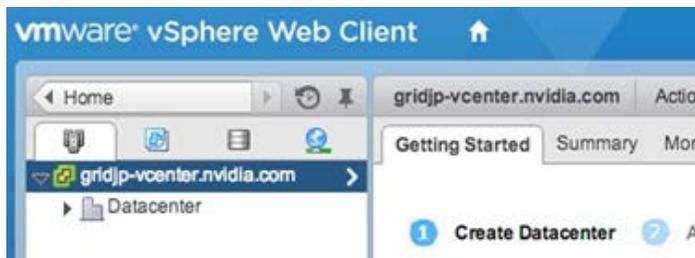
Select the **Getting Started** tab and select **Add a Host**.
The *Name and location* dialog opens.
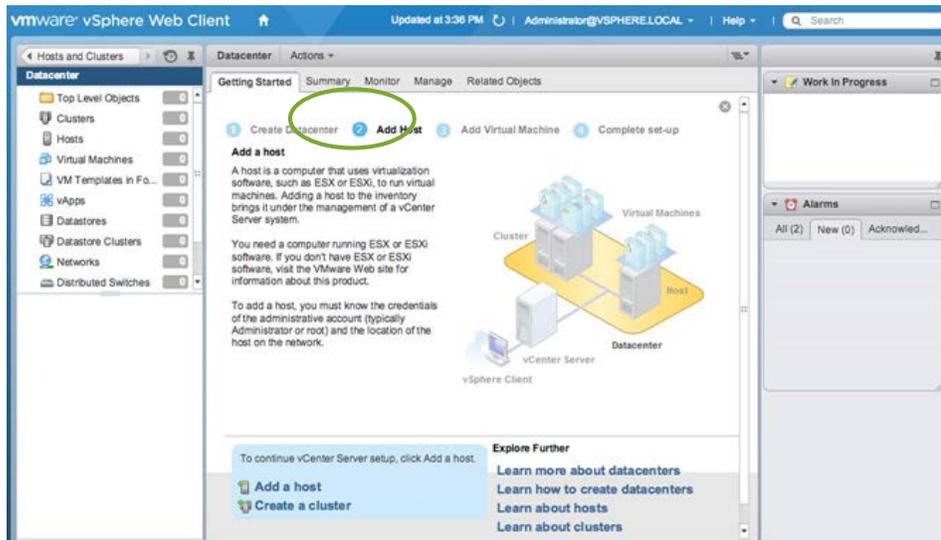
6. Enter the host name or IP address of the vSphere host and click **Next**.
   The *Connection settings* dialog displays.

7. Enter the administrator account credentials in the **User name** and **Password** fields and
   click **Next**. The *Security Alert* popup displays.

8. Click **Yes** to replace the host certificate.
   The *Host summary* dialog displays.

9. Review the settings and click **Next** to proceed.
   The *Assign license* dialog displays.

10. Confirm the license selection and click **Next**.
    The *Lockdown mode* dialog displays.

11. Accept the default setting (Disabled) and click **Next**.
    The *VM location* dialog displays.

12. Select a cluster or accept the default option and click **Next** to proceed.
    The *Ready to complete* dialog displays.

13. Click **Finish** to complete adding the new host.
    The new host is now visible in the left panel when you click the datacenter name.
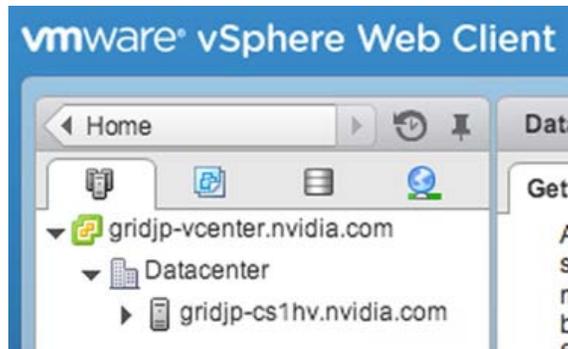
4.2.4   Setting the NTP service on a host

1. Click a host object in the menu on the left, click **Manage**➔**Settings**➔**Time Configuration**➔**Edit**
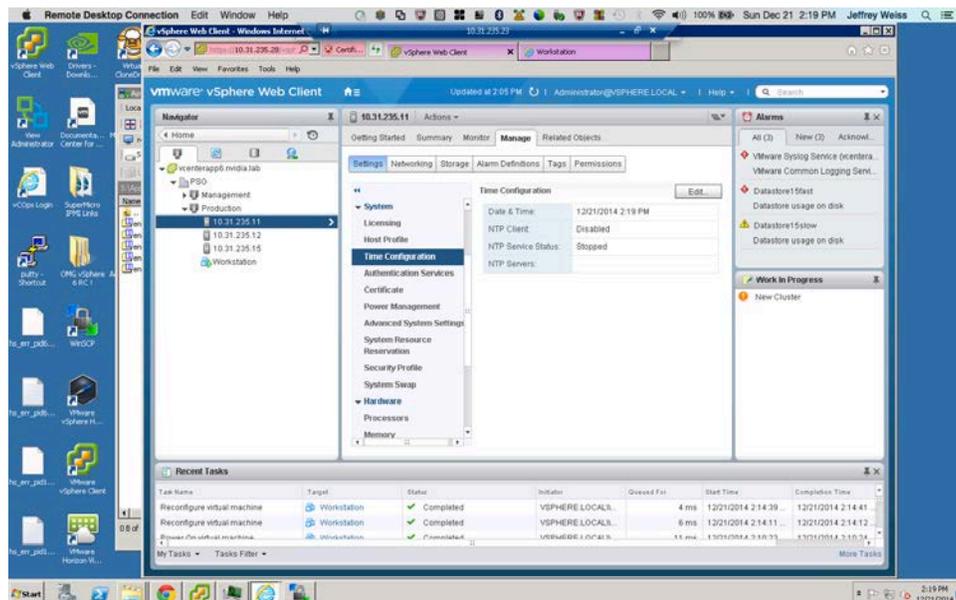
2. Enter a valid time server and click **OK**:
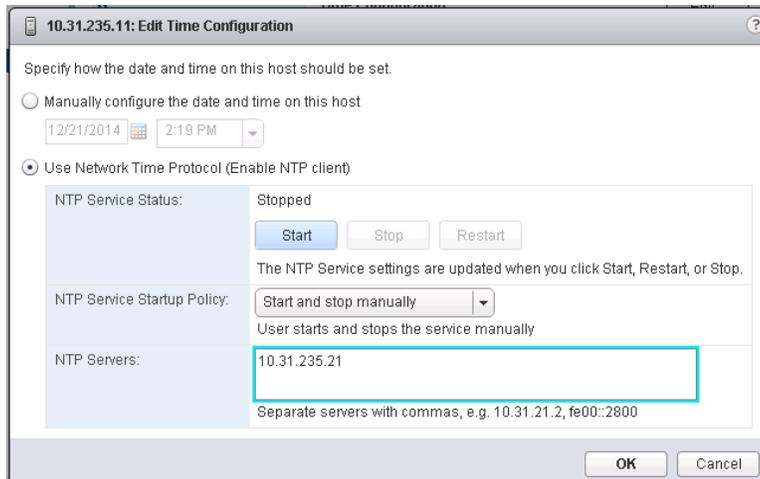
3. Do this for each host to ensure time is accurate for all guests.

4.2.5   Setting a vCenter Appliance to Auto-Start

Use the following procedure to set a vCenter Appliance to start automatically:

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

2. Click the **Edit** button.
    The *Edit VM Startup and Shutdown* window displays.
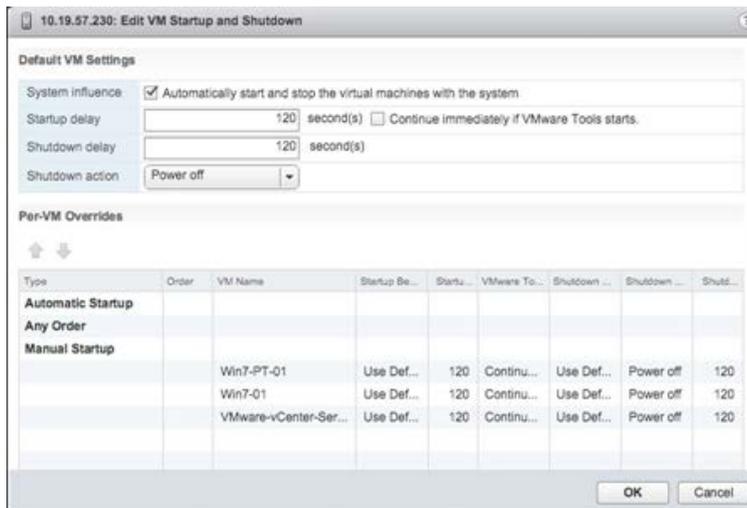
3. Select the **vCenter Appliance** and click the **Up** arrow to move that virtual machine up to the **Automatic Startup** section

4. 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
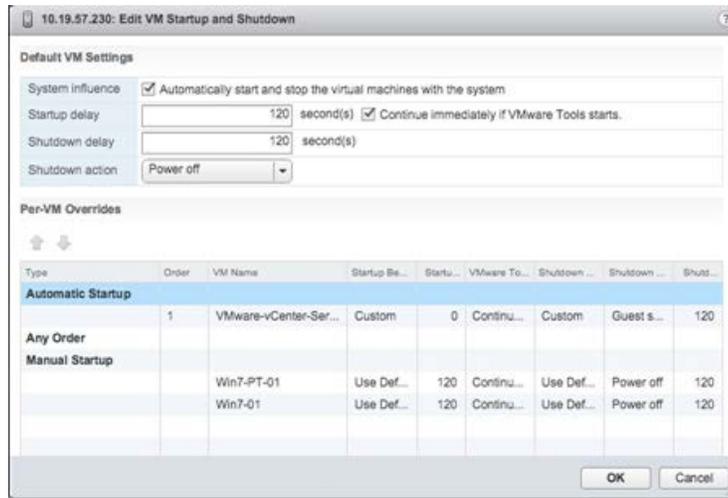
5. Click **OK** to apply the configuration.

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

4.2.6   Mounting an NFS ISO Data Store

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.
   The *New Datastore* window displays with the **Type** tab selected.
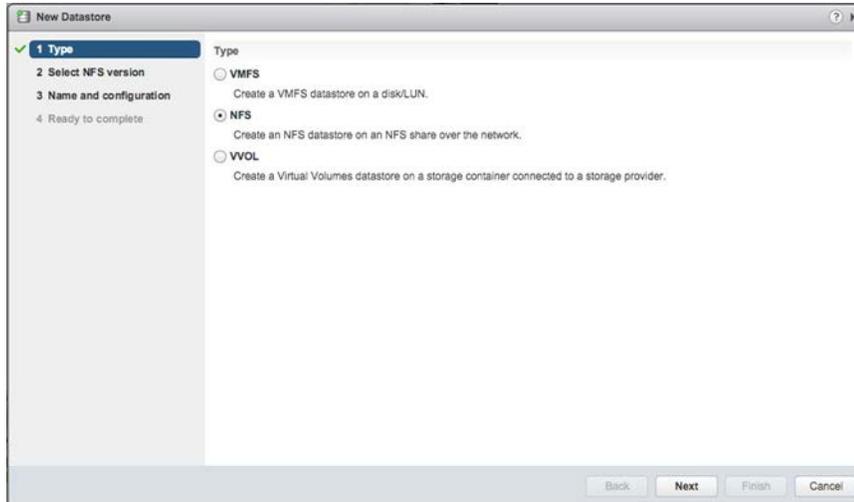
3. Select **NFS** and click **Next** to proceed.
   The **Select NFS version** tab displays.

4. Select the correct NFS version and click **Next** to proceed.
   The *Name and configuration* tab displays.

5. Enter the NFS exported folder path and the NFS server address in the **Folder** and **Address** fields, respectively.
   Since this an ISO data store, consider mounting it as read-only by checking the **Mount NFS** as read-only checkbox.

6. Click **Next** to proceed.
   The **Host accessibility** tab displays.

7. Select the host that will use the new data store,

8. Select **Next** to proceed.

   The **Ready to complete** tab displays.

9. Review the settings.

10. 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.

Chapter 5.
NVIDIA VGPU MANAGER VIB INSTALLATION

This chapter discusses the following NVIDIA vGPU Manager `VIB` installation topics:

▸ Step-by-step guide to downloading and preparing the VIB file for installation on the host.

- VIB file upload using either vSphere Web Client (see section 5.1.1, Uploading VIB in vSphere on page 58)
- WinSCP (see Chapter 15, *Using WINSCP*), on page 206)

▸ Installing the `VIB`.

▸ Uninstalling the `VIB`.

▸ NOTE: Upgrading from previous versions of the VIB is not supported.

5.1 PREPARING THE VIB FOR INSTALL

Use the following procedure to prepare the `VIB` file:

1. Download the archive containing the `VIB` file.

2. Extract the contents of the archive to a folder. The file ending with `VIB` is the file that you must copy to the host data store for installation.

5.1.1   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 206 if you are using WinSCP.

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

1. Click the **Related Objects** tab for the desired server

2. Select **Datastores**.

3. Either right click the data store and then select **Browse Files,** or click the icon in the toolbar. The *Datastore Browser* window displays.
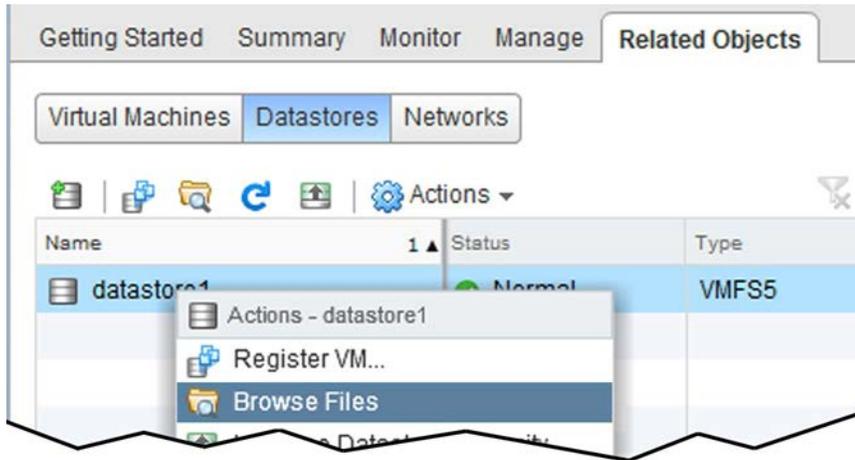
4. Click the **New Folder** icon.
    The *Create a new folder* window displays.

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

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

7. Click the **Upload** icon.
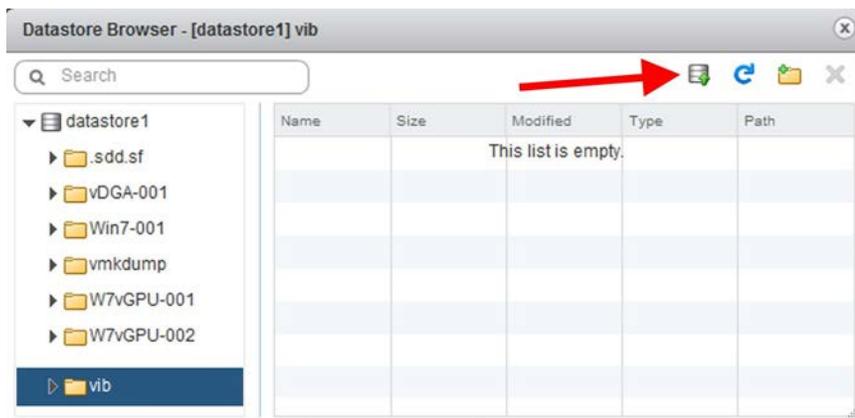
The *Client Integration Access Control* window displays.

8. Select **Allow**.

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.

5.2 PREPARE AND INSTALL THE `VIB` FILE

This section guides you through preparing, installing, and confirming the `VIB` file.

1. Upload the `VIB` file to the host using either the vSphere Web Client Utility (see the following sections for more detailed information: see section 5.1.1, Uploading VIB in vSphere on page 58; Chapter 15, *Using WINSCP*, on page 206)

2. Place the host into Maintenance mode by right-clicking it and then selecting **Maintenance Mode➔Enter Maintenance Mode**.
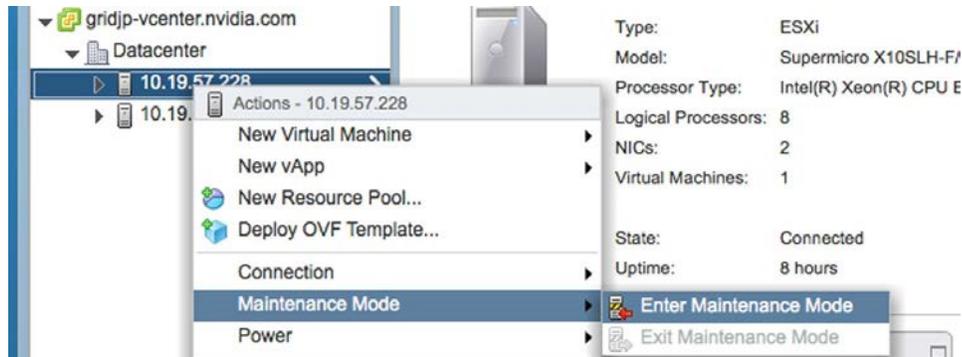
> 💬 Alternatively, you can place the host into Maintenance mode using the command prompt by entering
> `$ esxcli system maintenanceMode set -- enable=true`
> This command will not return a response. Making this change using the command prompt will not refresh the vSphere Web Client UI. Click the **Refresh** 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.

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

5.2.1 Installing the VIB

Use the following procedure to install the **VIB** file, NVIDIA-vgx-VMware-346.27-1OEM.600.0.0.2159203.x86_64.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 --no-sig-check
-v /<path>/NVIDIA-vgx-VMware-346.27-1OEM.600.0.0.2159203.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-vgx-
VMware_vSphere_6_Host_Driver_346.27-1OEM.600.0.0.2159203
VIBs Removed: VIBs Skipped:
```

> 💬 Although the display states "**Reboot Required: false**", a reboot is necessary for the vib to load and xorg to start.

2. From the vSphere Web Client, exit **Maintenance Mode** by right-clicking the host and selecting **Exit Maintenance Mode**.

> 💬 Alternatively, you may exit from Maintenance mode via the command prompt by entering:
> **$ esxcli system maintenanceMode set -- enable=false**
>
> This command will not return a response.
>
> Making this change via 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.

3. Reboot the host from the vSphere Web Client by right-clicking the host and then selecting **Reboot**.

> 💬 You can reboot the host by entering the following at the command prompt:
> **$ reboot**
> This command will not return a response. The Reboot Host window displays.

When rebooting from the vSphere Web Client, enter a descriptive reason for the reboot in the **Log a reason for this reboot operation** field, and then click **OK** to proceed.

5.2.2 Confirm the 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-vgx-VMware_vSphere_6_Host_Driver 346.27-1OEM.600.0.0.2159203
NVIDIA VMwareAccepted 2014-12-30
```

2. Confirm GPU detection using the following command:

```
$ nvidia-smi
```

This command returns output similar to the following (output shown below is for a pair of K2 cards in a single host):

```
+------------------------------------------------------------------------+
| NVIDIA-SMI 346.27  Driver Version: 346.27   |
|------------- -------------+----------------+----------------------+
|GPU Name      Persistence-M |  Bus-Id Disp.A |  Volatile Uncorr. ECC |
|Fan Temp Perf Pwr:Usage/Cap|    Memory-Usage | GPU-Util Compute M.|

|========================+================+======================|
|   0 GRID K2   Off|0000:09:00.0 Off|        Off |
|N/A 52C P8    30W / 117W |917MiB/4095MiB | 7%   Default |
+--------------------------+----------------+------------------------+
|   1 GRID K2   Off|0000:0A:00.0 Off|        Off |
| N/A 43C      P0    50W / 117W |11MiB/4095MiB    |    0%   Default |
+--------------------------+----------------+------------------------+
|   2 GRID K2   Off|0000:85:00.0 Off|        Off |
| N/A 42C      P0    51W / 117W|11MiB /4095MiB  | 0%   Default |
+--------------------------+----------------+------------------------+
|   3 GRID K2   Off|0000:86:00.0 Off|   Off |
| N/A 37C      P0    39W / 117W| 11MiB /4095MiB | 0%   Default |
+--------------------------+----------------+------------------------+


+------------------------------------------------------------------------+
| Compute processes: GPU Memory |
| GPU    PID  Process name Usage |
|=====================================================================|
| No running compute processes found   |
+------------------------------------------------------------------------+
```

If the **nvidia-smi** command does not return any output (see Chapter 14, *Troubleshooting*, on page 193)

3. SMI also allows GPU monitoring using the following command:

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

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

5.3 UNINSTALLING THE VIB

This section describes how to uninstall the **VIB**. NOTE: Upgrading is not supported.

5.3.1   Uninstalling VIB

Use the following procedure to uninstall **VIB**:

1. Determine the name of the vGPU driver bundle as described in section 5.2.2, *Confirm the VIB Installation*, on page 61

2. Run the following command to uninstall the driver package:

```
$ esxcli software vib remove -n NVIDIA-vgx-
VMware_vSphere_6_Host_Driver --maintenance-mode
```

The following message displays when installation is successful:

```
Installation Result:
Message: Operation finished successfully. Reboot Required: false
VIBs Installed:
VIBs Removed: NVIDIA_bootbank_NVIDIA-vgx-
VMware_vSphere_6_Host_Driver_340.49- 1OEM.600.0.0.1951118
VIBs Skipped:
```

3. Reboot the host to complete the uninstallation process.

Chapter 6.
BUILDING CITRIX XENDESKTOP 7.6

This chapter covers installing Citrix XenDesktop 7.6, including:

▶ Building XenDesktop 7.6

▶ Registering and allocating the trial license

▶ Installing a XenDesktop

▶ Importing vSphere and vSphere SDK Certificates

▶ Using Citrix Studio to performing site setup

▶ Configuring Citrix license server

▶ Verify Citrix StoreFront Configuration

6.1 REGISTERING AND ALLOCATING THE TRIAL LICENSE

Use the following procedure to obtain and allocate a Citrix XenDesktop 7.6 License:

1. Using an internet browser. Navigate to the
   http://**www.citrix.com/tryxendesktop** website. Select the Download Now button
   to register for a new trial or login to your existing account

2. Enter your **Login ID** and **Password** and click the **Login** button if you already have a Citrix account. If you do not have an account you can create one by selecting the **Create an Account** tab.
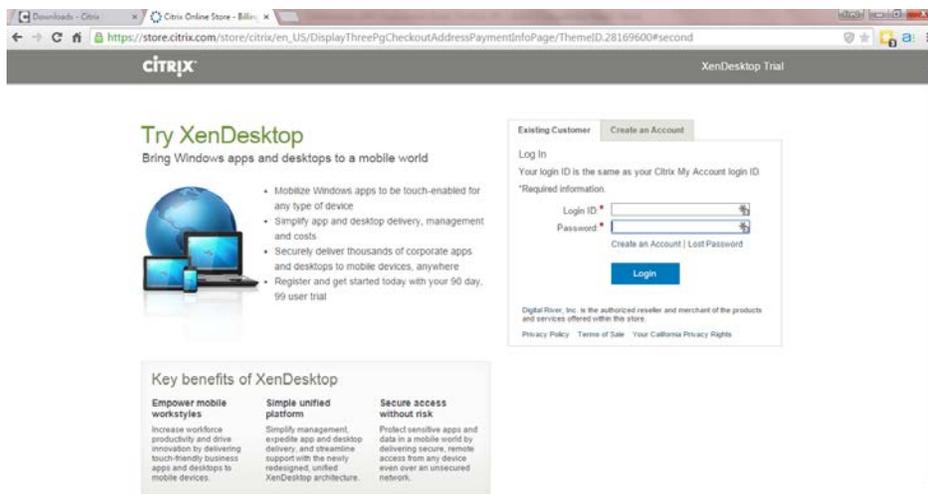
3. To create your account enter all required information as directed, select the checkbox confirming that **I have read and accept the Citrix End User License Agreement** and select **Proceed to Download** to continue.
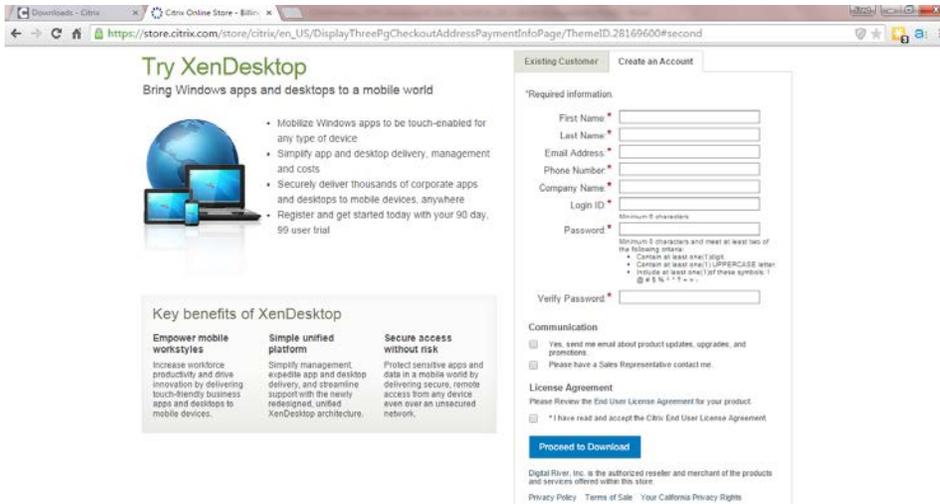
4. Existing customers will be required to confirm their account information and select the checkbox confirming that **I have read and accept the Citrix End User License Agreement.** Select the **Proceed to Download** button to continue**.**
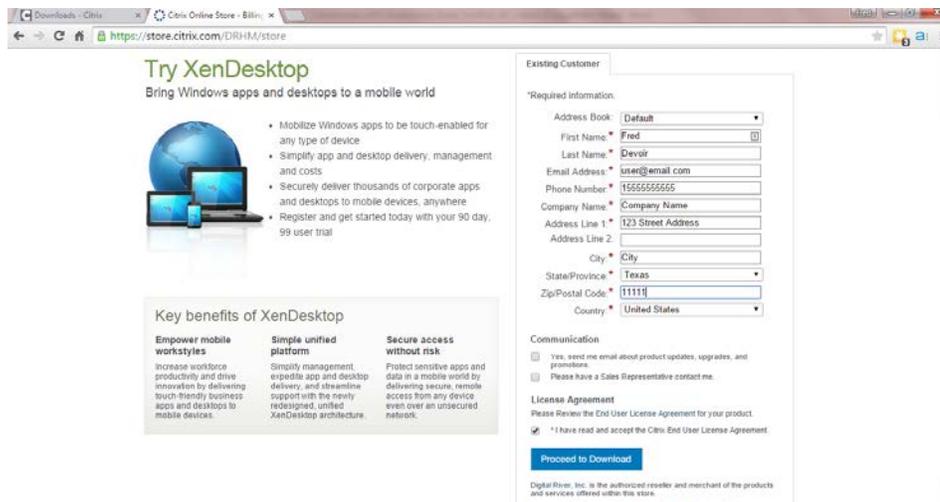
5. Select the **Download XenDesktop 7.6** link to download the .iso image needed to install the Citrix XenDesktop software. There is also a helpful Reviewer's Guide on this page, which walks you through the basic deployment in a matter of hours. Select **My account** link under **Step 2** do proceed to your account and allocate the license.
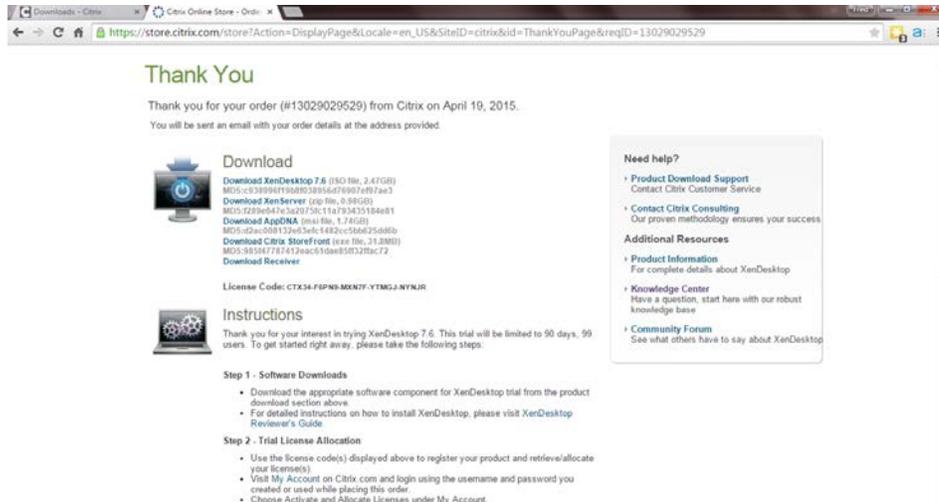
6. Open the email that you received from **citrix.cs@digitalriver.com** as it contains the **License Code** needed to allocate your trial license. This code should match the one displayed in your Citrix account.
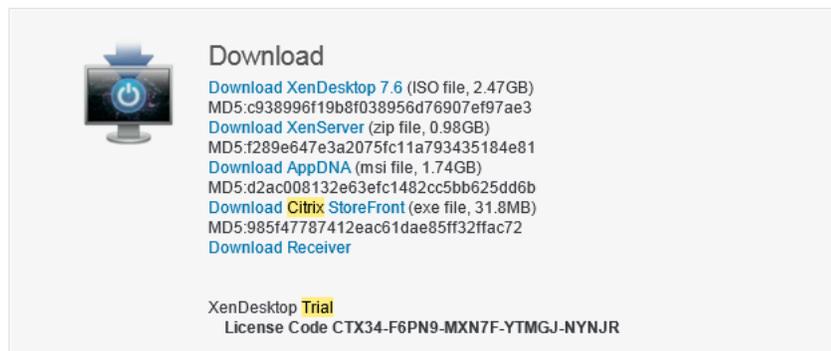
> You may choose to enter the license key later. Even without a license key, all features are enabled and the evaluation will run successfully for up to 30 days

7. Under the **Activate and Allocate Licenses** tab, select the **check box** on the left side of the line that contains the Code from your email. Select the **Continue** button to proceed.
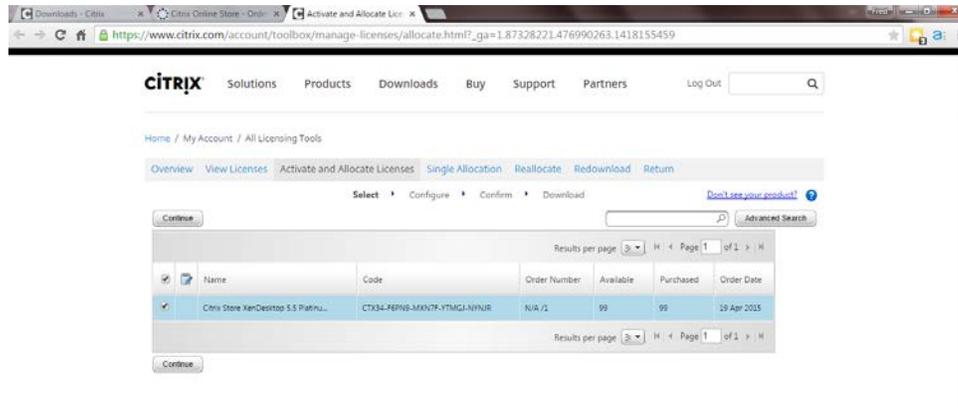
8. Type the NETBIOS hostname of the server where you will install XenDesktop 7.6 into the Host ID field. Select the continue button to proceed.
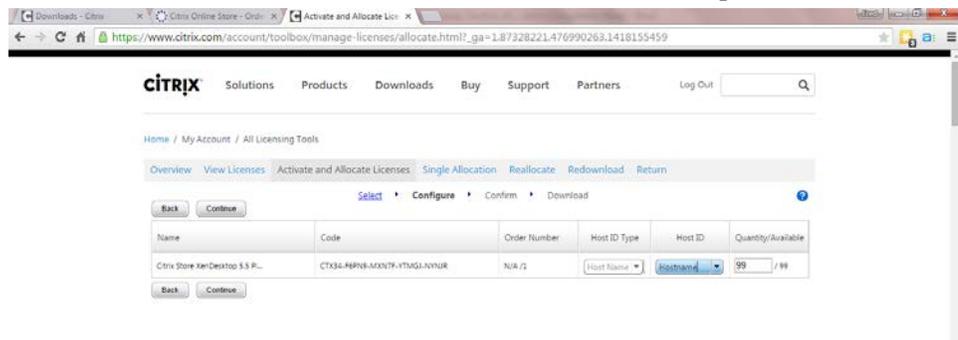
💬 The Host ID field is case sensitive. If you do not enter your Host ID exactly the same as your NETBIOS Hostname it will not work properly. Do not enter the FQDN name of your server.

9. Confirm your selection and Host ID, select the Confirm button to proceed to download your license file.
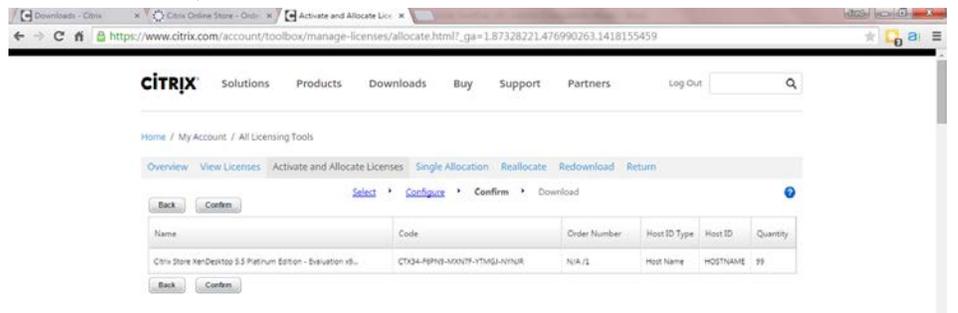
10. A confirmation dialogue will be displayed, select **OK** to proceed to download your license file.
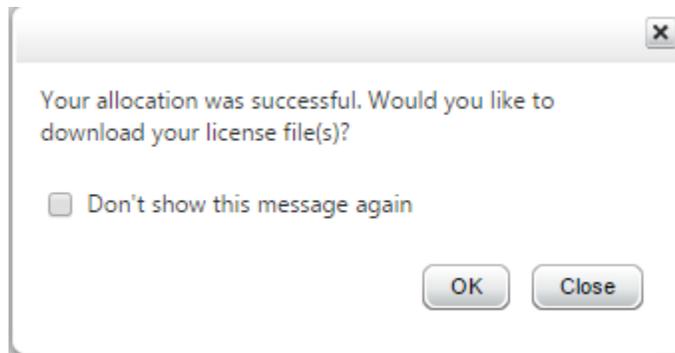
11. Select the **checkbox** on the left side of the license you with to download. Select the **download** button to start your download.
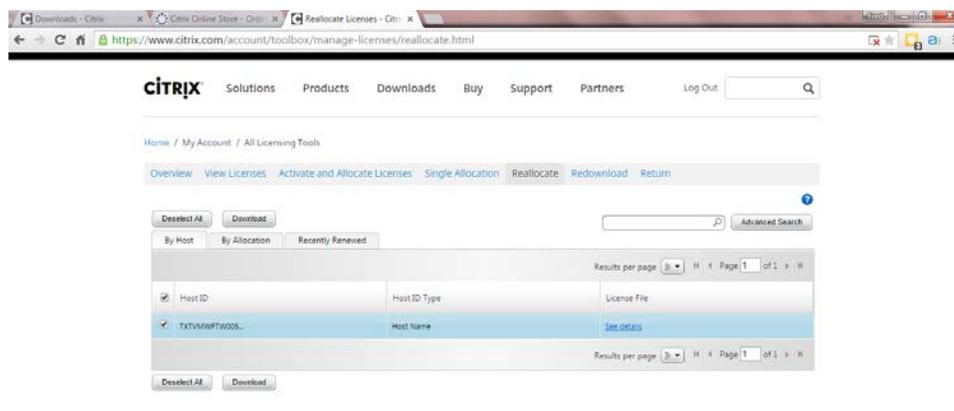
12. Once the download completes, copy the license file to the server where you will install XenDesktop 7.6.

6.2 INSTALLING XENDESKTOP 7.6

The server on which you are installing Citrix XenDesktop 7.6 must meet the requirements listed in "General Prerequisites" on page 14. As of the publication date of this manual the current version of the Citrix XenDesktop is:

Version: 7.6.0.5029
XenApp_and_XenDesktop7_6.iso

Use the following procedure to install XenDesktop 7.6:

1. Mount the .iso image using Virtual Clone Drive or similar. Alternately you can use vCenter Client to mount the .iso from an ISO data store to the virtual machine CD drives.
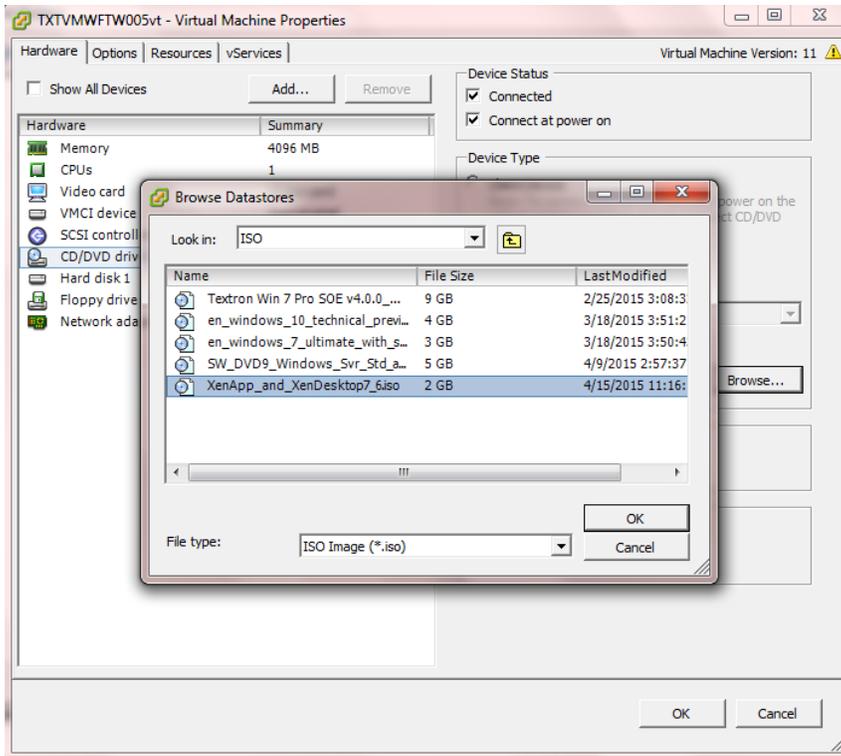
2. Launch the CD with AutoPlay or select the AutoSelect.exe application to start the Citrix XenApp and XenDesktop installer. Select **Start** button next to XenDesktop Deliver applications and desktops.
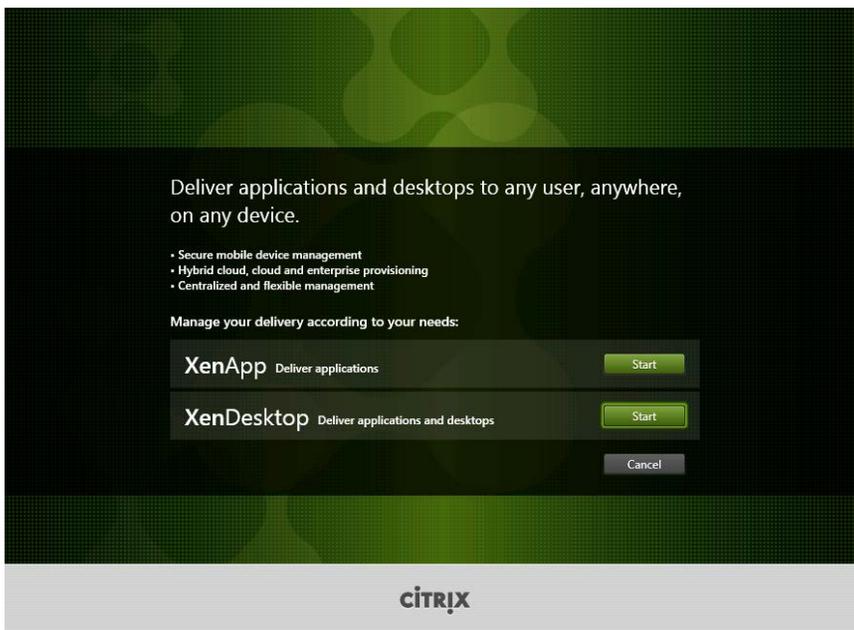
3. Select **Delivery Controller** under Get Started on the *XenDesktop 7.6* window.
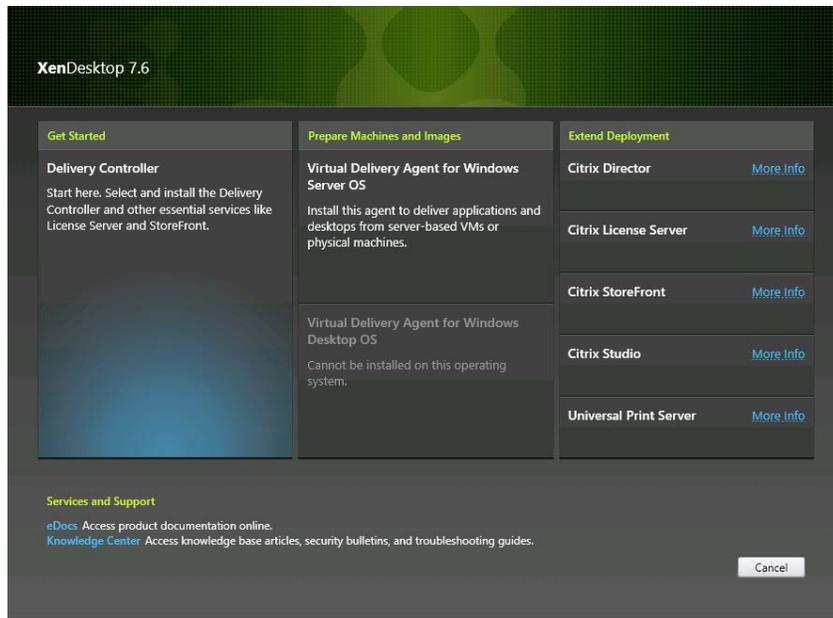
4. Select **Yes** to agree to the User Account Control prompt if displayed.
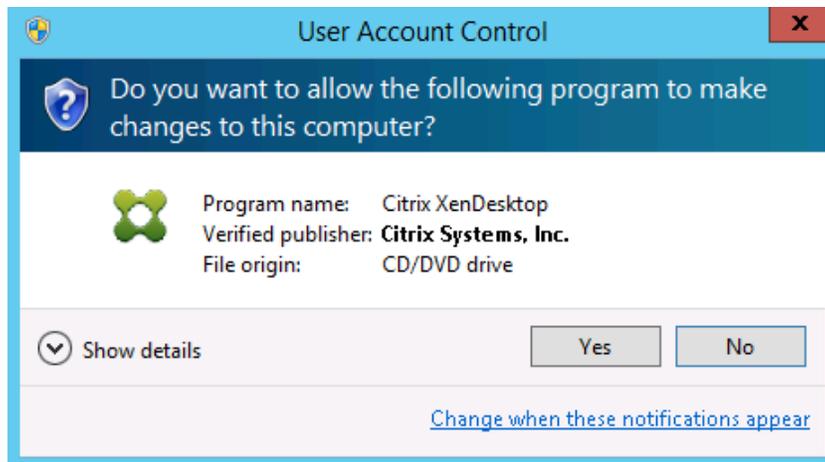
5. Check I have read, understand, and accept the terms of the license agreement radio button to accept the agreement. Select Next to continue.
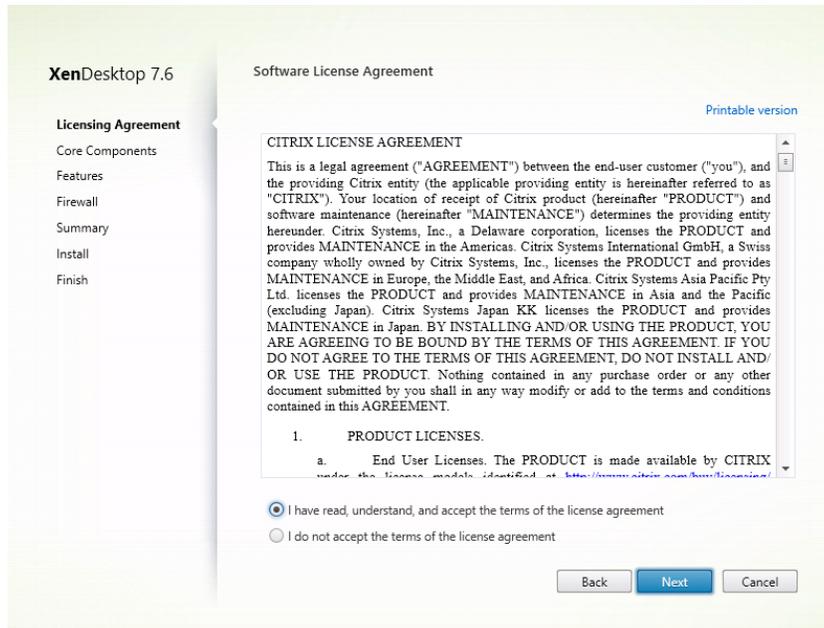
6.Select the Core Components that you wish to install, by default all components are
   selected. For Single Server installation select **Next** with all Core Components
   checked to continue.
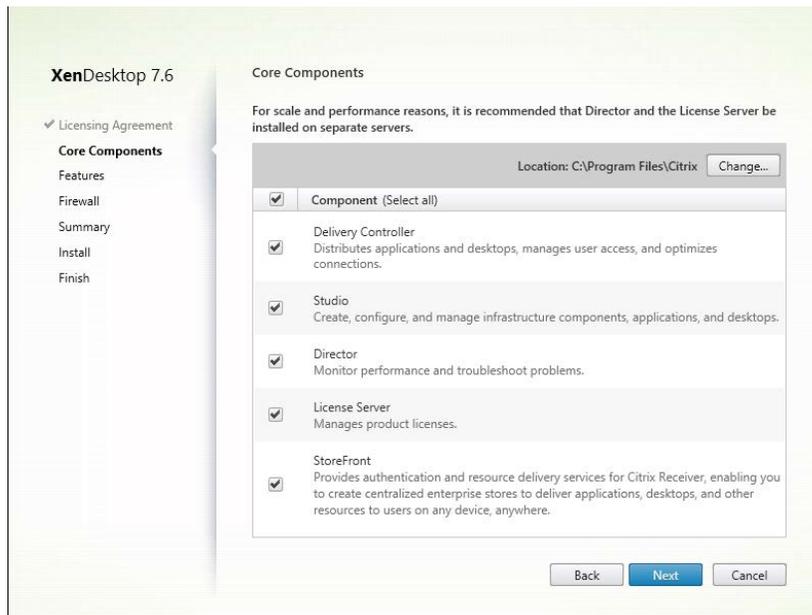
7.Select **Features** needed for this installation. Citrix best practice is to use a SQL Server
   separate from the XenDesktop server. For single server installation accept the
   default selections and select **Next** to continue.
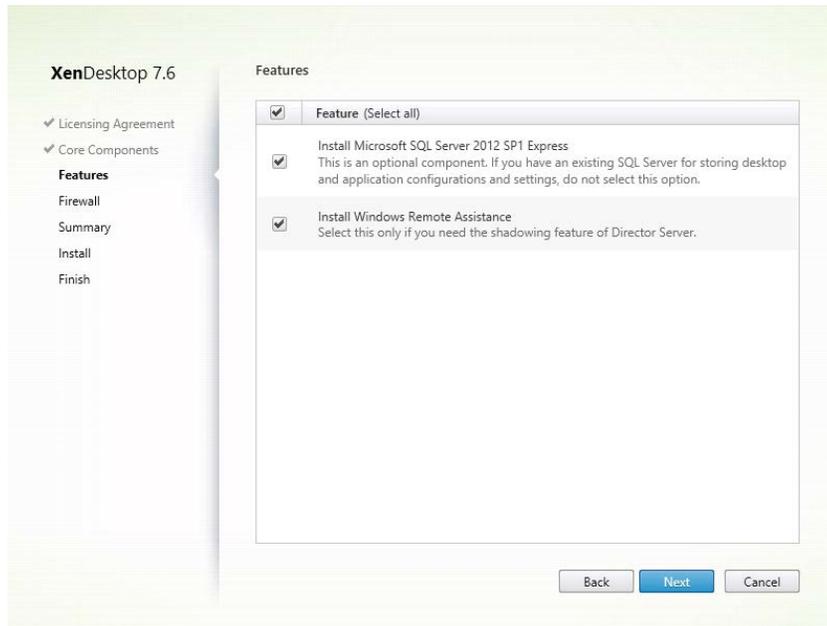
8. Automatically create the firewall rules in Windows Firewall, unless you are not using Windows Firewall. Select Next to continue.
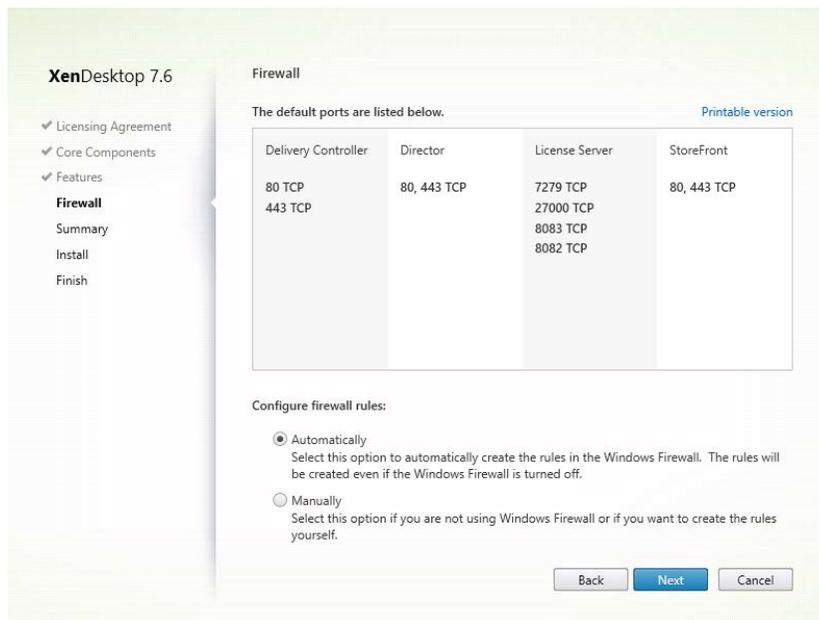
9. Review Summary and Select **Install** to begin installation.
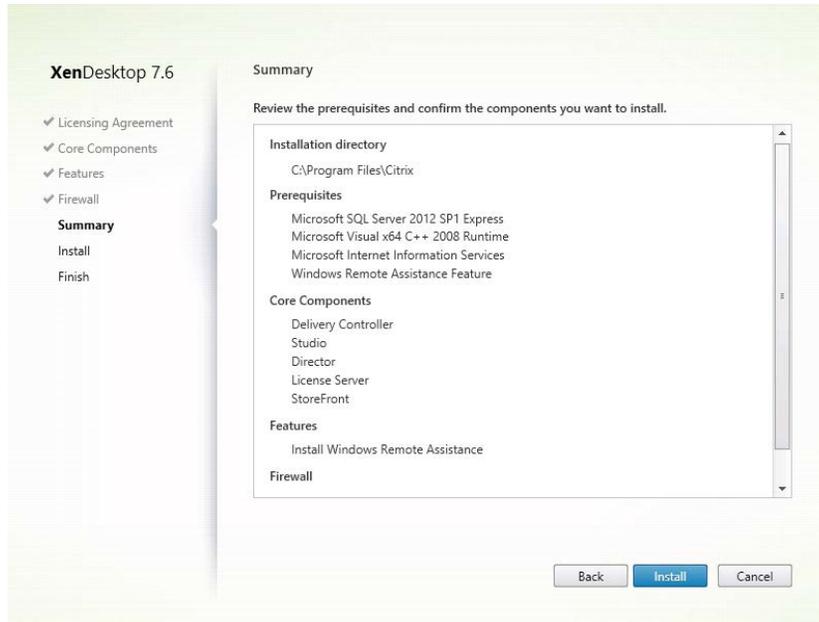
10. Select **Finish** when the *Finish Installation* dialogue is displayed.
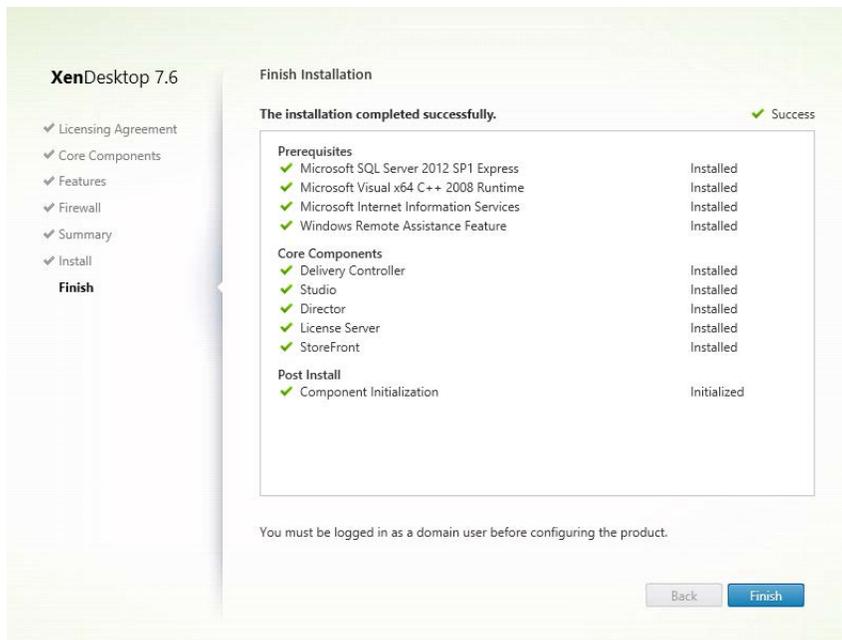
6.3 IMPORTING VSPHERE AND VSPHERE WEB SERVICE SDK CERTIFICATES

Before you can continue with the XenDesktop configuration import the VMware vSphere 6.0 server certificate and vSphere Web Service SDK Certificates. Using Internet Explorer or another web browser, type the web address of the vCenter server into the address bar.

6.3.1 Importing vSphere Certificate

1. Select **Continue to this website (not recommended).**
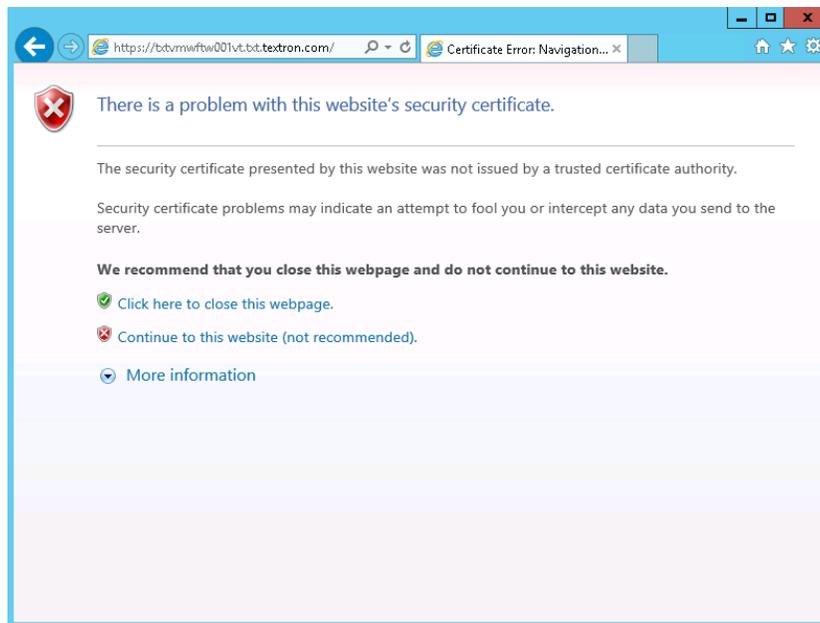
2. Click on the **certificate warning** next to the URL in the address bar and select **View certificates**.
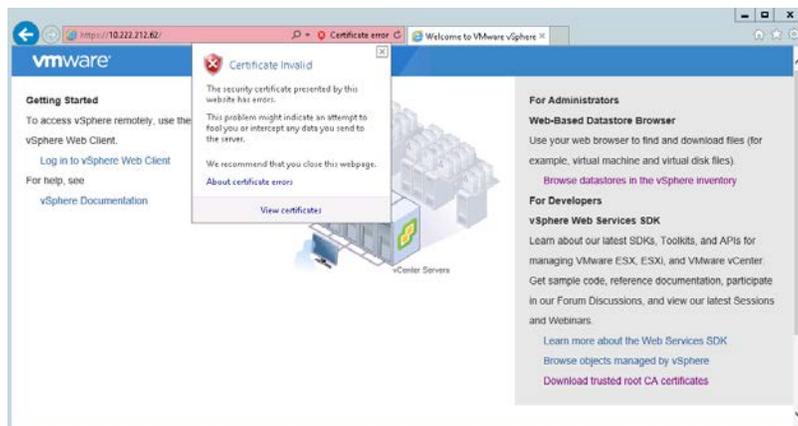
3. Select the **Install Certificate** button to import the certificate into the local certificate store.
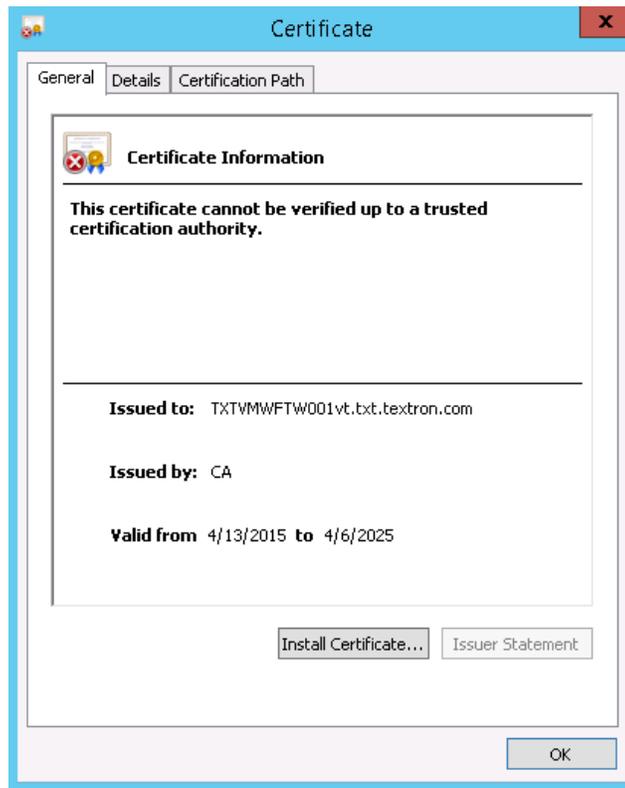
4. Select **Local Machine** radio button on the **Certificate Import Wizard** and then select **Next** to import the server certificate into the local machine certificate store.

5. Select **Yes** to agree to the User Account Control prompt if displayed.
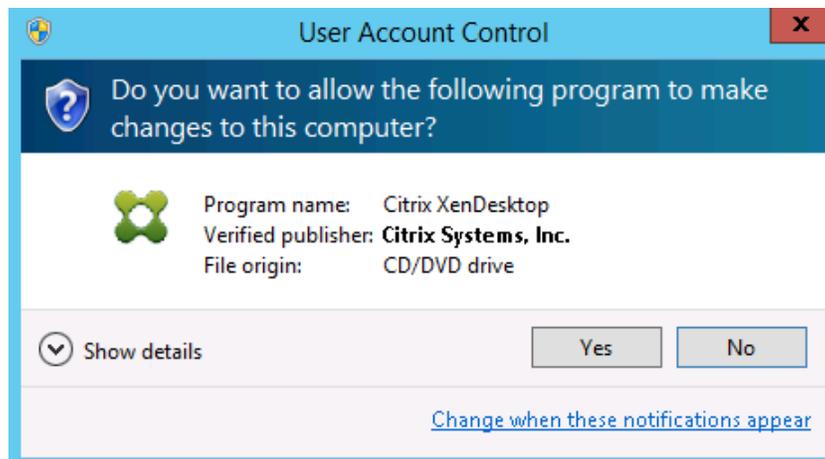
6. Select **Next** to accept the default to automatically select the certificate store based on the certificate type.
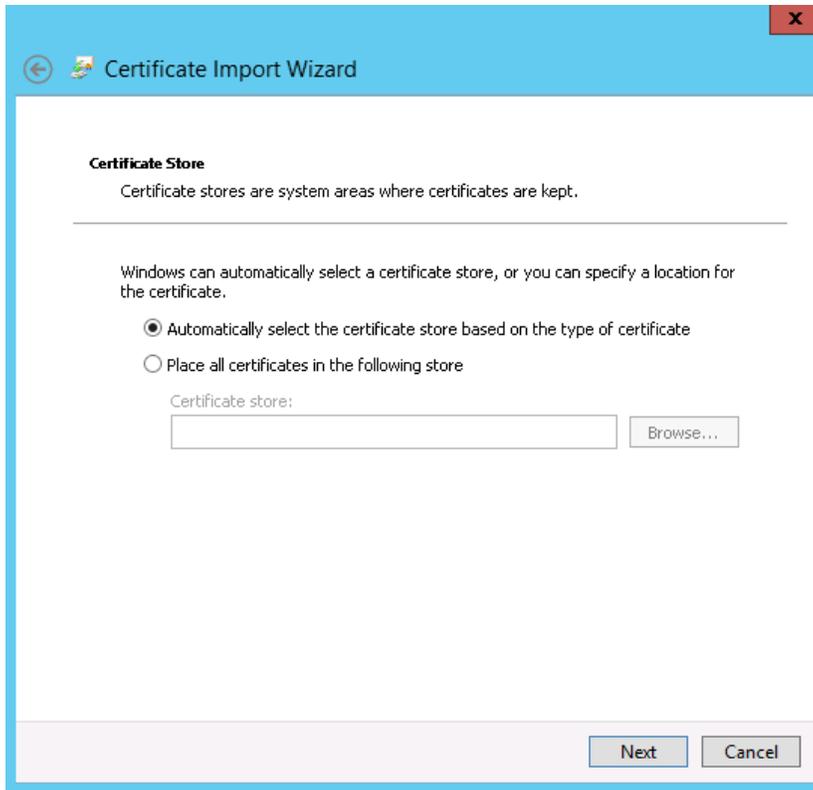
7.Select **Finish** to continue.
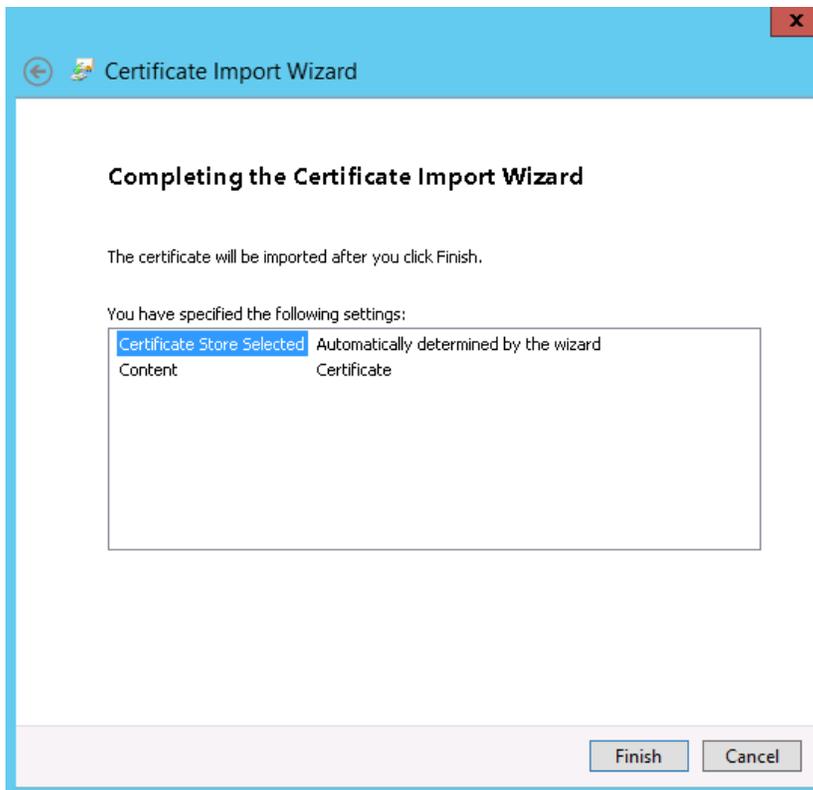
8. Select **OK** to close the notice that the certificate was imported successfully.
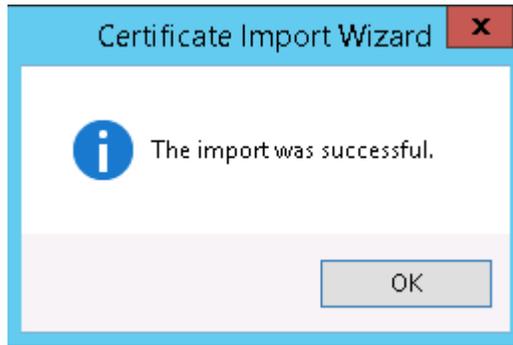
6.3.2   Importing vSphere Web Services SDK Certificates

1. Select **Download the trusted root CA certificates** from the vSphere Web Services SDK section.
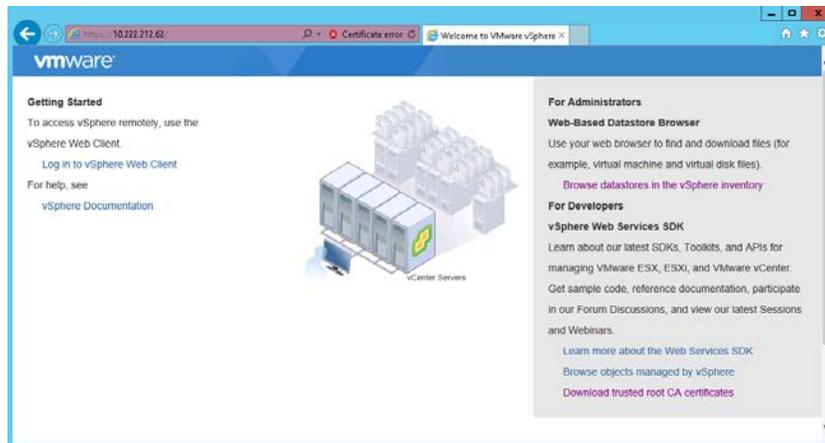
2. Save and then extract the certs from the download.zip file.

3. Launch MMC control panel from the **Start➔Run** menu.

4. Select **Yes** to agree to the User Account Control prompt if displayed.
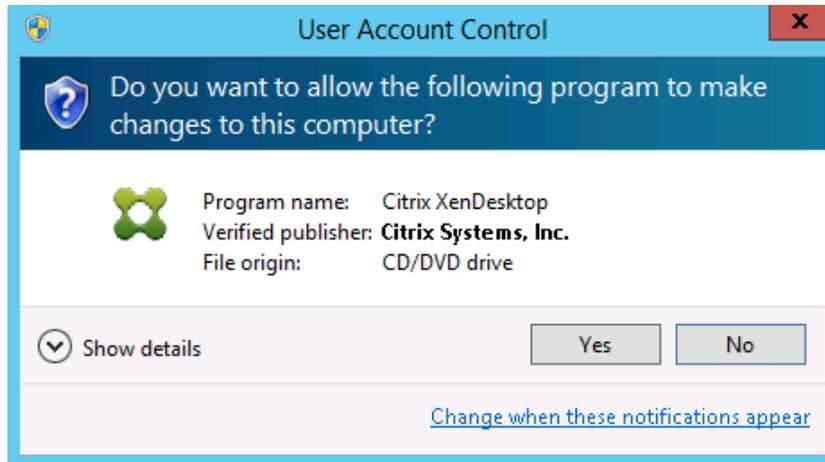
5. Select **Add/Remove Snap-in** from the **File** menu.
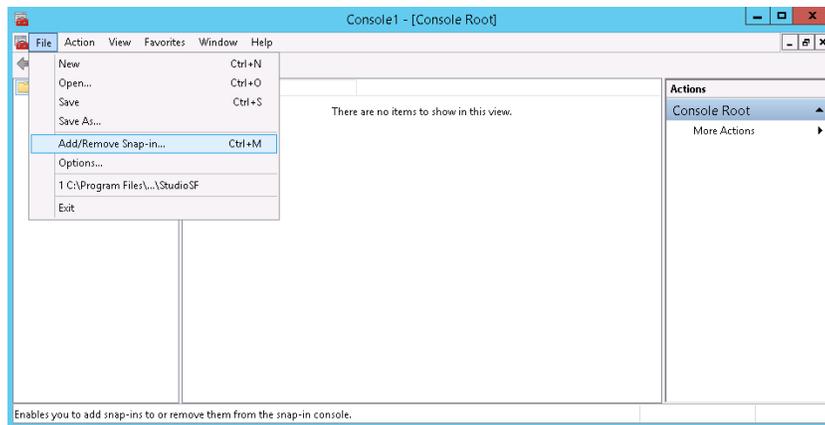
6. Select **Certificates** from the **Available snap-ins** and then click **Add** button.
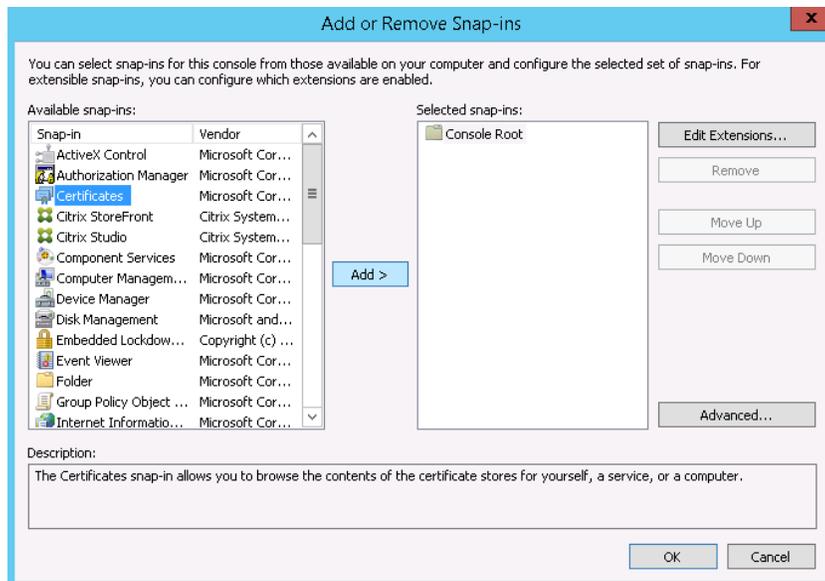
7. Select **Computer Account** radio button and then select **Next** to continue.
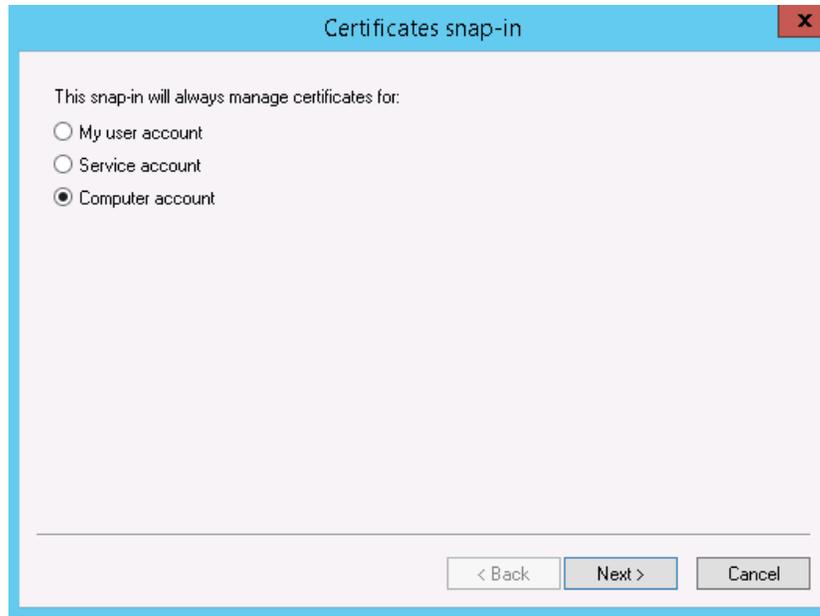
8. Accept the default of **Local computer** as the computer you want to manage, and then select **Finish** to continue.
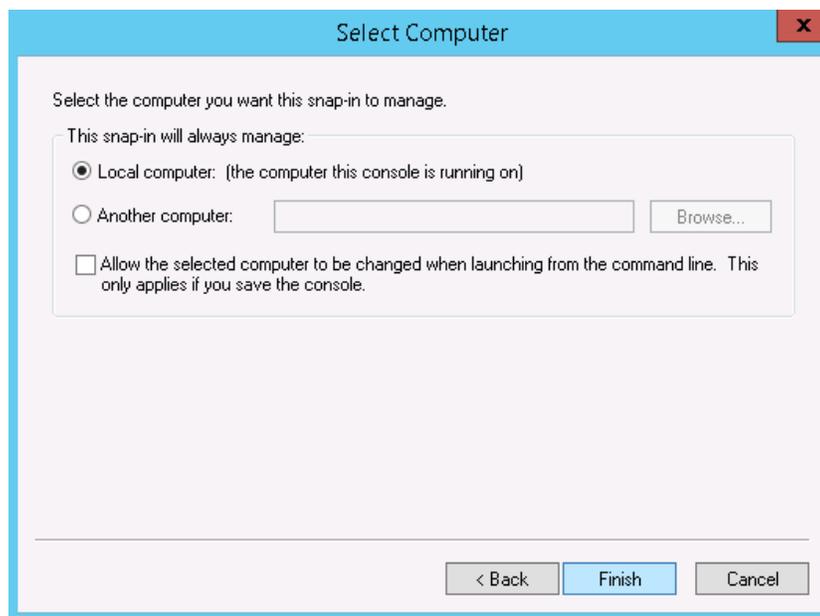
9. Select **OK** to load the snap-in.
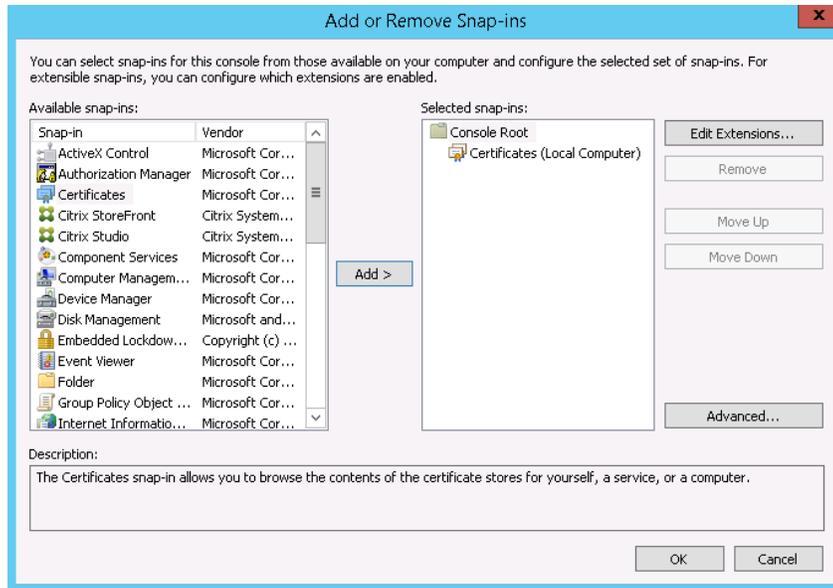
10. Expand **Certificates**, and then expand **Trusted Root Certificates**. Right click on **Certificates** under **Trusted Root Certificates** and select **All Tasks→Import**.
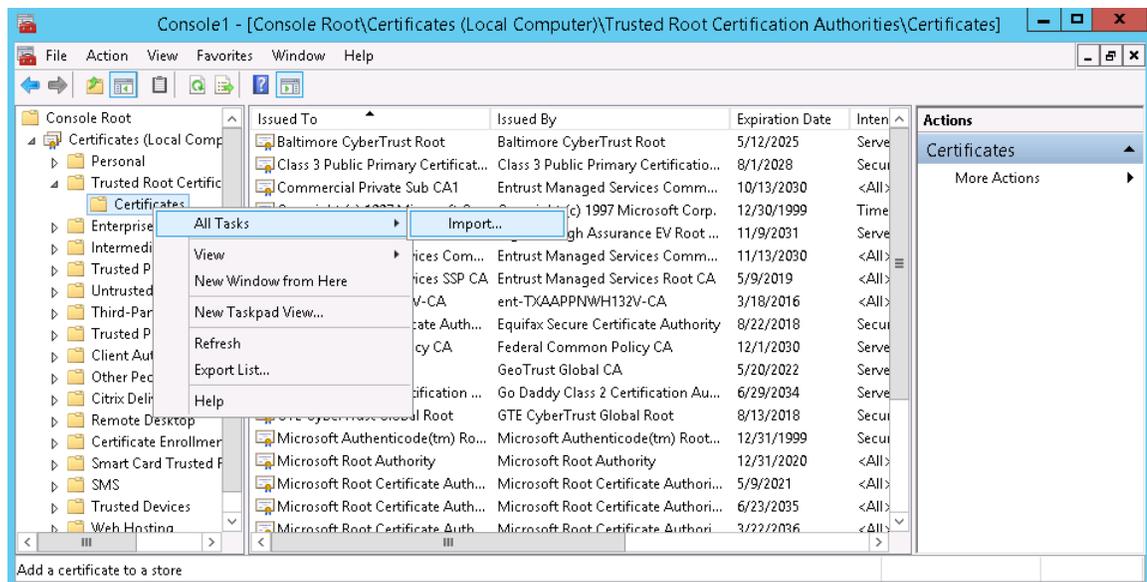
11. Select **Next** to continue to import certificates using the **Certificate Import Wizard**.

12. Click on the **Browse** button to locate the recently downloaded certificates.
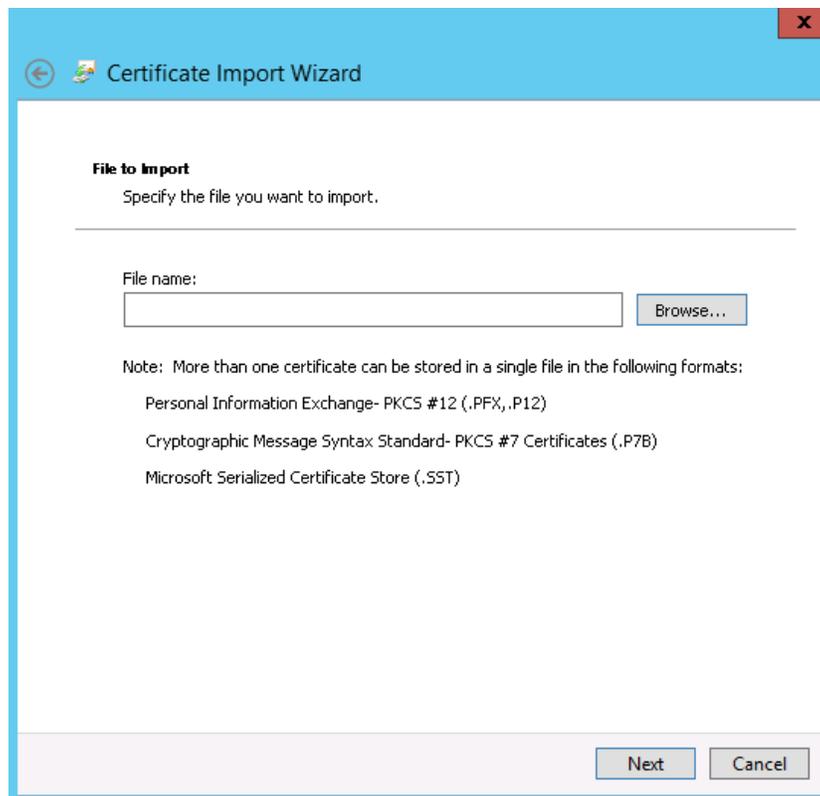
13. Locate the certs folder where you recently extracted it from the download.zip. Select All Files (*.*) file type and then select the first certificate .0 file. Select Open to choose this certificate file.
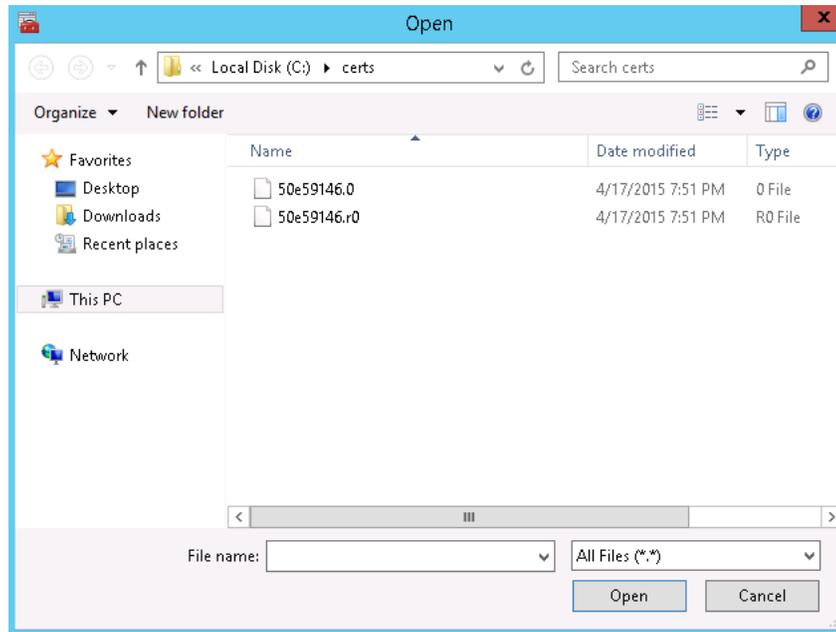
14. Select **Next** to continue.
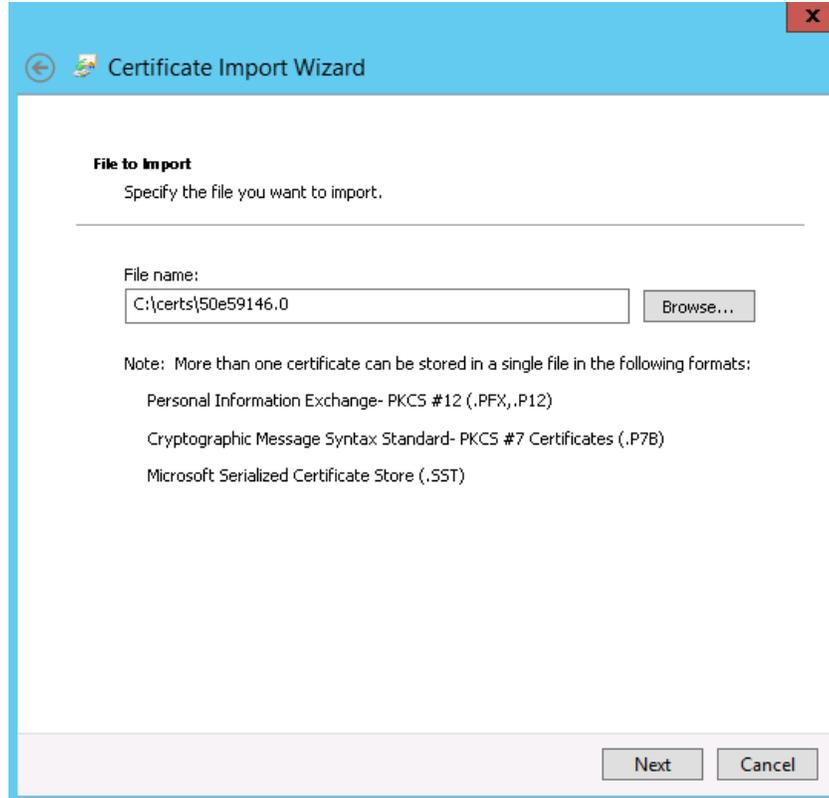
15. Select **Next** to continue with importing the certificate into the **Trusted Root Certification Authorities**.
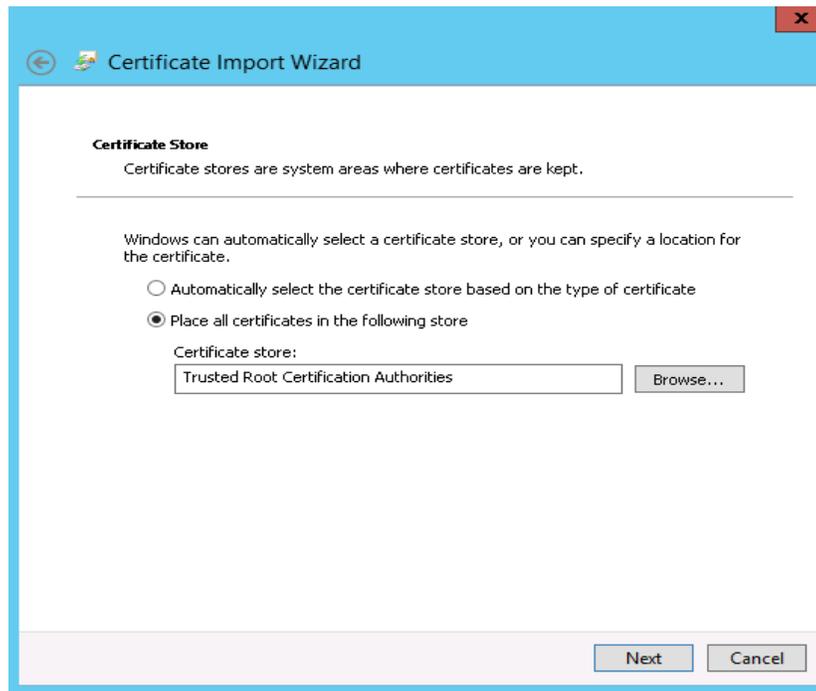
16. Select **Finish** to import the certificate.
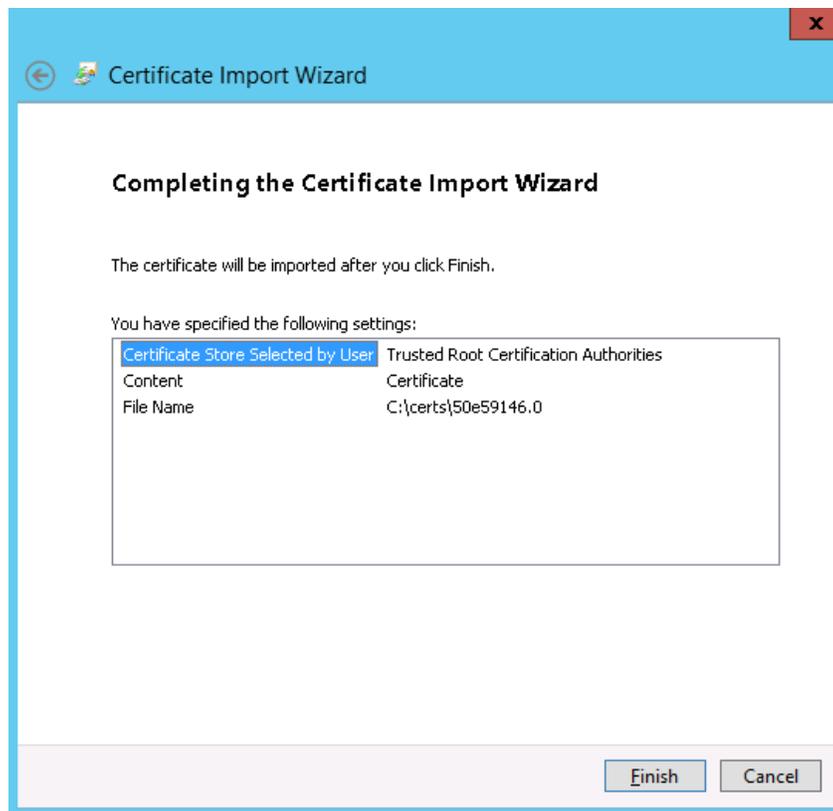
17. Select **OK** to close the notice that the certificate was imported successfully.
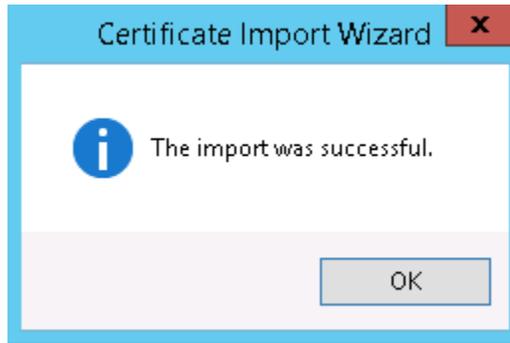
18. Repeat to import process for the vSphere Web Services SDK certificate on the second certificate .r0 file extracted from download.zip.

19. Close the MMC window.

20. Select **No,** to not save console settings.

6.4 USING CITRIX STUDIO TO PERFORM SITE SETUP

Use the following procedure to configure Citrix Studio which will be used to manage the XenDesktop 7.6 deployment:

1. On the **Start Menu** search for and select **Citrix Studio**.
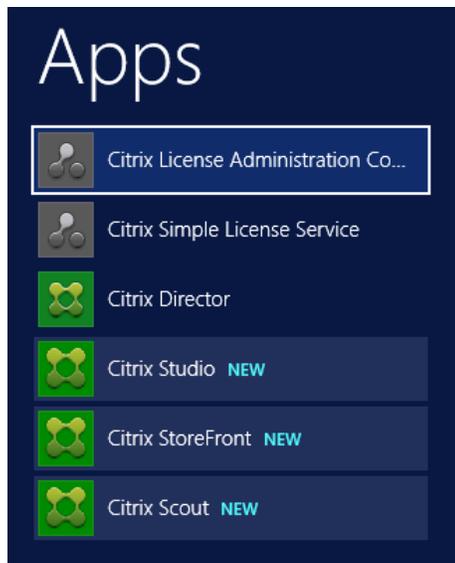
2. Select **Yes** to agree to the User Account Control prompt if displayed.
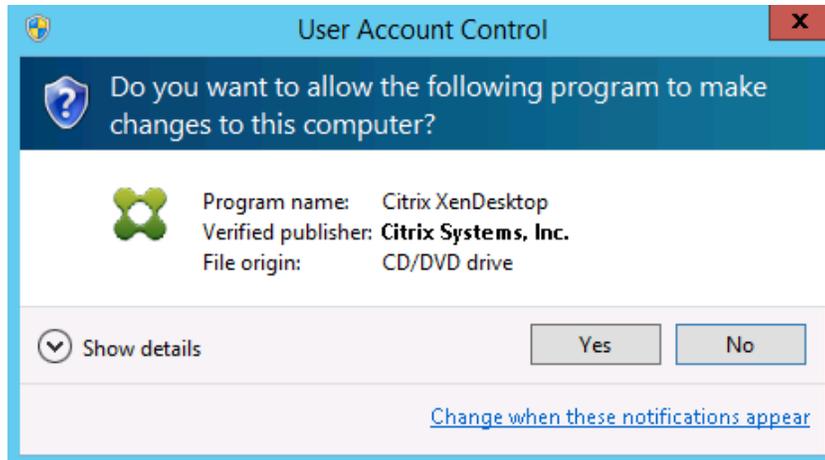
3. Under **Site Setup**, Select Deliver applications and desktops to your users.
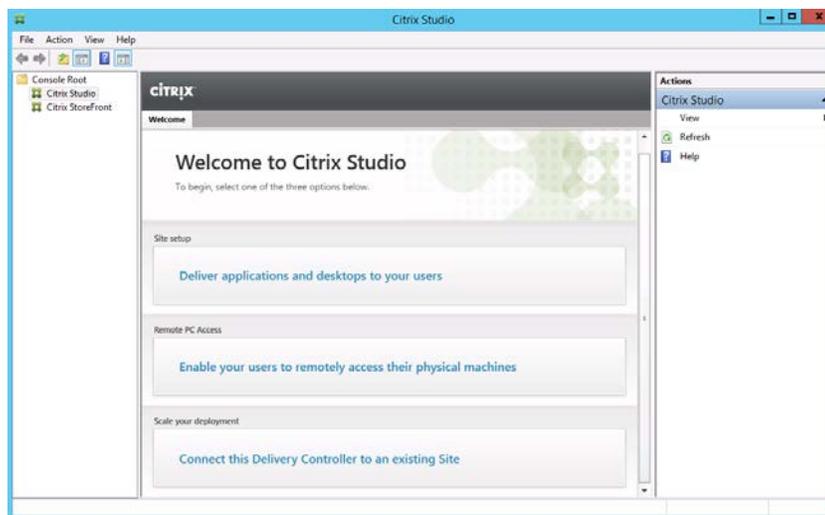
If you have not joined the server to the domain you will not be able to run the Site setup process. The following message will be displayed. If you have received this error, uninstall XenDesktop, join the server to the domain and reinstall XenDesktop before continuing.

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.
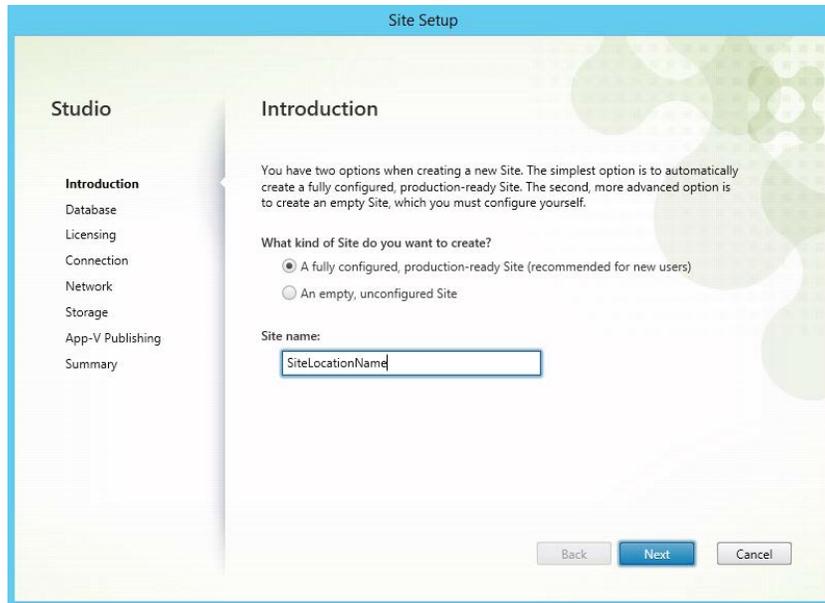
5. For a single server deployment accept the default settings of '.\SQLEXPRESS' and the Database Name as stated. For a multi-server deployment with SQL Server 2008 R2 or newer enter the appropriate Database server location and database name. Select Test Connection to verify the database server connection.
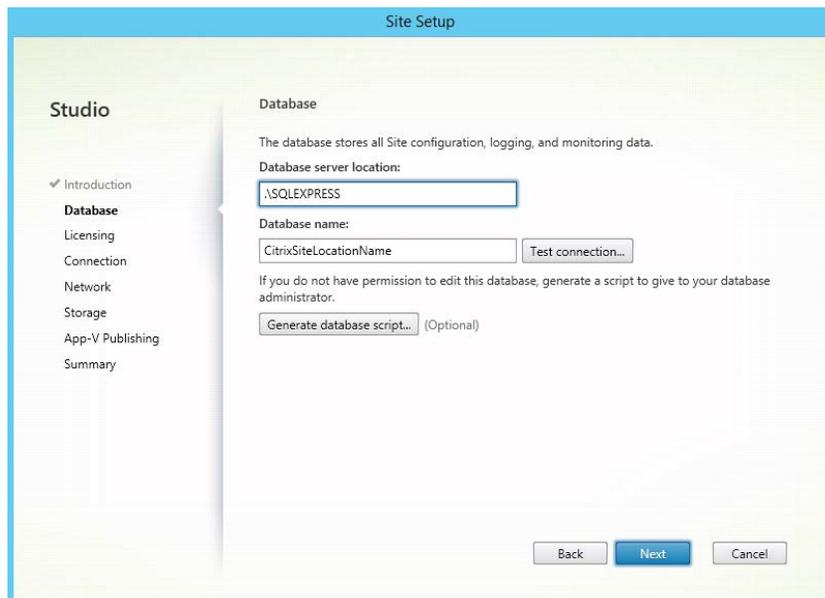
6. If the database was not found on the database server you will be prompted to create the database automatically. Select **Ok** to continue.
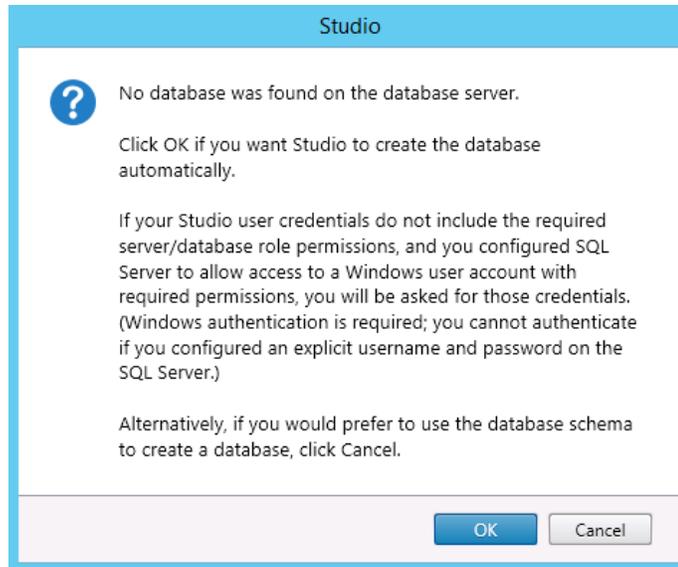
7. A notice will be displayed that the database connection passed. Select **Close** to continue.
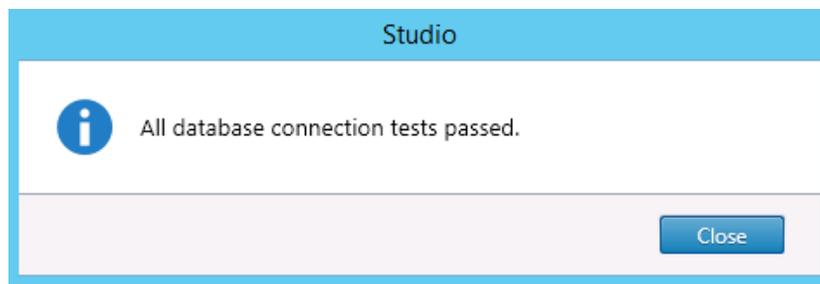
8. Select **Next** to continue.

9. On the **Licensing** dialogue, accept the default settings which will **Use the free 30-day trial**. Select **Next** to continue.
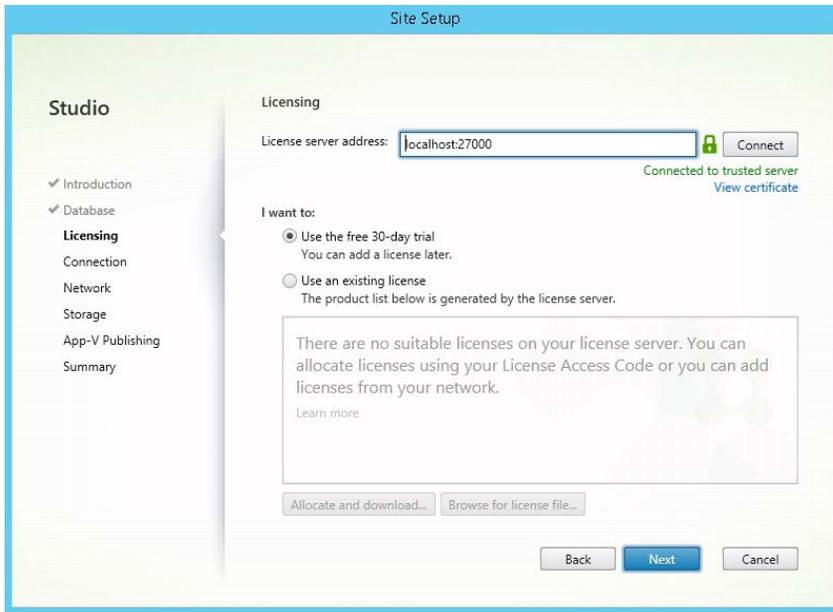
10. Enter the **vCenter Web Services SDK url, User name, password** and a **Connection Name**. Accept the default for **Studio Tools (Machine Creation Services)**. Select **Next** to continue.
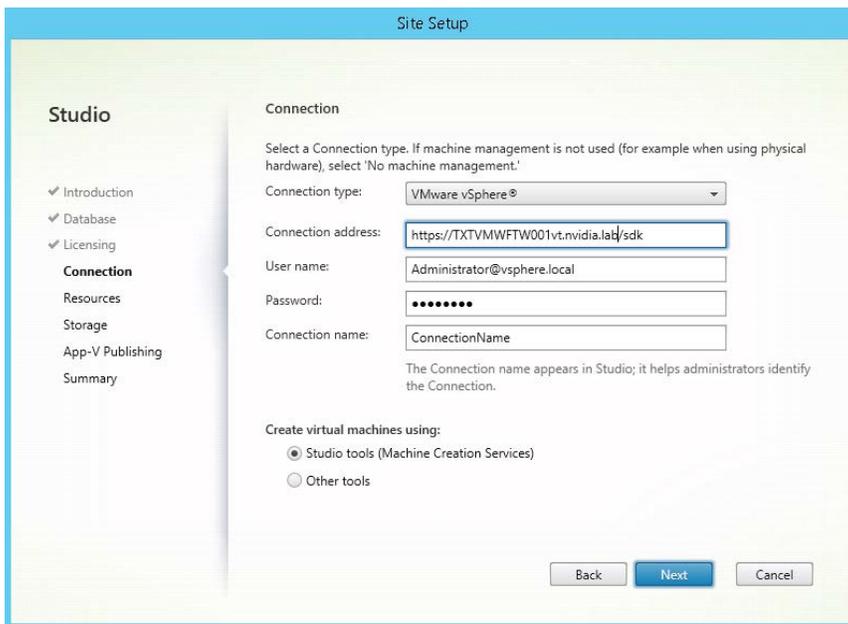
💬 If you receive a certificate error, ensure that you have imported the vSphere Certificate, and vSphere Web Services SDK certificates into the Local Machines Trusted Root Certificate Authorities.
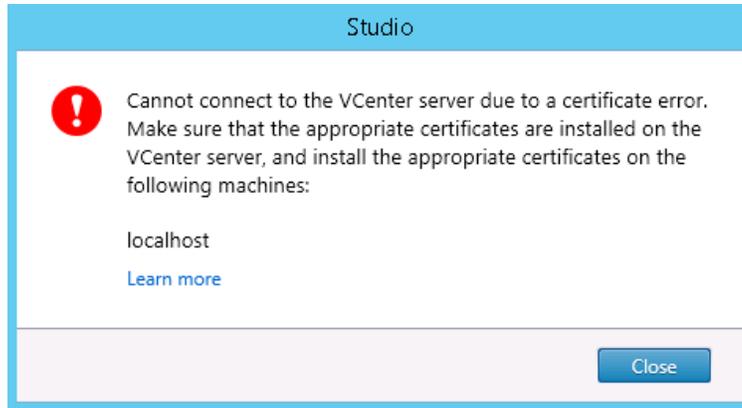
11. Enter a **Resource Name** for the connection to vSphere. Select **Browse** to select a cluster for new virtual machines.
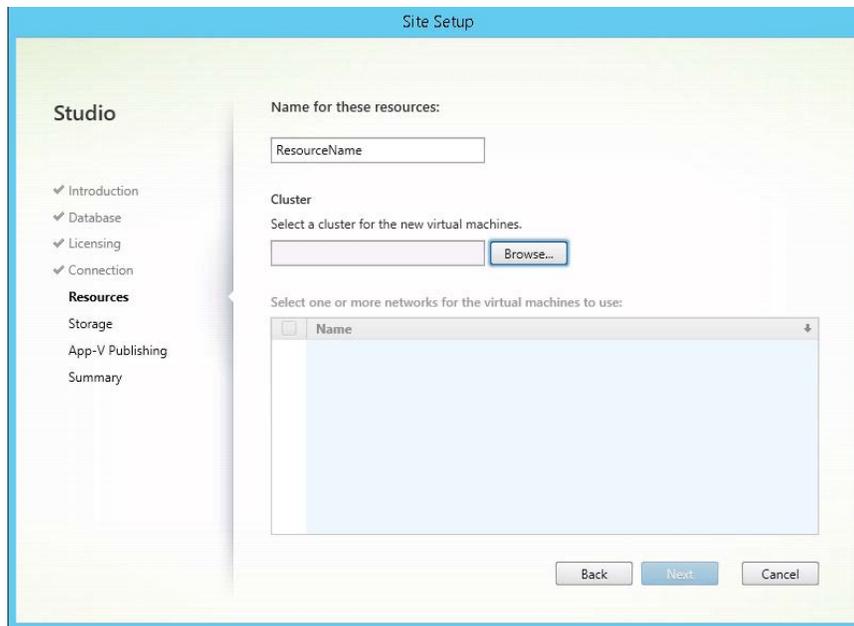
12. Expand and **Select** the cluster you want to provision virtual machines into. Click **OK** to Continue.
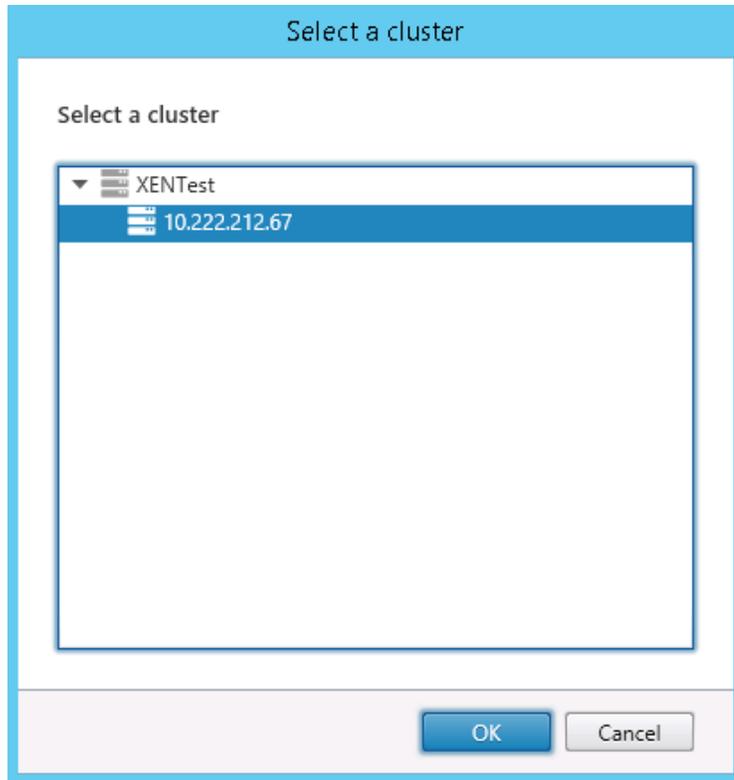
13. Select the Network that you wish the virtual machines to use. Select Next to continue.
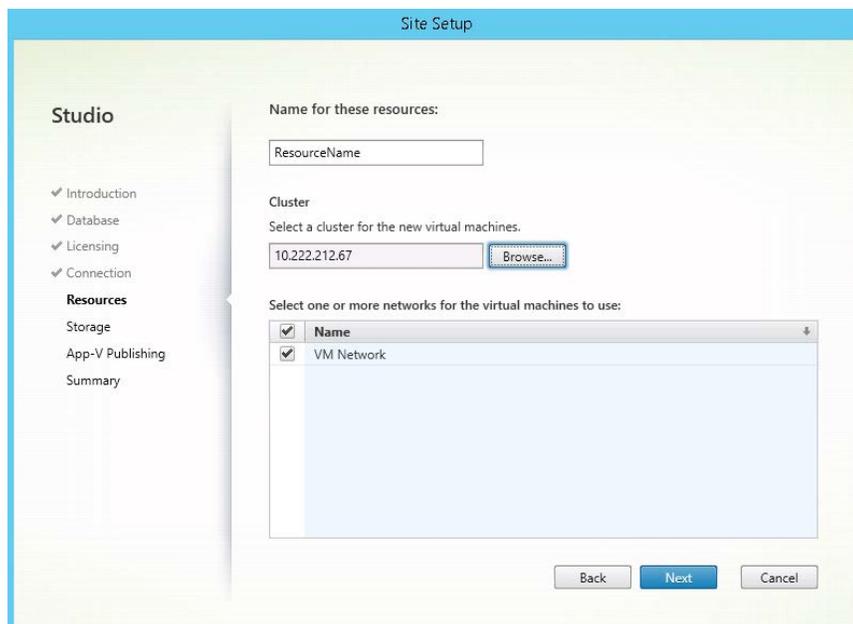
14. Select a **Datastore** in which to store the virtual machine files and any personal vDisks. Select **Next** to continue.
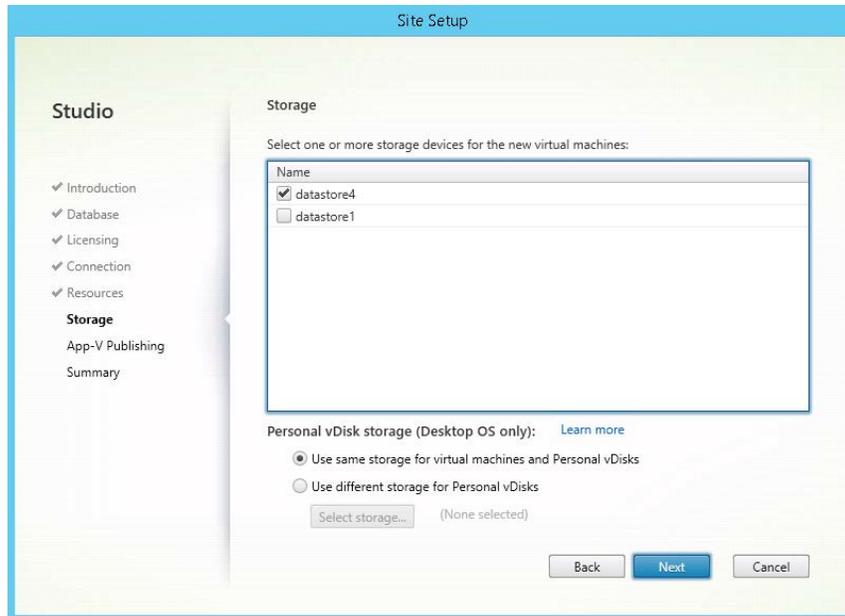
15. Accept the default of **No** radio button on the App-V Publishing dialogue. Select **Next** to continue.
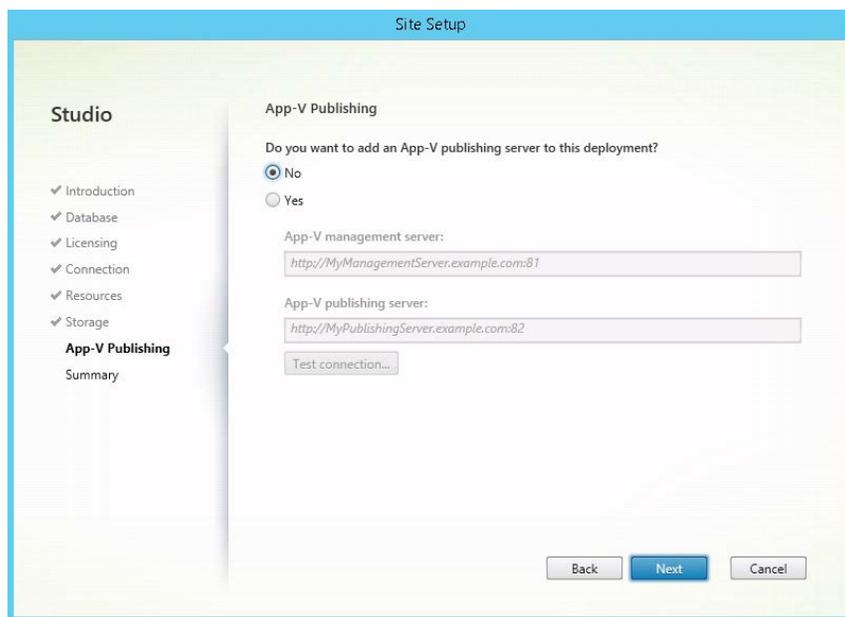
16. Select **Finish** to complete the Site Setup.
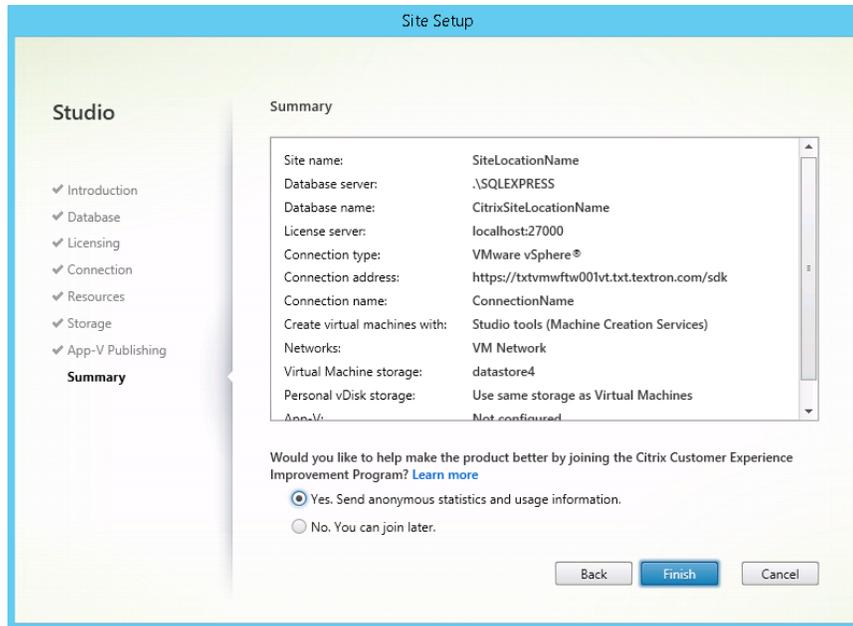
17. Select **Test Site Configuration** to verify everything is configured properly.
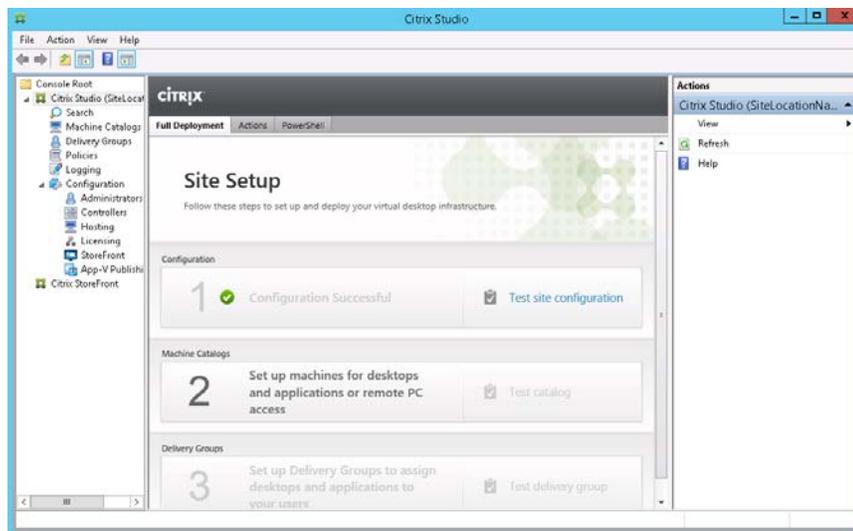
18. Select **Show Report** to view any warnings or failed tests. Select **Close** to exit the **Site Configuration Testing** dialogue.
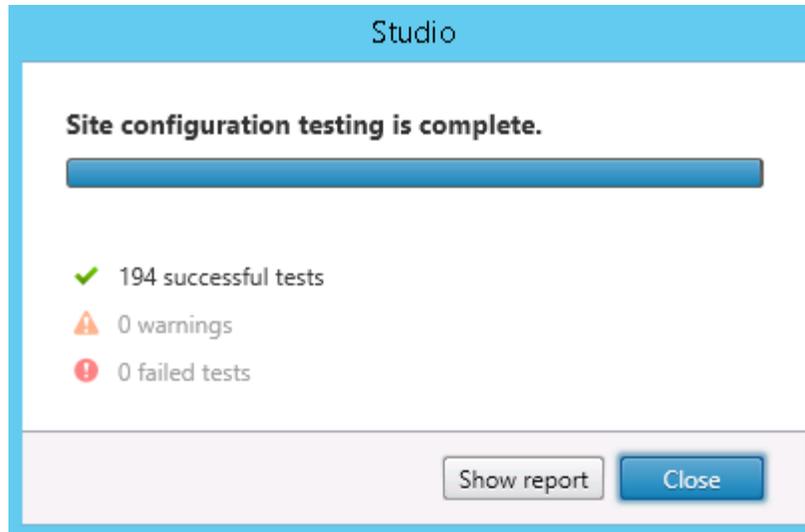
6.5 CONFIGURING CITRIX LICENSES SERVER

Use the following procedure to configure the Citrix XenDesktop 7.6 Enterprise license:

19. 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.
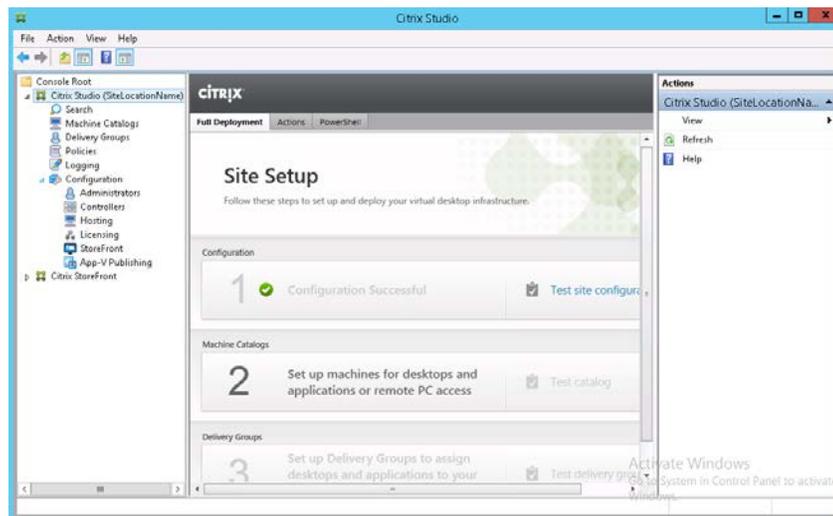
20. Select **Add Licenses** from the **Action** pane.
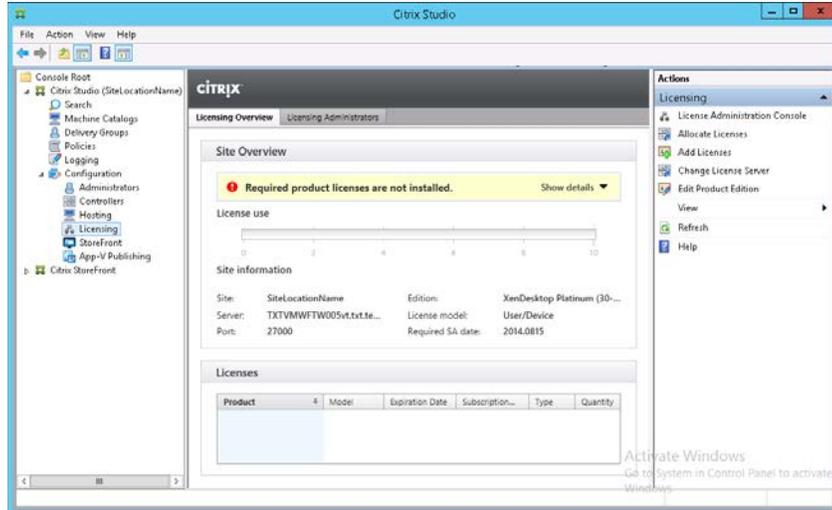
21. Browse the file system on the XenDesktop server to where you previously saved the .lic file downloaded from you're Citrix My Account. Select the **file**, and then click **Open** button to continue.
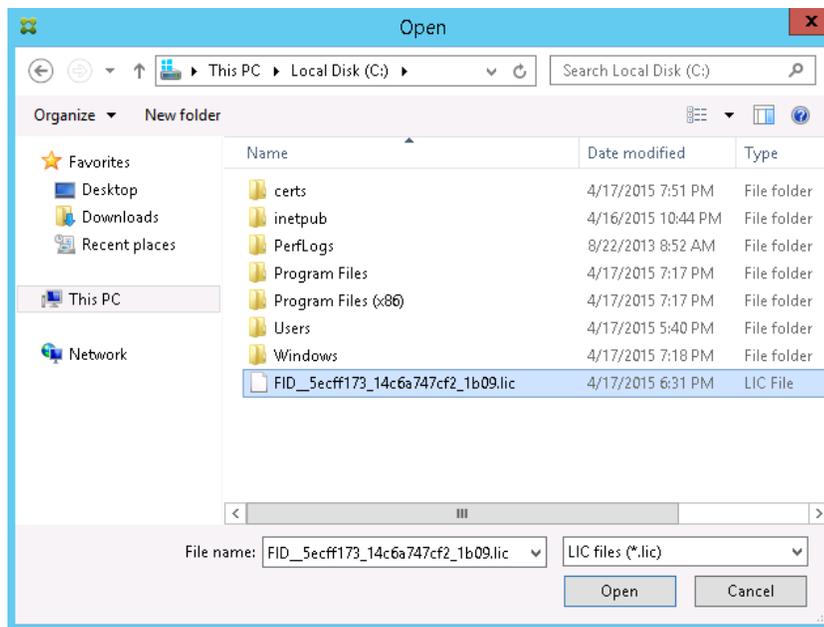
22. Select **Yes** to confirm to add the license file selected.

23. The license will be displayed in the center pane of the license screen.
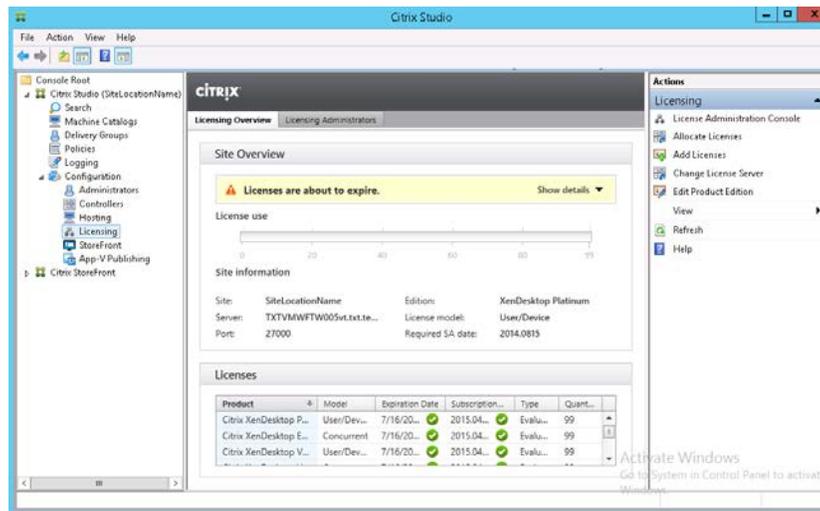
6.6 VERIFY CITRIX STOREFRONT CONFIGURATION

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** dialogue.
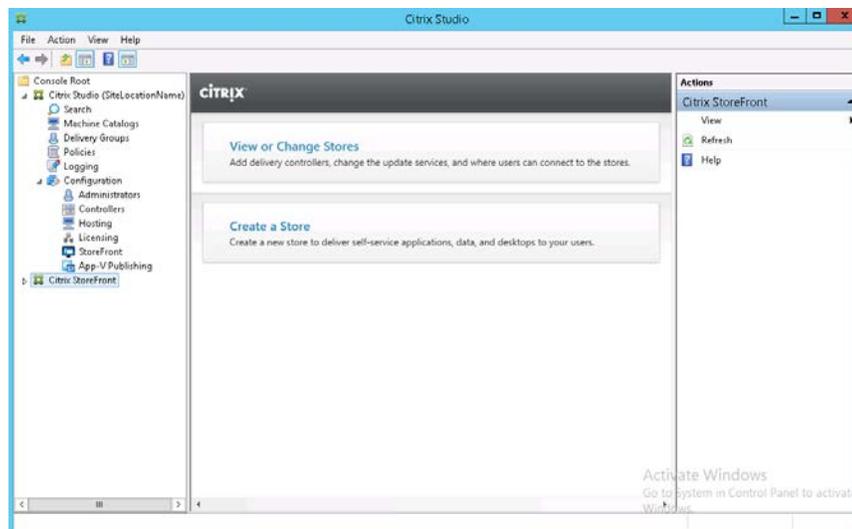
2. The center pane should display information about the currently configured Store. Note that the status of the StoreFront is using HTTP not HTTPS.
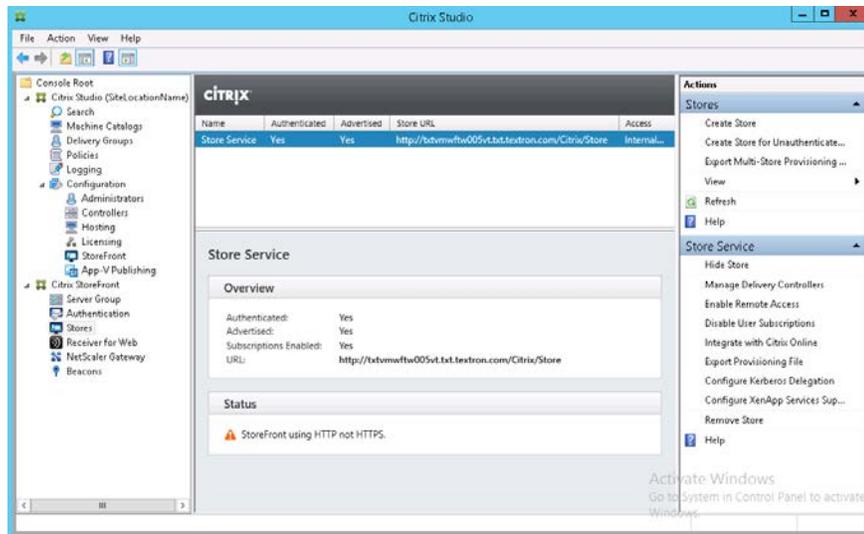
3. Select **Receiver for Web** from the left pane. Make note of the **Website URL** that is displayed on the center pane. This is the **Citrix Web Receiver** url and will be used to login to XenDesktop from a web browser.
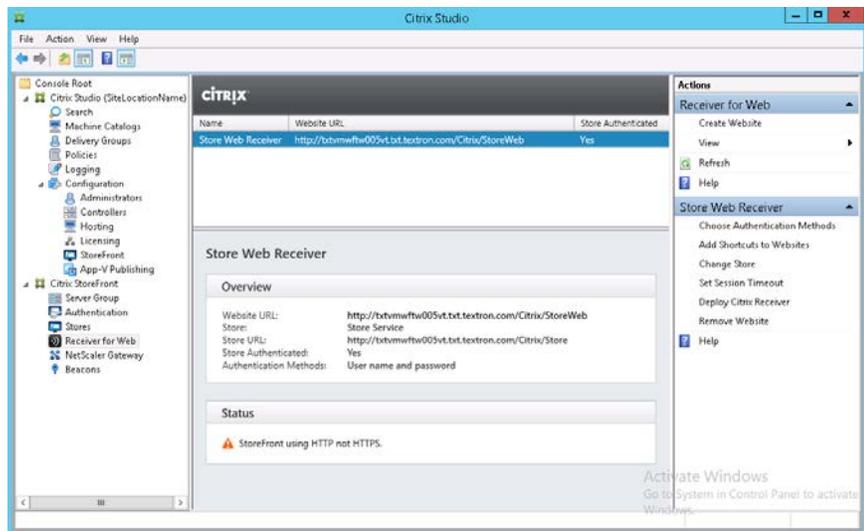
4. This completes the initial setup of XenDesktop.

Chapter 7.
BUILDING A VGPU VSPHERE HOST CLUSTER

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 vGPU-enabled desktops across clustered vSphere hosts with NVIDIA GPUs.

7.1 ADDING VSPHERE HOSTS TO MAKE A CLUSTER

To make a cluster you first create a Datacenter, then a cluster inside that cluster, and finally add hosts to the cluster. Then the hosts can be managed in a group.

7.1.1 Creating a DataCenter in vCenter

1. Log in to vCenter Server's vSphere Web Client and click on **Home**.
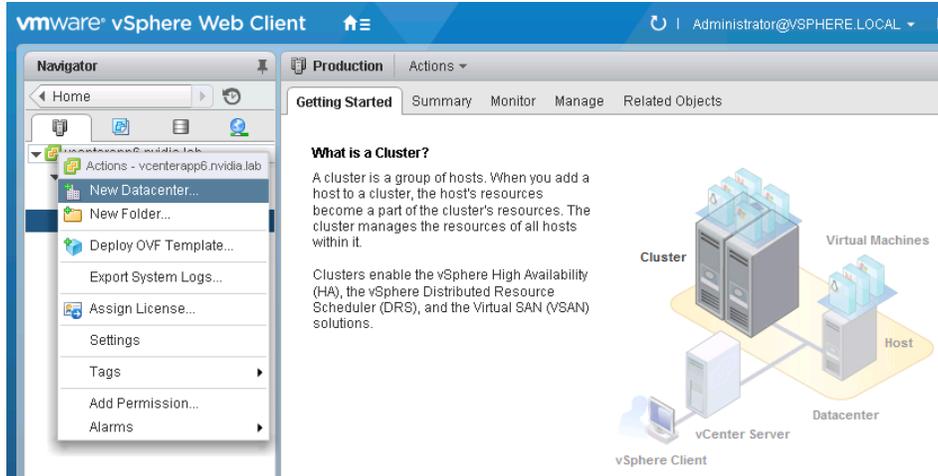
2. On the resulting pop-up, enter a name for the new Datacenter, and click OK:

7.1.2   Creating a Cluster in the Datacenter

1. Right click on the Datacenter object, then on the menu that pops up click New Cluster…
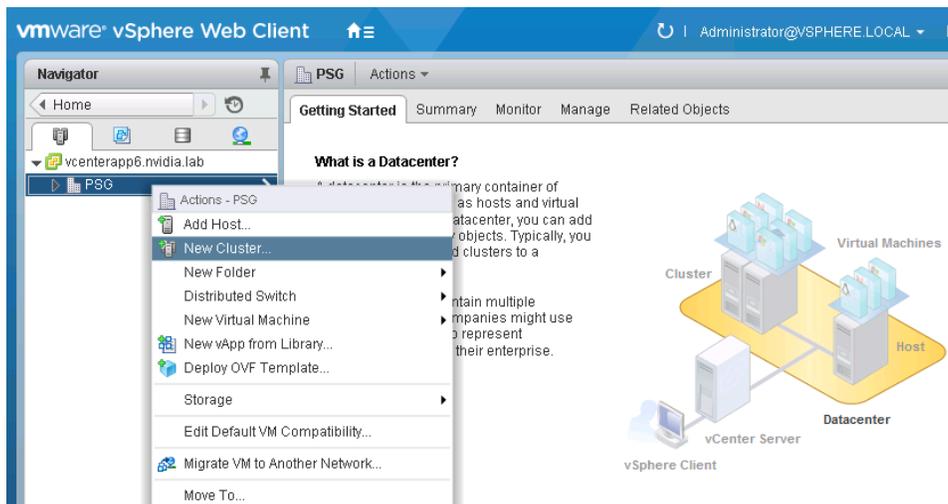
2. A pop-up appears, enter the name of the new Cluster and click OK:
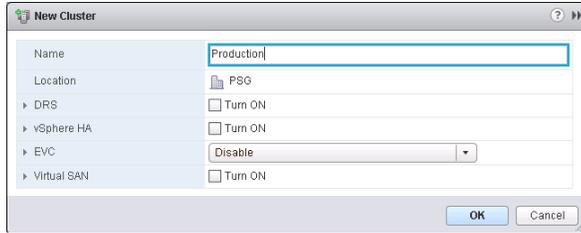
3. You should now have a logical organizational structure of at least a cluster in a datacenter. In our lab we used two clusters in a datacenter, one for management, the other for production VDI.

7.2 ADDING VGPU HOSTS TO A CLUSTER

To create a VGPU Cluster, just add VGPU enabled hosts to your virtual desktop cluster. This cluster must only contain hosts that have NVIDIA GPUs that support NVIDIA GRID VGPU. This is required in order to support better distribution of resources for VGPU.

This chapter covers VGPU profiles and considers in choosing them for your needs including:

- ▶ List of available VGPU profiles
- ▶ Matching profiles to user needs

8.1 THE ROLE OF THE VGPU MANAGER

vGPU profiles assign custom amounts of dedicated graphics memory. NVIDIA GPU Manager assigns the correct amount of memory to meet the specific needs of each user. Every virtual machine has dedicated graphics memory (just like any desktop or laptop device), thus ensuring that it has the resources needed to handle the expected graphics load.

NVIDIA vGPU Manager allows up to eight users to share each physical GPU by assigning the graphics resources of the available GPUs to virtual machines using a balanced approach. Each NVIDIA GRID K1 card has four GPUs, allowing up to 32 users to share a single card, a GRID K2 has two GPUs allowing up to 16 users per card.

8.2 THE FULL LIST OF VGPU PROFILES

The profiles represent a line up of virtual GPUs, varying in size of frame buffer memory and number of heads. The division of frame buffer is what defines the number of users possible per 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 vGPU Graphics Memory Options

NVIDIA GRID Board	Virtual GPU Profile	Application Certifications	Graphics Memory	Max Display per User	Max Resolution per Display	Max Users per Board	Use Case
GRID K2	K280Q	YES	4,096 MB	4	2560x1600	2	Designer
	K260Q	YES	2,048 MB	4	2560x1600	4	Designer/Power User
	K240Q	YES	1,024 MB	2	2560x1600	8	Designer/Power User
	K220Q	YES	512 MB	2	2560x1600	16	Power User
GRID K1	K180Q	YES	4,096 MB	4	2560x1600	4	Power User
	K160Q	YES	2,048 MB	4	2560x1600	8	Power User
	K140Q	YES	1,024 MB	2	2560x1600	16	Knowledge Worker
	K120Q	YES	512 MB	2	2560x1600	32	Knowledge Worker

8.3 WHY THE "Q" ON THE PROFILE?

Like the NVIDIA Quadro cards for professional graphics applications, the vGPU profiles ending in **Q** undergo the same rigorous application certification process as those -class processors. 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. For a list of certified applications, visit www.nvidia.com/gridcertifications.

8.4 MATCHING PROFILES TO USER NEEDS

As stated earlier in the deployment guide, you need to define your user needs and match them to available NVIDIA GRID VGPU graphics card profiles.

If we take the three example Revit user use cases and their computing needs defined earlier, 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. See below:

▶ Entry Level Engineer / Design Reviewer
  - 4 GB RAM
  - 4 VCPUs (2.5 GHz)
  - 1 GB video RAM – K240Q or K140Q
  - 2 2560x1600 displays

▶ Mid Level Engineer
  - 8 GB RAM
  - 4-8 VCPUs (2.5 GHz)
  - 2 GB video RAM – K260Q or K160Q
  - Up to 4 2560x1600 displays

▶ Advanced Engineer
  - 16 GB RAM
  - 8 VCPUs (2.5 GHz)
  - 2-4 GB video RAM – K260Q - K280Q
  - Up to 4 2560x1600 displays

8.5 POWER USER & DESIGNER PROFILES

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

Selecting vGPU Profiles

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 multiple workstation users can be placed on a single host, even combined with other less demanding users to increase density. Suddenly there were choices when designing a solution.

The key to picking a profile for these users is compatibility and performance. Leveraging the NVIDIA platform, built of 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.

If your users leverage physical GPUs in the range of a K600, then adding K1 cards and then adjusting profiles based on frame buffer requirements allows you the flexibility and density to meet the performance needs of these user groups. If your application demands the highest performance then the K2 with its K5000 equivalent GPU allows you to adjust frame buffer without sacrificing that performance.

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

8.6 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 across the vast majority of end users now leverage graphics to one degree or another. These include the underlying operating system's desktop, such as Windows 7 with Aero enabled, and Windows 8.1 takes that to a higher level. 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 latter, and specifically Microsoft Office 2013, has a check box to enable its ability to leverage graphics hardware (GPU) acceleration.

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 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.

8.7 K280Q & K180Q PROFILES VS VDGA

With the addition of these two profiles, each assigned the full 4GB of Frame Buffer memory assigned to a GPU, you no longer have to pull valuable GPUs out and manage them separately as assigned with vDGA. Now you can manage all your user groups, from high density Knowledge Workers needing the rich and responsive experience of 512MB of frame buffer and access to a shared K600 equivalent GPU (K120Q), to the most demanding of workstation end users needing 4GB of frame buffer and access to a full K5000 equivalent GPU (K280Q). When needs change, just shut down your virtual desktop and assign a new profile, which is far more simple than having to shut down a physical machine, change the PCI card, reboot a physical machine.

8.8 LEGACY PROFILES

The GRID K100 and K200 GPU profiles were originally designed for the lighter graphics workloads associated with the applications that knowledge workers and power users use most of their time. They were based on what was typically seen for GPU and frame buffer at the time. As the graphics needs of even basic apps continues to increase, plus the benefit of a Quadro certified driver, the K120Q and K220Q profiles have replaced the 100 and 200 profiles.

Chapter 9.
APPLICATION BASED VDI HOST SIZING EXAMPLE – ESRI ARCGIS

9.1 HOW MANY USERS CAN I GET ON A HOST?

Using one application, ESRI ArcGIS Pro, a typical dialogue with a customer often sounds like this…

Customer: How many users can I get on a host?

NVIDIA: What is their primary application?

Customer: ESRI ArcGIS Pro.

NVIDIA: Are they primarily 3D or 2D data users?

Customer: 3D mostly.

NVIDIA: Would you describe them as light, medium, or heavy users?

Customer: Medium to heavy.

NVIDIA: Power users to designers then.

Customer: I need performance AND scalability numbers that I can use to justify the project…

9.2 AN EXAMPLE APPLICATION: ESRI ARCGIS PRO

ArcGIS Pro is a premier application for GIS mapping visualizing, editing, and analysis. It requires a GPU, and creates heavy CPU load that needs to be considered in architecting your vGPU solution. The size of your map data, the concurrency of your users, and the level of interaction with 3D data need to be factored into defining your user groups. ESRI classifies its users as follows:

User Classification Matrix						
NVIDIA User Classifications	Knowledge Workers		Power User		Designer	
ESRI User Classifications	Light 2D	Medium 2D	Heavy 2D	Light 3D	Medium 3D	Heavy 3D

Given they are the most graphics intensive users, we focused our tests on the designer and power user groups. Based on our findings, NVIDIA GRID provides the following performance and scalability:

9.3 APPLICATION BASED SIZING

This section contains an overview of the testing, the methodology, and the results that support the findings in this deployment guide. We also detail the lab environment used to test the contents of this guide.

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

9.4 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.

9.4.1 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

For the test 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.

9.5 UX – THE VDI USER EXPERIENCE

To define user experience (UX) requires defining elements of application and user interaction. This can be obvious like the rendering time for an image to appear, or to smoothly pan across that image, or more subtle 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 is much harder to measure.

Users also 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.

9.5.1 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 metrics. This gave us a single value to track, and then validate the rest of the results were within acceptable ranges as well.

9.5.2 Real life experience verses 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.

9.5.3 The importance of eyes on!

Its 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.

9.6 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 testing 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.

- ▶ 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.
- ▶ Rendering vs Analytics:

▶ 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. In general we run 1, then 8, then 16, to get a baseline of results and accompanying logs (CPU, GPU, RAM, networking, storage IOPS, etc.).

9.7 RESULTS OVERVIEW

▶ **vCPU:**

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

        o *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.

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

▶ **GPU**

ESRI states that a GPU is required for 3D workloads, we ran tests with servers hosting either pairs of K1 or K2 cards. A test with the full GPU and all 4GB of frame buffer, the K280Q profile, offered high performance but lacked scalability, limited to 4 virtual desktops in our two K2 host. We then tested with the K260Q profile and its 2GB of frame buffer, up to its maximum of 8 users (4/K2 card) it was still within the metrics provided by ESRI. This meant there was room for more users if we lowered frame buffer to 1GB via the K240Q profile. We also tried the K220Q profile and its 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, the K240Q 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.

9.8 RESULTS

The following are the full results of our testing. Beginning with a graph showing the region of acceptability as a green bar, times above this region are a worsening UX, below would be a better UX. Looking for both performance and scalability, as the line for the profile moves above the region of acceptability this becomes the greatest scalability while still within performance expectations. Its important to note that your users, your data, and your hardware, will impact these results and you may decide a different level of performance or scalability is required to meet your individual business needs.
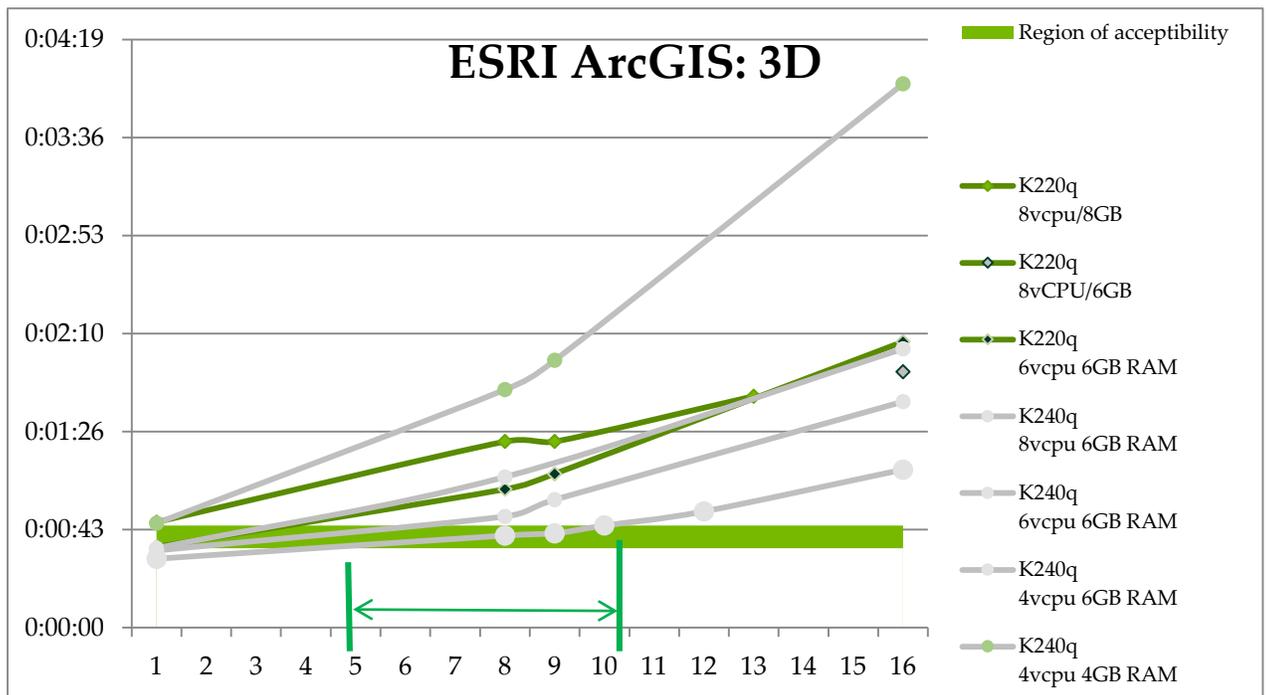
Detailed results:

VM Config	VM count	DrawTime (min:sec)	FPS	Min FPS	Std deviation
K220q 8vcpu/8GB	1	00:46.5	52.25		
	8	1:22	39.33		
	9	01:45.1	16.07	3.12	12.3
	13	1:42	30.87		
K220q 8vCPU/6GB	16	01:52.8	31.45	3.2	17.9
K220q 6vcpu 6GB RAM	1	00:35.6	52.25	12.25	
	8	1:01	34.62	5.8	7.2
	9	01:07.8	31.87	4.87	13.32
	16	2:06	20.77	3.34	23.45
K240q 8vcpu 6GB RAM	1	0:00:35	59.95	13.86	
	8	0:01:06	53.34	8.09	21.99
	16	0:02:03	47.79	3.8	4.74
K240q 6vcpu 6GB RAM	1	34.41	60.03		
	8	00:49.0	51.48	7.3	3.6
	9	00:56.4	50.77	6.8	3.48
	16	01:39.6	46.60	4.52	6.27
K240q 4vcpu 6GB RAM	1	00:30.3	61.84	17.75	
	8	00:40.6	49.76	11.45	1.45
	9	00:41.6	46.87	10.21	3.2
	10	**00:45.0**	**45.76**	**9.46**	**2.62**
	12	00:51.2	40.31	7.54	6.3
	16	01:09.6	37.52	5.3	2.7
K240q 4vcpu 4GB RAM	1	0:00:46	59.94	16.7	
	8	0:01:45	52.67	14.51	8.67
	9	0:01:58	51.39	15.4	9.23
	16	0:04:00	45.84	15.48	25.8
K240q 4vcpu 6GB RAM Think Time 5 Seconds	1	00:30.2	60.02	24.81	
	8	00:34.0	57.05	14.15	0.79
	12	00:38.1	52.87	12.00	1.2
	16	00:46.0	49.55	9.85	2.8
K240q 4vcpu 6GB RAM Think Time 5 Seconds Navigation Time 5 Second (Panning and Zooming)	1	01:11.2	62.34	21.96	
	8	01:15.9	53.45	14.15	1.9
	12	01:22.6	45.35	8.65	2.9
	16	01:32.8	40.65	6.00	3.7

Chapter 10.
CREATING YOUR FIRST VGPU VIRTUAL DESKTOP

This chapter describes how to:

▶ Create and configure a virtual machine in vSphere

▶ Install Windows and VMware Tools on the VM

▶ Customize Windows settings

▶ Install Citrix Virtual Desktop Agent on the VM

▶ Adding the Gold Master image to the domain

▶ Adjust some additional VM settings and enable VM console access

▶ Enable the NVIDIA vGPU and finalizing the installation

10.1   CREATING A VIRTUAL MACHINE

These instructions are to assist in making a VM from scratch that will support NVidia vGPU. Later the VM will be used as a gold master image. Use the following procedure to configure a vGPU for a single guest desktop:

1. Browse to the host or cluster using the *vSphere Web Client*.
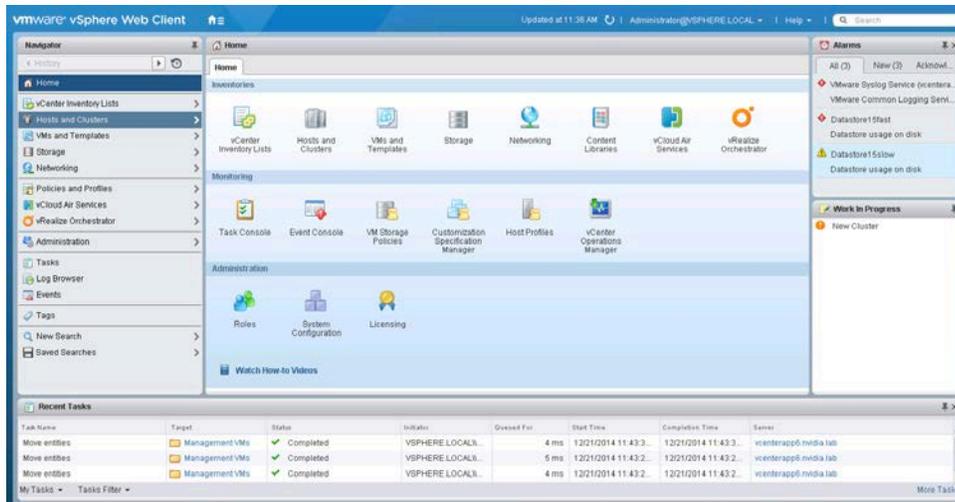
2. Right-click the desired host or cluster and select **New Virtual Machine**.

The *New Virtual Machine* wizard begins.
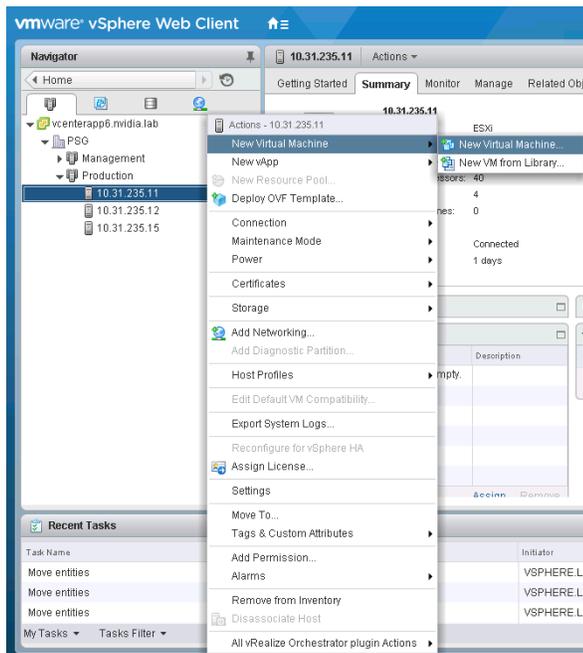
3. Select **Create a new virtual machine** and click **Next**.
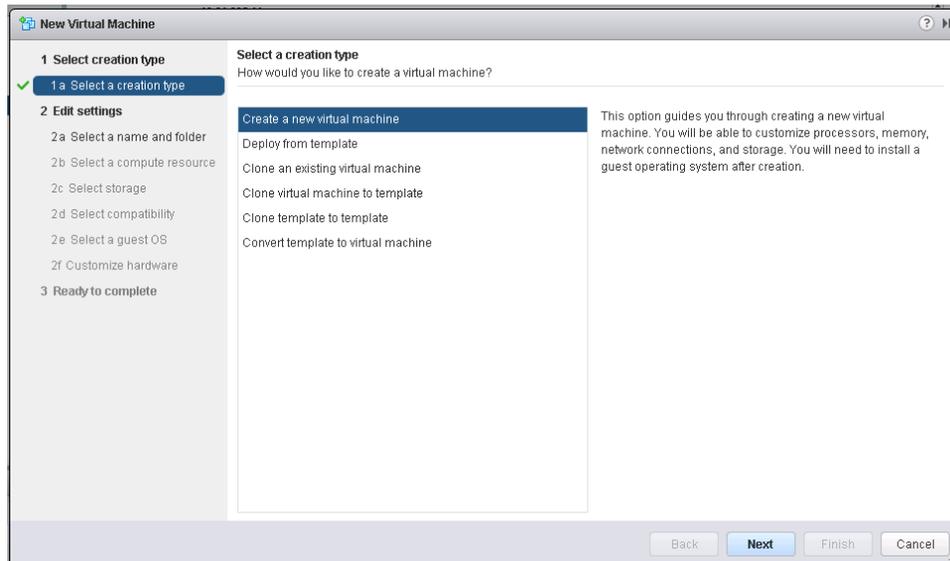
4. Enter a name for the virtual machine. Choose the location to host the virtual machine using the **Select a location for the virtual machine** section. Click Next to continue.
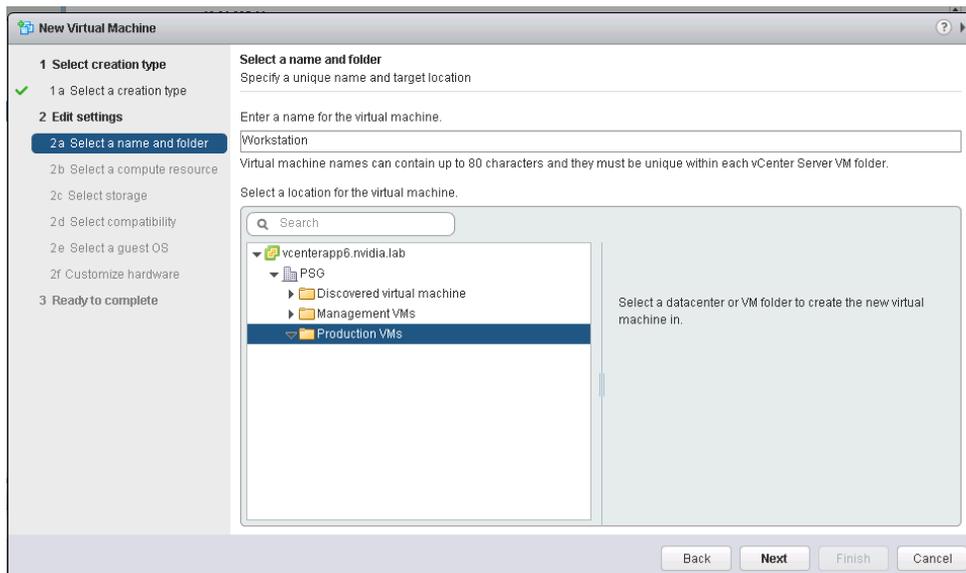
5. Select a compute resource to run the VM. Click Next to continue. NOTE: This compute resource should include an NVIDIA GRID K1 or K2 card installed and be correctly configured.
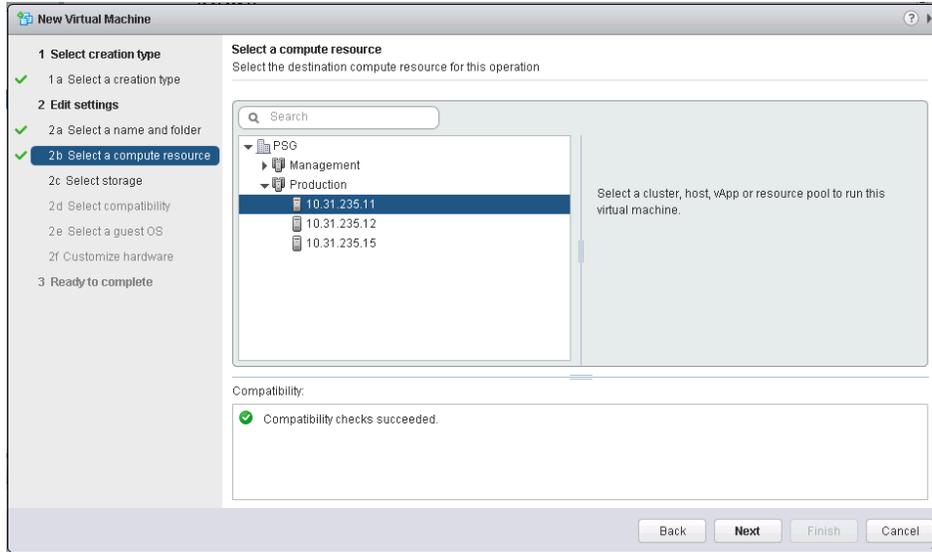
6. Select the datastore to host the virtual machine. Click **Next** to continue.
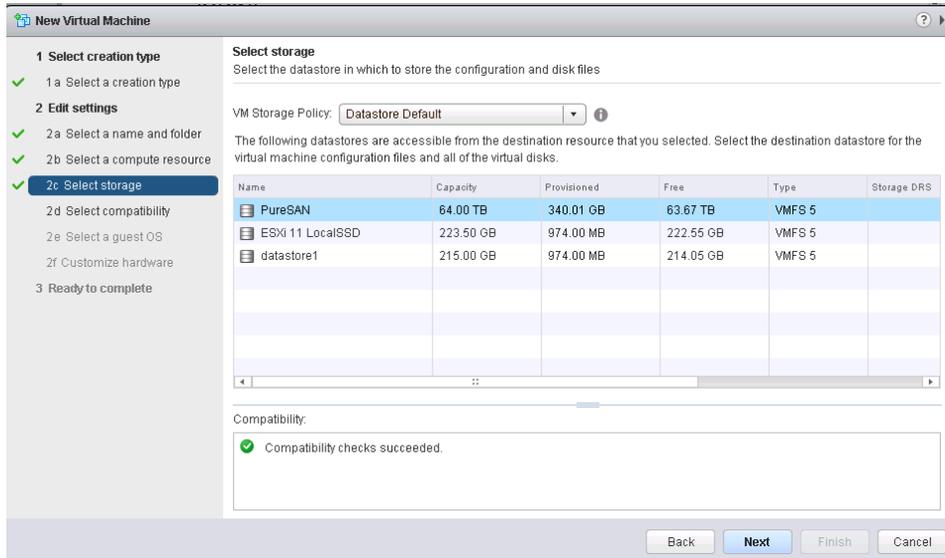
7. Select compatibility for the virtual machine. This allows VMs to run on different versions of vSphere. In order to run vGPU select ESXi 6.0 and later. Click **Next** to continue.
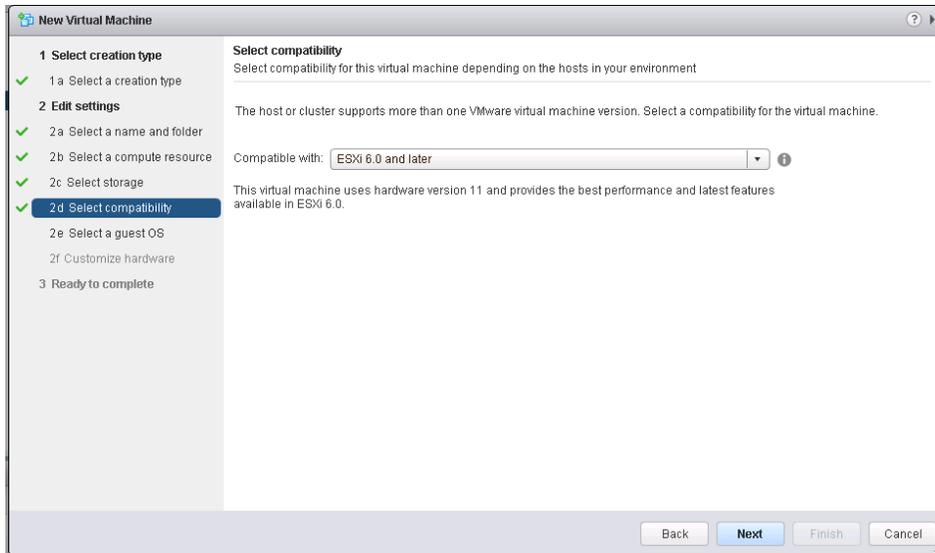
8. Select the appropriate Windows OS from the **Guest OS Family** and **Guest OS Version** pull-down menus. Click Next to continue.
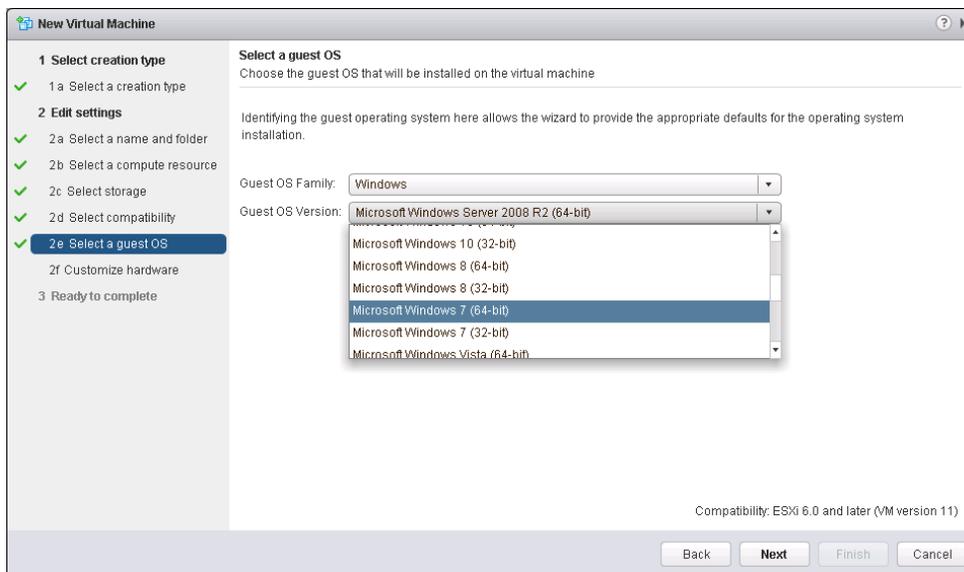
9. Customize hardware is next. Set the virtual hardware based on your desktop workload requirements. Click Next to continue.
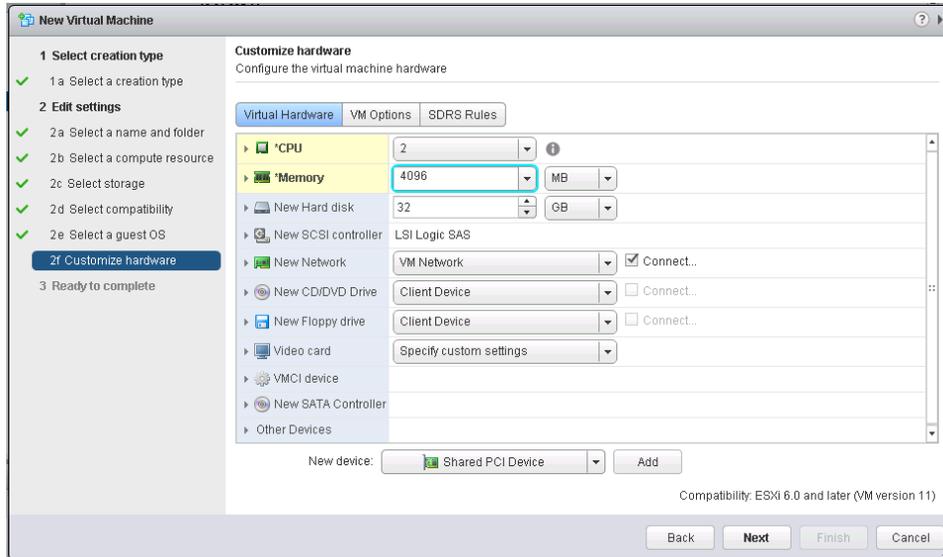
10. Review the New Virtual Machine configuration prior to completion. Click **Finish** when ready.
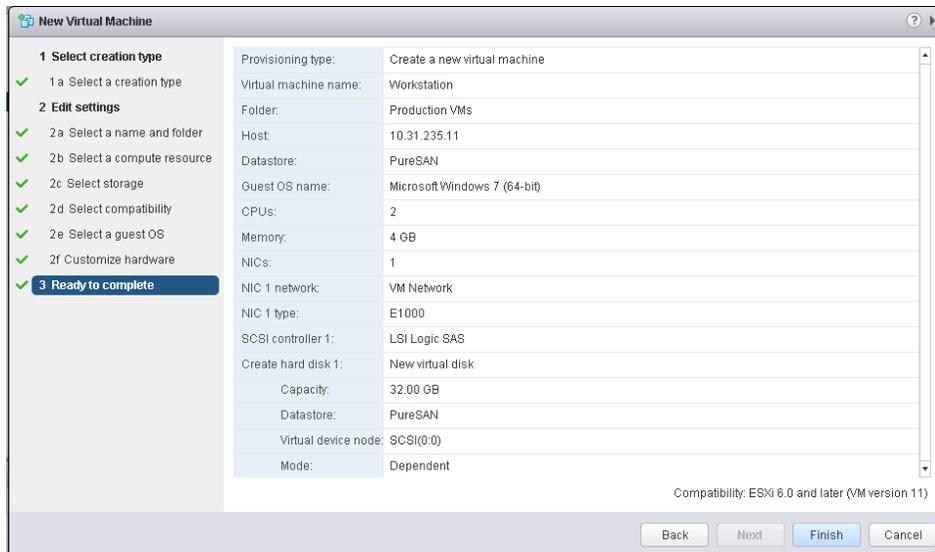
The new virtual machine container has now been created.
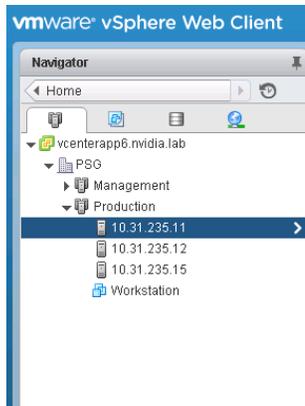
Installation of the guest OS inside the VM is next.

10.2   INSTALLING WINDOWS

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.
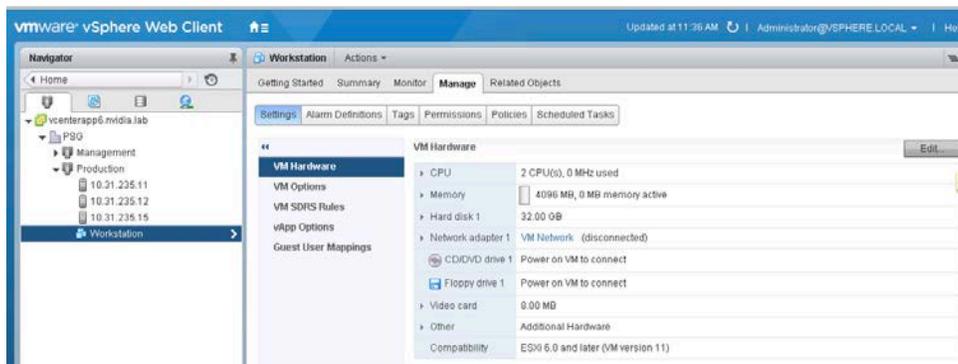
2. Locate the CD/DVD entry under the Virtual Hardware tab. Select the arrow drop down to reveal data sources for the CD/DVD media. (In this example a Datastore ISO file will be used.) Check the Connect checkbox for CD/DVD drive 1. This will connect the ISO file to the VMs virtual CD/DVD drive.
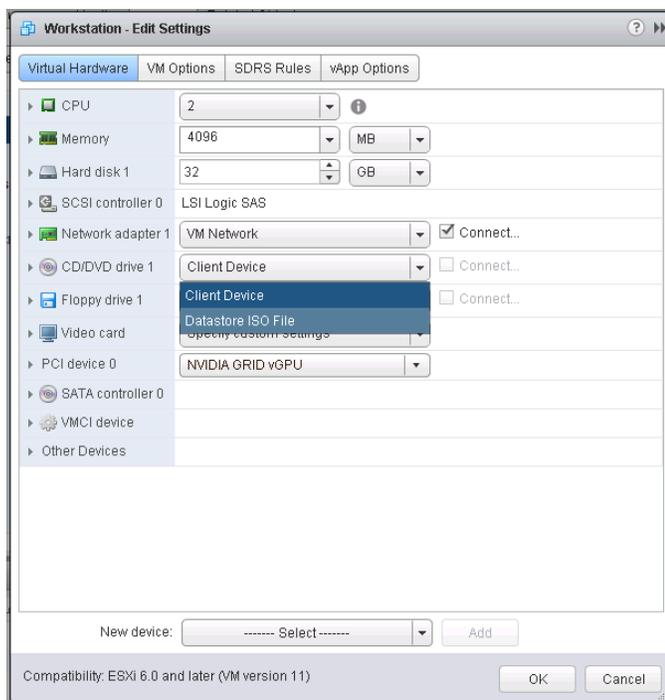
3. Toggle the carrot next to the CD/DVD drive 1 icon to reveal the details of the virtual device. For Status check the **Connect At Power On** checkbox. This will connect the ISO file to the VM's virtual CD/DVD drive during boot up. Next Click on the Browse button for the CD/DVD Media.
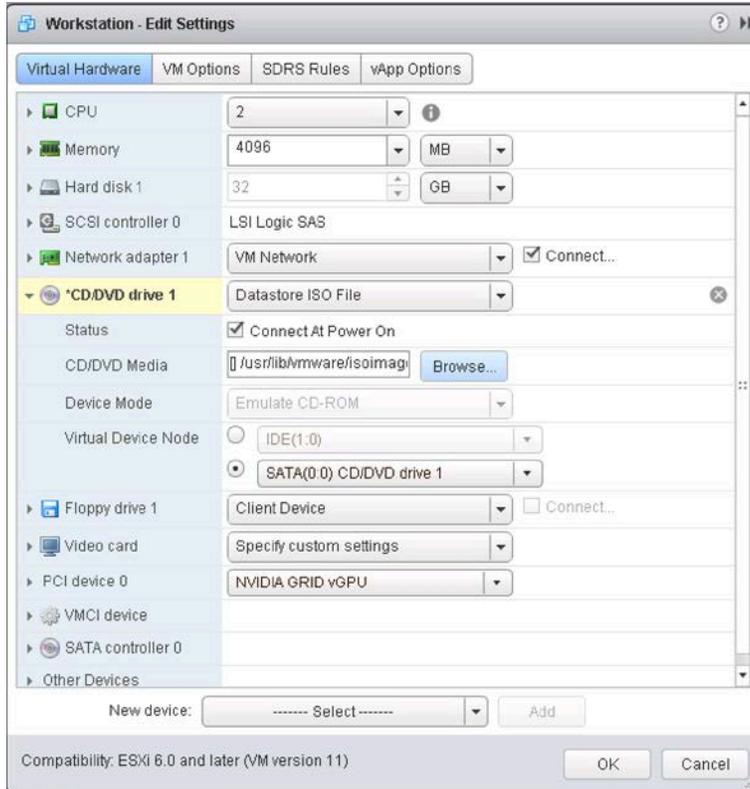
4. Navigate to and select the OS ISO file to be used for install. Click OK to select the ISO.
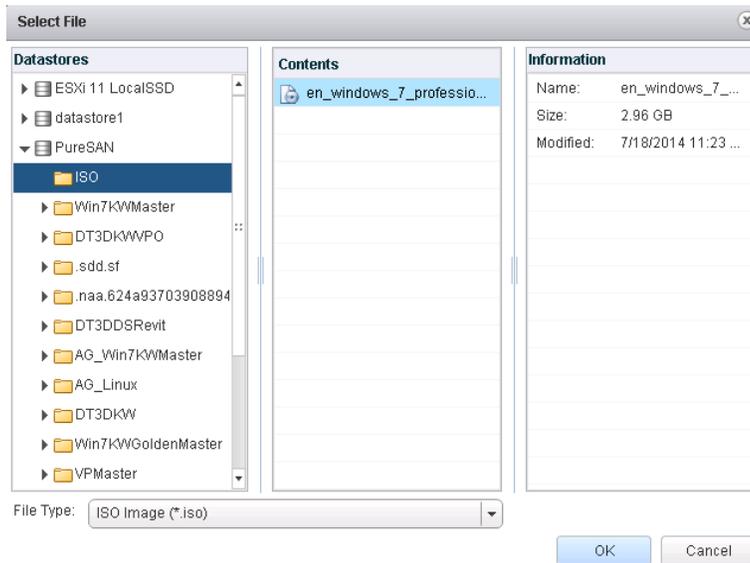
5. Right-click the virtual machine, and then select **Power>Power On** to start the virtual machine and boot from the .ISO to install the operating system.

The virtual machine boots from the selected .ISO.

> 💬 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 the both the operating system and VMware tools have been installed.

6. Perform a Custom (fresh) installation of Windows 7 on the virtual machine.
   During installation, Windows reboots the VM several times.

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

8. Go through the initial Windows setup wizard to name the computer, create a local account, set the timezone, choose update installation policy, etc.

   Windows 7 is now installed on 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.

1. Select the **Summary** tab from the virtual machine console.

2. Click the **Install VMware Tools** link in the yellow bar.
   The *Install VMware Tools* window displays.

3. Click **Install VMware Tools**.
   Back in the virtual machine console, Windows 7 detects the CD image and the *AutoPlay* window should open. If not, browse to the virtual CD-ROM in the virtual machine and access it manually.

32-bit installer:

64-bit installer:

4. Click **Run setup.exe or setup64.exe.**

   The *User Account Control* popup may display.
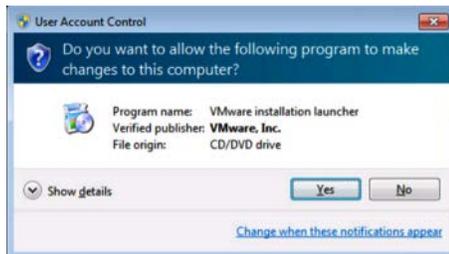
5. If UAC prompts, then click **Yes.**

6. The installer begins, click **Next.**
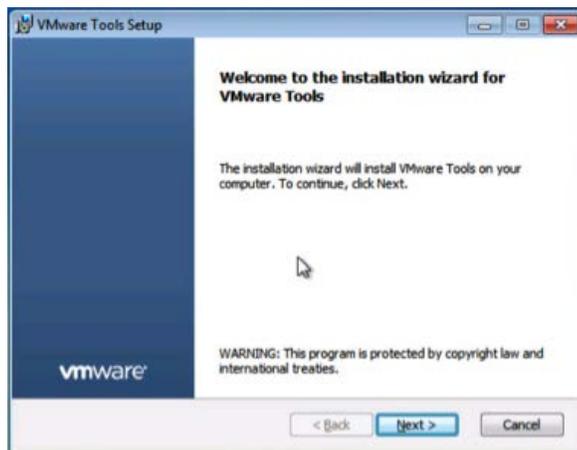
7. When prompted, select **Complete** installation, click **Next**, and accept all defaults.
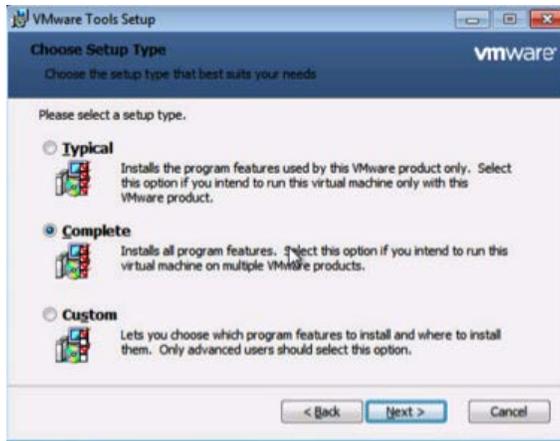
8. When prompted, select **Install** to begin installation:
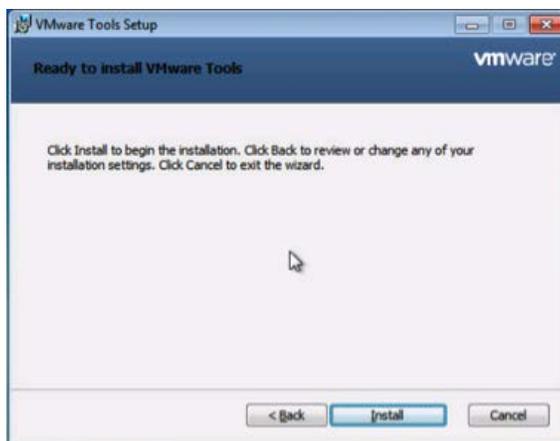
9. Check the **Always trust software from "VMware, Inc."** checkbox and click **Install** to continue.

10. The *Set Network Location* window may display, if so then select **Work Network**.

11. Click **Finish** in the VMware Tools installer and reboot the virtual machine when prompted. This reboot is critical to ensure the tools are now the loaded drivers.
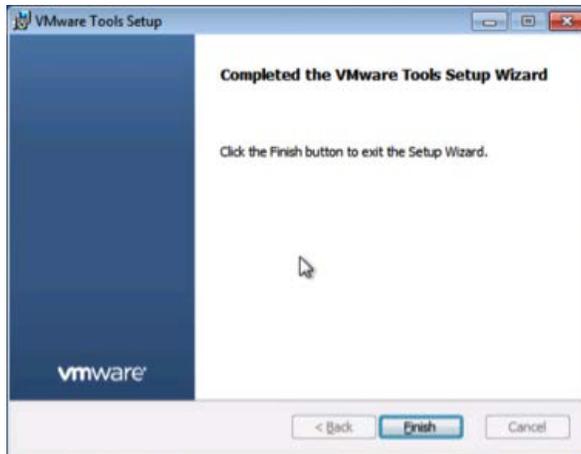
VMware Tools is now installed on the virtual machine

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

Adding a VM to the domain:

1. On the VM, click Start, then right click on Computer to bring up its menu, click Properties:
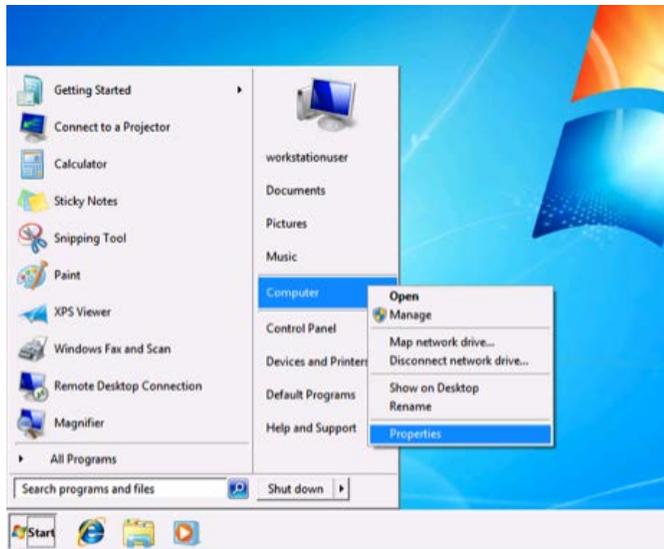
2. On the Properties window, click Change settings:
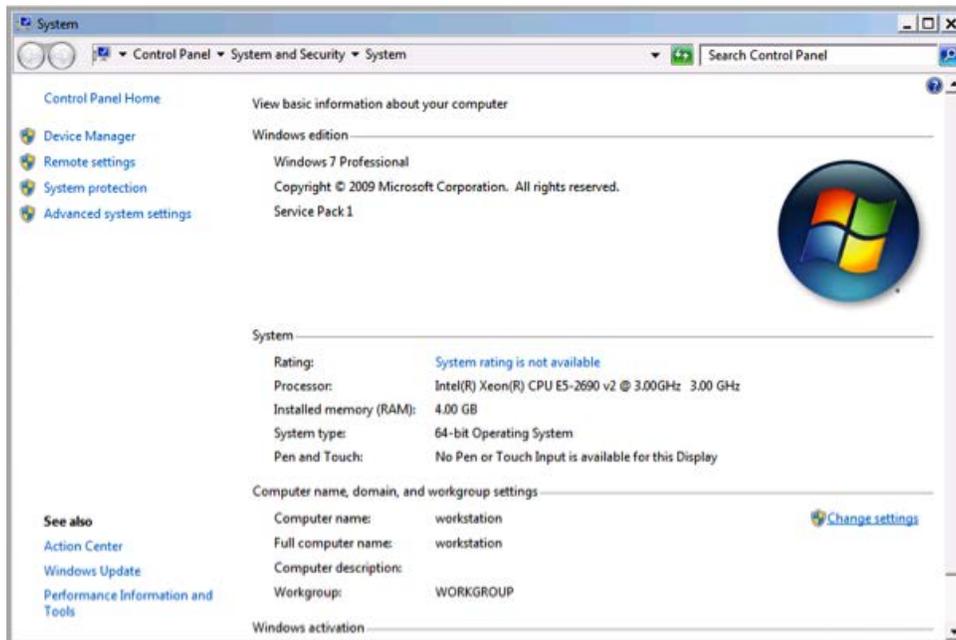
3. This brings up the System Properties window, on the Computer Name tab click Change:
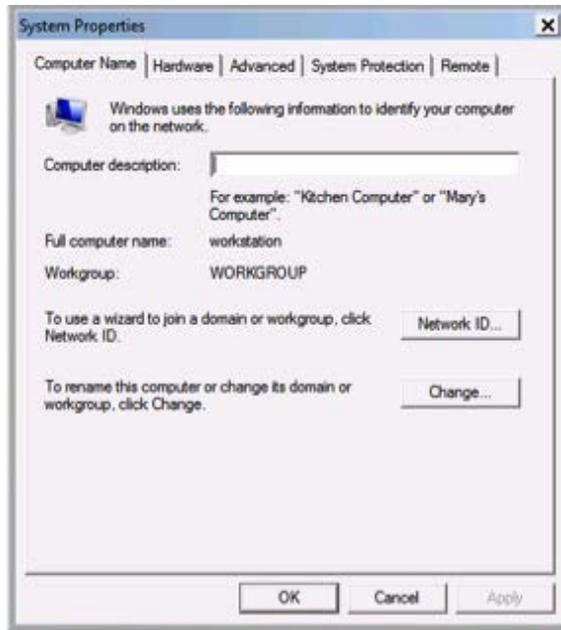
4. On the Computer Name/Domain Changes window, enter in an appropriate Computer name, then Domain name, and click OK. Our chosen naming is shown below, use what is appropriate for your POC/trial.

5. A security window pops up, fill in your specific domain administrator credentials and click OK:
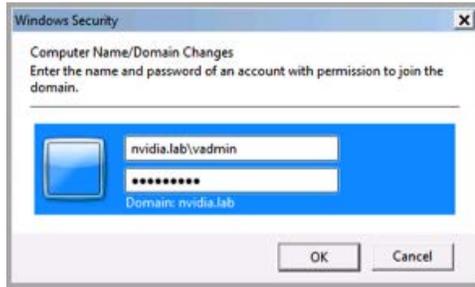
6.On successful authentication you will see the following welcome pop-up showing your VM is now on the domain (the domain name should reflect your domain information):
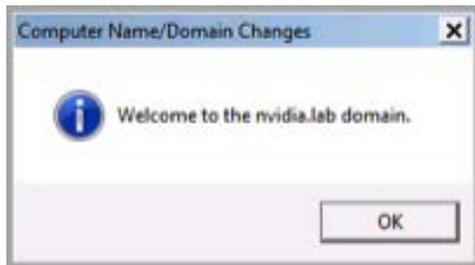
7.Click OK and the VM needs to reboot to complete the process, click OK again and the VM reboots immediately.
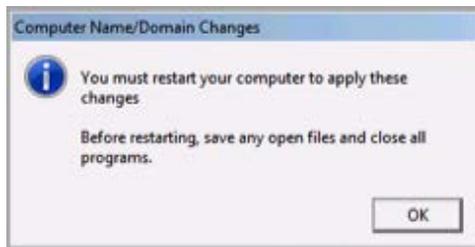
10.5   INSTALLING CITRIX VIRTUAL DELIVERY AGENT

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 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 data store to the virtual machine CD drives.
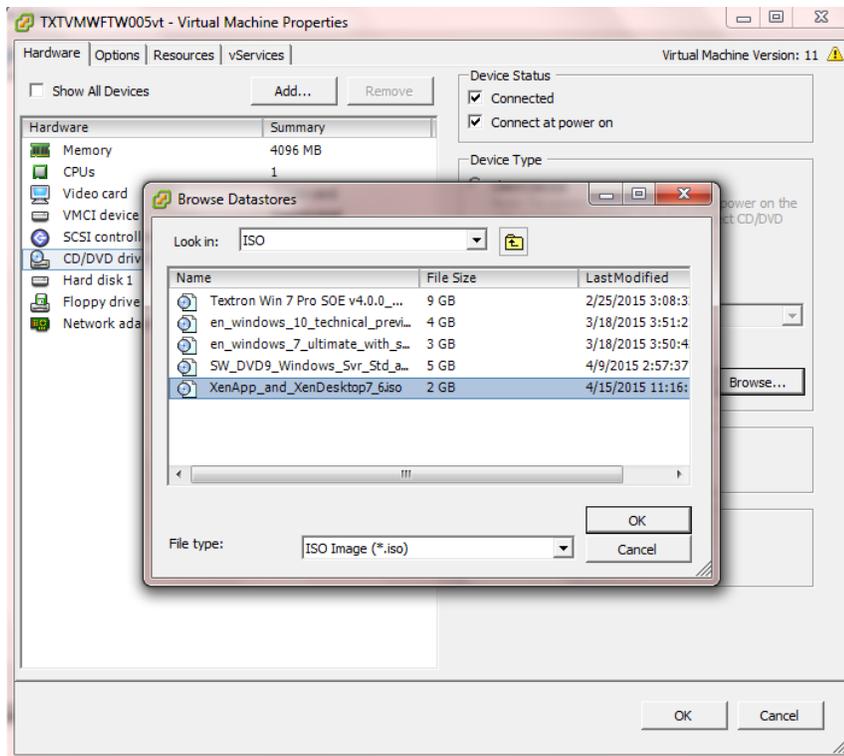
2. Launch the CD with AutoPlay or select the AutoSelect.exe application to start the Citrix XenApp and XenDesktop installer. Select **Start** button next to XenDesktop Deliver applications and desktops.
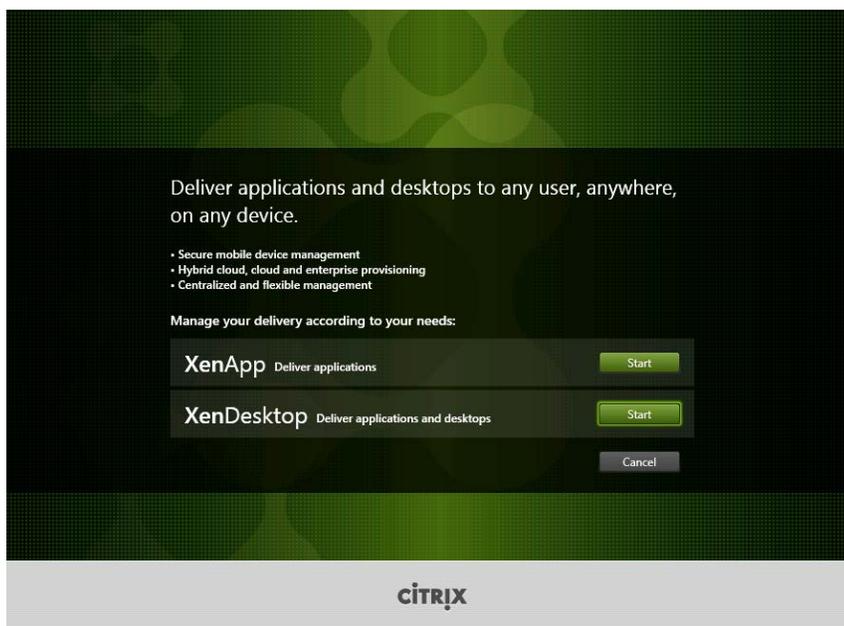
3. Select **Windows Delivery Agent for Windows Desktop OS** under **Prepare Machines and Images** on the XenDesktop 7.6 window.
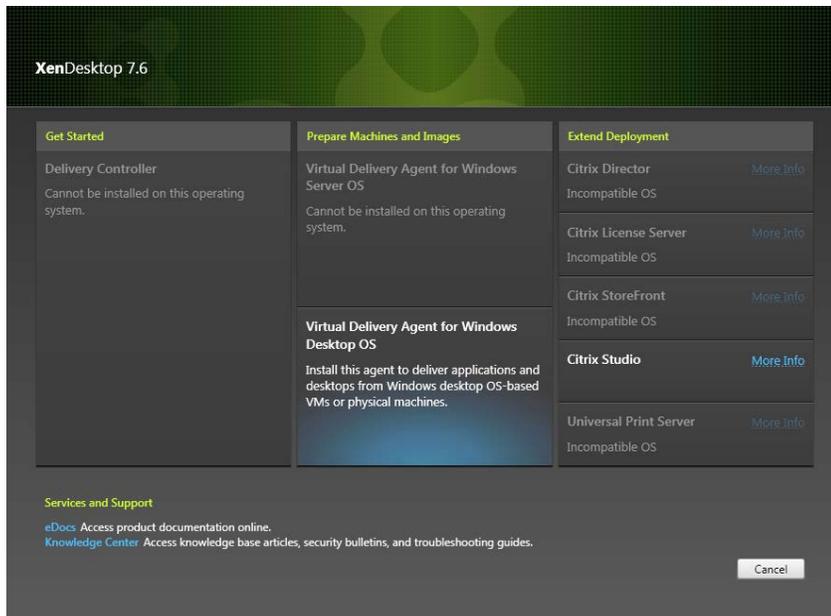
4.Select **Yes** to agree to the User Account Control prompt if displayed.
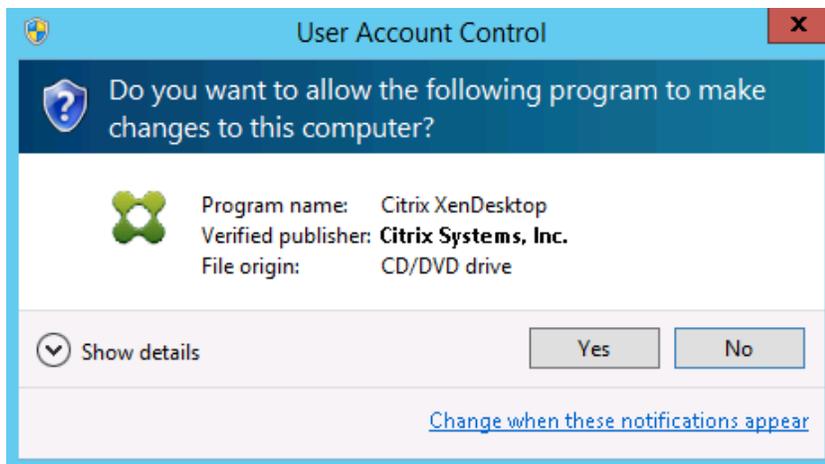
5.Accept the default **Environment** value of **Create a Master Image** radio button, and select **Next** to continue.
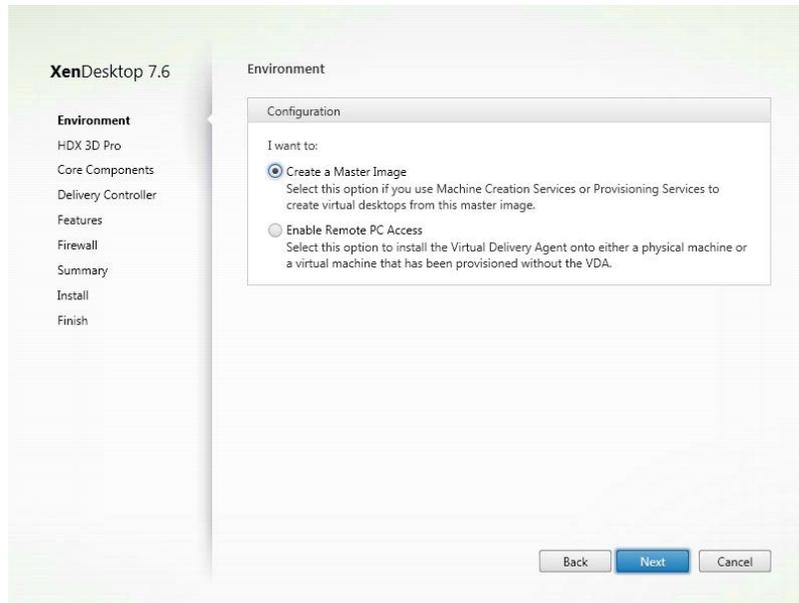
6. Select the **Yes, install the VDA for HDX 3D Pro** radio button on the HDX 3D Pro Dialogue, and select **Next** to continue.
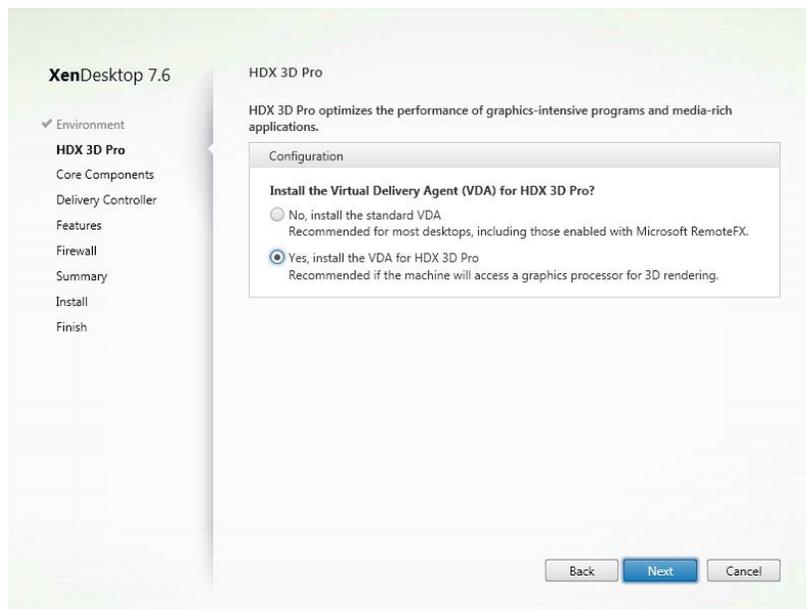
7. Deselect the **Citrix Receiver** check box on the **Core Components** dialogue, and select **Next** to continue..
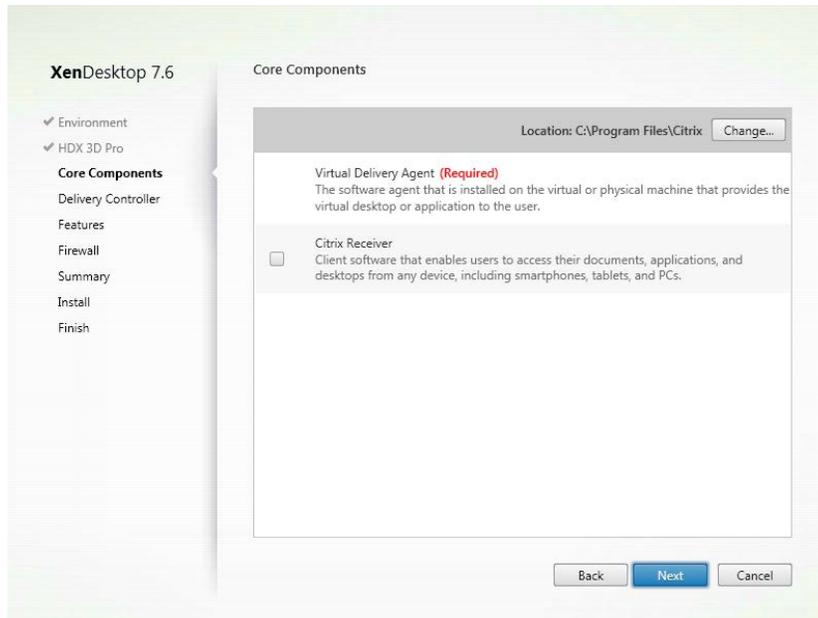
💬 Citrix Receiver can be installed separately later and would only be used if connecting to another XenApp or XenDesktop Store.

8. Enter the XenDesktop delivery controller address. This is the FQDN of the XenDesktop server. Click on the Text connection button to validate the address.
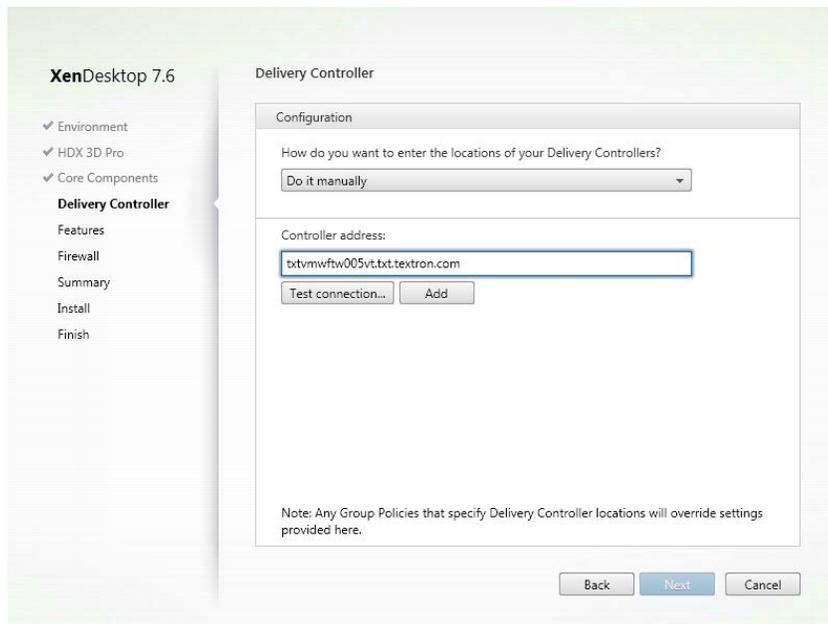
💬 IP Addresses are not supported as a controller address, only Fully Qualified Domain Names are allowed.

9. 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.
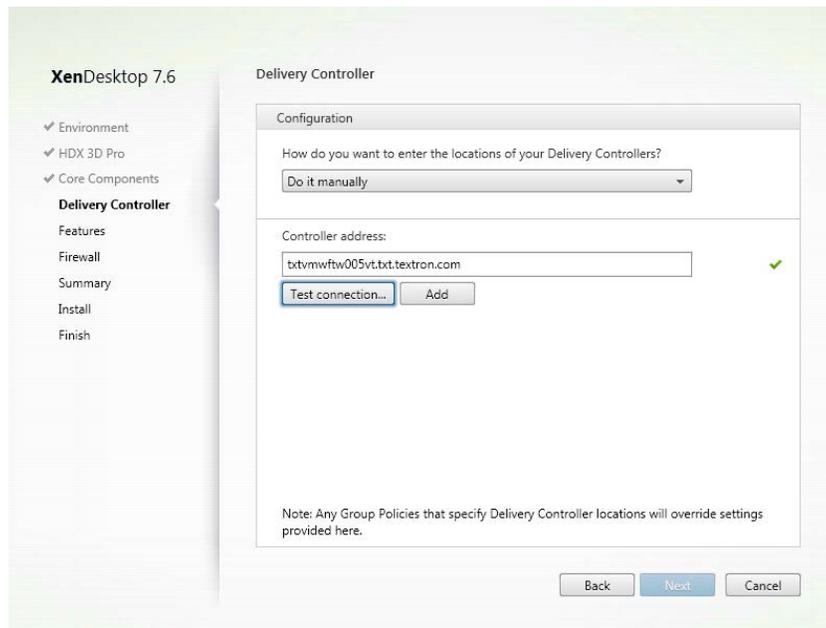
10. The delivery controller address will be displayed above the delivery controller address entry box once it has been added. Select **Next** to continue.
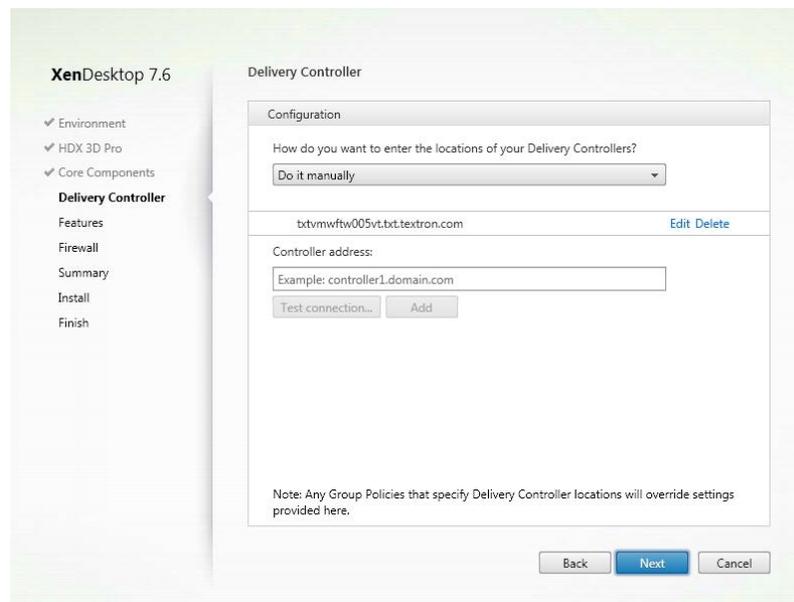
11. Accept the defaults on the **Features** dialogue and select **Next** to continue.
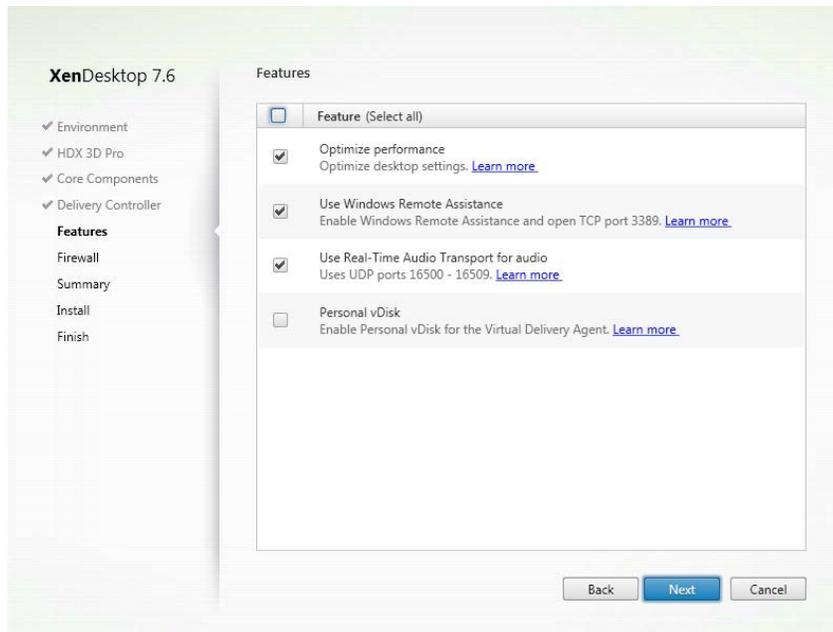
> Personal vDisk can be enabled at a later date if desired

12. Accept the default value to automatically create the Firewall entries. Select **Next** to continue.
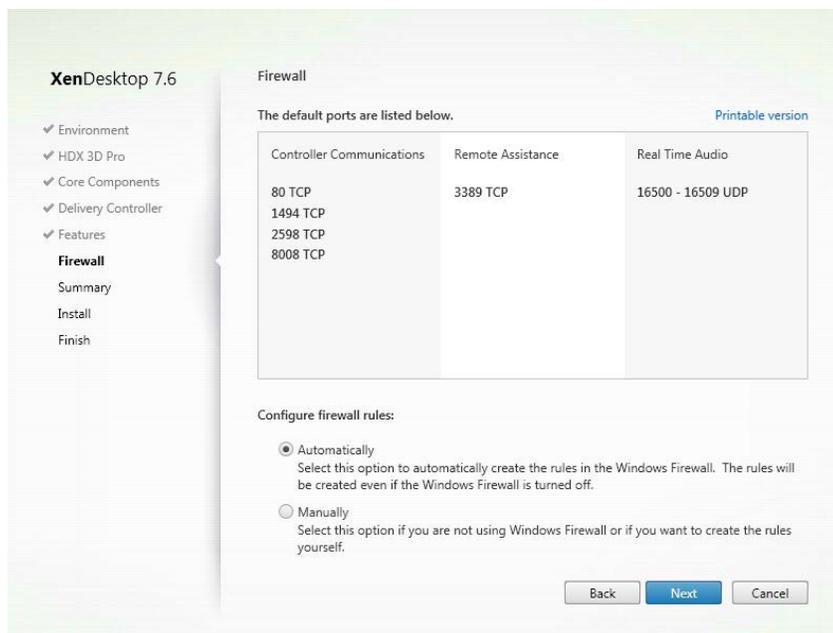
13. Review the summary and select **Install** to continue.
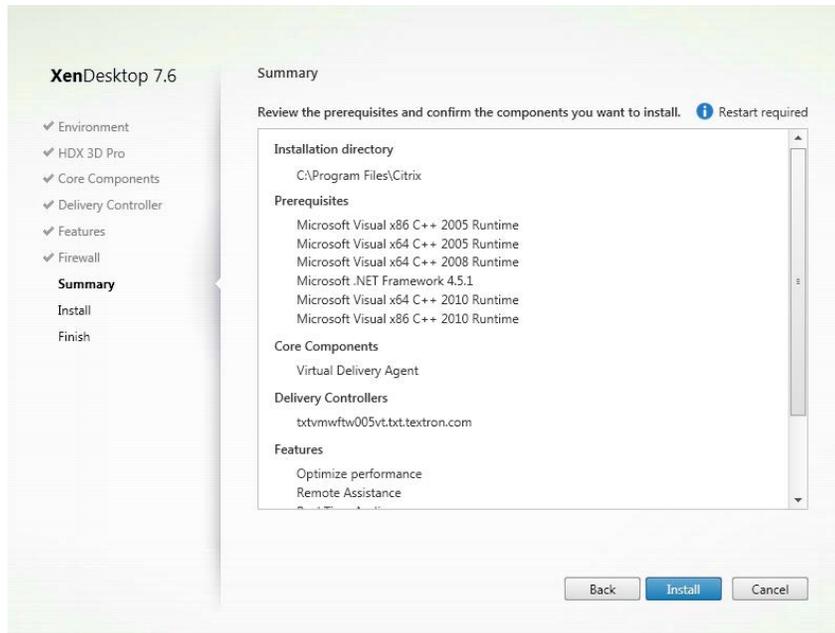
14. If the Installation completed successfully accept the default to restart the machine and select **Finish** to continue.
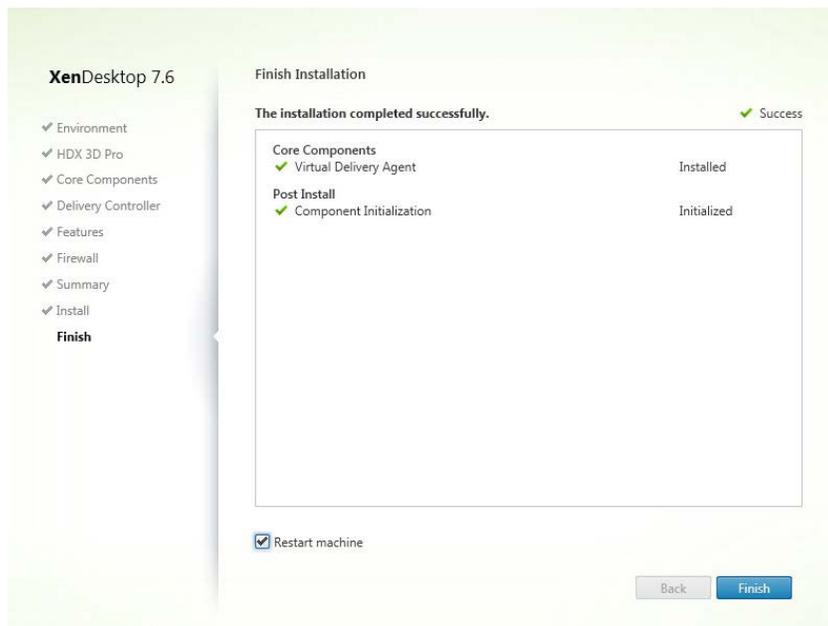
10.6 ADDITIONAL VIRTUAL MACHINE SETTINGS

1.Perform the following additional tasks on the virtual machine as required:

- OPTIONAL: Upgrade Microsoft Internet Explorer to Version 11 from:
  http://windows.microsoft.com/en-us/internet-explorer/download-ie.
  Reboot the virtual machine when prompted.

> 💬 To evaluate browser-based HTML5 applications, consider using newer browsers that utilize hardware acceleration within virtual desktop environments.

- Turn Off Windows Firewall for all network types.

> ❗ **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.

2. Shut down the virtual machine once this is completed

3. Close the remote console; this will not be functional when 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

10.7 ENABLING THE NVIDIA VGPU

Use the following procedure to enable vGPU support for your virtual machine (you must edit the virtual machine settings):

1. Power down the virtual machine.

2. Click on the VM in the Navigator window. Select the Manage tab and settings. Click on the Edit button.
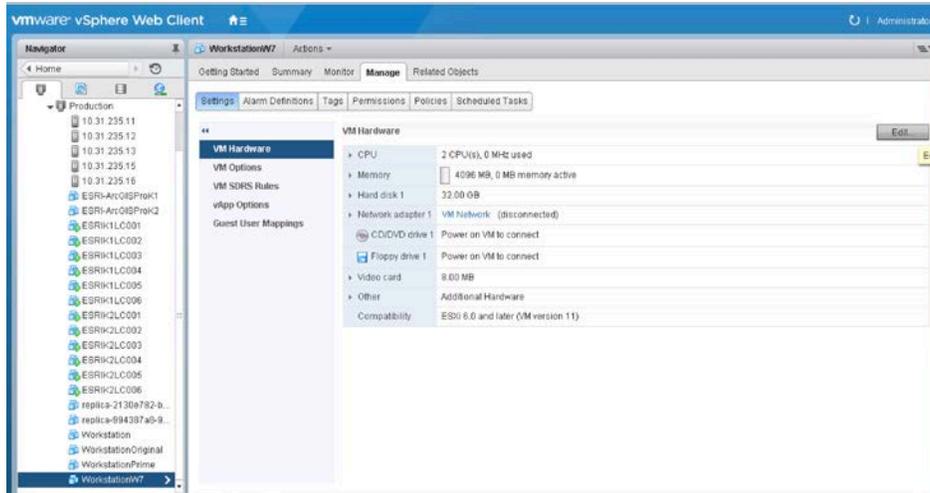
3. The Edit Settings dialog appears.
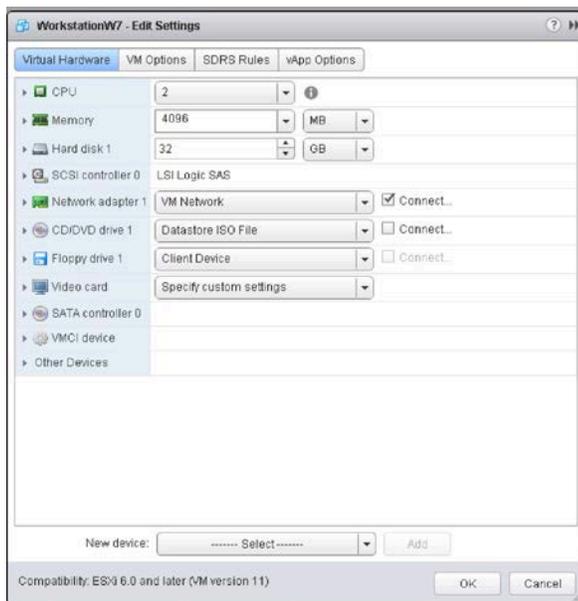
4. Click on the New Device bar and select Shared PCI device.

5. Click on Add to continue

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

7. Use the GPU Profile selection bar to configure the GPU.
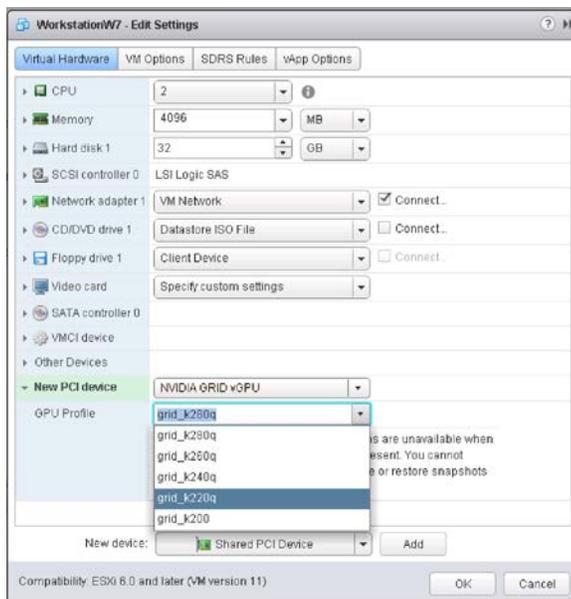
> The k2xx options will appear in servers with K2 GPUs, and the k1xx options will appear in servers with K1 GPUs. The K100 and K200 profiles are not Quadro certified and are being phased out.

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

10.8   INSTALLING NVIDIA DRIVER IN WINDOWS VIRTUAL DESKTOP

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.

> **!** **CAUTION:** DO NOT REBOOT THE VIRTUAL MACHINE IF YOU HAVE OLDER NVIDIA DRIVERS INSTALLED. DOING THIS WILL CAUSE A BLUE SCREEN.

3. Log into Windows and open the Device Manager.
   The **Standard VGA Graphics 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.
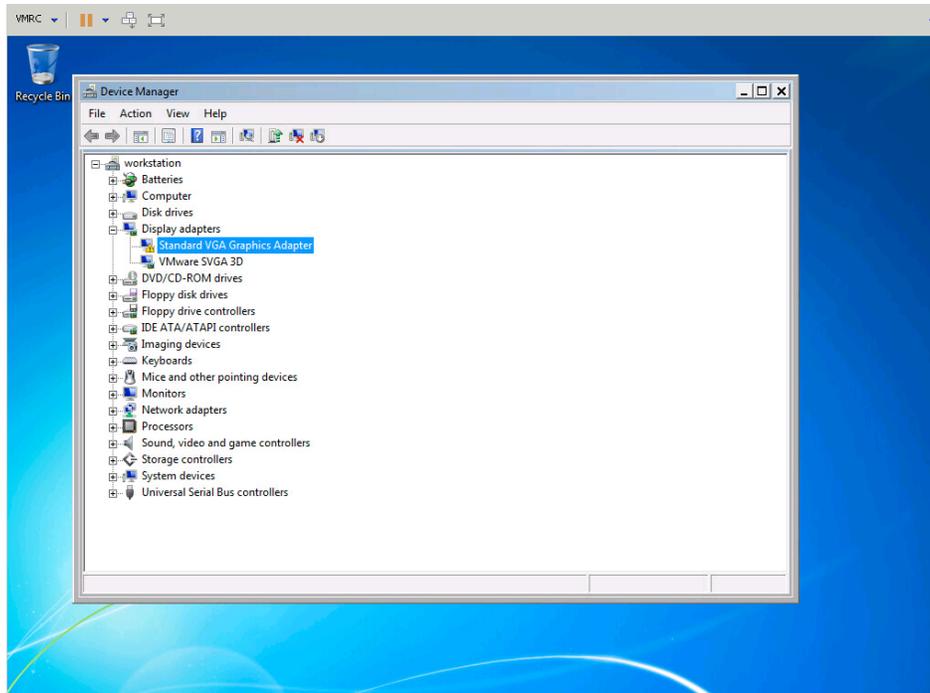
4. Locate the NVIDIA driver where you extracted it earlier and right-click the install application icon to launch it. (Recommendation: Have the software on a share volume that can be mounted by the VM for quick access and extract the drivers).
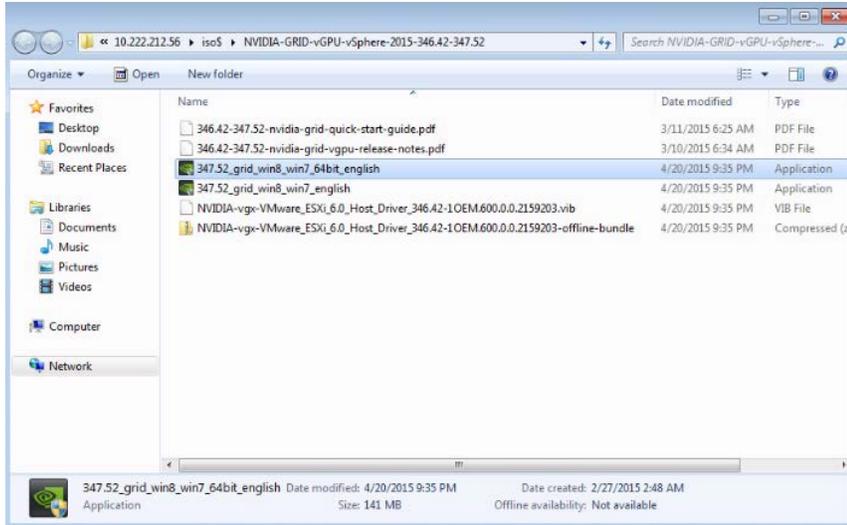
> ⚠ **CAUTION**: NVIDIA vGPU vSphere VIB version and NVIDIA driver for Windows version need to match.

5. Right-click on the installer application and select **Run as administrator**.

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

Creating your first vGPU Virtual Desktop
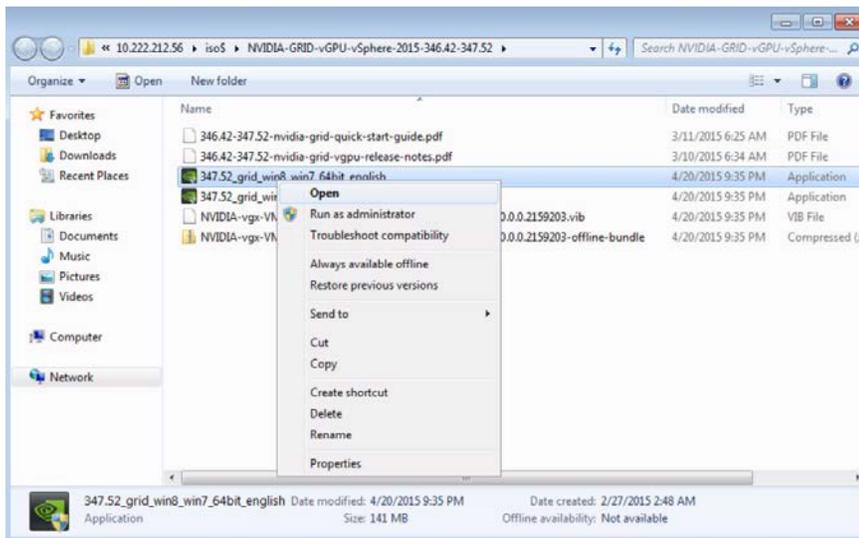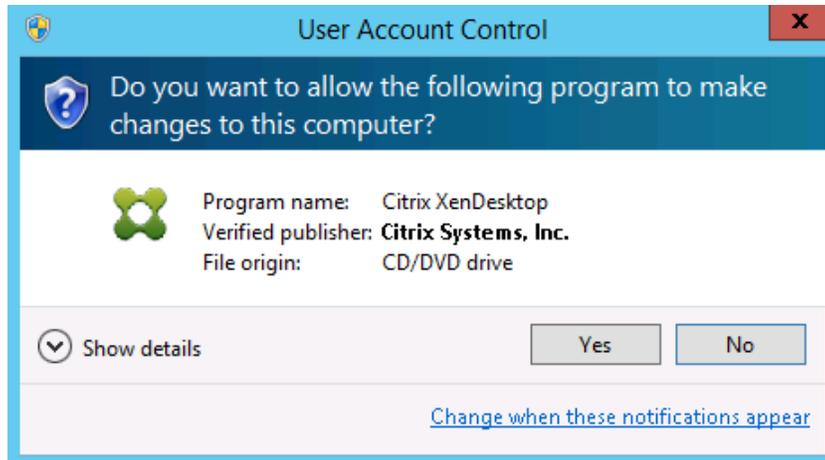
7. Select **OK** to accept the default location where the driver files will be saved.

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

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

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

11. A window displays when the NVIDIA Graphics Driver installation is complete. Select **RESTART NOW** to Reboot the VM to complete the install.

> After restarting, the mouse cursor may not track properly using VNC or vSphere console. If so, use Remote Desktop.

10.9 REMOVING THE 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.

1.Shutdown the virtual machine, and using the vSphere Web Client **Navigator** with the left hand pane to the **VMs and Templates** tab. **Right-Click** on the master image VM and select **Edit Settings** menu item.
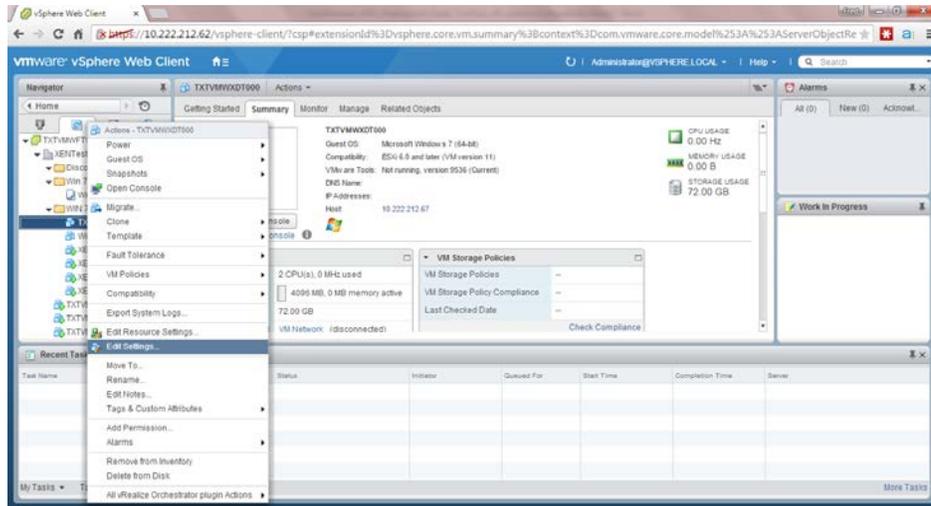
2. Select the **grey-x** on the far right on the **PCI Device** item that is connected to the
   NVIDIA GRID vGPU to select the device for removal.
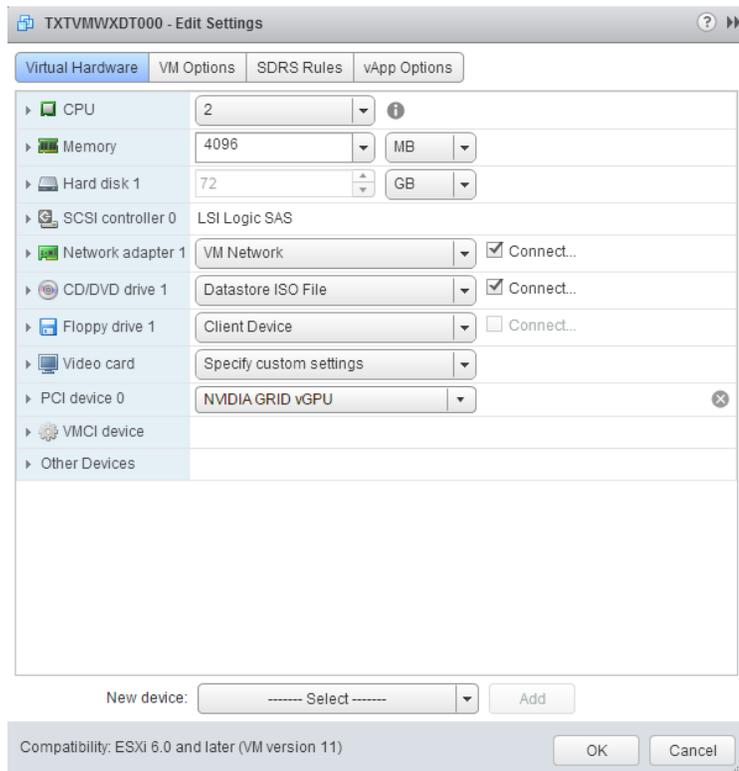
3. Select the **OK** button at the bottom of the **Edit Settings** dialogue to save the settings.
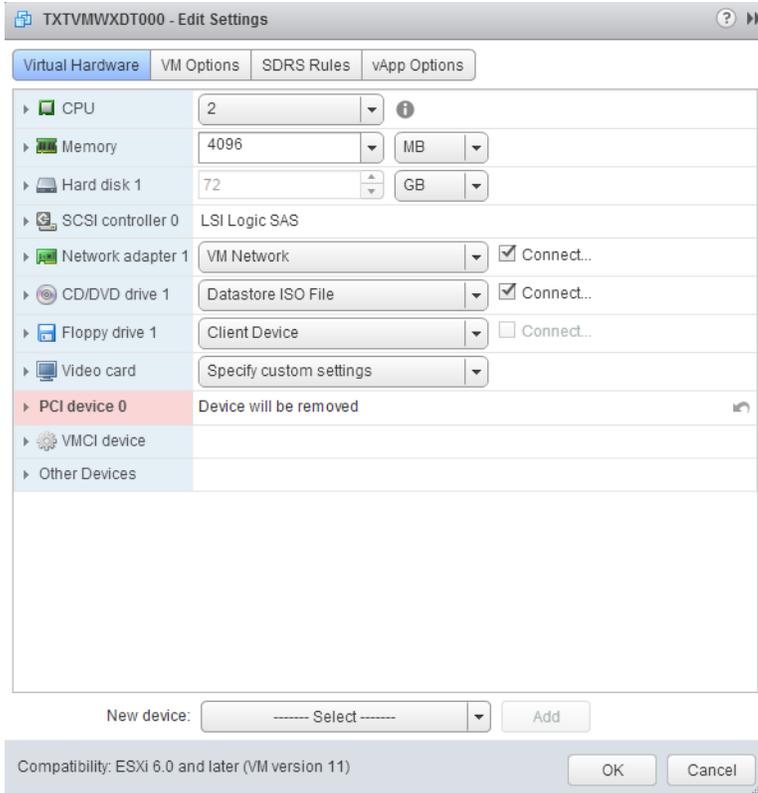
4. The Recent Tasks pane at the bottom will display the status of Complete once the device has been removed. A XenDesktop machine catalog can be safely configured now.
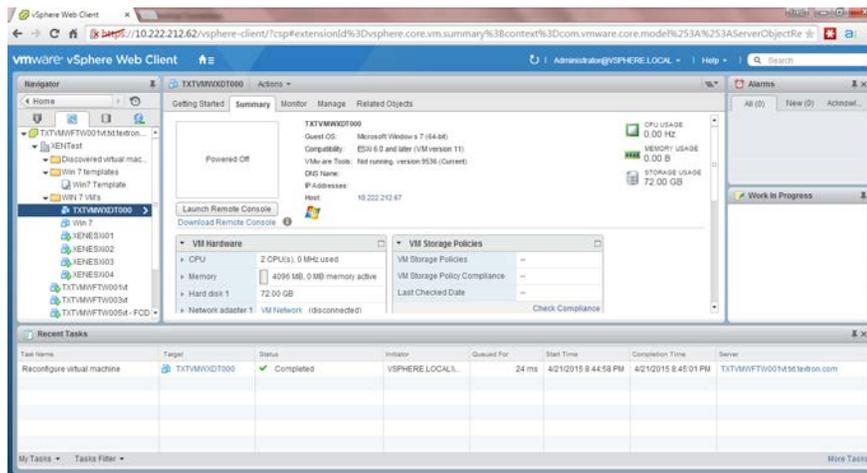
Chapter 11.
CREATING XENDESKTOP MACHINE CATALOGS AND DELIVERY GROUPS

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 Virtual Machine 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 can reduce evaluation time.

11.1   CREATING A SNAPSHOT

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

1. Right-click a virtual machine and then select **Snapshot**➔**Take Snapshot**…

    The **Take VM Snapshot for** **<vmname>** popup displays.
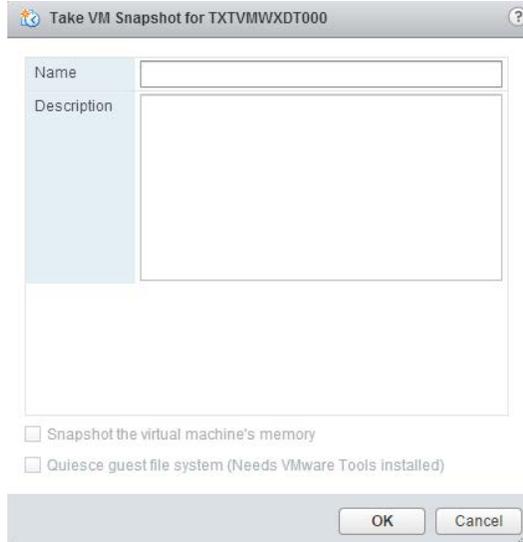
2. Enter a name for the snapshot.

3. Optionally enter a description for the snapshot.

4. 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.

3.Right-Click on **Machine Catalogs** menu item and select **Create Machine Catalog** menu item.
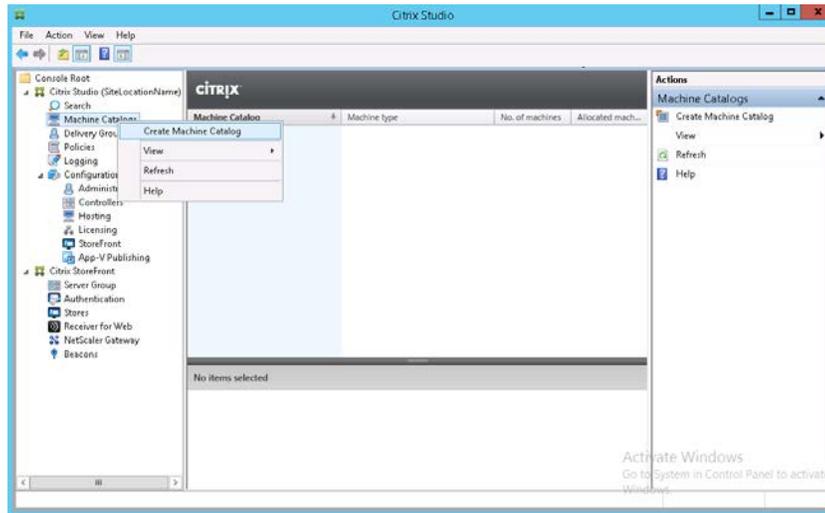
4.Select **Next** button on the **Machine Catalog Setup** dialogue to begin.
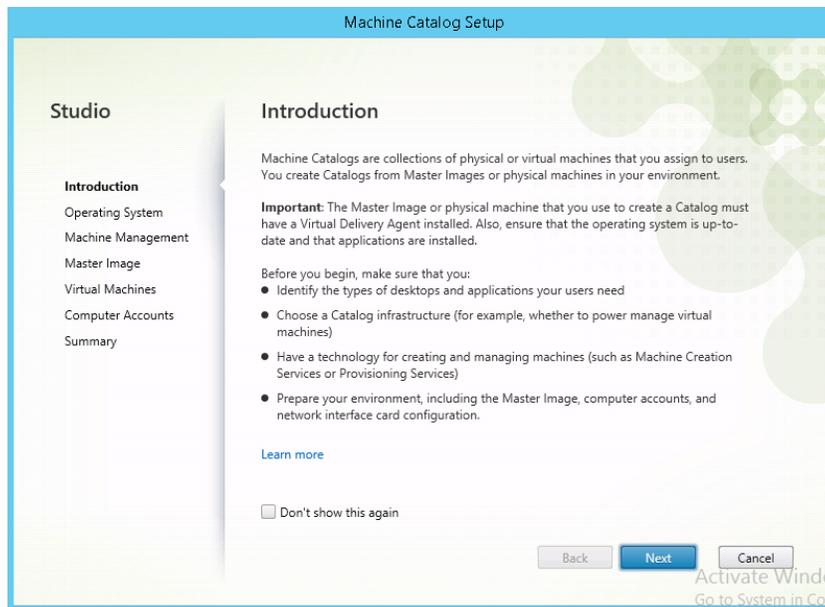

5.On the **Operating System** dialogue, select the **Windows Desktop OS** radio button and then select **Next** to continue.
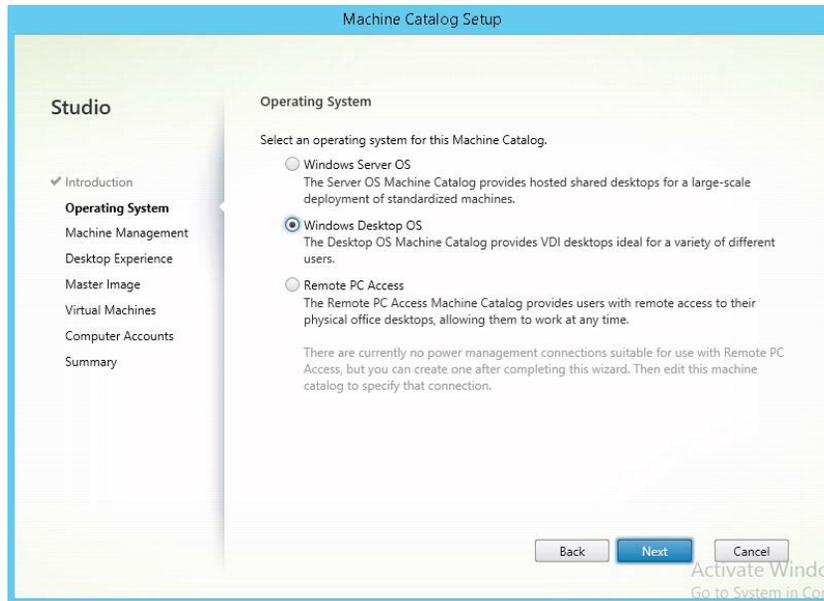
6. On the **Machine Management** dialogue select the **Another Service or technology** radio button for the **Deploy machines using** section. Select **Next** to continue.
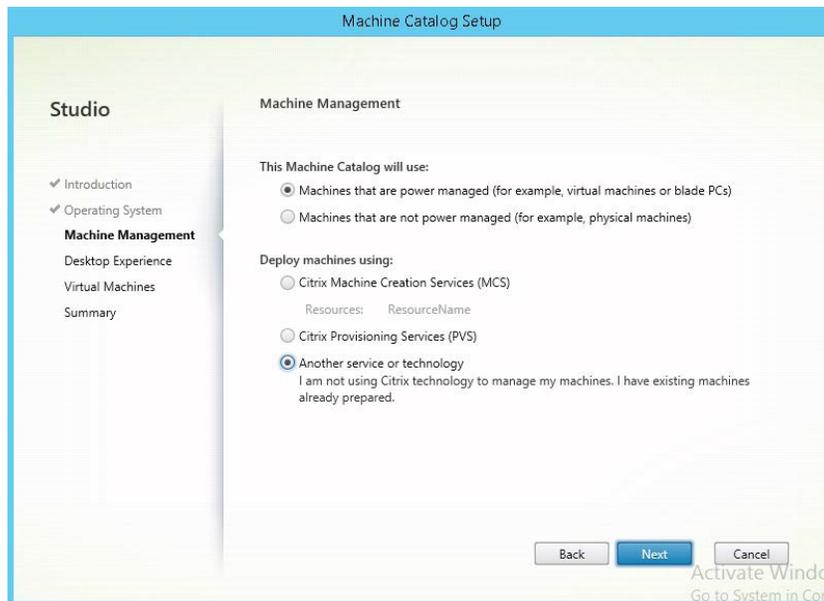
💬 This is only used for the master image, the random pool will use Citrix Machine Creation Services (MCS)

7. On the **Desktop Experience** dialogue, accept the default and select **Next** to continue.
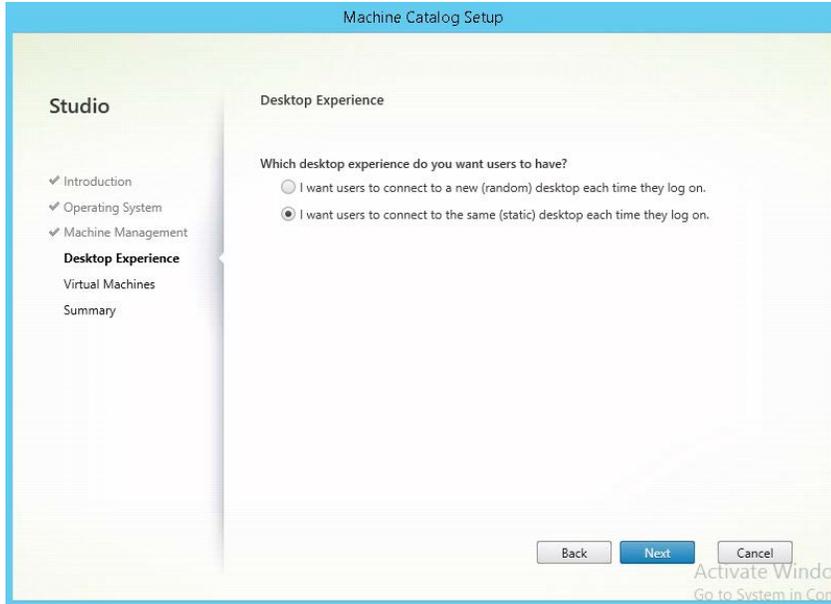
8. On the **Virtual Machines and Users** dialogue, select the Add VMs button to begin to select the master image to the machine catalog.
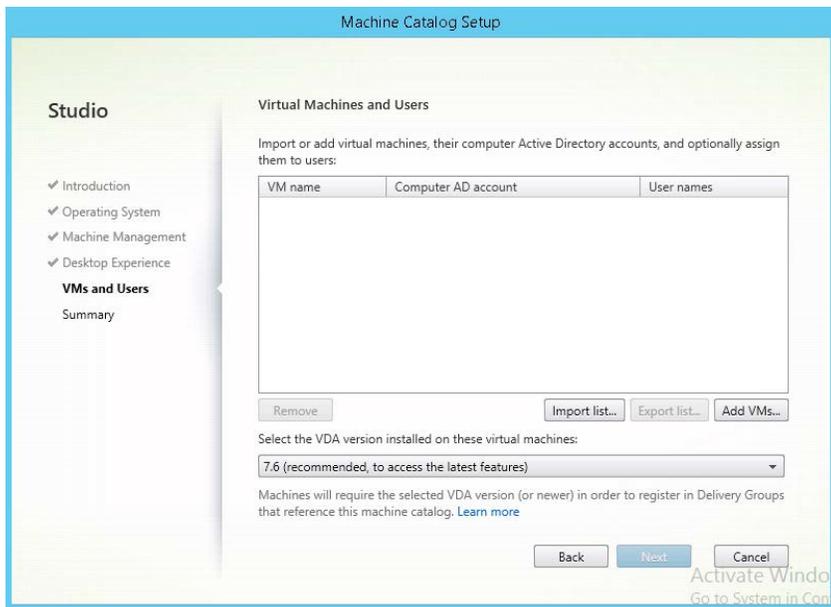
9. On the **Select VMs** dialogue, navigate through the vCenter VMs site tree and select **check box** next to the machine or machines to add to the master image machine catalog. Select the **OK** button to continue.
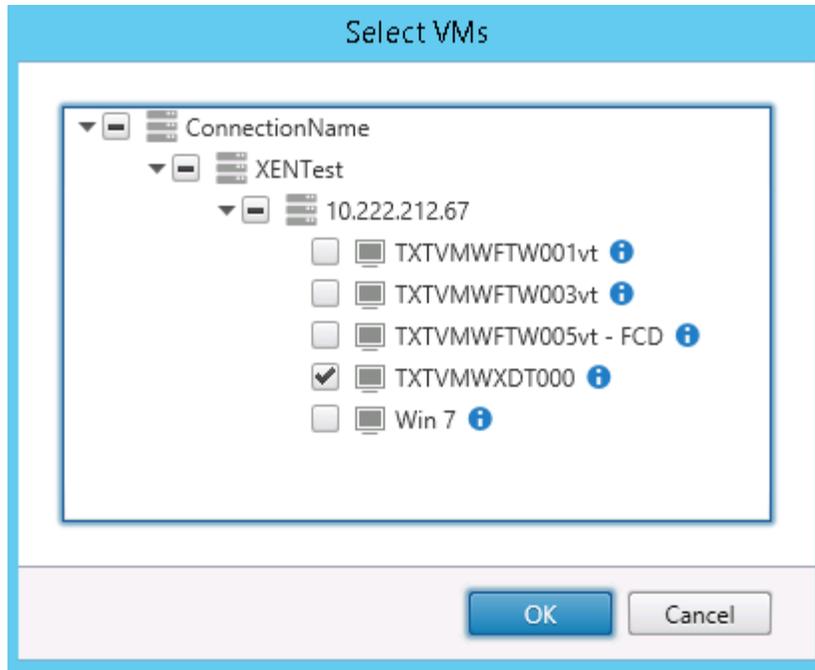
10. Select the **elipse** button next to the **Computer AD Account** field to search Active Directory for the machine account to add.
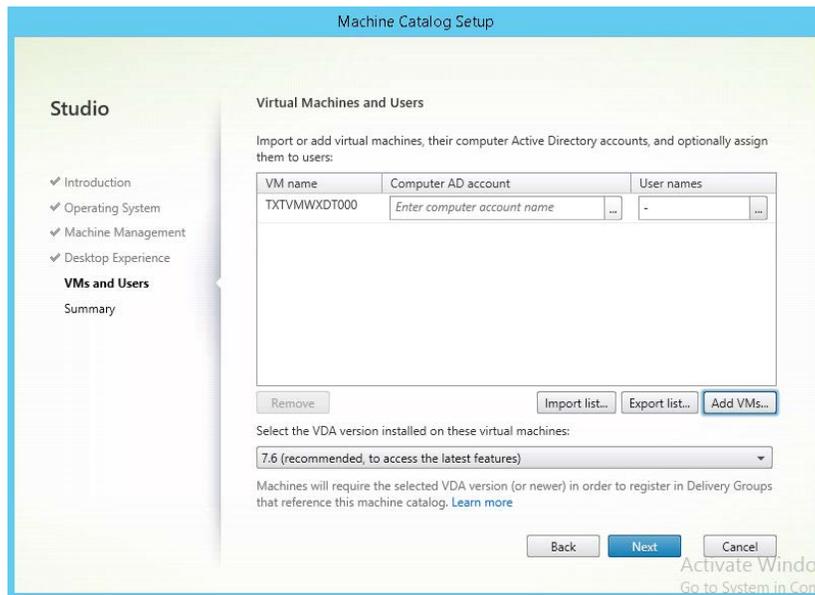
11. Enter the computer name into the **Active Directory Select Computer** dialogue. Select **Check Names** button to verify the computer account exists and then select **OK** button to continue.
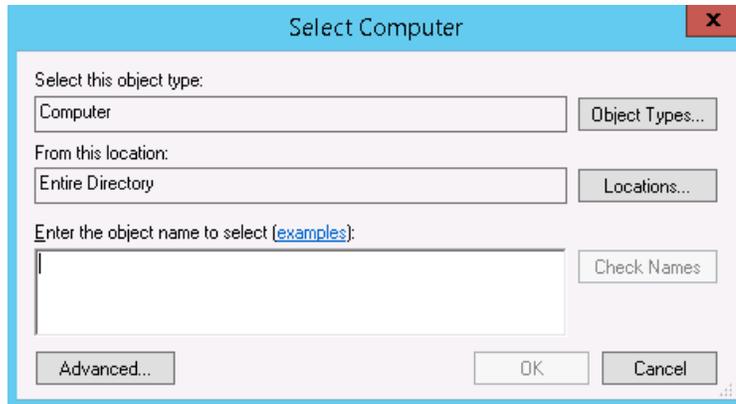
12. Select the **elipse** next to **User Names** field to search Active Directory for the user account to add.
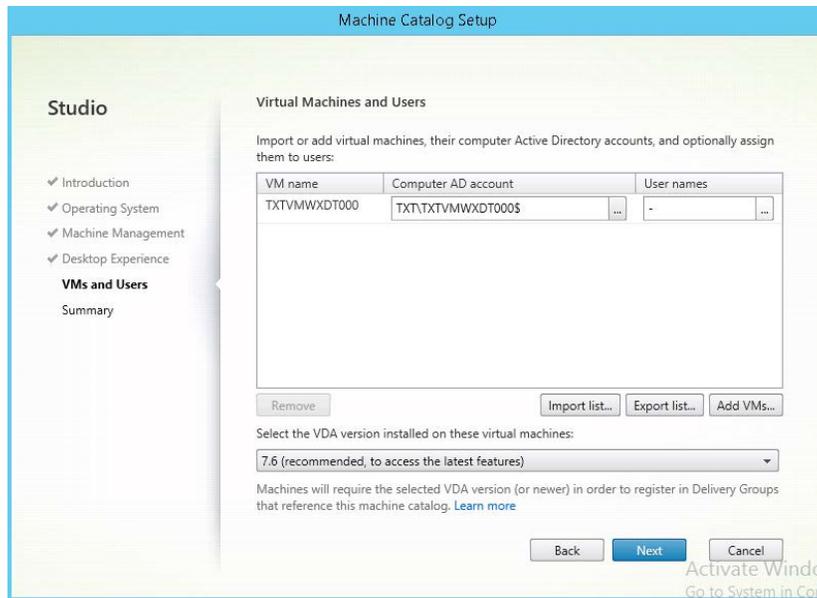
13. Enter the users name into the **Active Directory Select Users** dialogue. Select the **Check Names** button to verify the users account then select the **OK** button to continue.
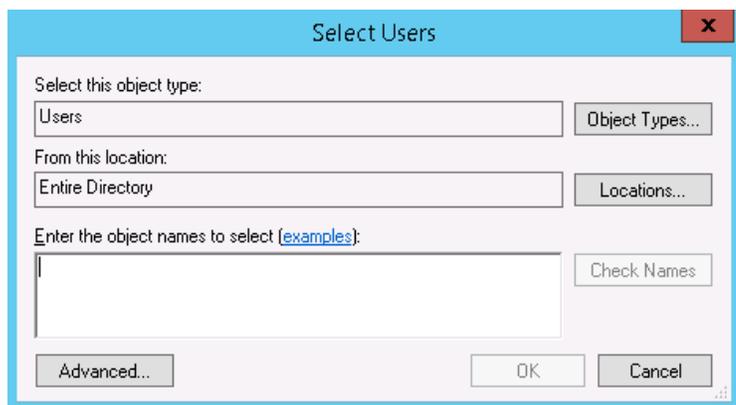
14. Once you have completed adding computer account and user account references for all machines in the catalog select **Next** to continue.

15. On the Summary dialogue enter a Machine Catalog name and then select Finish button to continue.
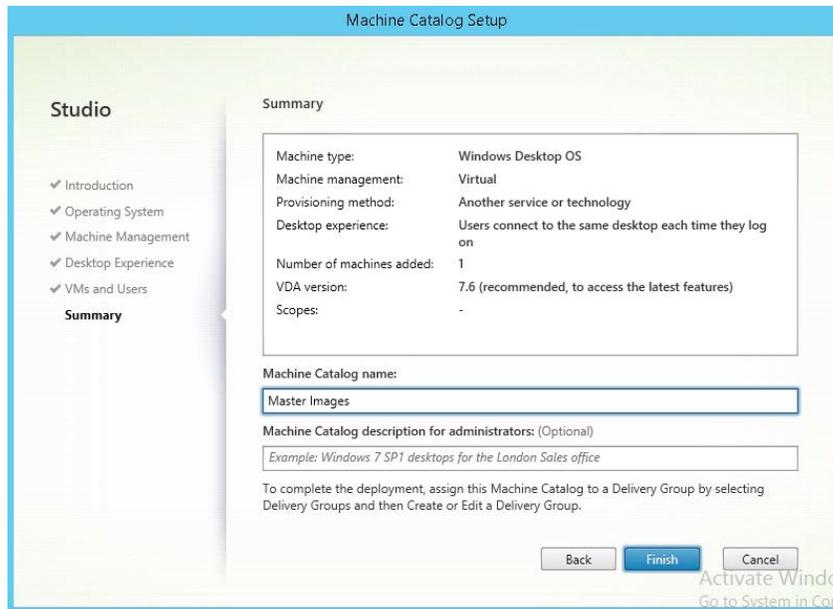
16. Once complete the machine catalog will be displayed in the center pane.
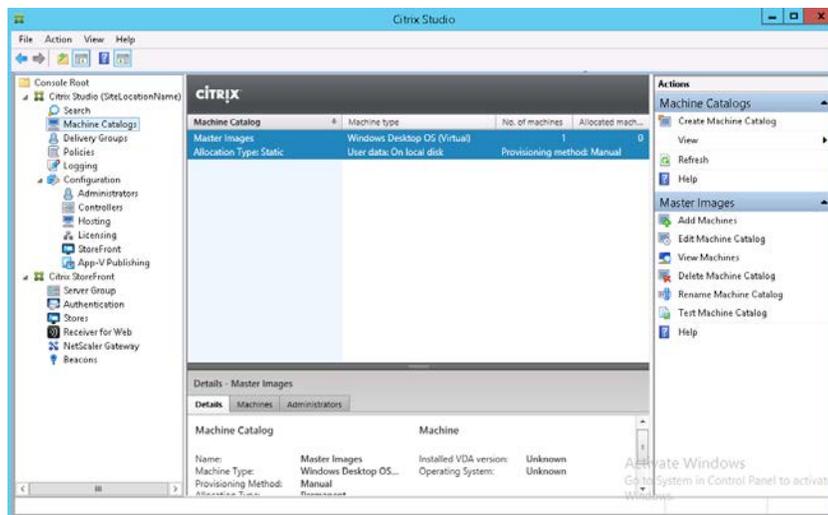
11.3   CREATING A XENDESKTOP RANDOM MACHINE CATALOG FROM A SNAPSHOT

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.

1. 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.

2. **Right-Click** on **Machine Catalog** from the left hand pane and select **Create Machine Catalog** menu item.
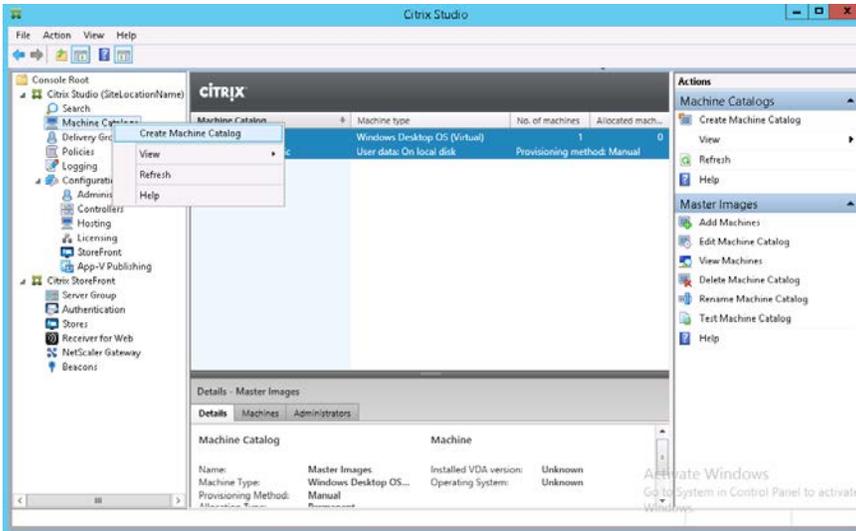
3. Select **Next** to continue.
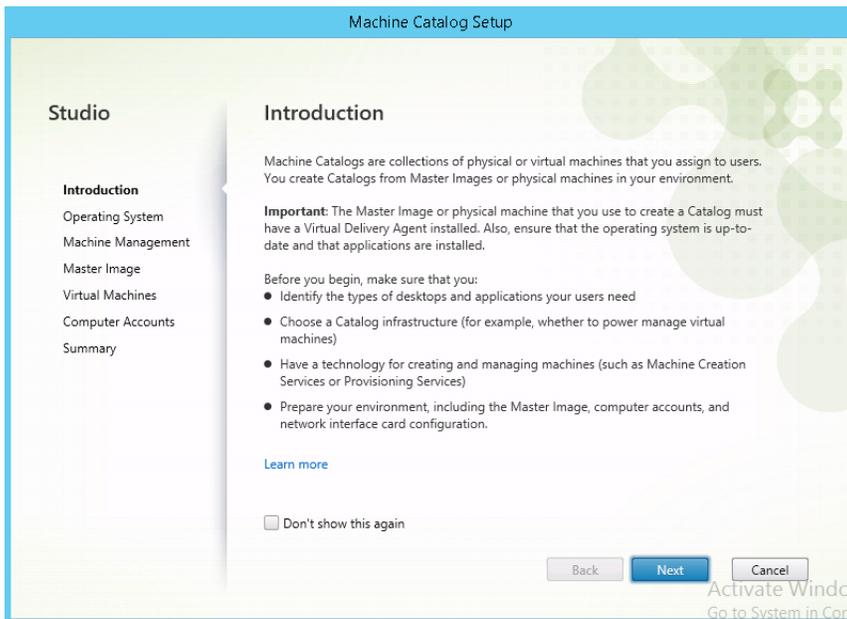
4. On **Operating System** dialogue select **Windows Desktop OS** radio button then select **Next** to continue.
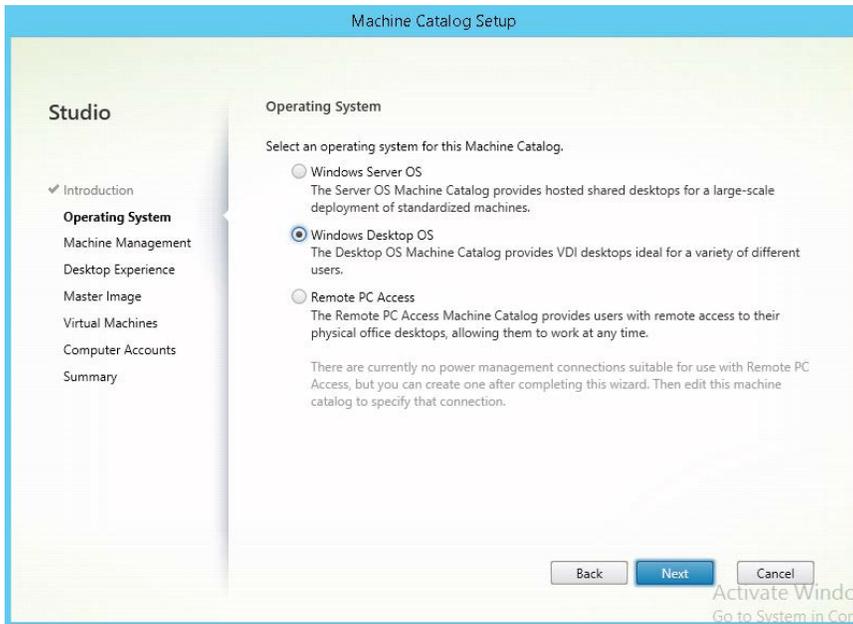
5. On **Machine Management** dialogue accept defaults for **Citrix Machine Creation Services (MCS)** and then select **Next** to continue..
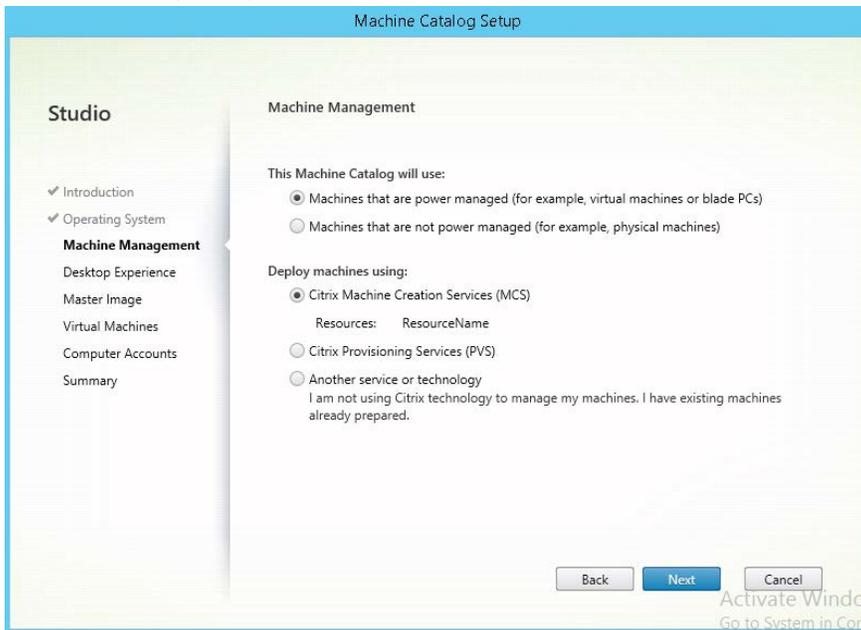
6. On the **Desktop Experience** dialogue select the **I want users to connect to a new (random) desktop each time they logon** radio button and then select **Next** to continue.
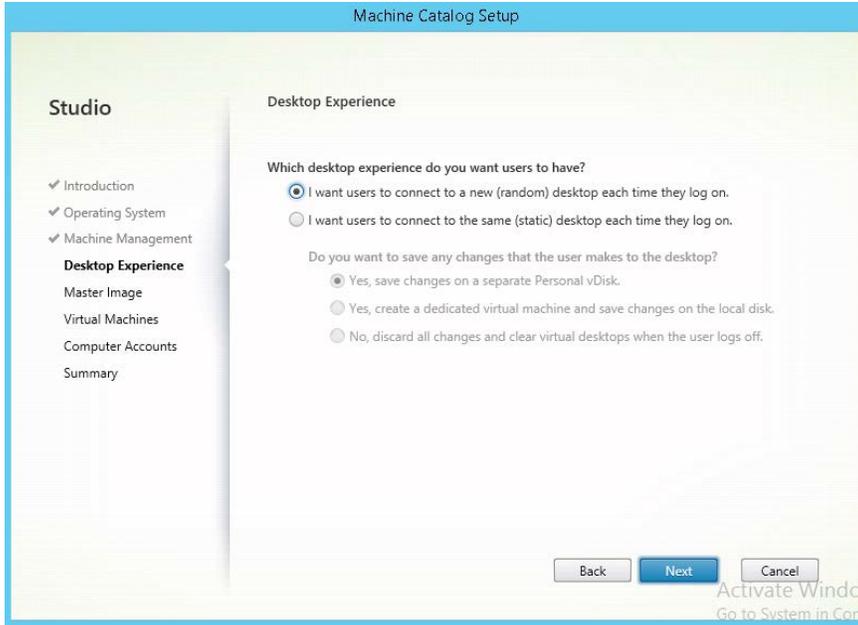
> If DHCP is not available use Static desktops, also make sure to use a dedicated virtual machine and save changes on the local disk.

On the **Master Image** dialogue 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.
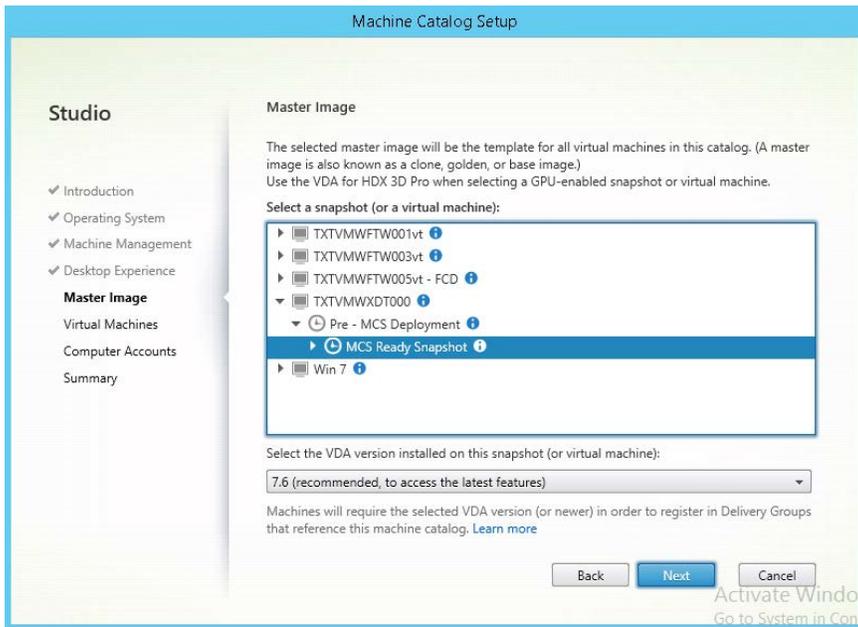
7. On the **Virtual Machines** dialogue specify the **Number of virtual machines needed** and then select **Next** to continue.
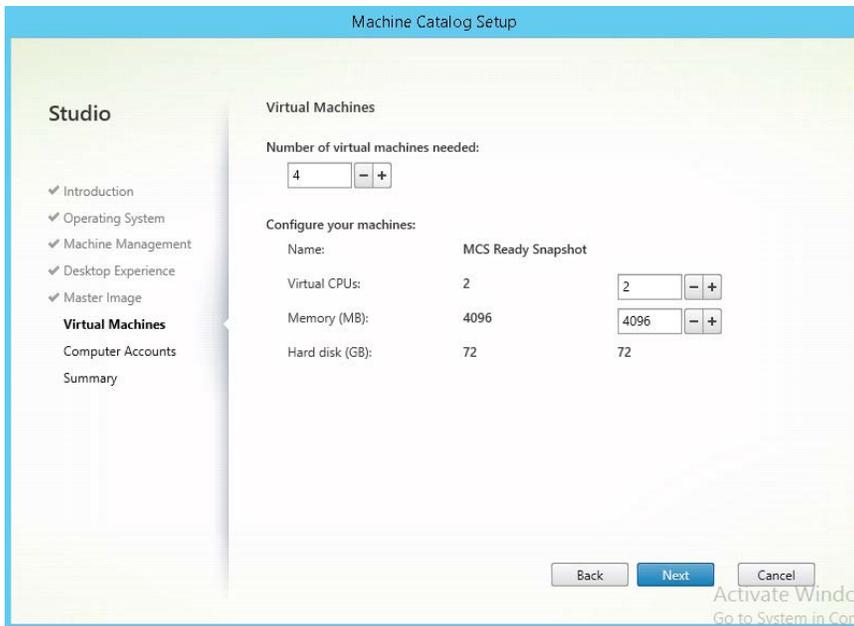
8. On the **Active Directory Computer Accounts** select the **Domain** to create machines in, browse the **OU tree** and select the OU to create machine accounts in, the OU should be displayed in the **Selected location** display field. Specify an **Account naming scheme** (append ## to the name for automatic numeric increments) then select **Next** to continue.
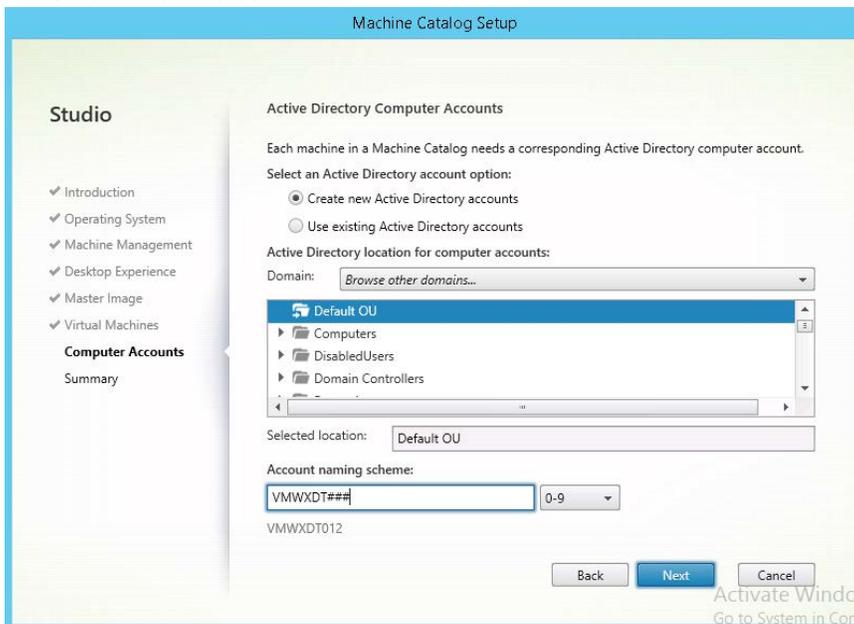
9. On the **Summary dialogue** enter the **Machine Catalog name** and then select **Finish** button to continue.
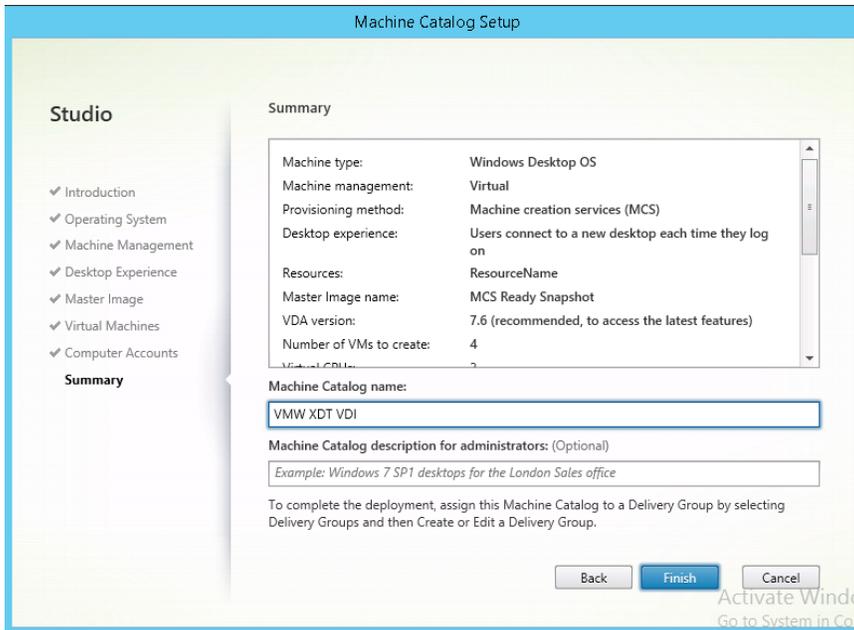
10. A progress bar will be displayed on the **Studio** dialogue while the machines are being provisioned. Select **Hide progress** button to allow the process to complete in the background.
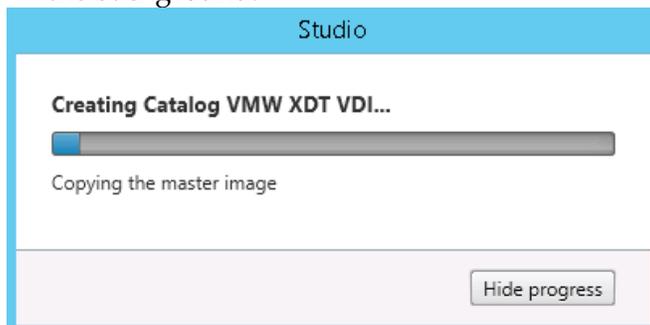
11. A progress bar continues to be displayed on the center pane of the **Machine Catalogs** screen in Citrix Studio.
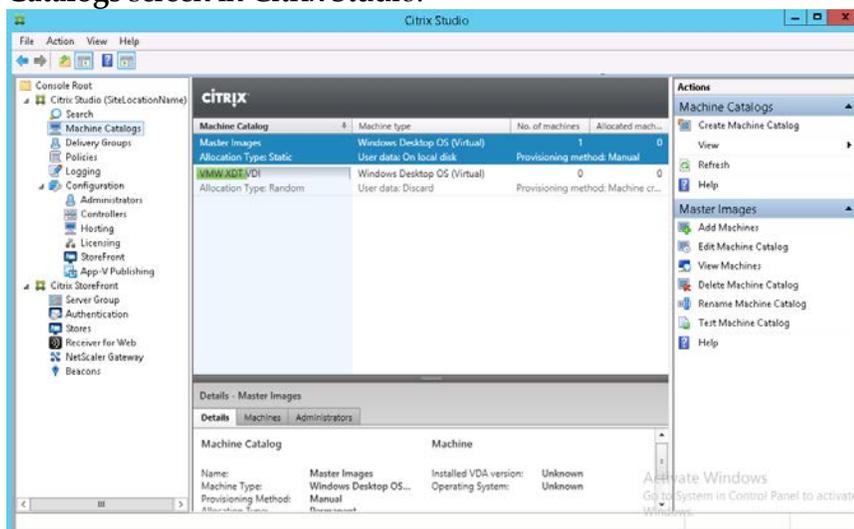
12. Once the machine catalog is created, we go back and re-enable the vGPU profiles in the next step

11.4 CONFIGURING VIRTUAL MACHINE TO ENABLE VGPU PROFILES

Perform the following procedure to each VM provisioned in either a Random or Static Machine Catalog. 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 mouse is not responsive and usable with vGPU enabled.

1. In the vSphere Web Client right click on a VM in the Catalog. Select Edit Setting to open the VM Settings dialogue.
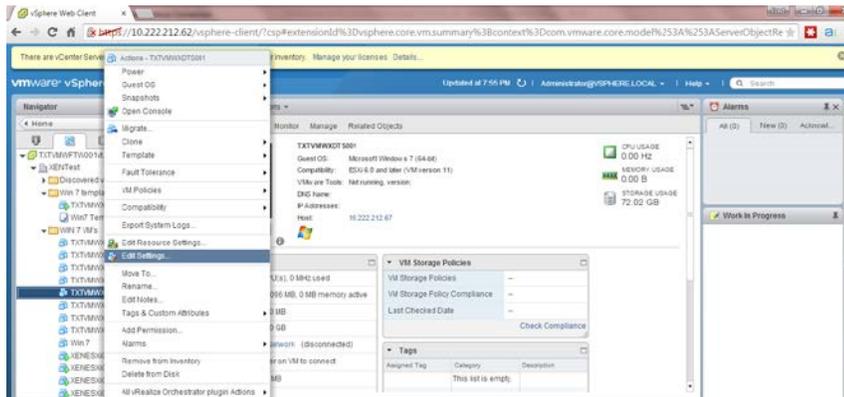
2. On the VM **Edit Settings** dialogue click on the **New device** drop down menu at the bottom and select **Share PCI device**.
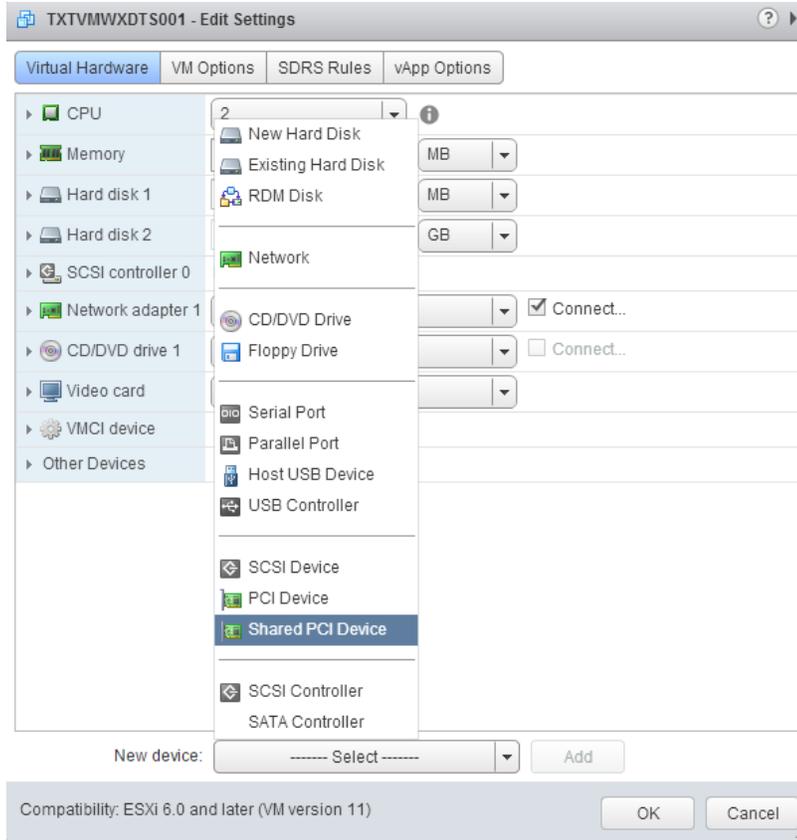
3. Select **Add** button to add the **Shared PCI Device** to the VM.
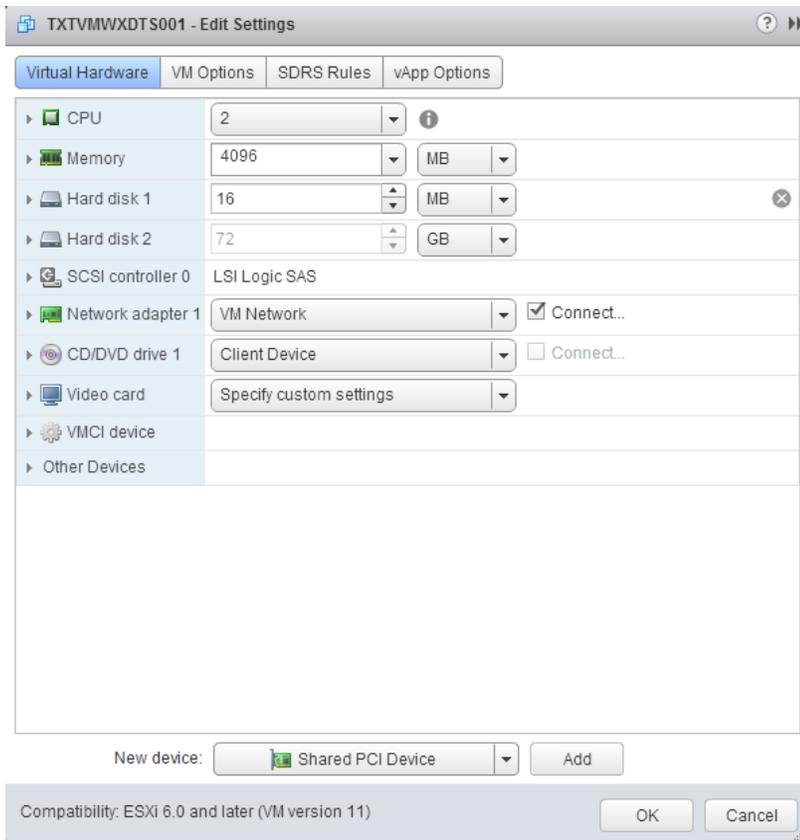
4. Select the **GPU Profile** for this VM and then Select **OK** to close the dialogue and
    save the VM.
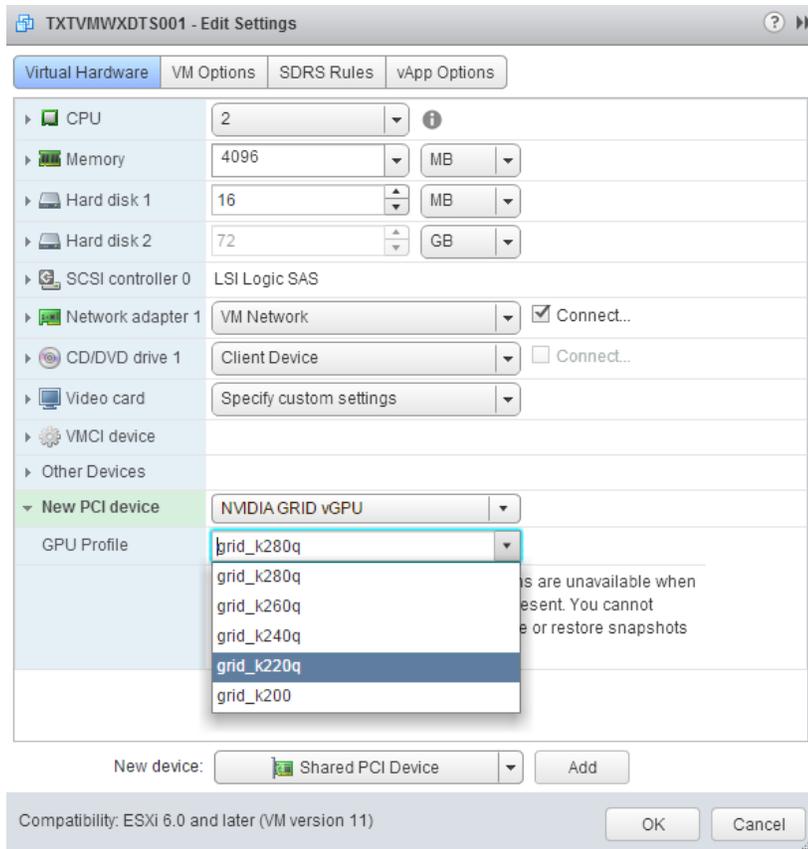
5. Repeat for all VM's provisioned on the Machine Catalog.

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**.
2. Right-click on **Delivery Groups** from the left hand pane and select the **Create Delivery Group** menu item.
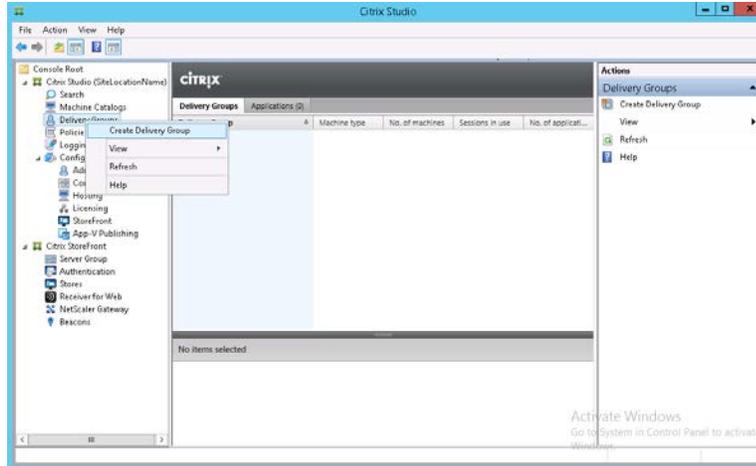
3. On the **Create Delivery Group** dialogue select the **Next** button to continue.
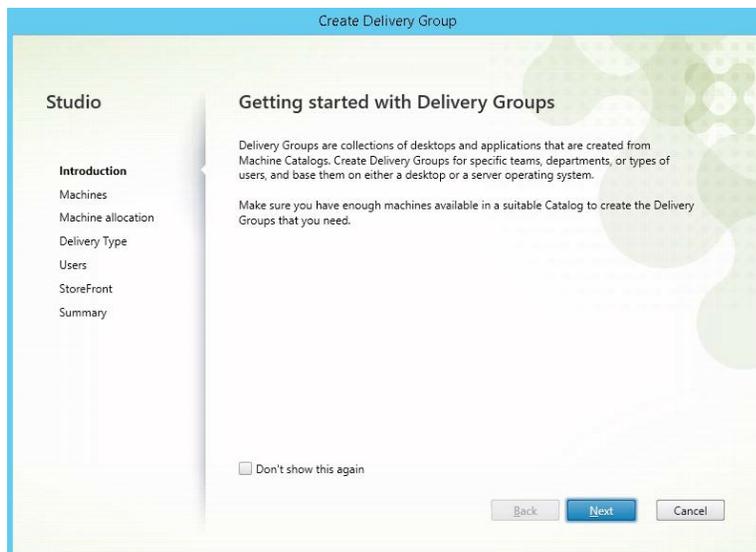
4. On the **Machines** dialogue select the master image catalog that we create earlier. Select the **Next** button to continue.
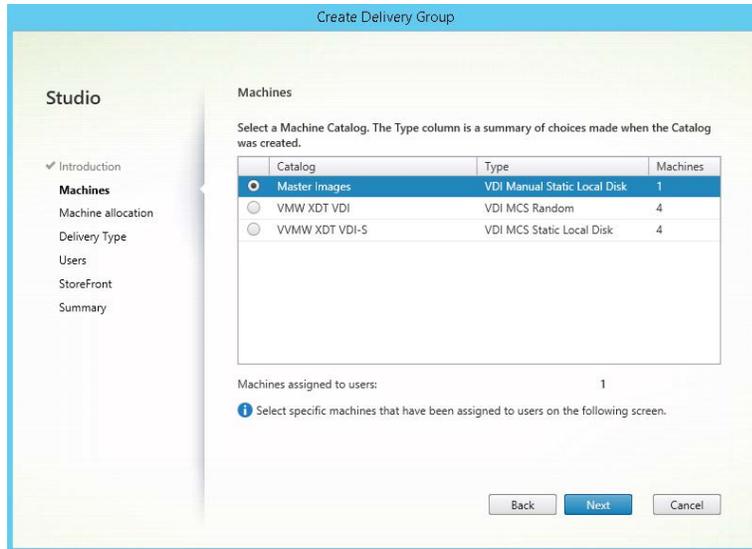
5. On the **User Assignments** dialogue optionally assign users to machines and then select the **Next** button to continue.
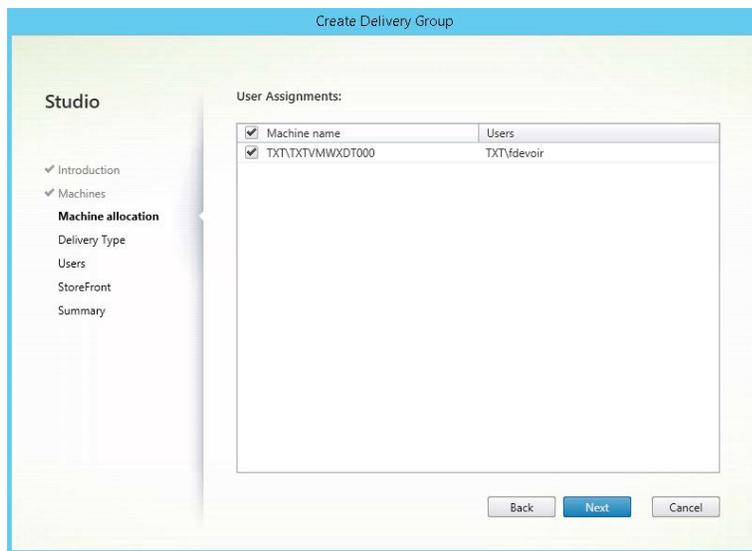
6. On the **Delivery Type** dialogue accept the default selection of **Desktops** radio button and then select the **Next** button to continue.
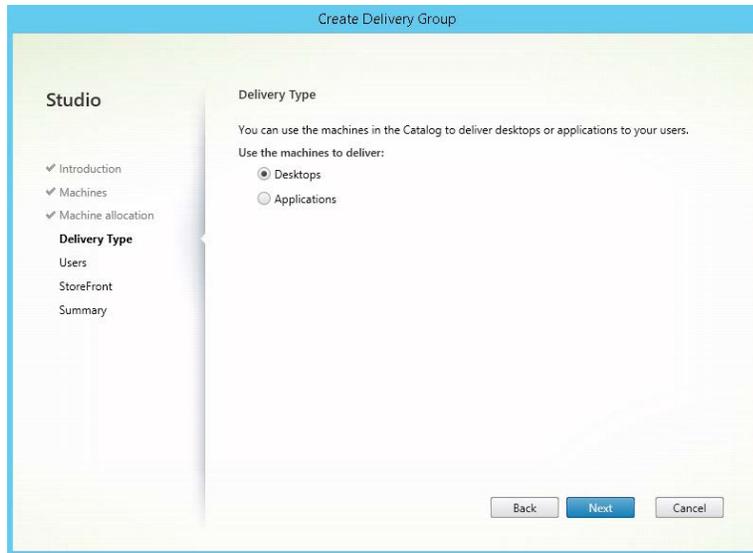
7. On the **Users** dialogue select the **Add** button to use the **Active Directory Select Users and Groups** to grant users or groups access to the delivery group.
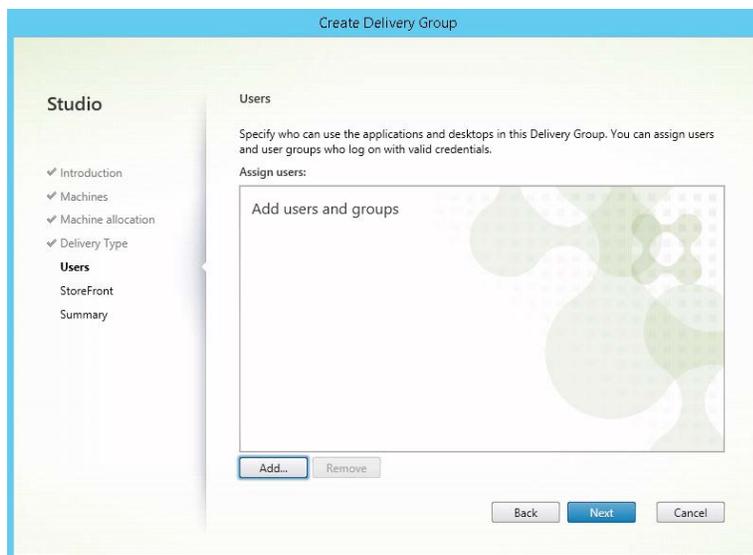
8. On the **Select Users and Groups** dialogue search for and add the users and groups that will be granted access to the delivery group. Select the **OK** button to continue.

9. On the **Users** dialogue select the **Next** button to continue.
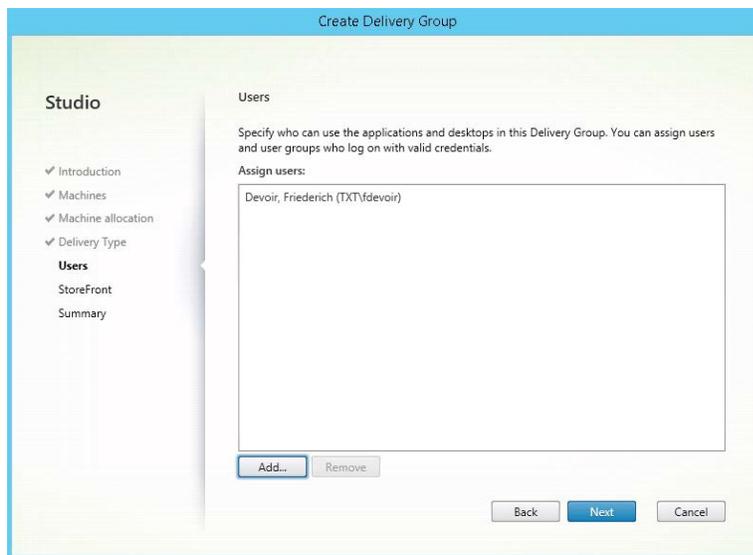
10. On the **StoreFront** dialogue accept the default to manually configure the StoreFront address on the VM installed Citrix Receiver client. Select the **Next** button to continue.
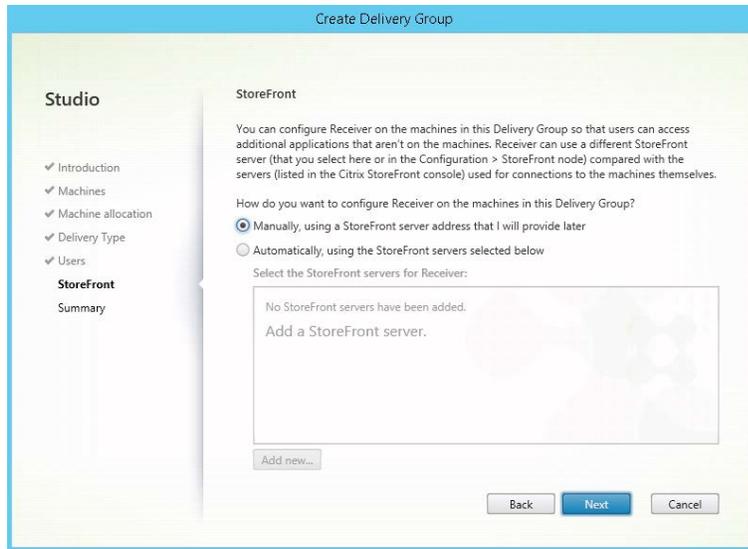
11. On the **Summary** dialogue enter a unique **Delivery Group** name, and **Display Name** for the delivery group. Select the **Finish** button to complete the creation of the delivery group.
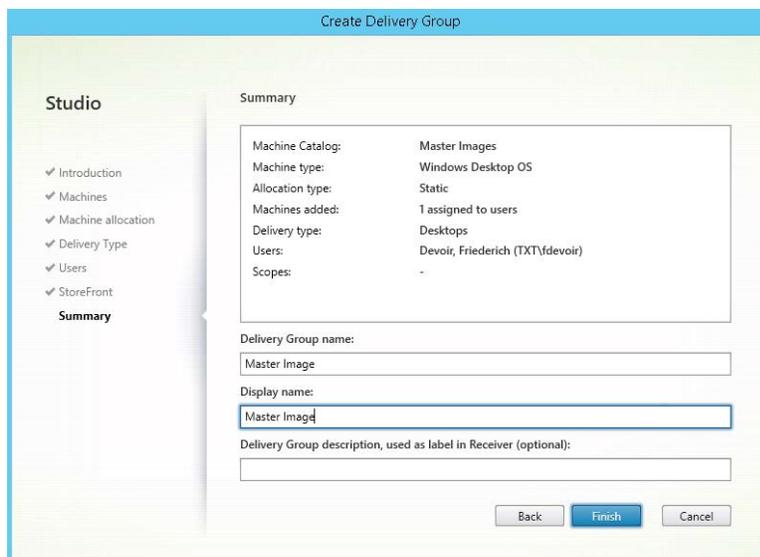
12. On the **Delivery Groups** center pane the newly created delivery group should now be available. Double-click on the **Delivery Group** row on the center pane to view the machines that are available for delivery.

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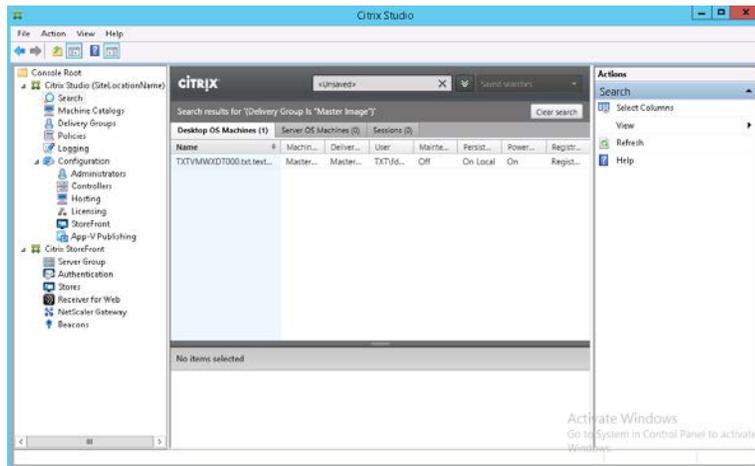

## 11.6 CREATING A XENDESKTOP DELIVERY GROUP FOR THE POOLED MACHINE CATALOG

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.
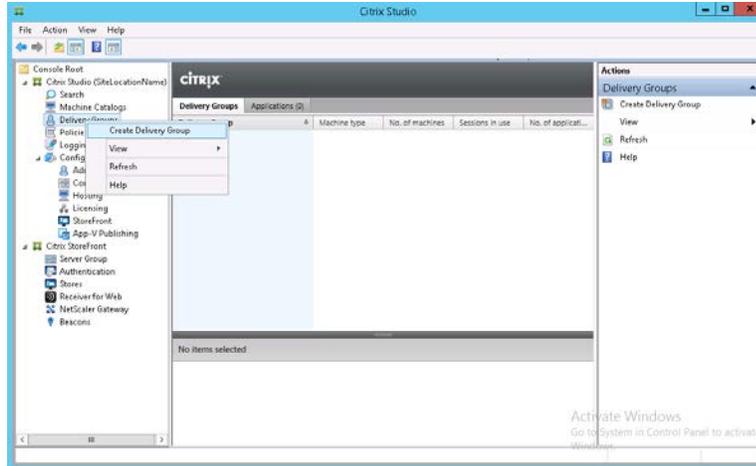
2. On the **Create Delivery Group** dialogue select the **Next** button to continue.
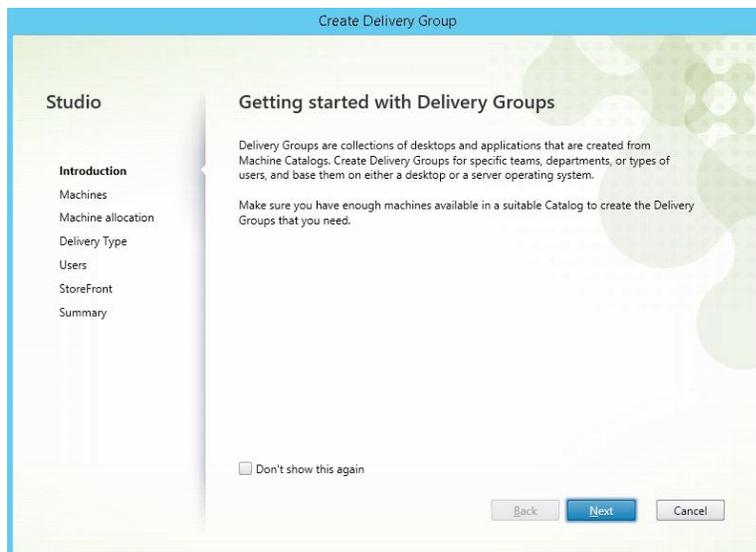
3. On the **Machines** dialogue select the random pool catalog that we create earlier. Select the **Next** button to continue.
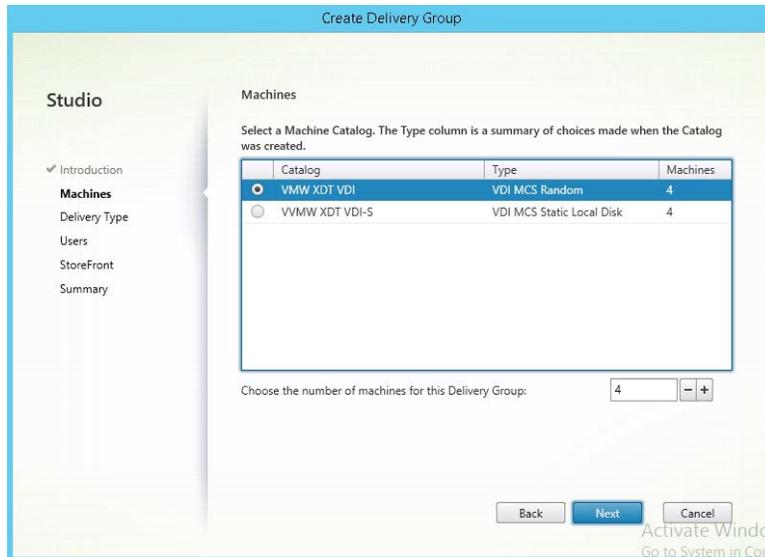
4. On the **Delivery Type** dialogue accept the default selection of **Desktops** radio button and then select the **Next** button to continue.
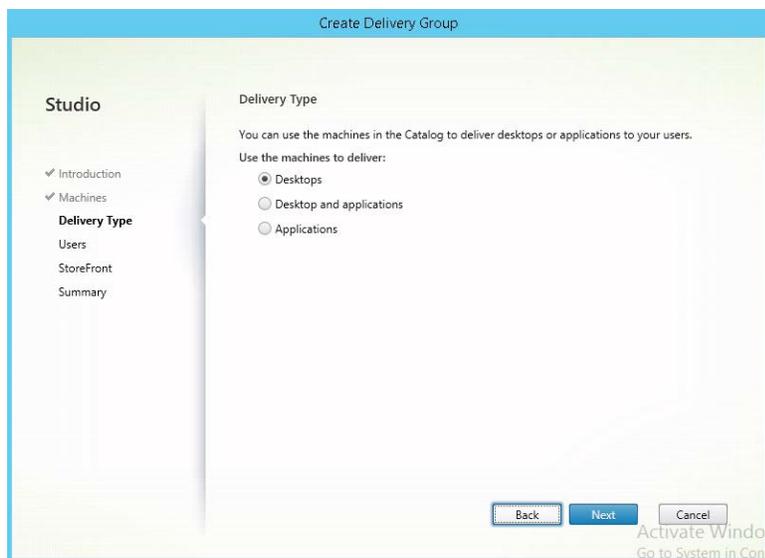
5. On the **Users** dialogue select the **Add** button to use the **Active Directory Select Users and Groups** to grant users or groups access to the delivery group.

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

7. On the **Users** dialogue select the **Next** button to continue.

8. On the **StoreFront** dialogue accept the default to manually configure the StoreFront address on the VM installed Citrix Receiver client. Select the **Next** button to continue.
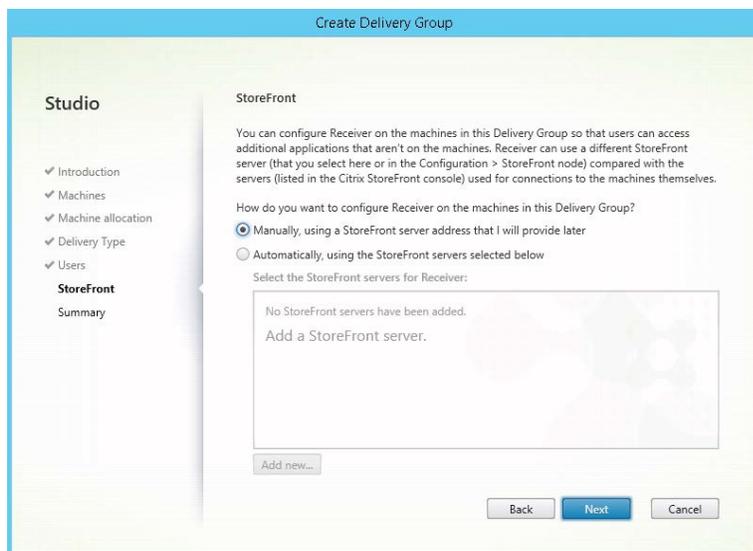
9. On the **Summary** dialogue enter a unique **Delivery Group** name, and **Display Name** for the delivery group. Select the **Finish** button to complete the creation of the delivery group.

10. 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.

14. 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 login to the XenDesktop server and open Citrix Studio. On the Left-hand pane select the **Receiver for Web** menu option. Note the **Website URL** displayed on the center pane.
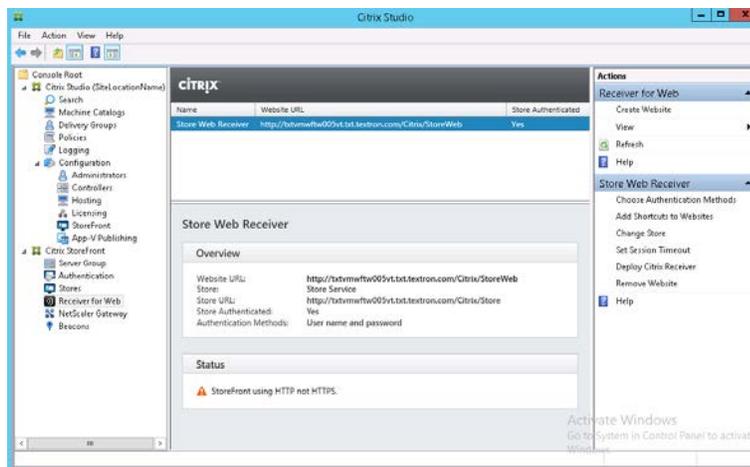
2. Open the Receiver for Web URL in a web browser. If Citrix Receiver client is not installed a prompt will be displayed to **Install** the client. Select the **I agree with the Citrix license agreement** check box. Select **Install** to begin

installation of the Citrix Receiver client. Alternately select the **Log on** link to skip the installation and continue to **Receiver for Web.**
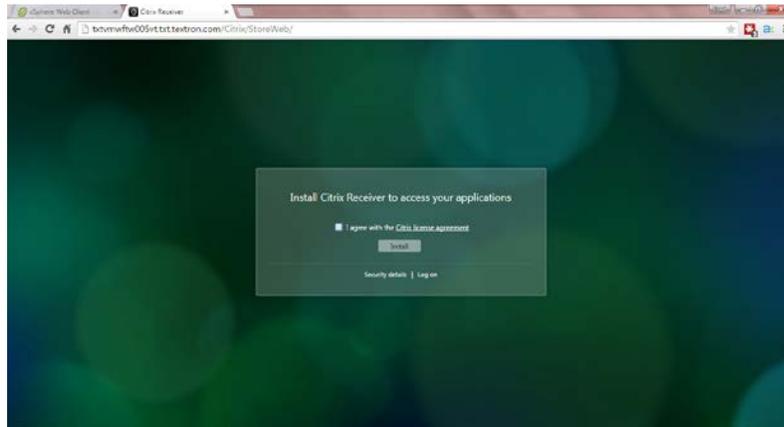
3. Enter **User name** domain account and **password**, then select **Log On** to continue.
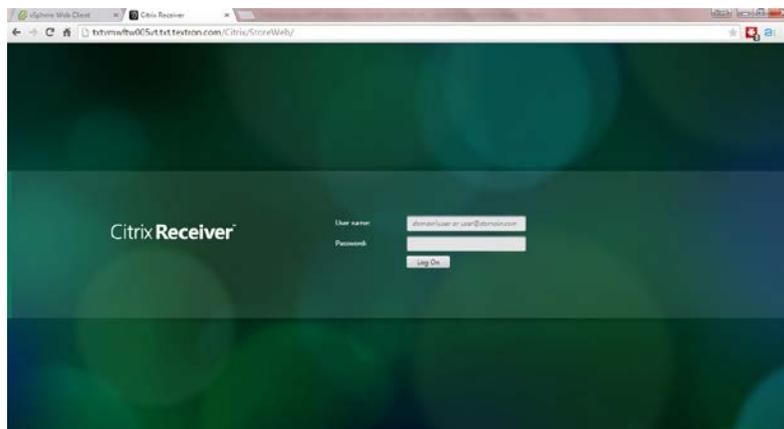
4. The Delivery Group resources that the logged in user has been granted access to will be displayed on the **Receiver for Web** catalog page. Select **the Master Image** icon, to test that XenDesktop functionality is configured properly.
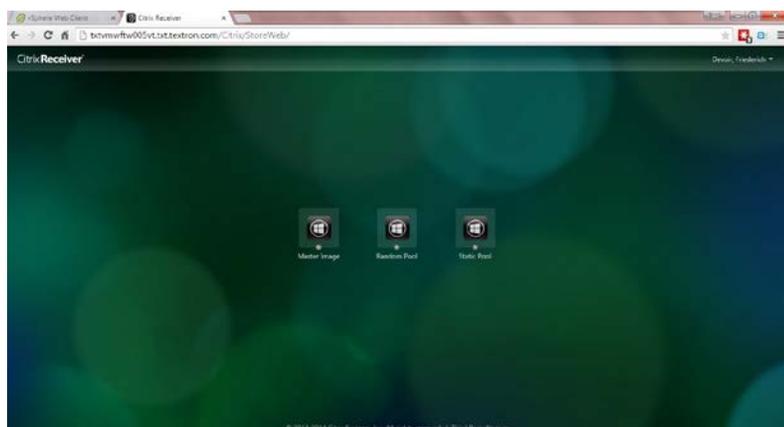
5. The **Desktop Viewer** will display with a connecting message.
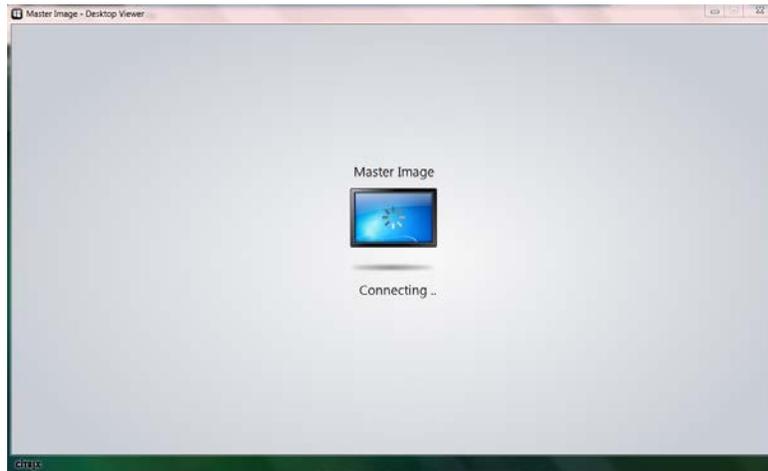
6. When the image is fully loaded the **Windows Desktop** will be displayed.

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

8. 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.

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

10. The NVIDIA Control Panel will indicate which of the GRID Profiles is loaded.
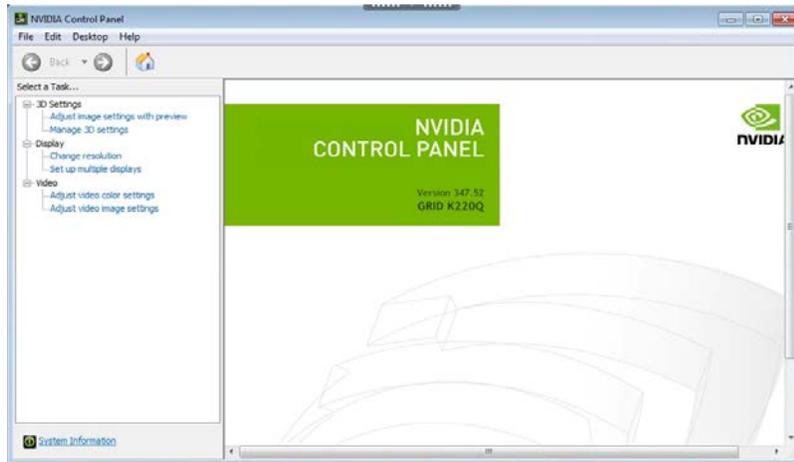
11. Repeat these steps to validate other delivery group configurations.

Chapter 12. TESTING, MONITORING, & BENCHMARKING

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

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

12.1 TESTING VIRTUAL DESKTOPS VS PHYSICAL ONES, AN IMPORTANT DIFFERENCE!

Its important to note that virtualization is a shared environment, allowing for efficiencies and administrative ease that are impossible with bare metal installs across the enterprise. Where a physical workstation or PC is one solution that meets a variety of business requirements, virtualizing those same workloads is a very different solution. As a result the testing and benchmarking must be looked at in a similar light. Where the goal on bare metal is to maximize performance at that desktop, with virtual desktops the goal is to maximize performance not only at the desktop, but also all the way out to whatever device over whatever bandwidth the end user is connecting with.

12.2   NVIDIA SYSTEM MANAGEMENT INTERFACE

The NVIDIA System Management Interface (NSMI) 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 *NSMI* documentation and see Chapter 14, Troubleshooting, on page 193 for more information.

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

NSMI ships with the graphics drivers for the following operating systems:

▶ 64-bit Windows Server 2008R2

▶ 64-bit Windows 7

▶ Linux

12.2.1   Querying NSMI

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 single virtual machine **Win7-c01** is utilizing a GRID K2 with the **K240Q vGPU** profile, and the GPU utilization for the GPU that is hosting the vGPU profile is **7%**.

> 💬 The NSMI documentation is available at
> http:// developer.download.nvidia.com/compute/cuda/6_0/rel/gdk/nvidia-smi.331.38.pdf.

```
+------------------------------------------------------------------+
| NVIDIA-SMI 340.49     Driver Version: 340.49   |
|------------- -------------+----------------+---------------------+
|GPU Name   Persistence-M | Bus-Id Disp.A |      Volatile Uncorr. ECC |
|Fan Temp Perf Pwr:Usage/Cap|  Memory-Usage |  GPU-Util    Compute M.|

|===========================+================+=======================|
|    0   GRID K2  Off|0000:09:00.0 Off|       Off |
|N/A 52C    P8    30W / 117W |917MiB/4095MiB | 7%    Default |
```

```
+--------------------------+---------------+----------------------+
|    1  GRID K2  Off|0000:0A:00.0 Off|         Off |
| N/A 43C  P0    50W / 117W |11MiB/4095MiB   |     0%    Default |
+--------------------------+---------------+----------------------+
|    2  GRID K2  Off|0000:85:00.0 Off|         Off |
| N/A 42C  P0    51W / 117W|11MiB /4095MiB  | 0%    Default |
+--------------------------+---------------+----------------------+
|    3  GRID K2  Off|0000:86:00.0 Off|         Off |
| N/A 37C  P0    39W / 117W| 11MiB /4095MiB | 0%    Default |
+--------------------------+---------------+----------------------+


+-------------------------------------------------------------------+
| Compute processes:    GPU Memory |
| GPU PID   Process name Usage |
|===================================================================|
|    0     37336   Win7-c01    896MiB |
+-------------------------------------------------------------------+
```

12.2.2 Monitoring

Adding command switches allows you to use NSMI 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 NSMI example on page 183, 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 **gpuvm** 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 date 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=000:09:00.0 --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 %, 4095 MiB, 915 MiB, 3180 MiB
26 %, 12 %, 4095 MiB, 915 MiB, 3180 MiB
9 %, 1 %, 4095 MiB, 916 MiB, 3179 MiB
17 %, 3 %, 4095 MiB, 916 MiB, 3179 MiB
34 %, 22 %, 4095 MiB, 916 MiB, 3179 MiB
34 %, 23 %, 4095 MiB, 916 MiB, 3179 MiB
39 %, 25 %, 4095 MiB, 916 MiB, 3179 MiB
29 %, 12 %, 4095 MiB, 916 MiB, 3179 MiB
25 %, 12 %, 4095 MiB, 916 MiB, 3179 MiB
55 %, 33 %, 4095 MiB, 916 MiB, 3179 MiB
35 %, 6 %, 4095 MiB, 916 MiB, 3179 MiB
30 %, 5 %, 4095 MiB, 916 MiB, 3179 MiB
37 %, 6 %, 4095 MiB, 916 MiB, 3179 MiB
```

```
15 %, 1 %, 4095 MiB, 916 MiB, 3179 MiB
11 %, 1 %, 4095 MiB, 916 MiB, 3179 MiB
21 %, 1 %, 4095 MiB, 916 MiB, 3179 MiB
25 %, 4 %, 4095 MiB, 916 MiB, 3179 MiB
26 %, 5 %, 4095 MiB, 916 MiB, 3179 MiB
27 %, 5 %, 4095 MiB, 916 MiB, 3179 MiB
16 %, 1 %, 4095 MiB, 916 MiB, 3179 MiB
10 %, 1 %, 4095 MiB, 916 MiB, 3179 MiB
26 %, 12 %, 4095 MiB, 916 MiB, 3179 MiB
28 %, 6 %, 4095 MiB, 916 MiB, 3179 MiB
29 %, 6 %, 4095 MiB, 916 MiB, 3179 MiB
31 %, 7 %, 4095 MiB, 916 MiB, 3179 MiB
33 %, 7 %, 4095 MiB, 916 MiB, 3179 MiB
20 %, 4 %, 4095 MiB, 916 MiB, 3179 MiB
37 %, 7 %, 4095 MiB, 916 MiB, 3179 MiB
24 %, 5 %, 4095 MiB, 916 MiB, 3179 MiB
24 %, 5 %, 4095 MiB, 916 MiB, 3179 MiB
9 %, 1 %, 4095 MiB, 916 MiB, 3179 MiB
51 %, 8 %, 4095 MiB, 916 MiB, 3179 MiB
52 %, 8 %, 4095 MiB, 916 MiB, 3179 MiB
53 %, 9 %, 4095 MiB, 916 MiB, 3179 MiB
56 %, 9 %, 4095 MiB, 916 MiB, 3179 MiB
54 %, 9 %, 4095 MiB, 916 MiB, 3179 MiB
44 %, 8 %, 4095 MiB, 916 MiB, 3179 MiB
44 %, 8 %, 4095 MiB, 916 MiB, 3179 MiB
46 %, 7 %, 4095 MiB, 916 MiB, 3179 MiB
10 %, 1 %, 4095 MiB, 916 MiB, 3179 MiB
29 %, 4 %, 4095 MiB, 917 MiB, 3178 MiB
43 %, 9 %, 4095 MiB, 917 MiB, 3178 MiB
44 %, 9 %, 4095 MiB, 917 MiB, 3178 MiB
45 %, 9 %, 4095 MiB, 917 MiB, 3178 MiB
2 %, 0 %, 4095 MiB, 915 MiB, 3180 MiB
```

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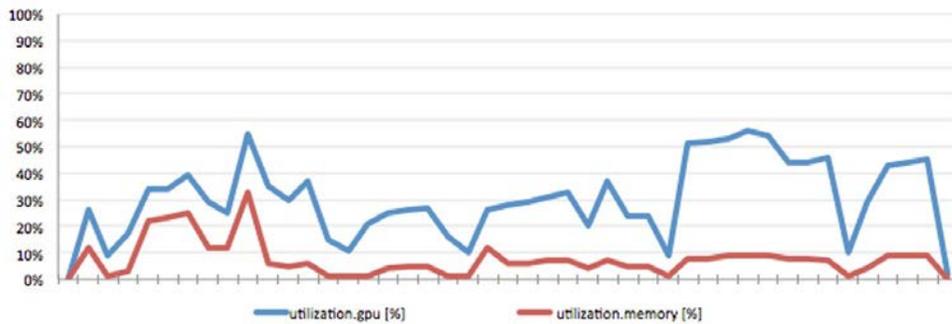

Figure 12-1.   Graphed Log Output

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   Browsers

▶ HTML 5 (Fishbowl): ie.microsoft.com/testdrive

▶ Internet Explorer: http://windows.microsoft.com/en-us/internet-explorer/download-ie

▶ Chrome: chrome.google.com

▶ Firefox: www.mozilla.org/firefox

▶ FaceBook: www.facebook.com

▶ NVIDIA: www.nvidia.com

▶ ESPN: www.espn.com

12.3.2   Microsoft

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

12.3.3   Dassault

▶ SolidWorks: http://www.solidworks.com/sw/purchase/solidworks-trial.htm

▶ eDrawings Horizoner (SolidWorks Horizoner): http://www.edrawingsHorizoner.com/ed/download.htm

▶ Samples (SeaScooter, etc.): http://www.edrawingsHorizoner.com/ed/edrawings-samples.htm

▶ Catia: www.3ds.com/products-services/catia

12.3.4   PTC

▶ Creo: www.ptc.com/product/creo

12.3.5   Siemens

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

12.3.6   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/products/tropics/
  - Sanctuary: http://unigine.com/products/sanctuary/

12.3.7   Lego Designer

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

12.3.8   Adobe

▶ Reader: get.adobe.com/reader

▶ Acrobat: www.adobe.com/Acrobat_XI

▶ Creative Cloud Suite: www.adobe.com/CreativeCloud

▶ PhotoShop (with high-resolution images): www.adobe.com/Photoshop

▶ Illustrator: www.adobe.com/Illustrator

12.3.9   BitSquid Stone Giant

▶ Bitsquid Stone Giant: http://www.bitsquid.se/gallery-stone_giant.html

12.3.10   Google Earth

▶ Locally installed: https://www.google.com/earth/explore/products/desktop.html

▶ Chrome delivered: https://www.google.com/earth/explore/products/plugin.html

12.3.11   Hulu/YouTube videos

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

12.3.12 Redway Turbine 3D

▶ Redway Turbine 3D: http://www.redway3d.com/news/news290609.php

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

▶ Inventor 2015 SP1 Update 2

12.3.14 Slicer3D

▶ Slicer 3D: 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 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
  Reboot Required: false VIBs Installed:
  VIBs Removed: NVIDIA_bootbank_NVIDIA-vgx-
  VMware_vSphere_6_Host_Driver_340.49- 1OEM.600.0.0.1951118 2014-10-06
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-vgx-VMware_vSphere_6_Host_Driver 340.49-1OEM.600.0.0.1951118
NVIDIA VMwareAccepted 2014-10-06
```

NVIDIA WDDM Issues

After installing the NVIDIA WDDM driver in a VM and connecting with Citrix Receiver, the desktop will not be visible from the HostClient console.

▶ Without a console, the vSphere Web Client can still provide the IP address of a VM.

# 13.2 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 of 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 256G of memory, these errors may occur if vGPU-enabled VMs are assigned more than 243G of memory. Reduce the total amount of system memory assigned to the VMs.

13.3   INSTALLATION ON A CISCO UCS240M3

The error message **Error loading /tools.t00 Fatal error: 10 (Out of resources)** may occur when installing vSphere 5.5. Use the following procedure to resolve this issue:

1. Enter the system BIOS by pressing **[F2]**.

2. Select the **Advanced** tab in CIMS.

3. Expand the **PCI Configuration** section.

```
              Aptio Setup Utility - Copyright (C) 2012 American Megatrends, Inc.
        Main  Advanced  Security  Server Management  Boot Options  Save & Exit

                                                    PCI, PCI-X and PCI Express
    ▶ Processor Configuration                        Settings.
    ▶ Memory Configuration
    ▶ QPI Configuration
    ▶ Intel TXT(LT-SX) Configuration
    ▶ South Bridge
    ▶ USB Configuration
    ▶ PCI Configuration
    ▶ ME Subsystem
    ▶ Serial Port Console Redirection
    ▶ Trusted Computing
    ▶ LOM and PCIe Slots Configuration

                                                    ↔    Select Screen
                                                    ↑↓   Select Item
                                                    Enter Select Field
                                                    +/-  Change Value
                                                    F1   General Help
                                                    F9   BIOS Defaults
                                                    F10  Save & Exit
                                                    ESC  Exit

              Version 2.15.1236. Copyright (C) 2012 American Megatrends, Inc.
```

4. Change the **MMCFG BASE** setting from **Auto** to **2GB**.

5. Save your changes and reboot the system.

6. Reinstall vSphere.

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.

▶ Confirm the GPU Installation on page 196

▶ Verify the .VIB Installation on page 197

▶ Confirm .VIB Loading on page 197

▶ Confirm PCI Bus Slot Order on page 198

14.1.1   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.

14.1.1.1 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 a specific host. Now 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:
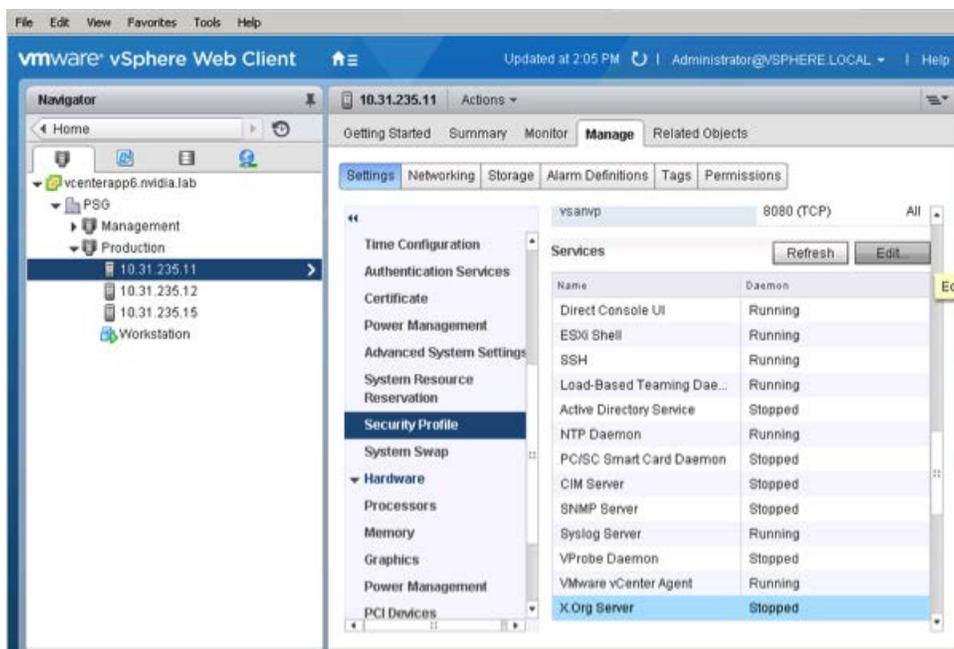
3. The service edit window appears, showing current reported status and allowing you access to starting, stopping, restarting, and managing service behavior on server boot. For a vGPU host you want the service to start and stop with the host. In the example below we show the service as stopped, click Start:
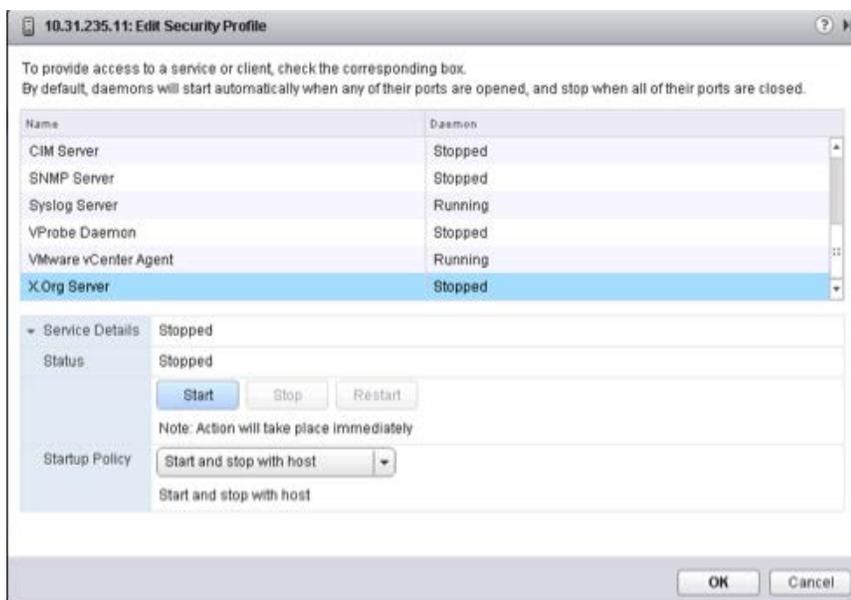
4. The window will refresh and the service should start.
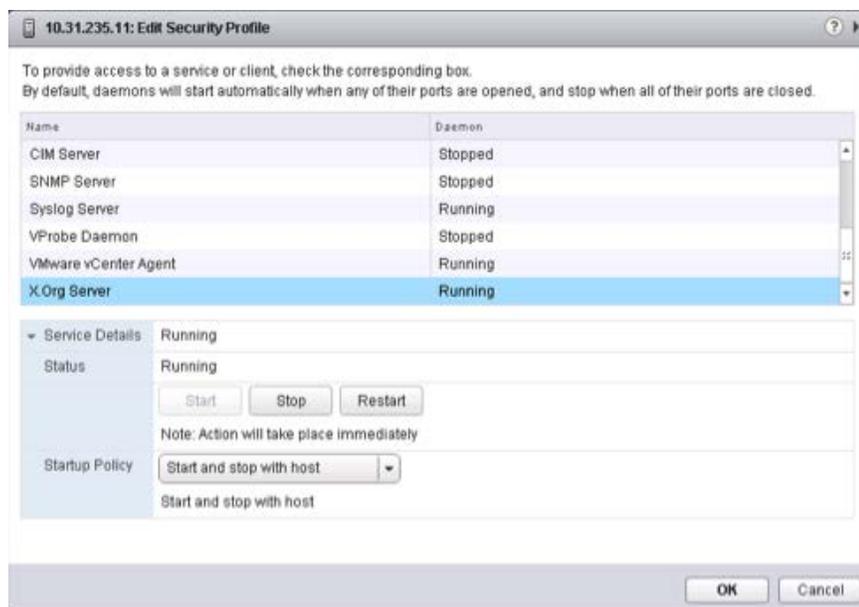
5. Check that your vGPU guests can now start.

▶ Managing X.Org via CLI on page 198

▶ Check NVIDIA-SMI on page 201

14.1.2   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:

```
0000:09:00.0 Display controller: NVIDIA Corporation NVIDIAGRID K2 [vmgfx0]
0000:0a:00.0 Display controller: NVIDIA Corporation NVIDIAGRID K2 [vmgfx1]
0000:0d:00.0 Display controller: Matrox Electronics Systems Ltd. MGA G200e
[Pilot] ServerEngines (SEP1)
0000:85:00.0 Display controller: NVIDIA Corporation NVIDIAGRID K2 [vmgfx2]
0000:86:00.0 Display controller: NVIDIA Corporation NVIDIAGRID K2 [vmgfx3]
```

In this example, the bottom two devices `85:00.0` and `86:00.0` are associated with one of the two K2 cards.

Run the following command to list details about the GPUs listed above:

```
$ esxcli hardware pci list -c 0x0300 -m 0xff
```

You should see output similar to the following for each of the device class `0x300`
`(Display)`:

```
0000:85:00.0
Address: 0000:85:00.0 Segment: 0x0000
Bus: 0x85 Slot: 0x00 Function: 0x0
VMkernel Name: vmgfx2
Vendor Name: NVIDIA Corporation Device Name: NVIDIAGRID K2 Configured Owner:
Unknown Current Owner: VMkernel
Vendor ID: 0x10de Device ID: 0x11bf SubVendor ID: 0x10de SubDevice ID: 0x100a
Device Class: 0x0300
Device Class Name: VGA compatible controller Programming Interface: 0x00
Revision ID: 0xa1 Interrupt Line: 0x0b IRQ: 11
Interrupt Vector: 0x3e PCI Pin: 0x00
Spawned Bus: 0x00 Flags: 0x0201 Module ID: 18 Module Name: nvidia Chassis: 0
Physical Slot: 4294967295
Slot Description: Chassis slot 5; function 0; relative bdf 01:00.0
Passthru Capable: true Parent Device: PCI 0:132:8:0
Dependent Device: PCI 0:133:0:0 Reset Method: Bridge reset
FPT Sharable: true
```

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 like something like this:
```
NVIDIA-vgx-VMware_vSphere_6_Host_Driver 340.49-1OEM.600.0.0.1951118
NVIDIA
VMwareAccepted    2014-10-06
```

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 4 10576
```

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

6. 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 a specific host. Now 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:
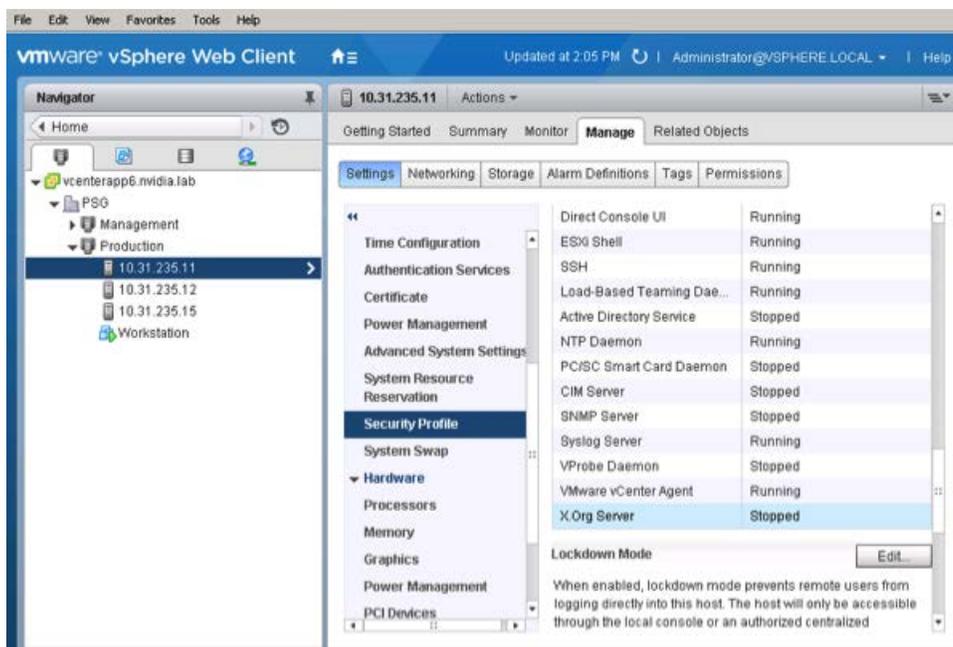
7. In the Services section you can see the current reported status of the X.Org Server service, to change or restart it single click on X.Org Server, then click Edit:
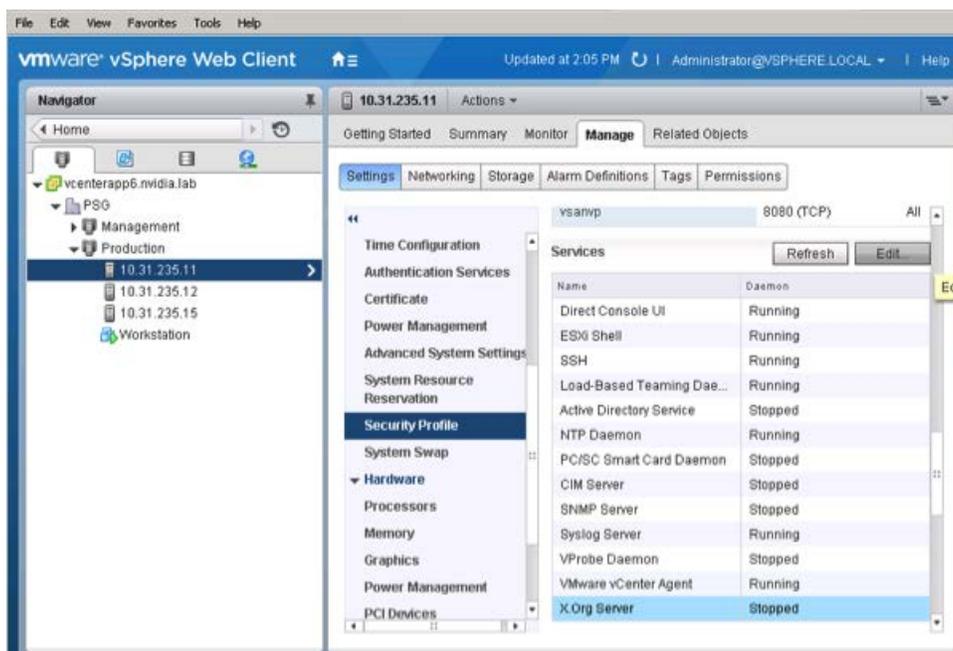
8. The service edit window appears, showing current reported status and allowing you access to starting, stopping, restarting, and managing service behavior on server boot. For a vGPU host you want the service to start and stop with the host. In the example below we show the service as stopped, click Start:
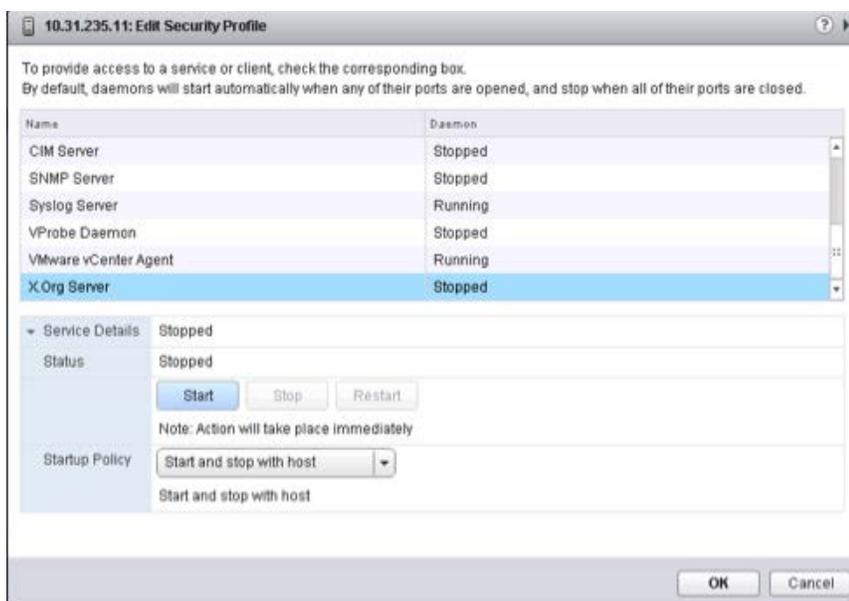
9. The window will refresh and the service should start.
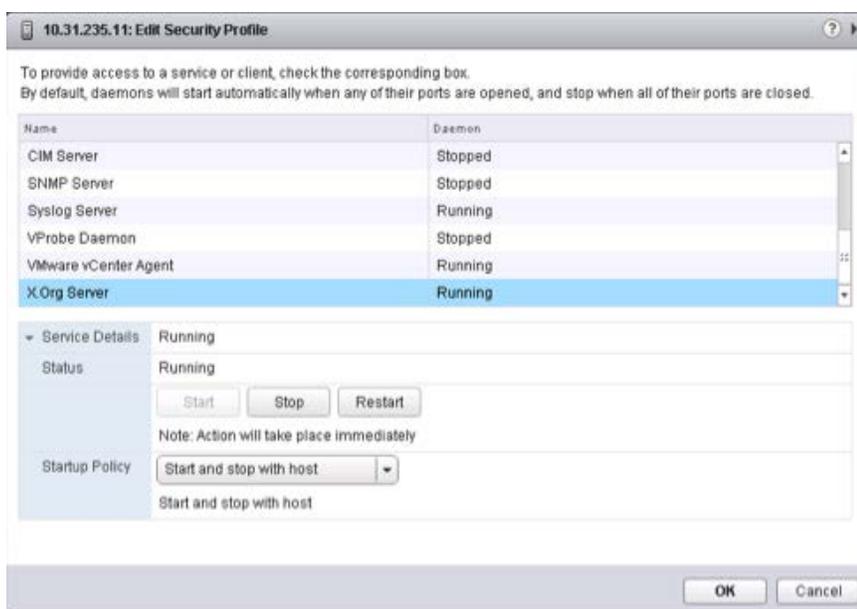
10. Check that your vGPU guests can now start.

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:

```
Xorg0 started
Xorg1 started
Xorg2 started
Xorg3 started
```

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:

```
$ /etc/init.d/xorg start
```

The system should return the following messages:

```
Xorg0 started
Xorg1 started
Xorg2 started
Xorg3 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 (NSMI) 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 NSMI 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:

```
+---------------------------------------------------------------------+
| NVIDIA-SMI 340.49     Driver Version: 340.49   |
|------------- -------------+----------------+----------------------+
|GPU Name   Persistence-M |  Bus-Id Disp.A |  Volatile Uncorr. ECC |
|Fan Temp Perf Pwr:Usage/Cap| Memory-Usage | GPU-Util Compute M.|
|===========================+================+======================|
|
|     0   GRID K2
|N/A 52C    P8     Off|0000:09:00.0 Off|
30W / 117W |917MiB/4095MiB   |
7%    Off |
Default |
+--------------------------+----------------+----------------------+
|     1   GRID K2   Off|0000:0A:00.0 Off|        Off |
| N/A 43C   P0     50W / 117W |11MiB/4095MiB    |     0%    Default |
+--------------------------+----------------+----------------------+
|     2   GRID K2   Off|0000:85:00.0 Off|        Off |
| N/A 42C   P0     51W / 117W|11MiB /4095MiB  |   0%    Default |
+--------------------------+----------------+----------------------+
|     3   GRID K2   Off|0000:86:00.0 Off|        Off |
| N/A 37C   P0     39W / 117W| 11MiB /4095MiB |   0%    Default |
+--------------------------+----------------+----------------------+


+---------------------------------------------------------------------+
| Compute processes:    GPU Memory |
| GPU PID   Process name Usage |
|=====================================================================|
|    0      37336   Win7-c01    896MiB |
+---------------------------------------------------------------------+
```

> 💬 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.

NSMI 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 NSMI will refresh every five seconds. See the NSMI 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.

14.2   ENABLING VM CONSOLE VNC ACCESS

Enabling VNC support in the virtual machine will negatively impact performance and should only be used for troubleshooting driver or 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.

14.3   FILING A BUG REPORT

When filing a bug 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

▶ NSMI 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 log` 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.

2. Enter the connection details required to connect to the vSphere host.

3. Select **Login** to display the *WinSCP Login* window.
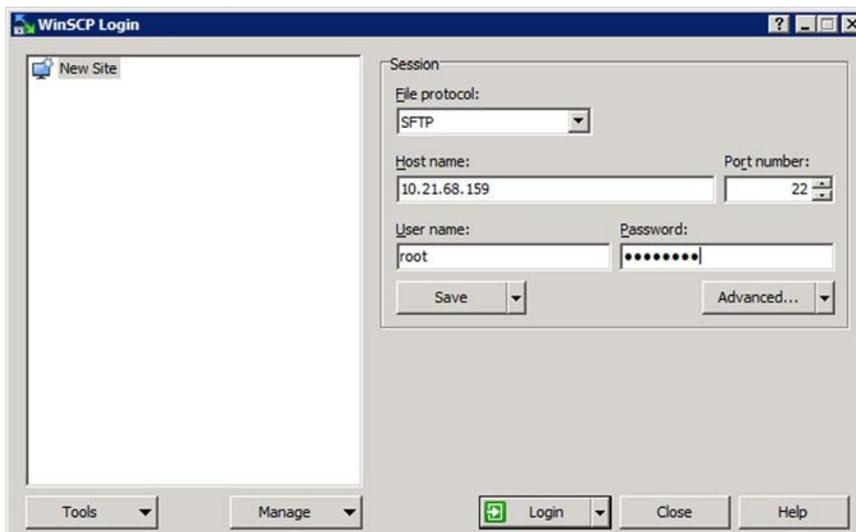
💬 If this is your first time connecting to this server, a warning dialog will appear asking

you to confirm the connection.

4. Navigate to the local folder containing the .VIB file that you want to upload to the vSphere host once the connection is established. Use the left pane of the WinSCP interface.

5. Navigate to a data store you want to upload the .VIB file to using the right pane of the WinSCP interface.
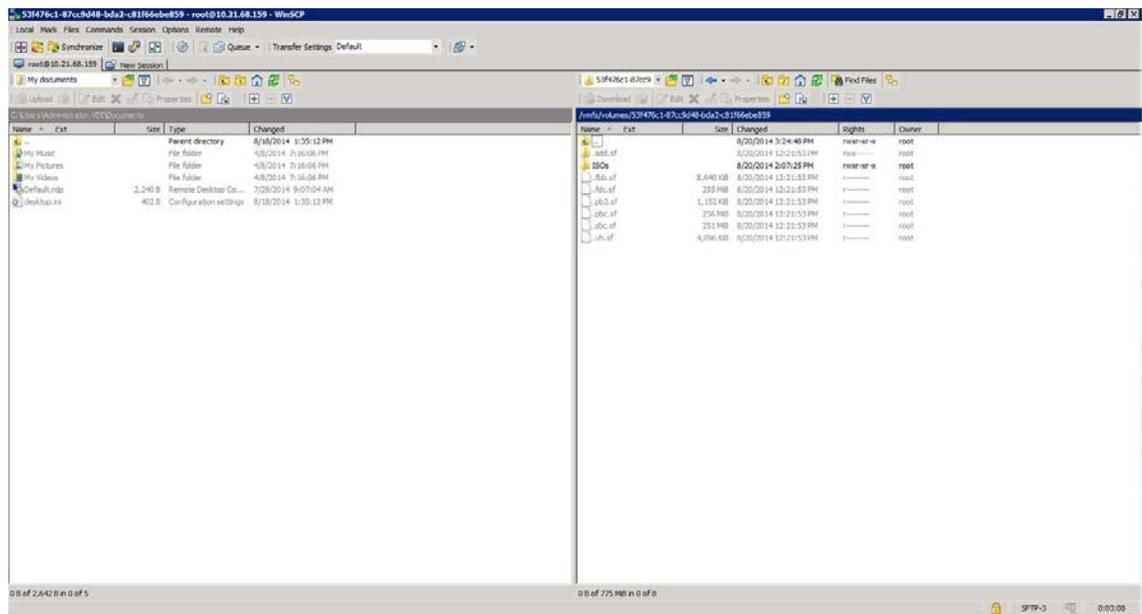
6. Right-click the .VIB file and select Upload to upload the file to the host.