White Paper

Cisco UCS C240-M3 Rack Server with NVIDIA GRID GPU cards on Citrix XenServer 6.2 and XenDesktop 7.5

July 2014

Contents

What You Will Learn

With the increased processor power of today's Cisco UCS B-Series and Cisco UCS C-Series servers, applications with demanding graphics components are now being considered for virtualization. Enhancing the capability to deliver these high-performance, graphics-rich applications are the addition of NVIDIA GRID K1 and K2 cards to the UCS portfolio of PCIe options for our C-Series rack servers.

With the addition of the new graphics processing capabilities, inclusion of the engineering, design, imaging and marketing departments of organizations can now enjoy the benefits that desktop virtualization brings to this space.

This new graphics capability gives organizations the ability to centralize their graphics workloads and data in the datacenter. This provides a very large benefit to organizations that need to be able to work-shift geographically. Until now, that has not been possible because the graphics files are too large to move and the files must be local to the person using them to be usable.

The benefit of PCIe graphics cards in Cisco UCS C-Series servers is four fold:

- Support for full length, full power NVIDIA GRID cards in a 2U form factor
- Cisco UCS Manager integration for management of the servers and GRID cards
- End to end integration with Cisco UCS management suite, including Cisco UCS Central and Cisco UCS Director
- Cisco UCS C240 M3s with two NVIDIA GRID K1 or K2 cards provide more efficient rack space than the two slot, 2.5 equivalent rack unit, HP WS460c workstation blade with the NVIDIA GRID card in a second chassis slot

The purpose of this document is to help partners and customers integrate NVIDIA GRID graphics processor cards with Cisco UCS C240 M3 rack servers on Citrix XenServer 6.2 SP1 and Citrix XenDesktop 7.5 in both pass through and virtual graphics processing unit (vGPU) modes.

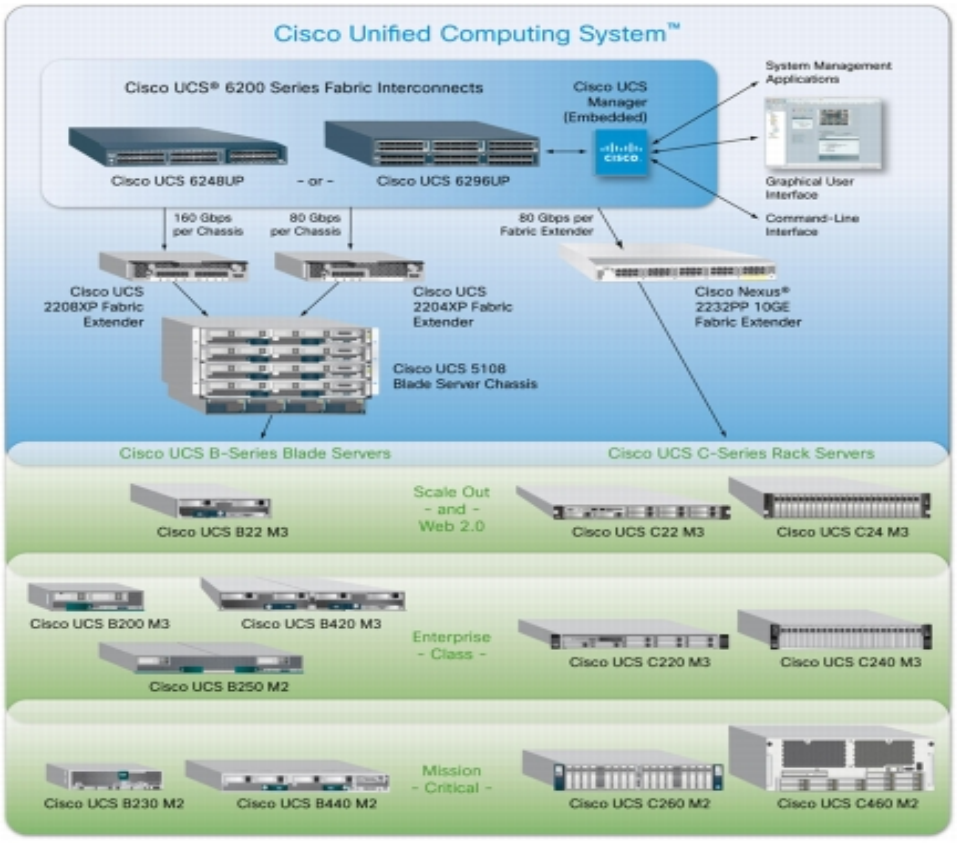
We believe that our partners, NVIDIA and Citrix, are in the best position to provide lists of applications that are supported by the card, their hypervisor and the desktop broker in pass through and vGPU modes.

Our objective is to provide the reader with specific methods for integrating Cisco UCS C240 M3 servers with NVIDIA GRID K1 or K2 cards with Citrix products so that the servers, the hypervisor and the desktop broker are ready for graphic application installation.

Cisco Unified Computing System

The Cisco Unified Computing System is a next-generation data center platform that unites compute, network, and storage access. The platform, optimized for virtual environments, is designed using open industry-standard technologies and aims to reduce total cost of ownership (TCO) and increase business agility. The system integrates a low-latency; lossless 10 Gigabit Ethernet unified network fabric with enterprise-class, x86-architecture servers. It is an integrated, scalable, multi chassis platform in which all resources participate in a unified management domain.

Figure 1. Cisco UCS Components



The main components of Cisco Unified Computing System are:

- **Computing**—The system is based on an entirely new class of computing system that incorporates blade servers based on Intel Xeon E5-2600/4600 and E7-2800 Series Processors.
- **Network**—The system is integrated onto a low-latency, lossless, 10-Gbps unified network fabric. This network foundation consolidates LANs, SANs, and high-performance computing networks which are separate networks today. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables, and by decreasing the power and cooling requirements.
- **Virtualization**—The system unleashes the full potential of virtualization by enhancing the scalability, performance, and operational control of virtual environments. Cisco security, policy enforcement, and diagnostic features are now extended into virtualized environments to better support changing business and IT requirements.
- **Storage access**—The system provides consolidated access to both SAN storage and Network Attached Storage (NAS) over the unified fabric. By unifying the storage access the Cisco Unified Computing System can access storage over Ethernet, Fibre Channel, Fibre Channel over Ethernet (FCoE), and iSCSI. This provides customers with choice for storage access and investment protection. In addition, the server administrators can pre-assign storage-access policies for system connectivity to storage resources, simplifying storage connectivity, and management for increased productivity.

- **Management**—The system uniquely integrates all system components which enable the entire solution to be managed as a single entity by the Cisco UCS Manager. The Cisco UCS Manager has an intuitive graphical user interface (GUI), a command-line interface (CLI), and a robust application programming interface (API) to manage all system configuration and operations.

The Cisco Unified Computing System is designed to deliver:

- A reduced Total Cost of Ownership and increased business agility.
- Increased IT staff productivity through just-in-time provisioning and mobility support.
- A cohesive, integrated system which unifies the technology in the data center. The system is managed, serviced and tested as a whole.
- Scalability through a design for hundreds of discrete servers and thousands of virtual machines and the capability to scale I/O bandwidth to match demand.
- Industry standards supported by a partner ecosystem of industry leaders.

Cisco UCS Manager

Cisco UCS Manager provides unified, embedded management of all software and hardware components of the Cisco Unified Computing System through an intuitive GUI, a command line interface (CLI), or an XML API. The Cisco UCS Manager provides unified management domain with centralized management capabilities and controls multiple chassis and thousands of virtual machines.

Cisco UCS Fabric Interconnect

The Cisco® UCS 6200 Series Fabric Interconnect is a core part of the Cisco Unified Computing System, providing both network connectivity and management capabilities for the system. The Cisco UCS 6200 Series offers line-rate, low-latency, lossless 10 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE) and Fibre Channel functions.

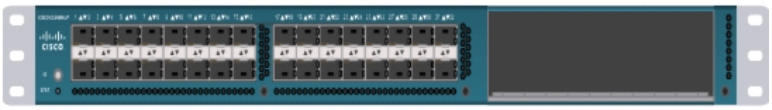
The Cisco UCS 6200 Series provides the management and communication backbone for the Cisco UCS B-Series Blade Servers and Cisco UCS 5100 Series Blade Server Chassis. All chassis, and therefore all blades, attached to the Cisco UCS 6200 Series Fabric Interconnects become part of a single, highly available management domain. In addition, by supporting unified fabric, the Cisco UCS 6200 Series provides both the LAN and SAN connectivity for all blades within its domain.

From a networking perspective, the Cisco UCS 6200 Series uses a cut-through architecture, supporting deterministic, low-latency, line-rate 10 Gigabit Ethernet on all ports, 1Tb switching capacity, 160 Gbps bandwidth per chassis, independent of packet size and enabled services. The product family supports Cisco low-latency, lossless 10 Gigabit Ethernet unified network fabric capabilities, which increase the reliability, efficiency, and scalability of Ethernet networks. The Fabric Interconnect supports multiple traffic classes over a lossless Ethernet fabric from a blade server through an interconnect. Significant TCO savings come from an FCoE-optimized server design in which network interface cards (NICs), host bus adapters (HBAs), cables, and switches can be consolidated.

Cisco UCS 6248UP Fabric Interconnect

The Cisco UCS 6248UP 48-Port Fabric Interconnect is a one-rack-unit (1RU) 10 Gigabit Ethernet, FCoE and Fiber Channel switch offering up to 960-Gbps throughput and up to 48 ports. The switch has 32 1/10-Gbps fixed Ethernet, FCoE and FC ports and one expansion slot.

Figure 2. Cisco UCS 6248UP Fabric Interconnect



Cisco UCS Fabric Extenders

Fabric Extenders are zero-management, low-cost, low-power consuming devices that distribute the system's connectivity and management planes into rack and blade chassis to scale the system without complexity. Designed never to lose a packet, Cisco fabric extenders eliminate the need for top-of-rack Ethernet and Fibre Channel switches and management modules, dramatically reducing infrastructure cost per server.

Cisco UCS 2232PP Fabric Extender

The Cisco Nexus® 2000 Series Fabric Extenders comprise a category of data center products designed to simplify data center access architecture and operations. The Cisco Nexus 2000 Series uses the Cisco® Fabric Extender architecture to provide a highly scalable unified server-access platform across a range of 100 Megabit Ethernet, Gigabit Ethernet, 10 Gigabit Ethernet, unified fabric, copper and fiber connectivity, rack, and blade server environments. The platform is ideal to support today's traditional Gigabit Ethernet while allowing transparent migration to 10 Gigabit Ethernet, virtual machine-aware unified fabric technologies.

The Cisco Nexus 2000 Series Fabric Extenders behave as remote line cards for a parent Cisco Nexus switch or Fabric Interconnect. The fabric extenders are essentially extensions of the parent Cisco UCS Fabric Interconnect switch fabric, with the fabric extenders and the parent Cisco Nexus switch together forming a distributed modular system. This architecture enables physical topologies with the flexibility and benefits of both top-of-rack (ToR) and end-of-row (EoR) deployments.

Today's data centers must have massive scalability to manage the combination of an increasing number of servers and a higher demand for bandwidth from each server. The Cisco Nexus 2000 Series increases the scalability of the access layer to accommodate both sets of demands without increasing management points within the network.

Figure 3. Cisco UCS 2232PP Fabric Extender

Cisco UCS C-Series Rack Servers

Cisco UCS C-Series Rack-Mount Servers keep pace with Intel Xeon processor innovation by offering the latest processors with an increase in processor frequency and improved security and availability features. With the increased performance provided by the Intel Xeon processor E5-2600 and E5-2600 v2 product families, Cisco UCS C-Series servers offer an improved price-to-performance ratio; extend Cisco Unified Computing System innovations to an industry standard rack-mount form factor, including a standards-based unified network fabric, Cisco VN-Link virtualization support, and Cisco Extended Memory Technology.

Designed to operate both in standalone environments and as part of the Cisco Unified Computing System, these servers enable organizations to deploy systems incrementally—using as many or as few servers as needed—on a schedule that best meets the organization's timing and budget. Cisco UCS C-Series servers offer investment protection through the capability to deploy them either as standalone servers or as part of the Cisco Unified Computing System.

One compelling reason that many organizations prefer rack-mount servers is the wide range of I/O options available in the form of PCI Express (PCIe) adapters. Cisco UCS C-Series servers supports spectrum of I/O options, which includes interfaces supported by Cisco as well as adapters from third parties.

Cisco UCS C240 M3 Rack Server

The Cisco UCS C240 M3 Rack Server (Figure X1) is designed for both performance and expandability over a wide range of storage-intensive infrastructure workloads, from big data to collaboration. The enterprise-class Cisco UCS C240 M3 server further extends the capabilities of the Cisco UCS portfolio in a 2RU form factor with the addition of the Intel® Xeon processor E5-2600 and E5-2600 v2 product families, which deliver an outstanding combination of performance, flexibility, and efficiency gains. The Cisco UCS C240 M3 offers up to two Intel Xeon processor E5-2600 or E5-2600 v2 processors, 24 DIMM slots, 24 disk drives, and four 1 Gigabit Ethernet LAN-on-motherboard (LOM) ports to provide exceptional levels of internal memory and storage expandability and exceptional performance.

The Cisco UCS C240 M3 interfaces with the Cisco UCS Virtual Interface Card. The Cisco UCS Virtual Interface Card is a virtualization-optimized Fibre Channel over Ethernet (FCoE) PCI Express (PCIe) 2.0 x8 10-Gbps adapter designed for use with Cisco UCS C-Series Rack Servers. The VIC is a dual-port 10 Gigabit Ethernet PCIe adapter that can support up to 256 PCIe standards-compliant virtual interfaces, which can be dynamically configured so that both their interface type (network interface card [NIC] or host bus adapter [HBA]) and identity (MAC address and worldwide name [WWN]) are established using just-in-time provisioning. In addition, the Cisco UCS VIC 1225 can support network interface virtualization and Cisco® Data Center Virtual Machine Fabric Extender (VM-FEX) technology. An additional five PCIe slots are made available for certified third party PCIe cards. The server is equipped to handle 24 on board SAS drives or SSDs along with shared storage solutions offered by our partners.

Cisco UCS C240 M3 server's disk configuration delivers balanced performance and expandability to best meet individual workload requirements. With up to 12 LFF (Large Form Factor) or 24 SFF (Small Form Factor) internal drives, the Cisco UCS C240 M3 optionally offers 10,000-RPM and 15,000-RPM SAS drives to deliver a high number of I/O operations per second for transactional workloads such as database management systems. In addition, high-capacity SATA drives provide an economical, large-capacity solution. Superfast SSDs are a third option for workloads that demand extremely fast access to smaller amounts of data. A choice of RAID controller options also helps increase disk performance and reliability.

The Cisco UCS C240 M3 further increases performance and customer choice over many types of storage-intensive applications such as:

- Collaboration
- Small and medium-sized business (SMB) databases
- Big data infrastructure
- Virtualization and consolidation
- Storage servers
- High-performance appliances

http://www.cisco.com/en/US/prod/collateral/ps10265/ps10493/ps12370/data_sheet_c78-700629.html

Figure 4. Cisco UCS C240 M3 Rack Server front view

Figure 5. Cisco UCS C240 M3 Rack Server Rear view with PCIe slot number

Table 1. Cisco UCS C240 M3 PCIe Slot Specification

PCIe Slot	Length	Lane
1	half	x8
2	half	**x16**
3	half	x8
4	3/4	x8
5	3/4	**x16**

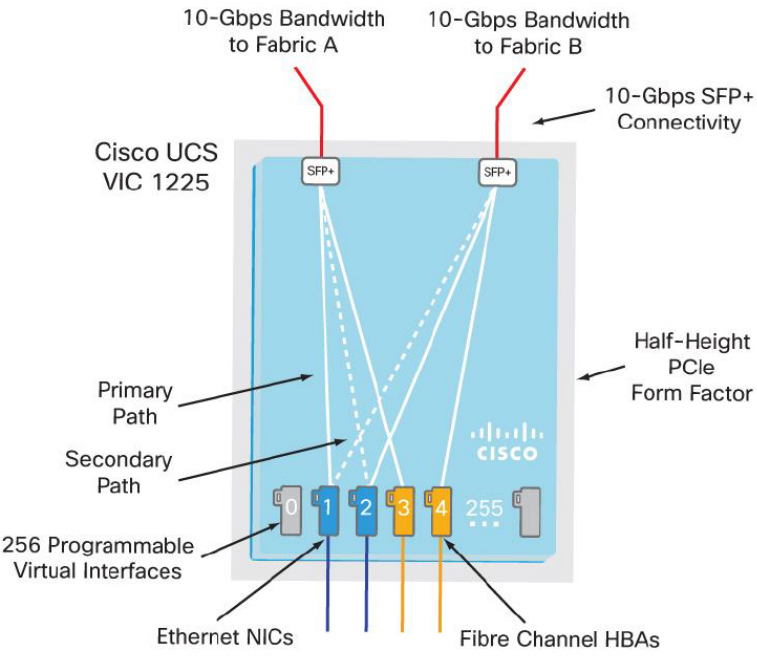
Cisco VIC 1225—10GE Option

A Cisco UCS Virtual Interface Card (VIC) 1225 (Figure 6) is a dual-port Enhanced Small Form-Factor Pluggable (SFP+) 10 Gigabit Ethernet and Fibre Channel over Ethernet (FCoE)-capable PCI Express (PCIe) card designed exclusively for Cisco UCS C-Series Rack Servers. With its half-height design, the card preserves full-height slots in servers for third-party adapters certified by Cisco. It incorporates next-generation converged network adapter (CNA) technology. The card enables a policy-based, stateless, agile server infrastructure that can present up to

256 PCIe standards-compliant interfaces to the host that can be dynamically configured as either network interface cards (NICs) or host bus adapters (HBAs) (Figure 7.) In addition, the Cisco UCS VIC 1225 supports Cisco Data Center Virtual Machine Fabric Extender (VM-FEX) technology, which extends the Cisco UCS fabric interconnect ports to virtual machines, simplifying server virtualization deployment.

Figure 6. Cisco UCS VIC 1225 CAN

Figure 7. Cisco UCS VIC 1225 CNA Architecture



Due to current limitation of UCS C240M3 platform, if you intend to use two NVIDIA GPU cards on the Cisco UCS C240 M3, you must choose the Emulex OneConnect OCe11102-F CNA card rather than the Cisco VIC 1225.

For this study we used OCe11102-F CNA.

Emulex OneConnect OCe11102-F CNA

The Emulex OneConnect® OCe11102-F is a dual-port 10Gb Ethernet (10GbE) PCIe 2.0 x8 adapter that consolidates network and storage traffic with high performance CPU offloads for Fibre Channel over Ethernet (FCoE) and Internet Small Computer System Interface (iSCSI) protocols (Figure 8).

A member of the Emulex OneConnect Universal Converged Network Adapter (UCNA) family, the OCe11102-F adapter supports a common infrastructure for networking and storage, reducing capital expenditures (CAPEX) for adapters, switches and cables, and operational expenditures (OPEX) for power, cooling and IT administration.

The OneConnect OCe11102-F supports protocol offloads for FCoE, iSCSI, TCP/IP and TCP Chimney to provide maximum bandwidth with minimum use of CPU resources. For virtualized server deployments, OneConnect protocol offloads enable more virtual machines (VMs) per server, providing greater cost saving to optimize return on investment.

Figure 8. Emulex CNA OCe11102-F

NVIDIA GRID cards

Network Delivered GPU Acceleration for Virtual Desktops

The NVIDIA GRID portfolio of technologies leverages the power of the GPU and the world's best graphics applications to deliver GPU-accelerated applications and games over the network to any user. NVIDIA GRID GPUs are based on the NVIDIA Kepler™ GPU architecture, delivering fast, reliable, energy-efficient performance. This architecture's virtualization capabilities lets multiple users simultaneously share GPUs with ultra-fast streaming display capability that eliminates lag, making a remote data center feel like it's next door. NVIDIA GRID software is a complete stack of GPU virtualization, remoting and session-management libraries that allows multiple users to experience graphics-intensive desktops, applications and games using GPUs. This enables exceptional capture, efficient compression, fast streaming, and low-latency display of high-performance enterprise applications.

Graphics Accelerated Virtual Desktops and Applications

NVIDIA GRID™ technology offers the ability to offload graphics processing from the CPU to the GPU in virtualized environments, allowing the data center manager to deliver true PC graphics-rich experiences to more users for the first time.

Figure 9. NVIDIA GRID K2 Card

Benefits of NVIDIA GRID for IT:

- Leverage industry-leading virtualization solutions, including Citrix, Microsoft, and VMware
- Add your most graphics-intensive users to your virtual solutions
- Improve the productivity of all users

Benefits of NVIDIA GRID for users:

- Highly responsive windows and rich multimedia experiences
- Access to all critical applications, including the most 3D-intensive
- Access from anywhere, on any device

See more at:
http://www.nvidia.com/object/enterprise-virtualization.html#sthash.fByArGhI.dpuf
http://www.nvidia.com/object/nvidia-grid.html#sthash.gFx6sIF7.dpuf

Table 2. NVIDIA GRID K1 and GRID K2 Comparison

	GRID K1	GRID K2
Number of GPUs	4 x entry Kelper GPUs	2 x high-end Kelper GPUs
Total NVIDIA CUDA cores	768	3072
Total memory size	16 GB DDR3	8 GB GDDR5
Max power	130 W	225 W
Board length	10.5	10.5
Board width	4.4	4.4
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
Technical Specifications	GRID K1 Board Specification	GRID K2 Board Specification

[1] NVIDIA GRID™ vGPU™ is only supported on compatible versions of Citrix XenServer. Consult Citrix for compatibility.
[2] Only compatible with VMware vSphere Hypervisor. Consult VMware for compatibility.

GRID vGPU Support for Citrix XenServer

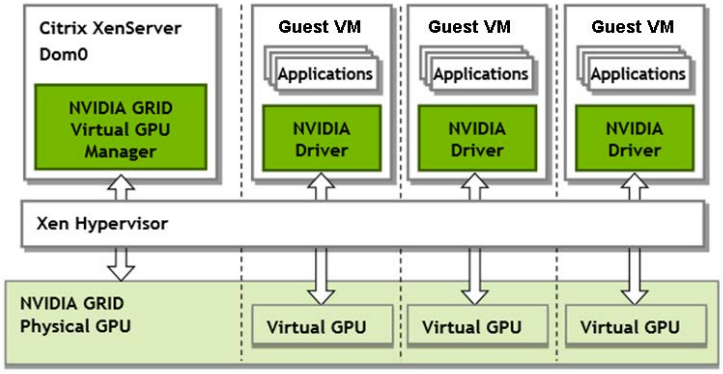
The NVIDIA GRID vGPU enables multiple virtual machines (VMs) to have simultaneous, direct access to a single physical GPU, using the same NVIDIA graphics drivers that are deployed on non-virtualized Operating Systems. By doing this, GRID vGPU provides VMs with unparalleled graphics performance and application compatibility, together with the cost-effectiveness and scalability brought about by sharing a GPU among multiple workloads.

ARCHITECTURE

Figure 10 shows the NVIDIA GRID vGPU high-level architecture is illustrated in the figure below. Under the control of NVIDIA's GRID Virtual GPU Manager running in XenServer dom0, NVIDIA GRID physical GPUs are capable of supporting multiple virtual GPU devices (vGPUs) that can be assigned directly to guest VMs.

Guest VMs use NVIDIA GRID virtual GPUs in the same manner as a physical GPU that has been passed through by the hypervisor: an NVIDIA driver loaded in the guest VM provides direct access to the GPU for performance-critical fast paths, and a paravirtualized interface to the NVIDIA GRID Virtual GPU Manager is used for non-performant management operations.

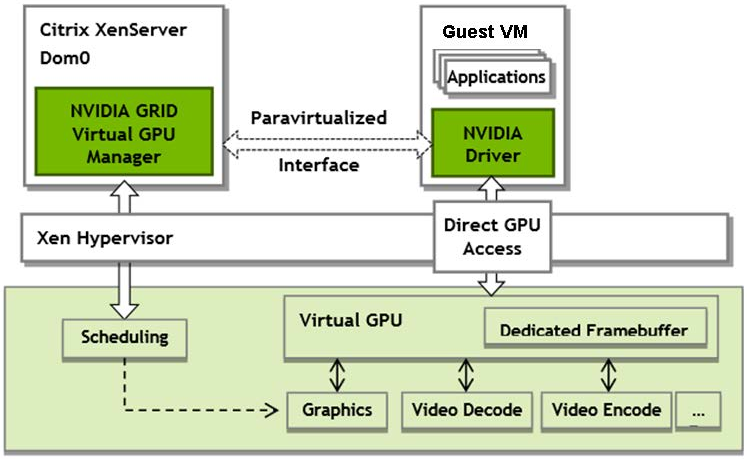
Figure 10. GRID vGPU System Architecture



NVIDIA GRID vGPUs are analogous to conventional GPUs, having a fixed amount of GPU frame buffer, and one or more virtual display outputs or "heads". The vGPU's frame buffer is allocated from the physical GPU's frame buffer at the time the vGPU is created, and 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), video decoding, and video encoding engines (Figure 11).

Figure 11. NVIDIA GRID vGPU Internal Architecture



Supported GPUs

NVIDIA GRID vGPU is supported on NVIDIA GRID K1 and K2 GPUs. Refer to the release notes for a list of recommended server platforms to use with NVIDIA GRID K1 and K2.

Virtual GPU (vGPU) types

NVIDIA GRID K1 and K2 each implement multiple physical GPUs. NVIDIA GRID K2 has two GPUs onboard and GRID K1 has four GPUs.

Each physical GPU can support several different types of virtual GPU. Virtual GPU types have a fixed amount of frame buffer, number of supported display heads and maximum resolutions, and are targeted at different classes of workloads.

Table 3 summarizes the vGPU types supported by GRID K1 and K2 are defined in Table xx.

Table 3. vGPU Types

Card	Physical GPUs	Virtual GPUs	Intended Use Case	Frame Buffer (Megabytes)	Virtual Display Heads	Max Resolution per Display Head	Maximum vGPUs	
							Per GPUs	Per Board
GRID K1	4	GRID K140Q	Power User	1024	2	2560x1600	4	16
		GRID K120Q	Power User	512	2	2560x1600	8	32
		GRID K100	Knowledge Worker	256	2	1920x1200	8	32
GRID K2	2	GRID K260Q	Power User, Designer	2048	4	2560x1600	2	4
		GRID K240Q	Power User, Designer	1024	2	2560x1600	4	8
		GRID K220Q	Power User, Designer	512	2	2560x1600	8	16
		GRID K200	Knowledge Worker	256	2	1920X1200	8	16

Due to their differing resource requirements, the maximum number of vGPUs that can be created simultaneously on a physical GPU varies according to the vGPU type. For example, a NVIDIA GRID K2 physical GPU can support up to 4 K240Q vGPUs on each of its two physical GPUs, for a total of 8 vGPUs, but only 2 K260Qs vGPUs, for a total of 4 vGPUs.

Citrix XenServer 6.2 with Service Pack 1

Citrix® XenServer® is an industry and value leading, open source virtualization platform for managing cloud, server and desktop virtual infrastructures and enabling a seamless path to the cloud. Key features include:

3D Graphics Pack (3DGP)

Support for hardware-accelerated vGPUs based on the NVIDIA GRID technology. Customers who have NVIDIA GRID K1 or GRID K2 cards installed in their systems can use this technology to share GPUs between multiple Virtual Machines. When combined with XenDesktop HDX 3D Pro, this enables the use of rich 3D applications, such as CAD, to be used by up to 64 concurrent VMs per server.

XenServer GPU Acceleration

The Virtual Graphical Processing Unit (vGPU) feature enables multiple virtual machines to directly access the graphics processing power of a single physical GPU. You can use hardware-accelerated vGPU access for Windows desktop VDI workloads. Both Windows client and server OS versions with true hardware-accelerated GPU sharing are suitable for users with complex and demanding design requirements. Supported for NVIDIA GRID K1 and K2 cards, vGPU sharing uses the same NVIDIA graphics drivers that are deployed on non-virtualized operating systems. Under the control of NVIDIA's GRID vGPU Manager, which runs in the XenServer Control Domain (dom0), compatible physical GPUs (pGPU) are capable of supporting multiple virtual GPU devices (vGPU) that can be assigned directly to VMs. Guest VMs use GRID vGPU in the same manner as a physical GPU that has been passed through by the hypervisor: an NVIDIA driver loaded in the guest VM provides direct access to the GPU for performance-critical fast paths, and a paravirtualized interface to the GRID vGPU Manager.

Datacenter automation suite

Automate key IT processes to improve service delivery and business continuity for virtual environments, resulting in both time and money savings while providing more responsive IT services. Capabilities include site recovery, high availability, and memory optimization.

High density cloud and desktop environment optimizations

Ensure the highest performance and data security, especially when integrated with other industry leading products including Citrix CloudPlatform and Citrix XenDesktop.

Advanced integration and management tools

Simplify operations with a complete suite of tools for role-based administration, performance reporting and integration with third-party storage.

High performance virtualization platform

Leverage best-in-class technologies, including the Xen Project hypervisor, XenMotion live migration, Storage XenMotion, the XenCenter management console and XenServer Conversion Manager for VMware to XenServer conversions.

Graphics Acceleration in Citrix XenDesktop

Citrix's HDX 3D Pro enables you to deliver desktops and applications that perform best with a graphics processing unit (GPU) for hardware acceleration, including 3D professional graphics applications based on OpenGL and DirectX. (The standard Citrix Virtual Desktop Agent (VDA) supports GPU acceleration of DirectX only.)

Example 3D professional applications include:

- Computer-aided design, manufacturing, and engineering (CAD/CAM/CAE) applications
- Geographical Information System (GIS) software
- Picture Archiving Communication System (PACS) for medical imaging
- Applications using the latest OpenGL, DirectX, NVIDIA CUDA, and OpenCL versions
- Computationally-intensive non-graphical applications that use NVIDIA Compute Unified Device Architecture (CUDA) GPUs for parallel computing

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

- On wide area network (WAN) connections: Deliver an interactive user experience over WAN connections with bandwidths as low as 1.5 Mbps.
- On local area network (LAN) connections: Deliver a user experience equivalent to that of a local desktop on LAN connections with bandwidths of 100 Mbps.

HDX 3D Pro enables you to replace complex and expensive workstations with much simpler user devices by moving the graphics processing into the data center for centralized management.

HDX 3D Pro provides GPU acceleration for Windows Desktop OS machines and Windows Server OS machines. When used with Citrix XenServer and NVIDIA GRID GPUs, HDX 3D Pro provides Virtual GPU (vGPU) acceleration for Windows Desktop OS machines.

GPU acceleration for Windows Desktop OS

With HDX 3D Pro you can deliver graphically intensive applications as part of hosted desktops or applications on Desktop OS machines, according to the requirements of your users. HDX 3D Pro supports physical host computers (including desktop, blade, and rack workstations) and XenServer VMs with GPU Passthrough and XenServer VMs with Virtual GPU (vGPU).

The XenServer GPU Passthrough feature enables you to create VMs with exclusive access to dedicated graphics processing hardware. You can install multiple GPUs on the hypervisor and assign VMs to each of these GPUs on a one-to-one basis.

The XenServer vGPU feature enables multiple virtual machines to directly access the graphics processing power of a single GPU.

GPU acceleration for Windows Server OS

HDX 3D Pro allows graphics-heavy applications running in Windows Server OS sessions to render on the server's graphics processing unit (GPU). By moving OpenGL, DirectX, Direct3D, and Windows Presentation Foundation (WPF) rendering to the server's GPU, the server's central processing unit (CPU) is not slowed by graphics rendering. Additionally, the server is able to process more graphics because the workload is split between the CPU and GPU.

When using HDX 3D Pro, multiple users can share graphics cards. When HDX 3D Pro is used with XenServer GPU Passthrough, a single server hosts multiple graphics cards, one per virtual machine.

GPU Sharing for RDS workloads

GPU Sharing enables GPU hardware rendering of OpenGL and DirectX applications in remote desktop sessions. GPU Sharing has the following characteristics:

- Can be used on bare metal or virtual machines to increase application scalability and performance.
- Enables multiple concurrent sessions to share GPU resources. (Most users do not require the rendering performance of a dedicated GPU.)
- Requires no special settings for basic operations.
  - To learn more: http://support.citrix.com/proddocs/topic/xenapp-xendesktop-75/hd-3d-gpu-acceleration-win-server-os.html

HDX 3D Pro integration

HDX 3D Pro integrates with your existing XenApp and XenDesktop infrastructure. You can deliver graphical applications as part of hosted applications or desktops on Desktop OS machines.

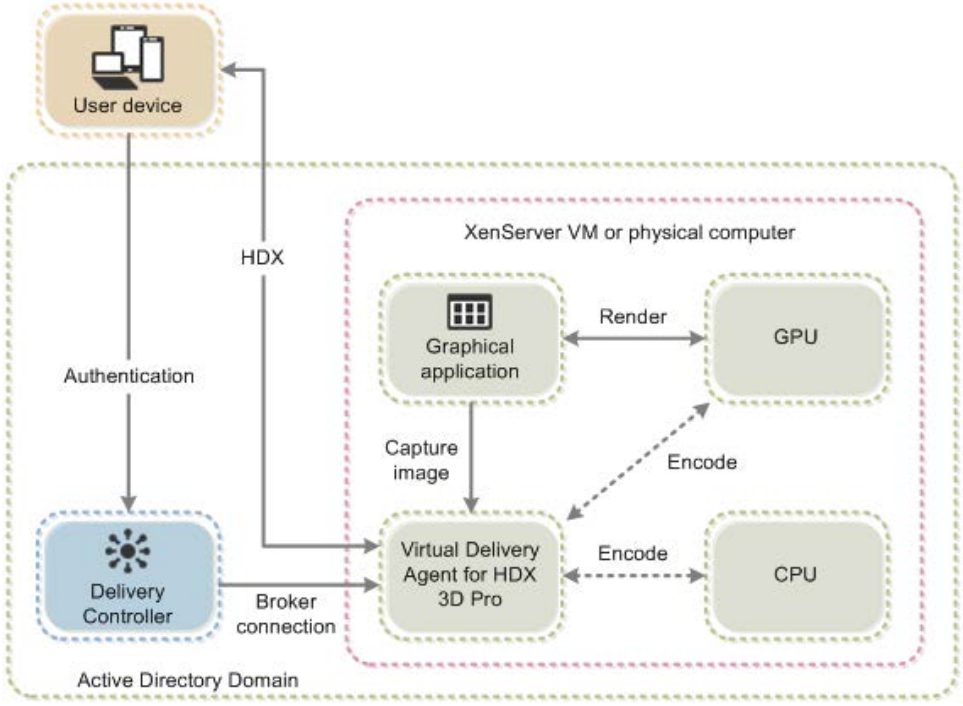
As shown in the following figure:

- The host computer must reside within the same Active Directory domain as your Delivery Controller.
- When a user logs on to Citrix Receiver and accesses the virtual application or desktop, the controller authenticates the user and contacts the VDA for HDX 3D Pro to broker a connection to the computer hosting the graphical application.

The VDA for HDX 3D Pro uses the appropriate hardware on the host to compress views of the complete desktop or just of the graphical application.

- The desktop or application views and the user interactions with them are transmitted between the host computer and the user device through a direct HDX connection between Citrix Receiver and the VDA for HDX 3D Pro.

Figure 12. Citrix XenServer and XenDesktop Graphics Processing Logical Flow



Software Requirement for GPU Acceleration Support

- NVIDIA GRID K1 or K2 cards: Customers who have installed hotfix XS62ESP1004 can use additional NVIDIA GRID cards and vGPU types. Hotfix XS62ESP1004 is available from https://support.citrix.com/article/CTX140417
  - For the most recent vGPU types refer to http://www.nvidia.com/object/virtual-gpus.html.
  - For the most recently supported NVIDIA cards refer to
    - XenServer Hardware Compatibility List—http://hcl.vmd.citrix.com/vGPUDeviceList.aspx
    - NVIDIA product information—http://www.nvidia.com/object/grid-boards.html
- A server capable of hosting XenServer and NVIDIA GRID cards. (Refer to the vGPU Release Notes at www.citrix.com/go/vgpu for details of recommended hardware)
- Citrix XenServer 6.2.0 Service Pack 1 or later
- The NVIDIA GRID vGPU software package for Citrix XenServer, consisting of the NVIDIA GRID Virtual GPU Manager for XenServer, and NVIDIA drivers for Windows 7 32-bit/64-bit Available from http://www.nvidia.com/vGPU
- To run Citrix XenDesktop with VMs running NVIDIA vGPU, you will also need:

  Citrix XenDesktop 7.1 or later full installation

Note: No other versions of Citrix XenServer or XenDesktop are currently supported for use with NVIDIA virtual GPUs.

Software Requirement for vGPU Guest Support

The following guests are supported for use with vGPU:

- Microsoft Windows 7 (32-bit/64-bit)
- Microsoft Windows Server 2008 R2 SP1

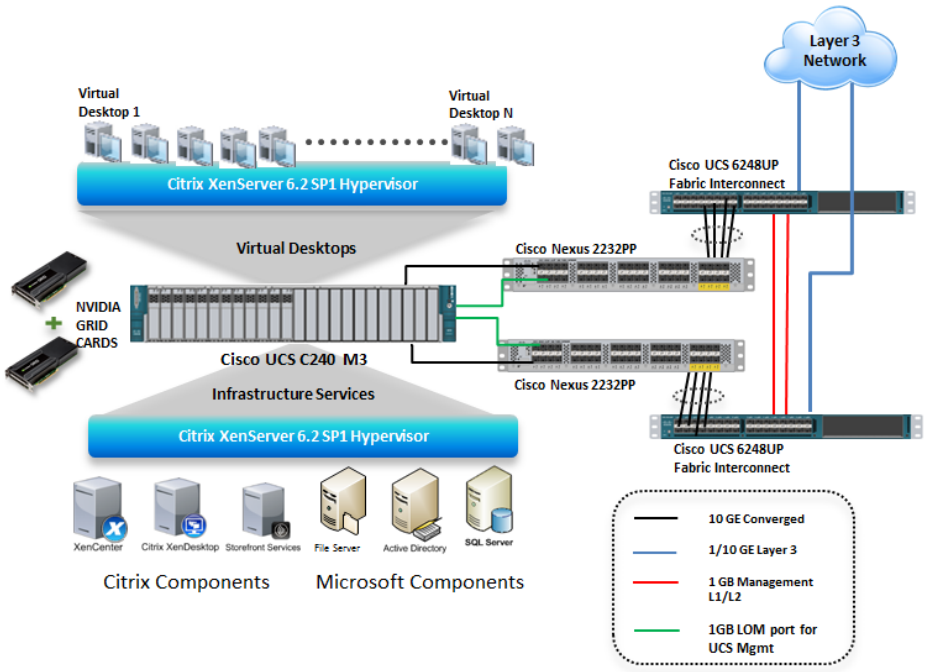
Customers who have installed Hotfix XS62ESP1004 can also use the following guests:

- Microsoft Windows 8 (32-bit/64-bit)
- Microsoft Windows 8.1 (32-bit/64-bit)
- Microsoft Windows Server 2012
- Microsoft Windows Server 2012 R2

Solution Configuration

Figure 13 illustrates the solution architecture.

Figure 13. Reference Architecture



Hardware Components:

- Cisco UCS C240-M3 Rack Server (2 X Intel Xeon processor E5-2680 v2 @ 2.70 GHz) with 256GB of memory (16 GB X 16 DIMMS @ 1866 MHz), hypervisor host
- Cisco UCS VIC1225 CNA/Rack Server—1 GPU scenario
- 2 Cisco Nexus 2232PP Fabric Extenders—2 GPU scenario
- 2 Cisco UCS 6248UP Fabric Interconnects

- 12-600GB SAS disks @ 10000 rpm
- 1 or 2 NVIDIA GRID K1/K2 cards
- Emulex OCe11102-F Gen 3 CNA card at10Gbps—2 GPU scenario

Software components:

- Cisco UCS firmware Version 2.2(2C)
- Citrix XenServer 6.2 SP1 for VDI Hosts
- Citrix XenDesktop 7.1
- Microsoft Windows 2012 R2
- Microsoft Windows 7 SP1 64 bit

UCS Configuration

Installing NVIDIA GRID GPU Card on C240 M3:

Table 4 lists the prerequisites for installing NVIDIA GRID card(s) in the Cisco UCS C240 M3.

Table 4. Minimum server firmware versions for the GPU cards

GPU	Minimum BIOS Version
NVIDIA GRID K1	1.5(1)
NVIDIA GRID K	1.5(1)

NVIDIA GPU Card Configuration Rules:

- Mixing different GPU cards in the same server is not supported.
- All GPU cards require two CPUs and two 1200 W power supplies in the server.
- It is not possible to use dual NVIDIA GPU cards and a Cisco virtual interface card (VIC) at the same time. This is because dual NVIDIA GPUs must be installed in slots 2 and 5 of the server, and a Cisco VIC must be installed in either slot 2 or slot 5. If you require two GPU cards and 10-Gb Ethernet connectivity, you must chooses a different supported adapter that can be used in a different slot. For supported adapters, see the Technical Specifications Sheet for the Cisco UCS C240 M3 server (Small Form Factor or Large Form Factor) at: http://www.cisco.com/en/US/products/ps10493/products_data_sheets_list.html

Figures 14 and 15 show the two scenarios for deploying NVIDIA GRID cards.

Figure 14. One GPU scenario

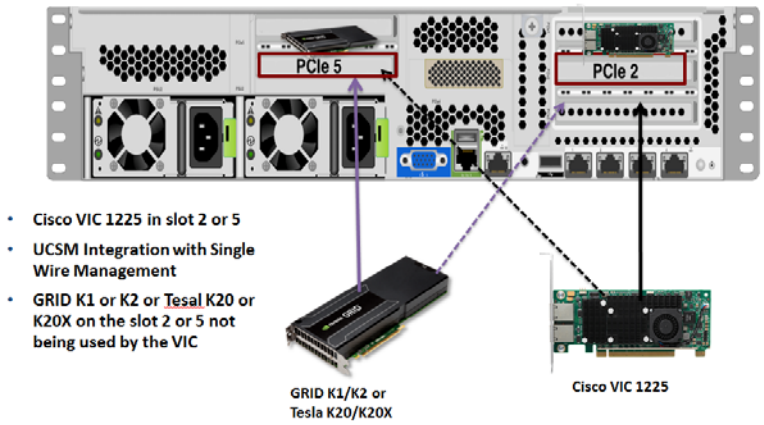
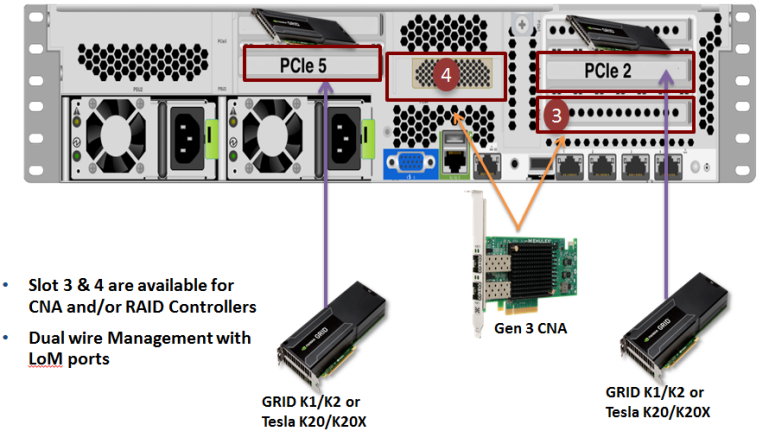


Figure 15. Two GPU scenario



Note: As shown in Figure 15, two GPU scenario, we used Emulex OCe11102-F Gen 3 CNA card.

Follow link given below for physical configuration of GRID cards in Riser cards slot 2 and 5:
http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/c/hw/C240/install/C240/replace.html#pgfId-1373451

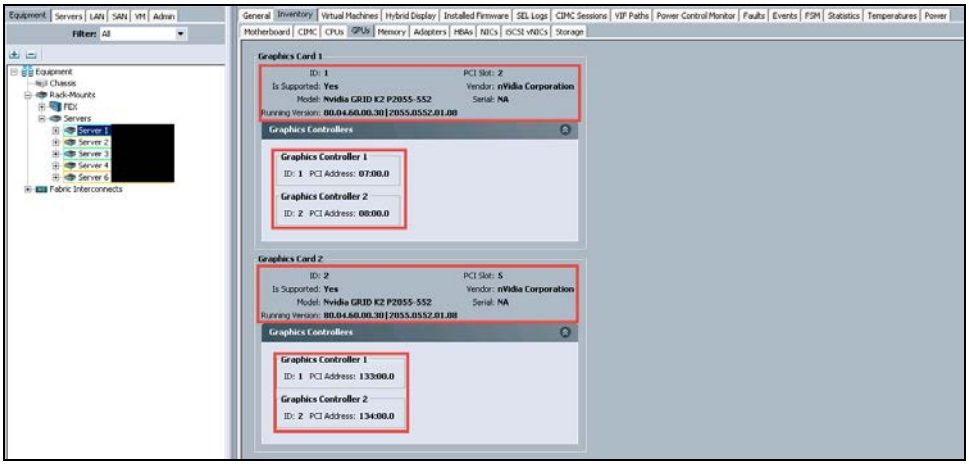
Base UCS System Configuration

For physical connectivity guidance and best practices for Cisco UCS C-Series server integration with Cisco UCS Manager, see: http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/c-series_integration/ucsm2-2/b_C-Series-Integration_UCSM2-2.html

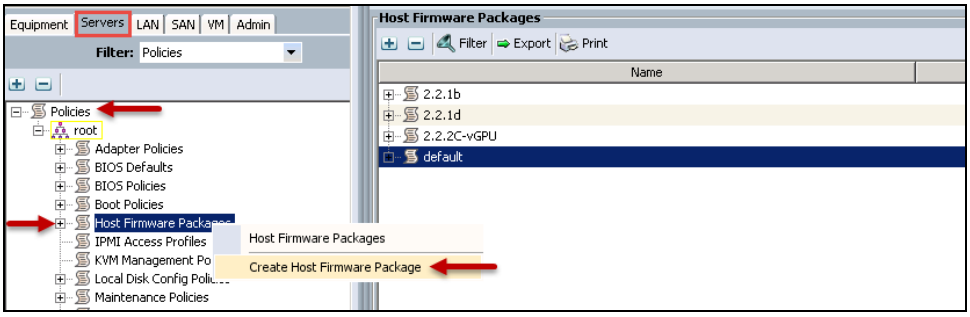
To perform the base Cisco UCS configuration, follow these steps:

1. Once server is discovered in Cisco UCS Manager (UCSM,) Select server → Inventory → GPUs.

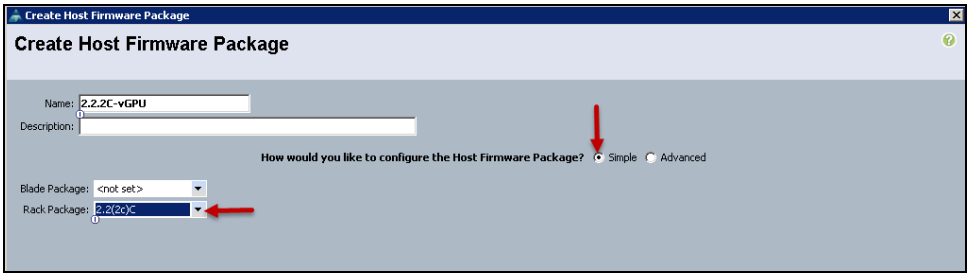
   As shown the screenshot below; PCIe slot 2 and 5 are utilized with two NVIDIA GRID K2 cards with running firmware version 80.04.60.00.30 | 2055.0552.01.08

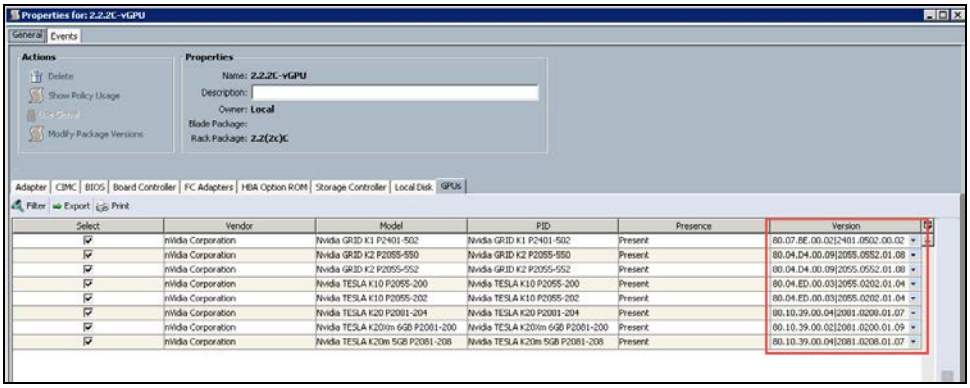


2. Create host firmware policy by selecting Server node on Cisco UCSM. Select Policies → Host Firmware Packages. Right click and select Create Host Firmware Package.

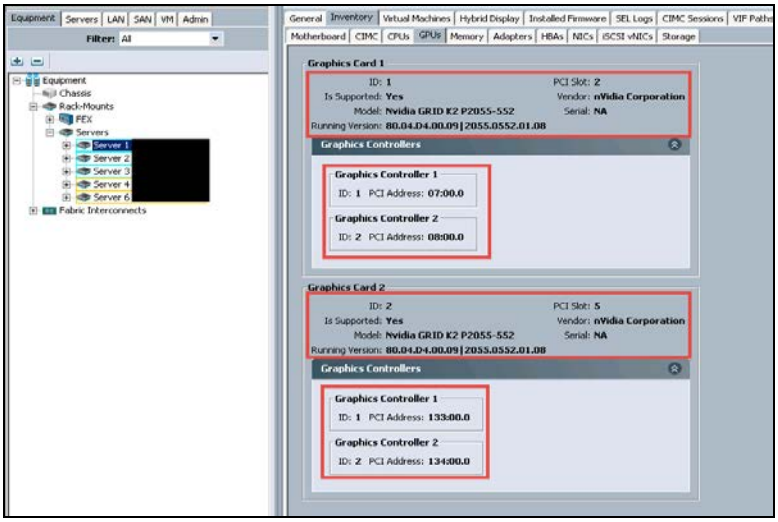


3. Select radio button for simple configuration of Host firmware package, select Rack package as 2.2.2C. Click OK.





4. Apply this host firmware package in Service Profile Template/Service Profiles firmware policy. Once completed upgrading firmware for server GPUs are running on firmware version selected.



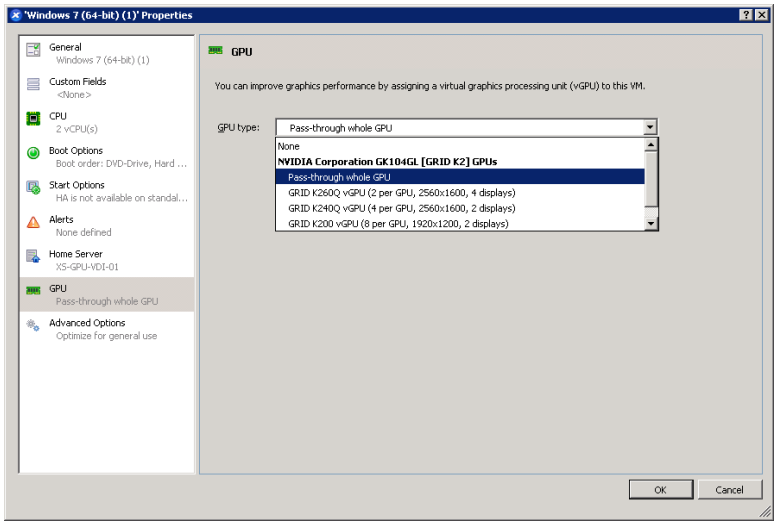
Enable Virtual Machines for Pass-Through configuration

With the NVIDIA GPU GRID cards installed in the server, follow the steps below to enable a VM for Pass-Through support.

1. Login to the Citrix XenServer using Citrix XenCenter.
2. Provision a test Windows7 Enterprise 64-bit virtual machine
   a. Create a VM with 4 vCPU and 16GB RAM (refer to CTX135811)
   b. Install XenServer Tools
   c. Add to Active Directory domain
   d. Enable Remote Desktop access
3. Right-click the test VM and select Properties
4. Go to the GPU tab, on right panel you should see the NVIDIA GPU listed in the drop-down.

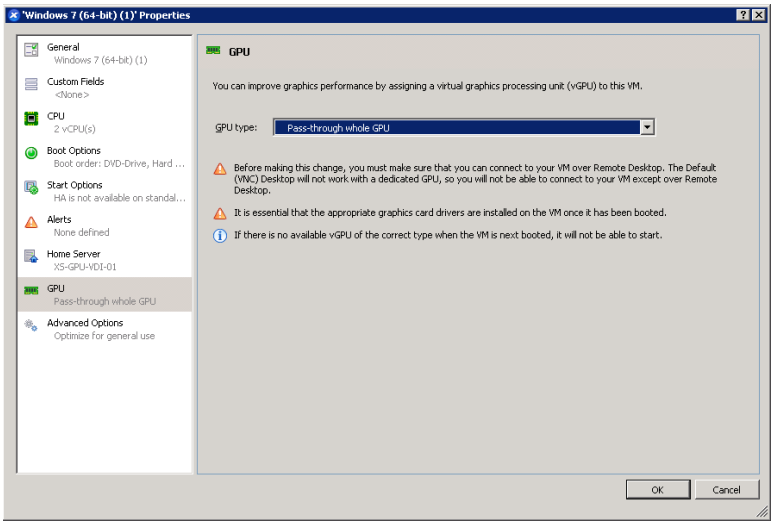
5. Select/Assign the NVIDIA GPU to the test VM.

   Example: NVIDIA Corporation GK104GL (GRID K2) GPUs, Pass-through whole GPU



   After selecting the pass-through GPU option, the following information is presented.

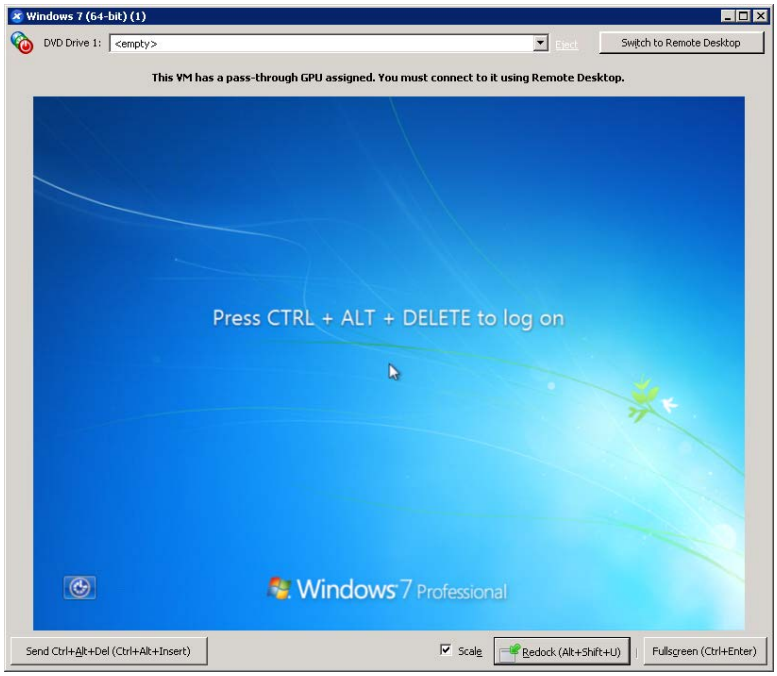
6. Power-On the test VM



   The test VM console in XenCenter states:

   This VM has a pass-through GPU assigned. You must connect to it using Remote Desktop'

7. Select Switch to Remote Desktop and login.

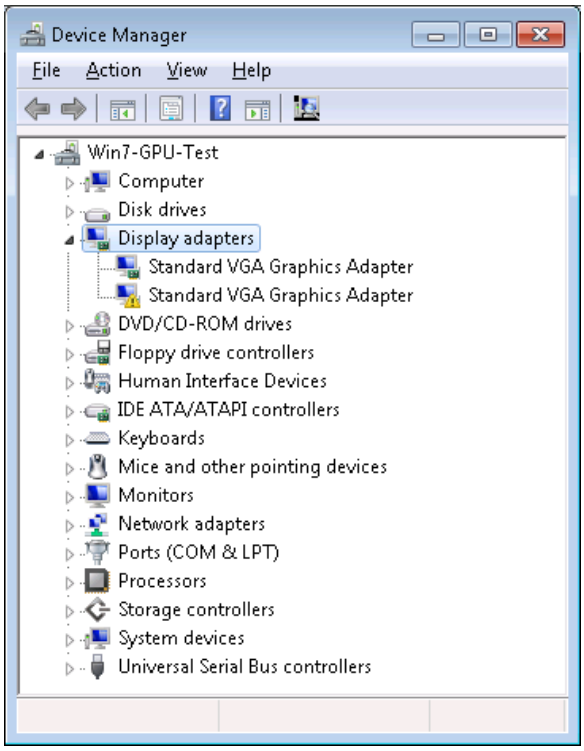


Using Windows Device Manager, review the graphic adapters within the test VM.

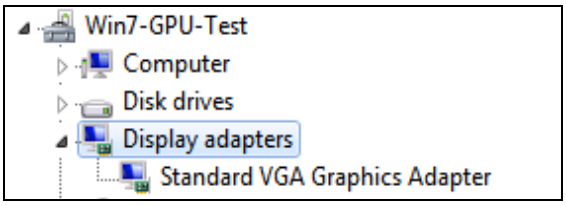
8. Go to Control Panel > Device Manager

   Optionally, you can open via a command line: Start > Run > type devmgmt.msc and hit Enter

9. Under Display adapters, you will see a second VGA Graphics Adapter with an exclamation (!) mark in yellow triangle. This indicates the NVIDIA driver is not yet installed.



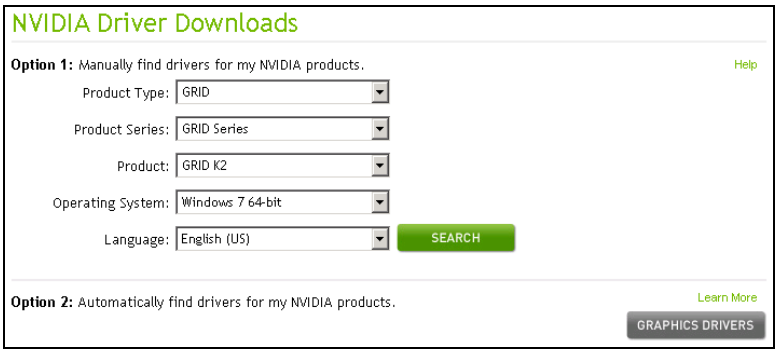
Note: Before a GPU is assigned to a VM. Device manager shows only one Standard VGA Graphics Adapter



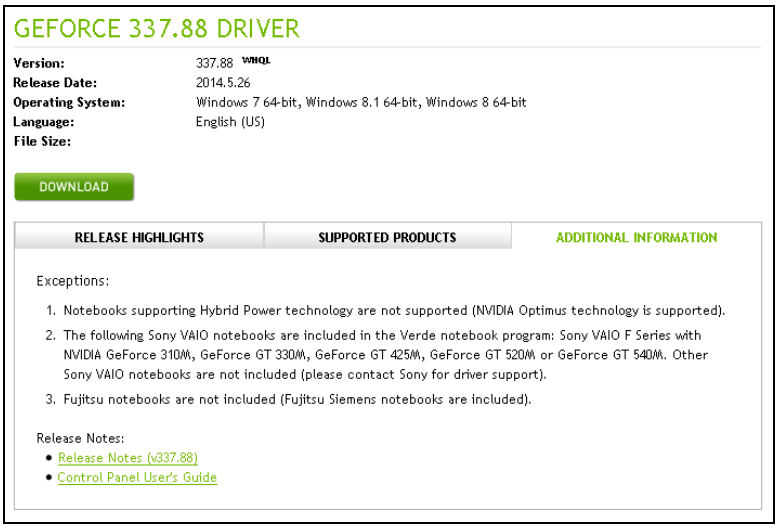
Pass-through GPU Software (driver) installation and configuration

The following steps will install the GPU driver on the guest operating system and deliver a GPU-powered XenServer VM (Virtual Machine).

1. From the VM where GPU is installed, go to www.nvidia.com/drivers
2. Select either Option 1 or 2 to download the driver



3. Select Download
4. Select the Additional Information tab for the driver release notes and User's Guide



Note: This driver is intended for use with a pass-through configuration. A separate driver type is available for a vGPU-enabled configuration (see section 6.3.4).

Install latest NVIDIA desktop drivers on virtual machine

1. Accept license terms and agreement click next.

2. Select radio button for custom installation. Click Next.

3. Select check box for each option. Click Next.

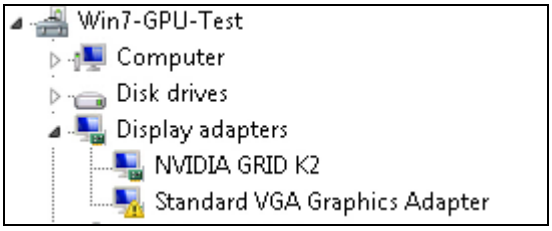
4. Click Finish. Reboot Virtual Machine.

5. From the test VM, open Device Manager expand Display adapters.

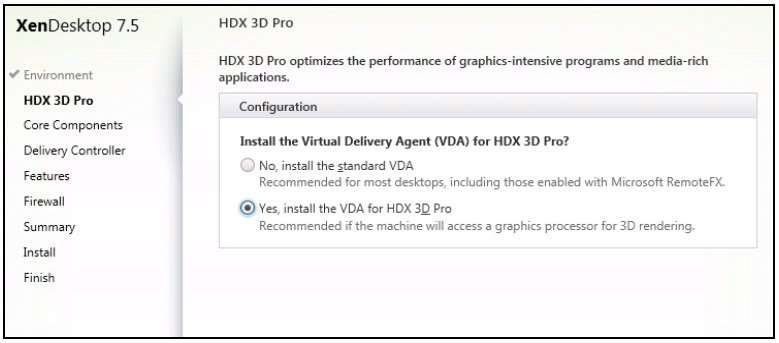
   Where previously we saw the exclamation triangle, it should show NVIDIA GRID Kx and no error icons.

Note: If you continue to see an exclamation mark, most likely reasons are:

- GPU driver service is not running
- Incompatible GPU driver



6. Install Citrix XenDesktop HDX 3D Pro Virtual Desktop Agent. Reboot when prompted.



Verify applications are ready to use the GPU Pass-Through Support

Using XenServer Command Line Interface (CLI)

1. Access your XenServer console using root credentials.
2. Type this command to verify if the GPU card is identified by the hypervisor:

   `lspci | grep VGA`

```
[root@XS-GPU-VDI-01 ~]# lspci | grep VGA
07:00.0 VGA compatible controller: NVIDIA Corporation GK104GL [GRID K2] (rev a1)
08:00.0 VGA compatible controller: NVIDIA Corporation GK104GL [GRID K2] (rev a1)
0b:00.0 VGA compatible controller: Matrox Electronics Systems Ltd. MGA G200e
[Pilot] ServerEngines (SEP1) (rev 02)
85:00.0 VGA compatible controller: NVIDIA Corporation GK104GL [GRID K2] (rev a1)
86:00.0 VGA compatible controller: NVIDIA Corporation GK104GL [GRID K2] (rev a1)
```

3. List of GPU Groups created by XenServer and their corresponding UUIDs:

   `xe gpu-group-list`

```
[root@XS-GPU-VDI-01 ~]# xe gpu-group-list
uuid ( RO)          : 31ab76c9-ebf3-ca6f-f1a7-86baacce75b7
   name-label ( RW): Group of Matrox Electronics Systems Ltd. MGA G200e [Pilot]
ServerEngines (SEP1) GPUs
   name-description ( RW):
uuid ( RO)        : 5d5f30bd-5839-f435-7c2f-4c440fddd4e8
 name-label ( RW): Group of NVIDIA Corporation GK104GL [GRID K2] GPUs
 name-description ( RW):
```

4. List of all GPU cards attached to the XenServer: **xe pgpu-list**

```
[root@XS-GPU-VDI-01 ~]# xe pgpu-list
uuid ( RO)             : 5d8f530d-ba9f-722f-b3a1-f28d5a432777
       vendor-name ( RO): NVIDIA Corporation
       device-name ( RO): GK104GL [GRID K2]
   gpu-group-uuid ( RW): 5d5f30bd-5839-f435-7c2f-4c440fddd4e8

uuid ( RO)             : 794f37e8-e5a4-0f3f-04ec-d110dc618b07
       vendor-name ( RO): NVIDIA Corporation
       device-name ( RO): GK104GL [GRID K2]
   gpu-group-uuid ( RW): 5d5f30bd-5839-f435-7c2f-4c440fddd4e8

uuid ( RO)             : ee2f2a4e-118c-bcc2-6533-1d7614e830cf
       vendor-name ( RO): NVIDIA Corporation
       device-name ( RO): GK104GL [GRID K2]
   gpu-group-uuid ( RW): 5d5f30bd-5839-f435-7c2f-4c440fddd4e8
```

```
uuid ( RO)             : 19cb8f0c-1c3c-b371-3ac5-5084adc8abc2
       vendor-name ( RO): NVIDIA Corporation
       device-name ( RO): GK104GL [GRID K2]
   gpu-group-uuid ( RW): 5d5f30bd-5839-f435-7c2f-4c440fddd4e8

List VMs hosted on the XenServer: xe vm-list
[root@XS-GPU-VDI-01 ~]# xe vm-list
uuid ( RO)             : 04a6f049-dcfa-b5cd-3b73-0e440f2609a1
       name-label ( RW): Windows 7 (64-bit) (1)
       power-state ( RO): running
```

5. Using the UUID listed above, verify the test VM has a GPU assigned:

```
   xe vgpu-list vm-uuid=<uuid of VM>
[root@XS-GPU-VDI-01 ~]# xe vgpu-list vm-uuid=04a6f049-dcfa-b5cd-3b73-0e440f2609a1
uuid ( RO)             : 2e7d457c-138b-d5e2-f42d-8bcc0122e3e2
          vm-uuid ( RO): 04a6f049-dcfa-b5cd-3b73-0e440f2609a1
   gpu-group-uuid ( RO): 55d5f30bd-5839-f435-7c2f-4c440fddd4e8
```

Prepare to Install the NVIDIA GRID vGPU Manager

Install XenServer and Apply Service Pack 1

1. Download XenServer 6.2.0 from https://www.citrix.com/downloads/xenserver/product-software/xenserver-62.html
2. Install the XenServer 6.2.0 Base Installation ISO and XenCenter 6.2.0 Windows Management Console.
   Refer to the XenServer Quick Start Guide for comprehensive details on installation located here http://support.citrix.com/article/CTX137827
3. Install XenServer 6.2.0 Service Pack 1 available from http://support.citrix.com/article/CTX139788
4. Reboot your host
5. Install XenServer hotfix updates, namely, XS62ESP1004 available from http://support.citrix.com/article/CTX140417

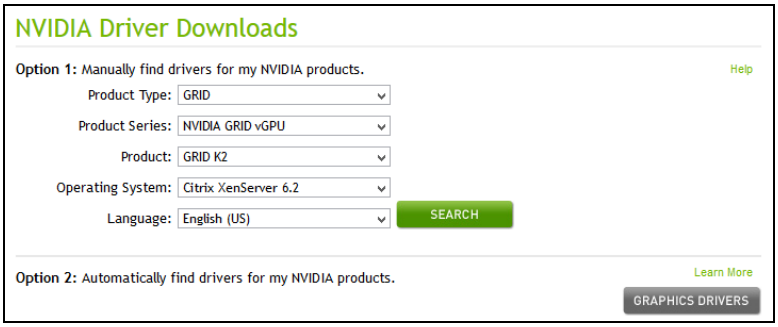
Note: Visit here for a complete list of required components with download links: https://www.citrix.com/go/private/vgpu.html

Install NVIDIA Virtual GPU Manager

vGPU Configuration

The NVIDIA Virtual GPU Manager runs in XenServer's dom0. It is provided as an RPM file, which must be copied to XenServer's dom0 and then installed.

1. Download the NVIDIA GRID vGPU software by visiting www.nvidia.com/drivers and selecting:
   - Product Type: GRID
   - Product Series: NVIDIA GRID vGPU
   - Product: GRID K1 or K2
   - Operating System: XenServer 6.2



Note: The software package includes: NVIDIA GRID vGPU Manager, release notes, user guide, and guest drivers for the NVIDIA GRID card.

2. Upload the NVIDIA driver to /tmp directory on the XenServer host using a tool such as WinSCP.
3. Place the XenServer host into maintenance mode.
4. Log into the XenServer host console through SSH as root.
5. Use the rpm command to install the package:

```
rpm -iv NVIDIA-vgx-xenserver-6.2-331.59.i386.rpm
[root@xenserver ~]# rpm -iv NVIDIA-vgx-xenserver-6.2-331.59.i386.rpm
Preparing packages for installation...
NVIDIA-vgx-xenserver-6.2-331.59
```

6. Reboot the XenServer host: shutdown –r now

```
[root@xenserver ~]# shutdown -r now
Broadcast message from root (pts/1):
The system is going down for reboot NOW!
```

7. Exit XenServer maintenance mode.

Verify Virtual GPU Manager Installation

1. After the XenServer platform has rebooted, verify that the NVIDIA GRID package installed and loaded correctly by checking for the NVIDIA kernel driver in the list of kernel loaded modules:

   lsmod | grep nvidia

   ```
   [root@XS-GPU-VDI-01 ~]# lsmod | grep nvidia
   nvidia                9656305 20
   i2c_core              20294 2 nvidia,i2c_i801
   ```

2. Verify that the NVIDIA kernel driver can successfully communicate with the GRID physical GPUs in your system by running the nvidia-smi command, which should produce a listing of the GPUs in your platform: **nvidia-smi**

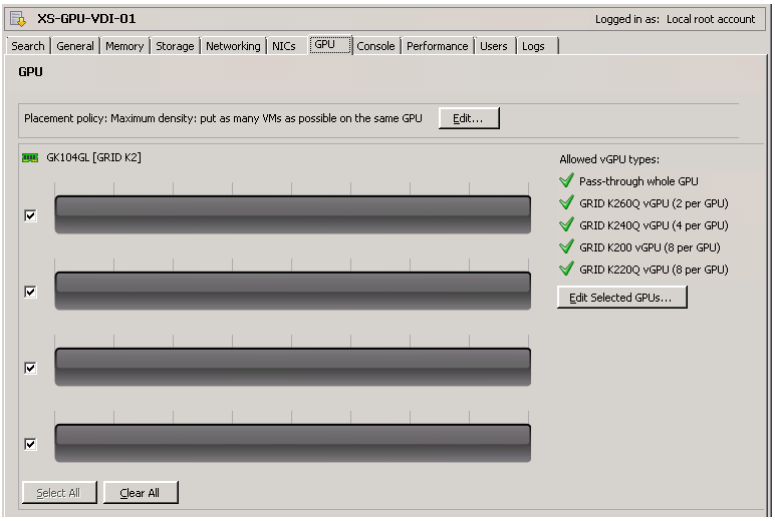
   ```
   [root@XS-GPU-VDI-01 ~]# nvidia-smi
   Wed Mar 19 02:48:32 2014
   +------------------------------------------------------+
   | NVIDIA-SMI 331.59 Driver Version: 331.59 |
   |-------------------------------+----------------------+----------------------+
   | GPU Name      Persistence-M   | Bus-Id        Disp.A | Volatile Uncorr. ECC |
   | Fan Temp Perf Pwr:Usage/Cap   |         Memory-Usage | GPU-Util Compute M.  |
   |===============================+======================+======================|
   | 0   GRID K2               On  | 0000:07:00.0     Off |                  Off |
   | N/A  45C   P8     27W / 117W  |      10MiB / 4095MiB | 0%           Default |
   +-------------------------------+----------------------+----------------------+
   | 1 GRID K2                 On  | 0000:08:00.0     Off |                  Off |
   | N/A  41C   P8     26W / 117W  |      10MiB / 4095MiB | 0%           Default |
   +-------------------------------+----------------------+----------------------+
   | 2 GRID K2                 On  | 0000:86:00.0     Off |                  Off |
   | N/A  37C   P8     27W / 117W  |      10MiB / 4095MiB | 0%           Default |
   +-------------------------------+----------------------+----------------------+

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

Note: The nvidia-smi command is only useful before GPUs are passed through to VMs or if you are using vGPU. The command will return an error when the GPUs are passed through and VMs are powered on because Dom0 can no longer see them. Error report as “Failed to initialize NVML: Unknown Error”

The vGPU types can also be verified after the rdiver installation using the command “xe vgpu-type-list”.

The GPUs can also be seen in XenCenter by selecting the host node and GPU tab. XenCenter shows two K2 cards with dual GPUs per card.



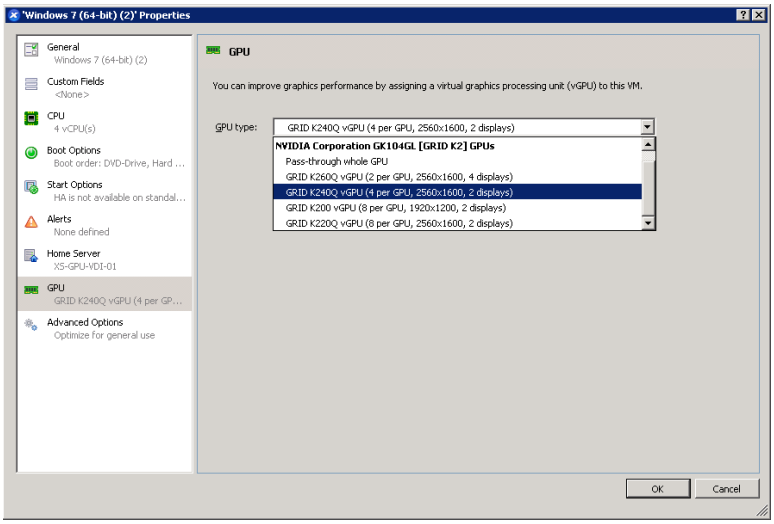
Enable Virtual Machine for vGPU configuration

XenServer supports configuration and management of virtual GPUs using XenCenter, or the xe command line tool that is run in a XenServer dom0 shell. Basic configuration using XenCenter is described in the following sections.

Follow these steps to enable a VM for vGPU GRID support.

1. Login to the XenServer using XenCenter.
2. Provision a test Windows7 Enterprise 64-bit virtual machine
   a. Create a VM with 4 vCPU and 16GB RAM (refer to CTX135811)
   b. Install XenServer Tools
   c. Add to Active Directory domain
3. Right-click the test VM and select Properties
4. Go to the GPU tab, on right panel you should see the NVIDIA GPU listed in the drop-down.
5. Select/Assign the NVIDIA GPU type to the test VM.
   Example: NVIDIA Corporation GK104GL (GRID K2) GPUs, GRID K240Q vGPU

Figure 16. Using XenCenter to configure a VM with a vGPU

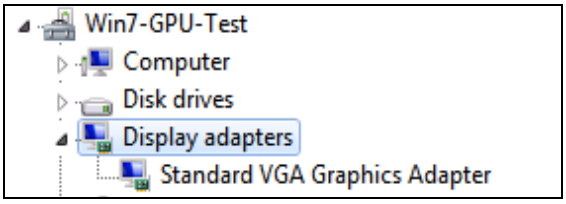


Virtual GPU Software (driver) installation and configuration

Once you have configured a VM with a vGPU, start the VM, using XenCenter. Viewing the VM's console in XenCenter the VM should boot to a standard Windows desktop in VGA mode at 800x600 resolution. The Windows screen resolution control panel may be used to increase the resolution to other resolutions, but to fully enable vGPU operation the NVIDIA driver must be installed.

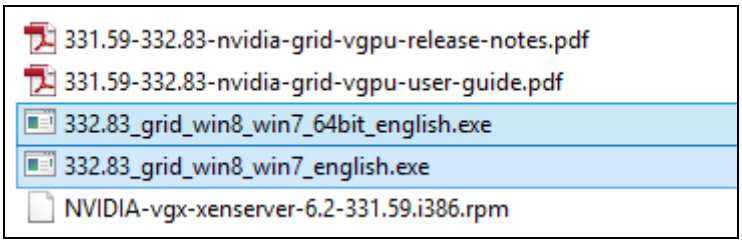
Before the NVIDIA driver is installed on the guest VM, Device Manager shows the Standard VGA Graphics Adapter (Figure 17).

Figure 17. Standard VGA Graphics Adapter in Device Manager



The following steps will install the GPU driver on the guest operating system and deliver a GPU-powered XenServer VM (Virtual Machine).

1. Copy the 32- or 64-bit NVIDIA Windows driver located from the vGPU driver pack to the test VM, and run setup.exe.



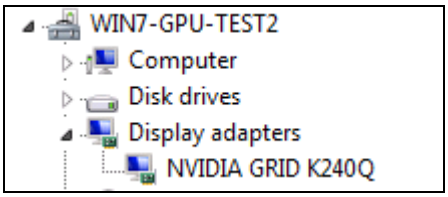
Note: The vGPU Manager and guest driver versions need to match. DO NOT attempt to use a newer guest driver with an older vGPU Manager or vice versa. The vGPU driver from NVIDIA is a different driver than the GPU Pass-through driver.

2. Install the NVIDIA Graphics Driver.
   a. Accept license terms and agreement click next.
   b. Select radial button for custom installation. Click Next.
   c. Select check box for each option. Click Next.
   d. Once complete, click Restart Now.

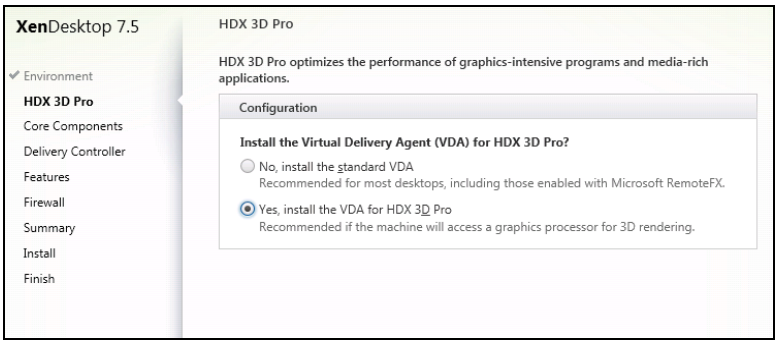
3. Verify the driver is correctly installed. Device Manager should show the NVIDIA GRID adapter with no error icons.

Note: If you continue to see an exclamation mark, most likely the reasons are:

- GPU driver service is not running
- Incompatible GPU driver



4. Install Citrix XenDesktop HDX 3D Pro Virtual Desktop Agent. Reboot when prompted.

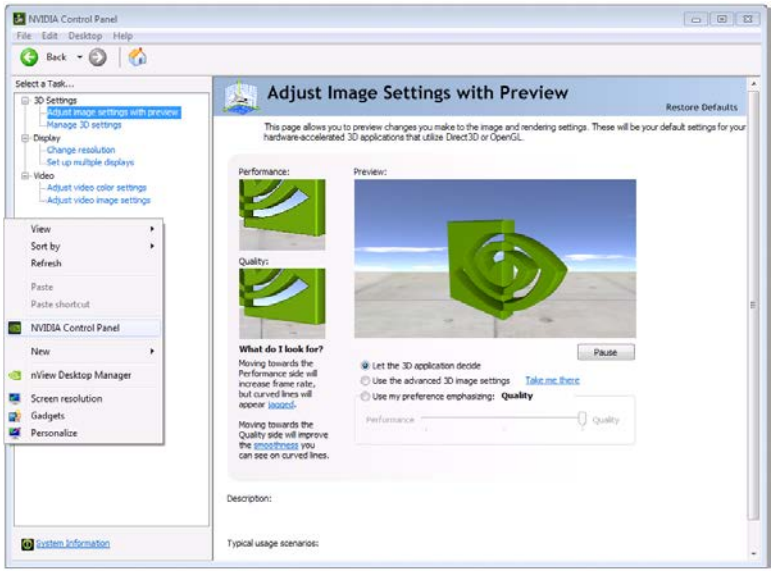


Note: If you install NVIDIA drivers after you install XenDesktop VDA with HDX 3D Pro, NVIDIA GRID is disabled. Enable NVIDIA GRID by using the Montereyenable tool provided by NVIDIA. To enable NVIDIA GRID, run the following command and restart the VDA: "Montereyenable.exe –enable –noreset"

Verify applications are ready to use the vGPU Support

Verify that the NVIDIA driver is running.

1. Right-click on the desktop. The NVIDIA Control Panel will be listed in the menu; select it to open the control panel.
2. Selecting "System Information" in the NVIDIA control panel will report the Virtual GPU that the VM is using, its capabilities, and the NVIDIA driver version that is loaded.



Virtual CPUs

HDX 3D Pro Graphics utilizes one virtual CPU completely for extracting desktop surfaces, typically 12 to 30 times per second. If Aero is enabled on Windows this requires a second virtual CPU. In addition, at least two vCPU is required for the graphics applications. Graphics applications are often CPU intensive, even with hardware acceleration of graphics rendering provided by the GPU. Hence it is important to configure the virtual machine running Citrix HDX 3D Pro Graphics with at least four virtual CPUs. Refer to the following CTX article for instructions on increasing the number of virtual CPUs on XenServer: http://support.citrix.com/article/CTX135811

This completes the process of setting up a single VM to use NVIDIA GRID vGPU. The VM is now capable of running the full range of DirectX and OpenGL graphics applications. In order to deliver the full performance and capabilities of vGPU, review the chapters on management and performance tuning in the GRID vGPU for Citrix XenServer User Guide.

Conclusion

The combination of Cisco UCS Manager, Cisco UCS C240 M3 rack servers, NVIDIA GRID cards, Citrix XenServer 6.2 SP1, and Citrix XenDesktop 7.5 provides a high performance platform for virtualizing graphics intensive applications.

The guidance in this document can help assure that our customers and partners are ready to host the growing list of graphics applications that are supported by Cisco, Citrix and NVIDIA.

References

Cisco Reference Documents & Support Sites

Cisco UCS C-Series Rack-Mount Servers: http://www.cisco.com/en/US/products/ps10265/

http://www.cisco.com/en/US/partner/products/ps12370/index.html

http://www.cisco.com/en/US/products/ps12571/index.html

Citrix Reference Documents & Support Sites

3D Graphics Pack: Configuring XenServer to use NVIDIA GRID v2.0 v1.2 April 2014:
http://support.citrix.com/servlet/KbServlet/download/37229-102-709582/Configuring%20XenServer%20to%20use%20NVIDIA%20GRID_3DGP_RTM.pdf

Citrix HDX 3D Pro technologies: http://support.citrix.com/proddocs/topic/xenapp-xendesktop-75/hd-3d-about.html

New Features and Improvements in XenServer 6.2.0 Service Pack 1:
http://support.citrix.com/servlet/KbServlet/download/36435-102-709189/XS620SP1_ReleaseNotes.pdf%20.pdf

Hotfix XS62ESP1004 download—For XenServer 6.2.0 Service Pack 1: http://support.citrix.com/article/CTX140417

Citrix 3D Graphics Pack download page: https://www.citrix.com/go/private/vgpu.html

NVIDIA Reference Documents & Support Sites
http://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-c-series-rack-servers/nvidia_grid_vgx.pdf

NVIDIA Product Specifications: http://www.nvidia.com/content/grid/resources/grid-cisco-datasheet.pdf

http://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-b-series-blade-servers/tesla_kseries_overview_lr.pdf

http://www.nvidia.com/content/grid/pdf/GRID_K1_BD-06633-001_v02.pdf

http://www.nvidia.com/content/grid/pdf/GRID_K2_BD-06580-001_v02.pdf

NVIDIA vGPU Driver Downloads: http://www.nvidia.com/Download/index.aspx?lang=en-us

GRID vGPU for Citrix XenServer User Guide:
http://us.download.nvidia.com/Windows/Quadro_Certified/GRID/332.83/331.59-332.83-nvidia-grid-vgpu-user-guide.pdf

Emulex CNA cards
http://www.emulex.com/products/ethernet-networking-storage-connectivity/converged-network-adapters/huawei-branded/oce11102-f/overview/

http://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-computing/ucs-c-series-rack-servers/OneConnect_dual-port_10Gbps_FCoE_Rack.pdf

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Printed in USA C11-732283-00 07/14