NVIDIA GRID™ ENTERPRISE SOFTWARE

QSG-07847-001_v02 | August 2016

Quick Start Guide

REQUIREMENTS

IMPORTANT: This Quick Start Guide is intended for those who are technically comfortable with minimal instructions.

Read the release notes to verify that you have the correct build numbers for all of the virtualization software that has been tested for the specific NVIDIA GRID™ Software release you are downloading. Other important version-specific information and known issues can also be found in the release notes.

You must be a registered on the NVIDIA® Licensing Portal to access the release notes.

Contents

This document contains the following information:

Requirements 2
Contents 2
Before You Begin 3
Product Activation Keys (PAK) 3
PAK Registration 4
Download GRID Software 5
License Server Creation, Installation, and Management 6
GPU Mode Switch Utility 10
Install Host Driver 11
Attach vGPU Profile to a VM 12
Install the NVIDIA driver into a VM 13
Configure a VM for an NVIDIA GRID vGPU License 14

Before You Begin

The NVIDIA GRID (version 2.0 and above) enterprise software is designed to run on select NVIDIA Tesla® GPUs. Read through the following list before you begin the software installation:

- NVIDIA GRID Certified Servers. For a complete list of GRID certified servers, go to: http://www.nvidia.com/object/grid-certified-servers.html
- Required GPU hardware:
  NVIDIA Tesla M10, M6 or M60
- A supported virtualization software stack

Product Activation Keys (PAK)

Once your order is processed, you receive a Welcome email containing your product activation keys.

Following is an example of the Welcome email:

Thank you for your NVIDIA GRID order!

The Product Activation Keys for your order are as follows:

AKHL-99840-09, Software, 25, Product Activation Key: 12345

Download your license file and NVIDIA GRID software by logging into the NVIDIA Licensing Portal here:

Register

If this is your first time visiting the NVIDIA Licensing Portal, you will be required to register.

If you already have an account, please log in

If you are registering more than one PAK, please complete the registration process for the first one, then click on "Register Additional Keys" link to register the additional PAKs.

Following registration, you will receive access information for the NVIDIA Enterprise Support Portal in a separate email.

Questions? Please visit our enterprise support portal or contact us at enterprisesupport@nvidia.com.

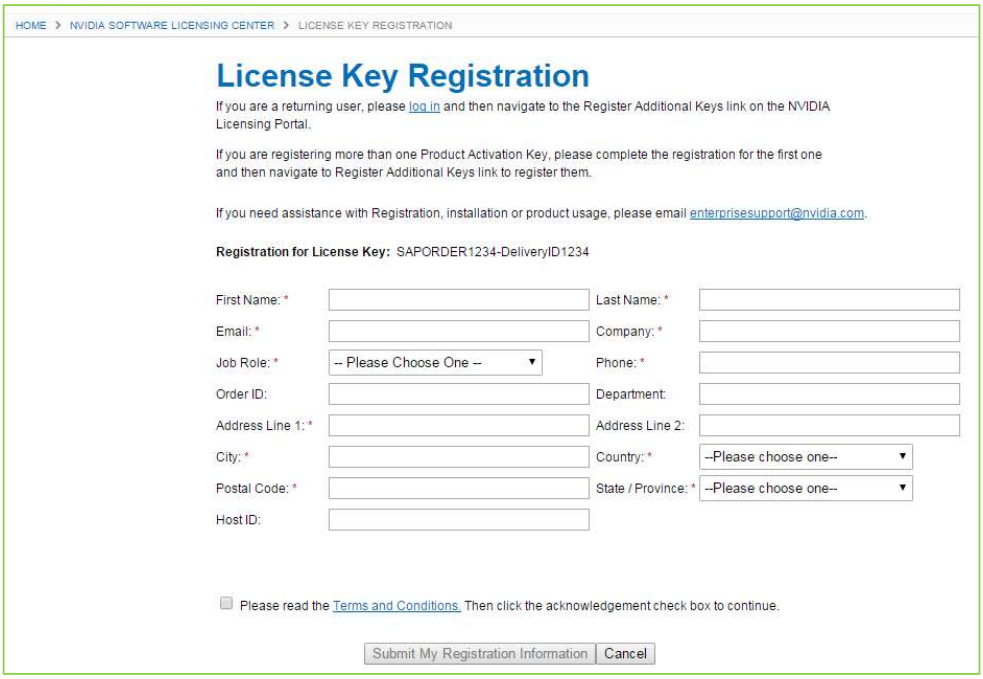
If you are an OEM or Reseller, please forward this message to the appropriate end customer for order number 12345 .

Thank you!

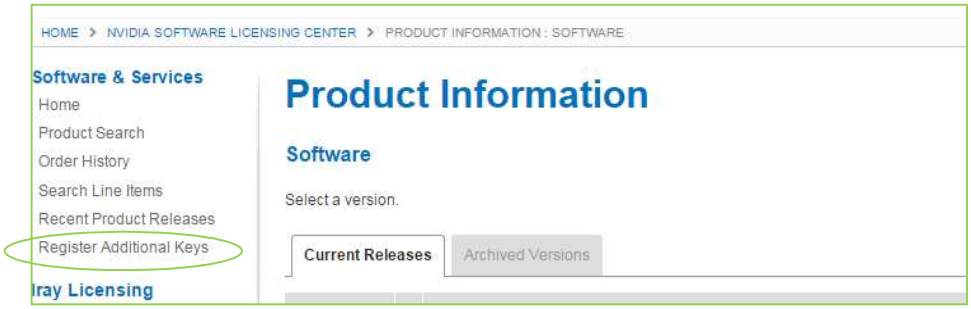
NVIDIA Customer Support Team

PAK Registration

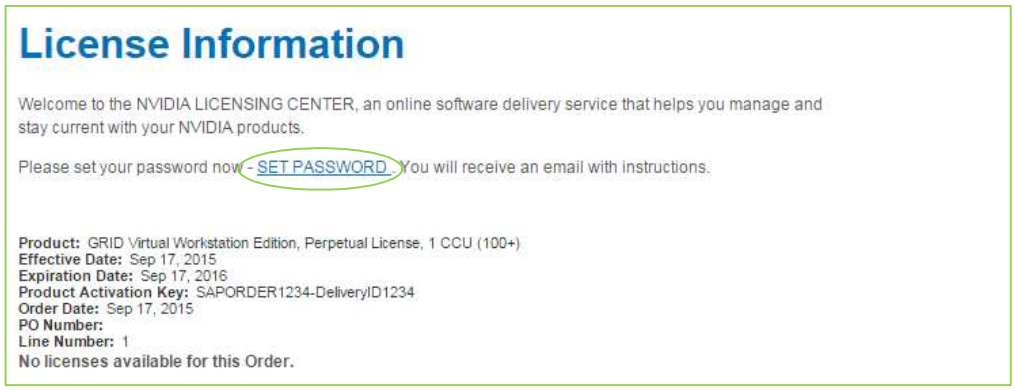
1. Select the log in link in the email if you already have an account on the NVIDIA Licensing Portal.
   If you do not have an account, select Register in the email to create an account. You must log in to gain access to your licenses and NVIDIA GRID software downloads.



2. Fill in the License Key Registration page and accept the Terms and Conditions.
3. Select Submit My Registration Information to display the **Product Information Software** page.
4. Register additional PAKs by selecting the Register Additional Keys link in the left pane of the **Product Information Software** page. The **License Key Registration** page displays for each additional key you request.

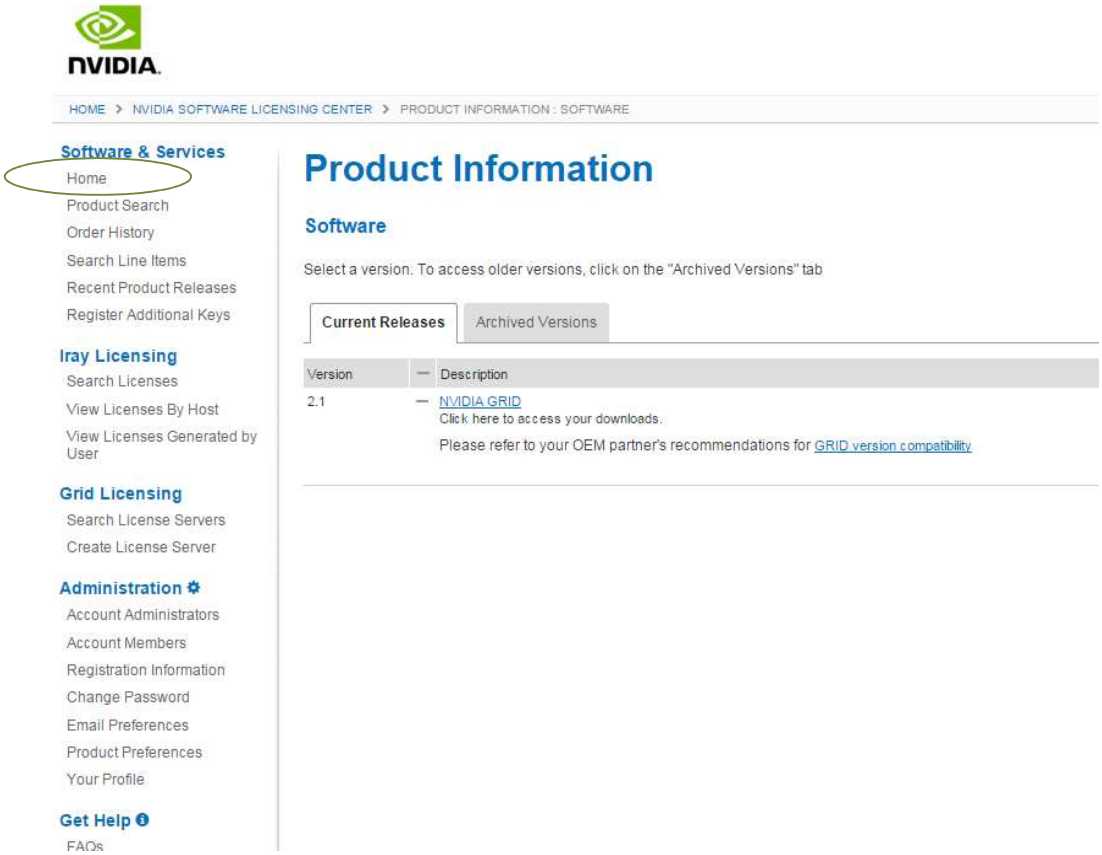


5. Select SET PASSWORD once registration completes and the License Information page displays.

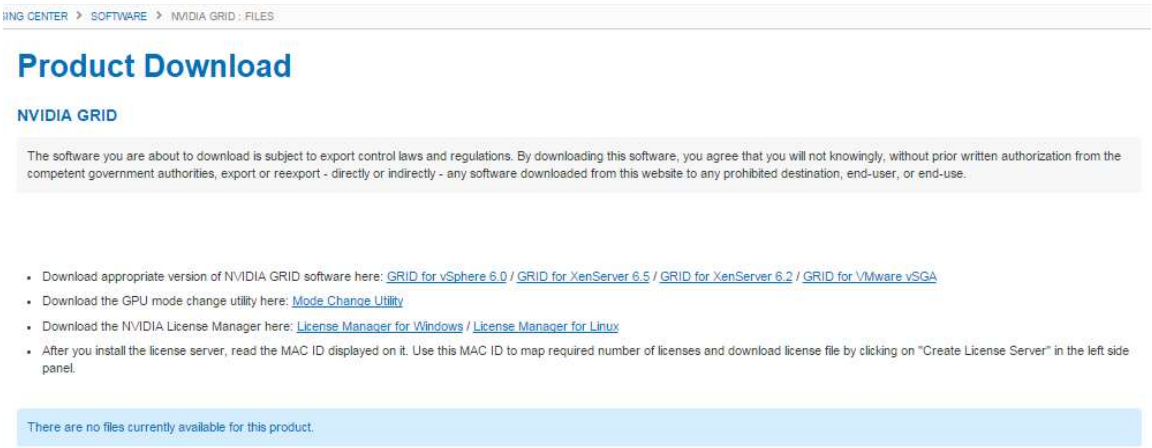


DOWNLOAD GRID SOFTWARE

1. Go back to the NVIDIA SOFTWARE LICENSING CENTER page and select Home from the list in the left pane to display the Product Information download page.



2. Select the Current Releases tab for latest GRID version. Older versions are available under the Archived Versions tab.
3. Select the NVIDIA GRID link to access the Product Download page containing the following software download links:
   - GRID Software: Select the host driver software based on the hypervisor brand and version you are running. NVIDIA GRID vGPU is supported on the latest versions of Citrix XenServer and VMware vSphere ESXi (refer to the release notes for the specific version).
   - GPU mode change utility: The Tesla M60/M6 boards are shipped in *Compute Mode*. In order to run NVIDIA GRID vGPU, the GPUs on the board must be changed to *Graphics Mode*. The GPU mode change utility is used to switch modes. Please refer to the *NVIDIA GRID gpumodeswitch User Guide* (DU-07765-001) for details.
   - NVIDIA License Manager Software: The license manager software is required to enable the Guest OS to utilize the NVIDIA GRID vGPU profiles by checking out a license at boot time. Please refer to the GRID License Server User Guide for details on installing and using the license manager.



License Server Creation, Installation, and Management

The NVIDIA GRID License Server manages a pool of floating licenses to enable NVIDIA GRID licensed products. It is installed onsite and configured with licenses generated and retrieved from NVIDIA's Licensing Portal website (see Figure 1). The NVIDIA GRID 2.0 vGPU profiles are licensed features on NVIDIA Tesla GPUs. These features are activated during a Guest OS boot by acquiring a software license issued from an NVIDIA GRID license server. The license is returned to the license pool on the server when the Guest OS shuts down.

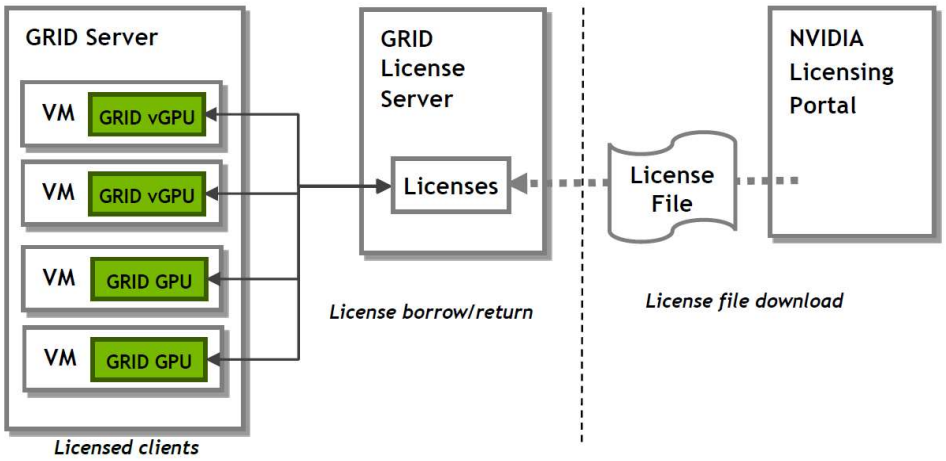


Figure 1. NVIDIA License Server

The license server requires a fixed IP address. The IP address may be assigned through DHCP or can be statically configured. The server's Ethernet MAC address is used as a unique identifier when registering the server and generating licenses in NVIDIA's licensing portal. The server runs on either Windows or Linux.

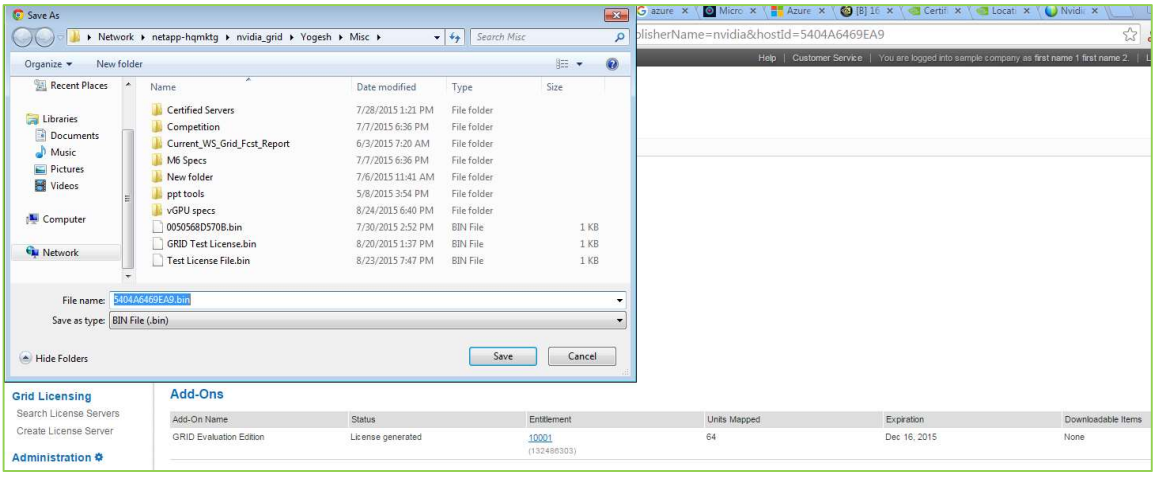
The system clock must to be accurate; NTP is recommended.

Use the following procedure to create a server interface:

1. Before proceeding, please ensure that you have downloaded and installed the GRID License Server.
2. Copy the MAC address of GRID license server. You will need this address for registering the license server and downloading license file.
3. Select Create License Server from under GRID Licensing in the left pane of the NVIDIA Software Licensing Center page to display the Create Server page.

4. Fill in your server details on the Create Server page. Enter the MAC address of the GRID license server in the License Server ID field.
5. Select the server to display View Server page. On this page, select Map Add-Ons to associate required number of licenses with this server. You can also create multiple servers and distribute your GRID licenses across them by mapping licenses to servers as necessary.

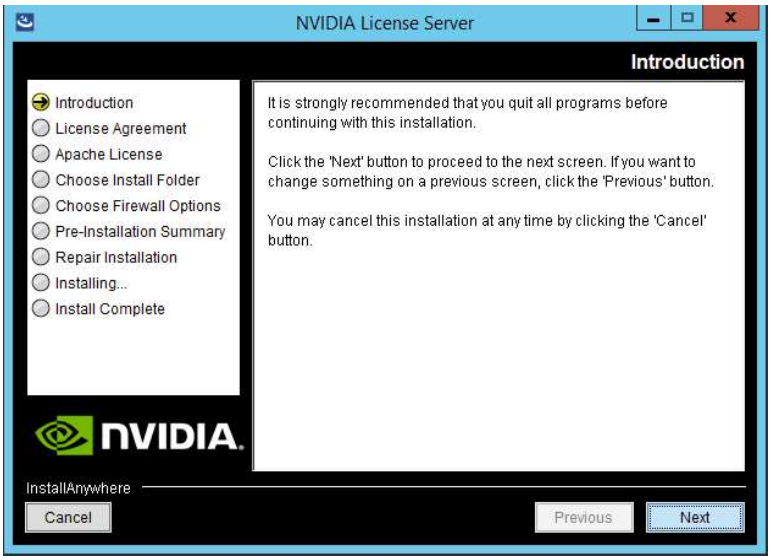
6. Select Download License File and save the **`.bin`** file onto your license server for installation. Java is required to install the NVIDIA GRID License Server. The package comes in a **`.zip`** file.



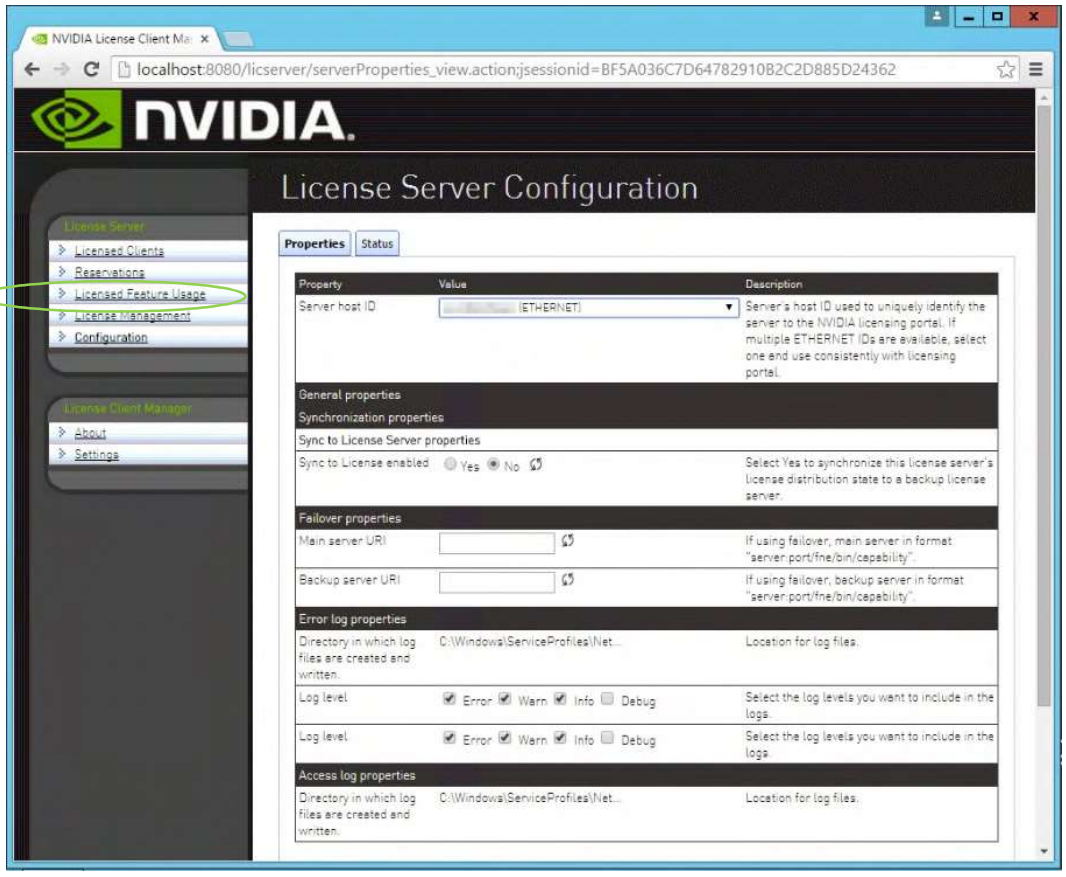
IMPORTANT: The **`.bin`** file downloaded to your server is valid for **24 hours**. If you do not install the **`.bin`** file within those 24 hours, you need to download the file again.

7. Unzip the license server installer.

8. Run **`setup.exe`** and follow the installation wizard.



9. Go to http://<FQDN of the license Server>:8080/licserver to display the License Server Configuration page. You will need the License Server's MAC Address to generate a license **`.bin`** file on the portal.



10. Select Configuration from the menu in the left pane.
11. Use the License Server Configuration menu to install the **.bin** file:
    a) Select Choose File.
    b) Use the file browser to locate the **.bin** file downloaded from the licensing portal web site.

GPU Mode Switch Utility

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

The utility can be run in a Windows 64-bit command prompt or in a Linux 32/64-bit shell, including Citrix XenServer dom0).

If the server platform hosting the Tesla GPUs does not natively run any of the supported environments, we recommend temporarily booting Linux on the server. The package includes bootable Linux images for this purpose.

To switch a GPU's mode, use the --**gpumode** command. For example:

- **gpumodeswitch --gpumode** graphics switches all the GPUs to graphics mode
- **gpumodeswitch --gpumode** compute switches all GPUs to compute mode.

A reboot is required after the mode has been changed.

NVIDIA Tesla M10 ships and support graphics mode only.

NVIDIA Tesla M60 ships in graphics mode but supports both graphics and compute modes.

INSTALL HOST DRIVER

Before NVIDIA GRID vGPU-enabled Guests can be configured, the NVIDIA GRID Host software must be installed into the Hypervisor. Citrix XenServer utilizes an **.rpm** file and VMware vSphere ESXi utilizes a **.vib** file.

Use the following procedure to install the host diver:

1. Copy the NVIDIA Driver file to the Host Server.
2. Install the Driver into the Hypervisor.
3. Reboot the Host Server.
4. Verify successful installation.

Hypervisor-specific Steps

1. Copy the **.vib** file for vSphere to the ESXi host and put the host in Maintenance Mode:
   ```
   $ esxcli system maintenanceMode set --enable true
   ```
2. Install the **Nvidia vib** file by running the following command:
   ```
   $ esxcli software vib install -v /tmp/NVIDIA**.vib
   ```

   NVIDIA** should be the full NVIDIA **vib** file name.

3. Exit out of the maintenance mode using the following command:
   ```
   $ esxcli system maintenanceMode set --enable false
   $ reboot
   ```
   The ESXi host is rebooted.
4. Verify the NVIDIA GPU card after the reboot using the following command:
   ```
   $ nvidia-smi
   ```

For XenServer, the GRID vGPU Manager runs in XenServer's Control Domain (dom0). It is installed as an RPM file which is copied to the XenServer and installed.

1. Copy the RPM file to the Host Server
2. Use the rpm command to install the package:
   ```
   [root@xenserver ~]# rpm -iv NVIDIA-**.rpm
   ```
3. Reboot the XenServer host
   ```
   [root@xenserver ~]# shutdown -r now
   ```
4. Verify that the GRID package has loaded after the XenServer host has rebooted
   ```
   [root@xenserver ~]#lsmod |grep nvidia
   nvidia 8152994 0
   i2c_core 20294 2 nvidia,i2c_
   ```

Refer to the appropriate *NVIDIA GRID vGPU Deployment Guide* for detailed instructions.

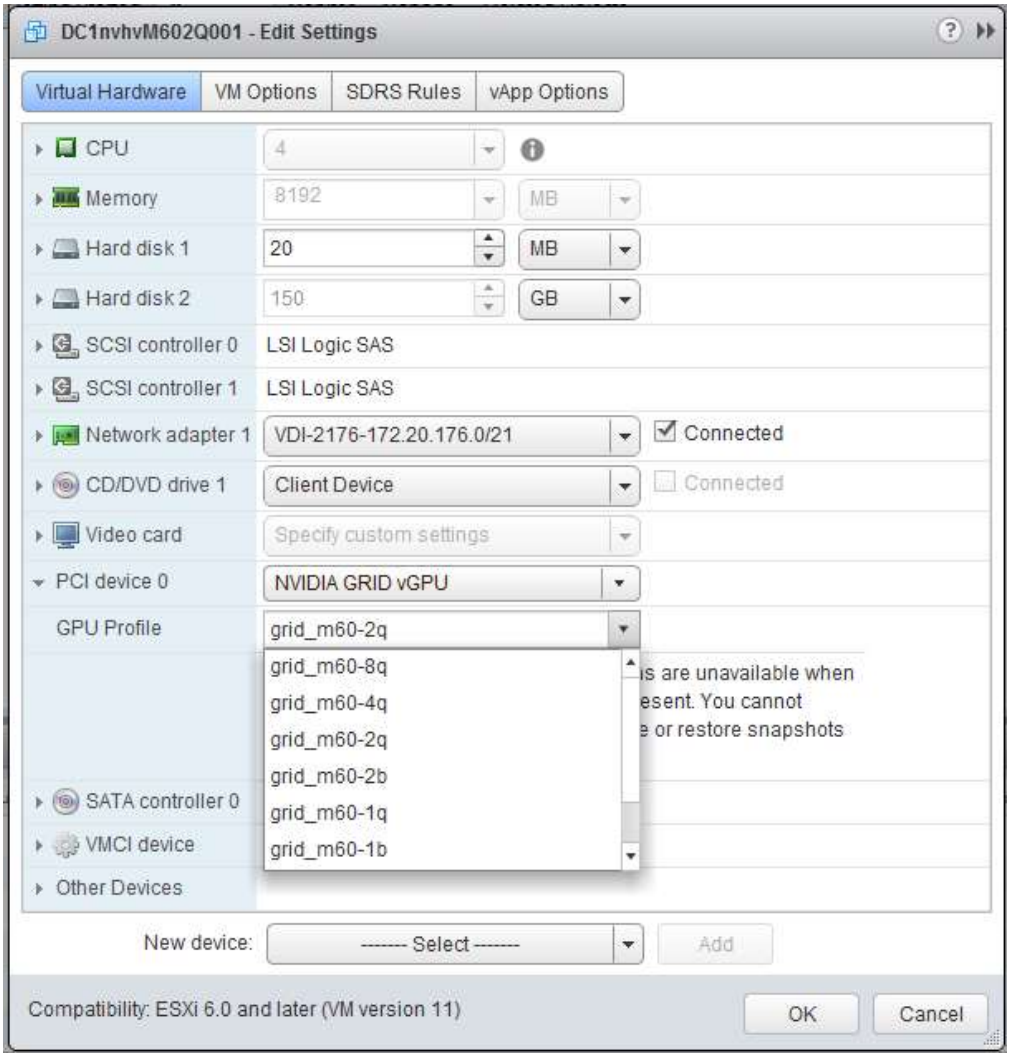
Attach GRID vGPU Profile to a VM

To have an attached GRID vGPU profile, you must configure the Virtual Machine Hardware.

vSphere

For vSphere, use the following procedure:

1. Open vCenter.
2. Go to **Edit VM** settings.
3. Add **Shared PCI Device** under **New Device**.
4. Select **Add**.
5. Select **GPU profile**, then select **OK**.
6. Start VM.



XenServer

For XenServer, use the following procedure:

1. Select New VM using XenCenter and step through the New VM wizard.
2. Select the Installation Media, Home Server, CPU & Memory option. GPU-enabled servers display a GPU configuration page
3. Select either Passthrough whole or a vGPU type from the GPU Type dropdown list (unavailable vGPU types are greyed out).
4. Click Next to configure Storage and Networking
5. Select Create Now when your configuration has completed.

Install the NVIDIA driver into a VM

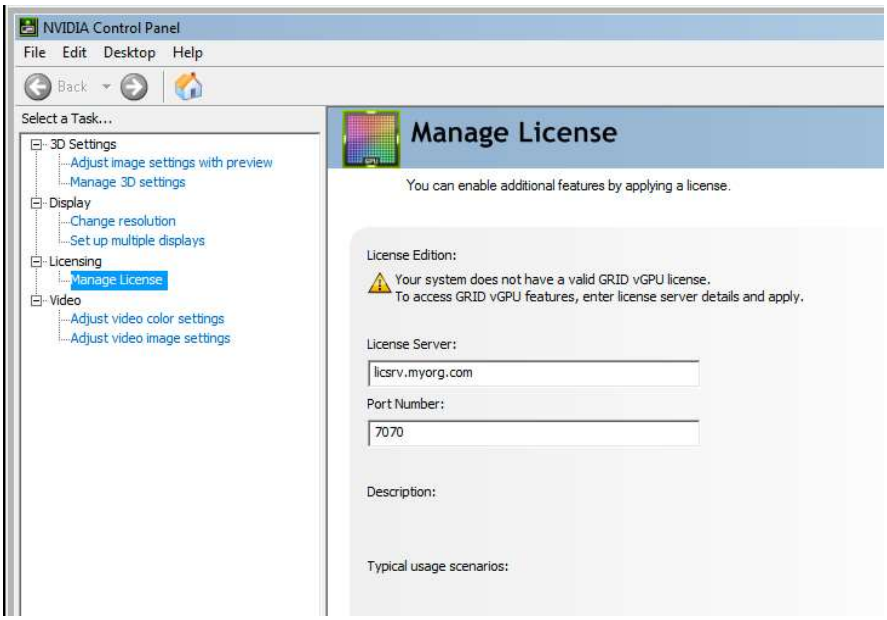
Before the VM can utilize the GRID vGPU in your VM, you need to install the NVIDIA Graphics Driver into the Guest OS.

1. Install GRID vGPU driver.
2. Select Custom (Advanced).
3. Select Clean installation.
4. Restart VM and connect to VM.

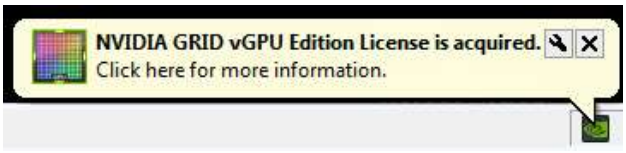
Configure VM for an NVIDIA GRID vGPU License

To enable VM access to an NVIDIA GRID vGPU license, you need to configure the Manage License Feature from the NVIDIA Control Panel (right-click on your desktop to access the NVIDIA Control Panel). Select Manage License from the left pane and enter the license server address and port. These settings can be preloaded through register keys. The license settings persist across reboots. A license is obtained prior to the user logging into the VM.

1. Go to the NVIDIA Control Panel (right-click on your desktop).



2. Select Manage License from the left pane.
3. Enter your license server address and port.
   These settings can be preloaded through register keys. The license settings persist across reboots. A license is obtained prior to the user logging into the VM.
4. Select Apply.



Congratulations. You have now completed the installation of the NVIDIA GRID Enterprise software.

Notice

The information provided in this specification is believed to be accurate and reliable as of the date provided. However, NVIDIA Corporation (“NVIDIA”) does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This publication supersedes and replaces all other specifications for the product that may have been previously supplied.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and other changes to this specification, at any time and/or to discontinue any product or service without notice. Customer should obtain the latest relevant specification before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer. NVIDIA hereby expressly objects to applying any customer general terms and conditions with regard to the purchase of the NVIDIA product referenced in this specification.

NVIDIA products are not designed, authorized or warranted to be suitable for use in medical, military, aircraft, space or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer’s own risk.

NVIDIA makes no representation or warranty that products based on these specifications will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer’s product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this specification. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this specification, or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this specification. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA. Reproduction of information in this specification is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, “MATERIALS”) ARE BEING PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, NVIDIA GRID, and NVIDIA Tesla are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2015, 2016 NVIDIA Corporation. All rights reserved.