gpumodeswitch

User Guide

DU-07764-001 _v12.0 through 12.4 Revision 02 | December 2021

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction to gpumodeswitch

gpumodeswitch is a command-line tool that is used to switch supported NVIDIA GPUs between compute and graphics mode. This chapter describes these modes and when to use them. Using gpumodeswitch describes how to use gpumodeswitch.

1.1. Compute and graphics mode

Tesla M60 and M6 GPUs support compute mode and graphics mode. NVIDIA vGPU requires GPUs that support both modes to operate in graphics mode.

Note: Even in compute mode, Tesla M60 and M6 GPUs do **not** support NVIDIA Virtual Compute Server vGPU types.

Recent Tesla M60 GPUs and M6 GPUs are supplied in graphics mode. However, your GPU might be in compute mode if it is an older Tesla M60 GPU or M6 GPU, or if its mode has previously been changed.

If your GPU supports both modes but is in compute mode, you must use the gpumodeswitch tool to change the mode of the GPU to graphics mode. If you are unsure which mode your GPU is in, use the gpumodeswitch tool to find out the mode as explained in Listing current GPU modes.

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

Compute mode is a configuration that is optimized for high-performance computing (HPC) applications as shown in Table 1.

While compute mode is optimal for HPC usage, it can cause compatibility problems with OS and hypervisors when the GPU is used primarily as a graphics device:

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

Graphics mode is a configuration that is optimized to address these problems as shown in Table 2.

Table 1. Compute mode settings

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

Table 2. Graphics mode settings

Setting	Value	Notes
Classcode	VGA Controller	This classcode indicates to OS that the GPU can function as a primary display device.
Memory BAR	256 megabytes	The GPUs expose a smaller memory BAR for direct access to the frame buffer.
I/O base BAR	Enabled	The GPU exposes an I/O BAR to claim the resources require to operate as a VGA controller.
ECC protection	Disabled	ECC protection is disabled by default, though it can still be enabled by use of the `nvidia-smi` management tool

1.2. When to use graphics mode

We recommend that graphics mode be used whenever supported Tesla products are used in the following scenarios:

- GPU pass-through with hypervisors that do not support large BARs. At the time of publication, this includes Citrix Hypervisor 6.2, 6.5, VMware ESXi 5.1, 5.5, 6.0, Red Hat Enterprise Linux 7.0, 7.1.
- GPU pass-through to Windows VMs on Xen and KVM hypervisors.
- NVIDIA vGPU deployments.

- VMware vSGA deployments.

Note: For the latest information on compatibility with compute and graphics modes, consult the release notes for supported hypervisors at NVIDIA Virtual GPU Software Documentation.

1.3. Supported products

The gpumodeswitch utility is supported **only** on the following products:

- Tesla M60
- Tesla M6

Note: Other GPUs that support NVIDIA vGPU software do not require or support mode switching.

Chapter 2. Using gpumodeswitch

gpumodeswitch is a command line utility that runs on Windows, Linux, or VMware ESXi. This chapter describes how to use gpumodeswitch, and the optional Linux boot packages included with gpumodeswitch.

2.1. Supported environments

gpumodeswitch can be run in the following environments

- Windows 64-bit command prompt
- Linux 32/64-bit shell (including Citrix Hypervisor dom0)
- VMware ESXi hypervisor

Note: 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 gpumodeswitch release package includes bootable Linux images for this purpose. These images are described in Booting from a Linux bootable image.

2.2. What's in the package

The gpumodeswitch package contains these files:

gpumodeswitch.exe
Windows executable

nvflsh64.sys
Windows 64-bit driver module

gpumodeswitch
Linux executable, also usable on Citrix Hypervisor dom0

NVIDIA-GpuModeSwitch-1OEM.600.0.0.2494585.x86_64.vib
vSphere Installation Bundle (VIB) for VMware ESXi

gpumodeswitch.iso
Bootable Linux ISO image

gpumodeswitch.zip
Bootable Linux image for use with USB storage (e.g. a flash key)

2.3. Installing the gpumodeswitch VIB

If you are using VMware ESXi as your hypervisor, you must install the gpumodeswitch VIB before attempting to run gpumodeswitch.

To install the VIB, you need to access the ESXi host through the ESXi Shell or secure shell (SSH). For information about how to enable ESXi Shell or SSH for an ESXi host, see the VMware documentation.

1. Put the ESXi host into maintenance mode.
   ```
   # vim-cmd hostsvc/maintenance_mode_enter
   ```
2. If an NVIDIA driver is already installed on the ESXi host, remove the driver.
   a). Get the name of the VIB package that contains the NVIDIA driver.
   ```
   # esxcli software vib list | grep -i nvidia
   ```
   b). Remove the VIB package that contains the NVIDIA driver.
   ```
   # esxcli software vib remove -n NVIDIA-driver-package
   ```
   NVIDIA-driver-package is the VIB package name that you got in the previous step.
3. Run the esxcli command to install the VIB.
   ```
   # esxcli software vib install --no-sig-check -v directory/NVIDIA-GpuModeSwitch-1OEM.600.0.0.2494585.x86_64.vib
   ```
   directory is the path to the directory that contains the VIB file.
4. Take the host out of maintenance mode.
   ```
   # vim-cmd hostsvc/maintenance_mode_exit
   ```
5. Reboot the ESXi host.
   ```
   # reboot
   ```

You can now run gpumodeswitch to switch the modes of your GPUs.
After switching the modes of your GPUs, continue with your NVIDIA vGPU set up by completing these tasks:

1. Uninstalling the gpumodeswitch VIB as explained in Uninstalling the gpumodeswitch VIB
2. Installing the NVIDIA virtual GPU manager for your hypervisor as explained in *Virtual GPU Software User Guide*

2.4. Running gpumodeswitch

gpumodeswitch supports these operations:

- Listing the current mode of GPUs in the host
- Changing the mode of individual or all GPUs

Note: gpumodeswitch lists and changes the modes only of GPUs that are not marked for passthrough to the VMs.

For each operation, gpumodeswitch writes a log file that contains information about the operation. For details, see Log files created by gpumodeswitch.

2.4.1. Prerequisites for running gpumodeswitch

Before running gpumodeswitch, ensure that the prerequisites for your environment are met:

- For Linux and VMware ESXi, ensure that you can run gpumodeswitch as root.
- For Citrix Hypervisor dom0 and Linux KVM hosts, ensure that no VMs are active on passthrough.
- For Windows, ensure that the following prerequisites are met:
  - You have administrator permissions (see Figure 2).
  - The nvflsh64.sys driver is in the same directory as the gpumodeswitch executable.

2.4.2. Listing current GPU modes

To list the current mode of all GPUs in the system, use --listgpumodes:

```
# gpumodeswitch --listgpumodes
NVIDIA GPU Mode Switch Utility Version 1.02
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.

PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
Adapter: PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (47,8747) : uChip 25AA320A 1.8-5.5V 4Kx8S, page
GPU Mode: N/A

Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
GPU Mode: Graphics

Tesla M60            (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:82:SP16
 S:00,B:84,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
GPU Mode: Compute
#
```

2.4.3. Switching GPU modes

To switch a GPU's mode, use the --gpumode command:

- **--gpumode graphics** switches to graphics mode

- **--gpumode compute** switches to compute mode

Note: After a GPU mode switch, the server platform should be rebooted to ensure that the modified resources of the GPU are correctly accounted for by any OS or hypervisor running on the platform.

2.4.3.1. Switching all GPUs interactively

By default, the command works on all supported GPUs in the host.

To switch all GPUs interactively, when prompted, type **y** to confirm the mode switch:

```
# gpumodeswitch --gpumode graphics
NVIDIA GPU Mode Switch Utility Version 1.02
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.

Update GPU Mode of all adapters to "graphics"?
Press 'y' to confirm or 'n' to choose adapters or any other key to abort:

y
Updating GPU Mode of all eligible adapters to "graphics"

PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
Adapter: PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (47,8747) : uChip 25AA320A 1.8-5.5V 4Kx8S, page
Cannot set GPU mode for this adapter

Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page

Programming UPR setting for requested mode..
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
 process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page

NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.7C.00.00 to 84.04.7C.00.00.

A reboot is required for the update to take effect.

InfoROM image updated successfully.

Tesla M60            (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:82:SP16
 S:00,B:84,PCI,D:00,F:00
```

```
Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page


Programming UPR setting for requested mode..
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
 process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
.................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.7C.00.00 to 84.04.7C.00.00.

A reboot is required for the update to take effect.

InfoROM image updated successfully.
#
```

2.4.3.2. Switching all GPUs without interaction

To switch all supported GPUs in the host without confirming the mode switch, use the `--auto` command.

```
# gpumodeswitch --gpumode graphics --auto

NVIDIA GPU Mode Switch Utility Version 1.23.0
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.

NOTE: Unconfigured display adapter found, device not accessible:
      PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:03,PCI,D:00,F:00
NOTE: Unconfigured display adapter found, device not accessible:
      PLX (8747h)          (10B5,8747,10B5,8747) H:82:SP8 S:00,B:83,PCI,D:00,F:00

Tesla M60            (10DE,13F2,10DE,113A) H:04:SP8 S:00,B:05,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:04:SP8 S:00,B:05,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page


Programming UPR setting for requested mode..
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
 process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
.................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.85.00.00 to 84.04.85.00.00.
```

```
A reboot is required for the update to take effect.

InfoROM image updated successfully.

Tesla M60            (10DE,13F2,10DE,113A) H:04:SP16 S:00,B:06,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:04:SP16
 S:00,B:06,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page


Programming UPR setting for requested mode..
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
 process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
.................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.85.00.00 to 84.04.85.00.00.

A reboot is required for the update to take effect.

InfoROM image updated successfully.

PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
Adapter: PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (47,8747) : uChip 25AA320A 1.8-5.5V 4Kx8S, page
Cannot set GPU mode for this adapter

Tesla M60            (10DE,13F2,10DE,113A) H:84:SP8 S:00,B:85,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:84:SP8 S:00,B:85,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page


Programming UPR setting for requested mode..
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
 process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
.................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.85.00.00 to 84.04.85.00.00.
```

```
A reboot is required for the update to take effect.

InfoROM image updated successfully.

Tesla M60            (10DE,13F2,10DE,113A) H:84:SP16 S:00,B:86,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:84:SP16
 S:00,B:86,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page


Programming UPR setting for requested mode..
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
 process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.85.00.00 to 84.04.85.00.00.

A reboot is required for the update to take effect.

InfoROM image updated successfully.

#
```

2.4.3.3. Switching individual GPUs

To switch the mode of an individual GPU, type **n** when prompted, then enter the index of the GPU you want to switch:

```
# gpumodeswitch --gpumode graphics

NVIDIA GPU Mode Switch Utility Version 1.02
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.

Update GPU Mode of all adapters to "graphics"?
Press 'y' to confirm or 'n' to choose adapters or any other key to abort:
n
Select display adapter:
<0> PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
<1> Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
<2> Tesla M60            (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
<3> PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:85,PCI,D:00,F:00
<4> PLX (8747h)          (10B5,8747,10B5,8747) H:86:SP8 S:00,B:87,PCI,D:00,F:00
<5> GRID K520            (10DE,118A,10DE,100D) H:88:SP8 S:00,B:89,PCI,D:00,F:00
<6> GRID K520            (10DE,118A,10DE,100D) H:88:SP16 S:00,B:8A,PCI,D:00,F:00
Select a number (ESC to quit): 1

Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
Adapter: Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page

Programming UPR setting for requested mode..
```

```
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
 process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.7C.00.00 to 84.04.7C.00.00.

A reboot is required for the update to take effect.

InfoROM image updated successfully.
#
```

2.4.4. Log files created by gpumodeswitch

For each operation that it performs, gpumodeswitch creates a log file that contains information about the operation. The file name and content of the log file depend on the operation.

Operation	Command	Log File Name	Log File Contents
List GPU modes	--listgpumode	listgpumodes.txt	A summary of GPU modes
Switch GPU modes	--gpumode	setgpumode.txt	A summary of the command executed

2.4.4.1. Log file for listing GPU modes

The --listgpumode command writes GPU mode information to a log file named listgpumodes.txt.

The log file summarizes GPU modes:

```
# more /tmp/listgpumodes.txt
GPU ID:
PLX (8747h)           (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
GPU Mode: N/A
GPU ID:
Tesla M60             (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
GPU Mode: Graphics
GPU ID:
Tesla M60             (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
GPU Mode: Graphics
#
```

2.4.4.2. Log file for switching GPU modes

The --gpumode command writes GPU mode update information to a log file named setgpumode.txt.

The log file summarizes the command execution:

```
# more /tmp/setgpumode.txt
GPU ID:
```

```
PLX (8747h)           (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
Cannot set GPU mode for this adapter
GPU ID:
Tesla M60             (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
Successfully updated GPU mode to graphics.
GPU ID:
Tesla M60             (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
Successfully updated GPU mode to graphics.
#
```

2.4.4.3. Changing the log file directory

By default, gpumodeswitch creates the log file in the current working directory if the current working directory is writable.

To change the directory in which the log file is created, use the --outfilepath option.

The following example commands create the log file in the /tmp directory:

- Listing current GPU modes:

```
# gpumodeswitch --listgpumodes --outfilepath /tmp/
```

- Switching GPUs to graphics mode:

```
# gpumodeswitch --gpumode graphics --outfilepath /tmp
```

Note: When using the bundled Linux bootable images (see Booting from a Linux bootable image), the system boots into a read-only file system. Log files can be generated by using the outfilepath option to write the log files into /tmp.

2.4.5. Next Steps

After switching the modes of your GPUs, continue with your NVIDIA vGPU set up as follows:

1. If you are using VMware ESXi as your hypervisor, uninstall the gpumodeswitch VIB as explained in Uninstalling the gpumodeswitch VIB.
2. Install the NVIDIA virtual GPU manager for your hypervisor explained in *Virtual GPU Software User Guide*.

2.5. Uninstalling the gpumodeswitch VIB

If you are using VMware ESXi as your hypervisor, you must uninstall the gpumodeswitch VIB after running gpumodeswitch.

To uninstall the VIB, you need to access the ESXi host through the ESXi Shell or secure shell (SSH). For information about how to enable ESXi Shell or SSH for an ESXi host, see the VMware documentation.

1. Put the ESXi host into maintenance mode.

```
# vim-cmd hostsvc/maintenance_mode_enter
```

2. Run the esxcli command to uninstall the VIB.

```
# esxcli software vib remove -n NVIDIA-VMware_ESXi_6.0_GpuModeSwitch_Driver
```

3. Take the host out of maintenance mode.

```
# vim-cmd hostsvc/maintenance_mode_exit
```

4. Reboot the ESXi host.

```
# reboot
```

After uninstalling the gpumodeswitch VIB, continue with your NVIDIA vGPU set up by installing the NVIDIA virtual GPU manager for your hypervisor as explained in *Virtual GPU Software User Guide*.

2.6. Booting from a Linux bootable image

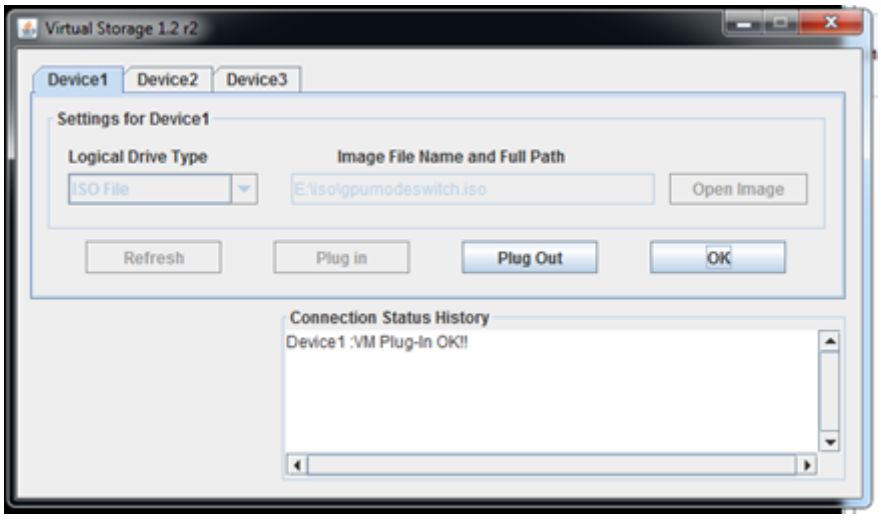
If your server platform is not running one of the environments that supports gpumodeswitch, use one of the Linux bootable images included in the release to boot your server to a minimal Linux environment that includes gpumodeswitch.

2.6.1. Booting directly from ISO

The gpumodeswitch.iso file in the release package is intended for direct boot on a server platform, using the server's remote management capability.

1. Connect the ISO file as an emulated storage device on the server.
2. Reboot the server.
3. Use the BIOS boot menu to select the emulated device for boot.

Figure 1. Connecting to gpumodeswitch.iso through server remote management



The ISO image boots to a Linux shell prompt from which gpumodeswitch can be run directly. For instructions, see Running gpumodeswitch.

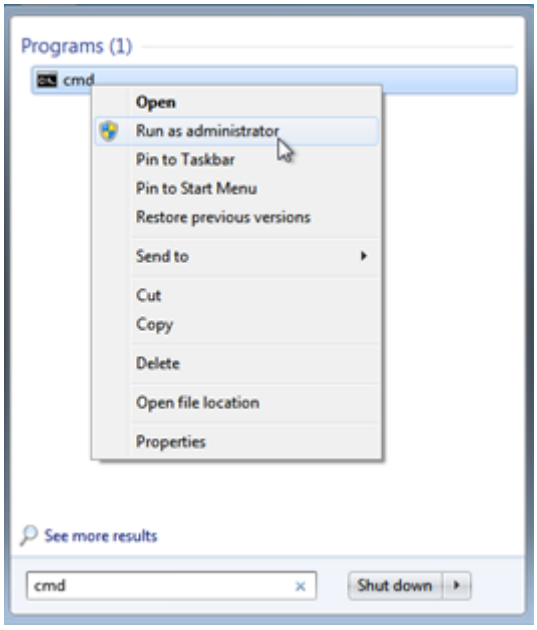
2.6.2. Booting from a USB flash key

The gpumodeswitch.zip file in the release package is intended to be unpacked onto a USB flash key, which can then be used to boot the server.

On a Windows system, follow these steps to create a bootable USB key:

1. Connect a USB flash key of at least 64 megabytes in size.
2. In **Windows Explorer**, right-click on the USB drive and select **Format** to format it.
3. Note the driver letter that is assigned to the USB drive (for example, E:).
4. Unzip gpumodeswitch.zip to the freshly formatted USB drive.
   a). Right-click the zip file and select **Extract All**.
   b). Browse to the USB drive's letter, and then click **OK** to unzip the contents.
5. Open a Windows **Command Prompt** window with Administrator privileges.
   a). Search for cmd.
   b). Right-click on the cmd program icon and select **Run as administrator**.

Figure 2. Launching an administrator command prompt



6. In the **Command Prompt** window, change to the USB drive by typing its drive letter, (for example, **e:**) and then pressing **Return**.
7. Run the following command, replacing **e:** with your USB drive's letter.

```
syslinux -m -a e:
```

8. Close the **Command Prompt** window, and eject and unplug the USB drive.
9. Boot the server from the USB drive.

The image boots to a Linux shell prompt from which gpumodeswitch can be run directly. For instructions, see Running gpumodeswitch.

2.7. Troubleshooting gpumodeswitch

Some common problems can cause the gpumodeswitch command to fail or to be unavailable.

2.7.1. Unloading an existing NVIDIA driver

You cannot run gpumodeswitch in a non-Windows environment where an existing NVIDIA driver is already loaded on the GPU.

In a Linux environment or a Citrix Hypervisor environment, the gpumodeswitch command fails and an error is reported:

```
[root@xenserver ~]# ./gpumodeswitch --listgpumodes
NVIDIA GPU Mode Switch Utility Version 1.02
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.
...
ERROR: In order to avoid the irreparable damage to your graphics
adapter it is necessary to unload the NVIDIA kernel driver first:
                rmmod nvidia
[root@xenserver ~]#
```

In a VMware ESXi environment, the gpumodeswitch command is not available.

2.7.1.1. Unloading an existing NVIDIA driver in a Linux environment

1. Halt any services that are using the GPU.
2. Unload the NVIDIA driver.

```
[root@linux ~]# rmmod nvidia
```

2.7.1.2. Unloading an existing NVIDIA driver in a Citrix Hypervisor environment

1. Halt any VMs that are using the GPU.
2. Stop the Citrix Hypervisor gpumon service.

```
[root@xenserver ~]# service xcp-rrdd-gpumon stop
Stopping XCP RRDD plugin xcp-rrdd-gpumon:                  [  OK  ]
```

3. Unload the NVIDIA kernel driver.

```
[root@xenserver ~]# rmmod nvidia
```

2.7.1.3. Unloading an existing NVIDIA driver in a VMware ESXi environment

Remove the driver and install the gpumodeswitch VIB as explained in Installing the gpumodeswitch VIB.

Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. 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 document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information 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 ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

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 this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform 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 document. NVIDIA accepts no 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 document 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 document. 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 document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND 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. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. 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 Terms of Sale for the product.

VESA DisplayPort

DisplayPort and DisplayPort Compliance Logo, DisplayPort Compliance Logo for Dual-mode Sources, and DisplayPort Compliance Logo for Active Cables are trademarks owned by the Video Electronics Standards Association in the United States and other countries.

HDMI

HDMI, the HDMI logo, and High-Definition Multimedia Interface are trademarks or registered trademarks of HDMI Licensing LLC.

OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.

Trademarks

NVIDIA, the NVIDIA logo, NVIDIA GRID, NVIDIA GRID vGPU, NVIDIA Maxwell, NVIDIA Pascal, NVIDIA Turing, NVIDIA Volta, GPUDirect, Quadro, and Tesla are trademarks 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

NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051
http://www.nvidia.com