GRID VGPU FOR VMWARE VSPHERE
Version 361.45.09 / 362.56

RN-07347-001 | May 23rd, 2016

Release Notes

CONTENTS

RELEASE NOTES

These Release Notes summarize current status, information on validated platforms, and known issues with NVIDIA GRID Virtual GPU software and hardware on VMware vSphere.

The following software is included in this release:

- NVIDIA GRID Virtual GPU Manager version 361.45.09 for VMware vSphere 6.0 Hypervisor (ESXi)
- NVIDIA Windows drivers for vGPU version 362.56
- NVIDIA Linux drivers for vGPU version 361.45.09

> **Note:** The GRID vGPU Manager and Windows guest VM drivers must be installed together. Older VM drivers will not function correctly with this release of GRID vGPU Manager. Similarly, older GRID vGPU Managers will not function correctly with this release of Windows guest drivers. See "VM running older NVIDIA vGPU drivers fails to initialize vGPU when booted," on page 4.

Updates in this release:

- Miscellaneous bug fixes

VALIDATED PLATFORMS

This release of virtual GPU provides support for NVIDIA GRID K1, GRID K2, Tesla M6 and Tesla M60 GPUs on VMware vSphere 6.0 Hypervisor (ESXi), running on validated server hardware platforms. For a list of validated server platforms, refer to http://www.nvidia.com/buygrid.

SOFTWARE VERSIONS

This release has been tested with the following software versions:

Software	Version tested
VMware vSphere Hypervisor (ESXi)	6.0 RTM build 2494585, 6.0 update 1, 6.0 update 2
VMware Horizon	6.2.1 RTM build 3268071 7.0 RTM build 3618085
VMware vCenter Server	6.0 RTM build 2562643

Linux support

GRID vGPU with Linux guest VMs is supported on Tesla M60 and M6, with the following distributions:

- Red Hat Enterprise Linux 6.6, 7
- CentOS 6.6, 7
- Ubuntu 12.04, 14.04 LTS

HARDWARE CONFIGURATION

Tesla M60 and M6 GPUs support compute and graphics modes, which can be configured using the `gpumodeswitch` tool provided with GRID software releases. GRID vGPU requires that M60 / M6 are configured in graphics mode.

KNOWN PRODUCT LIMITATIONS

Known product limitations for this release of NVIDIA GRID are described in the following sections:

VM RUNNING OLDER NVIDIA VGPU DRIVERS FAILS TO INITIALIZE VGPU WHEN BOOTED

Description

A VM running older NVIDIA drivers, such as those from a previous vGPU release, will fail to initialize vGPU when booted on a `vSphere` platform running the current release of GRID Virtual GPU Manager.

In this scenario, the VM boots in standard VGA mode with reduced resolution and color depth. The NVIDIA GRID GPU is present in Windows Device Manager but displays a warning sign, and a device status of "Windows has stopped this device because it has reported problems. (Code 43)".

The VM's log file reports one of the following errors:

- A version mismatch between guest and host drivers:

```
vthread-10| E105: vmiop_log: Guest VGX version(2.0) and Host VGX
version(2.1) do not match
```

- A signature mismatch:

```
vthread-10| E105: vmiop_log: VGPU message signature mismatch.
```

Resolution

Install the latest NVIDIA vGPU release drivers in the VM.

VIRTUAL GPU FAILS TO START IF ECC IS ENABLED

Description

GRID K2, Tesla M60, and Tesla M6 support ECC (error correcting code) for improved data integrity. If ECC is enabled, virtual GPU fails to start. The following error is logged in the VM's log file:

```
vthread10|E105: Initialization: VGX not supported with ECC Enabled.
```

Virtual GPU is not currently supported with ECC active. GRID K2 cards and Tesla M60, M6 cards in graphics mode ship with ECC disabled by default, but ECC may subsequently be enabled using `nvidia-smi`.

Resolution

Use `nvidia-smi` to list the status of all GPUs, and check for ECC noted as enabled on GPUs. Change the ECC status to off on a specific GPU by executing the following command:

```
nvidia-smi -i id -e 0
```

id is the index of the GPU as reported by `nvidia-smi`.

SINGLE VGPU BENCHMARK SCORES ARE LOWER THAN PASSTHROUGH GPU

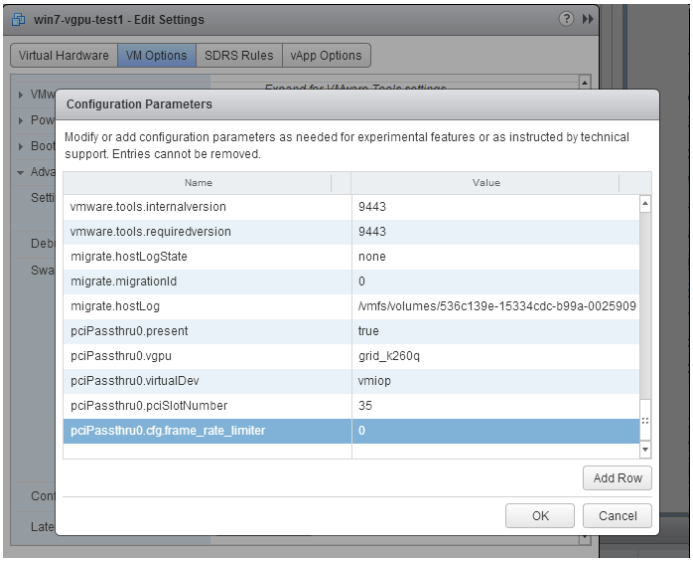
Description

A single vGPU configured on a physical GPU produces lower benchmark scores than the physical GPU run in passthrough mode.

Aside from performance differences that may be attributed to a vGPU's smaller framebuffer size, vGPU incorporates a performance balancing feature known as Frame Rate Limiter (FRL), which is enabled on all vGPUs. FRL is used to ensure balanced performance across multiple vGPUs that are resident on the same physical GPU. The FRL setting is designed to give good interactive remote graphics experience but may reduce scores in benchmarks that depend on measuring frame rendering rates, as compared to the same benchmarks running on a passthrough GPU.

Resolution

FRL is controlled by an internal vGPU setting. NVIDIA does not validate vGPU with FRL disabled, but for validation of benchmark performance, FRL can be temporarily disabled by adding the configuration parameter `pciPassthru0.cfg.frame_rate_limiter` in the VM's advanced configuration options (select `Edit Settings`, select the `VM Options` tab, expand the `Advanced` dropdown, select `Edit Configuration`, select `Add Row`, then manually enter the parameter name and set it to `0`. (This setting can only be changed when the VM is powered off.)



With this setting in place, the VM's vGPU will run without any frame rate limit. The FRL can be reverted back to its default setting by setting `pciPassthru0.cfg.frame_rate_limiter` to `1` or by removing the parameter from the advanced settings.

VMS CONFIGURED WITH LARGE MEMORY FAIL TO INITIALIZE VGPU WHEN BOOTED

Description

When starting multiple VMs configured with large amounts of RAM (typically more than 32GB per VM), a VM may fail to initialize vGPU. In this scenario, the VM boots in VMware SVGA mode and doesn't load the NVIDIA driver. The NVIDIA GRID GPU is present in Windows Device Manager but displays a warning sign, and a device status of "Windows has stopped this device because it has reported problems. (Code 43)".

The VM's log file contains these error messages:

```
vthread10|E105: NVOS status 0x29
vthread10|E105: Assertion Failed at 0x7620fd4b:179
```

```
vthread10|E105: 8 frames returned by backtrace
...
vthread10|E105: VGPU message 12 failed, result code: 0x29
...
vthread10|E105: NVOS status 0x8
vthread10|E105: Assertion Failed at 0x7620c8df:280
vthread10|E105: 8 frames returned by backtrace
...
vthread10|E105: VGPU message 26 failed, result code: 0x8
```

Resolution

vGPU reserves a portion of the VM's framebuffer for use in GPU mapping of VM system memory. The reservation is sufficient to support up to 32GB of system memory, and may be increased to accommodate up to 64GB by adding the configuration parameter `pciPassthru0.cfg.enable_large_sys_mem` in the VM's advanced configuration options (select `Edit Settings`, select the `VM Options` tab, expand the `Advanced` dropdown, select `Edit Configuration`, select `Add Row`, then manually enter the parameter name and set it to `1`. (This setting can only be changed when the VM is powered off.)

With this setting in place, less GPU framebuffer is available to applications running in the VM. To accommodate system memory larger than 64GB, the reservation can be further increased by adding `pciPassthru0.cfg.extra_fb_reservation` in the VM's advanced configuration options, and setting its value to the desired reservation size in megabytes. The default value of 64M is sufficient to support 64GB of RAM. We recommend adding 2M of reservation for each additional 1GB of system memory. For example, to support 96GB of RAM, set `pciPassthru0.cfg.extra_fb_reservation` to `128`.

The reservation can be reverted back to its default setting by setting `pciPassthru0.cfg.enable_large_sys_mem` to `0`, or by removing the parameter from the advanced settings.

RESOLVED ISSUES

Bug ID	Summary and Description
200184005	**Driver upgrade from 352.83 to 361.40 fails on bare metal** Driver upgrade fails on a bare metal setup with this error message: `An NVIDIA kernel module 'nvidia' appears to already be loaded in your kernel. This may be because it is in use .....` The upgrade fails because it does not stop the nvidia-`gridd` service.
200182826	**4096x2160 resolution is not available in Windows Control Panel** 4096x2160 resolution cannot be applied through NVIDIA Control Panel and is not listed in Windows Control Panel.
200144667	**MPlayer, or other video players, fail to start when using hardware acceleration on Linux VMs running vGPU** VDPAU is currently not supported on Linux VMs running vGPU.
200138540	**Horizon does not scale over multiple monitors when connecting to Ubuntu VMs** When connecting to an Ubuntu VM in full screen mode, Horizon fails to scale over multiple monitors
1749323	**Error Powering on 4 M60_8A VMs with VIB 361.40** When a host with 4 VMs with M60_8A 8 GB profiles is powered on, only 3 VMs start. The 4th VM fails to start.
1721555	**The screen is not updated in full-screen mode with pop-up media player controls automatically hidden** If the Xbox app is used to stream video from an Xbox 360 device, the screen is correctly updated in windowed mode. But in full-screen mode, the screen freezes until the mouse is moved.

KNOWN ISSUES

NVIDIA Control Panel is killed during reconnection with a View

Description	If NVIDIA Control Panel is running while a View session is disconnected and then reconnected, NVIDIA Control Panel is killed before the View session is reconnected.
Version	
Fix	
Status	Open
Ref. #	200176969

GNOME Display Manager (GDM) fails to start on Red Hat Enterprise Linux 7.2

Description	GDM fails to start on Red Hat Enterprise Linux 7.2.
Version	

Workaround	Permanently enable permissive mode for Security Enhanced Linux (SELinux).
	1. As root, edit the `/etc/selinux/config` file to set `SELINUX` to `permissive`.
	`SELINUX=permissive`
	2. Reboot the system.
	`~]# reboot`
	For more information, see "Permissive Mode," in *Red Hat Enterprise Linux 7 SELinux User's and Administrator's Guide*.

Status	Not an NVIDIA bug
Ref. #	200167868

NVIDIA Control Panel complains that "you are not currently using a display that is attached to an Nvidia GPU"	
Description	When you launch NVIDIA Control Panel on a VM configured with vGPU, it fails to start and complains about not using a display attached to an NVIDIA GPU. This happens because Windows is using VMware's SVGA device instead of NVIDIA vGPU.
Version	
Fix	Make NVIDIA vGPU the primary display adapter. Use Windows screen resolution control panel to make the second display, identified as "2" and corresponding to NVIDIA vGPU, to be the active display and select the `Show desktop only on 2` option. Click `Apply` to accept the configuration. You may need to click on the `Detect` button for Windows to recognize the display connected to NVIDIA vGPU. **Note**: if the VMware Horizon/View agent is installed in the VM, the NVIDIA GPU is automatically selected in preference to the SVGA device.
Status	Open
Ref. #	

VM configured with more than one vGPU fails to initialize vGPU when booted	
Description	Using the current VMware vCenter user interface, it is possible to configure a VM with more than one vGPU device. When booted, the VM boots in VMware SVGA mode and doesn't load the NVIDIA driver. The additional vGPU devices are present in Windows Device Manager but display a warning sign, and a device status of "Windows has stopped this device because it has reported problems. (Code 43)".
Version	
Workaround	GRID vGPU currently supports a single virtual GPU device per VM. Remove any additional vGPUs from the VM configuration before booting the VM.
Status	Open
Ref. #	

A VM configured with both a vGPU and a passthrough GPU fails to start the passthrough GPU

Description	Using the current VMware vCenter user interface, it is possible to configure a VM with a vGPU device and a passthrough (direct path) GPU device. This is not a currently supported configuration for vGPU. The passthrough GPU appears in Windows Device Manager with a warning sign, and a device status of "Windows has stopped this device because it has reported problems. (Code 43)".
Version	
Workaround	Do not assign vGPU and passthrough GPUs to a VM simultaneously.
Status	Open
Ref. #	1735002

vGPU allocation policy fails when multiple VMs are started simultaneously

Description	If multiple VMs are started simultaneously, vSphere may not adhere to the placement policy currently in effect. For example, if the default placement policy (breadth-first) is in effect, and 4 physical GPUs are available with no resident vGPUs, then starting 4 VMs simultaneously should result in one vGPU on each GPU. In practice, more than one vGPU may end up resident on a GPU.
Version	
Workaround	Start VMs individually.
Status	Not an NVIDIA bug
Ref. #	200042690

Prior to installing Horizon agent inside a VM, the Start menu's sleep option is available.

Description	When a VM is configured with a vGPU, the Sleep option remains available in the Windows Start menu. Sleep is not supported on vGPU and attempts to use it will lead to undefined behavior.
Version	
Workaround	Do not use Sleep with vGPU. Installing the VMware Horizon agent will disable the Sleep option.
Status	Closed
Ref. #	200043405

vGPU-enabled VMs fail to start, nvidia-smi fails when VMs are configured with too high a proportion of the server's memory.

Description	If vGPU-enabled VMs are assigned too high a proportion of the server's total memory, one of more of the VMs may fail to start with the error "The available Memory resources in the parent resource pool are insufficient for the operation", and nvidia-smi run in the host shell returns this error: `-sh: can't fork` For example, on a server configured with 256G of memory, these errors may occur if vGPU-enabled VMs are assigned more than 243G of memory.
Version	
Workaround	Reduce the total amount of system memory assigned to the VMs.
Status	Closed
Ref. #	200060499

On reset/restart VMs fail to start with the error "VMIOP: no graphics device is available for vGPU…"

Description	On a system running a maximal configuration, i.e. maximum number of vGPU VMs the server can support, some VMs might fail to start post a reset or restart operation.
Version	
Fix	Upgrade to ESXi 6.0 Update 1
Status	Closed
Ref. #	200097546

nvidia-smi shows high GPU utilization for vGPU VMs with active Horizon sessions

Description	vGPU VMs with an active Horizon connection utilize a high percentage of the GPU on the ESXi host. The GPU utilization remains high for the duration of the Horizon session even if there are no active applications running on the VM.
Version	
Workaround	None
Status	Open
Ref. #	1735009

Multiple WebGL tabs in Microsoft Internet Explorer may trigger TDR on Windows VMs

Description	Running intensive WebGL applications in multiple IE tabs may trigger a TDR on Windows VMs.
Version	
Workaround	Disable hardware acceleration in IE. To enable software rendering in IE, refer Microsoft's knowledge base article at https://support.microsoft.com/en-us/kb/2528233.
Status	Open
Ref. #	200148377

BSOD with 361.40/362.13 drivers on GRID K1 cards with Windows 10

Description	If Windows 10 is the guest OS on a server with a GRID K1 card running the 361.40/362.13 drivers, the OS crashes. If the 352.83/354.80 drivers are used, the system drops the connection, but the guest OS continues to function. This issue does not occur if the guest OS is Windows 7.
Version	
Workaround	
Status	Open
Ref. #	1756897

Notice

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.

Information furnished is believed to be accurate and reliable. However, NVIDIA Corporation assumes no responsibility for the consequences of use of such information or for any infringement of patents or other rights of third parties that may result from its use. No license is granted by implication of otherwise under any patent rights of NVIDIA Corporation. Specifications mentioned in this publication are subject to change without notice. This publication supersedes and replaces all other information previously supplied. NVIDIA Corporation products are not authorized as critical components in life support devices or systems without express written approval of NVIDIA Corporation.

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 and the NVIDIA logo 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