GRID VGPU FOR VMWARE VSPHERE
Version 352.103/356.60

RN-07347-001 | March 2017

Release Notes

CONTENTS

Release Notes **1**

Validated Platforms **2**

Software versions 2

Linux support 2

Hardware configuration 3

Known Issues **4**

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.

Included in this release is NVIDIA GRID Virtual GPU Manager version 352.103 for VMware vSphere 6.0 Hypervisor (ESXi), NVIDIA Windows drivers for vGPU version 356.60, and NVIDIA Linux drivers for vGPU version 352.103.

> **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 issue "VM running older NVIDIA vGPU drivers fails to initialize vGPU when booted," on page 5.

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
VMware Horizon	6.1.1 RTM build 2769635 6.2 RTM build 2975366
VMware vCenter Server	6.0 RTM build 2562643

Linux support

GRID vGPU supports the following Linux distributions as a guest OS **only** on Tesla M60 and Tesla M6 on VMware vSphere:

- Red Hat Enterprise Linux 6.6 and later compatible 6.x versions
- Red Hat Enterprise Linux 7.0-7.2 and later compatible 7.x versions
- CentOS 6.6 and later compatible 6.x versions
- CentOS 7.0-7.2 and later compatible 7.x versions
- Ubuntu 12.04 LTS
- Ubuntu 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 ISSUES

NVENC requires at least 1 Gbyte of frame buffer	
Description	Using the frame buffer for the NVIDIA hardware-based H.264/HEVC video encoder (NVENC) may cause memory exhaustion with vGPU profiles that have 512 Mbytes or less of frame buffer. To reduce the possibility of memory exhaustion, NVENC is disabled on profiles that have 512 Mbytes or less of frame buffer. Application GPU acceleration remains fully supported and available for all profiles, including profiles with 512 Mbytes or less of frame buffer. NVENC support from VMware is a recent feature and, if you are using an older version, you should experience no change in functionality The following vGPU profiles have 512 Mbytes or less of frame buffer: • Tesla M6-0B, M6-0Q • Tesla M60-0B, M60-0Q • GRID K100, K120Q • GRID K200, K220Q
Workaround	If you require NVENC to be enabled, use a profile that has at least 1 Gbyte of frame buffer.
Status	Closed
Ref. #	1816861

The VMware VIB installer incorrectly reports that reboot is not required after installing the vGPU Manager VIB

Description	After installing the NVIDIA Virtual GPU Manager VIB for vSphere on the ESXi host, the `esxcli` command to install the VIB incorrectly reports `Reboot Required: false` in the installation result message. The instructions in *GRID Virtual GPU User Guide* for installing and updating the NVIDIA Virtual GPU Manager VIB are correct.
Workaround	Reboot the ESXi host and remove it from maintenance mode as instructed in *GRID Virtual GPU User Guide*. Ignore the line `Reboot Required: false` in the installation result message.
Status	Open
Ref. #	1816290

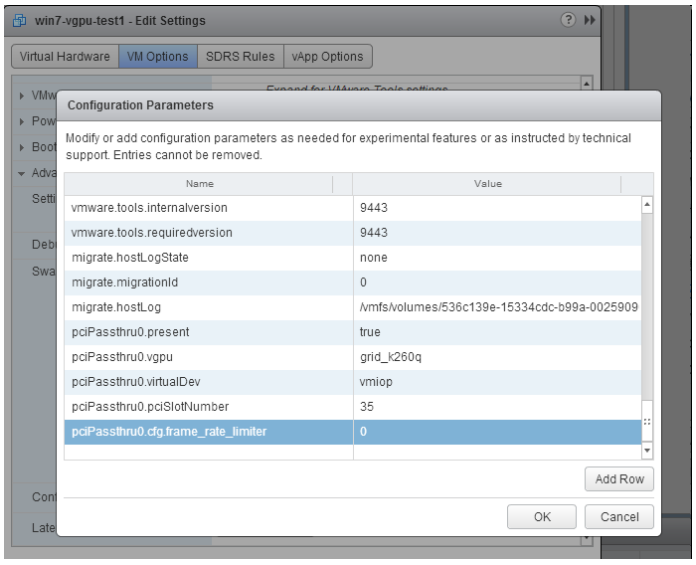
VM running older NVIDIA vGPU drivers fails to initialize vGPU when booted

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 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`
	
	Or, The VM's log file reports a signature mismatch:
	
	`vthread-10| E105: vmiop_log: VGPU message signature mismatch.`

Fix	Install the latest NVIDIA vGPU release drivers in the VM.
Status	Open
Ref. #	

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`.
Version	
Workaround	Use `nvidia-smi` to list status on all GPUs, and check for ECC noted as enabled on GPUs. Change the ECC status to off on a specific GPU by executing '`nvidia-smi -i <id> -e 0`', where `<id>` is the index of the GPU as reported by `nvidia-smi`.
Status	Open
Ref. #	

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.
Version	
Workaround	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.
Status	Open
Ref. #	

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`

Version	
Fix	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 settings 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.
Status	Fixed
Ref. #	1510886

Horizon session is unable to scale to resolutions higher than 1280x1024 on VMs running vGPU on Tesla M6 or Tesla M60	
Description	GRID vGPU is a licensed feature on Tesla M6 and M60. A software license is required to enable full vGPU features within the guest VM.
Version	
Fix	Please refer to `GRID Licensing User Guide` on how to license your vGPU VM.
Status	Open
Ref. #	

Horizon sessions run at a low frame rate on VMs running vGPU on Tesla M6 or Tesla M60	
Description	GRID vGPU is a licensed feature on Tesla M6 and M60. A software license is required to enable full vGPU features within the guest VM.
Version	
Fix	Please refer to `GRID Licensing User Guide` on how to license your vGPU VM.
Status	Open
Ref. #	

Applications fail to start on VMs running vGPU on Tesla M6 or Tesla M60	
Description	Applications fail to start within the VM, and the following error is logged in the VM's log file: `vthread-10| E105: vmiop_log: Guest is unlicensed. Cannot allocate more than 0x20 channels!` GRID vGPU is a licensed feature on Tesla M6 and M60. A software license is required to enable full vGPU features within the guest VM.
Version	
Fix	Please refer to `GRID Licensing User Guide` on how to license your vGPU VM.
Status	Open
Ref. #	

Memory exhaustion can occur with vGPU profiles that have 512 Mbytes or less of frame buffer

Description	Memory exhaustion can occur with vGPU profiles that have 512 Mbytes or less of frame buffer. This issue typically occurs when multiple display heads are used with Citrix XenDesktop or VMware Horizon on a Windows 10 guest VM. When this error occurs, the NVIDIA host driver reports Xid error 31 and Xid error 43 in the VMware vSphere log file `vmware.log` in the guest VM's storage directory. The following vGPU profiles have 512 Mbytes or less of frame buffer: ▶ Tesla M6-0B, M6-0Q ▶ Tesla M60-0B, M60-0Q ▶ GRID K100, K120Q ▶ GRID K200, K220Q
Version	
Workaround	
Status	Open
Ref. #	200130864

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 the error ""-sh: can't fork". For example, on a server configured with 256G of memory, these errors may occur if vGPU-enabled VMs are assigned more than 243G of memory.
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

MPlayer, or other video players, fail to start when using hardware acceleration on Linux VMs running vGPU

Description	VDPAU is currently not supported on Linux VMs running vGPU.
Version	
Workaround	
Status	Fixed
Ref. #	200144667

After configuring license on a vGPU VM, Horizon session fails to scale to resolutions higher than 1280x1024

Description	After configuring license on a vGPU VM over a Horizon session, the Horizon window still fails to scale to resolutions beyond 1280x1024.
Version	
Workaround	Disconnect and reconnect the Horizon session.
Status	Open
Ref. #	200138866

Horizon does not scale over multiple monitors when connecting to Ubuntu VMs

Description	Horizon when connecting to an Ubuntu VM in full screen mode fails to scale over multiple monitors.
Version	
Workaround	
Status	Fixed
Ref. #	200138540

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

After a reboot of the VM, the task bar moves to the secondary monitor when the view session is reconnected

Description	After a reboot of the VM, the task bar moves to the secondary monitor when the view session is reconnected. The issue occurs only on the first connection after the reboot while VM is rebooted from the view session.
Version	
Workaround	Change the view session from full-screen mode to window mode and then change it back to full-screen mode: 1. In the view session screen, click the  (**Restore Down**) icon to change the view session to window mode. 2. In the title bar of the view session window, click the  (**Maximize**) icon to change the view session back again to full-screen mode.
Status	Closed
Ref. #	200170401

NVIDIA driver forces Tesla M60 to be the primary display adapter	
Description	After the NVIDIA driver is installed on a Tesla M60 GPU on a server running baremetal Windows (no hypervisor), a black or blue screen is observed as the primary display switches from iGPU to Tesla M60. ▶ If the Windows default WDDM driver is installed for an iGPU device such as Matrox, a blue screen is observed. ▶ If the iGPU driver is installed, a black screen is observed.
Version	
Workaround	1. Before installing the NVIDIA driver on your server, ensure that the server is connected to the network and is accessible through remote access software such as VNC. 2. Connect to the server through IPMI or VNC and install the NVIDIA driver. 3. When the installation of the NVIDIA driver is complete, reboot the server. After you reboot the server, the IPMI display may display only a black or a blue screen. In this situation, display the OS desktop on the IPMI display by configuring the first display to be the active display. The first display is identified as **1** and corresponds to the server's onboard VGA device. 1. Connect to the server through the remote access software that you are using. 2. Open Windows Control Panel. 3. In Windows Control Panel, click **Appearance and Personalization** and then **Connect to an external display**. 4. If Windows Control Panel does not show the display that is connected to the server's onboard VGA device, click **Detect**. 5. In the **Multiple displays** list, click **Show desktop only on 1** and then click **Apply**.
Status	Not an NVIDIA bug
Ref. #	1727289

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