GRID SOFTWARE FOR VMWARE VSPHERE VERSION 367.132/370.39

Release Notes

TABLE OF CONTENTS

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | ii

www.nvidia.com
GRID Software for VMware vSphere Version   RN-07347-001 _v4.9 | iii
367.132/370.39

Chapter 1.
RELEASE NOTES

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

> The most current version of the documentation for this release of NVIDIA GRID Software can be found online at GRID 4.9 Software Documentation.

This release includes the following software:

▶ NVIDIA GRID Virtual GPU Manager version 367.132 for the VMware vSphere releases listed in Hypervisor Software Releases
▶ NVIDIA Windows drivers for vGPU version 370.39
▶ NVIDIA Linux drivers for vGPU version 367.133

> **Caution**
>
> 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.

Updates in this release:

▶ Miscellaneous bug fixes
▶ Security updates

www.nvidia.com
GRID Software for VMware vSphere Version   RN-07347-001 _v4.9 | 1
367.132/370.39

Chapter 2.
VALIDATED PLATFORMS

This release of NVIDIA GRID software provides support for several NVIDIA GPUs on validated server hardware platforms, VMware vSphere hypervisor software versions, and guest operating systems.

2.1. Supported NVIDIA GPUs and Validated Server Platforms

This release of NVIDIA GRID software provides support for the following NVIDIA GPUs on VMware vSphere, running on validated server hardware platforms:

▸ GRID K1
▸ GRID K2
▸ Tesla M6
▸ Tesla M10
▸ Tesla M60

For a list of validated server platforms, refer to NVIDIA GRID Certified Servers.

> Tesla M60 and M6 GPUs support compute mode and graphics mode. GRID vGPU requires GPUs that support both modes to operate in graphics mode.
>
> 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.
>
> To configure the mode of Tesla M60 and M6 GPUs, use the `gpumodeswitch` tool provided with GRID software releases.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 2

2.2. Hypervisor Software Releases

Supported VMware vSphere Hypervisor (ESXi) Releases

This release is supported on the VMware vSphere Hypervisor (ESXi) releases listed in the table.

> Updates to a base release of VMware vSphere Hypervisor (ESXi) are compatible with the base release and can also be used with this version of NVIDIA GRID Software unless expressly stated otherwise.

Software	Release Supported
VMware vSphere Hypervisor (ESXi)	6.5 and compatible updates
	6.0 and compatible updates

Supported Management Software and Virtual Desktop Software Releases

This release supports **only** the management software and virtual desktop software releases listed in the table.

> If a specific release, even an update release, is not listed, it's **not** supported.

Software	Version Tested
VMware Horizon	7.9 and compatible 7.9.*x* updates
	7.7 and compatible 7.7.*x* updates
	7.5 and compatible 7.5.*x* updates
	6.2.1 RTM build 3268071
VMware vCenter Server	6.5.0 RTM build 4602587
	6.0 RTM build 2562643

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 3

2.3. Guest OS Support

NVIDIA GRID software supports several Windows releases and Linux distributions as a guest OS.

> Use only a guest OS release that is listed as supported by NVIDIA GRID software with your virtualization software. To be listed as supported, a guest OS release must be supported not only by NVIDIA GRID software, but also by your virtualization software. NVIDIA **cannot** support guest OS releases that your virtualization software does not support.

2.3.1. Windows Guest OS Support

NVIDIA GRID software supports **only** the following Windows releases as a guest OS on VMware vSphere:

> If a specific release, even an update release, is not listed, it's **not** supported.

- ▸ Windows Server 2016 1607, 1709
- ▸ Windows Server 2012 R2
- ▸ Windows Server 2008 R2
- ▸ Windows 10 RTM (1507), November Update (1511), Anniversary Update (1607), Creators Update (1703) (32/64-bit)
- ▸ Windows 8.1 (32/64-bit)
- ▸ Windows 8 (32/64-bit)
- ▸ Windows 7 (32/64-bit)

2.3.2. Linux Guest OS Support

NVIDIA GRID software supports only the following Linux distributions as a guest OS **only** on supported Tesla GPUs on VMware vSphere:

> If a specific release, even an update release, is not listed, it's **not** supported.

- ▸ Red Hat Enterprise Linux 7.0-7.4 and later compatible 7.x versions
- ▸ CentOS 7.0-7.4 and later compatible 7.x versions
- ▸ Red Hat Enterprise Linux 6.6 and later compatible 6.x versions
- ▸ CentOS 6.6 and later compatible 6.x versions
- ▸ Ubuntu 16.04 LTS
- ▸ Ubuntu 14.04 LTS

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 4

▶ Ubuntu 12.04 LTS

> GRID K1 and GRID K2 do not support vGPU on a Linux guest OS.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 5

Chapter 3.
KNOWN PRODUCT LIMITATIONS

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

3.1. vGPU profiles with 512 Mbytes or less of frame buffer support only 1 virtual display head on Windows 10

Description

To reduce the possibility of memory exhaustion, vGPU profiles with 512 Mbytes or less of frame buffer support only 1 virtual display head on a Windows 10 guest OS.

The following vGPU profiles have 512 Mbytes or less of frame buffer:

▸ Tesla M6-0B, M6-0Q
▸ Tesla M10-0B, M10-0Q
▸ Tesla M60-0B, M60-0Q
▸ GRID K100, K120Q
▸ GRID K200, K220Q

Workaround

Use a profile that supports more than 1 virtual display head and has at least 1 Gbyte of frame buffer.

www.nvidia.com
GRID Software for VMware vSphere Version   RN-07347-001 _v4.9 | 6
367.132/370.39

3.2. 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 both Citrix and 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 M10-0B, M10-0Q
▸ 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.

3.3. VM failures or crashes on servers with 1 TB or more of system memory

Description

Support for vGPU and vSGA is limited to servers with less than 1 TB of system memory. On servers with 1 TB or more of system memory, VM failures or crashes may occur. For example, when Citrix XenDesktop is used with a Windows 7 guest OS, a blue screen crash may occur. However, support for vDGA is not affected by this limitation.

Resolution

Limit the amount of system memory on the server to less than 1 TB.

Set `memmapMaxRAMMB` to 1048064, which is equal to 1048576 minus 512.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 7

If the problem persists, contact your server vendor for the recommended system memory configuration with NVIDIA GPUs.

3.4. 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 VMware 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 the following device status:

```
Windows has stopped this device because it has reported problems. (Code 43)
```

Depending on the versions of drivers in use, the VMware vSphere 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.

3.5. Virtual GPU fails to start if ECC is enabled

Description

Tesla M60 and Tesla M6 GPUs support error correcting code (ECC) memory for improved data integrity. Tesla M60 and M6 GPUs in graphics mode are supplied with ECC memory disabled by default, but it may subsequently be enabled using `nvidia-smi`.

However, NVIDIA GRID vGPU does not support ECC memory. If ECC memory is enabled, NVIDIA GRID vGPU fails to start. The following error is logged in the VMware vSphere VM's log file:

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

Resolution

Ensure that ECC is disabled on all GPUs.

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

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

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

3.6. 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.

> This setting can only be changed when the VM is powered off.

1. Select **Edit Settings**.
2. In **Edit Settings** window, select the **VM Options** tab.
3. From the **Advanced** drop-down list, select **Edit Configuration**.
4. In the **Configuration Parameters** dialog box, click **Add Row**.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 9

5. In the **Name** field, type the parameter name
    `pciPassthru0.cfg.frame_rate_limiter`, in the **Value** field type 0, and click
    OK.
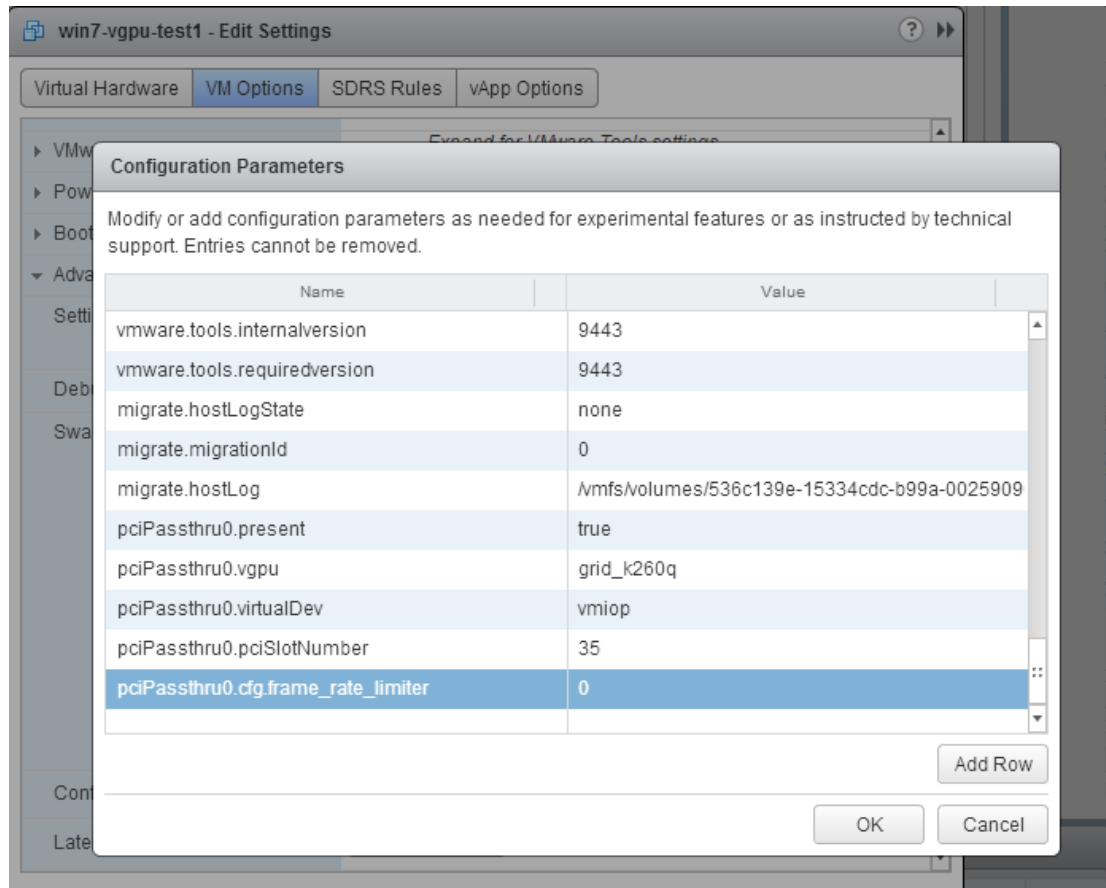
With this setting in place, the VM's vGPU will run without any frame
rate limit. The FRL can be reverted back to its default setting by setting
`pciPassthru0.cfg.frame_rate_limiter` to 1 or by removing the parameter from
the advanced settings.

3.7. GRID K1 and GRID K2 cards do not support monitoring of vGPU engine usage

Description

GRID K1 and GRID K2 cards do not support monitoring of vGPU engine usage. All
tools and APIs for any vGPU running on GRID K1 or GRID K2 cards report 0 for the
following usage statistics:

▸ 3D/Compute
▸ Memory controller bandwidth

www.nvidia.com
GRID Software for VMware vSphere Version   RN-07347-001 _v4.9 | 10
367.132/370.39

▸ Video encoder
▸ Video decoder

3.8. 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 the following device status:

```
Windows has stopped this device because it has reported problems. (Code 43)
```

The VMware vSphere 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

> 💬 This setting can only be changed when the VM is powered off.

1. Select **Edit Settings**.
2. In **Edit Settings** window, select the **VM Options** tab.
3. From the **Advanced** drop-down list, select **Edit Configuration**.
4. In the **Configuration Parameters** dialog box, click **Add Row**.
5. In the **Name** field, type the parameter name `pciPassthru0.cfg.enable_large_sys_mem`, in the **Value** field type 1, and click **OK**.

www.nvidia.com
GRID Software for VMware vSphere Version   RN-07347-001 _v4.9 | 11
367.132/370.39

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 64 GB of RAM. We recommend adding 2 M of reservation for each additional 1 GB of system memory. For example, to support 96 GB 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.

www.nvidia.com
GRID Software for VMware vSphere Version   RN-07347-001 _v4.9 | 12
367.132/370.39

Chapter 4.
RESOLVED ISSUES

No resolved issues are reported in this release for VMware vSphere.

www.nvidia.com
GRID Software for VMware vSphere Version   RN-07347-001 _v4.9 | 13
367.132/370.39

Chapter 5.
SECURITY UPDATES

5.1. Restricting Access to GPU Performance Counters

The NVIDIA graphics driver contains a vulnerability (CVE-2018-6260) that may allow access to application data processed on the GPU through a side channel exposed by the GPU performance counters. To address this vulnerability, update the driver and restrict access to GPU performance counters to allow access only by administrator users and users who need to use CUDA profiling tools.

The GPU performance counters that are affected by this vulnerability are the hardware performance monitors used by the CUDA profiling tools such as CUPTI, Nsight Graphics, and Nsight Compute. These performance counters are exposed on the hypervisor host and in guest VMs only as follows:

▸ On the hypervisor host, they are always exposed. However, the Virtual GPU Manager does not access these performance counters and, therefore, is not affected.
▸ In Windows and Linux guest VMs, they are exposed **only** in VMs configured for GPU pass through. They are not exposed in VMs configured for NVIDIA vGPU.

5.1.1. Windows: Restricting Access to GPU Performance Counters for One User by Using **NVIDIA Control Panel**

Perform this task from the guest VM to which the GPU is passed through.

Ensure that you are running **NVIDIA Control Panel** version 8.1.950.

1. Open **NVIDIA Control Panel**:

    ▸ Right-click on the Windows desktop and select **NVIDIA Control Panel** from the menu.
    ▸ Open **Windows Control Panel** and double-click the **NVIDIA Control Panel** icon.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 14

2. In **NVIDIA Control Panel**, select the **Manage GPU Performance Counters** task in the **Developer** section of the navigation pane.

3. Complete the task by following the instructions in the **Manage GPU Performance Counters** > **Developer** topic in the **NVIDIA Control Panel** help.

5.1.2. Windows: Restricting Access to GPU Performance Counters Across an Enterprise by Using a Registry Key

You can use a registry key to restrict access to GPU Performance Counters for all users who log in to a Windows guest VM. By incorporating the registry key information into a script, you can automate the setting of this registry for all Windows guest VMs across your enterprise.

Perform this task from the guest VM to which the GPU is passed through.

> 💬 **Caution** Only **enterprise administrators** should perform this task. Changes to the Windows registry must be made with care and system instability can result if registry keys are incorrectly set.

1. Set the `RmProfilingAdminOnly` Windows registry key to 1.

   ```
   [HKEY_LOCAL_MACHINE\SOFTWARE\NVIDIA Corporation\Global\NVTweak]
   Value: "RmProfilingAdminOnly"
   Type: DWORD
   Data: 00000001
   ```

   The data value 1 restricts access, and the data value 0 allows access, to application data processed on the GPU through a side channel exposed by the GPU performance counters.

2. Restart the VM.

5.1.3. Linux Guest VMs: Restricting Access to GPU Performance Counters

On systems where unprivileged users don't need to use GPU performance counters, restrict access to these counters to system administrators, namely users with the `CAP_SYS_ADMIN` capability set. By default, the GPU performance counters are not restricted to users with the `CAP_SYS_ADMIN` capability.

Perform this task from the guest VM to which the GPU is passed through.

This task requires `sudo` privileges.

1. Log in to the guest VM.

2. Set the kernel module parameter `NVreg_RestrictProfilingToAdminUsers` to 1 by adding this parameter to the `/etc/modprobe.d/nvidia.conf` file.

   ▶ If you are setting only this parameter, add an entry for it to the `/etc/modprobe.d/nvidia.conf` file as follows:

   ```
   options nvidia
    NVreg_RegistryDwords="NVreg_RestrictProfilingToAdminUsers=1"
   ```

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39
RN-07347-001 _v4.9 | 15

- ▸ If you are setting multiple parameters, set them in a single entry as in the following example:

```
options nvidia NVreg_RegistryDwords="RmPVMRL=0x0"
 "NVreg_RestrictProfilingToAdminUsers=1"
```

If the `/etc/modprobe.d/nvidia.conf` file does not already exist, create it.

3. Restart the VM.

5.1.4. Hypervisor Host: Restricting Access to GPU Performance Counters

On systems where unprivileged users don't need to use GPU performance counters, restrict access to these counters to system administrators. By default, the GPU performance counters are not restricted to system administrators.

Perform this task from your hypervisor host machine.

1. Open a command shell as the root user on your hypervisor host machine.

2. Set the kernel module parameter `NVreg_RestrictProfilingToAdminUsers` to 1 by using the `esxcli set` command.

   - ▸ If you are setting only this parameter, set it as follows:

   ```
   # esxcli system module parameters set -m nvidia -p
    "NVreg_RestrictProfilingToAdminUsers=1"
   ```

   - ▸ If you are setting multiple parameters, set them in a single command as in the following example:

   ```
   # esxcli system module parameters set -m nvidia -p
    "NVreg_RegistryDwords=RmPVMRL=0x0 NVreg_RestrictProfilingToAdminUsers=1"
   ```

3. Reboot your hypervisor host machine.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 16

Chapter 6.
KNOWN ISSUES

6.1. NVOS errors might be logged when the NvFBC state is changed

Description

When the NvFBC state is changed in the VM, the following error messages might be written to the log files:

```
NVOS status 0x56
```

```
Failed to reset guest's license info in host
```

If you see these errors in the log files, ignore them.

Workaround

None required.

Status

Open

Ref. #

200494421

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 17

6.2. 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 in the following situations:

▸ Full screen 1080p video content is playing in a browser. In this situation, the session hangs and session reconnection fails.
▸ Multiple display heads are used with Citrix XenDesktop or VMware Horizon on a Windows 10 guest VM.
▸ Higher resolution monitors are used.
▸ Applications that are frame-buffer intensive are used.
▸ NVENC is in use.

To reduce the possibility of memory exhaustion, NVENC is disabled on profiles that have 512 Mbytes or less of frame buffer.

When memory exhaustion 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 M10-0B, M10-0Q
▸ Tesla M60-0B, M60-0Q
▸ GRID K100, K120Q
▸ GRID K200, K220Q

The root cause is a known issue associated with changes to the way that recent Microsoft operating systems handle and allow access to overprovisioning messages and errors. If your systems are provisioned with enough frame buffer to support your use cases, you should not encounter these issues.

Workaround

▸ Use an appropriately sized vGPU to ensure that the frame buffer supplied to a VM through the vGPU is adequate for your workloads.
▸ Monitor your frame buffer usage.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 18

- ▸ If you are using Windows 10, consider these workarounds and solutions:

  - ▸ Use a profile that has 1 Gbyte of frame buffer.
  - ▸ Optimize your Windows 10 resource usage.

    To obtain information about best practices for improved user experience using Windows 10 in virtual environments, complete the NVIDIA GRID vGPU Profile Sizing Guide for Windows 10 download request form.

    Additionally, you can use the VMware OS Optimization Tool to make and apply optimization recommendations for Windows 10 and other operating systems.

Status

Open

Ref. #

- ▸ 200130864
- ▸ 1803861

6.3. vGPU VM fails to boot in ESXi 6.5 if the graphics type is Shared

Description

On VMware vSphere Hypervisor (ESXi) 6.5, after vGPU is configured, VMs to which a vGPU is assigned may fail to start and the following error message may be displayed:

```
The amount of graphics resource available in the parent resource pool is
 insufficient for the operation.
```

The vGPU Manager VIB provides vSGA and vGPU functionality in a single VIB. After this VIB is installed, the default graphics type is Shared, which provides vSGA functionality. To enable vGPU support for VMs in VMware vSphere 6.5, you must change the default graphics type to Shared Direct. If you do not change the default graphics type you will encounter this issue.

Version

VMware vSphere Hypervisor (ESXi) 6.5

Workaround

Change the default graphics type to Shared Direct as explained in *GRID Software User Guide*.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 19

Status

Open

Ref. #

200256224

6.4. ESXi 6.5 web client shows high memory usage even when VMs are idle

Description

On VMware vSphere Hypervisor (ESXi) 6.5, the web client shows a memory usage alarm with critical severity for VMs to which a vGPU is attached even when the VMs are idle. When memory usage is monitored from inside the VM, no memory usage alarm is shown. The web client does not show a memory usage alarm for the same VMs without an attached vGPU.

Version

VMware vSphere Hypervisor (ESXi) 6.5

Workaround

Avoid using the VMware vSphere Hypervisor (ESXi) 6.5 web client to monitor memory usage for VMs to which a vGPU is attached.

Status

Not an NVIDIA bug

Ref. #

200191065

6.5. GRID Virtual GPU Manager must not be on a host in a VMware DRS cluster

Description

The ESXi host on which the NVIDIA Virtual GPU Manager for vSphere is installed must not be a member of a VMware Distributed Resource Scheduler (DRS) cluster. The installer for the NVIDIA driver for GRID Virtual GPU cannot locate the GRID GPU card

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 20

on a host in a VMware DRS Cluster. Any attempt to install the driver on a VM on a host in a DRS cluster fails with the following error:

```
NVIDIA Installer cannot continue
This graphics driver could not find compatible graphics hardware.
```

Version

Workaround

Move GRID Virtual GPU Manager to a host outside the DRS cluster.

1. Remove GRID Virtual GPU Manager from the host in the DRS cluster.
2. Create a cluster of VMware ESXi hosts outside the DRS domain.
3. Install the GRID Virtual GPU Manager on an ESXi host in the cluster that you created in the previous step.
4. Create a vSphere VM for use with GRID Virtual GPU.
5. Configure the vSphere VM with GRID Virtual GPU.
6. Boot the vSphere VM and install the NVIDIA driver for GRID Virtual GPU.

For instructions for performing these tasks, refer to *GRID Software User Guide*.

Status

Open

Ref. #

1933449

6.6. GNOME Display Manager (GDM) fails to start on Red Hat Enterprise Linux 7.2 and CentOS 7.0

Description

GDM fails to start on Red Hat Enterprise Linux 7.2 and CentOS 7.0 with the following error:

```
Oh no! Something has gone wrong!
```

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.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39
RN-07347-001 _v4.9 | 21

```
~]# 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

6.7. NVIDIA Control Panel fails to start and reports 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 reports that you are not using a display attached to an NVIDIA GPU. This happens because Windows is using VMware's SVGA device instead of NVIDIA vGPU.

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.

> 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

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39
RN-07347-001 _v4.9 | 22

Ref. #

6.8. 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 the following device status:

```
Windows has stopped this device because it has reported problems. (Code 43)
```

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. #

6.9. 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 the following device status:

```
Windows has stopped this device because it has reported problems. (Code 43)
```

Workaround

Do not assign vGPU and passthrough GPUs to a VM simultaneously.

Status

Open

Ref. #

1735002

6.10. 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.

Workaround

Start VMs individually.

Status

Not an NVIDIA bug

Ref. #

200042690

6.11. Before Horizon agent is installed 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.

Workaround

Do not use Sleep with vGPU.

Installing the VMware Horizon agent will disable the **Sleep** option.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39
RN-07347-001 _v4.9 | 24

Status

Closed

Ref. #

200043405

6.12. 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, the following errors occur:

▶ One or more of the VMs may fail to start with the following error:

```
The available Memory resources in the parent resource pool are insufficient
 for the operation
```

▶ When run in the host shell, the `nvidia-smi` utility 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.

Workaround

Reduce the total amount of system memory assigned to the VMs.

Status

Closed

Ref. #

200060499

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 25

6.13. On reset or restart VMs fail to start with the error `VMIOP: no graphics device is available for vGPU…`

Description

On a system running a maximal configuration, that is, with the maximum number of vGPU VMs the server can support, some VMs might fail to start post a reset or restart operation.

Fix

Upgrade to ESXi 6.0 Update 1.

Status

Closed

Ref. #

200097546

6.14. `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.

Workaround

None

Status

Open

Partially resolved for Horizon 7.0.1:

▸ For Blast connections, GPU utilization is no longer high.
▸ For PCoIP connections, utilization remains high.

www.nvidia.com
GRID Software for VMware vSphere Version
367.132/370.39

RN-07347-001 _v4.9 | 26

Ref. #

1735009

6.15. 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.

Workaround

Disable hardware acceleration in IE.

To enable software rendering in IE, refer to the Microsoft knowledge base article How to enable or disable software rendering in Internet Explorer.

Status

Open

Ref. #

200148377

www.nvidia.com
GRID Software for VMware vSphere Version   RN-07347-001 _v4.9 | 27
367.132/370.39