Virtual GPU Software R510 for Citrix Hypervisor

Release Notes
Table of Contents

Chapter 1. Release Notes.................................................................................................... 1
  1.1. NVIDIA vGPU Software Driver Versions...........................................................................1
  1.2. Compatibility Requirements for the NVIDIA vGPU Manager and Guest VM Driver...........2
  1.3. Updates in Release 14.0....................................................................................................3

Chapter 2. Validated Platforms............................................................................................5
  2.1. Supported NVIDIA GPUs and Validated Server Platforms..............................................5
     2.1.1. Switching the Mode of a GPU that Supports Multiple Display Modes.........................7
     2.1.2. Switching the Mode of a Tesla M60 or M6 GPU............................................................7
  2.2. Hypervisor Software Releases..........................................................................................8
  2.3. Guest OS Support...........................................................................................................10
     2.3.1. Windows Guest OS Support..........................................................................................10
     2.3.2. Linux Guest OS Support..............................................................................................11
  2.4. NVIDIA CUDA Toolkit Version Support............................................................................12
  2.5. XenMotion with vGPU Support..........................................................................................12
  2.6. Multiple vGPU Support...................................................................................................14
  2.7. Peer-to-Peer CUDA Transfers over NVLink Support.........................................................16
  2.8. Unified Memory Support.................................................................................................17
  2.9. NVIDIA Deep Learning Super Sampling (DLSS) Support..................................................18

Chapter 3. Known Product Limitations..............................................................................20
  3.1. vGPUs of different types on the same GPU are not supported.........................................20
  3.2. NVENC does not support resolutions greater than 4096×4096........................................20
  3.3. Issues occur when the channels allocated to a vGPU are exhausted................................21
  3.4. Total frame buffer for vGPUs is less than the total frame buffer on the physical GPU....22
  3.5. Issues may occur with graphics-intensive OpenCL applications on vGPU types with limited frame buffer..................................................................................................................24
  3.6. In pass through mode, all GPUs connected to each other through NVLink must be assigned to the same VM.................................................................25
  3.7. vGPU profiles with 512 Mbytes or less of frame buffer support only 1 virtual display head on Windows 10.............................................................................................................25
  3.8. NVENC requires at least 1 Gbyte of frame buffer............................................................26
  3.9. VM running an incompatible NVIDIA vGPU guest driver fails to initialize vGPU when booted..........................................................................................................................26
  3.10. Single vGPU benchmark scores are lower than pass-through GPU................................27
  3.11. nvidia-smi fails to operate when all GPUs are assigned to GPU pass-through mode....28
  3.12. Windows Aero is disabled on Citrix Virtual Apps and Desktops session using 3 or 4 monitors in 2560×1600 resolution.................................................................30
3.13. VMs configured with large memory fail to initialize vGPU when booted............................30
3.14. vGPU host driver RPM upgrade fails...................................................................................31

Chapter 4. Resolved Issues................................................................................................33

Chapter 5. Known Issues................................................................................................... 34

5.1. For some license deployments, GSP firmware remains enabled with unsupported
products.....................................................................................................................................34
5.2. After an upgrade of the Linux graphics driver from an RPM package in a licensed VM,
licensing fails..........................................................................................................................35
5.3. After an upgrade of the Linux graphics driver from a Debian package, the driver is not
loaded into the VM....................................................................................................................36
5.4. Desktop session freezes when a Linux VM is migrated to or from a host running an
NVIDIA vGPU software 14 release...........................................................................................36
5.5. The reported NVENC frame rate is double the actual frame rate........................................37
5.6. Hypervisor host reboots when multiple cloned VMs are simultaneously powered on or
migrated.....................................................................................................................................38
5.7. NVENC does not work with Teradici Cloud Access Software on Windows........................38
5.8. A licensed client might fail to acquire a license if a proxy is set............................................39
5.9. Session connection fails with four 4K displays and NVENC enabled on a 2Q, 3Q, or 4Q
vGPU...........................................................................................................................................40
5.10.Disconnected sessions cannot be reconnected or might be reconnected very slowly
with NWMI installed...................................................................................................................41
5.11. Linux VM hangs after vGPU migration to a host running a newer vGPU manager
version.........................................................................................................................................41
5.12. Idle Teradici Cloud Access Software session disconnects from Linux VM............................42
5.13. No virtual GPU types are listed in Citrix XenCenter..................................................................43
5.14. NVIDIA vGPU software graphics driver fails to load after upgrade on Citrix Hypervisor......44
5.15. Windows guest VMs with vGPUs or GPUs with large BAR memory settings fail to boot
to the desktop in UEFI mode...................................................................................................45
5.16. Migrated VM with more than two vGPUs crashes on destination host.................................46
5.17. Driver upgrade in a Linux guest VM with multiple vGPUs might fail.................................46
5.18. NVIDIA Control Panel fails to start if launched too soon from a VM without licensing
information.................................................................................................................................47
5.19. VNC client session goes blank and console VNC is corrupted when the guest driver
is uninstalled...............................................................................................................................47
5.21. Suspend and resume between hosts running different versions of the vGPU manager
fails...............................................................................................................................................48
5.22. On Linux, the frame rate might drop to 1 after several minutes............................................49
5.23. Citrix XenCenter displays misleading information about vGPU types................................49
5.24. Citrix Virtual Apps and Desktops connection freezes initially............................................50

Virtual GPU Software R510 for Citrix Hypervisor  RN-06927-001_v14.0 Revision 03  iii
Virtual GPU Software R510 for Citrix Hypervisor

5.25. DWM crashes randomly occur in Windows VMs .................................................................51
5.26. NVIDIA Control Panel fails to launch in a platform layer or published image ..................52
5.27. Remote desktop session freezes with assertion failure and XID error 43 after migration .................................................................52
5.28. Citrix Virtual Apps and Desktops session freezes when the desktop is unlocked .............53
5.29. NVIDIA vGPU software graphics driver fails after Linux kernel upgrade with DKMS enabled .....................................................................54
5.30. On Citrix Hypervisor, all vGPUs in a VM must be of the same type .................................55
5.31. Console VGA cannot be disabled ..................................................................................55
5.32. Console VNC is unusable with Xorg on multiple vGPUs in a VM ..................................56
5.33. RAPIDS cuDF merge fails on NVIDIA vGPU .................................................................57
5.34. Migration of VMs configured with vGPU stops before the migration is complete .............57
5.35. ECC memory settings for a vGPU cannot be changed by using NVIDIA X Server Settings ........................................................................58
5.36. Changes to ECC memory settings for a Linux vGPU VM by nvidia-smi might be ignored .....................................................................58
5.37. Incorrect GPU type shown for Quadro RTX 8000 GPUs in Citrix XenCenter ..................59
5.38. NVIDIA Notification Icon prevents log off of Citrix Published Application user sessions ........................................................................60
5.39. Host core CPU utilization is higher than expected for moderate workloads ..................60
5.40. Frame capture while the interactive logon message is displayed returns blank screen ....61
5.41. RDS sessions do not use the GPU with some Microsoft Windows Server releases .........62
5.42. When the scheduling policy is fixed share, GPU utilization is reported as higher than expected ........................................................................62
5.43. nvidia-smi reports that vGPU migration is supported on all hypervisors .........................63
5.44. NVIDIA Control Panel Crashes in a VM connected to two 4K displays ..........................64
5.45. vGPU guest VM driver not properly loaded on servers with more than 512 GB or 1 TB or more of system memory ..................................................64
5.46. Luxmark causes a segmentation fault on an unlicensed Linux client ............................65
5.47. A segmentation fault in DBus code causes nvidia-gridd to exit on Red Hat Enterprise Linux and CentOS ................................................................ 66
5.48. No Manage License option available in NVIDIA X Server Settings by default ...............66
5.49. Licenses remain checked out when VMs are forcibly powered off ..................................67
5.50. Memory exhaustion can occur with vGPU profiles that have 512 Mbytes or less of frame buffer ........................................................................68
5.51. VM bug checks after the guest VM driver for Windows 10 RS2 is installed .....................69
5.52. On Citrix Hypervisor 7.0, VMs unexpectedly reboot and Citrix Hypervisor crashes or freezes ........................................................................69
5.53. With no NVIDIA driver installed, Citrix Hypervisor misidentifies Tesla M10 cards ........71
5.54. GNOME Display Manager (GDM) fails to start on Red Hat Enterprise Linux 7.2 and CentOS 7.0 ................................................................................71
5.55. Video goes blank when run in loop in Windows Media Player ......................................72
5.56. Local VGA console is momentarily unblanked when Citrix Virtual Apps and Desktops changes resolution of the VM desktop.................................................................................... 73
5.57. VM bugchecks on shutdown/restart when Citrix Virtual Apps and Desktops is installed and NVIDIA driver is uninstalled or upgraded........................................................................ 73
5.58. Application frame rate may drop when running Citrix Virtual Apps and Desktops at 2560×1600 resolution................................................................................................................ 74
5.59. Windows VM BSOD............................................................................................................... 75
5.60. Windows VM BSOD when upgrading NVIDIA drivers over a Citrix Virtual Apps and Desktops session.................................................................................................................................. 76
5.61. XenCenter does not allow vGPUs to be selected as a GPU type for Linux VMs.............. 76
5.62. If X server is killed on a RHEL7 VM running vGPU, XenCenter console may not automatically switch to text console........................................................................................................ 77
5.63. Citrix Virtual Apps and Desktops shows only a black screen when connected to a vGPU VM............................................................................................................................................... 78
Chapter 1. Release Notes

These Release Notes summarize current status, information on validated platforms, and known issues with NVIDIA vGPU software and associated hardware on Citrix Hypervisor.

Note: The most current version of the documentation for this release of NVIDIA vGPU software can be found online at NVIDIA Virtual GPU Software Documentation.

1.1. NVIDIA vGPU Software Driver Versions

Each release in this release family of NVIDIA vGPU software includes a specific version of the NVIDIA Virtual GPU Manager, NVIDIA Windows driver, and NVIDIA Linux driver.

<table>
<thead>
<tr>
<th>NVIDIA vGPU Software Version</th>
<th>NVIDIA Virtual GPU Manager Version</th>
<th>NVIDIA Windows Driver Version</th>
<th>NVIDIA Linux Driver Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.0</td>
<td>510.47.03</td>
<td>511.65</td>
<td>510.47.03</td>
</tr>
</tbody>
</table>

For details of which Citrix Hypervisor releases are supported, see Hypervisor Software Releases.
1.2. Compatibility Requirements for the NVIDIA vGPU Manager and Guest VM Driver

The releases of the NVIDIA vGPU Manager and guest VM drivers that you install must be compatible. If you install an incompatible guest VM driver release for the release of the vGPU Manager that you are using, the NVIDIA vGPU fails to load.

See VM running an incompatible NVIDIA vGPU guest driver fails to initialize vGPU when booted.

**Note:** This requirement does not apply to the NVIDIA vGPU software license server. All releases in this release family of NVIDIA vGPU software are compatible with all releases of the license server.

### Compatible NVIDIA vGPU Manager and Guest VM Driver Releases

The following combinations of NVIDIA vGPU Manager and guest VM driver releases are compatible with each other.

- NVIDIA vGPU Manager with guest VM drivers from the same release
- NVIDIA vGPU Manager from a later major release branch with guest VM drivers from the previous branch

**Note:**
When NVIDIA vGPU Manager is used with guest VM drivers from the previous branch, the combination supports only the features, hardware, and software (including guest OSes) that are supported on both releases.

For example, if vGPU Manager from release 14.0 is used with guest drivers from release 13.1, the combination does not support Red Hat Enterprise Linux 8.1 because NVIDIA vGPU software release 14.0 does not support Red Hat Enterprise Linux 8.1.

The following table lists the specific software releases that are compatible with the components in the NVIDIA vGPU software 14 major release branch.

<table>
<thead>
<tr>
<th>NVIDIA vGPU Software Component</th>
<th>Release</th>
<th>Compatible Software Releases</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVIDIA vGPU Manager</td>
<td>14.0</td>
<td>- Guest VM driver release 14.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- All guest VM driver 13.x releases</td>
</tr>
<tr>
<td>Guest VM drivers</td>
<td>14.0</td>
<td>NVIDIA vGPU Manager release 14.0</td>
</tr>
</tbody>
</table>
Incompatible NVIDIA vGPU Manager and Guest VM Driver Releases

The following combinations of NVIDIA vGPU Manager and guest VM driver releases are incompatible with each other.

- NVIDIA vGPU Manager from a later major release branch with guest VM drivers from a production branch two or more major releases before the release of the vGPU Manager
- NVIDIA vGPU Manager from an earlier major release branch with guest VM drivers from a later branch

The following table lists the specific software releases that are incompatible with the components in the NVIDIA vGPU software 14 major release branch.

<table>
<thead>
<tr>
<th>NVIDIA vGPU Software Component</th>
<th>Release</th>
<th>Incompatible Software Releases</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVIDIA vGPU Manager</td>
<td>14.0</td>
<td>All guest VM driver releases 11.x and earlier</td>
</tr>
<tr>
<td>Guest VM drivers</td>
<td>14.0</td>
<td>All NVIDIA vGPU Manager releases 13.x and earlier</td>
</tr>
</tbody>
</table>

1.3. Updates in Release 14.0

New Features in Release 14.0

- Support for Tesla Compute Cluster (TCC) mode for Q-series vGPUs on Windows guest VMs
- Support for GPU System Processor (GSP) in GPU pass through and bare-metal configurations on Linux with vCS

Note: If you are using a product other than vCS, you must disable GSP as explained in Virtual GPU Software User Guide.

- Enhanced NVIDIA CUDA Toolkit support:
  - NVIDIA CUDA Toolkit profilers can be enabled when unified memory is enabled.
  - Nsight Systems GPU context switch trace is supported.
- Enhancements to the NVIDIA Management Library (NVML) to determine whether a vGPU type supports GPUDirect technology and peer-to-peer CUDA transfers over NVLink
- Addition of RPM and Debian packages for the NVIDIA vGPU software graphics drivers for Linux
- Security updates - see Security Bulletin: NVIDIA GPU Display Driver - February 2022, which is posted shortly after the release date of this software and is listed on the NVIDIA Product Security page
- Miscellaneous bug fixes
Hardware and Software Support Introduced in Release 14.0

- Support for the following GPUs:
  - NVIDIA A2
  - NVIDIA A30X
  - NVIDIA A100X
- Support for Red Hat Enterprise Linux 8.5 as a guest OS
- Support for Citrix Virtual Apps and Desktops version 7 2112

Feature Support Withdrawn in Release 14.0

- Red Hat Enterprise Linux 8.1 is no longer supported as a guest OS.
- Red Hat Enterprise Linux 7.8 and 7.7 are no longer supported as a guest OS.
- Windows Server 2012 R2 is no longer supported as a guest OS.

Features Deprecated in Release 14.0

The following table lists features that are deprecated in this release of NVIDIA vGPU software. Although the features remain available in this release, they might be withdrawn in a future release. In preparation for the possible removal of these features, use the preferred alternative listed in the table.

<table>
<thead>
<tr>
<th>Deprecated Feature</th>
<th>Preferred Alternative</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legacy NVIDIA vGPU software license server</td>
<td>NVIDIA License System</td>
<td>NVIDIA Virtual GPU Software License Server End of Life Notice</td>
</tr>
</tbody>
</table>

Notice
Chapter 2. Validated Platforms

This release family of NVIDIA vGPU software provides support for several NVIDIA GPUs on validated server hardware platforms, Citrix Hypervisor hypervisor software versions, and guest operating systems. It also supports the version of NVIDIA CUDA Toolkit that is compatible with R510 drivers.

2.1. Supported NVIDIA GPUs and Validated Server Platforms

This release of NVIDIA vGPU software provides support for the following NVIDIA GPUs on Citrix Hypervisor, running on validated server hardware platforms:

- GPUs based on the NVIDIA Maxwell™ graphic architecture:
  - Tesla M6
  - Tesla M10
  - Tesla M60
- GPUs based on the NVIDIA Pascal™ architecture:
  - Tesla P4
  - Tesla P6
  - Tesla P40
  - Tesla P100 PCIe 16 GB (XenMotion with vGPU is not supported.)
  - Tesla P100 SXM2 16 GB (XenMotion with vGPU is not supported.)
  - Tesla P100 PCIe 12GB (XenMotion with vGPU is not supported.)
- GPUs based on the NVIDIA Volta architecture:
  - Tesla V100 SXM2
  - Tesla V100 SXM2 32GB
  - Tesla V100 PCIe
  - Tesla V100 PCIe 32GB
  - Tesla V100S PCIe 32GB
- Tesla V100 FHHL
- GPUs based on the NVIDIA Turing™ architecture:
  - Tesla T4
  - Quadro RTX 6000 in displayless mode
  - Quadro RTX 6000 passive in displayless mode
  - Quadro RTX 8000 in displayless mode
  - Quadro RTX 8000 passive in displayless mode

In displayless mode, local physical display connectors are disabled.

- GPUs based on the NVIDIA Ampere architecture:
  - NVIDIA A100 PCIe 80GB (supports only compute workloads on Linux with GPU pass through; graphics acceleration is not supported)
  - NVIDIA A100X (supports only compute workloads on Linux with GPU pass through; graphics acceleration is not supported)
  - NVIDIA A100 HGX 80GB (supports only compute workloads on Linux with GPU pass through; graphics acceleration is not supported)
  - NVIDIA A100 PCIe 40GB (supports only compute workloads on Linux with GPU pass through; graphics acceleration is not supported)
  - NVIDIA A100 HGX 40GB (supports only compute workloads on Linux with GPU pass through; graphics acceleration is not supported)
  - NVIDIA A40 in displayless mode
  - NVIDIA A30 (supports only compute workloads on Linux with GPU pass through; graphics acceleration is not supported)
  - NVIDIA A30X (supports only compute workloads on Linux with GPU pass through; graphics acceleration is not supported)
  - NVIDIA A16
  - NVIDIA A10
  - NVIDIA A2
  - NVIDIA RTX A6000 in displayless mode
  - NVIDIA RTX A5000 in displayless mode

In displayless mode, local physical display connectors are disabled.

For a list of validated server platforms, refer to NVIDIA GRID Certified Servers.
2.1.1. Switching the Mode of a GPU that Supports Multiple Display Modes

Some GPUs support displayless and display-enabled modes but must be used in NVIDIA vGPU software deployments in displayless mode.

The GPUs listed in the following table support multiple display modes. As shown in the table, some GPUs are supplied from the factory in displayless mode, but other GPUs are supplied in a display-enabled mode.

<table>
<thead>
<tr>
<th>GPU</th>
<th>Mode as Supplied from the Factory</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVIDIA A40</td>
<td>Displayless</td>
</tr>
<tr>
<td>NVIDIA RTX A5000</td>
<td>Display enabled</td>
</tr>
<tr>
<td>NVIDIA RTX A6000</td>
<td>Display enabled</td>
</tr>
</tbody>
</table>

A GPU that is supplied from the factory in displayless mode, such as the NVIDIA A40 GPU, might be in a display-enabled mode if its mode has previously been changed.

To change the mode of a GPU that supports multiple display modes, use the `displaymodeselector` tool, which you can request from the NVIDIA Display Mode Selector Tool page on the NVIDIA Developer website.

**Note:**
Only the following GPUs support the `displaymodeselector` tool:
- NVIDIA A40
- NVIDIA RTX A5000
- NVIDIA RTX A6000

Other GPUs that support NVIDIA vGPU software do not support the `displaymodeselector` tool and, unless otherwise stated, do not require display mode switching.

2.1.2. Switching the Mode of a Tesla M60 or M6 GPU

Tesla M60 and M6 GPUs support compute mode and graphics mode. NVIDIA 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 NVIDIA vGPU software releases. If you are unsure which mode your GPU is in, use the `gpumodeswitch` tool to find out the mode.

**Note:**
Validated Platforms

Only Tesla M60 and M6 GPUs support the `gpumodeswitch` tool. Other GPUs that support NVIDIA vGPU do not support the `gpumodeswitch` tool and, except as stated in Switching the Mode of a GPU that Supports Multiple Display Modes, do not require mode switching.

Even in compute mode, Tesla M60 and M6 GPUs do not support NVIDIA Virtual Compute Server vGPU types. Furthermore, vCS is not supported on any GPU on Citrix Hypervisor.

For more information, refer to `gpumodeswitch User Guide`.

2.2. Hypervisor Software Releases

Changes to Citrix Product Names and Release Numbers

Citrix has changed the product names of its hypervisor and virtual desktop software.

<table>
<thead>
<tr>
<th>Old Name</th>
<th>New Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>XenServer</td>
<td>Citrix Hypervisor</td>
</tr>
<tr>
<td>XenApp and XenDesktop</td>
<td>Citrix Virtual Apps and Desktops</td>
</tr>
<tr>
<td>Citrix Receiver</td>
<td>Citrix Workspace App</td>
</tr>
</tbody>
</table>

The release number of Citrix Virtual Apps and Desktops releases after 7.18 is in the form `7 yymm`, where `yymm` denotes the year and month that the product was released. For example, for a product released in August 2018, `yymm` is 1808.

For more information, see these topics in the Citrix documentation:

- New product names
- New product and component version numbers

Supported Citrix Hypervisor Releases

This release family of NVIDIA vGPU software is supported on the Citrix Hypervisor releases listed in the table.

Note:

Support for NVIDIA vGPU software requires the Premium Edition (previously Enterprise Edition) of Citrix Hypervisor. For details, see Licensing in the Citrix documentation.

Cumulative update releases for a base release of Citrix Hypervisor are compatible with the base release and can also be used with this version of NVIDIA vGPU software unless expressly stated otherwise.

<table>
<thead>
<tr>
<th>Software</th>
<th>Releases Supported</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrix Hypervisor 8.2</td>
<td>RTM build and compatible cumulative update releases</td>
<td>This release supports XenMotion with vGPU on suitable GPUs as</td>
</tr>
</tbody>
</table>
The following GPUs are supported in GPU pass through mode only:

- NVIDIA A100 PCIe 80GB
- NVIDIA A100X
- NVIDIA A100 HGX 80GB
- NVIDIA A100 PCIe 40GB
- NVIDIA A100 HGX 40GB
- NVIDIA A30
- NVIDIA A30X

vCS is not supported.

<table>
<thead>
<tr>
<th>Software</th>
<th>Releases Supported</th>
<th>Notes</th>
</tr>
</thead>
</table>
| Citrix XenServer 7.1 | RTM build and compatible cumulative update releases | Not supported on the following GPUs:  
  - Tesla T4  
  - Quadro RTX 6000 passive  
  - Quadro RTX 8000 passive  
  - NVIDIA RTX A6000  
  - NVIDIA RTX A5000  
  - NVIDIA A40  
  - NVIDIA A30  
  - NVIDIA A30X  
  - NVIDIA A16  
  - NVIDIA A10  
  - NVIDIA A2  
  - NVIDIA A100 PCIe 80GB  
  - NVIDIA A100X  
  - NVIDIA A100 HGX 80GB  
  - NVIDIA A100 PCIe 40GB  
  - NVIDIA A100 HGX 40GB  

Multiple vGPUs on a single VM are not supported.

vCS is not supported.
Validated Platforms

**Supported Virtual Desktop Software Releases**

This release supports only the virtual desktop software releases listed in the table. Except where otherwise stated, HDX 3D Pro mode is supported but not required.

<table>
<thead>
<tr>
<th>Software</th>
<th>Releases Supported</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrix Virtual Apps and Desktops</td>
<td>Versions 7 2112, 7 2109, 7 2106, 7 2103, 7 2012, 7 2009, 7 2006, 7 2003, 7 1912, 7 1909, 7 1906, and 7 1903</td>
<td>XenMotion with vGPU is not supported.</td>
</tr>
<tr>
<td>XenApp/XenDesktop</td>
<td>Version 7.15</td>
<td>Note: Version 7.15 is supported only in HDX 3D Pro mode. HDX 3D Pro mode is required to ensure that the NVIDIA GPU can be used unrestricted.</td>
</tr>
</tbody>
</table>

**2.3. Guest OS Support**

NVIDIA vGPU software supports several Windows releases and Linux distributions as a guest OS. The supported guest operating systems depend on the hypervisor software version.

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

NVIDIA vGPU software supports only 64-bit guest operating systems. No 32-bit guest operating systems are supported.

**2.3.1. Windows Guest OS Support**

NVIDIA vGPU software supports only the 64-bit Windows releases listed in the table as a guest OS on Citrix Hypervisor. The releases of Citrix Hypervisor for which a Windows release is supported depend on whether NVIDIA vGPU or pass-through GPU is used.

Note:
If a specific release, even an update release, is not listed, it’s not supported.
XenMotion with vGPU is supported on supported Windows guest OS releases.
### Validated Platforms

#### Virtual GPU Software R510 for Citrix Hypervisor

<table>
<thead>
<tr>
<th>Guest OS</th>
<th>NVIDIA vGPU - Citrix Hypervisor Releases</th>
<th>Pass-Through GPU - Citrix Hypervisor Releases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2022</td>
<td>8.2</td>
<td>8.2, 7.1 cumulative update 2</td>
</tr>
<tr>
<td>Windows Server 2019</td>
<td>8.2, 7.1 cumulative update 2</td>
<td>8.2, 7.1 cumulative update 2</td>
</tr>
<tr>
<td>Windows Server 2016 1709, 1607</td>
<td>8.2, 7.1</td>
<td>8.2, 7.1</td>
</tr>
<tr>
<td>Windows 10 November 2021 Update (21H2) and all Windows 10 releases supported by Microsoft up to and including this release See Note [1]</td>
<td>8.2, 7.1</td>
<td>8.2, 7.1</td>
</tr>
</tbody>
</table>

**Note:**
1. The hardware-accelerated GPU scheduling feature introduced in Windows 10 May 2020 Update (2004) is **not** supported on GPUs based on the Maxwell architecture and is supported only in pass-through mode on GPUs based on later architectures.

#### Linux Guest OS Support

NVIDIA vGPU software supports **only** the Linux distributions listed in the table as a guest OS on Citrix Hypervisor. The releases of Citrix Hypervisor for which a Linux release is supported depend on whether NVIDIA vGPU or pass-through GPU is used.

**Note:**
If a specific release, even an update release, is not listed, it’s **not** supported.

XenMotion with vGPU is **not** supported on any Linux guest OS release.

<table>
<thead>
<tr>
<th>Guest OS</th>
<th>NVIDIA vGPU - Citrix Hypervisor Releases</th>
<th>Pass-Through GPU - Citrix Hypervisor Releases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Hat Enterprise Linux 8.5</td>
<td>8.2</td>
<td>8.2</td>
</tr>
<tr>
<td>Red Hat Enterprise Linux 8.4</td>
<td>8.2</td>
<td>8.2</td>
</tr>
<tr>
<td>Red Hat Enterprise Linux 8.2</td>
<td>8.2</td>
<td>8.2</td>
</tr>
<tr>
<td>CentOS Linux 8 (1911)</td>
<td>8.2</td>
<td>8.2</td>
</tr>
<tr>
<td>CentOS 8.0</td>
<td>8.2</td>
<td>8.2</td>
</tr>
<tr>
<td>Red Hat Enterprise Linux 7.9</td>
<td>8.2, 7.1</td>
<td>8.2, 7.1</td>
</tr>
<tr>
<td>CentOS 7.6-7.8</td>
<td>8.2, 7.1</td>
<td>8.2, 7.1</td>
</tr>
</tbody>
</table>
Validated Platforms

<table>
<thead>
<tr>
<th>Guest OS</th>
<th>NVIDIA vGPU - Citrix Hypervisor Releases</th>
<th>Pass-Through GPU - Citrix Hypervisor Releases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubuntu 20.04 LTS</td>
<td>8.2</td>
<td>8.2</td>
</tr>
<tr>
<td>Ubuntu 18.04 LTS</td>
<td>8.2, 7.1 cumulative update 2</td>
<td>8.2, 7.1 cumulative update 2</td>
</tr>
<tr>
<td>Ubuntu 16.04 LTS</td>
<td>8.2, 7.1</td>
<td>8.2, 7.1</td>
</tr>
<tr>
<td>Ubuntu 14.04 LTS</td>
<td>8.2, 7.1</td>
<td>8.2, 7.1</td>
</tr>
</tbody>
</table>

2.4. NVIDIA CUDA Toolkit Version Support

The releases in this release family of NVIDIA vGPU software support NVIDIA CUDA Toolkit 11.6.

For more information about NVIDIA CUDA Toolkit, see [CUDA Toolkit 11.6 Documentation](#).

**Note:**
If you are using NVIDIA vGPU software with CUDA on Linux, avoid conflicting installation methods by installing CUDA from a distribution-independent runfile package. Do not install CUDA from a distribution-specific RPM or Deb package.

To ensure that the NVIDIA vGPU software graphics driver is not overwritten when CUDA is installed, deselect the CUDA driver when selecting the CUDA components to install.

For more information, see [NVIDIA CUDA Installation Guide for Linux](#).

2.5. XenMotion with vGPU Support

XenMotion with vGPU is supported only on a subset of supported GPUs, Citrix Hypervisor releases, and guest operating systems.

**Note:** vGPU migration is disabled for a VM for which any of the following NVIDIA CUDA Toolkit features is enabled:
- Unified memory
- Debuggers
- Profilers

**Supported GPUs**
- Tesla M6
- Tesla M10
- Tesla M60
Validated Platforms

- Tesla P4
- Tesla P6
- Tesla P40
- Tesla V100 SXM2
- Tesla V100 SXM2 32GB
- Tesla V100 PCIe
- Tesla V100 PCIe 32GB
- Tesla V100S PCIe 32GB
- Tesla V100 FHHL
- Tesla T4
- Quadro RTX 6000
- Quadro RTX 6000 passive
- Quadro RTX 8000
- Quadro RTX 8000 passive
- NVIDIA A2
- NVIDIA A10
- NVIDIA A16
- NVIDIA A40
- NVIDIA RTX A5000
- NVIDIA RTX A6000

Supported Citrix Hypervisor Releases
Citrix Hypervisor 8.2 only.

Supported Guest OS Releases
Windows only. XenMotion with vGPU is not supported on Linux.

Known Issues with XenMotion with vGPU Support

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Affected GPUs</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration to or from a host running an NVIDIA vGPU software 14 release</td>
<td>Tesla T4, Tesla V100</td>
<td>Desktop session freezes when a Linux VM is migrated to or from a host running an NVIDIA vGPU software 14 release</td>
</tr>
<tr>
<td>Simultaneous migration of multiple VMs</td>
<td>All GPUs that support XenMotion with vGPU</td>
<td>Hypervisor host reboots when multiple cloned VMs are simultaneously powered on or migrated</td>
</tr>
</tbody>
</table>
Validated Platforms

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Affected GPUs</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration between hosts with different ECC memory configuration</td>
<td>All GPUs that support XenMotion with vGPU</td>
<td>Migration of VMs configured with vGPU stops before the migration is complete</td>
</tr>
</tbody>
</table>

2.6. Multiple vGPU Support

To support applications and workloads that are compute or graphics intensive, multiple vGPUs can be added to a single VM. The assignment of more than one vGPU to a VM is supported only on a subset of vGPUs and Citrix Hypervisor releases.

Supported vGPUs

Only Q-series vGPUs that are allocated all of the physical GPU’s frame buffer are supported.

<table>
<thead>
<tr>
<th>GPU Architecture</th>
<th>Board</th>
<th>vGPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampere</td>
<td>NVIDIA A40</td>
<td>A40-48Q See Note [1].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A40-48C See Note [1].</td>
</tr>
<tr>
<td></td>
<td>NVIDIA A16</td>
<td>A16-16Q See Note [1].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A16-16C See Note [1].</td>
</tr>
<tr>
<td></td>
<td>NVIDIA A10</td>
<td>A10-24Q See Note [1].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A10-24C See Note [1].</td>
</tr>
<tr>
<td></td>
<td>NVIDIA A2</td>
<td>A2-16Q See Note [1].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A2-16C See Note [1].</td>
</tr>
<tr>
<td></td>
<td>NVIDIA RTX A6000</td>
<td>A6000-48Q See Note [1].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A6000-48C See Note [1].</td>
</tr>
<tr>
<td></td>
<td>NVIDIA RTX A5000</td>
<td>A5000-24Q See Note [1].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A5000-24C See Note [1].</td>
</tr>
<tr>
<td>Turing</td>
<td>Tesla T4</td>
<td>T4-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T4-16C</td>
</tr>
<tr>
<td></td>
<td>Quadro RTX 6000</td>
<td>RTX6000-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RTX6000-24C</td>
</tr>
<tr>
<td></td>
<td>Quadro RTX 6000 passive</td>
<td>RTX6000P-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RTX6000P-24C</td>
</tr>
<tr>
<td></td>
<td>Quadro RTX 8000</td>
<td>RTX8000-48Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RTX8000-48C</td>
</tr>
<tr>
<td></td>
<td>Quadro RTX 8000 passive</td>
<td>RTX8000P-48Q</td>
</tr>
<tr>
<td>GPU Architecture</td>
<td>Board</td>
<td>vGPU</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Volta</td>
<td>Tesla V100 SXM2 32GB</td>
<td>V100DX-32Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V100D-32C</td>
</tr>
<tr>
<td></td>
<td>Tesla V100 PCIe 32GB</td>
<td>V100D-32Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V100D-32C</td>
</tr>
<tr>
<td></td>
<td>Tesla V100S PCIe 32GB</td>
<td>V100S-32Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V100S-32C</td>
</tr>
<tr>
<td></td>
<td>Tesla V100 SXM2</td>
<td>V100X-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V100X-16C</td>
</tr>
<tr>
<td></td>
<td>Tesla V100 PCIe</td>
<td>V100-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V100-16C</td>
</tr>
<tr>
<td></td>
<td>Tesla V100 FHHL</td>
<td>V100L-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V100L-16C</td>
</tr>
<tr>
<td>Pascal</td>
<td>Tesla P100 SXM2</td>
<td>P100X-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P100X-16C</td>
</tr>
<tr>
<td></td>
<td>Tesla P100 PCIe 16GB</td>
<td>P100-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P100-16C</td>
</tr>
<tr>
<td></td>
<td>Tesla P100 PCIe 12GB</td>
<td>P100C-12Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P100C-12C</td>
</tr>
<tr>
<td></td>
<td>Tesla P40</td>
<td>P40-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P40-24C</td>
</tr>
<tr>
<td></td>
<td>Tesla P6</td>
<td>P6-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P6-16C</td>
</tr>
<tr>
<td></td>
<td>Tesla P4</td>
<td>P4-8Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4-8C</td>
</tr>
<tr>
<td>Maxwell</td>
<td>Tesla M60</td>
<td>M60-8Q</td>
</tr>
<tr>
<td></td>
<td>Tesla M10</td>
<td>M10-8Q</td>
</tr>
<tr>
<td></td>
<td>Tesla M6</td>
<td>M6-8Q</td>
</tr>
</tbody>
</table>

**Note:**

1. This type of vGPU cannot be assigned with other types of vGPU to the same VM.
Maximum vGPUs per VM
NVIDIA vGPU software supports up to a maximum of 16 vGPUs per VM on Citrix Hypervisor.

Supported Hypervisor Releases
Citrix Hypervisor 8.2 only.

2.7. Peer-to-Peer CUDA Transfers over NVLink Support

Peer-to-peer CUDA transfers enable device memory between vGPUs on different GPUs that are assigned to the same VM to be accessed from within the CUDA kernels. NVLink is a high-bandwidth interconnect that enables fast communication between such vGPUs. Peer-to-Peer CUDA transfers over NVLink are supported only on a subset of vGPUs, Citrix Hypervisor releases, and guest OS releases.

Supported vGPUs
Only Q-series and C-series vGPUs that are allocated all of the physical GPU’s frame buffer on physical GPUs that support NVLink are supported.

<table>
<thead>
<tr>
<th>GPU Architecture</th>
<th>Board</th>
<th>vGPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampere</td>
<td>NVIDIA A40</td>
<td>A40-48Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A40-48C</td>
</tr>
<tr>
<td></td>
<td>NVIDIA A10</td>
<td>A10-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A10-24C</td>
</tr>
<tr>
<td></td>
<td>NVIDIA RTX A6000</td>
<td>A6000-48Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A6000-48C</td>
</tr>
<tr>
<td></td>
<td>NVIDIA RTX A5000</td>
<td>A5000-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A5000-24C</td>
</tr>
<tr>
<td>Turing</td>
<td>Quadro RTX 6000</td>
<td>RTX6000-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RTX6000-24C</td>
</tr>
<tr>
<td></td>
<td>Quadro RTX 6000 passive</td>
<td>RTX6000P-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RTX6000P-24C</td>
</tr>
<tr>
<td></td>
<td>Quadro RTX 8000</td>
<td>RTX8000-48Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RTX8000-48C</td>
</tr>
<tr>
<td></td>
<td>Quadro RTX 8000 passive</td>
<td>RTX8000P-48Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RTX8000P-48C</td>
</tr>
</tbody>
</table>
**GPU Architecture** | **Board** | **vGPU**
---|---|---
Volta | Tesla V100 SXM2 32GB | V100DX-32Q
| | | V100DX-32C
| Tesla V100 SXM2 | V100X-16Q
| | V100X-16C
Pascal | Tesla P100 SXM2 | P100X-16Q
| | P100X-16C

**Supported Hypervisor Releases**

Peer-to-Peer CUDA Transfers over NVLink are supported on all hypervisor releases that support the assignment of more than one vGPU to a VM. For details, see Multiple vGPU Support.

**Supported Guest OS Releases**

Linux only. Peer-to-Peer CUDA Transfers over NVLink are not supported on Windows.

**Limitations**

- Only direct connections are supported. NVSwitch is not supported.
- PCIe is not supported.
- SLI is not supported.

### 2.8. Unified Memory Support

Unified memory is a single memory address space that is accessible from any CPU or GPU in a system. It creates a pool of managed memory that is shared between the CPU and GPU to provide a simple way to allocate and access data that can be used by code running on any CPU or GPU in the system. Unified memory is supported only on a subset of vGPUs and guest OS releases.

**Note:** Unified memory is disabled by default. If used, you must enable unified memory individually for each vGPU that requires it by setting a vGPU plugin parameter. NVIDIA CUDA Toolkit profilers are supported and can be enabled on a VM for which unified memory is enabled.

**Supported vGPUs**

Only Q-series and C-series vGPUs that are allocated all of the physical GPU’s frame buffer on physical GPUs that support unified memory are supported.
<table>
<thead>
<tr>
<th>GPU Architecture</th>
<th>Board</th>
<th>vGPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampere</td>
<td>NVIDIA A40</td>
<td>A40-48Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A40-48C</td>
</tr>
<tr>
<td></td>
<td>NVIDIA A16</td>
<td>A16-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A16-16C</td>
</tr>
<tr>
<td></td>
<td>NVIDIA A10</td>
<td>A10-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A10-24C</td>
</tr>
<tr>
<td></td>
<td>NVIDIA A16</td>
<td>A2-16Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A2-16C</td>
</tr>
<tr>
<td></td>
<td>NVIDIA RTX A6000</td>
<td>A6000-48Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A6000-48C</td>
</tr>
<tr>
<td></td>
<td>NVIDIA RTX A5000</td>
<td>A5000-24Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A5000-24C</td>
</tr>
</tbody>
</table>

**Supported Guest OS Releases**

Linux only. Unified memory is **not** supported on Windows.

**Limitations**

- When unified memory is enabled for a VM, XenMotion with vGPU is disabled for the VM.

### 2.9. NVIDIA Deep Learning Super Sampling (DLSS) Support

NVIDIA vGPU software supports NVIDIA DLSS on NVIDIA RTX Virtual Workstation.

**Supported DLSS versions:** 2.0. Version 1.0 is **not** supported.

**Supported GPUs:**

- NVIDIA A40
- NVIDIA A16
- NVIDIA A2
- NVIDIA A10
- NVIDIA RTX A6000
- NVIDIA RTX A5000
- Tesla T4
- Quadro RTX 8000
Validated Platforms

- Quadro RTX 8000 passive
- Quadro RTX 6000
- Quadro RTX 6000 passive

Note: NVIDIA graphics driver components that DLSS requires are installed only if a supported GPU is detected during installation of the driver. Therefore, if the creation of VM templates includes driver installation, the template should be created from a VM that is configured with a supported GPU while the driver is being installed.

Supported applications: only applications that use nvngx_dlss.dll version 2.0.18 or newer
Chapter 3. Known Product Limitations

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

3.1. vGPUs of different types on the same GPU are not supported

Citrix Hypervisor does not support different vGPU types on the same GPU. All vGPUs on a single GPU must be of the same type.

3.2. NVENC does not support resolutions greater than 4096×4096

Description

The NVIDIA hardware-based H.264 video encoder (NVENC) does not support resolutions greater than 4096×4096. This restriction applies to all NVIDIA GPU architectures and is imposed by the GPU encoder hardware itself, not by NVIDIA vGPU software. The maximum supported resolution for each encoding scheme is listed in the documentation for NVIDIA Video Codec SDK. This limitation affects any remoting tool where H.264 encoding is used with a resolution greater than 4096×4096. Most supported remoting tools fall back to software encoding in such scenarios.

Workaround

If your GPU is based on a GPU architecture later than the NVIDIA Maxwell® architecture, use H.265 encoding. H.265 is more efficient than H.264 encoding and has a maximum resolution
of 8192×8192. On GPUs based on the NVIDIA Maxwell architecture, H.265 has the same
maximum resolution as H.264, namely 4096×4096.

Note: Resolutions greater than 4096×4096 are supported only by the H.265 decoder that 64-bit
client applications use. The H.265 decoder that 32-bit applications use supports a maximum
resolution of 4096×4096.

If you are using Citrix Virtual Apps and Desktops, which is a 32-bit application, switch to using
H.264 hardware encoding with the Use video codec for compression Citrix graphics policy set
to Actively Changing Regions. This policy setting encodes only actively changing regions of
the screen (for example, a window in which a video is playing). Provided that the number of
pixels along any edge of the actively changing region does not exceed 4096, H.264 encoding is
offloaded to the NVENC hardware encoder.

### 3.3. Issues occur when the channels allocated to a vGPU are exhausted

**Description**

Issues occur when the channels allocated to a vGPU are exhausted and the guest VM to
which the vGPU is assigned fails to allocate a channel to the vGPU. A physical GPU has a
fixed number of channels and the number of channels allocated to each vGPU is inversely
proportional to the maximum number of vGPUs allowed on the physical GPU.

When the channels allocated to a vGPU are exhausted and the guest VM fails to allocate a
channel, the following errors are reported on the hypervisor host or in an NVIDIA bug report:

Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0): Guest attempted to
allocate channel above its max channel limit 0xfb
Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0): VGPU message 6
failed, result code: 0x1a
Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0): 0xc1d004a1, 0xff0400fb, 0x36f,
Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0): 0xff1fe314, 0xff1fe038, 0x100b6f000, 0x1000,
Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0): , 0x1, 0x0
Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0): 0x80000000, 0xff0e0200, 0x0, 0x0, (Not logged),
Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0): 0x1, 0x0
Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0): , 0x0

**Workaround**

Use a vGPU type with more frame buffer, thereby reducing the maximum number of vGPUs
allowed on the physical GPU. As a result, the number of channels allocated to each vGPU is
increased.
3.4. **Total frame buffer for vGPUs is less than the total frame buffer on the physical GPU**

Some of the physical GPU’s frame buffer is used by the hypervisor on behalf of the VM for allocations that the guest OS would otherwise have made in its own frame buffer. The frame buffer used by the hypervisor is not available for vGPUs on the physical GPU. In NVIDIA vGPU deployments, frame buffer for the guest OS is reserved in advance, whereas in bare-metal deployments, frame buffer for the guest OS is reserved on the basis of the runtime needs of applications.

If error-correcting code (ECC) memory is enabled on a physical GPU that does not have HBM2 memory, the amount of frame buffer that is usable by vGPUs is further reduced. All types of vGPU are affected, not just vGPUs that support ECC memory.

On all GPUs that support ECC memory and, therefore, dynamic page retirement, additional frame buffer is allocated for dynamic page retirement. The amount that is allocated is inversely proportional to the maximum number of vGPUs per physical GPU. All GPUs that support ECC memory are affected, even GPUs that have HBM2 memory or for which ECC memory is disabled.

The approximate amount of frame buffer that NVIDIA vGPU software reserves can be calculated from the following formula:

\[
\text{max-reserved-fb} = \frac{\text{vgpu-profile-size-in-mb}}{16} + 16 + \text{ecc-adjustments} + \text{page-retirement-allocation} + \text{compression-adjustment} 
\]

- **max-reserved-fb**
  - The maximum total amount of reserved frame buffer in Mbytes that is not available for vGPUs.

- **vgpu-profile-size-in-mb**
  - The amount of frame buffer in Mbytes allocated to a single vGPU. This amount depends on the vGPU type. For example, for the T4-16Q vGPU type, vgpu-profile-size-in-mb is 16384.

- **ecc-adjustments**
  - The amount of frame buffer in Mbytes that is not usable by vGPUs when ECC is enabled on a physical GPU that does not have HBM2 memory.
    - If ECC is enabled on a physical GPU that does not have HBM2 memory ecc-adjustments is \(fb\)-without-ecc/16, which is equivalent to 64 Mbytes for every 8 byte of frame buffer assigned to the vGPU. \(fb\)-without-ecc is total amount of frame buffer with ECC disabled.
    - If ECC is disabled or the GPU has HBM2 memory, ecc-adjustments is 0.

- **page-retirement-allocation**
  - The amount of frame buffer in Mbytes that is reserved for dynamic page retirement.
    - On GPUs based on the NVIDIA Maxwell GPU architecture, page-retirement-allocation = \(4 \div \text{max-vgpus-per-gpu}\).
On GPUs based on NVIDIA GPU architectures after the Maxwell architecture, page-retirement-allocation = 128+max-vgpus-per-gpu

**max-vgpus-per-gpu**

The maximum number of vGPUs that can be created simultaneously on a physical GPU. This number varies according to the vGPU type. For example, for the T4-16Q vGPU type, max-vgpus-per-gpu is 1.

**compression-adjustment**

The amount of frame buffer in Mbytes that is reserved for the higher compression overhead in vGPU types with 12 Gbytes or more of frame buffer on GPUs based on the Turing architecture.

**compression-adjustment** depends on the vGPU type as shown in the following table.

<table>
<thead>
<tr>
<th>vGPU Type</th>
<th>Compression Adjustment (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4-16Q</td>
<td>28</td>
</tr>
<tr>
<td>T4-16C</td>
<td></td>
</tr>
<tr>
<td>T4-16A</td>
<td></td>
</tr>
<tr>
<td>RTX6000-12Q</td>
<td>32</td>
</tr>
<tr>
<td>RTX6000-12C</td>
<td></td>
</tr>
<tr>
<td>RTX6000-12A</td>
<td></td>
</tr>
<tr>
<td>RTX6000-24Q</td>
<td>104</td>
</tr>
<tr>
<td>RTX6000-24C</td>
<td></td>
</tr>
<tr>
<td>RTX6000-24A</td>
<td></td>
</tr>
<tr>
<td>RTX6000P-12Q</td>
<td>32</td>
</tr>
<tr>
<td>RTX6000P-12C</td>
<td></td>
</tr>
<tr>
<td>RTX6000P-12A</td>
<td></td>
</tr>
<tr>
<td>RTX6000P-24Q</td>
<td>104</td>
</tr>
<tr>
<td>RTX6000P-24C</td>
<td></td>
</tr>
<tr>
<td>RTX6000P-24A</td>
<td></td>
</tr>
<tr>
<td>RTX8000-12Q</td>
<td>32</td>
</tr>
<tr>
<td>RTX8000-12C</td>
<td></td>
</tr>
<tr>
<td>RTX8000-12A</td>
<td></td>
</tr>
<tr>
<td>RTX8000-16Q</td>
<td>64</td>
</tr>
<tr>
<td>RTX8000-16C</td>
<td></td>
</tr>
<tr>
<td>RTX8000-16A</td>
<td></td>
</tr>
<tr>
<td>RTX8000-24Q</td>
<td>96</td>
</tr>
<tr>
<td>RTX8000-24C</td>
<td></td>
</tr>
<tr>
<td>RTX8000-24A</td>
<td></td>
</tr>
<tr>
<td>vGPU Type</td>
<td>Compression Adjustment (MB)</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>RTX8000-48Q</td>
<td>238</td>
</tr>
<tr>
<td>RTX8000-48C</td>
<td></td>
</tr>
<tr>
<td>RTX8000-48A</td>
<td></td>
</tr>
<tr>
<td>RTX8000P-12Q</td>
<td>32</td>
</tr>
<tr>
<td>RTX8000P-12C</td>
<td></td>
</tr>
<tr>
<td>RTX8000P-12A</td>
<td></td>
</tr>
<tr>
<td>RTX8000P-16Q</td>
<td>64</td>
</tr>
<tr>
<td>RTX8000P-16C</td>
<td></td>
</tr>
<tr>
<td>RTX8000P-16A</td>
<td></td>
</tr>
<tr>
<td>RTX8000P-24Q</td>
<td>96</td>
</tr>
<tr>
<td>RTX8000P-24C</td>
<td></td>
</tr>
<tr>
<td>RTX8000P-24A</td>
<td></td>
</tr>
<tr>
<td>RTX8000P-48Q</td>
<td>238</td>
</tr>
<tr>
<td>RTX8000P-48C</td>
<td></td>
</tr>
<tr>
<td>RTX8000P-48A</td>
<td></td>
</tr>
</tbody>
</table>

For all other vGPU types, compression-adjustment is 0.

Note: In VMs running Windows Server 2012 R2, which supports Windows Display Driver Model (WDDM) 1.x, an additional 48 Mbytes of frame buffer are reserved and not available for vGPUs.

3.5. Issues may occur with graphics-intensive OpenCL applications on vGPU types with limited frame buffer

Description

Issues may occur when graphics-intensive OpenCL applications are used with vGPU types that have limited frame buffer. These issues occur when the applications demand more frame buffer than is allocated to the vGPU.

For example, these issues may occur with the Adobe Photoshop and LuxMark OpenCL Benchmark applications:

- When the image resolution and size are changed in Adobe Photoshop, a program error may occur or Photoshop may display a message about a problem with the graphics hardware and a suggestion to disable OpenCL.
Known Product Limitations

- When the LuxMark OpenCL Benchmark application is run, XID error 31 may occur.

**Workaround**

For graphics-intensive OpenCL applications, use a vGPU type with more frame buffer.

### 3.6. In pass through mode, all GPUs connected to each other through NVLink must be assigned to the same VM

**Description**

In pass through mode, all GPUs connected to each other through NVLink must be assigned to the same VM. If a subset of GPUs connected to each other through NVLink is passed through to a VM, unrecoverable error XID 74 occurs when the VM is booted. This error corrupts the NVLink state on the physical GPUs and, as a result, the NVLink bridge between the GPUs is unusable.

**Workaround**

Restore the NVLink state on the physical GPUs by resetting the GPUs or rebooting the hypervisor host.

### 3.7. 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
Workaround
Use a profile that supports more than 1 virtual display head and has at least 1 Gbyte of frame buffer.

3.8. 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

Workaround
If you require NVENC to be enabled, use a profile that has at least 1 Gbyte of frame buffer.

3.9. VM running an incompatible NVIDIA vGPU guest driver fails to initialize vGPU when booted

Description
A VM running a version of the NVIDIA guest VM driver that is incompatible with the current release of Virtual GPU Manager will fail to initialize vGPU when booted on a Citrix Hypervisor platform running that release of Virtual GPU Manager.

A guest VM driver is incompatible with the current release of Virtual GPU Manager in either of the following situations:
The guest driver is from a release in a branch two or more major releases before the current release, for example release 9.4.

In this situation, the Citrix Hypervisor VM’s /var/log/messages log file reports the following error:

```plaintext
vmiop_log: (0x0): Incompatible Guest/Host drivers: Guest VGX version is older than the minimum version supported by the Host. Disabling vGPU.
```

The guest driver is from a later release than the Virtual GPU Manager.

In this situation, the Citrix Hypervisor VM’s /var/log/messages log file reports the following error:

```plaintext
vmiop_log: (0x0): Incompatible Guest/Host drivers: Guest VGX version is newer than the maximum version supported by the Host. Disabling vGPU.
```

In either situation, the VM boots in standard VGA mode with reduced resolution and color depth. The NVIDIA virtual GPU is present in Windows Device Manager but displays a warning sign, and the following device status:

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

**Resolution**

Install a release of the NVIDIA guest VM driver that is compatible with current release of Virtual GPU Manager.

### 3.10. Single vGPU benchmark scores are lower than pass-through GPU

**Description**

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

Aside from performance differences that may be attributed to a vGPU’s smaller frame buffer size, vGPU incorporates a performance balancing feature known as Frame Rate Limiter (FRL). On vGPUs that use the best-effort scheduler, FRL is enabled. On vGPUs that use the fixed share or equal share scheduler, FRL is disabled.

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 pass-through GPU.

**Resolution**

FRL is controlled by an internal vGPU setting. On vGPUs that use the best-effort scheduler, NVIDIA does not validate vGPU with FRL disabled, but for validation of benchmark
performance, FRL can be temporarily disabled by specifying `frame_rate_limiter=0` in the vGPU’s plugin parameters.

- **Citrix Hypervisor 8.1 or later**: Specify `frame_rate_limiter=0` in the `extra_args` parameter for the group to which the vGPU belongs:

  ```
  [root@xenserver ~]# xe vgpu-param-set uuid=vgpu-uuid extra_args=frame_rate_limiter=0
  ```

- **Citrix Hypervisor earlier than 8.1**: Specify `frame_rate_limiter=0` in the VM’s `platform:vgpu_extra_args` parameter:

  ```
  [root@xenserver ~]# xe vm-param-set uuid=vm-uuid
  platform:vgpu_extra_args="frame_rate_limiter=0"
  ```

The setting takes effect the next time the VM is started or rebooted.

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 in one of the following ways:

- Removing the vGPU plugin’s parameters:
  - **Citrix Hypervisor 8.1 or later**: Removing the `extra_args` key the from group to which the vGPU belongs
  - **Citrix Hypervisor earlier than 8.1**: Removing the `vgpu_extra_args` key from the `platform` parameter

- Removing `frame_rate_limiter=0` from the `extra_args` or `vgpu_extra_args` key

- Setting `frame_rate_limiter=1`. For example:

  ```
  - **Citrix Hypervisor 8.1 or later**:
    ```
    [root@xenserver ~]# xe vgpu-param-set uuid=vgpu-uuid
    extra_args=frame_rate_limiter=1
    ```
  - **Citrix Hypervisor earlier than 8.1**:
    ```
    [root@xenserver ~]# xe vm-param-set uuid=vm-uuid
    platform:vgpu_extra_args="frame_rate_limiter=1"
    ```

### 3.11. `nvidia-smi` fails to operate when all GPUs are assigned to GPU pass-through mode

#### Description

If all GPUs in the platform are assigned to VMs in pass-through mode, `nvidia-smi` will return an error:

```
[root@xenserver-vgx-test ~]# nvidia-smi
Failed to initialize NVML: Unknown Error
```

This is because GPUs operating in pass-through mode are not visible to `nvidia-smi` and the NVIDIA kernel driver operating in the Citrix Hypervisor dom0.
To confirm that all GPUs are operating in pass-through mode, use XenCenter's GPU tab to review current GPU assignment:

Resolution
N/A
3.12. Windows Aero is disabled on Citrix Virtual Apps and Desktops session using 3 or 4 monitors in 2560×1600 resolution

Description

Windows Aero may be disabled when Citrix Virtual Apps and Desktops is connected to a VM with a vGPU or passthrough GPU, with 3 or 4 monitors at 2560×1600 resolution.

This limitation is a limitation of Windows 7. For details, see the Microsoft knowledge base article Desktop background disappears with very large extended desktop on Windows 7.

3.13. 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 standard VGA mode with reduced resolution and color depth. The NVIDIA vGPU software 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 Citrix Hypervisor VM’s /var/log/messages log file contains these error messages:

vmiop_log: error: NVOS status 0x29
vmiop_log: error: Assertion Failed at 0x7620fd4b:179
vmiop_log: error: 8 frames returned by backtrace
...  
vmiop_log: error: VGPU message 12 failed, result code: 0x29
...  
vmiop_log: error: NVOS status 0x8
vmiop_log: error: Assertion Failed at 0x7620c8df:280
vmiop_log: error: 8 frames returned by backtrace
...  
vmiop_log: error: 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 specifying enable_large_sys_mem=1 in the vGPU’s plugin parameters.
Known Product Limitations

- **Citrix Hypervisor 8.1 or later:** Specify `enable_large_sys_mem=1` in the `extra_args` parameter for the group to which the vGPU belongs:
  
  ```
  [root@xenserver ~]# xe vgpu-param-set uuid=vgpu-uuid extra_args=enable_large_sys_mem=1
  ```

- **Citrix Hypervisor earlier than 8.1:** Specify `enable_large_sys_mem=1` in the VM’s `platform:vgpu_extra_args` parameter:
  
  ```
  [root@xenserver ~]# xe vm-param-set uuid=vm-uuid
  platform:vgpu_extra_args="enable_large_sys_mem=1"
  ```

The setting takes effect the next time the VM is started or rebooted. With this setting in place, less GPU FB is available to applications running in the VM. To accommodate system memory larger than 64GB, the reservation can be further increased by specifying `extra_fb_reservation` in the VM’s `platform:vgpu_extra_args` parameter, 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 `extra_fb_reservation` to 128:

- **Citrix Hypervisor 8.1 or later:**
  
  ```
  extra_args="enable_large_sys_mem=1,extra_fb_reservation=128"
  ```

- **Citrix Hypervisor earlier than 8.1:**
  
  ```
  platform:vgpu_extra_args="enable_large_sys_mem=1, extra_fb_reservation=128"
  ```

The reservation can be reverted back to its default setting in one of the following ways:

- Removing the vGPU plugin’s parameters:
  
  - **Citrix Hypervisor 8.1 or later:** Removing the `extra_args` key from group to which the vGPU belongs
  - **Citrix Hypervisor earlier than 8.1:** Removing the `vgpu_extra_args` key from the `platform` parameter
  - Removing `enable_large_sys_mem` from the `extra_args` or `vgpu_extra_args` key
  - Setting `enable_large_sys_mem=0`

3.14. **vGPU host driver RPM upgrade fails**

**Description**

Upgrading vGPU host driver RPM fails with an error message about failed dependencies on the console.

```
[root@xenserver ~]# rpm -U NVIDIA-vGPU-xenserver-6.5-352.46.x86_64.rpm
error: Failed dependencies:
  NVIDIA-vgx-xenserver conflicts with NVIDIA-vGPU-xenserver-6.5-352.46.x86_64
[root@xenserver ~]#
```

**Resolution**

Uninstall the older vGPU RPM before installing the latest driver.

Use the following command to uninstall the older vGPU RPM:
[root@xenserver ~]# rpm -e NVIDIA-vgx-xenserver
Chapter 4. Resolved Issues

Only resolved issues that have been previously noted as known issues or had a noticeable user impact are listed. The summary and description for each resolved issue indicate the effect of the issue on NVIDIA vGPU software before the issue was resolved.

Issues Resolved in Release 14.0

No resolved issues are reported in this release for Citrix Hypervisor.
Chapter 5.  Known Issues

5.1.  For some license deployments, GSP firmware remains enabled with unsupported products

Description

If GPU System Processor (GSP) firmware is enabled with an unsupported product in a GPU pass through or bare-metal configuration on Linux, the VM or bare-metal host should fail to acquire a license. This behavior is implemented to prevent the VM or host from being in an unsupported configuration. However, only license acquisition from a networked legacy NVIDIA vGPU software license server fails. For all other license deployments, the VM or host acquires a license and is in an unsupported configuration.

GSP is supported only for vCS in GPU pass through and bare-metal deployments on Linux. If you are using any other product in a GPU pass through or bare-metal deployment on Linux, you must disable the GSP firmware.

Note: For NVIDIA vGPU deployments on Linux and all NVIDIA vGPU software deployments on Windows, GSP is also not supported but GSP firmware is already disabled. For these deployments, this issue does not arise.

The GSP firmware might be enabled with an unsupported product in a GPU pass through or bare-metal configuration on Linux. In this situation, the following error message is written to the licensing event log file when the VM or host attempts to acquire a license:

Invalid feature requested for the underlying GSP firmware configuration.
Disable GSP firmware to use this feature.

For the location of the licensing event log file, refer to Virtual GPU Client Licensing User Guide.

Workaround

Ensure that the GSP firmware is disabled as explained in Virtual GPU Software User Guide.
5.2. After an upgrade of the Linux graphics driver from an RPM package in a licensed VM, licensing fails

Description
After the NVIDIA vGPU software graphics driver for Linux is upgraded from an RPM package in a licensed VM, licensing fails. The `nvidia-smi vgpu -q` command shows the driver version and license status as N/A. Restarting the `nvidia-gridd` service fails with a `Unit not found` error.

Workaround
Perform a clean installation of the NVIDIA vGPU software graphics driver for Linux from an RPM package.

1. Remove the currently installed driver.
2. Install the new version of the driver.
   
   ```bash
   rpm -iv nvidia-linux-grid-510_510.47.03_amd64.rpm
   ```

Status
Open

Ref. #
3512766
5.3. After an upgrade of the Linux graphics driver from a Debian package, the driver is not loaded into the VM

Description
After the NVIDIA vGPU software graphics driver for Linux is upgraded from a Debian package, the driver is not loaded into the VM.

Workaround
Use one of the following workarounds to load the driver into the VM:

- Reboot the VM.
- Remove the `nvidia` module from the Linux kernel and reinsert it into the kernel.

1. Remove the `nvidia` module from the Linux kernel.
   
   \$ `sudo rmmod nvidia`

2. Reinsert the `nvidia` module into the Linux kernel.
   
   \$ `sudo modprobe nvidia`

Status
Not a bug

Ref. #
200748806

5.4. Desktop session freezes when a Linux VM is migrated to or from a host running an NVIDIA vGPU software 14 release

Description
When a Linux VM configured with a Tesla V100 or Tesla T4 vGPU is migrated between a host running an NVIDIA vGPU software 14 release and a host running a an NVIDIA vGPU software...
13 release, the remote desktop session freezes. After the session freezes, the VM must be rebooted to recover the session. This issue occurs only when the NVIDIA hardware-based H.264/HEVC video encoder (NVENC) is enabled.

Version

The issue affects migrations between a host running an NVIDIA vGPU software 14 release and a host running an NVIDIA vGPU software 13 release.

Workaround

Disable NVENC.

Status

Open

Ref. #

3512790

5.5. The reported NVENC frame rate is double the actual frame rate

Description

The frame rate in frames per second (FPS) for the NVIDIA hardware-based H.264/HEVC video encoder (NVENC) reported by the `nvidia-smi encodersessions` command and NVWMI is double the actual frame rate. Only the reported frame rate is incorrect. The actual encoding of frames is not affected.

This issue affects only Windows VMs that are configured with NVIDIA vGPU.

Status

Open

Ref. #

2997564
5.6. Hypervisor host reboots when multiple cloned VMs are simultaneously powered on or migrated

Description
When multiple cloned VMs are simultaneously powered on or migrated, the hypervisor host reboots. No crash dump is generated and no error messages related to the reboot are written to the log files daemon.log or kern.log.

Version
This issue affects only Citrix Hypervisor 8.2.

Status
Open

Ref. #
200726850

5.7. NVENC does not work with Teradici Cloud Access Software on Windows

Description
The NVIDIA hardware-based H.264/HEVC video encoder (NVENC) does not work with Teradici Cloud Access Software on Windows. This issue affects NVIDIA vGPU and GPU pass through deployments.

This issue occurs because the check that Teradici Cloud Access Software performs on the DLL signer name is case sensitive and NVIDIA recently changed the case of the company name in the signature certificate.

Status
Not an NVIDIA bug

This issue is resolved in the latest 21.07 and 21.03 Teradici Cloud Access Software releases.
5.8. A licensed client might fail to acquire a license if a proxy is set

Description
If a proxy is set with a system environment variable such as `HTTP_PROXY` or `HTTPS_PROXY`, a licensed client might fail to acquire a license.

Workaround
Perform this workaround on each affected licensed client.

1. Add the address of the NVIDIA vGPU software license server to the system environment variable `NO_PROXY`.
   - The address must be specified exactly as it is specified in the client’s license server settings either as a fully-qualified domain name or an IP address. If the `NO_PROXY` environment variable contains multiple entries, separate the entries with a comma (,).
   - If high availability is configured for the license server, add the addresses of the primary license server and the secondary license server to the system environment variable `NO_PROXY`.
2. Restart the NVIDIA driver service that runs the core NVIDIA vGPU software logic.
   - On Windows, restart the `NVIDIA Display Container` service.
   - On Linux, restart the `nvidia-gridd` service.

Status
Closed

Ref. #
200749065
5.9. Session connection fails with four 4K displays and NVENC enabled on a 2Q, 3Q, or 4Q vGPU

Description

Desktop session connections fail for a 2Q, 3Q, or 4Q vGPU that is configured with four 4K displays and for which the NVIDIA hardware-based H.264/HEVC video encoder [NVENC] is enabled. This issue affects only Teradici Cloud Access Software sessions on Linux guest VMs.

This issue is accompanied by the following error message:

This Desktop has no resources available or it has timed out

This issue is caused by insufficient frame buffer.

Workaround

Ensure that sufficient frame buffer is available for all the virtual displays that are connected to a vGPU by changing the configuration in one of the following ways:

- Reducing the number of virtual displays. The number of 4K displays supported with NVENC enabled depends on the vGPU.

<table>
<thead>
<tr>
<th>vGPU</th>
<th>4K Displays Supported with NVENC Enabled</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Q</td>
<td>1</td>
</tr>
<tr>
<td>3Q</td>
<td>2</td>
</tr>
<tr>
<td>4Q</td>
<td>3</td>
</tr>
</tbody>
</table>

- Disabling NVENC. The number of 4K displays supported with NVENC disabled depends on the vGPU.

<table>
<thead>
<tr>
<th>vGPU</th>
<th>4K Displays Supported with NVENC Disabled</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Q</td>
<td>2</td>
</tr>
<tr>
<td>3Q</td>
<td>2</td>
</tr>
<tr>
<td>4Q</td>
<td>4</td>
</tr>
</tbody>
</table>

- Using a vGPU type with more frame buffer. Four 4K displays with NVENC enabled on any Q-series vGPU with at least 6144 MB of frame buffer are supported.

Status

Not an NVIDIA bug
5.10. Disconnected sessions cannot be reconnected or might be reconnected very slowly with NVWMI installed

**Description**
Disconnected sessions cannot be reconnected or might be reconnected very slowly when the NVIDIA Enterprise Management Toolkit (NVWMI) is installed. This issue affects Citrix Virtual Apps and Desktops and VMware Horizon sessions on Windows guest VMs.

**Workaround**
Uninstall NVWMI.

**Status**
Open

Ref. #
3262923

5.11. Linux VM hangs after vGPU migration to a host running a newer vGPU manager version

**Description**
When a Linux VM configured with a Tesla V100 or Tesla T4 vGPU is migrated from a host that is running a vGPU manager 11 release before 11.6 to a host that is running a vGPU manager 13 release, the VM hangs. After the migration, the destination host and VM become unstable. When this issue occurs, XID error 31 is written to the log files on the destination hypervisor host.
Known Issues

Version

This issue affects migration from a host that is running a vGPU manager 11 release before 11.6 to a host that is running a vGPU manager 13 release.

Workaround

If the VM is configured with a Tesla T4 vGPU, perform the following sequence of steps before attempting the migration:

1. Upgrade the host that is running a vGPU manager 11 release to release 11.6 or a later vGPU manager 11 release.
2. Disconnect any remoting tool that is using NVENC.

Note: You cannot use this workaround for a VM that is configured with a Tesla V100 vGPU.

Status

Open

Ref. #

200691445

5.12. Idle Teradici Cloud Access Software session disconnects from Linux VM

Description

After a Teradici Cloud Access Software session has been idle for a short period of time, the session disconnects from the VM. When this issue occurs, the error messages NVOS status 0x19 and vGPU Message 21 failed are written to the log files on the hypervisor host. This issue affects only Linux guest VMs.

Status

Open

Ref. #

200689126
5.13. No virtual GPU types are listed in Citrix XenCenter

Description
No virtual GPU (vGPU) types are listed on the Select virtual GPU type drop-down list in Citrix XenCenter. However, the command `nvidia-smi vgpu -s` lists all supported vGPU types. This issue occurs because the driver version number of the Virtual GPU Manager contains multiple decimal points and is misinterpreted by Citrix XenCenter.

Version
Citrix Hypervisor 8.2

Resolution
Apply Hotfix XS82E002 - For Citrix Hypervisor 8.2 from Citrix.

Status
Not an NVIDIA bug. Resolved by Citrix in hotfix XS82E002 for Citrix Hypervisor 8.2

Ref. #
200653755
5.14. **NVIDIA vGPU software graphics driver fails to load after upgrade on Citrix Hypervisor**

**Description**

NVIDIA vGPU software graphics driver fails to load after an upgrade of NVIDIA vGPU software from an 11.x release to a 12.x release on Citrix Hypervisor. This issue occurs because the hypervisor name in the RPM package name has changed from `xenserver` to `CitrixHypervisor`.

The attempted upgrade fails its dependency checks as follows:

```bash
[root@localhost ~]# rpm -Uvh NVIDIA-vGPU-CitrixHypervisor-8.2-510.47.03 .x86_64.rpm
error: Failed dependencies:
NVIDIA-vGPU-xenserver conflicts with NVIDIA-vGPU-CitrixHypervisor-1:8.2-510.47.03 .x86_64
```

**Workaround**

1. Determine the full name of the installed RPM package for the Virtual GPU Manager.
   ```bash
   [root@localhost ~]# rpm -qa | grep NV
   NVIDIA-vGPU-xenserver-8.2-470.82.x86_64
   ```

2. Remove the installed RPM package for the Virtual GPU Manager.
   ```bash
   [root@localhost ~]# rpm -ev NVIDIA-vGPU-xenserver-8.2-470.82.x86_64
   Preparing packages...
   NVIDIA-vGPU-xenserver-1:8.2-470.82.x86_64
   ```

3. Perform a fresh installation of the latest RPM package for the Virtual GPU Manager.
   ```bash
   [root@localhost ~]# rpm -ivh NVIDIA-vGPU-CitrixHypervisor-8.2-510.47.03 .x86_64.rpm
   Preparing...                          ################################# [100%]
   Updating / installing...
   1:NVIDIA-vGPU-CitrixHypervisor-1:8.2-510.47.03 .x86_64 [100%]
   ```

4. Reboot the hypervisor host.
   ```bash
   [root@localhost ~]# reboot
   ```

**Status**

Open

**Ref. #**

200682984
5.15. Windows guest VMs with vGPUs or GPUs with large BAR memory settings fail to boot to the desktop in UEFI mode

Description
Windows guest VMs configured with vGPUs or physical GPUs that have large BAR memory settings fail to boot to the desktop in UEFI mode. The VM can be reached through its IP address but the desktop session is blank after the connection to the VM is established.

By default, Citrix Hypervisor creates a VM in UEFI boot mode and the boot mode of a VM cannot be changed after the VM is created.

Version
Citrix Hypervisor 8.2

Workaround
Delete the VM and re-create it, selecting BIOS Boot mode when you specify the installation method for the OS software on the new VM.

Status
Open

Ref. #
200676622
5.16. Migrated VM with more than two vGPUs crashes on destination host

Description
When a VM that is configured with more than two vGPUs is migrated, the VM crashes on the destination host. When this issue occurs, the log file on the hypervisor host is flooded with NVOS status messages 0x23, 0x33, 0x5, and 0x21 and vGPU failure messages 4, 6, 7, 9, 12, 14, 19, 21, 23, 26, 32, 43, 54, and 56.

Status
Open

Ref. #
200660221

5.17. Driver upgrade in a Linux guest VM with multiple vGPUs might fail

Description
Upgrading the NVIDIA vGPU software graphics driver in a Linux guest VM with multiple vGPUs might fail. This issue occurs if the driver is upgraded by overinstalling the new release of the driver on the current release of the driver while the nvidia-gridd service is running in the VM.

Workaround
1. Stop the nvidia-gridd service.
2. Try again to upgrade the driver.

Status
Open

Ref. #
200633548
5.18. **NVIDIA Control Panel** fails to start if launched too soon from a VM without licensing information

**Description**
If NVIDIA licensing information is not configured on the system, any attempt to start **NVIDIA Control Panel** by right-clicking on the desktop within 30 seconds of the VM being started fails.

**Workaround**
Restart the VM and wait at least 30 seconds before trying to launch **NVIDIA Control Panel**.

**Status**
Open

**Ref. #**
200623179

5.19. **VNC client session goes blank and console VNC is corrupted when the guest driver is uninstalled**

**Description**
When the NVIDIA vGPU software graphics driver is uninstalled from a Windows 10 guest VM for which the boot mode is UEFI, the VNC client session goes blank and console VNC is corrupted.

This issue affects only Windows 10 guest VMs for which the boot mode is UEFI. It does **not** affect Windows 10 guest VMs for which the boot mode is BIOS.

**Workaround**
Reboot the VM by using Citrix XenCenter.

**Status**
Not an NVIDIA bug
5.20. **Citrix Virtual Apps and Desktops session corruption occurs in the form of residual window borders**

**Description**
When a window is dragged across the desktop in a Citrix Virtual Apps and Desktops session, corruption of the session in the form of residual window borders occurs.

**Version**
This issue affects only Citrix Virtual Apps and Desktops version 7 2003

**Workaround**
Use Citrix Virtual Apps and Desktops version 7 1912 or 2006.

**Status**
Not an NVIDIA bug

**Ref. #**
200608675

5.21. **Suspend and resume between hosts running different versions of the vGPU manager fails**

**Description**
Suspending a VM configured with vGPU on a host running one version of the vGPU manager and resuming the VM on a host running a version from an older main release branch fails. For example, suspending a VM on a host that is running the vGPU manager from release 14.0 and resuming the VM on a host running the vGPU manager from release 13.2 fails. When this issue occurs, the error *There were no servers available to complete the specified operation* is reported on Citrix XenCenter.
Status
Not an NVIDIA bug

Ref. #
200602087

5.22. On Linux, the frame rate might drop to 1 after several minutes

Description
On Linux, the frame rate might drop to 1 frame per second (FPS) after NVIDIA vGPU software has been running for several minutes. Only some applications are affected, for example, glxgears. Other applications, such as Unigine Heaven, are not affected. This behavior occurs because Display Power Management Signaling (DPMS) for the Xorg server is enabled by default and the display is detected to be inactive even when the application is running. When DPMS is enabled, it enables power saving behavior of the display after several minutes of inactivity by setting the frame rate to 1 FPS.

Workaround
1. If necessary, stop the Xorg server.
   
   ```
   # /etc/init.d/xorg stop
   ```
2. In a plain text editor, edit the `/etc/X11/xorg.conf` file to set the options to disable DPMS and disable the screen saver.

   a). In the `Monitor` section, set the `DPMS` option to `false`.
   ```
   Option "DPMS" "false"
   ```

   b). At the end of the file, add a `ServerFlags` section that contains option to disable the screen saver.
   ```
   Section "ServerFlags"
   Option "BlankTime" "0"
   EndSection
   ```

   c). Save your changes to `/etc/X11/xorg.conf` file and quit the editor.
3. Start the Xorg server.
   ```
   # etc/init.d/xorg start
   ```

Status
Open
5.23.  Citrix XenCenter displays misleading information about vGPU types

Description

The **GPU type** drop-down list in the Citrix XenCenter management GUI displays misleading information about the maximum combined resolution supported by Q-series and B-series vGPU types. The GUI lists the maximum number of displays supported and the highest resolution supported in a manner that implies a maximum combined resolution higher than is actually supported. These vGPU types support the number of displays listed only at resolutions lower than the resolution listed and support the resolution listed only with fewer displays than the number listed.

For example, as shown in the following screen capture, the GUI lists a resolution of 5120×2880 and four displays for the V100-1Q vGPU type. However, this vGPU type supports only one display at a resolution of 5120×2880 and supports a maximum of four displays only at resolutions of 2560×1600 and lower.

![Screen capture showing misleading vGPU type information](image)

Version

All supported Citrix Hypervisor versions.

Workaround

For information about display configurations supported by Q-series and B-series vGPU types, see [Virtual GPU Software User Guide](#).
5.24. Citrix Virtual Apps and Desktops connection freezes initially

Description
When -0B and -0Q vGPU types are used with Citrix Virtual Apps and Desktops version 7 1903 and later versions, the session freezes or a black screen is seen when the connection is first made. When this issue occurs, the error message DXGI_ERROR_DEVICE_REMOVED is displayed. The affected versions of Citrix Virtual Apps and Desktops use Microsoft DDAPI.

Version
Citrix Virtual Apps and Desktops version 7 1903 and later versions

Workaround
Disable the use hardware encoding for video codec Citrix policy, which is enabled by default.

Status
Open

Ref. #
200494400

5.25. DWM crashes randomly occur in Windows VMs

Description
Desktop Windows Manager (DWM) crashes randomly occur in Windows VMs, causing a blue-screen crash and the bug check CRITICAL_PROCESS_DIED. Computer Management shows problems with the primary display device.
Version
This issue affects Windows 10 1809, 1903 and 1909 VMs.

Status
Not an NVIDIA bug

Ref. #
2730037

5.26. NVIDIA Control Panel fails to launch in a platform layer or published image

Description
When NVIDIA vGPU software is used with Citrix App Layering and a platform layer for a vGPU is being configured, NVIDIA Control Panel might fail to launch. The driver might be working normally or it might fail with code 31 or code 43.

For more information and a workaround for this issue, see NVIDIA Control Panel fails to launch in platform layer or published image in the Citrix Support Knowledge Center.

5.27. Remote desktop session freezes with assertion failure and XID error 43 after migration

Description
After multiple VMs configured with vGPU on a single hypervisor host are migrated simultaneously, the remote desktop session freezes with an assertion failure and XID error 43. This issue affects only GPUs that are based on the Volta GPU architecture. It does not occur if only a single VM is migrated.

When this error occurs, the following error messages are logged to the Citrix Hypervisor log file /var/log/messages:

```
Jan  3 14:35:48 ch81-m1 vgpu-12[8050]: error: vmiop_log: NVOS status 0x1f
Jan  3 14:35:48 ch81-m1 vgpu-12[8050]: error: vmiop_log: Assertion Failed at
0x4b8cacf6:286
...```
Jan 3 14:35:59 ch81-m1 vgpu-12[8050]: error: vmiop_log: (0x0): XID 43 detected on physical_chid:0x174, guest_chid:0x14

Status
Open

Ref. #
200581703

5.28. Citrix Virtual Apps and Desktops session freezes when the desktop is unlocked

Description
When a Citrix Virtual Apps and Desktops session that is locked is unlocked by pressing Ctrl+Alt+Del, the session freezes. This issue affects only VMs that are running Microsoft Windows 10 1809 as a guest OS.

Version
Microsoft Windows 10 1809 guest OS

Workaround
Restart the VM.

Status
Not an NVIDIA bug

Ref. #
2767012
5.29. **NVIDIA vGPU software graphics driver fails after Linux kernel upgrade with DKMS enabled**

**Description**

After the Linux kernel is upgraded (for example by running `sudo apt full-upgrade`) with Dynamic Kernel Module Support (DKMS) enabled, the `nvidia-smi` command fails to run. If DKMS is enabled, an upgrade to the Linux kernel triggers a rebuild of the NVIDIA vGPU software graphics driver. The rebuild of the driver fails because the compiler version is incorrect. Any attempt to reinstall the driver fails because the kernel fails to build.

When the failure occurs, the following messages are displayed:

```
-> Installing DKMS kernel module:
  ERROR: Failed to run `/usr/sbin/dkms build -m nvidia -v 510.47.03 -k 5.3.0-28-generic`:
  Kernel preparation unnecessary for this kernel. Skipping...
  Building module:
  cleaning build area...
  'make' -j8 NV_EXCLUDE_BUILD_MODULES='' KERNEL_UNAME=5.3.0-28-generic IGNORE_CC_MISMATCH=''
  modules...(bad exit status: 2)
  ERROR (dkms apport): binary package for nvidia: 510.47.03 not found
  Error! Bad return status for module build on kernel: 5.3.0-28-generic
(x86_64)
  Consult /var/lib/dkms/nvidia/ 510.47.03/build/make.log for more information.
  -> error.
  ERROR: Failed to install the kernel module through DKMS. No kernel module was installed;
  please try installing again without DKMS, or check the DKMS logs for more information.
  ERROR: Installation has failed. Please see the file '/var/log/nvidia-installer.log' for details.
  You may find suggestions on fixing installation problems in the README available on the Linux driver download page at www.nvidia.com.
```

**Workaround**

When installing the NVIDIA vGPU software graphics driver with DKMS enabled, use one of the following workarounds:

- Before running the driver installer, install the `dkms` package, then run the driver installer with the `-dkms` option.
- Run the driver installer with the `--no-cc-version-check` option.

**Status**

Not a bug.
5.30. On Citrix Hypervisor, all vGPUs in a VM must be of the same type

Description

When a VM on Citrix Hypervisor is configured with multiple vGPUs, all vGPUs must be of the same type. Citrix Hypervisor does not support the assignment of multiple vGPUs of different types to a single VM.

What happens when you try to assign a different type of vGPU than the vGPUs already assigned to a VM depends on the Citrix utility that you are using.

- If you are using Citrix XenCenter, only vGPUs of the same type as the vGPUs already assigned are available for selection.
- If you are using the `xe` command, the command fails and the following error message is displayed:

```
Cannot create a virtual GPU that is incompatible with the existing types on the VM.
```

Status

Not an NVIDIA bug.

Ref. #

2836271

5.31. Console VGA cannot be disabled

Description

In Citrix Hypervisor releases since 8.1, a VM’s console VGA interface cannot be disabled by using the command `xe vm-param-set uuid=vm-uuid platform:vgpu_extra_args="disable_vnc=1"`.

This behavior is the result of the following changes that Citrix introduced in Citrix Hypervisor release 8.1:

- The command `xe vm-param-set` for assigning plugin configuration parameters has been withdrawn.
Known Issues

- Plugin parameters can no longer be set for vGPUs that were created by using Citrix XenCenter. They can be set only for vGPUs that were created from the command line by using the `xe` command.

**Version**

Since Citrix Hypervisor release 8.1

**Workaround**

Create the vGPU by using the `xe` command and specify plugin parameters for the group to which the vGPU belongs as explained in Virtual GPU Software User Guide.

**Status**

Not a bug

**Ref. #**

200569785

5.32. Console VNC is unusable with Xorg on multiple vGPUs in a VM

**Description**

If a Linux VM is configured with multiple vGPUs and the Xorg display server is running on more than one vGPU in the VM, severe corruption in console VNC is observed.

**Workaround**

Ensure that the Xorg display server is running on only one vGPU in the VM.

**Status**

Not an NVIDIA bug.

**Ref. #**

200568261
5.33. **RAPIDS cuDF merge fails on NVIDIA vGPU**

**Description**

The `merge` function of the RAPIDS cuDF GPU data frame library fails on NVIDIA vGPU. This function fails because RAPIDS uses the Unified Memory feature of CUDA, which NVIDIA vGPU does not support.

**Status**

Open

**Ref. #**

2642134

5.34. **Migration of VMs configured with vGPU stops before the migration is complete**

**Description**

When a VM configured with vGPU is migrated to another host, the migration stops before it is complete. When this issue occurs, the message `The VGPU is not compatible with any PGPU in the destination.` appears in the Citrix XenCenter management GUI.

This issue occurs if the ECC memory configuration (enabled or disabled) on the source and destination hosts are different. The ECC memory configuration on both the source and destination hosts must be identical.

**Workaround**

Before attempting to migrate the VM again, ensure that the ECC memory configuration on both the source and destination hosts are identical.

**Status**

Not an NVIDIA bug
Known Issues

Ref. #
200520027

5.35. ECC memory settings for a vGPU cannot be changed by using NVIDIA X Server Settings

Description
The ECC memory settings for a vGPU cannot be changed from a Linux guest VM by using NVIDIA X Server Settings. After the ECC memory state has been changed on the ECC Settings page and the VM has been rebooted, the ECC memory state remains unchanged.

Workaround
Use the nvidia-smi command in the guest VM to enable or disable ECC memory for the vGPU as explained in Virtual GPU Software User Guide.

If the ECC memory state remains unchanged even after you use the nvidia-smi command to change it, use the workaround in Changes to ECC memory settings for a Linux vGPU VM by nvidia-smi might be ignored.

Status
Open

Ref. #
200523086

5.36. Changes to ECC memory settings for a Linux vGPU VM by nvidia-smi might be ignored

Description
After the ECC memory state for a Linux vGPU VM has been changed by using the nvidia-smi command and the VM has been rebooted, the ECC memory state might remain unchanged.
This issue occurs when multiple NVIDIA configuration files in the system cause the kernel module option for setting the ECC memory state RMGuestECCState in `/etc/modprobe.d/nvidia.conf` to be ignored.

When the `nvidia-smi` command is used to enable ECC memory, the file `/etc/modprobe.d/nvidia.conf` is created or updated to set the kernel module option `RMGuestECCState`. Another configuration file in `/etc/modprobe.d/` that contains the keyword `NVreg_RegistryDwordsPerDevice` might cause the kernel module option `RMGuestECCState` to be ignored.

**Workaround**

This workaround requires administrator privileges.

1. Move the entry containing the keyword `NVreg_RegistryDwordsPerDevice` from the other configuration file to `/etc/modprobe.d/nvidia.conf`.
2. Reboot the VM.

**Status**

Open

**Ref. #**

200505777

5.37. Incorrect GPU type shown for Quadro RTX 8000 GPUs in Citrix XenCenter

**Description**

On the **GPU** tab in **Citrix XenCenter**, Quadro RTX 8000 GPUs are incorrectly shown as `[QUADRO RTX 6000]`.

However, the virtual GPU types for a Quadro RTX 8000 GPU are listed correctly.

NVIDIA vGPU software does not support systems with a mixture of GPU types. Therefore, you can assume that any GPU in a system that you know to contain Quadro RTX 8000 GPUs that is shown as `[QUADRO RTX 6000]` is, in fact, a Quadro RTX 8000 GPU.

**Status**

Not an NVIDIA bug

**Ref. #**

200492607
5.38. **NVIDIA Notification Icon** prevents log off of Citrix Published Application user sessions

**Description**

By default on Windows Server operating systems, the **NVIDIA Notification Icon** application is started with every Citrix Published Application user session. This application might prevent the Citrix Published Application user session from being logged off even after the user has quit all other applications.

**Resolution**

Disable the **NVIDIA Notification Icon** application for Citrix Published Application user sessions as explained in *Virtual GPU Software User Guide*.

**Status**

Resolved by the provision of Windows registry keys for disabling the **NVIDIA Notification Icon** application for Citrix Published Application user sessions.

**Ref. #**

2206368

5.39. **Host core CPU utilization is higher than expected for moderate workloads**

**Description**

When GPU performance is being monitored, host core CPU utilization is higher than expected for moderate workloads. For example, host CPU utilization when only a small number of VMs are running is as high as when several times as many VMs are running.

**Workaround**

Disable monitoring of the following GPU performance statistics:

- vGPU engine usage by applications across multiple vGPUs
Encoder session statistics
Frame buffer capture (FBC) session statistics
Statistics gathered by performance counters in guest VMs

Status
Open

Ref. #
2414897

5.40. Frame capture while the interactive logon message is displayed returns blank screen

Description
Because of a known limitation with NvFBC, a frame capture while the interactive logon message is displayed returns a blank screen.

An NvFBC session can capture screen updates that occur after the session is created. Before the logon message appears, there is no screen update after the message is shown and, therefore, a black screen is returned instead. If the NvFBC session is created after this update has occurred, NvFBC cannot get a frame to capture.

Workaround
See Black Screen at Logon with Nvidia HDX 3D Pro enabled in the Citrix knowledge base.

Status
Not a bug

Ref. #
2115733
5.41. RDS sessions do not use the GPU with some Microsoft Windows Server releases

Description

When some releases of Windows Server are used as a guest OS, Remote Desktop Services (RDS) sessions do not use the GPU. With these releases, the RDS sessions by default use the Microsoft Basic Render Driver instead of the GPU. This default setting enables 2D DirectX applications such as Microsoft Office to use software rendering, which can be more efficient than using the GPU for rendering. However, as a result, 3D applications that use DirectX are prevented from using the GPU.

Version

- Windows Server 2019
- Windows Server 2016
- Windows Server 2012

Solution

Change the local computer policy to use the hardware graphics adapter for all RDS sessions.

2. Set the Use the hardware default graphics adapter for all Remote Desktop Services sessions option.

5.42. When the scheduling policy is fixed share, GPU utilization is reported as higher than expected

Description

When the scheduling policy is fixed share, GPU engine utilization can be reported as higher than expected for a vGPU.
For example, GPU engine usage for six P40-4Q vGPUs on a Tesla P40 GPU might be reported as follows:

```
[root@localhost:] nvidia-smi vgpu
Mon Aug 20 10:33:18 2018
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 390.42                 Driver Version: 390.42                    |
|-------------------------------+--------------------------------+------------|
| GPU  Name                     | Bus-Id                         | GPU-Util   |
|      vGPU ID    Name          | VM ID    VM Name               | vGPU-Util  |
|===============================+================================+============|
|   0  Tesla P40                | 00000000:81:00.0               |  99%       |
|   85109      GRID P40-4Q   | 85110    win7-xmpl-146048-1    |     32%    |
|   87195      GRID P40-4Q   | 87196    win7-xmpl-146048-2    |     39%    |
|   88095      GRID P40-4Q   | 88096    win7-xmpl-146048-3    |     26%    |
|   90475      GRID P40-4Q   | 90476    win7-xmpl-146048-4    |      0%    |
|   93363      GRID P40-4Q   | 93364    win7-xmpl-146048-5    |      0%    |
|   1  Tesla P40                | 00000000:85:00.0               |   0%       |
+-------------------------------+--------------------------------+------------+

```

The vGPU utilization of vGPU 85109 is reported as 32%. For vGPU 87195, vGPU utilization is reported as 39%. And for 88095, it is reported as 26%. However, the expected vGPU utilization of any vGPU should not exceed approximately 16.7%.

This behavior is a result of the mechanism that is used to measure GPU engine utilization.

**Status**

Open

**Ref. #**

2227591

5.43. `nvidia-smi` reports that vGPU migration is supported on all hypervisors

**Description**

The command `nvidia-smi vgpu -m` shows that vGPU migration is supported on all hypervisors, even hypervisors or hypervisor versions that do not support vGPU migration.

**Status**

Closed

**Ref. #**

200407230
5.44. **NVIDIA Control Panel** Crashes in a VM connected to two 4K displays

**Description**

When **NVIDIA Control Panel** is started in a VM connected through a pass-through GPU to two 4K displays, a **Citrix HDX 3D Pro Warning** pop-up window opens containing the following error message:

> Full screen applications are not supported on Citrix HDX 3D Pro. Press Ok to exit

**NVIDIA Control Panel** then crashes.

After the VM is rebooted, **NVIDIA Control Panel** is available again.

**Version**

Citrix Virtual Apps and Desktops 7.16

**Status**

Not an NVIDIA bug

**Ref. #**

200393766

5.45. **vGPU guest VM driver not properly loaded on servers with more than 512 GB or 1 TB or more of system memory**

**Description**

If PV IOMMU is enabled, support for vGPU is limited to servers with a maximum of 512 GB of system memory. On servers with more than 512 GB of system memory and PV IOMMU enabled, the guest VM driver is not properly loaded. **Device Manager** marks the vGPU with a yellow exclamation point.

If PV IOMMU is disabled, support for vGPU is limited to servers with less than 1 TB of system memory. This limitation applies only to systems with supported GPUs based on the Maxwell architecture: Tesla M6, Tesla M10, and Tesla M60. On servers with 1 TB or more of system memory.
memory, VMs configured with vGPU fail to power on. However, support for GPU pass through is not affected by this limitation.

**Resolution**

1. If PV IOMMU is enabled, disable it.
   
   ```
   [root@xenserver ~]# /opt/xensource/libexec/xen-cmdline --set-xen iommu=dom0-passthrough
   ```
2. If the server has 1 TB or more of system memory, limit the amount of system memory on the server to 1 TB minus 16 GB.
   
   If the server has less than 1 TB or more of system memory, omit this step.
3. Reboot the server.

**Status**

Not an NVIDIA bug

**Ref. #**

1799582

**5.46. Luxmark causes a segmentation fault on an unlicensed Linux client**

**Description**

If the Luxmark application is run on a Linux guest VM configured with NVIDIA vGPU that is booted without acquiring a license, a segmentation fault occurs and the application core dumps. The fault occurs when the application cannot allocate a CUDA object on NVIDIA vGPUs where CUDA is disabled. On NVIDIA vGPUs that can support CUDA, CUDA is disabled in unlicensed mode.

**Status**

Not an NVIDIA bug.

**Ref. #**

200330956
5.47. **A segmentation fault in DBus code causes `nvidia-gridd` to exit on Red Hat Enterprise Linux and CentOS**

**Description**

On Red Hat Enterprise Linux 6.8 and 6.9, and CentOS 6.8 and 6.9, a segmentation fault in DBus code causes the `nvidia-gridd` service to exit.

The `nvidia-gridd` service uses DBus for communication with **NVIDIA X Server Settings** to display licensing information through the **Manage License** page. Disabling the GUI for licensing resolves this issue.

To prevent this issue, the GUI for licensing is disabled by default. You might encounter this issue if you have enabled the GUI for licensing and are using Red Hat Enterprise Linux 6.8 or 6.9, or CentOS 6.8 and 6.9.

**Version**

Red Hat Enterprise Linux 6.8 and 6.9

CentOS 6.8 and 6.9

**Status**

Open

**Ref. #**

- 200358191
- 200319854
- 1895945

5.48. **No Manage License option available in NVIDIA X Server Settings by default**

**Description**

By default, the **Manage License** option is not available in **NVIDIA X Server Settings**. This option is missing because the GUI for licensing on Linux is disabled by default to work around the...
issue that is described in A segmentation fault in DBus code causes nvidia-gridd to exit on Red Hat Enterprise Linux and CentOS.

**Workaround**

This workaround requires `sudo` privileges.

```
[Note: Do not use this workaround with Red Hat Enterprise Linux 6.8 and 6.9 or CentOS 6.8 and 6.9. To prevent a segmentation fault in DBus code from causing the nvidia-gridd service from exiting, the GUI for licensing must be disabled with these OS versions.]
```

If you are licensing a physical GPU for vCS, you **must** use the configuration file `/etc/nvidia/gridd.conf`.

1. If **NVIDIA X Server Settings** is running, shut it down.
2. If the `/etc/nvidia/gridd.conf` file does not already exist, create it by copying the supplied template file `/etc/nvidia/gridd.conf.template`.
3. As root, edit the `/etc/nvidia/gridd.conf` file to set the `EnableUI` option to `TRUE`.
4. Start the `nvidia-gridd` service.

```
# sudo service nvidia-gridd start
```

When **NVIDIA X Server Settings** is restarted, the **Manage License** option is now available.

**Status**

Open

5.49. **Licenses remain checked out when VMs are forcibly powered off**

**Description**

NVIDIA vGPU software licenses remain checked out on the license server when non-persistent VMs are forcibly powered off.

The NVIDIA service running in a VM returns checked out licenses when the VM is shut down. In environments where non-persistent licensed VMs are not cleanly shut down, licenses on the license server can become exhausted. For example, this issue can occur in automated test environments where VMs are frequently changing and are not guaranteed to be cleanly shut down. The licenses from such VMs remain checked out against their MAC address for seven days before they time out and become available to other VMs.
Resolution

If VMs are routinely being powered off without clean shutdown in your environment, you can avoid this issue by shortening the license borrow period. To shorten the license borrow period, set the `LicenseInterval` configuration setting in your VM image. For details, refer to Virtual GPU Client Licensing User Guide.

Status

Closed

Ref. #

1694975

5.50. 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 Virtual Apps and Desktops 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 Citrix Hypervisor `/var/log/messages` file.

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
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.
- 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.

For more information, see also Windows 10 Optimization for XenDesktop on the Citrix blog.

Status

Open

Ref. #

- 200130864
- 1803861

5.51. VM bug checks after the guest VM driver for Windows 10 RS2 is installed

Description

When the VM is rebooted after the guest VM driver for Windows 10 RS2 is installed, the VM bug checks. When Windows boots, it selects one of the standard supported video modes. If Windows is booted directly with a display that is driven by an NVIDIA driver, for example a vGPU on Citrix Hypervisor, a blue screen crash occurs.

This issue occurs when the screen resolution is switched from VGA mode to a resolution that is higher than 1920×1200.
Fix
Download and install Microsoft Windows Update KB4020102 from the Microsoft Update Catalog.

Workaround
If you have applied the fix, ignore this workaround.
Otherwise, you can work around this issue until you are able to apply the fix by not using resolutions higher than 1920×1200.
1. Choose a GPU profile in Citrix XenCenter that does not allow resolutions higher than 1920×1200.
2. Before rebooting the VM, set the display resolution to 1920×1200 or lower.

Status
Not an NVIDIA bug

Ref. #
200310861

5.52. On Citrix Hypervisor 7.0, VMs unexpectedly reboot and Citrix Hypervisor crashes or freezes

Description
On Citrix Hypervisor 7.0, VMs to which a vGPU is attached unexpectedly reboot and Citrix Hypervisor crashes or freezes.

The event log in the Citrix Hypervisor /var/log/crash/xen.log file lists the following errors:

‣ A fatal bus error on a component at the slot where the GPU card is installed
‣ A fatal error on a component at bus 0, device 2, function 0

This issue occurs when page-modification logging (PML) is enabled on Intel Broadwell CPUs running Citrix Hypervisor 7.0. Citrix is aware of this issue and is working on a permanent fix.

Workaround
Disable page-modification logging (PML) as explained in XenServer 7 host crash while starting multiple virtual machines in the Citrix Support Knowledge Center.
5.53. With no NVIDIA driver installed, Citrix Hypervisor misidentifies Tesla M10 cards

Description
An erroneous entry in the `pci.ids` database causes Citrix Hypervisor to identify Tesla M10 cards as GRID M40 when no NVIDIA driver is installed.

Version
Citrix Hypervisor 6.5 and 7.0

Workaround
None

Status
Not an NVIDIA bug

Ref. #
1853248

5.54. 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!
Known Issues

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

5.55. Video goes blank when run in loop in Windows Media Player

Description
When connected to a vGPU-enabled VM using Citrix Virtual Apps and Desktops, a video played back in looping mode on Windows Media Player goes blank or freezes after a few iterations.

Workaround
None

Status
Not an NVIDIA bug

Ref. #
1306623
5.56. Local VGA console is momentarily unblanked when Citrix Virtual Apps and Desktops changes resolution of the VM desktop

Description
When Citrix Virtual Apps and Desktops establishes a remote connection to a VM using vGPU, the VM’s local VGA console display in XenCenter is blanked (assuming the VM local console has not been disabled by setting `disable_vnc=1` in the vGPU plugin’s parameters). If the Citrix Virtual Apps and Desktops session changes resolution of the VM’s desktop, the local VGA console momentarily unblanks, allowing a XenCenter user to briefly view the desktop.

Workaround
Disable the VM’s local VGA console.

- **Citrix Hypervisor 8.1 or later:**
  
  ```bash
  [root@xenserver ~]# xe vgpu-param-set uuid=vgpu-uuid extra_args=disable_vnc=1
  ```

- **Citrix Hypervisor earlier than 8.1:**
  
  ```bash
  xe vm-param-set uuid=vm-uuid platform:vgpu_extra_args="disable_vnc=1"
  ```

Status
Open

Ref. #
NVIDIA-145/1375164

5.57. VM bugchecks on shutdown/restart when Citrix Virtual Apps and Desktops is installed and NVIDIA driver is uninstalled or upgraded.

Description
If the Citrix Virtual Apps and Desktops agent is installed in a VM before any NVIDIA GPU driver is installed, the VM will bugcheck [bluescreen] when the NVIDIA
driver is subsequently upgraded or uninstalled. The bugcheck code is 0x7E, SYSTEM_THREAD_EXCEPTION_NOT_HANDLED.

Workaround

Use one of the following workarounds:

‣ Do a force shutdown of the VM and restart it.
‣ Install the NVIDIA driver in guest VMs before installing Citrix Virtual Apps and Desktops.

Status

Open

Ref. #

NVIDIA-295/200018125

5.58. Application frame rate may drop when running Citrix Virtual Apps and Desktops at 2560×1600 resolution.

Description

An application’s rendering frame rate may drop when running Citrix Virtual Apps and Desktops at 2560×1600 resolution, relative to the frame rate obtained at lower resolutions.

Fix

Using the Windows regedit utility within the VM, open the HKLM\SOFTWARE\Citrix\Graphics registry key and create a new DWORD value, EncodeSpeed, with a value of 2. Reboot the VM. This setting may improve the delivered frame rate at the expense of a reduction in image quality.

Status

Open

Ref. #

NVIDIA-190/1416336
5.59.  Windows VM BSOD

Description

Windows VM bugchecks on Citrix Hypervisor when running a large number of vGPU based VMs.

The Citrix Hypervisor /var/log/messages file contains these error messages:

```
NVRM: Xid (PCI:0000:08:00): 31, Ch 0000001e, engmask 00000111, intr 10000000
NVRM: Xid (PCI:0000:08:00): 31, Ch 00000016, engmask 00000111, intr 10000000

... vmiop_log: error: Assertion Failed at 0xb5b898d8:4184
vmiop_log: error: 8 frames returned by backtrace
vmiop_log: error: /usr/lib/libnvidia-vgx.so(_nv000793vgx+0x69d) [0xb5b8064d]
vmiop_log: error: /usr/lib/libnvidia-vgx.so(_nv000479vgx+0x118) [0xb5b898d8]
vmiop_log: error: /usr/lib/libnvidia-vgx.so(_nv000782vgx+0x59) [0xb5b85f49]
vmiop_log: error: /usr/lib/libnvidia-vgx.so(_nv000347vgx+0x3db) [0xb5b932db]
vmiop_log: error: /usr/lib/libnvidia-vgx.so [0xb5b78e4a]
vmiop_log: error: /usr/lib/xen/bin/vgpu [0x80554be]
vmiop_log: error: /lib/libpthread.so.0(clone+0x5e) [0xb76fc5ee]
vmiop_log: error: failed to initialize guest PTE entries
vmiop_log: error: failed to fill up guest PTE entries 3
vmiop_log: error: VGPU message 27 failed, result code: 0xff000003
vmiop_log: error: 0xc1d00001, 0xff010000, 0x1a77ba000, 0x0, 0x1,
vmiop_log: error: 0x1, 0x1000, 0x10202, 0xc1d00001, 0xff000003,
vmiop_log: error: 0xc1d00001, 0x0
vmiop_log: error: Timeout occurred, reset initiated.
```

Version

Citrix Hypervisor 6.2

Fix

Ensure that you are running the latest OEM firmware for your NVIDIA vGPU software boards.

Status

Closed

Ref. #

NVIDIA-327/1632120
5.60. Windows VM BSOD when upgrading NVIDIA drivers over a Citrix Virtual Apps and Desktops session

Description

Windows VM bugchecks when NVIDIA guest drivers are upgraded over a Citrix Virtual Apps and Desktops session.

If the VM is restarted after the bugcheck, the upgraded driver loads correctly and full functionality is available.

Fix

Upgrade Citrix Virtual Apps and Desktops to 7.6 Feature Pack 3

Status

Closed

Ref. #

NVIDIA-370/200130780

5.61. XenCenter does not allow vGPUs to be selected as a GPU type for Linux VMs

Description

When creating a new Linux VM or editing the properties of an existing Linux VM, XenCenter does not allow vGPUs to be selected as a GPU type.

vGPU on Linux VMs is supported as a technical preview on Citrix Hypervisor 6.5, and does include XenCenter integration.

Version

Affects the XenCenter integration with Citrix Hypervisor 6.5 only.

Resolved in the XenCenter integration with Citrix Hypervisor 7.0.
Workaround

Refer to XenServer vGPU Management in Virtual GPU Software User Guide for how to configure vGPU by using the xe CLI.

Status

Closed

Ref. #

NVIDIA-360

5.62. If X server is killed on a RHEL7 VM running vGPU, XenCenter console may not automatically switch to text console

Description

If X server is killed on a RHEL7 VM running vGPU, XenCenter console may display a corrupted image and fail to switch over to text console.

The failure to switch over to text console is due to a bug in RHEL7, which causes X server to not start correctly under certain configurations.

Workaround

Use CTRL+ALT+F1, F2, or F3 to switch between Linux terminals.

Status

Closed

Ref. #

NVIDIA-350/200123378
5.63. Citrix Virtual Apps and Desktops shows only a black screen when connected to a vGPU VM

Description
Citrix Virtual Apps and Desktops sometimes displays only a black screen when it is connected to an NVIDIA vGPU VM. The probable cause is that the display that is connected to the NVIDIA vGPU is entering a lower power state.

Fix
Disable all display-related power management settings.
For detailed instructions, visit Microsoft power plans frequently asked questions and from the list, select your OS version.

Status
Not an NVIDIA bug

Ref. #
1719877
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

© 2013-2022 NVIDIA Corporation & affiliates. All rights reserved.