

Virtual GPU Software R418 for Nutanix AHV

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 8.10	2
1.4. Updates in Release 8.9	3
1.5. Updates in Release 8.8	3
1.6. Updates in Release 8.7	3
1.7. Updates in Release 8.6	3
1.8. Updates in Release 8.5	4
1.9. Updates in Release 8.4	4
1.10. Updates in Release 8.3	4
1.11. Updates in Release 8.2	5
1.12. Updates in Release 8.1	5
1.13. Updates in Release 8.0	5
Chapter 2. Validated Platforms	7
2.1. Supported NVIDIA GPUs and Validated Server Platforms	7
2.2. Hypervisor Software Releases	8
2.3. Guest OS Support	9
2.3.1. Windows Guest OS Support	9
2.3.2. Linux Guest OS Support	10
2.4. NVIDIA CUDA Toolkit Version Support	12
Chapter 3. Known Product Limitations	13
3.1. Issues occur when the channels allocated to a vGPU are exhausted	13
3.2. Total frame buffer for vGPUs is less than the total frame buffer on the physical GPU	14
3.3. Issues may occur with graphics-intensive OpenCL applications on vGPU types with	
limited frame buffer	14
3.4. vGPU profiles with 512 Mbytes or less of frame buffer support only 1 virtual display head	
on Windows 10	
3.5. NVENC requires at least 1 Gbyte of frame buffer	
3.6. VM running older NVIDIA vGPU drivers fails to initialize vGPU when booted	
3.7. Virtual GPU fails to start if ECC is enabled	
3.8. Single vGPU benchmark scores are lower than pass-through GPU	
3.9. nvidia-smi fails to operate when all GPUs are assigned to GPU pass-through mode	
Chapter 4. Resolved Issues	20
Chapter 5. Security Updates	23

5.1. Restricting Access to GPU Performance Counters	23
5.1.1. Windows: Restricting Access to GPU Performance Counters for One User by Usir NVIDIA Control Panel	9
5.1.2. Windows: Restricting Access to GPU Performance Counters Across an Enterprise by Using a Registry Key	
5.1.3. Linux Guest VMs and Hypervisor Host: Restricting Access to GPU Performance Counters	
Chapter 6. Known Issues	26
6.1. Since 8.9: NVENC does not work with Teradici Cloud Access Software on Windows	
6.2. A licensed client might fail to acquire a license if a proxy is set	
6.3. 8.0-8.6 Only: NVIDIA vGPU software graphics driver installation fails in Ubuntu gue	st
6.4. 8.0-8.5 Only: Desktop sessions disconnect after the server becomes unresponsive	
6.5. 8.0-8.4 Only: Application responsiveness degrades over time	28
6.6. On systems with more than 1 TiB of system memory, application performance degrade over time	
6.7. 8.0-8.3 Only: When the VMs to which 16 vGPUs on a single GPU are assigned are started simultaneously, one VM fails to boot	
6.8. 8.0, 8.1 Only: VMs fail to boot with failed assertions	30
6.9. NVIDIA Control Panel fails to start if launched too soon from a VM without licensing information	
6.10. DWM crashes randomly occur in Windows VMs	32
6.11. Citrix Virtual Apps and Desktops session freezes when the desktop is unlocked	32
6.12. NVIDIA vGPU software graphics driver fails after Linux kernel upgrade with DKM enabled	
6.13. 8.0 Only: Incorrect NVIDIA vGPU software Windows graphics driver version in the installer	
6.14. Vulkan applications crash in Windows 7 guest VMs configured with NVIDIA vGPU	35
6.15. Host core CPU utilization is higher than expected for moderate workloads	35
6.16. Frame capture while the interactive logon message is displayed returns blank scree	e.n. 36
6.17. RDS sessions do not use the GPU with some Microsoft Windows Server releases	37
6.18. Cloned VMs configured with a vGPU type different than the type in the master image fail to start	
6.19. 8.0-8.2 Only: Even when the scheduling policy is equal share, unequal GPU utilization is reported	
6.20. When the scheduling policy is fixed share, GPU utilization is reported as higher the expected	
6.21. Benign warnings during Virtual GPU Manager installation or uninstallation after hypervisor upgrade	

6.22. Benign not in gzip format messages during Virtual GPU Manager installation or	
uninstallation	. 41
6.23. License is not acquired in Windows VMs	. 41
6.24. nvidia-smi reports that vGPU migration is supported on all hypervisors	. 42
6.25. Resolution is not updated after a VM acquires a license and is restarted	.42
6.26. A segmentation fault in DBus code causes nvidia-gridd to exit on Red Hat Enterprise Linux and CentOS	43
6.27. No Manage License option available in NVIDIA X Server Settings by default	. 44
6.28. Licenses remain checked out when VMs are forcibly powered off	. 44
6.29. Memory exhaustion can occur with vGPU profiles that have 512 Mbytes or less of frame buffer	. 45
6.30. GNOME Display Manager (GDM) fails to start on Red Hat Enterprise Linux 7.2 and CentOS 7.0	. 46

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 Nutanix AHV.



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.

NVIDIA vGPU Software Version	NVIDIA Virtual GPU Manager Version	NVIDIA Windows Driver Version	NVIDIA Linux Driver Version
8.10	418.240	427.71	418.240
8.9	418.226.00	427.60	418.226.00
8.8	418.213	427.48	418.211.00
8.7	418.196	427.33	418.197.02
8.6	418.181	427.11	418.181.07
8.5	418.165.01	426.94	418.165.01
8.4	418.149	426.72	418.149
8.3	418.130	426.52	418.130
8.2	418.109	426.26	418.109
8.1	418.92	426.04	418.92
8.0	418.66	425.31	418.70

For details of which Nutanix AHV 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 older NVIDIA vGPU drivers fails to initialize vGPU when booted.

Different versions of the vGPU Manager and quest VM driver from within the same main release branch can be used together. For example, you can use the vGPU Manager from release 8.1 with quest VM drivers from release 8.0. However, versions of the vGPU Manager and quest VM driver from different main release branches cannot be used together. For example, you cannot use the vGPU Manager from release 8.1 with guest VM drivers from release 7.2.



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.

1.3. Updates in Release 8.10

New Features in Release 8.10

- 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 8.10

Support for Red Hat Enterprise Linux 8.5 as a guest OS

Feature Support Withdrawn in Release 8.10

- 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.

Updates in Release 8.9 1.4.

New Features in Release 8.9

- ▶ Security updates see Security Bulletin: NVIDIA GPU Display Driver October 2021, which is posted shortly after the release date of this software and is listed on the NVIDIA Product Security page
- Miscellaneous bug fixes

Updates in Release 8.8

New Features in Release 8.8

- Security updates see Security Bulletin: NVIDIA GPU Display Driver July 2021
- Miscellaneous bug fixes

Hardware and Software Support Introduced in Release 8.8

Support for Red Hat Enterprise Linux 8.4 as a guest OS

1.6. Updates in Release 8.7

New Features in Release 8.7

- Security updates see <u>Security Bulletin: NVIDIA GPU Display Driver April 2021</u>
- Miscellaneous bug fixes

Updates in Release 8.6

New Features in Release 8.6

- Security updates see <u>Security Bulletin: NVIDIA GPU Display Driver January 2021</u>
- Miscellaneous bug fixes

Updates in Release 8.5 1.8.

New Features in Release 8.5

- Security updates- see Security Bulletin: NVIDIA GPU Display Driver September 2020
- Miscellaneous bug fixes

Feature Support Withdrawn in Release 8.5

Nutanix AHV 5.5 is no longer supported.

Updates in Release 8.4

New Features in Release 8.4

- Miscellaneous bug fixes
- Security updates see <u>Security Bulletin: NVIDIA GPU Display Driver June 2020</u>

Hardware and Software Support Introduced in Release 8.4

- Support for Nutanix AHV 5.15
- Support for the following guest OS releases
 - Red Hat Enterprise Linux 8.2 and 7.8
 - CentOS 7 8

Feature Support Withdrawn in Release 8.4

- Nutanix AHV 5.9 and 5.8 are no longer supported
- The following OS releases are no longer supported as a guest OS:
 - Red Hat Enterprise Linux 7.5
 - CentOS 7.5

Updates in Release 8.3

New Features in Release 8.3

- Miscellaneous bug fixes
- Security updates see <u>Security Bulletin: NVIDIA GPU Display Driver February 2020</u>

Hardware and Software Support Introduced in Release 8.3

- Support for the following guest OS releases
 - Red Hat Enterprise Linux 8.1
 - CentOS Linux 8 (1911)

Feature Support Withdrawn in Release 8.3

- ► The following OS releases are no longer supported as a guest OS:
 - Windows Server 2008 R2
 - Red Hat Enterprise Linux 7.0-7.4
 - CentOS 7.0-7.4

1.11. Updates in Release 8.2

New Features in Release 8.2

- Miscellaneous bug fixes
- Security updates see <u>Security Updates</u>

1.12. Updates in Release 8.1

New Features in Release 8.1

- Security updates
- Miscellaneous bug fixes

Hardware and Software Support Introduced in Release 8.1

Support for Red Hat Enterprise Linux 7.7 as a guest OS

1.13. Updates in Release 8.0

New Features in Release 8.0

- ► New -1B4 virtual GPU types
- Security updates
- Miscellaneous bug fixes

Hardware and Software Support Introduced in Release 8.0

- Support for the Tesla T4 GPU
- Support for the following OS releases as a guest OS:
 - Windows 10 October 2018 Update (1809)
 - ► Windows Server 2019
- ► Support for Ubuntu 18.04 as a guest OS

Chapter 2. Validated Platforms

This release family of NVIDIA vGPU software provides support for several NVIDIA GPUs on validated server hardware platforms, Nutanix AHV hypervisor software versions, and quest operating systems. It also supports the version of NVIDIA CUDA Toolkit that is compatible with R418 drivers.

2.1. Supported NVIDIA GPUs and Validated Server Platforms

This release of NVIDIA vGPU software provides support for the following NVIDIA GPUs on Nutanix AHV, running on validated Nutanix NX series server hardware platforms:

- ▶ GPUs based on the NVIDIA Maxwell $^{\mathsf{TM}}$ graphic architecture:
 - Tesla M10
 - ► Tesla M60
- GPUs based on the NVIDIA Pascal[™] architecture:
 - ► Tesla P4
 - ► Tesla P40
- GPUs based on the NVIDIA Volta architecture:
 - ► Tesla V100 PCIe
 - ► Tesla V100 PCIe 32GB
- ► GPUs based on the NVIDIA Turing architecture:
 - Tesla T4

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



Note:

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 <code>gpumodeswitch</code> tool provided with NVIDIA vGPU software releases.

Hypervisor Software Releases

This release is supported on the Nutanix AHV releases listed in the table.



Note: Updates to a base release of Nutanix AHV are compatible with the base release and can also be used with this version of NVIDIA vGPU software unless expressly stated otherwise.

Software	Releases Supported	Notes
Since 8.4: Nutanix AOS Hypervisor (AHV) 5.15	5.15 RTM and compatible 5.15. <i>x</i> updates	
Nutanix AOS Hypervisor (AHV) 5.10	5.10 RTM and compatible 5.10.x updates	Support for the following GPUs is introduced in release 5.10.1:
		▶ Tesla P4
		► Tesla V100 PCIe 32GB
		5.10 RTM does not support Tesla P4 or Tesla V100 PCIe 32GB.
8.0-8.3 only: Nutanix AOS Hypervisor (AHV) 5.9	5.9 RTM and compatible 5.9.x updates	Not supported on the following GPUs:
		▶ Tesla P4
		► Tesla V100 PCIe 32GB
8.0-8.3 only: Nutanix AOS Hypervisor (AHV) 5.8	5.8 RTM and compatible 5.8.x updates	Not supported on the following GPUs:
		► Tesla P4
		► Tesla V100 PCIe 32GB
8.0-8.4 only: Nutanix AOS Hypervisor (AHV) 5.5	5.5 RTM and compatible 5.5.x updates	Not supported on the following GPUs:
		▶ Tesla P4
		► Tesla V100 PCIe 32GB

2.3. **Guest OS Support**

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



Note:

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

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

Windows Guest OS Support 2.3.1.

NVIDIA vGPU software supports **only** the 64-bit Windows releases listed in the table as a guest OS on Nutanix AHV. The releases of Nutanix AHV 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.

Guest OS	NVIDIA vGPU - Nutanix AHV Releases	Pass-Through GPU - Nutanix AHV Releases
Windows Server 2019	Since 8.4: 5.15, 5.10	Since 8.4: 5.15, 5.10
	8.0-8.3 only: 5.10	8.0-8.3 only: 5.10
Windows Server 2016 1709, 1607	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5
Windows Server 2012 R2	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5
8.0-8.2 only: Windows Server 2008 R2	5.10, 5.9, 5.8, 5.5	5.10, 5.9, 5.8, 5.5
Windows 10 October 2018 Update (1809) and all	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
Windows 10 releases supported by Microsoft up to and including this release	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5

Guest OS	NVIDIA vGPU - Nutanix AHV Releases	Pass-Through GPU - Nutanix AHV Releases
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5
Windows 8.1 Update	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5
Windows 8.1	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5
Windows 8	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5
Windows 7	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5

2.3.2. Linux Guest OS Support

NVIDIA vGPU software supports only the Linux distributions listed in the table as a guest OS on Nutanix AHV. The releases of Nutanix AHV 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.

Guest OS	NVIDIA vGPU - Nutanix AHV Releases	Pass-Through GPU - Nutanix AHV Releases
Since 8.10: Red Hat Enterprise Linux 8.5	5.15, 5.10	5.15, 5.10
Since 8.8: Red Hat Enterprise Linux 8.4	5.15, 5.10	5.15, 5.10
Since 8.4: Red Hat Enterprise Linux 8.2	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
8.3-8.9 only: Red Hat Enterprise Linux 8.1	8.5-8.9 only: 5.15, 5.10	8.5-8.9 only: 5.15, 5.10

Guest OS	NVIDIA vGPU - Nutanix AHV Releases	Pass-Through GPU - Nutanix AHV Releases
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.3 only: 5.10, 5.9, 5.8, 5.5	8.3 only: 5.10, 5.9, 5.8, 5.5
Since 8.3: CentOS Linux 8 (1911)	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.3 only: 5.10, 5.9, 5.8, 5.5	8.3 only: 5.10, 5.9, 5.8, 5.5
8.6-8.9 only: Red Hat Enterprise Linux 7.7-7.8	5.15, 5.10	5.15, 5.10
8.4, 8.5 only: Red Hat Enterprise Linux 7.6-7.8	8.5 only: 5.15, 5.10	8.5 only: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
8.3 only: Red Hat Enterprise Linux 7.5-7.7	5.10, 5.9, 5.8, 5.5	5.10, 5.9, 5.8, 5.5
8.1, 8.2 only: Red Hat Enterprise Linux 7.0-7.7	5.10, 5.9, 5.8, 5.5	5.10, 5.9, 5.8, 5.5
8.0 only: Red Hat Enterprise Linux 7.0-7.6	5.10, 5.9, 5.8, 5.5	5.10, 5.9, 5.8, 5.5
Since 8.6: CentOS 7.7-7.8	5.15, 5.10	5.15, 5.10
8.4, 8.5 only: CentOS 7.6-7.8	8.5 only: 5.15, 5.10	8.5 only: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
8.3 only: CentOS 7.5, 7.6	5.10, 5.9, 5.8, 5.5	5.10, 5.9, 5.8, 5.5
8.0-8.2 only: CentOS 7.0-7.6	5.10, 5.9, 5.8, 5.5	5.10, 5.9, 5.8, 5.5
Ubuntu 18.04 LTS	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5
Ubuntu 16.04 LTS	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5
Ubuntu 14.04 LTS	Since 8.5: 5.15, 5.10	Since 8.5: 5.15, 5.10
	8.4 only: 5.15, 5.10, 5.5	8.4 only: 5.15, 5.10, 5.5
	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5	8.0-8.3 only: 5.10, 5.9, 5.8, 5.5

NVIDIA CUDA Toolkit Version Support 2.4.

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

For more information about NVIDIA CUDA Toolkit, see CUDA Toolkit 10.1 Documentation. This documentation applies to the base NVIDIA CUDA Toolkit 10.1 release and updates to the base release.



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.

Chapter 3. Known Product Limitations

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

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):
Oxcld004a1, Oxff0e0000, Oxff0400fb, Oxc36f,
Jun 26 08:01:25 srvxen06f vgpu-3[14276]: error: vmiop_log: (0x0):
Oxff1fe314, Oxff1fe038, Ox100b6f000, Ox1000,
                                                                                     0x1,
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.

3.2. 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 quest OS is reserved on the basis of the runtime needs of applications.

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

max-reserved-fb = vgpu-profile-size-in-mb÷16 + 16

max-reserved-fb

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

vqpu-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.



Note: In VMs running a Windows guest OS that supports Windows Display Driver Model (WDDM) 1.x, namely, Windows 7, Windows 8.1, Windows Server 2008, and Windows Server 2012, an additional 48 Mbytes of frame buffer are reserved and not available for vGPUs.

3.3. Issues may occur with graphicsintensive 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.

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.

3.4. 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.

3.5. 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.

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

Description

A VM running a version of the NVIDIA guest VM drivers from a previous main release branch, for example release 4.4, will fail to initialize vGPU when booted on a Nutanix AHV platform running the current release of Virtual GPU Manager.

In this scenario, 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:

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

Depending on the versions of drivers in use, the Nutanix AHV VM's /var/log/messages log file reports one of the following errors:

- An error message:
 - vmiop log: error: Unable to fetch Guest NVIDIA driver information
- A version mismatch between guest and host drivers:
 - vmiop log: error: Guest VGX version(1.1) and Host VGX version(1.2) do not match
- ► A signature mismatch:
 - vmiop log: error: VGPU message signature mismatch.

Resolution

Install the current NVIDIA guest VM driver in the VM.

3.7. Virtual GPU fails to start if ECC is enabled

Description

Tesla M60, Tesla M6, and GPUs based on the Pascal GPU architecture, for example Tesla P100 or Tesla P4, support error correcting code (ECC) memory for improved data integrity. Tesla

M60 and M6 GPUs in graphics mode are supplied with ECC memory disabled by default, but it may subsequently be enabled using nvidia-smi. GPUs based on the Pascal GPU architecture are supplied with ECC memory enabled.

However, NVIDIA vGPU does not support ECC memory. If ECC memory is enabled, NVIDIA vGPU fails to start.

The following error is logged in the Nutanix AHV host's /var/log/messages log file: vmiop log: error: Initialization: VGX not supported with ECC Enabled.

Resolution

Ensure that ECC is disabled on all GPUs.

Before you begin, ensure that NVIDIA Virtual GPU Manager is installed on your hypervisor.

1. Use nvidia-smi to list the status of all GPUs, and check for ECC noted as enabled on GPUs.

```
# nvidia-smi -q
========NVSMI LOG=========
Timestamp
                                : Tue Dec 19 18:36:45 2017
Driver Version
                                : 384.99
Attached GPUs
                               : 1
GPU 0000:02:00.0
[...]
   Ecc Mode
                               : Enabled
     Current
       Pending
                                : Enabled
[...]
```

- 2. Change the ECC status to off on each GPU for which ECC is enabled.
 - If you want to change the ECC status to off for all GPUs on your host machine, run this command:

```
# nvidia-smi -e 0
```

If you want to change the ECC status to off for a specific GPU, run this command:

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

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

This example disables ECC for the GPU with index 0000:02:00.0.

```
# nvidia-smi -i 0000:02:00.0 -e 0
```

- 3. Reboot the host.
- 4. Confirm that ECC is now disabled for the GPU.

```
# nvidia-smi -q
========NVSMI LOG=========
                            : Tue Dec 19 18:37:53 2017
Timestamp
Driver Version
                               : 384.99
```

```
Attached GPUs
                                     : 1
GPU 0000:02:00.0
    Ecc Mode
                                     : Disabled
        Current
        Pending
                                    : Disabled
[...]
```

If you later need to enable ECC on your GPUs, run one of the following commands:

If you want to change the ECC status to on for all GPUs on your host machine, run this command:

```
# nvidia-smi -e 1
```

If you want to change the ECC status to on for a specific GPU, run this command:

```
# nvidia-smi -i id -e 1
```

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

This example enables ECC for the GPU with index 0000:02:00.0.

```
# nvidia-smi -i 0000:02:00.0 -e 1
```

After changing the ECC status to on, reboot the host.

3.8. 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 setting the extra vGPU parameter frame rate limiter=0 through the ACLI when a vGPU type is assigned to a VM.

For example:

```
vm.gpu_assign a7f24fc0-dd05-4595-a76f-8b1901942aae gpu=Nvidia_GRID_M60-4Q
extra_param=frame_rate_limiter=0
```

The setting takes effect the next time any VM using the given vGPU type is started.

With this setting in place, the VM's vGPU will run without any frame rate limit.

The FRL can be reverted back to its default setting by setting the extra vGPU parameter frame rate limiter=1

3.9. nvidia-smi fails to operate when all GPUs are assigned to GPU passthrough mode

Description

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

```
[root@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 Nutanix AHV host.

Resolution

N/A

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 8.10

No resolved issues are reported in this release for Nutanix AHV.

Issues Resolved in Release 8.9

No resolved issues are reported in this release for Nutanix AHV.

Issues Resolved in Release 8.8

No resolved issues are reported in this release for Nutanix AHV.

Issues Resolved in Release 8.7

Bug ID	Summary and Description
3226853	8.0-8.6 Only: NVIDIA vGPU software graphics driver installation fails in Ubuntu guest VMs
	Installation of the NVIDIA vGPU software graphics driver from a .run file fails in Ubuntu guest VMs that are running Linux kernel version 5.8 or later.

Issues Resolved in Release 8.6

Bug ID	Summary and Description
2925629	8.0-8.5 Only: Desktop sessions disconnect after the server becomes unresponsive
	Desktop sessions disconnect after the server becomes unresponsive. Before the sessions disconnect, the error message VGPU message 32 failed, result code: 0x59 is written to the log files on the hypervisor host.

Issues Resolved in Release 8.5

Bug ID	Summary and Description
3051614	8.0-8.4 Only: Application responsiveness degrades over time
	Application responsiveness degrades over time, causing slow application performance and stutter when users switch between applications. This issue occurs because the GPU driver is not setting the Linux kernel PCI coherent_dma_mask for NVIDIA GPU devices. If the coherent_dma_mask is not set, IOMMU IOVA space is restricted to the default size of 32 bits for DMA allocations performed in the NVIDIA GPU device context. Furthermore, for hosts on which iommu=pt is set, the default coherent_dma_mask causes IOMMU mappings to always be created. When IOMMU mappings are always created, performance degradation can occur because all host to device
	accesses require translation by hardware IOMMU.

Issues Resolved in Release 8.4

Bug ID	Summary and Description
200594274	8.0-8.3 Only: When the VMs to which 16 vGPUs on a single GPU are assigned are started simultaneously, one VM fails to boot
	When the VMs to which 16 vGPUs on a single GPU (for example 16 T4-1Q vGPUs on a Tesla T4 GPU) are assigned are started simultaneously, only 15 VMs boot and the remaining VM fails to start. When the other VMs are shut down, the VM that failed to start boots successfully.

Issues Resolved in Release 8.3

Bug ID	Summary and Description
2175888	8.0-8.2 Only: Even when the scheduling policy is equal share, unequal GPU utilization is reported
	When the scheduling policy is equal share, unequal GPU engine utilization can be reported for the vGPUs on the same physical GPU.

Issues Resolved in Release 8.2

Bug ID	Summary and Description
2644858	8.0, 8.1 Only: VMs fail to boot with failed assertions
	In some scenarios with heavy workloads running on multiple VMs configured with NVIDIA vGPUs on a single pysical GPU, additional VMs configured with NVIDIA vGPU on the same GPU fail to boot. The failure of the VM to boot is

Bug ID	Summary and Description
	followed by failed assertions. This issue affects GPUs based on the NVIDIA
	Volta GPU architecture and later architectures.

Issues Resolved in Release 8.1

Bug ID	Summary and Description
200534988	Error XID 47 followed by multiple XID 32 errors
	After disconnecting Citrix Virtual Apps and Desktops and clicking the power button in the VM, error XID 47 occurs followed by multiple XID 32 errors. When these errors occur, the hypervisor host becomes unusable.
-	8.0 Only: Incorrect NVIDIA vGPU software Windows graphics driver version in the installer
	The NVIDIA vGPU software Windows graphics driver version in the installer is incorrect. The driver version incorrectly appears as 325.31 instead of 425.31 in the Extraction path field and the title of the NVIDIA Graphics Driver (325.31) Package Window.

Issues Resolved in Release 8.0

Bug ID	Summary and Description
1971698	NVIDIA vGPU encoder and process utilization counters don't work with Windows Performance Counters
	GPU encoder and process utilization counter groups are listed in Windows Performance Counters, but no instances of the counters are available. The counters are disabled by default and must be enabled.

Chapter 5. Security Updates

5.1. Restricting Access to GPU Performance Counters

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

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

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

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

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

Ensure that you are running NVIDIA Control Panel version 8.1.950.

- 1. Open NVIDIA Control Panel:
 - ▶ Right-click on the Windows desktop and select **NVIDIA Control Panel** from the menu.
 - Open Windows Control Panel and double-click the NVIDIA Control Panel icon.
- 2. In NVIDIA Control Panel, select the Manage GPU Performance Counters task in the **Developer** section of the navigation pane.

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

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

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

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



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

1. Set the RmProfilingAdminOnly Windows registry key to 1.

[HKLM\SYSTEM\CurrentControlSet\Services\nvlddmkm\Global\NVTweak]

Value: "RmProfilingAdminOnly"

Type: DWORD Data: 00000001

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

2. Restart the VM.

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

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

Perform this task from the guest VM to which the GPU is passed through or from your hypervisor host machine.

In Linux guest VMs, this task requires sudo privileges. On your hypervisor host machine, this task must be performed as the root user on the machine.

- 1. Log in to the guest VM or open a command shell on your hypervisor host machine.
- 2. Set the kernel module parameter NVreg RestrictProfilingToAdminUsers to 1 by adding this parameter to the /etc/modprobe.d/nvidia.conf file.

- ▶ If you are setting only this parameter, add an entry for it to the /etc/modprobe.d/ nvidia.conf file as follows:
 - options nvidia NVreg RegistryDwords="NVreg RestrictProfilingToAdminUsers=1"
- ▶ If you are setting multiple parameters, set them in a single entry as in the following example:

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

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

3. Restart the VM or reboot your hypervisor host machine.

Chapter 6. Known Issues

6.1. Since 8.9: 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.

Ref.

200749065

6.2. 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.

200704733

6.3. 8.0-8.6 Only: NVIDIA vGPU software graphics driver installation fails in Ubuntu guest VMs

Description

Installation of the NVIDIA vGPU software graphics driver from a .run file fails in Ubuntu quest VMs that are running Linux kernel version 5.8 or later.

Version

Ubuntu 18.04 LTS with Linux kernel version 5.8 or later

Workaround

Revert to a Linux kernel version earlier than 5.8.

Status

Resolved in NVIDIA vGPU software 8.7

Ref.

3226853

8.0-8.5 Only: Desktop sessions 6.4. disconnect after the server becomes unresponsive

Description

Desktop sessions disconnect after the server becomes unresponsive. Before the sessions disconnect, the error message VGPU message 32 failed, result code: 0x59 is written to the log files on the hypervisor host.

Status

Resolved in NVIDIA vGPU software 8.6

Ref.

2925629

6.5. 8.0-8.4 Only: Application responsiveness degrades over time

Description

Application responsiveness degrades over time, causing slow application performance and stutter when users switch between applications. This issue occurs because the GPU driver is not setting the Linux kernel PCI coherent dma mask for NVIDIA GPU devices. If the coherent dma mask is not set, IOMMU IOVA space is restricted to the default size of 32 bits for DMA allocations performed in the NVIDIA GPU device context. Furthermore, for hosts on which iommu=pt is set, the default coherent dma mask causes IOMMU mappings to always be created. When IOMMU mappings are always created, performance degradation can occur because all host to device accesses require translation by hardware IOMMU.

Status

Resolved in NVIDIA vGPU software 8.5



Note: On systems with more than 1 TiB of system memory and GPUs based on GPU architectures earlier than the NVIDIA Ampere architecture, a related issue might still cause application performance to degrade over time. For details, see On systems with more than 1 TiB of system memory, application performance degrades over time.

Ref.

3051614

On systems with more than 1 TiB 6.6. of system memory, application performance degrades over time

Description

On systems with more than 1 TiB of system memory, application performance degrades over time. As a result, application performance is slow and stutter occurs when users switch between applications. This issue occurs because the virtual GPU manager temporarily limits the dma mask and the coherent dma mask to 40 bits while the vGPU is being initialized. On systems with more than 1 TiB of system memory, the coherent dma mask addressing capability is less than the amount of system memory. As a result, IOMMU mappings are always created, which can cause performance degradation because all host to device accesses require translation by hardware IOMMU.

Workaround

Reduce the amount of system memory to 1 TiB or less.

Status

Open

Ref.

3063042

6.7. 8.0-8.3 Only: When the VMs to which 16 vGPUs on a single GPU are assigned are started simultaneously, one VM fails to boot

Description

When the VMs to which 16 vGPUs on a single GPU (for example 16 T4-1Q vGPUs on a Tesla T4 GPU) are assigned are started simultaneously, only 15 VMs boot and the remaining VM fails to start. When the other VMs are shut down, the VM that failed to start boots successfully.

The log file on the hypervisor host contains these error messages:

```
2020-03-04T16:01:13.626Z| vmx| E110: vmiop_log: NVOS status 0x51
2020-03-04T16:01:13.626Z| vmx| E110: vmiop_log: Assertion Failed at 0x5e530d8c:303
2020-03-04T16:01:13.626Z| vmx| E110: vmiop log: (0x0): Failed to alloc quest FB
memory
2020-03-04T16:01:13.626Z| vmx| E110: vmiop log: (0x0): init device instance failed
for inst 0 with error 2 (vmiop-display: error allocating framebuffer)
2020-03-04T16:01:13.626Z| vmx| E110: vmiop_log: (0x0): Initialization: init_device_instance failed error 2
2020-\overline{0}3-04T1\overline{6}:01:13.629Z \mid vmx \mid E110: vmiop log: display init failed for inst: 0
```

Status

Resolved in NVIDIA vGPU software 8.4

Ref.

200594274

6.8. 8.0, 8.1 Only: VMs fail to boot with failed assertions

Description

In some scenarios with heavy workloads running on multiple VMs configured with NVIDIA vGPUs on a single pysical GPU, additional VMs configured with NVIDIA vGPU on the same GPU fail to boot. The failure of the VM to boot is followed by failed assertions. This issue affects GPUs based on the NVIDIA Volta GPU architecture and later architectures.

When this error occurs, error messages similar to the following examples are logged to the Nutanix AHV log file:

```
nvidia-vgpu-mgr[31526]: error: vmiop_log: NVOS status 0x1e
```

```
nvidia-vgpu-mgr[31526]: error: vmiop log: Assertion Failed at 0xb2d3e4d7:96
nvidia-vgpu-mgr[31526]: error: vmiop_log: 12 frames returned by backtrace
nvidia-vgpu-mgr[31526]: error: vmiop_log: /usr/lib64/libnvidia-vgpu.so(_nv003956vgpu +0x18) [0x7f4bb2cfb338] vmiop_dump_stack nvidia-vgpu-mgr[31526]: error: vmiop_log: /usr/lib64/libnvidia-vgpu.so(_nv004018vgpu
+0xd4) [0x7f4bb2d09ce4] vmiopd alloc pb channel
nvidia-vgpu-mgr[31526]: error: vmiop_log: /usr/lib64/libnvidia-vgpu.so(_nv002878vgpu
+0x137) [0x7f4bb2d3e4d7] vgpufceInitCopyEngine_GK104
nvidia-vgpu-mgr[31526]: error: vmiop log: /usr/lib64/libnvidia-vgpu.so(+0x80e27)
[0x7f4bb2cd0e27]
nvidia-vgpu-mgr[31526]: error: vmiop log: /usr/lib64/libnvidia-vgpu.so(+0x816a7)
[0x7f4bb2cd16a7]
nvidia-vgpu-mgr[31526]: error: vmiop_log: vgpu() [0x413820]
nvidia-vgpu-mgr[31526]: error: vmiop_log: vgpu() [0x413a8d]
nvidia-vgpu-mgr[31526]: error: vmiop_log: vgpu() [0x40e11f]
nvidia-vgpu-mgr[31526]: error: vmiop_log: vgpu() [0x40bb69]
nvidia-vgpu-mgr[31526]: error: vmiop_log: vgpu() [0x40b51c]
nvidia-vgpu-mgr[31526]: error: vmiop log: /lib64/libc.so.6( libc start main+0x100)
nvidia-vgpu-mgr[31526]: error: vmiop log: vgpu() [0x4033ea]
nvidia-vgpu-mgr[31526]: error: vmiop log: (0x0): Alloc Channel (Gpfifo) for device
failed error: 0x1e
nvidia-vgpu-mgr[31526]: error: vmiop_log: (0x0): Failed to allocate FCE channel
nvidia-vgpu-mgr[31526]: error: vmiop_log: (0x0): init_device_instance failed for
 inst 0 with error 2 (init frame copy engine)
nvidia-vgpu-mgr[31526]: error: vmiop log: (0x0): Initialization:
init device instance failed error 2
nvidia-vgpu-mgr[31526]: error: vmiop_log: display_init failed for inst: 0
nvidia-vgpu-mgr[31526]: error: vmiop env log: (0 \times \overline{0}): vmiope process configuration:
 plugin registration error
nvidia-vgpu-mgr[31526]: error: vmiop env log: (0x0): vmiope process configuration
failed with 0x1a
kernel: [858113.083773] [nvidia-vgpu-vfio] ace3f3bb-17d8-4587-920e-199b8fed532d:
start failed. status: 0x1
```

Status

Resolved in NVIDIA vGPU software 8.2

Ref.

2644858

6.9. **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

Wait at least 30 seconds before trying to launch NVIDIA Control Panel.

Status

Open

Ref.

200623179

6.10. DWM crashes randomly occur in Windows VMs

Description

Desktop Windows Manager (DWM) crashes randomly occur in Windows VMs, causing a bluescreen 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

6.11. 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

6.12. 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 418.70 -k 5.3.0-28-
       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: 418.70 not found
       Error! Bad return status for module build on kernel: 5.3.0-28-generic
 (x86 64)
       Consult /var/lib/dkms/nvidia/ 418.70/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.

Ref.

2836271

6.13. 8.0 Only: Incorrect NVIDIA vGPU software Windows graphics driver version in the installer

Description

The NVIDIA vGPU software Windows graphics driver version in the installer is incorrect. The driver version incorrectly appears as 325.31 instead of 425.31 in the Extraction path field and the title of the NVIDIA Graphics Driver (325.31) Package Window.



Version

NVIDIA vGPU software Windows graphics driver version 425.31 in NVIDIA vGPU software release 8.0

Workaround

To simplify future administration of your system, you can correct the driver version in the folder name when the installer prompts you for the extraction path. However, if you do not change the name of the folder in the extraction path, the installation succeeds and the driver functions correctly.

Resolved in NVIDIA vGPU software release 8.1

6.14. Vulkan applications crash in Windows 7 guest VMs configured with NVIDIA **vGPU**

Description

In Windows 7 quest VMs configured with NVIDIA vGPU, applications developed with Vulkan APIs crash or throw errors when they are launched. Vulkan APIs require sparse texture support, but in Windows 7 guest VMs configured with NVIDIA vGPU, sparse textures are not enabled.

In Windows 10 quest VMs configured with NVIDIA vGPU, sparse textures are enabled and applications developed with Vulkan APIs run correctly in these VMs.

Status

Open

Ref

200381348

6.15. 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

Open

Ref.

2414897

6.16. 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

Press **Enter** or wait for the screen to update for NvFBC to capture the frame.

Status

Not a bug

Ref.

2115733

6.17. RDS sessions do not use the GPU with some Microsoft Windows Server releases

Description

When some releases of Windows Server are used as a quest 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.

- 1. Choose Local Computer Policy > Computer Configuration > Administrative Templates > Windows Components > Remote Desktop Services > Remote Desktop Session Host > Remote Session Environment.
- 2. Set the Use the hardware default graphics adapter for all Remote Desktop Services sessions option.

6.18. Cloned VMs configured with a vGPU type different than the type in the master image fail to start

Description

Cloned VMs configured with a vGPU type different than the type in the master image fail to start.

When a Windows 10 VM is booted, the VM becomes stuck in a loop and alternately displays Getting devices ready: 50% and Preparation in progress.

Workaround

Create one master image for each vGPU type that you want to use. Do not attempt to configure a cloned VM with a vGPU type different than the type in the master image.

Status

Not an NVIDIA bug

Ref.

2285306

6.19. 8.0-8.2 Only: Even when the scheduling policy is equal share, unequal GPU utilization is reported

Description

When the scheduling policy is equal share, unequal GPU engine utilization can be reported for the vGPUs on the same physical GPU.

For example, GPU engine usage for three P40-8Q vGPUs on a Tesla P40 GPU might be reported as follows:

root@localhost:~] nvidia-smi vgpu ded Jun 27 10:33:18 2018							
NVIDIA-SMI 390.59	Driver Version: 390.59						
GPU Name vGPU ID Name	Bus-Id VM ID VM Name	GPU-Util vGPU-Util					
	00000000:81:00.0 2122682 centos7.4-xmpl-211 2122692 centos7.4-xmpl-211 2122664 centos7.4-xmpl-211	0%					
1 Tesla P40 2122662 GRID P40-8Q 2122658 GRID P40-8Q 2122660 GRID P40-8Q	00000000:85:00.0 2122689 centos7.4-xmpl-211 2122667 centos7.4-xmpl-211						

The vGPU utilization of the vGPU 2122658 is reported as 59%. However, the expected vGPU utilization should not exceed 33%

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

Resolved in NVIDIA vGPU software 8.3

Ref.

2175888

6.20. 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:

_	localhost:~ g 20 10:33:	nvidia-smi vgr 18 2018	ou			
NVIDIA-SMI 390.42			Driver	Driver Version: 390.42		
GPU	Name vGPU ID	Name	Bus-Id VM ID		GPU-Util vGPU-Util	
0	Tesla P40 85109 87195 88095 89170 90475 93363	GRID P40-4Q GRID P40-4Q GRID P40-4Q	00000000 85110 87196 88096 89171	win7-xmpl-146048-1	99% 32% 39% 39% 39%	
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

6.21. Benign warnings during Virtual GPU Manager installation or uninstallation after hypervisor upgrade

Description

If the Virtual GPU Manager is installed or uninstalled after the hypervisor is upgraded, several warning messages about missing files are displayed.

The messages are similar to the following examples:

```
WARNING: Can't read module /lib/modules/4.4.77-1.el6.nutanix.20170830.124.x86 64/
weak-updates/ixgbevf.ko: No such file or directory
WARNING: Can't read module /lib/modules/4.4.77-1.el6.nutanix.20170830.124.x86_64/
weak-updates/i40evf.ko: No such file or directory
egrep: /lib/modules/4.4.77-1.el6.nutanix.20170830.124.x86 64//weak-updates/
ixgbevf.ko: No such file or directory
egrep: /lib/modules/4.4.77-1.el6.nutanix.20170830.124.x86 64//weak-updates/
i40evf.ko: No such file or directory
gzip: /boot/initramfs-4.4.77-1.el6.nutanix.20170830.124.x86 64.tmp: not in gzip
format.
WARNING: Can't read module /lib/modules/4.4.77-1.el6.nutanix.20170830.124.x86 64/
weak-updates/ixgbevf.ko: No such file or directory
WARNING: Can't read module /lib/modules/4.4.77-1.el6.nutanix.20170830.124.x86 64/
weak-updates/i40evf.ko: No such file or directory
```

After a hypervisor upgrade, files that are the targets of some symbolic links no longer exist. Although these missing files cause warning messages to be displayed during the installation or uninstallation of the Virtual GPU Manager, the installation or uninstallation is completed without errors.

Workaround

Ignore these messages as they are benign.

Status

Open

Ref.

200423757

6.22. Benign not in gzip format messages during Virtual GPU Manager installation or uninstallation

Description

During the installation or uninstallation of the Virtual GPU Manager, the warning messages are displayed:

```
gzip: /boot/initramfs-4.4.77-1.el6.nutanix.20170830.100726.x86 64.img: not in gzip
gzip: /boot/initramfs-4.4.77-1.el6.nutanix.20170830.100726.x86_64.tmp: not in gzip
```

Workaround

Ignore these messages as they are benign.

Status

Open

Ref.

200405700

6.23. License is not acquired in Windows VMs

Description

When a windows VM configured with a licensed vGPU is started, the VM fails to acquire a license.

Error messages in the following format are written to the NVIDIA service logs:

```
[000000020.860152600 sec] - [Logging.lib] ERROR: [nvGridLicensing.FlexUtility]
353@FlexUtility::LogFneError : Error: Failed to add trusted storage. Server
URL : license-server-url
[1,7E2,2,1[7000003F,0,9B00A7]]
System machine type does not match expected machine type..
```

Workaround

This workaround requires administrator privileges.

- 1. Stop the **NVIDIA Display Container LS** service.
- 2. Delete the contents of the folder %SystemDrive%:\Program Files\NVIDIA Corporation\Grid Licensing.
- 3. Start the NVIDIA Display Container LS service.

Closed

Ref.

200407287

6.24. 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

6.25. Resolution is not updated after a VM acquires a license and is restarted

Description

In a Red Enterprise Linux 7.3 guest VM, an increase in resolution from 1024×768 to 2560×1600 is not applied after a license is acquired and the gridd service is restarted. This issue occurs if the multimonitor parameter is added to the xorg.conf file.

Version

Red Enterprise Linux 7.3

Open

Ref.

200275925

6.26. 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

6.27. 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.

- 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

6.28. 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 quaranteed 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

6.29. 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 quest 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.

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.

Status

Open

Ref.

- 200130864
- 1803861

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

Description

GDM fails to start on Red Hat Enterprise Linux 7.2 and CentOS 7.0 with the following error: Oh no! Something has gone wrong!

Workaround

Permanently enable permissive mode for Security Enhanced Linux (SELinux).

- 1. As root, edit the /etc/selinux/config file to set SELINUX to permissive. SELINUX=permissive
- 2. Reboot the system.

~] # reboot

For more information, see <u>Permissive Mode</u> in *Red Hat Enterprise Linux 7 SELinux User's and* Administrator's Guide.

Status

Not an NVIDIA bug

Ref.

200167868

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. All rights reserved.

