

NVIDIA Data Center GPU Driver version 620.70.10 (Linux)

Release Notes

RN-08625-620 _v2.0 | November 2026

Table of Contents

Chapter 1. Version Highlights 1
1.1. Software Versions 1
1.2. Fixed Issues 1
1.3. Known Issues 1
Chapter 2. Virtualization 4
Chapter 3. Hardware and Software Support 6

Chapter 1. Version Highlights

This section provides highlights of the NVIDIA Data Center GPU R620 Driver (version 620.70.10 Linux).

For changes related to the 620 release of the NVIDIA display driver, review the file "NVIDIA_Changelog" available in the .run installer packages.

- Linux driver release date: 11/23/2026

1.1. Software Versions

For this release, the software versions are as follows:

- CUDA Toolkit 13: 13.x

  Note that starting with CUDA 11, individual components of the toolkit are versioned independently. For a full list of the individual versioned components (for example, nvcc, CUDA libraries, and so on), see the CUDA Toolkit Release Notes.
- NVIDIA Data Center GPU Driver: 620.70.10 (Linux)
- Fabric Manager: Use `nv-fabricmanager -v`
- NVFlash: 5.919.0

For more information on getting started with the NVIDIA Fabric Manager on NVSwitch-based systems (for example, NVIDIA HGX A100), refer to the Fabric Manager User Guide.

1.2. Fixed Issues

- None in this release.

1.3. Known Issues

- NVIDIA Confidential Computing has an ABI breakage in its attestation flows. Driver 620.70.10 is incompatible with Attestation SDK versions prior to 2026.10.02. More details can be found on the Confidential Computing Release Notes page.

This ABI breakage is limited to Attestation flows initiated via the Attestation SDK. It does not require any user-application recompilations, and has no effect on non-CC users.

- Package installation of the proprietary kernel module flavor is deprecated and starting in 615 is no longer provided. Debian transition packages for cuda-drivers are provided and RPM packages Obsolete cuda-drivers to migrate to nvidia-open. To install or lock to the proprietary driver packages follow the Version Locking instructions:

  On Ubuntu and Debian:

  ```
  sudo apt install nvidia-driver-pinning-620
  sudo apt install cuda-drivers
  ```

  On RHEL:

  ```
  sudo dnf versionlock \*nvidia\*620\*
  sudo dnf install cuda-drivers
  ```

- In certain rare scenarios, immediately after driver upgrades a launched CUDA process may hang with an infinite loop and become unkillable. Rebooting your system after upgrading from driver branch 610 to 615, as a one-time event, prevents this behavior from occurring.
- To conform to OpenCL 3.0 spec, starting with 610.xx drivers, NVIDIA OpenCL implementation will return the error `CL_INVALID_BUFFER_SIZE` when clCreateBuffer and related APIs are called with size greater than the value returned by `CL_DEVICE_MAX_MEM_ALLOC_SIZE`.

  With this change, earlier behavior where the corresponding OpenCL calls that would return success for the above case, will now see the error `CL_INVALID_BUFFER_SIZE` returned.
- NVIDIA OpenCL 3.0 drivers do not claim conformance for `cl_khr_fp16`. Device and platform queries do not report it as a supported extension. It should be considered experimental, without functional, conformance, or performance guarantees.
- Applications using `cuTensorMapEncodeTiled()` or `cuTensorMapEncodeIm2col*()` on Blackwell GPUs may encounter sporadic illegal memory accesses (IMA) or MMU faults (like XID 13) due to driver incorrectly configuring TMA descriptors when the tensor's backing memory allocation is less than 128KB and it is not a dense non-overlapping tensor. To workaround this issue, add the following after descriptor creation for such tensors.

  ```
  CUtensorMap tensorMap;
  cuTensorMapEncode*(&tensorMap, ...); // Create descriptor
  ((uint64_t*)&tensorMap)[1] &= ~((uint64_t)1 << 21); // Workaround : clear bit 85
  ```

  This workaround is specific to Blackwell's tensor map descriptor layout and may not be portable to future GPU architectures. It should be removed once updated drivers with the fix are deployed.
- This version of the GPU driver will fail to initialize on systems with Hopper GPUs subrevision = 3 and VBIOS versions older than 96.00.68.00.xx. Please ensure the system is using a VBIOS version 96.00.68.00.xx or newer before upgrading to this version of the driver.
- The default TCC mode in the NVIDIA driver does not support IOMMU-based isolation (necessary for Windows features such as DMA protection, kernel DMA guard,

virtualization-based security, etc.). The impacted GPUs are NVIDIA L40, NVIDIA L40S, NVIDIA L20, NVIDIA L4, and NVIDIA RTX PRO 6000 Blackwell Server Edition.

- This GPU Driver release is compatible only with Data Center GPU Manager (DCGM) versions 4.3.x or newer. Earlier versions of DCGM are not compatible.
- On RHEL 10 x86_64 with kernel 6.12.0-55.29.1.el10_0.x86_64 the `doca-ofed` package does not include the `ib_umad.ko` module, causing Fabric Manager to fail at startup. This issue is not present on the older kernel 6.12.0-55.9.1.el10_0.x86_64.
- If your environment meets all of the following conditions:
  - You are using an RPM based distribution.
  - You are not using the online CUDA package repository.
  - You manually installed `nvidia-fabricmanager-devel-620.70.10-1`.

  In this case, uninstall the package manually before installing the new version. This does not prevent the other driver packages from being upgraded successfully.

Chapter 2. Virtualization

To make use of GPU passthrough with virtual machines running Windows and Linux, the hardware platform must support the following features:

- A CPU with hardware-assisted instruction set virtualization: Intel VT-x or AMD-V.
- Platform support for I/O DMA remapping.
- On Intel platforms, the DMA remapper technology is called Intel VT-d.
- On AMD platforms, it is called AMD IOMMU.

Support for these features varies by processor family, product, and system, and should be verified at the manufacturer's website.

The following hypervisors are supported for virtualization:

Hypervisor	Notes
Citrix XenServer	Version 6.0 and later
VMware vSphere (ESX / ESXi)	Version 5.1 and later.
Red Hat KVM	Red Hat Enterprise Linux 9 with KVM
Microsoft Hyper-V	Windows Server 2019 Hyper-V Generation 2

Data Center products now support one display of up to 2560x1600 resolution.

The following GPUs are supported for device passthrough for virtualization:

GPU Family	Boards Supported
NVIDIA Rubin	NVIDIA VR NVL72
NVIDIA Blackwell	NVIDIA GB300, NVIDIA HGX B300, NVIDIA RTX 6000D, NVIDIA H20BFX, NVIDIA RTX 6000 PRO, NVIDIA GB200 NVL, NVIDIA HGX B200
NVIDIA Grace Hopper	NVIDIA GH200
NVIDIA Hopper	NVIDIA H100, NVIDIA H800
NVIDIA Ada Lovelace	NVIDIA L40, L4, L2, L20

GPU Family	Boards Supported
NVIDIA Ampere GPU Architecture	NVIDIA A800, A100, A40, A30, A16, A10, A10G, A2, AX800
NVIDIA Turing	NVIDIA T4, NVIDIA T4G

Chapter 3. Hardware and Software Support

Support for these features varies by processor family, product, and system, and should be verified at the manufacturer's website.

Coherent Driver-Based Memory Management (CDMM)

With R620 on all NVIDIA hardware-coherent platforms, the NVIDIA driver now defaults to Coherent Driver-based Memory Management (CDMM) instead of onlining GPU memory to the operating system as a NUMA node. NUMA mode remains fully supported. The mode is a node-wide setting controlled by a kernel module parameter and takes effect after a driver reload or reboot, so it should be selected before upgrading.

To select CDMM mode:

```
echo 'options nvidia NVreg_CoherentGPUMemoryMode=driver' | sudo tee
/etc/modprobe.d/nvidia-openrm.conf
```

To select NUMA mode:

```
# Loading NVIDIA module instructions

echo 'options nvidia NVreg_CoherentGPUMemoryMode=numa' \
| sudo tee /etc/modprobe.d/nvidia-openrm.conf

# NVIDIA-UVM module parameter

echo 'options nvidia-uvm uvm_disable_sam_migration=false' \
| sudo tee -a /etc/modprobe.d/nvidia-openrm.conf
```

Reload the driver or reboot for the change to take effect. To verify the active mode:

```
grep Coherent /proc/driver/nvidia/params

#check for SAM Migration
cat /sys/module/nvidia_uvm/parameters/uvm_disable_sam_migration
```

For a comparison of both modes, see blog post.



Note:

1. If there is already a configuration file for the NVIDIA driver, merge the options into a single options line.
2. To remove the configuration, undo its addition to the configuration file

3. To use GDRCopy with CDMM, please use version 2.5.1 or later of GDRCopy.

Supported Operating Systems for NVIDIA Data Center GPUs

The Release 620 driver is supported on the following operating systems:

- The following table summarizes the supported Linux 64-bit distributions. For a complete list of distributions, kernel versions supported, see the CUDA Linux System Requirements documentation.

Distribution	x86_64	Arm64 Server (Generic)	NVIDIA Grace Arm64 Server	NVIDIA Vera Arm64 Server
Debian 13.x (where x <= 7)	Yes	No	Yes	No
Debian 12.x (where x <= 15)**	Yes	No	Yes	No
OpenSUSE Leap 16.x (where x <= 1)	Yes	No	No	No
OpenSUSE Leap 15.x (where y = 6)**	Yes	No	No	No
Fedora 44	Yes	No	No	No
Red Hat Enterprise Linux for NVIDIA 26.03	No	No	No	Yes
Red Hat Enterprise Linux 10.x (where x <= 3)	Yes	Yes	Yes	No
Red Hat Enterprise Linux 9.y (where y <= 9)	Yes	Yes	Yes	No
Rocky Linux 10.x (where x <= 3)	Yes	No	No	No
Rocky Linux 9.y (where y <= 9)	Yes	No	No	No
Red Hat Enterprise Linux 8.y (where y <= 10)	Yes	Yes	No	No
Rocky Linux 8.y (where y <= 10)	Yes	No	No	No

Distribution	x86_64	Arm64 Server (Generic)	NVIDIA Grace Arm64 Server	NVIDIA Vera Arm64 Server
SUSE Linux Enterprise Server 16	Yes	Yes	Yes	No
SUSE Linux Enterprise Server 15.y (where y = 7)	Yes	Yes	Yes	No
Ubuntu 26.04.z LTS (where z <= 1)	Yes	Yes	Yes	No
Ubuntu 24.04.z LTS (where z <= 5)	Yes	Yes	Yes	No
Ubuntu 22.04.z LTS (where z <= 5)	Yes	Yes	Yes	No
KylinOS V11 SP3 2403	Yes	Yes	No	No
Amazon Linux AL2023	Yes	No	Yes	Yes
Microsoft Azure Linux 3.0	Yes	No	Yes	Yes
Oracle Linux 10	Yes	No	No	No
Oracle Linux 9	Yes	No	No	No
Oracle Linux 8	Yes	No	No	No

** The 620 branch will be the last release to support this OS.

Supported Operating Systems and CPU Configurations for NVIDIA RTX 6000D

Linux 64-bit distributions:

- Debian 12.15
- Ubuntu 24.04
- Ubuntu 22.04
- SUSE Linux Enterprise Server 16
- RedHat Enterprise Linus 10.3
- RedHat Enterprise Linux 9.9
- RedHat Enterprise Linux 8.10

Supported Operating Systems and CPU Configurations for NVIDIA RTX PRO 4500 Blackwell Server Edition

- Linux 64-bit distributions:
  - Ubuntu 24.04
  - RedHant Enterprise Linux 9.9
  - SUSE Linux Enterprise Server 15.7

Supported Operating Systems and CPU Configurations for NVIDIA H20 BFX

Linux 64-bit distributions:

- Ubuntu 22.04
- SUSE Linux Enterprise Server 15.7

Supported Operating Systems and CPU Configurations for NVIDIA DRIVE P2021

Linux 64-bit distributions:

- Ubuntu 24.04

Supported Operating Systems and CPU Configurations for NVIDIA RTX Pro 6000 Blackwell Server Edition

The Release 620 driver is validated with NVIDIA RTX Pro 6000 Blackwell Server Edition on the following operating systems and CPU configurations:

- Linux 64-bit distributions:
  - Debian 12.15
  - RedHat Enterprise Linux 10.3
  - RedHat Enterprise Linux 9.9
  - Rocky Linux 9.9
  - RedHat Enterprise Linux 8.10
  - OpenSUSE Leap 15.7
  - Fedora 44
  - Ubuntu 24.04
  - Ubuntu 22.04

Supported Virtualization Configurations

The Release 620 driver is validated with NVIDIA HGX A100, HGX A800, H100, and H800 on the following configurations:

- Passthrough (full visibility of GPUs and NVSwitches to guest VMs):
  - 8-GPU configurations with Ubuntu 22.04.5

API Support

This release supports the following APIs:

- NVIDIA® CUDA® 13.x for NVIDIA® Turing™, Hopper™, NVIDIA Ampere architecture, NVIDIA Ada Lovelace architecture, and NVIDIA Blackwell architecture GPUs
- OpenGL® 4.6
- Vulkan® 1.3
- Open Computing Language (OpenCL™ software) 3.0

Note that for using graphics APIs on Windows (such as OpenGL, Vulkan, DirectX 11, and DirectX 12) or any WDDM 2.0+ based functionality on Data Center GPUs, vGPU is required. See the vGPU documentation for more information.

Supported NVIDIA Data Center GPUs

The NVIDIA Data Center GPU driver package is designed for systems that have one or more Data Center GPU products installed. This release of the driver supports CUDA C/C++ applications and libraries that rely on the CUDA C Runtime and/or CUDA Driver API.

Attention: Release 470 was the last driver branch to support Data Center GPUs based on the NVIDIA Kepler architecture. This includes discontinued support for the following compute capabilities:

- sm_30 (NVIDIA Kepler)
- sm_32 (NVIDIA Kepler)
- sm_35 (NVIDIA Kepler)
- sm_37 (NVIDIA Kepler)

For more information on GPU products and compute capability, see https://developer.nvidia.com/cuda-gpus.

NVIDIA Server Platforms	
Product	**Architecture**
NVIDIA VR NVL72	NVR NVL72 and NVSwitch
NVIDIA GB300 NVL72	GB300 and NVSwitch
NVIDIA HGX B300 8-GPU	B300 and NVSwitch

NVIDIA Server Platforms	
Product	**Architecture**
NVIDIA HGX GB200 NVL	GB200 and NVLink
NVIDIA GB200 NVL4	GB200 and NVLink
NVIDIA HGX B200 8-GPU	B200 and NVSwitch
NVIDIA GH200	GH200 and NVLink
NVIDIA HGX H20-3e 8-GPU	H20 and NVSwitch
NVIDIA HGX H20 8-GPU	H20 and NVSwitch
NVIDIA HGX H200 8-GPU	H200 and NVSwitch
NVIDIA HGX H100 8-GPU	H100 and NVSwitch
NVIDIA HGX H800 8-GPU	H800 and NVSwitch
NVIDIA HGX H100 4-GPU	H100 and NVLink
NVIDIA HGX A800 8-GPU	A800 and NVSwitch
NVIDIA HGX A100 8-GPU	A100 and NVSwitch
NVIDIA HGX A100 4-GPU	A100 and NVLink

Data Center H-Series Products	
Product	**GPU Architecture**
NVIDIA H100 PCIe	NVIDIA Hopper
NVIDIA H100 NVL	NVIDIA Hopper
NVIDIA H200 NVL	NVIDIA Hopper
NVIDIA H800 PCIe	NVIDIA Hopper
NVIDIA H800 NVL	NVIDIA Hopper

Data Center L-Series Products	
Product	**GPU Architecture**
NVIDIA L2	NVIDIA Ada Lovelace
NVIDIA L20	NVIDIA Ada Lovelace
NVIDIA L40	NVIDIA Ada Lovelace
NVIDIA L40S	NVIDIA Ada Lovelace
NVIDIA L4	NVIDIA Ada Lovelace

RTX-Series / T-Series Products	
Product	**GPU Architecture**
NVIDIA RTX PRO 6000 Blackwell Server Edition	NVIDIA Blackwell
NVIDIA RTX PRO 6000 Blackwell Workstation Edition	NVIDIA Blackwell
NVIDIA RTX PRO 6000 Blackwell Max-Q Workstation Edition	NVIDIA Blackwell
NVIDIA RTX PRO 5500 Blackwell Workstation Edition	NVIDIA Blackwell
NVIDIA RTX PRO 5000 Blackwell 72GB	NVIDIA Blackwell
NVIDIA RTX PRO 5000 Blackwell	NVIDIA Blackwell
NVIDIA RTX PRO 4500 Blackwell Server Edition	NVIDIA Blackwell
NVIDIA RTX PRO 4500 Blackwell	NVIDIA Blackwell
NVIDIA RTX PRO 4000 Blackwell	NVIDIA Blackwell
NVIDIA RTX PRO 4000 SFF Blackwell	NVIDIA Blackwell
NVIDIA RTX PRO 2000 Blackwell	NVIDIA Blackwell
NVIDIA RTX 6000 Ada Generation	NVIDIA Ada Lovelace
NVIDIA RTX 5880 Ada Generation	NVIDIA Ada Lovelace
NVIDIA RTX 5000 Ada Generation	NVIDIA Ada Lovelace
NVIDIA RTX 4500 Ada Generation	NVIDIA Ada Lovelace
NVIDIA RTX 4000 Ada Generation	NVIDIA Ada Lovelace
NVIDIA RTX 4000 SFF Ada Generation	NVIDIA Ada Lovelace
NVIDIA RTX 2000 Ada Generation	NVIDIA Ada Lovelace
NVIDIA RTX 2000E Ada Generation	NVIDIA Ada Lovelace
NVIDIA RTX A6000	NVIDIA Ampere architecture
NVIDIA RTX A5500	NVIDIA Ampere architecture
NVIDIA RTX A5000	NVIDIA Ampere architecture
NVIDIA RTX A4500	NVIDIA Ampere architecture
NVIDIA RTX A4000H	NVIDIA Ampere architecture
NVIDIA RTX A4000	NVIDIA Ampere architecture
NVIDIA RTX A2000 12GB	NVIDIA Ampere architecture
NVIDIA RTX A2000	NVIDIA Ampere architecture

RTX-Series / T-Series Products	
Product	**GPU Architecture**
NVIDIA RTX A1000	NVIDIA Ampere architecture
NVIDIA RTX A400	NVIDIA Ampere architecture
NVIDIA RTX A800 40GB Active	NVIDIA Ampere architecture
Quadro RTX 8000	NVIDIA Turing
Quadro RTX 6000	NVIDIA Turing
Quadro RTX A6000	NVIDIA Turing
Quadro RTX 5000	NVIDIA Turing
Quadro RTX A5000	NVIDIA Turing
Quadro RTX 4000	NVIDIA Turing
Quadro RTX A4000	NVIDIA Turing
NVIDIA T1000 8GB	NVIDIA Turing
NVIDIA T600	NVIDIA Turing
NVIDIA T400 4GB	NVIDIA Turing
NVIDIA T400	NVIDIA Turing
NVIDIA T400E	NVIDIA Turing

Data Center A-Series Products	
Product	**GPU Architecture**
NVIDIA A2	NVIDIA Ampere architecture
NVIDIA A800, AX800	NVIDIA Ampere architecture
NVIDIA A100X	NVIDIA Ampere architecture
NVIDIA A100 NVIDIA A100 80 GB PCIe	NVIDIA Ampere architecture
NVIDIA A40	NVIDIA Ampere architecture
NVIDIA A30, A30X	NVIDIA Ampere architecture
NVIDIA A16	NVIDIA Ampere architecture
NVIDIA A10, A10M, A10G	NVIDIA Ampere architecture

Data Center T-Series Products	
Product	**GPU Architecture**
NVIDIA T4, T4G	NVIDIA Turing

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.

Trademarks

NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the United States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2026 NVIDIA Corporation & affiliates. All rights reserved.

NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051
https://www.nvidia.com