NVIDIA Data Center GPU Driver version 582.16 (Windows)

Release Notes

RN-08625-565 _v5.0 | January 2026

Table of Contents

Chapter 1. Version Highlights

This section provides highlights of the NVIDIA Data Center GPU R580 Driver (version 582.16 Windows).

For changes related to the 580 release of the NVIDIA display driver, review the file "NVIDIA_Changelog" available in the .run installer packages.

- Linux driver release date: 01/08/2026
- Windows driver release date: 01/08/2026

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: 582.16 (Windows)
- NVFlash: 5.791

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

- Specific Linux kernel versions and configurations will take a long time to scrub large memory buffers without encountering scheduling points. The fix: explicit scheduling points are inserted into a buffer populating loop.
- Because FSP has logic to conditionally reset depending on the PLM, the source isolating it to remove FSP access will fix this issue. The fix CL therefore does that; the source isolates the RESET PLM so FSP cannot access it.
- Move the allocation of `migrate_vma_state_t` to use `kvmalloc()` which will fall back to a virtually contiguous allocation if a physically contiguous allocation from `kmalloc()` is not possible.

Note that on 4K kernels, because the allocation is above `UVM_KMALLOC_THRESHOLD` so `vmalloc()` will be used. On 64K kernels `kmalloc()` will be used.

- Available multicast groups move some code to make sure we check the partition exists first before any other return code.
- Add a check before attempting to free the NULL pointer.
- Immediately cast each operand in `get_global_id` to `size_t` before the multiply/adds.
- Implemented the new functionality that reports the `WAITING_FOR_CONN_RECOVERY` state for connections that are blocking initialization.
- Improved/enhanced logging around connection/disconnected times.
- `cuGetDeviceCount` will fail and report `cudaErrorNoDevice` if GPU is disabled. The driver should return this error to hybrid_runtime but during the deinit flow for `cudaErrorNoDevice`, the driver frees the hybrid_runtime causing the hybrid_runtime return address to be invalidated on stack (stack corruption). Hybrid runtime is loaded by static loader via etlb and should be freed by static loader via etlb freeLibrary call.
- Set GPU health after the link state has been set to down for remove operations.
- Add proper synchronization for ISINK controller code.
- Enhanced GFM API SDK to include more detailed information of GPU, Switch, Partition, compute and switch nodes.
- Support operation of the fabric with a single switch tray failure/removal in the NVL domain. The GPUs will continue to operate with reduced bandwidth.
- SINK XIDs are only logged when power smoothing is enabled.
- Partitioning API calls are failed/reverted only when the call results in the partition becoming unhealthy.
- Do not advertise non-graphics MIG devices as physical GPUs when in MIG mode.
- Remove the old workaround in the NVIDIA driver to promote all spinlocks to raw spinlocks under `PREEMPT_RT`, which are no longer necessary.
- Introduced Multi-tenancy NVLink Partition Mode and Trusted NVLink Partition Mode.

  Multi-tenancy NVLink Partition Mode (default): GPU reset is needed when it is moved across partitions.

  Trusted NVLink Partition Mode: GPUs can move across partitions and start working without reset or service restart.

  The mode is controlled by a new FM config option `MNNVL_NVLINK_PARTITION_MODE`. The update to config option requires NMX-C restart to take effect.
- Remove WPR mappings from reserved memory mapping.

1.3. Known Issues

- 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.

- When upgrading from ClosedRM to OpenRM, nvidia-smi may fail.

  Workaround

  Run the following commands:

  ```
  sudo rpm -e  nvidia-open-driver-G06-kmp-default --nodeps
  sudo zypper in nvidia-driver-G06-kmp-default
  sudo zypper install -y nvidia-open-570
  ```

- 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-580.65.06-1`.

  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 Blackwell	NVIDIA HGX B300, NVIDIA RTX 6000D, NVIDIA H20BFX, NVIDIA RTX 6000 PRO, NVIDIA HGX GB200 NVL, NVIDIA HGX B200
NVIDIA Grace Hopper	NVIDIA GH200
NVIDIA Hopper	NVIDIA H100, NVIDIA H800
NVIDIA Ada Lovelace	NVIDIA L40, L4, L2, L20
NVIDIA Ampere GPU Architecture	NVIDIA A800, A100, A40, A30, A16, A10, A10G, A2, AX800

GPU Family	Boards Supported
NVIDIA Turing	NVIDIA T4, NVIDIA T4G
NVIDIA Volta	NVIDIA V100
NVIDIA Pascal	Quadro: P2000, P4000, P5000, P6000, GP100
	Tesla: P100, P40, P4
NVIDIA Maxwell	Quadro: K2200, M2000, M4000, M5000, M6000, M6000 24GB
	Tesla: M60, M40, M6, M4

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)

The R580 Driver introduces Coherent Driver-Based Memory Management (CDMM) for GB200 platforms. With CDMM, the driver manages GPU memory instead of the OS. CDMM avoids OS onlining of the GPU memory and the exposing of the GPU memory as a NUMA node to the OS. It is recommended that Kubernetes clusters enable CDMM to resolve potential memory over-reporting.

To set up the driver in CDMM mode, run the following commands and then reload the driver:

```
echo options nvidia NVreg_CoherentGPUMemoryMode=driver >
/etc/modprobe.d/nvidia-openrm.conf
```

Note:

1. If there is already a configuration file for the nvidia driver, please 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.
4. GPU Direct Storage is not supported with CDMM.

Supported Operating Systems for NVIDIA Data Center GPUs

The Release 580 driver is supported on the following operating systems:

- Windows x86_64 operating systems:
  - Microsoft Windows® Server 2025 24H2
  - Microsoft Windows® Server 2022 21H2
  - Microsoft Windows® 11 25H2

- Microsoft Windows® 11 24H2 - SV4
- Microsoft Windows® 11 23H2
- Microsoft Windows® 11 22H2 - SV2
- Microsoft Windows® 10 22H2

- The HGX platform also includes support for the Windows OS 64-bit distributions:
  - Microsoft Windows® Server 2025
  - Microsoft Windows® Server 2022
- Windows is supported only in shared NVSwitch virtualization configurations.

Supported Operating Systems and CPU Configurations for NVIDIA HGX B300

Windows 64-bit distributions:

- Microsoft Windows® Server 2025
- Microsoft Windows® Server 2022

Supported Operating Systems and CPU Configurations for NVIDIA RTX 6000D

Windows 64-bit distributions:

- Microsoft Windows® Server 2025
- Microsoft Windows® 11 24H2 - SV4
- Microsoft Windows® 11 23H2 - SV3

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

The Release 580 driver is validated with NVIDIA RTX Pro 6000 Blackwell Server Edition on the following operating systems and CPU configurations:

- Windows 64-bit distributions:
  - Microsoft Windows® Server 2025
  - Microsoft Windows® 11 24H2 - SV4
  - Microsoft Windows® 11 23H2 - SV3
  - Microsoft Windows® 10 22H2
  - Microsoft Windows® 10 21H2

API Support

This release supports the following APIs:

- NVIDIA® CUDA® 13.x for NVIDIA® Maxwell™, Pascal™, Volta™, Turing™, Hopper™, NVIDIA Ampere architecture, NVIDIA Ada Lovelace architecture, and NVIDIA Blackwell architecture GPUs
- OpenGL® 4.6
- Vulkan® 1.3
- DirectX 11
- DirectX 12 (Windows 10)
- 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 HGX GB200 NVL	GB200 and NVLink
NVIDIA HGX B200 8-GPU	B200 and NVSwitch
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 Server Platforms	
Product	**Architecture**
NVIDIA HGX A100 8-GPU	A100 and NVSwitch
NVIDIA HGX A100 4-GPU	A100 and NVLink
NVIDIA HGX-2	V100 and NVSwitch

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 5000 Blackwell	NVIDIA Blackwell
NVIDIA RTX PRO 4500 Blackwell	NVIDIA Blackwell
NVIDIA RTX PRO 4000 Blackwell	NVIDIA Blackwell
NVIDIA RTX PRO 2000 Blackwell	NVIDIA Blackwell
NVIDIA RTX PRO 4000 SFF Blackwell	NVIDIA Blackwell

RTX-Series / T-Series Products	
Product	**GPU Architecture**
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
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

Data Center V-Series Products	
Product	**GPU Architecture**
NVIDIA V100	Volta

Data Center P-Series Products	
Product	**GPU Architecture**
NVIDIA Tesla P100	NVIDIA Pascal
NVIDIA Tesla P40	NVIDIA Pascal
NVIDIA Tesla P4	NVIDIA Pascal

Data Center M-Class Products	
Product	**GPU Architecture**
NVIDIA Tesla M60	Maxwell
NVIDIA Tesla M40 24 GB	Maxwell
NVIDIA Tesla M40	Maxwell
NVIDIA Tesla M6	Maxwell
NVIDIA Tesla M4	Maxwell

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 Unites States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

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