# NVIDIA Data Center GPU Driver version 552.74 (Windows)

## Release Notes

RN-08625-550 _v4.0 | June 2024

# Table of Contents

# Chapter 1. Version Highlights

This section provides highlights of the NVIDIA Data Center GPU R550 Driver (version 552.74 Windows).

For changes related to the 550 release of the NVIDIA display driver, review the file "NVIDIA_Changelog" available in the .run installer packages.

- Windows driver release date: 07/09/2024

## 1.1. Software Versions

For this release, the software versions are as follows:

- CUDA Toolkit 12: 12.4

  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: 552.74 (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

- Security updates: see Security Bulletin: NVIDIA GPU Display Driver – July 2024, which is listed on the NVIDIA Product Security page. 4564626

## 1.3. Known Issues

- None.

# 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. |
| 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 Ada Lovelace | NVIDIA L40, L4 |
| NVIDIA Grace Hopper | NVIDIA GH200 |
| NVIDIA Hopper | NVIDIA H100, NVIDIA H800 |
| NVIDIA Ampere GPU Architecture | NVIDIA A800, A100, A40, A30, A16, A10, A10G, A2, AX800 |
| NVIDIA Turing | NVIDIA T4, NVIDIA T4G |
| NVIDIA Volta | NVIDIA V100 |
| NVIDIA Pascal | Quadro: P2000, P4000, P5000, P6000, GP100 |

| GPU Family | Boards Supported |
| --- | --- |
| NVIDIA Maxwell | Tesla: P100, P40, P4<br>Quadro: K2200, M2000, M4000, M5000, M6000, M6000 24GB<br>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.

## Supported Operating Systems for NVIDIA Data Center GPUs

The Release 550 driver is supported on the following operating systems:

- Windows x86_64 operating systems:
  - Microsoft Windows® Server 2022
  - Microsoft Windows® 11 21H2 - SV1
  - Microsoft Windows® 11 22H2 - SV2
  - Microsoft Windows® 11 23H2
  - Microsoft Windows® 10 21H2
  - Microsoft Windows® 10 22H2

## Supported Operating Systems and CPU Configurations for NVIDIA HGX H20

- Windows 64-bit distributions:
  - Windows Server 2022

## Supported Operating Systems and CPU Configurations for NVIDIA HGX H200

The Release 550 driver is validated with NVIDIA HGX H200 on the following operating systems and CPU configurations:

- Windows 64-bit distributions:
  - Windows Server 2022
  - Windows is supported only in shared NVSwitch virtualization configurations.

## Supported Operating Systems and CPU Configurations for NVIDIA HGX H100/H800

The Release 550 driver is validated with NVIDIA HGX H100 on the following operating systems and CPU configurations:

- Windows 64-bit distributions:
  - Windows Server 2022
  - Windows is supported only in shared NVSwitch virtualization configurations.

## Supported Operating Systems and CPU Configurations for NVIDIA HGX A100/A800

The Release 550 driver is validated with NVIDIA HGX A100 on the following operating systems and CPU configurations:

- Windows 64-bit distributions:
  - Windows Server 2022
  - Windows is supported only in shared NVSwitch virtualization configurations.
- CPU Configurations:
  - AMD Rome in PCIe Gen4 mode
  - Intel Skylake/Cascade Lake (4-socket) in PCIe Gen3 mode

## API Support

This release supports the following APIs:

- NVIDIA® CUDA® 12.4 for NVIDIA® Maxwell™, Pascal™, Volta™, Turing™, Hopper™, NVIDIA Ampere architecture, and NVIDIA Ada Lovelace GPU 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 H20 8-GPU | H20 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 |
| NVIDIA HGX-2 | V100 and NVSwitch |

| **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 |

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

| RTX-Series / T-Series Products | |
|---|---|
| **Product** | **GPU Architecture** |
| NVIDIA RTX 6000 Ada Generation | NVIDIA Ada Lovelace |
| NVIDIA RTX 4000 SFF Ada Generation | NVIDIA Ada Lovelace |
| NVIDIA RTX A6000 | NVIDIA Ampere architecture |
| NVIDIA RTX A5000 | NVIDIA Ampere architecture |
| NVIDIA RTX A4000 | NVIDIA Ampere architecture |
| Quadro RTX 8000 | NVIDIA Turing |
| Quadro RTX 6000 | NVIDIA Turing |
| Quadro RTX 4000 | NVIDIA Turing |
| NVIDIA T1000 | NVIDIA Turing |
| NVIDIA T600 | NVIDIA Turing |
| NVIDIA T400 | 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<br>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