NVIDIA DGX GB300 NVL72

Release Notes v1.0.0 for Multi-Node System Stack Firmware and Software

RN-11874-001_1.0.0 | September 2025

Document History

Version	Date	Description of Change
01	Oct 10, 2025	Release 1.0.0

Table of Contents

List of Tables

Overview

This document contains detailed information for the NVIDIA® DGX GB300 NVL72 v1.0.0 software/firmware release.

This document includes dependencies and instructions that are specific to this release, the versions provided as part of the release, and a list of known issues and improvements.

Legal Disclaimer: The System SW 1.0 GA Release is only for the target product DGX GB300 NVL72 and cannot be used for any other product.

NVIDIA DGX GB300 NVL72 Features

This section outlines the DGX GB300 NVL72 features included in this release, which provides essential operational capabilities to support end-to-end system qualification and validation.

This release includes the following features:

- NVLink Domain
  - 72x1 GPUs
- Fabric Manager
  - Support for NVLink multi-node deployment.
  - Peer-to-Peer Traffic routing/configuration.
  - Single default partition and multiple user partitions, with all available GPUs.
- IMEX Service
  - IMEX Service Peer-to-Peer Memory Import/Export support.
  - IMEX Dynamic Reconnect support.
  - IMEX Service multicast (NVLink SHARP) Import/Export support.
- RM Multi-User/Multi-Job Isolation support.
- Platform information for the compute and switch trays.
- NVIDIA multi-node CUDA®
  - Unicast traffic and multicast.
  - OpenMPI will be supported with NVLink Intranode and MNNVL internode.
  - Multi-Node NvBandwidth.
- NCCL
  - NCCL: All2All, AllReduce, and Unicast.
  - NVSHMEM
- CX8
  - CX8 Telemetry
- DCGM with multi-node testing support

Multi-Node System Software Stack Package Contents

The tables in this section provide the contents of the system software stack package for the NVIDIA DGX GB300 NVL72 reference design. Refer to the *FW update guide* for more information about installation, configuration, and workarounds.

This release includes public posting links necessary for creating the Multi-Node systems stack.

Compute Tray Components

Refer to the *NVIDIA FW update guide* for more information about the HMC and HMC NO-BIOS firmware bundle and installation instructions.

Note: The following are the host software versions tested with DGX GB300 Release 1.0.0.

Host Software Components

Component	Version
DOCA_Host	3.1.0-091513
MFT Tools	4.33.0-3002
GPU Driver	580.95.05
IMEX	580.95.05
CUDA	13.0.1
BF3_BFB	3.1.0-81
BF3_NIC	32.46.3048

HMC

Bundle file: nvfw_HGX-GBX00_0026_250904.1.0_custom_prod-signed.fwpkg

Component	Version
CPLD	0.22
GPU	97.10.4A.00.05
EROT	01.04.0031.0000_n04
HMC	GB200Nvl-25.08-B
SBIOS	02.05.06
SMR	1.56

Component	Version
CoRIM	HGX-GBX00_0026_250904.1.0_custom_prod-signed.corim
Recovery Firmware	nvfw_HGX-GBX00_0026_250904.1.0_custom_recovery_prod-signed.fwpkg
CRT	nvfw_GB300-P4059_fwpkg_prod_cert.crt
PEM	nvfw_GB300-P4059_fwpkg_prod_pub.pem
VBIOS Only Firmware	g548_0301_893__97104A0005.fwpkg

BMC

Bundle file: nvfw_DGX-GBX00_0026_250904.1.1_custom_prod-signed.fwpkg

Component	Version
EROT	01.04.0031.0000_n04
BMC	GB200Nvl-25.08-9
SMA	0003.00.0220.0001
CoRIM	DGX-GBX00_0026_250904.1.1_custom_prod-signed.corim
CRT	nvfw_GB300-p4058_corim_prod_cert.crt
PEM	nvfw_GB300-p4058_corim_prod_pub.pem

CX8

Component	Version
SP0 Firmware	fw-ConnectX8-rel-40_46_3048-900-9X86E-00CX-SP0_Ax-UEFI-14.39.14-FlexBoot-3.8.100.signed-MT_0000001226.fwpkg
SP0 CoRIM	fw-ConnectX8-rel-40_46_3048-900-9X86E-00CX-SP0_Ax-UEFI-14.39.14-FlexBoot-3.8.100.signed.corim
ST0 Firmware	fw-ConnectX8-rel-40_46_3048-900-9X86E-00CX-ST0_Ax-UEFI-14.39.14-FlexBoot-3.8.100.signed-MT_0000001513.fwpkg
ST0 CoRIM	fw-ConnectX8-rel-40_46_3048-900-9X86E-00CX-ST0_Ax-UEFI-14.39.14-FlexBoot-3.8.100.signed.corim

GB300 PS / MP NVSwitch Components

NVOS

Component	Version
NVOS	25.02.4282
SM	2025.06.5
GFM	580.82.11
NMX-C	3.0.216
NMX-T	3.0.12
QM3	35.2014.4686
CoRIM	GB300-P4093_0005_250825.1.0_prod-signed.corim

BMC+FPGA+EROT

Bundle file: nvfw_GB300-P4093_0004_250825.1.0_prod-signed.fwpkg

Component	Version
EROT	01.04.0031.0000_n04
BMC	88.0002.1955
SMR	0.24
CoRIM	GB300-P4093_0004_250825.1.0_prod-signed.corim
Recovery Firmware	nvfw_GB300-P4093_0004_250825.1.0_recovery_prod-signed.fwpkg

SBIOS+EROT

Bundle file: nvfw_GB300-P4093_0006_250731.1.0_prod-signed.fwpkg

Component	Version
EROT	01.04.0031.0000_n04
SBIOS	0ACTV_00.01.020

Component	Version
CoRIM	GB300-P4093_0006_250731.1.0_prod-signed.corim
Recovery Firmware	nvfw_GB300-P4093_0006_250731.1.0_recovery_prod-signed.fwpkg

CPLD

Bundle file: nvfw_GB300-P4093_0007_250824.1.0_prod-signed.fwpkg

Component	Version
CPLD1	CPLD000420_REV0201
CPLD2	CPLD000419_REV0200
CPLD3	CPLD000418_REV0200

NVIDIA/gdrcopy

NVIDIA/gdrcopy is a fast GPU memory copy library-based on NVIDIA GPUDirect RDMA technology.

GDRCOPY version 2.5 is a publicly available release.

- The code is released on (NVIDIA/gdrcopy).
- The prebuilt packages release is available on Index of /compute/redist/gdrcopy.
- Refer to Magnum IO GDRCopy for more information.

Component	Version
Stars	1006
Language	C++

System Software, Tools, and Drivers

Table 1. Public Release Links Associated with this Release

The following table provides a list of the public links associated with this release.

Description	Link
Datacenter Driver Version **580.82.07**	Datacenter Driver 580 downloads
CUDA Toolkit **13.0.1**	https://developer.nvidia.com/cuda-downloads **Note**: This is the link for the current release which might not match 13.0.1. If the current release does not match, locate release13.0.1 in the archives at Archive of Previous CUDA Releases.

Table 3. cuDNN

The following table provides the download information for the cuDNN release.

Component	Version	Location
cuDNN	**9.13.0**	NVIDIA cuDNN release notes

Table 4. NCCL

The following table provides information about where the NCCL release can be downloaded and links to the latest documentation.

Component	Location
NCCL Homepage	NVIDIA Collective Communications Library (NCCL) Download Page
PIP Wheels	NVIDIA Collective Communication Library (NCCL) Runtime
Github release	NVIDIA NCCL GitHub release
Devzone binaries Current release **2.28.3**	NVIDIA NCCL download page (current release) NVIDIA NCCL legacy download page (Legacy releases)
Release notes	NCCL release notes 2.28.3

Table 5. NVSHMEM

The following table provides information about where the NVSHEM release can be downloaded and links to the latest documentation.

Component	Location
NVSHMEM Homepage	NVIDIA NVSHMEM Homepage
Release **3.4.5**	NVIDIA NVSHMEM Downloads
Legacy releases	NVIDIA NVSHMEM Archive of legacy releases
Documentation	• Release Notes/Installation Guide: NVIDIA NVSHMEM Release Notes • API Documentation: NVIDIA OpenSHMEM Library (NVSHMEM) Documentation • Best Practice Guide: NVIDIA NVSHMEM Best Practice Guide

Table 6. DLFW Containers

The following table provides a list of the DL Frameworks containers.

Component	Versions	Location
PyTorch	Pytorch version **2.8.0a0** with Nvidia additions as part of DLFW 25.08 container	NGC catalog for PyTorch
Tensor RT	TensorRT Version **10.13.2.2**	NGC catalog for TensorRT
JAX	JAX version **0.6.2** with Nvidia additions as a part of DLFW **25.08** Container	NGC catalog for JAX
DL Frameworks Release Notes	**25.08**	NVIDIA Optimized Frameworks - NVIDIA Docs

Table 7. NVIDIA TensorRT

The following table provides information about where TensorRT can be downloaded and links to the latest documentation.

Attention: TensorRT version 10.13 includes functional support for GB300. TensorRT (to be released at a future date) will offer performance improvements that utilize GB300 system capabilities.

Component	Location
TensorRT Homepage	NVIDIA TensorRT Getting Started Documentation Nvonline: 1139359, 1139351,1139347,1139356
Release **10.13**	NVIDIA TensorRT current releases

Component	Location
Legacy releases	NVIDIA TensorRT legacy releases
Documentation	• Release Notes/Installation Guide: NVIDIA TensorRT Overview and Release Notes • API Documentation: NVIDIA TensorRT Documentation
TRT-LLM	nvcr.io/nvidia/tensorrt-llm/release:gb300-ea nvcr.io/nvidia/tensorrt-llm/devel:gb300-ea Source code https://github.com/NVIDIA/TensorRT-LLM/tree/feat/gb300-ea Early build with Beta quality .Please expect functional and performance issues since the support is in progress.

Table 8. MFT/Tools Firmware Package Contents

The following table provides the versions of the necessary MFT/tools to download the firmware bundle.

Component	Version	Location
DCGM	**4.4.1**	NVIDIA DCGM
NVSSVT	**1.6**	NVOnline: **1108364**
nvdebug	**1.7.0**	NVOnline: **1109504**
nvfwupd	**2.0.7**	**NVES Announcement**
nvrastool	**1.5**	NVOnline: **1112947**

Known Issues

Release 1.0.0

1. **GB300 CPU needs to clear the dpc_trigger_status.**

   In hotplug scenarios that involve CX8 downstream devices that enable DPC, some DPC events that are triggered by the hotplug attempts are not correctly handled by RAS firmware.

   Workaround: Disable DPC for CX8 downstream links before you attempt to hotplug the devices. This issue will be fixed in the next release.

2. **A GPU that is part of a multicast group cannot be removed from the partition used in MultiCast can be reused prematurely**

   A GPU participating in an active Multicast Team, it cannot be removed from a partition.

   Workaround:
   - Reset or power cycle all partition GPUs – On startup, the probe request will release any previously allocated multicast groups. Once the probe request completes, the GPU can then be removed from the partition.
   - Delete and recreate the partition – Delete the current partition and create a new one that excludes the GPUs intended for removal. NVIDIA recommends using a different partition ID for the new partition.

3. **Compute replacement with swap can cause workload to fail**

   If a compute tray is replaced and the old compute tray is re-added into the cluster into a different slot, it can lead to workloads failure.

   Workaround: Perform NVOS factory reset. Then reset all GPUs or reboot all compute trays. This issue will be fixed in a future release.

4. **NeighborMTUDiscards property is not defined in MetricReportDefinitions**

   NeighborMTUDiscards property is not defined in MetricReportDefinitions but exists in the Metric Reports.

   Workaround: This property will be removed in a future release. Property read will fail - do not use.

5. **Oneshot EDPp Scaling Factor (0x09) Returns Incorrect Low Value**

   The out-of-band NSM type 3 command 0x09 (Get Programmable EDPp Scaling Factor) can return an incorrect one-shot scaling factor value of 64 when one-shot scaling factor is cleared by the out-of-band admin. Redfish or any other tools that use this command would return an incorrect EDPp scaling factor.

   Workaround: There is currently no workaround, and this issue will be fixed in a future NVOS release.

6. **When an incorrect static topology file is specified on a 1x72 topology, FM log displays duplicate control plane error states**

Workaround: There is currently no workaround, and this issue will be fixed in a future NVOS release.

7. **GPU Status state missing from metric reports**

   Status/Status URI is missing from */redfish/v1/TelemetryService/MetricReports/HGX_ProcessorMetrics_0*

   Workaround: Status can be read from telemetry service processor metrics (/redfish/v1/Systems/HGX_Baseboard_0/Processors/GPU_{GpuId}#/Status/State). This issue will be fixed in a future release.

8. **Inconsistent GPU memory reported by NSM, nvidia-smi, and Redfish**

   Multiple management interfaces report different total GPU memory values: NSM Type 3 (0x0C) and nvidia-smi (-q -d MEMORY) return 284,208 MiB, while Redfish TotalMemorySizeMiB returns 285,324 MiB. This occurs when querying inventory or memory capacity from NSM, nvidia-smi, MODS, and Redfish.

   There is no data-loss or security impact identified.

   Workaround: There is currently no workaround, and this issue will be fixed in a future release.

9. **BMC PCIe link reset causes SBIOS exception.**

   When the PCIe link between the Grace CPU and BMC is reset at runtime, the system might take a fatal exception, which generates a Fatal CPER event.

   Workaround: Customers should avoid ungraceful PCIe link resets to the BMC while systems are operational.

10. No mechanism to disable host IPMI interface

This firmware release does not provide a mechanism to disable the host IPMI interface and any privileged user running on a compute tray can send IPMI commands to the BMC on the same compute tray.

Workaround

A future BMC firmware release will provide an interface to restrict host IPMI commands. In the meantime, necessary access controls need to be implemented on the compute trays to limit host privileged access by users not authorized to access the BMC.

Notice

The information provided in this specification is believed to be accurate and reliable as of the date provided. However, NVIDIA Corporation (“NVIDIA”) does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information. 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 publication supersedes and replaces all other specifications for the product that may have been previously supplied.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and other changes to this specification, at any time and/or to discontinue any product or service without notice. Customer should obtain the latest relevant specification 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. NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this specification.

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 these specifications will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do 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 specification. NVIDIA does not accept any 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 specification, 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 specification. 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 specification is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

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. 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 NVIDIA terms and conditions of sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, HGX, NVSwitch, and NVLink are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company 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
http://www.nvidia.com