



# DGX OS 5 Software

## Release Notes

# Table of Contents

Chapter 1. DGX OS Releases and Versioning.....	1
Chapter 2. DGX OS 5 Releases.....	3
2.1. New Features in DGX OS Release 5.4.....	5
2.2. New Features in DGX OS Release 5.3.....	6
2.2.1. Rotating the GPG Keys.....	6
2.2.1.1. Rotating the GPG Key For a Default Installation or After Reimaging.....	6
2.2.1.2. Rotating the GPG Keys for the DGX Software Stack.....	7
2.3. New Features in DGX OS Release 5.2.....	7
2.4. New Features in DGX OS Release 5.1.....	8
2.5. New Features in DGX OS Release 5.0.....	9
2.6. Update History.....	10
2.6.1. Update: November 22, 2022.....	10
2.6.2. Update: October 14, 2022.....	11
2.6.3. Update: June 7, 2022.....	13
2.6.4. Update: May 17, 2022.....	15
2.6.5. Update: April 28, 2022.....	15
2.6.6. Update: February 17, 2022.....	15
2.6.7. Update: December 14, 2021.....	17
2.6.8. Update: October 26 , 2021.....	18
2.6.9. DGX OS 5.1 Release: August 26, 2021.....	18
2.6.10. Update: June 30 , 2021.....	20
2.6.11. Update: June 20 , 2021.....	20
2.6.12. Update: June 2, 2021.....	20
2.6.13. Update: May 27, 2021.....	20
2.6.14. Update: May 06, 2021.....	20
2.6.15. Update: April 20, 2021.....	21
2.6.16. Update: April 13, 2021.....	21
2.6.17. Update: March 30, 2021.....	21
2.6.18. Update: March 2, 2021.....	21
2.6.19. Update: February 23, 2021.....	22
2.6.20. Update: January 20, 2021.....	22
2.6.21. Update: December 11, 2020.....	23
2.6.22. DGX OS 5.0 Release: October 31, 2020.....	23
Chapter 3. Known Issues Summary.....	25
3.1. Known Issues Details.....	26

3.1.1. Regression of CUDA application startup performance.....	26
3.1.2. NVSM Stress Test Logs Do Not Contain Summary Information.....	26
3.1.3. Unsupported Installation Options Appear in the ISO GRUB Menu.....	27
3.1.4. nvidia-release-upgrade May Report That Not All Updates Have Been Installed and Exit.....	27
3.1.5. EFI Boot Manager Lists Ubuntu as a Boot Option.....	27
3.1.6. Duplicate EFI Variable May Cause efibootmgr to Fail.....	28
3.1.7. Erroneous Insufficient Power Error May Occur for PCIe Slots.....	28
3.1.8. AMD Crypto Co-processor is not Supported.....	29
3.1.9. nvsm show alerts Reports NVSwitch PCIe Link Width is Degraded.....	29
3.1.10. nvsm show health Reports Firmware as Not Authenticated.....	30
3.1.11. Running NGC Containers Older than 20.10 May Produce "Incompatible MOFED Driver" Message.....	30
3.1.12. System May Slow Down When Using mpirun.....	31
3.1.13. Software Power Cap Not Reported Correctly by nvidia-smi.....	31
3.1.14. Forced Reboot Hangs the OS.....	32
3.1.15. Applications that call the cuCTXCreate API Might Experience a Performance Drop.....	32
3.1.16. NVIDIA Desktop Shortcuts Not Updated After a DGX OS Release Upgrade.....	32
3.1.17. Unable to Set a Separate/xinerama Mode through the xorg.conf File or through nvidia-settings.....	33
3.2. DGX OS Resolved Issues Details.....	34
3.2.1. NVSM Platform Displays as Unsupported.....	35
3.2.2. cudaMemFree CUDA API Performance Regression.....	35
3.2.3. NVSM Enumerates NVSwitches as 8-13 Instead of 0-5.....	35
3.2.4. BMC is not Detectable After Restoring BMC to Default.....	36
3.2.5. A System with Encrypted rootfs May Fail to Boot if one of the M.2 drives is Corrupted.....	36
3.2.6. NVSM Fails to Show CPU Information on Non-English Locales.....	37
3.2.7. Driver Version Mismatch Reported.....	37
3.3. Known Limitations Details.....	37
3.3.1. No RAID Partition Created After ISO Install.....	37
3.3.2. NSCQ Library and Fabric Manager Might Not Install When Installing a New NVIDIA Driver.....	38
3.3.3. System Services Startup Messages Appear Upon Completion of First-Boot Setup.....	38
3.3.4. [DGX A100]: Hot-plugging of Storage Drives not Supported.....	39
3.3.5. [DGX A100]: Syslog Contains Numerous "SM LID is 0, maybe no SM is running" Error Messages.....	39
3.3.6. [DGX-2]: Serial Over LAN Does not Work After Cold Resetting the BMC.....	39
3.3.7. [DGX-2]: Some BMC Dashboard Quick Links Appear Erroneously.....	40
3.3.8. [DGX-2]: Applications Cannot be Run Immediately Upon Powering on the DGX-2.....	40

3.3.9. [DGX-1]: Script Cannot Recreate RAID Array After Re-inserting a Known Good SSD.	41
3.3.10. [DGX Station A100] Suspend and Power Button Section Appears in Power Settings.	41
3.3.11. [DGX-2] NVSM Does not Detect Downgraded GPU PCIe Link.....	41
<b>Appendix A. Downgrading Firmware for Mellanox ConnectX-4 Cards.....</b>	<b>43</b>
A.1. Checking the Device Type.....	43
A.2. Downgrading the Firmware.....	43
<b>Appendix B. DGX Software Stack.....</b>	<b>45</b>

---

# Chapter 1. DGX OS Releases and Versioning

This information helps you understand the DGX OS release numbering convention and your options to upgrade your DGX OS software.

## DGX OS Releases

DGX OS is a customized Linux distribution that is based on Ubuntu Linux. It includes platform-specific configurations, diagnostic and monitoring tools, and the drivers that are required to provide the stable, tested, and supported OS to run AI, machine learning, and analytics applications on DGX systems.

DGX OS is released twice a year, typically around February and August, for two years after the first release. Updates are provided between releases and thereafter for the entire support duration.

### Release Versions

The DGX OS release numbering convention is MAJOR.MINOR, and it defines the following types of releases:

- ▶ **Major releases** are typically based on Ubuntu releases, which include new kernel versions and new features that are not always backwards compatible.

For example:

- ▶ DGX OS 6.x releases are based on Ubuntu 22.04
- ▶ DGX OS 5.x releases are based on Ubuntu 20.04
- ▶ DGX OS 4.x is based on Ubuntu 18.04
- ▶ **Minor releases** include mostly new NVIDIA features and accumulated bug fixes and security updates.

These releases are incremental and always include all previous software changes.

- ▶ In DGX OS 4 and earlier, minor releases were also typically aligned with NVIDIA Graphics Drivers for Linux releases.
- ▶ In DGX OS 5 and OS 6, you now have the option to install newer NVIDIA Graphic Drivers independently of the DGX OS release.

## DGX OS Release Mechanisms

This section provides information about the DGX OS release mechanisms that are available to install or upgrade to the latest version of the DGX OS.

### The ISO Image

DGX OS is released as an ISO image that includes the necessary packages and an autonomous installer. Updated versions of the ISO image are also released that:

- ▶ Provide bug fixes and security mitigations.
- ▶ Improve the installation experience.
- ▶ Provide hardware configuration support.

You should always use the latest ISO image, except when you need to restore the system to an earlier version.



**WARNING:** This image allows you to install or reimage a DGX system to restore the system to a default state, but the process erases all of the changes that you applied to the OS.

### The Linux Software Repositories

Upgrades to DGX OS are provided through the software repositories. Software repositories are storage locations from which your system retrieves and installs OS updates and applications. The repositories used by DGX OS are hosted by Canonical for the Ubuntu OS and NVIDIA for DGX specific software and other NVIDIA software. Each repository is a collection of software packages that are intended to install additional software and to update the software on DGX systems.

New versions of these packages, which contain bug fixes and security updates, provide an update to DGX OS releases. The repositories are also updated to include hardware enablement, which might add support for a new system or a new hardware component, such as a network card or disk drive. This update does not affect existing hardware configurations.

System upgrades are cumulative, which means that your systems will **always** receive the **latest version** of all of the updated software components. You cannot select which upgrades to make or limit upgrades to the non-latest DGX OS 6.x release.



**Important:** We recommend that you **do not** update **only** individual components.

**Before** you update a system, refer to the [DGX OS Software Release Notes](#) for a list of the available updates. For more information on displaying available updates and upgrade instructions, refer to the [DGX OS 5 User Guide](#).

---

# Chapter 2. DGX OS 5 Releases

The following are the key features of DGX OS Release 5:

- ▶ Supports all NVIDIA servers, DGX Station, and DGX Station A100 in one ISO image.
- ▶ Based on Ubuntu 20.04
- ▶ Includes drive encryption for added security.

## UPDATE ADVISEMENT

### ▶ NVIDIA KVM not Supported

This release does not support the Linux Kernel-based Virtual Mode (KVM) on DGX systems.



**Note:** NVIDIA KVM is available only with DGX-2 systems. DGX-2 customers that require this feature should stay with the latest DGX OS Server 4.x release.

### ▶ Update DGX OS on DGX A100 before updating VBIOS

DGX A100 systems running DGX OS earlier than version 4.99.8 should be updated to the latest version before updating the VBIOS to version 92.00.18.00.0 or later. Failure to do so will result in the GPUs not getting recognized.

### ▶ NGC Containers

With DGX OS 5, customers should update their NGC containers to container release 20.10.17 or later if they are using multi-node training. For all other use cases, refer to the [NGC Framework Containers Support Matrix](#).


Refer to the [NVIDIA Deep Learning Frameworks](#) documentation for information about the latest container releases and how to access the releases.

### ▶ Ubuntu Security Updates

Customers are responsible for keeping the DGX server up to date with the latest Ubuntu security updates using the 'apt full-upgrade' procedure. See the [Ubuntu Wiki Upgrades](#) web page for more information. Also, the Ubuntu Security Notice site (<https://usn.ubuntu.com/>) lists known Common Vulnerabilities and Exposures (CVEs), including those that can be resolved by updating the DGX OS software.

## CURRENT VERSIONS

Here is a current list of the main DGX software stack component versions in the software repositories:

Component	Version	Additional Information
CUDA Toolkit	11.4	<p>Refer to the <a href="#">NVIDIA CUDA Toolkit Release Notes</a>.</p> <div>  <b>Note:</b> For DGX servers, CUDA is updated only if it has been previously installed. </div>
GPU Driver	R450: 450.216.04 R470: 470.161.03 R515: 515.86.01	Refer to the <a href="#">NVIDIA Data Center GPU documentation</a>
NCCL	<a href="#">2.15.1</a>	
cuDNN	<a href="#">8.4.1</a>	
DCGM	<a href="#">2.4.7</a>	Refer to the <a href="#">DCGM Release Notes</a> .
Mellanox OFED	MLNX 5.4-3.5.8.0	Refer to <a href="#">MLNX 5.4-3.5.8.0</a>
MLNX FW	ConnectX-4 12.28.2006 ConnectX-5 16.31.2006 ConnectX-6 20.31.2354 ConnectX-7 28.34.4000	
NVSM	22.09.03	Refer to the <a href="#">NVIDIA System Management Documentation</a> .
Docker Engine	docker-ce: 20.10.21  <b>Note:</b> If necessary, the following components require separate installation via <code>sudo apt install</code> : <ul style="list-style-type: none"> <li>▶ docker-ce-rootless-extras 20.10.18</li> <li>▶ docker-scan-plugin 0.9.0</li> </ul>	Refer to <a href="#">v20.10.21</a>
NVIDIA Container Toolkit	nvidia-container-toolkit: 1.10.0-1  nvidia-docker2: 2.11.0-1	Refer to the <a href="#">NVIDIA Container Toolkit</a> documentation.



Component	Version	Additional Information
	libnvidia-container1: 1.10.0-1 libnvidia-container-tools: 1.10.0-1	
NGC CLI	2.2.0-1	Refer to the <a href="#">NGC CLI Documentation</a>
GPUDirect Storage (GDS)	v1.0	Refer to <a href="#">GDS Documentation</a>
MIG Configuration Tool	nvidia-mig-manager 0.4.3	Refer to the following NVIDIA mig-parted github pages: <a href="https://github.com/NVIDIA/mig-parted">https://github.com/NVIDIA/mig-parted</a> and <a href="https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd">https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd</a>
nvipmitool	1.0.6.0	
nvidia-peer-memory/nvidia-peer-memory DKMS	1.3.0	

In addition to upgrading to the versions described in this section, performing an over-the-network update will also upgrade the Ubuntu 20.04 LTS version and Ubuntu kernel, depending on when the upgrade is performed.

For a list of updates in DGX OS 5, see [Update History](#).

## 2.1. New Features in DGX OS Release 5.4

Here are the new features in DGX OS 5.4 (see also the [Update History](#) for important changes made since the initial release):

- ▶ GPUDirect Storage 1.0 was added

Upgraded Software packages:

- ▶ R450, R470 GPU drivers (see Software Contents below for versions)
- ▶ NCCL 2.15.1
- ▶ NVSM to 22.06.02
- ▶ DCGM to 2.4.7
- ▶ MLNX OFED to 5.4-3.5.8.0
- ▶ docker-ce: 20.10.18
- ▶ MIG Configuration Tool: 0.4.3

The newest version of `nvidia-mig-parted` now contains a set of checkpoint/restore commands. These allow one to checkpoint (and later restore) the MIG configuration applied across all GPUs on a node, regardless of what tool was used to set up those MIG configurations.

In previous versions of `nvidia-mig-parted`, all MIG configurations had to be done via `nvidia-mig-parted` itself in order for it to recognize and subsequently reconfigure the MIG state on set of GPUs. With this new checkpoint/restore feature, tools such as `nvidia-smi` can be used to configure MIG as well.

To use this feature, one would run (for example):

```
$ sudo nvidia-smi mig -C -cgi 1g.5gb,1g.5gb,1g.5gb,1g.5gb,1g.5gb,1g.5gb,1g.5gb
$ sudo -E nvidia-mig-parted checkpoint
```

This will save a checkpoint of the current MIG state to the default location of `/var/lib/nvidia-mig-manager/checkpoint.json`. Later (after a reboot, for example) users can run `restore` to ensure that the checkpointed MIG configuration is properly restored:

```
$ sudo -E nvidia-mig-parted restore
```

## 2.2. New Features in DGX OS Release 5.3



**Important:** The features and component versions in DGX OS 5.3 are identical to the versions in DGX OS 5.2. In DGX OS 5.3, the GPG keys that are used to sign the packages and metadata in those repositories need to be rotated. Refer to [Rotating the GPG Keys](#) for more information.

See also the [Update History](#) for important changes made since the initial release.

### 2.2.1. Rotating the GPG Keys

NVIDIA constantly evaluates and improves security implementations. As part of these improvements, we are rolling out changes to harden the security and reliability of our repositories. These changes require rotating the GPG keys that are used to sign the metadata and packages in those repositories.

#### 2.2.1.1. Rotating the GPG Key For a Default Installation or After Reimaging

This section provides information about how to rotate the GPG keys for a default DGX OS installation from the factory or after you reimage with the DGX OS ISO.

1. Download the new repository setup packages.

```
wget https://repo.download.nvidia.com/baseos/ubuntu/focal/x86_64/pool/common/n/
nvidia-repo-keys/nvidia-repo-keys_22.04-1_all.deb
wget https://repo.download.nvidia.com/baseos/ubuntu/focal/x86_64/pool/dgx/n/
nvidia-repos/dgx-repo_21.07-1_amd64.deb
wget https://repo.download.nvidia.com/baseos/ubuntu/focal/x86_64/pool/common/n/
nvidia-repos/cuda-compute-repo_21.07-1_amd64.deb
```

2. Directly install the `.deb` packages, which skips the GPG check performed in `apt`.



**Note:** If prompted, ensure that you accept the maintainer's version for all files.

```
$ sudo dpkg --force-confnew -i ./nvidia-repo-keys_22.04-1_all.deb ./dgx-  
repo_21.07-1_amd64.deb ./cuda-compute-repo_21.07-1_amd64.deb
```

3. Manually revoke the previous DGX and CUDA GPG keys.

```
sudo apt-key del 629C85F2  
sudo apt-key del 7FA2AF80
```

OTA updates can now occur as normal.

### 2.2.1.2. Rotating the GPG Keys for the DGX Software Stack

This section provides information about how to rotate the GPG keys if you installed Ubuntu and the DGX Software Stack.

1. Download the updated dgx-repo-files tarball and extract its contents onto the root filesystem.

```
curl https://repo.download.nvidia.com/baseos/ubuntu/focal/dgx-repo-files.tgz |  
sudo tar xzf - -C /
```

2. Manually revoke the previous DGX and CUDA GPG keys.

```
$ sudo apt-key del 629C85F2  
$ sudo apt-key del 7FA2AF80
```

OTA updates can now occur as normal.

## 2.3. New Features in DGX OS Release 5.2

Here are the new features in DGX OS 5.2 (see also the [Update History](#) for important changes made since the initial release):

- ▶ Updated NVSM to 21.09.14
- ▶ Updated DCGM to 2.3.2
- ▶ Added DGX Software Stack installation method

The DGX Software Stack provides the option to install a vanilla version of Ubuntu 20.04 and then separately install the additional NVIDIA software (NVIDIA DGX Software Stack). This option is available for DGX servers (DGX A100, DGX-2, DGX-1). The DGX Software Stack is a stream-lined version of the software stack incorporated into the DGX OS ISO image, and includes meta-packages to simplify the installation process. Refer to the [DGX Software Stack for Ubuntu Installation Guide \(http://docs.nvidia.com/dgx/dgx-ubuntu-install-guide/index.html\)](http://docs.nvidia.com/dgx/dgx-ubuntu-install-guide/index.html) for instructions.

### UPDATE ADVISEMENT

- ▶ **IMPORTANT:** This release incorporates the following updates.
  - ▶ NVIDIA MLNX\_OFED 5.4

Customers are advised to consider these updates and any effect they may have on their application. For example, some MOFED-dependent applications may be affected.

A best practice is to upgrade on select systems and verify that your applications work as expected before deploying on more systems.

## 2.4. New Features in DGX OS Release 5.1

Here are the new features in DGX OS 5.1 (see also the [Update History](#) for important changes made since the initial release):

- ▶ Added NVIDIA GPU driver Release 470.



**Note:** When upgrading DGX OS, the system remains on the installed GPU driver branch. For example, the GPU driver branch on the system does not automatically switch from R450 to R470. Refer to the [Changing Your GPU Branch](#) section of the DGX OS User Guide for instructions on switching GPU driver branches.

- ▶ Supports the CUDA Toolkit up to 11.4 natively, or newer versions via the compatibility module.
- ▶ Updated the Docker Engine to 20.10.
- ▶ Incorporates NVIDIA MLNX\_OFED 5.4.
- ▶ Updated NVSM
  - ▶ Added ability to generate a test alert/email.
  - ▶ NVSM `dump/show health` includes firmware version information (incorporates 'nvsm show -level all' in the command).
  - ▶ NVSM binds port 273 to 127.0.0.1 to limit external communications.  
To open other ports for IPV4 or IPV6, edit `nvsm.config` (bindaddress) and then restart NVSM
- ▶ Added NVML libraries
- ▶ Includes MOFED 5.4
- ▶ Added NGC CLI
- ▶ Added MIG Configuration Tool to define MIG partitions and provide a systemd service to make MIG partitions persist across reboots.
  - ▶ MIG is disabled by default
  - ▶ The MIG configuration file overrides any MIG-related `nvidia-smi` commands. Use `nvidia-mig-parted` instead of `nvidia-smi` for MIG configuration.
- ▶ `arp_ignore=1` and `arp_announce=2` are now set on all InfiniBand configured interfaces.
- ▶ Added LLDPd for validating network cabling  
The default configuration is now set to use the PortID of the interface name rather than the MAC address.
- ▶ Supports GPUDirect Storage 1.0 (Refer to [GDS Documentation](#) for installation instructions)

## UPDATE ADVISEMENT

- ▶ **IMPORTANT:** This release incorporates the following updates.

- ▶ NVIDIA MLNX\_OFED 5.4

Customers are advised to consider these updates and any effect they may have on their application. For example, some MOFED-dependent applications may be affected.

A best practice is to upgrade on select systems and verify that your applications work as expected before deploying on more systems.

## 2.5. New Features in DGX OS Release 5.0

Here are the new features in DGX OS 5.0 (see also the [Update History](#) for important changes made since the initial release):

- ▶ NVIDIA GPU driver Release 450.
- ▶ Supports the CUDA Toolkit up to 11.0 natively, or newer versions via the compatibility module.
- ▶ Incorporates NVIDIA MLNX\_OFED 5.1.
- ▶ Added rootfs encryption option, configurable during the re-imaging process.
- ▶ Added option to password protect the GRUB menu, configurable during the first boot process.
- ▶ Updated NVSM
- ▶ Added support for custom drive partitioning
- ▶ Added monitoring of firmware health
- ▶ Updated the default InfiniBand network naming policy.

The InfinBand interfaces, enumerated as `ibx` in previous releases, now enumerate as `ibpxsy` (similar to Ethernet (`enpxsy`)). Refer to the [DGX A100 User Guide](#) for the new naming.

## UPDATE ADVISEMENT

- ▶ **IMPORTANT:** This release incorporates the following updates.

- ▶ NVIDIA MLNX\_OFED 5.1

Customers are advised to consider these updates and any effect they may have on their application. For example, some MOFED-dependent applications may be affected.

A best practice is to upgrade on select systems and verify that your applications work as expected before deploying on more systems.

## 2.6. Update History

This section provides information about the updates to DGX OS 5.

The updates listed include:

- ▶ Major component updates in the Ubuntu repositories.
- ▶ NVIDIA driver updates in the Ubuntu repository

Refer to [Installing the DGX OS \(Reimaging the System\)](#) for instructions on how to install DGX OS from the ISO image,

Refer to [Performing Package Updates](#) for instructions on how to update DGX OS with all the latest DGX OS 5 updates from the network repositories.

### 2.6.1. Update: November 22, 2022

- ▶ GPUDirect Storage 1.0 was added.
- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ R515 NVIDIA GPU Driver: 515.86.01
  - ▶ R470 NVIDIA GPU Driver: 470.161.03
  - ▶ R450 NVIDIA GPU Driver: 450.216.04



**Note:** When upgrading DGX OS, the system remains on the installed GPU driver branch. For example, the GPU driver branch on the system does not automatically switch from R450 to R470. Refer to the [Changing Your GPU Branch](#) section of the DGX OS User Guide for instructions on switching GPU driver branches.

Here are the contents of the DGX OS 5.4.1 ISO:

Component	Version	Additional Information
CUDA Toolkit	11.4	Refer to the <a href="#">NVIDIA CUDA Toolkit Release Notes</a> .  <b>Note:</b> For DGX servers, CUDA is updated only if it has been previously installed.
GPU Driver	R450: 450.216.04 R470: 470.161.03 R515: 515.86.01	Refer to the <a href="#">NVIDIA Data Center GPU documentation</a>
NCCL	<a href="#">2.15.1</a>	
cuDNN	<a href="#">8.4.1</a>	


Component	Version	Additional Information
DCGM	<a href="#">2.4.7</a>	Refer to the <a href="#">DCGM Release Notes</a> .
Mellanox OFED	MLNX 5.4-3.5.8.0	Refer to <a href="#">MLNX 5.4-3.5.8.0</a>
MLNX FW	ConnectX-4 12.28.2006 ConnectX-5 16.31.2006 ConnectX-6 20.31.2354 ConnectX-7 28.34.4000	
NVSM	22.09.03	Refer to the <a href="#">NVIDIA System Management Documentation</a> .
Docker Engine	docker-ce: 20.10.21 <b>Note:</b> If necessary, the following components require separate installation via <code>sudo apt install</code> : <ul style="list-style-type: none"> <li>▶ docker-ce-rootless-extras 20.10.18</li> <li>▶ docker-scan-plugin 0.9.0</li> </ul>	Refer to <a href="#">v20.10.21</a>
NVIDIA Container Toolkit	nvidia-container-toolkit: 1.10.0-1 nvidia-docker2: 2.11.0-1 libnvidia-container1: 1.10.0-1 libnvidia-container-tools: 1.10.0-1	Refer to the <a href="#">NVIDIA Container Toolkit</a> documentation.
NGC CLI	2.2.0-1	Refer to the <a href="#">NGC CLI Documentation</a>
GPUDirect Storage (GDS)	v1.0	Refer to <a href="#">GDS Documentation</a>
MIG Configuration Tool	nvidia-mig-manager 0.4.3	Refer to the following NVIDIA mig-parted github pages: <a href="https://github.com/NVIDIA/mig-parted">https://github.com/NVIDIA/mig-parted</a> and <a href="https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd">https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd</a>
nvipmitool	1.0.6.0	
nvidia-peer-memory/nvidia-peer-memory DKMS	1.3.0	

## 2.6.2. Update: October 14, 2022

- ▶ GPUDirect Storage 1.0 was added.

- ▶ The following changes were made to the Ubuntu repositories:


- ▶ **R470 NVIDIA GPU Driver:** 470.129.06
- ▶ **R450 NVIDIA GPU Driver:** 450.203.03

- ▶  **Note:** When upgrading DGX OS, the system remains on the installed GPU driver branch. For example, the GPU driver branch on the system does not automatically switch from R450 to R470. Refer to the [Changing Your GPU Branch](#) section of the DGX OS User Guide for instructions on switching GPU driver branches.

- ▶ The following changes were made to the Ubuntu repositories:

- ▶ NCCL 2.15.1
- ▶ DCGM 2.4.7
- ▶ MOFED 5.4-3.5.8.0
- ▶ NVSM 22.06.02
- ▶ Docker-ce 20.10.18
- ▶ MIG Configuration Tool: 0.4.3


- ▶ Here are the contents of the DGX OS 5.4.1 ISO:

Component	Version	Additional Information
CUDA Toolkit	11.4	Refer to the <a href="#">NVIDIA CUDA Toolkit Release Notes</a> .   <b>Note:</b> For DGX servers, CUDA is updated only if it has been previously installed.
GPU Driver	R470: 470.129.06 R450: 450.203.03	Refer to the <a href="#">NVIDIA Data Center GPU documentation</a>
DCGM	<a href="#">2.4.7</a>	Refer to the <a href="#">DCGM Release Notes</a> .
Mellanox OFED	<b>MLNX 5.4-3.5.8.0</b>	Refer to <a href="#">MLNX 5.4-3.5.8.0</a>
MLNX FW	ConnectX-4 12.28.2006 ConnectX-5 16.31.2006 <b>ConnectX-6 20.31.2354</b> <b>ConnectX-7 28.34.4000</b>	
NVSM	22.06.02	Refer to the <a href="#">NVIDIA System Management Documentation</a> .
Docker Engine	docker-ce: 20.10.18 <b>Note:</b> If necessary, the following components require	Refer to <a href="#">v20.10.18</a>




Component	Version	Additional Information
	separate installation via <code>sudo apt install</code> : <ul style="list-style-type: none"> <li>▶ <code>docker-ce-rootless-extras 20.10.18</code></li> <li>▶ <code>docker-scan-plugin 0.9.0</code></li> </ul>	
NVIDIA Container Toolkit	<code>nvidia-container-toolkit: 1.10.0-1</code> <code>nvidia-docker2: 2.11.0-1</code> <code>libnvidia-container1: 1.10.0-1</code> <code>libnvidia-container-tools: 1.10.0-1</code>	Refer to the <a href="#">NVIDIA Container Toolkit</a> documentation.
NGC CLI	2.2.0-1	Refer to the <a href="#">NGC CLI Documentation</a>
GPUDirect Storage (GDS)	v1.0	Refer to <a href="#">GDS Documentation</a>
MIG Configuration Tool	<code>nvidia-mig-manager 0.4.3</code>	Refer to the following NVIDIA mig-parted github pages: <a href="https://github.com/NVIDIA/mig-parted">https://github.com/NVIDIA/mig-parted</a> and <a href="https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd">https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd</a>
<code>nvipmitool</code>	1.0.6.0	
<code>nvidia-peer-memory/nvidia-peer-memory DKMS</code>	1.3.0	

### 2.6.3. Update: June 7, 2022

- ▶ Installer version updated to 5.3.1.
- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **R470 NVIDIA GPU Driver:** 470.129.06
  - ▶ **R450 NVIDIA GPU Driver:** 450.191.01
  - ▶  **Note:** When upgrading DGX OS, the system remains on the installed GPU driver branch. For example, the GPU driver branch on the system does not automatically switch from R450 to R470. Refer to the [Changing Your GPU Branch](#) section of the DGX OS User Guide for instructions on switching GPU driver branches.
- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **DCGM:** 2.3.6
  - ▶ **NVSM:** 22.03.05
  - ▶ **Docker CE:** 20.10.16
  - ▶ **nvidia-peer-memory/nvidia-peer-memory DKMS:** 1.3.0

- The **DGX OS 5.3.1 ISO** has been released.

Here are the contents of the DGX OS 5.3.1 ISO:

Component	Release with R450	Release with R470	Additional Information
Ubuntu	20.04 LTS20.04 LTS		Refer to the <a href="#">Ubuntu 20.04 Desktop Guide</a> .
Ubuntu kernel	<b>5.4.0-113-generic</b>		See <a href="#">Linux 5.4.0-113-generic</a> .
GPU Driver	450.191.01	470.129.06   <b>Note:</b> Updating from R450 to R470 does not happen automatically when updating DGX OS 5, but requires separate steps. Refer to the <a href="#">Changing Your GPU Branch</a> section of the DGX OS User Guide for instructions.	Refer to the <a href="#">NVIDIA Data Center GPU documentation</a> .
CUDA Toolkit	<b>11.4</b>		Refer to the <a href="#">NVIDIA CUDA Toolkit Release Notes</a> .  <b>Note:</b> CUDA is installed from the ISO only on DGX Station systems, including DGX Station A100.
Docker Engine	<b>20.10.16</b>		Refer to <a href="#">v20.10.11-3</a> .
NVIDIA Container Toolkit	nvidia-container-runtime: 2.8.0-1 nvidia-container-toolkit: 1.7.0-1 nvidia-docker2: 2.8.0-1 libnvidia-container1: 1.7.0-1 libnvidia-container-tools: 1.7.0-1		Refer to the <a href="#">NVIDIA Container Toolkit</a> documentation.

Component	Release with R450	Release with R470	Additional Information
NVSM	<b>22.03.05</b>		Refer to the <a href="#">NVIDIA System Management Documentation</a> .
DCGM	<b>2.3.6</b>		Refer to the <a href="#">DCGM Release Notes</a> .
NGC CLI	2.2.0-1		Refer to the <a href="#">NGC CLI Documentation</a>
Mellanox OFED	MLNX 5.4-3.1.0.0		Refer to <a href="#">MLNX OFED v5.4-1.0.3.0</a>
MIG Configuration Tool	<b>0.1.2-1</b>		Refer to the following NVIDIA mig-parted github pages: <a href="https://github.com/NVIDIA/mig-parted">https://github.com/NVIDIA/mig-parted</a> and <a href="https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd">https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd</a>
nvipmitool	1.0.6.0		
nvidia-peer-memory/ nvidia-peer-memory DKMS	<b>1.3.0</b>		

## 2.6.4. Update: May 17, 2022

- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **NVIDIA GPU R470 Driver:** 470.129.06
  - ▶ **NVIDIA GPU R450 Driver:** 450.191.01

## 2.6.5. Update: April 28, 2022



**Important:** In DGX OS 5.3, the GPG keys that are used to sign the packages and metadata in those repositories need to be rotated. Refer to [Rotating the GPG Keys](#) for more information.

## 2.6.6. Update: February 17, 2022

- ▶ Installer version updated to 5.2.0.
- ▶ Added DGX Software Stack installation method


The DGX Software Stack provides the option to install a vanilla version of Ubuntu 20.04 and then separately install the additional NVIDIA software (NVIDIA DGX Software Stack). This option is available for DGX servers (DGX A100, DGX-2, DGX-1). The DGX Software Stack is a stream-lined version of the software stack incorporated into the DGX OS ISO image, and

includes meta-packages to simplify the installation process. Refer to the [DGX Software Stack for Ubuntu Installation Guide \(http://docs.nvidia.com/dgx/dgx-ubuntu-install-guide/index.html\)](http://docs.nvidia.com/dgx/dgx-ubuntu-install-guide/index.html) for instructions.

- The following changes were made to the Ubuntu repositories:

- **R470 NVIDIA GPU Driver:** 470.103.01

- **R450 NVIDIA GPU Driver:** 470.172.01

-  **Note:** When upgrading DGX OS, the system remains on the installed GPU driver branch. For example, the GPU driver branch on the system does not automatically switch from R450 to R470. Refer to the [Changing Your GPU Branch](#) section of the DGX OS User Guide for instructions on switching GPU driver branches.

- The following changes were made to the Ubuntu repositories:

- **DCGM:** 2.3.2


- **NVSM:** 21.09.14

- **Docker CE:** 20.10.11

- **nvidia-peer-memory/nvidia-peer-memory DKMS:** 1.3.0

- The **DGX OS 5.2.0 ISO** has been released.

Here are the contents of the DGX OS 5.2.0 ISO

Component	Version	Additional Information
Ubuntu	20.04 LTS	Refer to the <a href="#">Ubuntu 20.04 Desktop Guide</a> .
Ubuntu kernel	<b>5.4.0-xx-generic</b>	See <a href="#">Linux 5.4.0-80.90</a> .
GPU Driver	<b>R450:</b> 450.172.01 <b>R470:</b> 470.103.01 <div>  <b>Note:</b> Updating from R450 to R470 does not happen automatically when updating DGX OS 5, but requires separate steps. Refer to the <a href="#">Changing Your GPU Branch</a> section of the DGX OS User Guide for instructions. </div>	Refer to the <a href="#">NVIDIA Data Center GPU documentation</a> .
CUDA Toolkit	<b>11.4</b>	Refer to the <a href="#">NVIDIA CUDA Toolkit Release Notes</a> .  <b>Note:</b> CUDA is installed from the ISO only on DGX Station systems, including DGX Station A100.
Docker Engine	<b>20.10.11</b>	Refer to <a href="#">v20.10.11</a> .

Component	Version	Additional Information
NVIDIA Container Toolkit	nvidia-container-runtime: 3.5.0-1 nvidia-container-toolkit: 1.7.0-1 nvidia-docker2: 2.8.0-1 libnvidia-container1: 1.7.0-1 libnvidia-container-tools: 1.7.0-1	Refer to the <a href="#">NVIDIA Container Toolkit</a> documentation.
NVSM	<b>21.09.14</b>	Refer to the <a href="#">NVIDIA System Management Documentation</a> .
DCGM	<b>2.3.2</b>	Refer to the <a href="#">DCGM Release Notes</a> .
NGC CLI	2.2.0	Refer to the <a href="#">NGC CLI Documentation</a>
Mellanox OFED	MLNX 5.4-1.0.3.0	Refer to <a href="#">MLNX OFED v5.4-1.0.3.0</a>
MIG Configuration Tool	<b>0.1.2-1</b>	Refer to the following NVIDIA mig-parted github pages: <a href="https://github.com/NVIDIA/mig-parted">https://github.com/NVIDIA/mig-parted</a> and <a href="https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd">https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd</a>
nvipmitool	1.0.60	
nvidia-peer-memory/nvidia-peer-memory DKMS	<b>1.3.0</b>	

## 2.6.7. Update: December 14, 2021

- ▶ Installer version updated to 5.1.1.
- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **R470 NVIDIA GPU Driver:** 470.82.01
- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **DCGM:** 2.3.1
  - ▶ **NVSM:** 21.09.10
  - ▶ **MOFED:** MLNX 5.4-3.1.0.0
  - ▶ **Docker CE:** 20.10.11
  - ▶ **nvidia-container stack:**
    - ▶ nvidia-docker2-2.8.0-1
    - ▶ nvidia-container-runtime-3.7.0-1

nvidia-container-toolkit-1.7.0-1  
 libnvidia-container-tools-1.7.0-1  
 libnvidia-container1-1.7.0-1

- ▶ **nvipmitool:** 1.0.6.0
- ▶ **nvidia-peer-memory/nvidia-peer-memory DKMS:** 1.2.0

## 2.6.8. Update: October 26 , 2021

- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **NVIDIA GPU Driver:** 450.156.00

## 2.6.9. DGX OS 5.1 Release: August 26, 2021

- ▶ The following updates were made to the Ubuntu repositories
  - ▶ Docker Engine: 20.10.7
  - ▶ NVSM: 21.07.15
  - ▶ DCGM: 2.2.9
  - ▶ nvidia-container-runtime: 3.5.0-1
  - ▶ NVIDIA MLNX\_OFED: 5.4-1.0.3.0
  - ▶ (New) NGC CLI: 2.2.0
  - ▶ (New) MIG Configuration Tool: 0.1.2-1
- ▶ The following changes were made to the Ubuntu repositories
  - ▶ Added the release 470 GPU Driver: 470.57.02




**Note:** When upgrading DGX OS, the system remains on the installed GPU driver branch. For example, the GPU driver branch on the system does not automatically switch from R450 to R470. Refer to the [Changing Your GPU Branch](#) section of the DGX OS User Guide for instructions on switching GPU driver branches.

- ▶ The **DGX OS 5.1.0 ISO** has been released.

Here are the contents of the DGX OS 5.1.0 ISO

Component	Version	Additional Information
Ubuntu	20.04 LTS	Refer to the <a href="#">Ubuntu 20.04 Desktop Guide</a> .
Ubuntu kernel	<b>5.4.0-81</b>	See <a href="#">Linux 5.4.0-80.90</a> .
GPU Driver	<b>R450:</b> 450.142.00	Refer to the <a href="#">NVIDIA Data Center GPU documentation</a> .

Component	Version	Additional Information
	<b>R470:</b> 470.57.02  <div>  <b>Note:</b> Updating from R450 to R470 does not happen automatically when updating DGX OS 5, but requires separate steps. Refer to the <a href="#">Changing Your GPU Branch</a> section of the DGX OS User Guide for instructions. </div>	
CUDA Toolkit	<b>11.4</b>	Refer to the <a href="#">NVIDIA CUDA Toolkit Release Notes</a> .  <b>Note:</b> CUDA is installed from the ISO only on DGX Station systems, including DGX Station A100.
Docker Engine	<b>20.10.7</b>	Refer to <a href="#">v20.10.7</a> .
NVIDIA Container Toolkit	nvidia-container-runtime: <b>3.5.0-1</b>  nvidia-container-toolkit: <b>1.5.1-1</b>  nvidia-docker2: <b>2 6.0-1</b>  libnvidia-container1: <b>1.4.0-1</b>  libnvidia-container-tools: <b>1.4.0-1</b>	Refer to the <a href="#">NVIDIA Container Toolkit</a> documentation.
NVSM	<b>21.07.15</b>	Refer to the <a href="#">NVIDIA System Management Documentation</a> .
DCGM	<b>2.2.9</b>	Refer to the <a href="#">DCGM Release Notes</a> .
NGC CLI	<b>2.2.0</b>	Refer to the <a href="#">NGC CLI Documentation</a>
Mellanox OFED	<b>MLNX 5.4-1.0.3.0</b>	Refer to <a href="#">MLNX_OFED v5.4-1.0.3.0</a>
MIG Configuration Tool	<b>0.1.2-1</b>	Refer to the following NVIDIA mig-parted github pages: <a href="https://github.com/NVIDIA/mig-parted">https://github.com/NVIDIA/mig-parted</a> and <a href="https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd">https://github.com/NVIDIA/mig-parted/tree/master/deployments/systemd</a>

## 2.6.10. Update: June 30 , 2021

- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **GPUDirect Storage**: Added support for [GPUDirect Storage v1.0](#) release. GPUDirect Storage requires manual installation. For more information and installation instructions, refer to <https://docs.nvidia.com/gpudirect-storage/troubleshooting-guide/index.html#troubleshoot-install>.
  - ▶ **NVSM** for GPUDirect Storage: Updated to 21.03.11 only when installing GPUDirect Storage
  - ▶ **MOFED** for GPUDirectStorage: Updated to 5.3-1.0.5.0 only when installing GPUDirect Storage.

## 2.6.11. Update: June 20 , 2021

- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **NVIDIA GPU Driver**: 450.142.00

## 2.6.12. Update: June 2, 2021

- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **NVIDIA GPU Driver**: 450.119.04

These are signed drivers and replace the unsigned drivers provided in the Ubuntu repositories.

## 2.6.13. Update: May 27, 2021

- ▶ The following changes were made to the Ubuntu repositories:
  - ▶ **NVSM**: 20.09.26
  - ▶ **MOFED**: MLNX 5.1-2.6.2.0

Incorporates mlnx-fw-updater 5.2-1.0.4.0. When the update is made, the Mellanox FW updater updates the ConnectX card firmware as follows:

Card	Firmware Version
ConnectX-4	12.28.2006  To force a downgrade, see <a href="#">Downgrading Firmware for Mellanox ConnectX-4 Cards</a> for more information.
ConnectX-5	16.29.1016
ConnectX-6	20.29.1016

## 2.6.14. Update: May 06, 2021

The following change was made in the DGX repositories:



► **NVIDIA GPU Driver:** 450.119.04

Unsigned precompiled 450.119.04 kernel modules have been added to the DGX repository which provides a fix for issue [Driver Version Mismatch Reported](#). They will be removed once signed precompiled 450.119.04 kernel modules are provided by Canonical.



**Important:** Do not update if your system has Secure Boot enabled. Since these are unsigned drivers, systems with Secure Boot enabled will fail to load the drivers.

## 2.6.15. Update: April 20, 2021

The following change was made in the Ubuntu repositories:

► **NVIDIA GPU Driver:** 450.119.03

See known issue [Driver Version Mismatch Reported](#).

## 2.6.16. Update: April 13, 2021

The following changes were made to the Ubuntu repositories:

- **GPUDirect Storage:** Added support for [GPUDirect Storage](#) as a Technical Preview. GPUDirect Storage requires manual installation. For more information and installation instructions, refer to <https://docs.nvidia.com/gpudirect-storage/troubleshooting-guide/index.html#troubleshoot-install>.
- **MOFED:** Updated to MLNX 5.1-2.6.2.0



**Note:** There is no need to manually uninstall previous MOFED versions before getting this update.

## 2.6.17. Update: March 30, 2021

The following changes were made to the Ubuntu repositories:

► **MOFED:** MLNX 5.1-2.5.8.0.47

If you have already updated to the latest Ubuntu kernel (`uname -a` reports 5.4.0-67 or later), then you need to uninstall MOFED and then reinstall it as follows.

```
$ apt-get purge mlnx-ofed-all mlnx-ofed-kernel-dkms --auto-remove
$ apt-get update
$ apt-get install mlnx-ofed-all nvidia-peer-memory-dkms
```

## 2.6.18. Update: March 2, 2021

- The following changes were made to the Ubuntu repositories:
  - Added support for the DGX Station A100.
  - **DCGM:** 2.0.14
  - **NVSM:** 20.09.20

- The following ISO has been released:

- **DGX OS 5.0.2**

Supports the DGX Station A100.

Here are the contents of the DGX OS 5.0.2 ISO:

Component	Version	Additional Information
Ubuntu	<b>20.04.1 LTS</b>	Refer to the <a href="#">Ubuntu 20.04 Desktop Guide</a> and <a href="#">Ubuntu Server Guide</a> .
Ubuntu kernel	<b>5.4.0-58-generic</b>	See <a href="#">linux 5.4.0-58-generic</a> .
GPU Driver	<b>450.80.02</b>	Refer to the <a href="#">NVIDIA Tesla documentation</a> .
CUDA Toolkit	<b>11.0</b>	Refer to the <a href="#">NVIDIA CUDA Toolkit Release Notes</a> .  <b>Note:</b> CUDA is installed from the ISO only on DGX Station systems, including DGX Station A100.
Docker Engine	<b>19.03.14</b>	Refer to <a href="#">v19.03.14</a> .
NVIDIA Container Toolkit	libnvidia-container1: <b>1.3.0-1</b> libnvidia-container-tools: <b>1.3.0-1</b> nvidia-container-runtime: <b>3.4.0-1</b> nvidia-container-toolkit: <b>1.3.0-1</b> nvidia-docker: <b>2 2.5.0-1</b>	Refer to the <a href="#">NVIDIA Container Toolkit</a> documentation.
NVSM	<b>20.09.17</b>	Refer to the <a href="#">NVIDIA System Management Documentation</a> .
DCGM	<b>2.0.14</b>	Refer to the <a href="#">DCGM Release Notes</a> .
NVIDIA System Tools	<b>20.09-1</b>	
Mellanox OFED	<b>MLNX 5.1-2.5.8.0</b>	Refer to <a href="#">MLNX_OFED v5.1-2.5.8.0</a> .

## 2.6.19. Update: February 23, 2021

The following change was made to Ubuntu repositories:

- **NVSM:** 20.09.17

## 2.6.20. Update: January 20, 2021

The following change was made in the Ubuntu repositories:

- **NVIDIA GPU Driver:** 450.102.04

## 2.6.21. Update: December 11, 2020

The following changes were made in the Ubuntu repositories:

- **MOFED:** MLNX 5.1-2.5.8.0

When the update is made, the Mellanox FW updater updates the ConnectX card firmware as follows:

Card	Firmware Version
ConnectX-4	12.28.2006  To force a downgrade, see <a href="#">Downgrading Firmware for Mellanox ConnectX-4 Cards</a> for more information.
ConnectX-5	16.28.4000
ConnectX-6	20.28.4000

- **Docker:** docker-ce 19.03.14

This addresses [CVE-2020-15257](#).

## 2.6.22. DGX OS 5.0 Release: October 31, 2020

DGX OS 5.0 was released with the DGX OS 5.0.0 ISO. Here are the contents of the DGX OS 5.0.0 ISO:

Component	Version	Additional Information
Ubuntu	<b>20.04</b> LTS	Refer to the <a href="#">Ubuntu 20.04 Desktop Guide</a> .
Ubuntu kernel	<b>5.4.0-52-generic</b>	See <a href="#">linux 5.4.0-52-generic</a> .
GPU Driver	<b>450.80.02</b>	Refer to the <a href="#">NVIDIA Tesla documentation</a> .
CUDA Toolkit	<b>11.0</b>	Refer to the <a href="#">NVIDIA CUDA Toolkit Release Notes</a> .  <b>Note:</b> CUDA is installed from the ISO only on DGX Station systems, including DGX Station A100.
Docker Engine	<b>19.03.13</b>	Refer to <a href="#">v19.03.13</a> .
NVIDIA Container Toolkit	libnvidia-container1: <b>1.3.0-1</b> libnvidia-container-tools: <b>1.3.0-1</b> nvidia-container-runtime: <b>3.4.0-1</b> nvidia-container-toolkit: <b>1.3.0-1</b> nvidia-docker: <b>2 2.5.0-1</b>	Refer to the <a href="#">NVIDIA Container Toolkit</a> documentation.

Component	Version	Additional Information
NVSM	20.07.40	Refer to the <a href="#">NVIDIA System Management Documentation</a> .
DCGM	2.0.13	Refer to the <a href="#">DCGM Release Notes</a> .
NVIDIA System Tools	20.09-1	
Mellanox OFED	MLNX 5.1-2.4.6.0	

---

# Chapter 3. Known Issues Summary

This section provides summaries of the issues in DGX OS 5

## **Known Issues for DGX Server:**

- ▶ [Regression of CUDA application startup performance](#)
- ▶ [Unsupported Installation Options Appear in the ISO GRUB Menu](#)
- ▶ [EFI Boot Manager Lists Ubuntu as a Boot Option](#)
- ▶ [Duplicate EFI Variable May Cause efibootmgr to Fail](#)
- ▶ [Erroneous Insufficient Power Error May Occur for PCIe Slots](#)
- ▶ [AMD Crypto Co-processor is not Supported](#)
- ▶ [nvsm show alerts Reports NVSwitch PCIe Link Width is Degraded](#)
- ▶ [nvsm show health Reports Firmware as Not Authenticated](#)
- ▶ [Running NGC Containers Older than 20.10 May Produce “Incompatible MOFED Driver” Message](#)
- ▶ [System May Slow Down When Using mpirun](#)
- ▶ [Software Power Cap Not Reported Correctly by nvidia-smi](#)
- ▶ [Forced Reboot Hangs the OS](#)
- ▶ [Applications that call the cuCTXCreate API Might Experience a Performance Drop](#)

## **Known Issues for DGX Station:**

- ▶ [Applications that call the cuCTXCreate API Might Experience a Performance Drop](#)
- ▶ [NVIDIA Desktop Shortcuts Not Updated After a DGX OS Release Upgrade](#)

## **Known Issues for DGX Station A100:**

- ▶ [Unable to Set a Separate/xinerama Mode through the xorg.conf File or through nvidia-settings](#)

## **Known Limitations (Issues that will not be fixed)**

- ▶ [No RAID Partition Created After ISO Install](#)
- ▶ [NSCQ Library and Fabric Manager Might Not Install When Installing a New NVIDIA Driver](#)
- ▶ [System Services Startup Messages Appear Upon Completion of First-Boot Setup](#)

- ▶ [\[DGX A100\]: Hot-plugging of Storage Drives is not Supported](#)
- ▶ [\[DGX A100\]: Syslog Contains Numerous "SM LID is 0, maybe no SM is running" Error Messages](#)
- ▶ [\[DGX-2\] Serial Over LAN Does not Work After Cold Resetting the BMC](#)
- ▶ [\[DGX-2\] Some BMC Dashboard Quick Links Appear Erroneously](#)
- ▶ [\[DGX-2\] Applications Cannot be Run Immediately Upon Powering on the DGX-2](#)
- ▶ [\[DGX-1\] Script Cannot Recreate RAID Array After Re-inserting a Known Good SSD](#)
- ▶ [\[DGX Station A100\] Suspend and Power Button Section Appears in Power Settings](#)
- ▶ [\[DGX-2\] NVSM Does not Detect Downgraded GPU PCIe Link](#)

## Resolve Issues

See [DGX OS Resolved Issues Details](#).

## 3.1. Known Issues Details

This section provides details for known issues in DGX OS 5.x.

### 3.1.1. Regression of CUDA application startup performance

#### Issue

Reported in 5.4.0 CUDA applications may experience longer CUDA binary loading.

#### Explanation

The CUBIN/FATBINARY loading execution time may be increased by up to ~15% with certain drivers and CUDA versions (experienced with R510 and CUDA 11.6). This impacts all CUDA module loading APIs, `cuModuleLoad*`, as well as CUDA modules loaded through the CUDA Runtime (CUDART). The issue is not expected to have an impact on the application once the modules are loaded.

### 3.1.2. NVSM Stress Test Logs Do Not Contain Summary Information

#### Issue

When you run an NVSM stress test, the log does not include the test summary.

#### Explanation

This issue currently under investigation.

### 3.1.3. Unsupported Installation Options Appear in the ISO GRUB Menu

#### Issue

When installing DGX OS from the ISO, selecting one of the following GRUB menu options results in errors beginning with `error : No such file or directory`.

- ▶ Install DGX OS 5.x.0 Without MLNX Drivers
- ▶ Install DGX OS 5.x.0 Without Nvidia Drivers

#### Explanation

Currently, the options **Without MLNX Drivers** and **Without NVIDIA Drivers** are not support and should not be selected.

### 3.1.4. nvidia-release-upgrade May Report That Not All Updates Have Been Installed and Exit

#### Issue

When running the `nvidia-release-upgrade` command on systems running DGX OS 4.99.x, it may exit and tell users: "Please install all available updates for your release before upgrading" even though all upgrades have been installed.

#### Explanation

To recover, issue the following command:

```
sudo apt install -y nvidia-fabricmanager-450/bionic-updates --allow-downgrades
```

After running the command, proceed with the regular upgrade steps:

```
sudo apt update
sudo apt full-upgrade -y
sudo apt install -y nvidia-release-upgrade
sudo nvidia-release-upgrade
```

### 3.1.5. EFI Boot Manager Lists Ubuntu as a Boot Option

#### Issue

Reported in release 5.1.0.

The GRUB menu may list "ubuntu" as a boot option in addition to the "DGX OS" boot option beginning with DGX OS 5.1.0. Either may be used to boot the DGX system to the DGX OS.

### Explanation

The "ubuntu" option will be removed in a future software release.

## 3.1.6. Duplicate EFI Variable May Cause efibootmgr to Fail

### Issue

Reported in release 5.1.0.

On some DGX-2 systems, the 'efibootmgr' command may fail with the following signature:

```
$ sudo efibootmgr
No BootOrder is set; firmware will attempt recovery
```

### Explanation

This happens when the SBIOS presents duplicate EFI variables. Because of this, `efivarfs` will not be fully populated which may ultimately cause `efibootmgr` to fail.

To work around:

1. Flash the BIOS with the latest SBIOS revision using the BMC.

Refer to <https://docs.nvidia.com/dgx/dgx2-fw-container-release-notes/sbios-update-from-bmc-ui.html#sbios-update-from-bmc-ui> for instructions.



**Important:** Do not power cycle the system after clicking **Cancel** at the **Firmware update completed** dialog.

2. From the command line, issue the following command to read the "Restore PLDM Flag".

```
$ sudo ipmitool raw 0x03 0x0D
```

This flag is cleared after reading, meaning that the system will not restore the PLDM table after the subsequent power cycle.

3. Power-cycle the system.

## 3.1.7. Erroneous Insufficient Power Error May Occur for PCIe Slots

### Issue

Reported in release 4.99.9.

The DGX A100 server reports "Insufficient power" on PCIe slots when network cables are connected.



### Explanation

This may occur with optical cables and indicates that the calculated power of the card + 2 optical cables is higher than what the PCIe slot can provide.

The message can be ignored.

## 3.1.8. AMD Crypto Co-processor is not Supported

### Issue

Reported in release 4.99.9.

The DGX A100 currently does not support the AMD Cryptograph Co-processor. When booting the system, you may see the following error message in the syslog:

```
ccp initialization failed
```

### Explanation

Even if the message does not appear, CCP is still not supported. The SBIOS makes zero CCP queues available to the driver, so CCP cannot be activated.

## 3.1.9. `nvsm show alerts` Reports NVSwitch PCIe Link Width is Degraded

### Issue

Reported in release 4.99.10.

NVSM raises alerts of Severity=Warning against PCIe links between NVSwitch and the Draco switch. The alert states "PCIe link width degraded" - the PCIe link width is expected to be x4 while the actual link width is x2.

There are six pairs of the PCIe links, so NVSM raises six such alerts in this condition.

### Explanation

The Broadcom firmware for the synthetic switch advertises the Draco switch has PCIe link width capability of x4. This synthesized information is not reflecting the hardware capability which is of width x2. NVSM raises alerts based on this incorrect information.

This issue will be resolved with updated firmware to be provided in the DGX A100 Firmware Update Container after version 20.05.12.3. See the [DGX A100 Firmware Update Container release notes](#) for the latest firmware status.

### 3.1.10. `nvsm show health` Reports Firmware as Not Authenticated

#### Issue

Reported in release 5.0.

When issuing `nvsm show health`, the output shows CEC firmware components as Not Authenticated, even when they have passed authentication.

Example:

```
CEC:
  CEC Version: 3.5
  EC_FW_TAG0: Not Authenticated
  EC_FW_TAG1: Not Authenticated
  BMC_FW authentication state: Not Authenticated
```

#### Explanation

The message can be ignored and does not affect the overall `nvsm health` output status.

### 3.1.11. Running NGC Containers Older than 20.10 May Produce “Incompatible MOFED Driver” Message

#### Issue

Reported in release 5.0.

DGX OS 5.0 incorporates Mellanox OFED 5.1 for high performance multi-node connectivity. Support for this version of OFED was added in NGC containers 20.10, so when running on earlier versions (or containers derived from earlier versions), a message similar to the following may appear.

```
ERROR: Detected MOFED driver 5.1-2.4.6, but this container has version 4.6-1.0.1.
  Unable to automatically upgrade this container.
  Multi-node communication may be unreliable or may result in crashes with this
  version.
  This incompatibility will be resolved in an upcoming release. .
```

#### Explanation

For applications that rely on OFED (typically those used in multi-node jobs), this is an indication that an update to NGC containers 20.10 or greater is required. For most other applications, this error can be ignored.

Some applications may return an error such as the following when running with NCCL debug messages enabled:

```
export NCCL_DEBUG=WARN
```

```
misc/ibvwrap.cc:284 NCCL WARN Call to ibv_modify_qp failed with error No such device
...
common.cu:777 'unhandled system error'
```

This may occur even for single-node training jobs. To work around this, issue the following:

```
export NCCL_IB_DISABLE=1
```

### 3.1.12. System May Slow Down When Using `mpirun`

#### Issue

Customers running Message Passing Interface (MPI) workloads may experience the OS becoming very slow to respond. When this occurs, a log message similar to the following would appear in the kernel log:

```
kernel BUG at /build/linux-fQ94TU/linux-4.4.0/fs/ext4/inode.c:1899!
```

#### Explanation

Due to the current design of the Linux kernel, the condition may be triggered when `get_user_pages` is used on a file that is on persistent storage. For example, this can happen when `cudaHostRegister` is used on a file path that is stored in an ext4 filesystem. DGX systems implement `/tmp` on a persistent ext4 filesystem.



**Note:** If you performed this workaround on a previous DGX OS software version, you do not need to do it again after updating to the latest DGX OS version.

In order to avoid using persistent storage, MPI can be configured to use shared memory at `/dev/shm` (this is a temporary filesystem).

If you are using Open MPI, then you can solve the issue by configuring the Modular Component Architecture (MCA) parameters so that `mpirun` uses the temporary file system in memory.

For details on how to accomplish this, see the Knowledge Base Article [DGX System Slows Down When Using `mpirun`](#) (requires login to the [NVIDIA Enterprise Support](#) portal).

### 3.1.13. Software Power Cap Not Reported Correctly by `nvidia-smi`

#### Issue

On DGX-1 systems with Pascal GPUs, `nvidia-smi` does not report Software Power Cap as "Active" when clocks are throttled by power draw.

#### Explanation

This issue is with `nvidia-smi` reporting and not with the actual functionality.

### 3.1.14. Forced Reboot Hangs the OS

#### Issue

When issuing `reboot -f` (forced reboot), I/O error messages appear on the console and then the system hangs.

The system reboots normally when issuing `reboot`.

#### Explanation

This issue will be resolved in a future version of the DGX OS.

### 3.1.15. Applications that call the cuCTXCreate API Might Experience a Performance Drop

#### Issue

Reported in release 5.0.

When some applications call `cuCtxCreate`, `cuGLCtxCreate`, or `cuCtxDestroy`, there might be a drop in performance.

#### Explanation

This issue occurs with Ubuntu 20.04, but not with previous versions. The issue affects applications that perform graphics/compute interoperations or have a plugin mechanism for CUDA, where every plugin creates its own context, or video streaming applications where computations are needed. Examples include `ffmpeg`, `Blender`, `simpleDrvRuntime`, and `cuSolverSp_LinearSolver`.

This issue is not expected to impact deep learning training.

### 3.1.16. NVIDIA Desktop Shortcuts Not Updated After a DGX OS Release Upgrade

#### Issue

Reported in release 4.0.4.

In DGX OS 4 releases, the NVIDIA desktop shortcuts have been updated to reflect current information about NVIDIA DGX systems and containers for deep learning frameworks. These desktop shortcuts are also organized in a single folder on the desktop.

After a DGX OS release upgrade, the NVIDIA desktop shortcuts for existing users are not updated. However, the desktop for a user added after the upgrade will have the current desktop shortcuts in a single folder.

## Explanation

If you want quick access to current information about NVIDIA DGX systems and containers from your desktop, replace the old desktop shortcuts with the new desktop shortcuts.

1. Change to your desktop directory.

```
$ cd /home/your-user-login-id/Desktop
```

2. Remove the existing NVIDIA desktop shortcuts.

```
$ rm dgx-container-registry.desktop \
  dgxstation-userguide.desktop \
  dgx-container-registry-userguide.desktop \
  nvidia-customer-support.desktop
```

3. Copy the folder that contains the new NVIDIA desktop shortcuts and its contents to your desktop directory.

```
$ cp -rf /etc/skel/Desktop/Getting\ Started/ .
```

## 3.1.17. Unable to Set a Separate/xinerama Mode through the xorg.conf File or through nvidia-settings

### Issue

Reported in release 5.0.2

In Station A100, in the BIOS, in OnBrd/Ext VGA Select=, when **Auto** or **External** is selected, the nvidia-conf-xconfig service sets up Xorg to use only the Display adapter.

### Explanation

Manually edit the existing the /etc/X11/xorg.conf.d/xorg-nvidia.conf file with the following settings:

```
--- xorg-nvidia.conf      2020-12-10 02:42:25.585721167 +0530
+++ /root/working-xinerama-xorg-nvidia.conf      2020-12-10 02:38:05.368218170 +0530
@@ -8,8 +8,10 @@
Section "ServerLayout"
    Identifier      "Layout0"
    Screen          0  "Screen0"
+   Screen          1  "Screen0 (1)" RightOf "Screen0"
    InputDevice     "Keyboard0" "CoreKeyboard"
    InputDevice     "Mouse0" "CorePointer"
+   Option          "Xinerama" "1"
EndSection

Section "Files"
@@ -43,6 +45,7 @@
    Driver          "nvidia"
    BusID           "PCI:2:0:0"
    VendorName      "NVIDIA Corporation"
```

```

+   Screen          0
+   EndSection

+   Section "Screen"
@@ -51,6 +54,25 @@
+   Monitor          "Monitor0"
+   DefaultDepth     24
+   Option           "AllowEmptyInitialConfiguration" "True"
+   SubSection       "Display"
+   Depth            24
+   EndSubSection
+EndSection
+
+Section "Device"
+ Identifier         "Device0 (1)"
+ Driver             "nvidia"
+ BusID              "PCI:2:0:0"
+ VendorName         "NVIDIA Corporation"
+ Screen             1
+EndSection
+
+Section "Screen"
+ Identifier         "Screen0 (1)"
+ Device             "Device0 (1)"
+ Monitor            "Monitor0"
+ DefaultDepth       24
+ Option             "AllowEmptyInitialConfiguration" "True"
+ SubSection         "Display"
+   Depth            24
+EndSubSection

```

## 3.2. DGX OS Resolved Issues Details

Here are the issues that are resolved in the latest release.

- ▶ [DGX A100/DGX-2] [Driver Version Mismatch Reported](#)
- ▶ [DGX A100] [BMC is not Detectable After Restoring BMC to Default](#)
- ▶ [DGX A100] [A System with Encrypted rootfs May Fail to Boot if one of the M.2 drives is Corrupted](#)
- ▶ [All DGX systems]: When starting the DCGM service, a version mismatch error message similar to the following will appear:
 

```
[78075.772392] nvidia-nvswitch: Version mismatch, kernel version 450.80.02 user version 450.51.06
```
- ▶ [All DGX systems]: When issuing `nvsm show health`, the `nvsmhealth_log.txt` log file reports that the `/proc/driver/` folders are empty.
- ▶ [DGX A100]: The Mellanox software that is included in the DGX OS installed on DGX A100 system does not automatically update the Mellanox firmware as needed when the Mellanox driver is installed.
- ▶ [DGX A100]: `nvsm stress-test` does not stress the system if MIG is enabled.  
Reported in 4.99.10
- ▶ [DGX A100]: With eight U.2 NVMe drives installed, the `nvsm-plugin-pcie` service reports "ERROR: Device not found in mapping table" for the additional four drives (for example, in response to `systemctl status nvsm*`).  
Reported in 4.99.11

- [DGX A100]: When starting the Fabric Manager service, the following error is reported: detected NVSwitch non-fatal error 10003 on NVSwitch pci.

Reported in 4.99.9

### 3.2.1. NVSM Platform Displays as Unsupported

#### Issue

Reported in release 5.0.

In DGX Station, when you run

```
$ nvsm show version
```

instead of displaying **DGX Station**, the **platform** field displays **Unsupported**.

#### Explanation

You can ignore this message.

### 3.2.2. cudaMemFree CUDA API Performance Regression

#### Issue

Reported in release 4.99.10.

In cases when NVLINK peers are enabled, there is a performance regression of cuMemFree CUDA API.

#### Explanation

The cuMemFree API is usually used during application teardown and is discouraged from being used in performance-critical paths, so the regression should not impact application end-to-end performance.

### 3.2.3. NVSM Enumerates NVSwitches as 8-13 Instead of 0-5

#### Issue

Reported in release 4.99.9. Fixed in release 5.1

NVSM commands that list the NVSwitches (such as `nvsm show nvswitches`) will return the switches with 8-13 enumeration.

Example:

```
nvsm show /systems/localhost/nvswitches
/systems/localhost/nvswitches
```

```
Targets:
NVSwitch10
NVSwitch11
NVSwitch12
NVSwitch13
NVSwitch8
NVSwitch9
```

### Explanation

Currently, NVSM recognizes NVSwitches as graphics devices, and enumerates them as a continuation of the GPU 0-7 enumeration.

## 3.2.4. BMC is not Detectable After Restoring BMC to Default

### Issue

Reported in release 4.99.8. Fixed with BMC 0.13.06

After using the BMC Web UI dashboard to restore the factory defaults (**Maintenance > Restore Factory Defaults**), the BMC can no longer be detected and the system is rendered unusable.

### Explanation

Do not attempt to restore the factory defaults using the BMC Web UI dashboard.

## 3.2.5. A System with Encrypted rootfs May Fail to Boot if one of the M.2 drives is Corrupted

### Issue

Reported in release 4.99.9. Fixed in 5.0.2.

On systems with encrypted rootfs, if one of the M.2 drives is corrupted, the system stops at the BusyBox shell when booting.

### Explanation

The inactive RAID array (due to the corrupted M.2 drive) is not getting converted to a degraded RAID array.

To work around, perform the following within the BusyBox.

1. Issue the following.

```
$ mdadm --run /dev/md?*
```

2. Wait a few seconds for the RAID and crypt to be discovered.
3. Exit.

```
$ exit
```



### 3.2.6. NVSM Fails to Show CPU Information on Non-English Locales

#### Issue

Reported in release 4.1.0 and 5.0 update 3

If the locale is other than English, the `nvsm show cpu` command reports the target processor does not exist.

```
$ sudo nvsm show cpu
ERROR:nvsm:Not Found for target address /systems/localhost/processors
ERROR:nvsm:Target address "/systems/*/processors/*" does not exist
```

#### Explanation

To work around, set the locale to English before issuing `nvsm show cpu`.

### 3.2.7. Driver Version Mismatch Reported

#### Issue

Reported in release 5.0: 4/20/21 update

Fixed in 5/06/21 update.

After updating the DGX OS, the `syslog/dmesg` reports the following version mismatch:

```
nvidia-nvswitch: Version mismatch, kernel version 450.119.03 user version 450.51.06
```

#### Explanation

This occurs with driver 450.119.03 on NVSwitch systems such as DGX -2 or DGX A100, and is due to a bug that causes the NSCQ library to fail to load. This will be resolved in an updated driver version.

## 3.3. Known Limitations Details

This section lists details for known limitations and other issues that will not be fixed.

### 3.3.1. No RAID Partition Created After ISO Install

#### Issue

After using the DGX OS ISO to install the DGX OS, there is no `/raid` partition created.

### Explanation

This occurs if you reboot the system right after the installation is completed. To create the data RAID, the DGX OS installer sets up a `systemd` service to create the `/raid` partition on first boot. If you reboot before you give that service a chance to finish, the `/raid` partition may not be properly set up.

To create the `/raid` partition, issue the following.

```
$ sudo configure RAID array.py -c -f
```

## 3.3.2. NSCQ Library and Fabric Manager Might Not Install When Installing a New NVIDIA Driver

### Issue

When you install a new NVIDIA Driver from the Ubuntu repository, the NSCQ library and Fabric Manager might not install.

### Explanation

The `libnvidia-nscq-xxx` packages provide the same `/usr/lib/x86_64-linux-gnu/libnvidia-nscq.so` file, so multiple packages cannot exist on your DGX system at the same time.

We recommend that you remove the old packages **before** installing the new driver branch. Refer to [Upgrading your NVIDIA Data Center GPU Driver to a Newer Branch](#) for instructions.

## 3.3.3. System Services Startup Messages Appear Upon Completion of First-Boot Setup

### Issue

After completing the first-boot setup process and getting to the login prompt, system services startup messages appear.

### Explanation

Some services cannot be started until after the initial configuration process is completed. Starting the services at the Ubuntu prompt avoids the need for an additional reboot to complete the setup process.

Once completed, the service messages do not appear at subsequent system reboots.

### 3.3.4. [DGX A100]: Hot-plugging of Storage Drives not Supported

#### Issue

Hot-plugging or hot-swapping one of the storage drives might result in system instability or incorrect device reporting.

#### Explanation and Workaround

Turn off the system before removing and replacing any of the storage drives.

### 3.3.5. [DGX A100]: Syslog Contains Numerous "SM LID is 0, maybe no SM is running" Error Messages

#### Issue

The system log (/var/log/syslog) contains multiple "SM LID is 0, maybe no SM is running" error message entries..

#### Explanation and Workaround

This issue is the result of the `srp_daemon` within the Mellanox driver. The daemon is used to discover and connect to InfiniBand SCSI RDMA Protocol (SRP) targets.

If you are not using RDMA, then disable the `srp_daemon` as follows.

```
$ sudo systemctl disable srp_daemon.service $ sudo systemctl disable
srptools.service
```

### 3.3.6. [DGX-2]: Serial Over LAN Does not Work After Cold Resetting the BMC

#### Issue

After performing a cold reset on the BMC (`ipmitool mc reset cold`) while serial over LAN (SOL) is active, you cannot restart the SOL session.

#### Explanation and Workaround

To re-active SOL, either

- ▶ Reboot the system, or
- ▶ Kill and then restart the process as follows.

1. Identify the Process ID of the SOL TTY process by running the following.

```
ps -ef | grep "/sbin/agetty -o -p -- \u --keep-baud 115200,38400,9600 ttyS0 vt220"
```

2. Kill the process.

```
kill <PID>
```

where <PID> is the Process ID returned by the previous command.

3. Either wait for the cron job to respawn the process or manually restart the process by running

```
/sbin/agetty -o -p -- \u --keep-baud 115200,38400,9600 ttyS0 vt220
```

### 3.3.7. [DGX-2]: Some BMC Dashboard Quick Links Appear Erroneously

#### Issue

On the BMC dashboard, the following Quick Links appear by mistake and should not be used.

- **Maintenance->Firmware Update**
- **Settings->NvMeManagement->NvMe P3700Vpd Info**

To recreate the array in this case,

1. Set the drive back to a good state.

```
# sudo /opt/MegaRAID/storcli/storcli64 /c0/e<enclosure_id>/s<drive_slot> set good
```

2. Run the script to recreate the array.

```
# sudo /usr/bin/configure RAID array.py -c -f
```

### 3.3.8. [DGX-2]: Applications Cannot be Run Immediately Upon Powering on the DGX-2

#### Issue

When attempting to run an application that uses the GPUs immediately upon powering on the DGX-2 system, you may encounter the following error.

```
CUDA_ERROR_SYSTEM_NOT_READY
```

#### Explanation and Workaround

The DGX-2 uses a fabric manager service to manage communication between all the GPUs in the system. When the DGX-2 system is powered on, the fabric manager initializes all the GPUs. This can take approximately 45 seconds. Until the GPUs are initialized, applications that attempt to use them will fail.

If you encounter the error, wait and launch the application again.

### 3.3.9. [DGX-1]: Script Cannot Recreate RAID Array After Re-inserting a Known Good SSD

#### Issue

When a good SSD is removed from the DGX-1 RAID 0 array and then re-inserted, the script to recreate the array fails.

#### Explanation and Workaround

After re-inserting the SSD back into the system, the RAID controller sets the array to offline and marks the re-inserted SSD as `Unconfigured_Bad` (`UBad`). The script will fail when attempting to rebuild an array when one or more of the SSDs are marked `UBad`.

To recreate the array in this case,

1. Set the drive back to a good state.

```
# sudo /opt/MegaRAID/storcli/storcli64 /c0/e<enclosure_id>/s<drive_slot> set good
```

2. Run the script to recreate the array.

```
# sudo /usr/bin/configure_raid_array.py -c -f
```

### 3.3.10. [DGX Station A100] Suspend and Power Button Section Appears in Power Settings

#### Issue

Reported in release 5.0.2.

In the Power Settings page of the DGX Station A100 GUI, the **Suspend & Power Button** section is displayed even though the options do not work.

#### Explanation

Suspend and sleep modes are not supported on the DGX Station A100.

### 3.3.11. [DGX-2] NVSM Does not Detect Downgraded GPU PCIe Link

#### Issue

If the GPU PCIe link is downgraded to Gen1, NVSM still reports the GPU health status as OK.

## Explanation

NVSM does not propagate the health status from the PCIe subsystem to other subsystems. For example: If there is a PCIe link degradation that is reported for Network Adapter, NVSM does not mark the Network adapter as unhealthy.

---

# Appendix A. Downgrading Firmware for Mellanox ConnectX-4 Cards

DGX OS 5.0.0 provides the `mlnx-fw-updater` package version 5.1-2.4.6.0 which automatically installs firmware version 12.28.2040 on ConnectX-4 devices.

Since 12.28.2006 is the recommended firmware version, on December 15 the updater package has been updated to install version 12.28.2006. However, if the firmware has already been updated to 12.28.2040, the updater will not install the downlevel firmware version since a newer version is already installed.

In this case, you will need to force the downgrade as explained in this section.

## A.1. Checking the Device Type

You can use the `mlxfwmanager` tool to verify whether ConnectX-4 devices are installed on your DGX system.

Run the following command.

```
:~$ sudo mlxfwmanager
Querying Mellanox devices firmware ...
Device #1:
-----
Device Type: ConnectX4
Part Number: MCX455A-ECA Ax
Description: ConnectX-4 VPI adapter card; EDR IB (100Gb/s) and
100GbE; single-port QSFP28; PCIe3.0 x16; ROHS R6
PSID: MT 2180110032
PCI Device Name: /dev/mst/mt4115_pciconf1
Base GUID: 248a070300945e60
Versions: Current Available
FW 12.28.2040 N/A
PXE 3.6.0102 N/A
UEFI 14.21.0017 N/A
```

## A.2. Downgrading the Firmware

If the output indicates that ConnectX-4 devices are installed, you need to downgrade the firmware.

To downgrade the firmware:

1. Determine the correct firmware package name.
  - a). Switch to the `/opt/Mellanox/mlnx-fw-updater/firmware` directory, where the updater installs the firmware files, and list the contents.

```
:/opt/mellanox/mlnx-fw-updater/firmware$ ls
```

- b). Identify the correct package from the output.

```
mlxfwmanager_sriov_dis_x86_64_4115 mlxfwmanager_sriov_dis_x86_64_4119
mlxfwmanager_sriov_dis_x86_64_4123 mlxfwmanager_sriov_dis_x86_64_4127
mlxfwmanager_sriov_dis_x86_64_41686 mlxfwmanager_sriov_dis_x86_64_4117
mlxfwmanager_sriov_dis_x86_64_4121 mlxfwmanager_sriov_dis_x86_64_4125
mlxfwmanager_sriov_dis_x86_64_41682
```

2. Execute the firmware package by using the `-f` flag.

```
:/opt/mellanox/mlnx-fw-updater/firmware$ sudo  
./mlxfwmanager_sriov_dis_x86_64_4115 -f
```

The software queries the current firmware and then updates (downgrades) the firmware.

```
Querying Mellanox devices firmware ...
...
-----
Found 2 device(s) requiring firmware update...
Device #1: Updating FW ...
  Initializing image partition - OK
  Writing Boot image component - OK
  Done
Device #2: Updating FW ...
  Initializing image partition - OK
  Writing Boot image component - OK
  Done
```

3. Reboot the system to allow the updates to take effect.

```
$ sudo reboot
```



---

## Appendix B. DGX Software Stack

This section lists the DGX software packages and kernel parameters in the DGX Software Stack.

### NVIDIA DGX Software Packages

This table lists all packages that are installed as part of the corresponding meta package:

DGX A100	DGX-2	DGX-1
<b><u>dgx-a100-system-configurations</u></b>	<b><u>dgx2-system-configurations</u></b>	<b><u>dgx1-system-configurations</u></b>
dgx-release	dgx-release	dgx-release
-	-	nv-ast-modeset
nvidia-crashdump	nvidia-crashdump	nvidia-crashdump
-	nv-enable-nvme-hot-plug	-
nv-hugepage	nv-hugepage	nv-hugepage
nv-iommu-pt	-	-
nv-ipmi-devintf	nv-ipmi-devintf	nv-ipmi-devintf
nv-limits	nv-limits	nv-limits
nv-update-disable	nv-update-disable	nv-update-disable
nvidia-acs-disable	nvidia-acs-disable	-
nvidia-kernel-defaults	nvidia-kernel-defaults	nvidia-kernel-defaults
nvidia-nvme-smartd	nvidia-nvme-smartd	-
nvidia-pci-bridge-power	nvidia-pci-bridge-power	nvidia-pci-bridge-power
nvidia-redfish-config	-	-
nvidia-relaxed-ordering-gpu	-	-
nvidia-relaxed-ordering-nvme	-	-
nvgpu-services-list	nvgpu-services-list	nvgpu-services-list

DGX A100	DGX-2	DGX-1
<b><u>dgx-a100-system-tools</u></b> dgx-release ipmitool nv-common-apis nv-env-paths nvidia-mig-manager nvidia-raid-config nvme-cli tpm2-tools	<b><u>dgx2-system-tools</u></b> dgx-release ipmitool nv-common-apis nv-env-paths - nvidia-raid-config nvme-cli tpm-tools	<b><u>dgx1-system-tools</u></b> dgx-release ipmitool nv-common-apis nv-env-paths - - - -
<b><u>dgx-a100-system-tools-extra</u></b> msecli	<b><u>dgx2-system-tools-extra</u></b> msecli	<b><u>dgx1-system-tools-extra</u></b> nvidia-raid-config storcli
<b><u>nvidia-mlnx-ofed-misc</u></b> mlnx-fw-updater mlnx-pxe-setup nvidia-mlnx-config nvidia-peer-memory   nvidia-peer-memory-dkms		
<b><u>Additional packages</u></b> nv-docker-options nvidia-logrotate nvidia-motd nvidia-ipmisol		

The following table lists all packages that will be installed as part of the system configuration package with more details:

Package	Description	1	2	A
dgx-release	Release information	R	R	R
nv-ast-modeset	Disable the Aspeed display driver. It can cause issues with connected monitors. The AST2xxx is the BMC used in our servers.  [DGX-1, DGX-2, DGX A100, DGX Station A100]	R	R	R

Package	Description	1	2	A
nv-enable-nvme-hot-plug	Configure kernel parameters for NVMe hot plug (see also kernel section below).		R	
nv-hugepage	Sets the "transparent_hugepage=madvise" kernel parameter.	R	R	R
nv-iommu-pt	Sets iommu=pt for AMD Rome platforms.			R
nv-ipmi-devintf	Add the ipmi_devintf module for accessing the BMC using the ipmi tool.	R	R	R
nv-limits	Increase the process resource limits for users (ulimits nofile 50000)	R	R	R
nv-update-disable	Disable automatic system upgrades. Users need to explicitly upgrade their systems using apt.	R	R	R
nvgpu-services-list	Lists GPU-consuming services in .json format, such as DCGM or NVSM, and required by the firmware update mechanism.	R	R	R
nvidia-acsc-disable	Disables the PCIe ACS capability to allow for better GPU-direct performance in bare-metal use cases on DGX A100.			R
nvidia-crashdump	Tools to manage kernel crash dumps. They are disabled by default.	R	R	R
nv-docker-options	Increases SHMEM and other resources.	R	R	R
nvidia-ipmisol [optional]	Enables serial output through the BMC (SOL - Serial over Lan)	O	O	O
nvidia-kernel-defaults	Disable ARP for security improvements net.ipv4.conf .all.arp_announce = 2 .all.arp_ignore = 1 .default.arp_announce = 2 .default.arp_ignore = 1	R	R	R
nvidia-logrotate	Modify the logrotate configuration	O	O	O
nvidia-motd	Modify message-of-the-day (MOTD) to display NVSM health monitoring alerts and release information.	O	O	O
nvidia-nvme-smartd	Enables SMART monitoring on NVME devices. By default, smartd will skip NVME devices.		R	R

Package	Description	1	2	A
nvidia-pci-bridge-power	Sets the bridge power control setting to “on” for all PCI bridges.	R	R	R
nvidia-relaxed-ordering-gpu	Sets a reg-key to enable PCIe relaxed-ordering in the GPUs			R
nvidia-relaxed-ordering-nvme	Installs a script that users can call to enable relaxed-order in NVME devices.			R
nvidia-redfish-config	Configures the redfish interface with an interface name and IP address. The interface name is “bmc_redfish0”, while the IP address is read from DMI type 42.			R

### Legend

- ▶ **1:** DGX-1
- ▶ **2:** DGX-2
- ▶ **A:** DGX A100
- ▶ **R:** Required package
- ▶ **O:** Optional package

### DGX Kernel Parameters

Kernel Parameter	Description	Package
ast.modeset=0	Disable the Aspeed display driver. The AST2xxx is the BMC used in our servers.  [DGX-1, DGX-2, DGX A100, DGX Station A100]	nv-ast-modeset
crashkernel=1G-:0M	Don't reserve any memory for crash dumps (when crash is disabled = default)	nvidia-crashdump
crashkernel=1G-:512M	Reserve 512MB for crash dumps (when crash is enabled)	nvidia-crashdump
pci=realloc=on	Allows kernel to reallocate PCI resources if allocations done by BIOS are insufficient.  This and pcie_ports=native are both required for NVME hot-plug on DGX2.	nv-enable-nvme-hot-plugth

Kernel Parameter	Description	Package
pcie_ports=native	Use Linux native services for PME, AER, DPC, PCIe hotplug. I.e. not firmware first.  This and pci=realloc=on are both required for NVME hot-plug on DGX2.	nv-enable-nvme-hot-plug
transparent_hugepage=madvise	Disable huge pages system-wide and only enable them inside MADV_HUGEPAGE madvise regions to prevent applications from allocating more memory resources than necessary.	nv-hugepage
iommu=pt	Enable pass through mode only and disable DMA translations. This enables optimizations for the CPU inside the DGX A100.	nv-iommu-pt
console=ttyS1,115200n8	Set console to serial port 1, using 115200 baud, no parity, 8 data bits  [DGX-2]	nvidia-ipmisol
console=ttyS0,115200n8	Set console to serial port 0, using 115200 baud, no parity, 8 data bits	nvidia-ipmisol

## Copyright

© 2023 NVIDIA Corporation & Affiliates. All rights reserved.

