



# DGX Station A100

## User Guide

# Table of Contents

About this Guide.....	vi
<b>Chapter 1. Introduction to the NVIDIA DGX Station™ A100.....</b>	<b>1</b>
1.1. Registering Your DGX Station A100.....	1
1.2. What's in the Box.....	2
1.3. DGX OS Software Summary.....	2
1.4. DGX Station A100 Hardware Summary.....	3
<b>Chapter 2. Getting Started with DGX Station A100.....</b>	<b>4</b>
2.1. Connecting and Powering on the DGX Station A100.....	4
2.2. Using DGX Station A100 as a Server Without a Monitor.....	8
2.3. Running Workloads on Systems with Mixed Types of GPUs.....	10
2.3.1. Running with Docker Containers.....	10
2.3.2. Running on Bare Metal.....	11
2.3.3. Using Multi-Instance GPUs.....	15
2.4. Completing the Initial Ubuntu OS Configuration.....	16
<b>Chapter 3. Using the BMC.....</b>	<b>17</b>
3.1. Understanding the BMC Controls.....	18
3.2. Configuring a Static IP Address for the BMC.....	19
3.3. Configuring a BMC Static IP Address Using ipmitool.....	19
3.3.1. Configuring a BMC Static IP Address Using the System BIOS.....	20
3.4. Logging into the BMC.....	20
3.5. Changing Your Default BMC Password.....	21
3.6. Logging in After Entering an Incorrect Password.....	22
3.7. Common BMC Tasks.....	22
3.7.1. Configuring the BMC Login Credentials.....	22
3.7.2. Using the Remote Control.....	23
3.7.3. Setting Up Active Directory or LDAP/E-Directory.....	23
3.7.4. Configuring Platform Event Filters.....	24
3.7.5. Uploading or Generating SSL Certificates.....	24
3.7.5.1. Viewing the SSL Certificate.....	24
3.7.5.2. Generating the SSL Certificate.....	25
3.7.5.3. Uploading the SSL Certificate.....	26
<b>Chapter 4. Enable MIG Mode in DGX Station A100.....</b>	<b>27</b>
<b>Chapter 5. Managing Self-Encrypting Drives on DGX Station A100.....</b>	<b>29</b>
5.1. Overview.....	29

5.2. Installing the nv-disk-encrypt Package.....	30
5.3. Configuring Trusted Computing.....	30
5.3.1. Determining Whether Drives Support Block SID.....	31
5.3.2. Enabling the TPM and Preventing the BIOS from Sending Block SID Requests.....	31
5.4. Initializing the System for Drive Encryption.....	32
5.5. Enabling Drive Locking.....	33
5.6. Initialization Examples.....	33
5.6.1. Example 1: Passing in the JSON File.....	33
5.6.1.1. Determining the Drives Can be Managed as Self Encrypting.....	33
5.6.1.2. Creating the Drive/Password Mapping JSON File.....	34
5.6.2. Example 2: Generating Random Passwords.....	35
5.6.3. Example 3: Specifying Passwords One at a Time When Prompted.....	35
5.7. Disabling Drive Locking.....	35
5.8. Exporting the Vault.....	36
5.9. Erasing Your Data.....	36
5.10. Clearing the TPM.....	37
5.11. Changing Disk Passwords, Adding Disks, or Replacing Disks.....	37
5.12. Recovering a Lost Key.....	37
<b>Chapter 6. Unpacking and Repacking the DGX Station A100.....</b>	<b>39</b>
6.1. Unpacking the DGX Station A100.....	39
6.2. Repacking the DGX Station A100 for Shipment.....	42
<b>Appendix A. Security.....</b>	<b>46</b>
<b>Appendix B. Safety.....</b>	<b>47</b>
B.1. Intended Application Uses.....	47
B.2. General Precautions.....	48
B.3. Electrical Precautions.....	48
B.4. Communications Cable Precautions.....	49
B.5. Other Hazards.....	50
<b>Appendix C. Connections, Controls, and Indicators.....</b>	<b>51</b>
C.1. Front-Panel Connections and Controls.....	51
C.2. Rear-Panel Connections and Controls.....	52
C.3. LAN Port Indicators.....	53
<b>Appendix D. Compliance.....</b>	<b>55</b>
D.1. DGX Station A100 Model Number.....	55
D.2. Australia/New Zealand.....	55
D.3. Brazil.....	55
D.4. Canada.....	56

D.5. China.....	56
D.6. European Union.....	57
D.7. India.....	58
D.8. Japan.....	59
D.9. Mexico.....	60
D.10. Russia/Kazakhstan/Belarus.....	60
D.11. South Africa.....	60
D.12. South Korea.....	61
D.13. Taiwan.....	61
D.14. United Kingdom.....	62
D.15. United States.....	63
D.16. United States/Canada.....	64
<b>Appendix E. DGX Station A100 Hardware Specifications.....</b>	<b>65</b>
E.1. Environmental Conditions.....	65
E.2. Component Specifications.....	65
E.3. Mechanical Specifications.....	65
E.4. Power Specifications.....	66
<b>Appendix F. Customer Support for the NVIDIA DGX Station™ A100.....</b>	<b>67</b>

## List of Tables

Table 1. BMC Navigation Controls .....	18
Table 2. Fields to Generate an SSL Certificate .....	25

---

# About this Guide

*DGX Station A100 User Guide* explains how to install, set up, and maintain the NVIDIA DGX Station™ A100 .

This guide is aimed at users and administrators who are familiar with the Ubuntu Desktop Linux OS, including the command line and the `sudo` command.



**Note:** The instructions in this guide for software administration apply only to the DGX OS. They do not apply if the DGX OS software that is supplied with the DGX Station A100 has been replaced with the DGX software for Red Hat Enterprise Linux or CentOS.

For additional information to help you use the DGX Station A100, see the following table.

Task	Additional Information
Use the Ubuntu Desktop Linux OS.	<a href="https://help.ubuntu.com/20.04/ubuntu-help/index.html">Ubuntu 20.04 Desktop Guide (<a href="https://help.ubuntu.com/20.04/ubuntu-help/index.html">https://help.ubuntu.com/20.04/ubuntu-help/index.html</a>)</a>
Use the DGX Station A100 to download and run containers for deep learning frameworks.	<a href="#">NGC Container Registry for DGX User Guide</a>
Use deep learning frameworks optimized for NVIDIA DGX systems.	<a href="#">NVIDIA Deep Learning Frameworks Documentation (<a href="https://docs.nvidia.com/deeplearning/dgx/">https://docs.nvidia.com/deeplearning/dgx/</a>)</a>
Use the tools and libraries in the DGX OS for development of deep learning frameworks.	<a href="#">NVIDIA Deep Learning SDK Documentation (<a href="https://docs.nvidia.com/deeplearning/sdk/">https://docs.nvidia.com/deeplearning/sdk/</a>)</a>

---

# Chapter 1. Introduction to the NVIDIA DGX Station™ A100

NVIDIA™ DGX Station™ A100 brings AI supercomputing to data science teams, offering data center technology without a data center or additional IT investment. Designed for multiple, simultaneous users, DGX Station A100 leverages server-grade components in an easy-to-place workstation form factor. It's the only system with four fully interconnected and Multi-Instance GPU (MIG)-capable NVIDIA A100 Tensor Core GPUs with up to 320 gigabytes (GB) of total GPU memory that can plug into a standard power outlet in the office or at home, resulting in a powerful AI appliance that you can place anywhere.



## 1.1. Registering Your DGX Station A100

To obtain support for your DGX Station A100, follow the instructions for registration in the Entitlement Certification email that was sent as part of the purchase.

Registration allows you to access the NVIDIA Enterprise Support Portal, obtain technical support, get software updates, and set up an NGC for DGX systems account. If you did not receive the information, open a case with the NVIDIA Enterprise Support Team at [Enterprise Support](#).

## 1.2. What's in the Box

- ▶ DGX Station A100
- ▶ NVIDIA DGX Station™ A100
- ▶ Accessory boxes containing the following items:
  - ▶ Quick Start Guide
  - ▶ AC power cable with a locale-specific grounded connector
  - ▶ Mini DisplayPort 1.2 to DisplayPort
  - ▶ USB recovery flash drive containing:
    - ▶ Source code of the open-source software that is installed on DGX Station™ A100
    - ▶ Toxic Substance Notice and Safety Instructions
    - ▶ Declaration of Conformity
    - ▶ Wheel locks

Wheel locks are provided to prevent the unit from rolling out of place. To install these locks, from the side, slide them under two of the wheels.

Inspect each piece of equipment in the packing box. If anything is missing or damaged, contact your supplier.

**Before** you get started, you need to first unpack DGX Station A100. See [Unpacking your DGX Station A100](#) for more information.

## 1.3. DGX OS Software Summary

The DGX OS software that is supplied with the DGX Station A100 includes the software that you need for downloading and running containers for deep learning frameworks. The software is already installed on the DGX Station A100, except where licensing requirements mandate that the software be supplied separately. Any software that must be supplied separately is installed automatically when the DGX Station A100 is first powered on.

For details about the DGX OS software, refer to the [DGX OS Software Release notes](#).

## 1.4. DGX Station A100 Hardware Summary

### Processors

Component	Qty	Description
CPU	1	Single AMD 7742, 64 cores, and 2.25 GHz (base)–3.4 GHz (max boost).
GPU	4	NVIDIA A100 with 80 GB (320GB total) or 40GB per GPU (160GB total) of GPU memory.

### System Memory and Storage

Component	Qty	Unit Capacity	Total Capacity	Description
System memory	8	64 GB	512 GB	System DDR4 RAM
Data storage	1	7.68 TB	7.68 TB	Cache/Data U.2 NVME drive
OS storage	1	1.92 TB	1.92 TB	Boot M.2 NVME drive

---

# Chapter 2. Getting Started with DGX Station A100

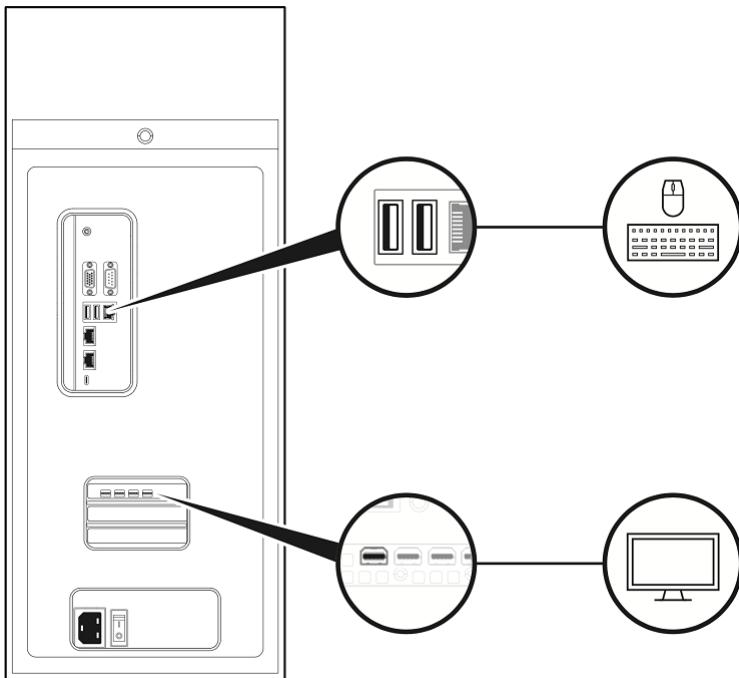
This section provides information about how to connect and power on the DGX Station A100.

## 2.1. Connecting and Powering on the DGX Station A100

To complete this task you need the following items, which are not supplied with the DGX Station A100:

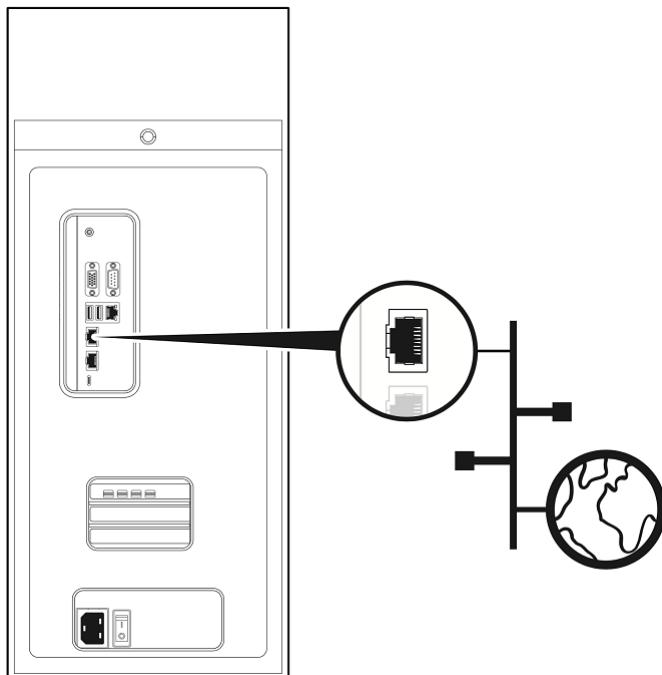
- ▶ Mini DisplayPort 1.2 to DisplayPort.
- ▶ USB keyboard
- ▶ USB mouse
- ▶ Ethernet cable

1. Connect a display to any DisplayPort connector and a keyboard and mouse to any two USB ports.



**Note:** For initial setup, connect only **one** display to the DGX Station A100. After you complete the initial Ubuntu OS configuration, you can configure the DGX Station A100 to use multiple displays. Refer to the [NVIDIA DGX OS 5 User Guide](#) for more information.

2. Use any of the two Ethernet ports to connect the DGX Station A100 to your LAN with Internet connectivity.





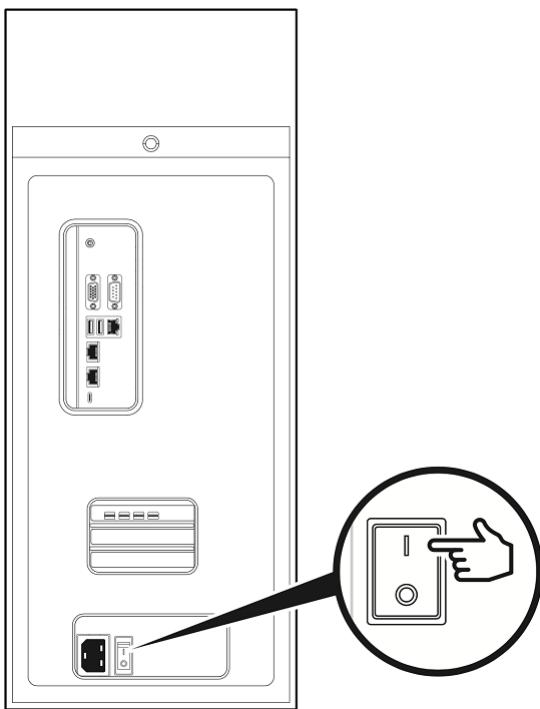
**Note:** Remember the following information:

- ▶ Connect only **one** Ethernet port on the DGX Station A100 to the Internet **unless** you plan to configure the ports manually and disable DHCP on at least one of the ports.
- ▶ By default, both Ethernet ports on the DGX Station A100 are configured for DHCP. If both the ports are connected simultaneously, each port will get its own IP address. The IP address that the Linux operating system (OS) uses will then alternate between these addresses, causing the OS and applications to malfunction.



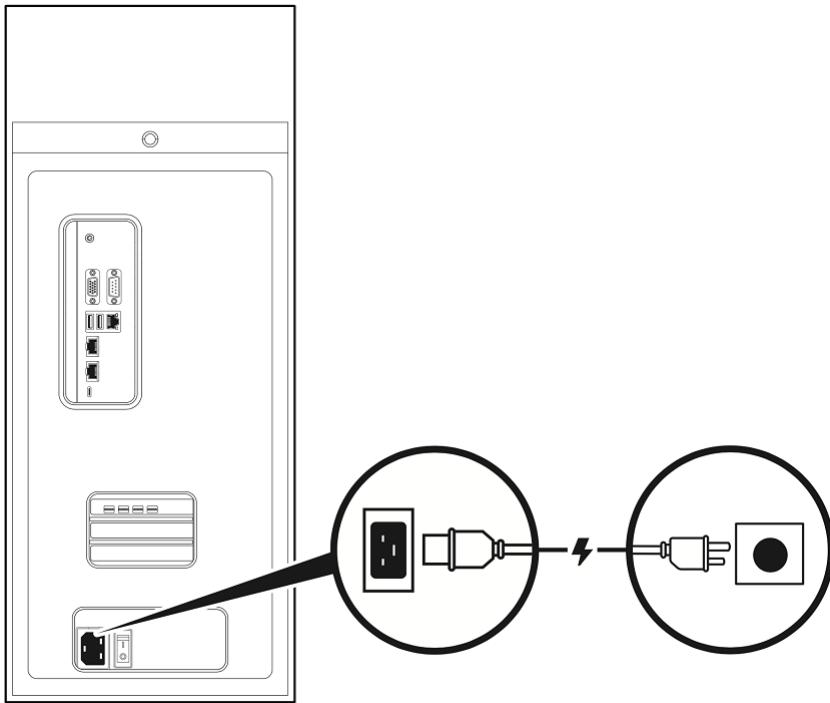
**Important:** After you boot the system and run through the initial configuration, **do not** edit the Network Manager Configuration file.

3. Make sure that the power supply rocker switch is in the OFF position.



4. Connect the supplied power cable from the power socket at the back of the unit to an appropriately rated, grounded AC outlet.

For details of the power consumption, input voltage, and current rating of the DGX Station A100, see [Power Specifications](#).



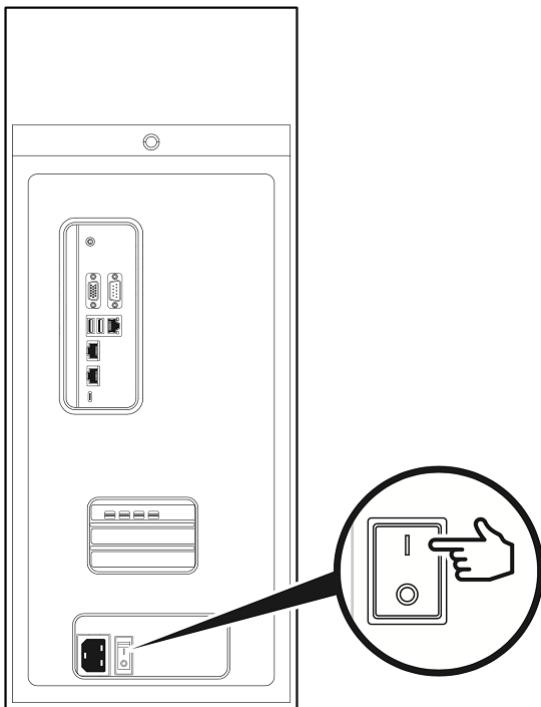
**RESTRICTION:** The power source for DGX Station A100 **must be 100V** and **cannot fall below 90V**.



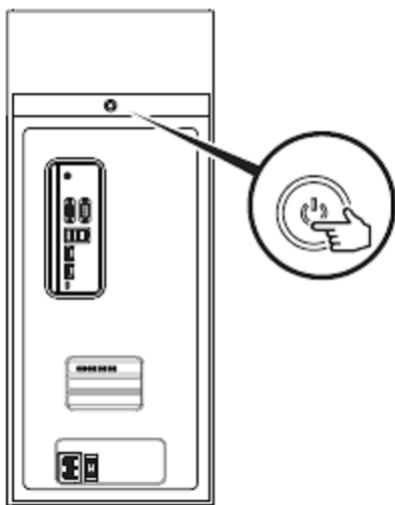
**CAUTION:** Remember the following information:

- ▶ Use **only** the supplied power cable and do not use this power cable with any other products or for any other purpose. Not all power cables have the same current ratings.
- ▶ Do **not** use household extension cables with your product. Household extension cables do not have overload protection and are not intended for use with computer systems.

5. Connect the display to a suitable AC outlet and power on the display.
6. Move the DGX Station A100 power supply rocker switch to the ON position.



7. Push the Power button on the front of the unit to power on the DGX Station A100.



## 2.2. Using DGX Station A100 as a Server Without a Monitor

By default, DGX Station A100 is shipped with the DP port automatically selected in the display.

To enter the SBIOS setup, see [Configuring a BMC Static IP Address Using the System BIOS](#).

- ▶ If you plan to use DGX Station A100 as a **desktop system**, use the information in this user guide to get started.

You do not need to make changes to the SBIOS.

- ▶ If you plan to use it as a **server without a monitor**, after the machine has booted in the Desktop GUI, the BMC remote console will not show a display.

In this case, return to the BIOS, and complete the following steps.



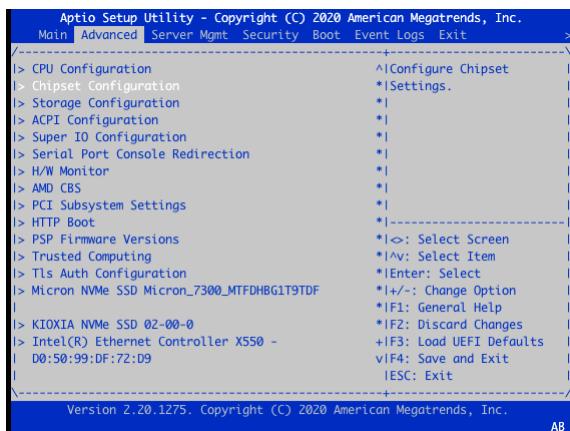
**Important:** If you do not change your SBIOS settings, after the machine has booted in the Desktop GUI, the BMC remote console will not show a display.

To change your SBIOS settings:

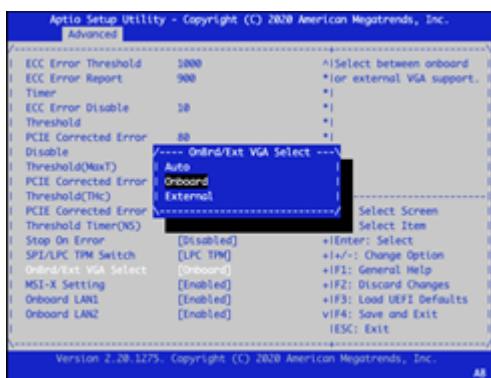


**Tip:** The SBIOS screen will show up on any monitor that is connected to the DP port, the VGA port, or the BMC remote console.

1. In the setup utility, click the **Advanced** tab.
2. Select **Chipset Configuration**.



3. In the **OnBrd/Ext VGA Select** dialog box, select **Onboard**.



4. To save and exit, press **F4**.

## 2.3. Running Workloads on Systems with Mixed Types of GPUs

The DGX Station A100 comes equipped with four high performance NVIDIA A100 GPUs and one DGX Display GPU. The NVIDIA A100 GPU is used to run high performance and AI workloads, and the DGX Display card is used to drive a high-quality display on a monitor.

When running applications on this system, it is important to identify the best method to launch applications and workloads to make sure the high performance NVIDIA A100 GPUs are used. You can achieve this in one of the following ways:

- ▶ [Running with Docker Containers](#)
- ▶ [Running on Bare Metal](#)
- ▶ [Using Multi-Instance GPUs](#)

When you log into the system and check which GPUs are available, you find the following:

```
lab@ro-dvt-058-80gb:~$ nvidia-smi -L
GPU 0: Graphics Device (UUID: GPU-269d95f8-328a-08a7-5985-ab09e6e2b751)
GPU 1: Graphics Device (UUID: GPU-0f2dff15-7c85-4320-da52-d3d54755d182)
GPU 2: Graphics Device (UUID: GPU-dc598de6-dd4d-2f43-549f-f7b4847865a5)
GPU 3: DGX Display (UUID: GPU-91b9d8c8-e2b9-6264-99e0-b47351964c52)
GPU 4: Graphics Device (UUID: GPU-e32263f2-ae07-f1db-37dc-17d1169b09bf)
```

A total of five GPUs are listed by nvidia-smi. This is because nvidia-smi is including the DGX Display GPU that is used to drive the monitor and high-quality graphics output.

When running an application or workload, the DGX Display GPU can get in the way because it does not have direct NVlink connectivity, sufficient memory, or the performance characteristics of the NVIDIA A100 GPUs that are installed on the system. As a result you should ensure that the correct GPUs are being used.

### 2.3.1. Running with Docker Containers

On the DGX OS, because Docker has already been configured to identify the high performance NVIDIA A100 GPUs and assign them to the container, this method is the simplest method.

A simple test is to run a small container with the `--gpus all` flag in the command and once in the container that is running `nvidia-smi`. The output shows that only the high-performance GPUs are available to the container:

```
lab@ro-dvt-058-80gb:~$ docker run --gpus all --rm -it ubuntu nvidia-smi -L
GPU 0: Graphics Device (UUID: GPU-269d95f8-328a-08a7-5985-ab09e6e2b751)
GPU 1: Graphics Device (UUID: GPU-0f2dff15-7c85-4320-da52-d3d54755d182)
GPU 2: Graphics Device (UUID: GPU-dc598de6-dd4d-2f43-549f-f7b4847865a5)
GPU 3: Graphics Device (UUID: GPU-e32263f2-ae07-f1db-37dc-17d1169b09bf)
```

This step will also work when the `--gpus n` flag is used, where `n` can be 1, 2, 3, or 4. These values represent the number of GPUs that should be assigned to that container. For example:

```
lab@ro-dvt-058-80gb:~ $ docker run --gpus 2 --rm -it ubuntu nvidia-smi -L
```

```
GPU 0: Graphics Device (UUID: GPU-269d95f8-328a-08a7-5985-ab09e6e2b751)
GPU 1: Graphics Device (UUID: GPU-0f2dff15-7c85-4320-da52-d3d54755d182)
```

In this example, Docker selected the first two GPUs to run the container, but if the `device` option is used, you can specify which GPUs to use:

```
lab@ro-dvt-058-80gb:~$ docker run --gpus '"device=GPU-dc598de6-dd4d-2f43-549f-f7b4847865a5, GPU-e32263f2-ae07-f1db-37dc-17d1169b09bf"' --rm -it ubuntu nvidia-smi -L
GPU 0: Graphics Device (UUID: GPU-dc598de6-dd4d-2f43-549f-f7b4847865a5)
GPU 1: Graphics Device (UUID: GPU-e32263f2-ae07-f1db-37dc-17d1169b09bf)
```

In this example, the two GPUs that were not used earlier are now assigned to run on the container.

### 2.3.2. Running on Bare Metal

To run applications by using the four high performance GPUs, the `CUDA_VISIBLE_DEVICES` variable must be specified before you run the application.



**Note:** This method does not use containers.

CUDA orders the GPUs by performance, so GPU 0 will be the highest performing GPU, and the last GPU will be the slowest GPU.



**Important:** If the `CUDA_DEVICE_ORDER` variable is set to `PCI_BUS_ID`, this ordering will be overridden.

In the following example, a CUDA application that comes with CUDA samples is run. In the output, GPU 0 is the fastest in a DGX Station A100, and GPU 4 (DGX Display GPU) is the slowest:

```
lab@ro-dvt-058-80gb:~$ sudo apt install cuda-samples-11-2
lab@ro-dvt-058-80gb:~$ cd /usr/local/cuda-11.2/samples/1_Utilities/p2pBandwidthLatencyTest
lab@ro-dvt-058-80gb:/usr/local/cuda-11.2/samples/1_Utilities/p2pBandwidthLatencyTest
$ sudo make
/usr/local/cuda/bin/nvcc -ccbin g++ -I../../common/inc -m64 --threads
0 -gencode arch=compute_35,code=sm_35 -gencode arch=compute_37,code=sm_37
-gencode arch=compute_50,code=sm_50 -gencode arch=compute_52,code=sm_52
-gencode arch=compute_60,code=sm_60 -gencode arch=compute_61,code=sm_61
-gencode arch=compute_70,code=sm_70 -gencode arch=compute_75,code=sm_75
-gencode arch=compute_80,code=sm_80 -gencode arch=compute_86,code=sm_86
-gencode arch=compute_86,code=compute_86 -o p2pBandwidthLatencyTest.o -c
p2pBandwidthLatencyTest.cu
nvcc warning : The 'compute_35', 'compute_37', 'compute_50', 'sm_35', 'sm_37' and
'sm_50' architectures are deprecated, and may be removed in a future release (Use -
Wno-deprecated-gpu-targets to suppress warning).
/usr/local/cuda/bin/nvcc -ccbin g++ -m64
-gencode arch=compute_35,code=sm_35 -gencode arch=compute_37,code=sm_37
-gencode arch=compute_50,code=sm_50 -gencode arch=compute_52,code=sm_52
-gencode arch=compute_60,code=sm_60 -gencode arch=compute_61,code=sm_61
-gencode arch=compute_70,code=sm_70 -gencode arch=compute_75,code=sm_75
-gencode arch=compute_80,code=sm_80 -gencode arch=compute_86,code=sm_86
-gencode arch=compute_86,code=compute_86 -o p2pBandwidthLatencyTest
p2pBandwidthLatencyTest.o
nvcc warning : The 'compute_35', 'compute_37', 'compute_50', 'sm_35', 'sm_37' and
'sm_50' architectures are deprecated, and may be removed in a future release (Use -
Wno-deprecated-gpu-targets to suppress warning).
mkdir -p ../../bin/x86_64/linux/release
cp p2pBandwidthLatencyTest ../../bin/x86_64/linux/release
lab@ro-dvt-058-80gb:/usr/local/cuda-11.2/samples/1_Utilities/p2pBandwidthLatencyTest
$ cd /usr/local/cuda-11.2/samples/bin/x86_64/linux/release
```

```
lab@ro-dvt-058-80gb:/usr/local/cuda-11.2/samples/bin/x86_64/linux/release $ ./p2pBandwidthLatencyTest
[P2P (Peer-to-Peer) GPU Bandwidth Latency Test]
Device: 0, Graphics Device, pciBusID: 1, pciDeviceID: 0, pciDomainID:0
Device: 1, Graphics Device, pciBusID: 47, pciDeviceID: 0, pciDomainID:0
Device: 2, Graphics Device, pciBusID: 81, pciDeviceID: 0, pciDomainID:0
Device: 3, Graphics Device, pciBusID: c2, pciDeviceID: 0, pciDomainID:0
Device: 4, DGX Display, pciBusID: c1, pciDeviceID: 0, pciDomainID:0
Device=0 CAN Access Peer Device=1
Device=0 CAN Access Peer Device=2
Device=0 CAN Access Peer Device=3
Device=0 CANNOT Access Peer Device=4
Device=1 CAN Access Peer Device=0
Device=1 CAN Access Peer Device=2
Device=1 CAN Access Peer Device=3
Device=1 CANNOT Access Peer Device=4
Device=2 CAN Access Peer Device=0
Device=2 CAN Access Peer Device=1
Device=2 CAN Access Peer Device=3
Device=2 CANNOT Access Peer Device=4
Device=3 CAN Access Peer Device=0
Device=3 CAN Access Peer Device=1
Device=3 CAN Access Peer Device=2
Device=3 CANNOT Access Peer Device=4
Device=4 CANNOT Access Peer Device=0
Device=4 CANNOT Access Peer Device=1
Device=4 CANNOT Access Peer Device=2
Device=4 CANNOT Access Peer Device=3
```

\*\*\*NOTE: In case a device doesn't have P2P access to other one, it falls back to normal memcpy procedure.  
So you can see lesser Bandwidth (GB/s) and unstable Latency (us) in those cases.

#### P2P Connectivity Matrix

D\D	0	1	2	3	4
0	1	1	1	1	0
1	1	1	1	1	0
2	1	1	1	1	0
3	1	1	1	1	0
4	0	0	0	0	1

#### Unidirectional P2P=Disabled Bandwidth Matrix (GB/s)

D\D	0	1	2	3	4
0	1323.03	15.71	15.37	16.81	12.04
1	16.38	1355.16	15.47	15.81	11.93
2	16.25	15.85	1350.48	15.87	12.06
3	16.14	15.71	16.80	1568.78	11.75
4	12.61	12.47	12.68	12.55	140.26

#### Unidirectional P2P=Enabled Bandwidth (P2P Writes) Matrix (GB/s)

D\D	0	1	2	3	4
0	1570.35	93.30	93.59	93.48	12.07
1	93.26	1583.08	93.55	93.53	11.93
2	93.44	93.58	1584.69	93.34	12.05
3	93.51	93.55	93.39	1586.29	11.79
4	12.68	12.54	12.75	12.51	140.26

#### Bidirectional P2P=Disabled Bandwidth Matrix (GB/s)

D\D	0	1	2	3	4
0	1588.71	19.60	19.26	19.73	16.53
1	19.59	1582.28	19.85	19.13	16.43
2	19.53	19.39	1583.88	19.61	16.58
3	19.51	19.11	19.58	1592.76	15.90
4	16.36	16.31	16.39	15.80	139.42

#### Bidirectional P2P=Enabled Bandwidth Matrix (GB/s)

D\D	0	1	2	3	4
0	1590.33	184.91	185.37	185.45	16.46
1	185.04	1587.10	185.19	185.21	16.37
2	185.15	185.54	1516.25	184.71	16.47

```

3 185.55 185.32 184.86 1589.52 15.71
4 16.26 16.28 16.16 15.69 139.43
P2P=Disabled Latency Matrix (us)
GPU 0 1 2 3 4
0 3.53 21.60 22.22 21.38 12.46
1 21.61 2.62 21.55 21.65 12.34
2 21.57 21.54 2.61 21.55 12.40
3 21.57 21.54 21.58 2.51 13.00
4 13.93 12.41 21.42 21.58 1.14

CPU 0 1 2 3 4
0 4.26 11.81 13.11 12.00 11.80
1 11.98 4.11 11.85 12.19 11.89
2 12.07 11.72 4.19 11.82 12.49
3 12.14 11.51 11.85 4.13 12.04
4 12.21 11.83 12.11 11.78 4.02
P2P=Enabled Latency (P2P Writes) Matrix (us)
GPU 0 1 2 3 4
0 3.79 3.34 3.34 3.37 13.85
1 2.53 2.62 2.54 2.52 12.36
2 2.55 2.55 2.61 2.56 12.34
3 2.58 2.51 2.51 2.53 14.39
4 19.77 12.32 14.75 21.60 1.13

CPU 0 1 2 3 4
0 4.27 3.63 3.65 3.59 13.15
1 3.62 4.22 3.61 3.62 11.96
2 3.81 3.71 4.35 3.73 12.15
3 3.64 3.61 3.61 4.22 12.06
4 12.32 11.92 13.30 12.03 4.05

```

**NOTE:** The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is enabled.

The example above shows the peer-to-peer bandwidth and latency test across all five GPUs, including the DGX Display GPU. The application also shows that there is no peer-to-peer connectivity between any GPU and GPU 4. This indicates that GPU 4 should not be used for high-performance workloads.

Run the example one more time by using the `CUDA_VISIBLE_DEVICES` variable, which limits the number of GPUs that the application can see.



**Note:** All GPUs can communicate with all other peer devices.

```

lab@ro-dvt-058-80gb:/usr/local/cuda-11.2/samples/bin/x86_64/linux/release$ 
CUDA_VISIBLE_DEVICES=0,1,2,3 ./p2pBandwidthLatencyTest
[P2P (Peer-to-Peer) GPU Bandwidth Latency Test]
Device: 0, Graphics Device, pciBusID: 1, pciDeviceID: 0, pciDomainID:0
Device: 1, Graphics Device, pciBusID: 47, pciDeviceID: 0, pciDomainID:0
Device: 2, Graphics Device, pciBusID: 81, pciDeviceID: 0, pciDomainID:0
Device: 3, Graphics Device, pciBusID: c2, pciDeviceID: 0, pciDomainID:0
Device=0 CAN Access Peer Device=1
Device=0 CAN Access Peer Device=2
Device=0 CAN Access Peer Device=3
Device=1 CAN Access Peer Device=0
Device=1 CAN Access Peer Device=2
Device=1 CAN Access Peer Device=3
Device=2 CAN Access Peer Device=0
Device=2 CAN Access Peer Device=1
Device=2 CAN Access Peer Device=3
Device=3 CAN Access Peer Device=0
Device=3 CAN Access Peer Device=1
Device=3 CAN Access Peer Device=2

```

\*\*\*NOTE: In case a device doesn't have P2P access to other one, it falls back to normal memcpy procedure.  
So you can see lesser Bandwidth (GB/s) and unstable Latency (us) in those cases.

## P2P Connectivity Matrix

D\D	0	1	2	3
0	1	1	1	1
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1

## Unidirectional P2P=Disabled Bandwidth Matrix (GB/s)

D\D	0	1	2	3
0	1324.15	15.54	15.62	15.47
1	16.55	1353.99	15.52	16.23
2	15.87	17.26	1408.93	15.91
3	16.33	17.31	18.22	1564.06

## Unidirectional P2P=Enabled Bandwidth (P2P Writes) Matrix (GB/s)

D\D	0	1	2	3
0	1498.08	93.30	93.53	93.48
1	93.32	1583.08	93.54	93.52
2	93.55	93.60	1583.08	93.36
3	93.49	93.55	93.28	1576.69

## Bidirectional P2P=Disabled Bandwidth Matrix (GB/s)

D\D	0	1	2	3
0	1583.08	19.92	20.47	19.97
1	20.74	1586.29	20.06	20.22
2	20.08	20.59	1590.33	20.01
3	20.44	19.92	20.60	1589.52

## Bidirectional P2P=Enabled Bandwidth Matrix (GB/s)

D\D	0	1	2	3
0	1592.76	184.88	185.21	185.30
1	184.99	1589.52	185.19	185.32
2	185.28	185.30	1585.49	185.01
3	185.45	185.39	184.84	1587.91

## P2P=Disabled Latency Matrix (us)

GPU	0	1	2	3
0	2.38	21.56	21.61	21.56
1	21.70	2.34	21.54	21.56
2	21.55	21.56	2.41	21.06
3	21.57	21.34	21.56	2.39

CPU	0	1	2	3
0	4.22	11.99	12.71	12.09
1	11.86	4.09	12.00	11.71
2	12.52	11.98	4.27	12.24
3	12.22	11.75	12.19	4.25

## P2P=Enabled Latency (P2P Writes) Matrix (us)

GPU	0	1	2	3
0	2.32	2.57	2.55	2.59
1	2.55	2.32	2.59	2.52
2	2.59	2.56	2.41	2.59
3	2.57	2.55	2.56	2.40

CPU	0	1	2	3
0	4.24	3.57	3.72	3.81
1	3.68	4.26	3.75	3.63
2	3.79	3.75	4.34	3.71
3	3.72	3.64	3.66	4.32

NOTE: The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is enabled.

For bare metal applications, the UUID can also be specified in the CUDA\_VISIBLE\_DEVICES variable as shown below:

```
lab@ro-dvt-058-80gb:/usr/local/cuda-11.2/samples/bin/x86_64/linux/release
$ CUDA_VISIBLE_DEVICES=GPU-0f2dff15-7c85-4320-da52-d3d54755d182, GPU-dc598de6-
dd4d-2f43-549f-f7b4847865a5 ./p2pBandwidthLatencyTest
```

The GPU specification is longer because of the nature of UUIDs, but this is the most precise way to pin specific GPUs to the application.

### 2.3.3. Using Multi-Instance GPUs

Multi-Instance GPUs (MIG) is a technology that is available on NVIDIA A100 GPUs. If MIG is enabled on the GPUs and if the GPUs have been partitioned already, then applications can be limited to run on these devices.

This works for both Docker containers and for bare metal using the `CUDA_VISIBLE_DEVICES` as shown in the examples below. For instructions on how to configure and use MIG, refer to the [NVIDIA Multi-Instance GPU User Guide](#).

Identify the MIG instances that will be used. Here is the output from a system that has GPU 0 partitioned into 7 MIGs:

```
lab@ro-dvt-058-80gb:~$ nvidia-smi -L
GPU 0: Graphics Device (UUID: GPU-269d95f8-328a-08a7-5985-ab09e6e2b751)
    MIG 1g.10gb Device 0: (UUID: MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/7/0)
    MIG 1g.10gb Device 1: (UUID: MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/8/0)
    MIG 1g.10gb Device 2: (UUID: MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/9/0)
    MIG 1g.10gb Device 3: (UUID: MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/11/0)
    MIG 1g.10gb Device 4: (UUID: MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/12/0)
    MIG 1g.10gb Device 5: (UUID: MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/13/0)
    MIG 1g.10gb Device 6: (UUID: MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/14/0)
GPU 1: Graphics Device (UUID: GPU-0f2dff15-7c85-4320-da52-d3d54755d182)
GPU 2: Graphics Device (UUID: GPU-dc598de6-dd4d-2f43-549f-f7b4847865a5)
GPU 3: DGX Display (UUID: GPU-91b9d8c8-e2b9-6264-99e0-b47351964c52)
GPU 4: Graphics Device (UUID: GPU-e32263f2-ae07-f1db-37dc-17d1169b09bf)
```

In Docker, enter the MIG UUID from this output, in which GPU 0 and Device 0 have been selected.

If you are running on DGX Station A100, restart the `nv-docker-gpus` and `docker` system services any time MIG instances are created, destroyed or modified by running the following:

```
lab@ro-dvt-058-80gb:~$ sudo systemctl restart nv-docker-gpus; sudo systemctl restart docker
```

`nv-docker-gpus` has to be restarted on DGX Station A100 because this service is used to mask the available GPUs that can be used by Docker. When the GPU architecture changes, the service needs to be refreshed.

```
lab@ro-dvt-058-80gb:~$ docker run --gpus '"device=MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/7/0"' --rm -it ubuntu nvidia-smi -L
GPU 0: Graphics Device (UUID: GPU-269d95f8-328a-08a7-5985-ab09e6e2b751)
    MIG 1g.10gb Device 0: (UUID: MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/7/0)
```

On bare metal, specify the MIG instances:



**Remember:** This application measures the communication across GPUs, and it is not relevant to read the bandwidth and latency with only one GPU MIG.

The purpose of this example is to illustrate how to use specific GPUs with applications, which is clearly illustrated below.

```
lab@ro-dvt-058-80gb: /usr/local/cuda-11.2/samples/bin/x86_64/linux/release
$ CUDA_VISIBLE_DEVICES=MIG-GPU-269d95f8-328a-08a7-5985-ab09e6e2b751/7/0 ./p2pBandwidthLatencyTest
[P2P (Peer-to-Peer) GPU Bandwidth Latency Test]
Device: 0, Graphics Device MIG 1g.10gb, pciBusID: 1, pciDeviceID: 0, pciDomainID: 0
```

```
***NOTE: In case a device doesn't have P2P access to other one, it falls back to  
normal memcpy procedure.  
So you can see lesser Bandwidth (GB/s) and unstable Latency (us) in those cases.  
  
P2P Connectivity Matrix  
D\D      0  
0       1  
Unidirectional P2P=Disabled Bandwidth Matrix (GB/s)  
D\D      0  
0   176.20  
Unidirectional P2P=Enabled Bandwidth (P2P Writes) Matrix (GB/s)  
D\D      0  
0   187.87  
Bidirectional P2P=Disabled Bandwidth Matrix (GB/s)  
D\D      0  
0   190.77  
Bidirectional P2P=Enabled Bandwidth Matrix (GB/s)  
D\D      0  
0   190.53  
P2P=Disabled Latency Matrix (us)  
GPU     0  
0    3.57  
  
CPU     0  
0    4.07  
P2P=Enabled Latency (P2P Writes) Matrix (us)  
GPU     0  
0    3.55  
  
CPU     0  
0    4.07
```

NOTE: The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is enabled.

## 2.4. Completing the Initial Ubuntu OS Configuration

When you power on the DGX Station A100 for the first time, you are prompted to accept end user license agreements for NVIDIA software. You are then guided through the process for completing the initial Ubuntu OS configuration.

For the complete procedure, refer to the [NVIDIA DGX OS 5 User Guide](#).

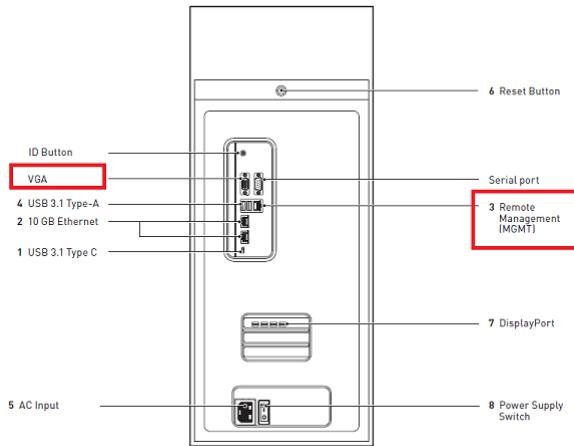
# Chapter 3. Using the BMC

The DGX Station A100 comes with a baseboard management controller (BMC) to monitor and control various hardware devices on the system, system sensors, and other parameters.



**CAUTION:** Before you can log into the BMC, see the security information in [Security](#).

BMC features its own management network port and VGA Display port.



The DGX Station A100 features two display devices, the DGX Display Adapter and the BMC Display Adapter. Depending on the BIOS settings you can direct the OS X display to either adapter. These settings can be adjusted to use one of the three modes listed before from the BIOS. The nvidia-conf-xconfig.service manages this function for the X-Windowing system.

Here is a list of the display select modes:

## Auto

Automatically configures the X-windowing system to use whichever compatible display adapter is currently present in the system. The default choice is to use the NVIDIA DGX Display Adapter (mini-DP), however if this adapter is not present the on-board BMC Display Adapter (15pin VGA) will be chosen. If neither of the adapters is detected for any reason, the service will exit without setup.

**On-Board**

Configures the X-windowing system to exclusively use the on-board BMC Display Adapter. If for any reason BMC Display Adapter is not present, it will fall back to use NVIDIA DGX Display Adapter (if present), otherwise the service will exit without any setup.

**External**

Configures the X-windowing system to exclusively use the DGX Display Adapter. The BMC Display Adapter is disabled. If for any reason NVIDIA DGX Display Adapter is not present, the service will exit without any setup.

See [Using DGX Station A100 as a Server Without a Monitor](#) for more information.

## 3.1. Understanding the BMC Controls

In the BMC dashboard, the left navigation pane on the BMC main page contains the primary controls.

Here is a list of the controls:

**Table 1. BMC Navigation Controls**

Control	Description
<b>Dashboard</b>	Displays the overall information about the status of the device.
<b>Sensor</b>	Provides status and readings for system sensors, such as SSD, PSUs, voltages, CPU temperatures, DIMM temperatures, and fan speeds.
<b>System Inventory</b>	Displays inventory information of the following system modules: System, Processor, Memory Controller, BaseBoard, Power, Thermal, PCIE Device, PCIE Function, and Storage.
<b>FRU Information</b>	Provides chassis, board, and product information.
<b>GPU Information</b>	Provides basic information on all the GPUs in the systems, including GUID, VBIOS version, InfoROM version, and number of retired pages for each GPU.
<b>Logs and Reports</b>	View, and if applicable, download and erase, the IPMI event log, and System, Audit, Video, and POST Code logs.
<b>Settings</b>	Configure the following settings: Captured BSOD, Date & Time, External User Services, KVM Mouse Setting, Log Settings, Media Redirection Settings, Network Settings, PAM Order Settings, Platform Event Filter, Services, SMTP Settings, SSL Settings, System Firewall,

Control	Description
	User Management, Video Recording, and IPMI Interfaces.
<b>Remote Control</b>	Opens the KVM Launch page for accessing the DGX A100 console remotely.
<b>Power Control</b>	Perform the following power actions: Power On, Power Off, Power Cycle, Hard Reset, and ACPI/Shutdown
<b>Maintenance</b>	Perform the following maintenance tasks: Backup Configuration, Firmware Image Location, Firmware Update, BIOS Update, Preserve Configuration, PSU Update, FPGA Update, Retimer Update, BackPlane Update, Preserve Configuration, Restore Configuration, Restore Factory Defaults, and Reset
<b>Sign out</b>	Sign out of the BMC web UI.

## 3.2. Configuring a Static IP Address for the BMC

Here is some information about how to configure a static IP address for the BMC.

## 3.3. Configuring a BMC Static IP Address Using ipmitool

Here is some information about how to set a static IP address for the BMC from the Ubuntu command line.



**Note:** If you cannot access the DGX Station A100 remotely, connect a display (1440x900 or lower resolution) and keyboard directly to the DGX Station A100.

To view the current settings, enter the following command.

```
$ sudo ipmitool lan print 1
```

To set a static IP address for the BMC, complete the following steps:

a). Set the IP address source to static.

```
$ sudo ipmitool lan set 1 ipsrc static
```

b). Set the appropriate address information.

- ▶ To set the IP address, in **Station IP address**, enter the following, and replace the italicized text with your information.

```
$ sudo ipmitool lan set 1 ipaddr<my-ip-address>
```

- ▶ To set the subnet mask, enter the following and replace the italicized text with your information.

```
$ sudo ipmitool lan set 1 netmask<my-netmask-address>
```

- ▶ To set the default gateway IP, in **Router IP address**, enter the following and replace the italicized text with your information.

```
$ sudo ipmitool lan set 1 defgw ipaddr<my-default-gateway-ip-address>
```

### 3.3.1. Configuring a BMC Static IP Address Using the System BIOS

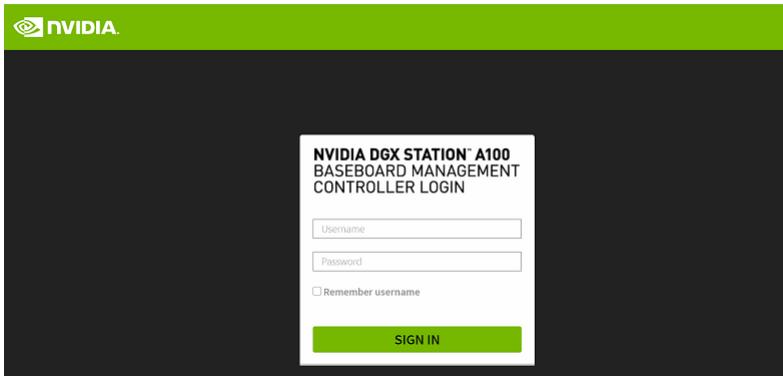
This section describes how to set a static IP address for the BMC when you cannot remotely access the DGX Station A100. This process involves setting the BMC IP address during system boot.

1. Connect a keyboard and display (1440 x 900 maximum resolution) to the DGX A100 System and power on the DGX Station A100.
2. When you see the SBIOS version screen, to enter the BIOS Setup Utility screen, press **Del** or **F2**.
3. In the BIOS Setup Utility screen, on the **Server Mgmt** tab, scroll to **BMC Network Configuration**, and press **Enter**.
4. Scroll to **Configuration address source** and press **Enter**.
5. In the **Configuration address source** dialog box, select **Static** and then press **Enter**.
6. Set the addresses for the Station IP address, Subnet mask, and Router IP address as needed by completing the following steps for each address:
  - a). Scroll to the specific item and press **Enter**.
  - b). Enter the appropriate information in the dialog box and press **Enter**.
  - c). After completing your changes, press **F4** to save and exit.

## 3.4. Logging into the BMC

Here are the steps to log into the BMC.

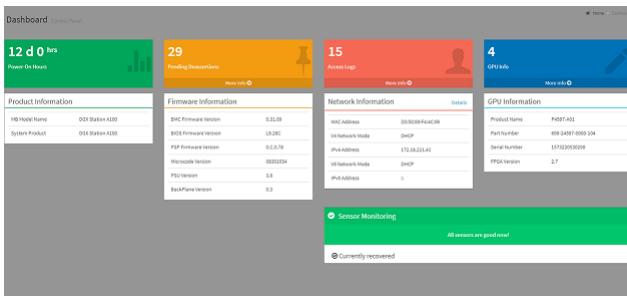
1. To log into the BMC, entered the configured IP address in a browser and press **Enter**.
2. Enter the BMC username and password that you configured earlier.



**WARNING:** If you enter the incorrect password 5 or more times, the **Login Failed** message is displayed, and you will be locked out. See [Logging in After Entering an Incorrect Password](#) for more information about when you can log back in.

3. Click **Sign in**.

Here is an example of the BMC dashboard:



## 3.5. Changing Your Default BMC Password

When you complete one of the following tasks:

- ▶ In the BMC web UI, click **Restore Factory Defaults** and **do not** click **Save** to preserve the settings.
- ▶ Run the `$ sudo ipmitool raw 0x32 0x66` command.

the passwords for the admin and Administrator users revert to the following default passwords:

- ▶ **admin** (for the admin role)
- ▶ **superuser** (for the Administrator role)

All other users on the system will be deleted.

We **strongly** recommend that you change your password. If you do not change your password, anyone can use the default passwords and log in to the BMC.

## 3.6. Logging in After Entering an Incorrect Password

After you enter an incorrect login BMC password at least five (5) times, you will be locked out. Here is some important information about how you can log in to the BMC again.

After entering incorrect passwords at least 5 times, you will be locked out for 10 minutes before you can attempt to log in again. During this 10-minute period, if you enter an incorrect password again, the lockout period is extended.

The lockout period is calculated in the following way:

**10 minutes \* {failure count - 5 tolerances}**

For example, if you entered incorrect passwords 7 times, you must wait for 20 minutes (**10 \* {7 - 5}**) before you can log in again.



**Important:** The start of the current lockout period is calculated from the **latest** incorrect password that you entered and **not** from the first lock time.

If you need to unlock users immediately, you must reset the BMC.

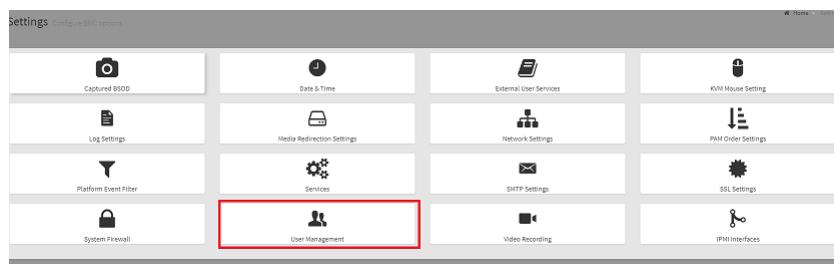
## 3.7. Common BMC Tasks

This section provides information about the most common tasks you can complete in the BMC.

### 3.7.1. Configuring the BMC Login Credentials

Here is some information about how to add or remove users from the BMC.

1. Log into the BMC.
2. In the left navigation pane, click **Settings**.
3. Click the **User Management** card.



- Click the **Help** icon for additional information about how to configure users and create a password.



**Important:** The password must be at least 13 characters long.

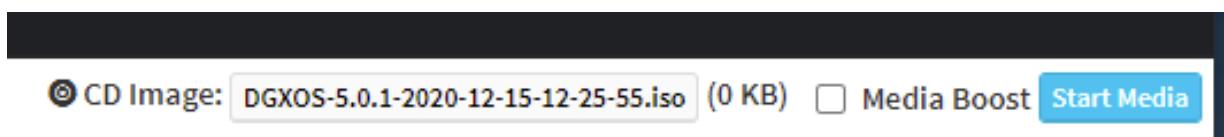
- Log out and log back in with the new credentials.

### 3.7.2. Using the Remote Control

Here is information about how to start the remove KVM and access the DGX A100 Station console.



**Important:** If you select the **Media Boost** checkbox, the processes that are related to media redirection will have a higher priority than other processes.

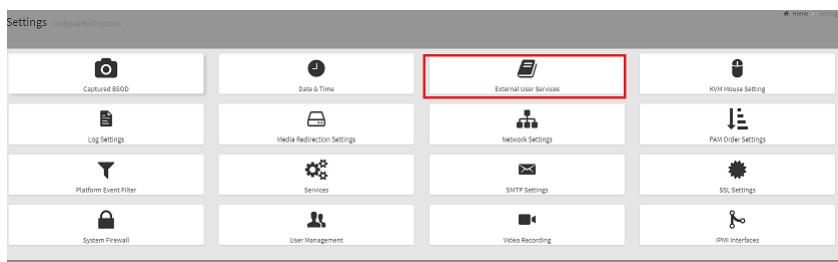


- In the left navigation, click **Remote Control**.
- Click **Launch KVM**.

### 3.7.3. Setting Up Active Directory or LDAP/E-Directory

Here is some information about setting up Active Directory or LDAP/E-Directory in BMC.

- In the left navigation pane, click **Settings**.
- Click **External User Services**.



- Click one of the following options and follow the instructions:



## 3.7.4. Configuring Platform Event Filters

Here is some information about how to configure event filters.

In the left navigation pane, click **Settings > Platform Event Filters**.

The Event Filters page shows all configured event filters and available slots. You can modify or add new event filter entry on this page.

- ▶ To view available configured and unconfigured slots, click **All** in the upper-left corner of the page.
- ▶ To view available configured slots, click **Configured** in the upper-left corner of the page.
- ▶ To view available unconfigured slots, click **UnConfigured** in the upper-left corner of the page.
- ▶ To delete an event filter from the list, click the **x** icon.

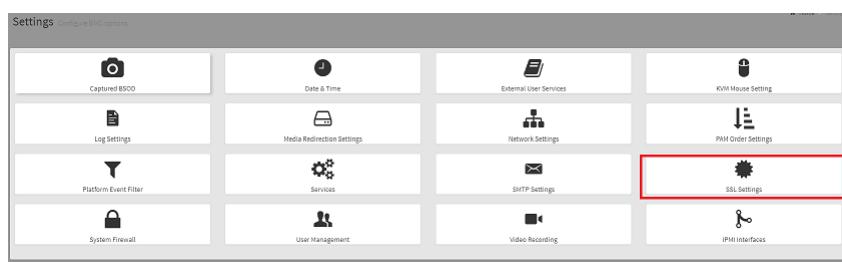
## 3.7.5. Uploading or Generating SSL Certificates

Here is some information about how you can upload or generate SSL certificates.

You can set up a new SSL certificate in one of the following ways:

- ▶ Generate a self-signed SSL.
- ▶ Upload an SSL, for example, to use a Trusted CA-signed certificate.

In the left navigation pane, click **Settings > External User Services**.



### 3.7.5.1. Viewing the SSL Certificate

You can display the SSL certificate that you generated or uploaded.

On the SSL Setting page, click **View SSL Certificate**.



The View SSL Certificate page displays the following basic information about the uploaded SSL certificate:

- ▶ Certificate Version, Serial Number, Algorithm, and Public Key
- ▶ Issuer information

- ▶ Valid Date range
- ▶ Issued to information

### 3.7.5.2. Generating the SSL Certificate

Here are the steps to generate an SSL certificate.

1. In the SSL Settings page, click **Generate SSL Certificate**.



2. Enter the appropriate information as described in the following table:

**Table 2. Fields to Generate an SSL Certificate**

Item	Description
Common Name (CN)	<p>The common name for which the certificate is to be generated.</p> <ul style="list-style-type: none"> <li>▶ Maximum length of 64 alpha-numeric characters.</li> <li>▶ The special characters, # and \$ are not allowed.</li> </ul>
Organization (O)	<p>The name of the organization for which the certificate is generated.</p> <ul style="list-style-type: none"> <li>▶ Maximum length of 64 alpha-numeric characters.</li> <li>▶ The special characters, # and \$ are not allowed.</li> </ul>
Organization Unit (OU)	<p>Overall organization section unit name for which the certificate is generated.</p> <ul style="list-style-type: none"> <li>▶ Maximum length of 64 alpha-numeric characters.</li> <li>▶ The special characters, # and \$ are not allowed.</li> </ul>
City or Locality (L)	<p><b>(Mandatory)</b> City or Locality of the organization</p> <ul style="list-style-type: none"> <li>▶ Maximum length of 64 alpha-numeric characters.</li> <li>▶ The special characters, # and \$ are not allowed.</li> <li>▶ Maximum length of 64 alpha-numeric characters.</li> </ul>

Item	Description
	<ul style="list-style-type: none"> <li>▶ The special characters, # and \$ are not allowed.</li> </ul>
State or Province (ST)	<p><b>(Mandatory)</b> State or Province of the organization</p> <ul style="list-style-type: none"> <li>▶ Maximum length of 64 alpha-numeric characters.</li> <li>▶ The special characters, # and \$ are not allowed.</li> </ul>
Country (C)	<p>Country code of the organization.</p> <ul style="list-style-type: none"> <li>▶ Only two characters are allowed.</li> <li>▶ Special characters are not allowed.</li> </ul>
Email Address	<b>(Mandatory)</b> Email address of the organization
Valid for	<ul style="list-style-type: none"> <li>▶ Validity of the certificate.</li> <li>▶ Enter a range from 1 to 3650 (days)</li> </ul>
Key Length	The key length bit value of the certificate (for example, 2048 bits)

3. To generate the certificate, click **Save**.

### 3.7.5.3. Uploading the SSL Certificate

In BMC, you can upload your SSL certificate.

Make sure the certificate and key meet the following requirements:

- ▶ SSL certificates and keys must both use the .pem file extension.
- ▶ Private keys must not be encrypted.
- ▶ SSL certificates and keys must each be less than 3584 bits in size.
- ▶ SSL certificates must be current (not expired).

1. On the SSL Setting page, click **Upload SSL Certificate**.



2. Click the **New Certificate** folder icon, browse to locate the appropriate file, and select it.
3. Click the **New Private Key** folder icon, browse and locate the appropriate file, and select it.
4. Click **Save**.

---

# Chapter 4. Enable MIG Mode in DGX Station A100

Here is some information about how you can enable the Multi-Instance GPU (MIG) mode.

1. By default, MIG mode is not enabled on the DGX Station A100.

For example, when you run `nvidia-smi`, the output shows that MIG mode is disabled:

```
$ nvidia-smi -i 0
+-----+
| NVIDIA-SMI 450.80.02      Driver Version: 450.80.02      CUDA Version: 11.0      |
+-----+
| GPU  Name      Persistence-M| Bus-Id      Disp.A  | Volatile Uncorr. ECC | | | | |
| Fan  Temp     Perf  Pwr:Usage/Cap| Memory-Usage | GPU-Util  Compute M.  |
| |          |          |              |             |          |          MIG M. |
+-----+
| 0  A100-SXM4-40GB     Off | 00000000:36:00.0 Off |          0 | | | | |
| N/A   29C     P0    62W / 400W |          0MiB / 40537MiB |       6%     Default |
| |          |          |              |             |          |          Disabled |
+-----+
```

2. To enable the MIG mode for each GPU, run the `nvidia-smi -i <GPU IDs> -mig 1` command.
3. Select the GPUs by using comma-separated GPU indexes, PCI Bus Ids, or UUIDs.

Here is some information to remember:

- ▶ If you do not specify a GPU ID, the MIG mode is applied to all the GPUs on the system.

```
$ sudo nvidia-smi -i 0 -mig 1
Enabled MIG Mode for GPU 00000000:36:00.0
All done.

$ nvidia-smi -i 0 --query-gpu=pci.bus_id,mig.mode.current --format=csv
pci.bus_id, mig.mode.current
00000000:36:00.0, Enabled
```

- ▶ If you are using MIG in a VM with GPU passthrough, you might need to reboot the VM to allow the GPU to be in MIG mode.

Sometimes, for security reasons, the GPU reset is not allowed via the hypervisor. Here is an example:

```
$ sudo nvidia-smi -i 0 -mig 1
Warning: MIG mode is in pending enable state for GPU 00000000:00:03.0:Not
Supported
Reboot the system or try nvidia-smi --gpu-reset to make MIG mode effective on
GPU 00000000:00:03.0
All done.
```

```
$ sudo nvidia-smi -i 0 -mig 1
$ sudo nvidia-smi --gpu-reset
Resetting GPU 00000000:00:03.0 is not supported.
```

- ▶ If you have agents on the system, such as monitoring agents that use the GPU, you might not be able to initiate a GPU reset.

On DGX systems, for example, you might encounter the following message:

```
$ sudo nvidia-smi -i 0 -mig 1
Warning: MIG mode is in pending enable state for GPU 00000000:07:00.0:In use
by another client
00000000:07:00.0 is currently being used by one or more other processes (e.g.
CUDA application or a monitoring application such as another instance of
nvidia-smi). Please first kill all processes using the device and retry the
command or reboot the system to make MIG mode effective.
All done.
```

4. Stop the nvsm, dcgm, and gdm3 services, enable MIG mode on the desired GPU, and restore the monitoring services:

```
$ sudo systemctl stop nvsm
$ sudo systemctl stop dcgm
$ sudo systemctl stop gdm3
$ sudo nvidia-smi -i 0 -mig 1
Enabled MIG Mode for GPU 00000000:07:00.0
All done.
```

The examples use super-user privileges. When you grant read access to `mig/config` capabilities, non-root users can also manage instances after the Station A100 has been configured in MIG mode. Refer to [Device Notes](#) for more information.

Here are the default file permissions on the `mig/config` file:

```
$ ls -l /proc/driver/nvidia/capabilities/*
/proc/driver/nvidia/capabilities/mig:
total 0
-r----- 1 root root 0 May 24 16:10 config
-r--r--r-- 1 root root 0 May 24 16:10 monitor
```

To ensure that the MIG instances are available in your containers, restart `nv-docker-gpus` and `docker`:

```
$ sudo systemctl restart nv-docker-gpus
$ sudo systemctl restart docker
```

---

# Chapter 5. Managing Self-Encrypting Drives on DGX Station A100

The DGX OS software supports the ability to manage self-encrypting drives (SEDs), including setting an Authentication Key to lock and unlock DGX Station A100 system drives.

You can manage only SED data drives, and the software cannot be used to manage OS drives, even if the drives are SED-capable.

## 5.1. Overview

The SED management software is provided in the `nv-disk-encrypt` package.

The software supports the following configurations:

- ▶ DGX Station A100 systems where all data drives are self-encrypting drives.
- ▶ Only SEDs used as data drives are supported.

The software will not manage SEDs that are OS drives.

The software provides the following functionality:

- ▶ Identifies eligible drives on the system.
- ▶ Allows you to assign authentication keys (passwords) for each SED as part of the initialization process.
  - ▶ The software can also generate random passwords for each drive.
  - ▶ The passwords are stored in a password-protected vault on the system.
- ▶ Once initialized, SEDs are locked upon power loss, such as a system shutdown or drive removal.  
Locked drives get unlocked after power is restored and the root file system is mounted.
- ▶ Provides functionality to export the vault.
- ▶ Provides functionality for erasing the drives.
- ▶ Provides the ability to revert the initialization.

## 5.2. Installing the nv-disk-encrypt Package

Use the package manager to install the nv-disk-encrypt package.

You can also optionally install the TPM2 tools package and reboot the system. The TPM tools package is required if you plan use the TPM2 to store security keys.

1. Update the packages.

```
$ sudo apt update
```

2. Install the nv-disk-encrypt package.

```
$ sudo apt install -y nv-disk-encrypt
```

3. **Optional:** Install tpm2-tools.

```
$ sudo apt install -y tpm2-tools
```

4. Reboot the system.

```
$ sudo reboot
```

If you plan to use TPM2, ensure that you enable it. See [Enabling the TPM](#) for more information.

## 5.3. Configuring Trusted Computing

This section provides information about how to configure trusted computing.

The DGX Station A100 system BIOS provides setup controls to configure the following Trusted Computing (TC) features:

- ▶ Trusted Platform Module

DGX Station A100 incorporates Trusted Platform Module 2.0 (TPM 2.0), which can be enabled from the system BIOS and used with the nv-disk-encrypt tool.

After this module is enabled, the tool uses the TPM for encryption and then stores the vault and SED authentication keys on the TPM instead of on the file system. Using the TPM is preferred because it allows the vault data to persist even if the system gets re-imaged.

- ▶ Block SID

Certain drives that are shipped with the DGX Station A100 system might support the Block SID authentication feature, which prevents malicious actors from taking ownership of drives and blocking others from using the drives. By default, the DGX BIOS will send the Block SID request.

In these setups, enable the **Disable Block Sid** feature in the BIOS **before** proceeding with the initialization steps.

### 5.3.1. Determining Whether Drives Support Block SID

The drive model is a good indicator of whether the drive supports this feature.

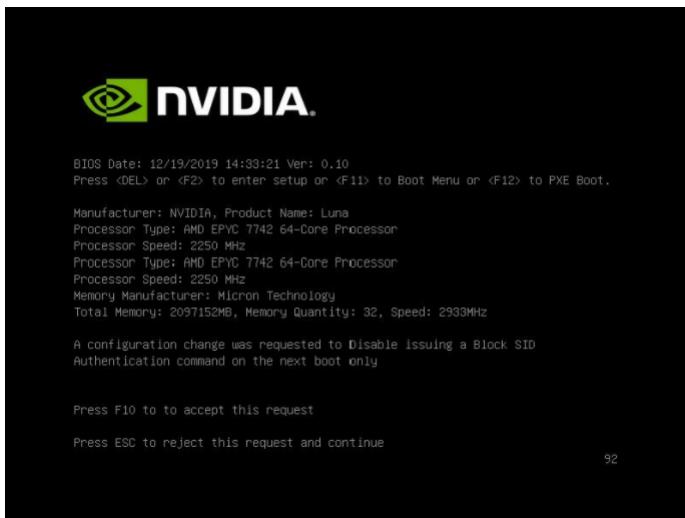
Run the following command and look for the KCM6DRUL3T84 model string.

Node	SN	Model	...
/dev/nvme0n1	30F0A004TN4R	KCM6DRUL3T84	...
/dev/nvme1n1	3070A013TN4R	KCM6DRUL7T68	...
/dev/nvme2n1	S4BGNY0N403766	SAMSUNG MZOLB7T6HMLA-00007	

### 5.3.2. Enabling the TPM and Preventing the BIOS from Sending Block SID Requests

Here are the instructions to enable the TPM and prevent the SBIOS from sending Block SID request. You can select which task to perform because each task is independent of the other.

1. Reboot the DGX Station A100.
2. To enter the BIOS Setup, in the NVIDIA splash screen, press **[Del]** or **[F2]**.
3. On the **Advanced** tab, scroll to **Trusted Computing** and press **[Enter]**.
4. Complete one of the following tasks:
  - ▶ To enable TPM, scroll to **Security Device** and select **Enabled**.
  - ▶ To disable Block SID, scroll to **Disable Block Sid** and select **Enabled**.
5. To continue the boot process, save and exit the BIOS Setup .
6. If you disabled Block SID, you will be prompted to disable issuing a Block SID Authentication command.



7. Press **F10** to confirm.

After the system boots, you can initialize the drive encryption.

## 5.4. Initializing the System for Drive Encryption

Here is some information about how you can initialize your DGX system for drive encryption.



**Note:** Before you initialize drive encryption, see [Configuring Trusted Computing](#) and, if necessary, complete the configuration instructions.

Initialize the system for drive encryption using the nv-disk-encrypt command.

```
$ sudo nv-disk-encrypt init [-k <your-vault-password>] [-f <path/to/ json-file>] [-g] [-r]
```

Here are the options:

- ▶ **-k** lets you create the vault password within the command. Otherwise, the software will prompt you to create a password before proceeding.
- ▶ **-f** lets you specify a JSON file that contains a mapping of passwords to drives.

See [Example 1: Passing in the JSON File](#) for more information.

- ▶ **-g** generates random salt values (stored in /etc/nv-disk-encrypt/.dgxenc.salt) for each drive password. NVIDIA strongly recommends using this option for best security, otherwise the software will use a default salt value instead of a randomly generated one.
- ▶ **-r** generates random passwords for each drive.

This avoids the need to create a JSON file or the need to enter a password one by one during the initialization.

## 5.5. Enabling Drive Locking

Here is some information about how to enable drive locking.

After initializing the system for SED management, use the nv-disk-encrypt command to enable drive locking by issuing the following command:

```
$ sudo nv-disk-encrypt lock
```

After initializing the system and enabling drive locking, when the drives lose power, the drives will be locked. After power is restored to the system, and the system is rebooted, the system will automatically unlock each drive.

## 5.6. Initialization Examples

This section provides information about some initialization examples.

### 5.6.1. Example 1: Passing in the JSON File

This section provides information about a method that specifies the drive/password mapping ahead of time.

By using the method in this example, you can simultaneously initialize several drives at a time, and after issuing the initialization command, and avoid entering a password for each drive.

#### 5.6.1.1. Determining the Drives Can be Managed as Self Encrypting

Review the storage layout of the DGX system to determine which drives are eligible to be managed as SEDs.

```
$ sudo nv-disk-encrypt info
```

The default output shows which drives **can** be used for encryption and which drives **cannot** be used.

The following example output snippet shows drives than **can** be used for encryption. Notice SED capable = Y and Boot disk = N.

```
Disk(s) that can be used for encryption
+-----+-----+
| Name   | Serial | Status
+-----+-----+
| /dev/nvme3n1 | xxxxx1 | SED capable = Y, Boot disk = N, Locked = N, Lock Enabled =
N, MBR done = N |
| /dev/nvme6n1 | xxxxx2 | SED capable = Y, Boot disk = N, Locked = N, Lock Enabled =
N, MBR done = N |
| /dev/nvme9n1 | xxxxx3 | SED capable = Y, Boot disk = N, Locked = N, Lock Enabled =
N, MBR done = N |
```

The following example output shows drives than **cannot** be used for encryption. SED capable = Y and Boot disk = Y, or SED capable = N.

```
Disk(s) that cannot be used for encryption
+-----+-----+
| Name | Serial | Status |
+-----+-----+
| /dev/nvme0n1 | xxxxx1 | SED capable = Y, Boot disk = Y, Locked = N, Lock Enabled = N, MBR done = N |
| /dev/sr0 | xxxxx2 | SED capable = N, Boot disk = N, Locked = N, Lock Enabled = N, MBR done = N |
| /dev/nvme1n1 | xxxxx3 | SED capable = Y, Boot disk = Y, Locked = N, Lock Enabled = N, MBR done = N |
| /dev/sda | unknown | SED capable = N, Boot disk = N, Locked = N, Lock Enabled = N, MBR done = N |
```

You also can specify the output be presented in JSON format by using the **-j** option.

```
$ sudo nv-disk-encrypt info -j
```

The drives that **can** be used for encryption are indicated by the following:

```
"sed_capable": true,
"used_for_boot": false
```

The drives that **can't** be used for encryption are indicated by both of the following options:

```
"sed_capable": true,
"used_for_boot": true
"sed_capable": false,
```

### 5.6.1.2. Creating the Drive/Password Mapping JSON File

You can create the drive/password mapping JSON file and use this file to initialize the system.

1. Create a JSON file that lists all the eligible SED-capable drives that you want to manage.

These are the list of drives that you obtained completing the task in [Determining Which Drives Can be Managed as Self-Encrypting](#).

The following example shows the format of the JSON file:

```
{
"/dev/nvme2n1":
"<your-password>",
"/dev/nvme3n1":
"<your-password>",
"/dev/nvme4n1":
"<your-password>",
"/dev/nvme5n1":
"<your-password>",
}
```



#### Remember:

- ▶ You **must** follow the syntax **exactly**.
- ▶ Passwords must consist of only upper-case letters, lower-case letters, digits, and/or these special characters: ~, :, @, %, ^, +, =, \_, and ,.

## 2. Initialize the system and enable locking.

The following command assumes you have placed the JSON file in the /tmp directory:

```
$ sudo nv-disk-encrypt init -f /tmp/<your-file>.json -g
$ sudo nv-disk-encrypt lock
```

## 3. When prompted, enter a password for the vault.

Passwords must consist of only upper-case letters, lower-case letters, digits, and/or these special characters: ~, :, @, %, ^, +, =, \_, and ,.

## 4. For security purposes, delete the JSON file in the temporary location.

### 5.6.2. Example 2: Generating Random Passwords

This section provides information about how to generate random passwords.

The following command uses the -k and -r options, so you will not be prompted to enter passwords. After you pass the vault password into the command, the command instructs the tool to generate random passwords for each drive.

```
$ sudo nv-disk-encrypt init -k <your-vault-password> -g -r
$ sudo nv-disk-encrypt lock
```

The vault password must consist of only upper-case letters, lower-case letters, digits, and/or these special characters: ~, :, @, %, ^, +, =, \_, and ,.

### 5.6.3. Example 3: Specifying Passwords One at a Time When Prompted

This example provides information about how you can specify passwords one at a time.

If there are a small number of drives or you don't want to create a JSON file, issue the following:

```
$ sudo nv-disk-encrypt init -g
$ sudo nv-disk-encrypt lock
```

The software prompts you to enter a password for the vault, and then a password for each eligible SED.

Passwords must consist of only upper-case letters, lower-case letters, digits, and/or these special characters: ~, :, @, %, ^, +, =, \_, and ,.

## 5.7. Disabling Drive Locking

Here is some information about how to disable drive locking.

To disable drive locking at any time after you initialize, run the following command:

```
$ sudo nv-disk-encrypt disable
```

- ▶ This command disables locking on all drives.
- ▶ You can run the initial set up again at any time after this process is complete.

## 5.8. Exporting the Vault

Here is some information about how you can export the vault.

To export all drive keys to a file, use the `export` function.



**Tip:** When you run this command, you must include the vault password.

```
$ sudo nv-disk-encrypt export -k <your-vault-password>
```

The `/tmp/secrets.out` file contains the mapping of disk serial numbers to drive passwords.

## 5.9. Erasing Your Data

Explain the benefits of the task, the purpose of the task, who should perform the task, and when to perform the task in 50 words or fewer.

Stop `cachefilesd` and unmount the RAID array



**CAUTION:** When you complete this task, **all data** will be lost. On DGX Station A100 systems, these drives generally form a RAID 0 array, which will also be destroyed when you perform an erase.

After you initialize the system for SED management, use the `nv-disk-encrypt` command to erase data on your drives .

1. To completely stop the RAID, issue the following commands:

```
$ systemctl stop cachefilesd
$ sudo umount /raid
$ sudo mdadm --stop /dev/md1
```

2. Complete the erase.

```
$ sudo nv-disk-encrypt erase
```

This command does the following:

- ▶ Sets the drives in an unlocked state.
- ▶ Disables locking on the drives.
- ▶ Removes the RAID 0 array configuration.

3. To rebuild the RAID, issue the following command:

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

## 5.10. Clearing the TPM

If you lost your TPM password, you cannot access the TPM contents. The only way to access TPM again is to clear the contents. After clearing the TPM, you need to initialize the vault and SED authentication keys again.

To clear the TPM, complete the following steps:

1. Reboot the DGX Station A100.
2. To enter the BIOS Setup, in the NVIDIA splash screen, press **[Del]** or **[F2]**.
3. On the **Advanced** tab, scroll to **Trusted Computing** and press **[Enter]**.
4. Clear TPM2.
5. Scroll to **Trusted Computing** and press **[Enter]**.
6. Scroll to **Pending Operation** and press **[Enter]**.
7. In the Pending Operation dialog box, select **TPM Clear**, and then press **[Enter]**.
8. Save and exit the BIOS Setup.

## 5.11. Changing Disk Passwords, Adding Disks, or Replacing Disks

The steps in this process can be used to change or rotate passwords, add disks, or replace disks.

1. Disable SED management.
2. Add or replace drives as needed and then rebuild the RAID array.  
Refer to [Recreating the Cache RAID 0 Volume](#) for instructions.
3. To enable SED management and assign passwords, see [Initializing the System for Drive Encryption](#).

## 5.12. Recovering a Lost Key

NVIDIA recommends that you back up your keys and store the keys in a secure location.

If you lose the key that was used to initialize and lock your drives, you cannot unlock the drive. The only way to recover is to perform a `factory-reset`, which will result in data loss.

SED drives come with a PSID printed on the label. This value can only be obtained by physically examining the drive as shown in the following image.



To specify the PSID and reset the drive, run the following `sedutil-cli` command:

```
$ sudo sedutil-cli --yesIreallywanttoERASEALLmydatausingthePSID <your- drive-PSID> /dev/nvme3n1
```

---

# Chapter 6. Unpacking and Repacking the DGX Station A100

This section provides information about unpacking the DGX Station A100 and repacking it for shipment.

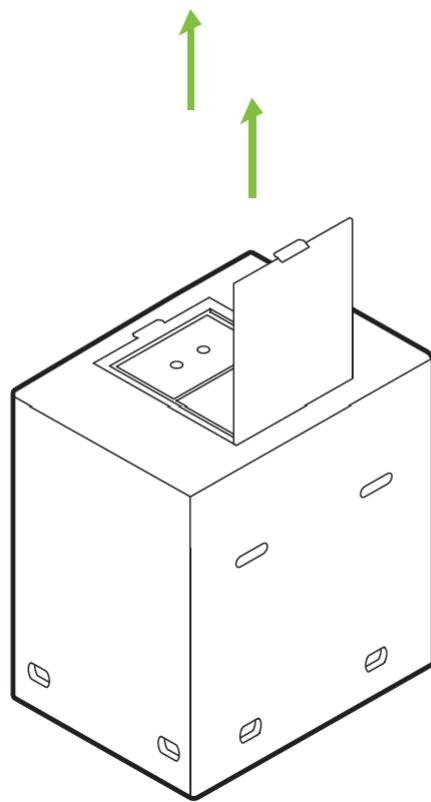
## 6.1. Unpacking the DGX Station A100

After you receive your DGX Station A100, carefully unpack it.

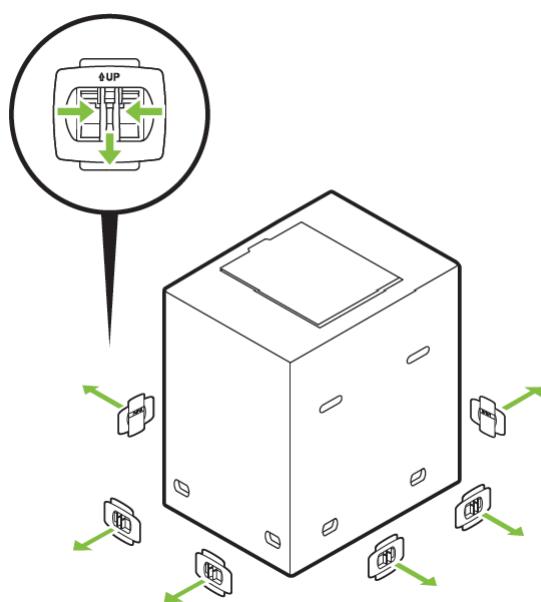


**CAUTION:** The DGX Station A100 weighs 91 lbs (43.1 kg). Do not attempt to lift the DGX Station A100. Instead, move it into position by rolling it on its fitted casters.

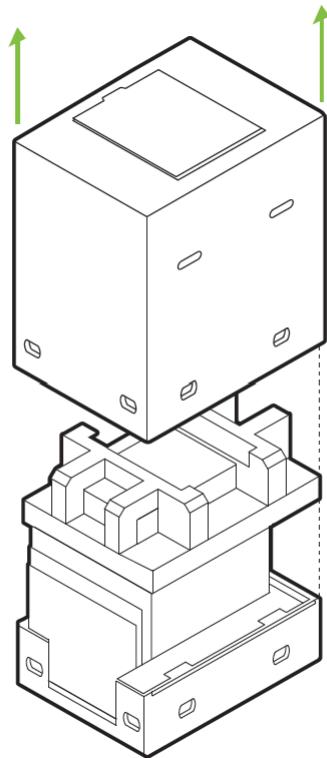
1. Open the top flap and remove the Accessory box and the power cord box.



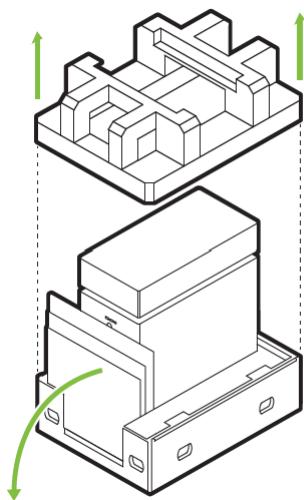
2. Disengage and remove the packing clasps from the cutouts in the shipping carton.  
Do not use excessive force when removing the clasps to prevent them from becoming jammed inside the shipping carton.



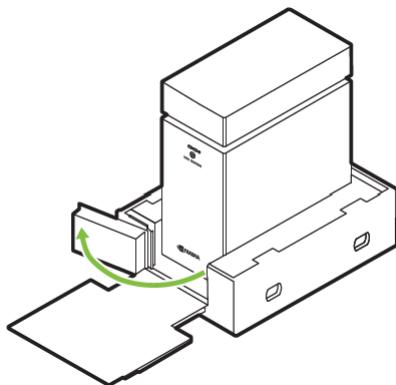
3. Raise the top cover of the shipping carton.



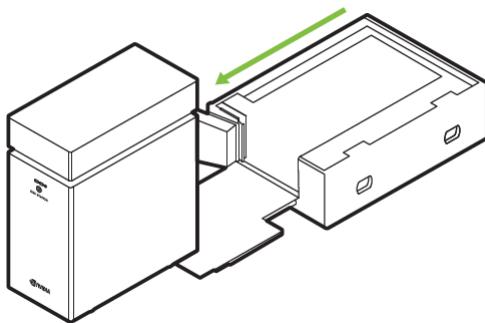
4. Fold down the ramp at the front of the bottom tray of the DGX Station A100 shipping carton and remove the packing material from the top.



5. Swing the door open.



6. Roll the DGX Station A100 off of the packaging by using the ramp and carefully roll the DGX Station A100 down the ramp.



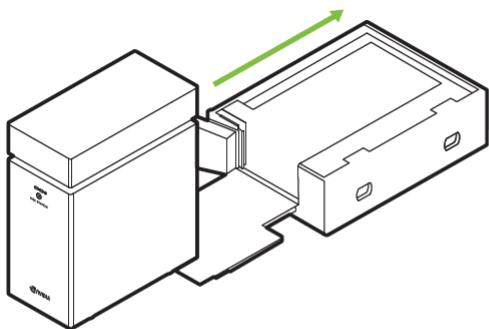
## 6.2. Repacking the DGX Station A100 for Shipment

If you are returning the DGX Station A100 to NVIDIA under an RMA, repack it in the packaging in which the replacement unit was advanced shipped to prevent damage during shipment.

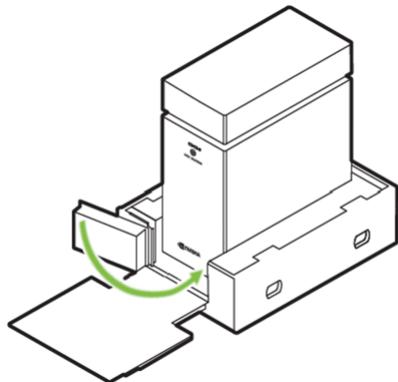


**CAUTION:** The DGX Station A100 weighs 91 lbs (43.1 kg). Do not attempt to lift the DGX Station A100. Instead, move it into position by rolling it on its fitted casters.

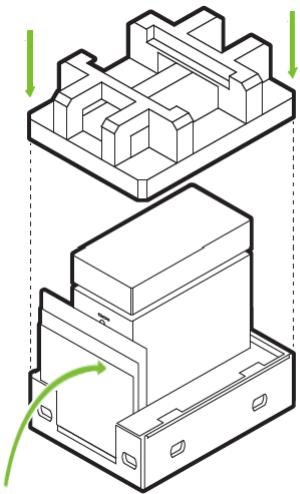
1. Carefully roll the DGX Station A100 up the ramp.



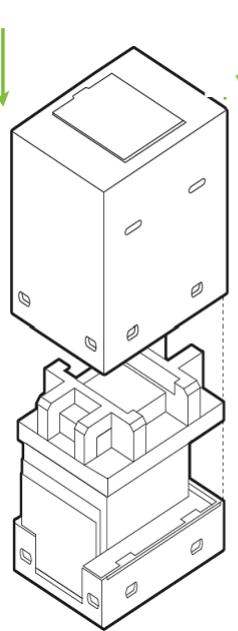
2. Swing the door closed.



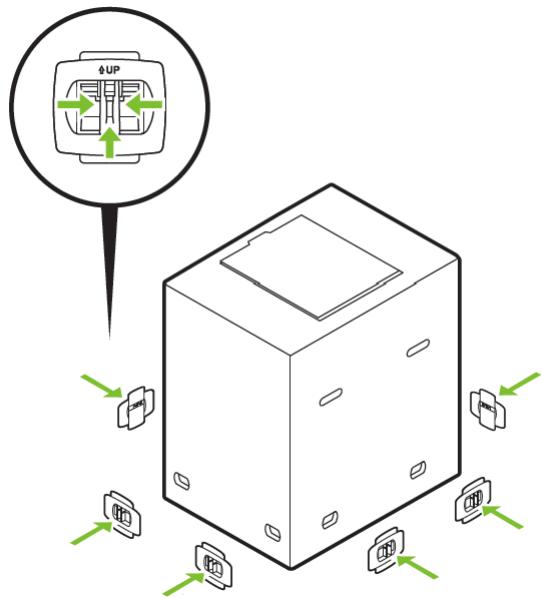
3. Fold up the ramp at the front of the bottom tray of the DGX Station A100 shipping carton and place the packing material on top.



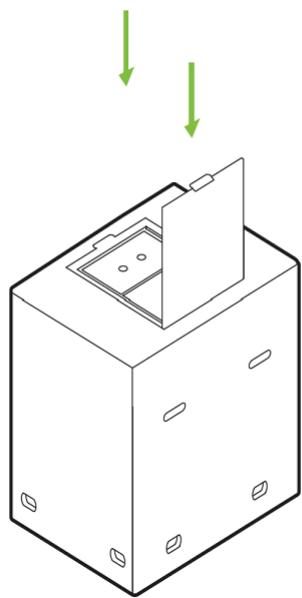
4. Lower the top cover of the shipping carton.



5. Re-engage and insert the packing clasps into the cutouts in the shipping carton.  
Do not use excessive force when inserting the clasps to prevent them from becoming jammed inside the shipping carton.



6. Close the top flap and place the Accessory box and the power cord box.



---

# Appendix A. Security

This section provides information about security in the DGX Station A100.

## Changing Your BMC Credentials

The DGX Station A100 comes with an embedded Baseboard Management Controller (BMC). It enables remote access and control of the workstation for authorized users. From the factory, the BMC ships with a default username and password (admin/admin), and for security reasons, you **must** change these credentials **before** you plug a network cable into the port on the rear IO panel.

To simplify this process, the first time the system is booted, you will be prompted for a username and password to configure the new BMC user and password. This process also disables the default user and scramble its password so it cannot be used to access the BMC.

The username for the new BMC user will be the same as the username that you entered for the operating system administrative user. The password will be unique because it is entered separately during the operating system configuration process.

## Plugging in the Network Cable

**After** you complete the initial configuration of the system, and the BMC username and password have been entered, complete the following steps:

1. Remove the yellow sticker from the rear IO panel of the DGX Station A100.
2. Remove and remove the dust cover to expose the Ethernet RJ45 port.
3. Plug in a network cable and the BMC will obtain an IP address from DHCP.
  - ▶ To configure a static IP address for the BMC, see [Configuring a Static IP Address for the BMC](#).
  - ▶ To log in to the BMC, see [Logging into the BMC](#).

See [Using the BMC](#) for more information about the BMC.

---

# Appendix B. Safety

To reduce the risk of bodily injury, electrical shock, fire, and equipment damage, read this document and observe all warnings and precautions in this guide before installing or maintaining your product. NVIDIA products are designed to operate safely when installed and used according to the product instructions and general safety practices. The guidelines included in this document explain the potential risks associated with computer operation and provide important safety practices designed to minimize these risks.

The product is designed and tested to meet IEC 60950-1 and IEC 62368-1, the Standard for the Safety of Information Technology Equipment. This also covers the national implementation of IEC 60950-1 or IEC 62368-1-based safety standards around the world, for example, UL 62368-1. These standards reduce the risk of injury from the following hazards:

- ▶ **Electric shock:** Hazardous voltage levels contained in parts of the product
- ▶ **Fire:** Overload, temperature, material flammability
- ▶ **Mechanical:** Sharp edges, moving parts, instability
- ▶ **Energy:** Circuits with high energy levels (240 volt amperes) or potential as burn hazards
- ▶ **Heat:** Accessible parts of the product at high temperatures
- ▶ **Chemical:** Chemical fumes and vapors
- ▶ **Radiation:** Noise, ionizing, laser, ultrasonic waves

Retain and follow all product safety and operating instructions. Always refer to the documentation supplied with your equipment. Observe all warnings on the product and in the operating instructions.



**WARNING: FAILURE TO FOLLOW THESE SAFETY INSTRUCTIONS COULD RESULT IN FIRE, ELECTRIC SHOCK OR OTHER INJURY OR DAMAGE. ELECTRICAL EQUIPMENT CAN BE HAZARDOUS IF MISUSED. OPERATION OF THIS PRODUCT, OR SIMILAR PRODUCTS, MUST ALWAYS BE SUPERVISED BY AN ADULT. DO NOT ALLOW CHILDREN ACCESS TO THE INTERIOR OF ANY ELECTRICAL PRODUCT AND DO NOT PERMIT THEM TO HANDLE ANY CABLES.**

## B.1. Intended Application Uses

This product was evaluated as Information Technology Equipment (ITE), which may be installed in offices, schools, computer rooms, and similar commercial type locations. The suitability of this product for other product categories and environments (such as medical,

industrial, residential, alarm systems, and test equipment), other than an ITE application, may require further evaluation.

## B.2. General Precautions

To reduce the risk of personal injury or damage to the equipment:

- ▶ Shut down the product and disconnect all AC power cables before installation.
- ▶ Do not connect or disconnect any cables when performing installation, maintenance, or reconfiguration of this product during an electrical storm.
- ▶ Never turn on any equipment when there is evidence of fire, water, or structural damage.
- ▶ Place the product away from radiators, heat registers, stoves, amplifiers, or other products that produce heat.
- ▶ Never use the product in a wet location.
- ▶ Avoid inserting foreign objects through openings in the product.
- ▶ Do not use conductive tools that could bridge live parts.
- ▶ Do not make mechanical or electrical modifications to the equipment.
- ▶ Use the product only with approved equipment.
- ▶ Follow all cautions and instructions marked on the equipment. Do not attempt to defeat safety interlocks (where provided).
- ▶ Operate the DGX Station A100 in a place where the temperature is always in the range 10°C to 35°C (50°F to 95°F).

## B.3. Electrical Precautions

### Power Cable

To reduce the risk of electric shock, fire, or damage to the equipment:

- ▶ Use only the supplied power cable and do not use this power cable with any other products or for any other purpose. Not all power cables have the same current ratings.
- ▶ Do not use household extension cables with your product. Household extension cables do not have overload protection and are not intended for use with computer systems.
- ▶ If you lose or damage the supplied power cable, or have to change the power cable for any reason, use a cable rated for your product and for the voltage and current marked on the electrical ratings label of the product. The voltage and current rating of the cable must be greater than the voltage and current rating marked on the product.
- ▶ Plug the power cable into a grounded (earthed) electrical outlet that is easily accessible at all times. The product is equipped with a three-wire electrical grounding-type plug which has a third pin for ground. This plug fits only into a grounded electrical power outlet.

- ▶ Do not disable the power cable grounding plug. The grounding plug is an important safety feature.
- ▶ Do not place objects on power cables. Arrange them so that no one may accidentally step on or trip over them.
- ▶ Do not pull on a cable. When unplugging the product from the electrical outlet, grasp the plug.
- ▶ When possible, use one hand only to connect or disconnect cables.
- ▶ Do not modify power cables or plugs. Consult a licensed electrician or your power company for site modifications.

## Power Supply

- ▶ Ensure that the voltage and frequency of your power source match the voltage and frequency inscribed on the equipment's electrical rating label. If you have a question about the type of power source to use, contact your authorized service provider.
- ▶ Connect the equipment to a properly wired and grounded electrical outlet and always follow your local or national wiring rules.
- ▶ Ensure that the socket outlet is near the equipment and is readily accessible for disconnection.
- ▶ To help protect your system from sudden, transient increases and decreases in electrical power, consider using a surge suppressor or line conditioner.
- ▶ Never force a connector into a port. Check for obstructions on the port. If the connector and port don't join with reasonable ease, they probably don't match. Make sure that the connector matches the port and that you have positioned the connector correctly in relation to the port.
- ▶ Do not open the power supply. Hazardous voltage, current and energy levels are present inside the power supply. The power supply in this product contains no user-serviceable parts. Return to manufacturer for servicing.

## B.4. Communications Cable Precautions

To reduce the risk of exposure to electrical shock hazards from communications cables:

- ▶ Do not connect communications cables during an electrical storm. There may be a risk of electric shock from lightning.
- ▶ Do not connect or use communications cables in a wet location.
- ▶ Disconnect the communications cables before opening a product enclosure, or touching or installing internal components.

## B.5. Other Hazards

### California Department of Toxic Substances Control

Perchlorate Material – special handling may apply. See [www.dtsc.ca.gov/hazardouswaste/perchlorate](http://www.dtsc.ca.gov/hazardouswaste/perchlorate).

Perchlorate Material: Lithium battery (CR2032) contains perchlorate. Please follow instructions for disposal.

### Nickel



The decorative metal foam on the DGX Station A100 casework contains some nickel. The metal foam is not intended for direct and prolonged skin contact. While nickel exposure is unlikely to be a problem, you should be aware of the possibility in case you're susceptible to nickel-related reactions.



#### CAUTION:

There is the danger of explosion if the battery is incorrectly replaced. When replacing the battery, use only the battery recommended by the equipment manufacturer. Dispose of batteries according to local ordinances and regulations. Do not attempt to recharge a battery.

Do not attempt to disassemble, puncture, or otherwise damage a battery.

更換電池警告:

警告

更換不正確之電池型式會有爆炸  
的風險

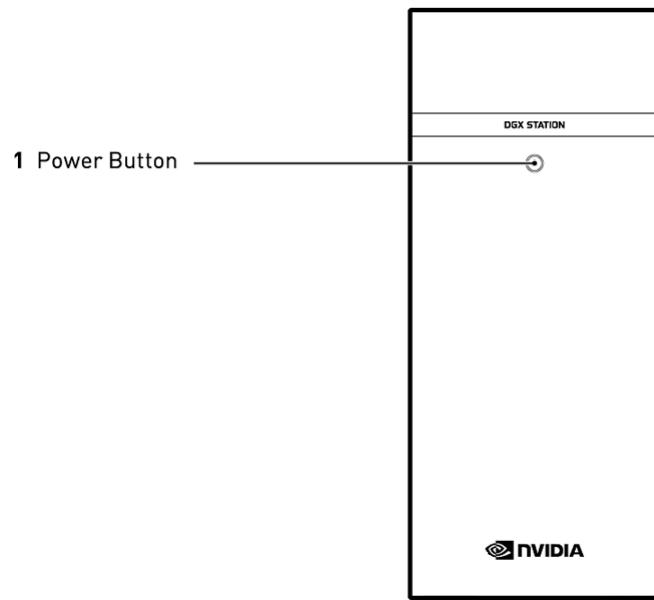
請依製造商說明書處理用過之電  
池。

---

# Appendix C. Connections, Controls, and Indicators

## C.1. Front-Panel Connections and Controls

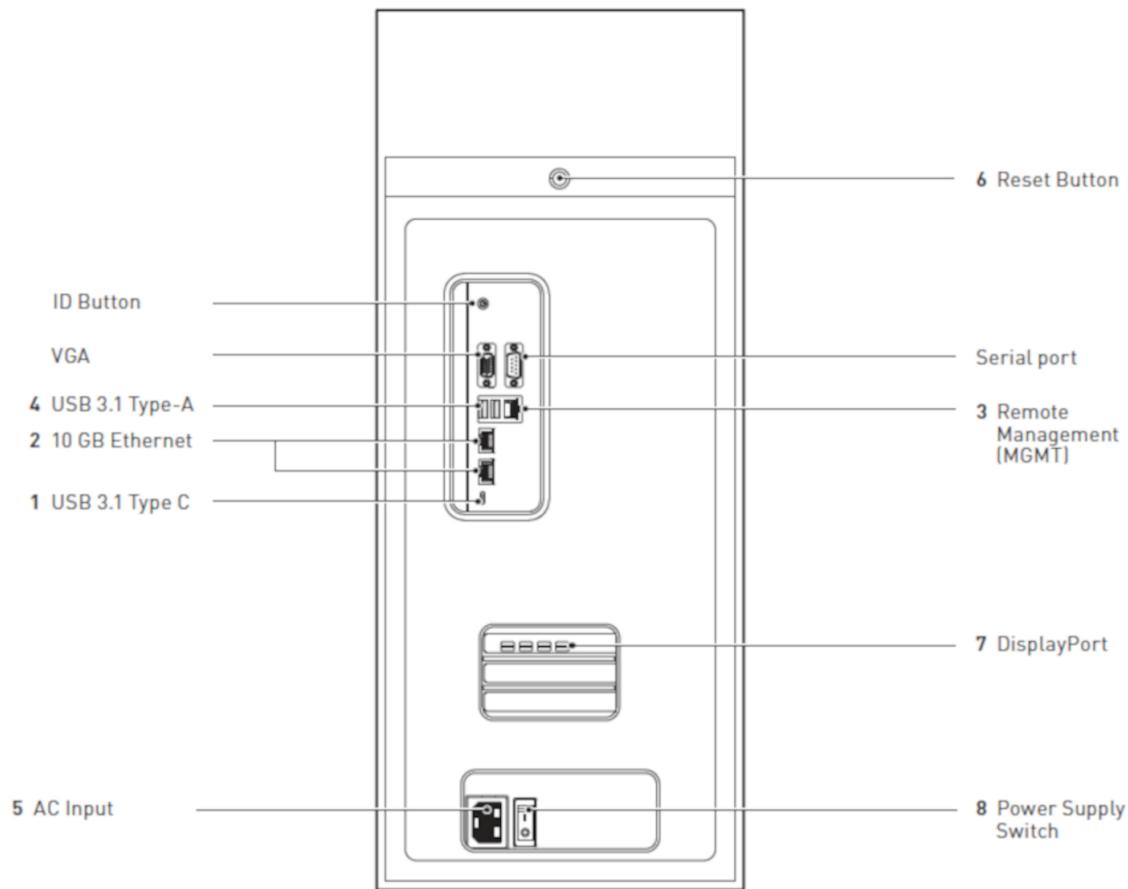
ID	Type	Qty	Description
1	Power Button	1	Press to turn the DGX Station A100 on or off.



## C.2. Rear-Panel Connections and Controls

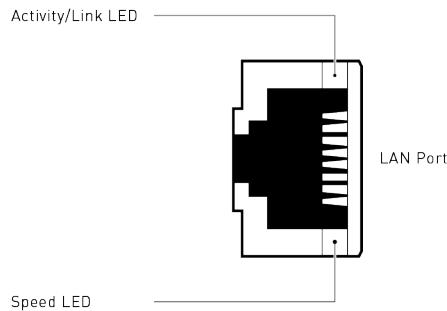
### Current Units

ID	Type	Qty	Description
1	USB 3.1 Type-C	1	USB 3.1 Type-C port
2	Ethernet	2	RJ45 10G LAN ports
	Remote Management	1	RJ45 1G LAN dedicated BMC Management (MGMT)
3	USB 3.1 x 2 Type-A	1	2 x USB 3.1 Type A port
6	AC Input	1	Power supply input
7	Reset Button	1	Press to reboot the system without turning off the system power
10	DisplayPort	3	Ports for connecting up to 3 displays
11	Power Supply Switch	1	Turn the power supply on and off



### C.3. LAN Port Indicators

LEDs on each Ethernet LAN port indicate the connection status as illustrated in the following figure and described in the following tables.



## Speed LED

Status	Description
Off	100 Mbps connection
Orange	1 Gbps connection
Green	10 Gbps connection

## Activity/Link LED

Status	Description
Off	No link
Green	Linked
Green (blinking)	Data activity

---

# Appendix D. Compliance

The NVIDIA DGX Station™ A100 is compliant with the regulations listed in this section.

## D.1. DGX Station A100 Model Number

Model: P3487

## D.2. Australia/New Zealand

Australian Communications and Media Authority (RCM)



This product meets the applicable EMC requirements for Class A, I.T.E equipment

## D.3. Brazil

Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO)



## D.4. Canada

Innovation, Science and Economic Development Canada (ISED)

CAN ICES-003(A)/NMB-003(A)

The Class A digital apparatus meets all requirements of the Canadian Interference-Causing Equipment Regulation.

Cet appareil numérique de la classe A respecte toutes les exigences du Règlement sur le matériel brouilleur du Canada.

## D.5. China

### RoHS Material Content

部件名称 Parts	有害物质 Hazardous Substances					
	铅 (Pb)	汞 (Hg)	镉 (Cd)	六价铬 (Cr(VI))	多溴联苯 (PBB)	多溴联苯醚 (PBDE)
机箱 Chassis	X	O	O	O	O	O
印刷电路部件 PCA	X	O	O	O	O	O
处理器 Processor	O	O	O	O	O	O
主板 Motherboard	X	O	O	O	O	O
电源设备 Power supply	X	O	O	O	O	O
存储设备 System storage	X	O	O	O	O	O
硬盘驱动器 Hard drive	X	O	O	O	O	O
机械部件 Mechanical parts	X	O	O	O	O	O
线材/连接器 Cables/Connectors	X	O	O	O	O	O
焊接金属 Soldering material	O	O	O	O	O	O
助焊剂、锡膏、标签及其他耗材 Flux, Solder Paste, label and other consumable materials	O	O	O	O	O	O

本表格依据SJ/T 11364-2014 的规定编制  
The table according to SJ/T 11364-2014

**O:** 表示该有害物质在该部件所有均质材料中的含量均在GB/T 26572-2011 标准规定的限量要求以下。

**O:** Indicates that this hazardous substance contained in all of the homogeneous materials for this part is below the limit requirement in GB/T 26572-2011.

**X:** 表示该有害物质至少在该部件的某一均质材料中的含量超出GB/T 26572-2011 标准规定的限量要求。

**X:** Indicates that this hazardous substance contained in at least one of the homogeneous materials used for this part is above the limit requirement in GB/T 26572-2011.

此表中所有名称中含“X”的部件均符合欧盟 RoHS 立法。

All parts named in this table with an “X” are in compliance with the European Union’s RoHS Legislation.

注：环保使用期限的参考标识取决于产品正常工作的温度和湿度等条件

Note: The referenced Environmental Protection Use Period Marking was determined according to normal operating use conditions of the product such as temperature and humidity.

## D.6. European Union

European Conformity; Conformité Européenne (CE)



This is a Class A product. In a domestic environment this product may cause radio frequency interference in which case the user may be required to take adequate measures.

The product has been marked with the CE Mark to illustrate its compliance.

This device complies with the following Directives:

- ▶ EMC Directive (2014/30/EU) for Class A, I.T.E equipment.
- ▶ Low Voltage Directive (2014/35/EU) for electrical safety.
- ▶ RoHS Directive (2011/65/EU) for hazardous substances.
- ▶ ErP Directive (2009/125/EC) for European Ecodesign.

A copy of the Declaration of Conformity to the essential requirements may be obtained directly from NVIDIA GmbH ("Bavaria Towers – Blue Tower, Einsteinstrasse 172, D-81677 Munich Germany")

## D.7. India

Bureau of Indian Standards (BIS)

**IS 13252 (Part 1)/  
IEC 60950-1**



R-xxxxxxxx

[www.bis.gov.in](http://www.bis.gov.in)

Self Declaration - Conforming to IS13252:2010, R-xxxxxxxx

### India RoHS compliance statement:

This product, as well as its related consumables and spares, complies with the reduction in hazardous substances provisions of the “India E-waste (Management and Handling) Rule 2016”.

It does not contain lead, mercury, hexavalent chromium, polybrominated biphenyls or polybrominated diphenyl ethers in concentrations exceeding 0.1 weight % and 0.01 weight % for cadmium, except for where allowed pursuant to the exemptions set in Schedule 2 of the Rule.”

## D.8. Japan

### Voluntary Control Council for Interference (VCCI)



この装置は、クラスA機器です。この装置を住宅環境で使用すると電波妨害を引き起こすことがあります。この場合には使用者が適切な対策を講ずるよう要求されることがあります。

VCCI - A

2008年、日本における製品含有表示方法、JIS C0950が公示されました。製造事業者は、2006年7月1日以降に販売される電気・電子機器の特定化学物質の含有に付きまして情報提供を義務付けられました。製品の部材表示に付きましては、以下をご覧ください。

A Japanese regulatory requirement, defined by specification JIS C 0950, 2008, mandates that manufacturers provide Material Content Declarations for certain categories of electronic products offered for sale after July 1, 2006.

To view the JIS C 0950 material declaration for this product, visit  
<http://www.nvidia.com>

### Japan RoHS Material Content Declaration

日本工業規格JIS C 0950:2008により、2006年7月1日以降に販売される特定分野の電気および電子機器について、製造者による含有物質の表示が義務付けられます。						
部品分類	特定化学物質記号					
	Pb	Hg	Cd	Cr(VI)	PBB	PBDE
筐体	除外項目	0	0	0	0	0
プリント基板	除外項目	0	0	0	0	0
プロセッサー	0	0	0	0	0	0
マザーボード	除外項目	0	0	0	0	0
電源	除外項目	0	0	0	0	0
システムストレージ	除外項目	0	0	0	0	0
ハードディスクドライブ	除外項目	0	0	0	0	0
機械部品	除外項目	0	0	0	0	0
ケーブル/コネクター	除外項目	0	0	0	0	0
はんだ付け材料	0	0	0	0	0	0
ラバックス、クリームはんだ ラベル、その他消耗品	0	0	0	0	0	0

注:

1. [0] は、特定化学物質の含有率が日本工業規格JIS C 0950:2008に記載されている含有率基準値より低いことを示します。

2. 「除外項目」は、特定化学物質が含有マークの除外項目に該当するため、特定化学物質について、日本工業規格JIS C 0950:2008に基づく含有マークの表示が不要であることを示します。

3. 「0.1wt%超」または「0.01wt%超」は、特定化学物質の含有率が日本工業規格JIS C 0950:2008に記載されている含有率基準値を超えることを示します。

## D.9. Mexico

Norma Official Mexicana (NOM)



## D.10. Russia/Kazakhstan/Belarus

Customs Union Technical Regulations (CU TR)



This device complies with the technical regulations of the Customs Union (CU TR)  
 ТЕХНИЧЕСКИЙ РЕГЛАМЕНТ ТАМОЖЕННОГО СОЮЗА О безопасности низковольтного  
 оборудования (TP TC 004/2011) ТЕХНИЧЕСКИЙ РЕГЛАМЕНТ ТАМОЖЕННОГО СОЮЗА  
 Электромагнитная совместимость технических средств (TP TC 020/2011) Технический  
 регламент Евразийского экономического союза "Об ограничении применения опасных  
 веществ в изделиях электротехники и радиоэлектроники" (TP TC 037/2016).

## D.11. South Africa

South African Bureau of Standards (SABS)

This device complies with the following SABS Standards:

SANS 2332: 2017/CISPR 32:2015

SANS 2335:2018/ CISPR 35:2016

National Regulator for Compulsory Specifications (NRCS)

This device complies with following standard under VC 8055

SANS IEC 60950-1

## D.12. South Korea

Korean Certification (KC)



R-R-NVA-P3487

A급 기기  
(업무용 방송통신기자재)

이 기기는 업무용(A급) 전자파적합기기로서 판매자 또는 사용자는 이 점을 주의하시기 바라며, 가정외의 지역에서 사용하는 것을 목적으로 합니다.

## D.13. Taiwan

Bureau of Standards, Metrology & Inspection (BSMI)



R33088  
**RoHS**

警告使用者：

此為甲類資訊技術設備，於居住環境中使用時，可能會造成射頻擾動，在此種情況下，使用者會被要求採取某些適當的對策

**Taiwan RoHS Material Content Declaration**

限用物質含有情況標示聲明書  
Declaration of the presence condition of the Restricted Substances Marking

設備名稱：伺服器  
Equipment Name: P3487

項元 Parts	限用物質及其化學符號 Restricted substances and its chemical symbols					
	鉛 (Pb)	汞 (Hg)	镉 (Cd)	六價铬 (Cr(VI))	多溴聯苯 (PBB)	多溴二英醚 (PBDE)
機殼 Chassis	-	0	0	0	0	0
印刷電路部件 PCA	-	0	0	0	0	0
處理器 Processor	0	0	0	0	0	0
主板 Motherboard	-	0	0	0	0	0
電源設備 Power supply	-	0	0	0	0	0
系統設備 System storage	-	0	0	0	0	0
硬盤驅動器 Hard drive	-	0	0	0	0	0
機械部件 Mechanical parts	-	0	0	0	0	0
線材/連接器 Cables/Connectors	-	0	0	0	0	0
焊接金屬 Soldering material	0	0	0	0	0	0
助焊劑、錫膏、標籤及其他耗材 Flux, Solder Paste, label and other consumable materials	0	0	0	0	0	0

備考1: 0: 指該限用物質未超出百分比含量基準值

Note 1: 0: Indicates that the percentage content of the restricted substance does not exceed the percentage of reference value of presence.

備考2: -: 表指該項限用物質為排除項目。

Note 2: -: indicates that the restricted substance corresponds to the exemption.

此表中所有名稱中含“-”的部件均符合歐盟 RoHS 立法。

All parts named in this table with an “-” are in compliance with the European Union's RoHS Legislation.

注: 環保使用期限的參考標識取決與產品正常工作的溫度和濕度等條件

Note: The referenced Environmental Protection Use Period Marking was determined according to normal operating use conditions of the product such as temperature and humidity.

## D.14. United Kingdom

UK Conformity Assessed (UKCA)



This product complies with following UK regulations:

SI 2016 No. 1091 "The Electromagnetic Compatibility Regulations 2016"

SI 2016 No. 1101 "The Electrical Equipment (Safety) Regulations 2016"

SI 2012 No. 3032 "The Restriction of the Use of Certain Hazardous Substances in Electrical and Electronic Equipment Regulations 2012"

SI 2010 No. 2617 "The Ecodesign for Energy-Related Products Regulations 2010"

## D.15. United States

Federal Communications Commission (FCC)

### FCC Marking (Class A)

This device complies with part 15 of the FCC Rules. Operation is subject to the following two conditions: (1) this device may not cause harmful interference, and (2) this device must accept any interference received, including any interference that may cause undesired operation of the device.

**NOTE:** This equipment has been tested and found to comply with the limits for a Class A digital device, pursuant to part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference when the equipment is operated in a commercial environment. This equipment generates, uses, and can radiate radio frequency energy and, if not installed and used in accordance with the instruction manual, may cause harmful interference to radio communications. Operation of this equipment in a residential area is likely to cause harmful interference in which case the user will be required to correct the interference at his own expense.

## D.16. United States/Canada

TUV Rheinland

**TUV Rheinland**



Energy Star



Energy Star qualified server

---

# Appendix E. DGX Station A100 Hardware Specifications

## E.1. Environmental Conditions

Condition	Operating Range	Nonoperating Range
Ambient temperature	10°C to 35°C (50°F to 95°F)	5°C to 40°C (41°F to 104°F)
Relative humidity	10% to 80% (non-condensing)	10% to 80% (non-condensing)

## E.2. Component Specifications

Component	Qty	Description
CPU	1	Single AMD 7742, 64 cores, 2.25 GHz (base)–3.4 GHz (max boost), and 2.25 GHz (base)–3.4 GHz (max boost)
GPU- current units	4	4x NVIDIA A100, 80 GB GPUs: <ul style="list-style-type: none"><li>▶ 5 petaOPS INT8</li><li>▶ 4×80 GB (320 GB total) GPU memory</li><li>▶ 2.5 petaFLOPS AI</li></ul>
System memory	8	512 GB DDR4
Data storage	1	7.68 TB and a cache/data U.2 NVME drive
OS storage	1	Boot M.2 NVME drive

## E.3. Mechanical Specifications

Specification	Value
Height	26.1" (639 mm)
Width	10.1" (256 mm)
Depth	20.4" (518 mm)

Specification	Value
Gross weight	91 lbs (43.1 kg)

## E.4. Power Specifications



**RESTRICTION:** The power source for DGX Station A100 **must be** 100V and **cannot** fall below 90V.

Input	Comments
100-115VAC/15A, 115-120VAC/12A, 200-240VAC/10A, and 50/60Hz	The DGX Station A100 power consumption can reach 1,500 W (ambient temperature 30°C) with all system resources under a heavy load.  Be aware of your electrical source's power capability to avoid overloading the circuit.

---

## Appendix F. Customer Support for the NVIDIA DGX Station™ A100

Contact NVIDIA Enterprise Support for assistance in reporting, troubleshooting, or diagnosing problems with your DGX Station A100 system. Also contact NVIDIA Enterprise Support for assistance in installing or moving the DGX Station A100 system.

For details on how to obtain support, visit the NVIDIA Enterprise Support web site (<https://www.nvidia.com/en-us/support/enterprise/>).

Our support team can help collect appropriate information about your issue and involve internal resources as needed.

**Copyright**

© 2022 NVIDIA Corporation & Affiliates. All rights reserved.

