DGX A100 System

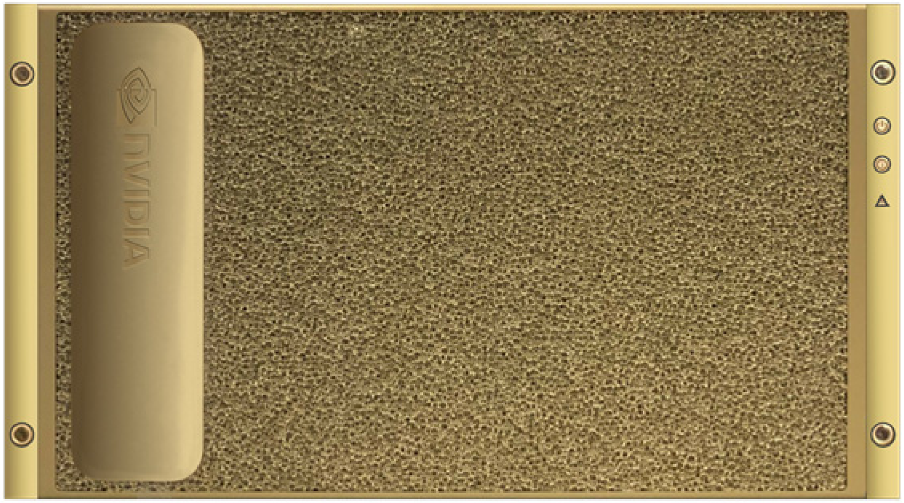
DU-09821-001_v01 | May 2020

User Guide

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

The NVIDIA DGX™ A100 system is the universal system purpose-built for all AI infrastructure and workloads, from analytics to training to inference. The system is built on eight NVIDIA A100 Tensor Core GPUs.



This document is for users and administrators of the DGX A100 system.

1.1 HARDWARE OVERVIEW

1.1.1 Component Specifications

Component	Qty	Description
GPU	8	NVIDIA A100 GPU 40 GB memory per GPU
CPU	2	2x AMD EPYC 7742 CPU w/64 cores
NVLink	12	600 GB/s GPU-to-GPU bandwidth
Storage (OS)	2	1.92 TB NVMe M.2 SSD (ea) in RAID 1 array
Storage (Data Cache)	4 (base config) 8 (optional with 4 additional drives installed)	3.84 TB NVMe U.2 SED (ea) in RAID 0 array
Network (cluster) card	8	Mellanox ConnectX-6 Single Port VPI 200 Gb InfiniBand (default)/Ethernet
Network (storage) card	1 (base config) 2 (optional with additional I/O card installed)	Mellanox ConnectX-6 Dual Port VPI 200 Gb Ethernet (default)/InfiniBand
System Memory (DIMM)	16 (base config) 32 (optional with 16 additional DIMMs installed)	1 TB total memory in base configuration 2 TB total memory in optional configuration
BMC (out-of-band system management)	1	1 GbE RJ45 interface, supports IPMI, SNMP, KVM, HTTPS
In-band system management	1	1 GbE RJ45 interface
Power Supply	6	3 kW

1.1.2 Mechanical Specifications

Feature	Description
Form Factor	6U Rackmount
Height	10.39” (264 mm)
Width	19" (482.3 mm)
Depth	35.32" (897.2 mm)
System Weight	271 lbs (123 kg)

1.1.3 Power Specifications

Input		Specification for Each Power Supply	Comments
200-240 volts AC	6.5 kW max.	3000 W @ 200-240 V, 16 A, 50-60 Hz	The DGX A100 system contains six load-balancing power supplies.

Support for N+N Redundancy

The DGX A100 includes six power supply units (PSU) configured for 3+3 redundancy. If three PSUs fail, the system will continue to operate at full power with the remaining three PSUs.

Note: The DGX A100 will not operate with less than three PSUs.

DGX A100 Locking Power Cords

The DGX A100 is shipped with a set of six (6) locking power cords that have been qualified for use with the DGX A100 to ensure regulatory compliance.

WARNING: To avoid electric shock or fire, do not connect other power cords to the DGX A100. For more details, see **“Electrical Precautions”**

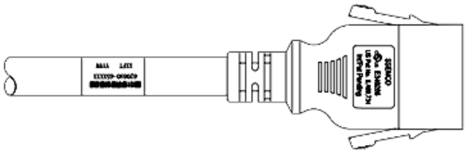
Power Cord Feature	Specification
Electrical	250VAC, 16A
Plug Standard	C19/C20
Dimension	1800mm length
Compliance	Cord: UL62, IEC60227 Connector/Plug: IEC60320-1

Follow these instructions for using the locking power cords.

Power Distribution Unit side

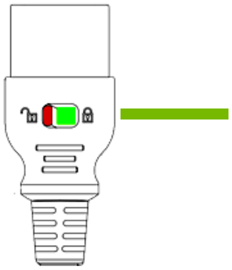
To INSERT, push the cable into the PDU socket

To REMOVE, press the clips together and pull the cord out of the socket

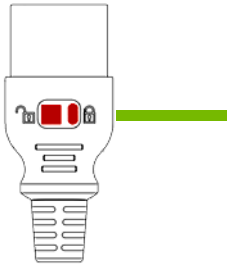


Power Supply (System) side

To INSERT or REMOVE make sure the cable is UNLOCKED and push/pull into/out of the socket



To UNLOCK the power cord, move the switch to the unlocked position (indicator will show GREEN)



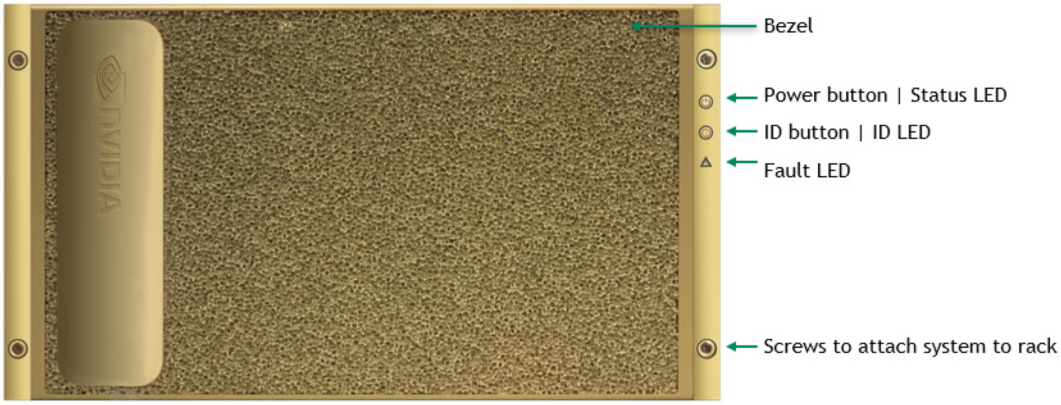
To LOCK the power cord, move the switch to the locked position (indicator should show only RED)

1.1.4 Environmental Specifications

Feature	Specification
Operating Temperature	5° C to 30° C (41° F to 86° F)
Relative Humidity	20% to 80% noncondensing
Airflow	840 CFM @ 80% fan PWM
Heat Output	22,179 BTU/hr

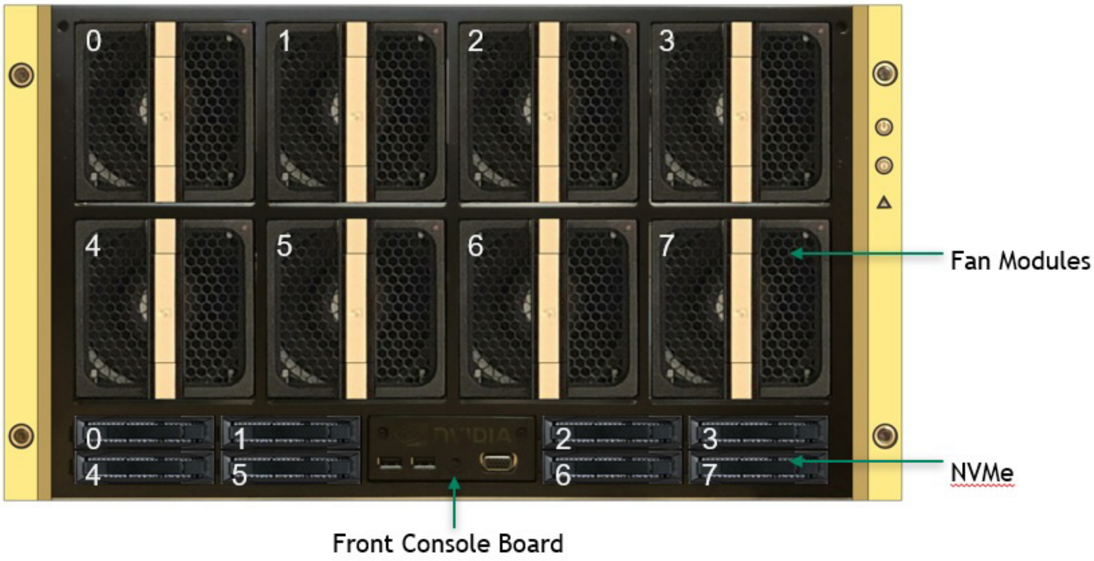
1.1.5 Front Panel Connections and Controls

1.1.5.1 With Bezel



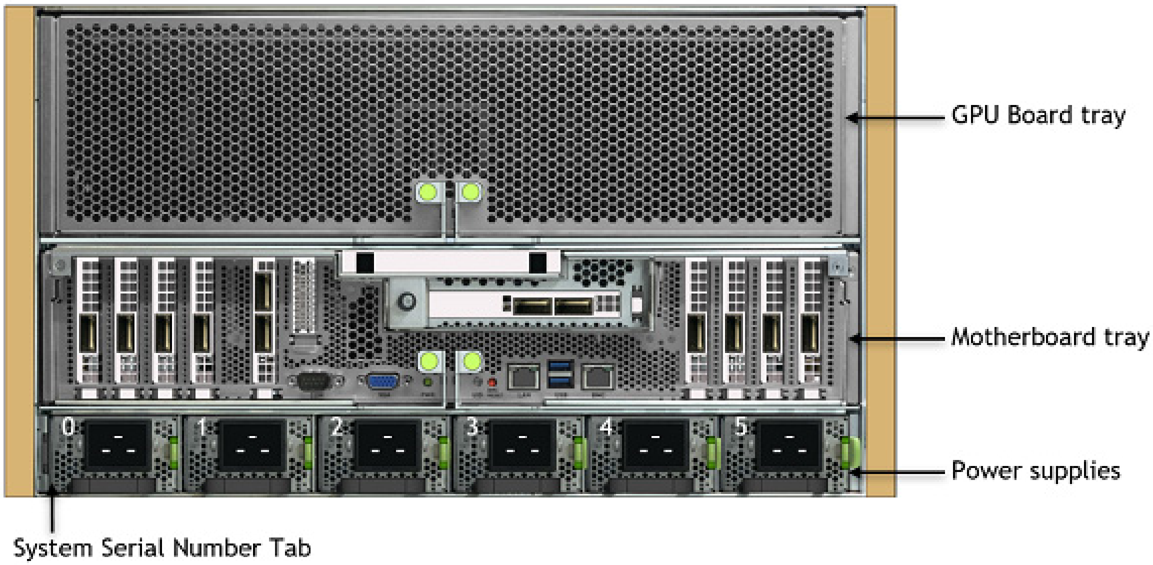
Control	Description
Power Button	Press to turn the DGX A100 system On or Off Green flashing (1 Hz): Standby (BMC booted) Green flashing (4 Hz): POST in progress Green solid On: Power On
ID Button	Press to cause the button blue LED to turn On or blink (configurable through the BMC) as an identifier during servicing. Also causes an LED on the back of the unit to flash as an identifier during servicing.
Fault LED	Amber On: System or component faulted

1.1.5.2 With Bezel Removed

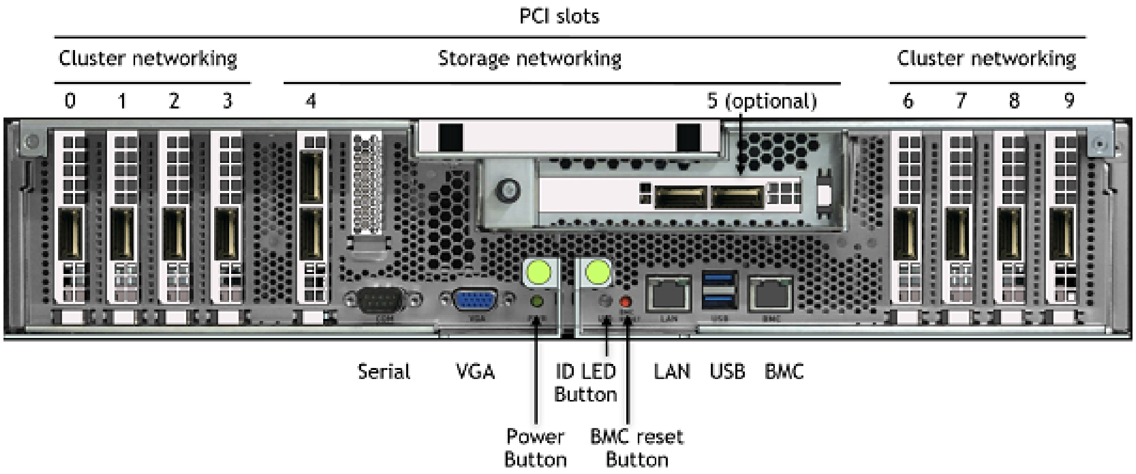


IMPORTANT:See the section "Turning DGX A100 On and Off" for instructions on how to properly turn the system on or off.

1.1.6 Rear Panel Modules



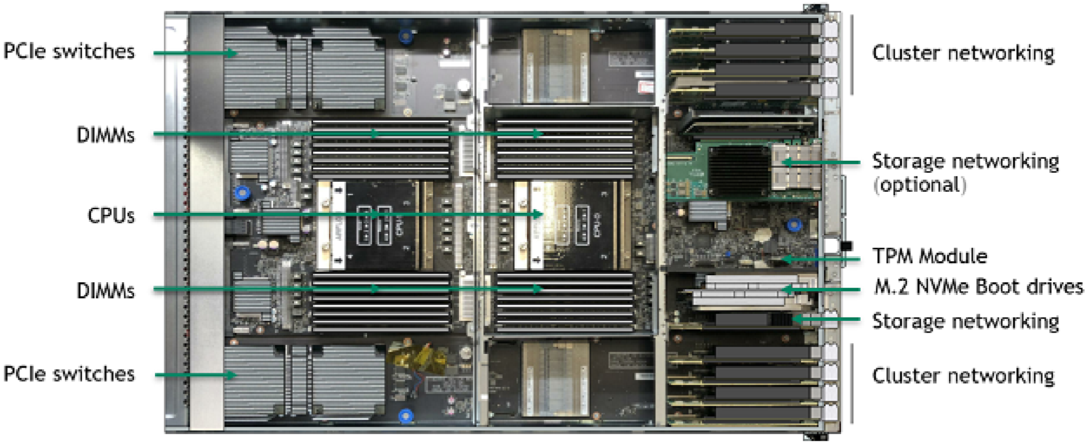
1.1.7 Motherboard Connections and Controls



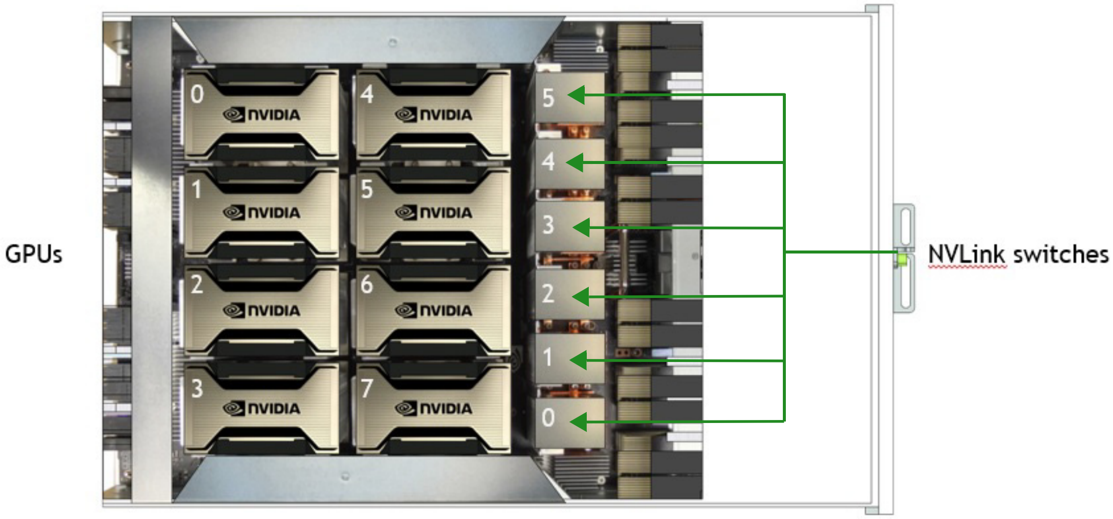
Control	Description
Power Button	Press to turn the system On or Off.
ID LED Button	Blinks when ID button is pressed from the front of the unit as an aid in identifying the unit needing servicing
BMC Reset button	Press to manually reset the BMC

See **"Configuring Network Proxies"** for details on the network connections.

1.1.8 Motherboard Tray Components



1.1.9 GPU Tray Components



1.2 NETWORK CONNECTIONS, CABLES, AND ADAPTORS

1.2.1 Network Ports

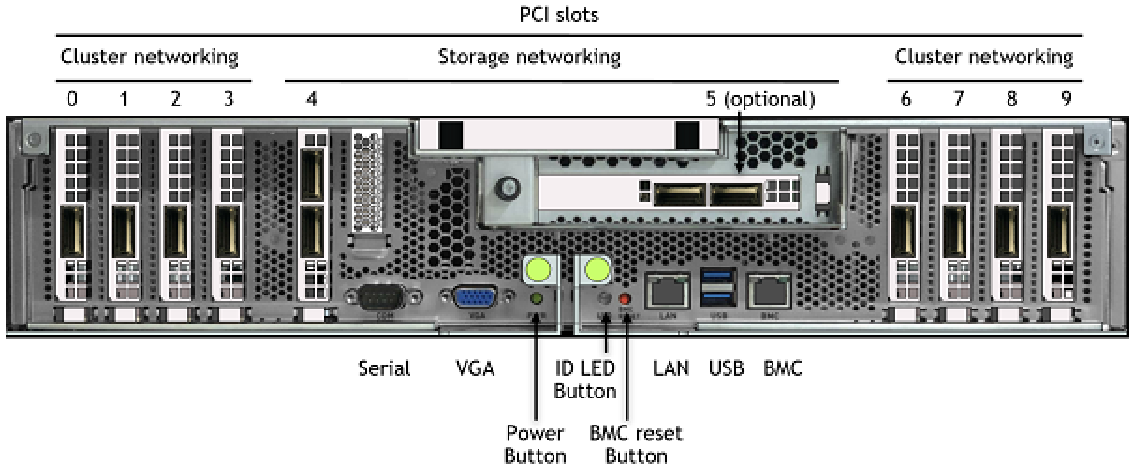


Table 1.1 Network Port Mapping

Slot	PCI Bus	Port Designation	
		Default	Optional
0	4b:00.0	ib2	enp75s0
1	54:00.0	ib3	enp84s0
2	ba:00.0	ib6	enp186s0
3	ca:00.0	ib7	enp202s0
4 port 0 (top)	e1:00.0	enp225s0f0	(See note)
4 port 1 (bottom)	e1:00.1	enp225s0f1	(See note)
5 port 0 (left)	61:00.0	enp97s0f0	(See note
5 port 1 (right)	61:00.1	enp97s0f1	(See note)
6	0c:00.0	ib0	enp12s0
7	12:00.0	ib1	enp18s0
8	8d:00.1	ib4	enp141s0
9	94:00.0	ib5	enp148s0
LAN	e2:00.0	enp226s0	N/A

Note: The InfiniBand port designations may change depending on individual port changes from InfiniBand to Ethernet or vice versa.

The Optional column lists the port designations after reconfiguring the default InfiniBand ports to Ethernet.

1.2.2 Supported Network Cables

The DGX A100 system is not shipped with network cables. You will need to purchase supported cables for your network.

For a list of cables compatible with the Mellanox ConnectX-6 VPI cards installed in the DGX A100 system, see the **Mellanox Firmware Compatible Products** page for the firmware release included in the DGX A100. The ConnectX-6 firmware determines which cables are supported.

1.2.3 Supported Network Adaptors

To connect the DGX A100 system to an existing 10 or 25 GbE network, you can purchase the following adaptors from NVIDIA.

Component	Mellanox MPN	Specification
Ethernet Cable Adapter	MAM1Q00A-QSA	40Gb/s to 10Gb/s, QSFP+ to SFP+
Passive Copper Hybrid Cable	MC2609130-003	Ethernet 40GbE to 4x10GbE, QSFP to 4xSFP+, 3m length
Passive Copper Hybrid Cable	MCP7F00-A003	Ethernet 100GbE to 4x25GbE, QSFP28 to 4xSFP28, 3m length, 28AWG

1.3 DGX OS SOFTWARE

The DGX A100 system comes pre-installed with a DGX software stack incorporating

- An Ubuntu server distribution with supporting packages
- The NVIDIA GPU driver
- Docker Engine
- NVIDIA Container Toolkit
- Mellanox OpenFabrics Enterprise Distribution for Linux (MOFED)
- Mellanox Software Tools (MST)
- cachefilesd (daemon for managing cache data storage)
- DGX A100 system support packages
- The following health monitoring software
  - NVIDIA System Management (NVSM)

    Provides active health monitoring and system alerts for NVIDIA DGX nodes in a data center. It also provides simple commands for checking the health of the DGX A100 system from the command line.
  - Data Center GPU Management (DCGM)

    This software enables node-wide administration of GPUs and can be used for cluster and data-center level management.

1.4 ADDITIONAL DOCUMENTATION

Note: Some of the documentation listed below is not available at the time of publication. See **https://docs.nvidia.com/dgx/** for the latest status.

- **NGC Container Registry for DGX**

  How to access the NGC container registry for using containerized deep learning GPU-accelerated applications on your DGX A100 system.

- **NVSM Software User Guide**

  Contains instructions for using the NVIDIA System Management software.

- **DCGM Software User Guide**

  Contains instructions for using the Data Center GPU Manager software.

1.5 CUSTOMER SUPPORT

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

1.5.1 NVIDIA Enterprise Support Portal

The best way to file an incident is to log on to the **NVIDIA Enterprise Support portal**.

1.5.2 NVIDIA Enterprise Support Email

You can also send an email to **enterprisesupport@nvidia.com**.

1.5.3 NVIDIA Enterprise Support - Local Time Zone Phone Numbers

Visit **NVIDIA Enterprise Support Phone Numbers.**

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

CHAPTER 2 CONNECTING TO THE DGX A100

2.1 CONNECTING TO THE CONSOLE

Connect to the DGX A100 console using either a direct connection or a remote connection through the BMC.

! **CAUTION:Connect directly to the DGX A100 console if the DGX A100 system is connected to a 172.17.xx.xx subnet.**

DGX OS Server software installs Docker Engine which uses the 172.17.xx.xx subnet by default for Docker containers. If the DGX A100 system is on the same subnet, you will not be able to establish a network connection to the DGX A100 system.

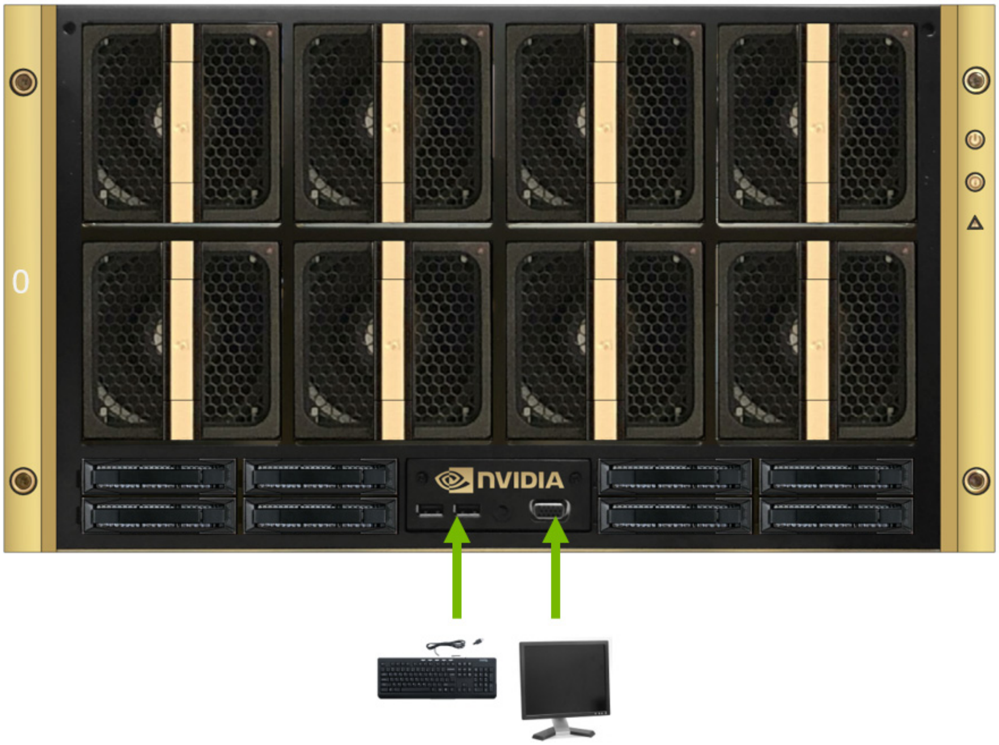
Refer to the section **"Configuring Docker IP Addresses"** for instructions on how to change the default Docker network settings.

2.1.1 Direct Connection

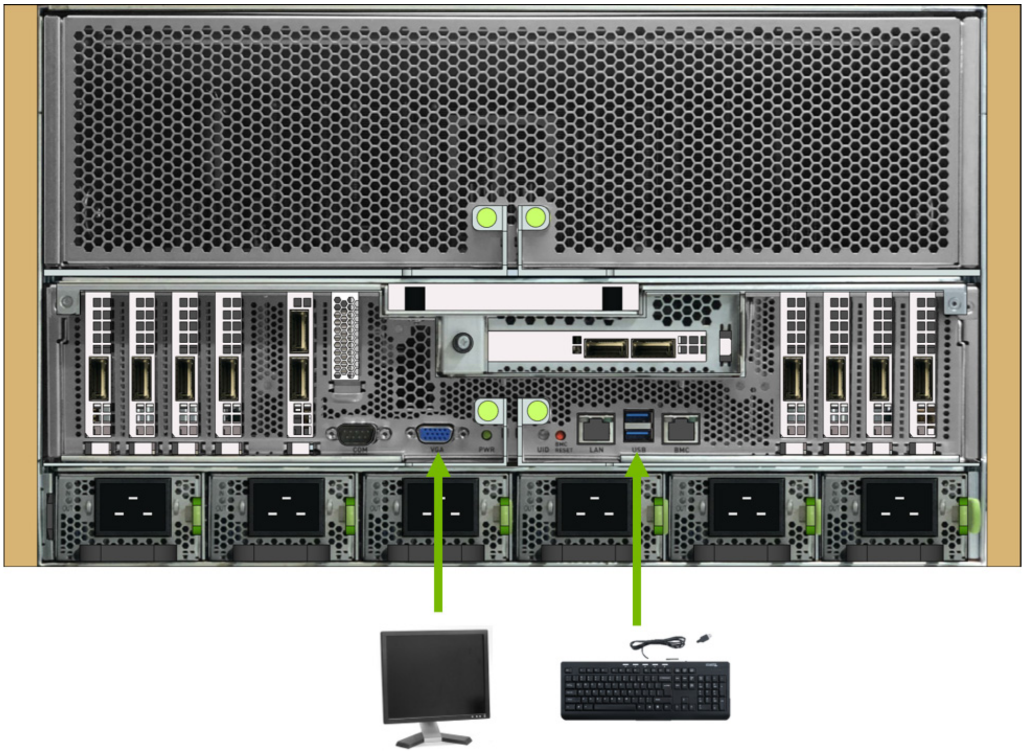
At either the front or the back of the DGX A100 system, connect a display to the VGA connector, and a keyboard to any of the USB ports.

Note: The display resolution must be 1440x900 or lower.

DGX A100 Server Front



DGX A100 Server Rear



2.1.2 Remote Connection through the BMC

Note: BMC Security

NVIDIA recommends that customers follow best security practices for BMC management (IPMI port). These include, but are not limited to, such measures as:

- Restricting the DGX A100 IPMI port to an isolated, dedicated, management network
- Using a separate, firewalled subnet
- Configuring a separate VLAN for BMC traffic if a dedicated network is not available

See the section **"Configuring Static IP Address for the BMC" if** you need to configure a static IP address for the BMC.

This method requires that you have the BMC login credentials. These credentials depend on the following conditions:

Prior to first time boot: The default credentials are

Username: admin

Password: dgxluna.admin

After first boot setup: During the first-boot procedure, you were prompted to configure an administrator username and password, and also a password for the BMC. The BMC username is the same as the administrator username

Username: <administrator-username>

Password: <bmc-password>

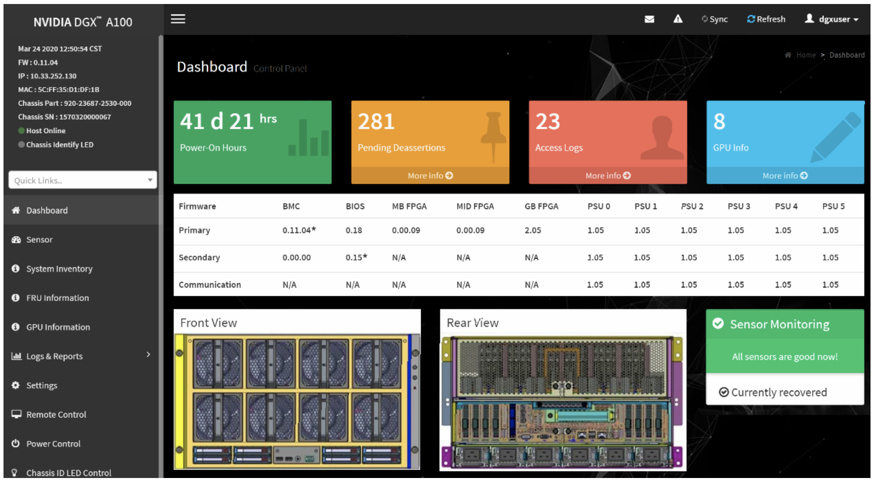
1 Make sure you have connected the BMC port on the DGX A100 system to your LAN.

2 Open a browser within your LAN and go to:

```
https://<bmc-ip-address>/
```

Make sure popups are allowed for the BMC address.

3 Log in.



Mar 24 2020 12:50:54 CST
FW : 0.11.04
IP : 10.33.252.130
MAC : 5C:FF:35:D1:DF:1B
Chassis Part : 920-23687-2530-000
Chassis SN : 1570320000067
Host Online
Chassis Identify LED

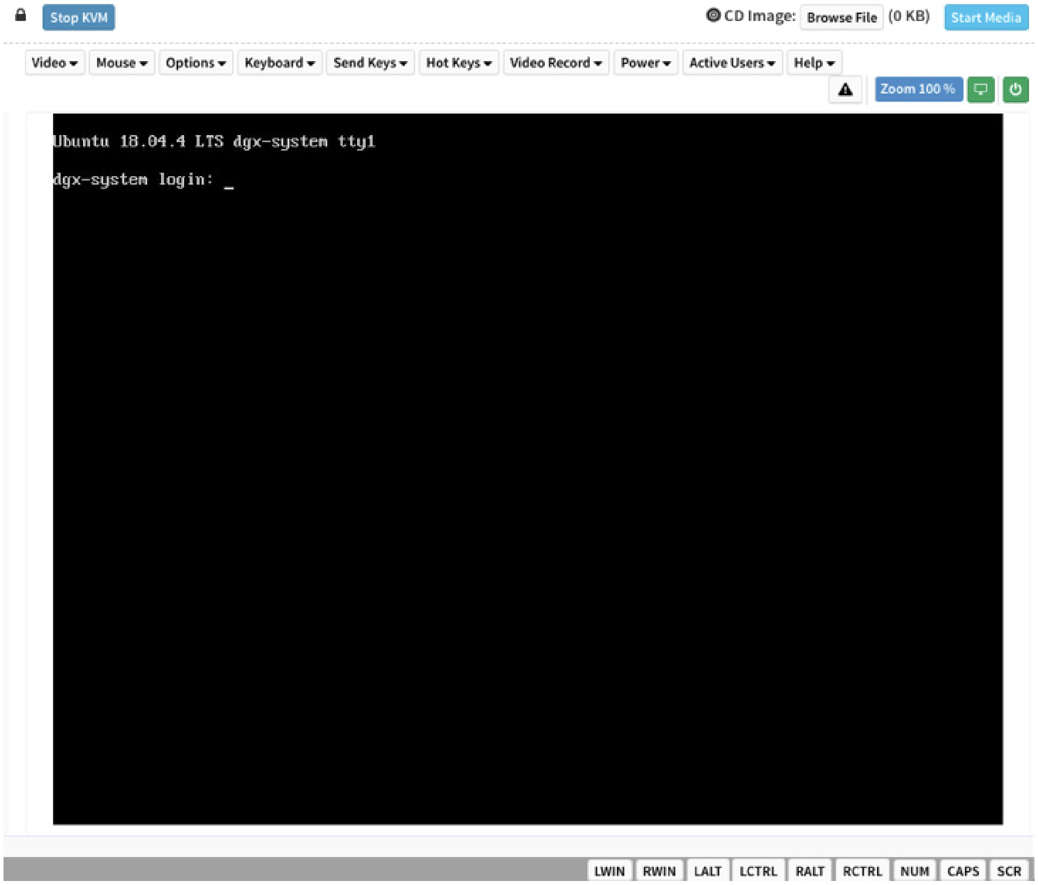
Firmware	BMC	BIOS	MB FPGA	MID FPGA	GB FPGA	PSU 0	PSU 1	PSU 2	PSU 3	PSU 4	PSU 5
Primary	0.11.04*	0.18	0.00.09	0.00.09	2.05	1.05	1.05	1.05	1.05	1.05	1.05
Secondary	0.00.00	0.15*	N/A	N/A	N/A	1.05	1.05	1.05	1.05	1.05	1.05
Communication	N/A	N/A	N/A	N/A	N/A	1.05	1.05	1.05	1.05	1.05	1.05

4 From the left-side navigation menu, click **Remote Control.**

The **Remote Control** page allows you to open a virtual Keyboard/Video/Mouse (KVM) on the DGX A100 system, as if you were using a physical monitor and keyboard connected to the front of the system.

5 Click **Launch KVM**.

The DGX A100 console appears in your browser.



```
Ubuntu 18.04.4 LTS dgx-system tty1

dgx-system login: _
```

2.2 SSH CONNECTION TO THE OS

You can also establish an SSH connection to the DGX A100 OS through the network port. See the section **“Network Ports”** to identify the port to use, and the section **“Configuring Static IP Addresses for the Network Ports”** if you need to configure a static IP address.

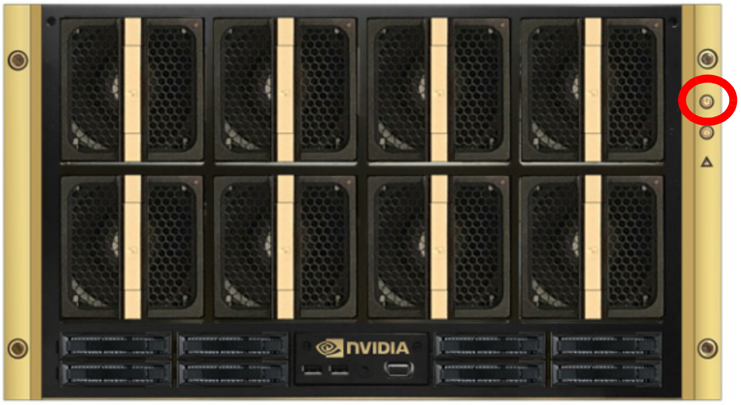
CHAPTER 3 FIRST-BOOT SETUP

While NVIDIA partner network personnel or NVIDIA field service engineers will install the DGX A100 system at the site and perform the first boot setup, the first boot setup instructions are provided here for reference and to support any re-imaging of the server.

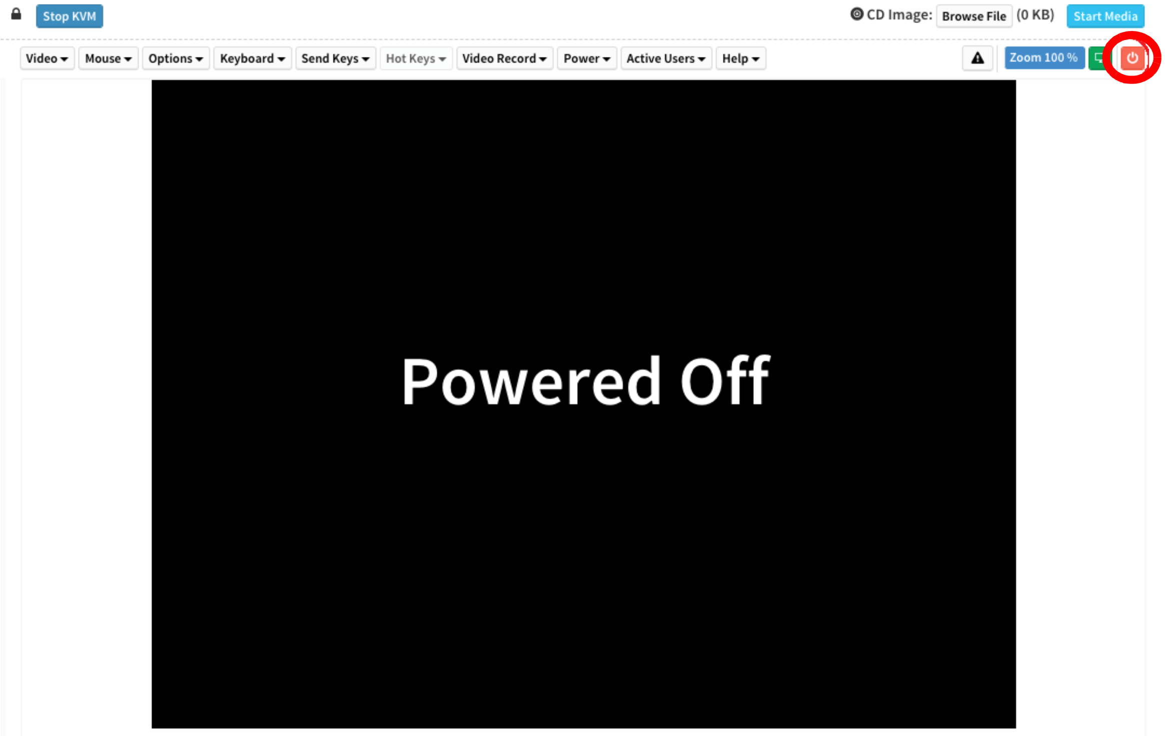
These instructions describe the setup process that occurs the first time the DGX A100 system is powered on after delivery or after the server is re-imaged.

Be prepared to accept all End User License Agreements (EULAs) and to set up your username and password. To preview the EULA, visit https://www.nvidia.com/en-us/data-center/dgx-systems/support/ and click the DGX EULA link.

1. Connect to the DGX A100 console as explained in **Connecting to the DGX A100 Console**.
2. Power on the DGX A100 system.
   - Using the physical power button



- Using the Remote BMC



Powered Off

The system will take a few minutes to boot.

You are presented with end user license agreements (EULAs) for the NVIDIA software.

3 Accept the EULA to proceed with the installation.

4 Perform the steps to configure the DGX A100 software.

- Select your language and locale preferences.
- Confirm the UTC clock setting.
- Create an administrative user account with your name, username, and password.

  The administrator username is also the BMC login username.
- Create a BMC password.

  The BMC password must consist of a minimum of 13 characters.

CAUTION: Once you create your login credentials, the default admin/dgxluna.admin login will no longer work.

Note: The BMC software will not accept "sysadmin" for a user name. If you create this user name for the system log in, "sysadmin" will not be available for logging in to the BMC.

- Choose a primary network interface for the DGX A100 system; for example, enp226s0.

  This should typically be the interface that you will use for subsequent system configuration or in-band management.

  After you select the primary network interface, the system attempts to configure the interface for DHCP and then asks you to enter the name server addresses.

  – If no DHCP is available, then click **OK** at the **Network autoconfiguration failed** dialog and configure the network manually.

  – If you want to configure a static address, then click **Cancel** at the dialog after the DHCP configuration completes to restart the network configuration steps.

  – If you need to select a different network interface, then click **Cancel** at the dialog after the DHCP configuration completes to restart the network configuration steps.

- Enter name server addresses.

- Choose a host name for the DGX A100 system.

  After completing the setup process, the DGX A100 system reboots automatically and then presents the login prompt.

5 Update the software to ensure you are running the latest version.

Updating the software ensures your DGX A100 system contains important updates, including security updates. The Ubuntu Security Notice site (https://usn.ubuntu.com/) lists known Common Vulnerabilities and Exposures (CVEs), including those that can be resolved by updating the DGX OS software.

a Run the package manager.

```
    $ sudo apt update
```

b Upgrade to the latest version.

```
    $ sudo apt full-upgrade
```

Note: RAID 1 Rebuild May Temporarily Affect System Performance - When the system is booted after restoring the image and running the first-boot setup, software RAID begins the process of rebuilding the RAID 1 array - creating a mirror of (or resynchronizing) the drive containing the software. System performance may be affected during the RAID 1 rebuild process, which can take an hour to complete.

During this time, the command “nvsm show health” will report a warning that the RAID volume is resyncing.

You can check the status of the RAID 1 rebuild process using “sudo nvsm show volumes”, and then inspecting the output under /systems/localhost/storage/volumes/md0/rebuild.

CHAPTER 4 QUICK START AND BASIC OPERATION

This chapter provides basic requirements and instructions for using the DGX A100 System, including how to perform a preliminary health check and how to prepare for running containers. Be sure to visit the DGX documentation website at **https://docs.nvidia.com/dgx/** for additional product documentation.

4.1 INSTALLATION AND CONFIGURATION

! **IMPORTANT:** It is mandatory that your DGX A100 System be installed by NVIDIA partner network personnel or NVIDIA field service engineers. If not performed accordingly, your DGX A100 hardware warranty will be voided.

Before installation, make sure you have given all relevant site information to your Installation Partner.

4.2 REGISTRATION

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

Registration allows you access to the NVIDIA Enterprise Support Portal, technical support, software updates and access to set up an **NGC for DGX** account.

If you did not receive the information, open a case with the NVIDIA Enterprise Support Team by going to **https://www.nvidia.com/en-us/support/enterprise/.** The site provides ways of contacting the NVIDIA Enterprise Services team for support without requiring an NVIDIA Enterprise Support account.

4.3 OBTAINING AN NGC ACCOUNT

NVIDIA GPU Cloud (NGC) provides simple access to GPU-optimized software for deep learning, machine learning and high-performance computing (HPC). An NGC account grants you access to these tools as well as the ability to set up a private registry to manage your customized software.

Work with NVIDIA Enterprise Support to set up an NGC enterprise account if you are the organization administrator for your DGX A100 purchase. See the NGC Container Registry for DGX User Guide (**https://docs.nvidia.com/dgx/ngc-registry-for-dgx-user-guide/**) for detailed instructions on getting an NGC enterprise account.

4.4 TURNING DGX A100 ON AND OFF

DGX A100 is a complex system, integrating a large number of cutting-edge components with specific startup and shutdown sequences. Observe the following startup and shutdown instructions.

4.4.1 Startup Considerations

In order to keep your DGX A100 running smoothly, allow up to a minute of idle time after reaching the login prompt. This ensures that all components are able to complete their initialization.

4.4.2 Shutdown Considerations

⚠ **WARNING:Risk of Danger** - Removing power cables or using Power Distribution Units (PDUs) to shut off the system while the Operating System is running may cause damage to sensitive components in the DGX A100 server.

When shutting down DGX A100, always initiate the shutdown from the operating system, momentary press of the power button, or by using Graceful Shutdown from the BMC, and wait until the system enters a powered-off state before performing any maintenance.

4.5 VERIFYING FUNCTIONALITY

4.5.1 Quick Health Check

This section walks you through the steps of performing a health check on the DGX A100 System, and verifying the Docker and NVIDIA driver installation.

1. Establish an SSH connection to the DGX A100 System.
2. Run a basic system check.

```
    $ sudo nvsm show health
```

   Verify that the output summary shows that all checks are Healthy and that the overall system status is Healthy.

3. Verify that Docker is installed by viewing the installed Docker version.

```
    $ sudo docker --version
```

   This should return the version as “`Docker version 19.03.5-ce`”, where the actual version may differ depending on the specific release of the DGX OS Server software.

4. Verify connection to the NVIDIA repository and that the NVIDIA Driver is installed.

```
$ sudo docker run --gpus all --rm nvcr.io/nvidia/cuda:11.0-runtime
nvidia-smi
```

   Docker pulls the nvidia/cuda container image layer by layer, then runs nvidia-smi.

   When completed, the output should show the NVIDIA Driver version and a description of each installed GPU.

See the NVIDIA Containers and Deep Learning Frameworks User Guide at **https://docs.nvidia.com/deeplearning/dgx/user-guide/index.html** for further instructions, including an example of logging into the NGC container registry and launching a deep learning container.

4.6 RUNNING NGC CONTAINERS WITH GPU SUPPORT

To obtain the best performance when running NGC containers on DGX A100 systems, two methods of providing GPU support for Docker containers have been developed:

- Native GPU support (included in Docker 19.03 and later, installed)
- NVIDIA Container Runtime for Docker (`nvidia-docker2` package)

The method implemented in your system depends on the DGX OS version installed.

DGX OS Release	Method included
4.99	Native GPU support NVIDIA Container Runtime for Docker (deprecated - availability to be removed in a future DGX OS release)

Each method is invoked by using specific Docker commands, described as follows.

4.6.1 Using Native GPU Support

- Use `docker run --gpus` to run GPU-enabled containers.
  - Example using all GPUs

```
$ docker run --gpus all ...
```

  - Example using two GPUs

```
$ docker run --gpus 2 ...
```

  - Examples using specific GPUs

```
$ docker run --gpus '"device=1,2"' ...
```

```
$ docker run --gpus '"device=UUID-ABCDEF,1"' ...
```

4.6.2 Using the NVIDIA Container Runtime for Docker

Note: The NVIDIA Container Runtime for Docker is deprecated and will be removed from the DGX OS in a future release.

Currently, the DGX OS also includes the NVIDIA Container Runtime for Docker (`nvidia-docker2`) which lets you run GPU-accelerated containers in one of the following ways.

- Use `docker run` and specify `runtime=nvidia`.

```
$ docker run --runtime=nvidia ...
```

- Use `nvidia-docker run`.

```
$ nvidia-docker run ...
```

  The `nvidia-docker2` package provides backward compatibility with the previous `nvidia-docker` package, so you can run GPU-accelerated containers using this command and the new runtime will be used.

- Use `docker run` with `nvidia` as the default runtime.

  You can set `nvidia` as the default runtime, for example, by adding the following line to the `/etc/docker/daemon.json` configuration file as the first entry.

```
"default-runtime": "nvidia",
```

  The following is an example of how the added line appears in the JSON file. Do not remove any pre-existing content when making this change.

```
{
 "default-runtime": "nvidia",
  "runtimes": {
     "nvidia": {
         "path": "/usr/bin/nvidia-container-runtime",
         "runtimeArgs": []
     }
 },
}
```

  You can then use `docker run` to run GPU-accelerated containers.

```
$ docker run ...
```

CAUTION: If you build Docker images while `nvidia` is set as the default runtime, make sure the build scripts executed by the Dockerfile specify the GPU architectures that the container will need. Failure to do so may result in the container being optimized only for the GPU architecture on which it was built. Instructions for specifying the GPU architecture depend on the application and are beyond the scope of this document. Consult the specific application build process for guidance.

4.7 MANAGING CPU MITIGATIONS

DGX OS Server includes security updates to mitigate CPU speculative side-channel vulnerabilities. These mitigations can decrease the performance of deep learning and machine learning workloads.

If your installation of DGX systems incorporates other measures to mitigate these vulnerabilities, such as measures at the cluster level, you can disable the CPU mitigations for individual DGX nodes and thereby increase performance.

4.7.1 Determining the CPU Mitigation State of the DGX System

If you do not know whether CPU mitigations are enabled or disabled, issue the following.

```
$ cat /sys/devices/system/cpu/vulnerabilities/*
```

- CPU mitigations are enabled if the output consists of multiple lines prefixed with Mitigation:.

  Example

```
KVM: Mitigation: Split huge pages
Mitigation: PTE Inversion; VMX: conditional cache flushes, SMT
vulnerable
Mitigation: Clear CPU buffers; SMT vulnerable
Mitigation: PTI
Mitigation: Speculative Store Bypass disabled via prctl and seccomp
Mitigation: usercopy/swapgs barriers and __user pointer sanitization
Mitigation: Full generic retpoline, IBPB: conditional, IBRS_FW, STIBP:
conditional, RSB filling
Mitigation: Clear CPU buffers; SMT vulnerable
```

- CPU mitigations are disabled if the output consists of multiple lines prefixed with Vulnerable.

  Example

```
KVM: Vulnerable
Mitigation: PTE Inversion; VMX: vulnerable
Vulnerable; SMT vulnerable
Vulnerable
Vulnerable
Vulnerable: __user pointer sanitization and usercopy barriers only; no
swapgs barriers
Vulnerable, IBPB: disabled, STIBP: disabled
Vulnerable
```

4.7.2 Disabling CPU Mitigations

CAUTION: Performing the following instructions will disable the CPU mitigations provided by the DGX OS Server software.

1 Install the nv-mitigations-off package.

```
$ sudo apt install nv-mitigations-off -y
```

2 Reboot the system.

3 Verify CPU mitigations are disabled.

```
$ cat /sys/devices/system/cpu/vulnerabilities/*
```

The output should include several Vulnerable lines. See **"Determining the CPU Mitigation State of the DGX System"** for example output.

4.7.3 Re-enabling CPU Mitigations

1 Remove the nv-mitigations-off package.

```
$ sudo apt purge nv-mitigations-off
```

2 Reboot the system.

3 Verify CPU mitigations are enabled.

```
$ cat /sys/devices/system/cpu/vulnerabilities/*
```

The output should include several Mitigations lines. See **"Determining the CPU Mitigation State of the DGX System"** for example output.

CHAPTER 5 ADDITIONAL FEATURES AND INSTRUCTIONS

This chapter describes specific features of the DGX A100 server to consider during setup and operation.

5.1 MANAGING THE DGX CRASH DUMP FEATURE

The DGX OS includes a script to manage this feature.

5.1.1 Using the Script

- To enable only dmesg crash dumps, enter

```
$ /usr/sbin/dgx-kdump-config enable-dmesg-dump
```

  This option reserves memory for the crash kernel.

- To enable both dmesg and vmcore crash dumps, enter

```
$ /usr/sbin/dgx-kdump-config enable-vmcore-dump
```

  This option reserves memory for the crash kernel.

- To disable crash dumps, enter

```
$ /usr/sbin/dgx-kdump-config disable
```

  This option disables the use of kdump and make sure no memory is reserved for the crash kernel.

5.1.2 Connecting to Serial Over LAN

While dumping vmcore, the BMC screen console goes blank approximately 11 minutes after the crash dump is started. To view the console output during the crash dump, connect to serial over LAN as follows:

```
$ ipmitool -I lanplus -H <bmc-ip-address> -U <BMC-USERNAME> -P <BMC-
PASSWORD> sol activate
```

CHAPTER 6 MANAGING THE DGX A100 SELF-ENCRYPTING DRIVES

The NVIDIA DGX™ OS software supports the ability to manage self-encrypting drives (SEDs), including setting an Authentication Key for locking and unlocking the drives on NVIDIA DGX™ A100 systems. You can manage only the SED data drives. The software cannot be used to manage OS drives even if they are SED-capable.

6.1 OVERVIEW

The self-encrypting drive (SED) management software is provided in the `nv-disk-encrypt` package.

The software supports the following configurations.

- NVIDIA DGX 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.
- Lets you assign Authentication Keys (passwords) for each SED as part of the initialization process.
  - Alternatively, the software can 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.

6.2 INSTALLING THE SOFTWARE

Use the package manager to install the `nv-disk-encrypt` package and then reboot the system.

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

6.3 INITIALIZING THE SYSTEM FOR DRIVE ENCRYPTION

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

Syntax

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

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 further instructions.
- `-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.

6.4 ENABLING DRIVE LOCKING

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

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

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

6.5 INITIALIZATION EXAMPLES

6.5.1 Example 1: Passing in the JSON File

The following instructions describe a method for specifying the drive/password mapping ahead of time. This method is useful for initializing several drives at a time and avoids the need to enter the password for each drive after issuing the initialization command, or if you want control of the passwords.

6.5.1.1 Determining Which 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.

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 snippet shows drives than **cannot** be used for encryption. Notice 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 |
```

Alternatively, you can specify the output be presented in JSON format by using the -j option.

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

In this case, drives that can be used for encryption are indicate by

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

And drives that cannot be used for encryption are indicated by either

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

Or

```
"sed_capable": false,
```

6.5.1.2 Creating the Drive/Password Mapping JSON File and Using it 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 from the section **“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>",
     }
```

- Be sure to follow the syntax exactly.
- Passwords must consist of only upper-case letters, lower-case letters, digits, and/or the following special-characters: `~ : @ % ^ + = _ ,`

2 Initialize the system and then 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
```

Provide a password for the vault when prompted.

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

3. Delete the JSON file in the temporary location for security.

6.5.2 Example 2: Generating Random Passwords

The following commands uses the -k and -r options so that you are not prompted to enter passwords. You pass the vault password into the command and then the command instructs the tool to generate random passwords for each drive.

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

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

6.5.3 Example 3: Specifying Passwords One at a Time When Prompted

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 the following special-characters: ~ : @ % ^ + = _ ,

6.6 DISABLING DRIVE LOCKING

You can disable drive locking at any time after initialization by issuing the following.

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

- This command disables locking on all drives.
- You can re-run the initial setup at any time after this.

6.7 EXPORTING THE VAULT

To export all drive keys out to a file, use the export function. This requires that you pass in the vault password.

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

```
Writing vault data to /tmp/secrets.out
```

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

6.8 ERASING YOUR DATA

! **CAUTION: Be aware when executing this that all data will be lost. On DGX A100 systems, these drives generally form a RAID 0 array - this will also be destroyed when performing an erase.**

After initializing the system for SED management, use the nv-disk-encrypt command to erase data on your drives by issuing the following.

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

This command

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

6.9 USING THE TRUSTED PLATFORM MODULE

The NVIDIA DGX A100 incorporates Trusted Platform Module 2.0 (TPM 2.0) which can be enabled from the system BIOS. Once enabled, the `nv-disk-encrypt` tool uses the module for encryption and storage of the vault and SED authentication keys.

You need to access the system BIOS to enable or disable the TPM.

6.9.1 Enabling the TPM

The DGX A100 system is shipped with the TPM disabled. To enable the TPM, do the following.

1. Reboot the DGX A100, then press **[Del]** or **[F2]** at the NVIDIA splash screen to enter the BIOS Setup.
2. Navigate to the **Advanced** tab on the top menu, then scroll to **Trusted Computing** and press **[Enter]**.
3. Scroll to **Security Device** and press **[Enter]**.
4. Select **Enable** at the *Security Device* popup, then press **[Enter]**.
5. Save and exit the BIOS Setup.

The `nv-disk-encrypt` tool can now use the TPM.

6.9.2 Clearing the TPM

If you've lost the password to your TPM, you will not be able to access its contents. In this case, the only way to regain access to the TPM is to clear the TPM's contents. After clearing the TPM, you will need to re-initialize the vault and SED authentication keys.

To clear the TPM, do the following.

1. Reboot the DGX A100, then press **[Del]** or **[F2]** at the NVIDIA splash screen to enter the BIOS Setup.
2. Disable Hidden Setup Options.
   a. Navigate to the **Advanced** tab on the top menu, then scroll to **Hidden Setup Options** and press **[Enter]**.
   b. Select **Disable** at the *Hidden Setup Options* popup, then press **[Enter]**.
3. Clear TPM2.
   a. Scroll to **Trusted Computing** and press **[Enter]**.
   b. Scroll to **Pending Operation** and press **[Enter]**.

   c Select **TPM Clear** at the *Pending Operation* popup, then press **[Enter]**.

4 Save and exit the BIOS Setup.

6.10 CHANGING DISK PASSWORDS, ADDING DISKS, OR REPLACING DISKS

The same requirements apply if you want to change or rotate passwords, add disks, or replace disks.

1 Disable SED management.

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

2 Add or replace drives as needed and then rebuild the RAID array.

   Refer to the NVIDIA DGX A100 Service Manual for instructions.

3 Enable SED management and assign passwords per the instructions in **"Initializing the System for Drive Encryption"** .

6.11 HOT REMOVAL AND RE-INSERTION

After removing and reinserting a drive, it will become locked. As the SED-unlock service only runs at upon system power on, it will not automatically unlock a hotly inserted drive. To unlock such drives, you can restart the service manually by issuing the following.

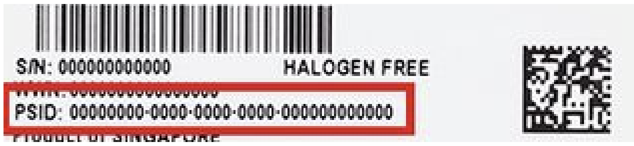
```
$ sudo systemctl start sed-unlock
```

Be sure to rebuild the RAID array after unlocking the drive.

6.12 RECOVERING FROM LOST KEYS

NVIDIA recommends backing up your keys and storing them in a secure location. If you've lost the key used to initialize and lock your drives, you will not be able to unlock the drive again. If this happens, the only way to recover is to perform a factory-reset, which will result in a loss of data.

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



Specify the PSID to reset the drive using the following `sedutil-cli` command:

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

CHAPTER 7 NETWORK CONFIGURATION

This chapter describes key network considerations and instructions for the DGX A100 System.

7.1 CONFIGURING NETWORK PROXIES

If your network requires use of a proxy server, you will need to set up configuration files to ensure the DGX A100 System communicates through the proxy.

7.1.1 For the OS and Most Applications

Edit the file `/etc/environment` and add the following proxy addresses to the file, below the PATH line.

```
http_proxy="http://<username>:<password>@<host>:<port>/"
ftp_proxy="ftp://<username>:<password>@<host>:<port>/";
https_proxy="https://<username>:<password>@<host>:<port>/";
no_proxy="localhost,127.0.0.1,localaddress,.localdomain.com"
HTTP_PROXY="http://<username>:<password>@<host>:<port>/"
FTP_PROXY="ftp://<username>:<password>@<host>:<port>/";
HTTPS_PROXY="https://<username>:<password>@<host>:<port>/";
NO_PROXY="localhost,127.0.0.1,localaddress,.localdomain.com"
```

Where username and password are optional.

Example:

```
http_proxy="http://myproxy.server.com:8080/"
ftp_proxy="ftp://myproxy.server.com:8080/";
https_proxy="https://myproxy.server.com:8080/";
```

7.1.2 For apt

Edit (or create) a proxy config file `/etc/apt/apt.conf.d/myproxy` and include the following lines

```
Acquire::http::proxy "http://<username>:<password>@<host>:<port>/";
Acquire::ftp::proxy "ftp://<username>:<password>@<host>:<port>/";
Acquire::https::proxy "https://<username>:<password>@<host>:<port>/";
```

Where username and password are optional.

Example:

```
Acquire::http::proxy "http://myproxy.server.com:8080/";
Acquire::ftp::proxy "ftp://myproxy.server.com:8080>/";
Acquire::https::proxy "https://myproxy.server.com:8080/";
```

7.1.3 For Docker

To ensure that Docker can access the NGC container registry through a proxy, Docker uses environment variables. For best practice recommendations on configuring proxy environment variables for Docker, see **https://docs.docker.com/engine/admin/systemd/#http-proxy**.

7.1.4 Configuring Docker IP Addresses

To ensure that the DGX A100 System can access the network interfaces for Docker containers, Docker should be configured to use a subnet distinct from other network resources used by the DGX A100 System.

By default, Docker uses the 172.17.0.0/16 subnet. Consult your network administrator to find out which IP addresses are used by your network. ***If your network does not conflict with the default Docker IP address range, then no changes are needed and you can skip this section.***

However, if your network uses the addresses within this range for the DGX A100 System, you should change the default Docker network addresses.

You can change the default Docker network addresses by either modifying the /etc/docker/daemon.json file or modifying the /etc/systemd/system/docker.service.d/docker-override.conf file. These instructions provide an example of modifying the `/etc/systemd/system/docker.service.d/docker-override.conf` to override the default Docker network addresses.

1 Open the docker-override.conf file for editing.

```
$ sudo vi /etc/systemd/system/docker.service.d/docker-override.conf
```

```
   [Service]
   ExecStart=
   ExecStart=/usr/bin/dockerd -H fd:// -s overlay2
   LimitMEMLOCK=infinity
   LimitSTACK=67108864
```

2 Make the changes indicated in bold below, setting the correct bridge IP address and IP address ranges for your network. Consult your IT administrator for the correct addresses.

```
[Service]
ExecStart=
ExecStart=/usr/bin/dockerd -H fd:// -s overlay2 --bip=192.168.127.1/24
      --fixed-cidr=192.168.127.128/25

LimitMEMLOCK=infinity
LimitSTACK=67108864
```

Save and close the `/etc/systemd/system/docker.service.d/docker-override.conf` file when done.

3 Reload the systemctl daemon.

```
$ sudo systemctl daemon-reload
```

4 Restart Docker.

```
$ sudo systemctl restart docker
```

7.1.5 Opening Ports

Make sure that the ports listed in the following table are open and available on your firewall to the DGX A100 System:

Port (Protocol)	Direction	Use
22 (TCP)	Inbound	SSH
53 (UDP)	Outbound	DNS
80 (TCP)	Outbound	HTTP, package updates
443 (TCP)	Outbound	For internet (HTTP/HTTPS) connection to NVIDIA GPU Cloud If port 443 is proxied through a corporate firewall, then WebSocket protocol traffic must be supported
443 (TCP)	Inbound	For BMC web services, remote console services, and cd-media service. If port 443 is proxied through a corporate firewall, then WebSocket protocol traffic must be supported

7.2 CONNECTIVITY REQUIREMENTS FOR NGC CONTAINERS

To run NVIDIA NGC containers from the NGC container registry, your network must be able to access the following URLs:

- **http://archive.ubuntu.com/ubuntu/**
- **http://security.ubuntu.com/ubuntu/**
- http://international.download.nvidia.com/dgx/repos/

  (To be accessed using apt-get, not through a browser.)
- **https://apt.dockerproject.org/repo/**
- **https://download.docker.com/linux/ubuntu/**
- https://nvcr.io/

  To verify connection to nvcr.io, run

```
$ wget https://nvcr.io/v2
```

You should see connecting verification followed by a 401 error.

```
--2018-08-01 19:42:58--  https://nvcr.io/v2
Resolving nvcr.io (nvcr.io)... 52.8.131.152, 52.9.8.8
Connecting to nvcr.io (nvcr.io)|52.8.131.152|:443... connected.
HTTP request sent, awaiting response... 401 Unauthorized
```

7.3 CONFIGURING STATIC IP ADDRESS FOR THE BMC

This section explains how to set a static IP address for the BMC. You will need to do this if your network does not support DHCP.

Use one of the methods described in the following sections:

- **"Configuring a BMC Static IP Address Using ipmitool"**
- **"Configuring a BMC Static IP Address Using the System BIOS"**

7.3.1 Configuring a BMC Static IP Address Using ipmitool

This section describes how to set a static IP address for the BMC from the Ubuntu command line.

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

To view the current settings, enter the following command.

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

To set a static IP address for the BMC, do the following.

1 Set the IP address source to **static**.

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

2 Set the appropriate address information.

- To set the IP address ("Station IP address" in the BIOS settings), 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 ("Router IP address" in the BIOS settings), enter the following and replace the italicized text with your information.

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

7.3.2 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 access the DGX A100 System remotely. 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, then turn on the DGX A100 System.
2. When you see the SBIOS version screen, press Del or F2 to enter the BIOS Setup Utility screen.

   Example setup screen – details may vary depending on SBIOS version.
3. At the BIOS Setup Utility screen, navigate to the Server Mgmt tab on the top menu, then scroll to BMC network configuration and press Enter.
4. Scroll to Configuration Address Source and press Enter, then at the Configuration Address source pop-up, select Static and then press Enter.
5. Set the addresses for the Station IP address, Subnet mask, and Router IP address as needed by performing the following for each:
   a Scroll to the specific item and press Enter.
   b Enter the appropriate information at the pop-up, then press Enter.
6. When finished making all your changes, press F10 to save & exit

7.4 CONFIGURING STATIC IP ADDRESSES FOR THE NETWORK PORTS

During the initial boot setup process for the DGX A100 System, you had an opportunity to configure static IP addresses for a single network interface. If you did not set this up at that time, you can configure the static IP addresses from the Ubuntu command line using the following instructions.

> **Note: If you are connecting to the DGX A100 console remotely, then connect using the BMC remote console. If you connect using SSH, then your connection will be lost when performing the final step. Also, the BMC connection will facilitate troubleshooting should you encounter issues with the config file.**
>
> **If you cannot access the DGX A100 System remotely, then connect a display (1440x900 or lower resolution) and keyboard directly to the DGX A100 System.**

1. Determine the port designation that you want to configure, based on the physical Ethernet port that you have connected to your network.

   See the section **"Configuring Network Proxies"** for the port designation of the connection you want to configure.

2. Edit the network configuration yaml file.

```
$ sudo vi /etc/netplan/01-netcfg.yaml
```

```
network:
  version: 2
  renderer: networkd
  ethernets:

     <port-designation>:
         dhcp4: no
         dhcp6: no
         addresses: [10.10.10.2/24]
         gateway4: 10.10.10.1
         nameservers:
           search: [<mydomain>, <other-domain>]
           addresses: [10.10.10.1, 1.1.1.1]
```

   Consult your network administrator for the appropriate information for the items in bold, such as network, gateway, and nameserver addresses, and use the port designations that you determined in step 1.

3. When finished with your edits, press **ESC** to switch to command mode, then save the file to the disk and exit the editor.

4. Apply the changes.

```
$ sudo netplan apply
```

> **Note:** If you are not returned to the command line prompt after a minute, then reboot the system.

For additional information, see **https://help.ubuntu.com/lts/serverguide/network-configuration.html.en**.

7.5 SWITCHING BETWEEN INFINIBAND AND ETHERNET

The NVIDIA DGX A100 System is equipped with eight QSFP56 network ports on the I/O board, typically used for cluster communications. By default these are configured as InfiniBand ports, but you have the option to convert these to Ethernet ports.

For these changes to work properly, the configured port must connect to a networking switch that matches the port configuration. In other words, if the port configuration is set to InfiniBand, then the external switch should be an InfiniBand switch with the corresponding InfiniBand cables. Likewise, if the port configuration is set to Ethernet, then the switch should also be Ethernet.

The DGX A100 is also equipped with one (and optionally two) dual-port connections typically used for network storage and configured by default for Ethernet. These can be configure for InfiniBand as well.

> **Note: On the dual-port cards, if one of the ports is configured for Ethernet and the other port is configured for InfiniBand, the following limitations apply.**
>
> **- FDR is not supported on the InfiniBand port (port 1 or 2).**
> **- If port 1 is InfiniBand, then port 2 (Ethernet) does not support 40 GbE/10GbE.**
> **- If port 1 is Ethernet, then port 2 (InfiniBand) does not support EDR.**

7.5.1 Starting the Mellanox Software Tools

1 To determine whether the Mellanox Software Tools (MST) services are running, enter the following.

```
$ sudo mst status -v
```

- "NA" in the MST column of the output indicates the services are ***not*** running, as shown in the following example.

  In this example, the additional storage network card is installed.

```
MST modules:
------------
MST PCI module is not loaded
MST PCI configuration module is not loaded
PCI devices:
------------
DEVICE_TYPE             MST      PCI       RDMA            NET                    NUMA
ConnectX6(rev:0)        NA       0c:00.0   mlx5_0          net-ib0                   3
ConnectX6(rev:0)        NA       ba:00.0   mlx5_8          net-ib6                   5
ConnectX6(rev:0)        NA       8d:00.0   mlx5_6          net-ib4                   7
ConnectX6(rev:0)        NA       54:00.0   mlx5_3          net-ib3                   1
ConnectX6(rev:0)        NA       4b:00.0   mlx5_2          net-ib2                   1
ConnectX6(rev:0)        NA       e1:00.1   mlx5_11         net-enp225s0f1            4
ConnectX6(rev:0)        NA       94:00.0   mlx5_7          net-ib5                   7
ConnectX6(rev:0)        NA       ca:00.0   mlx5_9          net-ib7                   5
ConnectX6(rev:0)        NA       61:00.1   mlx5_5          net-enp97s0f1             0
ConnectX6(rev:0)        NA       12:00.0   mlx5_1          net-ib1                   3
ConnectX6(rev:0)        NA       e1:00.0   mlx5_10         net-enp225s0f0            4
ConnectX6(rev:0)        NA       61:00.0   mlx5_4          net-enp97s0f0             0
```

- The device path in the MST column output indicates the services are running, as shown in the following example.

```
MST modules:
------------
    MST PCI module is not loaded
    MST PCI configuration module loaded
PCI devices:
------------
DEVICE_TYPE         MST                       PCI       RDMA       NET           NUMA
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf9.1  e1:00.1   mlx5_11    net-enp225s0f1   1
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf9    e1:00.0   mlx5_10    net-enp225s0f0   1
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf8    cd:00.0   mlx5_9     net-ib7          1
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf7    bb:00.0   mlx5_8     net-ib6          1
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf6    94:00.0   mlx5_7     net-ib5          1
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf5    8d:00.0   mlx5_6     net-ib4          1
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf4.1  61:00.1   mlx5_5     net-enp97s0f1    0
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf4    61:00.0   mlx5_4     net-enp97s0f0    0
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf3    54:00.0   mlx5_3     net-ib3          0
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf2    4b:00.0   mlx5_2     net-ib2          0
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf1    14:00.0   mlx5_1     net-ib1          0
ConnectX6(rev:0)  /dev/mst/mt4123_pciconf0    0d:00.0   mlx5_0     net-ib0          0
$
```

2 If necessary, start the **mst** driver.

```
$ sudo mst start
```

7.5.2 Determining the Current Port Configuration

To determine the current port configuration, enter the following:

```
$ sudo mlxconfig -e query | egrep -e Device\|LINK_TYPE
```

The following example shows the output for one of the port devices, showing the device path and the default, current, and next boot configuration.

```
Device #2:
Device type:    ConnectX6
Device:         /dev/mst/mt4123_pciconf8
Configurations:            Default         Current         Next Boot
*   LINK_TYPE_P1           IB(1)           IB(1)           IB(1)
```

- IB(1) indicates the port is configured for InfiniBand.
- ETH(2) indicates the port is configured for Ethernet.

Determine the Device path bus numbers for the slot number of the port you want to configure. See the diagram and table in **"Configuring Network Proxies"** for the mapping.

7.5.3 Switching the Port Configuration

Make sure that you have started the Mellanox Software Tools (MST) services as explain in the section **"Starting the Mellanox Software Tools"** , and have identified the correct ports to change.

Issue mlxconfig for each port you want to configure.

Syntax:

```
$ sudo mlxconfig -y -d <device-path> set LINK_TYPE_P1=<config-
number>
```

where

<device-path> corresponds to the port you want to configure

<config-number> is '1' for InfiniBand and '2' for Ethernet.

Example setting slot 0 to Ethernet

```
$ sudo mlxconfig -y -d /dev/mst/mt4123_pciconf2 set LINK_TYPE_P1=2
```

Example setting slot 1 to InfiniBand

```
$ sudo mlxconfig -y -d /dev/mst/mt4123_pciconf3 set LINK_TYPE_P1=1
```

CHAPTER 8 CONFIGURING STORAGE

By default, the DGX A100 System includes four SSDs in a RAID 0 configuration. These SSDs are intended for application caching, so you must set up your own NFS storage for long term data storage. The following instructions describe how to mount the NFS onto the DGX A100 System, and how to cache the NFS using the DGX A100 SSDs for improved performance.

Disabling cachefilesd

The DGX A100 system uses `cachefilesd` to manage caching of the NFS. If you do not want `cachefilesd` enabled, you can disable it as follows.

```
$ sudo systemctl stop cachefilesd
$ sudo systemctl disable cachefilesd
```

Using cachefilesd

The following instructions describe how to mount the NFS onto the DGX A100 system, and how to cache the NFS using the DGX A100 SSDs for improved performance.

Make sure that you have an NFS server with one or more exports with data to be accessed by the DGX A100 System, and that there is network access between the DGX A100 System and the NFS server.

1 Configure an NFS mount for the DGX A100 System.

   a Edit the filesystem tables configuration.

```
    $ sudo vi /etc/fstab
```

   b Add a new line for the NFS mount, using the local mount point of /mnt.

```
     <nfs_server>:<export_path> /mnt nfs
     rw,noatime,rsize=32768,wsize=32768,nolock,tcp,intr,fsc,nofail 0 0
```

- /mnt is used here as an example mount point.
- Consult your Network Administrator for the correct values for <nfs_server> and <export_path>.
- The nfs arguments presented here are a list of recommended values based on typical use cases. However, "fsc" must always be included as that argument specifies use of FS-Cache.

c Save the changes.

2 Verify the NFS server is reachable.

```
$ ping <nfs_server>
```

Use the server IP address or the server name provided by your network administrator.

3 Mount the NFS export.

```
$ sudo mount /mnt
```

`/mnt` is an example mount point.

4 Verify caching is enabled.

```
$ cat /proc/fs/nfsfs/volumes
```

Look for the text FSC=yes in the output.

The NFS will be mounted and cached on the DGX A100 System automatically upon subsequent reboot cycles.

CHAPTER 9 UPDATING AND RESTORING THE SOFTWARE

9.1 UPDATING THE DGX A100 SOFTWARE

You must register your DGX A100 system in order to receive email notification whenever a new software update is available.

These instructions explain how to update the DGX A100 software through an internet connection to the NVIDIA public repository. The process updates a DGX A100 system image to the latest QA'd versions of the entire DGX A100 software stack, including the drivers, for the latest update within a specific release; for example, to update to the latest Release 4.5 update from an earlier Release 4.5 version.

For instructions on ugrading from one Release to another (for example, from Release 4 to Release 5), consult the release notes for the target release.

9.1.1 Connectivity Requirements For Software Updates

Before attempting to perform the update, verify that the DGX A100 system network connection can access the public repositories and that the connection is not blocked by a firewall or proxy.

Enter the following on the DGX A100 system.

```
$ wget -O f1-changelogs http://changelogs.ubuntu.com/meta-release-lts
```

```
$ wget -O f2-archive http://archive.ubuntu.com/ubuntu/dists/bionic/
Release
```

```
$ wget -O f3-usarchive http://us.archive.ubuntu.com/ubuntu/dists/bionic/
Release
```

```
$ wget -O f4-security http://security.ubuntu.com/ubuntu/dists/bionic/
Release
```

```
$ wget -O f5-download http://download.docker.com/linux/ubuntu/dists/
bionic/Release
```

```
$ wget -O f6-international http://international.download.nvidia.com/dgx/
repos/bionic/dists/bionic/Release
```

All the `wget` commands should be successful and there should be six files in the directory with non-zero content.

9.1.2 Update Instructions

> **CAUTION:** These instructions update all software for which updates are available from your configured software sources, including applications that you installed yourself. If you want to prevent an application from being updated, you can instruct the Ubuntu package manager to keep the current version. For more information, see **Introduction to Holding Packages** on the Ubuntu Community Help Wiki.

Perform the updates using commands on the DGX A100 console.

1. Run the package manager.

```
$ sudo apt update
```

2. Check to see which software will get updated.

```
$ sudo apt full-upgrade -s
```

   To prevent an application from being updated, instruct the Ubuntu package manager to keep the current version. See **Introduction to Holding Packages**.

3. Upgrade to the latest version.

```
$ sudo apt full-upgrade
```

   Answer any questions that appear.

   Most questions require a Yes or No response. If asked to select the grub configuration to use, select the current one on the system.

   Other questions will depend on what other packages were installed before the update and how those packages interact with the update. Typically, you can accept the default option when prompted.

4. Reboot the system.

9.2 RESTORING THE DGX A100 SOFTWARE IMAGE

If the DGX A100 software image becomes corrupted (or the OS NVMe drives are replaced), restore the DGX A100 software image to its original factory condition from a pristine copy of the image.

The process for restoring the DGX A100 software image is as follows:

1. Obtain an ISO file that contains the image from NVIDIA Enterprise Support as explained in **"Obtaining the DGX A100 Software ISO Image and Checksum File"** .
2. Restore the DGX A100 software image from this file either remotely through the BMC or locally from a bootable USB flash drive.
   - If you are restoring the image remotely, follow the instructions in **"Re-Imaging the System Remotely"** .
   - If you are restoring the image locally, prepare a bootable USB flash drive and restore the image from the USB flash drive as explained in the following topics:
     - Creating a Bootable Installation Medium
     - Re-Imaging the System From a USB Flash Drive

> **Note:** The DGX OS Server software is restored on one of the two NMVe M.2 drives. When the system is booted after restoring the image, software RAID begins the process rebuilding the RAID 1 array - creating a mirror of (or resynchronizing) the drive containing the software. System performance may be affected during the RAID 1 rebuild process, which can take an hour to complete.

9.2.1 Obtaining the DGX A100 Software ISO Image and Checksum File

To ensure that you restore the latest available version of the DGX A100 software image, obtain the current ISO image file from NVIDIA Enterprise Support. A checksum file is provided for the image to enable you to verify the bootable installation medium that you create from the image file.

1. Log on to the **NVIDIA Enterprise Support** site.
2. Click the Announcements tab to locate the download links for the DGX A100 software image.
3. Download the ISO image and its checksum file and save them to your local disk.

   The ISO image is also available in an archive file. If you download the archive file, be sure to extract the ISO image before proceeding.

9.2.2 Re-Imaging the System Remotely

These instructions describe how to re-image the system remotely through the BMC. For information about how to restore the system locally, see **"Re-Imaging the System From a USB Flash Drive"** .

Before re-imaging the system remotely, ensure that the correct DGX A100 software image is saved to your local disk. For more information, see **"Obtaining the DGX A100 Software ISO Image and Checksum File"** .

1. Log in to the BMC.
2. Click Remote Control and then click Launch KVM.
3. Set up the ISO image as virtual media.
   a. From the top bar, click **Browse File** and then locate the re-image ISO file and click Open.
   b. Click Start Media.
4. Reboot, install the image, and complete the DGX A100 system setup.
   a. From the top menu, click Power and then select **Reset Server**.
   b. Click **OK** at the Power Control dialogs, then wait for the system to power down and then come back online.
   c. As the system boots, press **[F11]** when the NVIDIA logo appears to get to the boot menu.
   d. Browse to locate the Virtual CD that corresponds to the inserted ISO, then boot the system from it.
   e. At the boot selection screen, select **Install DGX Server**.

      If you are an advanced user who is not using the RAID disks as cache and want to keep data on the RAID disks, then select **Install DGX Server without formatting RAID**. See the section **"Retaining the RAID Partition While Installing the OS"** for more information.
   f. Press Enter.

      The DGX A100 system will reboot from ISO image and proceed to install the image. This can take approximately 15 minutes.

Note: The Mellanox InfiniBand driver installation may take up to 30 minutes, depending on how many cards undergo a firmware update.

After the installation is completed, the system ejects the virtual CD and then reboots into the OS.

Refer to **"First-Boot Setup"** for the steps to take when booting up the DGX A100 system for the first time after a fresh installation.

9.2.3 Creating a Bootable Installation Medium

After obtaining an ISO file that contains the software image from NVIDIA Enterprise Support, create a bootable installation medium, such as a USB flash drive or DVD-ROM, that contains the image.

Note: If you are restoring the software image remotely through the BMC, you do not need a bootable installation medium and you can omit this task.

- If you are creating a bootable USB flash drive, follow the instructions for the platform that you are using:
  - On a text-only Linux distribution, see**“Creating a Bootable USB Flash Drive by Using the dd Command”**
  - On Windows, see **“Creating a Bootable USB Flash Drive by Using Akeo Rufus”**
- If you are creating a bootable DVD-ROM, you can use any of the methods described in **Burning the ISO on to a DVD** on the Ubuntu Community Help Wiki.

9.2.4 Creating a Bootable USB Flash Drive by Using the dd Command

On a Linux system, you can use the **dd** command to create a bootable USB flash drive that contains the DGX A100 software image.

Note: To ensure that the resulting flash drive is bootable, use the dd command to perform a device bit copy of the image. If you use other commands to perform a simple file copy of the image, the resulting flash drive may not be bootable.

Ensure that the following prerequisites are met:

- The correct DGX A100 software image is saved to your local disk. For more information, see **“Obtaining the DGX A100 Software ISO Image and Checksum File”** .
- The USB flash drive capacity is at least 4 GB.

1 Plug the USB flash drive into one of the USB ports of your Linux system.

2 Obtain the device name of the USB flash drive by running the lsblk command.

```
lsblk
```

You can identify the USB flash drive from its size.

3 As root, convert and copy the image to the USB flash drive.

```
$ sudo dd if=path-to-software-image bs=2048 of=usb-drive-device-
name
```

! **CAUTION:** The dd command erases all data on the device that you specify in the of option of the command. To avoid losing data, ensure that you specify the correct path to the USB flash drive.

9.2.5 Creating a Bootable USB Flash Drive by Using Akeo Rufus

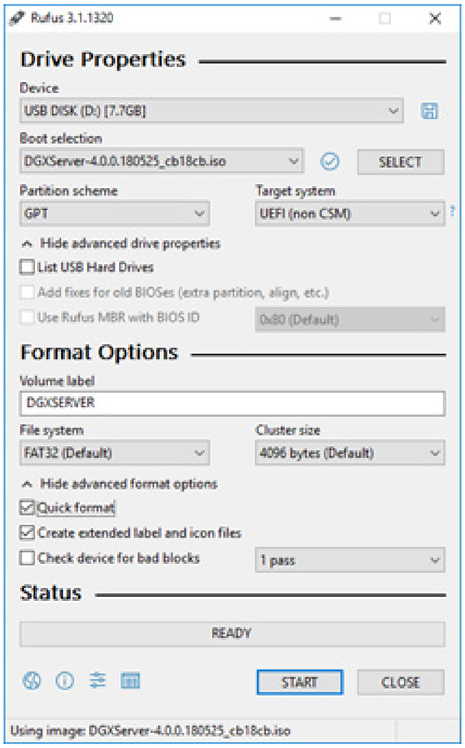
On a Windows system, you can use the **Akeo Reliable USB Formatting Utility (Rufus)** to create a bootable USB flash drive that contains the DGX A100 software image.

Ensure that the following prerequisites are met:

- The correct DGX A100 software image is saved to your local disk. For more information, see **"Obtaining the DGX A100 Software ISO Image and Checksum File"** .
- The USB flash drive has a capacity of at least 4 GB.

1 Plug the USB flash drive into one of the USB ports of your Windows system.

2 Download and launch the **Akeo Reliable USB Formatting Utility (Rufus)**



Drive Properties

Device: USB DISK (D:) [7.7GB]

Boot selection: DGXServer-4.0.0.180525_cb18cb.iso

Partition scheme: GPT

Target system: UEFI (non CSM)

Hide advanced drive properties

□ List USB Hard Drives

□ Add fixes for old BIOSes (extra partition, align, etc.)

□ Use Rufus MBR with BIOS ID 0x80 (Default)

Format Options

Volume label: DGXSERVER

File system: FAT32 (Default)

Cluster size: 4096 bytes (Default)

Hide advanced format options

☑ Quick format

☑ Create extended label and icon files

□ Check device for bad blocks 1 pass

Status

READY

Using image: DGXServer-4.0.0.180525_cb18cb.iso

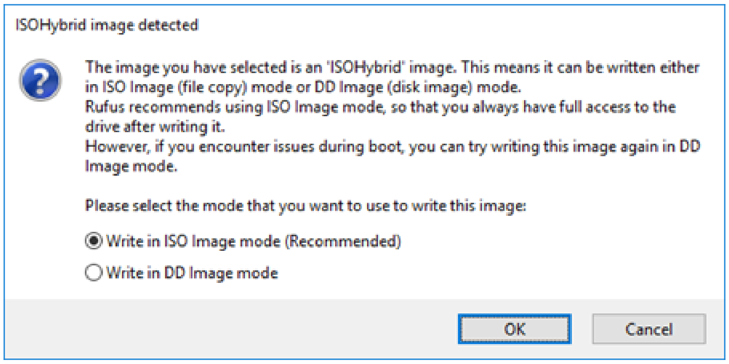
.

3 Under Boot selection, click SELECT and then locate and select the ISO image.

4 Under Partition scheme, select GPT.

5 Under **File system**, select **FAT32**.

6 Click Start. Because the image is a hybrid ISO file, you are prompted to select whether to write the image in ISO Image (file copy) mode or DD Image (disk image) mode.



ISOHybrid image detected

The image you have selected is an 'ISOHybrid' image. This means it can be written either in ISO Image (file copy) mode or DD Image (disk image) mode.
Rufus recommends using ISO Image mode, so that you always have full access to the drive after writing it.
However, if you encounter issues during boot, you can try writing this image again in DD Image mode.

Please select the mode that you want to use to write this image:

● Write in ISO Image mode (Recommended)

○ Write in DD Image mode

7 Select Write in ISO Image mode and click OK.

9.2.6 Re-Imaging the System From a USB Flash Drive

These instructions describe how to re-image the system from a USB flash drive. For information about how to restore the system remotely, see **"Re-Imaging the System Remotely"** .

Before re-imaging the system from a USB flash drive, ensure that you have a bootable USB flash drive that contains the current DGX A100 software image.

1 Plug the USB flash drive containing the OS image into the DGX A100 system.

2 Connect a monitor and keyboard directly to the DGX A100 system.

3 Boot the system and press F11 when the NVIDIA logo appears to get to the boot menu.

4 Select the USB volume name that corresponds to the inserted USB flash drive, and boot the system from it.

5 When the system boots up, select Install DGX Server on the startup screen.

If you are an advanced user who is not using the RAID disks as cache and want to keep data on the RAID disks, then select Install DGX Server without formatting RAID. See the section **"Retaining the RAID Partition While Installing the OS"** for more information.

6 Press Enter.

The DGX A100 system will reboot and proceed to install the image. This can take more than 15 minutes.

Note: The Mellanox InfiniBand driver installation may take approximately 30 minutes, depending on how may cards undergo a firmware update.

After the installation is completed, the system then reboots into the OS.

Refer to **"First-Boot Setup"** for the steps to take when booting up the DGX A100 system for the first time after a fresh installation.

9.2.7 Retaining the RAID Partition While Installing the OS

The re-imaging process creates a fresh installation of the DGX OS. During the OS installation or re-image process, you are presented with a boot menu when booting the installer image. The default selection is Install DGX Software. The installation process then repartitions all the SSDs, including the OS SSD as well as the RAID SSDs, and the RAID array is mounted as /raid. This overwrites any data or file systems that may exist on the OS disk as well as the RAID disks.

Since the RAID array on the DGX A100 system is intended to be used as a cache and not for long-term data storage, this should not be disruptive. However, if you are an advanced user and have set up the disks for a non-cache purpose and want to keep the data on those drives, then select the Install DGX Server without formatting RAID option at the boot menu during the boot installation. This option retains data on the RAID disks and performs the following:

- Installs the cache daemon but leaves it disabled by commenting out the RUN=yes line in `/etc/default/cachefilesd`.
- Creates a `/raid` directory, leaves it out of the file system table by commenting out the entry containing "/raid" in `/etc/fstab`.
- Does not format the RAID disks.

When the installation is completed, you can repeat any configurations steps that you had performed to use the RAID disks as other than cache disks.

You can always choose to use the RAID disks as cache disks at a later time by enabling cachefilesd and adding /raid to the file system table as follows:

1. Uncomment the #RUN=yes line in `/etc/default/cachefiled`.
2. Uncomment the /raid line in `etc/fstab`.
3. Run the following:
   a. Mount /raid.

```
$ sudo mount /raid
```

b Start the cache daemon.

```
$ systemctl start cachefilesd
```

These changes are preserved across system reboots.

CHAPTER 10 USING THE BMC

The NVIDIA DGX A100 system comes with a baseboard management controller (BMC) for monitoring and controlling various hardware devices on the system. It monitors system sensors and other parameters.

10.1 CONNECTING TO THE BMC

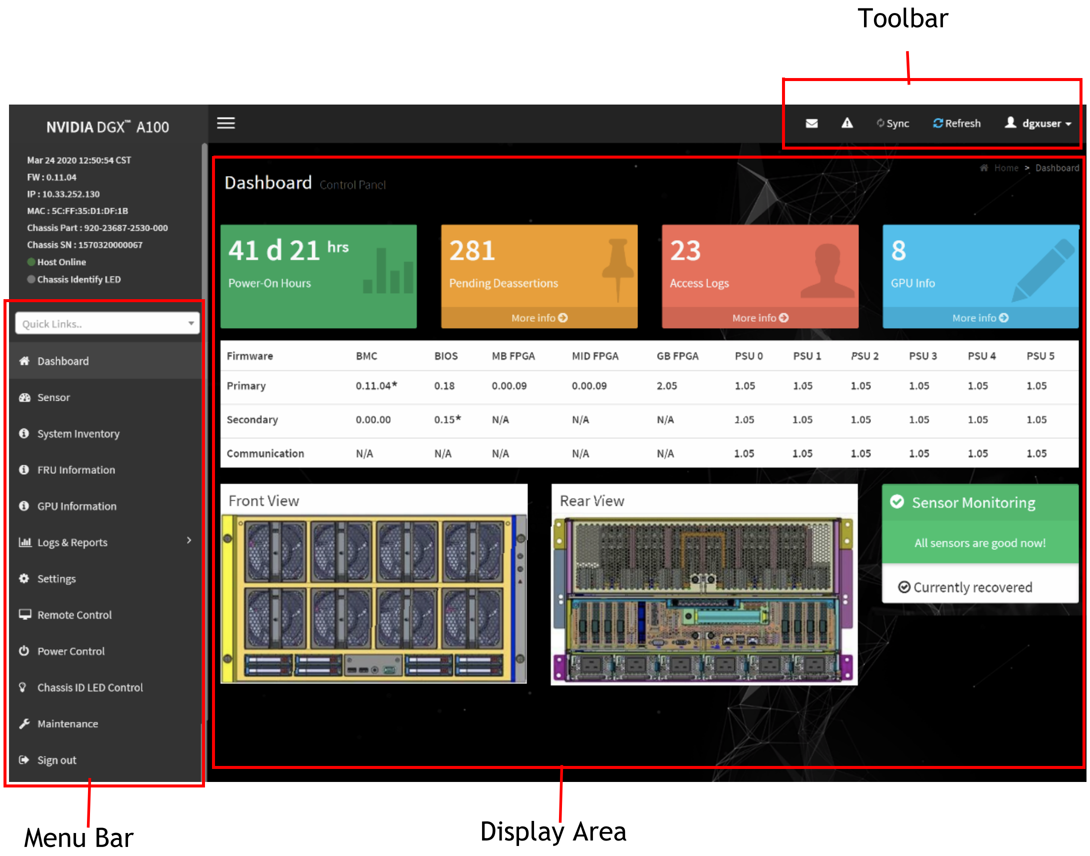
1. Make sure you have connected the BMC port on the DGX A100 system to your LAN.
2. Open a browser within your LAN and go to:

   ```
   https://<bmc-ip-address>/
   ```

   The BMC is supported on the following browsers:

   - Internet Explorer 11 and later
   - Firefox 29.0 (64-bit) and later
   - Google Chrome 7.0.3396.87 (64-bit) and later
3. Log in.

The BMC dashboard opens.



Mar 24 2020 12:50:54 CST
FW : 0.11.04
IP : 10.33.252.130
MAC : 5C:FF:35:D1:DF:1B
Chassis Part : 920-23687-2530-000
Chassis SN : 1570320000067
● Host Online
● Chassis Identify LED

41 d 21 hrs Power-On Hours

281 Pending Deassertions

23 Access Logs

8 GPU Info

Firmware	BMC	BIOS	MB FPGA	MID FPGA	GB FPGA	PSU 0	PSU 1	PSU 2	PSU 3	PSU 4	PSU 5
Primary	0.11.04*	0.18	0.00.09	0.00.09	2.05	1.05	1.05	1.05	1.05	1.05	1.05
Secondary	0.00.00	0.15*	N/A	N/A	N/A	1.05	1.05	1.05	1.05	1.05	1.05
Communication	N/A	N/A	N/A	N/A	N/A	1.05	1.05	1.05	1.05	1.05	1.05

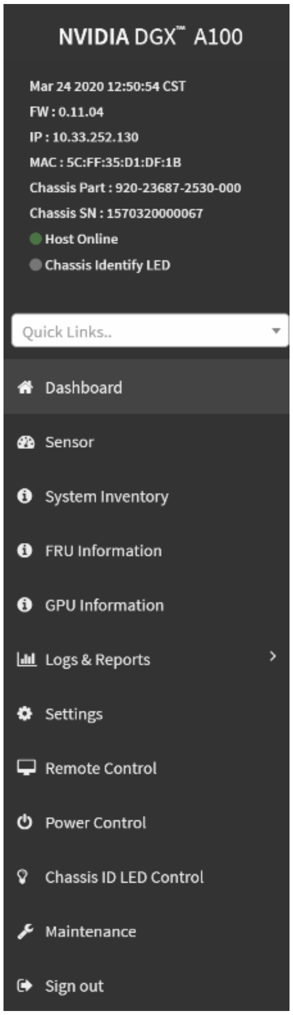
Sensor Monitoring

All sensors are good now!

Currently recovered

10.2 OVERVIEW OF BMC CONTROLS

The left-side navigation menu bar on the BMC main page contains the primary controls.



NVIDIA DGX™ A100

Mar 24 2020 12:50:54 CST

FW : 0.11.04

IP : 10.33.252.130

MAC : 5C:FF:35:D1:DF:1B

Chassis Part : 920-23687-2530-000

Chassis SN : 1570320000067

● Host Online

● Chassis Identify LED

Table 10.1 BMC Main Controls

Control	Descsription
Quick Links ...	Provides quick access to several tasks.
Dashboard	Displays the overall information about the status of the device.
Sensor	Provides status and readings for system sensors, such as SSD, PSUs, voltages, CPU temperatures, DIMM temperatures, and fan speeds.
System Inventory	Displays inventory information of system modules: System, Processor, Memory Controller, BaseBoard, Power, Thermal, PCIE Device, PCIE Function, Storage.
FRU Information	Provides, chassis, board, and product information for each field-replaceable unit (FRU) device.

Table 10.1 BMC Main Controls (Continued)

Control	Descsription
GPU Information	Provides basic information on all the GPUs in the systems, including GUID, VBIOS version, InfoROM version, and number of retired pages for each GPU.
Logs and Reports	View, and if applicable, download and erase, the IPMI event log, and System, Audit, Video, and POST Code logs.
Settings	Configure the following settings: Captured BSOD, 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, User Management, Video Recording
Remote Control	Opens the KVM Launch page for accessing the DGX A100 console remotely.
Power Control	Perform the following power actions: Power On, Power Off, Power Cycle, Hard Reset, ACP/Shutdown
Chassis ID LED Control	Lets you to change the chassis ID LED behavior: Off, Solid On, Blinking On (select from 5 to 255 second blinking intervals)
Maintenance	Perform the following maintenance tasks: Backup Configuration, Firmware Image Location, Firmware Update, Preserve Configuration, Restore Configuration, Restore Factory Defaults, System Administrator
Sign out	Sign out of the BMC web UI.

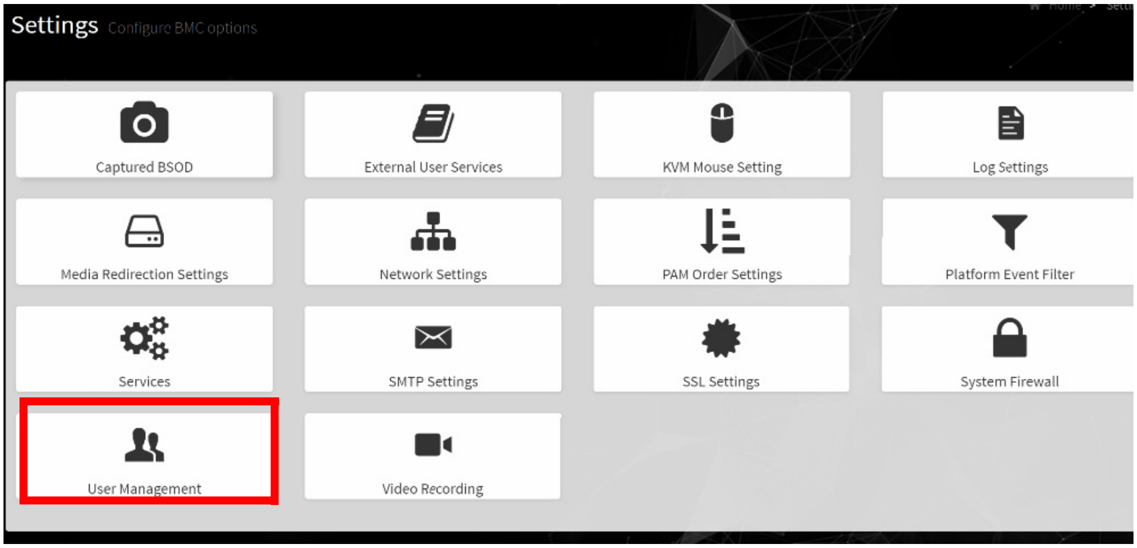
10.3 COMMON BMC TASKS

10.3.1 Changing BMC Login Credentials

Adding/Removing Users

1 Select **Settings** from the left-side navigation menu.

2 Select the **User Management** card.

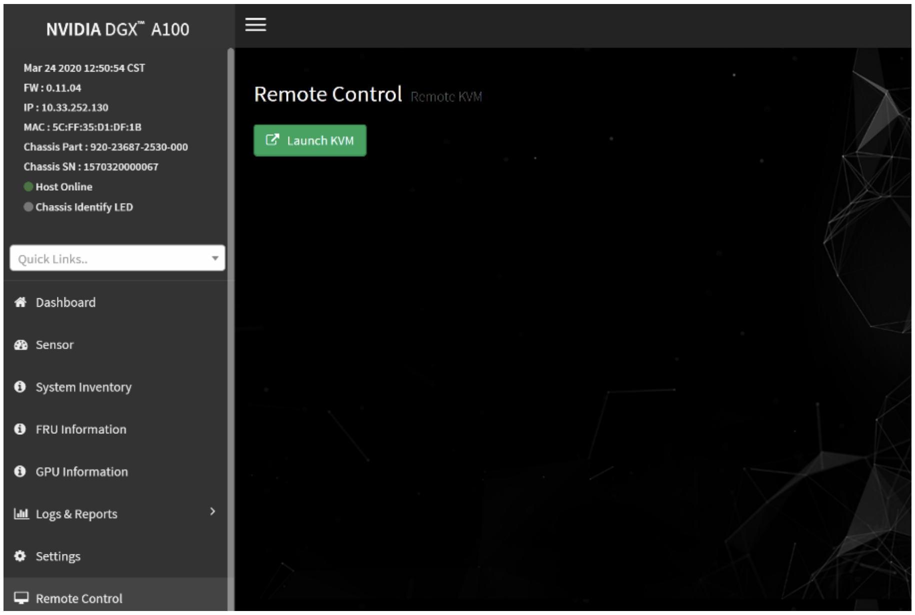


3 Click the Help icon (?) for information about configuring users and creating a password.

4 Log out and then log back in with the new credentials.

10.3.2 Using the Remote Console

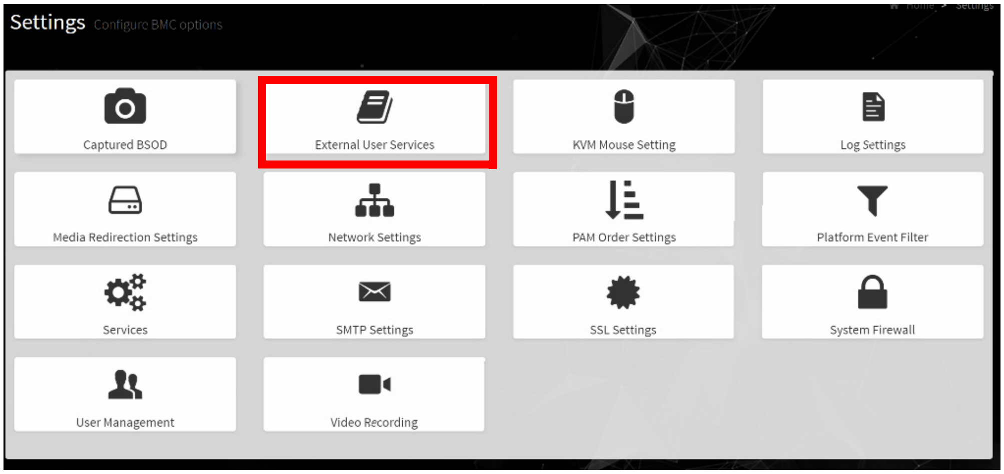
1. Click **Remote Control** from the left-side navigation menu.
2. Click **Launch KVM** to start the remote KVM and access the DGX A100 console.



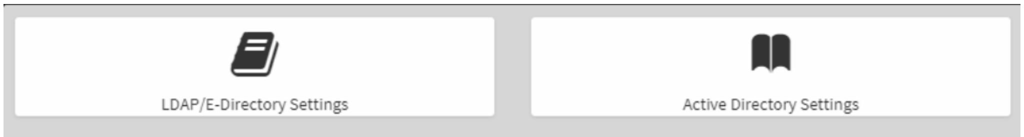
Mar 24 2020 12:50:54 CST
FW : 0.11.04
IP : 10.33.252.130
MAC : 5C:FF:35:D1:DF:1B
Chassis Part : 920-23687-2530-000
Chassis SN : 1570320000067
● Host Online
● Chassis Identify LED

10.3.3 Setting Up Active Directory or LDAP/E-Directory

1. From the side navigation menu, click **Settings** and then click **External User Services**.

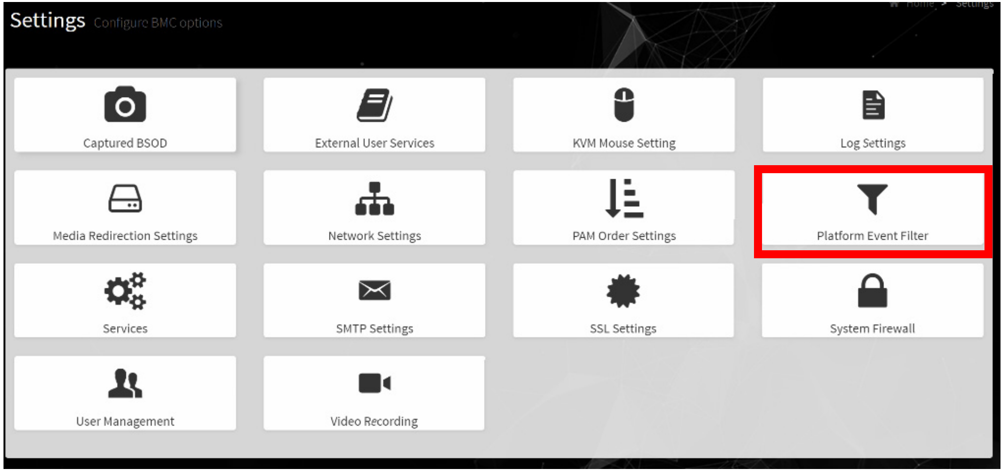


2. Click either **Active Directory Settings** or **LDAP/E-Directory Settings** and then follow the instructions.



10.3.4 Configuring Platform Event Filters

From the side navigation menu, click **Settings** and then click **Platform Event Filters.**



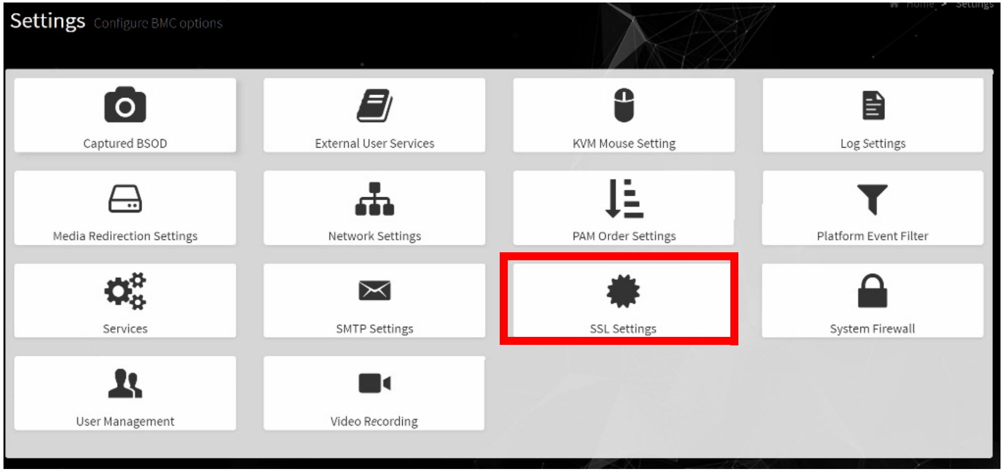
The Event Filters page shows all configured event filters and available slots. You can modify or add new event filter entry on this page.

- 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.

10.3.5 Uploading or Generating SSL Certificates

Two methods are available for setting up a new certificate - generating a (self-signed) SSL, or uploading an SSL (for example, to use a Trusted CA-signed certificate).

From the side navigation menu, click **Settings** and then click **External User Services**.

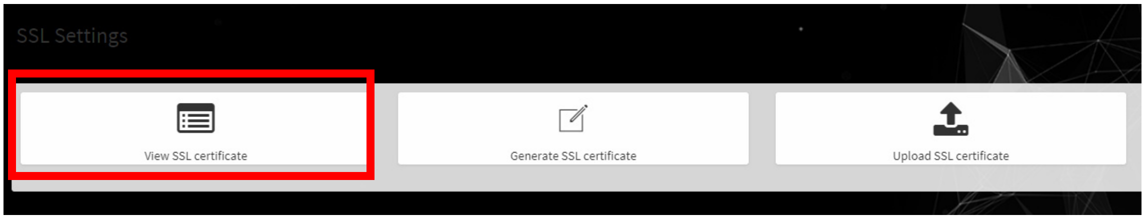


Refer to the following sections for instructions on

- Viewing the SSL Certificate
- Generating an SSL Certificate
- Uploading an SSL Certificate

10.3.5.1 Viewing the SSL Certificate

From the SSL Setting page, select **View SSL Certificate**.

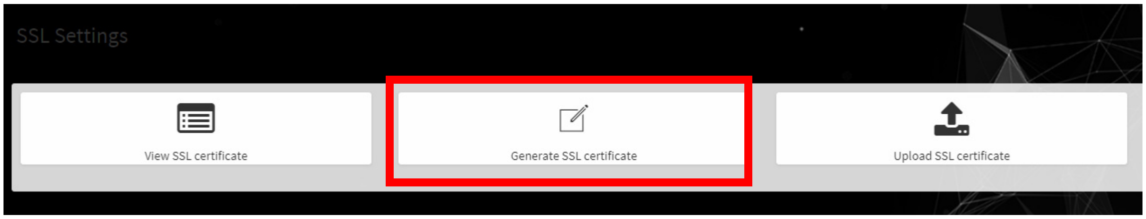


The View SSL Certificate page displays the basic information about the uploaded SSL certificate.

- Certificate Version, Serial Number, Algorithm, and Public Key
- Issuer information
- Valid Date range
- Issued to information

10.3.5.2 Generating the SSL Certificate

1 From the SSL Setting page, select **Generate SSL Certificate**.



2 Fill in the information as described in the following table.

Item	Description/Requirements
Common Name (CN)	The common name for which the certificate is to be generated. •§ Maximum length of 64 alpha-numeric characters. •§ Special characters '#' and '$' are not allowed.
Organization (O)	The name of the organization for which the certificate is generated. •Maximum length of 64 alpha-numeric characters. •§ Special characters '#' and '$' are not allowed.
Organization Unit (OU)	Overall organization section unit name for which the certificate is generated. •Maximum length of 64 alpha-numeric characters. •§ Special characters '#' and '$' are not allowed.
City or Locality (L)	City or Locality of the organization (mandatory) •Maximum length of 64 alpha-numeric characters. •§ Special characters '#' and '$' are not allowed.
State or Province (ST)	State or Province of the organization (mandatory) •Maximum length of 64 alpha-numeric characters. •§ Special characters '#' and '$' are not allowed.
Country (C)	Country code of the organization. •Only two characters are allowed. •Special characters are not allowed.
Email Address	Email address of the organization (mandatory)
Valid for	Validity of the certificate. Enter a range from 1 to 3650 (days)
Key Length	The key length bit value of the certificate (Ex. 2048 bits)

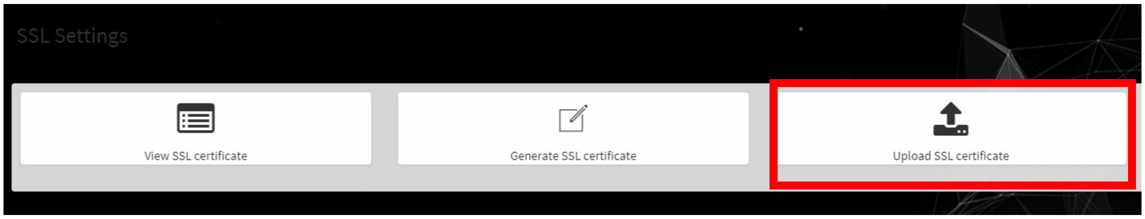
3 Click **Save** to generate the new certificate.

10.3.5.3 Uploading the 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 From the SSL Setting page, select **Upload SSL Certificate**.



2 Click the **New Certificate** folder icon, then browse to locate the appropriate file and select it.

3 Click the **New Private Key** folder icon, then browse and locate the appropriate file and select it.

4 Click **Save**.

CHAPTER 11 MULTI-INSTANCE GPU

Multi-Instance GPU (MIG) is a new capability of the NVIDIA A100 GPU. MIG uses spatial partitioning to carve the physical resources of a single A100 GPU into as many as seven independent GPU instances. These instances run simultaneously, each with its own memory, cache, and compute streaming multiprocessors. MIG enables the A100 GPU to deliver guaranteed quality of service at up to 7X higher utilization compared to non-MIG enabled GPUs.

MIG enables:

- GPU memory isolation among parallel GPU workloads
- Physical allocation of resources used by parallel GPU workloads

This chapter describes the basic instructions for using MIG when running NGC containers.

11.1 ENABLING MIG ON THE DGX A100 SYSTEM

Management of MIG instances is accomplished using the NVIDIA Management Library (NVML) APIs or its command-line utility (nvidia-smi). Enablement of MIG requires a GPU reset and hence some system services that manage GPUs should be terminated before enabling MIG.

To enable MIG on all eight GPUs in the system, issue the following.

1 Stop the NVSM and DCGM services.

```
$ sudo systemctl stop nvsm
$ sudo systemctl stop dcgm
```

2 Enable MIG on all eight GPUs.

```
$ sudo nvidia-smi -mig 1
```

3 Restart the NVSM and DCGM services.

```
$ sudo systemctl start nvsm
$ sudo systemctl start dcgm
```

11.2 VIEWING AVAILABLE PROFILES

11.2.1 Viewing GPU Profiles

You will need to know the profile ID when setting up MIG on a specific GPU. To view the available profiles (configurations), issue the following.

```
$ nvidia-smi mig -i 0 -lgip
```

Example output:

```
+-----------------------------------------------------------------------
+
| GPU instance profiles:                                                |
| GPU  Name       ID    Instances   Memory     P2P    SM    DEC   ENC   |
|                       Free/Total   GiB              CE    JPEG  OFA   |
|=======================================================================
|
|   0  1_SLICE    19     7/7        4.83       No     14     0     0    |
|                                                       1     0     0    |
+-----------------------------------------------------------------------
+
|   0  2_SLICE    14     3/3        9.65       No     28     1     0    |
|                                                       2     0     0    |
+-----------------------------------------------------------------------
+
|   0  3_SLICE     9     2/2        19.31      No     42     2     0    |
|                                                       3     0     0    |
+-----------------------------------------------------------------------
+
|   0  4_SLICE     5     1/1        19.31      No     56     2     0    |
|                                                       4     0     0    |
+-----------------------------------------------------------------------
+
|   0  7_SLICE     0     1/1        38.61      No     98     5     0    |
|                                                       7     1     1    |
+-----------------------------------------------------------------------
+
```

11.2.2 Viewing Compute Profiles

You will need to know the compute profile ID when setting up compute instances on a specific GPU instance. To view the available profiles (configurations) for a particular GPG instance, issue the following.

Syntax:

```
nvidia-smi mig -i <gpu-id> -gi <gpu-instance-id> -lcip
```

where

`<gpu-id>` is the GPU device ID (0,1,2,3,4,5,6,7)

`<gpu-instance-id>` is the GPU instance ID

Example:

The following example lists the available MIG compute profiles on GPU0/GPU instance ID 0:

```
root# nvidia-smi mig -i 0 -gi 0 -lcip
+-----------------------------------------------------------------------------+
| Compute instance profiles:                                                  |
| GPU     GPU      Profile   Profile  Instances   Exclusive     Shared        |
|       Instance    Name       ID     Free/Total     SM     DEC   ENC   OFA   |
|         ID                                                CE    JPEG        |
|=============================================================================|
|   0      0       7_1_SLICE    0      7/7           14      5     0     1    |
|                                                             7     1          |
+-----------------------------------------------------------------------------+
|   0      0       7_2_SLICE    1      3/3           28      5     0     1    |
|                                                             7     1          |
+-----------------------------------------------------------------------------+
|   0      0       7_3_SLICE    2      2/2           42      5     0     1    |
|                                                             7     1          |
+-----------------------------------------------------------------------------+
|   0      0       7_4_SLICE    3      1/1           56      5     0     1    |
|                                                             7     1          |
+-----------------------------------------------------------------------------+
|   0      0       7_7_SLICE    4*     1/1           98      5     0     1    |
|                                                             7     1          |
+-----------------------------------------------------------------------------+
```

11.3 CREATING MIG INSTANCES

MIG instances include GPI instances and compute instances.

Creating GPU instances can be thought of as splitting one big GPU into multiple smaller GPUs, with each GPU instance having dedicated compute and memory resources. Each GPU Instance behaves like a smaller, fully capable independent GPU that includes a predefined number of streaming multiprocessors (SM), L2 cache slices, memory controllers, and frame buffer memory.

A *GPU compute instance* is another grouping that can configure different levels of compute power created within a GPU instance, encapsulating all the compute resources (such as number of Copy Engines and NVDEC units) that can execute work in the GPU instance. By default, a single GPU compute instance is created under each GPU instance, exposing all the GPU compute resources available within the GPU instance. A GPU instance can be subdivided into multiple smaller GPU compute instances to further split its compute resources.

11.3.1 Creating a GPU instance

Syntax

```
nvidia-smi mig -i <gpu-id> -cgi <profile>[,<profile>...]
```

where

<gpu-id> is the GPU device ID (0, 1, 2, 3, 4, 5, 6, 7)

<profile> is the MIG profile ID

Example

The following example creates a 7-slice GPU instance on GPU 0.

```
nvidia-smi mig -i 0 -cgi 0
```

This can be verified by issuing the following:

```
root# nvidia-smi mig -i 0 -lgi
```

```
+------------------------------------------------+
| GPU instances:                                 |
| GPU  Profile    Profile  Instance   Placement  |
|       Name        ID       ID       Start:Size |
|================================================|
|   0  7_SLICE       0        0          0:8     |
+------------------------------------------------+
```

11.3.2 Creating a Compute Instance

By default, a compute instance is created for each GPU instance, but you can create additional compute instances if needed. When creating compute instance, consider the number of streaming multiprocessors that are available according the compute profile ID.

Syntax:

```
nvidia-smi mig -i <gpu-id> -gi <gpu-instance-id> -cci <compute-mig-
profile-id>
```

Examples:

The following example creates a compute instance corresponding to CI profile name "7_3_SLICE" (ID 2)

```
root# nvidia-smi mig -i 0 -gi 0 -cci 2
```

```
Successfully created compute instance on GPU  0 GPU instance ID  0 using
profile ID  2
```

The following example creates another compute instance corresponding to CI profile ID 2

```
root# nvidia-smi mig -i 0 -gi 0 -cci 2
```

```
Successfully created compute instance on GPU  0 GPU instance ID  0 using
profile ID  2
```

The following example creates a compute instance corresponding to CI profile name "7_1_SLICE" (ID 0)

```
root# nvidia-smi mig -i 0 -gi 0 -cci 0
```

```
Successfully created compute instance on GPU  0 GPU instance ID  0 using
profile ID  0
```

The following command lists the resulting GPU MIG devices that can be used in CUDA applications and services.

```
root# nvidia-smi mig -i 0 -gi 0 -lci
```

11.4 USING MIG WITH DOCKER CONTAINERS

The NVIDIA Container Toolkit (v1.0.6 or later) has been extended to support running containers with Docker when the A100 GPU is in MIG mode.

To run containers, use the native GPU support provided with Docker 19.03 and later (included in the DGX OS installed on the DGX A100 system). Specify the GPU instance or compute instance using the "`device=`" parameter of the `--gpu` option as shown in the syntax below.

Syntax:

```
docker run --gpus '"device=<gpu-id>:<mig-id>"' <container path>/
<repository>:<tag> <command>
```

Example

```
docker run --gpus '"device=0:1"' nvidia/cuda:11.0-base nvidia-smi
```

11.5 DELETING MIG INSTANCES

MIG instances can be deleted only if they are idle, meaning that there are no CUDA applications running on the instance. Deleeting a MIG instance does not affect other instances being used on the same GPU.

11.5.1 MIG Instance Deletion Process

First remove any compute instances.

Syntax:

```
root# nvidia-smi mig -i <gpu-id> -gi <gpu-instance-id> -ci <compute-
instance-id> -dci
```

Then you can remove the GPU instances.

Syntax:

```
root# nvidia-smi mig -i <gpu-id> -gi <gpu-instance-id> -dgi
```

11.5.2 MIG Instance Deletion Examples

Example of attempting to delete a GPU instance without first deleting the compute instances.

```
root# nvidia-smi mig -i 0 -gi 0 -dgi
Unable to destroy GPU instance ID 0 from GPU 0: In use by another client
Failed to destroy GPU instances: In use by another client
```

Example showing MIG deletion of a GPU instance 0 where two compute instances (0 and 1) were created from it.

1. Delete compute instance 1 on GPU instance 0

```
root# nvidia-smi mig -i 0 -gi 0 -ci 1 -dci
Successfully destroyed compute instance ID 1 from GPU 0 GPU instance ID
0
```

2. Delete compute instance 0 on GPU instance 0.

```
root# nvidia-smi mig -i 0 -gi 0 -ci 0 -dci
Successfully destroyed compute instance ID 0 from GPU 0 GPU instance ID 0
```

3. Delete GPU instance 0.

```
root# nvidia-smi mig -i 0 -gi 0 -dgi
Successfully destroyed GPU instance ID 0 from GPU 0
```

CHAPTER 12 SECURITY

12.1 USER SECURITY MEASURES

The NVIDIA DGX A100 system is a specialized server designed to be deployed in a data center. It must be configured to protect the hardware from unauthorized access and unapproved use. The DGX A100 system is designed with a dedicated BMC Management Port and multiple Ethernet network ports.

When installing the DGX A100 system in the data center, follow best practices as established by your organization to protect against unauthorized access.

12.1.1 Securing the BMC Port

NVIDIA recommends that the BMC port of the DGX A100 system be connected to a dedicated management network with firewall protection. If remote access to the BMC is required (such as for a system hosted at a co-location provider), it should be accessed through a secure method that provides isolation from the internet, such as through a VPN server.

12.2 SYSTEM SECURITY MEASURES

The NVIDIA DGX A100 system incorporates the following security measures.

12.2.1 Secure Flash of DGX A100 Firmware

Secure Flash is implemented for the DGX A100 to prevent unsigned and unverified firmware images from being flashed onto the system.

12.2.1.1 Encryption

- System firmware is encrypted during over-the-network upgrades.
- The firmware encryption algorithm is AES-CBC.
- The firmware encryption key strength is 128 bits or higher.
- Each firmware class uses a unique encryption key.
- Firmware decryption is performed either by the same agent that performs signature check or a more trusted agent in the same COT

12.2.1.2 Signing

- The firmware signature is validated upon each boot of the DGX A100.

  This is not implemented for the PSU firmware and CPLD.
- The firmware signature is validated on every update before the firmware image is updated in non-volatile storage.

12.2.2 NVSM Security

See **Configuring NVSM Security**.

12.3 SECURE DATA DELETION

This section explains how to securely delete data from the NVIDIA DGX A100 system SSDs to permanently destroy all the data that was stored there. This performs a more secure SSD data deletion than merely deleting files or reformatting the SSDs.

12.3.1 Prerequisite

Prepare a bootable installation medium that contains the current DGX OS Server ISO image.

See:

- **“Obtaining the DGX A100 Software ISO Image and Checksum File”**
- **“Creating a Bootable Installation Medium”**

12.3.2 Instructions

1. Boot the system from the ISO image, either remotely or from a bootable USB key.
2. At the GRUB menu, choose '**Rescue a broken system'**, then configure the locale and network information.
3. When asked to choose a root file system, choose

   '**`Do not use a root file system`**'

   and then

   '**`Execute a shell in the installer environment'`**
4. Log in.
5. Run the following command to identify the devices available in the system:

```
$ sudo nvme list
```

6. Run `nvme format -s1` on all storage devices listed.

   Syntax:

```
$ sudo nvme format -s1 <device-path>
```

   where

   `<device-path>` is the specific storage node as listed in the previous step.

APPENDIX A INSTALLING SOFTWARE ON AIR-GAPPED DGX A100 SYSTEMS

For security purposes, some installations require that systems be isolated from the internet or outside networks. Since most DGX A100 software updates are accomplished through an over-the-network process with NVIDIA servers, this section explains how updates can be made when using an over-the-network method is not an option. It includes a process for installing Docker containers as well.

A.1 INSTALLING NVIDIA DGX A100 SOFTWARE

One method for updating DGX A100 software on an air-gapped DGX A100 system is to download the ISO image, copy it to removable media and then re-image the DGX A100 System from the media. This method is available only for software versions that are available as ISO images for download.

Alternately, you can update the DGX A100 software by performing a network update from a local repository. This method is available only for software versions that are available for over-the-network updates.

A.2 RE-IMAGING THE SYSTEM

! **CAUTION:** This process destroys all data and software customizations that you have made on the DGX A100 System. Be sure to back up any data that you want to preserve and push any Docker images that you want to keep to a trusted registry.

1 Obtain the ISO image from the NVIDIA Enterprise Services.

   a Log on to the **NVIDIA Enterprise Support** site and click the Announcements tab to locate the DGX OS Server image ISO file.

   b Download the image ISO file.

2 Refer to the instructions in the **"Restoring the DGX A100 Software Image"** section for additional instructions.

A.3 CREATING A LOCAL MIRROR OF THE NVIDIA AND CANONICAL REPOSITORIES

The procedure below describes how to download all the necessary packages to create a mirror of the repositories that are needed to update NVIDIA DGX A100 systems. For more information on DGX OS versions and the release notes available, visit **https://docs.nvidia.com/dgx/dgx-os-server-release-notes/index.html#dgx-os-release-number-scheme**.

Note: These procedures apply only to upgrades within the same major release, such as 4.x → 4.y. It does not support upgrades across major releases, such as 3.x → 4.x..

A.3.1 Create Mirrors

The instructions in this section are to be performed on a system with network access.

Prerequisites

- A system installed with Ubuntu OS is needed to create the mirror because there are several Ubuntu tools that need to be used.
- The system must contain enough storage space to replicate the repositories to a filesystem; the space requirement could be as high as 250GB.
- An efficient way to move large amount of data; for example, shared storage in a DMZ, or portable USB drives that can be brought into the air-gapped area.

The data will need to be moved to the systems that need to be updated. Make sure the portable drive is formatted using ext4 or FAT32.

1. Make sure the storage device is attached to the system with network access and identify the mount point.

   Example mount point: `/media/usb/repository`

2. Once the space requirement has been met, install the apt-mirror package.

   Make sure the target directory is owned by the user apt-mirror or the replication will not work.

```
$ sudo apt update
$ sudo apt install apt-mirror
$ sudo chown apt-mirror:apt-mirror /media/usb/repository
```

3. Configure the path of the destination directory in `/etc/apt/mirror.list` and use the included list of repositories below to retrieve the packages for both Ubuntu base OS as well as the NVIDIA DGX OS packages:

```
############# config ##################
#
set base_path   /media/usb/repository  #/your/path/here
#
# set mirror_path  $base_path/mirror
# set skel_path    $base_path/skel
# set var_path     $base_path/var
# set cleanscript $var_path/clean.sh
# set defaultarch  <running host architecture>
# set postmirror_script $var_path/postmirror.sh
set run_postmirror 0
set nthreads     20
set _tilde 0
#
############# end config ##############
```

```
# Standard Canonical package repositories:
deb http://security.ubuntu.com/ubuntu bionic-security main
deb http://security.ubuntu.com/ubuntu bionic-security universe
deb http://security.ubuntu.com/ubuntu bionic-security multiverse
deb http://archive.ubuntu.com/ubuntu/ bionic main multiverse universe
deb http://archive.ubuntu.com/ubuntu/ bionic-updates main multiverse
universe
#
deb-i386 http://security.ubuntu.com/ubuntu bionic-security main
deb-i386 http://security.ubuntu.com/ubuntu bionic-security universe
deb-i386 http://security.ubuntu.com/ubuntu bionic-security multiverse
deb-i386 http://archive.ubuntu.com/ubuntu/ bionic main multiverse
universe
deb-i386 http://archive.ubuntu.com/ubuntu/ bionic-updates main
multiverse universe
#
```

```
# DGX specific repositories:
deb http://international.download.nvidia.com/dgx/repos/bionic bionic
main restricted universe multiverse
deb http://international.download.nvidia.com/dgx/repos/bionic bionic-
updates main restricted universe multiverse
deb http://international.download.nvidia.com/dgx/repos/bionic bionic-
r418+cuda10.1 main multiverse restricted universe
```

```
#
deb-i386 http://international.download.nvidia.com/dgx/repos/bionic
bionic main restricted universe multiverse
deb-i386 http://international.download.nvidia.com/dgx/repos/bionic
bionic-updates main restricted universe multiverse
# Only for DGX OS 4.1.0
deb-i386 http://international.download.nvidia.com/dgx/repos/bionic
bionic-r418+cuda10.1 main multiverse restricted universe
```

```
# Clean unused items
clean http://archive.ubuntu.com/ubuntu
clean http://security.ubuntu.com/ubuntu
```

4 Run apt-mirror and wait for it to finish downloading content. This will take a long time depending on the network connection speed.

```
$ sudo apt-mirror
```

5 Eject the removable storage with all packages.

```
$ sudo eject /media/usb/repository
```

A.3.2 Configure the Target System

The instructions in this section are to be performed on the target system.

Prerequisites

- The target DGX A100 system is installed, has gone through the first boot process, and is ready to be updated with the latest packages.
- A USB storage device is attached to the target DGX A100 system.

  There are other ways to transfer the data that are not covered in this document as they will depend on the data center policies for the air-gapped environment.

1 Mount the storage device on the air-gapped system to `/media/usb/repository` for consistency.

2 Configure apt to use the filesystem as the repository in the file `/etc/apt/sources.list` by modifying the following lines.

```
deb file:///media/usb/repository/mirror/security.ubuntu.com/ubuntu
bionic-security main
deb file:///media/usb/repository/mirror/security.ubuntu.com/ubuntu
bionic-security universe
deb file:///media/usb/repository/mirror/security.ubuntu.com/ubuntu
bionic-security multiverse
deb file:///media/usb/repository/mirror/archive.ubuntu.com/ubuntu/
bionic main multiverse universe
deb file:///media/usb/repository/mirror/archive.ubuntu.com/ubuntu/
bionic-updates main multiverse universe
```

3 Configure apt to use the NVIDIA DGX OS packages in the file `/etc/apt/sources.list.d/dgx.list`.

```
deb file:///media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic main
multiverse restricted universe
```

4 If present, remove the file `/etc/apt/sources.list.d/docker.list` as it is no longer needed and it will eliminate error messages during the update process.

5 Configure apt to use the NVIDIA DGX OS packages in the file `/etc/apt/sources.list.d/dgx-bionic-r418-cuda10-1-repo.list`

```
deb file:///media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic/ bionic-
r418+cuda10.1 main multiverse restricted universe
```

6 Edit the file `/etc/apt/preferences.d/nvidia` to update the Pin parameter as follows.

```
Package: *
#Pin: origin international.download.nvidia.com
Pin: release o=DGX Server
Pin-Priority: 600
```

7 Update the apt repository and confirm there are no errors.

```
$ sudo apt update
```

```
Get:1 file:/media/usb/repository/mirror/security.ubuntu.com/ubuntu
bionic-security InRelease [88.7 kB]
Get:1 file:/media/usb/repository/mirror/security.ubuntu.com/ubuntu
bionic-security InRelease [88.7 kB]
Get:2 file:/media/usb/repository/mirror/archive.ubuntu.com/ubuntu
bionic InRelease [242 kB]
Get:2 file:/media/usb/repository/mirror/archive.ubuntu.com/ubuntu
bionic InRelease [242 kB]
Get:3 file:/media/usb/repository/mirror/archive.ubuntu.com/ubuntu
bionic-updates InRelease [88.7 kB]
Get:4 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic-r418+cuda10.1
InRelease [13.0 kB]
Get:5 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic InRelease
[13.1 kB]
Get:3 file:/media/usb/repository/mirror/archive.ubuntu.com/ubuntu
bionic-updates InRelease [88.7 kB]
Get:4 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic-r418+cuda10.1
InRelease [13.0 kB]
Get:5 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic InRelease
[13.1 kB]
Hit:6 https://download.docker.com/linux/ubuntu bionic InRelease
Get:7 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic-
r418+cuda10.1/multiverse amd64 Packages [10.1 kB]
Get:8 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic-
r418+cuda10.1/restricted amd64 Packages [10.3 kB]
Get:9 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic-
r418+cuda10.1/restricted i386 Packages [516 B]
Get:10 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic/multiverse
amd64 Packages [44.5 kB]
Get:11 file:/media/usb/repository/mirror/
```

```
international.download.nvidia.com/dgx/repos/bionic bionic/multiverse
i386 Packages [8,575 B]
Get:12 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic/restricted
i386 Packages [745 B]
Get:13 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic/restricted
amd64 Packages [8,379 B]
Get:14 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic/universe amd64
Packages [2,946 B]
Get:15 file:/media/usb/repository/mirror/
international.download.nvidia.com/dgx/repos/bionic bionic/universe i386
Packages [496 B]
Reading package lists... Done
Building dependency tree
Reading state information... Done
249 packages can be upgraded. Run 'apt list --upgradable' to see them.
$
```

8. Upgrade the system using the newly configured local repositories.

```
$ sudo apt full-upgrade
```

A.4 INSTALLING DOCKER CONTAINERS

This method applies to Docker containers hosted on the NVIDIA NGC Container Registry, and requires that you have an active NGC account.

1. On a system with internet access, log in to the NGC Container Registry by entering the following command and credentials.

```
$ docker login nvcr.io
```

```
Username: $oauthtoken
Password: apikey
```

   Type “$oauthtoken” exactly as shown for the Username. This is a special username that enables API key authentication. In place of **apikey**, paste in the API Key text that you obtained from the NGC website.

2. Enter the docker pull command, specifying the image registry, image repository, and tag:

```
$ docker pull nvcr.io/nvidia/repository:tag
```

3. Verify the image is on your system using docker images.

```
$ docker images
```

4. Save the Docker image as an archive. .

```
$ docker save nvcr.io/nvidia/repository:tag > framework.tar
```

5. Transfer the image to the air-gapped system using removable media such as a USB flash drive.

6 Load the NVIDIA Docker image.

```
$ docker load –i framework.tar
```

7 Verify the image is on your system.

```
$ docker images
```

APPENDIX B SAFETY

B.1 SAFETY INFORMATION

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 server product.

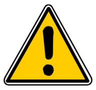
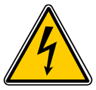
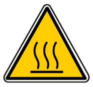
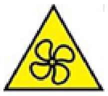
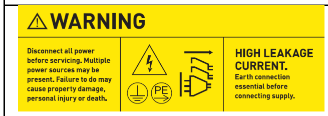
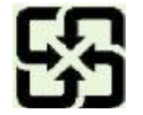
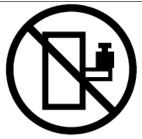
In the event of a conflict between the information in this document and information provided with the product or on the website for a particular product, the product documentation takes precedence.

Your server should be integrated and serviced only by technically qualified persons.

You must adhere to the guidelines in this guide and the assembly instructions in your server manuals to ensure and maintain compliance with existing product certifications and approvals. Use only the described, regulated components specified in this guide. Use of other products I components will void the UL Listing and other regulatory approvals of the product, and may result in noncompliance with product regulations in the region(s) in which the product is sold.

B.2 SAFETY WARNINGS AND CAUTIONS

To avoid personal injury or property damage, before you begin installing the product, read, observe, and adhere to all of the following safety instructions and information. The following safety symbols may be used throughout the documentation and may be marked on the product and/or the product packaging.

Symbol	Meaning
CAUTION	Indicates the presence of a hazard that may cause minor personal injury or property damage if the CAUTION is ignored.
WARNING	Indicates the presence of a hazard that may result in serious personal injury if the WARNING is ignored.
	Indicates potential hazard if indicated information is ignored.
	Indicates shock hazards that result in serious injury or death if safety instructions are not followed
	Indicates hot components or surfaces.
	Indicates do not touch fan blades, may result in injury.
 ⚠WARNING Disconnect all power before servicing. Multiple power sources may be present. Failure to do may cause property damage, personal injury or death. HIGH LEAKAGE CURRENT. Earth connection essential before connecting supply.	Shock hazard - Product might be equipped with multiple power cords. To remove all hazardous voltages, disconnect all power cords.
	High leakage current ground(earth) connection to the Power Supply is essential before connecting the supply.
	Recycle the battery.
	The rail racks are designed to carry only the weight of the server system. Do not use rail-mounted equipment as a workspace. Do not place additional load onto any rail-mounted equipment.

B.3 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.4 SITE SELECTION

Choose a site that is:

- Clean, dry, and free of airborne particles (other than normal room dust).
- Well-ventilated and away from sources of heat including direct sunlight and radiators.
- Away from sources of vibration or physical shock.
- In regions that are susceptible to electrical storms, we recommend you plug your system into a surge suppressor and disconnect telecommunication lines to your modem during an electrical storm.
- Provided with a properly grounded wall outlet.
- Provided with sufficient space to access the power supply cord(s), because they serve as the product's main power disconnect.

B.5 EQUIPMENT HANDLING PRACTICES

Reduce the risk of personal injury or equipment damage:

- Conform to local occupational health and safety requirements when moving and lifting equipment.
- Use mechanical assistance or other suitable assistance when moving and lifting equipment.

B.6 ELECTRICAL PRECAUTIONS

Power and Electrical Warnings

Caution: The power button, indicated by the stand-by power marking, DOES NOT completely turn off the system AC power; standby power is active whenever the system is plugged in. To remove power from system, you must unplug the AC power cord from the wall outlet. Make sure all AC power cords are unplugged before you open the chassis, or add or remove any non hot-plug components.

Do not attempt to modify or use an AC power cord if it is not the exact type required. A separate AC cord is required for each system power supply.

Some power supplies in servers use Neutral Pole Fusing. To avoid risk of shock use caution when working with power supplies that use Neutral Pole Fusing.

The power supply in this product contains no user-serviceable parts. Do not open the power supply. Hazardous voltage, current and energy levels are present inside the power supply. Return to manufacturer for servicing.

When replacing a hot-plug power supply, unplug the power cord to the power supply being replaced before removing it from the server.

To avoid risk of electric shock, tum off the server and disconnect the power cords, telecommunications systems, networks, and modems attached to the server before opening it.

Power Cord Warnings

Use certified AC power cords to connect to the server system installed in your rack.

Caution: To avoid electrical shock or fire, check the power cord(s) that will be used with the product as follows:

- Do not attempt to modify or use the AC power cord(s) if they are not the exact type required to fit into the grounded electrical outlets.
- The power cord(s) must meet the following criteria:
  - The power cord must have an electrical rating that is greater than that of the electrical current rating marked on the product.
  - The power cord must have safety ground pin or contact that is suitable for the electrical outlet.
  - The power supply cord(s) is/are the main disconnect device to AC power. The socket outlet(s) must be near the equipment and readily accessible for disconnection.
  - The power supply cord(s) must be plugged into socket-outlet(s) that is /are provided with a suitable earth ground.

B.7 SYSTEM ACCESS WARNINGS

Caution: To avoid personal injury or property damage, the following safety instructions apply whenever accessing the inside of the product:

- Turn off all peripheral devices connected to this product.

- Turn off the system by pressing the power button to off.
- Disconnect the AC power by unplugging all AC power cords from the system or wall outlet.
- Disconnect all cables and telecommunication lines that are connected to the system.
- Retain all screws or other fasteners when removing access cover(s). Upon completion of accessing inside the product, refasten access cover with original screws or fasteners.
- Do not access the inside of the power supply. There are no serviceable parts in the power supply.
- Return to manufacturer for servicing.
- Power down the server and disconnect all power cords before adding or replacing any non hot-plug component.
- When replacing a hot-plug power supply, unplug the power cord to the power supply being replaced before removing the power supply from the server.

Caution: If the server has been running, any installed processor(s) and heat sink(s) may be hot.

Unless you are adding or removing a hot-plug component, allow the system to cool before opening the covers. To avoid the possibility of coming into contact with hot component(s) during a hot-plug installation, be careful when removing or installing the hot-plug component(s).

Caution: To avoid injury do not contact moving fan blades. Your system is supplied with a guard over the fan, do not operate the system without the fan guard in place.

B.8 RACK MOUNT WARNINGS

Note: *The following installation guidelines are required by UL for maintaining safety compliance when installing your system into a rack.*

The equipment rack must be anchored to an unmovable support to prevent it from tipping when a server or piece of equipment is extended from it. The equipment rack must be installed according to the rack manufacturer's instructions.

Install equipment in the rack from the bottom up with the heaviest equipment at the bottom of the rack.

Extend only one piece of equipment from the rack at a time.

You are responsible for installing a main power disconnect for the entire rack unit. This main disconnect must be readily accessible, and it must be labeled as controlling power to the entire unit, not just to the server(s).

To avoid risk of potential electric shock, a proper safety ground must be implemented for the rack and each piece of equipment installed in it.

Elevated Operating Ambient- If installed in a closed or multi-unit rack assembly, the operating ambient temperature of the rack environment may be greater than room ambient. Therefore, consideration should be given to installing the equipment in an environment compatible with the maximum ambient temperature (T_{ma}) specified by the manufacturer.

Reduced Air Flow -Installation of the equipment in a rack should be such that the amount of air flow required for safe operation of the equipment is not compromised.

Mechanical Loading- Mounting of the equipment in the rack should be such that a hazardous condition is not achieved due to uneven mechanical loading.

Circuit Overloading- Consideration should be given to the connection of the equipment to the supply circuit and the effect that overloading of the circuits might have on overcurrent protection and supply wiring. Appropriate consideration of equipment nameplate ratings should be used when addressing this concern.

Reliable Earthing- Reliable earthing of rack-mounted equipment should be maintained.

Particular attention should be given to supply connections other than direct connections to the branch circuit (e.g. use of power strips).

B.9 ELECTROSTATIC DISCHARGE (ESD)

Caution: ESD can damage drives, boards, and other parts. We recommend that you perform all procedures at an ESD workstation. If one is not available, provide some ESD protection by wearing an antistatic wrist strap attached to chassis ground -- any unpainted metal surface -- on your server when handling parts.

Always handle boards carefully. They can be extremely sensitive to ESO. Hold boards only by their edges. After removing a board from its protective wrapper or from the server, place the board component side up on a grounded, static free surface. Use a conductive foam pad if available but not the board wrapper. Do not slide board over any surface.

B.10 OTHER HAZARDS

CALIFORNIA DEPARTMENT OF TOXIC SUBSTANCES CONTROL:

Perchlorate Material – special handling may apply. See **www.dtsc.ca.gov/perchlorate**.

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

NICKEL

Ni
NICKEL

NVIDIA Bezel. The bezel's decorative metal foam contains some nickel. The metal foam is not intended for direct and prolonged skin contact. Please use the handles to remove, attach or carry the bezel. 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.

Battery Replacement

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.

Cooling and Airflow

Caution: Carefully route cables as directed to minimize airflow blockage and cooling problems. For proper cooling and airflow, operate the system only with the chassis covers installed. Operating the system without the covers in place can damage system parts. To install the covers:

- Check first to make sure you have not left loose tools or parts inside the system.
- Check that cables, add-in cards, and other components are properly installed.
- Attach the covers to the chassis according to the product instructions.

The equipment is intended for installation only in a Server Room/ Computer Room where both these conditions apply:

- Access can only be gained by SERVICE PERSONS or by USERS who have been instructed about the reasons for the restrictions applied to the location and about any precautions that shall be taken; and
- Access is through the use of a TOOL or lock and key, or other means of security, and is controlled by the authority responsible for the location

APPENDIX C COMPLIANCE

The NVIDIA Luna Server is compliant with the regulations listed in this section.

C.1 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.

California Department of Toxic Substances Control: Perchlorate Material - special handling may apply. See www.dtsc.ca.gov/perchlorate.

C.2 UNITED STATES / CANADA

cTUVus Mark

TÜVRheinland
C
US

C.3 CANADA

Innovation, Science and Economic Development Canada (ISED)

CAN ICES-3(A)/NMB-3(A)

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

Cet appareil numerique de la class A respecte toutes les exigences du Reglement sur le materiel brouilleur du Canada.

C.4 CE

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.

This device bears the CE mark in accordance with Directive 2014/53/EU.

This device complies with the following Directives:

- EMC Directive A, I.T.E Equipment.
- Low Voltage Directive for electrical safety.
- RoHS Directive for hazardous substances.
- Energy-related Products Directive (ErP).

The full text of EU declaration of conformity is available at the following internet address: www.nvidia.com/support

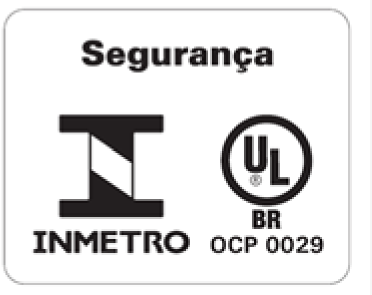
A copy of the Declaration of Conformity to the essential requirements may be obtained directly from NVIDIA GmbH (Bavaria Towers – Blue Tower, Einsteinstrasse 172, D-81677 Munich, Germany).

C.5 AUSTRALIA AND NEW ZEALAND

Australian Communications and Media Authority

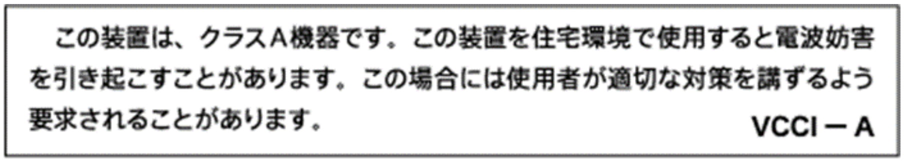
This product meets the applicable EMC requirements for Class A, I.T.E equipment

C.6 BRAZIL



C.7 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

This is a Class A product.

In a domestic environment this product may cause radio interference, in which case the user may be required to take corrective actions. VCCI-A

2008年、日本における製品含有表示方法、JISC0950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

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。

機器名称：サーバ

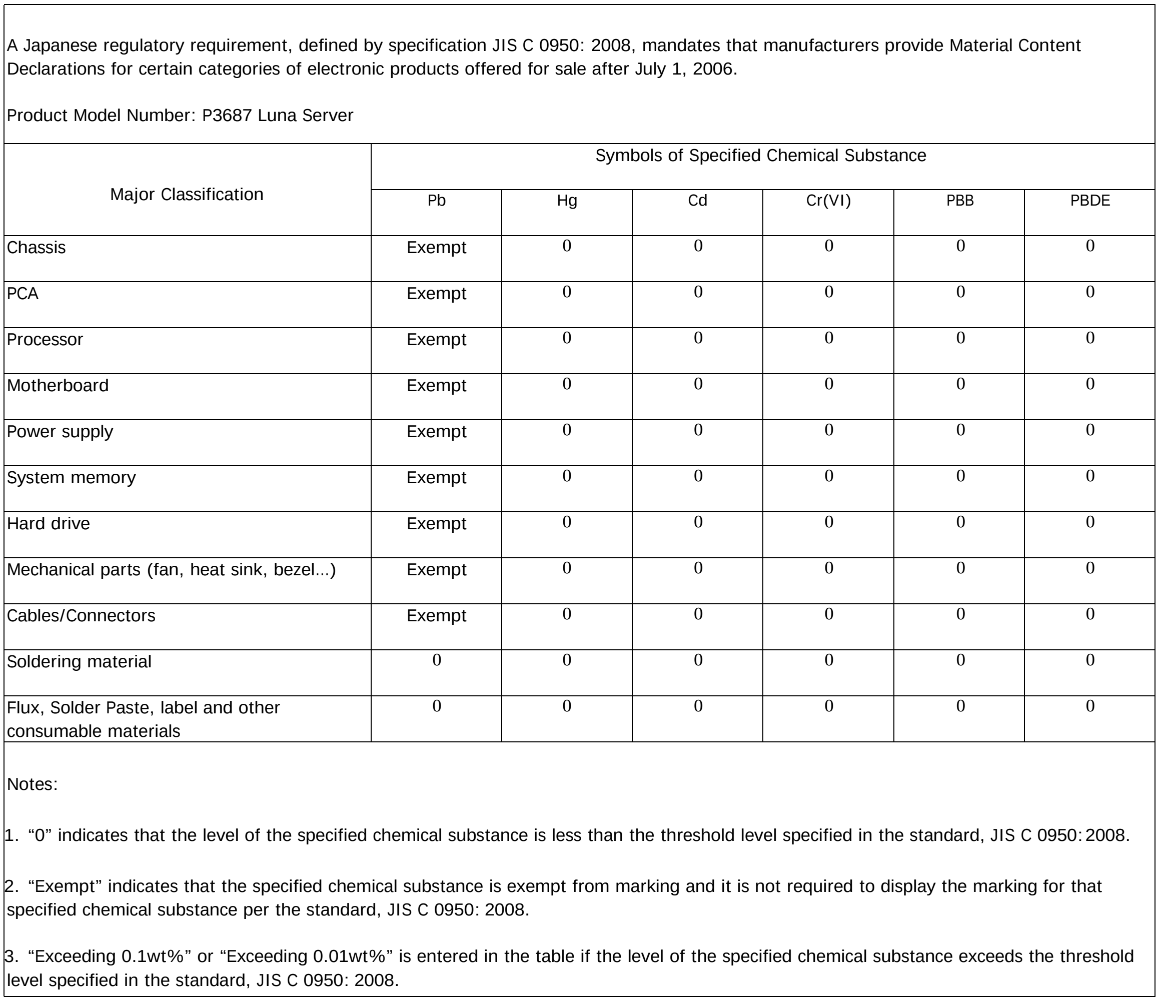
主な分類	特定化学物質記号					
	Pb	Hg	Cd	Cr(VI)	PBB	PBDE
筐体	除外項目	0	0	0	0	0
プリント基板	除外項目	0	0	0	0	0
プロセッサー	除外項目	0	0	0	0	0
マザーボード	除外項目	0	0	0	0	0
電源	除外項目	0	0	0	0	0
システムメモリ	除外項目	0	0	0	0	0
ハードディスクドライブ	除外項目	0	0	0	0	0
機械部品 (ファン、ヒートシンク、ベゼル ..)	除外項目	0	0	0	0	0
ケーブル / コネクター	除外項目	0	0	0	0	0
はんだ付け材料	0	0	0	0	0	0
フラックス、クリームはんだ、ラベル、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ていることを示します。



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.

Product Model Number: P3687 Luna Server

Major Classification	Symbols of Specified Chemical Substance					
	Pb	Hg	Cd	Cr(VI)	PBB	PBDE
Chassis	Exempt	0	0	0	0	0
PCA	Exempt	0	0	0	0	0
Processor	Exempt	0	0	0	0	0
Motherboard	Exempt	0	0	0	0	0
Power supply	Exempt	0	0	0	0	0
System memory	Exempt	0	0	0	0	0
Hard drive	Exempt	0	0	0	0	0
Mechanical parts (fan, heat sink, bezel...)	Exempt	0	0	0	0	0
Cables/Connectors	Exempt	0	0	0	0	0
Soldering material	0	0	0	0	0	0
Flux, Solder Paste, label and other consumable materials	0	0	0	0	0	0

Notes:

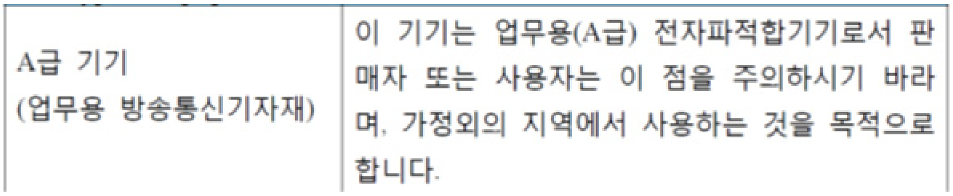
1. “0” indicates that the level of the specified chemical substance is less than the threshold level specified in the standard, JIS C 0950: 2008.

2. “Exempt” indicates that the specified chemical substance is exempt from marking and it is not required to display the marking for that specified chemical substance per the standard, JIS C 0950: 2008.

3. “Exceeding 0.1wt%” or “Exceeding 0.01wt%” is entered in the table if the level of the specified chemical substance exceeds the threshold level specified in the standard, JIS C 0950: 2008.

C.8 SOUTH KOREA

R-R-WT1-P3687



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

Class A Equipment (Industrial Broadcasting & Communication Equipment). This equipment Industrial (Class A) electromagnetic wave suitability equipment and seller or user should take notice of it, and this equipment is to be used in the places except for home.

Korea RoHS Material Content Declaration

:

확인 및 평가 양식은 제품에 포함 된 유해 물질의 허용 기준의 준수에 관한				
문 준비	상호 :	엔비디아홍콩홀딩즈 리미티드(영업소)	법인등록번호	110181-0036373
	대표자성명	카렌테레사번즈	사업자등록번호:	120-84-06711
	주소	서울특별시 강남구 영동대로 511, 2101호 (삼성동,		

제품 내용			
제품의 종류	해당없음	제품명(규격)	해당없음
세부모델명(번호):	해당없음	제품출시일	해당없음
제품의 중량	해당없음	제조, 수입업자	엔비디아

엔비디아의 그래픽 카드제품은 전기 전자제품 및 자동차의 자원순환에 관한 법률 시행령 제 11조 제 1항에 의거한 법 시행행규칙 제 3조에에따른 유해물질함유 기준을 확인 및 평가한 결과, 이를 준수하였음을 공표합니다.

구비서류 : 없음

작성방법

① 제품의 종류는 "전기.전자제품 및 자동차의 자원순환에관한 법률 시행령" 제 8조 제 1항 및 제 2항에 따른 품목별로 구분하여 기재합니다.

② 전기 전자 제품의 경우 모델명 (번호), 자동차의 경우, 제원관리번호를 기재합니다.

③ 해당제품의 제조업자 또는 수입업자를 기재합니다.

Confirmation and Evaluation Form Concerning the Adherence to Acceptable Standards of Hazardous Materials Contained in Products

Statement Prepared by	Company Name:	Nvidia HongKong Holding Ltd.Korea branch	Corporate Identification Number:	110181-0036373
	Name of Company Representative:	Karen Theresa Burns	Business Registration Number:	120-84-06711
	Address	2788 San Tomas Expressway, Santa Clara, CA 95051		

Product Information

Product Category:	N/A	Name of Product:	N/A
Detailed Product Model Name (Number):	N/A	Date of first market release:	N/A
Weight of Product:	N/A	Manufacturer and/or Importer:	NVIDIA Corporation

This for is publicly certify That NVIDIA Company has undergone the confirmation and evaluation procedures for the acceptable amounts of hazardous materials contained in graphic card according to the regulations stipulated in Article 3 of the 'Status on the Recycling of Electrical and Electronic Products, and Automobiles' and that company has graphic card adhered to the Enforcement Regulations of Article 11, Item 1 of the statute.

Attachment: None

*Preparing the Form

① Please indicate the product category according to the categories listed in Article 8, Items 1and 2 of the ' Enforcement Ordinance of the Statute on the Recycling of Electrical, Electronic and Automobile Materials'

② For electrical and electronic products, please indicate the Model Name (and number). For automobiles, please indicate the Vehicle Identification Number.

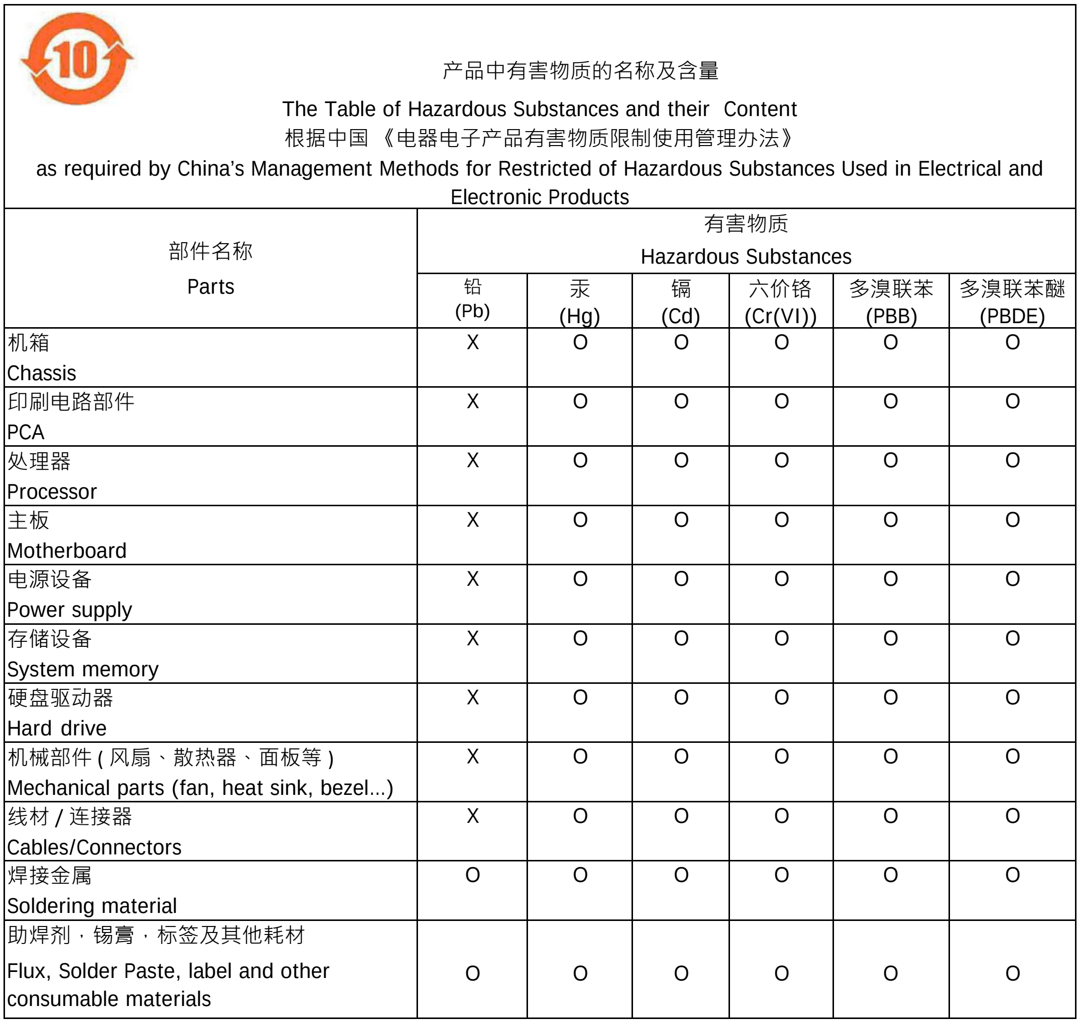
③ Please indicate the name of manufacturer and/or importer of the product.

C.9 CHINA

China Compulsory Certificate

No certification is needed for China. The NVIDIA DGX A100 is a server with power consumption greater than 1.3 kW.

China RoHS Material Content Declaration



产品中有害物质的名称及含量

The Table of Hazardous Substances and their Content

根据中国《电器电子产品有害物质限制使用管理办法》

as required by China's Management Methods for Restricted of Hazardous Substances Used in Electrical and Electronic Products

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	X	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 memory	X	O	O	O	O	O
硬盘驱动器 Hard drive	X	O	O	O	O	O
机械部件 (风扇、散热器、面板等) Mechanical parts (fan, heat sink, bezel...)	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.

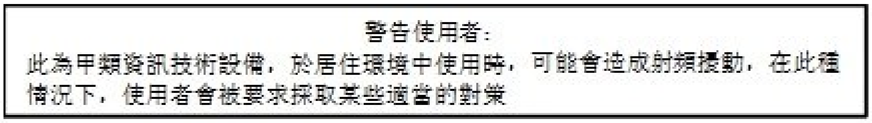
注：环保使用期限的参考标识取决于产品正常工作的温度和湿度等条件

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

C.10 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

報驗義務人：

香港商輝達香港控股有限公司台灣分公司 統一編號：80022300

臺北市內湖區基湖路**8**號.

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 Sustances Marking						
設備名稱：DGX 伺服器 Equipment Name: DGX Server						
單元 Parts	限用物質及其化學符號 Restricted substances and its chemical symbols					
	铅 (Pb)	汞 (Hg)	镉 (Cd)	六價铬 (Cr(VI))	多溴聯苯 (PBB)	多溴二苯醚 (PBDE)
機箱 Chassis	-	O	O	O	O	O
印刷電路部件 PCA	-	O	O	O	O	O
處理器 Processor	-	O	O	O	O	O
主板 Motherboard	-	O	O	O	O	O
電源設備 Power supply	-	O	O	O	O	O
存儲設備 System memory	-	O	O	O	O	O
硬盤驅動器 Hard drive	-	O	O	O	O	O
機械部件 (風扇、散熱器、面板等) Mechanical parts (fan, heat sink, bezel…)	-	O	O	O	O	O
線材/連接器 Cables/Connectors	-	O	O	O	O	O
焊接金屬 Soldering material	O	O	O	O	O	O
助焊劑，錫膏，標籤及其他耗材 Flux, Solder Paste, label and other consumable materials	O	O	O	O	O	O

備考1：O：系指該限用物質未超出百分比含量基準值
Note 1： O：indicates that the percentage content of the restricted substance does not exceed the percentage of reference value of presence.
備考2： -：系指該项限用物質为排外项目。
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决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.

C.11 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 (ТР ТС 004/2011)

ТЕХНИЧЕСКИЙ РЕГЛАМЕНТ ТАМОЖЕННОГО СОЮЗА Электромагнитная совместимость технических средств (ТР ТС 020/2011)

Технический регламент Евразийского экономического союза "Об ограничении применения опасных веществ в изделиях электротехники и радиоэлектроники" (ТР ЕАЭС 037/2016)

Federal Agency of communication (FAC)

This device complies with the rules set forth by Federal Agency of Communications and the Ministry of Communications and Mass Media

Federal Security Service notification has been filed.

Notice

THE INFORMATION IN THIS GUIDE AND ALL OTHER INFORMATION CONTAINED IN NVIDIA DOCUMENTATION REFERENCED IN THIS GUIDE IS PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE INFORMATION FOR THE PRODUCT, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the product described in this guide shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

THE NVIDIA PRODUCT DESCRIBED IN THIS GUIDE IS NOT FAULT TOLERANT AND IS NOT DESIGNED, MANUFACTURED OR INTENDED FOR USE IN CONNECTION WITH THE DESIGN, CONSTRUCTION, MAINTENANCE, AND/OR OPERATION OF ANY SYSTEM WHERE THE USE OR A FAILURE OF SUCH SYSTEM COULD RESULT IN A SITUATION THAT THREATENS THE SAFETY OF HUMAN LIFE OR SEVERE PHYSICAL HARM OR PROPERTY DAMAGE (INCLUDING, FOR EXAMPLE, USE IN CONNECTION WITH ANY NUCLEAR, AVIONICS, LIFE SUPPORT OR OTHER LIFE CRITICAL APPLICATION). NVIDIA EXPRESSLY DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY OF FITNESS FOR SUCH HIGH RISK USES. NVIDIA SHALL NOT BE LIABLE TO CUSTOMER OR ANY THIRD PARTY, IN WHOLE OR IN PART, FOR ANY CLAIMS OR DAMAGES ARISING FROM SUCH HIGH RISK USES.

NVIDIA makes no representation or warranty that the product described in this guide will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this guide. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this guide, or (ii) customer product designs.

Other than the right for customer to use the information in this guide with the product, no other license, either expressed or implied, is hereby granted by NVIDIA under this guide. Reproduction of information in this guide is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

Trademarks

NVIDIA, the NVIDIA logo, and DGX are trademarks and/or registered trademarks of NVIDIA Corporation in the Unites States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright