



NVIDIA DGX-1

User Guide

Table of Contents

Chapter 1. Introduction to the NVIDIA DGX-1 Deep Learning System.....	1
1.1. Using the DGX-1: Overview.....	1
1.2. Hardware Specifications.....	1
1.2.1. Components.....	1
1.2.2. Mechanical.....	2
1.2.3. Environmental.....	3
1.2.4. Power Requirements.....	3
1.2.5. Connections and Controls.....	3
1.2.6. Rear Panel Power Controls.....	4
1.2.7. LAN LEDs.....	5
1.2.8. IPMI Port LEDs.....	5
1.2.9. Hard Disk Indicators.....	6
1.2.10. Power Supply Unit (PSU) LED.....	7
Chapter 2. Installation and Setup.....	8
2.1. Registering Your DGX-1.....	8
2.2. Choosing a Setup Location / Site Preparation.....	8
2.3. Unpacking the DGX-1.....	11
2.4. What's In the Box.....	12
2.5. Installing the DGX-1 Into a Rack.....	12
2.5.1. Installing the Rails.....	13
2.5.2. Mounting the DGX-1.....	13
2.6. Attaching the Bezel.....	14
2.7. Connecting the Power Cables.....	15
2.8. Connecting the Network Cables.....	16
2.9. Setting Up the DGX-1.....	17
2.10. Post Setup Instructions for DGX OS Server Software Version 2.x and Earlier.....	19
2.11. Updating the DGX-1 Software.....	21
2.12. Managing CPU Mitigations.....	22
2.12.1. Determining the CPU Mitigation State of the DGX System.....	22
2.12.2. Disabling CPU Mitigations.....	22
2.12.3. Re-enabling CPU Mitigations.....	23
Chapter 3. Preparing for Using Docker Containers.....	24
3.1. Installing Docker and the Docker Engine Utility for NVIDIA GPUs on DGX OS Server Software 2.x or Earlier.....	24
3.2. Configuring Docker IP Addresses.....	25

3.2.1. Configuring Docker IP Addresses for DGX OS Server Software Version 2.x and Earlier.....	26
3.2.2. Configuring Docker IP Addresses for DGX OS Server Software Version 3.1.1 and Later.....	26
3.3. Letting Users Issue Docker Commands.....	27
3.3.1. Checking if a User is in the Docker Group.....	27
3.3.2. Creating a User.....	28
3.3.3. Adding a User to the Docker Group.....	28
3.4. Enabling GPU Support for NGC Containers.....	28
3.5. Configuring a System Proxy.....	30
3.6. Configuring NFS Mount and Cache.....	30
Chapter 4. Configuring and Managing the DGX-1.....	32
4.1. Using the BMC.....	32
4.1.1. Creating a Unique BMC Password for Remote Access.....	33
4.1.2. Viewing System Information.....	34
4.1.3. Submitting BMC Log Files.....	34
4.1.4. Determining Total Power Consumption.....	34
4.1.5. Accessing the DGX-1 Console.....	35
4.1.6. Powering Off / Power Cycling the System Remotely.....	35
4.1.6.1. From the DGX-1 Console Window.....	35
4.1.6.2. From the BMC UI.....	35
4.1.7. Setting Up Active Directory.....	36
4.2. Configuring a Static IP Address for the BMC.....	38
4.2.1. Configuring a BMC Static IP Address Using ipmitool.....	38
4.2.2. Configuring a BMC Static IP Address Using the System BIOS.....	39
4.2.3. Configuring a BMC Static IP Address Using the BMC Dashboard.....	43
4.3. Configuring Static IP Addresses for the Network Ports.....	44
4.4. Obtaining MAC Addresses.....	45
4.5. Resetting GPUs in the DGX-1.....	48
4.6. Changing the Mellanox Card Port Type.....	48
4.6.1. Switching the Port from InfiniBand to Ethernet.....	50
4.6.2. Switching the Port from Ethernet to InfiniBand.....	51
4.7. Enabling USB 3.0.....	51
Chapter 5. Security.....	56
Chapter 6. Maintaining and Servicing the NVIDIA DGX-1.....	57
6.1. Problem Resolution and Customer Care.....	57
6.2. Restoring the DGX-1 Software Image.....	57
6.2.1. Obtaining the DGX-1 Software ISO Image and Checksum File.....	58

6.2.2. Re-Imaging the System Remotely.....	58
6.2.3. Creating a Bootable Installation Medium.....	61
6.2.3.1. Creating a Bootable USB Flash Drive by Using the dd Command.....	62
6.2.3.2. Creating a Bootable USB Flash Drive by Using Akeo Rufus.....	63
6.2.4. Re-Imaging the System From a USB Flash Drive.....	64
6.2.5. Retaining the RAID Partition While Installing the OS.....	65
6.3. Updating the System BIOS.....	66
6.4. Updating the BMC.....	68
6.5. Updating Component Firmware Using the Firmware Update Container.....	70
6.5.1. Obtaining the Firmware Update Container.....	71
6.5.2. Querying the Firmware Manifest.....	72
6.5.3. Querying the Currently Installed Firmware Versions.....	72
6.5.4. Updating the Firmware.....	73
6.5.4.1. Updating All Firmware Components.....	73
6.5.4.2. Updating Specific Firmware Components.....	74
6.5.4.3. Additional Options.....	74
6.5.4.4. Command Summary.....	75
6.5.5. Removing the Container.....	75
6.5.6. Using the .run File.....	75
6.6. Replacing the System and Components.....	76
6.6.1. Replacing the System.....	76
6.6.2. Replacing an SSD.....	76
6.6.3. Recreating the Virtual Drives.....	78
6.6.3.1. Access the BIOS Setup Utility.....	78
6.6.3.2. Clear the Drive Group Configuration.....	81
6.6.3.3. Recreate the OS Virtual Drive.....	85
6.6.3.4. Recreate the RAID0 Virtual Drive.....	93
6.6.4. Recreating the RAID 0 Array.....	105
6.6.5. Replacing the Power Supplies.....	106
6.6.6. Replacing the Fan Module.....	107
6.6.7. Replacing the Battery.....	108
6.6.8. Replacing the DIMMs.....	112
6.6.9. Installing/Replacing the 10GbE Mezzanine SPF+ NIC.....	116
6.6.10. Replacing the InfiniBand Cards.....	119
6.6.11. Setting Up the InfiniBand Cards.....	123
6.7. Secure Data Deletion of the SSDs.....	127
6.7.1. Method A: Using the MegaRAID Controller.....	127
6.7.2. Method B: Using the StorCLI Secure Erase Commands.....	127

Chapter 7. Installing Software on Air-Gapped NVIDIA DGX-1 Systems.....	129
7.1. Installing NVIDIA DGX-1 Software.....	129
7.1.1. Re-Imaging the System.....	129
7.1.2. Creating a Local Mirror of the NVIDIA and Canonical Repositories.....	130
7.1.2.1. Creating the Mirror in a DGX OS 4 System.....	130
7.1.2.2. Configuring the Target Air-Gapped DGX OS 4 System.....	132
7.2. Installing Docker Containers.....	134
Chapter 8. Customer Support for the NVIDIA DGX-1.....	135
Chapter 9. King Slide - AH61-500 Instructions.....	136
Chapter 10. Safety.....	138
10.1. Safety Warnings and Cautions.....	138
10.2. Intended Application Uses.....	139
10.3. Site Selection.....	139
10.4. Equipment Handling Practices.....	139
10.5. Electrical Precautions.....	140
10.6. System Access Warnings.....	141
10.7. Rack Mount Warnings.....	141
10.8. Electrostatic Discharge.....	142
10.9. Other Hazards.....	142
Chapter 11. Compliance.....	144
11.1. United States.....	144
11.2. United States / Canada.....	144
11.3. Canada.....	144
11.4. CE.....	145
11.5. Japan.....	145
11.6. Australia.....	146
11.7. China.....	146
11.8. Israel.....	148
11.9. South Korea.....	148
11.10. India.....	149

List of Figures

Figure 1. Example label for model CX455A	122
Figure 2. Example label for model CX555A	123

List of Tables

Table 1. SR Modules	10
Table 2. LR Modules	10

Chapter 1. Introduction to the NVIDIA DGX-1 Deep Learning System

The NVIDIA® DGX-1™ Deep Learning System is the world's first purpose-built system for deep learning with fully integrated hardware and software that can be deployed quickly and easily.



1.1. Using the DGX-1: Overview

The NVIDIA DGX-1 comes with a base operating system consisting of an Ubuntu OS, Docker, Docker Engine Utility for NVIDIA GPUs, and NVIDIA drivers. This system is designed to run a number of NVIDIA-optimized deep learning framework applications packaged in Docker containers. You can use your own scheduling and management software to run jobs, and also build and run your own applications on the DGX-1.

1.2. Hardware Specifications

1.2.1. Components

Component	Qty	Description
Base Server	1	Dual Intel® Xeon® CPU motherboard with x2 9.6 GT/s QPI, 8 Channel with 2 DPC DDR4, Intel® C610 Chipset, AST2400 BMC

Component	Qty	Description
	1	GPU Baseboard supporting 8 SXM2 modules (Cube Mesh) and 4 PCIe x16 slots for InfiniBand/Ethernet NICs
	1	Chassis with 3+1 1600W Power supply and support for up to five 2.5 inch drives
	1	10/100BASE-T IPMI Port
	1	RS232 Serial Port
	2	USB 3.0 Ports (set as USB 2.0 by default. To enable USB 3.0, see Enabling USB 3.0 for instructions.)
Power Supply	4	1600 W each.
CPU	2	Intel® Xeon® E5-2698 v4, 20-core, 2.2GHz, 135W
GPU	8	<p>(Pascal) Tesla P100, featuring</p> <ul style="list-style-type: none"> ▶ 170 teraflops, FP16 ▶ 16 GB memory per GPU ▶ 28,672 NVIDIA CUDA® Cores <p>(Volta) Tesla V100, featuring</p> <ul style="list-style-type: none"> ▶ 1 petaflop, Mixed Precision ▶ 16 GB memory per GPU ▶ 40,960 NVIDIA CUDA® Cores ▶ 5120 NVIDIA Tensor Cores <p>(Volta 32 GB) Tesla V100, featuring</p> <ul style="list-style-type: none"> ▶ 1 petaflop, Mixed Precision ▶ 32 GB memory per GPU ▶ 40,960 NVIDIA CUDA® Cores ▶ 5120 NVIDIA Tensor Cores
System Memory	16	32 GB DDR4 LRDIMM (512 GB total)
SAS Raid Controller	1	8 port LSI SAS 3108 RAID Mezzanine
Storage (RAID 0) (Data)	4	1.92 TB, 6 Gb/s, SATA 3.0 SSD
Storage (OS)	1	480 GB, 6 Gb/s, SATA 3.0 SSD
10 GbE NIC	1	Dual port, 10GBASE-T, network adapter Mezzanine
InfiniBand EDR/100GbE NIC	4	<p>Single port, x16 PCIe, Mellanox ConnectX-4 VPI MCX455A-ECAT</p> <p>or</p> <p>Single port, x16 PCIe, Mellanox ConnectX-5 VPI MCX555A-ECAT</p>

1.2.2. Mechanical

Feature	Description
Form Factor	3U Rackmount

Feature	Description
Height	5.16" (13.1 cm)
Width	17.5" (44.4 cm)
Depth	34.1" (86.6 cm)
Gross Weight	134 lbs (61 kg)

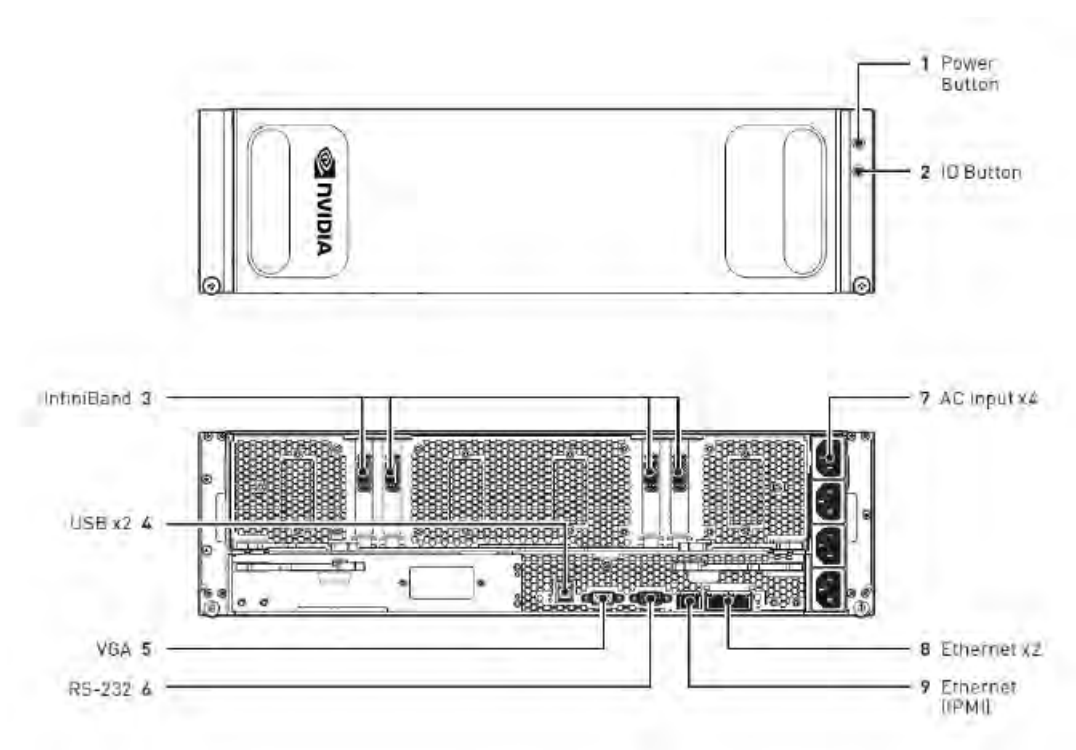
1.2.3. Environmental

Feature	Description
Operating Temperature	5° C to 35° C (41° F to 95° F)
Relative Humidity	20% to 85% noncondensing
Airflow	340 CFM @ 35° C
Heat Output	12000 BTU/hr

1.2.4. Power Requirements

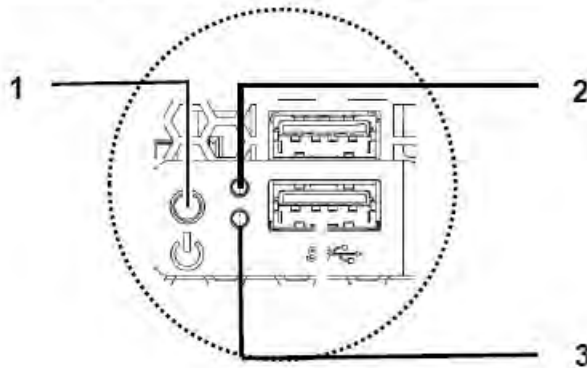
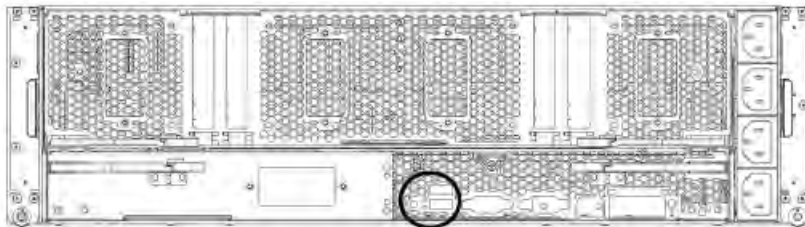
Input		Specification for Each Power Supply	Comments
200-240 V (ac)	3500 W max.	1600 W @ 200-240 V, 8 A, 50-60 Hz	The DGX-1 contains four load-balancing power supplies, with 3+1 redundancy.

1.2.5. Connections and Controls



ID	Type	Qty	Description
1	Power button	1	Press to turn the DGX-1 on or off. Blue: System power on Off: System power off Amber (blinking): DC power Off, fault reported in BMC SEL Amber and blue (blinking): DC power On and fault reported in BMC SEL
2	ID button	1	Press to cause an LED on the back of the unit to flash as an identifier during servicing.
3	InfiniBand/ Ethernet (QSFP28)	4	EDR IB/100GbE
4	USB	2	USB 3.0 ports are available to connect a keyboard.
5	VGA	1	The VGA port connects to a VGA capable monitor for local viewing of the DGX-1 setup console or base OS.
6	DB9	1	RS232 serial port for internal debugging
7	AC input	4	Power supply inputs
8	Ethernet (RJ45)	2	10GBASE-T dual port network adapter Mezzanine
9	IPMI (RJ45)	1	10/100BASE-T Intelligent Platform Management Interface (IPMI) port

1.2.6. Rear Panel Power Controls



ID	Type	Qty	Description
1	Power button	1	Press and immediately release the power button for a graceful shutdown of the host OS.

ID	Type	Qty	Description
			Press and hold the power button for at least four seconds to shut down the system immediately. The BMC remains live.
2	Power LED	1	Off: Power off Blue (steady): Power on Blue (blinking): BMC reports system health fault.
3	Main Board Status LED	1	Off: Normal Amber (blinking): BMC reports system health fault.

1.2.7. LAN LEDs

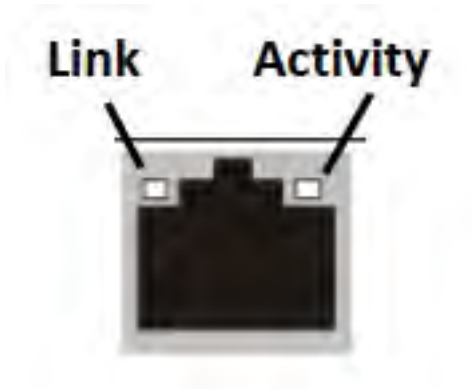
LEDs next to each Ethernet port indicate the connection status as described in the table below:



LED	Status	Description
1 (Port 1 Link/Activity)	Amber (steady)	LAN link
	Amber (blinking)	LAN access (off when there is traffic)
	Off	Disconnected
2 (Port 1 Speed)	Green	10 Gb/s
	Amber	1 Gb/s
	Off	100 Mb/s
3 (Port 0 Link/Activity)	Amber (steady)	LAN link
	Amber (blinking)	LAN access (off when there is traffic)
	Off	Disconnected
4 (Port 0 Speed)	Green	10 Gb/s
	Amber	1 Gb/s
	Off	100 Mb/s

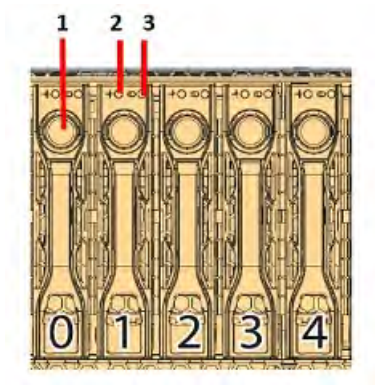
1.2.8. IPMI Port LEDs



LEDs on the IPMI port indicate the connection status as described in the table below:



Link	Activity	Description
Off	Off	Unplugged
Green (steady)	Green (blinking)	100M active link
Off	Green (blinking)	10M active link

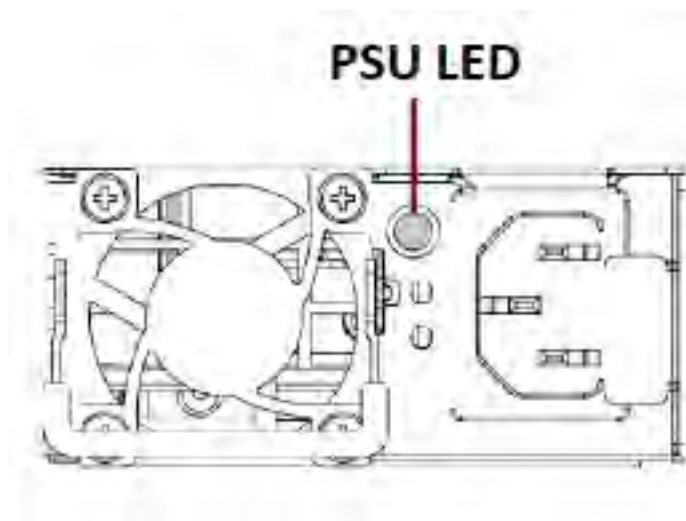
1.2.9. Hard Disk Indicators



ID	Feature	Description
1	Button and release lever for removing the HDD	
2	 HDD present LED	Blue (Steady): Drive present Blue (Blinking twice/sec): Identification (such as when initializing or locating through the SBIOS) Blue (Blinking once/sec): Rebuilding (such as when creating a RAID array) Amber (Steady): Warning/failure Off: Slot empty
3	 HDD activity LED	Blue: Access

1.2.10. Power Supply Unit (PSU) LED

The PSU LED indicates the operation status of the PSU as described in the table below:



Activity	Description
Green	Normal operation
Amber (blinking)	Power off; Fault
Green (blinking)	Power on; Standby mode

Chapter 2. Installation and Setup

This chapter provides the basic instructions for installing and setting up the NVIDIA DGX-1.

2.1. Registering Your DGX-1

To obtain support for your DGX-1 system, follow the instructions for registration in the Entitlement Certification email that was sent as part of the purchase.

Registration allows you to access the NVIDIA Enterprise Support Portal, obtain technical support, get software updates, and set up an NGC for DGX systems account. If you did not receive the information, open a case with the NVIDIA Enterprise Support Team at <https://www.nvidia.com/en-us/support/enterprise/>.

Refer to the [Customer Support](#) chapter for customer support contact information.

2.2. Choosing a Setup Location / Site Preparation

Decide on a suitable location for setting up and operating the DGX-1. The location should be clean, dust-free, and well ventilated.

General Conditions

- ▶ Prepare a sufficiently wide aisle to accommodate the unboxed chassis (chassis dimensions - 5.16"H x 17.5"W x 34.1"D).
- ▶ The rack must accommodate a 134 lb, 3U rack mount system (chassis dimensions - 5.16"H x 17.5"W x 34.1"D).
- ▶ The rack must have square mounting holes.
- ▶ Leave enough clearance in front of the rack (36" (91.4 cm)) to enable you to install the unit into the rack.
- ▶ Leave approximately 30" (76.2cm) of clearance in the back of the rack to allow for sufficient airflow and ease in servicing.

- ▶ Always make sure the rack is secured and stable before adding or removing the appliance or any other component.
- ▶ Prepare adequate sound-proofing: The equipment fans can generate 72-100 dBA.

Environmental Conditions

- ▶ Operating environment
 - ▶ Temperature: 5 # C to 35 # C (41 # F to 95 # F)
 - ▶ Relative humidity: 20% to 85% noncondensing
- ▶ Air flow
 - ▶ The chassis fans can produce a maximum of 340 CFM of air flow.
 - ▶ Do not block the ventilation areas at the front and rear of the chassis.
 - ▶ Minimize any restrictions on air flow around the chassis.

Connections

- ▶ Power:
 - ▶ The DGX-1 is powered through four 1600W power supply units, each rated at 200-240VAC, 8A, 50/60 Hz. Total system power requirement: 3500W
 - ▶ C13/C14 cables provided for each power supply to connect to a compatible PDU.

IMPORTANT: Do not use the provided cables with any other product or for any other purpose.
- ▶ Network: Dual 10GBASE-T RJ45 connection

Use industry standard CAT6 Ethernet cables for connecting to the network ports. (Cables not included.)
- ▶ IPMI: 10/100BASE-T RJ45 connection

Use industry standard CAT6 Ethernet cables for connecting to the network ports. (Cables not included.)
- ▶ InfiniBand: Qty 4 - QSFP28 ports, InfiniBand and Ethernet compliant

Use Mellanox-compliant InfiniBand cables for connecting to the InfiniBand ports. (Cables not included.)

The DGX-1 System is not shipped with network cables. For a list of cables compatible with the Mellanox ConnectX-4 or ConnectX-5 VPI cards installed in the DGX-1 system, visit the Mellanox [ConnectX-4](#) or [ConnectX-5](#) Firmware Download page, select the appropriate FW version, OPN (model), and PSID, and then select Release Notes from the Documentation column.

If you have replaced the Dual port, 10GBASE-T, network adapter Mezzanine with the 10GbE SPF+ NIC,

- ▶ All passive and active limiting direct attach cables that comply with SFF-8431 v4.1 and SFF-8472 v10.4 specifications are supported.
- ▶ The following Intel optics and/or direct attach cables are supported:

Table 1. SR Modules

Supplier	Type	Part Numbers
Intel	DUAL RATE 1G/10G SFP+ SR (bailed)	AFBR-703SDZ-IN2
Intel	DUAL RATE 1G/10G SFP+ SR (bailed)	FTLX8571D3BCV-IT
Intel	DUAL RATE 1G/10G SFP+ SR (bailed)	AFBR-703SDDZ-IN1

Table 2. LR Modules

Supplier	Type	Part Numbers
Intel	DUAL RATE 1G/10G SFP+ LR (bailed)	FTLX1471D3BCV-IT
Intel	DUAL RATE 1G/10G SFP+ LR (bailed)	AFCT-701SDZ-IN2
Intel	DUAL RATE 1G/10G SFP+ LR (bailed)	AFCT-701SDDZ-IN1

- ▶ The following lists the 3rd party SFP+ modules and direct attach cables that have received some testing. Not all modules are applicable to all devices.

Supplier	Type	Part Numbers
Finisar	SFP+ SR bailed, 10g single rate	FTLX8571D3BCL
Avago	SFP+ SR bailed, 10g single rate	AFBR-700SDZ
Finisar	SFP+ LR bailed, 10g single rate	FTLX1471D3BCL
Finisar	DUAL RATE 1G/10G SFP+ SR (No Bail)	FTLX8571D3QCV-IT
Avago	DUAL RATE 1G/10G SFP+ SR (No Bail)	AFBR-703SDZ-IN1
Finisar	DUAL RATE 1G/10G SFP+ LR (No Bail)	FTLX1471D3QCV-IT
Avago	DUAL RATE 1G/10G SFP+ LR (No Bail)	AFCT-701SDZ-IN1

Preparing for Network Access

- ▶ The IPMI port and Ethernet ports can be connected to your local LAN.
These ports are configured for DHCP by default.
 - ▶ To use DHCP, connect the port to a local DHCP server which should provide an IP address and assign a DNS configuration to the DGX-1.
 - ▶ If DHCP is not available, then you will need to set up a static IP for each Ethernet port.
- ▶ 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-1 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
- ▶ Make sure your network can connect to the following:
 - ▶ <http://us.archive.ubuntu.com/ubuntu/>
 - ▶ <http://security.ubuntu.com/ubuntu>
 - ▶ <http://international.download.nvidia.com/dgx1/repos/> (Base OS Software 2.x or earlier; to be accessed using apt-get, not a browser)
 - ▶ <http://international.download.nvidia.com/dgx/repos/> (Base OS Software 3.1 or later; to be accessed using apt-get, not a browser)
 - ▶ <https://apt.dockerproject.org/repo>
 - ▶ <https://download.docker.com/linux/ubuntu>
 - ▶ <https://nvcr.io>

Test using

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

The output should include

```
Connecting to nvcr.io (nvcr.io)|52.8.131.152|:443...connected.  
HTTP request sent, awaiting response... 401 Unauthorized
```

If access to those URLs requires use of a proxy, refer to [Setting Up a System Proxy](#) for setup instructions.

2.3. Unpacking the DGX-1

1. Remove the shrink wrap.
2. Collapse the yellow "Do not stack" cone, if included.
3. Open the main DGX-1 box, then remove the accessory and rail kit boxes.



CAUTION: At least four people, or a mechanical assist, are required to remove the DGX-1 from the box. To reduce the risk of personal injury or damage to the equipment, always observe local occupational health and safety requirements and guidelines for material handling.



DO NOT use the handles at the front of the DGX-1 to lift the unit. The handles are designed for sliding the unit out of a rack, and not for carrying the full weight of the DGX-1.

4. Remove the protective plastic sheet from the top of the DGX-1.
5. Preserve and retain packaging.
6. Be sure to inspect each piece of equipment shipped in the packing box. If anything is missing or damaged, contact your supplier.

2.4. What's In the Box

The NVIDIA DGX-1 shipping box includes the following:

- ▶ NVIDIA DGX-1
 - ▶ Bezel
 - ▶ Rail hardware kit
 - ▶ Accessory Box
 - ▶ AC Power Cables (qty 4 – IEC 60320 C13/14, 3.3-ft./1-meter, compatible with data center PDUs)
- IMPORTANT: Do not use the provided cables with any other product or for any other purpose.**
- ▶ Hard disk bay screws
 - ▶ Toxic Substance Notice & Safety Instructions
 - ▶ Quick Start Guide
 - ▶ DVD containing source files for open source software



Note: The four power cables included in the box are not optional. All power cables are necessary and must be plugged into individual 10 A capable sockets for optimal DGX-1 operation. Failure to do so can result in a reduction in power redundancy, a reduction in performance, or a complete system failure.

2.5. Installing the DGX-1 Into a Rack



CAUTION: To prevent bodily injury when mounting or servicing the DGX-1 in a rack, you must take special precautions to ensure that the system remains stable. The following guidelines are provided to ensure your safety.

- The DGX-1 should be mounted at the bottom of the rack if it is the only unit in the rack.
- When mounting the DGX-1 in a partially filled rack, load the rack from the bottom to the top with the heaviest component at the bottom of the rack.
- If the rack is provided with stabilizing devices, install the stabilizers before mounting or servicing the DGX-1 in the rack.
- The DGX-1 weighs approximately 134 lbs, so an equipment lift is required to safely lift the unit and then accurately align the chassis rails with the rack rails.
- **DO NOT** use the handles at the front of the DGX-1 to lift the unit. The handles are designed for sliding the unit out of a rack, and not for carrying the full weight of the DGX-1.

2.5.1. Installing the Rails



Note: The rail assemblies shipped with the appliance fit into a standard 19" rack between 26-inches and 33.5-inches deep (66 cm to 85 cm). The outer rail is adjustable from approximately 23.5" to 34" (59.7 cm to 86.4 cm)

Refer to the instructions in the rail packaging for details on installing the rails onto the rack and chassis. The instructions are also provide in the chapter [King Slide - AH61-500](#) in this document.

The following are supplemental instructions:

1. Use a Phillips screwdriver to assist in mounting the rails to the rack.
2. If necessary, detach the inner rails from the outer slide rails.
3. Follow any designations on the inner rail (or its outer rail mate) to determine the proper orientation and positioning to connect to the chassis, then secure to the chassis.



Important: Make sure that the reinforced hole at the front end of the rail is positioned on the bottom side of the rail, and that it aligns with the thumbscrew on the front of the DGX-1. If the hole is positioned on the top side, then the rail is on the wrong side of the DGX-1 and the DGX-1 will not fit properly in the rack.

4. Follow any designations on the outer slide rail to determine front/back and left-side/right-side positioning against the rack.
5. Secure the back of one of the slide rails to the rack, then extend the rail until it fits securely to the front of the rack.
6. Secure the slide rail to the front of the rack.
7. Repeat steps 4-6 for the other slide rail.

2.5.2. Mounting the DGX-1



Note: CAUTION: Stability hazard — The rack stabilizing mechanism must be in place, or the rack must be bolted to the floor before you slide the DGX-1 out for servicing. Failure to stabilize the rack can cause the rack to tip over.

1. Confirm that the DGX-1 has the inner rails attached and that you have already mounted the outer rails into the rack.
2. With the front of the unit facing away from the rack, use an equipment lift to assist in sliding the unit into the rack as follows:

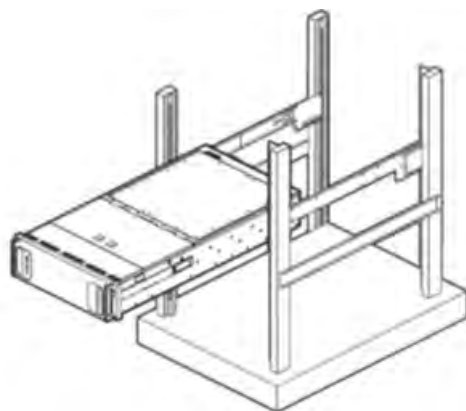


CAUTION: The DGX-1 weighs approximately 134 lbs, so an equipment lift is required to safely lift the unit and then accurately align the chassis rails with the rack rails.

- a). Align the inner chassis rails with the front of the outer rack rails.

- b). Slide the inner rails into the outer rails, keeping the pressure even on both sides (you may have to depress the locking tabs when inserting).

When the DGX-1 has been pushed completely into the rack, you should hear the locking tabs "click" into the locked position.



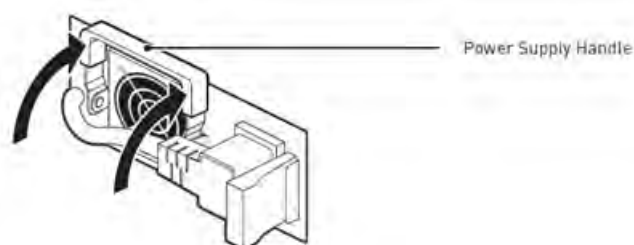
3. Lock the unit in place using the thumb screws located on the front of the unit.



2.6. Attaching the Bezel

The bezel is designed to attach easily to the front of the DGX-1.

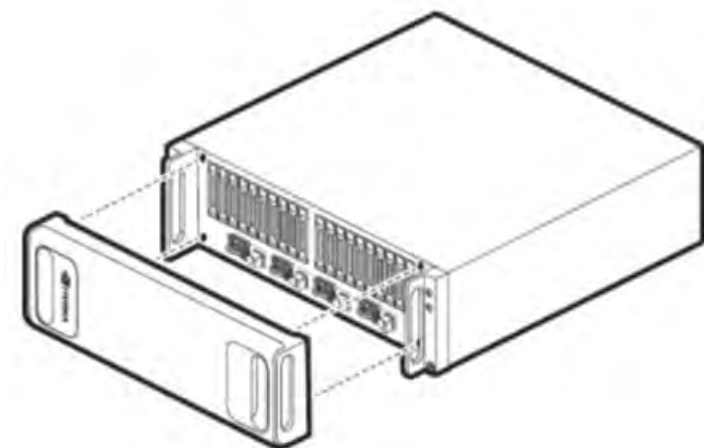
1. Prepare the DGX-1 by making sure that the power supply handles (located at the power supply fans) are flipped up.



2. Move any other obstructions, such as cable ties, away from the outer edge of the DGX-1.
3. With the bezel positioned so that the NVIDIA logo is visible from the front and is on the left hand side, line up the pins near the corners of the DGX-1 with the holes in back of the bezel, then gently press the bezel against the DGX-1.



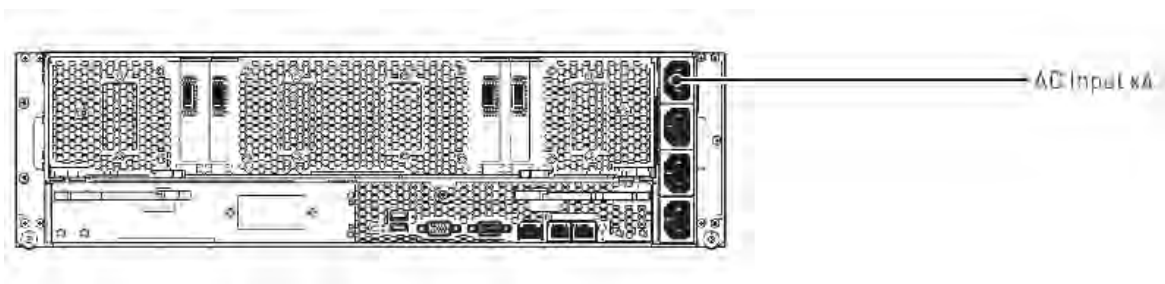
Note: CAUTION: Be careful not to accidentally press the power button that is on the right edge of the DGX-1 when removing or installing the bezel.



The bezel is held in place magnetically .

2.7. Connecting the Power Cables

1. Open the accessory box and remove the four C13/C14 power cables.
2. Use the cables to connect each of the four plugs at the right-rear of the DGX-1 to a PDU.



Note: NVIDIA recommends connecting power through a smart PDU which provides the ability to power-cycle the system remotely.

- a). Secure each cable to the DGX-1, using the power cable retention clips attached to the power plugs.

- b). Connect each cable to the PDU.

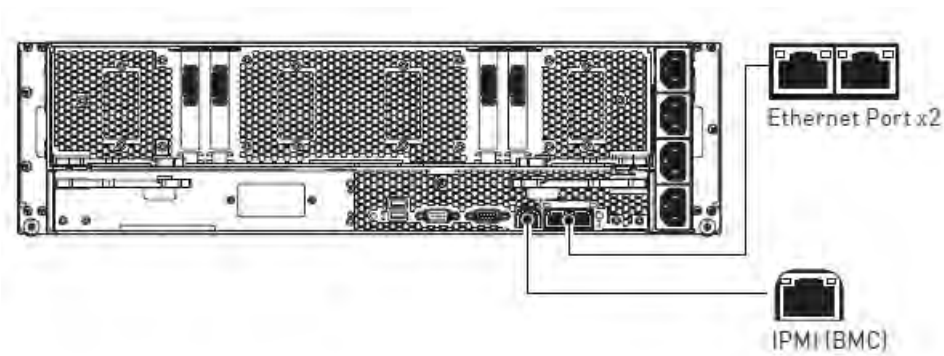
Ensure that the cables are distributed over at least two circuits and, if using 3-phase PDUs, they are balanced across all phases as much as possible. Ideally, each cable should connect to a different PDU.

- c). Verify that each cable is firmly inserted into the PDU.

There is usually a click to indicate full insertion.

2.8. Connecting the Network Cables

1. Using an Ethernet cable, connect one of the dual Ethernet ports (em1 or em2) to your LAN for internet access to the NVIDIA DGX website, remote access to launched application containers on the DGX-1, or to connect to the DGX-1 using SSH.



The left-side/right-side ethernet port designation depends on the Base OS software version installed on the DGX-1 as listed in the table below.

Ethernet Port Position	Port Designation: Base OS Software 2.x and earlier	Port Designation: Base OS Software 3.x and later
Right Side	em1	enp1s0f0
Left Side	em2	enp1s0f1



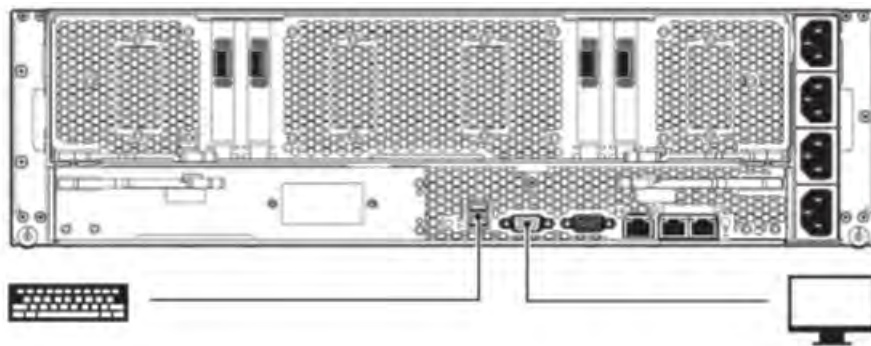
Note: NVIDIA recommends connecting only one of the Ethernet ports to your LAN. If you are connecting both Ethernet ports, they must each be connected to separate networks. The DGX-1 is not configured from the factory to have multiple Ethernet interfaces on the same network.

2. Using an Ethernet cable, connect the IPMI (BMC) port to your LAN for remote access to the base management controller (BMC).
Verify that all network cables are firmly inserted into the DGX-1 and the associated network switch.

2.9. Setting Up the DGX-1

These instructions describe the setup process that occurs the first time the DGX-1 is powered on after delivery. Be prepared to accept all EULAs and to set up your username and password.

1. Connect a display to the VGA connector, and a keyboard to any of the USB ports.



For best display results, use a monitor with a native resolution of 1024x768 or lower.

2. Power on the DGX-1.



The system will take a few minutes to boot.

You may be presented with end user license agreements (EULAs) for the NVIDIA software at this point in the setup, depending on the DGX-1 software version. Accept all EULAs to proceed with the installation.

You are prompted to configure the DGX-1 software.

3. Perform the steps to configure the DGX-1 software.

T

he steps are based on DGX OS 5.0. There are some variations in the process if the DGX system is installed with a previous DGX OS release.

- ▶ Select your language and locale preferences.
- ▶ Select the country for your keyboard.
- ▶ Select your time zone.

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

The administrator username is used also for the BMC login username and GRUB username.

DGX OS 4 and earlier: When logging in to the BMC, enter your username for both the User ID as well as the password. Be sure to create a unique BMC password at the first opportunity.



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.

- Create a BMC admin password.
The BMC password must consist of a minimum of 13 characters.
- [Available starting with DGX OS 5.0] Create a GRUB password.
You can click OK without entering a password which will disable this step, but NVIDIA recommends setting the GRUB password for security hardening.
- [Available starting with DGX OS 5.0] Create a root filesystem password.
You will need the new passphrase to unlock the root filesystem when the system boots.
This step appears only if you installed the system with an encrypted root filesystem during re-imaging.
- Choose a primary network interface for the DGX-1.
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.



Note: After you select the primary network interface, the system attempts to configure the interface for DHCP and then asks you to enter a hostname for the system. If DHCP is not available, you will have the option to configure the network manually. If you need to configure a static IP address on a network interface connected to a DHCP network, select **Cancel** at the *Network configuration – Please enter the hostname for the system* screen. The system will then present a screen with the option to configure the network manually.

- 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.
- If prompted, fill in the requested networking information such as name server or domain name.
- Choose a host name for the DGX-1.
- Choose to install predefined software.

Press the space bar to select or deselect the software to install.



Note: By default, the DGX-1 installs only minimal software packages necessary to ensure system functionality. You can deselect the OpenSSH package; however, NVIDIA recommends that you keep this package selected, and uninstall it only if required by your IT security policy.

4. Select **OK** to continue.

You may be presented with end user license agreements (EULAs) for the NVIDIA software at this point in the setup, depending on the DGX OS software version. Accept all EULAs to complete the installation.

The system completes the installation, reboots automatically, then presents the system login prompt:

```
<hostname> login:
Password:
```

5. Log in.



Note: Before using the DGX-1, check the [DGX OS Server release notes](#) for the availability of important over-the-network software updates, including security updates.

2.10. Post Setup Instructions for DGX OS Server Software Version 2.x and Earlier

These instructions apply if your DGX-1 is installed with software version 2.x or earlier.

To determine the DGX OS Server software version on your system, enter the following command.

```
$ grep VERSION /etc/dgx-release
DGX_SWBUILD_VERSION="2.0.4"
```

1. If your network is configured for DHCP, then make sure that dynamic DNS updates are enabled.

Check whether `/etc/resolv.conf` is a link to `/run/resolvconf/resolv.conf`.

```
$ ls -l /etc/resolv.conf
```

Expected output:

```
lrwxrwxrwx 1 root root 29 Dec 1 21:19 /etc/resolv.conf ->
../run/resolvconf/resolv.conf
```

- ▶ If the expected output appears, then skip to step 2.
- ▶ If this does not appear, then enable dynamic DNS updates as follows:

- a). Launch the *Resolvconf Reconfigure* package.

```
$ sudo dpkg-reconfigure resolvconf
```

The *Configuring resolvconf* screen appears.

- b). Select **<Yes>** when asked whether to prepare `/etc/resolv.conf` for dynamic updates.
- c). Select **<No>** when asked whether to append original file to dynamic file.
- d). Select **<OK>** at the *Reboot recommended* screen.

You do not need to reboot.

You are returned to the command line.

- e). Bring down the interface, where `<network interface>` is `em1` or `em2`, whichever you have set up as your primary network interface.

```
$ sudo ifdown <network interface>
```

Expected output:

```
ifdown: interface <network interface> not configured
```

- f). Bring up the interface, where `<network interface>` is `em1` or `em2`, whichever you have set up as your primary network interface.

```
$ sudo ifup <network interface>
```

Expected output (last line):

```
...
bound to <IP address> -- renewal in ...
```

- g). Repeat step 1 to confirm that `/etc/resolv.conf` is a link to `/run/resolvconf/resolv.conf`.

2. Make sure that the `nvidia-peer-memory` module is installed.

```
$ lsmod | grep nv_peer_mem
```

If the following output appears, then your DGX-1 setup is complete and you do not need to perform the next steps.

```
nv_peer_mem          16384    0
nvidia               11911168  30
nv_peer_mem,nvidia_modeset,nvidia_uv
mib_core             143360   13
rdma_cm,ib_cm,ib_sa,iw_cm,nv_peer_mem,mlx4_ib,mlx5_ib,
ib_mad,ib_ucm,ib_umad,ib_uverbs,rdma_ucm,ib_ipoib
```

3. If there is no output to the `lsmod` command, then build and install the `nvidia-peer-memory` module.

- a). Get and install the module.

```
$ sudo apt-get update
```

```
$ sudo apt-get install --reinstall mlnx-ofed-kernel-dkms nvidia-peer-memory-dkms
```

Expected output.

```
DKMS: install completed.
Processing triggers for initramfs-tools (0.103ubuntu4.2) ...
update-initramfs: Generating /boot/initrd.img-4.4.0-45-generic
```

- b). Add the module to the Linux kernel.

```
$ sudo modprobe nv_peer_mem
```

There is no expected output for this command.

- c). Repeat step 2 to confirm that the `nvidia-peer-memory` module has been added.

2.11. Updating the DGX-1 Software

You must register your DGX-1 in order to receive software updates. Once registered, you will receive an email notification whenever a new software update is available.



Note: Updating the software ensures your DGX-1 server 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.

These instructions explain how to update the DGX-1 software through an internet connection to the NVIDIA public repository. The process updates a DGX-1 system image to the latest QA'd versions of the entire DGX-1 software stack, including the drivers.

The instructions apply to a point update from the latest software release – for example, from 3.1.4 to 3.1.6. To upgrade from DGX OS Server release 1.0, 2.0, or 2.1, refer to the release notes for detailed 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-1 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. When 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.

4. Reboot the system.

2.12. 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. This capability is available starting with DGX OS Server release 4.4.0.

2.12.1. 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
```

2.12.2. 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.

2.12.3. 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 3. Preparing for Using Docker Containers

This chapter presents an overview of the prerequisites for accessing Docker containers from the command line for use on the NVIDIA® DGX-1™ in base OS mode. These containers include NVIDIA DGX-1 specific software to ensure the best performance for your applications. Using these containers as a basis for your applications should provide the best single-GPU performance and multi-GPU scaling.

- ▶ [Installing Docker and the Docker Engine Utility for NVIDIA GPUs on DGX OS Server Software 2.x or Earlier](#)
- ▶ [Configuring Docker IP Addresses](#)
- ▶ [Letting Users Issue Docker Commands](#)
- ▶ [Configuring a System Proxy](#)
- ▶ [Configuring NFS Mount and Cache](#)

3.1. Installing Docker and the Docker Engine Utility for NVIDIA GPUs on DGX OS Server Software 2.x or Earlier

To enable portability in Docker images that leverage GPUs, NVIDIA developed the Docker Engine Utility for NVIDIA GPUs (`nvidia-docker`), an open-source project that provides a command line tool to mount the user mode components of the NVIDIA driver and the GPUs into the Docker container at launch.

As of DGX OS Server software version 3.1.1 and later, Docker and the `nvidia-docker` utility are part of the base software installation and you do not need to perform the steps in this section. However, if your DGX-1 is installed with software version 2.x or earlier, then follow these instructions to install Docker and `nvidia-docker` on the system.

To determine the DGX OS Server software version on your system, enter the following command.

```
$ grep VERSION /etc/dgx-release
DGX_SWBUILD_VERSION="2.0.4"
```

Ensure your environment meets the prerequisites before installing Docker. For more information, see [Getting Started with Docker](#).

1. Install [Docker](#).

```
$ sudo apt-key adv --keyserver
hkp://p80.pool.sks-keyservers.net:80 --recv-keys
58118E89F3A912897C070ADB76221572C52609D
$ echo deb https://apt.dockerproject.org/repo ubuntu-trusty main
| sudo tee /etc/apt/sources.list.d/docker.list
$ sudo apt-get update
$ sudo apt-get -y install docker-engine=1.12.6-0~ubuntu-trusty
```

2. Edit the `/etc/default/docker` file to use the Overlay2 storage driver.

- a). Open the `/etc/default/docker` file for editing.

```
$ sudo vi /etc/default/docker
```

- b). Add the following line:

```
DOCKER_OPTS="--storage-driver=overlay2"
```

If there is already a `DOCKER_OPTS` line, then add the parameters (text between the quote marks) to the `DOCKER_OPTS` environment variable.

- c). Save and close the `/etc/default/docker` file when done.
- d). Restart Docker with the new configuration.

```
$ sudo service docker restart
```

3. Install the [NVIDIA Docker Engine Utility for NVIDIA GPUs](#).

The following example installs both `nvidia-docker` and the `nvidia-docker-plugin`.

```
$ wget -P /tmp
https://github.com/NVIDIA/nvidia-docker/releases/download/v1.0.1/nvidia-
docker_1.0.1-1_amd64.deb

$ sudo dpkg -i /tmp/nvidia-docker*.deb && rm
/tmp/nvidia-docker*.deb
```

3.2. Configuring Docker IP Addresses

To ensure that the DGX-1 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-1.

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-1, you should change the default Docker network addresses. The method for accomplishing this depends on the Base OS software version installed on the DGX-1.

1. If you don't know the Base OS software version installed on the DGX-1, then enter the following and inspect the VERSION entry.

```
$ grep VERSION /etc/dgx-release
DGX_SWBUILD_VERSION="3.1.1"
```

2. Follow the instructions in the section appropriate for the software version installed.

- ▶ [Configuring Docker IP Addresses for DGX OS Server Software Version 2.x and Earlier](#)
- ▶ [Configuring Docker IP Addresses for DGX OS Server Software Version 3.1.1 and Later](#)

3.2.1. Configuring Docker IP Addresses for DGX OS Server Software Version 2.x and Earlier

1. Open the `/etc/default/docker` file for editing.

```
$ sudo vi /etc/default/docker
```

2. Modify the `/etc/default/docker` file, specifying the correct bridge IP address and IP address ranges for your network. Consult your IT administrator for the correct addresses.

For example, if your DNS server exists at IP address `10.10.254.254`, and the `192.168.0.0/24` subnet is not otherwise needed by the DGX-1, you can add the following line to the `/etc/default/docker` file:

```
DOCKER_OPTS="--dns 10.10.254.254 --bip=192.168.0.1/24 --fixedcidr=192.168.0.0/24"
```

If there is already a `DOCKER_OPTS` line, then add the parameters (text between the quote marks) to the `DOCKER_OPTS` environment variable.

3. Save and close the `/etc/default/docker` file when done.
4. Restart Docker with the new configuration.

```
$ sudo service docker restart
```

3.2.2. Configuring Docker IP Addresses for DGX OS Server Software Version 3.1.1 and Later

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
```

3.3. Letting Users Issue Docker Commands

To prevent the `docker` daemon from running without protection against escalation of privileges, the NVIDIA Docker software requires `sudo` privileges to run containers.

You can grant the required privileges to users who will run containers on the DGX-1 in one of the following ways:

- ▶ Add each user as an administrator user with `sudo` privileges.
- ▶ Add each user as a standard user without `sudo` privileges and then add the user to the `docker` group.

This section provides instructions for adding users to the `docker` group.



Note: WARNING: Only add users to the `docker` group whom you would trust with root privilege. These instructions make it more convenient for users to access Docker containers; however, the resulting `docker` group is equivalent to the root user, because once a user is able to send commands to the Docker engine, they are able to escalate privilege and run root level operations. This may violate your organization's security policies. See the [Docker Daemon Attack Surface](#) for information on how this can impact security in your system. Always consult your IT department to make sure the installation is in accordance with the security policies of your data center.



Note: The commands in this section require `sudo` access, and should be performed by a system administrator.

3.3.1. Checking if a User is in the Docker Group

To check whether a user is already part of the `docker` group, enter the following:

```
$ groups username
```

The output shows all the groups of which that user is a member. If `docker` is not listed, then add that user.

3.3.2. Creating a User

To create a new user in order to add them to the `docker` group, perform the following:

```
$ sudo adduser username
```

Follow the prompts for creating a password and other user configuration settings.

3.3.3. Adding a User to the Docker Group

For each user you want to add to the `docker` group, enter the following command:

```
$ sudo usermod -a -G docker username
```

3.4. Enabling GPU Support for NGC Containers

To obtain the best performance when running NGC containers, three methods of providing GPU support for Docker containers have been developed:

- ▶ Native GPU support (included with Docker-ce 19.03 or later)
- ▶ NVIDIA Container Runtime for Docker (`nvidia-docker2` package)
- ▶ Docker Engine Utility for NVIDIA GPUs (`nvidia-docker` package)

The method implemented in your system depends on the DGX OS version installed (for DGX systems), the specific NGC Cloud Image provided by a Cloud Service Provider, or the software that you have installed in preparation for running NGC containers on TITAN PCs, Quadro PCs, or vGPUs.

Refer to the following table to assist in determining which method is implemented in your system.

DGX OS Release	Method Included
4.2	Native GPU support NVIDIA Container Runtime for Docker
4.1	NVIDIA Container Runtime for Docker
4.0	NVIDIA Container Runtime for Docker
3.1 (versions 3.1.8, 3.1.7, and 3.1.6)	Docker Engine Utility for NVIDIA GPUs but can be updated to the NVIDIA Container Runtime for Docker
3.1 (up to version 3.1.5)	Docker Engine Utility for NVIDIA GPUs
2.1	Docker Engine Utility for NVIDIA GPUs
2.0	Docker Engine Utility for NVIDIA GPUs

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

Using Native GPU support



Note: If Docker is updated to 19.03 on a system which also has `nvidia-docker2` installed, then instructions for using the NVIDIA Container Runtime for Docker can still be used.

- 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" ...
```

Using the NVIDIA Container Runtime for Docker

With the NVIDIA Container Runtime for Docker installed (`nvidia-docker2`), you can 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 new package provides backward compatibility, so you can still run GPU-accelerated containers by 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.

Using the Docker Engine Utility for NVIDIA GPUs

With the Docker Engine Utility for NVIDIA GPUs installed (`nvidia-docker`), run GPU-enabled containers as follows.

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

3.5. Configuring a System Proxy

If you will be using the DGX-1 in base OS mode, and your network requires use of a proxy, then edit the file `/etc/apt/apt.conf.d/proxy.conf` and make sure the following lines are present, using the parameters that apply to your network:

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

To ensure that Docker is able to access the DGX-1 Container Registry through the 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>.

3.6. Configuring NFS Mount and Cache

The DGX-1 includes four SSDs in a RAID 0 configuration. These SSDs are intended for application caching, so you must set up your own NFS drives for long term data storage.

Disabling `cachedfilesd`

The DGX-1 system uses `cachedfilesd` to manage caching of the NFS. If you do not want `cachedfilesd` enabled, you can disable it as follows.

```
sudo systemctl stop cachedfilesd
sudo systemctl disable cachedfilesd
```

Using `cachedfilesd`

The following instructions describe how to mount the NFS onto the DGX-1, and how to cache the NFS using the DGX-1 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-1, and that there is network access between the DGX-1 and the NFS server.

1. Configure an NFS mount for the DGX-1.
 - 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,wsiz=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 the example mount point used in step 1.

4. Verify caching is enabled.

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

Look for the text FSC=yes in the output.

Upon rebooting, the NFS should be mounted and cached on the DGX-1.

Chapter 4. Configuring and Managing the DGX-1

This chapter describes the following DGX-1 configuration and management tasks:

- ▶ [Using the BMC](#)
- ▶ [Configuring a Static IP Address for the BMC](#)
- ▶ [Configuring Static IP Addresses for the Network Ports](#)
- ▶ [Obtaining MAC Addresses](#)
- ▶ [Resetting GPUs](#)
- ▶ [Changing the Mellanox Port Type](#)
- ▶ [Enabling USB 3.0](#)
- ▶

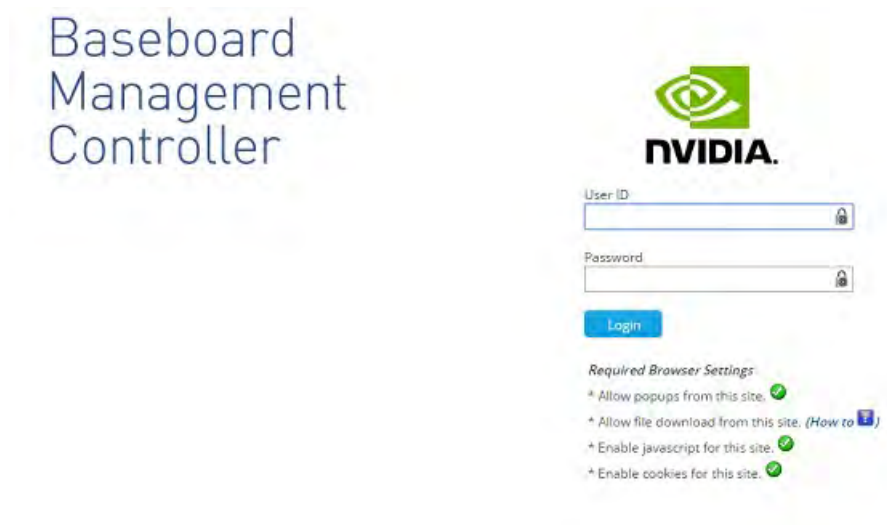
4.1. Using the BMC

The DGX-1 includes a baseboard management controller (BMC) that lets you manage and monitor the DGX-1 independently of the CPU or operating system. You can access the BMC remotely through the Ethernet connection to the IPMI port.

This section describes how to access the BMC, and describes a few common tasks that you can accomplish through the BMC. It is not meant to be a comprehensive description of all the BMC capabilities.

To access the BMC remotely:

1. Make sure you have connected the IPMI port on the DGX-1 to your LAN.
2. Open a Java-enabled browser within your LAN and go to `http://<IPMI IP Address>/`.
Use Firefox or Internet Explorer. Google Chrome is not officially supported by the BMC.



3. Log in.

Your initial log in credentials are based on the ones you created when you first set up the DGX-1. Enter your username for both the User ID as well as the Password.

User ID: <your username>

Password: <your username>.

4. Be sure to change your password immediately to ensure the security of the BMC.

See the next section for instructions on how to change your BMC password.

4.1.1. Creating a Unique BMC Password for Remote Access

When you set up the DGX-1 upon powering it on for the first time, you set up a username and password for the system. These credentials are also used to log in to the BMC remotely, except that the BMC password is the username.

It is strongly recommended that you create a unique password as soon as possible.

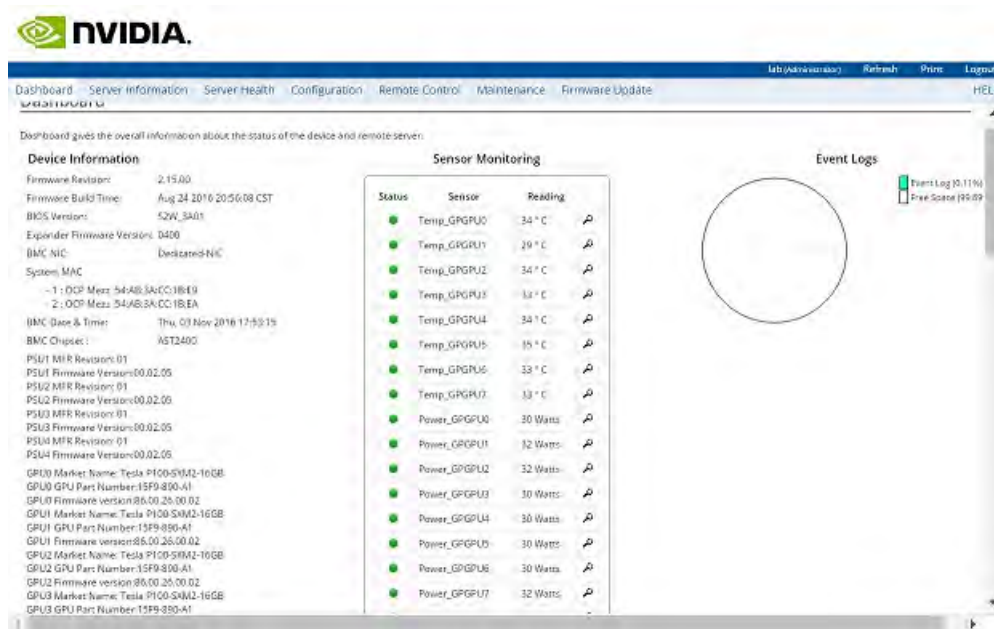
Create a unique BMC password as follows:

1. Open a Java-enabled web browser within your LAN and go to `http://<IPMI IP address>/`.
Use Firefox or Internet Explorer. Google Chrome is not officially supported by the BMC.
2. Log in with the username that you created when you first set up the DGX-1.
Enter your username for both the User ID as well as the password:
User ID: <your username>
Password: <your username>
3. From the top menu, click **Configuration** and then select **User**.
4. Select your username and then click **Modify User**.
5. In the *Modify User* dialog, select **Change Password**, and then enter your new password in the *Password* and *Confirm Password* boxes.

- Click **Modify** when finished.

4.1.2. Viewing System Information

The BMC opens to the dashboard, which shows information about the system and system components, such as temperatures and voltages.



4.1.3. Submitting BMC Log Files

The BMC provides automatic logging of system activities and status. The NVIDIA Enterprise Support team uses the log files to assist in troubleshooting. Follow these instructions to obtain the log files to send to NVIDIA Enterprise Support.

- Log into the BMC, then click **Server Health** from the top menu and select **Event Log**.
- Make sure that **Text** is selected at *Format of Download Event Logs*.









Format of Download Event Logs: ☒ Text ☐ Hex

- Click **Save Event Logs** to download the event logs.

4.1.4. Determining Total Power Consumption

You can use the BMC dashboard to determine total power consumption of the DGX-1 as follows:

- Log into the BMC.
- From the BMC dashboard, locate the *Sensor Monitoring* area and then scroll down the page until you see the *PSU Input* rows.

	PSU1 Input	216 Watts	
	PSU2 Input	216 Watts	
	PSU3 Input	135 Watts	
	PSU4 Input	27 Watts	

3. Add the values for all the PSUs.

In this example, the total power consumption would be $216+216+135+27 = 594$ watts.

4.1.5. Accessing the DGX-1 Console

The DGX-1 BMC uses JViewer to view the console.



Note: To use JViewer with Mac OS, first install XQuartz, available from <https://www.xquartz.org>.

1. Log into the BMC.
2. From the top menu, click **Remote Control** and then select **Console Redirection**.
3. Click **Java Console** to open the popup window.

The window provides interactive control of the DGX-1 console.

4.1.6. Powering Off / Power Cycling the System Remotely

4.1.6.1. From the DGX-1 Console Window

If you have opened the Java Viewer (*Remote Control->Console Redirection*) to view the console window, then you can power cycle, reset, or shutdown the DGX-1 as follows:



Note: To use JViewer with Mac OS, first install XQuartz, available from <https://www.xquartz.org>.

1. From the JViewer top menu, click **Power** and then select from the available options, depending on what you want to do.



2. Click **Yes** and then **OK** at the *Power Control* dialog, then wait for the system to perform the intended action.

4.1.6.2. From the BMC UI

1. Log into the BMC.
2. From the top menu, click **Remote Control** and then select **Server Power Control**.



The screenshot shows the NVIDIA DGX-1 BMC interface. The top navigation bar includes 'Dashboard', 'Server Information', 'Server Health', 'Configuration', 'Remote Control', 'Maintenance', 'Firmware Update', and 'HELP'. The 'Remote Control' menu is open, showing options: 'Console Redirection', 'Server Power Control' (highlighted with a red circle), and 'Java SOL'. Below the menu, the 'Power Control and Status' section displays the current server power status as 'Host is currently on'. It lists several power control options: 'Reset Server', 'Power Off Server - Immediate', 'Power Off Server - Orderly Shutdown' (selected), 'Power On Server', and 'Power Cycle Server'. A 'Perform Action' button is located at the bottom right.

3. Select from the available options according to what you want the system to do, then click **Perform Action**.

4.1.7. Setting Up Active Directory

You can access the NVIDIA DGX-1 system using active directory (AD) authentication for your organization. The simplest method is to use the BMC dashboard.

1. Log into the BMC.
2. From the top menu, click **Configuration** and then select **Active Directory**.
3. Click **Advanced Settings** to open the **Advanced Active Directory Settings** pop-up window, then fill out the information for your network.

The screenshot shows the 'Advanced Active Directory Settings' pop-up window. It contains the following fields and controls:

- Active Directory Authentication:** A checkbox labeled 'Enable' which is checked.
- Secret Username:** An empty text input field.
- Secret Password:** An empty text input field.
- User Domain Name:** A text input field containing 'ad.example.com'.
- Domain Controller Server Address1:** A text input field containing 'xxx.xxx.xxx.xxx' with a dropdown arrow on the right.
- Domain Controller Server Address2:** An empty text input field.
- Domain Controller Server Address3:** An empty text input field.
- Buttons:** 'Save' and 'Cancel' buttons at the bottom right.

- a). Select the **Active Directory Authentication Enable** check box.
- b). **Do not** fill out the Secret Username and Secret Password boxes.

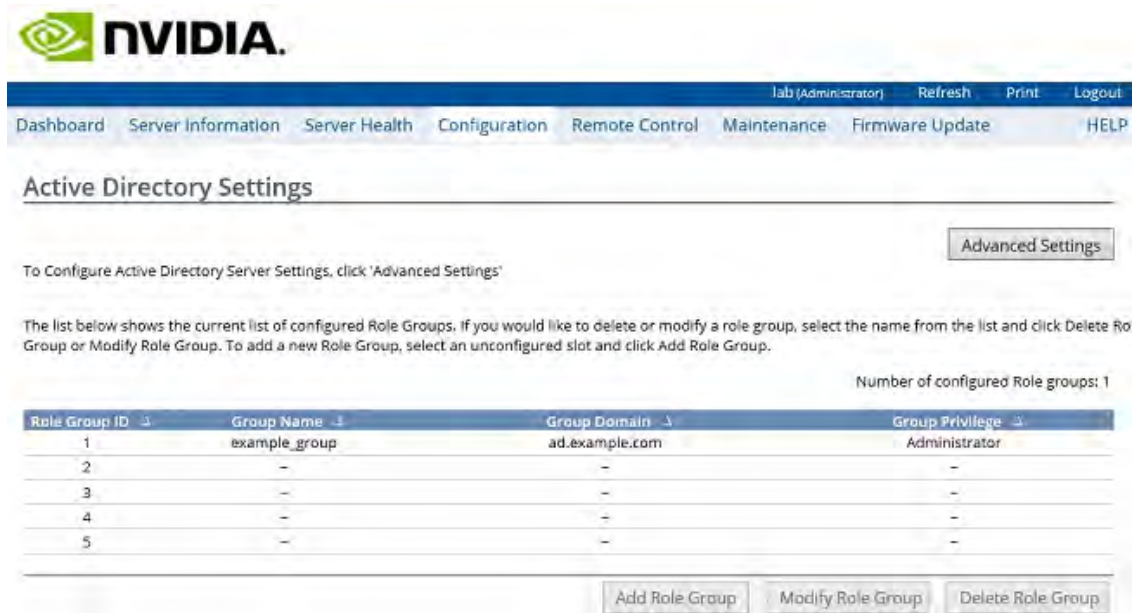
- c). Fill out the **User Domain Name** box and only one of the **Domain Controller Server Address** boxes.
- d). Click **Save** when done.
- 4. Add role groups as needed.
 - a). Click the first empty **Role Group** line and then click **Add Role Group**.
 - b). Enter your role group information in the **Add Role Group** dialog box.

- c). Be aware that the **User** role group privilege does not allow access to the BMC dashboard.

The following are descriptions of each role group privilege option.

Role Group Privilege	Description
Administrator	Access to all BMC dashboard functions
Operator	Read-only access to BMC dashboard controls
User	No access to the BMC dashboard; ipmitool supported
Callback	No access to the BMC dashboard; ipmitool supported
OEM Proprietary	Read-only access to BMC dashboard controls except for User, DNS, Network, and PEF features

- d). Enable extended privileges as needed.
- e). Click **Add** when done.



4.2. Configuring a 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 the System BIOS](#)
- ▶ [Configuring the BMC Static IP Address Using ipmitool](#)
- ▶ [Configuring the BMC Static IP Address Using the BMC User Interface](#)

4.2.1. 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-1 remotely, then connect a display (1024x768 or lower resolution) and keyboard directly to the DGX-1.

To view the current settings, enter the following command.

```
$ sudo ipmitool lan print 1
Set in Progress      : Set Complete
Auth Type Support    : MD5
Auth Type Enable     : Callback : MD5
```

```

: User : MD5
: Operator : MD5
: Admin : MD5
: OEM : MD5
IP Address Source : DHCP Address
IP Address : 10.31.241.190
Subnet Mask : 255.255.255.0
MAC Address : 54:ab:3a:72:08:a9
SNMP Community String : Quanta
IP Header : TTL=0x40 Flags=0x40 Precedence=0x00 TOS=0x10
BMC ARP Control : ARP Responses Enabled, Gratuitous ARP Disabled
Gratuitous ARP Intrvl : 0.0 seconds
Default Gateway IP : 10.31.241.1
Default Gateway MAC : 00:00:00:00:00:00
Backup Gateway IP : 0.0.0.0
Backup Gateway MAC : 00:00:00:00:00:00
802.1q VLAN ID : Disabled
802.1q VLAN Priority : ORMCp+ Cipher Suites : 0,1,2,3,6,7,8,11,12,15,16,17
Cipher Suite Priv Max : XaaaaaaaaaXXX
: X=Cipher Suite Unused
: c=CALLBACK
: u=USER
: o=OPERATOR
: a=ADMIN
: O=OEM

```

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 10.31.241.190
```

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

```
$ sudo ipmitool lan set 1 netmask 255.255.255.0
```

- ▶ 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 10.31.241.1
```

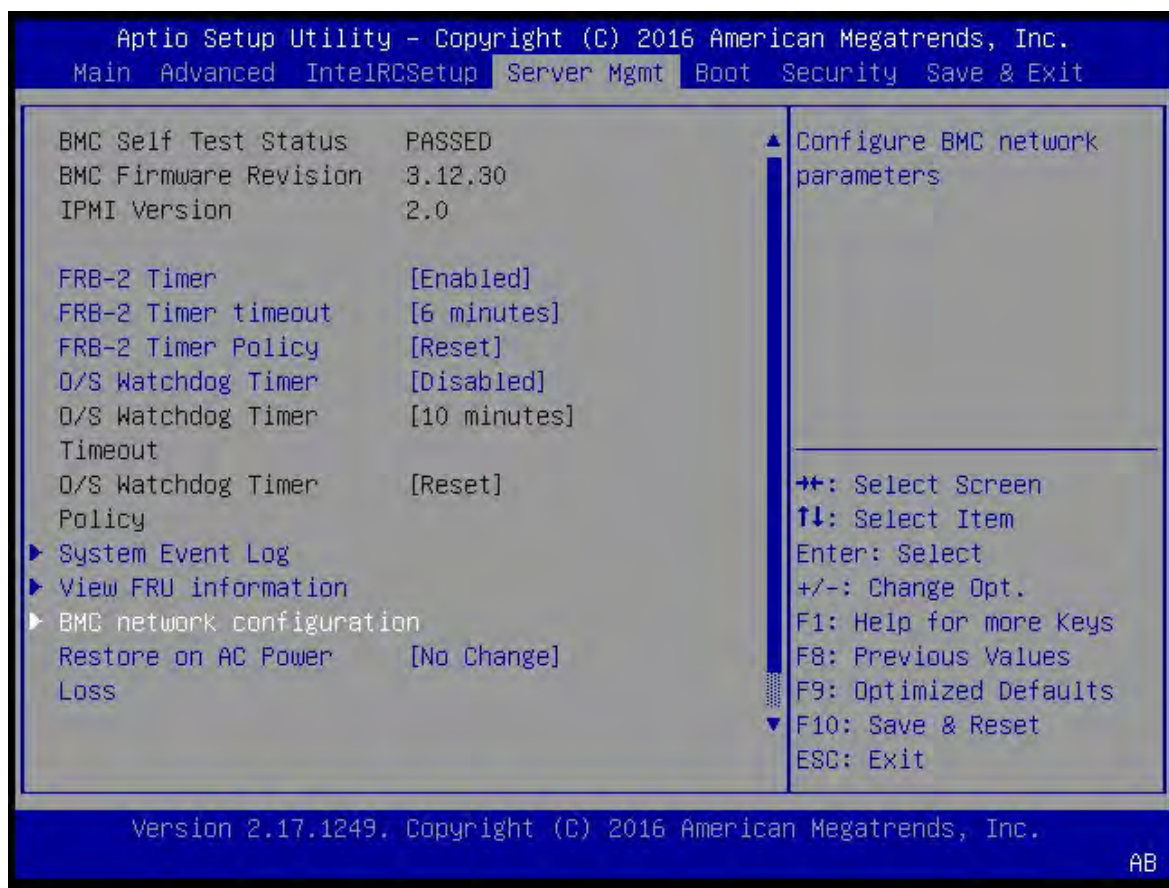
4.2.2. Configuring a BMC Static IP Address Using the System BIOS

This section describes how to set a static IP address for the BMC when you cannot access the DGX-1 remotely. This process involves setting the BMC IP address during system boot.

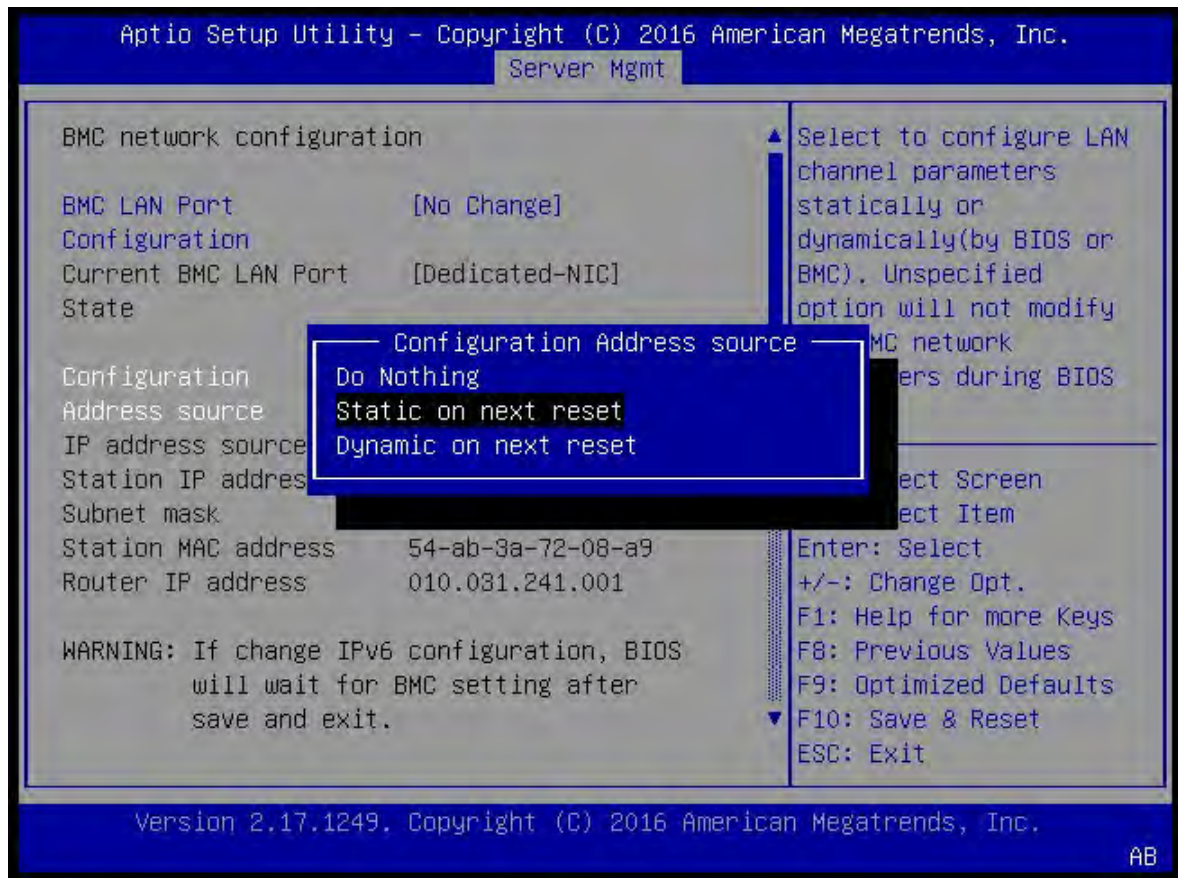
1. Connect a keyboard and display (1024x768 or lower resolution) to the DGX-1, then turn on the DGX-1.
2. When you see the NVIDIA logo, press **Del** to enter the BIOS Utility Setup Screen.



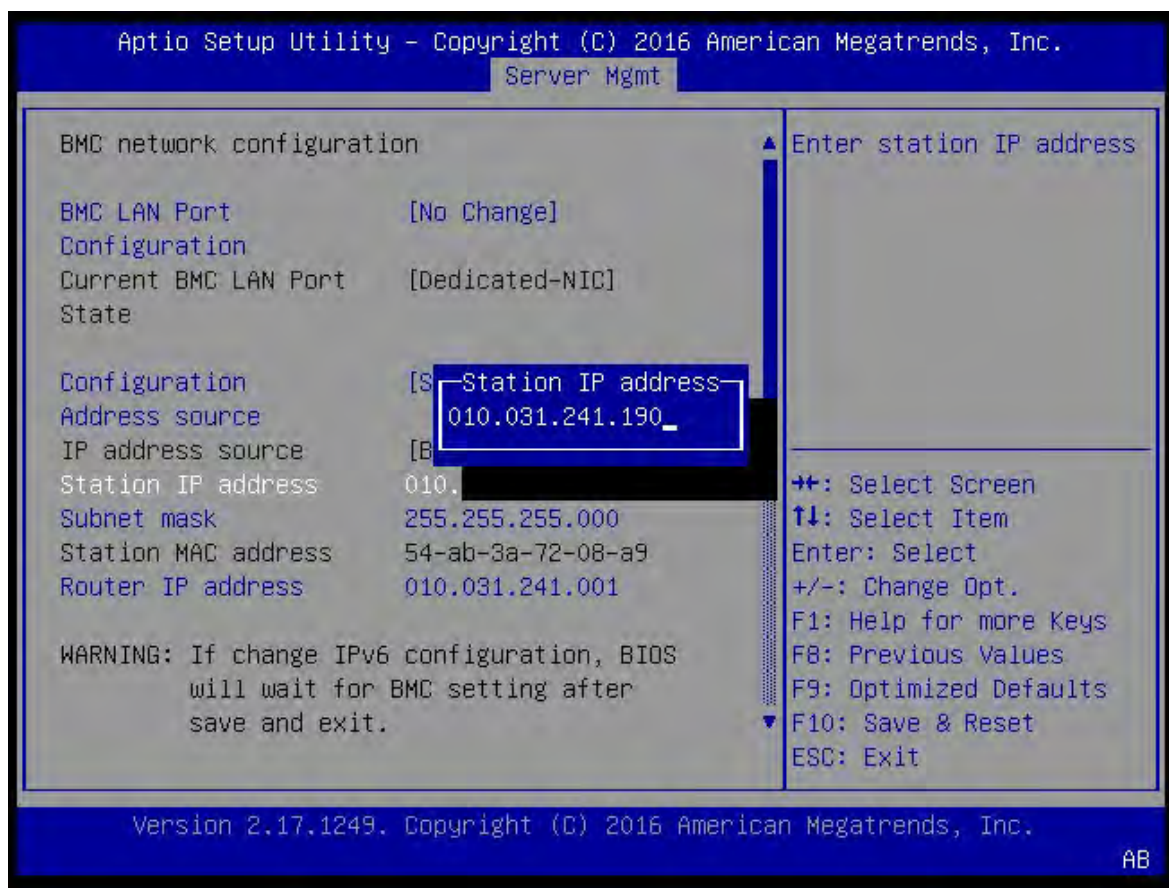
3. At the BIOS Setup Utility screen, navigate to the **Server Mgmt** tab on the top menu, then scroll to **BMC network configuration** and press **Enter**.



4. Scroll to **Configuration Address Source** and press **Enter** , then at the **Configuration Address source** pop-up, select **Static on next reset** 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.

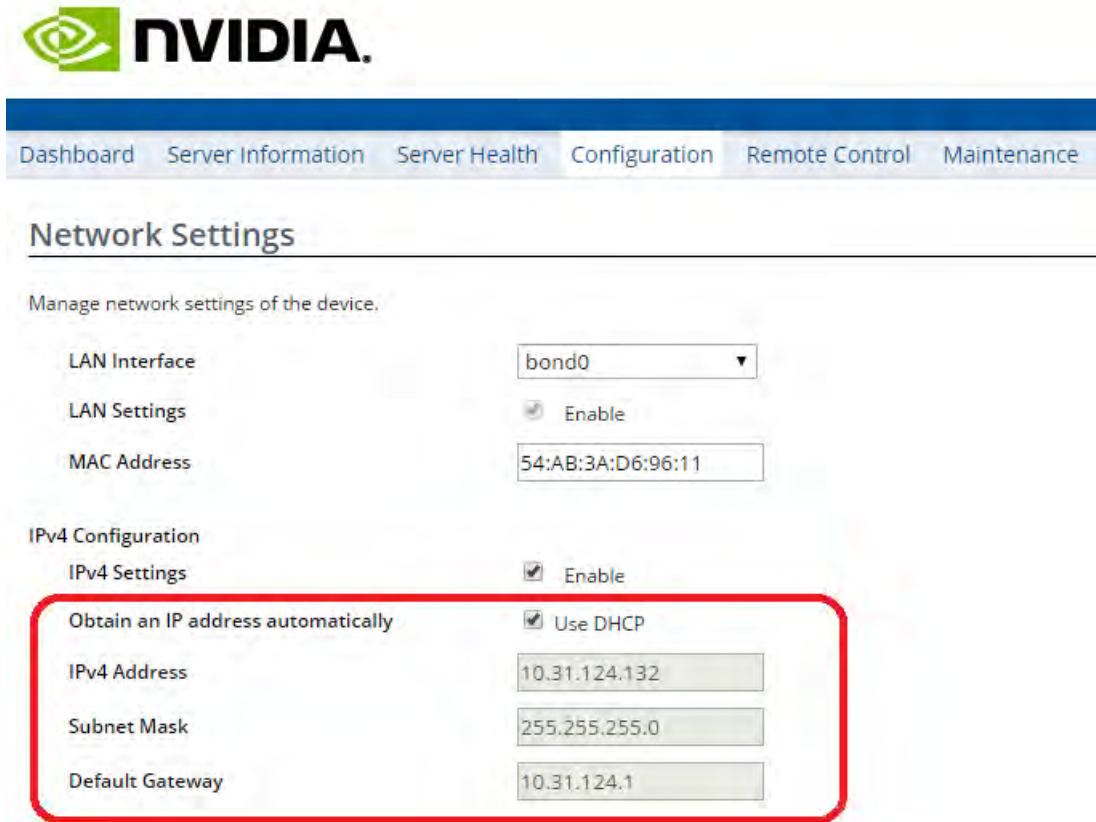


- When finished making all your changes, press **F10** to Save & Reset, then select **Yes** at the confirmation pop-up and press **Enter**.

You can now access the BMC over the network.

4.2.3. Configuring a BMC Static IP Address Using the BMC Dashboard

- Log into the BMC, then click **Configuration** from the top menu and select **Network Settings**.
- In the *IPv4 Configuration* section of the Network Settings page, clear the **Use DHCP** check box, and then enter the appropriate values for the **IPv4 Address** , **Subnet Mask** , and **Default Gateway** fields.



NVIDIA

Dashboard Server Information Server Health Configuration Remote Control Maintenance

Network Settings

Manage network settings of the device.

LAN Interface

LAN Settings ☒ Enable

MAC Address

IPv4 Configuration

IPv4 Settings ☒ Enable

Obtain an IP address automatically ☒ Use DHCP

IPv4 Address

Subnet Mask

Default Gateway

3. Click **Save** when done.

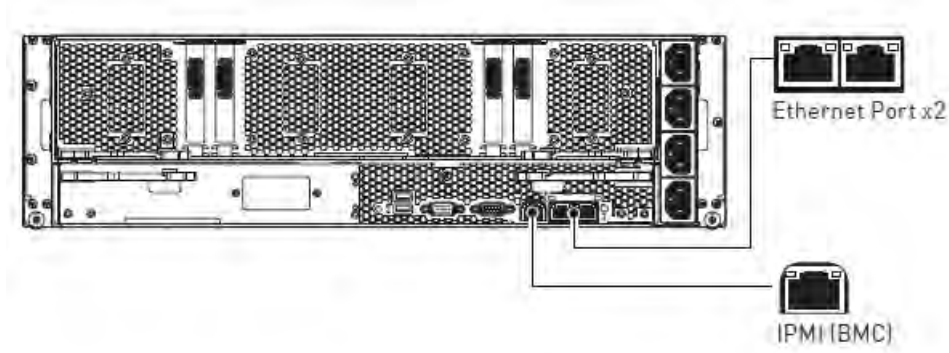
4.3. Configuring Static IP Addresses for the Network Ports

During the initial boot setup process for the DGX-1, you had an opportunity to configure static IP addresses for the network ports. If you did not set this up at that time, you can configure the static IP addresses from the Ubuntu command line according to the following instructions.



Note: If you cannot access the DGX-1 remotely, then connect a display (1024x768 or lower resolution) and keyboard directly to the DGX-1.

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



Use the following port designations according to the DGX-1 Base OS software version installed on the DGX-1:

Ethernet Port Position	Port Designation: Base OS Software 2.x and earlier	Port Designation: Base OS Software 3.x and later
Right Side	em1	enp1s0f0
Left Side	em2	enp1s0f1

2. Edit the interfaces file.

```
$ sudo vi /etc/network/interfaces
## Configure a static IP
auto em1
iface em1 inet static
    address 192.168.1.14
    gateway 192.168.1.1
    netmask 255.255.255.0
    network 192.168.1.0
    broadcast 192.168.1.255
```

Consult your network administrator for the appropriate addresses for your network, 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.

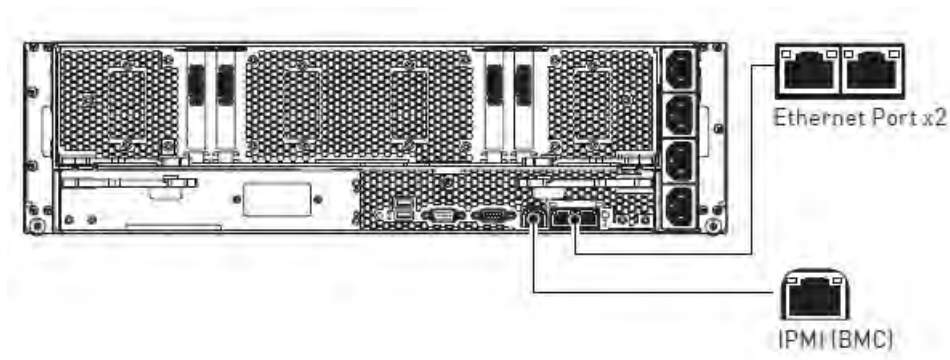
```
:wq
```

4. Restart the network services to put the changes into effect.

```
$ sudo /etc/init.d/networking restart
```

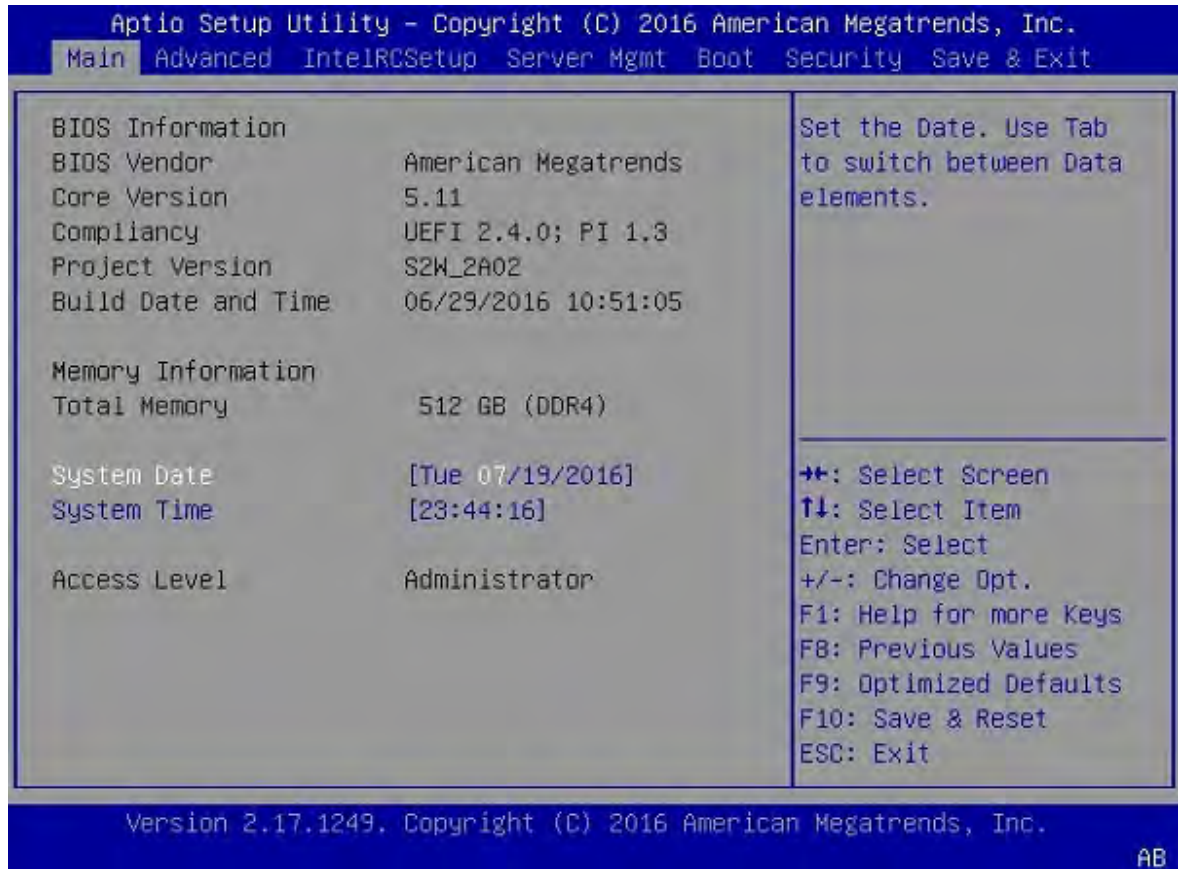
4.4. Obtaining MAC Addresses

These instructions explain how to determine the MAC addresses for the IPMI port (BMC) as well as both ethernet ports of the DGX-1.

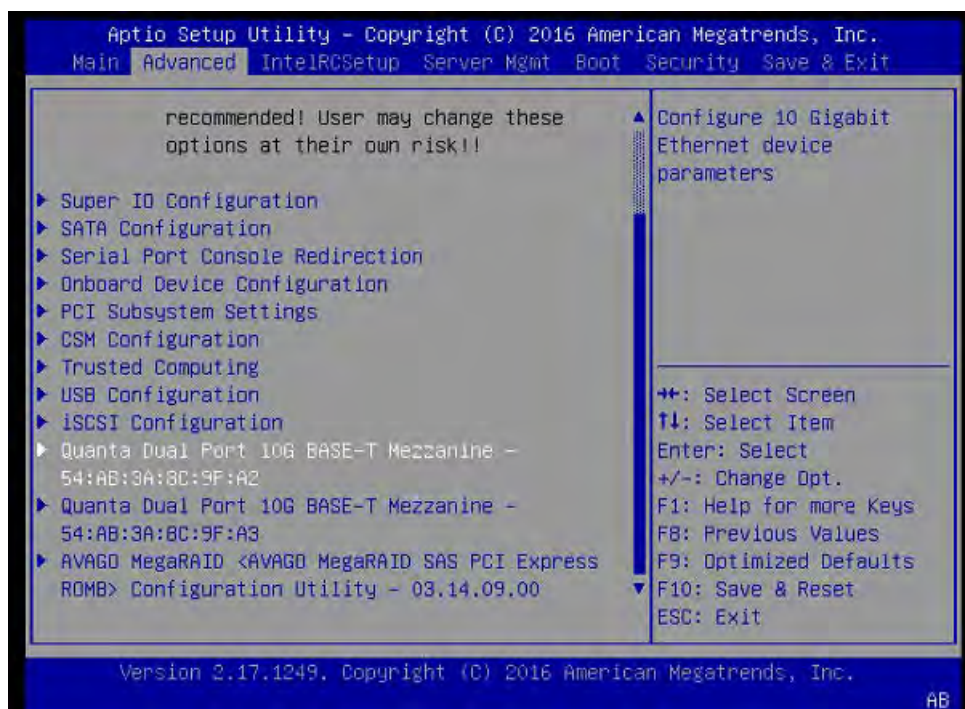


The ports are, from left to right, IPMI (BMC), em2 (or enp1s0f1), em1 (or enp1s0f0).

1. Connect a display (1024x768 or lower resolution) and keyboard to the DGX-1.
2. Turn the DGX-1 on or reboot.
3. At the NVIDIA logo boot screen, press **[F2]** or **[Del]** to enter the BIOS setup screen.

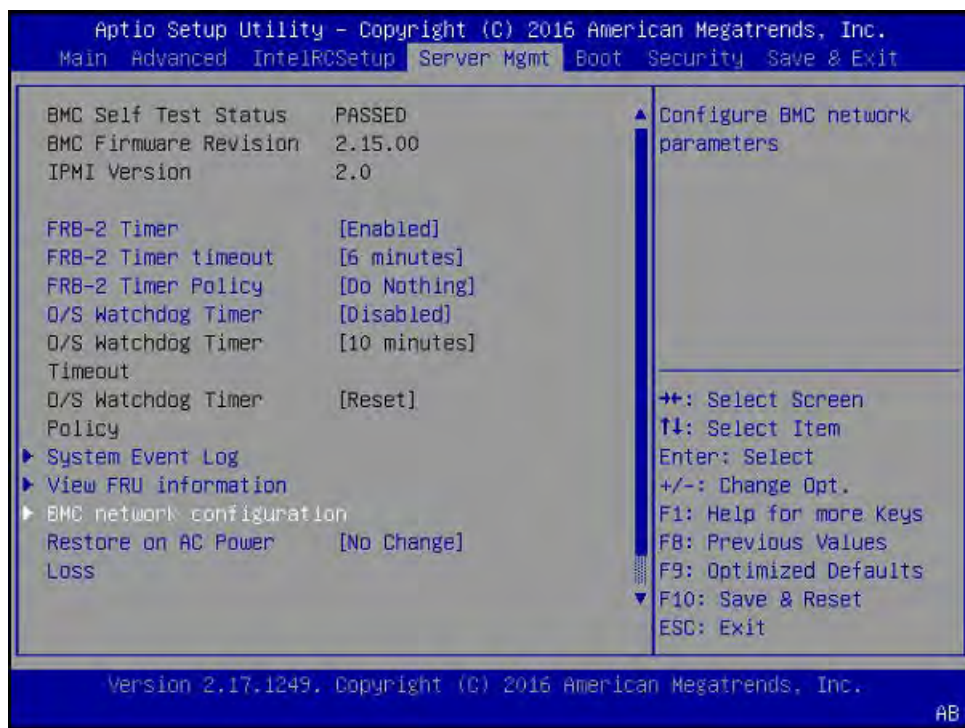


4. Select the **Advanced** tab from the top menu, then scroll down to view the two **Quanta Dual Port 10G BASE-T Mezzanine** items.



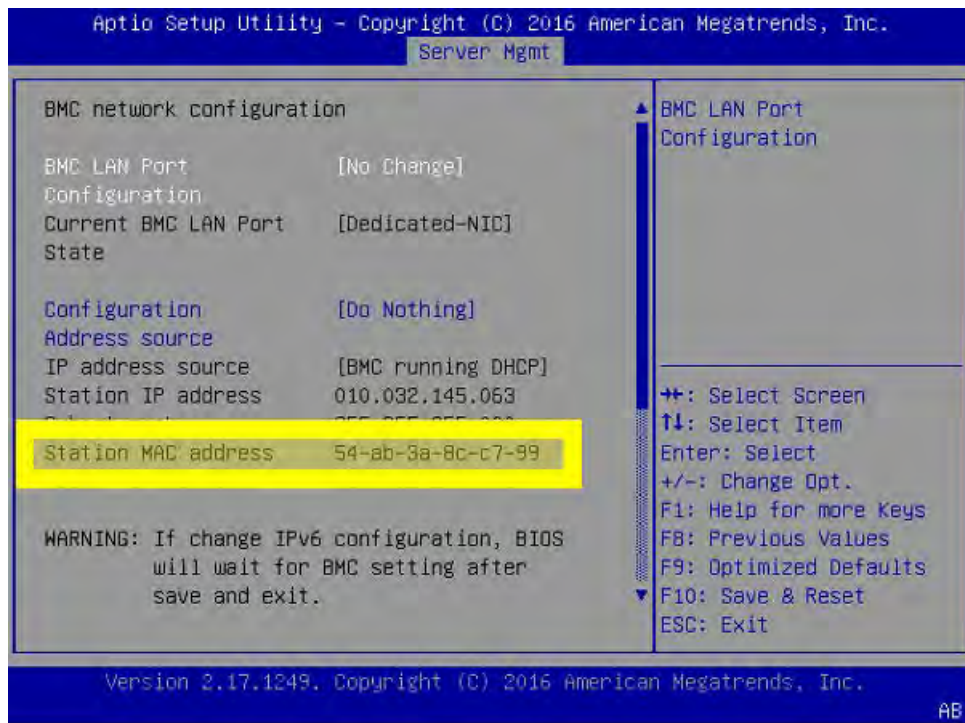
The first item shows the MAC address for ethernet port em1, and the second item shows the MAC address for em2.

5. Navigate to and select **Server Mgmt** from the top menu, then scroll down to and select **BMC network configuration**.



6. Scroll down to view the Station MAC address.

This shows the MAC address for the BMC.



4.5. Resetting GPUs in the DGX-1

The NVIDIA® DGX-1™ GPUs may need to be reset during maintenance or repair work. For example, when a GPU presents Double Bit Errors or repeat Single Bit Errors in the same location, pages are retired. For the retired pages to be excluded (unavailable to the user/application), the GPU needs to be reset. In the case of the DGX-1 platforms, individual GPUs cannot be reset because they are linked via NVLink, so all the GPUs must be reset simultaneously.

The only way to reset all the GPUs in the system is to reboot the system.

4.6. Changing the Mellanox Card Port Type

Data center environments can vary in the type of network connectivity used. The NVIDIA DGX-1 system is equipped with either 4 Single port Mellanox ConnectX-4 VPI MCX455A-ECAT modules or 4 Single port Mellanox ConnectX-5 VPI MCX555A-ECAT modules. By default these are configured as Infiniband ports, but you have the option to convert these to Ethernet ports. In order 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.

Downloading the Mellanox Software Tools

These instructions apply only to DGX-1 servers installed with DGX OS Server release 3.1 or earlier. DGX OS Server release 4.0 and later already include the Mellanox software tools (MST).

Do not install the Mellanox software tools (MST) with DGX OS Server release 4.0 and later; installing the MST will corrupt the included MST installation.

1. Download http://www.mellanox.com/downloads/MFT/mft-4.6.0-48-x86_64-deb.tgz.
2. Untar the downloaded package.
3. Switch to the download directory `mft-4.6.0-48-x86_64-deb/` and then run the installation script.

```
$ install.sh
```

Starting the Mellanox Software Tools

1. Start the mst driver:

```
$ sudo mst start
```

2. To verify that the Mellanox Software Tools (MST) services are running, enter the following.

```
~$ sudo mst status
```

- ▶ The following output indicates the services are **not** running.

```
MST modules:
-----
MST PCI module is not loaded
MST PCI configuration module is not loaded

PCI Devices:
-----
05:00.0
84:00.0
0c:00.0
8b:00.0
```

- ▶ The following output indicates the services are running.

```
MST modules:
-----
MST PCI module is not loaded
MST PCI configuration module loaded
MST devices:
-----
/dev/mst/mt4115_pciconf0 - PCI configuration cycles access.
                           domain:bus:dev.fn=0000:05:00.0 addr.reg=88
data.reg=92
                           Chip revision is: 00
/dev/mst/mt4115_pciconf1 - PCI configuration cycles access.
                           domain:bus:dev.fn=0000:0c:00.0 addr.reg=88
data.reg=92
                           Chip revision is: 00
/dev/mst/mt4115_pciconf2 - PCI configuration cycles access.
```

```

data.reg=92                                domain:bus:dev.fn=0000:84:00.0 addr.reg=88
                                           Chip revision is: 00
/dev/mst/mt4115_pciconf3 - PCI configuration cycles access.
data.reg=92                                domain:bus:dev.fn=0000:8b:00.0 addr.reg=88
                                           Chip revision is: 00
-1:~$

```

Determining the Current Port Configuration

To determine the current port configuration, enter the following,

```
~$ ibv_devinfo | grep -e "hca_id\\|link_layer"
```

- The following output indicates the ports are configured for InfiniBand.

```

hca_id: mlx5_3  link_layer:  InfiniBand
hca_id: mlx5_2  link_layer:  InfiniBand
hca_id: mlx5_1  link_layer:  InfiniBand
hca_id: mlx5_0  link_layer:  InfiniBand

```

- The following output indicates the ports are configured for Ethernet.

```

hca_id: mlx5_3  link_layer:  Ethernet
hca_id: mlx5_2  link_layer:  Ethernet
hca_id: mlx5_1  link_layer:  Ethernet
hca_id: mlx5_0  link_layer:  Ethernet

```

4.6.1. Switching the Port from InfiniBand to Ethernet

Make sure that you have installed the Mellanox Software Tools (MST) services and have started running them as explain in the section [Downloading and Starting the Mellanox Software Tools](#).

1. Change the configuration for all four ports to Ethernet.

```

~$ sudo mlxconfig -y -d /dev/mst/mt4115_pciconf0 set LINK_TYPE_P1=2
~$ sudo mlxconfig -y -d /dev/mst/mt4115_pciconf1 set LINK_TYPE_P1=2
~$ sudo mlxconfig -y -d /dev/mst/mt4115_pciconf2 set LINK_TYPE_P1=2
~$ sudo mlxconfig -y -d /dev/mst/mt4115_pciconf3 set LINK_TYPE_P1=2

```

2. Verify the configuration changes have been applied.

```

~$ sudo mlxconfig query |grep -e "LINK_TYPE\\|PCI\\ device"
PCI device: /dev/mst/mt4115_pciconf3 LINK_TYPE_P1 ETH(2)
PCI device: /dev/mst/mt4115_pciconf2 LINK_TYPE_P1 ETH(2)
PCI device: /dev/mst/mt4115_pciconf1 LINK_TYPE_P1 ETH(2)
PCI device: /dev/mst/mt4115_pciconf0 LINK_TYPE_P1 ETH(2)

```



Note: These changes will not take place until the server is rebooted.

3. Reboot the DGX-1.
4. After the DGX-1 is back online, make sure that Ethernet is the active link layer.

```

~$ ibv_devinfo |grep -e "hca_id\\|link_layer"
hca_id: mlx5_3  link_layer:  Ethernet
hca_id: mlx5_2  link_layer:  Ethernet
hca_id: mlx5_1  link_layer:  Ethernet
hca_id: mlx5_0  link_layer:  Ethernet

```

4.6.2. Switching the Port from Ethernet to InfiniBand

Make sure that you have installed the Mellanox Software Tools (MST) services and have started running them as explain in the section [Downloading and Starting the Mellanox Software Tools](#).

1. Change the configuration for all four ports to InfiniBand.

```
~$ sudo mlxconfig -y -d /dev/mst/mt4115_pciconf3 set LINK_TYPE_P1=1
~$ sudo mlxconfig -y -d /dev/mst/mt4115_pciconf2 set LINK_TYPE_P1=1
~$ sudo mlxconfig -y -d /dev/mst/mt4115_pciconf1 set LINK_TYPE_P1=1
~$ sudo mlxconfig -y -d /dev/mst/mt4115_pciconf0 set LINK_TYPE_P1=1
```

2. Verify the configuration changes have been applied.

```
~$ sudo mlxconfig query |grep -e "LINK_TYPE\|PCI\ device"
PCI device: /dev/mst/mt4115_pciconf3 LINK_TYPE_P1 IB(1)
PCI device: /dev/mst/mt4115_pciconf2 LINK_TYPE_P1 IB(1)
PCI device: /dev/mst/mt4115_pciconf1 LINK_TYPE_P1 IB(1)
PCI device: /dev/mst/mt4115_pciconf0 LINK_TYPE_P1 IB(1)
```



Note: These changes will not take place until the server is rebooted.

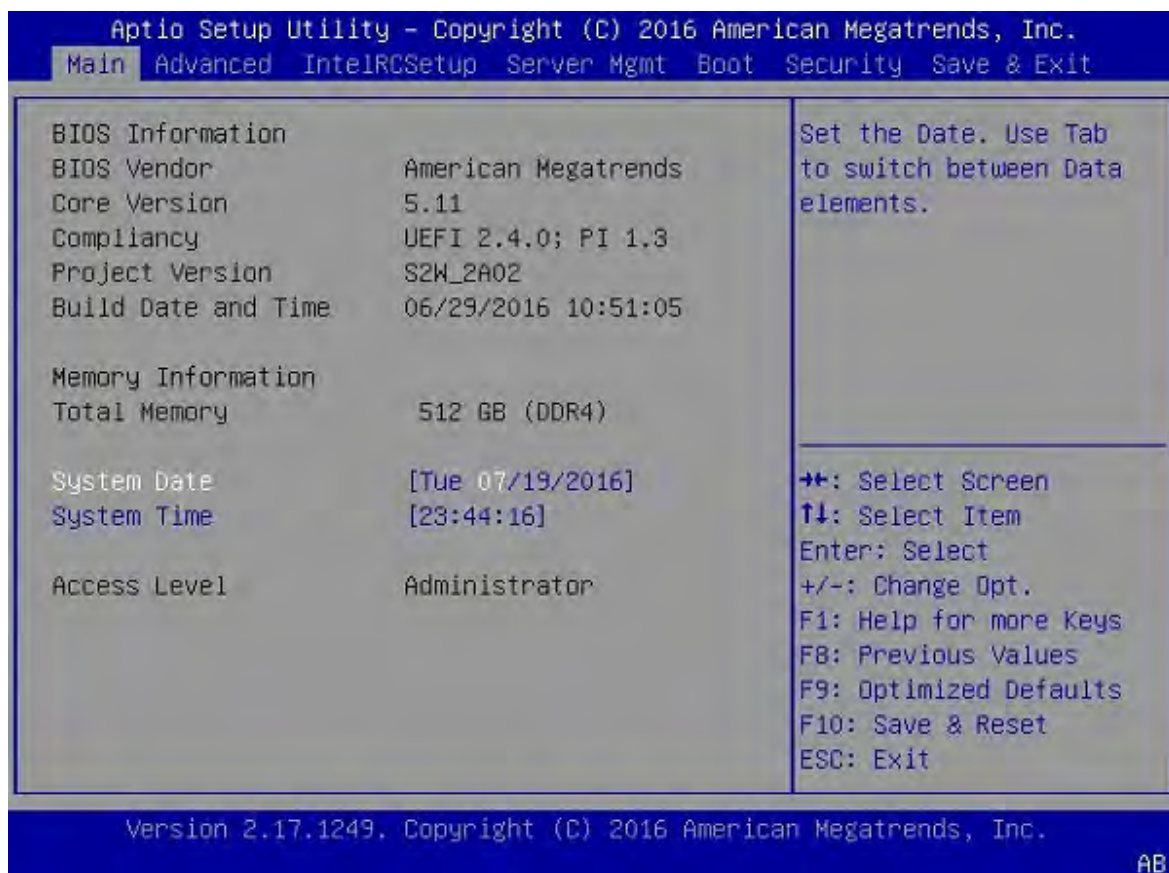
3. Reboot the DGX-1.
4. After the DGX-1 is back online, make sure that InfiniBand is the active link layer.

```
~$ ibv_devinfo |grep -e "hca_id\|link_layer"
hca_id: mlx5_3 link_layer: InfiniBand
hca_id: mlx5_2 link_layer: InfiniBand
hca_id: mlx5_1 link_layer: InfiniBand
hca_id: mlx5_0 link_layer: InfiniBand
```

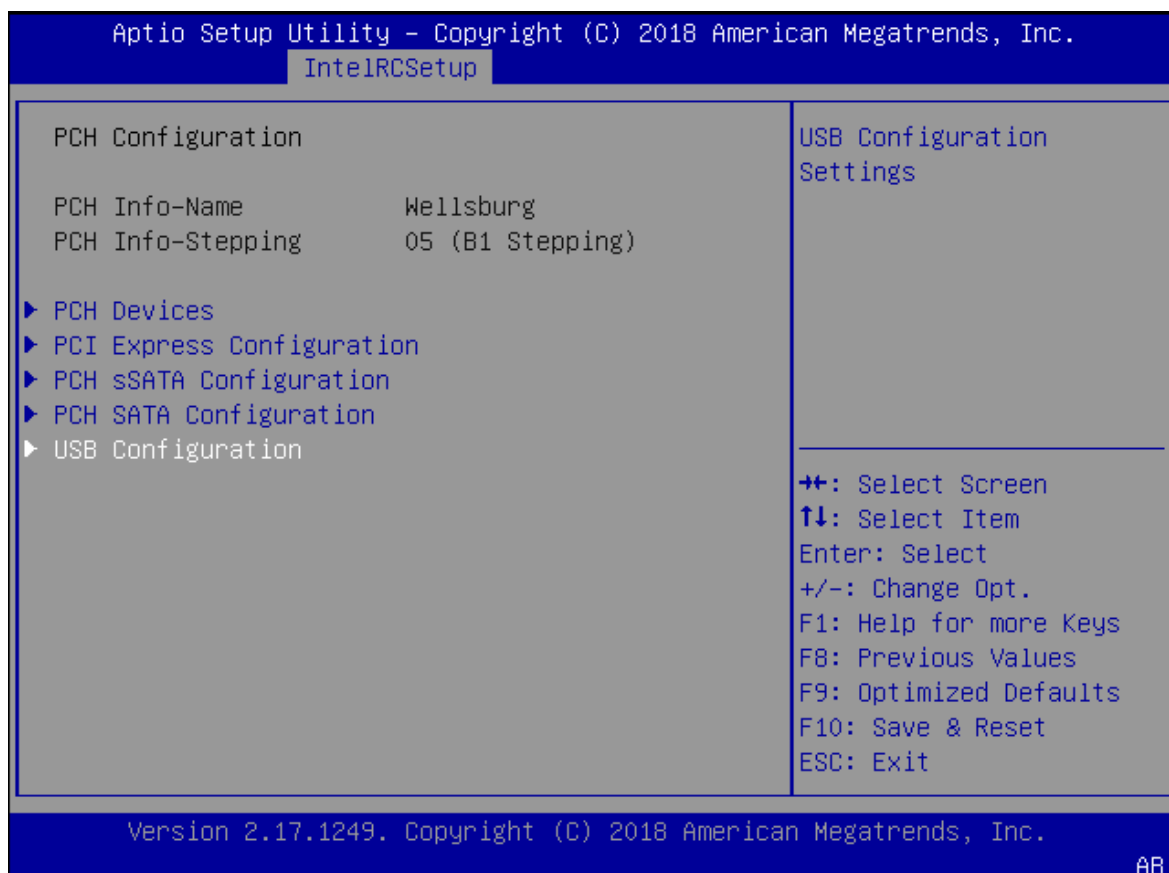
4.7. Enabling USB 3.0

There are two USB 3.0 ports on the DGX-1. However, they are set up as USB 2.0 by default in the system BIOS. To obtain USB 3.0 performance, you must enable USB 3.0 in the system BIOS as explained in this section.

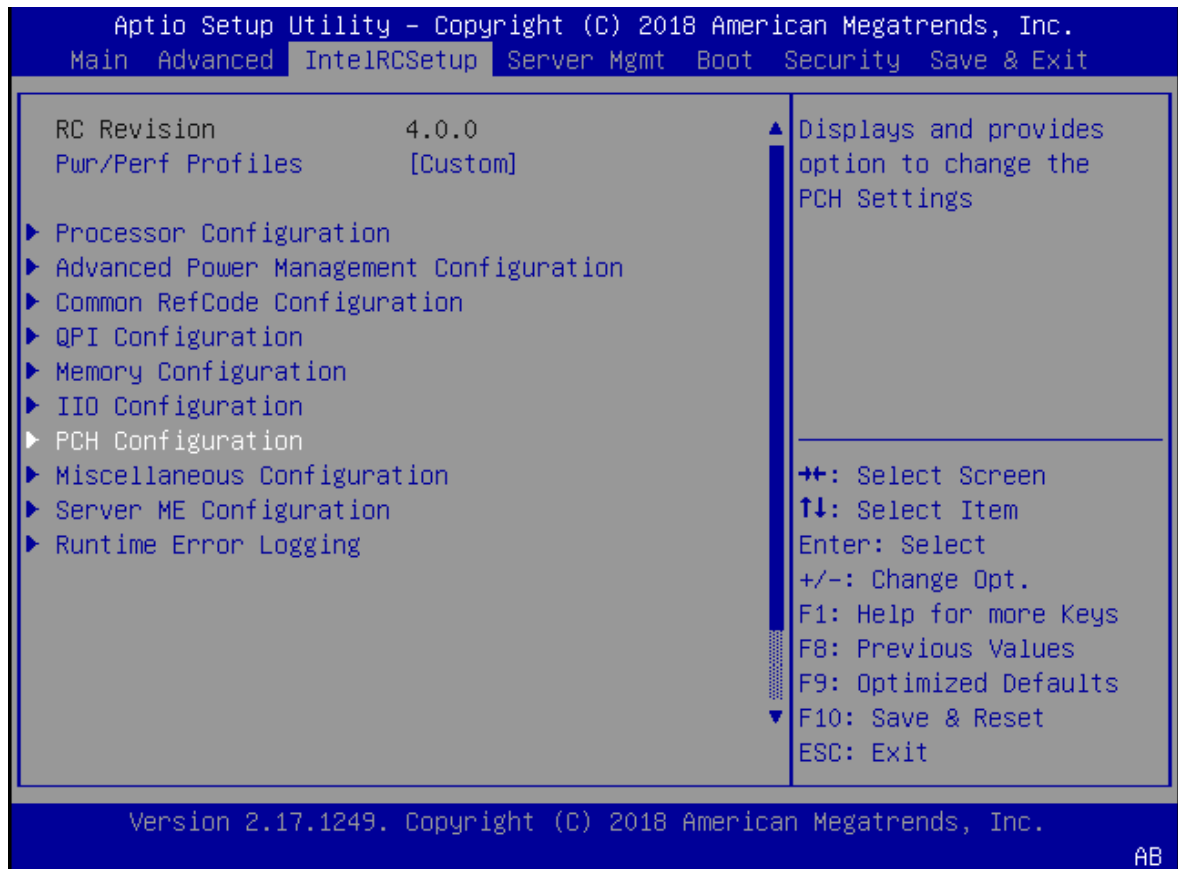
1. Connect a display (1024x768 or lower resolution) and keyboard to the DGX-1.
2. Turn the DGX-1 on or reboot.
3. At the NVIDIA logo boot screen, press **[F2]** or **[Del]** to enter the BIOS setup screen.



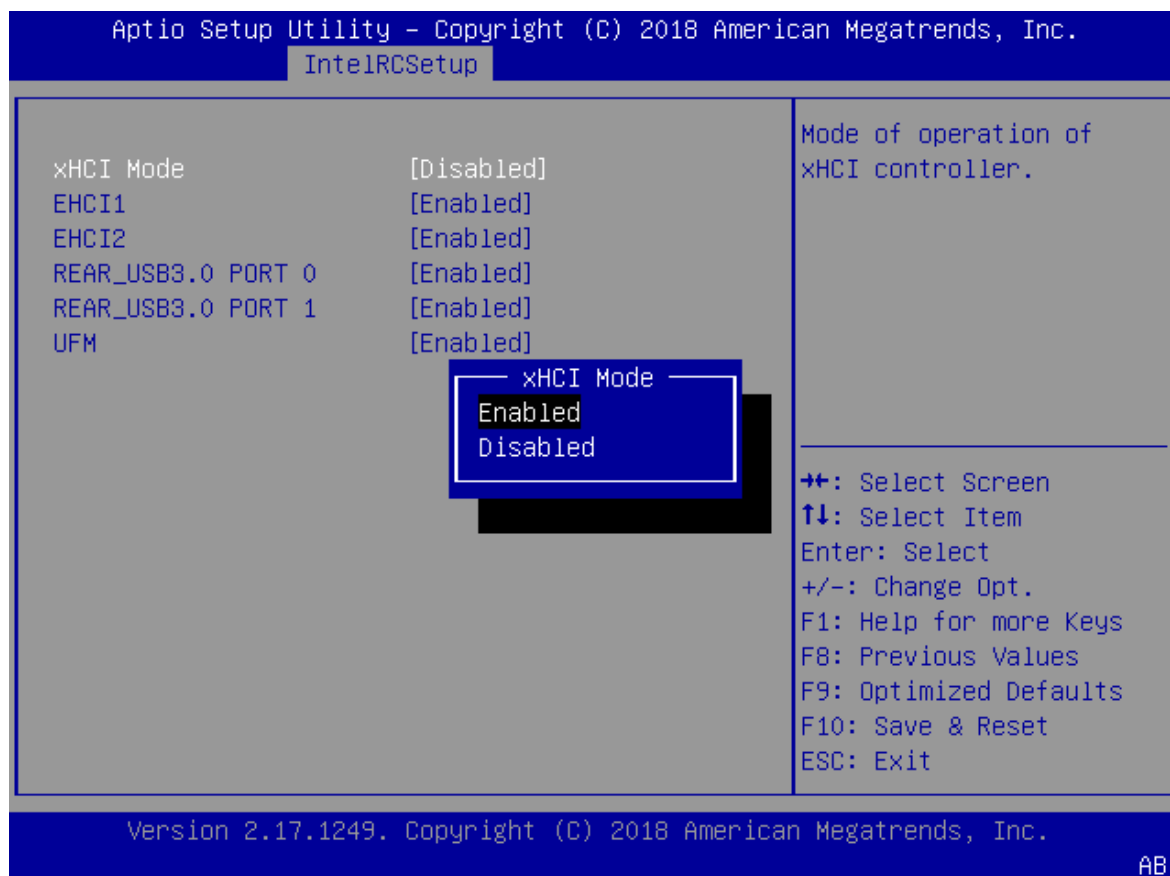
4. Select the **IntelRCSetup** tab from the top menu, then scroll down to **USB Configuration** and press **Enter**.



5. From the **IntelRCSetup** screen, scroll down to **PCH Configuration** and press **Enter**.



6. Scroll to **xHCI Mode** and press **Enter**, then select **Enabled** at the **xHCI Mode** dialog and press **Enter**.



7. Press **F10**, then select **Yes** at the confirmation pop-up and press **Enter**.
USB 3.0 is enabled upon system reboot.

Chapter 5. Security

The NVIDIA DGX-1 Server 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-1 Server is designed with a dedicated BMC Management Port and multiple Ethernet network ports.

When installing the DGX-1 Server in the data center, follow best practices as established by your organization to protect against unauthorized access.

Updating BMC Username and Password

If you have not already updated the default BMC username and passwords, NVIDIA recommends the BMC username and passwords be updated immediately to protect your system or nodes from unauthorized access. Follow the instructions in the section [Creating a Unique BMC Password for Remote Access](#) to update the default password.

Securing SNMP

Administrators must also confirm that the SNMP Read and Read/Write community strings are changed to a secure phrase or that you have disabled SNMP access. These settings can be found in the BMC dashboard under **Configuration > SNMP**.

Securing the BMC Port

NVIDIA recommends that the BMC port of the DGX-1 Server 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.

Chapter 6. Maintaining and Servicing the NVIDIA DGX-1

Be sure to familiarize yourself with the NVIDIA Terms & Conditions documents before attempting to perform any modification or repair to the DGX-1. These Terms & Conditions for the DGX-1 can be found through the [NVIDIA DGX Systems Support \(http://www.nvidia.com/object/dgxsystems-support.html\)](http://www.nvidia.com/object/dgxsystems-support.html) page.

6.1. Problem Resolution and Customer Care

Log on to the [NVIDIA Enterprise Support \(https://nvid.nvidia.com/dashboard/\)](https://nvid.nvidia.com/dashboard/) site for assistance with troubleshooting, diagnostics, or to report problems with your DGX-1.

Refer to [Customer Support for the NVIDIA DGX-1](#) for additional contact information.

Refer to [Submitting BMC Log Files](#) for instructions on how to obtain the BMC log files to assist in troubleshooting.

6.2. Restoring the DGX-1 Software Image

If the DGX-1 software image becomes corrupted or the OS SSD was replaced after a failure, restore the DGX-1 software image to its original factory condition from a pristine copy of the image.

The process for restoring the DGX-1 software image is as follows:

1. Obtain an ISO file that contains the image from NVIDIA Enterprise Support as explained in [Obtaining the DGX-1 Software ISO Image and Checksum File](#).
2. Restore the DGX-1 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](#)

6.2.1. Obtaining the DGX-1 Software ISO Image and Checksum File

To ensure that you restore the latest available version of the DGX-1 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 \(https://nvid.nvidia.com/dashboard/\)](https://nvid.nvidia.com/dashboard/) site.
2. Click the **Announcements** tab to locate the download links for the DGX-1 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.

6.2.2. Re-Imaging the System Remotely

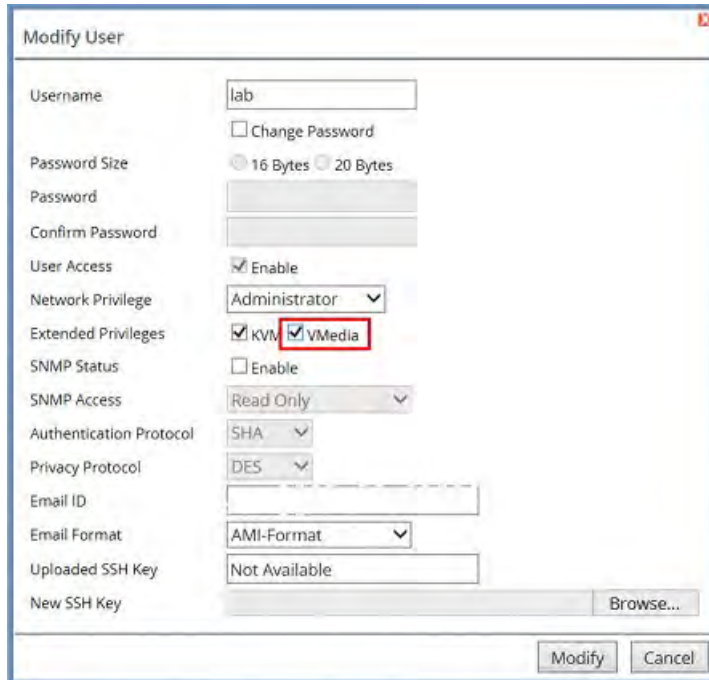
These instructions describe how to re-image the system remotely through the BMC. For information about how to restore the system locally, see [Re-Imaging the System From a USB Flash Drive](#).

Before re-imaging the system remotely, ensure that the correct DGX-1 software image is saved to your local disk. For more information, see [Obtaining the DGX-1 Software ISO Image and Checksum File](#).

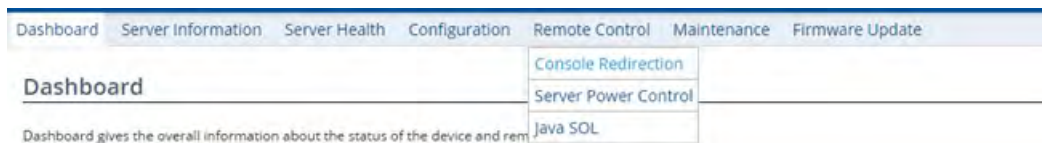


Note: The DGX-1 BMC uses JViewer to view the console. To use JViewer with Mac OS, first install XQuartz, available from <https://www.xquartz.org>.

1. Connect to the BMC and change user privileges.
 - a). Open a Java-enabled web browser within your LAN and go to `http://IPMI-IP-address/`, then log in.
Use Firefox or Internet Explorer. Google Chrome is not officially supported by the BMC.
 - b). From the top menu, click **Configuration** and then select **User Management**.
 - c). Select the user name that you created for the BMC, then click **Modify User**.
 - d). In the **Modify User** dialog, select the **VMedia** check box to add it to the extended privileges for the user, then click **Modify**.



2. Set up the ISO image as virtual media.
 - a). From the top menu, click **Remote Control** and select **Console Redirection**.



- b). Click **Java Console** to open the remote JViewer window.
Make sure pop-up blockers are disabled for this site.
 - c). From the JViewer top menu bar, click **Media** and then select **Virtual Media Wizard**.

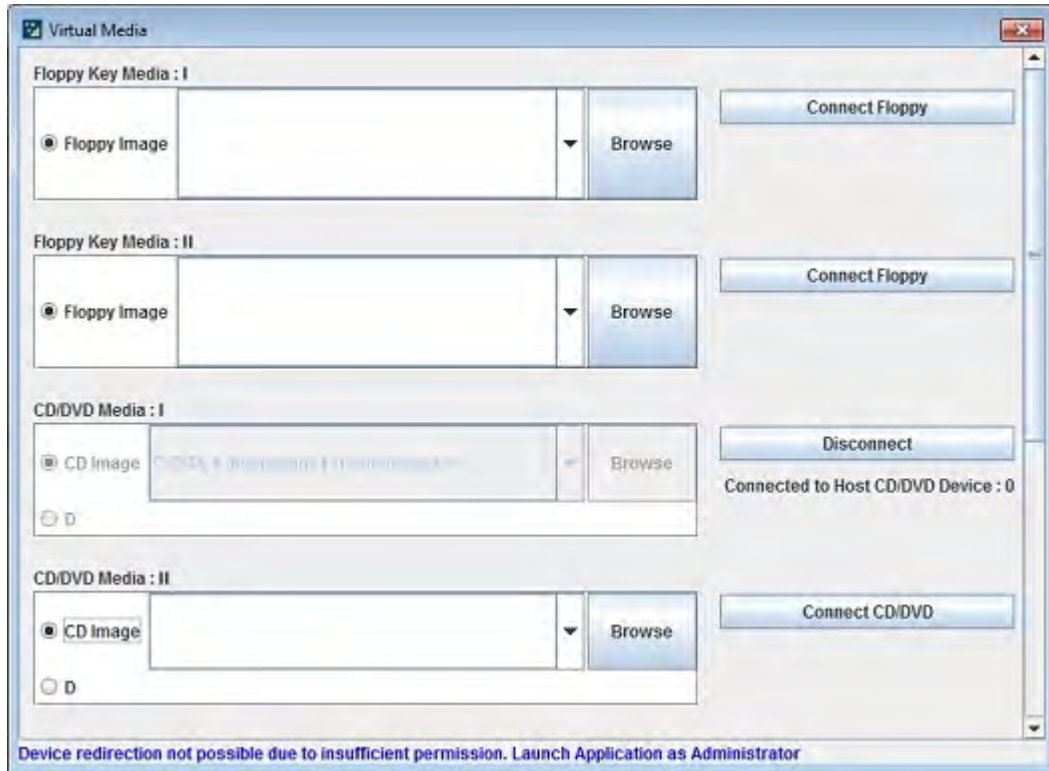


- d). From the **CD/DVD Media: I** section of the **Virtual Media** dialog, click **Browse** and then locate the re-image ISO file and click **Open**.

You can ignore the device redirection warning at the bottom of the **Virtual Media** wizard as it does not affect the ability to re-image the system.

- e). Click **Connect CD/DVD**, then click **OK** at the **Information** dialog.

The **Virtual Media** window shows that the ISO image is connected.



f). Close the window.

The CD ROM icon in the menu bar turns green to indicate that the ISO image is attached.



3. Reboot, install the image, and complete the DGX-1 setup.
 - a). From the top menu, click **Power** and then select **Reset Server**.



- b). Click **Yes** and then **OK** at the **Power Control** dialogs, then wait for the system to power down and then come back online.
 - c). 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.

d). Press **Enter**.

The DGX-1 will reboot from CDR0M0 1.00, and proceed to install the image. This can take approximately 15 minutes.



Note: The Mellanox InfiniBand driver installation may take up to 10 minutes.

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

Refer to [Setting Up the DGX-1](#) for the steps to take when booting up the DGX-1 for the first time after a fresh installation.

6.2.3. Creating a Bootable Installation Medium

After obtaining an ISO file that contains the DGX OS Server 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 \(https://help.ubuntu.com/community/BurningIsoHowto#Burning_the_ISO_on_to_a_DVD\)](https://help.ubuntu.com/community/BurningIsoHowto#Burning_the_ISO_on_to_a_DVD) on the Ubuntu Community Help Wiki.



Note: The ISO file that contains software image for some DGX OS Server releases is greater than the 4.7 GB capacity of a single-layer DVD-ROM. You cannot install these releases from a bootable DVD-ROM because installation of DGX OS Server from a dual-layer DVD-ROM is **not** supported. Check the size of the ISO file that contains the DGX OS Server software image before creating a bootable DVD-ROM.

6.2.3.1. Creating a Bootable USB Flash Drive by Using the `dd` Command

On a Linux system, you can use the `dd` (<http://manpages.ubuntu.com/manpages/bionic/en/man1/dd.1.html>) command to create a bootable USB flash drive that contains the DGX-1 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-1 software image is saved to your local disk. For more information, see [Obtaining the DGX-1 Software ISO Image and Checksum File](#).
- ▶ The USB flash drive meets these requirements:
 - ▶ The USB flash drive has a capacity of at least 8 GB.
 - ▶ The partition scheme on the USB flash drive is a GPT partition scheme for UEFI.

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` (<http://manpages.ubuntu.com/manpages/bionic/man8/lsblk.8.html>) command.

lsblk

You can identify the USB flash drive from its size, which is much smaller than the size of the SSDs in the DGX-1, and from the mount points of any partitions on the drive, which are under `/media`.

In the following example, the device name of the USB flash drive is `sde`.

```
~$ lsblk
NAME        MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sda           8:0    0   1.8T  0 disk
└─sda1        8:1    0   121M  0 part /boot/efi
└─sda2        8:2    0   1.8T  0 part /
sdb           8:16   0   1.8T  0 disk
└─sdb1        8:17   0   1.8T  0 part
sdc           8:32   0   1.8T  0 disk
sdd           8:48   0   1.8T  0 disk
sde           8:64   1    7.6G  0 disk
└─sde1        8:65   1    7.6G  0 part /media/deeplearner/DGXSTATION
~$
```

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.

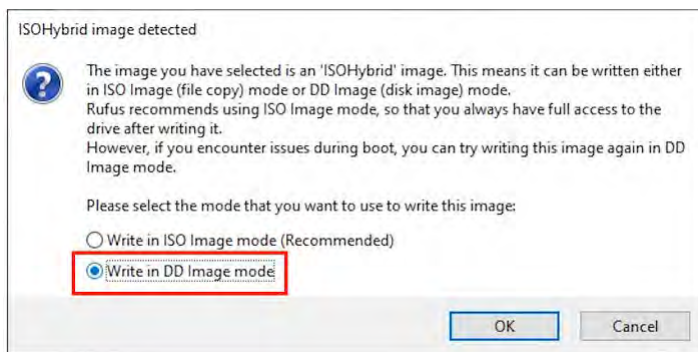
6.2.3.2. Creating a Bootable USB Flash Drive by Using Akeo Rufus

On a Windows system, you can use the [Akeo Reliable USB Formatting Utility \(Rufus\)](https://rufus.akeo.ie/) (<https://rufus.akeo.ie/>) to create a bootable USB flash drive that contains the DGX OS software image.

Ensure that the following prerequisites are met:

- ▶ The correct DGX OS software image is saved to your local disk. For more information, see [Obtaining the DGX-1 Software ISO Image and Checksum File](#).
- ▶ The USB flash drive has a capacity of at least 8 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 \(Rufus\)](https://rufus.akeo.ie/) (<https://rufus.akeo.ie/>).
3. In **Drive Properties**, select the following options:
 - a). In **Boot selection**, click **SELECT**, locate, and select the DGX OS software image.
 - b). In **Partition scheme**, select **GPT**.
 - c). In **Target System**, select **UEFI (non CSM)**.
4. In **Format Options**, select the following options:
 - a). In **File system**, select **NTFS**.
 - b). In **Cluster Size**, select **4096 bytes (Default)**.
5. 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.



6. Select **Write in ISO Image mode** and click **OK**.

6.2.4. 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-1 software image.



CAUTION: Installing the DGX-1 software image erases all data stored on the OS SSD. The partition where all users' documents, software settings, bookmarks, and other personal files are stored, resides on the OS SSD and will be erased. However, if you choose to install the DGX-1 software and preserve the RAID array contents, persistent data stored in the RAID array will not be unaffected.

1. Plug the USB flash drive containing the OS image into the DGX-1.
2. Connect a monitor and keyboard directly to the DGX-1.
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 one of the following options from the GRUB menu.
 - ▶ Install DGX OS <version>: Install DGX OS and reformat data RAID
 - ▶ Install DGX OS <version> Without Reformatting Data RAID
 - ▶ Advanced Installation Options: Select if you want to install with an encrypted root filesystem, then select one of the following options.
 - ▶ Install DGX OS <version> With Encrypted Root
 - ▶ Install DGX OS <version> With Encrypted Root and Without Reformatting Data RAID

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 the option that contains **Without Reformatting RAID**. See the section [Retaining the RAID Partition While Installing the OS](#) for more information.

6. Press **Enter**.

The DGX-1 will reboot and proceed to install the image. This can take more than 15 minutes.



Note: The Mellanox InfiniBand driver installation may take up to 10 minutes.

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

Refer to [Setting Up the DGX-1](#) for the steps to take when booting up the DGX-1 for the first time after a fresh installation.

6.2.5. Retaining the RAID Partition While Installing the OS

This information describes an installation option that is available starting with DGX OS Server 3.1.1.

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-1 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/cachefilesd`.
2. Uncomment the `/raid` line in `etc/fstab`.
3. Run the following:
 - a). Mount `/raid`.

```
sudo mount /raid
```

- b). Reload the systemd manager configuration.

```
systemctl daemon-reload
```

- c). Start the cache daemon.

```
systemctl start cachefilesd.server
```

These changes are preserved across system reboots.

6.3. Updating the System BIOS

You can update the system BIOS remotely through the BMC. Before updating the system BIOS, the system must be turned off through the BMC according to the instructions in this section.



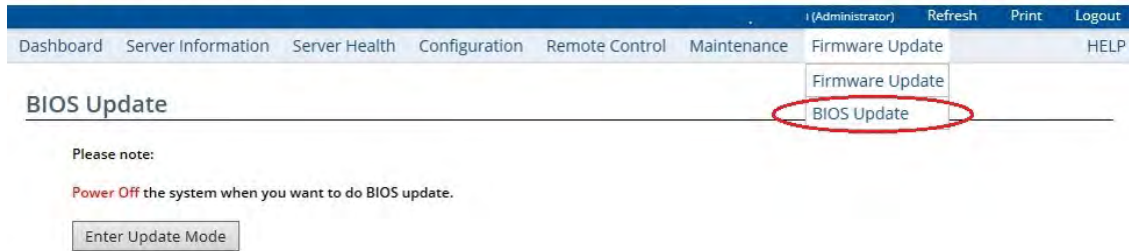
Note: NVIDIA recommends updating the SBIOS using an NVIDIA-provided Docker container. For instructions, see the section [Using the Firmware Update Container](#).

1. Obtain the BIOS image.
 - a). Log on to [NVIDIA Enterprise Support \(https://nvid.nvidia.com/dashboard/\)](https://nvid.nvidia.com/dashboard/) and click the Announcements tab to locate the DGX-1 software image archive.
 - b). Download the image archive and then extract the .bin file.
2. Log on to the BMC and shut down the DGX-1.
 - a). Open a Java-enabled web browser within your LAN and go to `http://<IPMI-IP-address>/`, then log in.
Use Firefox or Internet Explorer. Google Chrome is not officially supported by the BMC.
 - b). From the top menu, click **Remote Control** and then select **Server Power Control**.
 - c). At the *Power Control and Status* screen, select the **Power Off Server - Orderly Shutdown** option, then click **Perform Action**.

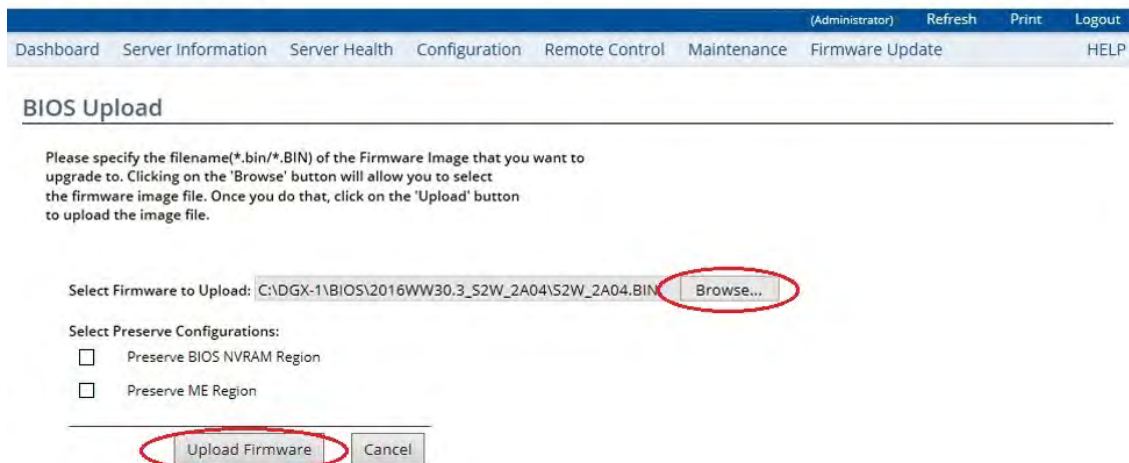


You can verify that the DGX-1 is shut down by noting that all the *Power Control and Status* options are grayed out except for the *Power On Server* option.

3. Update the system BIOS.
 - a). From the top menu, click **Firmware Update**, select **BIOS Update**, and then click **Enter Update Mode**.



- b). Click **OK** at the *Are you sure to enter update mode?* dialog.
- c). From the *BIOS Upload* screen, click **Browse** at the *Select Firmware to Upload* step, then navigate the explorer windows to locate the file you downloaded and select it.



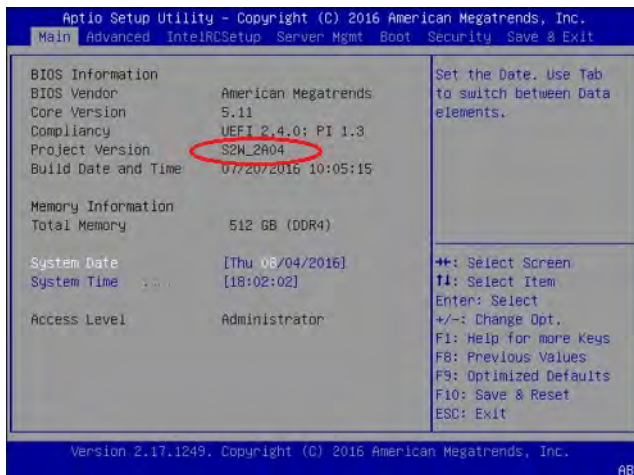
- d). Be sure all the check boxes under *Select Preserve Configuration* are cleared.
This ensures that the BIOS reverts to its fail-safe default settings for a reliable update.
- e). Click **Upload Firmware** to start the process of installing the updated BIOS.
You are asked to wait while the image is verified.
- f). Click **OK** at the *Proceed?* dialog to start the actual upgrade process.
The *BIOS Flash Status* screen shows the upgrade progress, which should take a couple of minutes to complete.



Note: Do not interrupt the upgrade process once it has started.

4. After the upgrade process has completed, you can use the top menu to turn the system back on.
 - a). From the top menu, click **Remote Control** and then select **Server Power Control**.
 - b). Select the **Power On Server** option, and then click **Perform Action**.

5. To verify that the BIOS was updated with the proper file, press **[F2]** or **[Del]** to enter the BIOS setup screen when the system reboots, then compare the Project Version with the update filename.



6.4. Updating the BMC

You can update the BMC remotely using the IPMI port. Applications can be left running. Power must be left on.



Note: NVIDIA recommends updating the BMC using an NVIDIA-provided Docker container. For instructions, see the section [Using the Firmware Update Container](#).

1. Obtain the BMC image.
 - a). Log on to [NVIDIA Enterprise Support \(https://nvid.nvidia.com/dashboard/\)](https://nvid.nvidia.com/dashboard/) and click the Announcements tab to locate the DGX-1 software image archive.
 - b). Download the image file.
2. Open a Java-enabled web browser within your LAN and go to `http://<IPMI IP address>/`, then log in to the BMC.

Use Firefox or Internet Explorer. Google Chrome is not officially supported by the BMC.

3. If you're using DHCP and choose not to preserve the network configuration, then obtain the MAC address for the BMC.

If the BMC is connected to a network via DHCP, the IP address could change after the update. Follow these substeps to obtain the MAC address in order to connect to the BMC after the update, in case the IP address changes. You can skip these steps if a static IP is used.

- a). From the top menu, click **Configuration** and then select **Network**.
 - b). Note the MAC address.
4. From the top menu, click **Firmware Update** and then select **Firmware Update** from the drop-down menu.

- Click **Enter Preserve Configuration**, then set the IPMI and Network preserve status to *Preserve* and all others to *Overwrite*.

Firmware Update

Upgrade firmware of the device. Press 'Enter Update Mode' to put the device in update mode.

The protocol information to be used for firmware image transfer during this update is as follows. To configure, choose 'Protocol Configuration' under Firmware Update menu.

Protocol Type : HTTP/HTTPS

WARNING: Please note that after entering the update mode, the widgets, other web pages and services will not work. All the open widgets will be automatically closed. If the upgradation is cancelled in the middle of the wizard, the device will be reset.

☐ Preserve all Configuration. This will preserve all the configuration settings during the firmware update - irrespective of the individual items marked as preserve/overwrite in the table below. All configuration items below will be preserved as default during the restore configuration operation. Click "Enter Preserve Configuration" to modify the Preserve status settings.

#	Preserve Configuration Item	Preserve Status
1	SDR	Overwrite
2	FRIJ	Overwrite
3	SEL	Overwrite
4	IPMI	Preserve
5	Network	Preserve
6	NTP	Overwrite
7	SNMP	Overwrite
8	SSH	Overwrite
9	KVM	Overwrite
10	Authentication	Overwrite

Note: Be sure to set IPMI to *Preserve* in order to preserve your BMC login credentials. If you fail to do this, the BMC username/password will be set to qct.admin/qct.admin. If this happens, then be sure to enter the BMC dashboard and go to Configuration->Users to add a new user account and disable the qct.admin account after updating the BMC.

- If necessary, click **Firmware Update** again from the top menu and then select **Firmware Update** from the drop-down menu to return to the Firmware Update page.

- Click **Enter Update Mode**, then click **OK** at the confirmation dialog.

After entering Update Mode, aborting the operation or even resizing the browser windows will terminate the session and reset the BMC. If this happens, you will need to close and then reopen the browser to initiate a new session. You may need to wait several minutes for the BMC to reset.

- At the *Upload Firmware* prompt, click **Browse** to locate and select the firmware image file.

Upload Firmware

Please select the firmware image to flash

e\S2W_BMC_v209\S2W_BMC_v209\S2Wv209.ima_enc

Select the encrypted file (the file with the "_enc" suffix on the file extension), as the BMC requires the firmware image to be encrypted.

- Click **Upload** to transfer the image to the BMC.
- At the *Select Based Firmware Update* prompt, select **Full Flash** and then click **Proceed**.

Section Based Firmware Update

All the module section versions in the existing image and uploaded image is same.

☐ Version Compare Flash ☒ Full Flash

#	Section Name	Existing Version	Uploaded Version	Upgradable/Non-Upgradable
1	boot	2.1.000000	2.1.000000	<input type="checkbox"/>
2	root	2.1.000000	2.1.000000	<input type="checkbox"/>
3	osimage	2.1.000000	2.1.000000	<input type="checkbox"/>
4	www	2.1.000000	2.1.000000	<input type="checkbox"/>
5	ast2400e	2.9.0	2.9.0	<input type="checkbox"/>



Note: IMPORTANT: Do not shut off power to the DGX-1 while updating the BMC. If the BMC update fails, keep the DGX-1 powered on and booted, and then contact NVIDIA Enterprise Support.

- ▶ When the BMC firmware update is completed, the BMC resets and the remote session terminates.
- ▶ To initiate a new BMC session, close and then reopen the browser.
- ▶ The BMC can take as much as 10 minutes to reset itself. During this time, the BMC will be unresponsive.

6.5. Updating Component Firmware Using the Firmware Update Container

This section provides instructions for updating firmware for the NVIDIA® DGX server BIOS, BMC, and power supplies using a Docker container. This is an alternate method to the instructions provided in the sections [Updating the BIOS](#) and [Updating the BMC](#).

General Firmware Update Guidelines

- ▶ Before updating the firmware, do the following to prevent corrupting the firmware due to a system crash or disruption to the update process.
 - ▶ Ensure the system is healthy
 - ▶ Stop system activities
 - ▶ Do not terminate the firmware update console while updating the firmware.
- Component firmware corruption may occur if the update process is interrupted.
- ▶ Certain components, such as the system BIOS, requires a system reboot for the new firmware to take effect.

Reboot the system if prompted.

- In the event of a firmware update failure, gather the follow logs for failure analysis:
 - /var/log/syslog
 - /var/log/nvidia-fw.log

6.5.1. Obtaining the Firmware Update Container

The NVIDIA DGX System Firmware Update utility is provided in a tar ball file and also as a .run file. Copy the files to the DGX system, then update the firmware using one of the following three methods:

- **NVSM** provides convenient commands to update the firmware using the firmware update container
- **Using Docker** to run the firmware update container
- **Using the .run file** which is a self-extracting package embedding the firmware update container tarball

Using NVSM

The NVIDIA DGX-1 system software includes Docker software required to run the container.

1. Copy the tar ball to a location on the DGX system.
2. From the directory where you copied the tarball file, enter the following command to load the container image.

```
$ sudo docker load -i nvfw-dgx1_19.10.7.tar.gz
```

3. To verify that the container image is loaded, enter the following.

```
$ sudo docker images
```

```
REPOSITORY    TAG
nvfw-dgx1     19.10.7
```

4. Using NVSM interactive mode, enter the firmware update module.

```
$ sudo nvsm
nvsm-> cd systems/localhost/firmware/install
```

5. Set the flags corresponding to the action you want to take.

```
$ nvsm(/system/localhost/firmware/install)-> set Flags=<option>
```

See the rest of this chapter for details on individual flags and arguments.

6. Run the command.

```
$ nvsm(/system/localhost/firmware/install)-> start
```

Using the Tar Ball File

The NVIDIA DGX-1 system software includes Docker software required to run the container.

1. Copy the tar ball to a location on the DGX system.

2. From the directory where you copied the tarball file, enter the following command to load the container image.

```
$ sudo docker load -i nvfw-dgx1_19.10.7.tar.gz
```

3. To verify that the container image is loaded, enter the following.

```
$ sudo docker images
```

```
REPOSITORY    TAG
nvfw-dgx1     19.10.7
```

4. Use the following syntax to run the container image.

```
$ sudo docker run --rm --privileged -ti -v /:/hostfs nvfw-dgx1:19.10.7 <command>
<[arg1] [arg2] ... [argn]
```

See the rest of this chapter for details on individual flags and arguments.

Using the .run File

Beginning with the firmware container version 19.04.1, a .run file is also available to run the firmware update container. The .run file is a self-extracting package embedding the firmware update container tarball. Using the .run file requires DGX OS Server 4.0.5 or later.

After obtaining the .run file, make the file executable.

```
$ chmod +x nvfw-dgx1_19.10.7.run
```

Use the following syntax to run the container image.

```
$ sudo nvfw-dgx1_19.10.7.run <command> <[arg1] [arg2] ... [argn]
```

See the rest of this chapter for details on individual flags and arguments.

6.5.2. Querying the Firmware Manifest

The manifest displays a listing of firmware components embedded in the containers that are qualified by NVIDIA.

To query the firmware manifest, enter the following:

```
# sudo docker run --rm --privileged -v /:/hostfs <container-repository:tag> show_fw_manifest
```

6.5.3. Querying the Currently Installed Firmware Versions

Display the onboard firmware version level of each component supported by the container. The output will show which component firmware is up to date, or whether it needs to be updated to the firmware level listed in the manifest.

To query the version information, enter the following.

```
# sudo docker run --privileged -v /:/hostfs <container-repository:tag> show_version
```

The output shows the onboard version, the version in the manifest, and whether the firmware is up-to-date.

6.5.4. Updating the Firmware

You can either update all the down-level firmware components at one time, or update just one or more components.

Command Syntax

```
sudo docker run [-e AUTO=1] --privileged -ti -v /:/hostfs <container-repository:tag>
update_fw [-f] <target>
```

Where *<target>* specifies the hardware to update, and is either

all

to update all firmware components (SBIOS, BMC, and PSU)

or one or more of the following:

SBIOS

to update the SBIOS

BMC

to update the BMC firmware

PSU

to update the power supply units' firmware



Note: Other components may be supported beyond those listed here. Query the firmware manifest to see all the components supported by the container.

The command will scan the specified firmware components and update any that are down-level.

See the section [Additional Options](#) for an explanation of the `[-e AUTO=1]` and `[-f]` options.

6.5.4.1. Updating All Firmware Components

The following instructions are an example of attempting to update all the firmware components using the container `nvfw-dgx1:19.10.7`. In this example, only the SBIOS and BMC require an update.

1. Enter the following.

```
$ sudo docker run --rm --privileged -ti -v /:/hostfs nvfw-dgx1:19.10.7 update_fw all
```

The container will scan the components and then prompt for confirmation before starting the update.

```
Following components will be updated with new firmware version:
```

```
SBIOS
```

```
BMC
```

```
IMPORTANT: Firmware update is disruptive and may require system reboot.
```

```
Stop system activities before performing the update.
```

```
Ok to proceed with firmware update? <Y/N>
```

2. Press **Y** to proceed.

When the update completes successfully, the following message is displayed.

```
Firmware update completed Component: SBIOS, update status: success, reboot
required: yes
Component: BMC, update status: success, new version: 3.20.30
```

3. If directed by the update message, reboot the system.

6.5.4.2. Updating Specific Firmware Components

The following is an example of updating the PSU firmware using the container `nvfw-dgx1:19.10.7`.

1. Enter the following.

```
$ sudo docker run --rm --privileged -ti -v /:/hostfs nvfw-dgx1:19.10.7 update_fw PSU
```

The container will scan the components and then prompt for confirmation before starting the update.

```
Following components will be updated with new firmware version:
```

```
PSU 1
PSU 2
PSU 3
PSU 4
```

```
IMPORTANT: Firmware update is disruptive and may require system reboot.
Stop system activities before performing the update.
Ok to proceed with firmware update? <Y/N>
```

2. Press **Y** to proceed.

When the update completes successfully, the following message is displayed.

```
Firmware update completed
Component: PSU, update status: success, new version: 00.03.07
```

3. If directed by the update message, reboot the system.

You can also update a subset of all the components. For example, to update both the power supply firmware and the system BIOS, enter the following:

```
$ sudo docker run --rm --privileged -ti -v /:/hostfs nvfw-dgx1:19.10.7 update_fw PSU SBIOS
```

6.5.4.3. Additional Options

Forcing the Firmware Update

To update the firmware regardless of whether it is down-level, use the `-f` option as follows.

```
$ sudo docker run --rm --privileged -ti -v /:/hostfs <container-repository:tag> update_fw -f
<target>
```

The container will not check the onboard versions against the manifest.

Updating the Firmware Non-interactively

The standard way to run the container is interactively (`-ti` option). The container will prompt you to confirm before initiating the update.

To update the firmware without encountering the prompt, use the `-e AUTO=1` and `-t` options as follows.

```
$ sudo docker run -e AUTO=1 --rm --privileged -ti -v /:/hostfs <container-repository:tag>
update_fw <target>
```

6.5.4.4. Command Summary

- Show the manifest.

```
$ sudo docker run --rm --privileged -v /:/hostfs <container-repository:tag>
show_fw_manifest
```

- Show version information.

```
$ sudo docker run --rm --privileged -v /:/hostfs <container-repository:tag> show_version
```

- Check the onboard firmware against the manifest and update any down-level firmware.

```
$ sudo docker run --rm --privileged -ti -v /:/hostfs <container-repository:tag> update_fw
<target>
```

- Bypass the version check and update the firmware.

```
$ sudo docker run --rm --privileged -ti -v /:/hostfs <container-repository:tag> update_fw
-f <target>
```

- Update the firmware in non-interactive mode.

```
$ sudo docker run --rm -e auto=1 --privileged -t -v /:/hostfs <container-repository:tag>
update_fw <target>
```

6.5.5. Removing the Container

Remove the container and image from the DGX server when it is no longer needed. To remove the container and image, enter the following:

```
$ sudo docker rmi -f <container-repository>
```

In this case, specify only the container repository and not the tag.

6.5.6. Using the .run File

Beginning with the firmware container version 19.04.1, a .run file is also available to run the firmware update container. The .run file is a self-extracting package embedding the firmware update container tarball. Using the .run file requires DGX OS Server 4.0.5 or later.

1. Before using, make the file executable.

```
$ chmod +x /<run-file-name>.run
```

2. Run the file.

```
$ sudo ./<run-file-name>.run
```

This command is the same as running the container with the update_fw all option.

The .run file accepts the same options that are used when running the container.

Examples:

Show the manifest.

```
$ sudo ./<run-file-name>.run show_fw_manifest
```

Show version information.

```
$ sudo ./<run-file-name>.run show_version
```

Check the onboard firmware against the manifest and update any down-level firmware.

```
$ sudo ./<run-file-name>.run update_fw <target>
```

Bypass the version check and update the firmware.

```
$ sudo ./<run-file-name>.run update_fw -f <target>
```

6.6. Replacing the System and Components

Be sure to familiarize yourself with the NVIDIA Terms & Conditions documents before attempting to perform any modification or repair to the DGX-1. These Terms & Conditions for the DGX-1 can be found through the [NVIDIA DGX Systems Support \(http://www.nvidia.com/object/dgxsystems-support.html\)](http://www.nvidia.com/object/dgxsystems-support.html) page.

Contact NVIDIA Enterprise Support to obtain an RMA number for any system or component that needs to be returned for repair or replacement. When replacing a component, use **only** the replacement supplied to you by NVIDIA.

The following components are customer-replaceable:

- ▶ Solid State Drives (SSDs)
- ▶ Power Supplies
- ▶ Fan Modules
- ▶ DIMMs
- ▶ Battery

Return failed high-value components to NVIDIA. You do not need to return low-cost items such as batteries, power supplies, and fans.

6.6.1. Replacing the System

When returning a DGX-1 under RMA, consider the following points.

SSDs

If necessary, you can remove and keep the SSDs prior to shipping the system back for replacement. If you already received a replacement system and you want to keep the original SSDs, install the new SSDs into the defective system when shipping it back.

Bezel

Be sure to include the bezel when returning the system.

6.6.2. Replacing an SSD

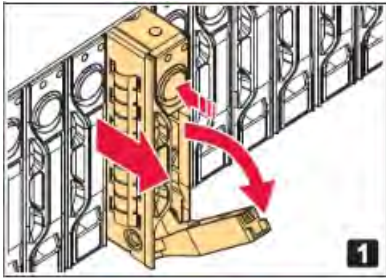
Access the SSDs from the front of the DGX-1. You can hot swap the SSDs as follows:

1. If not already removed, remove the bezel by grasping the bezel by the side handles and then pulling the bezel straight off the front of the DGX-1.

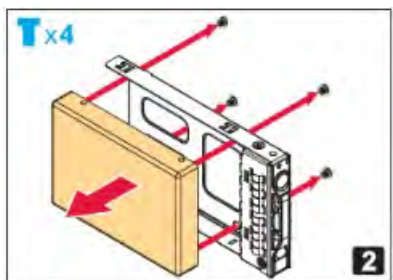


Note: CAUTION: Be careful not to accidentally press the power button that is on the right edge of the DGX-1 when removing or installing the bezel.

2. Locate the SSD that you want to replace, then press the round button at the top edge to release the latch.
3. Pull the latch down and then out to unseat the SSD assembly.



4. Continue pulling the SSD assembly to completely remove it from the unit.
5. Using a phillips screwdriver, remove the four screws attaching the SSD to the hot-swap tray.



6. Save the screws for the replacement.
7. Mount the replacement SSD to the hot-swap tray using the four screws.
Make sure that the connector is on the open edge side of the tray.
8. With the round button at the top, insert the assembly into the appropriate bay, then push the assembly all the way in.
9. Press the latch against the assembly to completely seat the assembly.
10. Reattach the bezel.

With the bezel positioned so that the NVIDIA logo is visible from the front and is on the left-hand side, line up the pins near the corners of the DGX-1 with the holes in back

of the bezel, then gently press the bezel against the DGX-1. The bezel is held in place magnetically.



Note: CAUTION: Be careful not to accidentally press the power button that is on the right edge of the DGX-1 when removing or installing the bezel.

6.6.3. Recreating the Virtual Drives

After you have replaced the OS SSD, with or without any of the cache SSDs, you need to recreate the virtual drives and then re-image the system in order to recreate the partitions on all the virtual drives.

The following is an overview of the process:

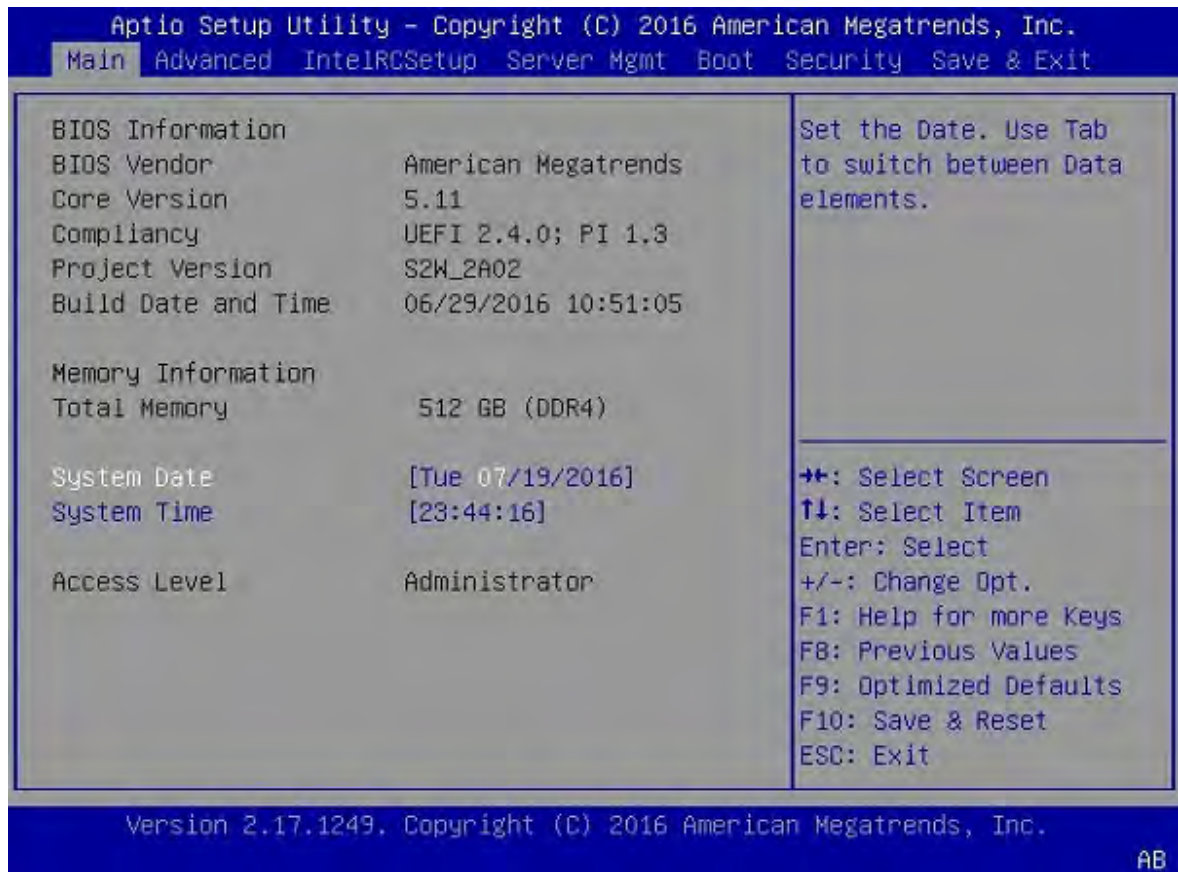
1. Clear the drive group configuration
2. Recreate the OS Virtual Drive
3. Recreate the Cache Virtual Drive
4. Re-image the System

These instructions apply only if you have replaced the OS SSD, with or without one or more of the cache SSDs. If you have replaced only one or more of the cache SSDs, and not the OS SSD, then follow the instructions in the section [Recreating the RAID 0 Array](#)

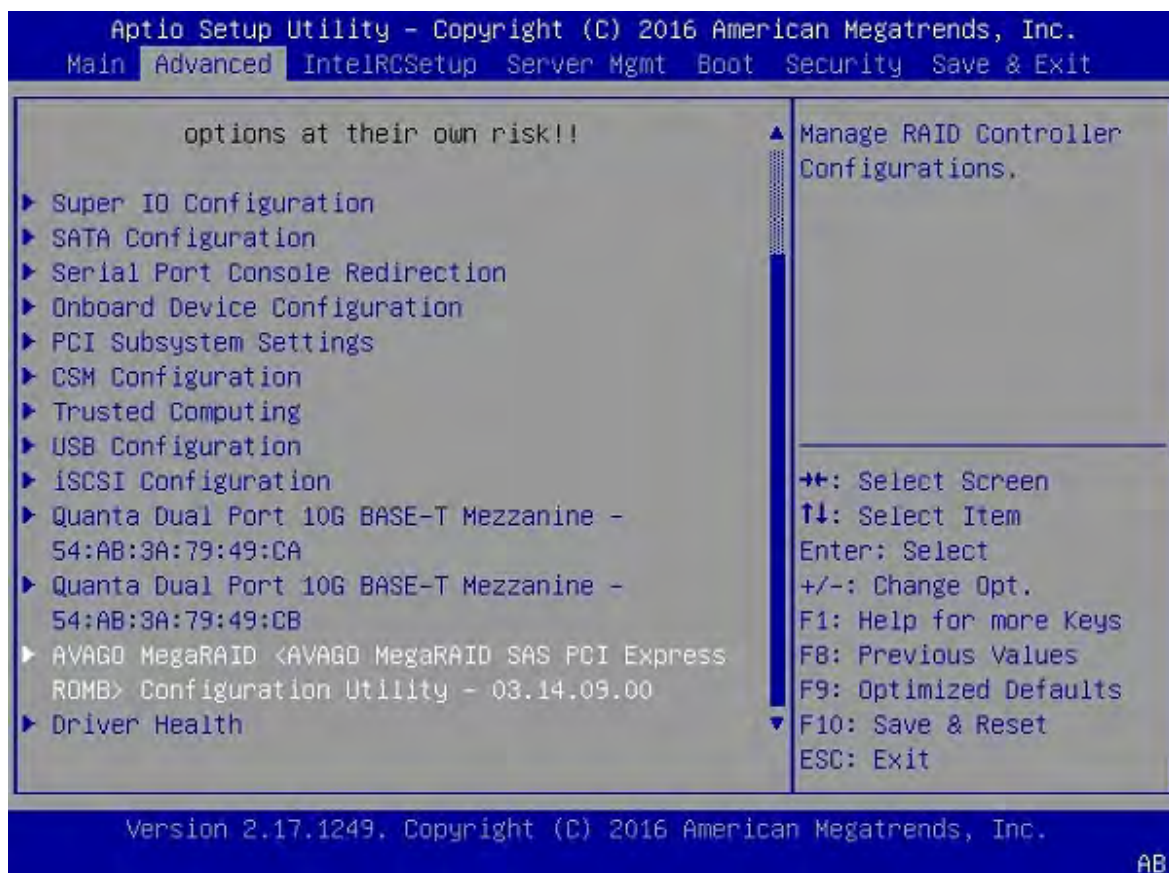
6.6.3.1. Access the BIOS Setup Utility

RAID configuration is accomplished through the BIOS setup utility.

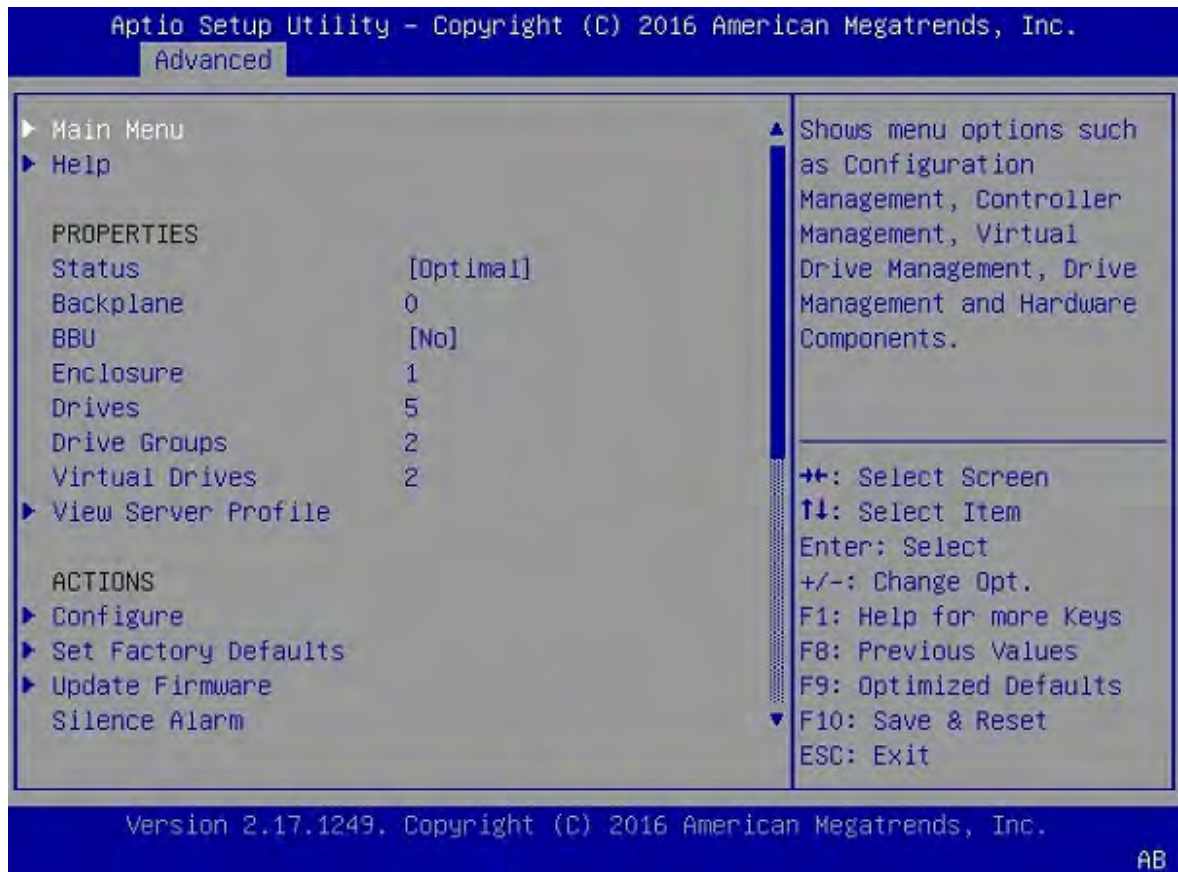
1. Connect a display (1024x768 or lower resolution) and keyboard to the DGX-1.
2. Turn the DGX-1 on or reboot.
3. At the NVIDIA logo boot screen, press **[F2]** or **[Del]** to enter the BIOS setup screen.



4. Select the **Advanced** tab from the top menu and then Scroll down and select the **MegaRAID Configuration Utility**.



The RAID Configuration menu appears.

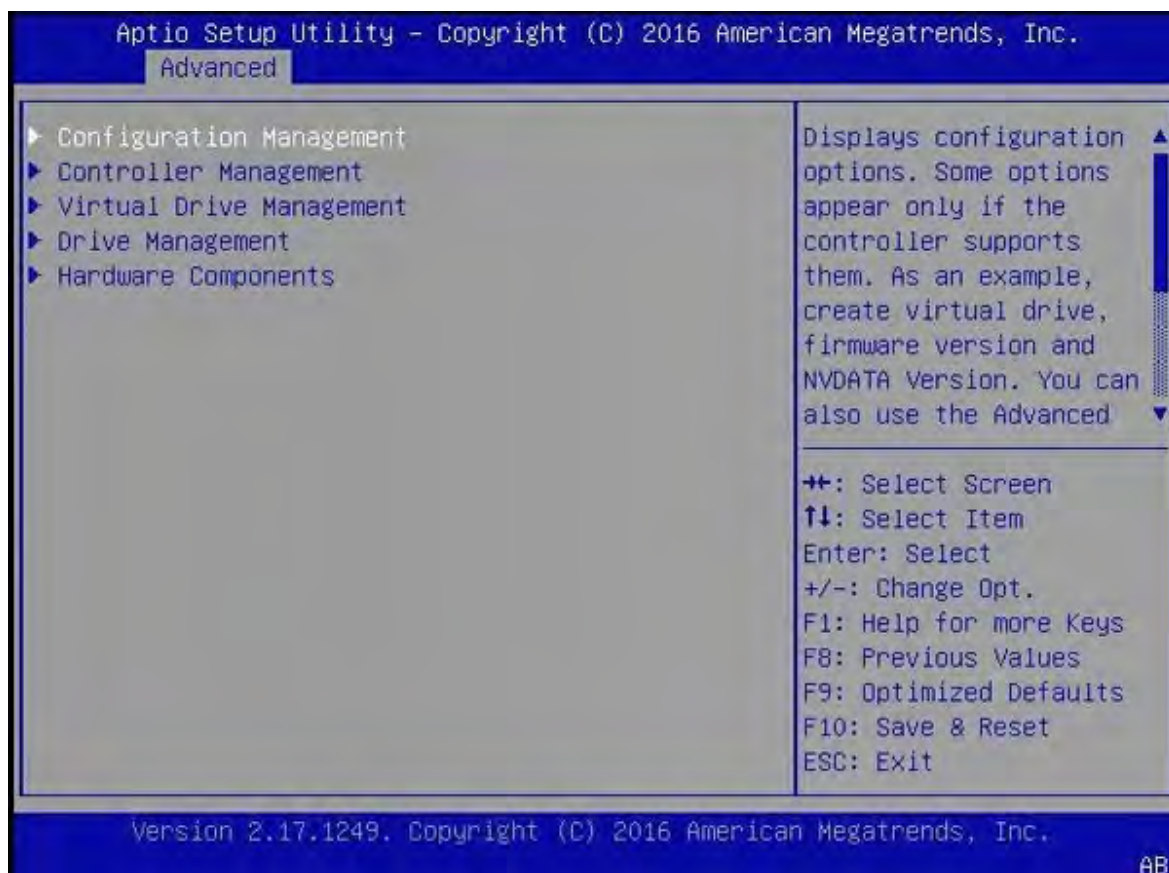


If you replaced the OS drive, follow the instructions in the section [Clear the Drive Group Configuration](#).

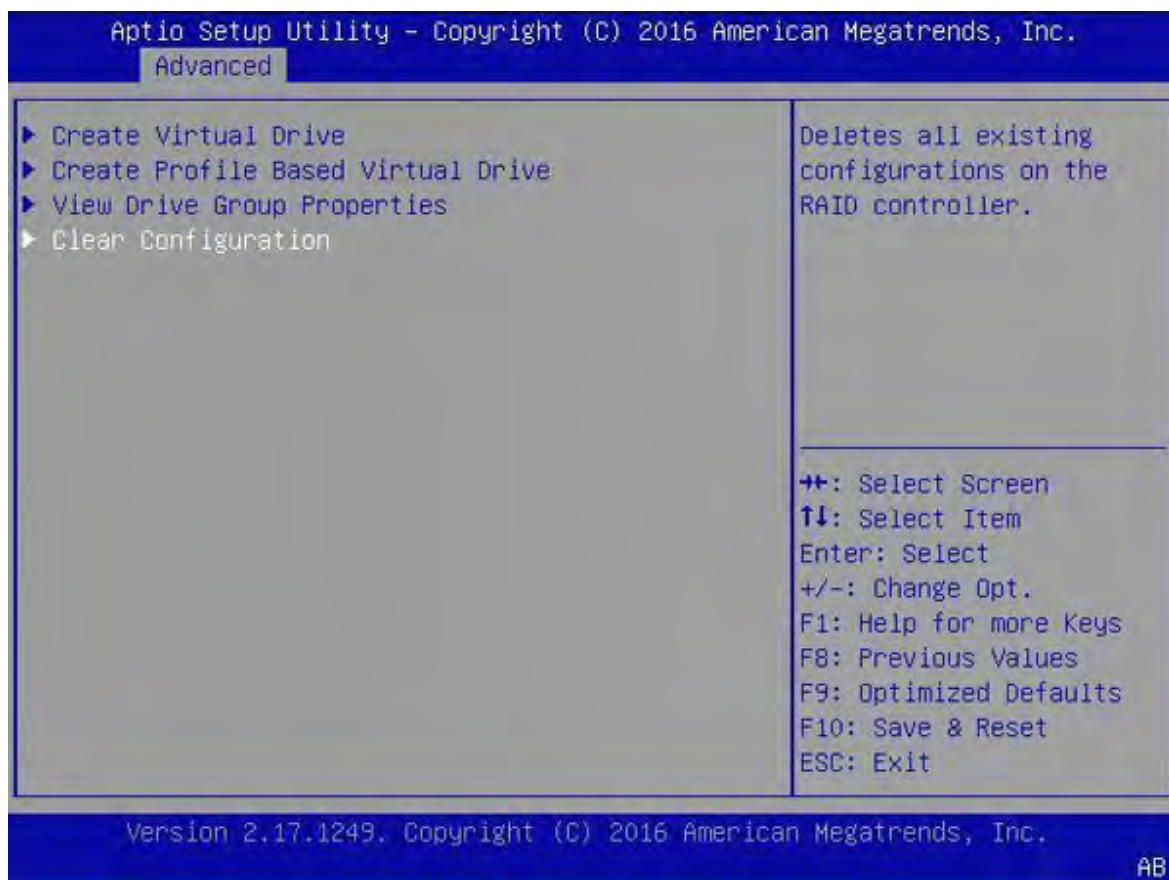
6.6.3.2. Clear the Drive Group Configuration

These instructions apply when you have replaced the OS drive.

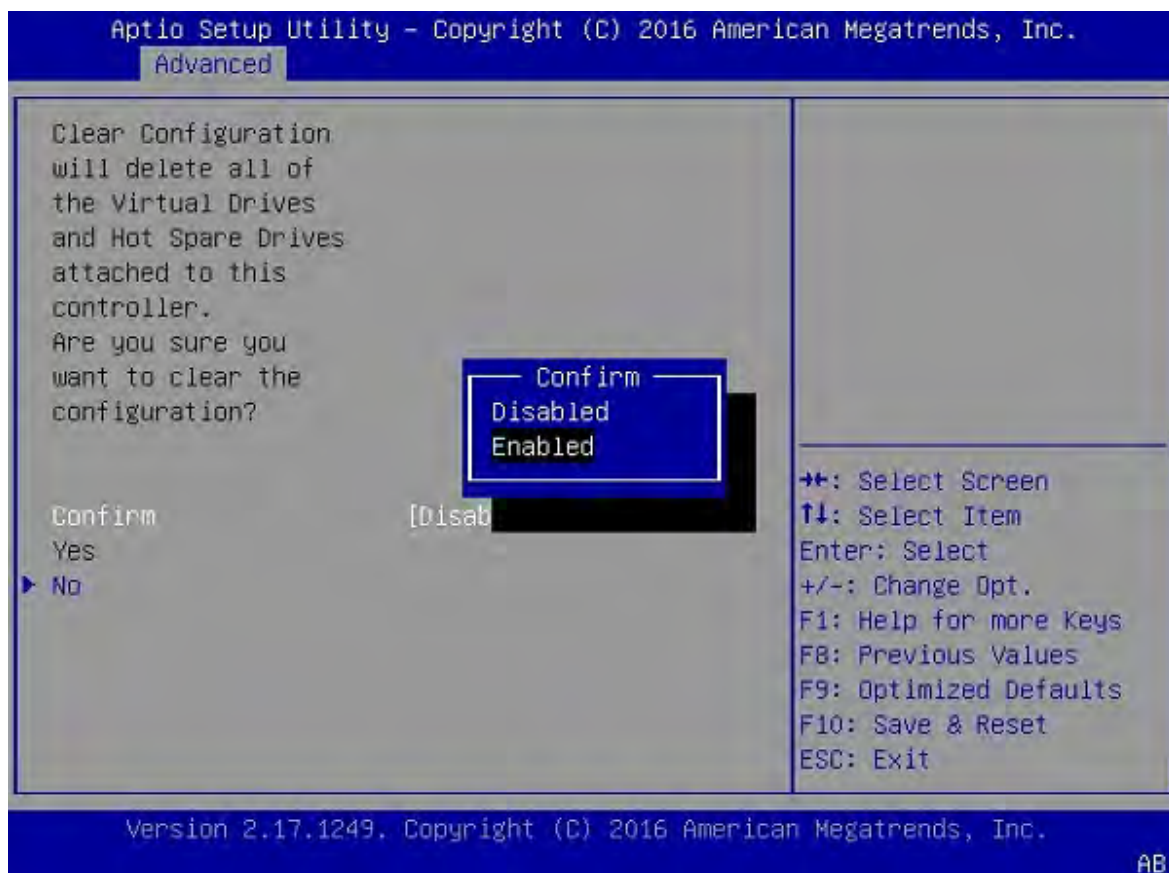
1. Select **Main Menu**, then select **Configuration Management**.



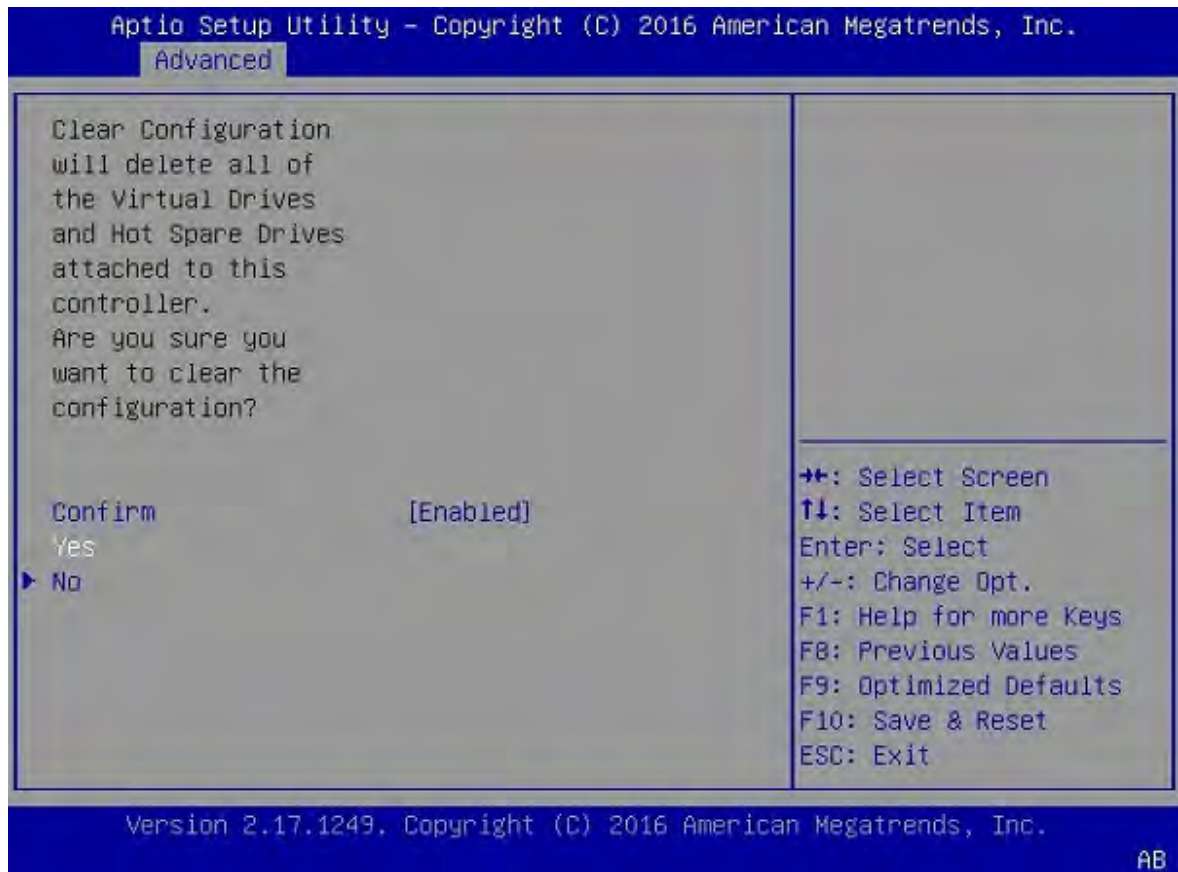
2. Select **Clear Configuration**.



3. Select **Confirm [Disabled]** and then select **Enabled** at the confirmation popup.



4. Select **Yes**, then select **OK** at the success screen.

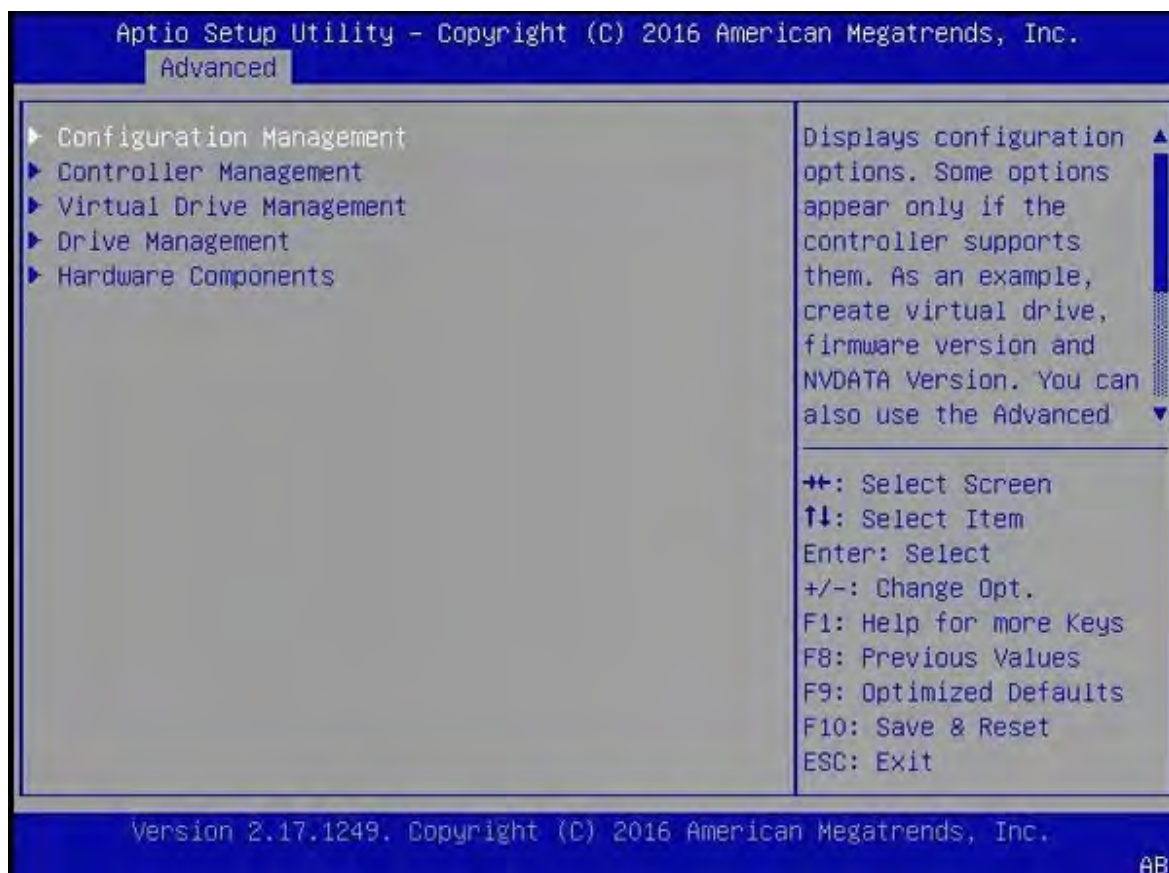


5. Follow the instructions in the sections [Recreate the OS Virtual Drive](#) and then [Recreate the RAID0 Virtual Drive](#).

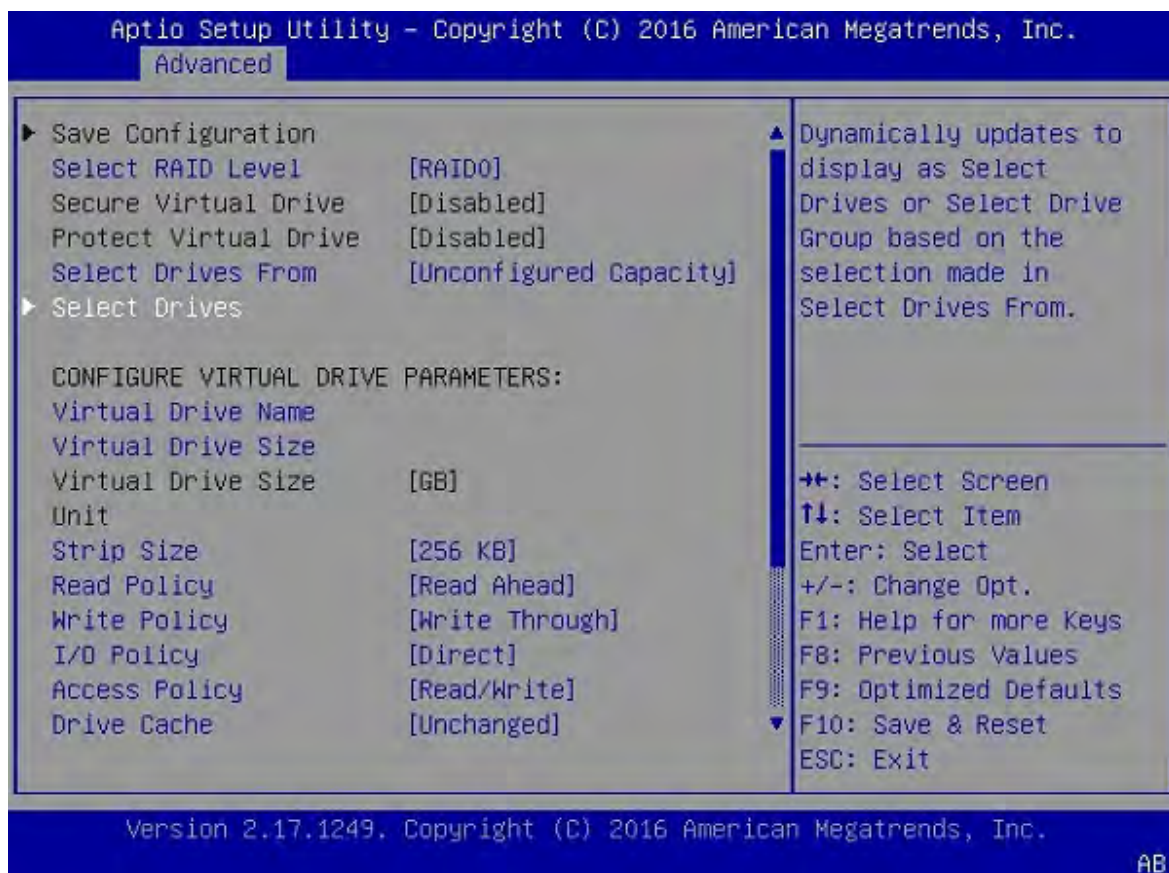
6.6.3.3. Recreate the OS Virtual Drive

These instructions apply when you have replaced the OS drive. Be sure to first complete the instructions in the section [Clear the Drive Group Configuration](#).

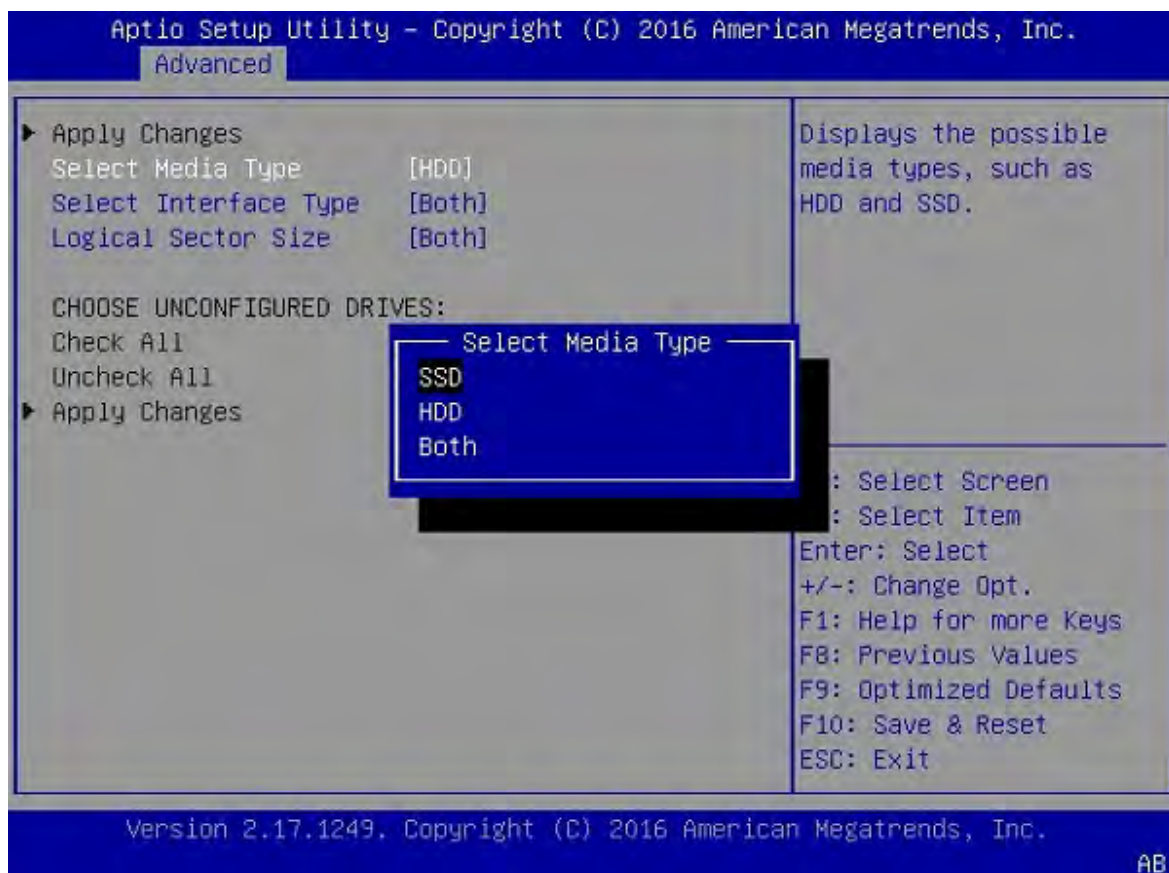
1. Navigate to the RAID Utility Main Menu, then under Actions, select Configure, then select **Configuration Management**.



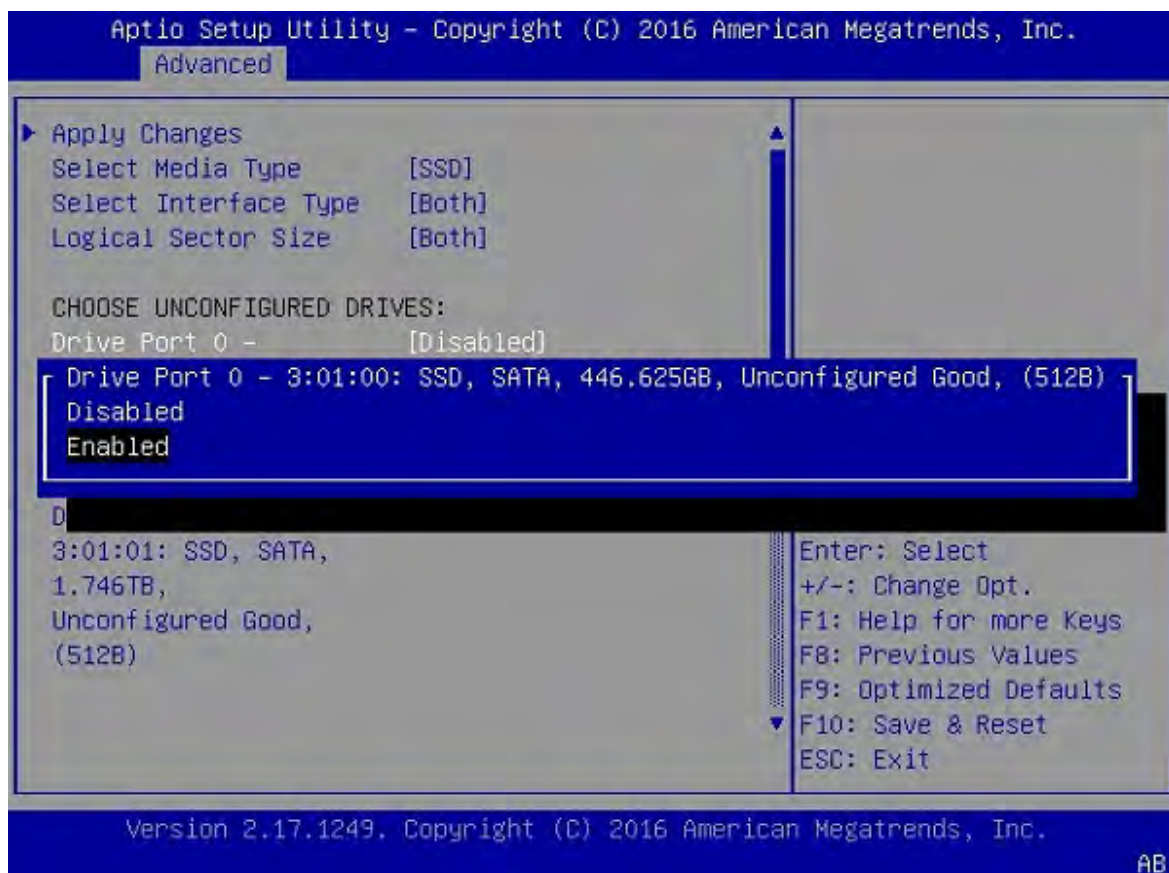
2. Select **Create Virtual Drive**, then select **Select Drives** at the next screen.
Leave all other options at their default settings as shown below.



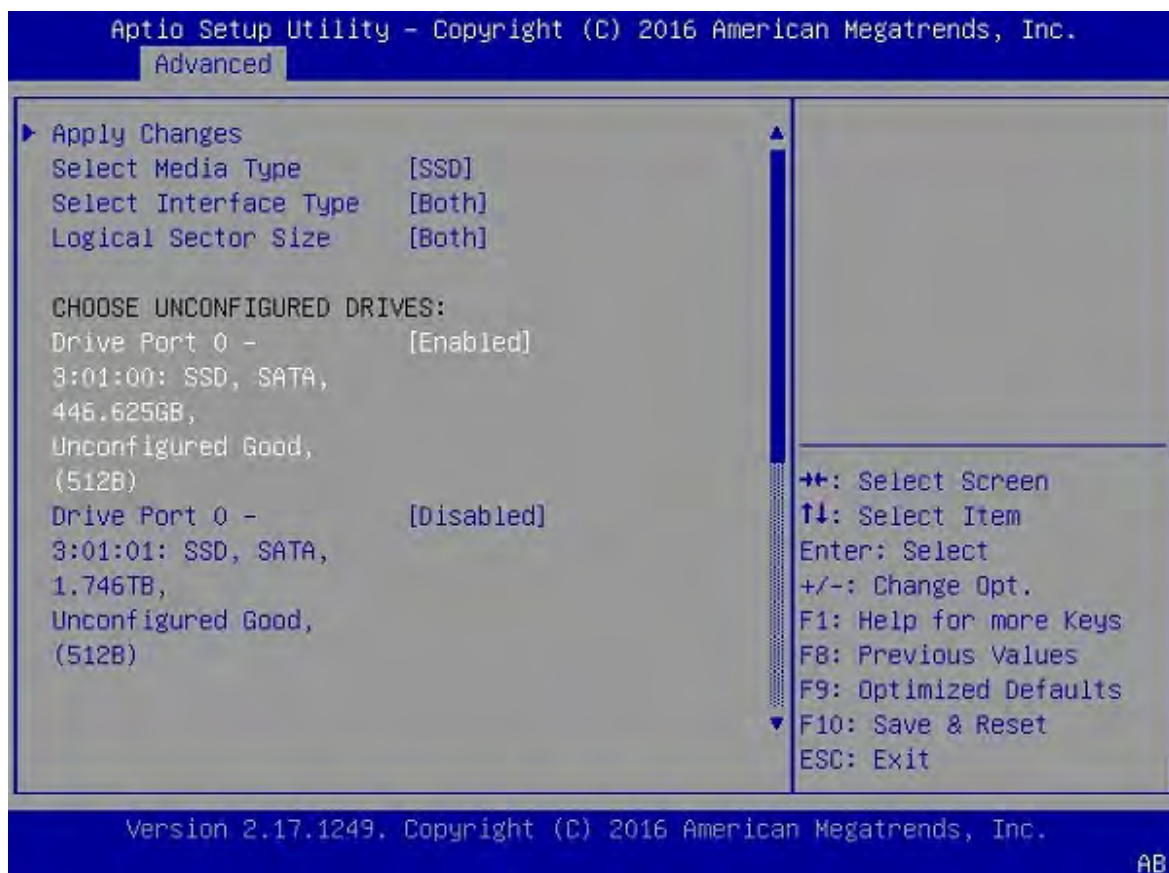
- The list of drives under **CHOOSE UNCONFIGURED DRIVES** will initially be empty.
- To view the available drives, select **Select Media Type [HDD]**, then change to **[SSD]**.



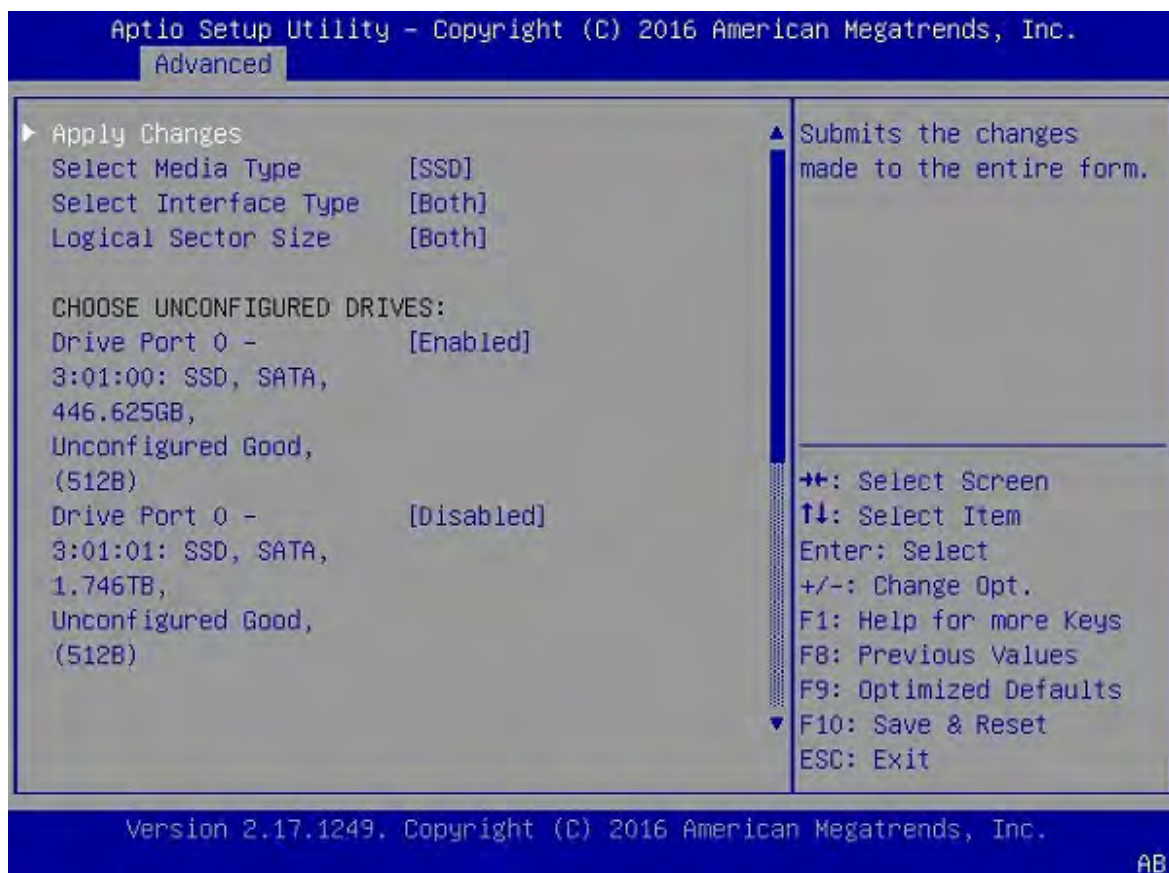
4. Under **CHOOSE UNCONFIGURED DRIVES**, select the 446 GB drive, then change to **[Enabled]** at the pop-up dialog.



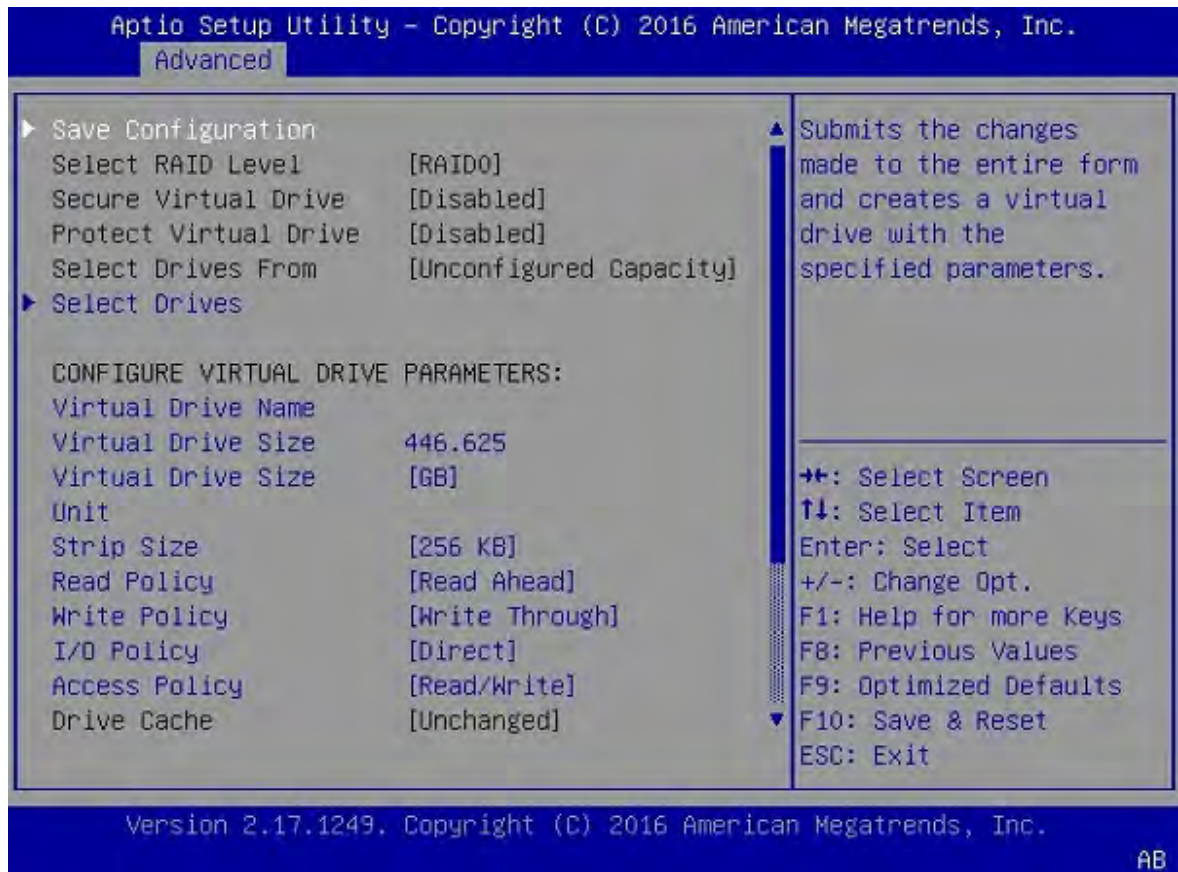
5. Confirm that only the first drive at **Drive Port 0 - 3:01:00** displays as **[Enabled]**.



6. Scroll up and select **Apply Changes**.

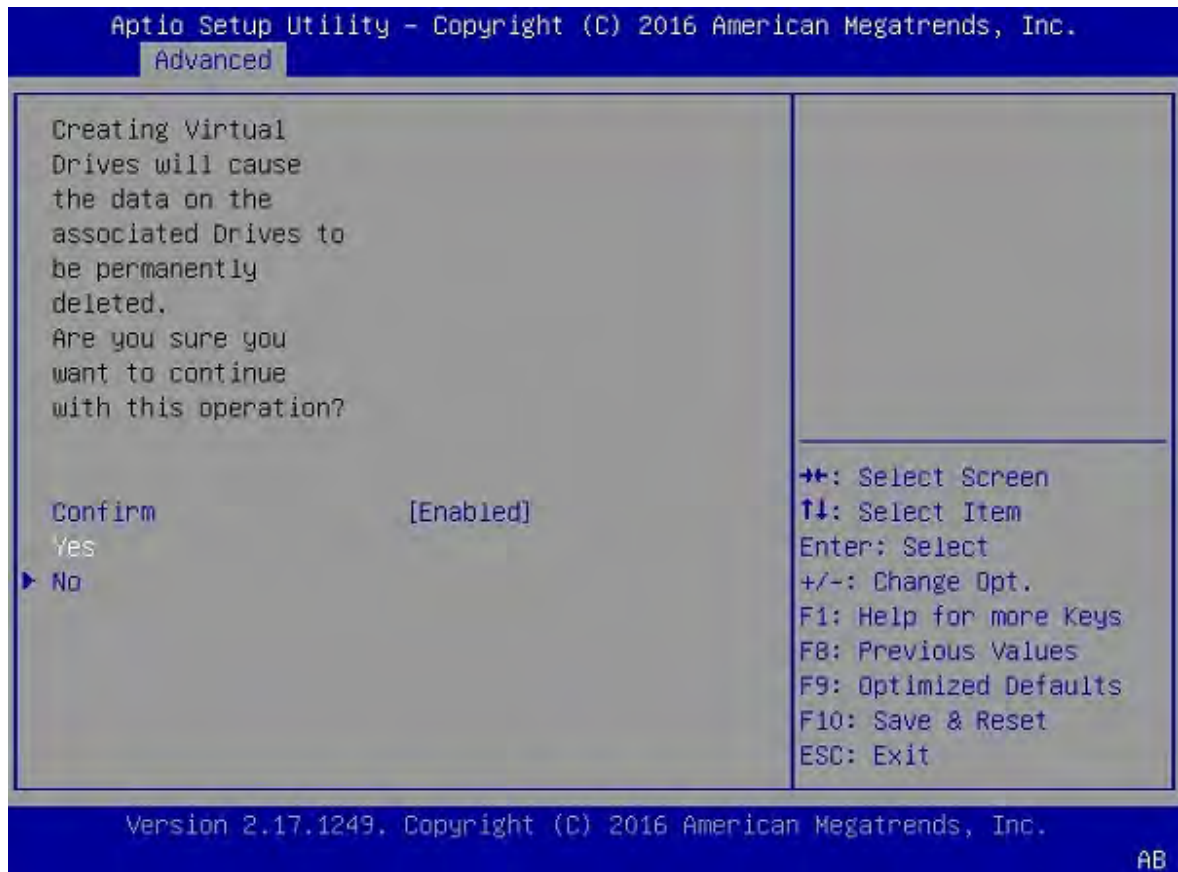


7. Select **OK** at the success screen.



The virtual drive creation page now displays a summary of your selection. The Virtual Drive Size should be approximately 446 GB.

8. Select **Save Configuration** at the top of the menu.
9. Change the **Confirm [Disabled]** field to **[Enabled]** and then select **[Yes]**.



10. Select **[OK]** at the success screen.

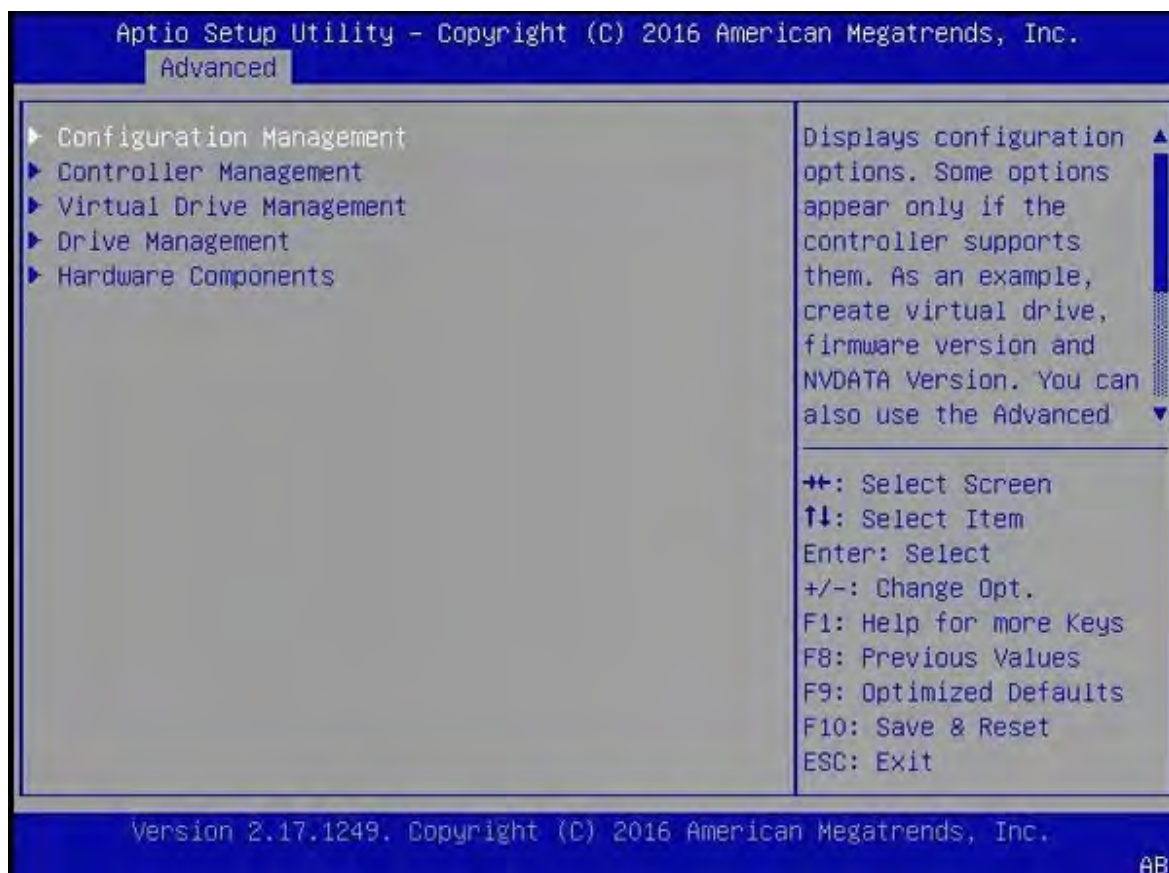
You have successfully re-created Virtual Drive 0, where the OS will be installed.

11. Follow the instructions in the section [Recreate the RAID0 Virtual Drive](#)

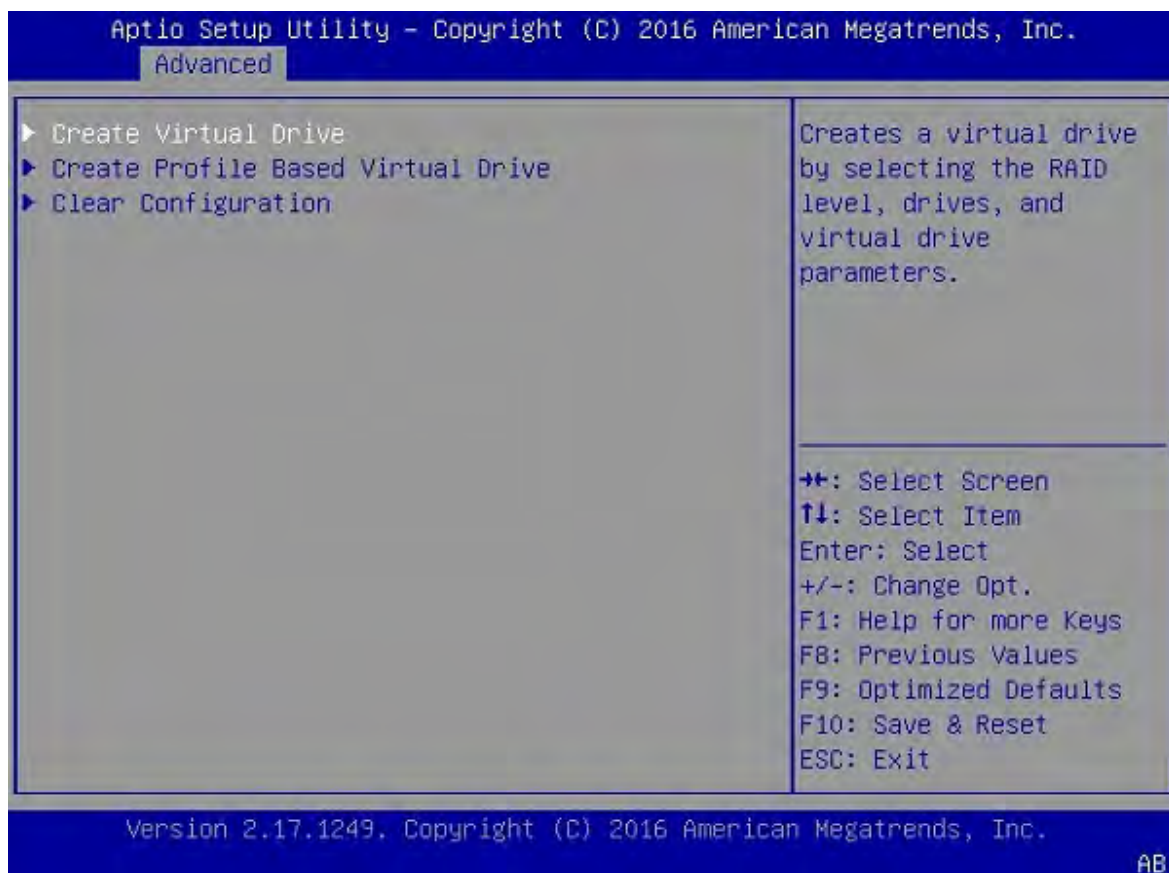
6.6.3.4. Recreate the RAID0 Virtual Drive

These instructions apply when you have replaced the OS drive and cleared the drive group configuration.

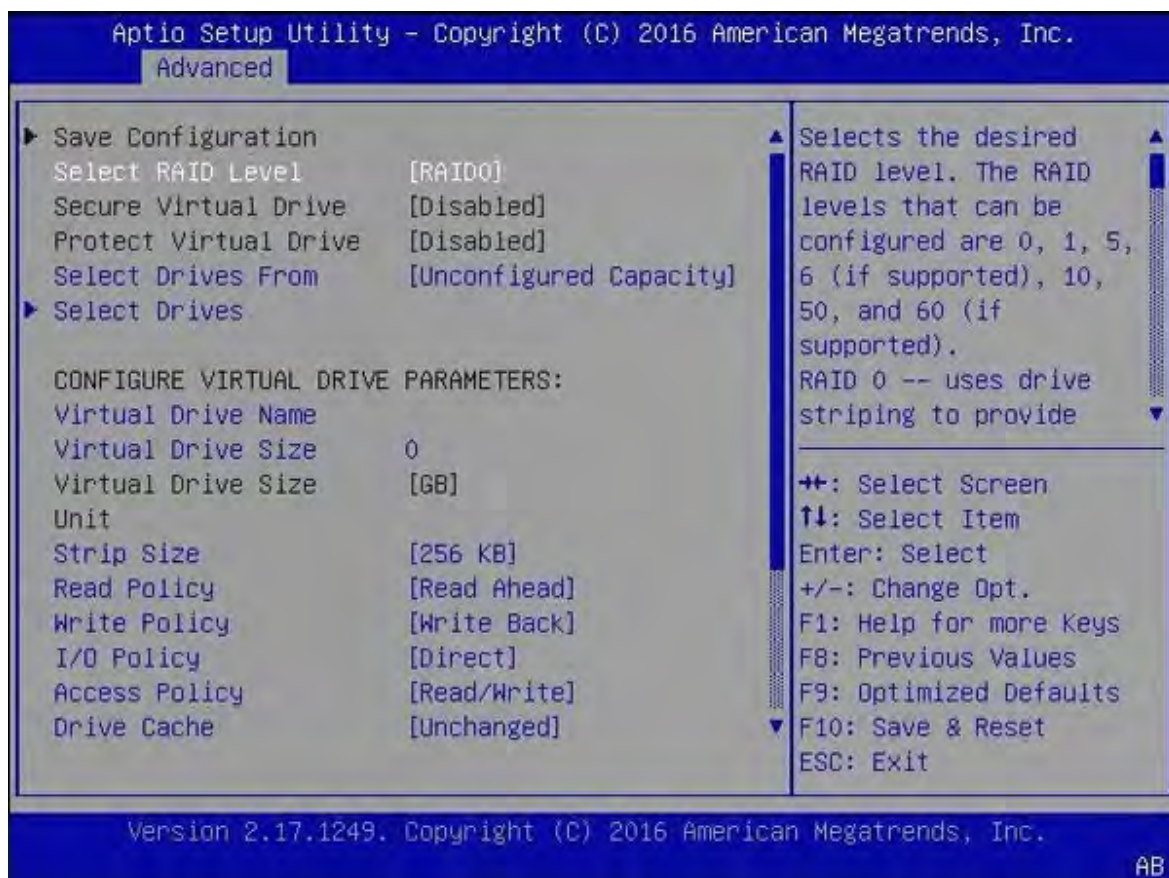
1. Navigate to the RAID Utility Main Menu, then under Action, select Configure, then select **Configuration Management**.



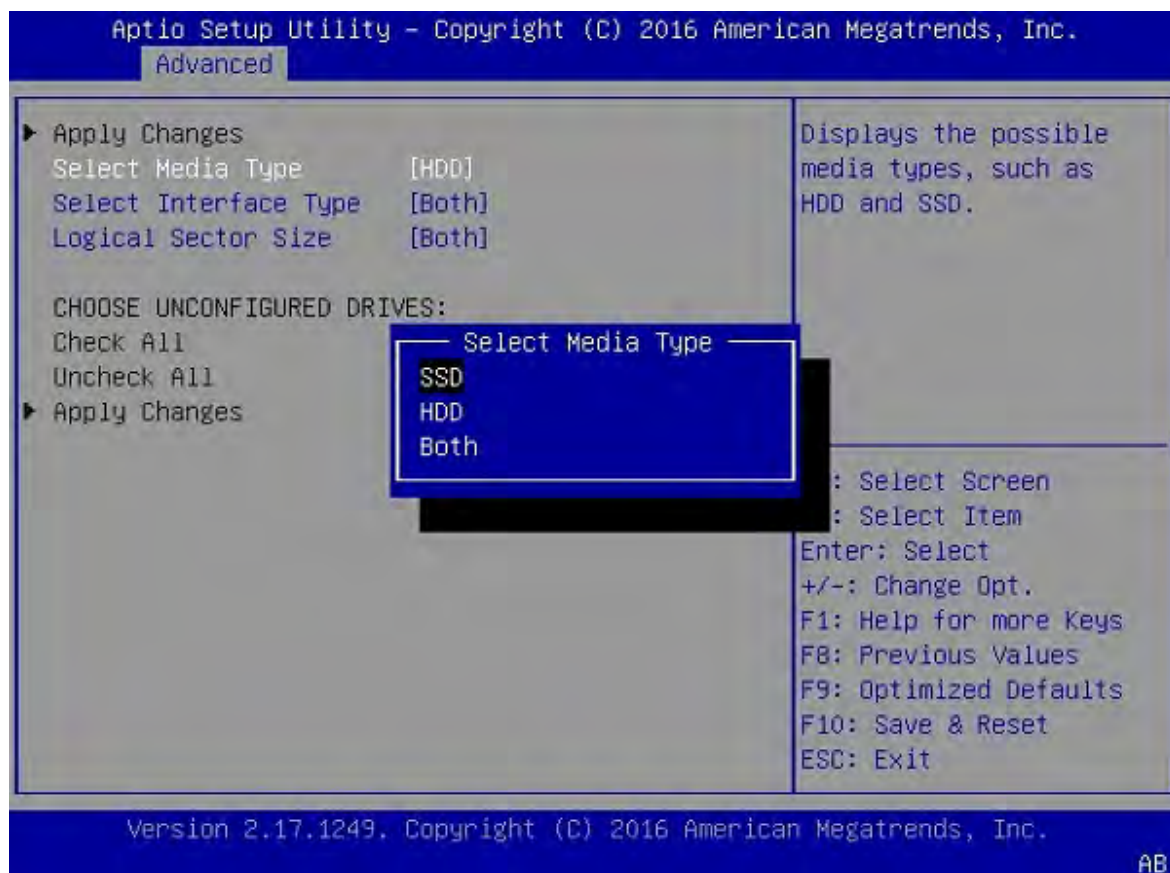
2. Select **Create Virtual Drive**.



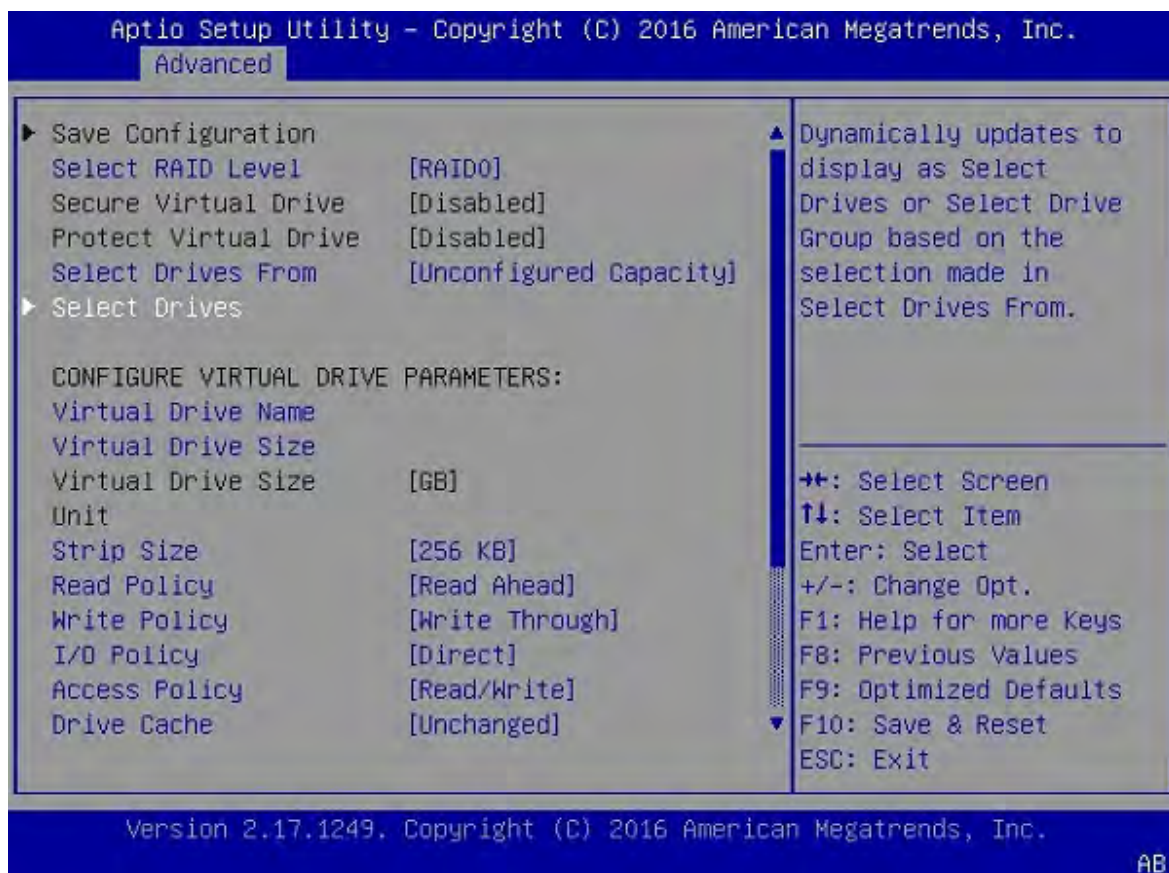
3. Scroll to **Select RAID Level** and switch to **[RAID0]**, if not already set.



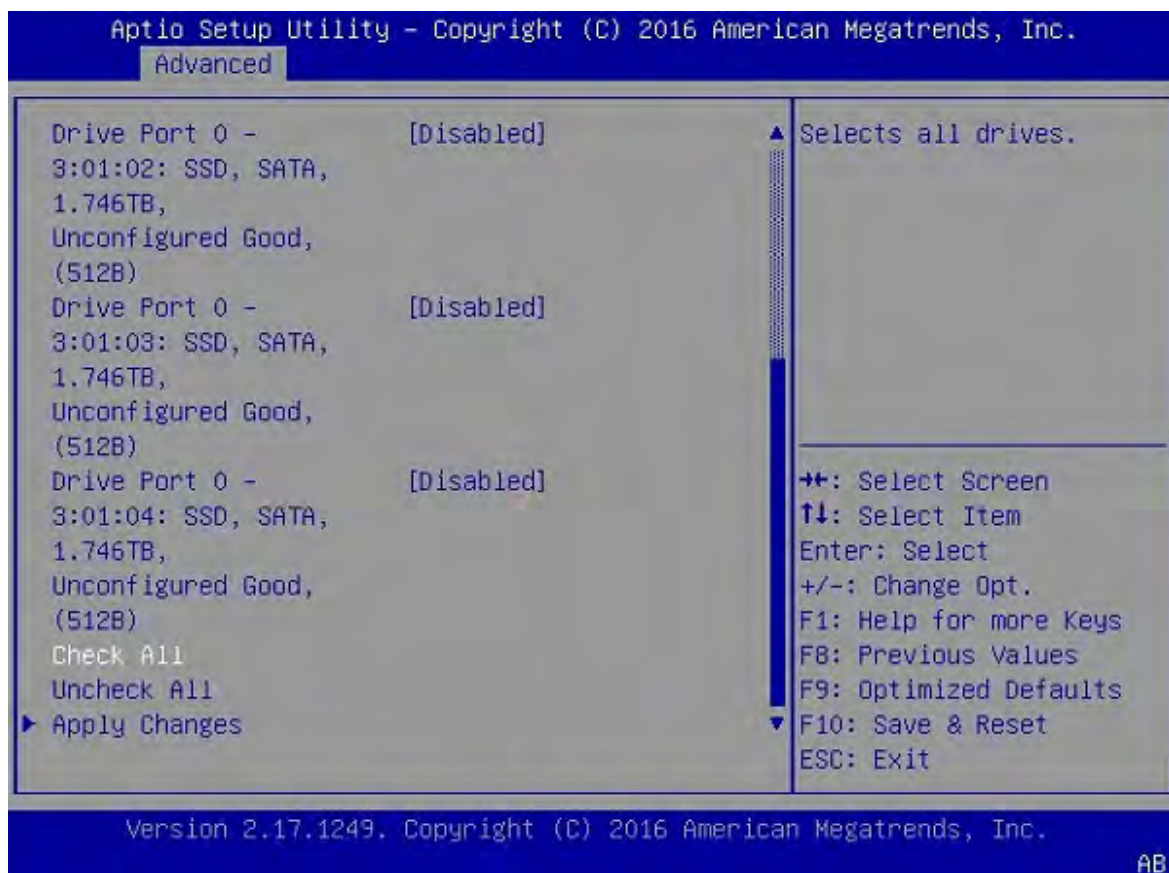
4. Scroll to **Select Media Type** and switch to **[SSD]**.



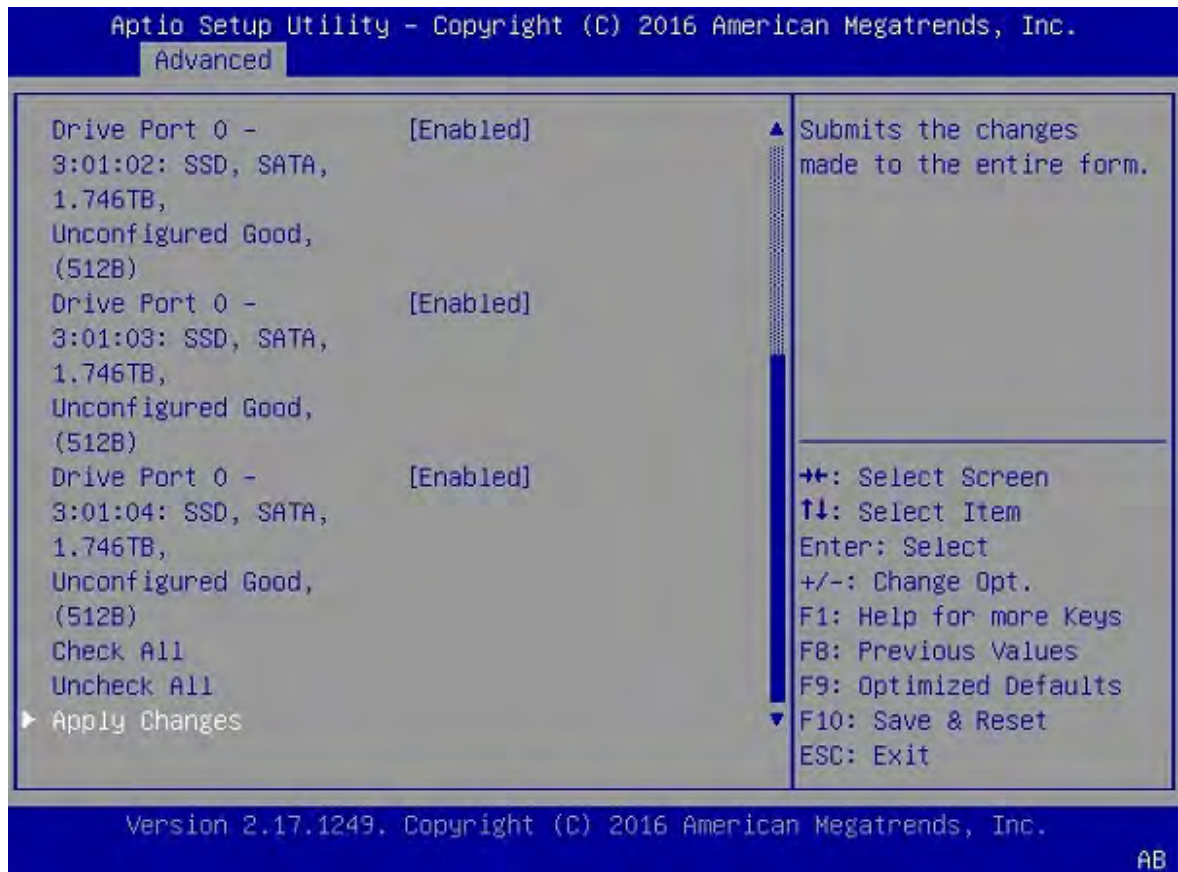
5. Select **Select Drives**.



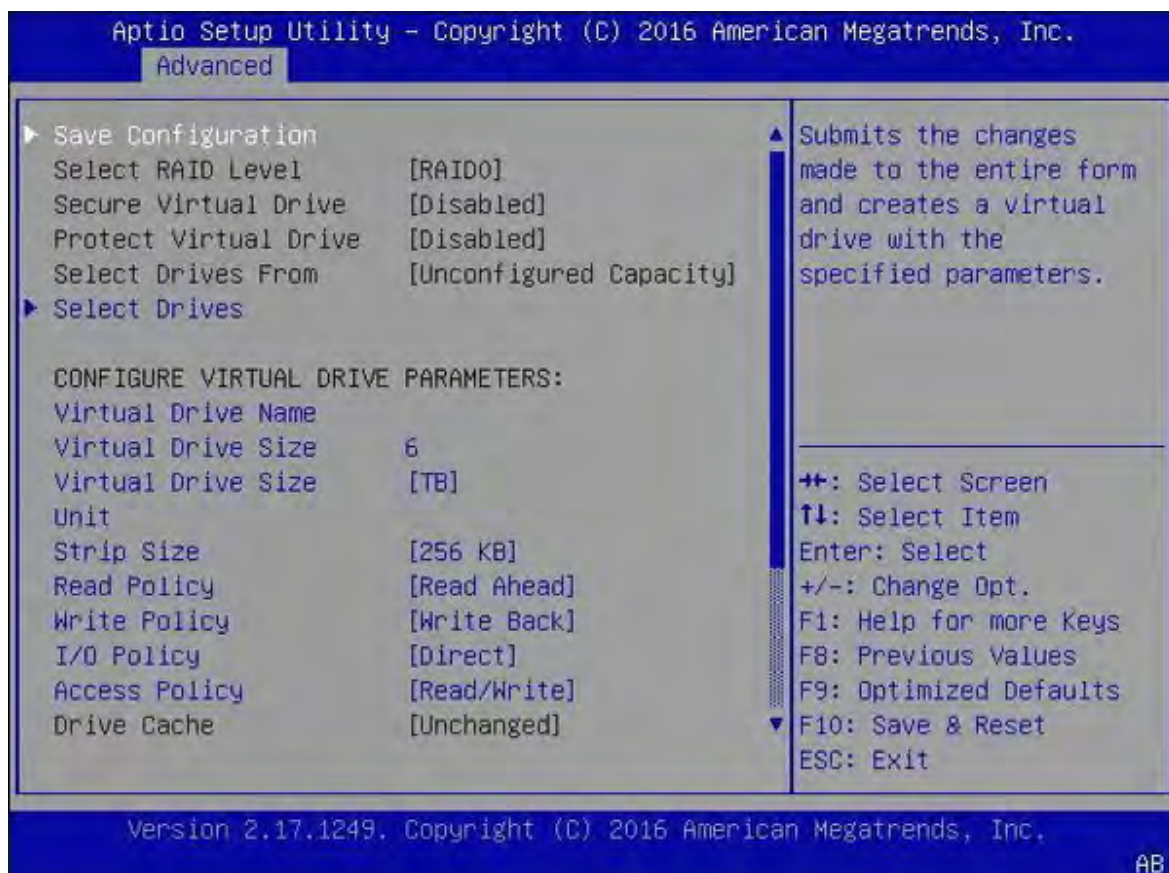
6. Switch all unconfigured 1TB drives to **[Enabled]**.



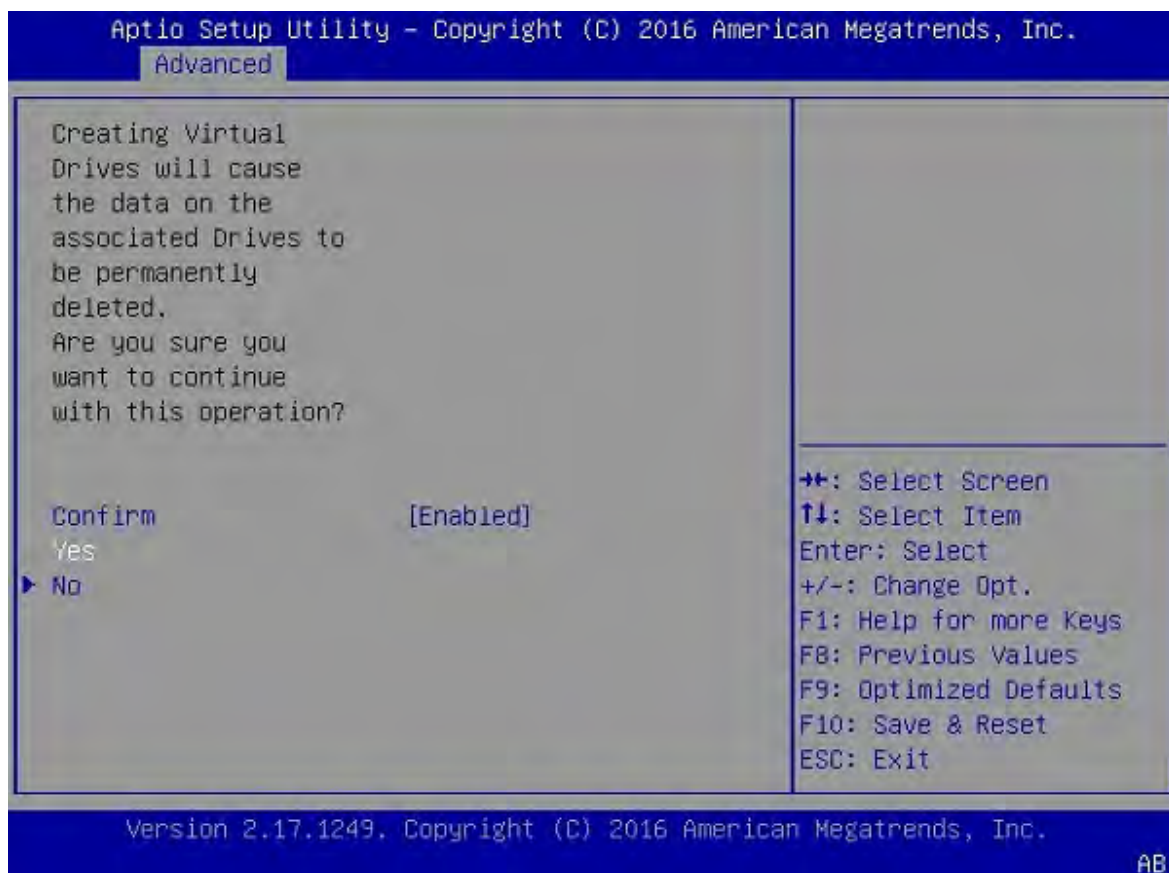
7. Select **Apply Changes**.



8. Change Confirm to **[Enabled]**, then select **Yes**.
9. Select **OK** at the success screen.
The *Create Virtual Drive* screen displays a summary of your selection.
10. Verify that the summary matches your selection, then select **Save Configuration**.



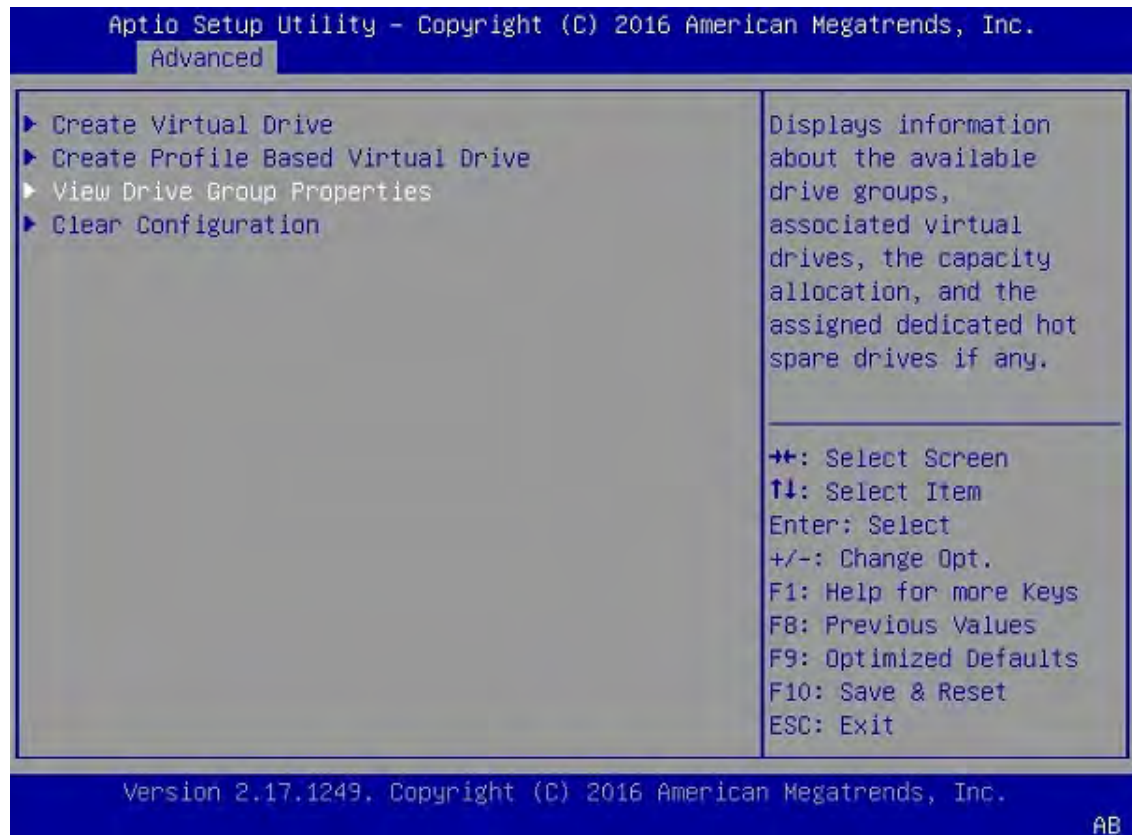
11. Make sure **Confirm** is set to **[Enabled]**, then select **Yes** to confirm the change.



12. Select **OK** at the success screen.

13. Confirm and exit.

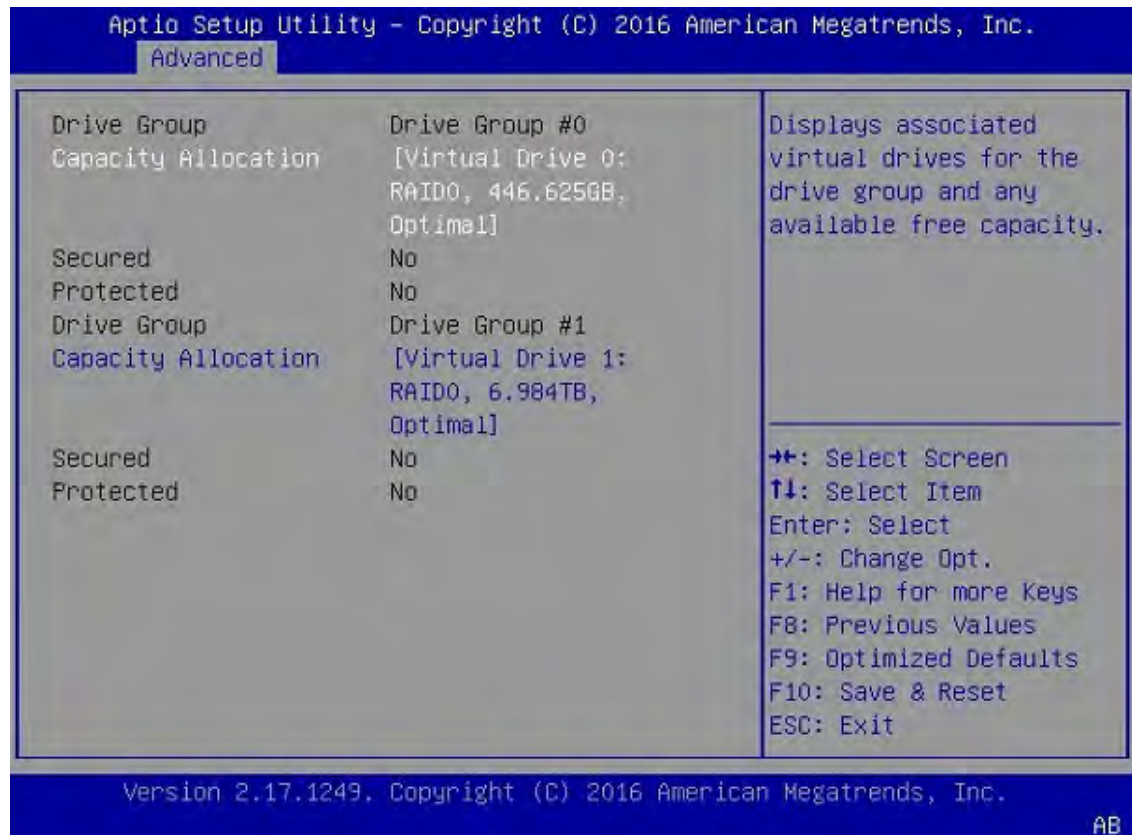
a). Select **View Drive Group Properties** to confirm the configuration.



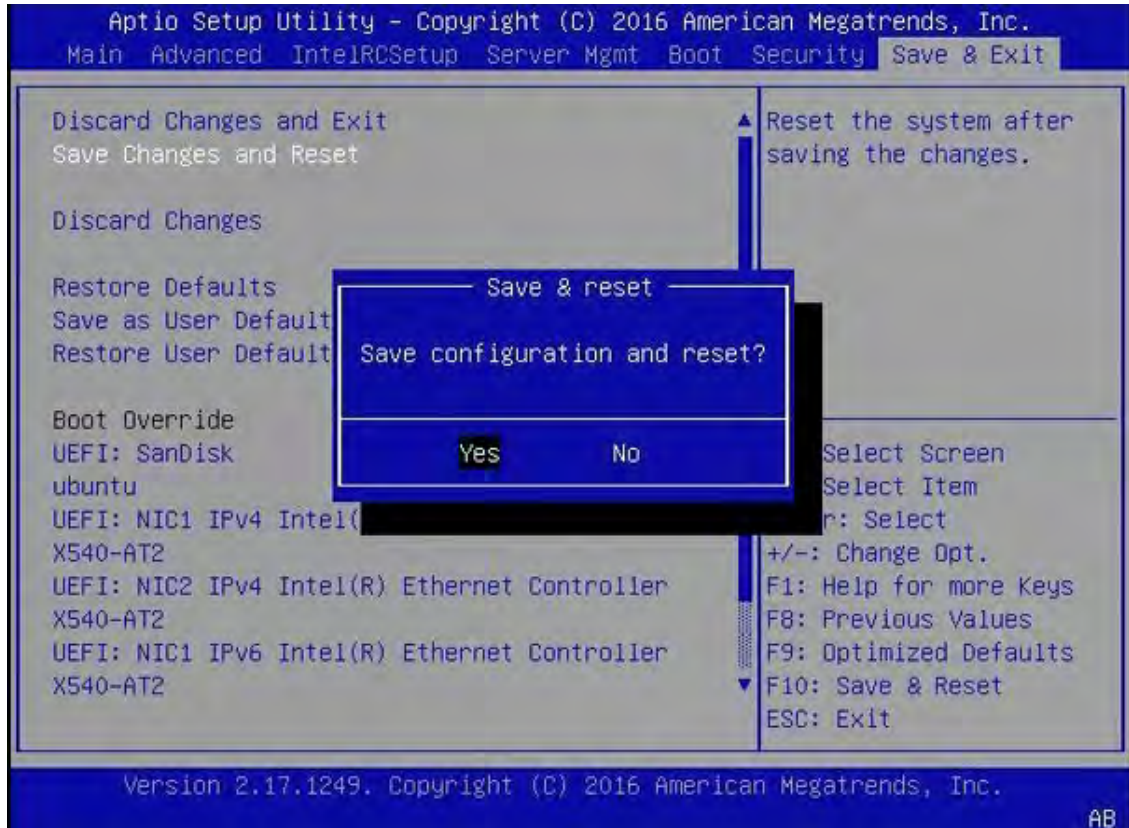
- b). Verify that your configuration screen shows that you have two virtual drives with the following properties:

Virtual Drive 0 of size 446 GB (or very similar)

Virtual Drive 1 of size 7 TB (or very similar).



- c). If your Drive Groups match the above, press **[F10]** to save these settings and reset the system.
- d). Select **Save Changes and Reset**, then select **Yes** at the confirmation prompt.



14. Follow the instructions in the section [Restoring the DGX-1 Software Image](#) to create the partitions.

6.6.4. Recreating the RAID 0 Array

After replacing one of the RAID 0 cache SSDs, you need to recreate the RAID 0 array. If you replaced only the cache and not the operating system SSD, then you can use a convenient script to recreate the RAID array. The script is part of the DGX-1 software as of version 2.0.4. The script requires that the Broadcom StorCLI utility is installed on the DGX system. As of DGX OS Server version 3.1.6, StorCLI is included in the DGX software. For instructions on installing the utility on DGX systems with DGX OS Server versions prior to 3.1.6, see the document *Using StorCLI to Recreate the NVIDIA DGX-1 RAID 0 Array*, available from the NVIDIA Enterprise Support site.



Note: Connect a display (1024x768 or lower resolution) and keyboard to the DGX-1 when booting the DGX-1 before recreating the RAID array. This is because the system may halt at the BIOS screen alerting you that the RAID array needs to be configured. Press C (or whichever key allows you to continue) to complete the boot process. You will be able to do this only if you are operating the DGX-1 through a direct display and keyboard connection.

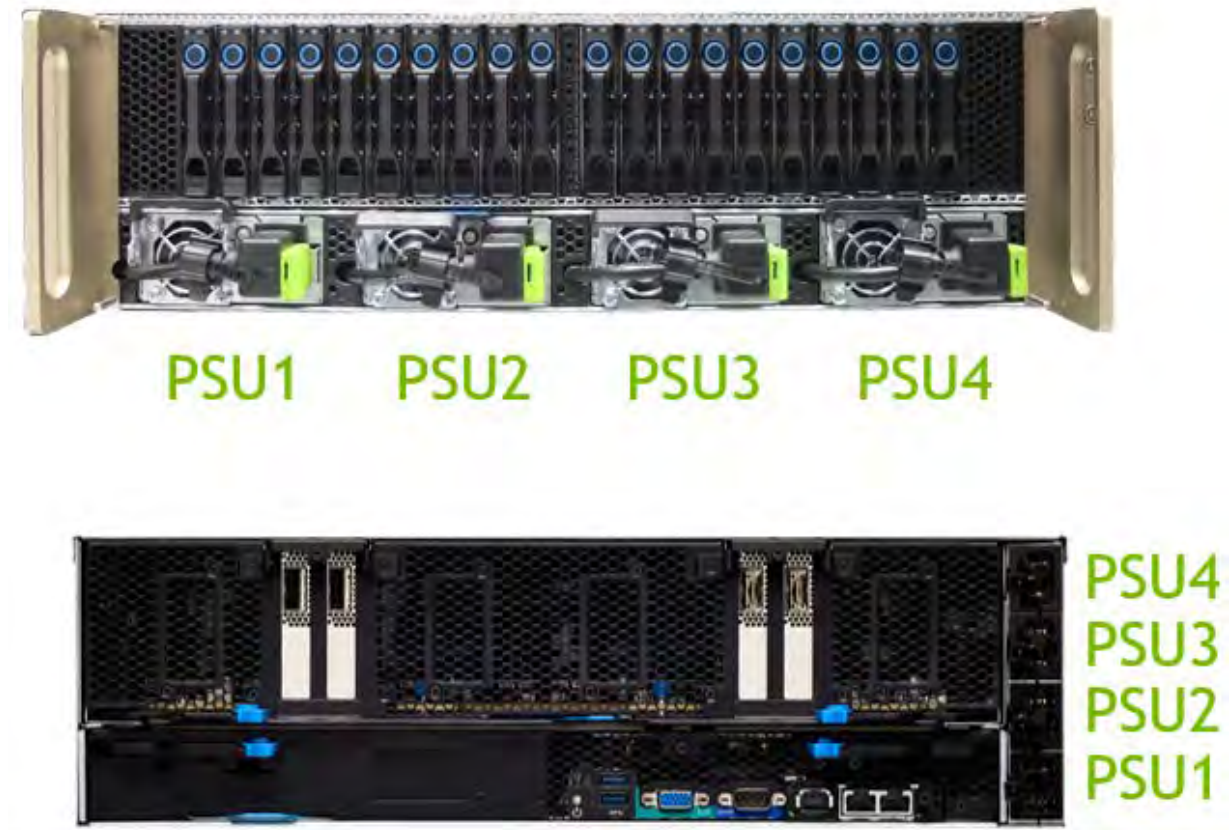
1. Run the script by entering the following on the command line:

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

2. After the script has finished recreating the RAID 0 array, reboot the DGX-1 to verify that `/raid` is mounted and usable.

6.6.5. Replacing the Power Supplies

Access the power supplies from the front of the DGX-1. Use the following diagram to assist in identifying each PSU - the IDs correspond to how the PSUs are identified in the BMC. The top image shows the PSUs and the bottom image shows the corresponding power cord position for each PSU.



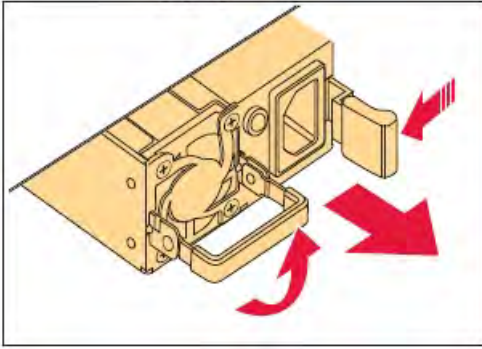
You can hot-swap the power supplies as follows:

1. If not already removed, remove the bezel by grasping the bezel by the side handles and then pulling the bezel straight off the front of the DGX-1.



CAUTION: Be careful not to accidentally press the power button that is on the right edge of the DGX-1 when removing or installing the bezel.

2. Unplug the power cord from the power connector on the fan assembly.
3. Flip the power supply handle out.



4. Push the green release lever to the left and simultaneously use the power supply handle to pull out the power supply.
5. Slide the replacement power supply into the bay and push until seated.
6. Flip the power supply handle up against the power supply.
7. Reconnect the power cord.



Note: IMPORTANT: Make sure that the end of the power cord cable tie is not inserted into the power supply fan. The cable tie can interfere with normal operation of the fan, resulting in failure of the power supply.

8. Reattach the bezel.

With the bezel positioned so that the NVIDIA logo is visible from the front and is on the left-hand side, line up the pins near the corners of the DGX-1 with the holes in back of the bezel, then gently press the bezel against the DGX-1. The bezel is held in place magnetically.

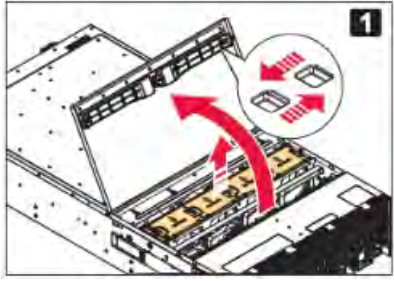


Note: CAUTION: Be careful not to accidentally press the power button that is on the right edge of the DGX-1 when removing or installing the bezel.

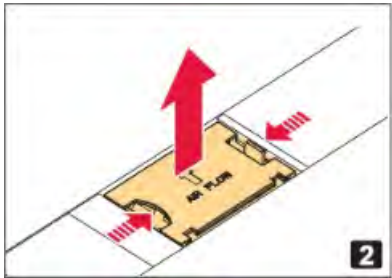
6.6.6. Replacing the Fan Module

CAUTION: To avoid overheating the system, the fan module should be replaced within 25 seconds after removal.

1. Unscrew the thumbscrews at the front of the DGX-1, then slide the DGX-1 about half way out from the rack.
2. Squeeze together the latches at the square access openings on the top of the chassis, then flip open the top of the chassis to expose the fan modules.



3. Squeeze the release tabs on the outer edge of the fan module you want to replace, then pull up to lift the fan module out of the unit.



4. Replace with a new fan module using the reverse steps.

6.6.7. Replacing the Battery

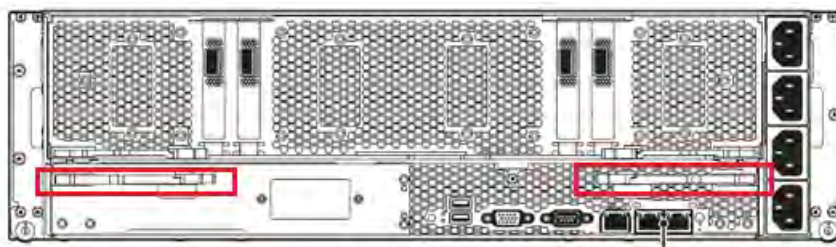
The battery is located on the motherboard tray, which is accessible from the rear of the DGX-1.



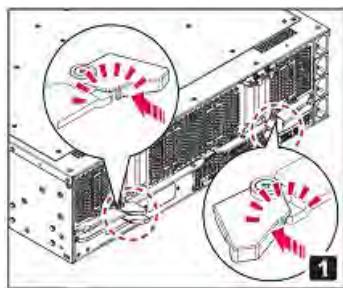
CAUTION: Static Sensitive Devices: - Be sure to observe best practices for electrostatic discharge (ESD) protection. This includes making sure personnel and equipment are connected to a common ground, such as by wearing a wrist strap connected to the chassis ground, and placing components on static-free work surfaces.

1. Turn off the DGX-1 and disconnect all network and power cabling.
2. Remove the motherboard tray.
 - a). Locate the locking levers for the motherboard tray at the rear of the DGX-1.

There are two sets of locking levers. The locking levers for the motherboard are the bottom set.

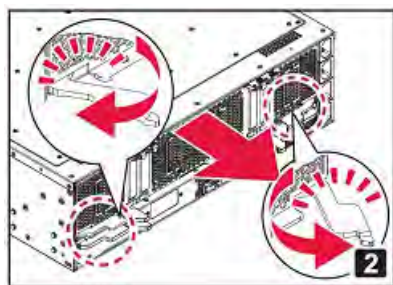


- b). Rotate the retention clasps inward towards the center of the unit.



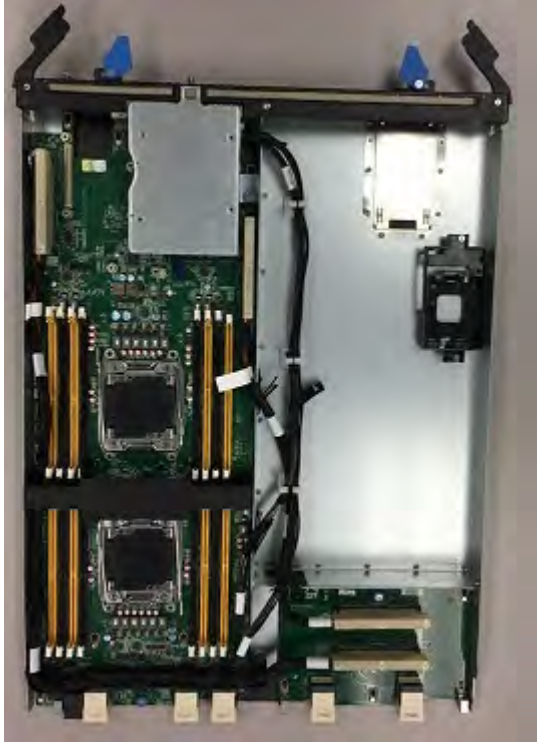
The retention clasps hold the locking levers in place. Rotating the clasps inward releases the locking levers.

- c). Swing the locking levers out and then use them to pull the motherboard tray out of the unit.



Do not pull the unit by the blue retention clasps; they may break.

- d). Set the motherboard tray on a clean work surface, and position it so that the locking levers are at the top as you look down on the tray.



The battery is located at the upper left of the tray, near the edge.

3. Remove the battery.

Gently push the battery to the right, then lift it off the board.



4. Install the new CR2032 battery.

With the "+" symbol on the right side of the battery facing the PCI bus cables, insert the battery at an angle into the slot, then gently push the battery to the left and into place.

5. Carefully insert the motherboard tray back into the unit, then swing the locking levers flat against the tray and secure them in place with the retention clasps.
6. Install all network and power cables.
7. Power on the system, then press **[F2]** or **[Del]** to enter the BIOS setup screen.
8. Navigate to the Main tab, then if necessary, set the System Date and System Time and then press **[F10]** to save the settings and reboot.

6.6.8. Replacing the DIMMs

Before attempting to replace any of the dual inline memory modules (DIMMs), make sure that you know the location of the faulty DIMM needing replacement. The location ID is an alpha-numeric designator, such as A0, A1, B0, B1, etc., and is reported in the BMC log files.

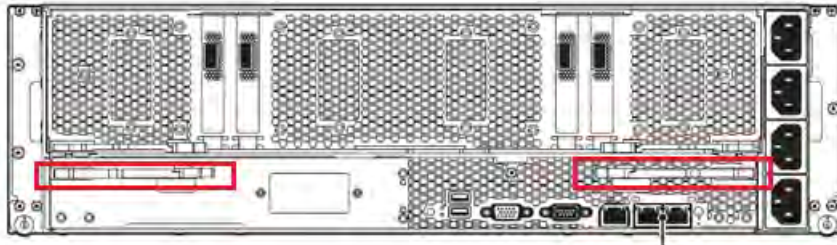


CAUTION: Static Sensitive Devices: - Be sure to observe best practices for electrostatic discharge (ESD) protection. This includes making sure personnel and equipment are connected to a common ground, such as by wearing a wrist strap connected to the chassis ground, and placing components on static-free work surfaces.

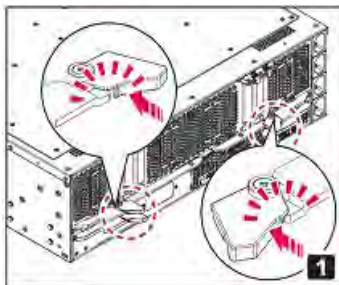
The DIMMs are located on the motherboard tray, which is accessible from the rear of the DGX-1.

1. Turn off the DGX-1 and disconnect all network and power cabling.
2. Remove the motherboard tray.
 - a). Locate the locking levers for the motherboard tray at the rear of the DGX-1.

There are two sets of locking levers. The locking levers for the motherboard are the bottom set.

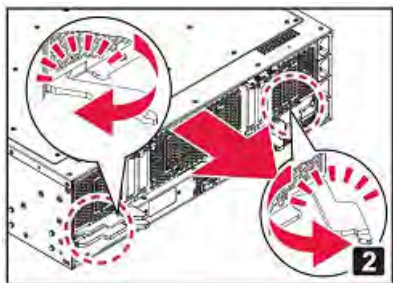


- b). Rotate the retention clasps inward towards the center of the unit.



The retention clasps hold the locking levers in place. Rotating the clasps inward releases the locking levers.

- c). Swing the locking levers out and then use them to pull the motherboard tray out of the unit.



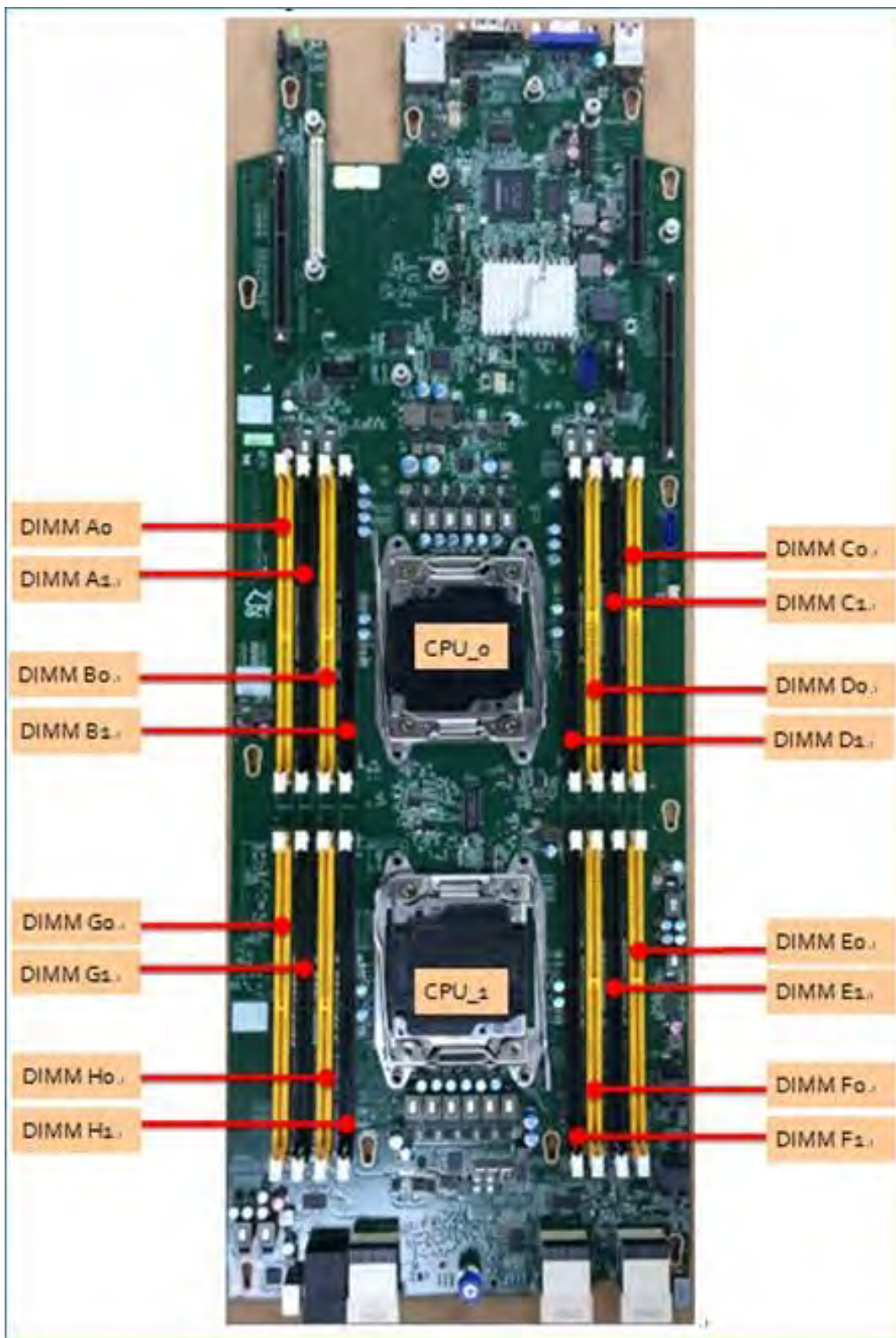
Do not pull the unit by the blue retention clasps; they may break.

- d). Set the motherboard tray on a clean work surface, and position it so that the locking levers are at the top as you look down on the tray.

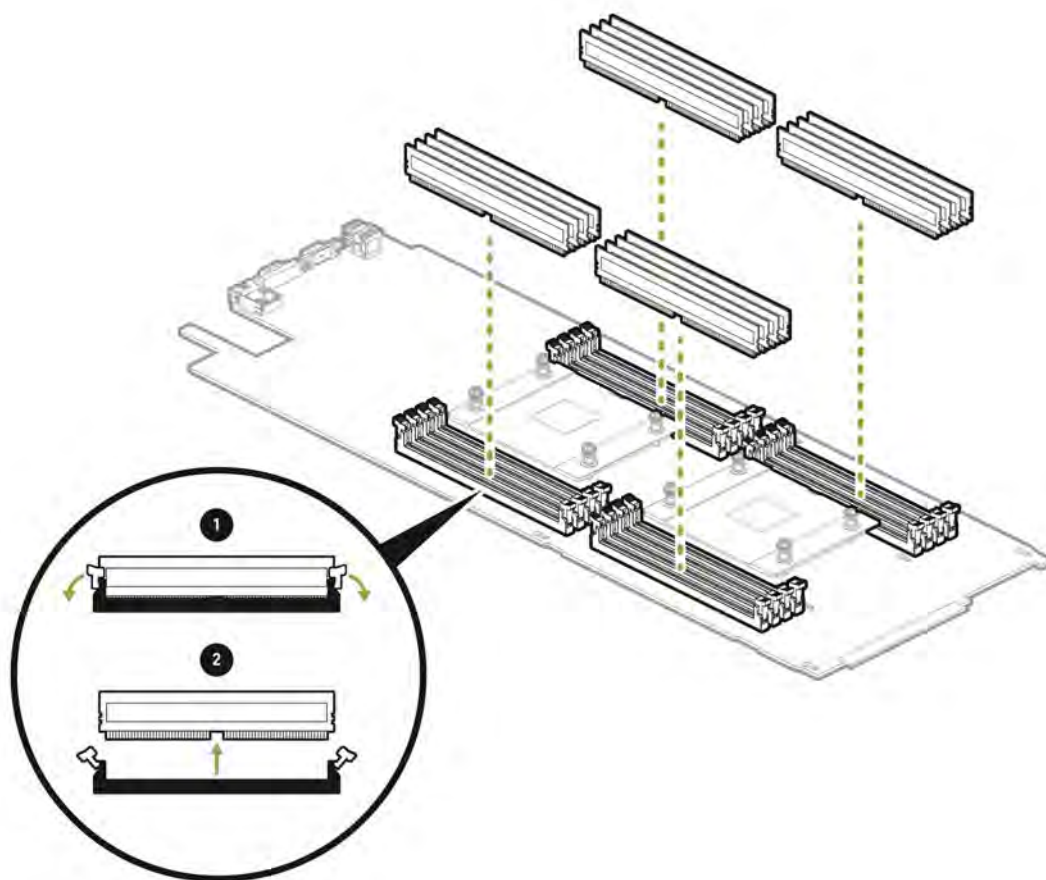


The DIMMs are on a printed circuit board on the left side of the tray.

3. Using the figure below as a guide, locate the DIMM corresponding to the ID of the faulty DIMM as reported in the BMC log.



4. Remove the DIMM.

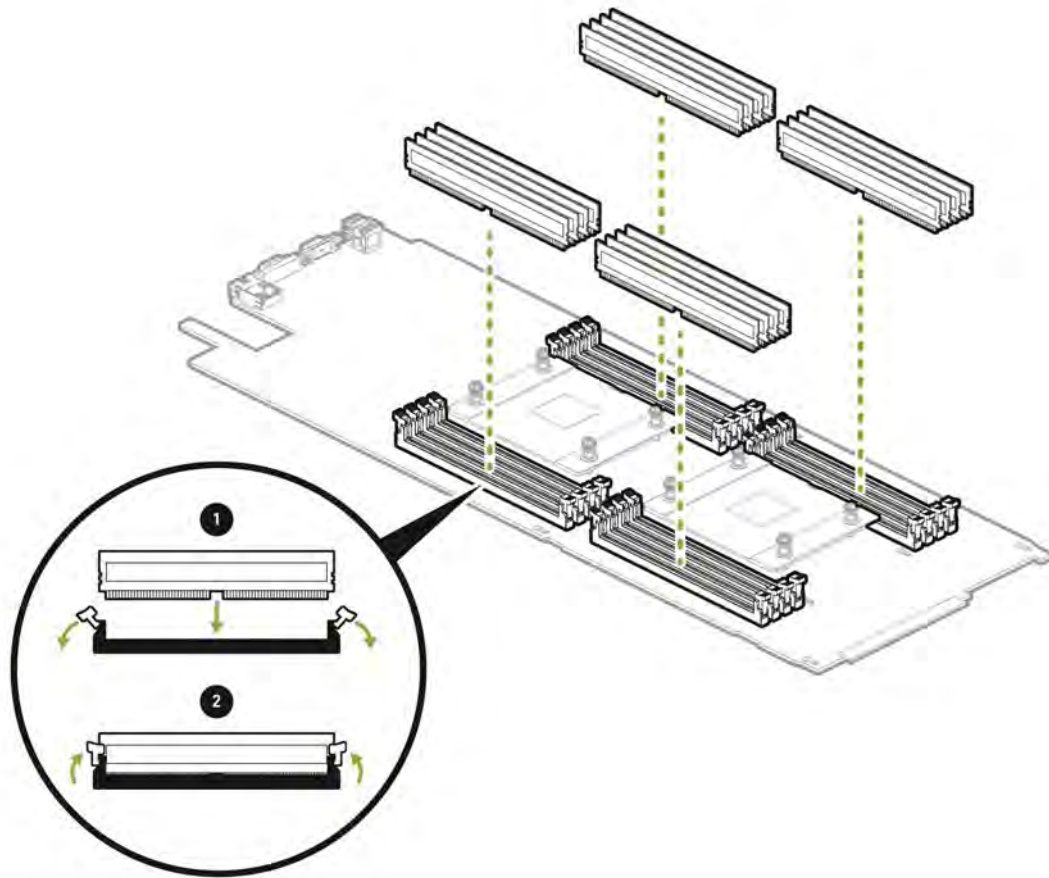


a). Press down on the side latches at both ends of the DIMM socket to push them away from the DIMM.

This should unseat the DIMM from the socket.

b). Pull the DIMM straight up to remove it from the socket.

5. Carefully insert the replacement DIMM.



- a). Make sure the socket latches are open.
 - b). Position the DIMM over the socket, making sure that the notch on the DIMM lines up with the key in the slot, then press the DIMM down into the socket until the side latches click in place.
 - c). Make sure that the latches are up and locked in place.
6. Carefully insert the motherboard tray back into the unit, then swing the locking levers flat against the tray and secure them in place with the retention clasps.

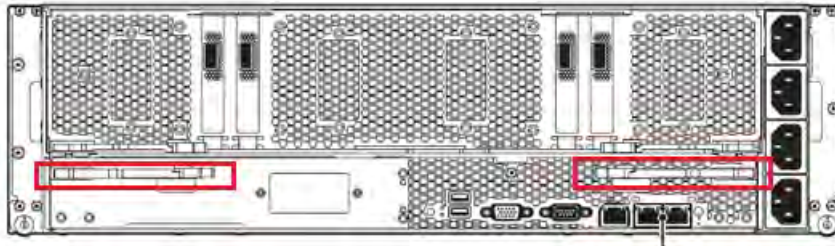
6.6.9. Installing/Replacing the 10GbE Mezzanine SPF+ NIC

The 10GbE SPF+ NIC is available as an optional replacement for the Dual port, 10GBASE-T, network adapter Mezzanine. It is located on the motherboard tray, which is accessible from the rear of the DGX-1 server.

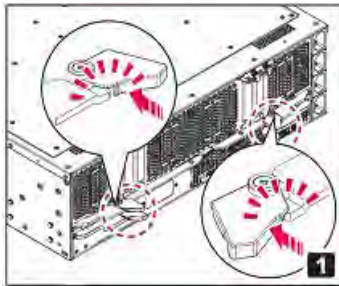


CAUTION: Static Sensitive Devices: - Be sure to observe best practices for electrostatic discharge (ESD) protection. This includes making sure personnel and equipment are connected to a common ground, such as by wearing a wrist strap connected to the chassis ground, and placing components on static-free work surfaces.

1. Turn off the DGX-1 server and disconnect all network and power cabling.
2. Remove the motherboard tray.
 - a). Locate the locking levers for the motherboard tray at the rear of the DGX-1.
There are two sets of locking levers. The locking levers for the motherboard are the bottom set.

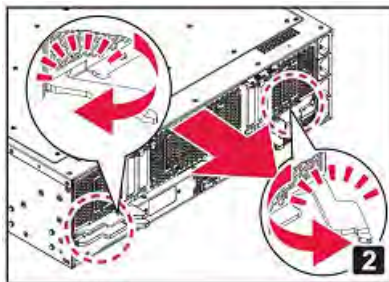


- b). Rotate the retention clasps inward towards the center of the unit.



The retention clasps hold the locking levers in place. Rotating the clasps inward releases the locking levers.

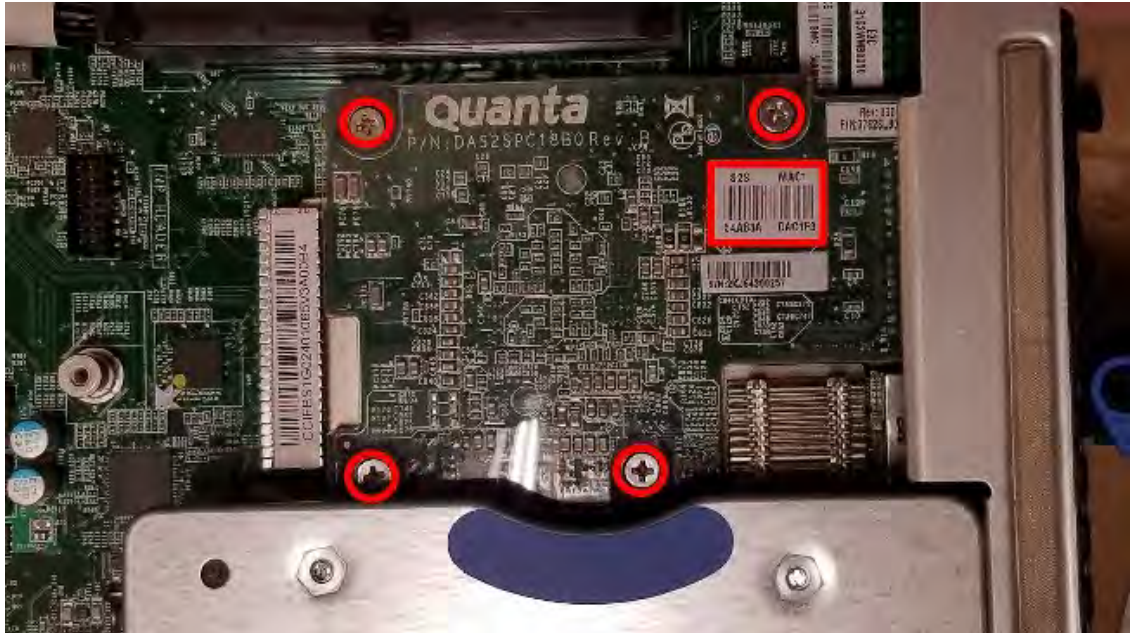
- c). Swing the locking levers out and then use them to pull the motherboard tray out of the unit.



Do not pull the unit by the blue retention clasps; they may break.

- d). Set the motherboard tray on a clean, static-free work surface.
3. Replace the NIC.
 - a). Locate the NIC.

The 10G NIC is located in the corner of the motherboard tray, by the locking levers. Look for the white Quanta label on the sixscreen.



- b). Remove the four screws as identified in the previous image, then gently lift the NIC off the motherboard..
 - c). Line up the new NIC so that the pins of the card are over the mezzanine connector on the motherboard, then gently press down until the card reaches the bottom..
 - d). Reinstall the four screws to secure the card to the motherboard.
 - e). Carefully insert the motherboard tray back into the unit, then swing the locking levers flat against the tray and secure them in place with the retention clasps.
 - f). Reconnect all cables and apply power to the system.
4. Validate the installation.
- a). Check that there are no BMC errors.
 - b). Validate Ethernet connectivity.

Check the Link and Speed LEDs on the 10G NIC card to see if the link is active (amber Link LED - left-side) and 10Gb/s (green Speed LED - right-side).

Use the following command to check if the 10G NIC card is recognized at the OS level.

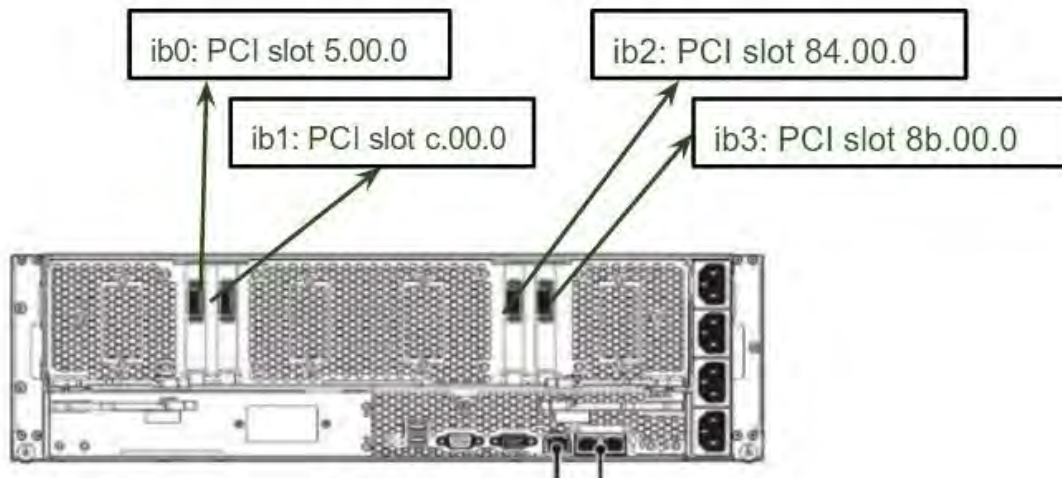
```
# lspci -nn | grep net
```

Use the following commands to check if the device is up, to check the status, and to verify that the MTU setting is correct.

```
# ip link show <device>
# ifconfig <device>
# ethtool -i <device>
```

6.6.10. Replacing the InfiniBand Cards

The InfiniBand cards are located on the GPU tray which is accessible from the rear of the DGX-1. Be sure you have identified the faulty InfiniBand card needing to be replaced. The slots are identified as indicated in the following image.



CAUTION: Static Sensitive Devices: - Be sure to observe best practices for electrostatic discharge (ESD) protection. This includes making sure personnel and equipment are connected to a common ground, such as by wearing a wrist strap connected to the chassis ground, and placing components on static-free work surfaces.

The InfiniBand cards installed in the DGX-1 system are either Mellanox model CX455A or CX555A. Before requesting the replacement InfiniBand card from NVIDIA, identify the model of the card that you need to order. To do this, enter the following on the DGX-1 console.

```
lspci | grep -i Mellanox
```

Output for model CX455A

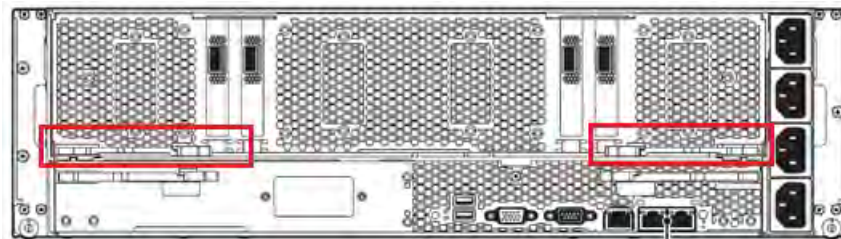
```
05:00.0 Infiniband controller: Mellanox Technologies MT27700 Family [ConnectX-4]
0c:00.0 Infiniband controller: Mellanox Technologies MT27700 Family [ConnectX-4]
84:00.0 Infiniband controller: Mellanox Technologies MT27700 Family [ConnectX-4]
8b:00.0 Infiniband controller: Mellanox Technologies MT27700 Family [ConnectX-4]
```

Output for model CX555A

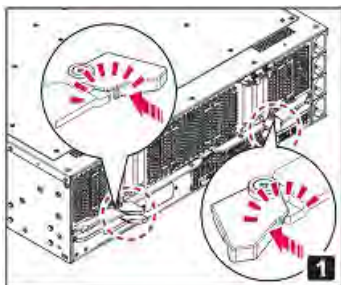
```
05:00.0 Infiniband controller: Mellanox Technologies MT27800 Family [ConnectX-5]
0c:00.0 Infiniband controller: Mellanox Technologies MT27800 Family [ConnectX-5]
84:00.0 Infiniband controller: Mellanox Technologies MT27800 Family [ConnectX-5]
8b:00.0 Infiniband controller: Mellanox Technologies MT27800 Family [ConnectX-5]
```

1. Turn off the DGX-1 and disconnect all network and power cabling.
2. Remove the GPU tray.
 - a). Locate the locking levers for the GPU tray at the rear of the DGX-1.

There are two sets of locking levers. The locking levers for the GPU tray are the top set.

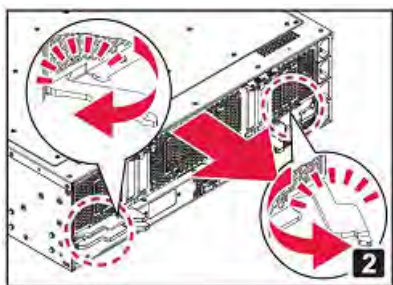


- b). Rotate the retention clasps inward towards the center of the unit.



The retention clasps hold the locking levers in place. Rotating the clasps inward releases the locking levers.

- c). Swing the locking levers out and then use them to pull the GPU tray out of the unit.



Do not pull the unit by the blue retention clasps; they may break.

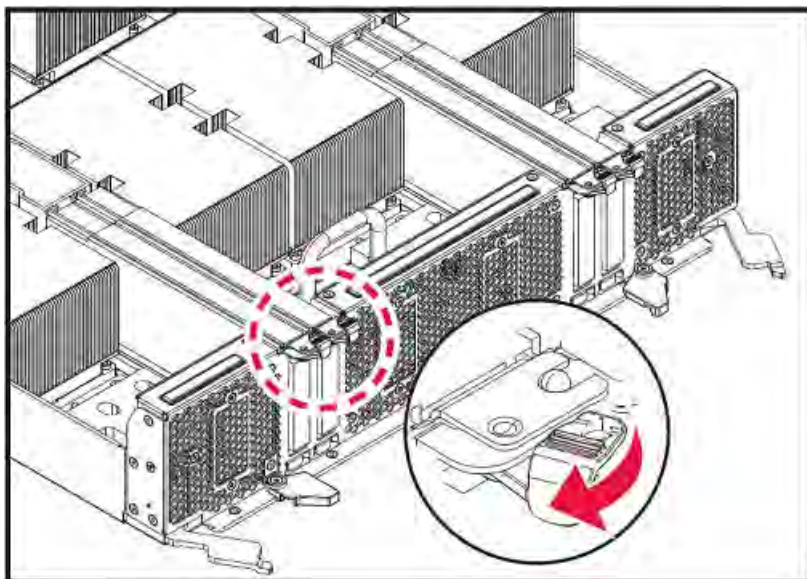
3. Set the GPU tray on a clean work surface.

WARNING: Do not attempt to move or lift the GPU tray by grabbing the U-bolts.

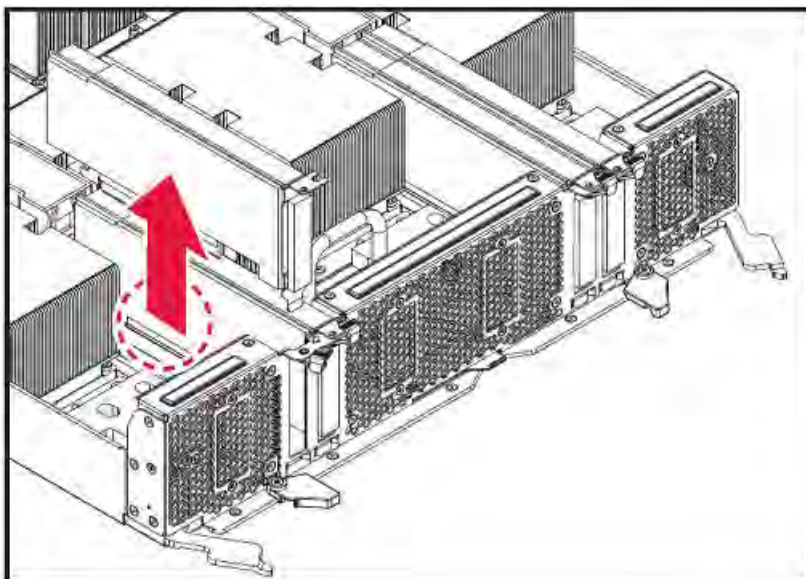


To properly move the GPU tray, grab the tray by the outer edges of the assembly and support it from underneath, taking care not to damage any components.

4. At the top edge of the bracket for the InfiniBand card that you want to replace, rotate the retention clasp to free the bracket.



5. Firmly grasp the InfiniBand card and lift it straight up out of the PCIe slot.



6. Compare the physical label on the removed card with the label on the replacement card to make sure they are the same model.

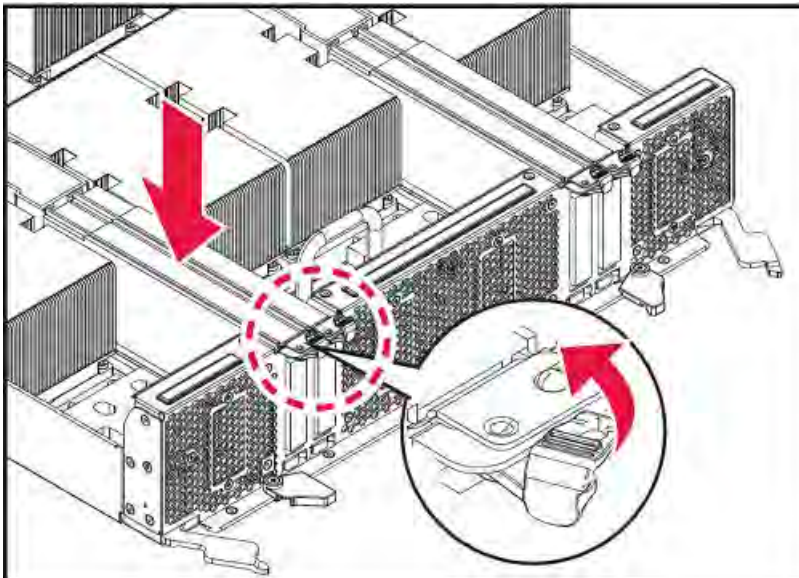
Figure 1. Example label for model CX455A



Figure 2. Example label for model CX555A



7. Position the replacement InfiniBand card over the empty PCIe slot and insert it into the slot.
8. Swing the retention clasp over the bracket to secure the bracket in place.



9. Carefully insert the GPU tray back into the unit, then swing the locking levers flat against the tray and secure them in place with the retention clasps.
10. Reconnect all connectors, boot the system, then perform the verification and setup steps described in the next section.

6.6.11. Setting Up the InfiniBand Cards

This section describes the steps needed to verify that the InfiniBand card has been replaced correctly.

1. With the DGX-1 turned on, verify that the card was installed correctly and is recognized by the system.

```
$ lspci | grep -i mellanox
```

The output should show all four InfiniBand cards.

Example

```
05:00.0 Infiniband controller: Mellanox Technologies MT27700 Family [ConnectX-4]
0c:00.0 Infiniband controller: Mellanox Technologies MT27700 Family [ConnectX-4]
84:00.0 Infiniband controller: Mellanox Technologies MT27700 Family [ConnectX-4]
8b:00.0 Infiniband controller: Mellanox Technologies MT27700 Family
[ConnectX-4]
```

If all four cards are not reported, then the card was not installed properly and should be resealed.

If a card other than the officially supported Mellanox family of adapters appears, contact NVIDIA Enterprise Support.

2. Verify that the InfiniBand drivers are present.

```
$ lsmod | grep -i ib_
```

The output should be a list of lb_ and mlx_ driver components.

Example

```
ib_ucm 20480 0
ib_ipoib 131072 0
ib_cm 45056 3 rdma_cm,ib_ucm,ib_ipoib
ib_uverbs 73728 2 ib_ucm,rdma_ucm
ib_umad 24576 0
mlx5_ib 192512 0
mlx4_ib 192512 0
ib_sa 36864 5 rdma_cm,ib_cm,mlx4_ib,rdma_ucm,ib_ipoib
ib_mad 57344 4 ib_cm,ib_sa,mlx4_ib,ib_umad
ib_core 143360 13
rdma_cm,ib_cm,ib_sa,iw_cm,nv_peer_mem,mlx4_ib,mlx5_ib,ib_mad,ib_ucm,ib_umad,ib_uverbs,rdma_ucm
ib_addr 20480 3 rdma_cm,ib_core,rdma_ucm
ib_netlink 16384 3 rdma_cm,iw_cm,ib_addr
mlx4_core 344064 2 mlx4_en,mlx4_ib
mlx5_core 524288 1 mlx5_ib
mlx_compat 16384 18
rdma_cm,ib_cm,ib_sa,iw_cm,mlx4_en,mlx4_ib,mlx5_ib,ib_mad,ib_ucm,ib_netlink,ib_addr,ib_core,ib_
```

3. Verify that the OFED software was installed correctly.

```
$ modinfo mlx5_core | grep -i version | head -1
```

Example output:

```
Version : 3.4-1.0.0
```



Note: DGX OS Server release 1.x should have OFED software 3.2.

DGX OS Server release 2.x should have OFED software 3.4.

DGX OS Server release 3.x should have OFED software 4.0.

DGX OS Server release 4.x should have OFED software 4.4.

4. Restart the InfiniBand services so that the new card is recognized.

a). Restart the InfiniBand service.

```
$ sudo service openibd restart
```

b). Verify that the service has started.

```
$ service openibd status
openibd start/running
```

c). If the services do not start, verify

- ▶ That the drivers are loaded according to step 3.
- ▶ That the associated cables are connected to the InfiniBand ports.
- ▶ The state of ibstat (refer to step 7)
- ▶ Whether errors are reported in /var/log/syslog.

If these steps do not indicate a problem and yet the services still do not start, contact NVIDIA Enterprise Support and obtain an RMA for the card.

5. Verify the firmware version.

```
$ cat /sys/class/infiniband/mlx5*/fw_ver
```

Example output:

```
12.17.1010
12.17.1010
12.17.1010
12.17.1010
```

The latest InfiniBand firmware version supported for each DGX OS Server release is as follows:

- ▶ Release 1.x: Firmware version 12.16.1020
- ▶ Release 2.x: Firmware version 12.17.1010
- ▶ Release 3.x: Firmware version 12.18.1000
- ▶ Release 4.x: Firmware version 12.23.1020

6. If you need to update the firmware, follow these steps:

a). Initiate the firmware update.

```
$ sudo /opt/mellanox/mlnx-fw-updater/mlnx_fw_updater.pl
```

The script will check the firmware version of each card and update where needed. If the firmware is updated for any card, you will need to reboot the system for the changes to take effect.

b). Reboot the system if instructed.

c). After rebooting the system, verify that all the Mellanox InfiniBand cards are using the current firmware.

```
$ cat /sys/class/infiniband/mlx5*/fw_ver
12.17.1010
12.17.1010
12.17.1010
12.17.1010
```

7. Verify the physical port state for the InfiniBand cards.

```
$ ibstat
```

In the output text, verify that the Physical State for each card with a cable connection is **LinkUp** and that the port for the card is configured with a GUID. The following example output shows one card in a non-connected state, and three cards in a connected state. Relevant text is highlighted in bold.

```
CA 'mlx5_0'
CA type: MT4115
```



```

Number of ports: 1
Firmware version: 12.17.1010
Hardware version: 0
Node GUID: 0x248a0703000de288
System image GUID: 0x248a0703000de288
Port 1:
    State: Down
    Physical state: Polling
    Rate: 10
    Base lid: 65535
    LMC: 0
    SM lid: 0
    Capability mask: 0x2651e848
    Port GUID: 0x248a0703000de288
    Link layer: InfiniBand
CA 'mlx5_1'
CA type: MT4115
Number of ports: 1
Firmware version: 12.17.1010
Hardware version: 0
Node GUID: 0x248a0703000de26c
System image GUID: 0x248a0703000de26c
Port 1:
    State: Initializing
    Physical state: LinkUp
    Rate: 100
    Base lid: 65535
    LMC: 0
    SM lid: 0
    Capability mask: 0x2651e848
    Port GUID: 0x248a0703000de26c
    Link layer: InfiniBand
CA 'mlx5_2'
CA type: MT4115
Number of ports: 1
Firmware version: 12.17.1010
Hardware version: 0
Node GUID: 0x248a0703001effde
System image GUID: 0x248a0703001effde
Port 1:
    State: Initializing
    Physical state: LinkUp
    Rate: 100
    Base lid: 65535
    LMC: 0
    SM lid: 0
    Capability mask: 0x2651e848
    Port GUID: 0x248a0703001effde
    Link layer: InfiniBand
CA 'mlx5_3'
CA type: MT4115
Number of ports: 1
Firmware version: 12.17.1010
Hardware version: 0
Node GUID: 0x7cfe900300118f22
System image GUID: 0x7cfe900300118f22
Port 1:
    State: Initializing
    Physical state: LinkUp
    Rate: 100
    Base lid: 65535
    LMC: 0
    SM lid: 0
    Capability mask: 0x2651e848
    Port GUID: 0x7cfe900300118f22
    Link layer: InfiniBand

```

See the section [Changing the InfiniBand Port Type](#) for instructions on switching the port to InfiniBand or Ethernet, if required.

6.7. Secure Data Deletion of the SSDs

This section explains how to securely delete data from the NVIDIA DGX-1 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.



Note: These instructions apply only to SSDs that are part of a virtual disk (RAID 0 array).

There are two methods you can use to securely delete data on the SSDs.

- ▶ Method A: Using the on-board MegaRAID controller
- ▶ Method B: Using the StorCLI Erase Command

6.7.1. Method A: Using the MegaRAID Controller

This method uses the on-board MegaRAID controller.

1. Boot the NVIDIA DGX-1 server and at the NVIDIA logo boot screen, press **[F2]** or **[Del]** to enter the BIOS setup screen.
2. Select the **Advanced** tab from the top menu and then scroll down and select the **MegaRAID Configuration Utility**.
3. Select **View Server Profile**.
4. Select **Virtual Drive Management**.
5. Select **Virtual Drive 0**.
6. Select **Operation** and press **[Enter]**, then select **Delete Virtual Drive**.
7. Select the "Thorough" erase mode, then select **Go** and press **[Enter]**.
8. Select **Confirm** and then select **Enabled** at the *Confirm* popup.
9. Return to the MegaRAID Configuration Utility and perform the same steps, selecting **Virtual Drive 1** at step 6.

6.7.2. Method B: Using the StorCLI Secure Erase Commands

This method uses the StorCLI Erase commands. StorCLI should be included in the ISO image that you'll need.

For instructions on preparing a bootable installable medium that contains the current DGX OS server ISO image, see

- ▶ [Obtaining the DGX-1 Software ISO Image and Checksum File](#)

► [Creating a Bootable Installation Medium](#)

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. Install the StorCLI package.

At the root prompt, issue the following.

```
# udpkg -i /cdrom/extras/pool/restricted/s/storcli/storcli_1.20.15_all.deb
```

5. Switch to the StorCLI folder.

```
# cd /opt/MegaRAID/storcli
```

6. Confirm that virtual drives 0 and 1 are in the system (boot and cache volumes respectively).

```
# ./storcli64 /c0 /vall show
```

Example output snippet.

```
Controller = 0
Status = Success
Description = None
```

```
Virtual Drives :
=====
```

```
-----
OC/VD  TYPE    State Access Consist Cache CaC sCC   Size Name
-----
0/0    RAID0   Opt1  RW      Yes    RWTD  -   ON   446.625 GB
0/1    RAID0   Opt1  RW      Yes    RWTD  -   ON    6.983 TB
...
```

7. Erase the volume `/c0 /v0`.

- a). Erase the volume.

```
# ./storcli64 /c0 /v0 start erase thorough
```

- b). Confirm the erase progress.

```
# ./storcli64 /c0 /v0 show erase
```

8. Erase the volume `/c0 /v1`.

This step can be performed even though the previous erase step is still in progress.

- a). Erase the volume.

```
# ./storcli64 /c0 /v1 start erase thorough
```

- b). Confirm the erase progress.

```
# ./storcli64 /c0 /v1 show erase
```

Chapter 7. Installing Software on Air-Gapped NVIDIA DGX-1 Systems

For security purposes, some installations require that systems be isolated from the internet or outside networks. Since most DGX-1 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.

7.1. Installing NVIDIA DGX-1 Software

One method for updating DGX-1 software on an air-gapped DGX-1 is to download the ISO image, copy it to removable media and then re-image the DGX-1 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-1 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.

7.1.1. Re-Imaging the System



Note: WARNING: This process destroys all data and software customizations that you have made on the DGX-1. 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 Enterprise Support site.
 - a). Log on to the [NVIDIA Enterprise Support \(https://nvid.nvidia.com/dashboard/\)](https://nvid.nvidia.com/dashboard/) site and click the Announcements tab to locate the DGX-1 Base OS Image ISO file.
 - b). Download the image ISO file.
2. Refer to the instructions in the [Restoring the DGX-1 Software Image](#) section for additional instructions.

7.1.2. 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 systems. The steps are specific to versions 4.0.X and 4.1.X, but they can be edited to work with other versions. 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>. For more information on how to upgrade from versions 4.0.x to 4.1.x, review the release notes: <https://docs.nvidia.com/dgx/pdf/DGX-OS-server-4.1-relnotes-update-guide.pdf>.



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.

7.1.2.1. Creating the Mirror in a DGX OS 4 System

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

Here are the prerequisites:

- ▶ A system installed with Ubuntu OS is needed to create the mirror because there are several Ubuntu tools that need to be used.
- ▶ You must be logged in to the system installed with Ubuntu OS as an administrator user because this procedure requires `sudo` privileges.
- ▶ The system must contain enough storage space to replicate the repositories to a file system. The space requirement could be as high as 250 GB.
- ▶ An efficient way to move large amount of data is needed, 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 that any portable drives are formatted using ext4 or FAT32.

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

Example mount point used in these instructions: `/media/usb/repository`

2. Install the `apt-mirror` package.

```
$ sudo apt update
$ sudo apt install apt-mirror
```

3. Change the ownership of the target directory to the `apt-mirror` user in the `apt-mirror` group.

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

The target directory must be owned by the user `apt-mirror` or the replication will not work.

4. 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 and 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 http://international.download.nvidia.com/dgx/repos/bionic bionic-
r450+cuda11.0 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
```

5. 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
```

6. Eject the removable storage with all packages.

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

7.1.2.2. Configuring the Target Air-Gapped DGX OS 4 System

The instructions in this section are to be performed on the target air-gapped DGX system.

Here are the prerequisites:

- ▶ The target air-gapped DGX system is installed, has gone through the first boot process, and is ready to be updated with the latest packages.
- ▶ The USB storage device on which the mirrors were created is attached to the target DGX 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 the `apt` (<http://manpages.ubuntu.com/manpages/bionic/en/man8/apt.8.html>) command to use the file system 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 removing it will eliminate error messages during the update process.

5. **(For DGX OS Release 4.1 and later only)** Configure `apt` to use the NVIDIA DGX OS packages in the file `/etc/apt/sources.list.d/dgx-bionic-r418-cuda10-1-repo.list`.

```
$ echo "deb file:///media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic/ bionic-r418+cuda10.1 main multiverse restricted universe" | sudo tee /etc/
apt/sources.list.d/dgx-bionic-r418-cuda10-1-repo.list
```

6. **Optional: (For DGX OS Release 4.5 and later only)** If you want to use the R450 NVIDIA graphics driver and CUDA Toolkit 11.0, configure `apt` to use the NVIDIA DGX OS packages in the file `/etc/apt/sources.list.d/dgx-bionic-r450-cuda11-0-repo.list`.

```
$ echo "deb file:///media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic/ bionic-r450+cuda11.0 main multiverse restricted universe" | sudo tee /etc/
apt/sources.list.d/dgx-bionic-r450-cuda11-0-repo.list
```



Note: If you want to continue using earlier releases, for example the R418 NVIDIA graphic driver and CUDA Toolkit 10.1, omit this step.

7. 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
```

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

```
$ sudo apt update
```

Output from this command is similar to the following example.

```
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-r450+cuda11.0 InRelease [7070 B]
Get:5 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic-r450+cuda11.0 InRelease [7070 B]
Get:6 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:6 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic InRelease [13.1 kB]
Hit:7 https://download.docker.com/linux/ubuntu bionic InRelease
Get:8 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic-r418+cuda10.1/multiverse amd64 Packages [10.1 kB]
Get:9 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic-r450+cuda11.0/multiverse amd64 Packages [17.4 kB]
Get:10 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic-r418+cuda10.1/restricted amd64 Packages [10.3 kB]
Get:11 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic-r450+cuda11.0/restricted amd64 Packages [26.4 kB]
Get:12 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic-r418+cuda10.1/restricted i386 Packages [516 B]
Get:13 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic/multiverse amd64 Packages [44.5 kB]
Get:14 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic/multiverse i386 Packages [8,575 B]
Get:15 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic/restricted i386 Packages [745 B]
Get:16 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic/restricted amd64 Packages [8,379 B]
Get:17 file:/media/usb/repository/mirror/international.download.nvidia.com/dgx/
repos/bionic bionic/universe amd64 Packages [2,946 B]
Get:18 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.
$
```

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

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

If you configured apt to use the NVIDIA DGX OS packages in the file `/etc/apt/sources.list.d/dgx-bionic-r450-cuda11-0-repo.list`, the NVIDIA graphics driver is upgraded to the R450 driver and the package sources are updated to obtain future updates from the R450 driver repositories.

10. **Optional: (For DGX OS Release 4.5 and later only)** If you configured apt to use the NVIDIA DGX OS packages in the file `/etc/apt/sources.list.d/dgx-bionic-r450-cuda11-0-repo.list` and want to use CUDA Toolkit 11.0, install it.

```
$ sudo apt install cuda-toolkit-11-0
```



Note: If you did not configure apt to use the NVIDIA DGX OS packages in the file `/etc/apt/sources.list.d/dgx-bionic-r450-cuda11-0-repo.list`, omit this step. If you try to install CUDA Toolkit 11.0, the attempt fails.

7.2. Installing Docker Containers

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

1. On a system with internet access, log in to the DGX 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 DGX 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
```

Chapter 8. Customer Support for the NVIDIA DGX-1

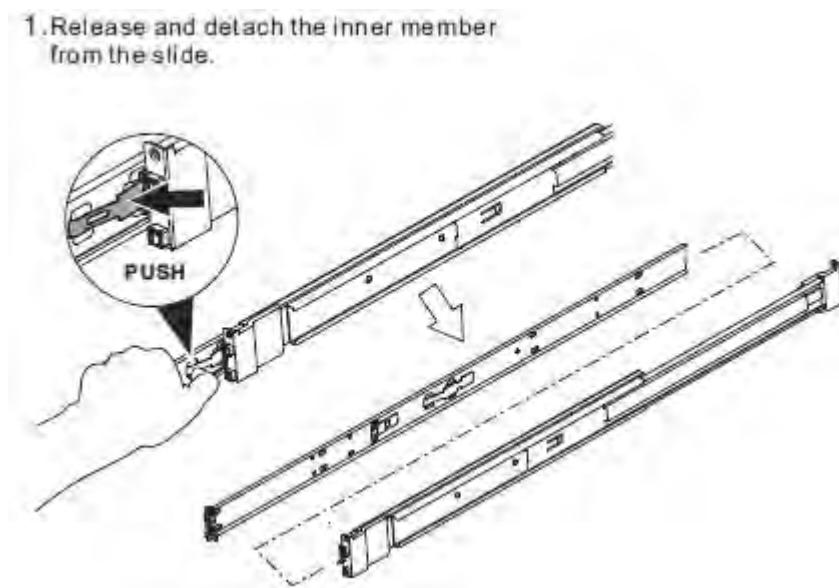
Contact NVIDIA Enterprise Support for assistance in reporting, troubleshooting, or diagnosing problems with your DGX-1 system. Also contact NVIDIA Enterprise Support for assistance in installing or moving the DGX-1 system.

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

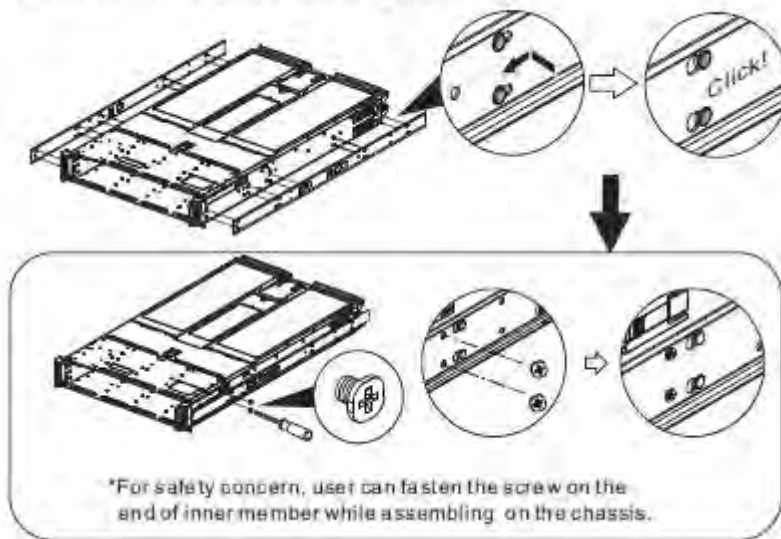
Our support team can help collect appropriate information about your issue and involve internal resources as needed.

Chapter 9. King Slide - AH61-500 Instructions

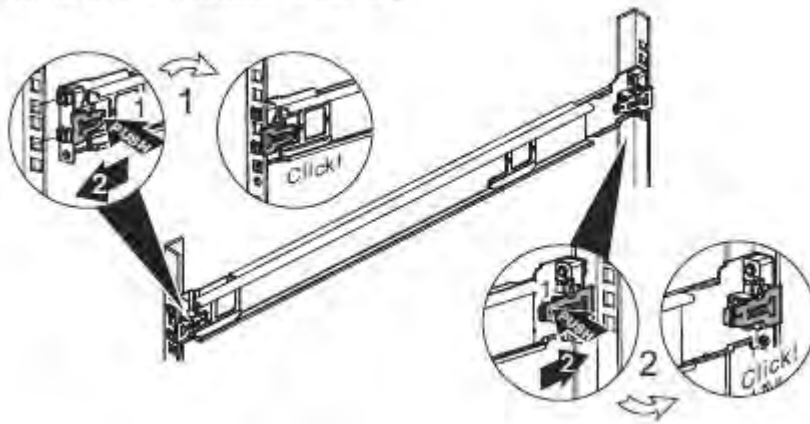
These are the rail assembly instructions for the rail kit provided with the DGX-1.



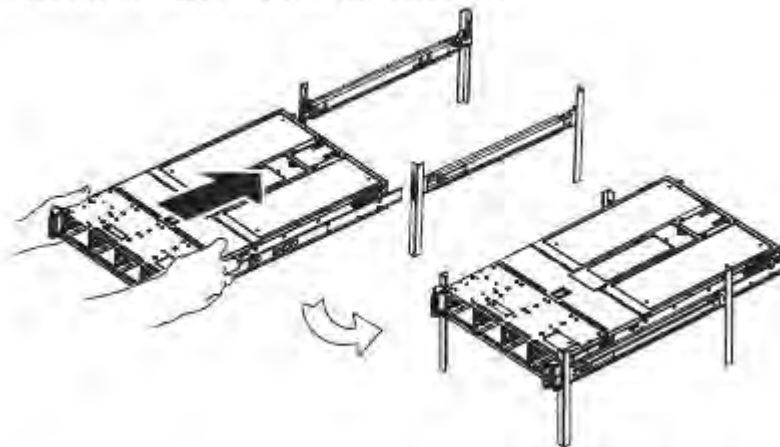
2. Attach inner member to the system.



3. Attach outer member to the rack.



4. Insert the unit to complete the installation.



All images copyright 2016 King Slide. All Rights Reserved

Chapter 10. Safety

To reduce the risk of bodily injury, electrical shock, fire, and equipment damage, read this document and observe all warnings and precautions in this guide before installing or maintaining your 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 or 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.

10.1. 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.

Symbol	Meaning
	Indicates do not touch fan blades, may result in injury.
	Indicates to unplug all AC power cord(s) to disconnect AC power.
	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.
	Indicates two people are required to safely handle the system

10.2. 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.

10.3. 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.

10.4. 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 .

10.5. 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, SV 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. Your system may use more than one AC power cord. Make sure all AC power cords are unplugged. Make sure the AC power cord(s) is/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, turn off the server and disconnect the power cord, telecommunications systems, networks, and modems attached to the server before opening it.

Power Cord Warnings

Use certified AC power cords to connect to both the power distribution unit (PDU) and server system installed in your rack.

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.

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

- ▶ 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.

10.6. 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.

10.7. 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).Topic paragraph

10.8. Electrostatic Discharge

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.

10.9. Other Hazards

PROPOSITION 65 WARNING

This product contains chemicals known to the State of California to cause cancer and birth defects or other reproductive harm.

CALIFORNIA DEPARTMENT OF TOXIC SUBSTANCES CONTROL

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

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

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.

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.

Chapter 11. Compliance

The NVIDIA DGX-1 is compliant with the regulations listed in this section.

11.1. 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.

11.2. United States / Canada

cULus Listing Mark



11.3. Canada

Industry Canada (IC)

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

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

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

11.4. 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.

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

This device complies with the following Directives:

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

A copy of the Declaration of Conformity to the essential requirements may be obtained directly from NVIDIA GmbH (Floessergasse 2, 81369 Munich, Germany).

11.5. Japan



この装置は、クラス A 情報技術装置です。この装置を家庭環境で使用する
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

11.6. Australia

RCM



11.7. China

RoHS Material Content



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					
	铅及其化合物 (Pb and its compounds)	汞及其化合物 (Hg and its compounds)	镉及其化合物 (Cd and its compounds)	六价铬化合物 (Cr(VI) compounds)	多溴联苯 (PBB)	多溴二苯醚 (PBDE)
机箱 Chassis	X	O	O	O	O	O
印刷电路部件 PCA	X	O	O	O	O	O
图形处理器 GPU	X	O	O	O	O	O
中央处理器 CPU	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.

11.8. Israel

SII

ודא שלמות ותקינות כבל החשמל והתקע אין להכניס או להוציא את התקע מרשת החשמל בידיים רטובות. אין לפתוח את המכשיר, במקרה של בעיה כלשהי יש לפנות למעבדת השיחות הקרובה. יש להרחיק את המכשיר מנזלים. במקרה של ריח מוזר, רעשים שמקורם במכשיר, יש לנתקו מיידית מרשת החשמל ולפנות למעבדת שיחות המכשיר מיועד לשימוש בתוך המבנה, ולא לשימוש חיצוני ולא לשימוש בסביבה לחה. אין לחתוך, לשבור, ולעקם את הכבל החשמל. אין להניח חפצים על הכבל החשמל או להניח לו להתחמם יתר על המידה, שכן עלול לגרום לנזק, דליקה או התחשמלות. יש להקפיד לחזק את התקן הניתוק במצב תפעולי מוכן לשימוש. אזהרה: אין להחליף את כבל הזינה בתחליפים לא מקוריים, חיבור לקוי עלול לגרום להתחשמלות המשתמש. בשימוש על כבל מאריך יש לוודא תקינות מוליך הארקה שבכבל.

11.9. South Korea

KC



MSIP-REM- QCT-P2787 (A)

<p>A급 기기 (업무용 방송통신기자재)</p>	<p>이 기기는 업무용(A급) 전자파적합기기로서 판매자 또는 사용자는 이 점을 주의하시기 바라며, 가정외의 지역에서 사용하는 것을 목적으로 합니다.</p>
--------------------------------	--

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.

11.10. India

BIS



Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA accepts no liability related to any default, damage, costs, or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, DGX, DGX-1, DGX-2, and DGX Station are trademarks and/or registered trademarks of NVIDIA Corporation in the United States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2020 NVIDIA Corporation. All rights reserved.