TQ8x00 MetroX®-2 HDR 200Gb/s InfiniBand Switch Systems User Manual

Table of Contents

Relevant for Models: MTQ8100 and MTQ8200

This manual describes the installation and basic use of the NVIDIA 1U HDR InfiniBand switch systems based on the NVIDIA Quantum™ switch ASIC. This manual is intended for IT managers and system administrators.

Ordering Information

System Model	NVIDIA SKU	Legacy OPN	Description	Lifecycle Phase
MTQ8100	MTQ8100-HS2F	920-9B110-00 FH-0M2	Quantum MetroX®-2 10km long reach 100Gb/s switch, 2 long reach QSFP28 ports, 8 standard HDR ports, 2 Power Supplies (AC), x86 dual core, standard depth, P2C airflow, Rail Kit	Mass Production
MTQ8200	MTQ8200-HS2F	920-9B110-00 FH-0M1	Quantum MetroX®-2 40km long reach 100Gb/s switch, 2 long reach QSFP28 ports, 8 standard HDR ports, 2 Power Supplies (AC), x86 dual core, standard depth, P2C airflow, Rail Kit	Mass Production

Related Documentation

Document	Description
InfiniBand Architecture Specification Release 1.3.1, November 2, 2016 and Vol. 2, Release 1.4	The InfiniBand Trade Association (IBTA) InfiniBand® specification at https:// www.infinibandta.org.

Document	Description
MLNX-OS® User Manual	This document contains information regarding the configuration and management of the MLNX-OS® software. See https://docs.nvidia.com/networking/category/mlnxos.
Hands-on workshops	https://academy.nvidia.com/en/infiniband-customized-training/
NVIDIA Networking Support	See https://www.nvidia.com/en-us/networking/support/.

Revision History

A list of the changes made to this document are provided in Document Revision History.

Introduction

NVIDIA Quantum MetroX®-2 systems enable seamless connectivity between remote InfiniBand data centers or storage platforms. MetroX-2 systems extend InfiniBand connectivity to 10 or 40 kilometers, enabling high data throughput, native remote direct memory access (RDMA) communications, advanced routing, and NVIDIA Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)™. MetroX-2 users can easily migrate application jobs from one InfiniBand center to another, or combine the compute power of multiple remote data centers together for higher overall performance and scalability.

Built with the NVIDIA Quantum InfiniBand switch device, the NVIDIA Quantum MetroX-2 solution provides up to 2 EDR InfiniBand long-reach ports and 8 EDR/HDR local InfiniBand ports, delivering up to 100Gb/s data throughout for distances of 10 and 40 kilometers. NVIDIA Quantum MetroX-2 enables aggregate data and storage networking over a single, consolidated fabric. MetroX-2 RDMA technology guarantees high-performance, high-volume data-sharing between distant sites, enabling existing data centers with expansion, disaster recovery, data mirroring and campus connectivity. Designed for today's business continuity and simplified disaster recovery design, NVIDIA Quantum MetroX-2 delivers zero Recovery Point Objective (RPO), providing the benefit of high bandwidth and simple high availability management connectivity for up to 40 kilometers.

NVIDIA TQ8100-HS2F supports up to 2 EDR QSFP28 MetroX-2 ports running 100Gb/s for up to 10 km, and up to 8 standard HDR local ports running 200Gb/s. NVIDIA Quantum MetroX-2 TQ8200-HS2F supports up to 2 EDR QSFP28 MetroX-2 ports running 100Gb/s for up to 40 km, and up to 8 standard HDR local ports running 200Gb/s.

TQ8100/TQ8200 Front View

TQ8100/TQ8200 Rear View

For additional airflow options, see Airflow.

Speed and Switching Capabilities

The table below describes maximum throughput and interface speed per system model.

System Model	Interfaces	Max Throughput
TQ8100	2 x QSFP28 ports	4Tb/s
	8 x QSFP56 ports	
TQ8200	2 x QSFP28 ports	
	8 x QSFP56 ports	

Management Interfaces, PSUs and Fans

The table below lists the various management interfaces and available replacement parts per system model.

System Model	USB	MGT	I²C	Console	Replaceable PSU	Replaceable Fan
TQ8100	Front (micro USB)	Front (1 port)	Front (through the micro USB port)	Front	Yes, 2	Yes, 6
TQ8200	Front (micro USB)	Front (1 port)	Front (through the micro USB port)	Front	Yes, 2	Yes, 6

Features

For a full feature list, please refer to the system's product brief. Go to http://www.mellanox.com. In the main menu, click on Products > InfiniBand/VPI Switch Systems, and select the desired product page.

Certifications

The list of certifications (such as EMC, Safety and others) per system for different regions of the world is located on the NVIDIA website at http://www.mellanox.com/page/environmental_compliance.

Installation

System Installation and Initialization

Installation and initialization of the system require attention to the normal mechanical, power, and thermal precautions for rack-mounted equipment.

> ⚠ The rack mounting holes conform to the EIA-310 standard for 19-inch racks. Take precautions to guarantee proper ventilation in order to maintain good airflow at ambient temperature.

> ⚠ • Unless otherwise specified, NVIDIA products are designed to work in an environmentally controlled data center with low levels of gaseous and dust (particulate) contamination.
> • The operation environment should meet severity level G1 as per ISA 71.04 for gaseous contamination and ISO 14644-1 class 8 for cleanliness level.

The installation procedure for the system involves the following phases:

Step	Procedure	See
1	Follow the safety warnings	Safety Warnings
2	Pay attention to the air flow consideration within the system and rack	Air Flow
3	Make sure that none of the package contents is missing or damaged	Package Contents
4	Mount the system into a rack enclosure	19" System Mounting Options
5	Power on the system	Initial Power On
6	Perform system bring-up	System Bring-Up of Managed Systems
7	[Optional] FRU replacements	FRU Replacements

Safety Warnings

Prior to the installation, please review the Safety Warnings. Note that some warnings may not apply to all models.

Air Flow

NVIDIA systems are offered with two air flow patterns:
- Power (rear) side inlet to connector side outlet - marked with blue power supplies/fans FRUs' handles.
  Air Flow Direction Marking - Power Side Inlet to Connector Side Outlet

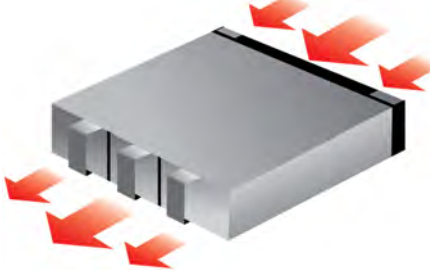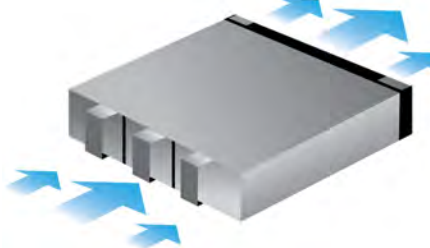- Connector (front) side inlet to power side outlet - marked with red power supplies/fans FRUs' handles.

Air Flow Direction Marking - Connector Side Inlet to Power Side Outlet

> ⚠ Certain systems are offered with one airflow pattern only (P2C/C2P). To view the airflow patterns offer, please refer to the Ordering Information.

> ❗ • All servers and systems in the same rack should be planned with the same airflow direction.
> • All FRU components need to have the same air flow direction. A mismatch in the air flow will affect the heat dissipation.

The table below provides an air flow color legend and respective OPN designation.

Direction	Description and OPN Designation
	Connector side inlet to power side outlet. Red latches are placed on the power inlet side. OPN designation is "-R".
	Power side inlet to connector side outlet. Blue latches are placed on the power inlet side. OPN designation is "-F".

Package Contents

Before installing your new system, unpack it and check against the parts list below that all the parts have been sent. Check the parts for visible damage that may have occurred during shipping.

The TQ8100 and TQ8200 package content is as follows:

- 1 – System

- 1 – Rail kit
- 2 – Power cables Type C13-C14
- 1 x Harness: HAR000631 – Harness RS232 2M cable – DB9 to RJ-45
- 2 – Cable retainers
- 1 – Quick Start Guide

> ⚠ If anything is damaged or missing, contact your sales representative at Networking-support@nvidia.com.

19" System Mounting Options

By default, the systems are shipped with the static rail kit described in Fixed Rail Kit.

Fixed Rail Kit

> ⓘ This document is preliminary and subject to change.

Kit OPN	Legacy Kit OPN	Rack Size and Rack Depth Range
930-9BRKT-00JF-000	MTEF-KIT-C	430-800 mm

The following parts are included in the fixed rail kit (see figure below):
- 2x Rack mount rails (A)
- 2x Rack mount blades (B)
- 2x Rack mount brackets (C)
- 8x M6 Standard cage nuts (D)
- 8x M6 Standard pan-head Phillips screws (E)
- 4x Flat Head Phillips 100 DEG 6-32X1/4" ST.ST PATCH 360 (F)

Rack Rail Kit Parts
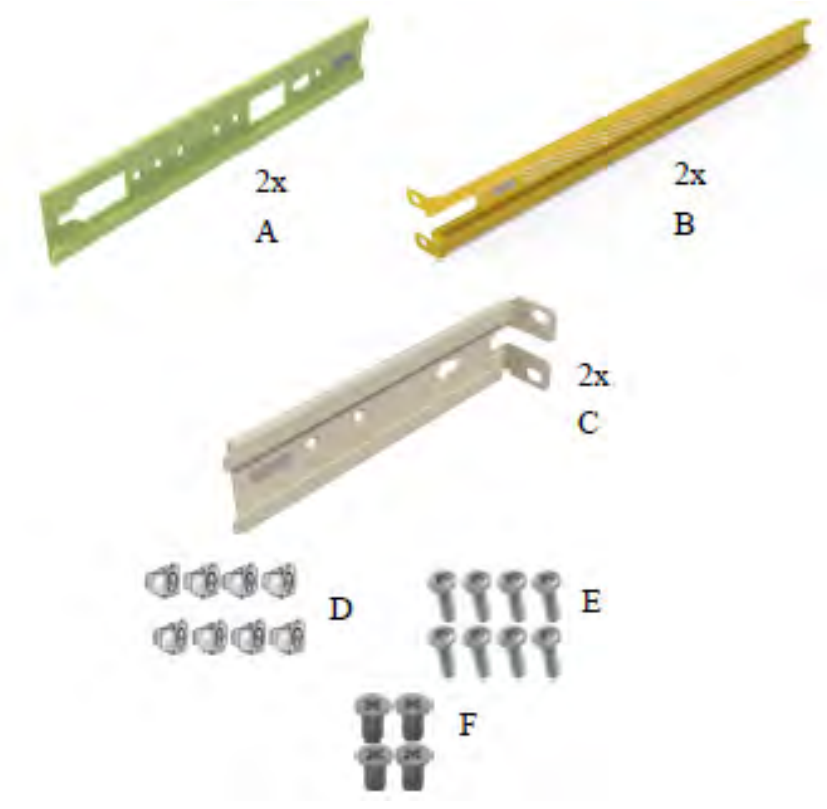
Prerequisites:

Before mounting the system to the rack, select the way you wish to place the system. Pay attention to the airflow within the rack cooling, connector and cabling options.

While planning how to place the system, consider the two installation options shown in the figures below, and review the following points:

- Make sure the system air flow is compatible with your installation selection. It is important to keep the airflow within the rack in the same direction.
- Note that the part of the system to which you choose to attach the rails (the front panel direction, as demonstrated in Option 1 or the FRUs direction, as demonstrated in Option 2) will determine the system's adjustable side. The system's part to which the brackets are attached will be adjacent to the cabinet.
- The FRU side is extractable. Mounting the rack brackets inverted to the FRU side (Option 2) will allow you to slide the FRUs, in and out.

Short Racks (430-580 mm) Installation Options

Front Side	Rear Side
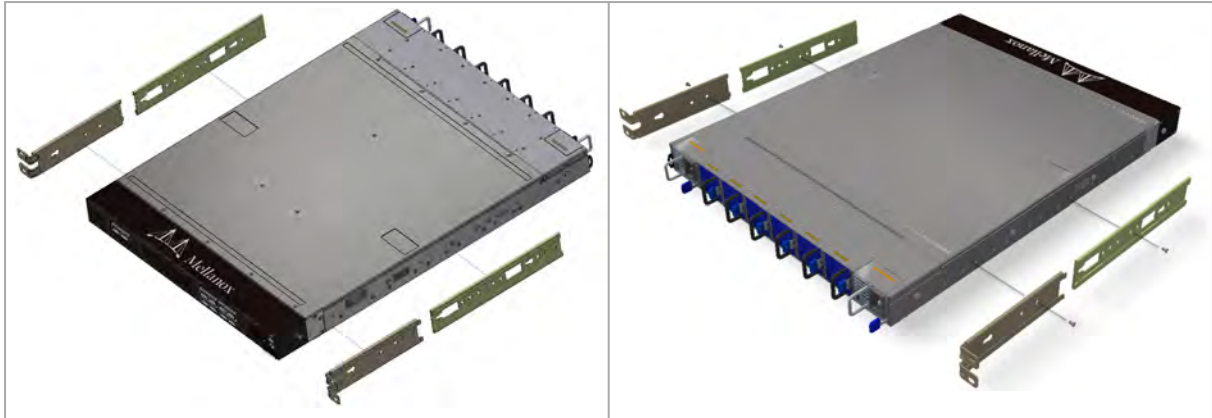
Standard Racks (580-800 mm) Installation Options

Front Side	Rear Side
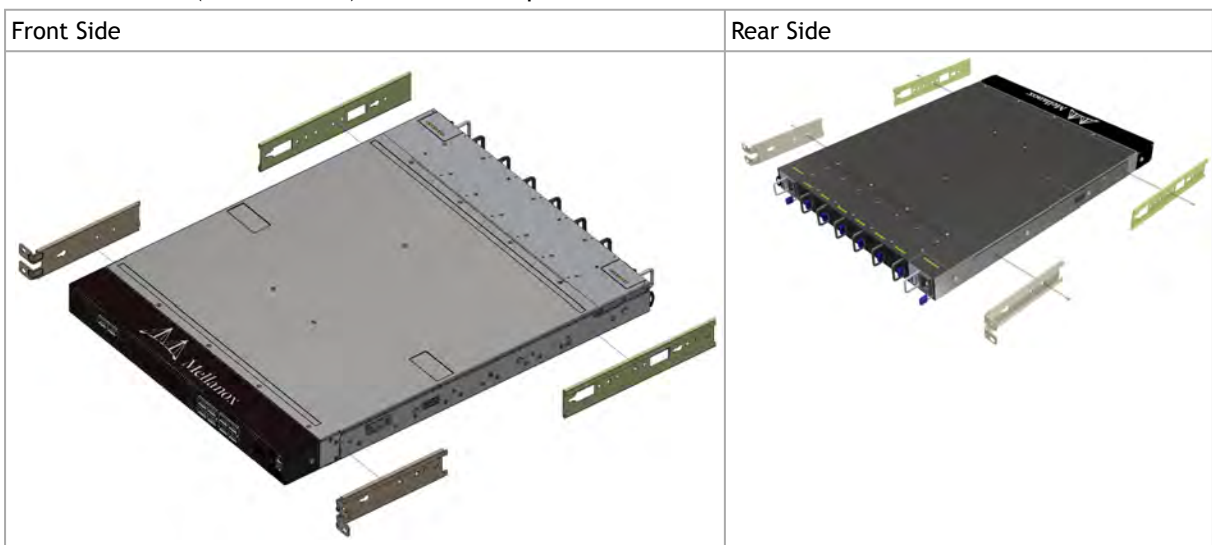
> ⚠ In short racks, the system's ventilation openings should be framed by the designated windows in the rails, as shown below.
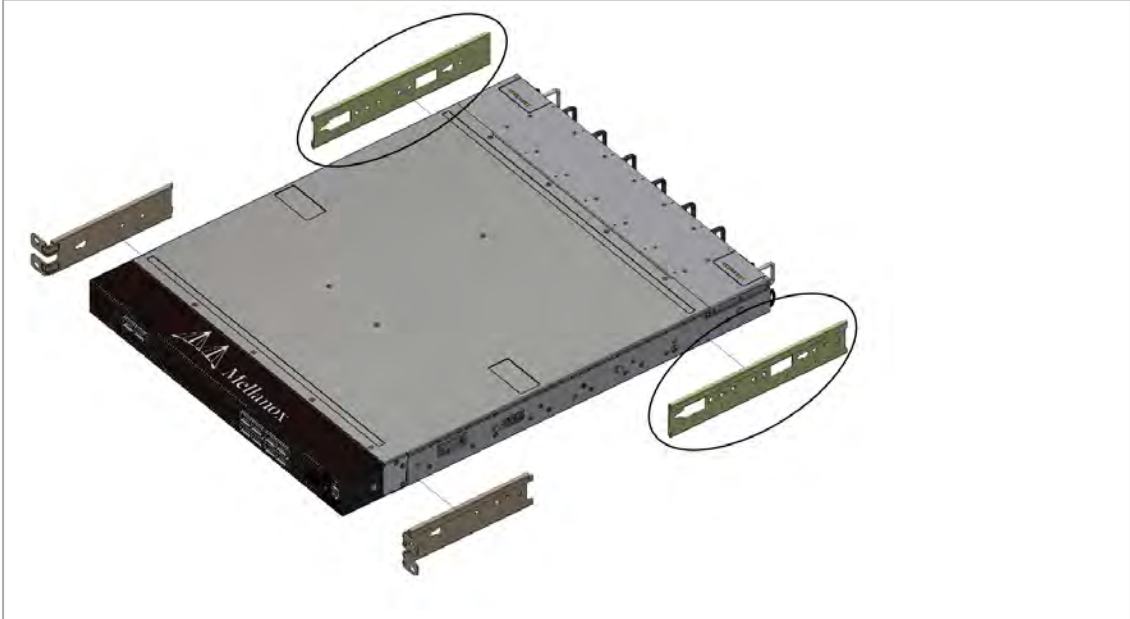
To mount the system into the rack:

> ❗ At least two people are required to safely mount the system in the rack.

> ⚠ The following steps include illustrations that show front side (ports) installation, yet all instructions apply to all installation options.

1. Attach the left and right rack mount rails (A) to the switch, by gently pushing the switch chassis' pins through the slider key holes, until locking occurs.
2. Secure the chassis in the rails by screwing 2 flat head Phillips screws (F) in the designated points with a torque of 1.5±0.2 Nm.

Attaching the Rails to the Chassis

3. Attach the left and right rack mount brackets (C) to the switch, by gently pushing the switch chassis' pins through the slider key holes, until locking occurs. Secure the system in the brackets by screwing the remaining 2 flat head Phillips screws (F) in the designated points with a torque of 1.5±0.2 Nm.

Attaching the Brackets to the Chassis
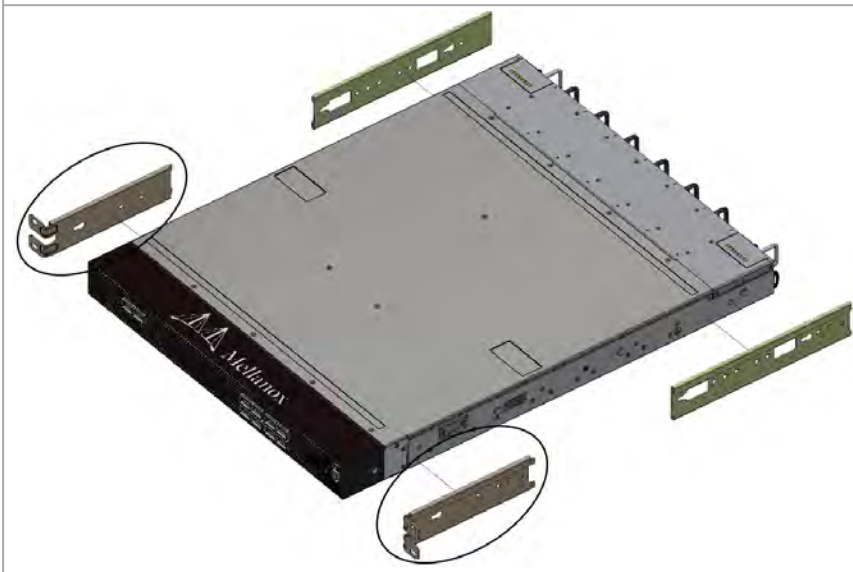
4. Install 8 cage nuts (D) in the desired 1U slots of the rack: 4 cage nuts in the non-extractable side and 4 cage nuts in the extractable side.

Installing the Cage Nuts
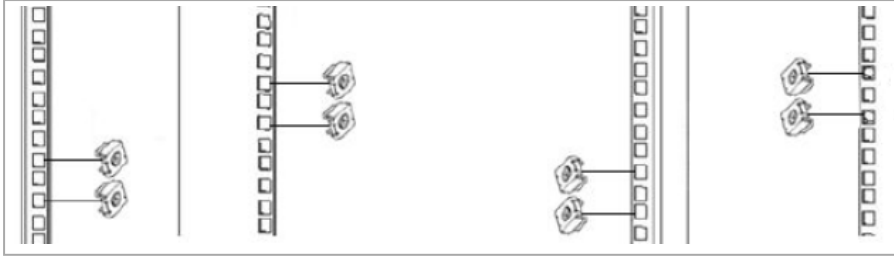
> ⚠ While each rack U (unit) consists of three holes, the cage nut should be installed vertically with its ears engaging the top and bottom holes only.

While your installation partner is supporting the system's weight, perform the following steps:

5. Attach the two rack mount blades (B) to the back side (FRU side) of the rack by inserting four M6 screws (E) in the designated cage nuts. Do not tighten the screws yet.

Attaching the Blades to the Rack
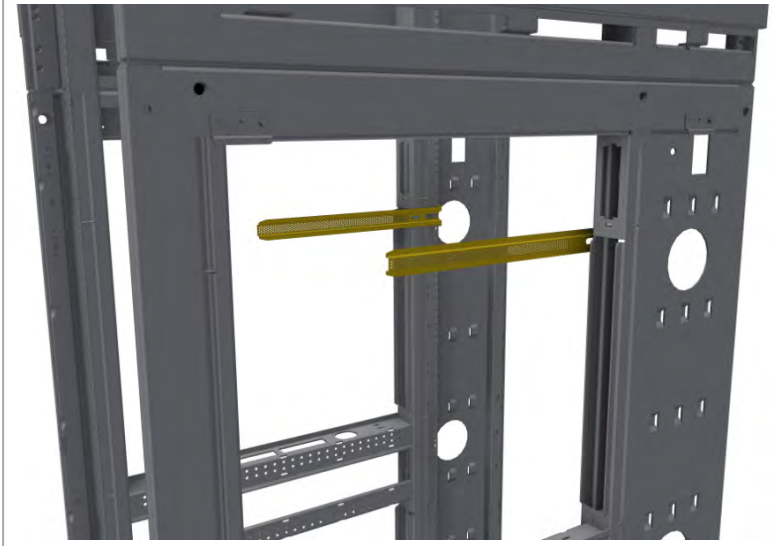
6. Slide the switch with the rails (A) and brackets (C) installed on it into the left and right blades (B) on the rack. Use four M6 screws (E) to fix the rack mount ears (C) to the rack. Do not tighten the screws yet.

Sliding the Blades in the Rails
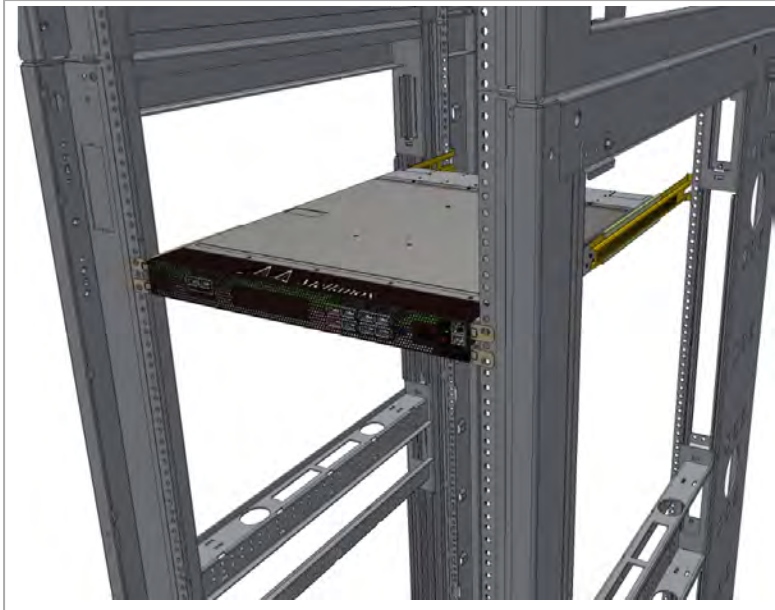
7. When fully inserted, fix the switch by tightening the 8 screws (E) inserted in Step 5 and Step 6 with a torque of 4.5±0.5.

Removing the System from the Rack

To remove a unit from the rack:

1. Turn off the system and disconnect it from peripherals and from the electrical outlet. While your installation partner is supporting the system's weight:
2. Loosen the screws attaching the rack mount brackets (C) to the rack. Do not remove them yet.
3. Loosen the screws attaching the rack mount blades (B) to the rack, and pull the blades towards you, while your partner is holding the system.
4. Extract the loosened screws from Step 2 and dismount the system from the rack.
5. Remove the rails and brackets from the chassis by unscrewing 4 screws.

Cable Installation

> ⓘ This document is preliminary and subject to change.

All cables can be inserted or removed with the unit powered on.

To insert a cable, press the connector into the port receptacle until the connector is firmly seated. The LED indicator, corresponding to each data port, will light when the physical connection is established. When a logical connection is made, the relevant port LED will turn on.

To remove a cable, disengage the locks and slowly pull the connector away from the port receptacle. The LED indicator for that port will turn off when the cable is unseated.

For full cabling guidelines, ask your NVIDIA representative for a copy of *NVIDIA Cable Management Guidelines and FAQs Application Note*.

For more information about port LEDs, refer to Port LEDs.

> ❗ Do not force the cable into the cage with more than 40 newtons / 9.0 pounds / 4kg force. Greater insertion force may cause damage to the cable or to the cage.

Cable Orientation

Splitter (Breakout) Cables and Adapters

> ⚠ The breakout option is intended for users planning to run HDR100 using ConnectX-6 only.

The breakout cable is a unique capability, where a single physical quad-lane QSFP port is divided into 2 dual-lane ports. It maximizes flexibility by enabling end users to use a combination of dual-lane and quad-lane interfaces according to the specific requirements of their network. All system ports may be split into 2-lane ports. Splitting a port changes the notation of that port from x/y to x/y/z with "x/y" indicating the previous notation of the port prior to the split and "z" indicating the number of the resulting single-lane port (1,2). Each sub-physical port is then handled as an individual port. For example, splitting port 5 into 2 lanes gives the following new ports: 1/5/1 & 1/5/2. For the systems splitting options, see "TQ8100/TQ8200 Splitting Options" below.

Sample Breakout Cable

> ✅ • Splitting the interface deletes all configuration on that interface.
> • This feature is available only for Quantum ASIC systems.

- In order to be able to use this feature, the system profile command must be activated with split-ready configuration (cross-reference to system profile command).

For more information on how to change the system's profile to allow Split-Ready configuration, how to change the module type to a split mode, and how to unsplit a split port, please refer to the "InfiniBand Switching" chapter in the latest *MLNX-OS® User Manual*.

TQ8100/TQ8200 Splitting Options

Ports 1-2 have no splitting capabilites (they opeate in EDR only), while ports 3-10 are splittable - Each of them can be split to 2xQSFP56 HDR100 ports, and there are no blocking requirements.
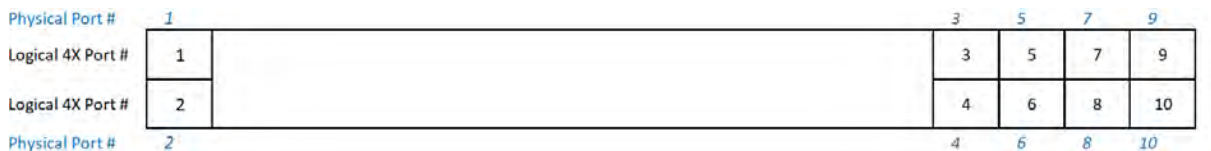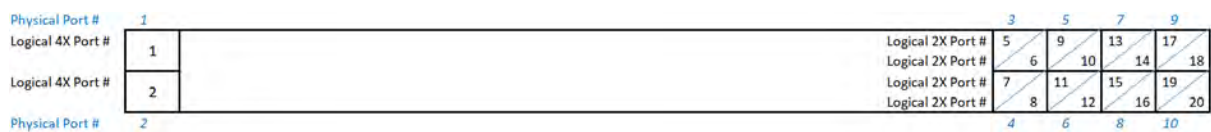

Logical Port Numbering Schematic

Two profiles can be selected for the TQ8x00 switch systems. The first one defines the system as a pure 10-port HDR200 switch. The other profile permits any or all QSFP ports to be split into two 2X (HDR100) ports.
The following diagrams attempt to show how the logical ports map onto the physical QSFP ports, as viewed by the IB tools (e.g. ibnetdiscover):

Switch Profile: Non-Splittable (Suitable for L2/Spine Switches)

Switch Profile: Splittable

⚠ Note: The IB tools report 21 logical ports. Port 21 is an internal port used for the SHARP Aggregation Node when SHARP is enabled.

⚠ Note: MLNX-OS will use Ethernet port numbering for split QSFP ports. For example, 2X ports 17 and 18 above will be named '1/9/1' and '1/9/2'.

Initial Power On

> (i) This document is preliminary and subject to change.

Each system's input voltage is specified in the Specifications chapter.

The power cords should be standard 3-wire AC power cords including a safety ground and rated for 15A or higher.

> ⚠ The system platform will automatically power on when AC power is applied. There is no power system. Check all boards, power supplies, and fan tray modules for proper insertion before plugging in a power cable.

1. Plug in the first power cable.
2. Plug in the second power cable.
3. Wait for the system upload process.

> ⚠ It may take up to five minutes to turn on the system. If the System Status LED shows amber after five minutes, unplug the system and call your NVIDIA representative for assistance.

4. Check the System Status LEDs and confirm that all of the LEDs show status lights consistent with normal operation (initially flashing, and then moving to a steady color) as shown below. For more information, refer to LED Notifications.

System Status LEDs 5 Minutes After Power On

> ⚠ After inserting a power cable and confirming the green System Status LED light is on, make sure that the Fan Status LED shows green. If the Fan Status LED is not green, unplug the power connection and check that the fan module is inserted properly and that the mating connector of the fan unit is free of any dirt and/or obstacles. If no obstacles were found and the problem persists, call your NVIDIA representative for assistance.

Two Power Inlets - Electric Caution Notifications:

> ⚠
> - Risk of electric shock and energy hazard. The two power supply units are independent. Disconnect all power supplies to ensure a powered down state inside of the switch platform.
> - ACHTUNG Gafahr des elektrischen Schocks. Entferrnen des Netzsteckers elnes Netzteils spannungsfrei. Um alle Einhieten spannungsfrei zu machen sind die Netzstecker aller Netzteile zu entfernen.
> - ATTENTION Risque de choc et de danger e'lectriques. Le de'branchment d'une seule alimentation stabilise'e ne de'branch uniquement qu'un module "Alimentation

Stabilise'e". Pour isoler completement le module en cause, Il faut de'brancher toutes les alimentations stabilise'es.

- 電擊與能源危害的危險。所有 PSU 均各自獨立。將所有電源供應器斷電，確保交換器平台內部在電源關閉狀態。

System Bring-Up of Managed Systems

⚠ The bring-up procedures described in this section do not apply to unmanaged/externally managed systems. Such systems are ready for operation after power-on.

In order to query the system, perform firmware upgrade or other firmware operation. Refer to the latest NVIDIA Firmware tools (MFT) located on the NVIDIA web site (https://network.nvidia.com/products/adapter-software/firmware-tools/).

In order to obtain the firmware version of the externally managed system:

1. Run the following command from a host:

```
# flint -d <device> q
```

2. Compare the results of this command with the latest version for your system posted on https://network.nvidia.com/products/adapter-software/firmware-tools/.
3. If the current version is not the latest version, follow the directions in the MFT User Manual to burn the new firmware.

Configuring Network Attributes

The procedures described in this chapter assume that you have already installed and powered-on the system according to the instructions in this document. Since the system comes with a pre-configured DHCP, you may find the explanation in Disable Dynamic Host Configuration Protocol (DHCP) sufficient. In case manual configuration is required, please refer to the instructions in Manual Host Configuration.

Disable Dynamic Host Configuration Protocol (DHCP)

DHCP is used for automatic retrieval of management IP addresses.

If a user connects through SSH, runs the wizard and turns off DHCP, the connection is immediately terminated, as the management interface loses its IP address. In such a case, the serial connection should be used.

⚠   `<localhost># ssh admin@<ip-address>`

```
Mellanox MLNX-OS Switch Management
Password:
Mellanox Switch
Mellanox configuration wizard
```

```
         Do you want to use the wizard for initial configuration? yes
         Step 1: Hostname? [my-switch]
         Step 2: Use DHCP on mgmt0 interface? [yes] no
         <localhost>#
```

Manual Host Configuration

To perform initial configuration of the system:

Step 1. Connect a host PC to the Console RJ45 port of the system, using the supplied harness cable (DB9 to RJ45).

⚠ Make sure to connect to the Console RJ45 port, and not to the (Ethernet) MGT port. Pay attention to the icons:

Console RJ45	IOIOI
Ethernet MGT	器

Step 2. Configure a serial terminal program (for example, HyperTerminal, minicom, or Tera Term) on your host PC with the settings described in the table below. Once you perform that, you should get the CLI prompt of the system.

Serial Terminal Program Configuration

Parameter	Setting
Baud Rate	115200
Data bits	8
Stop bits	1
Parity	None
Flow Control	None

Step 3. Login as admin and use admin as password. On the first login, the MLNX-OS configuration wizard will start.

Step 4. To configure network attributes and other initial parameters to the system, follow the configuration wizard as shown in the Configuration Wizard Session table below.

Configuration Wizard Session

Wizard Session Display	Comments
NVIDIA configuration wizard Do you want to use the wizard for initial configuration? yes	You must perform this configuration the first time you operate the system or after resetting the system. Type 'y' and then press <Enter>.
Step 1: Hostname? [switch]	If you wish to accept the default hostname, press <Enter>. Otherwise, type a different hostname and press <Enter>.

Wizard Session Display	Comments
Step 2: Use DHCP on mgmt0 interface? [no] yes	Perform this step to obtain an IP address for the system. (mgmt0 is the management port of the system). If you wish the DHCP server to assign the IP address, type 'yes' and press <Enter>. If you type 'no' (no DHCP), then you will be asked whether you wish to use the 'zeroconf' configuration or not. If you enter 'no' (no Zeroconf), you must enter a static IP, and the session will continue.
Step 3: Enable IPv6? [yes]	The management interface will be able to use IPv6 addresses. If you enter "no" (no IPv6), you will automatically be referred to Step 6.
Step 4: Enable IPv6 auto-config (SLAAC) on mgmt0 interface? [no]	This turns on auto-configuration of the IPv6 addresses. This is unsuitable for DHCPv6.
Step 5: Enable DHCPv6 on mgmt0 interface? [no]	To enable DHCPv6 on the MGMT0 interface.
Step 6: Admin password (Press <Enter> to leave unchanged)? <new_password> Step 6: Confirm admin password? <new_password>	To avoid illegal access to the machine, please type a password and then press <Enter>. Then confirm the password by re-entering it. Note that password characters are not printed.
You have entered the following information: <A summary of the configuration is now displayed.> To change an answer, enter the step number to return to or hit <enter> to save changes and exit. Choice: <Enter> Configuration changes saved.	The wizard displays a summary of your choices and then asks you to confirm the choices or to re-edit them. Either press <Enter> to save changes and exit, or enter the configuration step number that you wish to return to. *Note: To re-run the configuration wizard, run the command "configuration jump-start" in Config mode.*

The table below shows an example of static IP configuration for mgmt0 interface.

Configuration Wizard Session - Static IP Configuration

```
NVIDIA configuration wizard

Do you want to use the wizard for initial configuration? yes

Step 1: Hostname? []
Step 2: Use DHCP on mgmt0 interface? [yes] no
Step 3: Use zeroconf on mgmt0 interface? [no]
Step 4: Primary IP address? [for example 192.168.10.4] 10.10.10.10
Mask length may not be zero if address is not zero (interface eth0)
Step 5: Netmask? [0.0.0.0] 255.255.255.0
Step 6: Default gateway? [for example 192.168.10.1] 10.10.10.255
Step 7: Primary DNS server?
Step 8: Domain name?
Step 9: Enable IPv6? [yes]
Step 10: Enable IPv6 autoconfig (SLAAC) on mgmt0 interface? [no]
Step 11: Admin password (Enter to leave unchanged)?

To change an answer, enter the step number to return to.
Otherwise hit <enter> to save changes and exit.
Choice:
Configuration changes saved.

To return to the wizard from the CLI, enter the "configuration jump-start"
command from configure mode. Launching CLI...
```

Step 5. Before attempting a remote (for example, SSH) connection to the system, check the mgmt0 interface configuration. Specifically, verify the existence of an IP address. To check the current mgmt0 configuration, enter the following command:

```
switch (config) # show interfaces mgmt0
Interface mgmt0 status:
  Comment:
  Admin up:          yes
  Link up:           yes
  DHCP running:      yes
  IP address:        10.209.28.50
  Netmask:           255.255.255.0
  IPv6 enabled:      yes
  Autoconf enabled:  no
  Autoconf route:    yes
  Autoconf privacy:  no
  DHCPv6 running:    no
  IPv6 addresses:    1
  IPv6 address:      fe80::202:c9ff:fe63:b55a/64
  Speed:             1000Mb/s (auto)
  Duplex:            full (auto)
  Interface type:    ethernet
  Interface source:  physical
  MTU:               1500
  HW address:        00:02:C9:63:B5:5A

  RX bytes:          968810197      TX bytes:          1172590194
  RX packets:        10982099       TX packets:        10921755
  RX mcast packets:  0              TX discards:       0
  RX discards:       0              TX errors:         0
  RX errors:         0              TX overruns:       0
  RX overruns:       0              TX carrier:        0
  RX frame:          0              TX collisions:     0
                                    TX queue len:      1000

switch (config) #
```

Step 6. Check the software version embedded in your system, using the command 'show version'. Compare this version to the latest version that can be retrieved from NVIDIA support site. To upgrade software, please refer to the *MLNX-OS User Manual*.

Remote Connection

Once the network attributes are set, you can access the CLI via SSH or the WebUI via HTTP/ HTTPs.

To access the CLI, perform the following steps:

1. Set up an Ethernet connection between the system and a local network machine using a standard RJ45 connector.
2. Start a remote secured shell (SSH) using the command: ssh -l <username> <IP_address>

```
# ssh -l <username> <ip_address>
Mellanox MLNX-OS Switch Management

Password:
```

3. Login as admin (default username is *admin*, password is *admin*).
4. Once you get the CLI prompt, you are ready to use the system.

For additional information about MLNX-OS, refer to the *MLNX-OS User Manual* located on the NVIDIA support web.

FRU Replacements

Power Supply

Mellanox systems equipped with two replaceable power supply units work in a redundant configuration. Either unit may be extracted without bringing down the system.

> ❗ Make sure that the power supply unit that you are NOT replacing is showing all green, for both the power supply unit and System Status LEDs.

> ❗ Power supply units have directional air flows similar to the fan module. The fan module airflow must coincide with the airflow of all of the power supply units. If the power supply unit airflow direction is different from the fan module airflow direction, the system's internal temperature will be affected. For power supply unit air flow direction, refer to Air Flow.

To extract a power supply unit:

1. Remove the power cord from the power supply unit.
2. Grasping the handle with your hand, push the latch release with your thumb while pulling the handle outward. As the power supply unit unseats, the power supply unit status LEDs will turn off.
3. Remove the power supply unit.

PS Unit Pulled Out
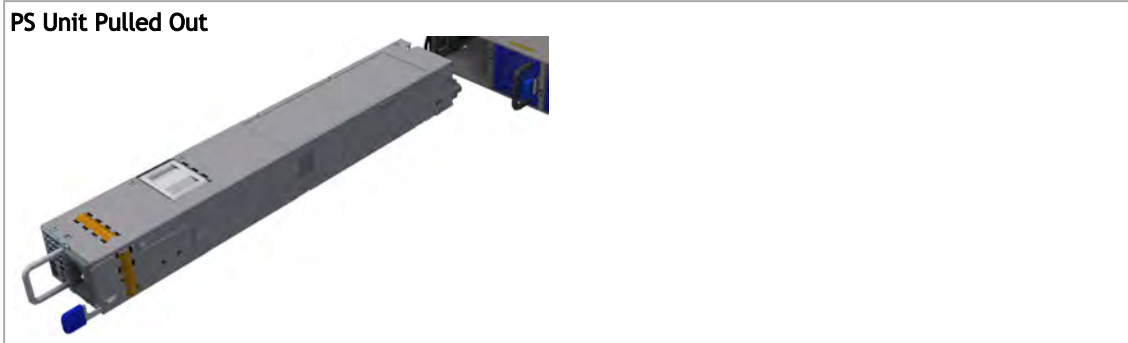
To insert a power supply unit:

1. Make sure the mating connector of the new unit is free of any dirt and/or obstacles.

> ❗ Do not attempt to insert a power supply unit with a power cord connected to it.

2. Insert the power supply unit by sliding it into the opening, until a slight resistance is felt.
3. Continue pressing the power supply unit until it seats completely. The latch will snap into place, confirming the proper installation.
4. Insert the power cord into the supply connector.
5. Insert the other end of the power cord into an outlet of the correct voltage.

> ❗ The green power supply unit indicator should light. If it does not, repeat the whole procedure to extract the power supply unit and re-insert it.

Fans

The system can fully operate if one fan FRU is dysfunctional or missing. Failure of more than one fan is not supported.

> ❗ Make sure that the fans have the air flow that matches the model number. An air flow opposite to the system design will cause the system to operate at a higher (less than optimal) temperature. For power supply unit air flow direction, refer to Air Flow.

To remove a fan unit:

> ⚠️ When replacing a faulty fan unit in an operational switch system, do not leave the slot unpopulated for more than 60 seconds.

1. Grasping the handle with your right hand, push the latch release with your thumb while pulling the handle outward. As the fan unit unseats, the fan unit status LEDs will turn off.
2. Remove the fan unit.

Fan Module Latches
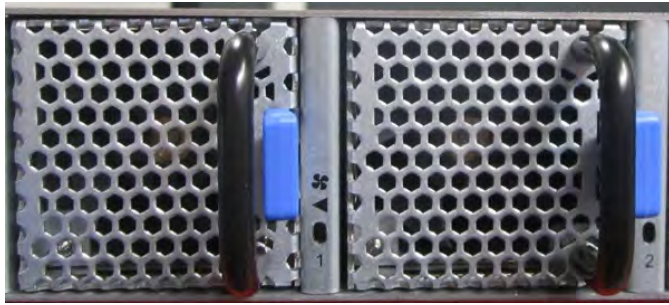
To remove or replace a fan unit, gently pull out its black handle while pushing the latch release with your thumb.

To insert a fan unit:

1. Make sure the mating connector of the new unit is free of any dirt and/or obstacles.
2. Insert the fan unit by sliding it into the opening until slight resistance is felt. Continue pressing the fan unit until it seats completely.

> ❗ The green Fan Status LED should light. If not, extract the fan unit and reinsert it. After two unsuccessful attempts to install the fan unit, power off the system before attempting any system debug.

Software Management

> ⓘ This document is preliminary and subject to change.

Managed systems come with an embedded management CPU card that runs MLNX-OS® management software. The MLNX-OS systems management package and related documentation can be downloaded at https://docs.nvidia.com/networking/category/mlnxos.

InfiniBand Subnet Manager

The InfiniBand Subnet Manager (SM) is a centralized entity running in the system. The SM applies network traffic related configurations such as QoS, routing, partitioning to the fabric devices. You can view and configure the Subnet Manager parameters via the CLI/WebUI. Each subnet needs one subnet manager to discover, activate and manage the subnet.

Each network requires a Subnet Manager to be running in either the system itself (system based) or on one of the nodes which is connected to the fabric (host based).

> ⚠ No more than two subnet managers are recommended for any single fabric.

The InfiniBand Subnet Manager running on the system supports up to 2048 nodes. If the fabric includes more than 2048 nodes, you may need to purchase NVIDIA's Unified Fabric Manager (UFM®) software package.

Each subnet needs one subnet manager to discover, activate and manage the subnet.

Each network requires a Subnet Manager to be running in either the system itself (system based) or on one of the nodes which is connected to the fabric (host based).

The subnet manager (OpenSM) assigns Local IDentifiers (LIDs) to each port connected to the fabric, and develops a routing table based on the assigned LIDs.

A typical installation using the OFED package will run the OpenSM subnet manager at system start up after the drivers are loaded. This automatic OpenSM is resident in memory, and sweeps the fabric approximately every 5 seconds for new adapters to add to the subnet routing tables.

Upgrading Software (on Managed Systems)

Software and firmware updates are available from the NVIDIA Support website. Check that your current revision is the same one that is on the NVIDIA website. If not upgrade your software. Copy the update to a known location on a remote server within the user's LAN.

Use the CLI or the GUI in order to perform software upgrades. For further information please refer to the *MLNX-OS Software User Manual* section Upgrading MLNX-OS® Software.

Be sure to read and follow all of the instructions regarding the updating of the software on your system.

Managed systems do not require Firmware updating. Firmware updating is done through the MLNX-OS management software. The system comes standard with a management software module for system management called NVIDIA Operating System (MLNX-OS). MLNX-OS® is installed on all NVIDIA

Quantum™ based managed systems. MLNX-OS® includes a CLI, WebUI, SNMP, system management software and IB management software (OpenSM).

> ⬥ The Ethernet ports for remote management connect to Ethernet systems. These systems must be configured to 100Mb/1Gb/s auto-negotiation.

Updating Firmware on Externally Managed Systems

There are two methods to update system firmware:

- (Typical) In-band via a switch network port across a cable connecting the server to the switch port.
- (Non-typical) Via the I²C port of the switch using an NVIDIA MTUSB-1 device connecting to a server's USB port on the one end and to the I²C port of the switch on the other.

Firmware updates should normally be conducted in-band. The use of the MTUSB-1 device is intended for cases of debug or firmware corruption and should be conducted by NVIDIA FAEs or Support engineers, or by trained users at the customer's site.

Both types of updates require the installation of NVIDIA Firmware Tools (MFT) package. The MFT package and user manual are available for download under https://network.nvidia.com/products/adapter-software/firmware-tools/. Please select the package that suits your operating system.

In order to obtain information regarding the externally managed system, you must download the NVIDIA MFT tools from https://network.nvidia.com/products/adapter-software/firmware-tools/.

Select and download the release that matches your system. Follow the instructions in the User Manual https://docs.nvidia.com/networking/category/mft to get the tools.

Updating Firmware In-band (Typical)

Check the currently programmed firmware on the system and compare it to the latest firmware available under https://network.nvidia.com/support/firmware/firmware-downloads/ (check under Quantum™ Switch Systems).

In order to obtain the firmware version of the externally managed system:

1. Obtain the LID of the target system. The following instructions use one of the utilities provided by the installed MFT package. (Other methods are described in the *MFT User Manual*) by performing the following:
   a. Mark the GUID printed on the inventory pull-out tab of the system.
   b. Run the command ibnetdiscover and search for the row starting with the word "Switch" and indicates the GUID of the system.
   c. Mark the displayed LID on that row (a decimal number).
2. Run the following command from a host:

```
# flint -d <device> q#
```

3. Compare the results of this command with the latest version for your system posted on https://network.nvidia.com/support/firmware/firmware-downloads/ (select the Quantum™ System page).

4. If the current version is not the latest version, follow the directions in the MFT User manual to burn the new firmware inband.

For further information, please refer to *MFT User Manual* at https://docs.nvidia.com/networking/category/mft.

Interfaces

> ⓘ This document is preliminary and subject to change.

The systems support the following interfaces:
- Data interfaces - InfiniBand
- 10/100/1000 MbE RJ45 management interface(s)
- USB
- RS232 Console port
- I²C interface
- Reset button
- Status and Port LEDs

In order to review the full configuration options matrix, refer to Management Interfaces, PSUs and Fans.

Data Interfaces

The data interfaces use QSFP28 and QSFP56 connectors. The full list of interfaces per system is provided in Speed and Switching Capabilities.

Each QSFP28 port can be connected with a QSFP28 cable or connector for 40/56/100Gb/s. Each QSFP56 port can be connected with QSFP56 cable or connector for 40/56/100/200Gb/s. The system offers High Power class 4 transceivers support in all ports. You may use the following NVIDIA transceivers, designed for InfiniBand EDR links on up to 10km and 40km of single mode fiber:

OPN	Description
MMA1L10-CR	NVIDIA optical transceiver, 100GbE, 100Gb/s, QSFP28, LC-LC, 1310nm, LR4 up to 10km
SPQ-CE-ER-CDFL-M	40km 100G QSFP28 ER Optical Transceiver

Speed

InfiniBand speed is auto-adjusted by the InfiniBand protocol. NVIDIA systems support QDR/FDR/EDR/HDR InfiniBand.
- FDR is an InfiniBand data rate, where each lane of a 4X port runs a bit rate of 14.0625Gb/s with 64b/66b encoding, resulting in an effective bandwidth of 56.25Gb/s.
- EDR is an InfiniBand data rate, where each lane of a 4X port runs a bit rate of 25Gb/s with 64b/66b encoding, resulting in an effective bandwidth of 100Gb/s.
- HDR is an InfiniBand data rate, where each lane of a 4X port runs a bit rate of 50Gb/s with 64b/66b encoding, resulting in an effective bandwidth of 200Gb/s.

RS232 (Console)

> ⚠ The RS232 serial "Console" port is labeled **IOIOI** .

The port labeled "Console" is an RS232 serial port on the front side of the chassis that is used for initial configuration and debugging. Upon first installation of the system, you need to connect a PC to this interface and configure network parameters for remote connections. Refer to Configuring Network Attributes to view the full procedure.

> ⚠ This interface is not found in externally managed systems.

Management

> ⚠ The RJ45 Ethernet "MGT" port is labeled 🖧 .

The RJ45 Ethernet port provides access for remote management. The management port is configured with auto-negotiation capabilities by default (100MbE to 1000GbE). The management port's network attributes (such as IP Address) must be pre-configured via the RS232 serial console port or by DHCP before use. Please refer to Configuring Network Attributes to view the full procedure.

> ❗ Make sure you use only FCC compliant Ethernet cables.

Micro USB

The micro USB interface located in the system's front side is USB 2.0 compliant and can be used by MLNX-OS software to connect to an external disk for software upgrade or file management. The connector comes in a standard micro USB shape. To view the full matrix of micro USB configuration options, refer to Management Interfaces, PSUs and Fans.

> ⚠ • USB 1.0 is not supported.
> • Do not use excessive force when inserting or extracting the USB disk to and from the connector.

I²C

The I²C connector is combined with the Console connector and is located on the front side of the system (the RJ45 connector). It can be used with the HAR255 I²C DB9 splitting harness.

Reset Button

The reset button is located on the front side of the system under the RJ45 connector. This reset button requires a tool to be pressed.

⬥ Do not use a sharp pointed object such as a needle or a push pin for pressing the reset button. Use a flat object to push the reset button.

- To reset the system and the CPU of its management board, push the Reset button and keep it pressed for up to 15 seconds.
- To reset the system, the CPU of its management board and the "admin" password, push the Reset button and keep it pressed for at least 15 seconds. You will then be able to enter without a password and set a new password for the user "admin". In the externally managed systems, the Reset button resets the device.

LEDs

See LED Notifications.

LED Notifications

ⓘ This document is preliminary and subject to change.

The system's LEDs are an important tool for hardware event notification and troubleshooting.

LEDs Symbols

Symbol	Name	Description	Normal Conditions
⚠	System Status LED	Shows the health of the system.	Green/Flashing green when booting
✹	Fan Status LED	Shows the health of the fans.	Green
⊘	Power Supply Units LEDs	Shows the health of the power supply units.	Green

Symbol	Name	Description	Normal Conditions
*	Unit Identifier LED	Lights up on command through the CLI.	Off or blue when identifying a port

System Status LED

System Status LED - Front Side

Front Panel	Description
	The LED in the red oval shows the system's status.

> ⬥ It may take up to five minutes to turn on the system. If the System Status LED shows red after five minutes, unplug the system and call your NVIDIA representative for assistance.

System Status LED Assignments

LED Behavior	Description	Action Required
Solid Green	The system is up and running normally.	N/A
Flashing Green	The system is booting up. This assignment is valid on managed systems only.	Wait up to five minutes for the end of the booting process.
Solid Amber	Major error has occurred. For example, corrupted firmware, system is overheated, etc.	If the System Status LED shows red five minutes after starting the system, unplug the system and call your NVIDIA representative for assistance.

Fan Status LED

Fan Status LED - Front and Rear Sides

Front Panel	Description	Rear Panel
	Both of these LEDs in the red ovals show the fans' status.	

Fan Status Front LED Assignments

LED Behavior	Description	Action Required
Solid Green	All fans are up and running.	N/A
Solid Amber	Error, one or more fans are not operating properly.	The faulty FRUs should be replaced.
Off	System boot	N/A

Fan Status Rear LED Assignments (One LED per Fan)

LED Behavior	Description	Action Required
Solid Green	A specific fan unit is operating.	N/A
Solid Amber	A specific fan unit is missing or not operating properly.	The fan unit should be replaced.
Off	System boot	N/A

> ⬥ Risk of Electric Shock! With the fan module removed, power pins are accessible within the module cavity. Do not insert tools or body parts into the fan module cavity.

Power Supply Status LEDs

There are two power supply inlets in the system (for redundancy). The system can operate with only one power supply connected. In case the power supply is an FRU, a second power supply unit can be added to support hot-swap ability. Each power supply unit has a single 2 color LED on the right side of the unit, that indicates the status of the unit.

Power Status LED

There are two power supply inlets in the system (for redundancy). The system can operate with only one power supply connected. In case the power supply is an FRU, a second power supply unit can be added to support hot-swap ability. Each power supply unit has a single 2 color LED on the right side of the unit, that indicates the status of the unit.

Rear Side Panel
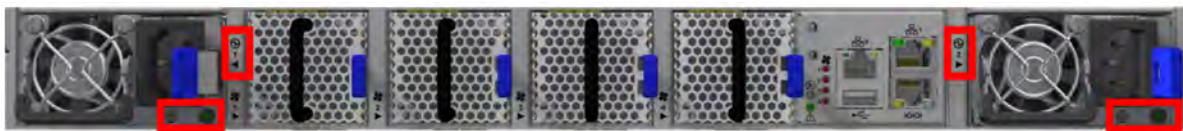
Power Supply Unit Status Front LED Assignments

LED Behavior	Description	Action Required
Solid Green	All plugged (one or two) power supplies are running normally.	N/A

LED Behavior	Description	Action Required
Solid Amber	One or both of the power supplies are not operational or not powered up/ the AC power cable is disconnected.	Make sure the AC cable is plugged in and active. If the problem resumes, the FRUs might be faulty, and should then be replaced.
Off	N/A	N/A

The power supply status LEDs on the rear side of the system are located on the PSUs themselves. Each PSU has one LED of its own.

Power Supply Unit Status Rear LED Assignments

LED Behavior	Description	Action Required
Solid Green	The PSU is running normally.	N/A
Flashing Green 1Hz	AC present / Only 12VSB on (PSU off) or PSU in Smart-on state.	Call your NVIDIA representative for assistance.
Amber	AC cord unplugged or AC power lost while the second power supply still has AC input power.	Plug in the AC cord of the faulty PSU.
	PS failure (including voltage out of range and power cord disconnected).	Check voltage. If OK, call your NVIDIA representative for assistance.
Flashing Amber	Power supply warning events where the power supply continues to operate; high temp, high power, high current, slow fan.	Call your NVIDIA representative for assistance.
Off	No AC power to all power supplies.	Call your NVIDIA representative for assistance.

Unit Identification LED

The UID LED is a debug feature, that the user can use to find a particular system within a cluster by turning on the UID blue LED.

To activate the UID LED on a switch system, run:

```
switch (config) # led MGMT uid on
```

To verify the LED status, run:

```
switch (config) # show leds
Module LED Status
----------------------------------------------------------------------
MGMT UID Blues
```

To deactivate the UID LED on a switch system, run:

```
switch (config) # led MGMT uid off
```

Port LEDs

By utilizing two pairs of two lanes per port, the systems can support up to 16 ports of 100G. You may switch between the two following states by pressing on the LED Splitting Control button:

- Displaying the link status of a single 4-lane port, or of the lower 2-lane split port (if a splitter cable is used). This is the default state.
- Displaying the link status of a single 4-lane port, or of the higher 2-lane split port (if a splitter cable is used).
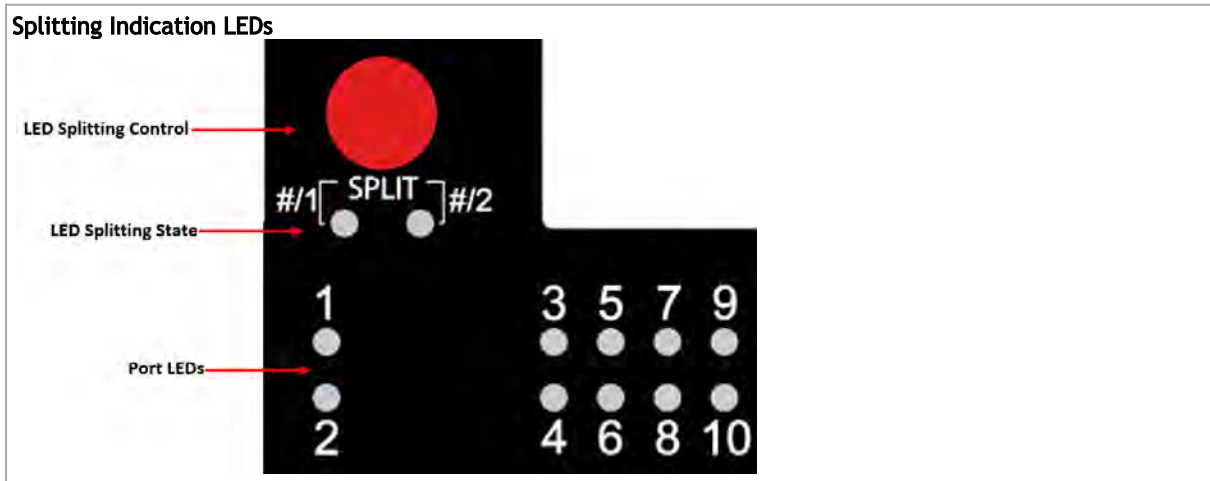
Each time you press on the Lane Select Button, the Port LEDs display will switch to a different state, as follows:

Lane Select Button States

State	Indication LED 1	Indication LED 2	Port LEDs Display
1	On	Off	1st 2x
2	Off	On	2nd 2x

The port LEDs behavior indicates the ports' state, as follows:

Port LEDs in InfinBand System Mode

LED Behavior	Description	Action Required
Off	Link is down.	Check the cable.
Solid Green	Link is up with no traffic.	N/A
Flashing Green	Link is up with traffic.	N/A
Solid Amber	Link is up.	Wait for the Logical link to raise. Check that the SM is up.
Flashing Amber	A problem with the link.	Check that the SM is up.

In InfiniBand system mode, the LED indicator, corresponding to each data port, will light orange when the physical connection is established (that is, when the unit is powered on and a cable is plugged into the port with the other end of the connector plugged into a functioning port). When a

logical connection is made the LED will change to green. When data is being transferred the light will blink green.

Inventory Pull-out Tab

> ⓘ This document is preliminary and subject to change.

The system's inventory parameters (such as serial number, part number and GUID address) can be extracted from the inventory pull-out tab on the lower right side of the front panel.

Pull-out Tab

Troubleshooting

> ⓘ This document is preliminary and subject to change.

Problem Indicator	Symptoms	Cause and Solution
LEDs	System Status LED is blinking for more than 5 minutes	Cause: MLNX-OS software did not boot properly and only firmware is running. Solution: Connect to the system via the console port, and check the software status. You might need to contact an FAE if the MLNX-OS software did not load properly.
	System Status LED is red	Cause: • Critical system fault (CPU error, bad firmware) • Over temperature Solution: • Check environmental conditions (room temperature)
	Fan Status LED is red	Cause: Possible fan issue Solution: • Check that the fan is fully inserted and nothing blocks the airflow • Replace the fan FRU if needed
	Front PSU Status LED is red	Cause: Possible PSU issue Solution: • Check/replace the power cable • Replace the PSU if needed
	The activity LED does not light up (InfiniBand)	Make sure that there is an SM running in the fabric.
System boot failure	The last software upgrade failed on x86 based systems	Solution: • Connect the RS232 connector (CONSOLE) to a laptop. • Push the system's reset button. • Press the ArrowUp or ArrowDown key during the system boot. GRUB menu will appear. For example: ``` Default image: 'SX_X86_64 SX_3.4.0008 2014-11-10 20:07:51 x86_64' Press enter to boot this image, or any other key for boot menu Booting default image in 3 seconds. Boot Menu --- 0: SX_X86_64 SX_3.4.0008 2014-11-10 20:07:51 x86_64 1: SX_X86_64 SX_3.4.0007 2014-10-23 17:27:34 x86_64 --- Use the ArrowUp and ArrowDown keys to select which entry is highlighted. Press enter to boot the selected image or 'p' to enter a password to unlock the next set of features. Highlighted entry is 0: " ``` • Select previous image to boot by pressing an arrow key and choosing the appropriate image.

Specifications

> ⓘ This document is preliminary and subject to change.

TQ8100 and TQ8200 Specifications

Feature		Value
Mechanical	Size:	1.7" (H) x 17" (W) x23.2" (D), 43.6mm (H) x 433.2mm (W) x 590.6mm (D)
	Mounting:	19" Rack mount
	Weight:	1 PSU: 11.4kg 2 PSUs: 12.488kg
	Speed:	**TQ8100:** 2 EDR QSFP28 long reach ports -100Gb/s for up to 10 km 8 standard HDR local ports - 40, 56, 100, 200 Gb/s per port **TQ8200:** 2 EDR QSFP28 long reach ports - 100Gb/s for up to 40 km 8 standard HDR local ports - 40, 56, 100, 200 Gb/s per port
	Connector cage:	10 QSFP56
Environmental	Temperature:	Operational: 0° to 40°C Non-Operational: -40° to 70°C
	Humidity:	Operational: 10%-85% non-condensing Non-Operational: 10%-90% non-condensing
	Altitude:	Operational: 3050m Non-Operational: 3050m
	Noise level:	TBD
Regulatory	Safety:	CB, cTUVus, CE, CU
	EMC:	EMC: CE, FCC, VCCI, ICES, RCM
	RoHS:	RoHS compliant
Power	Input Voltage:	1x/2x, 10A/100-127Vac, 50/60Hz, 4.4A/ 200-240Vac, 50/60Hz
	Global Power Consumption:	**Global Power Consumption** • Max power consumption with all ports connected to AOCs: 600W • Max power consumption with all ports connected to DACs: 538W
Main Devices	CPU:	x86 ComEx Broadwell D-1508
	Switch:	NVIDIA Quantum™
Throughput	Switching:	4Tb/s

Appendix

The document contains the following appendixes:

- Accessory and Replacement Parts
- Thermal Threshold Definitions
- Interface Specifications
- Disassembly and Disposal

Accessory and Replacement Parts

ⓘ This document is preliminary and subject to change.

Ordering Part Numbers for Replacement Parts

OPN	Legacy OPN	Part Description
930-9BRKT-00JF-000	MTEF-KIT-C	Static rack installation kit for 200G 1U systems to be mounted into 430-800mm depth racks
930-9BRKT-00JJ-000	MTEF-KIT-F	Telescopic rack installation kit for 200G 1U systems to be mounted into 600-800mm depth racks
930-9BFAN-00IN-000	MTEF-FANF-C	200G 1U systems Fan Module w/ P2C airflow
930-9BFAN-00J0-000	MTEF-FANR-C	200G 1U systems Fan Module w/ C2P airflow
930-9BPSU-00JZ-000	MTEF-PSF-AC-C	200G 1U systems 1100W AC Power Supply w/ P2C airflow
930-9BPSU-00JG-000	MTEF-PSR-AC-C	200G 1U systems 1100W AC Power Supply w/ C2P airflow
HAR000631	-	Harness RS232 2M cable – DB9 to RJ-45 (for managed switches only)
ACC000501	-	NVIDIA 1U switch black power cord, 250V, 10A, 1830MM and C14 TO C13. Note: Can be purchased as a stand-alone product with PN ACC000501-BUY.

Thermal Threshold Definitions

ⓘ This document is preliminary and subject to change.

Three thermal threshold definitions are measured by the Quantum™ ASICs, and impact the overall switch system operation state as follows:

- Warning – 105°C: On managed systems only: When the ASIC device crosses the 100°C threshold, a Warning Threshold message will be issued by the management software, indicating to system administration that the ASIC has crossed the Warning threshold. Note that this temperature threshold does not require nor lead to any action by hardware (such as switch shutdown).
- Critical – 120°C: When the ASIC device crosses this temperature, the switch firmware will automatically shut down the device.

- Emergency – 130°C: In case the firmware fails to shut down the ASIC device upon crossing its Critical threshold, the device will auto-shutdown upon crossing the Emergency (130°C) threshold.

Interface Specifications

> ⓘ This document is preliminary and subject to change.

QSFP Pin Description

QSFP Pin Description

Connector Pin Number	Pin Name	Signal Description
1	GND	Ground
2	Tx2n	Transmitter Inverted Data Input
3	Tx2p	Transmitter Non-Inverted Data Input
4	GND	Ground
5	Tx4n	Transmitter Inverted Data Input
6	Tx4p	Transmitter Non-Inverted Data Input
7	GND	Ground
8	Mod-SelL	Module Select
9	ResetL	Module Reset

Connector Pin Number	Pin Name	Signal Description
10	Vcc Rx	+3.3 V Power supply receiver
11	SCL	2-wire serial interface clock
12	SDA	2-wire serial interface data
13	GND	Ground
14	Rx3p	Receiver Non-Inverted Data Output
15	Rx3n	Receiver Inverted Data Output
16	GND	Ground
17	Rx1p	Receiver Non-Inverted Data Output
18	Rx1n	Receiver Inverted Data Output
19	GND	Ground
20	GND	Ground
21	Rx2n	Receiver Inverted Data Output 3
22	Rx2p	Receiver Non-Inverted Data Output 3
23	GND	Ground
24	Rx4n	Receiver Inverted Data Output 3
25	Rx4p	Receiver Non-Inverted Data Output 3
26	GND	Ground
27	ModPrsL	Module Present
28	IntL	Interrupt
29	Vcc Tx	+3.3 V Power supply transmitter
30	Vcc 1	+3.3 V Power Supply
31	LPMode	Low Power Mode
32	GND	Ground
33	Tx3p	Transmitter Non-Inverted Data Input
34	Tx3n	Transmitter Inverted Data Input
35	GND	Ground
36	Tx1p	Transmitter Non-Inverted Data Input
37	Tx1n	Transmitter Inverted Data Input
38	GND	Ground

QSFP Connector Male and Female Views
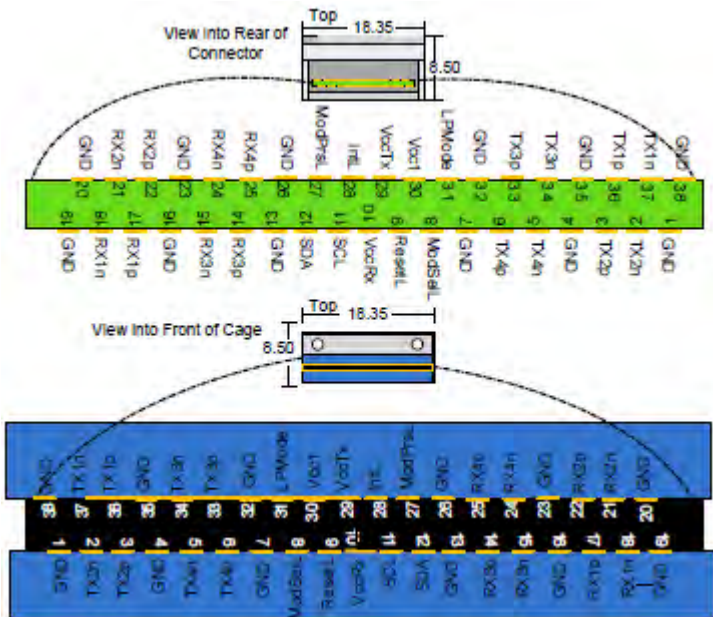
RJ45 to DB9 Harness Pinout

In order to connect a host PC to the Console RJ45 port of the system, a RS232 harness cable (DB9 to RJ45) is supplied.

RJ45 to DB9 Harness Pinout
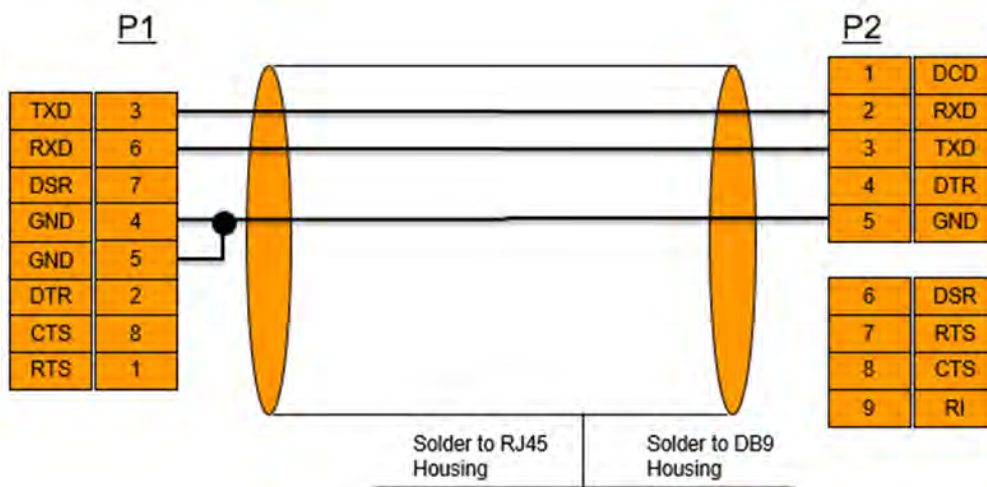
Disassembly and Disposal

> (i) This document is preliminary and subject to change.

Disassembly Procedure

To disassemble the system from the rack:

1. Unplug and remove all connectors.
2. Unplug all power cords.
3. Remove the ground wire.
4. Unscrew the center bolts from the side of the system with the bracket.

> ⚠ Support the weight of the system when you remove the screws so that the system does not fall.

5. Slide the system from the rack.
6. Remove the rail slides from the rack.
7. Remove the caged nuts.

Disposal

According to the WEEE Directive 2002/96/EC, all waste electrical and electronic equipment (EEE) should be collected separately and not disposed of with regular household waste. Dispose of this product and all of its parts in a responsible and environmentally friendly way.

Document Revision History

Date	Revision	Description
August 2022	1.2	Updated OPNs in: • Ordering Information • Installation • Accessory and Replacement Parts
August 2021	1.1	Updated Software Management.
August 2020	1.0	Initial release