FLEXTRONICS®

14.5U 4X SDR/DDR InfiniBand Switch Platform User's Manual

Rev 0.10

Contents

About this Manual

This manual provides an overview of the 14.5U Chassis 4X InfiniBand Switch Platform and guidelines for its operation.

Intended Audience

This manual is intended for users and system administrators responsible for installing and setting up the switch platform.

The manual assumes familiarity with the InfiniBand™ architecture specification.

Related Documentation

The documentation set accompanying the 14.5U 4X InfiniBand Switch platform includes the following:

Table 1 - Reference Documents and Web Sites

Switch Firmware and Firmware Update Tools	See http://www.mellanox.com under Firmware downloads. Note that the Switch System described in this manual is based on Mellanox Technologies' MT47396 InfiniScale^{TM} III switch device.
InfiniBand Administration (IBADM) Tools User's Manual Document no. 2130UM (Mellanox Technologies document)	Describes the installation and use of IBADM tools for management and firmware updates of platforms based on Mellanox Technologies' InfiniBand devices. Available from http://www.mellanox.com under Management Tools.
IB Switch Platforms Embedded Management Tools Document no. 2141UM	Describes the Embedded Management Tools (EMGT) package, which provides Out-of-Band management capabilities for InfiniBand switches. Available from http://docs.mellanox.com (requires a Mellanox Technologies customer account).

Revision History

Table 2 - Revision History of this User's Manual

Revision	Date	Details
0.10	July 2007	First release

1 Overview

This User's Manual provides an overview of the 14.5U 4X SDR/DDR[1] Modular InfiniBand Switch Platform ('the switch') and the operational environment. The switch is based on Mellanox Technologies' MT47396 InfiniScale[TM] III switch device.

Table 3 - Switch Rack Mechanical and Environmental Requirements (Fully Populated Chassis)

Rack Height	Rack Width	Rack Depth	Weight	Max. Power (288 4X Ports)	Ambient Temp.
14.5U (644mm)	19" (EIA-310)	23.6" (600mm)	~195 lb (88.1 kg)	1900 [W] @ 220VAC Up to four Power Supplies of 100-240VAC, 50-60Hz, 13-6.3A	Max: 55ºC Min: 0ºC

This *User's Manual* includes the following chapters:

- This "Overview" chapter
- "Switch Installation and Basic Operation" (page 9)
- "Switch Management Tools" (page 21)
- "Mechanical Drawings Appendix" (page 23)

The switch ships in a minimum base configuration plus additional modules depending on the chosen customer configuration. Optional modules include Leaf boards, a redundant management capability of Spine boards, and redundant power supply units. Figure 1, "Switch Views" shows the various hardware components required for installation.

Figure 1: Switch Views

---
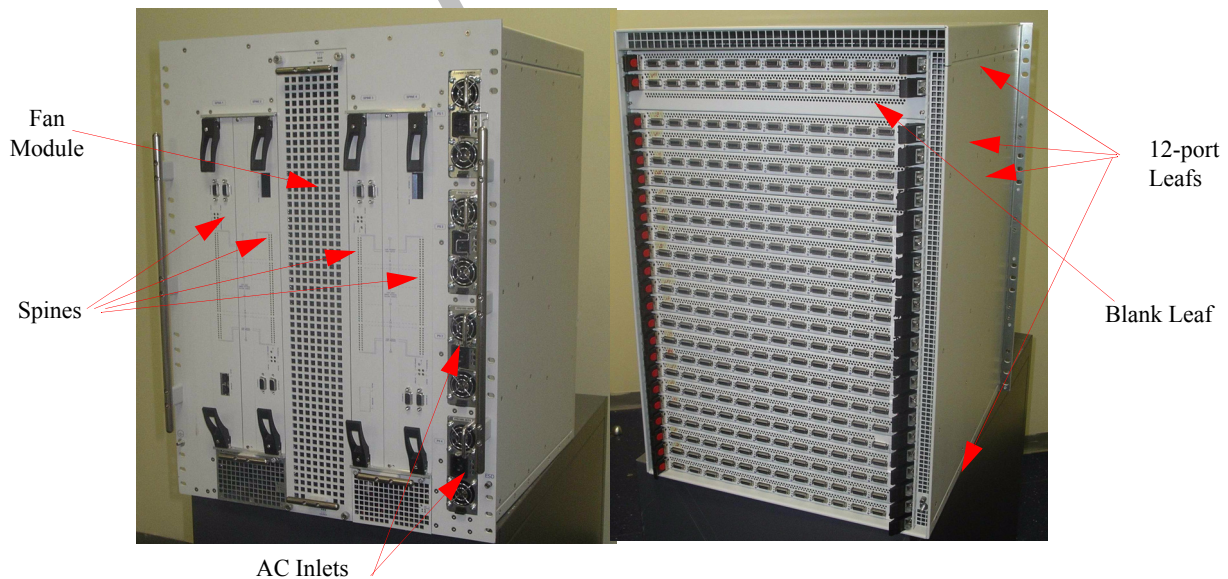
1. 4X IB SDR ports transfer data at 10Gb/s, and 4X IB DDR ports transfer data at 20Gb/s.

1.1 Ordering Information

Table 4 - List of Part Numbers and Switch Product Descriptions

Part Number	Description
F-XR430084	288 Modular InfiniBand DDR Switch in a 288-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430085	288 Modular InfiniBand DDR Switch in a 276-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430086	288 Modular InfiniBand DDR Switch in a 264-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430087	288 Modular InfiniBand DDR Switch in a 252-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430088	288 Modular InfiniBand DDR Switch in a 240-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430089	288 Modular InfiniBand DDR Switch in a 228-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430090	288Modular InfiniBand DDR Switch in a 216-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430091	288 Modular InfiniBand DDR Switch in a 204-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430092	288 Modular InfiniBand DDR Switch in a 192-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430093	288 Modular InfiniBand DDR Switch in a 180-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430094	288 Modular InfiniBand DDR Switch in a 168-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430095	288 Modular InfiniBand DDR Switch in a 156-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430096	288 Modular InfiniBand DDR Switch in a 144-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430097	288 Modular InfiniBand DDR Switch in a 132-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430098	288 Modular InfiniBand DDR Switch in a 120-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430099	288 Modular InfiniBand DDR Switch in a 108-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430100	288 Modular InfiniBand DDR Switch in a 96-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430101	288 Modular InfiniBand DDR Switch in a 84-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430102	288 Modular InfiniBand DDR Switch in a 72-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430103	288 Modular InfiniBand DDR Switch in a 60-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430104	288 Modular InfiniBand DDR Switch in a 48-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430105	288 Modular InfiniBand DDR Switch in a 36-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430106	288 Modular InfiniBand DDR Switch in a 24-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device
F-XR430107	288 Modular InfiniBand DDR Switch in a 12-Port 4X Configuration; 14.5U Chassis; Based on the InfiniScale III Device

2 Switch Installation and Basic Operation

2.1 Installation Safety Warnings

1. Installation Instructions

Read all installation instructions before connecting the equipment to the power source.

2. Bodily Injury Due to Weight

This equipment is very heavy and should be carried by at least **four** people to avoid back and/or other bodily injuries.

3. Over-temperature

This equipment should not be operated in an area with an ambient temperature exceeding the maximum recommended: 55°C (131°F). Moreover, to guarantee proper air flow, allow at least 8cm (3 inches) of clearance around the ventilation openings.

4. Stacking the Chassis

The chassis should not be stacked on any other equipment. If the chassis falls, it can cause bodily injury and equipment damage.

5. During Lightning - Electrical Hazard

During periods of lightning activity, do not work on the equipment or connect or disconnect cables.

6. Copper InfiniBand Cable Connecting/Disconnecting

Copper InfiniBand cables are heavy and not flexible, as such they should be carefully attached to or detached from the connectors. Refer to the cable manufacturer for special warnings/instructions.

7. Rack Mounting and Servicing

When this product is mounted or serviced in a rack, special precautions must be taken to ensure that the system remains stable. In general you should fill the rack with equipment starting from the bottom to the top.

8. Equipment Installation

This equipment should be installed, replaced, or serviced only by trained and qualified personnel.

9. Equipment Disposal

Disposal of this equipment should be in accordance to all national laws and regulations.

10. Local and National Electrical Codes

This equipment should be installed in compliance with local and national electrical codes.

11. Norway Electrical Codes - Electrical Hazard

The switch platform must be connected to an earthed mains socket-outlet.

12. Norway Electrical Codes - Electrical Hazard

In Norway, this system should be connected to the IT power distribution system only.

2.2 Installation Steps

Installation and initialization of the switch is a simple process requiring attention to the normal mechanical, power, and thermal precautions for rack-mounted equipment. The switch does not require any programming or configuration to begin operation in an InfiniBand fabric using standard compliant Subnet Management software. The following steps provide a basic overview of the installation process:

1. Physical Installation - Section 2.2.1

2. Power Connections - Section 2.2.2

3. InfiniBand Copper Cable Installation - Section 2.2.3

4. Switch Power Up - Section 2.2.4

5. InfiniBand Fabric Initialization and Management - Section 2.2.5

6. Hot Swap Extraction/ Insertion of Leaf and Spine Boards - Section 2.2.6

7. Hot Swap Extraction/ Insertion of Power Modules - Section 2.2.9

8. Hot Swap Extraction/ Insertion of Fan Module - Section 2.2.10

2.2.1 Physical Installation

Warning: This equipment is very heavy and should be carried by at least **four** people to avoid back and/ or other bodily injuries.

The switch uses 14.5U of rackspace in a standard 19" rack. As shown in Figure 1,"Switch Views", the switch ships from the factory with the mounting racks on the front side. In addition to the mounting rails, the switch requires support underneath the unit to support the un-mounted side (in this case, the rear side).

Please make sure to choose a rack which is able to support the mechanical and environmental characteristics of a fully populated switch chassis as listed in Table 3, "Switch Rack Mechanical and Environmental Requirements (Fully Populated Chassis)".

Note: A 15U mounting space is required for proper installation of the switch platform.

The rack mounting holes conform to the IEA-310 standard for 19-inch racks. Take precautions to guarantee proper ventilation to ensure air intake at the front of the chassis and exhaust in the rear; this is crucial for maintaining good airflow at ambient temperature. In particular, route cables such that they do not impede the air exhaust from the chassis.

2.2.2 Power Connections

The switch includes integrated hot-swap power supplies which support up to four load-sharing 1000W supply units. The slots for the four power supply units (PSUs) are on the front panel and are labelled PS 1-4. Each PSU has a dedicated AC inlet. This design enables the optional use of separate main and backup AC feeds. The input voltage is Autorange, 100-240 VAC, 50Hz or 60Hz. The power cords should be standard 3-wire AC power cords including a safety ground.

The switch system requires at least two PSUs to operate a fully populated platform. The third PSU provides a failover protection mechanism such that the system will continue operating if one PSU fails. Connecting the third and fourth PSUs to a different AC line provides AC failover protection.

Figure 2: Four Power Inlets - Electric Caution Notification

CAUTION	ACHTUNG	ATTENTION
Risk of electric shock and energy hazard. Disconnecting one power supply disconnects only one power supply module. To isolate the unit completely disconnect all power supplies.	Gafahr des elektrischen Schocks.Entfernen des Netzsteckers eines Netzteils spannungsfrei. Um alle Einheiten spannungsfrei zu machen, sind die Netzstecker aller Netzteile zu entfernen.	Risque de choc et de danger e'lectriques.Le de'branchement d'une seule alimentation stabilise'e ne de'branche uniquement qu'un module "Alimentation Stabilise'e". Pour isoler completement le module en cause, il faut de'brancher toutes les alimentations stabilise'es.

2.2.3 InfiniBand Copper Cable Installation

The switch uses industry standard 4X (CX4) InfiniBand copper cables which are available from several vendors. The standard 4X cables support full 20+20Gb/s (DDR) and 10+10Gb/s (SDR) bidirectional wire speed of the switch ports. All InfiniBand 4X connections are made to the Leaf Boards. Each Leaf has 12 InfiniBand 4X copper connectors which are numbered 1-12 from left to right.

All cables can be inserted or removed with the unit powered on. To insert a cable, press the connector onto the port receptacle until the connector is firmly seated. The green LED indicator to the right of the port will light when the physical connection is established (when unit is powered on). Then lock the connector using the latching mechanism particular to the cable vendor. To remove, disengage the lock and slowly pull the connector away from the port receptacle.

Table 5 is a decoder of the ports yellow and green LEDs states.

Table 5 - Port LEDs Decoder Table

Port Green LED State	Port Yellow LED State	Description
ON (steady)	OFF	Physical link connection of port is established, but the logical link connection is not
ON (steady)	ON (Steady)	Both the physical and logical link connections of port are established
Blinking	--	Physical link error
ON (Steady)	Blinking	Transfer activity is on-going. Blinking rate is proportional to traffic rate
OFF	OFF	Cable is unseated / unconnected

Note: The switch does not provide a visual means that indicates the port speed configuration (SDR or DDR) and/or the link width (1X or 4X). The speed and link width configurations can be retrieved using software.

Warning: Take care to not impede the air exhaust flow through the ventilation next to the InfiniBand ports. Use cable lengths which allow routing horizontally around to the side of the chassis before bending upward or downward in the rack.

2.2.4 Switch Power Up

Check all boards, power supply, and fan tray modules for proper insertion and seating before connecting the AC power cords (read Section 2.2.6 through Section 2.2.7 to ensure all modules are properly installed). Insert thermal blanks in unused leaf slots to maintain balanced air flow. Also, mounting screws should all be tightened.

2.2.4.1 Power Supply and Spine Board Indicator Status at Power ON

As the power is turned on, you should observe the following conditions for normal operation:

1. Power Supply Unit(s) AC OK and DC OK indicators are ON and Alarm indicator is OFF.

2. All LED indicators on the Spine and Leaf Boards light up briefly and then turn off.

3. Spine Board indicators will display status of internal links to the installed Leaf Boards. All PHY links to present Leaf Boards should be ON.

4. There is a *green* Status LED that indicates per Spine board and per Leaf board that all its power supplies are good. Make sure that all are on.

5. There is a *yellow* Status LED per Leaf board and per Spine board. For a Leaf board it should always be *off*. For a Spine board it is *on* in case one of its power supplies are faulty. Make sure they are all off.

6. For Spine boards only: There is a *yellow* System Fail LED that will light to indicate one of the following failure situations: a fan failure, over-temperature, or a main AC power supply failure. On power-on this LED will blink for several seconds and then turn off if none of the above situations has occurred.

7. There is a *blue* LED per Spine board and per Leaf board. It should be *off* for all boards.

The Spine Board has two sets of indicators. One shows which Leaf Boards are present. The second set shows the internal link status to the Leaf Boards. The green PHYSICAL LINK LED indicators on the Spine (and Leaf) Boards will indicate the state representing the link status of their corresponding ports. The green PHY LED is (steady) ON to indicate the physical link is up. It blinks to indicate a physical link error. The yellow LOGICAL LINK LED indicates the logical link is active and will turn ON only when the link is enabled by subnet management software in the InfiniBand subnet. It blinks during activities, and the blinking rate is proportional to traffic rate.

2.2.5 InfiniBand Fabric Initialization and Management

Subnet Management and Subnet Administration are standard components defined by the InfiniBand Specification which implement fabric initialization, discovery, and configuration. The switch's switching fabric can be managed by any third party Subnet Management software running on an InfiniBand Host that is connected to any switch port. The switch is also compatible with the Open Source Subnet Manager OpenSM.

The switch is basically an InfiniBand fabric built out of individual InfiniBand switch devices and appears so to external Subnet Managers. The switch supports the SystemImage GUID feature of the InfiniBand specification which means that software management tools that support this feature can recognize all the internal devices as a single switch system.

A fully populated 288 4X port system features up to 5.76 Tbit/sec (IB ports in DDR mode)[1] Constant Bisectional Bandwidth (CBB) Fat Tree topology that the Subnet Manager will discover and configure to enable a full cluster to operate. For more information refer to the documentation of the Subnet Manager being used.

2.2.6 Hot-Swap Extraction/Insertion of Leaf Boards

Switch Leafs are located on the back of the system. Each switch Leaf board has a pair of ejectors that lock the board in place and serve as a lever for seating or extracting (see Figure 4). The ejectors snap inward to lock the board into

1. 2.88 Tbit/sec when IB ports are in SDR mode

place and include a red button to disengage the lock and electrically notify the system that an extraction has been requested.

2.2.6.1 Extracting a Leaf Board

1. Disconnect all cables connected to the Leaf board.

2. Unscrew the retaining screws on the outside of the connectors. These screws are captive and will remain hanging in place. See Figure 3.

3. Press the red buttons to unlock the ejectors. See Figure 4 (a).

4. Press the ejectors outward so that the board pops out, then slowly slide the board out. *Note that the board is short, therefore do not let go of it while sliding it out.* See Figure 4 (b).

Note: The Leaf board does not have a base plate, thus the print side components are exposed. Please handle the board carefully.
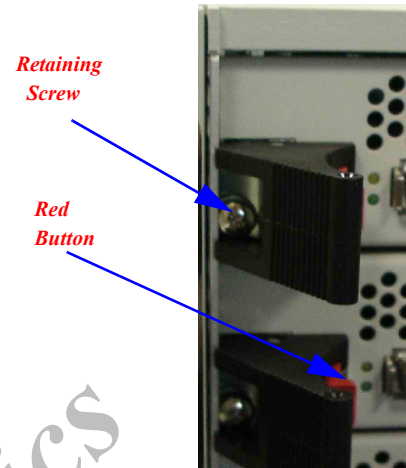
Figure 3: Retaining Screw and Red Button

Figure 4: Leaf Board Extraction

(a) Pressing the red buttons unlocks the ejectors

(b) Sliding out the Leaf board

2.2.6.2 Inserting a Leaf Board

1. Verify the ejectors are unlocked by pressing the red buttons and pulling the ejectors outward on both sides of the board slot.

2. Carefully seat the Leaf board into the side guide rails and slowly slide the board into the chassis until the ejectors begin to engage on the chassis edge.

3. Press the ejectors inward until the locks snap. This indicates that the board is fully seated.

4. Screw in the captive retaining screws on the outside of each ejector to lock the Leaf board in place.

Note: The Leaf board does not have a base plate, thus the print side components are exposed. Please handle the board carefully.

2.2.7 Hot-Swap Extraction/Insertion of Spine Boards

Each switch Spine board has a pair of ejectors that lock the board in place and serve as levers for seating or extracting (see Figure 5).
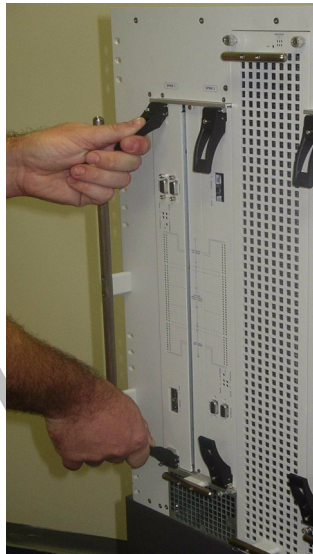
2.2.7.1 Extracting a Spine Board

1. Disconnect all cables connected to the spine.

2. Unscrew both (top and bottom) retaining screws of the spine. See Figure 5(a).

3. Hold the spine ejectors and press both of them outward simultaneously. Expect some resistance before the board pops out. See Figure 5(b).

4. Pull the ejectors out till they form 90 degrees with the panel of the system, then slide out the Spine board. See Figure 5(c).

5. Hold the board with one hand supporting its bottom and the other grasping its middle. Then slide the board all the way out. *Note not to damage the GBX connectors on the board and/or the EMI gasket of the spine-slot in the system.*

Figure 5: Spine Board Extraction

(a) Step 2 of Spine Extraction   (b) Step 3 of Spine Extraction   (c) Step 4 of Spine Extraction

2.2.8 Inserting a Spine Board

1. Hold the Spine board with one hand supporting its bottom and the other grasping its middle.

2. Insert the board into its rail. The board should slide in *without* resistance.

3. Once the board is about half-way inside its rail, hold the spine ejectors and push the board in until it cannot move.

4. Make sure the spine ejectors are open all the way (90 degrees). Now press both ejectors inward simultaneously. You should feel the ejectors are hooked into the front panel.

5. Press both ejectors applying some force to close them completely.

6. Screw in the captive retaining screws of the two ejectors to lock the Spine board in place.

2.2.9 Hot-Swap Extraction/Insertion Power Supply Modules

A PSU (Power Supply Unit) in a redundant configuration (third or fourth PSU) can be extracted without bringing down the system. A PSU should not be extracted with the system on if the system is *not* in redundant configuration.

2.2.9.1 Extracting a PSU

1. Disconnect the AC cord.

2. Press up the black latch at the bottom of the PSU and simultaneously pull PSU handle. See Figure 6(a).

3. Slide out the PSU completely. See Figure 6(b).

Figure 6: PSU Extraction
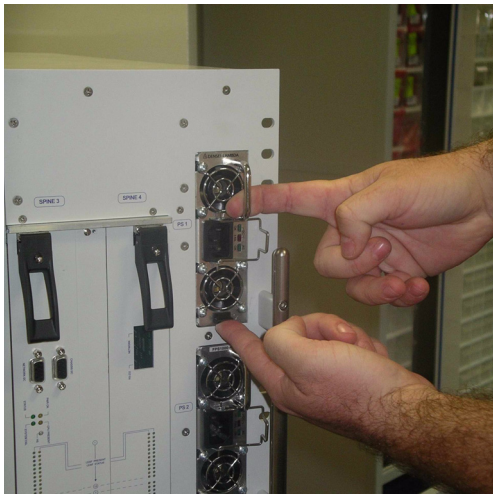
(a) Releasing the PSU

(b) Sliding out the PSU

2.2.9.2 Inserting a PSU

1. Insert the PSU by sliding it into the PSU slot until it seats completely. Make sure the black latch pops out.

2. Connect the AC cord.

3. Make sure the green AC OK and DC OK LEDs should turn on.

2.2.10 Hot Swap Replacement of Fan Modules

There are three fan modules on the switch: One Leaf fan module, and two Spine fan modules. The Leaf fan module of the switch platform is located in the center of the front and includes six fans. See Figure 7 on page 18.

One Spine fan module is located on the left side beneath Spines 1 and 2, and the other module on the right side beneath Spines 3 and 4. Each such module includes two fans.

Each of the fan modules above has a dedicated green LED per fan that lights steady green as long as the fan rotates. If the LED turns off, this indicates that the fan has stopped and the system temperature may rise—in this case, the fan module should be replaced as soon as possible.

2.2.10.1 Extracting the Leaf Fan Module

1. Unscrew the four captive screws that hold the Leaf fan module to the chassis. See Figure 7(a).
2. Slowly slide out the fan module while holding both its upper and lower handles. See Figure 7(b).

Figure 7: Leaf Fan Module Extraction

Retaining Screw

(a) Unscrewing a Retaining Screw

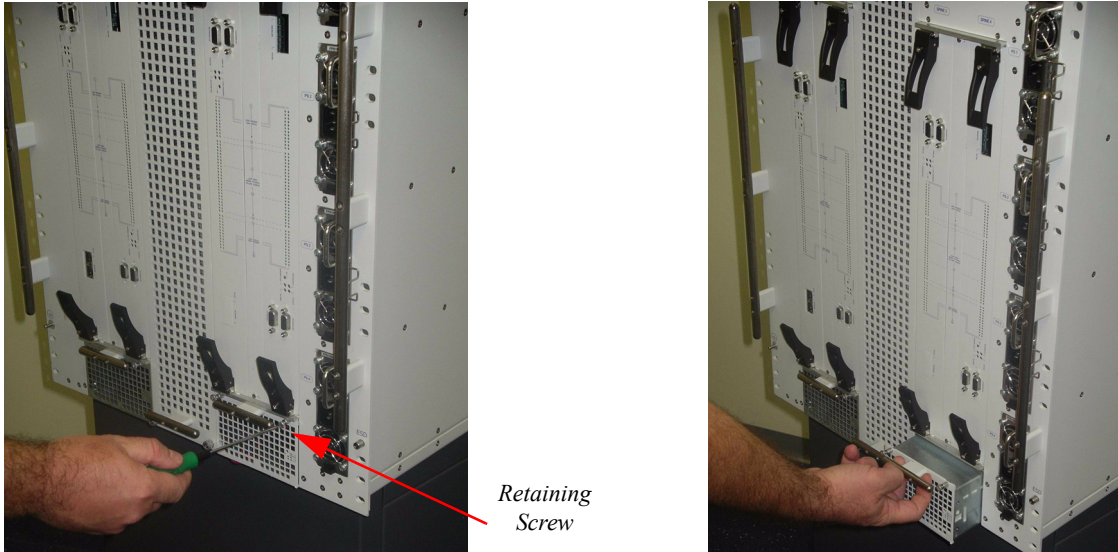(b) Sliding out the Leaf Fan Module
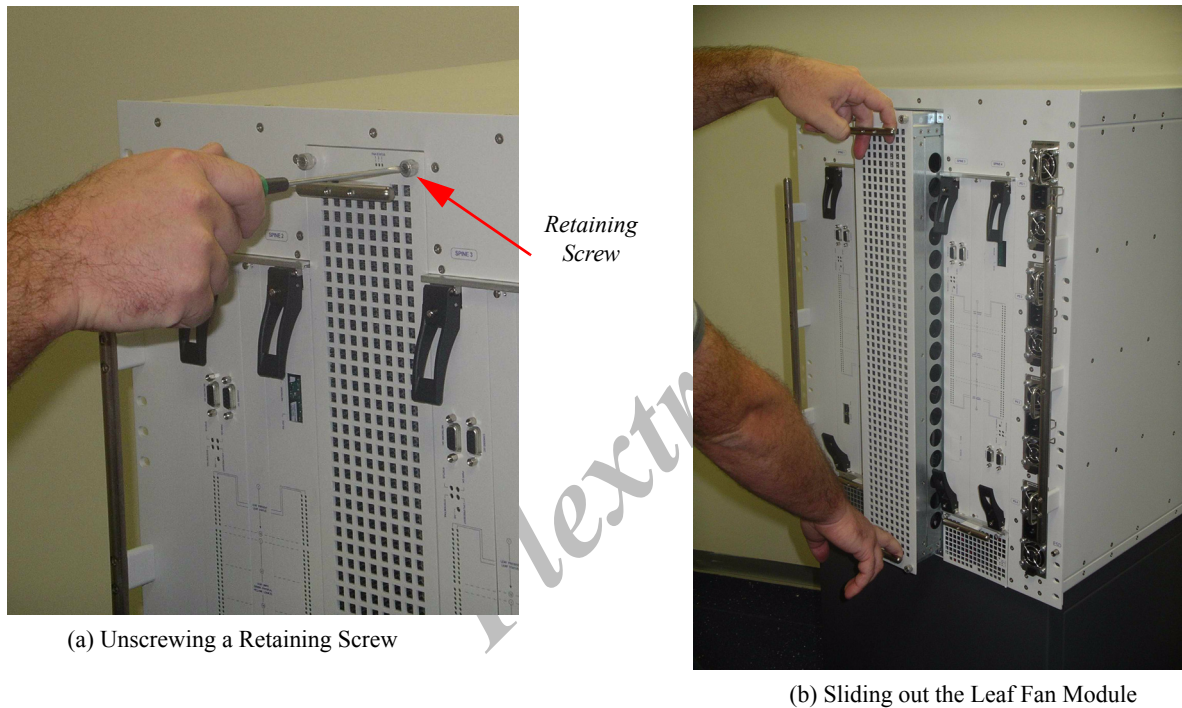
2.2.10.2 Inserting the Leaf Fan Module

1. Slowly slide in the new Leaf fan module while holding both its upper and lower ejectors.
2. Screw back the four captive screws.
3. Make sure that the six green LEDs are turned on (indicating that all fans are rotating).

2.2.10.3 Extracting a Spine Fan Module

1. Unscrew the two captive screws that hold the Spine fan module to the chassis. See Figure 8(a).
2. Pull out the fan module from its handle. See Figure 8(b).

Figure 8: Spine Fan Module Extraction

Retaining Screw

(a) Unscrewing a Retaining Screw   (b) Sliding out the Spine Fan Module

2.2.10.4 Inserting a Spine Fan Module

1. Slide in the new Spine fan module.

2. Screw back the two captive screws.

3. Make sure that the two green LEDs are turned on (indicating that the fans are rotating)

2.2.11 Over-temperature Shutdown Mechanism

Each Leaf card and each Spine card has its own on-board thermal sensor to track PCB temperature. If a PCB's temperature exceeds 70°C, the thermal sensor shuts down the over-heated card.

3 Switch Management Tools

This chapter describes:

- In-Band management of the switch system using the IBADM package (Section 3.1)
- Out-of-Band management of the switch system via Embedded Management Tools (Section 3.2)

3.1 IBADM Features

Maintenance and configuration of the switch is done in-band through the InfiniBand fabric using the Mellanox Infini-Band Administration (IBADM) tools package (see "Related Documentation" on page 5). This package provides the ability to make firmware upgrades and monitor the temperature, voltage, port utilization, and other status parameters in the switch.

The IBADM tools enable the system administrator to manage one or more switches from a single remote InfiniBand host. The features include the following:

- Full In-Band Management of multiple switch systems and Mellanox HCA Systems from a single host
- Simple default configuration to get started quickly
- Name-based subnet browsing and topology verification
- Event monitoring of port statistics, link status and system status for all ports in the switch
- Checking and updating the firmware
- Intuitive Command Line Interfaces
- Extensible and customizable via common scripting languages

  Note: IBADM provides no baseboard management capabilities.

Figure 9 shows the default configuration model for these tools:

Figure 9: Default Configuration Model
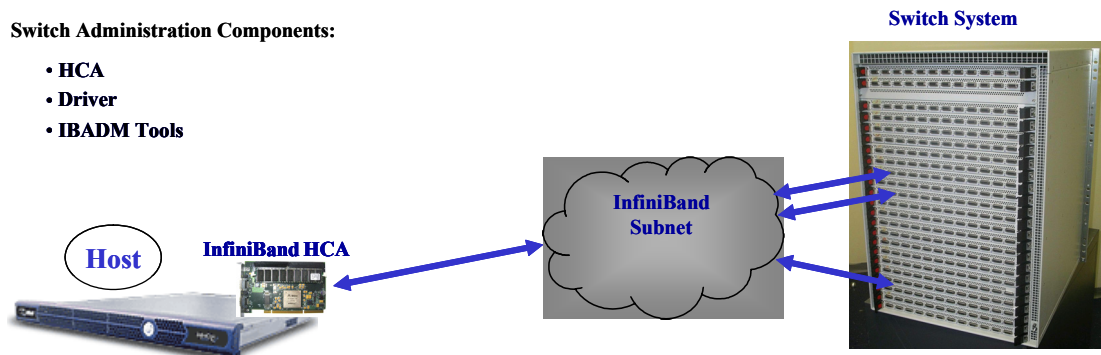
3.1.1 IBADM Requirements

The general requirements for installing the IBADM software are listed below. Please see the IBADM release notes for details on platform, OS, Driver and Subnet Management support.

1. Computer Platform with any Mellanox Technologies host channel adapter (HCA) card installed

2. HCA Driver

3. InfiniBand compliant Subnet Management. The Open source Subnet Manager (OpenSM) is recommended.

3.1.2 How to Get IBADM

The IBADM tool and documentation are available for download via http://www.mellanox.com. See "Related Documentation" on page 5.

3.2 Embedded Management Tools

Embedded Management Tools are based on the I2C interface of the switch system. These tools can be run from the embedded Power-PC CPU card installed in the switch, or from the Linux host PC via a USB to I2C adapter (an MTUSB-x device available from Mellanox).

These tools enable basic chassis management functions such as

• monitoring temperature, voltages, power supplies and fans

• querying for board serial numbers and their revisions

In addition, the tools enable firmware management capabilities such as

• querying for existing firmware versions

• burning new firmware (from scratch or for recovery from damaged firmware)

• querying for and changing system GUIDs

• checking for duplicate or bad GUIDs

For further details, see *Embedded Management Tools User's Manual*.