



Mellanox Hosts & BridgeX® Gateways Ethernet over InfiniBand Quick Start Guide

Rev 1.0

www.mellanox.com

NOTE:

THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT (“PRODUCT(S)”) AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES “AS-IS” WITH ALL FAULTS OF ANY KIND AND SOLELY FOR THE PURPOSE OF AIDING THE CUSTOMER IN TESTING APPLICATIONS THAT USE THE PRODUCTS IN DESIGNATED SOLUTIONS. THE CUSTOMER’S MANUFACTURING TEST ENVIRONMENT HAS NOT MET THE STANDARDS SET BY MELLANOX TECHNOLOGIES TO FULLY QUALIFY THE PRODUCTO(S) AND/OR THE SYSTEM USING IT. THEREFORE, MELLANOX TECHNOLOGIES CANNOT AND DOES NOT GUARANTEE OR WARRANT THAT THE PRODUCTS WILL OPERATE WITH THE HIGHEST QUALITY. ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL MELLANOX BE LIABLE TO CUSTOMER OR ANY THIRD PARTIES FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING, BUT NOT LIMITED TO, PAYMENT FOR PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY FROM THE USE OF THE PRODUCT(S) AND RELATED DOCUMENTATION EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.



Mellanox Technologies, Inc.
350 Oakmead Parkway Suite 100
Sunnyvale, CA 94085
U.S.A.
www.mellanox.com
Tel: (408) 970-3400
Fax: (408) 970-3403

Mellanox Technologies Ltd
PO Box 586 Hermon Building
Yokneam 20692
Israel
Tel: +972-4-909-7200
Fax: +972-4-959-3245

© Copyright 2011. Mellanox Technologies. All rights reserved.

Mellanox®, BridgeX®, ConnectX®, CORE-Direct®, InfiniBlast®, InfiniBridge®, InfiniHost®, InfiniRISC®, InfiniScale®, InfiniPCI®, PhyX®, Virtual Protocol Interconnect and Voltaire are registered trademarks of Mellanox Technologies, Ltd.

FabricIT and SwitchX are a trademark of Mellanox Technologies, Ltd.

All other trademarks are property of their respective owners.

Contents

1	Introduction	7
1.1	Basic Setup.....	7
1.2	Software and Firmware Versions	9
2	System Configuration.....	10
2.1	Server (Host) Infiniband.....	10
2.1.1	Install the BXOFED on Your Host	10
2.2	Ethernet Host (Mellanox ConnectX/ConnetX2 10GE-NIC)	10
2.2.1	Install mlx4_en Software on Your Host	10
3	System Loading	11
3.1	Server (Host) Infiniband.....	11
3.2	Server (Host) Ethernet.....	11
3.3	BridgeX.....	12
4	End-to-End Connectivity Verification	13
5	Sanity Test.....	15
6	Troubleshooting.....	16
Appendix A	FabricIT BXM Installation.....	17
A.1	BridgeX Firmware Update	18
A.2	Mellanox HCA/NIC Firmware Version Verification	19
A.3	Mellanox HCA/NIC New Firmware Image Programming	19

List of Figures

Figure 1: Ethernet over InfiniBand Gateway Environment 8

List of Tables

Table 1: Reference Documents	6
Table 2: Hardware Specification	8
Table 3: Software and Firmware Versions	9

Preface

This Preface provides general information concerning the scope and organization of this Quick Start Guide. It includes the following sections:

1. Introduction
2. System Configuration
 - a. Server (Host) Infiniband
 - b. Ethernet Host (Mellanox ConnectX/ConnetX2 10GE-NIC)
3. System Loading
4. End-to-End Connectivity Verification
5. Sanity Test
6. Troubleshooting
7. FabricIT BXM Installation
 - a. BridgeX Firmware Update
 - b. Mellanox HCA/NIC Firmware Version Verification
 - c. Mellanox HCA/NIC New Firmware Image Programming

Intended Audience

This guide is intended for network administrators who are responsible for configuring and setting up Mellanox Technologies environments.

Related Documentation

The following table lists the documents referenced in this Quick Start Guide.

Table 1: Reference Documents

Document Name	Description
FabricIT BXM Management Software CLI User's Manual	Mellanox FabricIT BridgeX Management (FabricIT BXM) is a software tool that enables the management and configuration of Mellanox Technologies' BX4010 Gateway Platforms in an Infini- Band cluster.
BXOFED User's Manual	Documentation of OFED with BridgeX support on the hosts.
BXOFED Installation Guide	This document describes how to install the various modules and test them in a Linux environment.
Mellanox Firmware Tools (MFT)	The Mellanox Firmware Tools (MFT) package is a set of firmware management tools for a single InfiniBand node. MFT can be used for: Generating a standard or customized Mellanox firmware image Querying for firmware information Burning a firmware image to a single Mellanox device

1 Introduction

This document describes the basic setup of an Ethernet over InfiniBand gateway environment. This setup enables enterprises to improve performance, flexibility and save energy and cooling costs. Consequently, allowing the data centers to operate in high-performance 40Gb/s network speed on the hosts and connecting lower speed Gigabit and 10GigE networks and 2, 4, 8Gb/s Fibre Channel SAN networks.

1.1 Basic Setup

The figure below describes Ethernet over InfiniBand gateway environment which allows the user to connect an existing InfiniBand cluster to an external Ethernet network.

- Each InfiniBand host must have either a Mellanox ConnectX® or ConnectX®-2 HCA connected to either the InfiniBand switch or a fabric of switches via an InfiniBand cable.
- BridgeX® gateway is connected to the same InfiniBand fabric via an InfiniBand cable.
- On the other side BridgeX® gateway is connected to Ethernet network via an Ethernet cable(s)

Figure 1: Ethernet over InfiniBand Gateway Environment

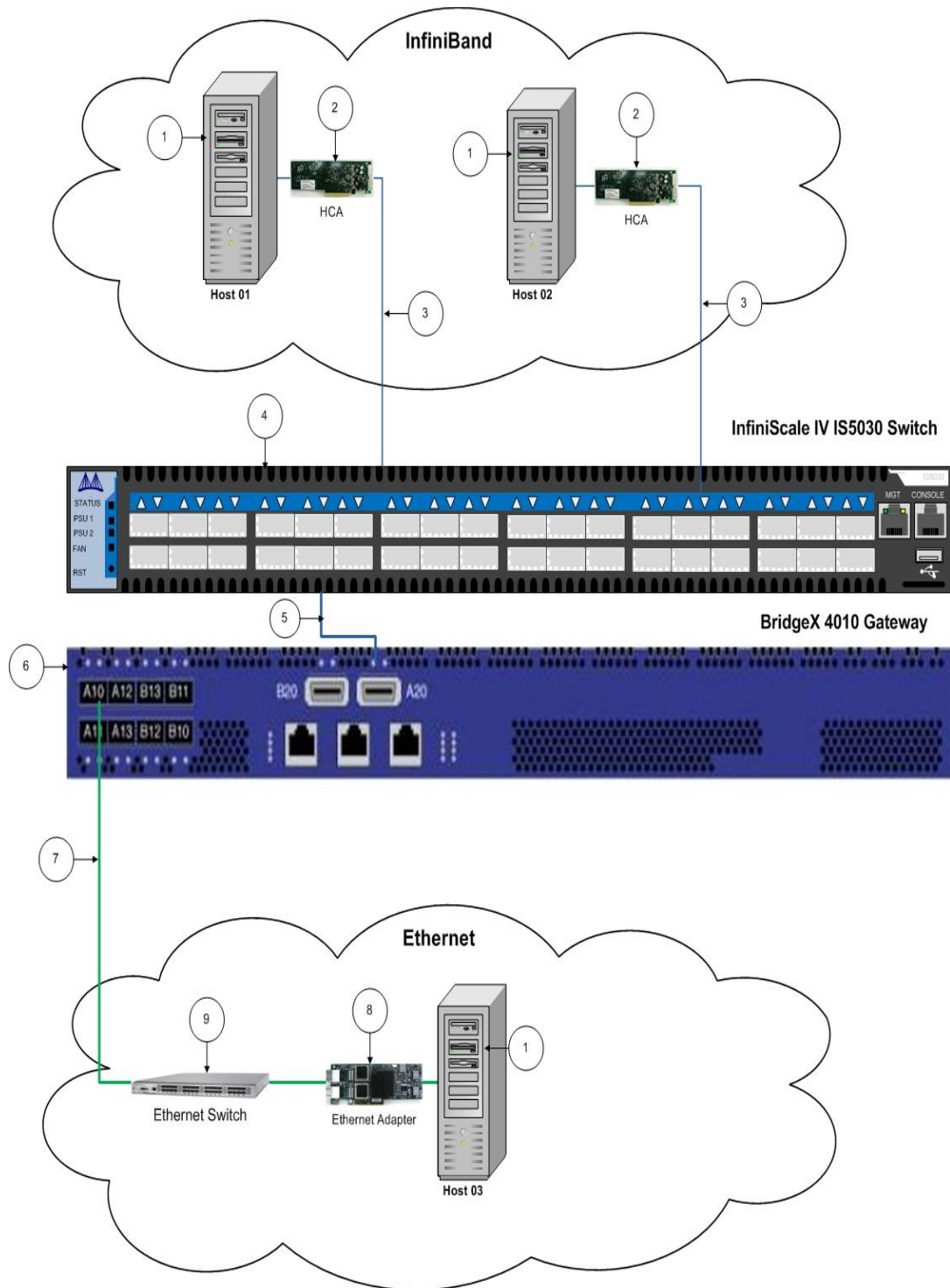


Table 2: Hardware Specification

#	Module	Description
1.	Host	A server that supports a PCI Express Gen1 interface connection.
2.	InfiniBand Host Channel Adapter Card	Provides data centers to consolidate communications, computing, management and storage traffic onto a single fabric. This HCA connects to the host system through a PCI Express Gen1 interface and supports

		InfiniBand connections. For a complete list of the supported HCA cards, see Mellanox website: http://www.mellanox.com
3.	InfiniBand Cable	Standard InfiniBand cable. For a complete list of the supported cables, see Mellanox website: http://www.mellanox.com
4.	InfiniScale® Switch	Switch systems based on InfiniScale IV. For a complete list of the supported switches, see Mellanox website: http://www.mellanox.com
5.	InfiniBand Cable	Standard InfiniBand cable. For a complete list of the supported cables, see Mellanox website: http://www.mellanox.com
6.	BridgeX® Gateway	BX4010 gateway allows data centers to deploy I/O consolidation solutions using InfiniBand as the convergence fabric of choice. For a complete list of the supported switches, see Mellanox website: http://www.mellanox.com
7.	Ethernet Cable	Standard Ethernet cable. For a complete list of the supported cables, see Mellanox website: http://www.mellanox.com
8.	Ethernet Adapter	Provides data centers to consolidate communications, computing, management and storage traffic onto a single fabric. This Adapter connects to the host system through a PCI Express Gen1 interface. For a complete list of the supported Ethernet adapters, see Mellanox website: http://www.mellanox.com
9.	Ethernet Switch	Standard Ethernet Switch

1.2 Software and Firmware Versions

The table below describes the basic requirements for the software and firmware versions of the various components.

Table 3: Software and Firmware Versions

Platform	Software Version
BridgeX Gateway	
FabricIT™ BXM	BXM_PPC_M460EX 1.3.6-5.img
BridgeX® Firmware	8.3.3160
InfiniBand Host	
BXOFED (Each Host)	BXOFED-1.5.1-1.3.6-4
HCA Firmware	Check the Mellanox website for the latest version compatible with your HCA card.
10 G Ethernet Host (other vendors can be used as well)	
MLx4_EN (Each Host)	mlnx_en_1_5_1
HCA Firmware	Check the Mellanox website for the latest version compatible with your HCA card.

2 System Configuration

2.1 Server (Host) Infiniband



Verify you have the updated firmware on each platform according to the [Mellanox](#) website. Otherwise, follow the steps described in the Mellanox HCA/NIC Firmware Version Verification and the Mellanox HCA/NIC New Firmware Image Programming section below.

2.1.1 Install the BXOFED on Your Host



Before installing the InfiniBand server, verify that your BXOFED supports your operating system and the kernel. For further information see `bxofed_release_notes.txt`

For detailed installation procedures, please refer to the BXOFED Installation Guide document for interactive menu or automatic installation.

1. Go to the Mellanox website and download BXOFED to your InfiniBand host.
2. Enter the Serial Number (S/N) when prompted. The S/N appears on the backplane of the BridgeX®.
3. Install the driver. Run:

```
#> tar xzvf BXOFED-1.5.1-1.3.6-4.tgz
#> cd BXOFED-1.5.1-1.3.6-4
#> ./install.pl
```

For detailed installation procedures, please refer to BXOFED Installation Guide document.

2.2 Ethernet Host (Mellanox ConnectX/ConnetX2 10GE-NIC)



Verify you have the updated firmware on each platform according to the [Mellanox](#) website. Otherwise, follow the steps described in the Mellanox HCA/NIC Firmware Version Verification and the Mellanox HCA/NIC New Firmware Image Programming section below.

2.2.1 Install mlx4_en Software on Your Host



Download the latest driver package from the website to your Ethernet host machines.

1. Go to the Mellanox website and download the `mlx_en` to your Ethernet host.
2. Install the driver. Run:

```
#> tar xzvf mlx_en-1.5.1.tgz file
#> cd mlx_en-1.5.1
#> ./install.sh
```

3 System Loading

The following subsection explains how to load the system from each component of the clusters. Prior to loading, please verify the following:

1. All the servers (switch and gateway) are powered up.
2. All the links are up.
3. You have the updated version and firmware on each platform.

3.1 Server (Host) Infiniband

1. Log into the IB hosts.
2. Verify all the required components are enabled (for EoIB, verify that MLX4_VNIC_LOAD is set to yes). Run from your host:

```
vi /etc/infiniband/openib.conf
# Load MLX4 modules
MLX4_LOAD=yes
# Load MLX4_VNIC module
MLX4_VNIC_LOAD=yes
```

3. Restart InfiniBand software stack to apply the changes made to the /etc/infiniband/openib.conf file. Run:

```
/etc/init.d/openibd restart
```

4. Verify that a Subnet Manager (SM) is running on the InfiniBand fabric. Run:

```
ibstat
```

The SM can run on a switch or from a single InfiniBand host.

If you do not have a managed switch with an embedded SM, use the OpenSM that is available with the BXOFED. To run the OpenSM:

- d. Log into one of the hosts.
- e. Type: \$ opensm start &
- f. Verify that the State is 'Active' and the Physical state is 'LinkUp'. Run:

```
ibstat or ibv_devinfo
```

5. Create a vnic on your host. For further information, see Step 2 in the End-to-End Connectivity Verification section.

3.2 Server (Host) Ethernet

1. Log into the Ethernet host.
2. Got to:

```
/etc/infiniband/openib.conf
```

3. Load the Ethernet modules. Run:

```
modprobe mlx4_en
```

4. Verify the link on the Ethernet NIC. Run:

```
ethtool <eth port>
```

Example:

```
# ethtool eth7
```

```
Settings for eth7:
Supported ports: [ ]
Supported link modes:
Supports auto-negotiation: No
Advertised link modes: 10000baseT/Full
Advertised auto-negotiation: No
Speed: 10000Mb/s
Duplex: Full
Port: Twisted Pair
PHYAD: 0
Transceiver: internal
Auto-negotiation: off
Supports Wake-on: d
Wake-on: d
Current message level: 0x00000014 (20)
Link detected: yes
```

5. Configure the Ethernet interface to use static or dynamic IP. Run:

```
ifconfig <interface> <ip address> netmask <netmask>
```

Example:

```
ifconfig eth7 50.50.50.1 255.255.255.0
```

3.3 BridgeX

1. Check you have the updated FabricIT and Firmware version according to the Software and Firmware Versions table. Run:

```
BridgeX-4010 > enable
BridgeX-4010 # configure terminal
BridgeX-4010 # show version
BridgeX-4010 # show asic-version
```

2. Log into the gateway and check the links are up and active, and the external ports are configured as Eth. Run:

```
BridgeX-4010 (config) # show port brief
Check all relevant links are up.
```

Example:

```
=====
Port  Type  Logical  Physical  Speed  Width  Rate
      State  State    (Gbps)           (Gbps)
=====
A20  IB    Active   LinkUp    10.0   4X     40.0
B20  IB    Down     Polling   N/A    4X     N/A
=====

Port  Type  Admin  Status  Speed  MTU  Duplex
      Mode           (Gbps)
=====
A10  Eth  Enabled  Up      10     9600  full (auto)
A11  Eth  Enabled  Down    10     9600  full (auto)
A12  Eth  Enabled  Down    10     9600  full (auto)
B10  Eth  Enabled  Down    10     9600  full (auto)
B11  Eth  Enabled  Down    10     9600  full (auto)
B12  Eth  Enabled  Down    10     9600  full (auto)
```

4 End-to-End Connectivity Verification

1. Check the available active ports. Run:

```
ibv_devinfo
```

Example:

```
# ibv_devinfo
hca_id:                mlx4_0
fw_ver:                2.7.000
node_guid:             0002:c903:0009:20d6
sys_image_guid:       0002:c903:0009:20d9
vendor_id:             0x02c9
vendor_part_id:       26428
hw_ver:                0xA0
board_id:              MT_0BB0120003
phys_port_cnt:        2
port:                  1
state:                 PORT_ACTIVE (4)
max_mtu:               2048 (4)
active_mtu:            2048 (4)
sm_lid:                1
port_lid:              14
port_lmc:              0x00
port:                  2
state:                 PORT_DOWN (1)
max_mtu:               2048 (4)
active_mtu:            2048 (4)
sm_lid:                0
port_lid:              0
port_lmc:              0x00
```

2. Create a vNIC on each InfiniBand host. Perform the following steps:

- a. Edit the `mlx4_vnic.conf` file. Run:

```
vi /etc/infiniband/mlx4_vnic.conf
```

- b. Add a vnic using the following command:

```
name=eth44 mac=00:25:8B:27:14:78 ib_port=mlx4_0:1 vid=-1 vnic_id=5
bx=00:00:00:00:00:00:03:B2 eport=A10
```

The table below describes the format of the vnic script.

Table 4: Fields Description

Field	Description
Name	The name of the logical interface that will be created.
Mac	An arbitrary mac address must be in the following format: xx:xx:xx:xx:xx:xx
Ib_port	Hang the vnic to a specific physical port. it can be viewed by running the <code>ibv_devinfo</code> command. It is determined by the HCA-ID(Line 1) and the active port(Line 11 for single port, lines 11 & 18 on dual ports).
Vid	The Vlan ID (default -1).
vNic_id	The vnic ID, the number must be unique per subnet
BX	The BridgeX GUID which can be viewed by running the <code>show bxm</code> command. Adds the vnic to a specific port.

Field	Description
	It is determined by the HCA-ID and the active port (it can be viewed by running the <code>ibv_devinfo</code> command).
Eport	The BridgeX port connected to the Ethernet cloud.

3. Start the vNIC. Run:

```
/etc/init.d/mlx4_vnic_confd start
```

4. Define an IP on the interface. Run:

```
ifconfig eth44 50.50.50.7 netmask 255.255.255.0
```

5. Repeat these steps (1 to 5) on each InfiniBand Host.

6. Verify that the vNIC was successfully configured. Run:

```
ifconfig -a
```

5 Sanity Test

Check connectivity. Ping from the IB subnet to the Ethernet Subnet and run:

1. On the InfiniHost side:

```
# ping 50.50.50.1
PING 50.50.50.1 (50.50.50.1) 56(84) bytes of data.
64 bytes from 50.50.50.1: icmp_seq=1 ttl=64 time=0.062 ms
64 bytes from 50.50.50.1: icmp_seq=2 ttl=64 time=0.049 ms
```

2. Verify the packet is received on the Ethernet Host. Either run:

```
Tcpdump
```

or use the interface counters below:

```
# ethtool -S eth7 | grep rx_pac
Rx_packets: 2673
```

6 Troubleshooting

For troubleshooting, see [FabricIT BXM Management Software CLI User's Manual](#).

Appendix A FabricIT BXM Installation

1. Log into the system to obtain the Serial Number. Run:

```
# show inventory
```

2. Download FabricIT BXM version from the Mellanox website.
3. Enter the Serial Number (S/N) when prompted.

To upgrade FabricIT BXM software on your system, perform the following steps:

1. Log into the BridgeX
2. Change to Config mode. Run:

```
BridgeX> enable
BridgeX # configure terminal
Check if you have available space for the new version by running the following
commands:
BridgeX (config) # show files system
BridgeX (config) # show images
```

If you have available space, continue to [Step 4](#).

3. Delete the old image from the “Images available to be installed” prior to fetching the new image. Run:

```
image delete
```

Example:

```
BridgeX (config) # image delete
image-EFM_PPC_M405EX-ppc-m405ex-20090531-190132.img
```

4. Fetch the new software image.

The image can be downloaded from the [Mellanox](#) website.

Example:

```
BridgeX(config)# Image fetch
scp://username:password@192.168.10.125/var/www/html/
<image_name>
Password (if required): *****
100.0%[#####]
#####]
BridgeX (config) #
```

5. Display the available images/ Run:

```
show images
```



There are two installed images on the system therefore, if one of the images gets corrupted (e.g., due to power interruption), in the next reboot the image will go up from the second partition as.

```
BridgeX (config) # show images
Images available to be installed:
new_image.img
BXN <new ver> 2009-05-13 16:52:50
Installed images:
Partition 1:
BXN <old ver> 2009-05-13 03:46:25
Partition 2:
BXN <new ver> 2009-05-13 03:46:25
Last boot partition: 1
Next boot partition: 1
No boot manager password is set.
BridgeX (config) #
```

6. Install the new image. Run:

```
image install
```



If FabricIT BXM version 1.0.2 is installed on your gateway system, first upgrade the image to version 1.1.0 and then to your target version (1.1.1 or higher).

The following steps describe how to upgrade from version 1.0.2 to 1.1.0

```
BridgeX (config) # image install <image_name> verify ignore-sig
Step 1 of 4: Verify Image
100.0%
[#####]
Step 2 of 4: Uncompress Image
100.0%
[#####]
Step 3 of 4: Create Filesystems
100.0%
[#####]
Step 4 of 4: Extract Image
100.0%
[#####]
BridgeX (config) #
```

7. Activate the new image to use it upon the next boot. Run:

```
image boot next
```

Example:

```
BridgeX (config) # image boot next
Run show images to review your images.
BridgeX (config) # show images
Images available to be installed:
new_image.img
BXM <new ver> 2009-05-13 16:52:50
Installed images:
Partition 1:
BXM <old ver> 2009-05-13 03:46:25
Partition 2:
BXM <new ver> 2009-05-13 16:52:50
Last boot partition: 1
Next boot partition: 2
No boot manager password is set.
```

A.1 BridgeX Firmware Update

1. Change to Config mode. Run:

```
BridgeX> enable
BridgeX # configure terminal
```

2. Display the list all of the modules and their corresponding FW version. Run:

```
show asic-version
```

3. Fetch the new FW bin file. Run:

```
image fetch
```

For further information on the command, see [Step 4](#) in the FabricIT BXM Installation section above.

4. Obtain the list of the modules requiring FW update. Run:

```
image install-chip-fw BX ?
```

5. Install the firmware. Run:

```
image install-chip-fw BX
```

Example:

```
image install-chip-fw BX <bin file>
```

6. Reboot the bridge system. Run:

```
reload
```

7. Verify the required FW was successfully updated. Run:

```
show asic-version
```



For more information on the update FW command see FabricIT BXM Management Software CLI User's Manual section 8.29.10.

A.2 Mellanox HCA/NIC Firmware Version Verification



To burn the new firmware on a Mellanox HCA, install first the Mellanox Firmware Tool (MFT). The package can be downloaded from the [Mellanox](#) website.

1. Load the mst module to burn the new firmware. Run:

```
> mst start
> mst status
```

The output of the mst status, shows the available Mellanox devices that can be burned.

2. Use the Device ID to update the firmware on the HCA. Run:

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

Example:

```
flint -d /dev/mst/mt26428_pci_cr1 q
Image type:          ConnectX
FW Version:          2.7.0
Device ID:           26428
Chip Revision:      A0
Description:
GUIDs:              0002c903000920d6 0002c903000920d7 0002c903000920d8
0002c903000920d9
MACs:               000000000000          000000000001
Board ID:            (MT_0BB0120003)
VSD:
PSID:               MT_0BB0120003
```

A.3 Mellanox HCA/NIC New Firmware Image Programming

1. Go to the Mellanox Firmware download page.
2. Use the PSID to find the bin file (Firmware)
3. Download the bin file and save it on your machine
4. Update your HCA firmware. Run:

```
flint -image <bin file> -dev <device> b
```

Example:

```
flint -image fw-25408-2_7_000-MHQH29-XTC_A2-A3.bin -dev
/dev/mst/mt26428_pci_cr1 b
```

5. Reboot the host.