Mellanox OFED for Linux Installation Guide

Rev 1.20 – June 12, 2008

Document Number: 2914

Mellanox Technologies, Inc.
2900 Stender Way
Santa Clara, CA 95054
U.S.A.
www.mellanox.com

Tel: (408) 970-3400
Fax: (408) 970-3403

Mellanox Technologies Ltd
P.O. Box 586 Hermon-Building
Yokneam 20692
Israel

Tel: +972-4-909-7200
Fax: +972-4-959-3245

Overview

This document describes how to install and test the Mellanox OFED for Linux package on a single host machine with Mellanox InfiniBand hardware installed. The chapter includes the following sections:

Hardware and Software Requirements

Hardware Requirements

Platforms

- A server platform with an adapter card based on one of the following Mellanox Technologies' InfiniBand HCA devices:
  - ConnectX™ IB (firmware: fw-25408)
  - InfiniHost™ III Ex (firmware: fw-25218 for Mem-Free cards, and fw-25208 for cards with memory)
  - InfiniHost™ III Lx (firmware: fw-25204)
  - InfiniHost™ (firmware: fw-23108)

Note: For the list of supported architecture platforms, please refer to the *Mellanox OFED for Linux Release Notes* file under `docs/`.

Required Disk Space for Installation

- 400MB

Software Requirements

Operating System

- Linux operating system

Note: For the list of supported operating system distributions and kernels, please refer to the *Mellanox OFED for Linux Release Notes* file under `docs/`.

Installer Privileges

- The installation requires administrator privileges on the target machine

Downloading Mellanox OFED

Step 1. Verify that the system has a Mellanox HCA installed by ensuring that you can see ConnectX or InfiniHost entries in the display.

The following example shows a system with an installed Mellanox HCA:

```
host1# lspci -v | grep Mellanox
06:01.0 PCI bridge: Mellanox Technologies MT23108 PCI Bridge (rev
a0) (prog-if 00[Normal decode])
07:00.0 InfiniBand: Mellanox Technologies MT23108 InfiniHost (rev
a0)
Subsystem: Mellanox Technologies MT23108 InfiniHost
```

Step 2. Download the ISO image to your host.

The image's name has the format MLNX_OFED_LINUX-<ver>-<OS label>.iso. You can download it from http://www.mellanox.com/ under Mellanox OFED.

Step 3. Use the md5sum utility to confirm the file integrity of your ISO image. Run the following command and compare the result to the value provided on the download page.

```
host1$ md5sum MLNX_OFED_LINUX-<ver>-<OS label>.iso
```

Installing Mellanox OFED

The installation script, `mlnxofedinstall`, performs the following:

- Discovers the currently installed kernel
- Uninstalls any InfiniBand stacks that are part of the standard operating system distribution or another vendor's commercial stack
- Installs the MLNX_OFED_LINUX binary RPMs (if they are available for the current kernel)
- Identifies the currently installed InfiniBand HCAs and automatically[1] upgrades the firmware

Pre-installation Notes

- The installation script removes all previously installed Mellanox OFED packages and re-installs from scratch. You will be prompted to acknowledge the deletion of the old packages.

  Note: Pre-existing configuration files will be saved with the extension ".conf.saverpm".

- If you need to install Mellanox OFED on an entire (homogeneous) cluster, a common strategy is to mount the ISO image on one of the cluster nodes and then copy it to a shared file system such as NFS. To install on all the cluster nodes, use cluster-aware tools (such as pdsh).

Installation Script

Mellanox OFED includes an installation script called `mlnxofedinstall`. Its usage is described next. You will use it during the installation procedure described in "Installation Procedure" below.

Usage

```
./mlnxofedinstall [OPTIONS]
```

Note: If no options are provided to the script, then all available RPMs are installed.

1. The firmware will not be updated if you run the install script with the '--without-fw-update' option.

Options

```
-c|--config <packages config_file>
                    Example of the configuration file can be found under
                    docs
-n|--net <network config file>
                    Example of the network configuration file can be
                    found under docs
-p|--print-available Print available packages for the current platform
                    and create a corresponding ofed.conf file. The
                    installation script exits after creating ofed.conf.
--with-32bit        Install 32-bit libraries (default). This is relevant
                    for x86_64 and ppc64 platforms.
--without-32bit     Skip 32-bit libraries installation.
--without-depcheck  Skip Distro's libraries check
--without-fw-update Skip firmware update
--force-fw-update   Force firmware update
--force             Force installation (without querying the user)
--all|--hpc|--basic Install all, hpc or basic packages respectively
-v|-vv|-vvv         Set verbosity level
-q                  Set quiet - no messages will be printed
```

Installation Procedure

Step 1. Login to the installation machine as root.

Step 2. Mount the ISO image on your machine

```
host1# mount -o ro,loop MLNX_OFED_LINUX-<ver>-<OS label>.iso /mnt
```

Note: After mounting the ISO image, `/mnt` will be a Read Only folder.

Step 3. Run the installation script

```
host1# /mnt/mlnxofedinstall
    This program will install the MLNX_OFED_LINUX package on your machine.
    Note that all other Mellanox, OEM, OFED, or Distribution IB packages will be
    removed.
    Do you want to continue?[y/N]:y

    Uninstalling the previous version of OFED

    Starting MLNX_OFED_LINUX installation

    Installing kernel-ib RPM
    Preparing...                ########################################### [100%]
       1:kernel-ib              ########################################### [100%]
    Installing kernel-ib-devel RPM
    Preparing...                ########################################### [100%]
       1:kernel-ib-devel        ########################################### [100%]
    Installing ib-bonding RPM
    Preparing...                ########################################### [100%]
```

```
   1:ib-bonding             ########################################### [100%]
Installing mft RPM
Preparing...                ########################################### [100%]
   1:mft                    ########################################### [100%]
Install user level RPMs:
Preparing...                ########################################### [100%]
   1:libibverbs             ###########################################  [  1%]
   2:libibcommon            ###########################################  [  3%]
   3:libibumad              ###########################################  [  4%]
   4:libibcommon            ###########################################  [  6%]
   5:libibumad              ###########################################  [  7%]
   6:libibverbs             ###########################################  [  9%]
   7:opensm-libs            ###########################################  [ 10%]
   8:libibmad               ###########################################  [ 12%]
   9:libibverbs-devel       ###########################################  [ 13%]
  10:librdmacm              ###########################################  [ 14%]
  11:opensm-libs            ###########################################  [ 16%]
  12:libmthca               ###########################################  [ 17%]
  13:libmlx4                ###########################################  [ 19%]
  14:libibcm                ###########################################  [ 20%]
  15:libibcommon-devel      ###########################################  [ 22%]
  16:libibumad-devel        ###########################################  [ 23%]
  17:libibmad               ###########################################  [ 25%]
  18:librdmacm              ###########################################  [ 26%]
  19:mpi-selector           ###########################################  [ 28%]
  20:libsdp                 ###########################################  [ 29%]
  21:mvapich_gcc            ###########################################  [ 30%]
  22:openmpi_gcc            ###########################################  [ 32%]
  23:ofed-scripts           ###########################################  [ 33%]
  24:libibverbs-devel       ###########################################  [ 35%]
  25:libibverbs-devel-static###########################################  [ 36%]
  26:libibverbs-devel-static###########################################  [ 38%]
  27:libibverbs-utils       ###########################################  [ 39%]
  28:libmthca               ###########################################  [ 41%]
  29:libmthca-devel-static  ###########################################  [ 42%]
  30:libmthca-devel-static  ###########################################  [ 43%]
  31:libmlx4                ###########################################  [ 45%]
  32:libmlx4-devel-static   ###########################################  [ 46%]
  33:libmlx4-devel-static   ###########################################  [ 48%]
  34:libibcm                ###########################################  [ 49%]
  35:libibcm-devel          ###########################################  [ 51%]
  36:libibcm-devel          ###########################################  [ 52%]
  37:libibcommon-devel      ###########################################  [ 54%]
  38:libibcommon-static     ###########################################  [ 55%]
  39:libibcommon-static     ###########################################  [ 57%]
  40:libibumad-devel        ###########################################  [ 58%]
  41:libibumad-static       ###########################################  [ 59%]
  42:libibumad-static       ###########################################  [ 61%]
  43:libibmad-devel         ###########################################  [ 62%]
  44:libibmad-devel         ###########################################  [ 64%]
```

```
  45:libibmad-static        ########################################### [ 65%]
  46:libibmad-static        ########################################### [ 67%]
  47:ibsim                  ########################################### [ 68%]
  48:librdmacm-utils        ########################################### [ 70%]
  49:librdmacm-devel        ########################################### [ 71%]
  50:librdmacm-devel        ########################################### [ 72%]
  51:libsdp                 ########################################### [ 74%]
  52:libsdp-devel           ########################################### [ 75%]
  53:libsdp-devel           ########################################### [ 77%]
  54:opensm                 ########################################### [ 78%]
  55:opensm-devel           ########################################### [ 80%]
  56:opensm-devel           ########################################### [ 81%]
  57:opensm-static          ########################################### [ 83%]
  58:opensm-static          ########################################### [ 84%]
  59:perftest               ########################################### [ 86%]
  60:mstflint               ########################################### [ 87%]
  61:sdpnetstat             ########################################### [ 88%]
  62:srptools               ########################################### [ 90%]
  63:rds-tools              ########################################### [ 91%]
  64:ibutils                ########################################### [ 93%]
  65:infiniband-diags       ########################################### [ 94%]
  66:qperf                  ########################################### [ 96%]
  67:ofed-docs              ########################################### [ 97%]
  68:mpitests_mvapich_gcc   ########################################### [ 99%]
  69:mpitests_openmpi_gcc   ########################################### [100%]

Installation finished successfully.
Programming HCA firmware...
Device: /dev/mst/mt25418_pci_cr0
Running: mlxburn -d /dev/mst/mt25418_pci_cr0 -fw ./firmware/fw-25408/fw-25408-
rel.mlx -no
-I- Image burn completed successfully.
Please reboot your system for the changes to take effect.
```

Note: In case your machine has an unsupported HCA device, no firmware update will occur and the error message below will be printed. Please contact your hardware vendor for help on firmware updates.

Error message:

```
-I- Querying device ...
-E- Can't auto detect fw configuration file: ...
```

Step 4. In case the installation script performed firmware updates to InfiniBand hardware, it will ask you to reboot your machine.

Step 5. The script adds the following lines to `/etc/security/limits.conf` for the userspace components such as MPI:

```
* soft memlock unlimited
* hard memlock unlimited
```

These settings unlimit the amount of memory that can be pinned by a user space application. If desired, tune the value unlimited to a specific amount of RAM.

Step 6. For your machine to be part of the InfiniBand fabric, a Subnet Manager must be running on one of the fabric nodes. At this point, Mellanox OFED for Linux has already installed the OpenSM Subnet Manager on your machine. For details on starting OpenSM, please refer to the "OpenSM – Subnet Manager" chapter in *Mellanox OFED for Linux User's Manual* under `docs/`.

Step 7. Run the `hca_self_test.ofed` utility to verify whether or not the InfiniBand link is up. The utility also checks for and displays additional information such as

- HCA firmware version
- Kernel architecture
- Driver version
- Number of active HCA ports along with their states
- Node GUID

Note: For more details on `hca_self_test.ofed`, see the file `hca_self_test.readme` under `docs/`.

```
host1# /usr/bin/hca_self_test.ofed

---- Performing InfiniBand HCA Self Test ----
Number of HCAs Detected ................ 1
PCI Device Check ....................... PASS
Kernel Arch ............................ x86_64
Host Driver Version .................... OFED-1.3 1.3-2.6.9_42.ELsmp
Host Driver RPM Check .................. PASS
HCA Firmware on HCA #0 ................. 2.3.0
HCA Firmware Check on HCA #0 ........... PASS
Host Driver Initialization ............. PASS
Number of HCA Ports Active ............. 0
Port State of Port #0 on HCA #0 ........ INIT
Port State of Port #0 on HCA #0 ........ DOWN
Error Counter Check on HCA #0 .......... PASS
Kernel Syslog Check .................... PASS
Node GUID on HCA #0 .................... 00:02:c9:03:00:00:10:e0
------------------ DONE ----------------------
```

Note: After the installer completes, information about the Mellanox OFED installation such as prefix, kernel version, and installation parameters can be retrieved by running the command `/etc/infiniband/info`.

Installation Results

Software

- The OFED and MFT packages are installed under the /usr directory.
- The kernel modules are installed under:

- InfiniBand subsystem:

  `/lib/modules/`uname -r`/updates/kernel/drivers/infiniband/`

- mlx4 driver:

  `/lib/modules/`uname -r`/updates/kernel/drivers/net/mlx4/mlx4_core.ko`

- RDS:

  `/lib/modules/`uname -r`/updates/kernel/net/rds/rds.ko`

- Bonding module:

  `/lib/modules/`uname -r`/updates/kernel/drivers/net/bonding/bonding.ko`

- The package kernel-ib-devel include files are placed under `/usr/src/ofa_kernel/include/`. These include files should be used when building kernel modules that use the stack. (Note that the include files, if needed, are “backported” to your kernel.)
- The raw package (un-backported) source files are placed under `/usr/src/ofa_kernel-<ver>`
- The script `openibd` is installed under `/etc/init.d/`. This script can be used to load and unload the software stack.
- The directory `/etc/infiniband` is created with the files `info` and `openib.conf`. The `info` script can be used to retrieve Mellanox OFED installation information. The `openib.conf` file contains the list of modules that are loaded when the `openibd` script is used.
- The file `90-ib.rules` is installed under `/etc/udev/rules.d/`
- The file `/etc/modprobe.conf` is updated to include the following:
  - **`alias ib<n> ib_ipoib`** (for each ib<n> interface)
  - **`alias net-pf-27 ib_sdp`** (for SDP)
- If OpenSM is installed, the daemon `opensmd` is installed under `/etc/init.d/` and `opensm.conf` is installed under `/etc`.
- If IPoIB configuration files are included, `ifcfg-ib<n>` files will be installed under:
  - `/etc/sysconfig/network-scripts/` on a RedHat machine
  - `/etc/sysconfig/network/` on a SuSE machine
- The installation process unlimits the amount of memory that can be pinned by a user space application. See <weblink>Step 5.
- Man pages will be installed under `/usr/share/man/`

Firmware

- The firmware of existing HCA devices will be updated if the following two conditions are fullfilled:
  1. You run the installation script in default mode; that is, *without* the option ‘--without-fw-update’.
  2. The firmware version of the HCA device is older than the firmware version included with the Mellanox OFED ISO image

  Note: If an HCA’s Flash was originially programmed with an Expansion ROM image, the automatic firmware update will also burn an Expansion ROM image.

- In case your machine has an unsupported HCA device, no firmware update will occur and the error message below will be printed. Please contact your hardware vendor for help on firmware updates.

  Error message:

```
-I- Querying device ...
```

```
Note:    -E- Can't auto detect fw configuration file: ...
```

Post-installation Notes

- Most of the Mellanox OFED components can be configured or reconfigured after the installation by modifying the relevant configuration files. See the relevant chapters in *Mellanox OFED for Linux User's Manual* under `docs/`.
- The list of the modules that will be loaded automatically upon boot can be found in the `/etc/infiniband/openib.conf` file.

Updating Firmware After Installation

In case you ran the `mlnxofedinstall` script with the '`--without-fw-update`' option and now you wish to (manually) update firmware on you adapter card(s), you need to perform the following steps:

Note: If you need to burn an Expansion ROM image, please refer to the section titled "Burning the Expansion ROM Image" in the *Mellanox OFED for Linux User's Manual* under `docs/`.

Note: The following steps are also appropriate in case you wish to burn newer firmware which you have downloaded from Mellanox Technologies' Web site (http://www.mellanox.com/support/firmware_download.php).

Step 1. Start `mst`.

```
host1# mst start
```

Step 2. Identify your target InfiniBand device for firmware update.

a. Get the list of InfiniBand device names on your machine.

```
host1# mst status
MST modules:
------------
    MST PCI module loaded
    MST PCI configuration module loaded
    MST Calibre (I2C) module is not loaded

MST devices:
------------
/dev/mst/mt25418_pciconf0        - PCI configuration cycles access.
                                   bus:dev.fn=02:00.0 addr.reg=88 data.reg=92
                                   Chip revision is: A0
/dev/mst/mt25418_pci_cr0         - PCI direct access.
                                   bus:dev.fn=02:00.0 bar=0xdef00000 size=0x100000
                                   Chip revision is: A0
/dev/mst/mt25418_pci_msix0       - PCI direct access.
                                   bus:dev.fn=02:00.0 bar=0xdeefe000 size=0x2000
/dev/mst/mt25418_pci_uar0        - PCI direct access.
                                   bus:dev.fn=02:00.0 bar=0xdc800000 size=0x800000
```

b. Your InfiniBand device is the one with the postfix "_pci_cr0". In the example listed above, this will be `/dev/mst/mt25418_pci_cr0`.

Step 3. Burn firmware.

a. Burning a firmware binary image using **`mstflint`** (that is already installed on your machine).

Please refer to `MSTFLINT_README.txt` under `docs/`.

b. Burning a firmware image from a .mlx file using the **`mlxburn`** utility (that is already installed on your machine).

The following command burns firmware onto the ConnectX IB device with the device name obtained in the example of Step 2.

```
host1$ mlxburn -dev /dev/mst/mt25418_pci_cr0 \
-fw /mnt/firmware/fw-25408/fw-25408-rel.mlx
```

Warning! Make sure that you have the correct *device name*, *firmware path*, and *firmware file name* before running this command. For help, please refer to the *Mellanox Firmware Tools (MFT) User's Manual* under `/mnt/docs/`.

Step 3. Reboot your machine after the firmware burning is completed.

Uninstalling Mellanox OFED

Use the script `/usr/sbin/ofed_uninstall.sh`. to uninstall the Mellanox OFED package. The script is part of the `ofed-scripts` RPM.

Additional Documentation

For additional information, please refer to the release notes and user manuals under the `docs/` folder of your Mellanox OFED for Linux installation.

You can also visit http://www.mellanox.com/products/ofed.php for FAQ, troubleshooting, future updates to this manual, etc.