Mellanox Messaging Accelerator (VMA)

Installation Guide

Version 6.6.4

Mellanox Technologies
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.
Beit Mellanox
PO Box 586 Yokneam 20692
Israel
www.mellanox.com
Tel: +972 (0)74 723 7200
Fax: +972 (0)4 959 3245

Contents

List of Tables

1 Introduction

The Mellanox's Messaging Accelerator (VMA) Installation Guide provides an introduction to installing and running VMA and using sockperf for UDP/TCP latency and benchmark testing.

1.1 System Requirements for VMA 6.6.4

The following table presents the currently certified combinations of stacks and platforms, and supported CPU architectures for VMA 6.6.4.

Table 1: System Requirements

Specification	Value
Network Adapter Cards	A Mellanox ConnectX®-3 network adapter card.
Firmware	ConnectX®-3 v2.31.5050 or newer
Driver Stack	MLNX-OFED 2.2-1.0.1
Tested Operating Systems and Kernels	RHEL 6 update 3 (2.6.32-279) RHEL 6 update 4 (2.6.32-358) Fedora 19 (3.9.5-301) SLES11 SP1 (2.6.32.12-0.7-default) SLES11 SP2 (3.0.13-0.27-default) SLES11 SP3 (3.0.76-0.11-default) Ubuntu 12.04 (3.2.0-39)
CPU Architecture	x86_64 (Intel Xeon)
Minimum memory requirements	1 GB of free memory for installation 800 MB per process running with VMA
Minimum disk space requirements	1 GB
Transport	Ethernet / InfiniBand / VPI

2 Installing VMA

The VMA library is delivered as a user-space library, and is called `libvma.X.Y.Z`. The VMA library is installed as any other RPM package.

Additionally, VMA is also integrated in MLNX_OFED and will be installed automatically with it.

If you intend to install the default VMA version compiled with MLNX_OFED, please refer to section Installing MLNX_OFED (on page 6).

➢ ***Before you begin, please verify the following***

- You are using a supported operating system and a supported CPU architecture for your operating system. See supported combinations in VMA System Requirements.
- Check whether VMA is installed using:

```
#rpm -qil libvma
```

  - If the VMA package is not installed, an appropriate message is displayed.
  - If a VMA package is installed, the RPM logs the VMA package information and the installed file list.

- Uninstall the current VMA software using:

```
#rpm -e libvma
```

2.1 Installing MLNX_OFED

1. Uninstall any previous MLNX_OFED versions:

```
/usr/sbin/ofed_uninstall.sh
```

2. Install the supported MLNX_OFED version for VMA:
   - To install packages required by VMA to support VPI:

```
./mlnxofedinstall --vma
```

   - To install packages required by VMA to support ETH:

```
./mlnxofedinstall –vma-eth
```

     For further information, please refer to the Installing Mellanox OFED (on page 9) section.

3. Verify MLNX_OFED installation completed successfully:

```
/etc/infiniband/info
```

4. Verify the following is configured in the `/etc/modprobe.d/mlnx.conf` file:

```
options ib_uverbs disable_raw_qp_enforcement=1
options mlx4_core fast_drop=1
options mlx4_core log_num_mgm_entry_size=-1
```

5. Start MLNX_OFED:

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

6. Verify that the supported version of ConnectX®-3 firmware is installed.

```
ibv_devinfo
```

NOTE: Configure ConnectX®-3 ports to work with desired transport using the `connectx_port_config` script. Make sure that ports are not configured with auto mode. For further information, please refer to the Port Type Management (on page 17) section.

7. Install the VMA software. Please refer to the Installing the VMA Software (on page 7) section for installation instructions.

2.2 Downloading the VMA Software

➢ ***To download the VMA software:***

1. Download the VMA software from the Mellanox site.
2. Click the **Downloads** tab to download the relevant package.
3. Save the file on your local drive.

2.3 Installing the VMA Software

➢ ***To install the VMA package:***

1. Go to the location where the package was saved.
2. Run the command below to start installation.

```
#rpm -i libvma.X.Y.Z-R.rpm
```

For Ubuntu OS, please run the following command:

```
#dpkg -i libvma.X.Y.Z-R.deb
```

During the installation process the:

- VMA library is copied to */usr/lib64/libvma.so*.
- VMA monitoring utility is installed at */usr/bin/vma_stat*.
- Sockperf, a latency and throughput benchmarking tool for multicast and UDP / TCP unicast traffic is installed at */usr/bin/sockperf*
- README.txt and version information (VMA_VERSION) are installed at */usr/share/doc/libvma-X.Y.Z-R/*
- VMA installs its configuration file, *libvma.conf*, to the following location: */etc/libvma.conf*

➢ ***After the installation process completes:***

It is recommended to manually add persistence for the following system parameters:

- Force IGMP to work in V2 on all interfaces:

  In the file, */etc/sysctl.conf*, add the following line:

```
net.ipv4.conf.all.force_igmp_version = 2
```

- When running over InfiniBand, configure the IPoIB mode to be datagram.

  a. Modify "`SET_IPOIB_CM=no`" in the file "`/etc/infiniband/openib.conf`"

  b. Verify that it is configured to work as UD mode

```
$cat /sys/class/net/ib0/mode
datagram
```

2.4 Uninstalling VMA

VMA is dependent on OFED; therefore, you must uninstall VMA prior to upgrading OFED. Following the OFED upgrade, you can reinstall VMA.

➢ ***To uninstall VMA:***

- For all OSs except for Ubuntu, run:

```
#rpm -e libvma
```

- For Ubuntu OS, run:

```
#dpkg -r libvma
```

When you uninstall VMA, the *libvma.conf* configuration file is saved with the existing configuration. The path of the saved path is displayed immediately after the uninstall completes.

2.5 Upgrading VMA

VMA 6.6.4 requires an updated version of OFED. Before you upgrade, you must first uninstall your current VMA version.

Uninstall the current VMA software using:

```
#rpm -e libvma
```

Follow the Installation Steps (on page 6) to install MLNX_OFED and the VMA software.

2.5.1 Upgrading libvma.conf

When you upgrade VMA, the *libvma.conf* configuration file is handled as follows:

- If the existing configuration file has been modified since it was installed and is different from the upgraded RPM, the modified version will be left in place, and the version from the new RPM will be installed with a *.rpmnew* suffix.
- If the existing configuration file has not been modified since it was installed, it will automatically be replaced by the version from the upgraded RPM.
- If the existing configuration file has been edited on disk, but is not actually different from the upgraded RPM, the edited version will be left in place; the version from the new RPM will not be installed.

2.6 Running VMA

Before running a user application, you must add the library `libvma.so` to the `env` variable `LD_PRELOAD`. For further information, please refer to the User Manual.

Example:

```
#LD_PRELOAD=libvma.so iperf -uc 224.22.22.22 -t 5
```

Appendix A: Installing Mellanox OFED

Download and install the latest OpenFabrics Enterprise Distribution (OFED) software package via the Mellanox web site at: http://www.mellanox.com/ => Products => Adapter IB/VPI SW => Linux SW/Drivers => Download.

The installation script, mlnxofedinstall, performs the following:

- Discovers the currently installed kernel
- Uninstalls any software 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 and Ethernet network adapters and automatically[1] upgrades the firmware

A.1 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).
- If your kernel version does not match with any of the offered pre-built RPMs, you can add your kernel version by using the "mlnx_add_kernel_support.sh" script located under the docs/ directory.

  Usage:

```
mlnx_add_kernel_support.sh -i|--iso <mlnx iso>[-t|--tmpdir <local work dir>][-v|--ver- bose]
```

NOTE: On Redhat and SLES distributions with errata kernel installed there is no need to use the mlnx_add_kernel_support.sh script. The regular installation can be performed and weak-updates mechanism will create symbolic links to the MLNX_OFED kernel modules.

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

A.2 Installation Script

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

Usage

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

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
-k|--kernel-version <kernel version>	Use provided kernel version instead of 'uname -r'
-p|--print-available	Print available packages for current platform and create corresponding ofed.conf file
--with-32bit	Install 32-bit libraries
--without-32bit	Skip 32-bit libraries installation (Default)
--without-depcheck	Skip Distro's libraries check
--without-fw-update	Skip firmware update
--fw-update-only	Update firmware. Skip driver installation
--force-fw-update	Force firmware update
--force	Force installation
--all|--hpc|--basic|--msm	Install all, hpc, basic or Mellanox Subnet manager packages correspondingly
--vma|--vma-vpi	Install packages required by VMA to support VPI
--vma-eth	Install packages required by VMA to work over Ethernet
--with-vma	Set configuration for VMA use (to be

```
                                    used with any installation parameter).

--guest                             Install packages required by guest os

--hypervisor                        Install packages required by hypervisor
                                    os

-v|-vv|-vvv                         Set verbosity level

--umad-dev-rw                       Grant non root users read/write
                                    permission for umad devices instead of
                                    default

--umad-dev-na                       Prevent from non root users read/write
                                    access for umad devices. Overrides
                                    '--umad-dev-rw'

--enable-affinity                   Run mlnx_affinity script upon boot

--disable-affinity                  Disable mlnx_affinity script (Default)


 --enable-sriov                     Burn SR-IOV enabled firmware - Note:
                                    Enable/Dis- able of SRIOV in a
                                    non-volatile configuration through
                                    uEFI and/or tool will override this
                                    flag.

--add-kernel-support                Add kernel support (Run
                                    mlnx_add_kernel_support.sh)

--skip-distro-check                 Do not check MLNX_OFED vs Distro
                                    matching

--hugepages-overcommit              Setting 80% of MAX_MEMORY as overcommit
                                    for huge page allocation

-q                                  Set quiet - no messages will be printed

--without-<package>                 Do not install package

--with-fabric-collector             Install fabric-collector package.
```

A.3 mlnxofedinstall Return Codes

The table below lists the mlnxofedinstall script return codes and their meanings.

Table 2: mlnxofedinstall Return Codes

Return Code	Meaning
0	The Installation ended successfully

Return Code	Meaning
1	The installation failed
2	No firmware was found for the adapter device
22	Invalid parameter
28	Not enough free space
171	Not applicable to this system configuration. This can occur when the required hardware is not present on the system.
172	Prerequisites are not met. For example, missing the required software installed or the hardware is not config- ured correctly.
173	Failed to start the mst driver

A.4 Installation Procedure

1. Login to the installation machine as root.
2. Mount the ISO image on your machine.

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

3. Run the installation script.

```
Logs dir: /tmp/MLNX_OFED_LINUX-2.2-0.0.9.10694.logs
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. Uninstalling the previous version of MLNX_OFED_LINUX

Starting MLNX_OFED_LINUX-2.2-0.0.9 installation ...

Installing mlnx-ofa_kernel RPM
Preparing...           ########################################################
mlnx-ofa_kernel        ########################################################
Installing kmod-mlnx-ofa_kernel 2.2 RPM
Preparing...           ########################################################
kmod-mlnx-ofa_kernel   ########################################################
Installing mlnx-ofa_kernel-devel RPM
Preparing...           ########################################################
mlnx-ofa_kernel-devel  ########################################################
Installing kmod-kernel-mft-mlnx 3.6.0 RPM
Preparing...           ########################################################
kmod-kernel-mft-mlnx   ########################################################
Installing knem-mlnx RPM
Preparing...           ########################################################
knem-mlnx              ########################################################
Installing kmod-knem-mlnx 1.1.1.90mlnx RPM
Preparing...           ########################################################
kmod-knem-mlnx         ########################################################
Installing ummunotify-mlnx RPM
Preparing...           ########################################################
ummunotify-mlnx        ########################################################
Installing kmod-ummunotify-mlnx 1.0 RPM
Preparing...           ########################################################
kmod-ummunotify-mlnx   ########################################################
Installing kmod-iser 1.2 RPM
Preparing...           ########################################################
kmod-iser              ########################################################
Installing kmod-srp 1.3.2 RPM
Preparing...           ########################################################
kmod-srp               ########################################################
Installing mpi-selector RPM
```

```
Preparing...           ##################################################
mpi-selector           ##################################################
Installing user level RPMs:
Preparing...           ##################################################
ofed-scripts           ##################################################
Preparing...           ##################################################
libibverbs             ##################################################
Preparing...           ##################################################
libibverbs-devel       ##################################################
Preparing...           ##################################################
libibverbs-devel-static ##################################################
Preparing...           ##################################################
libibverbs-utils       ##################################################
Preparing...           ##################################################
libmlx4                ##################################################
Preparing...           ##################################################
libmlx4-devel          ##################################################
Preparing...           ##################################################
libmlx5                ##################################################
Preparing...           ##################################################
libmlx5-devel          ##################################################
Preparing...           ##################################################
libibcm                ##################################################
Preparing...           ##################################################
libibcm-devel          ##################################################
Preparing...           ##################################################
libibumad              ##################################################
Preparing...           ##################################################
libibumad-devel        ##################################################
Preparing...           ##################################################
libibumad-static       ##################################################
Preparing...           ##################################################
libibmad               ##################################################
Preparing...           ##################################################
libibmad-devel         ##################################################
Preparing...           ##################################################
libibmad-static        ##################################################
Preparing...           ##################################################
ibsim                  ##################################################
Preparing...           ##################################################
ibacm                  ##################################################
Preparing...           ##################################################
librdmacm              ##################################################
Preparing...           ##################################################
librdmacm-utils        ##################################################
Preparing...           ##################################################
librdmacm-devel        ##################################################
Preparing...           ##################################################
opensm-libs            ##################################################
Preparing...           ##################################################
opensm                 ##################################################
Preparing...           ##################################################
opensm-devel           ##################################################
Preparing...           ##################################################
opensm-static          ##################################################
Preparing...           ##################################################
dapl                   ##################################################
Preparing...           ##################################################
dapl-devel             ##################################################
Preparing...           ##################################################
dapl-devel-static      ##################################################
Preparing...           ##################################################
dapl-utils             ##################################################
Preparing...           ##################################################
perftest               ##################################################
Preparing...           ##################################################
mstflint               ##################################################
```

```
Preparing...            ##################################################
mft                     ##################################################
Preparing...            ##################################################
srptools                ##################################################
Preparing...            ##################################################
rds-tools               ##################################################
Preparing...            ##################################################
rds-devel               ##################################################
Preparing...            ##################################################
ibutils2                ##################################################
Preparing...            ##################################################
ibutils                 ##################################################
Preparing...            ##################################################
cc_mgr                  ##################################################
Preparing...            ##################################################
dump_pr                 ##################################################
Preparing...            ##################################################
ar_mgr                  ##################################################
Preparing...            ##################################################
ibdump                  ##################################################
Preparing...            ##################################################
infiniband-diags        ##################################################
Preparing...            ##################################################
infiniband-diags-compat ##################################################
Preparing...            ##################################################
qperf                   ##################################################
Preparing...            ##################################################
fca                     ##################################################
Preparing...            ##################################################
mxm                     ##################################################
Preparing...            ##################################################
openmpi                 ##################################################
Preparing...            ##################################################
openmpi                 ##################################################
Preparing...            ##################################################
bupc                    ##################################################
Preparing...            ##################################################
infinipath-psm          ##################################################
Preparing...            ##################################################
infinipath-psm-devel    ##################################################
Preparing...            ##################################################
mvapich2                ##################################################
Preparing...            ##################################################
hcoll                   ##################################################
Preparing...            ##################################################
libibprof               ##################################################
Preparing...            ##################################################
mlnxofed-docs           ##################################################
Preparing...            ##################################################
mpitests_mvapich2_2_0rc1##################################################
Preparing...            ##################################################
mpitests_openmpi_1_6_5  ##################################################
Preparing...            ##################################################
mpitests_openmpi_1_8    ##################################################
Device (06:00.0):
06:00.0 Network controller: Mellanox Technologies MT27500 Family [ConnectX-3]
Link Width: 8x
PCI Link Speed: 5Gb/s Installation finished successfully.

Attempting to perform Firmware update... Querying Mellanox devices firmware
...

Device #1:
----------
Device Type:    ConnectX3Pro Part Number:  MCX354A-FCC_Ax
```

```
Description:    ConnectX-3 Pro VPI adapter card; dual-port QSFP; FDR IB
(56Gb/s) and 40GigE;PCIe3.0 x8 8GT/s;RoHS R6
PSID:    MT_1090111019
PCI Device Name: 0000:05:00.0
Versions: Current      Available FW 2.30.8000    2.31.5000
PXE      3.4.0224     3.4.0224

Status:  Update required
Found 1 device(s) requiring firmware update... Device #1: Updating FW ... Done
A restart is needed for updates to take effect.
Log File: /tmp/MLNX_OFED_LINUX-2.2-0.0.9.10694.logs/fw_update.log To load
the new driver, run:
/etc/init.d/openibd restart
```

NOTE: In case your machine has the latest firmware, no firmware update will occur and the installation script will print at the end of installation a message similar to the following:

```
Device #1:

----------

  Device Type:      ConnectX3Pro

  Part Number:      MCX354A-FCC_Ax

  Description:      ConnectX-3 Pro VPI adapter card; dual-port QSFP;
                    FDR IB (56Gb/s) and 40GigE;PCIe3.0 x8
                    8GT/s;RoHS R6

  PSID:       MT_1090111019

  PCI Device Name:  0000:05:00.0

  Versions:   Current      Available

  FW         2.31.5000     2.31.5000

  PXE        3.4.0224      3.4.0224

Status:       Up to date
```

NOTE: In case your machine has an unsupported network adapter 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:

```
Device #1:

----------

  Device:         0000:05:00.0

  Part Number:

  Description:

  PSID:           MT_0DB0110010

  Versions:       Current        Available

    FW           2.9.1000       N/A

  Status:         No matching image found
```

4. Reboot the machine **if** the installation script performed firmware updates to your network adapter hardware. Otherwise, restart the driver by running: "/etc/init.d/openibd restart"

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

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

The installation process set the amount of memory that can be pinned by a user space application to unlimited. If desired, tune the value unlimited to a specific amount of RAM.

Note: For your machine to be part of the InfiniBand/VPI 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.

5. **(InfiniBand only)** 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 informa- tion such as

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

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

Appendix B: Port Type Management

ConnectX®-3 ports can be individually configured to work as InfiniBand or Ethernet ports. By default both ConnectX-3 ports are initialized as InfiniBand ports. If you wish to change the port type use the connectx_port_config script after the driver is loaded.

NOTE: When changing port type using the "connectx_port_config" utility, all the HCA's ports and interfaces are brought down, and then brought back up with a new port configuration.

Running "`/sbin/connectx_port_config -s`" will show current port configuration for all ConnectX devices.

Port configuration is saved in the file: /etc/infiniband/connectx.conf. This saved configuration is restored at driver restart only if restarting via "/etc/init.d/openibd restart".

Possible port types are:

- eth – Ethernet
- ib – Infiniband
- auto – Link sensing mode - Detect port type based on the attached network type. If no link is detected, the driver retries link sensing every few seconds.

The table below lists the ConnectX port configurations supported by VPI.

Table 3: Supported ConnectX®-3 Port Configurations

Port 1 Configuration	Port 2 Configuration
ib	ib
ib	eth
eth	eth

The port link type can be configured for each device in the system at run time using the "`/sbin/connectx_port_config`" script. This utility will prompt for the PCI device to be modified (if there is only one it will be selected automatically).

In the next stage the user will be prompted for the desired mode for each port. The desired port configuration will then be set for the selected device.

This utility also has a non-interactive mode:

```
/sbin/connectx_port_config [[-d|--device <PCI device ID>] -c|--conf
<port1,port2>]"
```