Connect. Accelerate. Outperform.™

Mellanox MLX4_EN Driver for VMware ESXi 5.1 and ESXi 5.5 User Manual

Rev. 1.9.10-0

www.mellanox.com

Connect. Accelerate. Outperform.™

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)4 909 7200 ; +972 (0)74 723 7200
Fax: +972 (0)4 959 3245

Mellanox MLX4_EN Driver for VMware ESXi 5.1 and ESXi 5.5 User Manual

Table of Contents

List of Tables

1 Overview

This document provides instructions for installing the MLX4_EN drivers for Mellanox Technologies ConnectX® family based network adapter cards in a VMware ESXi-5.1 and ESXi 5.5 server environment. The ConnectX® family adapters identify on the PCI bus as 25408, 25418, 25448, 26418, 26428, 26438, 26448, 26468, 26478, 26488, 25400, 27500, 27510, 27511, 27520, 27521, 27530, 27531, 27540, 27541, 27550, 27551, 27560, 27561.

1.1 Main Features Overview

- Single/Dual port
- MSI-X / Int-X
- NetQueue support
- Multiple Tx/Rx rings
- Hardware Tx/Rx checksum offload
- Large Send Offload (TCP Segmentation Offload)
- VLAN Tx/Rx acceleration (HW VLAN stripping/insertion)
- Ethtool support
- NAPI support
- Wake-on-Lan (WoL) (only on supported hardware)
- NC-SI
- Auto moderation
- Net dump (Only for ESXi5.5Ux)
- RSS Queues
- Fixed Pass-Through
- 10G
- 40G (Only for ESXi5.5Ux)
- iSER
- RoCE

2 Driver Software

VMware uses a file package called a VIB (VMware Installation Bundle) as the mechanism for installing or upgrading software packages on an ESXi server.

MLX4_EN driver consists of several dependant kernel modules, each with its own .vib file. In order to install the driver, the VIBs need to be installed together.

For this, MLX4_EN driver provides a bundle file, a zip file that contain each module VIB file and metadata file that describes the dependencies between them.

The following steps describe how to download, install, and run the driver.

2.1 Installing and Running the offline_bundle Driver on ESXi-5.1 and ESXi-5.5

1. Copy the offline_bundle zip file to the ESXi 5.1 or ESXi 5.5 machine.
2. Extract its contents.
3. Install the driver.
   a. Remove any earlier version of the driver from your VMware ESXi server machine prior to installing the new version (see Section 2.2, “Removing the offline_bundle Driver,” on page 4).
   b. Install the MLX4_EN driver offline_bundle package.

```
#> esxcli software vib install -d <path>/<bundle_file>
```

   For example:

```
#> esxcli software vib install -d <path>/MLNX-OFED-ESX-1.9.10.0-10EM-
   550.0.0.1331820.zip
```

   c. Reboot ESXi server. (The driver will be loaded automatically).

➢ ***To verify that the driver is loaded:***

```
#> esxcli system module list | grep <module_name>
```

For example:

```
#> esxcli system module list | grep mlx4_core
```

➢ ***To query network uplinks installed on your machine:***

```
#> esxcli network nic list
```

The number of uplinks claimed by MLX4_EN module should be displayed.

2.2 Removing the offline_bundle Driver

➢ ***To remove the offline_bundle driver package from the ESXi server machine:***

```
#> esxcli software vib remove -n net-mlx4-ib
#> esxcli software vib remove -n scsi-ib-iser
#> esxcli software vib remove -n net-rdma-cm
#> esxcli software vib remove -n net-ib-addr
#> esxcli software vib remove -n net-ib-cm
#> esxcli software vib remove -n net-ib-umad
#> esxcli software vib remove -n net-ib-sa
#> esxcli software vib remove -n net-ib-mad
#> esxcli software vib remove -n net-ib-core
#> esxcli software vib remove -n net-mlx4-en
#> esxcli software vib remove -n net-mlx4-core
```

To remove the offline_bundle driver, the command must be run in the same order as shown in the example above.

2.3 Loading/Unloading Driver Kernel Modules

➢ ***To unload the driver:***

```
#> /opt/mellanox/bin/openibd.sh stop
```

➢ ***To load the driver:***

```
#> /opt/mellanox/bin/openibd.sh start
```

➢ ***To restart the driver:***

```
#> /opt/mellanox/bin/openibd.sh restart
```

2.4 Driver Default Values

The below are mlx4_en and mlx4_core module parameters.

Some of these values can be changed by using module parameters, which can be obtained by running:

```
#> esxcli system module parameters list -m <module name>
```

For further information, please refer to Section 4.1, “Changing the Driver’s Module Parameters Default Values,” on page 7.

Table 1 - mlx4_core Module Parameters

Parameter	Description	Values
debug_level	Enables debug tracing.	• 1=enabled • 0=disabled [default: 0]
enable_64b_cqe_eqe	Enables 64 byte CQEs/EQEs when it is supported by the firmware.	• 1=enabled • 0=disabled [default: 0]

Table 1 - mlx4_core Module Parameters

Parameter	Description	Values
enable_qos	Enables Quality of Service support in the HCA.	• 1=enabled • 0=disabled [default: 0]
log_mtts_per_seg	Log2 number of MTT entries per segment.	1-7 [default: 3]
log_num_mgm_entry_size	Log2 MGM entry size, that defines the number of QPs per MCG. Not in use with device managed flow steering.	9, 10, 11, 12 [default: 12]
msi_x	Enables MSI-X,	• 1=enabled • 0=disabled [default: 1]
mtu_4k	Enables configuration of 4k MTU.	• 1=enabled • 0=disabled [default: 0]

Table 2 - mlx4_en Module Parameters

Parameter	Description	Values
inline_thold	Threshold for using inline data.	0-104 [default: 104]
netq	Uses netqueue.	• 1=enabled • 0=disabled [default: 1]
netq_num_rings_per_rss	Number of rings per RSS netqueue.	0, 2, 4 [default: 0]
pfcrx	Priority based Flow Control policy on RX. Per priority bit mask. It is 8 bits bit mask, each bit indicates priority [0-7]. Bit value: • 1 - respect incoming pause frames on the specified priority. • 0 - ignore incoming pause frames on the specified priority.	0-255 [default: 0]
pfctx	Priority based Flow Control policy on TX. Per priority bit mask. It's 8 bits bit mask, each bit indicates priority [0-7]. Bit value: • 1 - generate pause frames according to the RX buffer threshold on the specified priority. • 0 - never generate pause frames on the specified priority.	0-255 [default: 0]
udp_rss	Performs RSS for incoming UDP traffic.	• 1=enable • 0=disable [default: 0]
use_rx_frags	Uses RX frags.	• 1=enable • 0=disable [default: 0]

3 Firmware Programming

1. Download the bootable binary image (md5sum: e7b3e9357ca4045fabe2e8a95d951343) from the Mellanox Firmware Tools (MFT) site.
2. Install the image according to the steps described in the README file.

The following procedure requires custom boot image downloading, mounting and booting from a USB device.

4 Additional Driver Settings

4.1 Changing the Driver's Module Parameters Default Values

➢ ***To change the module parameters default values:***

1. Query the available module parameters.

```
#> esxcli system module parameters list -m <module name>
```

2. Set the driver with the required parameters.

```
#> esxcli system parameters set -m <module name> -p <parameters list>
```

For example:

```
#> esxcli system module parameters set -m mlx4_en -p 'netq=1 udp_rss=1'
```

3. Verify that the parameters are set correctly.

```
#> esxcli system module parameters list -m <module name>
```

4. Restart the driver module to apply new default kernel module parameters (see Section 2.3).

4.2 Disabling/Enabling Automatic Load of the Driver upon Boot

1. Query the driver auto load status.

```
#> esxcli system module list
```

2. Disable auto load on boot.

```
#> esxcli system module set -enabled=false -m <module_name>
```

For example:

```
#> esxcli system module set -enabled=false -m mlx4_core
```

3. Enable auto load on boot.

```
#> esxcli system module set -enabled=true -m <module_name>
```

For example:

```
#> esxcli system module set -enabled=true -m mlx4_core
```

4.3 Adding the Device as an uplink to an Existing Vswitch using the CLI

4.3.1 Locally

1. Log into the ESXi server with root permissions.
2. Find your device uplink_name under the "name" column.

```
#> esxcli network nic list
```

3. Add an uplink from a vSwitch.

```
#> esxcli network vswitch standard uplink add -u <uplink_name> -v
<vswitch_name>
```

Once you add a device via the CLI, it is visible in the vSphere client console, thus removing it can be performed via the UI.

4. Check that the uplink was added successfully.

```
#> esxcli network vswitch standard list -v <vswitch_name>
```

➢ ***To remove the device locally:***

1. Log into the ESXi server with root permissions.
2. Remove an uplink from a vSwitch.

```
#> esxcli network vswitch standard uplink remove -u <uplink_name> -v
<vswitch_name>
```

For additional documents, please refer to the VMware site:

https://pubs.vmware.com/vsphere-50/index.jsp?topic=%2Fcom.vmware.vcli.ref.doc_50%2Fesxcli_network.html

4.3.2 Remotely

1. Download and install VMware vSphere Management Assistant (vMA) from:

   https://my.vmware.com/web/vmware/info/slug/datacenter_cloud_infrastructure/vmware_vsphere/5_5
2. Use the command “vicfg-vswitch” from the vMA environment.

For additional documents, please refer to the VMware site:

https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-pubs.html

4.3.3 Renaming the uplink Name

The uplink naming format is in an increasing order, e.g. when working in either Multifunction Mode or Flex10 "vmnic0" to "vmnic7".

If the order is disrupted/inconsistent and you wish to correct, please follow the procedure below:

1. Log into the ESXi server with root permissions.
2. Open the “vi /etc/vmware/esx.conf” file.
3. Locate the /device/<PCi device>/vmkname = "vmnicX".
4. Change the vmnic numbers to the desired order.
5. Save the file.
6. Reboot the server.

Two vmnics with the same number cannot exist in the same ESXi server.

4.4 Configuring ESXi iSER

4.4.1 Installing the Driver Bundle

1. Activate the SSH in ESXi.
   Configuration -> Security Profile -> Services Properties -> SSH -> Options -> Start.
2. Copy the driver bundle using SCP into the machine.
3. Install the MLNX-OFED-ESX driver.

```
#> esxcli software vib install -d /<bundle_file>
```

   For example:

```
# esxcli software vib install -d /MLNX-OFED-ESX-1.9.8.1.zip
```

4. Reboot the machine.
5. Verify the ib_iser module is loaded.

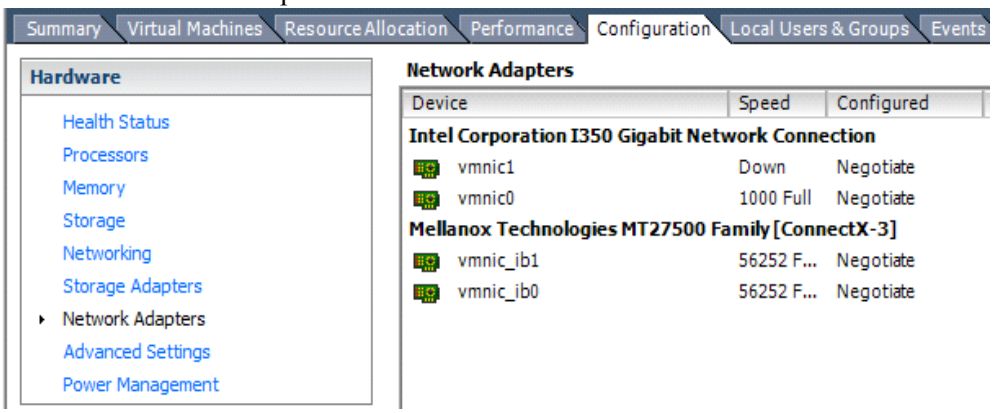
```
# vmkload_mod -l | grep ib_iser
```

➢ ***To load the ib_iser driver:***

```
# vmkload_mod ib_iser
```

4.4.2 Verifying the Mellanox Adapters are Recognized

- `vmnic_ibX` means the device port is in InfiniBand mode.
- `vmnicX` means the device port is in Ethernet mode.



4.4.3 Changing the Port Mode to Ethernet (RoCE)

1. Change the Port Mode to Ethernet.

```
#> esxcli system parameters set -m <module name> -p <port type>
```

   For example:

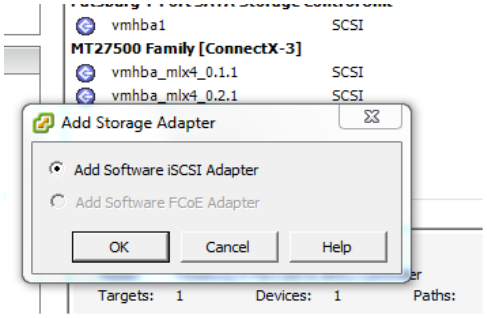
```
# esxcli system module parameters set -m mlx4_core -p port_type_array=2,2
```

2. Reboot the machine.
3. Verify the network adapters are now set as `vmnicX` and not as `vmnic_ibX`.

4.4.4 Adding iSCSI Storage Adapter

1. Add iSCSI Storage Adapter

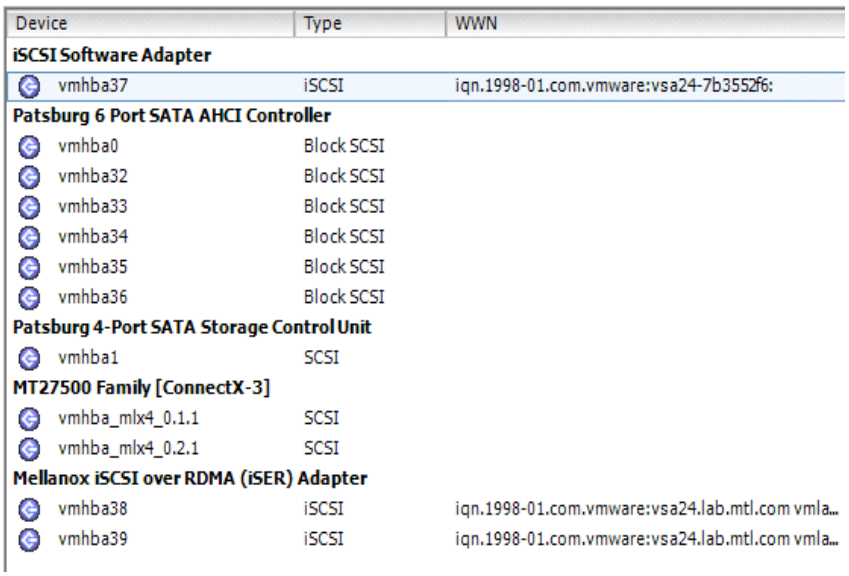
   Go to Configuration -> Storage Adapters -> Add -> OK



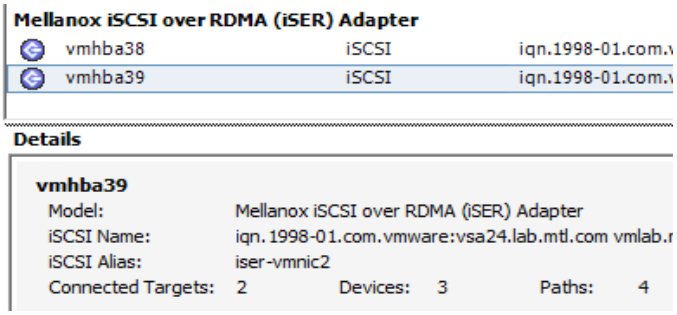
2. Click Refresh.

   The iSCSI Software adapter and the Mellanox iSER adapters, one for each port will be displayed.

   For example:



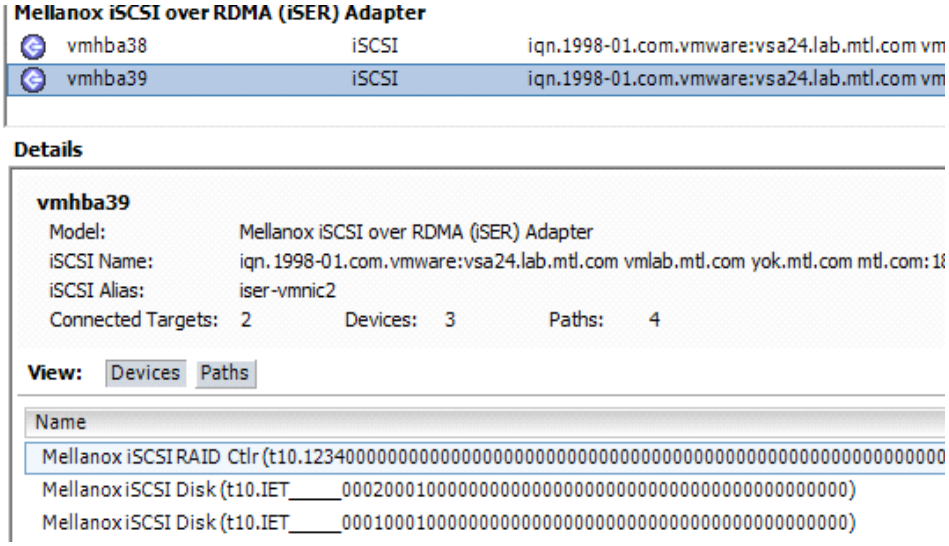
   Each port of the HCA is a `vmnic` in ESXi and each `vmhba` adapter can be bound to a single `vmnic`. Clicking the adapter iSCSI alias will display to which vmnic it can be bound.



4.4.5 Scanning for Targets

Prior to scanning for targets, verify a vSwitch is created with VMKernel for the vmnic needed.

1. Choose the correct `vmhba` for the same `vmnic`.
   Go to Properties -> Network Configuration -> Add.
2. Click OK.
3. Wait several seconds for ESXi to refresh itself and go to Dynamic Discovery -> Add.
4. Enter the IP address of the target machine.
5. Click OK.
   After several seconds the targets are shown in the Static Discovery.
6. Close the window and choose Yes to refresh.
   If the targets are not shown when clicking the `vmhba` adapter, right click the `vmhba` adapter and rescan.



4.4.6 Enabling Flow Control in an Ethernet Switch

Working with RoCE and Ethertnet switch requires enabling Flow Control on the ports. Additionally, working with RoCE and using an Ethernet switch requires having Flow Control enabled in the switch.

1. Connect to the switch using an SSH connection and enter the configuration mode.

```
en
configure terminal
```

2. Enable Flow Control for each port.

```
interface ethernet 1/22
shutdown
flowcontrol receive on
flowcontrol send on
exit
```

4.4.7 Using LUN as a VMFS Datastore

1. Add a storage.
   Go to Configuration -> Storage ->Add Storage -> Disk\LUN
2. Choose LUN.
3. Choose the VMFS version.
4. Enter the datastore name.
5. Start using the VMFS.
   - Edit the VM and add a disk that is saved on the newly created datastore
     or
   - Create/migrate a VM to that datastore