



Connect. Accelerate. Outperform.™

Mellanox MLNX-OFED ESX Driver for VMware ESXi 5.1 and ESXi 5.5 User Manual

Rev. 1.9.10-0

Last Updated: December 23, 2014

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

© Copyright 2014. Mellanox Technologies. All Rights Reserved.

Mellanox®, Mellanox logo, BridgeX®, ConnectX®, Connect-IB®, CoolBox®, CORE-Direct®, InfiniBridge®, InfiniHost®, InfiniScale®, MetroX®, MLNX-OS®, TestX®, PhyX®, ScalableHPC®, SwitchX®, UFM®, Virtual Protocol Interconnect® and Voltaire® are registered trademarks of Mellanox Technologies, Ltd.

ExtendX™, FabricIT™, HPC-X™, Mellanox Open Ethernet™, Mellanox PeerDirect™, Mellanox Virtual Modular Switch™, MetroDX™, Unbreakable-Link™ are trademarks of Mellanox Technologies, Ltd.

All other trademarks are property of their respective owners.

Table of Contents

Chapter 1	Overview	2
1.1	Main Features Overview	2
Chapter 2	Driver Software	3
2.1	Installing and Running the offline_bundle Driver on ESXi-5.1 and ESXi-5.5	3
2.2	Removing the offline_bundle Driver	4
2.3	Loading/Unloading Driver Kernel Modules	4
2.4	Driver Default Values	4
Chapter 3	Firmware Programming	6
Chapter 4	Additional Driver Settings	7
4.1	Changing the Driver's Module Parameters Default Values	7
4.2	Disabling/Enabling Automatic Load of the Driver upon Boot	7
4.3	Adding the Device as an uplink to an Existing Vswitch using the CLI	7
4.3.1	Locally	7
4.3.2	Remotely	8
4.3.3	Renaming the uplink Name	8
4.4	Configuring ESXi iSER	9
4.4.1	Installing the Driver Bundle	9
4.4.2	Verifying the Mellanox Adapters are Recognized	9
4.4.3	Adding iSCSI Storage Adapter	10
4.4.4	Scanning for Targets	11
4.4.5	Enabling Flow Control in an Ethernet Switch	11
4.4.6	Using LUN as a VMFS Datastore	12

List of Tables

Table 1:	mlx4_core Module Parameters	4
Table 2:	mlx4_en Module Parameters	5

1 Overview

This document provides instructions for installing the MLNX-OFED ESX 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.

MLNX-OFED ESX 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, MLNX-OFED ESX 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.	<ul style="list-style-type: none"> • 1=enabled • 0=disabled [default: 0]
enable_64b_cqe_eqe	Enables 64 byte CQEs/EQEs when it is supported by the firmware.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 1=enabled • 0=disabled [default: 0]
log_mmts_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,	<ul style="list-style-type: none"> • 1=enabled • 0=disabled [default: 1]
mtu_4k	Enables configuration of 4k MTU.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 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: <ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 1=enable • 0=disable [default: 0]
use_rx_frags	Uses RX frags.	<ul style="list-style-type: none"> • 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

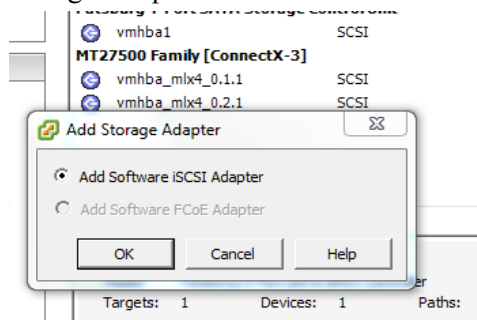
- `vmnic_ibX` means the device port is in InfiniBand mode.
- `vmnicX` means the device port is in Ethernet mode.

Device	Speed	Configured
Intel Corporation I350 Gigabit Network Connection		
vmnic1	Down	Negotiate
vmnic0	1000 Full	Negotiate
Mellanox Technologies MT27500 Family [ConnectX-3]		
vmnic_ib1	56252 F...	Negotiate
vmnic_ib0	56252 F...	Negotiate

4.4.3 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:

Device	Type	WWN
iSCSI Software Adapter		
vmhba37	iSCSI	iqn.1998-01.com.vmware:vsa24-7b3552f6:
Patsburg 6 Port SATA AHCI Controller		
vmhba0	Block SCSI	
vmhba32	Block SCSI	
vmhba33	Block SCSI	
vmhba34	Block SCSI	
vmhba35	Block SCSI	
vmhba36	Block SCSI	
Patsburg 4-Port SATA Storage Control Unit		
vmhba1	SCSI	
MT27500 Family [ConnectX-3]		
vmhba_mlx4_0.1.1	SCSI	
vmhba_mlx4_0.2.1	SCSI	
Mellanox iSCSI over RDMA (iSER) Adapter		
vmhba38	iSCSI	iqn.1998-01.com.vmware:vsa24.lab.mtl.com vmla...
vmhba39	iSCSI	iqn.1998-01.com.vmware:vsa24.lab.mtl.com vmla...

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.

Mellanox iSCSI over RDMA (iSER) Adapter

vmhba38	iSCSI	iqn.1998-01.com.v
vmhba39	iSCSI	iqn.1998-01.com.v

Details

vmhba39			
Model:	Mellanox iSCSI over RDMA (iSER) Adapter		
iSCSI Name:	iqn.1998-01.com.vmware:vsa24.lab.mtl.com vmlab.r		
iSCSI Alias:	iser-vmnic2		
Connected Targets:	2	Devices:	3
		Paths:	4

4.4.6 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