



Mellanox ConnectX-4/ConnectX-5 NATIVE ESXi Driver for VMware vSphere 6.5 User Manual

Rev 4.16.8.8

NOTE:

THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT ("PRODUCT(S)") AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES "ASIS" 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 2017. Mellanox Technologies Ltd. All Rights Reserved.

Mellanox®, Mellanox logo, Accelio®, BridgeX®, CloudX logo, CompustorX®, Connect-IB®, ConnectX®, CoolBox®, CORE-Direct®, EZchip®, EZchip logo, EZappliance®, EZdesign®, EZdriver®, EZsystem®, GPUDirect®, InfiniHost®, InfiniBridge®, InfiniScale®, Kotura®, Kotura logo, Mellanox CloudRack®, Mellanox CloudXMellanox®, Mellanox Federal Systems®, Mellanox HostDirect®, Mellanox Multi-Host®, Mellanox Open Ethernet®, Mellanox OpenCloud®, Mellanox OpenCloud Logo®, Mellanox PeerDirect®, Mellanox ScalableHPC®, Mellanox StorageX®, Mellanox TuneX®, Mellanox Connect Accelerate Outperform logo, Mellanox Virtual Modular Switch®, MetroDX®, MetroX®, MLNX-OS®, NP-1c®, NP-2®, NP-3®, Open Ethernet logo, PhyX®, PlatformX®, PSIPHY®, SiPhy®, StoreX®, SwitchX®, Tilera®, Tilera logo, TestX®, TuneX®, The Generation of Open Ethernet logo, UFM®, Unbreakable Link®, Virtual Protocol Interconnect®, Voltaire® and Voltaire logo are registered trademarks of Mellanox Technologies, Ltd.

All other trademarks are property of their respective owners.

For the most updated list of Mellanox trademarks, visit <http://www.mellanox.com/page/trademarks>

Table of Contents

Table of Contents	3
List Of Tables	4
Release Update History	5
Chapter 1 Overview	6
1.1 Content of MLNX-NATIVE-ESX Driver Package.	6
1.2 Supported HCAs Firmware Versions.	6
1.3 Tested Hypervisors in Paravirtualized and SR-IOV Environments	7
Chapter 2 Changes and New Features in Rev 4.16.8.8	8
Chapter 3 Known Issues	9
Chapter 4 Bug Fixes History	12
Chapter 5 Change Log History	13

List Of Tables

Table 1:	Release Update History	5
Table 2:	Supported Uplinks to Servers	6
Table 3:	Supported HCAs Firmware Versions	6
Table 4:	Tested Hypervisors in Paravirtualized and SR-IOV Environments	7
Table 5:	Changes and New Features.	8
Table 6:	Known Issues	9
Table 7:	Fixed Bugs List	12
Table 8:	Change Log History	13

Release Update History

Table 1 - Release Update History

Release	Date	Description
Rev 4.16.8.8	January 31, 2017	Initial release of this MLNX-NATIVE-ESXi-ConnectX-4 version for ESXi 6.5

1 Overview

These are the release notes of Mellanox ConnectX-4/ConnectX-5 NATIVE ESXi Driver for VMware vSphere 6.5. Mellanox ConnectX-4/ConnectX-5 NATIVE ESXi Driver for VMware vSphere 6.5 supports the following uplinks to servers

Table 2 - Supported Uplinks to Servers

Version	OS	Uplink Speed
4.16.8.8	ESXi 6.5	10/25/40/50/100GbE

1.1 Content of MLNX-NATIVE-ESX Driver Package

The MLNX-NATIVE-ESX-ConnectX-4/ConnectX-5 driver package is distributed as an offline bundle (.zip file) and contains:

- **ESXi 6.5:**

MLNX-NATIVE-ESX-ConnectX-4-5_4.16.8.8-10EM-650.0.0.4240417.zip - Hypervisor bundle for ESXi 6.5 contains the following kernel modules:

- nmlx5_core
- nmlx5_rdma

1.2 Supported HCAs Firmware Versions

MLNX-NATIVE-ESX-ConnectX-4 Rev 4.16.8.8 supports the following Mellanox Ethernet HCA:

Table 3 - Supported HCAs Firmware Versions

HCAs	Recommended Firmware Rev.
ConnectX-4	12.18.1000
ConnectX-4 Lx	14.18.1000
ConnectX-5	16.18.1000
ConnectX-5 Ex	16.18.1000

Please note that older firmware versions were not tested with this release.

For the latest firmware versions, visit:

- http://www.mellanox.com/page/vmware_matrix
- or
- <http://www.mellanox.com/supportdownloader/>

1.3 Tested Hypervisors in Paravirtualized and SR-IOV Environments

Table 4 - Tested Hypervisors in Paravirtualized and SR-IOV Environments

Tested Hypervisors	HCAs	Operating System
SR-IOV	ConnectX-4/ConnectX-4 Lx ConnectX-5/ConnectX-5 Ex	Windows Server 2012 R2
		RedHat 6.5
		RedHat 6.6
		RedHat 7.0
		RedHat 7.1
		RedHat 7.2
		SLES 11.3
Paravirtualized	ConnectX-4/ConnectX-4 Lx ConnectX-5/ConnectX-5 Ex	Windows Server 2012 R2
		RedHat 6.5
		RedHat 6.6
		RedHat 7.0
		RedHat 7.1
		RedHat 7.2
		SLES 11.3

2 Changes and New Features in Rev 4.16.8.8

Table 5 - Changes and New Features

Feature/Change	Description
Adapter Cards	Added support for ConnectX-5/ConnectX-5 Ex adapter cards. Note: ConnectX-5/ConnectX-5 Ex cards are currently at beta level.

3 Known Issues

The following is a list of general limitations and known issues of the various components of this MLNX-NATIVE-ESX-ConnectX-4 release.

Table 6 - Known Issues (Sheet 1 of 3)

Internal Ref.	Description
-	Description: This driver support Ethernet cards only, VPI cards can be configured as Ethernet only mode when burning the firmware. For further information on firmware burning, please refer to the MFT User Manual. (www.mellanox.com > Products > Software > InfiniBand/VPI Drivers > Firmware Tools) WA: N/A Keywords: Protocols, InfiniBand, Ethernet
-	Description: Multicast and IPv6 traffic might be unstable over SR-IOV. WA: N/A Keywords: SR-IOV, IPv6, Multicast
653679	Description: Firmware VF configuration must be N+1 (while N is the required VF number). For example: If your configuration requires 10 VFs, the firmware must be set to support 11 VFs (ESXi Limitation). WA: N/A Keywords: VFs
778371	Description: Wake-on-LAN does not notify when invalid parameters are provided. WA: N/A Keywords: WoL
778572	Description: Nested ESXi might not function properly. WA: N/A Keywords: Nested ESXi
765008	Description: Device RSS fails to hash traffic to sufficient RX rings with Broadcast traffic. WA: N/A Keywords: RSS, RX rings
852883	Description: In stress condition ‘Watchdog’ may appear, leading to uplink going up and down. WA: N/A Keywords: uplink, watchdog
698142/637104	Description: Traffic loss of large packets might occur after MTU change. WA: Reboot the driver Keywords: MTU, Traffic loss

Table 6 - Known Issues (Sheet 2 of 3)

Internal Ref.	Description
775668	Description: VMs can get Call Trace upon MTU change during heavy traffic.
	WA: N/A
	Keywords: VM, MTU, heavy traffic
776274	Description: Reloading the driver while the SR-IOV VFs are ON will result in a Purple Screen of Death (PSOD).
	WA: N/A
	Keywords: Driver reload, SR-IOV
784293	Description: VGT traffic over VXLAN interfaces is currently not supported.
	WA: N/A
	Keywords: VGT traffic, VXLAN
862168	Description: VMs with SR-IOV cannot be powered on when running low on available vectors.
	WA: N/A
	Keywords: VM, SR-IOV
777205	Description: When setting a trunk range in an SR-IOV VF that operates in VGT+ mode, untagged traffic will pass.
	WA: N/A
	Keywords: VM, SR-IOV, VLAN trunk
685558	Description: There is no traffic between PV and SR-IOV VF connected to different ports on the same HCA.
	WA: N/A
	Keywords: PV, SR-IOV VF, HCA
858972	Description: Setting the "Allow Guest MTU Change" option in vSphere Client is currently not functional. Although guest MTU changes in SR-IOV are allowed, they do not affect the port's MTU and the guest's MTU remains the same as the PF MTU.
	WA: N/A
	Keywords: MTU, SR-IOV
910292	Description: Running with ConnectX-4/ConnectX-4 Lx older firmware versions, might result in the following internal firmware errors: <ul style="list-style-type: none"> • Device health compromised • synd 0x1: firmware internal error • extSync 0x94ee
	WA: Upgrade your firmware to the latest version 12.17.2020/14.17.2020
	Keywords: Firmware

Table 6 - Known Issues (Sheet 3 of 3)

Internal Ref.	Description
-	Description: Geneve options length support is limited to 56B. The behavior of received packets with options length bigger than 56B is undefined.
	WA: N/A
	Keywords: Geneve
939198	Description: VST mode in SR-IOV in ConnectX-5 is currently not functional.
	WA: N/A
	Keywords: VST Mode, SR-IOV
958154	Description: NetQ RSS for encapsulated traffic is currently not supported. Encapsulated traffic (VXLAN/Geneve) directed to NetQ RSS queue will not be distributed through all queues' channels, thus will not utilize the RSS feature.
	Note: It is highly recommended to avoid requesting RSS for encapsulated interfaces, i.e. refrain from defining the following in the VM configuration file: <iface_name>.pniceFeatures=4
	WA: N/A
	Keywords: NetQ RSS, encapsulated traffic

4 Bug Fixes History

Table 7 lists the bugs fixed in this release.

Table 7 - Fixed Bugs List

Internal Ref.	Issue
846359	Description: Fixed an issue which caused the adapter card to get stuck in Down state after setting the ring size to 8192.
	Keywords: Ring size
	Discovered in Release: 4.16.7.8
	Fixed in Release: 4.16.8.10

5 Change Log History

Table 8 - Change Log History (Sheet 1 of 2)

Feature/Change	Description
Rev. 4.16.7.8	
Geneve Stateless Offload	Geneve network protocol is encapsulated into IP frame (L2 tunneling). Encapsulation is suggested as a means to alter the normal IP routing for datagrams, by delivering them to an intermediate destination that would otherwise not be selected based on the (network part of the) IP Destination Address field in the original IP header.
Remote Direct Memory Access (RDMA)	Remote Direct Memory Access (RDMA) is the remote memory management capability that allows server-to-server data movement directly between application memory without any CPU involvement.
Set Link Speed	Enables you to set the link speed to a specific link speed supported by ESXi. For further information, see the User Manual section “ <i>Set Link Speed</i> ”.
Priority Flow Control (PFC)	Applies pause functionality to specific classes of traffic on the Ethernet link. For further information, see the User Manual section “ <i>Priority Flow Control (PFC)</i> ”.
NetQ RSS	Allows the user to configure multiple hardware queues backing up the single RX queue. NetQ RSS improves vMotion performance and multiple streams of IPv4/IPv6 TCP/UDP/IPSEC bandwidth over single interface between the Virtual Machines. For further information, see the User Manual section “ <i>NetQ RSS</i> ”.
Default Queue RSS (DRSS)	Allows the user to configure multiple hardware queues backing up the default RX queue. DRSS improves performance for large scale multicast traffic between hypervisors and Virtual Machines interfaces. For further information, see the User Manual section “ <i>Default Queue Receive Side Scaling (DRSS)</i> ”.
SR-IOV	Single Root IO Virtualization (SR-IOV) is a technology that allows a physical PCIe device to present itself multiple times through the PCIe bus. Support for up to 8 ConnectX-4 ports and up to 16 VFs. For further information, refer to the User Manual
RX/TX Ring Resize	Allows the network administrator to set new RX\TX ring buffer size.
VXLAN Hardware Stateless Offloads for ConnectX®-4	VXLAN hardware offload enables the traditional offloads to be performed on the encapsulated traffic.
NetDump	Enables a host to transmit diagnostic information via the network to a remote netdump service, which stores it on disk. Network-based core-dump collection can be configured in addition to or instead of disk-based coredump collection.
NetQueue	NetQueue is a performance technology in VMware ESXi that significantly improves performance in Ethernet virtualized environments.
Wake-on-LAN	Allows a network administrator to remotely power on a system or to wake it up from sleep mode

Table 8 - Change Log History (Sheet 2 of 2)

Feature/Change	Description
Hardware Offload	<ul style="list-style-type: none"> • Large Send Offload (TCP Segmentation Offload) • RSS (Device RSS)
Hardware Capabilities	<ul style="list-style-type: none"> • Multiple Tx/Rx rings • Fixed Pass-Through • Single/Dual port • MSI-X
Ethernet Network	<ul style="list-style-type: none"> • TX/RX checksum • Auto moderation and Coalescing • VLAN stripping offload