NVIDIA ConnectX-4 onwards NICs NATIVE ESXi 6.7 Driver for VMware vSphere Release Notes v4.17.71.1

Exported on Feb/16/2021 12:49 PM
https://docs.mellanox.com/x/Xh2bAg

Table of Contents

Overview **3**

Release Notes Revision History 3

General Information **4**

Content of MLNX-NATIVE-ESX Driver Package 4

Supported HCAs Firmware Versions 4

Tested Hypervisors in Paravirtualized and SR-IOV Environments 5

Changes and New Features **6**

Bug Fixes in the Version **7**

Known Issues **8**

Change Log History **12**

Bug Fixes History **15**

Overview

NVIDIA® Networking native ESXi drivers enable industry-leading performance and efficiency as non-virtualized environments using hardware offloads such as RDMA over Converged Ethernet (RoCE) on VMware vSphere. NVIDIA® ConnectX®-4 onwards deliver 10/25/40/50 and 100GbE network speeds with ESXi 6.5 onwards, allowing the highest port rate on ESXi today.

Release Notes Revision History

Date	Revision	Description
February 16 2020	4.17.71.1	Initial release of this Release Notes version.

General Information

These are the release notes of NVIDIA® ConnectX®-4 onwards adapter cards for NATIVE ESXi Driver for VMware vSphere 6.7. This driver supports the following uplinks to servers.

Version	OS	Uplink Speed
4.17.71.1	ESXi 6.7, ESXi 6.7 U1	10/25/40/50/100bE

Content of MLNX-NATIVE-ESX Driver Package

ESXi 6.7:
Mellanox-nmlx5_4.17.71.1-1OEM.670.0.0.8169922.zip - Hypervisor bundle for ESXi 6.7 contains the following kernel modules:

- nmlx5_core
- nmlx5_rdma

Supported HCAs Firmware Versions

MLNX-NATIVE-ESX Rev 4.17.71.1 supports the following Mellanox Ethernet HCA and their corresponding firmware version:

HCAs	Minimal Recommended Firmware Rev.
ConnectX-4	12.28.2006
ConnectX-4 Lx	14.29.1016
ConnectX-5 / ConnectX-5 Ex	16.29.1016
ConnectX-6	20.29.1016
ConnectX-6 Dx	22.29.1016
ConnectX-6 Lx	26.29.1016

For the latest firmware versions, visit: https://www.mellanox.com/support/firmware/firmware-downloads

Tested Hypervisors in Paravirtualized and SR-IOV Environments

Tested Hypervisors	HCAs	Guest Operating System
SR-IOV	ConnectX-4 ConnectX-4 Lx ConnectX-5/ConnectX-5 Ex ConnectX-6 ConnectX-6 Dx ConnectX-6 Lx	Windows Server 2016 DC RedHat 8.0 RedHat 7.5 RedHat 7.3 RedHat 6.10 RedHat 6.3 SLES 12 SP4 SLES 12 SP3
Paravirtualized[a] (Ethernet Only)	ConnectX-4 ConnectX-4 Lx ConnectX-5/ConnectX-5 Ex ConnectX-6 ConnectX-6 Dx ConnectX-6 Lx	Windows Server 2016 DC RedHat 8.0 RedHat 7.5 RedHat 7.3 RedHat 6.10 RedHat 6.3 SLES 12 SP4 SLES 12 SP3

a. Paravirtualized RDMA is supported only in Linux Operating Systems and in this release it was tested for RedHat 7.5 only.

Changes and New Features

Feature/Change	Description
4.17.71.1	
Adapter Cards	Added support for ConnectX-6 Lx devices.
RDMA 10K QP support	Scaled support for up to 10K connections over RDMA networks.
SR-IOV InfiniBand	SR-IOV InfiniBand is at GA level.
supported_num_ports module parameter	Updated the supported_num_ports default value to 1 to lower memory constraints. **Note:** The user must set a value corresponding to the amount of ports installed in the system.
sriov_mc_isolation	Added sriov_mc_isolation module parameter to isolate multicast traffic to SR-IOV interfaces. Default value is OFF.

Bug Fixes in the Version

The table below lists the bugs fixed in this release. For older issues, please refer to Bug Fixes History.

Internal Ref.	Description
2139652	**Description:** SRI-OV is not supported while ENS is enabled.
	Keywords: SRI-OV, ENS
	Discovered in Version: 4.17.70.1
	Fixed in Release: 4.17.71.1
1712298	**Description:** Live unload of the driver is not supported. Doing so may cause a PSOD if the max_vfs parameter is set.
	Keywords: Driver load
	Discovered in Version: 4.17.15.6
	Fixed in Release: 4.17.71.1
2120216	**Description:** The maximum number of established active RDMA connections (QPs) is currently 5000.
	Keywords: QPs, RDMA
	Discovered in Version: 4.17.70.1
	Fixed in Release: 4.17.71.1
2125202	**Description:** ENS is currently not supported in ConnectX-6 Dx adapter cards.
	Keywords: ENS
	Discovered in Version: 4.17.70.1
	Fixed in Release: 4.17.71.1
2130911	**Description:** Setting ETS value to 0 may cause WQE timeout.
	Keywords: ETS, QOS
	Discovered in Version: 4.17.70.1
	Fixed in Release: 4.17.71.1

Known Issues

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

Internal Ref.	Description
2429623	**Description:** Enabling sriov_mc_isolation module parameter may result in vmknic and emulated NICs multicast and IPv6 traffic loss.
	Workaround: Unset or set the module parameter to 0.
	Keywords: Multicast, IPv6, SR-IOV
	Discovered in Version: 4.17.71.1
2139469	**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.
	Workaround: N/A
	Keywords: MTU, SR-IOV
	Discovered in Version: 4.17.71.1
1340255	**Description:** ECN statistic counters accumulatorsPeriod and ecnMarkedRocePackets display wrong values and cannot be cleared.
	Workaround: N/A
	Keywords: nmlx5 ecn nmlxcli
	Discovered in Version: 4.17.71.1
1340275	**Description:** ECN tunable parameter initialAlphaValue for the Reaction Point protocol cannot be modified.
	Workaround: N/A
	Keywords: nmlx5 ecn nmlxcli
	Discovered in Version: 4.17.71.1
2430662	**Description:** ConnectX-6 Dx speed remains zero after port goes down and reboot is performed.
	Workaround: Turn the down and then up again
	Keywords: ConnectX-6 Dx, link speed
	Discovered in Version: 4.17.71.1

Internal Ref.	Description
2449578	**Description:** When in ENS mode, changing the scheduler to HCLK, may cause traffic loss.
	Workaround: N/A
	Keywords: ENS, HCLK scheduler
	Discovered in Version: 4.17.71.1
1671303	**Description:** A PSOD may occur during vMotion over ENS VMK. This issue is pending VMWare investigation.
	Workaround: N/A
	Keywords: ENS, vMotion
	Discovered in Version: 4.17.15.16
1682956	**Description:** During ENS uplink detachment from the ENS DVS, the below error message regarding the queue still being allocated or that the requested queue is not in use may appear. *"Driver covers for OS issue and the messages are for information only."*
	Workaround: N/A
	Keywords: ENS, queue
	Discovered in Version: 4.17.15.16
1446060	**Description:** Although the max_vfs module parameter range is "0-128", due to firmware limitations, the following are the supported VFs per single port devices: • ConnectX-4 / ConnectX-5: up to 127
	Workaround: N/A
	Keywords: SR-IOV, VFs per port
	Discovered in Version: 4.17.14.2
1340275	**Description:** ECN tunable parameter initialAlphaValue for the Reaction Point protocol cannot be modified.
	Workaround: N/A
	Keywords: nmlx5 ecn nmlxcli
	Discovered in Version: 4.17.13.8
1340255	**Description:** ECN statistic counters accumulatorsPeriod and ecnMarkedRocePackets display wrong values and cannot be cleared.
	Workaround: N/A
	Keywords: nmlx5 ecn nmlxcli

Internal Ref.	Description
	Discovered in Version: 4.17.13.8
-	**Description:** The hardware can offload only up to 256B of headers.
	Workaround: N/A
	Keywords: Hardware offload
	Discovered in Version: 4.17.13.8
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.
	Workaround: N/A
	Keywords: MTU, SR-IOV
746100	**Description:** The 'esxcli mellanox uplink link info -u <vmnic_name>' command reports the 'Auto negotiation' capability always as 'true'.
	Workaround: N/A
	Keywords: 'Auto negotiation' capability
1068621	**Description:** SMP MADs (ibnetdiscover, sminfo, iblinkinfo, smpdump, ibqueryerr, ibdiagnet and smpquery) are not supported on the VFs.
	Workaround: N/A
	Keywords: SMP MADs
778371	**Description:** Wake-on-LAN does not notify when invalid parameters are provided.
	Workaround: N/A
	Keywords: WoL
778572	**Description:** Nested ESXi might not function properly.
	Workaround: N/A
	Keywords: Nested ESXi
765008	**Description:** Device RSS fails to hash traffic to sufficient RX rings with Broadcast traffic.
	Workaround: N/A
	Keywords: RSS, RX rings
852883	**Description:** In stress condition 'Watchdog' may appear, leading to uplink going up and down.
	Workaround: N/A

Internal Ref.	Description
	Keywords: uplink, watchdog

Change Log History

Feature/Change	Description
4.17.70.1	
Power Limitation	Disabled the option of shutting down the link due to power limitation.
Differentiated Services Code Point (DSCP)	Added support for trusting Differentiated Services Code Point (DSCP) and setting default value for RoCE traffic.
SR-IOV VF Counters	Added a new counter that enables the user to query per Virtual Function counters.
RX Counters	Added the RX out-of-buffer counter to indicate any lack of software receive buffers.
SR-IOV	SR-IOV InfiniBand is at beta level.
Rev. 4.17.15.16	
Enhanced Network Stack (ENS)	Enhanced data path is a networking stack mode, which when configured provides superior network performance. It is primarily targeted for NFV workloads, which requires the performance benefits provided by this mode. ENS utilizes DPDK Poll Mode driver model and significantly improves packet rate and latency for small message sizes. This feature is compliant with NSX-T version 2.3.1.0.0.11294271. **Note:** This driver version maintains support of all previous Ethernet functionalities, and can operate in both ENS and slow-path network stack mode, based on the DVS configuration.
Rev. 4.17.14.2	
Virtualization	Removed a VF driver limitation. Now the driver can support the maximum number of VFs supported by the firmware.
Bug Fixes	See "Bug Fixes History" section.
Rev. 4.17.13.8	
Explicit Congestion Notification (ECN)	Explicit Congestion Notification (ECN) is an extension to the Internet Protocol and to the Transmission Control Protocol. ECN allows end-to- end notification of network congestion without dropping packets. To configure ECN behavior, download the nmlxcli tool from the Mellanox site. For further information, refer to the User Manual section *Explicit Congestion Notification (ECN)*.
Dynamic RSS	Improves network performance by allowing OS Load Balancer better RSS RX queue utilization during heavy traffic of the same type. For further information, refer to the User Manual section *Dynamic RSS*.

Multiple RSS Engines	Improves network performance by exposing multiple RSS RX queues to the hypervisor network stack. For further information, refer to the User Manual section *Multiple RSS Engines*.
Packet Capture Utility	Packet Capture utility duplicates all traffic, including RDMA, in its raw Ethernet form (before stripping) to a dedicated "sniffing" QP, and then passes it to an ESX drop capture point. It allows gathering of Ethernet and RoCE bidirectional traffic via pktcap- uw and viewing it using regular Ethernet tools, e.g. Wireshark To enable/disable packet capture, download the nmlxcli tool from the Mellanox site. For further information, refer to the User Manual section Packet Capture Utility.
SR-IOV max_vfs module parameter Type Modification	Changed the type of the SR-IOV max_vfs module parameter from a single integer value to an array of unsigned integers. For further information, refer to the User Manual.
InfiniBand SR-IOV	Enables the creation of InfiniBand virtual functions, allowing the guests to operate over an InfiniBand fabric.
DCBX Negotiation Support for PFC	PFC port configuration can now be auto-negotiated with switches that support the DCBX protocol.
ESXi CLI	Added ESXi CLI support for ESXi 6.7
Adapter Cards	Added support for ConnectX-5/ConnectX-5 Ex adapter cards. **Note:** ConnectX-5/ConnectX-5 Ex cards are currently at beta level.
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. **Note:** It is recommended to use RoCE with PFC enabled in driver and network switches. For how to enable PFC in the driver see section *Priority Flow Control (PFC)* in the User Manual.
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 *"Set Link Speed"*.

Priority Flow Control (PFC)	Applies pause functionality to specific classes of traffic on the Ethernet link. For further information, see the User Manual section *"Priority Flow Control (PFC)"*.
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 *"NetQ RSS"*.
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 *"Default Queue Receive Side Scaling (DRSS)"*.
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 coredump 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
Hardware Offload	• Large Send Offload (TCP Segmentation Offload) • RSS (Device RSS)
Hardware Capabilities	• Multiple Tx/Rx rings • Fixed Pass-Through • Single/Dual port • MSI-X
Ethernet Network	• TX/RX checksum • Auto moderation and Coalescing • VLAN stripping offload

Bug Fixes History

The table below lists the bugs fixed in this release.

Internal Ref.	Description
1677627	**Description:** IPv6 as inner packet is not supported by VMWare yet. This capability will be added in next NSX-T release.
	Keywords: ENS, IPv6, Geneve
	Discovered in Version: 4.17.15.16
	Fixed in Release: 4.17.70.1
1668029	**Description:** There is no Geneve traffic on NSX-T 2.3 when DVS is in standard (non- ENS) mode.
	Keywords: ENS, Geneve
	Discovered in Version: 4.17.15.16
	Fixed in Release: 4.17.70.1
781277	**Description:** The "esxcli network sriovnic vf stats" command is not supported. When running this command on a vmknic, a failure message is displayed.
	Keywords: esxcli SR-IOV
	Discovered in Version: 4.17.13.8
	Fixed in Release: 4.17.70.1
1358381	**Description:** Fixed an issue that prevented ESXi from being discovered via the CDP protocol on ConnectX-4 Lx adapter cards.
	Keywords: CDP protocol, ConnectX-4 Lx
	Discovered in Release: 4.17.13.8
	Fixed in Release: 4.17.14.2

Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. Neither NVIDIA Corporation nor any of its direct or indirect subsidiaries (collectively: "NVIDIA") make any representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.
NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.
Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.
NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.
NVIDIA products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer's own risk.
NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA accepts no liability related to any default, damage, costs, or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document or (ii) customer product designs.
No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.
Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.
THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, "MATERIALS") ARE BEING PROVIDED "AS IS." NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA's aggregate and cumulative

liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, and Mellanox are trademarks and/or registered trademarks of Mellanox Technologies Ltd. and/or NVIDIA Corporation in the U.S. and in other countries. Other company and product names may be trademarks of the respective companies with which they are associated.
For the complete and most updated list of Mellanox trademarks, visit http://www.mellanox.com/page/trademarks

Copyright

© 2021 Mellanox Technologies Ltd. All rights reserved.