SUSE Linux Enterprise Server (SLES) 15 SP2 Inbox Driver User Manual

SLES 15 SP2

Document History

Version	Date	Description of Change
SLES 15 SP2	November 2020	Initial release of this document

Table of Contents

1 Firmware Burning

1. Identify the adapter card's PSID.

```
# mstflint -d 81:00.0 q
Image type:          FS2
FW Version:          2.36.5000
FW Release Date:     26.1.2016
Rom Info:            type=PXE version=3.4.718 devid=4103
Device ID:           4103
Description:         Node             Port1            Port2
Sys image
GUIDs:               e41d2d0300b3f590 e41d2d0300b3f591 e41d2d0300b3f592
e41d2d0300b3f593
MACs:                                     e41d2db3f591     e41d2db3f592
VSD:
PSID:                MT_1090111019
```

2. Download the firmware BIN file from the Mellanox website that matches your card's PSID: www.mellanox.com → Support →Support → Firmware Download
3. Burn the firmware.

```
# mstflint -d <lspci-device-id> -i <image-file> b
```

4. Reboot your machine after the firmware burning is completed.

2 Port Type Management

2.1 Port Type Management/VPI Cards Configuration

Note: This tool is supported in the following devices:

- 4th generation devices: ConnectX- and ConnectX-3 Pro (FW 2.31.5000 and above).
- 5th generation devices: Connect-IB, ConnectX-4, ConnectX-4 Lx, ConnectX-5, ConnectX-6 and ConnectX-6 Dx.

Device ports can be individually configured to work as InfiniBand or Ethernet ports. By default, device ports are initialized as InfiniBand ports. If you wish to change the port type, use the mstflint tool after the driver is loaded.

1. Install mstflint tools: Zypper install mstflint.
2. Check the PCI address.

```
lspci | grep Mellanox
00:06.0 Infiniband controller: Mellanox Technologies MT27520
Family [ConnectX-3 Pro]
```

3. Use mstconfig to change the link type as desired IB – for InfiniBand, ETH – for Ethernet.

   ```
   mstconfig –d <device pci> s LINK_TYPE_P1/2=<ETH|IB|VPI>
   ```

   Example:

   ```
   mstconfig -d 82:00.1 s LINK_TYPE_P1=ETH
   ```

4. Reboot your machine.

3 Modules Loading and Unloading

Mellanox modules for ConnectX®-3/ConnectX®-3 Pro are:

- mlx4_en, mlx4_core, mlx4_ib

Mellanox modules for Connect-IB/ConnectX®-4/ConnectX®-4 Lx/ConnectX®-5/ ConnectX®-6/ ConnectX®-6 Dx are:

- mlx5_core, mlx5_ib
- To load and unload the modules, use the commands below:
  - Loading the driver: `modprobe <module name>`

    ```
    modprobe mlx5_ib
    ```

  - Unloading the driver: `modprobe –r <module name>`

    ```
    modprobe –r mlx5_ib
    ```

4 Important Packages and Their Installation

rdma-core

rdma-core	RDMA core userspace libraries and daemons

libibmad: Low layer InfiniBand diagnostic and management programs

libibmad	OpenFabrics Alliance InfiniBand MAD library

opensm: InfiniBand Subnet Manager

opensm-libs	Libraries used by OpenSM and included utilities
opensm	OpenIB InfiniBand Subnet Manager and management utilities

Ibutils: OpenIB Mellanox InfiniBand Diagnostic Tools

ibutils-libs	Shared libraries used by ibutils binaries
ibutils	OpenIB Mellanox InfiniBand Diagnostic Tools
infiniband-diags: OpenFabrics Alliance InfiniBand Diagnostic Tools	
infiniband-diags	OpenFabrics Alliance InfiniBand Diagnostic Tools
perftest: IB Performance tests	
perftest	IB Performance Tests
mstflint: Mellanox Firmware Burning and Diagnostics Tools	
mstflint	Mellanox firmware burning tool

- To install the packages above run:

```
# zypper –n install <package-name>
```

5 SR-IOV Configuration

5.1 Setting up SR-IOV in ConnectX Adapters

1. Download mstflint tools: zypper install mstflint.
2. Check the device's PCI.

```
lspci | grep Mellanox
```

3. Check if SR-IOV is enabled in the firmware.

```
mstconfig -d <device pci> q
```

Example:

```
# mstconfig -d 81:00.0 q

Device #1:
----------

Device type:    ConnectX3Pro
PCI device:     81:00.0

Configurations:                                 Current
         SRIOV_EN                               True(1)
         NUM_OF_VFS                             0
         LINK_TYPE_P1                           VPI(3)
         LINK_TYPE_P2                           VPI(3)
         LOG_BAR_SIZE                           3
         BOOT_PKEY_P1                           0
         BOOT_PKEY_P2                           0
         BOOT_OPTION_ROM_EN_P1                  True(1)
         BOOT_VLAN_EN_P1                        False(0)
```

```
        BOOT_RETRY_CNT_P1                   0
        LEGACY_BOOT_PROTOCOL_P1             PXE(1)
        BOOT_VLAN_P1                        1
        BOOT_OPTION_ROM_EN_P2               True(1)
        BOOT_VLAN_EN_P2                     False(0)
        BOOT_RETRY_CNT_P2                   0
        LEGACY_BOOT_PROTOCOL_P2             PXE(1)
        BOOT_VLAN_P2                        1
        IP_VER_P1                           IPv4(0)
        IP_VER_P2                           IPv4(0)…
```

4. Check SRIOV_EN and NUM_OF_VFS configurations.
5. Enable SR-IOV:

```
mstconfig –d <device pci> s SRIOV_EN=<False|True>
```

6. Configure the needed number of VFs.

```
mstconfig –d <device pci> s NUM_OF_VFS=<NUM>
```

Note: This file will be generated only if IOMMU is set in the grub.conf file (by adding ¨intel_iommu=on¨ to /boot/grub/grub.conf file).

7. [**mlx4 devices only**] Edit the file /etc/modprobe.d/mlx4.conf:

```
options mlx4_core num_vfs=[needed num of VFs] port_type_array=[1/2
for IB/ETH],[ 1/2 for IB/ETH]
```

Example:

```
options mlx4_core num_vfs=8 port_type_array=1,1
```

8. [**mlx5 devices only**] Write to the sysfs file the number of needed VFs.

```
echo [num_vfs] > /sys/class/infiniband/mlx5_0/device/sriov_numvfs
```

9. Reboot the driver.
10. Load the driver and verify that the VFs were created.

```
lspci | grep mellanox
```

Example:

```
dev-r-vrt-214:~ # lspci | grep nox
82:00.0 Ethernet controller: Mellanox Technologies MT27700 Family
[ConnectX-4]
82:00.1 Ethernet controller: Mellanox Technologies MT27700 Family
[ConnectX-4]
82:00.2 Ethernet controller: Mellanox Technologies MT27700 Family
[ConnectX-4 Virtual Function]
82:00.3 Ethernet controller: Mellanox Technologies MT27700 Family
[ConnectX-4 Virtual Function]
82:00.4 Ethernet controller: Mellanox Technologies MT27700 Family
[ConnectX-4 Virtual Function]
82:00.5 Ethernet controller: Mellanox Technologies MT27700 Family
[ConnectX-4 Virtual Function]
```

For further information, refer to section Setting Up SR-IOV MLNX_OFED User Manual.

Default RoCE Mode Setting for RDMA_CM Application

1. Create a directory for the mlx4/mlx5 device.

```
mkdir -p /sys/kernel/config/rdma_cm/mlx4_0/
```

2. Validate what is the used RoCE mode in the default_roce_mode configfs file.

```
# cat /sys/kernel/config/rdma_cm/mlx4_0/ports/1/default_roce_mode
IB/RoCE v1
```

3. Change the default RoCE mode,
   - For RoCE v1: IB/RoCE v1
   - For RoCE v2: RoCE v2

```
# echo "RoCE v2" >
/sys/kernel/config/rdma_cm/mlx4_0/ports/1/default_roce_mode
# cat /sys/kernel/config/rdma_cm/mlx4_0/ports/1/default_roce_mode
RoCE v2
```

```
# echo "IB/RoCE v1" >
/sys/kernel/config/rdma_cm/mlx4_0/ports/1/default_roce_mode
# cat /sys/kernel/config/rdma_cm/mlx4_0/ports/1/default_roce_mode
IB/RoCE v1
```

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. NVIDIA Corporation (“NVIDIA”) makes no 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.

Trademarks

NVIDIA, the NVIDIA logo, and Mellanox are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and 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

NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051
http://www.nvidia.com