IB Fabric congestion Analysis v16

**HyungKwang Choi (2023.5)**

Table of Contents

[**1. Introduction** 3](#_Toc157060088)

[**1.1 Revision History** 3](#_Toc157060089)

[**1.2 Glossary** 3](#_Toc157060090)

[**1.3 Reference** 3](#_Toc157060091)

[**2. Script running** 4](#_Toc157060092)

[**3. Before running.** 5](#_Toc157060093)

[**3.1 What it does** 5](#_Toc157060094)

[**3.2 Python installation.** 5](#_Toc157060095)

[**3.2 Usage.** 5](#_Toc157060096)

[**4. Fabric Congestion key factors** 6](#_Toc157060097)

[**4.1 LinkDownCounter > 0** 6](#_Toc157060098)

[**4.2 PortXmitDiscards > 0** 6](#_Toc157060099)

[**4.3 port\_fec\_uncorrectable\_block\_counter > 0** 7](#_Toc157060100)

[**4.4 ExcessiveBufferOverrunErrors > 0** 8](#_Toc157060101)

[**4.5 max\_retransmission\_rate > 500** 8](#_Toc157060102)

[**4.6 BER (Symbol/Effective/Raw)** 8](#_Toc157060103)

[**4.7 CongestionIndex >= 10 (will be deprecated )** 10](#_Toc157060104)

[**4.8 Lost Bandwidth** 11](#_Toc157060105)

[**4.9 (Tier1 & Tier4) Switch -> Servers By TX/RX Bandwidth** 11](#_Toc157060106)

[**5. PM tips for switches “IB/Ethernet/SX6036” from sysdump** 14](#_Toc157060107)

[**5.1 From : Ethernet/IB switch** 14](#_Toc157060108)

[**5.1 From : SX6036** 15](#_Toc157060109)

[**6. Causing Fabric degradation** 20](#_Toc157060110)

[**6.1 Link speed downgrade** 20](#_Toc157060111)

[6.1.1 “Unexpected actual link speed” 20](#_Toc157060112)

# **1. Introduction**

## **1.1** **Revision History**

Table 1 tracks the revision history for this specification.

Table : Revision History

| Revision | Date | Author | Comments |
| --- | --- | --- | --- |
| 1.0 | Sep 28th, 2021 | Jie | Initial Draft |
| 2.0 | Mar 15th, 2022 | HyungKwang Choi | Fixed script bug (FEC Uncorrectable) |
| 3.0 | Apr 14th, 2022 | HyungKwang Choi | Fixed script bug (Symbol BER/Err) |
| 4.0 | Apr 30th, 2022 | HyungKwang Choi | Modifying BER threshold from ‘1e-12’ to ‘1e-13’ |
| 5.0 | May 1th, 2022 | HyungKwang Choi | Added "Max Retransmission\_rate" |
| 5.1 | May 11th, 2022 | HyungKwang Choi | modifed "Running command" parsing code |
| 6.1 | May 13th, 2022 | HyungKwang Choi & Sam | In the Ibdiagnet "2.8.1", ibdiagnet2\_db.csv files , data written like "-1". Which triggers an exception |
| 6.2 | May 26th, 2022 | HyungKwang Choi | added the lib "is\_integer\_dtype" which help to distingush PortXmitWaitExt data Int and Hex |
| 7.2 | June 2th, 2022 | HyungKwang Choi | added "Lost Bandwitdh" |
| 8.2 | June 8th, 2022 | HyungKwang Choi | added "Raw BER" |
| 8.3 | June 14th, 2022 | HyungKwang Choi | Fixed "unsupported extended counters (in older FDR device and old SX6710 GW device)" |
| 9.3 | June 15th, 2022 | HyungKwang Choi | Added  "Switch -> Servers By RX Bandwidth, Switch <-> Switch By RX Bandwidth" |
| 10 | June 15th, 2022 | HyungKwang Choi | Changed Script name to "check\_ib\_link\_status\_v10.py" |
| 11 | June 16th, 2022 | HyungKwang Choi | Changed output layout of "Fabric link Rate in details + Lost Bandwidth" |
| 12 | June 22th, 2022 | HyungKwang Choi | Changed some of output layout, and removed unused libarary. |
| 13 | Mar 20th, 2023 | HyungKwang Choi | df\_pm.replace({'LSA': {'FDR10':'14'}},  inplace = True) |
| 14 | Apr 20th, 2023 | HyungKwang Choi | Added new function for "df\_all\_port\_TX\_RX\_Bandwidth" |
| 15 | May 11th, 2023 | HyungKwang Choi | Added new function for "Checking Invalid LID number" |
| 16 | May 15th, 2023 | HyungKwang Choi | Modified lid\_checkig syntax |
| 17 | May 16th, 2023 | HyungKwang Choi | Fixed a bug about "Checking Invalid LID number" |
| 18 | June 28th, 2023 | HyungKwang Choi | Fixed a bug about "Checking Invalid LID number" |
| 19 | Jan 25th, 2023 | HyungKwang Choi | integrated existing script "show\_ib\_inventory.py" |

## **1.2 Glossary**

Table 2 lists the terms and acronyms used in this Document.

Table 2: Terminology and Acronyms

| Term | Definition |
| --- | --- |
|  |  |

## **1.3 Reference**

Table 3 lists the reference referred at this Document.

Table 3: Reference

| Subject | Site |
| --- | --- |
| BER (Bit Error Rate) and Troubleshoot v1.0.pdf | My previous knowledge sharing for BER |

# **2. Script running**

C:\Python39\study>check\_ib\_link\_status\_v19.py -h

usage: usage: check\_ib\_link\_status\_v19.py [-h] -i IBDIAGNET\_FOLDER [-o OUTPUT\_FILE] [-n TOP\_N] [-s] [-a] [-v]

Examples (How to run) > :

#python check\_ib\_link\_status\_v15.py -i ./tem/ibdiagnet2

#python check\_ib\_link\_status\_v15.py -i ./tem/ibdiagnet2 -a

#python check\_ib\_link\_status\_v15.py -i ./tem/ibdiagnet2 -a -n 20

#python check\_ib\_link\_status\_v15.py -i ./tem/ibdiagnet2 -a -n 20 -o C:\Python39\study\save\_backdata.xlsx

#python check\_ib\_link\_status\_v15.py -i ./tem/ibdiagnet2 -a -n 20 -s -o C:\Python39\study\save\_backdata.xlsx

optional arguments:

-h, --help show this help message and exit

-i IBDIAGNET\_FOLDER, --ibdiagnet-folder IBDIAGNET\_FOLDER

read data from ibdiagnet2 output folder

-o OUTPUT\_FILE, --output-file OUTPUT\_FILE

write results to xlsx file

-n TOP\_N, --top-n TOP\_N

print first N entries(by default, n = 10)

-s, --skip-lid-checking

skip\_lid\_checking : Skip Invalid LID & Mcast LID checking

-a, --all Display all TX/RX Bandwidth & Lost Bandwidth in details

-v, --version print current script version

-a, --all Display all TX/RX Bandwidth & Lost Bandwidth in details -v, --version print current script version

C:\Python39\study>check\_ib\_link\_status\_v19.py -i ./temp/ibdiagnet2\_06081228/var/tmp/ibdiagnet2 -a

Running: ibdiagnet -r --pc --pm\_pause\_time 300 -P all=1 --extended\_speeds all --pm\_per\_lane --reset\_phy\_info --get\_phy\_info --get\_cable\_info

INFO: The --top-n option is not set, only the first 10 records will be listed here.

LinkDowned Counters:

##################################################

SrcDevice SrcPort SrcGUID LinkDownedCounter DstDevice DstPort DstGUID

MF0;se02ibsw70848:MQM8700/U1 9 0xb8cef603001dae9a 1 MT4123 ConnectX6 Mellanox Technologies 1 0xb8cef60300a1bb2a

MF0;se02ibsw70848:MQM8700/U1 10 0xb8cef603001dae9a 1 MT4123 ConnectX6 Mellanox Technologies 1 0xb8cef60300a1bad2

MF0;se02ibsw70848:MQM8700/U1 11 0xb8cef603001dae9a 1 MT4123 ConnectX6 Mellanox Technologies 1 0xb8cef60300a1bb1a

MF0;se02ibsw70848:MQM8700/U1 12 0xb8cef603001dae9a 1 MT4123 ConnectX6 Mellanox Technologies 1 0xb8cef60300a1bb4a

XmitDidscard Counters:

##################################################

SrcDevice SrcPort SrcGUID PortXmitDiscards DstDevice DstPort DstGUID

MF0;se02ibsw70848:MQM8700/U1 10 0xb8cef603001dae9a 122 MT4123 ConnectX6 Mellanox Technologies 1 0xb8cef60300a1bad2

MF0;se02ibsw70848:MQM8700/U1 9 0xb8cef603001dae9a 1 MT4123 ConnectX6 Mellanox Technologies 1 0xb8cef60300a1bb2a

MF0;se02ibsw70848:MQM8700/U1 11 0xb8cef603001dae9a 1 MT4123 ConnectX6 Mellanox Technologies 1 0xb8cef60300a1bb1a

MF0;se02ibsw70848:MQM8700/U1 12 0xb8cef603001dae9a 1 MT4123 ConnectX6 Mellanox Technologies 1 0xb8cef60300a1bb4a

Port FEC Uncorrectable Counters:

##################################################

SrcDevice SrcPort SrcGUID PortFECUncorrectableBlockCounter DstDevice DstPort DstGUID

MF0;se02ibsw33748:MQM8700/U1 40 0x08c0eb030053a17c 12655 MF0;se02mr05-mgmt01:MCS8500/L20/U1 7 0x08c0eb030073bbb6

MF0;se02ibsw33648:MQM8700/U1 26 0x08c0eb03005cff40 7390 MF0;se02mr02-mgmt01:MCS8500/L05/U1 6 0x043f720300fee418

MF0;se02ibsw43848:MQM8700/U1 30 0x1070fd030029b39a 440 MF0;se02mr05-mgmt01:MCS8500/L10/U1 15 0x1070fd03003274ce

~

~

~

# **3. Before running.**

## **3.1 What it does**

The script is used to collect & sort up which causes Fabric Congestion from ibdiagnet.

Parsing 4 files (ibdiagnet2.log, ibdiagnet2.net\_dump, ibdiagnet2.db\_csv, ibdiagnet2.net\_dump\_ext) from Ibdiagnet, then extract PM info which causes Fabric Congestion.

LinkDownedCounter > 0

PortXmitDiscards > 0

ExcessiveBufferOverrunErrors > 0

PortFECUncorrectableBlockCounter > 0

Max Retransmission\_rate Counters > 500

Lost BandWidth (Gbps)

Raw BER Counters > 1e-5

Effective BER > 1e-13

Symbol BER > 1e-13 or (Symbol Err > 0)

congestion indexes = PortXmitWaitExt / PortXmitPktsExtended > 10

Lost Bandwidth Summary

(Tier1 & Tier4) Switch -> Servers By RX Bandwidth

(Tier1 & Tier4) Servers -> Switch By TX Bandwidth

(Tier2 & Tier3) Switch <-> Switch By RX Bandwidth

(Tier2 & Tier3) Switch <-> Switch By TX Bandwidth

## **3.2 Python installation.**

[root@mtbc-r740-06 ~]# python3 --version

Python 3.6.8

Prerequisites:

Python 3.8+

Pandas library

OpenPyXL library

[root@mtbc-r740-06 ~]# pip3 install pandas

[root@mtbc-r740-06 ~]# pip3 install openpyxl

[root@mtbc-r740-06 ~]# pip3 list

C:\Python39\study>pip3 list

Package Version

--------------- --------

et-xmlfile 1.1.0

fsspec 2022.5.0

numpy 1.22.3

openpyxl 3.0.9

pandas 1.4.1

pip 21.1.3

python-dateutil 2.8.2

pytz 2022.1

setuptools 56.0.0

six 1.16.0

## **3.2 Usage.**

First command, It clears all Fabric PM stats and port Phy info.

The second, it collects the count accumulated.

ibdiagnet -r --pc --pm\_pause\_time 600 -P all=1 --extended\_speeds all --pm\_per\_lane --reset\_phy\_info --get\_phy\_info --get\_cable\_info

ibdiagnet -r -P all=1 --extended\_speeds all --pm\_per\_lane --get\_phy\_info --get\_cable\_info

# **4. Fabric Congestion key factors**

## **4.1 LinkDownCounter > 0**

* File location & Command

./ibdiagnet2.pm:link\_down\_counter=0x00000000

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery 5 13

# Port counters: Lid 5 port 13 (CapMask: 0x5300)

PortSelect:......................13

CounterSelect:...................0x0000

SymbolErrorCounter:..............0

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0 🡸================

PortRcvErrors:...................0

PortRcvRemotePhysicalErrors:.....0

PortRcvSwitchRelayErrors:........0

PortXmitDiscards:................0

PortXmitConstraintErrors:........0

PortRcvConstraintErrors:.........0

CounterSelect2:..................0x00

LocalLinkIntegrityErrors:........0

ExcessiveBufferOverrunErrors:....0

QP1Dropped:......................0

VL15Dropped:.....................0

PortXmitData:....................1596816

PortRcvData:.....................1596888

PortXmitPkts:....................22178

PortRcvPkts:.....................22179

PortXmitWait:....................0

## **4.2 PortXmitDiscards > 0**

* File location & Command

./ibdiagnet2.pm:port\_xmit\_discard=0x00000000

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery 5 13

# Port counters: Lid 5 port 13 (CapMask: 0x5300)

PortSelect:......................13

CounterSelect:...................0x0000

SymbolErrorCounter:..............0

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0

PortRcvErrors:...................0

PortRcvRemotePhysicalErrors:.....0

PortRcvSwitchRelayErrors:........0

PortXmitDiscards:................0 🡸================

PortXmitConstraintErrors:........0

PortRcvConstraintErrors:.........0

CounterSelect2:..................0x00

LocalLinkIntegrityErrors:........0

ExcessiveBufferOverrunErrors:....0

QP1Dropped:......................0

VL15Dropped:.....................0

PortXmitData:....................1596816

PortRcvData:.....................1596888

PortXmitPkts:....................22178

PortRcvPkts:.....................22179

PortXmitWait:....................0

ibportstate 5 79 reset

ibportstate 5 79 query

ibportstate 5 79 disable

ibportstate 5 79 enable

## **4.3 port\_fec\_uncorrectable\_block\_counter > 0**

* File location & Command

./ibdiagnet2.pm:port\_fec\_uncorrectable\_block\_counter=0x00000000

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery -T 5 13

# PortExtendedSpeedsCounters counters: Lid 5 port 13

PortSelect:......................13

CounterSelect:...................0x0000000000000000

SyncHeaderErrorCounter:..........0

UnknownBlockCounter:.............0

ErrorDetectionCounterLane0:......0

ErrorDetectionCounterLane1:......0

ErrorDetectionCounterLane2:......0

ErrorDetectionCounterLane3:......0

ErrorDetectionCounterLane4:......0

ErrorDetectionCounterLane5:......0

ErrorDetectionCounterLane6:......0

ErrorDetectionCounterLane7:......0

ErrorDetectionCounterLane8:......0

ErrorDetectionCounterLane9:......0

ErrorDetectionCounterLane10:.....0

ErrorDetectionCounterLane11:.....0

FECCorrectableBlockCtrLane0:.....0

FECCorrectableBlockCtrLane1:.....0

FECCorrectableBlockCtrLane2:.....0

FECCorrectableBlockCtrLane3:.....0

FECCorrectableBlockCtrLane4:.....0

FECCorrectableBlockCtrLane5:.....0

FECCorrectableBlockCtrLane6:.....0

FECCorrectableBlockCtrLane7:.....0

FECCorrectableBlockCtrLane8:.....0

FECCorrectableBlockCtrLane9:.....0

FECCorrectableBlockCtrLane10:....0

FECCorrectableBlockCtrLane11:....0

FECUncorrectableBlockCtrLane0:...0

FECUncorrectableBlockCtrLane1:...0

FECUncorrectableBlockCtrLane2:...0

FECUncorrectableBlockCtrLane3:...0

FECUncorrectableBlockCtrLane4:...0

FECUncorrectableBlockCtrLane5:...0

FECUncorrectableBlockCtrLane6:...0

FECUncorrectableBlockCtrLane7:...0

FECUncorrectableBlockCtrLane8:...0

FECUncorrectableBlockCtrLane9:...0

FECUncorrectableBlockCtrLane10:..0

FECUncorrectableBlockCtrLane11:..0

# mlxlink -d lid-3 -pc 36

# mlxlink -d lid-3 -m -c -e -p 36 --show\_device --show\_serdes\_tx --show\_fec --show\_ber\_monitor

* Tips

 > to get ‘fec\_uncorrectable\_block\_counter’, you have to add below options while running ibdiagnet

# --extended\_speeds all

> below extra fields are collected when you add the options “--extended\_speeds all”

sync\_header\_error\_counter=0x00000000

unknown\_block\_counter=0x00000000

fec\_corrected\_symbol\_counter\_total=0x0000000000000000

fec\_corrected\_symbol\_counter\_lane[0]=0x00000000

fec\_corrected\_symbol\_counter\_lane[1]=0x00000000

fec\_corrected\_symbol\_counter\_lane[2]=0x00000000

fec\_corrected\_symbol\_counter\_lane[3]=0x00000000

port\_fec\_correctable\_block\_counter=0x00008281

port\_fec\_uncorrectable\_block\_counter=0x00000000

port\_fec\_corrected\_symbol\_counter=0x00008292

 > Clear FEC counters

#perfquery -x -R 0x20 1 (lid 20)

> with ‘Perfquery’ command, you will see  ‘FECUncorrectableBlockCtrLaneXXXX’. Those are sum of ..

FECCorrectableBlockCtrLaneX/FECUncorrectableBlockCtrLaneX vs. port\_fec\_correctable\_block\_counter / port\_fec\_uncorrectable\_block\_counter

## **4.4 ExcessiveBufferOverrunErrors > 0**

* File location & Command

./ibdiagnet2.pm: port\_fec\_uncorrectable\_block\_counter=0x0000044c

./ibdiagnet2.db\_csv

## **4.5 max\_retransmission\_rate > 500**

## **4.6 BER (Symbol/Effective/Raw)**

* BER threshold

Table

Description automatically generated

* File location & Command

ibdiagnet2.net\_dump\_ext

ibdiagnet2.db\_csv

\*\* For details, please refer to my previous knowledge sharing (BER (Bit Error Rate) and Troubleshoot v1.0.pdf)

**> on a switch.**

TestLab [standalone: master] # show interfaces ib 10/1/14

IB15/1/4 state:

Logical port state : Active

Physical port state : LinkUp

Current line rate : 200.0 Gbps

Supported speeds : sdr, qdr, fdr, edr, hdr

Speed : hdr

Supported widths : 1X, 2X, 4X

Width : 4X

Max supported MTUs : 4096

MTU : 4096

VL capabilities : VL0 - VL7

Operational VLs : VL0 - VL3

Description :

IB Subnet : infiniband-default

Phy-profile : high-speed-ber

Width reduction mode : Not supported

Telemetry sampling : Disabled

Telemetry threshold : Disabled

Telemetry record : Disabled

Telemetry threshold level: N/A bytes

RX:

Bytes : 3480

Packets : 13

Errors : 7

Symbol errors : 735 <====

VL15 dropped packets: 0

TX:

Bytes : 11232

Packets : 39

Wait : 0

Discarded packets: 0

TestLab [standalone: master] # show interfaces ib 10/1/14 link-diagnostics

-----------------------------------------------------------------------

Interface Code Status

-----------------------------------------------------------------------

IB10/1/14 15 Bad signal integrity

[root@NVIDIA]# perfquery 199 14

# Port counters: Lid 199 port 14 (CapMask: 0x5300)

PortSelect:......................14

CounterSelect:...................0x0000

SymbolErrorCounter:..............38 <=======

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0

PortRcvErrors:...................0

[root@NVIDIA]# mlxlink -d lid-199 -p 14 -c -e -m

Operational Info

----------------

State : Active

Physical state : LinkUp

Speed : IB-HDR

Width : 4x

FEC : LL-FEC (271,257) + PLR

Loopback Mode : No Loopback

Auto Negotiation : ON

Supported Info

--------------

Enabled Link Speed : 0x00000061 (HDR,EDR,SDR)

Supported Cable Speed : 0x00000061 (HDR,EDR,SDR)

Troubleshooting Info

--------------------

Status Opcode : 15

Group Opcode : PHY FW

Recommendation : Bad signal integrity. <======

Physical Counters and BER Info

------------------------------

Time Since Last Clear [Min] : 143.1 <=== important to math BER rate

Effective Physical Errors : 122 <=========== This used for symbol error calculation

Raw Physical Errors Per Lane : 73506956266,43017712645,42717472449,58570852076

Effective Physical BER : 7E-8

Raw Physical BER : 1E-4

Link Down Counter : 0

Link Error Recovery Counter : 0

EYE Opening Info

----------------

Physical Grade : 1894, 2031, 2412, 2097

Height Eye Opening [mV] : N/A, N/A, N/A, N/A

Phase Eye Opening [psec] : N/A, N/A, N/A, N/A

Module Info

-----------

Identifier : QSFP28

Compliance : N/A

Cable Technology : 850 nm VCSEL

Cable Type : Active cable (active copper / optics)

OUI : Mellanox

Vendor Name : Mellanox

Vendor Part Number : P35346-001

Vendor Serial Number : THY1120057

Rev : A1

Wavelength [nm] : 850

Transfer Distance [m] : 30

Attenuation (5g,7g,12g) [dB] : N/A

FW Version : 38.100.59

Digital Diagnostic Monitoring : Yes

Power Class : 5.0 W max

CDR RX : ON,ON,ON,ON

CDR TX : ON,ON,ON,ON

LOS Alarm : N/A

Temperature [C] : 53 [-10..80]

Voltage [mV] : 3232.9 [3100..3500]

Bias Current [mA] : 7.240,7.326,7.346,7.258 [5.492..8.5]

Rx Power Current [dBm] : -2,-2,-2,-2 [-12..6] <== check if RX level is proper or not.

Tx Power Current [dBm] : -2,-2,-2,0 [-14..6]

* How to troubleshoot

These are typically able to be cleared by reseating or replacing the cable.

**> Port reset**

# mlxlink -d lid-3 -p 1 –port\_state DN //(DN: Down, UP : Up, TG : Toggle)

# mlxlink -d lid-3 -p 36 -c -e -m //verify

#(ibportstate <lid number> <port number> reset)

Ex)

#Ibportstate 5 79 query

#ibportstate 5 79 disable

#ibportstate 5 79 enable

**> After Cable replacement or port change, Please check the status.**

    Mlxlink -d lid-<lid number> <portnumber>

Ex)

# mlxlink -d lid-3 -pc 36 <=== port counter clear

# mlxlink -d lid-3 -p 36 --show\_device  --show\_serdes\_tx     --show\_fec   --show\_ber\_monitor

# mlxlink -d lid-3 -m -c -e -p 36  ç== FEC uncorrectable/physical error counter

## **4.7 CongestionIndex >= 10 (will be deprecated )**

* File location & Command

./ibdiagnet2.pm: port\_xmit\_wait, port\_xmit\_pkts

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery 5 13 -x

# Port counters: Lid 5 port 13 (CapMask: 0x5300)

PortSelect:......................13

CounterSelect:...................0x0000

SymbolErrorCounter:..............0

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0

PortRcvErrors:...................0

PortRcvRemotePhysicalErrors:.....0

PortRcvSwitchRelayErrors:........0

PortXmitDiscards:................0

PortXmitConstraintErrors:........0

PortRcvConstraintErrors:.........0

CounterSelect2:..................0x00

LocalLinkIntegrityErrors:........0

ExcessiveBufferOverrunErrors:....0

QP1Dropped:......................0

VL15Dropped:.....................0

PortXmitData:....................1596816

PortRcvData:.....................1596888

PortXmitPkts:....................22178

PortRcvPkts:.....................22179

PortXmitWait:....................471366584

* Explanation on fabric ‘Congestion Indexes’

[PortXmitWait and Symbol Errors | Salesforce](https://mellanox.lightning.force.com/lightning/r/Knowledge_Article__kav/ka21T000000bpWSQAY/view?ws=%2Flightning%2Fr%2FCase%2F50050000016gclzAAA%2Fview)

**PortXmitWait** - This is not an error counter. It is a very high-resolution counter that is incremented every internal clock tick of the device whenever there is a packet queued on an output port, and this packet cannot be sent. This can happen whenever there is some temporary congestion in the fabric. For instance, if two nodes are sending to one node you can see this counter increment, or if there is a node mismatch in rates you can see this counter increment. When the counter increments it does not mean that any packets dropped. Packets are just stored in the buffer of the port until they can be sent. Therefore, it is natural to see these PortXmitWait counters increment under normal conditions.

* How to math PortXmitWait

CongestionIndex >= 10

   => PortXmitWaitExt/PortXmitPktsExtended

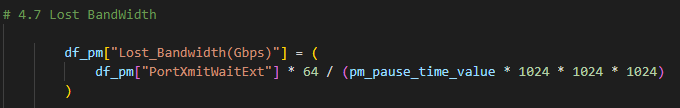
## **4.8 Lost Bandwidth**

* How to math

Lost Bandwidth/sec = (xmit\_wait\_ext\_end-xmit\_wait\_ext\_start/test\_time\_in\_sec) *\*64*

For Gb/s : lost\_bw /(1024 x 1024 x 1024)

* From script code.



## **4.9 (Tier1 & Tier4) Switch -> Servers By TX/RX Bandwidth**

* File location & Command

./ibdiagnet2.pm: port\_xmit\_wait, port\_xmit\_pkts

./ibdiagnet2.db\_csv

[root@My\_test\_lab ~]# perfquery 5 13 -x

# Port counters: Lid 5 port 13 (CapMask: 0x5300)

PortSelect:......................13

CounterSelect:...................0x0000

SymbolErrorCounter:..............0

LinkErrorRecoveryCounter:........0

LinkDownedCounter:...............0

PortRcvErrors:...................0

PortRcvRemotePhysicalErrors:.....0

PortRcvSwitchRelayErrors:........0

PortXmitDiscards:................0

PortXmitConstraintErrors:........0

PortRcvConstraintErrors:.........0

CounterSelect2:..................0x00

LocalLinkIntegrityErrors:........0

ExcessiveBufferOverrunErrors:....0

QP1Dropped:......................0

VL15Dropped:.....................0

PortXmitData:....................1596816

PortRcvData:.....................1596888

PortXmitPkts:....................22178

PortRcvPkts:.....................22179

PortXmitWait:....................471366584

* knowledge

Text

Description automatically generated

* How to math

Table

Description automatically generated

* From script code.

Text

Description automatically generated

# **5. PM tips for switches “IB/Ethernet/SX6036” from sysdump**

## **5.1 From : Ethernet/IB switch**

* From sysdump ‘Sysinfo.txt’

1. Please find sysinfo.txt file from sysdump

* **Ethernet switch**

Eth1/1:

Admin state : Enabled

Operational state : Up

Last change in operational status : 2w 1d and 2:07:50 ago (1 oper change)

Boot delay time : 0 sec

Description : N\A

Mac address : 24:8a:07:81:04:ec

MTU : 9000 bytes (Maximum packet size 9022 bytes)

Fec : auto

Flow-control : receive off send off

Actual speed : 40 Gbps

Auto-negotiation : Enabled

Width reduction mode : Unknown

Switchport mode : trunk

MAC learning mode : Enabled

Last clearing of "show interface" counters: 00:20:48

60 seconds ingress rate : 506180208 bits/sec, 63272526 bytes/sec, 41173 packets/sec 🡸= See ingress/ egress difference

60 seconds egress rate : 2349264 bits/sec, 293658 bytes/sec, 4457 packets/sec

Rx:

53228689 packets

53095932 unicast packets

132309 multicast packets

448 broadcast packets

80359739790 bytes

809713 discard packets 🡸==== check if it’s increases or not.

0 error packets 🡸====

0 fcs errors

0 undersize packets

0 oversize packets

0 pause packets

0 unknown control opcode

0 symbol errors 🡸====

Tx:

4493032 packets

4317677 unicast packets

175355 multicast packets

0 broadcast packets

1056276187 bytes

0 discard packets 🡸====

0 error packets

0 hoq discard packets 🡸====

* **IB Switch**

IB1/13 state:

Logical port state : Active

Physical port state : LinkUp

Current line rate : 56.0 Gbps

Supported speeds : sdr, ddr, qdr, fdr10, fdr

Speed : fdr

Supported widths : 1X, 4X

Width : 4X

Max supported MTUs : 4096

MTU : 4096

VL capabilities : VL0 - VL7

Operational VLs : VL0 - VL7

Description : ibs1#1

IB Subnet : infiniband-default

Phy-profile : high-speed-ber

Width reduction mode : Disabled

RX bytes : 631437720

RX packets : 7998549

RX errors : 0 🡸====

Symbol errors : 0 🡸====

VL15 dropped packets : 0

TX bytes : 7724808536

TX packets : 5211828

TX wait : 260003439 🡸===

TX discarded packets : 0

## **5.1 From : SX6036**

* From sdkdump

..................................................

Port Performance Counters

..................................................

Port 0x12300 - IEEE 802.3 Counters Group

==================================================

frames\_transmitted\_ok 0

frames\_received\_ok 0

frame\_check\_sequence\_errors 0

alignment\_errors 0

octets\_transmitted\_ok 0

octets\_received\_ok 0

multicast\_frames\_xmitted\_ok 0

broadcast\_frames\_xmitted\_ok 0

multicast\_frames\_received\_ok 0

broadcast\_frames\_received\_ok 0

in\_range\_length\_errors 0

out\_of\_range\_length\_field 0

frame\_too\_long\_errors 0

symbol\_error\_during\_carrier 0

mac\_control\_frames\_transmitted 0

mac\_control\_frames\_received 0

unsupported\_opcodes\_received 0

pause\_mac\_ctrl\_frames\_received 0

pause\_mac\_ctrl\_frames\_transmitted 0

Port 0x12300 - RFC 2863 Counters Group

==================================================

if\_in\_octets 0

if\_in\_ucast\_pkts 0

if\_in\_discards 0

if\_in\_errors 0

if\_in\_unknown\_protos 0

if\_out\_octets 0

if\_out\_ucast\_pkts 0

if\_out\_discards 0

if\_out\_errors 0

if\_in\_multicast\_pkts 0

if\_in\_broadcast\_pkts 0

if\_out\_multicast\_pkts 0

if\_out\_broadcast\_pkts 0

Port 0x12300 - RFC 2819 Counters Group

==================================================

ether\_stats\_drop\_events 0

ether\_stats\_octets 0

ether\_stats\_pkts 0

ether\_stats\_broadcast\_pkts 0

ether\_stats\_multicast\_pkts 0

ether\_stats\_crc\_align\_errors 0

ether\_stats\_undersize\_pkts 0

ether\_stats\_oversize\_pkts 0

ether\_stats\_fragments 0

ether\_stats\_jabbers 0

ether\_stats\_collisions 0

ether\_stats\_pkts64octets 0

ether\_stats\_pkts65to127octets 0

ether\_stats\_pkts128to255octets 0

ether\_stats\_pkts256to511octets 0

ether\_stats\_pkts512to1023octets 0

ether\_stats\_pkts1024to1518octets 0

ether\_stats\_pkts1519to2047octets 0

ether\_stats\_pkts2048to4095octets 0

ether\_stats\_pkts4096to8191octets 0

ether\_stats\_pkts8192to10239octets 0

Port 0x12300 - RFC 3635 Counters Group

==================================================

dot3stats\_alignment\_errors 0

dot3stats\_fcs\_errors 0

dot3stats\_single\_collision\_frames 0

dot3stats\_multiple\_collision\_frames 0

dot3stats\_sqe\_test\_errors 0

dot3stats\_deferred\_transmissions 0

dot3stats\_late\_collisions 0

dot3stats\_excessive\_collisions 0

dot3stats\_internal\_mac\_transmit\_errors 0

dot3stats\_carrier\_sense\_errors 0

dot3stats\_frame\_too\_longs 0

dot3stats\_internal\_mac\_receive\_errors 0

dot3stats\_symbol\_errors 0

dot3control\_in\_unknown\_opcodes 0

dot3in\_pause\_frames 0

dot3out\_pause\_frames 0

Port 0x12300 - CLI Counters Group

==================================================

port\_rx\_octets 0

port\_rx\_frames 0

port\_rx\_jumbo 0

port\_rx\_unicast 0

port\_rx\_multicast 0

port\_rx\_broadcast 0

port\_rx\_no\_buffer 0

port\_rx\_fcs\_errors 0

port\_rx\_runt 0

port\_rx\_other\_errors 0

port\_tx\_octets 0

port\_tx\_frames 0

port\_tx\_jumbo 0

port\_tx\_unicast 0

port\_tx\_multicast 0

port\_tx\_broadcast 0

port\_tx\_errors 0

Port 0x12300 - EXTENDED Counters Group

==================================================

tx\_wait 0

ecn\_marked 0

no\_buffer\_discard\_mc 0

rx\_ebp 0

tx\_ebp 0

rx\_buffer\_almost\_full 0

rx\_buffer\_full 0

tx\_stats\_pkts64octets 0

tx\_stats\_pkts65to127octets 0

tx\_stats\_pkts128to255octets 0

tx\_stats\_pkts256to511octets 0

tx\_stats\_pkts512to1023octets 0

tx\_stats\_pkts1024to1518octets 0

tx\_stats\_pkts1519to2047octets 0

tx\_stats\_pkts2048to4095octets 0

tx\_stats\_pkts4096to8191octets 0

tx\_stats\_pkts8192to10239octets 0

Port 0x12300 - DISCARD Counters Group

==================================================

ingress\_general 0

ingress\_policy\_engine 0

ingress\_vlan\_membership 0

ingress\_tag\_frame\_type 0

egress\_vlan\_membership 0

loopback\_filter 0

egress\_general 0

egress\_link\_down 0

egress\_hoq 0

port\_isolation 0

egress\_policy\_engine 0

ingress\_tx\_link\_down 0

egress\_stp\_filter 0

egress\_hoq\_stall 0

egress\_sll 0 🡸== “ingress\_discard\_all ” Please find below

Port 0x12300 - PER PRIO Counters Group

==================================================

Prio 0

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 1

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 2

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 3

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 4

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 5

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 6

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Prio 7

..............................

rx\_octets 0

rx\_uc\_frames 0

rx\_mc\_frames 0

rx\_bc\_frames 0

rx\_frames 0

tx\_octets 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

rx\_pause 0

rx\_pause\_duration 0

tx\_pause 0

tx\_pause\_duration 0

rx\_pause\_transition 0

rx\_discard 0

Port 0x12300 - PER TC Counters Group

==================================================

TC 0

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 1

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 2

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 3

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 4

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 5

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 6

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

TC 7

..............................

tx\_octet 0

tx\_uc\_frames 0

tx\_mc\_frames 0

tx\_bc\_frames 0

tx\_frames 0

tx\_queue 0

tx\_no\_buffer\_discard\_uc 0

tx\_wred\_discard 0

prsunny@SYD26-0101-0708-15T1:~$ show int counters 🡸= check ingress\_discard\_all   
IFACE STATE RX\_OK RX\_BPS RX\_UTIL RX\_ERR RX\_DRP RX\_OVR TX\_OK TX\_BPS TX\_UTIL TX\_ERR TX\_DRP TX\_OVR  
----------- ------- --------- -------- --------- -------- -------- -------- --------- -------- --  
Ethernet0 U 77401644 N/A N/A 0 0 0 612546543 N/A N/A 0 0 0  
Ethernet4 U 82013052 N/A N/A 0 0 0 548287581 N/A N/A 0 0 0  
Ethernet8 U 75445567 N/A N/A 0 0 0 640693267 N/A N/A 0 0 0  
Ethernet12 U 75069284 N/A N/A 0 0 0 590180253 N/A N/A 0 0 0  
Ethernet16 U 81110674 N/A N/A 0 0 0 512705461 N/A N/A 0 0 0  
Ethernet20 U 74303484 N/A N/A 0 0 0 560971127 N/A N/A 0 0 0  
Ethernet24 U 83767004 N/A N/A 0 0 0 567398414 N/A N/A 0 0 0  
Ethernet28 U 89491970 N/A N/A 0 0 0 527621910 N/A N/A 0 0 0  
Ethernet32 U 77686352 N/A N/A 0 0 0 529482801 N/A N/A 0 0 0  
Ethernet36 U 77047095 N/A N/A 0 0 0 446032780 N/A N/A 0 0 0  
Ethernet40 U 81180399 N/A N/A 0 0 0 522866000 N/A N/A 0 0 0  
Ethernet44 U 83906899 N/A N/A 0 0 0 521436128 N/A N/A 0 0 0  
Ethernet48 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet52 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet56 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet60 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet64 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet68 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet72 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet76 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet80 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet84 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet88 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet92 X 0 N/A N/A 0 0 0 0 N/A N/A 0 0 0  
Ethernet96 U 869161966 N/A N/A 0 0 0 100500628 N/A N/A 0 0 0  
Ethernet100 U 860994567 N/A N/A 0 0 0 104002217 N/A N/A 0 0 0  
Ethernet104 U 901131138 N/A N/A 0 0 0 97657185 N/A N/A 0 0 0  
Ethernet108 U 815075863 N/A N/A 0 0 0 114106192 N/A N/A 0 0 0  
Ethernet112 U 823447607 N/A N/A 0 0 0 92292313 N/A N/A 0 0 0  
Ethernet116 U 857305589 N/A N/A 0 0 0 99577287 N/A N/A 0 0 0  
Ethernet120 U 828557210 N/A N/A 0 0 0 103516688 N/A N/A 0 0 0  
Ethernet124 U 728265885 N/A N/A 0 0 0 103008124 N/A N/A 0 0 0  
  
 Port 0x10300 - DISCARD Counters Group   
==================================================  
ingress\_general                          0  
ingress\_policy\_engine                    0  
ingress\_vlan\_membership                  0  
ingress\_tag\_frame\_type                   97164  
egress\_vlan\_membership                   0  
loopback\_filter                          0  
egress\_general                           0  
egress\_link\_down                         0  
egress\_hoq                               0  
port\_isolation                           0  
egress\_policy\_engine                     4158013  
ingress\_tx\_link\_down                     0  
egress\_stp\_filter                        0  
egress\_hoq\_stall                         0  
egress\_sll                               0  
ingress\_discard\_all                      346428 🡸=== SN2700s with the reason of egress\_policy\_engine. If possible, please modify the default ACL rule to permit all.

# **6. Causing Fabric degradation**

## **6.1 Link speed downgrade**

### 6.1.1 “Unexpected actual link speed”

* From ibdiagnet2.log

With ‘ibdiagnet2.log’, sometimes we can see ‘link speed downgrade’. This can cause congestion.

**Speed / Width checks**

**-I- Link Speed Check (Compare to supported link speed)**

**-E- Links Speed Check finished with errors**

**-E- Link: S248a070300bcc330/N248a070300bcc330/P10<-->Sb8599f0300f990c6/mlx5\_3/P1 - Unexpected actual link speed 10 (enable\_speed1="2.5 or 5 or 10 or 14 or 25 or FDR10", enable\_speed2="2.5 or 5 or 10 or 14 or 25 or 50" therefore final speed should be 25)**

**[root@mtbc-r740-07 ~]# mlxlink -d lid-1 -p 9 -e -e -m**

**Operational Info**

**----------------**

**State : Active**

**Physical state : LinkUp**

**Speed : IB-QDR**

**Width : 4x**

**FEC : No FEC**

**Loopback Mode : No Loopback**

**Auto Negotiation : ON**

**Supported Info**

**--------------**

**Enabled Link Speed : 0x0000001f (FDR,FDR10,QDR,DDR,SDR)**

**Supported Cable Speed : 0x0000001f (FDR,FDR10,QDR,DDR,SDR)**

**Troubleshooting Info**

**--------------------**

**Status Opcode : 34**

**Group Opcode : PHY FW**

**Recommendation : The active speed was degraded from maximal possible speed due to local signal integrity issue.**

* Troubleshooting

# mlxlink -d /dev/mst/mt4123\_pciconf0.1 -m -c -e

# mlxlink -d /dev/mst/mt4123\_pciconf0.1 --link\_mode\_force --speeds QDR

**> port reset on both ports**

#mlxlink -d /dev/mst/mt4123\_pciconf0.1 -a tg

#mlxlink -d lid-1 -p 9 -a tg

#mlxlink -d lid-1 -p 9 -c

**> SW FW reset : service impact..be careful**

#flint -d lid-1 swreset

**> Physically reseat/replace the cable**

[root@mtbc-r740-07 ~]# ibportstate 3 40

Switch PortInfo:

# Port info: Lid 3 port 40

LinkState:.......................Active

PhysLinkState:...................LinkUp

Lid:.............................0

SMLid:...........................0

LMC:.............................0

LinkWidthSupported:..............1X or 4X or 2X

LinkWidthEnabled:................1X or 4X or 2X

LinkWidthActive:.................4X

LinkSpeedSupported:..............2.5 Gbps or 10.0 Gbps

LinkSpeedEnabled:................2.5 Gbps or 10.0 Gbps

LinkSpeedActive:.................10.0 Gbps

LinkSpeedExtSupported:...........14.0625 Gbps or 25.78125 Gbps or 53.125 Gbps

LinkSpeedExtEnabled:.............14.0625 Gbps or 25.78125 Gbps or 53.125 Gbps

LinkSpeedExtActive:..............14.0625 Gbps

# MLNX ext Port info: Lid 3 port 40

StateChangeEnable:...............0x00

LinkSpeedSupported:..............0x00

LinkSpeedEnabled:................0x00

LinkSpeedActive:.................0x00

Peer PortInfo:

# Port info: Lid 3 DR path slid 6; dlid 65535; 0,40 port 36

LinkState:.......................Active

PhysLinkState:...................LinkUp

Lid:.............................0

SMLid:...........................0

LMC:.............................0

LinkWidthSupported:..............1X or 4X

LinkWidthEnabled:................1X or 4X

LinkWidthActive:.................4X

LinkSpeedSupported:..............2.5 Gbps or 5.0 Gbps or 10.0 Gbps

LinkSpeedEnabled:................2.5 Gbps or 5.0 Gbps or 10.0 Gbps

LinkSpeedActive:.................10.0 Gbps

LinkSpeedExtSupported:...........14.0625 Gbps or 25.78125 Gbps

LinkSpeedExtEnabled:.............14.0625 Gbps or 25.78125 Gbps

LinkSpeedExtActive:..............14.0625 Gbps

# MLNX ext Port info: Lid 3 DR path slid 6; dlid 65535; 0,40 port 36

StateChangeEnable:...............0x00

LinkSpeedSupported:..............0x01

LinkSpeedEnabled:................0x01

LinkSpeedActive:.................0x00

ibwarn: [8655] validate\_extended\_speed: Peer ports operating at active extended speed 1 rather than 2 (25.78125 Gbps)

[root@mtbc-r740-07 ~]# ibportstate 1 36

Switch PortInfo:

# Port info: Lid 1 port 36

LinkState:.......................Active

PhysLinkState:...................LinkUp

Lid:.............................0

SMLid:...........................0

LMC:.............................0

LinkWidthSupported:..............1X or 4X

LinkWidthEnabled:................1X or 4X

LinkWidthActive:.................4X

LinkSpeedSupported:..............2.5 Gbps or 5.0 Gbps or 10.0 Gbps

LinkSpeedEnabled:................2.5 Gbps or 5.0 Gbps or 10.0 Gbps

LinkSpeedActive:.................10.0 Gbps

LinkSpeedExtSupported:...........14.0625 Gbps or 25.78125 Gbps

LinkSpeedExtEnabled:.............14.0625 Gbps or 25.78125 Gbps

LinkSpeedExtActive:..............14.0625 Gbps 🡸===========

# MLNX ext Port info: Lid 1 port 36

StateChangeEnable:...............0x00

LinkSpeedSupported:..............0x01

LinkSpeedEnabled:................0x01

LinkSpeedActive:.................0x00

Peer PortInfo:

# Port info: Lid 1 DR path slid 6; dlid 65535; 0,36 port 40

LinkState:.......................Active

PhysLinkState:...................LinkUp

Lid:.............................0

SMLid:...........................0

LMC:.............................0

LinkWidthSupported:..............1X or 4X or 2X

LinkWidthEnabled:................1X or 4X or 2X

LinkWidthActive:.................4X

LinkSpeedSupported:..............2.5 Gbps or 10.0 Gbps

LinkSpeedEnabled:................2.5 Gbps or 10.0 Gbps

LinkSpeedActive:.................10.0 Gbps

LinkSpeedExtSupported:...........14.0625 Gbps or 25.78125 Gbps or 53.125 Gbps

LinkSpeedExtEnabled:.............14.0625 Gbps or 25.78125 Gbps or 53.125 Gbps

LinkSpeedExtActive:..............14.0625 Gbps

# MLNX ext Port info: Lid 1 DR path slid 6; dlid 65535; 0,36 port 40

StateChangeEnable:...............0x00

LinkSpeedSupported:..............0x00

LinkSpeedEnabled:................0x00

LinkSpeedActive:.................0x00

ibwarn: [11087] validate\_extended\_speed: Peer ports operating at active extended speed 1 rather than 2 (25.78125 Gbps)

[root@mtbc-r740-07 ~]# mlxcables -d SW\_MT54000\_BD-BDDWDC4-SPG-383-QM8790-SL20N3\_lid-0x0003,mlx5\_0,1\_cable\_40

Querying Cables ....

Cable #1:

---------

Cable name : SW\_MT54000\_BD-BDDWDC4-SPG-383-QM8790-SL20N3\_lid-0x0003,mlx5\_0,1\_cable\_40

>> No FW data to show

-------- Cable EEPROM --------

Identifier : QSFP+ (0dh)

Technology : Copper cable unequalized (a0h)

Compliance : 40GBASE-CR4, FDR,QDR,DDR,SDR

Attenuation: 2.5GHz : 6dB

5.0GHz : 10dB

7.0GHz : 13dB

12.9GHz : 0dB

25.78GHz : 0dB

OUI : 0x0002c9

Vendor : Mellanox

Serial number : MT1335VS01041

Part number : MC2207128-003

Revision : A3

Temperature [c] : N/A

Digital Diagnostic Monitoring : NO

Length [m] : 3 m

[root@mtbc-r740-07 ~]# mlxcables -d SW\_MT54000\_BD-BDDWDC4-SPG-383-QM8790-SL20N3\_lid-0x0003,mlx5\_0,1\_cable\_39

Querying Cables ....

Cable #1:

---------

Cable name : SW\_MT54000\_BD-BDDWDC4-SPG-383-QM8790-SL20N3\_lid-0x0003,mlx5\_0,1\_cable\_39

>> No FW data to show

-------- Cable EEPROM --------

Identifier : QSFP+ (0dh)

Technology : Copper cable unequalized (a0h)

Compliance : 40GBASE-CR4, FDR,QDR,DDR,SDR

Attenuation: 2.5GHz : 5dB

5.0GHz : 8dB

7.0GHz : 11dB

12.9GHz : 0dB

25.78GHz : 0dB

OUI : 0x0002c9

Vendor : Mellanox

Serial number : MT1534VS06704

Part number : MC2207130-002

Revision : A3

Temperature [c] : N/A

Digital Diagnostic Monitoring : NO

Length [m] : 2 m

[root@mtbc-r740-07 ~]# mlxcables -d SW\_MT53000\_IB\_SW\_SB7890\_lid-0x0001,mlx5\_0,1\_cable\_35

Querying Cables ....

Cable #1:

---------

Cable name : SW\_MT53000\_IB\_SW\_SB7890\_lid-0x0001,mlx5\_0,1\_cable\_35

>> No FW data to show

-------- Cable EEPROM --------

Identifier : QSFP+ (0dh)

Technology : Copper cable unequalized (a0h)

Compliance : 40GBASE-CR4, FDR,QDR,DDR,SDR

Attenuation: 2.5GHz : 5dB

5.0GHz : 8dB

7.0GHz : 11dB

12.9GHz : 0dB

25.78GHz : 0dB

OUI : 0x0002c9

Vendor : Mellanox

Serial number : MT1524VS03303

Part number : MC2207130-002

Revision : A3

Temperature [c] : N/A

Digital Diagnostic Monitoring : NO

Length [m] : 2 m

[root@mtbc-r740-07 ~]# mlxlink -d lid-3 -p 40 -c -e -m

Operational Info

----------------

State : Active

Physical state : LinkUp

Speed : IB-FDR

Width : 4x

FEC : Firecode FEC

Loopback Mode : No Loopback

Auto Negotiation : ON

Supported Info

--------------

Enabled Link Speed : 0x00000015 (FDR,QDR,SDR)

Supported Cable Speed : 0x0000003f (**EDR**,FDR,FDR10,QDR,DDR,SDR) 🡸= Please check the peer side Speed

Troubleshooting Info

--------------------

Status Opcode : 0

Group Opcode : N/A

Recommendation : No issue was observed.

Tool Information

----------------

Firmware Version : 27.2008.3328

amBER Version : 2.00

MFT Version : mft 4.20.0-34

Physical Counters and BER Info

------------------------------

Time Since Last Clear [Min] : 28.5

Symbol Errors : 0

Symbol BER : 15E-255

Effective Physical Errors : 0

Effective Physical BER : 15E-255

Raw Physical Errors Per Lane : 0,0,0,0

Raw Physical BER : 15E-255

Link Down Counter : 0

Link Error Recovery Counter : 0

EYE Opening Info

----------------

Physical Grade : 0, 0, 0, 0

Height Eye Opening [mV] : 72, 96, 114, 55

Phase Eye Opening [psec] : N/A, N/A, N/A, N/A

Module Info

-----------

Identifier : QSFP+

Compliance : N/A

Cable Technology : Copper cable unequalized

Cable Type : Passive copper cable

OUI : Mellanox

Vendor Name : Mellanox

Vendor Part Number : MC2207128-003

Vendor Serial Number : MT1335VS01041

Rev : A3

Wavelength [nm] : N/A

Transfer Distance [m] : 3

Attenuation (5g,7g,12g) [dB] : 10,13,0

FW Version : N/A

Digital Diagnostic Monitoring : No

Power Class : N/A

CDR RX : N/A

CDR TX : N/A

LOS Alarm : N/A

Temperature [C] : N/A

Voltage [mV] : N/A

Bias Current [mA] : N/A

Rx Power Current [dBm] : N/A

Tx Power Current [dBm] : N/A

IB Cable Width : 1x,2x,4x

Memory Map Revision : 2

Linear Direct Drive : 0

Cable Breakout : Channels implemented [1,2,3,4]/Far end is unspecified

SMF Length : N/A

MAX Power : 0

Cable Rx AMP : N/A

Cable Rx Emphasis : N/A

Cable Rx Post Emphasis : N/A

Cable Tx Equalization : N/A

Wavelength Tolerance : N/A

Module State : N/A

DataPath state [per lane] : N/A

Rx Output Valid [per lane] : N/A

Rx Input Valid [per lane] : N/A

Nominal bit rate : 0.000Gb/s

Rx Power Type : OMA

Manufacturing Date : 03\_09\_13

Active Set Host Compliance Code : N/A

Active Set Media Compliance Code: N/A

Error Code Response : N/A

Module FW Fault : N/A

DataPath FW Fault : N/A

Tx Fault [per lane] : N/A

Tx LOS [per lane] : N/A

Tx CDR LOL [per lane] : N/A

Rx LOS [per lane] : N/A

Rx CDR LOL [per lane] : N/A

Tx Adaptive EQ Fault [per lane] : N/A

[root@mtbc-r740-07 ~]# mlxlink -d lid-1 -p 36 -c -e -m

Operational Info

----------------

State : Active

Physical state : LinkUp

Speed : IB-FDR

Width : 4x

FEC : Firecode FEC

Loopback Mode : No Loopback

Auto Negotiation : ON

Supported Info

--------------

Enabled Link Speed : 0x0000001f (FDR,FDR10,QDR,DDR,SDR)

Supported Cable Speed : 0x0000001f (FDR,FDR10,QDR,DDR,SDR)

Troubleshooting Info

--------------------

Status Opcode : 0

Group Opcode : N/A

Recommendation : No issue was observed.

Tool Information

----------------

Firmware Version : 15.2008.3328

MFT Version : mft 4.20.0-34

Physical Counters and BER Info

------------------------------

Time Since Last Clear [Min] : 28.7

Effective Physical Errors : 0

Raw Physical Errors Per Lane : 0,0,0,0

Effective Physical BER : 15E-255

Raw Physical BER : 15E-255

Link Down Counter : 0

Link Error Recovery Counter : 0

EYE Opening Info

----------------

Physical Grade : 39730, 56648, 45507, 53863

Height Eye Opening [mV] : 140, 182, 149, 180

Phase Eye Opening [psec] : 0, 0, 0, 0

Module Info

-----------

Identifier : QSFP+

Compliance : N/A

Cable Technology : Copper cable unequalized

Cable Type : Passive copper cable

OUI : Mellanox

Vendor Name : Mellanox

Vendor Part Number : MC2207128-003

Vendor Serial Number : MT1335VS01041

Rev : A3

Wavelength [nm] : N/A

Transfer Distance [m] : 3

Attenuation (5g,7g,12g) [dB] : 10,13,0

FW Version : N/A

Digital Diagnostic Monitoring : No

Power Class : N/A

CDR RX : N/A

CDR TX : N/A

LOS Alarm : N/A

Temperature [C] : N/A

Voltage [mV] : N/A

Bias Current [mA] : N/A

Rx Power Current [dBm] : N/A

Tx Power Current [dBm] : N/A