Western Digital.

Configuration Guide

Configuring a Mellanox® Network Switch for Lossless Networking with OpenFlex™ Platforms

Abstract

This configuration guide provides an overview of how to configure lossless Ethernet settings on Mellanox based Ethernet network switches with Western Digital® OpenFlex platforms.

April 2025

Table of Contents

Introduction

NVMe-oF™ based storage offers the promise of low latency shared storage. To obtain the performance potential of this technology, Ethernet switches in the network topology must be configured for lossless networking using standard Data Center Bridging (DCB) technologies. While these settings may be unfamiliar to many readers, they are not complicated to understand or follow.

Configuration Process Summary

The process of enabling lossless networking functionality on Mellanox® based Ethernet switches can be broken down into the following six steps:

1. Configure Ports
2. Configure MTU
3. Enable DSCP
4. Configure PFC
5. Configure ECN
6. Show Pertinent Switch Counters

Example Hardware Specifications

In this guide, an SN2100 with Onyx version 3.10.4302 was used. Terminal output shown in this guide may vary based on the product and firmware version. For additional information, see the Data24 Compatibility Matrix. As these instructions will be making many changes to the switch remember to write memory to ensure no settings are lost during a power outage event.

Configuration Table

Included below for convenience is a table to record the lossless configuration values for deployment.

Description	Variable	Example	Deployment Value
PFC Priority	<ROCE_PRI>	3	
CNP Priority	<CNP_PRI>	6	
Port	<PORT>	1/1	

Configure Ports

The purpose of this section is to describe how to configure the ports on the switch so that they will successfully connect to the target product.

OpenFlex Data24

For the OpenFlex Data24, it is recommended to disable speed auto negotiation at the switch port and force the port speed to 100 Gb.

1. Physically connect the OpenFlex Data24 to the switch with a QSFP28 cable.
2. Enter Configuration mode on the switch.

```
# enable
# configure terminal
```

3. Force the link to connect at a 100 Gb link speed.

```
(config) # interface ethernet <PORT> speed 100G force
```

4. Verify the link speed configuration.

```
(config) # show interfaces ethernet <PORT>
```

Enable DSCP

1. Configure trust mode to L3 on each desired port.

```
(config) # interface ethernet <PORT> qos trust L3
```

2. Verify the trust mode is configured properly with this show command.

```
(config) # show qos interface ethernet <PORT>
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show qos interface ethernet 1/1

Eth1/1:
  Trust mode             : L3
  Default switch-priority: 0
  Default PCP            : 0
  Default DEI            : 0
  PCP,DEI rewrite        : disabled
  IP PCP;DEI rewrite     : enable
  DSCP rewrite           : disabled
```

Configure PFC

1. Disable Ethernet Flow Control (Global Pause) system wide.

```
(config) # interface ethernet 1/1-1/32 flowcontrol send off force
(config) # interface ethernet 1/1-1/32 flowcontrol receive off force
```

2. Verify that Global Pause is disabled with this show command.

```
(config) # show interfaces ethernet <PORT>
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show interfaces ethernet 1/1

Eth1/1:
  Admin state                       : Enabled
  Operational state                 : Down
  Last change in operational status: Never
  Boot delay time                   : 0 sec
  Description                       : N/A
  Mac address                       : b8:59:9f:69:99:98
  MTU                               : 9216 bytes (Maximum packet size 9238 bytes)
  Fec                               : auto
  Operational Fec                   : N/A
  Flow-control                      : receive off send off
  Supported speeds                  : 1G 10G 25G 40G 50G 56G 100G
  Advertised speeds                 : 100G
  Actual speed                      : Unknown
  Auto-negotiation                  : Enabled
  Width reduction mode              : Unknown
  Switchport mode                   : access
  MAC learning mode                 : Enabled
  Forwarding mode                   : inherited cut-through
```

3. Enable PFC system wide.

```
(config) # dcb priority-flow-control enable
```

4. Verify that PFC have been enabled with this show command.

```
(config) # show dcb priority-flow-control
```

Note: Verification that PFC has been enabled can be done on a per port basis with this show command.

```
(config) # show dcb priority-flow-control interface ethernet <PORT>
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show dcb priority-flow-control interface ethernet 1/1
PFC: enabled
Priority Enabled List:
Priority Disabled List: 0 1 2 3 4 5 6 7


-----------------------------------------------
Interface         PFC admin         PFC oper
-----------------------------------------------
Eth1/1            Auto              Disabled
```

5. Create RoCE traffic pool with desired priority to setup lossless queues.

```
(config) # traffic pool RoCE type lossless
(config) # traffic pool RoCE memory percent 50.00
(config) # traffic pool RoCE map switch-priority <ROCE _ PRI>
```

6. Verify the traffic pool is configured properly with this show command.

```
(config) # show traffic pool
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show traffic pool
------------------------------------------------------------------------------------------------
Traffic                    Type      Memory    Switch         Memory actual    Usage     Max Usage
Pool                                   [%]       Priorities
------------------------------------------------------------------------------------------------
lossless-default (RO)     lossless  auto                      0                0         0
roce                       lossless  50.00     3               5.1M             0         0
lossy-default              lossy     auto      0, 1, 2, 4,     5.1M             0         0
                                               5, 6, 7
roce-reserved              lossy     auto                       0                0         0

Exception list:
N/A
```

7. Enable PFC on interface.

```
(config) # interface ethernet <PORT> dcb priority-flow-control mode on force
```

8. Persist changes.

```
(config)# write memory
```

Summary of CLI Commands

```
interface ethernet <PORT> qos trust L3
interface ethernet 1/1-1/32 flowcontrol send off force
interface ethernet 1/1-1/32 flowcontrol receive off force
dcb priority-flow-control enable
traffic pool RoCE type lossless
traffic pool RoCE memory percent 50.00
traffic pool RoCE map switch-priority <ROCE _ PRI>
interface ethernet <PORT> dcb priority-flow-control mode on force

write memory
```

Configure ECN

Note: ECN Configuration is not required for the Data24.

1. Enable ECN with desired priority on each desired port.

```
(config) # interface ethernet <PORT> traffic-class <ROCE _ PRI> congestion-control ecn minimum-absolute 150 maximum-absolute 1500
```

2. Verify that ECN has been enabled with this show command.

```
(config) # show interfaces ethernet <PORT> congestion-control
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show interfaces ethernet 1/1 congestion-control


Interface ethernet: 1/1:
  ECN marked packets: 0


  TC-0:
    Mode: none
  TC-1:
    Mode: none
  TC-2:
    Mode: none
  TC-3:
    Mode                : ECN
    Threshold mode      : absolute
    Minimum threshold   : 150 KB
    Maximum threshold   : 1500 KB
    RED dropped packets: 0
  TC-4:
    Mode: none
  TC-5:
    Mode: none
  TC-6:
    Mode: none
  TC-7:
    Mode: none
```

3. Configure the desired CNP priority to be ETS Strict on each desired port.

```
(config) # interface ethernet <PORT> traffic-class <CNP _ PRI> dcb ets strict
```

4. Verify that the CNP Priority has been set to ETS Strict with this show command.

```
(config) # show dcb ets interface ethernet <PORT>
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show dcb ets interface ethernet 1/1

Eth1/1:
  Interface Bandwidth Shape [Mbps]: N/A
  Multicast unaware mapping       : disabled

  Flags:
    S.Mode: Scheduling Mode [Strict/WRR]
    D     : -
    W     : Weight
    Bw.Sh : Bandwidth Shaper
    Bw.Gr : Bandwidth Guaranteed

  ETS per TC:
    ------------------------------------------------------------------
    TC   S.Mode      W      W(%)     BW Sh.(Mbps)     BW Gr.(Mbps)
    ------------------------------------------------------------------
    0    WRR         12     14       N/A              0
    1    WRR         13     15       N/A              0
    2    WRR         12     14       N/A              0
    3    WRR         13     15       N/A              0
    4    WRR         12     14       N/A              0
    5    WRR         13     14       N/A              0
    6    Strict      0      0        N/A              0
    7    WRR         13     14       N/A              0
```

5. Persist changes.

```
(config)# write memory
```

Summary of CLI Commands

```
interface ethernet <PORT> traffic-class <ROCE _ PRI> congestion-control ecn minimum-absolute 150 maximum-absolute 1500
interface ethernet <PORT> traffic-class <CNP _ PRI> dcb ets strict


write memory
```

Show Pertinent Switch Counters

1. Show PFC counters for a specific interface.

```
(config) # show interfaces ethernet <PORT> counters pfc prio all
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show interfaces ethernet 1/1 counters pfc prio all

Eth1/1:
  PFC 0:
    Rx:
      0                     pause packets
      0                     pause duration
    Tx:
      0                     pause packets
      0                     pause duration
  PFC 1:
    Rx:
      0                     pause packets
      0                     pause duration
    Tx:
      0                     pause packets
      0                     pause duration
```

2. Show Traffic Class queue depth and queue drops on a specific interface.

```
(config) # show interfaces ethernet <PORT> counters tc all
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show interfaces ethernet 1/1 counters tc all

 Eth1/1:
   TC 0:
     0                     packets
     0                     bytes
     0                     unicast queue depth
     0                     multicast queue depth
     0                     unicast no buffer discard
     0                     WRED discard

   TC 1:
     0                     packets
     0                     bytes
     0                     unicast queue depth
     0                     multicast queue depth
     0                     unicast no buffer discard
     0                     WRED discard

   TC 2:
     0                     packets
     0                     bytes
     0                     unicast queue depth
     0                     multicast queue depth
     0                     unicast no buffer discard
     0                     WRED discard
```

3. Show ECN counters on a specific interface.

```
(config) # show interfaces ethernet <PORT> congestion-control
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show interfaces ethernet 1/1 congestion-control

Interface ethernet: 1/1:
  ECN marked packets: 0

  TC-0:
    Mode: none
  TC-1:
    Mode: none
  TC-2:
    Mode: none
  TC-3:
    Mode                : ECN
    Threshold mode      : absolute
    Minimum threshold   : 150 KB
    Maximum threshold   : 1500 KB
    RED dropped packets: 0
  TC-4:
    Mode: none
  TC-5:
    Mode: none
  TC-6:
    Mode: none
  TC-7:
    Mode: none
```

4. Show interface counters.

```
(config) # show interface ethernet <PORT> counters
```

Example Output:

```
pfemlsn2100-4 [standalone: master] (config) # show interfaces ethernet 1/1 counters

Eth1/1:
  Rx:
    0                   packets
    0                   unicast packets
    0                   multicast packets
    0                   broadcast packets
    0                   bytes
    0                   packets of 64 bytes
    0                   packets of 65-127 bytes
    0                   packets of 128-255 bytes
    0                   packets of 256-511 bytes
    0                   packets of 512-1023 bytes
    0                   packets of 1024-1518 bytes
    0                   packets Jumbo
    0                   discard packets
    0                   error packets
    0                   fcs errors
    0                   undersize packets
    0                   oversize packets
    0                   pause packets
    0                   unknown control opcode
    0                   symbol errors
    0                   discard packets by storm control

  Tx:
    0                   packets
    0                   unicast packets
    0                   multicast packets
    0                   broadcast packets
    0                   bytes
    0                   discard packets
    0                   error packets
    0                   hoq discard packets
    0                   pause packets
    0                   pause duration
    0                   ECN marked packets
```

5. Clear port and PFC counters.

```
(config) # show interface ethernet <PORT> counters
```


5601 Great Oaks Parkway
San Jose, CA 95119, USA
www.westerndigital.com