Dell EMC Networking – Deploying Data Center Bridging

Guide for deploying DCB in a typical Data Center

Abstract

A short Data Center Bridging (DCB) cheat sheet configuration with Dell EMC Networking OS10

September 2018

Revisions

Date	Description
September 2018	Initial release

Acknowledgements

This paper was produced by the following members of the Dell EMC networking technical marketing engineering team:

Author: Mario Chow

Support:

Other: N/A

Table of contents

Executive summary

The data center has always been a critical component of the IT infrastructure of enterprise organizations. Ethernet has been, and will most likely continue to be, the dominant network protocol in data centers for server-to-server communication.

Data centers are facing increasing performance demands and are also embracing convergence to address consolidation. These factors require traffic patterns that ensure lossless delivery of packets in the network with no packet loss. However, Ethernet was originally created to be a best-effort network protocol. To meet the new network quality demands, a set of enhancements to the existing Ethernet protocol stack was created.

With the introduction of Data Center Bridging (DCB), key Ethernet protocol enhancements such as Priority-Based Flow Control, Enhanced Transmission Selection, and Congestion Notification were introduced.

The implementation of DCB delivers a true converged environment, allowing for specific treatment of traffic flows by tagging each traffic pattern with a unique Class of Service (0-7) or Differentiated Services Code Point (0-63).

This technical white paper provides information about configuring DCB in a typical leaf and spine topology.

1 Data Center Bridging

Data-Center Bridging (DCB) is an extension to the Ethernet protocol that makes dedicated traffic flows possible in a converged network scenario. DCB distinguishes traffic flows by tagging the traffic with a specific value (0-7) for Class of Service (CoS) or (0-63) for Differentiated Services Code Point (DSCP).

For selected traffics, DCB aims to

- Eliminate loss due to queue overflow; also known as **lossless Ethernet**
- Provide the option to allocate bandwidth on links

To some extent, DCB enables the treatment of different priorities as if they were different pipes. To meet these goals, new standards are being developed, or have already been developed, that either extend the existing set of Ethernet protocols or emulate the connectivity offered by Ethernet protocols.

There are four distinct components that make up DCB:

- Congestion Notification (802.1Qau)
- Priority Flow Control (802.1Qbb)
- Enhanced Transmission Selection (802.1Qaz)
- DCBx

1.1 Congestion Notification (802.1Qau)

The first enhancement is a standard that specifies protocols, procedures, and managed objects that support congestion management of long-lived data flows within network domains of limited bandwidth delay product. Using this enhancement, the bridges can signal congestion information to the end stations that can limit the transmission rate to avoid frame loss.

This mechanism enables support for higher layer protocols that are highly sensitive to loss or latency.

Data center networks and backplane fabrics employ applications that depend on the delivery of data packets with a lower latency. These applications also require a much lower probability of packet loss as compared to typical VLAN bridged networks. This enhancement supports the use of a single bridged LAN for these applications as well as traditional LAN applications.

Although not widely used, this component of DCB is available for use when required.

1.2 Priority Flow Control (802.1Qbb)

The second enhancement is intended to eliminate frame loss due to congestion. This enhancement is achieved by a mechanism like the IEEE 802.3x PAUSE but operates on individual priorities. This mechanism, along with other Data Center Bridging technologies, enables support for higher layer protocols that are highly sensitive to loss, without affecting the operation of traditional LAN protocols that utilize other priorities. Priority Flow Control (PFC) also complements Congestion Notification in DCB networks. Operation of PFC is limited to a domain that is controlled by a DCB control protocol, which controls the application of Priority-Based Flow Control, Enhanced Transmission Selection, and Congestion Notification.

This amendment enables multiple data-center networks, including those serving loss-sensitive protocols such as inter-processor communication, storage, and so on, to be converged onto an IEEE 802 network.

1.3 Enhanced Transmission Selection (802.1Qaz)

The third enhancement specifies enhancement of transmission selection to support allocation of bandwidth amongst traffic classes. When the offered load in a traffic class does not use its allocated bandwidth, enhanced transmission selection allows other traffic classes to use the available bandwidth. The-bandwidth allocation priorities coexist with strict priorities.

Networks prioritize traffic to provide different service characteristics to traffic classes. It is desirable to be able to share bandwidth between priorities carrying bursts of high offered-loads rather than servicing them with strict priority, while allowing strict priority for time-sensitive and management traffic that requires minimum latency. Also, when traffic at one priority level does not use its allocation, it is desirable to allow other priorities to use that bandwidth.

1.4 DCBx

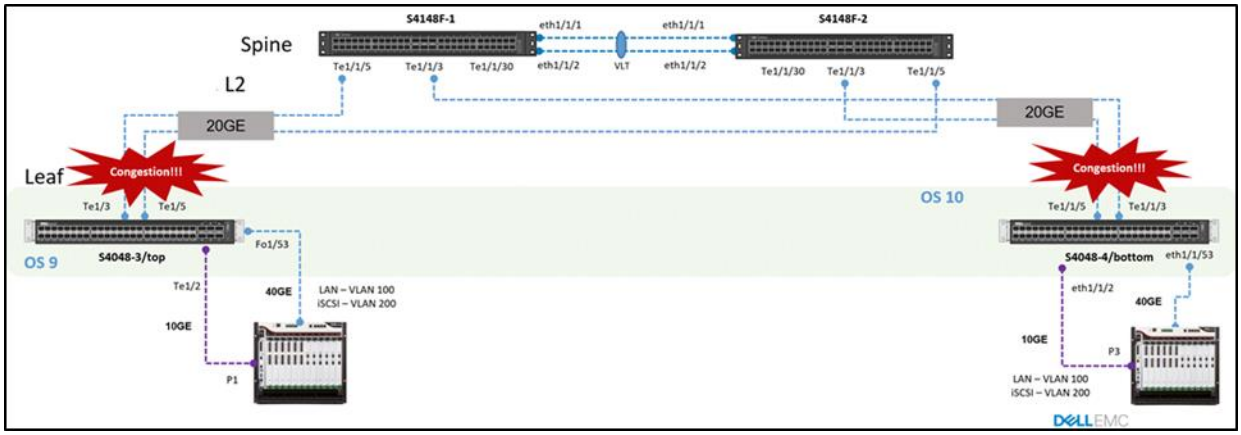
The fourth enhancement is an exchange protocol. This protocol is a discovery and capability-exchange protocol that is used for conveying capabilities and configuration of the features between neighbors to ensure consistent configuration across the network. This protocol is expected to leverage functionality that IEEE 802.1AB (LLDP) provides.

2 Test setup

Figure 1 shows the setup used to test DCB, PFC, and ETS. The following list defines the setup:

- Traffic is bi-directional.
- Traffic is between two operating systems – OS9 {9.13(0.0) & OS10 (10.4.0E(X2)}.
- Priority Flow Control is “ON” with iSCSI traffic and “OFF” with LAN traffic.
- ETS is set to 1% for LAN traffic, 99% for iSCSI traffic.
- DCB is on at both ends.
- Ixia traffic generator transmitting two types of traffic:
  - LAN CoS 0 @ 10GE
  - iSCSI CoS 4 @ 40GE

Figure 1 Dell EMC Networking DCB Test Setup



A Appendix

A.1 Configuration details

This section provides the details about the configurations applicable on all the switches.

Table 1 Component table example

Component	Description
Operating system	Dell OS 9 and 10
Application	IXExplorer 6.70 - IXIA
Switch	S4048-ON, S4128F-ON

OS 9 based switch configuration – Switch on the left side labeled “S4048-ON-3”

1. Trust the dot1p Class of service values to queue assignment configured by the operating system.
   a. **Switch(config)#** service-class dynamic dot1p

2. Create the dcb map configuration. In this configuration pfc and ets are enabled.
   a. **Switch(config)#** dcb-map <name>
   b. **Switch(config-dcbmap-name)#** priority-group 0 bandwidth 1 pfc off
      LAN traffic is given 1% of its 10GE bandwidth and it will not generate PFC grames towards the sender (Ixia).
   c. **Switch(config-dcbmap-name)#** priority-group 1 bandwidth 99 pfc on
      iSCSI traffic is given 99% of its 40GE bandwidth and it will generate PFC frames towards the sender (Ixia).
   d. **Switch(config-dcbmap-name)#** priority-pgid 0 0 0 0 1 0 0 0
      Assign the priority group to the corresponding class of service. In this case, CoS 4 is configured with priority group 1.

3. Assign the dcb map to the all the interfaces receiving and transmitting these two types of traffic.
   a. **Switch(config)#** inter Te1/3
   b. **Switch(config-if-te-1/3)#** dcb-map <name>
   c. **Switch(config-if-te-1/3)#** end

Remember to configure the switchports connected to the ixia traffic generator as spanning-tree edge port. This configuration tells the switch that edge ports are to transition into forwarding mode straight away and not participate into any spanning-tree negotiations.

OS10 based switch configuration – All other switches besides "S4048-ON-3"

1. Turn on DCBx on the switch.
    a. **Switch#** conf t
    b. **Switch(config#** dcbx en
2. Turn on ETS.
    a. **Switch(config)#** system qos
    b. **Switch(config-sys-qos)#** ets mode on
    c. **Switch(config-sys-qos)#** end
    d. Switch#
3. Configure the qos-map to match the queues to the CoS values
    a. **Switch#** conf t
    b. **Switch(config)#** qos-map traffic-class qmap
    c. **Switch(config-qos-map)#** queue 0 qos-group 0-3,5-7
    d. **Switch(config-qos-map)#** queue 4 qos-group 4

4. Start configuring the class maps to be used with all respective policy maps – first one "queueing".
    4a. **Queueing –** Refers to the queues configured in step 3.
    a. Switch# conf t
    b. Switch(config)# class-map type queueing q0
    c. Switch(config-cmap-queuing)# match queue 0
    d. Switch(config-cmap-queuing)# exit
    e. Switch(config)# class-map type queuing q4
    f. Switch(config-cmap-queuing)# match queue 4
    g. Switch(config-cmap-queuing)# exit
    h. Switch(config)#

    4b**. Network-qos** – This class map is used by the service policy input and applied as ingress traffic. (See "Interface configuration" in step 6)
    a. Switch# conf t
    b. Switch(config)# class-map type network-qos temp
    c. Switch(config-cmap-nqos)# match qos-group 4

5. Start configuring the relevant policy maps that will use the class-maps defined in steps 3-4 and self-generated class "class-trust".
    a. Switch# conf t
    b. Switch(config)# policy-map type qos trust_dot1p
    c. Switch(config-pmap-qos)# class class-trust
    d. Switch(config-pmap-c-qos)# trust dot1p

    5a**. ETS policy map –** Assigning bandwidth percentage to each queue.
    a. Switch# conf t
    b. Switch(config)# policy-map type queuing q1
    c. Switch(config-pmap-queuing)# class q0
    d. Switch(config-pmap-c-que)# bandwidth percent 1
    e. Switch(config-pmap-c-que)# exit
    f. Switch(config-pmap-queuing)# class q4
    g. Switch(config-pmap-c-que)# bandwidth percent 99

    5b. **PFC policy map –** Policy map turning pfc "ON" on CoS 4 or iSCSI flows.

a. **Switch#** conf t
b. **Switch(config)#** policy-map type network-qos temp
c. **Switch(config-pmap-network-qos)#** class temp
d. **Switch(config-pmap-c-nqos)#** pause
e. **Switch(config-pmap-c-nqos)#** pfc-cos 4

6. **Applying policy maps to interfaces**
   a. **Switch#** conf t
   b. **Switch(config#** inte ethe1/1/53 – interface part of iSCSI flows.
   c. **Switch(config-if-eth1/1/53)#** switchport mode trunk
   d. **Switch(config-if-eth1/1/53)#** switchport trunk allowed vlan 100,200 – LAN and iSCSI traffic allowed in this interface.
   e. **Switch(config-if-eth1/1/53)#** service-policy input type network-qos temp
   Service policy that turns on PFC on CoS = 4. Policy map in step 5b.
   f. **Switch(config-if-eth1/1/53)#** service-policy input type qos trust_dot1p
   QoS policy that trusts the system's dot1p settings
   g. **Switch(config-if-eth1/1/53)#** service-policy output type queuing q1
   Egress service policy applying ETS on output queues defined in step 5a.
   h. **Switch(config-if-eth1/1/53)#** ets mode on
   Turns "ON" ETS
   i. **Switch(config-if-eth1/1/53)#** qos-map traffic-class qmap
   Assigns the system queues to the proper class of service value on ingress
   j. **Switch(config-if-eth1/1/53)#** priority-flow-control mode on
   PFC is turned "ON" at the interface
   k. **Switch(config-if-eth1/1/53)#** end
   l. **Switch#**

Note: Apply configurations in step 6 on all switches (OS10) and interfaces that are participating in the transmission and reception of LAN and iSCSI traffic.

A.2 Appendix

The two following configurations apply to the switches in Figure 1.

S4048-ON-3 (OS9)

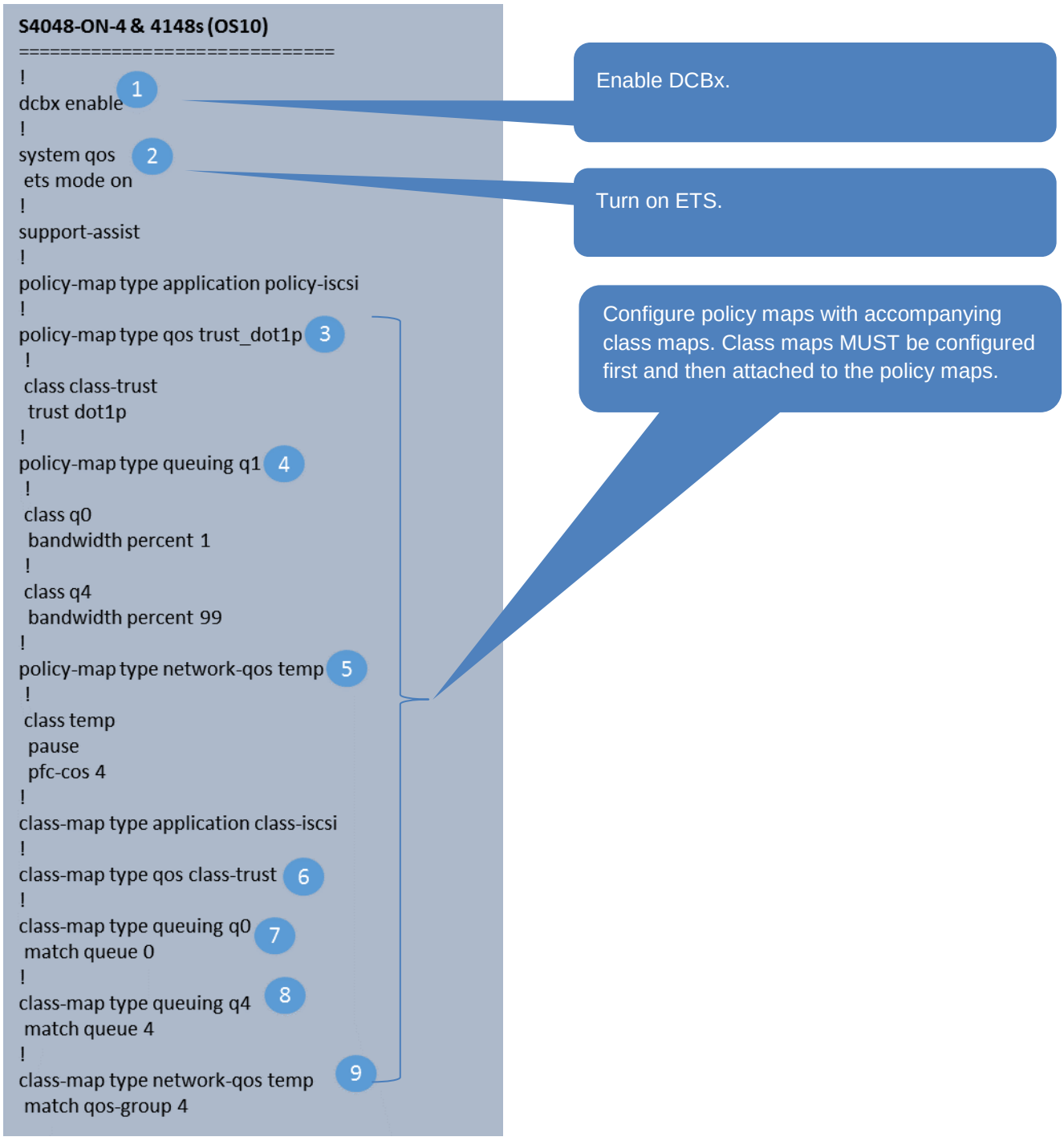
```
S4048-ON-3 (OS9)
============================
service-class dynamic dot1p 1
!
dcb enable pfc-queues 4 2
!
dcb-map training 3
 priority-group 0 bandwidth 1 pfc on
 priority-group 1 bandwidth 99 pfc on
 priority-pgid 0 0 0 0 1 0 0 0
!
interface TenGigabitEthernet 1/3 4
 description Po10_channel_member_links
 no ip address
 mtu 9216
 dcb-map training
!
 port-channel-protocol LACP
  port-channel 10 mode active
!
 protocol lldp
  advertise management-tlv system-description system-name
  advertise interface-port-desc
 shutdown
mkt-s4048-3#sh run int te1/5
!
interface TenGigabitEthernet 1/5
 description Po10_channel_member_links
 no ip address
 mtu 9216
 dcb-map training
!
 port-channel-protocol LACP
  port-channel 10 mode active
```

1 Trust system wide 802.1p settings mappings to CoS values.

2 Enable all queues as Priority Flow Control-enabled.

3 Create the DCB map specifying PFC and ETS. Then assign the priority group to the proper class of service.

4 Assign the DCB map to all active interfaces.


S4048-ON-4 & 4148s (OS10)
==========
!
dcbx enable 1
!
system qos 2
ets mode on
!
support-assist
!
policy-map type application policy-iscsi
!
policy-map type qos trust_dot1p 3
!
class class-trust
trust dot1p
!
policy-map type queuing q1 4
!
class q0
bandwidth percent 1
!
class q4
bandwidth percent 99
!
policy-map type network-qos temp 5
!
class temp
pause
pfc-cos 4
!
class-map type application class-iscsi
!
class-map type qos class-trust 6
!
class-map type queuing q0 7
match queue 0
!
class-map type queuing q4 8
match queue 4
!
class-map type network-qos temp 9
match qos-group 4
Enable DCBx.
Turn on ETS.
Configure policy maps with accompanying class maps. Class maps MUST be configured first and then attached to the policy maps.