

# AI Network Fabric Deployment Guide

The following section gives a comprehensive view of a RoCEv2 topology, design, configurations, and key takeaways from a successful proof of concept.

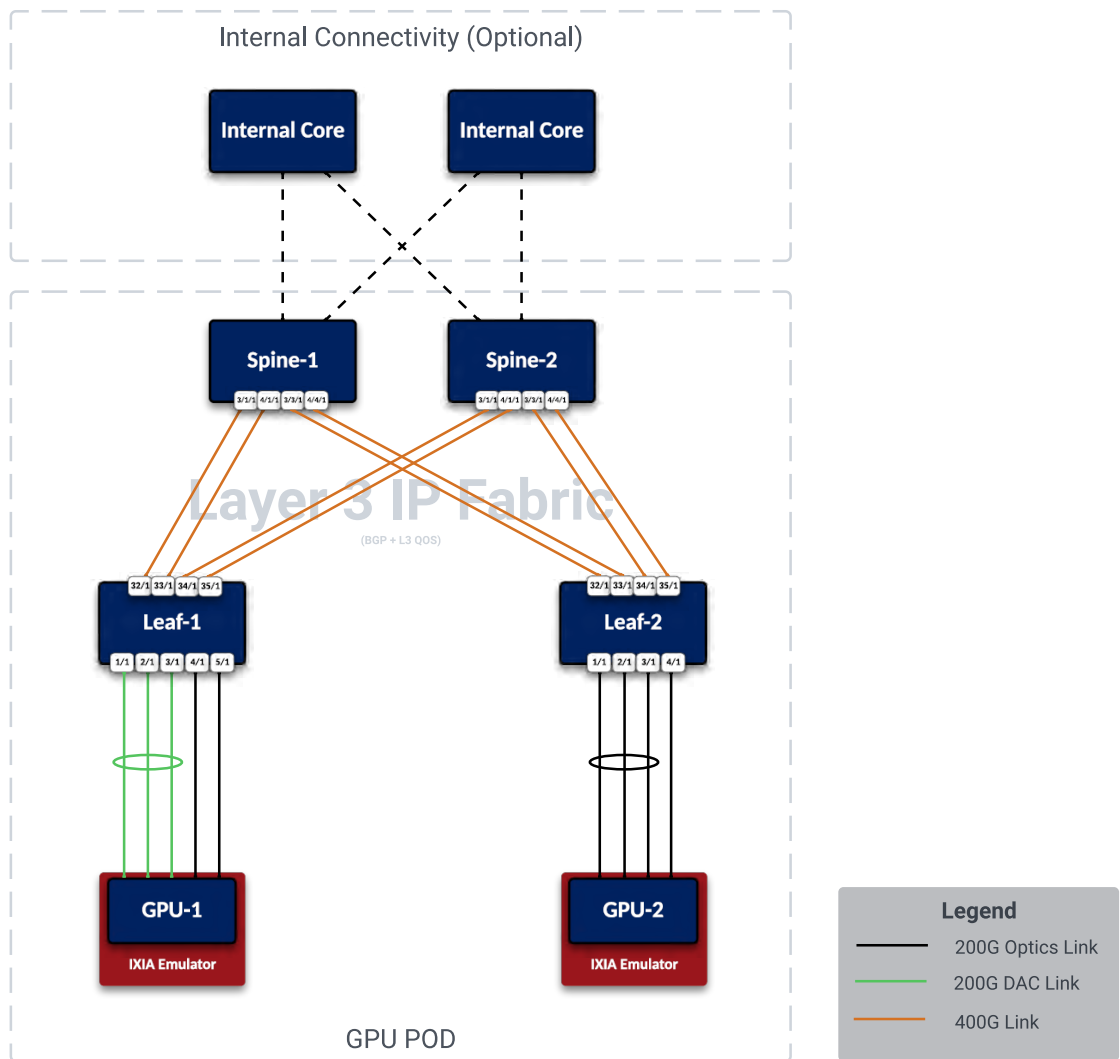


Figure 1: Physical Topology

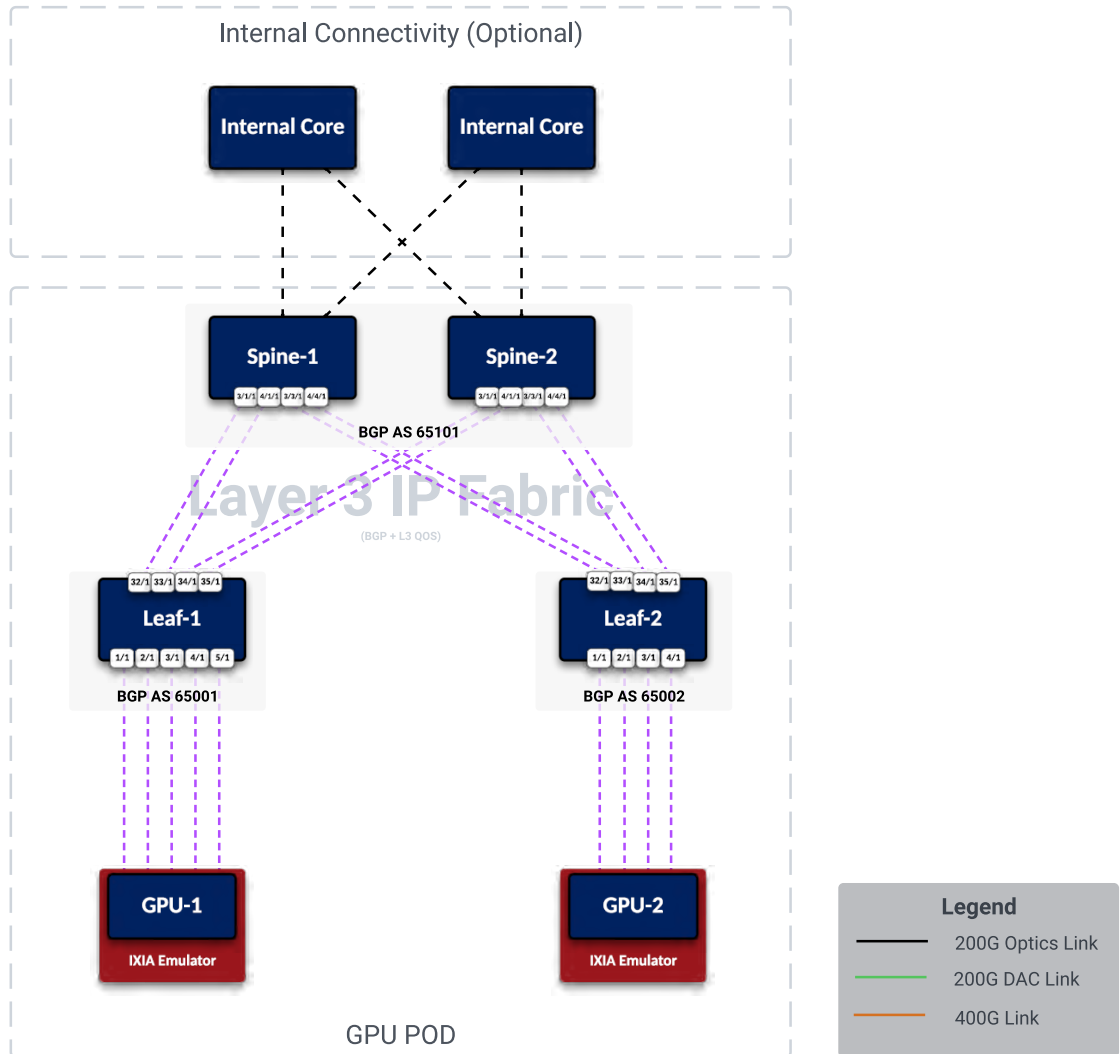


Figure 2: BGP Logical Topology

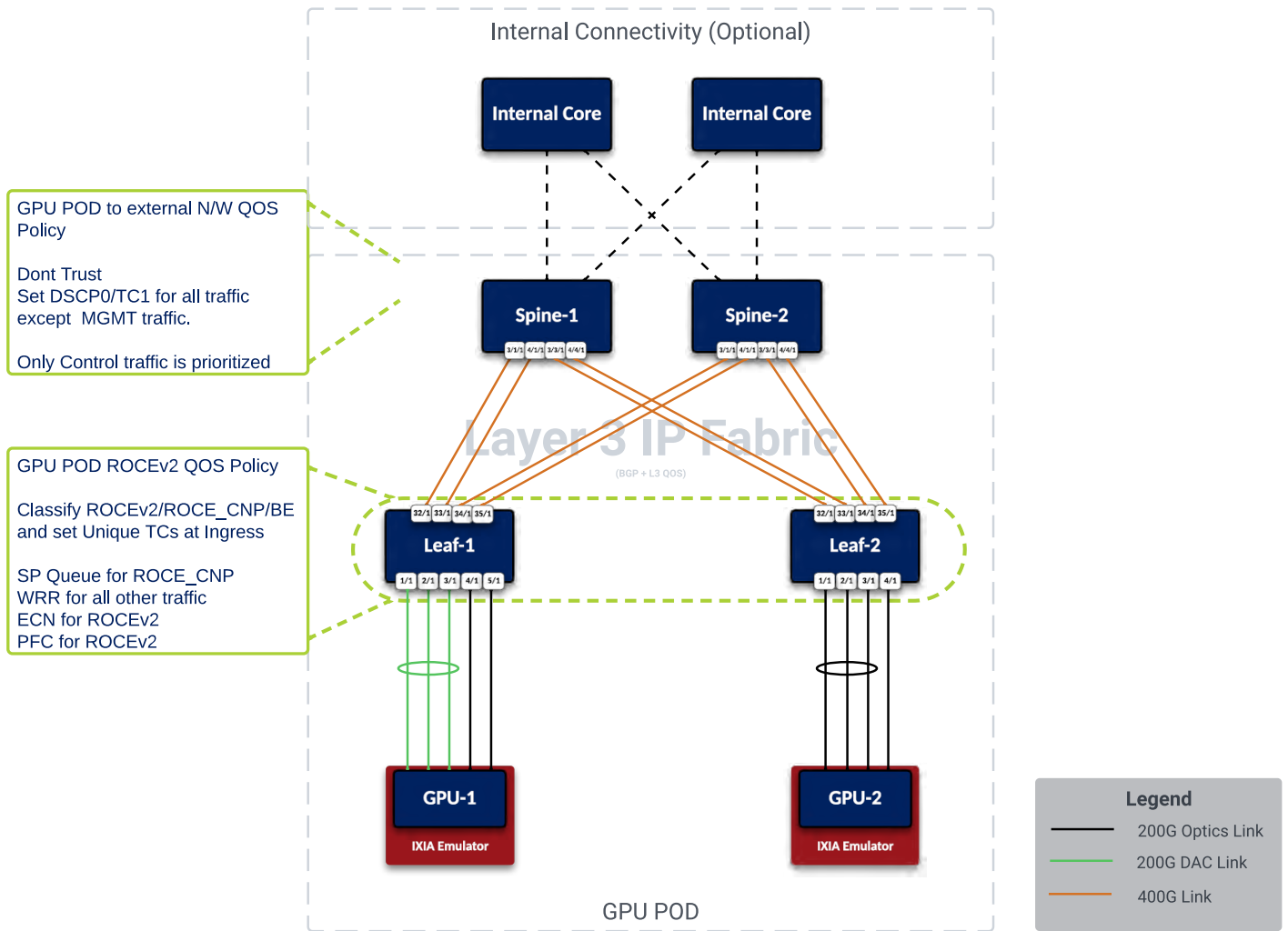


Figure 3: QoS Logical Topology

**ROCEv2 QoS Policy plan on GPU POD Leafs**

Traffic Types	Marking Method	Input Markings	Queueing Behavior	ECN Enabled	PFC Enabled	TC/Queue
Best Effort	Trust Application	CS0, CS1, CS4, CS5	WRR 5 Percent	✗	✗	1
ROCEv2	Trust Application	CS3	WRR	✓	✓	3
ROCE CNP	Trust Application	CS2	SP	✗	✗	6
Control Traffic	Trust Application	CS7	SP	✗	✗	7

**ECN Markings Will Vary:** ECN was enabled and configured with the following values (256k/512k), but these values will change dependent on deployment

### ROCEv2 QoS Policy plan on GPU POD Spines

Traffic Types	Marking Method	Input Markings	Queueing Behavior	ECN Enabled	PFC Enabled	TC/Queue
Control Traffic	Trust Application	CS7	SP	✘	✘	7
ROCE CNP	Trust Application	CS2	SP	✘	✘	6
ROCEv2	Trust Application	CS3	WRR	✔	✔	3
Best Effort	Trust Application	CS0, CS1, CS4, CS5	WRR 5 Percent	✘	✘	1

**ECN Markings Will Vary:** ECN was enabled and configured with the following values (256k/512k), but these values will change dependent on deployment

### GPU Spine link to Internal Core QoS Plan

Optional QoS Plan to Internal Core							
Traffic Types	Marking Method	Input Markings	Marking Action	Queueing Behavior	ENC Config	PFC Enabled	Suggested TC
Best Effort	Don't Trust/ Down Mark	CS0, CS2, CS3	Mark CS0	WRR		✘	1
Control Traffic	Trust Application	CS7		SP		✘	7

### RoCEv2 Reference Configs

#### Global QoS Mapping Config

Sample QoS mapping config for a GPU POD Leaf and Spine.

You will notice we do not need to configure a qos map for ROCEv2 traffic, because by default EOS will map CS3 to traffic-class 3 (TC3).

Leaf

```
qos map DSCP 8 9 10 11 12 13 14 15 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 to
traffic-class 1
qos map DSCP 16 17 18 19 20 21 22 23 to traffic-class 6
```

Spine

```
qos map DSCP 8 9 10 11 12 13 14 15 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 to
traffic-class 1
qos map DSCP 16 17 18 19 20 21 22 23 to traffic-class 6
```

#### QoS Profile Config

Sample QoS mapping config for a GPU POD Leaf and Spine.

Leaf

```
qos profile ai-scheduler
  priority-flow-control on
  priority-flow-control priority 3 no-drop
  !
  uc-tx-queue 1
    no priority
    bandwidth percent 5
```

```
!  
uc-tx-queue 3  
    no priority  
    bandwidth percent 95  
    random-detect ecn minimum-threshold 256 kbytes maximum-threshold 512 kbytes  
max-mark-probability 100 weight 0
```

Spine

### Interface Config

```
qos profile ai-scheduler  
    priority-flow-control on  
    priority-flow-control priority 3 no-drop  
!  
tx-queue 1  
    no priority  
    bandwidth percent 5  
!  
tx-queue 3  
    no priority  
    bandwidth percent 95  
    random-detect ecn minimum-threshold 512 kbytes maximum-threshold 768 kbytes  
max-mark-probability 100  
!
```

Sample interface configuration for the GPU POD leafs and spines referenced in the [topology](#).

```
hardware counter feature ecn out
```

### ECN Counter Platform Support

Global config is needed on some Arista platforms to enable ECN counter feature. See the [TOI for more information](#).

Leaf-1

```
!  
hardware counter feature ecn out  
!  
interface Port-Channel20  
    description GPU1-CHANNEL  
    mtu 9214  
    no switchport  
    ip address 11.20.1.1/30  
!  
interface Ethernet1/1  
    description GPU1-Port1  
    mtu 9214  
    speed 200g-4  
    no switchport  
    channel-group 20 mode active  
    service-profile ai-scheduler
```

```
!
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet2/1
    description GPU1-Port2
    mtu 9214
    speed 200g-4
    no switchport
    channel-group 20 mode active
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet3/1
    description GPU1-Port3
    mtu 9214
    speed 200g-4
    no switchport
    channel-group 20 mode active
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet4/1
    description GPU1-Port4
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.20.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet5/1
    description GPU1-Port5
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.20.3.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet32/1
    description SPINE1-Et3/1/1
```

```
mtu 9214
speed 400g-8
no switchport
ip address 11.23.1.1/30
service-profile ai-scheduler
!
uc-tx-queue 3
    random-detect ecn count
!
interface Ethernet33/1
    description SPINE1-Et4/1/1
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.23.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet34/1
    description SPINE2-Et3/2/1
    shutdown
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.24.1.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet35/1
    description SPINE2-Et4/2/1
    shutdown
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.24.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
```

## Leaf-2

```
!  
hardware counter feature ecn out  
!  
interface Port-Channel10  
    description GPU2-CHANNEL  
    mtu 9214  
    no switchport  
    ip address 11.10.1.1/30  
!  
interface Ethernet1/1  
    description GPU1-Port1  
    mtu 9214  
    speed 200g-4  
    no switchport  
    channel-group 10 mode active  
    service-profile ai-scheduler  
    !  
    uc-tx-queue 3  
        random-detect ecn count  
!  
interface Ethernet2/1  
    description GPU1-Port2  
    mtu 9214  
    speed 200g-4  
    no switchport  
    channel-group 10 mode active  
    service-profile ai-scheduler  
    !  
    uc-tx-queue 3  
        random-detect ecn count  
!  
interface Ethernet3/1  
    description GPU1-Port3  
    mtu 9214  
    speed 200g-4  
    no switchport  
    channel-group 10 mode active  
    service-profile ai-scheduler  
    !  
    uc-tx-queue 3  
        random-detect ecn count  
!  
interface Ethernet4/1  
    description GPU1-Port4  
    mtu 9214  
    speed 200g-4  
    no switchport  
    ip address 11.10.2.1/30
```



```
service-profile ai-scheduler
!
uc-tx-queue 3
    random-detect ecn count
!
interface Ethernet5/1
    description GPU1-Port5
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.10.3.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet32/1
    description SPINE1-Et3/1/1
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.13.1.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet33/1
    description SPINE1-Et4/1/1
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.13.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet34/1
    description SPINE2-Et3/2/1
    shutdown
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.14.1.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
```

```
interface Ethernet35/1
  description SPINE2-Et4/2/1
  shutdown
  mtu 9214
  speed 400g-8
  no switchport
  ip address 11.14.2.1/30
  service-profile ai-scheduler
  !
  uc-tx-queue 3
    random-detect ecn count
  !
```

#### Spine-1

```
!
hardware counter feature ecn out
!
interface Ethernet3/1/1
  description LEAF1-Et32/1
  mtu 9214
  no switchport
  ip address 11.13.1.2/30
  service-profile ai-scheduler
  !
  tx-queue 3
    random-detect ecn count
!
interface Ethernet3/3/1
  description LEAF2-Et32/1
  mtu 9214
  no switchport
  ip address 11.23.1.2/30
  service-profile ai-scheduler
  !
  tx-queue 3
    random-detect ecn count
!
interface Ethernet4/1/1
  description LEAF1-Et33/1
  mtu 9214
  no switchport
  ip address 11.13.2.2/30
  service-profile ai-scheduler
  !
  tx-queue 3
    random-detect ecn count
!
interface Ethernet4/4/1
  description LEAF2-Et33/1
  shutdown
```

```
mtu 9214
no switchport
ip address 11.23.2.2/30
service-profile ai-scheduler
!
tx-queue 3
    random-detect ecn count
!
```

## Spine-2

```
!
hardware counter feature ecn out
!
interface Ethernet3/1/1
    description LEAF1-Et34/1
    mtu 9214
    no switchport
    ip address 11.14.1.2/30
    service-profile ai-scheduler
    !
    tx-queue 3
        random-detect ecn count
!
interface Ethernet3/3/1
    description LEAF2-Et34/1
    mtu 9214
    no switchport
    ip address 11.24.1.2/30
    service-profile ai-scheduler
    !
    tx-queue 3
        random-detect ecn count
!
interface Ethernet4/1/1
    description LEAF1-Et35/1
    mtu 9214
    no switchport
    ip address 11.14.2.2/30
    service-profile ai-scheduler
    !
    tx-queue 3
        random-detect ecn count
!
interface Ethernet4/4/1
    description LEAF2-Et35/1
    shutdown
    mtu 9214
    no switchport
    ip address 11.24.2.2/30
```

### Priority Flow Control watchdog

Priority Flow Control (PFC) Watchdog feature monitors interfaces for priority-flow-control pause storm. If such a storm is detected on no-drop enabled priorities, it takes actions such as:

- Disable reacting to received pause frames
- Stop sending packets to these interfaces and drop any incoming packets from these interfaces
- Error Disable the port

PFC pause storm reception is usually an indication of a misbehaving node downstream, and propagating this congestion upstream is not desired.

```
priority-flow-control pause watchdog default timeout 0.20
priority-flow-control pause watchdog default recovery-time 0.20
priority-flow-control pause watchdog default polling-interval 0.100
priority-flow-control pause watchdog action drop
```

### Operational Bandwidth Calculation when BRR >100%

- When bandwidth percentages are allotted to queues in RR priority group (using bandwidth percent command), the output of `show qos interface <interface>` will show the bandwidth allocation in terms of percentage but the operational values may not sum up to 100 in all cases since, at the hardware level, weight is purely a natural number and not assigned as a percentage. The effective weight percentage can be found using.
  - › Effective Percentage of Bandwidth allocated to queue = Weight of this queue / Sum of weight of all WRR queues **When total configured bandwidth exceeds 100%.**
- **Operational Values at CLI level**
  - › The queues that have a configured bandwidth, the operational value assumes the configured values scaled down by a factor of (Total Configured Bandwidth)/100
  - › For queues that have no configured value of bandwidth, operational bandwidth at the CLI level is considered invalid.
- If there is still some unprogrammed bandwidth left after steps 1 and 2, then the last queue with unconfigured bandwidth percentage is assigned that bandwidth at the CLI level. If there is no queue with unconfigured bandwidth percentage, the residual bandwidth is not assigned to any of the queue and in such cases the percentages will not add up to 100%.

Example for above explanation

```
qos profile ai-scheduler
  priority-flow-control on
  priority-flow-control priority 3 no-drop
  !
  uc-tx-queue 1
    no priority
    bandwidth percent 20
  !
  uc-tx-queue 3
    no priority
    bandwidth percent 95
    random-detect ecn minimum-threshold 256 kbytes maximum-threshold 512 kbytes
  max-mark-probability 100 weight 0
  !
```

At the CLI level the ScaleDownFactor=(TotalConfiguredBandwidth)/100=115(50+95)/100=1.15

Queue	Configured Bandwidth	Operational Bandwidth
uc-tx-queue 4	Unconfigured	1% (residual bandwidth given)
uc-tx-queue 1	20%	[ 20/1.15 ] = 17%
uc-tx-queue 3	95%	[ 95/1.15 ] = 82% (Rounded)
mc-ux-queue 0	Unconfigured	Invalid
mc-ux-queue 1	Unconfigured	Invalid
uc-tx-queue 0	Unconfigured	Invalid
uc-tx-queue 2	Unconfigured	Invalid
uc-tx-queue 5	Unconfigured	Invalid

### RoCEv2 Useful CLI

CLI Command	Description
<code>show class-map</code>	To verify the DSCP to TC mappings
<code>show qos profile</code>	
<code>show qos interfaces &lt;interface-name&gt;</code>	To verify the Queues & BRR Percent Allocation for each Queue
<code>show interfaces &lt;interface-name&gt; counters queue detail</code>	To check the Queue counters. Use Watch 1 diff to check real time counters
<code>show qos interfaces ethernet &lt;interface-name&gt; ecn</code>	To verify the ECN enabled Queue & the configured Minimum/Maximum ecn Thresholds
<code>show qos interfaces ethernet &lt;interface-name&gt; ecn counters queue</code>	To verify the ECN packets marked during congestion on ROCEv2 Queue. Use Watch 1 diff to check real time counters
<code>show priority-flow-control-status</code>	To verify the PFC enabled interfaces and the Queues/Verify PFC watchdog configuration
<code>show priority-flow-control counters</code>	To verify the PFC received/transmitted frames during congestion
<code>show priority-flow-control counters watchdog</code>	To verify the PFC watchdog counters per interface, during the event of a PFC Flood
<code>show monitor session</code>	To check Monitor session capturing packets on incoming interfaces and mirrored to CPU
<code>ssh admin@&lt;switch&gt; \"bash tcpdump -s 0 -Un -w - -i mirror0\" /Applications/Wireshark.app/Contents/MacOS/Wireshark -k -i -</code>	To show the packet capture from the Switch real time in Wireshark

### Key Takeaways

For enabling ECN counters, there is no config in the QoS profile config hierarchy. ECN counters need to be enabled on each interface. The recommendation for in-place QoS modification is to make configuration changes via config-session to keep the changes clean. Behavior when the total BRR allocation is > 100%, EOS does an Effective Percentage calculation, which could cause undesired allocation.

Matching on CS and AF values matches only the exact code points and not a range within that CS. Use the range option to match more than one value. Always have a class-default explicitly defined to capture all the unintended unmatched values and take them into a low priority Queue. By default, they will take the Queue as per the `DSCP -> tc qos` mapping in the system which might be undesirable.

## Full Configurations

The configurations below are intended to be leveraged in a physical topology, due to features that are not supported on cEOS-lab or vEOS-lab.

Leaf-1

```
!
no aaa root
!
username admin privilege 15 role network-admin secret
sha512 $6$eucN5ngreuExDgwS$xnD7T8jO..GBDX0DUlp.
hn.W7yW94xTjSanqgaQGBzPIhDAsyAl9N4oSChvOMvf07uVBFI4mKMxwdVEUVKgY/.
!
prompt %H.%D{%H:%M:%S}%P
!
service routing protocols model multi-agent
!
queue-monitor length
!
hostname Leaf-1
ip name-server vrf MGMT 8.8.8.8
!
qos profile ai-scheduler
  priority-flow-control on
  priority-flow-control priority 3 no-drop
  !
  uc-tx-queue 1
    no priority
    bandwidth percent 5
  !
  uc-tx-queue 3
    no priority
    bandwidth percent 95
    random-detect ecn minimum-threshold 256 kbytes maximum-threshold 512 kbytes
max-mark-probability 100 weight 0
!
spanning-tree mode mstp
!
system ll
  unsupported speed action error
  unsupported error-correction action error
!
interface Port-Channel10
  description GPU1-CHANNEL
  mtu 9214
  no switchport
  ip address 11.10.1.1/30
!
interface Ethernet1/1
  description GPU1-Port1
  mtu 9214
```

```
speed 200g-4
no switchport
channel-group 10 mode active
service-profile ai-scheduler
!
uc-tx-queue 3
    random-detect ecn count
!
interface Ethernet2/1
    description GPU1-Port2
    mtu 9214
    speed 200g-4
    no switchport
    channel-group 10 mode active
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet3/1
    description GPU1-Port3
    mtu 9214
    speed 200g-4
    no switchport
    channel-group 10 mode active
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet4/1
    description GPU1-Port4
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.10.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet5/1
    description GPU1-Port5
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.10.3.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
```

```
        random-detect ecn count
!
interface Ethernet32/1
    description SPINE1-Et3/1/1
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.13.1.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet33/1
    description SPINE1-Et4/1/1
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.13.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet34/1
    description SPINE2-Et3/2/1
    shutdown
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.14.1.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet35/1
    description SPINE2-Et4/2/1
    shutdown
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.14.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Loopback0
    ip address 192.168.101.1/32
```



```

!
interface Vxlan1
    vxlan udp-port 4789
    vxlan qos ecn propagation
!
ip routing
!
ip prefix-list LOOPBACK
    seq 10 permit 192.168.0.0/16 ge 32
!
ntp server 0.north-america.pool.ntp.org
!
qos map dscp 8 9 10 11 12 13 14 15 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 to
traffic-class 1
qos map dscp 16 17 18 19 20 21 22 23 to traffic-class 6
!
route-map LOOPBACKS permit 10
    match ip address prefix-list LOOPBACK
!
router bgp 65001
    router-id 192.168.101.1
    graceful-restart restart-time 300
    graceful-restart
    maximum-paths 4 ecmp 4
    neighbor GPU-SERVER peer group
    neighbor GPU-SERVER send-community
    neighbor GPU-SPINE peer group
    neighbor GPU-SPINE remote-as 65101
    neighbor GPU-SPINE send-community
    neighbor 11.10.1.2 peer group GPU-SERVER
    neighbor 11.10.1.2 remote-as 65151
    neighbor 11.10.2.2 peer group GPU-SERVER
    neighbor 11.10.2.2 remote-as 65152
    neighbor 11.13.1.2 peer group GPU-SPINE
    neighbor 11.13.2.2 peer group GPU-SPINE
    neighbor 11.14.1.2 peer group GPU-SPINE
    neighbor 11.14.2.2 peer group GPU-SPINE
    redistribute connected route-map LOOPBACKS
!
end

```

#### Leaf-2

```

!
no aaa root
!
username admin privilege 15 role network-admin secret
sha512 $6$eucN5ngreuExDgws$xnD7T8jO..GBDX0DUlp.
hn.W7yW94xTjSanqgaQGBzPIhDAsyAl9N4oSChvOMvf07uVBFI4mKMxwdVEUVKgY/.
!
prompt %H.%D{%H:%M:%S}%P

```

```
!  
service routing protocols model multi-agent  
!  
queue-monitor length  
!  
hostname Leaf-2  
ip name-server vrf MGMT 8.8.8.8  
!  
qos profile ai-scheduler  
  priority-flow-control on  
  priority-flow-control priority 3 no-drop  
  !  
  uc-tx-queue 1  
    no priority  
    bandwidth percent 5  
  !  
  uc-tx-queue 3  
    no priority  
    bandwidth percent 95  
    random-detect ecn minimum-threshold 256 kbytes maximum-threshold 512 kbytes  
max-mark-probability 100 weight 0  
!  
spanning-tree mode mstp  
!  
system l1  
  unsupported speed action error  
  unsupported error-correction action error  
!  
interface Port-Channel20  
  description GPU2-CHANNEL  
  mtu 9214  
  no switchport  
  ip address 11.20.1.1/30  
!  
interface Ethernet1/1  
  description GPU1-Port1  
  mtu 9214  
  speed 200g-4  
  no switchport  
  channel-group 20 mode active  
  service-profile ai-scheduler  
  !  
  uc-tx-queue 3  
    random-detect ecn count  
!  
interface Ethernet2/1  
  description GPU1-Port2  
  mtu 9214  
  speed 200g-4  
  no switchport
```

```
channel-group 20 mode active
service-profile ai-scheduler
!
uc-tx-queue 3
    random-detect ecn count
!
interface Ethernet3/1
    description GPU1-Port3
    mtu 9214
    speed 200g-4
    no switchport
    channel-group 20 mode active
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet4/1
    description GPU1-Port4
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.20.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet5/1
    description GPU1-Port5
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.20.3.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet32/1
    description SPINE1-Et3/1/1
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.23.1.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
```

```
channel-group 20 mode active
service-profile ai-scheduler
!
uc-tx-queue 3
    random-detect ecn count
!
interface Ethernet3/1
    description GPU1-Port3
    mtu 9214
    speed 200g-4
    no switchport
    channel-group 20 mode active
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet4/1
    description GPU1-Port4
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.20.2.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet5/1
    description GPU1-Port5
    mtu 9214
    speed 200g-4
    no switchport
    ip address 11.20.3.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
interface Ethernet32/1
    description SPINE1-Et3/1/1
    mtu 9214
    speed 400g-8
    no switchport
    ip address 11.23.1.1/30
    service-profile ai-scheduler
    !
    uc-tx-queue 3
        random-detect ecn count
!
```

```
interface Ethernet33/1
  description SPINE1-Et4/1/1
  mtu 9214
  speed 400g-8
  no switchport
  ip address 11.23.2.1/30
  service-profile ai-scheduler
  !
  uc-tx-queue 3
    random-detect ecn count
  !
interface Ethernet34/1
  description SPINE2-Et3/2/1
  shutdown
  mtu 9214
  speed 400g-8
  no switchport
  ip address 11.24.1.1/30
  service-profile ai-scheduler
  !
  uc-tx-queue 3
    random-detect ecn count
  !
interface Ethernet35/1
  description SPINE2-Et4/2/1
  shutdown
  mtu 9214
  speed 400g-8
  no switchport
  ip address 11.24.2.1/30
  service-profile ai-scheduler
  !
  uc-tx-queue 3
    random-detect ecn count
  !
interface Loopback0
  ip address 192.168.102.1/32
  !
interface Vxlan1
  vxlan udp-port 4789
  vxlan qos ecn propagation
  !
ip routing
  !
ip prefix-list LOOPBACK
  seq 10 permit 192.168.0.0/16 ge 32
  !
ntp server 0.north-america.pool.ntp.org
  !
qos map dscp 8 9 10 11 12 13 14 15 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 to
```

```

traffic-class 1
qos map dscp 16 17 18 19 20 21 22 23 to traffic-class 6
!
route-map LOOPBACKS permit 10
    match ip address prefix-list LOOPBACK
!
router bgp 65002
    router-id 192.168.102.1
    graceful-restart restart-time 300
    graceful-restart
    maximum-paths 4 ecmp 4
    neighbor GPU-SERVER peer group
    neighbor GPU-SERVER send-community
    neighbor GPU-SPINE peer group
    neighbor GPU-SPINE remote-as 65101
    neighbor GPU-SPINE send-community
    neighbor 11.20.1.2 peer group GPU-SERVER
    neighbor 11.20.1.2 remote-as 65151
    neighbor 11.20.2.2 peer group GPU-SERVER
    neighbor 11.20.2.2 remote-as 65152
    neighbor 11.23.1.2 peer group GPU-SPINE
    neighbor 11.23.2.2 peer group GPU-SPINE
    neighbor 11.24.1.2 peer group GPU-SPINE
    neighbor 11.24.2.2 peer group GPU-SPINE
    redistribute connected route-map LOOPBACKS
!
end

```

## Spine-1

```

!
no aaa root
!
username admin privilege 15 role network-admin secret
sha512 $6$eucN5ngreuExDgws$xnD7T8jO..GBDX0DUlp.
hn.W7yW94xTjSanqgaQGBzPIhDAsyAl9N4oSChvOMvf07uVBFI4mKMxwdVEUVKgY/.
!
prompt %H.%D{%H:%M:%S}%P
!
service routing protocols model multi-agent
!
queue-monitor length
!
hostname Spine-1
ip name-server vrf MGMT 8.8.8.8
!
qos profile ai-scheduler
    priority-flow-control on
    priority-flow-control priority 3 no-drop
!
tx-queue 1

```

```
        no priority
        bandwidth percent 5
    !
    tx-queue 3
        no priority
        bandwidth percent 95
        random-detect ecn minimum-threshold 512 kbytes maximum-threshold 768 kbytes
max-mark-probability 100
    !
spanning-tree mode mstp
    !
system 11
    unsupported speed action error
    unsupported error-correction action error
    !
queue-monitor streaming
    no shutdown
    !
management api http-commands
    no shutdown
    !
aaa authorization exec default local
    !
interface Ethernet3/1/1
    description LEAF1-Et32/1
    mtu 9214
    no switchport
    ip address 11.13.1.2/30
    service-profile ai-scheduler
    !
    tx-queue 3
        random-detect ecn count
    !
interface Ethernet3/3/1
    description LEAF2-Et32/1
    mtu 9214
    no switchport
    ip address 11.23.1.2/30
    service-profile ai-scheduler
    !
    tx-queue 3
        random-detect ecn count
    !
interface Ethernet4/1/1
    description LEAF1-Et33/1
    mtu 9214
    no switchport
    ip address 11.13.2.2/30
    service-profile ai-scheduler
    !
```

```
    tx-queue 3
      random-detect ecn count
!
interface Ethernet4/4/1
  description LEAF2-Et33/1
  shutdown
  mtu 9214
  no switchport
  ip address 11.23.2.2/30
  service-profile ai-scheduler
!
  tx-queue 3
    random-detect ecn count
!
interface Loopback0
  ip address 192.168.103.1/32
!
interface Vxlan1
  vxlan udp-port 4789
!
ip routing
!
ip prefix-list LOOPBACK
  seq 10 permit 192.168.0.0/16 ge 32
!
ntp server 0.north-america.pool.ntp.org
!
qos map dscp 8 9 10 11 12 13 14 15 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 to
traffic-class 1
qos map dscp 16 17 18 19 20 21 22 23 to traffic-class 6
!
route-map LOOPBACKS permit 10
  match ip address prefix-list LOOPBACK
!
router bgp 65101
  router-id 192.168.103.1
  graceful-restart restart-time 300
  graceful-restart
  maximum-paths 4 ecmp 4
  neighbor GPU-LEAF peer group
  neighbor GPU-LEAF send-community
  neighbor 11.13.1.1 peer group GPU-LEAF
  neighbor 11.13.1.1 remote-as 65001
  neighbor 11.13.2.1 peer group GPU-LEAF
  neighbor 11.13.2.1 remote-as 65001
  neighbor 11.23.1.1 peer group GPU-LEAF
  neighbor 11.23.1.1 remote-as 65002
  neighbor 11.23.2.1 peer group GPU-LEAF
  neighbor 11.23.2.1 remote-as 65002
  redistribute connected route-map LOOPBACKS
!
end
```



## Spine-2

```
!
no aaa root
!
username admin privilege 15 role network-admin secret
sha512 $6$eucN5ngreuExDgwS$xnD7T8jO..GBDX0DUlp.
hn.W7yW94xTjSanqgaQGBzPIhDAsyAl9N4oSChvOMvf07uVBFI4mKMxwdVEUVKgY/.
!
prompt %H.%D{%H:%M:%S}%P
!
service routing protocols model multi-agent
!
queue-monitor length
!
hostname Spine-2
ip name-server vrf MGMT 8.8.8.8
!
qos profile ai-scheduler
  priority-flow-control on
  priority-flow-control priority 3 no-drop
  !
  tx-queue 1
    no priority
    bandwidth percent 5
  !
  tx-queue 3
    no priority
    bandwidth percent 95
    random-detect ecn minimum-threshold 512 kbytes maximum-threshold 768 kbytes
max-mark-probability 100
!
spanning-tree mode mstp
!
system ll
  unsupported speed action error
  unsupported error-correction action error
!
queue-monitor streaming
  no shutdown
!
management api http-commands
  no shutdown
!
aaa authorization exec default local
!
interface Ethernet3/1/1
  description LEAF1-Et34/1
  mtu 9214
  no switchport
```

```
ip address 11.14.1.2/30
service-profile ai-scheduler
!
tx-queue 3
    random-detect ecn count
!
interface Ethernet3/3/1
    description LEAF2-Et34/1
    mtu 9214
    no switchport
    ip address 11.24.1.2/30
    service-profile ai-scheduler
    !
    tx-queue 3
        random-detect ecn count
!
interface Ethernet4/1/1
    description LEAF1-Et35/1
    mtu 9214
    no switchport
    ip address 11.14.2.2/30
    service-profile ai-scheduler
    !
    tx-queue 3
        random-detect ecn count
!
interface Ethernet4/4/1
    description LEAF2-Et35/1
    shutdown
    mtu 9214
    no switchport
    ip address 11.24.2.2/30
    service-profile ai-scheduler
    !
    tx-queue 3
        random-detect ecn count
!
interface Loopback0
    ip address 192.168.104.1/32
!
interface Vxlan1
    vxlan udp-port 4789
!
ip routing
!
ip prefix-list LOOPBACK
    seq 10 permit 192.168.0.0/16 ge 32
!
ntp server 0.north-america.pool.ntp.org
!
```

```
qos map dscp 8 9 10 11 12 13 14 15 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 to
traffic-class 1
qos map dscp 16 17 18 19 20 21 22 23 to traffic-class 6
!
route-map LOOPBACKS permit 10
    match ip address prefix-list LOOPBACK
!
router bgp 65101
    router-id 192.168.104.1
    graceful-restart restart-time 300
    graceful-restart
    maximum-paths 4 ecmp 4
    neighbor GPU-LEAF peer group
    neighbor GPU-LEAF send-community
    neighbor 11.14.1.1 peer group GPU-LEAF
    neighbor 11.14.1.1 remote-as 65001
    neighbor 11.14.2.1 peer group GPU-LEAF
    neighbor 11.14.2.1 remote-as 65001
    neighbor 11.24.1.1 peer group GPU-LEAF
    neighbor 11.24.1.1 remote-as 65002
    neighbor 11.24.2.1 peer group GPU-LEAF
    neighbor 11.24.2.1 remote-as 65002
    redistribute connected route-map LOOPBACKS
!
end
```

**Santa Clara—Corporate Headquarters**

5453 Great America Parkway,  
Santa Clara, CA 95054

Phone: +1-408-547-5500

Fax: +1-408-538-8920

Email: [info@arista.com](mailto:info@arista.com)

**Ireland—International Headquarters**

3130 Atlantic Avenue  
Westpark Business Campus  
Shannon, Co. Clare  
Ireland

**Vancouver—R&D Office**

9200 Glenlyon Pkwy, Unit 300  
Burnaby, British Columbia  
Canada V5J 5J8

**San Francisco—R&D and Sales Office 1390**

Market Street, Suite 800  
San Francisco, CA 94102

**India—R&D Office**

Global Tech Park, Tower A, 11th Floor  
Marathahalli Outer Ring Road  
Devarabeesanahalli Village, Varthur Hobli  
Bangalore, India 560103

**Singapore—APAC Administrative Office**

9 Temasek Boulevard  
#29-01, Suntec Tower Two  
Singapore 038989

**Nashua—R&D Office**

10 Tara Boulevard  
Nashua, NH 03062



Copyright © 2025 Arista Networks, Inc. All rights reserved. CloudVision, and EOS are registered trademarks and Arista Networks is a trademark of Arista Networks, Inc. All other company names are trademarks of their respective holders. Information in this document is subject to change without notice. Certain features may not yet be available. Arista Networks, Inc. assumes no responsibility for any errors that may appear in this document. January 13, 2025 07-00010-08