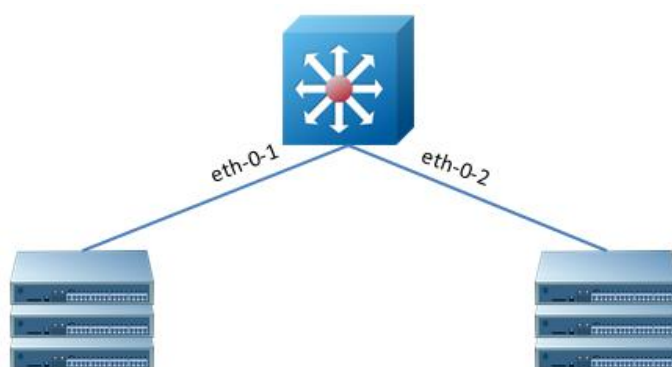


1 示例场景介绍

1.1 业务需求

- RoCEv2 业务按照要求划分到优先级 3 和 6
- 开启 PFC 和 ECN 功能，实现无损网络

1.2 RDMA 组网示例



1.3 交换机在 RDMA 网络中的作用

- 交换机利用 MQC 功能，对业务报文进行精确匹配后灵活分类，并划分优先级。
- 根据映射后的优先级，实现基于优先级的流控，并基于 WRED 机制完成 ECN 标记，保障 RoCEv2 报文的无损转发。常见的映射规则为 COS 和 DSCP 映射。

2 基于 COS 的 RDMA 配置示例

2.1 交换机配置

2.1.1 开启 ECN 功能

```
#配置优先级 3 及优先级 6 的 ECN 策略

switch#configure terminal

switch(config)#class-map type traffic-class ecn-pri-3

switch(config-cmap-tc)#match traffic-class 3

switch(config-cmap-tc)#exit

switch(config)#class-map type traffic-class ecn-pri-6

switch(config-cmap-tc)#match traffic-class 6

switch(config-cmap-tc)#exit

switch(config)#policy-map type traffic-class ecn-pmap

switch(config-pmap-tc)#class type traffic-class ecn-pri-3

switch(config-pmap-tc-c)#random-detect maximum-threshold 8192 minimum-threshold 64

switch(config-pmap-tc-c)#exit

switch(config-pmap-tc)#class type traffic-class ecn-pri-6

switch(config-pmap-tc-c)#random-detect maximum-threshold 8192 minimum-threshold 64

switch(config-pmap-tc-c)#end


#将 ECN 策略应用至对应的物理接口

switch#configure terminal

switch(config)#interface eth-0-1

switch(config-if)#service-policy type traffic-class ecn-pmap

switch(config-if)#exit

switch(config)#interface eth-0-2

switch(config-if)#service-policy type traffic-class ecn-pmap
```

```
switch(config-if)#end
```

2.1.2 创建业务 Vlan

```
#创建业务 Vlan10,并将对应物理接口加入到业务 Vlan

switch#configure terminal
switch(config)#vlan database
switch(config-vlan)#vlan 10
switch(config-vlan)#exit
switch(config)#interface eth-0-1
switch(config-if)#switchport mode trunk
switch(config-if)#switchport trunk allowed vlan add 10
switch(config)#interface eth-0-2
switch(config-if)#switchport mode trunk
switch(config-if)#switchport trunk allowed vlan add 10
```

2.1.3 开启 PFC 功能

```
#开启对应物理接口优先级 3 和优先级 6 的 PFC 功能

switch#configure terminal
switch(config)#interface eth-0-1
switch(config-if)#priority-flow-control mode on
switch(config-if)#priority-flow-control enable priority 3 6
switch(config)#interface eth-0-2
switch(config-if)#priority-flow-control mode on
switch(config-if)#priority-flow-control enable priority 3 6
switch(config-if)#end
```

2.2 服务器网卡（Mellanox）配置

```
#配置网卡（名称 eth6）到 vlan 10

modprobe 8021q
vconfig add eth6 10
ifconfig eth6.10 192.168.105.194/24 up
ifconfig eth6 0
```

#配置流控相关的参数:

```
mlnx_qos -i eth6 --pfc 0,0,0,1,0,0,0,0  
  
echo 1 > /sys/class/net/eth6/ecn/roce_np/enable/3  
  
echo 1 > /sys/class/net/eth6/ecn/roce_rp/enable/3  
  
echo 6 > /sys/class/net/eth6/ecn/roce_np/cnp_802p_prio  
  
sysctl -w net.ipv4.tcp_ecn=1  
  
cma_roce_mode -d mlx5_2 -p 1 -m 2  
  
sysctl -w net.ipv4.tcp_ecn=1  
  
cma_roce_mode -d mlx5_2 -p 1 -m 2  
  
cma_roce_tos -d mlx5_2 -t 24  
  
vconfig set_egress_map eth6.10 4 3
```

说明:

服务器网卡配置请以服务器厂商配置手册为准

3 基于 DSCP 的 RDMA

3.1 交换机配置

3.1.1 开启 ECN 功能

```
#配置优先级 3 及优先级 6 的 ECN 策略

switch#configure terminal
Enter configuration commands, one per line. End with CNTL/Z.
switch(config)#class-map type traffic-class ecn-pri-3
switch(config-cmap-tc)#match traffic-class 3
switch(config-cmap-tc)#exit
switch(config)#class-map type traffic-class ecn-pri-6
switch(config-cmap-tc)#match traffic-class 6
switch(config-cmap-tc)#exit
switch(config)#policy-map type traffic-class ecn-pmap
switch(config-pmap-tc)#class type traffic-class ecn-pri-3
switch(config-pmap-tc-c)#random-detect maximum-threshold 8192 minimum-threshold 64
switch(config-pmap-tc-c)#exit
switch(config-pmap-tc)#class type traffic-class ecn-pri-6
switch(config-pmap-tc-c)#random-detect maximum-threshold 8192 minimum-threshold 64
switch(config-pmap-tc-c)#end

#将 ECN 策略应用至对应的物理接口

switch#configure terminal
switch(config)#interface eth-0-1
switch(config-if)#service-policy type traffic-class ecn-pmap
switch(config-if)#exit
switch(config)#interface eth-0-2
```

```
switch(config-if)#service-policy type traffic-class ecn-pmap
switch(config-if)#end
```

3.1.2 创建业务 Vlan

```
#创建业务 Vlan10,并将对应物理接口加入到业务 Vlan

switch#configure terminal
switch(config)#vlan database
switch(config-vlan)#vlan 10
switch(config-vlan)#exit
switch(config)#interface eth-0-1
switch(config-if)#switchport access vlan 10
switch(config)#interface eth-0-2
switch(config-if)#switchport access vlan 10
```

3.1.3 配置 PFC

```
#将 DSCP=26 (af31) 映射到优先级 3, 将 DSCP=48 (CS6) 映射到优先级 6

switch#configure terminal
Enter configuration commands, one per line.  End with CNTL/Z.
switch(config)#class-map type qos pfc-pri-3
switch(config-cmap-qos)#match dscp af31
switch(config-cmap-qos)#exit
switch(config)#class-map type qos pfc-pri-6
switch(config-cmap-qos)#match dscp cs6
switch(config-cmap-qos)#exit
switch(config)#policy-map type qos pfc-pmap
switch(config-pmap-qos)#class type qos pfc-pri-3
switch(config-pmap-qos-c)#set traffic-class 3
switch(config-pmap-qos-c)#exit
switch(config-pmap-qos)#class type qos pfc-pri-6
switch(config-pmap-qos-c)#set traffic-class 6
switch(config-pmap-qos-c)#end
```

#将映射策略应用到对应的物理接口，并开启对应物理接口优先级 3 和优先级 6 的 PFC 功能

```
switch#configure terminal
switch(config)# interface eth-0-1
switch(config-if)#priority-flow-control mode on
switch(config-if)#priority-flow-control enable priority 3 6
switch(config-if)#service-policy type qos input pfc-pmap
switch(config)# interface eth-0-2
switch(config-if)#priority-flow-control mode on
switch(config-if)#priority-flow-control enable priority 3 6
switch(config-if)#service-policy type qos input pfc-pmap
```

3.2 服务器网卡（Mellanox）配置

```
#网卡名称 eth7
// Enable DCQCN in priority 3
echo 1 > /sys/class/net/eth7/ecn/roce_np/enable/3
echo 1 > /sys/class/net/eth7/ecn/roce_rp/enable/3
// Configure CNP to use DSCP over priority 6.
echo 48 > /sys/class/net/eth7/ecn/roce_np/cnp_dscp
// Enable PFC at priority 3 and 6.
mlnx_qos -i eth7 --trust=dscp --pfc=0,0,0,1,0,0,1,0
```

说明：

服务器网卡配置请以服务器厂商配置手册为准