August 2020

Congestion control for Ethernet traffic on IBM Spectrum Virtualize

Prioritize and manage different types of Ethernet traffic

Table of contents

Overview

Challenge
IBM Spectrum Virtualize supports various Ethernet traffic types or classes such as iSCSI host attach, iSCSI storage, iSER host attach, RDMA-based clustering for node-to-node communication and so on. Managing the entire traffic at a single Ethernet port needs prioritization of traffic based on traffic type.

Solution
IBM Storwize provides a mechanism that enables you to manage Ethernet traffic on every Ethernet port.

Priority flow control on IBM Spectrum Virtualize

Overview of PFC

Priority flow control (PFC) is an Ethernet protocol that supports the ability to assign priorities to different types of traffic within the network. With PFC configured, administrators can reduce network disruption due to congestion by slowing down or pausing specific classes of traffic on ports, thus providing better bandwidth for more important traffic.

IBM® Spectrum Virtualize systems support priority flow control for iSCSI host attach traffic since the 7.5.0.0 release. Since then till release 8.2.x.x, PFC was supported on Emulex 10 Gb ports only.

In Spectrum Virtualize 8.3.0.0, IBM introduced PFC support for more traffic classes and for additional Ethernet adapters. PFC support has been added for new protocols such as iSER on the faster 25Gb Ethernet ports. Because 25Gb ports interoperate with 10Gb ports, mismatched speeds increase the chances of congestion on the fabric if not configured correctly. This paper includes configuration details for setting up PFC before and after version 8.3.0.0 along with troubleshooting tips.

Architecture

Software
- IBM Spectrum Virtualize7.5.0.0 onwards only for iSCSI
- IBM Spectrum Virtualize 8.2.1 onward for separate traffic classes

Hardware
List of IBM controllers that support PFC:
- IBM FlashSystem 9100, FlashSystem 9150, FlashSystem 200
- Storwize V7000 Gen2, Storwize V7000 Gen3
- SV1, SV2

Network
- DCBX supported switches
- LLDP enabled on switch

End-to-end configuration for congestion control on IBM SVC 8.2.x.x or earlier

From release 7.5.0.0 till release 8.2.x.x PFC was supported only on Emulex 10Gb ports for iSCSI host attach traffic only. This section provides the steps to configure PFC for these releases explained as flowchart.

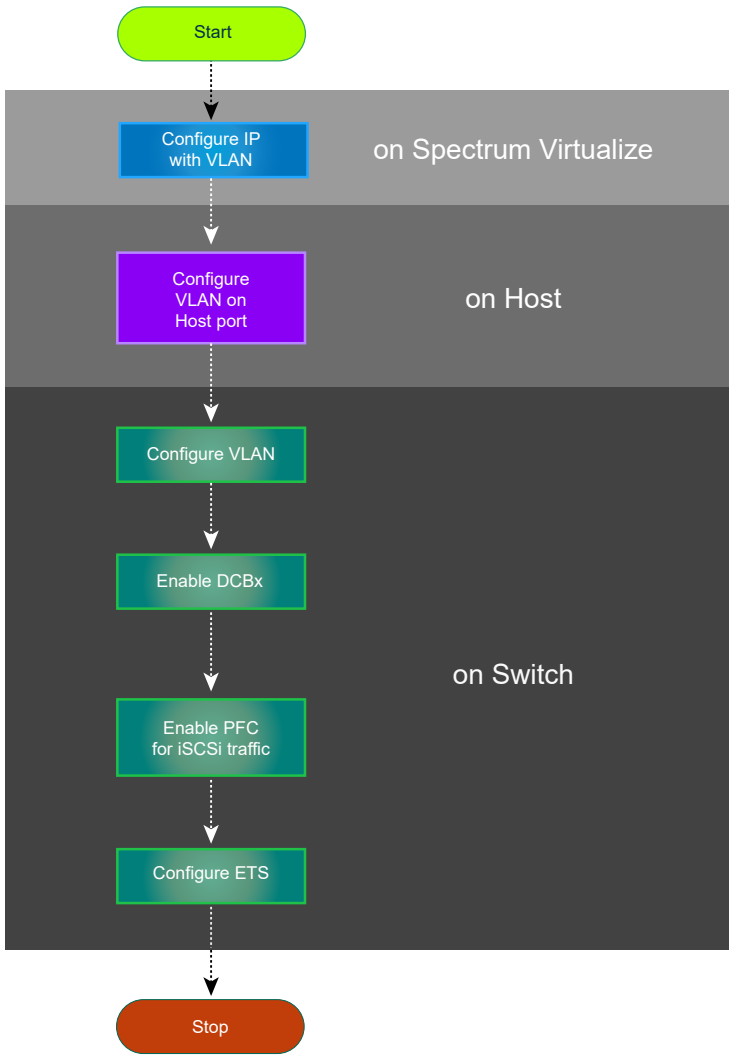

Figure 1. Flow chart depicting flow of settings required to configure PFC on IBM Storwize 8.2.x.x or earlier

1. Configure IP addresses with virtual local area network (VLAN) on Spectrum Virtualize for host attach ports using the `svctask cfgportip` command.
   Note: VLAN is compulsory to configure PFC.
2. Ensure that VLAN is correctly configured on the host ports.

3. Ensure that VLAN is correctly configured on the switches and routers present in fabric by using vendor-specific commands.
4. Make sure that the host and storage connectivity is correct, and they can ping each other.
5. Enable Data Center Bridging Capability Exchange (DCBX) on switch (this is required for PFC to work).
6. Define and enable PFC for iSCSI traffic tag on the switch. Your switch must support iSCSI Type Length Value (TLV) to support PFC with SVC. Check vendor-specific guidelines for the correct steps.
7. Optionally, configure Enhance Transmission Selection (ETS) settings on switch according to the workload by using vendor-specific commands.

Things to remember:

- Port speed should be 10Gb or higher to configure PFC settings.
- To enable priority flow for host attachment traffic on a port, make sure that the host flag is set to *yes* on the configured IP on that port.
- To enable priority flow for storage traffic on a port, make sure that storage flag is set to *yes* on the configured IP on that port.
- PFC on RDMA over Converged Ethernet (RoCE) adapters will be supported only from version 8.4.0.0 onwards.

End-to-end configuration for congestion cotrol on IBM Spectrum Virtualize 8.3.0.0 onwards

From 8.3.0.0 release, the steps to support PFC have changed from earlier. SVC now supports various types of traffic that can share the same port, for example, iSCSI host attach traffic, iSCSI storage traffic, iSER host attach traffic can now share a single Ethernet port.

In SVC, Ethernet traffic is divided in to three different traffic classes:

- Host attach
- Storage
- Node-to-node communication (or clustering or system)

These traffic classes are system-wide and pertain to the function rather than the protocol. For example, once a priority class of service is assigned to host attach traffic, all iSCSI and iSER host attach traffic connected to PFC-enabled ports will use the same traffic class.

Perform the following steps to configure PFC 8.3.0.0 or later for the traffic classes explained as flow chart:

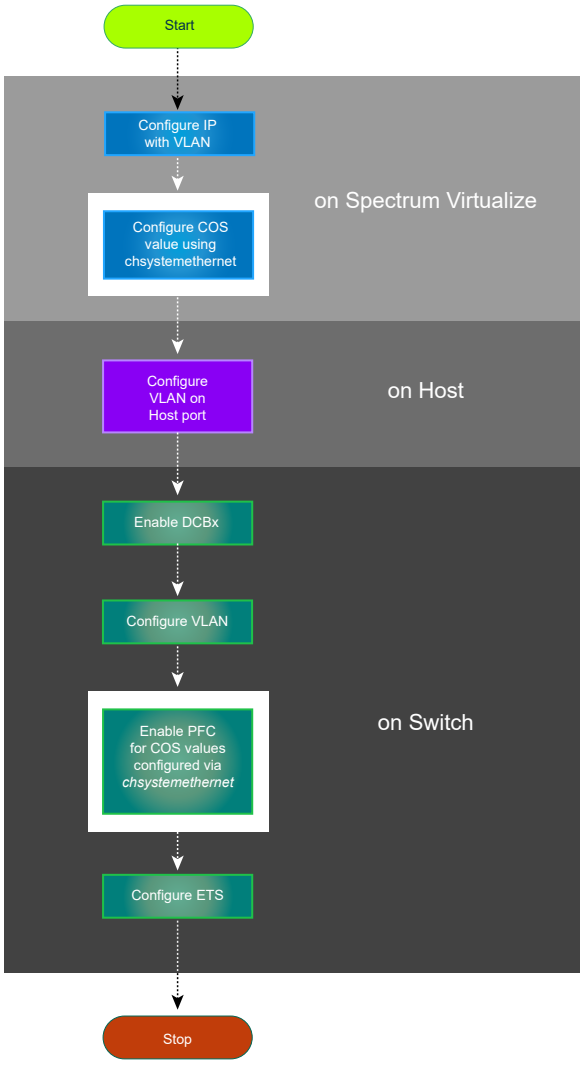

Figure 2. Flow chart depicting the flow of settings required to configure PFC on IBM Storwize for 8.3.0.0 or later versions

Steps that are different from releases before version 8.3.0.0 are highlighted using purple color. This section provides details about each step.

1. Configure IP addresses with VLAN on Spectrum Virtualize ports. **Note**: VLAN is compulsory to configure PFC.
2. Ensure that VLAN is correctly configured on the host ports.
3. Ensure that VLAN is correctly configured on the switches and routers by using vendor-specific commands.
4. Make sure that the host and storage connectivity is correct, and they can ping each other.
5. Configure the *Class of Service* values for different classes of traffic on SVC as described earlier by using the `svctask`

`chsystemethernet` command.

6. Enable DCBX on switch (this is required for PFC to work.
7. Define and enable PFC on the switch for the traffic classes defined in step 3.
8. Optionally, configure the ETS settings on the switch according to the workload by using vendor-specific commands.

Things to remember:

- Set the priority tag in the range 0 to 7.
- For each supported traffic class, enable the same priority tag on the switch. For example, if you plan to have a priority tag 3 for host attach traffic, ensure that the priority tag 3 is enabled on the switch.
- If you are planning on using the same port for different types of traffic, ensure that the ETS settings are configured on the network.

Configuring PFC for host attach traffic on IBM Spectrum Virtualize

This section explains the configuration steps to be done on IBM Spectrum Virtualize for configuring PFC for host attach traffic.

Assign a new IP address for host attach traffic

You can assign a new IP address for host attach communicationif it is not set using the `svctask cfgportip` command. Along with IP, assign VLAN on the port.

CLI example: Assign a new IP on port 7 with VLAN 30 for host attach traffic

svctask cfgportip -node node1 -ip 192.X.X.X -mask 255.255.255.0 -gw 192.X.X.1 **-vlan 30 -host yes 7**

Note: If IPv6 is used then `host_6` and `vlan_6` parameters will be used instead of `host` and `vlan`.

GUI example:

1. Assign a new IP on port 7 with VLAN 30 for host attach traffic by clicking **Settings → Network → Ethernet Ports**.

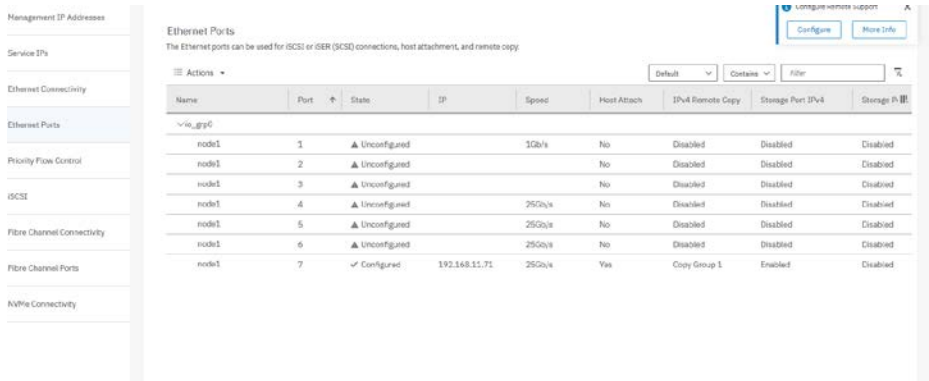

Figure 3. Identify Ethernet port

2. Right click an unconfigured port for which you want to configure the IP and click **Modify IP Settings**.

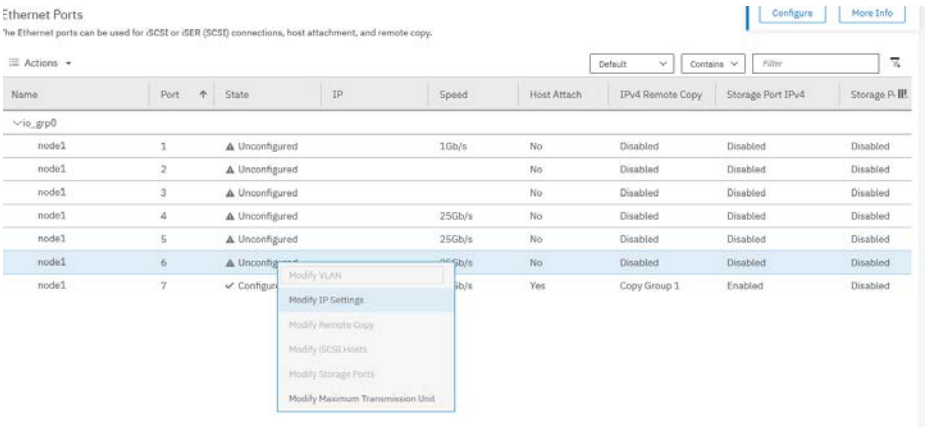

Figure 4. Modify IP setting

3. Enter the IP and click **Modify**.

4. To use this IP for host attach traffic, right-click the IP and click **Modify iSCSI Hosts**.

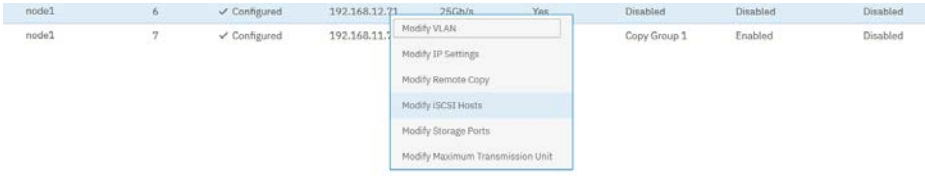

Figure 5. Update VLAN

5. Make sure that the iSCSI host is enabled. If it is not, then select **Enabled** and click **Modify**.

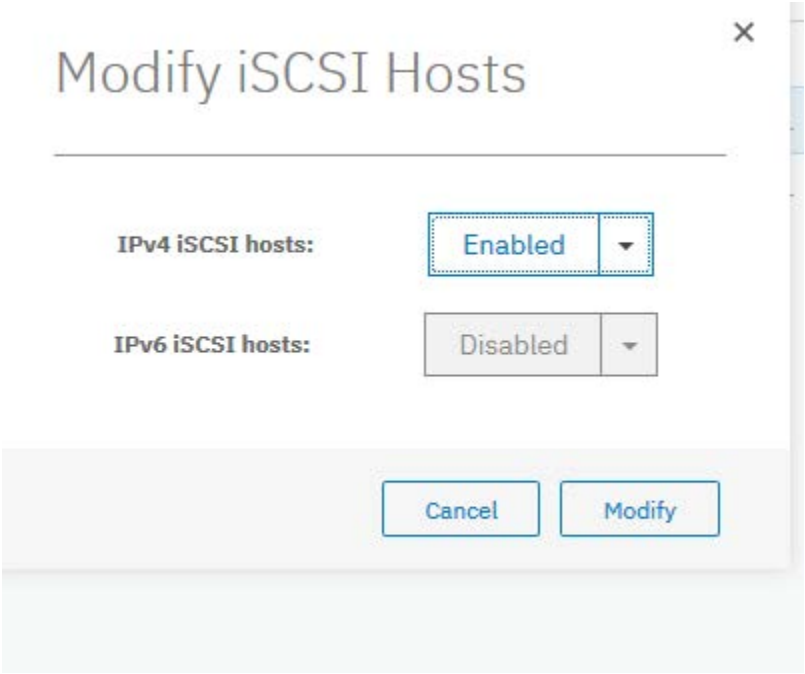

Figure 6. Modify iSCSI hosts

Assign VLAN on a port with already configured IP

If IPs are already assigned, but VLAN is not enabled then configure VLAN for the existing IPs using the `svctask cfgportip` command.

Exampe CLI: Assign VLAN 30 on port 7 for node1 of the storage controller

```
svctask cfgportip -node node1 -vlan 30 7
```

The `svcinfo lsportip` command can be used to verify the changes.

Example GUI: Assign VLAN 30 on port 7 for node1 of the storage controller using GUI

1. In the IBM Spectrum Virtualize GUI, click **Settings → Network → Ethernet Ports**.

Things to remember:

- VLAN configuration is mandatory for PFC to function
- VLAN configuration must be used by all entities (such as nodes, hosts, backend nodes, and switches)
- Only DCBX capable switches can support PFC configuration
- DCBX enables switch and adapter ports to exchange parameters that describe traffic classes and PFC capabilities. For these steps in detail, check your switch documentation.

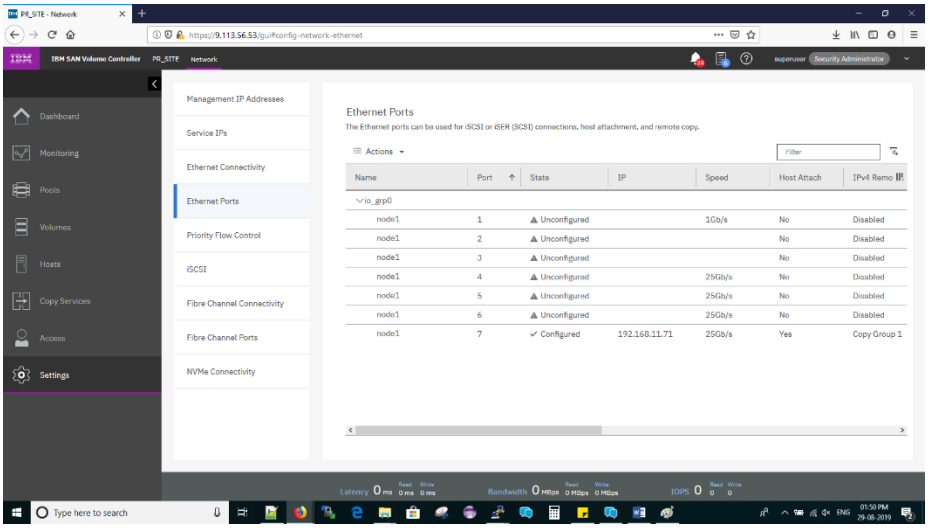

Figure 7: Identify ethernet port

2. Right-click the port and click **Modify VLAN**.

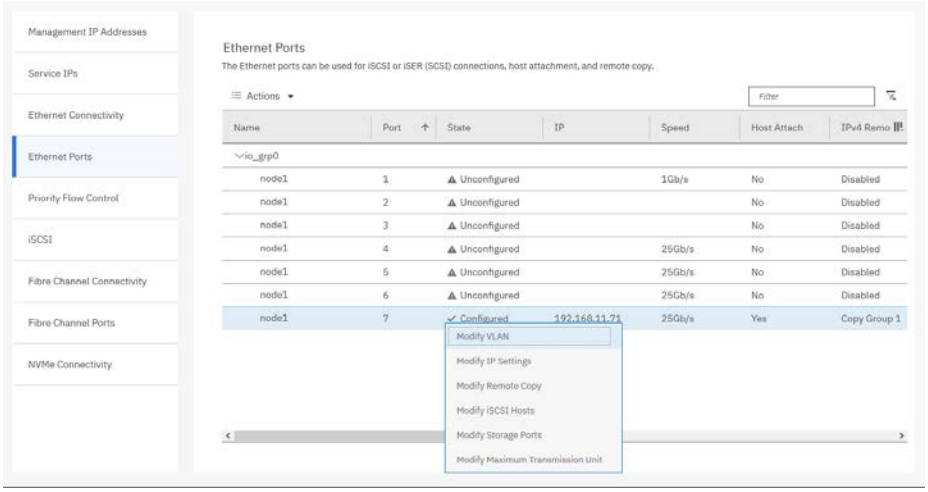

Figure 8. Update VLAN

3. In the dialog box that is opened, select the **Enable** check box.

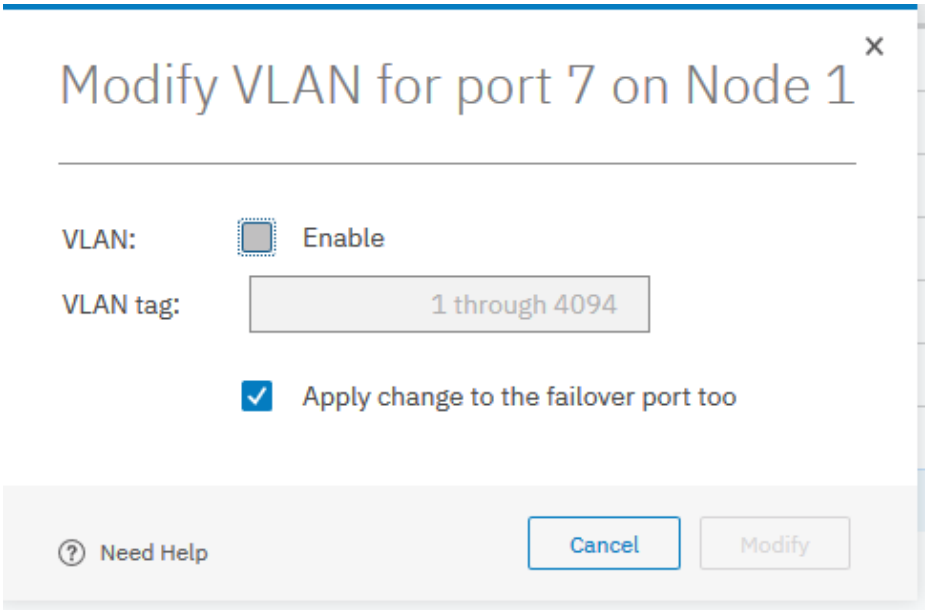

Figure 9: Apply changes

4. Specify the VLAN tag and click **Modify**.

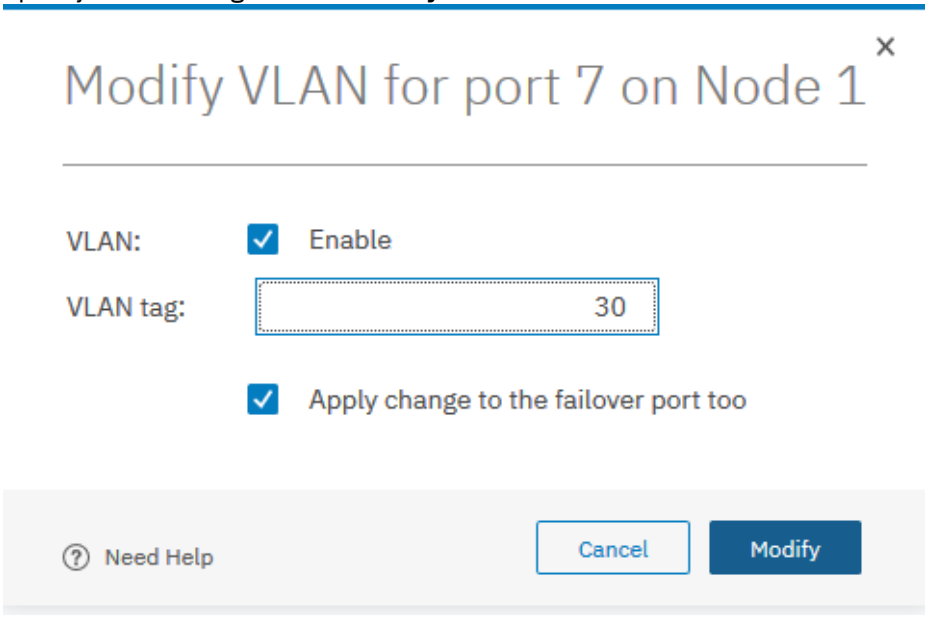

Figure 10: Specify VLAN ID

5. A warning message indicating that if you modify VLAN of an existing IP address then all connections through that IP will be lost is displayed. Click **Yes**.

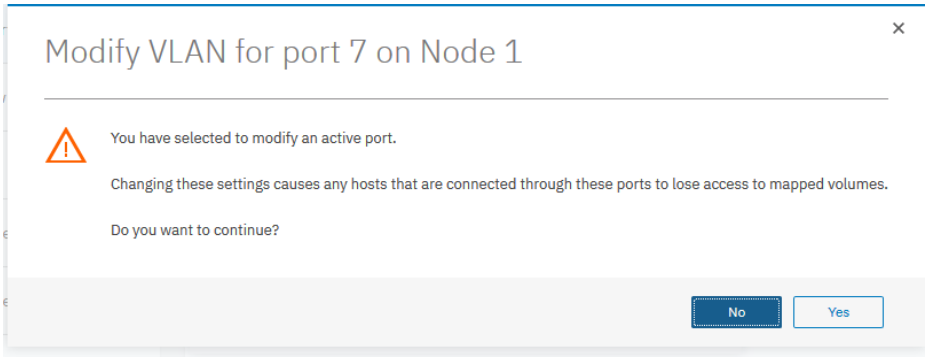

Figure 11: Update the port

Assign COS value for host attach traffic on SVC

After the IBM Storwize ports are configured with the correct VLAN tag, configure the priority code point (PCP) value or the class of service (COS) value for host attach traffic on SVC using the `chsystemethernet` command.

Things to remember for PCP settings:

- Set the priority tag in the range 0 to 7.
- For each supported traffic class, enable the same priority tag on the switch. For example, if you plan to have a priority tag 3 for storage traffic, ensure that the priority tag 3 is enabled on the switch.
- If you are planning on using the same port for different types of traffic, ensure that the ETS settings are configured on the network.

CLI example:

```
svctask chsystemethernet -hostattachcos 4
```

GUI example: Defining priority tags for defined traffic classes using GUI

1. Click **Settings → Network → Priority Flow Control**.
2. Provide priority tag values to host attach host attach traffic and click **Save**.

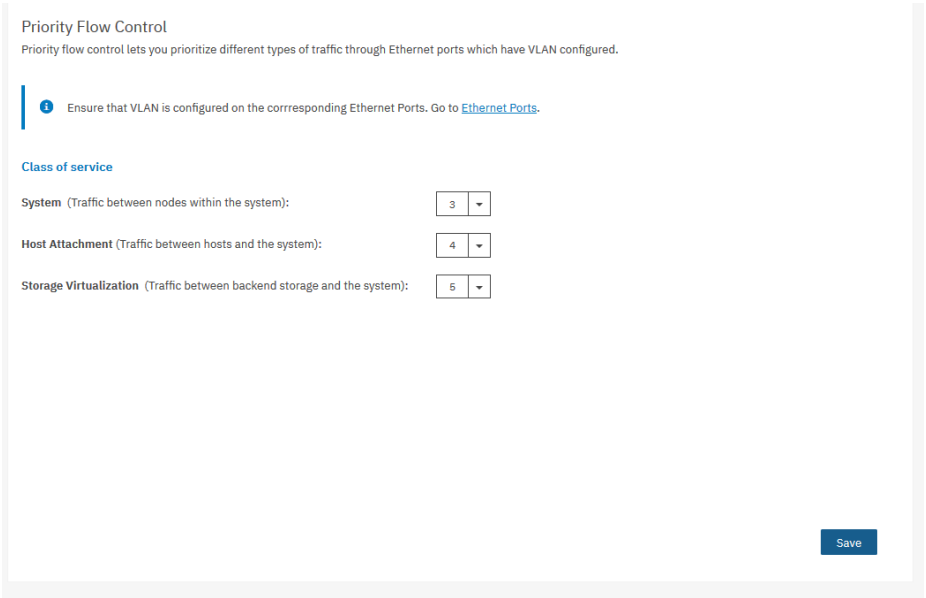

Figure 12. Specify COS value

Only single traffic class can also be assigned priority keeping others as 0. Rescanning or reestablishing iSER host attach sessions are required after changes are made for PFC tag settings.

Configuring PFC for system or clustering traffic on IBM Spectrum Virtualize

This section explains the configuration steps to enable PFC for system or clustering traffic. (system or clustering traffic is often called as Remote Direct Memory Access (RDMA)-based node-to-node traffic).

Assign a new IP address for node-to-node communication

New IP address for clustering between IBM Spectrum Virtualize nodes can be assigned, if it is not set using the `satask chnodeip` CLI, along with IP assign VLAN on the port.

Example CLI: Assign a new IP address with VLAN 30 on port 5 for node-to-node or system communication

satask chnodeip -ip 192.168.59.14 -gw 192.168.59.1 -mask 255.255.255.0 -vlan 30 -port_id 5

This can be verified using the `sainfo lsnodeip` command:

```
IBM_2145:ibmcluster:superuser>sainfo lsnodeip
port_id    rdma_type port_speed vlan link_state state        node_IP_address gateway      subnet_mask
1                    1Gb/s           active     unconfigured
2                    1Gb/s           active     unconfigured
3                                    inactive   unconfigured
4          RoCE      25Gb/s          active     configured   10.10.12.17     10.10.12.1   255.255.255.0
5          RoCE      25Gb/s     30   active     configured   192.168.59.14   192.168.59.1 255.255.255.0
6                    10Gb/s          active     unconfigured
7                    10Gb/s          active     unconfigured
IBM_2145:ibmcluster:superuser>sainfo lsnodeipconnectivity
status          local_port_id local_vlan local_rdma_type local_ip_addr remote_port_id remote_vlan remote_rdma_type remote_ip_addr
remote_wwnn     remote_panel_name cluster_id       error_data
Connected:RoCE  4                     RoCE           10.10.12.17   4                       RoCE             10.10.12.18
500507680C008B01 78FNML0          000020321C1160C
Connected:RoCE  5             30      RoCE           192.168.59.14 5             30        RoCE             192.168.59.13
500507680C008B01 78FNML0          000020321C1160C
IBM_2145:ibmcluster:superuser>
```

Figure 13. Output of sainfo lsnodeip

Example GUI: To configure a system IP login to service assistance GUI:

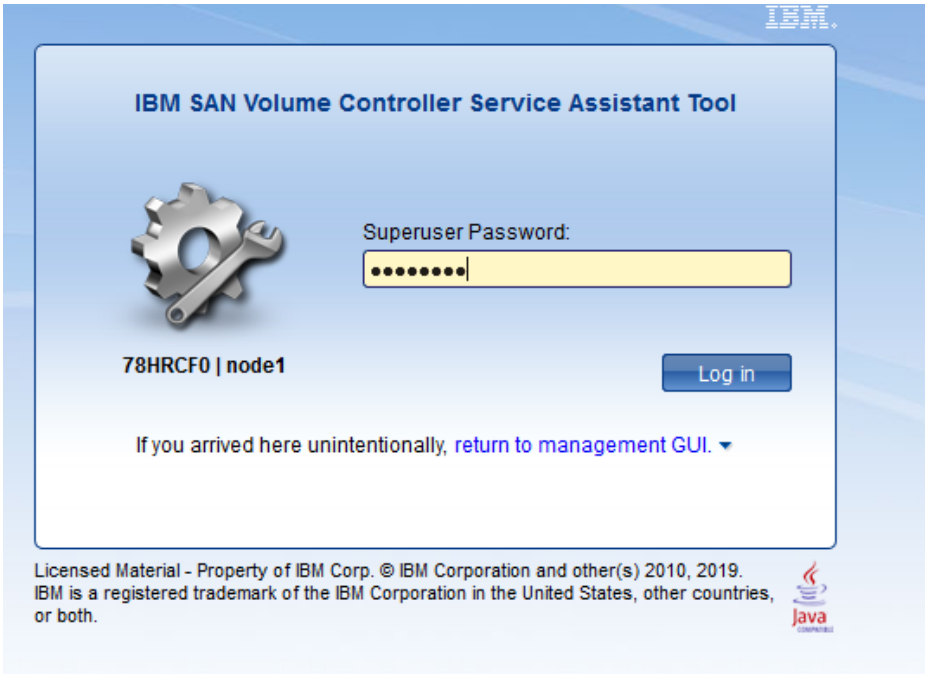

Figure 14. Log in to Service Assistance GUI

1. On the left panel, click **Change Node IP**.

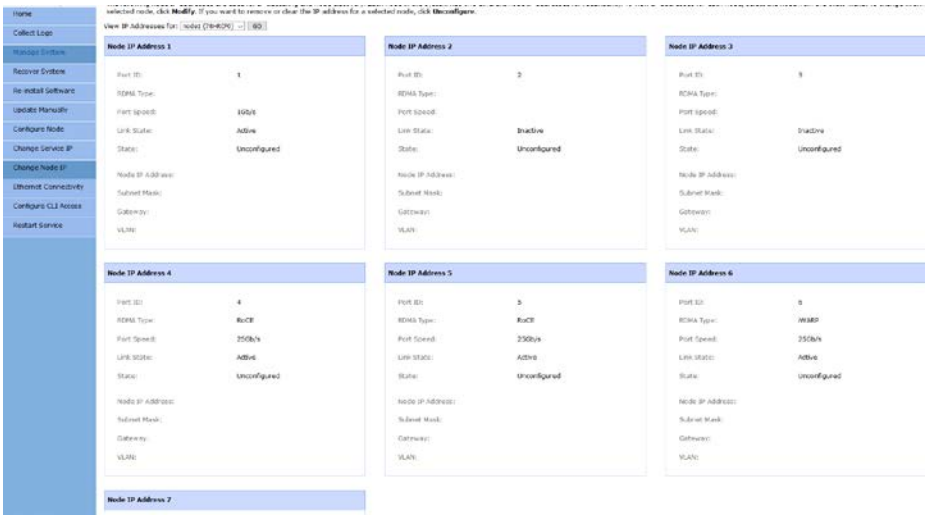

Figure 15 Select the Ethernet port

2. Select a port on which you want to configure system IP and click **Modify**.

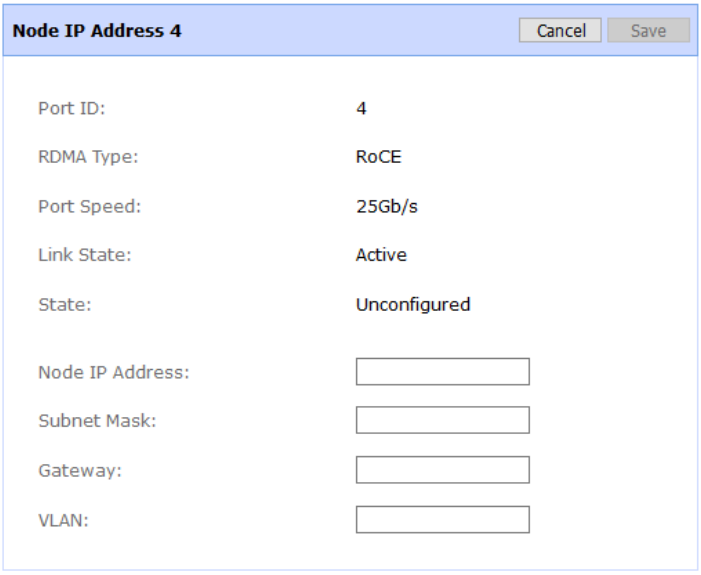

Figure 16 Modify settings

3. Enter node IP, subnet, gateway, and VLAN details, and then click **Save**. Notice that the node IP details are updated.

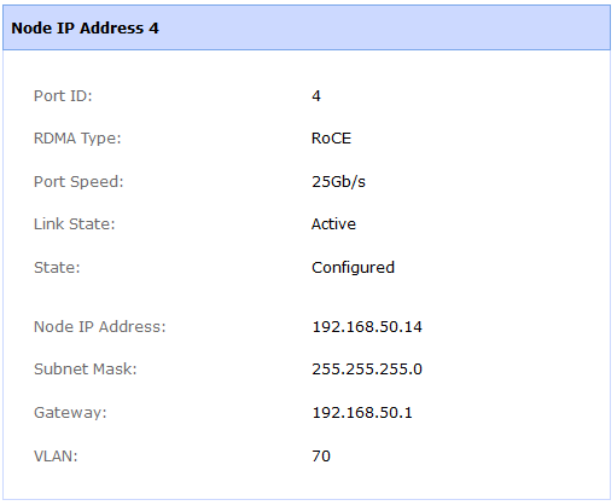

Figure 17. Set node IP

4. Similarly, configure the other ports as required.

Things to remember:

- VLAN configuration is mandatory for PFC to function.
- VLAN configuration must be used by all entities (such as nodes, backend nodes, hosts, and switches)
- Only DCBX-capable switches can support PFC configuration.
- DCBX enables switch and adapter ports to exchange parameters that describe traffic classes and PFC capabilities. For detailed steps, check your switch documentation.

Assign VLAN on a port with an already configured IP

If IPs are already assigned but VLAN is not enabled, then configure VLAN for the existing IPs using the same CLI command.

Example GUI: To configure a VLAN on an existing system IP login to service assistance GUI

On the left panel, click **change Node IP**, and select the node IP details which you want to modify. In this example, modify port number 5 (as VLAN is not configured here).

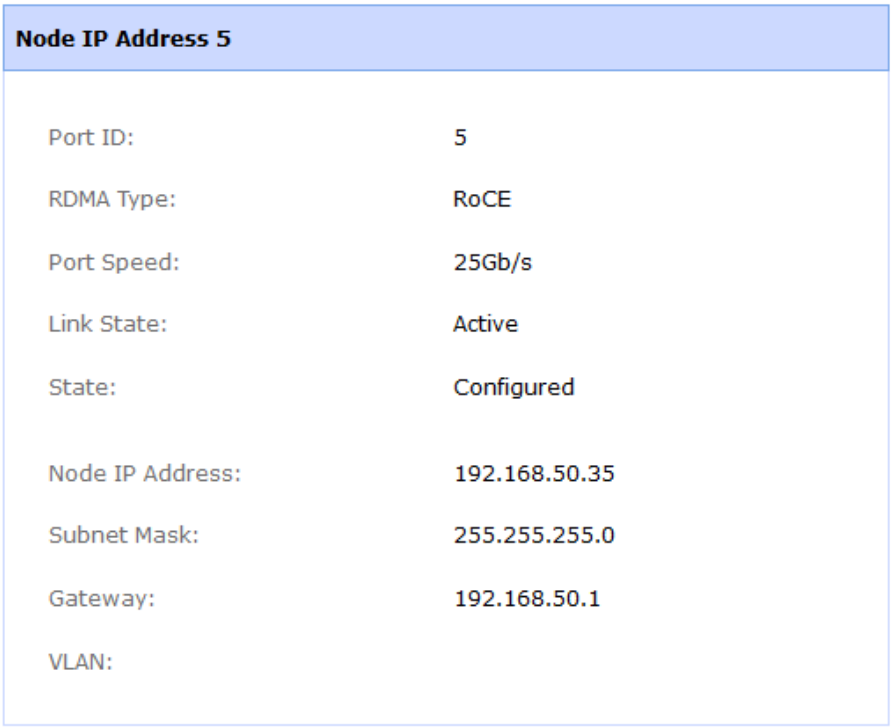

Figure 18. Select port

Select the node IP and click **Modify**. Specify VLAN and click **Save**.

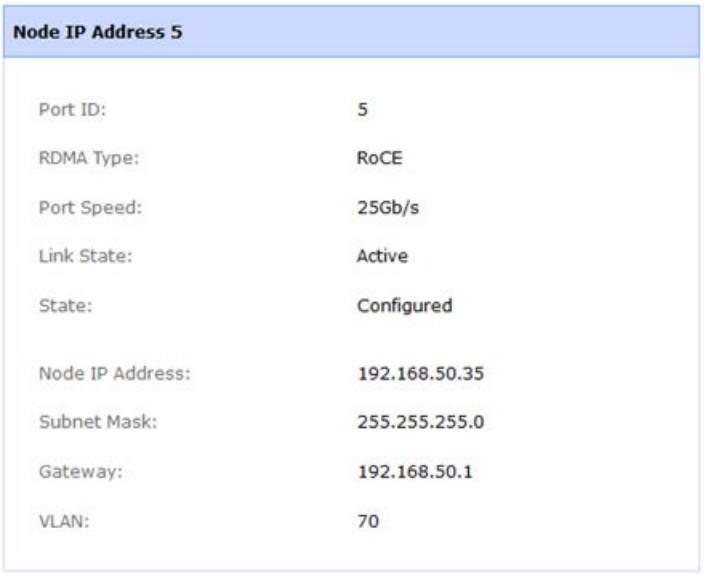

Figure 19. Update VLAN ID

Assign COS value for node-to-node traffic on SVC

After configuring the IBM Storwize ports with the correct VLAN tag, configure the PCP value or the COS value for system traffic on SVC using the `chsystemethernet` command.

CLI example:
svctask chsystemethernet -systemcos 3

GUI example: Defining priority tags for defined traffic classes using GUI

1. Click **Settings** → **Network** → **Priority Flow Control**.

2. Enter the priority tag values for system traffic and click **Save**.

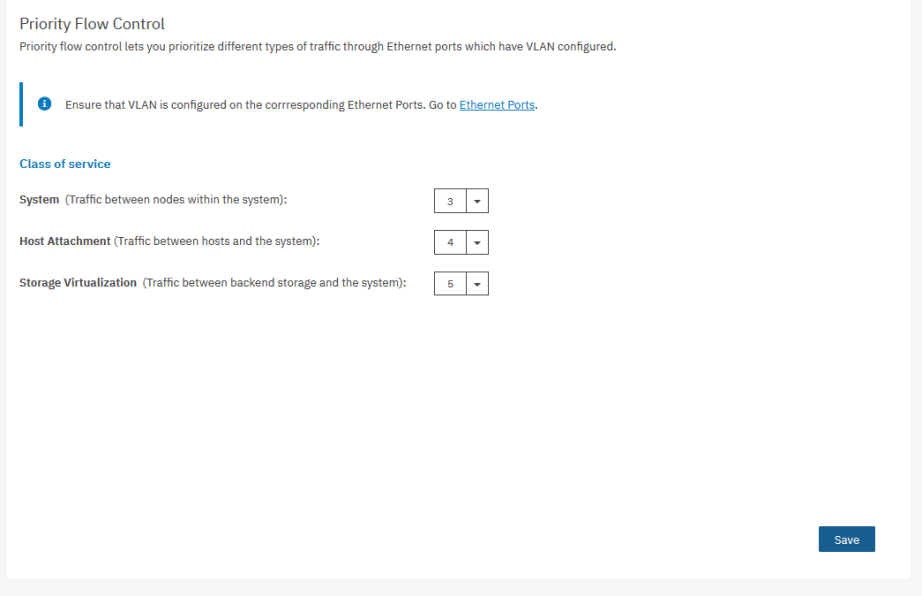

Figure 20. Specify COS value

Things to remember for PCP settings:

- Set the priority tag in the range 0 to 7.
- For each supported traffic class, enable the same priority tag on the switch. For example, if you plan to have a priority tag 3 for storage traffic, ensure that the priority tag 3 is enabled on the switch.
- If you are planning on using the same port for different types of traffic, ensure that the ETS settings are configured on the network.

Note:
Only single traffic class can also be assigned priority keeping others as 0. Changing PFC COS value for system or clustering traffic requires re-establishing RDMA-based node to node sessions, after PFC tag settings are done on cluster and after any node joins the cluster. To re-establish the session, you must have at least two RDMA-based node-to-node sessions.

Once the pre-requites are met, close one of the sessions and open that session again, and then close the second session and then reopen it and so on. One way to open and close the session again is just by removing that node IP (using `satask chnodeip -noip -port_id <port id>`) and configuring it back again.

Configuring PFC for storage traffic on IBM Spectrum Virtualize

This section explains the configuration steps to be done on IBM Spectrum Virtualize for configuring PFC for back-end storage traffic.

Assign a new IP address for back-end storage traffic

A new IP address for back-end storage traffic can be assigned, if it is not set using the `svctask cfgportip` command. Along with the IP, assign VLAN on the port.

Example CLI: Assign a new IP address on a port for only storage traffic with VLAN 40 on port 6

svctask cfgportip -node node1 -ip 192.X.X.X -mask 255.255.255.0 -gw

192.X.X.1 **-vlan 40 -storage yes** 6

To enable both host and storage traffic on the same port, use following command:

svctask cfgportip -node node1 -ip 192.X.X.X -mask 255.255.255.0 -gw 192.X.X.1 **-vlan 40 -host yes -storage yes** 5

Example GUI: Assign a new IP address for configuring storage traffic

1. To configure a new host attach or storage IP, click **Settings → Network → Ethernet Ports**.

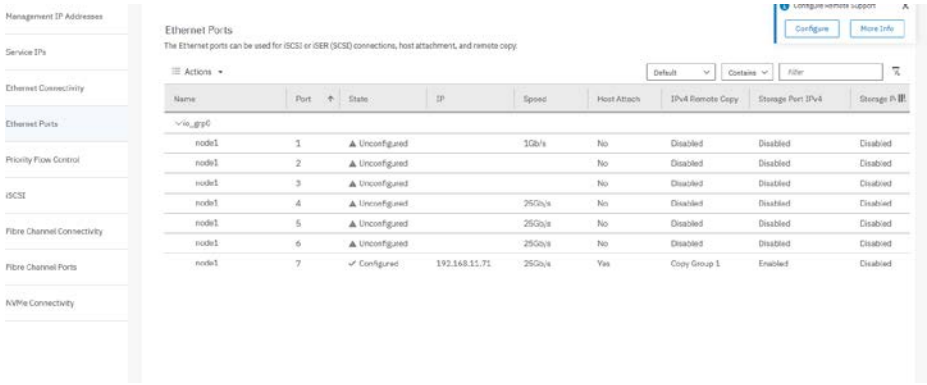

Figure 21. Select port and assign IP

2. Right-click the port for which you need to configure the IP and click **Modify IP Settings**.

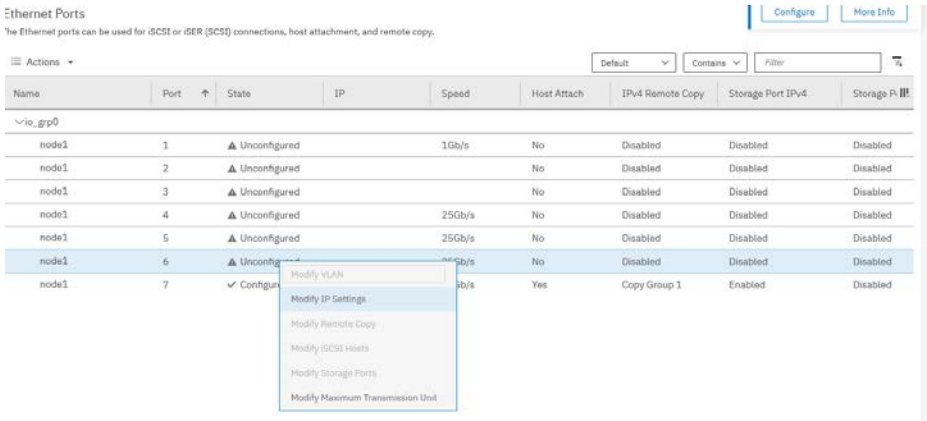

Figure 22. Modify IP settings

3. Specify the IP and click **Modify**.

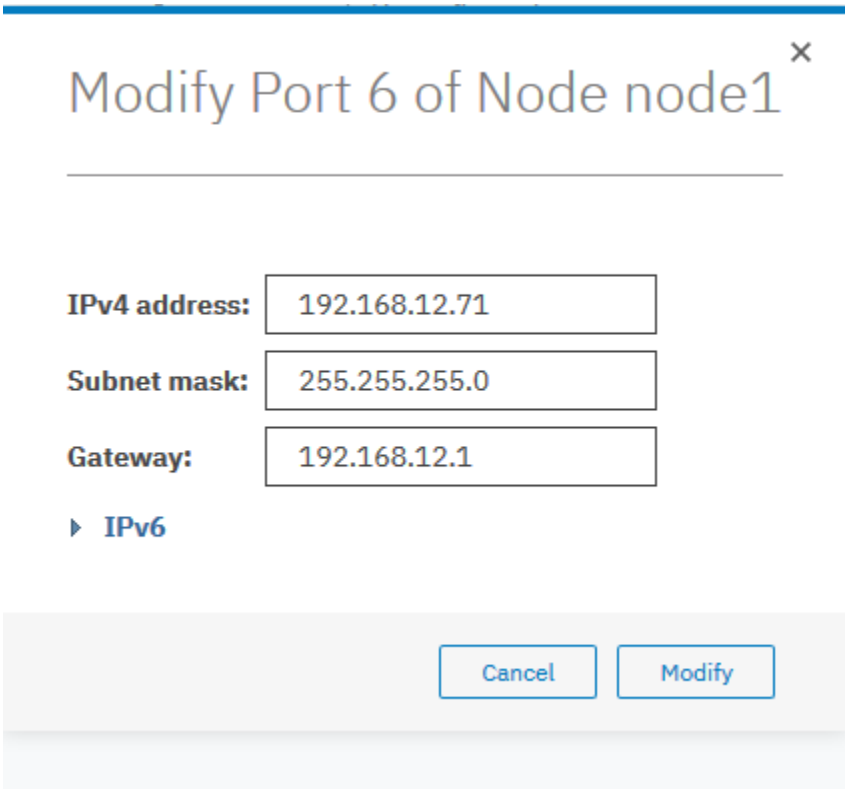

Figure 23. Set IP

4. To use this IP for storage connectivity, right-click the IP and click **Modify Storage ports**.

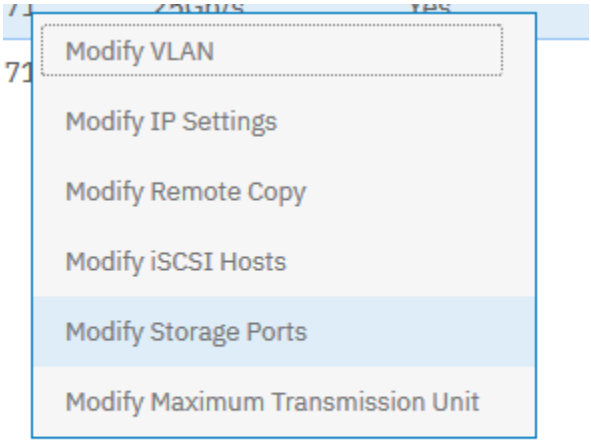

Figure 24. Set VLAN

5. If it is not enabled, select **Enabled** and click **Modify**.

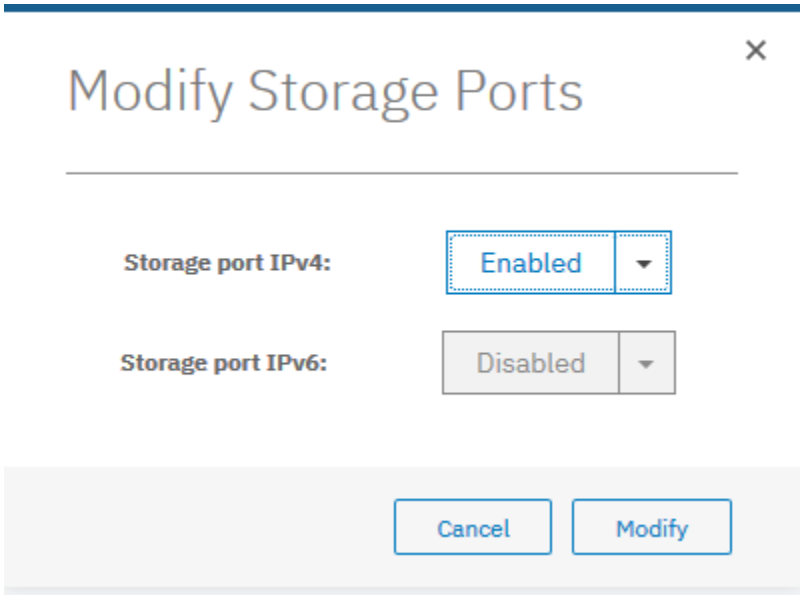

Figure 25. Modify storage ports

Assign VLAN on a port with an already configured IP

If IPs are already assigned, but VLAN is not enabled, then configure VLAN for the existing IPs using the `satask cfgportip` command.

Example CLI: Enabling VLAN 30 on port 7 which is already being used for storage connections.

```
svctask cfgportip -node node1 -vlan 30 7
```

Example GUI: Configure VLAN on an existing host attach IP or storage IP using GUI

1. In the IBM Spectrum Virtualize GUI, click **Settings → Network → Ethernet Ports**.

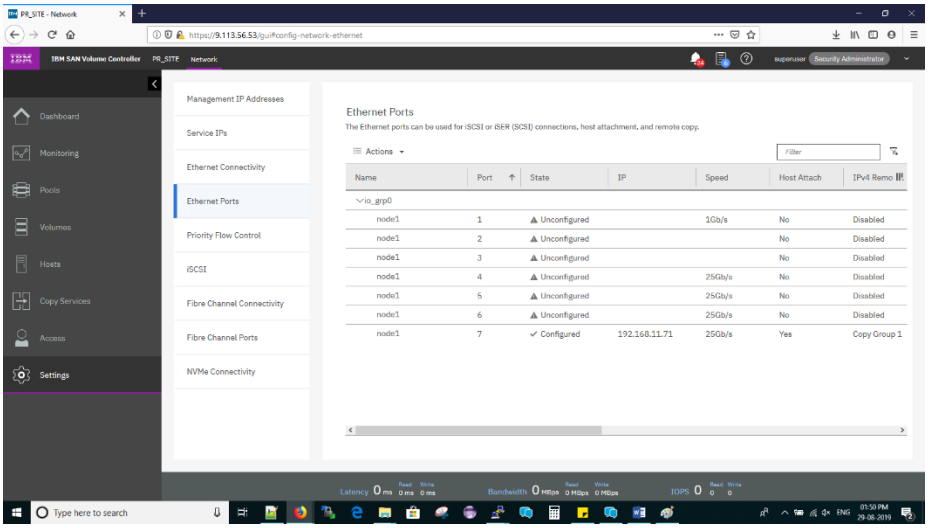

Figure 26. Select Ethernet port

2. Right-click the port and click **Modify VLAN**.

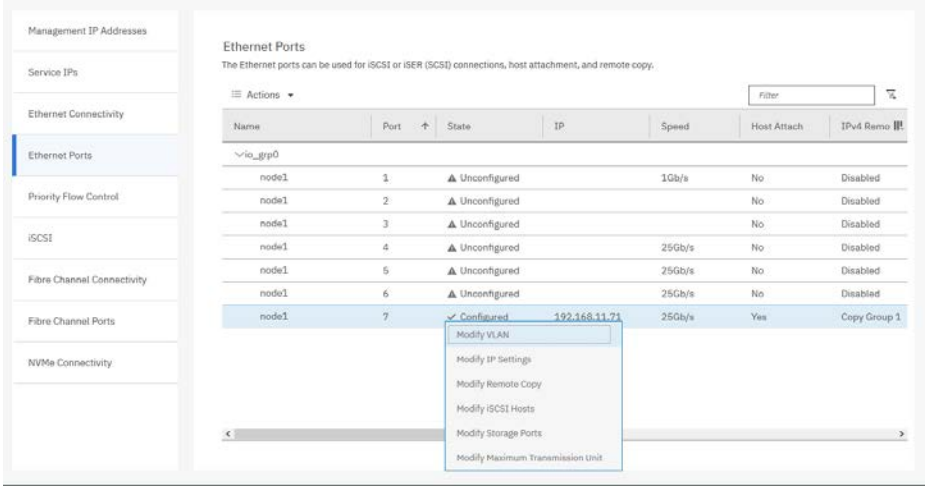

Figure 27: Modify VLAN

3. In the dialog box that is displayed, select the **Enable** check box.

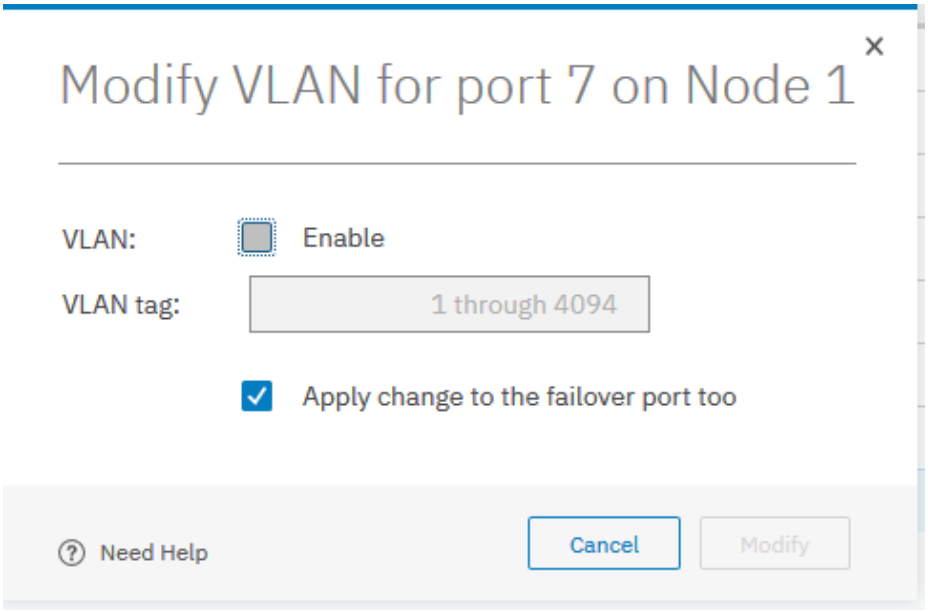

Figure 28. Specify VLAN ID

4. Specify the VLAN tag and click **Modify**.

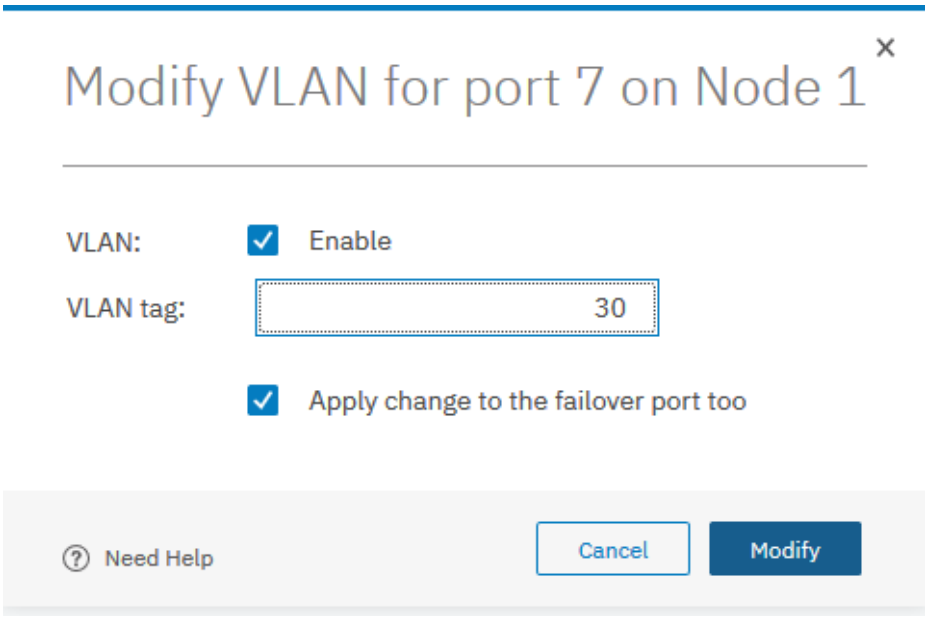

Figure 29. Apply settings

Things to remember:

- VLAN configuration is mandatory for PFC to function
- VLAN configuration must be used by all entities (such as nodes, backend nodes, hosts, and switches).
- Only DCBX capable switches can support the PFC configuration.
- DCBX enables switch and adapter ports to exchange parameters that describe traffic classes and PFC capabilities. For detailed steps, check your switch documentation.

Note that if you modify VLAN of an existing IP address then all connections through that IP will be lost.

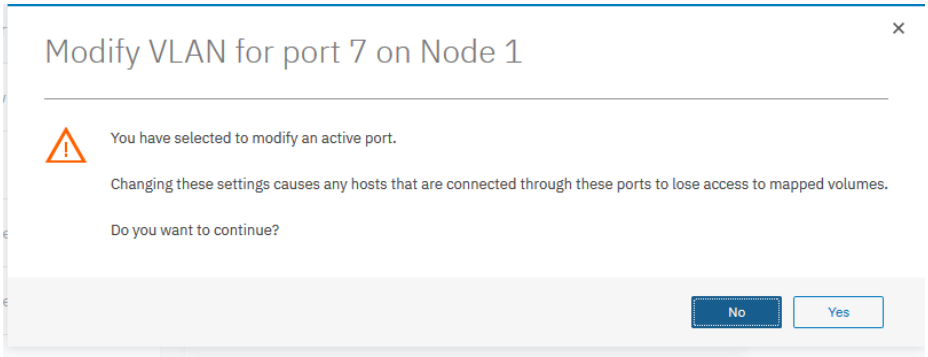

Figure 30. Update the settings

Assign COS value for storage traffic on SVC

After configuring the IBM Storwize ports with the correct VLAN tag, you can configure the PCP value or the COS value for storage traffic on SVC using the `svctask chsystemethernet` command.

CLI example:
`svctask chsystemethernet -storagecos 5`

GUI example: Defining priority tags for defined traffic classes using GUI

1. Click **Settings** → **Network** → **Priority Flow Control**.
2. Specify the priority tag values for storage traffic and click **Save**.

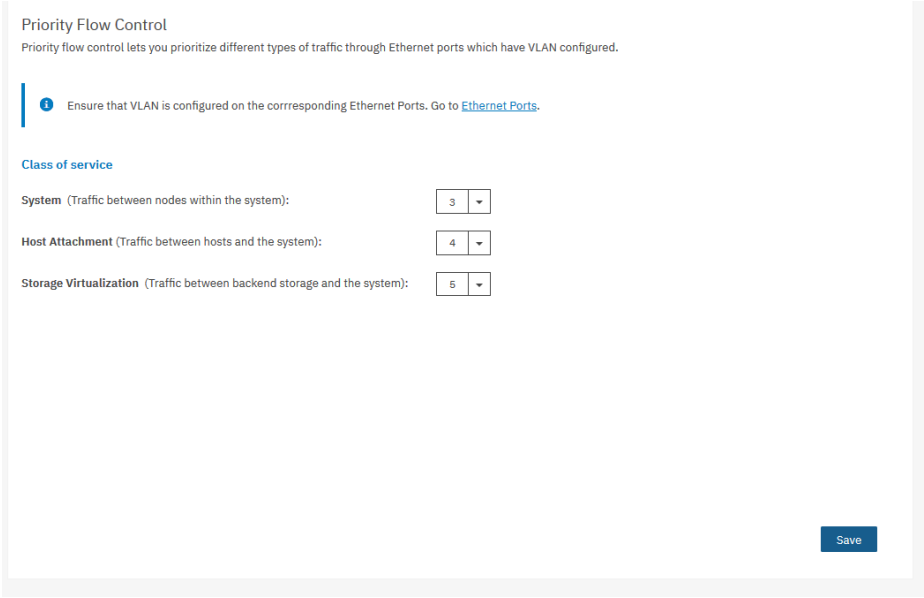

Figure 31. Set COS value

Note: Only single traffic class can also be assigned priority, keeping others as 0.

Validate and troubleshoot the configuration

Current limitations:

- A working VLAN configuration is mandatory.
- Rescanning and reestablishing iSER host attach sessions are required if any changes are made for PFC tag settings.
- Rescanning and reestablishing RDMA-based node-to-node traffic sessions are required if any changes are made for PFC tag settings.
- PFC will work only for Ethernet ports with 10G or more speed.

This section provides detailed steps to validate if PFC is working or not, along with the steps to fix issues (if any).

`svcinfo lsportip <port id>` lists a detailed view of the overall configuration of the specified port.

You need to consider the following parameters for PFC configuration:

`dcbx_state`

This is the most basic parameter that you should check before configuring or validating PFC. This should be *enabled*. If it is disabled, then check the switch settings. It should have DCBX and Link Layer Discovery Protocol (LLDP) enabled. If it is *unsupported*, then that means that the port does not accept PFC configuration from the switch port, even if the switch is DCBX-capable. All ports that are less than 10 Gbps have this value.

Lossless parameters for all traffic classes

Lossless parameters indicate whether PFC is enabled on not on an Ethernet port. There are five fields:

- **`lossless_host_attach`** or **`lossless_host_attach6`**: Indicates whether PFC is enabled or not for host attach traffic on that port.
- **`lossless_system`**: Indicates whether PFC is enabled or not for system traffic on that port.
- **`lossless_storage`** or **`lossless_storage6`**: Indicates whether PFC is enabled or not for storage traffic on that port.

`pfc_enabled_tags`

This field displays a list of priority tags that are enabled on the switch. If no priority tags are enabled on the switch, this field is blank. If priority tags are enabled on the switch but PFC is not enabled for those priority tags, this field is blank.

Frequently asked questions (FAQs)

How a correct configuration for host attach look like?

If all the settings of PFC are correctly done for an IP address, then `lossless_hostattach` (or `lossless_hostattach6` for IPv6) field will be

displayed. Refer to one such following example from GUI:

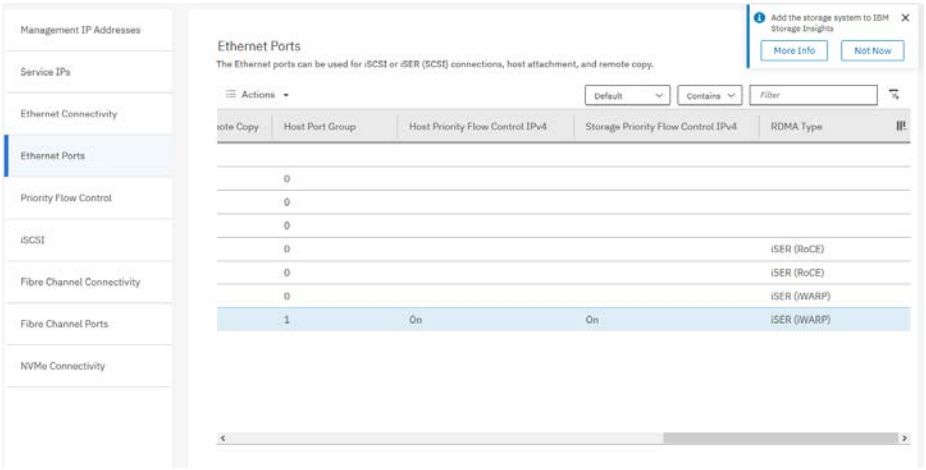

Figure 32. Displaying PFC as ON for host attach and storage traffic

From CLI, it can be checked using the `lsportip` command:

```
bash# lsportip 3 | grep lossless_host_attach
lossless_host_attach on
lossless_host_attach6
lossless_host_attach
lossless_host_attach6
lossless_host_attach on
lossless_host_attach6
lossless_host_attach
lossless_host_attach6
```

How to check and make priority flow control work for host attach traffic?

Refer to the flow chart in the following figure to check if PFC is enabled for host attach traffic for an IPv4 address. This flow chart also guides you through the steps to make priority flow work for that IPv4 address.

Note: For IPv6 address, check the `lossless_hostattach6` field instead of `lossless_hostattach` and follow the same flow chart.
Your switch must support iSCSI Type Length Value (TLV)

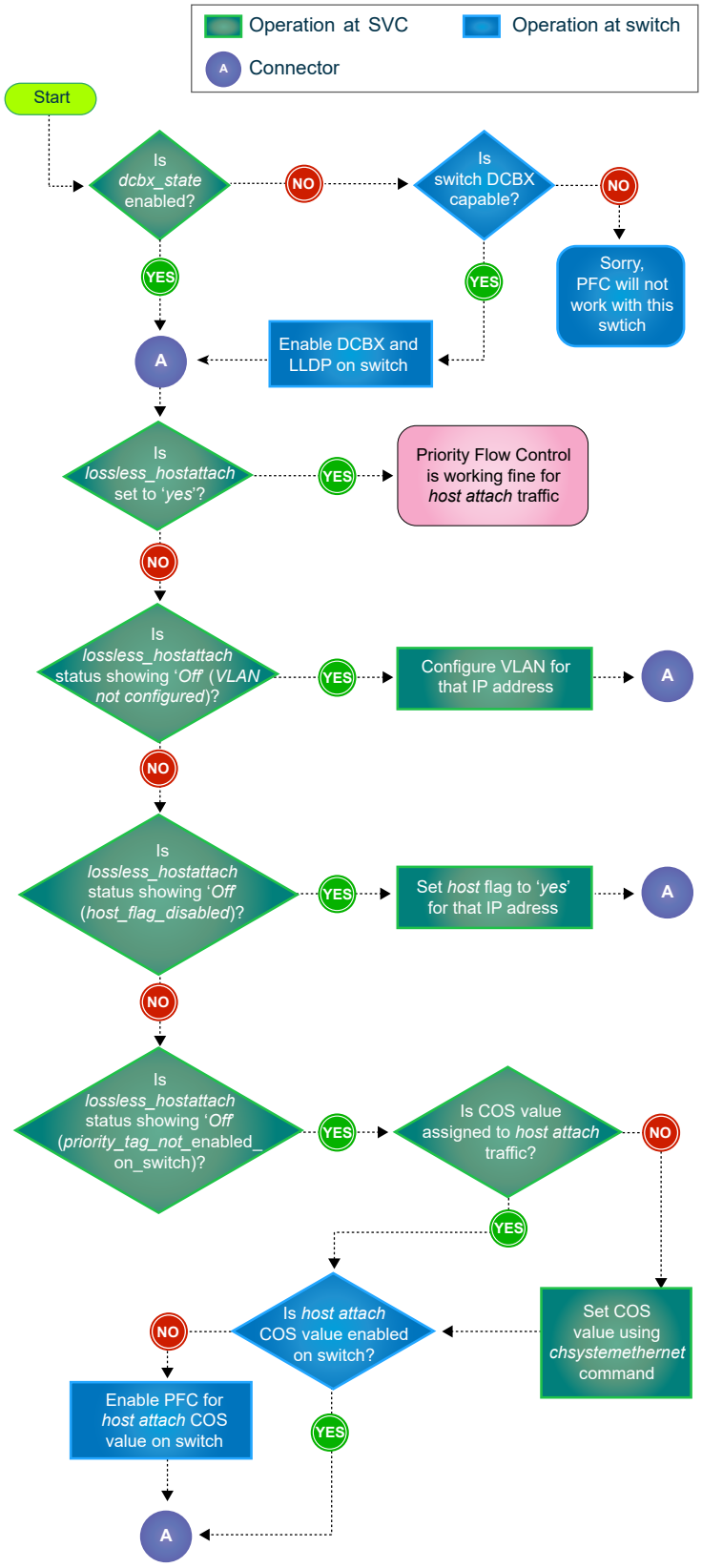

Figure 33: Steps for checking if PFC for host attach and storage traffic is enabled

Figure 34 shows that PFC status as **Off** and **VLAN is not configured** for host attach traffic. Follow the process in the flow chart in figure 33 to correct it.

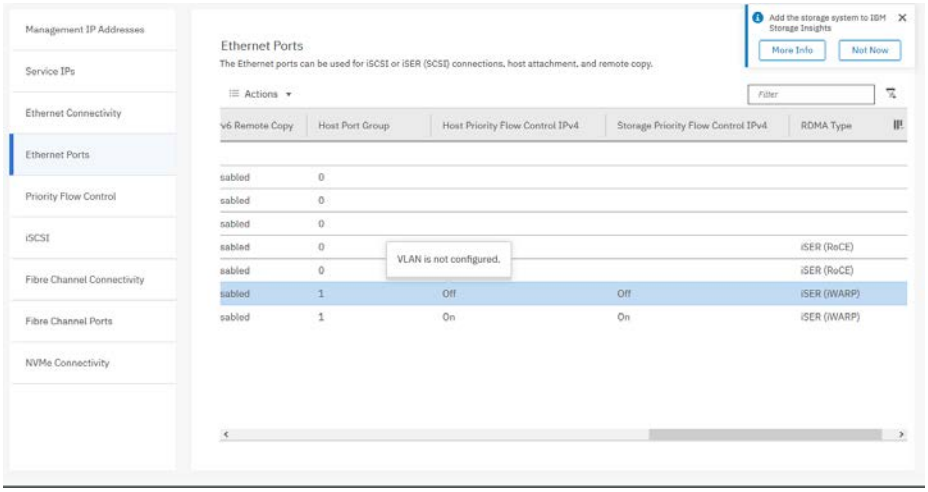

Figure 34: Displaying that PFC is ON for host attach and storage traffic

How a correct configuration for storage looks like?

If all the settings of priority flow control are correctly done for an IP address, then the `lossless_storage` (or `lossless_storage6`) field will be dispalyed. Refer Figure 34 for one such example from GUI:

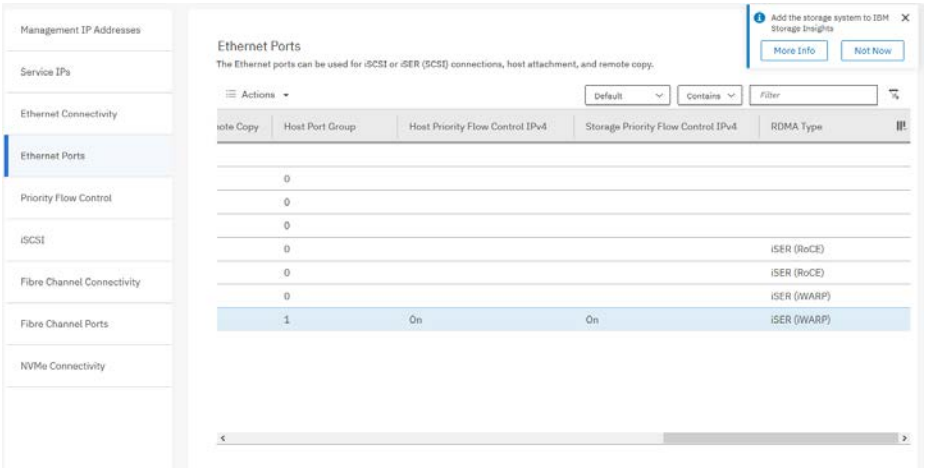

Figure 35. PFC is ON for host attach and storage traffic

From CLI, it can be checked using `lsportip`:

```
bash# lsportip 3 | grep lossless_storage
lossless_storage on
lossless_storage6
lossless_storage
lossless_storage6
lossless_storage on
lossless_storage6
lossless_storage
lossless_storage6
```

Figure 36 shows that the PFC status is **Off** and **VLAN is not configured** for the host attach traffic. Follow the process in the flow chart to correct it.

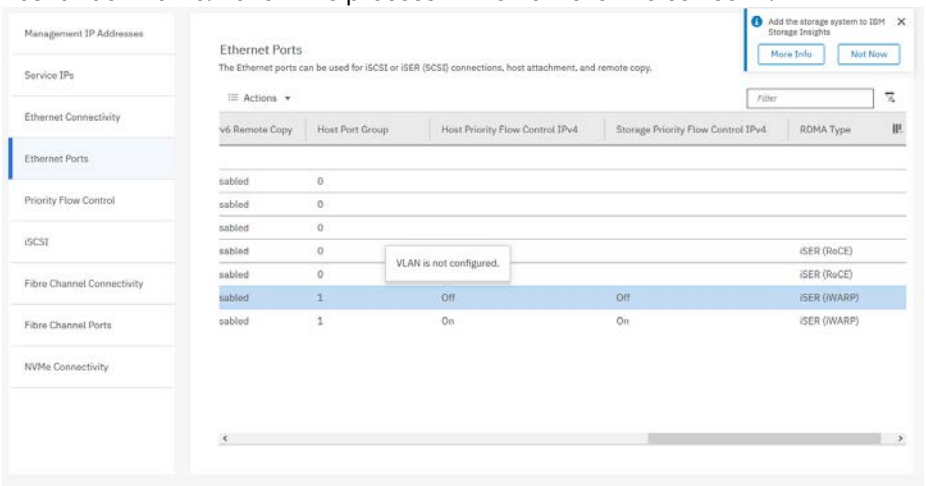

Figure 36. PFC is Off for both host and storage traffic

How to check and make PFC work for storage traffic?

You must follow the steps shown in the flow chart (in Figure 36) to check if PFC is enabled for storage traffic for an IPv4 address. This flow chart also guides you through the steps to make priority flow work for that IPv4 address.

Note: For the IPv6 address, check the `lossless_storage6` field instead of `lossless_storage` and follow the steps in the flow chart.

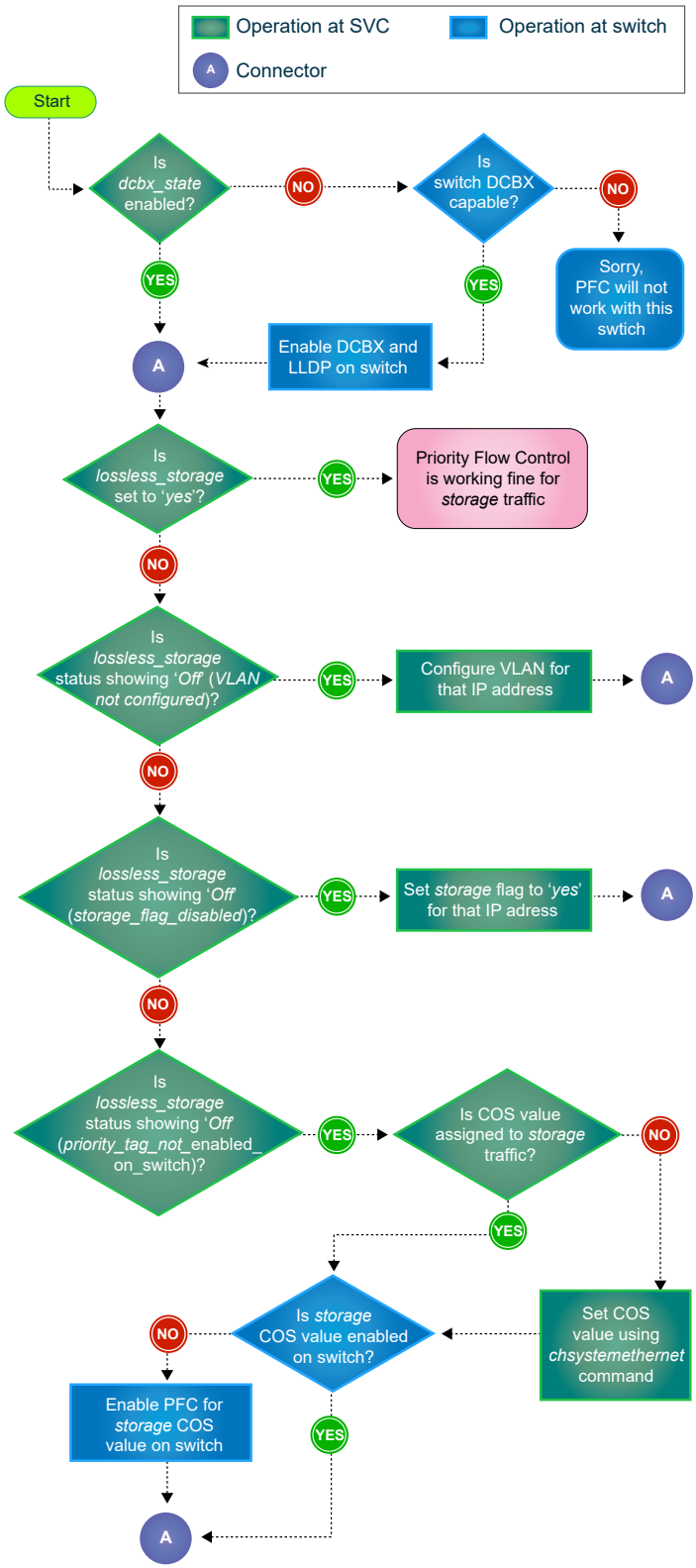

Figure 37. Steps to check if PFC for host attach and storage traffic is enabled

How a correct configuration for clustering traffic looks like?

If all the settings of PFC are correctly done for an IP address then the `lossless_system` field will be displayed. From the GUI, validate that PFC is enabled on the clustering ports.

Ethernet Ports

The Ethernet ports can be used for iSCSI or iSER (SCSI) connections, host attachment, and remote copy.

Actions

Name	Port	Speed	System Priority Flow Control	RDMA Type
io_grp0				
node1	1	1Gb/s		
node1	2			
node1	3			
node1	4	25Gb/s		iSER (RoCE)
node1	5	25Gb/s		iSER (RoCE)
node1	6	25Gb/s	On	iSER (iWARP)
node1	7	25Gb/s	On	iSER (iWARP)

Figure 38. PFC is On for system or clustering ports

From the CLI, it can be checked using `lsportip`:

```
# lsportip 5 |grep lossless_system
lossless_system on
lossless_system on
```

How to check and make PFC work for clustering traffic?

Follow the steps in the flow chart (shown in Figure 38) to check if priority flow control is enabled for clustering (system) traffic for an IPv4 or IPv6 address. This flow chart also guides you through the steps to make priority flow work for that IP address.

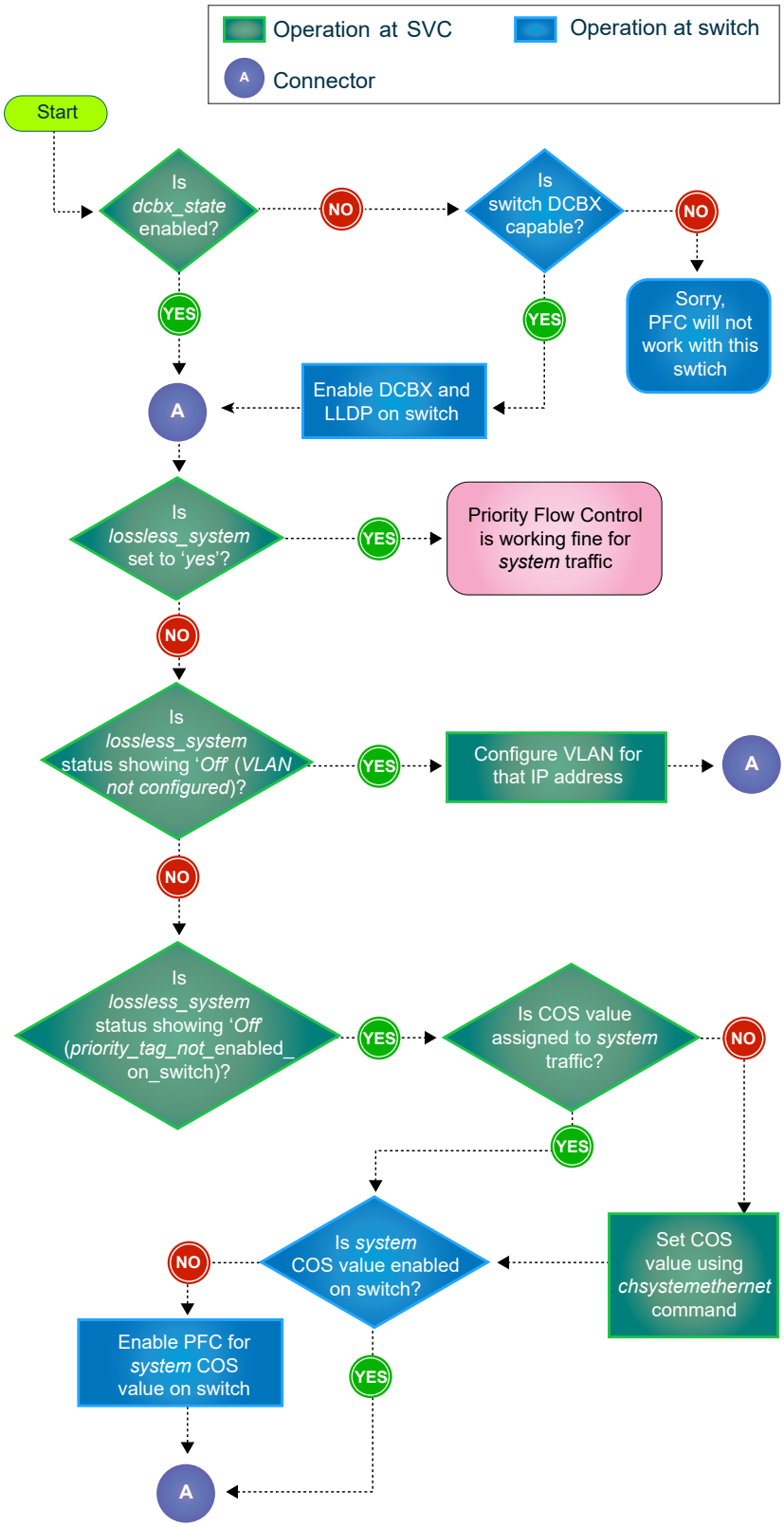

Figure 39. Steps to check if PFC for system or clustering traffic is enabled

Summary

This paper helps in configuring and managing different Ethernet traffic of IBM Spectrum Virtualize. It contains separate sections for each traffic type explaining the configuring steps in detail.

Also, it provides a section of validation and troubleshooting that can help in identifying a problem faced during the configuration with flow charts for better understanding.

Get more information

To learn more about IBM Spectrum Virtualize, refer to the following websites:
IBM Storwize V7000 Support page
ibm.com/storage/support/storwize/v7000
IBM Storwize V7000 Knowledge Center
ibm.com/support/knowledgecenter/ST3FR7
iSCSI Implementation and Best Practices on IBM Storwize Storage Systems
https://www.redbooks.ibm.com/redbooks/pdfs/sg248327.pdf

About the authors

Aakanksha Mathur leads the test effort for IBM Spectrum Virtualize product chain at IBM System Labs, India. She has been working in the storage domain for more than 12 years including 9 years with IBM System Labs. She can be reached at aamathur@in.ibm.com.

Bharti Soni is a senior system staff software engineer in the SVC development team at IBM Systems Labs, India. She holds a master's degree in Computer Science from IIT Roorkee. You can reach Bharti at: bharsoni@in.ibm.com.

IBM Systems
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of America