HOL-0709-02 - Enterprise SONiC - Building a layer 3 fabric with BGP

Dell Technologies DEMO CENTER

Table of Contents

Introduction

This document is intended for technical professionals whom want to learn about the Enterprise SONiC Distribution by Dell Technologies network operating system.

Note: The virtual lab environment runs the Enterprise SONiC operating system. As this is a virtual lab, it simulates physical hardware, therefore the performance may be slower than labs that have a physical infrastructure. Some elements of the lab may take longer to complete than would normally be the case in a physical environment.

Software for Open Networking in the Cloud (SONiC)

SONiC is an open-source network operating system based on Debian Linux that runs on switches from multiple vendors. SONiC offers a full feature set that has been in use in the data centers of some of the largest cloud service providers. Since it is open-source, SONiC helps users leverage the collective strength and take advantage of the vast ecosystem and community.

Enterprise SONiC Distribution by Dell Technologies

Enterprise SONiC Distribution by Dell Technologies (also known as Enterprise SONiC) is an enhancement of the SONiC community version. Enterprise SONiC is based on the open-source SONiC OS, which includes feature enhancements, hardening, and support targeted for the demanding data center leaf and spine fabrics while maintaining the open-source spirit.

Enterprise SONiC provides the benefits of our partnering with the SONiC community on a highly scalable open-source network operating system (NOS). SONiC is based on a modern containerized architecture that accelerates software innovation and evolution. The customers will also benefit from Dell Technologies enhancements, including hardening and validation of enterprise-ready data center use cases and support on select Dell EMC's PowerSwitch family of industry-leading open networking switches.

The Enterprise SONiC operating system supports the following system models:

- Dell EMC PowerSwitch N3248TE -ON
- Dell EMC PowerSwitch S5112F-ON • Dell EMC PowerSwitch S5224F-ON • Dell EMC PowerSwitch S5232F-ON • Dell EMC PowerSwitch S5248F-ON
- Dell EMC PowerSwitch S5296F-ON
- Dell EMC PowerSwitch Z9264F-ON
- Dell EMC PowerSwitch Z9332F-ON
- Dell EMC PowerSwitch Z9100-ON

Dell Technologies has released a virtual appliance version of the Enterprise SONiC network operating system for the Dell EMC PowerSwitch S5248F-ON. Customers or other interested people can use this virtual machine to simulate, experiment with and test various network topologies. It enables the provisioning of a robust and

proven network operating system across production and development platforms with a uniform Enterprise SONiC distribution and single-image consistency. The Enterprise vSONiC appliance is used to test various network configurations and integration in a simulated environment.

Lab Overview

Lab guidance

Architecture

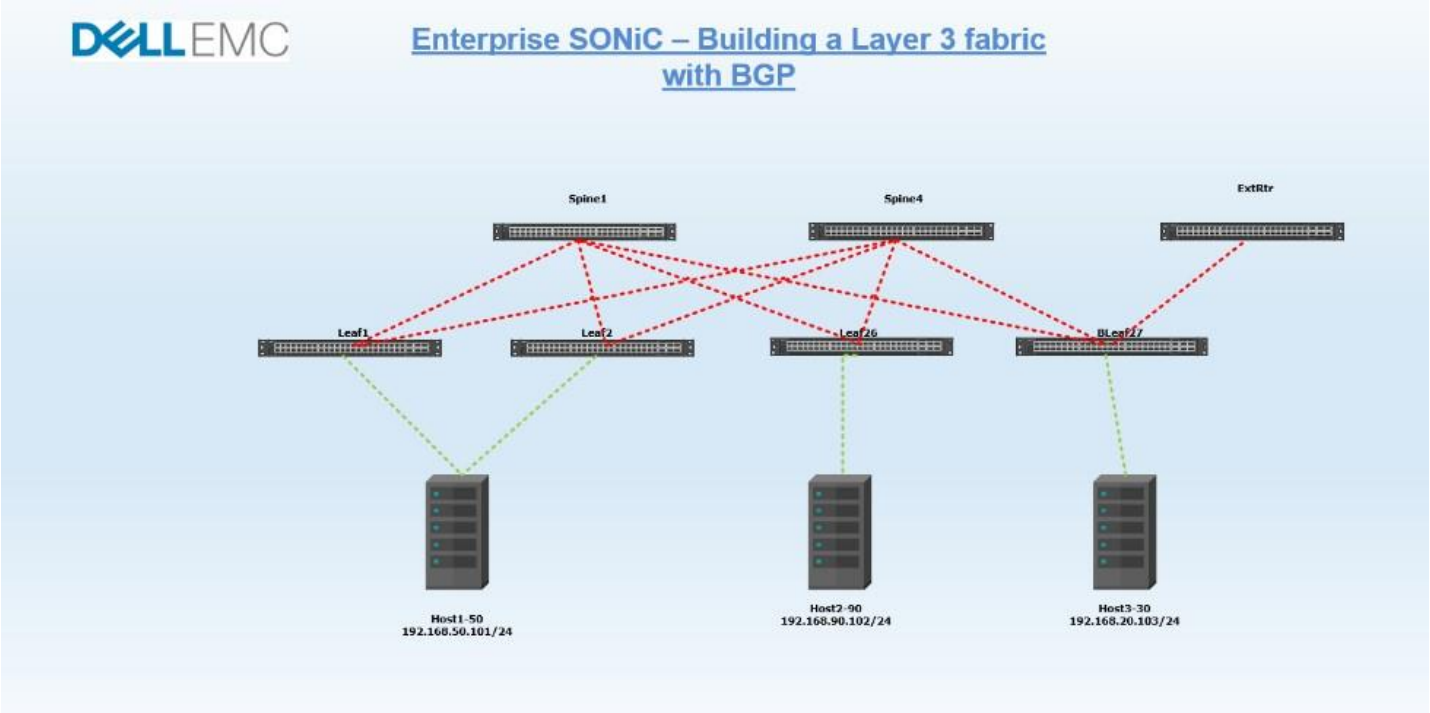
Building a layer 3 fabric with unnumbered BGP (40 minutes): Advanced - This module demonstrates a layer 3 fabric with unnumbered BGP.

This lab may take more than 40 minutes to complete.

Enterprise SONiC Distribution operating system

The Enterprise SONiC Distribution by Dell Technologies operating system enables dis-aggregation of the operating system from the underlying hardware. You will gain firsthand experience with the configuration of VLANs, assignment of IP addresses, and the basic Enterprise SONiC operating system commands. Using the Management Framework command-line interface (MF-CLI), you can configure leaf pairs with multi-chassis link aggregation, a leaf-spine network using Border Gateway Protocol (BGP). This lab introduces you to the core capabilities of the Enterprise SONiC operating system and its key concepts.

Network Diagram



Lab Details

Switches/Host	Use	IP Address	Gateway	Credentials	Shutdown Procedure
Spine1	Spine	192.168.1.11/24	192.168.1.1	admin/admin123	sudo shutdown
Spine4	Spine	192.168.1.12/24	192.168.1.1	admin/admin123	sudo shutdown
Leaf1	Leaf	192.168.1.13/24	192.168.1.1	admin/admin123	sudo shutdown
Leaf2	Leaf	192.168.1.14/24	192.168.1.1	admin/admin123	sudo shutdown
Leaf26	Leaf	192.168.1.15/24	192.168.1.1	admin/admin123	sudo shutdown
BLeaf27	Border Leaf	192.168.1.16./24	192.168.1.1	admin/admin123	sudo shutdown
Host150	Server	N/A		admin/admin	sudo poweroff
Host290	Server	N/A		admin/admin	sudo poweroff
Host330	Server	N/A		admin/admin	sudo poweroff
ExtRtr	External Router	192.168.1.17/24	192.168.1.1	admin/admin123	N/A

Building a layer 3 fabric with BGP - module overview

Build a layer 3 fabric with BGP (40 Minutes) - Advanced - This module will demonstrate the use of Border Gateway Protocol (BGP) in the Enterprise SONiC networking operating system. BGP enables routing/reachability information of networks among autonomous systems and is classified as path vector protocol. There are benefits in enabling this protocol which include scaling, prefix filtering, traffic engineering, tagging, and stability when connected to other vendor networks.

Benefits Summary

BGP easily scales in large environments with thousands of switches; extensive traffic tagging and engineering capabilities; and is very stable between switches from different vendors. The use of BGP in the data center supports and exceeds the requirements when it is designed to run an overlay network in the data center.

In this lab you will:

- Configure a MCLAG leaf switch
- Configure unnumbered BGP in a L3 leaf-spine network • Verify the configuration

Lesson 1 - Building a Layer 3 fabric with unnumbered BGP

Lesson 1 Overview

In this lesson you will configure switches **Spine1**, **Leaf1**, **Leaf26** for a Layer 3 BGP fabric.

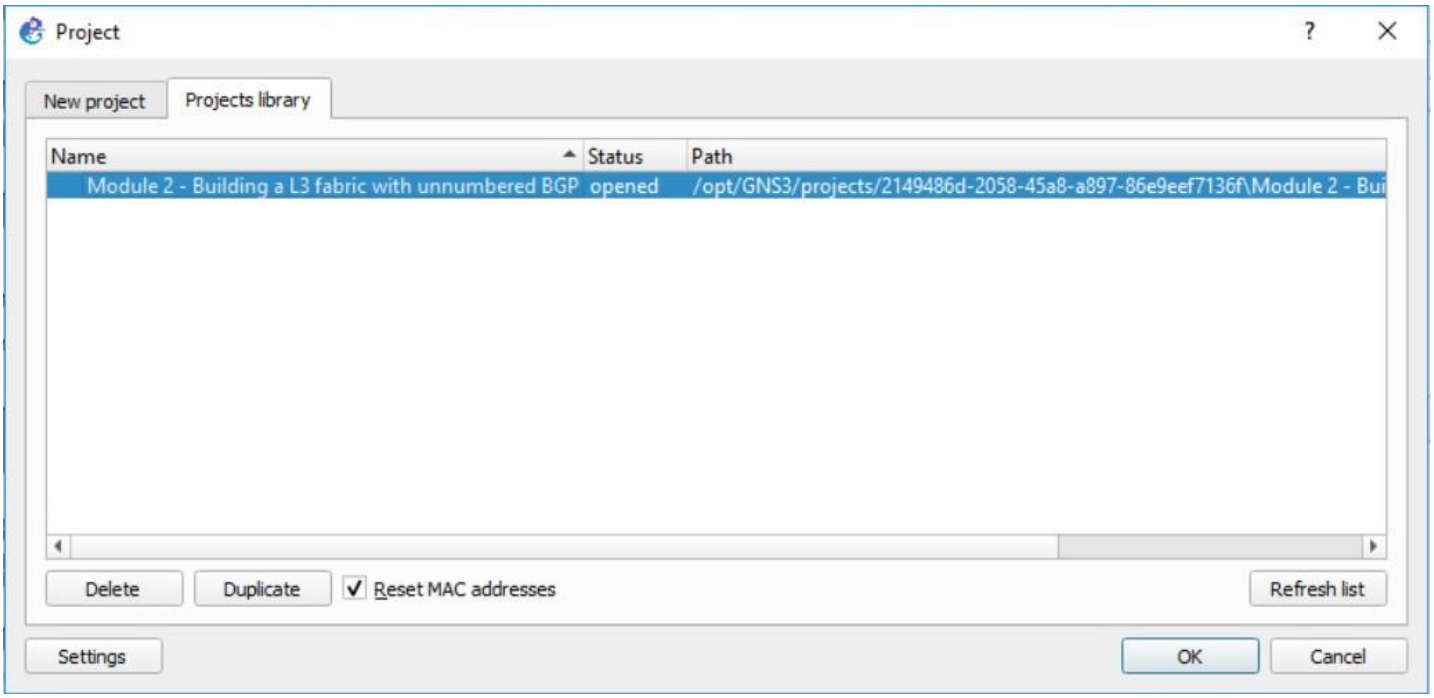
The remaining switches **Leaf2**, **BLeaf27**, **Spine4** and **ExtRtr** are already configured complementing what is in this lesson.

The module begins with launching the GNS3 network emulator.

Open GNS3 application

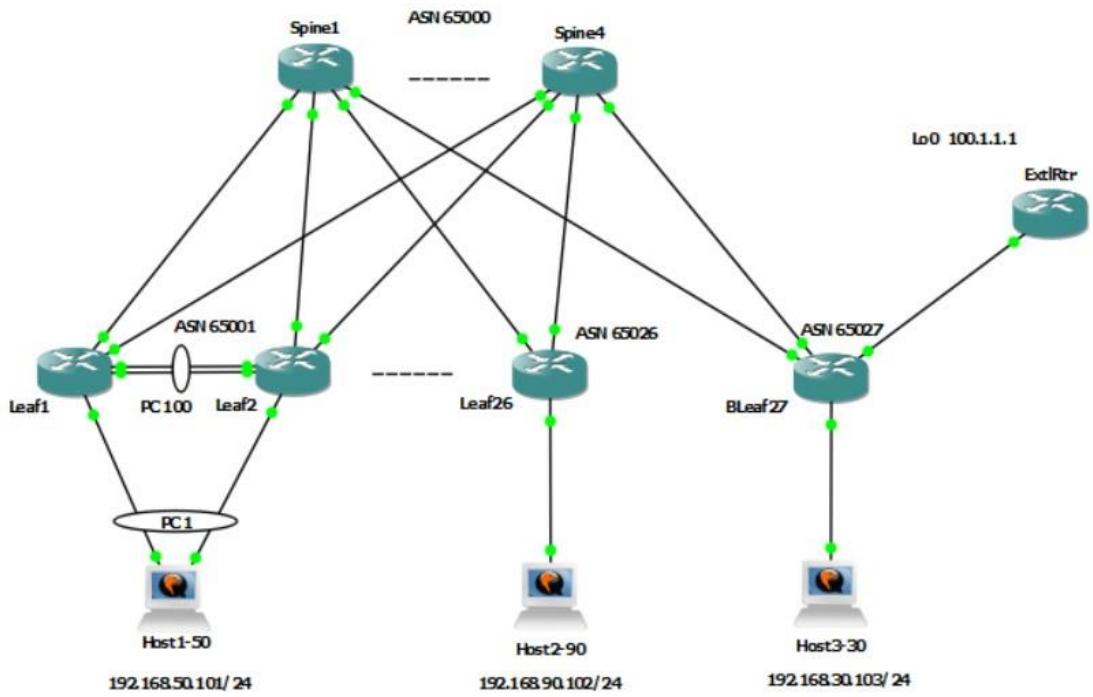
To open the GNS3 application, double click the GNS3 icon on the desktop.

When the **Project** window opens, select **Project library** tab, click the **Building a L3 fabric with unnumbered BGP** entry and click **OK**.



GNS3 Topology view

The topology should come up with all connections green. If not, please **END** the lab and try again.

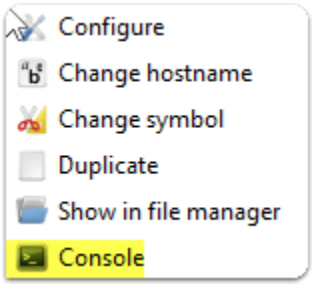


Note: Names for host devices include the VLAN ID as part of the name.

Configure Spine1

Right-click on node **Spine1.**

Select **Console.**



Spine1 login

On switch **Spine1** please log in with the following credentials:

- login: **admin**
- password: **admin**

Enter the Management Framework shell.

```
sonic-cli
```

Configure interface naming mode

Enable Standard interface-naming mode and define hostname.

```
configure terminal
interface-naming standard
hostname Spine1
end
write memory
exit
```

Wait for the **Interface naming mode** message.

Exit the Management Framework CLI session and re-enter.

```
sonic-cli
```

Configure interfaces connecting to leaf switches

Configure IPv6 link-local address on the interfaces connected to leaf switches. On switch **Spine1**, please enter the following commands:

```
configure terminal
interface Eth 1/1
description Leaf1
ipv6 enable
no shutdown
exit

interface Eth 1/2
description Leaf2
ipv6 enable
```

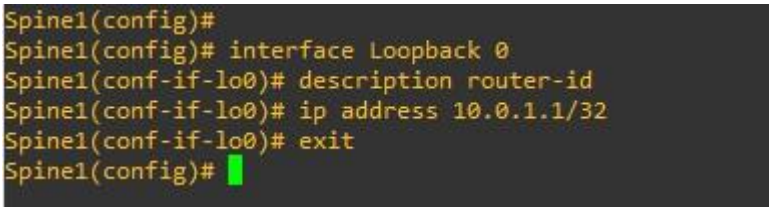
```
no shutdown
exit

interface Eth 1/3
description Leaf26
ipv6 enable
no shutdown
exit

interface Eth 1/4
description BLeaf27
ipv6 enable
no shutdown
exit
```

```
Spine1# configure terminal
Spine1(config)# interface Eth 1/1
Spine1(conf-if-Eth1/1)# description Leaf1
Spine1(conf-if-Eth1/1)# ipv6 enable
Spine1(conf-if-Eth1/1)# no shutdown
Spine1(conf-if-Eth1/1)# exit
Spine1(config)# interface Eth 1/2
Spine1(conf-if-Eth1/2)# description Leaf2
Spine1(conf-if-Eth1/2)# ipv6 enable
Spine1(conf-if-Eth1/2)# no shutdown
Spine1(conf-if-Eth1/2)# exit
Spine1(config)#
Spine1(config)# interface Eth 1/3
Spine1(conf-if-Eth1/3)# description Leaf26
Spine1(conf-if-Eth1/3)# ipv6 enable
Spine1(conf-if-Eth1/3)# no shutdown
Spine1(conf-if-Eth1/3)# exit
Spine1(config)# interface Eth 1/4
Spine1(conf-if-Eth1/4)# description BLeaf27
Spine1(conf-if-Eth1/4)# ipv6 enable
Spine1(conf-if-Eth1/4)# no shutdown
Spine1(conf-if-Eth1/4)# exit
Spine1(config)#
```

Configure Loopback interface

On switch **Spine1** , enter the following commands to configure the router-ID loopback:

```
interface Loopback 0
description router-id
ip address 10.0.1.1/32
exit
```



Configure BGP

Configure unnumbered BGP using peer-groups. The **remote-as external** and **capability extendednexthop** enable unnumbered BGP. On **Spine1**, please enter the following commands:

```
router bgp 65000
router-id 10.0.1.1
bestpath as-path multipath-relax

 peer-group LEAF
  advertisement-interval 5
  remote-as external
  capability extended-nexthop
  address-family ipv4 unicast
  activate
  exit
 exit

 address-family ipv4 unicast
 redistribute connected
 maximum-paths 2
 exit

neighbor interface Eth 1/1
 peer-group LEAF
 exit

neighbor interface Eth 1/2
 peer-group LEAF
 exit

neighbor interface Eth 1/3
 peer-group LEAF
 exit

neighbor interface Eth 1/4
 peer-group LEAF
 exit

end
```

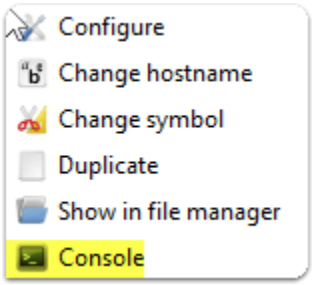
```
write memory
```

Note: BFD and aggressive timers are removed to reduce load on the GNS3 server. These will be typically enabled in a production environment.

Configure Leaf1

Right-click on node **Leaf1.**

Select **Console.**



Leaf1 login

On switch **Leaf1** please log in with the following credentials:

- login: **admin**
- password: **admin**

Enter the Management Framework shell.

```
sonic-cli
```

Enable Standard interface naming mode

Enable the Standard interface naming mode and the hostname.

```
configure terminal
interface-naming standard
hostname Leaf1
end
write memory
exit
```

Wait for the **Interface naming mode** message. Exit the Management Framework CLI session and reenter.

```
sonic-cli
```

Configure upstream interfaces

Configure IPv6 link-local address on the interfaces connected to spine switches. On switch **Leaf1**, please enter the following commands:

```
configure terminal
interface Eth 1/8
description Spine1
ipv6 enable
no shutdown
exit

interface Eth 1/9
description Spine4
ipv6 enable
no shutdown
exit
```

Configure loopback interface

On switch **Leaf1** , configure **router-id** loopback. Each leaf switch will use IP address 10.0.2.X, where X is the leaf number.

```
interface Loopback 0
description router-id
ip address 10.0.2.1/32
exit
```

Configure MCLAG PortChannel

Configure the MCLAG peer-link interface **PortChannel 100** and MCLAG peer-link members.

```
interface Portchannel 100 mode on
description MCLAG_Peer_Link
exit

interface Eth 1/5
description MCLAG_Peer_Link
channel-group 100
no shutdown
exit

interface Eth 1/6
description MCLAG_Peer_Link
channel-group 100
no shutdown
exit
```

Note: Static port channels are used in GNS3 to reduce the GNS3 server CPU load.

Configure MCLAG

Configure MCLAG domain 1 using **Leaf1** and **Leaf2** loopback address as the keep-alive link and **PortChannel100** as the peer-link. Set the MCLAG **delay-restore** timer to 90 seconds.

```
mclag domain 1
source-ip 10.0.2.1
peer-ip 10.0.2.2
peer-link PortChannel 100
delay-restore 90
exit
```

Enable link state tracking to support the MCLAG **delay-restore** timer.

```
link state track MCLAG
downstream all-mclag
```

Configure Host VLAN

Assign **Vlan 50** with IP address of 192.168.50.254 as the host gateway. Assign a unique secondary IP address if desired. On switch **Leaf1** enter the following commands:

```
interface vlan 50
ip address 192.168.50.254/24
ip address 192.168.50.253/24 secondary
exit
```

Note: MCLAG does not yet support VRRP. Use a common IP address for both MCLAG Peers to support the host gateway address.

Configure port channel interface to **Host1** and add **Vlan 50** and assign to MCLAG domain**.** Assign **Vlan 50** to **PortChannel 100** to allow host traffic failover to MCLAG peer-links.

```
interface PortChannel 1 mode on
description Host1-portchannel
switchport trunk allowed vlan add 50
mclag 1
exit

interface Eth 1/1
description Host1_Portchannel
channel-group 1
no shutdown
exit

interface PortChannel 100
switchport trunk allowed vlan add 50
exit
```

Configure route filters

Configure route filters to advertise the Host networks and router-id using prefix list by entering the following commands:

```
ip prefix-list CONN seq 10 permit 10.0.2.0/24 ge 32
ip prefix-list CONN seq 20 permit 192.168.0.0/16 ge 24

route-map REDIST permit 10
match ip address prefix-list CONN
exit
```

Configure Unnumbered BGP

Configure unnumbered BGP using peer-groups. On switch **Leaf1**, please enter the following commands:

```
router bgp 65001
 router-id 10.0.2.1
 bestpath as-path multipath-relax

 peer-group SPINE
  advertisement-interval 5
  remote-as external
  capability extended-nexthop
  address-family ipv4 unicast
  activate
  allowas-in 2
  exit
 exit

 address-family ipv4 unicast
 redistribute connected route-map REDIST
 maximum-paths 4
 exit

neighbor interface Eth 1/8
 peer-group SPINE
exit

neighbor interface Eth 1/9
 peer-group SPINE
exit
!
end
write mem
```

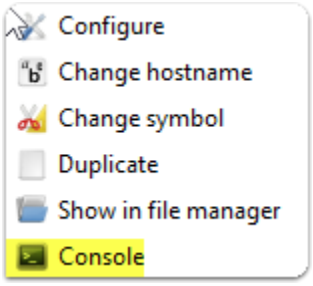
Note: The BGP parameter **allowas-in** is needed to allow MCLAG keep-alive messages to originate and terminate from same ASN of Leaf1 and Leaf2 over the spine network.

Leaf1 configuration is complete.

Configure Leaf26

Right-click on node **Leaf26.**

Select **Console.**



Leaf26 Login

On switch **Leaf26,** please log in with the following credentials:

- login: **admin**
- password: **admin**

Enter the Management Framework CLI shell.

```
sonic-cli
```

Note: Leaf26 is configured as a single-homed leaf that may be applicable to some fabrics.

Enable Standard interface mode

Enable standard interface naming mode and hostname.

```
configure terminal
interface-naming standard
hostname Leaf26
end
write memory
exit
```

Wait for **Interface naming mode** message, then exit the Management Framework CLI session and reenter.

```
sonic-cli
```

Configure Upstream Interfaces

Configure IPv6 link-local address on the interfaces connected to spine switches. On switch **Leaf26**, please enter the following commands:

```
configure terminal
interface Eth 1/8
```

```
description Spine1
ipv6 enable
no shutdown
exit

interface Eth 1/9
description Spine4
ipv6 enable
no shutdown
exit
```

Configure Router-ID Loopback

On switch **Leaf26** , configure the router-id loopback. Each leaf switch will use IP address 10.0.2.X, where X is the leaf number.

```
interface Loopback 0
description router-id
ip address 10.0.2.26/32
exit
```

Configure Host VLAN and Interface

Assign **Vlan 90** with IP address. Assign **Host2** interface (Eth1/1) . On switch **Leaf26** enter the following commands:

```
interface vlan 90
ip address 192.168.90.254/24
exit

interface Eth 1/1
switchport trunk allowed Vlan add 90
no shutdown
exit
```

Configure route filters

Configure prefix list and route map to filter BGP route advertisements.

```
ip prefix-list CONN seq 10 permit 10.0.2.0/24 ge 32
ip prefix-list CONN seq 20 permit 192.168.0.0/16 ge 24
```

```
route-map REDIST permit 10
match ip address prefix-list CONN
exit
```

Configure Unnumbered BGP

Configure unnumbered BGP using peer-groups. On switch **Leaf26**, please enter the following commands:

```
router bgp 65026
router-id 10.0.2.26
bestpath as-path multipath-relax
peer-group SPINE
 advertisement-interval 5
 remote-as external
 capability extended-nexthop
 address-family ipv4 unicast
 activate
 exit
 exit

 address-family ipv4 unicast
 redistribute connected route-map REDIST
 maximum-paths 4
 exit

neighbor interface Eth 1/8
 peer-group SPINE
exit

neighbor interface Eth 1/9
 peer-group SPINE
exit
!
end
write memory
```

Leaf26 configuration is complete.

Verification

Verification - Check BGP sessions

On switch **Spine1** please log in with the following credentials:

- login: **admin**

- password: **admin**

Enter the Management Framework CLI shell.

To verify the configuration and validate BG operation, please run the following command on switch **Spine1.** All neighbors should be established and show 2 -4 prefixes.

```
show bgp ipv4 unicast summary
```

```
admin@Spine1:~$ sonic-cli
Spine1# show bgp ipv4 unicast summary
BGP router identifier 10.0.1.1, local AS number 65000
Neighbor    V    AS      MsgRcvd    MsgSent    InQ      OutQ     Up/Down          State/PfxRcd
Eth1/1      4    65001   42         37         0        0        00:29:36         2
Eth1/2      4    65001   42         37         0        0        00:29:36         2
Eth1/3      4    65026   40         37         0        0        00:29:29         2
Eth1/4      4    65027   42         37         0        0        00:29:34         4

Total number of neighbors 4
Total number of neighbors established 4
Spine1#
```

Additional information about the configuration and neighbors can be viewed with the following command.

```
show bgp ipv4 unicast neighbors
```

```
Spine1# show bgp ipv4 unicast neighbors

BGP neighbor is Eth1/1, remote AS 65001, local AS 65000, external link
  BGP version 4, remote router ID 10.0.2.1 , local router ID 10.0.1.1
  BGP state = Established, up for 00:38:03
  Last read 00:00:00, Last write 00:00:53
  Hold time is 180 seconds, keepalive interval is 60 seconds
  Minimum time between advertisement runs is 5 seconds
  Neighbor capabilities:
    4 Byte AS: advertised and received
    AddPath: advertised and received
    Route refresh: advertised and received
    Multiprotocol Extension: advertised and received
    Graceful restart: advertised and received
  Message statistics:
    InQ depth is 0
    OutQ depth is 0
                         Sent       Rcvd
    Opens:               1          1
    Notifications:       0          0
    Updates:             6          11
    Keepalive:           38         39
    Route Refresh:       0          0
    Capability:          0          0
    Total:               45         51
--more--
```

Verification - Check BGP Routes

To verify the configuration and check the route database in routers please run the following command on **Leaf1**, **Leaf2**, **Leaf26**, **BLeaf27.** With ECMP enabled, there should be 2 paths listed per leaf network.

```
show ip route bgp
```

```
Leaf1# show ip route bgp
Codes:  K - kernel route, C - connected, S - static, B - BGP, O - OSPF
        > - selected route, * - FIB route, q - queued route, r - rejected route, # - not installed in hardware
      Destination                Gateway                                                  Dist/Metric   Uptime
-------------------------------------------------------------------------------------------------------------
 B>*   0.0.0.0/0                 via fe80::eb7:d1ff:fe76:4100  Eth1/9                    20/0          00:19:30
   *                             via fe80::eb7:d1ff:fe8b:d200  Eth1/8
 B>*   10.0.1.1/32               via fe80::eb7:d1ff:fe8b:d200  Eth1/8                    20/0          00:19:35
 B>*   10.0.1.4/32               via fe80::eb7:d1ff:fe76:4100  Eth1/9                    20/0          00:19:30
 B>*   10.0.2.2/32               via fe80::eb7:d1ff:fe76:4100  Eth1/9                    20/0          00:19:30
   *                             via fe80::eb7:d1ff:fe8b:d200  Eth1/8
 B>*   10.0.2.26/32              via fe80::eb7:d1ff:fe76:4100  Eth1/9                    20/0          00:10:30
   *                             via fe80::eb7:d1ff:fe8b:d200  Eth1/8
 B>*   10.0.2.27/32              via fe80::eb7:d1ff:fe76:4100  Eth1/9                    20/0          00:19:30
   *                             via fe80::eb7:d1ff:fe8b:d200  Eth1/8
 B>*   192.168.11.0/24           via fe80::eb7:d1ff:fe76:4100  Eth1/9                    20/0          00:19:30
   *                             via fe80::eb7:d1ff:fe8b:d200  Eth1/8
 B>*   192.168.30.0/24           via fe80::eb7:d1ff:fe76:4100  Eth1/9                    20/0          00:19:30
   *                             via fe80::eb7:d1ff:fe8b:d200  Eth1/8
 B>*   192.168.90.0/24           via fe80::eb7:d1ff:fe76:4100  Eth1/9                    20/0          00:10:30
   *                             via fe80::eb7:d1ff:fe8b:d200  Eth1/8
Leaf1#
```

Verify MCLAG

Verify **Leaf1** and **Leaf2** MCLAG operation with the following commands.

```
show mclag brief
show mclag interface 1 1
```

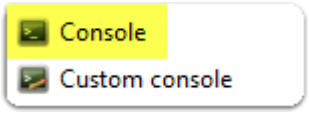
```
Leaf1# show mclag brief

Domain ID              : 1
Role                   : active
Session Status         : up
Peer Link Status       : up
Source Address         : 10.0.2.1
Peer Address           : 10.0.2.2
Peer Link              : PortChannel100
Keepalive Interval     : 1 secs
Session Timeout        : 30 secs
Delay Restore          : 90 secs
System Mac             : 0c:13:6f:93:b7:00
Mclag System Mac       :


Number of MLAG Interfaces:1
------------------------------------------------------------
MLAG Interface       Local/Remote Status
------------------------------------------------------------
PortChannel1               up/up
Leaf1#
```

```
Leaf1# show mclag interface 1 1

Local/Remote Status   : up/up
TrafficDisable        : No
IsolateWithPeerLink   : Yes
Leaf1#
```

Note: One leaf should be **active** and the other **standby**. Both **Session Status** and **Peer Link Status** should be **up**. The **standby** unit will use the MAC address of the **active** peer.

Verification - Ping from Host1 to Host2

Right-click on the **Host1** node and select **Console**



Log in to Host1

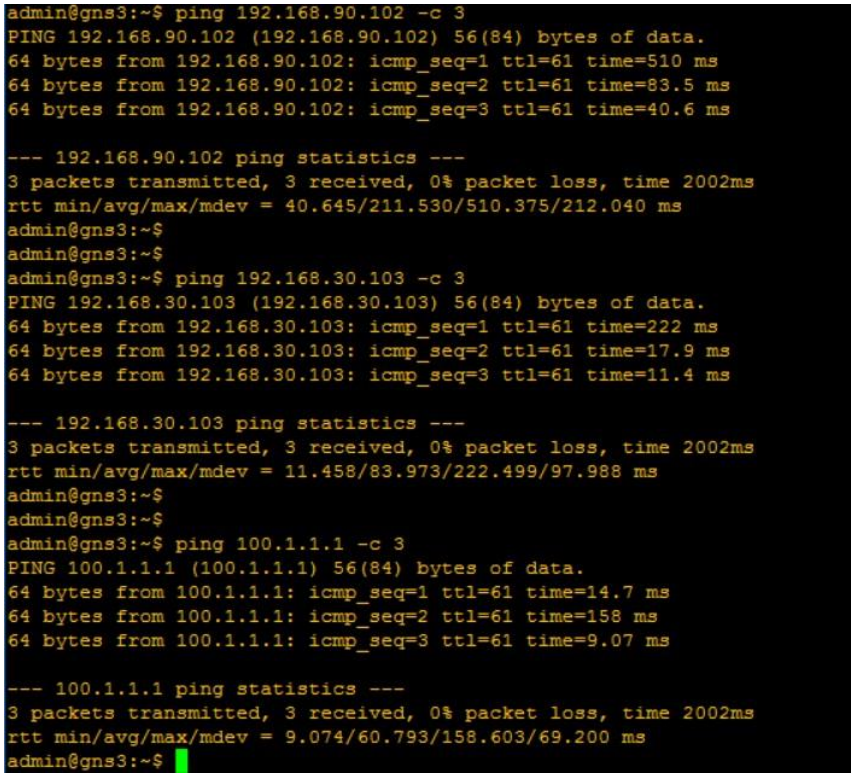
On the console of **Host1-50** login with the following credentials:

- username: **admin**
- password: **admin**

Ping hosts

On **Host1-50,** validate the network by pinging **Host2-90, Host3-30 and ExtRtr loopback** using the following commands:

```
ping 192.168.90.102 -c 3
ping 192.168.30.103 -c 3
ping 100.1.1.1 -c 3
```



This concludes this lesson.

Module conclusion

In this module you learned about Enterprise SONiC Distribution by Dell Technologies which can be used to provide Data Center fabrics using industry standard protocols. During the module you:

- Configured MCLAG with static addresses.
- Configured a BGP fabric with unnumbered interfaces. • Advertised networks with route-maps and prefix-lists.
- Validated the fabric.
- Tested Layer 3 connectivity across the fabric with end hosts.