Enterprise SONiC Distribution by Dell Technologies Proof of Concept

Deployment Guide

H18645.1

Abstract

This guide demonstrates a proof-of-concept deployment of a leaf-spine network with a BGP EVPN overlay using Enterprise SONiC. It details the topology requirements and processes for rapid deployment of a test network.

Dell Technologies Solutions

June 2021

Notes, cautions, and warnings

NOTE: A NOTE indicates important information that helps you make better use of your product.

CAUTION: A CAUTION indicates either potential damage to hardware or loss of data and tells you how to avoid the problem.

WARNING: A WARNING indicates a potential for property damage, personal injury, or death.

Contents

1

Introduction

This chapter presents the following topics:

Topics:

Introduction to Enterprise SONiC

Enterprise SONiC Distribution by Dell Technologies is a hardened, validated, and supported version of the Software for Open Networking in the Cloud (SONiC) operating system for switch configuration and monitoring. It includes distribution of open-source community SONiC and additional features to support the ecosystem and partners.

Enterprise SONiC supports an industry standard, intuitive Management Framework command-line interface (MF-CLI), and supports programmatic interfaces such as REST API, and the Google Remote Procedure Call (gRPC) Network Management Interface (gNMI). This guide uses MF-CLI to configure the network nodes.

The objective of this guide is to demonstrate the procedures to configure a BGP leaf-spine topology with an EVPN overlay to support L2 multirack and multitenancy capabilities. Symmetric routing is used between networks in same VRF. This topology meets many of the requirements for a modern data center to support segmentation and mobility of virtual machines.

2

POC Prerequisites

This chapter presents the following topics:

Topics:

Hardware requirements

This POC uses the following Dell EMC data center switches.

Dell EMC PowerSwitch S5232F-ON

The PowerSwitch S5232F-ON switch is a 1-Rack Unit (1U), multilayer switch with 32 QSFP28 ports, as shown in the following figure. The high-performance S5232F-ON switch is an optimal leaf or spine switch for environments requiring connectivity for 10/25/40/50/100 GbE ports.

In this guide, the PowerSwitch S5232-ON is used as a spine switch.

Figure 1. PowerSwitch S5232F-ON switch

Dell EMC PowerSwitch S5248F-ON

The PowerSwitch S5248F-ON switch is a 1-Rack Unit (1U), multilayer switch with 48 SFP28 25 GbE ports, two QSFP28-DD ports (two 100 GbE interfaces per port), and four QSFP28 100 GbE ports, as shown in the following figure. The high-performance S5248F-ON switch is an optimal leaf switch for environments that require connectivity for 25 GbE and 10 GbE compute and storage.

In this guide, the PowerSwitch S5248F-ON is used as a leaf switch.

Figure 2. PowerSwitch S5248F-ON switch

POC requirements

The general requirements to set up the proof of concept (POC) topology include the following:

- The following Dell EMC PowerSwitch switches with Open Network Install Environment (ONIE) installed, racked, and powered on.
  - Two S5232F-ON as spines
  - Four S5248F-ON as leafs

- One N3248TE-ON as a 1 GB management switch (or other suitable 1 GB switch)
- An optional DHCP/DNS server for automatic installation or manual installation with a static IP address
- A valid Dell Digital Locker (DDL) account (see https://www.dell.com/support/software) with entitlement to Enterprise SONiC Distribution by Dell Technologies
- An HTTP, TFTP, or FTP server connected to the management network
- Two servers used as fabric host devices with two 10 GbE or 25 GbE network interfaces
- Supported cables and optics (see Support Matrix for Enterprise SONiC Distribution by Dell Technologies 3.2.)
- Console connection to a terminal server or laptop
  - Required serial port settings are 115200, 8 data bits, and no parity

The following figure shows the ideal physical topology for the management network.

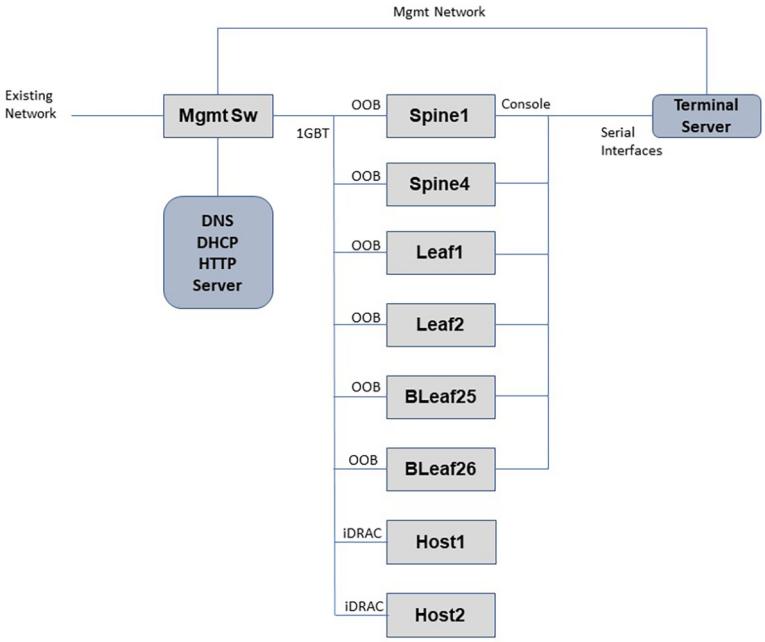

Figure 3. Management network topology

NOTE: This topology reflects the interface and node naming convention used in the setup.

POC reference deployment

This POC uses a typical leaf and spine topology. The leaf switches are PowerSwitch S5248F-ON switches, and the spine switches are PowerSwitch S5232F-ON switches.

MultiChassis Link Aggregation (MCLAG) is configured on the leaf switches to provide link and device redundancy at the leaf layer.

The following figure shows the physical connections used in the configurations for this POC.

NOTE: The external router is shown only for reference and is not used in this deployment.

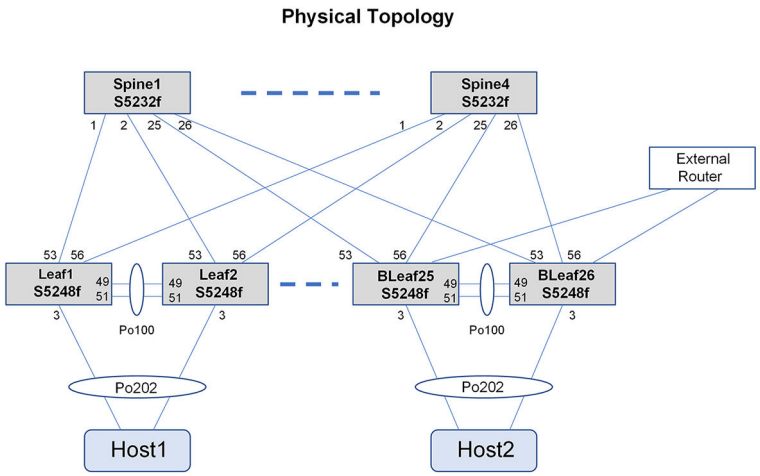

Figure 4. POC physical reference deployment setup

NOTE: This topology reflects the interface and node naming convention used in the setup.

The following figure shows the logical layout of the underlay and overlay networks. Two Virtual Network Instances (VNIs) are deployed between the two leaf pairs depicting the overlay network. The L2 VNI supports stretching the L2 network (Vlan 140) between the leaf pairs. The L3 VNI supports symmetric routing between networks in the same VRF.

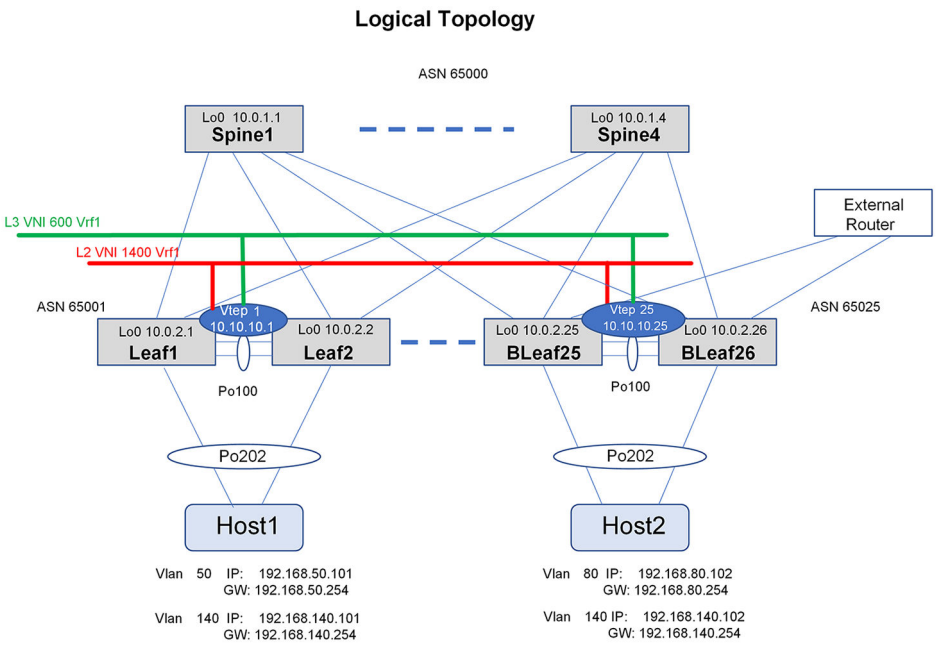

Figure 5. POC logical reference deployment setup

In the preceding figures:

- The leaf and border leaf devices are configured to act as Virtual Tunnel End Points (VTEPs). One Layer 3 VNI is created per Virtual Routing and Forwarding (VRF). IPv4 prefixes learned on the VRF are redistributed into the overlay. The leaf devices are Layer 2 with the hosts.
- Border leaf 25 (BLeaf25) and border leaf 26 (BLeaf26) are peered to the gateway router that is then used to advertise the default route.
- Each host has two paths that connect to two leaf switches through a port channel.
- Each leaf has two paths that connect to the spine.
- Each leaf has a backup path that connects to the peer leaf.

POC deployment features

The following features are configured for this POC:

- The interfaces used in this POC are 100 GbE except the interfaces to the servers, which are 10/25 GbE.
- LLDP is enabled by default.
- Layer 3:
  - All Layer 3 interfaces and VLANs with IP addresses are configured with an MTU of 9100 bytes.
  - The fabric is configured with BGP unnumbered.
  - eBGP is configured in the fabric.
  - ECMP is configured between the leaf and spine switches.
  - ASN allocations are unique on the leaf and spine switches:
    - Leaf 1 and Leaf 2—ASN# 65001
    - BLeaf 25 and BLeaf 26—ASN# 65025
    - Spine 1 and Spine 4—ASN# 65000
- VRF
  - VRF in leaf, single tenant setup:
    - Vrf-tenant1 (Tenant1) for overlay (EVPN) routes
    - Default VRF for underlay routes and connectivity
- Monitoring:
  - Syslog—Interface logs, BGP session logs, and chassis-related (PSU, fan, temperature) syslog alerts
  - SNMP—Enabled SNMPv2 queries and destination for SNMP traps
- IP prefixes

The following table shows the node addressing information.

Table 1. IP Prefixes – Switches and VXLAN VTEP tunnels

Node	BGP ASN	Router-ID Loopback0	VTEP IP Loopback1	Stretched L2 Anycast-address	L3 VLAN Anycast-address	Anycast-MAC-address
Leaf1	65001	10.0.2.1	10.10.10.1	192.168.140.254/24	192.168.50.254/24	00:00:00:00:00:01:02
Leaf2	65001	10.0.2.2	10.10.10.1	192.168.140.254/24	192.168.50.254/24	00:00:00:00:00:01:02
BLeaf25	65025	10.0.2.25	10.10.10.25	192.168.140.254/24	192.168.80.254/24	00:00:00:00:00:01:02
BLeaf26	65025	10.0.2.26	10.10.10.25	192.168.140.254/24	192.168.80.254/24	00:00:00:00:00:01:02
Spine1	65000	10.0.1.1	NA	NA	NA	NA
Spine4	65000	10.0.1.4	NA	NA	NA	NA

3

POC Deployment Procedure

This chapter presents the following topics:

Topics:

- POC deployment steps

POC deployment steps

The following steps provide the framework for the POC deployment.

Steps

1. Download the Enterprise SONiC Distribution code:
   a. Go to https://www.dell.com/support/software and log in with your Dell registered credentials (DDL).
   b. Select the appropriate entry:

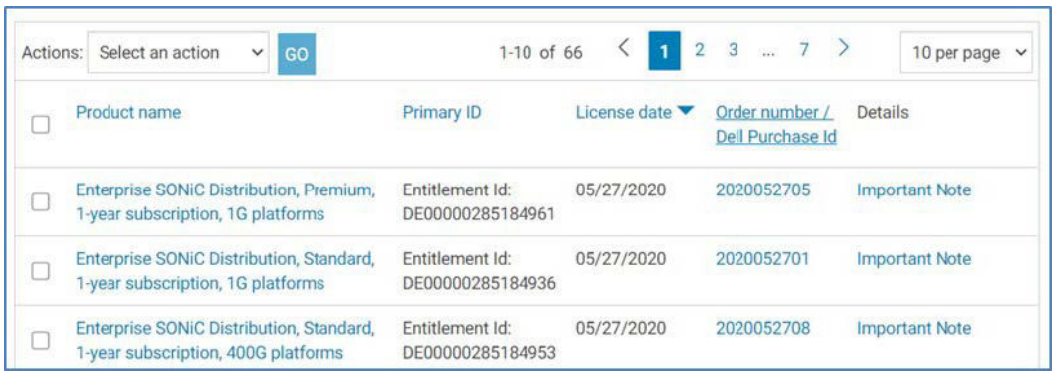

Figure 6. Enterprise SONiC entitlement

   c. Select the appropriate Enterprise SONiC entitlement for your switches and download the installer:

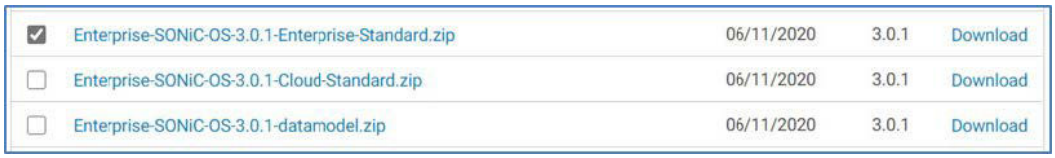

Figure 7. Enterprise SONiC images

   To run this POC, select **Enterprise_SONiC_OS_3.2.0_Enterprise_Standard** version, or later.

   NOTE: The **Enterprise_Premium** version is also supported for this POC.

   d. Extract the `.zip` file and copy the `.bin` file to your HTTP, TFTP, or FTP server.
2. Uninstall any existing images.

   When you reboot the switches, a startup menu is displayed. Use the arrow keys on the keyboard to select the `ONIE: Uninstall OS` option, as shown in the following figure:

```
+-----------------------------------------------------------------------------+
|*ONIE: Install OS                                                            |
| ONIE: Rescue                                                                |
| ONIE: Uninstall OS                                                          |
| ONIE: Update ONIE                                                           |
| ONIE: Embed ONIE                                                            |
+-----------------------------------------------------------------------------+
```

Figure 8. ONIE startup switch menu: Uninstall OS

3. Install the newly downloaded Enterprise SONiC image:
   a. Use the arrow keys on the keyboard to select the `ONIE: Install OS` option, as shown in the following figure:

```
+-----------------------------------------------------------------------------+
|*ONIE: Install OS                                                            |
| ONIE: Rescue                                                                |
| ONIE: Uninstall OS                                                          |
| ONIE: Update ONIE                                                           |
| ONIE: Embed ONIE                                                            |
+-----------------------------------------------------------------------------+
```

Figure 9. ONIE startup switch menu: Install OS

   The switch reboots.

   b. At the prompt, enter the `onie-discovery-stop` command to stop the ONIE discovery process.

```
ONIE:/ # onie-discovery-stop
```

   c. If a DHCP server is not available, manually configure an IP address (for example, 192.168.0.2). To configure an IP address on the switch's management interface, run the following commands at the `ONIE#` prompt:

```
ONIE:/ # ifconfig eth0 192.168.0.2 netmask 255.255.255.0
ONIE:/ # ip route add default via 192.168.0.1
```

   d. Ensure that the switches can contact the HTTP server hosting the image before attempting the installation. Pinging the HTTP server verifies this.

```
ONIE:/ # ping 192.168.2.210
```

   e. Run the following command to install the Enterprise SONiC 3.2 code on the switches:

```
ONIE:/ # onie-nos-install https://192.168.2.210/
Enterprise_SONiC_OS_3.2.0_Enterprise_Standard.bin
```

   When the image has been installed, the switches reboot with the correct image.

4. Log in to the switches:
   a. Use username **`admin`** and password **`YourPaSsWoRd`**, as shown in the following figure:

```
At Console:
Debian GNU/Linux 9 sonic ttyS1

sonic login: admin
Password: YourPaSsWoRd

SSH from any remote server to sonic can be done by connecting to the IP address of the
Management interface
user@debug:~$ ssh admin@sonic_ip_address(or SONIC DNS Name)
admin@sonic's password:
```

Figure 10. Logging in to the switch

   b. Change the default password, as shown in the following figure:

```
You are required to change your password immediately (root enforced)
Changing password for admin.
(current) UNIX password: YourPaSsWoRd
Enter new UNIX password: *********
Retype new UNIX password: ********
Linux sonic 4.9.0-11-2-amd64 #1 SMP Debian 4.9.189-3+deb9u2 (2019-11-11) x86_64
You are on

 ____  ___  _   _ _____ ____
|___ \|_ _|| \ | |_ _|/ ___|
  __) | | |  \| | | | \___ \
 / __/ | | | |\  | | |  ___) |
|_____|___|_| \_|_|_|\____|

-- Software for Open Networking in the Cloud --

Unauthorized access and/or use are prohibited.
All access and/or use are subject to monitoring.

Help:    http://azure.github.io/SONiC/

admin@sonic:~$
```

Figure 11. Changing the password from the default

5. Disable the Zero-Touch Provisioning (ZTP) process:
   a. Run the following command at the switch prompt:

   ```
   admin@sonic:~$ sudo config ztp disable -y
   ```

   The switches run ZTP when they first start up. This process must be disabled when the ZTP server is not being used to configure the switches. Allow a few minutes for the switch restart and allow manual configuration.

   b. Verify that the switches are ready by running the following command at the switch prompt:

   ```
   admin@sonic:~$ show system status
   System is ready
   ```

6. Perform the actions in this step only if Enterprise SONiC has not been recently loaded from ONIE.
   a. Ensure that the switches have set the default port profile.
      - The correct HwSku for a PowerSwitch S5248F-ON is: DellEMC-S5248f-P-25G-DPB (Present).
      - The correct HwSku for a PowerSwitch S5232F-ON is: DellEMC-S5232f-C32 (Present).
   b. To view the effective port profile, run the HwSku command:

   ```
   admin@sonic:~$ sudo config-hwsku list
   ```

   c. To change the port profile, run the appropriate command:
      - For the PowerSwitch S5248F-ON, run the following command:

      ```
      admin@sonic:~$ sudo config-hwsku set DellEMC-S5248f-P-25G-DPB
      ```

      - For the PowerSwitch S5232F-ON, run the following command:

      ```
      admin@sonic:~$ sudo config-hwsku set DellEMC-S5232f-C32
      ```

   The switch configuration is cleared, and a reboot occurs.

7. Verify Layer 2 or Layer 3 mode on the ASIC:

   NOTE: After installing Enterprise SONiC from ONIE, the default switching profile for a switch running Enterprise SONiC is Layer 3.

   a. If Enterprise SONiC was not installed from ONIE, verify the switching profile configuration:

   ```
   admin@sonic:~$ sonic-cli
   sonic# show running-configuration | grep factory
   factory default profile l3 confirm
   ```

b. If the switching profile is set to Layer 2, change the switching profile:

```
admin@sonic:~$ sonic-cli
sonic# configure terminal
sonic(config)# factory default profile l3 confirm
```

The switch reboots after the appropriate command is run.

8. If necessary, configure breakout ports:

NOTE: This guide uses the default port breakouts. If any of the high-speed ports (Eth1/49 – Eth1/56 on a PowerSwitch S5248F-ON or Eth1/1 – Eth1/32 on a PowerSwitch S5232F-ON) must be broken out (4 x 25 GbE, 4x 10 GbE), do so before configuring the switch.

a. Run the following commands:

```
admin@sonic:~$ sonic-cli
sonic# configure terminal
sonic(config)# interface breakout port 1/53 mode 4x25G
```

The Dynamic Port Breakout in-progress, use `show interface breakout port 1/53` to check status message displays.

b. Verify the success of the breakout.

```
sonic# show interface breakout port 1/53
-----------------------------------------------
Port   Breakout Mode   Status        Interfaces
-----------------------------------------------
1/53   4x25G           Completed     Eth1/53/1
                                     Eth1/53/2
                                     Eth1/53/3
                                     Eth1/53/4
```

9. If using 10 GbE server interfaces on a S5200 series switch, change the port group to support 10 GbE interfaces. These are applied to groups of four, for example, Eth1/1-1/4.

NOTE: Only the S5212F-ON, S5224F-ON, S5248F-ON, and S5296F-ON switches have port groups.

For example, change interfaces Eth1/1 to Eth 1/4 from 25 GbE to 10 GbE.

```
admin@sonic:~$ sonic-cli
sonic# configure terminal
sonic(config)# port-group 1 speed 10000
sonic(config)# end
sonic# write memory
```

4

POC Configuration

This chapter presents the following topics:

Topics:

Spine configuration steps

The following diagram highlights the devices configured in this section. Spine1 and Spine4 are configured demonstrating the use of four spines.

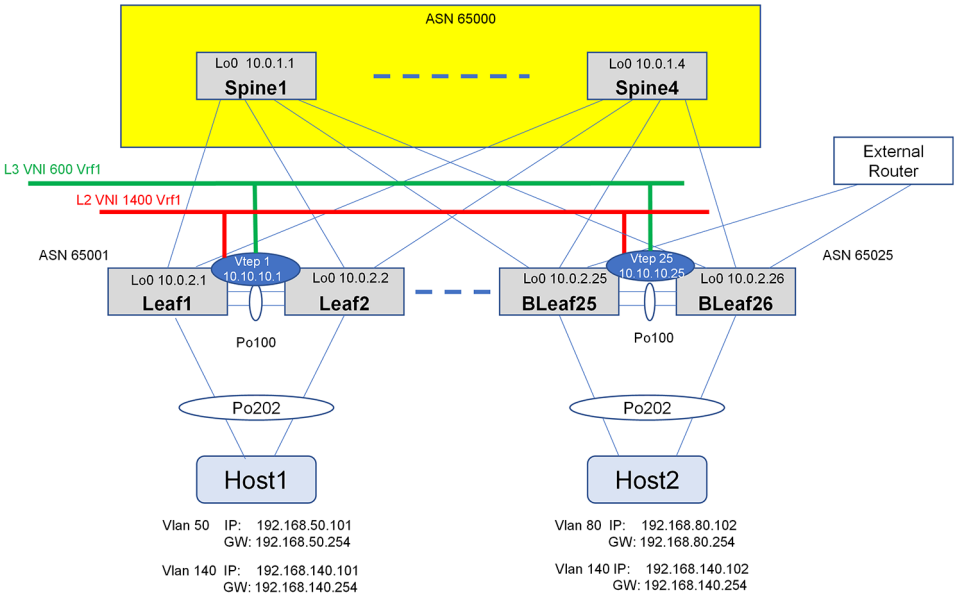

Figure 12. Device configuration

NOTE: Only Spine1 and Spine 4 are configured in this guide. More spines can be added if needed to increase fabric bandwidth.

1. Enter Management Framework CLI (MF-CLI) and configuration mode.
2. Change interface naming mode to Standard and add hostname.
3. Exit back to the Linux shell to activate the changes.

Spine1	Spine4

`sonic-cli`
`configure terminal`
 	
`interface-naming standard`
	`sonic-cli`
	`configure terminal`
	
	`interface-naming standard`

Spine1	Spine4

`hostname Spine1`
`end`
 	
`exit`
	`hostname Spine4`
	`end`
	
	`exit`

4. Re-enter the MF-CLI and configuration mode.
5. Assign a loopback interface with a unique router ID for each spine.
6. Enable IPv6 on the spine downlinks to support unnumbered BGP point-to-point links and enable the interfaces.

Spine1	Spine4

`sonic-cli`
`configure terminal`
 	
`interface loopback 0`
` description router-id`
` ip address 10.0.1.1/32`
` exit`
 	
`interface Eth 1/1`
` description Leaf1`
` ipv6 enable`
` no shutdown`
 	
`interface Eth 1/2`
` description Leaf2`
` ipv6 enable`
` no shutdown`
 	
`interface Eth 1/25`
` description BLeaf25`
` ipv6 enable`
` no shutdown`
 	
`interface Eth 1/26`
` description BLeaf26`
` ipv6 enable`
` no shutdown`
` exit`
	`sonic-cli`
	`configure terminal`
	
	`interface loopback 0`
	` description router-id`
	` ip address 10.0.1.4/32`
	` exit`
	
	`interface Eth 1/1`
	` description Leaf1`
	` ipv6 enable`
	` no shutdown`
	
	`interface Eth 1/2`
	` description Leaf2`
	` ipv6 enable`
	` no shutdown`
	
	`interface Eth 1/25`
	` description BLeaf25`
	` ipv6 enable`
	` no shutdown`
	
	`interface Eth 1/26`
	` description BLeaf26`
	` ipv6 enable`
	` no shutdown`
	` exit`

7. Configure the BGP router with ECMP enabled.
8. Add the IPv4 unicast address-family to advertise the router-ID and paths for two leafs.
9. Configure the BGP leaf peer-group. Set timers and enable unnumbered BGP and BFD.
10. Activate both IPv4 unicast and EVPN address families for the peer group.
11. Assign the leaf neighbor interfaces to the peer-group.
12. Save the configuration.

Spine1	Spine4

`router bgp 65000`
` router-id 10.0.1.1`
` bestpath as-path multipath-relax`
 	
` address-family ipv4 unicast`
`  redistribute connected`
`  maximum-paths 2`
`  exit`
 	
`peer-group LEAF`
`  advertisement-interval 5`
`  timers 3 9`
`  remote-as external`
`  capability extended-nexthop`
`  bfd`
`  address-family ipv4 unicast`
	`router bgp 65000`
	` router-id 10.0.1.4`
	` bestpath as-path multipath-relax`
	
	` address-family ipv4 unicast`
	`  redistribute connected`
	`  maximum-paths 2`
	`  exit`
	
	`peer-group LEAF`
	`  advertisement-interval 5`
	`  timers 3 9`
	`  remote-as external`
	`  capability extended-nexthop`
	`  bfd`
	`  address-family ipv4 unicast`

Spine1	Spine4

`activate`
`exit`
`address-family l2vpn evpn`
`activate`
`exit`
`exit`
 	
`neighbor interface Eth 1/1`
`peer-group LEAF`
`exit`
`neighbor interface Eth 1/2`
`peer-group LEAF`
`exit`
`neighbor interface Eth 1/25`
`peer-group LEAF`
`exit`
`neighbor interface Eth 1/26`
`peer-group LEAF`
`exit`
`exit`
`end`
 	
`write memory`
	`activate`
	`exit`
	`address-family l2vpn evpn`
	`activate`
	`exit`
	`exit`
	
	`neighbor interface Eth 1/1`
	`peer-group LEAF`
	`exit`
	`neighbor interface Eth 1/2`
	`peer-group LEAF`
	`exit`
	`neighbor interface Eth 1/25`
	`peer-group LEAF`
	`exit`
	`neighbor interface Eth 1/26`
	`peer-group LEAF`
	`exit`
	`exit`
	`end`
	
	`write memory`

Leaf1 and Leaf2 configuration steps

The following diagram highlights the leaf swiches configured in this section. Leaf1 and Leaf2 are configured as a MCLAG pair.

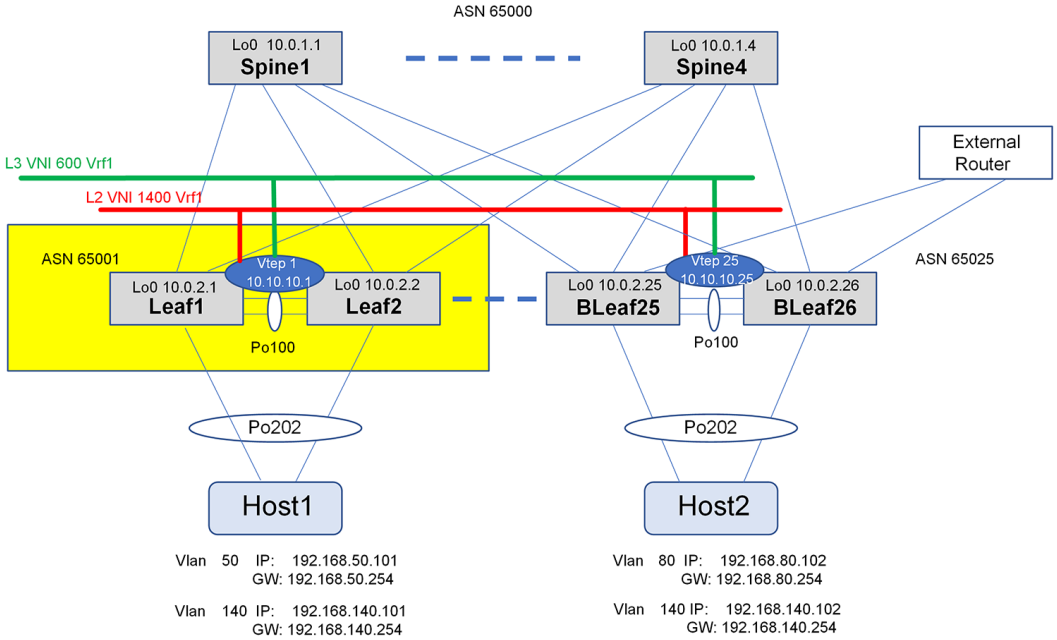

Figure 13. Leaf1 and Leaf2 configuration

1. Enter Management Framework CLI (MF-CLI) and then configuration mode.
2. Change the interface naming mode to Standard and add a hostname.
3. Exit back to Linux shell to activate the changes.

Leaf1	Leaf2

`sonic-cli`
`configure terminal`
`interface-naming standard`
`hostname Leaf1`
`end`
 	
`exit`
	`sonic-cli`
	`configure terminal`
	`interface-naming standard`
	`hostname Leaf2`
	`end`
	
	`exit`

4. Re-enter the MF-CLI and configuration mode
5. Assign a loopback interface with a unique router ID on each leaf.
6. Then enable IPv6 on spine uplinks to support unnumbered BGP point-to-point links and enable the interfaces.

Leaf1	Leaf2

`sonic-cli`
`configure terminal`
`interface loopback 0`
` description router-id`
` ip address 10.0.2.1/32`
` exit`
 	
`interface Eth 1/53`
` description Spine1`
` ipv6 enable`
` no shutdown`
 	
`interface Eth 1/56`
` description Spine4`
` ipv6 enable`
` no shutdown`
` exit`
	`sonic-cli`
	`configure terminal`
	`interface loopback 0`
	` description router-id`
	` ip address 10.0.2.2/32`
	` exit`
	
	`interface Eth 1/53`
	` description Spine1`
	` ipv6 enable`
	` no shutdown`
	
	`interface Eth 1/56`
	` description Spine4`
	` ipv6 enable`
	` no shutdown`
	` exit`

7. Assign a VRF for each tenant. This POC only has one tenant, but will support multiple tenants with additional VRFs.
8. Configure VLANs 50 and 140 for the hosts, bind to the tenant's VRF and assign the same anycast-address to each host VLAN.
9. Assign a dedicated VLAN (Vlan60) for a L3 VNI and bind to the tenant's VRF.

   NOTE: The VLAN assigned to the L3 VNI does not need an IP address.

Leaf1	Leaf2

`ip vrf Vrf-tenant1`
 	
`interface Vlan 50`
` ip vrf forwarding Vrf-tenant1`
` ip anycast-address 192.168.50.254/24`
` neigh-suppress`
` exit`
 	
`interface Vlan 140`
` ip vrf forwarding Vrf-tenant1`
` ip anycast-address 192.168.140.254/24`
` neigh-suppress`
` exit`
 	
`interface Vlan 60`
` ip vrf forwarding Vrf-tenant1`
` exit`
	`ip vrf Vrf-tenant1`
	
	`interface Vlan 50`
	` ip vrf forwarding Vrf-tenant1`
	` ip anycast-address 192.168.50.254/24`
	` neigh-suppress`
	` exit`
	
	`interface Vlan 140`
	` ip vrf forwarding Vrf-tenant1`
	` ip anycast-address 192.168.140.254/24`
	` neigh-suppress`
	` exit`
	
	`interface Vlan 60`
	` ip vrf forwarding Vrf-tenant1`
	` exit`

10. Configure the MCLAG peer-link and port channel members.
11. Configure all VLANs on the MCLAG's peer-link port channel.

    NOTE: The MCLAG peer-link must include the L3 VNI's VLAN.

Leaf1	Leaf2

`interface PortChannel 100`
` description MCLAG-Peer-Link`
` switchport trunk allowed vlan add 50,60,140`
` exit`
 	
`interface Eth 1/49`
` description MCLAG-Peer-Link`
` channel-group 100`
` no shutdown`
` exit`
 	
`interface Eth 1/51`
` description MCLAG-Peer-Link`
` channel-group 100`
` no shutdown`
` exit`
	`interface PortChannel 100`
	` description MCLAG-Peer-Link`
	` switchport trunk allowed vlan add 50,60,140`
	` exit`
	
	`interface Eth 1/49`
	` description MCLAG-Peer-Link`
	` channel-group 100`
	` no shutdown`
	` exit`
	
	`interface Eth 1/51`
	` description MCLAG-Peer-Link`
	` channel-group 100`
	` no shutdown`
	` exit`

CAUTION: When copying and pasting the content from the table above, ensure that the underlined monospace text remains on one line. Failure to keep the underlined content on one line results in a line error on the switch.

12. Configure the MCLAG domain.
13. The loopback IP address used for the router-id is also used as the MCLAG source and peer IP address.
14. Assign the host's port channel interfaces and assign to the MCLAG domain.

    NOTE: By default, port channels are LACP. Use the `interface PortChannel <num> mode on` command for static LAGs. If changing LAG modes, you must completely remove and reassign the port channel.

    NOTE: The spine switches are used for the MCLAG peer keep-alive link communication

Leaf1	Leaf2

`mclag domain 1`
` source-ip 10.0.2.1`
` peer-ip 10.0.2.2`
` peer-link PortChannel 100`
` delay-restore 90`
` exit`
 	
`interface PortChannel 202`
` description Host1_PortChannel`
` switchport trunk allowed vlan add 50,140`
` mclag 1`
` exit`
 	
`interface Eth 1/3`
` description Host1_PortChannel`
` channel-group 202`
` no shutdown`
` exit`
	`mclag domain 1`
	` source-ip 10.0.2.2`
	` peer-ip 10.0.2.1`
	` peer-link PortChannel 100`
	` delay-restore 90`
	` exit`
	
	`interface PortChannel 202`
	` description Host1_PortChannel`
	` switchport trunk allowed vlan add 50,140`
	` mclag 1`
	` exit`
	
	`interface Eth 1/3`
	` description Host1_PortChannel`
	` channel-group 202`
	` no shutdown`
	` exit`

15. Enable static anycast-address and assign a unique MAC to be used across all leaf switches.
16. Create a loopback interface for the VTEP IP address. This address must be the same for each MCLAG peer.
17. Create a VXLAN interface and assign the same VTEP source IP address to each peer.
18. Assign a unique primary-ip to optimize routing for orphan ports and active-standby hosts. The router-id can be used for this purpose.
19. Map the L2 VNI to the host VLANs that are stretched across the fabric.
20. Create a L3 VNI by mapping the dedicated VLAN (Vlan60) to the non-default VRF.

    NOTE: The Network Virtualization Overlay (NVO) is automatically assigned to the VXLAN interface.

Leaf1	Leaf2
```ip anycast-address enable ip anycast-mac-address 00:00:00:00:01:02 interface Loopback 1 description LogicalVTEP ip address 10.10.10.1/32 exit interface vxlan vtep1 source-ip 10.10.10.1 primary-ip 10.0.2.1 map vni 1400 vlan 140 map vni 600 vlan 60 map vni 600 vrf Vrf-tenant1 exit```	```ip anycast-address enable ip anycast-mac-address 00:00:00:00:01:02 interface Loopback 1 description LogicalVTEP ip address 10.10.10.1/32 exit interface vxlan vtep1 source-ip 10.10.10.1 primary-ip 10.0.2.2 map vni 1400 vlan 140 map vni 600 vlan 60 map vni 600 vrf Vrf-tenant1 exit```

21. Configure the BGP router with ECMP enabled.
22. Configure the underlay's address-family and redistribute the connected networks.
23. Add VNI advertisements to the EVPN address family and assign the MCLAG peer's primary-ip.

Leaf1	Leaf2
```router bgp 65001 router-id 10.0.2.1 bestpath as-path multipath-relax address-family ipv4 unicast redistribute connected maximum-paths 2 exit address-family l2vpn evpn advertise-all-vni advertise-pip peer-ip 10.0.2.2 exit```	```router bgp 65001 router-id 10.0.2.2 bestpath as-path multipath-relax address-family ipv4 unicast redistribute connected maximum-paths 2 exit address-family l2vpn evpn advertise-all-vni advertise-pip peer-ip 10.0.2.1 exit```

24. Configure the BGP spine peer-group.
25. Set timers and enable unnumbered BGP and BFD on the peer-group.
26. Activate both IPv4 unicast and EVPN address families for the spine peer group.
27. Set `allowas-in` to support MCLAG peer's ASN to be received.

Leaf1	Leaf2
```peer-group SPINE advertisement-interval 5 timers 3 9 remote-as external capability extended-nexthop bfd address-family ipv4 unicast activate allowas-in 1 exit address-family l2vpn evpn activate exit exit```	```peer-group SPINE advertisement-interval 5 timers 3 9 remote-as external capability extended-nexthop bfd address-family ipv4 unicast activate allowas-in 1 exit address-family l2vpn evpn activate exit exit```

28. Assign the spine neighbor interfaces to the peer-group.
29. Create the BGP router for each tenant's VRF.
30. Save the configuration.

Leaf1	Leaf2

`neighbor interface Eth 1/53`
`  peer-group SPINE`
`  exit`
`  neighbor interface Eth 1/56`
`  peer-group SPINE`
`  exit`
`  exit`
 	
`router bgp 65001 vrf Vrf-tenant1`
`  address-family ipv4 unicast`
`  redistribute connected`
`  exit`
`  address-family l2vpn evpn`
`  advertise ipv4 unicast`
`  exit`
`  exit`
`end`
 	
`write memory`
	`neighbor interface Eth 1/53`
	`  peer-group SPINE`
	`  exit`
	`  neighbor interface Eth 1/56`
	`  peer-group SPINE`
	`  exit`
	`  exit`
	
	`router bgp 65001 vrf Vrf-tenant1`
	`  address-family ipv4 unicast`
	`  redistribute connected`
	`  exit`
	`  address-family l2vpn evpn`
	`  advertise ipv4 unicast`
	`  exit`
	`  exit`
	`end`
	
	`write memory`

Border leafs configuration steps

The following diagram highlights the switches configured as border leafs. BLeaf25 and BLeaf26 are configured as an MCLAG pair and are also supporting hosts.

NOTE: This configuration does not include the details for the External Router.

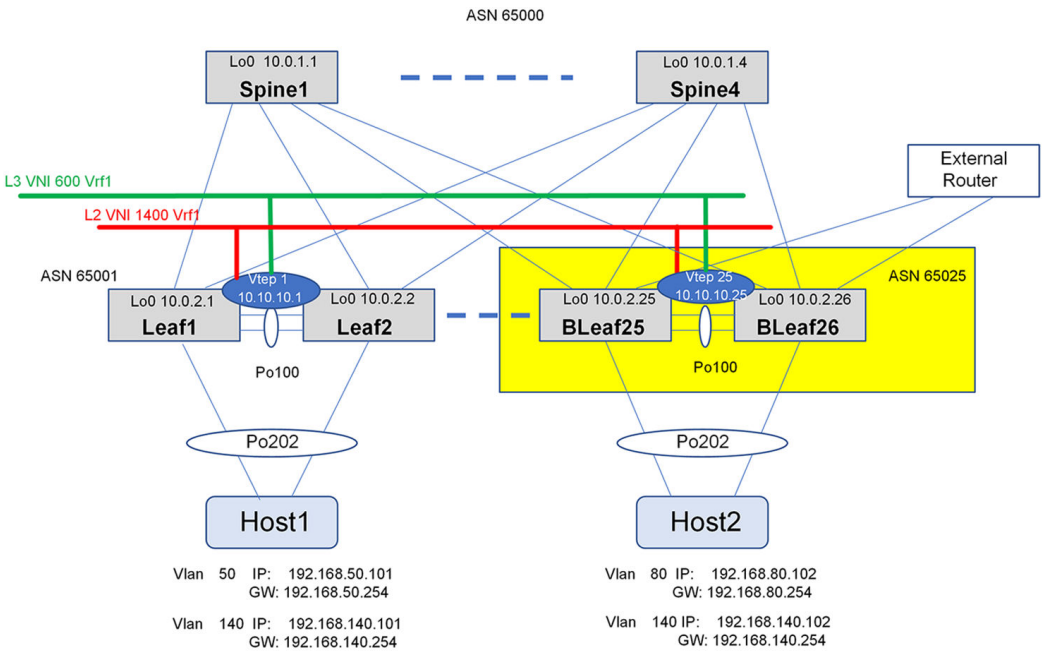

Figure 14. Configuration of border leafs

1. Enter Management Framework CLI (MF-CLI) and then configuration mode.
2. Change the interface naming mode to Standard, add hostname, save configuration and exit back to Linux shell to activate the changes.

BLeaf25	BLeaf26

`sonic-cli`
`configure terminal`
 	
`interface-naming standard`
`hostname BLeaf25`
`end`
`write memory`
`exit`
	`sonic-cli`
	`configure terminal`
	
	`interface-naming standard`
	`hostname BLeaf26`
	`end`
	`write memory`
	`exit`

3. Re-enter the MF-CLI and configuration mode.
4. Assign a unique router ID on each leaf.
5. Then enable IPv6 on spine uplinks to support unnumbered BGP point-to-point links and enable the interfaces.

BLeaf25	BLeaf26

`sonic-cli`
`configure terminal`
 	
`interface loopback 0`
` description router-id`
` ip address 10.0.2.25/32`
` exit`
 	
`interface Eth 1/53`
` description Spine1`
` ipv6 enable`
` no shutdown`
 	
`interface Eth 1/56`
` description Spine4`
` ipv6 enable`
` no shutdown`
` exit`
	`sonic-cli`
	`configure terminal`
	
	`interface loopback 0`
	` description router-id`
	` ip address 10.0.2.26/32`
	` exit`
	
	`interface Eth 1/53`
	` description Spine1`
	` ipv6 enable`
	` no shutdown`
	
	`interface Eth 1/56`
	` description Spine4`
	` ipv6 enable`
	` no shutdown`
	` exit`

6. Assign a VRF for each tenant. This POC only has one tenant, but will support multiple tenants with additional VRFs.
7. Configure VLANs 80 and 140 for the hosts, bind to the tenant's VRF and assign the same anycast-address to each host VLAN.
8. Assign a dedicated VLAN (Vlan60) for a L3 VNI and bind to the tenant's VRF.

   NOTE: The VLAN assigned to the L3 VNI does not need an IP address.

BLeaf25	BLeaf26

`ip vrf Vrf-tenant1`
 	
`interface Vlan 80`
` ip vrf forwarding Vrf-tenant1`
` ip anycast-address 192.168.80.254/24`
` neigh-suppress`
` exit`
 	
`interface Vlan 140`
` ip vrf forwarding Vrf-tenant1`
` ip anycast-address 192.168.140.254/24`
` neigh-suppress`
` exit`
 	
`interface Vlan 60`
` ip vrf forwarding Vrf-tenant1`
` exit`
	`ip vrf Vrf-tenant1`
	
	`interface Vlan 80`
	` ip vrf forwarding Vrf-tenant1`
	` ip anycast-address 192.168.80.254/24`
	` neigh-suppress`
	` exit`
	
	`interface Vlan 140`
	` ip vrf forwarding Vrf-tenant1`
	` ip anycast-address 192.168.140.254/24`
	` neigh-suppress`
	` exit`
	
	`interface Vlan 60`
	` ip vrf forwarding Vrf-tenant1`
	` exit`

9. Configure the MCLAG peer-link and port channel members.
10. Configure all VLANs on the MCLAG's peer-link port channel.

BLeaf25	BLeaf26

`interface PortChannel 100`
` description MCLAG-Peer-Link`
` switchport trunk allowed vlan add 60,80,140`
` exit`
 	
`interface Eth 1/49`
` description MCLAG-Peer-Link`
` channel-group 100`
` no shutdown`
` exit`
 	
`interface Eth 1/51`
` description MCLAG-Peer-Link`
` channel-group 100`
` no shutdown`
` exit`
	`interface PortChannel 100`
	` description MCLAG-Peer-Link`
	` switchport trunk allowed vlan add 60,80,140`
	` exit`
	
	`interface Eth 1/49`
	` description MCLAG-Peer-Link`
	` channel-group 100`
	` no shutdown`
	` exit`
	
	`interface Eth 1/51`
	` description MCLAG-Peer-Link`
	` channel-group 100`
	` no shutdown`
	` exit`

CAUTION: When copying and pasting the content from the table above, ensure that the underlined monospace text remains on one line. Failure to keep the underlined content on one line results in a line error on the switch.

11. Configure the MCLAG domain.
12. The loopback IP address used for the router-id is also used as the MCLAG source and peer IP address.
13. Assign the host's port channel interfaces and assign to the MCLAG domain.

    NOTE: By default, port channels are LACP. Use the `interface PortChannel <num> mode on` command for static LAGs. If changing LAG modes, you must completely remove and reassign the port channel.

    NOTE: The spine switches are used for the MCLAG peer's keep-alive link communication.

BLeaf25	BLeaf26

`mclag domain 1`
` source-ip 10.0.2.25`
` peer-ip 10.0.2.26`
` peer-link PortChannel 100`
` delay-restore 90`
` exit`
 	
`interface PortChannel 202`
` description Host1_PortChannel`
` switchport trunk allowed vlan add 80,140`
` mclag 1`
` exit`
 	
`interface Eth 1/3`
` description Host1_PortChannel`
` channel-group 202`
` no shutdown`
` exit`
	`mclag domain 1`
	` source-ip 10.0.2.26`
	` peer-ip 10.0.2.25`
	` peer-link PortChannel 100`
	` delay-restore 90`
	` exit`
	
	`interface PortChannel 202`
	` description Host1_PortChannel`
	` switchport trunk allowed vlan add 80,140`
	` mclag 1`
	` exit`
	
	`interface Eth 1/3`
	` description Host1_PortChannel`
	` channel-group 202`
	` no shutdown`
	` exit`

14. Enable static anycast-address and assign a unique MAC to be used across all leaf switches.
15. Create a loopback interface for the VTEP IP address. This address must be the same for each MCLAG peer.
16. Create a VXLAN interface and assign the same VTEP source IP address to each peer.
17. Assign a unique primary-ip to optimize routing for orphan ports and active-standby hosts. The router-id can be used for this purpose.
18. Map the L2 VNI to the host VLANs that are stretched across the fabric.
19. Create a L3 VNI by mapping the dedicated VLAN (Vlan60) to the non-default VRF. The Network Virtualization Overlay (NVO) is automatically assigned to the VXLAN interface.

    NOTE: The Network Virtualization Overlay (NVO) is automatically assigned to the VXLAN interface

BLeaf25	BLeaf26
```ip anycast-address enable ip anycast-mac-address 00:00:00:00:01:02 interface Loopback 1 description LogicalVTEP ip address 10.10.10.25/32 exit interface vxlan vtep25 source-ip 10.10.10.25 primary-ip 10.0.2.25 map vni 1400 vlan 140 map vni 600 vlan 60 map vni 600 vrf Vrf-tenant1 exit```	```ip anycast-address enable ip anycast-mac-address 00:00:00:00:01:02 interface Loopback 1 description LogicalVTEP ip address 10.10.10.25/32 exit interface vxlan vtep25 source-ip 10.10.10.25 primary-ip 10.0.2.26 map vni 1400 vlan 140 map vni 600 vlan 60 map vni 600 vrf Vrf-tenant1 exit```

20. Configure the BGP router with ECMP enabled.
21. Configure the underlay's address-family and redistribute the connected networks.
22. Add VNI advertisements to the EVPN address family and assign the MCLAG peer's primary-ip.

BLeaf25	BLeaf26
```router bgp 65025 router-id 10.0.2.25 bestpath as-path multipath-relax address-family ipv4 unicast redistribute connected maximum-paths 2 exit address-family l2vpn evpn advertise-all-vni advertise-pip peer-ip 10.0.2.26 exit```	```router bgp 65025 router-id 10.0.2.26 bestpath as-path multipath-relax address-family ipv4 unicast redistribute connected maximum-paths 2 exit address-family l2vpn evpn advertise-all-vni advertise-pip peer-ip 10.0.2.25 exit```

23. Configure the BGP spine peer-group.
24. Set timers and enable unnumbered BGP and BFD on the peer-group.
25. Activate both IPv4 unicast and EVPN address families for the spine peer group.
26. Set `allowas-in` to support MCLAG peer's ASN to be received.

BLeaf25	BLeaf26
```peer-group SPINE advertisement-interval 5 timers 3 9 remote-as external capability extended-nexthop bfd address-family ipv4 unicast activate allowas-in 1 exit address-family l2vpn evpn activate exit exit```	```peer-group SPINE advertisement-interval 5 timers 3 9 remote-as external capability extended-nexthop bfd address-family ipv4 unicast activate allowas-in 1 exit address-family l2vpn evpn activate exit exit```

27. Assign the spine neighbor interfaces to the peer-group.
28. Create the BGP router for each tenant's VRF.
29. Save the configuration.

BLeaf25	BLeaf26

`neighbor interface Eth 1/53`
`  peer-group SPINE`
`  exit`
`  neighbor interface Eth 1/56`
`  peer-group SPINE`
`  exit`
`  exit`
 	
`router bgp 65025 vrf Vrf-tenant1`
`  address-family ipv4 unicast`
`  redistribute connected`
`  exit`
`  address-family l2vpn evpn`
`  advertise ipv4 unicast`
`  exit`
`  exit`
`end`
 	
`write memory`
	`neighbor interface Eth 1/53`
	`  peer-group SPINE`
	`  exit`
	`  neighbor interface Eth 1/56`
	`  peer-group SPINE`
	`  exit`
	`  exit`
	
	`router bgp 65025 vrf Vrf-tenant1`
	`  address-family ipv4 unicast`
	`  redistribute connected`
	`  exit`
	`  address-family l2vpn evpn`
	`  advertise ipv4 unicast`
	`  exit`
	`  exit`
	`end`
	
	`write memory`

Monitoring the POC environment

Enterprise SONiC includes several monitoring capabilities.

About this task

For this POC, syslog and SNMP are used to monitor the environment. Apply the following configurations to each switch node.

Steps

1. If DHCP is not used, configure the out-of-band (OOB) management IP interface and the gateway IP interface for each switch.

```
admin@sonic:~$ sonic-cli
sonic# configure terminal
sonic(config)# interface management 0
sonic(conf-if-eth0)# ip address A.B.C.D/len gwaddr A.B.C.D
```

2. Configure the syslog server.

```
logging server A.B.C.D remote-port 514
```

3. Enable the SNMP agent on the switches:

```
sonic(config)# snmp-server agentaddress A.B.C.D
sonic(config)# snmp-server community public
```

4. Configure one or more SNMP server destinations for SNMPv2 traps (UDP/1024):

```
sonic(config)# snmp-server enable trap
sonic(config)# snmp-server host A.B.C.D community public port 1024
```

5. Save the configuration, and optionally, restart the SNMP service:

```
end
write memory
exit

$ sudo systemctl restart snmp
$ systemctl status snmp
```

If SNMP service is restarted, expect 30 seconds for the service to recover.

Optional POC features

Network Time Protocol (NTP)

```
ntp vrf <mgmt | default>
    ntp server <A.B.C.D>|<A::B>|<hostname>
```

TACACS+

```
tacacs-server host <A.B.C.D> key **** timeout 10 type pap [vrf mgmt]
    aaa authentication login default group tacacs+ local
    aaa authentication failthrough enable
```

RADIUS

```
radius-server host <A.B.C.D> key **** timeout 10 auth-type pap [vrf mgmt]
    aaa authentication login default group radius local
    aaa authentication failthrough enable
```

Verifying the POC

Verify that the POC procedure is successful.

The following table shows the different test cases and the MF-CLI show commands used to verify the POC:

Table 2. Proof of Concept features and show commands verification

Feature	Validation step

EBGP Underlay unnumbered	Verify unnumbered BGP neighbors between spine and leaf:
	`show bgp ipv4 unicast summary`
	`show bgp ipv4 unicast neighbors`

	Verify IPv4 loopback advertisement over unnumbered BGP:
	`show ip route bgp`
	`show ip route connected`

	Verify the usage of peer-group in BGP:
	`show run bgp`
	`show bgp peer-group`

	Verify that the ECMP path is established for the advertised loopback interfaces: `show ip route bgp`
EBGP Overlay	Verify that the VXLAN tunnel is formed between the VTEPs:

Table 2. Proof of Concept features and show commands verification (continued)

Feature	Validation step
	NOTE: A vxlan tunnel may not form until an active host is attached. `show vxlan tunnel`
	Verify L3 VNIs are getting advertised between the VTEPs `show bgp l2vpn evpn`
	Verify the directly connected IPv4 networks on the VRFs are advertised as well as learned on the associated L3 VNI `show ip route vrf Vrf-tenant1`
	Verify the IPv4 routes learned on the VRF are redistributed into the overlay network and learn on the peer VTEPs `show ip route vrf Vrf-tenant1`

	Verify the IPv4 data traffic from the connected interface in one VRF in a leaf device to the same VRF across another leaf device:
	`ping from Host1 to Host2`
	From host1, ping host2 over VLAN 140 and VLAN 50:
	`Host1$ ping -I vlan50 192.168.80.102`
	`Host1$ ping -I vlan140 192.168.80.102`
	`Host1$ ping -I vlan140 192.168.140.102`
	`Host1$ ping -I vlan50 192.168.140.102`

	Commands for additional VXLAN information:
	`show vxlan interface`
	`show evpn`
	`show evpn vni 1400`
	`show evpn vni 600`

High Availability	Verify that the traffic from the host is switched from one leaf to another leaf during a link failure.
	Verify that the traffic in the leaf switches from one spine to another spine if there is a link failure on the link with active traffic.
	Verify that the traffic in the leaf switches to the peer leaf using the connected link when the link to all spines fails.
SNMP	Verify snmpwalk for the interface MIB.

A

Host Configuration

This deployment guide assumes ESXI servers with VMs attached to the leaf switches. However, this appendix provides alternative configurations for bare metal servers using Ubuntu operating system.

Topics:

- Ubuntu host configuration

Ubuntu host configuration

Example of Ubuntu hosts with LACP and tagged VLAN interfaces. Edit the `/etc/netplan/50-cloud-init.yaml` file.

```
$ sudo vi  /etc/netplan/50-cloud-init.yaml

network:
   version: 2
   renderer: network

   ethernets:
      eports:
         match:
            name: ens*

   bonds:
      bond0:
         interfaces: [eports]
         parameters:
            mode: 802.3ad
            lacp-rate: slow
            mii-monitor-interval: 100

   vlans:
      bond0.140:
         id: 140
         link: bond0
         dhcp4: no
         addresses: [192.168.140.102/24]
         gateway4: 192.168.140.254

      bond0.50:
         id: 50
         link: bond0
         dhcp4: no
         addresses: [192.168.50.102/24]
         gateway4: 192.168.50.254

:wq!
```

Follow with:

```
sudo netplan apply-d
sudo reboot
```

B

POC End Host Configurations

This appendix lists the configurations of Dell switches as end hosts to demonstrate end host connectivity into the POC environment and to show connection between the end hosts. For successful validation, the end hosts must be active.

This appendix presents the following topics:

Topics:

Dell switch (Host1)

If servers are not available, then hosts may be emulated with OS10 switches. The following configurations are available for reference.

- Host1
- Host2

```
config t
hostname Host1
spanning-tree disable

interface vlan 50
 ip address 192.168.50.101/24
 exit

interface vlan 140
 ip address 192.168.140.101/24
 exit

interface portchannel 1
 switchport mode trunk
 switchport trunk allowed vlan 50,140
 no shut
 exit

interface ethernet 1/1/1
 no switchport
 channelgroup 1 mode auto
 no shut
 exit

interface ethernet 1/1/2
 no switchport
 channelgroup 1 mode auto
 no shut
 exit

ip route 0.0.0.0/0 192.168.50.254

end
write memory
```

Dell switch (Host2)

```
config t
hostname Host2
spanning-tree disable

interface vlan 80
 ip address 192.168.50.102/24
 exit

interface vlan 140
 ip address 192.168.140.102/24
 exit

interface portchannel 1
 switchport mode trunk
 switchport trunk allowed vlan 80,140
 no shut
 exit

interface ethernet 1/1/1
 no switchport
 channelgroup 1 mode auto
 no shut
 exit

interface ethernet 1/1/2
 no switchport
 channelgroup 1 mode auto
 no shut
 exit

ip route 0.0.0.0/0 192.168.80.254

end
write memory
```

C

Additional Information

Topics:

Dell Technologies documentation

The following Dell Technologies documentation provides additional and relevant information. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell Technologies representative.

- Networking Solutions Info Hub
- Enterprise SONiC Documents (Employee or Partner account login credentials are required)

Support and feedback

For technical support, visit http://www.dell.com/support or call (USA) 1-800-945-3355.

We encourage readers to provide feedback on the quality and usefulness of this publication by sending an email to Dell_Networking_Solutions@Dell.com.