Dell EMC Networking Layer 3 Leaf-Spine Deployment and Best Practices with OS10

Deploying layer 3 leaf-spine networks in the data center with Dell EMC Networking OS10 switches

Dell EMC Networking Infrastructure Solutions
March 2018

Revisions

Date	Rev.	Description	Authors
March 2018	1.0	Initial release	Andrew Waranowski, Curtis Bunch

Table of contents

1   Introduction

Data center networks have traditionally been built in a three-layer hierarchical tree consisting of access, aggregation and core layers as shown in Figure 1.
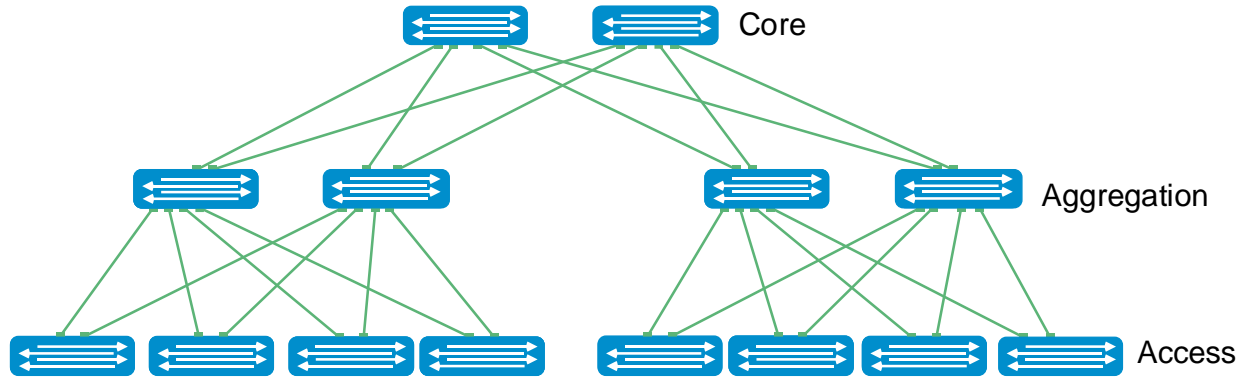
Figure 1   Hierarchical networking model

Due to increasing east-west traffic within the data center (server-server, server-storage, etc.), an alternative to the traditional access-aggregation-core network model is becoming more widely used. This architecture, shown in Figure 2, is known as a leaf-spine network and is a non-blocking network where all devices are exactly the same number of hops away.
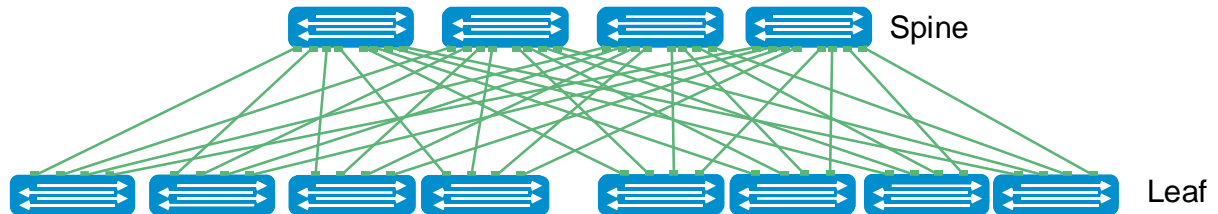
Figure 2   Leaf-spine architecture

In a leaf-spine architecture, the access layer is referred to as the leaf layer. Servers and storage devices connect to leaf switches at this layer. At the next level, the aggregation and core layers are collapsed into a single spine layer. Every leaf switch connects to every spine switch to ensure that all leaf switches are no more than one hop away from one another. This minimizes latency and the likelihood of bottlenecks in the network.

A leaf-spine architecture is highly scalable. As administrators add racks to the data center, a pair of leaf switches are added to each new rack. Spine switches may be added as bandwidth requirements increase. If the initial spine layer is exhausted an additional layer can be deployed creating a 3-tier model.

The interconnections between leaf and spine are dynamically routed. This deployment guide provides step-by-step configuration examples using eBGP or OSPF to provide dynamic routing. It includes examples using Dell EMC Networking switches at both the leaf and spine layers. The objective is to enable a network administrator or engineer to deploy a layer 3 leaf-spine architecture using the examples provided.

DELLEMC

1.1   Typographical Conventions

The command line examples in this document use the following conventions:

`Monospace Text`	CLI examples
<u>`Underlined Monospace Text`</u>	CLI examples that wrap the page. This text is entered as a single command.
`Italic Monospace Text`	Variables in CLI examples
`Bold Monospace Text`	Used to distinguish CLI examples from surrounding text.

DELLEMC

2 Hardware overview

This section briefly describes the hardware used to validate the examples in this guide. A complete listing of hardware and components used is provided in Appendix B.

2.1 Dell EMC Networking S4148F-ON

The Dell EMC Networking S4148F-ON is a 1-Rack Unit (RU) switch with forty-eight 10GbE ports and four 10/25/40/50/100GbE ports. In this guide, Two S4148F-ON switches are used as a leaf switches in the examples in this guide.
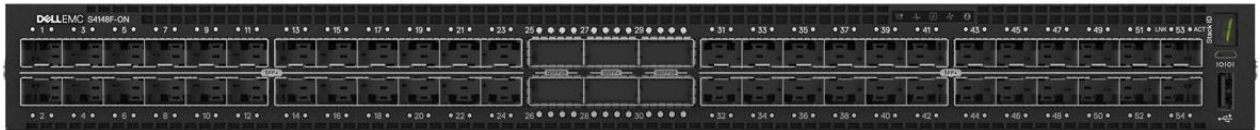
Figure 3   Dell EMC Networking S4148F-ON

2.2 Dell EMC Networking S4248FB-ON

The Dell EMC Networking S4248FB-ON is a 1-RU, multilayer switch with forty 10GbE ports, two 40GbE ports, and six 10/25/40/50/100GbE ports. Two S4248FB-ON switches are used as leaf switches in the examples in this guide.

Figure 4   Dell EMC Networking S4248FB-ON

2.3 Dell EMC Networking Z9100-ON

The Dell EMC Networking Z9100-ON is a 1-RU, multilayer switch with thirty-two 10/25/40/50/100GbE ports plus two 10GbE ports. The Z9100-ON is used as a spine switch in examples in this guide.

Figure 5   Dell EMC Networking Z9100-ON

DELLEMC

3   Leaf-spine overview

The following concepts apply to layer 3 leaf-spine topologies:

- Each leaf switch connects to every spine switch in the topology.
- Servers, storage arrays, edge routers and similar devices always connect to leaf switches, never to spines.

Layer 3 topologies use two leaf switches at the top of each rack configured as a Virtual Link Trunking (VLT) pair. VLT allows all connections to be active while also providing fault tolerance. As administrators add racks to the data center, two leaf switches configured for VLT are added to each new rack.

The total number of leaf-spine connections is equal to the number of leaf switches multiplied by the number of spine switches. To increase fabric bandwidth additional connections between leaf and spine switches can be implemented as long as the spine layer has the capacity for the additional connections.

3.1   Layer 3 leaf-spine topology

In a layer 3 leaf-spine network, traffic between leaf and spine switches is routed. The layer 3 / layer 2 boundary is at the leaf switches. This means at the leaf layer and below (leaf switches and hosts), communication is achieved at layer 2. However, communication at and above the leaf switches is achieved at layer 3. Spine switches are never connected to each other in a layer 3 topology. Equal cost multi-path routing (ECMP) is used to load balance traffic across the layer 3 network. Connections within racks from hosts to leaf switches are layer 2. Connections to external networks are made from a pair of edge or border leaf switches as shown in Figure 6. Connections from the data center core can also be made directly to the spines.
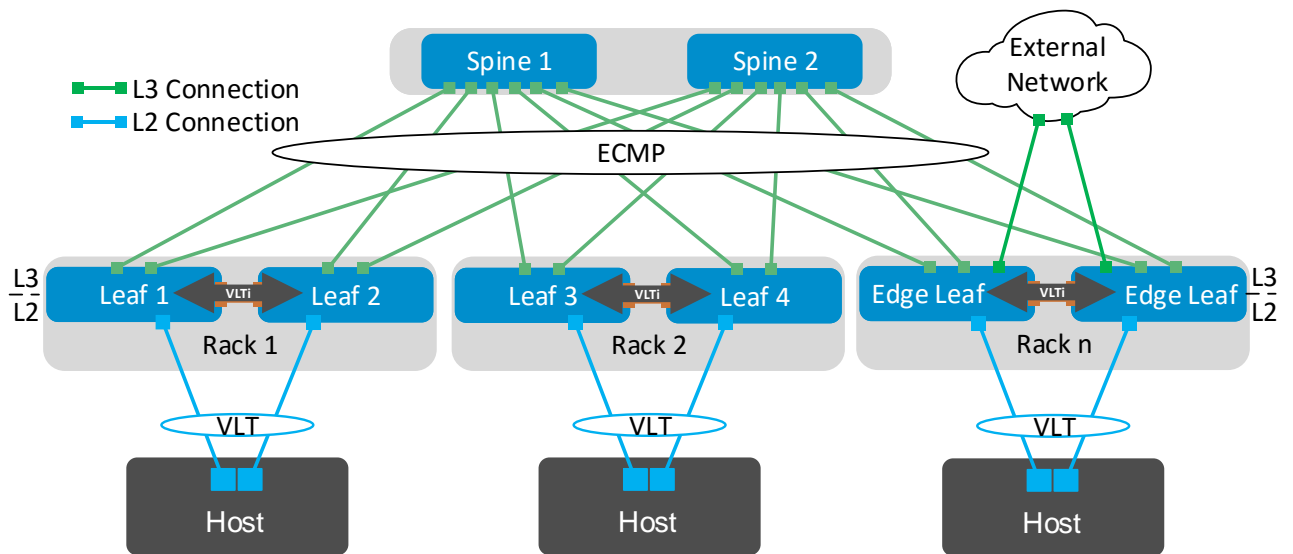
Figure 6   Layer 3 leaf-spine network

D&LLEMC

4 Protocols used in the leaf-spine examples

This section provides an overview of the protocols used in constructing the leaf-spine network examples in this guide.

- VLT, Section 4.1
- LACP/LAG, Section 4.2
- Uplink Failure Detection (UFD), Section 4.3
- RSTP, Section 4.4
- Routing protocols, Section 4.5
    - Border Gateway Protocol (BGP)
    - Open Shortest Path First (OSPF)
- VRRP, Section 4.6
- ECMP, Section 4.7

4.1 Virtual Link Trunking (VLT)

VLT allows link aggregation group (LAG) terminations on two separate switches and supports a loop-free topology. The two switches are referred to as VLT peers and are kept synchronized via an inter-switch link (ISL) called the VLT interconnect (VLTi). A separate backup link maintains heartbeat messages across the OOB management network or using a point-to-point link between peers.

VLT provides layer 2 multipathing and load-balances traffic. VLT offers the following additional benefits:

- Eliminates spanning tree-blocked ports
- Uses all available uplink bandwidth
- Enables fast path switchover if either a link or device fails
- Ensures high availability

Note: Downstream connections from leaf switches configured for VLT do not necessarily have to be configured as LAGs if other fault tolerant methods are preferred (e.g. multipath IO). In this guide, examples 1 and 2 use LAGs to downstream servers while examples 3 and 4 do not.

4.2 LACP/LAG

Link Aggregation Group (LAG) bundles multiple links into a single interface to increase bandwidth between two devices. LAGs also provide redundancy via the multiple paths. In a leaf-spine network, LAGs are typically used to attach servers or storage devices to the VLT leaf pairs.

Link Aggregation Control Protocol (LACP) is an improvement over static LAGs in that the protocol will automatically failover if there is a connectivity issue. This is especially important if the links traverse a media converter where it is possible to lose Ethernet connectivity while links remain in an Up state.

DELLEMC

4.3 Uplink Failure Detection (UFD)

If a leaf switch loses all connectivity to the spine layer, by default the attached hosts continue to send traffic to that leaf without a direct path to the destination. The VLTi link to the peer leaf switch handles traffic during such a network outage, but this is not considered a best practice.

Dell EMC recommends enabling UFD, which detects the loss of upstream connectivity. An uplink state group is configured on each leaf switch, which creates an association between the uplinks to the spines and the downlink interfaces.

In the event all uplinks fail on a switch, UFD automatically shuts down the downstream interfaces. This propagates to the hosts attached to the leaf switch. The host then uses its link to the remaining switch to continue sending traffic across the leaf-spine network.

4.3.1 UFD vs. iBGP at the leaf layer

Some leaf and spine implementations make use of iBGP between leaf switches instead of UFD. If UFD is not used, it is possible, due to hashing, for packets to enter a leaf switch which does not have functioning uplinks. If iBGP is enabled between the leaf switches, it is possible for them to route packets to each other, provided there are valid routes. However, the designs in this document make use of UFD, because it reduces network and configuration complexity and simplifies path determination.

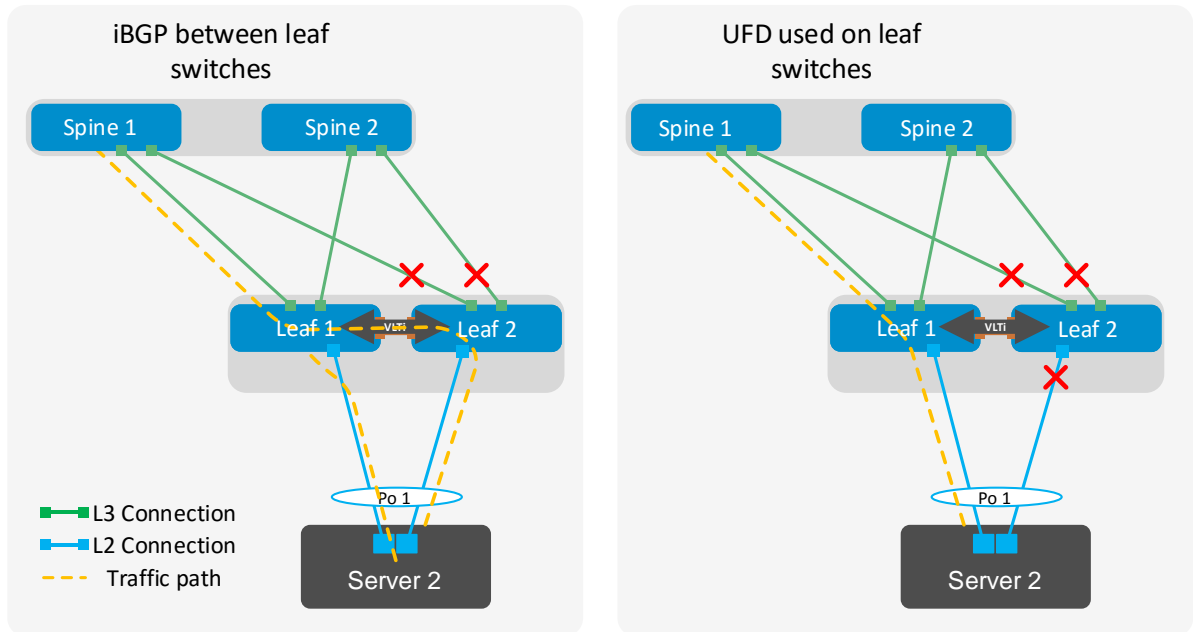Figure 7   Using iBGP between leaf switches vs UFD on leaf switches provides more path options at the cost of greater network and configuration complexity

4.4 Rapid Spanning Tree Protocol (RSTP)

As a precautionary measure, Dell EMC recommends enabling RSTP on all switches that have layer 2 interfaces. Even though VLT environments are loop-free, STP should be configured as a best practice in case

DELLEMC

of switch misconfiguration or improperly connected cables. In properly configured and connected leaf-spine networks, there are no ports blocked by Spanning Tree Protocol.

4.5 Routing protocols

Any of the following routing protocols may be used on layer 3 connections when designing a leaf-spine network:

- OSPF
- BGP

4.5.1 Border Gateway Protocol (BGP)

BGP may be selected for scalability and is well suited for very large networks. BGP can be configured as External BGP (eBGP) to route between autonomous systems or Internal BGP (iBGP) to route within a single autonomous system.

Layer 3 leaf-spine networks use ECMP routing. eBGP and iBGP handle ECMP differently. By default, eBGP supports ECMP without any adjustments. To keep configuration complexity to a minimum, Dell EMC recommends eBGP in leaf-spine fabric deployments.

BGP tracks IP reachability to the peer remote address and the peer local address. Whenever either address becomes unreachable, BGP brings down the session with the peer. To ensure fast convergence with BGP, Dell EMC recommends enabling Neighbor fall-over with BGP. Neighbor fall-over terminates BGP sessions of any directly adjacent peer if the link to reach the peer goes down without waiting for the hold-down timer to expire.

4.5.2 Open Shortest Path First (OSPF)

OSPF is an interior gateway protocol that provides routing inside an autonomous network. OSPF routers send link-state advertisements to all other routers within the same autonomous system areas. While generally more memory and CPU intensive than BGP, OSPF may offer faster convergence. OSPF is often used in smaller networks (100 OSPF routers, depending on various factors).

4.6 Virtual Router Redundancy Protocol (VRRP)

VRRP is a first hop redundancy protocol. It provides gateway redundancy by enabling a pair of VRRP routers to coordinate and act as one gateway. This way, if one VRRP router fails, the other router will detect its missing peer, and become the active gateway.

Although VRRP is traditionally an active/standby protocol, wherein only one VRRP router does the forwarding at a time, when used between VLT peers, it becomes active/active. Even the backup VRRP router will forward packets for hosts and end devices.

In this guide, VRRP version 3 is specified. VRRP version 3 supports IPv6. VRRP version 3 requires higher priority for preemption (with version 2, if both routers have the same priority, the router with the higher IP address will preempt). Version 3 also uses millisecond timers for improved tunability.

DELLEMC

4.7 Equal Cost Multi-Path (ECMP)

The nature of a leaf-spine topology is that leaf switches are no more than one hop away from each other. As shown in Figure 8, Leaf 1 has two equal cost paths to Leaf 4, one through each spine. The same is true for all leaf switches.
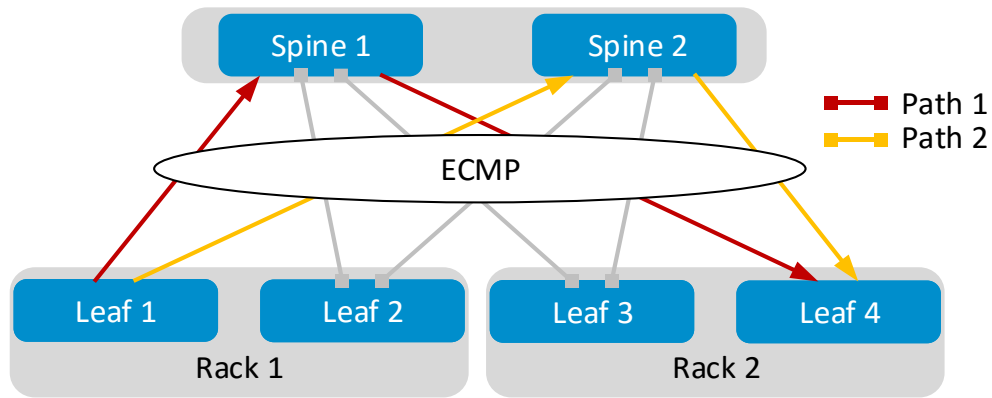
Figure 8　　Use of ECMP in a layer 3 topology

ECMP is a routing technique used in a layer 3 leaf-spine topology for load balancing packets along these multiple equal cost paths. ECMP is enabled on all leaf and spine switches, allowing traffic between leaf switches to be load balanced across the spine switches.

5 Layer 3 configuration planning

5.1 BGP ASN configuration

When eBGP is used, an autonomous system number (ASN) is assigned to each switch. Valid private, 2-byte ASNs range from 64512 through 65534. Figure 9 shows the ASN assignments used for leaf and spine switches in the BGP examples in this guide.
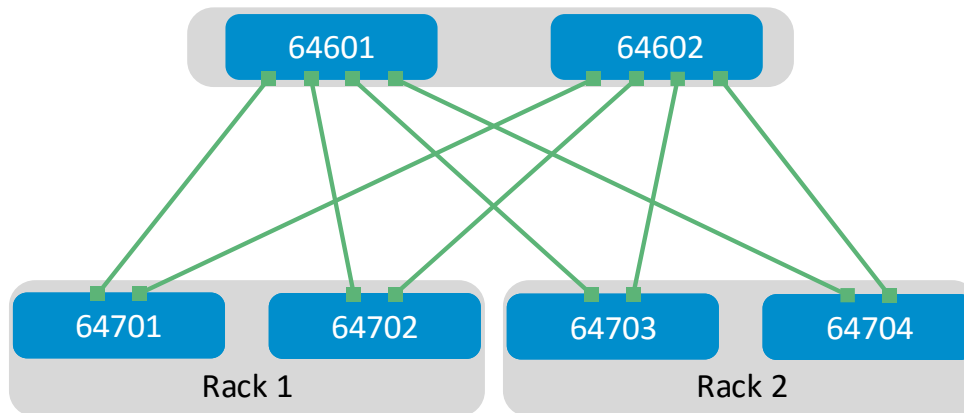
Figure 9   BGP ASN assignments

ASNs should follow a logical pattern for ease of administration and allow for growth as additional leaf and spine switches are added. In this example, an ASN with a "6" in the hundreds place represents a spine switch (e.g. 64**6**01) and an ASN with a "7" in the hundreds place represents a leaf switch (e.g. 64**7**01).

Note: The same ASN can be used across all tier-2 spine switches if the growth plans do not require an additional layer of spine switches.

5.2 IP addressing

Establishing a logical, scalable IP address scheme is important before deploying a leaf-spine topology. This section covers the IP addressing used in the layer 3 examples in this guide.

5.2.1 Loopback addresses

Loopback addresses may be used as router IDs when configuring routing protocols. As with ASNs, loopback addresses should follow a logical pattern that will make it easier for administrators to manage the network and allow for growth. Figure 10 shows the loopback addresses used as router IDs in the BGP and OSPF examples in this guide.

DELLEMC
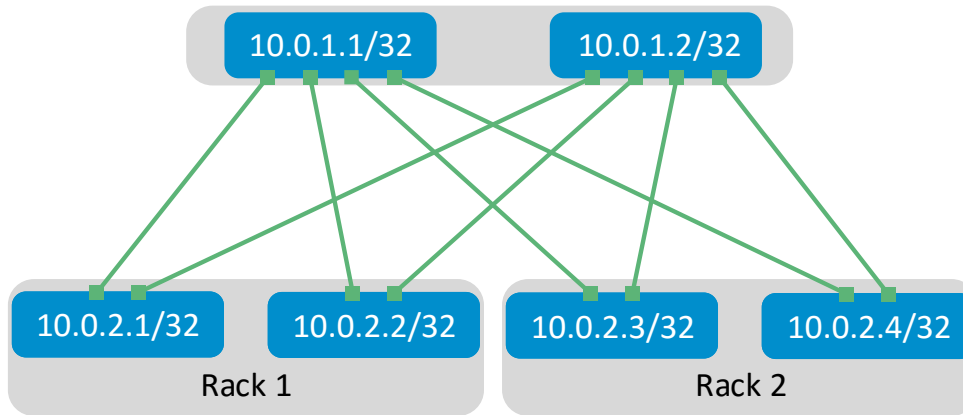
Figure 10   Loopback addressing

All loopback addresses used are part of the 10.0.0.0/8 address space with each address using a 32-bit mask. In this example, the third octet represents the layer, "1" for spine and "2" for leaf. The fourth octet is the counter for the appropriate layer. For example, 10.0.**1**.1/32 is the first spine switch in the topology while 10.0.**2**.4/32 is the fourth leaf switch.

5.2.2   Point-to-point addresses

Table 1 lists layer 3 connection details for each leaf and spine switch.

All addresses come from the same base IP prefix, 192.168.0.0/16 with the third octet representing the spine number. For example, 192.168.**1**.0/31 is a two-host subnet connected to Spine 1 while 192.168.**2**.0/31 is connected to Spine 2. The /31 net mask is used to conserve IPs. This IP scheme is easily extended as leaf and spine switches are added to the network.

Link labels are provided in the table for quick reference with Figure 11.

Table 1   Interface and IP configuration

Link Label	Source switch	Source interface	Source IP	Network	Destination switch	Destination interface	Destination IP
A	Leaf 1	Eth1/1/25	.1	192.168.1.0/31	Spine 1	Eth1/1/1	.0
B	Leaf 1	Eth1/1/26	.1	192.168.2.0/31	Spine 2	Eth1/1/1	.0
C	Leaf 2	Eth1/1/25	.3	192.168.1.2/31	Spine 1	Eth1/1/2	.2
D	Leaf 2	Eth1/1/26	.3	192.168.2.2/31	Spine 2	Eth1/1/2	.2
E	Leaf 3	Eth1/1/43	.5	192.168.1.4/31	Spine 1	Eth1/1/3	.4
F	Leaf 3	Eth1/1/44	.5	192.168.2.4/31	Spine 2	Eth1/1/3	.4
G	Leaf 4	Eth1/1/43	.7	192.168.1.6/31	Spine 1	Eth1/1/4	.6
H	Leaf 4	Eth1/1/44	.7	192.168.2.6/31	Spine 2	Eth1/1/4	.6

DELLEMC

Note: Link labels A-D refer to S4148F switch ports, while link labels E-H refer to S4248FB switch ports.

The point-to-point IP addresses used in this guide are shown in Figure 11:
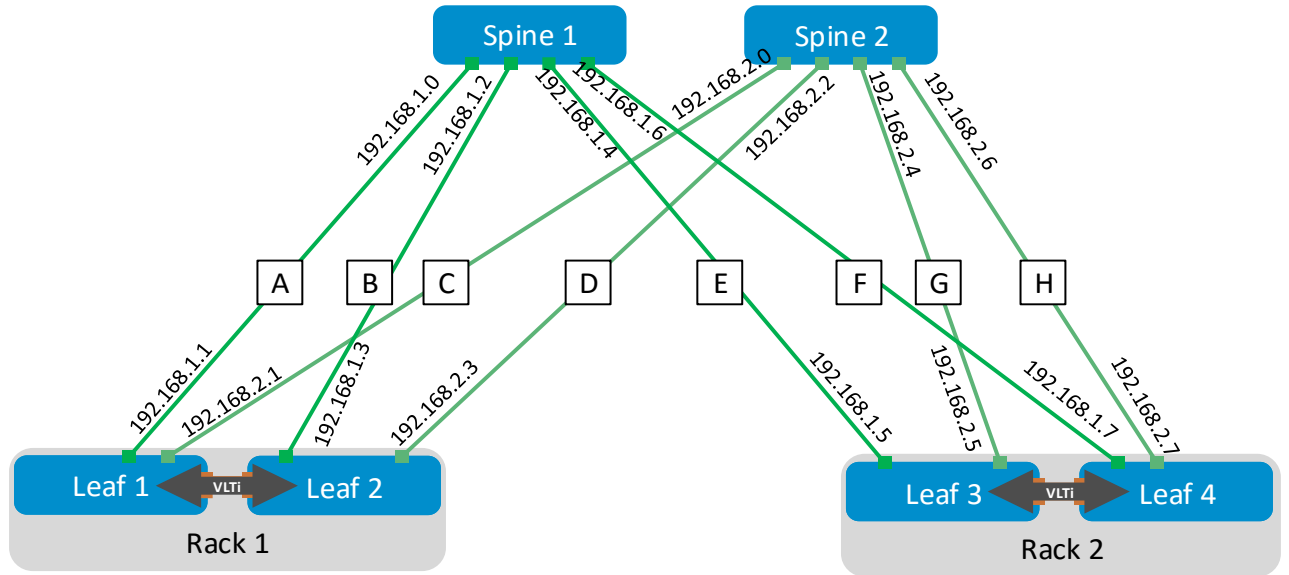
Figure 11   Point-to-point IP addresses

6 Example 1: Layer 3 with Dell EMC leaf and spine switches using OSPF

This section covers deploying a leaf-spine environment using OSPF with S4148F-ON and S4248FB-ON switches at the leaf layer and Z9100-ON switches used at the spine layer. While the S4148F-ON configuration is shown in this section the configuration provides OSPF configuration examples to build the layer 3 leaf-spine topology shown in Figure 12. While the S4148F-ON configuration is shown in this section the configuration details for the S4248FB-ON switches are attached.
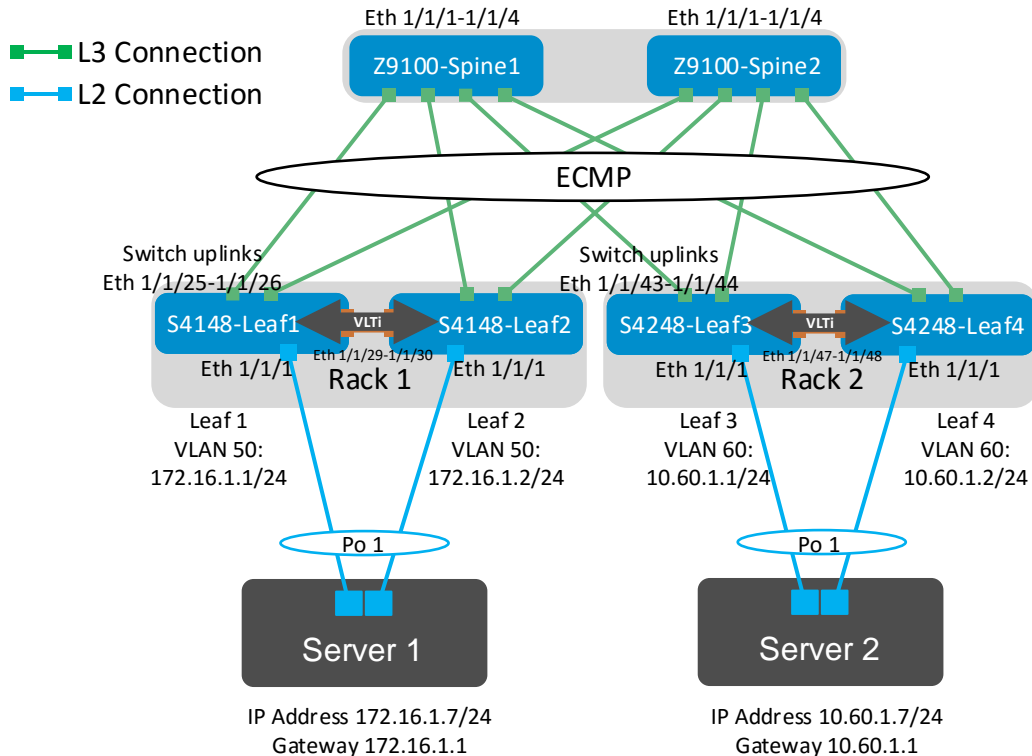
Figure 12   Example 1: Layer 3 leaf-spine topology

In Rack 1, VLAN 50 is used and devices in VLAN 50 are assigned IP addresses on the 172.16.1.0/24 network. Because VRRP with VLT is configured on S4148F-Leaf1 and S4148F-Leaf2, Server 1 may specify the virtual IP address (VIP) assigned to VLAN 50 on both leaf switch, 172.16.1.254, as its default gateway. Traffic is load balanced across both leaf switches.

Rack 2 is configured in an identical manner, except VLAN 60 is used and devices in VLAN 60 are assigned IP addresses on the 10.60.1.0/24 network. Server 2 may specify the VLAN 60 VIP address configured on the leaf switches 10.60.1.254, as its default gateway.

Note: All switch configuration files for the topology in Figure 12 are contained in the attachment named **Example1_config_files.pdf.** The files may be edited as needed in a plain text editor and commands pasted directly into switch consoles. Dell EMC Networking switches start at their factory default settings per Appendix A.

6.1   S4148F-ON leaf switch configuration

The following configuration details are for S4148F-Leaf1 and S4148F-Leaf2 in Figure 12. The configuration commands for S4248FB-Leaf3 and S4248FB-Leaf4 are similar and are provided in the attachments.

Note: On S4148F-ON switches running OS10, Telnet is disabled and SSH is enabled by default. The S4148F-ON has default credentials of admin/admin. Therefore, the switch is accessible by default via SSH with those credentials. Changing the admin password to something more complex is recommended.

Set the host name, configure the OOB management interface and default gateway. Enable RSTP as a precaution. S4148F-Leaf1 is configured as the primary RSTP root bridge using the `spanning-tree priority 0` command. S4148F-Leaf2 is configured as the secondary RSTP root bridge using the `spanning-tree priority 4096` command.

S4148F-Leaf1	S4148F-Leaf2

`enable`
`configure t`
 	
`hostname S4148F-Leaf1`
 	
`interface mgmt 1/1/1`
`no ip address`
`ip address 100.67.170.30/24`
`no shutdown`
<u>`management route 0.0.0.0/0`</u>
<u>`100.67.170.254`</u>
 	
`spanning-tree mode rstp`
`spanning-tree rstp priority 0`
	`enable`
	`configure t`
	
	`hostname S4148F-Leaf2`
	
	`interface mgmt 1/1/1`
	`no ip address`
	`ip address 100.67.170.29/24`
	`no shutdown`
	<u>`management route 0.0.0.0/0`</u>
	<u>`100.67.170.254`</u>
	
	`spanning-tree mode rstp`
	`spanning-tree rstp priority 4096`

Configure the VLT interconnect between S4148F-Leaf1 and S4148F-Leaf2. In this configuration, remove each interface from layer 2 mode with the `no switchport` command for interfaces eth 1/1/29-1/1/30. Then enter the VLT domain and add the backup destination. Next add the interfaces participating in VLTi. Finally, enable VLT peer routing.

S4148F-Leaf1	S4148F-Leaf2
| ```
interface range ethernet 1/1/29-1/1/30
no switchport

vlt-domain 127
backup destination 100.67.170.29
discovery-interface ethernet 1/1/29
discovery-interface ethernet 1/1/30
peer-routing
``` | ```
interface range ethernet 1/1/29-1/1/30
no switchport

vlt-domain 127
backup destination 100.67.170.30
discovery-interface ethernet 1/1/29
discovery-interface ethernet 1/1/30
peer-routing
``` |

Create a server-facing VLAN interface. Use the same VLAN ID on both leaf switches. Assign an IP address to the VLAN interface. The address must be unique but on the same network on both leaf switches. Configure VRRP to use VRRP version 3. Create a VRRP group and specify the group's virtual IP address.

Note: In this example, Server 1's NIC is configured as an LACP NIC team. It is assigned the IP address 172.16.1.7/24. The VRRP VIP address, 172.16.1.254, is specified as Server 1's default gateway.

S4148F-Leaf1	S4148F-Leaf2
| ```
interface Vlan 50
ip address 172.16.1.1/24
no shutdown

exit
vrrp version 3
interface vlan 50
vrrp-group 50
virtual-address 172.16.1.254
``` | ```
interface Vlan 50
ip address 172.16.1.2/24
no shutdown

exit
vrrp version 3
interface vlan 50
vrrp-group 50
virtual-address 172.16.1.254
``` |

Configure each downstream server-facing interface with an LACP port channel. Configure each port channel for VLT. Port channel 10 connects downstream to Server 1 and is configured as an RSTP edge port.

S4148F-Leaf1	S4148F-Leaf2
| ```
interface port-channel 10
description "Server 1"
switchport access vlan 50
vlt-port-channel 10
spanning-tree port type edge

interface ethernet 1/1/1
description "Server 1"
channel-group 10 mode active
no shutdown
``` | ```
interface port-channel 10
description "Server 1"
switchport access vlan 50
vlt-port-channel 10
spanning-tree port type edge

interface ethernet 1/1/1
description "Server 1"
channel-group 10 mode active
no shutdown
``` |

The two upstream layer 3 interfaces connected to the spine switches are configured. Assign IP addresses per Table 1. Configure a loopback interface to be used as the router ID. This is used with BGP or OSPF.

Note: If multiple loopback interfaces exist on a system, the interface with the highest numbered IP address is used as the router ID. This configuration only uses one loopback interface.

S4148F-Leaf1	S4148F-Leaf2
| ```
interface ethernet 1/1/25
description "Z9100-Spine1 eth 1/1/1"
no switchport
ip address 192.168.1.1/31
ip ospf 1 area 0
no shutdown

interface ethernet 1/1/26
description "Z9100-Spine2 eth 1/1/1"
no switchport
ip address 192.168.2.1/31
ip ospf 1 area 0
no shutdown

interface loopback 0
description "Router ID"
ip address 10.0.2.1/32
no shutdown
``` | ```
interface ethernet 1/1/25
description "Z9100-Spine1 eth 1/1/2"
no switchport
ip address 192.168.1.3/31
ip ospf 1 area 0
no shutdown

interface ethernet 1/1/26
description "Z9100-Spine2 eth 1/1/2"
no switchport
ip address 192.168.2.3/31
ip ospf 1 area 0
no shutdown

interface loopback 0
description "Router ID"
ip address 10.0.2.2/32
no shutdown
``` |

Configure a route map and IP prefix-list to redistribute all loopback addresses and leaf networks via OSPF.

The command `seq 10 permit 10.0.0.0/8 ge 24` includes all addresses in the 10.0.0.0/8 address range with a mask greater than or equal to 24. This includes all loopback addresses used as router IDs as well as the 10.60.1.0/24 network used on leaf switches 3 and 4 as shown in Figure 12.

The command `seq 20 permit 172.16.0.0/16 ge 24` includes the 172.16.1.0/24 network used on leaf switches 1 and 2 as shown in Figure 12.

S4148F-Leaf1	S4148F-Leaf2
| ```
route-map spine-leaf permit 10
match ip address prefix-list spine-
leaf

ip prefix-list spine-leaf seq 10
permit 10.0.0.0/8 ge 24

ip prefix-list spine-leaf seq 20
permit 172.16.0.0/16 ge 24
``` | ```
route-map spine-leaf permit 10
match ip address prefix-list spine-
leaf

ip prefix-list spine-leaf seq 10
permit 10.0.0.0/8 ge 24

ip prefix-list spine-leaf seq 20
permit 172.16.0.0/16 ge 24
``` |

Configure UFD. This shuts the downstream interfaces if all uplinks fail. The hosts attached to the switch use the remaining LACP LAG member to continue sending traffic across the fabric.

DELLEMC

S4148F-Leaf1	S4148F-Leaf2
uplink-state-group 1 description "Disable downstream ports in event all uplinks fail" downstream port-channel 10 upstream ethernet 1/1/25 upstream ethernet 1/1/26 enable	uplink-state-group 1 description "Disable downstream ports in event all uplinks fail" downstream port-channel 10 upstream ethernet 1/1/25 upstream ethernet 1/1/26 enable

Enable OSPF globally and redistribute connected networks according to the previously defined route map. Finally, exit configuration mode and save the configuration.

S4148F-Leaf1	S4148F-Leaf2
router ospf 1 redistribute connected route-map spine-leaf end write memory	router ospf 1 redistribute connected route-map spine-leaf end write memory

6.2   Z9100-ON spine switch configuration

The following configuration details are for Z9100-ON-Spine1 and Z9100-ON-Spine2 in Figure 12.

Note: On S4148F-ON switches running OS10, Telnet is disabled and SSH is enabled by default. The S4148F-ON has default credentials of admin/admin. Therefore, the switch is accessible by default via SSH with those credentials. Changing the admin password to something more complex is recommended.

Set the host name, configure the OOB management interface and default gateway.

Z9100-Spine1	Z9100-Spine2
enable configure t hostname Z9100-Spine1 interface mgmt 1/1/1 no ip address ip address 100.67.173.36/24 no shutdown management route 0.0.0.0/0 100.67.173.254	enable configure t hostname Z9100-Spine2 interface mgmt 1/1/1 no ip address ip address 100.67.173.35/24 no shutdown management route 0.0.0.0/0 100.67.173.254

DELLEMC

Configure the four point-to-point interfaces connected to leaf switches. Assign IP addresses per Table 1. Configure a loopback interface to be used as the router ID. Enable OSPF on the interfaces with process 1 area 0.

> **Note**: If multiple loopback interfaces exist on a system, the interface with the highest numbered IP address is used as the router ID. This configuration only uses one loopback interface.

Z9100-Spine1	Z9100-Spine2
interface ethernet 1/1/1 description "S4148F-Leaf1 eth 1/1/29" no switchport ip address 192.168.1.0/31 ip ospf 1 area 0 no shutdown interface ethernet 1/1/2 description "S4148F-Leaf2 1/1/29" no switchport ip address 192.168.1.2/31 ip ospf 1 area 0 no shutdown interface ethernet 1/1/3 description "S4248FB-Leaf3 eth 1/1/47" no switchport ip address 192.168.1.4/31 ip ospf 1 area 0 no shutdown interface ethernet 1/1/4 description "S4248FB-Leaf4 eth 1/1/47" no switchport ip address 192.168.1.6/31 ip ospf 1 area 0 no shutdown interface loopback 0 description "Router ID" ip address 10.0.1.1/32 no shutdown	interface ethernet 1/1/1 description "S4148F-Leaf1 eth 1/1/30" no switchport ip address 192.168.2.0/31 ip ospf 1 area 0 no shutdown interface ethernet 1/1/2 description "S4148F-Leaf2 eth 1/1/30" no switchport ip address 192.168.2.2/31 ip ospf 1 area 0 no shutdown interface ethernet 1/1/3 description "S4248FB-Leaf3 eth 1/1/48" no switchport ip address 192.168.2.4/31 ip ospf 1 area 0 no shutdown interface ethernet 1/1/4 description "S4248FB-Leaf4 eth 1/1/48" no switchport ip address 192.168.2.6/31 ip ospf 1 area 0 no shutdown interface loopback 0 description "Router ID" ip address 10.0.1.2/32 no shutdown

Configure a route map and IP prefix-list to redistribute all loopback addresses and leaf networks via BGP or OSPF.

The command `seq 10 permit 10.0.0.0/8 ge 24` includes all addresses in the 10.0.0.0/8 address range with a mask greater than or equal to 24. This includes all loopback addresses used as router IDs as well as the 10.60.1.0/24 network used on leaf switches 3 and 4 as shown in Figure 12.

The command `seq 20 permit 172.16.0.0/16 ge 24` includes the 172.16.1.0/24 network used on leaf switches 1 and 2 as shown in Figure 12.

Z9100-Spine1	Z9100-Spine2
route-map spine-leaf permit 10 match ip address prefix-list spine-leaf ip prefix-list spine-leaf seq 10 permit 10.0.0.0/8 ge 24 ip prefix-list spine-leaf seq 20 permit 172.16.0.0/16 ge 24	route-map spine-leaf permit 10 match ip address prefix-list spine-leaf ip prefix-list spine-leaf seq 10 permit 10.0.0.0/8 ge 24 ip prefix-list spine-leaf seq 20 permit 172.16.0.0/16 ge 24

Enable OSPF globally and redistribute connected networks according to the previously defined route map. Finally, exit configuration mode and save the configuration.

Z9100-Spine1	Z9100-Spine2
router ospf 1 redistribute connected route-map spine-leaf end write memory	router ospf 1 redistribute connected route-map spine-leaf end write memory

DELLEMC

6.3 Example 1 validation

In addition to sending traffic between hosts, the configuration shown in Figure 12 can be validated with the commands shown in this section. For more information on commands and output, see the Command Line Reference Guide for the applicable switch (links to documentation are provided in Appendix C).

Command and output examples are provided for one spine and one leaf switch. Command output on other switches is similar.

6.3.1 show ip ospf neighbor

This command shows the state of all connected OSPF neighbors. In this configuration, each spine has four neighbors (the four leaf switches) and each leaf switch has two neighbors (the two spine switches).

```
Z9100-Spine1#show ip ospf neighbor
Neighbor ID  Pri    State     Dead Time Address        Interface        Area
--------------------------------------------------------------------------------
10.0.2.1  1    FULL/DR      00:00:30  192.168.1.1    ethernet1/1/1      0.0.0.0
10.0.2.2  1    FULL/DR      00:00:33  192.168.1.3    ethernet1/1/2      0.0.0.0
10.0.2.3  1    FULL/DR      00:00:36  192.168.1.5    ethernet1/1/3      0.0.0.0
10.0.2.4  1    FULL/DR      00:00:39  192.168.1.7    ethernet1/1/4      0.0.0.0

S4148F-Leaf1#show ip ospf neighbor
Neighbor ID    Pri         State      Dead Time Address  Interface        Area
--------------------------------------------------------------------------------
10.0.1.1  1    FULL/BDR     00:00:30  192.168.1.0    ethernet1/1/25     0.0.0.0
10.0.1.2  1    FULL/BDR     00:00:33  192.168.2.0    ethernet1/1/26     0.0.0.0
```

DELLEMC

6.3.2 show ip route ospf

This command is used to verify the OSPF entries in the Routing Information Base (RIB). Entries with multiple paths shown are used with ECMP. The two server networks in this example, 10.60.1.0 and 172.16.1.0, each have two paths from Z9100-Spine1, one through each leaf switch.

The first set of routes with a subnet mask of /32 are the IPs configured for router IDs.

```
Z9100-Spine1#show ip route ospf
Codes: C - connected
       S - static
       B - BGP, IN - internal BGP, EX - external BGP
       O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
       N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
       E2 - OSPF external type 2, * - candidate default,
       + - summary route, > - non-active route
Gateway of last resort is not set
  Destination          Gateway                          Dist/Metric Last Change
--------------------------------------------------------------------------------
  O  E2 10.0.1.2/32    via 192.168.1.1    ethernet1/1/1      110/20 00:15:27
                       via 192.168.1.3    ethernet1/1/2
                       via 192.168.1.5    ethernet1/1/3
                       via 192.168.1.7    ethernet1/1/4
  O  E2 10.0.2.1/32    via 192.168.1.1    ethernet1/1/1      110/20 00:15:32
  O  E2 10.0.2.2/32    via 192.168.1.3    ethernet1/1/2      110/20 00:15:32
  O  E2 10.0.2.3/32    via 192.168.1.5    ethernet1/1/3      110/20 00:15:32
  O  E2 10.0.2.4/32    via 192.168.1.7    ethernet1/1/4      110/20 00:15:32
  O  E2 10.60.1.0/24   via 192.168.1.5    ethernet1/1/3      110/20 00:15:32
                       via 192.168.1.7    ethernet1/1/4
  O  E2 172.16.1.0/24  via 192.168.1.1    ethernet1/1/1      110/20 00:15:32
                       via 192.168.1.3    ethernet1/1/2
  O     192.168.2.0/31 via 192.168.1.1    ethernet1/1/1      110/2  00:15:32
  O     192.168.2.2/31 via 192.168.1.3    ethernet1/1/2      110/2  00:15:32
  O     192.168.2.4/31 via 192.168.1.5    ethernet1/1/3      110/2  00:15:32
  O     192.168.2.6/31 via 192.168.1.7    ethernet1/1/4      110/2  00:15:32
```

S4148F-Leaf1 has two paths to all other leaf switches and two paths to Server 2's network, 10.60.1.0. There is one path through each spine switch.

```
S4148F-Leaf1# show ip route ospf
Codes: C - connected
       S - static
       B - BGP, IN - internal BGP, EX - external BGP
       O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
       N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
```

```
          E2 - OSPF external type 2, * - candidate default,
          + - summary route, > - non-active route
Gateway of last resort is not set
    Destination          Gateway                         Dist/Metric Last Change
  -------------------------------------------------------------------------------
    O  E2 10.0.1.1/32     via 192.168.1.0      ethernet1/1/25      110/20   00:15:34
    O  E2 10.0.1.2/32     via 192.168.2.0      ethernet1/1/26      110/20   00:15:27
    O  E2 10.0.2.2/32     via 192.168.1.0      ethernet1/1/25      110/20   00:15:27
                          via 192.168.2.0      ethernet1/1/26
    O  E2 10.0.2.3/32     via 192.168.1.0      ethernet1/1/25      110/20   00:15:27
                          via 192.168.2.0      ethernet1/1/26
    O  E2 10.0.2.4/32     via 192.168.1.0      ethernet1/1/25      110/20   00:15:27
                          via 192.168.2.0      ethernet1/1/26
    O  E2 10.60.1.0/24    via 192.168.1.0      ethernet1/1/25      110/20   00:15:27
                          via 192.168.2.0      ethernet1/1/26
    O     192.168.1.2/31  via 192.168.1.0      ethernet1/1/25      110/2    00:15:34
    O     192.168.1.4/31  via 192.168.1.0      ethernet1/1/25      110/2    00:15:34
    O     192.168.1.6/31  via 192.168.1.0      ethernet1/1/25      110/2    00:15:34
    O     192.168.2.2/31  via 192.168.2.0      ethernet1/1/26      110/2    00:15:27
    O     192.168.2.4/31  via 192.168.2.0      ethernet1/1/26      110/2    00:15:27
    O     192.168.2.6/31  via 192.168.2.0      ethernet1/1/26      110/2    00:15:27
```

Note: The command `show ip route <cr>` can also be used to verify the information above as well as static routes and direct connections.

6.3.3   show vlt

This command is used to validate VLT configuration status on leaf switches in this topology. The Inter-chassis link (ICL) Link Status, Heart Beat Status and VLT Peer Status must all be up. The role for one switch in the VLT pair is primary and its peer switch (shown here) is assigned the secondary role. Ensure Peer-Routing is shown as enabled.

```
S4148F-Leaf1#show vlt 127
Domain ID                 : 127
Unit ID                   : 2
Role                      : secondary
Version                   : 1.0
Local System MAC address  : 14:18:77:25:49:b9
VLT MAC address           : 14:18:77:25:56:b9
IP address                : fda5:74c8:b79e:1::2
Delay-Restore timer       : 90 seconds
Peer-Routing              : Enabled
Peer-Routing-Timeout timer : 0 seconds
VLTi Link Status
    port-channel1000      : up
```

DELLEMC

```
VLT Peer Unit ID      System MAC Address      Status      IP Address           Version
--------------------------------------------------------------------------------
    1                 14:18:77:25:56:b9       up          fda5:74c8:b79e:1::1    1.0
```

6.3.4   show vlt vlt-port-detail

This command is used to validate VLT LAG status on leaf switches in this topology. This command shows the status and active VLANs of all VLT LAGs (Port channel 10 in this example). The local and peer status must both be up.

```
S4148F-Leaf1#show vlt 127 vlt-port-detail
vlt-port-channel ID : 10
VLT Unit ID      Port-Channel      Status      Configured ports      Active ports
--------------------------------------------------------------------------------
   1             port-channel10    up          1                     0
 * 2             port-channel10    up          1                     0
```

6.3.5   show vlt mismatch

This command highlights configuration issues between VLT peers. Mismatch examples include incompatible VLT configuration settings, VLAN differences, different switch operating system versions and spanning-tree inconsistencies. There should be no mismatches reported on any switch configured for VLT. If there is, resolve the mismatch.

```
S4148F-Leaf1#show vlt 127 mismatch
Peer-routing mismatch:
No mismatch

VLAN mismatch:
No mismatch

VLT VLAN mismatch:
No mismatch
```

6.3.6   show uplink-state-group

This command is used to validate the UFD status on leaf switches in this topology. `Status: Enabled, Up` indicates UFD is enabled and no interfaces are currently disabled by UFD.

```
S4148F-Leaf1#show uplink-state-group

Uplink State Group: 1, Status: Enabled,up
```

If an interface happens to be disabled by UFD, the `show uplink-state-group` command output will appear as follows:

```
Uplink State Group: 1, Status: Enabled,down
```

DELLEMC

> **Note:** When an interface has been disabled by UFD, the `show interfaces` *interface* command for affected interfaces indicates it is error-disabled as follows:
> ```
> S4148F-Leaf-1#show interfaces te 1/4
> TenGigabitEthernet 1/4 is up, line protocol is down(error-disabled[UFD])
> ```
> -- Output truncated --

6.3.7   show spanning-tree rstp brief

This command validates that Spanning Tree Protocol is enabled on the leaf switches. All interfaces are forwarding (`Sts` column shows `FWD`). One of the leaf switches (S4148F-Leaf1 in this example) is the root bridge and sever-facing interfaces (Po 1 in this example) are edge ports.

```
S4148F-Leaf1#show spanning-tree active
Spanning tree enabled protocol rstp with force-version rstp
Executing IEEE compatible Spanning Tree Protocol
Root ID    Priority 0, Address 1418.7725.49b9
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID    Priority 0, Address 1418.7725.49b9
We are the root
Configured hello time 2, max age 20, forward delay 15
Flush Interval 200 centi-sec, Flush Invocations 89
Flush Indication threshold 0 (MAC flush optimization is disabled)
Interface                                               Designated
Name            PortID  Prio  Cost   Sts   Cost      Bridge ID      PortID
-------------------------------------------------------------------------
ethernet1/1/53  128.468  128   2000   FWD    0     0   1418.7725.49b9  128.468
Interface
Name            Role    PortID   Prio   Cost    Sts    Cost   Link-type   Edge
-------------------------------------------------------------------------
ethernet1/1/53  Desg    128.468  128    2000    FWD    0      AUTO        No
```

7 Example 2: Layer 3 with Dell EMC leaf and spine switches using eBGP

This section provides eBGP configuration examples to build the layer 3 leaf-spine topology shown in Figure 13. Dell EMC Networking S4148F-ON and S4248FB-ON switches are used at the leaf layer and Dell EMC Networking Z9100-ON switches are used at the spine layer. While the S4148F-ON configuration is shown in this section, the configuration details for the S4248FB-ON switches are attached.

Note: The BGP ASNs and IP addresses defined in Section 5 are used here.
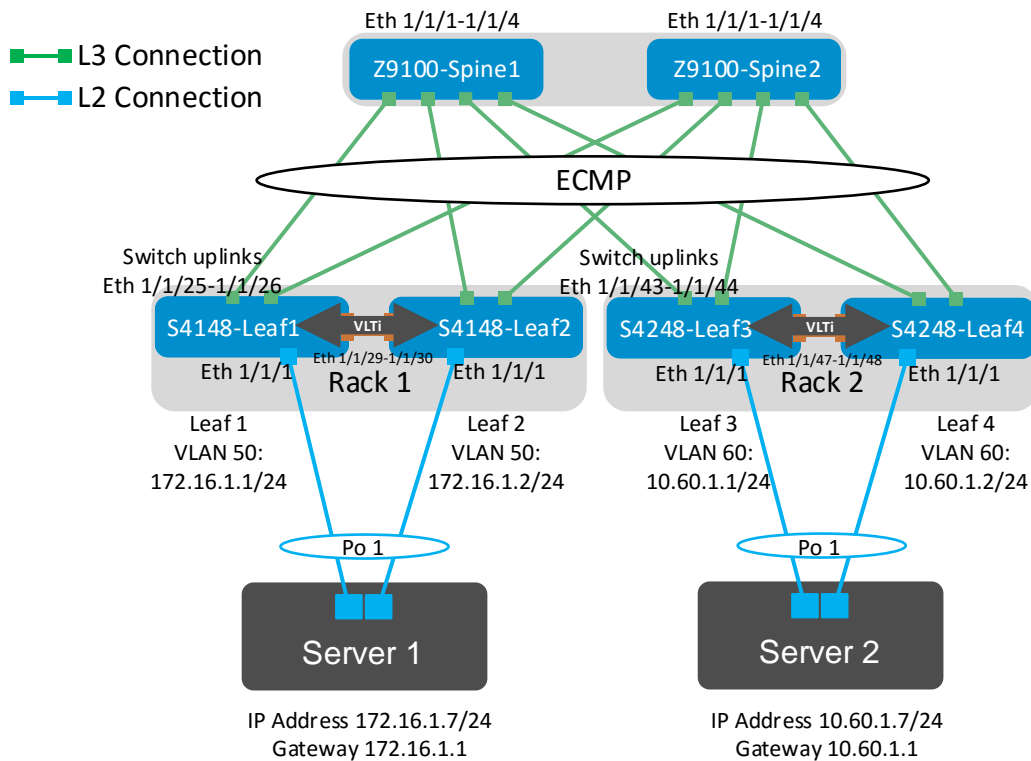

Figure 13   Example 2: Layer 3 leaf-spine topology

Note: All switch configuration files for the topology in Figure 13 are contained in the attachment named **Example2_config_files.pdf**. The files may be edited as needed in a plain text editor and commands pasted directly into switch consoles. Dell EMC Networking switches start at their factory default settings per Appendix A.

7.1   S4148F-ON leaf switch configuration

The following configuration details are for S4148F-Leaf1 and S4148F-Leaf2 in Figure 13. The configuration commands for S4248FB-Leaf3 and S4248FB-Leaf4 are similar and are provided in the attachments.

> **Note:** On S4148F-ON switches running OS10, Telnet is disabled and SSH is enabled by default. The S4148F-ON has default credentials of admin/admin. Therefore, the switch is accessible by default via SSH with those credentials. Changing the admin password to something more complex is recommended.

Set the host name, configure the OOB management interface and default gateway. Enable RSTP as a precaution. S4148F-Leaf1 is configured as the primary RSTP root bridge using the `spanning-tree priority 0` command. S4148F-Leaf2 is configured as the secondary RSTP root bridge using the `spanning-tree priority 4096` command.

S4148F-Leaf1	S4148F-Leaf2

`enable`
`configure t`
 	
`hostname S4148F-Leaf1`
 	
`interface mgmt 1/1/1`
`no ip address`
`ip address 100.67.170.30/24`
`no shutdown`
`management route 0.0.0.0/0`
`100.67.170.254`
 	
`spanning-tree mode rstp`
`spanning-tree rstp priority 0`
	`enable`
	`configure t`
	
	`hostname S4148F-Leaf2`
	
	`interface mgmt 1/1/1`
	`no ip address`
	`ip address 100.67.170.29/24`
	`no shutdown`
	`management route 0.0.0.0/0`
	`100.67.170.254`
	
	`spanning-tree mode rstp`
	`spanning-tree rstp priority 4096`

Configure the VLT interconnect between S4148F-Leaf1 and S4148F-Leaf2. In this configuration, remove each interface from layer 2 mode with the `no switchport` command for interfaces eth 1/1/29-1/1/30. Then enter the VLT domain and add the backup destination. Next add the interfaces participating in VLTi. Finally enable VLT peer routing.

S4148F-Leaf1	S4148F-Leaf2

`interface range ethernet 1/1/29-1/1/30`
`no switchport`
 	
`vlt-domain 127`
`backup destination 100.67.170.29`
`discovery-interface ethernet 1/1/29`
`discovery-interface ethernet 1/1/30`
`peer-routing`
	`interface range ethernet 1/1/29-1/1/30`
	`no switchport`
	
	`vlt-domain 127`
	`backup destination 100.67.170.30`
	`discovery-interface ethernet 1/1/29`
	`discovery-interface ethernet 1/1/30`
	`peer-routing`

DELLEMC

Create a server-facing VLAN interface. Use the same VLAN ID on both leaf switches. Assign an IP address to the VLAN interface. The address must be unique but on the same network on both leaf switches. Configure VRRP to use VRRP version 3. Create a VRRP group and specify the group's virtual IP address.

Note: In this example, Server 1's NIC is configured as an LACP NIC team. It is assigned the IP address 172.16.1.7/24. The VRRP VIP address, 172.16.1.254, is specified as Server 1's default gateway.

S4148F-Leaf1	S4148F-Leaf2
| ```
interface Vlan 50
ip address 172.16.1.1/24
no shutdown

exit
vrrp version 3
interface vlan 50
vrrp-group 50
virtual-address 172.16.1.254
``` | ```
interface Vlan 50
ip address 172.16.1.2/24
no shutdown

exit
vrrp version 3
interface vlan 50
vrrp-group 50
virtual-address 172.16.1.254
``` |

Configure each downstream server-facing interface with an LACP port channel. Configure each port channel for VLT. Port channel 10 connects downstream to Server 1 and is configured as an RSTP edge port.

S4148F-Leaf1	S4148F-Leaf2
| ```
interface port-channel 10
description "Server 1"
switchport access vlan 50
vlt-port-channel 10
spanning-tree port type edge

interface ethernet 1/1/1
description "Server 1"
channel-group 10 mode active
no shutdown
``` | ```
interface port-channel 10
description "Server 1"
switchport access vlan 50
vlt-port-channel 10
spanning-tree port type edge

interface ethernet 1/1/1
description "Server 1"
channel-group 10 mode active
no shutdown
``` |

The two upstream layer 3 interfaces connected to the spine switches are configured. Assign IP addresses per Table 1. Configure a loopback interface to be used as the router ID. This is used with BGP or OSPF.

Note: If multiple loopback interfaces exist on a system, the interface with the highest numbered IP address is used as the router ID. This configuration only uses one loopback interface.

S4148F-Leaf1	S4148F-Leaf2
interface ethernet 1/1/25 description "Z9100-Spine1 eth 1/1/1" no switchport ip address 192.168.1.1/31 no shutdown interface ethernet 1/1/26 description "Z9100-Spine2 eth 1/1/1 no switchport ip address 192.168.2.1/31 no shutdown interface loopback 0 description "Router ID" ip address 10.0.2.1/32 no shutdown	interface ethernet 1/1/25 description "Z9100-Spine1 eth 1/1/2" no switchport ip address 192.168.1.3/31 no shutdown interface ethernet 1/1/26 description "Z9100-Spine2 eth 1/1/2" no switchport ip address 192.168.2.3/31 no shutdown interface loopback 0 description "Router ID" ip address 10.0.2.2/32 no shutdown

Configure a route map and IP prefix-list to redistribute all loopback addresses and leaf networks via BGP.

The command `seq 10 permit 10.0.0.0/8 ge 24` includes all addresses in the 10.0.0.0/8 address range with a mask greater than or equal to 24. This includes all loopback addresses used as router IDs as well as the 10.60.1.0/24 network used on leaf switches 3 and 4 as shown in Figure 12.

The command `seq 20 permit 172.16.0.0/16 ge 24` includes the 172.16.1.0/24 network used on Leaf switches 1 and 2 as shown in Figure 13.

S4148F-1	S4148F-Leaf2
route-map spine-leaf permit 10 match ip address prefix-list spine-leaf ip prefix-list spine-leaf seq 10 permit 10.0.0.0/8 ge 24 ip prefix-list spine-leaf seq 20 permit 172.16.0.0/16 ge 24	route-map spine-leaf permit 10 match ip address prefix-list spine-leaf ip prefix-list spine-leaf seq 10 permit 10.0.0.0/8 ge 24 ip prefix-list spine-leaf seq 20 permit 172.16.0.0/16 ge 24

Configure UFD. This shuts the downstream interfaces if all uplinks fail. The hosts attached to the switch use the remaining LACP LAG member to continue sending traffic across the fabric. Finally, exit configuration mode and save the configuration.

S4148F-Leaf1	S4148F-Leaf2
uplink-state-group 1	uplink-state-group 1

DELLEMC

description "Disable downstream ports in event all uplinks fail" downstream port-channel 10 upstream ethernet 1/1/25 upstream ethernet 1/1/26 enable	description "Disable downstream ports in event all uplinks fail" downstream port-channel 10 upstream ethernet 1/1/25 upstream ethernet 1/1/26 enable

Use these commands to configure eBGP.

First, enable eBGP with the `router bgp` *ASN* command. The ASN is from Figure 9.

The bgp `bestpath as-path multipath-relax` enables ECMP. The `maximum-paths eBGP 2` command specifies the maximum number of parallel paths to a destination to add to the routing table. In this topology, there are two equal cost best paths from a spine to a host, one to each leaf that the host is connected.

BGP neighbors are configured and Neighbor fall-over is enabled. Graceful restart enables the data plane to continue forwarding traffic for a time if the BGP process fails or quits.

BGP hello and hold down timers are set to three and nine seconds, respectively. Neighbor fall-over will trigger route withdrawal when IP connectivity between BGP peers is lost. The more aggressive timers (default is 60 seconds hello and 180 seconds hold down) are for backup. The advertisement interval is set to one second. This is to prevent BGP speakers from advertising updates immediately upon receipt. Instead, they will advertise them in batched intervals of one second. This delay is to prevent overhead.

Finally, exit configuration mode and save the configuration.

S4148F-Leaf1	S4148F-Leaf2
router bgp 64701 bestpath as-path multipath-relax maximum-paths ebgp 2 address-family ipv4 unicast redistribute connected route-map spine-leaf exit graceful-restart role receiver-only template spine-leaf fall-over advertisement-interval 1 timers 3 9 exit neighbor 192.168.1.0 remote-as 64601 inherit template spine-leaf no shutdown exit neighbor 192.168.2.0 remote-as 64602	router bgp 64702 bestpath as-path multipath-relax maximum-paths ebgp 2 address-family ipv4 unicast redistribute connected route-map spine-leaf exit graceful-restart role receiver-only template spine-leaf fall-over advertisement-interval 1 timers 3 9 exit neighbor 192.168.1.2 remote-as 64601 inherit template spine-leaf no shutdown exit neighbor 192.168.2.2 remote-as 64602

| ```
inherit template spine-leaf
no shutdown

end
write memory
``` | ```
inherit template spine-leaf
no shutdown

end
write memory
``` |

7.2   Z9100-ON spine switch configuration

The following configuration details are for Z9100-ON-Spine1 and Z9100-ON-Spine2 in Figure 13.

Set the host name, configure the OOB management interface and default gateway.

Z9100-Spine1	Z9100-Spine2
| ```
enable
configure t

hostname Z9100-Spine1

interface mgmt 1/1/1
no ip address
ip address 100.67.173.36/24
no shutdown
management route 0.0.0.0/0
100.67.173.254
``` | ```
enable
configure t

hostname Z9100-Spine2

interface mgmt 1/1/1
no ip address
ip address 100.67.173.35/24
no shutdown
management route 0.0.0.0/0
100.67.173.254
``` |

Configure the four point-to-point interfaces connected to leaf switches. Assign IP addresses per Table 1.
Configure a loopback interface to be used as the router ID. This is used with BGP or OSPF.

Z9100-Spine1	Z9100-Spine2
| ```
interface ethernet 1/1/1
description "S4148F-Leaf1 eth 1/1/29"
no switchport
ip address 192.168.1.0/31
no shutdown

interface ethernet 1/1/2
description "S4148F-Leaf2 eth 1/1/29"
no switchport
ip address 192.168.1.2/31
no shutdown

interface ethernet 1/1/3
description "S4248FB-Leaf3 eth 1/1/47"
no switchport
ip address 192.168.1.4/31
no shutdown
``` | ```
interface ethernet 1/1/1
description "S4148F-Leaf1 eth 1/1/30"
no switchport
ip address 192.168.2.0/31
no shutdown

interface ethernet 1/1/2
description "S4148F-Leaf2 eth 1/1/30"
no switchport
ip address 192.168.2.2/31
no shutdown

interface ethernet 1/1/3
description "S4248FB-Leaf3 eth 1/1/48"
no switchport
ip address 192.168.2.4/31
no shutdown
``` |

```
interface ethernet 1/1/4                  interface ethernet 1/1/4
description "S4248FB-Leaf4 eth 1/1/47"     description "S4248FB-Leaf4 eth 1/1/48"
no switchport                             no switchport
ip address 192.168.1.6/31                 ip address 192.168.2.6/31
no shutdown                               no shutdown

interface loopback 0                      interface loopback 0
description "Router ID"                    description "Router ID"
ip address 10.0.1.1/32                     ip address 10.0.1.2/32
no shutdown                               no shutdown
```

Configure a route map and IP prefix-list to redistribute all loopback addresses and leaf networks via BGP or OSPF.

The command `seq 10 permit 10.0.0.0/8 ge 24` includes all addresses in the 10.0.0.0/8 address range with a mask greater than or equal to 24. This includes all loopback addresses used as router IDs as well as the 10.60.1.0/24 network used on leaf switches 3 and 4 as shown in Figure 12.

The command `seq 20 permit 172.16.0.0/16 ge 24` includes the 172.16.1.0/24 network used on Leaf switches 1 and 2 as shown in Figure 12.

Z9100-Spine1	Z9100-Spine2
route-map spine-leaf permit 10 match ip address prefix-list spine- leaf ip prefix-list spine-leaf seq 10 permit 10.0.0.0/8 ge 24 ip prefix-list spine-leaf seq 20 permit 172.16.0.0/16 ge 24	route-map spine-leaf permit 10 match ip address prefix-list spine- leaf ip prefix-list spine-leaf seq 10 permit 10.0.0.0/8 ge 24 ip prefix-list spine-leaf seq 20 permit 172.16.0.0/16 ge 24

Use these commands to configure BGP.

First, enable eBGP with the `router bgp` *ASN* command. The ASN is from Figure 9.

The bgp `bestpath as-path multipath-relax` command enables ECMP. The `maximum-paths eBGP 2` command specifies the maximum number of parallel paths to a destination to add to the routing table. In this topology, there are two equal cost best paths from a spine to a host, one to each leaf that the host is connected.

BGP neighbors are configured and Neighbor fall-over is enabled. Graceful restart enables the data plane to continue forwarding traffic for a time if the BGP process fails or quits.

BGP hello and hold down timers are set to three and nine seconds, respectively. Neighbor fall-over will trigger route withdrawal when IP connectivity between BGP peers is lost. The more aggressive timers (default is 60

seconds hello and 180 seconds hold down) are for backup. The advertisement interval is set to one second. This is to prevent BGP speakers from advertising updates immediately upon receipt. Instead, they will advertise them in batched intervals of one second. This delay is to prevent overhead.

Finally, exit configuration mode and save the configuration.

Z9100-Spine1	Z9100-Spine2
| ```
router bgp 64601
bestpath as-path multipath-relax
maximum-paths ebgp 2
address-family ipv4 unicast
redistribute connected route-map
spine-leaf
exit
graceful-restart role receiver-only
template spine-leaf
fall-over
advertisement-interval 1
timers 3 9
exit
neighbor 192.168.1.1
remote-as 64701
inherit template spine-leaf
no shutdown
exit
neighbor 192.168.1.3
remote-as 64702
inherit template spine-leaf
no shutdown
exit
neighbor 192.168.1.5
remote-as 64703
inherit template spine-leaf
no shutdown
exit
neighbor 192.168.1.7
remote-as 64704
inherit template spine-leaf
no shutdown

end
write memory
``` | ```
router bgp 64602
bestpath as-path multipath-relax
maximum-paths ebgp 2
address-family ipv4 unicast
redistribute connected route-map
spine-leaf
exit
graceful-restart role receiver-only
template spine-leaf
fall-over
advertisement-interval 1S
timers 3 9
exit
neighbor 192.168.2.1
remote-as 64701
inherit template spine-leaf
no shutdown
exit
neighbor 192.168.2.3
remote-as 64702
inherit template spine-leaf
no shutdown
exit
neighbor 192.168.2.5
remote-as 64703
inherit template spine-leaf
no shutdown
exit
neighbor 192.168.2.7
remote-as 64704
inherit template spine-leaf
no shutdown

end
write memory
``` |

7.3   Example 2 validation

In addition to sending traffic between hosts, the configuration shown in Figure 13 can be validated with the commands shown in this section. For more information on commands and output, see the Command Line Reference Guide for the applicable switch (links to documentation are provided in Appendix C).

DELLEMC

Command and output examples are provided for one spine and one leaf. Command output on other switches is similar.

7.3.1 show ip bgp summary

This command shows the status of all BGP connections. Each spine has four neighbors (the four leaf switches) and each leaf switch has two neighbors (the two spines switches). As can be seen, the leaf switches each use different ASNs, as the environment uses eBGP.

```
Z9100-Spine1#show ip bgp summary
BGP router identifier 10.0.1.1 local AS number 64601
Neighbor         AS          MsgRcvd     MsgSent     Up/Down        State/Pfx
192.168.1.1      64701       43          45          00:01:38       8
192.168.1.3      64702       43          46          00:01:38       8
192.168.1.5      64703       42          46          00:01:38       8
192.168.1.7      64704       43          45          00:01:37       8


S4148F-Leaf1#show ip bgp summary
BGP router identifier 10.0.2.1 local AS number 64701
Neighbor         AS          MsgRcvd     MsgSent     Up/Down        State/Pfx
192.168.1.0      64601       110         106         00:04:22       12
192.168.2.0      64602       108         110         00:04:19       12
```

7.3.2 show ip route bgp

This command is used to verify the BGP entries in the Routing Information Base (RIB). Entries with multiple paths shown are used with ECMP. The two server networks in this example, 10.60.1.0 and 172.16.1.0, each have two best paths from Z9100-Spine1, one through each leaf switch.

Note: The first set of routes with a subnet mask of /32 are the IPs configured for router IDs.

```
Z9100-Spine1#show ip route bgp
Codes: C - connected
       S - static
       B - BGP, IN - internal BGP, EX - external BGP
       O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
       N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
       E2 - OSPF external type 2, * - candidate default,
       + - summary route, > - non-active route
Gateway of last resort is not set
  Destination                 Gateway              Dist/Metric     Last Change
---------------------------------------------------------------------------
  B  EX 10.0.1.2/32           via 192.168.1.1      20/0            00:09:01
                              via 192.168.1.3
  B  EX 10.0.2.1/32           via 192.168.1.1      20/0            00:09:04
  B  EX 10.0.2.2/32           via 192.168.1.3      20/0            00:09:04
```

DELLEMC

```
B  EX 10.0.2.3/32          via 192.168.1.5        20/0             00:09:04
B  EX 10.0.2.4/32          via 192.168.1.7        20/0             00:09:03
B  EX 10.60.1.0/24         via 192.168.1.5        20/0             00:09:03
                           via 192.168.1.7
B  EX 172.16.1.0/24        via 192.168.1.1        20/0             00:09:04
                           via 192.168.1.3
B  EX 192.168.2.0/31       via 192.168.1.1        20/0             00:09:04
B  EX 192.168.2.2/31       via 192.168.1.3        20/0             00:09:04
B  EX 192.168.2.4/31       via 192.168.1.5        20/0             00:09:04
B  EX 192.168.2.6/31       via 192.168.1.7        20/0             00:09:03
```

S4148F-Leaf1 has two paths to all other leaf switches and two paths to Server 2's network, 10.60.1.0. There is one path through each spine switch.

```
S4148F-Leaf1#show ip route bgp
Codes: C - connected
       S - static
       B - BGP, IN - internal BGP, EX - external BGP
       O - OSPF, IA - OSPF inter area, N1 - OSPF NSSA external type 1,
       N2 - OSPF NSSA external type 2, E1 - OSPF external type 1,
       E2 - OSPF external type 2, * - candidate default,
       + - summary route, > - non-active route
Gateway of last resort is not set
  Destination                Gateway             Dist/Metric      Last Change
---------------------------------------------------------------------------
  B  EX 10.0.1.1/32          via 192.168.1.0     20/0             00:09:04
  B  EX 10.0.1.2/32          via 192.168.2.0     20/0             00:09:01
  B  EX 10.0.2.2/32          via 192.168.1.0     20/0             00:09:01
                             via 192.168.2.0
  B  EX 10.0.2.3/32          via 192.168.1.0     20/0             00:09:01
                             via 192.168.2.0
  B  EX 10.0.2.4/32          via 192.168.1.0     20/0             00:09:00
                             via 192.168.2.0
  B  EX 10.60.1.0/24         via 192.168.1.0     20/0             00:09:01
                             via 192.168.2.0
  B  EX 192.168.1.2/31       via 192.168.1.0     20/0             00:09:04
  B  EX 192.168.1.4/31       via 192.168.1.0     20/0             00:09:04
  B  EX 192.168.1.6/31       via 192.168.1.0     20/0             00:09:04
  B  EX 192.168.2.2/31       via 192.168.2.0     20/0             00:09:01
  B  EX 192.168.2.4/31       via 192.168.2.0     20/0             00:09:01
  B  EX 192.168.2.6/31       via 192.168.2.0     20/0             00:09:01
```

7.3.3   Dell EMC Networking leaf switch validation commands previously covered

The following commands previously covered may be run to validate the Dell EMC Networking leaf switches for this configuration. The output is the same or similar to that shown in the referenced sections.

`show vlt` – see Section 6.3.3

`show vlt vlt-port-detail`– see Section 6.3.4

`show vlt mismatch` – see Section 6.3.5

`show uplink-state-group` – see Section 6.3.6

`show spanning-tree rstp brief` – see Section 6.3.7

DELLEMC

A Dell EMC Networking ONIE switch factory default settings

> **Note:** This set of instructions brings ONIE switches to their factory default settings, which includes removing the operating system. If only removing the configuration is desired, this can be achieved with `delete startup-config` proceeded with `reload`.

All Dell EMC Networking ONIE switches in this guide can be reset to factory defaults with the following instructions:

1. Reboot the switch
2. Use the up and down arrows to select ONIE: Rescue, then press Enter. The highlighted entry (*) runs automatically in the operating system.

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

3. Press **Enter** again to enable the console
4. Use the `onie-uninstall` command to return to the default ONIE settings.

```
ONIE:/ # onie-uninstall
Erasing unused NOR flash region Erasing 128 Kibyte @ 20000 - 100% complete.
Erasing internal mass storage device: /dev/mmcblk0 (7832MB) Percent complete:
100%
```

DELLEMC

B   Validated hardware and operating systems

The following table includes the hardware and operating systems used to validate the examples in this guide.

Table 2   Switches and operating systems used in this guide

Switch	OS / Version
Dell EMC Networking S4148F-ON	10.4.0E(R3)
Dell EMC Networking S4248FB-ON	10.4.0E(R3)
Dell EMC Networking Z9100-ON	10.4.0E(R3)

C   Technical support and resources

Dell EMC TechCenter is an online technical community where IT professionals have access to numerous resources for Dell EMC software, hardware and services.

Dell EMC TechCenter Networking Guides

Manuals and documentation for Dell EMC Networking S4148F-ON

Manuals and documentation for Dell EMC Networking S4248FB-ON

Manuals and documentation for Dell EMC Networking Z9100-ON

D Support and Feedback

Contacting Technical Support

Support Contact Information   Web: Dell.com/support

                                              Telephone: USA: 1-800-945-3355

Feedback for this document

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