VMware Validated Design 4.1 Networking Supplement

Deploying Dell EMC Leaf and Spine Networks in a VVD 4.1 Environment

Dell EMC Networking Infrastructure Solutions
December 2017

Revisions

Date	Rev.	Description	Authors
Dec 2017	1.0	Initial release	Andrew Waranowski, Curtis Bunch, Jordan Wilson

Table of contents

1 Introduction

The purpose of this document is to provide best practices for setting up the Dell EMC leaf and spine networks in the context of an NSX-enabled VMware Validated Design environment (VVD). The leaf and spine networks detailed in this document facilitate and enable the function of the production traffic running in the NSX overlay. This document is a supplement to the *Leaf-Spine Deployment and Best Practices Guide* that details how the Dell EMC leaf and spine networks can be adapted to support the NSX overlay in a VVD environment.

In this guide, two examples are detailed: one, in which the layer 2/3 boundary is extended to the leaf layer, and one in which the layer 2/3 boundary is extended to the spine layer. Because this guide is a supplement, there is quite a bit of information about protocols, equipment, and design decisions pertaining to the leaf and spine documentation which is left out, as that information is already explained in detail in the *Leaf-Spine Deployment and Best Practices Guide*. This guide explains the configuration that is addition to what is shown in the *Leaf-Spine Deployment and Best Practices Guide*, which is specific to supporting the overlay production network.

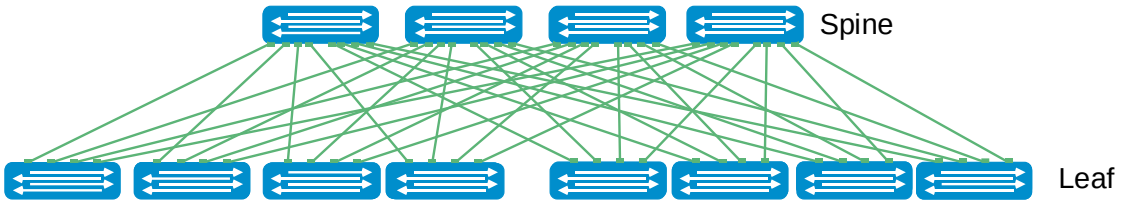

Figure 1 Leaf-spine architecture

1.1 Typographical Conventions

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

`Monospace Text`	CLI examples
`Underlined Monospace Text`	CLI examples that wrap the page - 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.

2 VVD 4.1 environment

The environment in this document consists of the minimum number of servers required for a VMware Validated Design 4.1 implementation. VVD 4.1 requires at least four servers for the Management pod and four servers for the Shared Edge and Compute. In addition to Management and Shared Edge and Compute pods, VMware defines two other pod types: Compute and Storage. However, a VVD 4.1 environment only requires Management and Shared Edge and Compute pods.

Note: A VVD 4.1 environment requires many different services in addition to a physical network underlay. These services include Active Directory, DHCP, FTP, NTP, FTP, and DNS. Adjustments will need to be made to the network in order to make these services accessible to the ESXi hosts.

2.1 Server to switch interconnect

Each server has two physical 10GbE connections: one to each leaf switch. The first NIC is connected to the first leaf switch and the second NIC is connected to the second leaf switch. When ESXi is installed, vmnic0 is paired with the first leaf switch, and vmnic1 is paired with the second leaf switch.

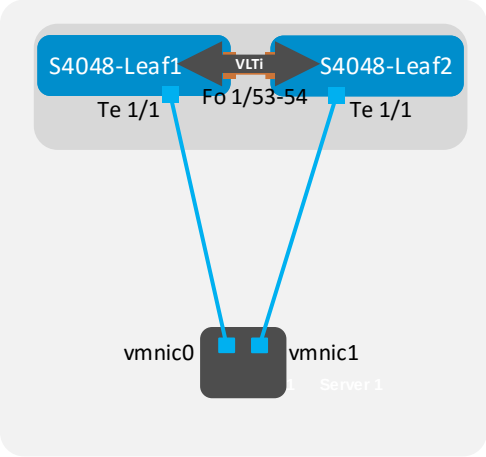

Figure 2 Host connections to leaf switches

2.2 Physical environment

The environment in this document consists of six switches and eight servers. For switches, there are two Dell Networking Z9100 switches at the spine layer and four Dell Networking S4048-ON switches at the leaf layer. The physical topology for layer 3 and layer 2 is the same, except for the links between the spines in the layer 2 topology. The links between the spine switches are required for enabling Virtual Link Trunking (VLT) between them.

For servers, there are four Dell PowerEdge R640 servers and four Dell PowerEdge R730xd servers. The R640 servers are used for the Management pod, and the R730xd servers are used for the Shared Edge and Compute pod.

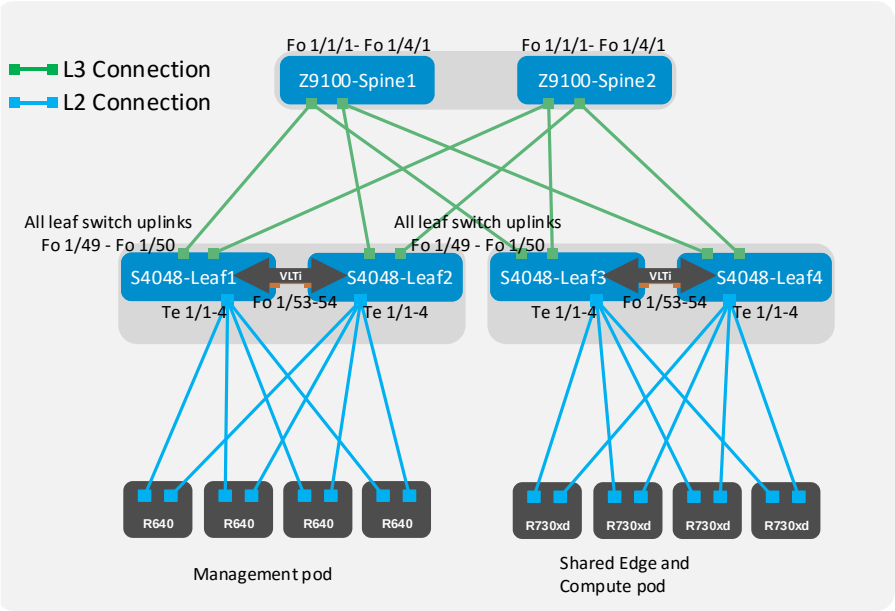

Figure 3 Layer 3 physical topology

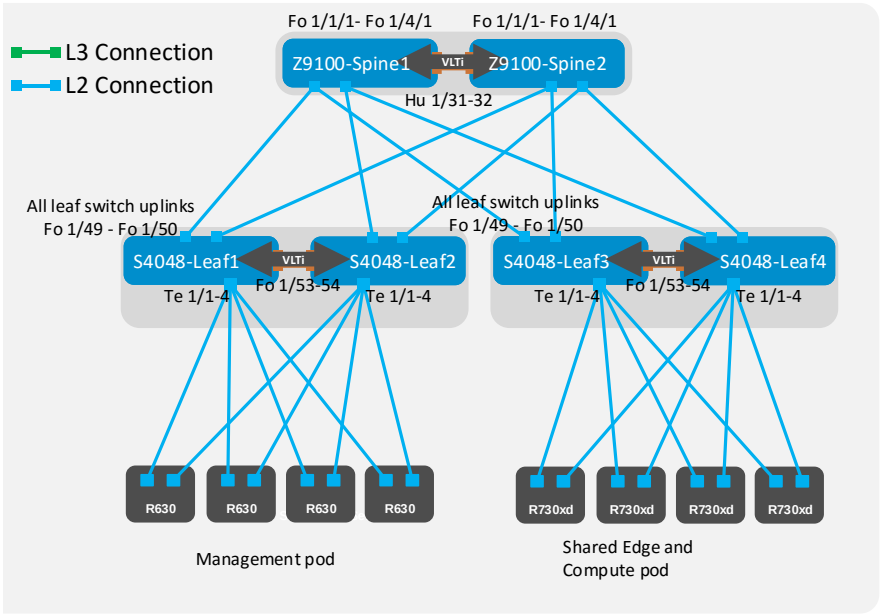

Figure 4 Layer 2 physical topology

2.3 Physical combined with virtual environments

A complete VVD environment consists of two regions: Region A and Region B. Regions support placing workloads closer to customers, and enables effective and efficient disaster recovery. With separate regions, each region can function as a backup if a disaster occurs. Figure 6 shows how complex and extensive the virtual infrastructure is.

For the purpose of this document, only one region is demonstrated. Much of the virtual infrastructure is left out and ignored as it is not relevant. As shown in Figure 5, the only virtual infrastructure that is relevant to this document is the Edge Services Gateway (ESG). ESGs provide the interconnection between the physical and virtual networks. The virtual networking devices underneath the ESGs, such as logical switches and distributed logical routers, function independently of the physical network. This document demonstrates how Dell Networking switches provide a robust, reliable, and scalable underlay for the NSX overlay, where the production traffic is. As a result, focus is placed on the physical network and how the physical network connects to the virtual network with little focus placed on the virtual networks themselves.

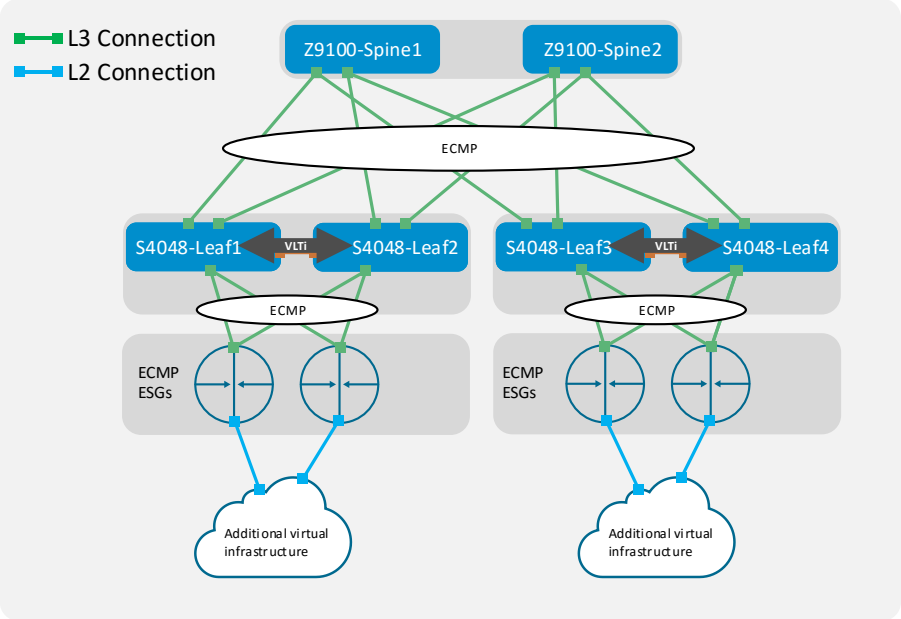

Figure 5 Combined physical and virtual elements

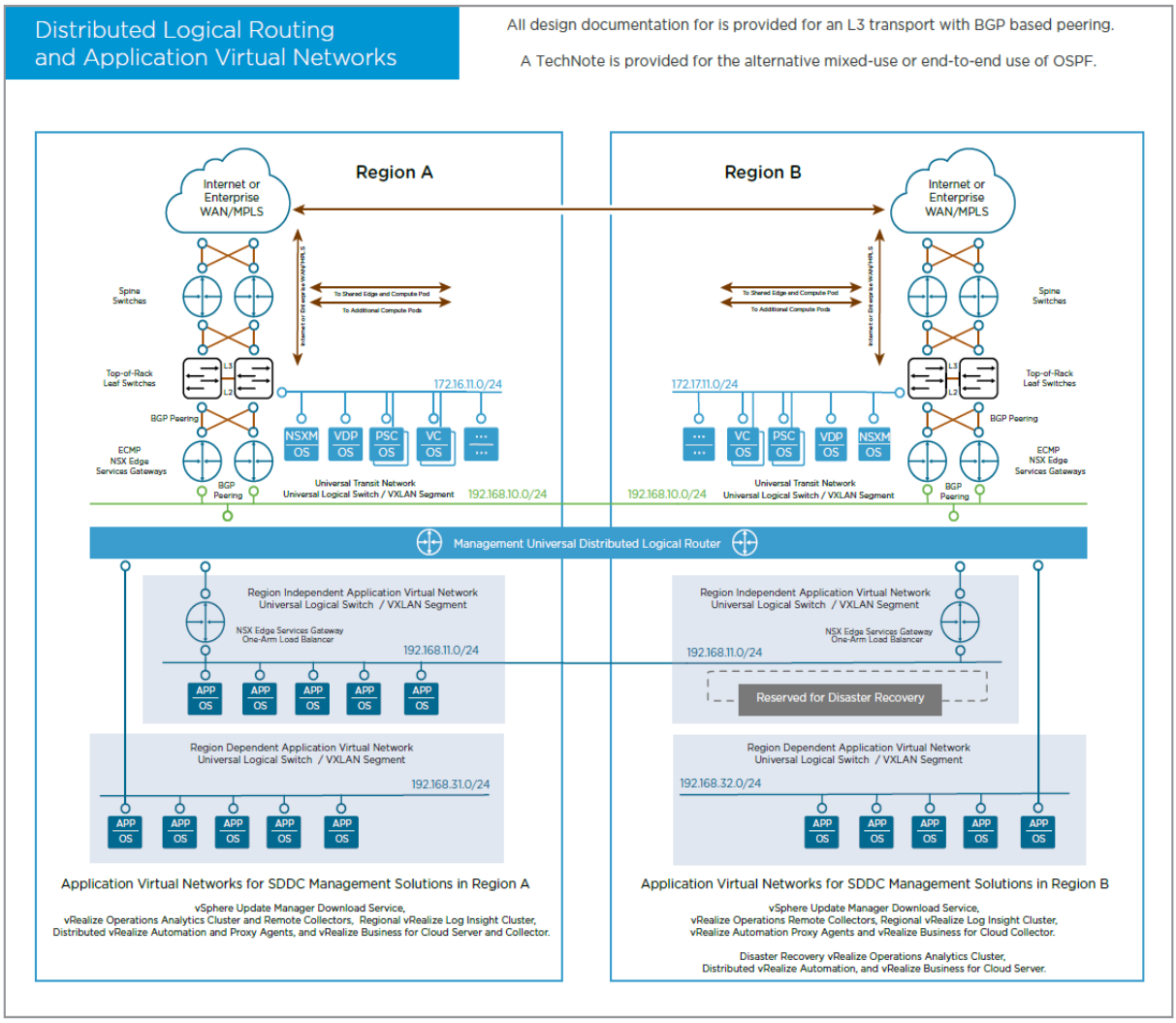

Figure 6 Detailed example of complete VVD 4.1 environment with physical and virtual elements

3 Example 1: Layer 2/3 boundary extended to the leaf switches

In this example the layer 2/3 boundary of the leaf and spine network is extended to the leaf switches. There are advantages to extending the layer 2/3 boundary to the leaf layer with the single greatest reason being scalability. As the leaf layer expands to accommodate growing data center demands, it is possible to add extra spines to maintain the desired leaf/spine oversubscription ratio. In networks where the layer 2/3 boundary is at the spines, the spine layer is limited to a single VLT pair. Another reason to extend layer 3 to the leaf switches is that it restricts broadcast and multicast traffic to smaller broadcast domains and constrains potential broadcast storms to smaller areas of the network.

In this scenario, each pod has its own set of VLANs for different vSphere services configured on the leaf switches. Those VLANs are not propagated through the spine layer. The spine layer provides connectivity between these leaf pairs via routed point-to-point links using External Border Gateway Protocol (eBGP).

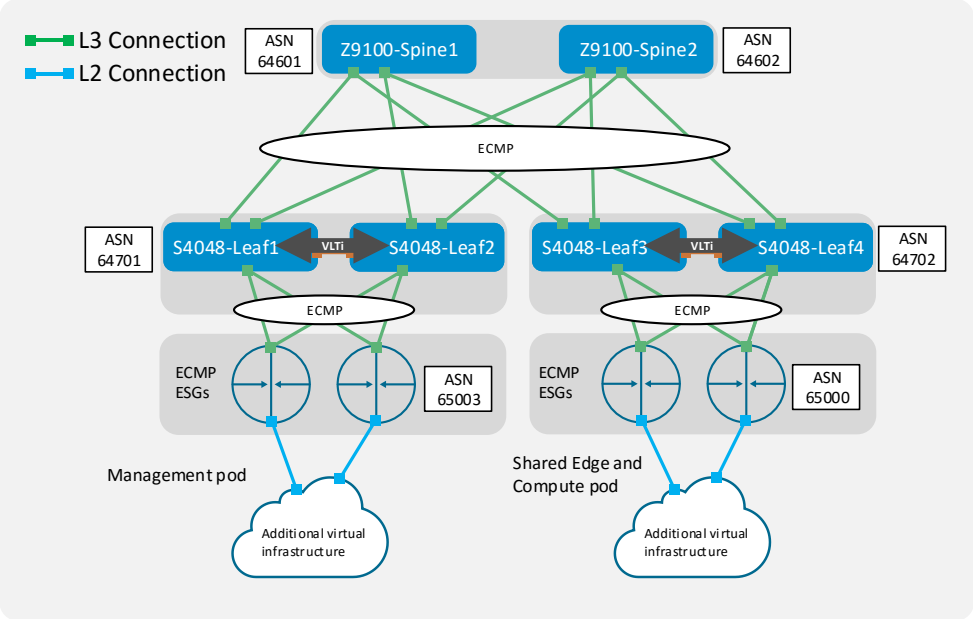

Figure 7 Example 1: Leaf-spine topology with the layer 2/3 boundary extended to the leaf switches

Note: All switch configuration files for the topology in Figure 7 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.

3.1 Design decisions

The following is a list of design decisions on top of the leaf and spine design which were made in order to support the VVD 4.1 environment. Leaf and spine design decisions that are not specific to supporting the VVD 4.1 environment can be found in the *Leaf-Spine Deployment and Best Practices Guide*.

Table 1 Design decisions

Design decision	
Jumbo frames	Jumbo frames are configured on all ports and VLANs on the switches because many vSphere services require jumbo frames. These services include vSAN, VXLAN, NFS, vSphere vMotion, and vSphere Replication.
QoS	All switch ports are configured to trust the DSCP markings inside the IP packets. The idea is to configure hosts and end stations to mark management traffic with a comparatively high DSCP value - 46 is common. When the switches intercept those packets, they honor those DSCP markings and assign the packets to a higher priority queue, and provide greater assurance that management traffic continues to flow if there is network contention.
NTP	In order to comply with VVD 4.1, a pair of infrastructure switches are synchronized with an NTP server that is external to the environment. Then all hosts and other network devices are synchronized to those switches. In Example 1, the leaf switches in the Management pod perform this function, while in Example 2, it is the spine switches.
IGMP snooping	Because vSAN and VXLAN require IGMP queriers to accommodate for multicast traffic, IGMP snooping is enabled globally on all four leaf switches. Also, unregistered multicast flooding is disabled to prevent unintentional multicast traffic from flooding.
IGMP querier for vSAN VLAN	Starting with version 6.6, vSAN no longer uses multicast. Because of this, vSAN 6.6 does not require IGMP querier. However, IGMP querier configuration on the vSAN VLAN is included for the purpose of reverse compatibility. Only IGMP querier, and not multicast routing, is required because vSAN multicast traffic stays within each rack.
IGMP querier for VXLAN VLAN	The VXLAN VLAN requires an IGMP querier, which accommodates NSX Hybrid mode and that VVD 4.1 specifies. In Hybrid mode, VTEPs use multicast for transporting broadcast, unknown unicast, and multicast (BUM) overlay traffic from one host to another within the rack, or local network. For BUM overlay transport across racks, VTEPs use unicast. The consequence of this is that, as with vSAN traffic, VXLAN multicast traffic stays in each rack, so only IGMP querier is required. If NSX Multicast mode is used, full multicast routing is required on all infrastructure switches since multicast is used for

	overlay transport across racks in addition to within the racks. If NSX Unicast mode is used, multicast configuration is not required.
Configured VLANs	Several VLANs are configured on the switch in order to accommodate for the VVD 4.1 architecture. These VLANs map to port groups within ESXi. They are: Ext management - provides access to individual ESXi hosts from outside the VVD 4.1 environment without using a jump box ESXi host management - enables communication between hosts, and between hosts and vCenter Server vMotion - enables and segregates vMotion traffic vSAN - provides transport for and segregates vSAN traffic VXLAN - provides support for NSX and VXLAN overlay traffic NFS - enables and segregates NFS, which VVD 4.1 requires in order to function as backup storage if the primary storage fails Replication - provides support and segregation for vSphere Replication Uplink1 - enables north/south traffic in the data center Uplink2 - enables north/south traffic in the data center In the example 1 environment, the VLANs are configured on leaf switches. In these networks, BGP peering occurs between the ESGs and the upstream leaf switches.
Virtual Router Redundancy Protocol (VRRP)	VRRP is used for gateway redundancy for all VLANs, with the exception of uplink VLANs and vSAN VLANs. This is a VVD 4.1 requirement. vSAN VLANs do not use VRRP because the vSANs in this environment do not use routing. Although VRRP is an active/standy first hop redundancy protocol (FHRP), it becomes active/active when it is used between VLT peers. Both VLT peers have the VRRP virtual MAC address in their Forwarding Information Base (FIB) table as a localda entry, and therefore, even the backup VRRP router will forward intercepted frames whose destination MAC address matches the VRRP virtual MAC address. Example 1 shows VRRP configured on the leaf switches.
No host-facing port channels	Another design decision is to not use host-facing port channels. To comply with the VVD 4.1, host-facing port channels are not used. Rather, route is used based on the physical NIC load on the distributed port groups, which does not necessitate any port channel configuration on the attached physical switch.
BGP	BGP is the chosen protocol in the SDDC for connecting all network devices, both physical and virtual. BGP allows for flexibility in network design in multi-site and mult-tenancy workloads, and is highly tunable.

Each cluster contains a pair of ESGs that peer with their respective ToR switches in a full-mesh ECMP configuration. Bidirectional Forwarding Detection (BFD) between ESGs and physical switches is not currently available, so it is disabled. VVD 4.1 specifies that BGP timers are set to four seconds for keepalive interval,

and 14 seconds for hold down for the peering relationship between the ESGs and their physical upstream peer switches.

eBGP is chosen because it simplifies configuration. With eBGP, there is no need to configure route reflectors, or to implement next hop self configuration. However, the ASN implementation in this guide diverges from the *Leaf-Spine Deployment and Best Practices Guide* in that leaf pairs share the same ASN. This saves on ASNs while maintaining a strictly eBGP network, since the leaf switches do not peer with each other. Also, it prevents ESGs from advertising routes that they learn from the leaf and spine network, back into the leaf and spine network. This is important since ESGs currently have an ASN-stripping behavior which causes those re-advertised routes to have higher preference than routes which have the true path originating in the leaf and spine network.

Default-originate and a route filter are applied to the leaf switches in order to advertise default routes and filter all other routes to the ESGs. Since there is only one way out of the NSX network, there is no need for the ESGs or DLRs to learn anything about the outside network except for how to get to it via default routes.

3.2 Layer 3 configuration planning

3.2.1 BGP ASN configuration

In VVD 4.1 environments, the leaf and spine design as shown in the *Leaf-Spine Deployment and Best Practices Guide* must be adapted in order to work properly with the ESGs. In terms of AS numbering, this means that each spine switch continues to get its own ASN, but leaf pairs share one ASN. The ESGs and other virtual networking devices use iBGP amongst themselves. Valid private, 2-byte ASNs range from 64512 through 65534. Figure 8 shows the ASN assignments used for leaf and spine switches in the BGP examples in this guide.

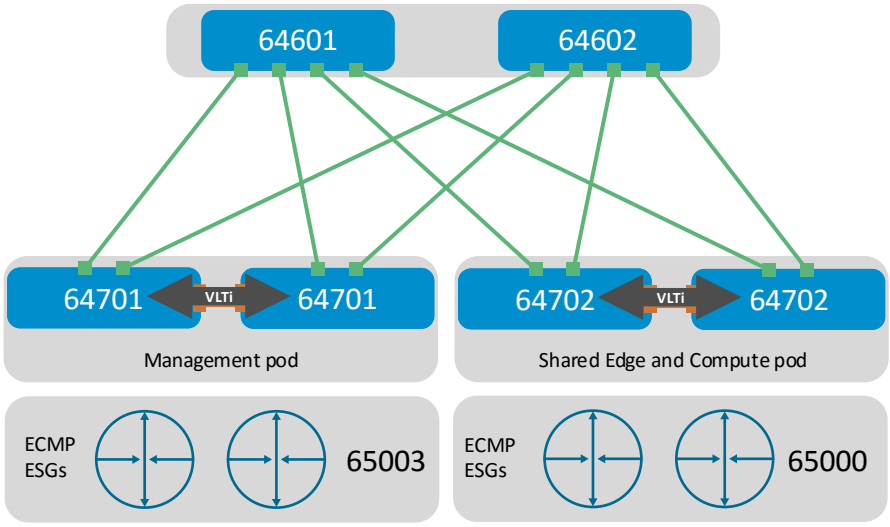

Figure 8 BGP ASN assignments

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, such as 64601, represents a spine and an ASN with a "7" in the hundreds place, such as 64701, represents a leaf switch.

3.2.2 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.

3.2.2.1 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 9 shows the loopback addresses used as router IDs for BGP i this guide.

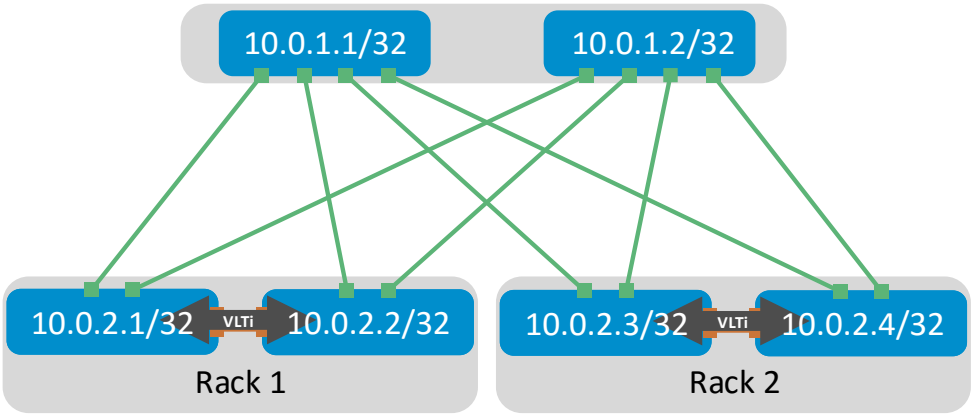

Figure 9 Loopback addressing

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.

3.2.2.2 Point-to-point addresses

Table 2 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. 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 10.

Table 2 Interface and IP configuration

Link Label	Source switch	Source interface	Source IP	Network	Destination switch	Destination interface	Destination IP
A	Leaf 1	fo1/49	.1	192.168.1.0/31	Spine 1	fo1/1/1	.0
B	Leaf 1	fo1/50	.1	192.168.2.0/31	Spine 2	fo1/1/1	.0
C	Leaf 2	fo1/49	.3	192.168.1.2/31	Spine 1	fo1/2/1	.2
D	Leaf 2	fo1/50	.3	192.168.2.2/31	Spine 2	fo1/2/1	.2
E	Leaf 3	fo1/49	.5	192.168.1.4/31	Spine 1	fo1/3/1	.4
F	Leaf 3	fo1/50	.5	192.168.2.4/31	Spine 2	fo1/3/1	.4
G	Leaf 4	fo1/49	.7	192.168.1.6/31	Spine 1	fo1/4/1	.6
H	Leaf 4	fo1/50	.7	192.168.2.6/31	Spine 2	fo1/4/1	.6

The point-to-point IP addresses used in this guide are shown in Figure 10:

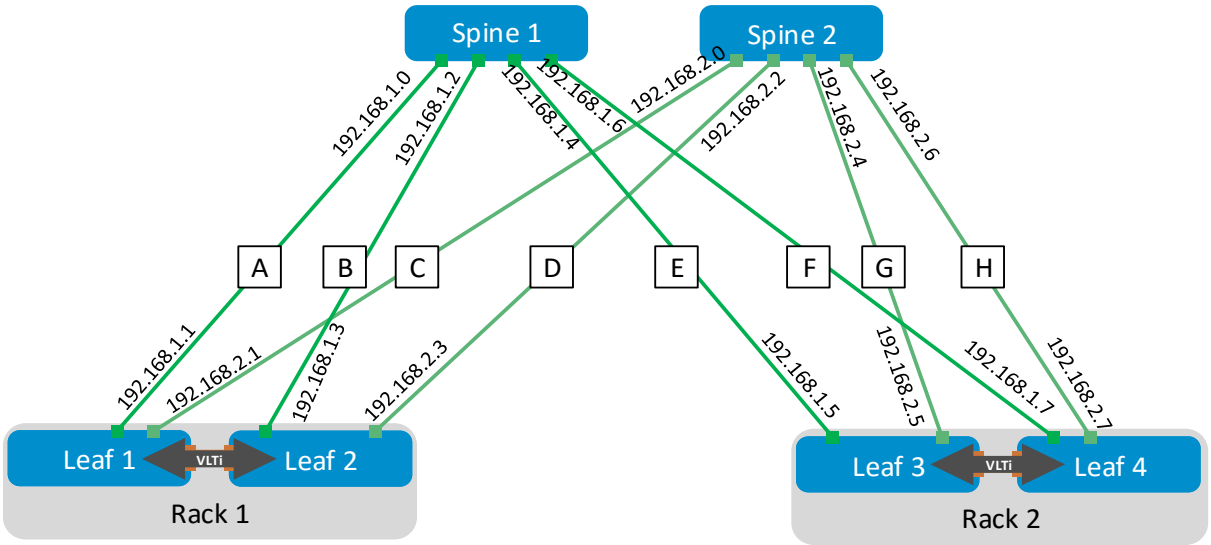

Figure 10 Point-to-point IP addresses

Note: The example point-to-point addresses use a 31-bit mask to save address space. This is optional and covered in RFC 3021. Below is an example when setting an IP address with a 31-bit mask on a Dell EMC S4048-ON. The warning message can be safely ignored on point-to-point interfaces.
S4048-Leaf-1(conf-if-fo-1/49)#**ip address 192.168.1.1/31**
% Warning: Use /31 mask on non point-to-point interface cautiously.

3.3 NSX configuration overview

The NSX VVD 4.1 environment demonstrated in this document was built using the VMware VVD 4.1 documentation, in particular, Region A Virtual Infrastructure Implementation, since the entire VVD 4.1 implementation is not required for the purposes of this document.

This section provides an overview of the virtual network infrastructure generated when one follows the VVD 4.1 documentation, along with how the virtual network infrastructure interconnects with the physical network infracture. Setting up NSX in a VVD 4.1-compliant way is covered in great detail by the VMware VVD 4.1 documentation.

Some of the generated virtual network infrastructure includes four ESGs which connect to the physical network: two for the Management pod and two for the Shared Edge and Compute pod. A diagram of the ESGs and their connections to their respective leaf switches can be seen in Figure 7. ESGs with their respective leaf switches in the Management pod can be seen in Figure 23. Figure 23 also shows specific IP addresses, VLANs, and ASNs that are used. The leaf switches share the same ASN, 64701 and the ESGs share the same ASN, 65003. More detailed connection information between the ESGs and the leaf switches can be found in Table 3.

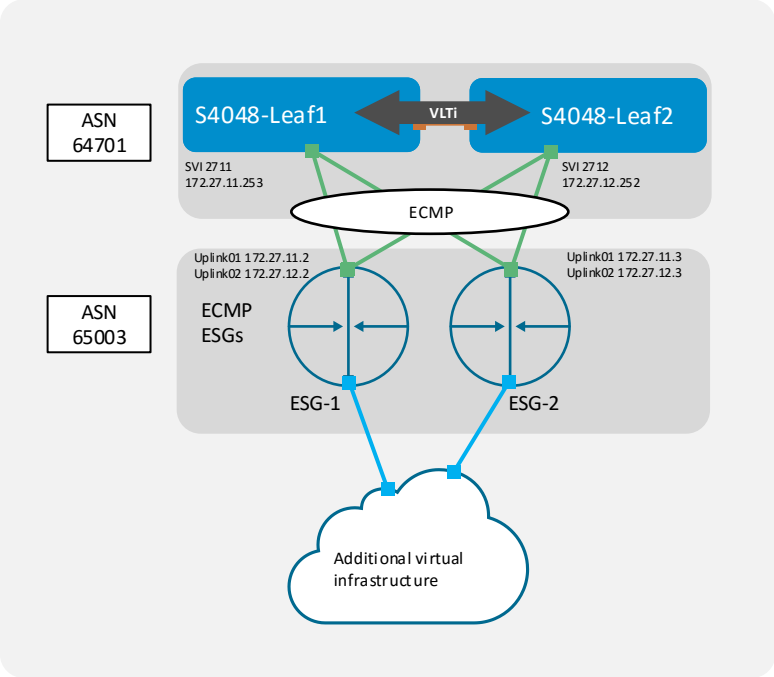

Figure 11 ESGs and leaf switches in the Management pod

Each ESG has two connections to the physical network: one to each leaf switch using the uplink networks. The ESG-1 Uplink01 interface peers to the Leaf-1 2711 VLAN interface, while the ESG-1 Uplink02 interface peers to the Leaf-2 2712 VLAN interface. The same is true for the ESG-2 uplinks. The ESG-2 Uplink01 interface peers to the Leaf-1 2711 VLAN interface, while the ESG-2 Uplink02 interface peers to the Leaf-2 2712 VLAN interface. Matching configuration is done for the Shared Edge and Compute pod. This information can be seen in Table 3.

All of the hosts in each pod share a virtual distributed switch (vDS) for interconnectivity. Each vDS has different port groups for different services, such as vMotion, vSAN, and VXLAN. For north/south connectivity, the Management pod's vDS has two port groups (Uplink01 and Uplink02), which are tagged for 2711 and 2712. These port groups map onto the uplink VLANs on the leaf switches named Uplink1 and Uplink2.

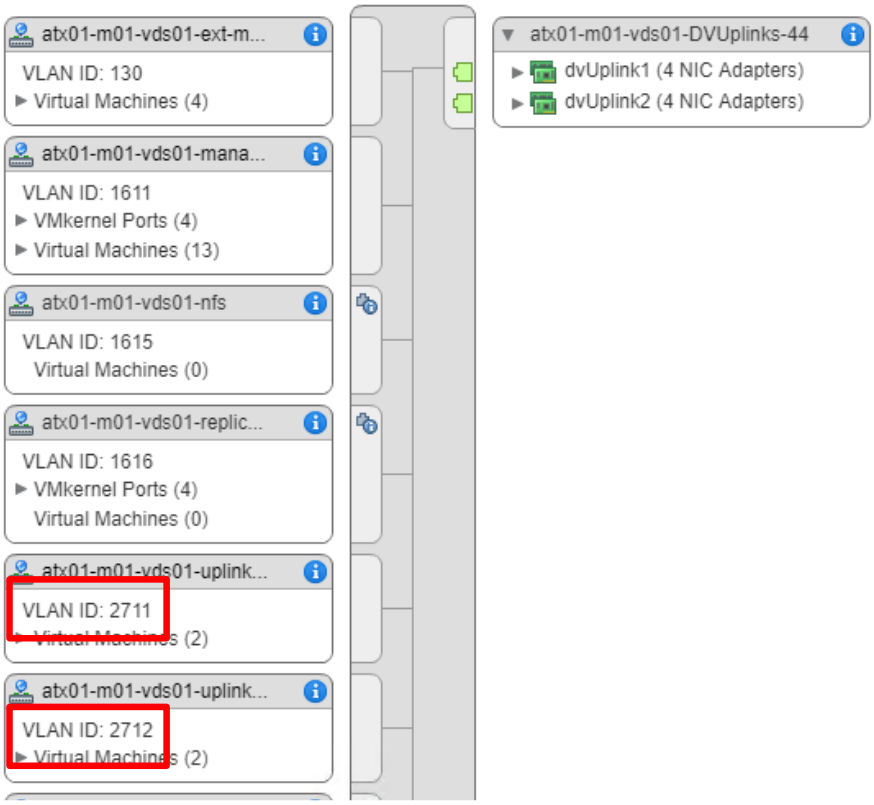

Figure 12 Topology view of the vDS in the Management pod

3.3.1 Uplink configuration

The Uplink01 port group is configured to use only dvUplink2 and the Uplink02 port group is configured to use only dvUplink1. The dvUplinks are, as the above diagram shows, how the virtual switch connects to the upstream physical switches. Those uplinks use the host vmnics. Each host uses two vmnics uplinks, so there is a total of eight connections to the leaf switches per pod.

One thing to be mindful about is how uplink port groups are configured. VVD 4.1 specifies that only one dvUplink be used for each uplink port group. Uplink1 is used to peer with Leaf 1 and Uplink2 is used to peer with Leaf 2. dvUplink2 has vmnic0 for all hosts so that is the chosen dvUplink for the Uplink1 port group. dvUplink1 has vmnic1 for all hosts so that is the chosen dvUplink for the Uplink2 port group. This way, it can be assured that vmnic0 is used to reach Leaf-1 directly, since that is how server/switch connectivity was chosen.

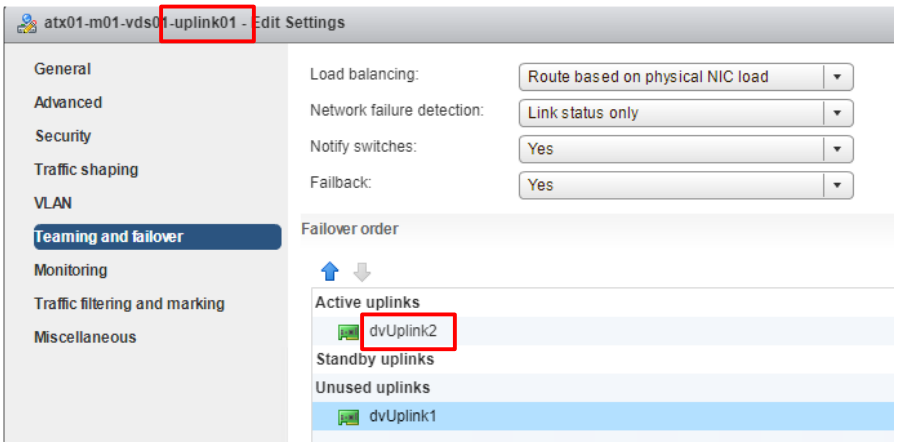

Figure 13 One distributed virtual uplink used per uplink port group

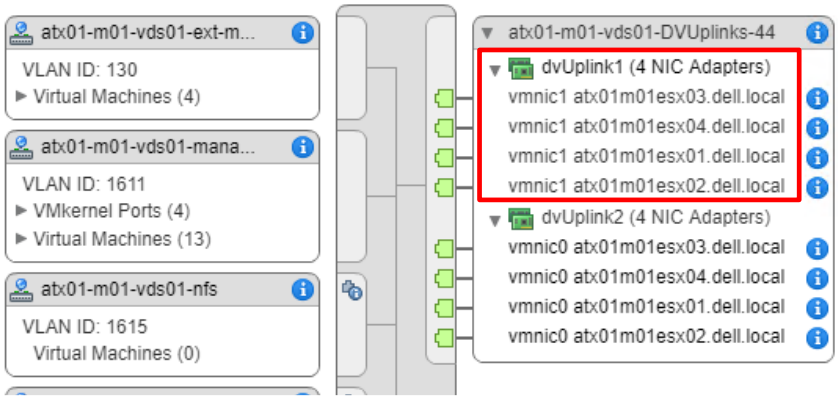

Figure 14 dvUplink1 contains vmnic1 for all hosts

3.3.2 BGP neighbor configuration

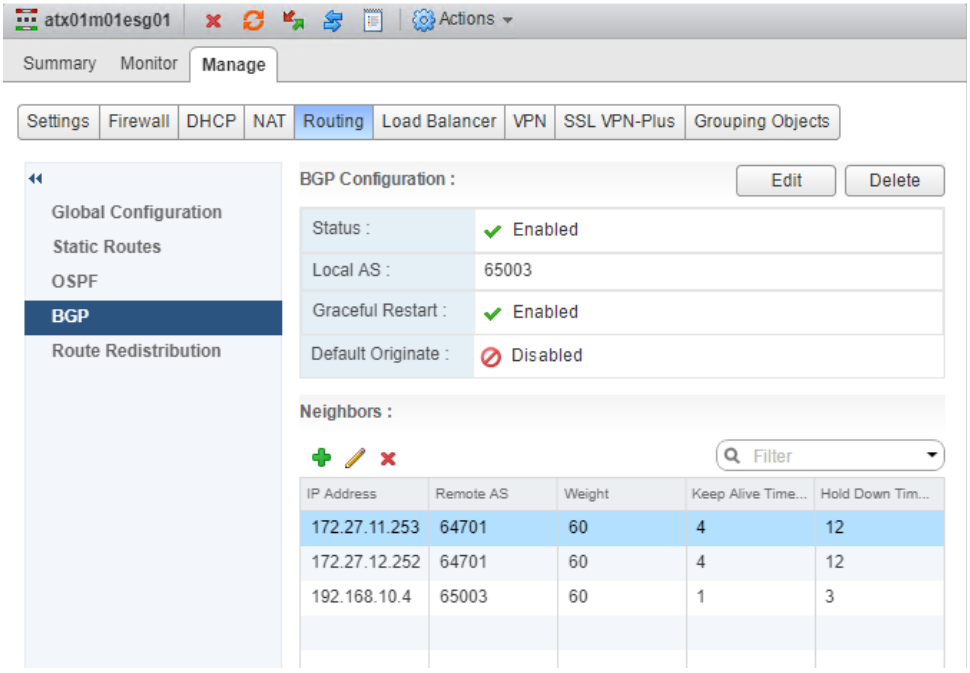

Figure 15 ESG-1 BGP neighbor configuration

As seen in Figure 15, ESG-1 peers with both upstream leaf switches with eBGP. Both uplink networks are used. Uplink01 corresponds to VLAN 2711 and Uplink02 corresponds to VLAN 2712 in the Management pod. This is reflected by the second and third octets in the IP addresses which use those networks. 172.27.11.253 refers to Leaf-1's 2711 VLAN interface and 172.27.12.252 refers to Leaf-2's 2712 VLAN interface. 192.168.10.4 refers to the interface of a downstream Universal Distributed Logical Router (UDLR), another virtual network device which is not included in the diagrams. Notice the BGP timers have been changed to four seconds for keepalive interval and 12 for hold down for the peering relationships to the leaf switches. Default timers are 60 seconds for keepalive interval and 180 seconds for hold down. This is to comply with VVD 4.1 specifications.

3.3.3 IP addressing

Table 3 shows the IP addresses used for peering between the ESGs and the physical switches in the Layer 3 environment.

Table 3 Interface and IP configuration for ESGs and leaf switches

Pod	Source switch	Source interface	Source IP	Network	ASN	Destination device	Destination interface	Destination IP	ASN
Management	Leaf 1	VLAN 2711	. 253	172.27.11.0/24	64701	ESG-1 ESG-2	Uplink01 Uplink01	.2 .3	65003 65003
Management	Leaf 1	VLAN 2712	. 253	172.27.12.0/24	64701				
Management	Leaf 2	VLAN 2711	. 252	172.27.11.0/24	64701				
Management	Leaf 2	VLAN 2712	. 252	172.27.12.0/24	64701	ESG-1 ESG-2	Uplink02 Uplink02	.2 .3	65003 65003
Shared Edge and Compute	Leaf 3	VLAN 2731	. 253	172.27.31.0/24	64702	ESG-3 ESG-4	Uplink01 Uplink01	.2 .3	65000 65000
Shared Edge and Compute	Leaf 3	VLAN 2732	. 253	172.27.32.0/24	64702				
Shared Edge and Compute	Leaf 4	VLAN 2731	. 252	172.27.31.0/24	64702				
Shared Edge and Compute	Leaf 4	VLAN 2732	. 252	172.27.32.0/24	64702	ESG-3 ESG-4	Uplink02 Uplink02	.2 .3	65000 65000

3.4 S4048-ON leaf switch configuration

The following configuration details are for S4048-Leaf1 and S4048-Leaf2 in Figure 7. The configuration commands for S4048-Leaf3 and S4048-Leaf4 are similar and are provided in the attachments.

Note: On S4048-ON switches, Telnet is enabled and SSH is disabled by default. Both services require the creation of a non-root user account to login. If needed, it is a best practice to use SSH instead of Telnet for security. SSH can optionally be enabled with the command:
`(conf)#ip ssh server enable.`
A user account can be created to access the switch via SSH with the command
`(conf)#username ssh_user sha256-password ssh_password`

First, configure the serial console enable password and disable Telnet.

S4048-Leaf1	S4048-Leaf2
enable configure enable sha256-password enable_password no ip telnet server enable	enable configure enable sha256-password enable_password no ip telnet server enable

Set the host name, configure the OOB management interface and default gateway. Enable LLDP and BFD. Enable RSTP as a precaution. S4048-Leaf1 is configured as the primary RSTP root bridge using the bridge-priority 0 command. S4048-Leaf2 is configured as the secondary RSTP root bridge using the bridge-priority 4096 command. Both leaf switches are configured to enable IGMP snooping globally, along with disallowing unregistered multicast flooding. Both leaf switches are configured with a policy map which instructs the switch to trust DSCP values in IP packets. An external NTP server is selected for time synchronization.

S4048-Leaf1	S4048-Leaf2
hostname S4048-Leaf1 interface ManagementEthernet 1/1 ip address 100.67.168.32/24 no shutdown management route 0.0.0.0/0 100.67.168.254 protocol lldp advertise management-tlv management- address system-description system-name advertise interface-port-desc bfd enable protocol spanning-tree rstp no disable bridge-priority 0 ip igmp snooping enable no ip igmp snooping flood policy-map-input TrustDSCPin trust diffserv ntp server 172.16.11.240	hostname S4048-Leaf2 interface ManagementEthernet 1/1 ip address 100.67.168.31/24 no shutdown management route 0.0.0.0/0 100.67.168.254 protocol lldp advertise management-tlv management- address system-description system-name advertise interface-port-desc bfd enable protocol spanning-tree rstp no disable bridge-priority 4096 ip igmp snooping enable no ip igmp snooping flood policy-map-input TrustDSCPin trust diffserv ntp server 172.16.11.240

Configure the VLT interconnect between S4048-Leaf1 and S4048-Leaf2. In this configuration, add interfaces fortyGigE 1/53-54 to static port channel 127 for the VLT interconnect. Enable jumbo frames and apply the DSCP policy map to the VLTi's physical interfaces. The backup destination is the management IP address of the VLT peer switch. Enable peer routing.

Note: Dell EMC recommends that the VLTi is configured as a static LAG, without LACP, per the commands shown below.

S4048-Leaf1	S4048-Leaf2
interface port-channel 127 description VLTi channel-member fortyGigE 1/53 - 1/54 mtu 9216 no shutdown interface range fortyGigE 1/53 - 1/54 description VLTi mtu 9216 service-policy input TrustDSCPin no shutdown vlt domain 127 peer-link port-channel 127 back-up destination 100.67.168.31 unit-id 0 peer-routing exit	interface port-channel 127 description VLTi channel-member fortyGigE 1/53 - 1/54 mtu 9216 no shutdown interface range fortyGigE 1/53 - 1/54 description VLTi mtu 9216 service-policy input TrustDSCPin no shutdown vlt domain 127 peer-link port-channel 127 back-up destination 100.67.168.32 unit-id 1 peer-routing exit

Configure each downstream server-facing interface with Switchport mode. Configure those interfaces as switch ports, as RSTP edge ports, apply the DSCP policy map, and enable the interfaces to accept and process jumbo frames.

S4048-Leaf1	S4048-Leaf2
interface tengigabitethernet 1/1 description Server 1 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/2 description Server 2 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/3 description Server 3 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216	interface tengigabitethernet 1/1 description Server 1 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/2 description Server 2 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/3 description Server 3 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216

no shutdown interface tengigabitethernet 1/4 description Server 4 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown	no shutdown interface tengigabitethernet 1/4 description Server 4 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown

Here, VLANs are created and applied to the interfaces. Additionally, IP addresses are applied to those VLANs. VRRP groups are created for all VLANs except the vSAN and Uplink VLANs. IGMP queriers are also configured.

S4048-Leaf1	**S4048-Leaf2**
interface Vlan 1611 description ESXi host management ip address 172.16.11.252/24 tagged te1/1 - 1/4 vrrp-group 11 description ESXi host management priority 254 virtual-address 172.16.11.253 no shutdown interface Vlan 1612 description vMotion ip address 172.16.12.252/24 tagged te1/1 - 1/4 vrrp-group 12 description vMotion priority 254 virtual-address 172.16.12.253 mtu 9216 no shutdown interface Vlan 1613 description vSAN ip address 172.16.13.252/24 tagged te1/1 - 1/4 ip igmp snooping querier ip igmp version 3 mtu 9216 no shutdown	interface Vlan 1611 description ESXi host management ip address 172.16.11.251/24 tagged te1/1 - 1/4 vrrp-group 11 description ESXi host management priority 100 virtual-address 172.16.11.253 no shutdown interface Vlan 1612 description vMotion ip address 172.16.12.251/24 tagged te1/1 - 1/4 vrrp-group 12 description vMotion priority 100 virtual-address 172.16.12.253 mtu 9216 no shutdown interface Vlan 1613 description vSAN ip address 172.16.13.251/24 tagged te1/1 - 1/4 ip igmp snooping querier ip igmp version 3 mtu 9216 no shutdown

```
interface Vlan 1614
description VXLAN
ip address 172.16.14.252/24
tagged te1/1 - 1/4
vrrp-group 14
description VXLAN
priority 254
virtual-address 172.16.14.253
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1615
description NFS
ip address 172.16.15.252/24
tagged te1/1 - 1/4
vrrp-group 15
description NFS
priority 254
virtual-address 172.16.15.253
mtu 9216
no shutdown

interface Vlan 1616
description Replication
ip address 172.16.16.252/24
tagged te1/1 - 1/4
vrrp-group 16
description Replication
priority 254
virtual-address 172.16.16.253
mtu 9216
no shutdown

interface Vlan 130
description Ext Management
ip address 10.158.130.252/24
tagged te1/1 - 1/4
vrrp-group 130
description Ext Management
priority 254
virtual-address 10.158.130.253
mtu 9216
no shutdown

interface Vlan 2711
description Uplink1
ip address 172.27.11.253/24
tagged te1/1 - 1/4
mtu 9216
```

```
interface Vlan 1614
description VXLAN
ip address 172.16.14.251/24
tagged te1/1 - 1/4
vrrp-group 14
description VXLAN
priority 100
virtual-address 172.16.14.253
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1615
description NFS
ip address 172.16.15.251/24
tagged te1/1 - 1/4
vrrp-group 15
description NFS
priority 100
virtual-address 172.16.15.253
mtu 9216
no shutdown

interface Vlan 1616
description Replication
ip address 172.16.16.251/24
tagged te1/1 - 1/4
vrrp-group 16
description Replication
priority 100
virtual-address 172.16.16.253
mtu 9216
no shutdown

interface Vlan 130
description Ext Management
ip address 10.158.130.251/24
tagged te1/1 - 1/4
vrrp-group 130
description Ext Management
priority 100
virtual-address 10.158.130.253
mtu 9216
no shutdown

interface Vlan 2711
description Uplink1
ip address 172.27.11.252/24
tagged te1/1 - 1/4
mtu 9216
```

no shutdown interface Vlan 2712 description Uplink2 ip address 172.27.12.253/24 tagged te1/1 - 1/4 mtu 9216 no shutdown	no shutdown interface Vlan 2712 description Uplink2 ip address 172.27.12.252/24 tagged te1/1 - 1/4 mtu 9216 no shutdown

The two upstream layer 3 interfaces connected to the spine switches are configured. Assign IP addresses per Table 2. Configure a loopback interface to be used as the router ID. This is used with BGP.

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.

S4048-Leaf1	**S4048-Leaf2**
interface fortyGigE 1/49 description Spine-1 ip address 192.168.1.1/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/50 description Spine-2 ip address 192.168.2.1/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface loopback 0 description Router ID ip address 10.0.2.1/32 mtu 9216 no shutdown	interface fortyGigE 1/49 description Spine-1 ip address 192.168.1.3/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/50 description Spine-2 ip address 192.168.2.3/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface loopback 0 description Router ID ip address 10.0.2.2/32 mtu 9216 no shutdown

Configure a route map and IP prefix list to redistribute all loopback addresses and leaf networks via BGP. Configure a route map and IP prefix list to filter routes to the ESGs.

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 all 172.16 networks configured on the leaf switches.

The command `seq 20 permit 172.16.0.0/16 ge 24` includes all 172.16 networks configured on all leaf switches.

The command `seq 5 permit 0.0.0.0/0` refers to no network in specific and is used by a route map to permit only a default route to be inlcuded in routing updates to the ESGs.

S4048-Leaf1	S4048-Leaf2
route-map spine-leaf permit 10 match ip address spine-leaf ip prefix-list spine-leaf description Redistribute loopback and leaf networks seq 10 permit 10.0.0.0/8 ge 24 seq 20 permit 172.16.0.0/16 ge 24 route-map default permit 10 match ip address default ip prefix-list default seq 5 permit 0.0.0.0/0	route-map spine-leaf permit 10 match ip address spine-leaf ip prefix-list spine-leaf description Redistribute loopback and leaf networks seq 10 permit 10.0.0.0/8 ge 24 seq 20 permit 172.16.0.0/16 ge 24 route-map default permit 10 match ip address default ip prefix-list default seq 5 permit 0.0.0.0/0

Include the point-to-point interfaces to each leaf pair in an ECMP group. Enable link bundle monitoring to report when traffic is unevenly distributed across multiple links.

Note: ECMP is not enabled until BGP is configured.

S4048-Leaf1	S4048-Leaf2
ecmp-group 1 interface fortyGigE 1/49 interface fortyGigE 1/50 link-bundle-monitor enable	ecmp-group 1 interface fortyGigE 1/49 interface fortyGigE 1/50 link-bundle-monitor enable

Configure UFD. This shuts down the downstream interfaces if all uplinks fail. The hosts attached to the switch uses its other port to continue sending traffic across the fabric.

Finally, exit configuration mode and save the configuration.

S4048-Leaf1	S4048-Leaf2
uplink-state-group 1 description Disable downstream ports in event all uplinks fail downstream TenGigabitEthernet 1/1-1/48	uplink-state-group 1 description Disable downstream ports in event all uplinks fail downstream TenGigabitEthernet 1/1-1/48

upstream fortyGigE 1/49,1/50 end write	upstream fortyGigE 1/49,1/50 end write

3.4.1 S4048-ON BGP configuration

Use these commands to configure BGP.

First, enable BGP with the `router bgp` *`ASN`* command. The ASN is from Figure 8.

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. This number should be equal to or greater than the number of spines, up to 64.

Four BGP neighbors are configured: two for spines and two for ESGs. All are configured for fast fallover. Unlike the spine neighbors, the ESG neighbors are not configured with BFD. The ESG neighbors also have their timers changed to four seconds for keepalive interval and 12 seconds for hold down in order to comply with VVD 4.1.

BFD settings are configured to 100 millisecond send/receive intervals. The multiplier is the number of packets that must be missed to declare a session down.

Finally, exit configuration mode and save the configuration with the `end` and `write` commands.

S4048-Leaf1	**S4048-Leaf2**
enable configure router bgp 64701 bgp bestpath as-path multipath-relax maximum-paths ebgp 2 redistribute connected route-map spine-leaf bgp graceful-restart neighbor spine-leaf peer-group neighbor spine-leaf fall-over neighbor spine-leaf advertisement- interval 1 neighbor spine-leaf no shutdown neighbor spine-leaf bfd neighbor spine-leaf password Dell!234 neighbor esg peer-group neighbor esg default-originate neighbor esg route-map default out	enable configure router bgp 64701 bgp bestpath as-path multipath-relax maximum-paths ebgp 2 redistribute connected route-map spine-leaf bgp graceful-restart neighbor spine-leaf peer-group neighbor spine-leaf fall-over neighbor spine-leaf advertisement- interval 1 neighbor spine-leaf no shutdown neighbor spine-leaf bfd neighbor spine-leaf password Dell!234 neighbor esg peer-group neighbor esg default-originate neighbor esg route-map default out

neighbor esg fall-over neighbor esg no shutdown neighbor esg password Dell!234 neighbor esg bfd disable neighbor esg timers 4 12 neighbor 192.168.1.0 remote-as 64601 neighbor 192.168.1.0 peer-group spine-leaf neighbor 192.168.1.0 no shutdown neighbor 192.168.2.0 remote-as 64602 neighbor 192.168.2.0 peer-group spine-leaf neighbor 192.168.2.0 no shutdown neighbor 172.27.11.2 remote-as 65003 neighbor 172.27.11.2 peer-group esg neighbor 172.27.11.2 no shutdown neighbor 172.27.11.3 remote-as 65003 neighbor 172.27.11.3 peer-group esg neighbor 172.27.11.3 no shutdown bfd all-neighbors interval 100 min_rx 100 multiplier 3 role active end write	neighbor esg fall-over neighbor esg no shutdown neighbor esg password Dell!234 neighbor esg bfd disable neighbor esg timers 4 12 neighbor 192.168.1.2 remote-as 64601 neighbor 192.168.1.2 peer-group spine-leaf neighbor 192.168.1.2 no shutdown neighbor 192.168.2.2 remote-as 64602 neighbor 192.168.2.2 peer-group spine-leaf neighbor 192.168.2.2 no shutdown neighbor 172.27.12.2 remote-as 65003 neighbor 172.27.12.2 peer-group esg neighbor 172.27.12.2 no shutdown neighbor 172.27.12.3 remote-as 65003 neighbor 172.27.12.3 peer-group esg neighbor 172.27.12.3 no shutdown bfd all-neighbors interval 100 min_rx 100 multiplier 3 role active end write

3.5 Z9100-ON spine switch configuration

The following configuration details are for Z9100-Spine1 and Z9100-Spine2 in Figure 7.

Note: On Z9100-ON switches, Telnet is enabled and SSH is disabled by default. Both services require the creation of a non-root user account to login. If needed, it is a best practice to use SSH instead of Telnet for security. SSH can optionally be enabled with the command:
`(conf)#`**`ip ssh server enable.`**
A user account can be created to access the switch via SSH with the command
`(conf)`**`#username ssh_user sha256-password ssh_password`**

First, configure the serial console enable password and disable Telnet.

Z9100-Spine1	Z9100-Spine2
enable configure enable sha256-password enable_password no ip telnet server enable	enable configure enable sha256-password enable_password no ip telnet server enable

Set the host name, configure the OOB management interface and default gateway. Enable LLDP and BFD. Set the port speed of the four ports connected to the leaf switches to 40GbE and create the DSCP policy map. The designated Management leaf switches are pointed at for time synchronization.

Z9100-Spine1	Z9100-Spine2
hostname Z9100-Spine1 interface ManagementEthernet 1/1 ip address 100.67.168.36/24 no shutdown management route 0.0.0.0/0 100.67.168.254 protocol lldp advertise management-tlv management- address system-description system-name advertise interface-port-desc bfd enable stack-unit 1 port 1 portmode single speed 40G no-confirm stack-unit 1 port 2 portmode single speed 40G no-confirm stack-unit 1 port 3 portmode single speed 40G no-confirm stack-unit 1 port 4 portmode single speed 40G no-confirm policy-map-input TrustDSCPin trust diffserv ntp server 172.16.11.252 ntp server 172.16.11.251	hostname Z9100-Spine2 interface ManagementEthernet 1/1 ip address 100.67.168.35/24 no shutdown management route 0.0.0.0/0 100.67.168.254 protocol lldp advertise management-tlv management- address system-description system-name advertise interface-port-desc bfd enable stack-unit 1 port 1 portmode single speed 40G no-confirm stack-unit 1 port 2 portmode single speed 40G no-confirm stack-unit 1 port 3 portmode single speed 40G no-confirm stack-unit 1 port 4 portmode single speed 40G no-confirm policy-map-input TrustDSCPin trust diffserv ntp server 172.16.11.252 ntp server 172.16.11.251

Configure the four point-to-point interfaces connected to leaf switches. Assign IP addresses per Table 2. Configure a loopback interface to be used as the router ID. This is used with BGP.

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 fortyGigE 1/1/1	interface fortyGigE 1/1/1

description Leaf 1 fo1/49 ip address 192.168.1.0/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/2/1 description Leaf 2 fo1/49 ip address 192.168.1.2/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/3/1 description Leaf 3 fo1/49 ip address 192.168.1.4/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/4/1 description Leaf 4 fo1/49 ip address 192.168.1.6/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface loopback 0 description Router ID ip address 10.0.1.1/32 mtu 9216 no shutdown	description Leaf 1 fo1/50 ip address 192.168.2.0/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/2/1 description Leaf 2 fo1/50 ip address 192.168.2.2/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/3/1 description Leaf 3 fo1/50 ip address 192.168.2.4/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/4/1 description Leaf 4 fo1/50 ip address 192.168.2.6/31 service-policy input TrustDSCPin mtu 9216 no shutdown interface loopback 0 description Router ID ip address 10.0.1.2/32 mtu 9216 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 7.

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 7.

Z9100-Spine1	**Z9100-Spine2**
route-map spine-leaf permit 10 match ip address spine-leaf ip prefix-list spine-leaf	route-map spine-leaf permit 10 match ip address spine-leaf ip prefix-list spine-leaf

description Redistribute loopback and leaf networks seq 10 permit 10.0.0.0/8 ge 24 seq 20 permit 172.16.0.0/16 ge 24	description Redistribute loopback and leaf networks seq 10 permit 10.0.0.0/8 ge 24 seq 20 permit 172.16.0.0/16 ge 24

Include the point-to-point interfaces to each leaf pair in an ECMP group. Enable link bundle monitoring to report when traffic is unevenly distributed across multiple links.

Note: ECMP is not actually enabled until BGP is configured.

Finally, exit configuration mode and save the configuration with the end and write commands.

Z9100-Spine1	Z9100-Spine2
ecmp-group 1 interface fortyGigE 1/1/1 interface fortyGigE 1/2/1 link-bundle-monitor enable ecmp-group 2 interface fortyGigE 1/3/1 interface fortyGigE 1/4/1 link-bundle-monitor enable end write	ecmp-group 1 interface fortyGigE 1/1/1 interface fortyGigE 1/2/1 link-bundle-monitor enable ecmp-group 2 interface fortyGigE 1/3/1 interface fortyGigE 1/4/1 link-bundle-monitor enable end write

3.5.1 Z9100-ON BGP configuration

Use these commands to configure BGP.

First, enabled BGP with the `router bgp` *`ASN`* command. The ASN is from Figure 8.

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 switch that the host is connected.

BGP neighbors are configured and fast fallover is enabled.

BFD settings are configured to 100 millisecond send/receive intervals. The multiplier is the number of packets that must be missed to declare a session down. Finally, exit configuration mode and save the configuration.

Z9100-Spine1	Z9100-Spine2
enable configure router bgp 64601 bgp bestpath as-path multipath-relax maximum-paths ebgp 2 redistribute connected route-map spine-leaf bgp graceful-restart neighbor spine-leaf peer-group neighbor spine-leaf fall-over neighbor spine-leaf advertisement-interval 1 neighbor spine-leaf no shutdown neighbor spine-leaf bfd neighbor 192.168.1.1 remote-as 64701 neighbor 192.168.1.1 peer-group spine-leaf neighbor 192.168.1.1 no shutdown neighbor 192.168.1.3 remote-as 64702 neighbor 192.168.1.3 peer-group spine-leaf neighbor 192.168.1.3 no shutdown neighbor 192.168.1.5 remote-as 64703 neighbor 192.168.1.5 peer-group spine-leaf neighbor 192.168.1.5 no shutdown neighbor 192.168.1.7 remote-as 64704 neighbor 192.168.1.7 peer-group spine-leaf neighbor 192.168.1.7 no shutdown bfd all-neighbors interval 100 min_rx 100 multiplier 3 role active end write	enable configure router bgp 64602 bgp bestpath as-path multipath-relax maximum-paths ebgp 2 redistribute connected route-map spine-leaf bgp graceful-restart neighbor spine-leaf peer-group neighbor spine-leaf fall-over neighbor spine-leaf advertisement-interval 1 neighbor spine-leaf no shutdown neighbor spine-leaf bfd neighbor 192.168.2.1 remote-as 64701 neighbor 192.168.2.1 peer-group spine-leaf neighbor 192.168.2.1 no shutdown neighbor 192.168.2.3 remote-as 64702 neighbor 192.168.2.3 peer-group spine-leaf neighbor 192.168.2.3 no shutdown neighbor 192.168.2.5 remote-as 64703 neighbor 192.168.2.5 peer-group spine-leaf neighbor 192.168.2.5 no shutdown neighbor 192.168.2.7 remote-as 64704 neighbor 192.168.2.7 peer-group spine-leaf neighbor 192.168.2.7 no shutdown bfd all-neighbors interval 100 min_rx 100 multiplier 3 role active end write

3.6 Example 1 validation

In addition to sending traffic between hosts, the configuration shown in Figure 7 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 switch and one leaf switch. Command output on other switches is similar.

3.6.1 show ip bgp summary

The `show ip route bgp summary` command shows the status of all BGP connections. Each spine switch has four neighbors (the four leaf switches) and each leaf switch has two neighbors (the two spine switches). This command also confirms BFD is enabled on the 6th line of output.

```
Z9100-Spine-1#show ip bgp summary
BGP router identifier 10.0.1.1, local AS number 64601
BGP local RIB : Routes to be Added 0, Replaced 0, Withdrawn 0
32 network entrie(s) using 2432 bytes of memory
64 paths using 6912 bytes of memory
BGP-RIB over all using 6976 bytes of memory
BFD is enabled, Interval 100 Min_rx 100 Multiplier 3 Role Active
37 BGP path attribute entrie(s) using 6192 bytes of memory
35 BGP AS-PATH entrie(s) using 350 bytes of memory
4 neighbor(s) using 32768 bytes of memory

Neighbor        AS     MsgRcvd  MsgSent     TblVer  InQ  OutQ Up/Down  State/Pfx
192.168.1.1     64701        7066     7044          0    0     0 4d:05:49:58 16
192.168.1.3     64701        7058     7041          0    0     0 4d:05:49:57 17
192.168.1.5     64702        7015     7004          0    0     0 4d:04:40:35 15
192.168.1.7     64702        7057     7045          0    0     0 4d:05:17:54 15


S4048-Leaf-1#show ip bgp summary
BGP router identifier 10.0.2.1, local AS number 64701
BGP local RIB : Routes to be Added 0, Replaced 0, Withdrawn 0
31 network entrie(s) using 2356 bytes of memory
52 paths using 5616 bytes of memory
BGP-RIB over all using 5668 bytes of memory
BFD is enabled, Interval 100 Min_rx 100 Multiplier 3 Role Active
20 BGP path attribute entrie(s) using 3268 bytes of memory
18 BGP AS-PATH entrie(s) using 180 bytes of memory
4 neighbor(s) using 32768 bytes of memory

Neighbor        AS     MsgRcvd  MsgSent     TblVer  InQ  OutQ Up/Down  State/Pfx
172.27.11.2     65003      105892   105588          0    0     0 4d:05:47:12 6
172.27.11.3     65003      105836   105584          0    0     0 4d:05:47:10 6
192.168.1.0     64601        7044     7066          0    0     0 4d:05:49:53 16
192.168.2.0     64602        7046     7060          0    0     0 4d:05:49:50 16
```

3.6.2 show ip route bgp

The `show ip route bgp` command validates the BGP entries in the Routing Information Base (RIB). Entries with multiple paths shown are used with ECMP. Leaf and spine networks for both Management and Shared Edge and Compute pods can be seen in this example (1611, 1612, 1613, etc.), as well as networks in the virtual infrastructure (192.168.10, 192.168.11, 192.168.100, 192.168.101 networks).

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

```
Z9100-Spine1#show ip route bgp
       Destination        Gateway                    Dist/Metric Last Change
       -----------        -------                    ----------- -----------
  B EX 10.0.1.2/32        via 192.168.1.1                  20/0        4d6h
                          via 192.168.1.3
  B EX 10.0.2.1/32        via 192.168.1.1                  20/0        4d4h
  B EX 10.0.2.2/32        via 192.168.1.3                  20/0        4d6h
  B EX 10.0.2.3/32        via 192.168.1.5                  20/0        4d5h
  B EX 10.0.2.4/32        via 192.168.1.7                  20/0        4d6h
  B EX 10.100.0.0/24      via 192.168.1.5                  20/0        4d5h
                          via 192.168.1.7
  B EX 10.100.1.0/24      via 192.168.1.5                  20/0        4d5h
                          via 192.168.1.7
  B EX 10.158.130.0/24    via 192.168.1.1                  20/0        4d6h
                          via 192.168.1.3
  B EX 10.158.150.0/24    via 192.168.1.5                  20/0        4d5h
                          via 192.168.1.7
  B EX 172.16.11.0/24     via 192.168.1.1                  20/0        4d6h
                          via 192.168.1.3
  B EX 172.16.12.0/24     via 192.168.1.1                  20/0        4d6h
                          via 192.168.1.3
  B EX 172.16.13.0/24     via 192.168.1.1                  20/0        4d6h
                          via 192.168.1.3
  B EX 172.16.14.0/24     via 192.168.1.1                  20/0        4d6h
                          via 192.168.1.3
  B EX 172.16.15.0/24     via 192.168.1.1                  20/0        4d6h
                          via 192.168.1.3
  B EX 172.16.16.0/24     via 192.168.1.1                  20/0        4d6h
                          via 192.168.1.3
  B EX 172.16.31.0/24     via 192.168.1.5                  20/0        4d5h
                          via 192.168.1.7
  B EX 172.16.32.0/24     via 192.168.1.5                  20/0        4d5h
                          via 192.168.1.7
  B EX 172.16.33.0/24     via 192.168.1.5                  20/0        4d5h
                          via 192.168.1.7
  B EX 172.16.34.0/24     via 192.168.1.5                  20/0        4d5h
                          via 192.168.1.7
```

```
  B EX 172.16.35.0/24     via 192.168.1.5                    20/0        4d5h
                          via 192.168.1.7
  B EX 172.16.36.0/24     via 192.168.1.5                    20/0        4d5h
                          via 192.168.1.7
  B EX 172.27.11.0/24     via 192.168.1.1                    20/0        4d6h
                          via 192.168.1.3
  B EX 172.27.12.0/24     via 192.168.1.1                    20/0        4d6h
                          via 192.168.1.3
  B EX 172.27.31.0/24     via 192.168.1.5                    20/0        4d5h
                          via 192.168.1.7
  B EX 172.27.32.0/24     via 192.168.1.5                    20/0        4d5h
                          via 192.168.1.7
  B EX 192.168.10.0/24    via 192.168.1.1                    20/0        4d6h
                          via 192.168.1.3
  B EX 192.168.11.0/24    via 192.168.1.1                    20/0        4d6h
                          via 192.168.1.3
  B EX 192.168.31.0/24    via 192.168.1.1                    20/0        4d6h
                          via 192.168.1.3
  B EX 192.168.32.0/24    via 192.168.1.1                    20/0        4d6h
                          via 192.168.1.3
  B EX 192.168.100.0/24   via 192.168.1.5                    20/0        4d5h
                          via 192.168.1.7
  B EX 192.168.101.0/24   via 192.168.1.5                    20/0        4d5h
                          via 192.168.1.7
```

Z9100-Spine1#S4048-Leaf-1 has two paths to all other leaf switches and two paths to the virtual infrastructure networks. There is one path through each spine switch. If all paths do not appear, make sure the `maximum-paths` statement in the BGP configuration is equal to or greater than the number of spine switches in the topology.

```
S4048-Leaf-1#show ip route bgp
       Destination        Gateway                    Dist/Metric Last Change
       -----------        -------                    ----------- -----------
  B EX 10.0.1.1/32        via 192.168.1.0                   20/0        4d6h
  B EX 10.0.1.2/32        via 192.168.2.0                   20/0        4d6h
  B EX 10.0.2.3/32        via 192.168.1.0                   20/0        4d5h
                          via 192.168.2.0
  B EX 10.0.2.4/32        via 192.168.1.0                   20/0        4d6h
                          via 192.168.2.0
  B EX 10.100.0.0/24      via 192.168.1.0                   20/0        4d6h
                          via 192.168.2.0
  B EX 10.100.1.0/24      via 192.168.1.0                   20/0        4d6h
                          via 192.168.2.0
```

```
B EX 10.158.150.0/24    via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 172.16.31.0/24     via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 172.16.32.0/24     via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 172.16.33.0/24     via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 172.16.34.0/24     via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 172.16.35.0/24     via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 172.16.36.0/24     via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 172.27.31.0/24     via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 172.27.32.0/24     via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 192.168.10.0/24    via 172.27.11.2                    20/0        4d6h
                        via 172.27.11.3
B EX 192.168.11.0/24    via 172.27.11.2                    20/0        4d6h
                        via 172.27.11.3
B EX 192.168.31.0/24    via 172.27.11.2                    20/0        4d6h
                        via 172.27.11.3
B EX 192.168.32.0/24    via 172.27.11.2                    20/0        4d6h
                        via 172.27.11.3
B EX 192.168.100.0/24   via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
B EX 192.168.101.0/24   via 192.168.1.0                    20/0        4d6h
                        via 192.168.2.0
```

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

3.6.3 show vrrp brief

The `show vrrp brief` command validates the status of the configured VRRP groups.

```
S4048-Leaf-1#show vrrp brief
Interface   Group    Pri Pre State  Master addr     Virtual addr(s)
Description
-----------------------------------------------------------------------------
-
Vl 130      IPv4 130 254 Y   Master 10.158.130.252  10.158.130.253
Ext Management
```

```
Vl 1611     IPv4 11  254 Y   Master 172.16.11.252   172.16.11.253
ESXi host management
Vl 1612     IPv4 12  254 Y   Master 172.16.12.252   172.16.12.253
vMotion
Vl 1614     IPv4 14  254 Y   Master 172.16.14.252   172.16.14.253
VXLAN
Vl 1615     IPv4 15  254 Y   Master 172.16.15.252   172.16.15.253
NFS
Vl 1616     IPv4 16  254 Y   Master 172.16.16.252   172.16.16.253
Replication

S4048-Leaf-2#sh vrrp brief
Interface   Group    Pri Pre State  Master addr     Virtual addr(s)
Description
-------------------------------------------------------------------------------
-
Vl 130      IPv4 130 100 Y   Backup 10.158.130.252  10.158.130.253
Ext Management
Vl 1611     IPv4 11  100 Y   Backup 172.16.11.252   172.16.11.253
ESXi host management
Vl 1612     IPv4 12  100 Y   Backup 172.16.12.252   172.16.12.253
vMotion
Vl 1614     IPv4 14  100 Y   Backup 172.16.14.252   172.16.14.253
VXLAN
Vl 1615     IPv4 15  100 Y   Backup 172.16.15.252   172.16.15.253
NFS
Vl 1616     IPv4 16  100 Y   Backup 172.16.16.252   172.16.16.253
Replication
```

3.6.4 show bfd neighbors

The `show bfd neighbors` command validates whether BFD is properly configured and sessions are established as indicated by Up in the State column.

Note: The output shown below is for BGP configurations as indicated by a B in the Clients column. On OSPF configurations, the output is identical except there is an O in the Clients column.

```
Z9100-Spine-1#show bfd neighbors

*       - Active session role
B       - BGP
O       - OSPF
```

```
  LocalAddr      RemoteAddr          Interface  State Rx-int Tx-int Mult Clients
* 192.168.1.0    192.168.1.1         Fo 1/1/1   Up    100    100    3    B
* 192.168.1.2    192.168.1.3         Fo 1/2/1   Up    100    100    3    B
* 192.168.1.4    192.168.1.5         Fo 1/3/1   Up    100    100    3    B
* 192.168.1.6    192.168.1.7         Fo 1/4/1   Up    100    100    3    B

S4048-Leaf-1#show bfd neighbors

*       - Active session role
B       - BGP
O       - OSPF

  LocalAddr       RemoteAddr          Interface  State Rx-int Tx-int Mult Clients
* 192.168.1.1     192.168.1.0         Fo 1/49    Up    100    100    3    B
* 192.168.2.1     192.168.2.0         Fo 1/50    Up    100    100    3    B
```

3.6.5 show vlt detail

The `show vlt detail` command validates VLT LAG status on leaf switches in this topology. This command shows the status and active VLANs of all VLT LAGs (Port channel 1 in this example). The local and peer status must both be up.

```
S4048-Leaf-1#show vlt detail
Local LAG Id  Peer LAG Id  Local Status  Peer Status  Active VLANs
------------  -----------  ------------  -----------  -------------
1             1            UP            UP            130, 1611-1616, 2711-2712
```

3.6.6 show uplink-state-group

The `show uplink-state-group` command validates the UFD status on leaf switches in this topology. Status: Enabled, Up indicates UFD is enabled and no interfaces are currently disabled by UFD.

```
S4048-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
```

Note: When an interface has been disabled by UFD, the show interfaces *interface* command for affected interfaces indicates it is error-disabled as follows:

```
S4048-Leaf-1#show interfaces te 1/4
TenGigabitEthernet 1/4 is up, line protocol is down(error-disabled[UFD])
```

-- Output truncated –-

3.6.7 show spanning-tree rstp brief

The `show spanning-tree rstp brief` 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 (S4048-Leaf-1 in this example) is the root bridge and sever-facing interface are edge ports.

```
S4048-Leaf-1#show spanning-tree rstp brief
Executing IEEE compatible Spanning Tree Protocol
Root ID    Priority 0, Address 1418.77e0.8931
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID    Priority 0, Address 1418.77e0.8931
We are the root
Configured hello time 2, max age 20, forward delay 15

Interface                                                    Designated
 Name       PortID   Prio Cost    Sts          Cost        Bridge ID         PortID
---------- -------- ---- ------- ----------- ------- -------------------- ------
--
Po 127     128.128  128  600     FWD(vltI)    0       0     1418.77e0.8931
128.128
Te 1/1     128.202  128  2000    FWD          0       0     1418.77e0.8931
128.202
Te 1/2     128.203  128  2000    FWD          0       0     1418.77e0.8931
128.203
Te 1/3     128.204  128  2000    FWD          0       0     1418.77e0.8931
128.204
Te 1/4     128.205  128  2000    FWD          0       0     1418.77e0.8931
128.205
Te 1/17    128.218  128  2000    FWD          0       0     1418.77e0.8931
128.218

Interface
 Name       Role   PortID   Prio Cost    Sts          Cost    Link-type Edge
---------- ------ -------- ---- ------- ----------- ------- --------- ----
Po 127     Desg   128.128  128  600     FWD          0       (vltI)P2P No
Te 1/1     Desg   128.202  128  2000    FWD          0       P2P       Yes
Te 1/2     Desg   128.203  128  2000    FWD          0       P2P       Yes
Te 1/3     Desg   128.204  128  2000    FWD          0       P2P       Yes
Te 1/4     Desg   128.205  128  2000    FWD          0       P2P       Yes
Te 1/17    Desg   128.218  128  2000    FWD          0       P2P       No
```

3.6.8 Validate VTEP functionality

NSX has an in-built tool for validating VTEP functionality.

From the vSphere web client:

1. Go to **Home > Networking & Security > Logical Switches**.
2. In the center pane, double click the Universal Transit Network logical switch.
3. Select source and destination hosts.
4. Click **Start Test**.

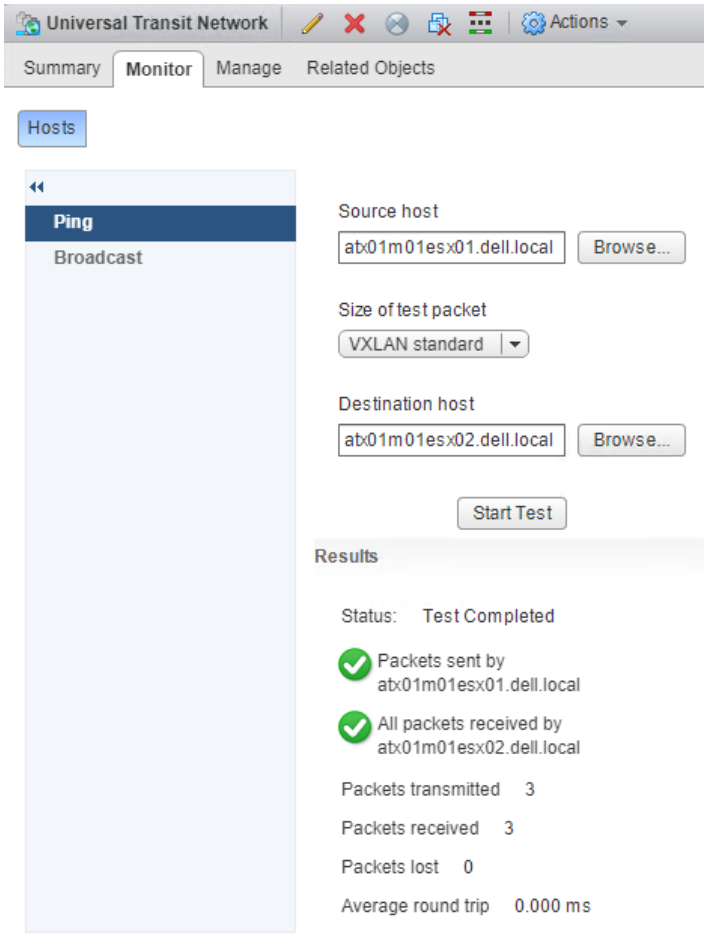

Figure 16 VTEP test

3.6.9 Validate VM-to-VM layer 2 VXLAN functionality

In this example, layer 2 functionality is tested between two VMs on different hosts.

Two Linux VMs are first migrated to two different hosts: atx01m01esx01 for lubuntu1, and atx01m01esx02 for lubuntu2 in the Management pod.

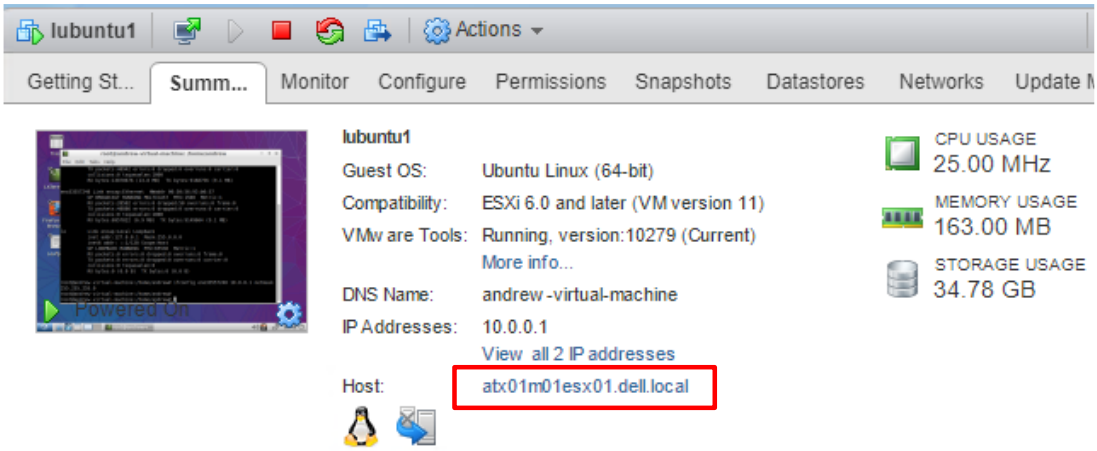

Figure 17 Lubuntu1 migrated to atx01m01esx01

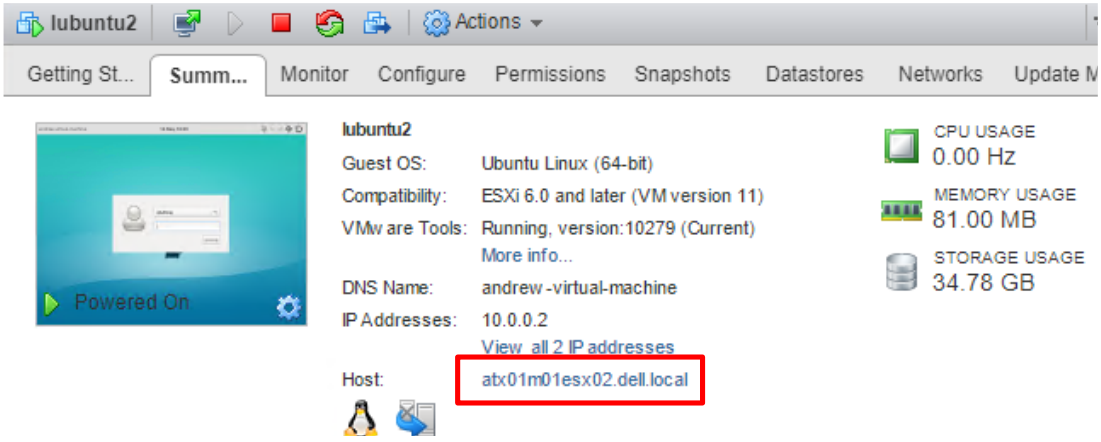

Figure 18 Lubuntu2 migrated to atx01m01esx02

Each VM is given an extra NIC and that NIC is attached to the Universal Transit Network universal logical switch (port group is **vwx-dvs-44-universalwire-1-sid-3000 Universal Transit Network**).

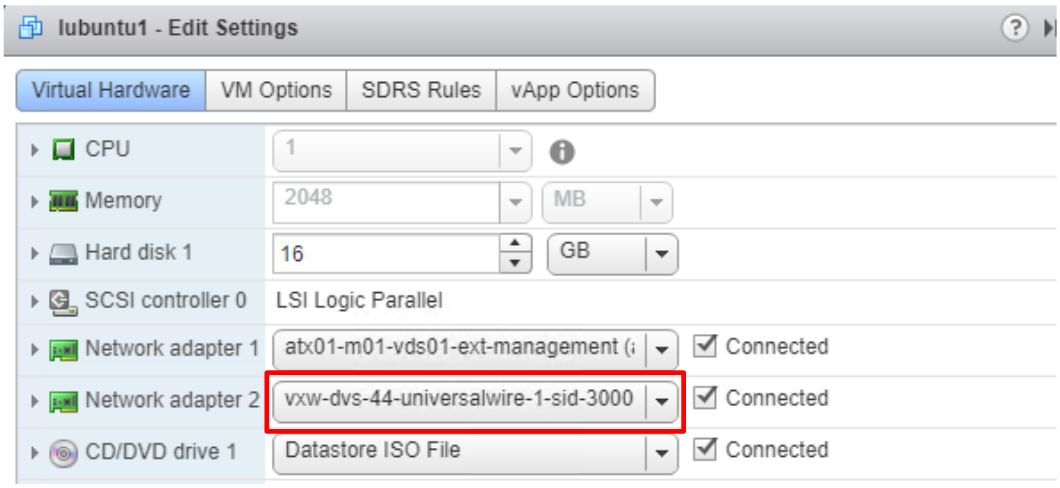

Figure 19 Lubuntu1 is given an extra NIC

Once both VMs are connected to the Universal Transit Network universal logical switch, they are given IP addresses in the same subnet. Lubuntu1 is given the IP 10.0.0.1 and lubuntu2 is given the IP 10.0.0.2. Then a ping test is issued.

```
          collisions:0 txqueuelen:1000
          RX bytes:13884014 (13.8 MB)  TX bytes:9179460 (9.1 MB)

eno33557248 Link encap:Ethernet  HWaddr 00:50:56:93:b0:57
          inet addr:10.0.0.1  Bcast:10.0.0.255  Mask:255.255.255.0
          inet6 addr: fe80::250:56ff:fe93:b057/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:20508 errors:0 dropped:50 overruns:0 frame:0
          TX packets:48900 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:6958296 (6.9 MB)  TX bytes:9150672 (9.1 MB)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          inet6 addr: ::1/128 Scope:Host
          UP LOOPBACK RUNNING  MTU:65536  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:0 (0.0 B)  TX bytes:0 (0.0 B)

root@andrew-virtual-machine:/home/andrew# ping 10.0.0.2
PING 10.0.0.2 (10.0.0.2) 56(84) bytes of data.
64 bytes from 10.0.0.2: icmp_seq=1 ttl=64 time=0.343 ms
64 bytes from 10.0.0.2: icmp_seq=2 ttl=64 time=0.247 ms
^C
--- 10.0.0.2 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 999ms
rtt min/avg/max/mdev = 0.247/0.295/0.343/0.048 ms
root@andrew-virtual-machine:/home/andrew#
```

Figure 20 Successful ping test between lubuntu1 and lubuntu2

Packet captures of the above network activity reveal a few things about how VTEPs encapsulate traffic when using Hybrid or Multicast modes. The first picture highlights an ARP packet. Since ARP is BUM traffic (it is broadcast), it is encapsulated with multicast. Once lubuntu1 resolves lubuntu2's MAC address, it proceeds to send unicast ping packets to lubuntu2, which are not BUM traffic. Since these packets are not BUM, they are encapsulated with unicast, in which the source and destination IP addresses are the IP addresses of the VTEP kernel ports for the hosts which contain the VMs. These are the tunnel endpoints.

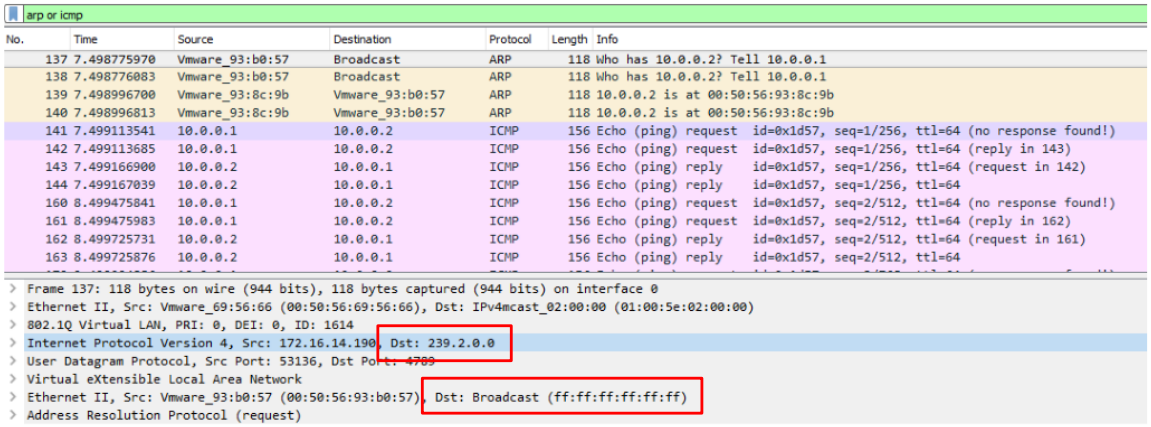

Figure 21 Selected ARP packet is BUM traffic encapsulated with multicast

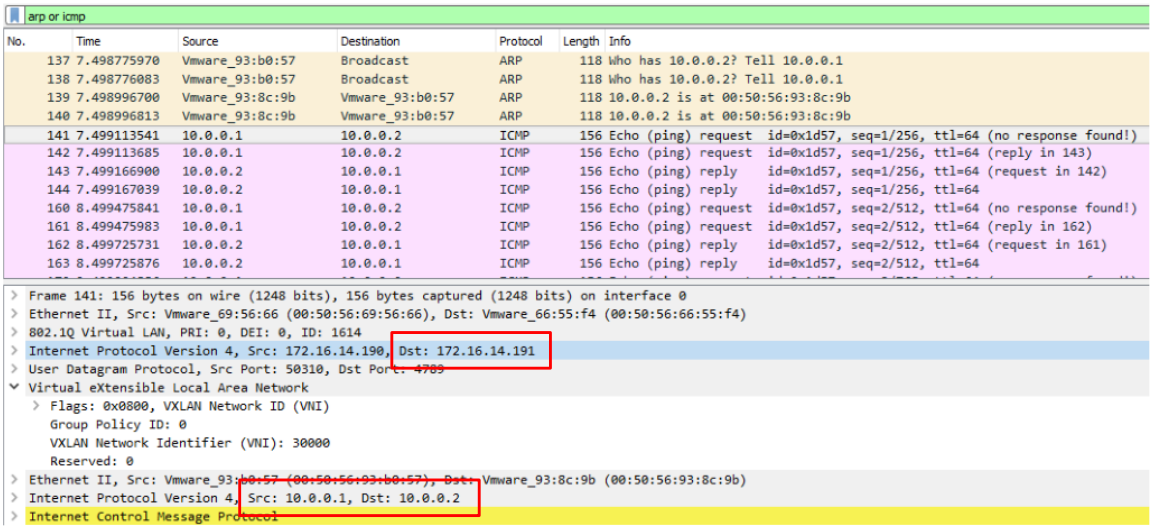

Figure 22 Selected ping packet is not BUM traffic, so it is encapsulated with unicast

4 Example 2: Layer 2/3 boundary extended to the spine switches

In this example the layer 2/3 boundary of the leaf and spine network is extended to the spine switches. There are a number of advantages to extending the layer 2/3 boundary to the spine layer. Perhaps the single greatest reason is simplicity. Routing only has to be configured on the spine switches. Frames are switched straight from the ESGs through the leaf switches and to the spine switches. Extending the layer 2/3 boundary to the spine layer also allows for some additional design freedom since VLANs and broadcast domains that are accessible to the ESGs are accessible to the spine switches. On the other hand, this type of configuration is not as scalable as the configuration detailed in Example 1.

In this scenario, each pod has its own set of VLANs for different vSphere services configured on its own leaf switches and spine switches, but not the other leaf pair. All VLANs are propagated to the spine layer. The spine layer provides connectivity between the pods via connected and eBGP routes.

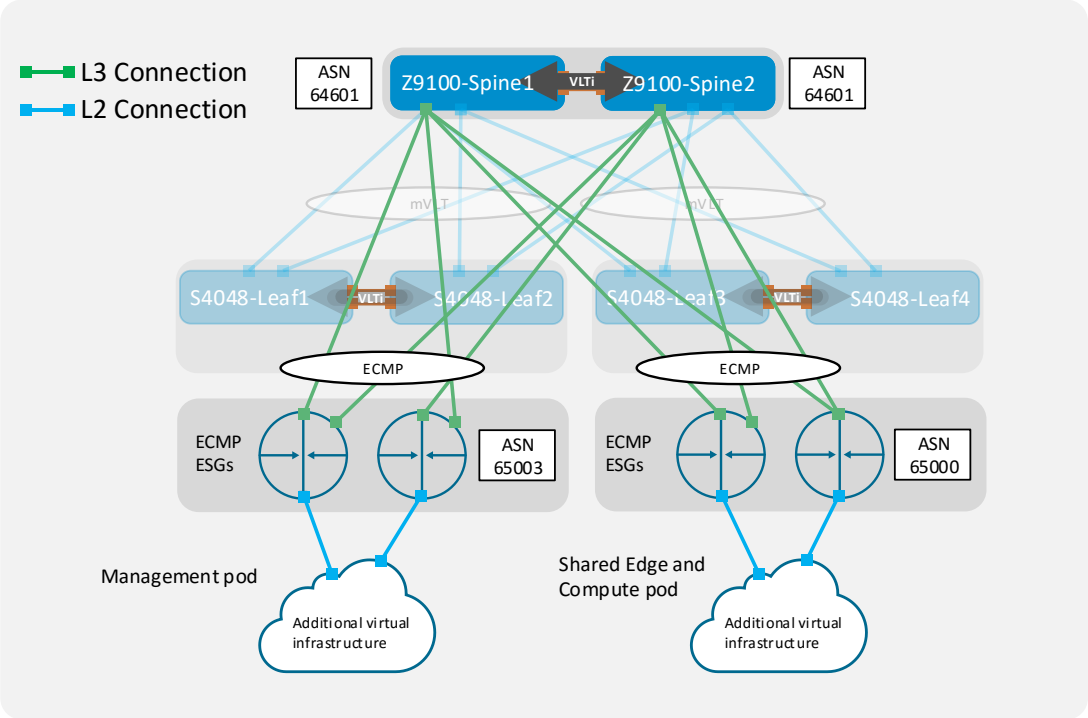

Figure 23 Leaf-spine topology with layer 2/3 boundary extended to the spine switches

Note: All switch configuration files for the topology in Figure 23 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 C.

4.1 Design decisions

The following is a list of design decisions on top of the leaf and spine design which were made in order to support the VVD 4.1 environment. Leaf and spine design decisions that are not specific to supporting the VVD 4.1 environment can be found in the *Leaf-Spine Deployment and Best Practices Guide*.

Table 4 Design decisions

Design decision	
Jumbo frames	Jumbo frames are configured on all ports and VLANs on the switches because many vSphere services require jumbo frames. These services include vSAN, VXLAN, NFS, vSphere vMotion, and vSphere Replication.
QoS	All switch ports are configured to trust the DSCP markings inside the IP packets. The idea is to configure hosts and end stations to mark management traffic with a comparatively high DSCP value - 46 is common. When the switches intercept those packets, they honor those DSCP markings and assign the packets to a higher priority queue, and provide greater assurance that management traffic continues to flow if there is network contention.
NTP	In order to comply with VVD 4.1, a pair of infrastructure switches are synchronized with an NTP server that is external to the environment. Then all hosts and other network devices are synchronized to those switches. In Example 1, the leaf switches in the Management pod perform this function, while in Example 2, it is the spine switches.
IGMP snooping	Because vSAN and VXLAN require IGMP queriers to accommodate multicast traffic, IGMP snooping is enabled globally on all six switches since the layer 2/3 boundary is at the spine layer. Also, unregistered multicast flooding is disabled to prevent unintentional multicast traffic from flooding.
IGMP querier for vSAN VLAN	Starting with version 6.6, vSAN no longer uses multicast. Because of this, vSAN 6.6 does not require IGMP querier. However, IGMP querier configuration on the vSAN VLAN is included for the purpose of reverse compatibility. Only IGMP querier, and not multicast routing, is required because vSAN multicast traffic stays within each VLAN. Queriers are set up on all six switches to serve as a back up if switches fail. Since the spine switches have the lowest IP addresses, they will win the querier election.
IGMP querier for VXLAN VLAN	The VXLAN VLAN requires an IGMP querier, which accommodates NSX Hybrid mode and that VVD specifies. In Hybrid mode, VTEPs use multicast for transporting broadcast, unknown unicast, and multicast (BUM) overlay traffic from one host to another within the rack, or local network. For BUM overlay transport across VLANs, VTEPs use unicast. The consequence of this is that, as with vSAN traffic, VXLAN multicast traffic stays in each VLAN, so only IGMP querier is required. If NSX Multicast mode is used, full multicast

	routing is required on all infrastructure switches since multicast is used for overlay transport across racks in addition to within the racks. If NSX Unicast mode is used, multicast configuration is not required. Just like the vSAN VLAN, queriers are set up on all six switches, but lower IPs are set up on the spine switches so they will win the querier election.
Configured VLANs	Several VLANs are configured on the switch in order to accommodate for the VVD 4.1 architecture. These VLANs map to port groups within ESXi. In the Example 2 environment, these VLANs are configured on the leaf and spine switches. They are: • Ext management - provides access to individual ESXi hosts from outside the VVD environment without using a jump box • ESXi host management - enables communication between hosts, and between hosts and vCenter Server • vMotion - enables and segregates vMotion traffic • vSAN - provides transport for and segregates vSAN traffic • VXLAN - provides support for NSX and VXLAN overlay traffic • NFS - enables and segregates NFS, which the VVD requires in order to function as backup storage if the primary storage fails • Replication - provides support and segregation for vSphere Replication • Uplink1 - enables north/south traffic in the data center • Uplink2 - enables north/south traffic in the data center In these networks, BGP peering occurs between the ESGs and the upstream leaf switches.
Virtual Router Redundancy Protocol (VRRP)	VRRP is used for gateway redundancy for all VLANs, with the exception of uplink VLANs and vSAN VLANs. This is a VVD requirement. vSAN VLANs do not use VRRP because the vSANs in this environment do not use routing. Although VRRP is an active/standy first hop redundancy protocol (FHRP), it becomes active/active when it is used between VLT peers. Both VLT peers have the VRRP virtual MAC address in their Forwarding Information Base (FIB) table as a localda entry, and therefore, even the backup VRRP router will forward intercepted frames whose destination MAC address matches the VRRP virtual MAC address. Example 2 shows VRRP configured on the spine switches.
No host-facing port channels	Another design decision is to not use host-facing port channels. To comply with the VVD, route based on physical NIC load on the distributed port groups is used. This does not necessitate port channel configuration on the attached physical switch.
BGP	BGP is the chosen protocol in the SDDC for connecting all network devices, both physical and virtual. BGP allows for flexibility in network design in multi-site and mult-tenancy workloads, and is highly tunable.

Each cluster contains a pair of ESGs which peer with the spine switches in a full-mesh ECMP configuration. Currently, BFD between ESGs and physical switches is not available, so it is disabled. VVD 4.1 specifies that BGP timers should be set to four seconds for keepalive interval and 14 seconds for hold down for the peering relationship between the ESGs and their physical upstream peer switches.

eBGP is chosen because it simplifies configuration. With eBGP, there is no need to configure route reflectors, or to implement next hop self configuration. However, the ASN implementation in this guide diverges from the *Leaf-Spine Deployment and Best Practices Guide* in that spine switches share the same ASN. This prevents ESGs from advertising routes that they learn from the leaf and spine network back into the leaf and spine network. This is important since ESGs currently have an ASN-stripping behavior which causes those re-advertised routes to have higher preference than routes which have the true path originating in the leaf and spine network.

iBGP is used between the spine switches. Because the spine switches have two mVLT connections to both leaf pairs, hashing will cause northbound traffic to be sent to both spine switches. Peer routing enables the spine switches to route on behalf of each other. iBGP ensures that packets are not lost in black holes by synchronizing routing knowledge.

Finally, both default-originate and a route filter are applied to the spine switches in order to advertise default routes and filter all other routes to the ESGs. Since there is only one way out the NSX network, there is no need for the ESGs or DLRs to learn anything about the outside network except for how to get to it via default routes.

4.2 Layer 3 configuration planning

4.2.1 BGP ASN configuration

In VVD 4.1 environments, the leaf and spine design as shown in the *Leaf-Spine Deployment and Best Practices Guide*, must be adapted in order to work properly with the ESGs. In terms of AS numbering in which the layer 2/3 boundary is at the spine layer, this means that the spine switches share one ASN and the leaf switches get none, since they are not routing. The ESGs and other virtual networking devices use iBGP amongst themselves. Valid private, 2-byte ASNs range from 64512 through 65534. Figure 8 shows the ASN assignments used for leaf and spine switches in the BGP examples in this guide.

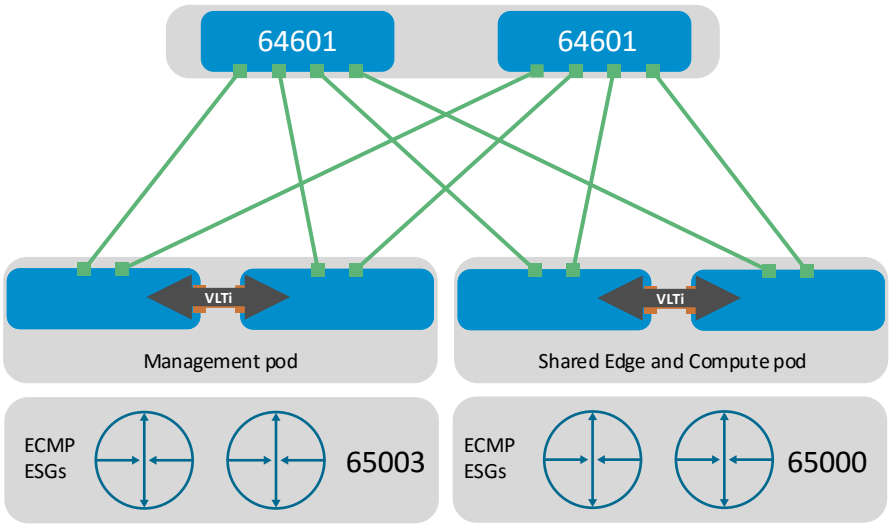

Figure 24 BGP ASN assignments

4.3 NSX configuration overview

The NSX VVD 4.1 environment demonstrated in this document was built using the VMware VVD 4.1 documentation, in particular, *Region A Virtual Infrastructure Implementation*, since the entire VVD 4.1 implementation is not required for the purposes of this document.

This section provides an overview of the virtual network infrastructure generated when one follows the VVD 4.1 documentation, along with how the virtual network infrastructure interconnects with the physical network infracture. Setting up NSX in a VVD 4.1-compliant way is covered in great detail the VMwareVVD 4.1 documentation.

Some of the generated virtual network infrastructure includes four ESGs which connect to the physical network: two for the Management pod and two for the Shared Edge and Compute pod. A diagram of the ESGs and their connections to the spine switches can be seen in Figure 23. ESGs with their respective leaf switches in the Management pod can be seen in Figure 25. Figure 25 also shows specific IP addresses, VLANs, and ASNs that are used. The spines share the same ASN, 64601 and the ESGs share the same ASN, 65003. More detailed connection information between the ESGs and the leaf switches can be found in Table 3.

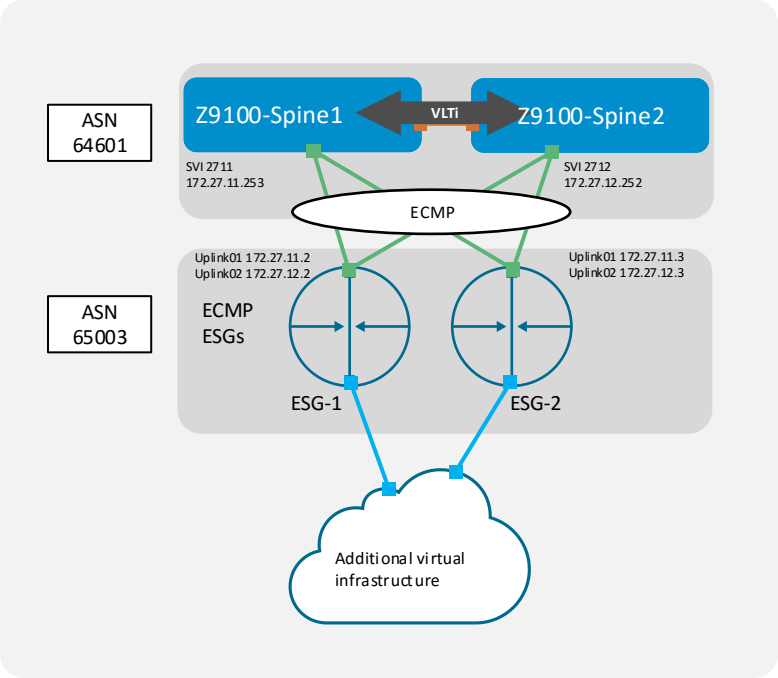

Figure 25 ESGs and spine switches in the Management pod

4.3.1 Uplink configuration

Each ESG has two connections to the physical network: one to each leaf switch using the uplink networks. The ESG-1 Uplink01 interface peers to the Spine-1 2711 VLAN interface, while the ESG-1 Uplink02 interface peers to the Spine-2 2712 VLAN interface. The same is true for the ESG-2 uplinks. The ESG-2 Uplink01 interface peers to the Spine-1 2711 VLAN interface, while the ESG-2 Uplink02 interface peers to the Spine-2 2712 VLAN interface. Matching configuration is done for the Shared Edge and Compute pod. This information can be seen in Table 3.

All of the hosts in each pod share a virtual distributed switch (vDS) for interconnectivity. Each vDS has different port groups for different services, such as vMotion, vSAN, and VXLAN. For north/south connectivity, the Management pod's vDS has two port groups (Uplink01 and Uplink02), which are tagged for 2711 and 2712. These port groups map onto the uplink VLANs on the leaf switches named Uplink1 and Uplink2.

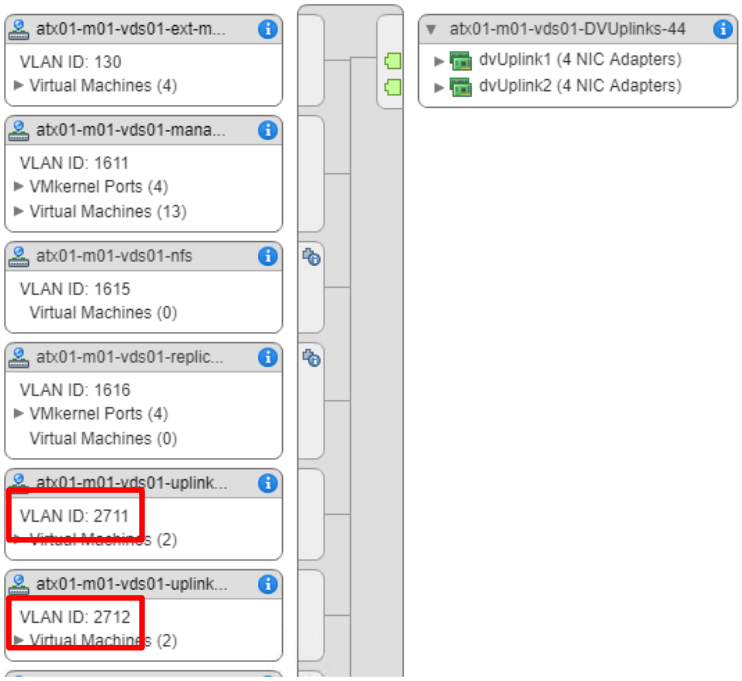

Figure 26 Topology view of the vDS in the Management pod

The Uplink01 port group is configured to use only dvUplink2 and the Uplink02 port group is configured to use only dvUplink1. The dvUplinks are, as the above diagram shows, how the virtual switch connects to the upstream physical switches. Those uplinks use the host vmnics. Each host uses two vmnics for uplinks, so there is a total of eight connections to the leaf switches per pod.

One thing to be mindful about is how uplink port groups are configured. VVD 4.1 specifies that only one dvUplink be used for each uplink port group. Port group Uplink1 is used to peer with Spine 1 and port group Uplink2 is used to peer with Spine 2. dvUplink2 has vmnic0 for all hosts, and so that is the chosen dvUplink for the Uplink1 port group. dvUplink1 has vmnic1 for all hosts, and so that is the chosen dvUplink for the Uplink2 port group. This way, it can be assured that vmnic0 is used to reach Leaf-1 directly, since that is how server/switch connectivity was chosen.

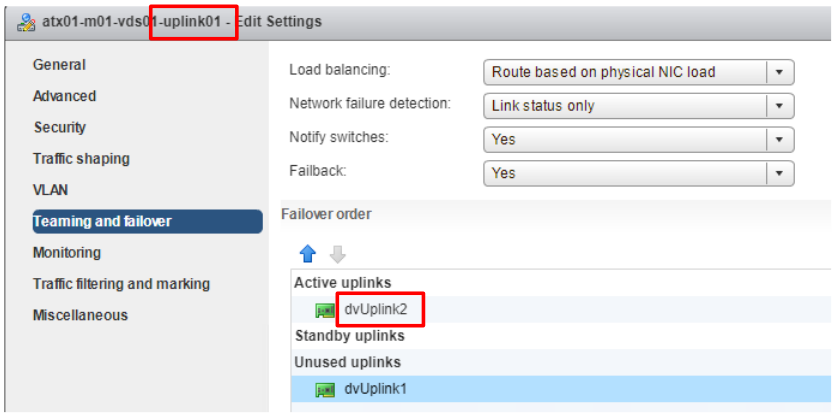

Figure 27 One distributed virtual uplink used per uplink port group

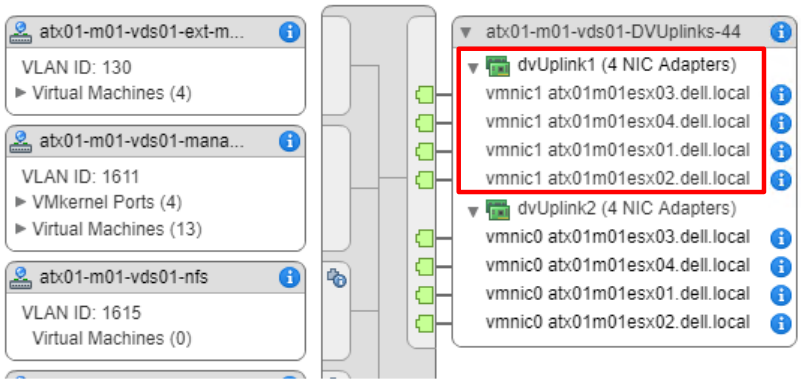

Figure 28 dvUplink1 with vmnic1 for all hosts

4.3.2 BGP neighbor configuration

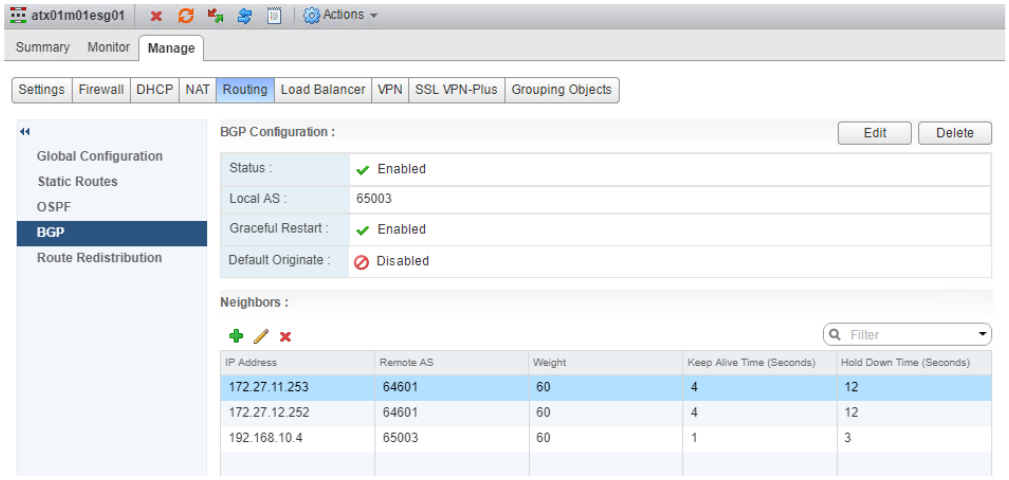

Figure 29 ESG-1 BGP neighbor configuration

As seen in Figure 29, ESG-1 peers with both upstream spine switches with eBGP. Both uplink networks are used. Uplink01 corresponds to VLAN 2711 and Uplink02 corresponds to VLAN 2712 in the Management pod. This is reflected by the second and third octets in the IP addresses which use those networks. 172.27.11.253 refers to Leaf-1's 2711 VLAN interface and 172.27.12.252 refers to Leaf-2's 2712 VLAN interface. 192.168.10.4 refers to the interface of a downstream Universal Distributed Logical Router (UDLR), another virtual network device which is not included in the diagrams. Notice the BGP timers have been changed to four seconds for keepalive interval and 12 for hold down for the peering relationships to the leaf switches. Default timers are 60 seconds for keepalive interval and 180 seconds for hold down. This is to comply with VVD 4.1 specifications.

4.3.3 IP addressing

Table 3 shows the IP addresses used for peering between the ESGs and the physical switches in the Layer 3 environment.

Table 5 Interface and IP configuration for ESGs and leaf switches

Pod	Source switch	Source interface	Source IP	Network	ASN	Destination device	Destination interface	Destination IP	ASN
Management	Spine1	VLAN 2711	. 253	172.27.11.0/24	64601	ESG-1 ESG-2	Uplink01 Uplink01	.2 .3	65003 65003
Management	Spine1	VLAN 2712	. 253	172.27.12.0/24	64601				
Management	Spine 2	VLAN 2711	. 252	172.27.11.0/24	64601				
Management	Spine 2	VLAN 2712	. 252	172.27.12.0/24	64601	ESG-1 ESG-2	Uplink02 Uplink02	.2 .3	65003 65003
Shared Edge and Compute	Spine 1	VLAN 2731	. 253	172.27.31.0/24	64601	ESG-3 ESG-4	Uplink01 Uplink01	.2 .3	65000 65000

Shared Edge and Compute	Spine 1	VLAN 2732	. 253	172.27.32.0/24	64601				
Shared Edge and Compute	Spine 2	VLAN 2731	. 252	172.27.31.0/24	64601				
Shared Edge and Compute	Spine 2	VLAN 2732	. 252	172.27.32.0/24	64601	ESG-3 ESG-4	Uplink02 Uplink02	.2 .3	65000 65000

4.4 S4048-ON leaf switch configuration

The following sections outline the configuration commands issued to the S4048-ON leaf switches to build the topology. The commands detailed below are for L2-Leaf1-S4048 and L2-Leaf2-S4048. The configuration commands for L2-Leaf3-S4048 and L2-Leaf4-S4048 are similar and are provided in the attachments.

Note: On S4048-ON switches, Telnet is enabled and SSH is disabled by default. Both services require the creation of a non-root user account to login. If needed, it is a best practice to use SSH instead of Telnet for security. SSH can optionally be enabled with the command:
`(conf)#ip ssh server enable.`
A user account can be created to access the switch via SSH with the command
`(conf)#username ssh_user sha256-password ssh_password`

First, configure the serial console enable password and disable Telnet.

L2-Leaf1-S4048	L2-Leaf2-S4048

`enable`
`configure`
 	
`enable sha256-password enable_password`
`no ip telnet server enable`
	`enable`
	`configure`
	
	`enable sha256-password enable_password`
	`no ip telnet server enable`

Set the host name, configure the OOB management interface and default gateway. Enable LLDP. Enable RSTP as a precaution. Both leaf switches are configured to enable IGMP snooping globally, along with disallowing unregistered multicast flooding. Both leaf switches are configured with a policy map which instructs the switch to trust DSCP values in IP packets. The designated switches are pointed at for time synchronization.

Note: In this layer 2 topology, the RSTP root bridge is configured at the spine layer.

L2-Leaf1-S4048	L2-Leaf2-S4048

hostname L2-S4048-Leaf1 interface ManagementEthernet 1/1 ip address 100.67.168.32/24 no shutdown management route 0.0.0.0/0 100.67.168.254 protocol lldp advertise management-tlv management- address system-description system-name advertise interface-port-desc protocol spanning-tree rstp no disable ip igmp snooping enable no ip igmp snooping flood policy-map-input TrustDSCPin trust diffserv ntp server 172.16.11.252 ntp server 172.16.11.251	hostname L2-S4048-Leaf2 interface ManagementEthernet 1/1 ip address 100.67.168.31/24 no shutdown management route 0.0.0.0/0 100.67.168.254 protocol lldp advertise management-tlv management- address system-description system-name advertise interface-port-desc protocol spanning-tree rstp no disable ip igmp snooping enable no ip igmp snooping flood policy-map-input TrustDSCPin trust diffserv ntp server 172.16.11.252 ntp server 172.16.11.251

Configure the VLT interconnect between S4048-Leaf1 and S4048-Leaf2. In this configuration, add interfaces fortyGigE 1/53-54 to static port channel 127 for the VLT interconnect. Enable jumbo frames and apply the DSCP policy map to the VLTi's physical interfaces. The backup destination is the management IP address of the VLT peer switch. Enable peer routing.

Note: Dell EMC recommends that the VLTi is configured as a static LAG (without LACP) per the commands shown below.

L2-Leaf1-S4048	**L2-Leaf2-S4048**
interface port-channel 127 description VLTi channel-member fortyGigE 1/53 - 1/54 mtu 9216 no shutdown interface range fortyGigE 1/53 - 1/54 description VLTi mtu 9216 service-policy input TrustDSCPin no shutdown	interface port-channel 127 description VLTi channel-member fortyGigE 1/53 - 1/54 mtu 9216 no shutdown interface range fortyGigE 1/53 - 1/54 description VLTi mtu 9216 service-policy input TrustDSCPin no shutdown

vlt domain 127 peer-link port-channel 127 back-up destination 100.67.168.31 unit-id 0 peer-routing exit	vlt domain 127 peer-link port-channel 127 back-up destination 100.67.168.32 unit-id 1 peer-routing exit

Configure each downstream server-facing interface with Switchport mode. Configure those interfaces as switch ports, as RSTP edge ports, apply the DSCP policy map, and enable the interfaces to accept and process jumbo frames.

L2-Leaf1-S4048	**L2-Leaf2-S4048**
interface tengigabitethernet 1/1 description Server 1 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/2 description Server 2 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/3 description Server 3 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/4 description Server 4 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown	interface tengigabitethernet 1/1 description Server 1 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/2 description Server 2 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/3 description Server 3 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown interface tengigabitethernet 1/4 description Server 4 switchport spanning-tree rstp edge-port service-policy input TrustDSCPin mtu 9216 no shutdown

Configure the upstream port channel going to the Z9100 VLT pair with LACP. Configure jumbo frames and apply the QoS policy map to the physical interfaces.

L2-Leaf1-S4048	L2-Leaf2-S4048
interface fortyGigE 1/49 description Spine-1 port-channel-protocol LACP port-channel 1 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/50 description Spine-2 port-channel-protocol LACP port-channel 1 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface port-channel 1 description Spine-1 switchport vlt-peer-lag port-channel 1 mtu 9216 no shutdown	interface fortyGigE 1/49 description Spine-1 port-channel-protocol LACP port-channel 1 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/50 description Spine-2 port-channel-protocol LACP port-channel 1 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface port-channel 1 description Spine-1 switchport vlt-peer-lag port-channel 1 mtu 9216 no shutdown

Here, VLANs are created and applied to the interfaces. IP addresses are applied to the VLANs that have queriers configured so that they can act as querier backups. IGMP queriers are configured.

L2-Leaf1-S4048	L2-Leaf2-S4048
interface Vlan 1611 description ESXi host management ip address 172.16.11.250/24 tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 1612 description vMotion tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 1613	interface Vlan 1611 description ESXi host management ip address 172.16.11.249/24 tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 1612 description vMotion tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 1613

description vSAN ip address 172.16.13.252/24 tagged te1/1 - 1/4 tagged port-channel 1 ip igmp snooping querier ip igmp version 3 mtu 9216 no shutdown interface Vlan 1614 description VXLAN ip address 172.16.14.252/24 tagged te1/1 - 1/4 tagged port-channel 1 ip igmp snooping querier ip igmp version 3 mtu 9216 no shutdown interface Vlan 1615 description NFS tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 1616 description Replication tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 130 description Ext Management tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 2711 description Uplink1 tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 2712	description vSAN ip address 172.16.13.251/24 tagged te1/1 - 1/4 tagged port-channel 1 ip igmp snooping querier ip igmp version 3 mtu 9216 no shutdown interface Vlan 1614 description VXLAN ip address 172.16.14.251/24 tagged te1/1 - 1/4 tagged port-channel 1 ip igmp snooping querier ip igmp version 3 mtu 9216 no shutdown interface Vlan 1615 description NFS tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 1616 description Replication tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 130 description Ext Management tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 2711 description Uplink1 tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 2712

description Uplink2 tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown	description Uplink2 tagged te1/1 - 1/4 tagged port-channel 1 mtu 9216 no shutdown

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

Finally, exit configuration mode and save the configuration with the end and write commands.

L2-Leaf1-S4048	**L2-Leaf2-S4048**
uplink-state-group 1 description Disable downstream ports in event all uplinks fail downstream TenGigabitEthernet 1/1-1/48 upstream fortyGigE 1/49,1/50 end write	uplink-state-group 1 description Disable downstream ports in event all uplinks fail downstream TenGigabitEthernet 1/1-1/48 upstream fortyGigE 1/49,1/50 end write

4.5 Z9100-ON spine configuration

The following sections outline the configuration commands issued to the Z9100-ON spine switches to build the topology in Figure 23.

Note: On Z9100-ON switches, Telnet is enabled and SSH is disabled by default. Both services require the creation of a non-root user account to login. If needed, it is a best practice to use SSH instead of Telnet for security. SSH can optionally be enabled with the command: (conf)#**ip ssh server enable.**
A user account can be created to access the switch via SSH with the command (conf)#**username *ssh_user* sha256-password *ssh_password***

First, configure the serial console enable password and disable Telnet.

L2-Spine1-Z9100	**L2-Spine2-Z9100**
enable configure enable sha256-password enable_password no ip telnet server enable	enable configure enable sha256-password enable_password no ip telnet server enable

Set the host name, configure the OOB management interface and default gateway. Enable LLDP. Enable RSTP as a precaution. L2-Spine1-Z9100 is configured as the primary RSTP root bridge using the bridge-

priority 0 command. Both spine switches are configured to enable IGMP snooping globally, along with disallowing unregistered multicast flooding. Both spine switches are configured with a policy map which instructs the switch to trust DSCP values in IP packets. An external NTP server is selected for time synchronization.

L2-Spine1-Z9100	L2-Spine2- Z9100
hostname L2-Z9100-Spine1 interface ManagementEthernet 1/1 ip address 100.67.168.36/24 no shutdown management route 0.0.0.0/0 100.67.168.254 protocol lldp advertise management-tlv management- address system-description system-name advertise interface-port-desc stack-unit 1 port 1 portmode single speed 40G no-confirm stack-unit 1 port 2 portmode single speed 40G no-confirm stack-unit 1 port 3 portmode single speed 40G no-confirm stack-unit 1 port 4 portmode single speed 40G no-confirm protocol spanning-tree rstp no disable bridge-priority 0 ip igmp snooping enable no ip igmp snooping flood policy-map-input TrustDSCPin trust diffserv ntp server 172.16.11.240	hostname L2-Z9100-Spine2 interface ManagementEthernet 1/1 ip address 100.67.168.35/24 no shutdown management route 0.0.0.0/0 100.67.168.254 protocol lldp advertise management-tlv management- address system-description system-name advertise interface-port-desc stack-unit 1 port 1 portmode single speed 40G no-confirm stack-unit 1 port 2 portmode single speed 40G no-confirm stack-unit 1 port 3 portmode single speed 40G no-confirm stack-unit 1 port 4 portmode single speed 40G no-confirm protocol spanning-tree rstp no disable bridge-priority 4096 ip igmp snooping enable no ip igmp snooping flood policy-map-input TrustDSCPin trust diffserv ntp server 172.16.11.240

Configure the VLT interconnect between Spine1 and Spine2. In this configuration, add interfaces fortyGigE 1/31-32 to static port channel 127 for the VLT interconnect. The backup destination is the management IP address of the VLT peer switch.

Configure the VLT interconnect between L2-Spine1-Z9100 and L2-Spine2-Z9100. In this configuration, add interfaces hundred 1/31 - 1/32 to static port channel 127 for the VLT interconnect. Enable jumbo frames and

apply the DSCP policy map to the VLTi's physical interfaces. The backup destination is the management IP address of the VLT peer switch. Enable peer routing.

Note: Dell EMC recommends that the VLTi is configured as a static LAG (without LACP) per the commands shown below.

L2-Spine1-Z9100	L2-Spine2-Z9100
interface port-channel 127 description VLTi channel-member hundred 1/31 - 1/32 mtu 9216 no shutdown interface range hundred 1/31 - 1/32 description VLTi mtu 9216 service-policy input TrustDSCPin no shutdown vlt domain 127 peer-link port-channel 127 back-up destination 100.67.168.35 unit-id 0 peer-routing exit	interface port-channel 127 description VLTi channel-member hundred 1/31 - 1/32 mtu 9216 no shutdown interface range hundred 1/31 - 1/32 description VLTi mtu 9216 service-policy input TrustDSCPin no shutdown vlt domain 127 peer-link port-channel 127 back-up destination 100.67.168.36 unit-id 1 peer-routing exit

Interfaces fortyGigE 1/1/1 – 1/4/1 connect to the leaf switches downstream via LACP port channels. Port channel 1 has members fortyGigE 1/1/1 and fortyGigE 1/2/1 and port channel 2 has members fortyGigE 1/3/1 and fortyGigE 1/4/1. The port channels are configured for VLT.

L2-Spine1-Z9100	L2-Spine2-Z9100
interface fortyGigE 1/1/1 description Leaf 1 fo1/49 port-channel-protocol LACP port-channel 1 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/2/1 description Leaf 2 fo1/49 port-channel-protocol LACP port-channel 1 mode active service-policy input TrustDSCPin mtu 9216 no shutdown	interface fortyGigE 1/1/1 description Leaf 1 fo1/50 port-channel-protocol LACP port-channel 1 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/2/1 description Leaf 2 fo1/50 port-channel-protocol LACP port-channel 1 mode active service-policy input TrustDSCPin mtu 9216 no shutdown

interface port-channel 1 portmode hybrid switchport vlt-peer-lag port-channel 1 mtu 9216 no shutdown interface fortyGigE 1/3/1 description Leaf 3 fo1/49 port-channel-protocol LACP port-channel 2 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/4/1 description Leaf 4 fo1/49 port-channel-protocol LACP port-channel 2 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface port-channel 2 portmode hybrid switchport vlt-peer-lag port-channel 2 mtu 9216 no shutdown	interface port-channel 1 portmode hybrid switchport vlt-peer-lag port-channel 1 mtu 9216 no shutdown interface fortyGigE 1/3/1 description Leaf 3 fo1/50 port-channel-protocol LACP port-channel 2 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface fortyGigE 1/4/1 description Leaf 4 fo1/50 port-channel-protocol LACP port-channel 2 mode active service-policy input TrustDSCPin mtu 9216 no shutdown interface port-channel 2 portmode hybrid switchport vlt-peer-lag port-channel 2 mtu 9216 no shutdown

Here, VLANs for both pods are created and applied to the interfaces. The Management pod's VLANs are tagged on port channel 1 and the Shared Edge and Compute pod's VLANs are tagged on port channel 2. Additionally, IP addresses are applied to those VLANs. VRRP groups are created for all VLANs except the vSAN and Uplink VLANs. IGMP queriers are also configured.

L2-Spine1-Z9100	**L2-Spine2-Z9100**
interface Vlan 1611 description ESXi host management ip address 172.16.11.252/24 tagged port-channel 1 vrrp-group 11 description ESXi host management priority 254 virtual-address 172.16.11.253 no shutdown	interface Vlan 1611 description ESXi host management ip address 172.16.11.251/24 tagged port-channel 1 vrrp-group 11 description ESXi host management priority 100 virtual-address 172.16.11.253 no shutdown

```
interface Vlan 1612
description vMotion
ip address 172.16.12.252/24
tagged port-channel 1
vrrp-group 12
description vMotion
priority 254
virtual-address 172.16.12.253
mtu 9216
no shutdown

interface Vlan 1613
description vSAN
ip address 172.16.13.250/24
tagged port-channel 1
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1614
description VXLAN
ip address 172.16.14.250/24
tagged port-channel 1
vrrp-group 14
description VXLAN
priority 254
virtual-address 172.16.14.253
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1615
description NFS
ip address 172.16.15.252/24
tagged port-channel 1
vrrp-group 15
description NFS
priority 254
virtual-address 172.16.15.253
mtu 9216
no shutdown

interface Vlan 1616
description Replication
ip address 172.16.16.252/24
tagged port-channel 1
vrrp-group 16
description Replication
```

```
interface Vlan 1612
description vMotion
ip address 172.16.12.251/24
tagged port-channel 1
vrrp-group 12
description vMotion
priority 100
virtual-address 172.16.12.253
mtu 9216
no shutdown

interface Vlan 1613
description vSAN
ip address 172.16.13.249/24
tagged port-channel 1
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1614
description VXLAN
ip address 172.16.14.249/24
tagged port-channel 1
vrrp-group 14
description VXLAN
priority 100
virtual-address 172.16.14.253
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1615
description NFS
ip address 172.16.15.251/24
vrrp-group 15
description NFS
priority 100
virtual-address 172.16.15.253
tagged port-channel 1
mtu 9216
no shutdown

interface Vlan 1616
description Replication
ip address 172.16.16.251/24
vrrp-group 16
description Replication
priority 100
```

priority 254 virtual-address 172.16.16.253 mtu 9216 no shutdown interface Vlan 130 description Ext Management ip address 10.158.130.252/24 tagged port-channel 1 vrrp-group 130 description Ext Management priority 254 virtual-address 10.158.130.253 no shutdown interface Vlan 2711 description Uplink1 ip address 172.27.11.253/24 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 2712 description Uplink2 ip address 172.27.12.253/24 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 1631 description ESXi host management ip address 172.16.31.252/24 tagged port-channel 2 vrrp-group 31 description ESXi host management priority 254 virtual-address 172.16.31.253 no shutdown interface Vlan 1632 description vMotion ip address 172.16.32.252/24 tagged port-channel 2 vrrp-group 32 description vMotion priority 254 virtual-address 172.16.32.253 mtu 9216	virtual-address 172.16.16.253 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 130 description Ext Management ip address 10.158.130.251/24 tagged port-channel 1 vrrp-group 130 description Ext Management priority 100 virtual-address 10.158.130.253 no shutdown interface Vlan 2711 description Uplink1 ip address 172.27.11.252/24 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 2712 description Uplink2 ip address 172.27.12.252/24 tagged port-channel 1 mtu 9216 no shutdown interface Vlan 1631 description ESXi host management ip address 172.16.31.251/24 tagged port-channel 2 vrrp-group 31 description ESXi host management priority 100 virtual-address 172.16.31.253 no shutdown interface Vlan 1632 description vMotion ip address 172.16.32.251/24 tagged port-channel 2 vrrp-group 32 description vMotion priority 100 virtual-address 172.16.32.253 mtu 9216

```
no shutdown

interface Vlan 1633
description vSAN
ip address 172.16.33.250/24
tagged port-channel 2
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1634
description VXLAN
ip address 172.16.34.250/24
tagged port-channel 2
vrrp-group 34
description VXLAN
priority 254
virtual-address 172.16.34.253
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1635
description NFS
ip address 172.16.35.252/24
tagged port-channel 2
vrrp-group 35
description NFS
priority 254
virtual-address 172.16.35.253
mtu 9216
no shutdown

interface Vlan 1636
description Replication
ip address 172.16.36.252/24
tagged port-channel 2
vrrp-group 36
description Replication
priority 254
virtual-address 172.16.36.253
mtu 9216
no shutdown

interface Vlan 150
description Ext Management
ip address 10.158.150.252/24
tagged port-channel 2
vrrp-group 150
```

```
no shutdown

interface Vlan 1633
description vSAN
ip address 172.16.33.249/24
tagged port-channel 2
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1634
description VXLAN
ip address 172.16.34.249/24
tagged port-channel 2
vrrp-group 34
description VXLAN
priority 100
virtual-address 172.16.34.253
ip igmp snooping querier
ip igmp version 3
mtu 9216
no shutdown

interface Vlan 1635
description NFS
ip address 172.16.35.251/24
tagged port-channel 2
vrrp-group 35
description NFS
priority 100
virtual-address 172.16.35.253
mtu 9216
no shutdown

interface Vlan 1636
description Replication
ip address 172.16.36.251/24
tagged port-channel 2
vrrp-group 36
description Replication
priority 100
virtual-address 172.16.36.253
mtu 9216
no shutdown

interface Vlan 150
description Ext Management
ip address 10.158.150.251/24
tagged port-channel 2
vrrp-group 150
```

description Ext Management priority 254 virtual-address 10.158.150.253 no shutdown interface Vlan 2731 description Uplink1 ip address 172.27.31.253/24 tagged port-channel 2 mtu 9216 no shutdown interface Vlan 2732 description Uplink2 ip address 172.27.32.253/24 tagged port-channel 2 mtu 9216 no shutdown	description Ext Management priority 100 virtual-address 10.158.150.253 no shutdown interface Vlan 2731 description Uplink1 ip address 172.27.31.252/24 tagged port-channel 2 mtu 9216 no shutdown interface Vlan 2732 description Uplink2 ip address 172.27.32.252/24 tagged port-channel 2 mtu 9216 no shutdown

Configure a route map and IP prefix list to redistribute all loopback addresses and leaf networks via BGP. Configure a route map and IP prefix list to filter routes to the ESGs.

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 all 172.16 networks configured on the leaf switches.

The command `seq 20 permit 172.16.0.0/16 ge 24` includes all 172.16 networks configured on all leaf switches.

The command `seq 5 permit 0.0.0.0/0` refers to no network in specific and is used by a route map to permit only a default route to be inlcuded in routing updates to the ESGs.

L2-Spine1-Z9100	**L2-Spine2-Z9100**
route-map spine-leaf permit 10 match ip address spine-leaf ip prefix-list spine-leaf description Redistribute loopback and leaf networks seq 10 permit 10.0.0.0/8 ge 24 seq 20 permit 172.16.0.0/16 ge 24 route-map default permit 10 match ip address default	route-map spine-leaf permit 10 match ip address spine-leaf ip prefix-list spine-leaf description Redistribute loopback and leaf networks seq 10 permit 10.0.0.0/8 ge 24 seq 20 permit 172.16.0.0/16 ge 24 route-map default permit 10 match ip address default

ip prefix-list default seq 5 permit 0.0.0.0/0	ip prefix-list default seq 5 permit 0.0.0.0/0

Include the interfaces to each leaf pair in an ECMP group. Enable link bundle monitoring to report when traffic is unevenly distributed across multiple links.

Note: ECMP is not enabled until BGP is configured.

L2-Spine1-Z9100	L2-Spine2-Z9100
ecmp-group 1 interface fortyGigE 1/1/1 interface fortyGigE 1/2/1 link-bundle-monitor enable ecmp-group 2 interface fortyGigE 1/3/1 interface fortyGigE 1/4/1 link-bundle-monitor enable	ecmp-group 1 interface fortyGigE 1/1/1 interface fortyGigE 1/2/1 link-bundle-monitor enable ecmp-group 2 interface fortyGigE 1/3/1 interface fortyGigE 1/4/1 link-bundle-monitor enable

4.5.1 Z9100-ON BGP configuration

Use the following commands commands to configure BGP.

First, enable BGP with the `router bgp ASN` command. The ASN is from Figure 11.

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. This number should be equal to or greater than the number of spine switches, up to 64.

Five BGP neighbors are configured: four for ESGs and one for each spine switch. All are configured for fast fallover. The ESG neighbors have their timers changed to four seconds for keepalive interval and 12 seconds for hold down.

Finally, exit configuration mode and save the configuration with the `end` and `write` commands.

L2-Spine1-Z9100	L2-Spine2-Z9100
router bgp 64601 bgp bestpath as-path multipath-relax maximum-paths ebgp 2 redistribute connected route-map spine-leaf bgp graceful-restart neighbor spine-leaf peer-group	router bgp 64601 bgp bestpath as-path multipath-relax maximum-paths ebgp 2 redistribute connected route-map spine-leaf bgp graceful-restart neighbor spine-leaf peer-group

```
neighbor spine-leaf fall-over
neighbor spine-leaf advertisement-
interval 1
neighbor spine-leaf password 7
7fa786736049da17
neighbor spine-leaf no shutdown
neighbor spine-leaf bfd
neighbor 172.27.11.252 remote-as 64601
neighbor 172.27.11.252 peer-group
spine-leaf
neighbor 172.27.11.252 no shutdown
neighbor esg peer-group
neighbor esg default-originate
neighbor esg route-map default out
neighbor esg fall-over
neighbor esg no shutdown
neighbor esg password Dell!234
neighbor esg timers 4 12
neighbor 172.27.11.2 remote-as 65003
neighbor 172.27.11.2 peer-group esg
neighbor 172.27.11.2 no shutdown
neighbor 172.27.11.3 remote-as 65003
neighbor 172.27.11.3 peer-group esg
neighbor 172.27.11.3 no shutdown
neighbor 172.27.31.3 remote-as 65000
neighbor 172.27.31.3 peer-group esg
neighbor 172.27.31.3 no shutdown
neighbor 172.27.31.2 remote-as 65000
neighbor 172.27.31.2 peer-group esg
neighbor 172.27.31.2 no shutdown

end
write
```

```
neighbor spine-leaf fall-over
neighbor spine-leaf advertisement-
interval 1
neighbor spine-leaf password 7
7fa786736049da17
neighbor spine-leaf no shutdown
neighbor spine-leaf bfd
neighbor 172.27.11.253 remote-as 64601
neighbor 172.27.11.253 peer-group
spine-leaf
neighbor 172.27.11.253 no shutdown
neighbor esg peer-group
neighbor esg default-originate
neighbor esg route-map default out
neighbor esg fall-over
neighbor esg no shutdown
neighbor esg password Dell!234
neighbor esg timers 4 12
neighbor 172.27.12.2 remote-as 65003
neighbor 172.27.12.2 peer-group esg
neighbor 172.27.12.2 no shutdown
neighbor 172.27.12.3 remote-as 65003
neighbor 172.27.12.3 peer-group esg
neighbor 172.27.12.3 no shutdown
neighbor 172.27.32.3 remote-as 65000
neighbor 172.27.32.3 peer-group esg
neighbor 172.27.32.3 no shutdown
neighbor 172.27.32.2 remote-as 65000
neighbor 172.27.32.2 peer-group esg
neighbor 172.27.32.2 no shutdown

end
write
```

4.6 Example 2 validation

In addition to sending traffic between hosts, the configuration shown in Figure 7 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 switch and one leaf switch. Command output on other switches is similar.

4.6.1 show ip bgp summary

The `show ip route bgp` command shows the status of all BGP connections. Each spine switch has four neighbors (the four ESGs). The `show ip bgp summary` command also confirms BFD is enabled on the 6th line of output.

```
L2-Z9100-Spine1#show ip bgp summary
BGP router identifier 172.27.32.253, local AS number 64601
BGP local RIB : Routes to be Added 0, Replaced 0, Withdrawn 0
28 network entrie(s) using 2128 bytes of memory
42 paths using 4536 bytes of memory
BGP-RIB over all using 4578 bytes of memory
8 BGP path attribute entrie(s) using 1204 bytes of memory
6 BGP AS-PATH entrie(s) using 60 bytes of memory
4 neighbor(s) using 32768 bytes of memory

Neighbor        AS            MsgRcvd  MsgSent     TblVer  InQ  OutQ Up/Down
State/Pfx
172.27.11.2     65003            8598     8122          0    0     0 03:11:17 7
172.27.11.3     65003            8558     8082          0    0     0 03:10:01 7
172.27.31.2     65000            8601     8540          0    0     0 03:12:01 7
172.27.31.3     65000            8544     8492          0    0     0 03:09:46 7
```

4.6.2 show ip route

The `show ip route` shows the contents of the routing table. Entries with multiple paths shown are used with ECMP. Leaf and spine networks for both Management and Shared Edge and Compute pods can be seen in this example (1611, 1612, 1613, etc.), as well as networks in the virtual infrastructure (192.168.10, 192.168.11, 192.168.100, 192.168.101 networks).

```
L2-Z9100-Spine1#show ip route

Codes: C - connected, S - static, R - RIP,
       B - BGP, IN - internal BGP, EX - external BGP,LO - Locally Originated,
       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, i - IS-IS, L1 - IS-IS level-1,
       L2 - IS-IS level-2, IA - IS-IS inter area, * - candidate default,
       > - non-active route, + - summary route

Gateway of last resort is not set

       Destination        Gateway                      Dist/Metric Last Change
       -----------        -------                      ----------- -----------
```

```
B EX 10.100.0.0/24      via 172.27.31.2                    20/0    04:04:33
                        via 172.27.31.3
B EX 10.100.1.0/24      via 172.27.31.2                    20/0    04:04:33
                        via 172.27.31.3
C    10.158.130.0/24    Direct, Vl 130                      0/0    14:59:49
C    10.158.150.0/24    Direct, Vl 150                      0/0    14:59:49
C    172.16.11.0/24     Direct, Vl 1611                     0/0    14:59:49
C    172.16.12.0/24     Direct, Vl 1612                     0/0    14:59:49
C    172.16.13.0/24     Direct, Vl 1613                     0/0    03:55:16
C    172.16.14.0/24     Direct, Vl 1614                     0/0    03:55:16
C    172.16.15.0/24     Direct, Vl 1615                     0/0    14:59:49
C    172.16.16.0/24     Direct, Vl 1616                     0/0    14:59:49
C    172.16.31.0/24     Direct, Vl 1631                     0/0    14:59:49
C    172.16.32.0/24     Direct, Vl 1632                     0/0    14:59:49
C    172.16.33.0/24     Direct, Vl 1633                     0/0    03:55:15
C    172.16.34.0/24     Direct, Vl 1634                     0/0    03:55:15
C    172.16.35.0/24     Direct, Vl 1635                     0/0    14:59:49
C    172.16.36.0/24     Direct, Vl 1636                     0/0    14:59:49
C    172.27.11.0/24     Direct, Vl 2711                     0/0    14:59:49
C    172.27.12.0/24     Direct, Vl 2712                     0/0    14:59:49
C    172.27.31.0/24     Direct, Vl 2731                     0/0    14:59:49
C    172.27.32.0/24     Direct, Vl 2732                     0/0    14:59:49
B EX 192.168.10.0/24    via 172.27.11.2                    20/0    04:04:47
                        via 172.27.11.3
B EX 192.168.11.0/24    via 172.27.11.2                    20/0    04:04:47
                        via 172.27.11.3
B EX 192.168.31.0/24    via 172.27.11.2                    20/0    04:04:47
                        via 172.27.11.3
B EX 192.168.32.0/24    via 172.27.11.2                    20/0    04:04:47
                        via 172.27.11.3
B EX 192.168.100.0/24   via 172.27.31.2                    20/0    04:04:33
                        via 172.27.31.3
B EX 192.168.101.0/24   via 172.27.31.2                    20/0    04:04:33
                        via 172.27.31.3
B EX 192.168.102.0/24   via 172.27.11.2                    20/0    04:04:47
                        via 172.27.11.3
B EX 192.168.103.0/24   via 172.27.31.2                    20/0    04:04:33
                        via 172.27.31.3
```

Z9100-Spine1#S4048-Leaf-1 has two paths to all other leaf switches and two paths to the virtual infrastructure networks. There is one path through each spine switch. If all paths do not appear, make sure

the maximum-paths statement in the BGP configuration is equal to or greater than the number of spine switches in the topology.

```
S4048-Leaf-1#show ip route bgp
       Destination        Gateway                      Dist/Metric Last Change
       -----------        -------                      ----------- -----------
  B EX 10.0.1.1/32        via 192.168.1.0                    20/0        4d6h
  B EX 10.0.1.2/32        via 192.168.2.0                    20/0        4d6h
  B EX 10.0.2.3/32        via 192.168.1.0                    20/0        4d5h
                          via 192.168.2.0
  B EX 10.0.2.4/32        via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 10.100.0.0/24      via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 10.100.1.0/24      via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 10.158.150.0/24    via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 172.16.31.0/24     via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 172.16.32.0/24     via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 172.16.33.0/24     via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 172.16.34.0/24     via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 172.16.35.0/24     via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 172.16.36.0/24     via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 172.27.31.0/24     via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 172.27.32.0/24     via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 192.168.10.0/24    via 172.27.11.2                    20/0        4d6h
                          via 172.27.11.3
  B EX 192.168.11.0/24    via 172.27.11.2                    20/0        4d6h
                          via 172.27.11.3
  B EX 192.168.31.0/24    via 172.27.11.2                    20/0        4d6h
                          via 172.27.11.3
  B EX 192.168.32.0/24    via 172.27.11.2                    20/0        4d6h
                          via 172.27.11.3
  B EX 192.168.100.0/24   via 192.168.1.0                    20/0        4d6h
                          via 192.168.2.0
  B EX 192.168.101.0/24   via 192.168.1.0                    20/0        4d6h
```

```
via 192.168.2.0
```

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

4.6.3 show vrrp brief

The `show vrrp brief` command validates the status of the configured VRRP groups.

```
L2-Z9100-Spine1#show vrrp brief

Interface   Group    Pri Pre State  Master addr     Virtual addr(s)
Description
-------------------------------------------------------------------------------
Vl 130      IPv4 130 254 Y   Master 10.158.130.252  10.158.130.253
Ext Management
Vl 150      IPv4 150 254 Y   Master 10.158.150.252  10.158.150.253
Ext Management
Vl 1611     IPv4 11  254 Y   Master 172.16.11.252   172.16.11.253
ESXi host management
Vl 1612     IPv4 12  254 Y   Master 172.16.12.252   172.16.12.253
vMotion
Vl 1614     IPv4 14  254 Y   Master 172.16.14.250   172.16.14.253
VXLAN
Vl 1615     IPv4 15  254 Y   Master 172.16.15.252   172.16.15.253
NFS
Vl 1616     IPv4 16  254 Y   Master 172.16.16.252   172.16.16.253
Replication
Vl 1631     IPv4 31  254 Y   Master 172.16.31.252   172.16.31.253
ESXi host management
Vl 1632     IPv4 32  254 Y   Master 172.16.32.252   172.16.32.253
vMotion
Vl 1634     IPv4 34  254 Y   Master 172.16.34.250   172.16.34.253
VXLAN
Vl 1635     IPv4 35  254 Y   Master 172.16.35.252   172.16.35.253
NFS
Vl 1636     IPv4 36  254 Y   Master 172.16.36.252   172.16.36.253
Replication
```

4.6.4 show vlt detail

The `show vlt detail` command validates VLT LAG status on leaf switches in this topology. This command shows the status and active VLANs of all VLT LAGs (Port channel 1 in this example). The local and peer status must both be up.

```
L2-Z9100-Spine1#show vlt detail
Local LAG Id  Peer LAG Id  Local Status  Peer Status  Active VLANs
------------  -----------  ------------  -----------  -------------
1             1            UP            UP            1, 130, 150, 1611-1616,
                                                       1631-1636, 2711-2712,
                                                       2731-2732
2             2            UP            UP            1, 130, 150, 1611-1616,
                                                       1631-1636, 2711-2712,
                                                       2731-2732
S4048-Leaf1#show vlt detail
Local LAG Id  Peer LAG Id  Local Status  Peer Status  Active VLANs
------------  -----------  ------------  -----------  -------------
1             1            UP            UP            130, 1611-1616, 2711-2712
```

4.6.5 show uplink-state-group

The `show uplink-state-group` command validates the UFD status on leaf switches in this topology. Status: Enabled, Up indicates UFD is enabled and no interfaces are currently disabled by UFD.

```
L2-S4048-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
```

Note: When an interface has been disabled by UFD, the show interfaces *interface* command for affected interfaces indicates it is error-disabled as follows:
S4048-Leaf-1#**show interfaces te 1/4**
TenGigabitEthernet 1/4 is up, line protocol is down(error-disabled[UFD])
-- Output truncated –-

4.6.6 show spanning-tree rstp brief

The `show spanning-tree rstp brief` 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 (S4048-Leaf-1 in this example) is the root bridge and sever-facing interface are edge ports.

```
L2-Z9100-Spine1#show spanning-tree rstp brief
Executing IEEE compatible Spanning Tree Protocol
Root ID    Priority 0, Address 4c76.25e8.d640
Root Bridge hello time 2, max age 20, forward delay 15
Bridge ID    Priority 0, Address 4c76.25e8.d640
We are the root
Configured hello time 2, max age 20, forward delay 15

Interface                                                  Designated
 Name      PortID   Prio Cost    Sts         Cost        Bridge ID        PortID
---------- -------- ---- ------- ----------- ------- -------------------- ------
--
Po 1       128.2    128  188     FWD(vlt)    0       0     4c76.25e8.d640 128.2
Po 2       128.3    128  188     FWD(vlt)    0       0     4c76.25e8.d640 128.3
Po 127     128.128  128  180     FWD(vltI)   0       0     4c76.25e8.d640
128.128

Interface
 Name      Role   PortID   Prio Cost    Sts         Cost    Link-type Edge
---------- ------ -------- ---- ------- ----------- ------- --------- ----
Po 1       Desg   128.2    128  188     FWD         0       (vlt) P2P No
Po 2       Desg   128.3    128  188     FWD         0       (vlt) P2P No
Po 127     Desg   128.128  128  180     FWD         0       (vltI)P2P No
```

4.6.7 Validate VTEP functionality

NSX has an in-built tool for validating VTEP functionality.

From the vSphere web client:

1. Go to **Home > Networking & Security > Logical Switches**.
2. In the center pane, double click the Universal Transit Network logical switch.
3. Select source and destination hosts.
4. Click **Start Test**.

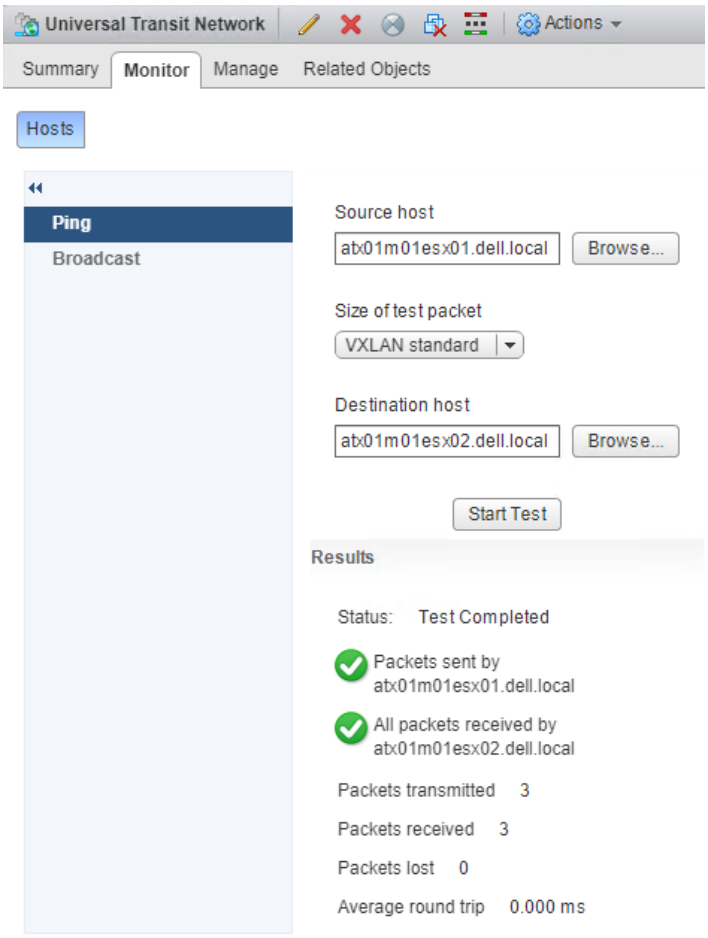

Figure 30 VTEP test

4.6.8 Validate VM-to-VM layer 2 VXLAN functionality

In this example, layer 2 functionality is tested between two VMs on different hosts.

Two Linux VMs are first migrated to two different hosts: atx01m01esx01 for lubuntu1, and atx01m01esx02 for lubuntu2 in the Management pod.

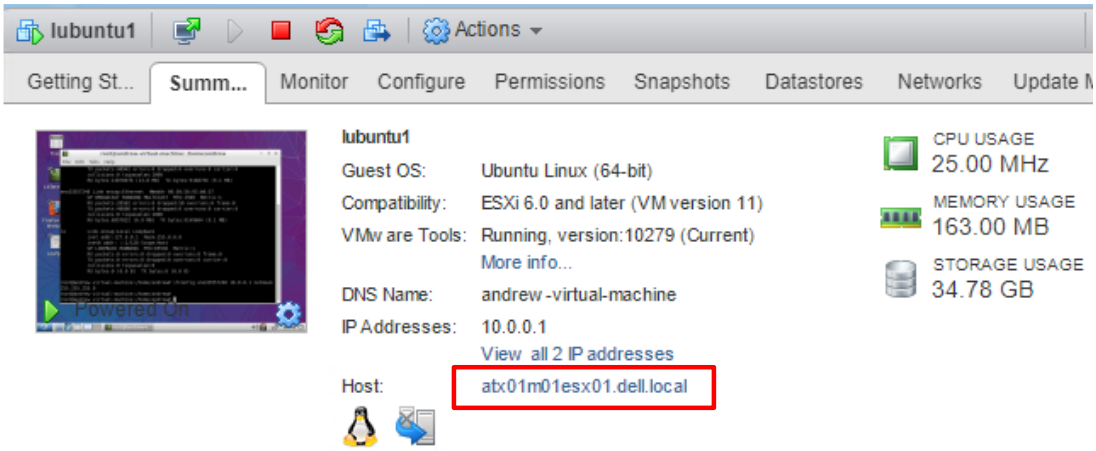

Figure 31 Lubuntu1 migrated to atx01m01esx01

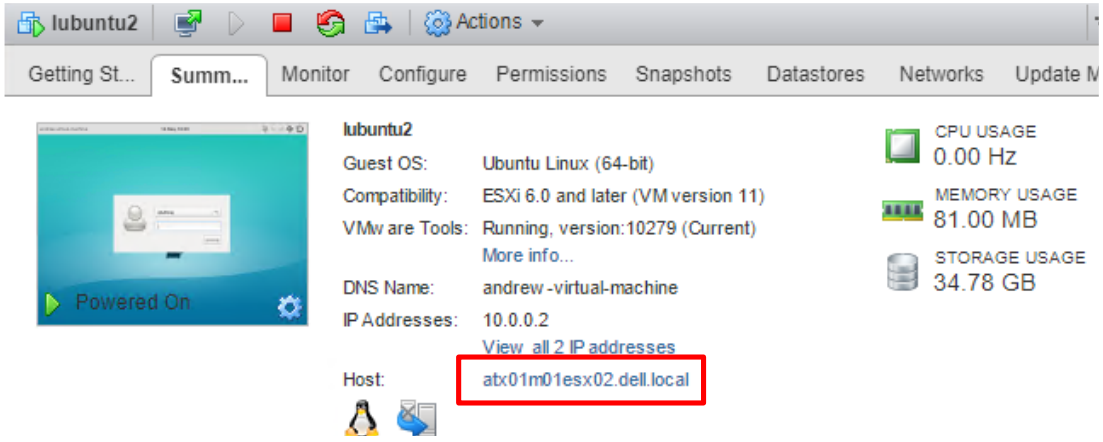

Figure 32 Lubuntu2 migrated to atx01m01esx02

Each VM is given an extra NIC and that NIC is attached to the Universal Transit Network universal logical switch (port group is **vwx-dvs-44-universalwire-1-sid-3000 Universal Transit Network**).

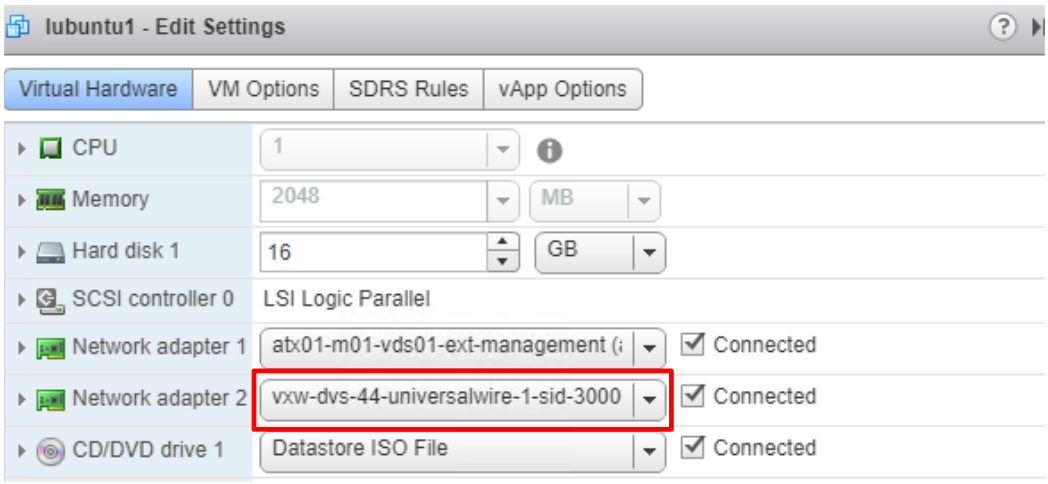

Figure 33 Lubuntu1 is given an extra NIC

Once both VMs are connected to the Universal Transit Network universal logical switch, they are given IP addresses in the same subnet. Lubuntu1 is given the IP 10.0.0.1 and lubuntu2 is given the IP 10.0.0.2. Then a ping test is issued.

```
          collisions:0 txqueuelen:1000
          RX bytes:13884014 (13.8 MB)  TX bytes:9179460 (9.1 MB)

eno33557248 Link encap:Ethernet  HWaddr 00:50:56:93:b0:57
          inet addr:10.0.0.1  Bcast:10.0.0.255  Mask:255.255.255.0
          inet6 addr: fe80::250:56ff:fe93:b057/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:20508 errors:0 dropped:50 overruns:0 frame:0
          TX packets:48900 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:6958296 (6.9 MB)  TX bytes:9150672 (9.1 MB)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          inet6 addr: ::1/128 Scope:Host
          UP LOOPBACK RUNNING  MTU:65536  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:0 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:0 (0.0 B)  TX bytes:0 (0.0 B)

root@andrew-virtual-machine:/home/andrew# ping 10.0.0.2
PING 10.0.0.2 (10.0.0.2) 56(84) bytes of data.
64 bytes from 10.0.0.2: icmp_seq=1 ttl=64 time=0.343 ms
64 bytes from 10.0.0.2: icmp_seq=2 ttl=64 time=0.247 ms
^C
--- 10.0.0.2 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 999ms
rtt min/avg/max/mdev = 0.247/0.295/0.343/0.048 ms
root@andrew-virtual-machine:/home/andrew#
```

Figure 34 Successful ping test between lubuntu1 and lubuntu2

Packet captures of the above network activity reveal a few things about how VTEPs encapsulate traffic when using Hybrid or Multicast modes. The first picture highlights an ARP packet. Since ARP is BUM traffic (it is broadcast), it is encapsulated with multicast. Once lubuntu1 resolves lubuntu2's MAC address, it proceeds to send unicast ping packets to lubuntu2, which are not BUM traffic. Since these packets are not BUM, they are encapsulated with unicast, in which the source and destination IP addresses are the IP addresses of the VTEP kernel ports for the hosts which contain the VMs. These are the tunnel endpoints.

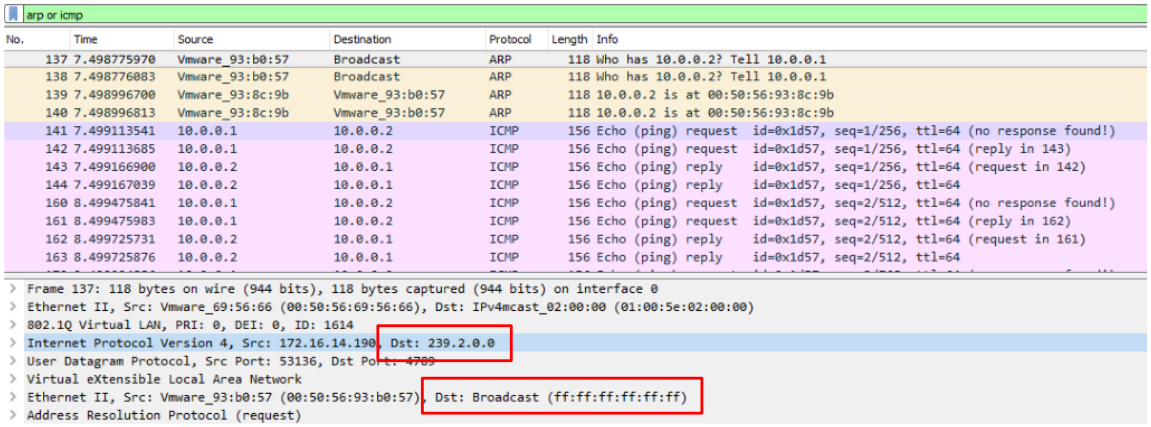

Figure 35 Selected ARP packet is BUM traffic encapsulated with multicast

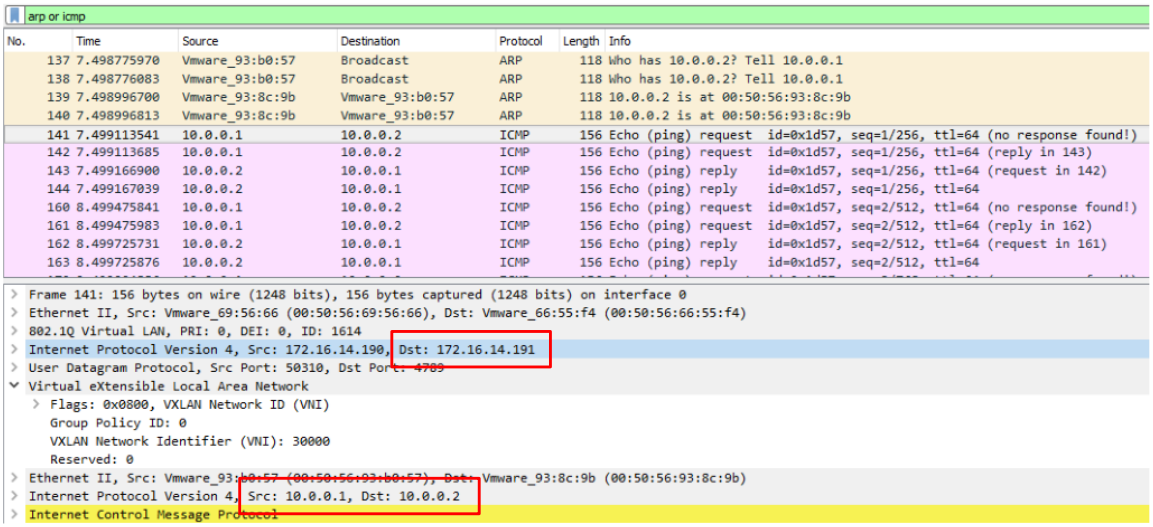

Figure 36 Selected ping packet is not BUM traffic, so it is encapsulated with unicast

A Dell EMC validated hardware and components

The following tables present the hardware and components used to configure and validate the example configurations in this guide.

A.1 Switches

Qty	Item	Firmware Version
2	Z9100-ON Spine switch	DNOS 9.11(2.4)
4	S4048-ON Leaf switch	DNOS 9.11(2.4)

A.2 PowerEdge R640 servers

This guide uses four PowerEdge R640 servers in the Management pod.

Qty per server	Item	Firmware Version
2	Intel(R) Xeon(R) Gold 6126 CPU @ 2.60GHz, 12 cores	-
64	GB RAM	-
8	447.13 GB SAS SSD	-
1	Dell HBA330 Mini (Embedded)	13.17.03.00
2	16 GB Internal SD Cards	-
1	Intel(R) 2P X710/2P I350 rNDC	18.0.16
-	R640 BIOS	1.1.7
-	iDRAC with Lifecycle Controller	3.00.00.00

A.3 PowerEdge R730xd servers

This guide uses four PowerEdge R730xd servers in the Shared Edge and Compute pod.

Qty per server	Item	Firmware Version
2	Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz, 12 cores	-
128	GB RAM	-
8	500 GB SAS SSD	-
1	PERC H730 Mini	25.5.2.0001
2	16 GB Internal SD Cards	-
1	Intel(R) 2P X520	18.0.16
-	R730xd BIOS	2.4.3
-	iDRAC with Lifecycle Controller	2.41.40.40

B Dell EMC validated software and required licenses

The following information provides a listing of the versions of the software components used to validate the example configurations in this guide, and the licenses required for the example configurations in this this guide.

B.1 Software

Item	Version
VMware ESXi	6.5.0 – build 5969303 - Update 1 - Dell EMC customized image version A00
VMware vCenter Server Appliance	6.5.0 - build 5973321
vSphere Web Client	6.5.0.1 - build 5973321
VMware NSX Manager	6.3.4 - build 6845891

B.2 Licenses

The VMware vCenter Server is licensed by instance. The remaining licenses are allocated based on the number of CPU sockets in the participating hosts.

Required licenses for the topology built in this guide are as follows:

- VMware vSphere 6 Enterprise Plus – 16 CPU sockets
- vCenter 6 Server Standard – 2 instances
- vSAN Advanced – 16 CPU sockets
- NSX for vSphere Enterprise - 16 CPU sockets

VMware product licenses can be centrally managed by going to the vSphere web client **Home** page and clicking **Licensing** in the center pane.

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 S4048-ON

Manuals and documentation for Dell EMC Networking Z9100-ON

D Support and Feedback

Contacting Technical Support

Support Contact Information

Web: http://support.dell.com/

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.