VCF on VxRail multirack deployment using BGP EVPN

Abstract

This document provides step-by-step deployment instructions for Dell EMC OS10 L2 VXLAN tunnel using BGP EVPN. This guide provides the foundation for multirack VxRail host discovery and deployment. Also, the VMware Cloud Foundation (VCF) is deployed and provides a VMware NSX environment to enable a Software Defined Data Center (SDDC).

June 2019

Revisions

Date	Description
June 2019	Initial publication

Table of contents

1 Introduction

Our vision at Dell EMC is to be the essential infrastructure company from the edge, to the core, and to the cloud. Dell EMC Networking ensures modernization for today's applications and for the emerging cloud-native world. Dell EMC is committed to disrupting the fundamental economics of the market with an open strategy that gives you the freedom of choice for networking operating systems and top-tier merchant silicon. The Dell EMC strategy enables business transformations that maximize the benefits of collaborative software and standards-based hardware, including lowered costs, flexibility, freedom, and security. Dell EMC provides further customer enablement through validated deployment guides which demonstrate these benefits while maintaining a high standard of quality, consistency, and support.

At the physical layer of a Software Defined Data Center (SDDC), the switching fabric can be provided by either Layer 2 or Layer 3 transport services. A leaf-spine architecture using Layer 3 IP supports a scalable data network. In a Layer 3 network fabric, the physical network configuration terminates Layer 2 networks at the leaf switch pair at the top of each rack. However, VxRail node discovery and NSX Controller instances and other virtual machines rely on VLAN-backed Layer 2 networks. Discovery or virtual machine migration cannot be completed because the IP subnet is available only in the rack where the virtual machine resides. One approach to solve this challenge is to implement Border Gateway Protocol (BGP) Ethernet VPN (EVPN) to create control plane backed tunnels between the separate IP subnets creating Layer 2 networks that span multiple racks or locations.

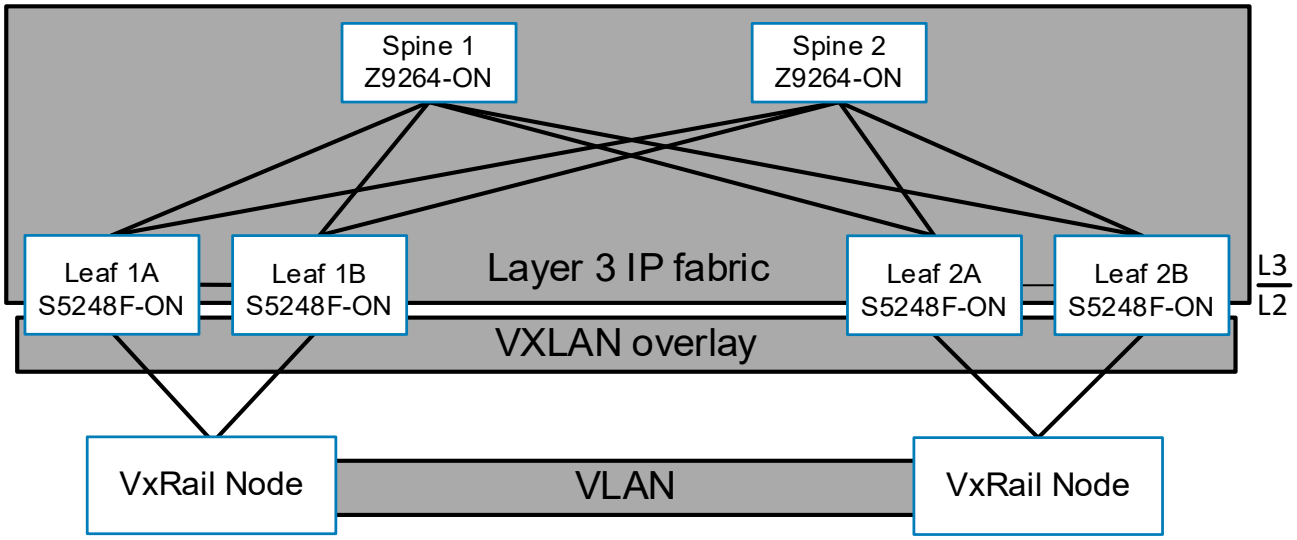


Figure 1 Illustration of a stretched layer 2 segment between VxRail nodes

smbiosDump | less

1.1 Objective

This guide is a supplement to the VMware Cloud Foundation on VxRail (VCF on VxRail) and VMware Validated Design on VxRail 5.0 (VVD on VxRail 5.0) documentation. It aims to provide design guidance to create a Layer 3 IP underlay with Layer 2 network overlays that enable the core services of the SDDC to operate across separate subnets such as ESXi management, vMotion, and vSAN.

This example uses a typical leaf-spine topology with BGP EVPN as a control plane mechanism orchestrating Virtual Extensible LANs (VXLANs) between the leaf switches. The individual switch configuration shows how to set up an end-to-end IP underlay network using BGP with an end-to-end virtual network using BGP EVPN L2 VXLAN configuration. The configuration provides a network virtual overlay (NVO) enabling east to west traffic flows for the core services referenced.

Once the network components are in place, the initial components of a Standard SDDC are deployed using VCF on VxRail. The deployment consists of a VVD management domain running on four Dell EMC VxRail E-series nodes. These four nodes are equally distributed in two racks in Availability Zone 1 (AZ1) in the primary data center (Region A). Using the network overlay for host discovery, both VxRail initialization tools and VMware Cloud Builder are used to complete the installation of the software components of the SDDC.

To accommodate north and southbound traffic into the SDDC, a pair of Equal-cost multi-path (ECMP)-enabled VMware NSX Edge Services Gateways (ESGs) are deployed. Using VM/host affinity rules, the ESGs are restricted to a single rack, enabling it as the edge rack, and BGP is configured between these ESGs and the leaf switches in the designated edge rack. In this example, WAN connectivity into the fabric is established between the spine switches and the data center core.

To provide site resiliency, a second availability zone (AZ2) is created and has connectivity to AZ1 through a Data Center Interconnection (DCI). The DCI is established between spine switches in both availability zones. VMware vSAN stretched clusters are established with the vSAN witness appliance located in Region B and are reachable through the DCI. Additional NSX components, including ESGs, are deployed in AZ2 and connected in a similar manner to AZ1.

Note: For more information about VCF on VxRail, see ***Dell EMC VxRail Hyperconverged Infrastructure***.

1.2 VMware Software-Defined Data Center

The VMware vision of the modern data center starts with a foundation of software-defined infrastructure. The foundation is based on the value that customers realize from a standardized architecture. It is a fully integrated hardware and software stack that is simple to manage, monitor, and operate. The VMware approach to the software-defined data center, or SDDC, delivers a unified platform that supports any application and provides flexible control. The VMware architecture for the SDDC enables companies to run private and hybrid clouds. The architecture uses unique capabilities to deliver key outcomes that enable efficiency, agility, and security.

The fully virtualized data center is automated and managed by intelligent, policy-based data center management software, vastly simplifying governance and operations. A unified management platform enables centralized monitoring and administration of all applications across physical geographies, heterogeneous infrastructure, and hybrid clouds. Workloads can be deployed and managed in physical, virtual, and cloud environments with a unified management experience. IT becomes agile, elastic, and responsive to a degree never before possible.

The VMware SDDC is based on well-established products from VMware. vSphere, vSAN and NSX provide compute, storage and networking virtualization to the SDDC and the vRealize Suite brings additional management, self-service, automation, intelligent operations and financial transparency. These VMware products form a solid foundation to host both traditional and cloud-native application workloads.

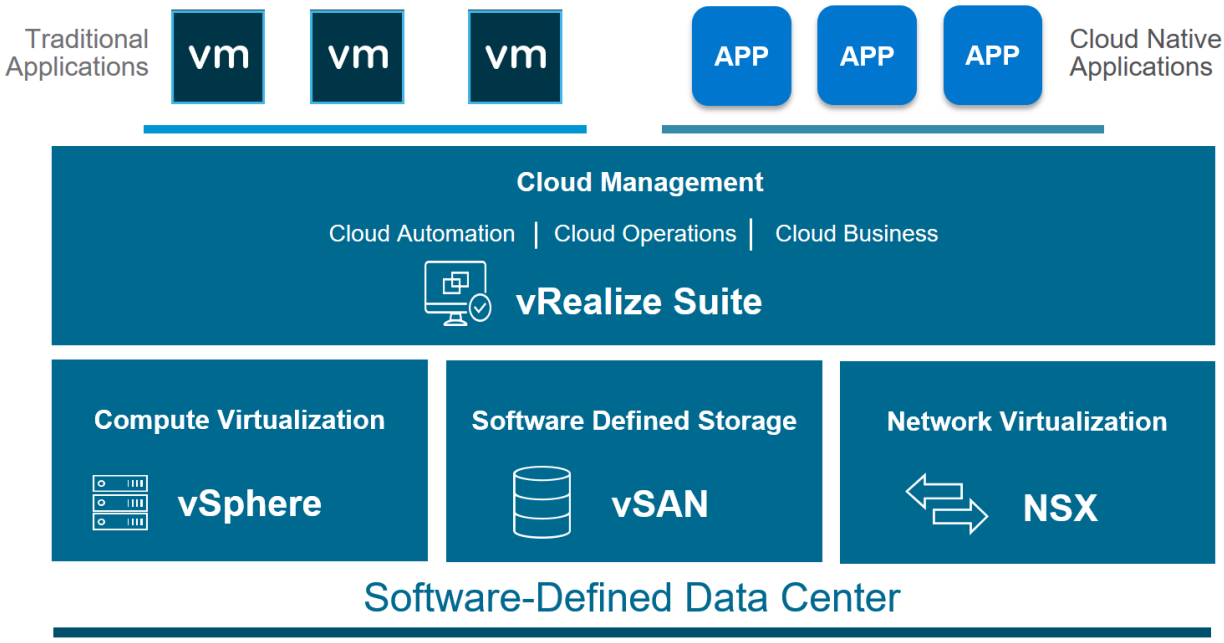


Figure 2 VMware software-defined data center high-level architecture

Note: To learn more about VMware SDDC, see *Dell EMC VxRail - Accelerating the Journey to VMware Software-Defined Data Center (SDDC)*.

1.3 VMware Cloud Foundation on VxRail

VMware Cloud Foundation (VCF) on VxRail is a new Dell EMC flagship offering for VMware SDDC. VCF builds upon native VxRail and Cloud Foundation capabilities with more unique Dell EMC and VMware jointly engineered integration features. These features simplify, streamline, and automate the operations of your entire SDDC from Day 0 through Day 2 operations.

VCF on VxRail provides the simplest path to the hybrid cloud through a fully integrated hybrid cloud platform. VCF on VxRail uses native VxRail hardware and software capabilities, and other VxRail unique integrations. vCenter plugins and Dell EMC networking integration work together to deliver a turnkey hybrid cloud user experience with full stack integration. Full stack integration means customers get an HCI infrastructure layer and cloud software stack in one, complete, automated life cycle, turnkey experience. The new platform delivers software defined services for compute using vSphere and vCenter, storage using vSAN, networking using NSX, and security and cloud management using vRealize Suite in both private or public environments. The services make it the operational hub for their hybrid cloud.

VCF on VxRail simplifies the data center by automating the public cloud in-house, and by deploying a standardized and validated network flexible architecture with integrated life cycle automation for the entire cloud infrastructure stack including hardware. SDDC Manager orchestrates the deployment, configuration, and life cycle management (LCM) of vCenter, NSX, and vRealize Suite above the ESXi and vSAN layers of VxRail. It unifies multiple VxRail clusters as workload domains or as multicluster workload domains.

Integrated with the SDDC Manager management experience, VxRail Manager is used to deploy and configure vSphere clusters that are powered by vSAN. It is also used to run the life cycle management of ESXi, vSAN, and hardware firmware using a fully integrated and unified SDDC Manager orchestrated process. It monitors the health of hardware components and provides remote service support as well. This level of integration is what gives customers a truly unique turnkey hybrid cloud experience not available on any other infrastructure.

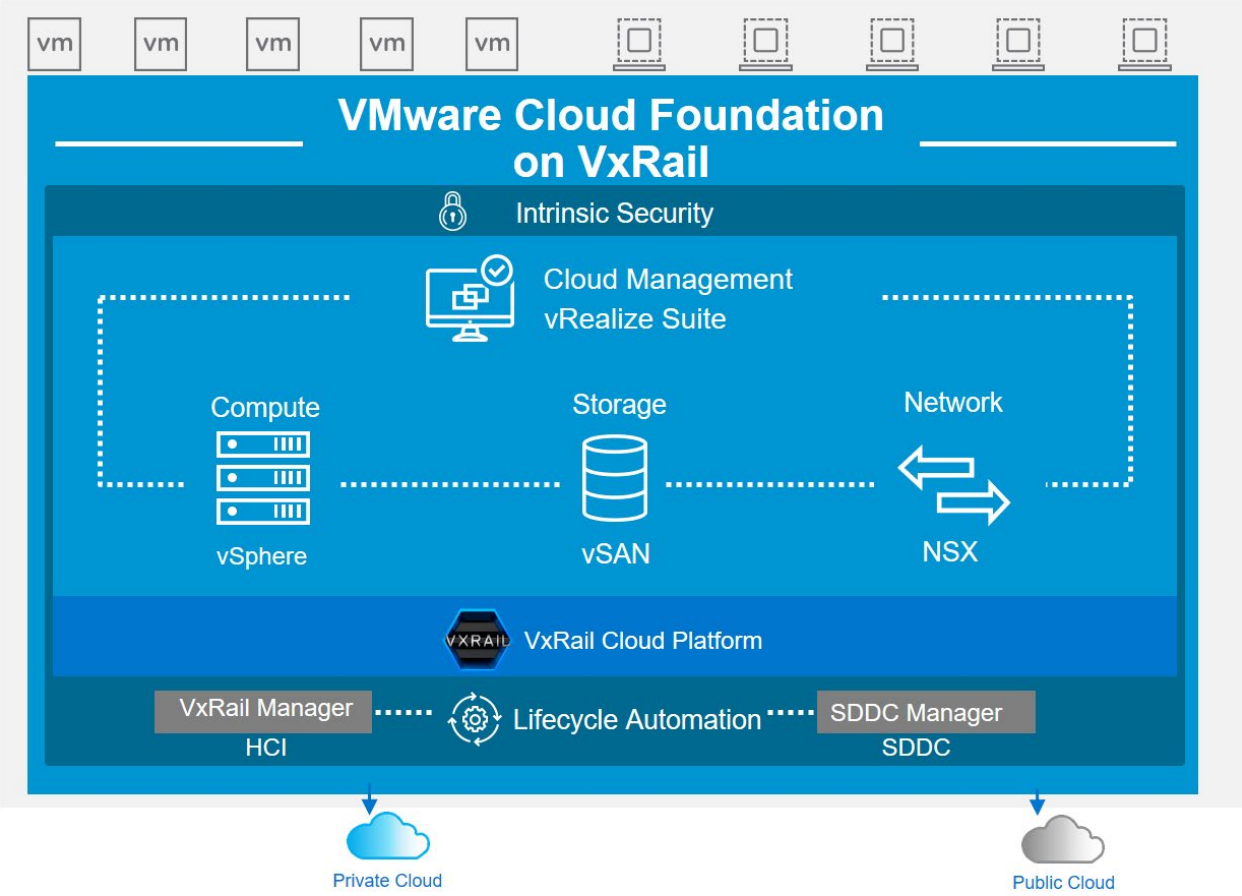


Figure 3 VMware Cloud Foundation on VxRail (VCF on VxRail) high-level architecture

To learn more about VMware Cloud Foundation on VxRail, see:

VMware Cloud Foundation on VxRail Administrator Guide

VMware Cloud Foundation on VxRail Planning and Preparation Guide

1.4 VMware Validated Design for SDDC on VxRail

VMware Validated Designs (VVD) simplify the process of deploying and operating an SDDC. They are comprehensive, solution-oriented designs that provide a consistent and repeatable production-ready approach to the SDDC. By definition, they are prescriptive blueprints that include comprehensive deployment and operational practices for the SDDC. It is an option available for customers, who are not ready or not value a complete approach to SDDC automation available in VCF on VxRail.

A VMware Validated Design is composed of a standardized, scalable architecture that is backed by the technical expertise of VMware and a software bill of materials (BOM) comprehensively tested for integration

and interoperability that spans compute, storage, networking, and management. Detailed guidance that synthetizes best practices on how to deploy, integrate, and operate the SDDC is provided to aid users to achieve performance, availability, security, and operational efficiency.

With the VVD for SDDC on VxRail, customers can easily architect, implement, and operate the complete SDDC faster and with less risk. Customers also get the benefits of best of breed HCI infrastructure platform. The latest available version at the time of writing this document is 5.0.

Customers can realize the following benefits by using VVD on VxRail:

- Accelerated time-to-market - streamline and simplify the complex design process of the SDDC, shortening deployment and provisioning cycles
- Increased efficiency – provide detailed, step-by-step guidance to reduce the time and the effort that is spent on operational tasks
- Lessen the uncertainty of deployments and operations - reduce uncertainty and potential risks that are associated with implementing and operating the SDDC
- IT agility – designed for expandability and to support a broad set of use cases and diverse types of applications that helps IT respond faster to the business needs

To learn more about VVD on VxRail, see *Dell EMC VxRail - Accelerating the Journey to VMware Software-Defined Data Center (SDDC)*.

1.5 Fabric Design Center

The Dell EMC Fabric Design Center (FDC) is a cloud-based application that automates the planning, design, and deployment of network fabrics that power Dell EMC compute, storage, and hyper-converged infrastructure solutions, including VxRail. The FDC is ideal for turnkey solutions and automation that is based on validated deployment guides like this one.

FDC enables design customization and flexibility to go beyond validated deployment guides. For additional information, see the Dell EMC Fabric Design Center.

1.6 Supported switches and operating systems

The examples provided in this Deployment Guide use VxRail 4.7.111 nodes that are connected to Dell EMC PowerSwitch S5248F-ON switches running the Dell EMC OS10 Enterprise Edition (OS10EE) 10.4.3.1. Any Dell PowerSwitch device can work but must be running OS10EE 10.4.3.1 or later to take advantage of NVO technologies referenced in this paper.

Dell EMC Networking supports the following switch and OS combinations for VxRail 4.7.111 and later using NVO technologies:

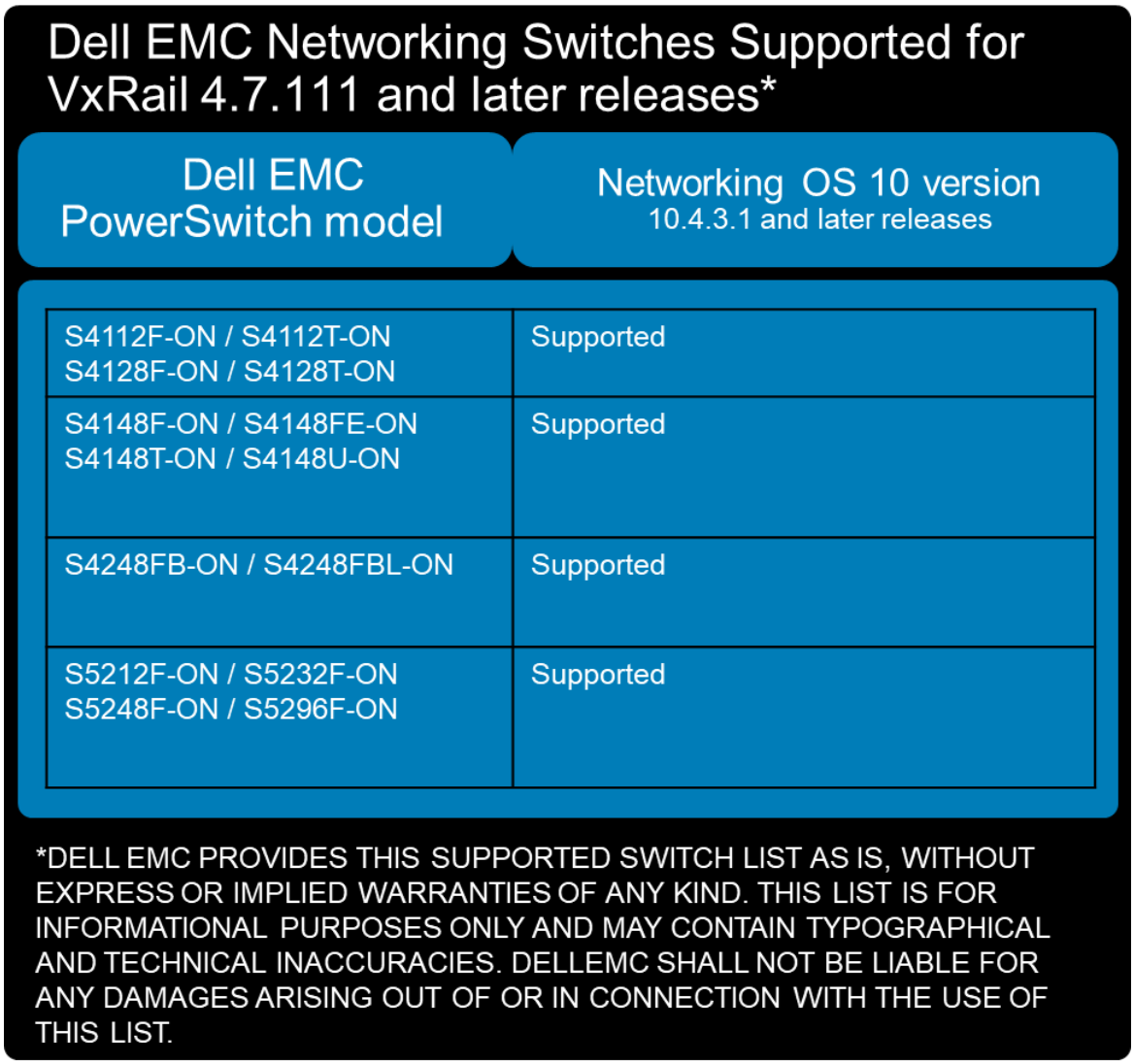


Figure 4 Supported Dell EMC Networking switches and operating systems

1.7 Typographical conventions

The CLI and GUI examples in this document use the following conventions:

`Monospace Text`	CLI examples
`Underlined Monospace Text`	CLI examples that wrap the page
`Italic Monospace Text`	Variables in CLI examples
`Bold Monospace Text`	Commands entered at the CLI prompt, or to highlight information in CLI output
Bold text	GUI fields and information entered in the GUI

1.8 Attachments

This document includes switch configuration file attachments. To access attachments in Adobe Acrobat Reader, click the ▶ icon in the left pane halfway down the page, then click the 📎 icon.

2 Physical infrastructure architecture

The architecture of the data center physical layer is based on logical hardware domains and the physical network topology. This section provides guidance on the hardware that is used as well as providing a general understanding of the physical network and the networking protocols used.

2.1 Hardware overview

This section provides an overview of the hardware that is used to validate this deployment. Appendix A contains a complete listing of hardware and software that is validated for this guide.

2.1.1 Dell EMC VxRail E560

The Dell EMC VxRail E series consists of nodes that are best suited for remote office or entry workloads. The E series nodes support up to 40 CPU cores, 1536GB memory, and 16TB hybrid or 30TB all-flash storage in a 1-Rack Unit (RU) form factor. The example within this document uses four VxRail E560 nodes, located in Region A, AZ1, and is part of the management cluster.

Figure 5 Dell EMC VxRail 1-RU node

2.1.2 Dell EMC VxRail P570

The Dell EMC VxRail P series consists of high-performance nodes that are optimized for heavy workloads, such as databases. Each appliance in the series has one node per 2-RU chassis. The models within this series are the Dell EMC VxRail P570 (hybrid), and the Dell EMC VxRail P570F (all-flash). The example within this document uses four VxRail P570 nodes, located in Region A, AZ2, and is part of the management cluster.

Figure 6 Dell EMC VxRail 2-RU node

2.1.3 Dell EMC PowerSwitch S5248F-ON

The Dell EMC PowerSwitch S5248F-ON is a 1-RU fixed switch with 48x 25 GbE, 4x multirate 100 GbE, and 2x 200 GbE ports. The S5248F-ON supports L2 static VXLAN with VLT. The example within this document uses eight S5248F-ON switches in VLT pairs as leaf switches. There are four pairs total, two in AZ1 and two in AZ2.

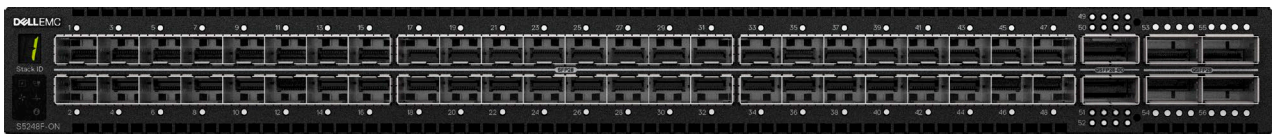


Figure 7 Dell EMC PowerSwitch S5248F-ON

2.1.4 Dell EMC PowerSwitch S5232F-ON

The Dell EMC PowerSwitch S5232F-ON is a 1-RU fixed switch with 32x 100 GbE and 2x10 GbE ports. The S5232F-ON is part of the S5200-ON series of switches. The example within this document uses two S5232F-ON switches as spine switches in Region A, AZ2.

Figure 8 Dell EMC PowerSwitch S5232F-ON

2.1.5 Dell EMC PowerSwitch Z9264F-ON

The Dell EMC PowerSwitch Z9264F-ON is a 2-RU 100 GbE aggregation/spine switch. The Z9264F-ON has up to 64 ports of multirate 100 GbE, or up to 128 ports of 10/25/40/50 GbE ports using supported breakout cables. The example within this document uses two Z9264F-ON switches as spine switches in Region A, AZ1.

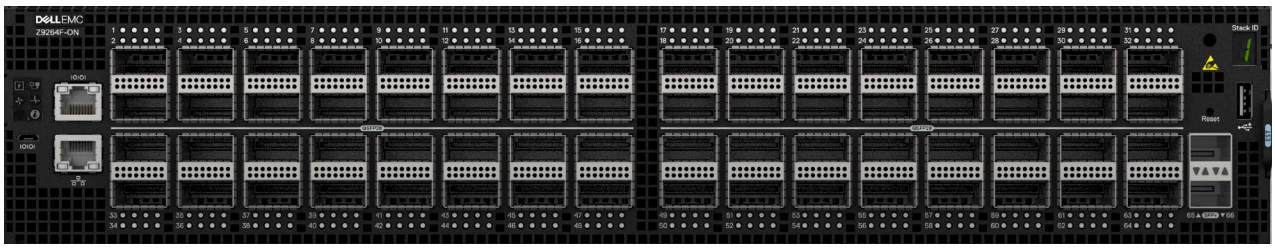


Figure 9 Dell EMC PowerSwitch Z9264F-ON

Note: Either the Dell EMC PowerSwitch S5232F-ON or the Z9264F-ON can be used as spine switches for either or both environments. A combination was used in this document to illustrate the capabilities of both platforms.

2.1.6 Dell EMC PowerSwitch S3048-ON

The Dell EMC PowerSwitch S3048-ON is a 1-RU switch with 48x1GbE BASE-T ports and 4x 10GbE SFP+ ports. This guide uses one S3048-ON switch for out-of-band (OoB) management traffic.

Figure 10 Dell EMC PowerSwitch S3048-ON

2.2 Network Transport

VMware Validated Design supports both Layer 2 and Layer 3 transports. In this section, the details of the Layer 3 leaf-spine topology are provided.

2.2.1 Layer 3 leaf and spine topology

In this document, a Clos leaf-spine topology is used for each availability zone. Individual switch configuration shows how to set up end-to-end Virtual Extensible Local Area Networks (VXLANs). External Borer Gateway Protocol (eBGP) is used for exchanging IP routes in the IP underlay network, and EVPN routes in the VXLAN overlay network. Virtual Link Trunking (VLT) is deployed between leaf pairs and internal BGP (iBGP) to provide Layer 3 path redundancy in the event a leaf switch loses connectivity to the spine switches.

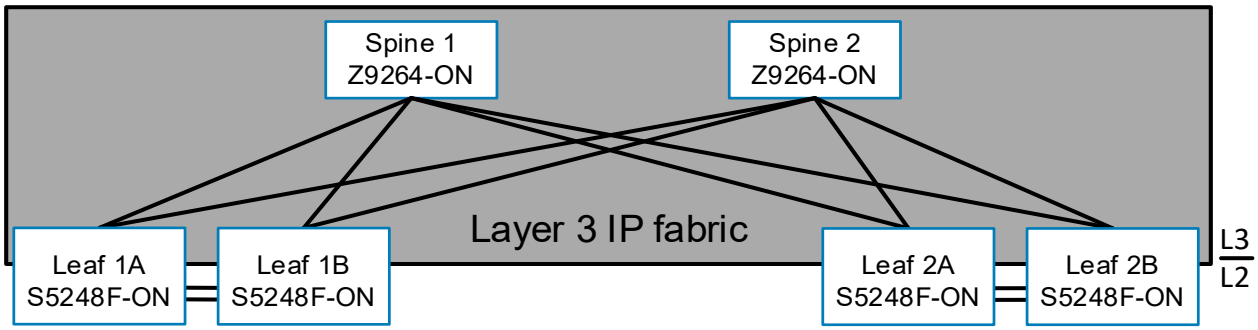


Figure 11 Layer 3 IP network transport

Note: For detailed instructions on creating a leaf-spine underlay, including considerations and alternative configurations, see *Dell EMC Networking Layer 3 Leaf-Spine Deployment and Best Practices with OS10EE.*

2.2.2 Border Gateway Protocol (BGP) Ethernet VPN (EVPN)

EVPN is a control plane for VXLAN tunnels over a Layer 3 infrastructure such as a leaf-spine network. EVPN is used to reduce flooding in the network and resolve scalability concerns when compared against Static VXLAN. EVPN uses multiprotocol BGP (MP-BGP) to exchange information between Virtual Tunnel Endpoints (VTEPs). EVPN was introduced in RFC 7432 and is based on BGP MPLS VPNs. RFC 8365 describes VXLAN-based EVPN.

The MP-BGP EVPN control plane provides protocol-based remote VTEP discovery, and MAC and ARP learning. As a result, flooding related to L2 unknown unicast traffic is reduced. The distribution of host MAC and IP reachability information supports virtual machine mobility and scalable VXLAN overlay network designs.

The BGP EVPN protocol groups MAC addresses and ARP/neighbor addresses under EVPN instances (EVIs) to exchange them between VTEPs. In OS10EE, each Virtual Network Interface (VNI) is mapped using switch scoped assignments where each VLAN is associated with a VNI in a 1:1 mapping. The benefits for deploying a BGP EVPN for VXLAN topology include:

- Eliminates the flood-and-learn method of VTEP discovery by enabling control-plane learning of end-host Layer 2 and Layer 3 reachability information
- Minimizes the network flooding of unknown unicast and broadcast traffic through EVPN-based MAC and IP route advertisements on local VTEPs
- Supports host mobility

OS10EE supports two types of VXLAN Network Virtual Overlay (NVO) networks:

- Static VXLAN
- BGP EVPN

Static VXLAN and BGP EVPN for VXLAN are configured and operate in the same ways:

- The overlay and underlay networks are manually configured
- Each virtual network and VNI are manually configured
- Access port membership in a virtual network is manually configured
- Underlay reachability to VTEPs peers is provisioned or learned using existing routing protocols

Static VXLAN and BGP EVPN for VXLAN configuration and operation differ as described in Table 1.

Table 1 Differences between Static VXLAN and VXLAN BGP EVPN

Static VXLAN	BGP EVPN for VXLAN
To start sending and receiving virtual-network traffic to and from a remote VTEP, manual configuration of VTEP must be completed.	No manual configuration is required. Each remote VTEP is automatically learned as a member of a virtual network from the EVPN routes received from the remote VTEP. After a remote VTEP's address is learned, VXLAN traffic is sent to, and received from, the VTEP.
Remote hosts are learned from data packets after decapsulation of the VXLAN header in the data plane.	Remote host MAC addresses are learned in the control plane using BGP EVPN Type 2 routes and MAC/IP advertisements

Note: For more information about static L2 VXLAN configuration, see *Dell EMC VxRail Multirack Deployment Guide*.

In this guide, BGP EVPN is used to create virtual networks to handle all the VxRail specific VLANs to create a management domain. This includes VMware ESXi management, vMotion, vSAN, and node discovery. Additionally, external services are in a separate virtual network to highlight the capabilities of Routing in and Out of Tunnels (RIOT).

All VMware NSX traffic, including NSX VTEPs and ESG traffic, is handled by traditional networking concepts and routed through the default VRF.

Terminology

- **VXLAN (Virtual Extensible LAN)** - The technology that provides the same Ethernet Layer 2 network services as VLAN does today, but with greater extensibility and flexibility
- **VNID (VXLAN Network Identifier)** - 24 bit segment ID that defines the broadcast domain.
- **VTEP (Virtual Tunnel Endpoint)** - This is the device that does the encapsulation and de-encapsulation
- **NVE (Network Virtual Interface)** - Logical interface where the encapsulation and de-encapsulation occur

2.2.3 EVPN instance and Autonomous System Number (ASN) considerations

An EVPN instance (EVI) spans across the VTEPs that participate in an EVPN. Each virtual network, or tenant segment, that is advertised using EVPN must be associated with an EVI. In OS10, configure EVIs in auto-EVI or manual configuration mode.

Auto-EVI configuration mode

When a virtual network is associated with a VTEP, auto-EVI mode automatically creates an EVPN instance. Auto EVI mode does the following:

- The EVI ID auto-generates with the same value as the virtual-network ID (VNID) configured on the VTEP and associates with the VXLAN network ID (VNI)
- A route distinguisher (RD) auto-generates for each EVI ID. An RD maintains the uniqueness of an EVPN route between different EVPN instances.
- A Route Target (RT) import and export auto-generates for each EVI ID. A RT determines how EVPN routes are distributed among EVPN instances.

Manual EVI configuration mode

To specify the RD and RT values, manually configure EVPN instances and associate each EVI with the overlay virtual network using the VXLAN VNI. The EVI activates only when the virtual network, RD, and RT values are configured.

Route distinguisher

A RD is an 8-byte identifier that uniquely identifies an EVI. Each EVPN route is prefixed with a unique RD and exchanged between BGP peers, making the tenant route unique across the network. In this way, overlapping address spaces among tenants are supported.

Route target

While the RD maintains the uniqueness of an EVPN route among different EVIs, an RT controls the way the EVPN routes are distributed among EVIs. Each EVI is configured with an import and export RT value. BGP EVPN routes advertise for an EVI and carry the export RT associated with the EVI. A receiving VTEP downloads information in the BGP EVPN route to EVIs that have a matching import RT value.

Autonomous System Number (ASN)

In this document, manual EVI configuration mode is used to ensure that ASN values are different between leaf switches participating in the same virtual networks. In this document, each leaf pair is associated with a unique ASN value regardless of the availability zone. Figure 12 shows the ASN values used in AZ1 and AZ2 will use different values. Using manual EVI lends itself to a brownfield environment where the ASN values have already been established.

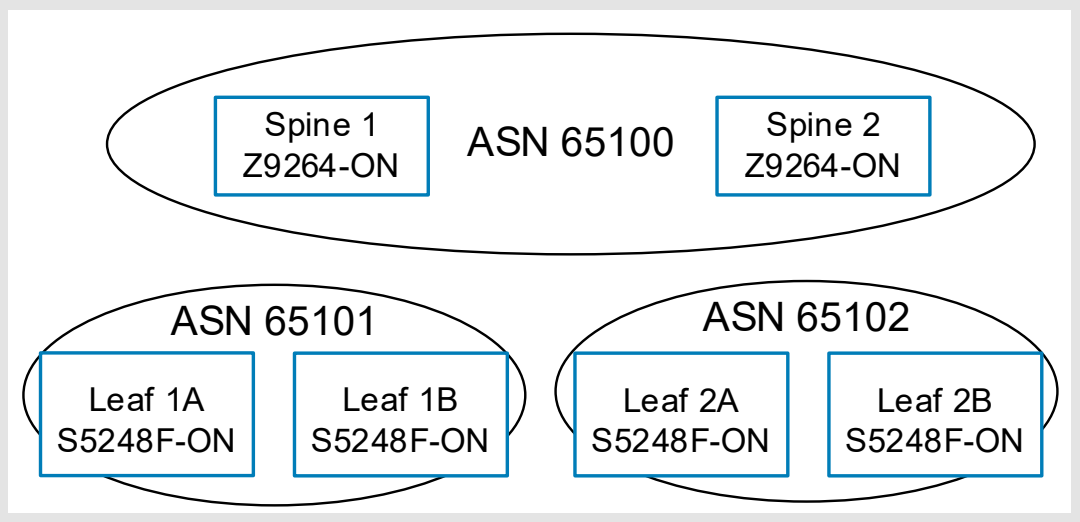


Figure 12 ASN values in AZ1

Note: In this example, private AS 2-byte values are used, and these should be changed to reflect the AS numbers and needs used in the environment. For example, in the *Dell EMC Networking Layer 3 Leaf-Spine Deployment and Best Practices with OS10EE* guide separate AS numbers are used for each spine switch, both models are accessible.

2.2.4 VXLAN frame format Maximum Transmission Unit considerations

VXLAN allows a Layer 2 network to scale across the data center by overlaying an existing Layer 3 network and is described in Internet Engineering Task Force document RFC 7348. Each overlay is referred to as a VXLAN segment.

Each segment is identified through a 24-bit segment ID referred to as a VNI. This allows up to 16 Million VNIs, far more than the traditional 4,094 VLAN IDs allowed on a physical switch.

VXLAN is a tunneling scheme that encapsulates Layer 2 frames in User Datagram Protocol (UDP) segments, as shown in Figure 13.

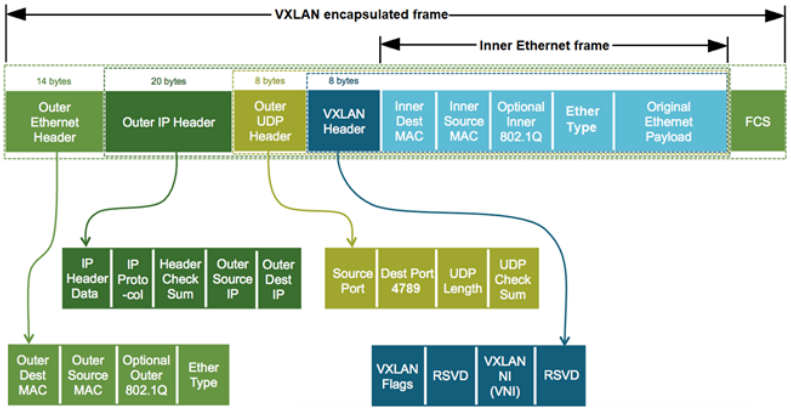


Figure 13 VXLAN encapsulated frame

VXLAN encapsulation adds approximately 50 bytes of overhead to each Ethernet frame. As a result, all switches in the underlay (physical) network must be configured to support an MTU of at least 1600 bcytes on all participating interfaces.

2.2.5 Routing in overlay networks

Integrated routing and bridging (IRB)

With EVPN, overlay routing occurs on leaf switches. As of OS10EE version 10.4.3.1, Dell EMC PowerSwitches support asymmetric IRB. This means that overlay routing, also known as RIOT, occurs on ingress leaf switches. The packets travel over the leaf and spine network on the destination VNI. When the packets arrive at the destination VTEP, they are bridged to the endpoint.

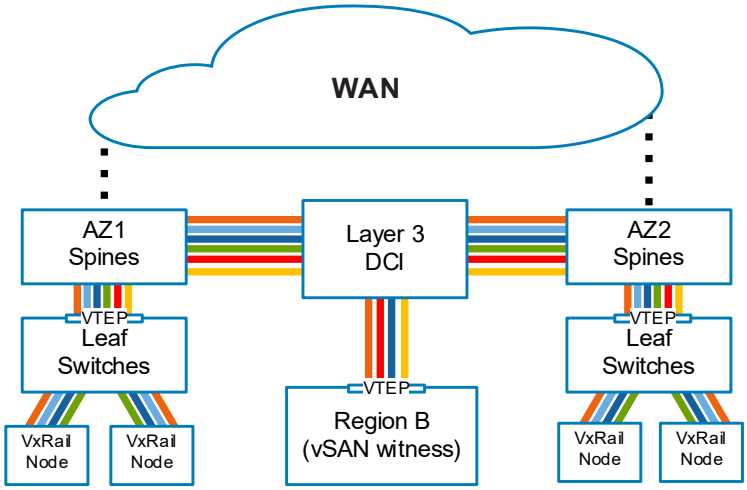
Anycast gateway

Anycast gateways with the same IP address are deployed to each leaf pair connected to servers. The anycast gateway IP address is set as the default gateway for all VMs on that virtual network. VMs on VNIs with anycast gateways use the same gateway information while behind different leaf pairs. When those VMs communicate with different networks, their local leaf switches always do the routing. This replaces Virtual Router Redundancy Protocol (VRRP) and enables VMs to migrate from one leaf pair to another without the need to change the network configuration.

Indirect gateway

As shown in Figure 6, VNI C does not have an anycast gateway configured. It uses an indirect gateway attached to the border leafs instead. In this case, the indirect gateway is a physical switch acting as a firewall/gateway to the Internet. When a VM on VNI C sends a packet destined for another network, it is tunneled to the border leaf pair where it is forwarded to the indirect gateway which makes the proper forwarding decisions according to its routing table.

In this deployment example, all VNIs use the anycast gateway model to provide IP reachability between both availability zones and Region B. While not discussed in this document, the indirect gateway model would be used to route any VNI to Internet.



Name	Location	VLAN
Management	AZ1/AZ2	1611
	RegionB	1711
vSAN	AZ1/AZ2	1613
	RegionB	1713
Discovery	AZ1/AZ2	3939
vMotion	AZ1/AZ2	1612

Figure 14 Distributed routing

2.2.6 Leaf and Spine switch characteristics

The leaf switches act as the VXLAN tunnel endpoints and perform VXLAN encapsulation and decapsulation. The leaf switches also participate in the MP-BGP EVPN to support control plane and data plane functions.

The control plane functions include:

- Initiate and maintain route adjacencies using any routing protocol in the underlay network
- Advertise locally learned routes to all MP-BGP EVPN peers
- Process the routes received from remote MP-BGP EVPN peers and install them in the local forwarding plane

The data plane functions include:

- Encapsulate server traffic with VXLAN headers and forward the packets in the underlay network
- Decapsulate VXLAN packets received from remote VTEPs and forward the native packets to downstream hosts
- Perform underlay route processing, including routing based on the outer IP address

The role of the spine switch changes based on its control plane and data plane functions. Spine switches participate in underlay route processing to forward packets and in the overlay network to advertise EVPN routes to all MP-BGP peers.

The control plane functions include:

- Initiate BGP peering with all neighbor leaf switches
- Advertise BGP routes to all BGP peers
- In the underlay network, initiate and maintain the routing adjacencies with all leaf and spine switches

The data plane functions include:

- Perform only the underlay route processing based on the outer header in VXLAN encapsulated packets
- Does not perform VXLAN encapsulation or decapsulation

2.2.7 Underlay network physical design AZ1

Figure 15 shows the wiring configuration for the six switches that comprise the leaf-spine network in AZ1. The colored solid lines are 100 GbE links and the light blue dashed lines are two QSFP28-DD 200 GbE cable pairs used for the VLT interconnect (VLTi). The use of QSFP28-DD offers a 400 GbE VLTi to handle any potential traffic increases resulting from failed interconnects to the spine layers. As a rule, it is suggested to maintain at minimum a 1:1 ratio between available bandwidth to the spine and bandwidth for the VLTi.

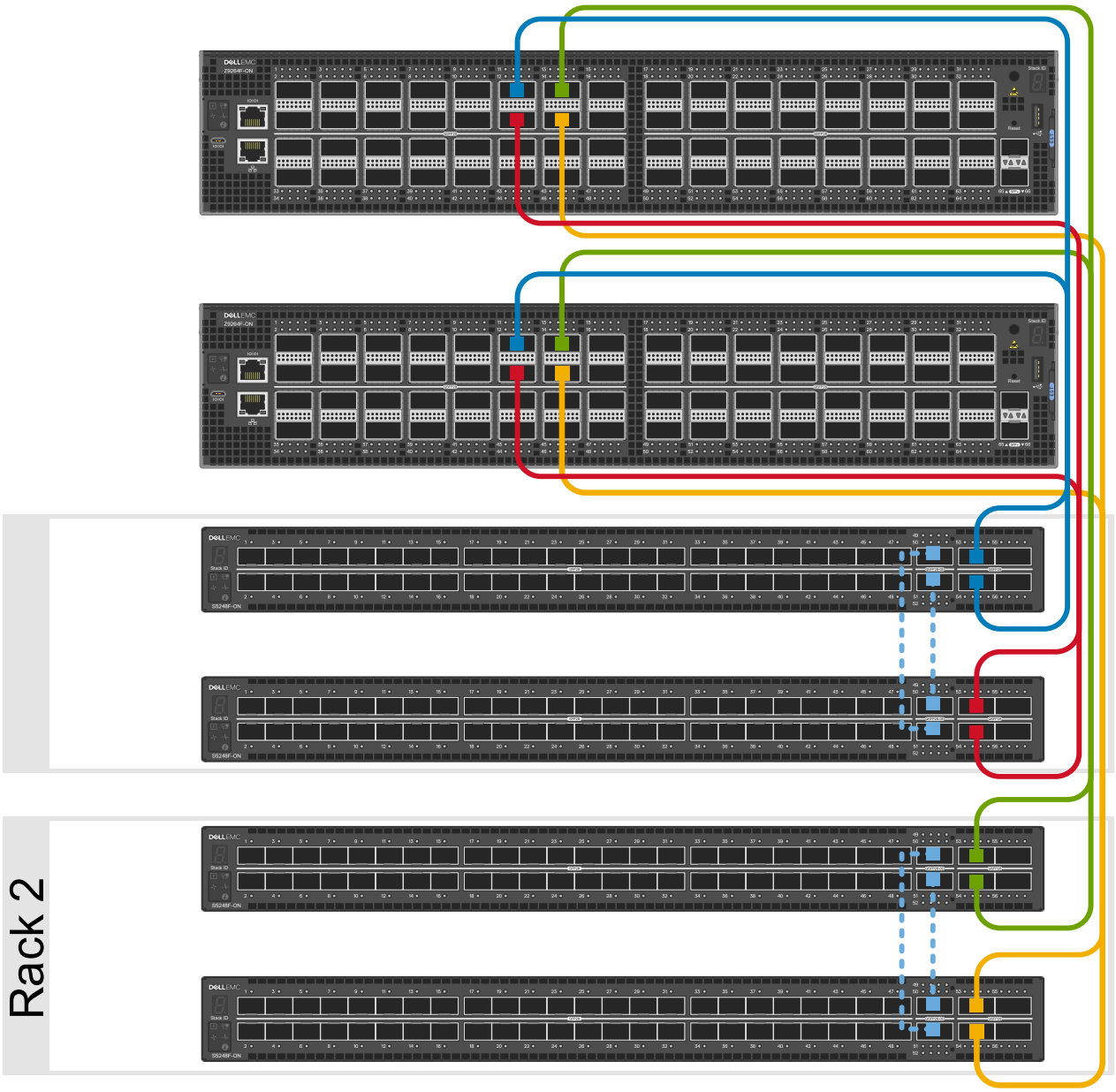


Figure 15 Physical switch topology AZ1

Note: All switch configuration commands are provided in the file attachments. See Section 1.8 for instructions on accessing the attachments.

2.3 Management domain architecture characteristics

Workload domains include combinations of ESXi hosts, and network equipment which can be set up with varying levels of hardware redundancy. Workload domains are connected to a network core that distributes data between them.

A workload domain represents a logical boundary of functionality that is managed by a single vCenter Server instance. Hard physical properties do not define the workload domain. Although a workload domain usually spans one rack, Figure 16 shows how BGP EVPN and VXLAN at the network switch level which enables a workload domain to span multiple racks.

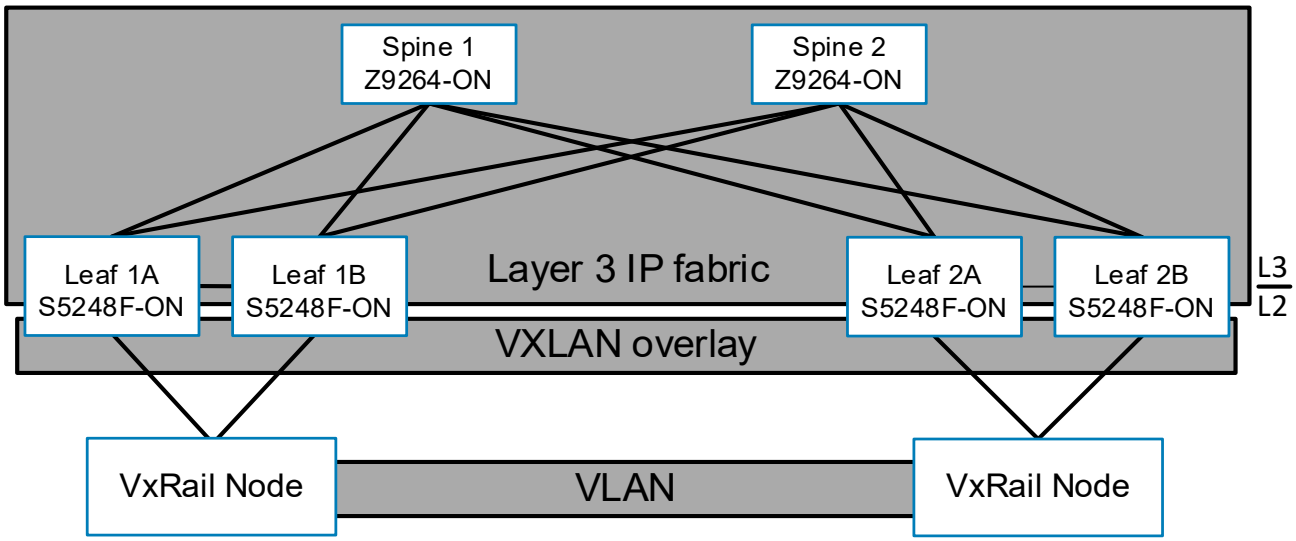


Figure 16 Management domain architecture using a VXLAN overlay in a single Availability Zone

In this document, four E-series VxRail nodes in Region A, Availability Zone 1 (AZ1) and four P-Series VxRail nodes in Region A, Availability Zone 2 (AZ2), create a management workload domain following VMware Validated Design (VVD) guidelines. The eight total nodes are equally divided between four racks (two in each AZ). The management cluster runs the virtual machines that manage the Standard SDDC. These virtual machines host:

- VxRail Manager
- vCenter Server
- vSphere Update Manager
- NSX Manager
- NSX Controllers
- vRealize suite
- Site Recovery Manager
- Other management components

All management, monitoring, and infrastructure services are provisioned to a vSphere cluster which provides high availability for these critical services

Figure 17 shows a single VxRail node and the associated VLANs required to complete a successful workload deployment. Each of these VLANs are mapped to a single VNI creating multiple stretched Layer 2 VLANs across the underlay, enabling east to west traffic flows between the nodes. The figure represents the minimum requirements for VxRail nodes only. Additional VLANs and IP subnets required to support VMware NSX VTEP and ESG are configured later in the document.

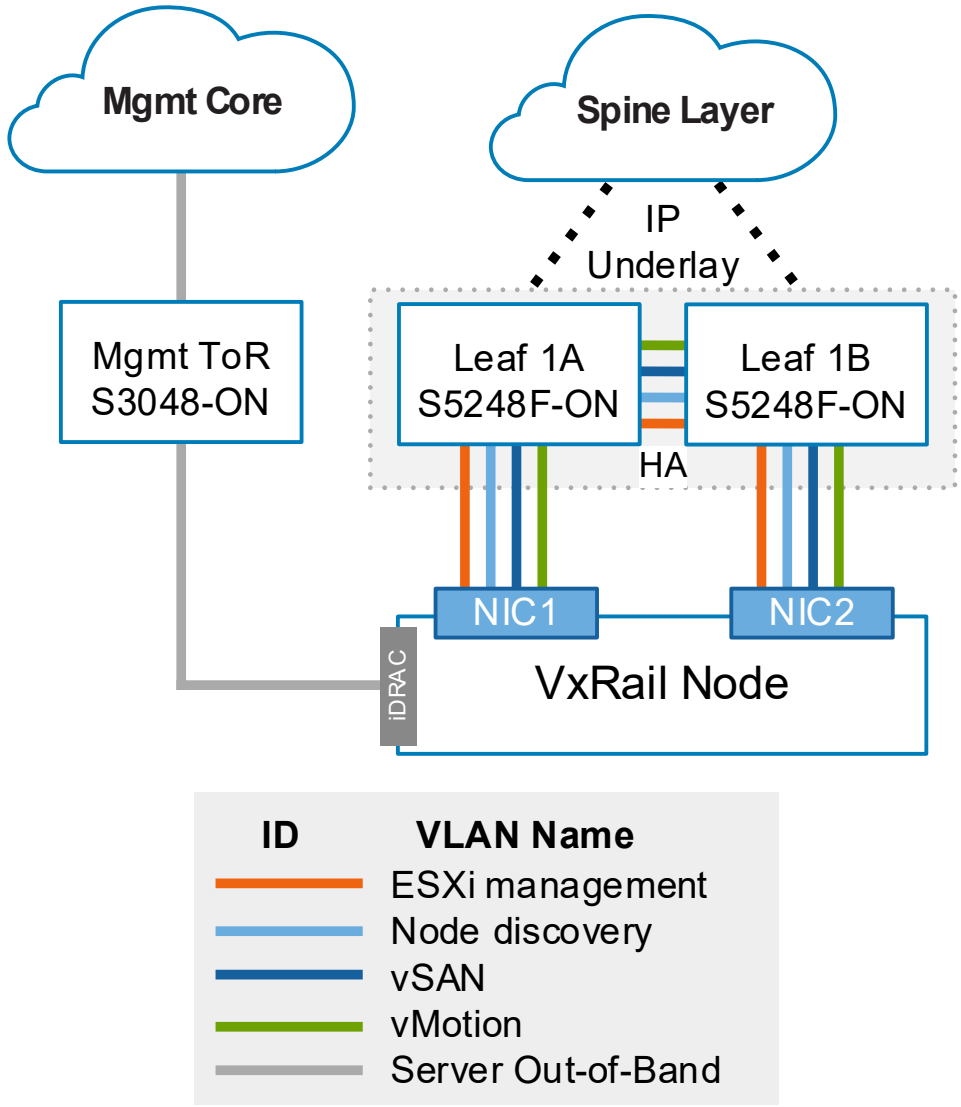


Figure 17 VxRail node connectivity to Dell EMC PowerSwitch S5248F-ON leaf switches

Figure 18 shows a physical view of Rack 1. The VxRail E series nodes sfo01w01vxrail01 through 04, each has 2x 25 GbE links with each being connected to one of the two S5248F-ON leaf switches in the rack. Each VxRail node has an iDRAC connected to a S3048-ON OOB management switch. This connection is used for the initial node configuration. The S5248F-ON leaf switches are connected using two QSFP28-DD 200 GbE direct access cables (DAC) forming a VLT interconnect (VLTi) for a total throughput of 400 GbE. Upstream connections to the spine switches are not shown but are configured using two QSFP28 100 GbE uplinks.

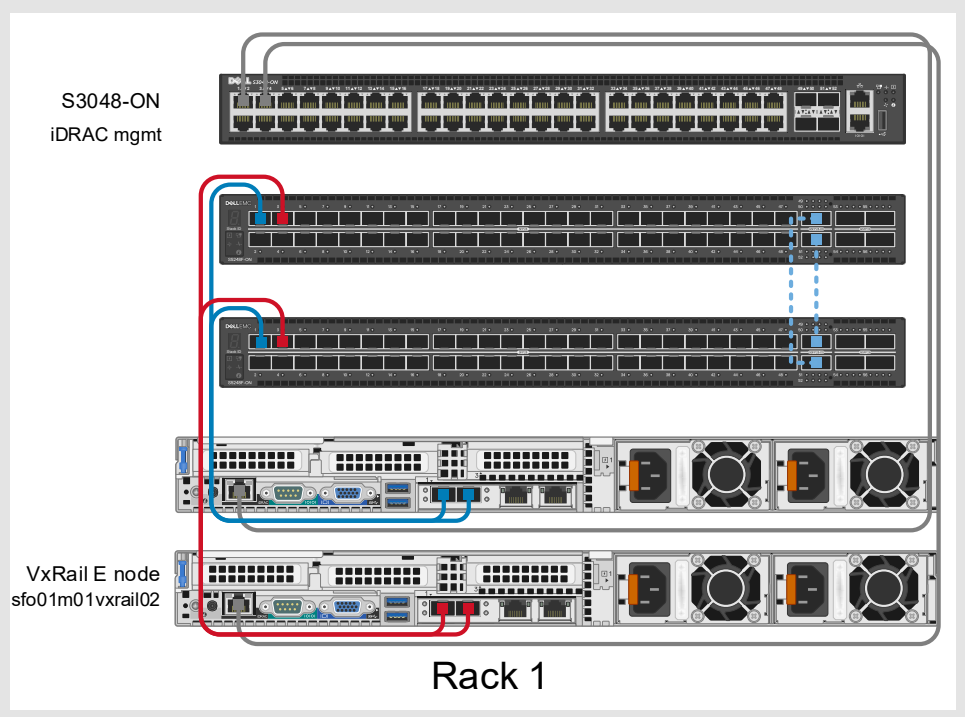


Figure 18 Dell EMC VxRail multirack Rack 1 physical connectivity

3 Planning and Preparation for AZ1

Before creating the IP underlay that drives the SDDC, it is important to plan out the networks, IP subnets, and external services required. Also, planning of the prerequisites on all required switching hardware is recommended.

3.1 VLAN IDs and IP subnets

VCF on VxRail requires that certain VLAN IDs and IP subnets for the traffic types in the SDDC are defined ahead of time. To meet these requirements Table 2 shows the values that are used in this document. The NSX VTEP VLAN is configured in a way that uses the VLAN ID in each rack while the IP subnet is changed based on the rack and AZ. For example, 172.20.101.0/24 would represent AZ1 Rack 1 with a Gateway of 172.20.101.253. In AZ2 172.20.202.0/24 would be Rack 2.

Table 2 VLAN and IP subnet configuration in Region A, AZ1

Cluster in Region A	VLAN function	VLAN ID	VNI	Subnet	Gateway
Management cluster	ESXi Management	1611	1611	172.16.11.0/24	172.16.11.253
	vSphere vMotion	1612	1612	172.16.12.0/24	172.16.12.253
	vSAN	1613	1613	172.16.13.0/24	172.16.13.253
	NSX VTEP	2000	-	172.20.*n*.0/24	172.20.*n*.253
	VxRail node discovery	3939	3939	-	-
	Uplink01	2711	-	172.27.11.0/24	172.27.11.1
	Uplink02	2712	-	172.27.12.0/24	172.27.12.1

Note: Use these VLAN IDs and IP subnets as samples. Configure the actual VLAN IDs and IP subnets according to the environment.

3.2 External Services

External services covered in the VVD include:

- Active Directory (AD)
- Dynamic Host Control Protocol (DHCP)
- Domain Name Services (DNS)
- Network Time Protocol (NTP)
- Simple Mail Transport Protocol (SMTP) Mail Relay
- File Transfer Protocol (FTP)
- Certificate Authority (CA)

In this section, the following are discussed along with the guidelines on placement:

- DNS
- NTP
- DHCP

3.2.1 DNS

In this document, the AD servers provide DNS services. Other DNS records that are used in this document follow the VVD examples and can be found in the **Host Names and IP Addresses in Region A** section under **Planning and Preparation**.

Table 3 Host names and IP addresses for the external services

Component group	Host name	DNS zone	IP address	Description
AD/DNS	dc01rpl	rainpole.local	172.16.11.4	Windows 2016 host containing AD and DNS server for rainpole.local
	dc01sfo	sfo01.rainpole.local	172.16.11.5	AD and DNS server in a child domain

3.2.2 NTP

Synchronized systems over NTP are essential for the validity of vCenter Single Sign-On and other certificates. Consistent system clocks are important for the proper operation of the components in the SDDC because in certain cases they rely on vCenter Single Sign-on. Using NTP also makes it easier to correlate logfiles from multiple sources during troubleshooting, auditing, or inspection of logfiles to detect attacks.

Table 4 shows the DNS Canonical Name (CNAME) record that maps the two time sources to one DNS name.

Table 4 NTP server FQDN and IP configuration in Region A

NTP server FQDN	Mapped IP Address
ntp.sfo01.rainpole.local	• 172.16.11.5 • 172.16.11.4
0.ntp.sfo01.rainpole.local	172.16.11.5
1.ntp.sfo01.rainpole.local	172.16.11.4

3.2.3 DHCP

DHCP is required for each VMkernel port of the ESXi hosts with an IPv4 address. A Microsoft Windows Server 2016 virtual machine that is associated with external services on subnet 10.10.14.0/24 is used in this deployment. DHCP relay (`ip help-address`) is used to route DHCP requests on behalf of the NSX VTEPs to the DHCP server. Table 5 outlines the DHCP values used in this document.

The VVD outlines the example usage of VLAN 1614 and the IP subnet of 172.16.14.0/24. In this paper, this has been modified to accommodate multiple subnets. VLAN ID 2000 is used and the corresponding IP subnets are reserved in the underlay network for these subnets. The third octet is increased by 1 to represent the rack ID with Region A, AZ1, rack 1 being 172.20.101.0/24 while Region A, AZ2, rack 16 could be 172.20.216.0/24.

Table 5 shows the IP address ranges used in this document. The DHCP servers in either availability zone is assumed to be configured correctly and are outside of the scope of this document.

Table 5 DHCP scope values for AZ1

ID	DHCP server IP address	Start IP address	End IP address	Gateway	Subnet Mask
RegionA-AZ1-R1	10.10.14.5	172.20.101.1	172.20.101.199	172.20.101.253	/24
RegionA-AZ1-R2	10.10.14.5	172.20.102.1	172.20.102.199	172.20.102.253	/24
RegionA-AZ2-R1	172.20.201.28	172.20.201.1	172.20.201.199	172.20.201.253	/24
RegionA-AZ2-R2	172.20.201.28	172.20.202.1	172.20.202.199	172.20.202.253	/24

Note: In this example the DHCP server in AZ2 was configured on the same Layer 2 segment as the infrastructure. This is not a requirement. A DHCP server requires IP reachability and can be located on the same segment or externally to the fabric to function.

3.3 Switch settings

Table 6 shows the unique values for the four S5248F-ON switches in AZ1. The table provides a summary of the configuration differences between each switch and each VLT switch pair.

Table 6 Unique switch settings for leaf switches in AZ1

Setting	S5248F-Leaf1A	S5248F-Leaf1B	S5248F-Leaf2A	S5248F-Leaf2B
Hostname	sfo01-Leaf01A	sfo01-Leaf01B	sfo01-Leaf02A	sfo01-Leaf02B
OOB IP address	100.67.198.32/24	100.67.198.31/24	100.67.166.36/24	100.67.166.35/24
Autonomous System Number (ASN)	• 65101	• 65101	• 65102	• 65102
Point-to-point interface IP addresses	• 192.168.1.1/31 • 192.168.2.1/31	• 192.168.1.3/31 • 192.168.2.3/31	• 192.168.1.5/31 • 192.168.2.5/31	• 192.168.1.7/31 • 192.168.2.7/31
Loopback0 address (router ID)	10.0.2.1/32	10.0.2.2/32	10.0.2.3/32	10.0.2.4/32
Loopback1 address (EVPN)	10.2.2.1/32	10.2.2.2/32	10.2.2.3/32	10.2.2.4/32
Loopback2 address (NVE)	10.222.222.1/32	10.222.222.1/32	10.222.222.2/32	10.222.222.2/32
VLAN 4000 IP address	192.168.3.0/31	192.168.3.1/31	192.168.3.2/31	192.168.3.3/31
VLAN 2000 IP addresses (interface and VIP)	• 172.20.101.251/24 • 172.20.101.253/24	• 172.20.101.252/24 • 172.20.101.253/24	• 172.20.102.251/24 • 172.20.102.253/24	• 172.20.102.252/24 • 172.20.102.253/24
VLAN 2711 IP addresses (ESG)	172.27.11.1/24	-	-	-
VLAN 2712 IP addresses (ESG)	-	172.27.12.1/24	-	-
virtual-network 1611 IP addresses (interface and anycast)	• 172.16.11.252/24 • 172.16.11.253/24	• 172.16.11.251/24 • 172.16.11.253/24	• 172.16.11.250/24 • 172.16.11.253/24	• 172.16.11.249/24 • 172.16.11.253/24
virtual-network 1612 IP addresses (interface and anycast)	• 172.16.12.252/24 • 172.16.12.253/24	• 172.16.12.251/24 • 172.16.12.253/24	• 172.16.12.250/24 • 172.16.12.253/24	• 172.16.12.249/24

				• 172.16.12.253/24
virtual-network 1613 IP addresses (interface and anycast)	• 172.16.13.252/24 • 172.16.13.253/24	• 172.16.13.251/24 • 172.16.13.253/24	• 172.16.13.250/24 • 172.16.13.253/24	• 172.16.13.249/24 • 172.16.13.253/24

Note: Use these VLAN IDs and IP subnets as samples. Configure the actual VLAN IDs and IP subnets according to your environment.

3.3.1 Verify OS10 version

The Dell EMC Networking S5248F-ON switches must have OS10EE version 10.4.3 or later to support Network Virtualization Overlays (NVOs). Run the **`show version`** command to check the operating system version.

Note: Dell EMC recommends upgrading to the latest release available on Dell Digital Locker (account required).

```
OS10# show version
Dell EMC Networking OS10-Enterprise
Copyright (c) 1999-2018 by Dell Inc. All Rights Reserved.
OS Version: 10.4.3.1
Build Version: 10.4.3.1.154
Build Time: 2019-03-20T18:16:25-0700
```

Note: Figure 4 provides a list of the supported switches and operating systems for VxRail deployments.

3.3.2 Verify license installation

Run the command `show license status` command to verify license installation. Locate the `License Type:` field and verify that `PERPETUAL` displays in the field.

```
OS10# show license status

System Information
---------------------------------------------------------
Vendor Name     :        Dell EMC
Product Name    :        S5248F-ON
Hardware Version:        A00
Platform Name   :        x86_64-dellemc_s5248f_c3538-r0
PPID            :        CN046MRJCES0085N0006
Service Tag     :        AAAAAAA
License Details
----------------
Software        :        OS10-Enterprise
Version         :        10.4.3.1
License Type    :        PERPETUAL
License Duration:        Unlimited
License Status  :        Active
License location:        /mnt/license/AAAAAAA.lic
---------------------------------------------------------
```

Note: If an evaluation license is installed, licenses purchased from Dell EMC are available for download on Dell Digital Locker. See the ***OS10 Enterprise Edition User Guide*** for installation instructions.

3.3.3 Factory default configuration

The configuration commands begin with the Dell EMC Networking switches at their factory default settings. Dell EMC Networking switches running the Dell EMC OS10 Enterprise Edition (OS10EE) can be reset to their default configuration using the following commands:

```
OS10# delete startup-configuration
Proceed to delete startup-configuration [confirm yes/no(default)]:y

OS10# reload
System configuration has been modified. Save? [yes/no]:n
Proceed to reboot the system? [confirm yes/no]:y
```

When complete, the switch reboots to the factory default configuration.

Note: By default, OS10EE has Telnet disabled, SSH enabled, and the OOB management interface that is configured to get an IP address using DHCP. The default username and password are both `admin`. Dell EMC recommends changing the admin password to a complex password when logging in for the first time.

4 Configure and verify the underlay network

4.1 Configure the first leaf switch in AZ1

The following sections cover the configuration for S5248F-ON switch with the hostname `sfo01-Leaf01a`, shown as the left switch in Figure 19. Virtual networks 1611 and 3939 are shown in the diagram as an example, all the required virtual networks are created during the switch configuration. While the remaining leaf switch configurations are not shown in the document, all switch configuration commands are provided in the file attachments. See Section 1.8 for instructions on accessing the attachments.

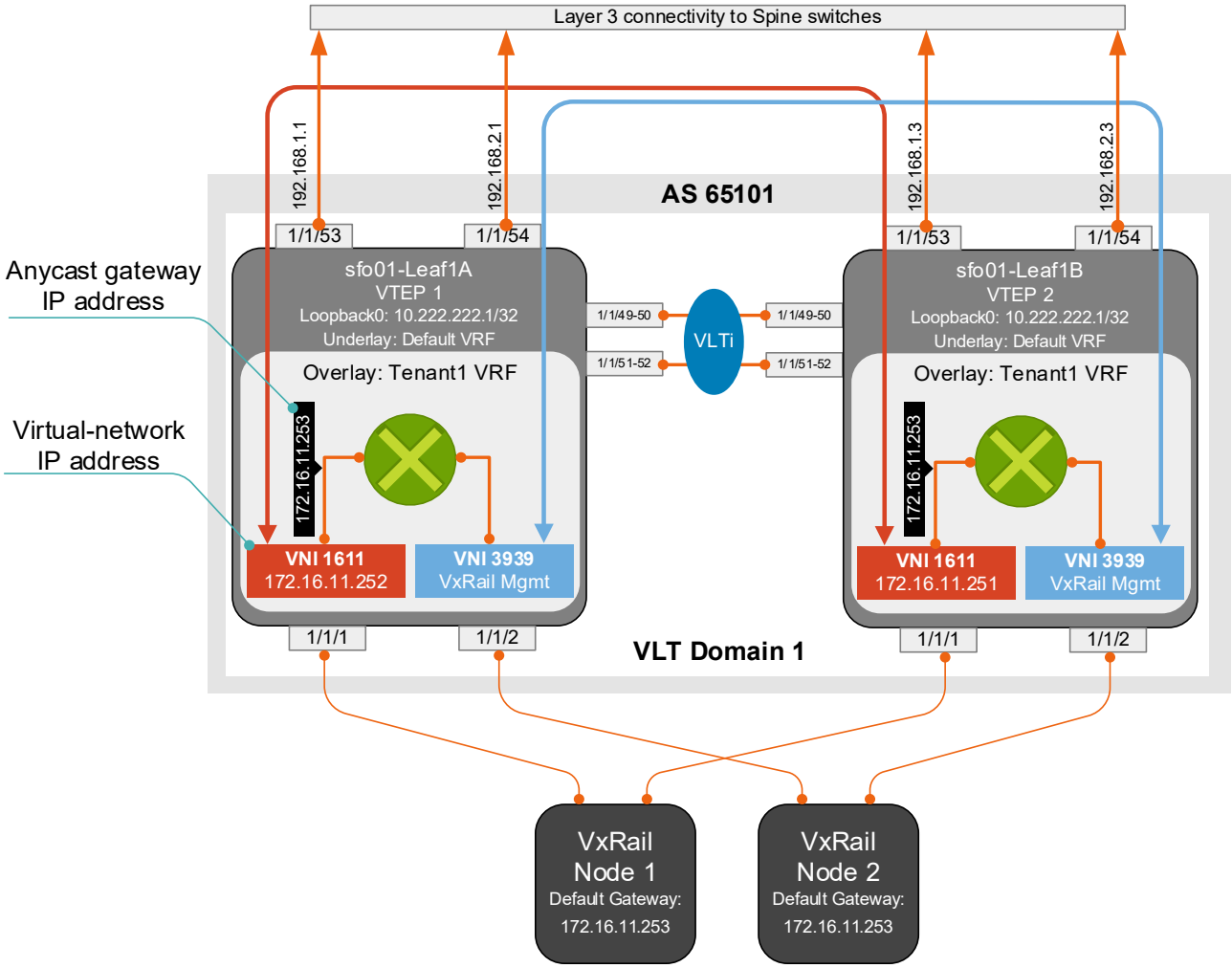


Figure 19 Region A, AZ1, Rack 1, leaf switch diagram

Note: In this example, private AS 2-byte values are used, and these should be changed to reflect the AS values used in the environment.

1. Configure general switch settings including management and NTP source.

```
OS10# configure terminal
OS10(config)# interface mgmt 1/1/1
OS10(conf-if-ma-1/1/1)# no ip address dhcp
OS10(conf-if-ma-1/1/1)# ip address 100.67.198.32/24
OS10(conf-if-ma-1/1/1)# exit
OS10(config)# management route 100.67.0.0/16 managementethernet
OS10(config)# hostname sfo01-Leaf01A
sfo01-Leaf01A(config)# no multicast snooping flood-restrict
sfo01-Leaf01A(config)# ntp server 100.67.10.20
sfo01-Leaf01A(config)# bfd enable
```

2. Configure a loopback interface for the Router ID

```
sfo01-Leaf01A(config)# interface loopback 0
sfo01-Leaf01A(conf-if-lo-0)# description Router-ID
sfo01-Leaf01A(conf-if-lo-0)# no shutdown
sfo01-Leaf01A(conf-if-lo-0)# ip address 10.0.2.1/32
sfo01-Leaf01A(conf-if-lo-0)# exit
```

3. Configure a loopback interface for NVE

```
sfo01-Leaf01A(config)# interface loopback 2
sfo01-Leaf01A(conf-if-lo-2)# description nve_loopback
sfo01-Leaf01A(conf-if-lo-2)# no shutdown
sfo01-Leaf01A(conf-if-lo-2)# ip address 10.222.222.1/32
sfo01-Leaf01A(conf-if-lo-2)# exit
```

4. Configure the loopback interface for the VXLAN source tunnel interface

```
sfo01-Leaf01A(config)# nve
sfo01-Leaf01A(config-nve)# source-interface loopback2
sfo01-Leaf01A(config-nve)# exit
```

5. Configure VXLAN virtual networks

```
sfo01-Leaf01A(config)# virtual-network 1611
sfo01-Leaf01A(config-vn)# vxlan-vni 1611
sfo01-Leaf01A(config-vn)# exit
sfo01-Leaf01A(config)# virtual-network 1612
sfo01-Leaf01A(config-vn)# vxlan-vni 1612
sfo01-Leaf01A(config-vn)# exit
sfo01-Leaf01A(config)# virtual-network 1613
sfo01-Leaf01A(config-vn)# vxlan-vni 1613
sfo01-Leaf01A(config-vn)# exit
sfo01-Leaf01A(config)# virtual-network 3939
sfo01-Leaf01A(config-vn)# vxlan-vni 3939
sfo01-Leaf01A(config-vn)# exit
```

6. Assign VLAN member interfaces to virtual networks

```
sfo01-Leaf01A(config)# interface vlan1611
sfo01-Leaf01A(config-if-vl-1611)# description sfo-mgmt
sfo01-Leaf01A(config-if-vl-1611)# virtual-network 1611
sfo01-Leaf01A(config-if-vl-1611)# no shutdown
sfo01-Leaf01A(config-if-vl-1611)# mtu 9216
sfo01-Leaf01A(config-if-vl-1611)# exit
sfo01-Leaf01A(config)# interface vlan1612
sfo01-Leaf01A(config-if-vl-1612)# virtual-network 1612
sfo01-Leaf01A(config-if-vl-1612)# description sfo-vmotion
sfo01-Leaf01A(config-if-vl-1612)# no shutdown
sfo01-Leaf01A(config-if-vl-1612)# mtu 9216
sfo01-Leaf01A(config-if-vl-1612)# exit
sfo01-Leaf01A(config)# interface vlan1613
sfo01-Leaf01A(config-if-vl-1613)# virtual-network 1613
sfo01-Leaf01A(config-if-vl-1613)# description sfo-vsan
sfo01-Leaf01A(config-if-vl-1613)# no shutdown
sfo01-Leaf01A(config-if-vl-1613)# mtu 9216
sfo01-Leaf01A(config-if-vl-1613)# exit
sfo01-Leaf01A(config)# interface vlan3939
sfo01-Leaf01A(config-if-vl-3939)# description vxrail-discovery
sfo01-Leaf01A(config-if-vl-3939)# virtual-network 3939
sfo01-Leaf01A(config-if-vl-3939)# no shutdown
sfo01-Leaf01A(config-if-vl-3939)# mtu 9216
sfo01-Leaf01A(config-if-vl-3939)# exit
```

7. Configure access ports as VLAN members for a switch-scoped VLAN-to-VNI mapping. VLAN 1611 is untagged to accommodate ESXi management, configured by default as untagged on the management VMkernel. The remaining VLANs are tagged to support the tagging from the default VxRail created vDS.

```
sfo01-Leaf01A(config)# interface ethernet1/1/1
sfo01-Leaf01A(conf-if-eth1/1/1)# description sfo01m01vxrail01
sfo01-Leaf01A(conf-if-eth1/1/1)# no shutdown
sfo01-Leaf01A(conf-if-eth1/1/1)# switchport mode trunk
sfo01-Leaf01A(conf-if-eth1/1/1)# switchport access vlan 1611
sfo01-Leaf01A(conf-if-eth1/1/1)# switchport trunk allowed vlan 1612-
1613,3939
sfo01-Leaf01A(conf-if-eth1/1/1)# mtu 9216
sfo01-Leaf01A(conf-if-eth1/1/1)# spanning-tree port type edge
sfo01-Leaf01A(conf-if-eth1/1/1)# flowcontrol receive on
sfo01-Leaf01A(conf-if-eth1/1/1)# flowcontrol transmit off
sfo01-Leaf01A(conf-if-eth1/1/1)# exit
sfo01-Leaf01A(config)# interface ethernet1/1/2
sfo01-Leaf01A(conf-if-eth1/1/2)# description sfo01m01vxrail02
sfo01-Leaf01A(conf-if-eth1/1/2)# no shutdown
sfo01-Leaf01A(conf-if-eth1/1/2)# switchport mode trunk
sfo01-Leaf01A(conf-if-eth1/1/2)# switchport access vlan 1611
sfo01-Leaf01A(conf-if-eth1/1/2)# switchport trunk allowed vlan 1612-
1613,3939
sfo01-Leaf01A(conf-if-eth1/1/2)# mtu 9216
```

```
sfo01-Leaf01A(conf-if-eth1/1/2)# spanning-tree port type edge
sfo01-Leaf01A(conf-if-eth1/1/2)# flowcontrol receive on
sfo01-Leaf01A(conf-if-eth1/1/2)# flowcontrol transmit off
sfo01-Leaf01A(conf-if-eth1/1/2)# exit
```

8. Configure upstream network-facing ports

```
sfo01-Leaf01A(config)# interface ethernet1/1/53
sfo01-Leaf01A(conf-if-eth1/1/53)# description sfo01-spine01
sfo01-Leaf01A(conf-if-eth1/1/53)# no shutdown
sfo01-Leaf01A(conf-if-eth1/1/53)# no switchport
sfo01-Leaf01A(conf-if-eth1/1/53)# mtu 9216
sfo01-Leaf01A(conf-if-eth1/1/53)# ip address 192.168.1.1/31
sfo01-Leaf01A(conf-if-eth1/1/53)# exit
sfo01-Leaf01A(config)# interface ethernet1/1/54
sfo01-Leaf01A(conf-if-eth1/1/54)# description sfo01-spine02
sfo01-Leaf01A(conf-if-eth1/1/54)# no shutdown
sfo01-Leaf01A(conf-if-eth1/1/54)# no switchport
sfo01-Leaf01A(conf-if-eth1/1/54)# mtu 9216
sfo01-Leaf01A(conf-if-eth1/1/54)# ip address 192.168.2.1/31
sfo01-Leaf01A(conf-if-eth1/1/54)# exit
```

9. Add a route map. This is an example route map is used to illustrate how to allow IP traffic to be passed on the switch.

```
sfo01-Leaf01A(config)# ip prefix-list spine-leaf seq 10 permit 10.0.2.0/24
ge 32
sfo01-Leaf01A(config)# ip prefix-list spine-leaf seq 20 permit 10.2.2.0/24
ge 32
sfo01-Leaf01A(config)# ip prefix-list spine-leaf seq 30 permit
10.222.222.0/24 ge 32
sfo01-Leaf01A(config)# ip prefix-list spine-leaf seq 40 permit
172.20.0.0/16
sfo01-Leaf01A(config)# route-map spine-leaf permit 10
sfo01-Leaf01A(config-route-map)# match ip address prefix-list spine-leaf
sfo01-Leaf01A(config-route-map)# exit
```

Note: See Section 7 for information on ESG configuration.

10. Configure eBGP

```
sfo01-Leaf01A(config)# router bgp 65101
sfo01-Leaf01A(config-router-bgp-65101)# router-id 10.0.2.1
sfo01-Leaf01A(config-router-bgp-65101)# bfd all-neighbors interval 200
min_rx 200 multiplier 3 role active
sfo01-Leaf01A(config-router-bgp-65101)# address-family ipv4 unicast
sfo01-Leaf01A(config-router-bgpv4-af)# redistribute connected route-map
spine-leaf
sfo01-Leaf01A(config-router-bgpv4-af)# exit
sfo01-Leaf01A(config-router-bgp-65101)# bestpath as-path multipath-relax
sfo01-Leaf01A(config-router-bgp-65101)# maximum-paths ebgp 2
```

Note: If more than two ESGs are being used update the `maximum-paths ebgp` value accordingly.

11. Configure eBGP for the IPv4 point-to-point peering

```
sfo01-Leaf01A(config-router-bgp-65101)# neighbor 192.168.1.0
sfo01-Leaf01A(config-router-neighbor)# advertisement-interval 5
sfo01-Leaf01A(config-router-neighbor)# bfd
sfo01-Leaf01A(config-router-neighbor)# fall-over
sfo01-Leaf01A(config-router-neighbor)# remote-as 65100
sfo01-Leaf01A(config-router-neighbor)# no shutdown
sfo01-Leaf01A(config-router-neighbor)# address-family ipv4 unicast
sfo01-Leaf01A(config-router-neighbor-af)# exit
sfo01-Leaf01A(config-router-neighbor)# exit
sfo01-Leaf01A(config-router-bgp-65101)# neighbor 192.168.2.0
sfo01-Leaf01A(config-router-neighbor)# advertisement-interval 5
sfo01-Leaf01A(config-router-neighbor)# bfd
sfo01-Leaf01A(config-router-neighbor)# fall-over
sfo01-Leaf01A(config-router-neighbor)# remote-as 65100
sfo01-Leaf01A(config-router-neighbor)# no shutdown
sfo01-Leaf01A(config-router-neighbor)# address-family ipv4 unicast
sfo01-Leaf01A(config-router-neighbor-af)# exit
sfo01-Leaf01A(config-router-neighbor)# exit
sfo01-Leaf01A(config-router-bgp-65101)# exit
```

12. Configure a loopback interface for BGP EVPN peering

```
sfo01-Leaf01A(config)# interface loopback 1
sfo01-Leaf01A(conf-if-lo-1)# description evpn_loopback
sfo01-Leaf01A(conf-if-lo-1)# no shutdown
sfo01-Leaf01A(conf-if-lo-1)# ip address 10.2.2.1/32
sfo01-Leaf01A(conf-if-lo-1)# exit
```

13. Configure BGP EVPN peering

```
sfo01-Leaf01A(config)# router bgp 65101
sfo01-Leaf01A(config-router-bgp-65101)# neighbor 10.2.1.1
sfo01-Leaf01A(config-router-neighbor)# remote-as 65100
sfo01-Leaf01A(config-router-neighbor)# ebgp-multihop 2
sfo01-Leaf01A(config-router-neighbor)# send-community extended
```

```
sfo01-Leaf01A(config-router-neighbor)# update-source loopback1
sfo01-Leaf01A(config-router-neighbor)# no shutdown
sfo01-Leaf01A(config-router-neighbor)# address-family ipv4 unicast
sfo01-Leaf01A(config-router-neighbor-af)# no activate
sfo01-Leaf01A(config-router-neighbor-af)# exit
sfo01-Leaf01A(config-router-neighbor)# address-family l2vpn evpn
sfo01-Leaf01A(config-router-neighbor-af)# activate
sfo01-Leaf01A(config-router-neighbor-af)# exit
sfo01-Leaf01A(config-router-neighbor)# exit
sfo01-Leaf01A(config-router-bgp-65101)# neighbor 10.2.1.2
sfo01-Leaf01A(config-router-neighbor)# remote-as 65100
sfo01-Leaf01A(config-router-neighbor)# ebgp-multihop 2
sfo01-Leaf01A(config-router-neighbor)# send-community extended
sfo01-Leaf01A(config-router-neighbor)# update-source loopback1
sfo01-Leaf01A(config-router-neighbor)# no shutdown
sfo01-Leaf01A(config-router-neighbor)# address-family ipv4 unicast
sfo01-Leaf01A(config-router-neighbor-af)# no activate
sfo01-Leaf01A(config-router-neighbor-af)# exit
sfo01-Leaf01A(config-router-neighbor)# address-family l2vpn evpn
sfo01-Leaf01A(config-router-neighbor-af)# activate
sfo01-Leaf01A(config-router-neighbor-af)# exit
sfo01-Leaf01A(config-router-neighbor)# exit
sfo01-Leaf01A(config-router-bgp-65101)# exit
```

14. Configure EVPN

```
sfo01-Leaf01A(config)# evpn
sfo01-Leaf01A(config-evpn)# evi 1611
sfo01-Leaf01A(config-evpn-evi-1611)# vni 1611
sfo01-Leaf01A(config-evpn-evi-1611)# rd 10.222.222.1:1611
sfo01-Leaf01A(config-evpn-evi-1611)# route-target 1611:1611 both
sfo01-Leaf01A(config-evpn-evi-1611)# exit
sfo01-Leaf01A(config-evpn)# evi 1612
sfo01-Leaf01A(config-evpn-evi-1612)# vni 1612
sfo01-Leaf01A(config-evpn-evi-1612)# rd 10.222.222.1:1612
sfo01-Leaf01A(config-evpn-evi-1612)# route-target 1612:1612 both
sfo01-Leaf01A(config-evpn-evi-1612)# exit
sfo01-Leaf01A(config-evpn)# evi 1613
sfo01-Leaf01A(config-evpn-evi-1613)# vni 1613
sfo01-Leaf01A(config-evpn-evi-1613)# rd 10.222.222.1:1613
sfo01-Leaf01A(config-evpn-evi-1613)# route-target 1613:1613 both
sfo01-Leaf01A(config-evpn-evi-1613)# exit
```

15. Configure dedicated L3 underlay path to reach VLT peer in case of network failure

```
sfo01-Leaf01A(config)# interface vlan4000
sfo01-Leaf01A(config-if-vl-4000)# no shutdown
sfo01-Leaf01A(config-if-vl-4000)# mtu 9216
sfo01-Leaf01A(config-if-vl-4000)# ip address 192.168.3.0/31
sfo01-Leaf01A(config-if-vl-4000)# exit
```

16. Configure VLTi member links

```
sfo01-Leaf01A(config)# interface range ethernet1/1/49-1/1/52
sfo01-Leaf01A(conf-range-eth1/1/49-1/1/52)# description VLTi
sfo01-Leaf01A(conf-range-eth1/1/49-1/1/52)# no shutdown
sfo01-Leaf01A(conf-range-eth1/1/49-1/1/52)# no switchport
sfo01-Leaf01A(conf-range-eth1/1/49-1/1/52)# exit
```

17. Configure the VLT domain

```
sfo01-Leaf01A(config)# vlt-domain 1
sfo01-Leaf01A(conf-vlt-1)# backup destination 100.67.198.31
sfo01-Leaf01A(conf-vlt-1)# discovery-interface ethernet1/1/49-1/1/52
sfo01-Leaf01A(conf-vlt-1)# peer-routing
sfo01-Leaf01A(conf-vlt-1)# vlt-mac 00:00:01:02:03:01
sfo01-Leaf01A(conf-vlt-1)# exit
```

18. Configure UFD with uplink VLT ports and downlink network ports

```
sfo01-Leaf01A(config)# uplink-state-group 1
sfo01-Leaf01A(conf-uplink-state-group-1)# enable
sfo01-Leaf01A(conf-uplink-state-group-1)# downstream ethernet1/1/1-1/1/8
sfo01-Leaf01A(conf-uplink-state-group-1)# upstream ethernet1/1/53
sfo01-Leaf01A(conf-uplink-state-group-1)# upstream ethernet1/1/54
sfo01-Leaf01A(conf-uplink-state-group-1)# exit
```

19. Configure iBGP IPv4 peering between VLT peers

```
sfo01-Leaf01A(config)# router bgp 65101
sfo01-Leaf01A(config-router-bgp-65101)# neighbor 192.168.3.1
sfo01-Leaf01A(config-router-neighbor)# remote-as 65101
sfo01-Leaf01A(config-router-neighbor)# no shutdown
sfo01-Leaf01A(config-router-neighbor)# exit
```

20. Create a tenant VRF. An OS10 best practice is to isolate any virtual network traffic in a non-default VRF

```
sfo01-Leaf01A(config)# ip vrf tenant1
sfo01-Leaf01A(conf-vrf)# exit
```

21. Configure the anycast gateway MAC address

```
sfo01-Leaf01A(config)# ip virtual-router mac-address 00:01:01:01:01:01
```

22. Configure routing on virtual networks

```
sfo01-Leaf01A(config)# interface virtual-network1611
sfo01-Leaf01A(conf-if-vn-1611)# no shutdown
sfo01-Leaf01A(conf-if-vn-1611)# mtu 9216
sfo01-Leaf01A(conf-if-vn-1611)# ip vrf forwarding tenant1
sfo01-Leaf01A(conf-if-vn-1611)# ip address 172.16.11.252/24
sfo01-Leaf01A(conf-if-vn-1611)# ip virtual-router address 172.16.11.253
sfo01-Leaf01A(conf-if-vn-1611)# exit
sfo01-Leaf01A(config)# interface virtual-network1612
sfo01-Leaf01A(conf-if-vn-1612)# no shutdown
sfo01-Leaf01A(conf-if-vn-1612)# mtu 9216
```

```
sfo01-Leaf01A(conf-if-vn-1612)# ip vrf forwarding tenant1
sfo01-Leaf01A(conf-if-vn-1612)# ip address 172.16.12.252/24
sfo01-Leaf01A(conf-if-vn-1612)# ip virtual-router address 172.16.12.253
sfo01-Leaf01A(conf-if-vn-1612)# exit
sfo01-Leaf01A(config)# interface virtual-network1613
sfo01-Leaf01A(conf-if-vn-1613)# no shutdown
sfo01-Leaf01A(conf-if-vn-1613)# mtu 9216
sfo01-Leaf01A(conf-if-vn-1613)# ip vrf forwarding tenant1
sfo01-Leaf01A(conf-if-vn-1613)# ip address 172.16.13.252/24
sfo01-Leaf01A(conf-if-vn-1613)# ip virtual-router address 172.16.13.253
sfo01-Leaf01A(conf-if-vn-1613)# exit
sfo01-Leaf01A(config)# interface virtual-network3939
sfo01-Leaf01A(conf-if-vn-3939)# no shutdown
sfo01-Leaf01A(conf-if-vn-3939)# ip vrf forwarding tenant1
sfo01-Leaf01A(conf-if-vn-3939)# exit
```

23. Repeat these, using the appropriate values from Section 3.3, for the remaining leaf switches in AZ1.

4.2 Configure the first spine switch in AZ1

This section covers the configuration of the Z9264-ON switch with the hostname `sfo01-spine01` shown in Figure 20.

Note: All switch configuration commands are provided in the file attachments. See Section 1.8 for instructions on accessing the attachments.

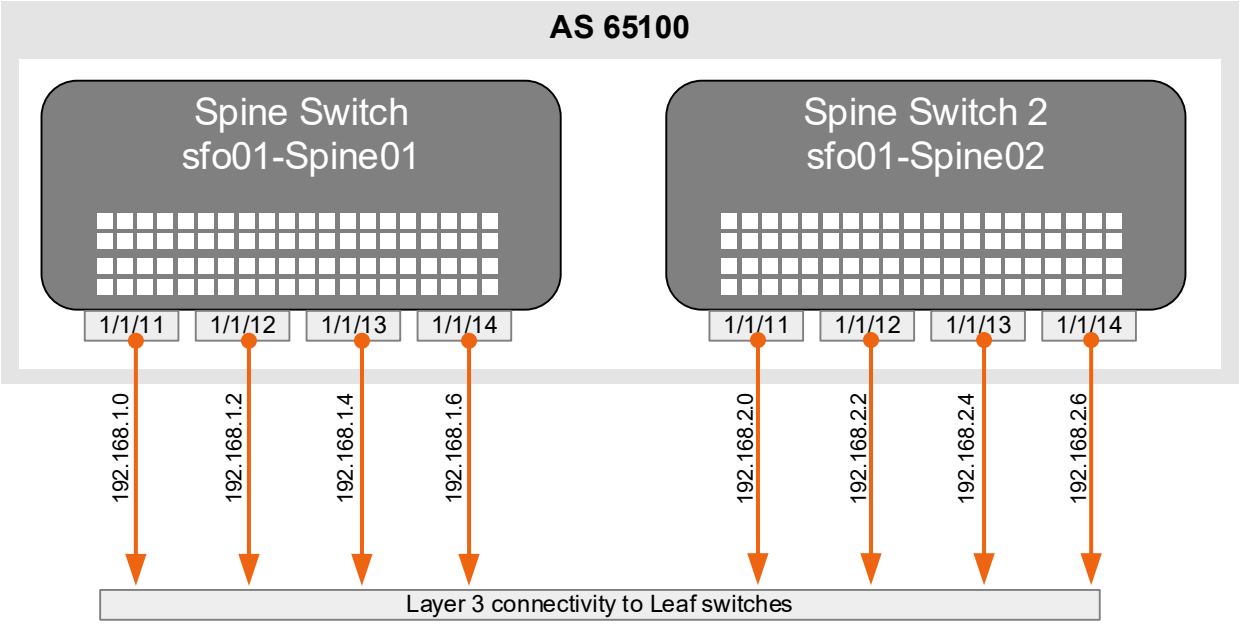


Figure 20 Region A, AZ1, spine layer diagram

1. Configure general switch settings including management and NTP source.

```
OS10# configure terminal
OS10(config)# interface mgmt 1/1/1
OS10(conf-if-ma-1/1/1)# no ip address dhcp
OS10(conf-if-ma-1/1/1)# ip address 100.67.198.36/24
OS10(conf-if-ma-1/1/1)# exit
OS10(config)# management route 100.67.0.0/16 managementethernet
OS10(config)# hostname sfo01-Spine01
sfo01-Spine01(config)# ntp server 100.67.10.20
sfo01-Spine01(config)# hardware forwarding-table mode scaled-l3-routes
sfo01-Spine01(config)# bfd enable
```

2. Configure a loopback interface for the Router ID

```
sfo01-Spine01(config)# interface loopback 0
sfo01-Spine01(conf-if-lo-0)# description Router-ID
sfo01-Spine01(conf-if-lo-0)# no shutdown
sfo01-Spine01(conf-if-lo-0)# ip address 10.0.1.1/32
sfo01-Spine01(conf-if-lo-0)# exit
```

3. Configure downstream ports on underlay links to leaf switches

```
sfo01-Spine01(config)# interface ethernet1/1/11
sfo01-Spine01(conf-if-eth1/1/11)# description sfo01-Leaf01A
sfo01-Spine01(conf-if-eth1/1/11)# no shutdown
sfo01-Spine01(conf-if-eth1/1/11)# no switchport
sfo01-Spine01(conf-if-eth1/1/11)# mtu 9216
sfo01-Spine01(conf-if-eth1/1/11)# ip address 192.168.1.0/31
sfo01-Spine01(conf-if-eth1/1/11)# exit
sfo01-Spine01(config)# interface ethernet1/1/12
sfo01-Spine01(conf-if-eth1/1/12)# description sfo01-Leaf01B
sfo01-Spine01(conf-if-eth1/1/12)# no shutdown
sfo01-Spine01(conf-if-eth1/1/12)# no switchport
sfo01-Spine01(conf-if-eth1/1/12)# mtu 9216
sfo01-Spine01(conf-if-eth1/1/12)# ip address 192.168.1.2/31
sfo01-Spine01(conf-if-eth1/1/12)# exit
sfo01-Spine01(config)# interface ethernet1/1/13
sfo01-Spine01(conf-if-eth1/1/13)# description sfo01-Leaf02A
sfo01-Spine01(conf-if-eth1/1/13)# no shutdown
sfo01-Spine01(conf-if-eth1/1/13)# no switchport
sfo01-Spine01(conf-if-eth1/1/13)# mtu 9216
sfo01-Spine01(conf-if-eth1/1/13)# ip address 192.168.1.4/31
sfo01-Spine01(conf-if-eth1/1/13)# exit
sfo01-Spine01(config)# interface ethernet1/1/14
sfo01-Spine01(conf-if-eth1/1/14)# description sfo01-Leaf02B
sfo01-Spine01(conf-if-eth1/1/14)# no shutdown
sfo01-Spine01(conf-if-eth1/1/14)# no switchport
sfo01-Spine01(conf-if-eth1/1/14)# mtu 9216
sfo01-Spine01(conf-if-eth1/1/14)# ip address 192.168.1.6/31
sfo01-Spine01(conf-if-eth1/1/14)# exit
```

4. Add a route map

```
sfo01-Spine01(config)# ip prefix-list spine-leaf seq 10 permit 10.0.1.0/24
ge 32
sfo01-Spine01(config)# ip prefix-list spine-leaf seq 20 permit 10.2.1.0/24
ge 32
sfo01-Spine01(config)# route-map spine-leaf permit 10
sfo01-Spine01(config-route-map)# match ip address prefix-list spine-leaf
sfo01-Spine01(config-route-map)# exit
```

5. Configure eBGP

```
sfo01-Spine01(config)# router bgp 65100
sfo01-Spine01(config-router-bgp-65100)# bfd all-neighbors interval 200
min_rx 200 multiplier 3 role active
sfo01-Spine01(config-router-bgp-65100)# router-id 10.0.1.1
sfo01-Spine01(config-router-bgp-65100)# address-family ipv4 unicast
sfo01-Spine01(config-router-bgpv4-af)# redistribute connected route-map
spine-leaf
sfo01-Spine01(config-router-bgp-65100)# bestpath as-path multipath-relax
sfo01-Spine01(config-router-bgp-65100)# maximum-paths ebgp 2
```

```
sfo01-Spine01(config-router-bgpv4-af)# exit
```

6. Configure eBGP for IPv4 point-to-point peering

```
sfo01-Spine01(config-router-bgp-65100)# neighbor 192.168.1.1
sfo01-Spine01(config-router-neighbor)# remote-as 65101
sfo01-Spine01(config-router-neighbor)# no shutdown
sfo01-Spine01(config-router-neighbor)# advertisement-interval 5
sfo01-Spine01(config-router-neighbor)# bfd
sfo01-Spine01(config-router-neighbor)# fall-over
sfo01-Spine01(config-router-neighbor)# exit
sfo01-Spine01(config-router-bgp-65100)# neighbor 192.168.1.3
sfo01-Spine01(config-router-neighbor)# remote-as 65101
sfo01-Spine01(config-router-neighbor)# no shutdown
sfo01-Spine01(config-router-neighbor)# advertisement-interval 5
sfo01-Spine01(config-router-neighbor)# bfd
sfo01-Spine01(config-router-neighbor)# fall-over
sfo01-Spine01(config-router-neighbor)# exit
sfo01-Spine01(config-router-bgp-65100)# neighbor 192.168.1.5
sfo01-Spine01(config-router-neighbor)# remote-as 65102
sfo01-Spine01(config-router-neighbor)# no shutdown
sfo01-Spine01(config-router-neighbor)# advertisement-interval 5
sfo01-Spine01(config-router-neighbor)# bfd
sfo01-Spine01(config-router-neighbor)# fall-over
sfo01-Spine01(config-router-neighbor)# exit
sfo01-Spine01(config-router-bgp-65100)# neighbor 192.168.1.7
sfo01-Spine01(config-router-neighbor)# remote-as 65102
sfo01-Spine01(config-router-neighbor)# no shutdown
sfo01-Spine01(config-router-neighbor)# advertisement-interval 5
sfo01-Spine01(config-router-neighbor)# bfd
sfo01-Spine01(config-router-neighbor)# fall-over
sfo01-Spine01(config-router-neighbor)# exit
sfo01-Spine01(config-router-bgp-65100)# exit
```

7. Configure a loopback interface for BGP EVPN peering

```
sfo01-Spine01(config)# interface loopback 1
sfo01-Spine01(conf-if-lo-1)# description evpn_loopback
sfo01-Spine01(conf-if-lo-1)# no shutdown
sfo01-Spine01(conf-if-lo-1)# ip address 10.2.1.1/32
sfo01-Spine01(conf-if-lo-1)# exit
```

8. Configure BGP EVPN peering

```
sfo01-Spine01(config)# router bgp 65100
sfo01-Spine01(config-router-bgp-65100)# neighbor 10.2.2.1
sfo01-Spine01(config-router-neighbor)# remote-as 65101
sfo01-Spine01(config-router-neighbor)# send-community extended
sfo01-Spine01(config-router-neighbor)# update-source loopback1
sfo01-Spine01(config-router-neighbor)# ebgp-multihop 2
sfo01-Spine01(config-router-neighbor)# no shutdown
sfo01-Spine01(config-router-neighbor)# address-family ipv4 unicast
sfo01-Spine01(config-router-neighbor-af)# no activate
sfo01-Spine01(config-router-neighbor-af)# exit
sfo01-Spine01(config-router-neighbor)# address-family l2vpn evpn
sfo01-Spine01(config-router-neighbor-af)# activate
sfo01-Spine01(config-router-neighbor-af)# exit
sfo01-Spine01(config-router-neighbor)# exit
sfo01-Spine01(config-router-bgp-65100)# neighbor 10.2.2.2
sfo01-Spine01(config-router-neighbor)# remote-as 65101
sfo01-Spine01(config-router-neighbor)# send-community extended
sfo01-Spine01(config-router-neighbor)# update-source loopback1
sfo01-Spine01(config-router-neighbor)# ebgp-multihop 2
sfo01-Spine01(config-router-neighbor)# no shutdown
sfo01-Spine01(config-router-neighbor)# address-family ipv4 unicast
sfo01-Spine01(config-router-neighbor-af)# no activate
sfo01-Spine01(config-router-neighbor-af)# exit
sfo01-Spine01(config-router-neighbor)# address-family l2vpn evpn
sfo01-Spine01(config-router-neighbor-af)# activate
sfo01-Spine01(config-router-neighbor-af)# exit
sfo01-Spine01(config-router-neighbor)# exit
sfo01-Spine01(config-router-bgp-65100)# neighbor 10.2.2.3
sfo01-Spine01(config-router-neighbor)# remote-as 65102
sfo01-Spine01(config-router-neighbor)# send-community extended
sfo01-Spine01(config-router-neighbor)# update-source loopback1
sfo01-Spine01(config-router-neighbor)# ebgp-multihop 2
sfo01-Spine01(config-router-neighbor)# no shutdown
sfo01-Spine01(config-router-neighbor)# address-family ipv4 unicast
sfo01-Spine01(config-router-neighbor-af)# no activate
sfo01-Spine01(config-router-neighbor-af)# exit
sfo01-Spine01(config-router-neighbor)# address-family l2vpn evpn
sfo01-Spine01(config-router-neighbor-af)# activate
sfo01-Spine01(config-router-neighbor-af)# exit
sfo01-Spine01(config-router-neighbor)# exit
sfo01-Spine01(config-router-bgp-65100)# neighbor 10.2.2.4
sfo01-Spine01(config-router-neighbor)# remote-as 65102
sfo01-Spine01(config-router-neighbor)# send-community extended
sfo01-Spine01(config-router-neighbor)# update-source loopback1
sfo01-Spine01(config-router-neighbor)# ebgp-multihop 2
sfo01-Spine01(config-router-neighbor)# no shutdown
sfo01-Spine01(config-router-neighbor)# address-family ipv4 unicast
```

```
sfo01-Spine01(config-router-neighbor-af)# no activate
sfo01-Spine01(config-router-neighbor-af)# exit
sfo01-Spine01(config-router-neighbor)# address-family l2vpn evpn
sfo01-Spine01(config-router-neighbor-af)# activate
sfo01-Spine01(config-router-neighbor-af)# exit
sfo01-Spine01(config-router-neighbor)# exit
sfo01-Spine01(config-router-bgp-65100)# exit
```

9. Repeat these, using the appropriate values from Section 3.3, for the remaining spine switches in AZ1.

4.3 Verify establishment of BGP between leaf and spine switches

The leaf switches must establish a connection to the spine switches before BGP updates can be exchanged. Verify that peering is successful and BGP routing has been established.

1. Run the `show ip bgp summary` command to display information about the BGP and TCP connection to neighbors. In Figure 21, all three BGP sessions for each leaf switch are shown. The last sessions, 192.168.3.1 is the iBGP sessions between the leaf pairs in the event of leaf to spine layer failure.

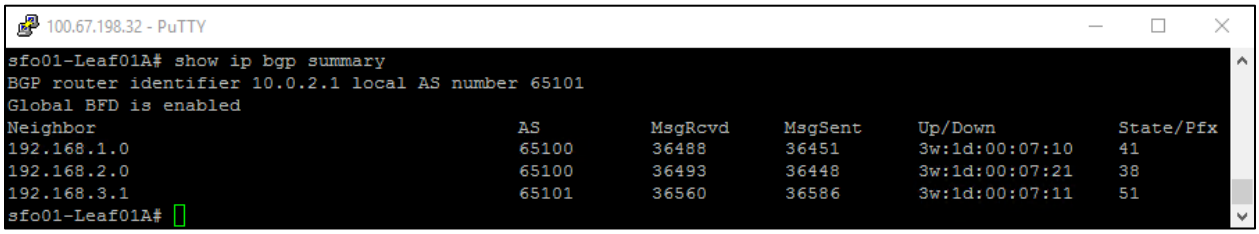


Figure 21 Output of show ip bgp summary

2. Run the `show ip route bgp` command to verify that all routes using BGP are being received, and that there are multiple routes to the BGP learned networks as indicated by the multiple gateway entries. Figure 22 shows two different routes to the remote loopback addresses for 10.0.2.3/32 and 10.2.2.3/32.

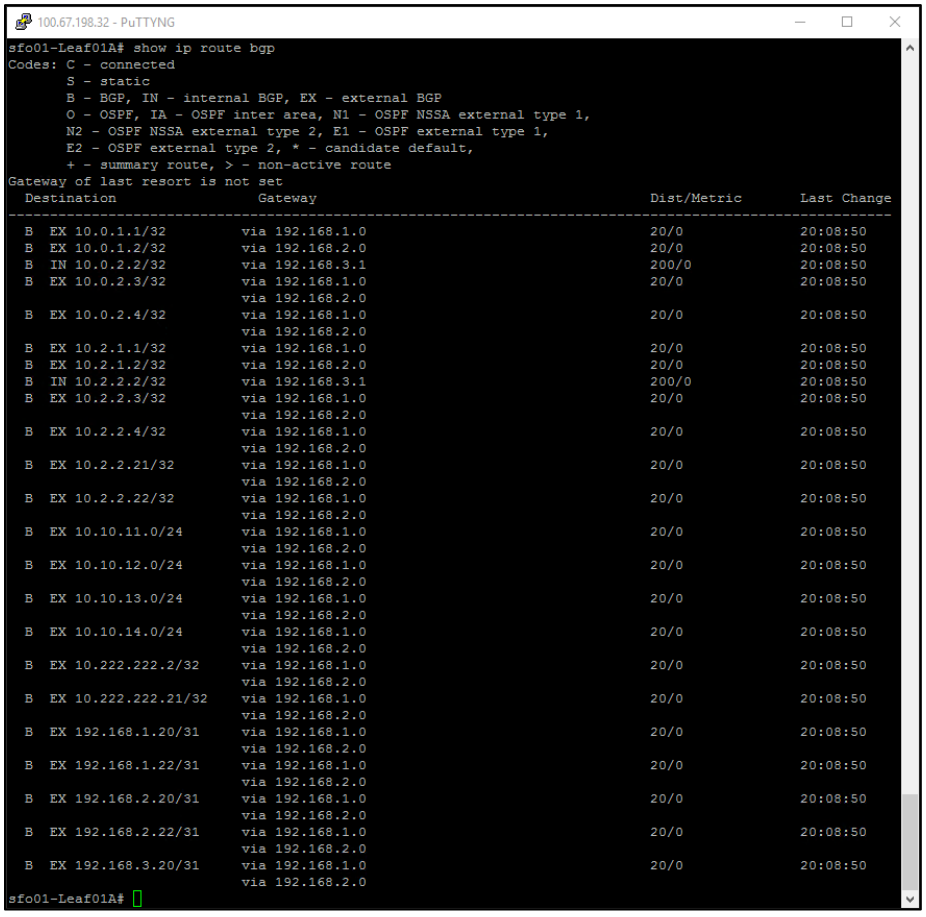


Figure 22 Output of show ip route

4.4 Verify BGP EVPN and VXLAN between leaf switches

For the L2 VXLAN virtual networks to communicate, each leaf must be able to establish a connection to the other leaf switches before host MAC information can be exchanged. Verify that peering is successful and BGP EVPN routing has been established.

1. Run the `show ip bgp l2vpn evpn summary` command to display information about the BGP EVPN and TCP connections to neighbors. Figure 23 shows the BGP states between leaf switch sfo01-Leaf01A and sfo01-spine01 (10.2.1.1) and sfo01-spine02 (10.2.1.2).

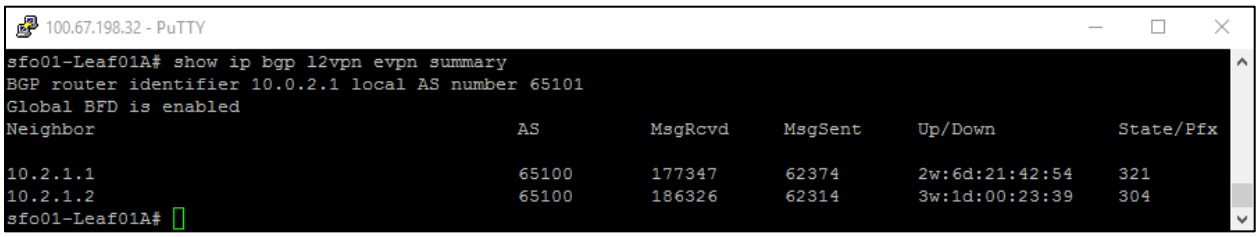


Figure 23 Output of show ip bgp l2vpn evpn neighbors

2. Run the `show evpn evi` command to verify the current state of all configured virtual networks. Figure 24 shows the state of each virtual network as `Up` and that the Integrated Routing and Bridging (IRB) VRF is set as `tenant1`.

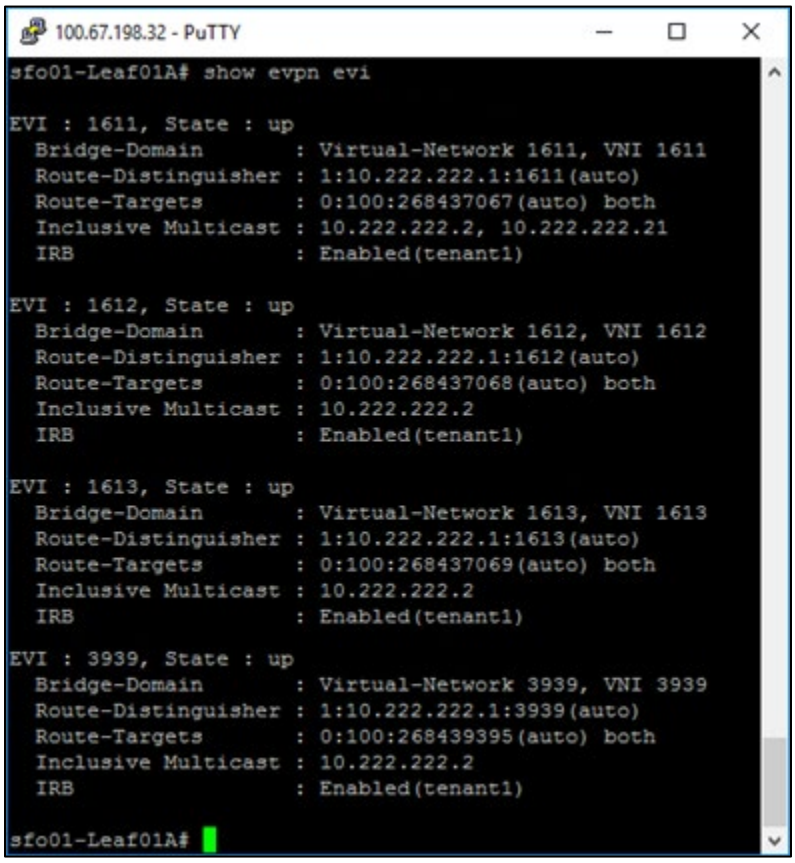


Figure 24 Output of show evpn evi

Note: For more validation and troubleshooting command, see the ***OS10 Enterprise Edition User Guide.***

5 Initialize the management VxRail cluster

This guide does not provide the detailed steps to initialize the VxRail cluster. The information provided lists the steps that are used to initialize the multirack cluster using four E-series VxRail nodes. This initialization creates the basis that the VCF Cloud Builder uses to create a management workload domain.

5.1 Configure laptop/workstation for VxRail initialization

A workstation/laptop with a web browser is used for the to perform the initialization process. It must be either plugged into one of the leaf switches, or able to logically reach the VxRail external management VLAN from elsewhere in the network. The following steps provide guidance on the configuration of a S5248F-ON switch to work with a 1 GbE connection from a workstation or laptop:

1. Enter the `show port-group` command to determine interface to port group mapping.

```
sfo01-Leaf01A# show port-group
Port-group              Mode            Ports           FEM
port-group1/1/1         Eth 25g-4x      1  2  3  4     -
port-group1/1/2         Eth 25g-4x      5  6  7  8     -
port-group1/1/3         Eth 25g-4x      9  10  11  12   -
port-group1/1/4         Eth 25g-4x      13  14  15  16  -
port-group1/1/5         Eth 25g-4x      17  18  19  20  -
port-group1/1/6         Eth 25g-4x      21  22  23  24  -
port-group1/1/7         Eth 25g-4x      25  26  27  28  -
port-group1/1/8         Eth 25g-4x      29  30  31  32  -
port-group1/1/9         Eth 25g-4x      33  34  35  36  -
port-group1/1/10        Eth 25g-4x      37  38  39  40  -
port-group1/1/11        Eth 25g-4x      41  42  43  44  -
port-group1/1/12        Eth 25g-4x      45  46  47  48  -
port-group1/1/13        Eth 100g-2x     49  50         -
port-group1/1/14        Eth 100g-2x     51  52         -
port-group1/1/15        Eth 100g-1x     53             -
port-group1/1/16        Eth 100g-1x     54             -
port-group1/1/17        Eth 100g-1x     55             -
port-group1/1/18        Eth 100g-1x     56             -
```

2. Configure the appropriate port group to 10 GbE. In this example port-group1/1/3 containing interface Ethernet 1/1/11 is used.

```
sfo01-Leaf01A# configure
sfo01-Leaf01A(config)# port-group 1/1/3
sfo01-Leaf01A(conf-pg-1/1/3)# mode Eth 10g-4x
sfo01-Leaf01A(conf-pg-1/1/3)# exit
```

3. Configure the breakout interface with the appropriate management VLAN.

```
sfo01-Leaf01A#(config)# interface ethernet 1/1/11:1
sfo01-Leaf01A#(conf-if-eth1/1/11:1)# switchport access vlan 1611
sfo01-Leaf01A#(conf-if-eth1/1/11:1)# description laptop
sfo01-Leaf01A#(conf-if-eth1/1/11:1)# end
```

4. Use the appropriate SFP-1GbE Base-T adapter, connect the workstation or laptop and verify connect speed using `show interface status`.

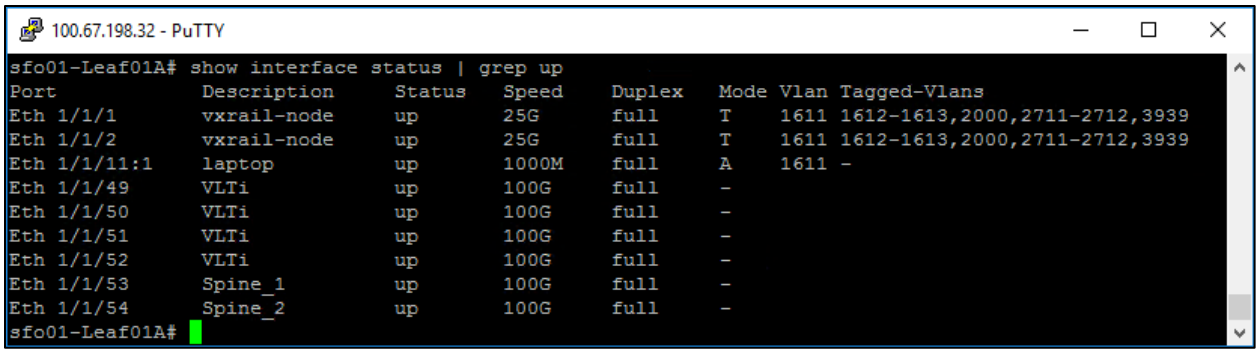


Figure 25 Output of show interface status confirming 1 GbE connectivity

5.2 VxRail initialization

This section outlines the general steps that are needed to initialize a VxRail cluster.

1. Install the VxRail nodes, by model, into the two racks in the data center.
2. Attach the appropriate cabling between the ports of the VxRail nodes and the switch ports.
3. Power on the four primary E-series nodes in both racks to form the initial VxRail cluster.
4. To access the VxRail external management on VLAN 1611, connect a workstation or laptop that is configured for VxRail.
5. Using a web browser, go to the default VxRail IP address to begin the VxRail initialization process.
6. Complete the steps provided within the initialization wizard.

Using the values provided, VxRail performs the verification process. Once the validation is complete, the initialization process builds a new VxRail cluster. The building progress of the cluster displays in the status window provided. When the **Hooray!** message displays, the VxRail initialization is complete and the new VxRail cluster is built.

5.3 VxRail deployment values

Table 7 lists the values that are used during the VxRail Manager initialization and expansion operation.

Note: The values are listed in order as they are entered in the GUI.

Table 7 VxRail network configuration values

	Parameter	Value
Appliance Settings	NTP server	172.16.11.5
	Domain	sfo01.rainpole.local
ESXi hostname and IP addresses	ESXi host name prefix	sfo01m01vxrail
	Separator	none
	Iterator	Num 0x
	Offset	1
	Suffix	none
	ESXi starting address	172.16.11.101
	ESXi ending address	172.16.11.104
vCenter Server	vCenter Server hostname (FQDN)	sfo01m01vc01.sfo01.rainpole.local
	admin username/password	administrator@vsphere.local

	vCenter Server SSO domain	vsphere.local
	IP address	172.16.11.62
VxRail Manager	VxRail Manager hostname	sfo01m01vxrail-mgr
	VxRail IP address	172.16.11.71
	Subnet mask	255.255.255.0
	Gateway	172.16.11.253
vMotion	Starting address for IP pool	172.16.12.101
	Ending address for IP pool	172.16.12.104
	Subnet mask	255.255.255.0
	VLAN ID	1612
vSAN	Starting address for IP pool	172.16.13.101
	Ending address for IP pool	172.16.13.104
	Subnet mask	255.255.255.0
	VLAN ID	1613
Solutions Logging	vRealize Log Insight hostname	sfo01vrli01

5.4 VxRail validation

To validate the initial VxRail cluster, follow these steps.

1. In the vSphere Web Client, browse to **Hosts and Clusters**
2. Select the cluster object **VxRail-Cluster**

    a. Click the **Monitor** tab
    b. Below the VxRail section select **Last Configuration Data Sent**

All the relevant information regarding the VxRail cluster can be viewed here including:

- Cluster ID
- Site ID
- Health state
- Product versions
- Chassis model and serial numbers

The values that are used in this deployment are included as an attachment. See Section 1.8 for instructions on accessing the attachments.

6 Deploy VMware Cloud Foundation for VxRail

To complete an automated SDDC deployment by using VCF Cloud Builder, a list of deployment specifications is provided as a set of Deployment Parameters XLS files. This file is used by professional services to deploy the VxRail management cluster into a VCF compatible management domain.

When the deployment is completed, the initial VxRail nodes are converted into a VCF management domain that contains the following virtual machines:

- Cloud Builder
- SDDC Manager
- vRealize Log Insight
- NSX Manager
- NSX controllers
- Workload Domain (WLD) Platform Services Controller (PSC)

Note: The Deployment Parameters Spreadsheet used in this deployment is included as an attachment. See Section 1.8 for instructions on accessing the attachments.

6.1 Configure first leaf switch for NSX VTEP traffic

The following sections cover the configuration for S5248F-ON switch with the hostname `sfo01-Leaf01a`. All switch configuration commands are provided in the file attachments. See Section 1.8 for instructions on accessing the attachments.

1. Configure interface VLAN 2000 to carry NSX VTEP traffic.

```
sfo01-Leaf01A(config)# interface vlan 2000
sfo01-Leaf01A(conf-if-vl-2000)# no shutdown
sfo01-Leaf01A(conf-if-vl-2000)# mtu 9216
sfo01-Leaf01A(conf-if-vl-2000)# ip address 172.20.1.251/24
sfo01-Leaf01A(conf-if-vl-2000)# ip helper-address 10.10.14.5
sfo01-Leaf01A(conf-if-vl-2000)# vrrp-group 200
sfo01-Leaf01A(conf-vlan2000-vrid-200)# virtual-address 172.20.1.253
sfo01-Leaf01A(conf-vlan2000-vrid-200)# exit
sfo01-Leaf01A(conf-if-vl-2000)# exit
```

2. Configure VxRail node ports as VLAN members for VLAN 2000.

```
sfo01-Leaf01A(config)# interface range ethernet1/1/1-1/1/4
sfo01-Leaf01A(conf-range-eth1/1/1-1/1/2)# switchport trunk allowed vlan
2000
sfo01-Leaf01A(conf-range-eth1/1/1-1/1/2)# end
```

Note: OS10EE adds the VLANs to the Trunk Allowed list and does not override existing VLAN membership.

6.2 Verify NSX VTEP DHCP IP addresses

For an NSX VTEP to properly communicate, IP address must be assigned using DHCP. Verify that DHCP IP address have been properly assigned to all NSX VTEP VMkernel interfaces.

1. Run the `esxcli network ip interface ipv4 get` command to display IPv4 information about all local interfaces on the VxRail host. Figure 26 shows the **Address Type** for vmk5 and vmk6 as DHCP. The IPv4 Address and Gateway are confirmed to be assigned by the DHCP server.

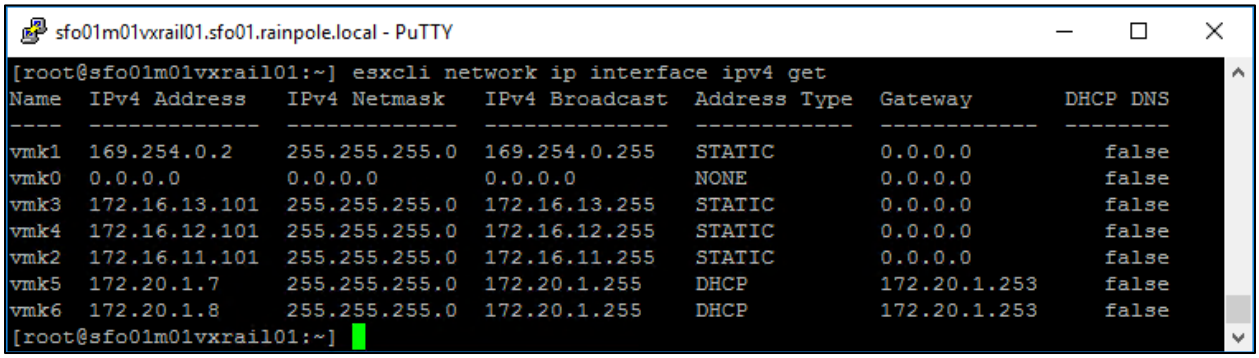


Figure 26 Verify DHCP assignment on NSX VTEP VMkernels (vmk5 & vmk6)

7 Configure Edge Service Gateways

After the VCF deployment is complete, the configuration of VMware NSX can be completed. VCF and VVD automated deployments require that ESG configuration is done manually. The steps are included in, *Configure NSX Dynamic Routing in the Shared Edge and Compute Cluster in Region A*.

For post-VCF deployment, two extra networks are needed to handle the north/south bound traffic for the SDDC. These networks provide the networks that are required for the Edge Services Gateway (ESG) to peer to leaf switches in the environment.

Table 8 shows the two VLANs configured for the two uplinks. These VLANs are only configured on the leaf switches associated to the rack where the ESGs are limited to run. One VLAN is configured on each leaf switch and redundancy is provided by having each ESG peer to each IP address.

Table 8 VLAND ID, VNI ID, Purpose

VLAN ID	Port-group name	Purpose	Leaf switch location
2711	sfo01-m01-vds01-uplink01	north/south uplink 01	Leaf1A
2712	sfo01-m01-vds01-uplink02	north/south uplink 02	Leaf1B

BGP peering is established between the ESG vNIC and two leaf switches to accommodate traffic leaving the SDDC. BGP peering between the ESGs and Leaf1A and Leaf1B is shown in in Figure 27. VMware VM affinity rules are deployed to pin the two ESGs to one physical rack in the data center. This allows the ESG to peer to the IP addresses associated with VLANs 2711 and 2712 on the leaf switches in that rack. Traffic is routed according to the rules of the leaf spine network. For example, northbound traffic, which is routed from the spines to R1 and R2. In this example, R1 and R2 are existing core routers in the data center.

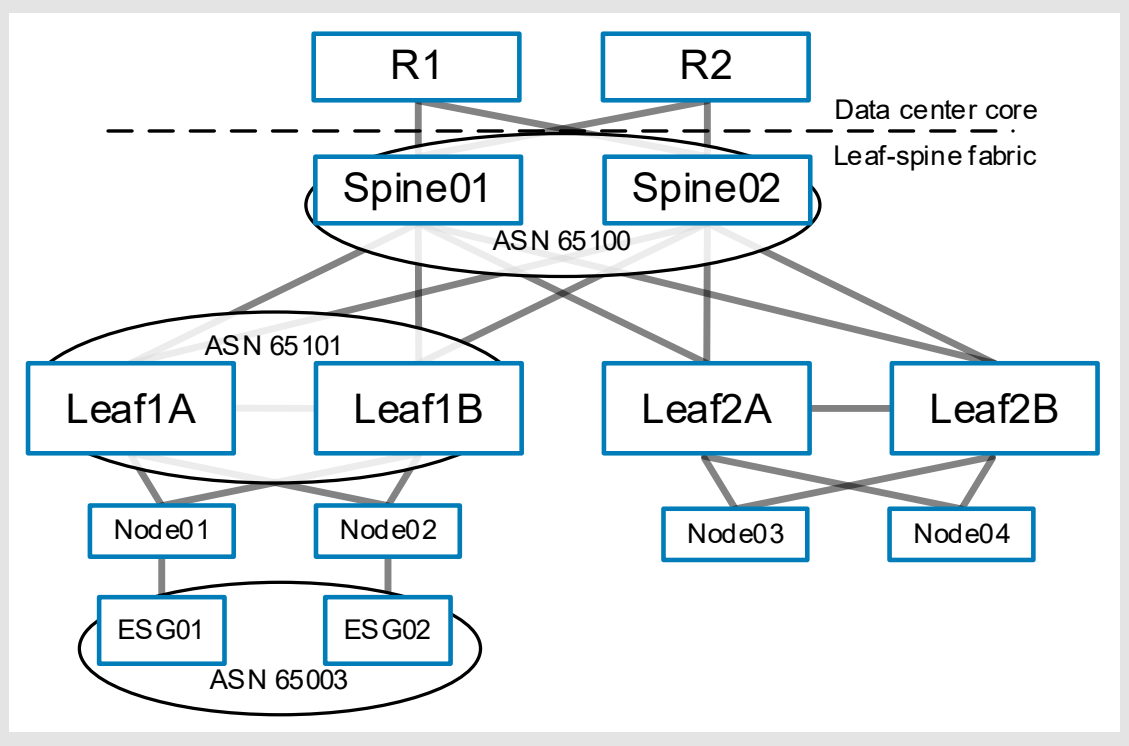


Figure 27 ESG and Edge topology in AZ1

7.1 Create a VM-Host affinity rule for ESGs

In this environment, the underlying VxRail hosts are potentially spread out among numerous racks in the data center. To establish BGP peering to switches Leaf1A and Leaf1B, a VM-Host affinity rule is defined to ensure the ESGs only run on hosts that are physically associated with the corresponding rack where the leaf switches reside.

1. Browse to the cluster in the vSphere Client.
2. Click the **Configure** tab, click **VM/Host Groups**.
3. Click **Add.**
4. In the **Create VM/Host Rules** dialog box, type a name for the rule.
5. From the **Type** drop down menu, select the appropriate type.
6. Click **Add** and in the **Add Group Member** window select the either virtual machines or VxRail nodes to which the rule applies and click **OK**.
7. Click **OK**.
8. Repeat the steps in this section for the remaining rule.

Table 9 VM/Host groups

VM/Host Group Name	Type	Members
Edge Hosts	Host Group	sfo01m01vxrail01 and sfo01m01vxrail02
Edge Service Gateways	VM Group	sfo01m01esg01-0 and sfo01m01esg02-0

Once the group rules are in place, create a VM/Host rule to bind the VM group Edge Service Gateways to the Host group Edge Hosts.

1. Browse to the cluster in the vSphere Client.
2. Click the **Configure** tab, click **VM/Host Rules**.
3. Click **Add.**
4. In the Create VM/Host Rules dialog box, type `host-group-rule-esg`.
5. From the **Type** drop down menu, select **Virtual Machines to Hosts**.
6. From the VM Group drop-down, select **Edge Service Gateways** and choose **Must run on hosts in group**.
7. From the **Host Group** drop-down, select **Edge Hosts**.
8. Click **OK**.

Anti-Affinity rules should be established to ensure both ESGs do not run on the same node. For instructions on creating anti-affinity rules see, *Deploy NSX Edge Devices for North-South Routing in the Shared Edge and Compute Cluster in Region A*.

7.2 Configure first leaf switch for north/south bound traffic

The following sections cover the configuration for S5248F-ON switch with the hostname `sfo01-Leaf01a`. All switch configuration commands are in the file attachments. See Section 1.8 for instructions on accessing the attachments.

1. Create VLAN 2711 and assign an IP address.

```
sfo01-Leaf01A(config)# interface vlan2711
sfo01-Leaf01A(config-if-vl-2711)# description sfo01-m01-esg-uplink01
sfo01-Leaf01A(config-if-vl-2711)# no shutdown
sfo01-Leaf01A(config-if-vl-2711)# mtu 9216
sfo01-Leaf01A(config-if-vl-2711)# ip address 172.27.11.1/24
sfo01-Leaf01A(config-if-vl-2711)# exit
```

2. Configure VxRail node ports as VLAN members for the two uplink VLANs.

```
sfo01-Leaf01A(config)# interface range ethernet1/1/1-1/1/2
sfo01-Leaf01A(conf-range-eth1/1/1-1/1/2)# switchport trunk allowed vlan
2711-2712
sfo01-Leaf01A(conf-range-eth1/1/1-1/1/2)# exit
```

3. Configure eBGP for the peering with the ESGs.

```
sfo01-Leaf01A(config)# router bgp 65101
sfo01-Leaf01A(config-router-bgp-65101)# neighbor 172.27.11.2
sfo01-Leaf01A(config-router-neighbor)# advertisement-interval 5
sfo01-Leaf01A(config-router-neighbor)# bfd
sfo01-Leaf01A(config-router-neighbor)# fall-over
sfo01-Leaf01A(config-router-neighbor)# password <bgp-password>
sfo01-Leaf01A(config-router-neighbor)# remote-as 65003
sfo01-Leaf01A(config-router-neighbor)# no shutdown
sfo01-Leaf01A(config-router-neighbor)# exit
sfo01-Leaf01A(config-router-bgp-65101)# neighbor 172.27.11.3
sfo01-Leaf01A(config-router-neighbor)# advertisement-interval 5
sfo01-Leaf01A(config-router-neighbor)# bfd
sfo01-Leaf01A(config-router-neighbor)# fall-over
sfo01-Leaf01A(config-router-neighbor)# password <bgp-password>
sfo01-Leaf01A(config-router-neighbor)# remote-as 65003
sfo01-Leaf01A(config-router-neighbor)# no shutdown
sfo01-Leaf01A(config-router-neighbor)# end
```

4. Update the existing route map.

```
sfo01-Leaf01A(config)# ip prefix-list spine-leaf seq 60 permit
172.27.11.0/24
sfo01-Leaf01A(config)# ip prefix-list spine-leaf seq 70 permit
172.27.12.0/24
```

Note: Additional networks will need to be added to the IP prefix list based on the tenant workload networks used in the environment.

7.3 Verify peering of NSX edge devices and establishment of BGP

The NSX Edge devices must establish a connection to each of the leaf switches before BGP updates can be exchanged. Verify that the NSX edge devices are successfully peering, and that BGP routing has been established.

1. From switch **sfo01-leaf01A** run the `show ip bgp summary` command to show a summarization of the neighbor adjacencies. Figure 28 shows both ESGs with ASN 65003.

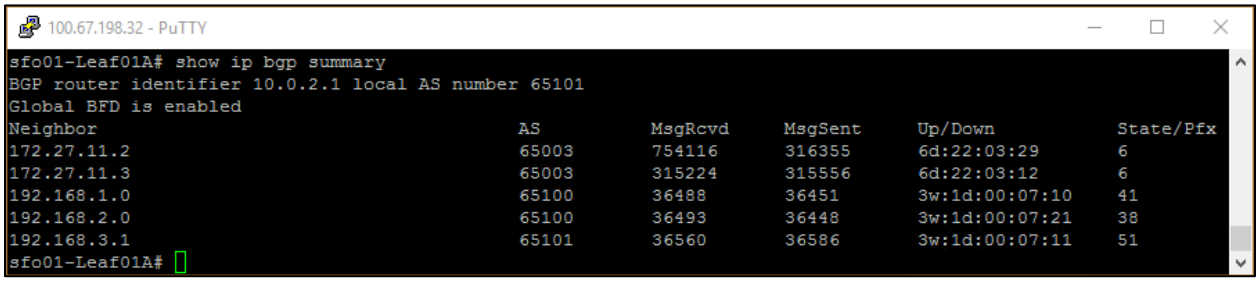


Figure 28 Output of `show ip bgp summary`

Note: For detailed step-by-step instructions on creating the required NSX edge devices and performing equivalent verification steps, see **Deploy NSX Edge Devices for North-South Routing in Region A**.

8 Stretching clusters to AZ2

VMware Cloud Foundation supports stretching a cluster across two availability zones within a region. In this section, Dell Networking OS10EE guidance is provided to create the network that is required for deploying a stretch cluster. For specific VCF steps on creating a stretched cluster including deploying the vSAN witness appliance, see the *VMware Cloud Foundation on Dell EMC VxRail Admin Guide*.

Note: Dell EMC VxRail does not support stretching clusters over Layer 3 networks. As a result, the cluster needs to be stretched over Layer 2.

8.1 Planning and Preparation

This section lists the prerequisites for preparing the clusters for stretching.

8.1.1 VLAN IDs and IP subnets for AZ2

Table 10 lists the VLANs and virtual networks that are required. The ESXi management VLAN, vSAN VLAN, vMotion VLAN, and the VxRail discovery VLAN must be stretched between the two availability zones. Also, the VLANs between the two zones must be identical.

IP reachability is required to stretch a Layer 2 network between the two availability zones using BGP EVPN and VXLAN EVPN is then configured to include all leaf pairs in both availability zones in the same virtual networks.

Table 10 VLAN and IP subnet configuration between AZ1 and AZ2

Cluster in Region A	VLAN function	VLAN ID	VNI	Subnet	Gateway
Management cluster	ESXi Management	1611	1611	172.16.11.0/24	172.16.11.253
	vSphere vMotion	1612	1612	172.16.12.0/24	172.16.12.253
	vSAN	1613	1613	172.16.13.0/24	172.16.13.253
	VxRail node discovery	3939	3939	-	-

Table 11 lists the additional IP subnets and VLANs are used to handle NSX VTEP traffic and NSX ESG traffic that is unique to availability zone 2.

Table 11 NSX VLAN and IP subnets for AZ2

Cluster in Region A	VLAN function	VLAN ID	VNI	Subnet	Gateway
Management cluster	VXLAN (NSX VTEP)	2000	-	172.20.n.0/24	172.20.<rack-id>.253
	Uplink01	1650	-	172.16.50.0/24	172.16.50.1
	Uplink02	1651	-	172.16.51.0/24	172.16.51.1

Note: Use these VLAN IDs and IP subnets as samples. Configure the actual VLAN IDs and IP subnets according to the environment.

8.1.2 vSAN witness traffic

A vSAN stretched cluster requires a witness host that is deployed in a third location, Region B, and is in a geographically separate location from Region A. Table 12 shows the two required networks for vSAN witness traffic. In this example, Region B is configured identically to Region A and uses virtual networks to provide both management and vSAN traffic across multiple racks at that location.

Table 12 vSAN Witness VLAN and IP subnets

Cluster in Region B	VLAN function	VLAN ID	VNI	Subnet	Gateway
Management cluster	Management	1711	1711	172.17.11.0/24	172.17.11.253
	vSAN	1713	1713	172.17.13.0/24	172.17.13.253

Note: The complete deployment of Region B is beyond the scope of this document and only the components related to the vSAN witness are shown here.

Figure 14 shows that the vSAN witness appliance is available through the Data Center Interconnect (DCI). The required VLANs are stretched between the two availability zones. The two extra VLANs for the vSAN witness location (1711 and 1713) are also shown.

The vSAN witness requires that both management and the vSAN network can route to both availability zones. In this document, both vSAN witness networks (management and vSAN) are associated with virtual networks.

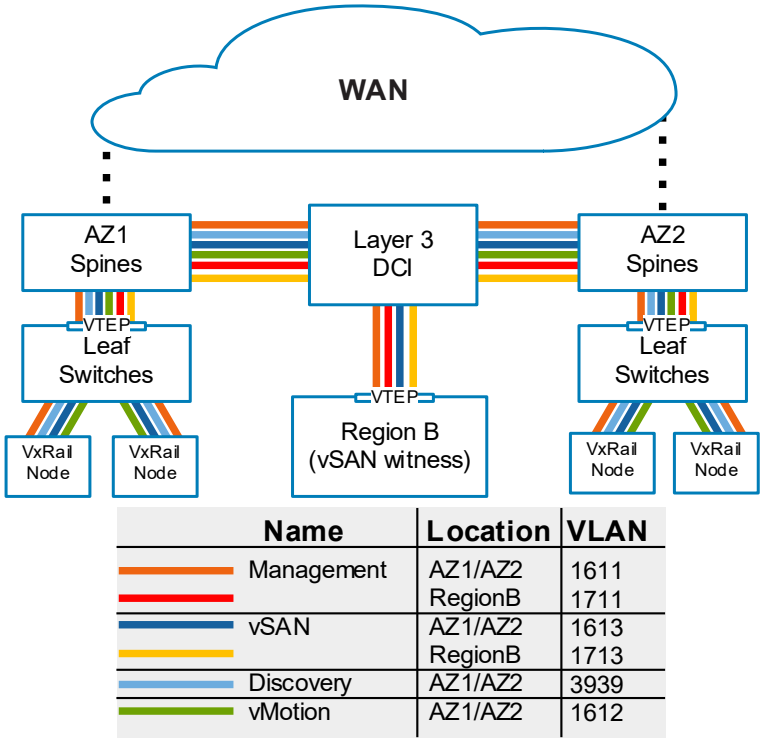


Figure 29 Using Anycast IP IRB routing connecting availability zones to Region B

Note: Centralized routing is not required for VM traffic that exists in an NSX virtual wire. This traffic is routed through the default VRF and is not encapsulated by the underlay. See section 7, Configure Edge Service Gateways.

8.1.3 Underlay network physical design for AZ2

Figure 30 shows the wiring configuration for the six switches that comprise the AZ2 leaf-spine network. The configuration is identical to AZ1 except a different spine switch model is shown here. The colored solid lines are 100 GbE links and the light blue dashed lines are two QSFP28-DD 200 GbE cable pairs are used for the VLTi. The use of QSFP28-DD offers a 400 GbE VLTi to handle any potential traffic increases resulting from failed interconnects to the spine layers. As a rule, it is suggested to maintain at minimum a 1:1 ratio between available bandwidth to the spine and bandwidth for the VLTi.

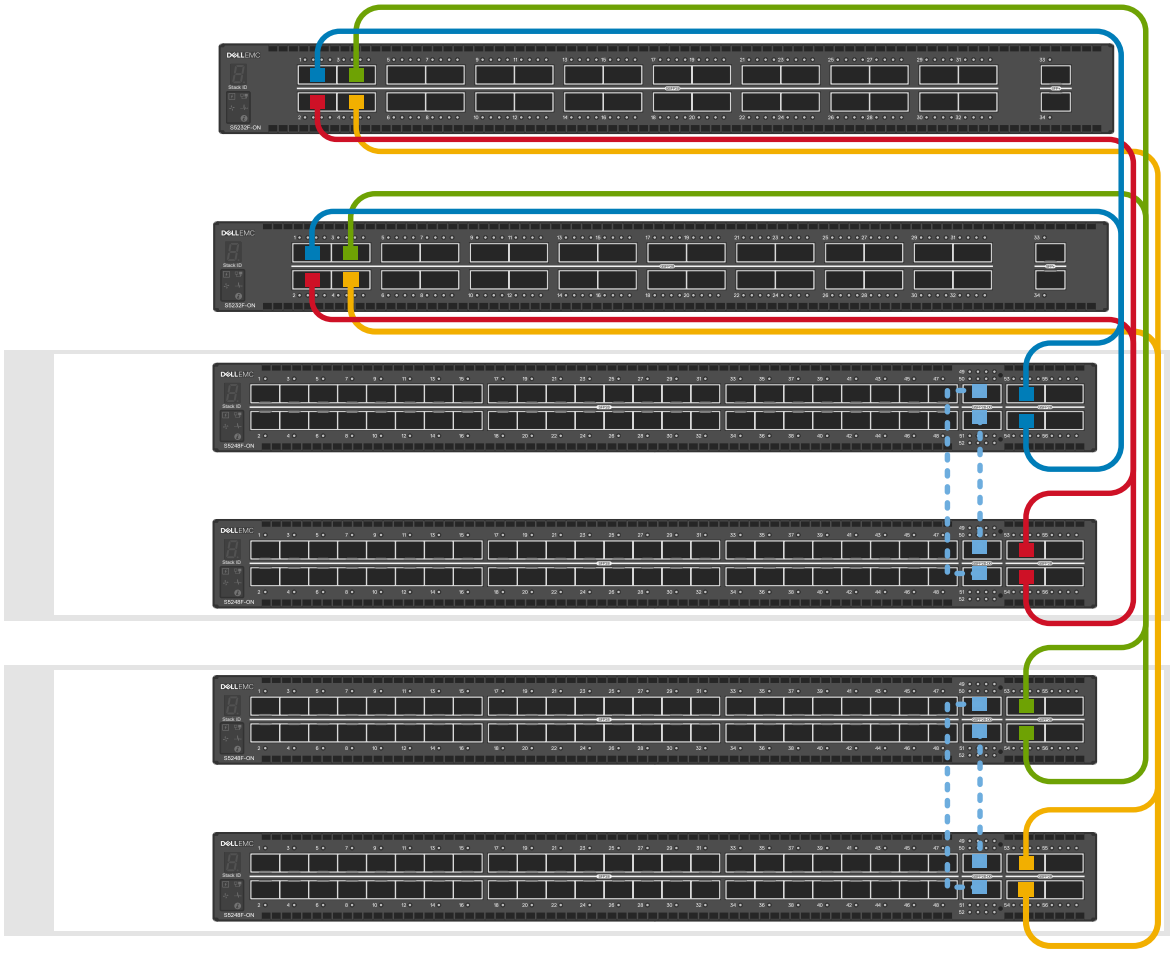


Figure 30 Physical topology for AZ2

Note: The S5232F-ON supports VTEP functionality. In a centralized routing model, the S5232F-ON can also take on the role of the border leaf, the concept was covered previously.

8.2 Configure and verify the underlay network

In this section network configuration is provided for availability zone 2. The configuration is identical to the configuration completed throughout previous sections that are related to available zone 1.

8.2.1 More switch configurations in AZ1

For a successful deployment, verify that the required vSAN witness networks are accessible from AZ1. Table 13 contains the additional networks that are needed to establish asymmetric routing between the availability zone and the Region B vSAN witness.

Table 13 More switch settings for leaf switches in AZ1

Setting	S5248F-Leaf1A	S5248F-Leaf1B	S5248F-Leaf2A	S5248F-Leaf2B
virtual-network 1711 IP addresses (interface and anycast)	• 172.17.11.252/24 • 172.17.11.253/24	• 172.17.11.251/24 • 172.17.11.253/24	• 172.17.11.250/24 • 172.17.11.253/24	• 172.17.11.249/24 • 172.17.11.253/24
virtual-network 1713 IP addresses (interface and anycast)	• 172.17.13.252/24 • 172.17.13.253/24	• 172.17.13.251/24 • 172.17.13.253/24	• 172.17.13.250/24 • 172.17.13.253/24	• 172.17.13.249/24 • 172.17.13.253/24

The sections listed below cover the configuration for S5248F-ON switch with the hostname `sfo01-Leaf01a`. This section provides the additional configuration information that is required to establish connectivity with the vSAN witness management and vSAN networks.

1. Add the additional VXLAN virtual networks

```
sfo01-Leaf01A(config)# virtual-network 1711
sfo01-Leaf01A(config-vn)# vxlan-vni 1711
sfo01-Leaf01A(config-vn)# exit
sfo01-Leaf01A(config)# virtual-network 1713
sfo01-Leaf01A(config-vn)# vxlan-vni 1713
sfo01-Leaf01A(config-vn)# exit
```

2. Assign VLAN member interfaces to two virtual networks

```
sfo01-Leaf01A(config)# interface vlan1711
sfo01-Leaf01A(config-if-vl-1711)# description lax-mgmt
sfo01-Leaf01A(config-if-vl-1711)# virtual-network 1711
sfo01-Leaf01A(config-if-vl-1711)# no shutdown
sfo01-Leaf01A(config-if-vl-1711)# mtu 9216
sfo01-Leaf01A(config-if-vl-1711)# exit
sfo01-Leaf01A(config)# interface vlan1713
sfo01-Leaf01A(config-if-vl-1713)# description lax-vsan
sfo01-Leaf01A(config-if-vl-1713)# virtual-network 1713
sfo01-Leaf01A(config-if-vl-1713)# no shutdown
sfo01-Leaf01A(config-if-vl-1713)# mtu 9216
sfo01-Leaf01A(config-if-vl-1713)# exit
```

3. Configure EVPN

```
sfo01-Leaf01A(config)# evpn
sfo01-Leaf01A(config-evpn)# evi 1711
sfo01-Leaf01A(config-evpn-evi-1711)# vni 1711
sfo01-Leaf01A(config-evpn-evi-1711)# rd 10.222.222.1:1711
sfo01-Leaf01A(config-evpn-evi-1711)# route-target 1711:1711 both
sfo01-Leaf01A(config-evpn-evi-1711)# exit
sfo01-Leaf01A(config-evpn)# evi 1713
sfo01-Leaf01A(config-evpn-evi-1713)# vni 1713
sfo01-Leaf01A(config-evpn-evi-1713)# rd 10.222.222.1:1713
sfo01-Leaf01A(config-evpn-evi-1713)# route-target 1713:1713 both
sfo01-Leaf01A(config-evpn-evi-1713)# exit
sfo01-Leaf01A(config-evpn)# exit
```

4. Configure routing on virtual networks

```
sfo01-Leaf01A(config)# interface virtual-network1711
sfo01-Leaf01A(conf-if-vn-1711)# no shutdown
sfo01-Leaf01A(conf-if-vn-1711)# mtu 9216
sfo01-Leaf01A(conf-if-vn-1711)# ip vrf forwarding tenant1
sfo01-Leaf01A(conf-if-vn-1711)# ip address 172.17.11.252/24
sfo01-Leaf01A(conf-if-vn-1711)# ip virtual-router address 172.17.11.253
sfo01-Leaf01A(conf-if-vn-1711)# exit
sfo01-Leaf01A(config)# interface virtual-network1713
sfo01-Leaf01A(conf-if-vn-1713)# no shutdown
sfo01-Leaf01A(conf-if-vn-1713)# mtu 9216
sfo01-Leaf01A(conf-if-vn-1713)# ip vrf forwarding tenant1
sfo01-Leaf01A(conf-if-vn-1713)# ip address 172.17.13.252/24
sfo01-Leaf01A(conf-if-vn-1713)# ip virtual-router address 172.17.13.253
sfo01-Leaf01A(conf-if-vn-1713)# exit
```

8.2.2 Switch settings for AZ2

Table 14 shows the unique values for the four S5248F-ON switches in Availability Zone 2. Table 14 provides a summary of the configuration differences between each switch and each VLT switch pair.

Table 14 Unique switch settings for leaf switches in AZ2

Setting	S5248F-Leaf1A	S5248F-Leaf1B	S5248F-Leaf2A	S5248F-Leaf2B
Hostname	sfo02-Leaf01A	sfo02-Leaf01B	sfo02-Leaf02A	sfo02-Leaf02B
Management IP address	100.67.167.32/24	100.67.167.31/24	100.67.167.30/24	100.67.167.29/24
Autonomous System Number (ASN)	65301	65301	65302	65302
P2P IP addresses	• 192.168.1.31/31 • 192.168.2.31/31	• 192.168.1.33/31 • 192.168.2.33/31	• 192.168.1.35/31 • 192.168.2.35/31	• 192.168.1.37/31 • 192.168.2.37/31
Loopback0 address (router ID)	10.0.2.31/32	10.0.2.32/32	10.0.2.33/32	10.0.2.34/32
Loopback1 address (EVPN)	10.2.2.31/32	10.2.2.32/32	10.2.2.33/32	10.2.2.34/32
Loopback2 address (NVE)	10.222.222.31/32	10.222.222.31/32	10.222.222.32/32	10.222.222.32/32
VLAN 4000 IP address	192.168.3.30/31	192.168.3.31/31	192.168.3.32/31	192.168.3.33/31

VLAN 2000 IP addresses (interface and VIP)	• 172.20.201.251/24 • 172.20.201.253/24	• 172.20.201.252/24 • 172.20.201.253/24	• 172.20.202.251/24 • 172.20.202.253/24	• 172.20.202.252/24 • 172.20.202.253/24
VLAN 1650 IP address (ESG)	172.16.50.1/24	-	-	-
VLAN 1651 IP address (ESG)	-	172.16.51.1/24	-	-
virtual-network 1611 IP addresses (interface and anycast)	• 172.16.11.248/24 • 172.16.11.253/24	• 172.16.11.247/24 • 172.16.11.253/24	• 172.16.11.246/24 • 172.16.11.253/24	• 172.16.11.245/24 • 172.16.11.253/24
virtual-network 1612 IP addresses (interface and anycast)	• 172.16.12.248/24 • 172.16.12.253/24	• 172.16.12.247/24 • 172.16.12.253/24	• 172.16.12.246/24 • 172.16.12.253/24	• 172.16.12.245/24 • 172.16.12.253/24
virtual-network 1613 IP addresses (interface and anycast)	• 172.16.13.248/24 • 172.16.13.253/24	• 172.16.13.247/24 • 172.16.13.253/24	• 172.16.13.246/24 • 172.16.13.253/24	• 172.16.13.245/24 • 172.16.13.253/24
virtual-network 1711 IP addresses (interface and anycast)	• 172.17.11.248/24 • 172.17.11.253/24	• 172.17.11.247/24 • 172.17.11.253/24	• 172.17.11.246/24 • 172.17.11.253/24	• 172.17.11.245/24 • 172.17.11.253/24
virtual-network 1713 IP addresses (interface and anycast)	• 172.17.13.248/24 • 172.17.13.253/24	• 172.17.13.247/24 • 172.17.13.253/24	• 172.17.13.247/24 • 172.17.13.253/24	• 172.17.13.247/24 • 172.17.13.253/24

Note: Use these VLAN IDs and IP subnets as exsamples. Configure the VLAN IDs and IP subnets according to your environment.

8.2.3 Configure the first leaf switch in AZ2

This section covers the configuration for S5248F-ON switch with the `sfo02-Leaf01a` hostname. Switch configuration commands are in the file attachments. See Section 1.8 for instructions on accessing the attachments.

1. Configure general switch settings including management and NTP source.

```
OS10# configure terminal
OS10(config)# interface mgmt 1/1/1
OS10(conf-if-ma-1/1/1)# no ip address dhcp
OS10(conf-if-ma-1/1/1)# ip address 100.67.167.32/24
OS10(conf-if-ma-1/1/1)# exit
OS10(config)# management route 100.67.0.0/16 managementethernet
OS10(config)# hostname sfo02-Leaf01A
sfo02-Leaf01A(config)# ntp server 100.67.10.20
sfo02-Leaf01A(config)# bfd enable
```

2. Configure a loopback interface for the Router ID

```
sfo02-Leaf01A(config)# interface loopback 0
sfo02-Leaf01A(conf-if-lo-0)# description Router-ID
sfo02-Leaf01A(conf-if-lo-0)# no shutdown
sfo02-Leaf01A(conf-if-lo-0)# ip address 10.0.2.31/32
sfo02-Leaf01A(conf-if-lo-0)# exit
```

3. Configure a loopback interface for NVE

```
sfo02-Leaf01A(config)# interface loopback 2
sfo02-Leaf01A(conf-if-lo-2)# description nve_loopback
sfo02-Leaf01A(conf-if-lo-2)# no shutdown
sfo02-Leaf01A(conf-if-lo-2)# ip address 10.222.222.31/32
sfo02-Leaf01A(conf-if-lo-2)# exit
```

4. Configure the loopback interface for the VXLAN source tunnel interface

```
sfo02-Leaf01A(config)# nve
sfo02-Leaf01A(config-nve)# source-interface loopback2
sfo02-Leaf01A(config-nve)# exit
```

5. Configure VXLAN virtual networks

```
sfo02-Leaf01A(config)# virtual-network 1611
sfo02-Leaf01A(config-vn)# vxlan-vni 1611
sfo02-Leaf01A(config-vn)# exit
sfo02-Leaf01A(config)# virtual-network 1612
sfo02-Leaf01A(config-vn)# vxlan-vni 1612
sfo02-Leaf01A(config-vn)# exit
sfo02-Leaf01A(config)# virtual-network 1613
sfo02-Leaf01A(config-vn)# vxlan-vni 1613
sfo02-Leaf01A(config-vn)# exit
sfo02-Leaf01A(config)# virtual-network 1711
sfo02-Leaf01A(config-vn)# vxlan-vni 1711
sfo02-Leaf01A(config-vn)# exit
```

```
sfo02-Leaf01A(config)# virtual-network 1713
sfo02-Leaf01A(config-vn)# vxlan-vni 1713
sfo02-Leaf01A(config-vn)# exit
sfo02-Leaf01A(config)# virtual-network 3939
sfo02-Leaf01A(config-vn)# vxlan-vni 3939
sfo02-Leaf01A(config-vn)# exit
```

6. Assign VLAN member interfaces to virtual networks

```
sfo02-Leaf01A(config)# interface vlan1611
sfo02-Leaf01A(config-if-vl-1611)# description sfo-mgmt
sfo02-Leaf01A(config-if-vl-1611)# virtual-network 1611
sfo02-Leaf01A(config-if-vl-1611)# no shutdown
sfo02-Leaf01A(config-if-vl-1611)# mtu 9216
sfo02-Leaf01A(config-if-vl-1611)# exit
sfo02-Leaf01A(config)# interface vlan1612
sfo02-Leaf01A(config-if-vl-1612)# virtual-network 1612
sfo02-Leaf01A(config-if-vl-1612)# description sfo-vmotion
sfo02-Leaf01A(config-if-vl-1612)# no shutdown
sfo02-Leaf01A(config-if-vl-1612)# mtu 9216
sfo02-Leaf01A(config-if-vl-1612)# exit
sfo02-Leaf01A(config)# interface vlan1613
sfo02-Leaf01A(config-if-vl-1613)# virtual-network 1613
sfo02-Leaf01A(config-if-vl-1613)# description sfo-vsan
sfo02-Leaf01A(config-if-vl-1613)# no shutdown
sfo02-Leaf01A(config-if-vl-1613)# mtu 9216
sfo02-Leaf01A(config-if-vl-1613)# exit
sfo02-Leaf01A(config)# interface vlan1711
sfo02-Leaf01A(config-if-vl-1711)# description lax-mgmt
sfo02-Leaf01A(config-if-vl-1711)# virtual-network 1711
sfo02-Leaf01A(config-if-vl-1711)# no shutdown
sfo02-Leaf01A(config-if-vl-1711)# mtu 9216
sfo02-Leaf01A(config-if-vl-1711)# exit
sfo02-Leaf01A(config)# interface vlan1713
sfo02-Leaf01A(config-if-vl-1713)# description lax-vsan
sfo02-Leaf01A(config-if-vl-1713)# virtual-network 1713
sfo02-Leaf01A(config-if-vl-1713)# no shutdown
sfo02-Leaf01A(config-if-vl-1713)# mtu 9216
sfo02-Leaf01A(config-if-vl-1713)# exit
sfo02-Leaf01A(config)# interface vlan3939
sfo02-Leaf01A(config-if-vl-3939)# description vxrail-discovery
sfo02-Leaf01A(config-if-vl-3939)# virtual-network 3939
sfo02-Leaf01A(config-if-vl-3939)# no shutdown
sfo02-Leaf01A(config-if-vl-3939)# mtu 9216
sfo02-Leaf01A(config-if-vl-3939)# exit
```

7. Configure access ports as VLAN members for a switch-scoped VLAN-to-VNI mapping

```
sfo02-Leaf01A(config)# interface ethernet1/1/1
sfo02-Leaf01A(conf-if-eth1/1/1)# description sfo02m01vxrail01
sfo02-Leaf01A(conf-if-eth1/1/1)# no shutdown
sfo02-Leaf01A(conf-if-eth1/1/1)# switchport mode trunk
sfo02-Leaf01A(conf-if-eth1/1/1)# switchport access vlan 1611
sfo02-Leaf01A(conf-if-eth1/1/1)# switchport trunk allowed vlan 1612-
1613,3939
sfo02-Leaf01A(conf-if-eth1/1/1)# mtu 9216
sfo02-Leaf01A(conf-if-eth1/1/1)# spanning-tree port type edge
sfo02-Leaf01A(conf-if-eth1/1/1)# flowcontrol receive on
sfo02-Leaf01A(conf-if-eth1/1/1)# flowcontrol transmit off
sfo02-Leaf01A(conf-if-eth1/1/1)# exit
sfo02-Leaf01A(config)# interface ethernet1/1/2
sfo02-Leaf01A(conf-if-eth1/1/2)# description sfo02m01vxrail02
sfo02-Leaf01A(conf-if-eth1/1/2)# no shutdown
sfo02-Leaf01A(conf-if-eth1/1/2)# switchport mode trunk
sfo02-Leaf01A(conf-if-eth1/1/2)# switchport access vlan 1611
sfo02-Leaf01A(conf-if-eth1/1/2)# switchport trunk allowed vlan 1612-
1613,3939
sfo02-Leaf01A(conf-if-eth1/1/2)# mtu 9216
sfo02-Leaf01A(conf-if-eth1/1/2)# spanning-tree port type edge
sfo02-Leaf01A(conf-if-eth1/1/2)# flowcontrol receive on
sfo02-Leaf01A(conf-if-eth1/1/2)# flowcontrol transmit off
sfo02-Leaf01A(conf-if-eth1/1/2)# exit
```

8. Configure upstream network-facing ports

```
sfo02-Leaf01A(config)# interface ethernet1/1/53
sfo02-Leaf01A(conf-if-eth1/1/53)# description sfo02-spine01
sfo02-Leaf01A(conf-if-eth1/1/53)# no shutdown
sfo02-Leaf01A(conf-if-eth1/1/53)# no switchport
sfo02-Leaf01A(conf-if-eth1/1/53)# mtu 9216
sfo02-Leaf01A(conf-if-eth1/1/53)# ip address 192.168.1.31/31
sfo02-Leaf01A(conf-if-eth1/1/53)# exit
sfo02-Leaf01A(config)# interface ethernet1/1/54
sfo02-Leaf01A(conf-if-eth1/1/54)# description sfo02-spine02
sfo02-Leaf01A(conf-if-eth1/1/54)# no shutdown
sfo02-Leaf01A(conf-if-eth1/1/54)# no switchport
sfo02-Leaf01A(conf-if-eth1/1/54)# mtu 9216
sfo02-Leaf01A(conf-if-eth1/1/54)# ip address 192.168.2.31/31
sfo02-Leaf01A(conf-if-eth1/1/54)# exit
```

9. Add a route map. This example route map is used to illustrate how to allow IP traffic to be passed on the switch.

```
sfo02-Leaf01A(config)# ip prefix-list spine-leaf seq 10 permit 10.0.2.0/24
ge 32
sfo02-Leaf01A(config)# ip prefix-list spine-leaf seq 20 permit 10.2.2.0/24
ge 32
sfo02-Leaf01A(config)# ip prefix-list spine-leaf seq 30 permit
10.222.222.0/24 ge 32
sfo02-Leaf01A(config)# ip prefix-list spine-leaf seq 40 permit
172.20.101.0/24
sfo02-Leaf01A(config)# ip prefix-list spine-leaf seq 50 permit
172.20.202.0/24
sfo02-Leaf01A(config)# ip prefix-list spine-leaf seq 60 permit
172.16.49.0/24
sfo02-Leaf01A(config)# ip prefix-list spine-leaf seq 70 permit
172.16.40.0/24
sfo02-Leaf01A(config)# route-map spine-leaf permit 10
sfo02-Leaf01A(config-route-map)# match ip address prefix-list spine-leaf
sfo02-Leaf01A(config-route-map)# exit
```

10. Configure eBGP

```
sfo02-Leaf01A(config)# router bgp 65301
sfo02-Leaf01A(config-router-bgp-65301)# router-id 10.0.2.31
sfo02-Leaf01A(config-router-bgp-65301)# bfd all-neighbors interval 200
min_rx 200 multiplier 3 role active
sfo02-Leaf01A(config-router-bgp-65301)# address-family ipv4 unicast
sfo02-Leaf01A(config-router-bgpv4-af)# redistribute connected route-map
spine-leaf
sfo02-Leaf01A(config-router-bgpv4-af)# exit
sfo02-Leaf01A(config-router-bgp-65301)# bestpath as-path multipath-relax
sfo02-Leaf01A(config-router-bgp-65301)# maximum-paths ebgp 2
```

11. Configure eBGP for the IPv4 point-to-point peering

```
sfo02-Leaf01A(config-router-bgp-65301)# neighbor 192.168.1.30
sfo02-Leaf01A(config-router-neighbor)# advertisement-interval 5
sfo02-Leaf01A(config-router-neighbor)# bfd
sfo02-Leaf01A(config-router-neighbor)# fall-over
sfo02-Leaf01A(config-router-neighbor)# remote-as 65300
sfo02-Leaf01A(config-router-neighbor)# no shutdown
sfo02-Leaf01A(config-router-neighbor)# address-family ipv4 unicast
sfo02-Leaf01A(config-router-neighbor-af)# exit
sfo02-Leaf01A(config-router-neighbor)# exit
sfo02-Leaf01A(config-router-bgp-65301)# neighbor 192.168.2.30
sfo02-Leaf01A(config-router-neighbor)# advertisement-interval 5
sfo02-Leaf01A(config-router-neighbor)# bfd
sfo02-Leaf01A(config-router-neighbor)# fall-over
sfo02-Leaf01A(config-router-neighbor)# remote-as 65300
sfo02-Leaf01A(config-router-neighbor)# no shutdown
sfo02-Leaf01A(config-router-neighbor)# address-family ipv4 unicast
sfo02-Leaf01A(config-router-neighbor-af)# exit
sfo02-Leaf01A(config-router-neighbor)# exit
sfo02-Leaf01A(config-router-bgp-65301)# exit
```

12. Configure a loopback interface for BGP EVPN peering

```
sfo02-Leaf01A(config)# interface loopback 1
sfo02-Leaf01A(conf-if-lo-1)# description evpn_loopback
sfo02-Leaf01A(conf-if-lo-1)# no shutdown
sfo02-Leaf01A(conf-if-lo-1)# ip address 10.2.2.31/32
sfo02-Leaf01A(conf-if-lo-1)# exit
```

13. Configure BGP EVPN peering

```
sfo02-Leaf01A(config)# router bgp 65301
sfo02-Leaf01A(config-router-bgp-65301)# neighbor 10.2.1.31
sfo02-Leaf01A(config-router-neighbor)# remote-as 65300
sfo02-Leaf01A(config-router-neighbor)# ebgp-multihop 2
sfo02-Leaf01A(config-router-neighbor)# send-community extended
sfo02-Leaf01A(config-router-neighbor)# update-source loopback1
sfo02-Leaf01A(config-router-neighbor)# no shutdown
sfo02-Leaf01A(config-router-neighbor)# address-family ipv4 unicast
sfo02-Leaf01A(config-router-neighbor-af)# no activate
sfo02-Leaf01A(config-router-neighbor-af)# exit
sfo02-Leaf01A(config-router-neighbor)# address-family l2vpn evpn
sfo02-Leaf01A(config-router-neighbor-af)# activate
sfo02-Leaf01A(config-router-neighbor-af)# exit
sfo02-Leaf01A(config-router-neighbor)# exit
sfo02-Leaf01A(config-router-bgp-65301)# neighbor 10.2.1.32
sfo02-Leaf01A(config-router-neighbor)# remote-as 65300
sfo02-Leaf01A(config-router-neighbor)# ebgp-multihop 2
sfo02-Leaf01A(config-router-neighbor)# send-community extended
sfo02-Leaf01A(config-router-neighbor)# update-source loopback1
sfo02-Leaf01A(config-router-neighbor)# no shutdown
```

```
sfo02-Leaf01A(config-router-neighbor)# address-family ipv4 unicast
sfo02-Leaf01A(config-router-neighbor-af)# no activate
sfo02-Leaf01A(config-router-neighbor-af)# exit
sfo02-Leaf01A(config-router-neighbor)# address-family l2vpn evpn
sfo02-Leaf01A(config-router-neighbor-af)# activate
sfo02-Leaf01A(config-router-neighbor-af)# exit
sfo02-Leaf01A(config-router-neighbor)# exit
sfo02-Leaf01A(config-router-bgp-65301)# exit
```

14. Configure EVPN

```
sfo02-Leaf01A(config)# evpn
sfo02-Leaf01A(config-evpn)# evi 1611
sfo02-Leaf01A(config-evpn-evi-1611)# vni 1611
sfo02-Leaf01A(config-evpn-evi-1611)# rd 10.222.222.31:1611
sfo02-Leaf01A(config-evpn-evi-1611)# route-target 1611:1611 both
sfo02-Leaf01A(config-evpn-evi-1611)# exit
sfo02-Leaf01A(config-evpn)# evi 1612
sfo02-Leaf01A(config-evpn-evi-1612)# vni 1612
sfo02-Leaf01A(config-evpn-evi-1612)# rd 10.222.222.31:1612
sfo02-Leaf01A(config-evpn-evi-1612)# route-target 1612:1612 both
sfo02-Leaf01A(config-evpn-evi-1612)# exit
sfo02-Leaf01A(config-evpn)# evi 1613
sfo02-Leaf01A(config-evpn-evi-1613)# vni 1613
sfo02-Leaf01A(config-evpn-evi-1613)# rd 10.222.222.31:1613
sfo02-Leaf01A(config-evpn-evi-1613)# route-target 1613:1613 both
sfo02-Leaf01A(config-evpn-evi-1613)# exit
sfo02-Leaf01A(config-evpn)# evi 1711
sfo02-Leaf01A(config-evpn-evi-1711)# vni 1711
sfo02-Leaf01A(config-evpn-evi-1711)# rd 10.222.222.31:1711
sfo02-Leaf01A(config-evpn-evi-1711)# route-target 1711:1711 both
sfo02-Leaf01A(config-evpn-evi-1711)# exit
sfo02-Leaf01A(config-evpn)# evi 1713
sfo02-Leaf01A(config-evpn-evi-1713)# vni 1713
sfo02-Leaf01A(config-evpn-evi-1713)# rd 10.222.222.31:1713
sfo02-Leaf01A(config-evpn-evi-1713)# route-target 1713:1713 both
sfo02-Leaf01A(config-evpn-evi-1713)# exit
```

15. Configure a dedicated L3 underlay path to reach VLT peer in case of network failure

```
sfo02-Leaf01A(config)# interface vlan4000
sfo02-Leaf01A(config-if-vl-4000)# no shutdown
sfo02-Leaf01A(config-if-vl-4000)# mtu 9216
sfo02-Leaf01A(config-if-vl-4000)# ip address 192.168.3.30/31
sfo02-Leaf01A(config-if-vl-4000)# exit
```

16. Configure VLTi member links

```
sfo02-Leaf01A(config)# interface range ethernet1/1/49-1/1/52
sfo02-Leaf01A(conf-range-eth1/1/49-1/1/52)# description VLTi
sfo02-Leaf01A(conf-range-eth1/1/49-1/1/52)# no shutdown
sfo02-Leaf01A(conf-range-eth1/1/49-1/1/52)# no switchport
sfo02-Leaf01A(conf-range-eth1/1/49-1/1/52)# exit
```

17. Configure the VLT domain

```
sfo02-Leaf01A(config)# vlt-domain 1
sfo02-Leaf01A(conf-vlt-1)# backup destination 100.67.167.31
sfo02-Leaf01A(conf-vlt-1)# discovery-interface ethernet1/1/49-1/1/52
sfo02-Leaf01A(conf-vlt-1)# peer-routing
sfo02-Leaf01A(conf-vlt-1)# vlt-mac 00:00:01:02:03:01
sfo02-Leaf01A(conf-vlt-1)# exit
```

18. Configure UFD with uplink VLT ports and downlink network ports

```
sfo02-Leaf01A(config)# uplink-state-group 1
sfo02-Leaf01A(conf-uplink-state-group-1)# enable
sfo02-Leaf01A(conf-uplink-state-group-1)# downstream ethernet1/1/1-1/1/8
sfo02-Leaf01A(conf-uplink-state-group-1)# upstream ethernet1/1/53
sfo02-Leaf01A(conf-uplink-state-group-1)# upstream ethernet1/1/54
sfo02-Leaf01A(conf-uplink-state-group-1)# exit
```

19. Configure iBGP IPv4 peering between VLT peers

```
sfo02-Leaf01A(config)# router bgp 65301
sfo02-Leaf01A(config-router-bgp-65301)# neighbor 192.168.3.31
sfo02-Leaf01A(config-router-neighbor)# remote-as 65301
sfo02-Leaf01A(config-router-neighbor)# no shutdown
sfo02-Leaf01A(config-router-neighbor)# exit
```

20. Create a tenant VRF. An OS10 best practice is to isolate any virtual network traffic in a nondefault VRF:

```
sfo02-Leaf01A(config)# ip vrf tenant1
sfo02-Leaf01A(conf-vrf)# exit
```

21. Configure the anycast gateway MAC address

```
sfo02-Leaf01A(config)# ip virtual-router mac-address 00:01:01:01:01:01
```

22. Configure routing on virtual networks

```
sfo02-Leaf01A(config)# interface virtual-network1611
sfo02-Leaf01A(conf-if-vn-1611)# no shutdown
sfo02-Leaf01A(conf-if-vn-1611)# mtu 9216
sfo02-Leaf01A(conf-if-vn-1611)# ip vrf forwarding tenant1
sfo02-Leaf01A(conf-if-vn-1611)# ip address 172.16.11.248/24
sfo02-Leaf01A(conf-if-vn-1611)# ip virtual-router address 172.16.11.253
sfo02-Leaf01A(conf-if-vn-1611)# exit
sfo02-Leaf01A(config)# interface virtual-network1612
sfo02-Leaf01A(conf-if-vn-1612)# no shutdown
sfo02-Leaf01A(conf-if-vn-1612)# mtu 9216
sfo02-Leaf01A(conf-if-vn-1612)# ip vrf forwarding tenant1
sfo02-Leaf01A(conf-if-vn-1612)# ip address 172.16.12.248/24
sfo02-Leaf01A(conf-if-vn-1612)# ip virtual-router address 172.16.12.253
sfo02-Leaf01A(conf-if-vn-1612)# exit
sfo02-Leaf01A(config)# interface virtual-network1613
sfo02-Leaf01A(conf-if-vn-1613)# no shutdown
sfo02-Leaf01A(conf-if-vn-1613)# mtu 9216
sfo02-Leaf01A(conf-if-vn-1613)# ip vrf forwarding tenant1
sfo02-Leaf01A(conf-if-vn-1613)# ip address 172.16.13.248/24
sfo02-Leaf01A(conf-if-vn-1613)# ip virtual-router address 172.16.13.253
sfo02-Leaf01A(conf-if-vn-1613)# exit
sfo02-Leaf01A(config)# interface virtual-network1711
sfo02-Leaf01A(conf-if-vn-1711)# no shutdown
sfo02-Leaf01A(conf-if-vn-1711)# mtu 9216
sfo02-Leaf01A(conf-if-vn-1711)# ip vrf forwarding tenant1
sfo02-Leaf01A(conf-if-vn-1711)# ip address 172.17.11.248/24
sfo02-Leaf01A(conf-if-vn-1711)# ip virtual-router address 172.17.11.253
sfo02-Leaf01A(conf-if-vn-1711)# exit
sfo02-Leaf01A(config)# interface virtual-network1713
sfo02-Leaf01A(conf-if-vn-1713)# no shutdown
sfo02-Leaf01A(conf-if-vn-1713)# mtu 9216
sfo02-Leaf01A(conf-if-vn-1713)# ip vrf forwarding tenant1
sfo02-Leaf01A(conf-if-vn-1713)# ip address 172.17.13.248/24
sfo02-Leaf01A(conf-if-vn-1713)# ip virtual-router address 172.17.13.253
sfo02-Leaf01A(conf-if-vn-1713)# exit
sfo02-Leaf01A(config)# interface virtual-network3939
sfo02-Leaf01A(conf-if-vn-3939)# no shutdown
sfo02-Leaf01A(conf-if-vn-3939)# ip vrf forwarding tenant1
sfo02-Leaf01A(conf-if-vn-3939)# exit
```

23. Repeat these, using the appropriate values from Section 8.2, for the remaining leaf switches in AZ2.

8.2.4 Configure the first spine switch in AZ2

This section covers the configuration of the S5232F-ON switch using the `sfo02-Spine01` hostname as shown in Figure 31.

Note: All switch configuration commands are provided in the file attachments. See Section 1.8 for instructions on accessing the attachments.

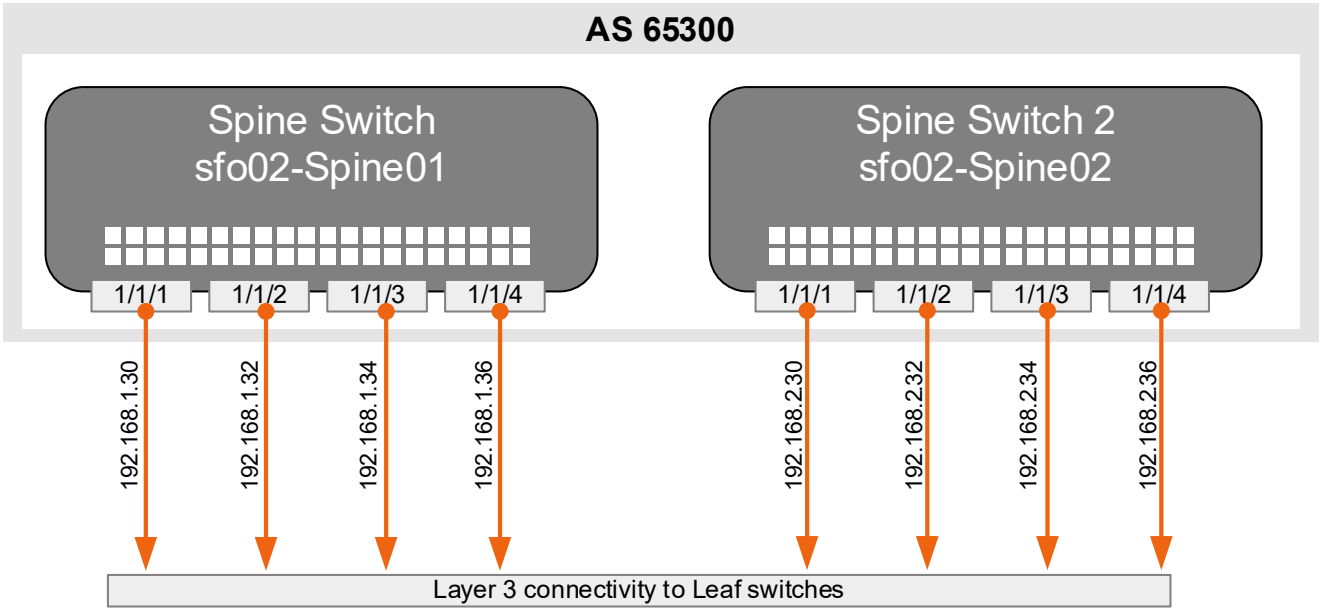


Figure 31 Region A, AZ2, Rack 1, spine layer diagram

1. Configure general switch settings including management and NTP source.

```
OS10# configure terminal
OS10(config)# interface mgmt 1/1/1
OS10(conf-if-ma-1/1/1)# no ip address dhcp
OS10(conf-if-ma-1/1/1)# ip address 100.67.167.33/24
OS10(conf-if-ma-1/1/1)# exit
OS10(config)# management route 100.67.0.0/16 managementethernet
OS10(config)# hostname sfo02-Spine01
sfo02-Spine01(config)# ntp server 100.67.10.20
sfo02-Spine01(config)# hardware forwarding-table mode scaled-l3-routes
sfo02-Spine01(config)# bfd enable
```

2. Configure a loopback interface for the Router ID

```
sfo02-Spine01(config)# interface loopback 0
sfo02-Spine01(conf-if-lo-0)# description Router-ID
sfo02-Spine01(conf-if-lo-0)# no shutdown
sfo02-Spine01(conf-if-lo-0)# ip address 10.0.1.31/32
sfo02-Spine01(conf-if-lo-0)# exit
```

3. Configure downstream ports on underlay links to leaf switches

```
sfo02-Spine01(config)# interface ethernet1/1/1
sfo02-Spine01(conf-if-eth1/1/1)# description sfo02-Leaf01A
sfo02-Spine01(conf-if-eth1/1/1)# no shutdown
sfo02-Spine01(conf-if-eth1/1/1)# no switchport
sfo02-Spine01(conf-if-eth1/1/1)# mtu 9216
sfo02-Spine01(conf-if-eth1/1/1)# ip address 192.168.1.30/31
sfo02-Spine01(conf-if-eth1/1/1)# exit
sfo02-Spine01(config)# interface ethernet1/1/2
sfo02-Spine01(conf-if-eth1/1/2)# description sfo02-Leaf01B
sfo02-Spine01(conf-if-eth1/1/2)# no shutdown
sfo02-Spine01(conf-if-eth1/1/2)# no switchport
sfo02-Spine01(conf-if-eth1/1/2)# mtu 9216
sfo02-Spine01(conf-if-eth1/1/2)# ip address 192.168.1.32/31
sfo02-Spine01(conf-if-eth1/1/2)# exit
sfo02-Spine01(config)# interface ethernet1/1/3
sfo02-Spine01(conf-if-eth1/1/3)# description sfo02-Leaf02A
sfo02-Spine01(conf-if-eth1/1/3)# no shutdown
sfo02-Spine01(conf-if-eth1/1/3)# no switchport
sfo02-Spine01(conf-if-eth1/1/3)# mtu 9216
sfo02-Spine01(conf-if-eth1/1/3)# ip address 192.168.1.34/31
sfo02-Spine01(conf-if-eth1/1/3)# exit
sfo02-Spine01(config)# interface ethernet1/1/4
sfo02-Spine01(conf-if-eth1/1/4)# description sfo02-Leaf02B
sfo02-Spine01(conf-if-eth1/1/4)# no shutdown
sfo02-Spine01(conf-if-eth1/1/4)# no switchport
sfo02-Spine01(conf-if-eth1/1/4)# mtu 9216
sfo02-Spine01(conf-if-eth1/1/4)# ip address 192.168.1.36/31
sfo02-Spine01(conf-if-eth1/1/4)# exit
```

4. Add a route map

```
sfo02-Spine01(config)# ip prefix-list spine-leaf seq 10 permit 10.0.1.0/24
ge 32
sfo02-Spine01(config)# ip prefix-list spine-leaf seq 20 permit 10.2.1.0/24
ge 32
sfo02-Spine01(config)# route-map spine-leaf permit 10
sfo02-Spine01(config-route-map)# match ip address prefix-list spine-leaf
sfo02-Spine01(config-route-map)# exit
```

5. Configure eBGP

```
sfo02-Spine01(config)# router bgp 65300
sfo02-Spine01(config-router-bgp-65300)# bfd all-neighbors interval 200
min_rx 200 multiplier 3 role active
sfo02-Spine01(config-router-bgp-65300)# router-id 10.0.1.31
sfo02-Spine01(config-router-bgp-65300)# address-family ipv4 unicast
sfo02-Spine01(config-router-bgpv4-af)# redistribute connected route-map
spine-leaf
sfo02-Spine01(config-router-bgp-65300)# bestpath as-path multipath-relax
sfo02-Spine01(config-router-bgp-65300)# maximum-paths ebgp 2
sfo02-Spine01(config-router-bgpv4-af)# exit
```

6. Configure eBGP for IPv4 point-to-point peering

```
sfo02-Spine01(config-router-bgp-65300)# neighbor 192.168.1.31
sfo02-Spine01(config-router-neighbor)# remote-as 65101
sfo02-Spine01(config-router-neighbor)# no shutdown
sfo02-Spine01(config-router-neighbor)# advertisement-interval 5
sfo02-Spine01(config-router-neighbor)# bfd
sfo02-Spine01(config-router-neighbor)# fall-over
sfo02-Spine01(config-router-neighbor)# exit
sfo02-Spine01(config-router-bgp-65300)# neighbor 192.168.1.33
sfo02-Spine01(config-router-neighbor)# remote-as 65101
sfo02-Spine01(config-router-neighbor)# no shutdown
sfo02-Spine01(config-router-neighbor)# advertisement-interval 5
sfo02-Spine01(config-router-neighbor)# bfd
sfo02-Spine01(config-router-neighbor)# fall-over
sfo02-Spine01(config-router-neighbor)# exit
sfo02-Spine01(config-router-bgp-65300)# neighbor 192.168.1.35
sfo02-Spine01(config-router-neighbor)# remote-as 65102
sfo02-Spine01(config-router-neighbor)# no shutdown
sfo02-Spine01(config-router-neighbor)# advertisement-interval 5
sfo02-Spine01(config-router-neighbor)# bfd
sfo02-Spine01(config-router-neighbor)# fall-over
sfo02-Spine01(config-router-neighbor)# exit
sfo02-Spine01(config-router-bgp-65300)# neighbor 192.168.1.37
sfo02-Spine01(config-router-neighbor)# remote-as 65102
sfo02-Spine01(config-router-neighbor)# no shutdown
sfo02-Spine01(config-router-neighbor)# advertisement-interval 5
sfo02-Spine01(config-router-neighbor)# bfd
sfo02-Spine01(config-router-neighbor)# fall-over
sfo02-Spine01(config-router-neighbor)# exit
sfo02-Spine01(config-router-bgp-65300)# exit
```

7. Configure a loopback interface for BGP EVPN peering

```
sfo02-Spine01(config)# interface loopback 1
sfo02-Spine01(conf-if-lo-1)# description evpn_loopback
sfo02-Spine01(conf-if-lo-1)# no shutdown
sfo02-Spine01(conf-if-lo-1)# ip address 10.2.1.31/32
sfo02-Spine01(conf-if-lo-1)# exit
```

8. Configure BGP EVPN peering

```
sfo02-Spine01(config)# router bgp 65300
sfo02-Spine01(config-router-bgp-65300)# neighbor 10.2.2.31
sfo02-Spine01(config-router-neighbor)# remote-as 65301
sfo02-Spine01(config-router-neighbor)# send-community extended
sfo02-Spine01(config-router-neighbor)# update-source loopback1
sfo02-Spine01(config-router-neighbor)# ebgp-multihop 2
sfo02-Spine01(config-router-neighbor)# no shutdown
sfo02-Spine01(config-router-neighbor)# address-family ipv4 unicast
sfo02-Spine01(config-router-neighbor-af)# no activate
sfo02-Spine01(config-router-neighbor-af)# exit
sfo02-Spine01(config-router-neighbor)# address-family l2vpn evpn
sfo02-Spine01(config-router-neighbor-af)# activate
sfo02-Spine01(config-router-neighbor-af)# exit
sfo02-Spine01(config-router-neighbor)# exit
sfo02-Spine01(config-router-bgp-65300)# neighbor 10.2.2.32
sfo02-Spine01(config-router-neighbor)# remote-as 65301
sfo02-Spine01(config-router-neighbor)# send-community extended
sfo02-Spine01(config-router-neighbor)# update-source loopback1
sfo02-Spine01(config-router-neighbor)# ebgp-multihop 2
sfo02-Spine01(config-router-neighbor)# no shutdown
sfo02-Spine01(config-router-neighbor)# address-family ipv4 unicast
sfo02-Spine01(config-router-neighbor-af)# no activate
sfo02-Spine01(config-router-neighbor-af)# exit
sfo02-Spine01(config-router-neighbor)# address-family l2vpn evpn
sfo02-Spine01(config-router-neighbor-af)# activate
sfo02-Spine01(config-router-neighbor-af)# exit
sfo02-Spine01(config-router-neighbor)# exit
sfo02-Spine01(config-router-bgp-65300)# neighbor 10.2.2.33
sfo02-Spine01(config-router-neighbor)# remote-as 65302
sfo02-Spine01(config-router-neighbor)# send-community extended
sfo02-Spine01(config-router-neighbor)# update-source loopback1
sfo02-Spine01(config-router-neighbor)# ebgp-multihop 2
sfo02-Spine01(config-router-neighbor)# no shutdown
sfo02-Spine01(config-router-neighbor)# address-family ipv4 unicast
sfo02-Spine01(config-router-neighbor-af)# no activate
sfo02-Spine01(config-router-neighbor-af)# exit
sfo02-Spine01(config-router-neighbor)# address-family l2vpn evpn
sfo02-Spine01(config-router-neighbor-af)# activate
sfo02-Spine01(config-router-neighbor-af)# exit
sfo02-Spine01(config-router-neighbor)# exit
sfo02-Spine01(config-router-bgp-65300)# neighbor 10.2.2.34
sfo02-Spine01(config-router-neighbor)# remote-as 65302
```

```
sfo02-Spine01(config-router-neighbor)# send-community extended
sfo02-Spine01(config-router-neighbor)# update-source loopback1
sfo02-Spine01(config-router-neighbor)# ebgp-multihop 2
sfo02-Spine01(config-router-neighbor)# no shutdown
sfo02-Spine01(config-router-neighbor)# address-family ipv4 unicast
sfo02-Spine01(config-router-neighbor-af)# no activate
sfo02-Spine01(config-router-neighbor-af)# exit
sfo02-Spine01(config-router-neighbor)# address-family l2vpn evpn
sfo02-Spine01(config-router-neighbor-af)# activate
sfo02-Spine01(config-router-neighbor-af)# exit
sfo02-Spine01(config-router-neighbor)# exit
sfo02-Spine01(config-router-bgp-65300)# exit
```

9. Repeat these, using the appropriate values from Section 3.3, for the remaining spine switches in AZ1.

8.2.5 Verify establishment of BGP and EVPN between availability zones

The leaf switches must establish a connection to the spine switches before BGP updates can be exchanged. Verify that peering is successful and BGP routing has been established.

1. Run the `show ip bgp summary` command to display information about the BGP and TCP connection to neighbors. In Figure 32 all three BGP sessions for each leaf switch are shown. The first two sessions represent the spine switches and the last session, 192.168.3.31 is the iBGP session between the leaf pairs in the event of leaf to spine layer failure.

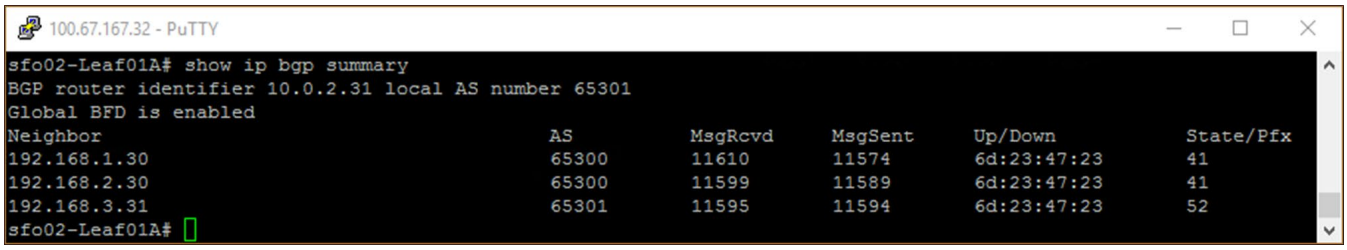


Figure 32 Output of show ip bgp summary

Note: ESG configuration is like AZ1 configuration. For information, see the *Deployment for Multiple Availability Zones* guide.

For the L2 VXLAN virtual networks to communicate, each leaf must be able to establish a connection to the other leaf switches before host MAC information can be exchanged. Verify that peering is successful and BGP EVPN routing has been established.

2. Run the `show evpn evi` command to verify the current state of all configured virtual networks. Figure 33 shows the state of each virtual network as `Up` and that IRB VRF is set as `tenant1`.

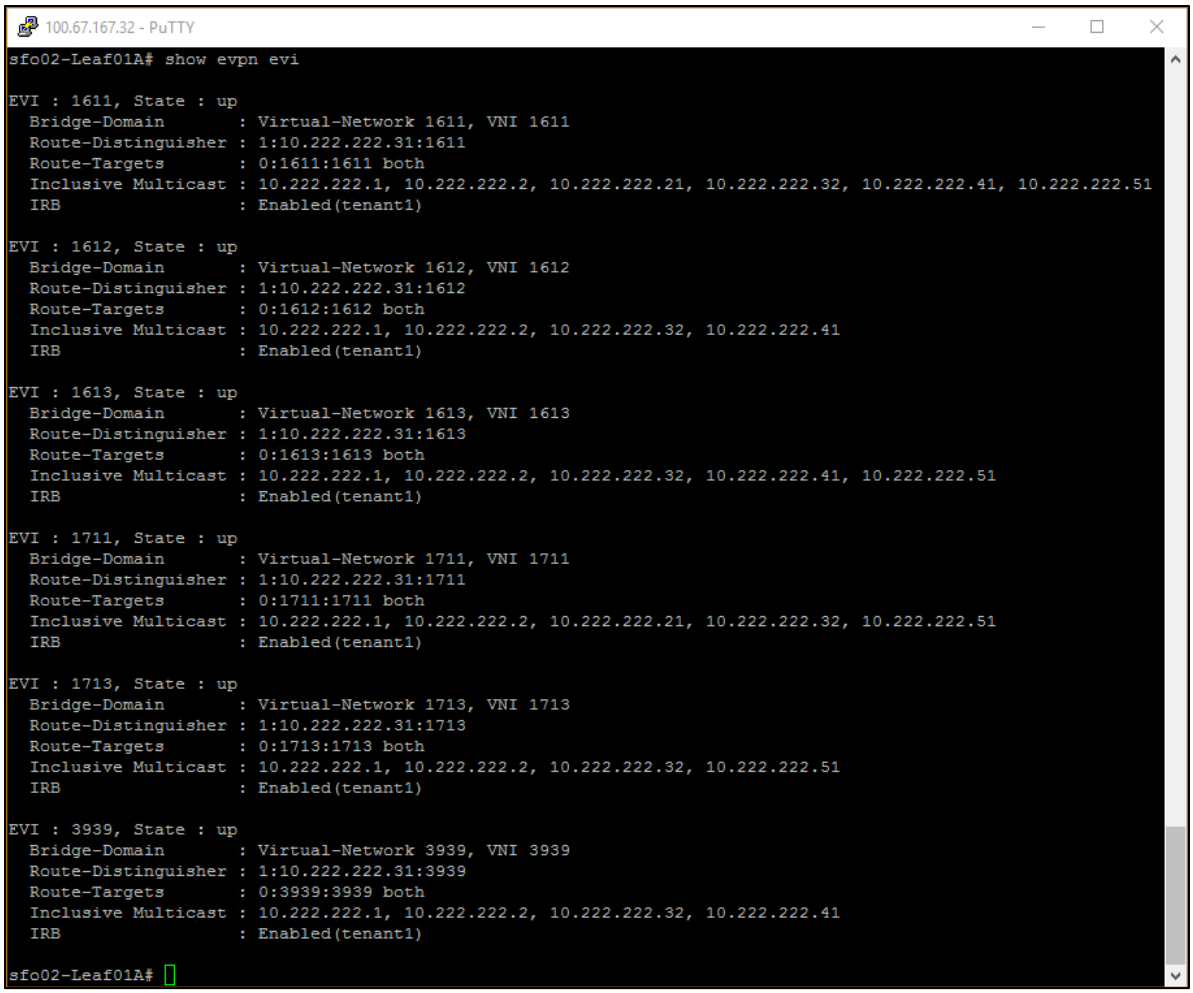


Figure 33 Output of show evpn evi from sfo02-Leaf01A

8.1 Next Steps

This section contains general guidance and validation of expanding a VxRail cluster to AZ2 as well as stretching vSAN and adding extra NSX components. The steps including:

1. Connect and power on the VxRail nodes to the leaf switches in AZ2
2. Expand the management cluster using the VxRail manager running on AZ1
3. Complete the steps for dual availability zones include:

    a. Deploying the vSAN witness in Region B
    b. Configuration of vSAN fault domains
    c. Deploy extra NSX ESGs to AZ2
    d. Create Anti-affinity rules for ESGs in AZ2

8.1.1 Expand the management cluster using the VxRail manager

Before expanding the cluster, ensure that:

- All appropriate networking configuration has been completed and validated
- All new hosts in AZ2 are properly connected and powered on

Expanding a VxRail VCF cluster in AZ1 can be done using the VxRail Manager vCenter Plugin. For information about using the VxRail manager to add more nodes, see the *Interactive Demo for VCF on Dell EMC VxRail.* Figure 34 shows the inventory view from VxRail Manager after the VxRail nodes in AZ2 have been added to the existing cluster in AZ1. The drop-down along the right side shows eight nodes between the two availability zones.

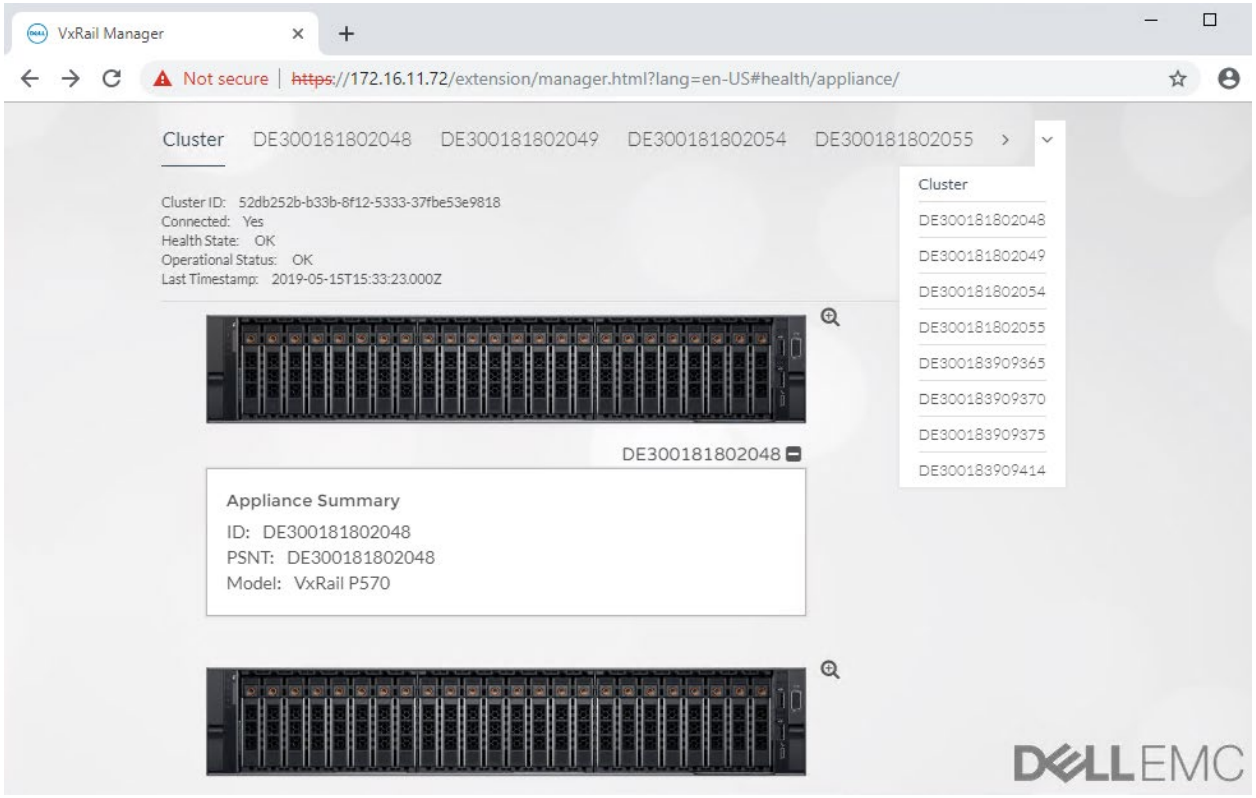


Figure 34 Expanded cluster including four E560 (AZ1) and four P570 (AZ2) nodes

8.1.2 Deployment for multiple availability zones

To complete the stretched vSAN deployment and configuring more NSX components, see the *Deployment for Multiple Availability Zones* guide.

Figure 35 shows the completed VxRail VCF SDDC management workload. The NSX networking components, including the four ESGs and two DLRs, are shown under `the sfo01-m01-sddc-edge` resource pool.

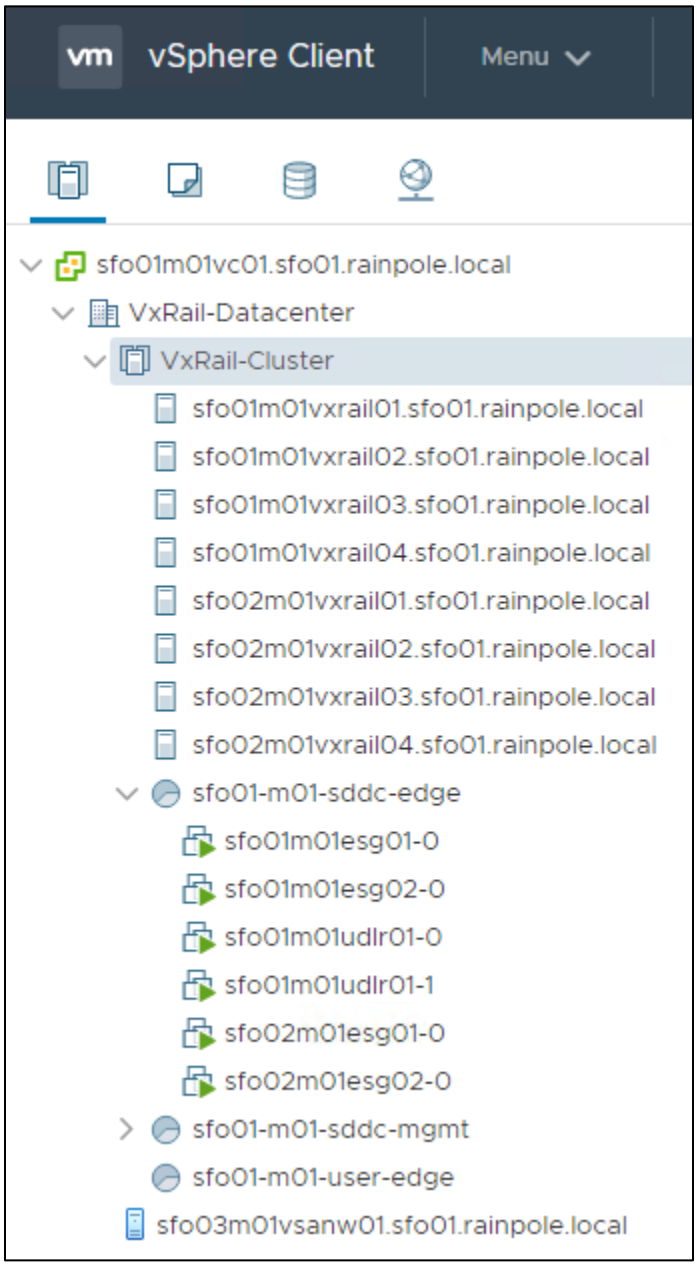


Figure 35 Stretched cluster with site-specific ESGs

A Validated components

This section contains the specific firmware and software revisions that are used in this guide.

A.1 Dell EMC PowerSwitch models

Table 15 Switches and operating system versions

Qty	Item	Version
8	Dell EMC PowerSwitch S5248F-ON leaf switches	10.4.3.1.154
2	Dell EMC PowerSwitch Z9264F-ON spine switches	10.4.3.1.154
2	Dell EMC PowerSwitch S5232F-ON spine switches	10.4.3.1.154
2	Dell EMC PowerSwitch S3048-ON OOB mgmt switches	10.4.3.1.154

A.2 VxRail E560 and P570 nodes

A cluster of four VxRail E560 nodes (AZ1) and four VxRail P570 (AZ2) was used to validate the examples in this guide. The nodes were each configured using the information that is provided in Table 16. The values below can also be found in the *VxRail Support Matrix.*

Table 16 VxRail node components

Qty per node	Item	Firmware version
2	Intel Xeon Gold 6136 CPU @ 3.00GHz, 12 cores	-
12	16GB DDR4 DIMMs (192GB total)	-
2	800GB SAS SSD	-
8	1.2TB SAS HDD	EF03
1	Dell HBA330 Storage Controller	15.17.09.06
1	Boot Optimized Storage Solution (BOSS) Controller w/ 2x240GB SATA SSDs	2.5.13.3016
1	Broadcom 57414 NDC – 2x25GbE SFP28 ports	21.40.16.60
-	BIOS	1.6.12
-	iDRAC with Lifecycle Controller	3.21.26.22

A.3 Appliance software

This deployment guide was developed using VxRail appliance software 4.7.111. The software consists of the component versions provided in Table 17.

Table 17 VxRail appliance software component versions

Item	Version
VxRail Manager	4.7.111.13048811
ESRS	3.28.00.06
Log Insight	4.6.0.8080673
VMware vCenter	6.7.0 build 11726888
VMware ESXi	6.7.0 build 13004448
Platform Service	4.7.111
NSX	6.4.411197766

B Technical resources

B.1 VxRail, VCF, and VVD Guides

VMware Cloud Foundation on VxRail Planning and Preparation Guide

VMware Cloud Foundation on VxRail Architecture Guide

Dell EMC VxRail Network Guide

VxRail Appliance 4.7.100 Administration Guide

VxRail Planning Guide for Virtual SAN Stretched Cluster

VMware Cloud Foundation on VxRail Technical FAQ

VMware Validated Designs v5.0 on Dell EMC VxRail for a Single Region: VMware Cloud Builder Deployment for Region A

VMware Cloud Foundation on VxRail Administrator Guide

VxRail Support Matrix

B.2 Dell EMC Networking Guides

Dell EMC PowerSwitch Guides

OS10 Enterprise Edition Users Guide 10.4.3.0 Edition

Dell EMC Networking Layer 3 Leaf-Spine Deployment and Best Practices with OS10EE

Dell EMC Networking Virtualization Overlay with BGP EVPN

Dell EMC OS10EE BGP EVPN Configuration Cheat Sheet

Dell EMC VxRail Multirack Deployment Guide

OS10 VLT Deployment and Best Practices Guides

C Support and feedback

Contacting Technical Support

Support Contact Information	Web: http://www.dell.com/support
	Telephone: USA: 1-800-945-3355

Feedback for this document

To provide feedback on the quality and usefulness of this publication, send an email to Dell_Networking_Solutions@Dell.com.