

# Dell EMC Validated System for Virtualization - NSX Reference Architecture

A step-by-step VMware NSX deployment on a leaf-spine data center network with FC630 compute nodes and iSCSI shared storage.

Dell Networking Solutions Engineering  
February 2017

## Revisions

Date	Revision	Description	Authors
February 2017	1.0	Initial Release	Jim Slaughter, Curtis Bunch

THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

Copyright © 2017 Dell Inc. or its subsidiaries. All rights reserved. Dell and the Dell EMC logo are trademarks of Dell Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

# Table of contents

Revisions.....	2
1 Introduction.....	8
1.1 Validated System for Virtualization.....	8
1.1.1 Addressing the need for flexibility.....	8
1.1.2 Differentiated approach addresses challenges and limitations .....	9
1.2 VMware NSX .....	9
1.3 The VXLAN protocol.....	10
2 Hardware overview.....	12
2.1 Dell PowerEdge FX2s enclosure and supported modules.....	12
2.1.1 PowerEdge FC630 server .....	13
2.1.2 PowerEdge FN410S I/O Module .....	13
2.2 PowerEdge R630 server .....	13
2.3 Dell Networking Z9100-ON.....	14
2.4 Dell Networking S4048-ON .....	14
2.5 Dell Networking S3048-ON .....	14
2.6 Dell Storage SC4020 storage array.....	15
2.7 Dell Storage SC220 expansion enclosure.....	15
3 Topology.....	16
3.1 Servers .....	16
3.2 Production network.....	16
3.2.1 Physical data center network (underlay) .....	17
3.2.2 NSX virtual network (overlay) .....	18
3.2.3 Combined physical and virtual networks.....	19
3.2.4 iSCSI SAN .....	20
3.3 Management network .....	21
4 Network connections .....	22
4.1 Production network connections.....	22
4.1.1 Management cluster – data center network .....	22
4.1.2 Compute cluster – data center network.....	23
4.1.3 Compute cluster – iSCSI SAN.....	24
4.1.4 Edge cluster – data center network .....	25

4.2	Management network connections.....	26
4.2.1	Management and edge clusters .....	26
4.2.2	Compute cluster.....	27
5	Leaf-spine topology .....	28
5.1	Routing protocol selection .....	28
5.2	BGP ASN configuration .....	29
5.3	BGP fast fall-over.....	29
5.4	IP Address Management.....	30
5.4.1	Loopback addresses .....	30
5.4.2	Point-to-point addresses.....	31
5.4.3	VLANs and IP addressing .....	33
5.5	VRRP.....	33
5.6	ECMP .....	34
5.7	VLT .....	34
5.8	Uplink Failure Detection .....	35
6	Configure physical switches .....	36
6.1	Factory default settings .....	36
6.2	FN410S switch configuration.....	37
6.3	S4048-ON leaf switch configuration .....	42
6.3.1	S4048-ON edge switch configuration.....	46
6.4	Z9100-ON spine switch configuration.....	48
6.5	S4048-ON iSCSI SAN switch configuration .....	51
6.6	S3048-ON management switch configuration.....	53
6.7	Verify switch configuration.....	53
6.7.1	Z9100-ON spine switch .....	53
6.7.2	S4048-ON leaf switch .....	55
6.7.3	FN410S I/O Module.....	57
6.7.4	S4048-ON iSCSI SAN switch .....	58
7	Prepare Servers .....	59
7.1	Confirm CPU virtualization is enabled in BIOS .....	59
7.2	Confirm network adapters are at factory default settings.....	59
7.3	Install ESXi .....	60

7.4	Configure the ESXi management network connection.....	60
8	Deploy VMware vCenter Server and add hosts .....	61
8.1	Deploy VMware vCenter Server .....	61
8.2	Connect to the vSphere Web Client .....	63
8.3	Install VMware licenses .....	63
8.4	Create a data center object and add hosts .....	64
8.5	Ensure hosts are configured for NTP .....	66
8.6	Create clusters and add hosts .....	67
8.7	Information on vSphere standard switches .....	68
9	Deploy vSphere distributed switches for vMotion .....	70
9.1	Create a VDS for each cluster .....	70
9.2	Add distributed port groups .....	71
9.3	Create LACP LAGs .....	73
9.4	Associate hosts and assign uplinks to LAGs.....	74
9.5	Configure teaming and failover on LAGs .....	78
9.6	Add VMkernel adapters for vMotion .....	79
9.7	Verify VDS configuration .....	81
9.8	Enable LLDP.....	82
9.8.1	Enable LLDP on each VDS and view information sent .....	82
9.8.2	View LLDP information received from physical switch .....	83
10	Configure iSCSI Storage .....	84
10.1	Enable iSCSI offloading on QLogic 57810 adapters .....	84
10.2	Storage Center SC4020 port configuration .....	85
10.2.1	Configure iSCSI port IP addresses.....	86
10.2.2	Create fault domains .....	87
10.2.3	Verify the Storage Center configuration .....	89
10.2.4	Tag the fault domain control ports in the correct VLANs .....	90
10.3	Verify iSCSI SAN switch configuration .....	91
10.4	Create an iSCSI VDS for the compute cluster .....	92
10.5	Set the iSCSI VDS MTU to 9000 and enable LLDP .....	92
10.6	Add distributed port groups .....	93
10.7	Configure teaming and failover.....	95

10.8	Associate hosts and assign uplinks .....	96
10.9	Add VMkernel adapters for iSCSI.....	97
10.10	Increase the MTU to 9000 on iSCSI VMkernel adapters .....	99
10.11	Bind iSCSI adapters with VMkernel ports .....	99
10.12	Configure dynamic discovery .....	100
10.13	SC4020 final configuration .....	101
10.13.1	Create servers in Storage Center .....	101
10.13.2	Create Storage Center server cluster .....	104
10.13.3	Create and map shared storage volume.....	105
10.14	Verify hosts are connected to storage.....	106
10.15	Create a datastore .....	107
11	Configure the NSX virtual network .....	109
11.1	NSX Manager .....	109
11.2	Register NSX Manager with vCenter Server .....	110
11.3	Deploy NSX controllers .....	111
11.4	Prepare host clusters for NSX .....	114
11.5	Configure clusters for VXLAN.....	115
11.6	Create a segment ID pool.....	117
11.7	Add a transport zone .....	117
11.8	Logical switch configuration.....	119
11.9	Distributed Logical Router configuration .....	121
11.9.1	Configure OSPF on the DLR .....	124
11.9.2	Firewall information .....	125
12	Verify NSX network functionality .....	126
12.1	Deploy virtual machines .....	126
12.2	Connect virtual wires .....	127
12.3	Configure networking in the guest OS.....	128
12.4	Test connectivity.....	128
13	Communicate outside the virtual network .....	129
13.1	Edge Services Gateway .....	129
13.1.1	Add a distributed port group .....	130
13.1.2	Create second LACP LAG.....	130

13.1.3 Assign uplinks to the second LAG .....	132
13.1.4 Configure port groups for teaming and failover .....	134
13.1.5 Deploy the Edge Services Gateway .....	135
13.1.6 Configure OSPF on the ESG.....	137
13.1.7 High Availability configuration.....	138
13.1.8 ESG validation .....	141
13.2 Hardware VTEP .....	143
13.2.1 Configure additional connections on spine switches .....	144
13.2.2 Configure the hardware VTEP and connect to NSX .....	145
13.2.3 Create a logical switch.....	150
13.2.4 Configure a replication cluster .....	152
13.2.5 Hardware VTEP Validation .....	153
14 Scaling guidance .....	156
14.1 Switch selection .....	156
14.2 iSCSI Storage sizing.....	156
14.3 Example – scale out to 3000 virtual machines .....	156
14.4 Port count and oversubscription (leaf-spine topology) .....	158
14.5 Rack diagrams.....	159
A Dell EMC validated hardware and components .....	161
A.1 Switches .....	161
A.2 PowerEdge R630 servers.....	161
A.3 PowerEdge FX2s chassis and components .....	162
A.4 Dell Storage Center SC4020 storage array.....	162
B Dell EMC validated software and required licenses .....	163
B.1 Software.....	163
B.2 Licenses.....	163
C Technical support and resources .....	164
C.1 Dell EMC product manuals and technical guides.....	164
C.2 VMware product manuals and technical guides.....	164
D Support and Feedback .....	165

# 1 Introduction

This guide covers an NSX deployment for the data center based on the Dell EMC Validated System for Virtualization.

The goal of this guide is to enable a network administrator or engineer with traditional networking and VMware ESXi experience to build a scalable NSX virtual network using the Dell EMC Validated System for Virtualization hardware and software outlined in this guide.

This document provides a best practice leaf-spine topology with configuration steps for all physical switches in the topology. It includes step-by-step configuration of a virtual network using VMware NSX that overlays the physical network. This includes configuration of logical switches, routers and options for communicating with external traditional networks using software and hardware solutions. It also includes steps to deploy ESXi on PowerEdge servers, deployment of a vSphere vCenter Server Appliance and configuration of shared storage on an iSCSI storage area network (SAN).

**Note:** See the appendices for product versions validated.

## 1.1 Validated System for Virtualization

The Dell EMC Validated System for Virtualization is the industry's most flexible converged system to date. The system enables network architects to choose which compute, storage and networking building blocks to test for integration and interoperability in support of virtualized environments.

The system incorporates a wide range of form factors, technology choices and deployment options, right-sized to fit each customer's needs. A fully-validated system can be configured, quoted and ordered in minutes. Automated lifecycle management tools allow customers to easily deploy, scale and update the system.

### 1.1.1 Addressing the need for flexibility

Customers face unprecedented pressures to improve efficiency and lower costs while balancing increasing business demands against decreasing IT budgets. The current operational model of delivering IT services--procuring technology from best-of-breed providers and managing them in isolation--proves to be time consuming and problematic. This approach typically burdens customers to make design decisions, validate components, set-up and configure components, and manage the environment going forward. In turn, this involves engaging multiple vendors for assistance with infrastructure elements that, over time, increase complexity and cost.

Existing integrated solutions to these challenges are either pre-integrated and prepackaged offers or traditional reference architectures. The former optimizes time-to-production and simplifies ongoing operations, with customers making a tradeoff on flexibility and choice. The latter provide some degree of flexibility but do not offer manageability or scalability benefits.

The Dell EMC Validated System for Virtualization bridges this gap by offering an integrated system that is tested and validated. The system is highly flexible, scalable and driven, using end-to-end automation throughout the infrastructure lifecycle.



### 1.1.2 Differentiated approach addresses challenges and limitations

To provide IT services faster, while lowering costs and streamlining operations, Dell EMC engineered the Validated System for Virtualization. This groundbreaking system enables greater operational efficiencies and savings--and unparalleled management simplicity--by giving you more power than ever to define and design it.

The system includes options from “do-it-yourself” using a deployment guide, to an on-site system integration by Dell EMC, to using your own integration vendor.

The Dell EMC Validated System for Virtualization is:

- Built on our best-of-breed products designed for virtualization across the ecosystem.
- Tested, validated and fully integrated, yet flexible enough to be tailored for your organization, removing risk and accelerating your time to value.
- Delivered with Dell EMC's Active System Manager (ASM) to simplify ongoing management.
- Delivered with Dell EMC's global reach, exceptional execution and delivery, providing consistent deployment, management and maintenance in every region of the world.
- Delivered with a single point-of-support for the complete system including hardware and software through Dell ProSupport Plus.

Information about the Dell EMC Validated System for Virtualization is available [here](#).

## 1.2 VMware NSX

VMware NSX enables network virtualization. With NSX, logical networks are created on top of a basic layer 2 (switched) or layer 3 (routed) physical infrastructure. This allows the physical and virtual environments to be decoupled, enabling agility and security in the virtual environment while allowing the physical environment to focus on throughput.

The NSX platform also provides for network services in the logical space. Some of these logical services include switching, routing, firewalling, load balancing and Virtual Private Network (VPN) services.

NSX benefits include the following:

- Simplified network service deployment, migration and automation
- Reduced provisioning and deployment time
- Scalable multi-tenancy across one or more data centers
- Distributed routing and a distributed firewall at the hypervisor allow for better east-to-west traffic flow and an enhanced security model
- Provides solutions for traditional networking problems such as limited VLANs, MAC address, FIB and ARP entries
- Application requirements do not require modification to the physical network
- Normalization of underlying hardware, enabling easier hardware migration and interoperability

## 1.3 The VXLAN protocol

NSX creates logical networks using the Virtual Extensible Local Area Network (VXLAN) protocol. The VXLAN protocol is described in Internet Engineering Task Force document [RFC 7348](#). VXLAN allows a layer 2 network to scale across the data center by overlaying a layer 3 network. Each overlay is referred to as a VXLAN segment and only virtual machines (VMs) within the same segment can communicate with each other.

Each segment is identified through a 24-bit segment ID referred to as a VXLAN Network Identifier (VNI). This allows up to 16 million VXLAN segment IDs, far greater 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 1:

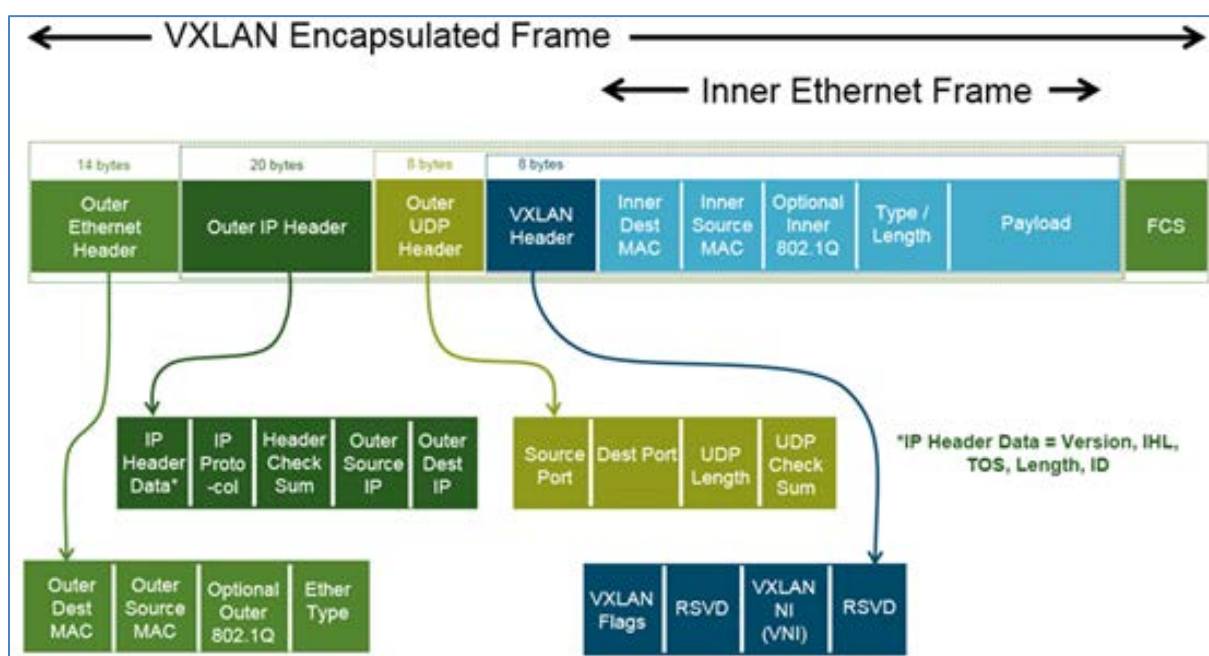


Figure 1 VXLAN encapsulated frame

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 bytes on all participating interfaces.

As part of the VXLAN configuration, each ESXi host is configured with a software VXLAN tunnel end point (VTEP). A software VTEP is a VMkernel interface where VXLAN encapsulation and de-encapsulation occurs.

A physical switch that supports VXLAN can act as a hardware VTEP, also referred to as a VXLAN Gateway (Section 13.2). This allows communication with servers inside the data center that are outside of the virtual network.

## 1.4 Typographical conventions

This document uses the following typographical conventions:

Monospace text

Command Line Interface (CLI) examples

**Monospace text**

Commands entered at the CLI prompt

*Monospace text*

Variables in CLI examples

## 2 Hardware overview

While the Dell EMC Validated System for Virtualization has flexibility and choice across servers, storage and networking, this guide is focused on a single instance of the system. This section briefly describes the primary hardware used to validate this deployment. A complete listing of hardware validated for this guide is provided in Appendix A.

### 2.1 Dell PowerEdge FX2s enclosure and supported modules

The PowerEdge FX2s enclosure is a 2-rack unit (RU) computing platform. It has capacity for two FC830 full-width servers, four FC630 half-width servers or eight FC430 quarter-width servers. The enclosure is also available with a combination of servers and storage sleds. The FX2s enclosure used in this guide contains four FC630 servers as shown in Figure 2.



Figure 2 Dell PowerEdge FX2s (front) with four PowerEdge FC630 servers

The back of the FX2s enclosure includes two I/O networking modules (IOMs) and eight PCIe expansion slots.



Figure 3 Dell PowerEdge FX2s (back) with two PowerEdge FN410S IOMs installed

### 2.1.1 PowerEdge FC630 server

The PowerEdge FC630 server is a half-width, 2-socket server. Four FC630 servers in the FX2s enclosure form the compute cluster for this deployment.



Figure 4 PowerEdge FC630

### 2.1.2 PowerEdge FN410S I/O Module

The PowerEdge FN410S IOM is a multilayer switch with eight internal, server-facing ports and four external, 10GbE SFP+ ports. Two FN410S IOMs installed in the FX2s enclosure provide fault tolerance.



Figure 5 PowerEdge FN410S

## 2.2 PowerEdge R630 server

The PowerEdge R630 server is a 2-socket, 1-RU server. The management and edge clusters in this guide use R630 servers.



Figure 6 PowerEdge R630

## 2.3 Dell Networking Z9100-ON

The Z9100-ON is a 1-RU, multilayer switch with thirty-two ports supporting 10/25/40/50/100GbE plus two 10GbE ports. The leaf-spine topology covered in this guide uses two Z9100-ON switches as spines.



Figure 7 Dell Networking Z9100-ON

## 2.4 Dell Networking S4048-ON

The S4048-ON is a 1-RU, layer 2/3 switch with forty-eight 10GbE SFP+ ports and six 40GbE QSFP+ ports. The leaf-spine topology covered in this guide uses six S4048-ON switches as leaf switches. Each compute cluster uses two additional S4048-ON switches for iSCSI traffic.



Figure 8 Dell Networking S4048-ON

## 2.5 Dell Networking S3048-ON

The S3048-ON is a 1-RU switch with forty-eight 1GbE Base-T ports and four 10GbE SFP+ ports. In this guide, one S3048-ON switch supports management traffic in each rack.



Figure 9 Dell Networking S3048-ON



## 2.6 Dell Storage SC4020 storage array

This guide uses an SC4020 storage array in the iSCSI SAN. It has four 10GbE ports (two per controller) for iSCSI traffic. The initial form factor is 2-RU with twenty-four 2.5" drive bays. It is expandable (using SC series expansion enclosures) to 192 drives with over 1 PB total storage capacity.



Figure 10 Dell Storage SC4020 (front)



Figure 11 Dell Storage SC4020 (back)

## 2.7 Dell Storage SC220 expansion enclosure

The SC220 expansion enclosure has a 2-RU form factor with twenty-four 2.5" drive bays. Up to seven SC220 enclosures may be added to the SC4020.



Figure 12 Dell Storage SC220 (front)



Figure 13 Dell Storage SC220 (back)

## 3 Topology

This section provides an overview of the physical and virtual topology used in this deployment.

### 3.1 Servers

The servers are grouped into three VMware vCenter clusters, with one cluster per physical rack:

- Rack 1 Management Cluster – contains three PowerEdge R630 servers
- Rack 2 Compute FC630 Cluster – contains one PowerEdge FX2s chassis with four FC630 servers.
- Rack 3 Edge Cluster – contains three PowerEdge R630 servers

The three clusters have been spread across three physical racks as shown in Figure 14 to illustrate the scalability of this design as additional servers and switches are added.

### 3.2 Production network

The production network used in this guide is divided into three major components:

- The physical, or underlay, data center network as shown in Figure 14.
- The NSX virtual, or overlay, network as shown in Figure 15.
- The iSCSI SAN as shown in Figure 17.



### 3.2.1 Physical data center network (underlay)

On the physical data center network, a leaf-spine topology is used for performance and scalability. Two leaf switches (S4048-ONs) are used in each rack for redundancy and increased performance. Dell Virtual-Link Trunking (VLT) connects each pair of leaf switches.

Each leaf switch has point-to-point connections to both spine switches (Z9100-ONs). Traffic between the leaf switches and spine switches is routed and Equal-Cost Multi-Path routing (ECMP) is leveraged to utilize all available bandwidth.

Leaf switch pairs are connected to downstream devices via VLT port channels. In Racks 1 and 3, these are direct connections to the QLogic 578xx adapters in the R630 servers. In Rack 2, Leaf Switches 3 and 4 are connected to a pair of FN410S switches in the FX2s chassis. The FN410S switches are configured for VLT, and are connected to FC630 servers inside the chassis via VLT port channels.

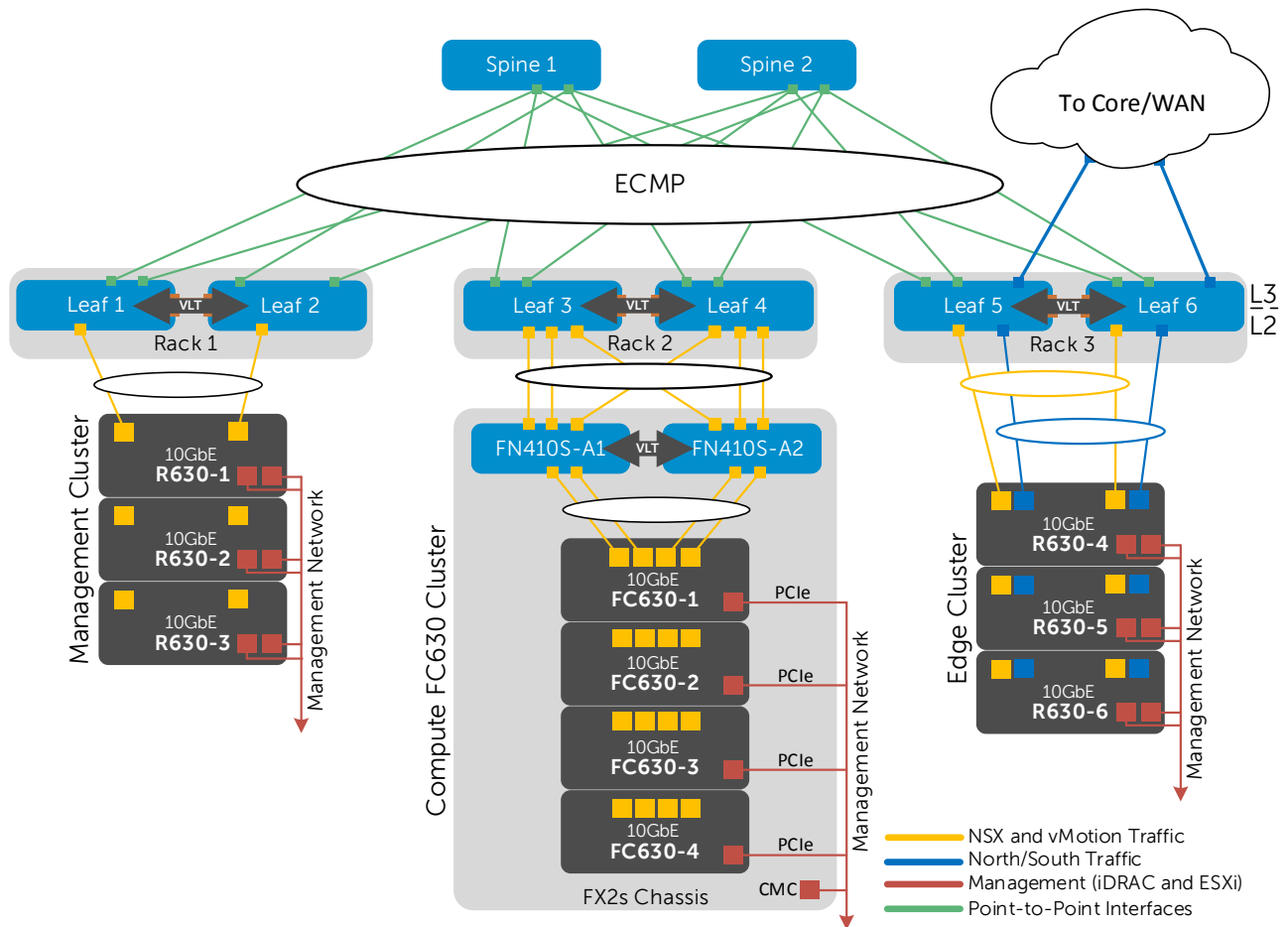


Figure 14 Physical data center network

### 3.2.2 NSX virtual network (overlay)

The virtual network, built with VMware NSX, overlays the data center network. All servers participating in the virtual network run VMware ESXi. VM-to-VM traffic is contained within the virtual network.

Traffic from the data center's virtual network to the network core or WAN (Wide Area Network) can be configured to pass through an Edge Services Gateway (ESG). This takes advantage of additional services provided by NSX, such as firewalling, load balancing and VPN services. ESG configuration is covered in Section 13.1.

Figure 15 shows the NSX virtual network built for this guide.

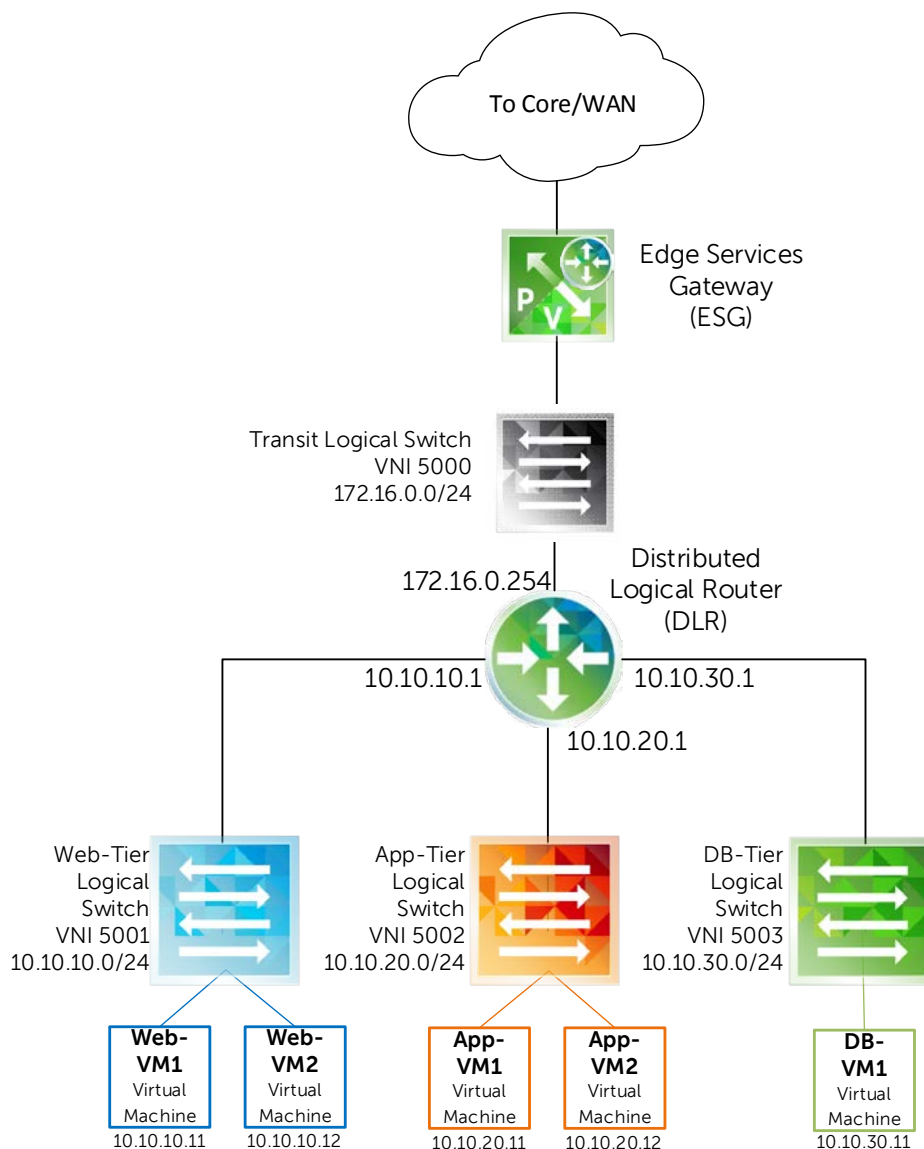


Figure 15 NSX virtual network

### 3.2.3 Combined physical and virtual networks

Figure 16 shows the combined physical data center network and virtual NSX network. All servers are running ESXi.

The management cluster in Rack 1 contains the vCenter Server Virtual Appliance (VCVA), NSX Manager, and NSX controllers.

The compute cluster in Rack 2 contains the production virtual machines. In this guide, the compute cluster includes VMs deployed to different virtual networks to represent web servers, application servers, and database servers.

The edge cluster in Rack 3 contains the ESG for connectivity to the network core or WAN. The edge cluster also contains the distributed logical router (DLR) for routing NSX traffic between networks.

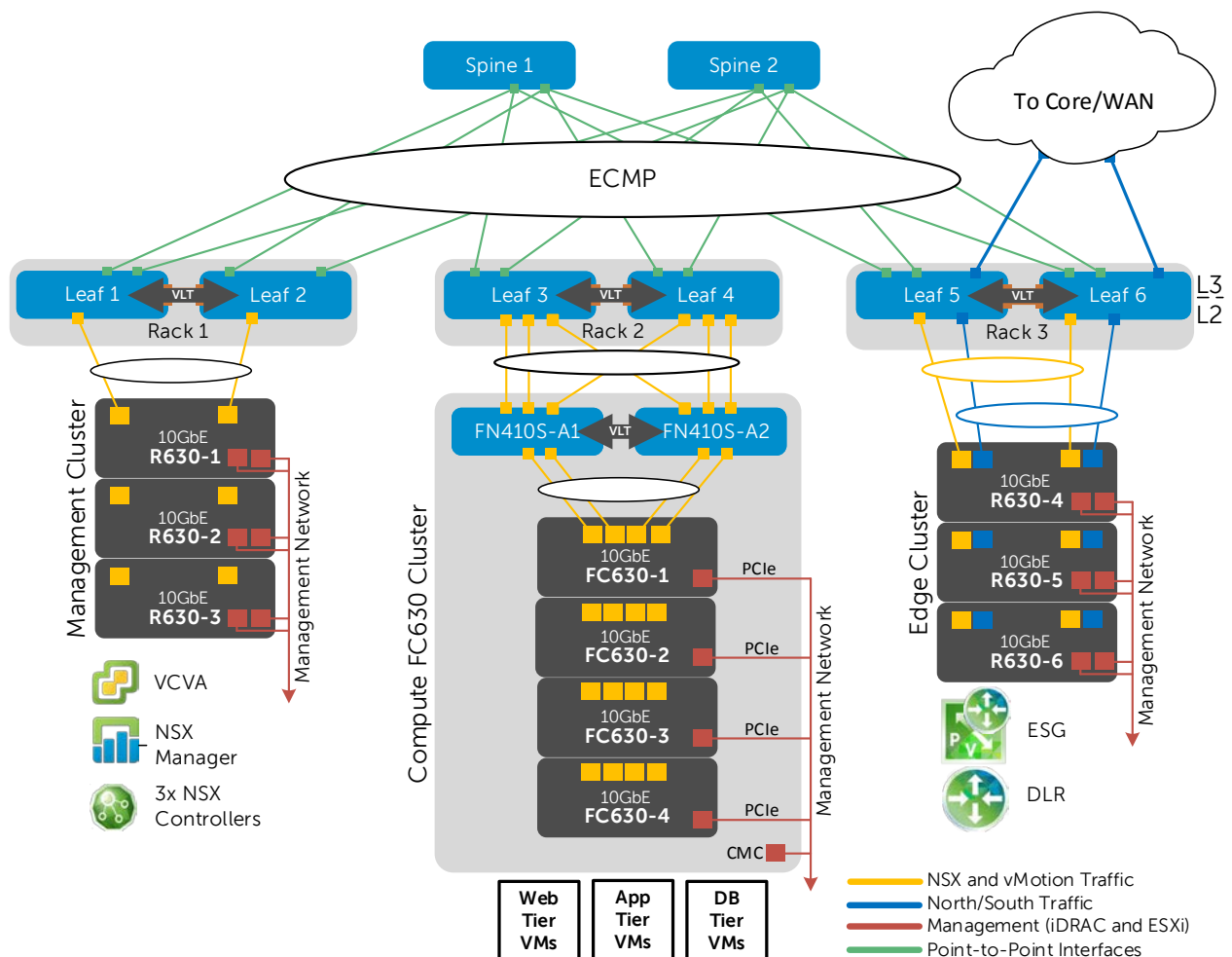


Figure 16 Combined physical and virtual networks

### 3.2.4 iSCSI SAN

Key vSphere features, such as vSphere High Availability and Distributed Resource Scheduler, require ESXi host clusters to be configured with shared storage. Shared storage can be provided by a variety of methods, including an iSCSI SAN, Fibre Channel SAN or VMware Virtual Storage Area Network (VSAN). The deployment example in this guide uses an iSCSI SAN in the Rack 2 Compute FC630 Cluster, as shown in Figure 17.

The iSCSI SAN is a dedicated storage network. Each FC630 server has a dual port QLogic 57810 Converged Network Adapter (CNA) connected to a pair of S4048-ON switches which are in turn connected to an SC4020 storage array. There are two S4048-ON iSCSI switches and one SC4020 storage array in each rack that contains a compute cluster (Rack 2 in this deployment).

**Note:** The S4048-ON switches used for iSCSI traffic are in addition to those used as leaf switches.

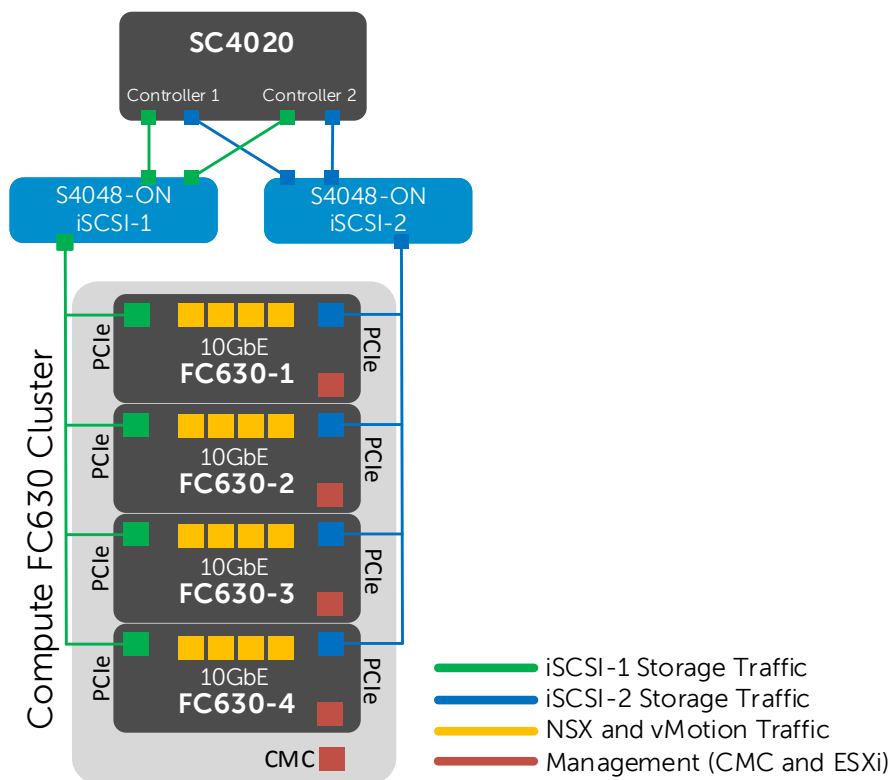


Figure 17 Compute cluster iSCSI SAN in Rack 2

**Note:** iSCSI SAN configuration instructions for the Rack 2 Compute FC630 Cluster are provided in Section 10 of this document. The iSCSI instructions can be extended to the Rack 1 Management and Rack 3 Edge clusters as well. Optionally, for VSAN configuration instructions, see [Dell EMC NSX Reference Architecture - FC430 Compute Nodes with VSAN Storage](#) or [Dell EMC NSX Reference Architecture - R730xd Compute Nodes with VSAN Storage](#) available on Dell TechCenter.

### 3.3 Management network

This guide uses a single management traffic network that is isolated from the production network. An S3048-ON switch installed in each rack provides connectivity to the management network.

Each R630 server has a 1GbE network adapter installed for ESXi host management and an iDRAC for out-of-band (OOB) server management. Each FX2s chassis has four 1GbE add-in PCIe network adapters (each connected internally to an FC630 server) for ESXi host management and a CMC for OOB management.

Each SC4020 storage array has two management ports, one connected to each controller.

These devices, in addition to the S3048-ON, S4048-ON and Z9100-ON switch management ports, are all connected to the management network as shown in Figure 18.

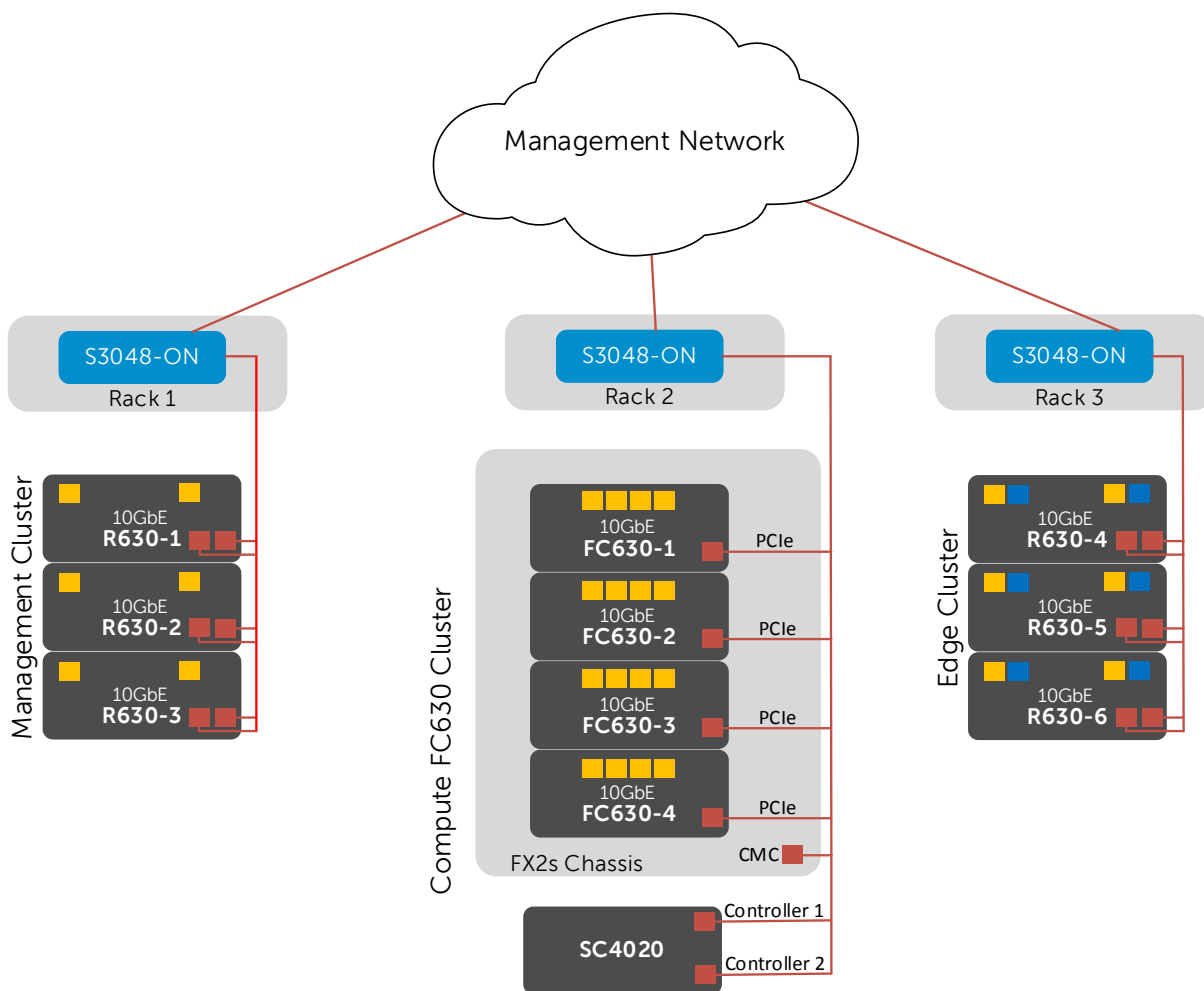


Figure 18 Physical layout of iDRAC, CMC, ESXi and Controller management interfaces

## 4 Network connections

This section details the physical network connections in each cluster.

### 4.1 Production network connections

#### 4.1.1 Management cluster – data center network

Figure 19 shows three PowerEdge R630 servers in the Rack 1 Management cluster connected to two S4048-ON switches (Leaf 1 and Leaf 2) via QLogic 57810 SFP+ dual port Network Daughter Cards (NDCs). The leaf switches are VLT peers and one NDC port from each server connects to each leaf.

**Note:** Optionally, QLogic 57840 SFP+ quad-port NDCs may be used in the management cluster R630 servers. Only two NDC ports are used in management cluster servers in this guide.

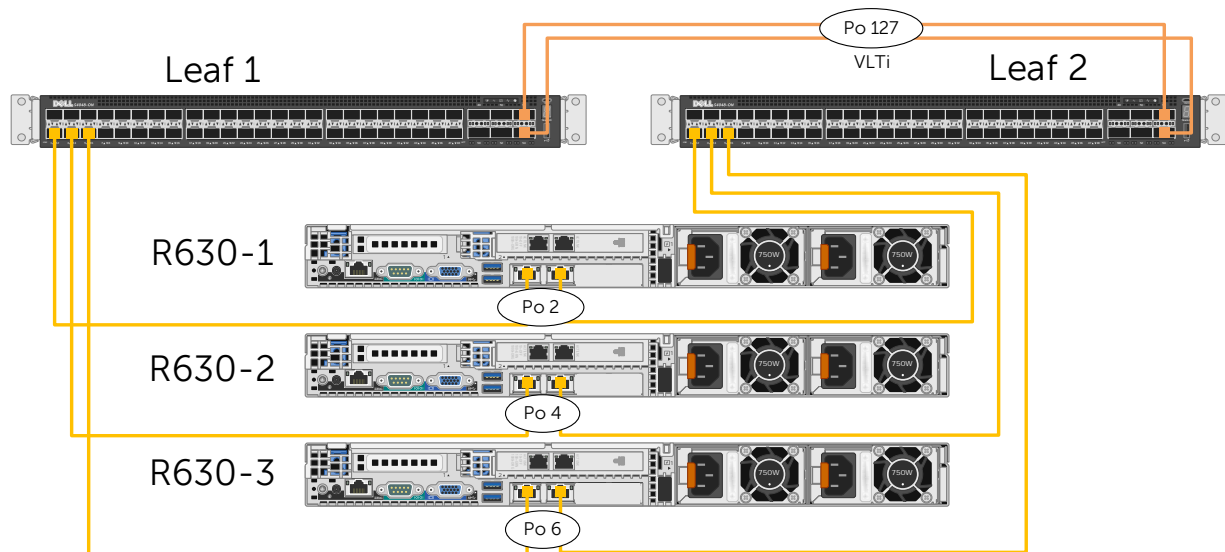


Figure 19 Production network connections for the management cluster

### 4.1.2 Compute cluster – data center network

Figure 20 shows the Rack 2 Compute cluster data center network connections from the FN410S switches in the FX2s chassis to Leaf 3 and Leaf 4. The leaf switches are VLT peers and three FN410S ports connect to each leaf switch. The FN410S switches are also VLT peers. The fourth FN410S port functions as the VLTi (VLT interconnect) between the switches.

Inside the FX2s chassis (not shown), four PowerEdge FC630 servers connect via QLogic 57840 quad-port network adapters to FN410S-A1 and A2. For each server, two links connect internally to FN410S-A1 and two connect to FN410S-A2.

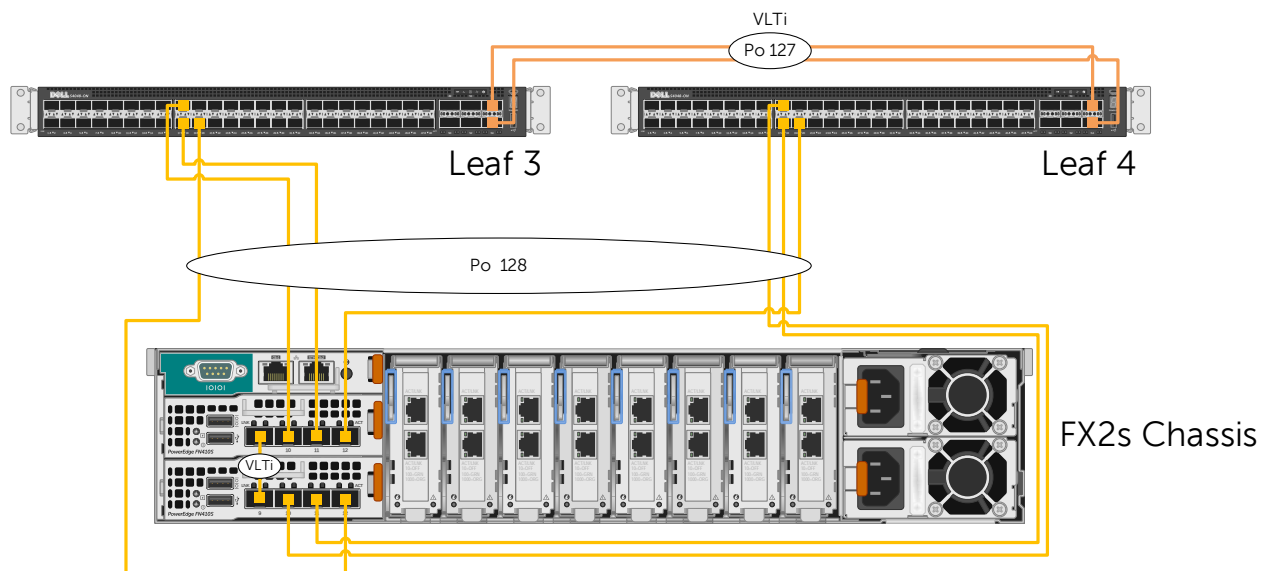


Figure 20 Production network connections for the compute cluster

### 4.1.3 Compute cluster – iSCSI SAN

Each FC630 server has one dual port QLogic 57810 SFP+ CNA installed in the back of the FX2s chassis for iSCSI traffic. FX2s chassis PCIe slots 1, 3, 5, & 7 are used. One port is connected to the S4048-ON switch designated iSCSI-1 and the other connected to iSCSI-2.

The SC4020 storage array has two controllers. The connections to each controller are split between iSCSI-1 and iSCSI-2 for fault tolerance.

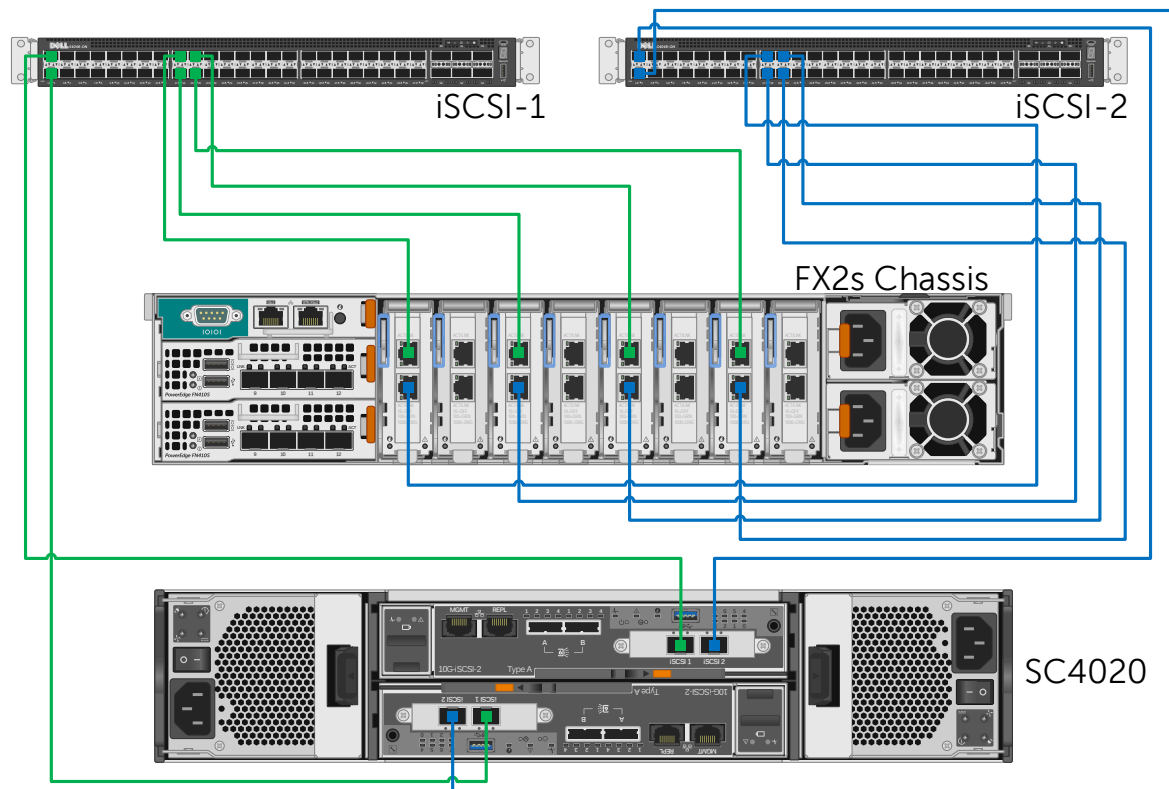


Figure 21 Compute cluster – iSCSI SAN connections



#### 4.1.4 Edge cluster – data center network

In the Rack 3 Edge cluster, three PowerEdge R630 servers connect to S4048-ON switches, Leaf 5 and Leaf 6, via QLogic 57840 quad-port NDCs. The yellow connections are used for East-West connections to Racks 1 and 2, and the blue connections are used for North-South connections to the network core or WAN.

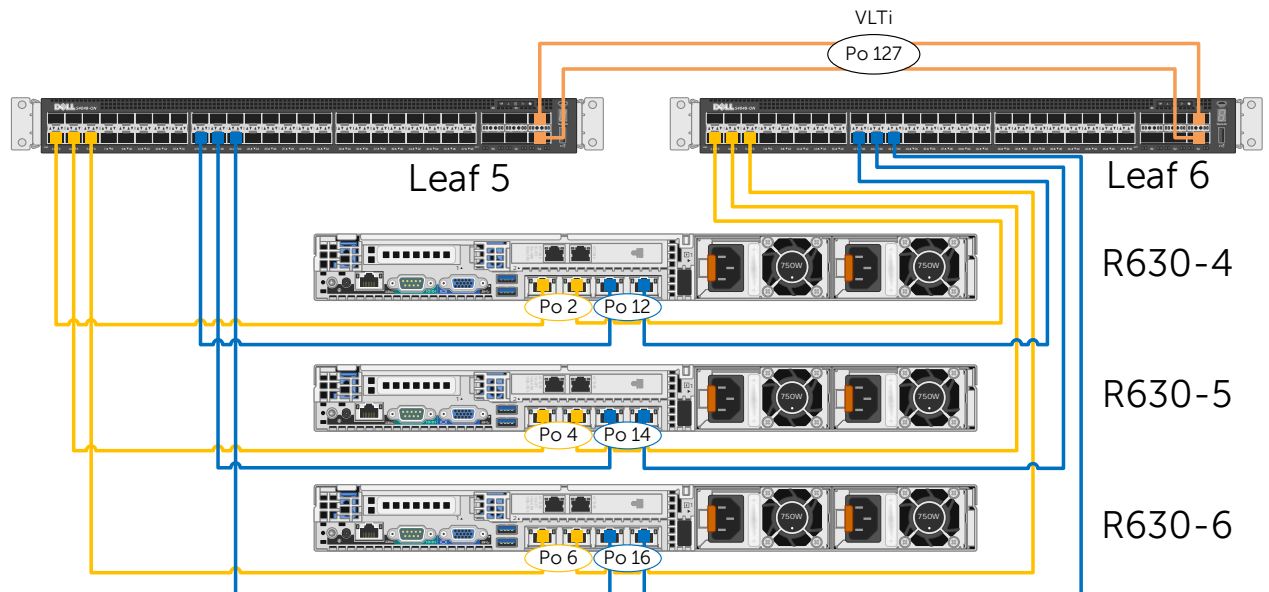


Figure 22 Production network connections for the edge cluster

## 4.2 Management network connections

These connections are used for non-production, management traffic.

### 4.2.1 Management and edge clusters

In the Rack 1 Management cluster, servers R630-1 through R630-3 are connected to an S3048-ON switch via add-in Intel I350-T dual port PCIe adapters. The R630 server iDRACs are connected to the same switch as shown in Figure 23. The Rack 3 Edge cluster is identical and uses servers R630-4 through R630-6.

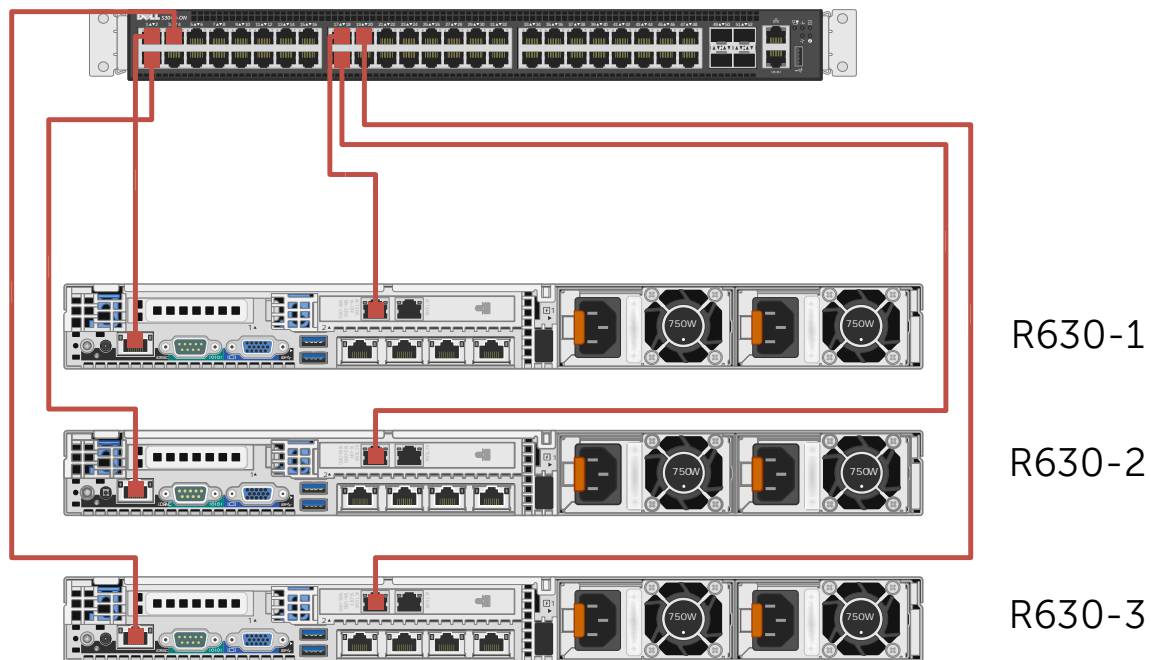


Figure 23 Management cluster – management network connections (edge cluster is identical)

## 4.2.2 Compute cluster

For management traffic in the Rack 2 Compute Cluster, each FC630 server has one Intel I350-T add-in adapter installed in the back of the FX2s chassis in PCIe slots 2, 4, 6, & 8. These connections, along with the CMC, are connected to an S3048-ON switch.

The management port from each SC4020 storage controller is also connected the S3048-ON.

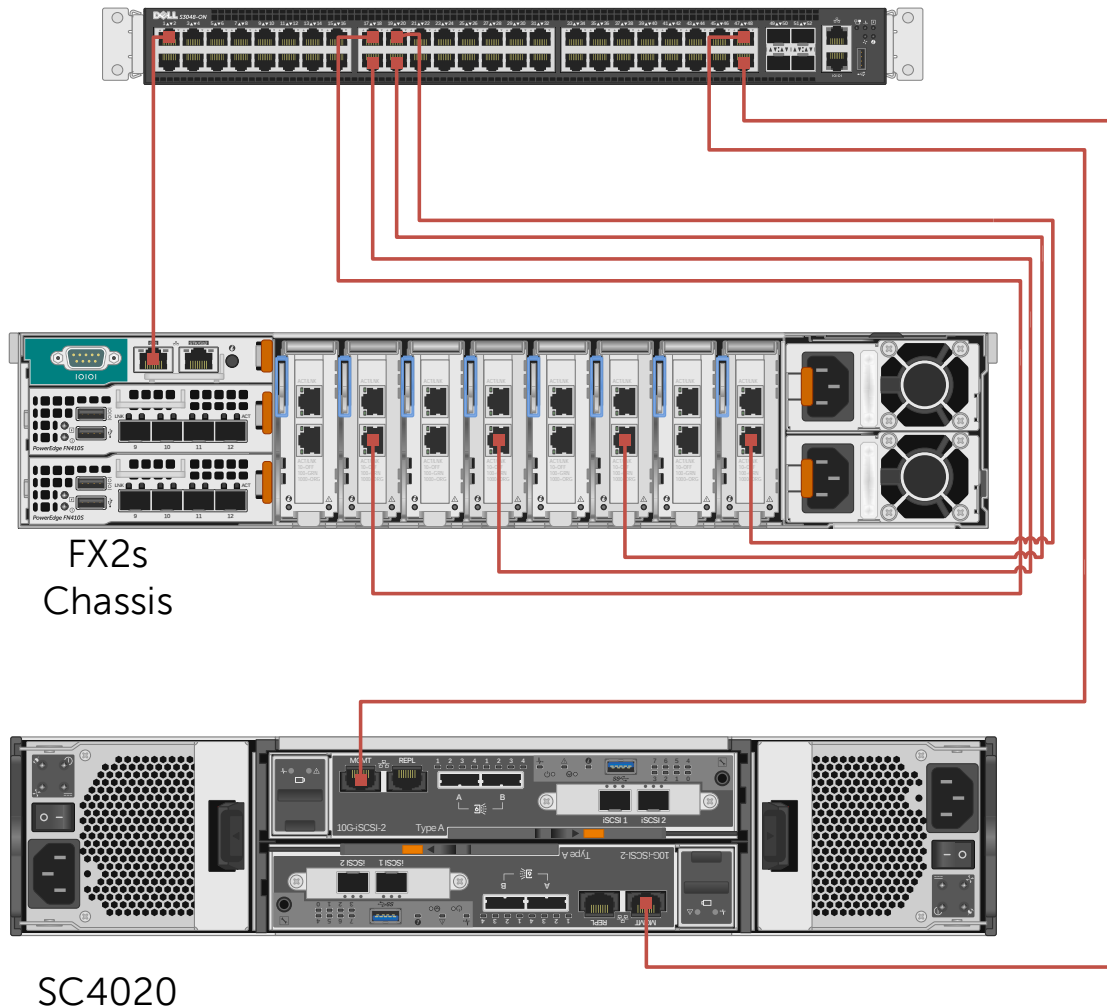


Figure 24 Compute cluster – management network connections

## 5 Leaf-spine topology

In a leaf-spine architecture, a series of access layer (top-of-rack) switches form the leaf switches. These switches are fully meshed to a series of spine switches. Each leaf connects to each spine, but the spines do not connect to one another. The total number of connections is equal to the number of leaf switches multiplied by the number of spine switches.

The mesh ensures that leaf switches are no more than one hop away from one another, minimizing latency and the likelihood of bottlenecks between leaf switches. Given any single-link failure scenario, all leaf switches retain connectivity to one another through the remaining links.

The connections between spine switches and leaf switches can be layer 2 or layer 3. The deployment scenario in this guide uses layer 3 connections. This limits layer 2 broadcast domains, resulting in improved network stability and scalability.

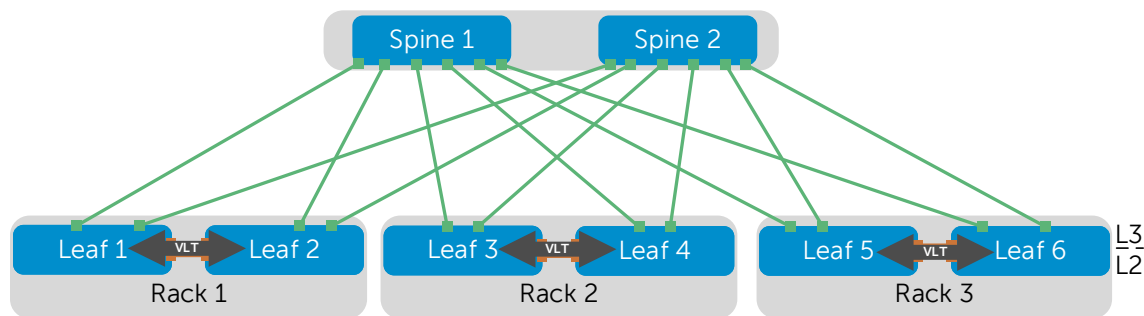


Figure 25 Leaf-spine topology example

Figure 25 shows a high-level diagram of the leaf-spine topology used in this guide with Z9100-ON switches as spines and S4048-ON switches as leaf switches.

The Z9100-ON supports a maximum number of 32 leaf switches. The example in this document uses six leaf switches in three racks. Two leaf switches are used in each rack for redundancy. The first rack contains the management cluster, the second rack contains the compute cluster and the edge cluster is in the third rack.

As administrators add racks to the data center, two leaf switches are added to each new rack. As bandwidth requirements increase, spine switches are added as needed. Scaling guidance is covered in Section 14.

### 5.1 Routing protocol selection

Choose from any of the following three routing protocols when designing a leaf-spine network:

- Border gateway protocol (External or Internal BGP)
- Open Shortest Path First (OSPF)
- Intermediate System to Intermediate System (IS-IS).

BGP was selected for this guide for scalability. BGP can be configured as *External* BGP (EBGP) to route between autonomous systems or *Internal* BGP (IBGP) to route within a single autonomous system.

EBGP excels at prefix filtering, traffic engineering and traffic tagging. This allows BGP to match on any attribute or prefix and prune prefixes between switches. Unlike EBGP, IBGP requires BGP add-path to support ECMP. To handle peering, IBGP requires route reflectors to mitigate the protocol's full-mesh requirement.

For scalability and the reasons described above, an EBGP deployment is used in this guide.

## 5.2 BGP ASN configuration

BGP has a reserved, private, 2-byte Autonomous System Number (ASN) range from 64,512 to 65,535. For this EBGP configuration, each switch is assigned a separate ASN. Figure 26 below shows the ASN assignments used in this guide.

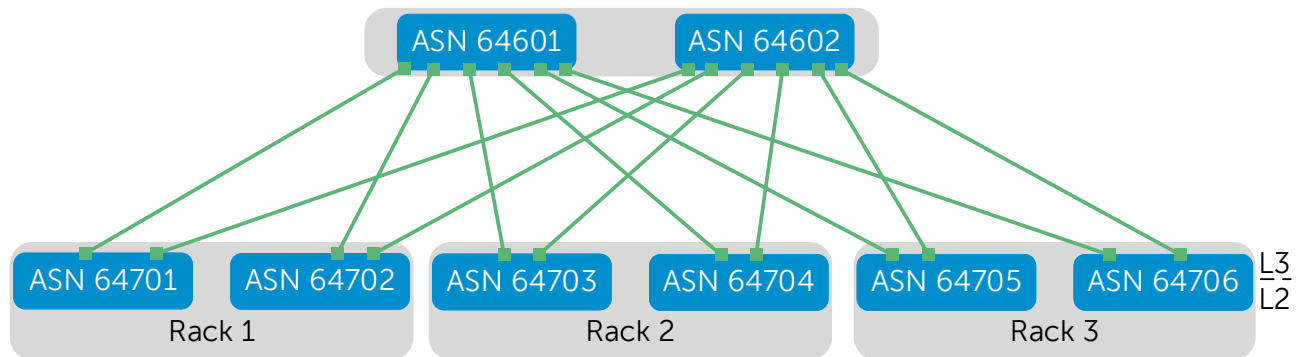


Figure 26 BGP ASN assignments

## 5.3 BGP fast fall-over

BGP tracks IP reachability to the peer remote address and the peer local address. Whenever either address becomes unreachable (for example, no active route exists in the routing table for the peer IPv4 destination/local address), BGP brings down the session with the peer. This feature is called fast fall-over. Dell EMC recommends enabling fast fall-over for EBGP settings.

## 5.4 IP Address Management

Proper IP address management is critical before deploying a leaf-spine topology. This section covers the IP addressing used on the physical network in this guide.

### 5.4.1 Loopback addresses

Figure 27 shows the loopback addresses used as router IDs. All loopback addresses are part of the 10.0.0.0/8 address space with each switch using a 32-bit mask.

In this scheme, the third octet represents the layer, 1 for spine and 2 for leaf. The fourth octet is the counter for the appropriate layer. For example, 10.0.1.1/32 is the first spine switch in the topology while 10.0.2.4/32 is the fourth leaf switch.

This address scheme helps with establishing BGP neighbor adjacencies as well as troubleshooting connectivity.

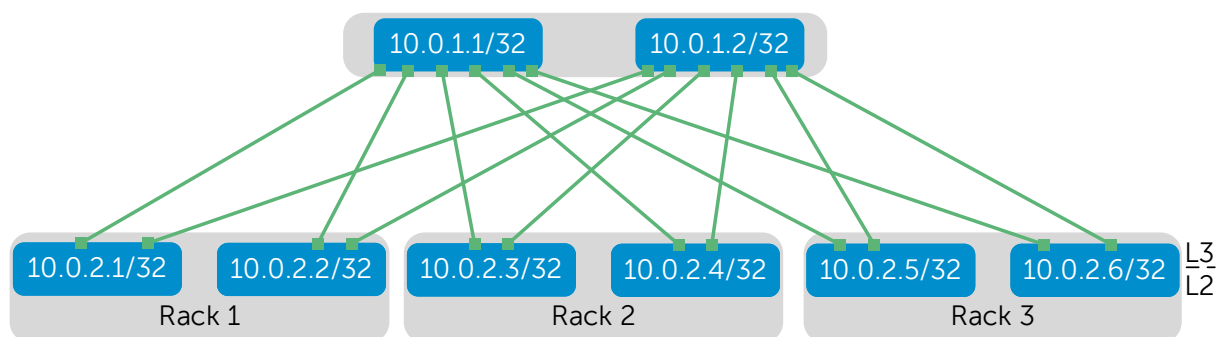


Figure 27 Loopback addressing

## 5.4.2 Point-to-point addresses

Table 1 below lists layer 3 connection details for each leaf and spine switch. The IP scheme below can be easily extended to account for additional leaf and spine switches.

All addresses come from the same base IP prefix, 192.168.0.0/16 with the 3<sup>rd</sup> octet representing the spine number. For instance 192.168.1.0/31 is a two host subnet that ties to Spine 1 while 192.168.2.0/31 ties to Spine 2.

Table 1 Interface and IP configuration

Source switch	Rack	Source interface	Source IP	Network	Destination switch	Destination interface	Destination IP	Label
Leaf 1	1	fo1/49	.1	192.168.1.0/31	Spine 1	fo1/1/1	.0	A
Leaf 1	1	fo1/50	.1	192.168.2.0/31	Spine 2	fo1/1/1	.0	B
Leaf 2	1	fo1/49	.3	192.168.1.2/31	Spine 1	fo1/2/1	.2	C
Leaf 2	1	fo1/50	.3	192.168.2.2/31	Spine 2	fo1/2/1	.2	D
Leaf 3	2	fo1/49	.5	192.168.1.4/31	Spine 1	fo1/3/1	.4	E
Leaf 3	2	fo1/50	.5	192.168.2.4/31	Spine 2	fo1/3/1	.4	F
Leaf 4	2	fo1/49	.7	192.168.1.6/31	Spine 1	fo1/4/1	.6	G
Leaf 4	2	fo1/50	.7	192.168.2.6/31	Spine 2	fo1/4/1	.6	H
Leaf 5	3	fo1/49	.9	192.168.1.8/31	Spine 1	fo1/5/1	.8	I
Leaf 5	3	fo1/50	.9	192.168.2.8/31	Spine 2	fo1/5/1	.8	J
Leaf 6	3	fo1/49	.11	192.168.1.10/31	Spine 1	fo1/6/1	.10	K
Leaf 6	3	fo1/50	.11	192.168.2.10/31	Spine 2	fo1/6/1	.10	L

Figure 28 shows the links from Table 1:

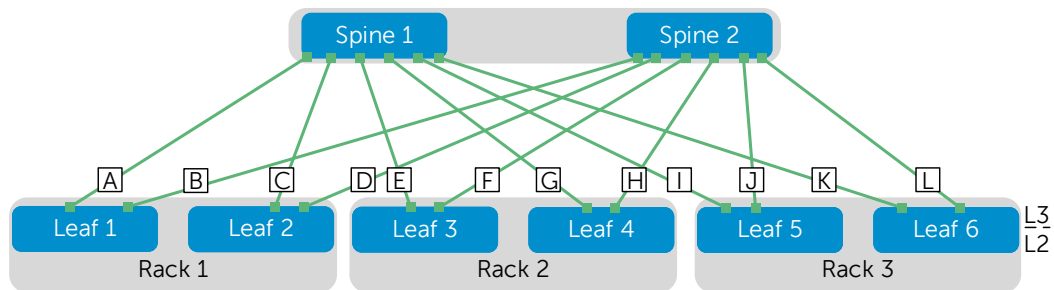


Figure 28 Point-to-point IP addressing

**Note:** The example point-to-point addresses use a 31-bit mask to save address space. This is optional, and covered in [RFC 3021](#). Below is an example when setting an IP address with a 31-bit mask on a Dell S4048-ON. The warning message can be safely ignored on point-to-point interfaces.

```
Leaf-1(conf-if-fo-1/49)#ip address 192.168.1.1/31
% Warning: Use /31 mask on non point-to-point interface cautiously.
```



### 5.4.3 VLANs and IP addressing

Table 2 outlines the VLAN IDs, network and gateway addresses used on the data center network. The "x" in each network address is replaced by the rack number to create a different network for each rack. The gateway address is the Virtual Router Redundancy Protocol (VRRP) group address, described in the next section. The VLANs and networks are advertised through the BGP instance at the same cost.

Table 2 VLAN and network examples

VLAN ID	Network	Gateway	Used For
22	10.22.x.0/24	10.22.x.254	vMotion
55	10.55.x.0/24	10.55.x.254	NSX

## 5.5 VRRP

VRRP is designed to eliminate a single point of failure in a routed network. VRRP is used to create a virtual router which is an abstraction of the two physical leaf switches. The virtual router is assigned an IP address that is used as the gateway address by the compute nodes. In the event that one of the leaf switches fails, the remaining leaf acts as the gateway until the failed unit recovers.

As illustrated in Figure 29, Node 1 is participating in VLAN 55 in Rack 2. The node has an IP address of 10.55.2.1. The node's gateway address is set to 10.55.2.254. This is the Virtual IP (VIP) provided by the VRRP instance running between leaf switches 3 and 4.

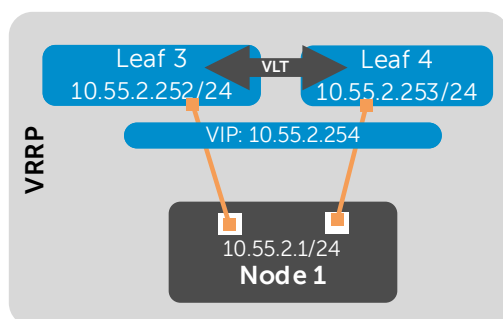


Figure 29 VRRP configuration example – VLAN 55 in Rack 2

A VRRP instance is created for each VLAN in each pair of leaf switches at the top of each rack.

Table 3 shows the VRRP IP addressing scheme for NSX VLAN 55 as an example. The numbering scheme is also used for the vMotion VLAN (VLAN 22), with the 2<sup>nd</sup> octet in the IP addresses replaced with the VLAN number.

Table 3 VRRP interface configuration for VLAN 55 – Racks 1-3

Rack ID	VLAN	First Leaf VLAN IP	Second Leaf VLAN IP	Virtual IP
Rack 1	55	10.55.1.252/24	10.55.1.253/24	10.55.1.254
Rack 2	55	10.55.2.252/24	10.55.2.253/24	10.55.2.254
Rack 3	55	10.55.3.252/24	10.55.3.253/24	10.55.3.254

## 5.6 ECMP

ECMP is the core protocol enabling the deployment of a layer 3 leaf-spine topology. ECMP gives each leaf and spine switch the ability to load balance flows across a set of equal next-hops. For example, when using two spine switches, each leaf has a connection to each spine. For every flow egressing a leaf switch, there exists two equal next-hops: one to each spine.

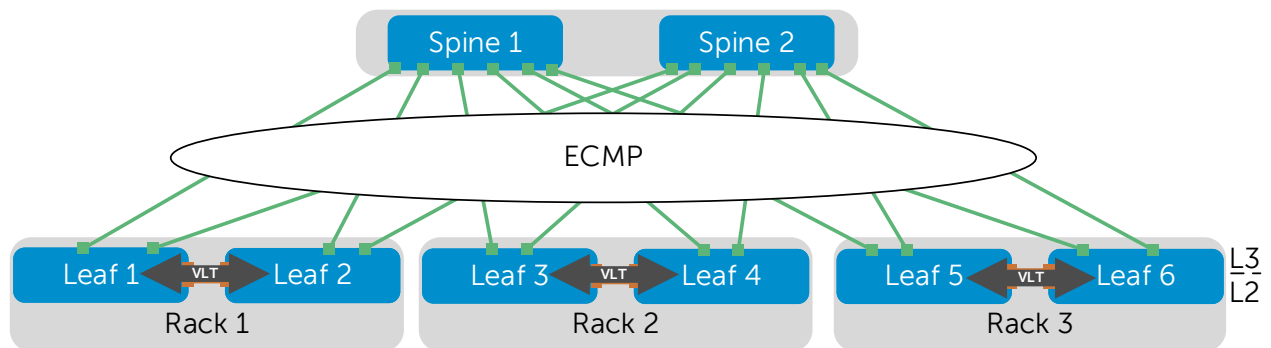


Figure 30 ECMP

## 5.7 VLT

A pair of leaf switches at the top of each rack provides redundancy. These switches' configurations include the Dell Networking Virtual Link Trunking (VLT) feature.

VLT reduces the role of spanning tree protocols (STPs) by allowing link aggregation group (LAG) terminations on two separate switches and supporting a loop-free topology. VLT provides Layer 2 multipathing and load-balances traffic where alternative paths exist. Virtual Link Trunking offers the following additional benefits:

- Allows a single device to use a LAG across two upstream devices
- Eliminates STP-blocked ports
- Uses all available uplink bandwidth
- Provides fast convergence if either the link or a device fails
- Provides link-level resiliency
- Assures high availability

## 5.8 Uplink Failure Detection

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

Dell EMC recommends enabling Uplink Failure Detection (UFD), which detects the loss of upstream connectivity. An uplink-state group is configured on each leaf switch, which creates an association between the spine uplinks and the downlink interfaces. An uplink-state group is also configured on each FN410S.

In the event of an uplink failure, UFD automatically shuts down the corresponding downstream interfaces. This propagates down to the hosts attached to the leaf or FN410S switch. The host then uses its remaining Link Aggregation Control Protocol (LACP) port member to continue sending traffic across the leaf-spine network.

## 6 Configure physical switches

This section contains switch configuration details with explanations for one switch in each major role on the production network. This chapter details the following switches:

- FN410S-A1
- S4048-ON: Leaf 1
- S4048-ON: Leaf 5 with edge configuration
- Z9100-ON: Spine 1
- S4048-ON: iSCSI 1

The remaining switches use configurations very similar to one of the five configurations above, with the applicable switches specified in each section. Complete configuration files for all switches used on the production network in this guide are provided as attachments.

**Notes: MTU** - The MTU is set to 9216 bytes on all switch interfaces in this guide. On the data center network, VXLAN protocol requirements require setting the MTU to at least 1600 bytes on all switches that will handle NSX traffic. On switches used in iSCSI SANs, Dell EMC recommends setting the MTU to 9216 for best performance. **Port Channel Numbering** – LACP port channel numbers may be any number in the range 1-128.

### 6.1 Factory default settings

The configuration commands in the sections below assume switches are at their factory default settings. All switches in this guide can be reset to factory defaults as follows:

```
switch#restore factory-defaults stack-unit unit# clear-all  
Proceed with factory settings? Confirm [yes/no]:yes
```

Factory settings are restored and the switch reloads. After reload, enter **A** at the [A/C/L/S] prompt as shown below to exit Bare Metal Provisioning mode.

```
This device is in Bare Metal Provisioning (BMP) mode.  
To continue with the standard manual interactive mode, it is necessary to  
abort BMP.
```

```
Press A to abort BMP now.  
Press C to continue with BMP.  
Press L to toggle BMP syslog and console messages.  
Press S to display the BMP status.  
[A/C/L/S]:A
```

```
% Warning: The bmp process will stop ...
```

```
Dell>
```

The switch is now ready for configuration.

## 6.2 FN410S switch configuration

The compute cluster includes a PowerEdge FX2s chassis with four FC630 servers and two FN410S switches.

Each FC630 server has an LACP-enabled port channel connected to internal interfaces of each FN410S. For clarity, only port channel 1 (for server FC630-1) is shown in Figure 31. The remaining port channels are numbered 2-4.

The two FN410S switches are configured as VLT peers. Three of the four FN410S external interfaces, tengigabitethernet 0/10–12, are configured in port channel 128 which is connected to leaf switches 3 and 4. The 4<sup>th</sup> external interface, tengigabitethernet 0/9, is used as the VLT interconnect between FN410S-A1 and FN410S-A2.

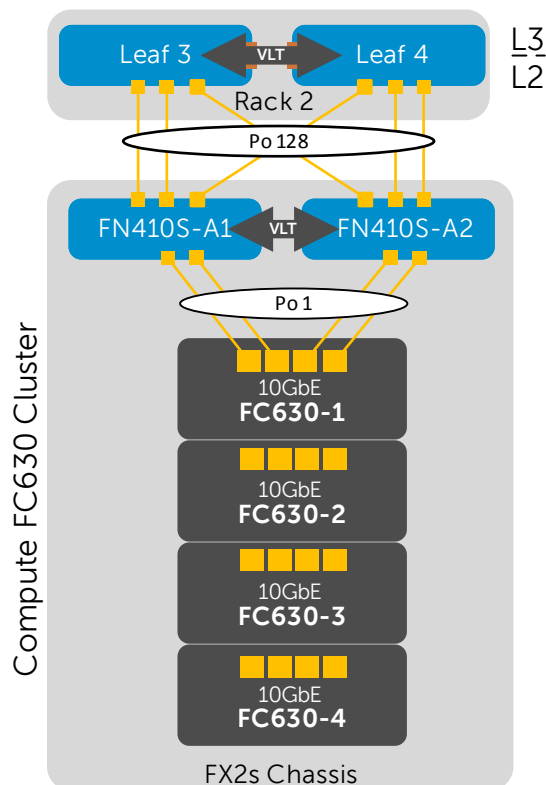


Figure 31 FN410S network connections (internal port channels to FC630-2 through 4 not shown)

The following section outlines the configuration commands issued to the FN410S switches. The switches start at their factory default settings per Section 6.1.

After FN410S switches boot to their default settings, place them in full-switch mode as follows:

```
Dell>enable
Dell#configure
Dell(conf)#stack-unit 0 iom-mode full-switch

% You are about to configure the Full Switch Mode.
```

Please reload to effect the changes

```
Dell(conf)#do reload
```

```
System configuration has been modified. Save? [yes/no]: yes
```

```
Proceed with reload [confirm yes/no]: yes
```

After FN410S switches boot to full-switch mode, enter the following commands to configure the FN410S-A1.

**Note:** Ensure FN410S switches have been placed in full-switch mode before proceeding. The following configuration details are specific to switch FN410S-A1. The configuration for FN410S-A2 is similar. See the FN410S-A1.txt and FN410S-A2.txt attachments.

Initial configuration involves setting the hostname, enabling Link Layer Discovery Protocol (LLDP) and disabling Data Center Bridging (DCB). LLDP is useful for troubleshooting (see Section 9.8). DCB is enabled by default on FN410S but is not used in this environment.

Finally, configure the management interface and default gateway.

```
enable
configure
```

```
hostname FN410S-A1
protocol lldp
advertise management-tlv management-address system-description system-name
advertise interface-port-desc
no dcb enable
```

```
interface ManagementEthernet 0/0
ip address 100.67.187.151/24
no shutdown
```

```
management route 0.0.0.0/0 100.67.187.254
```

Next, the VLT interface between the two switches is configured. In this configuration, interface tengigabitethernet 0/9 is used for the VLTi interface. It is added to static port-channel 127. The backup destination is the management IP address of the VLT peer switch, FN410S-A2. The VLT unit-id is set to 0 (and is set to 1 on FN410S-A2).

```
interface port-channel 127
description VLTi
mtu 9216
channel-member tengigabitethernet 0/9
no shutdown
```

```

interface tengigabitethernet 0/9
description VLTi
no shutdown

vlt domain 127
peer-link port-channel 127
back-up destination 100.67.187.162
unit-id 0

```

The upstream interfaces to the two leaf switches are configured in this section. External interfaces tengigabitethernet 0/10-12 are used and placed in LACP-enabled port channel 128. The port channel is configured for VLT and jumbo frames are enabled for VXLAN traffic.

```

interface range tengigabitethernet 0/10-12
description To Leaf switches 3 and 4 te 1/45-47
mtu 9216
port-channel-protocol LACP
port-channel 128 mode active
no shutdown

interface port-channel 128
description To Leaf switches 3 and 4 te 1/45-47
mtu 9216
portmode hybrid
switchport
vlt-peer-lag port-channel 128
no shutdown

```

The downstream interfaces are configured in the next set of commands. Internal interfaces are added to a port channel to each FC630. The port channels are configured for VLT and jumbo frames are enabled on all interfaces for VXLAN traffic.

```

interface tengigabitethernet 0/1
description To FC630-1
mtu 9216
port-channel-protocol LACP
port-channel 1 mode active
no shutdown

interface tengigabitethernet 0/2
description To FC630-1
mtu 9216
port-channel-protocol LACP
port-channel 1 mode active
no shutdown

interface port-channel 1
description To FC630-1

```

```

mtu 9216
portmode hybrid
switchport
vlt-peer-lag port-channel 1
no shutdown

interface tengigabitethernet 0/3
description To FC630-2
mtu 9216
port-channel-protocol LACP
port-channel 2 mode active
no shutdown

interface tengigabitethernet 0/4
description To FC630-2
mtu 9216
port-channel-protocol LACP
port-channel 2 mode active
no shutdown

interface port-channel 2
description To FC630-2
mtu 9216
portmode hybrid
switchport
vlt-peer-lag port-channel 2
no shutdown

interface tengigabitethernet 0/5
description To FC630-3
mtu 9216
port-channel-protocol LACP
port-channel 3 mode active
no shutdown

interface tengigabitethernet 0/6
description To FC630-3
mtu 9216
port-channel-protocol LACP
port-channel 3 mode active
no shutdown

interface port-channel 3
description To FC630-3
mtu 9216
portmode hybrid
switchport

```



```

vlt-peer-lag port-channel 3
no shutdown

interface tengigabitethernet 0/7
description To FC630-4
mtu 9216
port-channel-protocol LACP
port-channel 4 mode active
no shutdown

interface tengigabitethernet 0/8
description To FC630-4
mtu 9216
port-channel-protocol LACP
port-channel 4 mode active
no shutdown

interface port-channel 4
description To FC630-4
mtu 9216
portmode hybrid
switchport
vlt-peer-lag port-channel 4

```

Finally, the two required VLAN interfaces are created. All downstream and upstream port channels are tagged in each VLAN.

```

interface Vlan 22
description vMotion
mtu 9216
tagged Port-channel 1-4,128
no shutdown

interface Vlan 55
description NSX
mtu 9216
tagged Port-channel 1-4,128
no shutdown

```

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

```

uplink-state-group 1
description Disable downstream ports in event all uplinks fail
downstream Port-channel 1-4
upstream Port-channel 128

```

Save the configuration.

```
end
write
```

## 6.3 S4048-ON leaf switch configuration

Each S4048-ON leaf switch has an LACP-enabled port channel connected to each of the downstream R630 servers, or in the case of the compute cluster, the downstream FN410S switches.

There are a total of six leaf switches in this guide, with two in each rack configured as VLT peers.

The following section outlines the configuration commands issued to S4048-ON leaf switches. The switches start at their factory default settings per Section 6.1.

**Note:** The following configuration details are specific to Leaf 1. The remaining leaf switches, 2-6, are similar. Leaf switches 3-4 have a different downstream port channel configuration. Leaf switches 5-6 have additional edge configuration steps that are covered in the next section. Complete configuration details for all six leaf switches are provided in the attachments named leaf1.txt through leaf6.txt.

Initial configuration involves setting the hostname, and enabling LLDP. LLDP is useful for troubleshooting (see Section 9.8). Finally, the management interface and default gateway are configured.

```
enable
configure

hostname Leaf-1
protocol lldp
advertise management-tlv management-address system-description system-name
advertise interface-port-desc

interface ManagementEthernet 1/1
ip address 100.67.187.35/24
no shutdown

management route 0.0.0.0/0 100.67.187.254
```

Next, the VLT interfaces between Leaf-1 and Leaf-2 are configured. In this configuration, interfaces fortyGigE 1/53-54 are used for the VLT interconnect. They are added to static port-channel 127. The backup destination is the management IP address of the VLT peer switch, Leaf-2.

```
interface port-channel 127
description VLTi
mtu 9216
channel-member fortyGigE 1/53 - 1/54
no shutdown
```

```

interface range fortyGigE 1/53 - 1/54
description VLTi
no shutdown

vlt domain 127
peer-link port-channel 127
back-up destination 100.67.187.34
unit-id 0
exit

```

The downstream interfaces, to the R630 servers in this case, are configured in the next set of commands. Each interface is added to a numerically corresponding port channel. The port channels are configured for VLT and jumbo frames are enabled on all interfaces for VXLAN traffic.

```

interface tengigabitethernet 1/2
description To R630-1 100.67.187.19
mtu 9216
port-channel-protocol LACP
port-channel 2 mode active
no shutdown

```

```

interface port-channel 2
description To R630-1 100.67.187.19
mtu 9216
portmode hybrid
switchport
vlt-peer-lag port-channel 2
no shutdown

```

```

interface tengigabitethernet 1/4
description To R630-2 100.67.187.18
mtu 9216
port-channel-protocol LACP
port-channel 4 mode active
no shutdown

```

```

interface port-channel 4
description To R630-2 100.67.187.18
mtu 9216
portmode hybrid
switchport
vlt-peer-lag port-channel 4
no shutdown

```

```

interface tengigabitethernet 1/6
description To R630-3 100.67.187.17
mtu 9216

```

```

port-channel-protocol LACP
port-channel 6 mode active
no shutdown

interface port-channel 6
description To R630-3 100.67.187.17
mtu 9216
portmode hybrid
switchport
vlt-peer-lag port-channel 6
no shutdown

```

The two required VLAN interfaces are created. All downstream port channels are tagged in each VLAN. Each interface is assigned to a VRRP group and a VRRP address is assigned. VRRP priority is set to 254 to make this switch the master. (On the VRRP peer switch, priority is set to 1).

```

interface Vlan 22
description vMotion
ip address 10.22.1.252/24
mtu 9216
tagged Port-channel 2,4,6
vrrp-group 22
description vMotion
priority 254
virtual-address 10.22.1.254
no shutdown

interface Vlan 55
description NSX
ip address 10.55.1.252/24
mtu 9216
tagged Port-channel 2,4,6
vrrp-group 55
description NSX
priority 254
virtual-address 10.55.1.254
no shutdown

```

The upstream layer 3 interfaces connected to the spines are configured. A loopback interface is configured as the router ID for BGP.

```

interface fortyGigE 1/49
description To Spine-1
ip address 192.168.1.1/31
mtu 9216
no shutdown

```

```

interface fortyGigE 1/50
description To Spine-2
ip address 192.168.2.1/31
mtu 9216
no shutdown

```

```

interface loopback 0
description Router ID
ip address 10.0.2.1/32

```

BGP is configured to allow routing to the IP fabric. Additionally, an IP prefix and route map are created to automatically redistribute all leaf subnets and loopback addresses from the leaf and spine switches.

```

route-map spine-leaf permit 10
match ip address spine-leaf

ip prefix-list spine-leaf
description BGP redistribute loopback and leaf networks
seq 5 permit 10.0.0.0/23 ge 32
seq 10 permit 10.0.0.0/8 ge 24
router bgp 64701
bgp bestpath as-path multipath-relax
maximum-paths ebgp 64
redistribute connected route-map spine-leaf
bgp graceful-restart
neighbor spine-leaf peer-group
neighbor spine-leaf fall-over
neighbor spine-leaf advertisement-interval 1
neighbor spine-leaf no shutdown
neighbor 192.168.1.0 remote-as 64601
neighbor 192.168.1.0 peer-group spine-leaf
neighbor 192.168.1.0 no shutdown
neighbor 192.168.2.0 remote-as 64602
neighbor 192.168.2.0 peer-group spine-leaf
neighbor 192.168.2.0 no shutdown

```

An ECMP group is created that includes the point-to-point interfaces to the two spine switches.

```

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

```

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

```

uplink-state-group 1
description Disable downstream ports in event all uplinks fail

```

```

downstream TenGigabitEthernet 1/1-1/48
upstream fortyGigE 1/49,1/50

```

Save the configuration.

```

end
write

```

### 6.3.1 S4048-ON edge switch configuration

The following section contains additional configuration steps required on leaf switches 5 and 6 connected to the core/WAN shown in Figure 32 below.

**Note:** Only the north/south (blue) links to the core/WAN are configured in this section. The remaining links were configured in the previous section. The following configuration details are specific to Leaf 5. Leaf 6 is similar. Complete configuration details are provided in the attachments named leaf5.txt and leaf6.txt.

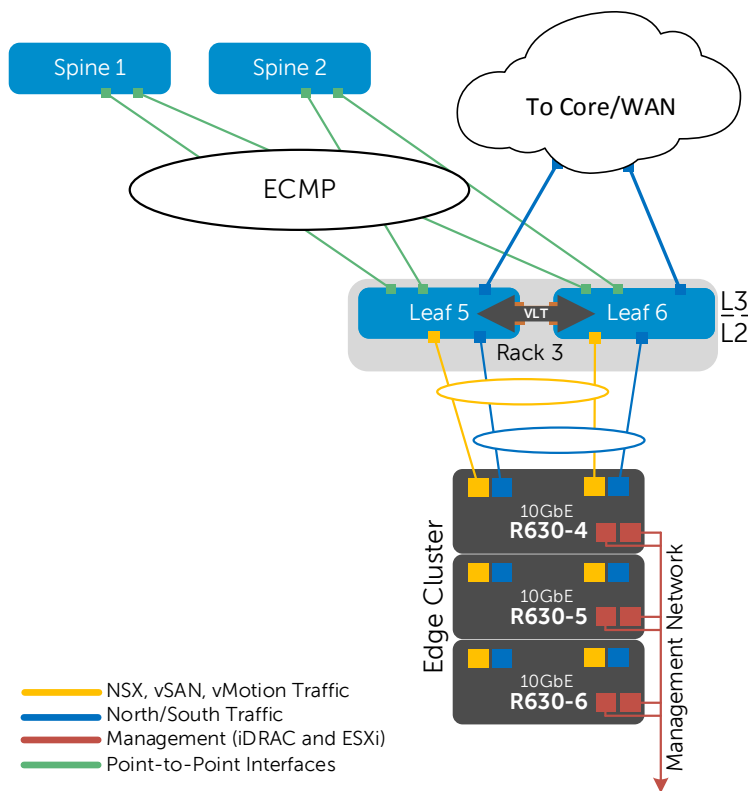


Figure 32 Edge cluster leaf switch configuration

Enable layer 3 VLT peer-routing. This will allow leaf 5 and leaf 6 to create an OSPF neighbor adjacency across the VLTi link.

```
vlt domain 127
peer-routing
```

Add the edge/wan/core links to the R630 edge servers.

```
interface tengigabitethernet 1/18
description Edge To R630-4 100.67.187.14
port-channel-protocol LACP
port-channel 12 mode active
no shutdown
```

```
interface port-channel 12
description Edge To R630-4 100.67.187.14
portmode hybrid
switchport
vlt port-channel 12
no shutdown
```

```
interface tengigabitethernet 1/20
description Edge To R630-5 100.67.187.15
port-channel-protocol LACP
port-channel 14 mode active
no shutdown
```

```
interface port-channel 14
description Edge To R630-5 100.67.187.15
portmode hybrid
switchport
vlt port-channel 14
no shutdown
```

```
interface tengigabitethernet 1/22
description Edge To R630-6 100.67.187.16
port-channel-protocol LACP
port-channel 16 mode active
no shutdown
```

```
interface port-channel 16
description Edge To R630-6 100.67.187.16
portmode hybrid
switchport
vlt port-channel 16
no shutdown
```

Add VLAN 66. This VLAN is dedicated to handling north/south traffic and does not use VRRP. The interface is used to create an Open Shortest Path First (OSPF) router adjacency and does not require a VRRP group address for forwarding.

```
interface vlan 66
description Edge
ip address 10.66.3.252/24
tagged Port-channel 12,14,16
no shutdown
```

Create an OSPF routing process to handle north to south traffic. A specific router ID is specified here to separate the neighbor relationship tables. This OSPF instance will create a neighbor relationship with Leaf 6 as well as the ESG, configured in Section 12.

```
router ospf 1
network 10.66.3.0/24 area 0
router-id 10.66.3.252
```

Save the configuration.

```
end
write
```

## 6.4 Z9100-ON spine switch configuration

**Note:** The following configuration details are specific to Spine 1. Spine 2 is similar. Complete configuration details are provided in the attachments named spine1.txt and spine2.txt.

Set the hostname, enable LLDP, and configure the management interface. Set the interface speed to 40GbE for all interfaces used for point-to-point links with the six leaf switches.

```
enable
configure

hostname Spine-1
protocol lldp
advertise management-tlv management-address system-description system-name
advertise interface-port-desc

interface ManagementEthernet 1/1
ip address 100.67.187.39/24
no shutdown
management route 0.0.0.0/0 100.67.187.254

stack-unit 1 port 1 portmode single speed 40G no-confirm
stack-unit 1 port 2 portmode single speed 40G no-confirm
```



```
stack-unit 1 port 3 portmode single speed 40G no-confirm
stack-unit 1 port 4 portmode single speed 40G no-confirm
stack-unit 1 port 5 portmode single speed 40G no-confirm
stack-unit 1 port 6 portmode single speed 40G no-confirm
```

The point to point interfaces and loop back interface are configured.

```
interface fortyGigE 1/1/1
description To Leaf 1 fo1/49
ip address 192.168.1.0/31
mtu 9216
no shutdown
```

```
interface fortyGigE 1/2/1
description To Leaf 2 fo1/49
ip address 192.168.1.2/31
mtu 9216
no shutdown
```

```
interface fortyGigE 1/3/1
description To Leaf 3 fo1/49
ip address 192.168.1.4/31
mtu 9216
no shutdown
```

```
interface fortyGigE 1/4/1
description To Leaf 4 fo1/49
ip address 192.168.1.6/31
mtu 9216
no shutdown
```

```
interface fortyGigE 1/5/1
description To Leaf 5 fo1/49
ip address 192.168.1.8/31
mtu 9216
no shutdown
```

```
interface fortyGigE 1/6/1
description To Leaf 6 fo1/49
ip address 192.168.1.10/31
mtu 9216
no shutdown
```

```
interface loopback 0
description Router ID
ip address 10.0.0.1/32
```

BGP is configured to allow routing to the IP fabric. Additionally, an IP prefix and route map are created to automatically redistribute all leaf subnets as well as loopback addresses from the leaf and spine switches.

```
route-map spine-leaf permit 10
match ip address spine-leaf

ip prefix-list spine-leaf
description BGP redistribute loopback and leaf networks
seq 5 permit 10.0.0.0/23 ge 32
seq 10 permit 10.0.0.0/8 ge 24
router bgp 64601
bgp bestpath as-path multipath-relax
maximum-paths ebgp 64
redistribute connected route-map spine-leaf
bgp graceful-restart
neighbor spine-leaf peer-group
neighbor spine-leaf fall-over
neighbor spine-leaf advertisement-interval 1
neighbor spine-leaf no shutdown
neighbor 192.168.1.1 remote-as 64701
neighbor 192.168.1.1 peer-group spine-leaf
neighbor 192.168.1.1 no shutdown
neighbor 192.168.1.3 remote-as 64702
neighbor 192.168.1.3 peer-group spine-leaf
neighbor 192.168.1.3 no shutdown
neighbor 192.168.1.5 remote-as 64703
neighbor 192.168.1.5 peer-group spine-leaf
neighbor 192.168.1.5 no shutdown
neighbor 192.168.1.7 remote-as 64704
neighbor 192.168.1.7 peer-group spine-leaf
neighbor 192.168.1.7 no shutdown
neighbor 192.168.1.9 remote-as 64705
neighbor 192.168.1.9 peer-group spine-leaf
neighbor 192.168.1.9 no shutdown
neighbor 192.168.1.11 remote-as 64706
neighbor 192.168.1.11 peer-group spine-leaf
neighbor 192.168.1.11 no shutdown
```

Create an ECMP group and include the point to point interfaces from the two spine switches.

```
ecmp-group 1
interface fortyGigE 1/1/1
interface fortyGigE 1/2/1
interface fortyGigE 1/3/1
interface fortyGigE 1/4/1
interface fortyGigE 1/5/1
```

```
interface fortyGigE 1/6/1
link-bundle-monitor enable
```

Save the configuration.

```
end
write
```

## 6.5 S4048-ON iSCSI SAN switch configuration

There are two iSCSI switches used in the deployment example in this guide. They are used in the Rack 2 compute cluster.

The following section outlines the configuration commands issued to the S4048-ON iSCSI switches. The switches start at their factory default settings per Section 6.1.

**Note:** The following configuration details are specific to iSCSI-1. iSCSI-2 is similar. Complete configuration details for both iSCSI SAN switches are provided in the attachments named `iscsi1.txt` and `iscsi2.txt`.

Initial configuration involves setting the hostname, and enabling LLDP. LLDP is useful for troubleshooting (see Section 9.8). Finally, the management interface and default gateway are configured.

```
enable
configure

hostname iSCSI-1
protocol lldp
advertise management-tlv management-address system-description system-name
advertise interface-port-desc

interface ManagementEthernet 1/1
ip address 100.67.187.6/24
no shutdown

management route 0.0.0.0/0 100.67.187.254
```

The four downstream interfaces (connected to the FC630 servers) are configured in the next set of commands. The interfaces are in placed in layer 2 mode with the `switchport` command and the MTU is set to 9216 for performance.

```
interface TenGigabitEthernet 1/17
description FC630 iSCSI
no ip address
mtu 9216
switchport
no shutdown

interface TenGigabitEthernet 1/18
```

```

description FC630 iSCSI
no ip address
mtu 9216
switchport
no shutdown

interface TenGigabitEthernet 1/19
description FC630 iSCSI
no ip address
mtu 9216
switchport
no shutdown

interface TenGigabitEthernet 1/20
description FC630 iSCSI
no ip address
mtu 9216
switchport
no shutdown

```

The two upstream interfaces connected to the SC4020 storage controllers are configured. The interfaces are in placed in layer 2 mode with the `switchport` command and the MTU is set to 9216 for performance.

```

interface TenGigabitEthernet 1/1
description To SC4020 CTRL1
no ip address
mtu 9216
switchport
no shutdown

interface TenGigabitEthernet 1/2
description To SC4020 CTRL2
no ip address
mtu 9216
switchport
no shutdown

```

VLAN interface 100 is created, and all interfaces used in the SAN are tagged in VLAN 100. An IP address on the iSCSI-1 network is assigned for troubleshooting. (Switch iSCSI-2 uses VLAN 200 with IP address 192.168.200.1).

```

interface Vlan 100
ip address 192.168.100.1/24
mtu 9216
tagged TenGigabitEthernet 1/1-1/2,1/17-1/20
no shutdown

```

Save the configuration.

```
end
write
```

## 6.6 S3048-ON management switch configuration

For the S3048-ON management switches, all ports used are in layer 2 mode and are in the default VLAN. No additional configuration is required.

## 6.7 Verify switch configuration

The following sections show commands and output to verify switches are configured and connected properly. Except where there are key differences, only output from one spine switch, one leaf switch, and one FN410S switch is shown to avoid repetition. Output from remaining devices will be similar.

### 6.7.1 Z9100-ON spine switch

#### 6.7.1.1 show ip bgp summary

This command verifies each BGP session to each of the six leaf switches is connected and sharing prefixes.

```
Spine-1#show ip bgp summary
```

```
BGP router identifier 10.0.0.1, local AS number 64601
BGP local RIB : Routes to be Added 0, Replaced 0, Withdrawn 0
16 network entrie(s) using 1216 bytes of memory
26 paths using 2808 bytes of memory
BGP-RIB over all using 2834 bytes of memory
41 BGP path attribute entrie(s) using 6880 bytes of memory
39 BGP AS-PATH entrie(s) using 390 bytes of memory
6 neighbor(s) using 49152 bytes of memory
```

Neighbor	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/Pfx
192.168.1.1	64701	5032	5013	0	0	0	3d:00:50:29	5
192.168.1.3	64702	17469	17469	0	0	0	00:00:00	Idle
192.168.1.5	64703	5031	5032	0	0	0	3d:00:50:44	5
192.168.1.7	64704	5030	5028	0	0	0	3d:00:50:29	5
192.168.1.9	64705	64	69	0	0	0	00:50:42	5
192.168.1.11	64706	62	66	0	0	0	00:49:33	5

### 6.7.1.2 show ip route bgp

This command is used to verify the BGP instance entries in the Routing Information Base (RIB) and ECMP. The first set of routes with a subnet mask of /32 are the IPs configured for router IDs.

The second set of routes with a /24 mask represents the 2 networks used, vMotion and NSX. Note that for each of these networks there are two routes. For example, 10.22.1.0/24 is reachable via 192.168.1.1 and 192.168.1.3, Leaf 1 and Leaf 2 respectively.

```
Spine-1#show ip route bgp
```

Destination	Gateway	Dist/Metric	Last Change
-----	-----	-----	-----
B EX 10.0.1.2/32	via 192.168.1.1	20/0	00:00:22
	via 192.168.1.3		
	via 192.168.1.9		
	via 192.168.1.7		
	via 192.168.1.5		
	via 192.168.1.11		
B EX 10.0.2.1/32	via 192.168.1.1	20/0	00:13:57
B EX 10.0.2.2/32	via 192.168.1.3	20/0	00:05:36
B EX 10.0.2.3/32	via 192.168.1.5	20/0	00:13:18
B EX 10.0.2.4/32	via 192.168.1.7	20/0	00:12:37
B EX 10.0.2.5/32	via 192.168.1.9	20/0	00:12:06
B EX 10.0.2.6/32	via 192.168.1.11	20/0	00:11:47
B EX 10.22.1.0/24	via 192.168.1.1	20/0	00:05:36
	via 192.168.1.3		
B EX 10.22.2.0/24	via 192.168.1.5	20/0	3d1h
	via 192.168.1.7		
B EX 10.22.3.0/24	via 192.168.1.9	20/0	01:41:59
	via 192.168.1.11		
B EX 10.55.1.0/24	via 192.168.1.1	20/0	00:05:36
	via 192.168.1.3		
B EX 10.55.2.0/24	via 192.168.1.5	20/0	3d1h
	via 192.168.1.7		
B EX 10.55.3.0/24	via 192.168.1.9	20/0	01:41:59
	via 192.168.1.11		

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

### 6.7.1.3 show ip route <network>

This command is used to verify that routes leading to the appropriate leaf switches are being propagated from BGP to the RIB. The commands for the 10.55.x.0 network are shown below as an example.

```
Spine-1#show ip route 10.55.1.0/24
```

```
Routing entry for 10.55.1.0/24
```

```
Known via "bgp 64601", distance 20, metric 0
```

```
Last update 00:23:33 ago
```

```
Tag value 64701
Routing Descriptor Blocks:
* via 192.168.1.1
* via 192.168.1.3
```

```
Spine-1#show ip route 10.55.2.0/24
Routing entry for 10.55.2.0/24
  Known via "bgp 64601", distance 20, metric 0
  Last update 3d2h ago
  Tag value 64703
  Routing Descriptor Blocks:
    * via 192.168.1.5
    * via 192.168.1.7
```

```
Spine-1#show ip route 10.55.3.0/24
Routing entry for 10.55.3.0/24
  Known via "bgp 64601", distance 20, metric 0
  Last update 02:00:32 ago
  Tag value 64705
  Routing Descriptor Blocks:
    * via 192.168.1.9
    * via 192.168.1.11
```

#### 6.7.1.4 Ping VRRP addresses

Both spine switches must be able to ping the VRRP addresses configured on each leaf switch pair.

```
Spine-1#ping 10.22.1.254
Sending 5, 100-byte ICMP Echos to 10.22.1.254, timeout is 2 seconds:
!!!!
Success rate is 100.0 percent (5/5), round-trip min/avg/max = 0/0/0 (ms)
```

Repeat for other VRRP addresses as needed:

```
10.22.1.254, 10.22.2.254, 10.22.3.254
10.55.1.254, 10.55.2.254, 10.55.3.254
```

**Note:** VRRP addresses in this document use the format 10.vlan#.rack#.254.

## 6.7.2 S4048-ON leaf switch

### 6.7.2.1 show vlt brief

The Inter-chassis link (ICL) Link Status, Heart Beat Status, and VLT Peer Status must all be up. The role for one switch in the VLT pair will be primary and its peer switch (not shown) will be assigned the secondary role.

```
Leaf-1#show vlt brief
```

#### VLT Domain Brief

```
-----
Domain ID:                127
Role:                     Primary
Role Priority:             32768
ICL Link Status:          Up
HeartBeat Status:         Up
VLT Peer Status:          Up
Local Unit Id:            0
Version:                  6(7)
Local System MAC address:  f4:8e:38:20:37:29
Remote System MAC address: f4:8e:38:20:54:29
Remote system version:    6(7)
Delay-Restore timer:      90 seconds
Delay-Restore Abort Threshold: 60 seconds
Peer-Routing :            Disabled
Peer-Routing-Timeout timer: 0 seconds
Multicast peer-routing timeout: 150 seconds
```

#### 6.7.2.2 show vlt detail

On leaf switches 1 and 2, downstream LAGs (port channels 2,4, and 6) will all be down until LAGs are configured on the directly connected ESXi hosts (covered in Section 9.4). VLANs 1, 22 and 55 are active.

Leaf-1#**show vlt detail**

Local LAG Id	Peer LAG Id	Local Status	Peer Status	Active VLANs
2	2	DOWN	DOWN	1, 22, 55
4	4	DOWN	DOWN	1, 22, 55
6	6	DOWN	DOWN	1, 22, 55

Leaf switches 5 and 6 have three additional downstream lags (port channels 12, 14 and 16) for edge traffic on VLAN 66. Port channels 12, 14, and 16 will be down until edge-lag2 is configured in Section 13.1.2.

Leaf-5#**show vlt detail**

Local LAG Id	Peer LAG Id	Local Status	Peer Status	Active VLANs
2	2	DOWN	DOWN	1, 22, 55
4	4	DOWN	DOWN	1, 22, 55
6	6	DOWN	DOWN	1, 22, 55
12	12	DOWN	DOWN	1, 66
14	14	DOWN	DOWN	1, 66
16	16	DOWN	DOWN	1, 66

On leaf switches 3 and 4, downstream port channel 128 is up because they are connected to properly configured FN410S switches. VLANs 1, 22 and 55 are active.



```
Leaf-3#show vlt detail
```

Local LAG Id	Peer LAG Id	Local Status	Peer Status	Active VLANs
128	128	UP	UP	1, 22, 55

### 6.7.2.3 show vrrp brief

The output from the `show vrrp brief` command should be similar to that shown below. The priority (Pri column) of the master router in the pair is 254 and the backup router (not shown) is assigned priority 1.

```
Leaf-1#show vrrp brief
```

Interface	Group	Pri	Pre	State	Master addr	Virtual addr(s)	Description
Vl 22	IPv4 22	254	Y	Master	10.22.1.252	10.22.1.254	vMotion
Vl 55	IPv4 55	254	Y	Master	10.55.1.252	10.55.1.254	NSX

## 6.7.3 FN410S I/O Module

### 6.7.3.1 show vlt brief

Like the S4048-ON switches above, the ICL Link Status, Heart Beat Status, and VLT Peer Status must all be up. One switch is primary and the peer (not shown) is the secondary.

```
FN410S-A1#show vlt brief
```

```
VLT Domain Brief
```

```
-----
Domain ID:                127
Role:                     Primary
Role Priority:             32768
ICL Link Status:          Up
HeartBeat Status:         Up
VLT Peer Status:          Up
Local Unit Id:            0
Version:                  6(7)
Local System MAC address:  f8:b1:56:6e:fc:5b
Remote System MAC address: f8:b1:56:76:b9:b5
Remote system version:    6(7)
Delay-Restore timer:      90 seconds
Delay-Restore Abort Threshold: 60 seconds
Peer-Routing :           Disabled
Peer-Routing-Timeout timer: 0 seconds
Multicast peer-routing timeout: 150 seconds
```

### 6.7.3.2 show vlt detail

Downstream LAGs (port channels 1-4) are down until LAGs are configured on the directly connected ESXi hosts running on the FC630 servers. This is covered in Section 9.4.

The upstream LAG (port channel 128) is currently up because it is connected to properly configured leaf switches (Leaf 3 and Leaf 4).

VLANs 1, 22 and 55 are active on all LAGs.

```
FN410S-A1#show vlt detail
```

Local LAG Id	Peer LAG Id	Local Status	Peer Status	Active VLANs
1	1	DOWN	DOWN	1, 22, 55
2	2	DOWN	DOWN	1, 22, 55
3	3	DOWN	DOWN	1, 22, 55
4	4	DOWN	DOWN	1, 22, 55
128	128	UP	UP	1, 22, 55

#### 6.7.4 S4048-ON iSCSI SAN switch

Verification of the iSCSI SAN switches is done in section 10.3.

## 7 Prepare Servers

This section covers basic PowerEdge server preparation and ESXi hypervisor installation. Installation of guest operating systems (Microsoft Windows Server, Red Hat Linux, etc.) is outside the scope of this document.

**Note:** Exact iDRAC console steps in this section may vary slightly depending on hardware, software and browser versions used. See your PowerEdge server documentation for steps to connect to the iDRAC virtual console.

### 7.1 Confirm CPU virtualization is enabled in BIOS

**Note:** CPU virtualization is typically enabled by default in PowerEdge server BIOS. These steps are provided for reference in case this required feature has been disabled.

1. Connect to the iDRAC in a web browser and launch the virtual console.
2. In the virtual console, from the **Next Boot** menu, select **BIOS Setup**.
3. Reboot the server.
4. From the **System Setup Main Menu**, select **System BIOS**, and then select **Processor Settings**.
5. Verify **Virtualization Technology** is set to **Enabled**.
6. To save the settings, click **Back**, **Finish**, and **Yes** if prompted to save changes.
7. If resetting network adapters to defaults, proceed to step 4, **System Setup Main Menu**, in the next section. Otherwise, reboot the server.

### 7.2 Confirm network adapters are at factory default settings

**Note:** These steps are only necessary if installed network adapters have been modified from their factory default settings.

1. Connect to the iDRAC in a web browser and launch the virtual console.
2. In the virtual console, from the **Next Boot** menu, select **BIOS Setup**.
3. Reboot the server.
4. From the **System Setup Main Menu**, select **Device Settings**.
5. From the **Device Settings** page, select the first port of the first NIC in the list.
6. From the **Main Configuration Page**, click the **Default** button followed by **Yes** to load the default settings. Click **OK**.
7. To save the settings, click **Finish** then **Yes** to save changes. Click **OK**.
8. Repeat for each NIC and port listed on the **Device Settings** page.
9. Reboot the server.

## 7.3 Install ESXi

Dell EMC recommends using the latest Dell EMC customized ESXi .iso image available on [support.dell.com](http://support.dell.com). The correct drivers for your PowerEdge hardware are built into this image.

Install ESXi on all servers that will be part of your deployment. For the example in this guide, ESXi is installed to redundant (mirrored) internal SD cards in the PowerEdge servers. This includes six R630 servers (in the management and edge clusters) and four FC630 servers (in the compute cluster).

A simple way to install ESXi on a PowerEdge server remotely is by using the iDRAC to boot the server directly to the ESXi .iso image. This is done as follows:

1. Connect to the iDRAC in a web browser and launch the virtual console.
2. In the virtual console, select **Virtual Media > Connect Virtual Media**.
3. Select **Virtual Media > Map CD/DVD** > browse to the Dell EMC customized ESXi .iso image > **Open > Map Device**.
4. Select **Next Boot > Virtual CD/DVD/ISO > OK**.
5. Select **Power > Reset System (warm boot)**. Answer **Yes** to reboot the server.
6. The server reboots to the ESXi .iso image and installation starts.
7. Follow the prompts to install ESXi. Select the server's Internal Dual SD Module (IDSMD) when prompted for a location.
8. After installation is complete, click **Virtual Media > Disconnect Virtual Media > Yes**.
9. Reboot the system when prompted.

## 7.4 Configure the ESXi management network connection

Be sure the host is physically connected to the management network. For this deployment, the Intel I350-T 1GbE add-in PCIe adapter provides this connection for R630 servers and FC630 servers.

1. Log in to the ESXi console and select **Configure Management Network > Network Adapters**.
2. Select the correct vmnic for the management network connection. Follow the prompts on the screen to make the selection.
3. Go to **Configure Management Network > IPv4 Configuration**. If DHCP is not used, specify a static IP address, mask, and default gateway for the management interface.
4. Optionally, configure DNS settings from the **Configure Management Network** menu if DNS is used on your network.
5. Press **Esc** to exit and answer **Y** to apply the changes.
6. From the ESXi main menu, select **Test Management Network**. Verify pings are successful. If there is an error, be sure you have configured the correct vmnic.
7. Optionally, under **Troubleshooting Options**, enable the ESXi shell and SSH to enable remote access to the CLI.
8. Log out of the ESXi console.

## 8 Deploy VMware vCenter Server and add hosts

### 8.1 Deploy VMware vCenter Server

VMware vCenter Server is required for managing clusters and NSX, as well as many other advanced vSphere features. vCenter Server can be installed as a Windows-based application or as a prepackaged SUSE Linux-based VM.

This guide uses the prepackaged VM, called the vCenter Server Appliance (VCSA) and its built in PostgreSQL database. VCSA 6.0 supports up to 1000 hosts and 10,000 VMs. VCSA is available for download at [my.vmware.com](http://my.vmware.com).

In this guide, VCSA is installed on a PowerEdge R630 server running ESXi. The server will be part of the management cluster.

**Note:** This section provides simplified VCSA installation instructions. Detailed instructions and information are provided in the VMware vCenter Server 6.0 Deployment Guide available at the following location: <https://www.vmware.com/files/pdf/techpaper/vmware-vcenter-server6-deployment-guide.pdf>

1. On a Windows workstation connected to the management network, mount the VCSA .iso image.
2. Install the Client Integration Plugin by running **\vcsa\VMWare-ClientIntegrationPlugin-6.0.0.exe**.
3. Open **\vcsa-setup.html** in a browser and accept the related warning prompts. Click **Install**.
  - a. Accept the license agreement and click **Next**.
  - b. Provide the ESXi host destination IP address, ESXi host username (root) and password. Click **Next**. Click **Yes** to accept the SSL certificate warning if prompted.
  - c. Provide a vCenter **Appliance name** (vctr01 for example), and **password**. Click **Next**.
  - d. Keep the default selection: **Install vCenter Server with an Embedded Platform Services Controller**. Click **Next**.
  - e. Select **Create a new SSO domain > Next**.
  - f. Provide an **SSO Password**, **SSO Domain name** (pct.lab for example), and **SSO Site name** (site for example).
  - g. Select an **Appliance size** depending on your requirements. For this guide **Medium (up to 400 hosts, 4000 VMs)** is selected. Click **Next**.
  - h. Select a **datastore**. Optionally, if space is limited, check the **Enable Thin Disk Mode** box. Click **Next**.
  - i. Keep the default selection: **Use an embedded database (PostgreSQL)**. Click **Next**.
  - j. Under **Network Settings**:
    - i. Keep the default network, **VMNetwork**.
    - ii. Select **IPv4** and the network type (**static or DHCP**). A static address is used in this guide.
    - iii. If **static** was selected, provide a **Network address**, **System name** (if not using a fully qualified domain name, retype the Network address), **Subnet mask**, **Network gateway**, and **DNS server**.
    - iv. Under **Configure time sync**, select **Synchronize appliance time with ESXi host**.

**Note:** If you select **Use NTP (Network Time Protocol) servers**, a warning appears at the bottom of the screen indicating deployment will fail if the ESXi host clock is not in sync with the NTP server. Since the

ESXi hosts are not yet configured for NTP, select **Synchronize appliance time with ESXi host**. ESXi hosts are configured for NTP in Section 8.5.

- v. Checking **Enable SSH** is optional. Click **Next**. Click **OK** if a fully qualified domain name (FQDN) recommendation box is displayed.
- k. Joining the **VMWare Customer Experience Improvement Program** is recommended but optional. Select an option and click **Next**.
- l. Review the summary page and click **Finish** if all settings are correct.

vCenter Server is installed as a virtual machine on the ESXi host. When complete, the message shown in Figure 33 is displayed.

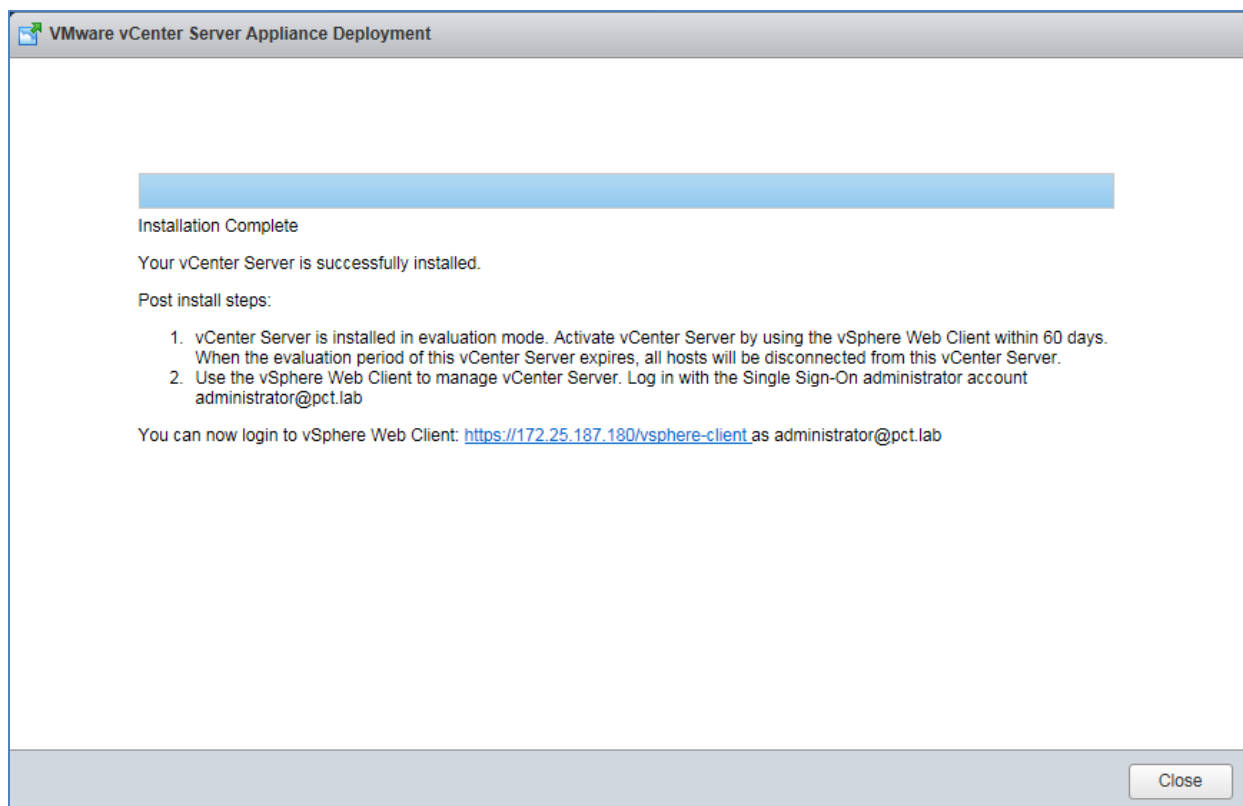


Figure 33 vCenter Server installation complete

## 8.2 Connect to the vSphere Web Client

**Note:** The vSphere Web Client is a service running on vCenter Server.

Connect to the vSphere Web Client in a browser by entering the following in the address bar:

**`https://<ip-address-or-hostname-of-vCenter-appliance>/vsphere-client`**

Log in with your vCenter credentials. After log in, the web client home page is displayed as shown in Figure 34.

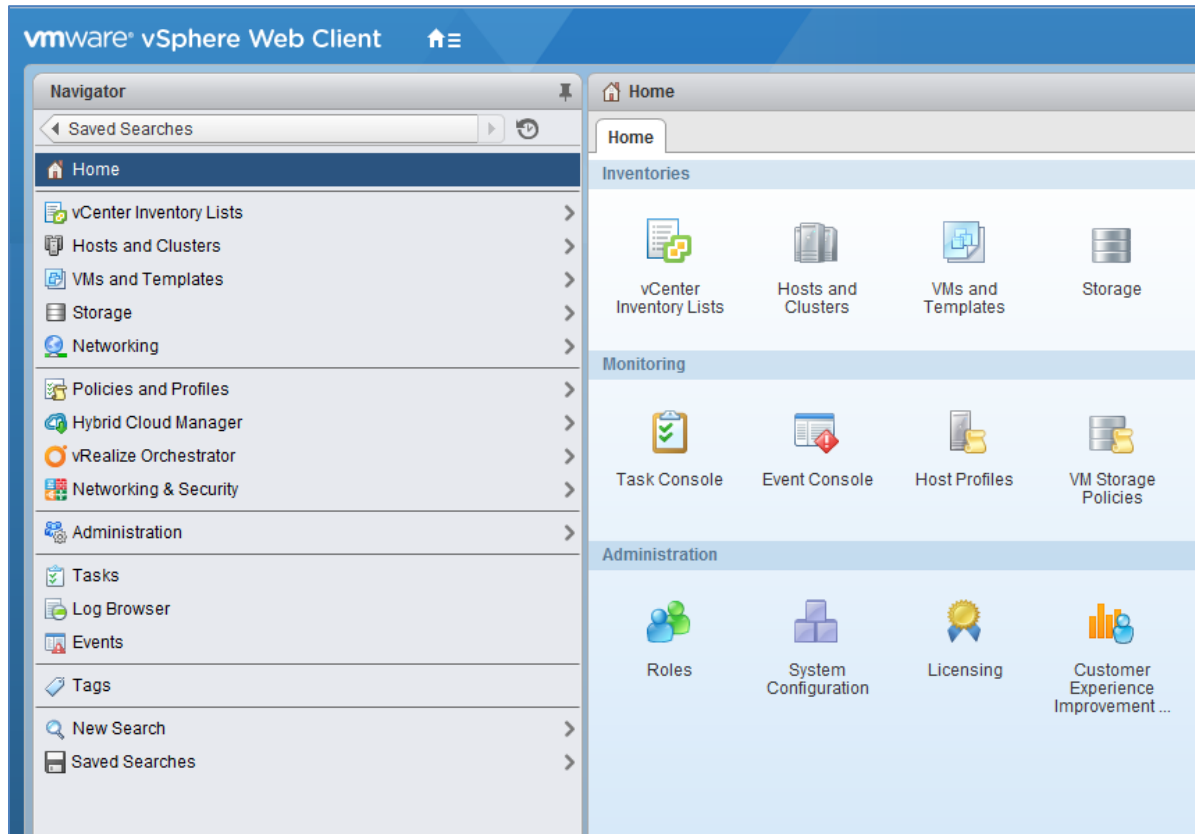


Figure 34 vSphere Web Client home page

The vast majority of management, configuration, and monitoring of your vSphere and NSX environment is done in the web client.

## 8.3 Install VMware licenses

The VMware licenses required for this deployment are listed in Appendix B.2. All VMware products used in this guide come with evaluation licenses that can be used for up to 60 days.

To install one or more product licenses:

1. From the web client **Home** page, select **Licensing** in the center pane.
2. On the **Licenses** page, select the **Licenses** tab and click the **+** icon, and type or paste license keys into the box provided. Click **Next**.
3. Provide **License names** for the keys or use the defaults. Click **Next > Finish**.

When complete, the Licenses page will look similar to Figure 35.

License	License Key	Product	Usage	Capacity
vcenter server		VMware vCenter Server 6 Standard (Instances)	1 Instances	1 Instances
NSX		NSX for vSphere Evaluation (CPUs)	0 CPUs	100 CPUs
vsphere-esxi		VMware vSphere 6 Enterprise Plus (CPUs)	16 CPUs	64 CPUs

Figure 35 Licenses tab

Licenses may then be assigned as needed from the **Assets** tab.

## 8.4 Create a data center object and add hosts

A data center object needs to be created before hosts can be added. This guide uses a single data center object named Datacenter.

To create a data center object:

1. On the web client **Home** screen, select **Hosts and Clusters**.
2. In the **Navigator** pane, right click the vCenter Server object and select **New Datacenter**.
3. Provide a name (Datacenter) and click **OK**.

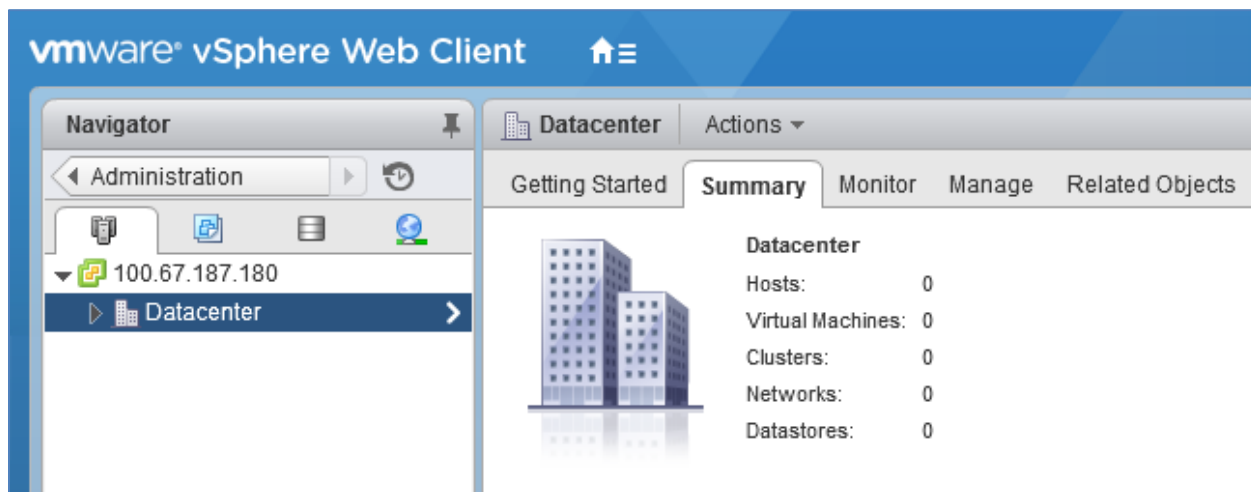


Figure 36 Datacenter created



To add ESXi hosts to the data center:

1. On the web client **Home** screen, select **Hosts and Clusters**.
2. In the **Navigator** pane, right click on **Datacenter** and select **Add Host**.
3. Specify the **IP address** of an ESXi host (or the **host name** if DNS is configured on your network). Click **Next**.
4. Enter the credentials for the ESXi host and click **Next**. If a security certificate warning box is displayed, click **Yes** to proceed.
5. On the **Host summary** screen, click **Next**.
6. Assign a license or select the evaluation license. This guide uses a VMware vSphere 6 Enterprise Plus license for ESXi hosts. Click **Next**.
7. Select a **Lockdown mode**. This guide uses the default setting, **Disabled**. Click **Next**.
8. For the VM location, select **Datacenter** and click Next.
9. On the **Ready to complete** screen, select **Finish**.

Repeat for all servers running ESXi that will be part of the NSX environment. This deployment example uses four FC630 servers and six R630 servers for a total of ten hosts running ESXi. When complete, all ESXi hosts will be added to the **Datacenter** object as shown in Figure 37.

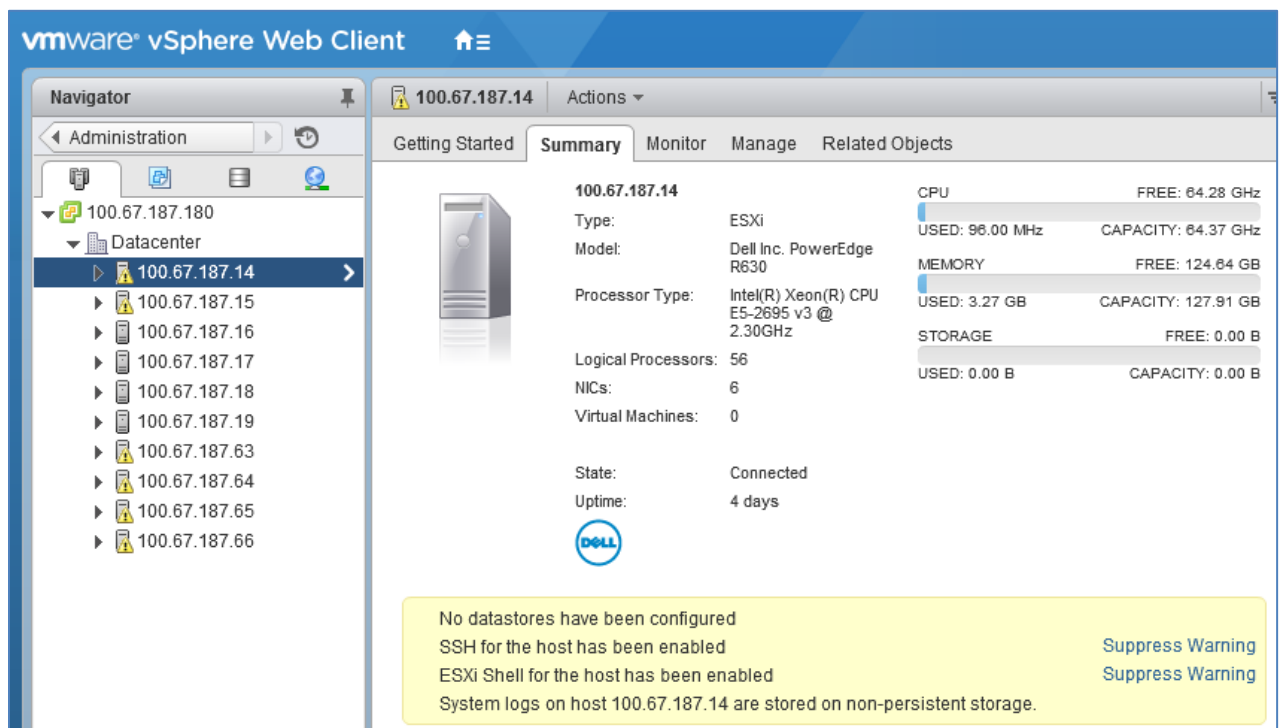



Figure 37 ESXi hosts added to the data center object

Some (or all) hosts may have a warning icon (  ) as shown in Figure 37. By selecting the host and going to the **Summary** tab, warning messages can be viewed.

The following warning messages may appear:

- *No datastores have been configured.* This message will be resolved when either a local datastore or a shared storage datastore is configured on the host. iSCSI datastore configuration (shared storage) is covered in Section 10, or see your ESXi documentation to create a local datastore.
- *SSH (or ESXi Shell) for the host has been enabled.* These messages will appear if either feature is enabled (as described in Section 7.4). If the behavior is desired, you may click **Suppress Warning** to remove the messages.
- *System logs on host are stored on non-persistent storage.* This message may appear when ESXi is installed to the redundant internal SD cards. This can be resolved by moving the system logs to either a local or shared storage datastore. iSCSI datastore configuration (shared storage) is covered in Section 10, or see your ESXi documentation to create a local datastore. Resolution is documented in VMware Knowledge Base article [2032823](#).

## 8.5 Ensure hosts are configured for NTP

It is a best practice to use NTP on the management network to keep time synchronized in an NSX environment. Ensure NTP is configured on ESXi hosts as follows:

1. On the web client **Home** screen, select **Hosts and Clusters**.
2. In the **Navigator** pane, select a host.
3. In the center pane, go to **Manage > Settings > Time Configuration**. If the information shown is correct (see Figure 38), skip to step 7. Otherwise, continue to step 4.
4. If NTP has not been configured properly, click **Edit**.
5. In the **Edit Time Configuration** dialog box:
  - a. Select **Use Network Time Protocol** radio button.
  - b. Next to **NTP Service Startup Policy**, select **Start and stop with host**.
  - c. Next to **NTP servers**, enter the IP address or FQDN of the NTP server.
  - d. Click **Start** to start the NTP client followed by **OK** to close the dialog box.
6. The **Time Configuration** page for the host should appear similar to Figure 38.

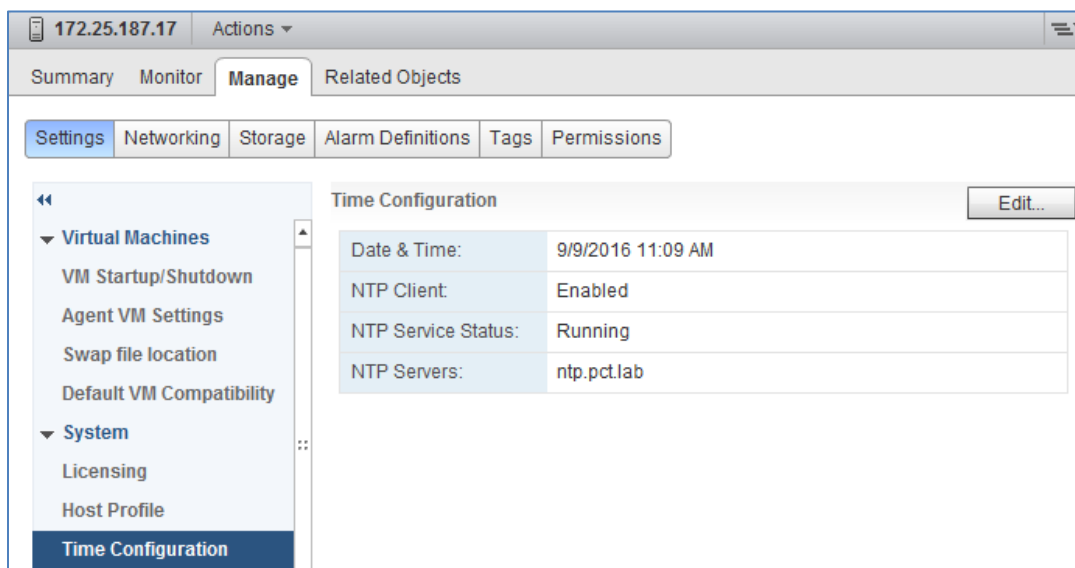


Figure 38 Proper NTP configuration on ESXi host

7. Repeat for remaining ESXi hosts as needed.

## 8.6 Create clusters and add hosts

When a host is added to a cluster, the host's resources become part of the cluster's resources. The cluster manages the resources of all hosts within it. Clusters enable features such as High Availability (HA), Distributed Resource Scheduler (DRS), and Virtual SAN (VSAN). For this guide, three clusters are created, with one cluster per rack:

- Rack 1 Management
- Rack 2 Compute FC630
- Rack 3 Edge

All ESXi hosts are added to one of the above clusters.

To add clusters to the data center object:

1. On the web client **Home** screen, select **Hosts and Clusters**.
2. In the **Navigator** pane, right click **Datacenter** and select **New Cluster**.
3. Name the cluster. For this example, the first cluster is named **Rack 1 Management**. Leave **DRS**, **vSphere HA**, **EVC** and **Virtual SAN** at their default settings (**Off/Disabled**). Click **OK**.

**Note:** vSphere DRS, HA, EVC, and Virtual SAN cluster features are outside the scope of this guide. For more information on DRS, HA, and EVC see the [VMware vSphere 6.0 Documentation](#). Dell EMC NSX guides with VSAN storage are listed in Appendix C.1.

Repeat for the remaining two clusters:

- Rack 2 Compute FC630
- Rack 3 Edge

In the Navigator pane, drag and drop ESXi hosts into the appropriate clusters. The three ESXi hosts on R630 servers in Rack 1 are placed in the **Rack 1 Management** cluster, the four ESXi hosts on FC630 servers in Rack 2 are placed in the **Rack 2 Compute FC630** cluster, and the three ESXi hosts on R630 servers in Rack 3 are placed in the **Rack 3 Edge** cluster.

When complete, each cluster (🖨️) contains its assigned hosts (🖨️) as shown in Figure 39:

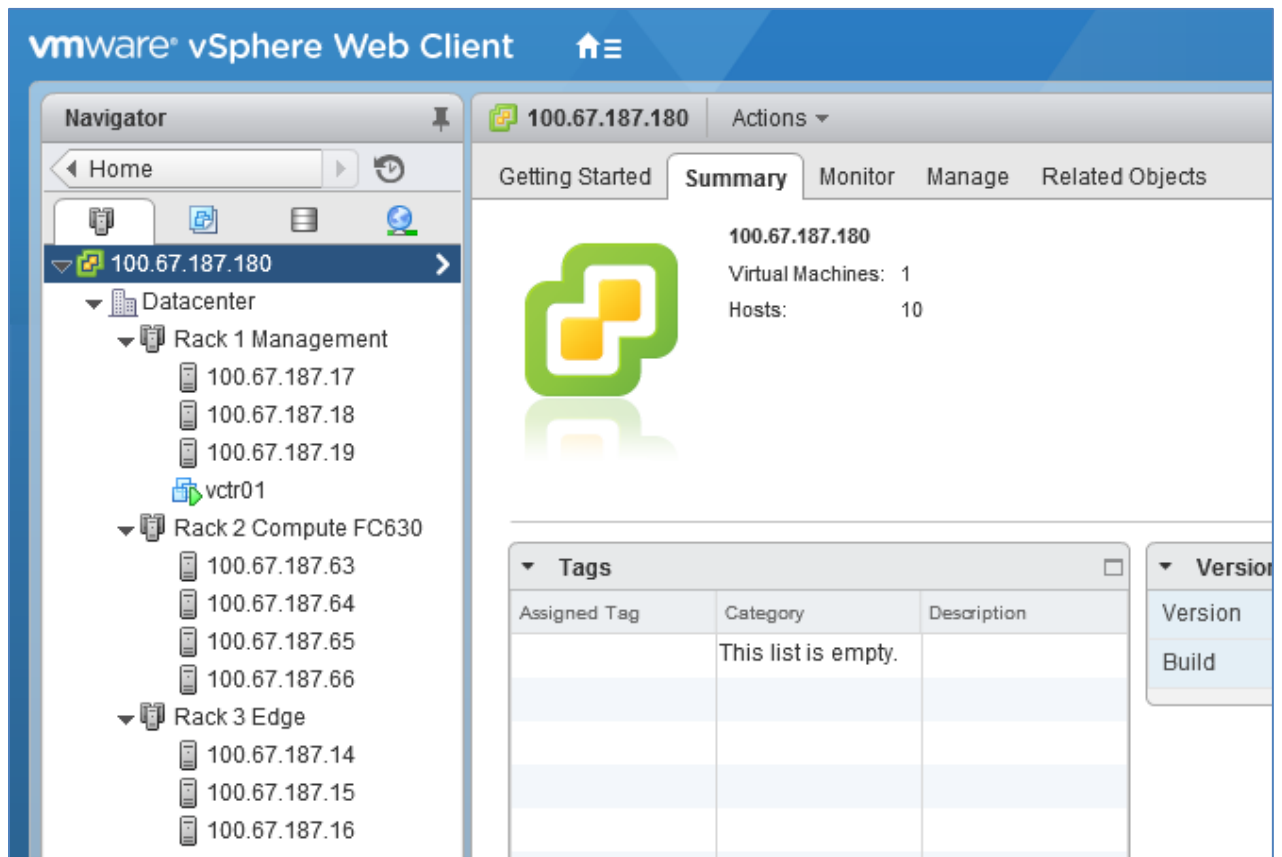


Figure 39 Clusters and hosts after initial configuration

**Note:** The vCenter Server Appliance, vctr01, is also shown in the Rack 1 Management cluster in Figure 39.

## 8.7 Information on vSphere standard switches

A vSphere standard switch (also referred to as a VSS or a standard switch) is a virtual switch that handles network traffic at the host level in a vSphere deployment. Standard switches provide network connectivity to hosts and virtual machines.

A standard switch named vSwitch0 is automatically created on each ESXi host during installation to provide connectivity to the management network.

Standard switches may be viewed, and optionally configured, as follows:

1. Go to the web client **Home** page, select **Hosts and Clusters**, and select a host in the **Navigator** pane.
2. In the center pane, select **Manage > Networking > Virtual switches**.
3. Standard switch **vSwitch0** appears in the list. Click on it to view details as shown in Figure 40.

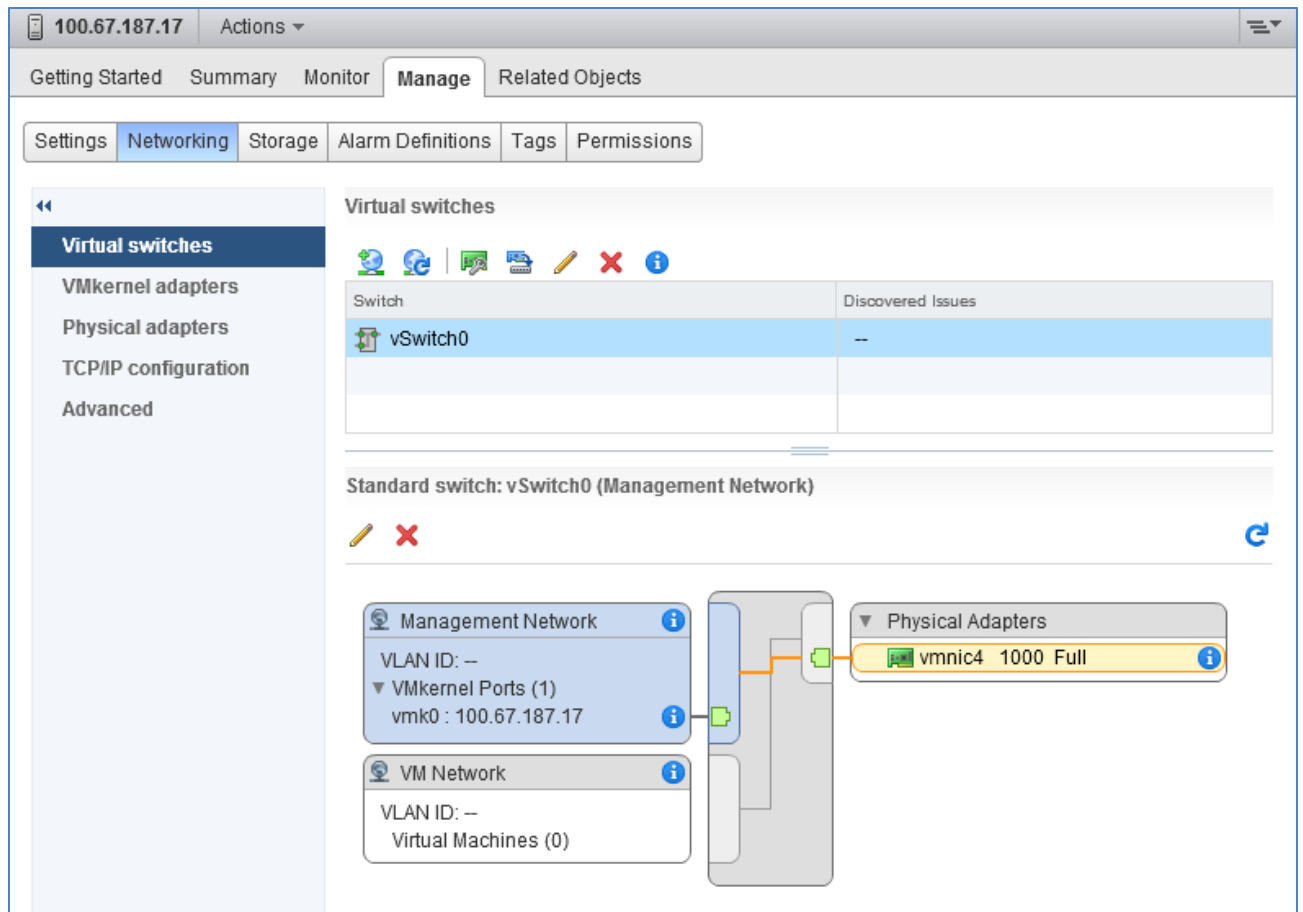


Figure 40 vSphere standard switch

**Note:** For this guide, only the default configuration is required on the standard switches. Standard switches are only used in this deployment for connectivity to the management network. Distributed switches, covered in the next section, are used for connectivity to the production network.

## 9 Deploy vSphere distributed switches for vMotion

A vSphere Distributed Switch (also referred to as a VDS or a distributed switch) is a virtual switch that provides network connectivity to hosts and virtual machines. Unlike vSphere standard switches, distributed switches act as a single switch across multiple hosts in a cluster. This lets virtual machines maintain consistent network configurations as they migrate across multiple hosts.

Distributed switches are configured in the web client and the configuration is populated across all hosts associated with the switch. They are used for connectivity to the Production network in this guide.

Distributed Switches contain two different port groups:

- **Uplink port group** – an uplink port group maps physical NICs on the hosts (vmnics) to uplinks on the VDS. Uplink port groups act as trunks and carry all VLANs by default.

**Note:** For consistent network configuration, you can connect the same physical NIC port on every host to the same uplink port on the distributed switch. For example, if you are adding two hosts, connect vmnic1 on each host to Uplink1 on the distributed switch.

- **Distributed port group** - Distributed port groups define how connections are made through the VDS to the network. In this guide, one distributed port group is created for each VLAN, and one for each VXLAN Network ID (VNI).

In this section, one VDS is created for each of the three clusters, and each VDS is shared by all hosts in the cluster. The three distributed switches used in this deployment are named:

- Rack 1 Management VDS
- Rack 2 Compute FC630 VDS
- Rack 3 Edge VDS

### 9.1 Create a VDS for each cluster

Create the first VDS named **Rack 1 Management VDS**:

1. On the web client **Home** screen, select **Networking**.
2. Right click on **Datacenter**. Select **Distributed switch > New Distributed Switch**.
3. Provide a name for the first VDS, **Rack 1 Management VDS**. Click **Next**.
4. On the **Select version** page, select **Distributed switch: 6.0.0 > Next**.
5. On the **Edit settings** page:
  - a. Leave the **Number of uplinks** set to **4** (this field to be replaced by LAGs later).
  - b. Leave **Network I/O Control** set to **Enabled**.
  - c. **Uncheck** the **Create a default port group** box.
6. Click **Next** followed by **Finish**.
7. The VDS is created with the uplink port group shown beneath it.

Repeat steps 1-7 above, substituting the switch name in step 3 to create the remaining two distributed switches, **Rack 2 Compute FC630 VDS** and **Rack 3 Edge VDS**.

When complete, the Navigator pane should look similar to Figure 41.

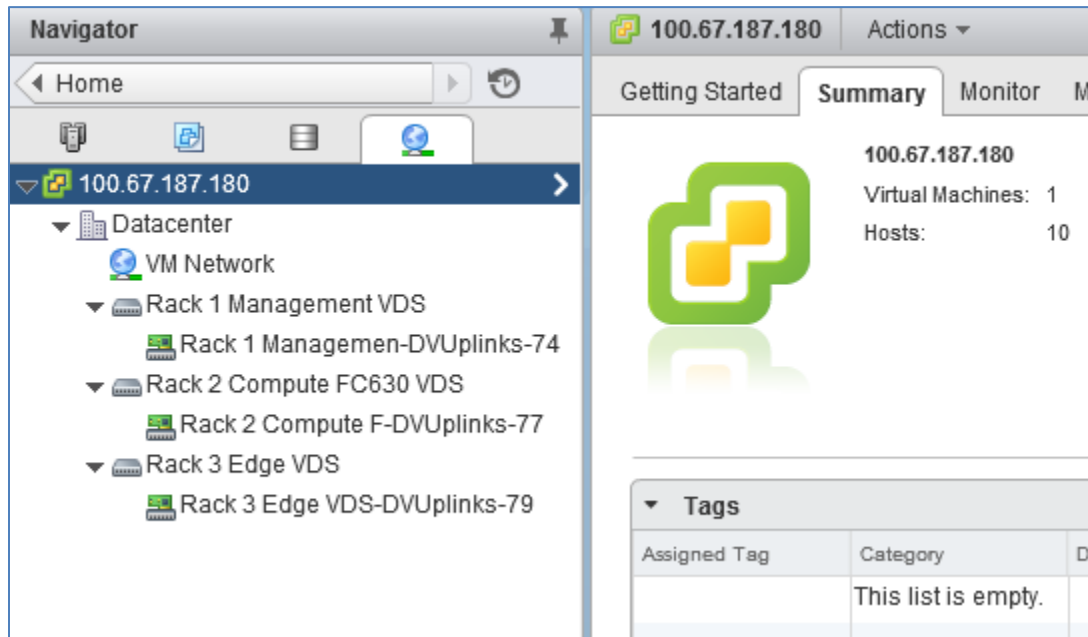


Figure 41 VDS created for each cluster

## 9.2 Add distributed port groups

In this section, a distributed port group for vMotion traffic is added to each VDS.

To create the port group for vMotion traffic on the **Rack 1 Management VDS**:

1. On the web client **Home** screen, select **Networking**.
2. Right click on **Rack 1 Management VDS**. Select **Distributed Port Group > New Distributed Port Group**.
3. On the **Select name and location** page, provide a name for the distributed port group, for example, **R1 Management vMotion**. Click **Next**.
4. On the **Configure settings** page, next to **VLAN type**, select **VLAN**. Set the **VLAN ID** to **22** for the vMotion port group. Leave other values at their defaults as shown in Figure 42.
5. Click **Next > Finish**.

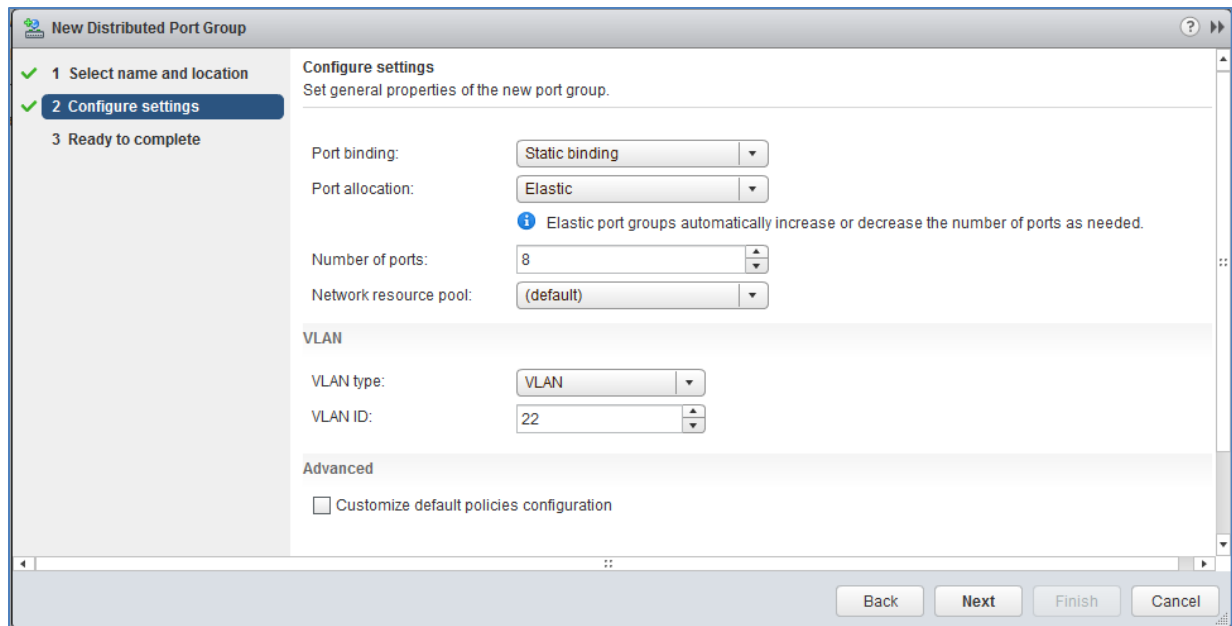


Figure 42 Distributed port group settings page – vMotion port group

Repeat the above for the remaining two distributed switches, **Rack 2 Compute FC630 VDS** and **Rack 3 Edge VDS**.

When complete, the Navigator pane will appear similar to Figure 43.

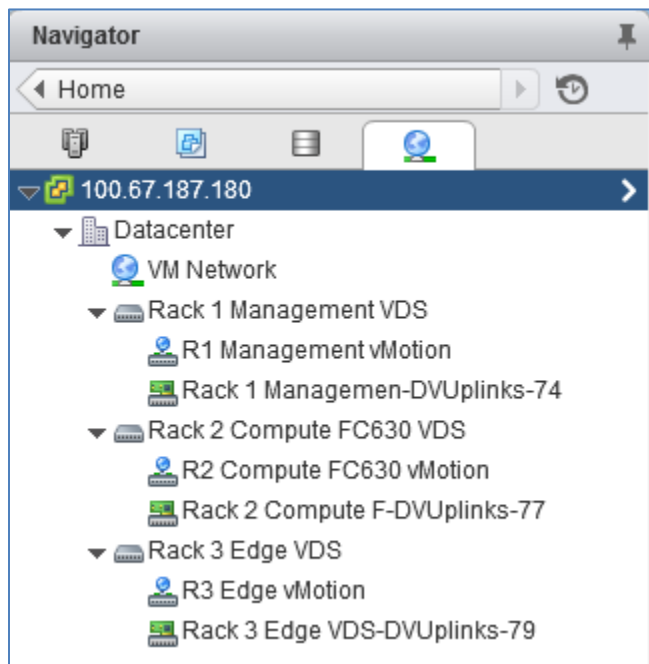


Figure 43 Distributed switches with vMotion port groups created

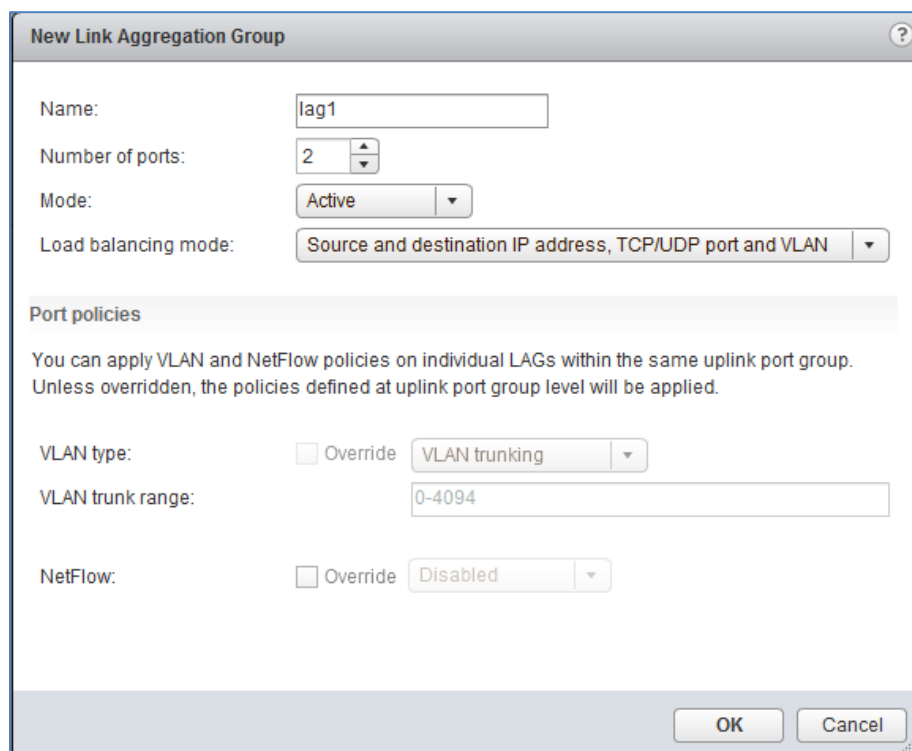


## 9.3 Create LACP LAGs

Since Link Aggregation Control Protocol (LACP) LAGs are used in the physical network between ESXi hosts and physical switches, LACP LAGs are also configured on each VDS.

To enable LACP on **Rack 1 Management VDS**:

1. On the web client **Home** screen, select **Networking**.
2. In the Navigator pane, select **Rack 1 Management VDS**.
3. In the center pane, select **Manage > Settings > LACP**.
4. Click the **+** icon. The **New Link Aggregation Group** dialog box opens.
5. Set **Number of ports** equal to the number of physical uplinks on each ESXi host. In this deployment, R630 hosts have two ports in a LAG connected to the upstream switches so this number is set to **2** for the Management VDS. (When configuring the other VDSs, the **Compute FC630 VDS** uses **4** ports and the **Edge VDS** uses **2**).
6. Set the **Mode** to **Active**. The remaining fields can be set to their default values as shown in Figure 44.




The image shows a 'New Link Aggregation Group' dialog box. It has a title bar with a question mark icon. The fields are as follows:

- Name:** A text box containing 'lag1'.
- Number of ports:** A spinner box set to '2'.
- Mode:** A dropdown menu set to 'Active'.
- Load balancing mode:** A dropdown menu set to 'Source and destination IP address, TCP/UDP port and VLAN'.
- Port policies:** A section with a description: 'You can apply VLAN and NetFlow policies on individual LAGs within the same uplink port group. Unless overridden, the policies defined at uplink port group level will be applied.'
- VLAN type:** A checkbox labeled 'Override' is unchecked, followed by a dropdown menu set to 'VLAN trunking'.
- VLAN trunk range:** A text box containing '0-4094'.
- NetFlow:** A checkbox labeled 'Override' is unchecked, followed by a dropdown menu set to 'Disabled'.

At the bottom right, there are 'OK' and 'Cancel' buttons.

Figure 44 LAG configuration

7. Click **OK** to close the dialog box.

This creates **lag1** on the VDS. The refresh icon () at the top of the screen may need to be clicked for the lag to appear in the table as shown in Figure 45.

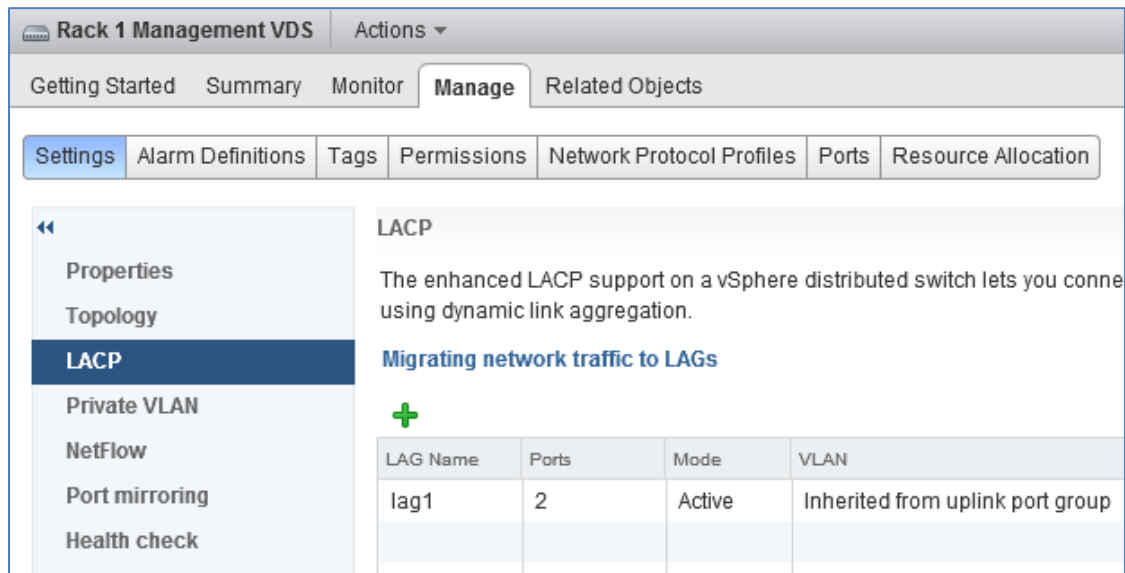


Figure 45 Lag1 created on Rack 1 Management VDS



Repeat steps 1-7 above for the remaining two distributed switches, **Rack 2 Compute FC630 VDS** and **Rack 3 Edge VDS**.



## 9.4 Associate hosts and assign uplinks to LAGs

Hosts and their vmnics must be associated with each vSphere distributed switch.

**Note:** Before starting this section, be sure you know the vmnic-to-physical adapter mapping for each host. This can be determined by going to **Home > Hosts and Clusters** and selecting the host in the **Navigator** pane. In the center pane select **Manage > Networking > Physical adapters**. In this example, vmnics used are numbered vmnic1 and vmnic3. Vmnic numbering will vary depending on adapters installed in the host.

To add hosts to Rack 1 Management VDS:

1. On the web client **Home** screen, select **Networking**.
1. Right click on **Rack 1 Management VDS** and select **Add and Manage Hosts**.
2. In the **Add and Manage Hosts** dialog box:
  - a. On the **Select task** page, make sure **Add hosts** is selected. Click **Next**.
  - b. On the **Select hosts** page, Click the  **New hosts** icon. Select the check box next to each host in the **Rack 1 Management** cluster. Click **OK > Next**.
  - c. On the **Select network adapters tasks** page, be sure the **Manage physical adapters** box is checked. Be sure all other boxes are unchecked. Click **Next**.
  - d. On the **Manage physical network adapters** page, each host is listed with its vmnics beneath it.
    - i. Select the first vmnic (vmnic1 in this example) on the first host and click  **Assign uplink**.

- ii. Select **lag1-0** > **OK**.
- iii. Select the second vmnic (vmnic3 in this example) on the first host and click  **Assign uplink**.
- iv. Select **lag1-1** > **OK**.
- e. Repeat steps i – iv for the remaining hosts. Click **Next** when done.
- f. On the **Analyze impact** page, **Overall impact status** should indicate  **No impact**.
- g. Click **Next** > **Finish**.

When complete, the **Manage > Settings > Topology** page for **Rack 1 Management VDS** should look similar to Figure 46.

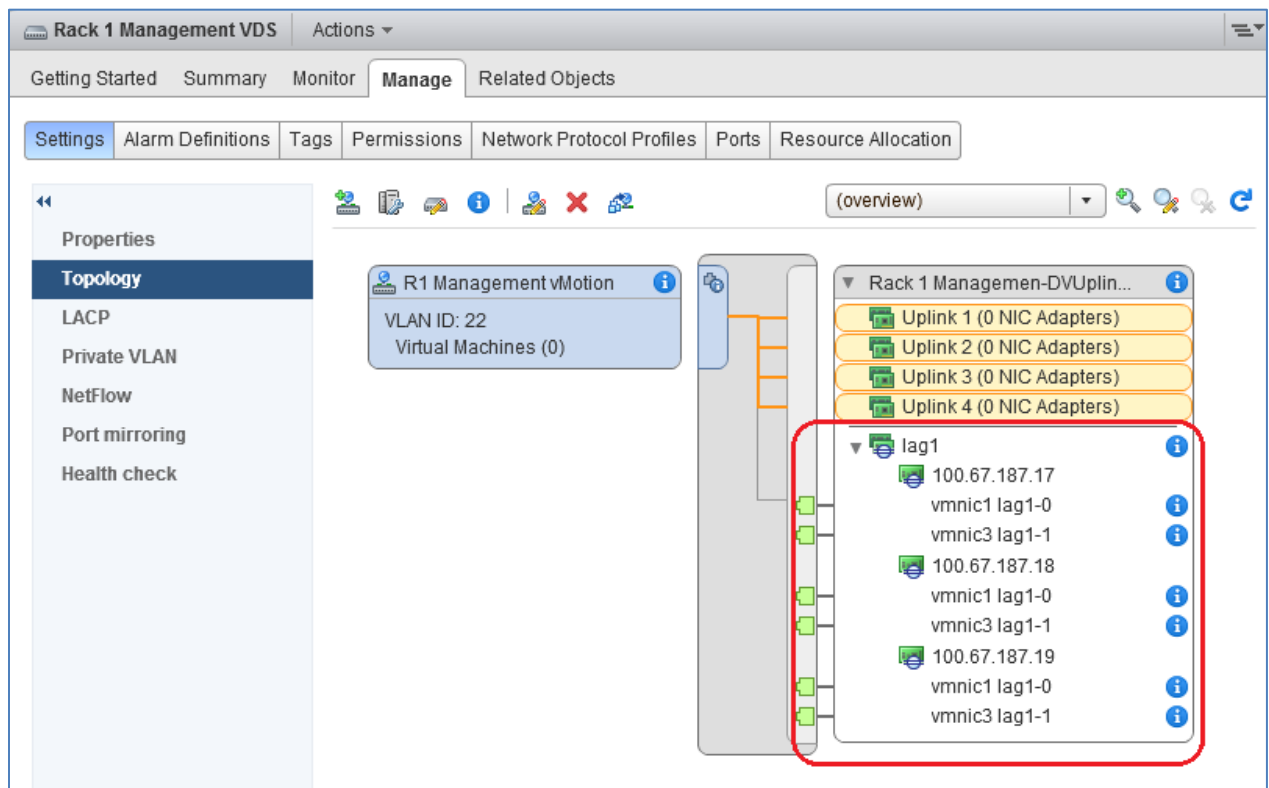


Figure 46 LAGs configured on Rack 1 Management VDS

Repeat steps 1-4 above for the remaining two distributed switches, Rack 2 Compute FC630 VDS and Rack 3 Edge VDS with one important change: Hosts connected to Rack 2 Compute FC630 VDS use four vmnics instead of two. In step 3.d., when configuring Rack 2 Compute FC630 VDS, assign the four vmnics on each FC630 to lags 1-0 through 1-3. Rack 3 Edge VDS uses two vmnics like Rack 1 Management VDS.

When complete, the **Manage > Settings > Topology** page for Rack 3 Edge VDS will look similar to Rack 1 Management VDS in Figure 46. The **Manage > Settings > Topology** page for Rack 2 Compute FC630 VDS will look similar to Figure 47.

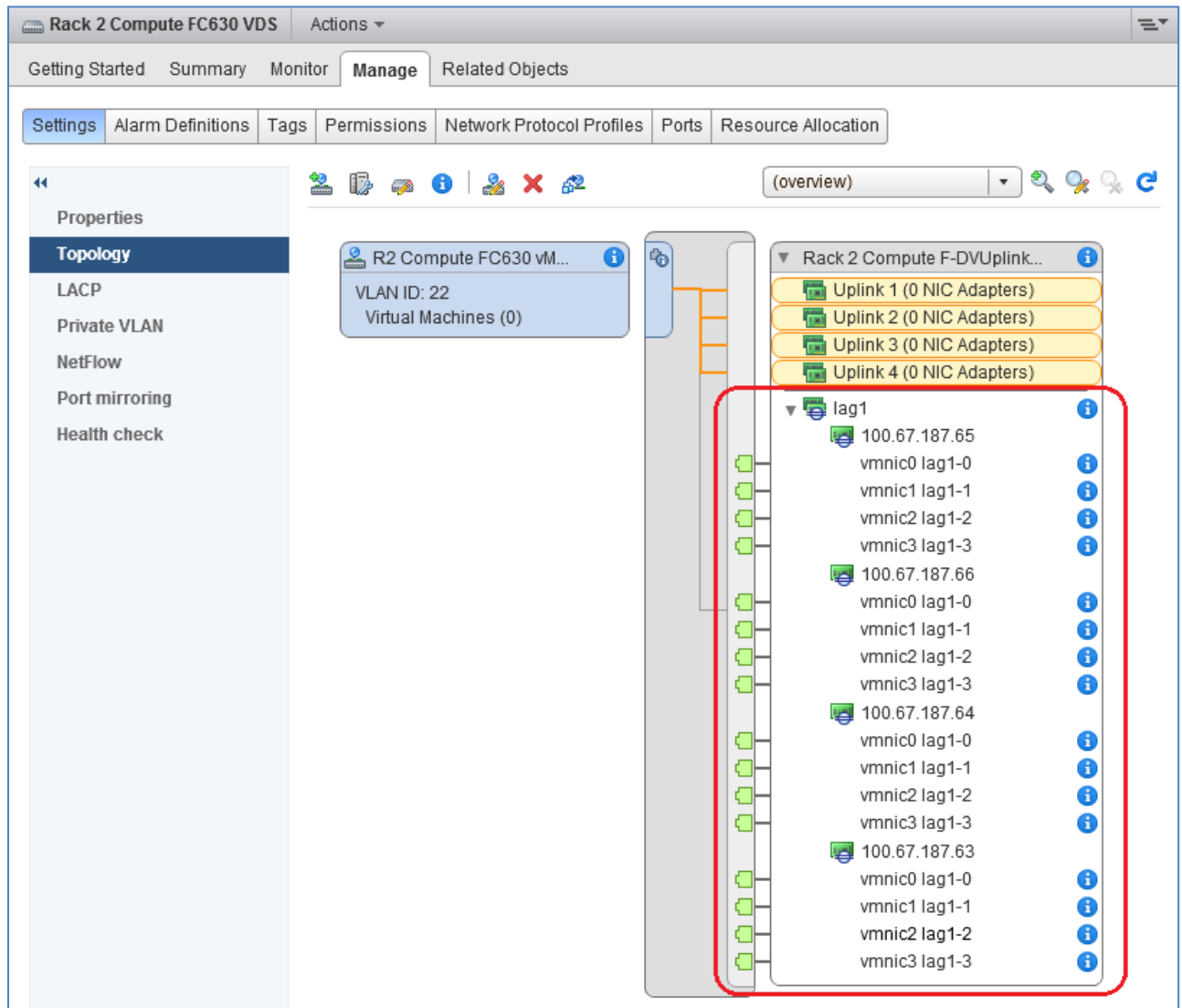


Figure 47 LAGs configured on Rack 2 Compute FC630 VDS

This configuration brings up the LAGs on the upstream switches. This can be confirmed by running the `show vlt detail` command on the upstream switches as shown in the examples from Leaf-1 (Management Cluster) and FN410S-A1 (Compute Cluster) below. The Local and Peer Status columns now indicate all LAGs are UP.

Leaf-1#**show vlt detail**

Local LAG Id	Peer LAG Id	Local Status	Peer Status	Active VLANs
-----	-----	-----	-----	-----
2	2	UP	UP	1, 22, 55
4	4	UP	UP	1, 22, 55
6	6	UP	UP	1, 22, 55

FN410S-A1#**show vlt detail**

Local LAG Id	Peer LAG Id	Local Status	Peer Status	Active VLANs
-----	-----	-----	-----	-----
1	1	UP	UP	1, 22, 55
2	2	UP	UP	1, 22, 55
3	3	UP	UP	1, 22, 55
4	4	UP	UP	1, 22, 55
128	128	UP	UP	1, 22, 55

## 9.5 Configure teaming and failover on LAGs

1. On the web client **Home** screen, select **Networking**.
2. Right click on **Rack 1 Management VDS**. Select **Distributed Port Group > Manage Distributed Port Groups**.
3. Select only the **Teaming and failover** checkbox. Click **Next**.
4. Click **Select distributed port groups**. Check the top box to select all port groups (is only vMotion in this case). Click **OK > Next**.
5. On the **Teaming and failover** page, click **lag1** and move it up to the **Active uplinks** section by clicking the up arrow. Move **Uplinks 1-4** down to the **Unused uplinks** section. Leave other settings at their defaults. The **Teaming and failover** page should look similar to Figure 48 when complete.

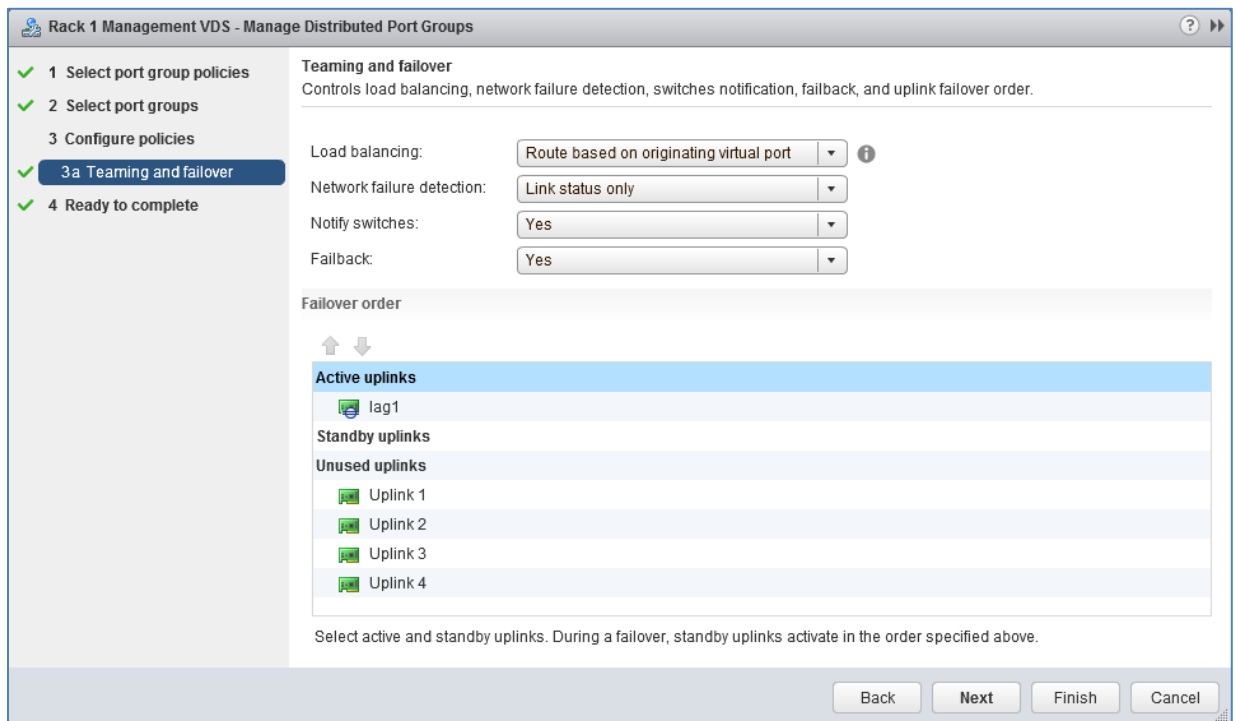


Figure 48 Teaming and failover settings

6. Click **Next** followed by **Finish** to apply the settings.

Repeat steps 1-6 above for the remaining two distributed switches, Rack 2 Compute FC630 VDS and Rack 3 Edge VDS.

## 9.6 Add VMkernel adapters for vMotion

In this section, a vMotion VMkernel adapter (also referred to as a VMkernel port) is added to each ESXi host to allow for vMotion traffic.




IP addresses can be statically assigned to VMkernel adapters upon creation, or DHCP may be used. Static IP addresses are used in this guide.

This deployment uses the following addressing scheme for the vMotion network, where "x" represents the rack number:

Table 4 VLAN and network examples

VLAN ID	Network	Used For
22	10.22.x.0/24	vMotion

To add a VMkernel adapter to each host connected to the Rack 1 Management VDS:

1. On the web client **Home** screen, select **Networking**.
2. Right click **Rack 1 Management VDS**, and select **Add and Manage Hosts**.
3. In the **Add and Manage Hosts** dialog box:
  - a. On the **Select task** page, make sure **Manage host networking** is selected. Click **Next**.
  - a. On the **Select hosts** page, click  **Attached hosts**. Select all hosts. Click **OK > Next**.
  - b. On the **Select network adapter tasks** page, make sure the **Manage VMkernel adapters** box is checked and all other boxes are unchecked. Click **Next**.
  - c. The **Manage VMkernel network adapters** page opens.
    - vMotion adapter
    - i. To add the vMotion adapter, select the first host and click  **New Adapter**.
    - ii. On the **Select target device** page, click the radio button next to **Select an existing network** and click **Browse**.
    - iii. Select the port group created for vMotion > **OK**. Click **Next**.
    - iv. On the **Port properties** page, leave **IPv4** selected and check only the **vMotion traffic** box. Click **Next**.
    - v. On the **IPv4 settings** page, if DHCP is not used, select **Use static IPv4 settings**. Set the IP address, for example 10.22.1.17, and subnet mask for the host on the vMotion network. Click **Next > Finish**.
  - d. Repeat steps i-v for the remaining hosts, then click **Next**.
  - e. On the **Analyze impact** page, **Overall impact status** should indicate  **No impact**.
  - f. Click **Next > Finish**.

Repeat the steps above for the remaining two distributed switches, Rack 2 Compute FC630 VDS and Rack 3 Edge VDS.

When complete, the VMkernel adapters page for each ESXi host in the vSphere data center should look similar to Figure 49. This page is visible by going to **Hosts and Clusters**, selecting a host in the **Navigator** pane, then selecting **Manage > Networking > VMkernel adapters** in the center pane.

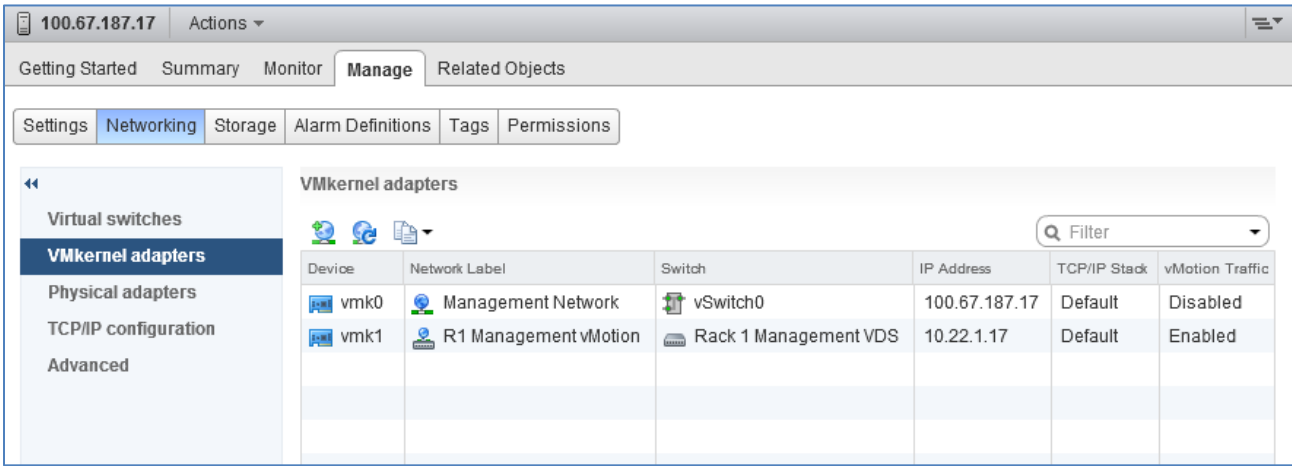


Figure 49 Host VMkernel adapters page

Adapter vmk0 was installed by default for host management. Adapter vmk1 was created in this section.

To verify the configuration, ensure the vMotion adapter, **vmk1** in this example, is shown as **Enabled** in the **vMotion Traffic** column and the VMkernel adapter IP addresses are correct on each host.



## 9.7 Verify VDS configuration

To verify the distributed switches have been configured correctly, the **Topology** page for each VDS provides a summary.

To view the **Topology** page for the **Rack 1 Management VDS**:

1. On the web client **Home** screen, select **Networking**.
2. In the Navigator pane, select **Rack 1 Management VDS**.
3. In the center pane, select **Manage > Settings > Topology** and click the ► icon next to **VMkernel Ports** to expand. The screen should look similar to Figure 50.

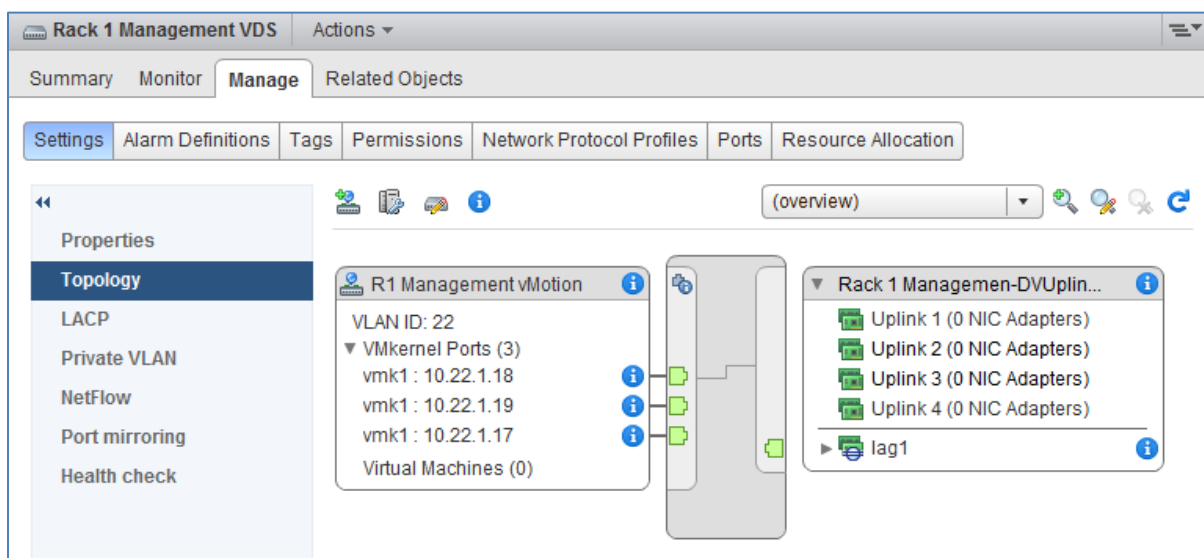


Figure 50 Rack 1 Management VDS VMkernel ports, VLANs, and IP addresses

Notice the distributed port group, **R1 Management vMotion** is shown in Figure 50 with its configured VLAN ID and VMkernel ports. Since VMkernel ports were configured for all three ESXi hosts in the Management cluster, there are three VMkernel ports in the distributed port group.

Repeat steps 1-3 above for the remaining two distributed switches, **Rack 2 Compute FC630 VDS** and **Rack 3 Edge VDS**, to verify they are properly configured.

## 9.8 Enable LLDP

Enabling Link Layer Discovery Protocol (LLDP) on vSphere distributed switches is optional but can be helpful for link identification and troubleshooting.

**Note:** LLDP works as described in this section with QLogic 57810 or QLogic 57840 adapters specified in Appendix A. LLDP functionality may vary with other adapters. LLDP must also be configured on the physical switches per the switch configuration instructions provided earlier in this guide.

### 9.8.1 Enable LLDP on each VDS and view information sent

Enabling LLDP on vSphere distributed switches enables them to send information such as vmnic numbers and MAC addresses to the physical switch connected to the ESXi host.

To enable LLDP on each VDS:

1. On the web client **Home** screen, select **Networking**.
2. Right click on a VDS, and select **Settings > Edit Settings**.
3. In the left pane of the **Edit Settings** page, click **Advanced**.
4. Under **Discovery protocol**, set **Type** to **Link Layer Discovery Protocol** and **Operation** to **Both**.
5. Click **OK**.

Repeat for remaining distributed switches.

To view LLDP information sent from the ESXi host adapters, run the following command from the CLI of a directly connected switch:

```
Leaf-1#show lldp neighbors
```

Loc PortID	Rem Host Name	Rem Port Id	Rem Chassis Id
Te 1/2	-	00:0a:f7:38:88:12	00:0a:f7:38:88:12
Te 1/2	r630-1	00:50:56:18:88:12	vmnic1
Te 1/4	-	00:0a:f7:38:96:62	00:0a:f7:38:96:62
Te 1/4	r630-2	00:50:56:18:96:62	vmnic1
Te 1/6	-	00:0a:f7:38:94:32	00:0a:f7:38:94:32
Te 1/6	r630-3	00:50:56:18:94:32	vmnic1
Fo 1/49	Spine-1	fortyGigE 1/1/1	4c:76:25:e7:41:40
Fo 1/50	Spine-2	fortyGigE 1/1/1	4c:76:25:e7:3b:40
Fo 1/53	Leaf-2	fortyGigE 1/53	f4:8e:38:20:54:29
Fo 1/54	Leaf-2	fortyGigE 1/54	f4:8e:38:20:54:29

The output above shows Leaf 1 is connected to vmnic1 of each host via interfaces Te 1/2, Te 1/4, and Te 1/6.

## 9.8.2 View LLDP information received from physical switch

LLDP configuration is part of the physical switch configurations covered in Section 6. The switches are configured to send information (host name, port number, etc.) via LLDP to the ESXi host network adapters.

To view LLDP information sent from the physical switch:

1. On the web client **Home** screen, select **Hosts and Clusters**.
2. In the **Navigator** pane, select a host.
3. In the center pane, select **Manage > Networking > Physical adapters**.
4. Select a connected physical adapter, **vmnic1** for example.
5. Below the adapter list, select the **LLDP** tab. Information similar to that shown in Figure 51 is provided by the switch.

The screenshot displays the VMware vSphere Web Client interface. The left-hand 'Navigator' pane shows the hierarchy: Virtual switches, VMkernel adapters, Physical adapters (selected), TCP/IP configuration, and Advanced. The main content area is titled 'Physical adapters' and contains a table with columns: Device, Actual Speed, Configured Speed, Switch, and MAC Address. The table lists four Broadcom Corporation QLogic 57840 10 Gigabit Ethernet Adapters (vmnic0, vmnic1, vmnic2, vmnic3). vmnic1 is selected and highlighted in blue. Below the table, the 'Physical network adapter: vmnic1' section is expanded, showing tabs for All, Properties, CDP, and LLDP. The LLDP tab is active, displaying the following information:

Link Layer Discovery Protocol	
Chassis ID	f4:8e:38:20:37:29
Port ID	TenGigabitEthernet 1/6
Time to live	120
TimeOut	60
Samples	318
Management Address	100.67.187.35
Port Description	TenGigabitEthernet 1/6
System Description	Dell Real Time Operating System Software. Dell Operating System Version: 2.0. Dell Application Software Version: 9.10(0.1P18) Copyright (c) 1999-2016 Dell Inc. All Rights Reserved. Build Time: Tue Oct 18 06:19:34 2016
System Name	Leaf-1

Figure 51 Information sent from physical switch to vmnic via LLDP

## 10 Configure iSCSI Storage

This section details configuring the Rack 2 Compute FC630 cluster with iSCSI SAN shared storage.

**Note:** These iSCSI instructions can be extended to the Rack 1 Management and Rack 3 Edge clusters or, for VSAN configuration instructions on the management and edge clusters, see [Dell EMC NSX Reference Architecture - FC430 Compute Nodes with VSAN Storage](#) or [Dell EMC NSX Reference Architecture - R730xd Compute Nodes with VSAN Storage](#) available on Dell TechCenter.

### 10.1 Enable iSCSI offloading on QLogic 57810 adapters

**Note:** For FC630 servers, the QLogic 57810 adapters used for iSCSI traffic are installed in PCIe slots in the back of the FX2s chassis. Do not modify the settings on the QLogic 57840 NDCs (located inside each FC630 server blade) because these are not used for iSCSI traffic in this deployment example.

For each FC630 in the compute cluster:

1. Connect to the server's iDRAC in a web browser and launch the virtual console.
2. In the virtual console, from the **Next Boot** menu, select **BIOS Setup**.
3. Reboot the server.
4. From the **System Setup Main Menu**, select **Device Settings**.
5. On the **Device Settings** page, click on the first port of the QLogic 57810 CNA to be used for iSCSI connectivity. This opens the **Main Configuration Page** for the port.
  - a. Select **Device Level Configuration** and change **Virtualization Mode** to **NPar**. Click **Back**.
  - b. Select **NIC Partitioning Configuration**.
    - i. Select **Partition 1 Configuration**. Set **NIC Mode** and **iSCSI Offload Mode** to **Enabled**. Leave **FCoE Mode** set to **Disabled**. Click **Back**.
    - ii. Select **Partition 2 Configuration** and set all three modes (NIC, iSCSI, FCoE) to **Disabled**. Click **Back**. Repeat to disable all modes on partitions 3 and 4.
  - c. When all partitions have been configured, click **Back > Finish**. Answer **Yes** when prompted to save changes followed by **OK** to return to the **Device Settings** page.
6. Click on the second port of the QLogic 57810 CNA to be used for iSCSI connectivity, and repeat step 5 above to configure partitioning.
7. Click **Finish > Finish** and answer the confirmation prompts as needed to save all changes and reboot the system. The system reboots to ESXi.

## 10.2 Storage Center SC4020 port configuration

This section covers initial configuration of the Dell Storage SC4020 array for the environment used in this guide as shown in Figure 52 and Table 5.

**Note:** For information on initial setup of the Dell Storage SC4020 array, refer to your Dell Storage SC Series documentation. The SC4020 used in this deployment is in virtual port mode.

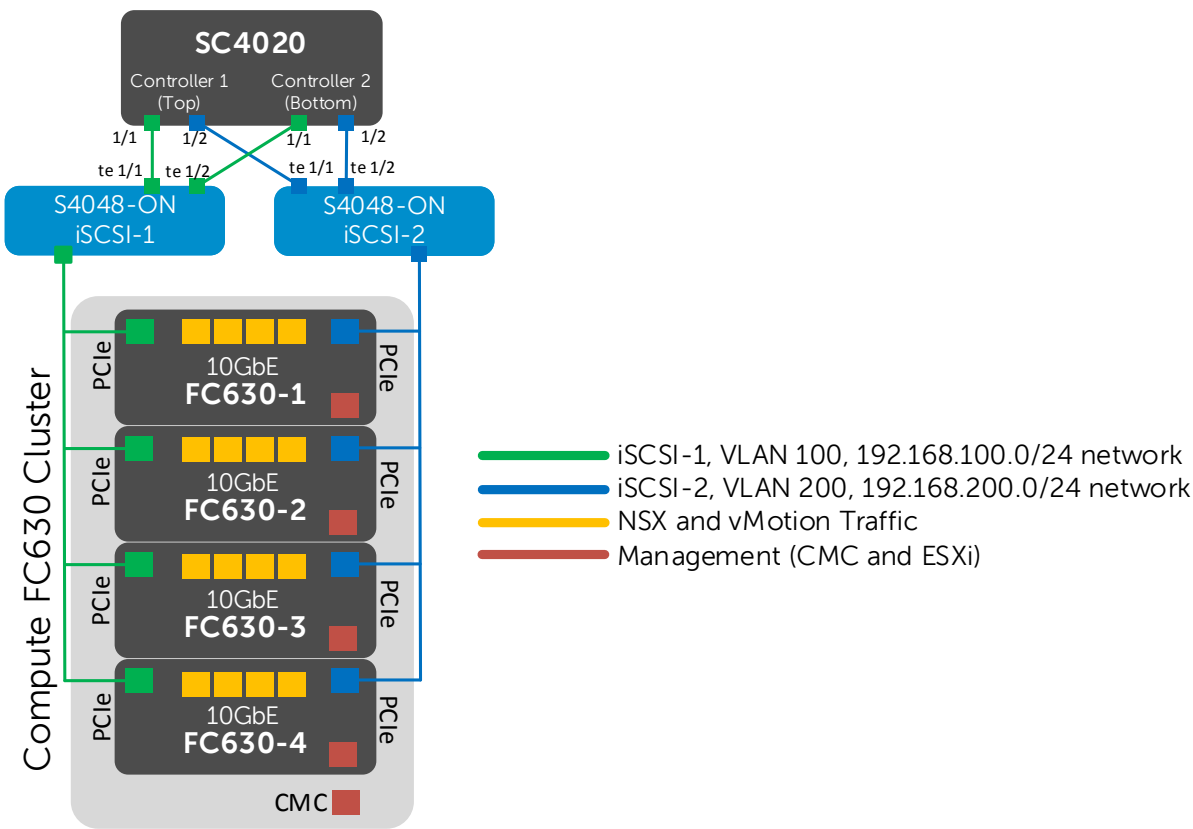


Figure 52 Compute cluster iSCSI SAN

Table 5 Switch to storage connections

Switch	Switch port #	VLAN	SC4020 controller	Cont. port #	Controller port IP address	SC4020 fault domain	Fault domain IP address
S4048-iSCSI-1	te 1/1	100	1 (Top)	1/1	192.168.100.11	iSCSI-1	192.168.100.10
S4048-iSCSI-1	te 1/2	100	2 (Bottom)	1/1	192.168.100.22	iSCSI-1	192.168.100.10
S4048-iSCSI-2	te 1/1	200	1 (Top)	1/2	192.168.200.11	iSCSI-2	192.168.100.20
S4048-iSCSI-2	te 1/2	200	2 (Bottom)	1/2	192.168.200.22	iSCSI-2	192.168.100.20

## 10.2.1 Configure iSCSI port IP addresses

1. Launch the SC4020 Storage Center GUI in a browser.
2. In the left pane, expand the items under **Controllers** and locate the **iSCSI** object under either controller, as shown in Figure 53.

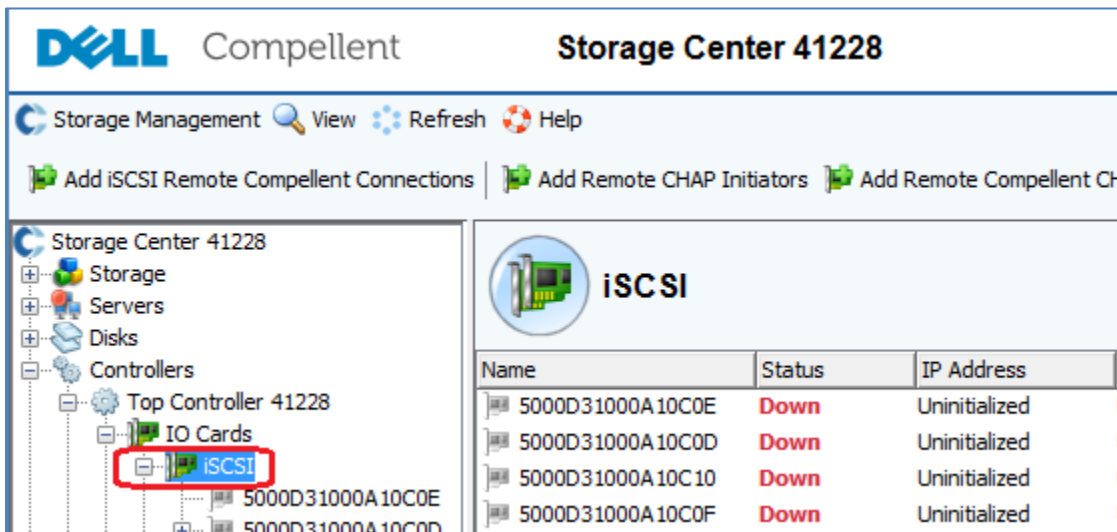


Figure 53 iSCSI object selected

3. Right click on the **iSCSI** object and select **Configure iSCSI IO Cards**.
4. This opens a window where iSCSI port IP addresses can be configured on both controllers. Configure the IP addresses per Table 5, as shown in Figure 54.

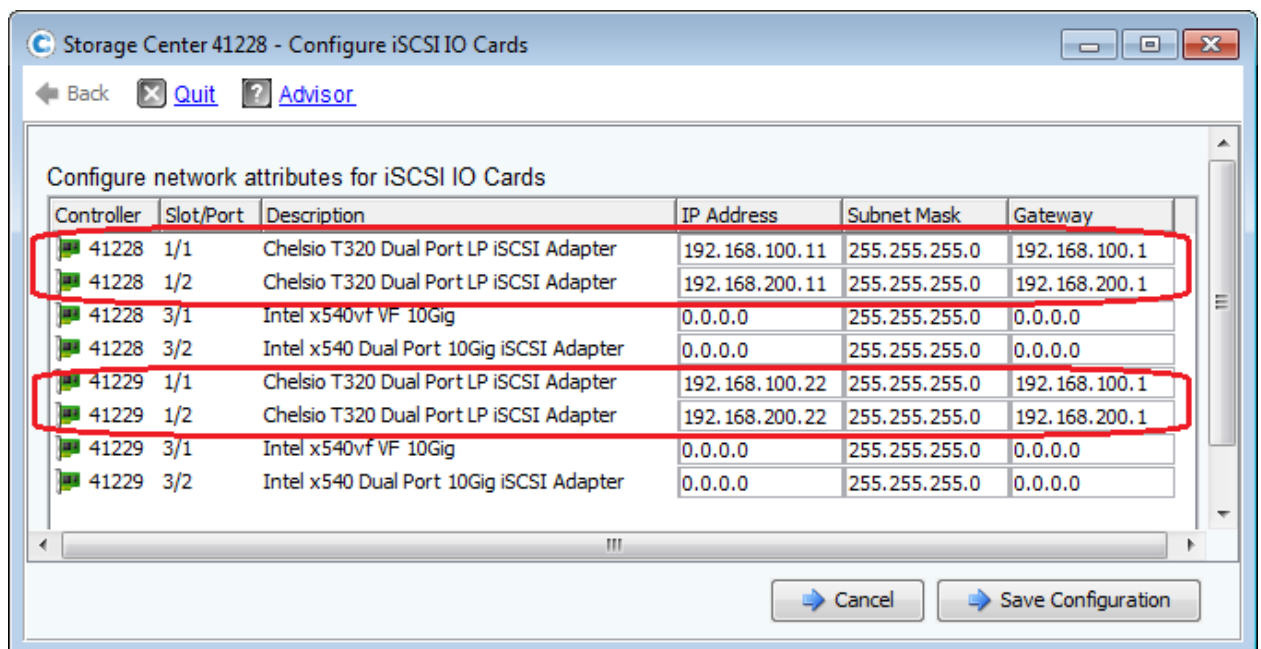


Figure 54 iSCSI port IP addresses configured

**Note:** Though not used in this deployment example, gateway addresses must be specified. The switch iSCSI VLAN interface addresses configured in section 6.5 (192.168.100.1 and 192.168.200.1) are provided as the gateway addresses.

5. Click **Save Configuration > Yes** (if warned about remaining unconfigured ports).

## 10.2.2 Create fault domains

Fault domains group iSCSI ports that are connected to the same network. Ports that belong to the same fault domain can fail over to each other because they have the same connectivity.

Create two fault domains as follows:

1. In the upper left corner of the GUI, select **Storage Management > System > Setup > Configure Local Ports**.
2. This opens the **Configure Local Ports** window. Make sure the **iSCSI** button is selected as shown in Figure 55.

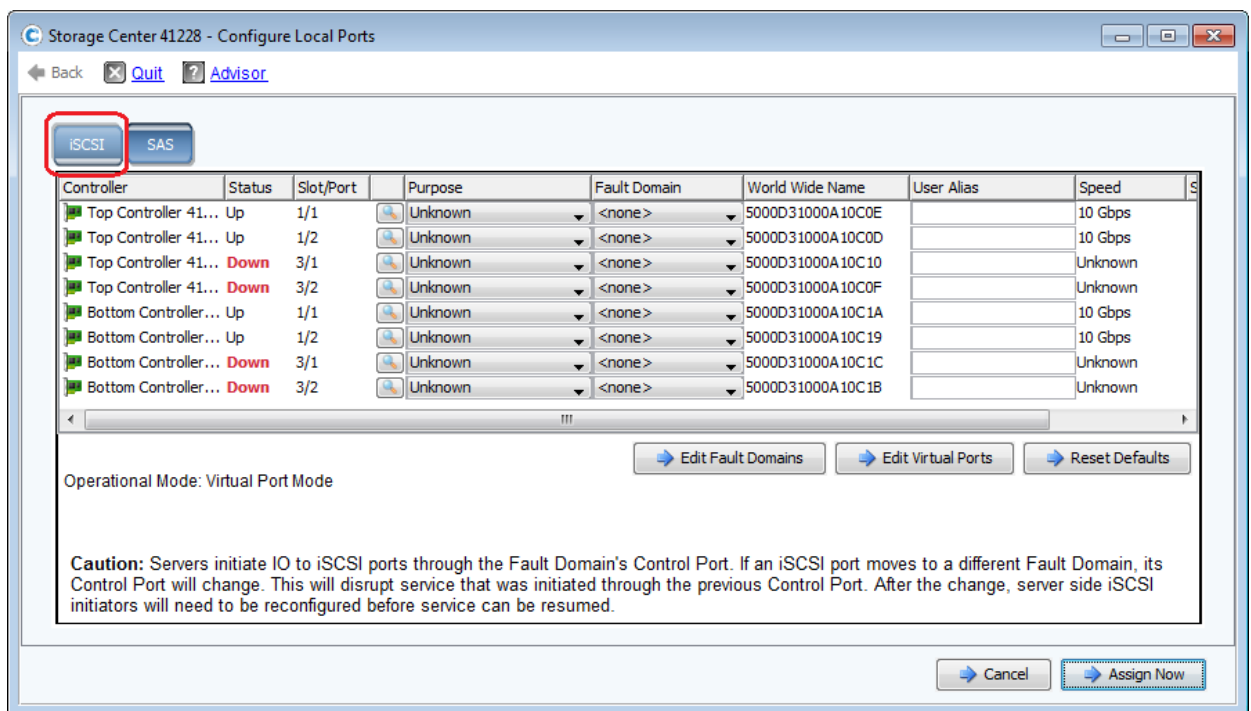


Figure 55 Configure Local Ports window

3. Select **Edit Fault Domains > Create Fault Domain**. Name the first domain iSCSI-1 and click **Continue**.
4. Enter the iSCSI-1 Fault Domain **IP Address** (192.168.100.10), **Net Mask** (255.255.255.0), and **Gateway** (192.168.100.1) as shown in Figure 56. Leave the **Port Number** at 3260.

Storage Center 41228 - Create Fault Domain

Back Quit Advisor

IP Address: 192.168.100.10

Net Mask: 255.255.255.0

Gateway: 192.168.100.1

Port Number: 3260

Continue

Figure 56 iSCSI-1 fault domain IP addressing and port number

5. Click **Continue > Create Now** to create the fault domain named iSCSI-1.
6. To create the second domain, select **Create Fault Domain >** name the second domain **iSCSI-2** and click **Continue**.
7. Enter the iSCSI-2 Fault Domain **IP Address** (192.168.200.10), **Net Mask** (255.255.255.0), and **Gateway** (192.168.200.1). Leave the **Port Number** at 3260.
8. Click **Continue > Create Now** to create the fault domain named iSCSI-2. The window should appear similar to Figure 57.

Storage Center 41228 - Edit Fault Domains - iSCSI

Back Quit Advisor

Name	Type	Used By Ports
iSCSI-1	iSCSI	No
iSCSI-2	iSCSI	No

Delete Domain Edit Fault Domain Create Fault Domain Return

Figure 57 Fault domains created

9. Click **Return** to return to the **Configure Local Ports** window.
10. In the **Purpose** column, make all four configured iSCSI ports **Front End** ports and assign the ports to the newly created fault domains as follows:
  - a. Assign the two storage ports connected to S4048-iSCSI-1 to fault domain **iSCSI-1** (port 1/1 on each controller in this example).



- b. Assign the two storage ports connected to S4048-iSCSI-2 to fault domain **iSCSI-2** (port 1/2 on each controller in this example).

The **Configure Local Ports** table should now appear similar to Figure 58.

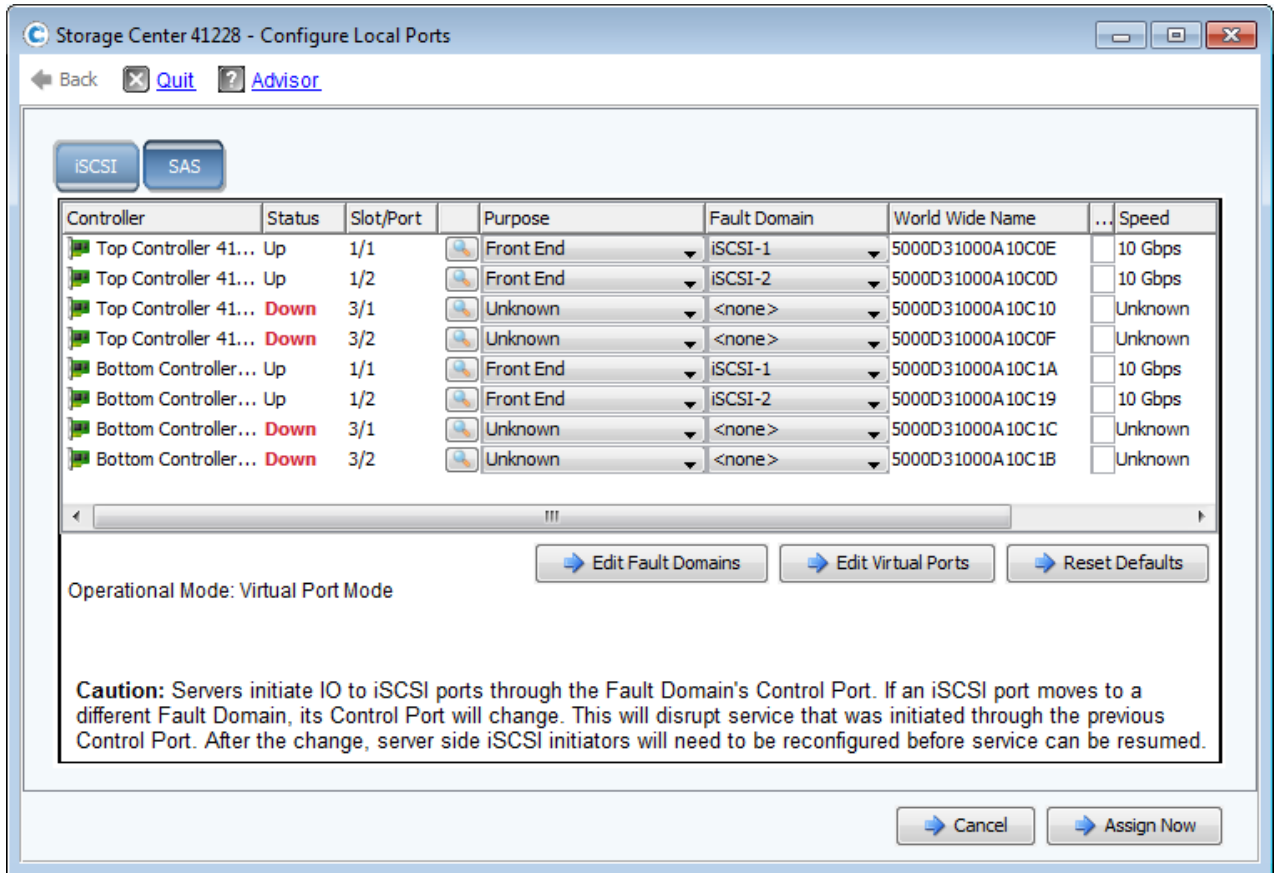


Figure 58 Port configuration and fault domain assignment

11. Click **Assign Now** to finish.

### 10.2.3 Verify the Storage Center configuration

In the left pane of the Storage Center GUI, expand **Controllers**. Under each controller, expand **IO Cards** then expand **iSCSI**. Expand each of the ports under iSCSI.

When properly configured, the tree should appear similar to Figure 59, with key items circled in red. Each controller will have two active iSCSI physical ports with a virtual iSCSI port under each. The two fault domain control ports will also appear under one of the two controllers.

**Storage Center 41228**

Storage Management
View
Refresh
Help

Rebalance Local Ports

**Storage Center 41228**

- Storage
- Servers
- Disks
- Controllers**
  - Top Controller 41228
    - IO Cards
      - ISCSI
        - 5000D31000A10C0E
        - 5000D31000A10C1F
        - 5000D31000A10C0D
        - 5000D31000A10C20
        - 5000D31000A10C10
        - 5000D31000A10C0F
        - ISCSI-1 Control Port
        - ISCSI-2 Control Port
      - SAS
      - Temps
      - Cache Card
    - Bottom Controller 41229
      - IO Cards
        - ISCSI
          - 5000D31000A10C1A
          - 5000D31000A10C21
          - 5000D31000A10C19
          - 5000D31000A10C22
          - 5000D31000A10C1C
          - 5000D31000A10C1B
        - SAS
        - Temps
        - Cache Card
    - UPS
    - Endlosures
    - Racks
    - Remote Systems
    - Users

**Controllers**

Name	Status	Local Port Condition	Leader
Top Controller 41228	Up	Balanced	True
Bottom Controller 41229	Up	Balanced	False

Figure 59 Storage Center iSCSI ports after configuration

#### 10.2.4 Tag the fault domain control ports in the correct VLANs

1. In the left pane of the Storage Center GUI, right click on **iSCSI-1 Control Port** and select **Properties**.
2. In the **Control Port Properties** window, check the **Enable VLAN Tagging** box and set the **VLAN ID** to **100** as shown in Figure 60.

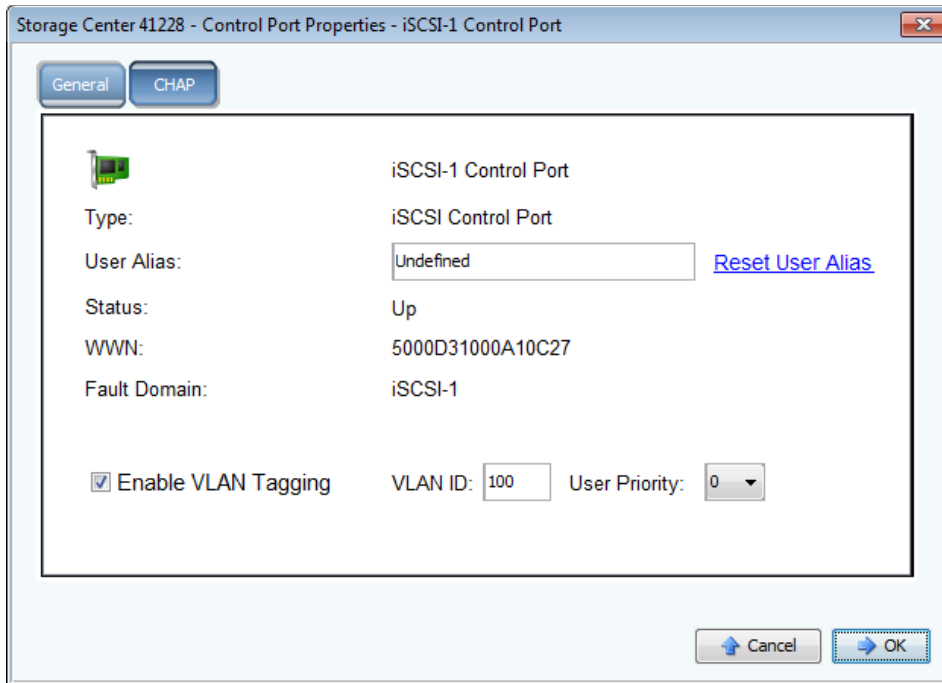


Figure 60 Control port properties window

3. Click **OK**.
4. Repeat steps 1-3 for the **iSCSI-2 Control Port**. It is tagged in VLAN **200**.

## 10.3 Verify iSCSI SAN switch configuration

**Note:** Be sure S4048-ON switches iSCSI-1 and iSCSI-2 are configured per section 6.5.

Connect to the console of the S4048-ON switch named iSCSI-1 and ping the IP address of the iSCSI-1 fault domain control port on the SC4020:

```
iSCSI-1#ping 192.168.100.10
```

```
Sending 5, 100-byte ICMP Echos to 192.168.100.10, timeout is 2 seconds:
```

```
!!!!
```

```
Success rate is 100.0 percent (5/5), round-trip min/avg/max = 0/0/0 (ms)
```

Repeat by connecting to the console of the S4048-ON switch named iSCSI-2 and pinging the IP address of the iSCSI-2 fault domain:

```
iSCSI-2#ping 192.168.200.10
```

```
Sending 5, 100-byte ICMP Echos to 192.168.200.10, timeout is 2 seconds:
```

```
!!!!
```

```
Success rate is 100.0 percent (5/5), round-trip min/avg/max = 0/0/0 (ms)
```

## 10.4 Create an iSCSI VDS for the compute cluster

**Note:** Return to the vSphere Web Client starting with this section.

To create a VDS for iSCSI traffic for the Rack 2 Compute FC630 Cluster:

1. On the web client **Home** screen, select **Networking**.
2. Right click on **Datacenter**. Select **Distributed switch > New Distributed Switch**.
3. Provide a name for the VDS, e.g. **Rack 2 iSCSI VDS**. Click **Next**.
4. On the **Select version** page, select **Distributed switch: 6.0.0** and click **Next**.
5. On the **Edit settings** page:
  - a. Change the **Number of uplinks** to **2**.
  - b. Leave **Network I/O Control** set to **Enabled**.
  - c. Uncheck the **Create a default port group** box.
6. Click **Next > Finish**.

The VDS is created with the uplink port group shown beneath it.

When complete, the **Navigator** pane should look similar to Figure 61.

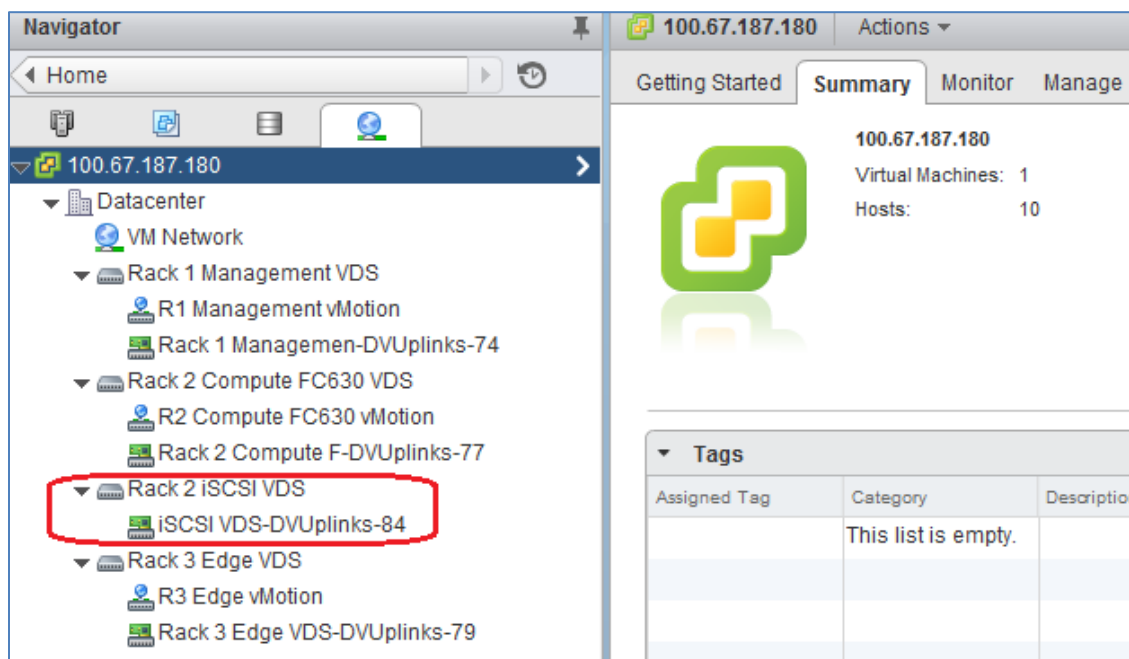


Figure 61 Rack 2 iSCSI VDS created

## 10.5 Set the iSCSI VDS MTU to 9000 and enable LLDP

Dell EMC recommends increasing the Maximum Transmission Unit (MTU) of devices handling storage traffic to 9000 bytes for best performance. LLDP may be configured at the same time (see Section 9.8 for more information on LLDP).

To configure the iSCSI VDS:

1. Go to **Home > Networking**.
2. Right click on **Rack 2 iSCSI VDS**, and select **Settings > Edit Settings**.
3. In the left pane of the **Edit Settings** page, click **Advanced**.
4. Set **MTU (Bytes)** to 9000.
5. Under **Discovery protocol**, set **Type** to **Link Layer Discovery Protocol** and **Operation** to **Both**.

The window should appear similar to Figure 62.

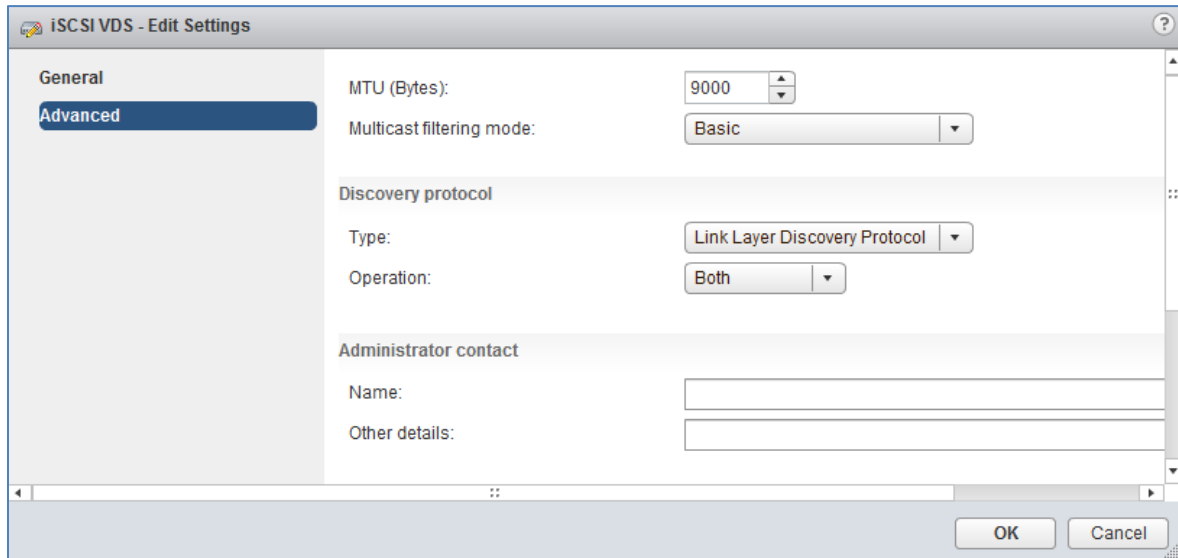


Figure 62 iSCSI VDS – Edit Settings window

6. Click **OK** to apply the settings.

## 10.6 Add distributed port groups

In this section, two distributed port groups for iSCSI traffic are added to the Rack 2 iSCSI VDS created in the previous section.

To create the distributed port groups on the **iSCSI VDS**:

1. On the web client **Home** screen, select **Networking**.
2. Right click on **Rack 2 iSCSI VDS**. Select **Distributed Port Group > New Distributed Port Group**.
3. On the **Select name and location** page, provide a name for the first distributed port group, e.g. **R2 iSCSI-1**. Click **Next**.
4. On the **Configure settings** page, next to **VLAN type**, select **VLAN**. Set the **VLAN ID** to **100**. Leave other values at their defaults as shown in Figure 63.

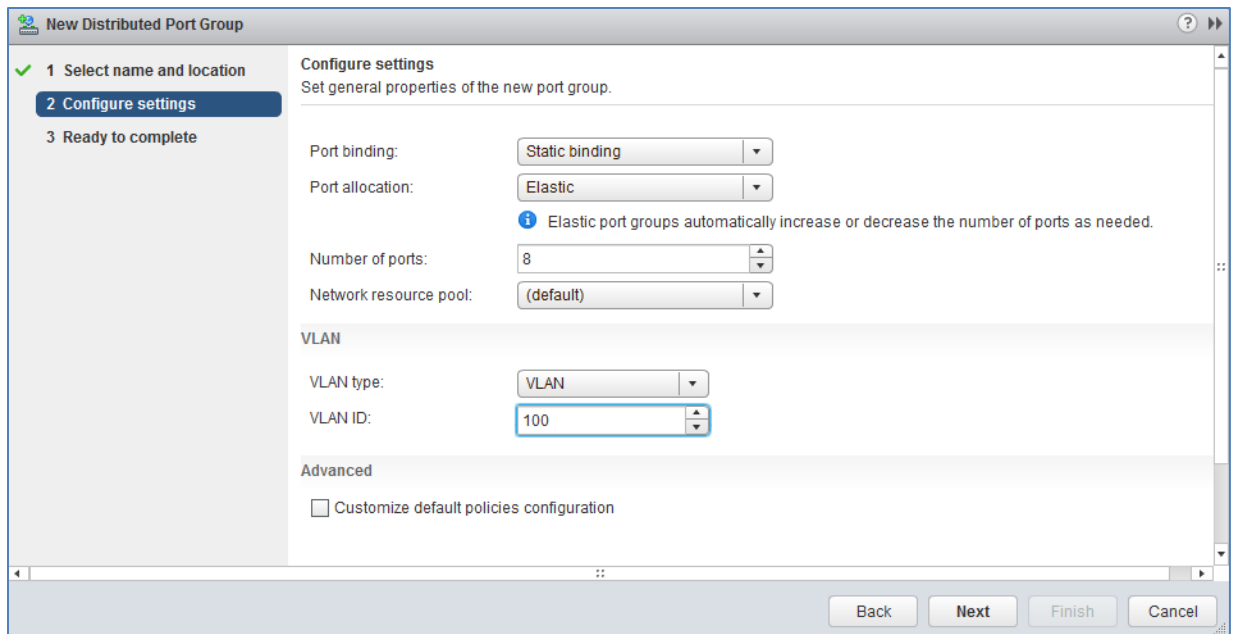


Figure 63 R2 iSCSI-1 distributed port group settings page

5. Click **Next > Finish**.

Repeat steps 2-5 above for the second port group. Provide a unique name, e.g. **R2 iSCSI-1** and set its VLAN ID to **200**.

When complete, the Navigator pane will appear similar to Figure 64.

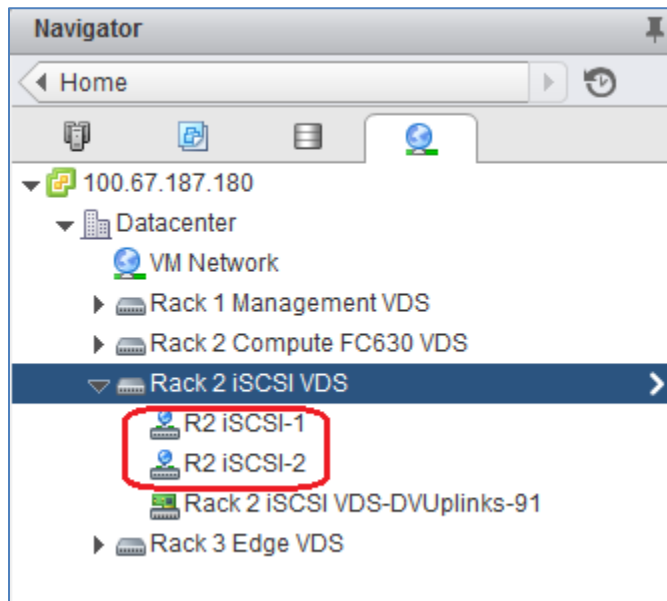


Figure 64 iSCSI distributed port groups created

## 10.7 Configure teaming and failover

1. On the web client **Home** screen, select **Networking**.
2. Right click on **Rack 2 Compute iSCSI VDS**. Select **Distributed Port Group > Manage Distributed Port Groups**.
3. Select only the **Teaming and failover** checkbox. Click **Next**.
4. Click **Select distributed port groups**. Check the box next to R2 iSCSI-1. Click **OK > Next**.
5. On the **Teaming and failover** page, click **Uplink 1** and move it up to the **Active uplinks** section by clicking the up arrow. Move **Uplink 2** down to the **Unused uplinks** section. Leave other settings at their defaults. The **Teaming and failover** page should look similar to Figure 65 when complete.

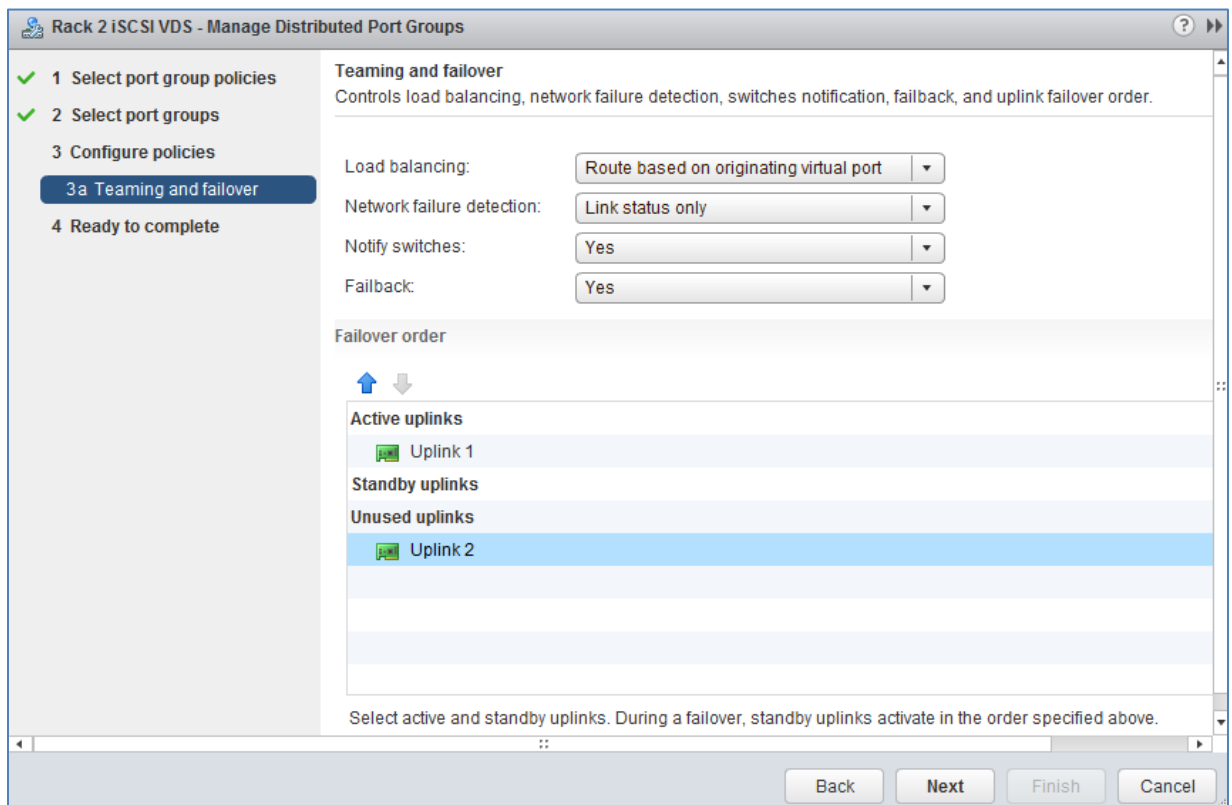


Figure 65 Teaming and failover settings for the R2 iSCSI-1 port group

6. Click **Next > Finish** to apply the settings.





Repeat steps 2-6 above for the **R2 iSCSI-2** port group. For the iSCSI-2 port group, make sure that **Uplink 2** is moved to **Active** and **Uplink 1** is moved to **Unused**.

## 10.8 Associate hosts and assign uplinks

Hosts and their vmnics must be associated with the Rack 2 iSCSI VDS.

**Note:** Before starting this section, be sure you know the vmnic-to-physical adapter mapping for each host. This can be determined by going to **Home > Hosts and Clusters** and selecting the host in the **Navigator** pane. In the center pane select **Manage > Networking > Physical adapters**. Adapter MAC addresses can be determined by connecting to the iDRAC. In this example, vmnics used are numbered vmnic4 and vmnic5. Vmnic numbering will vary depending on adapters installed in the host.

To add hosts to the Rack 2 iSCSI VDS:

1. On the web client **Home** screen, select **Networking**.
2. Right click **Rack 2 iSCSI VDS** and select **Add and Manage Hosts**.
3. In the **Add and Manage Hosts** dialog box:
  - a. On the **Select task** page, make sure **Add hosts** is selected. Click **Next**.
  - b. On the **Select hosts** page, Click the  **New hosts** icon. Select the check box next to each host in the **Rack 2 Compute FC630** cluster. Click **OK > Next**.
  - c. On the **Select network adapters tasks** page, be sure the **Manage physical adapters** box is checked. Be sure all other boxes are unchecked. Click **Next**.
  - d. On the **Manage physical network adapters** page, each host is listed with its vmnics beneath it.
    - i. Select the first vmnic (vmnic4 in this example) on the first host and click  **Assign uplink**.
    - ii. Select **Uplink 1 > OK**.
    - iii. Select the second vmnic (vmnic5 in this example) on the first host and click  **Assign uplink**.
    - iv. Select **Uplink 2 > OK**.
4. Repeat steps i – iv for the remaining hosts. Click **Next** when done.
5. On the **Analyze impact** page, **Overall impact status** should indicate  **No impact**.
6. Click **Next > Finish**.

When complete, the **Manage > Settings > Topology** page for **Rack 2 iSCSI VDS** should look similar to Figure 66.



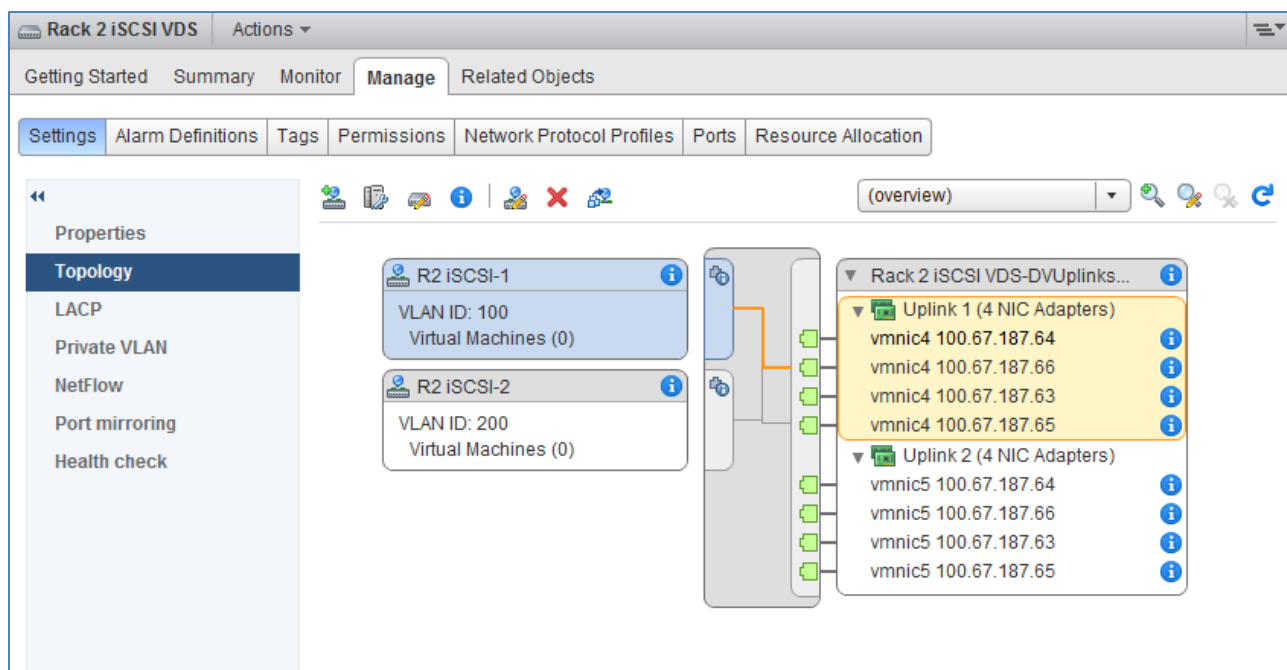


Figure 66 R2 iSCSI VDS vmnic configuration

## 10.9 Add VMkernel adapters for iSCSI

In this section, two iSCSI VMkernel adapters (also referred to as VMkernel ports) are added to each ESXi host to allow for multipath iSCSI traffic.

IP addresses can be statically assigned to VMkernel adapters upon creation, or DHCP may be used. Static IP addresses are used in this guide.

This deployment uses the following addressing scheme for the iSCSI networks:




Table 6 iSCSI VLANs and networks

VLAN ID	Network	Used For
100	192.168.100.0/24	iSCSI-1
200	192.168.200.0/24	iSCSI-2

Table 7 Host iSCSI IP Addresses

Host	iSCSI-1	iSCSI-2
FC630-1	192.168.100.63 /24	192.168.200.63 /24
FC630-2	192.168.100.64 /24	192.168.200.64 /24
FC630-3	192.168.100.65 /24	192.168.200.65 /24
FC630-4	192.168.100.66 /24	192.168.200.66 /24

To add a VMkernel adapter to each host connected to the iSCSI VDS:

1. On the web client **Home** screen, select **Networking**.
2. Right click on **iSCSI VDS**, and select **Add and Manage Hosts**.
3. In the **Add and Manage Hosts** dialog box:
4. On the **Select task** page, make sure **Manage host networking** is selected. Click **Next**.
  - a. On the **Select hosts** page, click  **Attached hosts**. Select all hosts. Click **OK > Next**.
  - b. On the **Select network adapter tasks** page, make sure the **Manage VMkernel adapters** box is checked and all other boxes are unchecked. Click **Next**.
  - c. The **Manage VMkernel network adapters** page opens.
    - i. To add the first iSCSI adapter, select the first host and click  **New Adapter**.
    - ii. On the **Select target device** page, click the radio button next to **Select an existing network** and click **Browse**.
    - iii. Select **R2 iSCSI-1**. Click **OK > Next**.
    - iv. On the **Port properties** page, leave **IPv4** selected and make sure no boxes are checked next to **Enable services**. Click **Next**.
    - v. On the **IPv4 settings** page, if DHCP is not used, select **Use static IPv4 settings**. Set the IP address, for example **192.168.100.63**, and subnet mask, **255.255.255.0**, for the host on the first iSCSI network. Click **Next > Finish**.
5. Repeat steps i-v to add a second iSCSI VMkernel adapter on network **R2 iSCSI-2** with an appropriate IP address, for example **192.168.200.63**. Repeat for the remaining hosts resulting in two VMkernel adapters per host, with one on each network. Click **Next**.
  - a. On the **Analyze impact** page, **Overall impact status** should indicate  **No impact**.
  - b. Click **Next > Finish**.

When complete, the **VMkernel adapters** page for each ESXi host in the vSphere data center should look similar to Figure 67. This page is visible by going to **Hosts and Clusters**, selecting a host in the **Navigator** pane, then selecting **Manage > Networking > VMkernel adapters** in the center pane.

Make sure the adapters are configured correctly on each host.

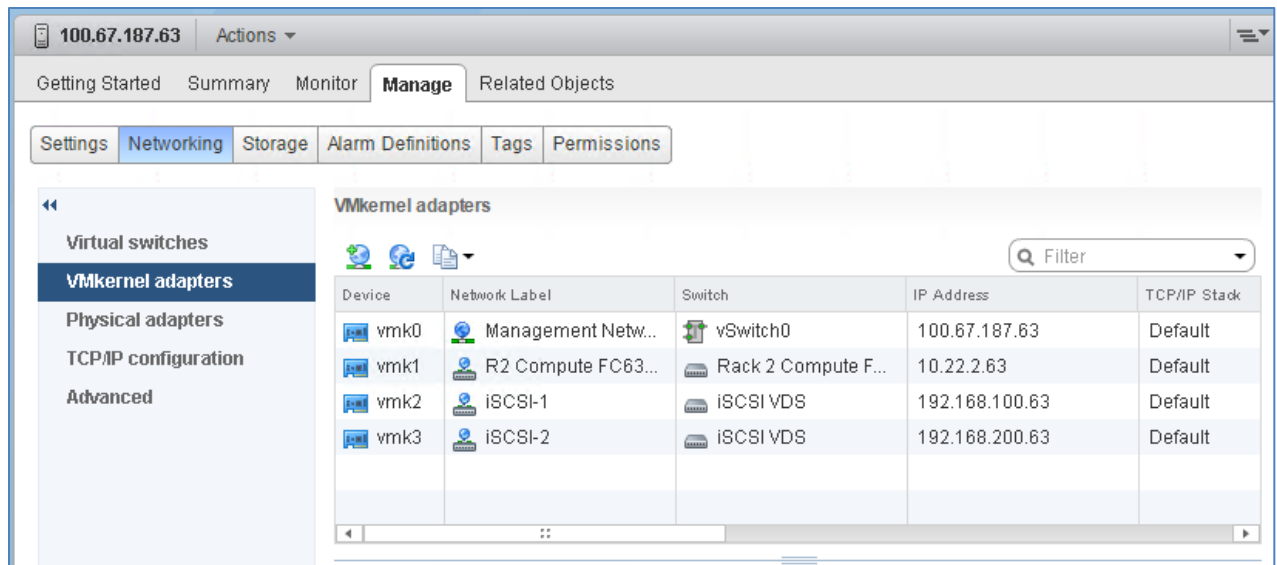


Figure 67 Host VMkernel adapters page with iSCSI networking configured

## 10.10 Increase the MTU to 9000 on iSCSI VMkernel adapters

Dell EMC recommends increasing the Maximum Transmission Unit (MTU) of devices handling storage traffic to 9000 bytes for best performance.


To set the MTU to 9000 bytes on iSCSI VMkernel adapters:

1. Go to **Home > Hosts and Clusters** and select the first host in the compute cluster.
2. In the center pane, select **Manage > Networking > VMkernel adapters**.
3. Select the **iSCSI-1** VMkernel adapter. Click **Edit settings**.
4. Select **NIC settings** and change the MTU to **9000**.
5. Click **OK**.
6. Repeat for the **iSCSI-2** VMkernel adapter.

Repeat steps 1-6 on the remaining hosts in the compute cluster.

## 10.11 Bind iSCSI adapters with VMkernel ports

1. Go to **Home > Hosts and Clusters**.
2. In the **Navigator** pane, select a host.
3. In the center pane, select **Manage > Storage > Storage adapters**.
4. Select a connected physical adapter listed under **QLogic 57XXX 1/10 Gigabit Ethernet Adapter**, e.g. vmhba32.
5. Under **Adapter Details**, click the **Network Port Binding** tab.
6. Click the **+** icon to open the **Bind vmhbaxx with VMkernel Adapter** window.
  - a. Select the **R2 iSCSI-1 (iSCSI VDS) port group**.
  - b. Click **OK**

- Click the  icon to rescan the host's storage adapter.

Repeat steps 4 through 7 for the second vmhba (e.g. vmhba33) on the host. In step 6.a., connect it to the iSCSI-2 port group.

Repeat the steps above for the remaining hosts in the Rack 2 Compute FC630 cluster.

When complete, each host's **Storage Adapters** page should look similar to Figure 68.

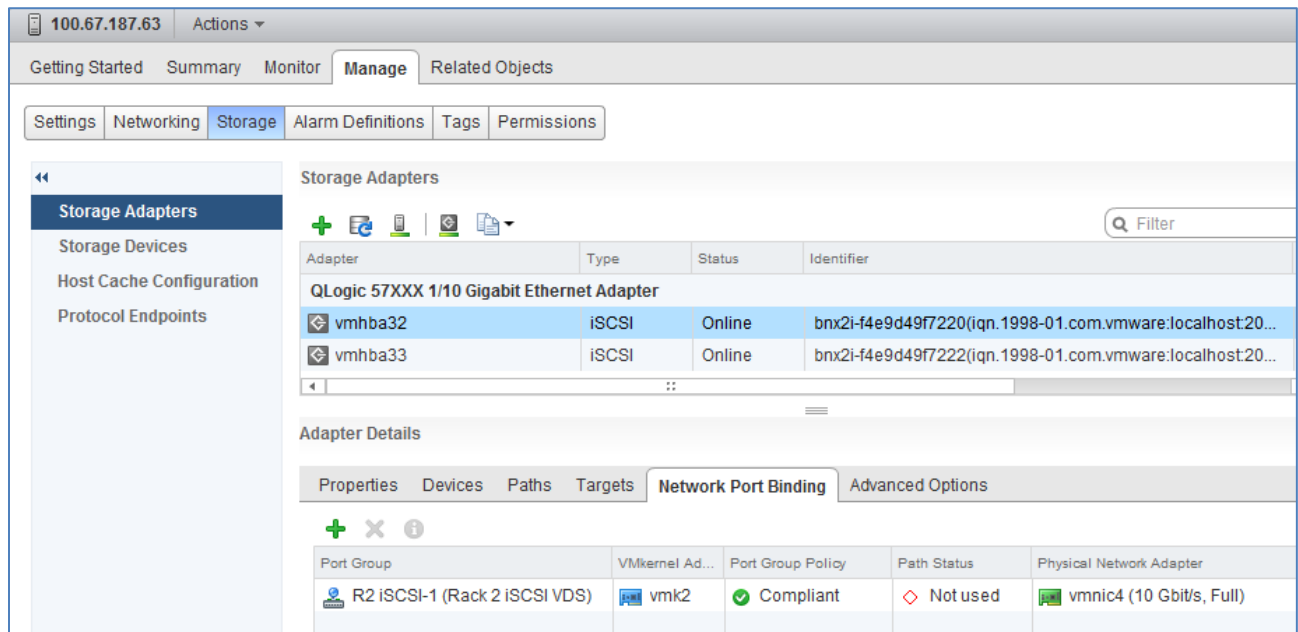


Figure 68 Storage adapters details page


## 10.12 Configure dynamic discovery

With dynamic discovery, each time the initiator contacts the SC4020 storage system, it sends a send targets request to the system. The SC4020 responds by supplying a list of available targets to the initiator.

**Note:** This feature enables the HBAs to automatically appear when creating servers on the SC4020.

To configure dynamic discovery for iSCSI:

- Go to **Home > Hosts and Clusters**.
- In the **Navigator** pane, select the first host in the **Rack 2 Compute FC630** cluster.
- In the center pane, select **Manage > Storage > Storage adapters**.
- Select host's first connected iSCSI adapter listed under **QLogic 57XXX 1/10 Gigabit Ethernet Adapter**, vmhba32 for example.
- Under **Adapter Details**, click the **Targets** tab
- Select **Dynamic Discovery** and click **Add** to open the **Add Send Target Server** window.

7. Next to **iSCSI Server**, enter the IP address for the iSCSI-1 fault domain on the SC4020, **192.168.100.10**. (Configured in section 10.2.2)
8. Leave the other settings at their default values and click **OK**.
9. Click the  icon to rescan the host's storage adapter.

Repeat steps 4-9 for the host's second connected iSCSI adapter. Use the IP address configured for the iSCSI-2 fault domain on the SC4020, **192.168.200.10**.

Repeat the above for the remaining hosts in the compute cluster.

## 10.13 SC4020 final configuration

### 10.13.1 Create servers in Storage Center

1. Connect to the Storage Center GUI in a browser and log in.
2. In the left pane, select **Servers > Create Server**. The configured iSCSI VMkernel ports (initiators) appear with their IP addresses as shown in Figure 69. In this guide, four FC630 servers are used with two iSCSI ports per server.

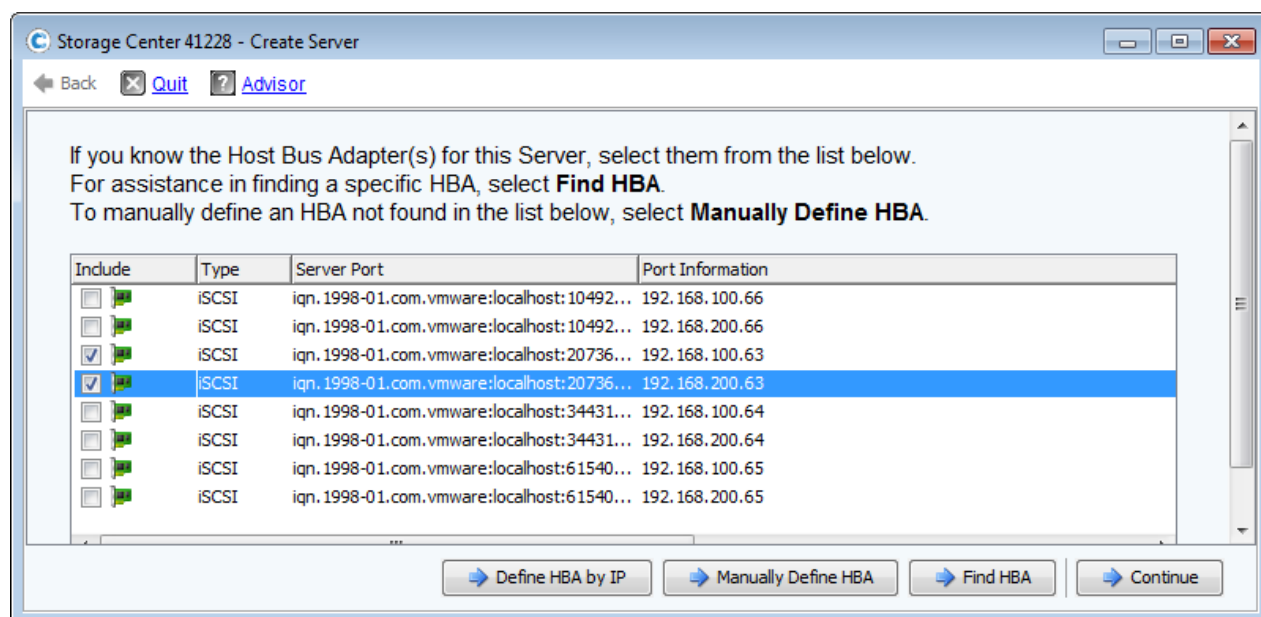


Figure 69 Host VMkernel adapters for iSCSI detected by Storage Center

3. For the first server, check the boxes for both of its iSCSI ports. In the figure above, the two ports for server FC630-1 (IP addresses ending in .63) are selected. Click **Continue**.
4. Give the first server a **Name**, e.g. **FC630-1**. Next to **Operating System**, select **VMware ESXi 6.0**, from the drop-down list as shown in Figure 70.

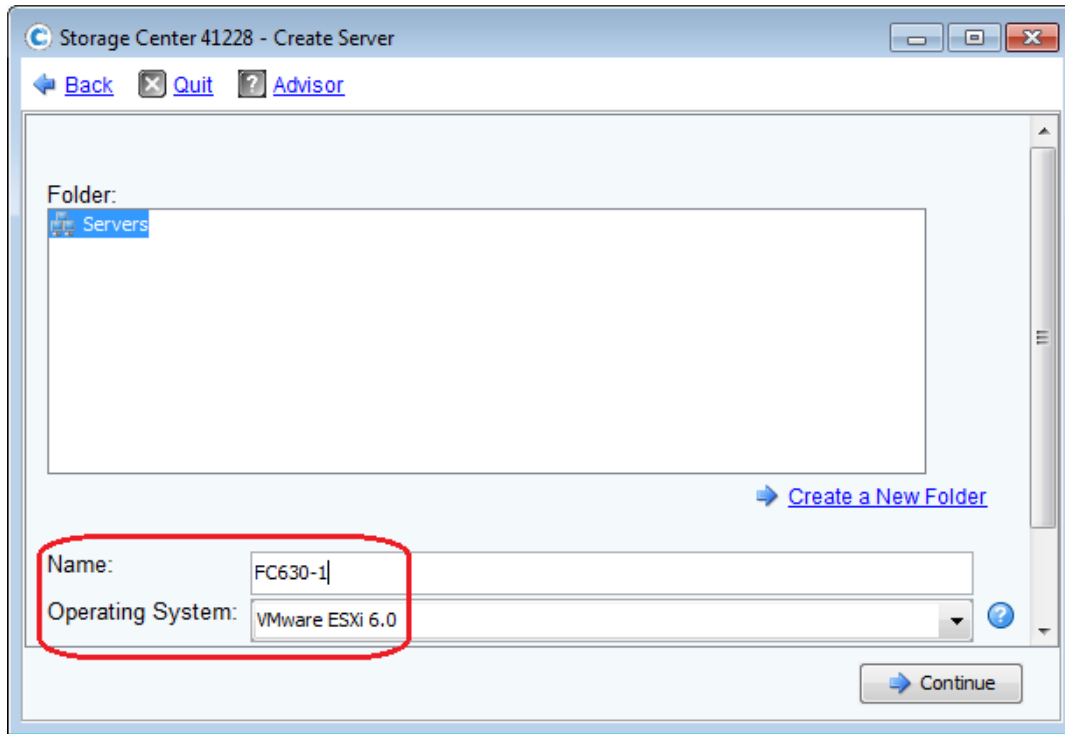


Figure 70 Name entered and operating system selected.

5. Click **Continue > Create Now**. The window indicates the server has been created and provides several options as shown in Figure 71.

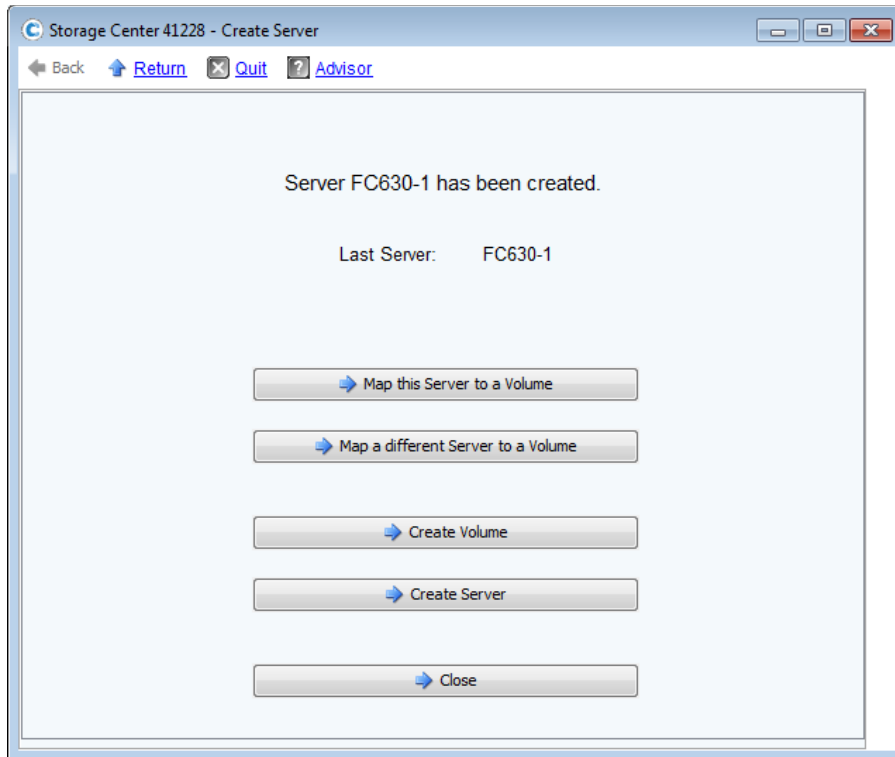


Figure 71 Options after server creation

6. Select **Create Server** to configure the next server in the compute cluster by following steps 3-5 above. Repeat as needed for the remaining servers.

When all servers in the vSphere compute cluster (FC630-1 through 4) are created, click **Close**. The Servers list in Storage Center appears as shown in Figure 72.

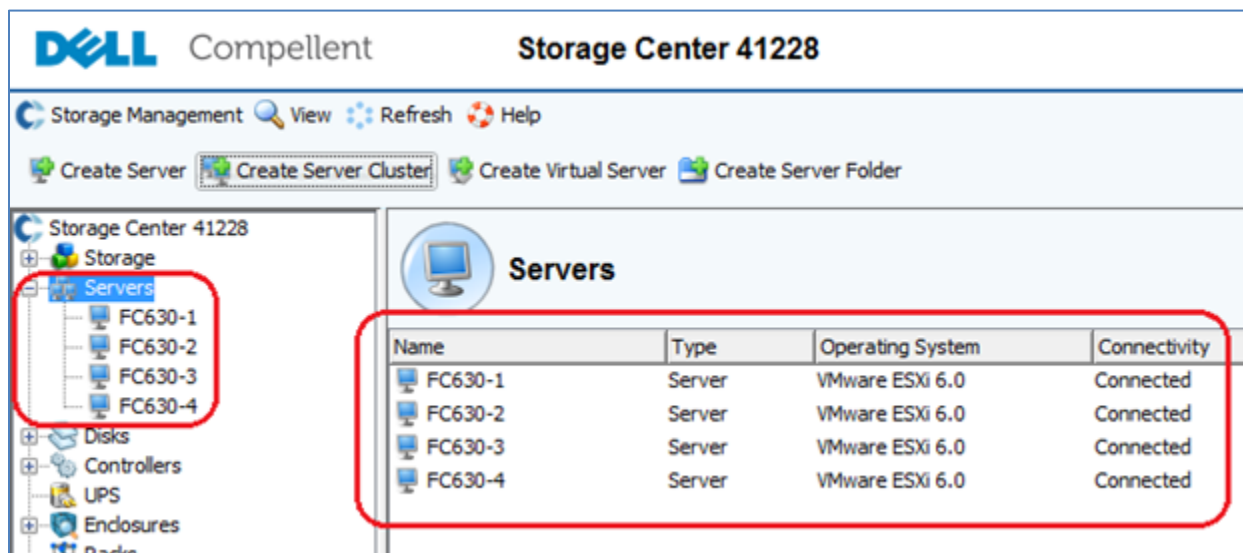


Figure 72 Servers created

### 10.13.2 Create Storage Center server cluster

A Storage Center server cluster enables the mapping of volumes to vSphere server clusters. In this section, the four FC630 servers are added to a cluster.

To create a server cluster:

1. In the left pane of the Storage Center GUI, right click on **Servers** and select **Create Server Cluster**.
2. In the **Create Server Cluster** dialog box, select **Add Existing Server**.
3. Select all servers in the cluster as shown in Figure 73.

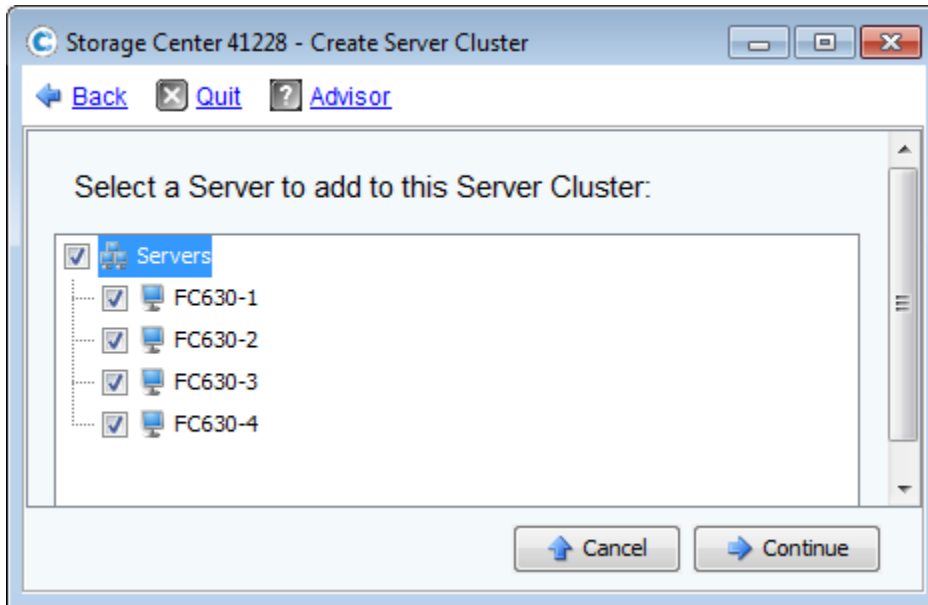


Figure 73 Select servers dialog box

4. Click **Continue > Continue** and enter a **Name** for the Storage Center server cluster in the box provided. The vSphere cluster name, **Rack 2 Compute FC630**, is used here for consistency.
5. Click **Continue > Create Now** to create the cluster. Click **Close** to return to the main screen.



When complete, the server cluster appears as shown in Figure 74.

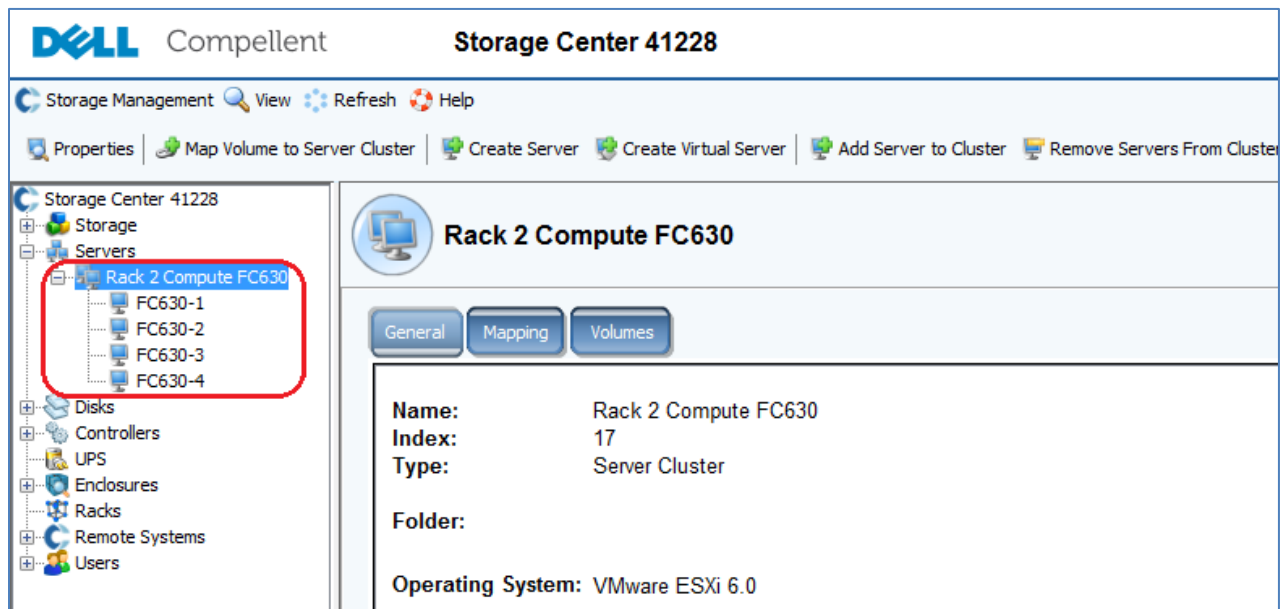


Figure 74 Storage Center cluster created

### 10.13.3 Create and map shared storage volume

In this section, a 2 TB volume is created and mapped to the cluster.

1. In the left pane of the Storage Center GUI, right click on **Storage** and select **Create Volume**.
2. Specify the **Size**. For this example, 2 TB is used. Click **Continue**.
3. Leave the **Replay Profiles** at their default values. Click **Continue**.
4. Provide a name, e.g. **Rack 2 Compute FC630 Vol 1**. Click **Continue > Create Now**.

The confirmation dialog will be similar to Figure 75.

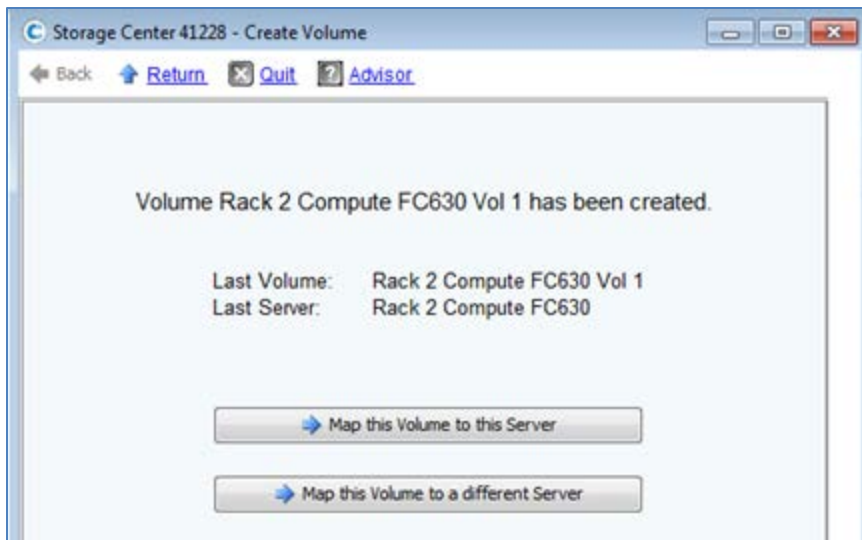



Figure 75 Volume created

If the correct server cluster is shown next to **Last Server**, as in Figure 75, select **Map this Volume to this Server**. Otherwise, select **Map this Volume to a different Server** and select the correct server cluster.

Follow the prompts to complete the mapping to the **Rack 2 Compute FC630** cluster. Leave advanced options at their defaults.

## 10.14 Verify hosts are connected to storage

**Note:** Return to the vSphere Web Client starting with this section.

1. On the vSphere Web Client **Home** screen, select **Hosts and Clusters**.
2. In the **Navigator** pane, select the first host in the **Rack 2 Compute FC630** cluster.
3. In the center pane, select **Manage > Storage > Storage adapters** and select the host's first storage adapter (e.g. vmhba32).
4. Click the  icon to rescan for newly added storage devices.
5. Under **Adapter Details**, select the **Devices** tab. The volume appears as shown in Figure 76.









































Adapter Details						
Properties Devices Paths Targets Network Port Binding Advanced Options						
<div>                                         </div>						
Name	Type	Capacity	Operational State	Hardware Acceleration	Drive Type	
COMPELNT iSCSI Disk (naa.6000d31000a10c00000000000000000d)	disk	2.00 TB	Attached	Supported	HDD	

Figure 76 Devices tab

6. Select the **Paths** tab. The target name, LUN number and status are shown. The status field is marked either **Active** or **Active (I/O)** as shown in Figure 77.

Adapter Details			
Properties	Devices	Paths	Targets
<div>Enable</div> <div>Disable</div>			
Runtime Name	Target	LUN	Status
vmhba32:C0:T0:L1	iqn.2002-03.com.compellent:5000d31000a10c1f:192.168.100.11:3260	1	◆ Active (I/O)

Figure 77 Paths tab


Repeat steps 3-6 above for the host's second storage adapter (e.g. vmhba33).

Repeat the above for the remaining hosts in the cluster. All hosts should have two active connections to the shared storage volume.

## 10.15 Create a datastore

Create a datastore that uses the shared storage volume.

To create the datastore:

1. Go to **Home > Storage**.
2. In the **Navigator** pane, right click on **Datacenter** and select **Storage > New Datastore**.
3. In the **New Datastore** window, for **Location**, **Datacenter** is selected. Click **Next**.
4. Leave the **Type** set to **VMFS** and click **Next**.
5. On the **Name and device selection** page:
  - a. Provide a **Datastore name**, e.g. **Rack 2 Compute FC630 Vol 1**.
  - b. From the dropdown menu, select *any* host in the cluster. When created, this datastore will be accessible to all configured hosts in the cluster (refer to the note on the screen next to the  icon). The screen will look similar to Figure 78.

New Datastore

✓ 1 Location


✓ 2 Type

3 Name and device selection

4 Partition configuration

5 Ready to complete

Datastore name: Rack 2 Compute FC630 Vol 1


 The datastore will be accessible to all the hosts that are configured with access to the selected disk/LUN. If you d  
disk/LUN that you are interested in, it might not be accessible to that host. Try changing the host or configure acc  
that disk/LUN.

Select a host to view its accessible disks/LUNs: 100.67.187.63

Name	LUN	Capacity	Hardware Acceler...	Drive Type
COMPELNT iSCSI Disk (naa.6000d31000a10c00000000...	1	2.00 TB	Supported	HDD

Figure 78 New datastore window with host selected

6. Click on the LUN and click **Next**.
7. Leave the **Partition configuration** at its default settings and click **Next > Finish** to create the datastore.

To verify the datastore has been mounted by all hosts in the compute cluster:

1. Go to **Home > Storage**.
2. In the **Navigator** pane, select the newly created datastore, **Rack 2 Compute FC630 Vol 1**.
3. In the center pane, select **Manage > Settings > Connectivity and Multipathing**.

All hosts in the compute cluster are listed with the datastore status shown as **Mounted** and **Connected** as per Figure 79.

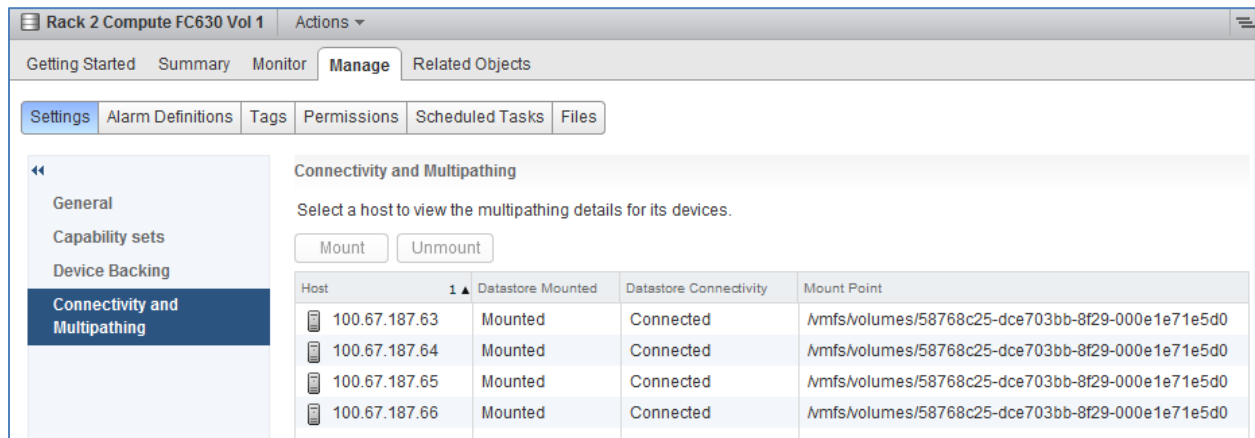


Figure 79 Datastore mounted and connected

The path status to the LUN is verified from each host by selecting a host (from the list in Figure 79) and expanding the **Paths** item near the bottom of the window. Repeat for each host. Each host has two active paths to the LUN as shown at the bottom of Figure 80.

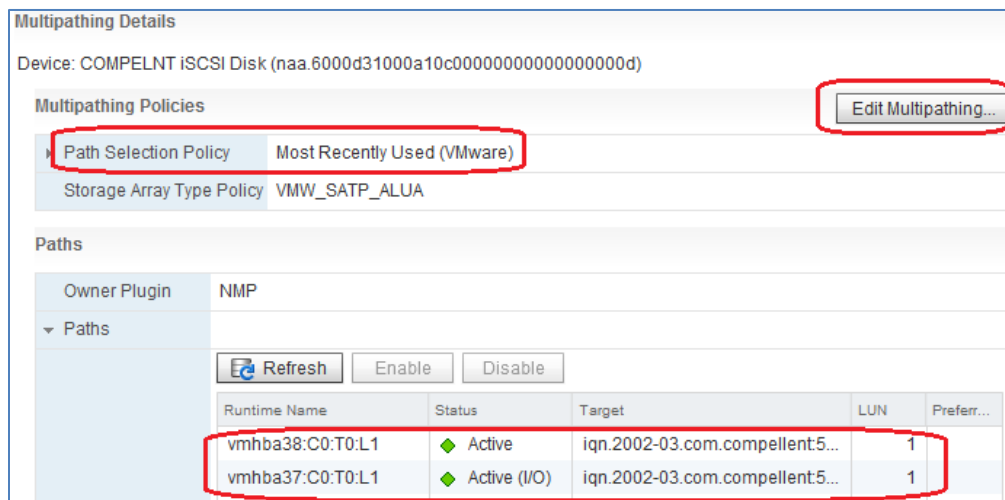


Figure 80 Path status to LUN

The multipathing policy selected will likely default to **Most Recently Used (VMware)** as shown in Figure 80. In this case, one path is marked **Active** and the other marked **Active (I/O)**. The path selection policy may be changed (e.g. to Round Robin) using the **Edit Multipathing** button circled.

## 11 Configure the NSX virtual network

This section covers the configuration steps and best practices to build the NSX topology used in this guide. For more information, refer to the [VMware NSX 6.2 Documentation Center](#).

### 11.1 NSX Manager

The NSX Manager is the centralized network management component of NSX. A single NSX Manager serves a single vCenter Server environment. It provides the means for creating, configuring, and monitoring NSX components such as controllers, logical switches and edge services gateways.

In this guide, NSX Manager is installed as a virtual appliance on an ESXi host in the management cluster. It is available from VMware as an Open Virtualization Appliance (.ova) file and is available for download at [my.vmware.com](http://my.vmware.com).

To install NSX Manager:

1. On the web client **Home** screen, select **Hosts and Clusters**.
2. In the **Navigator** pane, right click on the target ESXi host in the Rack 1 Management cluster and select **Deploy OVF Template**.
3. Select **Local file** and **Browse** to the .ova file (the current naming format is VMware-NSX-Manager-version#.ova). Select the file, click **Open > Next**.
4. Check the **Accept extra configuration options** box and click **Next**.

**Note:** The extra configuration options include IP address, default gateway, DNS, NTP, and SSH.

5. Click **Accept** on the **Accept license agreements** page. Click **Next**.
6. Keep the default name, **NSX Manager**. Select **Datacenter** and click **Next**.
7. On the **Select storage** screen, select a datastore and click **Next**.

**Note:** It is a best practice to use a shared storage datastore to allow for a High Availability (HA) cluster configuration, so that the NSX Manager appliance can be restarted on another host if the original host fails. See the [VMware vSphere 6.0 Documentation Center](#) for HA cluster configuration instructions.

8. On the **Setup networks** screen, select the management network, named **VM Network** by default. Click **Next**.
9. On the **Customize template** screen:
  - a. Enter the **CLI admin** and **CLI Privilege Mode** passwords to be used at the NSX Manager CLI.
  - b. Expand **Network properties**. Provide a **hostname** (for example, nsxmanager). The IP address and gateway information may be filled out or supplied by a DHCP server on your network.
  - c. Fill out the **DNS** section if used on your network and if not provided by DHCP.
  - d. Under **Services Configuration**, Provide the **NTP server** host name or IP address. It is a best practice to use NTP on your NSX management network. Optionally, check the box to **Enable SSH**.
  - e. Click **Next**.
10. The **Ready to complete** screen provides a summary of the installation as shown in Figure 81. Review your settings, check the **Power on after deployment** box, and click **Finish**.

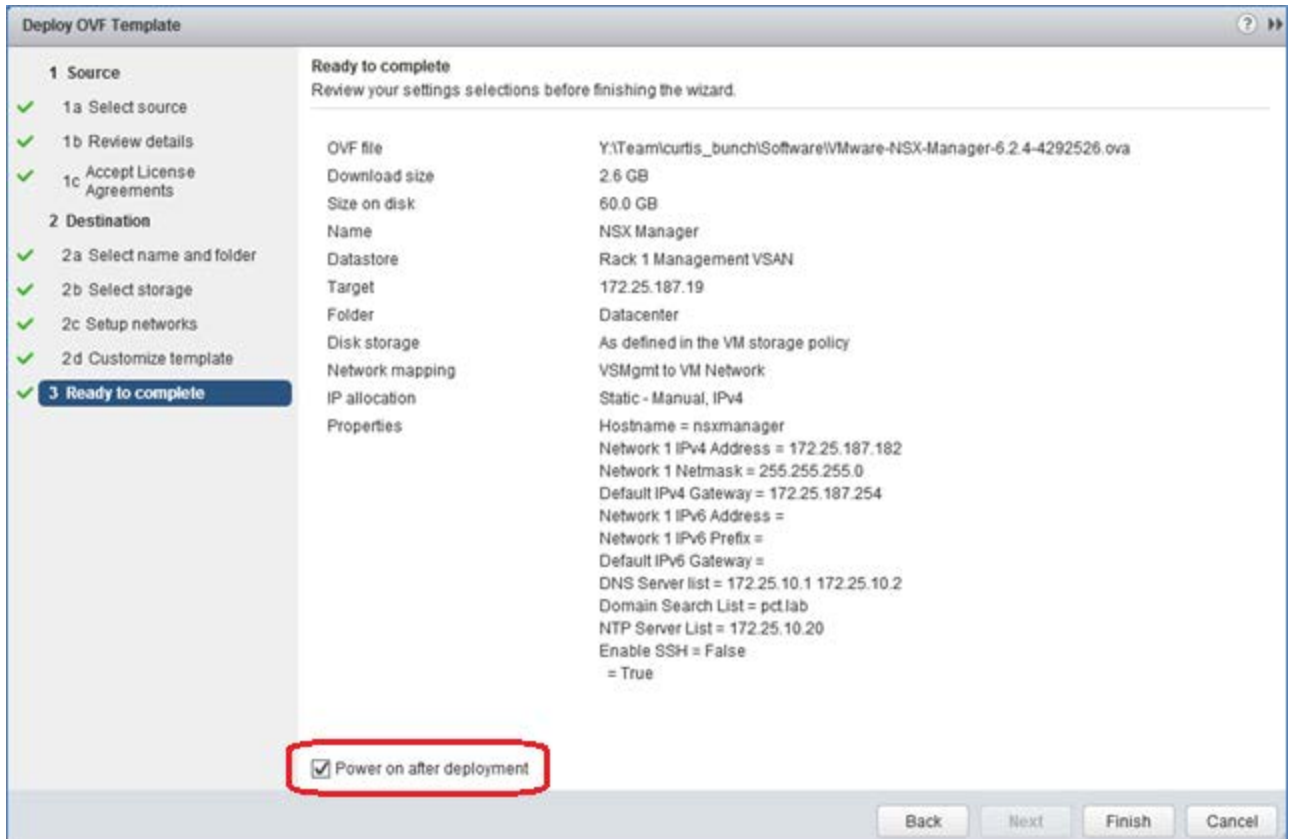


Figure 81 NSX Manager summary screen

The NSX Manager appliance is deployed and boots its Linux OS.

**Note:** If needed, the NSX Manager appliance console is accessible by going to **Hosts and Clusters** and expanding the **Rack 1 Management** cluster. Right click on the **NSX Manager** virtual machine and select **Open Console**.

## 11.2 Register NSX Manager with vCenter Server

**Note:** Only one NSX Manager can be registered with a vCenter Server.

To register the vCenter Server with NSX Manager:

1. After the NSX Manager appliance has booted, go to **[https://<ip\\_address\\_or\\_nsx\\_manager\\_hostname>](https://<ip_address_or_nsx_manager_hostname>)** in a web browser.
2. Login as **admin** with the NSX Manager password specified in the previous section.
3. Click the **Manage vCenter Registration** button.
4. Next to **vCenter Server**, click the **Edit** button.
5. Enter the IP address or host name of the vCenter Server, the vCenter Server user name (for example, administrator@pct.lab), password, and click **OK**.
6. Click **Yes** to trust the certificate when prompted.

7. Verify the vCenter Server status is **Connected** as shown in Figure 82.

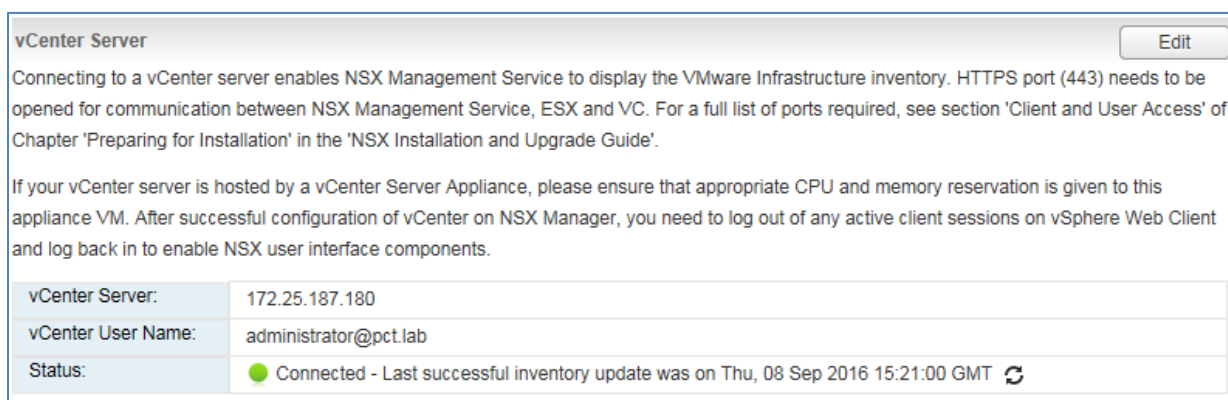


Figure 82 NSX Manager is connected to vCenter Server

The following steps are done in the web client:

1. Log out of the web client if logged in.
2. Log into the web client using the same credentials used to register NSX Manager.
3. The **Networking & Security** icon now appears on the **Home** page as shown in Figure 83.

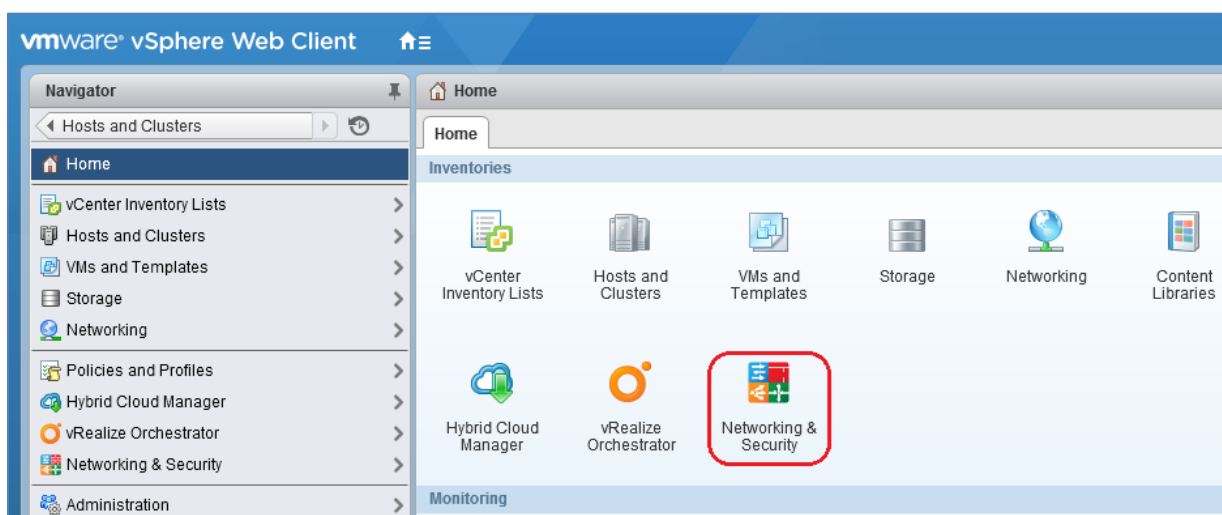


Figure 83 Networking & Security icon

## 11.3 Deploy NSX controllers

NSX controllers are responsible for managing the distributed switching and routing modules in the hypervisors. Three controllers are required in a supported configuration and can tolerate one controller failure while still providing for controller functionality.

NSX controllers communicate on the management network, and do not have any data plane traffic passing through them. Therefore, data forwarding will continue even if all NSX controllers are off line.

As a best practice, each NSX controller should be deployed on a different ESXi host so that a single host failure will not bring down more than one controller. In this guide, there are three hosts in the management cluster with one NSX controller deployed to each host.

To deploy the NSX controllers:

1. In the web client, go to **Home > Networking & Security**.
2. In the **Navigator** pane, select **Installation**.
3. On the **Management** tab under **NSX Controller nodes**, select **+** to open the **Add Controller** dialog box.
4. In the **Add Controller** dialog box:
  - a. Provide a **name** for the first controller, such as NSX Controller 1.
  - b. The **NSX Manager** and **Datacenter** should be selected by default. If not, select them.
  - c. Next to **Cluster/Resource Pool**, select the **Rack 1 Management** cluster.
  - d. Next to **Datastore**, select a previously configured datastore (shared storage is recommended).
  - e. Next to **Host**, select the ESXi host where the NSX controller VM will reside. Use a different host for each controller.
  - f. The **Folder** is optional and is skipped for this guide.
  - g. Next to **Connected To**, click **Select**. Next to **Object Type** select **Network**. Select the management network, **VM Network**, and click **OK**.
  - h. Next to **IP Pool**, click **Select**. When creating the first controller, an IP pool will need to be created. Click **New IP Pool**, and fill out the **Add Static IP Pool** fields similar to the example in Figure 84. Use an address range containing at least 3 available addresses on your management network (one address will be used for each NSX controller).

**Add Static IP Pool**

Name: \* NSX controller pool

Gateway: \* 100.67.187.254  
A gateway can be any IPv4 or IPv6 address.

Prefix Length: \* 24

Primary DNS:

Secondary DNS:

DNS Suffix:

Static IP Pool: \* 100.67.187.183-100.67.187.185  
for example 192.168.1.2-192.168.1.100 or  
abcd:87:87::10-abcd:87:87::20

OK Cancel

Figure 84 Add Static IP Pool dialog box



- i. Click **OK** to create the pool.
- j. In the **Select IP Pool** dialog box, select the pool and click **OK**.
- k. Type and confirm a complex password for the controller cluster. This field only appears when setting up the first controller.
- l. Click **OK** again to deploy the NSX controller.

**Note:** Wait for the deployment to complete as shown in the Status column under NSX Controller nodes before deploying the next controller.

5. Repeat Steps 3 and 4 above for the remaining two controllers, except use the existing IP pool instead of creating a new one in step h.

When all three controllers are deployed, the **Networking & Security > Installation > Management** page appears similar to Figure 85. Each controller's status is shown as **Connected** and each controller has two green boxes (representing status of each controller peer) in the **Peers** column.



Installation

ManagementHost PreparationLogical Network PreparationService Deployments

NSX Managers

⚙️ Actions

🔍 Filter










NSX Manager	IP Address	vCenter	Version
 100.67.187.182	100.67.187.182	 100.67.187.180	6.2.4.4292526

1 items

NSX Controller nodes

➕ ✖ 📄 ⚙️ Actions

🔍 Filter

Name	Controller Node	NSX Manager	Status	Peers	Software Version
NSX Controller 1	100.67.187.183 controller-1	 100.67.187.182	✓ Connected	 	6.2.47844
NSX Controller 2	100.67.187.184 controller-2	 100.67.187.182	✓ Connected	 	6.2.47844
NSX Controller 3	100.67.187.185 controller-3	 100.67.187.182	✓ Connected	 	6.2.47844


3 items

Figure 85 NSX controllers deployed

## 11.4 Prepare host clusters for NSX

Host preparation is the process in which the NSX Manager installs NSX kernel modules on each ESXi host in a cluster. This only needs to be done on clusters that will send and receive traffic on the NSX (virtual) network. In this deployment example, this includes the Rack 2 Compute FC630 and Rack 3 Edge clusters. The Rack 1 Management cluster will not be part of the virtual network.

To prepare the Compute cluster:

1. On the web client **Home** screen, select **Networking & Security**.
2. Select **Installation > Host Preparation**.
3. Click on the row containing **Rack 2 Compute FC630** cluster. Click  **Actions** and click **Install > Yes**.
4. When complete, the cluster's **Installation Status** and **Firewall** columns will both show a green check mark.

**Note:** The VXLAN column will still indicate **Not Configured**. VXLAN will be configured in the next section.

Repeat steps 1-4 above for the Edge cluster.

When complete, the host preparation tab will appear similar to Figure 86.





Installation			
Management	Host Preparation	Logical Network Preparation	Service Deployments
NSX Manager: 100.67.187.182			
NSX Component Installation on Hosts			
 Actions			
Clusters & Hosts	Installation Status	Firewall	VXLAN
▶  Rack 3 Edge	✓ 6.2.4.4292526	✓ Enabled	Not Configured
▶  Rack 2 Compute FC630	✓ 6.2.4.4292526	✓ Enabled	Not Configured
▶  Rack 1 Management	Not Installed	Not Configured	Not Configured

Figure 86 Host Preparation status

The host preparation process installs two vSphere Installation Bundles (VIBs) to each host in the cluster, esx-vsip and esx-vxlan. This can be confirmed running the following command from an ESXi host CLI:

```
esxcli software vib list | grep esx-v
```

The command output will include esx-vsip and esx-vxlan if installed successfully:

esx-vsip	6.0.0-0.0.4249023	VMware	VMwareCertified	2016-09-02
esx-vxlan	6.0.0-0.0.4249023	VMware	VMwareCertified	2016-09-02

**Note:** If additional hosts are later added to the prepared clusters, the required NSX components will be automatically deployed to those hosts.

## 11.5 Configure clusters for VXLAN

VXLAN is configured on a per-cluster basis with each cluster mapped to a VDS. VXLAN configuration creates a VMkernel interface on each host that serves as the software VTEP. This enables virtual network functionality on each host in the cluster.


Before starting, plan an IP addressing scheme for the VTEPs. The number of IP addresses in the pool should be enough to cover all ESXi hosts in the cluster participating in NSX. In this guide, VLAN 55 is used for VXLAN traffic and the IP addressing scheme is shown in Table 8.

Table 8 VTEP IP pool addresses

Cluster Name	IP Pool Name	Network	Gateway	IP Pool Range
Rack 2 Compute FC630	R2 VTEP Pool	10.55.2.0/24	10.55.2.254	10.55.2.1-100
Rack 3 Edge	R3 VTEP Pool	10.55.3.0/24	10.55.3.254	10.55.3.1-100

**Note:** The Rack 1 Management cluster is not configured because its hosts are not part of the virtual network in this guide.

To configure the Rack 2 Compute FC630 cluster for VXLAN:

1. Go to **Home > Networking & Security > Installation** and select the **Host Preparation** tab.
2. In the center pane, select the **Rack 2 Compute FC630** cluster. Click  **Actions > Configure VXLAN**.
3. In the **Configure VXLAN Networking** dialog box:
  - a. Next to **Switch**, ensure the correct VDS is selected (for example, Rack 2 Compute FC630 VDS).
  - b. Set the **VLAN** to **55**.
  - c. Leave the **MTU** set to **1600**.
  - d. Next to **VMKNic IP Addressing**, select **Use IP Pool** and select **New IP Pool** from the drop-down menu. This opens the **Add Static IP Pool** dialog box.
    - i. Next to **Name**, enter **R2 VTEP Pool**.
    - ii. Set the **Gateway** to **10.55.2.254**.
    - iii. Set the **Prefix Length** to **24** (number of bits in the subnet mask).
    - iv. Fill out the DNS information if used on your network.
    - v. Set the **Static IP Pool** to **10.55.2.1-10.55.2.100**.
    - vi. Click **OK**.
4. On the **Configure VXLAN Networking** window, set the **VMKNic Teaming Policy** to **Enhanced LACP**.
5. Click **OK**.

It may take a few minutes for VXLAN configuration to complete. When done, the VXLAN column should indicate **Configured** with a green check mark as shown in Figure 87:

Installation			
Management	Host Preparation	Logical Network Preparation	Service Deployments
NSX Manager: 100.67.187.182			
NSX Component Installation on Hosts			
Actions			
Clusters & Hosts	Installation Status	Firewall	VXLAN
▶ Rack 3 Edge	✓ 6.2.4.4292526	✓ Enabled	Not Configured
▶ Rack 2 Compute FC630	✓ 6.2.4.4292526	✓ Enabled	✓ Configured
▶ Rack 1 Management	Not Installed	Not Configured	Not Configured

Figure 87 VXLAN successfully configured on Rack 2 Compute FC630 cluster

Repeat steps 1-5 above for the Rack 3 Edge cluster with the following changes:

- Step 3.a. - ensure **Rack 3 Edge VDS** is selected.
- Step 3.d. - Replace the pool name and IP addressing as needed per Table 8.

When VXLAN configuration is complete, verify the configuration by viewing the network topology as follows:

5. Go to **Home > Networking**.
6. Select a VDS in a cluster configured for VXLAN, such as **Rack 2 Compute FC630 VDS**.
7. In the center pane, select **Manage > Settings > Topology**.
8. In the topology diagram, there is a new port group with the prefix **vxw-vmknicPg-dvs-**. It is on VLAN 55 and has one VMkernel port for each host in the cluster. Each port has an IP address from the VTEP pool for the cluster, as shown in Figure 88.

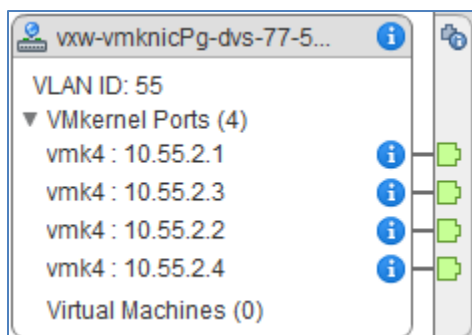


Figure 88 VXLAN VMkernel ports with VTEP IP addresses assigned

## 11.6 Create a segment ID pool

VXLAN tunnels are built between VTEPs. An ESXi host is an example of a typical software VTEP. Each VXLAN tunnel must have a segment ID, which is pulled from a segment ID pool that you create. Segment IDs are used as VNI's. The range of valid IDs is 5000-16777215.

**Note:** Do not configure more than 10,000 VNIs in a single vCenter. vCenter limits the number of distributed port groups to 10,000.

To create a segment ID pool:

On the web client **Home** screen, select **Networking & Security > Installation** and select the **Logical Network Preparation** tab.

1. Click **Segment ID** and click the **Edit** button.
2. Enter a contiguous range for **Segment ID pool**, for example **5000-5999**.
3. Leave the remaining items at their defaults, as shown in Figure 89, and click **OK**.

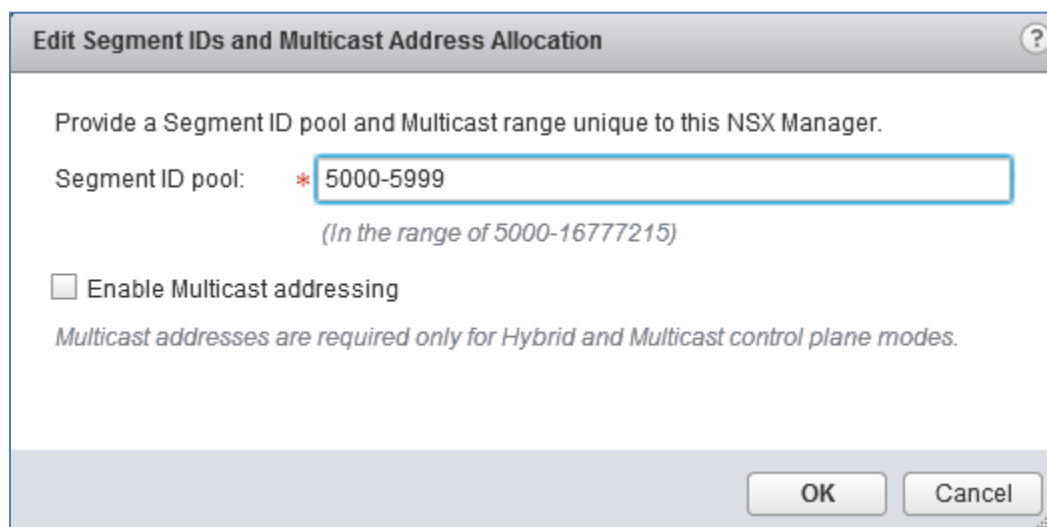


Figure 89 Segment ID pool dialog box

## 11.7 Add a transport zone

A transport zone controls which hosts a logical switch can reach. It can span one or more clusters. Transport zones dictate which clusters and, therefore, which VMs can use a particular virtual network.

An NSX environment can contain one or more transport zones. A cluster can belong to multiple transport zones while a logical switch can belong to only one transport zone. A single transport zone is used in this guide for all NSX-enabled clusters.

To create a transport zone:

1. Go to **Home > Networking & Security > Installation** and select the **Logical Network Preparation** tab.
2. Select **Transport Zones** and click the **+** icon.
3. Name the zone **Transport Zone 1**.
4. Leave **Replication mode** set to **Unicast**.
5. Select the clusters to add to the transport zone. **Rack 2 Compute FC630** and **Rack 3 Edge** are selected as shown in Figure 90:

**New Transport Zone**

Name:

Description:

Replication mode:

- ☐ Multicast  
*Multicast on Physical network used for VXLAN control plane.*
- ☒ Unicast  
*VXLAN control plane handled by NSX Controller Cluster.*
- ☐ Hybrid  
*Optimized Unicast mode. Offloads local traffic replication to physical network.*

Select clusters that will be part of the Transport Zone

	Name	NSX vSwitch	Status
<input checked="" type="checkbox"/>	Rack 3 Edge	Rack 3 Edge VDS	Normal
<input checked="" type="checkbox"/>	Rack 2 Compute FC430	Rack 2 Compute FC430 VDS	Normal
<input type="checkbox"/>			
<input type="checkbox"/>			

OK Cancel

Figure 90 Transport zone with two attached clusters

6. Click **OK** to create the zone.

## 11.8 Logical switch configuration

An NSX logical switch creates a broadcast domain similar to a physical switch or a VLAN. This deployment creates four logical switches, each of which is associated with a unique VNI. The VNI is automatically assigned from the segment ID pool created in Section 11.6.

Table 9 shows the four logical switches used in this deployment:

**Table 9** Logical Switch and VNI Assignment

Logical switch name	VXLAN network ID (VNI)	Network	Used for
Transit Network	5000	172.16.0.0/24	Transit network
Web-Tier	5001	10.10.10.0/24	Web network
App-Tier	5002	10.10.20.0/24	Application network
DB-Tier	5003	10.10.30.0/24	Database network

Figure 91 shows the logical connectivity between the three logical switches used for VM traffic (Web-Tier, App-Tier, and DB-Tier), the Distributed Logical Router (DLR) and the transit logical switch. The DLR acts as the default gateway for each VM connected to its respective logical switch and is configured in Section 11.9.

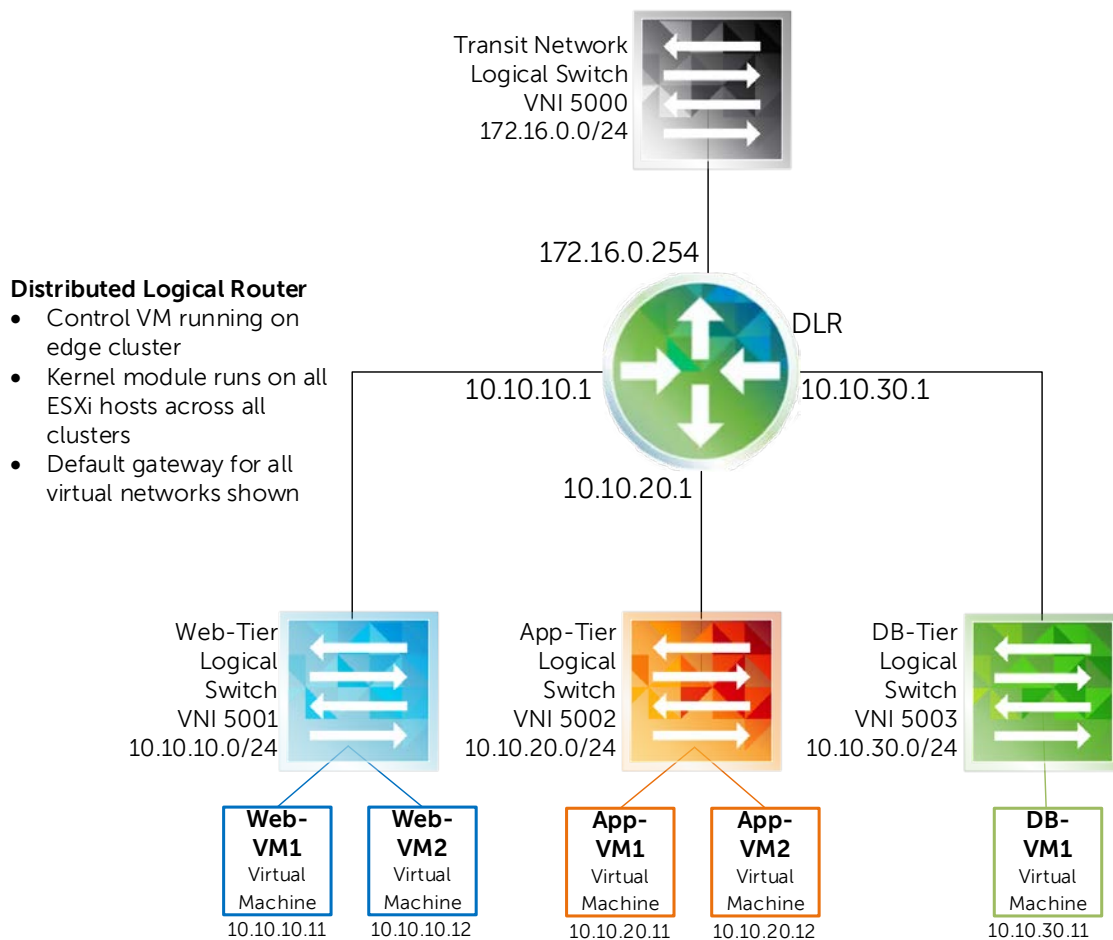


Figure 91 Logical switches and DLR topology

To deploy the four logical switches:

1. On the web client **Home** screen, select **Networking & Security > Logical Switches**.
2. Click the **+** icon to add a new logical switch.
3. In the **New Logical Switch** dialog box:
  - a. Type the first switch name, **Transit Network**.
  - b. Next to **Transport Zone**, **Transport Zone 1** should already be selected. If not, click **Change** and select it.
  - c. Leave **Replication mode** set to **Unicast**, **Enable IP Discovery** checked and **Enable MAC Learning** unchecked.
  - d. Click **OK**.

Repeat the steps above for the remaining logical switches and substitute the proper switch name in step 3a. Ensure that all logical switches are placed in Transport Zone 1.



Figure 92 shows the four logical switches after creation:

Logical Switches							
NSX Manager: 172.25.187.182							
Segment ID	Name	Status	Transport Zone	Hardware Ports Binding	Scope	Control Plane Mode	Tenant
5000	Transit Network	✓ Normal	Transport Zone 1	0	Global	Unicast	virtual wire tenant
5001	Web-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast	virtual wire tenant
5002	App-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast	virtual wire tenant
5003	DB-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast	virtual wire tenant

Figure 92 Logical switches after creation

## 11.9 Distributed Logical Router configuration

A Distributed Logical Router (DLR) is a virtual appliance that provides routing between VXLAN networks. It is installed on a host in the Rack 3 Edge cluster.

Figure 91 shows the DLR's location in the virtual network. All four logical switches connect to it. Table 10 shows the DLR interface IP addresses used in this guide:

Table 10 DLR IP addressing

Interface name	IP address/Subnet prefix
Transit Network	172.16.0.254/24
Web-Tier	10.10.10.1/24
App-Tier	10.10.20.1/24
DB-Tier	10.10.30.1/24

To configure the DLR:

1. Go to **Home > Networking & Security > NSX Edges** and click the icon.
2. In the **New NSX Edge** dialog box:
  - a. Select **Logical (Distributed) Router**, provide a name (**DLR1**, for example). Verify that the **Deploy Edge Appliance** box is checked and click **Next**.
  - b. Provide **CLI credentials** for the DLR, leave other values at their defaults and click **Next**.
  - c. Under **NSX Edge Appliances**, click the icon to create an edge appliance.

- d. In the **Add NSX Edge Appliance** dialog box:
  - i. Set **Cluster/Resource Pool** to **Rack 3 Edge** and select a **Datastore**.
  - ii. The **Host** and **Folder** fields may be left blank. The host is automatically assigned from the cluster.

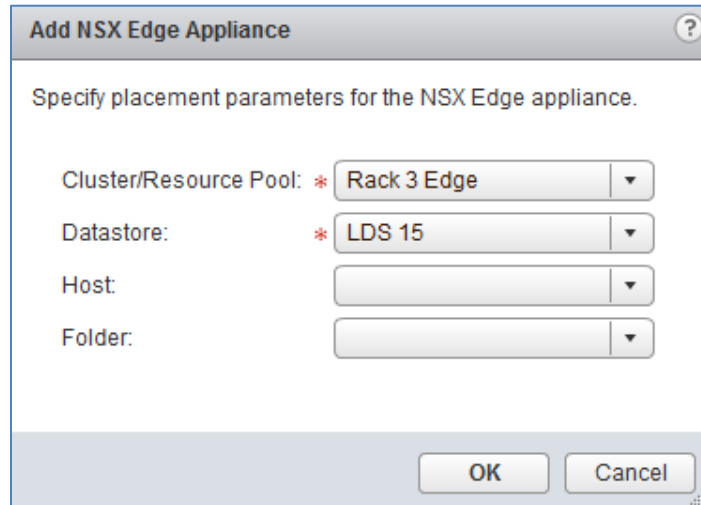






Figure 93 Add NSX Edge Appliance dialog box

- e. Click **OK** to close the **Add NSX Edge Appliance** dialog box and click **Next**.
- f. On the **Configure Interfaces** page, next to **Connected To**, click **Select**.
- g. Be sure **Logical Switch** is selected at the top, and select **Transit Network**. Click **OK**.
- h. To create the DLR uplink interface:
  - i. Under **Configure interfaces of this NSX Edge**, click the  icon to open the **Add Interface** dialog box.
  - ii. Name the interface **Transit Network**.
  - iii. Set **Type** to **Uplink**.
  - iv. Next to **Connected To** click **Select**.
  - v. Be sure **Logical Switch** is selected at the top and select **Transit Network**. Click **OK**.
  - vi. Under **Configure Subnets**, click the .
  - vii. Type **172.16.0.254** for the **Primary IP Address** and set the **Subnet Prefix Length** to **24**.
  - viii. Leave the remaining values at their defaults and click **OK** to close.
- i. To create the DLR internal interfaces:
  - i. Under **Configure interfaces of this NSX Edge**, click the  icon to open the **Add Interface** dialog box.
  - ii. Name the interface **Web-Tier**.
  - iii. Set **Type** to **Internal**.
  - iv. Next to **Connected To** click **Select**.
  - v. Be sure **Logical Switch** is selected at the top and select **Web-Tier**. Click **OK**.
  - vi. Under **Configure Subnets**, click the .
  - vii. Type **10.10.10.1** for the **Primary IP Address** and set the **Subnet Prefix Length** to **24**.
  - viii. Leave the remaining values at their defaults and click **OK**.

- j. Repeat steps i-viii under letter i above to create the remaining DLR internal interfaces (App-Tier and DB-Tier in this example). Substitute **Name**, **Connected To**, and **IP Address** values accordingly as Figure 94 shows:

Name	IP Address	Subnet Prefix Length	Connected To
Transit Network	172.16.0.254*	24	Transit Network
Web-Tier	10.10.10.1*	24	Web-Tier
App-Tier	10.10.20.1*	24	App-Tier
DB-Tier	10.10.30.1*	24	DB-Tier

Figure 94 DLR interfaces configured

- k. Click **Next** when complete.
- l. Uncheck **Configure Default Gateway** and click **Next**.
- m. Click **Finish** to deploy the DLR. It may take a few minutes to complete.

To validate DLR settings and status:

1. Go to **Home > Networking & Security > NSX Edges**.
2. In the center pane, double click on the DLR to open the DLR summary and management page.
3. Select **Manage > Settings > Interfaces**. Verify all settings are correct as shown in Figure 95. If any changes need to be made, click the pencil icon to edit.

vNIC#	Name	IP Address	Subnet Prefix Length	Connected To	Type	Status
2	Transit Network	172.16.0.254*	24	Transit Network	Uplink	✓
10	Web-Tier	10.10.10.1*	24	Web-Tier	Internal	✓
11	App-Tier	10.10.20.1*	24	App-Tier	Internal	✓
12	DB-Tier	10.10.30.1*	24	DB-Tier	Internal	✓

Figure 95 Configured interfaces on the distributed logical router

### 11.9.1 Configure OSPF on the DLR

This topology uses OSPF to provide dynamic routing to the ESG. BGP can be used instead, but for this guide OSPF was selected to provide a distinct demarcation between the physical underlay and the virtual overlay. The ESG serves as the next-hop router in this environment, and is configured in Section 13.1. The NSX default Area 51, which is a not-so-stubby area (NSSA), will be used between the DLR and the ESG.

Configure the Router ID:


1. Go to **Home > Networking & Security > NSX Edges**.
2. In the center pane, double click on the DLR to open the DLR summary and management page.
3. Select **Manage > Routing > Global Configuration**. Click **Edit** in the **Dynamic Routing Configuration** section.
4. Next to **Router ID**, choose the default (**Transit Network – 172.16.0.254**) and click **OK**.
5. Click **Publish Changes** near the top of the screen.

Enable OSPF and configure OSPF features:

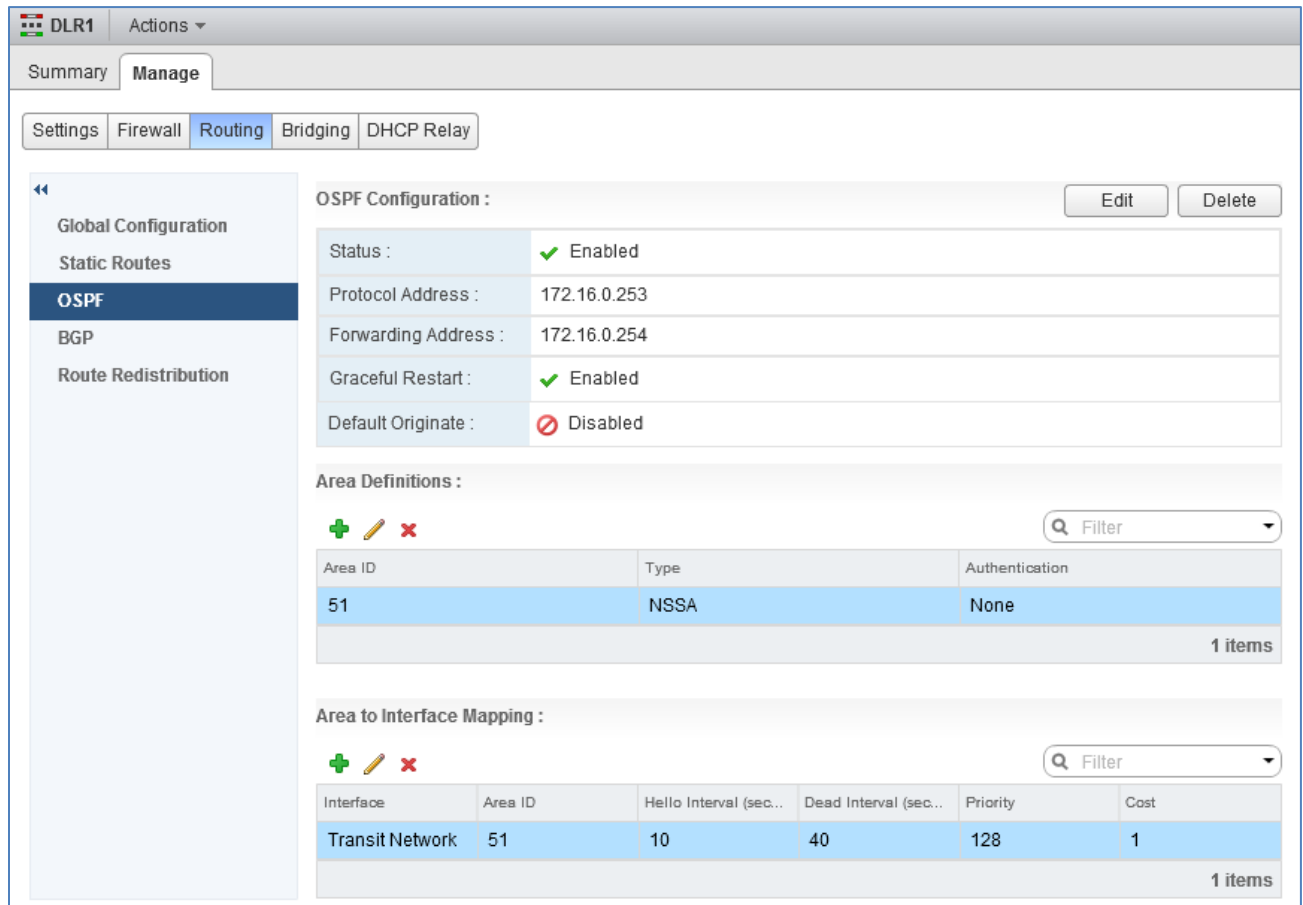
1. On the **Routing** page, select **OSPF**.
2. Click **Edit** at the top right corner of the window and check the **Enable OSPF** box.
3. Set the **Protocol Address** to 172.16.0.253.
4. Set the **Forwarding Address** to 172.16.0.254.
5. Leave the **Enable Grateful Restart** box checked and click **OK**.
6. Click **Publish Changes**.

**Note:** The protocol and forwarding addresses should be from the same subnet. The protocol address is used to form OSPF adjacencies. The forwarding address is the DLR interface IP address.

Enable interfaces to participate in their respective OSPF areas:

1. Click the  icon under **Area to Interface Mapping**.
2. Next to **Interface**, select **Transit Network**.
3. Set the **Area** to **51** (default).
4. Leave all other values at their defaults and click **OK**.
5. Click **Publish Changes**.

When complete, the OSPF page for the DLR should appear similar to Figure 96.



**DLR1** Actions ▾

Summary **Manage**




Settings Firewall **Routing** Bridging DHCP Relay

Global Configuration  
Static Routes  
**OSPF**  
BGP  
Route Redistribution

**OSPF Configuration :** Edit Delete

Status :	✓ Enabled
Protocol Address :	172.16.0.253
Forwarding Address :	172.16.0.254
Graceful Restart :	✓ Enabled
Default Originate :	✗ Disabled




**Area Definitions :**

   Filter

Area ID	Type	Authentication
51	NSSA	None

1 items

**Area to Interface Mapping :**

   Filter

Interface	Area ID	Hello Interval (sec...)	Dead Interval (sec...)	Priority	Cost
Transit Network	51	10	40	128	1

1 items

Figure 96 OSPF configuration complete on the DLR

## 11.9.2 Firewall information

The DLR firewall can be accessed by going to **Home > Networking & Security > NSX Edges**. Double click on the DLR and go to **Manage > Firewall**.

**Note:** Configuration of firewall rules is outside the scope of this document. For more information, refer to the [VMware NSX 6.2 Documentation Center](#).

## 12 Verify NSX network functionality

In this section, a small number of virtual machines are deployed to different clusters to verify connectivity within the NSX network.

**Note:** Virtual machine/guest operating system deployment steps are not included in this document. For instructions, see the [Deploying Virtual Machines](#) section of the vSphere 6.0 online documentation. Guest operating systems can be any supported by ESXi 6.0. Microsoft Windows Server 2012 R2 was used as the guest operating system for each virtual machine deployed in this section.

Shared storage is required to take advantage of advanced VMware features such as DRS and HA. For example, when creating VMs in the **Rack 2 Compute FC630** cluster, use the datastore named **Rack 2 Compute FC630 Vol 1** created in Section 10.15 of this guide.

### 12.1 Deploy virtual machines

For this example, three VMs are deployed in the Rack 2 Compute FC630 cluster. The first two represent application servers and are named App-VM1 and App-VM2. The third represents a web server and is named Web-VM1.

A fourth VM, App-VM3, is deployed in the Rack 3 Edge cluster to validate communication between clusters. The added VMs are shown in Figure 97.

**Note:** The Rack 1 Management cluster is not configured for VXLAN traffic and therefore is not part of the virtual network validation.

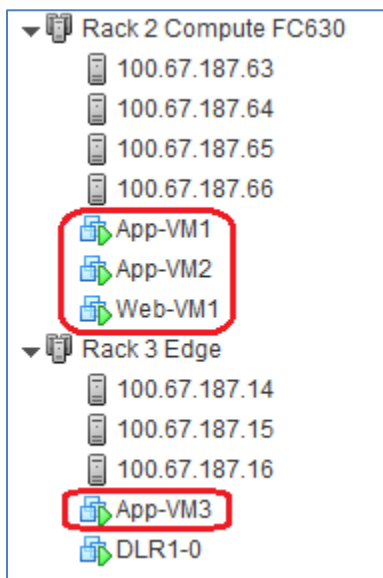


Figure 97 Hosts and Clusters view - virtual machines deployed

## 12.2 Connect virtual wires

A virtual wire is a distributed port group that is automatically created on each VDS as logical switches are created. The virtual wire descriptor contains the name of the logical switch and the logical switch's segment ID.

To connect a VM to a virtual wire:

1. Go to **Home > Hosts and Clusters**.
2. Right click on the first VM, **App-VM1**, and select **Edit Settings**.
3. Next to the **Network adapter**, select the virtual wire on the **App-Tier** network as shown in Figure 98.

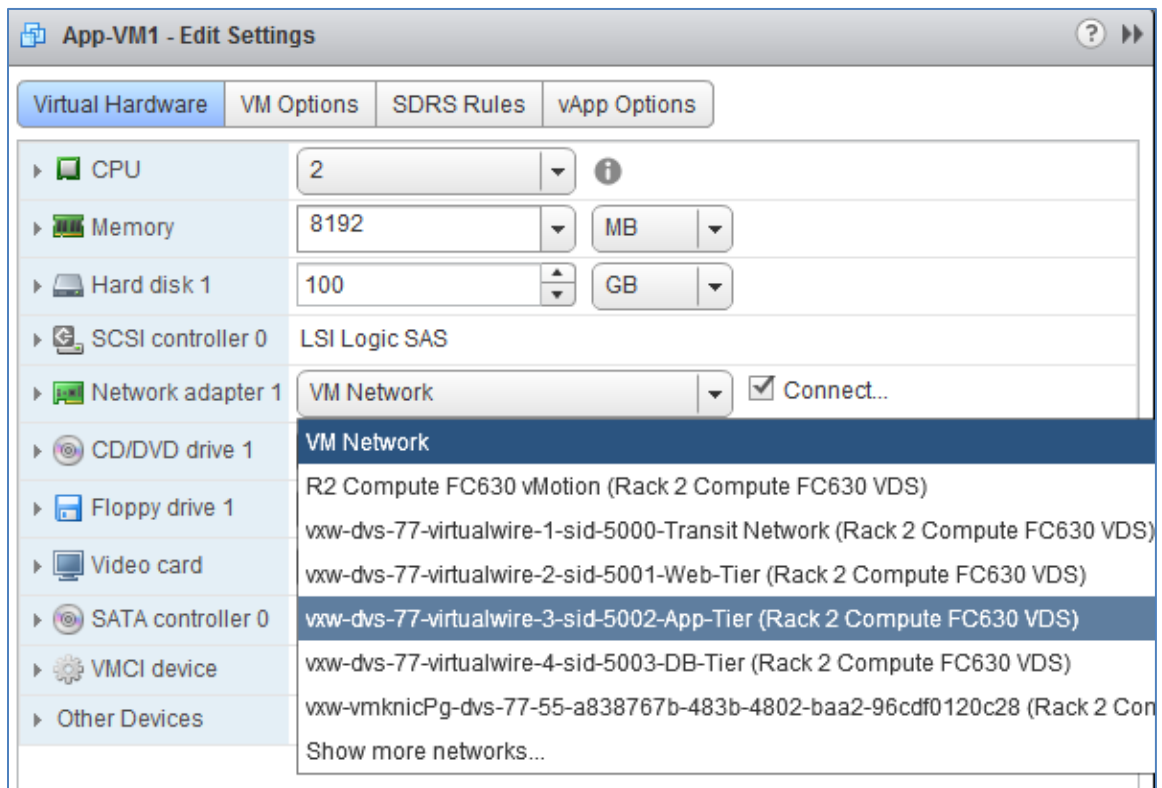


Figure 98 Virtual wire on App-Tier network selected

Repeat steps 1-3 for remaining VMs to be placed on the **App-Tier** segment (**App-VM2** and **App-VM3** in this example.)

Repeat for the VM named **Web-VM1**, except select the virtual wire on the **Web-Tier** segment in step 3.

## 12.3 Configure networking in the guest OS

Power on the virtual machines. Log in to a guest OS by right clicking on the VM and selecting **Open Console**. Use the normal procedure in the guest OS to configure networking.

Using the virtual networking IP address scheme covered in Section 11.8, the IP addresses, subnet masks, and gateway are configured on each VM per Table 11. The gateway addresses are the DLR internal interfaces.

Table 11 Virtual machine IP addressing

Virtual Machine	Cluster	IP Address	Gateway
Web-VM1	Rack 2 Compute FC630	10.10.10.11/24	10.10.10.1
App-VM1	Rack 2 Compute FC630	10.10.20.11/24	10.10.20.1
App-VM2	Rack 2 Compute FC630	10.10.20.12/24	10.10.20.1
App-VM3	Rack 3 Edge	10.10.20.13/24	10.10.20.1

## 12.4 Test connectivity

**Note:** Guest operating system firewalls may need to be temporarily disabled or modified to allow responses to ICMP ping requests for this test. By default, the firewall settings on the DLR allow this type of internal traffic.

Within the source guest operating system, ping the destination VMs using Table 12 as a guide. Successful pings validate the segment tested is configured properly.

Table 12 Test examples to validate connectivity

Source	Destination	Validates
App-VM1	App-VM2	Connectivity within the cluster on same segment.
App-VM1	App-VM3	Connectivity between clusters on the same segment.
App-VM1	Web-VM1	Connectivity within the cluster on different segments.
Web-VM1	App-VM3	Connectivity between clusters on different segments.

**Note:** The 2nd test in Table 12 (App-VM1 to App-VM3) is a good example of virtual layer 2 (switched) traffic over a physical layer 3 (routed) network. Both VMs are on the 10.10.20.0/24 virtual network but they are physically located on ESXi hosts in different racks. Therefore, traffic between these VMs is routed through the spine switches in the physical network.



## 13 Communicate outside the virtual network

In most cases, some virtual machines on the NSX network need to communicate with machines on external traditional networks. Two devices designed to handle this traffic are Edge Services Gateways (ESGs) and hardware VTEPs.

For communication between systems on virtual and physical networks, Dell EMC recommends using an ESG for north-south traffic entering and leaving the data center and a hardware VTEP for east-west traffic within the data center.

### 13.1 Edge Services Gateway

An ESG is an NSX virtual appliance similar to a DLR. Dell EMC recommends using an ESG to handle north-south traffic between the data center's virtual network and the WAN or network core. This allows the administrator to take advantage of additional features provided by the ESG, such as load balancing and VPN services.

The physical topology for the edge cluster is shown in Figure 99. The edge cluster contains the DLR and ESG virtual appliances.

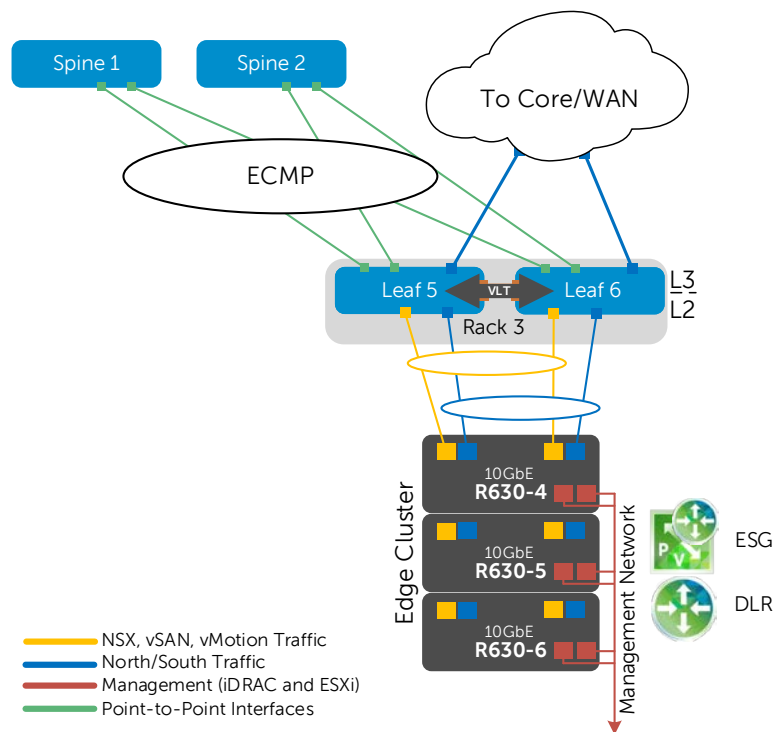


Figure 99 Rack 3 Edge Cluster physical topology

In Figure 99, the yellow connections leading into the edge cluster represent two 10GbE NICs per host, with each pair configured as a port channel. These connections handle NSX and vMotion traffic within the leaf-spine network.

The blue links are in a separate port channel that handles north and southbound traffic. Leaf switches 5 and 6 use OSPF to create an adjacency with the ESG virtual machine. Leaf switch edge configuration was covered in Section 6.3.1.

### 13.1.1 Add a distributed port group

Before deploying an ESG, an additional, VLAN-backed, distributed port group needs to be created on the edge cluster VDS. This additional port group handles all north and southbound traffic between the NSX environment and the network core or WAN.

**Note:** VLAN 66 was configured on Leaf 5 and Leaf 6 in Section 6.3.1.

To create the port group:


1. In the web client, go to **Home > Networking**.
2. Right click on **Rack 3 Edge VDS**. Select **Distributed Port Group > New Distributed Port Group**.
3. In the **New Distributed Port Group** wizard:
  - a. Provide the name **R3 Edge VLAN 66** and click **Next**.
  - b. On the **Configure Settings** page, set **VLAN type** to **VLAN** and set the **VLAN ID** to **66**.
  - c. Click **Next > Finish**.

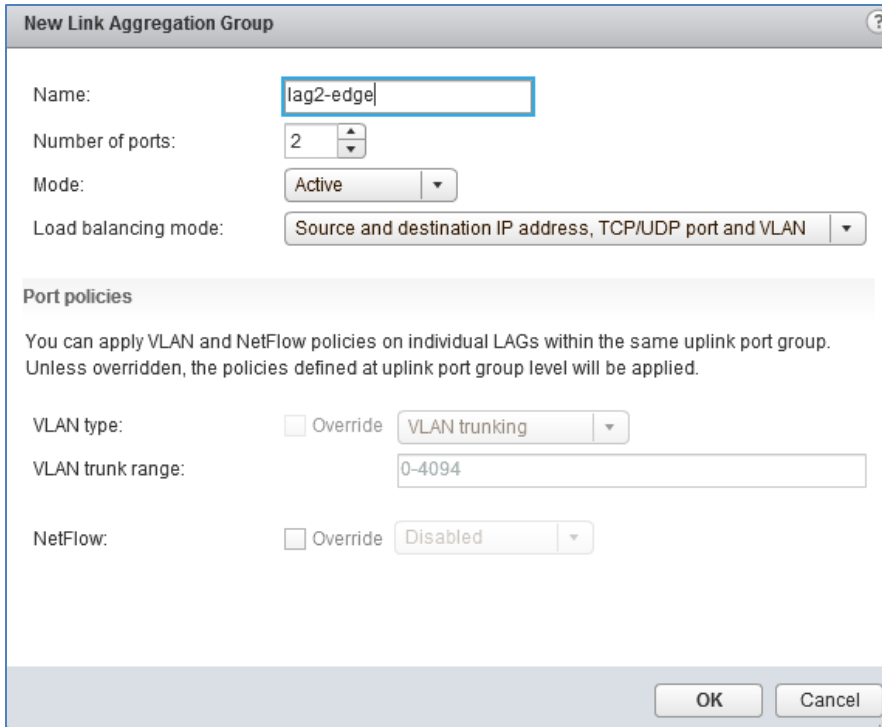
This creates the port group named R3 Edge VLAN 66.

### 13.1.2 Create second LACP LAG

On the Rack 3 Edge VDS, two LAGs are needed as shown in Figure 99 above. One is for traffic within the data center shown in yellow (**lag1**, created earlier in Section 9.3), and one for edge traffic to the WAN/network Core shown in blue (**lag2-edge**, to be created here).

To configure the edge LAG on **Rack 3 Edge VDS**:

1. Go to **Home > Networking**.
2. In the **Navigators** pane, select **Rack 3 Edge VDS**.
3. In the center pane, select **Manage > Settings > LACP**.
4. On the **LACP** page, click the  icon. The **New Link Aggregation Group** (LAG) dialog box opens.
5. Set the **Name** to **lag2-edge**.
6. Set the **Number of ports** equal to the number of physical uplinks in the LAG on each ESXi host. In this deployment, R630 hosts use two links for the edge LAG so this number is set to **2**.
7. Set the **Mode** to **Active**. The remaining fields can be set to their default values as shown in Figure 100.



**New Link Aggregation Group**

Name:

Number of ports:

Mode:

Load balancing mode:

**Port policies**

You can apply VLAN and NetFlow policies on individual LAGs within the same uplink port group. Unless overridden, the policies defined at uplink port group level will be applied.

VLAN type: ☐ Override

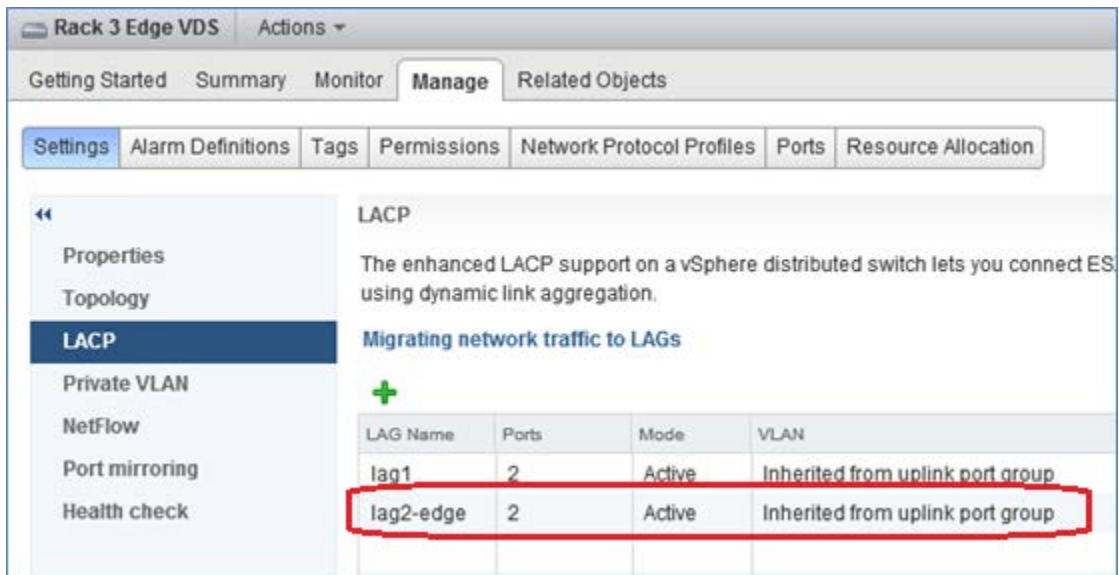
VLAN trunk range:

NetFlow: ☐ Override

Figure 100 Lag2-edge configuration

8. Click **OK** to close the dialog box. This creates **lag2-edge** on the VDS.

The refresh icon (🔄) at the top of the screen may need to be clicked for the lag to appear in the table as shown in Figure 101.



**Rack 3 Edge VDS** Actions ▾

Getting Started Summary Monitor **Manage** Related Objects

Settings Alarm Definitions Tags Permissions Network Protocol Profiles Ports Resource Allocation

« Properties Topology **LACP** Private VLAN NetFlow Port mirroring Health check

**LACP**

The enhanced LACP support on a vSphere distributed switch lets you connect ESX using dynamic link aggregation.

Migrating network traffic to LAGs

+





LAG Name	Ports	Mode	VLAN
lag1	2	Active	Inherited from uplink port group
lag2-edge	2	Active	Inherited from uplink port group

Figure 101 Lag2-edge created on Rack 3 Edge VDS

### 13.1.3 Assign uplinks to the second LAG

**Note:** Before starting this section, be sure you know the vmnic-to-physical adapter mapping for each host. This can be determined by going to **Home > Hosts and Clusters** and selecting the host in the **Navigator** pane. In the center pane select **Manage > Networking > Physical adapters**. These are the vmnics connected to port channels 12, 14 and 16 on Leaf Switches 5 and 6.

To assign uplinks to lag2-edge:

1. Go to **Home > Networking**.
2. Right click on **Rack 3 Edge VDS**, and select **Add and Manage Hosts**.
3. In the **Add and Manage Hosts** dialog box:
  - a. On the **Select task** page, make sure **Manage host networking** is selected. Click **Next**.
  - b. On the **Select hosts** page, click the  **Attached hosts** icon. Select all hosts in the Rack 3 Edge cluster. Click **OK > Next**.
  - c. On the **Select network adapters tasks** page, be sure the **Manage physical adapters** box is checked. Be sure all other boxes are unchecked. Click **Next**.
  - d. On the **Manage physical network adapters** page, each host is listed with its vmnics beneath it.
    - i. Select the first vmnic on the first host and click  **Assign uplink**.
    - ii. Select **lag2-edge-0 > OK**.
    - iii. Select the second vmnic on the first host and click  **Assign uplink**.
    - iv. Select **lag2-edge-1 > OK**.
  - e. Repeat steps i – iv for the remaining hosts. Click **Next** when done.
  - f. On the **Analyze impact** page, **Overall impact status** should indicate  **No impact**.
  - g. Click **Next > Finish**.

When complete, the **Manage > Settings > Topology** page for **Rack 3 Edge VDS** should look similar to Figure 102.

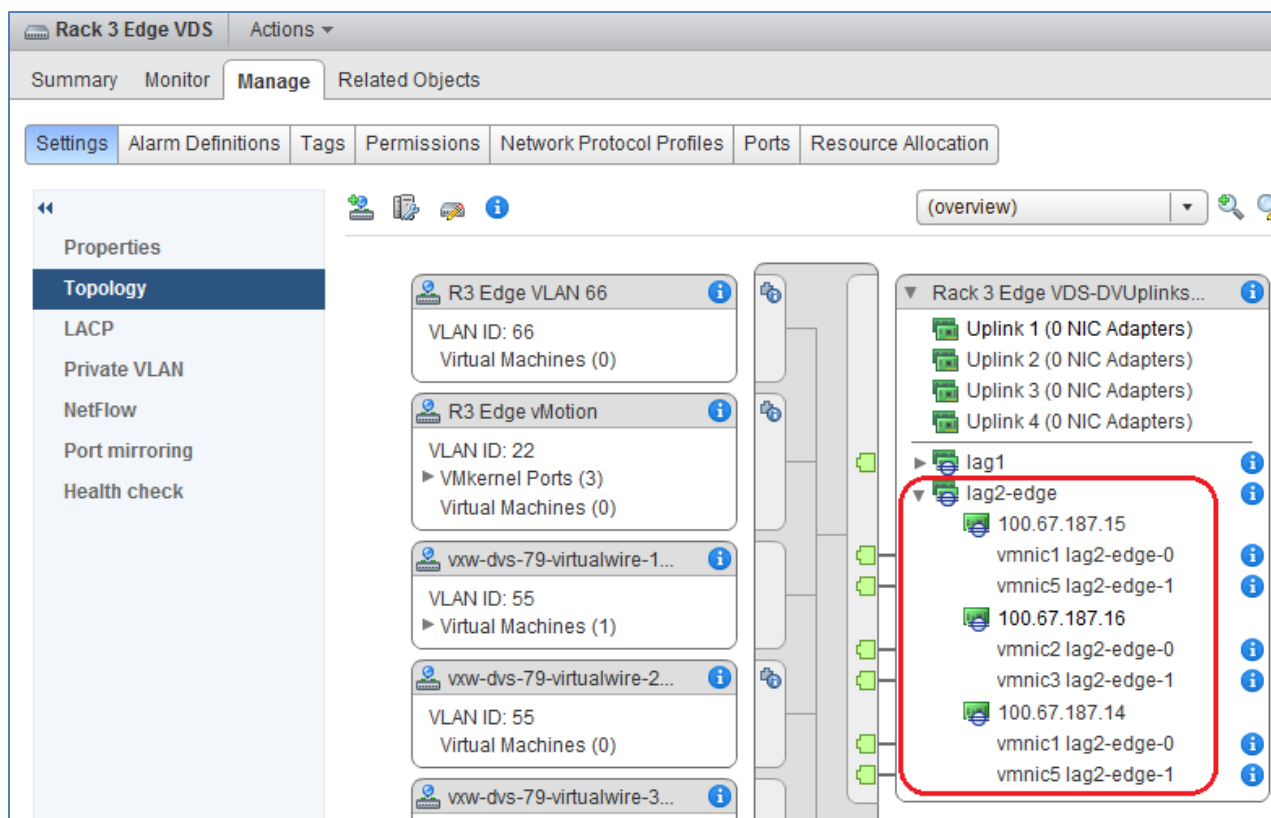


Figure 102 Lag2-edge configured on Rack 3 Edge VDS

This configuration brings up the edge LAGs (port channels 12, 14 and 16) on the upstream switches. This can be confirmed by running the `show vlt detail` command on the upstream switches (Leaf 5 and Leaf 6).

The Local and Peer Status columns indicate UP for all port channels.

Leaf-5#`show vlt detail`

Local LAG Id	Peer LAG Id	Local Status	Peer Status	Active VLANs
2	2	UP	UP	1, 22, 55
4	4	UP	UP	1, 22, 55
6	6	UP	UP	1, 22, 55
12	12	UP	UP	1, 66
14	14	UP	UP	1, 66
16	16	UP	UP	1, 66

### 13.1.4 Configure port groups for teaming and failover

1. On the web client **Home** screen, select **Networking**.
2. Right click on **Rack 3 Edge VDS**. Select **Distributed Port Group > Manage Distributed Port Groups**.
3. Select the **Teaming and failover** checkbox. Click **Next**.
4. Click **Select distributed port groups**.
5. Check the box next to the **R3 Edge VLAN 66** port group. Click **OK > Next**.
6. On the **Teaming and failover** page, move **lag2-edge** up to the **Active uplinks** section. Move **Uplinks 1-4** down to the **Unused uplinks** section as shown in Figure 103.

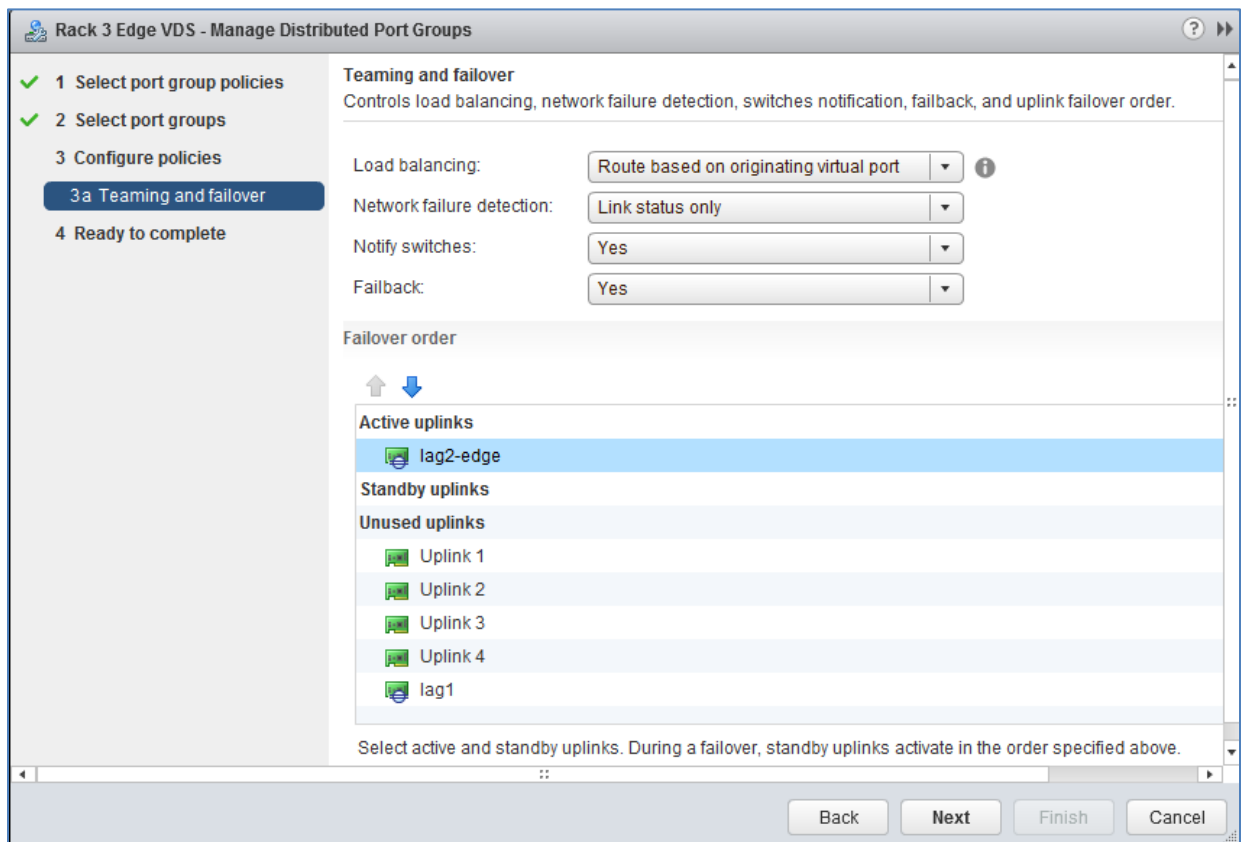



Figure 103 Rack 3 Edge VDS teaming and failover settings

7. Click **Next > Finish**.


### 13.1.5 Deploy the Edge Services Gateway

Now that layer 2 connectivity between the edge cluster and the two leaf switches has been established, the ESG appliance is added and configured.

To deploy the ESG:

1. Go to **Home > Networking & Security > NSX Edges** and click the  icon.
2. In the **New NSX Edge** dialog box:
  - a. Select **Edge Services Gateway** and name it **ESG**.
  - b. Verify the **Deploy NSX Edge** box is checked and click **Next**.
  - c. Provide **CLI credentials** for the ESG, leave other settings at defaults, and click **Next**.
  - d. Next to **Appliance Size**, select a size. **Large** is selected for this guide.

**Note:** See [System Requirements for NSX](#) for ESG sizing specifications.

- e. Click the  icon to create an edge appliance
- f. In the **Add NSX Edge Appliance** dialog box:
  - i. Set **Cluster/Resource Pool** to **Rack 3 Edge** and select a **Datastore**.
  - ii. The **Host** and **Folder** fields may be left blank. The host is automatically assigned from the cluster. Click **OK > Next**.

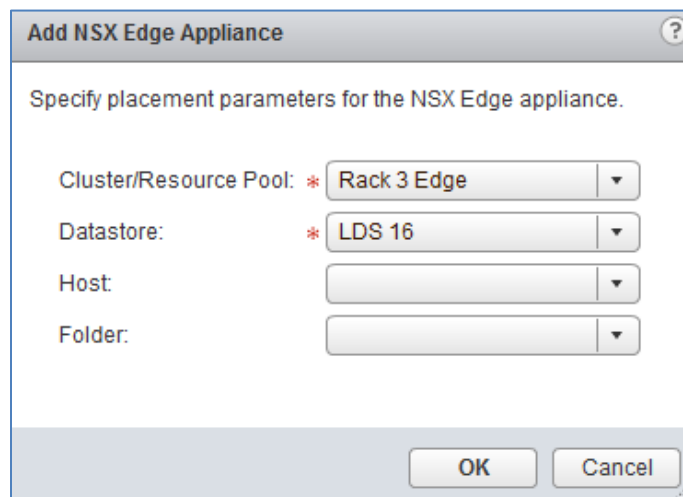




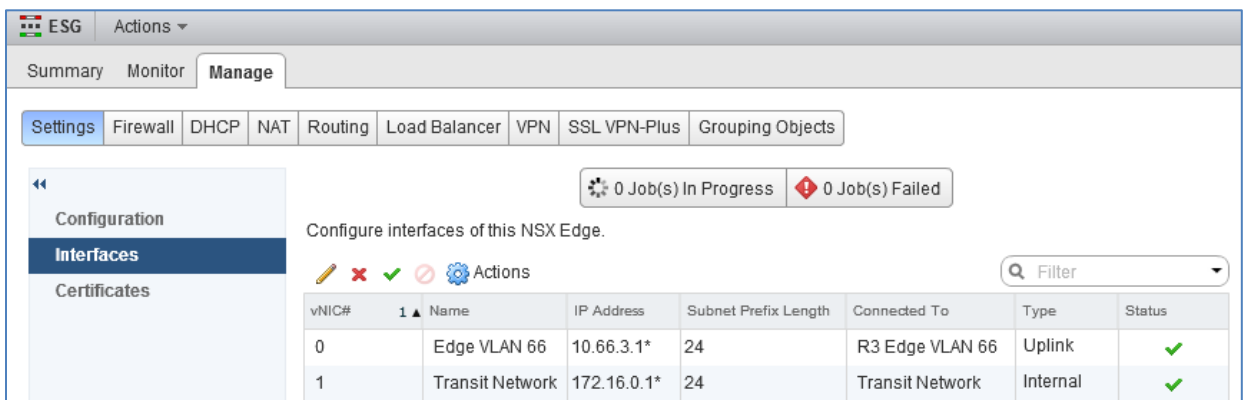
Figure 104 Add NSX edge appliance dialog box

- g. On the **Configure Interfaces** page, click the  icon to open the **Add NSX Edge Interface** dialog box.
  - i. **Name** the interface **Edge VLAN 66**
  - ii. Set **Type** to **Uplink**
  - iii. Next to **Connected To**, click **Select**.
  - iv. Be sure **Distributed Portgroup** is selected at the top and select **R3 Edge VLAN 66**. Click **OK**.
  - v. Click the  icon above **Primary IP Address**.

- vi. Enter **10.66.3.1** for the **Primary IP Address** and set the **Subnet Prefix Length** to **24**.
- vii. Leave the remaining values at their defaults and click **OK** to close.
- h. Click the **+** icon to open the **Add NSX Edge Interface** dialog box again.
  - i. **Name** the interface **Transit Network**.
  - ii. Set **Type** to **Internal**
  - iii. Next to **Connected To**, click **Select**.
  - iv. Be sure **Logical Switch** is selected and select **Transit Network (5000)**. Click **OK**.
  - v. Click the **+** icon above **Primary IP Address**.
  - vi. Type **172.16.0.1** for the **Primary IP Address** and set the **Subnet Prefix Length** to **24**.
  - vii. Leave the remaining values at their defaults and click **OK > Next**.
- i. Uncheck **Configure Default Gateway** and click **Next**.
- j. Leave **Configure Firewall default policy** unchecked and click **Next**.
- k. Click **Finish** to deploy the ESG. It may take a few minutes to complete.

To validate ESG settings:

1. Go to **Home > Networking & Security > NSX Edges**.
2. In the center pane, double click on the ESG to open the ESG summary and management page.
3. Select **Manage > Settings > Interfaces**. Verify settings are correct as shown in Figure 105.



The screenshot shows the ESG management console. The left sidebar has a tree view with 'Configuration', 'Interfaces' (selected), and 'Certificates'. The main area has tabs for 'Settings', 'Firewall', 'DHCP', 'NAT', 'Routing', 'Load Balancer', 'VPN', 'SSL VPN-Plus', and 'Grouping Objects'. The 'Settings' tab is active, and the 'Interfaces' sub-tab is selected. A table displays the configured interfaces for the NSX Edge.

vNIC#	Name	IP Address	Subnet Prefix Length	Connected To	Type	Status
0	Edge VLAN 66	10.66.3.1*	24	R3 Edge VLAN 66	Uplink	✓
1	Transit Network	172.16.0.1*	24	Transit Network	Internal	✓

Figure 105 Configured interfaces on the ESG



### 13.1.6 Configure OSPF on the ESG

Configuring OSPF on the ESG enables the ESG to learn and advertise routes from the core network/WAN upstream. This deployment defines two tasks for OSPF participation. The first task defines Area 0 and creates route adjacencies to the two leaf switches, Leaf 5 and Leaf 6. The second task adds the NSX default Area 51 for use between the ESG and DLR.



Configure the Router ID:

1. Go to **Home > Networking & Security > NSX Edges**.
2. In the center pane, double click on the ESG to open the ESG summary and management page.
3. Select **Manage > Routing > Global Configuration**. Next to **Dynamic Routing Configuration** click **Edit**.
4. Next to **Router ID**, keep the default, **Edge VLAN 66 – 10.66.3.1**, and click **OK**.
5. Click **Publish Changes** near the top of the screen

Enable OSPF:

1. On the **Routing** tab, select **OSPF**.
2. Next to **OSPF Configuration**, click **Edit**.
3. Check the **Enable OSPF** box and leave the **Enable Grateful Restart** box checked. Click **OK**.
4. Click **Publish Changes**.

Enable interfaces to participate in their respective OSPF areas:

1. Click the  icon under **Area to Interface Mapping**.
2. Next to **vNIC**, select **Edge VLAN 66**.
3. Set the **Area** to **0**.
4. Leave all other values at their defaults and click **OK**.
5. Click the  icon under **Area to Interface Mapping**.
6. Next to **vNIC**, select **Transit Network**.
7. Set the **Area** to **51** (default).
8. Leave all other values at their defaults and click **OK**.
9. Click **Publish Changes**.

When complete, the OSPF page for the ESG should be similar to Figure 106. Two interfaces are mapped to two separate OSPF areas. The external interface is mapped to Area 0 and the internal interface is mapped to Area 51.

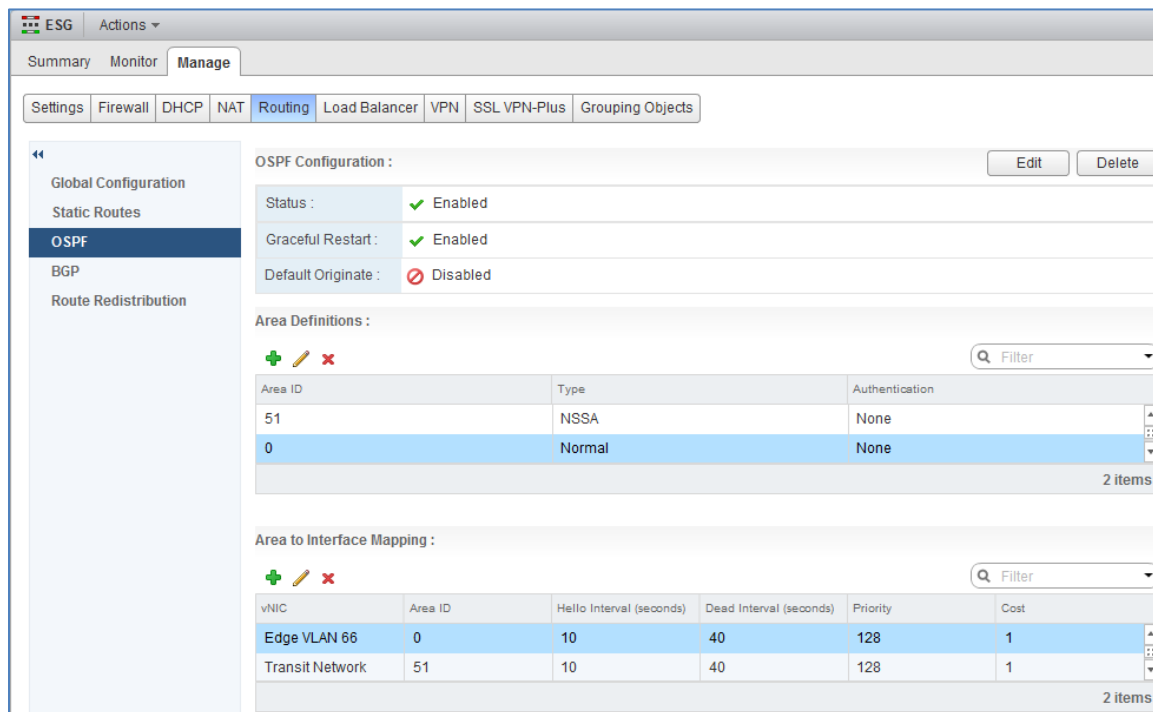


Figure 106 ESG OSPF configuration complete

At this point, all OSPF area 0 adjacencies are established. Run the command `show ip ospf neighbor` on leaf switches 5 and 6 to validate this.

```
Leaf-5#show ip ospf neighbor
```

Neighbor ID	Pri	State	Dead Time	Address	Interface	Area
10.66.3.1	128	FULL/DR	00:00:38	10.66.3.1	Vl 66	0
10.66.3.253	1	FULL/DROTHER	00:00:30	10.66.3.253	Vl 66	0

## 13.1.7 High Availability configuration

**Note:** For more information, see the [NSX Edge High Availability](#) section of the VMware NSX 6.2 documentation.

Enabling HA deploys a backup copy of the ESG (as an additional VM) to another host in the Edge Cluster to act as a standby ESG. The standby provides backup in case of a failure with the active ESG.

To enable HA for the ESG:

1. Go to **Home > Networking & Security > NSX Edges**.

2. In the center pane, double click on the ESG to open the ESG summary and management page.
3. Select **Manage > Settings > Configuration**.
4. Next to **HA Configuration**, click **Change**.
  - a. In the **Change HA configuration** box, set **HA Status** to **Enable**.
  - b. Set the **vNIC** to **Transit Network**.
  - c. Leave the remaining values at their defaults and click **OK**.

After a few minutes, the standby ESG is deployed. The ESG **Configuration** page appears similar to Figure 107. **ESG-0 (Active)** and **ESG-1** are shown with **Deployed** status. It may take a few minutes for (Active) to appear next to ESG-0.

The screenshot shows the ESG Configuration page with the following sections:

- Details:**
  - Size: Large
  - Host Name: NSX-edge-3
  - Auto generate rules: Enabled
  - Syslog servers: (Change)
  - Server 1:
  - Server 2:
- HA Configuration:**
  - HA Status: Enabled
  - vNIC: 1
  - Declare Dead Time: 15
  - Logging: Disabled
  - Log level: Info
- DNS Configuration:**
  - DNS Server 1:
  - DNS Server 2:
  - Cache Size: 16
  - Logging: Disabled
  - Log level: Info
- NSX Edge Appliances:**

Name	Status	Host	Datastore	Folder	Resource Pool
ESG-0 (Active)	Deployed		Rack 3 Edge VSAN		Rack 3 Edge
ESG-1	Deployed		Rack 3 Edge VSAN		Rack 3 Edge

Figure 107 ESG high availability enabled with shared storage

**Note:** The Datastore shown in Figure 107 is shared storage (named **Rack 3 Edge VSAN**). If shared storage, e.g. a VSAN or SAN, is *not* configured on the edge cluster, ESG-1 will be deployed to the same local datastore (and same ESXi host) as ESG-0. In this case, for fault tolerance, move ESG-1's local datastore and host as follows: Click on **ESG-1** and click the pencil icon. Leave the **Datacenter** and **Cluster** selections as is. Change **Datastore** to a different local datastore in the edge cluster (the new host is automatically selected). Click **OK**. After a few minutes, ESG-1 will be migrated to the new datastore and host.

North-south access has been established using OSPF as the dynamic routing protocol. Figure 108 illustrates this configuration.

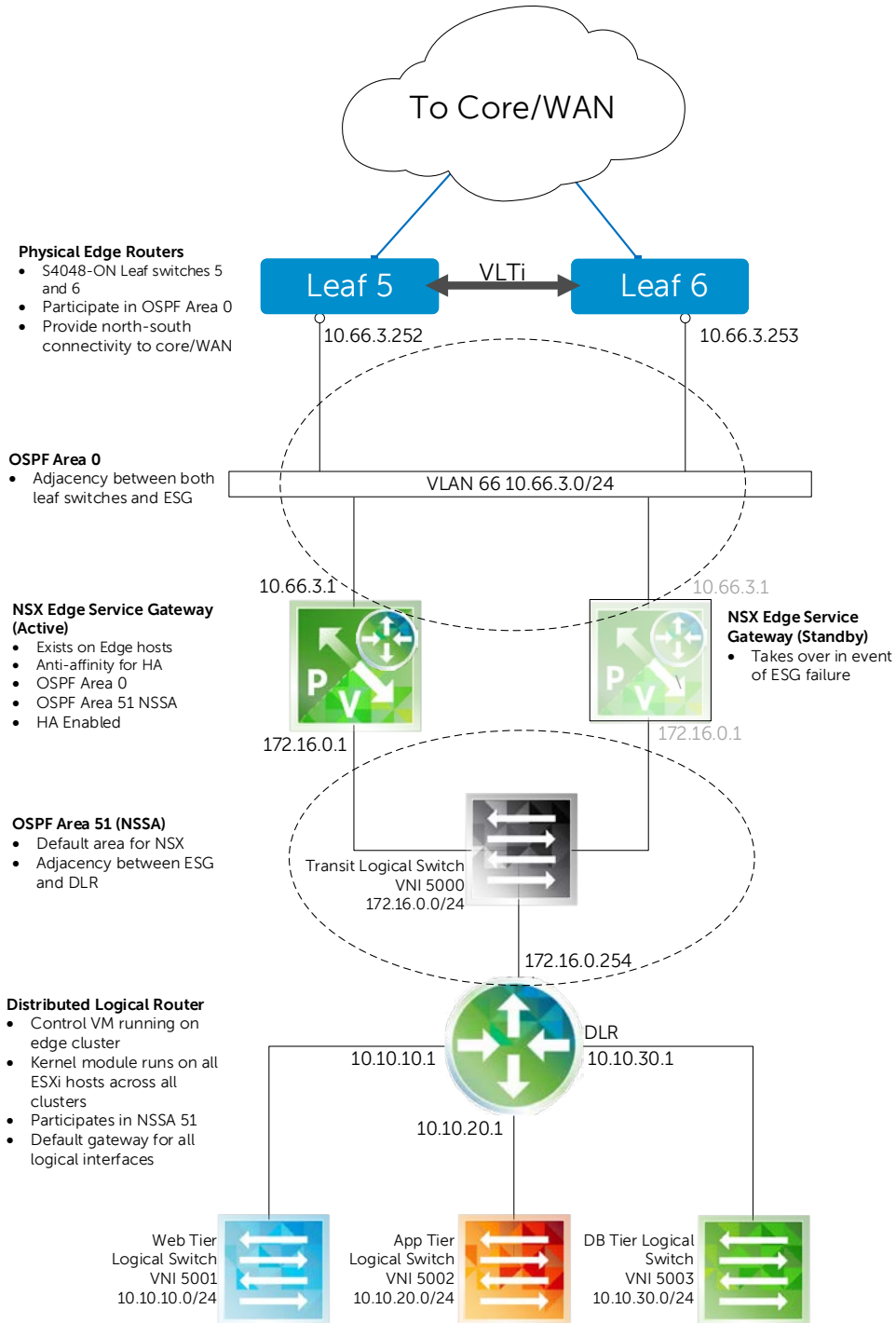


Figure 108 Logical overview of NSX edge

**Note:** Extending leaf switches 5 and 6 to the network core or WAN is outside the scope of this document.

## 13.1.8 ESG validation

### 13.1.8.1 Commands and output

Access the ESG console by going to **Hosts & Clusters**, right clicking on the active ESG VM (**ESG-0**), and selecting **Open Console**. Login using the credentials specified when the ESG was created in Section 13.1.5 (default user name is **admin**).

Basic troubleshooting commands and output run from the ESG CLI are as follows:

```
NSX-edge-3-0> traceroute 10.10.10.11 ← IP address of Web-VM1
traceroute to 10.10.10.11 (10.10.10.11), 30 hops max, 60 byte packets
 1  172.16.0.254 (172.16.0.254)  0.059 ms  1002.063 ms  1002.069 ms
 2  10.10.10.11 (10.10.10.11)  0.534 ms  *  *
```

```
NSX-edge-3-0> show ip route
```

Codes: O - OSPF derived, i - IS-IS derived, B - BGP derived,  
C - connected, S - static, L1 - IS-IS level-1, L2 - IS-IS level-2,  
IA - OSPF inter area, E1 - OSPF external type 1, E2 - OSPF external type 2,  
N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2

Total number of routes: 7

O	N2	10.10.10.0/24	[110/1]	via 172.16.0.254	←Web-Tier Network
O	N2	10.10.20.0/24	[110/1]	via 172.16.0.254	←App-Tier Network
O	N2	10.10.30.0/24	[110/1]	via 172.16.0.254	←DB-Tier Network
C		10.66.3.0/24	[0/0]	via 10.66.3.1	
C		169.254.1.0/30	[0/0]	via 169.254.1.1	
C		172.16.0.0/24	[0/0]	via 172.16.0.1	

```
NSX-edge-3-0> show ip ospf neighbors
```

Neighbor ID	Priority	Address	Dead Time	State	Interface
10.66.3.252 ←Leaf 5	1	10.66.3.252	31	Full/BDR	vNic_0
10.66.3.253 ←Leaf 6	1	10.66.3.253	34	Full/DROTHER	vNic_0
172.16.0.254 ←DLR	128	172.16.0.253	39	Full/DR	vNic_1

### 13.1.8.2 Traffic test

**Note:** The ESG Firewall denies external traffic by default and must be configured or temporarily disabled for traffic to pass. Access the ESG firewall by going to **Home > Networking & Security > NSX Edges**. Double-click on the ESG and go to **Manage > Firewall**.

To validate functionality of the ESG, send traffic between a system on the network core/WAN network and the VMs on the NSX network.

For a simplified test, a compute node (running a Windows OS in this example) with an IP address 8.0.0.1/24 and gateway set to 8.0.0.2 is directly connected to Leaf 6, port tengigabitethernet 1/48. The following configuration is added to Leaf 6:

```
interface TenGigabitEthernet 1/48
description To Compute Node
ip address 8.0.0.2/24
no shutdown
exit
```

```
router ospf 1
network 8.0.0.0/24 area 0
```

Provided the compute node, ESG, and VM firewalls are properly configured, the VMs can now be pinged from the compute node connected to the leaf switch.

The compute node at 8.0.0.1 pings Web-VM1 at 10.10.10.11 as follows:

```
C:\Windows\system32>ping 10.10.10.11
```

```
Pinging 10.10.10.11 with 32 bytes of data:
Reply from 10.10.10.11: bytes=32 time<1ms TTL=124
Reply from 10.10.10.11: bytes=32 time<1ms TTL=124
```

A trace route command issued from the compute node at 8.0.0.1 to Web-VM1 at 10.10.10.11 returns the following:

```
C:\Windows\system32>tracert 10.10.10.11
```

```
Tracing route to WIN-U3U892VR1IJ [10.10.10.11]
over a maximum of 30 hops:
```

1	<1 ms	<1 ms	<1 ms	8.0.0.2	←Leaf 6
2	<1 ms	<1 ms	<1 ms	10.66.3.1	←ESG
3	<1 ms	<1 ms	<1 ms	172.16.0.254	←DLR
4	1 ms	<1 ms	<1 ms	WIN-U3U892VR1IJ [10.10.10.11]	←Web-VM1

Trace complete.

## 13.2 Hardware VTEP

For communication between systems on virtual and physical networks within the data center (east-west traffic), a hardware VTEP provides the best performance. It is not considered a best practice to use an ESG for east-west traffic within the data center's leaf-spine network; it can become a bottleneck under heavy loads.

**Note:** The hardware VTEP feature requires an NSX for vSphere Enterprise license.

The switch acting as the hardware VTEP must support VXLAN, such as Dell Networking S4048-ON or S6000-ON switches. This guide uses an S4048-ON.

The hardware VTEP connects upstream to the same two spine switches used in the NSX network. This enables communication between the virtual and physical networks. The added hardware VTEP and physical server are outlined in red in Figure 109:

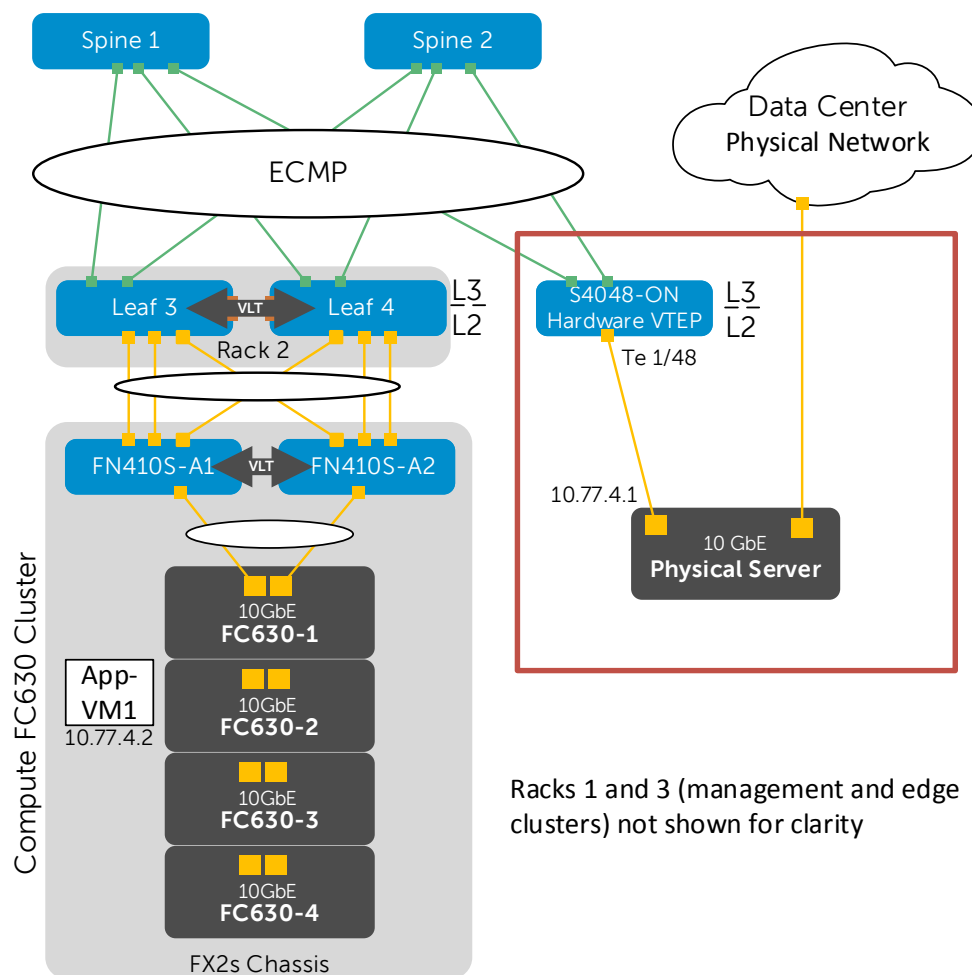


Figure 109 Hardware VTEP and physical server location in leaf-spine network

**Note:** Interface tengigabitethernet 1/48 is connected to a single server in this example. Any available interfaces on the hardware VTEP may be configured and connected to physical servers as needed.

## 13.2.1 Configure additional connections on spine switches

**Note:** These configuration steps are in addition to the spine switch configurations provided in the attachments named spine1.txt and spine2.txt.

On each spine switch, BGP and an additional interface, fortyGigE 1/7/1, are configured to connect to the hardware VTEP as shown in the following two sections.

### 13.2.1.1 Spine 1 – additional configuration steps

```
enable
configure

stack-unit 1 port 7 portmode single speed 40G no-confirm

interface fortyGigE 1/7/1
description To HW VTEP fo1/49
ip address 192.168.1.12/31
mtu 9216
no shutdown

router bgp 64601
neighbor 192.168.1.13 remote-as 64707
neighbor 192.168.1.13 peer-group spine-leaf
neighbor 192.168.1.13 no shutdown

ecmp-group 1
interface fortyGigE 1/7/1

end
write
```

### 13.2.1.2 Spine 2 – additional configuration steps

```
enable
configure

stack-unit 1 port 7 portmode single speed 40G no-confirm

interface fortyGigE 1/7/1
description To HW VTEP fo1/50
ip address 192.168.2.12/31
mtu 9216
no shutdown

router bgp 64602
neighbor 192.168.2.13 remote-as 64707
```



```

neighbor 192.168.2.13 peer-group spine-leaf
neighbor 192.168.2.13 no shutdown

ecmp-group 1
interface fortyGigE 1/7/1

end
write

```

## 13.2.2 Configure the hardware VTEP and connect to NSX

**Note:** The S4048-ON starts at its factory default settings. To reset to factory defaults, see Section 6.1. The switch configuration is provided in the hw-vtep.txt attachment.

Initial configuration involves setting the hostname, enabling LLDP and configuring the management interface and default gateway as follows:

```

enable
configure
hostname HW-VTEP
protocol lldp
advertise management-tlv management-address system-description system-name
advertise interface-port-desc

interface ManagementEthernet 1/1
ip address 100.67.187.36/24
no shutdown
management route 0.0.0.0/0 100.67.187.254

```

Next, configure the upstream layer 3 interfaces connected to the spines. Configure a loopback interface as the router ID for BGP. Complete these actions as follows:

```

interface fortyGigE 1/49
description To Spine-1
ip address 192.168.1.13/31
mtu 9216
no shutdown

interface fortyGigE 1/50
description To Spine-2
ip address 192.168.2.13/31
mtu 9216
no shutdown

interface loopback 0
description Router ID
ip address 10.0.2.7/32

```

Enable the BGP processes to allow routing to the IP fabric. Additionally, create an IP prefix and route map to automatically redistribute all leaf subnets and loopback addresses from the leaf and spine switches as follows:

```
route-map spine-leaf permit 10
match ip address spine-leaf

ip prefix-list spine-leaf
description BGP redistribute loopback and leaf networks
seq 5 permit 10.0.0.0/23 ge 32
seq 10 permit 10.0.0.0/8 ge 24
router bgp 64707
bgp bestpath as-path multipath-relax
maximum-paths ebgp 64
redistribute connected route-map spine-leaf
bgp graceful-restart
neighbor spine-leaf peer-group
neighbor spine-leaf fall-over
neighbor spine-leaf advertisement-interval 1
neighbor spine-leaf no shutdown
neighbor 192.168.1.12 remote-as 64601
neighbor 192.168.1.12 peer-group spine-leaf
neighbor 192.168.1.12 no shutdown
neighbor 192.168.2.12 remote-as 64602
neighbor 192.168.2.12 peer-group spine-leaf
neighbor 192.168.2.12 no shutdown
```

Create an ECMP group and include the interfaces to the two spine switches as follows:

```
ecmp-group 1
interface fortyGigE 1/49
interface fortyGigE 1/50
link-bundle-monitor enable
```

Enable the VXLAN feature and BFD. Create a loopback interface and assign an address to be used as the HW VTEP address as follows:

```
feature vxlan
bfd enable

interface Loopback 77
ip address 10.77.4.254/32
no shutdown
```

Create a VXLAN instance. The gateway address is the hardware VTEP address (same address as loopback 77 above). For controller 1, use the IP address of NSX Controller 1.

**Note:** NSX controller addresses can be determined in the web client by going to **Home > Networking & Security > Installation > Management**.

```
vxlan-instance 1
gateway-ip 10.77.4.254
fail-mode secure
controller 1 100.67.187.183 port 6640 ssl
no shutdown
```

Configure an interface connected to a physical server and place it in the VXLAN instance as follows:

```
interface te 1/48
description To Physical Server
vxlan-instance 1
no shutdown

end
write
```

Create a secure management connection between the S4048-ON and VMware NSX by generating a certificate on the switch:

```
HW-VTEP#crypto cert generate self-signed cert-file flash://vtep-cert.pem key-
file flash://vtep-privkey.pem
```


Generating self signed certificate. This might take a few minutes.

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
Certificate generated successfully.
```

View the certificate from the CLI by running the following command:

```
HW-VTEP#show file flash://vtep-cert.pem
-----BEGIN CERTIFICATE-----
MIIDmTCCAoGgAwIBAgICAKswDQYJKoZIhvcNAQEFBQAwfjELMAkGA1UEBhMCVVMx
HjAcBgNVBAMMFVixODdVMzAtUzQwNDgtUjMtVG9SMjENMAsGA1UECgwERGVSbDEY
MBYGA1UECwwPRGVsbCB0ZXR3b3JraW5nMREwDwYDVQQHDAhTQU4gSm9zZTETMBEG
A1UECAwKQ2FsaWZvcn5pYTAeFw0xNjA5MDMxNzMDMzhaFw0yNjA5MDExNzMDMzha
MH4xCzAJBgNVBAYTAlVTMR4wHAYDVQQDEBVSMTg3VTMwLVMDMDQ4LVIzLVRvUjIx
DTALBgNVBAoMBERlbGwxGDAWBgNVBAsMD0RlbGwzTmV0d29ya2luZzERMA8GA1UE
BwwIU0FOIEpvc2UxEzARBgNVBAGMCkNhbGlm3JuaWEwgGgEiMA0GCSqGSIb3DQEB
AQUAA4IBDwAwggEKAoIBAQC+DF9S0vHVbUv0ZuxY5r08nEqxXiUXYmJCyzhlW06I
LHYs3UBO/dFAgdxPh8ddRNL0zGXNoAYTU1Q6YeIou46xKgriWLCaw1CbK2QluiVn
5DeuvmBd4JcSsSzUj5jCCeX7sdjy3CVhzmL+pHUY1+FDDlyVi9cs5KapOqHHRDI
MDt0ZCFp9q8hdpmt6xfMtD2/Ml7DaUrmymGNNWh3xt+YewkYOBvJuydR2czUosRy
qCUhylIRcB+RhFlsd9kKTqIJNgE7ouxG90na94+KuofOVFcZho3dHSpIUv1fhNz
Q307EfJIIpFufsxDRZWfhrOMtpJka9Qxp1GWCvjPnkPAGMBAAGjITafMB0GA1Ud
DgQWBBAOaPuXmtLDTJVV++VYBiQr9gHCTANBgkqhkiG9w0BAQUFAAOCAQEA09GD
DfipIcOfi+/L011V63x6eXuVaLp1SPAYrgAIxFbepj7sHWWGe2UZixmEhSmY9of
+AkqXY4sC1C4GQER4EaokF3FW0j/n35onNKeXY/78CkJx+lp60qD0DATE7/L5Ew+
OGhXDVCrI5y2Cw4sZ3gYoVXgQasAv8QGRU7tBV67ezdTu6Ur55DlOf+PpHpY8Dl1
fGjdTbzpN90HpZyFDoBaQbXgPuW9riVS8xNUj/M4sPTm81p4GdqgivilMo4CR/6Z
ilwhD2I++gm+VJa0xVnUf1nmjKnpX/YHxpdHvsVxs1ZLdh7uMfGhEsTsrpxqZTX
9IEsmi3+H+O4FW2FGQ==
-----END CERTIFICATE-----
```

Copy the certificate output above including the BEGIN and END CERTIFICATE statements.

1. In the web client, go to **Home > Networking & Security > Service Definitions > Hardware Devices** and click the .
2. In the Edit Hardware Device box:
  - a. Next to **Name**, enter **Hardware VTEP**.
  - b. In the **Certificate** box, paste the certificate output from the S4048-ON as shown in Figure 110.

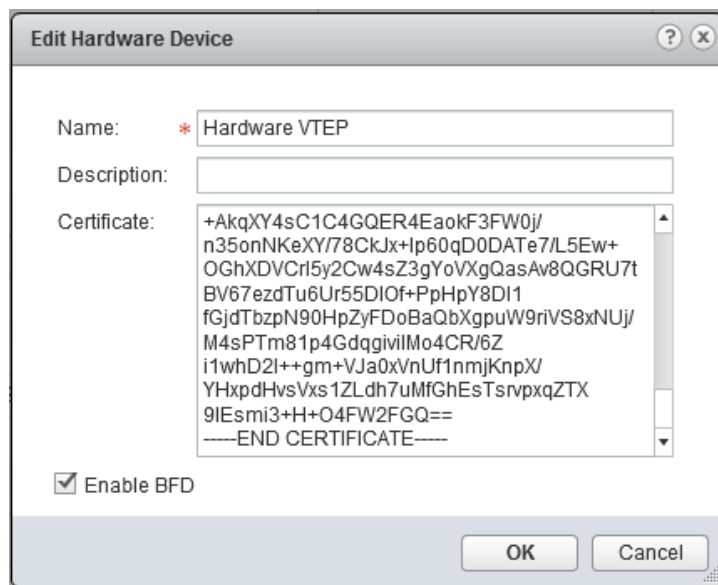


Figure 110 Creating a Hardware Device in VMware NSX

- c. Leave the **Enable BFD** box checked and click **OK**.

After a minute or two, the following output is logged on the S4048-ON:

```
Jan 30 19:42:41: %STKUNIT1-M:CP %OVSDBSVR-5-SESSION_CONNECTED: Instance 1
session 100.67.187.185 is connected
Jan 30 19:42:41: %STKUNIT1-M:CP %OVSDBSVR-5-SESSION_CONNECTED: Instance 1
session 100.67.187.184 is connected
Jan 30 19:42:40: %STKUNIT1-M:CP %OVSDBSVR-5-SESSION_CONNECTED: Instance 1
session 100.67.187.183 is connected
```

This confirms that the hardware VTEP is connected to NSX and an Open vSwitch Database (OVSDB) session is established.

**Note:** The following optional debug command can be issued to view additional VXLAN connection information:

```
S4048#debug vxlan ovbdb-json-rpc packet-type all vxlan-instance 1
```

To confirm that the hardware device has been added to NSX and has proper connectivity, refresh the **Hardware Devices** screen in the web client by clicking the refresh icon (🔄).

The screen should appear similar to Figure 111, with **Connectivity Up** and a **green checkmark** under **BFD Enabled**. The **Management IP Address** shown is the management address of the S4048-ON used as the hardware VTEP.

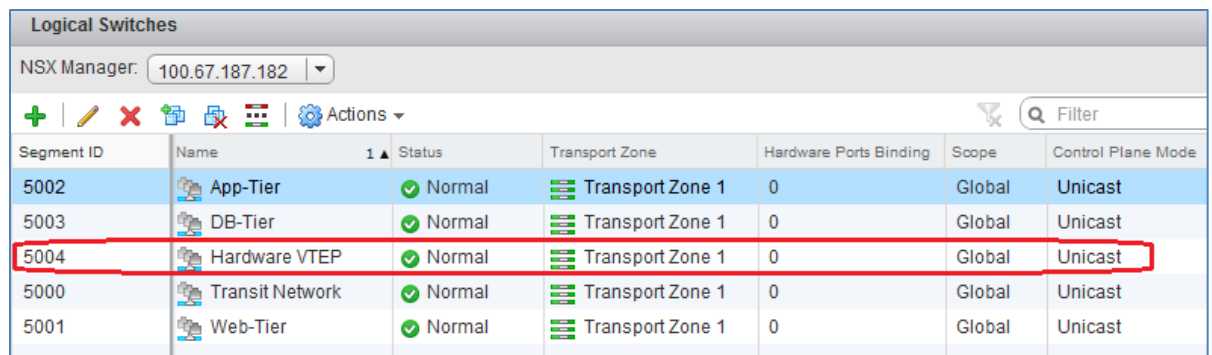
Service Definitions				
<div> <a href="#">Services</a> <a href="#">Service Managers</a> <a href="#">Hardware Devices</a> </div>				
NSX Manager: <input type="text" value="100.67.187.182"/>				
<div> <div>Hardware Devices</div> <div> +   ✎   ✖   🔄   ⚙️ Actions </div> <div>🔍 Filter</div> </div>				
Name	Management IP Address	Connectivity	BFD Enabled	Logical Switches
Hardware VTEP	100.67.187.36	Up	✓	0

Figure 111 Hardware Device Status

### 13.2.3 Create a logical switch

Create a logical switch as follows:

1. In the web client, go to **Home > Networking & Security > Logical Switches**.
2. Click the **+** icon to add a new logical switch.
3. In the **New Logical Switch** dialog box:
  - a. For **Name**, enter **Hardware VTEP**.
  - b. Next to **Transport Zone**, **Transport Zone 1** should already be selected. If not, click **Change** and select it.
  - c. Leave the **Replication mode** set to **Unicast**, **Enable IP Discovery** checked and **Enable MAC Learning** unchecked.
  - d. Click **OK**. A new logical switch named Hardware VTEP is created as shown in Figure 112:



The screenshot shows the 'Logical Switches' page in the NSX Manager. At the top, there's a header 'Logical Switches' and a dropdown for 'NSX Manager' set to '100.67.187.182'. Below the header is a toolbar with icons for adding, editing, deleting, and other actions, along with an 'Actions' dropdown and a search filter. The main table lists logical switches with columns: Segment ID, Name, Status, Transport Zone, Hardware Ports Binding, Scope, and Control Plane Mode. The row for '5004 Hardware VTEP' is highlighted with a red border, indicating it is the newly created switch. Other switches listed include 5002 App-Tier, 5003 DB-Tier, 5000 Transit Network, and 5001 Web-Tier.

Segment ID	Name	Status	Transport Zone	Hardware Ports Binding	Scope	Control Plane Mode
5002	App-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast
5003	DB-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast
5004	Hardware VTEP	✓ Normal	Transport Zone 1	0	Global	Unicast
5000	Transit Network	✓ Normal	Transport Zone 1	0	Global	Unicast
5001	Web-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast

Figure 112 Hardware VTEP logical switch created

4. Select the new logical switch, **Hardware VTEP**, and select **Actions > Manage Hardware Bindings**.
  - a. Expand **Hardware VTEP (0 Bindings)** and click the **+** icon. The IP address of the S4048 is automatically filled out (100.67.187.36 in this example).
  - b. In the **Port** column, click **Select**. Ports that are assigned to vxlan-instance 1 on the S4048-ON switch appear as shown in Figure 113. In this example, it is port **Te 1/48**.

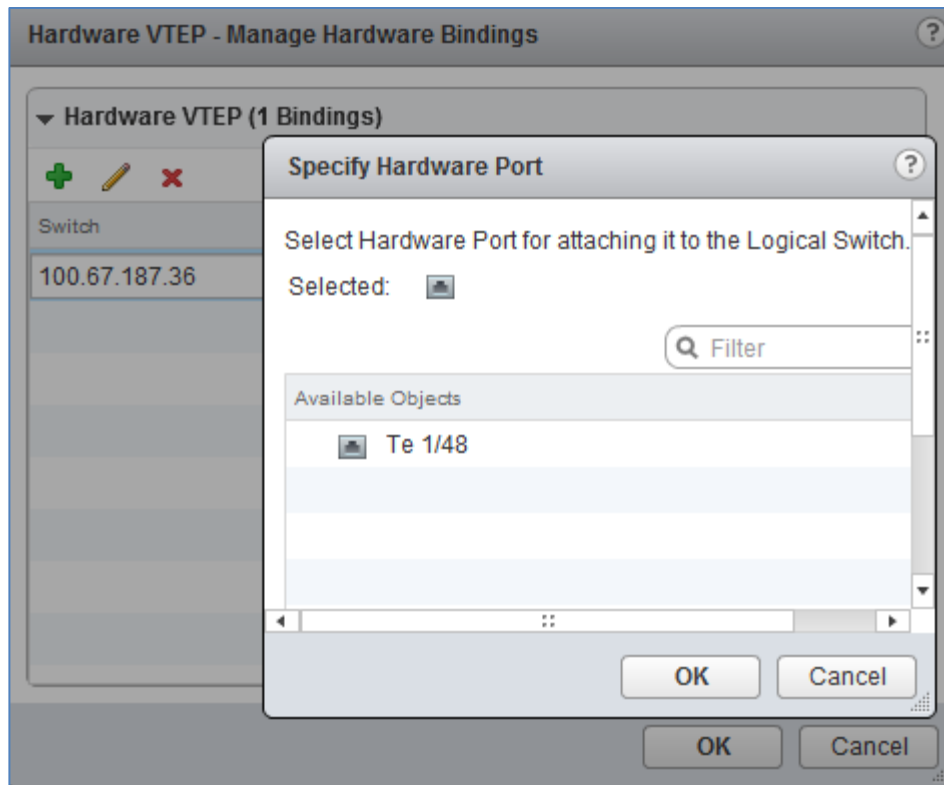


Figure 113 Manage Hardware Bindings – Specify Hardware Port window

- c. Select the port and click **OK**.
- d. Enter **0** in the VLAN box and click **OK**.

When complete, the **Logical Switches** page will look similar to Figure 114. The **Hardware Ports Binding** column indicates one port, Te 1/48 in this example, is configured.

Logical Switches						
NSX Manager: 100.67.187.182						
<div>+</div> <div>Filter</div>						
Segment ID	Name	Status	Transport Zone	Hardware Ports Binding	Scope	Control Plane Mode
5002	App-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast
5003	DB-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast
5004	Hardware VTEP	✓ Normal	Transport Zone 1	1	Global	Unicast
5000	Transit Network	✓ Normal	Transport Zone 1	0	Global	Unicast
5001	Web-Tier	✓ Normal	Transport Zone 1	0	Global	Unicast

Figure 114 Hardware port bound to logical switch

### 13.2.4 Configure a replication cluster

The hardware VTEP is not capable of handling Broadcast, Unknown unicast and Multicast (BUM) traffic and requires at least one NSX-enabled host to process these requests.

On the **Home > Networking & Security > Service Definitions > Hardware Devices** page, next to **Replication Cluster**, click **Edit** and select up to 10 hosts. Only one host is active and the rest serve as backups. Figure 115 shows the four hosts in the compute cluster, Rack 2 Compute FC630, are selected.

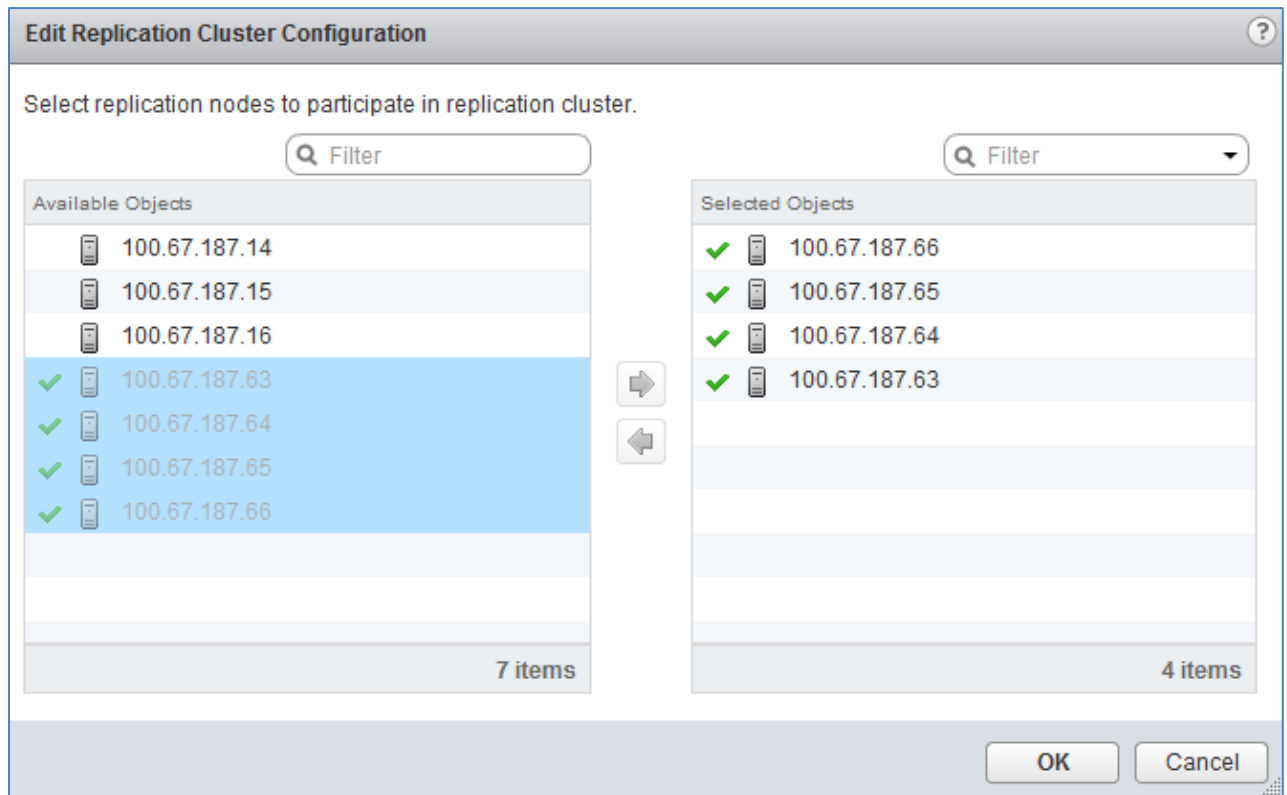


Figure 115 Creating a Replication Cluster

Click **OK** to add the hosts. When complete, the bottom half of the **Hardware Devices** page appear similar to Figure 116. The hosts configured in the replication cluster are shown and BFD is enabled.

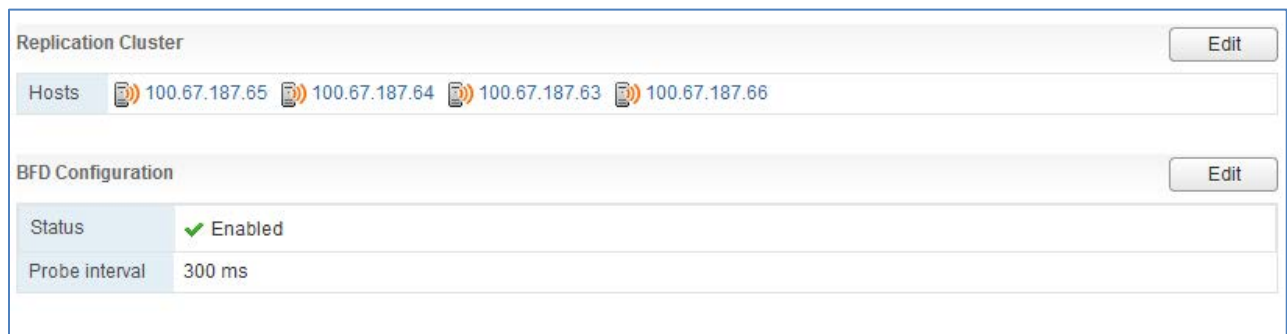


Figure 116 Replication cluster configured



## 13.2.5 Hardware VTEP Validation

### 13.2.5.1 Switch commands and output

Use the following commands and output to verify the hardware VTEP configuration on the S4048-ON.

The `show vxlan vxlan-instance 1` command should return the information shown below. The managers shown in the output below are the three NSX controllers. All three should be connected.

```
HW-VTEP#sh vxlan vxlan-instance 1
Instance           : 1
Mode               : Controller
Admin State        : enabled
Management IP      : 100.67.187.36
Gateway IP         : 10.77.4.254
MAX Backoff        : 30000
Controller 1       : 100.67.187.183:6640 ssl
Managers           :
                   : 100.67.187.183:6640 ssl (connected)
                   : 100.67.187.184:6640 ssl (connected)
                   : 100.67.187.185:6640 ssl (connected)
Fail Mode          : secure
Port List          : Te 1/48
```

Use the `show vxlan vxlan-instance 1 logical-network` command to obtain the logical network name, to be used in the subsequent command.

```
HW-VTEP#show vxlan vxlan-instance 1 logical-network
Instance           : 1
Total LN count     : 1

* - No VLAN mapping exists and yet to be installed
Name               VNID
3202111c-f90e-3c81-aa47-2aaceb72b0df  5004
```

Note that the VXLAN Network Identifier above, 5004 in this example, matches the NSX Hardware VTEP logical switch segment ID shown earlier in Figure 112.

The `show vxlan vxlan-instance 1 logical-network name <name>` command indicates the establishment of a MAC tunnel for each of the four hosts in the replication cluster, along with the software VTEP IP address of each host and the configured hardware port (Te 1/48).

```
HW-VTEP#show vxlan vxlan-instance 1 logical-network name 3202111c-f90e-3c81-aa47-2aaceb72b0df
```

```
Name           : 3202111c-f90e-3c81-aa47-2aaceb72b0df
Description    :
Type           : ELAN
Tunnel Key     : 5004
VFI            : 28674
```

Unknown Multicast MAC Tunnels:

```
10.55.2.1 : vxlan_over_ipv4 (up)
10.55.2.2 : vxlan_over_ipv4 (up)
10.55.2.3 : vxlan_over_ipv4 (up)
10.55.2.4 : vxlan_over_ipv4 (up)
```

Port Vlan Bindings:

```
Te 1/48: VLAN: 0 (0x80000001),
```

The hardware VTEP should be able to ping the IP address of all software VTEPs. The valid software VTEP addresses configured in Section 11.5 are 10.55.2.1-4 (Rack 2 FC630 Compute Cluster) and 10.55.3.1-3 (Rack 3 Edge Cluster). The `gateway-ip` address configured on the hardware VTEP must be specified as the source in the command syntax. The following examples ping one software VTEP from each cluster.

```
HW-VTEP#ping 10.55.2.1 source ip 10.77.4.254
```

```
Sending 5, 100-byte ICMP Echos to 10.55.2.1 from 10.77.4.254, timeout is 2
seconds:
```

```
!!!!
```

```
Success rate is 100.0 percent (5/5), round-trip min/avg/max = 0/0/0 (ms)
```

```
HW-VTEP#ping 10.55.3.3 source ip 10.77.4.254
```

```
Sending 5, 100-byte ICMP Echos to 10.55.3.3 from 10.77.4.254, timeout is 2
seconds:
```

```
!!!!
```

```
Success rate is 100.0 percent (5/5), round-trip min/avg/max = 0/0/0 (ms)
```

### 13.2.5.2 Traffic test

To validate functionality, send traffic between a physical server in the data center (running a Linux or Windows Server operating system for example) and a VM on the NSX network.

Connect the physical server to the configured port on the hardware VTEP (interface `tengigabitethernet 1/48` in this example, shown in Figure 109 at the beginning of this section). The server's network adapter is assigned the address **10.77.4.1/24**.

In the web client, add a second network adapter to App-VM1 in the Rack 2 Compute FC630 cluster and connect it to the Hardware VTEP logical switch as follows:

1. In the web client, go to **Home > Hosts and clusters**.
2. Right click on the VM, **App-VM1**, and click **Edit Settings**.
3. Next to **New device**, select **Network** and click **Add**.
4. Next to **New Network**, expand the drop-down menu and select **Show more networks**. This opens the **Select Network** box shown in Figure 117:

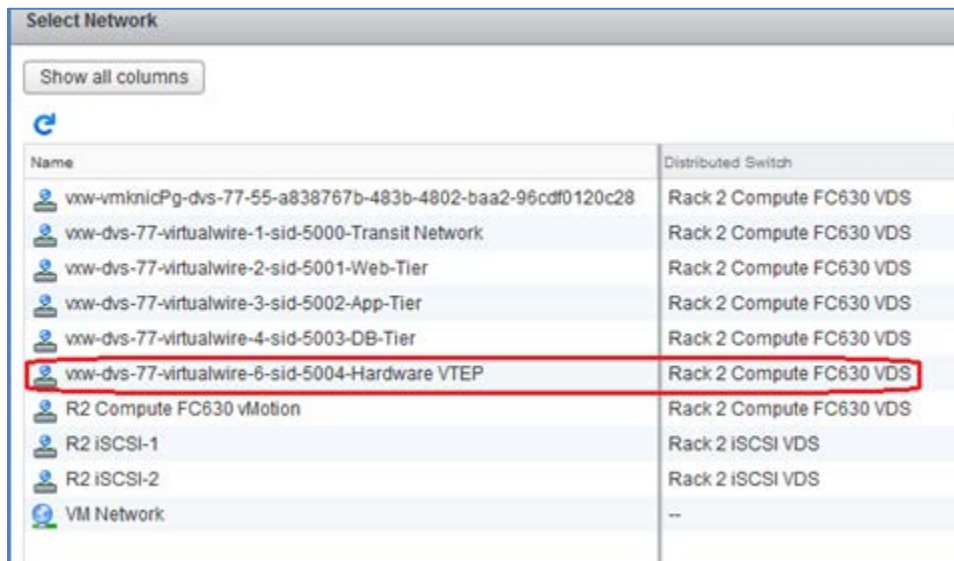


Figure 117 Select Network dialog box

5. Select the virtual wire labeled **Hardware VTEP** and click **OK** to return to the **Edit Settings** box.
6. In the **Edit Settings** box, expand **New Network**. Change the **Adapter Type** from E1000E to **VMXNET3** (since 10GbE adapters are used). Click **OK**.
7. Right click on **App-VM1** and select **Open Console**.
8. Log in to App-VM1 and set the IP address of the newly added adapter to **10.77.4.2/24**. The default gateway is not set or changed for this configuration. App-VM1's original adapter with IP address 10.10.20.11 and gateway remains configured for connectivity with other VMs.

Provided operating system firewalls are properly configured, App-VM1 (10.77.4.2) successfully pings the physical server (10.77.4.1) through the hardware VTEP using the new adapter.

App-VM1 continues to have connectivity to other VMs (Web-VM1, App-VM3, etc.) on their virtual networks (10.10.10.0, 10.10.20.0, etc.) as before.

## 14 Scaling guidance

### 14.1 Switch selection

The leaf layer in this deployment uses the Dell Networking S4048-ON because of its ability to provide a low-latency, non-blocking, layer-2 network architecture. It provides for growth and performance with 48 x 10GbE and six 40GbE ports.

The spine layer uses the Z9100-ON because it provides for substantial growth and outstanding performance with thirty-two 40/100GbE ports per switch. The solution outlined in this document provides scalability out to 16 racks without adding additional spine switches.

### 14.2 iSCSI Storage sizing

The iSCSI storage component in this deployment uses one Dell Storage SC4020 all-in-one array in each compute rack to provide linear scaling as the needs of the deployment grow.

In addition to one SC4020 array, up to seven SC220 storage expansion enclosures can be added to provide a total of 192 drives per rack. SC220 expansion enclosures are optional based on storage requirements of the virtual machines in the rack.

Two additional Dell Networking S4048-ON switches are used per rack to provide two separate iSCSI networks that are isolated from the existing leaf switches.

**Note:** The two HDDs in each individual FC630 blade server are optional and are not included in the storage calculations.

### 14.3 Example – scale out to 3000 virtual machines

The goal of this section is to extend the solution outlined in this deployment guide to accommodate approximately 3000 virtual machines in compute clusters.

This is done as a mathematical exercise and there are many variables to consider when determining hardware requirements for a required number of VMs. To help estimate hardware needs based on the number of VMs required, VM specifications and storage requirements, see the scaling calculator attachment, **scaling\_calc\_fc630\_iscsi.xlsx**.

The following tables include the virtual machine profile, PowerEdge FC630 hardware and storage hardware used in this example:

**Table 13** Virtual machine profile

Virtual CPUs	Virtual Memory (GB)	Virtual Disk Size (GB)
2	8	100

**Table 14** Hardware per PowerEdge FC630

Sockets	Cores per socket	DIMM count	DIMM size (GB)	Total memory (GB)
2	14	8	32	256

**Table 15** SC4020 / SC220 storage hardware

Number of SC4020s per rack (fixed at 1 in calculator)	Number of SC220s per rack	Disk Size (GB)	Disks per SC4020 / SC220	RAID Level
1	2	1200	24	6

The virtual machine requirements in Table 13, the FC630 server hardware in Table 14, and the storage hardware in Table 15 are entered into scaling\_calc\_fc630\_iscsi.xlsx. For the number of VMs required, 3000 is entered into the spreadsheet.

After entering the data above, the **Final Counts** section at the bottom of the spreadsheet indicates a requirement of 30 FX2s chassis containing 121 FC630 servers. These numbers are rounded up to 32 FX2s chassis containing 128 FC630 servers. At eight chassis per rack, 32 FX2s chassis divide equally across four racks.

Table 16 shows the final numbers for the compute clusters in a 3000 VM deployment.

Table 16 3000+ compute node example

FX2s chassis	FC630 servers	Racks	VMs	vCPUs	Total memory (TB)	Usable storage (TB)
32	128	4	3168	7168	32.8	316.8

With four racks for four compute clusters and allowing one rack per management and edge cluster, this 3000 VM solution example uses six racks.

The leaf-spine network for this solution consists of twelve S4048-ON leaf switches (two per rack) and two Z9100-ON spine switches. Its storage network consists of eight S4048-ON iSCSI switches (two per compute rack).

## 14.4 Port count and oversubscription (leaf-spine topology)

The following table outlines the connections for six racks with two spine switches with 40Gb interconnect speeds.

Table 17 Oversubscription Information

	PowerEdge FC630	FN410S IOM (2 per FX2s)	IOM links to leaf switches (8 FX2s per rack)	Leaf links to spine switches per rack	Total links for leaf switches to two spine switches
<b>Connections</b>	2 NIC ports	6 uplink interfaces	8 chassis * 6 = 48 uplinks	2 per leaf switch, 2 leaf switches per rack = 4 links	7 racks * 4 links = 28 uplinks
<b>Port bandwidth</b>	10Gb	10Gb	10Gb	40Gb	40Gb
<b>Total theoretical bandwidth</b>	2 * 10 = 20Gb	60Gb	48 * 10Gb = 480Gb per rack (240Gb per leaf switch)	4 * 40Gb = 160Gb per rack (80Gb per leaf switch)	28 * 40Gb = 1120 Gb

This example provides for an oversubscription rate of 3:1 for 40Gb connectivity. To lower the subscription rate, make additional connections from the leaf switches to the spine switches as needed.

## 14.5 Rack diagrams

Figure 118 shows the management cluster in Rack 1. It includes, from top to bottom, one S3048-ON management switch, two Z9100-ON spine switches, two S4048-ON leaf switches, and three PowerEdge R630 servers. The edge cluster in Rack 3 is identical, with the two spine switches located in either rack.

Adequate space is available to allow for additional spine switches to be added as bandwidth requirements dictate. The management and edge clusters can also be combined in the same rack if preferred.



Figure 118 Rack containing management or edge cluster and spine switches

Figure 119 illustrates a rack containing a compute cluster. It includes, from top to bottom, one S3048-ON management switch, two S4048-ON leaf switches, eight PowerEdge FX2s chassis (containing 32 FC630 servers), two S4048-ON iSCSI SAN switches, three SC220 expansion enclosures and one SC4020 storage array.

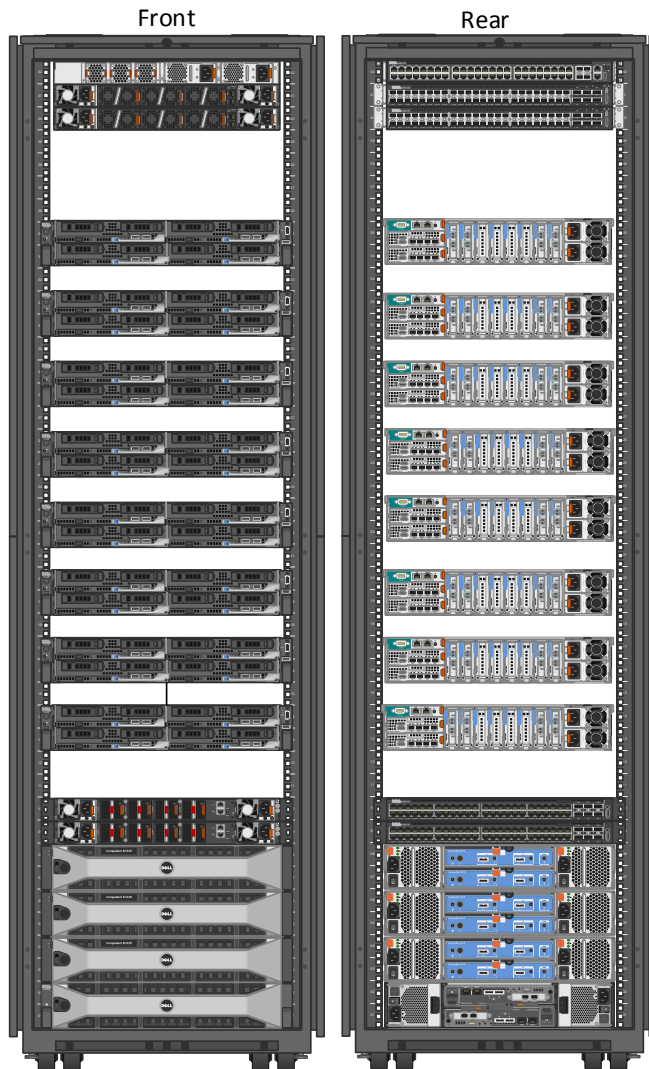


Figure 119 Rack containing compute cluster with storage and switches



## A Dell EMC validated hardware and components

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

### A.1 Switches

Qty	Item	Firmware Version
2	Z9100-ON Spine switch	DNOS 9.11.0.0 P2
6	S4048-ON Leaf switch	DNOS 9.11.0.0 P2
2	S4048-ON iSCSI SAN switch	DNOS 9.11.0.0 P2
1	S4048-ON Hardware VTEP	DNOS 9.11.0.0 P2
3	S3048-ON Management switch	DNOS 9.11.0.0

### A.2 PowerEdge R630 servers

This guide uses six PowerEdge R630 servers, three in the Management cluster and three in the Edge cluster.

Qty per server	Item	Firmware Version
2	Intel Xeon E5-2695 v3 2.3GHz CPU, 14 cores	-
128	GB RAM	-
8	400 GB SAS SSD	-
1	PERC H730 Mini Storage Controller	25.2.1.0037
2	16 GB Internal SD Cards	-
1	QLogic 57840 SFP+ 10GbE QP rNDC (Required for Edge cluster, may substitute with QLogic 57810 SFP+ 10GbE DP rNDC in Mgmt. cluster)	7.12.19
1	Intel I350-T Base-T 1GbE DP PCIe adapter	17.5.10
1	QLogic 57810 SFP+ 10GbE DP PCIe adapter for connection to iSCSI storage (Not required if VSAN or local storage is used).	7.12.19
-	R630 BIOS	2.1.7
-	iDRAC with Lifecycle Controller	2.30.30.30

## A.3 PowerEdge FX2s chassis and components

This guide uses one FX2s chassis with four FC630 servers in the Compute cluster.

Qty per chassis	Item	Firmware Version
1	FX2s Chassis Management Controller	1.32.200
4	FC630 servers. Each server contains: <ul style="list-style-type: none"><li>• 2 - Intel Xeon E5-2695 v3 2.3GHz CPU, 14 cores</li><li>• 8 - 32GB DIMMS (256 GB total)</li><li>• 2 - 300 GB SAS HDD</li><li>• 2 - 16 GB Internal SD Cards</li><li>• 1 - PERC H330 Mini Storage Controller</li><li>• 1 - QLogic 57840 10GbE QP bNDC</li><li>• FC630 BIOS</li><li>• FC630 iDRAC with Lifecycle Controller</li></ul>	<ul style="list-style-type: none"><li>• -</li><li>• -</li><li>• -</li><li>• -</li><li>• 25.4.0.0017</li><li>• 7.12.19</li><li>• 2.1.7</li><li>• 2.30.30.30</li></ul>
2	FN410S IOM	DNOS 9.11.0.0
4	Intel I350-T Base-T 1GbE DP LP PCIe adapter (for Management)	17.5.10
4	QLogic 57810 SFP+ LP PCIe adapter (for iSCSI storage)	7.12.19

## A.4 Dell Storage Center SC4020 storage array

Qty	Item	Firmware Version
2	Storage controllers	6.6.5
2	Chelsio T320 10GbE Dual Port LP iSCSI adapters	07.08.00.00
24	480 GB SAS SSD	-

## B Dell EMC validated software and required licenses

The Software table presents the versions of the software components used to validate the example configurations in this guide. The Licenses section presents the licenses required for the example configurations in this this guide.

### B.1 Software

Item	Version
VMware ESXi	6.0.0 Update 2 - Dell EMC customized image version A00
VMware vSphere Desktop Client	6.0.0 build 3562874
VMware vCenter Server Appliance	6.0.0 Update 2 - build 3634788
vSphere Web Client	6.0.0 build 3617395 (included with VCSA above)
VMware NSX Manager	6.2.4 build 4292526

### B.2 Licenses

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

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

- VMware vSphere 6 Enterprise Plus – 20 CPU sockets
- vCenter 6 Server Standard – 1 instance
- NSX for vSphere Enterprise - 14 CPU sockets

## C Technical support and resources

[Dell.com/support](http://Dell.com/support) is focused on meeting customer needs with proven services and support.

[Dell TechCenter](http://Dell TechCenter) is an online technical community where IT professionals have access to numerous resources for Dell EMC software, hardware and services.

### C.1 Dell EMC product manuals and technical guides

[Manuals and documentation for Dell Networking S3048-ON](#)

[Manuals and documentation for Dell Networking S4048-ON](#)

[Manuals and documentation for Dell Networking Z9100-ON](#)

[Manuals and Documentation for PowerEdge FX2/FX2s and Modules](#)

[Manuals and documentation for PowerEdge R630](#)

[Manuals and documentation for Dell Storage SC4020](#)

[Dell TechCenter Networking Guides](#)

[PowerEdge FX2 – FN I/O Module – VLT Deployment Guide](#)

[Dell EMC NSX Reference Architecture - FC430 Compute Nodes with VSAN Storage](#)

[Dell EMC NSX Reference Architecture - R730xd Compute Nodes with VSAN Storage](#)

### C.2 VMware product manuals and technical guides

[VMware vSphere 6.0 Documentation Center](#)

[VMware NSX 6.2 Documentation Center](#)

[VMware vCenter Server 6.0 Deployment Guide](#)

[VMware Compatibility Guide](#)

[VMware KB Article – Dell Networking VXLAN Hardware Gateway with NSX](#)

## D Support and Feedback

### Contacting Technical Support

Support Contact Information

Web: <http://Support.Dell.com/>

Telephone: USA: 1-800-945-3355

### Feedback for this document

We encourage readers to provide feedback on the quality and usefulness of this publication by sending an email to [Dell\\_Networking\\_Solutions@Dell.com](mailto:Dell_Networking_Solutions@Dell.com).