DELL EMC READY BUNDLE FOR VIRTUALIZATION — WITH VMWARE AND ISCSI INFRASTRUCTURE

Design Guide

APRIL 2017

TABLE OF CONTENTS

Executive Sumamry

This Design Guide outlines the architecture of the Dell EMC Ready Bundle for Virtualization with VMware and iSCSI infrastructure. The Dell EMC Ready Bundle for Virtualization is the most flexible Ready Bundle in the industry. This converged iSCSI solution can be built to include Dell PowerEdge rack servers, PowerEdge FX2 converged platforms, Dell SC Storage SC4020 for mid-tier enterprise storage and Dell SC Storage SCv2020 for entry-level storage. Virtualization with VMware vSphere 6 completes the solution to provide a robust platform for enterprise workloads.

Introduction

The Dell EMC Ready Bundle for Virtualization with VMware is a family of converged and hyper-converged systems that combine servers, storage, networking, and infrastructure management into an integrated and optimized system that provides a platform for general-purpose virtualized workloads. The components undergo testing and validation to ensure full compatibility with VMware vSphere. The Ready Bundle uses infrastructure best practices, and converged server architecture. This document provides an overview of the design and implementation of the Dell EMC Ready Bundle for Virtualization with VMware and iSCSI storage.

Audience and scope

This document is intended for stakeholders in IT infrastructure and service delivery who have purchased or are considering purchase of Dell EMC Ready Bundle for Virtualization with VMware and iSCSI infrastructure. This document provides information and diagrams to familiarize you with the components that make up the bundle.

Supported Configuration

The following table provides an overview of the supported components:

Table 1. Supported Configuration

Components	Entry level configuration	Mid-tier configuration
Server platforms	Choice of: • Dell EMC PowerEdge R630, • Dell EMC PowerEdge R730	Choice of: • Dell EMC PowerEdge R630, • Dell EMC PowerEdge R730, • Dell EMC PowerEdge FX2 with Dell EMC PowerEdge FC630 or Dell EMC PowerEdge FC430
LAN connectivity	(2) Dell Networking S4048-T	(2) Dell Networking S4048-ON
SAN connectivity	(2) Dell Networking S4048T-T	(2) Dell Networking S4048-ON
Out-of-band (OOB) connectivity	None. LAN switch is used for OOB	(1) Dell Networking S3048-ON
Storage array	Dell Storage SCv2020	Dell Storage SC4020
Management server platform	(2) Dell EMC PowerEdge R630	(2) Dell EMC PowerEdge R630
Management software components	• Dell Storage Manager • VMware vCenter Server Appliance • VMware vRealize Automation • VMware vRealize Business • VMware vRealize Log Insight • VMware vRealize Operations Manager with optional Dell Storage Management Pack	• Dell Storage Manager • VMware vCenter Server Appliance • VMware vRealize Automation • VMware vRealize Business • VMware vRealize Log Insight • VMware vRealize Operations Manager with optional Dell Storage Management Pack

Compute server

Ready bundle for Virtualization offers ccustomers a choice of rack servers or modular enclosures for their compute infrastructure. The following table details the supported compute servers:

Table 2. Compute server configurations

Platform model	PowerEdge R630 and PowerEdge R730	PowerEdge FX2 with PowerEdge FC630	PowerEdge FX2 with PowerEdge FC430
Processor	(2) Intel® Xeon® E5 v4 Broadwell processors	(2) Intel® Xeon® E5 v4 Broadwell processors	(2) Intel® Xeon® E5 v4 Broadwell processors
Memory	2400 MT/s RDIMMs	2400 MT/s RDIMMs	2400 MT/s RDIMMs
Network adapter	Intel® X710 quad port 10Gb network adapter or QLogic® 57800 10 Gb quad port network adapter	Intel® X710 dual port 10Gb network adapter or QLogic® 57810 10 Gb dual port network adapter	QLogic® 57810 10 Gb dual port network adapter
Host bus adapter	Intel® X710 dual port 10Gb network adapter or QLogic® 57810 10 Gb dual port network adapter	Intel® X710 dual port 10Gb network adapter or QLogic® 57810 10 Gb dual port network adapter	QLogic® 57810 10 Gb dual port network adapter
Boot device	Internal Dual SD Module with 16 GB SD card for ESXi	Internal Dual SD Module with 16 GB SD card for ESXi	Internal Dual SD Module with 16 GB SD card for ESXi
OOB management	iDRAC8 Enterprise	iDRAC8 Enterprise	iDRAC8 Enterprise
Hypervisor	VMware ESXi 6.5	VMware ESXi 6.5	VMware ESXi 6.5
FX2 chassis configuration	Not applicable	CMC and (2) Dell Networking FN IOAs	CMC and (2) Dell Networking FN IOAs

Management server

The management cluster consist of three PowerEdge R630 servers will the following configuration:

Table 3. Management server configuration

Components	Details
Number of servers	(2) Dell EMC PowerEdge R630
Processor	(2) Intel® Xeon® E5 v4 Broadwell processors
Memory	2400MT/s RDIMMs
Network adapter	Intel® X710 quad port 10Gbnetwork adapter or QLogic® 57800 10 Gb quad port network adapter
Host bus adapter	Intel® X710 dual port 10Gb network adapter or QLogic® 57810 10 Gb dual port network adapter
Boot device	Internal Dual SD Module with 16 GB SD card for ESXi
OOB management	iDRAC8 Enterprise

Design Principles

The following principles are central to the design and architecture of the Dell EMC Ready Bundle for Virtualization. The Ready Bundle for Virtualization is built and validated using these design principles.

- No single point-of-failure: Redundancy is incorporated in the critical aspects[1] of the solution, including server high availability features, networking, and storage.
- Integrated Management: Provide integrated management using vRealize Operations Manager with associated plugins.
- Hardware configuration for virtualization: The system is designed for general use case virtualization. Each server is equipped with appropriate processor, memory, host bus, and network adapters as required for virtualization.
- Best practices adherence: Storage, networking and vSphere best practices of the corresponding components are incorporated into the design to ensure availability, serviceability and optimal performance.
- vRealize Suite Enabled: The system supports VMware vRealize Suite for managing heterogeneous and hybrid cloud.
- Flexible configurations: Ready Bundle for Virtualization can be configured to suit most customer needs for a virtualized infrastructure. The solution supports options, such as configuring server model, server processors, server memory, type of storage (Fibre Channel, iSCSI and SDS), and network switches based on customer needs.

Design Overview

This section provides an overview of the architecture including computer, storage, network, and management architecture. The following figure provides a high-level overview of the architecture, including compute servers (showing flexible compute nodes), management servers, LAN switches, SAN switches, storage arrays and out-of-band switches:

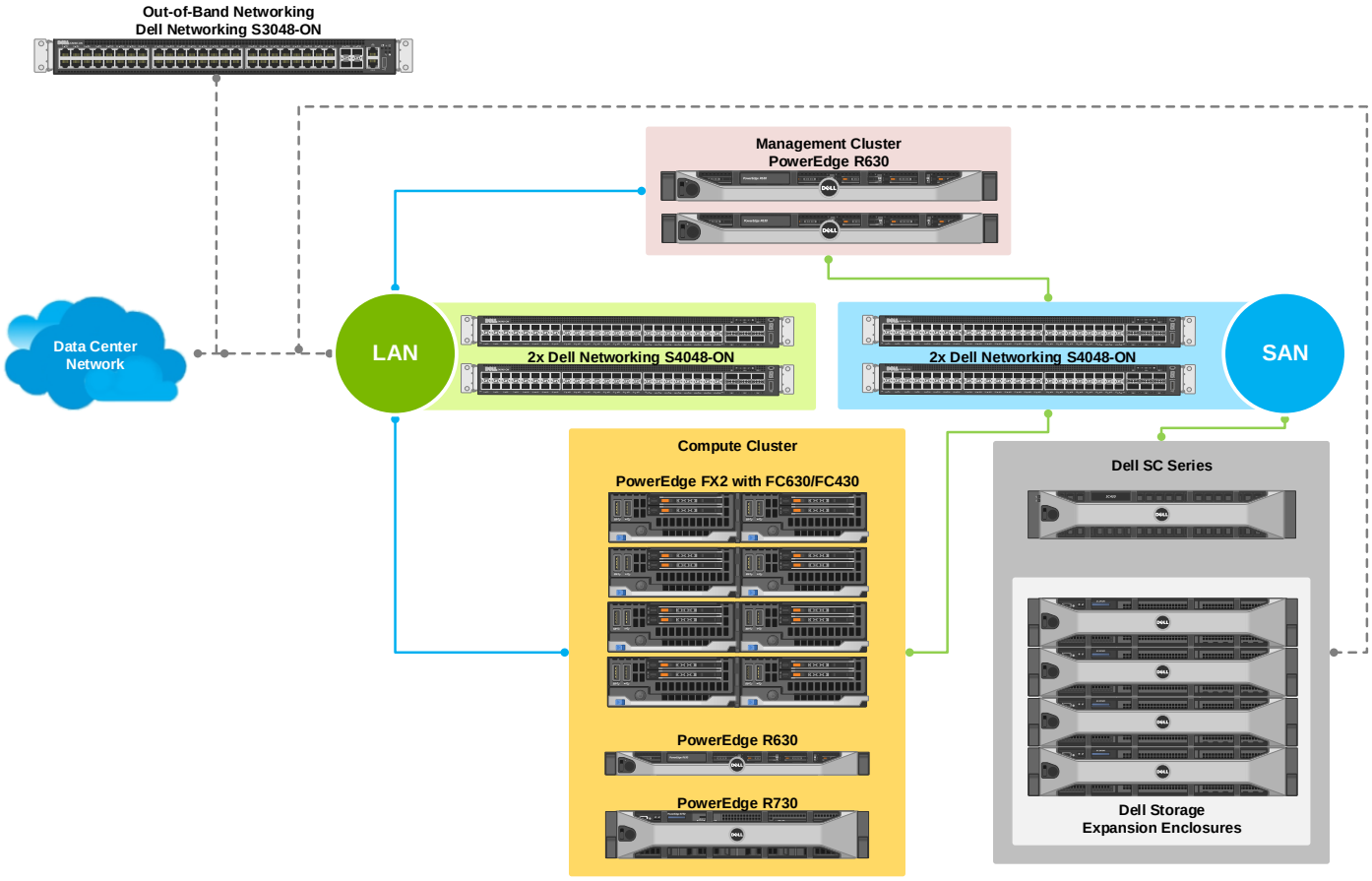

Figure 1. Ready Bundle for Virtualization with VMware and iSCSI infrastructure – Design Overview

[1] Out of band management is not considered critical to user workload and does not have redundancy

Compute Design

The latest Intel® Xeon v4 Broadwell generation processors power the Ready Bundle for Virtualization with VMware. With up to 22 cores per CPU and clock speeds of up to 3.5Ghz in the Dell PowerEdge rack servers you can reduce CPU socket-based licensing costs and achieve greater VM density. The Dell EMC converged FX2 platform provides a powerful balance of resource density and power consumption in a dense form factor. These server nodes are configured with processors that have a lower thermal design power (TDP) value resulting in lower cooling and electrical costs.

The Dell PowerEdge rack platforms support the RDIMM and LRDIMM memory types. Load Reduced DIMM (LRDIMM) uses an iMB buffer to isolate electrical loading from the host memory controller. This buffer and isolation allows for the use of quad ranked DIMM to increase overall memory capacity. For general-purpose virtualization solutions, 2400 MT/s RDIMMs are recommended. Memory can be configured in various modes from within the BIOS. Optimizer mode is the default mode and is recommended for most virtualization use cases to provide the optimized memory performance. For improved reliability and resiliency, other modes such as mirror mode and Dell fault-resilient mode are available.

The Intel® X710 or QLogic® network adapters are supported to provide 10 Gb network connectivity to the top of rack LAN switches. Additional Intel® X710 or QLogic® network adapters are supported to provide 10 Gb network iSCSI connectivity to the SAN switches

Dell PowerEdge servers support various BIOS configuration profiles that control the processor, memory, and other configuration options. Dell recommends the default profile of performance for the Dell EMC Ready Bundle for Virtualization with VMware.

Network Design

This section provides an overview of the network architecture including compute and management server connectivity. Details around the Top-of-Rack (ToR) and virtual switch configuration are provided.

Dell Networking S4048-ON

The network architecture employs Virtual Link Trunking (VLT) connection between the two Top-of-Rack (ToR) switches. In a non-VLT environment, redundancy requires idle equipment which drives up infrastructure costs and increases risks. In a VLT environment, all paths are active, adding immediate value and throughput while still protecting against hardware failures. VLT technology allows a server or bridge to uplink a physical trunk into more than one Dell Networking S4048-ON switch by treating the uplink as one logical trunk. A VLT connected pair of switches acts as a single switch to a connecting bridge or server. Both links from the bridge network can actively forward and receive traffic. VLT provides a replacement for Spanning Tree Protocol (STP) based networks by providing both redundancy and full bandwidth utilization using multiple active paths. Major benefits of VLT technology are:

- Dual control plane for highly available, resilient network services
- Full utilization of the active LAG interfaces
- Active / Active design for seamless operations during maintenance events

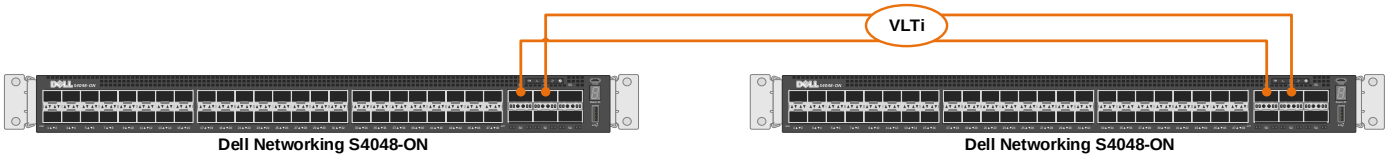

Figure 2. Dell Networking S4048-ON Virtual Link Trunk Interconnect (VLTi) Configuration

The Dell Networking S4048-ON switches each provide six 40 GbE uplink ports. The Virtual Link Trunk Interconnect (VLTi) configuration in this architecture uses two 40 GbE ports from each Top-of-Rack (ToR) switch to provide an 80 Gb data path between the switches. The remaining four 40Gb ports allow for high speed connectivity to spine switches or directly to the data center core network infrastructure. They can also be used to extend connectivity to other racks.

Network connectivity of PowerEdge FX2 Servers

This section describes the network connectivity when using PowerEdge FX2 modular infrastructure blade servers. The compute cluster Dell EMC PowerEdge FC630 and the Dell EMC PowerEdge FC430 server nodes connect to the Dell Networking S4048-ON switches through the Dell EMC PowerEdge FN410S modules shown in the following PowerEdge FX2 blade chassis figure:

- Connectivity between the PowerEdge FX server nodes and the PowerEdge FN410S: The internal architecture of PowerEdge FX2 chassis provides connectivity between an Intel or QLogic® 10 GbE Network Daughter Card (NDC) in each PowerEdge FX2 server node and the internal ports of the PowerEdge FN410S module. The FN410 has eight 10 GbE internal ports. Four Dell EMC PowerEdge FC630 server nodes connect to four of the internal ports in each FN410. Eight Dell EMC PowerEdge FC430 server nodes connect to all eight of the internal ports in each PowerEdge FN410S.

- Connectivity between the Dell EMC PowerEdge FN401S and S4048-ON switches: Two Dell EMC PowerEdge FN410S I/O Aggregators (IOA) in the PowerEdge FX2 architecture provides the Top-of-Rack (ToR) connectivity for the PowerEdge FX2 server nodes. Each IOA provides four external ports. The FN IOA provides a low-touch configuration mode called VLT mode in which Port 9 from each IOA in the chassis forms a Virtual Link Trunk Interconnect (VLTi). Ports 10 and 11 from both the FN IOA form a single port channel which in turn connects to the Top-of-Rack (ToR) switches.

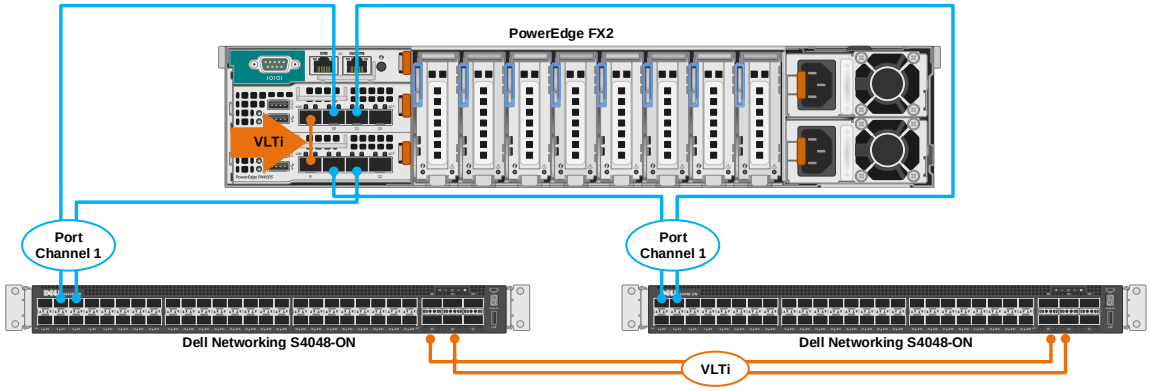

Figure 3. PowerEdge FX2 connectivity

Network configuration of PowerEdge rack servers

The compute cluster can either be Dell EMC PowerEdge FX2 servers or the Dell PowerEdge rack servers. This section describes the network connectivity if rack servers are used for compute servers, and the management servers. The following image is an example of the connectivity between the compute and management Dell PowerEdge rack servers and Dell Networking S4048-ON switches: The compute and management rack servers have two 10 GbE connections to S4048-ON switches through one Intel® or QLogic® dual port 10 GbE network daughter card (NDC).

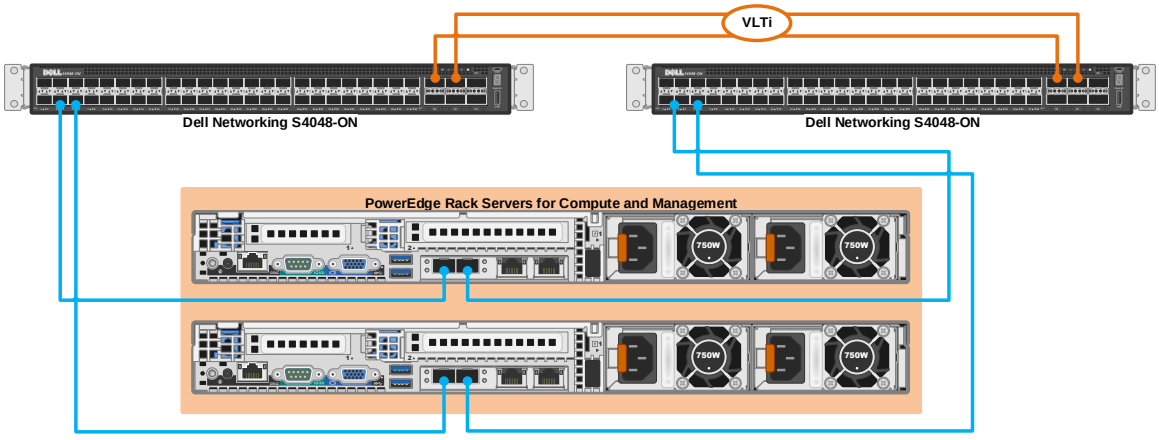

Figure 4. PowerEdge rack server connectivity

Network configuration for LAN traffic with VMware vSphere Distributed Switch (vDS)

Customers can achieve bandwidth prioritization for different traffic classes such as host management, vMotion, and VM network using VMware Distributed Virtual Switches. The VMware vSphere Distributed Switch (vDS) can be configured, managed, and monitored from a central interface and provides:

- Simplified virtual machine network configuration
- Enhanced network monitoring and troubleshooting capabilities
- Support for network bandwidth partitioning when NPAR is not available

The following figures show the virtual distributed switch configuration for the management and compute servers.

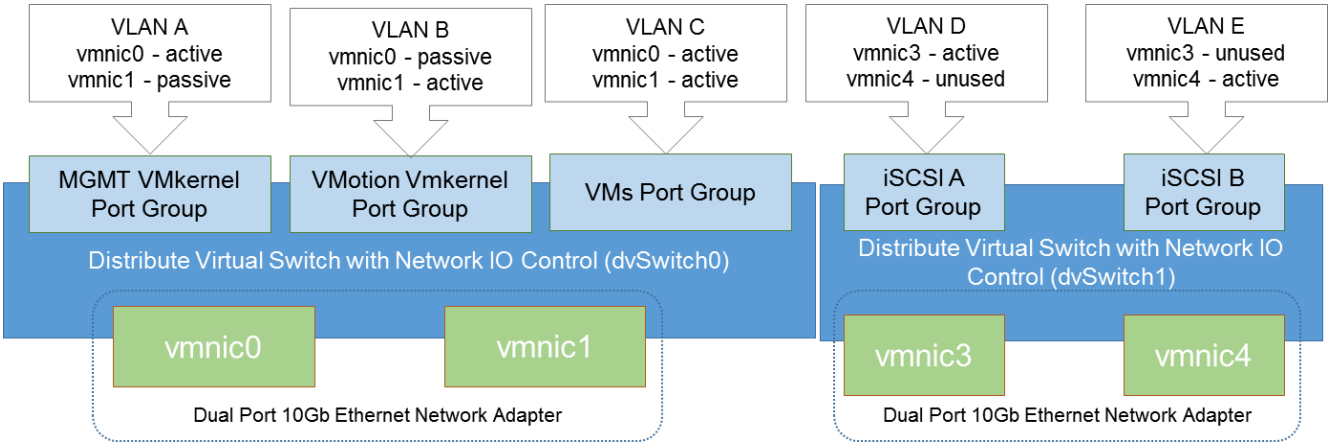

Figure 5. Distributed Virtual Switch for dual port configuration

For configuring iSCSI vSwitch, configure two vmkernal ports in the iSCSI vSwitch and bind the corresponding adapter. Each vmkernal port is bound to a physical adapter and then attached to the software or hardware iSCSI adapter.

Note: When using software iSCSI adapter it is recommended to not to bind port. For more information refer:
https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2038869

Multipathing to an ESXi host is automatic when the server object has more than one HBA or iSCSI initiator posts assigned to it. To configure iSCSI initiator for multipathing consider following key points:

- Verify that there is one VMKernal interface for each physical NIC to be used for storage traffic, and that they follow the virtual switch port binding recommendations mentioned in above link.
- Adjust the failover order on each VMKernal interface for 1:1 VMKernal to physical NIC ratio.
- Add both VMKernal interfaces to the iSCSI adapter network port binding when using hardware iSCSI offload.
- Rescan the iSCSI adapter for the new volumes.
- From within the DSM (Dell Storage Manager) client, create the server object for the host.

For more details refer:
https://pubs.vmware.com/vsphere-65/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-65-storage-guide.pdf

Storage Design

The Ready Bundle uses the Dell Storage Center connected to a pair of 10 Gb Ethernet switches for storage traffic. The Dell SC Storage SC4020 provides a shared storage mechanism to enable a highly available datastore for midtier enterprise storage workloads. Fault domains ensure that data paths are distributed in a way that ensures a component failure does not impact virtual machine operations. The Dell SC Storage SCv2020 series provides the same high availability and redundancy to workloads for entry level enterprise storage. Two Dell Networking S4048-ON or S4048T network switches connect the iSCSI fabric.

Storage fabric configuration

This solution is configured with two fault domains for high availability. For the Dell SC Storage arrays, Port 1 from each controller connects to iSCSI switch 1, while Port 2 connects to iSCSI switch 2, as shown in the following figure: Each iSCSI switch represents a fault domain for the iSCSI storage and enables port failover.

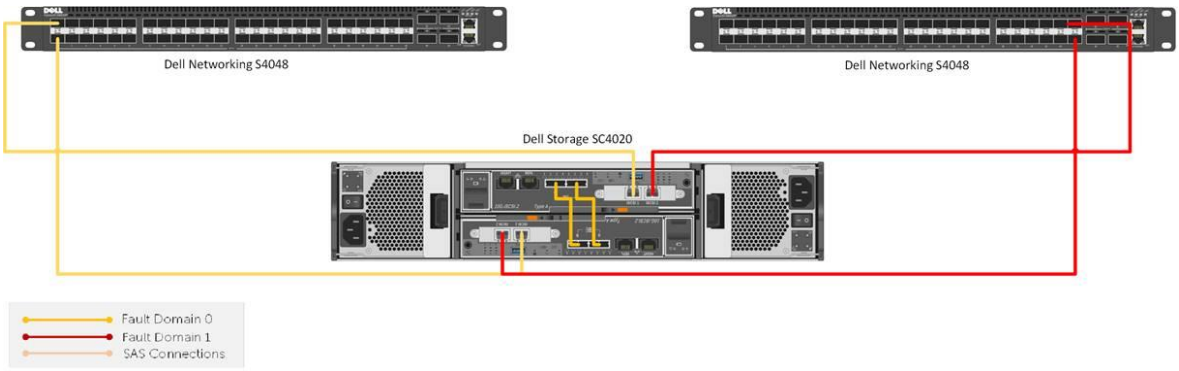

Figure 6. Storage fabric configuration with Dell SC Storage SC4020

Storage connectivity for compute and management servers

Each server has a 10 Gb Ethernet network adapter for connecting to the iSCSI storage fabrics. For Dell FX2 platform servers the PCIe network adapter is inserted in to the rear of the chassis and mapped to the appropriate server node using the chassis controller software. When using the PowerEdge FC630 servers, the PCIE slots 2, 4, 6 and 8 are populated with the 10 Gb network adapter as shown in the diagram below. The PowerEdge FC430 server requires the population of all eight slots.

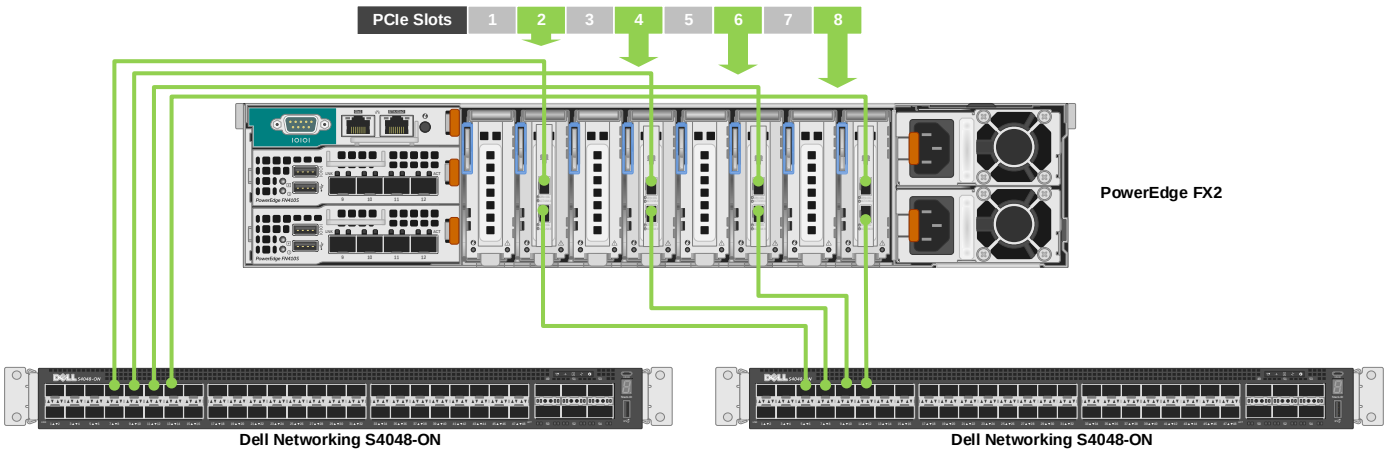

Figure 7. PowerEdge FX2 SAN connectivity

For management servers and rack compute servers, each server is configured with a dual port 10 Gb network adapters with each port connecting to the SAN switch as shown.

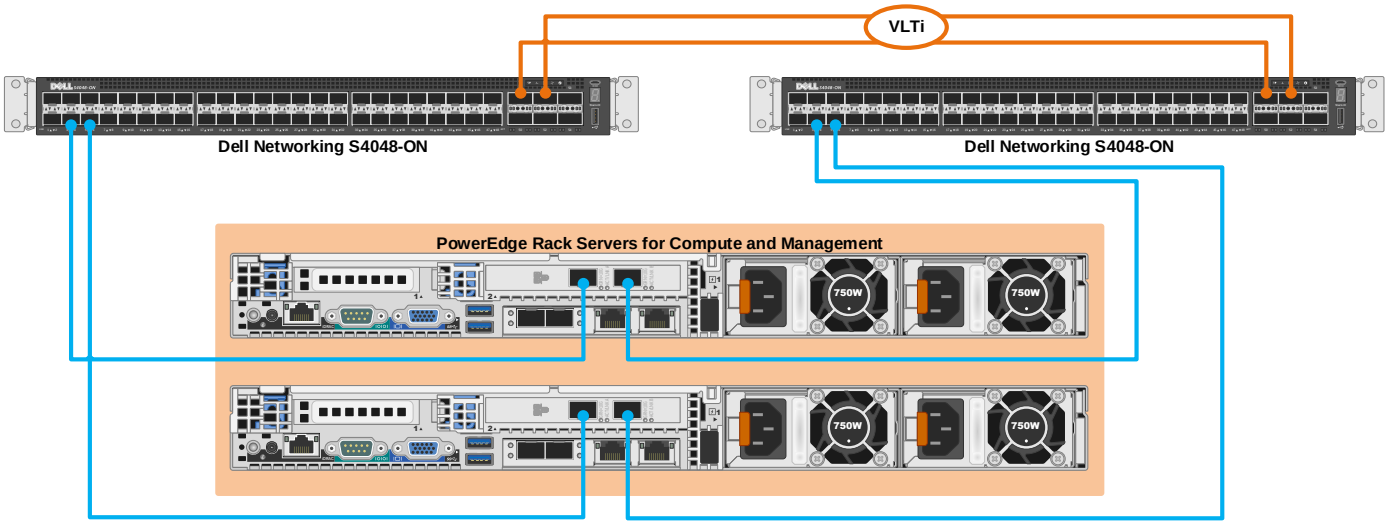

Figure 8. PowerEdge Rack Server SAN connectivity

Storage configuration

Multiple datastores within the VMware vSphere Cluster enable vSphere HA Datastore Heartbeating. This ensures that partitioning, or isolated host networks do not trigger VM movement within the cluster. By default, the vSphere cluster selects up to five datastores for the Datastore Heartbeating.

Management Design

Management Infrastructure

The management infrastructure consists for two R630 servers as part of a management cluster. Management components are virtualized to provide high availability. The bundle further protects these components using the dedicated management cluster. Redundant 10Gb Ethernet uplinks to the network infrastructure and storage fabric, combined with vSphere High Availability ensure that management components stay online. A Dell Networking S3048 switch is used for OOB connectivity. iDRAC ports in each management and compute cluster connect to this switch.

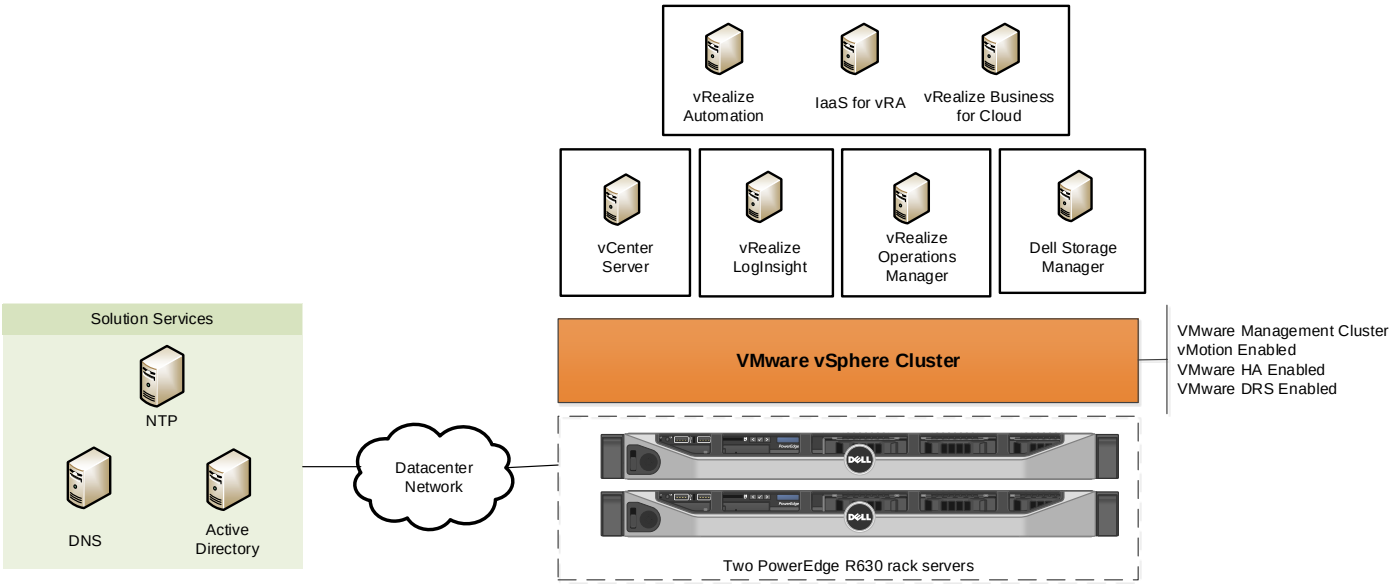

Figure 9. Management Design

Management Components

The following are management components:

- VMware vCenter Server Appliance
- VMware vRealize Automation
- vRealize IaaS
- VMware vRealize Business
- VMware vRealize Log Insight
- VMware vRealize Operations Manager
- Dell Storage Manager

The management software components run on virtual machines contained in your management cluster. The following table lists the management components in the bundle and the VM sizing of those components:

Table 4.

Component	VMs	CPU cores	RAM (GB)	OS (GB)	Data (GB)	NIC
VMware vCenter Server	1	4	12	12	256	1
vRealize Operations Manager	1	4	16	20	254	1
vRealize Automation (vRA)	1	4	18	50	90	1
vRealize Business	1	4	8	50	0	1
vRealize Log Insight	1	4	8	20	511	1
vRealize IaaS (for vRA)	1	4	6	80	0	1
Dell Storage Manager	1	2	8	80	0	1

VMware vRealize Automation Design

The VMware vRealize Automation architecture can be deployed using several models to provide flexibility in resource consumption, availability, and fabric management. The small, medium, and large deployment models all scale up as needed.

Note: Do not use the Minimal Deployment model described in the vRealize Automation documentation. This model is for proof of concept deployments only and cannot scale with the environment.

vRealize Automation uses DNS entries for all components. When scaling, vRealize Automation uses load balancers to distribute the workload across multiple automation appliances, web servers, and infrastructure managers. These components can be scaled as needed and distributed in the same data center or geographically disperse. In addition to scaling for workload, vRealize Automation with load balancers is highly available. This feature ensures that users who consume vRealize Automation XaaS blueprints are not affected by an outage.

vRealize Automation components

The following are the components for vRealize Automation:

- VMware vRealize Automation Appliance - Contains core VRA services and the vRealize Orchestrator service, AD sync connectors, and an internal appliance database. Clusters A/A except DB failover is manual.
- Infrastructure web server - Runs IIS for user consumption of vRealize services and is fully active-active.
- Infrastructure Manager Service - Manages the infrastructure deployed with VRA. For small/medium deployment, this service runs on the Infrastructure Web Server and is active-passive with manual failover.
- Agents - Agents are used for integration with external systems such as Citrix, Hyper-V, and ESXi.
- Distributed Execution Manager - Performs the orchestration tasks necessary. When using multiple DEM worker servers, Distributed Execution Orchestrators are used to assign, track, and manage the tasks give to the workers.

- MSSQL - Tracks infrastructure components, supports Always ON for SQL Server 2016, otherwise a SQL failover cluster is necessary for HA.

VMware vRealize Automation deployment model sizing

VMware vRealize Automation deployment can be small, medium, or large based on the requirements. The following table provides a summary of the three deployment options:

Table 5. vRealize Automation deployment options

	Small	Medium	Large
Managed Machines	10,000	30,000	50,000
Catalog Items	500	1,000	2,500
Concurrent Provisions	10	50	100
vRA Appliances	1	2	2
Windows VMs	1	6	8
MSSQL DB	Single server	Failover cluster	Failover cluster
vRealize Business Appliance	1	1	1
Load Balancers	0	3	3

The following images illustrate a large deployment:

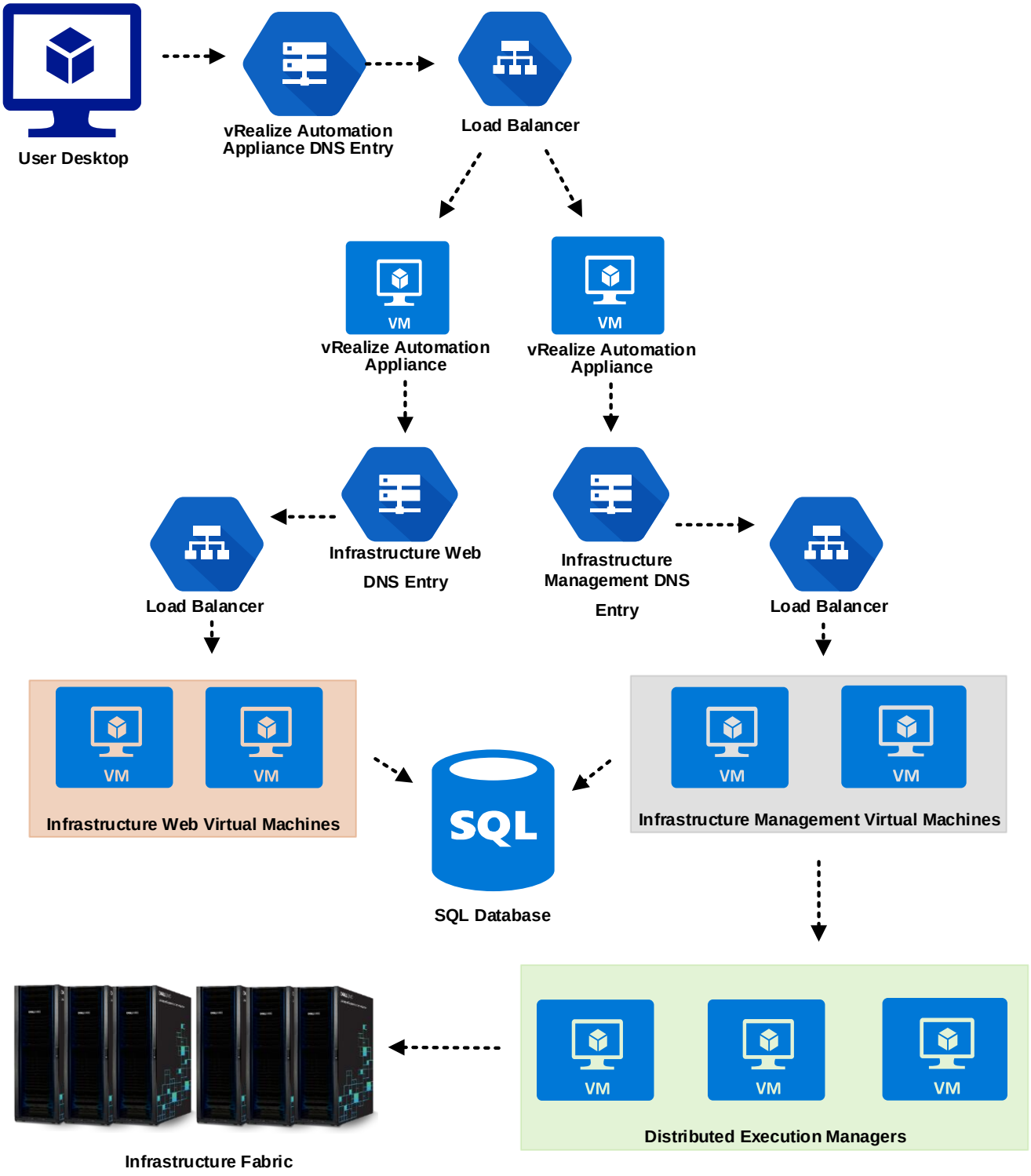

Figure 10. vRealize Automation – Large Deployment

Scaling the Ready Bundle

The solution can be scaled by adding multiple compute nodes (pods) in the customer data center. The Dell Networking Z9100 switch can be used to create a simple yet scalable network. The Z9100 switches serve as the spine switches in the leaf-spine architecture. The Z9100 is a multiline rate switch supporting 10 GbE, 25 GbE, 40 GbE, 50 GbE, and 100 GbE connectivity and can aggregate multiple racks with little or no oversubscription. When connecting multiple racks, using the 40 GbE uplinks from the rack, you can build a large fabric that supports multi-terabit clusters. The density of the Z9100 allows flattening the network tiers and creating an equal-cost fabric from any point to any other point in the network.

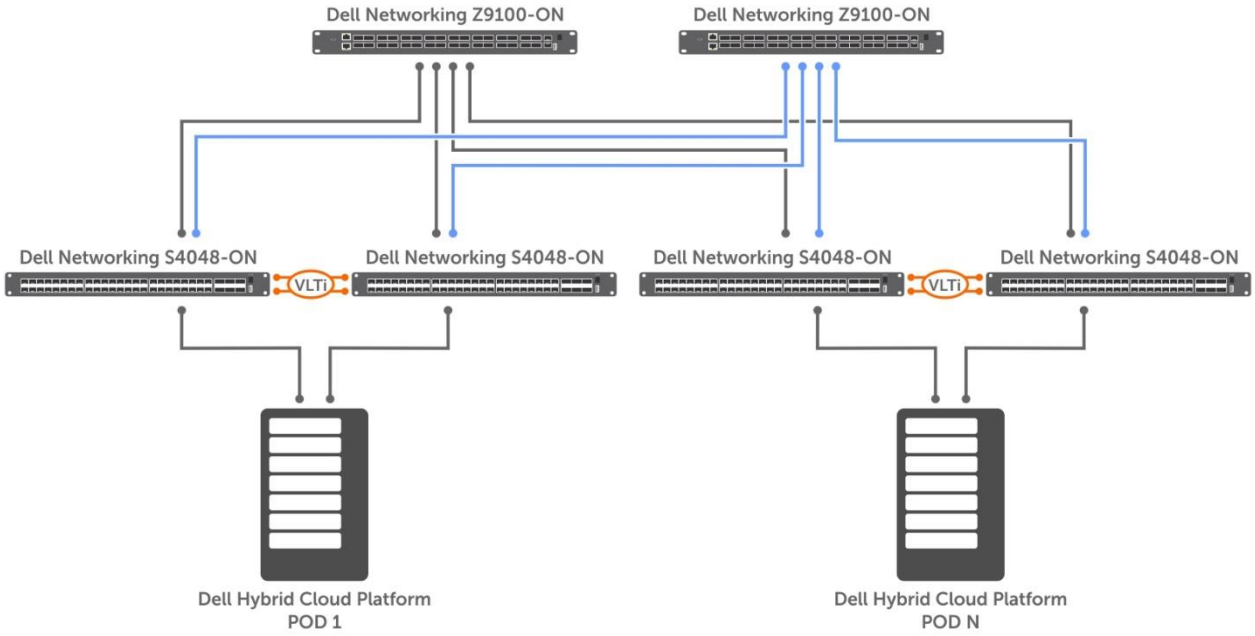

Figure 11. Multiple Compute PODs scaled out using leaf spine architecture

For large domain layer-2 requirements the Extended Virtual Link Trunking (eVLT) can be used on the Z9100, as shown in the following figure. The VLT pair formed can scale in terms of hundreds of servers inside multiple racks. Each rack has four 40 GbE links to the core network providing enough bandwidth for all the traffic between each rack.

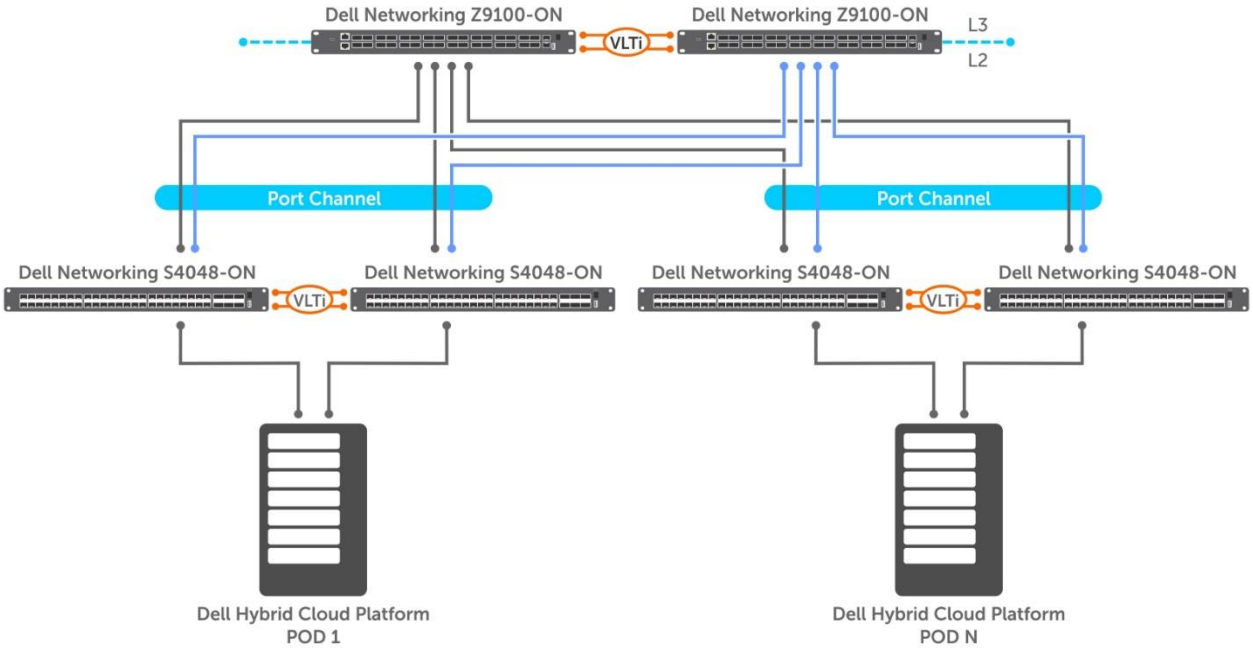

Figure 12. Multiple Compute PODs scaled out using eVLT

References

- Dell EMC SC Series Best Practices with VMware vSphere 5.x–6.x
- vSphere storage for ESXi 6.0
- vRealize Automation – Reference Architecture
- Dell VLT Reference Architecture
- Dell Configuration Guide for S4048-ON