Network Verified Architecture

VMware Validated Design for NetApp HCI

VVD 4.2 Architecture Design

Sean Howard, NetApp
May 2019 | NVA-1128-DESIGN | Version 1.1

Abstract

This document describes the high-level design criteria for a VMware Validated Design (VVD, version 4.2) using VMware cloud-enablement products layered on top of NetApp® HCI components.

vmware READY CERTIFIED PARTNER ARCHITECTURE

NetApp®

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

This VVD architecture and design document describes a validated model for a software-defined data center (SDDC) and provides a detailed design of each management component of the SDDC stack.

The architecture overview in this document discusses the building blocks and the main principles of each SDDC management layer that, when combined, produce a VMware Private Cloud. The detailed design description provides the design options that are compatible with the design objective and the design decisions that are required to build each SDDC component.

NetApp has also developed a NetApp Verified Architecture (NVA) that provides additional guidance on a VMware Private Cloud. This additional verification includes references for supporting automation, third-party plug-ins, expanded discussion on NetApp Element Software benefits, and use case verification for a VMware Private Cloud.

For more information on the NetApp HCI and the VMware Private Cloud see:

- NVA-1122-DEISGN: NetApp HCI for VMware Private Cloud - NVA Design
- NVA-1122-DEPLOY: NetApp HCI for VMware Private Cloud - NVA Deployment

1.1 Intended Audience

This document is intended for cloud architects, infrastructure administrators, and cloud administrators. It assumes that the reader is familiar with and wants to use VMware software to deploy and manage an SDDC that meets the requirements for capacity, scalability, backup and restore, and extensibility for disaster recovery.

2 Required VMware Software

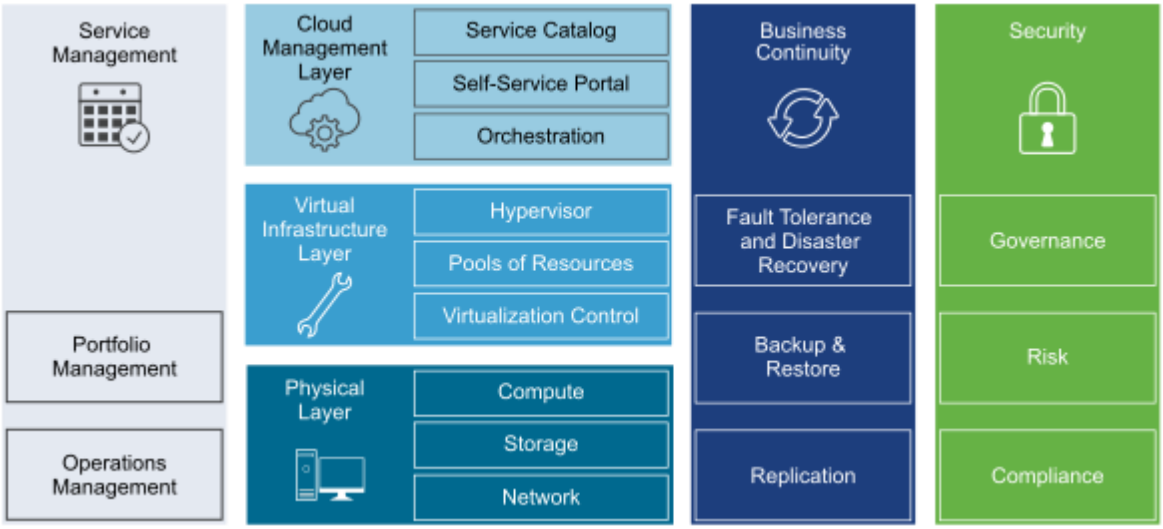
This VVD architecture and design has been validated with certain specific product versions. See the VMware Validated Design Release Notes for full information about supported product versions.

3 Solution Architecture Overview

The SDDC described in this VVD enables an IT organization to automate the provisioning of common repeatable requests and to respond to business needs with agility and predictability. Traditionally this ability has been referred to as infrastructure as a service (IaaS); however, this VVD for an SDDC extends the typical IaaS solution to include a broader and more complete IT solution.

This architecture is based on a number of layers and modules, which allows interchangeable components in the end solution. If a particular component design doesn't fit your business or technical requirements, you should be able to swap it out for a similar component. This VVD has been rigorously tested to provide stability, scalability, and compatibility to help you achieve your desired IT outcome.

Figure 1) Architecture overview.



3.1 Physical Layer

The lowest layer of the solution is the physical layer, sometimes referred to as the core layer, which consists of the compute, network, and storage components. The compute component is composed of x86-based servers that run the management, edge, and tenant compute workloads. This design gives some guidance for the physical capabilities required to run this architecture, but it does not make recommendations for a specific type or brand of hardware.

Note: All components must be supported. See the VMware Compatibility Guide.

3.2 Virtual Infrastructure Layer

The virtual infrastructure layer sits on top of the physical layer components. This layer controls access to the underlying physical infrastructure and controls and allocates resources to the management and tenant workloads. The management workloads consist of elements in the virtual infrastructure layer itself, along with elements in the cloud management, service management, business continuity, and security layers.

3.3 Cloud Management Layer

The cloud management layer is the top layer of the stack. Service consumption occurs at this layer.

This layer calls for resources and orchestrates the actions of the lower layers, most commonly by means of a user interface or application programming interface (API). Although the SDDC can stand on its own without other ancillary services, other supporting components are needed for a complete SDDC experience. The service management, business continuity, and security layers complete the architecture by providing this support.

3.4 Service Management Layer

When building any type of IT infrastructure, portfolio and operations management plays a key role in maintaining continuous day-to-day service delivery. The service management layer of this architecture primarily focuses on operations management and in particular monitoring, alerting, and log management.

3.5 Operations Management Layer

The architecture of the operations management layer includes management components that provide support for the main types of operations in an SDDC. For the microsegmentation use case, you can perform monitoring and logging with vRealize Log Insight.

Within the operations management layer, the underlying physical infrastructure and the virtual management and tenant workloads are monitored in real time. Information is collected in the form of structured data (metrics) and unstructured data (logs). The operations management layer also knows about the SDDC topology (physical and virtual compute, networking, and storage resources), which are key in intelligent and dynamic operational management. The operations management layer consists primarily of monitoring and logging functionality.

3.6 Business Continuity Layer

An enterprise-ready system must contain elements to support business continuity by providing data backup, restoration, and disaster recovery. If data loss occurs, the right elements must be in place to prevent permanent loss to the business. This design provides comprehensive guidance on how to operate backup and restore functions and includes run books with detailed information on how to fail over components in the event of a disaster.

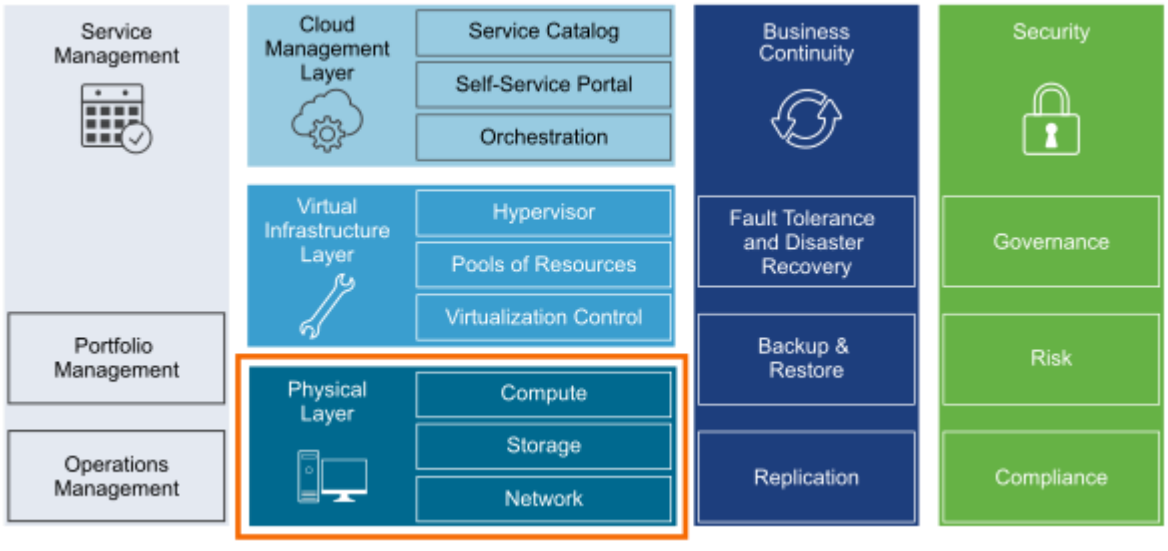
3.7 Security Layer

All systems must be secure by design. A secure design reduces risk and increases compliance while providing a governance structure. The security layer outlines what is needed to make sure that the entire SDDC is resilient to both internal and external threats.

3.8 Physical Infrastructure Architecture

The architecture of the data center physical layer (Figure 2) is based on logical hardware domains and the physical network topology.

Figure 2) Physical infrastructure design.



3.9 Workload Domain Architecture

This VVD for SDDC uses a set of common building blocks called workload domains.

Workload Domain Architecture Characteristics

Workload domains can include different combinations of servers and network equipment that can be set up with varying levels of hardware redundancy and component quality. Workload domains are connected to a network core that distributes data between them. The workload domain is not defined by any physical properties. Rather, it is a standard unit of connected elements in the SDDC.

A workload domain is a logical boundary of functionality, managed by a single vCenter server, for the SDDC platform. Although each workload domain typically spans one rack, it is possible to aggregate multiple workload domains into a single rack in smaller setups. For both small and large setups, homogeneity and easy replication are important.

Different workload domains of the same type can provide different characteristics for varying requirements. For example, one virtual infrastructure workload domain could use full hardware redundancy for each component (power supply through memory chips) for increased availability. At the same time, another virtual infrastructure workload domain in the same setup could use low-cost hardware without any hardware redundancy. These variations make the architecture suitable for the different workload requirements in the SDDC.

Workload Domain-to-Rack Mapping

Workload domains are not mapped one-to-one to data center racks. Although a workload domain is an atomic unit of a repeatable building block, a rack is merely a unit of size. Because workload domains can have different sizes, the way workload domains are mapped to data center racks depends on the use case.

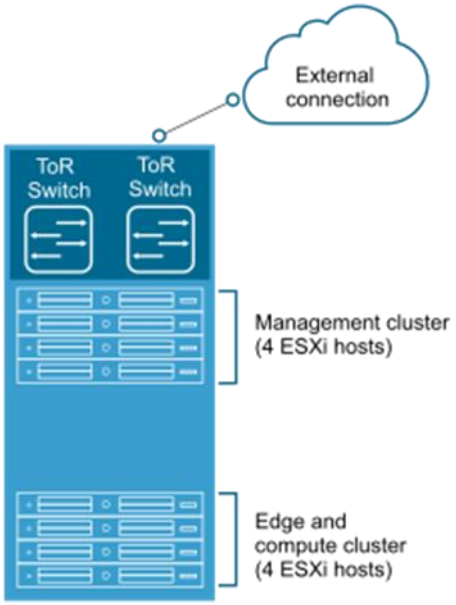
Note: When using a layer 3 network fabric, the management and the shared edge and compute clusters cannot span racks. NSX Controller instances and other virtual machines rely on VLAN-backed networks. The physical network configuration terminates layer 2 networks in each rack at the top-of-rack (ToR) switch. Therefore, you cannot migrate a virtual machine to a different rack because the IP subnet is available only in the rack where the virtual machine currently resides.

- **One workload domain in one rack.** One workload domain can occupy exactly one rack.
- **Multiple workload domains in one rack.** Two or more workload domains can occupy a single rack. For example, one management workload domain and one virtual infrastructure workload domain can be deployed to a single rack.
- **Single workload domain across multiple racks.** A single workload domain can stretch across multiple adjacent racks. For example, a virtual infrastructure workload domain can have more ESXi hosts than a single rack can support.

3.10 Cluster Types

The SDDC differentiates between different types of clusters, including management clusters, compute clusters, edge clusters, and shared edge and compute clusters.

Figure 3) Clusters in the SDDC.



Management Cluster

The management cluster lives in the management workload domain and runs the virtual machines that manage the SDDC. These virtual machines host vCenter Server, vSphere Update Manager, NSX Manager, NSX Controller, vRealize Operations Manager, vRealize Automation, vRealize Log Insight, and other management components. Because the management cluster contains critical infrastructure, consider implementing a basic level of hardware redundancy for this cluster.

Management cluster components must not have tenant-specific addressing.

Shared Edge and Compute Cluster

The shared edge and compute cluster is the first cluster in the virtual infrastructure workload domain; it hosts the tenant virtual machines (sometimes referred to as workloads or payloads). This shared cluster also runs the required NSX services to enable north-south routing between the SDDC tenant virtual machines and the external network, and east-west routing inside the SDDC. As the SDDC expands, you can add more compute-only clusters to support a mix of different types of workloads for different types of service level agreements (SLAs).

Compute Cluster

Compute clusters live in a virtual infrastructure workload domain and host the SDDC tenant workloads. An SDDC can contain different types of compute clusters and provide separate compute pools for different types of SLAs.

External Storage

External storage is delivered through the NFS protocol by NetApp ONTAP® Select appliances. The option to automatically deploy and configure these appliances is presented by the NetApp Deployment Engine.

3.11 Physical Network Architecture

The VVD for SDDC can use most physical network architectures.

Network Transport

You can implement the physical layer switch fabric for an SDDC by offering layer 2 or layer 3 transport services. For a scalable and vendor-neutral data center network, use a layer 3 transport.

This VVD supports both layer 2 and layer 3 transports. When deciding whether to use layer 2 or layer 3, consider the following:

- NSX equal-cost multipath (ECMP) edge devices establish layer 3 routing adjacency with the first upstream layer 3 device to provide equal-cost routing for management and workload virtual machine traffic.
- The investment you have today in your current physical network infrastructure.
- The following benefits and drawbacks for both layer 2 and layer 3 designs.

Benefits and Drawbacks of Layer 2 Transport

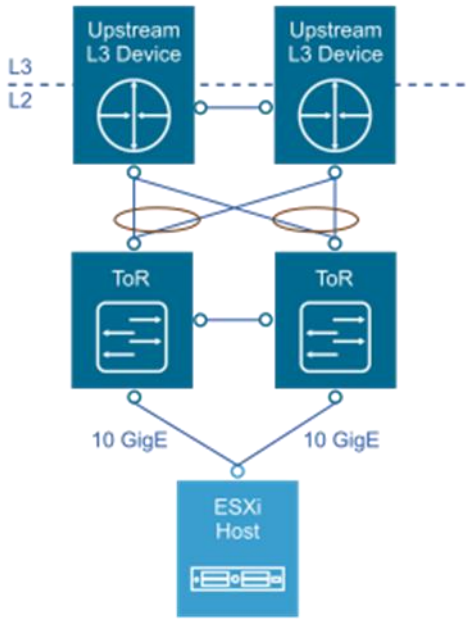
- A design that uses layer 2 transport requires these considerations: In a design that uses layer 2 transport, top-of-rack (ToR) switches and upstream layer 3 devices, such as core switches or routers, form a switched fabric.
- The upstream layer 3 devices terminate each VLAN and provide default gateway functionality.
- Uplinks from the ToR switch to the upstream layer 3 devices are 802.1Q trunks carrying all required VLANs.

Using a layer 2 transport has the benefits and drawbacks described in Table 1.

Table 1) Benefits and drawbacks of layer 2 transport.

Characteristic	Description
Benefits	• More design freedom. • You can span VLANs, which can be useful in some circumstances.
Drawbacks	• The size of such a deployment is limited because the fabric elements have to share a limited number of VLANs. • You might have to rely on a specialized data center switching fabric product from a single vendor.

Figure 4) Example layer 2 transport.



Benefits and Drawbacks of Layer 3 Transport

A design using layer 3 transport requires these considerations:

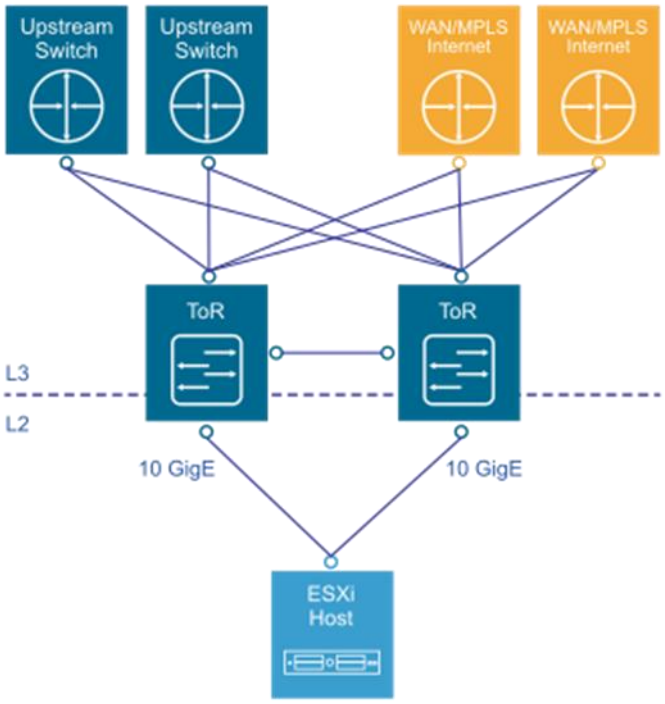
- Layer 2 connectivity is limited in the data center rack up to the ToR switches.
- The ToR switch terminates each VLAN and provides default gateway functionality. That is, it has a switch virtual interface (SVI) for each VLAN.
- Uplinks from the ToR switch to the upstream layer are routed point-to-point links. VLAN trunking on the uplinks is not allowed.
- A dynamic routing protocol, such as OSPF, IS-IS, or the Border Gateway Protocol (BGP), connects the ToR switches and upstream switches. Each ToR switch in the rack advertises a small set of prefixes, typically one per VLAN or subnet. In turn, the ToR switch calculates equal-cost paths to the prefixes it receives from other ToR switches.

Using layer 3 routing has the benefits and drawbacks described in Table 2.

Table 2) Benefits and drawbacks for layer 3 transport.

Characteristic	Description
Benefits	You can choose from a wide array of layer 3-capable switch products for the physical switching fabric. You can mix switches from different vendors because of the general interoperability between implementation of OSPF, IS-IS, or BGP. This approach is typically more cost effective because it makes use of only the basic functionality of the physical switches.
Drawbacks	VLANs are restricted to a single rack. This restriction can affect vSphere fault tolerance and storage networks. To overcome this limitation, use layer 2 bridging in NSX.

Figure 5) Sample layer 3 transport.



Infrastructure Network Architecture

A key goal of network virtualization is to provide a virtual-to-physical network abstraction. To achieve this, the physical fabric must provide a robust IP transport with the following characteristics:

- Simplicity
- Scalability
- High bandwidth
- Fault-tolerant transport
- Support for different levels of quality of service (QoS)

Simplicity and Scalability

Simplicity and scalability are the first and most critical requirements for networking.

Simplicity

Switch configuration in a data center must be simple. General or global configurations such as AAA, SNMP, syslog, NTP, and others should be replicated line by line, independent of the position of the switches. A central management capability to configure all switches at once is an alternative. Restrict configurations that are unique to the switches such as multichassis link aggregation groups, VLAN IDs, and dynamic routing protocol configurations.

Scalability

Scalability factors include, but are not limited to, the following:

- The number of racks supported in a fabric.
- The amount of bandwidth between any two racks in a data center.
- The number of paths between racks.

The total number of ports available across all switches and the oversubscription that is acceptable determine the number of racks supported in a fabric. Different racks might host different types of infrastructure, which can result in different bandwidth requirements.

- Racks with IP storage systems might receive or source more traffic than other racks.
- Compute racks, such as racks hosting hypervisors with virtual machines, might have different bandwidth requirements than shared edge and compute racks, which provide connectivity to the outside world.

Link speed and the number of links vary to satisfy different bandwidth demands. You can vary them for each rack.

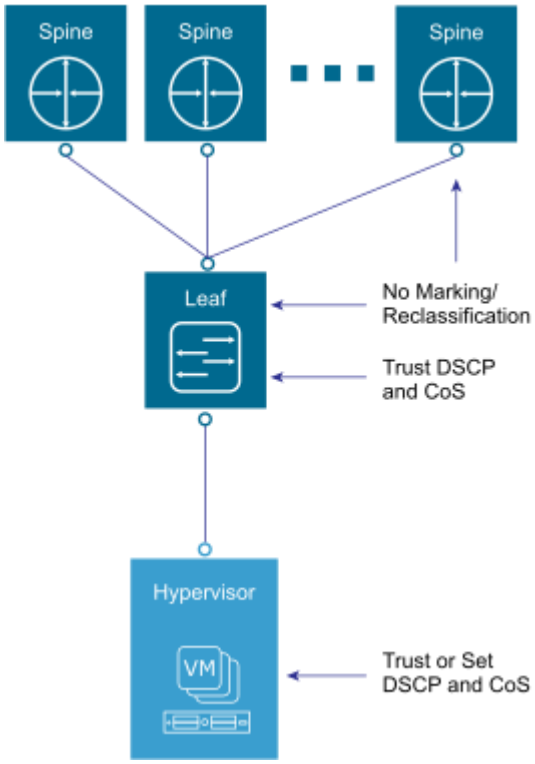
Quality of Service Differentiation

Virtualized environments carry different types of traffic, including tenant, storage, and management traffic, across the switching infrastructure. Each traffic type has different characteristics and makes different demands on the physical switching infrastructure.

- Management traffic, although typically low in volume, is critical for controlling the physical and virtual network state.
- IP storage traffic is typically high in volume and generally stays within a data center.

For virtualized environments, the hypervisor sets the QoS values for the different traffic types. The physical switching infrastructure must trust the values set by the hypervisor. No reclassification is necessary at the server-facing port of a top-of-rack switch. If there is a congestion point in the physical switching infrastructure, the QoS values determine how the physical network sequences, prioritizes, or potentially drops traffic.

Figure 6) Quality of service trust point.



Two types of QoS configuration are supported in the physical switching infrastructure.

- Layer 2 QoS, also called class of service
- Layer 3 QoS, also called DSCP marking

vSphere Distributed Switch supports both class of service and DSCP marking. Users can mark the traffic based on the traffic type or packet classification. When the virtual machines are connected to the VXLAN-based logical switches or networks, the QoS values from the internal packet headers are copied to the VXLAN-encapsulated header. This enables the external physical network to prioritize the traffic based on the tags on the external header.

Physical Network Interfaces

If the server has more than one physical network interface card (NIC) of the same speed, use two as uplinks, with VLANs trunked to the interfaces.

vSphere Distributed Switch supports several NIC teaming options. Load-based NIC teaming supports optimal use of available bandwidth and supports redundancy in case of a link failure. Use two 10GbE connections for each server in combination with a pair of ToR switches. 802.1Q network trunks can support a small number of VLANs; for example, management, storage, VXLAN, vSphere Replication, and VMware vSphere vMotion traffic.

3.12 Availability Zones and Regions

In an SDDC, availability zones are collections of infrastructure components. Regions support disaster recovery solutions and allow you to place workloads closer to your customers. In this design, each region houses a single availability zone.

This VVD uses two regions, with a single availability zone in Region A and single availability zone in Region B.

Figure 7) Availability zones and regions.

Regions

Multiple regions support placing workloads closer to your customers. For example, you could operate one region on the U.S. East Coast and one on the U.S. West Coast, or you could operate one region in Europe and another in the United States.

Regions are helpful in several ways:

- Regions can support disaster recovery solutions. One region can be the primary site and another region can be the recovery site.
- You can use multiple regions to address data privacy laws and restrictions in certain countries by keeping tenant data within a region in the same country.

Although the distance between regions can be rather large, the latency between regions must be less than 150ms.

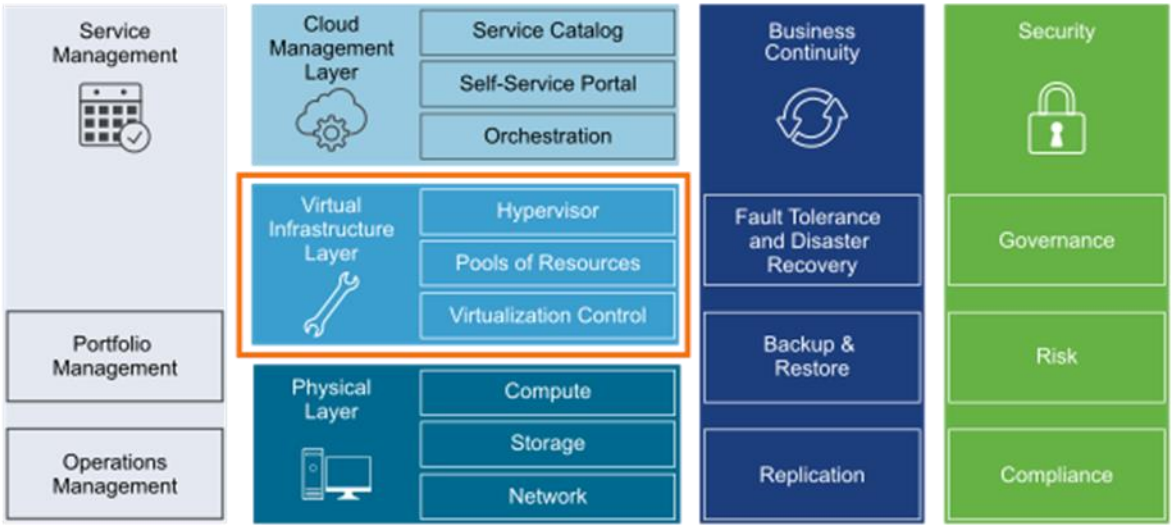
This validated design uses two example regions: Region A in San Francisco (SFO) and Region B in Los Angeles (LAX).

4 Virtual Infrastructure Architecture

The virtual infrastructure is the foundation of SDDC. It contains the software-defined infrastructure, software-defined networking, and software-defined storage. The virtual infrastructure layer runs the operations management layer and the cloud management platform (CMP).

In the virtual infrastructure layer, access to the underlying physical infrastructure is controlled and allocated to the management and tenant workloads. The virtual infrastructure layer consists of the hypervisors on the physical hosts and the control of these hypervisors. SDDC management components consist of elements in the virtual management layer, elements in the cloud management layer, and elements in the operations management, business continuity, and security areas.

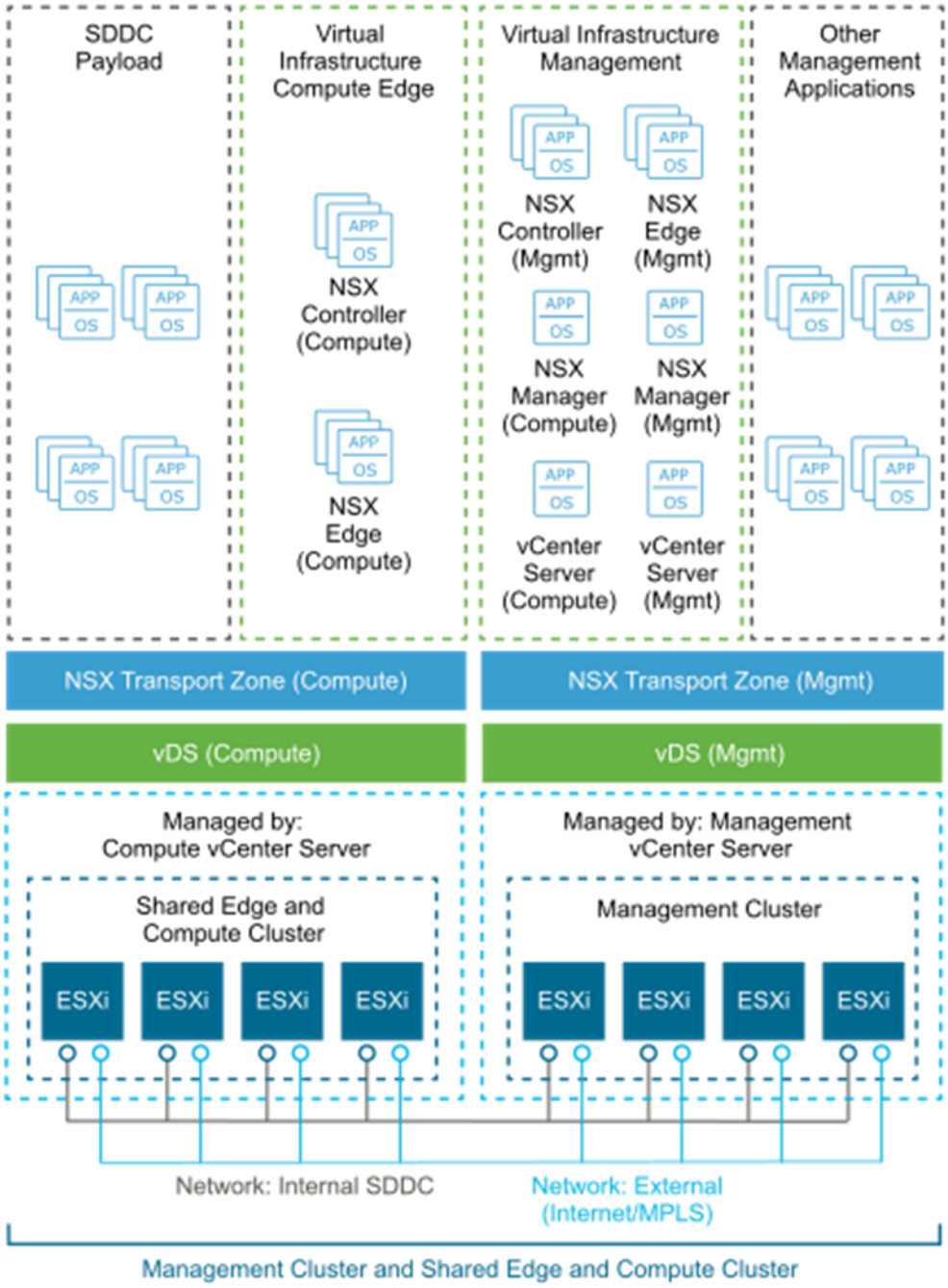
Figure 8) Virtual infrastructure layer in the SDDC.



4.1 Virtual Infrastructure Overview

The SDDC virtual infrastructure consists of two regions. Each region includes a management workload domain that contains the management cluster and a virtual infrastructure workload domain that contains the shared edge and compute cluster.

Figure 9) SDDC logical design.



Management Cluster

The management cluster runs the virtual machines that manage the SDDC. These virtual machines host vCenter Server, vSphere Update Manager, NSX Manager, NSX Controller instances, vRealize Operations Manager, vRealize Log Insight, vRealize Automation, Site Recovery Manager, and other management components. All management, monitoring, and infrastructure services are provisioned to a vSphere cluster that provides high availability for these critical services. Permissions on the management cluster limit access only to administrators. This limitation protects the virtual machines that are running the management, monitoring, and infrastructure services from unauthorized access.

Shared Edge and Compute Cluster

The shared edge and compute cluster runs the following components:

- NSX services that are required for north-south routing between the SDDC tenant workloads and the external network, and east-west routing in the SDDC
- Tenant workloads

As the SDDC expands, you can add more compute-only clusters to support a mix of different types of workloads for different types of SLAs.

4.2 Network Virtualization Components

VMware NSX for vSphere, a network virtualization platform, is a key solution in the SDDC architecture. The NSX for vSphere platform consists of several components that are relevant to the network virtualization design.

NSX for vSphere Platform

NSX for vSphere creates a network virtualization layer. All virtual networks are created on top of this layer, which is an abstraction between the physical and virtual networks. Several components are required to create this network virtualization layer:

- vCenter Server
- NSX Manager
- NSX Controller
- NSX Virtual Switch

These components are separated into different planes to create communications boundaries and to isolate workload data from system control messages.

- **Data plane.** Workload data is contained wholly within the data plane. NSX logical switches segregate unrelated workload data. The data is carried over designated transport networks in the physical network. The NSX vSwitch, distributed routing, and distributed firewall are also implemented in the data plane.
- **Control plane.** Network virtualization control messages are located in the control plane. Control plane communication should be carried on secure physical networks (VLANs) that are isolated from the transport networks used for the data plane. Control messages are used to set up networking attributes on NSX Virtual Switch instances, as well as to configure and manage disaster recovery and distributed firewall components on each ESXi host.
- **Management plane.** The network virtualization orchestration happens in the management plane. In this layer, CMPs such as VMware vRealize Automation can request, consume, and destroy networking resources for virtual workloads. Communication is directed from the CMP to vCenter Server to create and manage virtual machines, and to NSX Manager to consume networking resources.

4.3 Network Virtualization Services

Network virtualization services include logical switches, logical routers, logical firewalls, and other components of NSX for vSphere.

Logical Switches

NSX for vSphere logical switches create logically abstracted segments to which tenant virtual machines can connect. A single logical switch is mapped to a unique VXLAN segment ID and is distributed across the ESXi hypervisors within a transport zone. This allows line-rate switching in the hypervisor without creating constraints of VLAN sprawl or spanning tree issues.

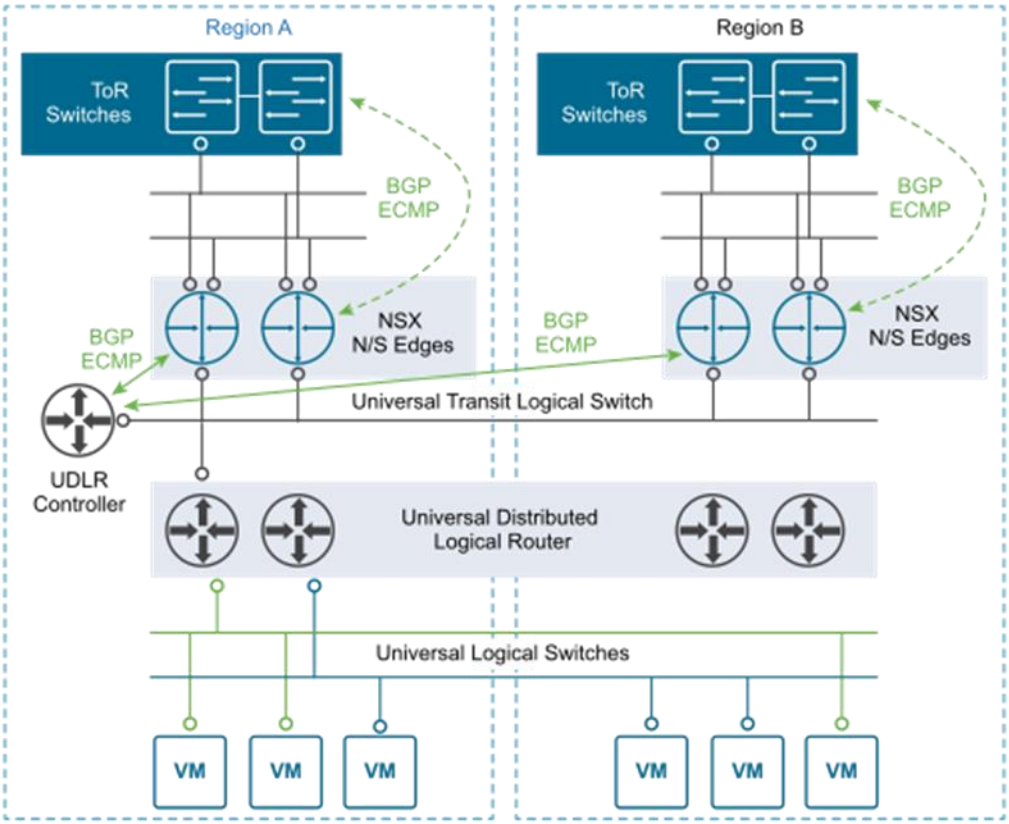
Universal Distributed Logical Router

The universal distributed logical router (UDLR) in NSX for vSphere is optimized for forwarding in the virtualized space (between VMs, on VXLAN-backed or VLAN-backed port groups). UDLR features include the following:

- High-performance, low-overhead first-hop routing
- Scaling the number of hosts
- Support for up to 1,000 logical interfaces (LIFs) on each distributed logical router (DLR)

A UDLR is installed in the kernel of every ESXi host and thus requires a virtual machine to provide the control plane. The control virtual machine of a UDLR is the control plane component of the routing process, providing communication between NSX Manager and an NSX Controller cluster through the User World Agent (UWA). NSX Manager sends LIF information to the Control virtual machine and NSX Controller cluster, and the Control virtual machine sends routing updates to the NSX Controller cluster.

Figure 10) Universal distributed logical routing with NSX for vSphere.



Designated Instance

The designated instance is responsible for resolving ARP on a VLAN LIF. There is one designated instance per VLAN LIF. The selection of an ESXi host as a designated instance is performed automatically by the NSX Controller cluster, and that information is pushed to all other ESXi hosts. Any ARP requests sent by the DLR on the same subnet are handled by the same ESXi host. In case of an ESXi host failure, the controller selects a new ESXi host as the designated instance and makes that information available to the other ESXi hosts.

User World Agent

UWA is a TCP and Secure Sockets Layer (SSL) client that enables communication between the ESXi hosts and NSX Controller nodes. UWA also enables the retrieval of information from NSX Manager through interaction with the message bus agent.

Edge Services Gateway

The UDLR provides virtual machine-to-virtual-machine or east-west routing, and the NSX Edge services gateway (ESG) provides north-south connectivity by peering with upstream top-of-rack switches, thereby enabling tenants to access public networks.

Logical Firewall

The NSX Logical Firewall provides security mechanisms for dynamic virtual data centers.

- The distributed firewall allows you to segment virtual data center entities like virtual machines. Segmentation can be based on virtual machine names and attributes, user identity, vCenter objects like data centers, and ESXi hosts. It can also be based on traditional networking attributes like IP addresses, port groups, and so on.
- The Edge Firewall component helps you meet key perimeter security requirements. These include building DMZs based on IP/VLAN constructs, tenant-to-tenant isolation in multitenant virtual data centers, Network Address Translation (NAT), partner (extranet) VPNs, and user-based SSL VPNs.

The Flow Monitoring feature displays network activity between virtual machines at the application protocol level. You can use this information to audit network traffic, define and refine firewall policies, and identify threats to your network.

Logical Virtual Private Networks (VPNs)

SSL VPN-Plus allows remote users to access private corporate applications. IPSec VPN offers site-to-site connectivity between an NSX Edge instance and remote sites. L2 VPN enables you to extend your data center by allowing virtual machines to retain network connectivity across geographical boundaries.

Logical Load Balancer

The NSX Edge load balancer enables network traffic to follow multiple paths to a specific destination. It distributes incoming service requests evenly among multiple servers in such a way that the load distribution is transparent to users. Load balancing thus helps to achieve optimal resource utilization, maximizing throughput, minimizing response time, and avoiding overload. NSX Edge provides load balancing up to layer 7.

Service Composer

Service Composer helps you provision and assign network and security services to applications in a virtual infrastructure. You map these services to a security group, and the services are applied to the virtual machines in the security group.

NSX Extensibility

VMware partners integrate their solutions with the NSX for vSphere platform to enable an integrated experience across the entire SDDC. Data center operators can provision complex, multitier virtual networks in seconds, independent of the underlying network topology or components.

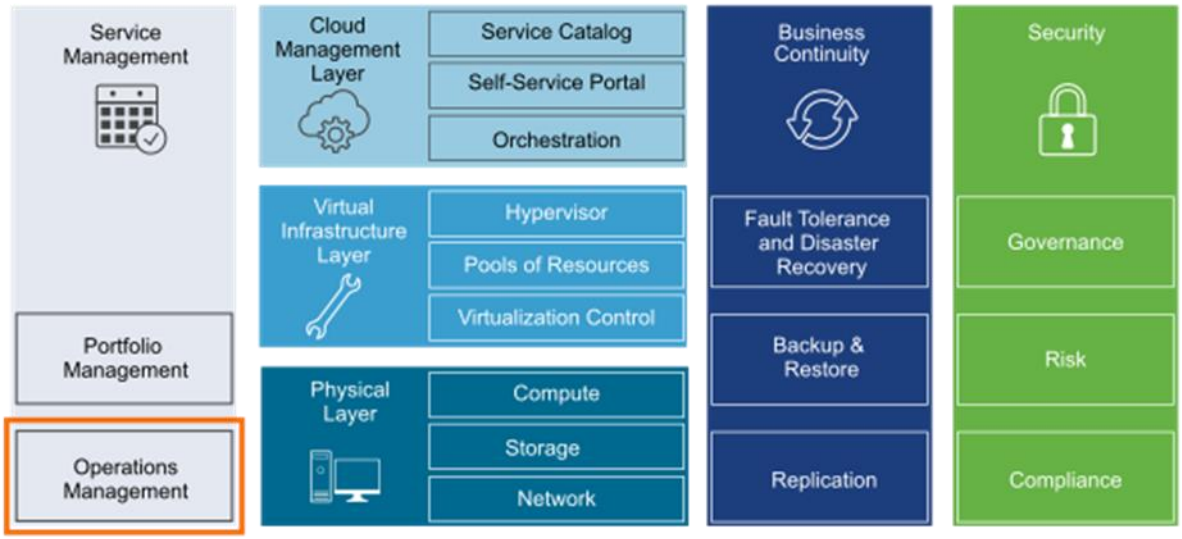
5 Operations Management Architecture

The architecture of the products of the operations management layer supports centralized monitoring of data and logging data about the other solutions in the SDDC. You use this architecture to deliver core operational procedures in the data center.

In the operations management layer, the physical infrastructure, virtual infrastructure, and tenant workloads are monitored in real time, collecting the following information for intelligent and dynamic operational management:

- Monitoring data, such as structured data (metrics) and unstructured data (logs)
- Topology data, such as physical and virtual compute, networking, and storage objects

Figure 11) Operations management layer of the SDDC.



5.1 Monitoring Architecture

vRealize Operations Manager tracks and analyzes the operation of multiple data sources in the SDDC by using specialized analytic algorithms. These algorithms help vRealize Operations Manager learn and predict the behavior of every object it monitors. Users access this information by using views, reports, and dashboards.

Deployment

vRealize Operations Manager is available as a preconfigured virtual appliance in Open Virtual Machine Format (OVF). By using the virtual appliance, you can easily create vRealize Operations Manager nodes with identical predefined sizes.

You deploy the OVF file of the virtual appliance once for each node. After node deployment, you access the product to set up cluster nodes according to their role and log in to configure the installation.

Deployment Models

You can deploy vRealize Operations Manager as a virtual appliance in one of the following configurations:

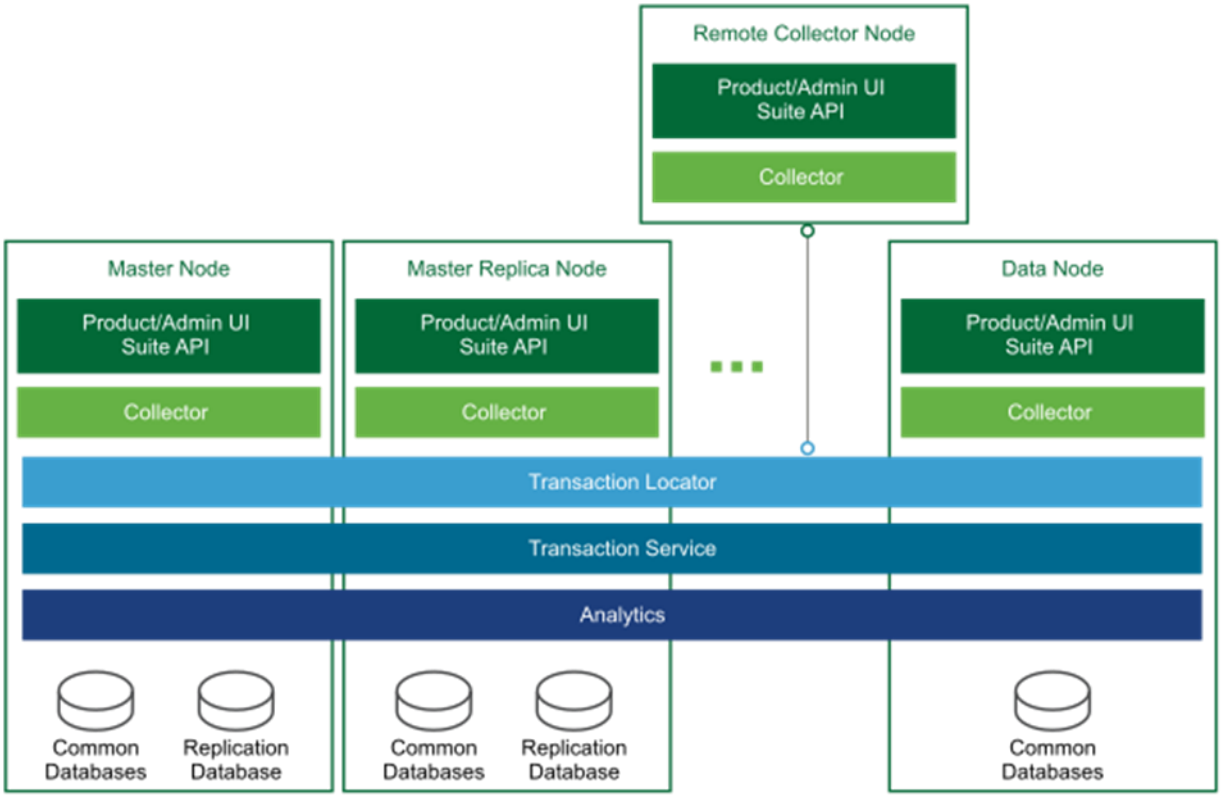
- Standalone node
- Cluster of one master and at least one data node, and optionally a group of remote collector nodes. You can establish high availability by using an external load balancer.

The compute and storage resources of the vRealize Log Insight instances can scale up as growth demands.

Architecture

vRealize Operations Manager contains functional elements that collaborate for data analysis and storage and that also support creating clusters of nodes with different roles.

Figure 12) vRealize Operations Manager architecture.



Types of Nodes

For high availability and scalability, you can deploy several vRealize Operations Manager instances in a cluster to track, analyze, and predict the operation of monitored systems. Cluster nodes can have one of the following roles:

- **Master node.** Required initial node in the cluster. In large-scale environments, the master node manages all other nodes. In small-scale environments, the master node is the single standalone vRealize Operations Manager node.
- **Master replica node.** Optional. Enables high availability of the master node.
- **Data node.** Optional. Enables scale out of vRealize Operations Manager in larger environments. Data nodes have adapters installed to perform collection and analysis. Data nodes also host vRealize Operations Manager management packs.
- **Remote collector node.** Overcomes data collection issues across the enterprise network, such as limited network performance. You can also use remote collector nodes to offload data collection from the other types of nodes.

Remote collector nodes only gather statistics about inventory objects and forward collected data to the data nodes. They do not store data or perform analysis. The master and master replica nodes are data nodes that have extended capabilities.

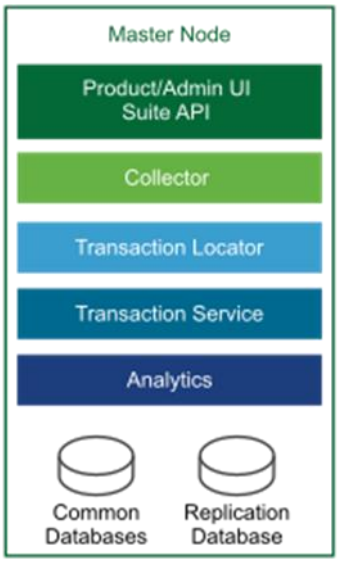
Types of Node Groups

- **Analytics cluster.** Tracks, analyzes, and predicts the operation of monitored systems. Consists of a master node, data nodes, and optionally of a master replica node.
- **Remote collector group.** Because it consists of remote collector nodes, only collects diagnostics data without storage or analysis. A vRealize Operations Manager deployment can contain several collector groups.
- Use collector groups to achieve adapter resiliency in case the collector experiences network interruption or becomes unavailable.

Application Functional Components

The functional components of a vRealize Operations Manager instance interact with each other to analyze diagnostics data from the data center and visualize the result in the web user interface.

Figure 13) Architecture of a vRealize Operations Manager node.



The components of a vRealize Operations Manager node perform these tasks:

- **Product/Admin UI and Suite API.** The UI server is a web application that serves as both a user and an administration interface, and it also hosts the API for accessing collected statistics.
- **Collector.** The collector collects data from all components in the data center.
- **Transaction locator.** The transaction locator handles the data flow between the master, master replica, and remote collector nodes.
- **Transaction service.** The transaction service is responsible for caching, processing, and retrieving metrics for the analytics process.
- **Analytics.** The analytics engine creates all associations and correlations between various datasets, handles all super-metric calculations, performs all capacity planning functions, and is responsible for triggering alerts.
- **Common databases.** Common databases store the following types of data that are related to all components of a vRealize Operations Manager deployment:

  - Collected metric data
  - User content, metric key mappings, licensing, certificates, telemetry data, and role privileges
  - Cluster administration data
  - Alerts and alarms, including the root cause, and object historical properties and versions
- **Replication database.** The replication database stores all resources, such as metadata, relationships, collectors, adapters, and collector groups, and the relationships between them.

Authentication Sources

You can configure vRealize Operations Manager user authentication to use one or more of the following authentication sources:

- vCenter Single Sign-On
- VMware Identity Manager
- OpenLDAP via LDAP
- Active Directory via LDAP

Management Packs

Management packs contain extensions and third-party integration software. They add dashboards, alert definitions, policies, reports, and other content to the inventory of vRealize Operations Manager. You can learn all about management packs, and download them, from VMware Solutions Exchange.

Backup

You back up each vRealize Operations Manager node by using traditional virtual machine backup solutions that are compatible with VMware vSphere Storage APIs – Data Protection (VADP).

Multiregion vRealize Operations Manager Deployment

The scope of this validated design can cover multiple regions, with a single availability zone per region.

This VVD for SDDC implements a large-scale vRealize Operations Manager deployment across multiple regions by using the following configuration:

- A load-balanced analytics cluster that runs multiple nodes is protected by Site Recovery Manager to fail over across regions.
- Multiple remote collector nodes that are assigned to a remote collector group in each region handle data coming from management solutions.

5.2 Logging Architecture

vRealize Log Insight provides real-time log management and log analysis with machine-learning-based intelligent grouping, high-performance searching, and troubleshooting across physical, virtual, and cloud environments.

Overview

vRealize Log Insight collects data from ESXi hosts by using the syslog protocol. vRealize Log Insight has the following capabilities:

- Connects to other VMware products, like vCenter Server, to collect events, tasks, and alarm data.
- Integrates with vRealize Operations Manager to send notification events and enable launch in context.
- Functions as a collection and analysis point for any system that is capable of sending syslog data.

To collect additional logs, you can install an ingestion agent on Linux or Windows servers, or you can use the preinstalled agent on certain VMware products. Preinstalled agents are useful for custom application logs and operating systems that do not natively support the syslog protocol, such as Windows.

Deployment

vRealize Log Insight is available as a preconfigured virtual appliance in OVF. By using the virtual appliance, you can create vRealize Log Insight nodes with predefined identical sizes.

You deploy the OVF file of the virtual appliance once for each node. After node deployment, you access the product to set up cluster nodes according to their role and log in to configure the installation.

Deployment Models

You can deploy vRealize Log Insight as a virtual appliance in one of the following configurations:

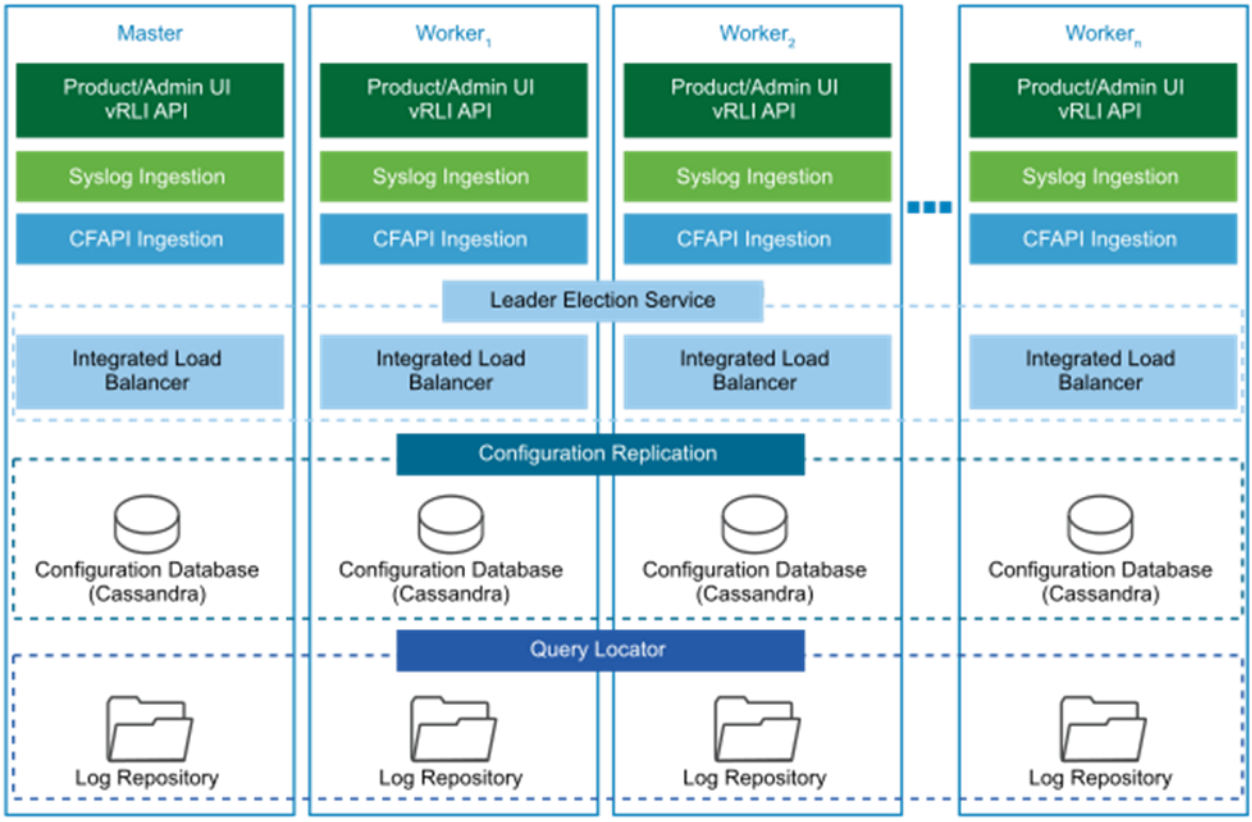
- A standalone node
- A cluster of one master and at least two worker nodes. You can establish high availability by using the integrated load balancer (ILB).

The compute and storage resources of the vRealize Log Insight instances can scale up as growth demands.

Architecture

The architecture of vRealize Log Insight in the SDDC enables several channels for the collection of log messages.

Figure 14) Architecture of vRealize Log Insight.



vRealize Log Insight clients connect to the ILB Virtual IP (VIP) address and use the syslog or the Ingestion API through the vRealize Log Insight agent to send logs to vRealize Log Insight. Users and administrators interact with the ingested logs by using the user interface or the API.

By default, vRealize Log Insight collects data from vCenter Server systems and ESXi hosts. You use content packs to forward logs from NSX for vSphere and vRealize Automation. They contain extensions or provide integration with other systems in the SDDC.

Types of Nodes

For functionality, high availability, and scalability, vRealize Log Insight supports the following types of nodes, which have inherent roles:

- **Master node.** Required initial node in the cluster. In standalone mode, the master node is responsible for all activities, including queries and log ingestion. The master node also handles operations that are related to the lifecycle of a cluster, such as performing upgrades and adding or removing worker nodes. In a scaled-out and highly available environment, the master node still performs such lifecycle operations. However, it functions as a generic worker about queries and log ingestion activities.

  The master node stores logs locally. If the master node is down, the logs stored on it become unavailable.
- **Worker node.** Optional. This component enables scale out in larger environments. As you add and configure more worker nodes in a vRealize Log Insight cluster for high availability (HA), queries and log ingestion activities are delegated to all available nodes. You must have at least two worker nodes to form a cluster with the master node.

  The worker node stores logs locally. If any of the worker nodes is down, the logs on the worker become unavailable.
- **Integrated load balancer.** In cluster mode, the ILB is the centralized entry point that enables vRealize Log Insight to accept incoming ingestion traffic. As nodes are added to the vRealize Log Insight instance to form a cluster, the ILB feature simplifies the configuration for high availability. The ILB balances the incoming traffic fairly among the available vRealize Log Insight nodes.

  The ILB runs on one of the cluster nodes at all times. In environments that contain several nodes, an election process determines the leader of the cluster. Periodically, the ILB performs a health check to determine whether reelection is required. If the node that hosts the ILB VIP address stops responding, the VIP address is failed over to another node in the cluster through an election process.

  All queries against data are directed to the ILB. The ILB delegates queries to a query master for the duration of the query. The query master queries all nodes, both master and worker nodes, for data and then sends the aggregated data back to the client.

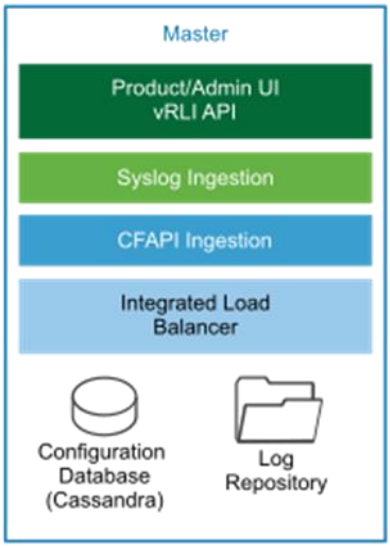
  Use the ILB for administrative activities unless you are performing administrative activities on individual nodes. The web user interface of the ILB presents data from the master and from the worker nodes in a scaled-out cluster in a unified display (single pane of glass).

Application Functional Components

The functional components of a vRealize Log Insight instance interact with each other to perform the following operations:

- Analyze logging data that is ingested from the components of a data center
- Visualize the results in a web browser, or support results query by using API calls

Figure 15) vRealize Log Insight logical node architecture.



vRealize Log Insight components perform these tasks:

- **Product/Admin UI and API.** The UI server is a web application that serves as both user and administration interface. The server hosts the API for accessing collected statistics.
- **Syslog ingestion.** Responsible for ingesting syslog logging data.
- **Log Insight native ingestion API (CFAPI) ingestion.** Responsible for ingesting logging data over the ingestion API by using one of the following methods:
  - vRealize Log Insight agent that has been deployed or preconfigured on SDDC components
  - Log Insight Importer that is used for ingestion of non-real-time data
- **Integration load balancing and election.** Responsible for balancing incoming UI, API, and data ingestion traffic.

  The ILB is a Linux Virtual Server that is built in the Linux kernel for layer 4 load balancing. Each node of vRealize Log Insight contains a service running the ILB, but only a single node functions as the leader at all times. In a single-node vRealize Log Insight instance, this is always the master node. In a scaled-out vRealize Log Insight cluster, this role can be inherited by any of the available nodes during the election process. The leader periodically performs health checks to determine whether a reelection process is required for the cluster.
- **Configuration database.** Stores configuration information about the vRealize Log Insight nodes and cluster. The information that is stored in the database is periodically replicated to all available vRealize Log Insight nodes.
- **Log repository.** Stores logging data that is ingested in vRealize Log Insight. The logging repository is local to each node and is not replicated. If a node is offline or removed, the logging data that is stored on that node becomes inaccessible. In environments where an ILB is configured, incoming logging data is evenly distributed across all available nodes.

  When a query arrives from the ILB, the vRealize Log Insight node that holds the ILB leader role delegates the query to any of the available nodes in the cluster.

Authentication Models

You can configure vRealize Log Insight user authentication to use one or more of the following authentication models:

- Microsoft Active Directory

- Local accounts
- VMware Identity Manager

Content Packs

Content packs add valuable troubleshooting information to vRealize Log Insight. They provide structure and meaning to raw logging data that is collected from either a vRealize Log Insight agent, vRealize Log Insight Importer, or a syslog stream. Content packs add vRealize Log Insight agent configurations, providing out-of-the-box parsing capabilities for standard logging directories and logging formats. They also add dashboards, extracted fields, alert definitions, query lists, and saved queries from the logging data related to a specific product in vRealize Log Insight. For more information, visit the Log Insight Content Pack Marketplace (found inside the administrative interface of the Log Insight appliance) or the VMware Solutions Exchange.

Integration with vRealize Operations Manager

The integration of vRealize Log Insight with vRealize Operations Manager provides data from multiple sources to a central place for monitoring the SDDC. This integration has the following advantages:

- vRealize Log Insight sends notification events to vRealize Operations Manager.
- vRealize Operations Manager can provide the inventory map of any vSphere object to vRealize Log Insight. In this way, you can view log messages from vRealize Log Insight in the vRealize Operations Manager Web user interface, taking you either directly to the object itself or to the location of the object in the environment.
- Access to the vRealize Log Insight user interface is embedded in the vRealize Operations Manager user interface.

Archiving

vRealize Log Insight supports data archiving on an NFS shared storage that the vRealize Log Insight nodes can access. However, vRealize Log Insight does not manage the NFS mount used for archiving, and it does not clean up the archival files.

The NFS mount for archiving can run out of free space or become unavailable for a period of time greater than the retention period of the virtual appliance. In that case, vRealize Log Insight stops ingesting new data until the NFS mount has enough free space or becomes available, or until archiving is disabled. If archiving is enabled, system notifications from vRealize Log Insight sends you an email when the NFS mount is about to run out of space or is unavailable.

Backup

You back up each vRealize Log Insight cluster by using traditional virtual machine backup solutions that are compatible with VADP.

Multiregion vRealize Log Insight Deployment

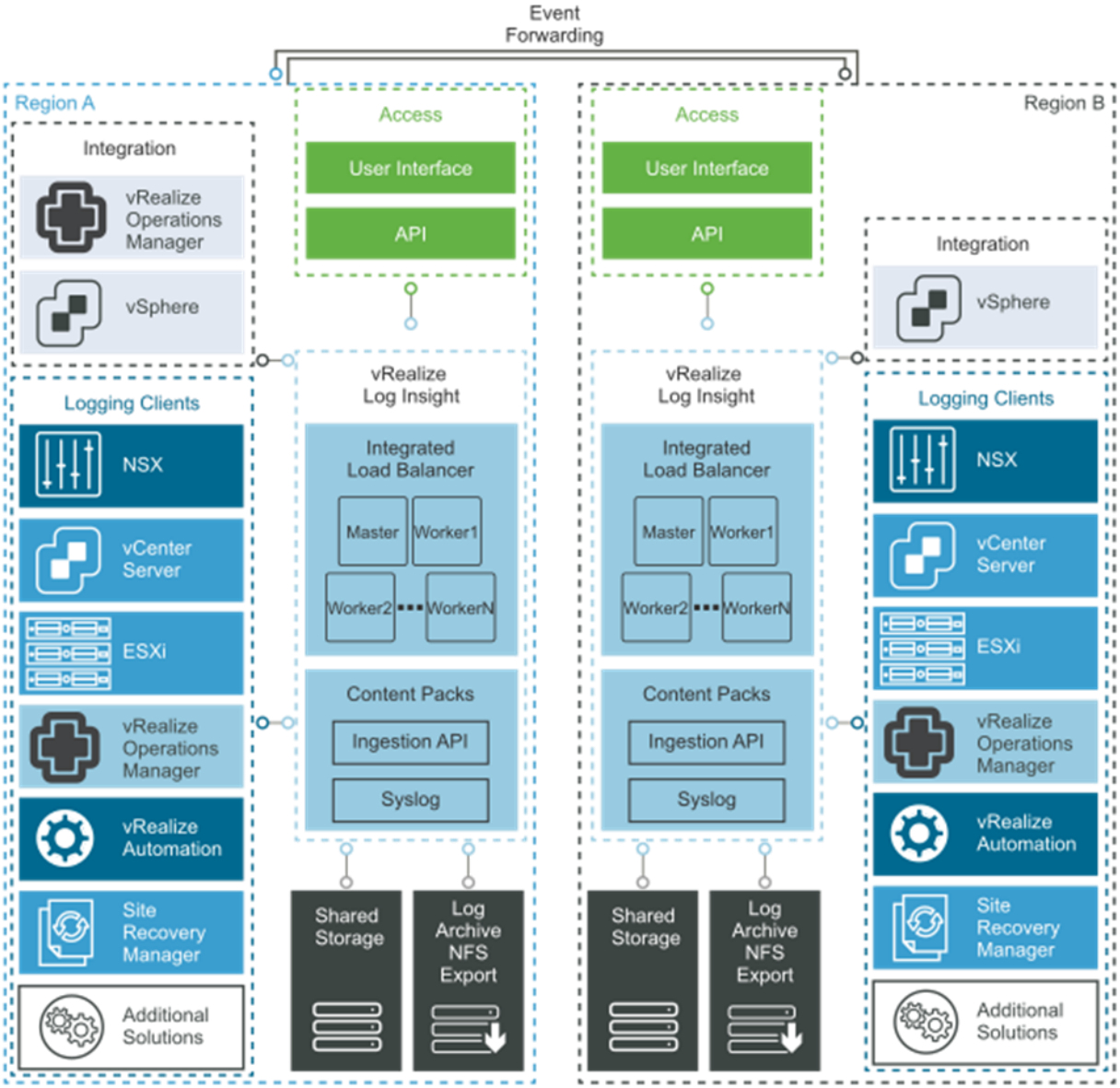
The scope of this validated design can cover multiple regions, with each region containing a single availability zone.

In a multiregion implementation, vRealize Log Insight provides a separate logging infrastructure in each region of the SDDC. Using vRealize Log Insight across multiple regions requires the following configuration:

- A cluster in each region
- Event forwarding to other vRealize Log Insight deployments across regions in the SDDC

Failover with vSphere Replication or disaster recovery with Site Recovery Manager is not necessary. The event forwarding feature adds tags to log messages that identify the source region. Event filtering prevents looping messages between the regions.

Figure 16) Event forwarding in vRealize Log Insight.



5.3 vSphere Update Manager Architecture

vSphere Update Manager provides centralized, automated patch and version management for VMware ESXi hosts and virtual machines on each vCenter Server.

Overview

vSphere Update Manager registers with a single vCenter Server instance so that an administrator can automate the following operations for the lifecycle management of the vSphere environment:

- Upgrade and patch ESXi hosts

- Install and upgrade third-party software on ESXi hosts
- Upgrade virtual machine hardware and VMware Tools

Use the vSphere Update Manager Download Service (UMDS) to deploy vSphere Update Manager on a secured, air-gapped network that is disconnected from other local networks and the internet. UMDS provides a bridge for internet access that is required to pull down upgrade and patch binaries.

Installation Models

The installation models of vSphere Update Manager are different according to the type of vCenter Server installation.

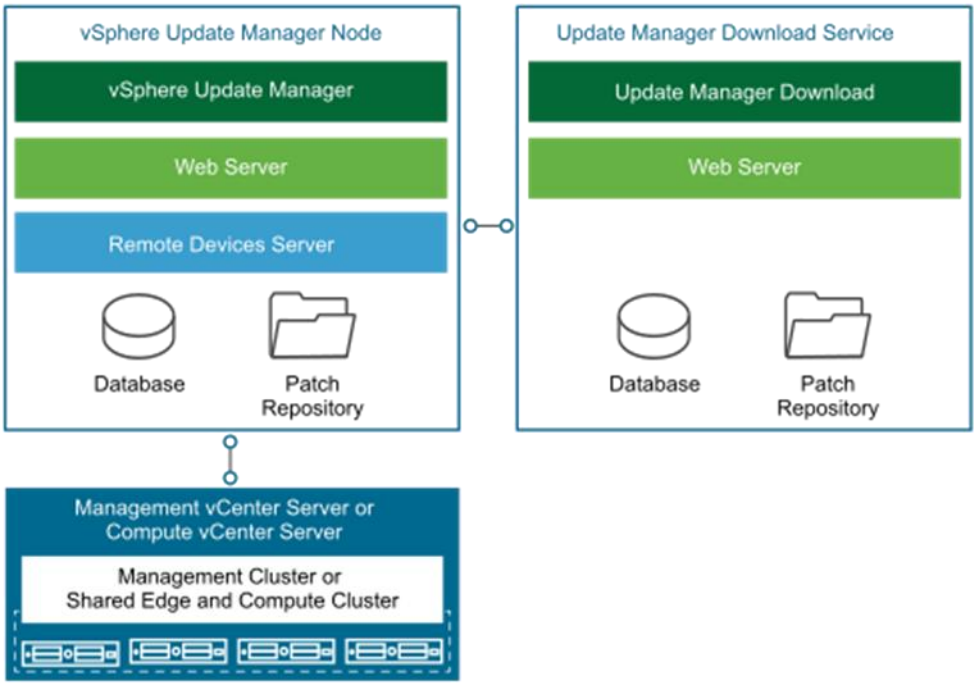
Table 3) Installation models of vSphere Update Manager and Update Manager Download Service.

Component	Installation Model	Description
vSphere Update Manager	Embedded in the vCenter Server appliance	vSphere Update Manager is automatically registered with the container vCenter Server appliance. You access vSphere Update Manager as a plug-in from vSphere Web Client. Use virtual appliance deployment to deploy vCenter Server and vSphere Update Manager as an all-in-one package in which sizing and maintenance for the latter is dictated by the former.
	Windows installable package for installation against a Microsoft Windows vCenter Server	You must run the vSphere Update Manager installation on either vCenter Server itself or an external Microsoft Windows Server. After installation and registration with vCenter Server, you access vSphere Update Manager as a plug-in from vSphere Web Client. Use the Windows installable deployment if you are using a vCenter Server instance for Windows. **Note:** In vSphere 6.5 and later, you can only pair a vSphere Update Manager instance for Microsoft Windows with a vCenter Server instance for Windows.
Update Manager Download Service	Installable package for Linux or Microsoft Windows Server	• For a Linux deployment, install UMDS on Ubuntu 14.0.4 or Red Hat Enterprise Linux 7.0. • For a Windows deployment, install UMDS on one of the supported host operating systems (Host OS) that are detailed in the VMware Knowledge Base article 2091273. You cannot install UMDS on the same system as vSphere Update Manager.

Architecture

vSphere Update Manager contains functional elements that collaborate for monitoring, notifying, and orchestrating the lifecycle management of your vSphere environment in the SDDC.

Figure 17) vSphere Update Manager and Update Manager Download Service architecture.



Types of Nodes

For functionality and scalability, vSphere Update Manager and Update Manager Download Service perform the following roles:

- **vSphere Update Manager.** Required node for integrated, automated lifecycle management of vSphere components. In environments ranging from a single instance to multiple vCenter Server instances, vSphere Update Manager is paired in a 1:1 relationship.
- **Update Manager Download Service.** In a secure environment in which there is an air gap between vCenter Server and vSphere Update Manager and internet access, UMDS provides the bridge for vSphere Update Manager to receive its patch and update binaries. In addition, you can use UMDS to aggregate downloaded binary data, such as patch metadata, patch binaries, and notifications. This binary data can then be shared across multiple instances of vSphere Update Manager to manage the lifecycle of multiple vSphere environments.

Backup

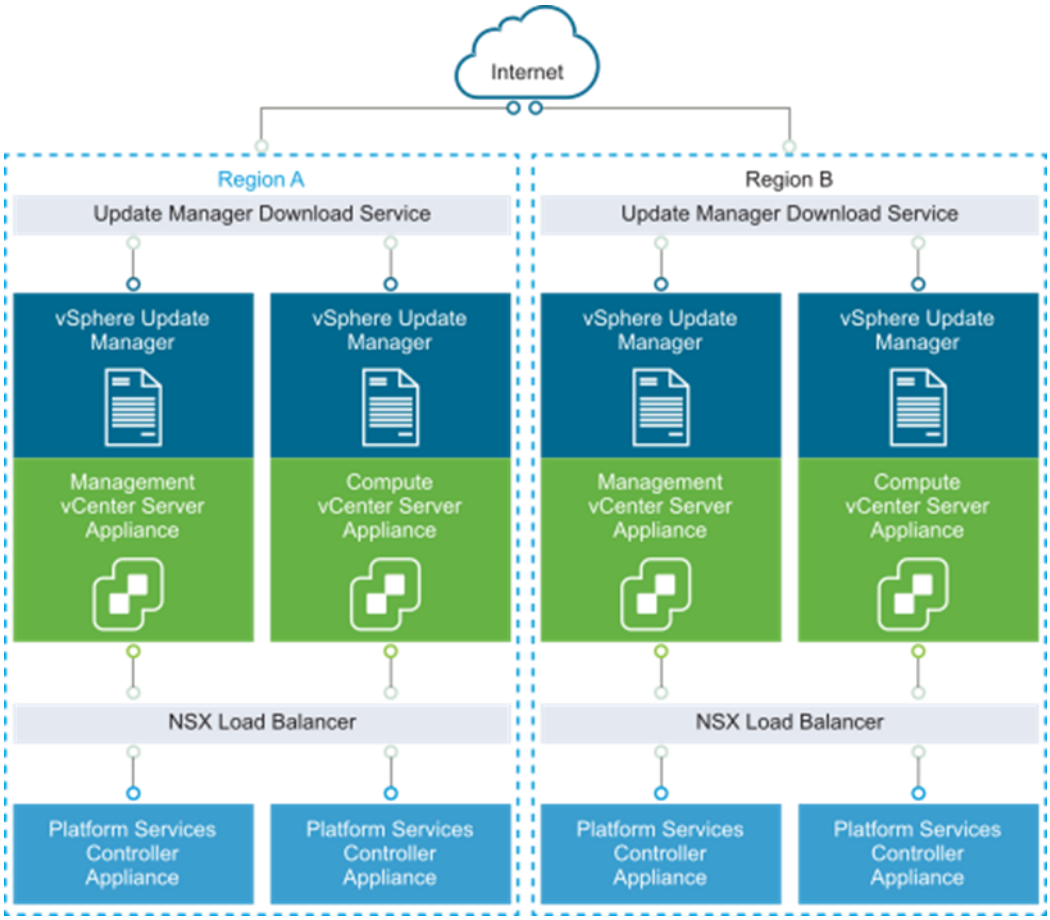
You can back up vSphere Update Manager in two ways – either as an embedded service on the vCenter Server appliance or deployed separately on a Microsoft Windows Server virtual machine. UMDS is backed up by using traditional virtual machine backup solutions. Such solutions are based on software that is compatible with VADP.

Multiregion Deployment of vSphere Update Manager and UMDS

Because of its multiregion scope, the VVD for SDDC uses vSphere Update Manager and UMDS in each region to provide automated lifecycle management of the vSphere components. If you have a vSphere Update Manager service instance with each vCenter Server deployed, you can deploy one UMDS instance per region. In this way, you have a central repository of aggregated patch binaries that are securely downloaded.

Failing over UMDS by using vSphere Replication and Site Recovery Manager is not necessary because each region contains its own UMDS instance.

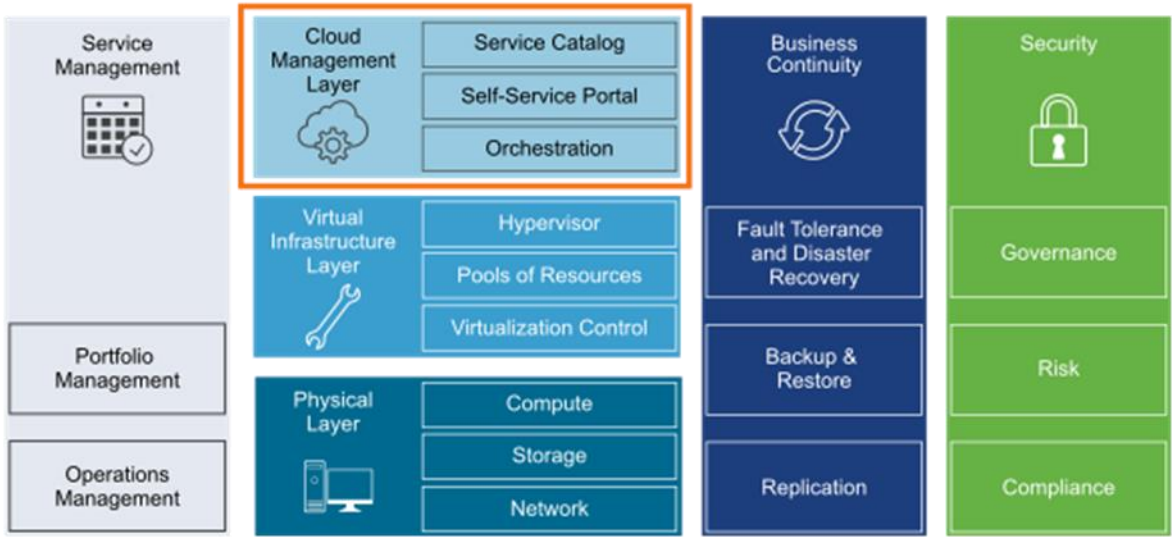
Figure 18) Dual-region interaction between vSphere Update Manager and Update Manager Download Service.



6 Cloud Management Architecture

The CMP is the primary consumption portal for the entire SDDC. Within the SDDC, you use vRealize Automation to author, administer, and consume virtual machine templates and blueprints.

Figure 19) Cloud management platform layer in the SDDC.



The CMP layer delivers the following multiplatform and multivendor cloud services:

- Comprehensive and purpose-built capabilities to provide standardized resources to global customers in a short time span
- Multiplatform and multivendor delivery methods that integrate with existing enterprise management systems
- Central user-centric and business-aware governance for all physical, virtual, private, and public cloud services
- Extensible architecture that meets customer and business needs

6.1 vRealize Automation Architecture of the Cloud Management Platform

vRealize Automation provides a secure web portal on which authorized administrators, developers, and business users can request new IT services and manage specific cloud and IT resources, while complying with business policies. Requests for IT service, including infrastructure, applications, desktops, and many others, are processed through a common service catalog to provide a consistent user experience.

vRealize Automation Installation Overview

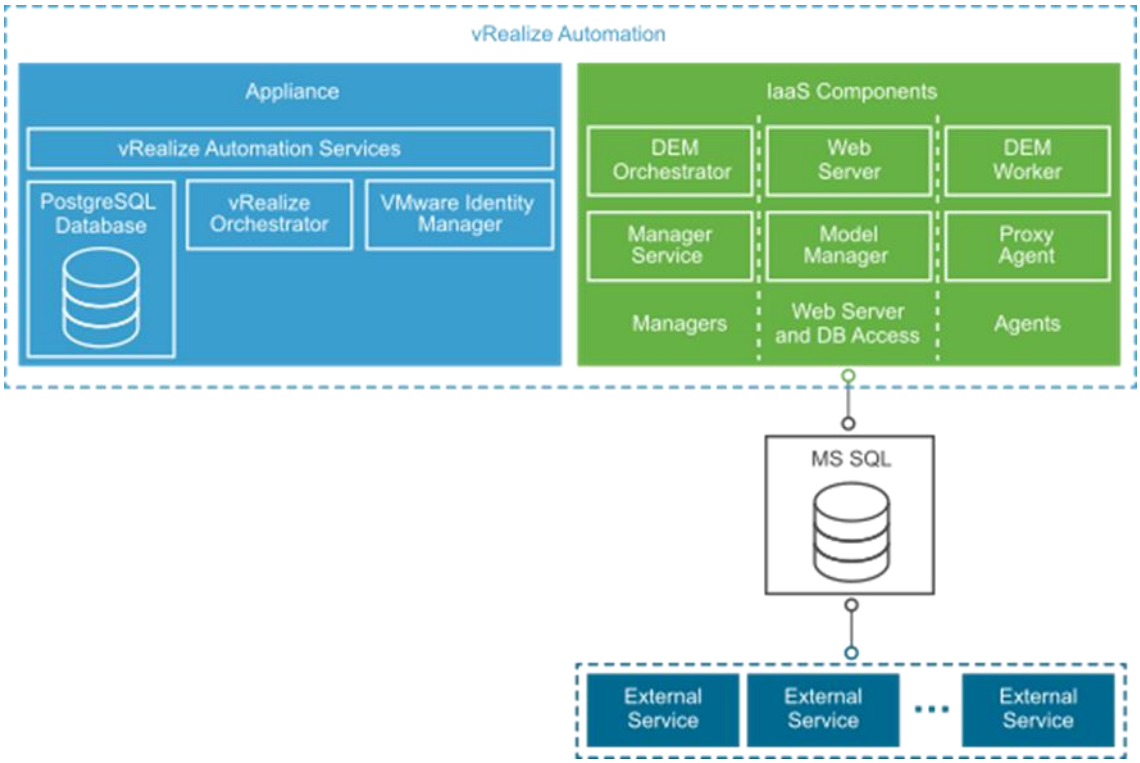
Installing vRealize Automation requires deploying the vRealize Automation appliance and the vRealize Automation IaaS components, which need to be installed on one more Windows servers. To install, you deploy a vRealize Automation appliance and then complete the bulk of the installation with one of the following options:

- A consolidated, browser-based installation wizard
- Separate browser-based appliance configuration and separate Windows installations for IaaS server components
- A command line-based, silent installer that accepts input from an answer properties file
- An installation REST API that accepts JSON-formatted input

vRealize Automation Architecture

vRealize Automation provides self-service provisioning, IT services delivery, and lifecycle management of cloud services across a wide range of multivendor, virtual, physical, and cloud platforms though a flexible and robust distributed architecture. The two main functional elements of the architecture are the vRealize Automation server and the IaaS components.

Figure 20) vRealize Automation architecture.



- **vRealize Automation server appliance.** The vRealize Automation server is deployed as a preconfigured Linux virtual appliance. The vRealize Automation server appliance is delivered as an open virtualization file (OVF) that you deploy on an existing virtualized infrastructure such as vSphere. It performs the following functions:
  - vRealize Automation product portal, where users log in to access self-service provisioning and management of cloud services
  - Single sign-on (SSO) for user authorization and authentication
  - Management interface for vRealize Automation appliance settings
- **Embedded vRealize Orchestrator.** The vRealize Automation appliance contains a preconfigured instance of vRealize Orchestrator. vRealize Automation uses vRealize Orchestrator workflows and actions to extend its capabilities.
- **PostgreSQL Database.** vRealize Server uses a preconfigured PostgreSQL database that is included in the vRealize Automation appliance. This database is also used by the instance of vRealize Orchestrator in the vRealize Automation appliance.
- **IaaS.** vRealize Automation IaaS consists of one or more Microsoft Windows servers that work together to model and provision systems in private, public, or hybrid cloud infrastructures.
- **Model Manager.** vRealize Automation uses models to facilitate integration with external systems and databases. The models implement business logic used by the Distributed Execution Manager (DEM).

The Model Manager provides services and utilities for persisting, versioning, securing, and distributing model elements. Model Manager is hosted on one of the IaaS web servers and communicates with DEMs, the SQL Server database, and the product interface web site.

- **IaaS web server.** The IaaS web server provides infrastructure administration and service authoring to the vRealize Automation product interface. The web server component communicates with the manager service, which provides updates from the DEM, SQL Server database, and agents.
- **Manager service.** Windows service that coordinates communication between IaaS DEMs, the SQL Server database, agents, and SMTP. The manager service communicates with the web server through the Model Manager and must be run under a domain account with administrator privileges on all IaaS Windows servers.
- **Distributed Execution Manager Orchestrator.** A DEM executes the business logic of custom models, interacting with the database and external databases and systems as required. A DEM orchestrator is responsible for monitoring DEM Worker instances, preprocessing workflows for execution, and scheduling workflows.
- **Distributed Execution Manager Worker.** The vRealize Automation IaaS DEM Worker executes provisioning and deprovisioning tasks initiated by the vRealize Automation portal. DEM Workers also communicate with specific infrastructure endpoints.
- **Proxy agents.** vRealize Automation IaaS uses agents to integrate with external systems and to manage information among vRealize Automation components. For example, vSphere proxy agent sends commands to and collects data from a vSphere ESX Server for the VMs provisioned by vRealize Automation.
- **VMware Identity Manager.** VMware Identity Manager is the primary identity provider for vRealize Automation, and manages user authentication, roles, permissions, and overall access into vRealize Automation by means of federated identity brokering. The following authentication methods are supported in vRealize Automation using VMware Identity Manager:
  - Username and password providing single-factor password authentication with basic Active Directory configuration for local users
  - Kerberos
  - Smart card/certificate
  - RSA SecurID
  - RADIUS
  - RSA Adaptive Authentication
  - SAML Authentication

VMware Validated Design Deployment Model

The scope of this VVD includes vRealize Automation appliance large-scale distributed deployment designed for a full-fledged, highly available Cloud Management Portal solution that includes:

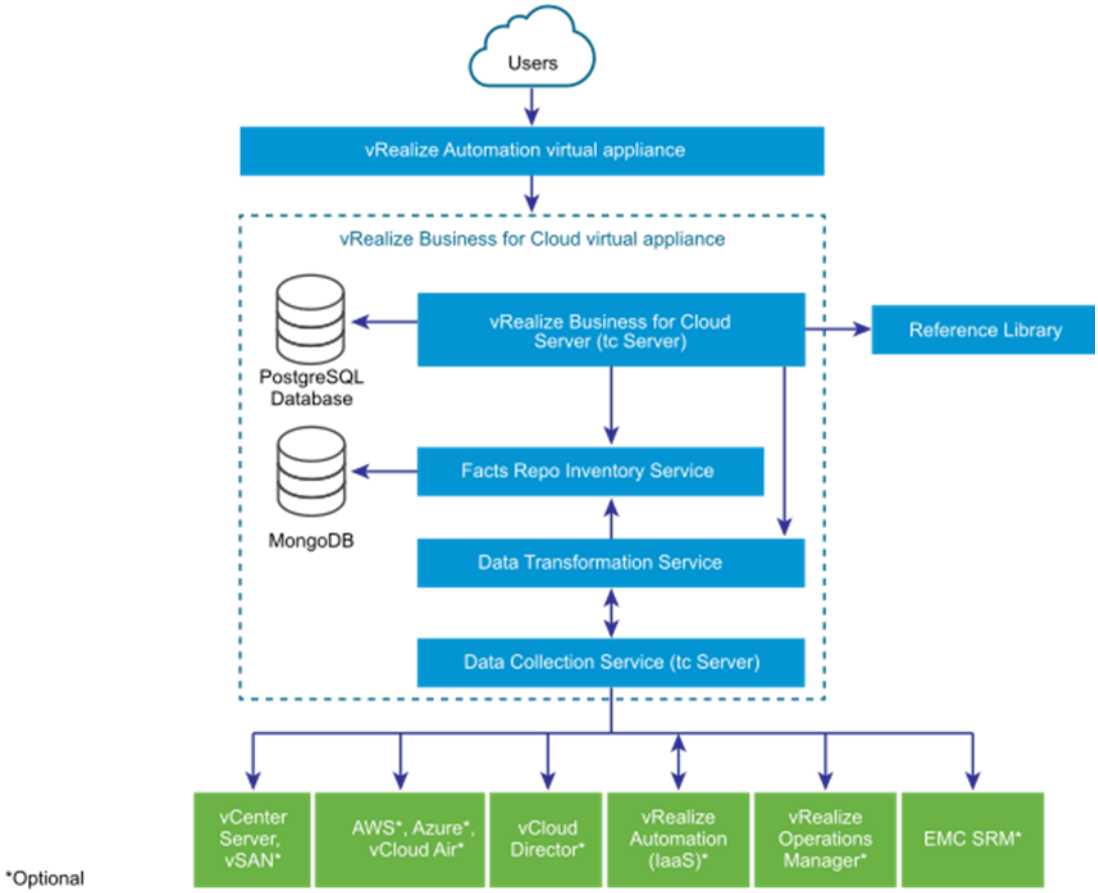
- Two vRealize Automation server appliances behind a load balancer
- Two vRealize Automation IaaS web servers behind a load balancer
- Two vRealize Automation Manager Service nodes (including DEM Orchestrator) behind a load balancer
- Two DEM Worker nodes
- Two IaaS proxy agent nodes

6.2 vRealize Business for Cloud Architecture

VMware vRealize Business for Cloud automates cloud costing, consumption analysis, and comparison, delivering the insight you need to efficiently deploy and manage cloud environments.

vRealize Business for Cloud tracks and manages the costs of private and public cloud resources from a single dashboard. It offers a comprehensive way to view, plan, and manage your cloud costs. vRealize Business for Cloud is tightly integrated with vRealize Automation. The architecture illustrates the main components of vRealize Business for Cloud: the server, the FactsRepo inventory service, the data transformation service, data collection services, and the reference database.

Figure 21) vRealize Business for Cloud.



- **Data collection services.** A set of services for each private and public cloud endpoint, such as vCenter Server, vCloud Director, Amazon Web Services (AWS), and vCloud Air. The data collection services retrieve both inventory information (servers, virtual machines, clusters, storage devices, and associations between them) and usage statistics (CPU and memory). The data collection services use the collected data for cost calculations.

  Note: You can deploy vRealize Business for Cloud in such a way that only its data collection services are enabled. This version of the vRealize Business appliance is known as a remote data collector. Remote data collectors reduce the data collection workload of vRealize Business for Cloud Servers and enable remote data collection from geographically distributed endpoints.

- **FactsRepo inventory service.** An inventory service built on MongoDB to store the collected data that vRealize Business for Cloud uses for cost computation.
- **Data transformation service.** Converts source-specific data from the data collection services into data structures for consumption by the FactsRepo inventory service. The data transformation service serves as is a single point of data aggregation from all data collectors.

- **vRealize Business for Cloud Server.** A web application that runs on Pivotal tc Server. vRealize Business for Cloud has multiple data collection services that run periodically, collecting inventory information and statistics, which are in turn stored in a PostgreSQL database as the persistent data store. Data collected from the data collection services is used for cost calculations.
- **Reference database.** Responsible for providing default, out-of-the-box costs for each of the supported cost drivers. The reference database is updated automatically or manually, and you can download the latest dataset and import it into vRealize Business for Cloud. The new values affect cost calculations. The reference data used depends on the currency you select at the time of installation.

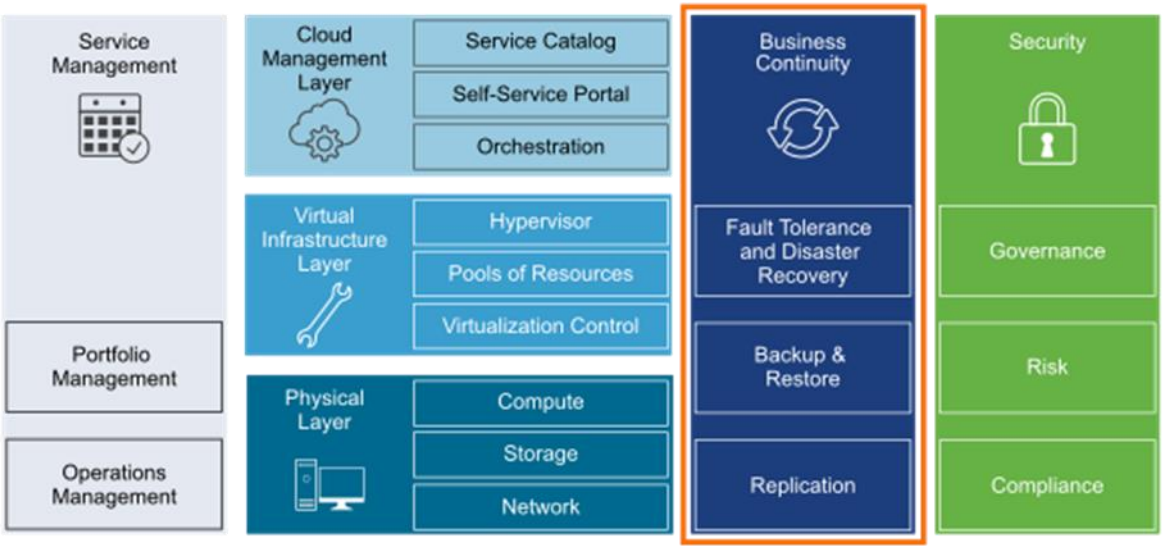
  Note: You cannot change the currency configuration after you deploy vRealize Business for Cloud.
- **Communication between the server and the reference database.** The reference database is a compressed and encrypted file that you can download and install manually or update automatically. For more information, see Update the Reference Database for vRealize Business for Cloud.
- **Other sources of information.** Other sources include vRealize Automation, vCloud Director, vRealize Operations Manager, AWS, Microsoft Azure and vCloud Air, and EMC Storage Resource Manager (SRM). These information sources are optional and are used only if installed and configured.
- **vRealize Business for Cloud operational model.** vRealize Business for Cloud continuously collects data from external sources and periodically updates the FactsRepo inventory service. You can view the collected data by using the vRealize Business for Cloud dashboard, or you can generate a report. The data synchronization and updates occur at regular intervals. However, you can manually trigger the data collection process when inventory changes occur. For example, you can trigger in response to the initialization of the system, or the addition of a private, public, or hybrid cloud account.
- **vRealize Business for Cloud deployment model.** VVD uses three virtual machines: a single vRealize Business for Cloud Server appliance, a single vRealize Business for Cloud remote data collector for Region A, and a single vRealize Business for Cloud remote data collector for Region B.

7 Business Continuity Architecture

The architecture of the business continuity layer includes management components that support backup, restore, and disaster recovery procedures. Within the business continuity layer, management components are implemented to handle the following business continuity requirements:

- Data protection
- Data replication
- Orchestrated disaster recovery

Figure 22) Business continuity layer of the SDDC.



7.1 Data Protection and Backup Architecture

You can implement a backup solution that uses the VADP to protect data in your SDDC management components and the tenant workloads that run in the SDDC.

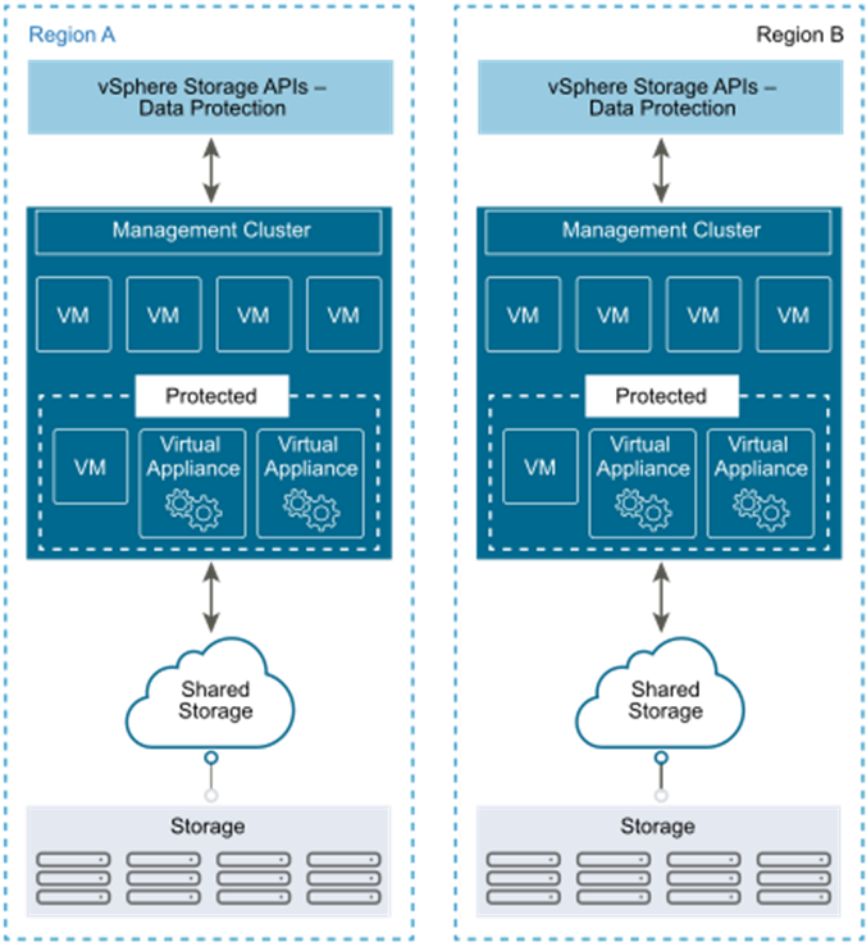
Data protection solutions provide the following functions in the SDDC:

- Backup and restore of virtual machines
- Organization of virtual machines into groups by VMware product
- Data storage according to company retention policies
- Informing administrators about backup and restore activities through reports
- Scheduling regular backups during nonpeak periods

Architecture

VADP instances provide data protection for management components of the SDDC.

Figure 23) Dual-region data protection architecture.



Multiregion Data Protection Deployment

Because of its multiregion scope, the VVD for SDDC calls for the deployment of a VADP-compatible backup solution in the management cluster for each region. Backup jobs are configured to provide recovery of a number of SDDC management components. A VADP-compatible backup solution stores the backups of the management virtual appliances on secondary storage according to a defined schedule.

7.2 Disaster Recovery Architecture

You use Site Recovery Manager in conjunction with vSphere Replication and their constructs to implement cross-region disaster recovery for the workloads of the management products in the SDDC.

Architecture

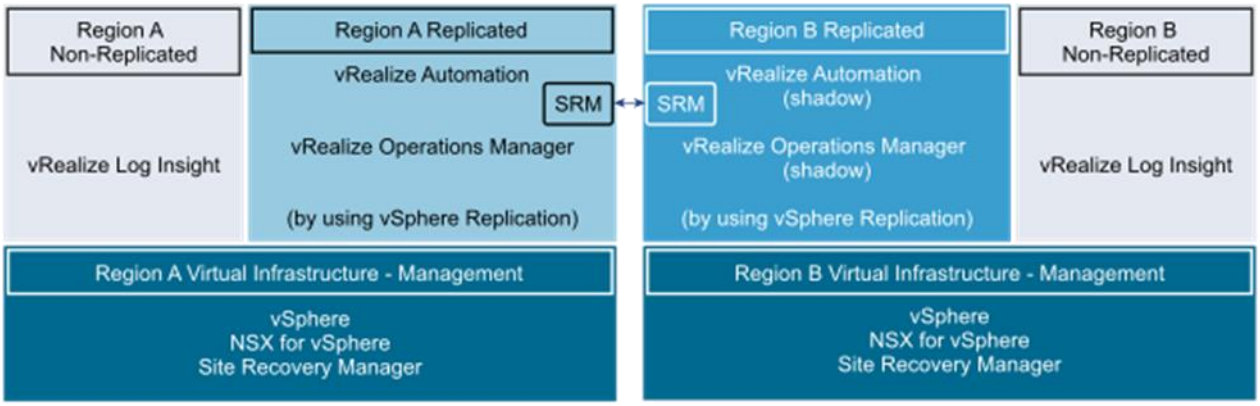
Disaster recovery that is based on Site Recovery Manager has the following main elements:

- **Dual-region configuration.** All protected virtual machines are initially located in Region A, the protected region, and are recovered in Region B, the recovery region. In a typical Site Recovery Manager installation, the protected region provides business-critical data center services. The recovery region is an alternative infrastructure to which Site Recovery Manager can relocate these services.

- **Replication of virtual machine data.** When you use array-based replication, one or more storage arrays at the protected region replicate data to peer arrays at the recovery region. To use array-based replication with Site Recovery Manager, you must first configure replication on the storage array and install a storage-specific adapter before you can configure Site Recovery Manager to use it.
- **vSphere Replication.** You configure vSphere Replication on virtual machines independently of Site Recovery Manager. Replication does not occur at the storage array level. The replication source and target storage can be any storage device.

  Note: You can configure vSphere Replication to use the multiple-point-in-time snapshot feature. This configuration offers greater flexibility for data recovery of protected virtual machines on the recovery region.
- **Protection groups.** A protection group is a group of virtual machines that fail over together at the recovery region during test and recovery. Each protection group protects one datastore group, and each datastore group can contain multiple datastores. However, you cannot create protection groups that combine virtual machines protected by array-based replication and vSphere Replication.
- **Recovery plans.** A recovery plan specifies how Site Recovery Manager recovers the virtual machines in the protection groups. You can include a combination of array-based replication protection groups and vSphere Replication protection groups in the same recovery plan.

Figure 24) Disaster recovery architecture.



Multiregion Site Recovery Manager Deployment

This VVD for SDDC pairs two Site Recovery Manager servers deployed on the management cluster. This design implements the following disaster recovery configuration:

- The following management applications are protected in the event of a disaster:
  - vRealize Automation and vRealize Business Server
  - The analytics cluster of vRealize Operations Manager
- The virtual infrastructure components that are not covered by disaster recovery protection, such as vRealize Log Insight, are available as separate instances in each region.

8 Detailed Design

This detailed SDDC design covers both physical and virtual infrastructure design. It includes numbered design decisions and the justification and implications of each decision.

- **Physical infrastructure design.** Focuses on the three main pillars of any data center: compute, storage, and network. This section contains information about availability zones and regions. It also

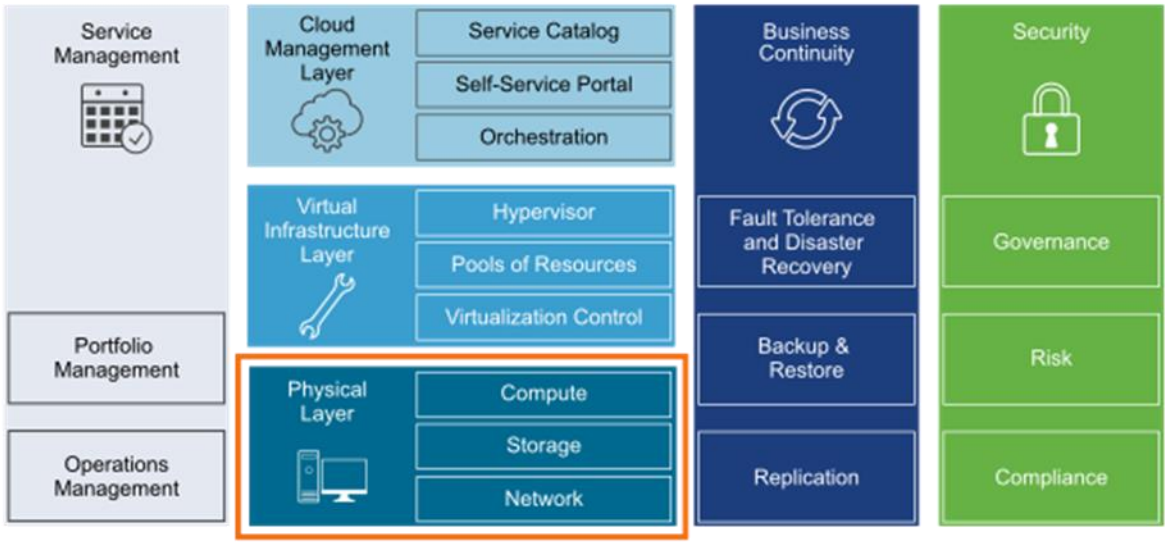
provides details on the rack and cluster configuration, and on physical ESXi hosts and the associated storage and network configurations.

- **Virtual infrastructure design.** Describes the core virtualization software configuration. This section has information about the ESXi hypervisor, vCenter Server, the virtual network design including VMware NSX, and software-defined storage for NetApp HCI. This section also includes details on business continuity (backup and restore) and disaster recovery.
- **CMP design.** Contains information about the consumption and orchestration layer of the SDDC stack, which uses vRealize Automation and vRealize Orchestrator. IT organizations can use the fully distributed and scalable architecture to streamline their provisioning and decommissioning operations.
- **Operations infrastructure design.** Explains how to architect, install, and configure vRealize Operations Manager and vRealize Log Insight. You learn how to make sure that service management within the SDDC is comprehensive. This section ties directly into the Operational Guidance section.

9 Physical Infrastructure Design

The physical infrastructure design includes decisions for availability zones and regions and the cluster layout in data center racks. Design decisions related to server, networking, and storage hardware are part of the physical infrastructure design.

Figure 25) Physical infrastructure design.



9.1 Physical Design Fundamentals

Physical design fundamentals include decisions about availability zones, regions, workload domains, clusters, and racks. The ESXi host physical design is also a part of design fundamentals.

- **Physical Networking Design.** VVD for SDDC can use most enterprise-grade physical network architectures.
- **Physical Storage Design.** This VVD uses different types of storage. Section 10.8, Shared Storage Design, contains background information and explains where the SDDC uses each type of storage.

Regions

This design deploys a protected region with one availability zone and a recovery region with a single availability zone.

Regions provide disaster recovery across different SDDC instances. This design uses two regions. Each region is a separate SDDC instance. The regions have a similar physical layer and virtual infrastructure designs but different naming.

The identifiers follow United Nations Code for Trade and Transport Locations(UN/LOCODE) and also contain a numeric instance ID.

Table 4) Region identifiers.

Region	Availability Zone and Region Identifier	Region-Specific Domain Name	Region Description
A	SFO01	sfo01.rainpole.local	San Francisco, CA, USA based data center
B	LAX01	lax01.rainpole.local	Los Angeles, CA, USA based data center

Note: Region identifiers might vary according to the locations used in the deployment.

Table 5) Availability zones and region design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-001	In Region A, deploy one availability zone to support all the SDDC management components and their SLAs.	A single availability zone can support all SDDC management and compute components for a region.	
SDDC-PHY-002	Use two regions.	Supports the technical requirement of multiregion failover capability according to the design objectives.	Having multiple regions requires an increased solution footprint and associated costs.
SDDC-PHY-003	In Region B, deploy a single availability zone that can support disaster recovery of the SDDC management components.	A single availability zone can support all SDDC management and compute components for a region. You can later add another availability zone to extend and scale the management and compute capabilities of the SDDC.	

Clusters and Racks

The SDDC functionality is split across multiple clusters. Each cluster can occupy one rack or multiple racks. The total number of racks for each cluster type depends on scalability needs.

Figure 26) SDDC cluster architecture.

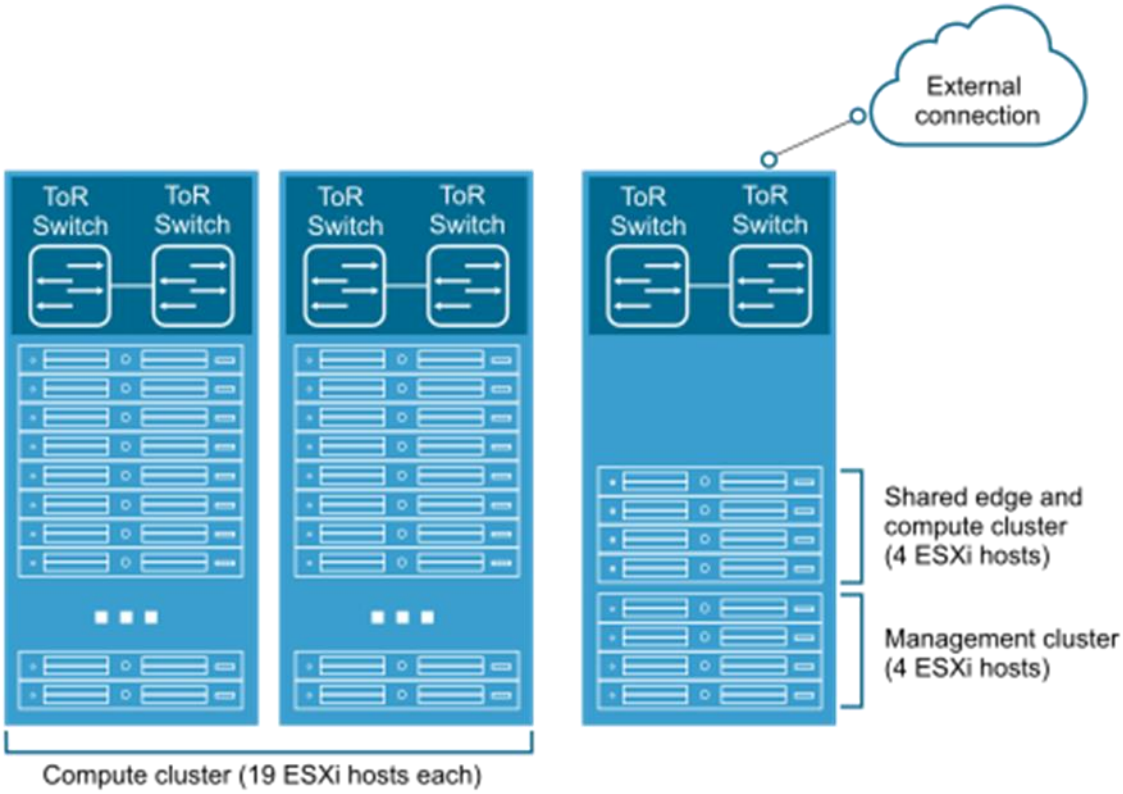


Table 6) Cluster and racks design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-004	In each region, place the management cluster and the shared edge and compute cluster in the same rack.	The number of required compute resources for the management cluster (four ESXi servers) and shared edge and compute cluster (four ESXi servers) is low and does not justify a dedicated rack for each cluster. On-ramp and off-ramp connectivity to physical networks (for example, north-south layer 3 routing on NSX Edge virtual appliances) can be supplied to both the management and compute clusters through this management and edge rack. Edge resources require external connectivity to physical network devices. Placing edge resources for management and compute in the same rack minimizes VLAN spread.	The data centers must include sufficient power and cooling to operate the server equipment. This depends on the selected vendor and products. If the equipment in this entire rack fails, a second region is needed to mitigate the downtime associated with such an event.
SDDC-PHY-005	External storage occupies one or more racks.	To simplify the scale out of the SDDC infrastructure, the storage-to-racks relationship has been standardized. It is possible that the storage system arrives from the manufacturer in a dedicated rack or set of racks. A storage system of this type is accommodated in the design.	Data centers must include sufficient power and cooling to operate the storage equipment. This depends on the selected vendor and products.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-006	Use two separate power feeds for each rack.	Redundant power feeds increase availability by making sure that failure of a power feed does not bring down all equipment in a rack. Combined with redundant network connections into a rack and within a rack, redundant power feeds prevent failure of equipment in an entire rack.	All equipment used must support two separate power feeds. The equipment must keep running if one power feed fails. If the equipment of an entire rack fails, the cause, such as flooding or an earthquake, also affects neighboring racks. Use a second region to reduce the downtime associated with such an event.
SDDC-PHY-007	Mount the compute resources (minimum of four ESXi hosts) for the management cluster together in a rack.	Mounting the compute resources for the management cluster together can ease physical data center design, deployment, and troubleshooting.	None
SDDC-PHY-008	Mount the compute resources for the shared edge and compute cluster (minimum of four ESXi hosts) together in a rack.	Mounting the compute resources for the shared edge and compute cluster together can ease physical data center design, deployment, and troubleshooting.	None

ESXi Host Physical Design Specifications

The physical design specifications of the ESXi host list the characteristics of the ESXi hosts that were used during deployment and testing of this VVD.

Physical Design Specification Fundamentals

The configuration and assembly process for each system is standardized, with all components installed in the same manner on each ESXi host. Standardizing the entire physical configuration of the ESXi hosts is crucial to providing an easily manageable and supportable infrastructure, because standardization eliminates variability. Deploy ESXi hosts with an identical configuration, including identical storage, and networking configurations, across all cluster members. For example, consistent PCI card slot placement, especially for network controllers, is essential for accurate alignment of physical to virtual I/O resources. Using identical configurations creates an even balance of virtual machine storage components across storage and compute resources.

All of these recommendations are met in this design through the use of NetApp HCI compute nodes. Select all nodes on a per-cluster basis from Table 7.

Table 7) Compute node model selection.

	H300E (Small)	H500E (Medium)	H700E (Large)
CPU	2x Intel E5-2620v4, 8 core, 2.1GHz	2x Intel E5-2650v4, 12 core, 2.2GHz	2x Intel E5-2695v4, 18 core, 2.1GHz
Total PCPU Cores	16	24	36
RAM in GB	384	512	768

Table 8) ESXi host design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-009	Use NetApp HCI E-Series compute nodes	Standardized compute node hardware configurations creates a seamless deployment, verifies VMware HCL compatibility, and eases troubleshooting.	Hardware choices might be limited.
SDDC-PHY-010	You must verify that all nodes have uniform configurations across a given cluster.	A balanced cluster delivers more predictable performance even during hardware failures. In addition, performance impact during maintenance mode evacuations and similar operations is minimal if the cluster is balanced.	None

ESXi Host Memory

The amount of memory required for compute clusters varies according to the workloads running in the cluster. When sizing memory for compute cluster hosts, it is important to remember the admission control setting (n+1), which reserves one host resource for failover or maintenance.

Table 9) Host memory design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-011	Use NetApp HCI H300E (small) compute nodes for the management cluster.	The management and edge VMs in this cluster require a total of 424GB RAM. 4x H300E compute nodes provide more than enough RAM and CPU resources to accommodate these needs, up to the N+2 level (for example, two hosts can be in maintenance mode).	None

ESXi Host Boot Device

Host boot devices are provided by two M.2 SATA solid-state drives (SSDs) that are automatically configured by the NetApp Deployment Engine.

9.2 Physical Networking Design

VVD for SDDC can use most enterprise-grade physical network architectures.

Switch Types and Network Connectivity

Setting up the physical environment requires careful consideration. Follow best practices for physical switches, switch connectivity, VLANs, subnets, and access port settings.

Top-of-Rack Physical Switches

When configuring top-of-rack (ToR) switches, consider the following best practices.

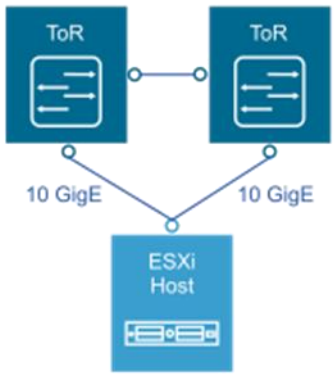
- Configure redundant physical switches to enhance availability.
- Configure switch ports that connect to ESXi hosts manually as trunk ports. Virtual switches are passive devices and do not support trunking protocols, such as the Dynamic Trunking Protocol.
- Modify the Spanning Tree Protocol on any port that is connected to an ESXi NIC to reduce the time it takes to transition ports over to the forwarding state, for example by using the Trunk PortFast feature found in a physical Cisco switch.
- Provide DHCP or DHCP Helper capabilities on all VLANs that are used by management and VXLAN VMkernel ports. This setup simplifies the configuration by using DHCP to assign IP address based on the IP subnet in use.
- Configure jumbo frames on all switch ports, interswitch links (ISLs), and switched virtual interfaces (SVIs).

Top-of-Rack Connectivity and Network Settings

Each ESXi host is connected redundantly to the SDDC network fabric ToR switches by means of two 10GbE ports. Configure the ToR switches to provide all necessary VLANs through an 802.1Q trunk. These redundant connections use features in vSphere Distributed Switch and NSX for vSphere so that no physical interface is overrun and redundant paths are used as long as they are available.

This validated design does not use hardware-based link aggregation for the compute nodes. However, it does use it for the 10/25GbE ports on the storage nodes. It is important that the switches chosen offer some form of multichassis link aggregation (MLAG) that supports LACP.

Figure 27) Host to ToR connectivity.



VLANs and Subnets

Each ESXi host uses VLANs and corresponding subnets.

Follow these guidelines:

- Use only /24 subnets to reduce confusion and mistakes when dealing with IPv4 subnetting.
- Use the IP address .253 as the floating interface with .251 and .252 for the Virtual Router Redundancy Protocol (VRPP) or the Hot Standby Routing Protocol (HSRP).
- Use the RFC1918 IPv4 address space for these subnets and allocate one octet by region and another octet by function. For example, the mapping `172.regionid.function.0/24` results in the following sample subnets.

Note: The following VLANs and IP ranges are samples. Your actual implementation depends on your environment.

Table 10) Sample values for VLANs and IP ranges.

Cluster	Function	Sample VLAN	Sample IP Range
Management	Management	1611 (native, stretched)	172.16.11.0/24
	vMotion	1612	172.16.12.0/24
	VXLAN	1614	172.16.14.0/24
	iSCSI	1613	172.16.13.0/24
Shared Edge and Compute	Management	1631 (native)	172.16.31.0/24
	vMotion	1632	172.16.32.0/24
	VXLAN	1634	172.16.34.0/24
	iSCSI	1633	172.16.33.0/24

Access Port Network Settings

Configure additional network settings on the access ports that connect the ToR switches to the corresponding servers.

- **Spanning Tree Protocol (STP).** Although this design does not use the STP, switches usually come with STP configured by default. Designate the access ports as trunk PortFast.
- **Trunking.** Configure the VLANs as members of an 802.1Q trunk with the management VLAN acting as the native VLAN.
- **MTU (Maximum Transmission Unit).** Set MTU for all VLANS and SVIs (management, vMotion, VXLAN, and storage) to jumbo frames for consistency.
- **DHCP Helper.** Configure the VIF of the management and VXLAN subnet as a DHCP proxy.
- **Multicast.** Configure Internet Group Management Protocol (IGMP) snooping on the ToR switches and include an IGMP querier on each VXLAN VLAN.

Region Interconnectivity

The SDDC management networks, the VXLAN kernel ports, and the edge and compute VXLAN kernel ports of the two regions must be connected. These connections can be over a VPN tunnel, point-to-point circuits, Multiprotocol Label Switching (MPLS), and so on. End users must be able to reach the public-facing network segments (public management and tenant networks) of both regions.

The region interconnectivity design must support jumbo frames and provide latency of less than 150ms. For full details on the requirements for region interconnectivity, see the Cross-vCenter NSX Installation Guide.

The design of a region connection solution is out of scope for this VVD.

Physical Network Design Decisions

The physical network design decisions determine the physical layout and use of VLANs. They also include decisions on jumbo frames and on other network-related requirements such as DNS and NTP.

Physical Network Design Decisions

- **Routing protocols.** Base the selection of the external routing protocol on your current implementation or on the expertise of the IT staff. Take performance requirements into consideration. Possible options are Open Shortest Path First (OSPF), the BGP, and Intermediate System to Intermediate System (IS-IS). Although each routing protocol has a complex set of advantages and disadvantages, this validated design uses BGP as its routing protocol.
- **DHCP proxy.** Set the DHCP proxy to point to a DHCP server by way of its IPv4 address. See the VVD Planning and Preparation document for details on the DHCP server.

Table 11) Physical network design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-NET-001	The physical network architecture must support the following requirements: • One 10GbE port on each ToR switch for ESXi host uplinks • No ether-channel (LAG/vPC) configuration for ESXi host uplinks • Layer 3 device that supports BGP and IGMP	Guarantees availability during a switch failure. This design uses vSphere host profiles that are not compatible with link-aggregation technologies. BGP is used as the dynamic routing protocol in this design. NSX Hybrid mode replication requires IGMP.	Hardware choices might be limited. Requires dynamic routing protocol configuration in the physical networking stack.
SDDC-PHY-NET-002	Use a physical network that is configured for BGP routing adjacency.	This design uses BGP as its routing protocol. Supports flexibility in network design for routing multisite and multitenancy workloads.	Requires BGP configuration in the physical networking stack.
SDDC-PHY-NET-003	Use two ToR switches for each rack.	This design uses two 10GbE links to each server to provide redundancy and reduce the overall design complexity.	Requires two ToR switches per rack, which can increase costs.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-NET-004	Use VLANs to segment physical network functions.	• Supports physical network connectivity without requiring many NICs. • Isolates the different network functions of the SDDC so that you can have differentiated services and prioritized traffic as needed.	Requires uniform configuration and presentation on all the trunks made available to the ESXi hosts.

Additional Design Decisions

Additional design decisions deal with static IP addresses, DNS records, and the required NTP time source.

Table 12) Additional network design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-NET-005	Assign static IP addresses to all management components in the SDDC infrastructure except for NSX VXLAN Tunnel Endpoints (VTEPs), which DHCP assigns.	Avoids connection outages due to DHCP availability or misconfiguration.	Requires accurate IP address management.
SDDC-PHY-NET-006	Create DNS records for all management nodes to enable forward, reverse, short, and fully qualified domain name (FQDN) resolution.	Ensures consistent resolution of management nodes by using both IP address (reverse lookup) and name resolution.	None
SDDC-PHY-NET-007	Use an NTP time source for all management nodes.	It is crucial to maintain accurate and synchronized time between management nodes.	None

Jumbo Frames Design Decisions

IP storage throughput can benefit from the configuration of jumbo frames. Increasing the per-frame payload from 1500 bytes to the jumbo frame setting improves the efficiency of data transfer. Jumbo frames must be configured end to end, which is feasible in a LAN environment. When you enable jumbo frames on an ESXi host, you have to select an MTU that matches the MTU of the physical switch ports.

The workload determines whether it makes sense to configure jumbo frames on a virtual machine. If the workload consistently transfers large amounts of network data, configure jumbo frames, if possible. In that case, confirm that both the virtual machine operating system and the virtual machine NICs support jumbo frames.

Using jumbo frames also improves the performance of vSphere vMotion.

Note: VXLAN requires an MTU value of at least 1600 bytes on the switches and routers that carry the transport zone traffic.

Table 13) Jumbo frames design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-NET-008	Configure the MTU size to at least 9000 bytes (jumbo frames) on the physical switch ports and distributed switch port groups that support the following traffic types. • NFS • iSCSI • vMotion • VXLAN • vSphere Replication	Improves traffic throughput. To support VXLAN, increase the MTU setting to a minimum of 1600 bytes. Setting this port group to 9000 bytes has no effect on VXLAN but ensures consistency across port groups that are adjusted from the default MTU size.	When adjusting the MTU packet size, you must also configure the entire network path (VMkernel port, distributed switch, physical switches, and routers) to support the same MTU packet size.

9.3 Physical Storage Design

This VVD uses different types of storage. Section 10.8, Shared Storage Design, explains where the SDDC uses each type and also provides background information. The focus of this section is physical storage design. All functional testing and validation of the designs was performed using NetApp HCI for guest VMs and datastores and NetApp ONTAP Select for file services.

NetApp Element Software

NetApp Element® software is designed for data centers in which rapid, modular growth or contraction is required for diverse workloads. Element is the storage infrastructure of choice for service providers, due to its flexible handling of permanent and transient workloads with various throughput and capacity requirements.

Element provides modular, scalable performance, with each storage node delivering guaranteed capacity and throughput to the environment. Each Element storage node added to a NetApp HCI environment provides a set amount of IOPS and capacity, allowing predictable, plannable growth.

Because each node provides a set throughput (IOPS) to the storage environment, QoS for each workload can be guaranteed. Minimum SLAs are assured with Element because the total throughput of the cluster is a known, quantifiable amount.

Element offers inline data deduplication and compression, enabling efficiency. Deduplication is performed at a cluster-wide scale, improving overall efficiency as used capacity increases.

The design of Element is based on ease of use. It has fewer “nerd knobs” because it offers the most efficient use with the fewest trade-offs. Element uses iSCSI configuration to eliminate the need for the elaborate tables of initiators and targets found in FC configurations. The use of iSCSI also provides automatic scaling capabilities and the inherent ability to take advantage of new Ethernet capabilities.

Finally, Element was designed for automation. All storage features are available through APIs, which are the only methods used by the UI to control the system. VMware private cloud utilities can consume these APIs to make self-service resources and fully integrated monitoring and debugging available to VMware management tools.

For more information, see the Element product page.

NetApp HCI Storage Physical Design

This design uses NetApp HCI storage nodes to implement software-defined storage for all clusters.

Software-defined storage is a key technology in the SDDC. NetApp HCI storage nodes use NetApp SolidFire® all-flash, shared-nothing technology to create a highly optimized, scale-out cluster of shared storage. This system delivers best-in-class QOS to provide deterministic performance in a mixed workload environment.

The physical design of a NetApp HCI storage cluster is simple. You don't need to make any decisions about controllers or physical disks. All disks in the storage nodes, regardless of model, are identical high-performance, enterprise-grade SSDs. Furthermore, there is no caching tier, so there is no need to do any sizing other than capacity in TB and IOPS. Finally, compression and deduplication are always enabled.

Despite this simplicity, I/O response times remain deterministic, even under heavy mixed workload conditions.

Requirements and Dependencies

- A minimum of four NetApp HCI storage nodes per region
- A maximum of 40 NetApp HCI storage nodes per region
- NetApp does not recommend any specific Ethernet switch vendor or models as long as they meet the following criteria
  - 10GbE or 25GbE switch ports
  - Nonblocking backplane
  - <1ms latency between any nodes

NetApp HCI Storage Nodes

NetApp HCI storage nodes are joined together in a scale-out cluster that is typically composed of matching node models for simplified capacity management. This is not a requirement, however; node models and generations can be mixed within a single cluster. This is particularly common during rolling cluster upgrades.

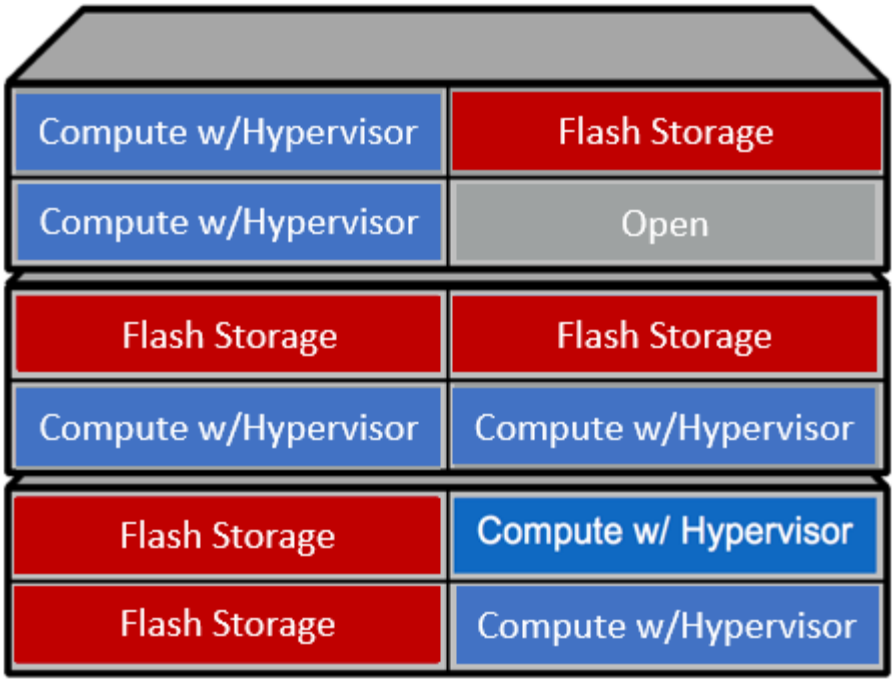
Two basic types of storage nodes are available to use with NetApp HCI:

- 1RU half-width for use inside a 4-node chassis alongside the compute nodes
- 1RU full-width for independent rack mounting separate from the compute nodes

Chassis-Based Storage Nodes

This type of storage node is installed alongside NetApp HCI compute nodes in a 2U/4-slot NetApp H-Series chassis, as shown in Figure 28:

Figure 28) Compute and storage nodes installed in H-Series chassis.



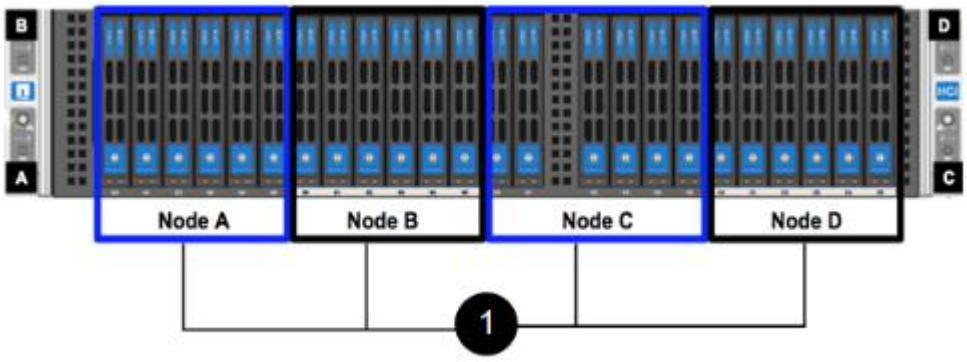
Note: The number of storage nodes in a NetApp HCI environment is independent of the number of compute nodes. Usually, deployments have a different number of storage nodes than compute nodes because storage requirements in an environment can vary greatly.

Chassis-based storage nodes require access to the hot-swappable 2.5-inch SSDs in the front of the NetApp HCI H-series chassis. Because of this, you must place a storage node in a chassis slot that corresponds to its six available drives. You can load compute nodes into any slot in the NetApp HCI chassis that does not have drives populated for that slot; compute nodes do not use external SSD storage.

The SSD drives for the storage nodes are arranged in groups of six across the front of the chassis. Each storage node has a fixed set of six SSD drives connected to it. If you move a storage node from one chassis slot to another, you must also move the corresponding drives in the front of the chassis. You must populate all six drives for each storage node.

Figure 29 shows the front of a 4-node storage chassis:

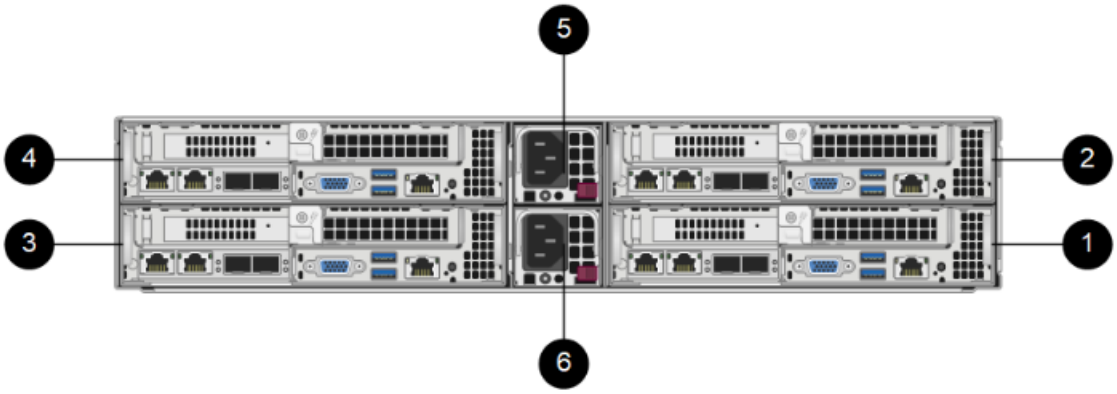
Figure 29) H-Series chassis drive mapping.



Item	Description
1	Six drives per node in a four-node chassis

Figure 30 shows the back of a NetApp HCI H-Series chassis. Each chassis includes two power supply units for power redundancy. This example contains four storage nodes.

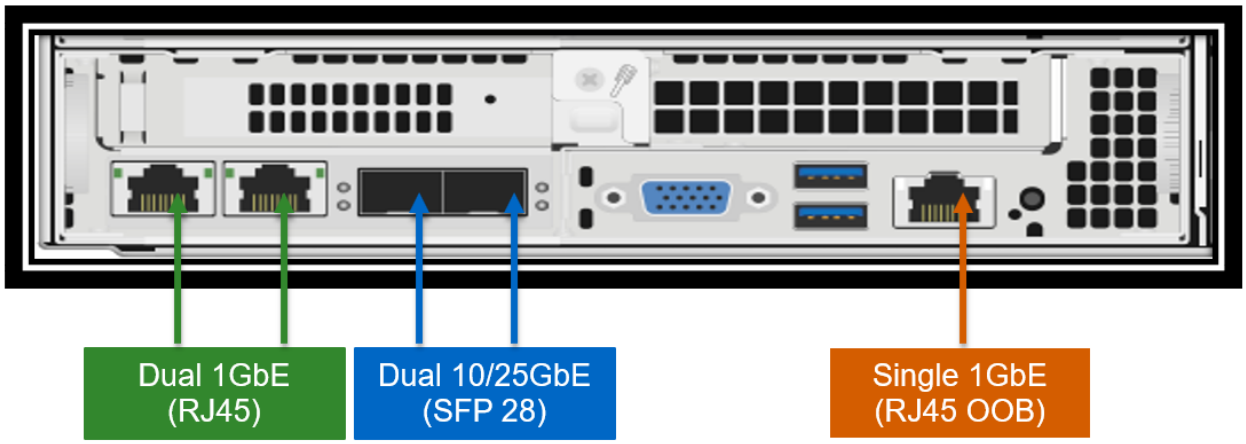
Figure 30) H-Series chassis rear view.



Item	Component	Drives Used
1	Node A	A0-A5
2	Node B	B0-B5
3	Node C	C0-C5
4	Node D	D0-D5
5	Power supply 1	N/A
6	Power supply 2	N/A

Figure 31 shows a detailed view of the back of a single storage node.

Figure 31) Chassis-based storage node detailed view.



Color	Port Type	Purpose
Green	Dual 1GbE (RJ45)	In-band management
Blue	Dual 10/25GbE (SFP 28)	iSCSI host traffic and intranode replication
Orange	Single 1GbE (RJ45)	Out-of-band IPMI management

Chassis-Based Storage Node Model Selection

Table 14 shows the available chassis-based storage nodes.

Table 14) Chassis-based storage node models.

	H300S	H500S	H700S
Form factor	1RU, half-width	1RU, half-width	1RU, half-width
CPU	E5-2620 v4: 8C at 2.1GHz	E5-2620 v4: 8C at 2.1GHz	E5-2620 v4: 8C at 2.1GHz
Boot device	1 x 240GB MLC	1 x 240GB MLC	1 x 240GB MLC
Networking	2x 10/25GbE SFP28/SFP+ 2x 1GbE RJ45	2x 10/25GbE SFP28/SFP+ 2x 1GbE RJ45	2x 10/25GbE SFP28/SFP+ 2x 1GbE RJ45
SSD	6x 480GB	6x 960GB	6x 1.9TB
IOPS	50,000	50,000	100,000
Effective block capacity (min) [1]	5.5TB	11TB	22TB

[1] NetApp HCI effective capacity calculations account for NetApp SolidFire Helix® data protection, system overhead, and global efficiency, including compression, deduplication, and thin provisioning. Element software customers typically achieve an effective capacity range of 5 to 10 times the (usable) capacity, depending on application workloads. The minimum numbers shown here reflect the efficiency rating that NetApp guarantees for VMware infrastructure workloads.

As an example, for a minimum-size starting solution, a configuration with four small (H300S) storage nodes has a total minimum effective capacity of 14.4TB. This configuration also supports a minimum of

150,000 IOPS at a deterministic sub-ms latency. Note that, in both cases, one node's worth of capacity and IOPS has been set aside to reflect N+1 operation.

Rackmount Storage Nodes

This type of storage node is intended for customers whose storage needs (either capacity or IOPS) exceed those of a typical NetApp HCI deployment. These nodes contain 12 SSD drives. The capacity of each node can be either 960GB, 1.92TB, or 3.84TB.

Other than their form factor, there is no difference between how these nodes are deployed or managed in NetApp HCI.

Rackmount Storage Node Model Selection

There is currently only one model of rackmount storage node available for NetApp HCI (Table 15).

Table 15) Rackmount storage node models.

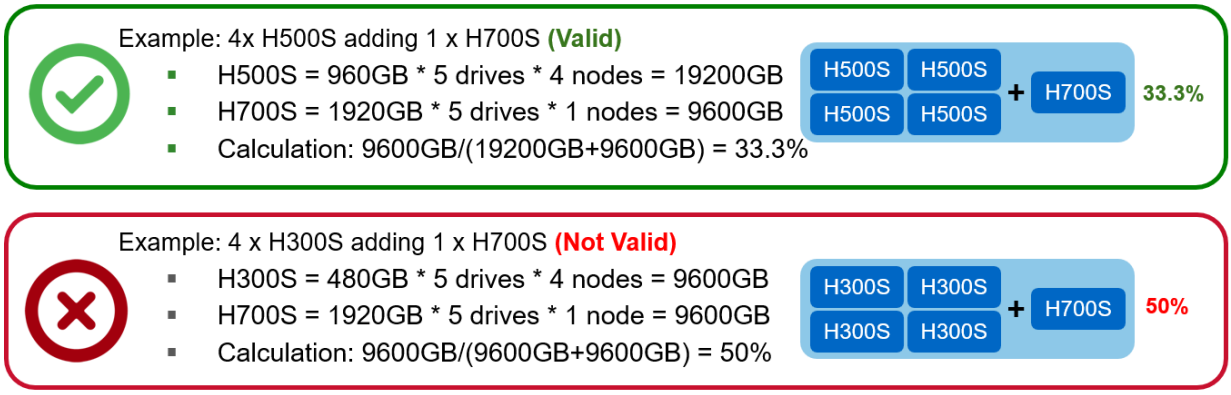
	H610S
Form factor	1RU, full-width
CPU	E5-2620 v4: 8C at 2.1GHz
Boot device	1x 240GB MLC
Networking	2x 10/25GbE SFP28/SFP+ 2x 1GbE RJ45
SSD	12x 960GB, 1.92TB, 3.84TB
IOPS	100,000
Effective block capacity (min) [1]	20TB, 40TB, 80TB

[1] NetApp HCI effective capacity calculations account for NetApp SolidFire Helix data protection, system overhead, and global efficiency, including compression, deduplication, and thin provisioning. Element software customers typically achieve an effective capacity range of 5 to 10 times the (usable) capacity, depending on application workloads. The minimum numbers shown here reflect the efficiency rating that NetApp guarantees for VMware Infrastructure workloads.

Mixing Storage Node Models

You can mix different models of storage nodes in the same cluster as long as no one node represents more than one-third of the total capacity of the cluster.

Figure 32) Calculating mixed storage node models.

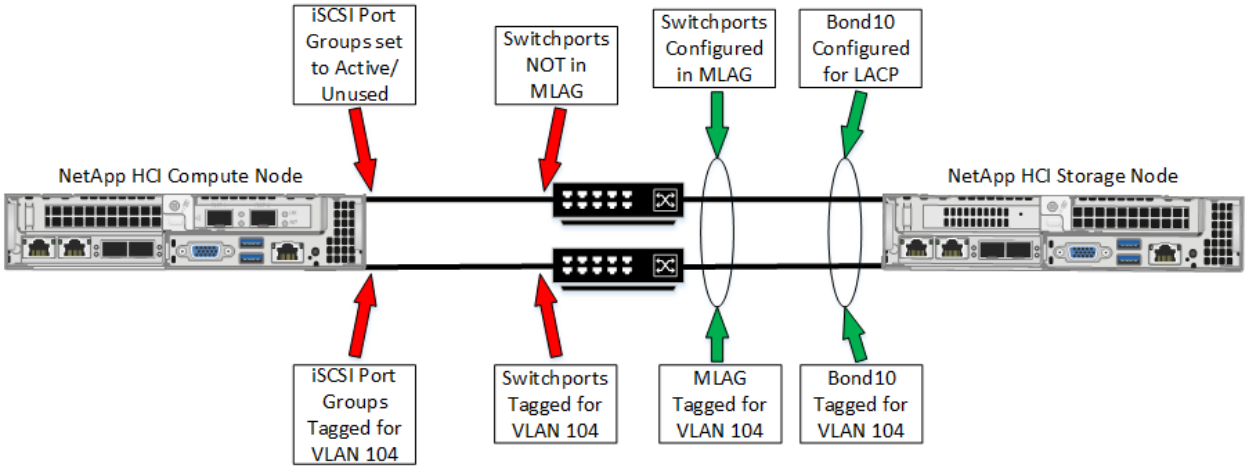


Storage Node Physical Networking

When you design the physical network, port bonding should be configured as follows:

- In this design, both 10/25GbE switch ports facing each NetApp HCI storage node should be bonded together in an MLAG (cross-switch link aggregation) that supports LACP.
- The 10/25GbE switch ports facing each NetApp HCI compute node should not be bonded together in any way. NIC teaming and load balancing for these ports is handled at the vSphere layer in this design.

Figure 33) Switch port bonding for compute and storage nodes.



Note: Only the two 10/25GbE ports on each storage node must be bonded with MLAG/LACP. The two 1Gb management ports do not need to be bonded.

Refer to your switch vendor's documentation for specific information about how to design and configure MLAG port bonding across switches. The details vary greatly because each switch vendor implements MLAG in its own proprietary way.

Second, the network used for iSCSI in this design must not be the native VLAN on either the compute or storage nodes. It must be the same VLAN on both compute and storage nodes, and it must be tagged on both as well.

NFS Physical Storage Design

Network File System (NFS) is a distributed file system protocol that allows a user on a client computer to access files over a network in much the same way that local storage is accessed. In this case, the client computer is an ESXi host, and the storage is provided via NFS from a NetApp ONTAP virtual machine.

The management cluster uses iSCSI-based NetApp SolidFire technology for primary storage and NetApp ONTAP Select for secondary storage. Primary storage in this design refers to all VMs running on the compute nodes. Secondary storage refers to ISOs, virtual machine templates, content libraries, file servers, and similar items.

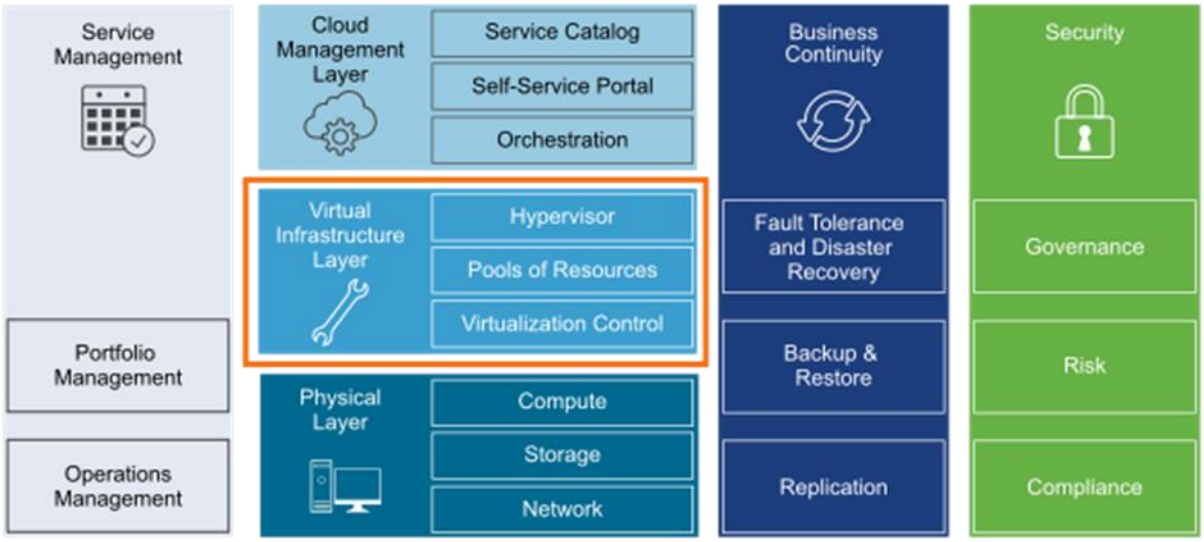
NetApp recommends that you use iSCSI-based HCI storage nodes for all VMs in the design to provide a deterministic experience with SolidFire IOPS QoS technology.

10 Virtual Infrastructure Design

The virtual infrastructure design includes the software components that make up the virtual infrastructure layer and that support business continuity for the SDDC.

These components include the software products that provide the virtualization platform hypervisor, virtualization management, storage virtualization, network virtualization, backup, and disaster recovery. VMware products in this layer include VMware vSphere and NSX for vSphere.

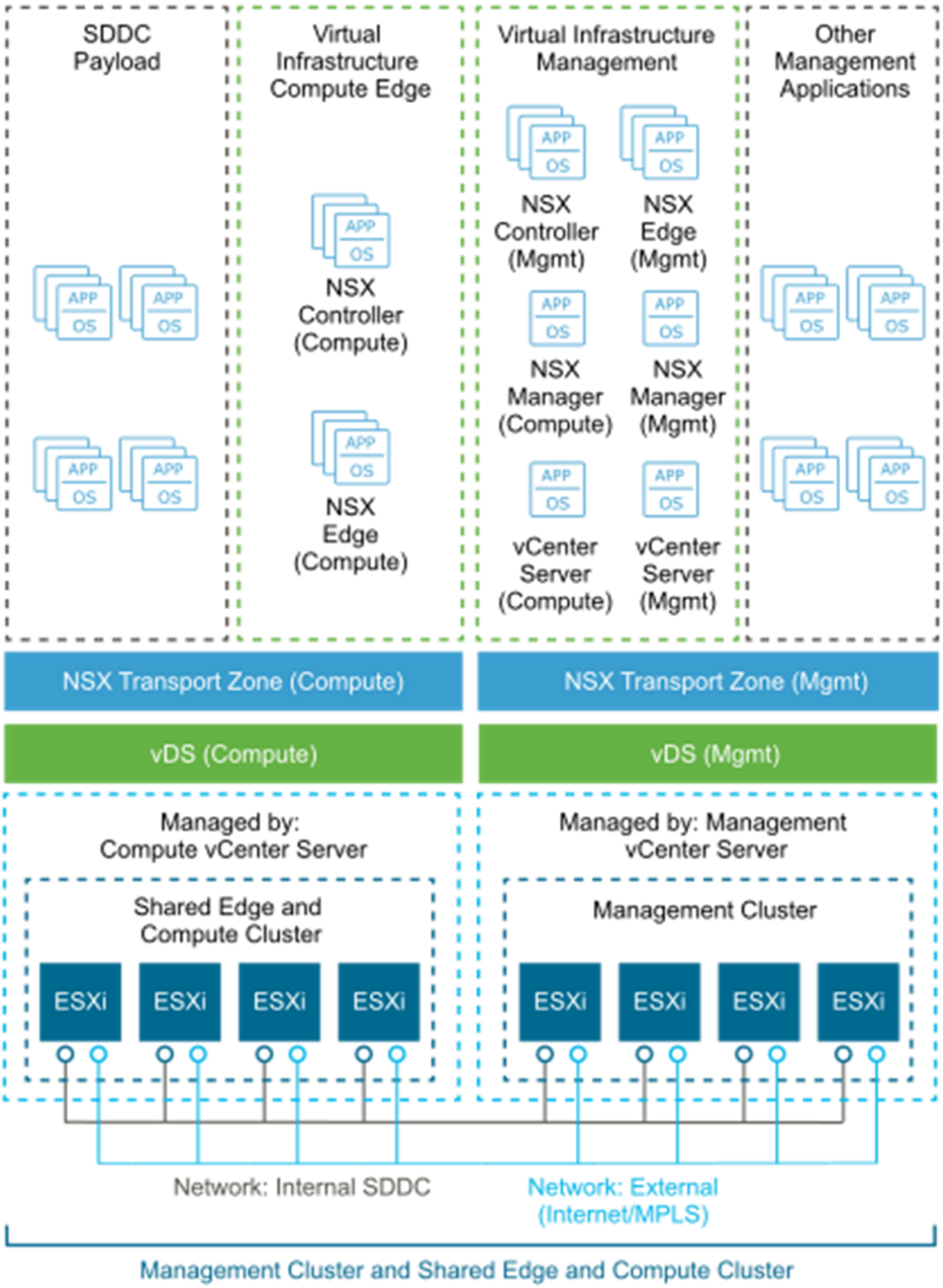
Figure 34) Virtual infrastructure layer in the SDDC.



10.1 Virtual Infrastructure Design Overview

The SDDC virtual infrastructure consists of two regions. Each region includes a management workload domain that contains the management cluster and a virtual infrastructure workload domain that contains the shared edge and compute cluster.

Figure 35) SDDC logical design.



10.2 Management Cluster

The management cluster runs the virtual machines that manage the SDDC. These virtual machines host vCenter Server, NSX Manager, NSX Controller, vRealize operations, vRealize Log Insight, vRealize Automation, Site Recovery Manager, and other shared management components. All management, monitoring, and infrastructure services are provisioned to a vSphere cluster that provides high availability for these critical services. Permissions on the management cluster limit access to administrators only. This protects the virtual machines running the management, monitoring, and infrastructure services.

10.3 Shared Edge and Compute Cluster

The virtual infrastructure design uses a shared edge and compute cluster. The shared cluster combines the characteristics of typical edge and compute clusters into a single cluster. It is possible to separate these in the future if necessary.

This cluster provides the following main functions:

- Supports on-ramp and off-ramp connectivity to physical networks
- Connects with VLANs in the physical world
- Hosts the SDDC tenant virtual machines

The shared edge and compute cluster connects the virtual networks (overlay networks) provided by NSX for vSphere and the external networks. An SDDC can mix different types of compute-only clusters and provide separate compute pools for different types of SLAs.

10.4 ESXi Design

The ESXi design includes design decisions for boot options, user access, and the virtual machine swap configuration.

ESXi Hardware Requirements

ESXi hardware requirements are described in section 9.1 “Physical Design Fundamentals.” The following design outlines the ESXi configuration.

ESXi Boot Disk and Scratch Configuration

For new installations of ESXi, the installer creates a 4GB VFAT scratch partition. ESXi uses this scratch partition to store log files persistently. By default, the VM-support output, which VMware uses to troubleshoot issues on the ESXi host, is also stored on the scratch partition.

Table 16) ESXi boot disk design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-ESXi-001	Install and configure all ESXi hosts to boot from internal M.2 SATA SSD devices by using the NetApp Deployment Engine.	Automatically configured and optimized by the NetApp Deployment Engine.	None

ESXi Host Access

After installation, ESXi hosts are added to a vCenter Server system and managed through that system.

Direct access to the host console is still available and is most commonly used for troubleshooting. You can access ESXi hosts directly with one of these three methods:

- **Direct Console User Interface (DCUI).** Graphical interface on the console. Allows basic administrative controls and troubleshooting options.
- **ESXi Shell.** A Linux-style bash login on the ESXi console.
- **Secure Shell (SSH) access.** Remote command-line console access.
- **VMware host client.** HTML5-based client that has a similar interface to vSphere Web Client but is only used to manage single ESXi hosts. You use the VMware host client to conduct emergency management when vCenter Server is temporarily unavailable.

You can enable or disable each method. By default, ESXi Shell and SSH are disabled to secure the ESXi host. The DCUI is disabled only if Strict Lockdown mode is enabled.

ESXi User Access

By default, root is the only user who can log in to an ESXi host directly. However, you can add ESXi hosts to an Active Directory domain. After the ESXi host has been added to an Active Directory domain, access can be granted through Active Directory groups. Auditing logins into the ESXi host also becomes easier.

Table 17) ESXi user access design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-ESXi-002	Add each ESXi host to the Active Directory domain for the region in which the ESXi host resides.	Using Active Directory membership allows greater flexibility in granting access to ESXi hosts. Ensuring that users log in with a unique user account allows greater visibility for auditing.	Adding ESXi hosts to the domain can add some administrative overhead.
SDDC-VI-ESXi-003	Change the default ESX Admins group to the SDDC-Admins Active Directory group. Add ESXi administrators to the SDDC-Admins group following standard access procedures.	Having an SDDC-Admins group is more secure because it removes a known administrative access point. In addition, different groups allow the separation of management tasks.	Additional changes to the ESXi hosts advanced settings are required.

Virtual Machine Swap Configuration

When a virtual machine is powered on, the system creates a VMkernel swap file to serve as a backing store for the virtual machine's RAM contents. The default swap file is stored in the same location as the virtual machine's configuration file. This simplifies the configuration; however, it can cause excessive unneeded replication traffic.

You can reduce the amount of traffic that is replicated by changing the swap file location to a user-configured location on the ESXi host. However, it can take longer to perform VMware vSphere vMotion operations when the swap file has to be recreated.

ESXi Design Decisions about NTP and Lockdown Mode Configuration

Table 18) Other ESXi host design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-ESXi-004	Configure all ESXi hosts to synchronize time with the central NTP servers.	Required because the deployment of a vCenter Server appliance on an ESXi host might fail if the host is not using NTP.	All firewalls located between the ESXi host and the NTP servers must allow NTP traffic on the required network ports.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-ESXi-005	Enable Lockdown mode on all ESXi hosts.	Increases the security of ESXi hosts by requiring administrative operations to be performed only from vCenter Server.	Lockdown mode settings are not part of host profiles and must be manually enabled on all hosts.

10.5 vCenter Server Design

The vCenter Server design includes both the design for the vCenter Server instance and the VMware Platform Services Controller instance.

A Platform Services Controller groups a set of infrastructure services including vCenter Single Sign-On, License Service, Lookup Service, and VMware Certificate Authority (VMCA). You can deploy the Platform Services Controller and the associated vCenter Server system on the same virtual machine (vCenter Server with an embedded Platform Services Controller). You can also deploy it on a different virtual machines (vCenter Server with an external Platform Services Controller).

- **vCenter Server Deployment.** The design decisions for vCenter Server deployment include the number of vCenter Server and Platform Services Controller instances, the type of installation, and the topology.
- **vCenter Server Networking**. As specified in the physical networking design, all vCenter Server systems must use static IP addresses and host names. The IP addresses must have valid internal DNS registration, including reverse name resolution.
- **vCenter Server Redundancy.** Protecting the vCenter Server system is important because it is the central point of management and monitoring for the SDDC. You protect vCenter Server according to the maximum downtime tolerated and whether failover automation is required.
- **vCenter Server Appliance Sizing.** Size resources and storage for the Management vCenter Server Appliance and the Compute vCenter Server Appliance must be specified to accommodate the expected number of management virtual machines in the SDDC.
- **vSphere Cluster Design.** The cluster design must take into account the workload that the cluster handles. Different cluster types in this design have different characteristics.
- **vCenter Server Customization.** vCenter Server supports a rich set of customization options, including monitoring, virtual machine fault tolerance, and so on.

By default, vSphere uses TLS/SSL certificates that are signed by VMCA (VMware Certificate Authority). These certificates are not trusted by end-user devices or browsers.

vCenter Server Deployment

The design decisions for vCenter Server deployment include the number of vCenter Server and Platform Services Controller instances, the type of installation, and the topology.

Table 19) vCenter Server design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-001	Deploy two vCenter Server systems: • One vCenter Server supporting the SDDC management components. • One vCenter Server supporting the edge	Isolates vCenter Server failures to management or compute workloads. Isolates vCenter Server operations between management and compute. Supports a scalable cluster design in which the management	Requires licenses for each vCenter Server instance.

Decision ID	Design Decision	Design Justification	Design Implication
	components and compute workloads.	components can be reused as additional compute resources are needed for the SDDC. Simplifies capacity planning for compute workloads by eliminating management workloads from consideration in the Compute vCenter Server. Improves the ability to upgrade the vSphere environment and related components by providing for explicit separation of maintenance windows: • Management workloads remain available while workloads in compute are being addressed. • Compute workloads remain available while workloads in management are being addressed. Ability to have clear separation of roles and responsibilities to make sure that only those administrators with proper authorization can attend to the management workloads. Facilitates quicker troubleshooting and problem resolution. Simplifies disaster recovery operations by supporting a clear demarcation between recovery of the management components and compute workloads. Enables the use of two NSX Manager instances, one for the management cluster and the other for the shared edge and compute cluster. Network separation of the clusters in the SDDC allows isolation of potential network issues.	

You can install vCenter Server as a Windows-based system or deploy the Linux-based VMware vCenter Server appliance. The appliance is preconfigured, enables fast deployment, and potentially results in reduced Microsoft licensing costs.

Table 20) vCenter Server platform design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-002	Deploy all vCenter Server instances as Linux-based vCenter Server appliances.	Allows rapid deployment, enables scalability, and reduces Microsoft licensing costs.	Operational staff might need Linux experience to troubleshoot the Linux-based appliances.

Platform Services Controller Design Decision Background

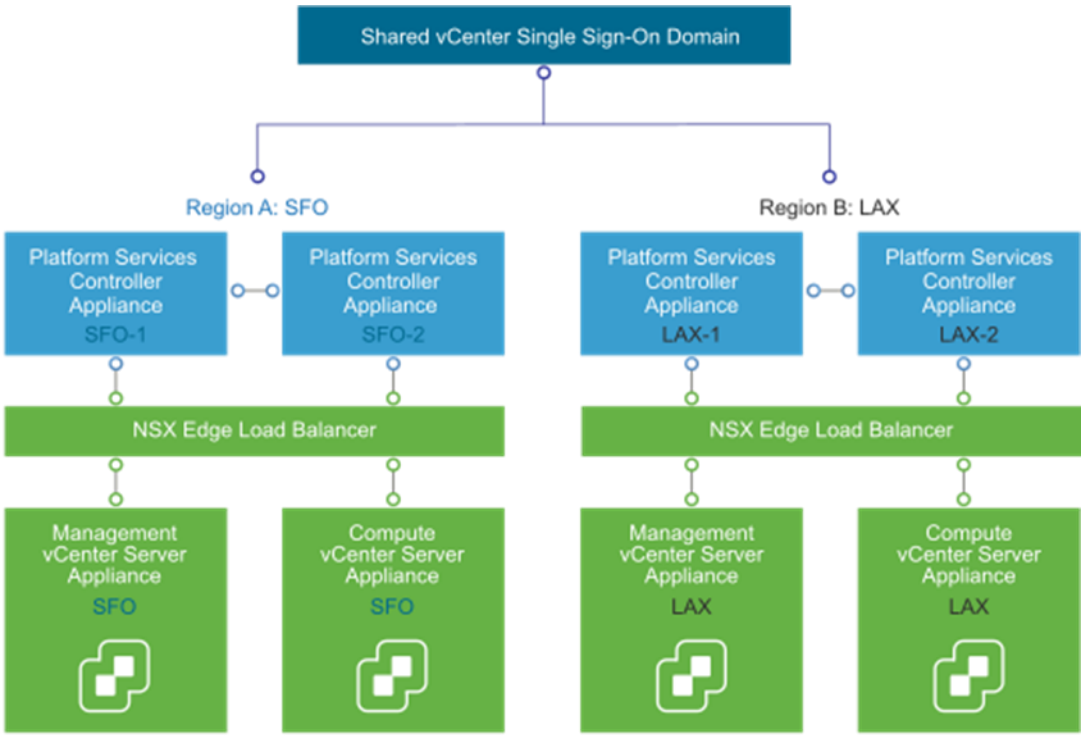
vCenter Server supports installation with an embedded Platform Services Controller (embedded deployment) or with an external Platform Services Controller.

- In an embedded deployment, vCenter Server and the Platform Services Controller run on the same virtual machine. Embedded deployments are recommended for standalone environments with only one vCenter Server system.
- Environments with an external Platform Services Controller can have multiple vCenter Server systems. The vCenter Server systems can use the same Platform Services Controller services. For example, several vCenter Server systems can use the same instance of vCenter Single Sign-On for authentication.
- You might need to replicate the Platform Services Controller instance with other Platform Services Controller instances; or the solution might include more than one vCenter Single Sign-On instance. In that case, you can deploy multiple external Platform Services Controller instances on separate virtual machines.

Table 21) Platform Services Controller design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-003	Deploy each vCenter Server with an external Platform Services Controller.	External Platform Services Controller instances are required for replication between Platform Services Controller instances.	The number of VMs that must be managed increases.
SDDC-VI-VC-004	Join all Platform Services Controller instances to a single vCenter Single Sign-On domain.	When all Platform Services Controller instances are joined into a single vCenter Single Sign-On domain, they can share authentication and license data across all components and regions.	Only one Single Sign-On domain exists.
SDDC-VI-VC-005	Create a ring topology for the Platform Services Controllers.	By default, Platform Services Controllers only replicate with one other Platform Services Controller, creating a single point of failure for replication. A ring topology makes sure that each Platform Services Controller has two replication partners and eliminates any single point of failure.	Command-line interface commands must be used to configure the ring replication topology.
SDDC-VI-VC-006	Use an NSX ESG as a load balancer for the Platform Services Controllers.	Using a load balancer increases the availability of the Platform Services Controller instances for all applications.	Configuring the load balancer and repointing vCenter Server to the load balancers VIP creates administrative overhead.

Figure 36) vCenter Server and Platform Services Controller deployment model.



vCenter Server Networking

As specified in the physical networking design, all vCenter Server systems must use static IP addresses and host names. The IP addresses must have valid internal DNS registration, including reverse name resolution.

The vCenter Server systems must maintain network connections to the following components:

- Systems running vCenter Server add-on modules
- Each ESXi host

vCenter Server Redundancy

Protecting the vCenter Server system is important because it is the central point of management and monitoring for the SDDC. You protect vCenter Server according to the maximum downtime tolerated and whether failover automation is required.

Table 22 lists methods available for protecting the vCenter Server system and the vCenter Server Appliance.

Table 22) Methods for protecting vCenter Server system and the vCenter Server appliance.

Redundancy Method	Protects vCenter Server (Windows)	Protects Platform Services Controller (Windows)	Protects vCenter Server (Virtual Appliance)	Protects Platform Services Controller (Virtual Appliance)
Automated protection using vSphere HA	Yes	Yes	Yes	Yes

Redundancy Method	Protects vCenter Server (Windows)	Protects Platform Services Controller (Windows)	Protects vCenter Server (Virtual Appliance)	Protects Platform Services Controller (Virtual Appliance)
Manual configuration and manual failover; for example, using a cold standby	Yes	Yes	Yes	Yes
HA cluster with external load balancer	Not available	Yes	Not available	Yes
vCenter Server HA	Not available	Not available	Yes	Not available

Table 23) vCenter Server protection design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-007	Protect all vCenter Server and Platform Services Controller appliances by using vSphere HA.	Supports availability objectives for vCenter Server appliances without a required manual intervention during a failure event.	vCenter Server becomes unavailable during a vSphere HA failover.

vCenter Server Appliance Sizing

Size resources and storage for the Management vCenter Server appliance and the Compute vCenter Server appliance to provide enough resources for the expected number of management virtual machines in the SDDC.

Table 24) Logical specification for the Management vCenter Server appliance.

Attribute	Specification
vCenter Server version	6.5 (vCenter Server appliance)
Physical or virtual system	Virtual (appliance)
Appliance size	Small (up to 100 hosts/1,000 VMs)
Platform Services Controller	External
Number of CPUs	4
Memory	16GB
Disk space	290GB

Table 25) Logical specification for the Compute vCenter Server appliance.

Attribute	Specification
vCenter Server version	6.5 (vCenter Server Appliance)
Physical or virtual system	Virtual (appliance)
Appliance size	Large (up to 1,000 hosts/10,000 VMs)

Attribute	Specification
Platform Services Controller	External
Number of CPUs	16
Memory	32GB
Disk space	640GB

Table 26) vCenter Server appliance sizing design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-008	Deploy Management vCenter Server appliances of a small deployment size or larger.	Depending on the number of management VMs that are running, a small vCenter Server appliance might be needed.	If the size of the management environment changes, you might need to increase the size of the vCenter Server appliance.
SDDC-VI-VC-009	Deploy Compute vCenter Server appliances of a large deployment size or larger.	Depending on the number of compute workloads and NSX Edge devices, running a vCenter Server appliance of a large size is best.	As the compute environment expands, you can resize to the extra-large size or add vCenter Server instances.

vSphere Cluster Design

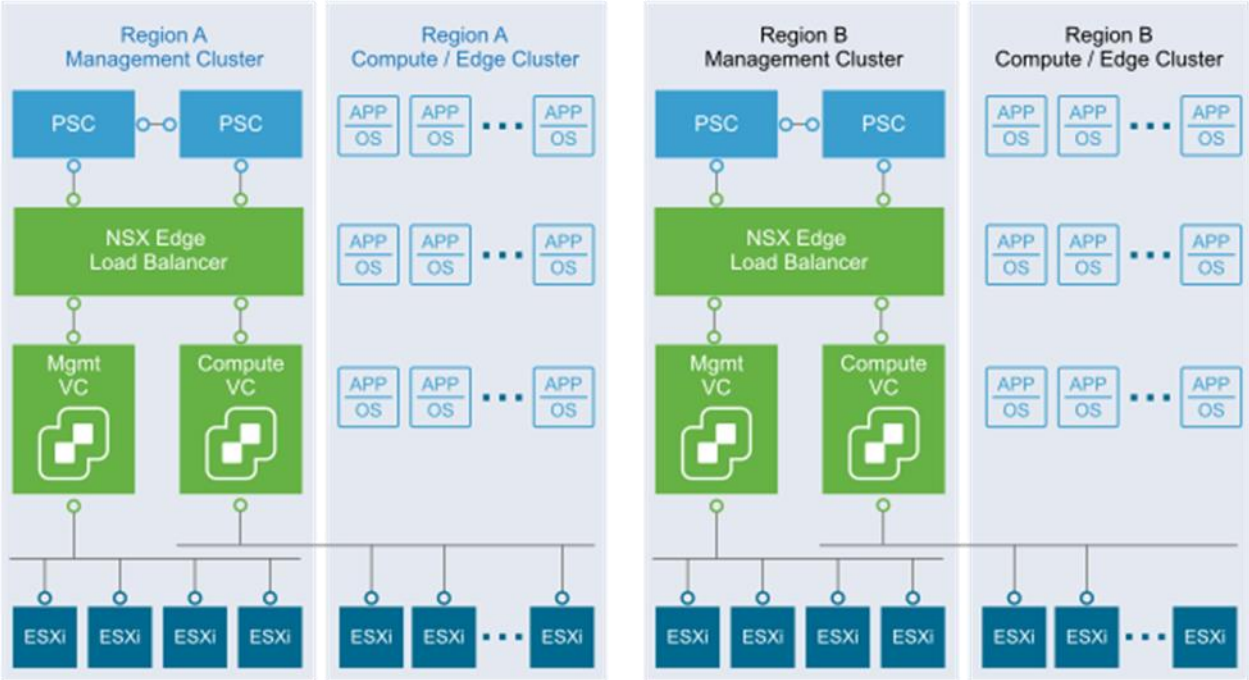
The cluster design must take into account the workload that the cluster handles. Different cluster types in this design have different characteristics.

vSphere Cluster Design Decision Background

The following heuristics help with cluster design decisions.

- Decide whether to use fewer, larger ESXi hosts, or more, smaller ESXi hosts.
  - A scale-up cluster has fewer, larger ESXi hosts.
  - A scale-out cluster has more, smaller ESXi hosts.
- Compare the capital costs of purchasing fewer, larger ESXi hosts with the costs of purchasing more, smaller ESXi hosts.
- Evaluate the operational costs of managing a few ESXi hosts with the costs of managing more ESXi hosts.
- Consider the purpose of the cluster.
- Consider the total number of ESXi hosts and cluster limits.

Figure 37) vSphere logical cluster layout.



vSphere High-Availability Design

VMware vSphere high availability (vSphere HA) protects your virtual machines in case of ESXi host failure by restarting virtual machines on other hosts in the cluster when an ESXi host fails.

vSphere HA Design Basics

During configuration of the cluster, the ESXi hosts elect a master ESXi host. The master ESXi host communicates with the vCenter Server system and monitors the virtual machines and secondary ESXi hosts in the cluster.

The master ESXi host detects different types of failure:

- ESXi host failure, for example an unexpected power failure
- ESXi host network isolation or connectivity failure
- Loss of storage connectivity
- Problems with virtual machine OS availability

Table 27) vSphere HA design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-010	Use vSphere HA to protect all clusters against failures.	vSphere HA supports a robust level of protection for both ESXi host and virtual machine availability.	You must provide enough resources on the remaining hosts so that you can migrate virtual machines to those hosts in the event of a host outage.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-011	Set vSphere HA Host Isolation Response to Leave Powered On.	Prevents problems with workloads in case of a false positive network isolation detection.	None

vSphere HA Admission Control Policy Configuration

The vSphere HA Admission Control Policy allows an administrator to configure how the cluster determines available resources. In a smaller vSphere HA cluster, a larger proportion of the cluster resources are reserved to accommodate ESXi host failures, based on the selected policy.

The following policies are available:

- **Host failures the cluster tolerates.** vSphere HA makes sure that a specified number of ESXi hosts can fail and sufficient resources remain in the cluster to fail over all the virtual machines from those ESXi hosts.
- **Percentage of cluster resources reserved.** vSphere HA reserves a specified percentage of aggregate CPU and memory resources for failover.
- **Specify failover hosts.** When an ESXi host fails, vSphere HA attempts to restart its virtual machines on any of the specified failover ESXi hosts. If restart is not possible—for example, if the failover ESXi hosts have insufficient resources or have failed as well—then vSphere HA attempts to restart the virtual machines on other ESXi hosts in the cluster.

vSphere Cluster Workload Design

The cluster workload design defines the vSphere clusters, cluster size and high-availability configuration, and the workloads that they handle.

Table 28) vSphere cluster workload design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-012	Create a single management cluster for each region. This cluster contains all management ESXi hosts.	Simplifies configuration by isolating management workloads from compute workloads. Makes sure that compute workloads have no impact on the management stack. You can add ESXi hosts to the cluster as needed.	Management of multiple clusters and vCenter Server instances increases operational overhead.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-013	Create a shared edge and compute cluster for each region. This cluster hosts compute workloads, NSX Controller instances, and associated NSX Edge gateway devices used for compute workloads.	Simplifies configuration and minimizes the number of ESXi hosts required for initial deployment. Makes sure that the management stack has no impact on compute workloads. You can add ESXi hosts to the cluster as needed.	Management of multiple clusters and vCenter Server instances increases operational overhead. Due to the shared nature of the cluster, when compute workloads are added, the cluster must be scaled out to maintain a high level of network performance. Due to the shared nature of the cluster, resource pools are required to makes sure that edge components receive all required resources.

Management Cluster Design

The management cluster design determines the number of hosts and vSphere HA settings for the management cluster.

Table 29) Management cluster design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-014	Create a management cluster in Region A with a minimum of four ESXi hosts.	Allocating four ESXi hosts provides full redundancy. Having four ESXi hosts guarantees redundancy during maintenance operations.	Additional ESXi host resources are required for redundancy.
SDDC-VI-VC-015	In Region B, create a management cluster with four ESXi hosts.	Allocating four ESXi hosts provides full redundancy for the cluster. Having four ESXi hosts guarantees redundancy during maintenance operations.	Additional ESXi host resources are required for redundancy.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-016	Configure admission control for one ESXi host failure and percentage-based failover capacity.	Using the percentage-based reservation works well in situations in which virtual machines have varying and sometimes significant CPU or memory reservations. vSphere 6.5 automatically calculates the reserved percentage based on ESXi host failures to tolerate and the number of ESXi hosts in the cluster.	In a four ESXi host management cluster, only the resources of three ESXi hosts are available for use.
SDDC-VI-VC-020	Create a host profile for the management cluster.	Using host profiles simplifies configuration of ESXi hosts and makes sure that settings are uniform across the cluster.	Any time an authorized change is made to an ESXi host, the host profile must be updated to reflect the change or the status shows as noncompliant.

Table 30 summarizes the attributes of the management cluster logical design.

Table 30) Management cluster logical design background.

Attribute	Specification
Number of ESXi hosts required to support management hosts with no overcommitment	2
Number of ESXi hosts recommended due to operational constraints (ability to take a host offline without sacrificing high-availability capabilities)	4
Capacity for ESXi host failures per cluster	25% reserved CPU RAM

Shared Edge and Compute Cluster Design

Tenant workloads run on the ESXi hosts in the shared edge and compute cluster. Due to the shared nature of the cluster, NSX Controller instances and edge devices run in this cluster. The design decisions determine the number of ESXi hosts and vSphere HA settings and several other characteristics of the shared edge and compute cluster.

Table 31) Shared edge and compute cluster design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-022	Create a shared edge and compute cluster for the NSX Controller instances and NSX Edge gateway devices.	An NSX Manager instance requires a 1:1 relationship with a vCenter Server system.	Each time you provision a compute vCenter Server system, a new NSX Manager instance is required. Set antiaffinity rules to keep each controller on a separate ESXi host. A 4-node cluster allows maintenance while making sure that the three controllers remain on separate ESXi hosts.
SDDC-VI-VC-023	Configure admission control for one ESXi host failure and percentage-based failover capacity.	vSphere HA protects the NSX Controller instances and Edge services gateway devices in the event of an ESXi host failure. vSphere HA powers on virtual machines from the failed ESXi hosts on any remaining ESXi hosts.	Only a single ESXi host failure is tolerated before potential resource contention.
SDDC-VI-VC-025	In Region A, create a shared edge and compute cluster with a minimum of four ESXi hosts.	Allocating four ESXi hosts provides full redundancy within the cluster. Having four ESXi guarantees redundancy during outages or maintenance operations.	Four ESXi hosts is the smallest starting point for the shared edge and compute cluster for redundancy and performance, thus increasing cost.
SDDC-VI-VC-026	In Region B, create a shared edge and compute cluster with a minimum of four hosts.	Three NSX Controller instances are required for sufficient redundancy and majority decisions. One ESXi host is available for failover and to allow for scheduled maintenance.	Four ESXi hosts is the smallest starting point for the shared edge and compute cluster for redundancy and performance, increasing complexity relative to a 3-node cluster.
SDDC-VI-VC-027	Set up VLAN-backed port groups for external access and management on the shared edge and compute cluster ESXi hosts.	Edge gateways require access to the external network in addition to the management network.	VLAN-backed port groups must be configured with the correct number of ports or with elastic port allocation.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-011	Set up VLAN-backed port groups for external access and management on the shared edge and compute cluster ESXi hosts.	Edge gateways need access to the external network in addition to the management network.	VLAN-backed port groups must be configured with the correct number of ports, or with elastic port allocation.
SDDC-VI-VC-028	Create a resource pool for the required SDDC NSX Controller instances and edge appliances with a CPU share level of High, a memory share of Normal, and a 16GB memory reservation.	The NSX components control all network traffic in and out of the SDDC and update route information for inter-SDDC communication. In a contention situation it is imperative that these virtual machines receive all the resources required.	During contention, SDDC NSX components receive more resources than all other workloads. Therefore, monitoring and capacity management must be a proactive activity.
SDDC-VI-VC-029	Create a resource pool for all user NSX Edge devices with a CPU share value of Normal and a memory share value of Normal.	NSX Edge instances for users created by vRealize Automation support functions such as load balancing for user workloads. These edge devices do not support the entire SDDC. Therefore, they receive a smaller amount of resources during contention.	During contention, these NSX Edge instances receive fewer resources than the SDDC edge devices. As a result, monitoring and capacity management must be a proactive activity.
SDDC-VI-VC-030	Create a resource pool for all user virtual machines with a CPU share value of Normal and a memory share value of Normal.	Creating virtual machines outside of a resource pool has a negative effect on all other virtual machines during contention. In a shared edge and compute cluster, the SDDC edge devices must be guaranteed resources above all other workloads in order not to impact network connectivity. Setting the share values to normal gives the SDDC edges more shares of resources during contention, ensuring that network traffic is not impacted.	During contention, user-workload virtual machines might be starved for resources and experience poor performance. It is critical that monitoring and capacity management must be a proactive activity and that capacity is added or a dedicated edge cluster is created before contention occurs.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-033	Create a host profile for the shared edge and compute cluster.	Using host profiles simplifies configuration of ESXi hosts and makes sure that settings are uniform across the cluster.	Any time an authorized change is made to an ESXi host, the host profile must be updated to reflect the change, or the status shows noncompliance.

Table 32 summarizes the attributes of the shared edge and compute cluster logical design. The number of VMs on the shared edge and compute cluster start low but grow quickly as user workloads are created.

Table 32) Shared edge and compute cluster logical design background.

Attribute	Specification
Minimum number of ESXi hosts required to support the shared edge and compute cluster	3
Number of ESXi hosts recommended due to operational constraints (ability to take an ESXi host offline without sacrificing high-availability capabilities)	4
Capacity for ESXi host failures per cluster	25% reserved CPU RAM

Compute Cluster Design

As the SDDC expands, additional compute-only clusters can be configured. Tenant workloads run on the ESXi hosts in the compute cluster instances. Multiple compute clusters are managed by the Compute vCenter Server instance. The design determines vSphere HA settings for the compute cluster.

Table 33) Compute cluster design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-035	Configure vSphere HA to use percentage-based failover capacity to ensure n+1 availability.	Using explicit host failover limits the total available resources in a cluster.	The resources of one ESXi host in the cluster are reserved, which can cause provisioning to fail if resources are exhausted.

vCenter Server Customization

vCenter Server supports a rich set of customization options, including monitoring, virtual machine fault tolerance, and so on.

Virtual Machine and Application Monitoring Service

When enabled, the virtual machine and application monitoring service, which uses VMware Tools, evaluates whether each virtual machine in the cluster is running. The service checks for regular heartbeats and I/O activity from the VMware Tools process running on guests. If the service receives no heartbeats or determines that I/O activity has stopped, it is likely that the guest operating system has failed or that VMware Tools has not been allocated time for heartbeats or I/O activity. In this case, the service determines that the virtual machine has failed and reboots the virtual machine.

Enable virtual machine monitoring for automatic restart of a failed virtual machine. The application or service running on the virtual machine must be capable of restarting successfully after a reboot, or the virtual machine restart is not sufficient.

Table 34) Monitor virtual machines design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-037	Enable virtual machine monitoring for each cluster.	virtual machine monitoring provides adequate in-guest protection for most virtual machine workloads.	There is no downside to enabling virtual machine monitoring.
SDDC-VI-VC-038	Create virtual machine groups for use in start-up rules in the management, shared edge, and compute clusters.	By creating virtual machine groups, you can create rules to configure the start-up order of the SDDC management components.	Creating the groups is a manual task and adds administrative overhead.
SDDC-VI-VC-039	Create virtual machine rules to specify the start-up order of the SDDC management components.	The rules enforce the start-up order of virtual machine groups to ensure the correct start-up order of the SDDC management components.	Creating these rules is a manual task and adds administrative overhead.

VMware vSphere Distributed Resource Scheduling

vSphere Distributed Resource Scheduling (DRS) provides load balancing of a cluster by migrating workloads from heavily loaded ESXi hosts to less utilized ESXi hosts in the cluster. vSphere DRS supports manual and automatic modes.

- **Manual.** Recommendations are made but an administrator must confirm the changes.
- **Automatic.** Automatic management can be set to five different levels. At the lowest setting, workloads are placed automatically at power on and only migrated to fulfill certain criteria, such as entering maintenance mode. At the highest level, any migration that provides even a slight improvement in balancing is executed.

Table 35) vSphere Distributed Resource Scheduling design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-040	Enable vSphere DRS on all clusters and set it to Fully Automated, with the default setting (Medium).	The default settings provide the best trade-off between load balancing and excessive migration with vSphere vMotion events.	If a vCenter Server outage occurs, mapping from virtual machines to ESXi hosts might be more difficult to determine.

Enhanced vMotion Compatibility

Enhanced vMotion Compatibility (EVC) works by masking certain features of newer CPUs to allow migration between ESXi hosts that contain older CPUs. EVC works only with CPUs from the same manufacturer, and there are limits to the version difference gaps between the CPU families.

If you set EVC during cluster creation, you can add ESXi hosts with newer CPUs at a later date without disruption. You can use EVC for a rolling upgrade of all hardware with zero downtime.

Set EVC to the highest level possible with the CPUs currently in use.

Table 36) VMware Enhanced vMotion Compatibility design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-045	Enable EVC on all clusters. Set the EVC mode to the lowest available setting supported for the hosts in the cluster.	Allows cluster upgrades without virtual machine downtime.	You can enable EVC only if clusters contain hosts with CPUs from the same vendor.

10.6 Virtualization Network Design

A well-designed network helps an organization meet its business goals. It prevents unauthorized access and provides timely access to business data. This network virtualization design uses vSphere and VMware NSX for vSphere to implement virtual networking.

- **Virtual Network Design Guidelines.** This VVD follows high-level network design guidelines and networking best practices.
- **Virtual Switches.** Virtual switches simplify the configuration process by providing a single-pane-of-glass view for performing virtual network management tasks.
- **NIC Teaming.** You can use NIC teaming to increase the network bandwidth available in a network path and to provide the redundancy that supports higher availability.
- **Network I/O Control.** When Network I/O Control is enabled, the distributed switch allocates bandwidth for the following system traffic types.
  - **VXLAN.** VXLAN enables you to create isolated, multitenant broadcast domains across data center fabrics and enables you to create elastic, logical networks that span physical network boundaries.
  - **vMotion TCP/IP Stack.** Use the vMotion TCP/IP stack to isolate traffic for vMotion and to assign a dedicated default gateway for vMotion traffic.

Virtual Network Design Guidelines

This VVD follows high-level network design guidelines and networking best practices.

Design Goals

The high-level design goals apply regardless of your environment.

- **Meet diverse needs.** The network must meet the diverse needs of many different entities in an organization. These entities include applications, services, storage, administrators, and users.
- **Reduce costs.** Reducing costs is one of the easier goals to achieve in the vSphere infrastructure. Server consolidation alone reduces network costs by reducing the number of required network ports and NICs, but a more efficient network design is desirable. For example, configuring two 10GbE NICs with VLANs might be more cost effective than configuring a dozen 1GbE NICs on separate physical networks.
- **Boost performance.** You can improve performance and decrease the time required to perform maintenance by providing sufficient bandwidth, which reduces contention and latency.
- **Improve availability.** A well-designed network improves availability, typically by providing network redundancy.

- **Support security.** A well-designed network supports an acceptable level of security through controlled access (where required) and isolation (where necessary).
- **Enhance infrastructure functionality.** You can configure the network to support vSphere features such as vSphere vMotion, vSphere High Availability, and vSphere Fault Tolerance.

Best Practices

Follow these networking best practices throughout your environment:

- Separate network services from one another to achieve greater security and better performance.
- Use Network I/O Control and traffic shaping to guarantee bandwidth to critical virtual machines. During network contention, these critical VMs receive a higher percentage of available bandwidth.
- Separate network services on a single vSphere Distributed Switch by attaching them to port groups with different VLAN IDs.
- Keep vSphere vMotion traffic on a separate network. When migration with vMotion occurs, the contents of the guest operating system's memory are transmitted over the network. You can put vSphere vMotion on a separate network by using a dedicated vSphere vMotion VLAN.
- When using pass-through devices with Linux kernel version 2.6.20 or an earlier guest OS, avoid MSI and MSI-X modes. These modes have a significant effect on performance.
- For best performance, use VMXNET3 virtual NICs.
- Make sure that physical network adapters that are connected to the same vSphere standard switch or vSphere Distributed Switch are also connected to the same physical network.

Network Segmentation and VLANs

Separating different types of traffic is required to reduce contention and latency. Separate networks are also required for access security.

High latency on any network can negatively affect performance. Some components are more sensitive to high latency than others. For example, reducing latency is important on the IP storage and the vSphere Fault Tolerance logging network because latency on these networks can negatively affect the performance of multiple virtual machines.

Depending on the application or service, high latency on specific virtual machine networks can also negatively affect performance. Use information gathered from the current state analysis and from interviews with key stakeholder and SMEs to determine which workloads and networks are especially sensitive to high latency.

Virtual Networks

Determine the number of networks or VLANs that are required depending on the type of traffic.

- vSphere operational traffic:
  - Management
  - vMotion
  - iSCSI
  - NFS storage
  - vSphere Replication
  - VXLAN
- Traffic that supports the organization's services and applications

Virtual Switches

Virtual switches simplify the configuration process by providing a single-pane-of-glass view for performing virtual network management tasks.

Virtual Switch Design Background

A vSphere Distributed Switch offers the following enhancements over standard virtual switches:

- **Centralized management.** Because distributed switches are created and managed centrally on a vCenter Server system, they make the switch configuration more consistent across ESXi hosts. Centralized management saves time, reduces mistakes, and lowers operational costs.
- **Additional features.** Distributed switches offer features that are not available on standard virtual switches. Some of these features can be useful to the applications and services that are running in the organization's infrastructure. For example, NetFlow and port mirroring provide monitoring and troubleshooting capabilities to the virtual infrastructure.

Consider the following caveat for distributed switches:

- Distributed switches are not manageable when vCenter Server is unavailable. vCenter Server therefore becomes a tier 1 application.

Number of Virtual Switches

Create fewer virtual switches, preferably just one. For each type of network traffic, configure a single port group to simplify configuration and monitoring.

Table 37) Virtual switch design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-001	Use vSphere Distributed Switches.	vSphere Distributed Switches simplify management.	Migration from a standard switch to a distributed switch requires a minimum of two physical NICs to maintain redundancy.
SDDC-VI-NET-002	Use a single vSphere Distributed Switch per cluster.	Reduces complexity of the network design. Reduces the size of the fault domain.	Increases the number of vSphere Distributed Switches that must be managed.
SDDC-VI-NET-003	Use an ephemeral binding for the management port group.	Ephemeral binding allows the recovery of the vCenter Server that is managing the vSphere Distributed Switch.	Port-level permissions and controls are lost across power cycles, so no historical context is saved.
SDDC-VI-NET-004	Use a static binding for all nonmanagement port groups.	Static binding enables a virtual machine to connect to the same port on the vSphere Distributed Switch. This allows historical data and port level monitoring.	None

Health Check

The health check service helps identify and troubleshoot the following common configuration errors in vSphere Distributed Switches:

- Mismatched VLAN trunks between an ESXi host and the physical switches it's connected to.
- Mismatched MTU settings between physical network adapters, distributed switches, and physical switch ports.
- Mismatched virtual switch teaming policies for the physical switch port-channel settings.

Health check monitors VLAN, MTU, and teaming policies.

- **VLANs.** Checks whether the VLAN settings on the distributed switch match the trunk port configuration on the connected physical switch ports.
- **MTU.** For each VLAN, health check determines whether the physical access switch port's MTU jumbo frame setting matches the distributed switch MTU setting.
- **Teaming policies.** Health check determines whether the connected access ports of the physical switch that participate in an EtherChannel are paired with distributed ports whose teaming policy is IP hash.

Health check is limited to the access switch port to which the ESXi hosts' NICs connects.

Table 38) vSphere Distributed Switch health check design decisions.

Design ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-002	Enable vSphere Distributed Switch health check on all virtual distributed switches.	vSphere Distributed Switch health check trunks all VLANS to all ESXi hosts that are attached to the vSphere Distributed Switch and makes sure that MTU sizes match the physical network.	You must have a minimum of two physical uplinks to use this feature.

Note: For VLAN and MTU checks, at least two physical NICs for the distributed switch are required. For a teaming policy check, at least two physical NICs and two hosts are required when applying the policy.

Management Cluster Distributed Switches

The management cluster uses a single vSphere Distributed Switch with the following configuration settings.

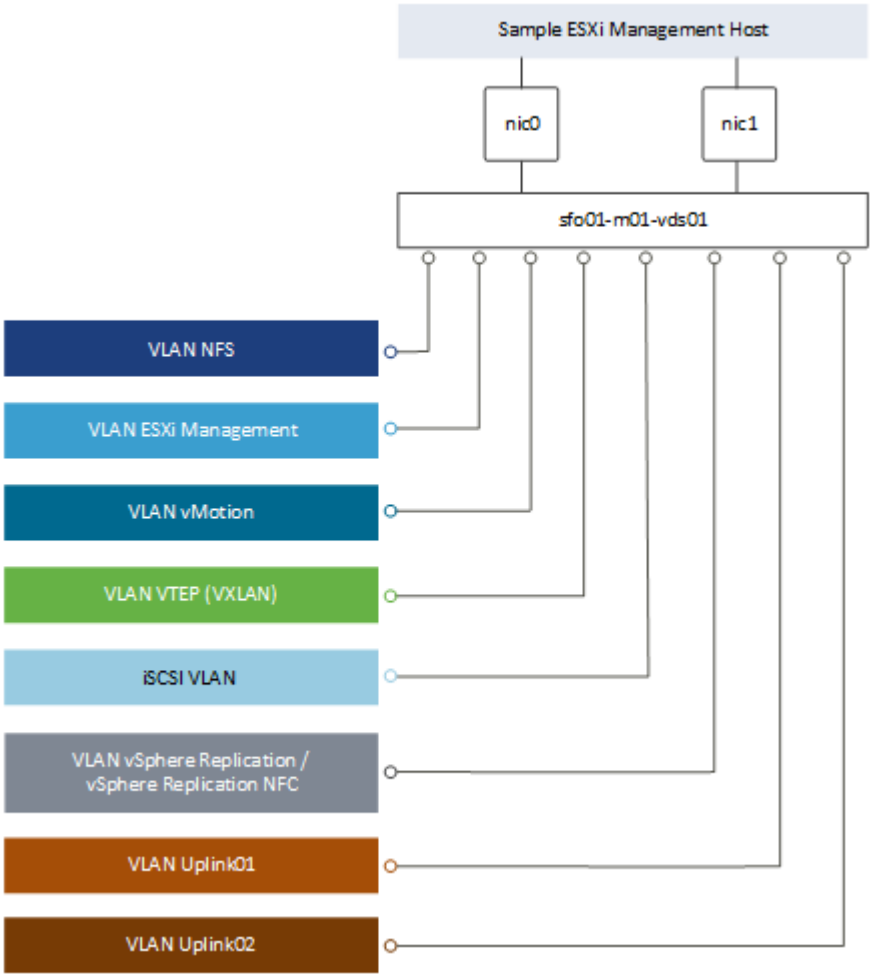
Table 39) Virtual switch for the management cluster.

vSphere Distributed Switch Name	Function	Network I/O Control	Number of Physical NIC Ports	MTU
sfo01-m01-vds01	• ESXi Management • Network IP Storage (NFS) • iSCSI • vSphere vMotion • VTEP • vSphere Replication/vSphere Replication NFC • Uplinks (2) to enable ECMP	Enabled	2	9000

Table 40) vDS-MgmtPort Group configuration settings.

Parameter	Setting
Failover detection	Link status only
Notify switches	Enabled
Failback	Yes
Failover order	Active uplinks: Uplink1, Uplink2

Figure 38) Network switch design for management ESXi Hosts.



This section expands on the logical network design by providing details on the physical NIC layout and physical network attributes.

Table 41) Management virtual switches by physical/virtual NIC.

vSphere Distributed Switch	vmnic	Function
sfo01-m01-vds01	0	Uplink
sfo01-m01-vds01	1	Uplink

Note: The following VLANs are meant as samples. Your actual implementation depends on your environment.

Table 42) Management virtual switch port groups and VLANs.

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
sfo01-m01-vds01	sfo01-m01-vds01-management	Route based on physical NIC load	1, 2	1611

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
sfo01-m01-vds01	sfo01-m01-vds01-vmotion	Route based on physical NIC load	1, 2	1612
sfo01-m01-vds01	sfo01-m01-vds01-iSCSI-A	Use explicit failover order	1	1613
sfo01-m01-vds01	sfo01-m01-vds01-iSCSI-B	Use explicit failover order	2	1613
sfo01-m01-vds01	sfo01-m01-vds01-uplink01	Route based on originating virtual port	1	2711
sfo01-m01-vds01	sfo01-m01-vds01-uplink02	Route based on originating virtual port	2	2712
sfo01-m01-vds01	sfo01-m01-vds01-nfs	Route based on physical NIC load	1, 2	1615
sfo01-m01-vds01	sfo01-m01-vds01-replication	Route based on physical NIC load	1, 2	1616
sfo01-m01-vds01	Auto Generated (NSX VTEP)	Route based on SRC-ID	1, 2	1614

Table 43) Management VMkernel adapter.

vSphere Distributed Switch	Network Label	Connected Port Group	Enabled Services	MTU
sfo01-m01-vds01	Management	sfo01-m01-vds01-management	Management traffic	1500 (Default)
sfo01-m01-vds01	vMotion	sfo01-m01-vds01-vmotion	vMotion traffic	9000
sfo01-m01-vds01	iSCSI-A	sfo01-m01-vds01-iSCSI-A	-	9000
sfo01-m01-vds01	iSCSI-B	sfo01-m01-vds01-iSCSI-B	-	9000
sfo01-m01-vds01	NFS	sfo01-m01-vds01-nfs	-	9000
sfo01-m01-vds01	Replication	sfo01-m01-vds01-replication	vSphere Replication traffic vSphere Replication NFC traffic	9000
sfo01-m01-vds01	Management	sfo02-m01-vds01-management	Management traffic	1500 (default)
sfo01-m01-vds01	vMotion	sfo02-m01-vds01-vmotion	vMotion traffic	9000

vSphere Distributed Switch	Network Label	Connected Port Group	Enabled Services	MTU
sfo01-m01-vds01	iSCSI-A	sfo02-m01-vds01-iSCSI-A	-	9000
sfo01-m01-vds01	iSCSI-B	sfo02-m01-vds01-iSCSI-B	-	9000
sfo01-m01-vds01	NFS	sfo02-m01-vds01-nfs	-	9000
sfo01-m01-vds01	Replication	sfo02-m01-vds01-replication	vSphere Replication traffic vSphere Replication NFC traffic	9000
sfo01-m01-vds01	VTEP	Autogenerated (NSX VTEP)	-	9000

For more information on the physical network design specifications, see section 9.2, “Physical Networking Design.”

Shared Edge and Compute Cluster Distributed Switches

The shared edge and compute cluster uses a single vSphere Distributed Switch with the following configuration settings.

Table 44) Virtual switch for the shared edge and compute cluster.

vSphere Distributed Switch Name	Function	Network I/O Control	Number of Physical NIC Ports	MTU
sfo01-w01-vds01	• ESXi Management • Network IP Storage (NFS) • vSphere vMotion • VTEP • Uplinks (2) to enable ECMP • iSCSI	Enabled	2	9000

Table 45) vDS-Comp01 port group configuration settings.

Parameter	Setting
Failover detection	Link status only
Notify switches	Enabled
Failback	Yes
Failover order	Active uplinks: Uplink1, Uplink2

Network Switch Design for Shared Edge and Compute ESXi Hosts

This section expands on the logical network design by providing details on the physical NIC layout and physical network attributes.

Figure 39) Network switch design for shared edge and compute ESXi hosts.

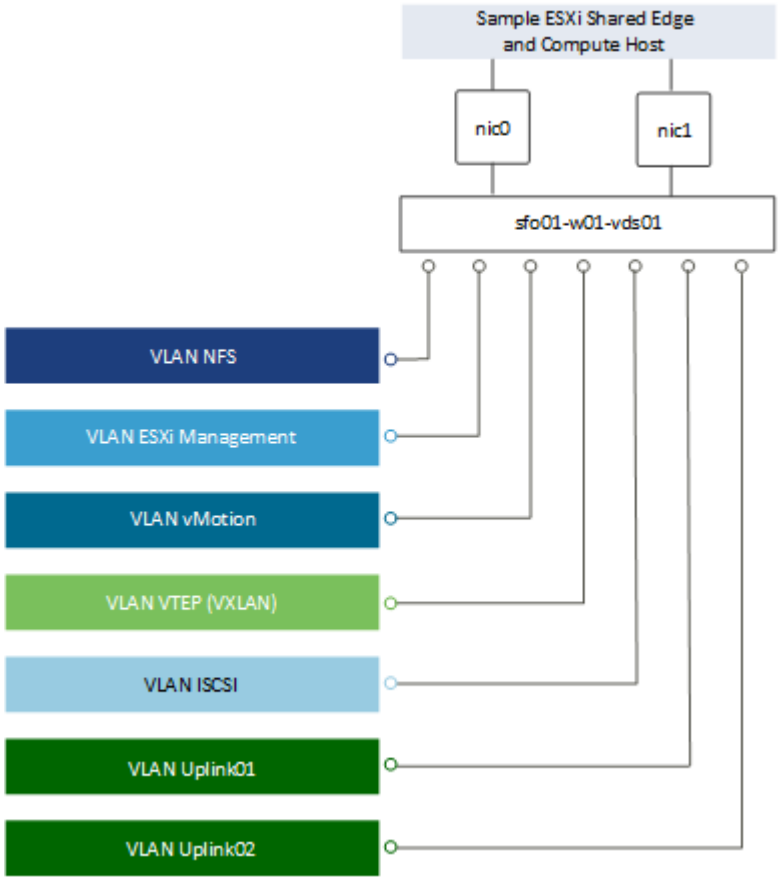


Table 46) Shared edge and compute cluster virtual switches by physical/virtual NIC.

vSphere Distributed Switch	vmnic	Function
sfo01-w01-vds01	0	Uplink
sfo01-w01-vds01	1	Uplink

Note: The following VLANs are meant as samples. Your actual implementation depends on your environment.

Table 47) Shared edge and compute cluster virtual switch port groups and VLANs.

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
sfo01-w01-vds01	sfo01-w01-vds01-management	Route based on physical NIC load	1, 2	1631
sfo01-w01-vds01	sfo01-w01-vds01-vmotion	Route based on physical NIC load	1, 2	1632

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
sfo01-w01-vds01	sfo01-w01-vds01-iSCSI-A	Use explicit failover order	1	1633
sfo01-w01-vds01	sfo01-w01-vds01-iSCSI-B	Use explicit failover order	2	1633
sfo01-w01-vds01	sfo01-w01-vds01-nfs	Route based on physical NIC load	1, 2	1615
sfo01-w01-vds01	sfo01-w01-vds01-uplink01	Route based on originating virtual port	1	1635
sfo01-w01-vds01	sfo01-w01-vds01-uplink02	Route based on originating virtual port	2	2713
sfo01-w01-vds01	sfo02-w01-vds01-management	Route based on physical NIC load	1, 2	1641
sfo01-w01-vds01	sfo02-w01-vds01-vmotion	Route based on physical NIC load	1, 2	1642
sfo01-w01-vds01	sfo02-w01-vds01-iSCSI-A	Use explicit failover order	1	1643
sfo01-w01-vds01	sfo02-w01-vds01-iSCSI-B	Use explicit failover order	2	1643
sfo01-w01-vds01	sfo02-w01-vds01-nfs	Route based on physical NIC load	1, 2	1625
sfo01-w01-vds01	sfo02-w01-vds01-uplink01	Route based on originating virtual port	1	1645
sfo01-w01-vds01	sfo02-w01-vds01-uplink02	Route based on originating virtual port	2	2723
sfo01-w01-vds01	Auto Generated (NSX VTEP)	Route based on SRC-ID	1, 2	1634

Table 48) Shared edge and compute cluster VMkernel adapter.

vSphere Distributed Switch	Network Label	Connected Port Group	Enabled Services	MTU
sfo01-w01-vds01	Management	sfo01-w01-vds01-management	Management traffic	1500 (default)
sfo01-w01-vds01	vMotion	sfo01-w01-vds01-vmotion	vMotion traffic	9000
sfo01-w01-vds01	iSCSI-A	sfo01-w01-vds01-iSCSI-A	-	9000

vSphere Distributed Switch	Network Label	Connected Port Group	Enabled Services	MTU
sfo01-w01-vds01	iSCSI-B	sfo01-w01-vds01-iSCSI-B	-	9000
sfo01-w01-vds01	NFS	sfo01-w01-vds01-nfs	-	9000
sfo01-w01-vds01	Management	sfo02-w01-vds01-management	Management traffic	1500 (default)
sfo01-w01-vds01	vMotion	sfo02-w01-vds01-vmotion	vMotion traffic	9000
sfo01-w01-vds01	iSCSI-A	sfo02-w01-vds01-iSCSI-A	-	9000
sfo01-w01-vds01	iSCSI-B	sfo02-w01-vds01-iSCSI-B	-	9000
sfo01-w01-vds01	NFS	sfo02-w01-vds01-nfs	-	9000
sfo01-w01-vds01	VTEP	Auto Generated (NSX VTEP)	-	9000

For more information on the physical network design, see section 9.2, “Physical Networking Design.”

Compute Cluster Distributed Switches

A compute cluster vSphere Distributed Switch uses the following configuration settings.

Table 49) Virtual switch for a dedicated compute cluster.

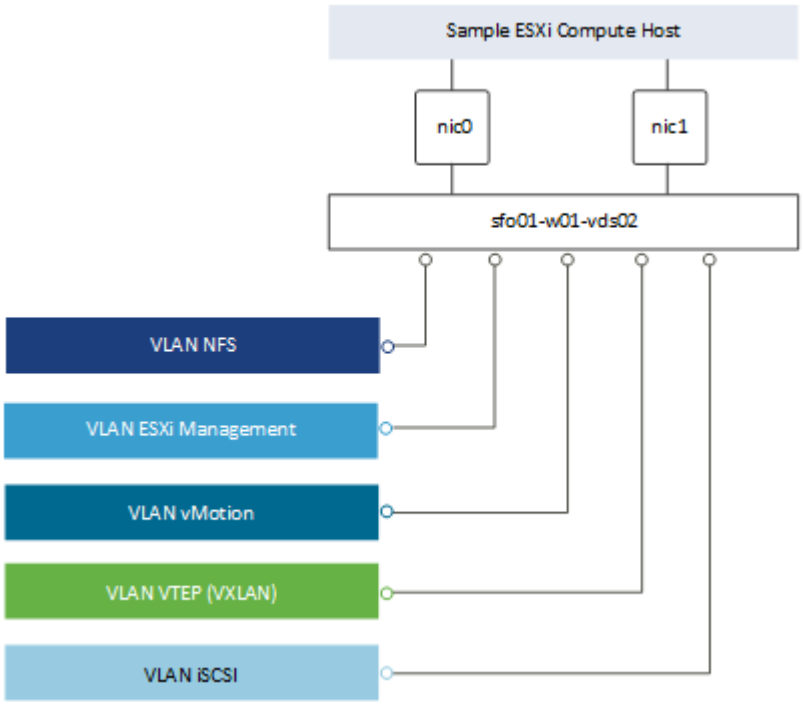
vSphere Distributed Switch Name	Function	Network I/O Control	Number of Physical NIC Ports	MTU
sfo01-w01-vds02	• ESXi management • Network IP storage (NFS) • vSphere vMotion • VTEP	Enabled	2	9000

Table 50) vDS-Comp02 port group configuration settings.

Parameter	Setting
Failover detection	Link status only
Notify switches	Enabled
Failback	Yes
Failover order	Active uplinks: Uplink1, Uplink2

Network Switch Design for Compute ESXi Hosts

Figure 40) Network switch design for compute ESXi hosts.



This section expands on the logical network design by providing details on the physical NIC layout and physical network attributes.

Table 51) Compute cluster virtual switches by physical/virtual NIC.

vSphere Distributed Switch	vmnic	Function
sfo01-w01-vds02	0	Uplink
sfo01-w01-vds02	1	Uplink

Note: The following VLANs are meant as samples. Your actual implementation depends on your environment.

Table 52) Compute cluster virtual switch port groups and VLANs.

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
sfo01-w01-vds02	sfo01-w01-vds02-management	Route based on physical NIC load	1, 2	1621
sfo01-w01-vds02	sfo01-w01-vds02-vmotion	Route based on physical NIC load	1, 2	1622
sfo01-w01-vds02	Autogenerated (NSX VTEP)	Route based on SRC-ID	1, 2	1624
sfo01-w01-vds02	sfo01-w01-vds02-nfs	Route based on physical NIC load	1, 2	1625

Table 53) Compute cluster VMkernel adapter.

vSphere Distributed Switch	Network Label	Connected Port Group	Enabled Services	MTU
sfo01-w01-vds02	Management	sfo01-w01-vds02-management	Management traffic	1500 (Default)
sfo01-w01-vds02	vMotion	sfo01-w01-vds02-vmotion	vMotion traffic	9000
sfo01-w01-vds02	NFS	sfo01-w01-vds02-nfs	-	9000
sfo01-w01-vds02	VTEP	Auto Generated (NSX VTEP)	-	9000

For more information on the physical network design specifications, see section 9.2, “Physical Networking Design.”

NIC Teaming

You can use NIC teaming to increase the network bandwidth available in a network path and to provide the redundancy that supports higher availability.

Benefits and Overview

NIC teaming helps avoid a single point of failure and offers options for load balancing of traffic. To further reduce the risk of a single point of failure, build NIC teams by using ports from multiple NIC and motherboard interfaces.

Create a single virtual switch with teamed NICs across separate physical switches. This VVD uses an active-active configuration that uses the route based on a physical NIC load algorithm for teaming. In this configuration, idle network cards do not wait for a failure to occur, and they aggregate bandwidth.

NIC Teaming Design Background

For a predictable level of performance, use multiple network adapters in one of the following configurations:

- An active-passive configuration that uses explicit failover when connected to two separate switches.
- An active-active configuration in which two or more physical NICs in the server are assigned the active role.

This validated design uses an active-active configuration.

Table 54) NIC teaming and policy.

Design Quality	Active-Active	Active-Passive	Comments
Availability	↑	↑	Regardless of the option chosen, use teaming to increase the availability of the environment.
Manageability	o	o	Neither design option impacts manageability.

Design Quality	Active-Active	Active-Passive	Comments
Performance	↑	o	An active-active configuration can send traffic across either NIC, thereby increasing the available bandwidth. This configuration provides a benefit if the NICs are being shared among traffic types and Network I/O Control is used.
Recoverability	o	o	Neither design option affects recoverability.
Security	o	o	Neither design option affects security.

Legend: ↑ = positive impact on quality; ↓ = negative impact on quality; o = no impact on quality.

Table 55) NIC teaming design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-006	Use the route based on a physical NIC load teaming algorithm for all port groups except for ones that carry VXLAN traffic and ones that carry iSCSI traffic. VTEP kernel ports and VXLAN traffic use routes based on SRC-ID. iSCSI ports use explicit failover with only one port active.	Reduce the complexity of the network design and increase resiliency and performance.	Because neither NSX nor iSCSI supports route based on physical NIC load, two different algorithms are necessary.

Network I/O Control

When Network I/O Control is enabled, the distributed switch allocates bandwidth for the following system traffic types:

- Fault tolerance traffic
- iSCSI traffic
- vSphere vMotion traffic
- Management traffic
- VMware vSphere Replication traffic
- NFS traffic
- Backup traffic
- Virtual machine traffic

How Network I/O Control Works

Network I/O Control enforces the share value specified for the different traffic types only when there is network contention. When contention occurs, Network I/O Control applies the share values set to each

traffic type. As a result, less important traffic, as defined by the share percentage, is throttled, allowing more important traffic types to gain access to more network resources.

Network I/O Control allows the reservation of bandwidth for system traffic based on the capacity of the physical adapters on an ESXi host. It also enables fine-grained resource control at the virtual machine network adapter level. Resource control is similar to the model for vCenter CPU and memory reservations.

Network I/O Control Heuristics

The following heuristics can help with design decisions:

- **Using shares or limits.** When you use bandwidth allocation, consider using shares instead of limits. Limits impose hard limits on the amount of bandwidth used by a traffic flow even when network bandwidth is available.
- **Limits on certain resource pools.** Consider imposing limits on a given resource pool. For example, if you put a limit on vSphere vMotion traffic, you can benefit in situations where multiple vSphere vMotion data transfers, initiated on different ESXi hosts at the same time, result in oversubscription at the physical network level. By limiting the available bandwidth for vSphere vMotion at the ESXi host level, you can prevent performance degradation for other traffic.
- **Teaming policy.** When you use Network I/O Control, use route based on physical NIC load teaming as a distributed switch teaming policy to maximize the networking capacity utilization. With load-based teaming, traffic might move among uplinks, and reordering of packets at the receiver can occasionally result.
- **Traffic shaping.** Use distributed port groups to apply configuration policies to different traffic types. Traffic shaping can help in situations where multiple vSphere vMotion migrations initiated on different ESXi hosts converge on the same destination ESXi host. The actual limit and reservation also depend on the traffic shaping policy for the distributed port group that the adapter is connected to.

Network I/O Control Design Decisions

Based on the heuristics, this design has the following decisions.

Table 56) Network I/O Control design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-007	Enable Network I/O Control on all distributed switches.	Increase resiliency and performance of the network.	If configured incorrectly, Network I/O Control might affect network performance for critical traffic types.
SDDC-VI-NET-008	Set the share value for vSphere vMotion traffic to Low.	During times of network contention, vMotion traffic is not as important as virtual machine or storage traffic.	During times of network contention, vMotion takes longer than usual to complete.
SDDC-VI-NET-009	Set the share value for vSphere Replication traffic to Low.	During times of network contention, vSphere Replication traffic is not as important as virtual machine or storage traffic.	During times of network contention, vSphere Replication takes longer and might violate the defined SLA.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-010	Set the share value for iSCSI traffic to High.	During times of network contention, iSCSI traffic requires guaranteed bandwidth to support virtual machine performance.	None
SDDC-VI-NET-011	Set the share value for management traffic to Normal.	By keeping the default setting of Normal, management traffic is prioritized higher than vSphere vMotion and vSphere Replication, but lower than iSCSI traffic. Management traffic is important because it makes sure that the hosts can still be managed during times of network contention.	None
SDDC-VI-NET-012	Set the share value for NFS traffic to Low.	Because NFS is used for secondary storage, such as backups and vRealize Log Insight archives, it is not as important as iSCSI traffic. By prioritizing it lower, iSCSI is not affected.	During times of network contention, backups are slower than usual.
SDDC-VI-NET-013	Set the share value for backup traffic to Low.	During times of network contention, it is more important for primary functions of the SDDC to continue to have access to network resources over backup traffic.	During times of network contention, backups are slower than usual.
SDDC-VI-NET-014	Set the share value for virtual machines to High.	Virtual machines are the most important asset in the SDDC. Leaving the default setting of High ensures that they always have access to the network resources they need.	None
SDDC-VI-NET-015	Set the share value for vSphere fault tolerance to Low.	This design does not use vSphere fault tolerance. Fault tolerance traffic can be set to the lowest priority.	None
SDDC-VI-NET-016	Set the share value for vSAN traffic to Low.	This design does not use vSAN. vSAN traffic can be set to the lowest priority.	None

VXLAN

VXLAN allows you to create isolated, multitenant broadcast domains across data center fabrics, and it also enables you to create elastic, logical networks that span physical network boundaries.

The first step in creating these logical networks is to abstract and pool the networking resources. Just as vSphere abstracts compute capacity from the server hardware to create virtual pools of resources that can be consumed as a service, vSphere Distributed Switches and VXLAN abstract the network into a generalized pool of network capacity and separate the consumption of these services from the underlying physical infrastructure. A network capacity pool can span physical boundaries, optimizing compute resource utilization across clusters, pods, and geographically separated data centers. The unified pool of network capacity can then be optimally segmented into logical networks that are directly attached to specific applications.

VXLAN works by creating layer 2 logical networks that are encapsulated in standard layer 3 IP packets. A segment ID in every frame differentiates the VXLAN logical networks from each other without any need for VLAN tags. As a result, large numbers of isolated layer 2 VXLAN networks can coexist on a common layer 3 infrastructure.

In the vSphere architecture, the encapsulation is performed between the virtual NIC of the guest virtual machine and the logical port on the virtual switch, making VXLAN transparent to both the guest virtual machines and the underlying layer 3 network. Gateway services between VXLAN and non-VXLAN hosts (for example, a physical server or the internet router) are performed by the NSX ESG appliance. The Edge gateway translates VXLAN segment IDs to VLAN IDs, so that non-VXLAN hosts can communicate with virtual machines on a VXLAN network.

The shared edge and compute cluster hosts all NSX Edge instances that connect to the internet or to corporate VLANs, so that the network administrator can manage the environment in a more secure and centralized way.

Table 57) VXLAN design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-017	Use NSX for vSphere to introduce VXLANs for the use of virtual application networks and tenant networks.	Simplify the network configuration for each tenant by using centralized virtual network management.	Requires additional compute and storage resources to deploy NSX components. Additional training on NSX for vSphere might be needed.
SDDC-VI-NET-018	Use VXLAN with NSX Edge gateways, the UDLR, and DLR to provide customer and tenant network capabilities.	Create isolated, multitenant broadcast domains across data center fabrics to create elastic, logical networks that span physical network boundaries.	Transport networks and MTU greater than 1600 bytes have to be configured in the reachability radius.
SDDC-VI-NET-019	Use VXLAN with NSX Edge gateways and the UDLR to provide management application network capabilities.	Leverage benefits of the network virtualization in the management cluster.	Requires installation and configuration of an NSX for vSphere instance in the management cluster.

vMotion TCP/IP Stack

Use the vMotion TCP/IP stack to isolate traffic for vMotion and to assign a dedicated default gateway for vMotion traffic.

By using a separate TCP/IP stack, you can manage vMotion and cold migration traffic according to the topology of the network, and as required for your organization.

- Route the traffic for the migration of virtual machines that are powered on or powered off by using a default gateway that is different from the gateway assigned to the default stack on the ESXi host.
- Assign a separate set of buffers and sockets.
- Avoid routing table conflicts that might otherwise appear when many features are using a common TCP/IP stack.
- Isolate traffic to improve security.

Table 58) vMotion TCP/IP stack design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-020	Use the vMotion TCP/IP stack for vSphere vMotion traffic.	By using the vMotion TCP/IP stack, vSphere vMotion traffic can be assigned a default gateway on its own subnet and can go over layer 3 networks.	The vMotion TCP/IP stack is not available in the vDS VMkernel creation wizard, and therefore the VMkernal adapter must be created directly on the ESXi host.

10.7 NSX Design

This design implements software-defined networking by using VMware NSX for vSphere. By using NSX for vSphere, virtualization delivers for networking what it has already delivered for compute and storage.

In much the same way that server virtualization programmatically creates, makes snapshots, deletes, and restores software-based virtual machines (VMs), NSX network virtualization programmatically creates, makes snapshots, deletes, and restores software-based virtual networks. The result is a transformative approach to networking that not only enables data center managers to achieve orders of magnitude better agility and economics, but also supports a vastly simplified operational model for the underlying physical network. NSX for vSphere is a nondisruptive solution because it can be deployed on any IP network from any vendor, including existing traditional networking models and next-generation fabric architectures.

When administrators provision workloads, network management is one of their most time-consuming tasks. Most of the time spent provisioning networks is consumed in configuring individual components in the physical infrastructure and verifying that network changes do not affect other devices that are using the same networking infrastructure.

The need to preprovision and configure networks is a major constraint to cloud deployments where speed, agility, and flexibility are critical requirements. Preprovisioned physical networks can allow the rapid creation of virtual networks and faster deployment times of workloads using the virtual network. As long as the physical network that you need is already available on the ESXi host where the workload is to be deployed, this works well. However, if the network is not available on a given ESXi host, you must find an ESXi host with the available network and spare capacity to run your workload in your environment.

To get around this bottleneck, you must decouple virtual networks from their physical counterparts. This in turn requires that you must programmatically recreate all physical networking attributes that are required by workloads in the virtualized environment. Because network virtualization supports the creation of virtual networks without modification of the physical network infrastructure, it allows more rapid network provisioning.

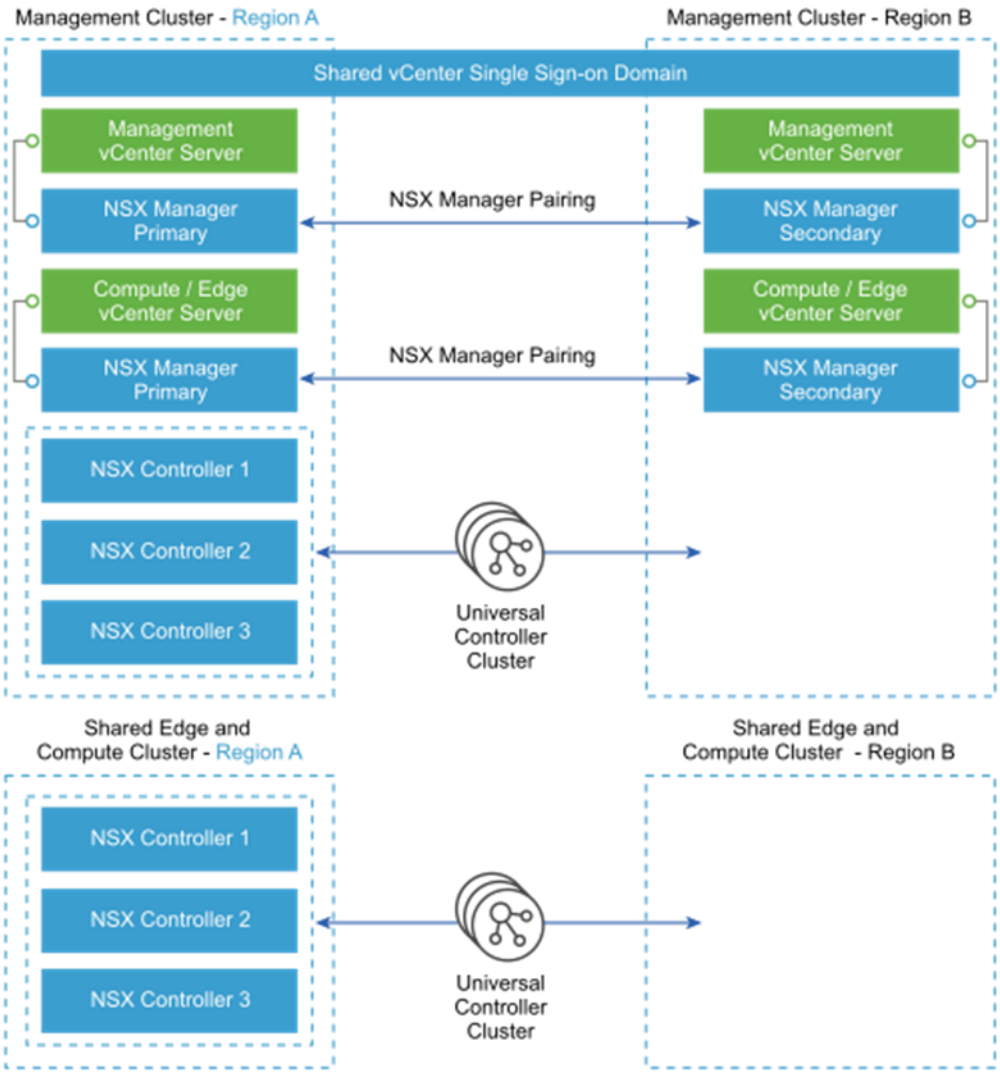
NSX for vSphere Design

Each NSX instance is tied to a vCenter Server instance. The design decision to deploy two vCenter Server instances per region (SDDC-VI-VC-001) requires deployment of two separate NSX instances per region.

Table 59) NSX for vSphere design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-001	Use two separate NSX instances per region. One instance is tied to the management vCenter Server, and the other instance is tied to the compute vCenter Server.	Software-defined networking (SDN) capabilities offered by NSX, such as load balancing and firewalls, are crucial for the compute/edge layer to support CMP operations. They are also crucial for the management applications in the management stack that need these capabilities.	You must install and perform initial configuration of multiple NSX instances separately.
SDDC-VI-SDN-002	Pair NSX Manager instances in a primary-secondary relationship across regions for both management and compute workloads.	NSX can extend the logical boundaries of the networking and security services across regions. As a result, workloads can be live-migrated and failed over between regions without reconfiguring the network and security constructs.	You must consider that you can pair up to eight NSX Manager instances.

Figure 41) Architecture of NSX for vSphere.



NSX Components

The following subsections describe the components of the solution and how they are relevant to the network virtualization design.

Consumption Layer

The CMP can consume NSX for vSphere, represented by vRealize Automation, by using the NSX REST API and the vSphere Web Client.

Cloud Management Platform

vRealize Automation consumes NSX for vSphere on behalf of the CMP. NSX offers self-service provisioning of virtual networks and related features from a service portal. Details of the service requests and their orchestration are outside the scope of this document; see the Cloud Management Platform Design document.

API

NSX for vSphere offers a powerful management interface through its REST API.

- A client can read an object by making an HTTP GET request to the object's resource URL.
- A client can write (create or modify) an object with an HTTP PUT or POST request that includes a new or changed XML document for the object.
- A client can delete an object with an HTTP DELETE request.

vSphere Web Client

NSX Manager provides a networking and security plug-in in vSphere Web Client. This plug-in is an interface for consuming virtualized networking from NSX Manager for users with sufficient privileges.

Table 60) Consumption method design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-003	For the shared edge and compute cluster NSX instance, end-user access is accomplished by using vRealize Automation services. Administrators use both vSphere Web Client and the NSX REST API.	vRealize Automation services are used for the customer-facing portal. vSphere Web Client consumes NSX for vSphere resources through the Network and Security plug-in. The NSX REST API offers the potential of scripting repeating actions and operations.	Customers typically interact only indirectly with NSX from the vRealize Automation portal. Administrators interact with NSX from vSphere Web Client and API.
SDDC-VI-SDN-004	For the management cluster NSX instance, consumption is only by provider staff using the vSphere Web Client and the API.	Makes sure that infrastructure components are not modified by tenants or nonprovider staff.	Tenants do not have access to the management stack workloads.

NSX Manager

NSX Manager provides the centralized management plane for NSX for vSphere and has a one-to-one mapping to vCenter Server workloads.

NSX Manager performs the following functions:

- Provides the single point of configuration and the REST API entry points for NSX in a vSphere environment.
- Deploys NSX Controller clusters, NSX Edge distributed routers, and NSX Edge service gateways in the form of OVF appliances, guest introspection services, and so on.
- Prepares ESXi hosts for NSX by installing VXLAN, distributed routing and firewall kernel modules, and the UWA.
- Communicates with NSX Controller clusters over REST and with ESXi hosts over the RabbitMQ message bus. This internal message bus is specific to NSX for vSphere and does not require setup of additional services.
- Generates certificates for the NSX Controller instances and ESXi hosts to secure control plane communications with mutual authentication.

NSX Controller

An NSX Controller instance performs the following functions:

- Provides the control plane to distribute VXLAN and logical routing information to ESXi hosts.

- Includes nodes that are clustered for scale out and high availability.
- Slices network information across cluster nodes for redundancy.
- Removes requirement of VXLAN layer 3 multicast in the physical network.
- Provides ARP suppression of broadcast traffic in VXLAN networks.

NSX control plane communication occurs over the management network.

Table 61) NSX Controller design decision.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-005	Deploy NSX Controller instances in Universal Cluster mode with three members to provide high availability and scale. Provision these three nodes through the primary NSX Manager instance.	The high availability of NSX Controller reduces the downtime period in case of failure of one physical ESXi host.	The secondary NSX Manager instance does not deploy controllers. The controllers from the primary NSX Manager instance manage all secondary resources.

NSX Virtual Switch

The NSX data plane consists of the NSX virtual switch, which is based on the vSphere Distributed Switch, with additional components to enable rich services. The add-on NSX components include kernel modules (VIBs), which run in the hypervisor kernel and provide services such as DLR, distributed firewall, and VXLAN capabilities.

The NSX virtual switch abstracts the physical network and provides access-level switching in the hypervisor. It is central to network virtualization because it enables logical networks that are independent of physical constructs such as VLAN. Using an NSX virtual switch offers the following benefits:

- Supports overlay networking and centralized network configuration. Overlay networking facilitates massive scale of hypervisors.
- Because the NSX virtual switch is based on VDS, it provides a comprehensive toolkit for traffic management, monitoring, and troubleshooting in a virtual network. Features include port mirroring, NetFlow/IPFIX, configuration backup and restore, network health check, QoS, and more.

Logical Switching

NSX logical switches create logically abstracted segments to which tenant virtual machines can be connected. A single logical switch is mapped to a unique VXLAN segment and is distributed across the ESXi hypervisors in a transport zone. The logical switch allows line-rate switching in the hypervisor without the constraints of VLAN sprawl or spanning tree issues.

Distributed Logical Router

NSX DLR is optimized for forwarding in the virtualized space; that is, forwarding between VMs on VXLAN-backed or VLAN-backed port groups. DLR has the following characteristics:

- High-performance, low-overhead first-hop routing
- Scales with the number of ESXi hosts
- Up to 1,000 LIFs on each DLR

Distributed Logical Router Control Virtual Machine

The DLR control virtual machine is the control plane component of the routing process, providing communication between NSX Manager and the NSX Controller cluster through the UWA. NSX Manager

sends LIF information to the control virtual machine and the NSX Controller cluster, and the control virtual machine sends routing updates to the NSX Controller cluster.

User World Agent

The UWA is a TCP (SSL) client that facilitates communication between the ESXi hosts and SX Controller instances, as well as the retrieval of information from NSX Manager through interaction with the message bus agent.

VXLAN Tunnel Endpoint

VTEPs are instantiated in the vSphere Distributed Switch to which the ESXi hosts that are prepared for NSX for vSphere are connected. VTEPs are responsible for encapsulating VXLAN traffic as frames in UDP packets and for the corresponding decapsulation. VTEPs take the form of one or more VMkernel ports with IP addresses and are used both to exchange packets with other VTEPs and to join IP multicast groups through the IGMP. If you use multiple VTEPs, then you must select a teaming method.

Edge Services Gateway

The primary function of the NSX ESG is north-south communication, but it also offers support for layer 2, layer 3, perimeter firewall, load balancing, and other services such as SSL-VPN and DHCP-relay.

Distributed Firewall

NSX includes a distributed kernel-level firewall (the distributed firewall). Security enforcement is done at the kernel and virtual machine network adapter level. The security enforcement implementation enables firewall rule enforcement in a highly scalable manner without creating bottlenecks on physical appliances. The distributed firewall has minimal CPU overhead and can perform at line rate.

The flow monitoring feature of the distributed firewall displays network activity between virtual machines at the application protocol level. This information can be used to audit network traffic, define and refine firewall policies, and identify botnets.

Logical Load Balancer

The NSX logical load balancer provides load balancing services up to layer 7, allowing distribution of traffic across multiple servers to achieve optimal resource utilization and availability. The logical load balancer is a service provided by the NSX ESG.

NSX for vSphere Requirements

NSX for vSphere requirements affect both physical and virtual networks.

Physical Network Requirements

Physical requirements determine the MTU size for networks that carry VLAN traffic, dynamic routing support, type synchronization through an NTP server, and forward and reverse DNS resolution.

Table 62) NSX for vSphere physical network requirements.

Requirement	Comments
Any network that carries VXLAN traffic must have an MTU size of 1600 or greater.	VXLAN packets cannot be fragmented. The MTU size must be large enough to support extra encapsulation overhead. This design uses jumbo frames and an MTU size of 9000 for VXLAN traffic.

Requirement	Comments
For the hybrid replication mode, IGMP snooping must be enabled on the layer 2 switches to which ESXi hosts that participate in VXLAN are attached. The IGMP querier must be enabled on the connected router or layer 3 switch.	IGMP snooping on layer 2 switches is a requirement of the hybrid replication mode, the recommended replication mode for broadcast, unknown unicast, and multicast traffic when deploying into an environment with large scale-out potential. The traditional requirement for Protocol Independent Multicast (PIM) is removed.
Dynamic routing support on the upstream layer 3 data center switches must be enabled.	Enable a dynamic routing protocol supported by NSX on the upstream data center switches to establish dynamic routing adjacency with the ESGs.
NTP server must be available.	NSX Manager requires NTP settings that synchronize it with the rest of the vSphere environment. Drift can cause problems with authentication. NSX Manager must be in sync with the vCenter Single Sign-On service on the Platform Services Controller.
Forward and reverse DNS resolution for all management VMs must be established.	The NSX Controller nodes do not require DNS entries.

NSX Component Specifications

Table 63 lists the components of the NSX for vSphere solution and the requirements for installing and running them. The compute and storage requirements have been taken into account when sizing resources to support the NSX for vSphere solution.

Note: NSX ESG sizing can vary with tenant requirements, so all options are listed.

Table 63) NSX component resource requirements.

VM	vCPU	Memory	Storage	Quantity per Stack Instance
NSX Manager	4	16GB	60GB	1
NSX Controller	4	4GB	20GB	3
NSX ESG	1 (Compact) 2 (Large) 4 (Quad Large) 6 (X-Large)	512MB (Compact) 1GB (Large) 1GB (Quad Large) 8GB (X-Large)	512MB 512MB 512MB 4.5GB (X-Large) (+4GB with swap)	Optional component. Deployment of the NSX ESG varies per use case.
DLR control VM	1	512MB	512MB	Optional component. Varies with use case. Typically, two per HA pair.
Guest introspection	2	1GB	4GB	Optional component. One per ESXi host.

VM	vCPU	Memory	Storage	Quantity per Stack Instance
NSX data security	1	512MB	6GB	Optional component. One per ESXi host.

NSX Edge Service Gateway Sizing

The Quad Large model is suitable for high-performance firewall, and the extra-large model is suitable for both high-performance load balancing and routing.

You can convert between NSX Edge service gateway sizes on demand by using a nondisruptive upgrade process, so NetApp recommends beginning with the large model and scaling up if necessary. A large NSX Edge service gateway is suitable for medium firewall performance but, as detailed later, the NSX Edge service gateway does not perform the majority of firewall functions.

Note: Edge service gateway throughput is influenced by the WAN circuit. NetApp recommends an adaptable approach, in which you convert as necessary.

Table 64) NSX Edge service gateway sizing design decision.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-006	Use large-size NSX Edge service gateways.	The large size provides all the performance characteristics needed even in the event of a failure. A larger size might also provide the required performance but at the expense of extra resources that cannot be used.	None

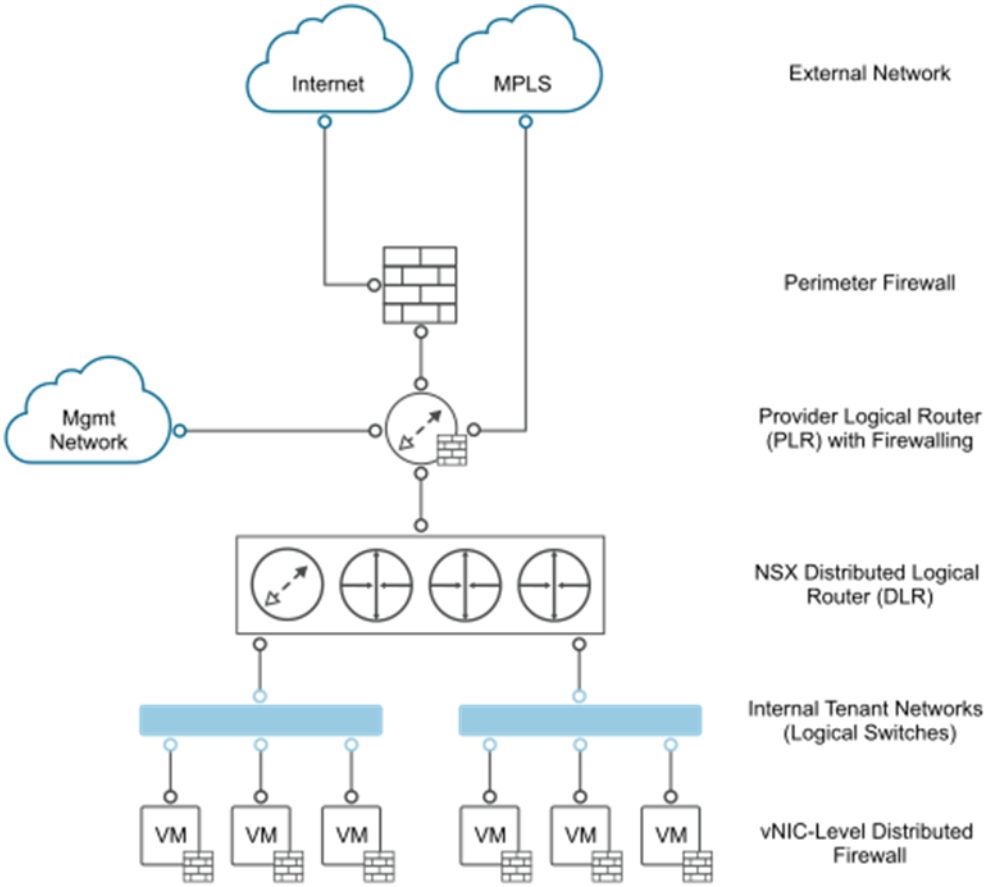
Network Virtualization Conceptual Design

This conceptual design description helps you to understand the network virtualization design.

The network virtualization conceptual design includes a perimeter firewall, a provider logical router, and the NSX for a vSphere logical router. It also includes the external network, internal tenant network, and internal non-tenant network.

Note: In this document, tenant refers to a tenant of the CMP in the compute/edge stack, or to a management application in the management stack.

Figure 42) Conceptual tenant overview.



The conceptual design has the following key components:

- **External networks.** Connectivity to and from external networks is through the perimeter firewall. The main external network is the internet.
- **Perimeter firewall.** The physical firewall exists at the perimeter of the data center. Each tenant receives either a full instance or a partition of an instance to filter external traffic.
- **Provider logical router (PLR).** The PLR exists behind the perimeter firewall and handles north-south traffic that is entering and leaving tenant workloads.
- **NSX for vSphere DLR.** This logical router is optimized for forwarding in the virtualized space—that is, between VMs, on VXLAN port groups, or on VLAN-backed port groups.
- **Internal non-tenant network.** A single management network, which sits behind the perimeter firewall but not behind the PLR. Enables customers to manage the tenant environments.
- **Internal tenant networks.** Connectivity for the main tenant workload. These networks are connected to a DLR, which sits behind the PLR. These networks take the form of VXLAN-based NSX for vSphere logical switches. Tenant virtual machine workloads are directly attached to these networks.

Cluster Design for NSX for vSphere

Following the vSphere design, the NSX for vSphere design consists of a management stack and a compute and edge stack in each region.

Management Stack

In the management stack, the underlying ESXi hosts are prepared for NSX for vSphere. The management stack has these components:

- NSX Manager instances for both stacks (management stack and compute and edge stack)
- NSX Controller cluster for the management stack
- NSX ESG and DLR control VMs for the management stack

Compute and Edge Stack

In the compute and edge stack, the underlying ESXi hosts are prepared for NSX for vSphere. The compute and edge stack has these components:

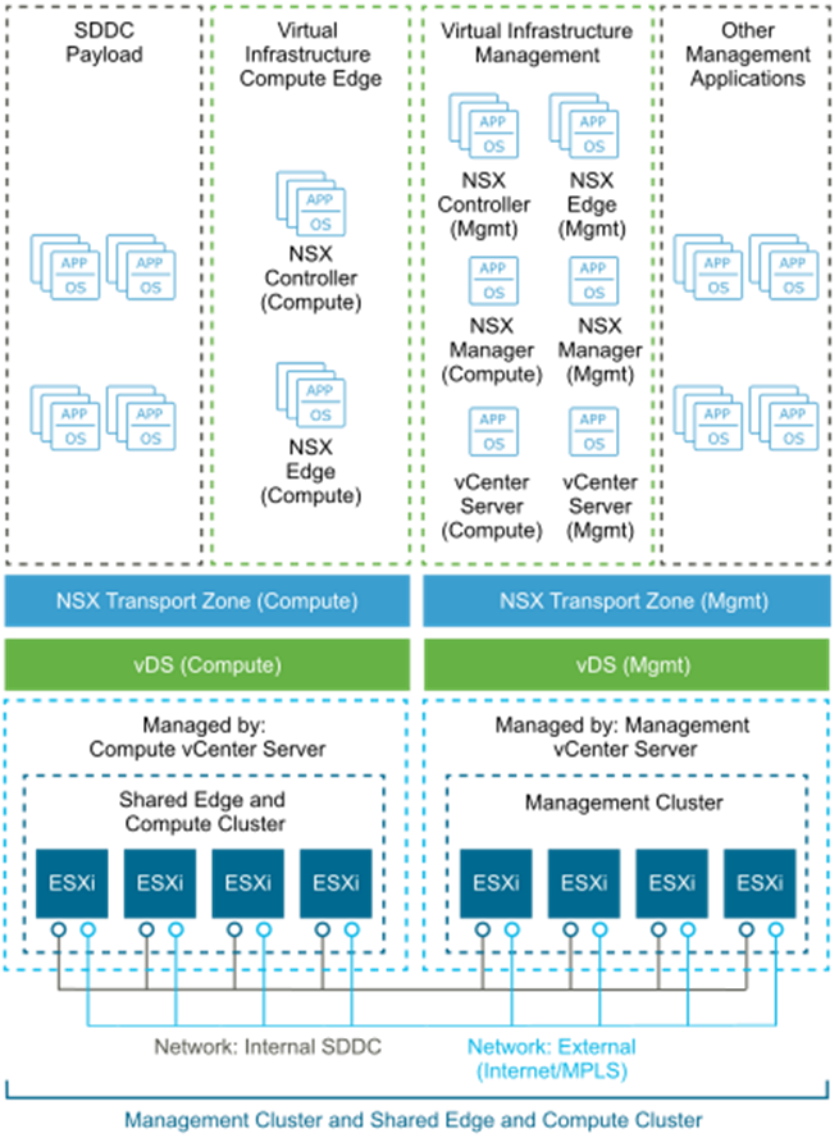
- NSX Controller cluster for the compute stack.
- All NSX Edge service gateways and DLR control VMs of the compute stack that are dedicated to handling the north-south traffic in the data center. A shared edge and compute stack helps prevent VLAN sprawl because any external VLANs only need be trunked to the ESXi hosts in this cluster.

Table 65) vSphere cluster design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-007	For the compute stack, do not use a dedicated edge cluster.	Simplifies configuration and minimizes the number of ESXi hosts required for initial deployment.	The NSX Controller instances, NSX ESGs, and DLR control VMs of the compute stack are deployed in the shared edge and compute cluster. Because of the shared nature of the cluster, you must scale out the cluster as compute workloads are added to avoid an effect on network performance.
SDDC-VI-SDN-008	For the management stack, do not use a dedicated edge cluster.	The number of supported management applications does not justify the cost of a dedicated edge cluster in the management stack.	The NSX Controller instances, NSX Edge service gateways, and DLR control VMs of the management stack are deployed in the management cluster.
SDDC-VI-SDN-009	Apply vSphere DRS anti-affinity rules to the NSX components in both stacks.	Using DRS prevents controllers from running on the same ESXi host and thereby risking their high-availability capability.	Additional configuration is required to set up anti-affinity rules.

The logical design of NSX considers the vCenter Server clusters and defines the place where each NSX component runs.

Figure 43) Cluster design for NSX for vSphere.



High Availability of NSX for vSphere Components

The NSX Manager instances from both stacks run on the management cluster. vSphere HA protects the NSX Manager instances by making sure that the NSX Manager virtual machine is restarted on a different ESXi host in the event of primary ESXi host failure. The data plane remains active during outages in the management and control planes, although the provisioning and modification of virtual networks is impaired until those planes become available again.

The NSX Edge service gateways and DLR control VMs of the compute stack are deployed on the shared edge and compute cluster. The NSX Edge service gateways and DLR control VMs of the management stack run on the management cluster.

NSX Edge components that are deployed for north-south traffic are configured in ECMP mode that supports route failover in seconds. NSX Edge components deployed for load balancing use NSX HA, which provides faster recovery than vSphere HA alone because NSX HA uses an active-passive pair of NSX Edge devices. By default, the passive Edge device becomes active within 15 seconds. All NSX Edge devices are also protected by vSphere HA.

Scalability of NSX Components

There is a one-to-one mapping between NSX Manager instances and vCenter Server instances. If the inventory of either the management stack or the compute stack exceeds the limits supported by a single vCenter Server, you can deploy a new vCenter Server instance, and you must also deploy a new NSX Manager instance. You can extend transport zones by adding more shared edge-compute and compute clusters until you reach the vCenter Server limits. Consider the limit of 100 DLRs per ESXi host, although the environment usually exceeds other vCenter Server limits before the DLR limit.

vSphere Distributed Switch Uplink Configuration

Each ESXi host uses two physical 10Gb Ethernet adapters associated with the uplinks on the vSphere Distributed Switches to which it is connected. Each uplink is connected to a different top-of-rack switch to mitigate the impact of a single top-of-rack switch failure and to provide two paths in and out of the SDDC.

Table 66) VTEP teaming and failover configuration design decision.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-010	Set up VTEPs to use a route based on SRC-ID for teaming and failover configuration.	Allows the use of the two uplinks of the distributed switch, resulting in better bandwidth utilization and faster recovery from network path failures.	None

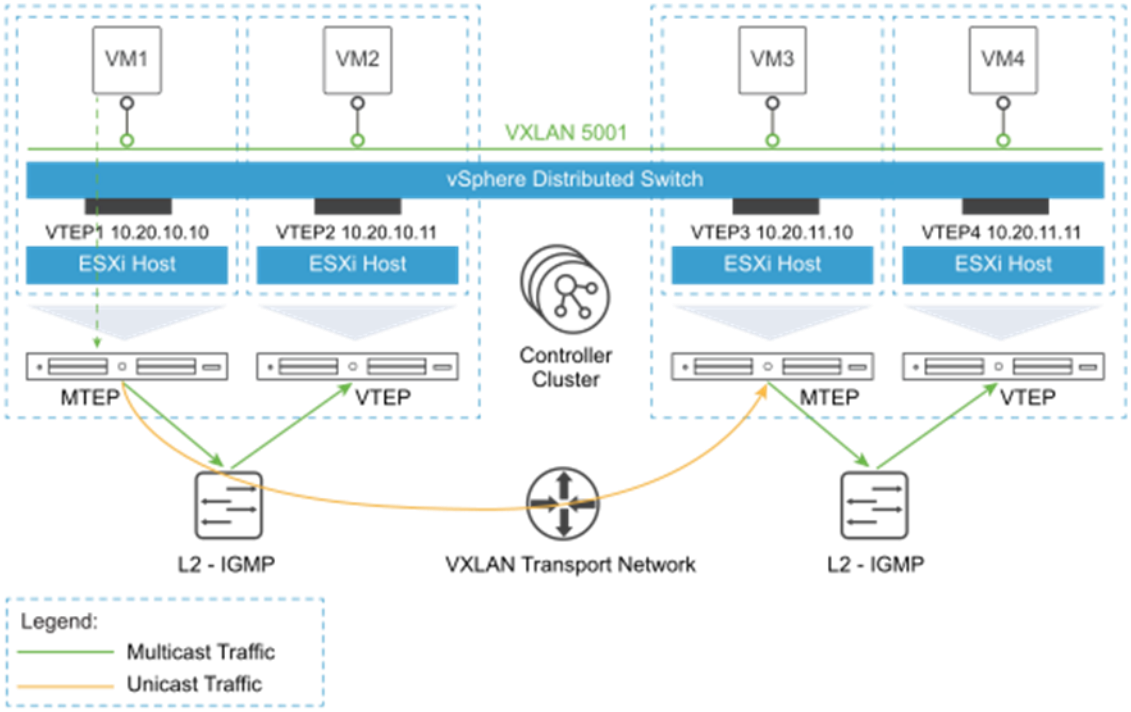
Logical Switch Control Plane Mode Design

The control plane decouples NSX for vSphere from the physical network and handles the broadcast, unknown unicast, and multicast traffic in the logical switches. The control plane is on top of the transport zone and is inherited by all logical switches that are created in it. It is possible to override aspects of the control plane.

The following options are available:

- **Multicast mode.** The control plane uses multicast IP addresses on the physical network. Use multicast mode only when upgrading from existing VXLAN deployments. In this mode, you must configure PIM/IGMP on the physical network.
- **Unicast mode.** The control plane is handled by the NSX Controller instances and all replication occurs locally on the ESXi host. This mode does not require multicast IP addresses or physical network configuration.
- **Hybrid mode.** This mode is an optimized version of the unicast mode where local traffic replication for the subnet is offloaded to the physical network. Hybrid mode requires IGMP snooping on the first-hop switch and access to an IGMP querier in each VTEP subnet. Hybrid mode does not require PIM.

Figure 44) Logical switch control plane in hybrid mode.



This design uses hybrid mode for control plane replication.

Table 67) Logical switch control plane mode design decision.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-011	Use hybrid mode for control plane replication.	Offloading multicast processing to the physical network reduces pressure on VTEPs as the environment scales out. For large environments, hybrid mode is preferable to unicast mode. Multicast mode is used only when migrating from existing VXLAN solutions.	IGMP snooping must be enabled on the ToR physical switch, and an IGMP querier must be available.

Transport Zone Design

A transport zone is used to define the scope of a VXLAN overlay network and can span one or more clusters within one vCenter Server domain. One or more transport zones can be configured in an NSX for vSphere solution. A transport zone is not meant to delineate a security boundary.

Table 68) Transport zone design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-012	For the compute stack, use a universal transport zone that encompasses all shared edge and compute clusters, and compute clusters from all regions for	A universal transport zone supports extending networks and security policies across regions. This allows seamless migration of applications across regions either by cross-vCenter	vRealize Automation is not able to deploy on-demand network objects against a secondary NSX Manager instance. You can pair up to eight NSX Manager instances. If the solution expands past eight NSX Manager

Decision ID	Design Decision	Design Justification	Design Implications
	workloads that require mobility between regions.	vMotion or by failover recovery with Site Recovery Manager.	instances, you must deploy a new primary manager and a new transport zone.
SDDC-VI-SDN-013	For the compute stack, use a global transport zone in each region that encompasses all shared edge and compute clusters and compute clusters for use with vRealize Automation on-demand network provisioning.	NSX manager instances with a secondary role cannot deploy universal objects. To allow all regions to deploy on-demand network objects, a global transport zone is required.	Shared edge and compute clusters and compute clusters have two transport zones.
SDDC-VI-SDN-014	For the management stack, use a single universal transport zone that encompasses all management clusters.	A single universal transport zone supports extending networks and security policies across regions. This allows seamless migration of the management applications across regions, either by cross-vCenter vMotion or by failover recovery with Site Recovery Manager.	You can pair up to eight NSX Manager instances. If the solution expands past eight NSX Manager instances, you must deploy a new primary manager and a new transport zone.
SDDC-VI-SDN-015	Enable Controller Disconnected Operation (CDO) mode in the management stack.	During times when the NSX Controller instances are unable to communicate with ESXi hosts, data plane updates, such as VNIs becoming active on an ESXi host, still occur.	Enabling CDO mode adds some overhead to the hypervisors when the control cluster is down.
SDDC-VI-SDN-016	Enable CDO mode on the shared edge and compute stack.	During times when the NSX Controller instances are unable to communicate with ESXi hosts, data plane updates, such as VNIs becoming active on an ESXi host, still occur.	Enabling CDO mode adds some overhead to the hypervisors when the control cluster is down.

Routing Design

The routing design considers different levels of routing in the environment to define a set of principles for a scalable routing solution.

- **North-south.** The PLR handles the north-south traffic to and from a tenant and management applications inside of application virtual networks.
- **East-west.** Internal east-west routing at the layer beneath the PLR deals with the application workloads.

Table 69) Routing model design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-017	Deploy a minimum of two NSX ESGs in an ECMP configuration for north-south routing in both management and shared edge clusters and compute clusters.	The NSX ESG is the appropriate device for managing north-south traffic. Using ECMP provides multiple paths in and out of the SDDC. This configuration results in faster failover times than deploying Edge service gateways in HA mode. Due to the upstream physical L3 devices, ECMP edges are required.	ECMP requires two VLANS in each region for uplinks, which adds an additional VLAN over traditional HA ESG configurations.
SDDC-VI-SDN-018	Deploy a single NSX UDLR for the management cluster to provide east-west routing across all regions.	Using the UDLR reduces to one the hop count between nodes attached to it. This reduces latency and improves performance.	UDLRs are limited to 1,000 LIFs. When that limit is reached, a new UDLR must be deployed.
SDDC-VI-SDN-019	Deploy a single NSX UDLR for the shared edge and compute clusters and compute clusters to provide east-west routing across all regions for workloads that require mobility across regions.	Using the UDLR reduces to one the hop count between nodes attached to it. This reduces latency and improves performance.	UDLRs are limited to 1,000 LIFs. When that limit is reached, a new UDLR must be deployed.
SDDC-VI-SDN-020	Deploy a DLR for the shared edge and compute clusters and compute clusters to provide east-west routing for workloads that require on-demand network objects from vRealize Automation.	Using the DLR reduces to one the hop count between nodes attached to it. This reduces latency and improves performance.	DLRs are limited to 1,000 LIFs. When that limit is reached, a new DLR must be deployed.
SDDC-VI-SDN-021	Deploy all NSX UDLRs without the local egress option enabled.	When local egress is enabled, control of ingress traffic is also necessary (for example, by using NAT). This becomes difficult to manage for little to no benefit.	All north-south traffic is routed through Region A until those routes are no longer available. At that time, all traffic dynamically changes to Region B.
SDDC-VI-SDN-022	Use BGP as the dynamic routing protocol inside the SDDC.	Using BGP as opposed to OSPF eases the implementation of dynamic routing. There is no need to plan and design access to OSPF area 0 inside the SDDC. OSPF area 0 varies based on customer configuration.	BGP requires configuring each ESG and UDLR with the remote router that it exchanges routes with.
SDDC-VI-SDN-023	Configure BGP Keep Alive Timer to 1 and Hold Down Timer to 3 between the UDLR and all ESGs that provide north-south routing.	With Keep Alive and Hold Timers between the UDLR and ECMP ESGs set low, a failure is detected quicker, and the routing table is updated faster.	If an ESXi host becomes resource constrained, the ESG running on that ESXi host might no longer be

Decision ID	Design Decision	Design Justification	Design Implications
			used, even though it is still up.
SDDC-VI-SDN-024	Configure BGP Keep Alive Timer to 4 and Hold Down Timer to 12 between the ToR switches and all ESGs providing north-south routing.	This provides a good balance between failure detection between the ToR switches and the ESGs and overburdening the ToRs with keep alive traffic.	By using longer timers to detect when a router is dead, a dead router stays in the routing table longer and continues to send traffic to a dead router.
SDDC-VI-SDN-025	Create one or more static routes on ECMP-enabled edges for subnets behind the UDLR and DLR with a higher admin cost than the dynamically learned routes.	When the UDLR or DLR control virtual machine fails over, router adjacency is lost and routes from upstream devices such as ToR switches to subnets behind the UDLR are lost.	This requires each ECMP edge device to be configured with static routes to the UDLR or DLR. If any new subnets are added behind the UDLR or DLR, the routes must be updated on the ECMP edges.

Transit Network and Dynamic Routing

Dedicated networks are needed to facilitate traffic between the universal dynamic routers and edge gateways, and to facilitate traffic between edge gateways and the top-of-rack switches. These networks are used for exchanging routing tables and for carrying transit traffic.

Table 70) Transit network design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-026	Create a universal virtual switch for use as the transit network between the UDLR and ESGs. The UDLR provides east-west routing in both compute and management stacks, while the ESGs provide north-south routing.	The universal virtual switch allows the UDLR and all ESGs across regions to exchange routing information.	Only the primary NSX Manager instance can create and manage universal objects, including this UDLR.
SDDC-VI-SDN-027	Create a global virtual switch in each region for use as the transit network between the DLR and ESGs. The DLR provides east-west routing in the compute stack, while the ESGs provide north-south routing.	The global virtual switch allows the DLR and ESGs in each region to exchange routing information.	A global virtual switch for use as a transit network is required in each region.
SDDC-VI-SDN-028	Create two VLANs in each region. Use those VLANs to enable ECMP between the north-south ESGs and the L3 device (ToR or upstream device). The ToR switches or upstream L3 devices have an SVI on one of the two VLANS, and each north-south ESG has an interface on each VLAN.	This enables the ESGs to have multiple equal-cost routes and provides more resiliency and better bandwidth utilization in the network.	Extra VLANs are required.

Firewall Logical Design

The NSX distributed firewall is used to protect all management applications that are attached to application virtual networks. To secure the SDDC, only other solutions in the SDDC and approved administration IPs can directly communicate with individual components. External-facing portals are accessible via a load balancer VIP.

This simplifies the design by having a single point of administration for all firewall rules. The firewall on individual ESGs is set to allow all traffic. An exception is ESGs that provide ECMP services, which requires the firewall to be disabled.

Table 71) Firewall design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-029	For all ESGs deployed as load balancers, set the default firewall rule to allow all traffic.	Restricting and granting access is handled by the distributed firewall. The default firewall rule does not have to do it.	Explicit rules to allow access to management applications must be defined in the distributed firewall.
SDDC-VI-SDN-030	For all ESGs deployed as ECMP north-south routers, disable the firewall.	Use of ECMP on the ESGs is required. Leaving the firewall enabled, even in allow all traffic mode, results in sporadic network connectivity.	Services such as NAT and load balancing cannot be used when the firewall is disabled.
SDDC-VI-SDN-031	Configure the Distributed Firewall to limit access to administrative interfaces in the management cluster.	Ensures that only authorized administrators can access the administrative interfaces of management applications.	Maintaining firewall rules adds administrative overhead.

Load-Balancer Design

The ESG implements load balancing in NSX for vSphere. The ESG has both a layer 4 and a layer 7 engine that offer different features, which are summarized in Table 72.

Table 72) Layer 4 vs. layer 7 load-balancer engine comparison.

Feature	Layer 4 Engine	Layer 7 Engine
Protocols	TCP	TCP HTTP HTTPS (SSL Pass-through) HTTPS (SSL Offload)
Load-balancing method	Round Robin Source IP Hash Least Connection	Round Robin Source IP Hash Least Connection URI
Health checks	TCP	TCP HTTP (GET, OPTION, POST) HTTPS (GET, OPTION, POST)

Feature	Layer 4 Engine	Layer 7 Engine
Persistence (keeping client connections to the same back-end server)	TCP: SourceIP	TCP: SourceIP, MSRDP HTTP: SourceIP, Cookie HTTPS: SourceIP, Cookie, ssl_session_id
Connection throttling	No	Client Side: Maximum concurrent connections, maximum new connections per second Server Side: Maximum concurrent connections
High availability	Yes	Yes
Monitoring	View VIP, Pool, and Server objects and stats by using CLI and API. View global stats for VIP sessions from vSphere Web Client.	View VIP, Pool, and Server objects and statistics by using CLI and API View global statistics about VIP sessions from vSphere Web Client
Layer 7 manipulation	No	URL block, URL rewrite, content rewrite

Table 73) NSX for vSphere load-balancer design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-032	Use the NSX load balancer.	The NSX load balancer can support the needs of the management applications. Using another load balancer increases cost and adds another component to be managed as part of the SDDC.	None
SDDC-VI-SDN-033	Use an NSX load balancer in HA mode for all management applications.	All management applications that require a load balancer are on a single virtual wire, and having a single load balancer keeps the design simple.	One management application owner might make changes to the load balancer that affect another application.
SDDC-VI-SDN-034	Use an NSX load balancer in HA mode for the Platform Services Controllers.	Using a load balancer increases the availability of the PSCs for all applications.	Configuring the Platform Services Controllers and the NSX load balancer adds administrative overhead.

Information Security and Access Control

Use a service account for authentication and authorization of NSX Manager for virtual network management.

Table 74) Authorization and authentication management design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-SDN-035	Configure a service account svc-	Provides the following access control features:	You must maintain the service account's

Decision ID	Design Decision	Design Justification	Design Implication
	nsxmanager in vCenter Server for application-to-application communication from NSX Manager with vSphere.	• NSX Manager accesses vSphere with the minimum set of permissions that are required to perform lifecycle management of virtual networking objects. • In the event of a compromised account, the accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	life cycle outside of the SDDC stack to preserve its availability.
SDDC-VI-SDN-036	Use global permissions when you create the `svc-nsxmanager` service account in vCenter Server.	• Simplifies and standardizes the deployment of the service account across all vCenter Server instances in the same vSphere domain. • Provides a consistent authorization layer.	All vCenter Server instances must be in the same vSphere domain.

Bridging Physical Workloads

NSX for vSphere offers VXLAN to layer 2 VLAN bridging capabilities with the data path contained entirely in the ESXi hypervisor. The bridge runs on the ESXi host where the DLR control virtual machine is located. Multiple bridges per DLR are supported.

Table 75) Virtual-to-physical-interface type design decision.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-037	Place all management and tenant virtual machines on VXLAN logical switches, unless you must satisfy an explicit requirement to use VLAN-backed port groups for these virtual machines. Where VLAN backed port groups are used, configure routing from VXLAN to VLAN networks. If a layer 2 adjacency between networks is a technical requirement, then connect VXLAN logical switches to VLAN-backed port groups by using NSX L2 Bridging.	Use NSX L2 Bridging only where virtual machines must be on the same network segment as VLAN-backed workloads and routing cannot be used, such as a dedicated backup network or physical resources. Both L2 Bridging and Distributed Logical Routing are supported on the same VXLAN logical switch.	Network traffic from virtual machines on VXLAN logical switches generally is routed. Where bridging is required, the data path occurs through the ESXi host that is running the active Distributed Logical Router Control VM. Therefore, all bridged traffic flows through this ESXi host at the hypervisor level. When scale out is required, you can add multiple bridges per DLR instance that share an ESXi host or multiple DLR instances to distribute bridging across ESXi hosts.

Region Connectivity

Regions must be connected to each other. Connection types can be point-to-point links, MPLS, VPN tunnels, and so on. This connection varies by customer and is out of scope for this design.

The region interconnectivity design must support jumbo frames and provide latency of less than 150ms. For full details on the requirements for region interconnectivity, see the Cross-VC NSX Design Guide.

Table 76) Intersite connectivity design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-038	Provide a connection between regions that is capable of routing between each cluster.	When NSX is configured for cross-vCenter to enable universal objects, connectivity between NSX manager instances, ESXi host VTEPs, and NSX Controller instances to ESXi hosts management interface is required. To support cross-region authentication, the vCenter Server and Platform Services Controller design requires a single vCenter Single Sign-On domain. Portability of management and compute workloads requires connectivity between regions.	Jumbo frames are required across regions.
SDDC-VI-SDN-039	Make sure that the latency between regions is less than 150ms.	A latency below 150ms is required for the following features: • Cross-vCenter vMotion • The NSX design for the SDDC	None

Application Virtual Network

Management applications, such as VMware vRealize Automation, VMware vRealize Operations Manager, and VMware vRealize Orchestrator, leverage a traditional three-tier client and server architecture. This architecture is composed of a presentation tier (user interface), a functional process logic tier, and a data tier. This architecture requires a load balancer for presenting end-user facing services.

Table 77) Isolated management applications design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-040	Place the following management applications on an application virtual network: • vRealize Automation • vRealize Automation Proxy Agents • vRealize Business • vRealize Business collectors • vRealize Operations Manager • vRealize Operations Manager remote collectors • vRealize Log Insight • Update Manager Download Service	Access to the management applications is only through published access points.	The application virtual network is fronted by an NSX Edge device for load balancing and the distributed firewall to isolate applications from each other and from external users. Direct access to application virtual networks is controlled by distributed firewall rules.
SDDC-VI-SDN-041	Create three application virtual networks. • Each region has a dedicated application virtual network for management applications in that region that does not require failover.	Using only three application virtual networks simplifies the design by	A single /24 subnet is used for each application virtual network. IP management is crucial to make sure that no shortage of IP addresses occurs.

Decision ID	Design Decision	Design Justification	Design Implications
	• One application virtual network is reserved for management application failover between regions.	sharing layer 2 networks with applications based on their needs.	

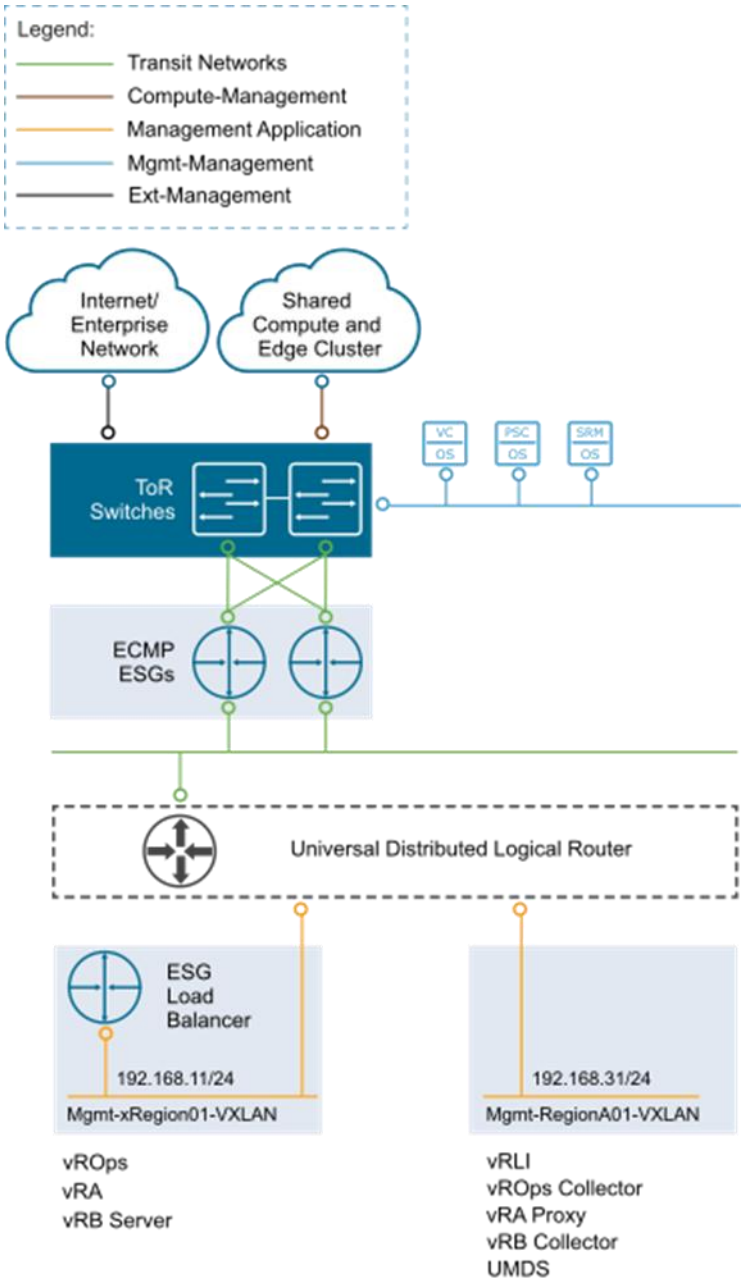
Table 78) Portable management applications design decisions.

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-042	The following management applications must be easily portable between regions: • vRealize Automation • vRealize Business • vRealize Operations Manager	Management applications must be easily portable between regions without requiring reconfiguration.	Unique addressing is required for all management applications.

Having software-defined networking based on NSX in the management stack makes all NSX features available to the management applications.

This approach to network virtualization service design improves security and mobility of the management applications and reduces the integration effort with existing customer networks.

Figure 45) Virtual application network components and design.



The following configuration choices might later facilitate the tenant onboarding process:

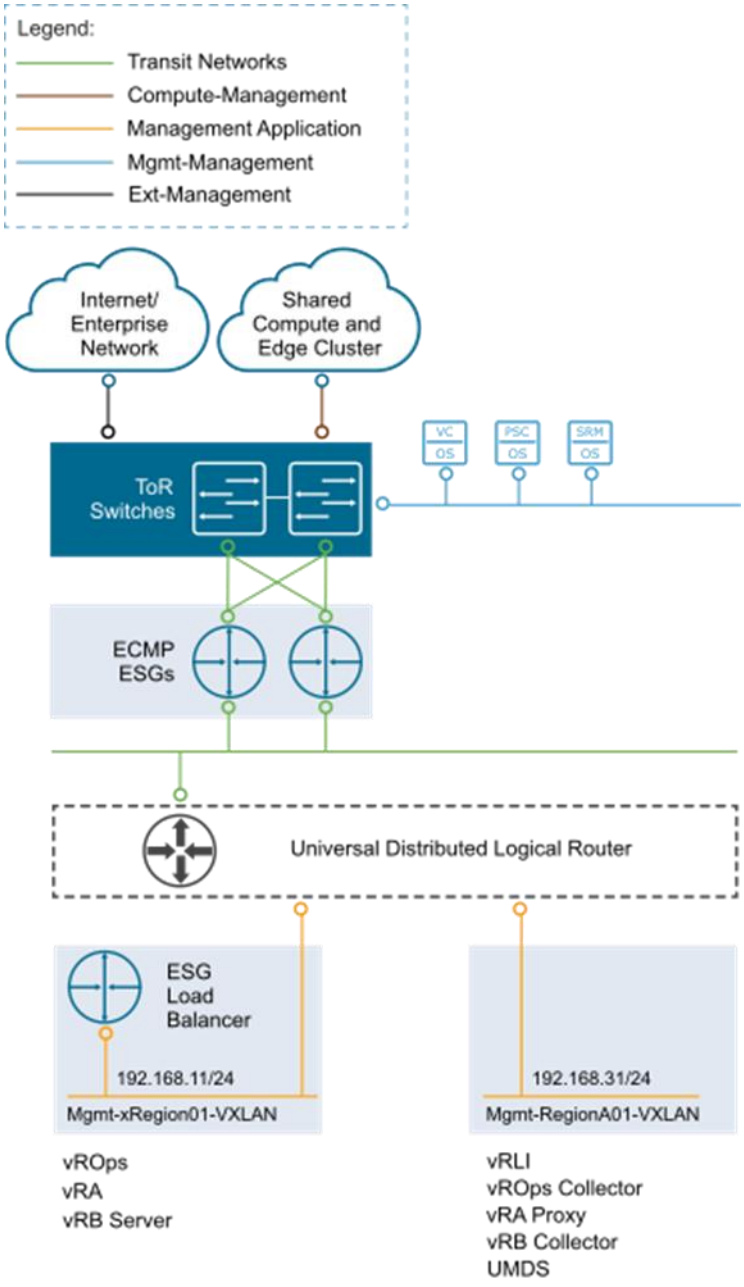
- Create the primary NSX ESG to act as the tenant PLR and the logical switch that forms the transit network for use in connecting to the UDLR.
- Connect the primary NSX ESG uplinks to the external networks.
- Connect the primary NSX ESG internal interface to the transit network.
- Create the NSX UDLR to provide routing capabilities for tenant internal networks and connect the UDLR uplink to the transit network.
- Create any tenant networks that are known up front and connect them to the UDLR.

Virtual Network Design Example

The virtual network design example illustrates an implementation for a management application virtual network.

Figure 46 shows an example of implementing a management application virtual network. The example service is vRealize Automation, but any other three-tier application would look similar.

Figure 46) Detailed example for vRealize Automation networking.



The example is set up as follows:

- Deploy vRealize Automation on the application virtual network that is used to fail over applications between regions. This network is provided by a VXLAN virtual wire (the orange network in Figure 46).

- The network used by vRealize Automation connects to external networks through NSX for vSphere. NSX ESGs and the UDLR route traffic between the application virtual networks and the public network.
- Services such as a web GUI, which must be available to the end users of vRealize Automation, are accessible through the NSX Edge load balancer.

The following table shows an example of a mapping from application virtual networks to IPv4 subnets. The actual mapping depends on the customer's environment and is based on available IP subnets.

Note: The following IP ranges are an example. Your actual implementation depends on your environment.

Table 79) Example IP ranges.

Application Virtual Network	Management Applications	Internal IPv4 Subnet
Mgmt-xRegion01-VXLAN	vRealize Automation (includes vRealize Orchestrator and vRealize Business) vRealize Operations Manager	192.168.11.0/24
Mgmt-RegionA01-VXLAN	vRealize Log Insight vRealize Operations Manager Remote Collectors vRealize Automation Proxy Agents	192.168.31.0/24
Mgmt-RegionB01-VXLAN	vRealize Log Insight vRealize Operations Manager Remote Collectors vRealize Automation Proxy Agents	192.168.32.0/24

Use of Secure Sockets Layer Certificates

By default, NSX Manager uses a self-signed SSL certificate. This certificate is not trusted by end-user devices or web browsers. When administrators or end users interact with the various management and consumption portals in the system, they must manually trust the self-signed certificates presented to them.

10.8 Shared Storage Design

The shared storage design includes design decisions for iSCSI-based NetApp HCI shared storage as well as NFS-based NetApp ONTAP Select shared storage.

Shared Storage Platform

Primary shared storage in the NetApp HCI platform is composed of a scale-out cluster of storage nodes based on the SolidFire all-flash shared-nothing technology.

Storage Cluster Architecture

Unlike other storage architectures that you may be familiar with, NetApp HCI storage clusters with Element software do not use any of the following:

- RAID
- Disk groups or aggregates
- Caching tiers
- Multiple drive types

- Spare disks
- Centralized controllers
- Drive shelves
- Postprocessing for deduplication or compression

Thus there are very few knobs to turn and little up-front architecture and planning that must be done. Simply add the number and size of nodes to the cluster appropriate for your performance (measured in IOPS) and capacity requirements (measured in usable TB).

SolidFire Helix Data Protection

A NetApp HCI storage cluster maintains data with the highest levels of availability, data protection, and security, architected specifically for cloud-scale, all-flash infrastructures.

SolidFire Helix data protection is a distributed replication algorithm that spreads at least two redundant copies of data across all drives in the system, delivering fault tolerance that is superior to traditional disk and all-flash array systems. This “RAID-les”’ approach has no single point of failure and allows the system to absorb multiple concurrent failures across all levels of the storage solution. Failures are isolated—avoiding performance effects on other aspects of the system—while all QoS settings remain enforced.

- No single point of failure
- Self-healing architecture
- Nondisruptive upgrades
- Fully automated
- Five 9s availability
- Rapid full mesh rebuilds
- Proven in the world’s most demanding data centers

Data Efficiencies

The always-on SolidFire architecture incorporates inline deduplication, compression, thin provisioning, and space-efficient snapshots that are performed across the entire data store with no effect on performance.

- **Global inline deduplication.** SolidFire always-on data deduplication works nondisruptively across the entire cluster to maximize data reduction.
- **Two-layer compression.** SolidFire uses both in-line and postprocess compression, optimized for space without sacrificing performance.
- **4k granular thin provisioning.** SolidFire uses 4k granular thin provisioning that does not require any reserve space, returning data immediately back to free space for use by other SolidFire volumes.

Scale Out

NetApp HCI storage cluster technology is an example of a scale-out architecture. This means that you can grow the system by adding more nodes, each of which contributes both disk capacity and controller throughput. This cluster of equal peers functions as a single logical unit.

A minimum of four nodes is required to build the initial cluster. As your need for capacity and performance grows, you simply add nodes to the cluster. The data is automatically rebalanced to take advantage of the new nodes, and iSCSI connections are seamlessly redistributed for the same reason. All active volumes running on the cluster remain the same, and no remaps or rescans are necessary.

It is also permissible to mix and match models and generations when adding new nodes to the cluster.

You can add nodes in this manner until you reach a supported maximum of 40 in a single NetApp HCI storage cluster.

Management and Storage Traffic Networks

The NetApp HCI storage cluster requires two separate networks. They must be on different subnets on different VLANs. You should reserve the iSCSI network for storage traffic, and you can share the management VLAN with other management constructs in your private cloud (for example, vCenter).

Figure 47) Required networks for storage cluster.

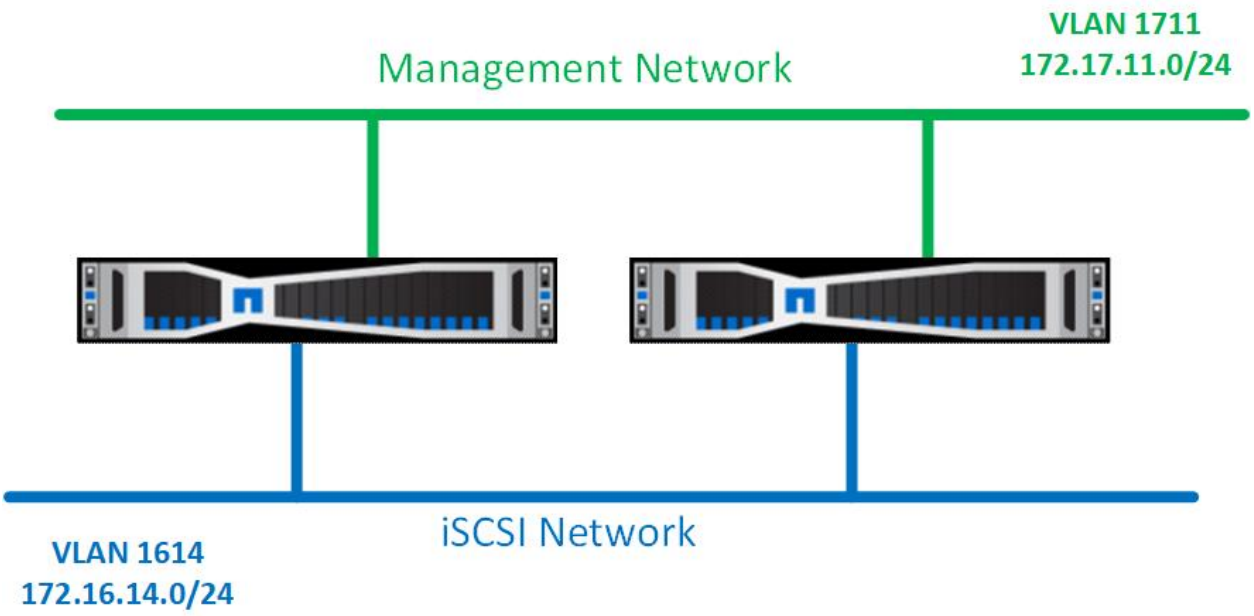


Table 80) Network functions.

Network	Function	Sample VLAN	Sample Subnet
iSCSI	• Traffic to or from ESXi hosts • Intranode replication traffic in the storage cluster	1614	172.16.14.0/24
Management	• In-band management traffic	1711	172.17.11.0/24

These two networks must be the same across all nodes of the storage cluster. All nodes must be L2 adjacent in this manner for the SVIP (storage virtual IP) and MVIP (management virtual IP) to be able to move to any node of the cluster.

Further, although it is possible to route iSCSI, the NetApp Deployment Engine (NDE) only supports a single VLAN/subnet for all iSCSI traffic. This means that you must have the iSCSI port groups of your ESXi hosts on the same subnet, as shown in Figure 48.

Figure 48) Compute node to storage node network mapping.

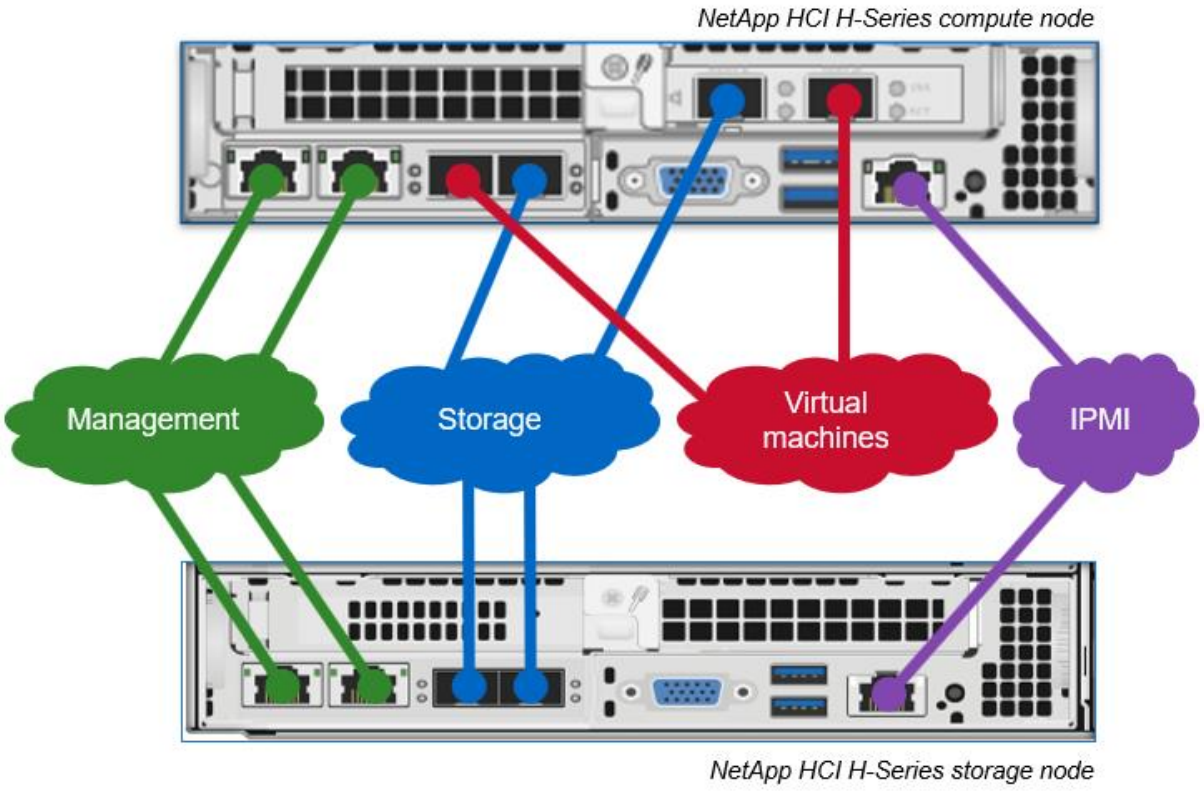
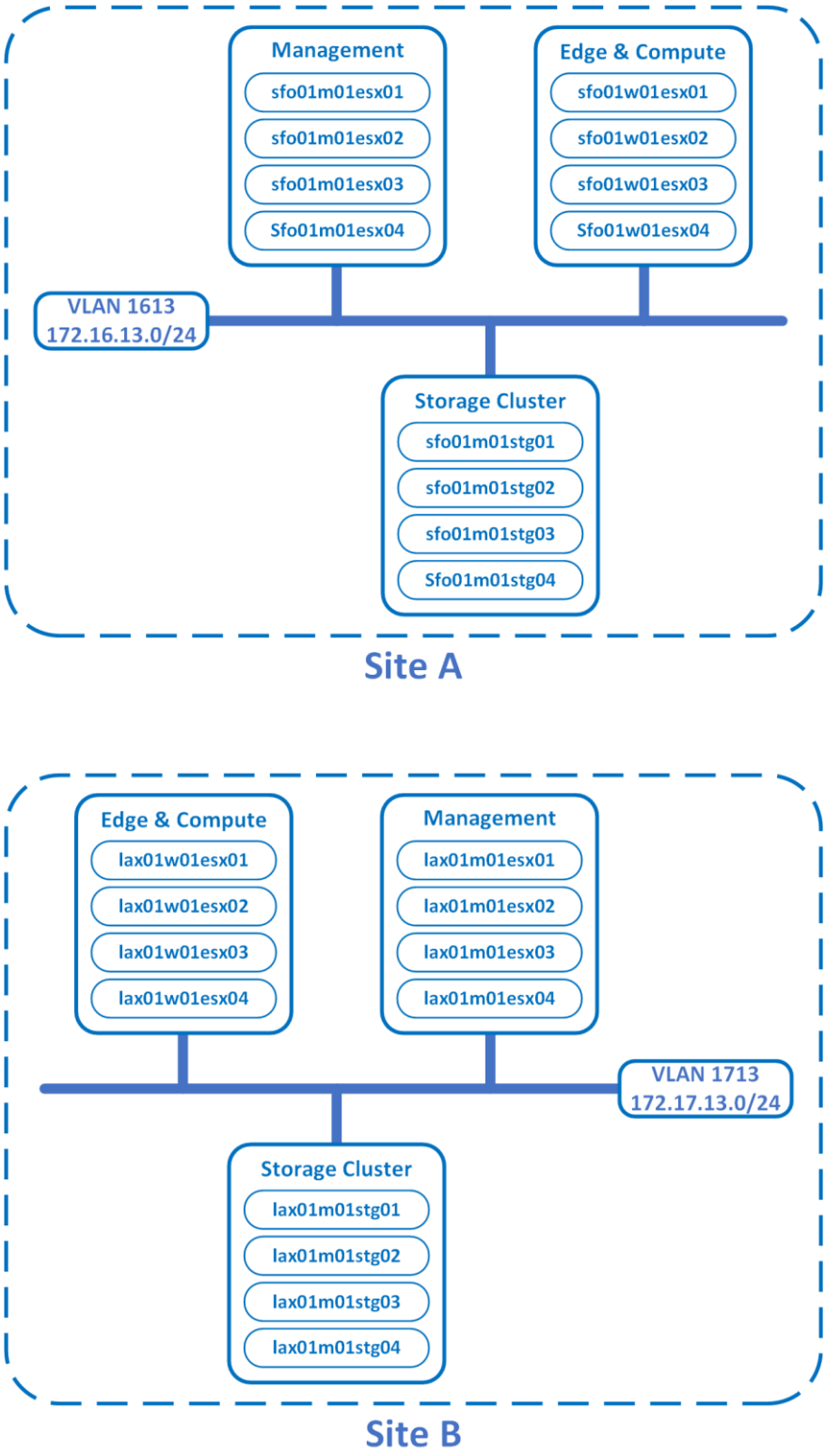


Figure 49) Compute cluster to storage cluster mapping.



Security

NetApp HCI full disk encryption provides comprehensive security for data at rest without sacrificing storage system performance or ease of use. The system uses a 256-bit password distributed across all the nodes in the system and prevents administrator access to the keys.

The self-encrypting drives work entirely at the hardware level, thus enabling physical data security with no effect on performance. No CPU cycles are taken from the host, and I/O transfer to the hosts is not interrupted.

The cluster-wide password that is used to lock drives is unique to each cluster. If any drives are removed from the system, the password is not present on the drives, and all data is encrypted. Therefore, a drive or a node that is removed ungracefully from a cluster cannot be accessed until it is securely erased or returned to its original cluster.

As an administrator, you can set security authorizations and apply them to all self-encrypting drives in the cluster. NetApp recommends that you do so, especially if you have any applications that are subject to compliance regimes, such as HIPAA or PCI.

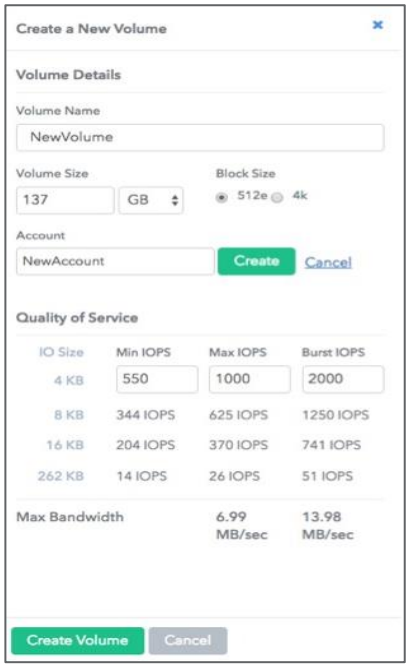
Quality of Service

NetApp HCI QoS technology enables IT administrators to easily assign and guarantee levels of performance (IOPS and bandwidth) to thousands of volumes residing in a single storage platform. The proper use of QoS makes sure that no workload is able to affect the performance of another. “Noisy neighbor” problems are particularly common in private cloud deployments, making it very difficult to deliver deterministic performance to all workloads in the system when using traditional storage platforms.

Through performance virtualization, NetApp HCI proactively provides applications with the performance they require throughout the life of their deployment. With guaranteed QoS from SolidFire, applications no longer compete for performance, and administrators no longer struggle with complex tiering systems or prioritization schemes. You allocate capacity and performance to each storage volume, which you can change dynamically without migrating data or affecting performance.

A NetApp HCI storage cluster can provide QoS parameters on a per-volume basis, as shown in Figure 50.

Figure 50) QoS parameters.



Cluster performance is measured in IOPS. QoS parameters are defined by three values:

- **Min IOPS.** The guaranteed level of performance for a volume. Performance does not drop below this level.
- **Max IOPS.** The maximum number of sustained IOPS that the NetApp HCI storage cluster provides to a volume.
- **Burst IOPS.** The maximum number of IOPS allowed in a short-burst scenario. NetApp HCI uses Burst IOPS when a cluster is running in a state of low cluster IOPS utilization. A single volume can accrue Burst IOPS and use the credits to burst above its Max IOPS up to its Burst IOPS level for a set burst period. A volume can burst for up to 60 seconds if the cluster has the capacity to accommodate the burst.

  A volume accrues 1 second of burst credit (up to a maximum of 60 seconds) for every second that the volume runs below its Max IOPS limit.

It is important to tune these parameters correctly in accordance with the needs of the various workloads deployed in your private cloud.

For more detailed information on how to architect QoS policies for particular workload types, refer to the technical report NetApp SolidFire Quality of Service.

SolidFire mNode

The SolidFire management node (mNode) is one of two virtual machines automatically deployed and configured on every NetApp HCI system during a new installation. The mNode is not a required component, but it enables the following features:

- NetApp HCI vCenter Plug-In
- NetApp Monitoring Agent (NMA)
- NetApp Active IQ® Collector
- SolidFire support reverse VPN tunnel
- SolidFire system upgrades

HCI Protection Domains

The Element software component of NetApp HCI 1.4 supports protection domains for storage clusters using chassis-based nodes. When storage is distributed evenly across multiple chassis, Element software automatically heals from certain types of chassis-wide failures, and alerts you when automatic healing is not possible. You can also use the Element API to integrate with third-party monitoring tools to inform you of protection status.

Connection Balancing

One of the key benefits of any scale-out storage solution is the elimination of a single controller (or controller pair) as a gateway to the storage on the back end. Instead, every node acts as a controller and is capable of servicing requests from hosts. This greatly increases the potential throughput of the system.

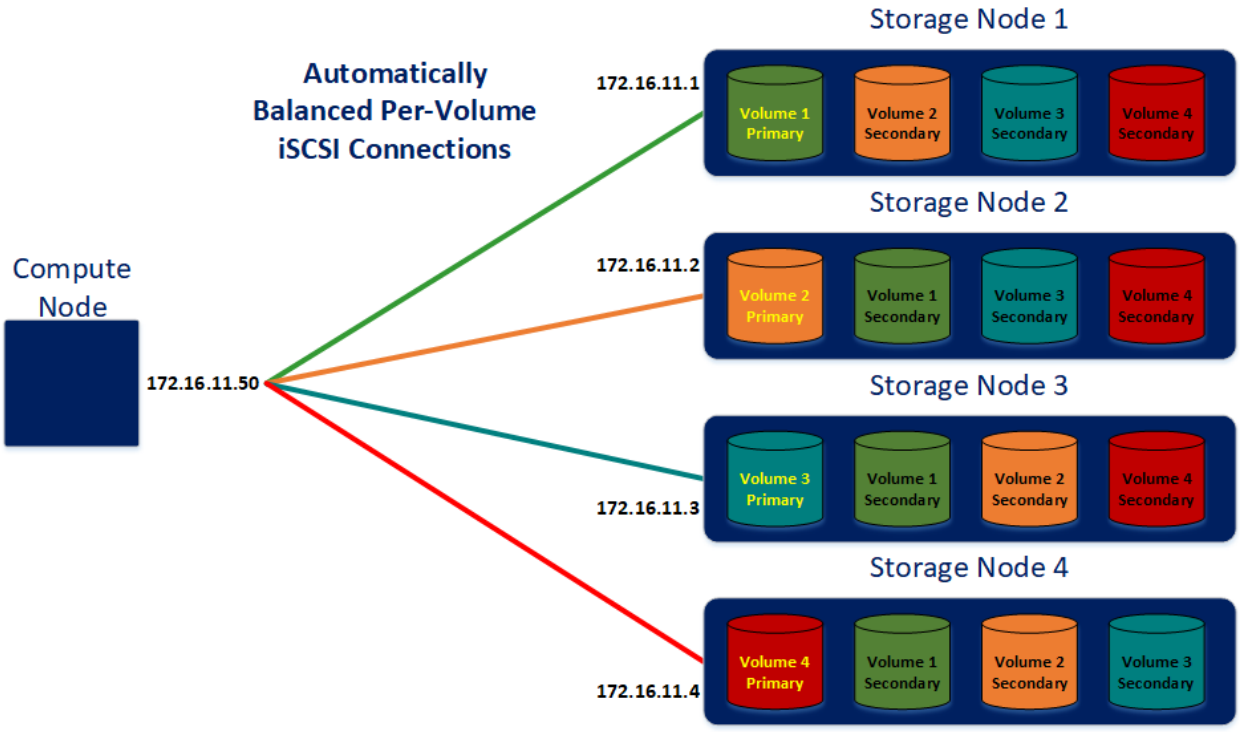
However, an automated mechanism is required to enable hosts to balance their connections to the storage cluster in such a way that traffic spans the nodes. Otherwise, this balancing would require painstaking and error-prone manual configuration. For example, one host is set to talk to controller 1, another host is set to talk to controller 2, and so on. The environment would also need to be reconfigured whenever new storage nodes are added.

In the case of a NetApp HCI storage cluster, balancing is achieved by using a feature in iSCSI called login redirection.

When a compute node wants to connect to the storage cluster, it starts by sending a login request to the IP that ESXi has configured as its iSCSI target, which is known as the cluster storage virtual IP (SVIP). This IP can move around between nodes, but it lives on only one node at any given time. When the login request comes in, the node with the SVIP decides which node should handle the traffic for the requested volume and then redirects the session to that node. This decision is based on the traffic load across the nodes. This way, the client sends all of its traffic to the node best suited from a load perspective.

If a single node is found to be handling too much traffic for its volumes, or if a new node is added and volumes are redistributed, the storage cluster asks the host to log out. Immediately after logging out, the host attempts to log on to the SVIP again, and it is redirected to whatever new node that volume’s primary metadata copy was moved to.

Figure 51) iSCSI connection balancing.



The same approach works if a node is removed from a cluster or if a node goes down. In the case of node removal, all of the volumes served by that node request a logout and the host is redirected to a different node when it attempts to log in. If a node goes down, the host attempts to log on again. As soon as the storage cluster realizes that the node is down, login requests are redirected to nodes that are up. In this way, iSCSI login redirection provides self-healing capabilities to the HCI storage cluster.

Volume Configuration

Capacity is presented to hosts from a NetApp HCI storage cluster in volumes, which are presented as block devices to the ESXi hosts. In this design, the relationship between volumes and datastores is 1:1. In other words, each volume provisioned contains a single VMFS-6 datastore.

The minimum volume configuration per region for this design is shown in Table 81.

Table 81) Minimum volume and datastore configuration per region.

Cluster	Minimum Number of Volumes	Capacity Per Volume	IOPS Min Per Volume	IOPS Max Per Volume	IOPS Burst Per Volume
Management	4 volumes/datastores	2TB	2,000	10,000	25,000
Edge and compute	8 volumes/datastores	4TB	10,000	20,000	50,000
TOTALS	12 volumes/datastores	40TB	88,000	200,000	500,000

This minimum volume capacity and IOPS configuration assumes the following:

- A minimum NetApp HCI storage cluster configuration of four H300S storage nodes.
- Specific workload IOPS requirements are unknown at deployment time.

- No individual workload's virtual disks require more than 10,000 IOPS on a consistent basis.
- An efficiency rating of 4:1 is achieved from the effects of deduplication, compression, and thin provisioning.
- 16TB of effective capacity is reserved for snapshots, backups, and ONTAP Select.
- A minimum 3:1 volume-to-storage-node ratio creates ideal connection and back-end balancing.

Larger storage clusters with similar assumptions would simply provision additional, identically configured 4TB datastores to the compute cluster or clusters, up to the limit of usable capacity in that particular cluster.

If specific individual workload virtual disk IOPS requirements exceed those provided by this configuration, NetApp recommends implementing a tiered setup that places high I/O workloads into a gold tier datastore, while leaving other less demanding workloads on silver tier datastores, as shown in Table 82.

Table 82) Example tiered volume configuration.

Volume Type	Total Number of Volumes	Capacity Per Volume	IOPS Min Per Volume	IOPS Max Per Volume	IOPS Burst Per Volume
Management cluster	4 volumes/datastores	2TB	2,000	10,000	25,000
Compute Gold	2 volumes/datastores	4TB	25,000	50,000	100,000
Compute Silver	6 volumes/datastores	4TB	10,000	20,000	50,000
Totals	12 volumes/datastores	40TB	118,000	260,000	600,000

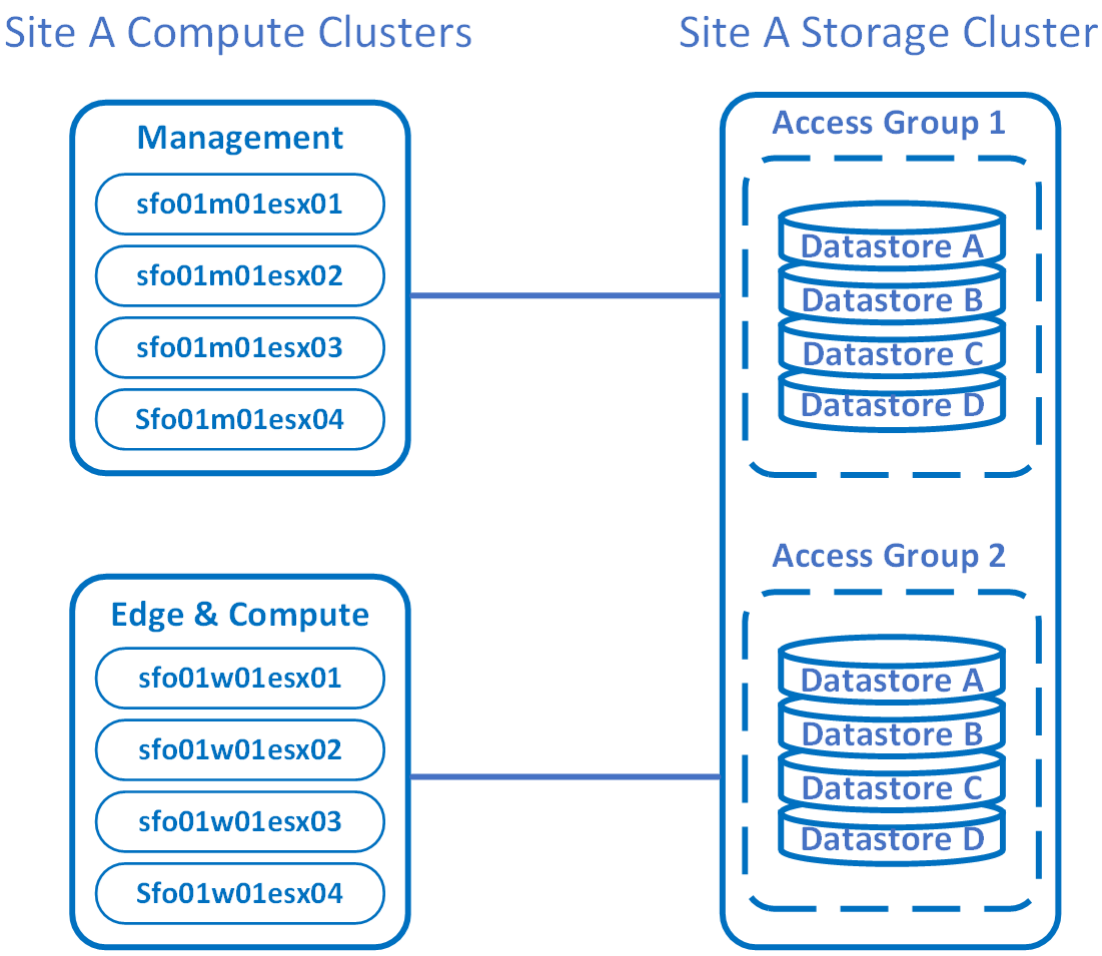
This example of volume capacity and IOPS configuration assumes the following:

- A NetApp HCI storage cluster configuration of six H300S storage nodes.
- No individual gold-tiered workload requires more than 25,000 IOPS on a consistent basis.
- No individual silver-tiered workload requires more than 10,000 IOPS on a consistent basis.
- A 4:1 efficiency rating is achieved from deduplication, compression, and thin provisioning.
- 16TB of effective capacity is reserved for snapshots, backups, and ONTAP Select.
- A minimum 2:1 volume-to-storage node ratio creates ideal connection and back-end balancing.

Datastore Access

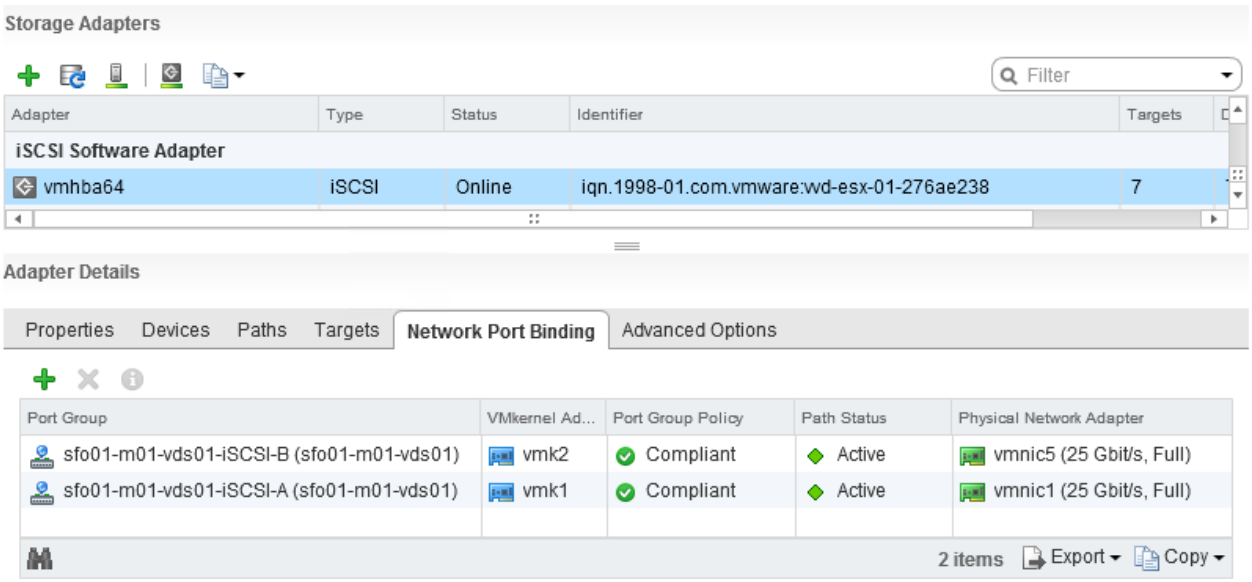
Although this design shares physical storage between the management and edge and compute clusters within a region, it is important to separate the VMFS datastores between them. In a NetApp HCI storage cluster, this is done through the use of Volume Access Groups. There should be two such groups per region: one that contains all of the IQNs for the ESXi hosts in the management cluster and one that contains the IQNs for all hosts in the compute clusters.

Figure 52) Datastore layout per region.



Volumes located on NetApp HCI storage clusters are accessed through the standard iSCSI software adapter available in vSphere. Software adapter creation and port bindings are automatically configured for you by the NDE when it sets up the hosts.

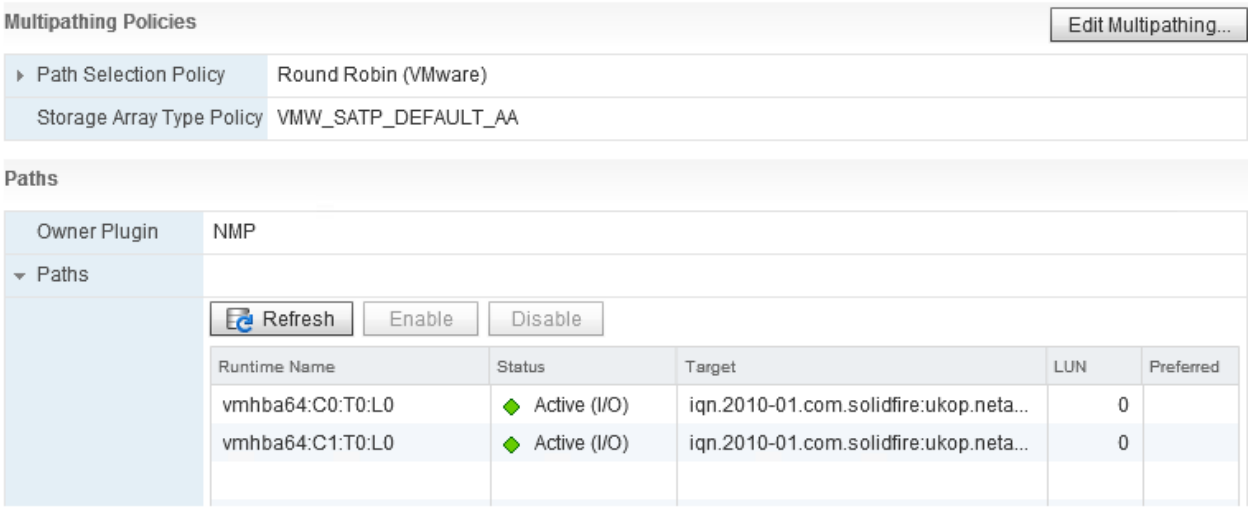
Figure 53) iSCSI software adapter port binding.



iSCSI Multipathing for Compute Nodes

Multipathing for NetApp HCI storage clusters uses the vSphere Native Multipathing Plug-in (NMP). This plug-in should use the round-robin path selection policy (VMware). Like the iSCSI software adapter and port bindings, this setting is automatically configured by NDE.

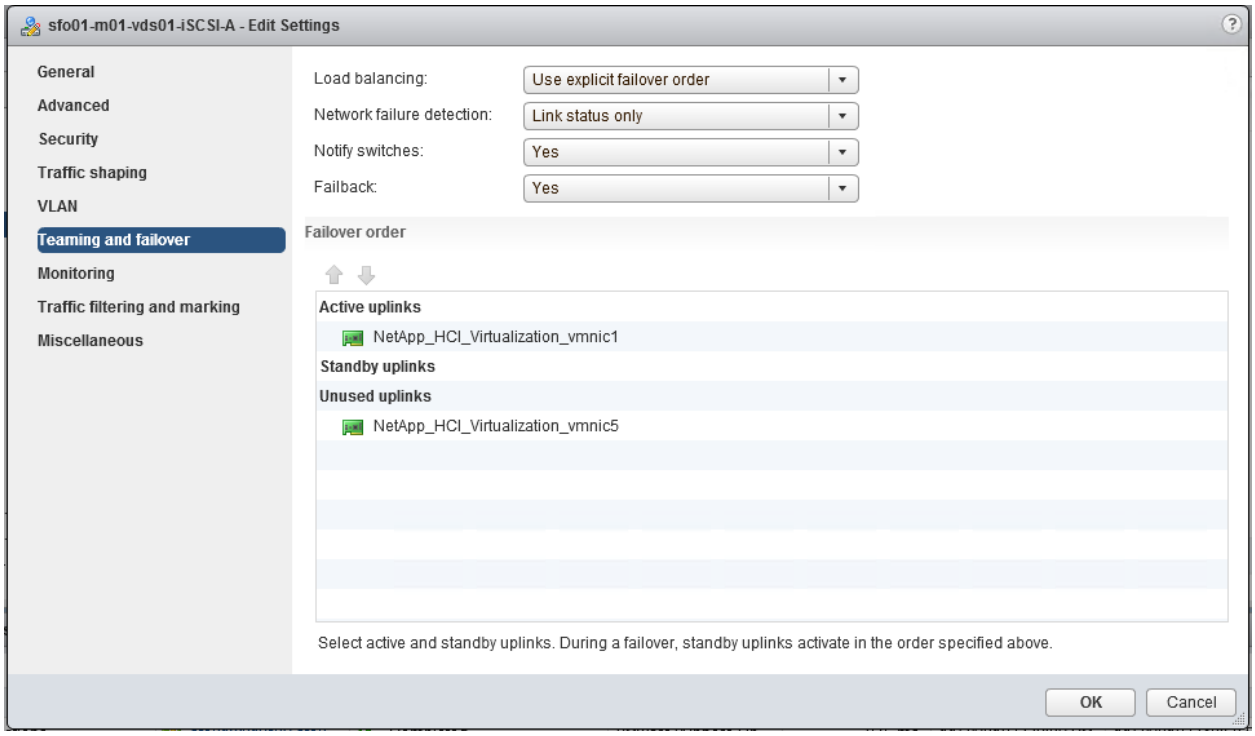
Figure 54) iSCSI device multipathing policy.



iSCSI Teaming and Failover Policy for Compute Nodes

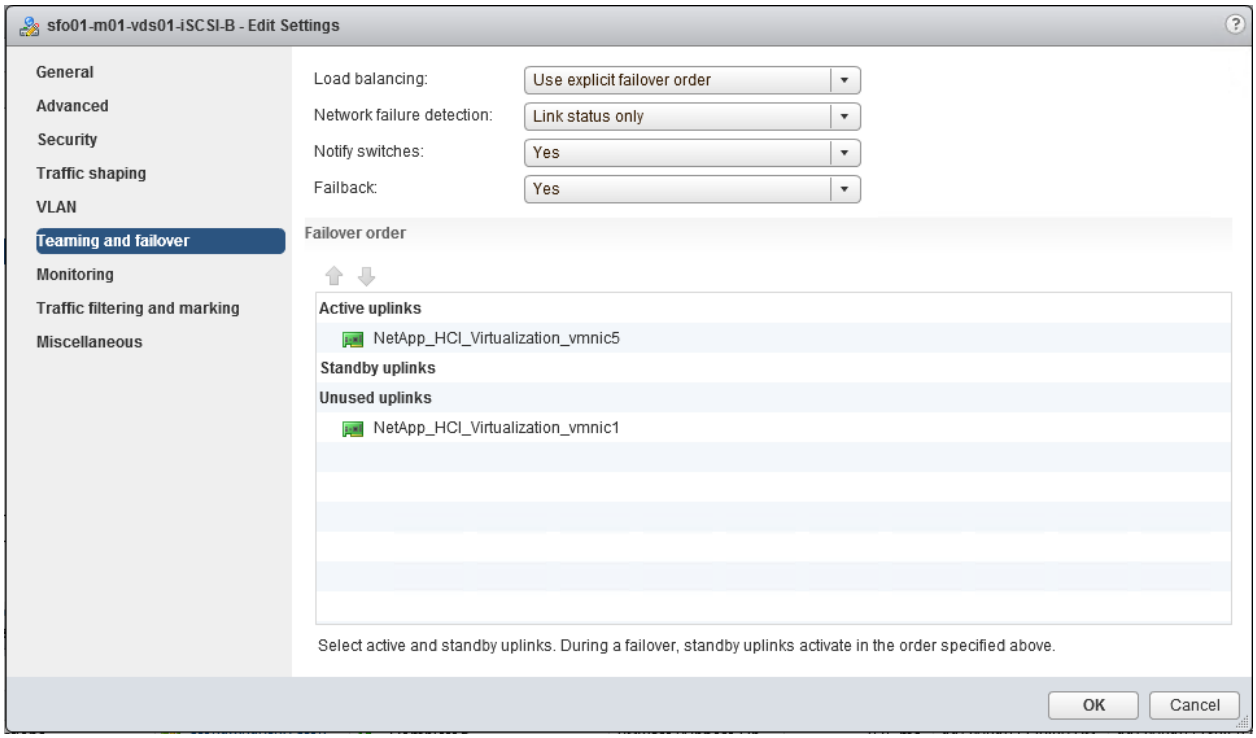
Teaming policy for the iSCSI port groups (*-iSCSI-A and *-iSCSI-B) should be configured such that *-iSCSI-A has only vmnic1 set as an active uplink, with vmnic5 set to unused.

Figure 55) Port group iSCSI-A teaming policy.



The teaming policy for iSCSI-A is set for explicit failover, and only uses vmnic1. *-iSCSI-B should be configured in the opposite way, with vmnic5 active and vmnic1 unused.

Figure 56) Port group iSCSI-B teaming policy.



iSCSI-B's teaming policy is set for explicit failover, and only uses vmnic5. Like the rest of the iSCSI configuration, the NDE configures this parameter automatically.

NetApp Element vCenter Plug-in

Although the NetApp HCI storage clusters have their own UI that can be accessed directly, in this design, most management tasks are performed with the NetApp Element vCenter Plug-in. This is a standard vCenter web client plug-in that provides a scalable, user-friendly interface to discover, configure, manage, and monitor NetApp HCI systems, both compute and storage nodes.

The plug-in bundles common administrative tasks into single operations. One example is datastore creation. Navigate to the NetApp Element Management widget, select Management > Datastores, and choose Create Datastore. When you click Submit on the form, the following tasks are all executed:

- Volume creation on the NetApp HCI storage cluster
- QOS policy application
- Presentation of the volume to relevant ESXi hosts' IQNs
- Creation of the VMFS6 file system
- A rescan of relevant ESXi hosts' storage

Figure 57) The NetApp Element vCenter Plug-in UI.

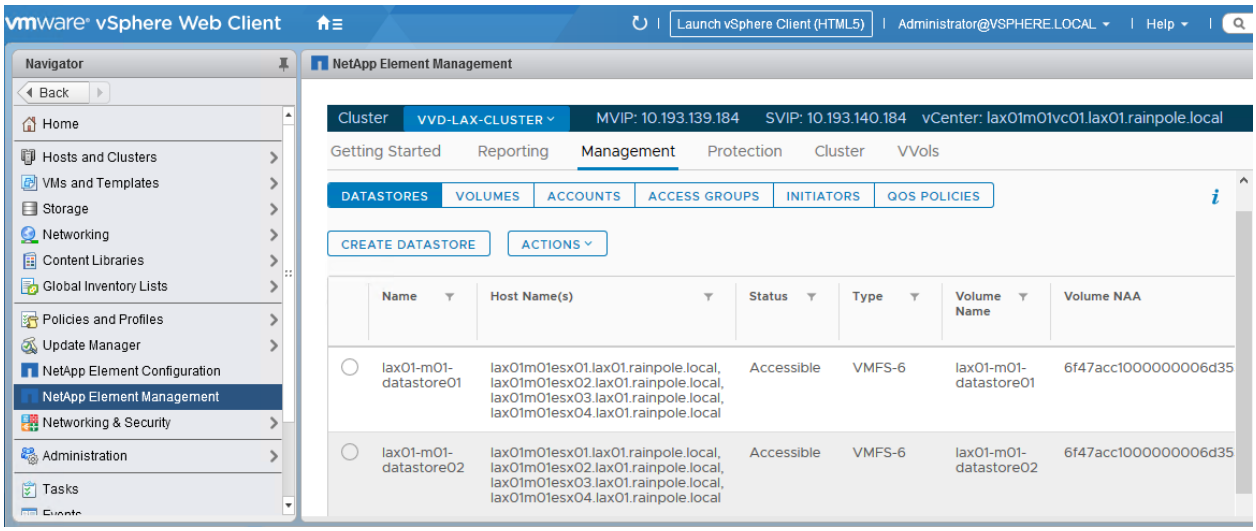


Figure 58 illustrates the paths by which data is communicated to and from the plug-in and various other components of the system.

Figure 58) NetApp HCI vSphere Plug-In telemetry paths.

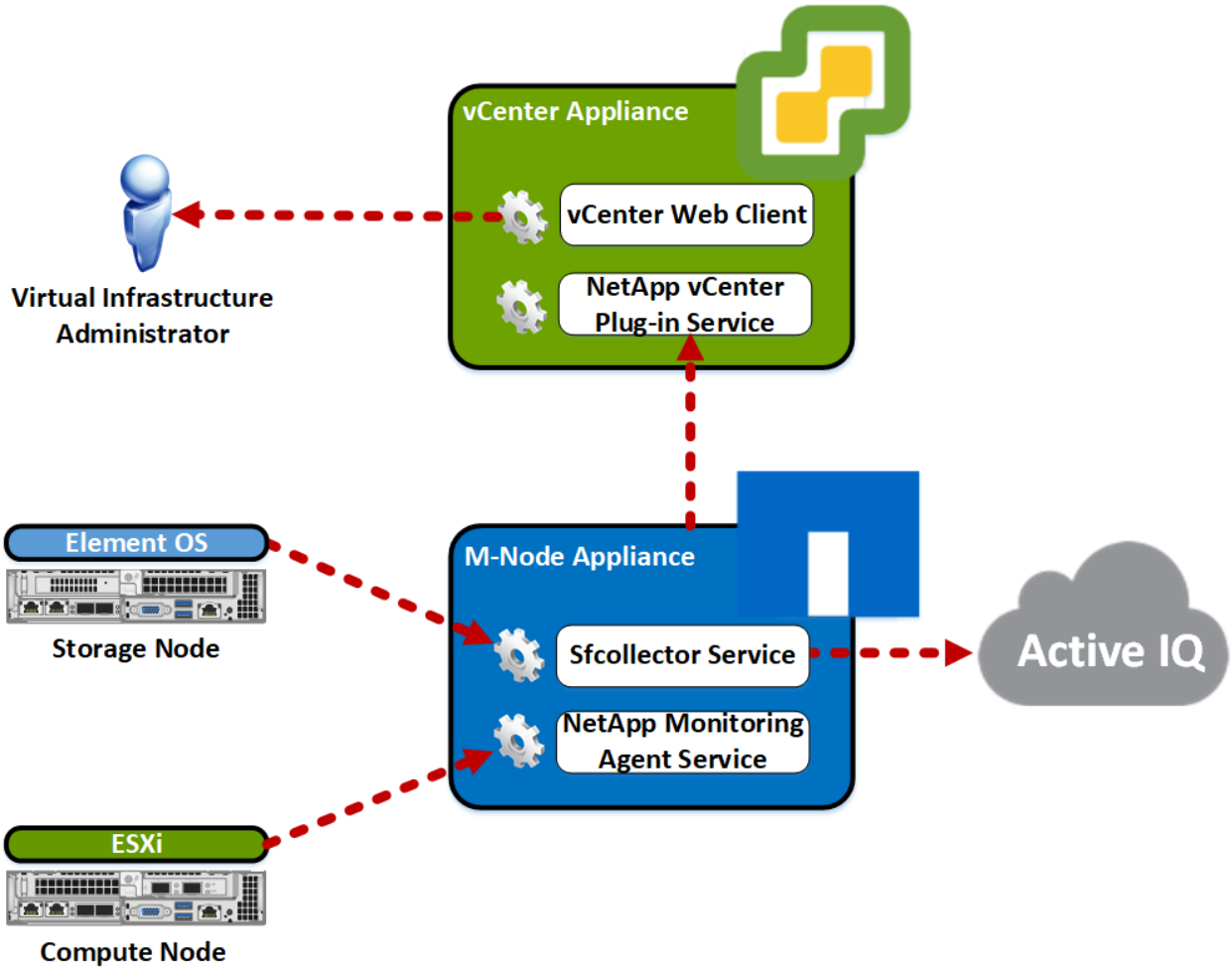
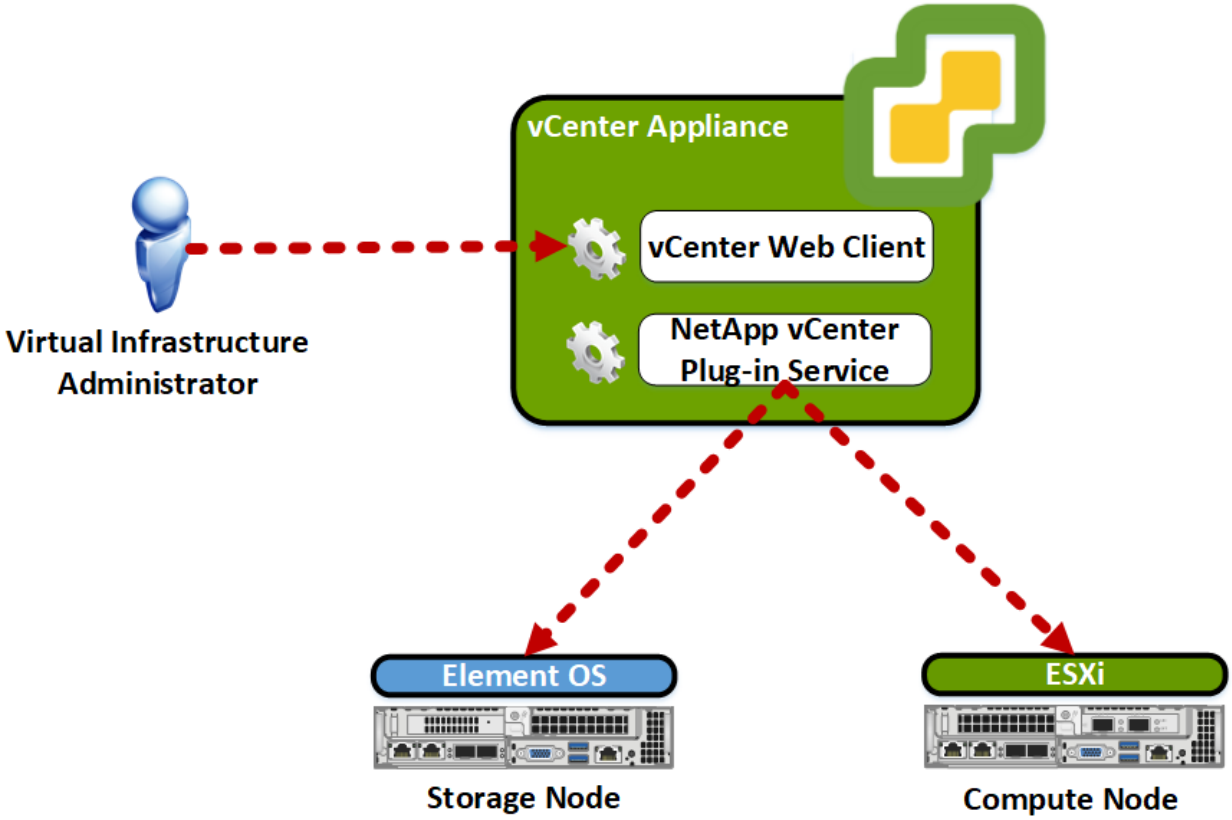


Figure 59 illustrates the paths by which commands are communicated to and from the plug-in and various other components of the system.

Figure 59) NetApp HCI vSphere Plug-in command paths.



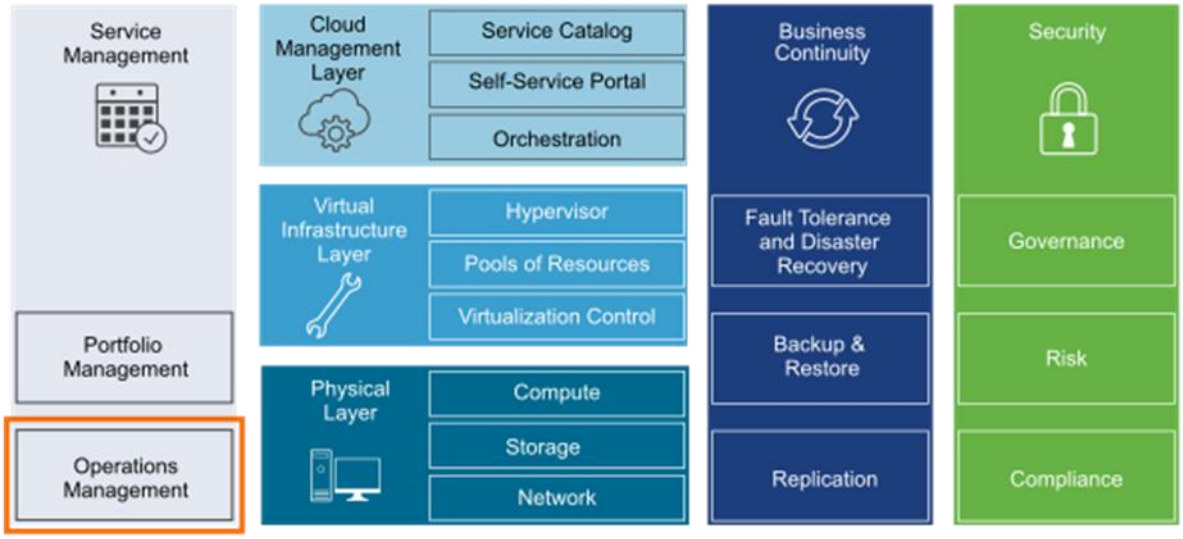
For more detailed information about the vCenter Plug-In for NetApp HCI, see the NetApp SolidFire Plug-in for VMware vCenter Server Web Client Version 4.1 User Guide.

11 Operations Management Design

The operations management design includes the software components that make up the operations management layer. The design offers guidance on the main elements of a product design, including deployment, sizing, networking, diagnostics, security, and integration with management solutions.

- Monitoring operations support in vRealize Operations Manager and vRealize Log Insight provides performance and capacity management of related physical and virtual infrastructure and cloud management components.
- Features of vSphere Update Manager support upgrading and patching of the ESXi hosts in the SDDC.

Figure 60) Operations management in the SDDC layered architecture.



- vRealize Operations Manager Design

  The foundation of vRealize Operations Manager is a single instance of a 3-node analytics cluster deployed in the protected region of the SDDC and a 2-node remote collector group in each region. The components run on the management cluster in each region.
- vRealize Log Insight Design

  vRealize Log Insight design enables real-time logging for all components that build up the management capabilities of the SDDC in a dual-region setup.
- vSphere Update Manager Design

  vSphere Update Manager supports patch and version management of ESXi hosts and virtual machines. vSphere Upgrade Manager is connected to a vCenter Server instance to retrieve information about and push upgrades to the managed hosts.

11.1 vRealize Operations Manager Design

The foundation of vRealize Operations Manager is a single instance of a 3-node analytics cluster that is deployed in the protected region of the SDDC, and a 2-node remote collector group in each region. The components run on the management cluster in each region.

- **Logical and Physical Design of vRealize Operations Manager**

  vRealize Operations Manager communicates with all management components in all regions of the SDDC to collect metrics that are presented through a number of dashboards and views.
- **Node Configuration of vRealize Operations Manager**

  The analytics cluster of the vRealize Operations Manager deployment contains the nodes that analyze and store data from the monitored components. Deploy a configuration of the analytics cluster that satisfies the requirements for monitoring the number of virtual machines in the design objectives of this validated design.
- **Networking Design of vRealize Operations Manager**

  Provide isolation and failover of the vRealize Operations Manager nodes by placing them in several network segments. This networking design also supports public access to the analytics cluster nodes.
- Information Security and Access Control in vRealize Operations Manager

  Protect the vRealize Operations Manager deployment by providing centralized role-based authentication and secure communication with the other components in the SDDC. Dedicate a set of

service accounts to the communication between vRealize Operations Manager and the management solutions in the data center.

- **Monitoring and Alerting in vRealize Operations Manager**

  Use vRealize Operations Manager to monitor the state of the SDDC management components in the SDDC by using dashboards. You can use the self-monitoring capability of vRealize Operations Manager to receive alerts about issues that are related to its operational state.

- **Management Packs in vRealize Operations Manager**

  The SDDC contains VMware products for network, storage, and cloud management. You can monitor and perform diagnostics on all of them in vRealize Operations Manager by using management packs.

- **Disaster Recovery of vRealize Operations Manager**

  To preserve monitoring functionality when a disaster occurs, the design of vRealize Operations Manager supports failing over a subset of the components between regions. Disaster recovery covers only the analytics cluster components, including the master, replica, and data nodes. The region-specific remote collector nodes remain in the affected region.

Logical and Physical Design of vRealize Operations Manager

vRealize Operations Manager communicates with all management components in all regions of the SDDC to collect metrics that are presented through a number of dashboards and views.

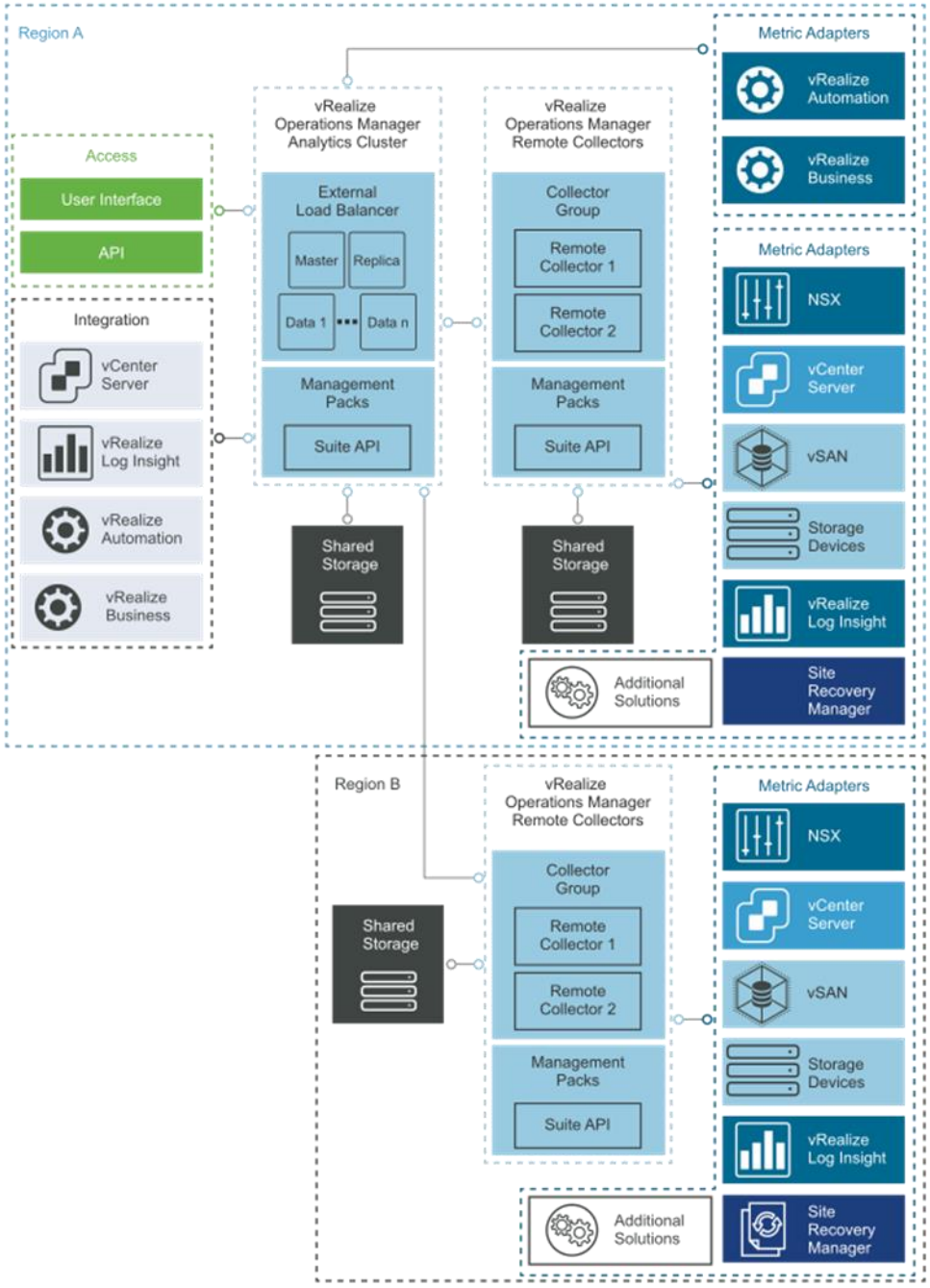
Logical Design

In a multiregion SDDC, deploy a vRealize Operations Manager configuration that consists of the following entities:

- A 3-node (medium-size) vRealize Operations Manager analytics cluster that is highly available. This topology provides high availability, scale-out capacity up to 16 nodes, and failover.
- A group of 2 remote collector nodes in each region. The remote collectors communicate directly with the data nodes in the vRealize Operations Manager analytics cluster. For load balancing and fault tolerance, 2 remote collectors are deployed in each region.

Each region contains its own pair of remote collectors that ease scalability by performing data collection from the applications that are not subject to failover and periodically sending collected data to the analytics cluster. You fail over the analytics cluster only because the analytics cluster is the construct that analyzes and stores monitoring data. This configuration supports failover of the analytics cluster by using Site Recovery Manager. In the event of a disaster, Site Recovery Manager migrates the analytics cluster nodes to the failover region.

Figure 61) Logical design of vRealize Operations Manager multiregion deployment.



Physical Design

The vRealize Operations Manager nodes run on the management cluster in each region of SDDC. For information about the types of clusters, see section 3.9, “Workload Domain Architecture.”

Data Sources

vRealize Operations Manager collects data from the following virtual infrastructure and cloud management components:

- Virtual infrastructure
  - Platform Services Controller instances
  - vCenter Server instances
  - ESXi hosts
  - NSX Manager instances
  - NSX Controller instances
  - NSX Edge instances
  - Shared storage
- vRealize Automation
  - vRealize Automation appliance
  - vRealize IaaS Web Server
  - vRealize IaaS Management Server
  - vRealize IaaS DEM
  - vRealize vSphere Proxy Agents
  - Microsoft SQL Server
- vRealize Business for Cloud
- vRealize Log Insight
- vRealize Operations Manager
- Site Recovery Manager

Node Configuration of vRealize Operations Manager

The analytics cluster of the vRealize Operations Manager deployment contains the nodes that analyze and store data from the monitored components. Deploy a configuration of the analytics cluster that satisfies the requirements for monitoring the number of virtual machines in the design objectives of this validated design.

Deploy a three-node vRealize Operations Manager analytics cluster in the cross-region application virtual network. The analytics cluster consists of one master node, one master replica node, and one data node to enable scale out and high availability.

Table 83) Node configuration of vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-001	Deploy initially vRealize Operations Manager as a cluster of three nodes: one master, one master replica, and one data node.	Provides the initial scale capacity required for monitoring up to 1,000 VMs and can be scaled up with additional data nodes.	• You must size all appliances identically, which increases the resources requirements in the SDDC. • You must manually install the additional data nodes according to the data-node scale guidelines.
SDDC-OPS-MON-002	Deploy two remote collector nodes per region.	Removes the metrics-collection load from the analytics cluster for applications that do not fail over between regions.	You must assign a collector group when configuring the monitoring of a solution.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-003	Apply vSphere DRS anti-affinity rules to the vRealize Operations Manager analytics cluster.	Using DRS prevents the vRealize Operations Manager analytics cluster nodes from running on the same ESXi host and risking the high availability of the cluster.	• You must perform additional configuration to set up an anti-affinity rule. • Of the four ESXi hosts in the management cluster, you can put only one in maintenance mode at a time.
SDDC-OPS-MON-004	Apply vSphere DRS anti-affinity rules to the vRealize Operations Manager remote collector group.	Using DRS prevents the vRealize Operations Manager remote collector nodes from running on the same ESXi host and risking the high availability of the cluster.	• You must perform additional configuration to set up an anti-affinity rule. • Of the four ESXi hosts in the management cluster, you can put only one in maintenance mode at a time.

Sizing Compute Resources for vRealize Operations Manager

You size compute resources for vRealize Operations Manager to accommodate the analytics operations that monitor the SDDC and the expected number of virtual machines contained in the SDDC.

Size the vRealize Operations Manager analytics cluster according to VMware Knowledge Base article 2093783. vRealize Operations Manager is also sized to accommodate the SDDC design by deploying a set of management packs. See “Management Packs” in vRealize Operations Manager.

The sizing of the vRealize Operations Manager instance is calculated using the options described in the rest of this section.

Table 84) vRealize Operations Manager sizing parameters.

Initial Setup	Scaled-Out Setup
4 vCenter Server appliances	4 vCenter Server appliances
4 NSX Manager instances	4 NSX Manager instances
6 NSX Controller instances	6 NSX Controller instances
50 ESXi hosts	100 ESXi hosts
1,000 virtual machines	10,000 virtual machines

Sizing Compute Resources for the Analytics Cluster Nodes

Deploying three medium-size virtual appliances supports the retention and monitoring of the expected number of objects and metrics for smaller environments of up to 1,000 virtual machines. As the environment expands, you should deploy more data notes to accommodate the larger number of objects and metrics.

Consider deploying additional vRealize Operations Manager data notes only if more ESXi hosts are added to the management pods. Doing so guarantees that the vSphere cluster has enough capacity to host these additional nodes without violating the vSphere DRS anti-affinity rules.

Table 85) Resources for a medium-size vRealize Operations Manager virtual appliance.

Attribute	Specification
Appliance size	Medium
vCPU	8
Memory	32GB
Single-node maximum objects	8,500
Single-node maximum collected metrics (*)	2,500,000
Multinode maximum objects per node (**)	6,250
Multinode maximum collected metrics per node (**)	1,875,000
Maximum number of end-point operations management agents per node	1,200
Maximum objects for a 16-node configuration	75,000
Maximum metrics for a 16-node configuration	19,000,000

(*) Metric numbers reflect the total number of metrics that are collected from all adapter instances in vRealize Operations Manager. To get this number, you can go to the Cluster Management page in vRealize Operations Manager and view the adapter instances of each node at the bottom of the page. You can view the number of metrics collected by each adapter instance. The estimations in the specification table represent the sum of these metrics.

Note: The number shown in the overall metrics on the Cluster Management page reflects the metrics that are collected from different data sources and the metrics that vRealize Operations Manager creates.

(**) This parameter represents a reduction in the maximum number of metrics to permit some head room.

Table 86) Compute size of the analytics cluster nodes for vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-005	Deploy each node in the analytics cluster as a medium-size appliance.	Provides the scale required to monitor the SDDC at full capacity. If you use a lower number of large-size vRealize Operations Manager nodes, you must increase the minimum host memory size to handle the increased performance that results from stretching NUMA node boundaries.	ESXi hosts in the management cluster must have physical CPUs with a minimum of eight cores per socket. In total, vRealize Operations Manager uses 24 vCPUs and 96GB of memory in the management cluster.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-006	Initially, deploy three medium-size nodes for the first 1,000 virtual machines in the shared edge and compute cluster.	Provides enough capacity for the metrics and objects generated by 100 hosts and 1,000 virtual machines. High availability is enabled in the analytics cluster and the metrics collection for the following components: • vCenter Server and Platform Services Controller • ESXi hosts • NSX for vSphere components • CMP components • vRealize Log Insight components	The first three medium-size nodes take more resources per 1,000 virtual machines because they must accommodate the requirements for high availability. Nodes that are deployed next can distribute this load out more evenly.
SDDC-OPS-MON-007	Add more medium-size nodes to the analytics cluster if the number of virtual machines in the SDDC exceeds 1,000.	• Provides the analytics cluster with enough capacity to meet the virtual machine object and metrics growth up to 10,000 virtual machines. • Provides the management cluster with enough physical capacity to take a host offline for maintenance or other reasons.	• The capacity of the physical ESXi hosts must be large enough to accommodate virtual machines that require 32GB of RAM without bridging NUMA node boundaries. • The management cluster must have enough ESXi hosts so that vRealize Operations Manager can run without violating vSphere DRS anti-affinity rules. • The number of nodes must not exceed the number of ESXi hosts in the management cluster minus one. For example, if the management cluster contains six ESXi hosts, you can deploy up to five vRealize Operations Manager nodes in the analytics cluster.

Sizing Compute Resources for the Remote Collector Nodes

Unlike the analytics cluster nodes, remote collector nodes only have the collector role. Deploying two remote collector nodes in each region does not increase the capacity for monitored objects.

Table 87) Size of a standard remote collector virtual appliance for vRealize Operations Manager.

Attribute	Specification
Appliance size	Remote Collector - Standard
vCPU	2
Memory	4 GB
Single-node maximum objects (*)	1,500
Single-node maximum collected metrics	600,000
Maximum number of end-point operations management agents per node	250

*The object limit for a remote collector is based on the VMware vCenter adapter.

Table 88) Compute size of the remote collector nodes of vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-008	Deploy the standard-size remote collector virtual appliances.	Enables metric collection for the expected number of objects in the SDDC when at full capacity.	You must provide four vCPUs and 8GB of memory in the management cluster in each region.

Sizing Storage in vRealize Operations Manager

You allocate storage capacity for analytics data collected from the management products and from the number of tenant virtual machines defined in the objectives of this SDDC design.

This design uses medium-size nodes for the analytics and remote collector clusters. To collect the required number of metrics, you must increase Disk 2 to a 1TB VMDK on each analytics cluster node.

Sizing Storage for the Analytics Cluster Nodes

The analytics cluster processes a large number of objects and metrics. As the environment expands, more data nodes need to be added to the analytics cluster. To plan the sizing requirements of your environment, see the vRealize Operations Manager sizing guidelines in VMware Knowledge Base article 2093783.

Table 89) Storage size of the analytics cluster of vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-009	Increase the size of Disk 2 to 1TB for each analytics cluster node.	Provides enough storage to meet the SDDC design objectives.	You must add the 1TB disk manually while the virtual machine for the analytics node is powered off.

Sizing Storage for the Remote Collector Nodes

Deploy the remote collector nodes with thin-provisioned disks. Because remote collectors do not perform analytics operations or store data, the default VMDK size is sufficient.

Table 90) Storage size of the remote collector nodes of vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-010	Do not provide additional storage for remote collectors.	Remote collectors do not perform analytics operations or store data on disk.	None

Networking Design of vRealize Operations Manager

You provide isolation and failover of the vRealize Operations Manager nodes by placing them in several network segments. This networking design also supports public access to the analytics cluster nodes.

For secure access, load balancing, and portability, deploy the vRealize Operations Manager analytics cluster in the shared cross-region application isolated network `Mgmt-xRegion01-VXLAN`. Then deploy the remote collector clusters in the shared local application isolated networks `Mgmt-RegionA01-VXLAN` and `Mgmt-RegionB01-VXLAN`.

Application Virtual Network Design for vRealize Operations Manager

The vRealize Operations Manager analytics cluster is installed in the cross-region shared application virtual network and the remote collector nodes are installed in their region-specific shared application virtual networks.

This networking design has the following features:

- The analytics nodes of vRealize Operations Manager are on the same network because they are failed over between regions. vRealize Automation also shares this network.
- All nodes have routed access to the vSphere management network through the NSX UDLR.
- Routing to the vSphere management network and other external networks is dynamic and is based on the BGP.

For more information about the networking configuration of the application virtual network, see section 10.6, “Virtualization Network Design” and section 10.7, “NSX Design.”

Table 91) Application virtual network for vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-011	Use the existing cross-region application virtual networks for the vRealize Operations Manager analytics cluster.	Supports disaster recovery by isolating the vRealize Operations Manager analytics cluster on the application virtual network `Mgmt-xRegion01-VXLAN`.	You must use an implementation in NSX to support this network configuration.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-012	Use the existing region-specific application virtual networks for vRealize Operations Manager remote collectors.	Provides local metrics collections for each region in the event of a cross-region network outage. Also colocalizes metrics collection to the SDDC applications in each region by using the virtual networks `Mgmt-RegionA01-VXLAN` and `Mgmt-RegionB01-VXLAN`.	You must use an implementation in NSX to support this network configuration.

IP Subnets for vRealize Operations Manager

You can allocate the following sample subnets for each cluster in the vRealize Operations Manager deployment.

Table 92) IP subnets in the application virtual network for vRealize Operations Manager.

vRealize Operations Manager Cluster Type	IP Subnet
Analytics cluster in Region A This cluster is also valid for Region B after failover.	192.168.11.0/24
Remote collectors in Region A	192.168.31.0/24
Remote collectors in Region B	192.168.32.0/24

Table 93) IP subnets for vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-013	Allocate separate subnets for each application virtual network.	Placing the remote collectors on their own subnet enables them to communicate with the analytics cluster and not be a part of the failover group.	None

DNS Names for vRealize Operations Manager

The FQDNs of the vRealize Operations Manager nodes follow these domain-name resolution rules:

- The IP addresses of the analytics cluster node and a load-balancer VIP are associated with names whose suffix is set to the root domain `rainpole.local`.

  From the public network, users access vRealize Operations Manager by using the VIP address. The traffic to vRealize Operation Manager is handled by a NSX Edge services gateway providing the load-balancer function.
- Name resolution for the IP addresses of the remote collector group nodes uses a region-specific suffix; for example, `sfo01.rainpole.local` or `lax01.rainpole.local`.
- The IP addresses of the remote collector group nodes are associated with names whose suffix is set to the region-specific domain; for example, `sfo01.rainpole.local` or `lax01.rainpole.local`.

Table 94) FQDNs for the vRealize Operations Manager nodes.

vRealize Operations Manager DNS Name	Node Type	Region
vrops01svr01.rainpole.local	Virtual IP of the analytics cluster	Region A (failover to Region B)
vrops01svr01a.rainpole.local	Master node in the analytics cluster	Region A (failover to Region B)
vrops01svr01b.rainpole.local	Master replica node in the analytics cluster	Region A (failover to Region B)
vrops01svr01c.rainpole.local	First data node in the analytics cluster	Region A (failover to Region B)
vrops01svr01x.rainpole.local	Additional data nodes in the analytics cluster	Region A (failover to Region B)
sfo01vropsc01a.sfo01.rainpole.local	First remote collector node	Region A
sfo01vropsc01b.sfo01.rainpole.local	Second remote collector node	Region A
lax01vropsc01a.lax01.rainpole.local	First remote collector node	Region B
lax01vropsc01b.lax01.rainpole.local	Second remote collector node	Region B

Table 95) DNS names for vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-014	Configure forward and reverse DNS records for all vRealize Operations Manager nodes and VIP address deployed.	All nodes are accessible by using the FQDN instead of by using only IP addresses.	You must manually provide DNS records for all vRealize Operations Manager nodes and the VIP address.

Networking for Failover and Load Balancing

By default, vRealize Operations Manager does not provide a solution for load-balanced UI user sessions across nodes in the cluster. You associate vRealize Operations Manager with the shared load balancer in the region.

The lack of load balancing for user sessions results in the following limitations:

- Users must know the URL of each node to access the UI. As a result, a single node might be overloaded if all users access it at the same time.
- Each node supports up to four simultaneous user sessions.
- Taking a node offline for maintenance might cause an outage. Users cannot access the UI of the node when the node is offline.

To avoid such problems, place the analytics cluster behind an NSX load balancer located in the `Mgmt-xRegion01-VXLAN` application virtual network. This load balancer is configured to allow up to four connections per node. The load balancer must distribute the load evenly to all cluster nodes. In addition, configure the load balancer to redirect service requests from the UI on port 80 to port 443.

Load balancing for the remote collector nodes is not required.

Table 96) Networking failover and load balancing for vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-015	Use an NSX Edge services gateway as a load balancer for the vRealize Operation Manager analytics cluster located in the `Mgmt-xRegion01-VXLAN` application virtual network.	Enables balanced access of tenants and users to the analytics services with the load being spread evenly across the cluster.	You must manually configure the NSX Edge devices to provide load-balancing services.
SDDC-OPS-MON-016	Do not use a load balancer for the remote collector nodes.	• Remote collector nodes must directly access the systems that they are monitoring. • Remote collector nodes do not require access to and from the public network.	None

Information Security and Access Control in vRealize Operations Manager

You can protect the vRealize Operations Manager deployment by providing centralized role-based authentication and secure communication with the other components in the SDDC. To do so, you can dedicate a set of service accounts to the communication between vRealize Operations Manager and the management solutions in the data center.

Authentication and Authorization

You can allow users to authenticate in vRealize Operations Manager in the following ways:

- **Import users or user groups from an LDAP database.** Users can use their LDAP credentials to log in to vRealize Operations Manager.
- **Use vCenter Server user accounts.** After a vCenter Server instance is registered with vRealize Operations Manager, the following vCenter Server users can log in to vRealize Operations Manager:
  - Users who have administration access in vCenter Server.
  - Users who have one of the vRealize Operations Manager privileges, such as PowerUser, assigned to the account that appears at the root level in vCenter Server.
- **Create local user accounts in vRealize Operations Manager.** vRealize Operations Manager performs local authentication by using the account information stored in its global database.

Table 97) Authorization and authentication management for vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-017	Use Active Directory authentication.	• Provides access to vRealize Operations Manager by using standard Active Directory accounts. • Authentication is available even if vCenter Server becomes unavailable.	You must manually configure Active Directory authentication.
SDDC-OPS-MON-018	Configure a service account (`svc-vrops-vsphere`) in vCenter Server for application-to-application communication from vRealize Operations Manager with vSphere.	Provides the following access control features: • The adapters in vRealize Operations Manager access vSphere with the minimum set of permissions that are required to collect metrics about vSphere inventory objects. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-019	Configure a service account (`svc-vrops-nsx`) in vCenter Server for application-to-application communication from vRealize Operations Manager with NSX for vSphere.	Provides the following access control features: • The adapters in vRealize Operations Manager access NSX for vSphere with the minimum set of permissions that are required for metrics collection and topology mapping. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.
SDDC-OPS-MON-020	Configure a service account (`svc-vrops-mpsd`) in vCenter Server for application-to-application communication from the Storage Devices Adapters in vRealize Operations Manager with vSphere.	Provides the following access control features: • The adapters in vRealize Operations Manager access vSphere with the minimum set of permissions that are required to collect metrics about vSphere inventory objects. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-022	Configure a service account (`svc-vrops-srm`) in vCenter Server for application-to-application communication from the Site Recovery Manager Adapters in vRealize Operations Manager with vSphere and Site Recovery Manager.	Provides the following access control features: • The adapters in vRealize Operations Manager access vSphere and Site Recovery Manager with the minimum set of permissions that are required to collect metrics. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.
SDDC-OPS-MON-023	Use global permissions when you create the `svc-vrops-vsphere`, `svc-vrops-nsx`, `svc-vrops-mpsd`, and `svc-vrops-srm` service accounts in vCenter Server.	• Simplifies and standardizes the deployment of the service accounts across all vCenter Server instances in the same vSphere domain. • Provides a consistent authorization layer.	All vCenter Server instances must be in the same vSphere domain.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-024	Configure a service account (`svc-vrops-vra`) in vRealize Automation for application-to-application communication from the vRealize Automation Adapter in vRealize Operations Manager with vRealize Automation.	Provides the following access control features: • The adapter in vRealize Operations Manager accesses vRealize Automation with the minimum set of permissions that are required for collecting metrics about provisioned virtual machines and capacity management. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	• You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability. • If you add more tenants to vRealize Automation, you must maintain the service account permissions to guarantee that metric uptake in vRealize Operations Manager is not compromised.
SDDC-OPS-MON-025	Configure a local service account (`svc-vrops-nsx`) in each NSX instance for application-to-application communication from the NSX-vSphere Adapters in vRealize Operations Manager with NSX.	Provides the following access control features: • The adapters in vRealize Operations Manager access NSX for vSphere with the minimum set of permissions that are required for metrics collection and topology mapping. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.

Encryption

Access to all vRealize Operations Manager Web interfaces requires an SSL connection. By default, vRealize Operations Manager uses a self-signed certificate.

Monitoring and Alerting in vRealize Operations Manager

You use vRealize Operations Manager to monitor the state of the SDDC management components in the SDDC by using dashboards. You can use the self-monitoring capability of vRealize Operations Manager to receive alerts about issues that are related to its operational state.

vRealize Operations Manager displays the following administrative alerts:

- **System alert.** A component of the vRealize Operations Manager application has failed.
- **Environment alert.** vRealize Operations Manager has stopped receiving data from one or more resources. Such an alert might indicate a problem with system resources or network infrastructure.
- **Log Insight log event.** The infrastructure on which vRealize Operations Manager is running has low-level issues. You can also use the log events for root-cause analysis.
- **Custom dashboard.** vRealize Operations Manager can show super metrics for data center monitoring, capacity trends, and a single-pane-of-glass overview.

Table 98) Monitoring vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-027	Configure vRealize Operations Manager for SMTP outbound alerts.	Enables administrators and operators to receive alerts from vRealize Operations Manager by email.	You must provide vRealize Operations Manager with access to an external SMTP server.
SDDC-OPS-MON-028	Configure vRealize Operations Manager custom dashboards.	Provides extended SDDC monitoring, capacity trends, and single-pane-of-glass overview.	You must manually configure the dashboards.

Management Packs in vRealize Operations Manager

The SDDC contains VMware products for network, storage, and cloud management. You can monitor and perform diagnostics on all of them in vRealize Operations Manager by using management packs.

Table 99) vRealize Operations Manager management packs in this VVD.

Management Pack	Installed by Default
Management Pack for VMware vCenter Server	X
Management Pack for NSX for vSphere	
Management Pack for Storage Devices	
Management Pack for vRealize Log Insight	X
Management Pack for vRealize Automation	X
Management Pack for vRealize Business for Cloud	X
Management Pack for Site Recovery Manager	

Table 100) Management packs in vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-029	Install the following management packs: • Management Pack for NSX for vSphere • Management Pack for Storage Devices • Management Pack for Site Recovery Manager	Provides additional granular monitoring for all virtual infrastructure and cloud management applications. You do not have to install the following management packs, because they are installed by default in vRealize Operations Manager: • Management Pack for VMware vCenter Server • Management Pack for vRealize Log Insight • Management Pack for vRealize Automation • Management Pack for vRealize Business for Cloud	You must install and configure each nondefault management pack manually.
SDDC-OPS-MON-030	Configure the following management pack adapter instances to the default collector group: • vRealize Automation • vRealize Business for Cloud	Provides monitoring of components during a failover.	The load on the analytics cluster, although minimal, increases.
SDDC-OPS-MON-031	Configure the following management pack adapter instances to use the remote collector group: • vCenter Server • NSX for vSphere • Network Devices • Storage Devices • vRealize Log Insight • Site Recovery Manager	Offloads data collection for local management components from the analytics cluster.	None

Disaster Recovery of vRealize Operations Manager

To preserve monitoring functionality when a disaster occurs, vRealize Operations Manager supports failing over a subset of the components between regions. Disaster recovery covers only the analytics cluster components, including the master, replica, and data nodes. The region-specific remote collector nodes remain in the affected region.

When a disaster occurs, use Site Recovery Manager and vSphere Replication for orchestrated recovery of the analytics cluster. You do not recover the remote collector nodes. Remote collector pairs collect

data only from local components, such as vCenter Server and NSX Manager, which are also not recovered during such an event. See “Recovery Plan for Site Recovery Manager and vSphere Replication” in section 13.2 for details.

11.2 vRealize Log Insight Design

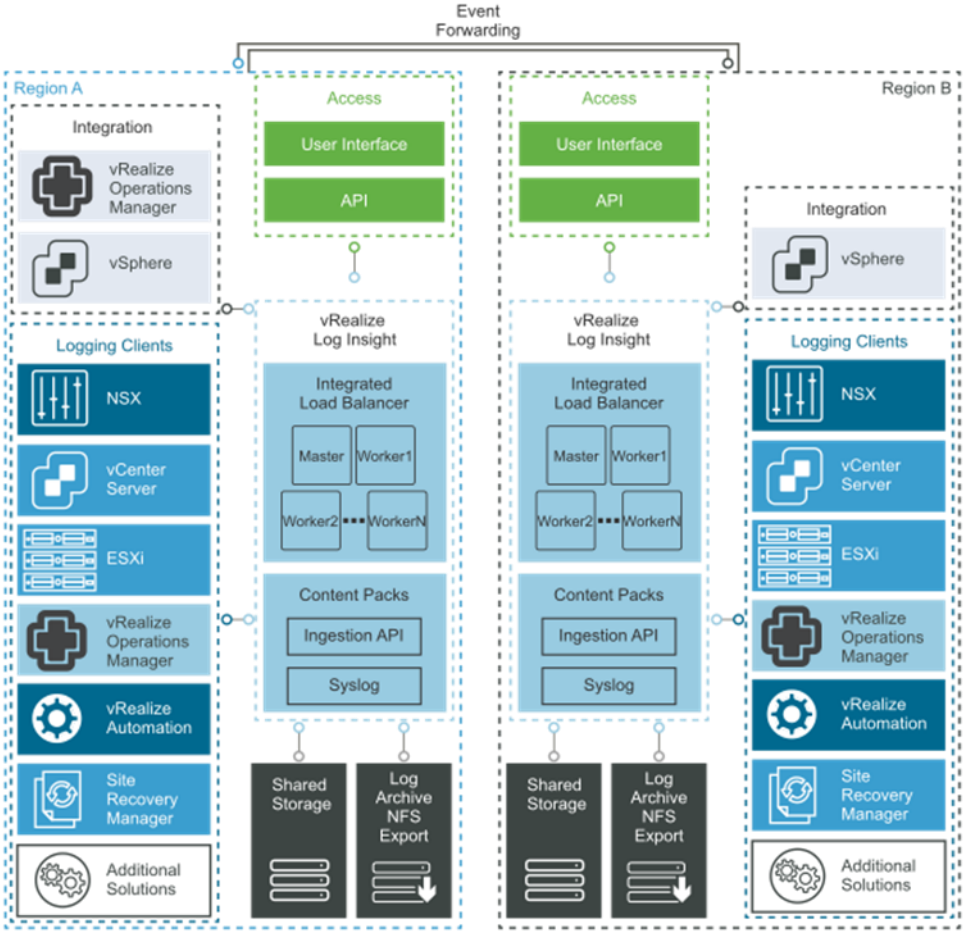
The vRealize Log Insight design enables real-time logging for all the management components of the SDDC in a dual-region setup.

Logical Design and Data Sources of vRealize Log Insight

vRealize Log Insight collects log events from all management components in both regions of the SDDC.

Logical Design

Figure 62) Logical design of vRealize Log Insight.



Sources of Log Data

vRealize Log Insight collects logs to help monitor information about the SDDC from a central location. It collects log events from the following virtual infrastructure and cloud management components:

- Management cluster
  - Platform Services Controller
  - vCenter Server

  - Site Recovery Manager
  - ESXi hosts
- Shared edge and compute cluster
  - Platform Services Controller
  - vCenter Server
  - ESXi hosts
- NSX for vSphere for the management cluster and for the shared compute and edge cluster
  - NSX Manager instances
  - NSX Controller instances
  - NSX Edge services gateway instances
  - NSX distributed logical router instances
  - NSX UDLR instances
  - NSX distributed firewall ESXi kernel module
- vRealize Automation
  - vRealize Automation Appliance
  - vRealize IaaS Web Server
  - vRealize IaaS Management Server
  - vRealize IaaS DEM
  - vRealize Agent Servers
  - vRealize Orchestrator (embedded in the vRealize Automation Appliance)
  - Microsoft SQL Server
- vRealize Business
  - vRealize Business server
  - vRealize Business data collectors
- vRealize Operations Manager
  - Analytics cluster nodes
  - Remote
- vRealize Log Insight instance in the other region as a result of event forwarding

Node Configuration of vRealize Log Insight

The vRealize Log Insight cluster consists of one master node and two worker nodes behind a load balancer.

You enable the integrated load balancer (ILB) on the 3-node cluster so that all log sources can address the cluster by its ILB. By using the ILB, you do not need to reconfigure all log sources with a new destination address in a future scale out. Using the ILB also ensures that vRealize Log Insight accepts all incoming ingestion traffic.

vRealize Log Insight users, using the web user interface or the API, and clients, ingesting logs through syslog or the Ingestion API, connect to vRealize Log Insight by using the ILB address.

A vRealize Log Insight cluster can scale out to 12 nodes, or 1 master and 11 worker nodes.

Table 101) Node configuration for vRealize Log Insight design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-001	In each region, deploy vRealize Log Insight in a cluster configuration of three nodes with an integrated load balancer: one master and two worker nodes.	• Provides high availability. • Using the ILB prevents a single point of failure. • Using the ILB simplifies Log Insight deployment and subsequent integration. • Using the ILB simplifies the Log Insight scale-out operations, reducing the need to reconfigure existing logging sources	• You must deploy a minimum of three medium nodes. • You must size each node identically. • If the capacity requirements for your vRealize Log Insight cluster expand, identical capacity must be added to each node.
SDDC-OPS-LOG-002	Apply vSphere DRS anti-affinity rules to the vRealize Log Insight cluster components.	Prevents the vRealize Log Insight nodes from running on the same ESXi host and risking the high availability of the cluster.	• You must perform additional configuration to set up anti-affinity rules. • You can put only a single ESXi host in maintenance mode at a time in the management cluster of four ESXi hosts.

Sizing Compute and Storage Resources for vRealize Log Insight

To accommodate all log data from the products in the SDDC, you must size the compute resources and storage for the Log Insight nodes correctly.

By default, the vRealize Log Insight virtual appliance uses the predefined values for small configurations, which have four vCPUs, 8GB of virtual memory, and 530.5GB of disk space provisioned. vRealize Log Insight uses 100GB of the disk space to store raw data, index, metadata, and other information.

Sizing Nodes

Select a size for the vRealize Log Insight nodes to collect and store log data from the SDDC management components and tenant workloads according to the objectives of this design.

Table 102) Compute resources for a medium-size vRealize Log Insight node.

Attribute	Specification
Appliance size	Medium
Number of CPUs	8
Memory	16GB
Disk capacity	530.5GB (490GB for event storage)
IOPS	1,000 IOPS

Attribute	Specification
Amount of processed log data when using log ingestion	75GB/day of processing per node
Number of processed log messages	5,000 event/second of processing per node
Environment	Up to 250 syslog connections per node

Sizing Storage

Sizing is based on IT organization requirements, but this design provides calculations based on a single-region implementation, and is implemented on a per-region basis. This sizing is calculated according to the following node configuration per region.

Table 103) Management systems whose log data is stored by vRealize Log Insight.

Category	Logging Sources	Quantity
Management cluster	Platform Services Controller	1
	vCenter Server	1
	Site Recovery Manager	1
	ESXi Hosts	4
Shared edge and compute cluster	Platform Services Controller	1
	vCenter Server	1
	ESXi Hosts	64
NSX for vSphere for the management cluster	NSX Manager	1
	NSX Controller instances	3
	NSX Edge services gateway instances: • Two ESGs for north-south routing • UDLR • Load balancer for vRealize Automation and vRealize Operations Manager • Load balancer for Platform Services Controllers	5
NSX for vSphere for the shared edge and compute cluster	NSX Manager	1
	NSX Controller instances	3
	NSX Edge services gateway instances: • UDLR • Distributed logical router • Two ESGs for north-south routing	4

Category	Logging Sources	Quantity
vRealize Automation	vRealize Automation Appliance with embedded vRealize Orchestrator	2
	vRealize IaaS Web Server	2
	vRealize IaaS Manager Server	2
	vRealize IaaS DEM	2
	vRealize Agent Servers	2
	Microsoft SQL Server	1
vRealize Business for Cloud	vRealize Business server appliance	1
	vRealize Business data collector	2
vRealize Operations Manager	Analytics nodes	3
	Remote collector nodes	2
Cross-region event forwarding		Total * 2

These components aggregate to approximately 109 syslog and vRealize Log Insight Agent sources per region, or 220 sources with a cross-region configuration.

Assuming that you want to retain 7 days of data, apply the following calculation:

vRealize Log Insight receives approximately 150MB to 190MB of log data per day per source, as follows:

- The rate of 150MB of logs per day is valid for Linux where 170bytes per message is the default message size.
- The rate of 190MB of logs per day is valid for Windows where 220bytes per message is the default message size.

```
170 bytes per message * 10 messages per second * 86400 seconds per day = 150 MB of logs per-day
per-source (Linux)
220 bytes per message * 10 messages per second * 86400 seconds per day = 190 MB of logs per-day
per-source (Windows)
```

To simplify calculation in this validated design, all calculations have been performed using the large 220byte size, which results in 190MB of log data expected each day for each source.

For 220 logging sources at a basal rate of approximately 190MB of logs that are ingested each day for each source over 7 days, you need the following storage space:

```
220 sources * 190 MB of logs per-day per-source * 1e-9 GB per byte ≈ 42 GB disk space per-day
```

To size the appliance for 7 days of log retention based on the amount of data stored in a day, use the following calculation:

```
(42 GB * 7 days) / 3 appliances ≈ 100 GB log data per vRealize Log Insight node

100 GB * 1.7 indexing overhead ≈ 170 GB log data per vRealize Log Insight Node
```

Based on this example, the storage space that is allocated per medium-size vRealize Log Insight virtual appliance is enough to monitor the SDDC.

When you need to increase the Log Insight capacity, consider the following approaches:

- If you need to maintain a log data retention for more than 7 days in your SDDC, consider adding more storage per node by adding a new virtual hard disk. vRealize Log Insight supports virtual hard disks of up to 2TB. If you need to add more than 2TB to a virtual appliance, add another virtual hard disk.

  When you add storage to increase the retention period, extend the storage for all virtual appliances. To increase the storage, only add new virtual hard disks. Do not extend existing retention virtual disks. After they are provisioned, to avoid data loss, do not reduce the size of or remove virtual disks.
- If you need to monitor more components by using log ingestion and exceed the number of syslog connections or ingestion limits defined in this design, you can:
  - Increase the size of the vRealize Log Insight node to a medium or large deployment size, as defined in the vRealize Log Insight documentation.
  - Deploy more vRealize Log Insight virtual appliances to scale your environment out. vRealize Log Insight can scale up to 12 nodes in an HA cluster.

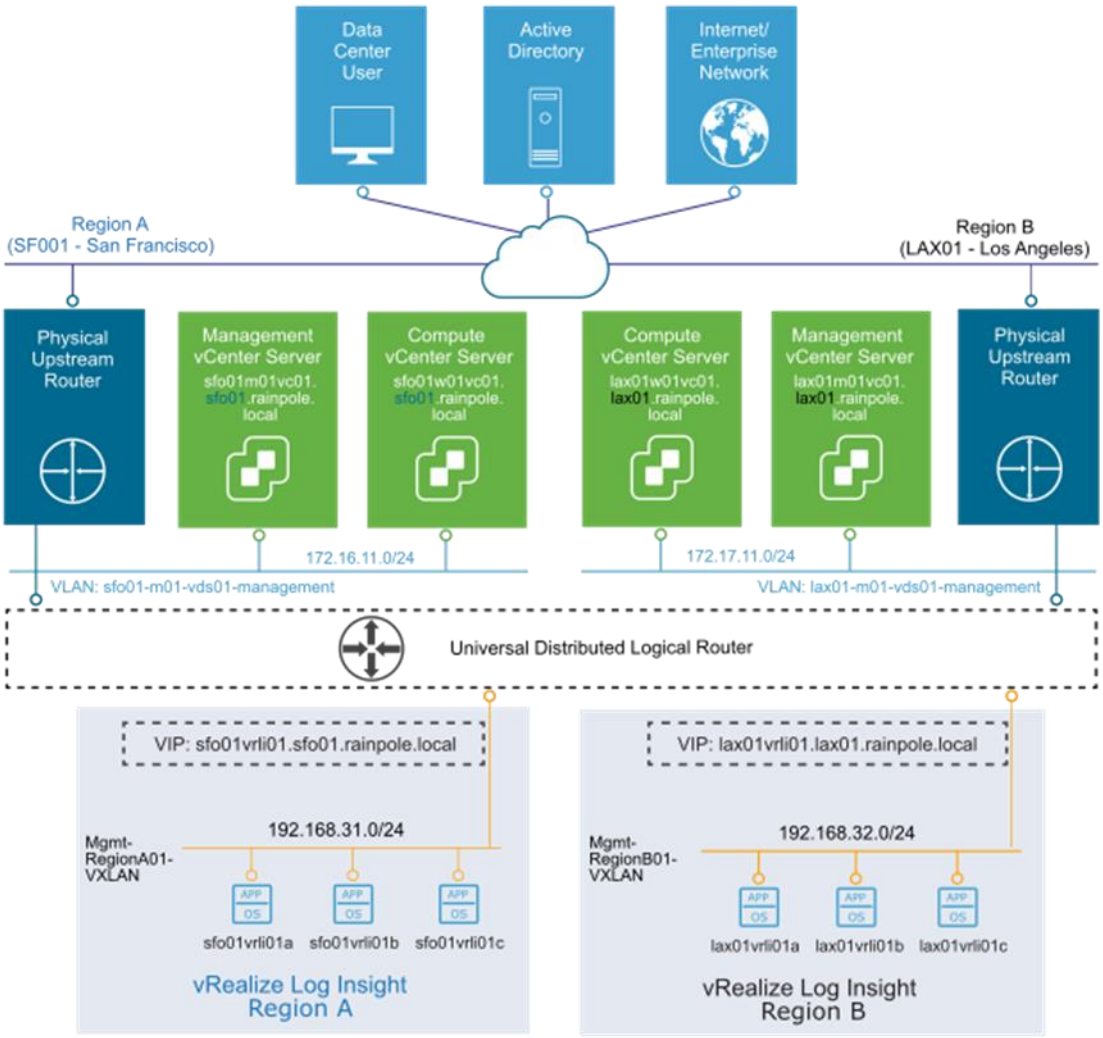
Table 104) Compute resources for the vRealize Log Insight nodes design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-003	Deploy vRealize Log Insight nodes of medium size.	Accommodates the number of expected syslog and vRealize Log Insight Agent connections from the following sources: • Management vCenter Server, Compute vCenter Server, and a connected Platform Services Controller pair • Management ESXi hosts, and shared edge and compute ESXi hosts • Management Site Recovery Manager components • Management and compute components of NSX for vSphere • vRealize Automation components • vRealize Business components • vRealize Operations Manager components • Cross-vRealize Log Insight cluster event forwarding These sources generate approximately 220 syslog and vRealize Log Insight Agent sources. Using a medium-size appliance provides adequate storage space for the vRealize Log Insight cluster for 7 days of data retention.	If you configure Log Insight to monitor additional syslog sources, you must increase the size of the nodes.

Networking Design of vRealize Log Insight

In both regions, the vRealize Log Insight instances are connected to the region-specific management VXLANs `Mgmt-RegionA01-VXLAN` and `Mgmt-RegionB01-VXLAN`. Each vRealize Log Insight instance is deployed in the shared management application isolated network.

Figure 63) Networking design for vRealize Log Insight deployment.



Application Network Design

This networking design has the following features:

- All nodes have routed access to the vSphere management network through the UDLR for the management cluster in the home region.
- Routing to the vSphere management network and the external network is dynamic and is based on the BGP.

For more information about the networking configuration of the application virtual networks for vRealize Log Insight, see the sections “Application Virtual Network” and “Virtual Network Design Example” in section 10.7

Table 105) Networking for vRealize Log Insight design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-004	Deploy vRealize Log Insight on the region-specific application virtual networks.	• Provides centralized access to log data in each region if a cross-region network outage occurs. • Colocates log collection to the region-local SDDC applications using the region-specific application virtual networks. • Provides a consistent deployment model for management applications.	• Interruption in the cross-region network can interfere with event forwarding between the vRealize Log Insight clusters and can cause gaps in log data. • You must use NSX to support this network configuration.

IP Subnets for vRealize Log Insight

You can allocate the following sample subnets to the vRealize Log Insight deployment.

Table 106) IP subnets in the application-isolated networks of vRealize Log Insight.

vRealize Log Insight Cluster	IP Subnet
Region A	192.168.31.0/24
Region B	192.168.32.0/24

DNS Names for vRealize Log Insight

vRealize Log Insight node name resolution, including the load balancer VIPs, uses a region-specific suffix, such as `sfo01.rainpole.local` or `lax01.rainpole.local`. The Log Insight components in both regions have the following node names.

Table 107) DNS names of the vRealize Log Insight nodes.

DNS Name	Role	Region
sfo01vrli01.sfo01.rainpole.local	Log Insight ILB VIP	Region A
sfo01vrli01a.sfo01.rainpole.local	Master node	Region A
sfo01vrli01b.sfo01.rainpole.local	Worker node	Region A
sfo01vrli01c.sfo01.rainpole.local	Worker node	Region A
sfo01vrli01x.sfo01.rainpole.local	Additional worker nodes (not deployed)	Region A
lax01vrli01.lax01.rainpole.local	Log Insight ILB VIP	Region B
lax01vrli01a.lax01.rainpole.local	Master node	Region B
lax01vrli01b.lax01.rainpole.local	Worker node	Region B
lax01vrli01c.lax01.rainpole.local	Worker node	Region B

DNS Name	Role	Region
lax01vrli01x.lax01.rainpole.local	Additional worker nodes (not deployed)	Region B

Table 108) DNS Names for vRealize Log Insight design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-005	Configure forward and reverse DNS records for all vRealize Log Insight nodes and VIPs.	All nodes are accessible by using FQDNs instead of by using only IP addresses.	You must manually provide a DNS record for each node and VIP.
SDDC-OPS-LOG-006	For all applications that fail over between regions (such as vRealize Automation and vRealize Operations Manager), use the FQDN of the vRealize Log Insight Region A VIP when you configure logging.	Support logging when not all management applications are failed over to Region B. For example, only one application is moved to Region B.	If vRealize Automation and vRealize Operations Manager are failed over to Region B and the vRealize Log Insight cluster is no longer available in Region A, update the A record on the child DNS server to point to the vRealize Log Insight cluster in Region B.

Retention and Archiving in vRealize Log Insight

Configure archive and retention parameters of vRealize Log Insight according to the company policy for compliance and governance. Each vRealize Log Insight virtual appliance has three default virtual disks and can use more virtual disks for storage.

Table 109) Virtual disk configuration in the vRealize Log Insight Virtual Appliance.

Hard Disk	Size	Usage
Hard disk 1	20GB	Root file system
Hard disk 2	510GB for medium-size deployment	Contains two partitions: • `/storage/var`. System logs • `/storage/core`. Storage for collected logs
Hard disk 3	512MB	First boot only

Calculate the storage space that is available for log data by using the following equation:

```
/storage/core = hard disk 2 space - system logs space on hard disk 2
```

Based on the size of the default disk, the storage core is equal to 490GB. If `/storage/core` is 490GB, vRealize Log Insight can use 475GB for retaining accessible logging data.

```
/storage/core = 510 GB - 20 GB = 490 GB
Retention = /storage/core – 3% * /storage/core
Retention = 490 GB - 3% * 490 ≈ 475 GB disk space per vRLI appliance
```

Calculate retention time by using the following equations:

```
GB per vRLI Appliance per day = (Amount in GB of disk space used per day / Number of vRLI
appliances) * 1.7 indexing
```

```
Retention in days = 475 GB disk space per vRLI appliance / GB per vRLI Appliance per day

(42 GB of logging data ingested per day / 3 vRLI appliances) * 1.7 indexing ≈ 24 GB per vRLI
Appliance per day
475 GB disk space per vRLI appliance / 24 GB per vRLI Appliance per Day ≈ 20 days of retention
```

Configure a retention period of 7 days for the medium-size vRealize Log Insight appliance.

Table 110. Retention period for vRealize Log Insight design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-007	Configure vRealize Log Insight to retain data for 7 days.	Accommodates logs from 220 syslog sources and vRealize Log Insight agents, as per the SDDC design.	None

Archiving

Configure vRealize Log Insight to archive log data only if you must retain logs for an extended period for compliance, auditability, or a customer-specific reason.

Table 111) vRealize Log Insight archiving.

Attribute of Log Archiving	Description
Archiving period	vRealize Log Insight archives log messages as soon as possible. At the same time, the logs are retained on the virtual appliance until the free local space is almost filled. Data exists on both the vRealize Log Insight appliance and the archive location for most of the retention period. The archiving period must be longer than the retention period.
Archive location	The archive location must be on NFS version 3 shared storage. The archive location must be available and must have enough capacity to accommodate the archives.

Apply an archive policy of 90 days for the medium-size vRealize Log Insight appliance. The vRealize Log Insight clusters each use approximately 400GB of shared storage, calculated with the following equations:

```
(Average Storage Utilization (GB) per Day sources * Days of Retention) / Number of vRLI
appliances ≈ Recommended Storage in GB
((((Recommended Storage Per Node * Number of vRLI appliances) / Days of Retention) * Days of
Archiving) * 10%) ≈ Archiving to NFS in GB

(42 GB * 7 Days) / 3 vRLI appliances = 98 GB ≈ 100 GB of Recommended Storage (rounded up)
((((100 GB * 3 vRLI appliances) / 7 Days of Retention) * 90 Days of Archiving) * 10%) = 386 GB ≈
400 GB of NFS
```

These sizes might change depending on the business compliance regulations of your organization.

Table 112) Log archive policy for vRealize Log Insight design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-008	Provide 400GB of NFS version 3 shared storage to each vRealize Log Insight cluster.	Accommodates log archiving from 220 logging sources for 90 days.	• You must manually maintain the vRealize Log Insight archive BLOBs stored on the NFS store, selectively cleaning the datastore as more space is required. • If you configure vRealize Log Insight to monitor more logging sources or add more vRealize Log Insight workers, you must increase the size of the NFS shared storage. • You must enforce the archive policy directly on the shared storage. • If the NFS mount does not have enough free space or is unavailable for a period greater than the retention period of the virtual appliance, vRealize Log Insight stops ingesting new data until the NFS mount has enough free space or becomes available, or archiving is disabled.

Alerting in vRealize Log Insight

vRealize Log Insight supports alerts that trigger notifications about its health and about the health of monitored solutions.

Alert Types

The following types of alerts exist in vRealize Log Insight:

- **System alerts.** vRealize Log Insight generates notifications when an important system event occurs. For example, if disk space is almost exhausted, vRealize Log Insight must start deleting or archiving old log files.
- **Content pack alerts.** Content packs contain default alerts that can be configured to send notifications. These alerts are specific to the content pack and are disabled by default.
- **User-defined alerts.** Administrators and users can define their own alerts based on data ingested by vRealize Log Insight, which handles alerts in two ways:
  - Sending an e-mail over SMTP.
  - Sending an alert to vRealize Operations Manager.

SMTP Notification

Enable e-mail notification for alerts in vRealize Log Insight.

Table 113) SMTP alert notification for vRealize Log Insight design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-009	Enable alerting over SMTP.	Enables administrators and operators to receive alerts by email from vRealize Log Insight.	Requires access to an external SMTP server.

Integration of vRealize Log Insight with vRealize Operations Manager

vRealize Log Insight supports integration with vRealize Operations Manager to provide a central location for monitoring and diagnostics.

Use the following integration points, which you can enable separately:

- **Notification events.** Forward notification events from vRealize Log Insight to vRealize Operations Manager.
- **Launch in context.** Launch vRealize Log Insight from the vRealize Operation Manager user interface.
- **Embedded vRealize Log Insight.** Access the integrated vRealize Log Insight user interface directly in the vRealize Operations Manager user interface.

Table 114) Integration of vRealize Log Insight with vRealize Operations Manager design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-010	Forward alerts to vRealize Operations Manager.	Provides monitoring and alerting information that is pushed from vRealize Log Insight to vRealize Operations Manager for centralized administration.	None
SDDC-OPS-LOG-011	Support launch in context with vRealize Operation Manager.	Provides access to vRealize Log Insight for context-based monitoring of an object in vRealize Operations Manager.	You can register only one vRealize Log Insight cluster with vRealize Operations Manager for launch in context at a time.
SDDC-OPS-LOG-012	Enable embedded vRealize Log Insight user interface in vRealize Operations Manager.	Provides central access to the vRealize Log Insight user interface for improved context-based monitoring on an object in vRealize Operations Manager.	You can register only one vRealize Log Insight cluster with vRealize Operations Manager at a time.

Information Security and Access Control in vRealize Log Insight

Protect the vRealize Log Insight deployment by providing centralized role-based authentication and secure communication with the other components in the SDDC.

Authentication

Enable role-based access control in vRealize Log Insight by using the existing `rainpole.local` Active Directory domain.

Table 115) Authorization and authentication management for vRealize Log Insight design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-013	Use Active Directory for authentication.	Provides fine-grained role and privilege-based access for administrator and operator roles.	You must provide access to Active Directory from all Log Insight nodes.
SDDC-OPS-LOG-014	Configure a service account (`svc-vrli-vsphere`) on vCenter Server for application-to-application communication from vRealize Log Insight with vSphere.	Provides the following access control features: • vRealize Log Insight accesses vSphere with the minimum set of permissions that are required to collect vCenter Server events, tasks, and alarms and to configure ESXi hosts for syslog forwarding. • If there is a compromised account, the accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.
SDDC-OPS-LOG-015	Use global permissions when you create the `svc-vrli-vsphere` service account in vCenter Server.	• Simplifies and standardizes the deployment of the service account across all vCenter Servers in the same vSphere domain. • Provides a consistent authorization layer.	All vCenter Server instances must be in the same vSphere domain.
SDDC-OPS-LOG-016	Configure a service account (`svc-vrli-vrops`) on vRealize Operations Manager for application-to-application communication from vRealize Log Insight for a two-way launch in context.	Provides the following access control features: • vRealize Log Insight and vRealize Operations Manager access each other with the minimum set of required permissions. • If there is a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.

Encryption

Replace default self-signed certificates with a CA-signed certificate to provide secure access to the vRealize Log Insight web user interface.

Table 116) CA-signed certificates for vRealize Log Insight design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-017	Replace the default self-signed certificates with a CA-signed certificate.	Configuring a CA-signed certificate makes sure that all communication to the externally facing web UI is encrypted.	The administrator must have access to a public key infrastructure (PKI) to acquire certificates.

Collecting Logs in vRealize Log Insight

As a part of vRealize Log Insight configuration, you configure syslog and vRealize Log Insight agents.

Client applications can send logs to vRealize Log Insight in one of the following ways:

- Directly to vRealize Log Insight by using the syslog TCP, syslog TCP over TLS/SSL, or syslog UDP protocol
- By using a vRealize Log Insight Agent
- By using vRealize Log Insight to directly query the vSphere Web Server APIs
- By using a vRealize Log Insight user interface

Table 117) Direct log communication to vRealize Log Insight design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-018	Configure syslog sources and vRealize Log Insight Agents to send log data directly to the VIP address of the vRealize Log Insight ILB.	• Allows future scale out without reconfiguring all log sources with a new destination address. • Simplifies the configuration of log sources in the SDDC.	• You must configure the ILB on the vRealize Log Insight cluster. • You must configure logging sources to forward data to the vRealize Log Insight VIP.
SDDC-OPS-LOG-019	Communicate with the vRealize Log Insight Agents by using the default ingestion API (`cfapi`), default disk buffer of 200MB, and nondefault No SSL.	• Supports multiline message transmissions from logs. • Provides ability to add metadata to events generated from system. • Provides client-side compression, buffering, and throttling capabilities, with minimal to no message loss during intermittent connection issues. • Provides server-side administration, metric collection, and configurations management of each deployed agent. • Supports disaster recovery of components in the SDDC.	• Transmission traffic is not secure. • Agent presence increases the overall resources used on the system.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-020	Deploy and configure the vRealize Log Insight agent for the vRealize Automation Windows servers.	• Windows does not natively support syslog. • vRealize Automation requires the use of agents to collect all vRealize Automation logs.	You must manually install and configure the agents on several nodes.
SDDC-OPS-LOG-021	Configure the vRealize Log Insight agent on the vRealize Automation appliance.	Simplifies configuration of log sources in the SDDC that are prepackaged with the vRealize Log Insight agent.	You must configure the vRealize Log Insight agent to forward logs to the vRealize Log Insight VIP.
SDDC-OPS-LOG-022	Configure the vRealize Log Insight agent for the vRealize Business appliances, including: • Server appliance • Data collectors	Simplifies configuration of log sources in the SDDC that are prepackaged with the vRealize Log Insight agent.	You must configure the vRealize Log Insight agent to forward logs to the vRealize Log Insight VIP.
SDDC-OPS-LOG-023	Configure the vRealize Log Insight agent for the vRealize Operation Manager appliances, including: • Analytics nodes • Remote collectors	Simplifies configuration of log sources in the SDDC that are prepackaged with the vRealize Log Insight agent.	You must configure the vRealize Log Insight agent to forward logs to the vRealize Log Insight VIP.
SDDC-OPS-LOG-024	Configure the NSX for vSphere components as direct syslog sources for vRealize Log Insight, including: • NSX Manager • NSX Controller instances • NSX Edge services gateways	Simplifies configuration of log sources in the SDDC that are syslog-capable.	• You must manually configure syslog sources to forward logs to the vRealize Log Insight VIP. • Not all operating system-level events are forwarded to vRealize Log Insight.
SDDC-OPS-LOG-025	Configure the vCenter Server Appliance and Platform Services Controller instances as direct syslog sources to send log data directly to vRealize Log Insight.	Simplifies configuration for log sources that are syslog-capable.	• You must manually configure syslog sources to forward logs to the vRealize Log Insight VIP. • Certain dashboards in vRealize Log Insight require the use of the vRealize Log Insight agent for proper ingestion. • Not all operating system-level events are forwarded to vRealize Log Insight.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-026	Configure vRealize Log Insight to ingest events, tasks, and alarms from the Management vCenter Server and Compute vCenter Server instances.	All tasks, events, and alarms generated across all vCenter Server instances in a specific region of the SDDC are captured and analyzed for the administrator.	• You must create a service account on vCenter Server to connect vRealize Log Insight for event, task, and alarm pulling. • Configuring vSphere Integration in vRealize Log Insight does not capture events that occur on the Platform Services Controller.
SDDC-OPS-LOG-027	Communicate with the syslog clients, such as ESXi, vCenter Server, and NSX for vSphere, by using the default syslog UDP protocol.	• Using the default UDP syslog protocol simplifies configuration for all syslog sources. • The UDP syslog protocol is the most common logging protocol available across products. • UDP has a lower performance overhead compared to TCP.	• If the network connection is interrupted, syslog traffic is lost. • UDP syslog traffic is not secure. • The UDP syslog protocol does not support reliability and retry mechanisms.
SDDC-OPS-LOG-028	Include the syslog configuration for vRealize Log Insight in the host profile for the following clusters: • Management • Shared edge and compute • Any additional compute	Simplifies the configuration of the hosts in the cluster and makes sure that settings are uniform across the cluster	Every time you make an authorized change to a host regarding the syslog configuration, you must update the host profile to reflect the change or the status shows Non-compliant.
SDDC-OPS-LOG-029	Deploy and configure the vRealize Log Insight agent for the Site Recovery Manager Windows servers.	• Windows does not natively support syslog. • VMware Site Recovery Manager requires the use of agents to collect all application-specific logs.	You must manually install and configure the agents on several nodes.
SDDC-OPS-LOG-030	Do not configure vRealize Log Insight to automatically update all deployed agents.	Manually install updated versions of the Log Insight Agents for each of the specified components in the SDDC for precise maintenance.	You must manually maintain the vRealize Log Insight Agents on each of the SDDC components.

Time Synchronization in vRealize Log Insight

Time synchronization is critical for the core functionality of vRealize Log Insight. By default, vRealize Log Insight synchronizes time with a predefined list of public NTP servers.

NTP Configuration

Configure consistent NTP sources on all systems that send log data (vCenter Server, ESXi, and vRealize Operation Manager). See Time Synchronization in the VMware Validated Design Planning and Preparation documentation.

Table 118) Time synchronization for vRealize Log Insight design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-031	Configure consistent NTP sources on all virtual infrastructure and cloud management applications for correct log analysis in vRealize Log Insight.	Guarantees accurate log time stamps.	All applications must synchronize time to the same NTP time source.

Content Packs in vRealize Log Insight

The SDDC contains several VMware products for networking, storage, and cloud management. Use content packs to have the logs generated from these components retrieved, extracted, and parsed into a human-readable format. In this way, Log Insight saves log queries and alerts, and you can use dashboards for efficient monitoring.

Table 119) Content packs for vRealize Log Insight design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-032	Install the following content packs: • VMware - Linux • VMware - NSX-vSphere • VMware - Orchestrator 7.0.1 • VMware - SRM • VMware - vRA 7 • Microsoft - SQL Server	Provides additional granular monitoring on the virtual infrastructure. Do not install the following content packs, because they are installed by default in vRealize Log Insight: • General • VMware - vSphere • VMware - vRops 6.x	Requires manual installation and configuration of each nondefault content pack.
SDDC-OPS-LOG-033	Configure the following agent groups that are related to content packs: • vRealize Automation (Linux) • vRealize Automation (Windows) • VMware Virtual Appliances • vRealize Operations Manager • vRealize Orchestrator • VMware Site Recover Manager • Microsoft SQL Server	• Provides a standardized configuration that is pushed to all of the vRealize Log Insight Agents in each of the groups. • Supports application-contextualized collection and parsing of the logs generated from the SDDC components, such as specific log directories, log files, and logging formats, by the vRealize Log Insight agent.	Adds minimal load to vRealize Log Insight.

Event Forwarding Between Regions with vRealize Log Insight

vRealize Log Insight supports event forwarding to other clusters and standalone instances. While forwarding events, the vRealize Log Insight instance still ingests, stores, and archives events locally.

Forward syslog data in vRealize Log Insight by using the Ingestion API or a native syslog implementation.

The vRealize Log Insight Ingestion API uses TCP communication. In contrast to syslog, the forwarding module supports the following features for the Ingestion API:

- Forwarding to other vRealize Log Insight instances
- Both structured and unstructured data; that is, multiline messages
- Metadata in the form of tags
- Client-side compression
- Configurable disk-backed queue to save events until the server acknowledges the ingestion

Table 120) Event forwarding across regions in vRealize Log Insight design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-034	Forward log events to the other region by using the Ingestion API.	Using the forwarding protocol supports the following operations: • Structured and unstructured data for client-side compression. • Event throttling from one vRealize Log Insight cluster to the other. Forwarding makes sure that, during a disaster recovery situation, the administrator has access to all logs from the two regions, although one region is offline.	• You must configure each region to forward log data to the other. The configuration requires administrative overhead to prevent recursion of logging between regions using inclusion and exclusion tagging. • Log forwarding adds more load on each region. You must consider log forwarding in the sizing calculations for the vRealize Log Insight cluster in each region. • You must configure and identically size both source and destination clusters.
SDDC-OPS-LOG-035	Configure log forwarding to use SSL.	Makes sure that the log forwarding operations from one region to the other are secure.	• You must set up a custom CA-signed SSL certificate. • Event forwarding with SSL does not work with the self-signed certificate that is installed on the destination servers by default. • If you add more vRealize Log Insight nodes to a region's cluster, the SSL certificate used by the vRealize Log Insight cluster in the other region must be installed in the Java keystore of the nodes before SSL can be used.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-036	Configure disk cache for event forwarding to 2,000MB (2GB).	Makes sure that log forwarding between regions has a buffer for approximately 2 hours if a cross-region connectivity outage occurs. The disk cache size is calculated at a base rate of 150MB each day for each syslog source with 110 syslog sources.	• If the event forwarder of vRealize Log Insight is restarted during the cross-region communication outage, messages that reside in the nonpersistent cache are cleared. • If a cross-region communication outage exceeds 2 hours, the newest local events are dropped and not forwarded to the remote destination, even after the cross-region connection is restored.

Disaster Recovery of vRealize Log Insight

Each region is configured to forward log information to the vRealize Log Insight instance in the other region.

Because of the forwarding configuration, an administrator of the SDDC can use either of the vRealize Log Insight clusters in the SDDC to query the available logs from one of the regions. As a result, you do not have to configure failover for the vRealize Log Insight clusters, and each cluster can remain associated with the region in which it was deployed.

11.3 vSphere Update Manager Design

vSphere Update Manager supports patch and version management of ESXi hosts and virtual machines. vSphere Upgrade Manager is connected to a vCenter Server instance to retrieve information about and push upgrades to the managed hosts.

vSphere Update Manager can remediate the following objects over the network:

- VMware Tools and VMware virtual machine hardware upgrade operations for virtual machines
- ESXi host patching operations
- ESXi host upgrade operations
- Physical Design of vSphere Update Manager

  You use the vSphere Update Manager service on each vCenter Server Appliance and deploy a vSphere UMDS in Region A and Region B to download and stage upgrade and patch data.
- Logical Design of vSphere Update Manager

  You configure vSphere Update Manager to apply updates on the management components of the SDDC according to the objectives of this design.

Physical Design of vSphere Update Manager

You use the vSphere Update Manager service on each vCenter Server Appliance and deploy a vSphere UMDS in Region A and Region B to download and stage upgrade and patch data.

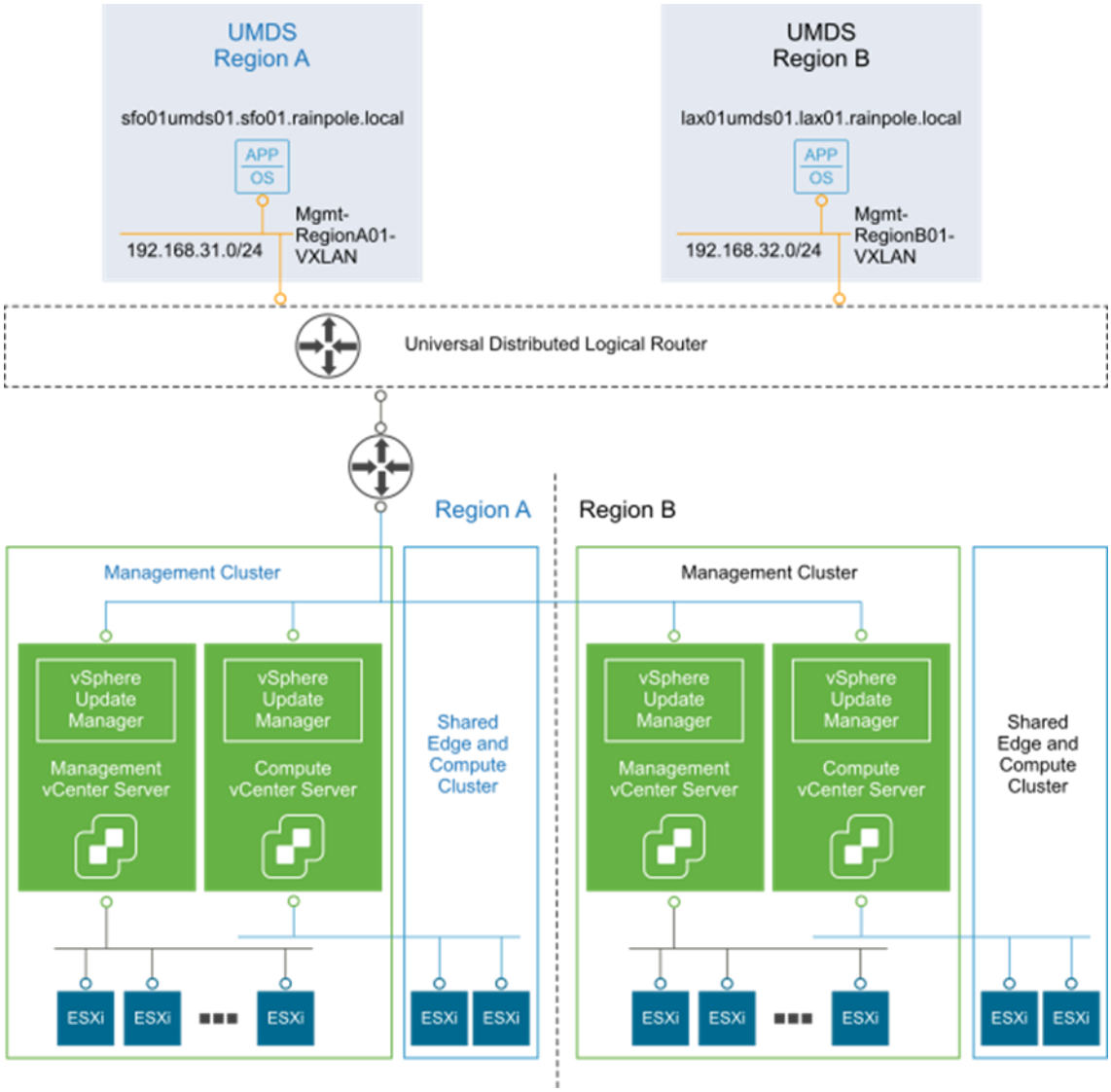
Networking and Application Design

You can use the vSphere Update Manager as a service of the vCenter Server Appliance. The Update Manager server and client components are part of the vCenter Server Appliance.

You can connect only one vCenter Server instance to a vSphere Update Manager instance.

Because this design uses multiple vCenter Server instances, you must configure a separate vSphere Update Manager for each vCenter Server. To avoid downloading updates on multiple vSphere Update Manager instances and to restrict access to the external network from vSphere Update Manager and vCenter Server, deploy a UMDS in each region. UMDS downloads upgrades, patch binaries, and patch metadata, and stages the downloaded data on a web server. The local Update Manager servers download the patches from UMDS.

Figure 64) vSphere Update Manager logical and networking design.



Deployment Model

vSphere Update Manager is preinstalled in the vCenter Server Appliance. After you deploy or upgrade the vCenter Server Appliance, the VMware vSphere Update Manager service starts automatically.

In addition to the vSphere Update Manager deployment, two models for downloading patches from VMware exist:

- **Internet-connected model.** The vSphere Update Manager server is connected to the VMware patch repository to download patches for ESXi hosts and virtual appliances. No additional configuration is required, other than to scan and remediate the hosts as needed.
- **Proxied access model.** For security reasons, vSphere Update Manager is placed on a safe internal network with no connection to the internet. Therefore, it cannot download patch metadata. You deploy UMDS to download and store patch metadata and binaries to a shared repository. vSphere Update Manager uses the shared repository as a patch datastore before remediating the ESXi hosts.

Table 121) Update Manager physical design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-001	Use the vSphere Update Manager service on each vCenter Server Appliance to provide a total of four vSphere Update Manager instances that you configure and use for patch management.	• Reduces the number of management virtual machines that need to be deployed and maintained in the SDDC. • Enables centralized, automated patch and version management for VMware vSphere and offers support for VMware ESXi hosts, virtual machines, and virtual appliances managed by each vCenter Server.	• All physical design decisions for vCenter Server determine the setup for vSphere Update Manager. • A one-to-one mapping of vCenter Server to vSphere Update Manager is required. Each Management vCenter Server or Compute vCenter Server instance in each region needs its own vSphere Update Manager.
SDDC-OPS-VUM-002	Use the embedded PostgreSQL of the vCenter Server Appliance for vSphere Update Manager.	• Reduces both overhead and licensing cost for external enterprise database systems. • Avoids problems with upgrades.	The vCenter Server Appliance has limited database management tools for database administrators.
SDDC-OPS-VUM-003	Use the network settings of the vCenter Server Appliance for vSphere Update Manager.	Simplifies network configuration because of the one-to-one mapping between vCenter Server and vSphere Update Manager. You configure the network settings once for both vCenter Server and vSphere Update Manager.	None
SDDC-OPS-VUM-004	Deploy and configure UMDS virtual machines for each region.	Limits direct access to the internet from vSphere Update Manager on multiple vCenter Server instances and reduces storage requirements on each instance.	You must maintain the host operating system and the database used by the UMDS.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-005	Connect the UMDS virtual machines to the region-specific application virtual network.	• Provides local storage and access to vSphere Update Manager repository data. • Avoids cross-region bandwidth usage for repository access. • Provides a consistent deployment model for management applications.	You must use NSX to support this network configuration.

Logical Design of vSphere Update Manager

You configure vSphere Update Manager to apply updates on the management components of the SDDC according to the objectives of this design.

UMDS Virtual Machine Specification

Allocate resources to and configure the virtual machines for UMDS according to the specifications shown in Table 122.

Table 122) UMDS virtual machine specifications.

Attribute	Specification
vSphere Update Manager Download Service	vSphere 6.5
Number of CPUs	2
Memory	2GB
Disk space	120GB
Operating system	Ubuntu 14.04 LTS

ESXi Host and Cluster Settings

When you perform updates by using the vSphere Update Manager, the update operation affects certain cluster and host base settings. Customize these settings according to your business requirements and use cases.

Table 123) Host and cluster settings that are affected by vSphere Update Manager.

Settings	Description
Maintenance mode	During remediation, updates might require the host to enter maintenance mode. However, virtual machines cannot run when a host is in maintenance mode. For availability during a host update, virtual machines are migrated to other ESXi hosts in a cluster before the host enters maintenance mode. However, putting a host in maintenance mode during update might cause issues with the availability of the cluster.

You can control an update operation by using a set of host and cluster settings in vSphere Update Manager.

Table 124) Host and cluster settings for updates.

Level	Setting	Description
Host settings	Virtual machine power state when entering maintenance mode.	You can configure vSphere Update Manager to power off, to suspend, or to not control virtual machines during remediation. This option applies only if vSphere vMotion is not available for a host.
	Retry maintenance mode in case of failure.	If a host fails to enter maintenance mode before remediation, vSphere Update Manager waits for a retry delay period and retries putting the host into maintenance mode as many times as you indicate.
	Allow installation of additional software on PXE-booted hosts.	You can install solution software on PXE-booted ESXi hosts. This option is limited to software packages that do not require a host reboot after installation.
Cluster settings	Disable vSphere Distributed Power Management (DPM), vSphere High Availability (HA) Admission Control, and Fault Tolerance (FT).	vSphere Update Manager does not remediate clusters with active DPM, HA, and FT.
	Enable parallel remediation of hosts.	vSphere Update Manager can remediate multiple hosts.
	Migrate powered-off or suspended virtual machines.	vSphere Update Manager migrates suspended and powered-off virtual machines from hosts that must enter maintenance mode to other hosts in the cluster. The migration is launched on virtual machines that do not prevent the host from entering maintenance mode.

Virtual Machine and Virtual Appliance Update Settings

vSphere Update Manager supports remediation of virtual machines and appliances. You can provide application availability upon virtual machine and appliance updates by performing the operations described in Table 125.

Table 125) vSphere Update Manager settings for remediation of virtual machines and appliances.

Configuration	Description
Take snapshots before virtual machine remediation.	If the remediation fails, use the snapshot to return the virtual machine to the state before the remediation.
Define the window in which a snapshot persists for a remediated virtual machine.	Automatically clean up virtual machine snapshots that are taken before remediation.
Enable smart rebooting for VMware vSphere vApps remediation.	Start virtual machines after remediation to maintain start-up dependencies even if some of the virtual machines are not remediated.

Baselines and Groups

vSphere Update Manager baselines and baseline groups are collections of patches that you can assign to a cluster or host in the environment. According to the business requirements, the default baselines might not be allowed until patches are tested or verified on development or preproduction hosts.

Baselines can be confirmed so that the tested patches are applied to hosts and updated only when appropriate.

Table 126) Baselines and baseline groups details.

Baseline or Baseline Group Feature		Description
Baselines	Types	There are four types of baselines: • **Dynamic baselines.** Change as items are added to the repository. • **Fixed baselines.** Remain the same. • **Extension baselines.** Contain additional software modules for ESXi hosts for VMware software or third-party software, such as device drivers. • **System-managed baselines.** Are automatically generated according to your vSphere inventory. A system-managed baseline is available in your environment for an ESXi patch, upgrade, or extension. You cannot add system-managed baselines to a baseline group or attach or detach them.
	Default baselines	vSphere Update Manager contains the following default baselines. Each of these baselines is configured for dynamic selection of new items. • **Critical host patches.** Upgrade hosts with a collection of critical patches that are high priority as defined by VMware. • **Noncritical host patches.** Upgrade hosts with patches that are not classified as critical. • **VMware Tools Upgrade to Match Host.** Upgrades the VMware Tools version to match the host version. • **VM Hardware Upgrade to Match Host.** Upgrades the VMware Tools version to match the host version. • **VA Upgrade to Latest.** Upgrades a virtual appliance to the latest version available.
Baseline groups	Definition	A baseline group consists of a set of nonconflicting baselines. You use baseline groups to scan and remediate objects against multiple baselines at the same time. Use baseline groups to construct an orchestrated upgrade that contains a combination of an upgrade baseline, a patch baseline, or extension baselines.
	Types	You can create two types of baseline groups according to the object type: • Baseline groups for ESXi hosts • Baseline groups for virtual machines

ESXi Image Configuration

You can store full images that you can use to upgrade ESXi hosts. These images cannot be automatically downloaded by vSphere Update Manager from the VMware patch repositories. You must obtain the image files from the VMware website or a vendor-specific source. The image can then be uploaded to vSphere Update Manager.

There are two ways to add packages to an ESXi image:

- **By using Image Builder.** If you use Image Builder, add the NSX software packages, such as `esx-vdpi`, `esx-vsip`, and `esx-vxlan`, to the ESXi upgrade image. You can then upload this slipstreamed ESXi image to vSphere Update Manager so that you can use the hosts being upgraded

in a software-defined networking setup. Such an image can be used for both upgrades and future fresh ESXi installations.

- **By using a baseline group.** If you use a baseline group, you can add additional patches and extensions, such as the NSX software packages `esx-vdpi`, `esx-vsip`, and `esx-vxlan`, to an upgrade baseline containing the ESXi image. In this way, vSphere Update Manager can orchestrate the upgrade while ensuring that the patches and extensions are nonconflicting. To do so, follow these steps:
  1. Download the NSX software package bundle from NSX Manager.
  2. Include the NSX software packages, such as `esx-vdpi`, `esx-vsip`, and `esx-vxlan`, in an extension baseline.
  3. Combine the extension baseline with the ESXi upgrade baseline in a baseline group so that you can use the hosts being upgraded in a software-defined networking setup.

vSphere Update Manager Logical Design Decisions

This design applies the decisions described in the following table on the logical design of vSphere Update Manager and update policy.

Table 127) vSphere Update Manager logical design decisions.

Design ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-006	Use the default patch repositories by VMware.	Simplifies the configuration because you do not configure additional sources.	None
SDDC-OPS-VUM-007	Set the virtual machine power state to Do Not Power Off.	Provides the most uptime for management components and compute workload virtual machines.	If the migration fails, you must manually intervene.
SDDC-OPS-VUM-008	Enable parallel remediation of hosts, assuming that enough resources are available to update multiple hosts at the same time.	Provides fast remediation of host patches.	More resources are unavailable at the same time during remediation.
SDDC-OPS-VUM-009	Enable migration of powered-off virtual machines and templates.	Makes sure that templates stored on all management hosts are accessible.	Increases the amount of time to start remediation for templates to be migrated.
SDDC-OPS-VUM-010	Use the default critical and noncritical patch baselines for the management cluster and for the shared edge and compute cluster.	Simplifies configuration because you can use the default baselines without customization.	All patches are added to the baselines as soon as they are released.
SDDC-OPS-VUM-011	Use the default schedule of a once-per-day check and patch download.	Simplifies configuration because you can use the default schedule without customization.	None

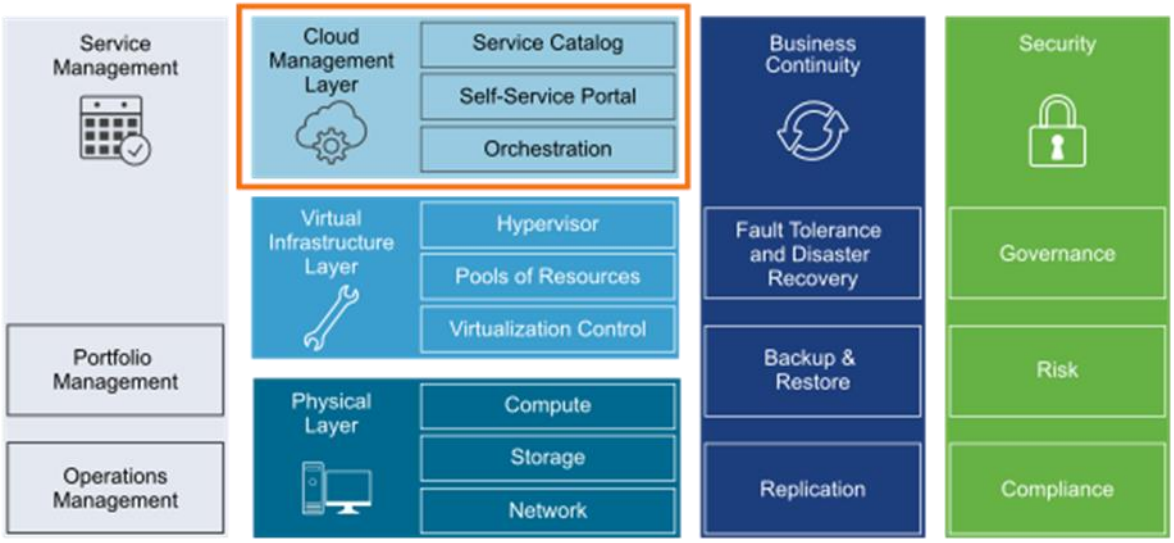
Design ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-012	Remediate hosts, virtual machines, and virtual appliances once a month, or according to business guidelines.	Aligns the remediation schedule with the business policies.	None
SDDC-OPS-VUM-013	Use a baseline group to add NSX for vSphere software packages to the ESXi upgrade image.	• Allows parallel remediation of ESXi hosts by making sure that the ESXi hosts are ready for software-defined networking immediately after the upgrade. • Prevents additional NSX remediation.	NSX for vSphere updates require periodic updates to the group baseline.
SDDC-OPS-VUM-014	Configure an HTTP web server on each UMDS service that the connected vSphere Update Manager servers must use to download patches from.	Enables the automatic download of patches on vSphere Update Manager from UMDS. The alternative is to manually copy media from one place to another.	You must be familiar with a third-party web service such as Nginx or Apache.

12 Cloud Management Platform Design

The CMP layer is the management component of the SDDC. The CMP layer allows you to deliver tenants with automated workload provisioning by using a self-service portal.

The CMP layer includes the components and functionality depicted in Figure 65.

Figure 65) The CMP layer in the SDDC.



- **Service catalog.** A self-service portal where users can browse and request the IT services and resources they need, such as a virtual machine or a machine on AWS. When you request a service catalog item, you provision the item to the designated cloud environment.
- **Self-service portal.** A unified interface for consuming IT services. Users can browse the service catalog to request IT services and resources, track their requests, and manage their provisioned items.
- **Orchestration.** Provides automated workflows to deploy service catalog items requested by users. You use the workflows to create and run automated, configurable processes to manage your SDDC infrastructure, as well as other VMware and third-party technologies.

  Note: vRealize Automation provides the self-service portal and the service catalog. Orchestration is enabled by an instance of vRealize Orchestrator that is internal to vRealize Automation.

12.1 vRealize Automation Design

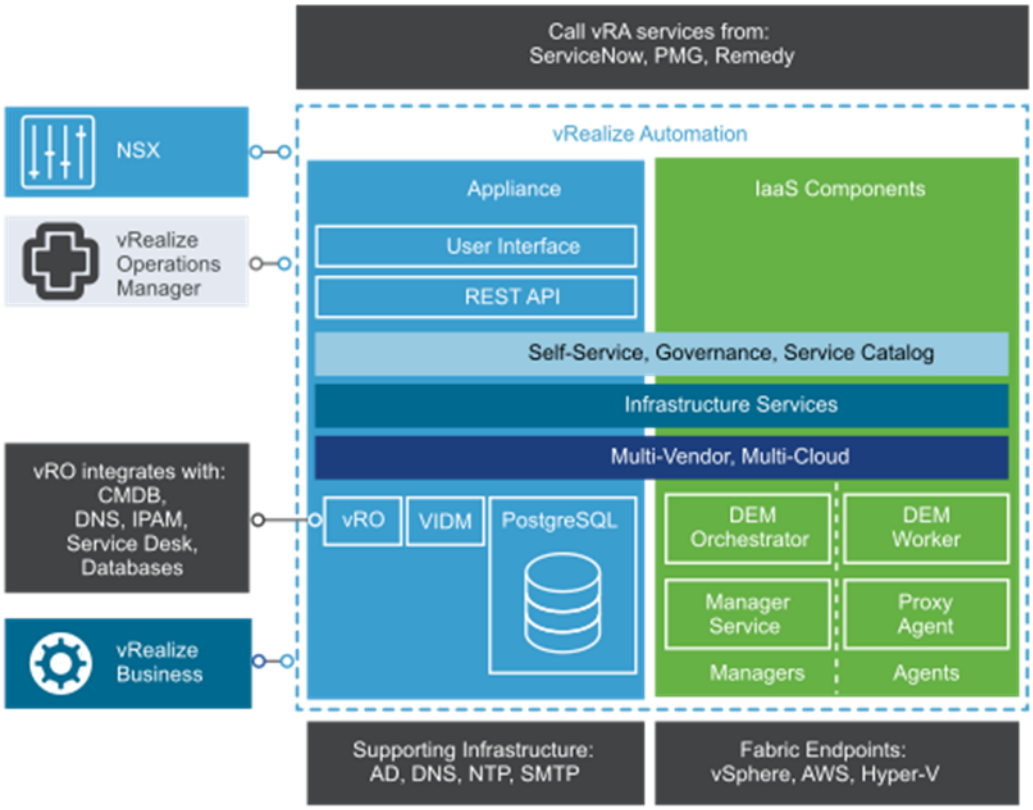
VMware vRealize Automation provides a service catalog from which tenants can deploy applications, and a portal that lets you deliver a personalized, self-service experience to end users.

vRealize Automation Logical Design

vRealize Automation offers several extensibility options that are designed to support a variety of use cases and integrations. In addition, the CMP, of which vRealize Automation is a central component, enables a usage model that includes interactions between users, the CMP itself, and integrations with the supporting infrastructure.

Figure 66 illustrates the vRealize Automation internal components and the integration of those components with other external components and the supporting infrastructure of the SDDC.

Figure 66) vRealize Automation logical architecture, extensibility, and external integrations.



- **Fabric endpoints.** vRealize Automation can leverage existing and future infrastructure representing multivendor and multicloud virtual, physical, and public cloud infrastructures. Each kind of supported infrastructure is represented by a fabric endpoint.
- **Call vRealize Automation services from existing applications.** vRealize Automation provides a RESTful API that can be used to call vRealize Automation application and infrastructure services from IT service management applications such as ServiceNow, the PMG Digital Business Platform, and BMC Remedy.
- **vRealize Business for Cloud.** vRealize Business for Cloud is tightly integrated with vRealize Automation. Together, they manage the resource costs of vRealize Automation by displaying this information during workload requests and on an ongoing basis with cost reporting by user, business group, or tenant. vRealize Business for Cloud supports pricing based on blueprints, endpoints, reservations, and reservation policies for the Compute Grouping Strategy. In addition, vRealize Business for Cloud supports the storage path and storage reservation policies for the storage grouping strategy.
- **vRealize Operations management.** The vRealize Automation management pack for vRealize Operation Manager provides comprehensive visibility into the performance and capacity metrics of a vRealize Automation tenant's business groups and underlying cloud infrastructure. By combining these new metrics with the custom dashboard capabilities of vRealize Operations, you can gain a high level of flexibility and insight when monitoring these complex environments.
- **Supporting infrastructure.** vRealize Automation integrates with the following supporting infrastructure:
  - Microsoft SQL Server to store data relating to the vRealize Automation IaaS elements
  - An NTP server to synchronize the time between the vRealize Automation components
  - Active Directory support of vRealize Automation tenant user authentication and authorization
  - Sending and receiving notification emails through SMTP for various actions that can be executed in the vRealize Automation console
- **NSX.** NSX and vRealize Automation integration offers several options for designing and authoring blueprints with the networking and security features provided by NSX. These blueprints take full advantage of all NSX network constructs, including switches, routers, and firewalls. This integration allows you to use an on-demand load balancer, on-demand NAT network, on-demand routed network, and on-demand security groups in a blueprint. When the blueprint is requested, it is automatically provisioned by vRealize Automation. Integration with NSX eliminates the need to provision networking as a separate activity outside of vRealize Automation.

Cloud Management Platform Usage Model

The CMP, of which vRealize Automation is a central component, enables interaction among users, the CMP itself, the supporting infrastructure, and the provisioning infrastructure. Figure 67 illustrates the usage model of the CMP in relation to these elements.

Figure 67) vRealize Automation usage model.

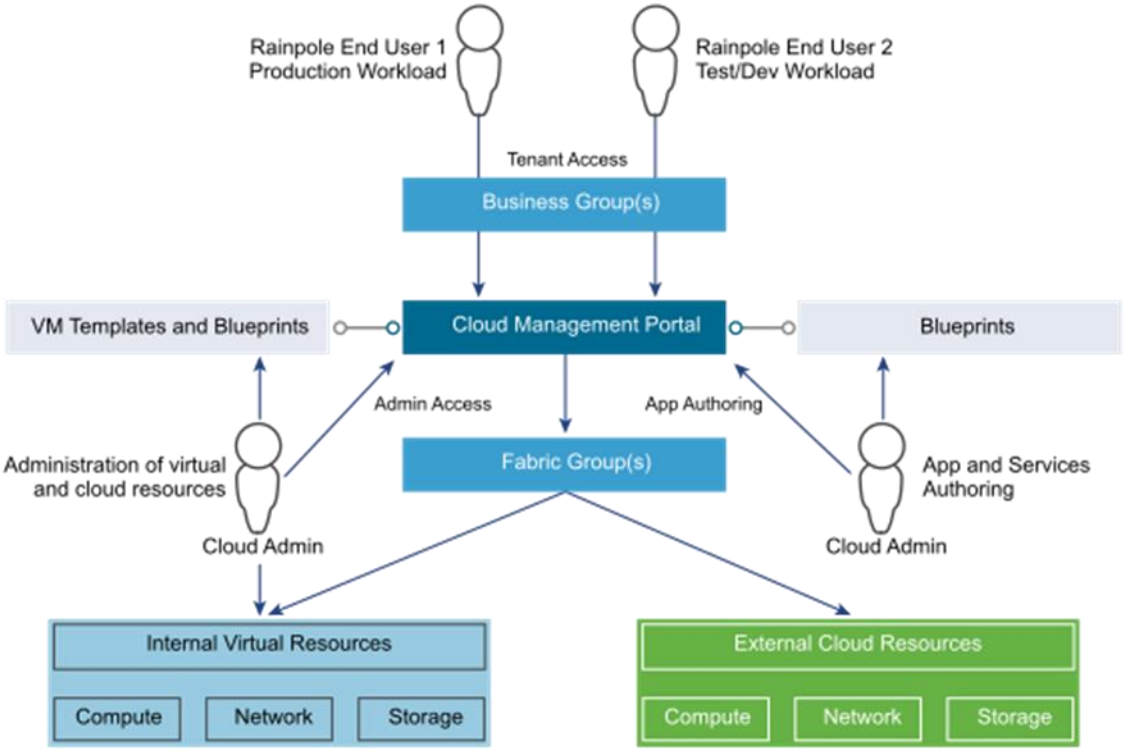


Table 128 lists the vRealize Automation elements and the components that make up each of these elements.

Table 128) vRealize Automation elements.

Element	Components	
Users	**Cloud administrators**	Tenant, group, fabric, infrastructure, service, and other administrators as defined by business policies and organizational structure.
	Cloud (or tenant) users	Users in an organization who can provision virtual machines and directly perform operations on them at the level of the operating system.
Tools and supporting infrastructure	Virtual machine templates and blueprints provide the foundation for the cloud. Virtual machine templates are used to author the blueprints that tenants (end users) use to provision their cloud workloads.	
Provisioning infrastructure	On-premises and off-premises resources that together form a hybrid cloud.	
	Internal virtual resources	Supported hypervisors and associated management tools.
	External cloud resources	Supported cloud providers and associated APIs.

Element	Components
Cloud management portal	A portal that provides self-service capabilities for users to administer, provision, and manage workloads. **vRealize Automation portal, admin access** — The default root tenant portal URL used to set up and administer tenants and global configuration options. **vRealize Automation portal, tenant access** — Refers to a subtenant that is accessed by using an appended tenant identifier. **Note:** In some configurations, a tenant portal might refer to the default tenant portal. In that case, the URLs match and the user interface is contextually controlled by the role-based access control permissions assigned to the tenant.

vRealize Automation Physical Design

The physical design consists of characteristics and decisions that support the logical design. The design objective is to deploy a fully functional cloud management portal with high availability and to provision to both Regions A and B.

To accomplish this design objective, deploy or leverage the following components in Region A to create a cloud management portal for the SDDC:

- Two vRealize Automation server appliances
- Two vRealize Automation IaaS web servers
- Two vRealize Automation Manager service nodes (including the DEM Orchestrator)
- Two DEM worker nodes
- Two IaaS Proxy Agent nodes
- One vRealize Business for Cloud server
- One vRealize Business for Cloud remote collector
- Supporting infrastructure such as Microsoft SQL Server, Active Directory, DNS, NTP, and SMTP

Place the vRealize Automation components in several network units for isolation and failover. All the components that make up the cloud management portal, along with their network connectivity, are shown in Figure 68 and Figure 69.

Figure 68) vRealize Automation design for Region A.

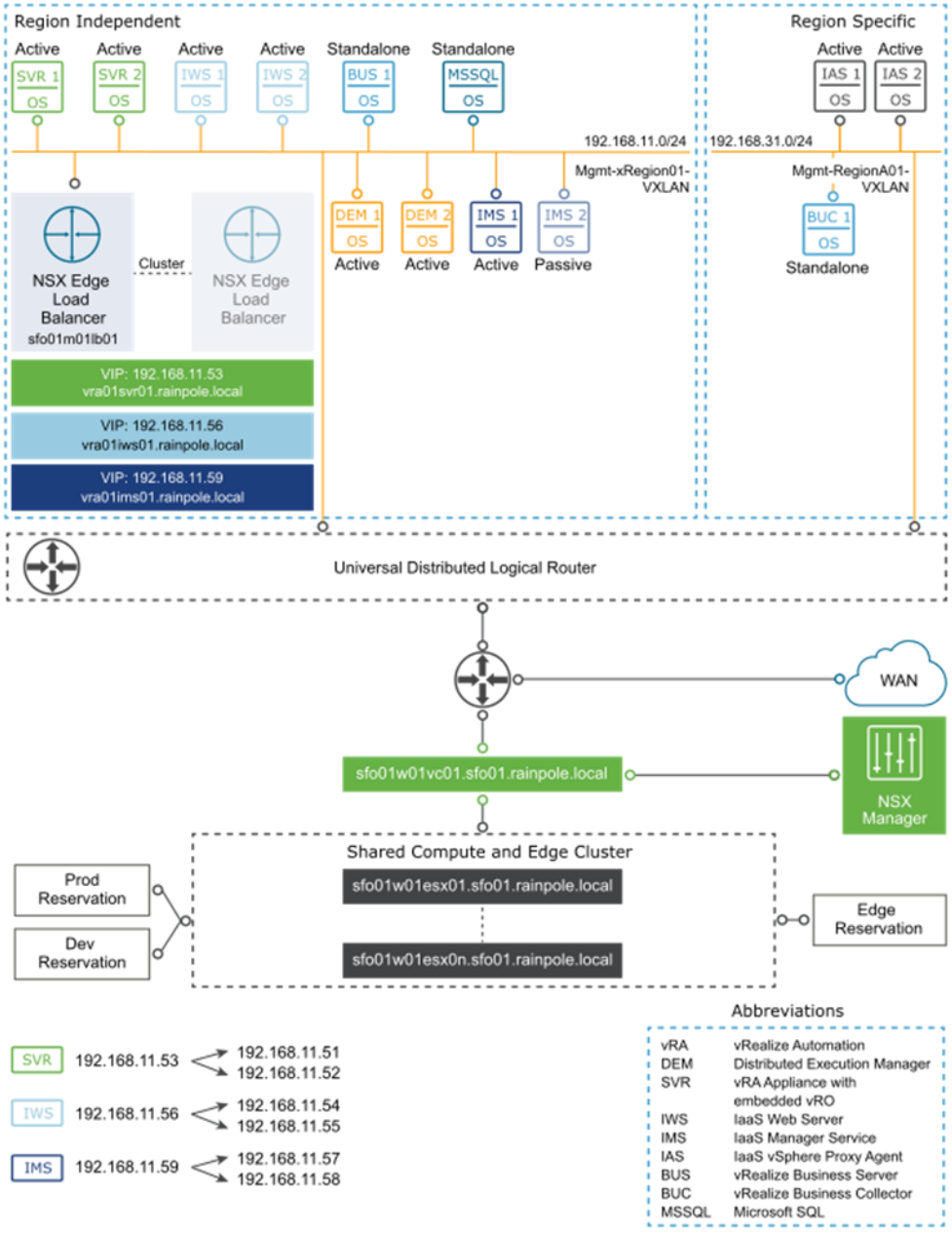
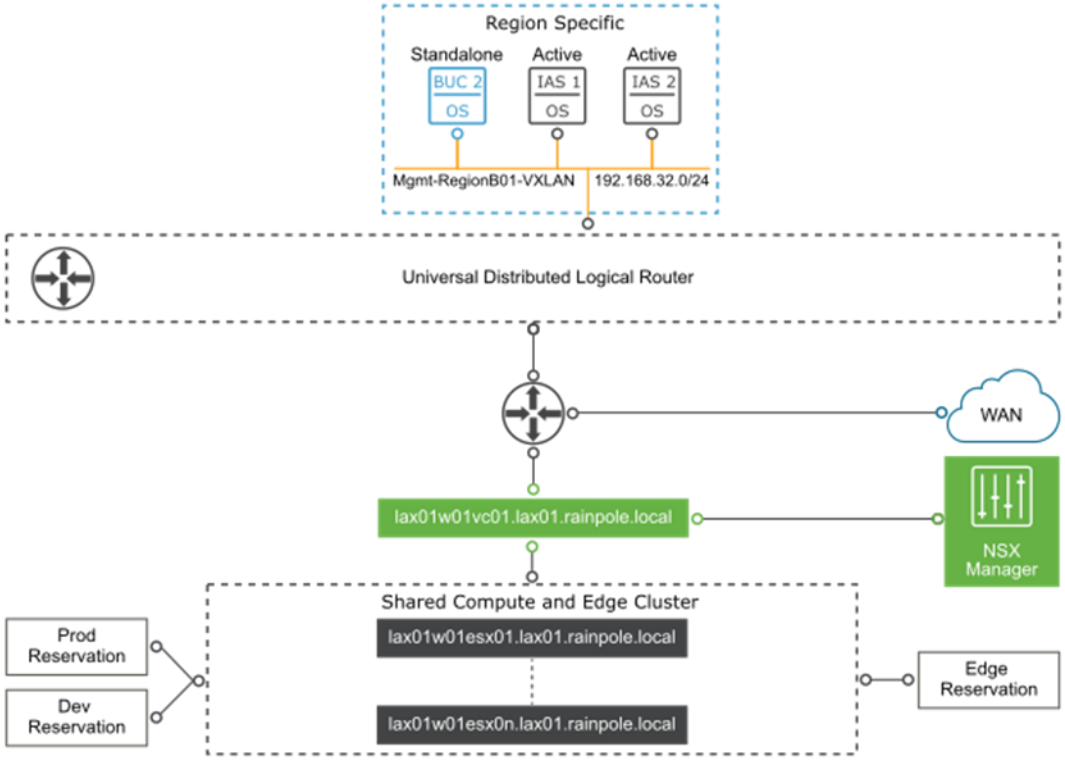


Figure 69) vRealize Automation design for Region B.



Deployment Considerations

This design uses NSX logical switches to abstract the vRealize Automation application and its supporting services. This abstraction allows the application to be hosted in any region regardless of the underlying physical infrastructure, including network subnets, compute hardware, or storage types. This design places the vRealize Automation application and its supporting services in Region A. The same instance of the application manages workloads in both Region A and Region B.

Table 129) vRealize Automation topology design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-001	Use a single vRealize Automation installation to manage both Region A and Region B deployments from a single instance.	vRealize Automation can manage one or more regions and provides a single consumption portal regardless of region. The abstraction of the vRealize Automation application over virtual networking allows it to be independent from any physical site locations or hardware.	You must size vRealize Automation to accommodate multiregion deployments.

Table 130) vRealize Automation anti-affinity rules design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-002	Apply vSphere DRS anti-affinity rules to the vRealize Automation components.	Using DRS prevents vRealize Automation nodes from residing on the same ESXi host and risking the cluster's high-availability capability.	Additional configuration is required to set up anti-affinity rules. Only a single ESXi host in the management cluster, of the four ESXi hosts, can be put into maintenance mode at a time.

Table 131) vRealize Automation IaaS Active Directory requirements design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-003	vRealize Automation IaaS machines are joined to Active Directory.	Active Directory access is a difficult requirement for vRealize Automation.	Active Directory access must be provided by using dedicated service accounts.

vRealize Automation Appliance

The vRealize Automation virtual appliance includes the cloud management web portal, an embedded vRealize Orchestrator instance, and database services. The vRealize Automation portal allows self-service provisioning and management of cloud services, as well as authoring blueprints, administration, and governance. The vRealize Automation virtual appliance uses an embedded PostgreSQL database for catalog persistence and database replication. The database is configured between two vRealize Automation appliances for high availability.

Table 132) vRealize Automation virtual appliance design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-004	Deploy two instances of the vRealize Automation virtual appliance to achieve redundancy. Each of these virtual appliances hosts an embedded vRealize Orchestrator instance.	Enable an active-active front-end portal for high availability.	None
SDDC-CMP-005	Deploy two appliances that replicate data by using the embedded PostgreSQL database.	Enable high availability for vRealize Automation. The embedded vRealize Orchestrator instance also uses this database.	In this active-passive configuration, manual failover between the two instances is required.
SDDC-CMP-006	During deployment, configure the vRealize Automation appliances with 18GB vRAM.	Supports deployment of vRealize Automation in environments with up to 25,000 Active Directory users.	For environments with more than 25,000 Active Directory users of vRealize Automation, vRAM must be increased to 22GB.

Table 133) vRealize Automation virtual appliance resource requirements per virtual machine.

Attribute	Specification
Number of vCPUs	4
Memory	18GB
vRealize Automation function	Portal website, Application, Orchestrator, service catalog, and Identity Manager.

vRealize Automation IaaS Web Server

The vRealize Automation IaaS web server provides a user interface in the vRealize Automation portal (a website) for the administration and consumption of IaaS components.

The IaaS website provides infrastructure administration and service authoring capabilities to the vRealize Automation console. The website component communicates with the Model Manager, which provides it with updates from the DEM, proxy agents, and database.

The Model Manager communicates with the database, the DEMs, and the portal website. The Model Manager is divided into two separately installable components: the Model Manager web service and the Model Manager data component.

Note: The vRealize Automation IaaS web server is a separate component from the vRealize Automation appliance.

Table 134) vRealize Automation IaaS web server design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-007	Install two vRealize Automation IaaS web servers.	vRealize Automation can support between 1,000 and 10,000 virtual machines. Two vRealize Automation IaaS web servers provide redundancy to the IaaS web server components.	Operational overhead increases as more servers are deployed.

Table 135) vRealize Automation IaaS web server resource requirements.

Attribute	Specification
Number of vCPUs	2
Memory	4GB
Number of vNIC ports	1
Number of local drives	1
vRealize Automation functions	Model Manager (web service)
Operating system	Microsoft Windows Server 2012 SP2 R2

vRealize Automation IaaS Manager Service and DEM Orchestrator Server

The vRealize Automation IaaS Manager Service and DEM server is at the core of the vRealize Automation IaaS platform. The vRealize Automation IaaS Manager Service and DEM server supports the following functions:

- Manages the integration of vRealize Automation IaaS with external systems and databases
- Provides business logic to the DEMs
- Manages business logic and execution policies
- Maintains all workflows and their supporting constructs

A DEM runs the business logic of custom models by interacting with other vRealize Automation components (repository) as required.

Each DEM instance acts in either an Orchestrator role or a Worker role. The DEM Orchestrator monitors the status of the DEM Workers. If a DEM Worker stops or loses the connection to the Model Manager or repository, the DEM Orchestrator puts the workflow back in the queue. It manages the scheduled workflows by creating new workflow instances at the scheduled time and allows only one instance of a particular scheduled workflow to run at a given time. It also preprocesses workflows before execution. Preprocessing includes checking preconditions for workflows and creating the workflow's execution history.

Note: The vRealize Automation IaaS Manager Service and DEM Orchestrator service are separate services but are installed on the same virtual machine.

Table 136) vRealize Automation IaaS Model Manager and DEM Orchestrator server design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-008	Deploy two virtual machines to run the vRealize Automation IaaS Manager Service and the DEM Orchestrator services in a load-balanced pool.	The vRealize Automation installer enables the automatic failover of IaaS Manager server in an outage of the first node. Automatic failover eliminates a single point of failure for the Manager Service. The DEM Orchestrator must have strong network connectivity to the model manager at all times.	You must provide more resources for these two virtual machines to accommodate the load of the two applications. If additional resources are required in the future, you can scale up these virtual machines at a later stage.

Table 137) vRealize Automation IaaS Model Manager and DEM Orchestrator server resource requirements per virtual machine.

Attribute	Specification
Number of vCPUs	2
Memory	4GB
Number of vNIC ports	1
Number of local drives	1
vRealize Automation functions	IaaS Manager Service, DEM Orchestrator
Operating system	Microsoft Windows Server 2012 SP2 R2

vRealize Automation IaaS DEM Worker Virtual Machine

vRealize Automation IaaS DEM Workers are responsible for executing provisioning and deprovisioning tasks initiated by the vRealize Automation portal. DEM Workers are also used to communicate with specific infrastructure endpoints.

Table 138) vRealize Automation IaaS DEM Worker design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-009	Install three DEM Worker instances per DEM host.	Each DEM Worker can process up to 30 concurrent workflows. Beyond this limit, workflows are queued for execution. If the number of concurrent workflows is consistently above 90, you can add additional DEM Workers on the DEM host.	If you add more DEM Workers, you must also provide additional resources to run them.

Table 139) vRealize Automation DEM Worker resource requirements per virtual machine.

Attribute	Specification
Number of vCPUs	2
Memory	6GB
Number of vNIC ports	1
Number of local drives	1
vRealize Automation functions	DEM Worker
Operating system	Microsoft Windows Server 2012 SP2 R2

vRealize Automation IaaS Proxy Agent

The vRealize Automation IaaS Proxy Agent is a windows service used to communicate with specific infrastructure endpoints. In this design, the vSphere Proxy agent is used to communicate with vCenter.

The IaaS Proxy Agent server provides the following functions:

- The vRealize Automation IaaS Proxy Agent can interact with different types of infrastructure components. For this design, only the vSphere proxy agent is used.
- vRealize Automation does not itself virtualize resources. Rather, it works with vSphere to provision and manage the virtual machines. It uses vSphere proxy agents to send commands to and collect data from vSphere.

Table 140) vRealize Automation IaaS Proxy Agent resource requirements per virtual machine.

Attribute	Specification
Number of vCPUs	2
Memory	4GB
Number of vNIC ports	1
Number of local drives	1
vRealize Automation functions	vSphere proxy agent
Operating system	Microsoft Windows Server 2012 SP2 R2

Load Balancer

Session persistence of a load balancer allows the same server to serve all requests after a session is established with that server. Session persistence is enabled on the load balancer to direct subsequent requests from each unique session to the same vRealize Automation server in the load balancer pool. The load balancer also handles failover for the vRealize Automation server (Manager Service) because only one Manager Service is active at any one time. Session persistence is not enabled because it is not a required component for the Manager Service.

Table 141) Load balancer design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-012	Set up an NSX edge device for load balancing the vRealize Automation services.	Required to enable vRealize Automation to handle a greater load and obtain a higher level of availability than without load balancers.	Additional configuration is required to configure the load balancer.
CSDDC-CMP-008	Set up an NSX edge device for load balancing the vRealize Automation services.	Enabling this design with a load balancer allows for future expansion of the CMP with application-level HA.	Additional configuration is required to configure the load balancers.
SDDC-CMP-013	Configure load balancer for vRealize Automation server appliance, Remote Console Proxy, and IaaS web server to use the round-robin algorithm with source-IP based persistence with a 1800-second timeout.	The round-robin algorithm provides a good balance of clients between both appliances, while source-IP makes sure that individual clients remain connected to the same appliance. A 1800-second timeout aligns with the vRealize Automation appliance server session timeout value. Sessions that transfer to a different vRealize Automation appliance might result in a poor user experience.	None
SDDC-CMP-014	Configure load balancer for vRealize IaaS Manager Service to use the round-robin algorithm without persistence.	The Manager Service does not need session persistence.	None

Consider the load-balancer characteristics described in Table 142 for vRealize Automation.

Table 142) Load-balancer application profile characteristics.

Server Role	Type	Enable SSL Pass-through	Persistence	Expires in (Seconds)
vRealize Automation - Persistence	HTTPS (443)	Enabled	Source IP	1800

Server Role	Type	Enable SSL Pass-through	Persistence	Expires in (Seconds)
vRealize Automation	HTTPS (443)	Enabled		

Table 143) Load-balancer service monitoring characteristics.

Monitor	Interval	Timeout	Max Retries	Type	Expected	Method	URL	Receive
vRealize Automation appliance	3	10	3	HTTPS	204	GET	/vcac/services/api/health	
vRealize Automation IaaS web	3	10	3	HTTPS		GET	/wapi/api/status/web	REGISTERED
vRealize Automation IaaS Manager	3	10	3	HTTPS		GET	/VMPSProvision	Provision Service
vRealize Orchestrator	3	10	3	HTTPS		GET	/vco-controlcenter/docs	

Table 144) Load-balancer pool characteristics.

Server Role	Algorithm	Monitor	Members	Port	Monitor Port
vRealize Automation appliance	Round robin	vRealize Automation appliance monitor	vRealize Automation appliance nodes	443	
vRealize Automation remote console proxy	Round robin	vRealize Automation appliance monitor	vRealize Automation appliance nodes	8444	443
vRealize Automation IaaS Web	Round robin	vRealize Automation IaaS web monitor	IaaS web nodes	443	
vRealize Automation IaaS Manager	Round robin	vRealize Automation IaaS Manager monitor	IaaS Manager nodes	443	
vRealize Automation Appliance	Round robin	Embedded vRealize Automation Orchestrator Control Center monitor	vRealize Automation appliance nodes	8283	

Table 145) Virtual server characteristics.

Protocol	Port	Default Pool	Application Profile
HTTPS	443	vRealize Automation appliance pool	vRealize Automation - persistence profile
HTTPS	443	vRealize Automation IaaS web pool	vRealize Automation - persistence profile
HTTPS	443	vRealize Automation IaaS Manager pool	vRealize Automation profile
HTTPS	8283	Embedded vRealize Orchestrator Control Center pool	vRealize Automation - persistence profile
HTTPS	8444	vRealize Automation remote console proxy pool	vRealize Automation - persistence profile

Information Security and Access Control in vRealize Automation

Use a service account for authentication and authorization of vRealize Automation to vCenter Server and vRealize Operations Manager to orchestrate and create virtual objects in the SDDC.

Table 146) Authorization and authentication management design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-015	Configure a service account (`svc-vra`) in vCenter Server for application-to-application communication from vRealize Automation with vSphere.	You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.
SDDC-CMP-016	Use local permissions when you create the `svc-vra` service account in vCenter Server.	The use of local permissions makes sure that only the Compute vCenter server instances are valid and accessible endpoints from vRealize Automation.	If you deploy more Compute vCenter server instances, you must make sure that the service account has been assigned local permissions in each vCenter server so that this vCenter server is a viable endpoint in vRealize Automation.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-017	Configure a service account (`svc-vra-vrops`) on vRealize Operations Manager for application-to-application communication from vRealize Automation for collecting health and resource metrics for tenant workload reclamation.	• vRealize Automation accesses vRealize Operations Manager with the minimum set of permissions that are required for collecting metrics to determine the workloads that are potential candidates for reclamation. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.

vRealize Automation Supporting Infrastructure

To satisfy the requirements of this SDDC design, configure additional components for vRealize Automation such as database servers to create a highly available database service and an email server for notification.

Microsoft SQL Server Database

vRealize Automation uses a Microsoft SQL Server database to store information about the vRealize Automation IaaS elements and the machines that vRealize Automation manages.

Table 147) vRealize Automation SQL Database design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-018	Set up a Microsoft SQL Server that supports the availability and I/O needs of vRealize Automation.	A dedicated or shared SQL Server can be used as long as it meets the requirements of vRealize Automation.	Requires additional resources and licenses.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-019	Locate the Microsoft SQL Server in the vRealize Automation virtual network or set it up to have global failover available.	For simple failover of the entire vRealize Automation instance from one region to another, the Microsoft SQL Server must be running as a virtual machine inside the vRealize Automation application virtual network. If the environment uses a shared SQL server, global failover provides connectivity from both primary and secondary regions.	Adds additional overhead to managing Microsoft SQL services.
SDDC-CMP-020	Set up Microsoft SQL Server with separate OS volumes for SQL Data, Transaction Logs, TempDB, and Backup.	Although each organization might have its own best practices in the deployment and configuration of Microsoft SQL Server, high-level best practices suggest separation of database data files and database transaction logs.	You might need to consult with the Microsoft SQL database administrators of your organization for guidance about production deployment in your environment.

Table 148) vRealize Automation SQL Database Server resource requirements per virtual machine.

Attribute	Specification
Number of vCPUs	8
Memory	16GB
Number of vNIC ports	1
Number of local drives	1 40GB (D:) (Application) 40GB (E:) Database data 20GB (F:) Database log 20GB (G:) TempDB 80GB (H:) Backup
vRealize Automation functions	Microsoft SQL Server Database
Microsoft SQL version	SQL Server 2012
Microsoft SQL Database version	SQL Server 2012 (110)
Operating system	Microsoft Windows Server 2012 R2

PostgreSQL Database Server

The vRealize Automation appliance uses a PostgreSQL database server to maintain the vRealize Automation portal elements and services and the information about the catalog items that the appliance manages. The PostgreSQL is also used to host data pertaining to the embedded instance of vRealize Orchestrator.

Table 149) vRealize Automation PostgreSQL database design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-021	Use the embedded PostgreSQL database in each vRealize Automation appliance. This database is also used by the embedded vRealize Orchestrator.	Simplifies the design and enables replication of the database across the two vRealize Automation appliances.	None
SDDC-CMP-022	Configure the embedded PostgreSQL database to use asynchronous replication.	Asynchronous replication offers a good balance between availability and performance.	Asynchronous replication provides a good level of availability in compliance with the design objectives.

Notification Email Server

vRealize Automation notification emails are sent using SMTP. These emails include notification of machine creation, expiration, and the notification of approvals received by users. vRealize Automation supports both anonymous connections to the SMTP server and connections using basic authentication. vRealize Automation also supports communication with or without SSL.

Create a global, inbound email server to handle inbound email notifications, such as approval responses. Only one global inbound email server, which appears as the default for all tenants, is needed. The email server provides accounts that you can customize for each user, providing separate email accounts, usernames, and passwords. Each tenant can override these settings. If tenant administrators do not override these settings before enabling notifications, vRealize Automation uses the globally configured email server. The server supports both the POP and the IMAP protocols, with or without SSL certificates.

Notifications

System administrators configure default settings for both the outbound and inbound email servers used to send system notifications. System administrators can create only one of each type of server that appears as the default for all tenants. If tenant administrators do not override these settings before enabling notifications, vRealize Automation uses the globally configured email server.

System administrators create a global outbound email server to process outbound email notifications and a global inbound email server to process inbound email notifications, such as responses to approvals.

Table 150) vRealize Automation email server configuration design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-023	Configure vRealize Automation to use a global outbound email server to handle outbound email notifications and a global inbound email server to handle inbound email notifications, such as approval responses.	Requirement to integrate vRealize Automation approvals and system notifications through emails.	You must prepare the SMTP/IMAP server and necessary firewall access and create a mailbox for inbound email (IMAP). Anonymous access can be used with outbound email.

vRealize Automation Cloud Tenant Design

A tenant is an organizational unit within a vRealize Automation deployment, and can represent a business unit within an enterprise or a company that subscribes to cloud services from a service provider. Each tenant has its own dedicated configuration, although some system-level configuration is shared across tenants.

Comparison of Single-Tenant and Multitenant Deployments

vRealize Automation supports deployments with a single tenant or multiple tenants. Systemwide configuration is always performed by using the default tenant, and this configuration can then be applied to one or more tenants. For example, systemwide configuration might specify defaults for branding and notification providers.

Infrastructure configuration, including the infrastructure sources that are available for provisioning, can be configured in any tenant and is shared among all tenants. The infrastructure resources, such as cloud or virtual compute resources or physical machines, can be divided into fabric groups managed by fabric administrators. The resources in each fabric group can be allocated to business groups within each tenant by using reservations.

- **Default-tenant deployment.** In a default-tenant deployment, all configuration occurs in the default tenant. Tenant administrators can manage users and groups and configure tenant-specific branding, notifications, business policies, and catalog offerings. All users log in to the vRealize Automation console at the same URL, but the features available to them are determined by their roles.
- **Single-tenant deployment.** In a single-tenant deployment, the system administrator creates a single new tenant for the organization that uses the same vRealize Automation instance. Tenant users log in to the vRealize Automation console at a URL that is specific to their tenant. Tenant-level configuration is segregated from the default tenant, although users with systemwide roles can view and manage both configurations. The IaaS administrator for the organization tenant creates fabric groups and appoints fabric administrators. Fabric administrators can create reservations for business groups in the organization tenant.
- **Multitenant deployment.** In a multitenant deployment, the system administrator creates new tenants for each organization that uses the same vRealize Automation instance. Tenant users log in to the vRealize Automation console at a URL that is specific to their tenant. Tenant-level configuration is segregated from other tenants and from the default tenant, although users with systemwide roles can view and manage configuration across multiple tenants. The IaaS administrator for each tenant creates fabric groups and appoints fabric administrators to their respective tenants. Although fabric administrators can create reservations for business groups in any tenant, in this scenario they typically create and manage reservations within their own tenants. If the same identity store is configured in multiple tenants, the same users can be designated as IaaS administrators or fabric administrators for each tenant.

Tenant Design

This design deploys a single tenant that contains two business groups:

- The first business group is designated for provisioning production workloads.
- The second business group is designated for development workloads.

Tenant administrators manage users and groups and configure tenant-specific branding, notifications, business policies, and catalog offerings. All users log in to the vRealize Automation console using the same URL, but the features available to them are determined by their roles.

Figure 70) Example Cloud Automation tenant design for two regions.

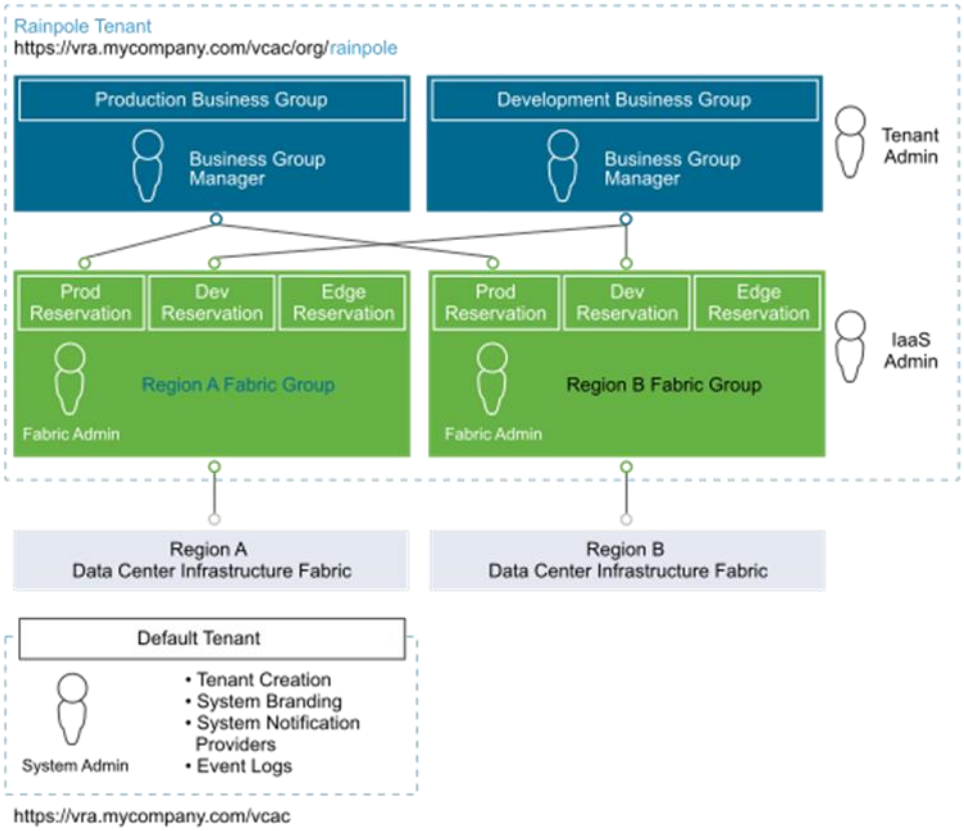


Table 151) Tenant design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-024	Uses vRealize Automation business groups for separate business units (instead of separate tenants).	Allows transparency across the environments and some level of sharing of resources and services such as blueprints.	Some elements, such as property groups, are visible to both business groups. The design does not provide full isolation for security or auditing.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-025	Create separate fabric groups for each deployment region. Each fabric group represent region-specific data center resources. Each of the business groups has reservations into each of the fabric groups.	Provides future isolation of fabric resources and potential delegation of duty to independent fabric administrators.	Initial deployment uses a single shared fabric that consists of one compute cluster.
SDDC-CMP-026	Allow access to the default tenant only by the system administrator for the purposes of managing tenants and modifying systemwide configurations.	Isolates the default tenant from individual tenant configurations.	Each tenant administrator is responsible for managing their own tenant configuration.
SDDC-CMP-027	Evaluate your internal organizational structure and workload needs. Configure business groups, reservations, service catalogs, and blueprints in the vRealize Automation instance based on your organization's needs.	vRealize Automation integrates with your organization's needs. In this design, guidance for Rainpole is provided as a starting point, but this guidance may not be appropriate for your specific business needs.	Partners and customers must evaluate their specific business needs.

Service Catalog

The service catalog provides a common interface for consumers of IT services to use to request and manage the services and resources they need.

A tenant administrator or service architect can specify information about the service catalog, such as the service hours, support team, and change window. Although the catalog does not enforce service-level agreements on services, service hours, the support team, and change window information are available to business users browsing the service catalog.

Catalog Items

Users can browse the service catalog for catalog items they are entitled to request. For some catalog items, a request results in the provisioning of an item that the user can manage. For example, the user can request a virtual machine with Windows 2012 preinstalled and then manage that virtual machine after it has been provisioned.

Tenant administrators define new catalog items and publish them to the service catalog. The tenant administrator can then manage the presentation of catalog items to the consumer and entitle new items to consumers. To make the catalog item available to users, a tenant administrator must entitle the item to the users and groups that should have access to it. For example, some catalog items might be available only to a specific business group, while other catalog items might be shared between business groups using the same tenant. The administrator determines what catalog items are available to different users based on their job function, department, or location.

Typically, a catalog item is defined in a blueprint, which provides a complete specification of the resource to be provisioned and the process to initiate when the item is requested. It also defines the options

available to a requester of the item. Options include virtual machine specifications, lease duration, or any additional information that the requester is prompted to provide when submitting the request.

Machine Blueprints

A machine blueprint is the complete specification for a virtual, cloud, or physical machine. A machine blueprint determines the machine's attributes, how it is provisioned, and its policy and management settings. Machine blueprints are published as catalog items in the service catalog.

Machine blueprints can be specific to a business group or shared among groups within a tenant. Tenant administrators can create shared blueprints that can be entitled to users in any business group within the tenant. Business group managers can create group blueprints that can only be entitled to users within a specific business group. A business group manager cannot modify or delete shared blueprints. Tenant administrators cannot view or modify group blueprints unless they also have the business group manager role for the appropriate group.

If a tenant administrator sets a shared blueprint's properties so that it can be copied, the business group manager can also copy the shared blueprint for use as a starting point to create a new group blueprint.

Table 152) Single-machine blueprints.

Name	Description
Base Windows Server (Development)	Standard Rainpole SOE deployment of Windows 2012 R2 available to the Development business group.
Base Windows Server (Production)	Standard Rainpole SOE deployment of Windows 2012 R2 available to the Production business group.
Base Linux Server (Development)	Standard Rainpole SOE deployment of Linux available to the Development business group.
Base Linux Server (Production)	Standard Rainpole SOE deployment of Linux available to the Production business group.
Windows Server + SQL Server (Production)	Base Windows 2012 R2 Server with silent SQL 2012 Server installation with custom properties. This server is available to the Production business group.
Windows Server + SQL Server (Development)	Base Windows 2012 R2 Server with silent SQL 2012 Server installation with custom properties. This server is available to the Development business group.

Blueprint Definitions

The following subsections provide details of each service definition that has been included as part of the current phase of cloud platform deployment.

Table 153) Base Windows Server requirements and standards.

Service Name	Base Windows Server
Provisioning method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter customizations.
Entitlement	Both Production and Development business group members.

Service Name	Base Windows Server
Approval process	No approval (preapproval assumed based on approved access to platform).
Operating system and version details	Windows Server 2012 R2
Configuration	Disk: single disk drive Network: standard vSphere networks
Lease and archival details	Lease: • Production blueprints: no expiration date • Development blueprints: minimum 30 days – maximum 270 days Archive: 15 days
Predevelopment and postdeployment requirements	Email sent to manager confirming service request (include description details)

Table 154) Base Windows Server blueprint sizing.

Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	4	60
Maximum	4	16	60

Table 155) Base Linux Server requirements and standards.

Service Name	Base Linux Server
Provisioning method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter customizations.
Entitlement	Both Production and Development business group members.
Approval process	No approval (preapproval assumed based on approved access to platform).
Operating system and version details	Red Hat Enterprise Server 6
Configuration	Disk: single disk drive Network: standard vSphere networks
Lease and archival details	Lease: • Production blueprints: no expiration date • Development blueprints: minimum 30 days – maximum 270 days Archive: 15 days
Predevelopment and postdeployment requirements	Email sent to manager confirming service request (include description details).

Table 156) Base Linux Server blueprint sizing.

Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	6	20
Maximum	4	12	20

Table 157) Base Windows Server with SQL Server installation requirements and standards.

Service Name	Base Windows Server
Provisioning method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter customizations.
Entitlement	Both Production and Development business group members.
Approval process	No approval (preapproval assumed based on approved access to platform).
Operating system and version details	Windows Server 2012 R2.
Configuration	Disk: single disk drive Network: standard vSphere networks Silent install: The blueprint calls a silent script using the vRealize Automation Agent to install SQL2012 Server with custom properties.
Lease and archival details	Lease: • Production blueprints: no expiration date • Development blueprints: minimum 30 days – maximum 270 days Archive: 15 days
Pre-development and post-deployment requirements	Email sent to manager confirming service request (include description details).

Table 158) Base Windows Server with SQL Server blueprint sizing.

Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	8	100
Maximum	4	16	400

Branding of the vRealize Automation Console

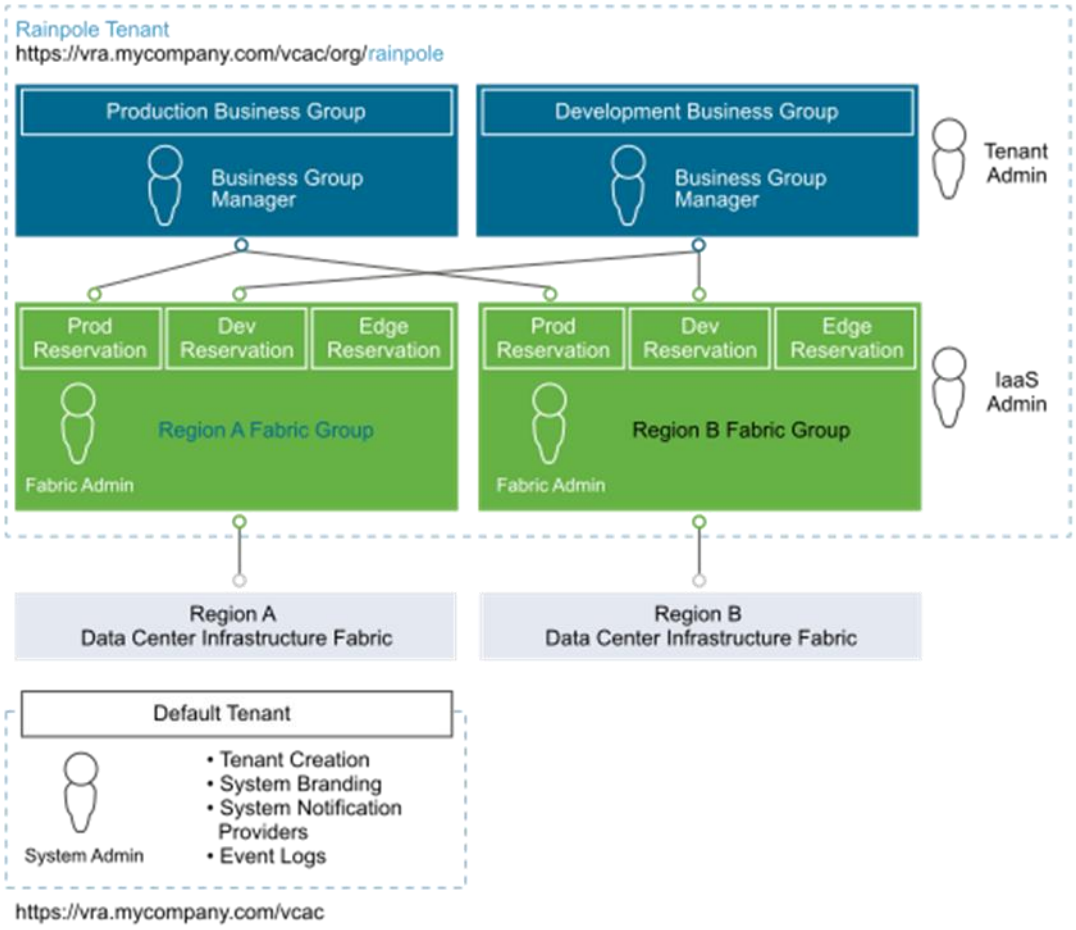
System administrators can change the appearance of the vRealize Automation console to meet site-specific branding guidelines by changing the logo, the background color, or information in the header and footer. System administrators control the default branding for tenants. Tenant administrators can use the default, or they can reconfigure branding for each tenant.

vRealize Automation Infrastructure as a Service Design

This topic introduces the integration of vRealize Automation with vSphere resources used to create an IaaS design for use with the SDDC.

Figure 71 illustrates the logical design of vRealize Automation groups and vSphere resources.

Figure 71) vRealize Automation logical design.



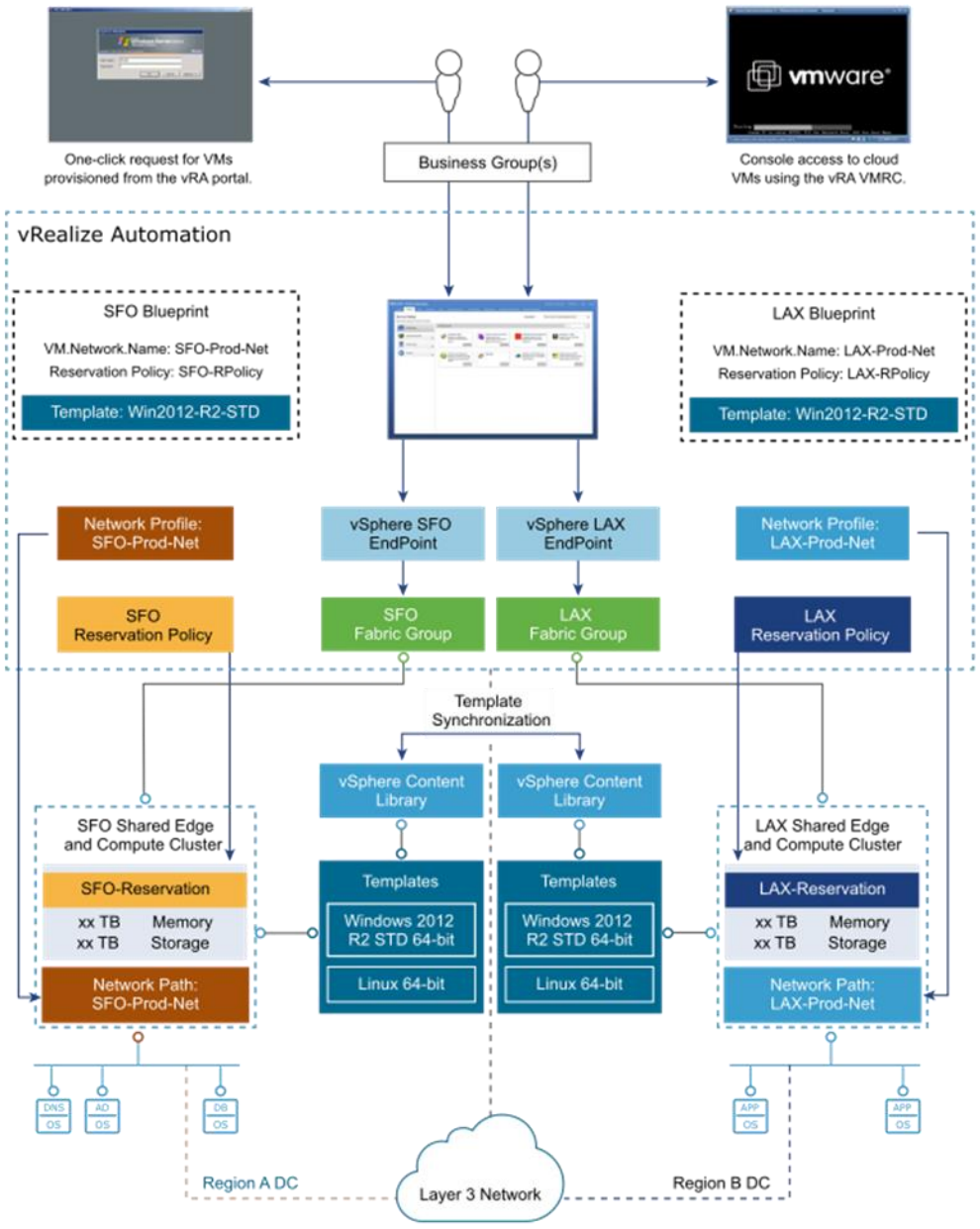
The terms defined in Table 159 apply to vRealize Automation when integrated with vSphere. These terms and their meaning might vary from the way they are used when referring only to vSphere.

Table 159) Definition of terms – vRealize Automation.

Term	Definition
vSphere (vCenter Server) endpoint	Provides information required by vRealize Automation IaaS to access vSphere compute resources.
Compute resource	Virtual object in vRealize Automation that represents a vCenter Server cluster or resource pool and datastores or datastore clusters. **Note:** Compute resources are CPU, memory, storage, and networks. Datastores and datastore clusters are part of the overall storage resources.
Fabric groups	vRealize Automation IaaS organizes compute resources into fabric groups.

Term	Definition
Fabric administrators	Fabric administrators manage compute resources, which are organized into fabric groups.
Compute reservation	A share of compute resources (vSphere cluster, resource pool, datastores, or datastore clusters), such as CPU and memory reserved for use by a particular business group for provisioning virtual machines. **Note:** vRealize Automation uses the term reservation to define resources (memory, storage, or networks) in a cluster. This is different from the use of reservation in vCenter Server, where a share is a percentage of total resources and reservation is a fixed amount.
Storage reservation	Similar to compute reservation (see above), but pertaining only to a share of the available storage resources. In this context, you specify a storage reservation in terms of gigabytes from an existing LUN or datastore.
Business groups	A collection of virtual machine consumers, usually corresponding to an organization's business units or departments. Only users in the business group can request virtual machines.
Reservation policy	vRealize Automation IaaS determines its reservation (also called virtual reservation), from which a particular virtual machine is provisioned. The reservation policy is a logical label or a pointer to the original reservation. Each virtual reservation can be added to one reservation policy.
Blueprint	The complete specification for a virtual machine, determining the machine attributes, the manner in which it is provisioned, and its policy and management settings. Blueprint allows the users of a business group to create virtual machines on a virtual reservation (compute resource) based on the reservation policy and using platform and cloning types. It also lets you specify or add machine resources and build profiles.

Figure 72) vRealize Automation integration with a vSphere endpoint.



Infrastructure Source Endpoints

An infrastructure source endpoint is a connection to the infrastructure that provides a set (or multiple sets) of resources that IaaS administrators can make available for consumption by end users. vRealize Automation IaaS regularly collects information about known endpoint resources and the virtual resources provisioned in them. Endpoint resources are referred to as compute resources or as compute pods; the terms are often used interchangeably.

Infrastructure data is collected through proxy agents that manage and communicate with the endpoint resources. This information about the compute resources on each infrastructure endpoint and the machines provisioned on each computer resource is collected at regular intervals.

During installation of the vRealize Automation IaaS components, you can configure the proxy agents and define their associated endpoints. Alternatively, you can configure the proxy agents and define their associated endpoints separately after primary vRealize Automation installation is complete.

Table 160) Endpoint design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-028	Create two vSphere endpoints.	One vSphere endpoint is required to connect to each vCenter Server instance in each region. Two endpoints are needed for two regions.	As additional regions are brought online, you must deploy additional vSphere endpoints.
SDDC-CMP-029	Create one vRealize Orchestrator endpoint that is configured to connect to the embedded vRealize Orchestrator instance.	vRealize Automation extensibility uses vRealize Orchestrator. The design includes one embedded vRealize Orchestrator cluster, which requires the creation of a single endpoint.	Requires configuration of a vRealize Orchestrator endpoint.
SDDC-CMP-030	Create one NSX endpoint and associate it with the vSphere endpoint.	The NSX endpoint is required to connect to NSX Manager and enable all the NSX-related operations supported in vRealize Automation blueprints.	None

Virtualization Compute Resources

A virtualization compute resource is a vRealize Automation object that represents an ESXi host or a cluster of ESXi hosts. When a group member requests a virtual machine, the virtual machine is provisioned on these compute resources. vRealize Automation regularly collects information about known compute resources and the virtual machines provisioned on them through the proxy agents.

Table 161) Compute resource design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-031	Create at least one compute resource for each deployed region.	Each region has one compute cluster, and one compute resource is required for each cluster.	As additional compute clusters are created, you must add them to the existing compute resources in their region or to a new resource, which must be created.
CSDDC-CMP-026	Assign the consolidated cluster as a compute resource in vRealize Automation.	This allows vRealize Automation to consume compute resources from the underlying virtual infrastructure.	None

Note: By default, compute resources are provisioned to the root of the compute cluster. In this design, use of vSphere resource pools is mandatory.

Fabric Groups

A fabric group is a logical container of several compute resources that can be managed by fabric administrators.

Table 162) Fabric group design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-032	Create a fabric group for each region and include all the compute resources and edge resources in that region.	To enable region-specific provisioning, you must create a fabric group in each region.	As additional clusters are added in a region, they must be added to the fabric group.

Business Groups

A business group is a collection of machine consumers that often corresponds to a line of business, a department, or another organizational unit. To request machines, a vRealize Automation user must belong to at least one business group. Each group has access to a set of local blueprints used to request machines.

Business groups have the following characteristics:

- A group must have at least one business group manager who maintains blueprints for the group and approves machine requests.
- Groups can contain support users who can request and manage machines on behalf of other group members.
- A vRealize Automation user can be a member of more than one business group and can have different roles in each group.

Reservations

A reservation is a share of one compute resource's available memory, CPU, and storage reserved for use by a particular fabric group. Each reservation is for one fabric group only, but the relationship is many-to-many. A fabric group might have multiple reservations on one compute resource, reservations on multiple compute resources, or both.

Converged Compute/Edge Clusters and Resource Pools

- Although reservations provide a method to allocate a portion of the cluster memory or storage in vRealize Automation, reservations do not control how CPU and memory are allocated during periods of contention on the underlying vSphere compute resources. vSphere resource pools are used to control the allocation of CPU and memory during times of resource contention on the underlying host. To fully use this feature, all VMs must be deployed into one of four resource pools: `sfo01-w01rp-sddc-edge` is dedicated to data center-level NSX Edge components and should not contain any user workloads.
- `sfo01-w01rp-sddc-mgmt` is dedicated to management VMs.
- `sfo01-w01rp-user-edge` is dedicated to any statically or dynamically deployed NSX components, such as NSX Edge gateways or load balancers that serve specific customer workloads.
- `sfo01-w01rp-user-vm` is dedicated to any statically or dynamically deployed virtual machines such as Windows, Linux, databases, and so on that contain specific customer workloads.

Table 163) Reservation design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-033	Create at least one vRealize Automation reservation for each business group at each region.	In the example, each resource cluster has two reservations, one for production and one for development, allowing both production and development workloads to be provisioned.	Because production and development share compute resources, the development business group must be limited to a fixed amount of resources.
SDDC-CMP-034	Create at least one vRealize Automation reservation for edge resources in each region.	An edge reservation in each region allows NSX to create edge services gateways on demand and place them on the edge cluster.	The workload reservation must define the edge reservation in the network settings.
SDDC-CMP-035	Configure all vRealize Automation workloads to use the `sfo01-w01rp-user-vm` resource pool.	To provide dedicated compute resources for NSX networking components, tenant-deployed workloads must be assigned to dedicated vSphere DRS resource pools. Workloads provisioned at the root resource-pool level receive more resources than those at child-resource pool levels. Such a configuration might starve those virtual machines in contention situations.	Cloud administrators must make sure that all workload reservations are configured with the appropriate resource pool. This may be a single resource pool for both production and development workloads or two resource pools, one dedicated for the Development business group and one dedicated for the Production business group.
SDDC-CMP-036	Configure vRealize Automation reservations for dynamically provisioned NSX Edge components (routed gateway) to use the `sfo01-w01rp-user-edge` resource pool.	To provide dedicated compute resources for NSX networking components, end-user-deployed NSX Edge components must be assigned to a dedicated end-user network component vCenter resource pool. Workloads provisioned at the root resource-pool level receive more resources than resource pools, which might starve those virtual machines in contention situations.	Cloud administrators must make sure that all workload reservations are configured with the appropriate resource pool.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-037	All vCenter resource pools used for Edge or Compute workloads must be created at the root level. Do not nest resource pools.	Nesting of resource pools can create administratively complex resource calculations that can result in unintended under or over allocation of resources during contention situations.	All resource pools must be created at the root resource pool level.

Reservation Policies

You can add each virtual reservation to one reservation policy. The reservation from which a particular virtual machine is provisioned is determined by vRealize Automation based on the reservation policy specified in the blueprint, if any, the priorities and current usage of the fabric group's reservations, and other custom properties.

Table 164) Reservation policy design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-038	Create at least one workload reservation policy for each region.	Reservation policies are used to target a deployment to a specific set of reservations in each region. Reservation policies are also used to target workloads into their appropriate region, compute cluster, and vSphere resource pool.	None
SDDC-CMP-039	Create at least one reservation policy for placement of dynamically created edge service gateways into the edge clusters.	Required to place the edge devices into their respective edge clusters and vSphere resource pools.	None

A storage reservation policy is a set of datastores that can be assigned to a machine blueprint to restrict disk provisioning to only those datastores. Storage reservation policies are created and associated with the appropriate datastores and assigned to reservations.

Table 165) Storage reservation policy design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-040	Storage tiers are not used in this design.	The underlying physical storage design does not use storage tiers.	Both business groups have access to the same storage. Customers using multiple datastores with different storage capabilities must evaluate the use of vRealize Automation storage reservation policies.

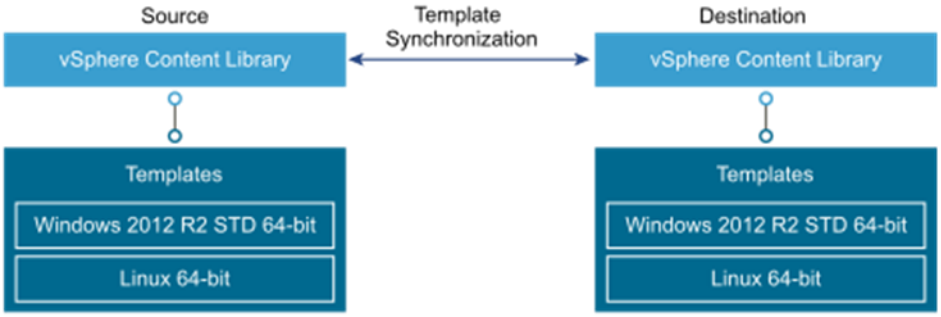
Template Synchronization

This dual-region design supports the provisioning of workloads across regions from the same portal using the same single-machine blueprints. A synchronization mechanism is required to have consistent templates across regions. There are multiple ways to achieve synchronization; for example, you can use the vSphere Content Library or external services like vCloud Connector or vSphere Replication.

Table 166) Template synchronization design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-041	This design uses the vSphere Content Library to synchronize templates across regions.	The vSphere Content Library is built into the version of vSphere being used and meets all the requirements to synchronize templates.	Storage space must be provisioned in each region. vRealize Automation cannot directly consume templates from the vSphere Content Library.

Figure 73) Template synchronization.



VMware Identity Management

VMware Identity Manager is integrated into the vRealize Automation appliance and provides tenant identity management.

The VMware Identity Manager synchronizes with the Rainpole Active Directory domain. Important users and groups are synchronized with VMware Identity Manager. Authentication uses the Active Directory domain, but searches are made against the local Active Directory mirror on the vRealize Automation appliance.

Figure 74) VMware Identity Manager proxies authentication between Active Directory and vRealize Automation.

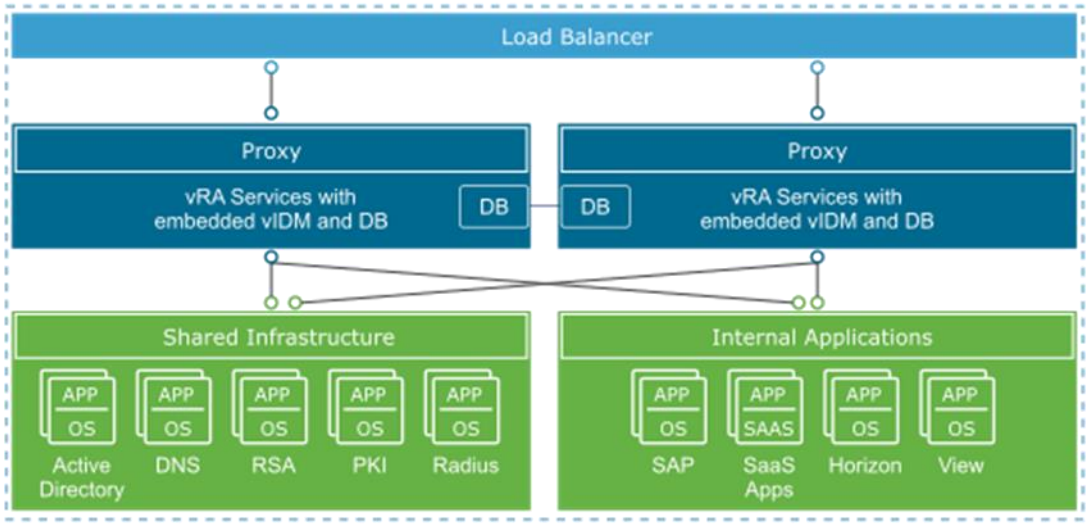


Table 167) Active Directory authentication decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-042	Choose Active Directory with Integrated Windows Authentication as the Directory Service connection option.	Rainpole uses a single-forest, multiple-domain Active Directory environment. Integrated Windows Authentication supports establishing trust relationships in a multidomain or multiforest Active Directory environment.	Requires that the vRealize Automation appliances are joined to the Active Directory domain.

By default, the vRealize Automation appliance is configured with 18GB of memory, which is enough to support a small Active Directory environment. An Active Directory environment is considered small if fewer than 25,000 users in the organizational unit must be synchronized. An Active Directory environment with more than 25,000 users is considered large and requires additional memory and CPU resources. For more information on sizing your vRealize Automation deployment, see the vRealize Automation documentation.

The connector is a component of the vRealize Automation service; it synchronizes users and groups between Active Directory and the vRealize Automation service. In addition, the connector is the default identity provider and authenticates users to the service.

Table 168) Connector configuration design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-043	To support Directories Service high availability, configure a second connector that corresponds to the second vRealize Automation appliance.	This design supports high availability by installing two vRealize Automation appliances and using load-balanced NSX Edge instances. Adding the second connector to the second vRealize Automation appliance provides redundancy and improves performance by load balancing authentication requests.	This design simplifies the deployment while using robust built-in HA capabilities. This design uses NSX for vSphere load balancing.

12.2 vRealize Business for Cloud Design

vRealize Business for Cloud provides end-user transparency for costs associated with operating workloads. A system such as vRealize Business gathers and aggregates the financial costs of workload operations. Such a system provides greater visibility both during a workload request and on a periodic basis. Visibility is improved regardless of whether the costs are "charged-back" to a specific business unit or are "showed-back" to illustrate the value that the SDDC provides.

vRealize Business integrates with vRealize Automation to display costing during workload requests and on an ongoing basis with costs reporting by the user, business group, or tenant. Additionally, tenant administrators can create a wide range of custom reports to meet the requirements of an organization.

Table 169) vRealize Business for Cloud Standard edition design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-044	Deploy vRealize Business for Cloud as part of the CMP and integrate it with vRealize Automation.	Tenant and workload costing is provided by vRealize Business for Cloud.	You must deploy more appliances for vRealize Business for Cloud and for remote collectors.
SDDC-CMP-045	Use the default vRealize Business for Cloud appliance size (8GB). For a vRealize Business for Cloud remote collector, use a reduced memory size of 2GB.	The default vRealize Business for Cloud appliance size supports up to 10,000 VMs. Remote collectors do not run a server service and can run on 2GB of RAM.	None
SDDC-CMP-046	Use the default vRealize Business reference costing database.	Default reference costing is based on industry information and is periodically updated.	Default reference costing might not accurately represent actual customer costs. vRealize Business Appliance requires internet access to periodically update the reference database.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-047	Deploy vRealize Business as a three-virtual-machine architecture with remote data collectors in Region A and Region B.	For best performance, the vRealize Business collectors should be regionally local to the resource that they are configured to collect. Because this design supports disaster recovery, the CMP can reside in Region A or Region B.	In a case in which the environment does not have disaster recovery support, you must deploy an additional appliance for the remote data collector. However, the vRealize Business server can handle the load on its own.
SDDC-CMP-048	Deploy the vRealize Business server virtual machine in the cross-region logical network.	vRealize Business deployment depends on vRealize Automation. During a disaster recovery event, vRealize Business migrates with vRealize Automation.	None
SDDC-CMP-049	Deploy a vRealize Business remote data collector virtual machine in each region-specific logical network.	vRealize Business remote data collector is a region-specific installation. During a disaster recovery event, the remote collector does not need to migrate with vRealize Automation.	The communication with vCenter Server involves an additional layer 3 hop through an NSX Edge device.

Table 170) vRealize Business for Cloud virtual appliance resource requirements per virtual machine.

Attribute	Specification
Number of vCPUs	4
Memory	8GB for a server and 2GB for a remote collector
vRealize Business function	Server or remote collector

12.3 vRealize Orchestrator Design

VMware vRealize Orchestrator is a development and process automation platform that provides a library of extensible workflows. This platform allows you to create and run automated, configurable processes to manage the VMware vSphere infrastructure as well as other VMware and third-party technologies.

In this VVD, vRealize Automation uses the vRealize Orchestrator plug-in to connect to vCenter Server for customized virtual machine provisioning and postprovisioning actions.

vRealize Orchestrator Logical Design

This VVD uses the vRealize Orchestrator instance that is embedded in the vRealize Automation appliance instead of using a dedicated or external vRealize Orchestrator instance.

Table 171) vRealize Orchestrator hardware design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-01	Use the internal vRealize Orchestrator instances that are embedded in the deployed vRealize Automation instances.	• The use of an embedded vRealize Orchestrator has the following advantages: • Provides faster time to value • Reduces the number of appliances to manage • Provides an easier upgrade path and better supportability • Improves performance • Removes the need for an external database	Overall simplification of the design leading to a reduced number of appliances and enhanced supportability.

vRealize Orchestrator Authentication

The embedded vRealize Orchestrator only supports vRealize Automation Authentication as the authentication method.

Table 172) vRealize Orchestrator directory service design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-02	Embedded vRealize Orchestrator uses vRealize Automation Authentication.	The only authentication method available.	None
SDDC-CMP-VRO-03	Configure vRealize Orchestrator to use the vRealize Automation customer tenant (Rainpole) for authentication.	The vRealize Automation default tenant users are only administrative users. By connecting to the customer tenant, workflows running on vRealize Orchestrator can run with end-user-granted permissions.	End users who run vRealize Orchestrator workflows are required to have permissions on the vRealize Orchestrator server. Some plug-ins might not function correctly using vRealize Automation Authentication.
SDDC-CMP-VRO-04	Each vRealize Orchestrator instance is associated with only one customer tenant.	To provide the best security and segregation between potential tenants, one vRealize Orchestrator instance is associated with a single tenant.	If additional vRealize Automation tenants are configured, additional vRealize Orchestrator installations are needed.

Network Ports

vRealize Orchestrator uses specific network ports to communicate with other systems. The ports are configured with a default value, but you can change the defaults at any time. When you make changes, verify that all ports are available for use by your host. If necessary, open these ports on any firewalls

through which network traffic for the relevant components flows. Verify that the required network ports are open before you deploy vRealize Orchestrator.

Default Communication Ports

Set default network ports and configure your firewall to allow incoming TCP connections. Other ports might be required if you are using custom plug-ins.

Table 173) vRealize Orchestrator default configuration ports.

Port	Number	Protocol	Source	Target	Description
HTTPS server port	443	TCP	End-user web browser	Embedded vRealize Orchestrator server	The SSL-secured HTTP protocol used to connect to the vRealize Orchestrator REST API.
Web configuration HTTPS access port	8283	TCP	End-user web browser	vRealize Orchestrator configuration	The SSL access port for the control center Web UI for vRealize Orchestrator configuration

External Communication Ports

Configure your firewall to allow outgoing connections using the external network ports so that vRealize Orchestrator can communicate with external services.

Table 174) vRealize Orchestrator default external communication ports.

Port	Number	Protocol	Source	Target	Description
LDAP	389	TCP	vRealize Orchestrator server	LDAP server	Lookup port for your LDAP authentication server.
LDAP using SSL	636	TCP	vRealize Orchestrator server	LDAP server	Lookup port for your secure LDAP authentication server.
LDAP using Global Catalog	3268	TCP	vRealize Orchestrator server	Global Catalog server	Port to which Microsoft Global Catalog server queries are directed.
DNS	53	TCP	vRealize Orchestrator server	DNS server	Name resolution.

Port	Number	Protocol	Source	Target	Description
VMware vCenter Single Sign-On server	7444	TCP	vRealize Orchestrator server	vCenter Single Sign-On server	Port used to communicate with the vCenter Single Sign-On server.
SQL Server	1433	TCP	vRealize Orchestrator server	Microsoft SQL Server	Port used to communicate with the Microsoft SQL Server or SQL Server Express instances that are configured as the vRealize Orchestrator database.
PostgreSQL	5432	TCP	vRealize Orchestrator server	PostgreSQL Server	Port used to communicate with the PostgreSQL Server that is configured as the vRealize Orchestrator database.
Oracle	1521	TCP	vRealize Orchestrator server	Oracle DB server	Port used to communicate with the Oracle Database Server that is configured as the vRealize Orchestrator database.
SMTP Server port	25	TCP	vRealize Orchestrator server	SMTP server	Port used for email notifications.
vCenter Server API port	443	TCP	vRealize Orchestrator server	VMware vCenter server	The vCenter Server API communication port used by vRealize Orchestrator to obtain virtual infrastructure and virtual machine information from the orchestrated vCenter Server instances.

Port	Number	Protocol	Source	Target	Description
vCenter Server	80	TCP	vRealize Orchestrator server	vCenter Server	Port used to tunnel HTTPS communication.
VMware ESXi	443	TCP	vRealize Orchestrator server	ESXi hosts	(Optional) Workflows using the vCenter Guest Operations API need direct connection between vRealize Orchestrator and the ESXi hosts the virtual machine is running on.

vRealize Orchestrator Server Mode

vRealize Orchestrator supports standalone mode and cluster mode. This design uses cluster mode.

vRealize Orchestrator supports the following server modes:

- **Standalone mode.** vRealize Orchestrator server runs as a standalone instance. This is the default mode of operation.
- **Cluster mode.** To increase availability of the vRealize Orchestrator services and to create a more highly available SDDC, you can configure vRealize Orchestrator to work in cluster mode. You can also start multiple vRealize Orchestrator instances in a cluster with a shared database. In cluster mode, multiple vRealize Orchestrator instances with identical server and plug-in configurations work together as a cluster and share a single database. When you cluster the vRealize Automation appliances, the vRealize Orchestrator instances embedded in them are automatically clustered.

All vRealize Orchestrator server instances communicate with each other by exchanging heartbeats at a certain time interval. Only active vRealize Orchestrator server instances respond to client requests and run workflows. If an active vRealize Orchestrator server instance fails to send heartbeats, it is considered to be nonresponsive, and one of the inactive instances takes over to resume all workflows from the point at which they were interrupted. The heartbeat is implemented through the shared database, so there are no implications in the network design for a vRealize Orchestrator cluster. If you have more than one active vRealize Orchestrator node in a cluster, concurrency problems can occur if different users use the different vRealize Orchestrator nodes to modify the same resource.

vRealize Orchestrator Load-Balancer Configuration

Configure load balancing for the vRealize Orchestrator instances embedded in the two vRealize Automation instances to provision network access to the vRealize Orchestrator control center.

Table 175) vRealize Orchestrator SDDC cluster design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-05	Configure the load balancer to permit network access to the embedded vRealize Orchestrator control center.	The control center allows customization of vRealize Orchestrator, such as changing the tenant configuration and certificates.	None

vRealize Orchestrator Information Security and Access Control

Use a service account for authentication and authorization of vRealize Orchestrator to vCenter Server for orchestrating and creating virtual objects in the SDDC.

Table 176) Authorization and authentication management design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-06	Configure a service account (`svc-vro`) in vCenter Server for application-to-application communication from vRealize Orchestrator with vSphere.	Introduces improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.
SDDC-CMP-VRO-07	Use local permissions when you create the `svc-vro` service account in vCenter Server.	The use of local permissions makes sure that only the Compute vCenter Server instances are valid and accessible endpoints from vRealize Orchestrator.	If you deploy more Compute vCenter Server instances, you must makes sure that the service account has been assigned local permissions in each vCenter Server so that this vCenter Server is a viable endpoint in vRealize Orchestrator.

vRealize Orchestrator Configuration

vRealize Orchestrator configuration includes guidance on client configuration, database configuration, SSL certificates, and plug-ins.

vRealize Orchestrator Client

The vRealize Orchestrator client is a desktop application that lets you import packages; create, run, and schedule workflows; and manage user permissions.

You can install the standalone version of the vRealize Orchestrator Client on a desktop system. Download the vRealize Orchestrator Client installation files from the vRealize Orchestrator appliance page at `https://vRA_hostname/vco`. Alternatively, you can run the vRealize Orchestrator Client using Java WebStart directly from the homepage of the vRealize Automation appliance console.

SSL Certificates

The vRealize Orchestrator configuration interface uses a secure connection to communicate with vCenter Server, relational database management systems, LDAP, vCenter Single Sign-On, and other servers.

You can import the required SSL certificate from a URL or a file. You can import the vCenter Server SSL certificate from the SSL Trust Manager tab in the vRealize Orchestrator configuration interface.

Table 177) vRealize Orchestrator SSL design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-08	The embedded vRealize Orchestrator instance uses the vRealize Automation appliance certificate.	Using the vRealize Automation certificate simplifies the configuration of the embedded vRealize Orchestrator instance.	None

vRealize Orchestrator Database

vRealize Orchestrator requires a database. This design uses the PostgreSQL database embedded in the vRealize Automation appliance.

Table 178) vRealize Orchestrator database design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-09	The embedded vRealize Orchestrator instance uses the PostgreSQL database embedded in the vRealize Automation appliance.	Using the embedded PostgreSQL database provides the following advantages: • Performance improvement • Design simplification	None

vRealize Orchestrator Plug-Ins

Plug-ins allow you to use vRealize Orchestrator to access and control external technologies and applications. Exposing an external technology in a vRealize Orchestrator plug-in allows you to incorporate objects and functions in workflows that access the objects and functions of the external technology. The external technologies that you can access using plug-ins can include virtualization management tools, email systems, databases, directory services, and remote-control interfaces. vRealize Orchestrator provides a set of standard plug-ins that allow you to incorporate technologies such as the vCenter Server API and email capabilities into workflows.

In addition, the vRealize Orchestrator open plug-in architecture allows you to develop plug-ins to access other applications. vRealize Orchestrator implements open standards to simplify integration with external systems. For information on developing custom content, see Developing with VMware vRealize Orchestrator.

vRealize Orchestrator and the vCenter Server Plug-In

You can use the vCenter Server plug-in to manage multiple vCenter Server instances. You can create workflows that use the vCenter Server plug-in API to automate tasks in your vCenter Server environment. The vCenter Server plug-in maps the vCenter Server API to JavaScript that you can use in workflows. The plug-in also provides actions that perform individual vCenter Server tasks that you can include in workflows.

The vCenter Server plug-in provides a library of standard workflows that automate vCenter Server operations. For example, you can run workflows that create, clone, migrate, or delete virtual machines. Before managing the objects in your VMware vSphere inventory with vRealize Orchestrator and running workflows on these objects, you must configure the vCenter Server plug-in. You must also define the

connection parameters between vRealize Orchestrator and the vCenter Server instances you want to orchestrate. You can configure the vCenter Server plug-in by using the vRealize Orchestrator configuration interface or by running the vCenter Server configuration workflows from the vRealize Orchestrator client. You can configure vRealize Orchestrator to connect to your vCenter Server instances for running workflows over the objects in your vSphere infrastructure.

To manage objects in your vSphere inventory by using vSphere Web Client, configure vRealize Orchestrator to work with the same vCenter Single Sign-On instance to which both vCenter Server and vSphere Web Client are pointing. Also, verify that vRealize Orchestrator is registered as a vCenter Server extension. Register vRealize Orchestrator as a vCenter Server extension when you specify a user (user name and password) who has privileges to manage vCenter Server extensions.

Table 179) vRealize Orchestrator vCenter Server plug-in design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-10	Configure the vCenter Server plug-in to control communication with the vCenter Servers.	Required for communication with vCenter Server instances and also for workflows.	None

vRealize Orchestrator Scalability

vRealize Orchestrator supports both system scale up and scale out.

Scale Up

A single vRealize Orchestrator instance allows up to 300 concurrent workflow instances in the running state. Workflow instances that are in the waiting or waiting-event state do not count toward this number. You can design long-running workflows that preserve resources by using the wait elements of the workflow palette. A single vRealize Orchestrator instance supports up to 35,000 managed virtual machines in its inventory. You can increase the memory and vCPU resources for the vRealize Automation appliance virtual machines to enable the scaling up of vRealize Orchestrator. For more information on increasing the memory allocated for the embedded vRealize Orchestrator to take advantage of the scaled up vRealize Automation appliance, see VMware Knowledge Base article 2147109.

Scale Out

In the current design, you can scale out vRealize Orchestrator by using a cluster of vRealize appliances that have the embedded vRealize Orchestrator appropriately configured using the same settings. Using a vRealize Orchestrator cluster allows you to increase the number of concurrent running workflows, but not the number of managed inventory objects. When clustering a vRealize Orchestrator server, choose the following cluster type:

- An active-active cluster with up to five active nodes. VMware recommends a maximum of three active nodes in this configuration.

In a clustered vRealize Orchestrator environment, you cannot change workflows while other vRealize Orchestrator instances are running. Stop all other vRealize Orchestrator instances before you connect the vRealize Orchestrator client and change or develop a new workflow.

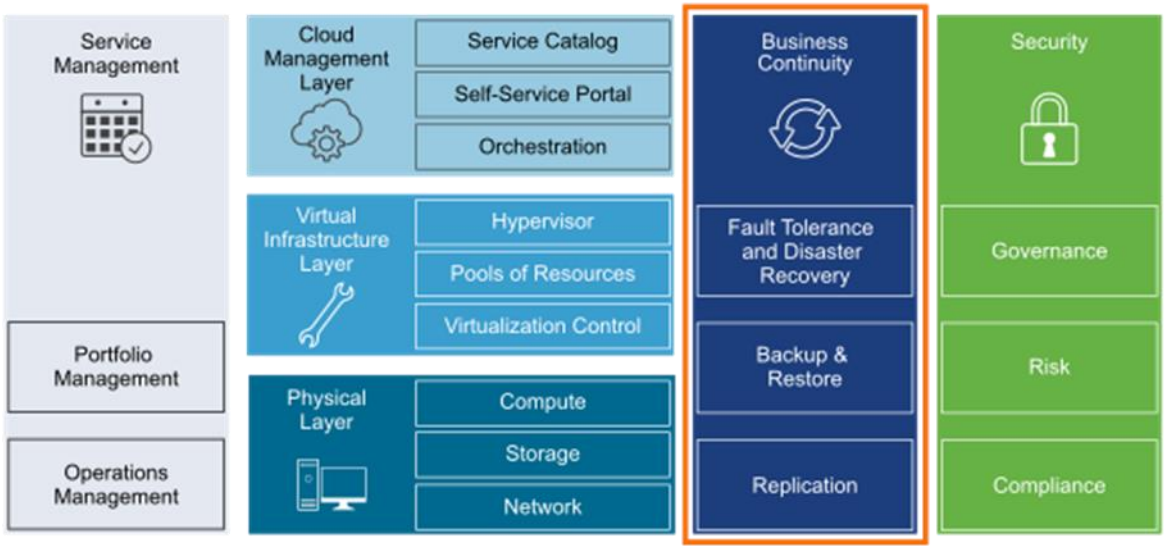
Table 180) vRealize Orchestrator scale-out design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-11	Configure vRealize Orchestrator in an active-active cluster configuration.	Active-active clusters allow both vRealize Orchestrator servers to equally balance workflow execution.	When you cluster the vRealize Automation appliances, the vRealize Orchestrator instances embedded in them are automatically clustered.

13 Business Continuity Design

Design for business continuity includes solutions for data protection and disaster recovery of critical management components of the SDDC. The design provides guidance on the main elements of a product design such as deployment, sizing, networking, diagnostics, and security.

Figure 75) Business continuity in the SDDC layered architecture.



13.1 Data Protection and Backup Design

For continuous operation of the SDDC if the data from a management application is compromised, you should design data protection of the management components in your environment.

Backup protects the data of your organization against loss due to hardware failure, accidental deletion, or other faults for each region. For consistent image-level backups, use backup software that is based on the VADP. You can use any VADP-compatible backup solution. Adapt and apply the design decisions to the backup software you use.

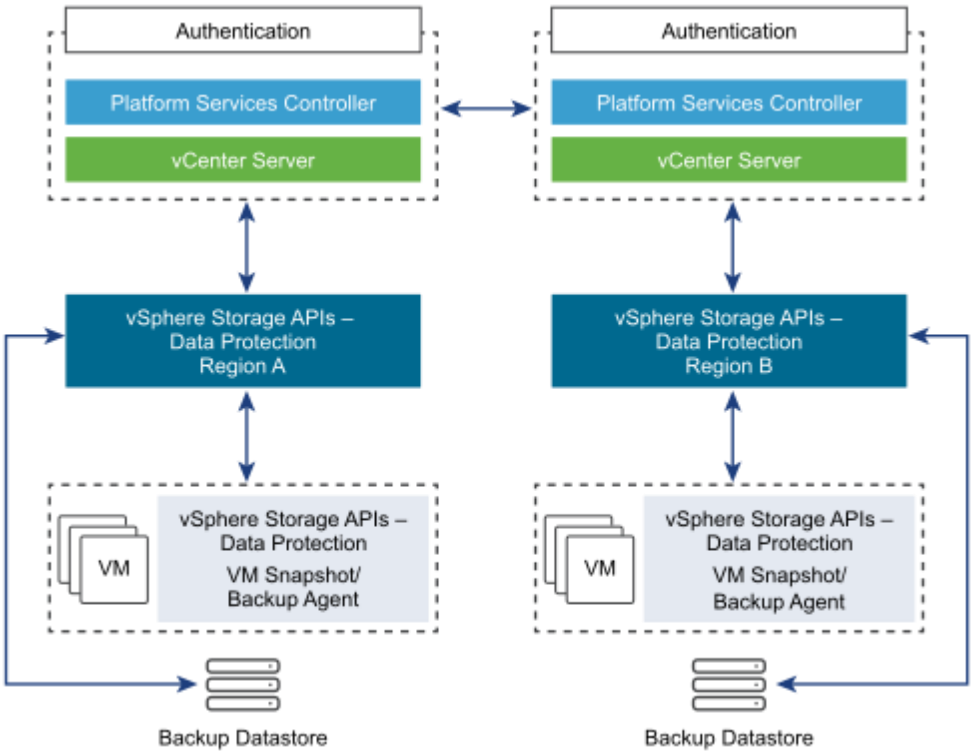
Table 181) VADP-compatible backup solution design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-001	Use a backup solution that is compatible with VADP and can perform image-level backups of the management components.	You can back up and restore most of the management components at the virtual machine image level.	None
SDDC-OPS-BKP-002	Use a VADP-compatible backup solution that can perform application-level backups of the management components.	Microsoft SQL Server requires application awareness when performing backup and restore procedures.	You must install application-aware agents on the virtual machine of the management component.

Logical Design for Data Protection

VADP-compatible backup solutions protect the virtual infrastructure at the VMware vCenter Server layer. Because the VADP-compatible backup solution is connected to the Management vCenter Server, it can access all management ESXi hosts and can detect the virtual machines that require backups.

Figure 76) Data protection logical design.



Backup Datastore for Data Protection

The backup datastore stores all data that is required to recover services according to a recovery point objective (RPO). Determine the target location based on the performance requirements.

VADP-compatible backup solutions can use deduplication technology to back up virtual environments at the data-block level for efficient disk utilization. To optimize backups and use the VMware vSphere Storage APIs, all ESXi hosts must have access to the production storage.

To back up the management components of the SDDC, size your secondary storage appropriately. You must provide 6TB of capacity without considering deduplication capabilities.

Table 182) Backup datastore design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-003	Allocate a dedicated datastore for the VADP-compatible backup solution and the backup data according to NFS physical storage design.	• Emergency restore operations are possible even when the primary datastore is not available because the VADP-compatible backup solution storage volume is separate from the primary datastore. • The amount of storage required for backups is greater than the amount of storage available in the datastore.	You must provide additional capacity by using a storage array.
SDDC-OPS-BKP-004	Provide secondary storage with a capacity of 6TB on-disk.	Secondary storage handles the backup of the management stack of a single region. The management stack consumes approximately 6TB of disk space, uncompressed and without deduplication.	You must provide more secondary storage capacity to accommodate increased disk requirements.

Backup Policies for Data Protection

Backup policies specify virtual machine backup options, the schedule window, and retention policies in this validated design.

Options for Virtual Machine Backup

VADP provides the following options for a virtual machine backup:

- **Network block device (NBD).** Transfers virtual machine data across the network so that the VADP-compatible solution can perform the backups.
  - The performance of the virtual machine network traffic might be lower.
  - NBD takes a quiesced snapshot. As a result, it might interrupt the I/O operations of the virtual machine to swap the `.vmdk` file or consolidate the data after the backup is complete.
  - The time to complete the virtual machine backup might be longer than the backup window.
  - NBD does not work in multiwriter disk mode.
- **Protection agent inside guest OS.** Provides backup of certain applications that are running in the guest operating system by using an installed backup agent.

- Enables application-consistent backup and recovery with Microsoft SQL Server, Microsoft SharePoint, and Microsoft Exchange support.
- Provides more granularity and flexibility to restore on the file level.

Table 183) Virtual machine transport mode design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-005	Use HotAdd to back up virtual machines.	HotAdd optimizes and speeds up virtual machine backups and does not affect the vSphere management network.	All ESXi hosts must have the same visibility for the virtual machine datastores.
SDDC-OPS-BKP-006	Use the VADP solution agent for backups of the Microsoft SQL Server.	You can restore application data instead of entire virtual machines.	You must install and maintain the VADP solution agent.

Schedule Window

Even though VADP uses changed-block-tracking technology to optimize the backup of data, to avoid any business impact, do not use a backup window when the production storage is in high demand.

Table 184) Backup schedule design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-007	Schedule daily backups.	You can recover virtual machine data that is at most a day old.	You lose data that has changed since the last backup 24 hours ago.
SDDC-OPS-BKP-008	Schedule backups outside the production peak times.	Backups occur when the system is under the lowest load. Make sure that backups are completed in the shortest time possible with the smallest risk of errors.	You must schedule backup to start between 8:00 PM and 8:00 AM or until the backup jobs are complete, whichever comes first.

Retention Policies

Retention policies are properties of a backup job. If you group virtual machines by business priority, you can set the retention requirements according to the business priority.

Table 185) Backup retention policies design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-009	Retain backups for at least 3 days.	Keeping 3 days of backups enables administrators to restore the management applications to a state within the last 72 hours.	Depending on the rate of change in virtual machines, backup retention policy can increase the storage target size.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-010	Retain backups for cross-region replicated backup jobs for at least 1 day.	Keeping 1 day of a backup for replicated jobs enables administrators, in the event of a disaster recovery situation in which failover was unsuccessful, to restore their region-independent applications to a state within the last 24 hours.	You lose data that has changed since the last backup 24 hours ago. This data loss also increases the storage requirements for the backup solution in a multiregion configuration.

Information Security and Access Control for Data Protection

You use a service account for authentication and authorization of a VADP-compatible backup solution for backup and restore operations.

Table 186) Authorization and authentication management for a VADP-compatible solution design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-011	Configure a service account (`svc-bck-vcenter`) in vCenter Server for application-to-application communication from the VADP-compatible backup solution with vSphere.	Provides the following access control features: • Provide the VADP-compatible backup solution with a minimum set of permissions that are required to perform backup and restore operations. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.
SDDC-OPS-BKP-012	Use global permissions when you create the `svc-bck-vcenter` service account in vCenter Server.	• Simplifies and standardizes the deployment of the service account across all vCenter Server instances in the same vSphere domain. • Provides a consistent authorization layer.	All vCenter Server instances must be in the same vSphere domain.

Component Backup Jobs for Data Protection

You can configure backup for each SDDC management component separately. This design does not suggest a requirement to back up the entire SDDC. Some products can perform internal configuration backups. Use those products in addition to the whole virtual machine component backups as appropriate.

Table 187) Component backup jobs design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-013	Use the internal configuration backup of NSX for vSphere.	Restoring small configuration files can be a faster and less destructive method to achieve a similar restoration of functionality.	You must provide space on an SFT or FTP server to store the NSX configuration backups.

13.2 Site Recovery Manager and vSphere Replication Design

To support disaster recovery in the SDDC, you must protect vRealize Operations Manager and vRealize Automation by using VMware Site Recovery Manager and VMware vSphere Replication. When failing over to a recovery region, these management applications continue to support operations management and CMP functionality.

This SDDC disaster recovery design includes two locations: Region A and Region B.

- **Protected Region A in San Francisco.** Region A contains the protected virtual workloads of the management stack. It is referred to as the protected region in this document.
- **Recovery Region B in Los Angeles.** Region B provides an environment to host virtual machines from the protected region if a disaster occurs. It is referred to as the recovery region.

Site Recovery Manager can automate the setup and execution of disaster recovery plans between these two regions.

Note: A region in the VVD is equivalent to the site construct in Site Recovery Manager.

Logical Design for Site Recovery Manager and vSphere Replication

Critical SDDC management applications and services must be available in case of disaster. These management applications are running as virtual machines and can have dependencies on applications and services that run in both regions.

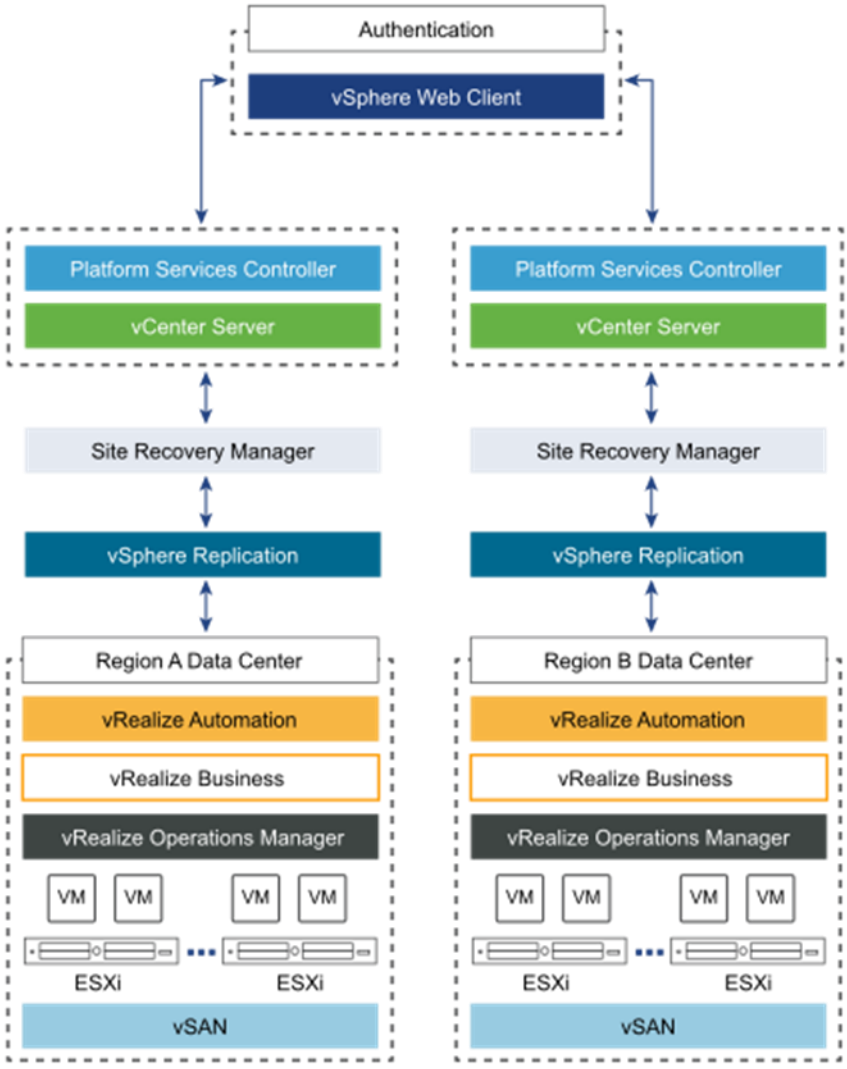
This validated design for disaster recovery defines the logical configuration of the SDDC management applications shown in Table 188.

Table 188) Logical configuration for disaster recovery in the SDDC.

Management Component	Logical Configuration for Disaster Recovery
Regions and ESXi hosts	• Region A has a management cluster of ESXi hosts that runs the virtual machines of the management application that must be protected. • Region B has a management cluster of ESXi hosts with sufficient free capacity to host the protected management applications from Region A.
vCenter Server	Each region has a vCenter Server instance for the management ESXi hosts in the region.

Management Component	Logical Configuration for Disaster Recovery
Site Recovery Manager	• Each region has a Site Recovery Manager server with an embedded database. • In each region, Site Recovery Manager is integrated with the Management vCenter Server instance.
vSphere Replication	• vSphere Replication provides hypervisor-based virtual machine replication between Region A and Region B. • vSphere Replication replicates data from Region A to Region B by using a dedicated VMkernel TCP/IP stack.

Figure 77) Disaster recovery logical design.



Deployment Design for Site Recovery Manager

A separate Site Recovery Manager instance is required for the protection and recovery of management components in the event of a disaster situation with your SDDC.

Install and configure Site Recovery Manager after you install and configure vCenter Server and Platform Services Controller in the region.

You have the following options for deployment and pairing of vCenter Server and Site Recovery Manager:

- vCenter Server options:
  - You can use Site Recovery Manager and vSphere Replication with a vCenter Server appliance or with vCenter Server for Windows.
  - You can deploy a vCenter Server appliance in one region and a vCenter Server for Windows instance in the other region.
- Site Recovery Manager options:
  - You can use either a physical system or a virtual system.
  - You can deploy Site Recovery Manager on a shared system, such as the system of vCenter Server for Windows, or on a dedicated system.

Table 189) Site Recovery Manager and vSphere replication deployment design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-001	Deploy Site Recovery Manager in a dedicated virtual machine.	All components of the SDDC solution must support the highest levels of availability. When Site Recovery Manager runs as a virtual machine, you can enable the availability capabilities of vCenter Server clusters.	Requires a Microsoft Windows server license.
SDDC-OPS-DR-002	Deploy each Site Recovery Manager instance in the management cluster.	All management components must be in the same cluster.	None
SDDC-OPS-DR-003	Deploy each Site Recovery Manager instance with an embedded PostgreSQL database.	• Reduces the dependence on external components. • Reduces potential database licensing costs.	You must assign database administrators who have the skills and tools to administer PostgreSQL databases.

Sizing Compute Resources for Site Recovery Manager

To support the orchestrated failover of the SDDC management components according to the objectives of this design, you must size the host operating system on which the Site Recovery Manager software runs.

Table 190) Compute resources for a Site Recovery Manager node.

Attribute	Specification
Number of vCPUs	2 (running at 2.0GHz or higher)
Memory	4GB
Number of virtual machine NIC ports	1
Number of disks	1

Attribute	Specification
Disk size	40GB
Operating system	Windows Server 2012 R2

Sizing is usually performed according to IT organization requirements. However, this design uses calculations that are based on the management components in a single region. The design then mirrors the calculations for the other region. Consider the management node configuration for each region shown in Table 191.

Table 191) SDDC nodes with failover support.

Management Component	Node Type	Number of Nodes
CMP	vRealize Automation appliance	2
	vRealize IaaS web server	2
	vRealize IaaS management server	2
	vRealize IaaS DEM	2
	Microsoft SQL Server	1
	vRealize Business for Cloud appliance	1
vRealize Operations Manager	vRealize Operations Manager master	1
	vRealize Operations Manager master replica	1
	vRealize Operations Manager data	1

You must protect a total of 13 virtual machines.

Use vSphere Replication as the replication solution between the Site Recovery Manager sites, and distribute the virtual machines in two protection groups.

Table 192) Compute resources for the Site Recovery Manager nodes design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-004	Deploy Site Recovery Manager on a Microsoft Windows Server guest OS according to the following specifications: • 2 vCPUs • 4GB memory • 40GB disk • 1GbE	Accommodates the protection of management components to supply the highest levels of availability. This size further accommodates the following setup: • The number of protected management virtual machines as defined in Table 190 • Two protection groups • Two recovery plans	You must increase the size of the nodes if you add more protection groups or virtual machines to protect or recover plans.

Placeholder Virtual Machines

Site Recovery Manager creates a placeholder virtual machine on the recovery region for every machine from the Site Recovery Manager protection group. Placeholder virtual machine files are small because they contain virtual machine configuration metadata but no virtual machine disks. Site Recovery Manager adds the placeholder virtual machines as recovery region objects to the vCenter Server inventory.

Networking Design for Site Recovery Manager and vSphere Replication

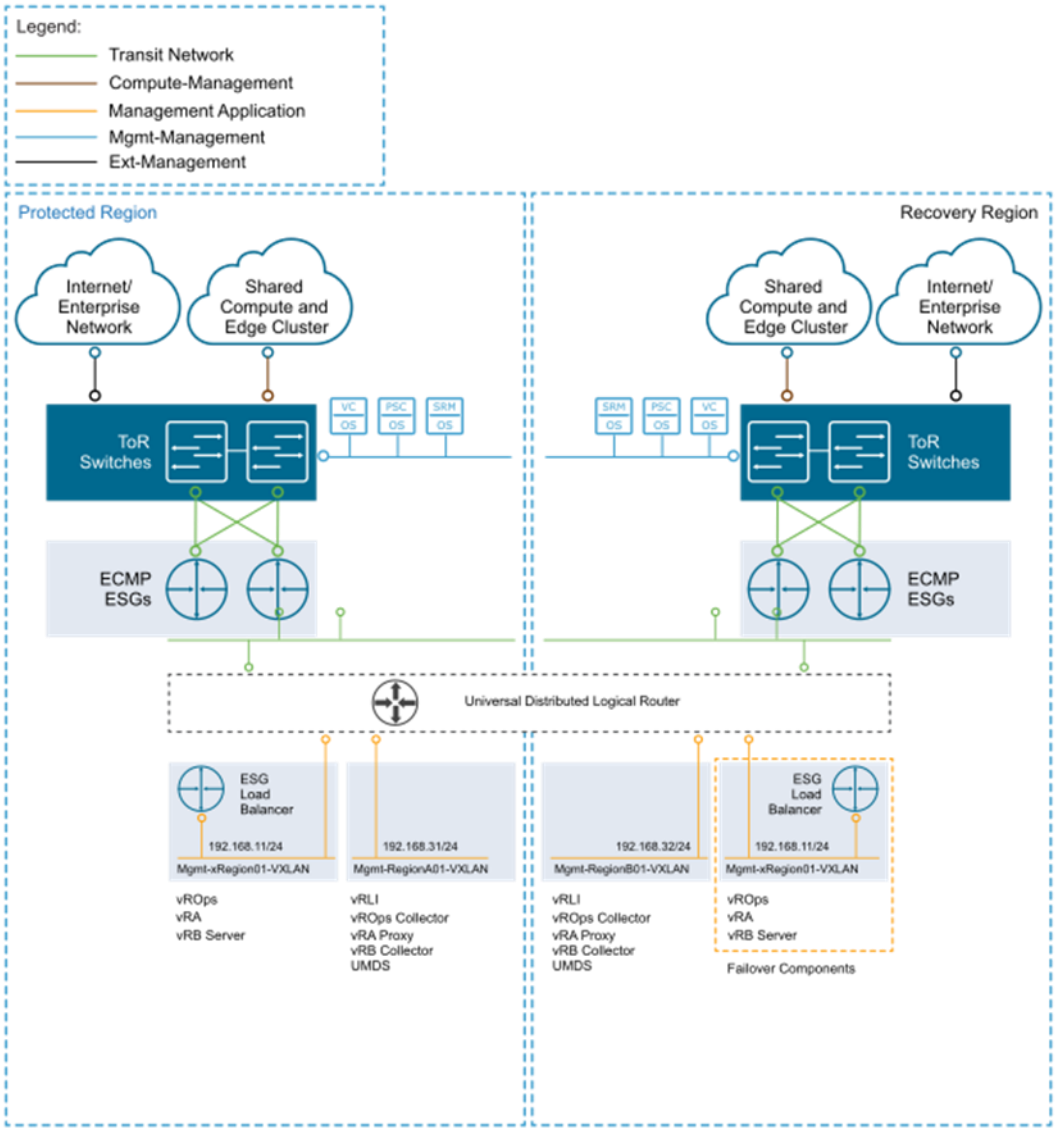
Moving an application physically from one region to another represents a networking challenge, especially if applications have hard-coded IP addresses. According to the requirements for the network address space and IP address assignment, use either the same or a different IP address at the recovery region. In many situations, you assign new IP addresses because VLANs might not stretch between regions.

This design uses NSX for vSphere to create virtual networks called application virtual networks. In application virtual networks, you can place workloads by using a single IP network address space that spans across data centers. Application virtual networks have the following benefits:

- Single IP network address space providing mobility between data centers
- Simplified disaster recovery procedures

After a failover, the recovered application is available under the same IPv4 address.

Figure 78) Logical network design for cross-region deployment with application virtual networks.



The IPv4 subnets (orange networks) are routed in the vSphere management network of each region. Nodes on these network segments are reachable from within the SDDC. IPv4 subnets, such as the subnet that contains the vRealize Automation primary components, overlap across a region.

Make sure that only the active IPv4 subnet is propagated in the region and beyond. The public-facing Ext-Management network of both regions (gray networks) is reachable by SDDC users and provides connection to external resources, such as Active Directory or DNS. For information about the design of application virtual networks, see “Application Virtual Network” in section 10.7.

NSX Edge devices provide load-balancing functionality, with each device fronting a network that contains the protected components of all management applications. In each region, you use the same configuration for the management applications and their Site Recovery Manager shadow. Active Directory and DNS services must be running in both the protected region and the recovery region.

Information Security and Access Control for Site Recovery Manager and vSphere Replication

You use a service account for authentication and authorization of Site Recovery Manager to vCenter Server for orchestrated disaster recovery of the SDDC.

Table 193) Authorization and authentication management for Site Recovery Manager and vSphere Replication design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-005	Configure a service account (`svc-srm`) in vCenter Server for application-to-application communication from Site Recovery Manager with vSphere.	Provides the following access control features: • Site Recovery Manager accesses vSphere with the minimum set of permissions that are required to perform disaster recovery failover orchestration and site pairing. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.
SDDC-OPS-DR-006	Configure a service account (`svc-vr`) in vCenter Server for application-to-application communication from vSphere Replication with vSphere	Provides the following access control features: • vSphere Replication accesses vSphere with the minimum set of permissions that are required to perform site-to-site replication of virtual machines. • In the event of a compromised account, accessibility in the destination application remains restricted. • You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's lifecycle outside of the SDDC stack to preserve its availability.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-007	Use global permissions when you create the `svc-srm` service account in vCenter Server.	• Simplifies and standardizes the deployment of the service account across all vCenter Server instances in the same vSphere domain. • Provides a consistent authorization layer. • If you deploy more Site Recovery Manager instances, reduces the effort in connecting them to the vCenter Server instances.	All vCenter Server instances must be in the same vSphere domain.
SDDC-OPS-DR-008	Use global permissions when you create the `svc-vr` service account in vCenter Server.	• Simplifies and standardizes the deployment of the service account across all vCenter Server instances in the same vSphere domain. • Provides a consistent authorization layer. • If you deploy more vSphere Replication instances, reduces the effort in connecting them to the vCenter Server instances.	All vCenter Server instances must be in the same vSphere domain.

Encryption

Replace default self-signed certificates with a CA-signed certificate to provide secure access and communication for vSphere Replication and Site Recovery Manager.

Table 194) CA-signed certificates for Site Recovery Manager and vSphere Replication design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-009	Replace the default self-signed certificates in each Site Recovery Manager instance with a CA-signed certificate.	Make sure that all communication to the externally facing Web UI of Site Recovery Manager and cross-product communication are encrypted.	Replacing the default certificates with trusted CA-signed certificates complicates installation and configuration.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-010	Replace the default self-signed certificates in each vSphere Replication instance with a CA-signed certificate.	Makes sure that all communications to the externally facing web UI for vSphere Replication and cross-product communication are encrypted.	Replacing the default certificates with trusted CA-signed certificates complicates installation and configuration.

Deployment Design for vSphere Replication

Deploy vSphere Replication for virtual machine replication in Site Recovery Manager. Consider the requirements for the operation of the management components that are failed over.

Replication Technology

You have the following options for replication technology when using Site Recovery Manager:

- Array-based replication through storage replication adapters with Site Recovery Manager
- vSphere Replication with Site Recovery Manager.

Table 195) Replication technology design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-011	Use vSphere Replication in Site Recovery Manager as the protection method for virtual machine replication.	• Allows flexibility in storage usage and vendor selection between the two disaster recovery regions. • Minimizes administrative overhead required to maintain storage replication adapter compatibility between the two regions of disaster recovery.	• All management components must be in the same cluster. • The total number of virtual machines configured for protection using vSphere Replication is reduced compared with the use of storage-based replication.

Networking Configuration of the vSphere Replication Appliances

vSphere Replication uses a VMkernel management interface on the ESXi host to send replication traffic to the vSphere Replication appliance in the recovery region. In this VVD, the vSphere Replication traffic has been isolated to its own dedicated port group and VLAN for each region to make sure that there is no effect on other vSphere management traffic. For more information about vSphere Replication traffic on the management ESXi hosts, see section 10.6, “Virtualization Network Design.”

vSphere Replication appliances and vSphere Replication servers are the target for the replication traffic that originates from the vSphere Replication VMkernel ports.

Table 196) vSphere Replication networking design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-012	Dedicate a distributed port group to vSphere Replication traffic.	Makes sure that vSphere Replication traffic does not affect other vSphere management traffic. The vSphere Replication servers potentially receive large amounts of data from the VMkernel adapters on the ESXi hosts.	You must allocate a dedicated VLAN for vSphere Replication.
SDDC-OPS-DR-013	Dedicate a distributed port group to vSphere Replication traffic	• VLANs provide traffic isolation. • Limits extraneous networking utilization across the ISL.	• Static routes on the ESXi hosts are required. • A sufficient number of VLANs are available in each cluster and should be used for traffic segregation. • Host profiles must be managed on a per-host level.
SDDC-OPS-DR-014	Dedicate a VMkernel adapter on the management ESXi hosts.	Makes sure that the ESXi server replication traffic is redirected to the dedicated vSphere Replication VLAN.	None
SDDC-OPS-DR-015	Attach a vNIC from the vSphere Replication VMs to the vSphere Replication port group.	Makes sure that the vSphere Replication VMs can communicate on the correct replication VLAN.	vSphere Replication VMs might require additional network adapters for communication on the management and replication VLANs.

Snapshot Space During Failover Tests

To perform failover tests, you must provide additional storage for the snapshots of the replicated VMs. This storage is minimal in the beginning but grows as test VMs write to their disks. Replication from the protected region to the recovery region continues during this time. The snapshots that are created during testing are deleted after the failover test is complete.

Sizing Resources for vSphere Replication

Select a size for the vSphere Replication nodes to facilitate virtual machine replication of the SDDC management components according to the objectives of this design.

Table 197) Compute resources for a vSphere Replication four vCPU node.

Attribute	Specification
Number of vCPUs	4
Memory	4GB

Attribute	Specification
Disk capacity	18
Environment	Up to 2,000 replications between nodes

Sizing is performed according to IT organization requirements. However, this design uses calculations for a single region. The design then mirrors the calculations for the other region. You must protect a total of 13 virtual machines. For information about the node configuration of the management components for each region that is used in the calculations, see Table 190.

Table 198) vSphere Replication deployment and size design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-017	Deploy each vSphere Replication appliance in the management cluster.	All management components must be in the same cluster.	None
SDDC-OPS-DR-018	Deploy each vSphere Replication appliance using the four vCPU size.	Accommodates the replication of the expected number of virtual machines that are a part of the following components: • vRealize Automation • vRealize Operations Manager	None

sMessages and Commands for Site Recovery Manager

You can configure Site Recovery Manager to present messages for notification and accept acknowledgement to users. Site Recovery Manager also provides a mechanism to run commands and scripts as necessary when running a recovery plan.

You can insert pre-power-on or post-power-on messages and commands in the recovery plans. These messages and commands are not specific to Site Recovery Manager. However, they pause the execution of the recovery plan to complete other procedures or running customer-specific commands or scripts to enable automation of recovery tasks.

Site Recovery Manager Messages

Some additional steps might be required before, during, and after running a recovery plan. For example, you might set up the environment so that a message appears when a recovery plan is initiated. In response, the administrator must acknowledge the message before the recovery plan continues. Messages are specific to each IT organization.

Consider the following sample messages and confirmation steps:

- Verify that IP address changes are made on the DNS server and that the changes are propagated.
- Verify that Active Directory services are available.
- After the management applications are recovered, perform application tests to verify that the applications are functioning correctly.

Additionally, confirmation steps can be inserted after every group of services that depend on other services. These conformations can be used to pause the recovery plan so that appropriate verification and testing be performed before subsequent steps are taken. These services are defined as follows:

- Infrastructure services
- Core services
- Database services
- Middleware services
- Application services
- Web services

Details about each message are specified in the workflow definition of the individual recovery plan.

Site Recovery Manager Commands

You can run custom scripts to perform infrastructure configuration updates or configuration changes on the environment of a virtual machine. The scripts that a recovery plan runs are located on the Site Recovery Manager server. The scripts can run against the Site Recovery Manager server or they can affect a virtual machine.

If a script must run on the virtual machine, Site Recovery Manager does not run it directly, but rather instructs the virtual machine to do it. The audit trail that Site Recovery Manager provides does not record the execution of the script, because the operation is on the target virtual machine.

Scripts or commands must be available in the path on the virtual machine according to the following guidelines:

- Use full paths to all executables. For example, `c:\windows\system32\cmd.exe` instead of `cmd.exe`.
- Call only `.exe` or `.com` files from the scripts. Command-line scripts can only call executables.
- To run a batch file, start the shell command with `c:\windows\system32\cmd.exe`.

The scripts that are run after powering on a virtual machine are executed under the Windows Servers Local Security Authority of the Site Recovery Manager server. Store post-power-on scripts on the Site Recovery Manager virtual machine. Do not store such scripts on a remote network share.

Recovery Plan for Site Recovery Manager and vSphere Replication

A recovery plan is the automated plan (runbook) for full or partial failover from Region A to Region B.

Recovery Time Objective

The RTO is the targeted duration of time and service level in which a business process must be restored as a result of an IT service or data loss issue, such as a natural disaster.

Table 199) Site Recovery Manager design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-019	Use Site Recovery Manager and vSphere Replication together to automate the recovery of the following management components: • vRealize Operations analytics cluster • vRealize Automation appliance instances • vRealize Automation IaaS components • vRealize Business Server	• Provides an automated run book for the recovery of the management components in the event of a disaster. • Makes sure that the recovery of management applications can be delivered in an RTO of 4 hours or less.	None

Replication and Recovery Configuration between Regions

You configure virtual machines in the Management vCenter Server in Region A to replicate to the Management vCenter Server in Region B in such a way that, in the event of a disaster in Region A, you have redundant copies of your virtual machines. During the configuration of replication between the two vCenter Server instances, the following options are available:

- **Guest OS quiescing.** Quiescing a virtual machine just before replication improves the reliability of recovering the virtual machine and its applications. However, any solution, including vSphere Replication, that quiesces an operating system and application might affect performance. For example, such an effect could appear in virtual machines that generate higher levels of I/O and where quiescing occurs often.
- **Network compression.** Network compression can be defined for each virtual machine to further reduce the amount of data transmitted between source and target locations.
- **Recovery point objective.** The RPO is configured on each virtual machine. The RPO defines the maximum acceptable age that the data stored and recovered in the replicated copy (replica) as a result of an IT service or data loss issue, such as a natural disaster, can have. The lower the RPO, the closer the replica's data is to the original. However, lower RPO requires more bandwidth between source and target locations and more storage capacity in the target location.
- **Point-in-time instance.** You define multiple recovery points (point-in-time instances, or PIT instances) for each virtual machine so that, when a virtual machine experiences data corruption, data integrity, or host OS infections, administrators can recover and revert to a recovery point before the compromising issue occurred.

Table 200) vSphere Replication design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-020	Do not enable guest OS quiescing in the policies for the management virtual machines in vSphere Replication.	Not all management virtual machines support the use of guest OS quiescing. Using the quiescing operation might result in an outage.	The replicas of the management virtual machines that are stored in the target region are crash-consistent rather than application-consistent.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-021	Enable network compression on the management virtual machine policies in vSphere Replication.	• Makes sure that the vSphere Replication traffic over the network has a reduced footprint. • Reduces the amount of buffer memory used on the vSphere Replication VMs.	To perform compression and decompression of data, a vSphere Replication virtual machine might require more CPU resources on the source site as more virtual machines are protected.
SDDC-OPS-DR-022	Enable an RPO of 15 minutes on the management virtual machine policies in vSphere Replication.	• Makes sure that the management application that is failing over after a disaster recovery event contains all data except any changes prior to 15 minutes of the event. • Achieves the availability and recovery target of 99% of this VVD.	Any changes that are made up to 15 minutes before a disaster recovery event are lost.
SDDC-OPS-DR-023	Enable point-in-time instances, keeping three copies over a 24-hour period on the management virtual machine policies in vSphere Replication.	Preserves application integrity for the management application that is failing over after a disaster recovery event occurs.	Increasing the number of retained recovery point instances increases the disk usage on the primary datastore.

Startup Order and Response Time

Virtual machine priority determines the virtual machine startup order.

- All priority 1 virtual machines are started before priority 2 virtual machines.
- All priority 2 virtual machines are started before priority 3 virtual machines.
- All priority 3 virtual machines are started before priority 4 virtual machines.
- All priority 4 virtual machines are started before priority 5 virtual machines.
- You can additionally set startup order of virtual machines within each priority group.

You can configure the following timeout parameters:

- Response time, which defines the time to wait after the first virtual machine powers on before proceeding to the next virtual machine in the plan.
- Maximum time to wait if the virtual machine fails to power on before proceeding to the next virtual machine.

You can adjust response time values as necessary during execution of the recovery plan test to determine the appropriate response time values.

Table 201) Site Recovery Manager startup order design decisions.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-024	Use a prioritized startup order for vRealize Operations Manager nodes.	• Makes sure that the individual nodes in the vRealize Operations Manager analytics cluster are started in such an order that the operational monitoring services are restored after a disaster. • Makes sure that the vRealize Operations Manager services are restored in the target of 4 hours.	• You must have VMware Tools running on each vRealize Operations Manager node. • If you increase the number of analytics nodes in the vRealize Operations Manager cluster, you must maintain the customized recovery plan.
SDDC-OPS-DR-025	Use a prioritized startup order for vRealize Automation and vRealize Business nodes.	• Makes sure that the individual nodes in vRealize Automation and vRealize Business are started in such an order that cloud provisioning and cost management services are restored after a disaster. • Makes sure that the vRealize Automation and vRealize Business services are restored within the target of 4 hours.	• You must have VMware Tools installed and running on each vRealize Automation and vRealize Business node. • If you increase the number of nodes in vRealize Automation, you must maintain the customized recovery plan.

Recovery Plan Test Network

When you create a recovery plan, you must configure test network options as follows:

- **Isolated network.** Created automatically. For a virtual machine that is being recovered, Site Recovery Manager creates an isolated private network on each ESXi host in the cluster. Site Recovery Manager creates a standard switch and a port group on it.

  A limitation of this automatic configuration is that a virtual machine that is connected to the isolated port group on one ESXi host cannot communicate with a virtual machine on another ESXi host. This option limits testing scenarios and provides an isolated test network only for basic virtual machine testing.
- **Port group.** Selecting an existing port group provides a more granular configuration to meet your testing requirements. If you want virtual machines across ESXi hosts to communicate, use a standard or distributed switch with uplinks to the production network, and create a port group on the switch that has tagging with a nonroutable VLAN enabled. In this way you isolate the network, because it is not connected to other production networks.

Because the application virtual networks for failover are fronted by a load balancer, as a recovery plan test network, you can use the recovery plan production network to provide realistic verification of a recovered management application.

Table 202) Recovery plan test network design decision.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-026	Use the target recovery production network for testing.	The design of the application virtual networks supports their use as recovery plan test networks.	During recovery testing, a management application is not reachable by using its production FQDN. Access the application by using its VIP address or assign a temporary FQDN for testing. Note that this approach results in certificate warnings because the assigned temporary host name and the host name in the certificate do not match.

Where to Find Additional Information

To learn more about the information that is described in this document, review to the following documents and/or websites:

- NetApp SolidFire Quality of Service
  https://www.netapp.com/us/media/tr-4644.pdf

Version History

Version	Date	Document Version History
Version 1.1	May 2019	Update to the Introduction section.
Version 1.0	October 2018	Initial release

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright Information

Trademark Information

NVA-1128-DESIGN-0519