Architecture and Design

Modified on 19 MAR 2019

VMware Validated Design 5.0

VMware Validated Design for Software-Defined Data Center 5.0

vmware®

You can find the most up-to-date technical documentation on the VMware website at:

https://docs.vmware.com/

Contents

About VMware Validated Design Architecture and Design

The *VMware Validated Design Architecture and Design* document contains a validated model of the Software-Defined Data Center (SDDC) and provides a detailed design of each management component of the SDDC stack.

Chapter 1 Architecture Overview discusses the building blocks and the main principles of each SDDC management layer. Chapter 2 Detailed Design provides the available design options according to the design objective, and a set of design decisions to justify selecting the path for building each SDDC component.

Intended Audience

VMware Validated Design Architecture and Design is intended for cloud architects, infrastructure administrators and cloud administrators who are familiar with and want to use VMware software to deploy in a short time and manage an SDDC that meets the requirements for capacity, scalability, backup and restore, and extensibility for disaster recovery support.

Required VMware Software

VMware Validated Design Architecture and Design is compliant and validated with certain product versions. See *VMware Validated Design Release Notes* for more information about supported product versions.

Before You Apply This Guidance

The sequence of the documentation of VMware Validated Design follows the stages for implementing and maintaining an SDDC. See Documentation Map for VMware Validated Design.

To use *VMware Validated Design Architecture and Design*, you must be acquainted with *Introducing VMware Validated Designs*.

Updated Information

This *VMware Validated Design Architecture and Design* document is updated with each release of the product or when necessary.

This table provides the update history of the *Architecture and Design* document.

Revision	Description
24 AUG 2020	At VMware, we value inclusion. To foster this principle within our customer, partner, and internal community, we are replacing some of the terminology in our content. We have updated this guide to remove instances of non-inclusive language.
12 FEB 2019	The VLAN of the management network across availability zones is now defined as native and stretched. See Switch Types and Network Connectivity .
22 JAN 2019	Initial release.

Architecture Overview

VMware Validated Design for Software-Defined Data Center (SDDC) enables an IT organization to automate the provisioning of common repeatable requests and to respond to business needs with agility and predictability. Usually, this use case is referred to as Infrastructure as a Service (IaaS). However, VMware Validated Design for Software-Defined Data Center extends the typical IaaS solution to provide an IT solution with features across many areas such as IaaS, operations management, business continuity and security.

The VMware Validated Design architecture is based on layers and modules. You can replace components to implement the end solution or outcome such as the SDDC. If a particular component design does not fit a business or technical requirement for some reason, you can swap it out for a similar one.

A VMware Validated Design is one way of assembling an architecture. It is tested for stability, scalability and compatibility. The design of the system ensures achieving best IT outcomes.

Figure 1-1. Architecture Overview

Physical Layer

The lowest layer of the solution is the physical layer which consists of the compute, network and storage components. The compute component contains the x86-based servers that run the management, edge and tenant workloads. This design provides some guidance about the physical capabilities that are required to run this architecture. However, you select a specific type or brand of hardware according to *VMware Compatibility Guide*.

Virtual Infrastructure Layer

The virtual infrastructure layer is on top of the physical layer components. The virtual infrastructure layer controls the 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, together with elements in the cloud management, service management, business continuity and security layers.

Cloud Management Layer

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

This layer requests resources and orchestrates the actions of the lower layers from a user interface or application programming interface (API).

Service Management Layer

When building any type of IT infrastructure, you use portfolio and operations management for continuous day-to-day service delivery. The service management area of this architecture focuses on operations management, in particular lifecycle management, monitoring, alerting, and log management.

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.

In the operations management layer, you monitor the underlying physical infrastructure and the virtual management and tenant workloads in real-time. Information is collected in the form of structured data (metrics) and unstructured data (logs). The operations management layer also retrieves the SDDC topology, that is 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.

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 of business critical data. This design provides 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.

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 the operations and setup that you must provide to implement an SDDC that is resilient to both internal and external threats.

This chapter includes the following topics:

- Physical Infrastructure Architecture
- Virtual Infrastructure Architecture
- Operations Management Architecture
- Cloud Management Architecture
- Business Continuity Architecture

Physical Infrastructure Architecture

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

Figure 1-2. Physical Infrastructure Design

Workload Domain Architecture

VMware Validated Design for Software-Defined Data Center uses a set of building blocks called workload domains. A workload domain consists of a set of ESXi hosts that are managed by one vCenter Server instance and network equipment for connection to the data center.

Workload Domain Architecture Characteristics

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

A workload domain represents a logical boundary of functionality, managed by a single vCenter Server instance. The workload domain is not defined by any hard physical properties. Although a workload domain usually spans one rack, you can aggregate multiple workload domains in a single rack in smaller setups. For both small and large setups, consider homogeneity and easy replication .

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

Workload Domain to Rack Mapping

Workload domains are not mapped one-to-one to data center racks. While a workload domain is an atomic unit of repeatable building blocks, a rack is a unit of size. Because workload domains can have different sizes, you map workload domains to data center racks according to 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 the 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 that has more ESXi hosts than a single rack can support.

Stretched Workload Domain Across Availability Zones

A cluster in a single workload domain can span across two availability zones by using vSAN stretched clustering.

Cluster Types

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

Figure 1-3. Clusters in a Single Availability Zone in the SDDC

Figure 1-4. Clusters in Two Availability Zones 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 and 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 provides non-vSAN storage using NFS, iSCSI or Fiber Channel. Different types of storage can provide different levels of SLA, ranging from just a bunch of disks (JBODs) using IDE drives with minimal to no redundancy, to fully redundant enterprise-class storage arrays.

Physical Network Architecture

VMware Validated Design for Software-Defined Data Center can use most physical network architectures. In an SDDC, you consider Layer 2 or Layer 3 transport, using quality of service tags for prioritized traffic handling on the network devices, NIC configuration on the physical servers, and VLAN port modes on both physical servers and network equipment.

Network Transport

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

VMware Validated Design supports both Layer 2 and Layer 3 transports. When you decide whether to use Layer 2 or Layer 3, consider the following factors:

- NSX ECMP Edge devices establish Layer 3 routing adjacency with the first upstream Layer 3 device to provide equal cost routing for management and workload traffic.

- The investment you have today in your current physical network infrastructure.

- The benefits and drawbacks for both Layer 2 and Layer 3 designs.

Benefits and Drawbacks of Layer 2 Transport

A design using Layer 2 transport has these considerations:

- In a design that uses Layer 2 transport, top of rack switches and upstream Layer 3 devices, such as core switches or routers, form a switched fabric.

- The upstream Layer 3 devices terminsates each VLAN and provide default gateway functionality.

- Uplinks from the top of rack switch to the upstream Layer 3 devices are 802.1Q trunks carrying all required VLANs.

Using a Layer 2 transport has the following benefits and drawbacks:

Table 1-1. 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 1-5. Example Layer 2 Transport

Benefits and Drawbacks of Layer 3 Transport

A design using Layer 3 transport has these considerations:

- Layer 2 connectivity is limited in the data center rack up to the top of rack switches.

- The top of rack switch terminates each VLAN and provides default gateway functionality. That is, it has a switch virtual interface (SVI) for each VLAN.

- Uplinks from the top of rack switch to the upstream layer are routed point-to-point links. You cannot use VLAN trunking on the uplinks.

- A dynamic routing protocol, such as OSPF, IS-IS, or BGP, connects the top of rack switches and upstream switches. Each top of rack switch in the rack advertises a small set of prefixes, typically one per VLAN or subnet. In turn, the top of rack switch calculates equal cost paths to the prefixes it receives from other top of rack switches.

Using Layer 3 routing has the following benefits and drawbacks:

Table 1-2. Benefits and Drawbacks of Layer 3 Transport

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

Figure 1-6. Example Layer 3 Transport

Infrastructure Network Architecture

An important goal of network virtualization is to provide a virtual-to-physical network abstraction.

To implement a virtual-to-physical network abstraction, 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 configuration 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 multi-chassis link aggregation groups, VLAN IDs, and dynamic routing protocol configuration.

Scalability

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

- Number of racks supported in a fabric.

- Amount of bandwidth between any two racks in a data center.

- 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 number of links vary to satisfy different bandwidth demands. You can modify 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 has different demands on the physical switching infrastructure.

- Management traffic, although typically low in volume, is critical for controlling physical and virtual network state.

- IP storage traffic is typically high in volume and generally remains in the boundaries of a data center.

For virtualized environments, the hypervisor sets the QoS values for the different traffic types. The physical switching infrastructure has to 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 1-7. 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 (CoS) marking.

- Layer 3 QoS, also called Differentiated Services Code Point (DSCP) marking.

A vSphere Distributed Switch supports both CoS 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 ESXi host 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 an optimal use of available bandwidth and redundancy in case of a link failure. Use a minimum of two 10-GbE connections, with two 25-GbE connections recommended, for each ESXi host in combination with a pair of top of rack switches. 802.1Q network trunks can support as many VLANs as required. For example, management, storage, VXLAN, vSphere Replication, and vSphere vMotion traffic.

Availability Zones and Regions

In an SDDC, availability zones are collections of infrastructure components. Availability zones are isolated from each other to prevent the propagation of failure or outage across the data center. Consider using regions to place workloads closer to your customers, comply with data privacy laws and restrictions, and support disaster recovery solutions for the entire SDDC.

This design uses a protected region for SDDC management components with one or two availability zones and recovery region with a single availability zone. You can place workloads in each availability zone and region. Usually, multiple availability zones form a single region.

This VMware Validated Design uses two regions, with the option to use one or two availability zones in Region A and single availability zone in Region B. You can expand the design to include multiple availability zones.

Figure 1-8. Availability Zones and Regions

Availability Zones

In a region, each availability zone is isolated from the other availability zones to prevent reproducing failure or outage across zone boundaries.

Using multiple availability zones provides high availability through redundancy.

Table 1-3. Characteristics of Availability Zones

Availability Zone Characteristic	Description
Outage prevention	You avoid outages and improve SLAs. An outage that is caused by external factors, such as power supply, cooling, and physical integrity, affects only one zone. These factors do not cause outage in other zones except in the case of major disasters.
Reliability	Each availability zone runs on its own physically distinct, independent infrastructure, and is engineered to be highly reliable. Each zone should have independent power, cooling, network, and security. Do not share common points of failures in a physical data center, like generators and cooling equipment, across availability zones. Additionally, these zones should be physically separate so that even uncommon disasters affect only one zone.
	Availability zones are either two distinct data centers in a metro distance, or two safety or fire sectors (data halls) in the same large-scale data center.
Distance between zones	Multiple availability zones belong to a single region. The physical distance between availability zones is short enough to offer low, single-digit latency (less than 5 ms) and large bandwidth (10 Gbps or greater) between the zones.
	You can operate workloads across multiple availability zones in the same region as if they were part of a single virtual data center. This architecture supports high availability that is suitable for mission critical applications.
	If the distance between two locations of equipment becomes too large, these locations can no longer function as two availability zones in the same region and must be treated as separate regions.

Regions

By using multiple regions, you can place workloads closer to your customers. For example, you can operate one region on the US East Coast and one region on the US West Coast, or operate a region in Europe and another region in the US.

Using regions has the following advantages:

- Support disaster recovery solutions. One region can be the primary site and another region can be the recovery site.

- Address data privacy laws and restrictions in certain countries by storing tenant data in a region in the same country.

The distance between regions can be large. The latency between regions must be less than 150 ms.

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

Virtual Infrastructure Architecture

The virtual infrastructure is the foundation of the 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.

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. The management components of the SDDC consist of elements in the virtual management layer itself, along with elements in the cloud management layer, or in the operations management, business continuity, or security areas.

Figure 1-9. Virtual Infrastructure Layer in the SDDC

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. Clusters in Region A can use two availability zones.

Figure 1-10. SDDC Logical Design for a Single Availability Zone

Figure 1-11. Logical Design for Two Availability Zones

Availability Zone 1 - Management and Shared Edge and Compute Clusters

Availability Zone 2 (Secondary) - Management and Shared Edge and Compute Clusters

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 Controllers, 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 which 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.

Network Virtualization Components

VMware NSX for vSphere creates the network virtualization layer 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

All virtual networks are created on top of the network virtualization layer, which is an abstraction between the physical and virtual networks. Creating this network virtualization layer requires the following components:

- vCenter Server
- NSX Manager
- NSX Controllers
- NSX logical switch

These components are separated in different planes to create communications boundaries and provide isolation of workload data from system control messages.

Data plane

The data plane handles the workload data only. NSX logical switches separate unrelated workload data. Data is carried over a designated transport network in the physical network. NSX logical switches, distributed routing, and distributed firewall are also implemented in the data plane.

Control plane

The control plane handles network virtualization control messages. Control messages are used to set up networking attributes on NSX logical switch instances, and to configure and manage distributed routing and firewall components on each ESXi host. Control plane communication is carried on secure physical networks (VLANs) that are isolated from the transport networks used for the data plane.

Management plane

The network virtualization orchestration occurs in the management plane. In this layer, cloud management platforms such as vRealize Automation can request, consume, and destroy networking resources for virtual workloads. The cloud management platform directs requests to vCenter Server to create and manage virtual machines, and to NSX Manager to consume networking resources.

Network Virtualization Services

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

Logical Switches

NSX 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 logical switch configuration provides support for line-rate switching in the hypervisor without creating constraints of VLAN sprawl or spanning tree issues.

Universal Distributed Logical Router

Universal distributed logical router (UDLR) in NSX for vSphere performs routing operations in the virtualized space (between VMs, on VXLAN- or VLAN-backed port groups). UDLR has the following features:

- 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

A UDLR is installed in the kernel of every ESXi host, as such it requires a VM for the control plane. The Control VM of a UDLR is the control plane component of the routing process, providing communication between NSX Manager and NSX Controller cluster through the User World Agent. NSX Manager sends logical interface information to the Control VM and NSX Controller cluster, and the Control VM sends routing updates to the NSX Controller cluster.

Figure 1-12. Universal Distributed Logical Routing by Using 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 distributed logical router 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

User World Agent (UWA) is a TCP and SSL client that enables communication between the ESXi hosts and NSX Controller nodes, and the retrieval of information from NSX Manager through interaction with the message bus agent.

Edge Services Gateway

While the UDLR provides VM-to-VM or east-west routing, the NSX Edge services gateway provides north-south connectivity, by peering with upstream layer 3 devices, thereby enabling tenants to access public networks.

Logical Firewall

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 VM names and attributes, user identity, vCenter Server objects like data centers, and ESXi hosts, or can be based on traditional networking attributes like IP addresses, port groups, and so on.

- The Edge Firewall component helps you meet important perimeter security requirements, such as building DMZs based on IP/VLAN constructs, tenant-to-tenant isolation in multi-tenant 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 allows you to extend your datacenter 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 in achieving optimal resource utilization, improving throughput, reducing 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, multi-tier virtual networks in seconds, independent of the underlying network topology or components.

Operations Management Architecture

The architecture of the products of the operations management layer supports centralized monitoring of 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 (metrics) and unstructured (logs) data

- Topology data, such as physical and virtual compute, networking, and storage objects

Figure 1-13. Operations Management Layer of the SDDC

ESXi Patching and Upgrade Architecture

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

Overview

vSphere Update Manager registers with a single vCenter Server instance where 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 1-4. 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 in the vSphere Client or vSphere Web Client.
		Use virtual appliance deployment to deploy easily vCenter Server and vSphere Update Manager as an all-in-one package. Sizing and maintenance for vSphere Update Manager is determined by the vCenter Server deployment.
	Windows installable package for installation against a Microsoft Windows vCenter Server	You must run the vSphere Update Manager installation on 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 in the vSphere Client and 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 pair a vSphere Update Manager instance for a Microsoft Windows only 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 in VMware Knowledge Base Article 2091273.
		UMDS and vSphere Update Manager must be running on different systems.

Architecture

The functional elements of vSphere Update Manager support monitoring, notifying and orchestrating the lifecycle management of your vSphere environment in the SDDC.

Figure 1-14. Architecture of vSphere Update Manager and Update Manager Download Service

Types of Nodes

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

vSphere Update Manager

Required node for integrated, automated lifecycle management of vSphere components. vSphere Update Manager and vCenter Server have a one-to-one relationship, regardless of the number of vCenter Server instances in the environment.

Update Manager Download Service

In a secure environment in which vCenter Server and vSphere Update Manager are isolated from the Internet, use UMDS as a bridge to provide patch and update binaries to vSphere Update Manager. In addition, to manage the lifecycle of multiple vSphere environments, you can use UMDS to aggregate downloaded binary data, such as patch metadata, patch binaries, and notifications, and share it across multiple instances of vSphere Update Manager.

Backup

You back up vSphere Update Manager, either as an embedded service on the vCenter Server Appliance or deployed separately on a Microsoft Windows Server virtual machine and UMDS by using traditional virtual machine backup solutions. Such solutions are based on software that is compatible with vSphere Storage APIs for Data Protection (VADP).

Multi-Region Deployment of vSphere Update Manager and UMDS

Because of its multi-region scope, VMware Validated Design for Software-Defined Data Center uses vSphere Update Manager and UMDS in each region to provide automated lifecycle management of the vSphere components. While you have a vSphere Update Manager service instance with each vCenter Server deployed, you 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 1-15. Dual-Region Interaction Between vSphere Update Manager and Update Manager Download Service

vRealize Suite Lifecycle Management Architecture

VMware vRealize Suite Lifecycle Manager automates the deployment, upgrade, and drift analysis of the VMware vRealize Suite solutions in this design.

Overview

vRealize Suite Lifecycle Manager automates the lifecycle management of the vRealize Suite through a Web-based management application and an API.

The vRealize Suite Lifecycle Manager solution supports the deployment and upgrade of following vRealize Suite products:

- vRealize Operations Manager

- vRealize Log Insight

- vRealize Automation (with embedded vRealize Orchestrator)

- vRealize Business for Cloud

Deployment

vRealize Suite Lifecycle Manager is available as a pre-configured virtual appliance in OVA format. You can access vRealize Lifecycle Manager using a Web-based application and a REST API.

After you deploy vRealize Suite Lifecycle Manager, you register one or more vCenter Server instances with it. Next, an administrator can automate the following operations for the lifecycle management of the vRealize Suite.

Similarly to the vSphere Update Manager Download Service, vRealize Suite Lifecycle Manager provides a bridge for Internet access that is required to download product binaries for install and upgrade from My VMware and content from the VMware Marketplace.

Lifecycle Management Features

vRealize Suite Lifecycle Manager provides the following features for management of multi-product environments:

- Manage a vRealize Suite product repository (installation and upgrade).

- Create environments

- Deployment and management of VMware Marketplace content

- Support existing vRealize Suite-based environments.

- Analyze the configuration drift in environments.

- Scale environments out.

- Upgrade environments.

Architecture

vRealize Suite Lifecycle Manager contains the functional elements that collaborate to orchestrate the lifecycle management operations of the vRealize Suite environment in this VMware Validated Design.

Figure 1-16. vRealize Suite Lifecycle Manager Architecture

Authentication Models

You can configure vRealize Lifecycle Manager user authentication to use the following authentication models:

- Local administrator account
- VMware Identity Manager

Product Binaries Repository

vRealize Suite Lifecycle Manager provides you two methods to retrieve and store product binaries, such as OVA, ISO, and PAK files, for install and upgrade of the vRealize Suite components.

1   Download product binaries to vRealize Suite Lifecycle Manager from My VMware. You can integrate vRealize Suite Lifecycle Manager directly with a My VMware account to access vRealize Suite entitlements. By using the My VMware integration, you can download vRealize Suite product binaries to the vRealize Suite Lifecycle Manager appliance repository.

2   Upload product binaries to vRealize Suite Lifecycle Manager. If your organization must restrict external traffic from the management components of the Software-Defined Data Center, you can download the vRealize Suite product binaries from My VMware and upload them to vRealize Suite Lifecycle Manager repository as an alternative approach. You can upload binaries from a local directory or an NFS endpoint share.

Marketplace Integration

By using vRealize Suite Lifecycle Manager, you can deploy vRealize Operations management packs, vRealize Log Insight content packs, and vRealize Automation blueprints and OVA files directly from the VMware Marketplace.

Backup

The vRealize Suite Lifecycle Manager virtual appliance is backed up using traditional virtual machine backup solutions that are compatible with VMware vSphere Storage APIs – Data Protection (VADP).

Multi-Region Deployment of vRealize Suite Lifecycle Manager

The scope of this design can cover both single region, multiple regions, and availability zones.

In a multi-region implementation, VMware Validated Design for Software-Defined Data Center implements a vRealize Suite Lifecycle Manager setup in multiple regions by using the following configuration:

- A single vRealize Suite Lifecycle Manager virtual appliance is replicated by vSphere Replication and recovered by Site Recovery Manager so that you can fail it over across regions in the event of planned migration or disaster recovery event.

- The vRealize Suite Lifecycle Manager instance manages the deployment, upgrade, and configuration drift analysis of the vRealize products across all regions.

In a multi-availability zone implementation, which is a subset of the multi-region design, vRealize Suite Lifecycle Manager continues to provide lifecycle services for the vRealize Suite based deployments in all regions of the SDDC. The virtual appliance of vRealize Suite Lifecycle Manager resides in Availability Zone 1 in Region A. If this zone becomes compromised, the instance it brought back online in Availability Zone 2.

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 pre-configured virtual appliance in OVF. By using the virtual appliance, you can easily create vRealize Operations Manager nodes with pre-defined 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 Operations Manager as a virtual appliance in one of the following configurations:

- Standalone node

- Cluster of one primary 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 support creating clusters of nodes with different roles.

Figure 1-17. 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 either of the following roles.

Primary Node

Required initial node in the cluster. In large-scale environments, manages all other nodes. In small-scale environments, the primary node is the single standalone vRealize Operations Manager node.

Primary Replica Node

Optional. Enables high availability of the primary 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. Remote collector nodes do not store data or perform analysis.

The primary and primary 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 primary node, data nodes, and optionally of a primary 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 cases where 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 1-18. Architecture of a vRealize Operations Manager Node

The components of vRealize Operations Manager node perform these tasks.

Product/Admin UI and Suite API

The UI server is a Web application that serves as both user and administration interface, and 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 primary, primary 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 data sets, 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 is 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, collector groups, and 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 more details about and download management packs from *VMware Solutions Exchange*.

Backup

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

Multi-Region vRealize Operations Manager Deployment

The scope of this validated design can cover both multiple regions and availability zones.

VMware Validated Design for Software-Defined Data Center implements a large-scale vRealize Operations Manager deployment across multiple regions by using the following configuration:

- 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 to handle data coming from management solutions

In a multi-availability zone implementation, which is a super-set of the multi-region design, vRealize Operations Manager continues to provide monitoring of the solutions in all regions of the SDDC. All components of vRealize Operations Manager reside in Availability Zone 1 in Region A. If this zone becomes compromised, all nodes are brought up in Availability Zone 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 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 pre-installed agent on certain VMware products. Using pre-installed agents is 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 pre-configured virtual appliance in OVF. By using the virtual appliance, you can easily create vRealize Log Insight nodes with pre-defined 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:

- Standalone node

- Cluster of one primary 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 1-19. 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 via the vRealize Log Insight agent to send logs to vRealize Log Insight. Users and administrators interact with the ingested logs using the user interface or the API.

By default, vRealize Log Insight collects data from vCenter Server systems and ESXi hosts. For forwarding logs from NSX for vSphere and vRealize Automation, use content packs. 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:

Primary Node

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

The primary node stores logs locally. If the primary 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 primary 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 (ILB)

In cluster mode, the ILB is the centralized entry point which ensures that vRealize Log Insight accepts 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 re-election is required. If the node that hosts the ILB Virtual IP (VIP) address stops responding, the VIP address is failed over to another node in the cluster via an election process.

All queries against data are directed to the ILB. The ILB delegates queries to a query primary for the duration of the query. The query primary queries all nodes, both primary 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 primary 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 using API calls.

Figure 1-20. vRealize Log Insight Logical Node Architecture

The 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 and API traffic, and incoming data ingestion traffic.

The Integrated Load Balancer is a Linux Virtual Server (LVS) that is built in the Linux Kernel for Layer 4 load balancing. Each node of vRealize Log Insight contains a service running the Integrated Load Balancer, but only a single node functions as the leader at all times. In a single-node vRealize Log Insight instance, this is always the primary 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 re-election 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 not replicated. If a node is offline or removed, the logging data which 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 holding 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 utilize one or more of the following authentication models:

- Microsoft Active Directory

- Local Accounts

- VMware Identity Manager

Content Packs

Content packs help add valuable troubleshooting information in 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. They add vRealize Log Insight agent configurations, providing out-of-the-box parsing capabilities for standard logging directories and logging formats, along with dashboards, extracted fields, alert definitions, query lists, and saved queries from the logging data related to a specific product in vRealize Log Insight. Visit *Log Insight Content Pack Marketplace* 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. The 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 within 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 purposes. vRealize Log Insight also does not perform cleanup of 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 using traditional virtual machine backup solutions that are compatible with VMware vSphere Storage APIs – Data Protection (VADP).

Multi-Region vRealize Log Insight Deployment

The scope of this validated design can cover both multiple regions and availability zones.

In a multi-region 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:

- Cluster in each region.

- Event forwarding to other vRealize Log Insight deployments across regions in the SDDC.

In a multi-availabilty zone implementation, which is a sub-set of the multi-region design, vRealize Log insight continues to provide a logging infrastructure in all regions of the SDDC. All components of the vRealize Log Insight cluster reside in Availability Zone 1 within Region A. If this zone becomes compromised, all nodes are brought up in the Availability Zone 2.

Failover by using vSphere Replication or disaster recovery by using 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 1-21. Event Forwarding in vRealize Log Insight

Cloud Management Architecture

The Cloud Management Platform (CMP) is the main consumption portal for the Software-Defined Data Center (SDDC). You use vRealize Automation to author, administer, and consume VM templates and blueprints and vRealize Business for Cloud to analyze the cost of deploying virtual infrastructure.

Figure 1-22. Cloud Management Layer in the SDDC

The cloud management layer provides the following multi-platform and multi-vendor cloud services:

- Capabilities to provide quickly standardized resources to global customers.

- Methods for multi-platform and multi-vendor delivery that integrate with existing enterprise management systems.

- Central governance for physical, virtual, private, and public cloud services that is user-centric and is aware of the business requirements.

- Extensible architecture

vRealize Automation Architecture

vRealize Automation provides a secure Web portal where authorized administrators, developers, and business users can request new IT services and manage specific cloud and IT resources according to business policies. Requests for IT services, including infrastructure, applications, desktops, and many others, are processed by using a common service catalog to provide a consistent user experience.

Deployment

Installing vRealize Automation requires deploying the vRealize Automation appliance, and the vRealize Automation Infrastructure as a Service (IaaS) components on one or more Windows Server virtual machines or physical servers. You deploy the vRealize Automation appliance and then complete the installation using 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 takes input from an answer properties file.

- An installation REST API that takes JSON formatted input.

vRealize Automation Architecture

vRealize Automation provides self-service provisioning, IT services delivery and lifecycle management of cloud services across a many multi-vendor virtual, physical, and cloud platforms using a flexible distributed architecture. The two main functional elements of the architecture are the vRealize Automation appliance and the IaaS components.

Figure 1-23. vRealize Automation Architecture

vRealize Automation Appliance

The vRealize Automation appliance is available as a pre-configured Linux virtual appliance in OVF. You deploy the appliance on existing virtualized infrastructure such as vSphere. The vRealize Automation appliance performs the following functions:

- vRealize Automation product portal, where users 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 pre-configured instance of vRealize Orchestrator. vRealize Automation uses vRealize Orchestrator workflows and actions to extend its provisioning capabilities.

PostgreSQL Database

vRealize Server uses a pre-configured PostgreSQL database that is included in the vRealize Automation appliance. The instance of vRealize Orchestrator in the vRealize Automation appliance also uses this database.

Infrastructure as a Service

vRealize Automation IaaS part consists of one or more Microsoft Windows Server instances 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. The Model Manager is hosted on one of the IaaS Web servers and communicates with DEMs, the Microsoft 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

The Manager Service is Windows service that coordinates the 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 Server instances.

Distributed Execution Manager Orchestrator

A Distributed Execution Manager (DEM) performs the business logic of custom models, interacting with the database and with external databases and systems as required. A DEM orchestrator is responsible for monitoring DEM worker instances, pre-processing workflows for execution, and scheduling workflows.

Distributed Execution Manager Worker

The IaaS DEM worker performs provisioning and de-provisioning 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, a vSphere proxy agent sends commands to and collects data from an ESXi host about the virtual machines that you provisioned from vRealize Automation.

VMware Identity Manager

VMware Identity Manager is the main identity provider for vRealize Automation. VMware Identity Manager manages user authentication, roles, permissions, and overall access to vRealize Automation using federated identity brokering. The following authentication methods are supported in vRealize Automation using VMware Identity Manager:

- User name-password providing single-factor password authentication with basic Active Directory configuration or for local users

- Kerberos

- Smart Card / Certificate

- RSA SecurID

- RADIUS

- RSA Adaptive Authentication

- SAML Authentication

Deployment Model

You can deploy vRealize Automation in one of the following configurations:

- Small deployment for up to 10,000 managed virtual machines

    - vRealize Automation appliance, potentially behind a load balancer

    - IaaS node

    - Microsoft SQL Server

- Medium deployment for up to 30,000 managed virtual machines

    - 3 vRealize Automation appliances behind a load balancer

    - 2 IaaS Web/Manager servers behind a load balancer

    - 2 IaaS DEM servers

    - 2 IaaS Proxy Agents

    - Microsoft SQL Server

- Large deployment for up to 50,000 managed virtual machine

  - 3 vRealize Automation appliances behind a load balancer

  - 2 IaaS Web servers behind a load balancer

  - 2 IaaS Manager servers behind a load balancer

  - 2 IaaS DEM servers

  - 2 IaaS Proxy Agents

  - Microsoft SQL Server

To address future growth of tenant workloads beyond 10,000 virtual machines without more operational overhead, this design implements a large deployment of vRealize Automation.

Multi-Region vRealize Automation Deployment

The scope of this validated design can cover both multiple regions and availability zones.

The scope of the VMware Validated Design includes vRealize Automation appliance large-scale distributed deployment designed for a complete and highly available cloud management solution that includes:

Table 1-5. vRealize Components that Are Failed Over

	Failed Over
3 vRealize Automation appliances behind a load balancer	X
2 IaaS Web servers behind a load balancer	X
2 IaaS Manager Service nodes (including DEM orchestrator) behind a load balancer	X
2 DEM Worker nodes	X
2 vSphere Proxy Agent nodes	
Microsoft SQL Server	X

In a multi-availability zone implementation, which is a super-set of the multi-region design, vRealize Automation continues to provide provisioning of tenant workloads in all regions of the SDDC. All components of the vRealize Automation reside in Availability Zone 1 in Region A. If this zone becomes compromised, all nodes are brought up in Availability Zone 2.

vRealize Business for Cloud Architecture

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

Use vRealize Business for Cloud to track and manage the costs of private and public cloud resources from a single dashboard. It offers a comprehensive way to see, plan and manage your cloud costs.

vRealize Business for Cloud is integrated with vRealize Automation. The architecture illustrates the main components of vRealize Business for Cloud, the server, FactsRepo inventory service, data transformation service, data collection services, and reference database.

Figure 1-24. vRealize Business Architecture

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 (CPU and memory) statistics. The data collection services use the collected data for cost calculations.

You can deploy only the data collection services by using the remote data collector version of the vRealize Business appliance. Remote data collectors reduce the data collection workload of the vRealize Business server, and enable remote data collection from geographically distributed endpoints.

Facts Repo Inventory Service

An inventory service that stores on MongoDB the collected data that vRealize Business uses for cost computation.

Data Transformation Service

A service that 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 aggregation of data from all data collectors.

vRealize Business Server

A Web application that runs on Pivotal tc Server. vRealize Business has multiple data collection services that run periodically, collecting inventory information and statistics, which are stored in a PostgreSQL database. Data that is collected from the data collection services is used for cost calculations.

Reference Database

A database that is 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 data set and import it into vRealize Business. The new values affect cost calculation. The reference data used depends on the currency you select at the time of installation.

Important You cannot change the currency configuration after you deploy vRealize Business for Cloud.

Communication Between Server and Reference Database

The reference database is a compressed and encrypted file, which you can download and install manually or update automatically. You can update the most current version of reference database.

Other Sources of Information

These information sources are optional. You use them only if installed and configured. The sources include vRealize Automation, vCloud Director, vRealize Operations Manager, Amazon Web Services (AWS), Microsoft Azure, vCloud Air, and EMC Storage Resource Manager (SRM).

Operational Model

vRealize Business for Cloud continuously collects data from external sources, and periodically updates the Facts Repo inventory service. You can view the collected data using the vRealize Business dashboard or generate a report. Data synchronization and updates occur at regular intervals. You can also manually trigger the data collection process when inventory changes occur. For example, in response to the initialization of the system, or addition of a cloud account.

Backup

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

Multi-Region vRealize Business Deployment

The scope of this validated design can cover both multiple regions and availability zones.

VMware Validated Design for Software-Defined Data Center implements a vRealize Business deployment across multiple regions by using the following configuration:

- To support its availability in vRealize Automation, vRealize Business server appliance is protected by Site Recovery Manager to fail over across regions

- One data collector node in each region to handle data coming from management solutions.

In a multi-availability zone implementation, which is a super-set of the multi-region design, vRealize Business continues to provide cost analysis for tenant workloads in all regions of the SDDC. All components of vRealize Business reside in Availability Zone 1 in Region A. If this zone becomes compromised, all nodes are brought up in Availability Zone 2.
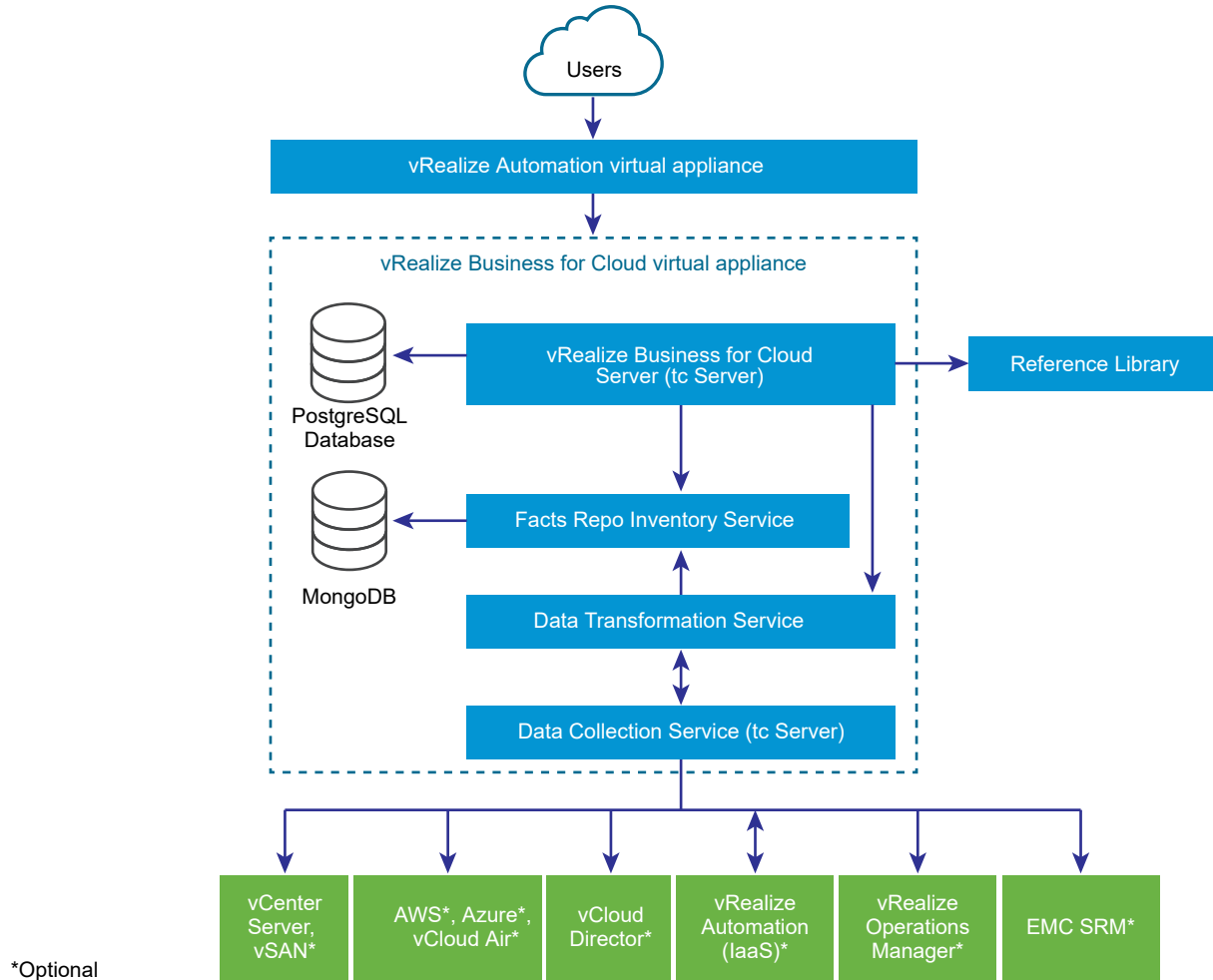
Business Continuity Architecture

The architecture of the business continuity layer includes management components that provide support for backup, restore, and disaster recovery operational procedures.

In the business continuity layer, management components are implemented to handle the following business continuity requirements.

- Data protection

- Data replication

- Orchestrated disaster recovery

Figure 1-25. Business Continuity Layer of the SDDC

Data Protection and Backup Architecture

You can implement a backup solution that uses the VMware vSphere Storage APIs – Data Protection (VADP), to protect the data of your SDDC management components, and of the tenant workloads that run in the SDDC.

Data protection solutions provide the following functions in the SDDC:

- Backup and restore virtual machines.

- Organization of virtual machines into groups by VMware product.

- Store data according to company retention policies.

- Inform administrators about backup and restore activities through reports.

- Schedule regular backups during non-peak periods.

Architecture

VADP instances provide data protection for the products that implement the management capabilities of the SDDC.

Figure 1-26. Dual-Region Data Protection Architecture

Multi-Region Data Protection Deployment

Because of its multi-region scope, the VMware Validated Design for Software-Defined Data Center supports the deployment of a VADP-compatible backup solution in the management cluster for each region. To provide recovery of a number of SDDC management components, you configure backup jobs. The VADP-compatible backup solution stores the backups of the management virtual appliances on a secondary storage according to a defined schedule.

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. To avoid outage reproduction in the protected region, use stretched storage clusters and implement an architecture with multiple availability zones.

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 which is considered as the protected region, and are recovered in Region B which is considered as 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

- Array-based replication. 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 configure replication first 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.

  You can configure vSphere Replication to use the multiple-point-in-time snapshot feature enabling more 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 1-27. Disaster Recovery Architecture

Multi-Region Site Recovery Manager Deployment

The scope of the VMware Validated Design 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 a subject of disaster recovery protection:

    - vRealize Automation, vRealize Business Server, and vRealize Suite Lifecycle Manager

    - Analytics cluster of vRealize Operations Manager

- The virtual infrastructure components that are not in the scope of the disaster recovery protection, such as vRealize Log Insight, are available as separate instances in each region.

Avoiding Disaster by using Multiple Availability Zones

To integrate stretched storage clusters for first-level disaster avoidance, use two availability zones in Region A: Availability Zone 1 and Availability Zone 2. If a disaster occurs, use the multi-region capabilities of Site Recovery Manager for orchestrated recovery.

Detailed Design

<div style="text-align: right; font-size: 3em; color: #999;">2</div>

The Software-Defined Data Center (SDDC) detailed design considers both physical and virtual infrastructure design. It includes numbered design decisions, and the justification and implications of each decision.

This design provides two design options for availability zone setup. In certain areas, configurations and design decision alternatives are specific to a single availability zone setup and to a multiple availability zone setup.

Physical Infrastructure Design

Focuses on the three main aspects of any data center: compute, storage, and network. In this section, you find information about availability zones and regions. The section also provides details on the rack and cluster configuration, and on physical ESXi hosts and the associated storage and network configurations.

Virtual Infrastructure Design

Provides details on the core virtualization software configuration. This section has information on the ESXi hypervisor, vCenter Server, the virtual network design including VMware NSX, and on software-defined storage for VMware vSAN. This section also includes details on business continuity (backup and restore) and on disaster recovery.

Operations Infrastructure Design

Explains how to architect, install, and configure vRealize Suite Lifecycle Manager, vSphere Update Manager, vRealize Operations Manager, vRealize Log Insight for lifecycle and service management.

Cloud Management Platform Design

Contains information on the consumption and orchestration layer of the SDDC stack, which uses vRealize Automation, vRealize Orchestrator, and vRealize Business. IT organizations can use the fully distributed and scalable architecture to streamline their provisioning and decommissioning operations.

This chapter includes the following topics:

- Physical Infrastructure Design
- Virtual Infrastructure Design

- [Operations Management Design](#)

- [Cloud Management Design](#)

- [Business Continuity Design](#)

Physical Infrastructure Design

The physical infrastructure design includes deciding on the configuration of availability zones and regions, and on the cluster layout in datacenter racks.

Design decisions related to server, networking, and storage hardware are part of the physical infrastructure design.

Figure 2-1. Physical Infrastructure Design

- [Physical Design Fundamentals](#)

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

- [Physical Networking Design](#)

- [Physical Storage Design](#)

  VMware Validated Design uses different types of storage. Consider storage mode, hardware compatibility for the selected storage, and I/O controllers.

Physical Design Fundamentals

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

Availability Zones and Regions

Availability zones and regions have different purposes. Availability zones protect against failures of individual hosts. You can consider regions to place workloads closer to your customers, comply with data privacy laws and restrictions, and support disaster recovery solutions for the entire SDDC.

This design uses a protected region for SDDC management components with one or two availability zones and recovery region with a single availability zone. You can place workloads in each availability zone and region. Usually, multiple availability zones form a single region.

Availability zones

An availability zone is the fault domain of the SDDC. Multiple availability zones can provide continuous availability of an SDDC, minimize down time of services and improve SLAs.

Regions

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

The SDDC in this validated design contains two regions.

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

Table 2-1. Availability Zones and Regions in the SDDC

Region	Availability Zone and Region Identifier	Region-Specific Domain Name	Region Description
A	SFO01	sfo01.rainpole.local	Availability Zone 1 in San Francisco, CA, USA based data center
A	SFO02	sfo01.rainpole.local	Availability Zone 2 in San Francisco, CA, USA based data center
B	LAX01	lax01.rainpole.local	Los Angeles, CA, USA based data center

Table 2-2. Design Decisions on Availability Zones and Regions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-001	In Region A, deploy one or two availability zones to support all SDDC management components and their SLAs.	Supports all SDDC management and compute components for a region.	■ Using a single availability zone results in limited redundancy of the overall solution. ■ The single availability zone can become a single point of failure and prevent high-availability design solutions in a region.
		Supports stretched clusters and application-aware failover for high availability between two physical locations.	Implementing two availability zones increases the solution footprint and can complicate the operational procedures.

Table 2-2. Design Decisions on Availability Zones and Regions (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-002	Use two regions.	Supports the technical requirement of multi-region 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 one availability zone to support disaster recovery of the SDDC management components.	Supports all SDDC management and tenant components for a region. You can later add another availability zone to extend and scale out the management and tenant capabilities of the SDDC.	■ Using a single availability zone results in limited redundancy of the overall solution. ■ The single availability zone can become a single point of failure and prevent high-availability design solutions in a region.

Clusters and Racks

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

Figure 2-2. SDDC Cluster Architecture for a Single Availability Zone

Figure 2-3. SDDC Cluster Architecture for Two Availability Zones

Table 2-3. Design Decisions on Clusters and Racks

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-004	In each availability zone, 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 (4 ESXi servers) and shared edge and compute cluster (4 ESXi servers) is low and does not justify a dedicated rack for each cluster. You provide on-ramp and off-ramp connectivity to physical networks (for example, north-south Layer 3 routing on NSX Edge virtual appliances) to both the management and shared edge and compute clusters by using this management and edge rack. Edge resources require external connectivity to physical network devices. Placing edge resources for management and tenants in the same rack minimizes VLAN spread.	The data centers must have sufficient power and cooling to operate the server equipment according to the selected vendor and products. If the equipment in the entire rack fails, to reduce downtime associated with such an event, you must have a second availability zone or region.
SDDC-PHY-005	Allocate one or more racks to external storage.	To simplify the scale-out of the SDDC infrastructure, standardize the storage-to-racks relationship. The storage system might arrive from the manufacturer in a dedicated rack or set of racks. In this way, a storage system of this type is accommodated for in the design.	The data centers must have sufficient power and cooling to operate the storage equipment according to the selected vendor and products.

Table 2-3. Design Decisions on Clusters and Racks (continued)

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 ensuring that failure of a power feed does not bring down all equipment in a rack. Combined with redundant network connections to a rack and in a rack, redundant power feeds prevent a failure of the 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 downtime associated with such an event.
SDDC-PHY-007	For each availability zone, mount the compute resources (minimum of 4 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	For each availability zone, mount the compute resources for the shared edge and compute cluster (minimum of 4 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 determine the characteristics of the ESXi hosts that you use to deploy this VMware Validated Design.

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. Because standardization of the physical configuration of the ESXi hosts removes variability, you operate an easily manageable and supportable infrastructure. Deploy ESXi hosts with identical configuration across all cluster members, including storage and networking configurations. For example, consistent PCI card slot placement, especially for network controllers, is essential for accurate alignment of physical to virtual I/O resources. By using identical configurations, you have an even balance of virtual machine storage components across storage and compute resources.

Select all ESXi host hardware, including CPUs, according to VMware Compatibility Guide.

The sizing of physical servers that run ESXi requires special considerations when you use vSAN storage. The design provides details on using vSAN as the primary storage system for the management cluster and the shared edge and compute cluster. This design also uses vSAN ReadyNodes. For information about the models of the physical servers, see the vSAN ReadyNode compatibility guide.

This design supports using a storage system other than vSAN as the primary storage system for the management and shared edge and compute clusters. However, using a non-vSAN storage is not a part of the documentation of this validated design. For details on vSAN capacity and performance characteristics to aid your selection of non-vSAN storage, see Physical Storage Design .

- An average-size VM has two vCPUs with 4 GB of RAM.

- A standard 2U server can host 60 average-size VMs on a single ESXi host.

Table 2-4. Design Decisions on the Physical Design of the ESXi Hosts

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-009	Use vSAN ReadyNodes with vSAN storage.	Using vSAN ReadyNodes ensures full compatibility with vSAN at deployment.	Hardware choices might be limited.
SDDC-PHY-010	Verify that all nodes have uniform configuration across a cluster.	A balanced cluster has more predictable performance even during hardware failures. In addition, the impact on performance during resync or rebuild is minimal if the cluster is balanced.	Apply vendor sourcing, budgeting, and procurement considerations for uniform server nodes, on a per cluster basis.

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, consider the admission control setting (n+1), which reserves the resources of one host for failover or maintenance.

The number of disk groups and disks that an ESXi host manages determines the memory requirements. To support the maximum number of disk groups, you must provide 32 GB of RAM. For more information about disk groups, including design and sizing guidance, see *Administering VMware vSAN* from the vSphere documentation.

Table 2-5. Design Decisions on the ESXi Host Memory

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-011	Set up each ESXi host in the management cluster with a minimum of 256 GB RAM.	The management and edge VMs in this cluster require a total of 453 GB RAM. The remaining RAM is available for new capabilities of the SDDC such as deployment of VMware NSX-T or VMware PKS.	In a four-node cluster, only 768 GB is available for use due to the n+1 vSphere HA setting.

ESXi Host Boot Device

Minimum boot disk size for ESXi in SCSI-based devices (SAS, SATA, or SAN) is greater than 5 GB. ESXi can be deployed using stateful local SAN SCSI boot devices, or by using vSphere Auto Deploy.

Selecting a boot device type and size for vSAN has the following considerations:

- vSAN does not support stateless vSphere Auto Deploy.

- Device types as ESXi boot devices

  - USB or SD embedded devices. The USB or SD flash drive must be at least 4 GB.

  - SATADOM devices. The size of the boot device per host must be at least 16 GB.

See *VMware vSAN Design and Sizing Guide* to select the option that best fits your hardware.

Physical Networking Design

Design of the physical SDDC network includes defining the network topology for connecting the physical switches and the ESXi hosts, determining switch port settings for VLANs and link aggregation, and designing routing. VMware Validated Design for Software-Defined Data Center can use most enterprise-grade physical network architectures.

Switch Types and Network Connectivity

Follow best practices for physical switches, switch connectivity, setup of VLANs and 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 Dynamic Trunking Protocol (DTP).

- Modify the Spanning Tree Protocol (STP) 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, using the Trunk PortFast feature on a Cisco physical switch.

- Provide DHCP or DHCP Helper capabilities on all VLANs that are used by the 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, inter-switch link (ISL) 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 using a minimum of two 10-GbE ports (two 25-GbE ports are recommended). Configure the ToR switches to provide all necessary VLANs via an 802.1Q trunk. These redundant connections use features of vSphere Distributed Switch and NSX for vSphere to guarantee no physical interface is overrun and redundant paths are used as long as they are available.

Figure 2-4. Host-to-ToR Connection

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 .254 as the (floating) interface with .252 and .253 for Virtual Router Redundancy Protocol (VRPP) or 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 2-6. Sample Values for VLANs and IP Ranges

Cluster	Availability Zone	Function	Sample VLAN	Sample IP range
Management	- Availability Zone 1 - Availability Zone 2	Management	1611 (Native, Stretched)	172.16.11.0/24
	Availability Zone 1	vMotion	1612	172.16.12.0/24
		VXLAN	1614	172.16.14.0/24
		vSAN	1613	172.16.13.0/24
	Availability Zone 2	Management	1621	172.16.21.0/24

Table 2-6. Sample Values for VLANs and IP Ranges (continued)

Cluster	Availability Zone	Function	Sample VLAN	Sample IP range
		vMotion	1622	172.16.22.0/24
		VXLAN	1614	172.16.24.0/24
		vSAN	1623	172.16.23.0/24
Shared Edge and Compute	■ Availability Zone 1 ■ Availability Zone 2	Management	1631 (Native, Stretched)	172.16.31.0/24
	Availability Zone 1	vMotion	1632	172.16.32.0/24
		VXLAN	1634	172.16.34.0/24
		vSAN	1633	172.16.33.0/24
	Availability Zone 2	Management	1641	172.16.41.0/24
		vMotion	1642	172.16.42.0/24
		VXLAN	1634	172.16.44.0/24
		vSAN	1643	172.16.43.0/25

Note Because NSX prepares the stretched cluster upon installation, the VXLAN VLANs have the same VLAN ID in both Availability Zone 1 and Availability Zone 2. The VLAN ID must exist in both availability zones but must map to a different routeable IP space in each zone so that NSX virtual tunnel end-points (VTEPs) can communicate.

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 a 802.1Q trunk with the management VLAN acting as the native VLAN.

MTU

> Set MTU for all VLANS and SVIs to the value for jumbo frames for consistency purposes.

DHCP helper

> Configure the VIF of the Management and VXLAN subnet as a DHCP proxy.

Multicast

Configure IGMP snooping on the ToR switches and include an IGMP querier on each VXLAN VLAN.

Connectivity Between Regions

The SDDC management networks, 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, 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 ensure that latency is less than 150 ms. For more details on the requirements for region interconnectivity, see the *Cross-VC NSX Design Guide*.

The design of a solution for region interconnectivity is out of scope for this VMware Validated Design.

Connectivity Between Availability Zones

Consider the following connectivity requirements for multiple availability zones:

- The latency between availability zones in the SDDC must be less than 5 ms.

- The network bandwidth must be 10 Gbps or higher.

- To support failover or VLAN-backed appliances such as vCenter Server, NSX Manager, and NSX Controller nodes, the management VLAN must be stretched between availability zones.

The design of a solution for availability zone interconnectivity is out of scope for this VMware Validated Design.

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 available expertise among the IT staff. Consider performance requirements. Possible options are OSPF, BGP, and IS-IS. Although each routing protocol has a complex set of advantages and disadvantages, this validated design utilizes BGP as its routing protocol.

DHCP proxy

Set the DHCP proxy to point to a DHCP server by IPv4 address. See the *VMware Validated Design Planning and Preparation* document for details on the DHCP server.

Table 2-7. Design Decisions on the Physical Network

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-NET-001	Implement the following physical network architecture: ■ A minimum of one 10-GbE port (one 25 GbE port recommended) on each ToR switch for ESXi host uplinks ■ No EtherChannel (LAG/vPC) configuration for ESXi host uplinks ■ Layer 3 device with BGP and IGMP support	■ Guarantees availability during a switch failure. ■ Provides compatibility with vSphere host profiles because they do not store link-aggregation settings. ■ Supports BGP as the dynamic routing protocol in the SDDC. ■ Provides compatibility with NSX hybrid mode replication because it 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 multi-site and multi-tenancy workloads.	Requires BGP configuration in the physical networking stack.
SDDC-PHY-NET-003	Use two ToR switches for each rack.	This design uses a minimum of two 10-GbE links, with two 25-GbE links recommended, to each ESXi host and provides redundancy and reduces the overall design complexity.	Requires two ToR switches per rack which can increase costs.
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 2-8. Additional Design Decisions on the Physical Network

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 VTEPs. NSX VTEPs are assigned by using a DHCP server. Set the lease duration for the VTEP DHCP scope to at least 7 days.	Ensures that interfaces such as management and storage always have the same IP address. In this way, you provide support for continuous management of ESXi hosts using vCenter Server and for provisioning IP storage by storage administrators. NSX VTEPs do not have an administrative endpoint. As a result, they can use DHCP for automatic IP address assignment. You are also unable to assign directly a static IP address to the VMkernel port of an NSX VTEP. IP pools are an option but the NSX administrator must create them. If you must change or expand the subnet, changing the DHCP scope is simpler than creating an IP pool and assigning it to the ESXi hosts. In a vSAN stretched configuration, the VLAN ID is the same in both availability zones. If you use IP pools, VTEPs in different availability zones will be unable to communicate. By using DHCP, each availability zone can have a different subnet associated with the same VLAN ID. As a result, the NSX VTEPs can communicate over Layer 3.	Requires precise IP address management.
SDDC-PHY-NET-006	Create DNS records for all management nodes to enable forward, reverse, short, and FQDN resolution.	Ensures consistent resolution of management nodes 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 critical 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 needs an MTU value of at least 1600 bytes on the switches and routers that carry the transport zone traffic.

Table 2-9. Design Decisions on Jumbo Frames

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. ■ vSAN ■ vMotion ■ VXLAN ■ vSphere Replication ■ NFS	Improves traffic throughput. To support VXLAN, the MTU setting must be increased to a minimum of 1600 bytes. Setting the MTU to 9000 bytes has no effect on VXLAN, but provides consistency across port groups that are adjusted from the default MTU size.	When you adjust 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.

Physical Storage Design

VMware Validated Design uses different types of storage. Consider storage mode, hardware compatibility for the selected storage, and I/O controllers.

All functional testing and validation of the design is on vSAN. Although VMware Validated Design uses vSAN, in particular for the clusters running management components, you can use any supported storage solution.

If selecting a storage solution other than vSAN, you must take into account that all the design, deployment, and Day-2 guidance in VMware Validated Design applies under the context of vSAN and adjust appropriately.

Your storage design must match or exceed the capacity and performance capabilities of the vSAN configuration in the design. For multiple availability zones, the vSAN configuration includes vSAN stretched cluster.

vSAN Physical Design

This design uses VMware vSAN to implement software-defined storage as the primary storage type for the management cluster. By using vSAN, you have a high level of control over the storage subsystem.

vSAN is a hyper-converged storage software that is fully integrated with the hypervisor. vSAN creates a cluster of local ESXi host hard disk drives and solid state drives, and presents a flash-optimized, highly-resilient, shared storage datastore to ESXi hosts and virtual machines. By using vSAN storage policies, you can control capacity, performance, and availability on a per virtual machine basis.

Requirements and Dependencies

The software-defined storage module has the following requirements and options.

Requirement Category	Requirements
Number of hosts	■ Minimum of 3 ESXi hosts providing storage resources to the vSAN cluster.
vSAN configuration	vSAN is configured as hybrid storage or all-flash storage. ■ A vSAN hybrid storage configuration requires both magnetic devices and flash caching devices. ■ An all-flash vSAN configuration requires flash devices for both the caching and capacity tiers.
Requirements for individual hosts that provide storage resources	■ Minimum of one SSD. The SSD flash cache tier should be at least 10% of the size of the HDD capacity tier. ■ Minimum of two HDDs for hybrid, or two additional flash devices for an all-flash configuration ■ RAID controller that is compatible with vSAN. ■ Minimum 10 Gbps network for vSAN traffic. ■ vSphere High Availability host isolation response set to power off virtual machines. With this setting, you prevent split-brain conditions if isolation or network partition occurs. In a split-brain condition, the virtual machine might be powered on by two ESXi hosts by accident. See design decision Table 2-33. Design Decisions on vSphere HA for more details.

Hybrid Mode and All-Flash Mode

vSAN has two modes of operation: all-flash and hybrid.

Hybrid Mode

In a hybrid storage architecture, vSAN pools server-attached capacity devices (in this case magnetic devices) and caching devices, typically SSDs or PCI-e devices, to create a distributed shared datastore.

All-Flash Mode

All-flash storage uses flash-based devices (SSD or PCI-e) as a write cache while other flash-based devices provide high endurance for capacity and data persistence.

Table 2-10. Design Decisions on the vSAN Mode

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-001	Configure vSAN in hybrid mode in the management cluster.	Provides performance that is good enough for the VMs in the management cluster that are hosted on vSAN.	vSAN hybrid mode does not provide the potential performance or additional capabilities such as deduplication of an all-flash configuration.

Sizing Storage

You usually base sizing on the requirements of the IT organization. However, this design provides calculations that are based on a single-region implementation, and is then implemented on a per-region basis. In this way, you can handle storage in a dual-region deployment that has failover capabilities enabled.

This sizing is calculated according to a certain node configuration per region. Although VMware Validated Design has enough memory capacity to handle N-1 host failures, and uses thin-provisioned swap for the vSAN configuration, the potential think-provisioned swap capacity is factored in the calculation.

Table 2-11. Management Layers and Hardware Sizes

Category	Quantity	Resource Type	Capacity Consumption
Physical Infrastructure (ESXi)	4	Memory	1024 GB
Virtual Infrastructure	17	Disk	1,199 GB
		Swap	108 GB
Operations Management	10	Disk	6,272 GB
		Swap	170 GB
Cloud Management	14	Disk	1,200 GB
		Swap	144 GB
Business Continuity	2	Disk	58 GB
		Swap	16 GB
Total	■ 43 management virtual machines ■ 4 ESXi hosts	Disk	8,729 GB
		Swap	438 GB
		Memory	1024 GB

Derive the storage space that is required for the capacity tier according to the following calculations. For vSAN memory consumption by management ESXi hosts, see VMware Knowledge Base article 2113954.

```
[Static Base Consumption + (Number of Disk Groups * (Static Disk Group Base Consumption + (Static
Flash Device Memory Overhead Per GB * Flash Device Capactiy))) + (Static Capacity Disk Base
Consumption * Number of Capacity Disks) ] * Host Quantity = vSAN  Memory Consumption
```

```
[5426 MB + (1 Disk Group * (636 MB + (8 MB * 300 GB Flash Storage))) + (70 MB * 2 Magnetic Disks)] *
4 ESXi Hosts
```

```
[5426 MB + (1 Disk Groups * (636 MB + 2400 MB)) + 140 MB] =  8602 MB / 1024 = 8.4 GB vSAN Memory
Consumption per host or 33.6 GB for a four node cluster
```

Derive the consumption of storage space by the management virtual machines according to the following calculations. See VMware vSAN Design and Sizing Guide.

```
VM Raw Storage Requirements (without FTT) + VM Swap (without FTT) = Virtual Machine Raw Capacity
Requirements

Virtual Machine Raw Capacity Requirements * FTT = Final Virtual Machine Raw Capacity Requirements
```

```
8,730 GB Disk + 430 GB Swap = 9,160 GB Virtual Machine Raw Capacity Requirements

9,160 GB Virtual Machine Raw Capacity Requirements * 2 (FTT=1, RAID1) = 18,320 GB Final Virtual
Machine Raw Capacity Requirements
```

Derive the requirements for total storage space for the capacity tier according to the following calculations:

```
vSAN Memory Consumption + Final Virtual Machine Raw Capacity Requirements = Total Raw Storage Capacity

Total Raw Storage Capacity * 30% Slack Overhead * 1% On-disk Format Overhead * 0.12% Checksum
Overhead = Raw Unformatted Storage Capacity
OR
Total Raw Storage Capacity * 30% Slack Overhead * 1% On-disk Format Overhead * 0.12% Checksum
Overhead * 20% Estimated Growth = Raw Unformatted Storage Capacity (with 20% Growth Capacity)

Raw Unformatted Storage Capacity / ESXi Quantity = Final Raw Storage Capacity per Host
```

Based on the calculations for the vSAN memory consumption and the management virtual machine consumption, calculate the final raw storage capacity required for the cluster and per the ESXi hosts.

```
34 GB vSAN Memory Consumption + 18,334 GB VM Raw Capactiy = 18,368 GB Total Raw Storage Capacity

18,368 GB Total Raw Storage Capacity * 30% Slack Overhead * 1% On-disk Format Overhead * 0.12%
Overhead ≈ 24,146 ≈ 24 TB Raw Unformatted Storage Capacity
24 TB Raw Unformatted Storage Capacity / 4 ESXi hosts ≈ 6 TB Final Raw Storage Capacity per host

18,368 GB Total Raw Storage Capacity * 30% Slack Overhead * 1% On-disk Format Overhead * 0.12%
Overhead * 20% Estimated Growth ≈ 28,975 GB ≈ 29 TB Raw Unformatted Storage Capacity (with 20% Growth
Capacity)
29 TB Raw Unformatted Storage Capacity / 4 ESXi hosts ≈ 8 TB Final Raw Storage Capacity per host
```

Determine the storage space that is required for the caching tier according to the following calculations:

```
Raw Unformatted Storage Capacity  * 50% * 10% = Total Flash Device Capacity

Total Flash Device Capacity / ESXi Quantity = Final Flash Device Capacity per Host
```

```
24 TB Raw Unformatted Storage Capacity * 50% * 10% Cache Required ≈ 1.2 TB Flash Device Capacity
1.2 TB Flash Device Storage Capacity / 4 ESXi Hosts ≈ 300 GB of Flash Device Capacity per Host

29 TB Raw Unformatted Storage Capacity (with 20% Growth Capacity) * 50% * 10% Cache Required ≈ 1.5 TB
```

```
Flash Device Capacity
1.5 TB Flash Device Storage Capacity / 4 ESXi Hosts ≈ 400 GB of Flash Device Capacity per Host
```

Table 2-12. Design Decisions on the vSAN Disk Configuration

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-002	Use one or more 400 GB or greater SSDs and two or more traditional 4 TB or greater HDDs to create at least a single disk group in the management cluster.	Provides enough capacity for the management VMs with a minimum of 10% flash-based caching, 30% of overhead, and 20% growth when using Failures to Tolerate = 1.	When using only a single disk group, you limit the amount of striping (performance) capability and increase the size of the fault domain.

vSAN Hardware Considerations

While VMware supports building your own vSAN cluster from compatible components, vSAN ReadyNodes are selected for this VMware Validated Design.

Build Your Own

Use hardware from the VMware Compatibility Guide for the following vSAN components:

- Solid state disks (SSDs)

- Magnetic hard drives (HDDs)

- I/O controllers, including vSAN certified driver and firmware combinations

Use VMware vSAN ReadyNodes

A vSAN ReadyNode is a server configuration that is validated in a tested, certified hardware form factor for vSAN deployment, jointly recommended by the server OEM and VMware. See the vSAN ReadyNode documentation. The vSAN ReadyNode documentation provides examples of standardized configurations, including supported numbers of VMs and estimated number of 4K IOPS delivered.

According to design decision SDDC-PHY-009, this design uses vSAN ReadyNodes.

SSD Endurance for vSAN

In a vSAN configuration, you use Solid-State Disks (SSDs) for the vSAN caching layer in hybrid deployments and for the capacity layer in all-flash deployments. You consider the endurance parameters, such as Drive Writes Per Day (DWPD) and Terabytes Written (TBW), of a certain SSD class to select best SSD devices according to the requirements for your environment.

Consider the following endurance criteria according to the vSAN configuration:

- For a hybrid deployment, the use of the SSD is split between a non-volatile write cache (approximately 30%) and a read buffer (approximately 70%). As a result, the endurance and the number of I/O operations per second that the SSD can sustain are important performance factors.

- For an all-flash model, endurance and performance have the same criteria. The caching tier performs many more write operations, as result, elongating or extending the life of the SSD capacity tier.

SSD Endurance

This design uses Class D SSDs for the caching tier.

SDDC Endurance Design Decision Background

For endurance of the SSDs used for vSAN, standard industry write metrics are the primary measurements used to gauge the reliability of the drive. No standard metric exists across all vendors. Drive Writes per Day (DWPD) or Terabytes Written (TBW) are normally used as measurements.

For vSphere 5.5, the endurance class was based on Drive Writes Per Day (DWPD). For VMware vSAN 6.0 and later, the endurance class has been updated to use Terabytes Written (TBW), based on the vendor's drive warranty. For more information about which vSAN versions that support TBW, see *VMware Compatibility Guide*.

Using TBW provides the option to use larger capacity drives with lower DWPD specifications.

If an SSD vendor uses DWPD as a measurement, calculate endurance in TBW with the following equation:

```
TBW (over 5 years) = Drive Size x DWPD x 365 x 5
```

For example, if a vendor specified a DWPD of 10 for an 800 GB capacity SSD, you can compute TBW by using the following equation:

```
TBW = 0.8 TB X 10 DWPD X 365 days X 5 yrs
TBW = 14600 TBW
```

As a result, the SSD supports 14600 TB writes over 5 years. The higher the TBW number, the higher the endurance class.

For SSDs that are designated for caching and all-flash capacity layers, consider the following endurance specifications for hybrid and for all-flash vSAN.

Endurance Class	TBW	Hybrid Caching Tier	All-Flash Caching Tier	All-Flash Capacity Tier
Class A	>= 365	No	No	Yes
Class B	>= 1825	Yes	No	Yes
Class C	>= 3650	Yes	Yes	Yes
Class D	>=7300	Yes	Yes	Yes

Table 2-13. Design Decisions on SSD Endurance Class

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-003	Use Class D (>= 7300TBW) SSDs for the caching tier of the management cluster.	If an SSD in the caching tier fails due to wear-out, the entire vSAN disk group becomes unavailable. The result is potential data loss or operational impact.	SSDs with higher endurance can be more expensive than lower endurance classes.

SSD Performance for vSAN

The SSD performance class and the level of vSAN performance are directly correlated. The highest-performing hardware results in the best performance of the solution. Cost is therefore the determining factor. A lower class of hardware that is more cost effective might be attractive even if the performance or size is not ideal.

For optimal performance of vSAN, select SSD Class E or greater . See the VMware Compatibility Guide for detail on the different classes.

SSD Performance Design Decision Background

Select a high class of SSD for optimal performance of vSAN. Before selecting a drive size, consider disk groups, sizing, and expected future growth. VMware defines classes of performance in the VMware Compatibility Guide as follows.

Table 2-14. SSD Performance Classes

Performance Class	Writes Per Second
Class A	2,500–5,000
Class B	5,000 – 10,000
Class C	10,000–20,000
Class D	20,000–30,000
Class E	30,000 – 100,000
Class F	100,000 +

Select an SSD size that is, at a minimum, 10% of the capacity tier drives, before failures to tolerate are considered. For example, select an SSD of at least 100 GB for a 1 TB disk group.

Caching Algorithm

Both hybrid and all-flash configurations adhere to the recommendation that 10% of consumed capacity is for the flash cache layer. Consider the following differences between the two configurations.

Hybrid vSAN

70% of the available cache is allocated for storing frequently read disk blocks, minimizing accesses to the slower magnetic disks. 30% of available cache is allocated to writes.

All-Flash vSAN

All-flash clusters have two types of flash: fast and durable write cache, and cost-effective capacity flash. In this configuration, cache is 100% allocated for writes, as read performance from capacity flash is sufficient.

Use Class E SSDs or greater for the highest possible level of performance from the vSAN volume.

Table 2-15. SSD Performance Class Selection

Design Quality	Option 1 Class E	Option 2 Class C	Comments
Availability	o	o	Neither design option impacts availability.
Manageability	o	o	Neither design option impacts manageability.
Performance	↑	↓	The higher storage class is used, the better the performance.
Recover-ability	o	o	Neither design option impacts recoverability.
Security	o	o	Neither design option impacts security.

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

Table 2-16. Design Decisions on SSD Performance Class

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-004	Use Class E SSDs (30,000-100,000 writes per second) for the management cluster.	Because of the storage I/O performance requirements in the management cluster, you need at least Class E SSDs.	Class E SSDs can be more expensive than lower class drives.

Magnetic HDD Characteristics for vSAN

The hard disk drives (HDDs) in a vSAN environment have two different purposes, capacity and object stripe width.

Capacity

Magnetic disks, or HDDs, unlike caching-tier SSDs, make up the capacity of a vSAN hybrid mode datastore.

Stripe Width

You can define stripe width at the virtual machine policy layer. vSAN might use additional stripes when making capacity and placement decisions outside a storage policy.

vSAN supports these disk types:

- Serial Attached SCSI (SAS)

- Near Line Serial Attached SCSI (NL-SCSI). Consider NL-SAS as enterprise SATA drives but with a SAS interface.

- Serial Advanced Technology Attachment (SATA). Use SATA magnetic disks only in capacity-centric environments where performance is not a priority.

Use SAS and NL-SAS for best performance. This VMware Validated Design uses 10,000 RPM drives for a balance between cost and availability.

HDD Capacity, Cost, and Availability Background Considerations

You can achieve the best results with SAS and NL-SAS.

The VMware vSAN design must consider the number of magnetic disks required for the capacity layer, and how well the capacity layer performs.

- SATA disks typically provide more capacity per individual drive, and tend to be less expensive than SAS drives. However, the trade-off is performance, because SATA performance is not as good as SAS performance due to lower rotational speeds (typically 7200 RPM)

- In environments where performance is critical, choose SAS magnetic disks instead of SATA magnetic disks.

Consider that failure of a larger capacity drive has operational impact on the availability and recovery of more components.

Rotational Speed (RPM) Background Considerations

HDDs tend to be more reliable, but that comes at a cost. SAS disks can be available up to 15,000 RPM speeds.

Table 2-17. vSAN HDD Environmental Characteristics

Characteristic	Revolutions per Minute (RPM)
Capacity	7,200
Performance	10,000
Additional Performance	15,000

Cache-friendly workloads are less sensitive to disk performance characteristics; however, workloads can change over time. HDDs with 10,000 RPM are the accepted norm when selecting a capacity tier.

For the software-defined storage module, use an HDD configuration that is suited to the characteristics of the environment. If there are no specific requirements, selecting 10,000 RPM drives achieves a balance between cost and availability.

Table 2-18. Design Decisions on HDD Selection

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-005	Use 10,000 RPM HDDs for the management cluster.	10,000 RPM HDDs provide a balance between performance and availability for the vSAN configuration. The performance of 10,000 RPM HDDs avoids disk drain issues. In vSAN hybrid mode, vSAN periodically flushes uncommitted writes to the capacity tier.	Limits hardware choice. Using slower and potentially cheaper HDDs is not recommended.

I/O Controllers for vSAN

The I/O controllers are as important to a vSAN configuration as the selection of disk drives. vSAN supports SAS, SATA, and SCSI adapters in either pass-through or RAID 0 mode. vSAN supports multiple controllers per ESXi host.

You select between single- and multi-controller configuration in the following way:

- Multiple controllers can improve performance, and mitigate a controller or SSD failure to a smaller number of drives or vSAN disk groups.

- With a single controller, all disks are controlled by one device. A controller failure impacts all storage, including the boot media (if configured).

Controller queue depth is possibly the most important aspect for performance. All I/O controllers in the *VMware vSAN Hardware Compatibility Guide* have a minimum queue depth of 256. Consider regular day-to-day operations and increase of I/O because of virtual machine deployment operations, or re-sync I/O activity as a result of automatic or manual fault remediation.

NFS Physical Design

You can use NFS as a secondary storage in the SDDC. When you design the physical NFS configuration, consider disk type and size, networking between the storage and the ESXi hosts, and volumes according to the data you are going to store. In this validated design, NFS stores VM templates, backups and log archives.

The management cluster uses vSAN for primary storage. When configured for a single availability zone, you use NFS for secondary storage. The shared edge and compute cluster and the compute clusters are not restricted to any particular storage technology. For compute clusters, decide on the technology for secondary storage according to the performance, capacity, and capabilities (replication, deduplication, compression, etc.) required by the workloads that are running in the clusters.

Table 2-19. Design Decisions on NFS Use

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-006	When using a single availability zone, provide NFS storage.	Separates primary virtual machine storage from backup data in case of primary storage failure. vRealize Log Insight archiving requires an NFS export. NFS storage provides the following features: ■ A datastore for backup data ■ An export for archive data ■ A datastore for templates and ISOs	An NFS capable external array is required. You cannot configure multiple availability zones to use an NFS array in the event an availability zone fails.
SDDC-PHY-STO-007	When using two availability zones, store templates and ISO files on the primary vSAN datastore.	To support an availability zone failure and continue provisioning operations, the templates and ISO files must be available in the surviving availability zone.	Templates and ISO files are on the primary vSAN storage.
SDDC-PHY-STO-008	When using two availability zones, provide virtual machine backups in both availability zones.	To support backup and restore operations in the event of an availability zone failure, backup targets must be available in both availability zones.	The cost of the backup solution increases.

NFS Physical Requirements

To use NFS storage in the VMware Validated Design, your environment must meet certain requirements for networking and bus technology.

■ All connections are made using a minimum of 10 Gb Ethernet, 25 Gb Ethernet recommended.

■ Jumbo frames are enabled.

■ 10K SAS (or faster) drives are used in the storage array.

You can combine different disk speeds and disk types in an array to create different performance and capacity tiers. The management cluster uses 10K SAS drives in the RAID configuration recommended by the array vendor to achieve the required capacity and performance.

Table 2-20. Design Decisions on NFS Hardware

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-009	Use 10K SAS drives for NFS volumes.	10K SAS drives provide a balance between performance and capacity. You can use faster drives. ■ vStorage API for Data Protection-based backups require high-performance datastores to meet backup SLAs. ■ vRealize Automation uses NFS datastores for its content catalog which requires high-performance datastores. ■ vRealize Log Insight uses NFS datastores for its archive storage which, depending on compliance regulations, can use a large amount of disk space.	10K SAS drives are more expensive than other alternatives.

NFS Volumes

Select a volume configuration for NFS storage in the SDDC according to the requirements of the management applications that use the storage.

Multiple datastores can be created on a single volume for applications that do not have a high I/O footprint.

■ For high I/O applications, such as backup applications, use a dedicated volume to avoid performance issues.

■ For other applications, set up Storage I/O Control (SIOC) to impose limits on high I/O applications so that other applications get the I/O they are requesting.

Table 2-21. Design Decisions on Volume Assignment

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-010	Use a dedicated NFS volume to support backup requirements.	The backup and restore process is I/O intensive. Using a dedicated NFS volume ensures that the process does not impact the performance of other management components.	Dedicated volumes add management overhead to storage administrators. Dedicated volumes might use more disks, according to the array and type of RAID.
SDDC-PHY-STO-011	Use a shared volume for other management component datastores.	Non-backup related management applications can share a common volume due to the lower I/O profile of these applications.	Enough storage space for shared volumes and their associated application data must be available.

Virtual Infrastructure Design

The virtual infrastructure design includes the software components that make up the virtual infrastructure layer for providing software-defined storage, networking and compute.

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

Figure 2-5. Virtual Infrastructure Layer in the SDDC

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 shared edge and compute cluster.

Figure 2-6. SDDC Single Availability Zone Logical Design

Management Cluster and Shared Edge and Compute Cluster

Figure 2-7. SDDC Logical Design with Two Availability Zones

Availability Zone 1 - Management and Shared Edge and Compute Clusters

Availability Zone 2 (Secondary) - Management and Shared Edge and Compute Clusters

Management Cluster

The management cluster runs the virtual machines that manage the SDDC. These virtual

machines host vCenter Server, NSX Manager, NSX Controllers, vRealize Lifecycle Manager, vRealize Operations Manager, 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 which provides high availability for these critical services. Permissions on the management cluster limit access to only administrators. In this way, you protect the virtual machines running the management, monitoring, and infrastructure services.

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 required.

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.

ESXi Design

For the design of the configuration of the ESXi hosts, consider boot options, user access, and the virtual machine swap configuration.

ESXi Hardware Requirements

For the ESXi hardware requirements, see Physical Design Fundamentals .

ESXi Manual Install and Boot Options

You can install or boot ESXi from the following storage systems:

SATA disk drives

SATA disk drives connected behind supported SAS controllers or supported on-board SATA controllers.

Serial-attached SCSI (SAS) disk drives

Supported for ESXi installation

SAN

Dedicated SAN disk on Fibre Channel or iSCSI.

USB devices

Supported for ESXi installation. Use a 16-GB SD card or larger.

FCoE

Dedicated FCoE LUN. You can use a VMware software FCoE adapter and a network adapter with FCoE capabilities. A dedicated FCoE HBA is not required.

ESXi can boot from a disk larger than 2 TB if the system firmware and the firmware on any add-in card support it. See the vendor documentation.

ESXi Boot Disk and Scratch Configuration

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

An ESXi installation on a USB media does not configure a default scratch partition. Specify a scratch partition on a shared datastore and configure remote syslog logging for the ESXi host.

Table 2-22. Design Decision on the ESXi Boot Disk

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-ESXi-001	Install and configure all ESXi hosts to boot using an SD device of 16 GB or greater.	SD cards are an inexpensive and easy to configure as an option for installing ESXi. Using SD cards allows allocation of all local HDDs to a vSAN storage system.	When you use SD cards, ESXi logs are not retained locally.

ESXi Host Access

After installation, you add ESXi hosts to a vCenter Server system and manage them by using the vCenter Server system.

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

Direct Console User Interface (DCUI)

Graphical interface on the console. Provides basic administrative controls and troubleshooting options.

ESXi Shell

A Linux-style bash login on the ESXi console itself.

Secure Shell (SSH) Access

Remote command-line console access.

VMware Host Client

HTML5-based client that has a similar interface to the vSphere Web Client but for managing individual ESXi hosts only. You use the VMware Host Client for emergency management when vCenter Server is temporarily unavailable

You can enable or disable each method. By default, the ESXi Shell and SSH are disabled to protect 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, you can grant access through Active Directory groups. Auditing logins in to the ESXi host also becomes easier.

Table 2-23. Design Decisions on ESXi User Access

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 provides greater flexibility in granting access to ESXi hosts. Ensuring that users log in with a unique user account provides 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, you can separate management tasks using different groups.	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 contents of the virtual machine's RAM. The default swap file is stored in the same location as the configuration file of the virtual machine. The colocation simplifies the configuration, however, it can cause an excess of replication traffic that is not needed.

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

ESXi Design Decisions about NTP and Lockdown Mode Configuration

Table 2-24. Other Design Decisions on the ESXi Host Configuration

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-ESXi-004	Configure all ESXi hosts to synchronize time with the central NTP servers.	The deployment of 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.
SDDC-VI-ESXi-005	Enable Lockdown mode on all ESXi hosts.	You increase the security of ESXi hosts by requiring that administrative operations be performed only from vCenter Server.	Lockdown mode settings are not part of vSphere host profiles and must be manually enabled on all hosts.

vCenter Server Design

The vCenter Server design includes the design for both vCenter Server and Platform Services Controller. Determine the number of instances, their size, networking configuration, cluster layout, redundancy, and security configuration.

A Platform Services Controller instance groups a set of infrastructure services including vCenter Single Sign-On, License service, Lookup Service, and VMware Certificate Authority (VMCA).

- vCenter Server Deployment

  A vCenter Server deployment can consist of one or several vCenter Server and Platform Services Controller instances according to the scale, number of virtual machines and continuity requirements for your environment.

- 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

  You size resources and storage for the Management vCenter Server Appliance and the Compute vCenter Server Appliance to provide enough resources for accommodating the expected number of management virtual machines in the SDDC.

- vSphere Cluster Design

  The cluster design must consider the workloads that the cluster handles. Different cluster types in this design have different characteristics.

- vCenter Server Customization

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

- Use of TLS Certificates in vCenter Server

  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

A vCenter Server deployment can consist of one or several vCenter Server and Platform Services Controller instances according to the scale, number of virtual machines and continuity requirements for your environment.

You also determine the type of installation and the topology of the vCenter Server and Platform Services Controller instances.

Table 2-25. Design Decisions on the Number of vCenter Server Instances

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-001	Deploy two vCenter Server systems in the first availability zone of each region. ■ One vCenter Server supporting the SDDC management components. ■ One vCenter Server supporting the edge components and tenant workloads.	Isolates vCenter Server failures to management or tenant workloads. Isolates vCenter Server operations between management and tenants. Supports a scalable cluster design where you might reuse the management components as more tenant workloads are added to the SDDC. Simplifies capacity planning for tenant workloads because you do not consider management workloads for 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 you are upgrading the tenant workloads ■ Tenant workloads remain available while you are upgrading the management nodes Supports clear separation of roles and responsibilities to ensure that only 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 one for the shared edge and compute cluster. Network separation of the clusters in the SDDC provides isolation of potential network issues.	Requires licenses for each vCenter Server instance.

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

Table 2-26. Design Decisions on the vCenter Server Platform

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

Platform Services Controller Design Decisions

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

- In an embedded deployment, the vCenter Server instance and the Platform Services Controller instance run on the same virtual machine.

- In an environment with an external Platform Services Controller, multiple vCenter Server systems can share the same Platform Services Controller services. For example, several vCenter Server systems can use the same instance of vCenter Single Sign-On for authentication.

Table 2-27. Design Decisions on Platform Service Controller

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-003	Deploy two external Platform Services Controller instances.	You can implement high availability of their services by placing a load balancer in front of the external Platform Services Controller instances. As a result, the replication between the Platform Services Controller instances is still available after you add or remove vCenter Servers.	The number of VMs that have to 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 in to 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 Service Controller instances.	By default, one Platform Service Controller instance replicates only with another Platform Services Controller instance. This setup creates a single point of failure for replication. A ring topology ensures that each Platform Service Controller instance has two replication partners and removes any single point of failure.	You use command-line interface commands to configure the ring replication topology.
SDDC-VI-VC-006	Use an NSX Edge services gateway as a load balancer for the Platform Services Controller instances.	Using a load balancer increases the availability of the Platform Services Controller instances for all applications.	Configuring the load balancer introduces administrative overhead.

Figure 2-8. 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:

- Platform Services Controller

- 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.

You can use the following methods for protecting a vCenter Server instance on Windows and a vCenter Server Appliance:

Table 2-28. Methods for Protecting the vCenter Server Node 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
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	Yes, but only if the Platform Services Controller instance is embedded

Table 2-29. Design Decisions on vCenter Server Protection

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 the 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

You size resources and storage for the Management vCenter Server Appliance and the Compute vCenter Server Appliance to provide enough resources for accommodating the expected number of management virtual machines in the SDDC.

Table 2-30. Resource Specification of the Management vCenter Server Appliance

Attribute	Specification
vCenter Server version	6.7 Update 1 (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	16 GB
Disk Space	340 GB

Table 2-31. Resource Specification of the Compute vCenter Server Appliance

Attribute	Specification
vCenter Server version	6.7 Update 1 (vCenter Server Appliance)
Physical or virtual system	Virtual (appliance)

Table 2-31. Resource Specification of the Compute vCenter Server Appliance (continued)

Attribute	Specification
Appliance Size	Large (up to 1,000 hosts / 10,000 VMs)
Platform Services Controller	External
Number of CPUs	16
Memory	32 GB
Disk Space	740 GB

Table 2-32. Design Decisions on vCenter Server Appliance Sizing

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-008	Deploy Management vCenter Server Appliances of a small deployment size or larger.	Based on the number of management VMs that are running, a vCenter Server Appliance of a small-size is sufficient.	If the size of the management environment grows, you might have to increase the vCenter Server Appliance size.
SDDC-VI-VC-009	Deploy Compute vCenter Server Appliances of a large deployment size or larger.	Based on the number of compute workloads and NSX Edge devices running, a vCenter Server Appliance of a large size is best.	As the tenant environment expands, you might have to resize to the X-Large size or add vCenter Server instances.

vSphere Cluster Design

The cluster design must consider the workloads that the cluster handles. Different cluster types in this design have different characteristics.

When you design the cluster layout in vSphere, consider the following guidelines:

- 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. Costs vary between vendors and models.

- Evaluate the operational costs for 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 2-9. vSphere Logical Cluster Layout with a Single Availability Zone

Figure 2-10. vSphere Logical Cluster Layout with Two Availability Zones

- **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 Cluster Workload Design**

- **Management Cluster Design**

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

- **Shared Edge and Compute Cluster Design**

  Tenant workloads run on the ESXi hosts in the shared edge and compute cluster. Because of the shared nature of the cluster, NSX Controllers and Edge devices run in this cluster. This cluster design determines the number of ESXi hosts, vSphere HA settings, and several other characteristics of the shared edge and compute cluster.

■ Compute Cluster Design

As the SDDC expands, you can configure additional compute-only clusters. 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.

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 primary ESXi host. The primary ESXi host communicates with the vCenter Server system and monitors the virtual machines and secondary ESXi hosts in the cluster.

The primary 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 2-33. Design Decisions on vSphere HA

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-010	Use vSphere HA to protect all virtual machines against failures.	vSphere HA supports a robust level of protection for both ESXi host and virtual machine availability.	You must provide sufficient resources on the remaining hosts so that virtual machines can be migrated to those hosts in the event of a host outage.
SDDC-VI-VC-011	Set vSphere HA Host Isolation Response to Power Off.	vSAN requires that the HA Isolation Response be set to Power Off and to restart VMs on available ESXi hosts.	VMs are powered off in case of a false positive and an ESXi host is declared isolated incorrectly.

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 ensures 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, 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 2-34. Design Decisions on vSphere Clusters

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-012	Create a single management cluster per region. This cluster contains all management ESXi hosts.	■ Simplifies configuration by isolating management workloads from tenant workloads. ■ Ensures that tenant 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.
SDDC-VI-VC-013	Create a shared edge and compute cluster per region. This cluster contains tenant workloads, NSX Controller nodes and associated NSX Edge gateway devices used for tenant workloads.	■ Creating a shared cluster for tenant edge devices and tenant virtual machines simplifies configuration and minimizes the number of ESXi hosts required for initial deployment. ■ 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. Because of the shared nature of the cluster, when you add tenant workloads, the cluster must be scaled out to keep high level of network performance. Due to the shared nature of the cluster, resource pools are required to provide all required resources to edge components .

Management Cluster Design

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

Table 2-35. Design Decisions on the Management Cluster

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-014	In Region A, create a management cluster with a minimum of 4 ESXi hosts for a single availability zone or with a minimum of 8 ESXi hosts for two availability zones (minimum of 4 ESXi hosts in each availability zone).	Allocating 4 ESXi hosts provides full redundancy for each availability zone in the cluster. Having 4 ESXi hosts in each availability zone guarantees vSAN and NSX redundancy during availability zone outages or maintenance operations.	Additional ESXi host resources are required for redundancy.
SDDC-VI-VC-015	In Region B, create a management cluster with a minimum of 4 ESXi hosts.	Allocating 4 ESXi hosts provides full redundancy for the cluster. Having four ESXi hosts guarantees vSAN and NSX redundancy during maintenance operations.	Additional ESXi host resources are required for redundancy.
SDDC-VI-VC-016	When using a single availability zone, configure Admission Control for 1 ESXi host failure and percentage-based failover capacity.	Using the percentage-based reservation works well in situations where virtual machines have varying and sometime significant CPU or memory reservations. vSphere automatically calculates the reserved percentage based on ESXi host failures to tolerate and the number of ESXi hosts in the cluster.	In a management cluster of 4 ESXi hosts, only the resources of 3 ESXi hosts are available for use.
SDDC-VI-VC-017	When using two availability zones, configure Admission Control for percentage-based failover based on half of the ESXi hosts in the cluster.	Allocating only half of a stretched cluster ensures that all VMs have enough resources if an availability zone outage occurs.	In a management cluster of 8 ESXi hosts, only the resources of 4 ESXi hosts are available for use. If you add more ESXi hosts to the management cluster, add them in pairs, one in each availability zone.
SDDC-VI-VC-018	When using two availability zones, set the cluster isolation addresses for the cluster to the gateway IP address for the vSAN network in both availability zones.	Allows vSphere HA to validate complete network isolation in the case of a connection failure between availability zones.	You must manually configure the isolation address.
SDDC-VI-VC-019	When using two availability zones, set the advanced cluster setting `das.usedefaultisolationaddress` to false.	Ensures that the manual isolation addresses are used instead of the default management network gateway address.	None.

Table 2-35. Design Decisions on the Management Cluster (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-020	When using a single availability zone, create a host profile for the management cluster.	Simplifies configuration of ESXi hosts and ensures that the settings are uniform across the cluster.	Anytime an authorized change to an ESXi host is made, you must update the host profile to reflect the change or the status will show non-compliant.
SDDC-VI-VC-021	When using two availability zones, create a host profile for each availability zone in the management cluster.	Simplifies configuration of ESXi hosts and ensures that the settings are uniform across the availability zones in the cluster.	Anytime an authorized change to an ESXi host is made, the host profile must be updated to reflect the change or the status will show non-compliant. Because of configuration differences between availability zones, two host profiles are required and must be applied on each ESXi host.

The management cluster logical design has the following attributes:

Table 2-36. Management Cluster Logical Design Background

Attribute	Specification
Number of ESXi hosts required to support management virtual machines with no memory over commitment	3
Number of ESXi hosts recommended because of operational constraints (ability to take a host offline without sacrificing high availability capabilities)	4
Number of ESXi hosts recommended because of operational constraints, while using vSAN (ability to take a host offline without sacrificing high availability capabilities) .	■ 4 (single availability zone) ■ 8 (two availability zones)
Capacity for ESXi host failures per cluster	■ 25% reserved CPU RAM (single availability zone) ■ 50% reserved CPU RAM (two availability zones)

Shared Edge and Compute Cluster Design

Tenant workloads run on the ESXi hosts in the shared edge and compute cluster. Because of the shared nature of the cluster, NSX Controllers and Edge devices run in this cluster. This cluster design determines the number of ESXi hosts, vSphere HA settings, and several other characteristics of the shared edge and compute cluster.

Table 2-37. Design Decisions on the Shared Edge and Compute Cluster

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-022	Create a shared edge and compute cluster for the NSX Controller nodes and NSX Edge gateway devices.	NSX Manager requires a one-to-one relationship with a vCenter Server system.	■ Each time you provision a Compute vCenter Server system, a new NSX Manager is required. ■ You set anti-affinity rules to keep each controller on a separate ESXi host. A 4-node cluster supports maintenance while ensuring that the 3 controllers remain on separate ESXi hosts.
SDDC-VI-VC-023	When using a single availability zone, configure Admission Control for 1 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-024	When using two availability zones, configure Admission Control for percentage-based failover based on half of the ESXi hosts in the cluster. For example, in a cluster with 8 ESXi hosts you configure admission control for 4 ESXi hosts 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 half of a stretched cluster should be used to ensure that all VMs have enough resources in an availability zone outage.	You must add ESXi hosts to the cluster in pairs, one in each availability zone.
SDDC-VI-VC-025	In Region A, create a shared edge and compute cluster with a minimum of 4 ESXi hosts for a single availability zone or with a minimum of 8 ESXi hosts for two availability zones (minimum of 4 ESXi hosts in each availability zone).	Allocating 4 ESXi hosts provides full redundancy for each availability zone within the cluster. Having 4 ESXi hosts in each availability zone guarantees vSAN and NSX redundancy during availability zone outages or maintenance operations.	4 ESXi hosts is the smallest starting point for a single availability zone and 8 ESXi hosts for two availability zones for the shared edge and compute cluster for redundancy and performance thus increasing cost.

Table 2-37. Design Decisions on the Shared Edge and Compute Cluster (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-026	In Region B, create a shared edge and compute cluster with a minimum of 4 hosts.	■ 3 NSX Controller nodes are required for sufficient redundancy and majority decisions. ■ 1 ESXi host is available for failover and to allow for scheduled maintenance.	4 ESXi hosts is the smallest starting point for the shared edge and compute cluster for redundancy and performance thus increasing cost over 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 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 nodes and edge appliances with a CPU share level of High, a memory share of Normal, and 16 GB 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, these virtual machines must receive all the resources required.	During contention, SDDC NSX components receive more resources than all other workloads. As a result, monitoring and capacity management of tenant workloads 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 edges for users, created by vRealize Automation, support functions such as load balancing for user workloads. These edge devices do not support the entire SDDC as such they receive a lower amount of resources during contention.	During contention, these NSX edges will 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 will have a negative impact on all other virtual machines during contention. In a shared edge and compute cluster, the SDDC edge devices must be guaranteed resources before all other workloads as to retain network connectivity. Setting the share values to Normal gives the SDDC edges more shares of resources during contention ensuring network traffic is not impacted.	During contention, tenant virtual machines might require resources and experience poor performance. Proactively perform monitoring and capacity management, add capacity or dedicate an edge cluster before contention occurs.
SDDC-VI-VC-031	When using two availability zones, set the cluster isolation addresses for the cluster to the gateway IP addresses for the vSAN network in both availability zones.	vSphere HA can validate complete network isolation in the case of a connection failure between availability zones.	You must manually configure the isolation address.

Table 2-37. Design Decisions on the Shared Edge and Compute Cluster (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-032	When using two availability zones, set the advanced cluster setting `das.usedefaultisolationaddress` to false.	Ensures that the manual isolation addresses are used instead of the default management network gateway address.	None.
SDDC-VI-VC-033	When using a single availability zone, create a host profile for the shared edge and compute cluster.	Simplifies the configuration of ESXi hosts and ensures that the settings are uniform across the cluster.	Anytime an authorized change to an ESXi host is made, you must update the host profile to reflect the change or the status will show non-compliant.
SDDC-VI-VC-034	When using two availability zones, create a host profile for each availability zone in the cluster.	Simplifies configuration of ESXi hosts and ensures that the settings are uniform across the availability zones in the cluster.	Anytime an authorized change to an ESXi host is made, you must update the host profile to reflect the change or the status will show non-compliant. Because of configuration differences between availability zones, two host profiles are required and must be applied on each ESXi host.

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

Table 2-38. 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 because of operational constraints (ability to take an ESXi host offline without sacrificing high availability capabilities)	4
Number of ESXi hosts recommended because of operational constraints, while using vSAN (ability to take an ESXi host offline without sacrificing high availability capabilities)	■ 4 (single availability zone) ■ 8 (two availability zones)
Capacity for ESXi host failures per cluster	■ 25% reserved CPU RAM (single availability zone) ■ 50% reserved CPU RAM (two availability zones)

Compute Cluster Design

As the SDDC expands, you can configure additional compute-only clusters. 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 2-39. Design Decisions on the Compute Cluster

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-035	For a single availability zone, 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 is reserved which can cause provisioning failure if resources are exhausted.
SDDC-VI-VC-036	When using two availability zones, configure Admission Control for percentage-based failover based on half of the ESXi hosts in the cluster. For example, in a cluster with 8 ESXi hosts you configure admission control for 4 ESXi hosts failure and percentage-based failover capacity.	Only half of a stretched cluster should be used to ensure that all VMs have enough resource in the event of an availability zone outage.	If you add more ESXi hosts to the compute cluster, you must add them in pairs, one in each availability zone.

vCenter Server Customization

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

VM 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 that is running on the guest OS. If the service receives no heartbeats or I/O activity, the guest operating system has likely failed or VMware Tools is not being 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 VM 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 2-40. Design Decisions on Monitoring and Startup Order Configuration for Virtual Machines

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-037	Enable VM Monitoring for each cluster.	VM Monitoring provides in-guest protection for most VM workloads.	None.
SDDC-VI-VC-038	Create virtual machine groups for use in startup rules in the management and shared edge and compute clusters.	By creating virtual machine groups, you can use rules to configure the startup 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 startup order of the SDDC management components.	Rules enforce the startup order of virtual machine groups, hence, the startup order of the SDDC management components.	Creating the rules is a manual task and adds administrative overhead.

VMware vSphere Distributed Resource Scheduling

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

Manual

vSphere DRA provides recommendations 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 can provide a slight improvement in balancing is performed.

When using two availability zones, enable vSphere DRS to create VM/Host group affinity rules for initial placement of VMs and impacting read locality. In this way, you avoid unnecessary vSphere vMotion migration of VMs between availability zones. Because the vSAN stretched cluster is still a single cluster, vSphere DRS is unaware that it stretches across different physical locations. As result, it might decide to move virtual machines between them. By using VM/Host group affinity rules, you can constrain virtual machines to an availability zone. Otherwise, if a virtual machine, VM1, that resides in Availability Zone 1, moves across availability zones, VM1 could eventually be running on Availability Zone 2. Because vSAN stretched clusters implement read locality, the cache for the virtual machine in Availability Zone 1 is warm whereas the cache in Availability Zone 2 is cold. This situation might impact the performance of VM1 until the cache for it in Availability Zone 2 is warmed up.

Table 2-41. Design Decisions on vSphere DRS

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).	Provides 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.
SDDC-VI-VC-041	When using two availability zones, create a host group and add the ESXi hosts in Region A - Availability Zone 1 to it.	Makes it easier to manage which virtual machines should run in which availability zone.	You must create and maintain VM/Host DRS group rules.
SDDC-VI-VC-042	When using two availability zones, create a host group and add the ESXi hosts in Region A - Availability Zone 2 to it.	Makes it easier to manage which virtual machines should run in which availability zone.	You must create and maintain VM/Host DRS group rules.
SDDC-VI-VC-043	When using two availability zones, create a virtual machine group and add the virtual machines in Region A - Availability Zone 1 to it.	Ensures that virtual machines are located only in the assigned availability zone.	You must add VMs to the allocated group manually to ensure they are not powered-on in or migrated to the wrong availability zone.
SDDC-VI-VC-044	When using two availability zones, create a virtual machine group and add the virtual machines in Region A - Availability Zone 2 to it.	Ensures that virtual machines are located only in the assigned availability zone.	You must add VMs to the allocated group manually to ensure they are not powered-on in or migrated to the wrong availability zone.

Enhanced vMotion Compatibility

Enhanced vMotion Compatibility (EVC) works by masking certain features of newer CPUs to allow migration between ESXi hosts containing 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 later without disruption. You can use EVC for a rolling upgrade of all hardware with zero downtime.

Set the cluster EVC mode to the highest available baseline that is supported for the lowest CPU architecture on the hosts in the cluster.

Table 2-42. Design Decisions on VMware Enhanced vMotion Compatibility

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-045	Enable Enhanced vMotion Compatibility (EVC) on all clusters.	Supports cluster upgrades without virtual machine downtime.	You can enable EVC only if clusters contain hosts with CPUs from the same vendor.
SDDC-VI-VC-046	Set the cluster EVC mode to the highest available baseline that is supported for the lowest CPU architecture on the hosts in the cluster.	Supports cluster upgrades without virtual machine downtime.	None.

Use of TLS Certificates in vCenter Server

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.

As a security best practice, replace at least all user-facing certificates with certificates that are signed by a third-party or enterprise Certificate Authority (CA). Certificates for machine-to-machine communication can remain VMCA-signed.

Table 2-43. Design Decisions on the TLS Certificates of vCenter Server

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-046	Replace the vCenter Server machine certificate and Platform Services Controller machine certificate with a certificate signed by a third-party Public Key Infrastructure.	Infrastructure administrators connect to both vCenter Server and the Platform Services Controller using a Web browser to perform configuration, management, and troubleshooting activities. Using the default certificate results in certificate warning messages.	Replacing and managing certificates is an operational overhead.
SDDC-VI-VC-047	Use a SHA-2 or higher algorithm when signing certificates.	The SHA-1 algorithm is considered less secure and has been deprecated.	Not all certificate authorities support SHA-2.

Virtualization Network Design

A well-designed network helps the 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

VMware Validated Design 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 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 traffic that is related to the main vSphere features.

- **VXLAN**

  VXLAN provides the capability to create isolated, multi-tenant broadcast domains across data center fabrics, and enables customers to create elastic, logical networks that span physical network boundaries.

- **vMotion TCP/IP Stack**

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

Virtual Network Design Guidelines

VMware Validated Design 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 simpler 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 25-GbE NICs might be more cost effective than configuring four 10-GbE NICs.

- Improve performance. You can achieve performance improvement and decrease the time that is required to perform maintenance by providing sufficient bandwidth, which reduces contention and latency.

- Improve availability. A well-designed network improves availability, usually by providing network redundancy.

- Support security. A well-designed network supports an acceptable level of security through controlled access and isolation, where required.

- 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 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 virtual machines will receive a higher percentage of the 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 a migration using vSphere vMotion occurs, the contents of the memory of the guest operating system is transmitted over the network. You can place 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 significant performance impact.

- For best performance, use VMXNET3 virtual machine NICs.

- Ensure 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, and 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.

According to 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

  - vSAN

  - NFS Storage

  - vSphere Replication

  - VXLAN

- Traffic that supports the services and applications in the organization.

Virtual Switches

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

Virtual Switch Design Background

A distributed switch offers several enhancements over a standard switch such as centralized control plane and support for traffic monitoring features.

Centralized management

Because distributed switches are created and managed centrally on a vCenter Server system, switch configuration is more consistent across ESXi hosts. Centralized management saves time, reduces mistakes, and reduces operational costs.

Additional features

Distributed switches have features that are not available on standard virtual switches.

NetFlow and port mirroring provide monitoring and troubleshooting capabilities to the virtual infrastructure.

To guarantee that traffic types with high priority have enough network capacity, you can assign shares to these traffic types by using Network I/O Control.

To ensure that every network adapter is used when the network traffic is high, you can use the Route Based on Physical NIC Load teaming policy. The distributed switch directs the traffic from one physical network adapter to another if the usage of an adapter remains at or above 75% for 30 seconds.

Disadvantages

If vCenter Server is unavailable, distributed switches are not manageable. Consider vCenter Server a Tier 1 application.

Virtual Switch Number and Configuration

Create a single virtual switch per cluster. For each type of network traffic, configure a port group to simplify configuration and monitoring.

Table 2-44. Design Decisions on Virtual Switch Type and Configuration

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-001	Use vSphere Distributed Switches (VDSs).	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.

Table 2-44. Design Decisions on Virtual Switch Type and Configuration (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-003	Use ephemeral port binding for the management port group.	Using ephemeral port binding provides the option for recovery of the vCenter Server instance that is managing the distributed switch.	Port-level permissions and controls are lost across power cycles, and no historical context is saved.
SDDC-VI-NET-004	Use static port binding for all non-management port groups.	Static binding ensures a virtual machine connects to the same port on the vSphere Distributed Switch. This allows for historical data and port level monitoring .	None.

Health Check

The health check service helps identify and troubleshoot configuration errors in vSphere distributed switches.

Health check helps identify the following common configuration errors.

- Mismatching VLAN trunks between an ESXi host and the physical switches it's connected to.

- Mismatching MTU settings between physical network adapters, distributed switches, and physical switch ports.

- Mismatching 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, determines whether the MTU size configuration for jumbo frames on the physical access switch port matches the distributed switch MTU setting.

Teaming policies

Determines whether the connected access ports of the physical switch that participate in an EtherChannel are paired with distributed ports whose teaming policy is Route based on IP hash.

Health check is limited to the access switch port to which the NICs of the ESXi hosts are connected.

Table 2-45. Design Decisions on Distributed Switch Health Check

Design ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-005	Enable vSphere Distributed Switch Health Check on all distributed switches.	vSphere Distributed Switch Health Check verifies that all VLANS are trunked to all ESXi hosts attached to the vSphere Distributed Switch and MTU sizes match the physical network.	You must have a minimum of two physical uplinks to use this feature. In a multiple availability zone configuration, some VLANs are not available to all ESXi hosts in the cluster which triggers alarms.

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 Switch

The cluster uses a single vSphere distributed switch whose design includes traffic types on the switch, the number of required NICs, jumbo frames configuration, port groups settings, and Network I/O Control settings.

Table 2-46. 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) - vSAN - vSphere vMotion - VXLAN Tunnel Endpoint (VTEP) - vSphere Replication/vSphere Replication NFC - Uplinks (2) for ECMP	Enabled	2	9000

Table 2-47. Configuration Settings of the Management Port Group

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

Figure 2-11. Network Switch Design for Management ESXi Hosts

Table 2-48. Management Virtual Switches by Physical and Virtual NICs

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

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

Table 2-49. 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
sfo01-m01-vds01	sfo01-m01-vds01-vmotion	Route based on physical NIC load	1, 2	1612
sfo01-m01-vds01	sfo01-m01-vds01-vsan	Route based on physical NIC load	1, 2	1613

Table 2-49. Management Virtual Switch Port Groups and VLANs (continued)

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
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	sfo02-m01-vds01-management	Route based on physical NIC load	1, 2	1621
sfo01-m01-vds01	sfo02-m01-vds01-vmotion	Route based on physical NIC load	1, 2	1622
sfo01-m01-vds01	sfo02-m01-vds01-vsan	Route based on physical NIC load	1, 2	1623
sfo01-m01-vds01	sfo02-m01-vds01-uplink01	Route based on originating port	1	2721
sfo01-m01-vds01	sfo02-m01-vds01-uplink02	Route based on originating port	2	2722
sfo01-m01-vds01	sfo02-m01-vds01-nfs	Route based on physical NIC load	1, 2	1625
sfo01-m01-vds01	sfo02-m01-vds01-replication	Route based on physical NIC load	1, 2	1626
sfo01-m01-vds01	Auto Generated (NSX VTEP)	Route based on SRC-ID	1, 2	1614

Table 2-50. Management VMkernel Adapters

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	vSAN	sfo01-m01-vds01-vsan	vSAN	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
sfo01-m01-vds01	vSAN	sfo02-m01-vds01-vsan	vSAN	9000
sfo01-m01-vds01	NFS	sfo02-m01-vds01-nfs	-	9000

Table 2-50. Management VMkernel Adapters (continued)

vSphere Distributed Switch	Network Label	Connected Port Group	Enabled Services	MTU
sfo01-m01-vds01	Replication	sfo02-m01-vds01-replication	vSphere Replication traffic vSphere Replication NFC traffic	9000
sfo01-m01-vds01	VTEP	Auto-generated (NSX VTEP)	-	9000

For more information on the physical network design specifications, see Physical Networking Design .

Shared Edge and Compute Cluster Distributed Switches

The shared edge and compute cluster uses a single vSphere distributed switch whose design includes traffic types on the switch, number of required NICs, jumbo frames configuration, port groups settings, and Network I/O Control settings.

Table 2-51. 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 ■ VXLAN Tunnel Endpoint (VTEP) ■ Uplinks (2) to enable ECMP ■ vSAN	Enabled	2	9000

Table 2-52. Configuration Settings of the Shared Edge and Compute Port Groups

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

When you design the switch configuration on the ESXi hosts, consider the physical NIC layout and physical network attributes.

Figure 2-12. Network Switch Design for Shared Edge and Compute ESXi Hosts

Table 2-53. Shared Edge and Compute Cluster Virtual Switches by Physical or 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 2-54. Shared Edge and Compute Cluster 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
sfo01-w01-vds01	sfo01-w01-vds01-vsan	Route based on physical NIC load	1, 2	1633

Table 2-54. Shared Edge and Compute Cluster Port Groups and VLANs (continued)

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
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-vsan	Route based on physical NIC load	1, 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 2-55. VMkernel Adapters for the Shared Edge and Compute Cluster

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	VSAN	sfo01-w01-vds01-vsan	vSAN	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	VSAN	sfo02-w01-vds01-vsan	vSAN	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 Physical Networking Design .

Compute Cluster Distributed Switches

A compute cluster uses a single vSphere distributed switch whose design includes traffic types on the switch, number of required NICs, jumbo frames configuration, port groups settings, and Network I/O Control settings.

Table 2-56. 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 ■ VXLAN Tunnel Endpoint (VTEP)	Enabled	2	9000

Table 2-57. Configuration Settings for Compute Port Groups

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 2-13. Network Switch Design for Compute ESXi Hosts

You also design the configuration of the physical NICs, NIC teaming and VLAN IDs.

Table 2-58. Virtual Switches for the Compute Cluster by Physical or 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 2-59. Port Groups and VLANs on the Virtual Switch for the Compute Cluster

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	Auto Generated (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 2-60. 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 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 provides options for load balancing of traffic. To reduce further 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.

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 2-61. NIC Teaming and Policy

Design Quality	Active-Active	Active-Passive	Comments
Availability	↑	↑	Using teaming regardless of the option increases the availability of the environment.
Manageability	o	o	Neither design option impacts manageability.
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 impacts recoverability.
Security	o	o	Neither design option impacts security.

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

Table 2-62. Design Decision on NIC Teaming

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-006	Use the Route based on physical NIC load teaming algorithm for all port groups except for ECMP uplinks and ones that carry VXLAN traffic. ECMP uplink port groups use Route based on originating virtual port. VTEP kernel ports and VXLAN traffic use Route based on SRC-ID.	Reduces the complexity of the network design and increases resiliency and performance.	Because NSX does not support 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 traffic that is related to the main vSphere features.

- Fault tolerance traffic

- iSCSI traffic

- vSphere vMotion traffic

- Management traffic

- VMware vSphere Replication traffic

- NFS traffic

- vSAN traffic

- Backup traffic

- Virtual machine traffic

Network I/O Control Heuristics

The following heuristics can help with design decisions for Network I/O Control.

Shares and 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 Network Resource Pools

Consider imposing limits on a given network 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 result occasionally.

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 where the adapter is connected to.

How Network I/O Control Works

Network I/O Control enforces the share value specified for the different traffic types when a network 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, granting access to more network resources to more important traffic types.

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

Network I/O Control Design Decisions

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

Table 2-63. Design Decisions on Network I/O Control

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-007	Enable Network I/O Control on all distributed switches.	Increases resiliency and performance of the network.	If configured incorrectly, Network I/O Control might impact 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, vSphere 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.
SDDC-VI-NET-010	Set the share value for vSAN traffic to High.	During times of network contention, vSAN traffic needs a 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 vSAN traffic. Management traffic is important because it ensures 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 vSAN traffic, by prioritizing it lower vSAN is not impacted.	During times of network contention, backups and log archiving are slower than usual.
SDDC-VI-NET-013	Set the share value for backup traffic to Low.	During times of network contention, the primary functions of the SDDC must continue to have access to network resources with priority 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.

Table 2-63. Design Decisions on Network I/O Control (continued)

Decision ID	Design Decision	Design Justification	Design Implication
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 the lowest priority.	None.
SDDC-VI-NET-016	Set the share value for iSCSI traffic to Low.	This design does not use iSCSI. iSCSI traffic can be set the lowest priority.	None.

VXLAN

VXLAN provides the capability to create isolated, multi-tenant broadcast domains across data center fabrics, and enables customers 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 Switch 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 in 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 VM 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 Edge services gateway 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 2-64. Design Decisions on the VXLAN Configuration

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.	Simplifies the network configuration for each tenant 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 Universal Distributed Logical Router (UDLR), and Distributed Logical Router (DLR) to provide tenant network capabilities.	Creates isolated, multi-tenant broadcast domains across data center fabrics to create elastic, logical networks that span physical network boundaries.	Transport networks and MTU greater than 1600 bytes has to be configured in the reachability radius.
SDDC-VI-NET-019	Use VXLAN with NSX Edge gateways and the Universal Distributed Logical Router (UDLR) to provide management application network capabilities.	Creates isolated broadcast domains across data center fabrics to create elastic, logical networks that span physical network boundaries for management applications. This approach also enables encapsulation and transportability of management components.	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 vSphere vMotion and to assign a dedicated default gateway for vSphere vMotion traffic.

By using a separate TCP/IP stack, you can manage vSphere vMotion and cold migration traffic according to the topology of the network, and as required by 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 2-65. Design Decisions on the vMotion TCP/IP Stack

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 as such the VMkernel adapter must be created directly on the ESXi host.

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, snapshots, deletes, and restores software-based virtual machines (VMs), NSX network virtualization programmatically creates, 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, including existing traditional networking models and next-generation fabric architectures, from any vendor.

When administrators provision workloads, network management is one of the most time-consuming tasks. Most of the time spent provisioning networks is consumed 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 pre-provision and configure networks is a major constraint to cloud deployments where speed, agility, and flexibility are critical requirements. Pre-provisioned physical networks can allow for the rapid creation of virtual networks and faster deployment times of workloads utilizing 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 decouple virtual networks from their physical counterparts. Decoupling, in turn, requires that you can 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

- NSX Components

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

- NSX for vSphere Requirements

  NSX for vSphere requirements impact both physical and virtual networks.

- Network Virtualization Conceptual Design

  This conceptual design provides you with an understanding of the network virtualization design.

- Cluster Design for NSX for vSphere

- **vSphere Distributed Switch Uplink Configuration**

  Each ESXi host uses two physical 10-GbE 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.

- **Logical Switch Control Plane Design**

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

- **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.

- **Routing Design**

  The routing design considers different levels of routing within the environment from which to define a set of principles for designing a scalable routing solution.

- **Firewall Logical Design**

  The NSX Distributed Firewall is used to protect all management applications 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 virtual IP (VIP).

- **Load Balancer Design**

  The NSX Edge services gateways (ESG) implement load balancing in NSX for vSphere.

- **Information Security and Access Control in NSX**

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

- **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 VM is located. Multiple bridges per DLR are supported.

- **Region Connectivity**

  Regions must be connected to each other. Connection types could be point-to-point links, MPLS, VPN Tunnels, etc. This connection is different according to your environment and is out of scope for this design.

- Application Virtual Network

  Management applications, such as VMware vRealize Automation, VMware vRealize Operations Manager, or VMware vRealize Orchestrator, leverage a traditional 3-tier client-server architecture with a presentation tier (user interface), functional process logic tier, and data tier. This architecture requires a load balancer for presenting end-user facing services.

- Virtual Network Design Example

  The virtual network design example illustrates an implementation of a management application virtual network for the management components in this validated design.

- Use of SSL Certificates in NSX

  By default, NSX Manager uses a self-signed Secure Sockets Layer (SSL) certificate. This certificate is not trusted by end-user devices or web browsers. It is a security best practice to replace these certificates with certificates that are signed by a third-party or enterprise Certificate Authority (CA).

NSX for vSphere Design

NSX Manager and vCenter Server have a one-to-one relationship. Per region, this design uses two vCenter Server instances and two NSX instances connected to them.

Table 2-66. Design Decisions on the Instances of NSX for vSphere

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 the cloud management platform operations, and also 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 2-14. Architecture of NSX for vSphere

NSX Components

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

Consumption Layer

The cloud management platform (CMP) can consume NSX for vSphere, represented by vRealize Automation, by using the NSX RESTful 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. See Cloud Management Design .

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 resource URL of the object.

- A client can write (create or modify) an object using 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

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

Table 2-67. Design Decisions on the NSX Consumption Method

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 the vSphere Web Client and the NSX REST API.	vRealize Automation services are used for the customer-facing portal. The 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.	End-users typically interact only indirectly with NSX from the vRealize Automation portal. Administrators interact with NSX from the 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.	Ensures that infrastructure components are not modified by tenants or non-provider 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, Edge distributed routers, and 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 User World Agent (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 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 2-68. Design Decisions on the NSX Controller Instances

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 does not deploy controllers. The controllers from the primary NSX Manager manage all secondary resources.

NSX Virtual Switch

The NSX data plane consists of the NSX virtual switch. This virtual switch is based on the vSphere Distributed Switch (VDS) with additional components to enable rich services. The add-on NSX components include kernel modules (VIBs) which run within the hypervisor kernel and provide services such as distributed logical router (DLR) and distributed firewall (DFW), 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 includes several benefits.

- Supports overlay networking and centralized network configuration. Overlay networking enables the following capabilities.

- 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 within a virtual network through features such as 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 within 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

The NSX distributed logical router (DLR) is optimized for forwarding in the virtualized space, that is, forwarding between VMs on VXLAN- or VLAN-backed port groups. DLR has the following characteristics.

- High performance, low overhead first hop routing

- Scales with number of ESXi hosts

- Up to 1,000 Logical Interfaces (LIFs) on each DLR

Distributed Logical Router Control Virtual Machine

The distributed logical router control virtual machine is the control plane component of the routing process, providing communication between NSX Manager and the NSX Controller cluster through the User World Agent (UWA). NSX Manager sends logical interface 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 User World Agent (UWA) is a TCP (SSL) client that facilitates communication between the ESXi hosts and the NSX Controller instances as well as the retrieval of information from the NSX Manager via interaction with the message bus agent.

VXLAN Tunnel Endpoint

VXLAN Tunnel Endpoints (VTEPs) are instantiated within 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 via Internet Group Membership Protocol (IGMP). If you use multiple VTEPs, then you must select a teaming method.

Edge Services Gateway

The NSX Edge services gateways (ESGs) primary function 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 known as the distributed firewall. Security enforcement is done at the kernel and VM 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 Edge service gateway.

NSX for vSphere Requirements

NSX for vSphere requirements impact both physical and virtual networks.

Physical Network Requirements

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

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, MTU size of 9000, for VXLAN traffic.
For the hybrid replication mode, Internet Group Management Protocol (IGMP) snooping must be enabled on the Layer 2 switches to which ESXi hosts that participate in VXLAN are attached. 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. You use hybrid replication mode for broadcast, unknown unicast, and multicast (BUM) 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

Determine the size of an NSX component according to your environment. Sizing resources for NSX according to storage requirements is a part of the physical storage design. See Table 2-12. Design Decisions on the vSAN Disk Configuration.

Size of NSX Edge services gateways might vary according to tenant requirements. Consider all options in such a case.

Table 2-69. Specifications of the NSX Components

VM	vCPU	Memory	Storage	Quantity per Stack Instance
NSX Manager	4	16 GB	60 GB	1
NSX Controller	4	4 GB	28 GB	3
NSX Edge	■ 1 (Compact) ■ 2 (Large) ■ 4 (Quad Large) ■ 6 (X-Large)	■ 512 MB (Compact) ■ 1 GB (Large) ■ 2 GB (Quad Large) ■ 8 GB (X-Large)	■ 1.1 GB (Compact) ■ 1.1 GB (Large) ■ 1.1 GB (Quad Large) ■ 4.84 GB (X-Large)	Optional component. Deployment of NSX ESG varies per use case.
DLR control VM	2	512 MB	1.1 GB	Optional component. Varies with use case. Typically 2 per HA pair.
Guest introspection	2	2 GB	6.26 GB	Optional component. 1 per ESXi host.
NSX data security	1	512 MB	6 GB	Optional component. 1 per ESXi host.

NSX Edge Service Gateway Sizing

The Quad Large size is suitable for high performance firewall abilities. The X-Large size is suitable for both high performance load balancing and routing.

You can convert between NSX Edge service gateway sizes upon demand using a non-disruptive upgrade process. Begin with the Large size and scale up if necessary. A Large NSX Edge service gateway is suitable for medium firewall performance. However, the NSX Edge service gateway does not perform the majority of firewall functions.

Note Edge service gateway throughput is influenced by the WAN circuit. Use an adaptable approach by converting as necessary.

Table 2-70. Design Decisions on Sizing the NSX Edge Service Gateways

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.

Network Virtualization Conceptual Design

This conceptual design provides you with an understanding of the network virtualization design.

The network virtualization conceptual design includes a perimeter firewall, a provider logical router, and the NSX for 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 cloud management platform within the compute/edge stack, or to a management application within the management stack.

Figure 2-15. 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 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 Distributed Logical Router (DLR)

This logical router is optimized for forwarding in the virtualized space, that is, between VMs, on VXLAN port groups or VLAN-backed port groups.

Management Network

The management network is a VLAN-backed network that supports all management components such as vCenter Server, Platform Services Controller, NSX Manager and NSX Controllers, and Update Manager Download Service (UMDS).

In a dual-region environment, this network also handles Site Recovery Manager traffic.

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 will be 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/edge stack in each region.

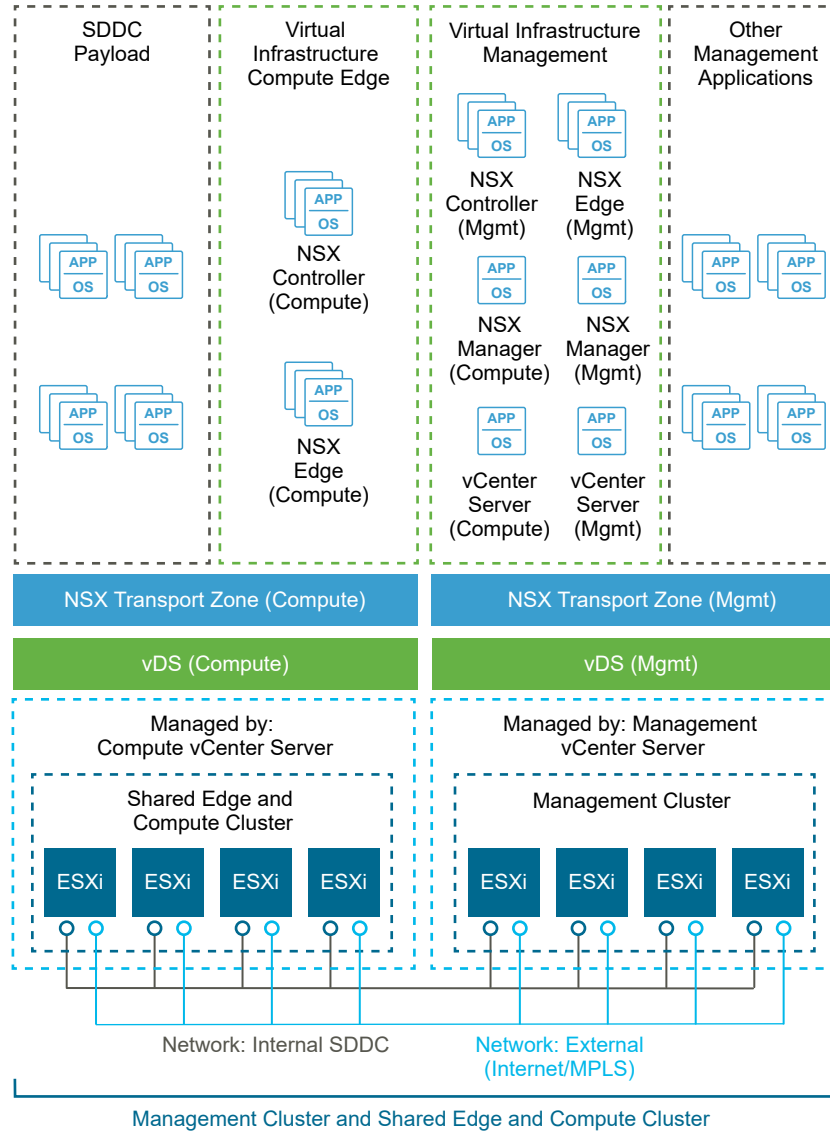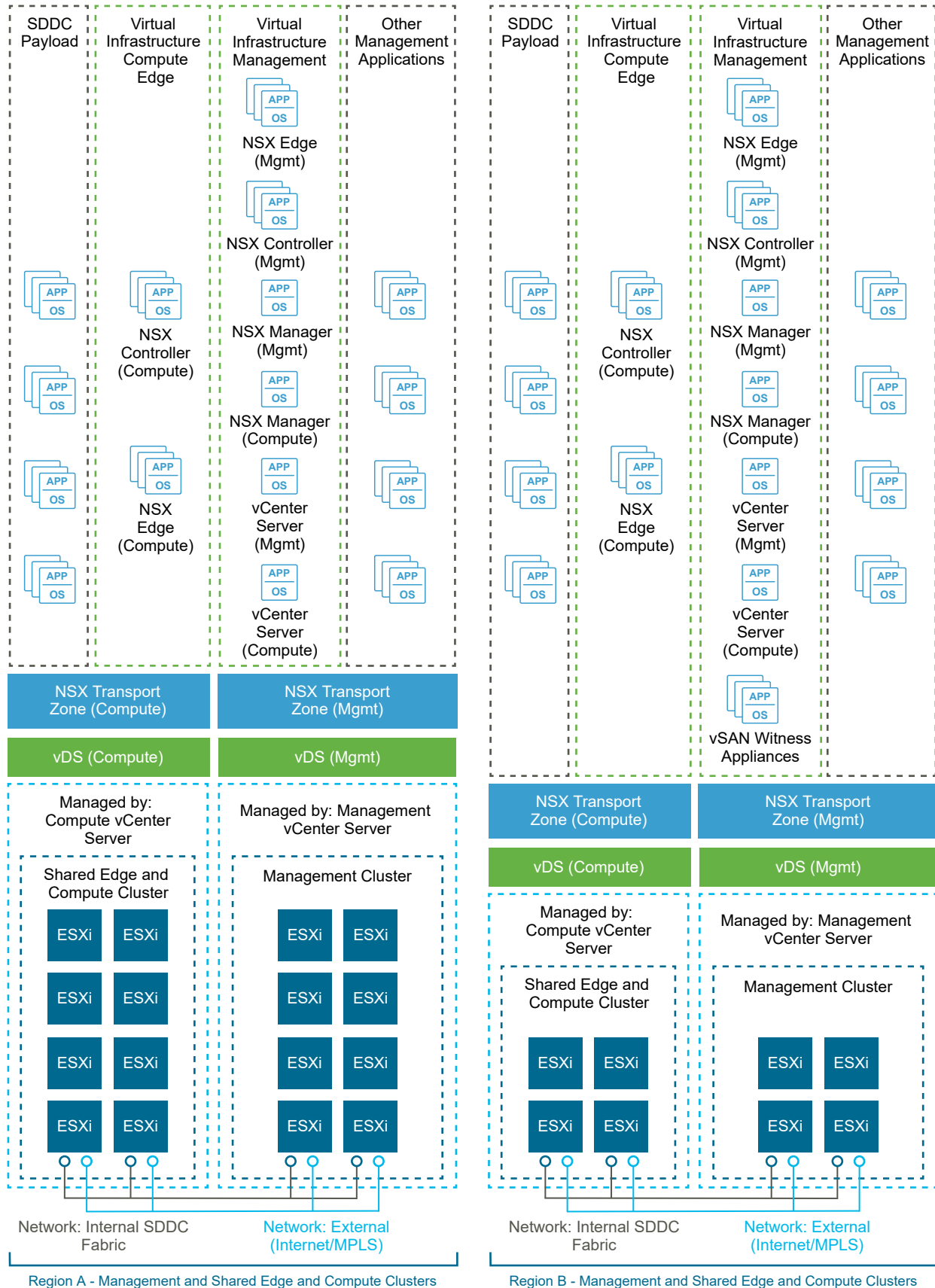## Figure 2-16. Cluster Design for NSX for vSphere with a Single Availability Zone

Figure 2-17. Cluster Design for NSX for vSphere with Two Availability Zones

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/edge stack)

- NSX Controller cluster for the management stack

- NSX ESG and DLR control VMs for the management stack

Edge and Compute Stack

In the edge and compute stack, the underlying ESXi hosts are prepared for NSX for vSphere. The edge and compute 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 need only be trunked to the ESXi hosts in this cluster.

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

Table 2-71. Design Decisions on Cluster Location of the NSX Edge Devices

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 Edge services gateways, 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 impact 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 Distributed Resource Scheduler (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.

High Availability of NSX for vSphere Components

The NSX Manager instances of both stacks run on the management cluster. vSphere HA protects each NSX Manager instance by ensuring that the NSX Manager VM is restarted on a different ESXi host in the event of primary ESXi host failure.

The NSX Controller nodes of the management stack run on the management cluster. The NSX for vSphere Controller nodes of the compute/edge stack run on the shared edge and compute cluster. In both clusters, vSphere Distributed Resource Scheduler (DRS) rules ensure that NSX for vSphere Controller nodes do not run on the same ESXi host.

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/edge 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 equal-cost multi-path (ECMP) mode that supports route failover in seconds. NSX Edge components for load balancing use NSX HA. NSX HA 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 15 seconds after the active device stops working. All NSX Edge devices are also protected by vSphere HA.

Scalability of NSX Components

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

vSphere Distributed Switch Uplink Configuration

Each ESXi host uses two physical 10-GbE 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 2-72. Design Decisions on VTEP Teaming and Failover Configuration

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-010	Set up VXLAN Tunnel Endpoints (VTEPs) to use Route based on SRC-ID for teaming and failover configuration.	Supports 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 Design

The control plane decouples NSX for vSphere from the physical network and handles the broadcast, unknown unicast, and multicast (BUM) traffic within the logical switches. The control plane is on top of the transport zone and is inherited by all logical switches that are created within 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 Controllers 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 2-18. Logical Switch Control Plane in Hybrid Mode

This design uses hybrid mode for control plane replication.

Table 2-73. Design Decisions on the Control Plane Mode of Logical Switches

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 2-74. Design Decisions on Transport Zones

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, and compute clusters from all regions for workloads that require mobility between regions.	A universal transport zone supports extending networks and security policies across regions. As a result, migration of applications across regions either by cross vCenter vMotion or by failover recovery with Site Recovery Manager is seamless.	vRealize Automation is not able to deploy on-demand network objects against a secondary NSX Manager. You must consider that 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 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, and compute clusters for use with vRealize Automation on-demand network provisioning.	NSX Manager instances with a secondary role cannot deploy universal objects. To enable all regions to deploy on-demand network objects, a global transport zone is required.	Shared edge and compute, 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. You implement seamless migration of the management applications across regions either by cross-vCenter vMotion or by failover recovery with Site Recovery Manager.	You must consider that you can pair up to eight NSX Manager instances. If the solution expands past eight NSX Manager instances, you must deploy a new primary NSX Manager and new transport zone.

Table 2-74. Design Decisions on Transport Zones (continued)

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-015	Enable Controller Disconnected Operation (CDO) mode in the management stack.	During times when the NSX controllers 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 an overhead to the hypervisors when the control cluster is down.
SDDC-VI-SDN-016	Enable Controller Disconnected Operation (CDO) mode on the shared edge and compute stack.	During times when the NSX controllers are unable to communicate with ESXi hosts, data plane updates, such as VNIs becoming active on an ESXi host, still occurs.	Enabling CDO mode adds an overhead to the hypervisors when the control cluster is down.

Routing Design

The routing design considers different levels of routing within the environment from which to define a set of principles for designing a scalable routing solution.

North-south

The Provider Logical Router (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 2-75. Design Decisions on the Routing Model of NSX

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-017	Deploy a minimum of two NSX Edge services gateways (ESGs) in an ECMP configuration for North-South routing in both management and shared edge and compute clusters.	■ You use an NSX ESG for directing North-South traffic. Using ECMP provides multiple paths in and out of the SDDC. ■ Failover is faster than deploying ESGs in HA mode.	ECMP requires 2 VLANS in each availability zone and region for uplinks which adds an extra VLAN over traditional HA ESG configurations.
	When using two availability zones, deploy a minimum of two NSX Edge services gateways in an ECMP configuration in each availability zone.	Because the upstream physical Layer 3 devices reside in a single availability zone, you must deploy ECMP edge devices in each availability zone for North-South routing.	
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 the hop count between nodes attached to it to 1. This reduces latency and improves performance.	UDLRs are limited to 1,000 logical interfaces. If that limit is reached, you must deploy a new UDLR.

Table 2-75. Design Decisions on the Routing Model of NSX (continued)

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-019	Deploy a single NSX UDLR for the shared edge and compute, and compute clusters to provide East-West routing across all regions for workloads that require mobility across regions.	Using the UDLR reduces the hop count between nodes attached to it to 1. It reduces latency and improves performance.	UDLRs are limited to 1,000 logical interfaces. If that limit is reached, you must deploy a new UDLR.
SDDC-VI-SDN-020	Deploy a single DLR for the shared edge and compute and compute clusters to provide East-West routing for workloads that require on demand network objects from vRealize Automation.	Using the DLR reduces the hop count between nodes attached to it to 1. It reduces latency and improves performance.	DLRs are limited to 1,000 logical interfaces. If that limit is reached, you must deploy a new DLR.
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 using NAT. This configuration is hard to manage for little benefit.	All North-South traffic is routed through Region A until those routes are no longer available. At that time, all traffic dynamically moves 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 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 layer 3 devices and all ESGs providing North-South routing.	This provides a good balance between failure detection between the layer 3 devices and the ESGs and overburdening the layer 3 devices 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 VM fails over router adjacency is lost and routes from upstream devices such as layer 3 devices to subnets behind the UDLR are lost.	This requires each ECMP edge device 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.

Table 2-75. Design Decisions on the Routing Model of NSX (continued)

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-026	Disable Graceful Restart on all ECMP Edges and Logical Router Control Virtual Machines.	Graceful Restart maintains the forwarding table which in turn will forward packets to a down neighbour even after the BGP timers have expired causing loss of traffic.	None.
SDDC-VI-SDN-027	In the management and shared edge and compute clusters, do not create anti-affinity rules to separate ECMP edges and Logical Router control virtual machines.	■ Because these clusters contain four hosts, creating an anti-affinity rule that contains four virtual machines results in not being able to enter maintenance mode to perform life cycle activities. ■ During a host failure, vSphere HA cannot restart the virtual machine because of the anti-affinity rule.	If the active Logical Router control virtual machine and an ECMP edge reside on the same host and that host fails, a dead path in the routing table appears until the standby Logical Router control virtual machine starts its routing process and updates the routing tables. To avoid this situation, add an additional host to the cluster and create an anti-affinity rule to keep these virtual machines separated.

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 the edge gateways and Layer 3 devices. These networks are used for exchanging routing tables and for carrying transit traffic.

Table 2-76. Design Decisions on the Transit Network

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-028	Create a universal virtual switch for use as the transit network between the UDLR and ESGs.	The UDLR and all ESGs across regions can exchange routing information. The UDLR provides East-West routing in both compute and management stacks while the ESGs provide North-South routing.	Only the primary NSX Manager can create and manage universal objects including this UDLR.
SDDC-VI-SDN-029	Create a global virtual switch in each region for use as the transit network between the DLR and ESGs.	The DLR and ESGs in each region can exchange routing information. The DLR provides East-West routing in the compute stack while the ESGs provide North-South routing.	A global virtual switch for use as a transit network is required in each region.
SDDC-VI-SDN-030	Create two VLANs in each availability zone. Use those VLANs to enable ECMP between the North-South ESGs and the Layer 3 device. The Layer 3 devices have an SVI on one of the two VLANS and each North-South ESG has an interface on each VLAN.	The ESGs can have multiple equal-cost routes. You also have more resiliency and better bandwidth use in the network.	Extra VLANs are required.

Firewall Logical Design

The NSX Distributed Firewall is used to protect all management applications 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 virtual IP (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 are ESGs that provide ECMP services, which require the firewall to be disabled.

Table 2-77. Design Decisions on Firewall Configuration

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-031	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-032	For all ESGs deployed as ECMP North-South routers, disable the firewall.	Use of ECMP on the ESGs is a requirement. 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-033	Configure the Distributed Firewall to limit access to administrative interfaces in the management cluster.	Only authorized administrators can access the administrative interfaces of management applications.	Maintaining firewall rules adds administrative overhead.

Load Balancer Design

The NSX Edge services gateways (ESG) implement load balancing in NSX for vSphere.

An ESG has both Layer 4 and Layer 7 engines that offer different features.

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)
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

Feature	Layer 4 Engine	Layer 7 Engine
Monitoring	View VIP (Virtual IP), Pool and Server objects and stats via CLI and API View global stats for VIP sessions from the vSphere Web Client	View VIP, Pool and Server objects and statistics by using CLI and API View global statistics about VIP sessions from the vSphere Web Client
Layer 7 manipulation	No	URL block, URL rewrite, content rewrite

Table 2-78. Design Decisions on Using an NSX Load Balancer

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-034	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-035	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, having a single load balancer keeps the design simple.	One management application owner might make changes to the load balancer that impact another application.
SDDC-VI-SDN-036	Use an NSX load balancer in HA mode for the Platform Services Controllers.	Using a load balancer increases the availability of the Platform Services Controllers for all applications.	Configuring the Platform Services Controllers and the NSX load balancer adds administrative overhead.

Information Security and Access Control in NSX

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

Table 2-79. Design Decisions on Authorization and Authentication Management in NSX

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-SDN-037	Configure a service account **svc-nsxmanager** in vCenter Server for application-to-application communication from NSX Manager with vSphere.	Provides the following access control features: ■ 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.	You must maintain the service account's life cycle outside of the SDDC stack to ensure its availability .
SDDC-VI-SDN-038	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 VM is located. Multiple bridges per DLR are supported.

Table 2-80. Design Decision on Virtual-to-Physical Interface Type

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-039	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 using NSX L2 Bridging.	Use NSX Layer 2 Bridging only where virtual machines need to 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 Layer 2 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 is through the ESXi host that is running the active Distributed Logical Router Control VM. As such, all bridged traffic flows through this ESXi host at the hypervisor level. As scale-out is required, you may 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 could be point-to-point links, MPLS, VPN Tunnels, etc. This connection is different according to your environment and is out of scope for this design.

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

Table 2-81. Design Decisions on Inter-Site Connectivity

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-040	Provide a connection between regions that is capable of routing between each cluster.	Configuring NSX for cross-vCenter to enable universal objects requires connectivity between NSX Manager instances, ESXi host VTEPs, and NSX Controllers to ESXi hosts management interface. Portability of management and tenant workloads requires connectivity between regions.	You must use jumbo frames across the regions.
SDDC-VI-SDN-041	Ensure that the latency between regions is less than 150 ms.	A latency below 150 ms 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, or VMware vRealize Orchestrator, leverage a traditional 3-tier client-server architecture with a presentation tier (user interface), functional process logic tier, and data tier. This architecture requires a load balancer for presenting end-user facing services.

Table 2-82. Design Decisions on Isolating Management Applications

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-042	Place the following management applications on an application virtual network. ■ Update Manager Download Service ■ vRealize Suite Lifecyle Manager ■ vRealize Operations Manager ■ vRealize Operations Manager remote collectors ■ vRealize Log Insight ■ vRealize Automation ■ vRealize Automation Proxy Agents ■ vRealize Business for Cloud ■ vRealize Business data collectors	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 external users. Direct access to application virtual networks is controlled by distributed firewall rules.
SDDC-VI-SDN-043	Create three application virtual networks. ■ Each region has a dedicated application virtual network for management applications in that region that do not require failover. ■ One application virtual network is reserved for management application failover between regions.	Using only three application virtual networks simplifies the design by sharing Layer 2 networks with applications based on their needs.	A single /24 subnet is used for each application virtual network. IP management becomes critical to ensure no shortage of IP addresses occurs.

Table 2-83. Design Decisions on Portable Management Applications

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-044	The following management applications must be easily portable between regions. ■ vRealize Suite Lifecyle Manager ■ vRealize Operations Manager ■ vRealize Automation ■ vRealize Business	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 2-19. Virtual Application Network Components and Design

Certain 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 of a management application virtual network for the management components in this validated design.

An example for implementing a management application virtual network is the network for vRealize Automation, but the setup of the application virtual networks of any other 3-tier application looks similar.

Figure 2-20. Detailed Example of vRealize Automation Networking

The example is set up as follows.

- You 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 (orange network in Figure 2-20. Detailed Example of vRealize Automation Networking).

- The failover network that is used by vRealize Automation connects to external networks by using 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 using the NSX Edge load balancer.

You map each application virtual network to an IPv4 subnet according to your environment and availability of IP subnets. For example, you can implement the following configuration:

Table 2-84. Example Application Virtual Networks

Application Virtual Network	Management Applications	Internal IPv4 Subnet
Mgmt-xRegion01-VXLAN	■ Cloud Management Platform (vRealize Automation with embedded vRealize Orchestrator, and vRealize Business for Cloud) ■ vRealize Operations Manager ■ vRealize Suite Lifecycle Manager	192.168.11.0/24
Mgmt-RegionA01-VXLAN	■ vRealize Log Insight ■ vRealize Operations Manager Remote Collectors ■ vRealize Automation Proxy Agents ■ vRealize Business Data Collectors	192.168.31.0/24
Mgmt-RegionB01-VXLAN	■ vRealize Log Insight ■ vRealize Operations Manager Remote Collectors ■ vRealize Automation Proxy Agents ■ vRealize Business Data Collectors	192.168.32.0/24

Use of SSL Certificates in NSX

By default, NSX Manager uses a self-signed Secure Sockets Layer (SSL) certificate. This certificate is not trusted by end-user devices or web browsers. It is a security best practice to replace these certificates with certificates that are signed by a third-party or enterprise Certificate Authority (CA).

Table 2-85. Design Decisions on CA-Signed SSL Certificates for NSX

Design ID	Design Decision	Design Justification	Design Implication
SDDC-VI-SDN-045	Replace the NSX Manager certificate with a certificate signed by a third-party Public Key Infrastructure.	Ensures communication between NSX administrators and the NSX Manager are encrypted by a trusted certificate.	Replacing and managing certificates is an operational overhead.

Shared Storage Design

The shared storage design includes the design for vSAN and NFS storage.

Well-designed shared storage provides the basis for an SDDC and has the following benefits.

- Prevents unauthorized access to business data.

- Protects data from hardware and software failures.

- Protects data from malicious or accidental corruption.

Follow these guidelines when designing shared storage for your environment.

- Optimize the storage design to meet the diverse needs of applications, services, administrators, and users.

- Strategically align business applications and the storage infrastructure to reduce costs, boost performance, improve availability, provide security, and enhance functionality.

- Provide multiple tiers of storage to match application data access to application requirements.

- Design each tier of storage with different performance, capacity, and availability characteristics. Because not every application requires expensive, high-performance, highly available storage, designing different storage tiers reduces cost.

- Shared Storage Platform

  You can decide between a traditional storage, VMware vSphere Virtual Volumes, and VMware vSAN storage.

- Shared Storage Logical Design

  The shared storage design selects the storage technology for each type of cluster.

- Datastore Cluster Design

  A datastore cluster is a collection of datastores with shared resources and a shared management interface. Datastore clusters are to datastores what clusters are to ESXi hosts. After you create a datastore cluster, you can use vSphere Storage DRS to manage storage resources.

- vSAN Storage Design

  VMware vSAN Storage design includes conceptual design, logical design, network design, cluster and disk group design, and policy design.

- NFS Storage Design

  This NFS design does not give specific vendor or array guidance. Consult your storage vendor for the configuration settings appropriate for your storage array.

Shared Storage Platform

You can decide between a traditional storage, VMware vSphere Virtual Volumes, and VMware vSAN storage.

Storage Types

Traditional Storage

Fibre Channel, NFS, and iSCSI are mature and viable options to support virtual machine needs.

VMware vSAN Storage

vSAN is a software-based distributed storage platform that combines the compute and storage resources of VMware ESXi hosts. When you design and size a vSAN cluster, hardware choices are more limited than for traditional storage.

VMware vSphere Virtual Volumes

This design does not use VMware vSphere Virtual Volumes because not all storage arrays have the same vSphere Virtual Volume feature sets enabled.

Traditional Storage and vSAN Storage

Fibre Channel, NFS, and iSCSI are mature and viable options to support virtual machine needs.

Your decision to implement one technology or another can be based on performance and functionality, and on considerations like the following:

- The organization's current in-house expertise and installation base

- The cost, including both capital and long-term operational expenses

- The organization's current relationship with a storage vendor

vSAN is a software-based distributed storage platform that combines the compute and storage resources of ESXi hosts. It provides a simple storage management experience for the user. However, you must carefully consider supported hardware options when sizing and designing a vSAN cluster.

Storage Type Comparison

ESXi hosts support a variety of storage types. Each storage type supports different vSphere features.

Table 2-86. Network Shared Storage Supported by ESXi Hosts

Technology	Protocols	Transfers	Interface
Fibre Channel	FC/SCSI	Block access of data/LUN	Fibre Channel HBA
Fibre Channel over Ethernet	FCoE/SCSI	Block access of data/LUN	Converged network adapter (hardware FCoE) NIC with FCoE support (software FCoE)
iSCSI	IP/SCSI	Block access of data/LUN	iSCSI HBA or iSCSI enabled NIC (hardware iSCSI) Network Adapter (software iSCSI)
NAS	IP/NFS	File (no direct LUN access)	Network adapter
vSAN	IP	Block access of data	Network adapter

Table 2-87. vSphere Features Supported by Storage Type

Type	vSphere vMotion	Datastore	Raw Device Mapping (RDM)	Application or Block-Level Clustering	vSphere HA and vSphere DRS	Storage APIs Data Protection
Local Storage	Yes	VMFS	No	Yes	No	Yes
Fibre Channel / Fibre Channel over Ethernet	Yes	VMFS	Yes	Yes	Yes	Yes
iSCSI	Yes	VMFS	Yes	Yes	Yes	Yes
NAS over NFS	Yes	NFS	No	No	Yes	Yes
vSAN	Yes	vSAN	No	Yes (using iSCSI Initiator)	Yes	Yes

Shared Storage Logical Design

The shared storage design selects the storage technology for each type of cluster.

The storage devices for use by each type of cluster are as follows.

- Management clusters use vSAN for primary storage and NFS for secondary storage.

- Shared edge and compute clusters can use FC/FCoE, iSCSI, NFS, or vSAN storage. No specific guidance is given as user workloads and other factors determine storage type and SLA for user workloads.

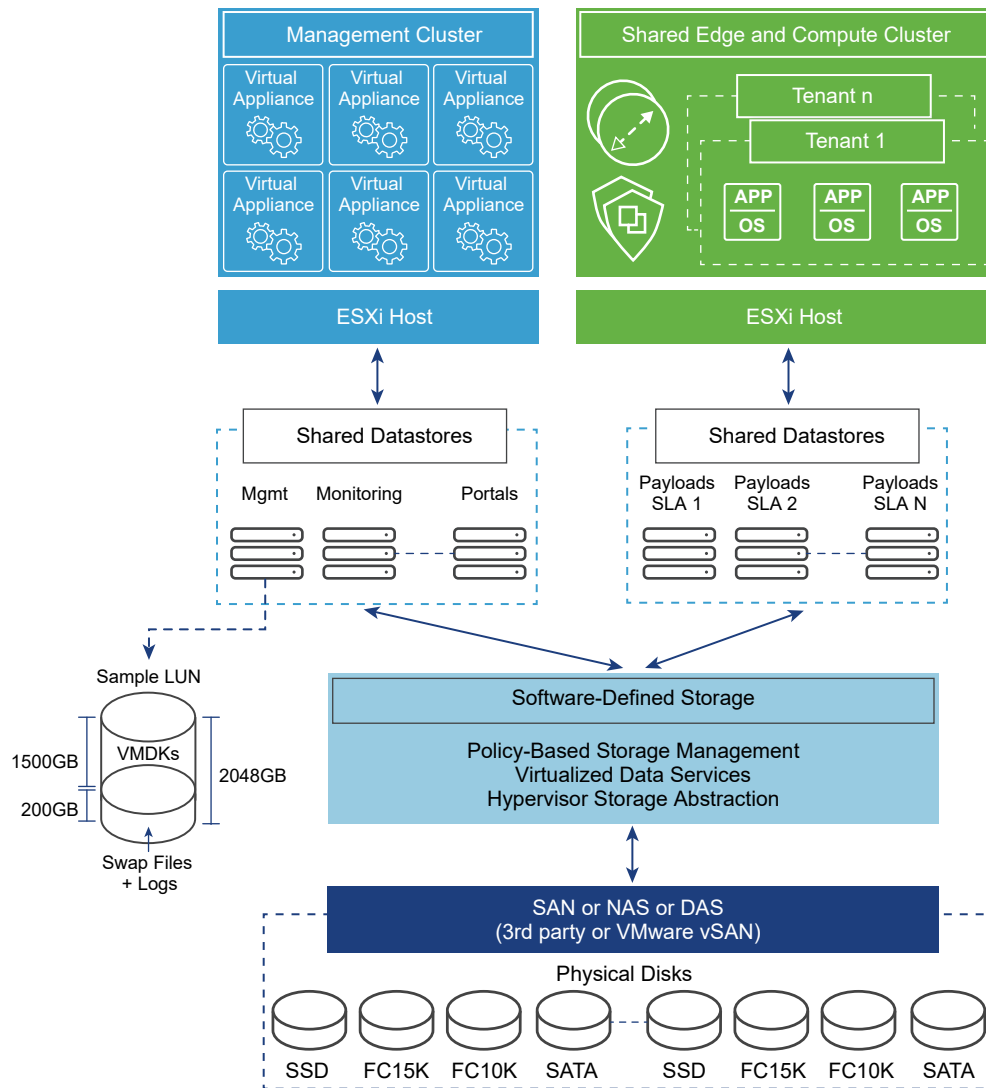## Figure 2-21. Logical Storage Design

Table 2-88. Design Decisions on Storage Type

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-001	When using a single availability zone in the management cluster, use vSAN and NFS shared storage: ■ Use vSAN as the primary shared storage platform. ■ Use NFS as the secondary shared storage platform for the management cluster.	By using vSAN as the primary shared storage solution, you can take advantage of more cost-effective local storage. NFS is primarily for archiving and to maintain historical data. Using NFS provides large, low-cost volumes that you can flexibly expand regularly according to capacity needs.	The use of two different storage technologies increases the complexity and operational overhead. You cannot configure multiple availability zones to use an NFS array in the event an availability zone fails.
SDDC-VI-Storage-002	In all clusters, ensure that at least 20% of free space is always available on all non-vSAN datastores.	If a datastore runs out of free space, applications and services in the SDDC, including but not limited to the NSX Edge core network services, the provisioning portal, and backup, fail.	Monitoring and capacity management must be proactive operations.

Storage Tiering

Not all application workloads have the same storage requirements. Storage tiering allows for these differences by creating multiple levels of storage with varying degrees of performance, reliability and cost, depending on the application workload needs.

Today's enterprise-class storage arrays contain multiple drive types and protection mechanisms. The storage, server, and application administrators face challenges when selecting the correct storage configuration for each application being deployed in the environment. Virtualization can make this problem more challenging by consolidating many different application workloads onto a small number of large devices.

The most mission-critical data typically represents the smallest amount of data and offline data represents the largest amount. Details differ for different organizations.

To determine the storage tier for application data, determine the storage characteristics of the application or service.

■ I/O operations per second (IOPS) requirements

■ Megabytes per second (MBps) requirements

■ Capacity requirements

- Availability requirements

- Latency requirements

After you determine the information for each application, you can move the application to the storage tier with matching characteristics.

- Consider any existing service-level agreements (SLAs).

- Move data between storage tiers during the application lifecycle as needed.

vSphere Storage APIs - Array Integration

The VMware vSphere Storage APIs - Array Integration (VAAI) supports a set of ESXCLI commands for enabling communication between ESXi hosts and storage devices. Using this API/CLI has several advantages.

The APIs define a set of storage primitives that enable the ESXi host to offload certain storage operations to the array. Offloading the operations reduces resource overhead on the ESXi hosts and can significantly improve performance for storage-intensive operations such as storage cloning, zeroing, and so on. The goal of hardware acceleration is to help storage vendors provide hardware assistance to speed up VMware I/O operations that are more efficiently accomplished in the storage hardware.

Without the use of VAAI, cloning or migration of virtual machines by the VMkernel data mover involves software data movement. The data mover issues I/O to read and write blocks to and from the source and destination datastores. With VAAI, the data mover can use the API primitives to offload operations to the array when possible. For example, when you copy a virtual machine disk file (VMDK file) from one datastore to another inside the same array, the data mover directs the array to make the copy completely inside the array. If you invoke a data movement operation and the corresponding hardware offload operation is enabled, the data mover first attempts to use hardware offload. If the hardware offload operation fails, the data mover reverts to the traditional software method of data movement.

Hardware data movement performs better than software data movement. It consumes fewer CPU cycles and less bandwidth on the storage fabric. Timing operations that use the VAAI primitives and use `esxtop` to track values such as CMDS/s, READS/s, WRITES/s, MBREAD/s, and MBWRTN/s of storage adapters during the operation show performance improvements.

Table 2-89. Design Decision on the Integration of vStorage APIs for Array

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-003	Select an array that supports vStorage APIs for Array Integration (VAAI) over NAS (NFS).	■ VAAI offloads tasks to the array itself, enabling the ESXi hypervisor to use its resources for application workloads and not become a bottleneck in the storage subsystem. ■ VAAI is required to support the desired number of virtual machine lifecycle operations.	Not all arrays support VAAI over NFS. For the arrays that support VAAI, to enable VAAI over NFS, you must install a plug-in from the array vendor .

Virtual Machine Storage Policies

You can create a storage policy for a virtual machine to specify which storage capabilities and characteristics are the best match for this virtual machine.

Note To specify the characteristics of virtual machines, use vSAN storage policies. You can define the policy at an individual disk level instead of at the volume level for vSAN.

You can identify the storage subsystem capabilities by using the VMware vSphere API for Storage Awareness (VASA) or by using a user-defined storage policy.

VMware vSphere API for Storage Awareness

With vSphere API for Storage Awareness, storage vendors can publish the capabilities of their storage to VMware vCenter Server, which can display these capabilities in its user interface.

User-defined storage policy

You define the storage policy using the VMware Storage Policy SDK, VMware vSphere PowerCLI, or vSphere Web Client.

You can assign a storage policy to a virtual machine and periodically check for compliance so that the virtual machine continues to run on storage with the correct performance and availability characteristics.

vSphere Storage I/O Control Design

VMware vSphere Storage I/O Control allows cluster-wide storage I/O prioritization, which results in better workload consolidation and helps reduce extra costs associated with over provisioning.

vSphere Storage I/O Control extends the constructs of shares and limits to storage I/O resources. You can control the amount of storage I/O that is allocated to virtual machines during periods of I/O congestion, so that more important virtual machines get preference over less important virtual machines for I/O resource allocation.

When vSphere Storage I/O Control is enabled on a datastore, the ESXi host monitors the device latency when communicating with that datastore. When device latency exceeds a threshold, the datastore is considered to be congested and each virtual machine that accesses that datastore is allocated I/O resources in proportion to their shares. Shares are set on a per-virtual machine basis and can be adjusted.

vSphere Storage I/O Control has several requirements, limitations, and constraints.

- Datastores that are enabled with vSphere Storage I/O Control must be managed by a single vCenter Server system.

- Storage I/O Control is supported on Fibre Channel-connected, iSCSI-connected, and NFS-connected storage. RDM is not supported.

- Storage I/O Control does not support datastores with multiple extents.

- Before using vSphere Storage I/O Control on datastores that are backed by arrays with automated storage tiering capabilities, verify that the storage array has been certified as compatible with vSphere Storage I/O Control. See *VMware Compatibility Guide*.

Table 2-90. Design Decisions on Storage I/O Control

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-004	Enable Storage I/O Control with the default values on all non-vSAN datastores.	Storage I/O Control ensures that all virtual machines on a datastore receive an equal amount of I/O.	Virtual machines that use more I/O access the datastore with priority. Other virtual machines can access the datastore only when an I/O contention occurs on the datastore.

Datastore Cluster Design

A datastore cluster is a collection of datastores with shared resources and a shared management interface. Datastore clusters are to datastores what clusters are to ESXi hosts. After you create a datastore cluster, you can use vSphere Storage DRS to manage storage resources.

vSphere datastore clusters group similar datastores into a pool of storage resources. When vSphere Storage DRS is enabled on a datastore cluster, vSphere automates the process of initial virtual machine file placement and balances storage resources across the cluster to avoid bottlenecks. vSphere Storage DRS considers datastore space usage and I/O load when making migration recommendations.

When you add a datastore to a datastore cluster, the datastore's resources become part of the datastore cluster's resources. The following resource management capabilities are also available for each datastore cluster.

Capability	Description
Space utilization load balancing	You can set a threshold for space use. When space use on a datastore exceeds the threshold, vSphere Storage DRS generates recommendations or performs migrations with vSphere Storage vMotion to balance space use across the datastore cluster.
I/O latency load balancing	You can configure the I/O latency threshold to avoid bottlenecks. When I/O latency on a datastore exceeds the threshold, vSphere Storage DRS generates recommendations or performs vSphere Storage vMotion migrations to help alleviate high I/O load.
Anti-affinity rules	You can configure anti-affinity rules for virtual machine disks to ensure that the virtual disks of a virtual machine are kept on different datastores. By default, all virtual disks for a virtual machine are placed on the same datastore.

You can enable vSphere Storage I/O Control or vSphere Storage DRS for a datastore cluster. You can enable the two features separately, even though vSphere Storage I/O control is enabled by default when you enable vSphere Storage DRS.

vSphere Storage DRS Background Information

vSphere Storage DRS supports automating the management of datastores based on latency and storage utilization. When configuring vSphere Storage DRS, verify that all datastores use the same version of VMFS and are on the same storage subsystem. Because vSphere Storage vMotion performs the migration of the virtual machines, confirm that all prerequisites are met.

vSphere Storage DRS provides a way of balancing usage and IOPS among datastores in a storage cluster:

- Initial placement of virtual machines is based on storage capacity.

- vSphere Storage DRS uses vSphere Storage vMotion to migrate virtual machines based on storage capacity.

- vSphere Storage DRS uses vSphere Storage vMotion to migrate virtual machines based on I/O latency.

- You can configure vSphere Storage DRS to run in either manual mode or in fully automated mode.

vSphere Storage I/O Control and vSphere Storage DRS manage latency differently.

- vSphere Storage I/O Control distributes the resources based on virtual disk share value after a latency threshold is reached.

- vSphere Storage DRS measures latency over a period of time. If the latency threshold of vSphere Storage DRS is met in that time frame, vSphere Storage DRS migrates virtual machines to balance latency across the datastores that are part of the cluster.

When making a vSphere Storage design decision, consider these points:

- Use vSphere Storage DRS where possible.

- vSphere Storage DRS provides a way of balancing usage and IOPS among datastores in a storage cluster:

  - Initial placement of virtual machines is based on storage capacity.

  - vSphere Storage vMotion is used to migrate virtual machines based on storage capacity.

  - vSphere Storage vMotion is used to migrate virtual machines based on I/O latency.

  - vSphere Storage DRS can be configured in either manual or fully automated modes

vSAN Storage Design

VMware vSAN Storage design includes conceptual design, logical design, network design, cluster and disk group design, and policy design.

- vSAN Conceptual Design and Logical Design

  This vSAN design is limited to the management cluster. The design uses the default storage policy to achieve redundancy and performance within the cluster.

- vSAN Network Design

  When performing network configuration, you have to consider the overall traffic and decide how to isolate vSAN traffic.

- **vSAN Cluster and Disk Group Design**

  When considering the cluster and disk group design, you have to decide on the vSAN datastore size, number of ESXi hosts per cluster, number of disk groups per ESXi host, and the vSAN policy.

- **vSAN Policy Design**

  After you enable and configure VMware vSAN, you can create storage policies that define the virtual machine storage characteristics. Storage characteristics specify different levels of service for different virtual machines.

vSAN Conceptual Design and Logical Design

This vSAN design is limited to the management cluster. The design uses the default storage policy to achieve redundancy and performance within the cluster.

VMware vSAN Conceptual Design

While vSAN can be used within the shared edge and compute cluster, this design currently gives no guidance for the implementation.

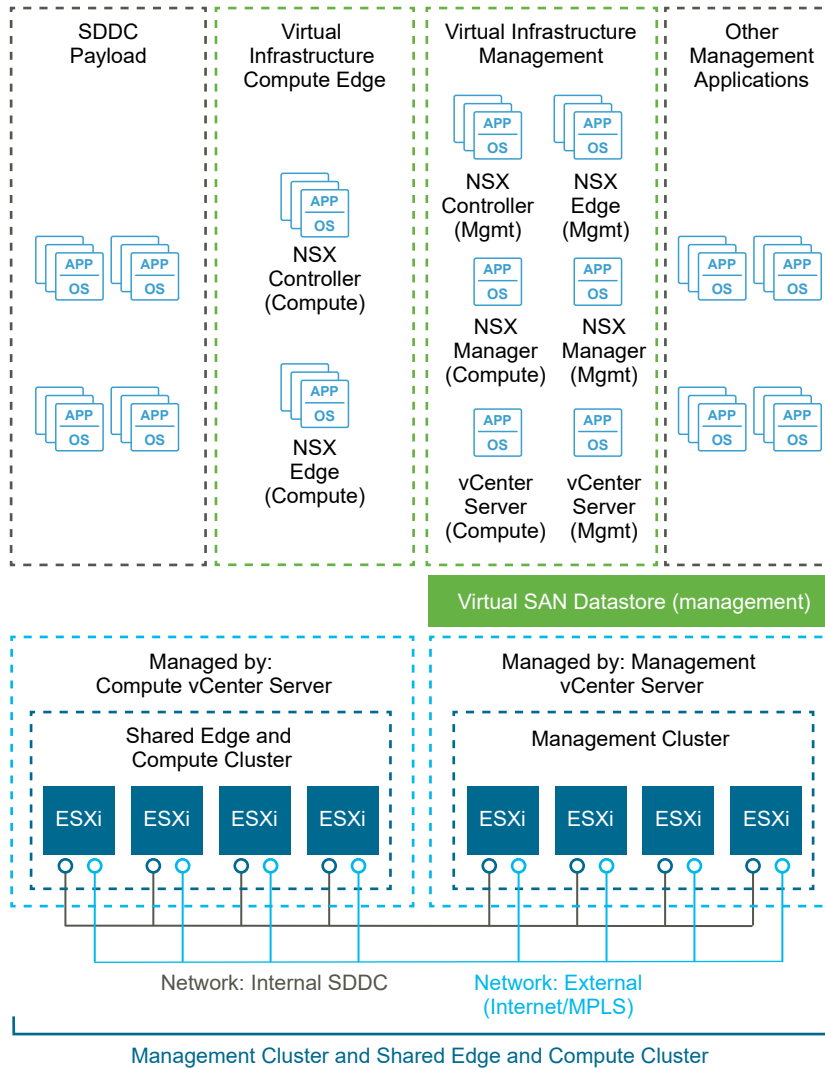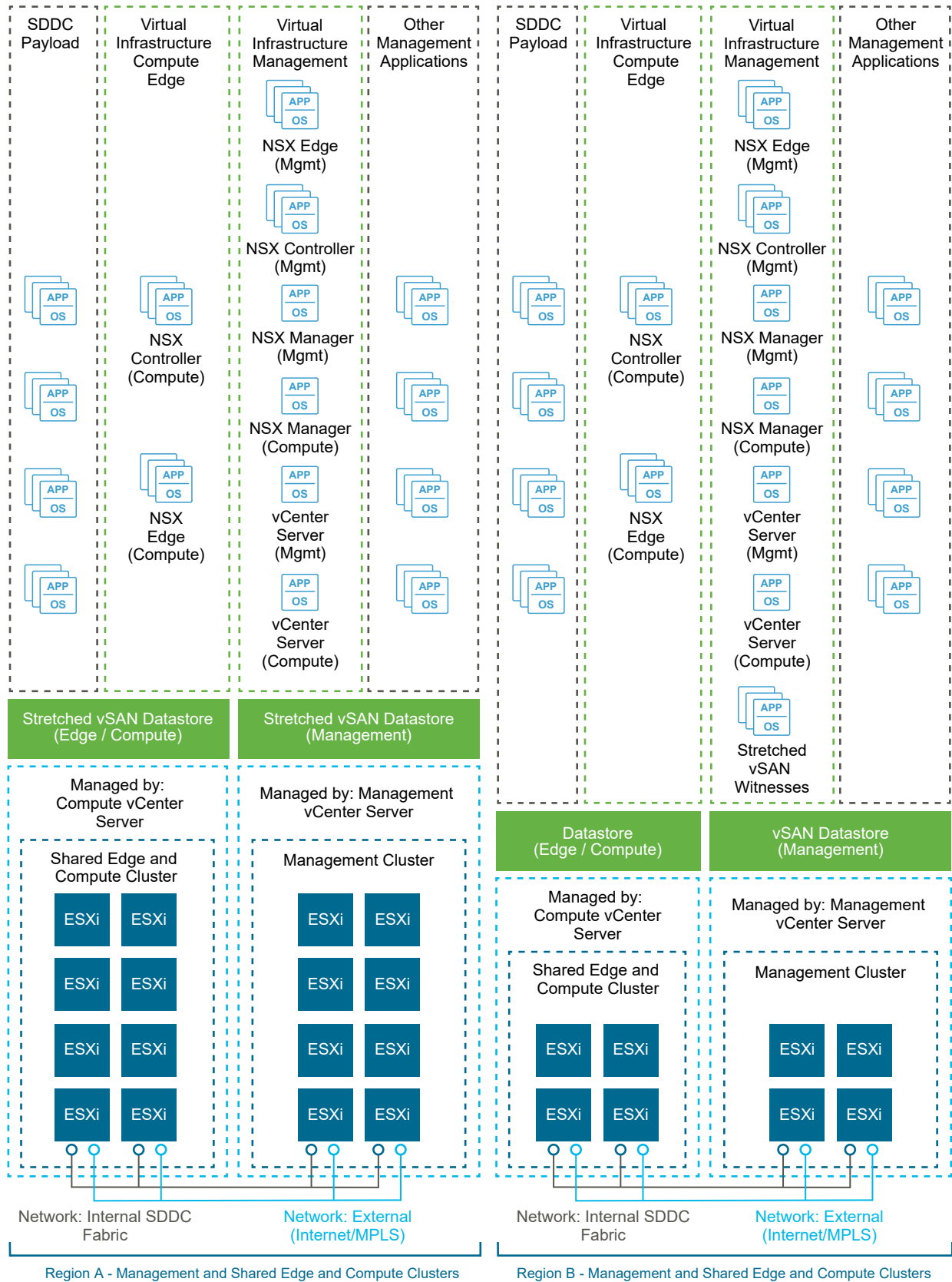## Figure 2-22. Conceptual vSAN Design with a Single Availability Zone

Figure 2-23. Conceptual vSAN Design with Two Availability Zones

vSAN Logical Design

In a cluster that is managed by vCenter Server, you can manage software-defined storage

resources just as you can manage compute resources. Instead of CPU or memory reservations, limits, and shares, you can define storage policies and assign them to virtual machines. The policies specify the characteristics of the storage and can be changed as business requirements change.

vSAN Network Design

When performing network configuration, you have to consider the overall traffic and decide how to isolate vSAN traffic.

vSAN Network Considerations

- Consider how much replication and communication traffic is running between ESXi hosts. With vSAN, the amount of traffic depends on the number of VMs that are running in the cluster, and on how write-intensive the I/O is for the applications running in the VMs.

The vSAN VMkernel port group is created as part of cluster creation. Configure this port group on all ESXi hosts in a cluster, even for ESXi hosts that are not contributing storage resources to the cluster.

Figure 2-24. vSAN Conceptual Network with a Single Availability Zone

Availability Zones Network Considerations

When using two availability zones, the management VLAN that vCenter Server and other VLAN backed management virtual machines use must be stretched across both availability zones. The technology used to stretch the VLAN is out of scope and varies according to your existing infrastructure.

The network infrastructure between availability zones must support jumbo frames, and ensure that latency is less than 5 ms.

Figure 2-25. VMware vSAN Conceptual Network with Two Availability Zones

Network Bandwidth Requirements

For solutions, use a minimum of a 10-GbE connection, with 25-GbE recommended, for use with vSAN to ensure the best and most predictable performance (IOPS) for the environment. Without it, a significant decrease in array performance appears.

A minimum of 10-Gb Ethernet also provides support for future use of vSAN all-flash configurations.

Table 2-91. Network Speed Selection

Design Quality	1 GbE	10 GbE or Greater	Comments
Availability	o	o	Neither design option impacts availability.
Manageability	o	o	Neither design option impacts manageability.
Performance	↓	↑	Faster network speeds increase vSAN performance (especially in I/O intensive situations).
Recoverability	↓	↑	Faster network speeds increase the performance of rebuilds and synchronizations in the environment. As a result, VMs are properly protected from failures.
Security	o	o	Neither design option impacts security.

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

Table 2-92. Design Decisions on Network Bandwidth for vSAN

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-001	Use a minimum of 10 GbE (25 GbE recommended) for vSAN traffic.	Performance with 10 GbE is sufficient, while with 25 GbE is optimal. If the bandwidth is less than 10 GbE, array performance significantly decreases.	The physical network must support 10 Gb or 25 Gb networking between every ESXi host in the vSAN cluster.

VMware vSAN Virtual Switch Type

vSAN supports the use of vSphere Standard Switch or vSphere Distributed Switch. The benefit of using vSphere Distributed Switch is that it supports Network I/O Control which supports prioritization of bandwidth if contention occurs.

This design uses a vSphere Distributed Switch for the vSAN port group to ensure that priority can be assigned using Network I/O Control to separate and guarantee the bandwidth for vSAN traffic.

Virtual Switch Design Background

Virtual switch type affects performance and security of the environment.

Table 2-93. Virtual Switch Types

Design Quality	vSphere Standard Switch	vSphere Distributed Switch	Comments
Availability	o	o	Neither design option impacts availability.
Manageability	↓	↑	The vSphere Distributed Switch is centrally managed across all ESXi hosts, unlike the standard switch which is managed on each ESXi host individually.
Performance	↓	↑	The vSphere Distributed Switch has added controls, such as Network I/O Control, which you can use to guarantee performance for vSAN traffic.
Recoverability	↓	↑	The vSphere Distributed Switch configuration can be backed up and restored, the standard switch does not have this functionality.
Security	↓	↑	The vSphere Distributed Switch has added built-in security controls to help protect traffic.

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

Table 2-94. Design Decisions on Virtual Switch Configuration for vSAN

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-002	Use the existing vSphere Distributed Switch instances in the management cluster in each region.	Provides guaranteed performance for vSAN traffic, if there is network contention, by using existing networking components.	All traffic paths are shared over common uplinks.

Jumbo Frames

vSAN supports jumbo frames for vSAN traffic.

A vSAN design should use jumbo frames only if the physical environment is already configured to support them, they are part of the existing design, or if the underlying configuration does not create a significant amount of added complexity to the design.

Table 2-95. Design Decisions on Jumbo Frames for vSAN

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-003	Configure jumbo frames on the VLAN dedicated to vSAN traffic.	Jumbo frames are already used to improve performance of vSphere vMotion and NFS storage traffic.	Every device in the network must support jumbo frames.

VLANs

Isolate vSAN traffic on its own VLAN. When a design uses multiple vSAN clusters, each cluster should use a dedicated VLAN or segment for its traffic. This approach prevents interference between clusters and helps with troubleshooting cluster configuration.

Table 2-96. Design Decisions on vSAN VLAN

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-004	When using a single availability zone, configure a dedicated VLAN for vSAN traffic for each vSAN enabled cluster.	VLANs provide traffic isolation.	VLANs span only a single cluster. Enough VLANs are available in each cluster and are to be used for traffic separation.
SDDC-VI-Storage-SDS-005	When using two availability zones, configure a dedicated VLAN in each availability zone for each vSAN enabled cluster.	VLANs provide traffic isolation. vSAN traffic between availability zones is routed. An additional stretched VLAN is not required.	Enough VLANs are available within each cluster and are to be used for traffic separation. Static routes on the ESXi hosts are required.

vSAN Witness

When using vSAN in a stretched cluster configuration, you must configure a vSAN stretched cluster witness host. This ESXi host must be configured in a third location that is not local to the ESXi hosts on either side of the stretched cluster.

This vSAN witness can be configured as a physical ESXi host or you can use the vSAN witness appliance.

Table 2-97. Design Decisions on the vSAN Witness Appliance for Multiple Availability Zones

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-006	Use a vSAN witness appliance located in the management cluster of Region B.	Region B is isolated from both availability zones in Region A and can function as an appropriate quorum location.	A third physically separate location is required when implementing a vSAN stretched cluster between two locations.

vSAN Cluster and Disk Group Design

When considering the cluster and disk group design, you have to decide on the vSAN datastore size, number of ESXi hosts per cluster, number of disk groups per ESXi host, and the vSAN policy.

vSAN Datastore Size

The size of the vSAN datastore depends on the requirements for the datastore. Consider cost versus availability to provide the appropriate sizing.

Table 2-98. Design Decisions on the vSAN Datastore

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-007	Provide the management cluster with a minimum of 29 TB of raw capacity for vSAN.	Management cluster virtual machines require at least 9 TB of raw storage (prior to FTT=1) and 18 TB when using the default vSAN storage policy. By allocating at least 29 TB, initially there is 20% free space that you can use for additional management virtual machines. NFS is used as secondary shared storage for some management components, for example, for backups and log archives.	None.
SDDC-VI-Storage-SDS-008	On all vSAN datastores, ensure that at least 30% of free space is always available.	When vSAN reaches 80% usage, a rebalance task is started which can be resource-intensive.	Increases the amount of available storage needed.

Number of ESXi Hosts Per Cluster

The number of ESXi hosts in the cluster depends on these factors:

- Amount of available space on the vSAN datastore

- Number of failures you can tolerate in the cluster

For example, if the vSAN cluster has only 3 ESXi hosts, only a single failure is supported. If a higher level of availability is required, additional hosts are required.

Cluster Size Design Background

Table 2-99. Number of Hosts Per Cluster

Design Quality	3 ESXi Hosts	32 ESXi Hosts	64 ESXi Hosts	Comments
Availability	↓	↑	↑↑	The more ESXi hosts in the cluster, the more failures the cluster can tolerate.
Manageability	↓	↑	↑	The more ESXi hosts in the cluster, the more virtual machines can run in the vSAN environment.
Performance	↑	↓	↓	Having a larger cluster can impact performance if there is an imbalance of resources. Consider performance as you make your decision.
Recoverability	o	o	o	Neither design option impacts recoverability.
Security	o	o	o	Neither design option impacts security.

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

Table 2-100. Design Decision on the Cluster Size for vSAN

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-009	When using a single availability zone, the management cluster requires a minimum of 4 ESXi hosts to support vSAN.	Having 4 ESXi hosts addresses the availability and sizing requirements, and allows you to take an ESXi host offline for maintenance or upgrades without impacting the overall vSAN cluster health.	The availability requirements for the management cluster might cause under utilization of the cluster's ESXi hosts.
SDDC-VI-Storage-SDS-010	When using two availability zones the Management cluster requires a minimum of 8 ESXi hosts (4 in each availability zone) to support a stretched vSAN configuration.	Having 8 ESXi hosts addresses the availability and sizing requirements, and allows you to take an availability zone offline for maintenance or upgrades without impacting the overall vSAN cluster health.	The availability requirements for the management cluster might cause underutilization of the cluster's ESXi hosts.

Number of Disk Groups Per ESXi Host

Disk group sizing is an important factor during volume design.

- If more ESXi hosts are available in the cluster, more failures are tolerated in the cluster. This capability adds cost because additional hardware for the disk groups is required.

- More available disk groups can increase the recoverability of vSAN during a failure.

Consider these data points when deciding on the number of disk groups per ESXi host:

- Amount of available space on the vSAN datastore

- Number of failures you can tolerate in the cluster

The optimal number of disk groups is a balance between hardware and space requirements for the vSAN datastore. More disk groups increase space and provide higher availability. However, adding disk groups can be cost-prohibitive.

Disk Groups Design Background

The number of disk groups can affect availability and performance.

Table 2-101. Number of Disk Groups Per ESXi Host

Design Quality	1 Disk Group	3 Disk Groups	5 Disk Groups	Comments
Availability	↓	↑	↑↑	The more ESXi hosts in the cluster, the more failures the cluster can tolerate. This capability adds cost because additional hardware for the disk groups is required.
Manageability	o	o	o	The more ESXi hosts in the cluster, more virtual machines can be managed in the vSAN environment.
Performance	o	↑	↑↑	If the flash percentage ratio to storage capacity is large, vSAN can deliver increased performance and speed.
Recoverability	o	↑	↑↑	More available disk groups can increase the recoverability of vSAN during a failure. Rebuilds complete faster because there are more places to place data and to copy data from.
Security	o	o	o	Neither design option impacts security.

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

Table 2-102. Design Decisions on the Disk Groups per ESXi Host

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-011	Configure vSAN with a minimum of one disk group per ESXi host in the management cluster.	Single disk group provides the required performance and usable space for the datastore.	Losing an SSD in an ESXi host takes the disk group offline. Using two or more disk groups can increase availability and performance.

vSAN Policy Design

After you enable and configure VMware vSAN, you can create storage policies that define the virtual machine storage characteristics. Storage characteristics specify different levels of service for different virtual machines.

The default storage policy tolerates a single failure and has a single disk stripe. Use the default policy. If you configure a custom policy, vSAN should guarantee its application. However, if vSAN cannot guarantee a policy, you cannot provision a virtual machine that uses the policy unless you enable force provisioning.

VMware vSAN Policy Options

A storage policy includes several attributes, which can be used alone or combined to provide different service levels. Policies can be configured for availability and performance conservatively to balance space consumed and recoverability properties. In many cases, the default system policy is adequate and no additional policies are required. Policies allow any configuration to become as customized as needed for the application's business requirements.

Policy Design Background

Before making design decisions, understand the policies and the objects to which they can be applied. The policy options are listed in the following table.

Table 2-103. VMware vSAN Policy Options

Capability	Use Case	Default Value	Maximum Value	Comments
Number of failures to tolerate	Redundancy	1	3	A standard RAID 1 mirrored configuration that provides redundancy for a virtual machine disk. The higher the value, the more failures can be tolerated. For n failures tolerated, n+1 copies of the disk are created, and 2n+1 ESXi hosts contributing storage are required. A higher n value indicates that more replicas of virtual machines are made, which can consume more disk space than expected.
Number of disk stripes per object	Performance	1	12	A standard RAID 0 stripe configuration used to increase performance for a virtual machine disk. This setting defines the number of HDDs on which each replica of a storage object is striped. If the value is higher than 1, increased performance can result. However, an increase in system resource usage might also result.

Table 2-103. VMware vSAN Policy Options (continued)

Capability	Use Case	Default Value	Maximum Value	Comments
Flash read cache reservation (%)	Performance	0%	100%	Flash capacity reserved as read cache for the storage is a percentage of the logical object size that is reserved for that object.
				Only use this setting for workloads if you must address read performance issues. The downside of this setting is that other objects cannot use a reserved cache.
				VMware recommends not using these reservations unless it is absolutely necessary because unreserved flash is shared fairly among all objects.

Table 2-103. VMware vSAN Policy Options (continued)

Capability	Use Case	Default Value	Maximum Value	Comments
Object space reservation (%)	Thick provisioning	0%	100%	The percentage of the storage object that will be thick provisioned upon VM creation. The remainder of the storage will be thin provisioned. This setting is useful if a predictable amount of storage will always be filled by an object, cutting back on repeatable disk growth operations for all but new or non-predictable storage use.
Force provisioning	Override policy	No	-	Force provisioning forces provisioning to occur even if the currently available cluster resources cannot satisfy the current policy. Force provisioning is useful in case of a planned expansion of the vSAN cluster, during which provisioning of VMs must continue. VMware vSAN automatically tries to bring the object into compliance as resources become available.

By default, policies are configured based on application requirements. However, they are applied differently depending on the object.

Table 2-104. Object Policy Defaults

Object	Policy	Comments
Virtual machine namespace	Failures-to-Tolerate: 1	Configurable. Changes are not recommended.
Swap	Failures-to-Tolerate: 1	Configurable. Changes are not recommended.

Table 2-104. Object Policy Defaults (continued)

Object	Policy	Comments
Virtual disks	User-Configured Storage Policy	Can be any storage policy configured on the system.
Virtual disk snapshots	Uses virtual disk policy	Same as virtual disk policy by default. Changes are not recommended.

Note If you do not specify a user-configured policy, vSAN uses a default system policy of 1 failure to tolerate and 1 disk stripe for virtual disks and virtual disk snapshots. To ensure protection for these critical virtual machine components, policy defaults for the VM namespace and swap are set statically and are not configurable. Configure policies according to the business requirements of the application. By using policies, vSAN can adjust the performance of a disk on the fly.

Policy Design Recommendations

Policy design starts with assessment of business needs and application requirements. Use cases for VMware vSAN must be assessed to determine the necessary policies. Start by assessing the following application requirements:

- I/O performance and profile of your workloads on a per-virtual-disk basis

- Working sets of your workloads

- Hot-add of additional cache (requires repopulation of cache)

- Specific application best practice (such as block size)

After assessment, configure the software-defined storage module policies for availability and performance in a conservative manner so that space consumed and recoverability properties are balanced. In many cases the default system policy is adequate and no additional policies are required unless specific requirements for performance or availability exist.

Table 2-105. Design Decisions on the vSAN Storage Policy

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-012	When using a single availability zone, use the default VMware vSAN storage policy.	Provides the level of redundancy that is needed in the management cluster. Provides the level of performance that is enough for the individual management components.	You might need additional policies for third-party VMs hosted in these clusters because their performance or availability requirements might differ from what the default VMware vSAN policy supports.
SDDC-VI-Storage-SDS-013	When using two availability zones, add the following setting to the default vSAN storage policy: Secondary Failures to Tolerate = 1	Provides the necessary protection for virtual machines in each availability zone, with the ability to recover from an availability zone outage.	You might need additional policies if third-party VMs are to be hosted in these clusters because their performance or availability requirements might differ from what the default VMware vSAN policy supports.
SDDC-VI-Storage-SDS-014	Leave the default virtual machine swap file as a sparse object on VMware vSAN.	Creates virtual swap files as a sparse object on the vSAN datastore. Sparse virtual swap files only consume capacity on vSAN as they are accessed. As a result, you can reduce the consumption on the vSAN datastore if virtual machines do not experience memory over-commitment which requires the use of the virtual swap file.	None.

NFS Storage Design

This NFS design does not give specific vendor or array guidance. Consult your storage vendor for the configuration settings appropriate for your storage array.

NFS Storage Concepts

NFS (Network File System) presents file devices to an ESXi host for mounting over a network. The NFS server or array makes its local file systems available to ESXi hosts. The ESXi hosts access the metadata and files on the NFS array or server using a RPC-based protocol. NFS is implemented using Standard NIC that is accessed using a VMkernel port (vmknic).

NFS Load Balancing

No load balancing is available for NFS/NAS on vSphere because it is based on single session connections. You can configure aggregate bandwidth by creating multiple paths to the NAS array, and by accessing some datastores via one path, and other datastores via another path. You can configure NIC Teaming so that if one interface fails, another can take its place. However, these load balancing techniques work only in case of a network failure and might not be able to handle error conditions on the NFS array or on the NFS server. The storage vendor is often the source for correct configuration and configuration maximums.

NFS Versions

vSphere is compatible with both NFS version 3 and version 4.1; however, not all features can be enabled when connecting to storage arrays that use NFS v4.1.

Table 2-106. Design Decisions on the NFS Version

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-NFS-001	Use NFS v3 for all NFS datastores.	NFS v4.1 datastores are not supported with Storage I/O Control and with Site Recovery Manager.	NFS v3 does not support Kerberos authentication.

Storage Access

NFS v3 traffic is transmitted in an unencrypted format across the LAN. Therefore, best practice is to use NFS storage on trusted networks only and to isolate the traffic on dedicated VLANs.

Many NFS arrays have some built-in security, which enables them to control the IP addresses that can mount NFS exports. A best practice is to use this feature to configure the ESXi hosts that can mount the volumes that are being exported and have read/write access to those volumes. Such a configuration prevents unapproved hosts from mounting the NFS datastores.

Exports

All NFS exports are shared directories that sit on top of a storage volume. These exports control the access between the endpoints (ESXi hosts) and the underlying storage system. Multiple exports can exist on a single volume, with different access controls on each.

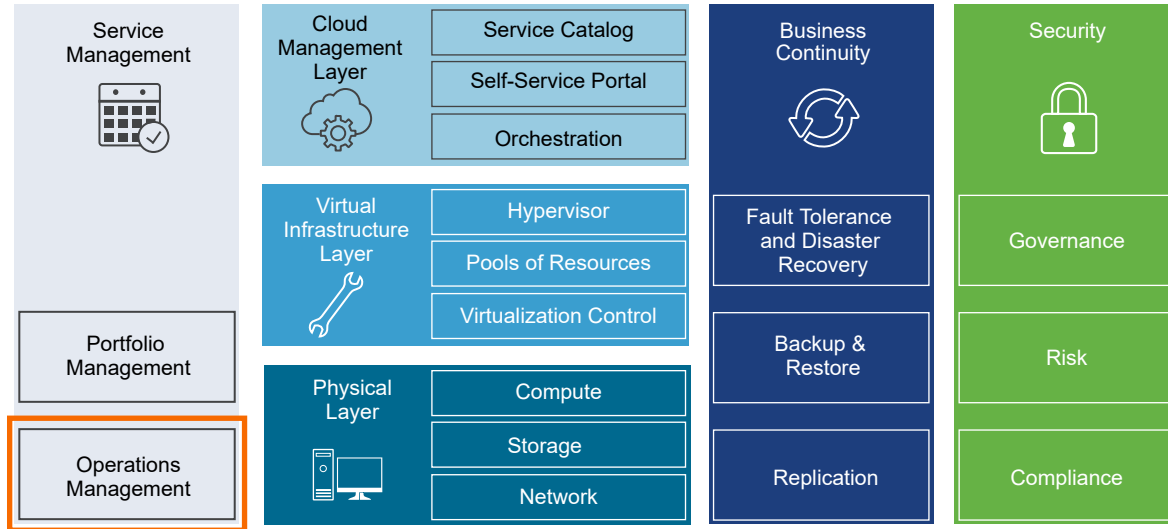| Export Size per Region	Size
vRealize Log Insight Archive	1 TB

Table 2-107. Design Decisions on the NFS Exports

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-NFS-002	Create one export to support the archival functionality of vRealize Log Insight for log persistence.	The storage requirements of these management components are separate from the primary storage.	Dedicated exports can add management overhead to storage administrators in the following areas: ■ Creating and maintaining the export ■ Maintaining access to the vRealize Log Insight nodes if you expand the cluster outside the original design.
SDDC-VI-Storage-NFS-003	Place the VADP-based backup export on its own separate volume as per Table 2-21. Design Decisions on Volume Assignment.	Backup activities are I/O intensive. Backup applications experience resource deficiency if they are placed on a shared volume.	Dedicated exports can add management overhead to storage administrators.
SDDC-VI-Storage-NFS-004	For each export, limit access to the application VMs or hosts requiring the ability to mount the storage only.	Limiting access helps ensure the security of the underlying data.	Securing exports individually can introduce operational overhead.

Operations Management Design

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

■ Features of vSphere Update Manager support upgrade and patching of the ESXi hosts in the SDDC.

■ Features of vRealize Suite Lifecycle Manager support initial installation and configuration of vRealize Suite products. Additional feaures support the Lifecycle Management capabilities and configuration drift analysisfor the vRealize Suite products.

■ Monitoring operations support in vRealize Operations Manager and vRealize Log Insight provides performance, capacity management and real-time logging of related physical and virtual infrastructure and cloud management components.

Figure 2-26. Operations Management in the SDDC Layered Architecture

- **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.

- **vRealize Suite Lifecycle Manager Design**

  vRealize Suite Lifecycle Manager provides lifecycle management capabilities for vRealize components including automated deployment, configuration, and upgrade. You deploy vRealize Suite Lifecycle Manager as a single virtual appliance. The appliance can be failed over across regions in a multi-region SDDC.

- **vRealize Operations Manager Design**

- **vRealize Log Insight Design**

  vRealize Log Insight design enables real-time logging for all components that build up the management capabilities of the SDDC.

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**

- 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. You deploy a vSphere Update Manager Download Service (UMDS) in Region A and Region B to download and stage upgrade and patch data.

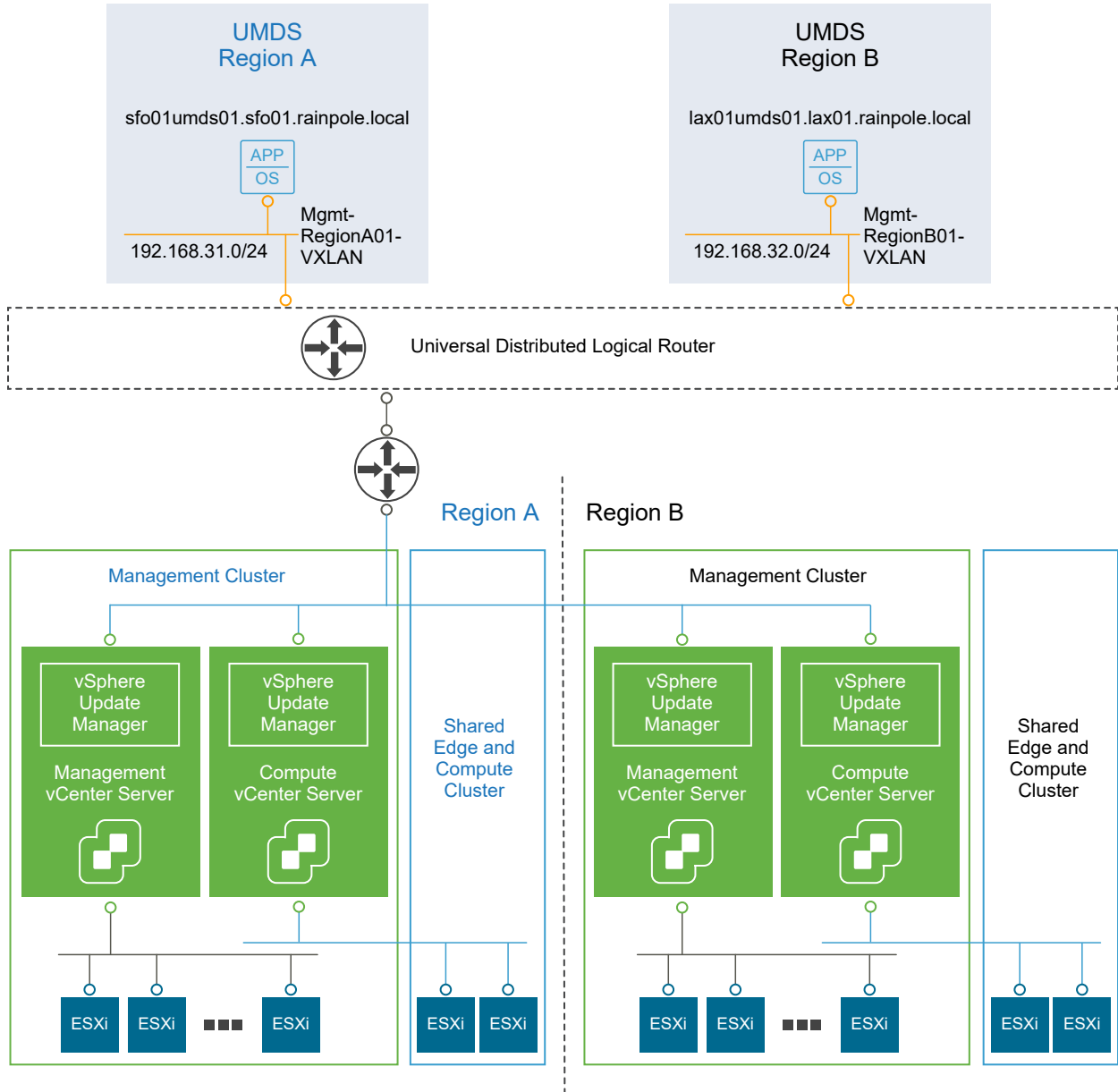### Networking and Application Design

You can use the vSphere Update Manager as a service of the vCenter Server Appliance. The Update Manager server and client components are a 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 save the overhead of downloading updates on multiple vSphere Update Manager instances and to restrict the 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 2-27. Logical and Networking Design of vSphere Update Manager

Deployment Model

vSphere Update Manager is pre-installed 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 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. 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 2-108. Design Decision on the Physical Design of vSphere Update Manager

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 you deploy and maintain 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 that are managed by each vCenter Server.	■ The physical design decisions for vCenter Server determine the setup for vSphere Update Manager. ■ The mapping between vCenter Server and vSphere Update Manager is one-to-one. Each Management vCenter Server or Compute vCenter Server instance in each region must have its own vSphere Update Manager.
SDDC-OPS-VUM-002	Use the embedded PostgreSQL server 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.

Table 2-108. Design Decision on the Physical Design of vSphere Update Manager (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-004	Deploy and configure a vSphere Update Manager Download Service (UMDS) virtual machine in each region.	Restricts the direct access to the Internet from vSphere Update Manager on multiple vCenter Server instances, and reduces the storage requirements on each instance.	You must maintain the host operating system used by the UMDS.
SDDC-OPS-VUM-005	Connect the UMDS virtual machines to the region-specific application virtual network.	■ Provides local storage and access to the repository data of vSphere Update Manager. ■ 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

You allocate resources to and configure the virtual machines for UMDS according to the following specification:

Table 2-109. UMDS Virtual Machine Specification

Attribute	Specification
vSphere Update Manager Download Service	vSphere 6.7 U1
Number of CPUs	2
Memory	2 GB
Disk Space	120 GB
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. You customize these settings according to your business requirements and use cases.

Table 2-110. 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. Virtual machines cannot run when a host is in maintenance mode. For availability during a host update, virtual machines are migrated to the 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.
vSAN	When using vSAN, consider the following factors when you update hosts by using vSphere Update Manager: ■ Host remediation might take a significant amount of time to complete because, by design, only one host from a vSAN cluster can be in maintenance mode at one time. ■ vSphere Update Manager remediates hosts that are a part of a vSAN cluster sequentially, even if you set the option to remediate the hosts in parallel. ■ If the number of failures to tolerate is set to 0 for the vSAN cluster, the host might experience delays when entering maintenance mode. The delay occurs because vSAN copies data between the storage devices in the cluster. To avoid delays, use the default vSAN policy where the number of failures to tolerate is 1.

You can control the update operation by using a set of host and cluster settings in vSphere Update Manager.

Table 2-111. Host and Cluster Settings for Updates

Level	Setting	Description
Host settings	VM power state when entering maintenance mode	You can configure vSphere Update Manager to power off, suspend, or do 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 can remediate only clusters with disabled vSphere DPM, vSphere HA, and vSphere FT.
	Enable parallel remediation of hosts	vSphere Update Manager can remediate multiple hosts. **Note** Parallel remediation is not supported if you use vSAN. Remediation is performed serially for the ESXi hosts.
	Migrate powered-off or suspended virtual machines	vSphere Update Manager migrates the 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 following operations:

Table 2-112. 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 startup dependencies no matter 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, you might need to allow the default baselines only after the patches are tested or verified on development or pre-production hosts. Confirm baselines so that the tested patches are applied to hosts and only updated when appropriate.

Table 2-113. Baseline and Baseline Groups

Baseline or Baseline Group Feature		Description
Baselines	Types	■ 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. Automatically generated according to your vSphere inventory. A system-managed baseline is available in your environment for a vSAN patch, upgrade, or extension. You cannot add system-managed baselines to a baseline group, or to attach or detach them.
	Default Baselines	vSphere Update Manager contains the following default baselines. Each of these baselines is configured with dynamic selection of new items. ■ Critical host patches. Upgrades hosts with a collection of critical patches that have high priority according to VMware. ■ Non-critical host patches. Upgrades 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 VM Hardware version to match the host version.

Table 2-113. Baseline and Baseline Groups (continued)

Baseline or Baseline Group Feature		Description
Baseline groups	Definition	A baseline group consists of a set of non-conflicting 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, 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 Web site or a vendor-specific source. You can then upload the image to vSphere Update Manager.

You can add packages to an ESXi image in the following ways:

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.

Using 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 the patches and extensions are not in conflict. Perform the following steps:

1   Download the NSX software packages bundle from the NSX Manager.

2   Include the NSX software packages, such as `esx-vdpi`, `esx-vsip` and `esx-vxlan`, in an extension baseline.

3   Place the extension baseline and 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 following decisions on the logical design of vSphere Update Manager and update policy:

Table 2-114. Design Decisions on the Logical Design of vSphere Update Manager

Design ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-006	Use the default patch repositories by VMware.	Simplifies the configuration because you use only the pre-defined sources.	None.
SDDC-OPS-VUM-007	Set the VM power state to Do Not Power Off.	Ensures longest uptime of management components and tenant workload virtual machines.	You must manually intervene if the migration fails.
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.	Less resources are available at the same time during remediation.
SDDC-OPS-VUM-009	Enable migration of powered-off virtual machines and templates.	Ensures 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 non-critical patch baselines for the management cluster and for the shared edge and compute cluster.	Simplifies the 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 the configuration because you can use the default schedule without customization.	None.
SDDC-OPS-VUM-012	Remediate hosts, virtual machines, and virtual appliances once a month or according to the business guidelines.	Aligns the remediation schedule with the business policies.	None.
SDDC-OPS-VUM-013	Use a baseline group to add NSX software packages to the ESXi upgrade image.	■ Supports remediation of ESXi hosts by ensuring that the ESXi hosts are ready for software-defined networking immediately after the upgrade. ■ Prevents from additional NSX remediation.	NSX updates require periodic updates of the group baseline.

Table 2-114. Design Decisions on the Logical Design of vSphere Update Manager (continued)

Design ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-014	On each UMDS virtual machine, install and configure an HTTP Web server to share patches with the connected vSphere Update Manager servers.	Enables the automatic download of patches on vSphere Update Manager from UMDS. The alternative is to copy media from one place to another manually.	You must be familiar with a third-party Web service such as Nginx or Apache.
SDDC-OPS-VUM-015	Configure the vSphere Update Manager integration with vSAN.	Enables the integration of vSphere Update Manager with the vSAN Hardware Compatibility List (HCL) for more precision and optimization when you patch with a specific vSphere release ESXi hosts that participate in a vSAN datastore.	■ You cannot perform upgrades between major revisions, for example, from ESXi 6.0 to ESXi 6.5, because of the NSX integration. You must maintain a custom baseline group when performing a major upgrade. ■ To access the available binaries, you must have an active account on *My VMware*.

vRealize Suite Lifecycle Manager Design

vRealize Suite Lifecycle Manager provides lifecycle management capabilities for vRealize components including automated deployment, configuration, and upgrade. You deploy vRealize Suite Lifecycle Manager as a single virtual appliance. The appliance can be failed over across regions in a multi-region SDDC.

vRealize Lifecycle Manager support the following products:

- vRealize Operations Manager

- vRealize Log Insight

- vRealize Automation (with embedded vRealize Orchestrator)

- vRealize Business for Cloud

- Logical Design of vRealize Suite Lifecycle Manager

  vRealize Suite Lifecycle Manager communicates with each Management vCenter Server in the SDDC to orchestrate the deployment, upgrade, and configuration drift analysis of vRealize Suite components in the SDDC.

- Physical Design for vRealize Suite Lifecycle Manager

  You deploy the vRealize Suite Lifecycle Manager virtual appliance and configure system settings, such as, product binaries and marketplace content to support vRealize Suite product deployments in the SDDC.

- Networking Design of vRealize Suite Lifecycle Manager

  For secure access to the UI and API of and failover of vRealize Suite Lifecycle Manager, you place the virtual appliance in the shared cross-region application virtual network.

- Environment Management Design for vRealize Suite Lifecycle Manager

  To deploy the vRealize Suite components by using vRealize Suite Lifecycle Manager, you configure a My VMware account, product binaries, Marketplace integration, data centers, environment structures, and product configuration drift.

- Information Security and Access Design in vRealize Suite Lifecycle Manager

  You protect the vRealize Suite Lifecycle Manager deployment by configuring the authentication and secure communication with the other components in the SDDC. You dedicate a service account to the communication between vRealize Suite Lifecycle Manager and vCenter Server.

- Disaster Recovery of vRealize Suite Lifecycle Manager

  To retain operation when a disaster occurs, this design supports failing vRealize Suite Lifecycle Manager over between regions.
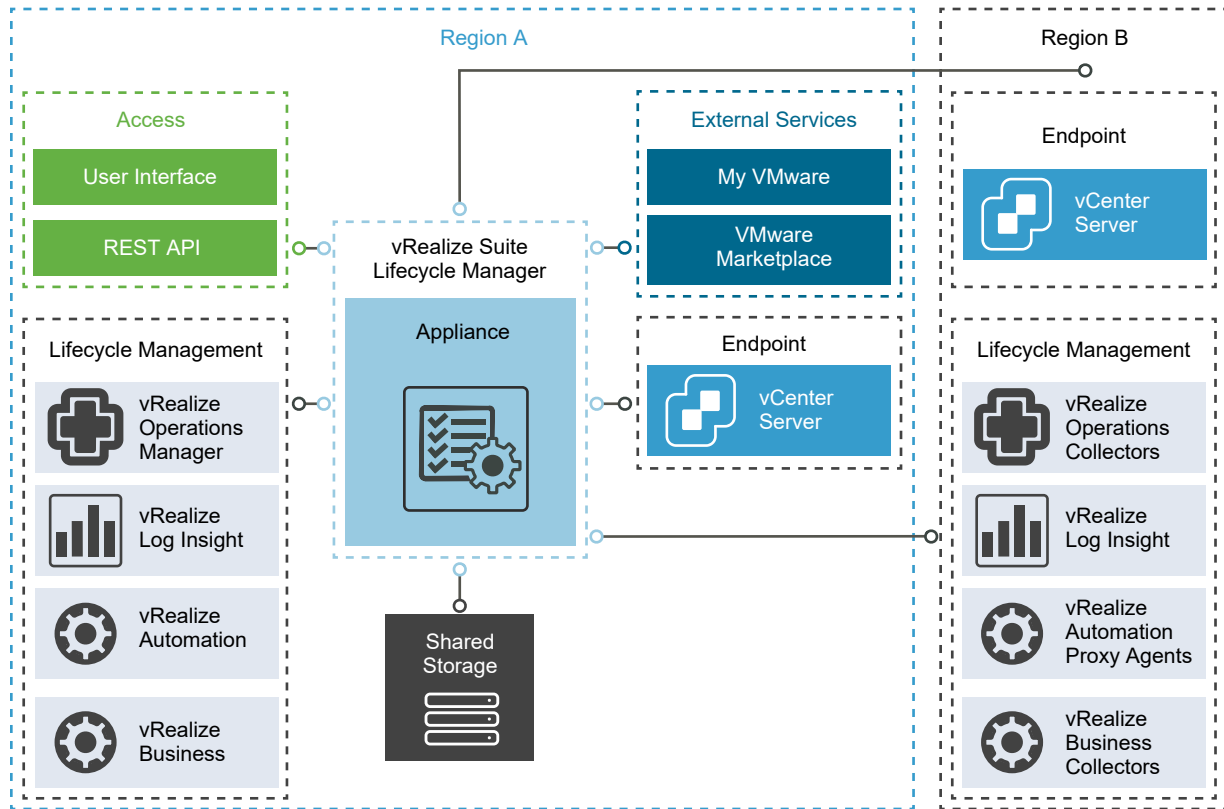
Logical Design of vRealize Suite Lifecycle Manager

vRealize Suite Lifecycle Manager communicates with each Management vCenter Server in the SDDC to orchestrate the deployment, upgrade, and configuration drift analysis of vRealize Suite components in the SDDC.

The vRealize Suite Lifecycle Manager setup consists of one virtual appliance that deploys and upgrades the vRealize components on a virtual infrastructure that is controlled by one or more vCenter Server instances.

vRealize Suite Lifecycle Manager controls separately the lifecycle of cross-region components, Region A specific components, and Region B specific components.

Figure 2-28. Logical Design of vRealize Lifecycle Manager in a Multi-Region Deployment

vRealize Suite Lifecycle Manager operates with the following elements and components in them:

Element	Components
Product Binaries	■ Installation 0VA files ■ Upgrade IS0 and PAK files
My VMware	■ Product Entitlement ■ Product Downloads ■ Product Licensing
VMware Marketplace	■ Marketplace Metadata ■ Download and Compatibility of vRealize Log Insight Content Packs ■ Download and Compatibility of vRealize Operations Manager Management Packs ■ Blueprints and 0VA files for vRealize Automation
Data Center	■ Geographic location ■ vCenter Server instances
Environment	■ vRealize Suite product path or solution path in a data center ■ Deployment configuration file in JSON format

Physical Design for vRealize Suite Lifecycle Manager

You deploy the vRealize Suite Lifecycle Manager virtual appliance and configure system settings, such as, product binaries and marketplace content to support vRealize Suite product deployments in the SDDC.

Deployment Model

In the design, you deploy a single vRealize Suite Lifecycle Manager virtual appliance in the management cluster of Region A. This configuration provides the ability to manage centrally the lifecycle of all vRealize Suite products deployed across the SDDC. In an SDDC that is multi-region and multi availability zone, the vRealize Suite Lifecycle Manager lifecycle management services can be ensured.

After you deploy the appliance, the vRealize Suite Lifecycle Manager services start automatically and the solution can be immediately configured.

Sizing Compute and Storage Resources

To accommodate products binaries and marketplace content for the SDDC, you must size the compute resources and storage for the vRealize Suite Lifecycle Manager.

The vRealize Suite Lifecycle Manager virtual appliance has the following resource requirements. Provide memory and CPUs for the operation of the appliance:

Table 2-115. Resource Specification of the vRealize Suite Lifecycle Manager Appliance

Attribute	Specification with Disabled Content Management	Specification with Enabled Content Management
Number of CPUs	2 vCPUs	4 vCPUs
Memory	16 GB	16 GB
Disk size	135 GB	135 GB
Networks	1 VM NIC	1 VM NIC

When you plan storage for lifecycle management using vRealize Suite Lifecycle Manager, consider the pre-defined disk size of the virtual appliance and storage for the following content:

- Product binaries for install and upgrade

- Marketplace content

- Application and operating system logs

Table 2-116. Design Decisions on the Compute Resources for vRealize Suite Lifecycle Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-001	Deploy vRealize Suite Lifecycle Manager virtual appliance with disabled content management.	■ Accommodates the resources required to support the deployment and upgrade of the vRealize Suite products that are used in the design. ■ Introduces a smaller footprint of the appliance because content management is not enabled. This design does not use the content management capabilities of vRealize Suite Lifecycle Manager.	If content management is required beyond the scope of the design, you must increase the CPU resources to accommodate these services.

Networking Design of vRealize Suite Lifecycle Manager

For secure access to the UI and API of and failover of vRealize Suite Lifecycle Manager, you place the virtual appliance in the shared cross-region application virtual network.

Networking Design of the vRealize Suite Lifecycle Manager Deployment

For secure access and portability, you deploy the vRealize Suite Lifecycle Manager virtual appliance in the shared cross-region application virtual network `Mgmt—xRegion01—VXLAN`.

Figure 2-29. Networking Design of the vRealize Suite Lifecycle Manager Deployment

This networking design has the following features:

- vRealize Suite Lifecycle Manager can be failed over between regions in the event of planned migration or disaster recovery without changing any IP addresses, DNS records, or routing configurations. vRealize Automation, vRealize Business, and vRealize Operations Manager also share this network for cross-region failover support.

- vRealize Suite Lifecycle Manager has routed access to the vSphere management network through the NSX Universal Distributed Logical Router.

- Routing to the vSphere management network, logical networks, and external networks is dynamic, and is based on the Border Gateway Protocol (BGP).

For more information about the networking configuration of the application virtual network, see Virtualization Network Design and NSX Design .

IP Subnets for vRealize Suite Lifecycle Manager

You can allocate the following example subnets for the cross-region VXLAN and use it for vRealize Suite Lifecycle Manager deployment.

Table 2-117. IP Subnets in the Application Virtual Network for vRealize Lifecycle Manager

Node	IP Subnet
vRealize Suite Lifecycle Manager in Region A	192.168.11.0/24

DNS Names for vRealize Suite Lifecycle Manager

The host name of the vRealize Suite Lifecycle Manager virtual appliance follows a specific domain name resolution:

- The IP addresses of the vRealize Suite Lifecycle Manager virtual appliance is associated with a fully qualified name whose suffix is set to the root domain `rainpole.local`.

Table 2-118. Domain Name Service Records for vRealize Suite Lifecycle Manager

DNS Name	Region
vrs01lcm01.rainpole.local	Region A(failover to Region B)

Table 2-119. Design Decision on the DNS Configuration of vRealize Suite Lifecycle Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-003	Configure forward and reverse DNS records for the vRealize Suite Lifecycle Manager virtual appliance.	vRealize Suite Lifecycle Manager is accessible by using fully qualified domain name instead of by using IP addresses only.	You must provide DNS records for the vRealize Suite Lifecycle Manager virtual appliance.

Environment Management Design for vRealize Suite Lifecycle Manager

To deploy the vRealize Suite components by using vRealize Suite Lifecycle Manager, you configure a My VMware account, product binaries, Marketplace integration, data centers, environment structures, and product configuration drift.

Product Binaries

vRealize Suite Lifecycle Manager provides two methods to retrieve and store product binaries for installation and upgrade of the vRealize Suite components.

Table 2-120. Methods for Uploading Product Binaries to vRealize Suite Lifecycle Manager

Method for Retrieving Product Binaries	Description
My VMware	vRealize Suite Lifecycle Manager can integrate directly with My VMware to access vRealize Suite entitlements. Use this method to simplify, automate, and organize the repository.
	You download OVA files for installation and upgrade directly to the vRealize Suite Lifecycle Manager appliance at the following locations:
	▪ `/data/myvmware/`*product*`/`*version*`/install/` for installation
	▪ `/data/myvmware/`*product*`/`*version*`/upgrade/` for upgrade
	If you remove individual binaries that are downloaded from My VMware from the vRealize Suite Lifecycle Manager user interface, the solution removes the related files and metadata from the repository.
	If you register a My VMware account with vRealize Suite Lifecycle Manager, you can provide license keys directly from an entitlement account or input a license key in the installation wizard during an environment creation.
Manual Upload	You can download vRealize Suite product binaries for installation and upgrade directly from My VMware, and then upload them to the vRealize Suite Lifecycle Manager appliance. Use this method if your organization restricts external traffic from the management components of the Software-Defined Data Center.
	You can upload the product binaries directly to the `/data/upload/` folder of the appliance or to an NFS share. vRealize Suite Lifecycle Manager discovers the upload and adds it to the repository.
	If you remove individual product binaries from the vRealize Suite Lifecycle Manager user interface, the solution removes the metadata from the repository but you must manually remove the file from the filesystem.

Table 2-121. Design Decisions on Downloading Product Binaries in vRealize Suite Lifecycle Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-004	Register the virtual appliance of vRealize Suite Lifecycle Manager with My VMware.	Provides the following capabilities: ■ Retrieve and use vRealize Suite licenses in the entitlement of the My VMware account during an SDDC deployment. ■ Download vRealize Suite marketplace content from the VMware Marketplace to the repository.	■ You must have a My VMware account with the required entitlement to access vRealize Suite product downloads and licenses. ■ You must manage a dedicated My VMware account for use with vRealize Suite Lifecycle Manager. ■ You must provide external HTTPS access to vRealize Suite Lifecycle Manager for My VMware and VMware Marketplace endpoints.
SDDC-OPS-LCM-005	Use a dedicated My VMware account for vRealize Suite Lifecycle Manager instead of a named user account.	Provides the following access control features: ■ Accessibility and privileges on the destination service remain restricted to an integration account. ■ Accountability in tracking interactions between the SDDC and My VMware.	■ You must have a My VMware account with the required entitlement to access vRealize Suite product downloads and licenses. ■ You must manage a dedicated My VMware account for use with vRealize Suite Lifecycle Manager.
SDDC-OPS-LCM-006	Download product binaries from My VMware to the virtual appliance of vRealize Suite Lifecycle Manager.	■ Supports download and organization of both installation and upgrade binaries (`.ova`, `.iso`, and `.pak`) for each vRealize Suite product in this design to the vRealize Suite Lifecycle Manager repository. ■ Reduces the administrative overhead of downloading, uploading, and importing media to the vRealize Suite Lifecycle Manager repository.	■ You must have a My VMware account with the required entitlement to access vRealize Suite product downloads and licenses. ■ You must provide more storage because the workflow downloads both installation and upgrade media for each product. ■ If My VMware is not accessible, you must upload the product binaries to the virtual appliance and register

Table 2-121. Design Decisions on Downloading Product Binaries in vRealize Suite Lifecycle Manager (continued)

Decision ID	Design Decision	Design Justification	Design Implication
		▪ Supports adding product upgrades.	them with vRealize Suite Lifecycle Manager. You must also enter manually the product licenses. ▪ Because you use the local repository of vRealize Suite Lifecycle Manager, you transfer the product binaries across the WAN during product deployment and upgrade across regions.

Deployment Paths

vRealize Suite Lifecycle Manager provides the following installation methods in the UI for an environment creation: using a product and using a JSON configuration file.

- Installation Wizard

  - Product Path

  - Solution Path

- Configuration File

Product Path

The default option in vRealize Suite Lifecycle Manager. Add individual vRealize Suite products to the SDDC environment as a clean installation or by importing an existing installation of the product.

A new environment has the following attributes:

- Target data center

- Environment type

- Environment name

When you add one or more products to an environment, you have the following deployment options:

- For a clean installation:

  - Product version

  - Deployment type

  - Node size

■ Additional Configuration

Solution Path

Deploy a use case-based solution for the SDDC environment. You do not use this option with this VMware Validated Design.

Configuration File

Deploy a new environment or import an existing one by using a configuration file in JSON format.

An environment has the following attributes:

■ Target data center

■ Environment type

■ Environment name

When you add one or more products to an environment, you provide a product configuration JSON file.

In this design, you use the product path option in the installation wizard for pre-deployment validation and simplified product deployment.

Table 2-122. Design Decisions on Deployment Paths in vRealize Suite Lifecycle Manager

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-007	Create environments using the product-based deployment path in the installation wizard.	■ Provides a modular path to deploy the vRealize Suite components in this design. ■ Aligns the deployment configurations and sizing. ■ Simplifies the preparation for deployment to a small set of configuration parameters. ■ Provides pre-deployment validations.	You must fill in the necessary configuration parameters in the UI during each installation operation.

Marketplace Integration

You can use vRealize Suite Lifecycle Manager to add and manage content from VMware Marketplace. Once downloaded, you can direct the content deployment to your SDDC directly from vRealize Suite Lifecycle Manager.

To use the integration with the VMware Marketplace, you must register the vRealize Suite Lifecycle Manager appliance with My VMware and the appliance must have Internet access.

You download the content packs from Table 2-165. vRealize Log Insight Content Packs in VMware Validated Design to the Marketplace in vRealize Suite Lifecycle Manager for integration in the SDDC. For information about the supported content pack versions, see *VMware Validated Design Release Notes*.

You download and install manually the vRealize Operations management packs in this design. Most management packs for the SDDC are pre-installed in the product. Some management packs are available only for manual download from My VMware, and you deploy and manage them using the vRealize Operations Manager user interface. See Management Packs in vRealize Operations Manager and *VMware Validated Design Release Notes*.

Table 2-123. Design Decisions on Marketplace Integration in vRealize Suite Lifecycle Manager

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-008	Download vRealize Log Insight content packs to vRealize Suite Lifecycle Manager using the VMware Marketplace integration.	Provides the ability to centrally download and manage vRealize Log Insight content packs across the SDDC.	■ The virtual appliance of vRealize Suite Lifecycle Manager must be registered with a My VMware account to enable the VMware Marketplace integration. ■ If My VMware and VMware Marketplace are inaccessible from the SDDC, you must manually download the vRealize Log Insight content packs from the VMware Solution Exchange and install the content packs using the vRealize Log Insight administrative user interface.
SDDC-OPS-LCM-009	Download manually vRealize Operations management packs to a location that is accessible from the SDDC.	The majority of vRealize Operations management packs for this design are pre-installed with the product and only some are available through My VMware.	You must download the vRealize Operations management packs that are not included in the product from the VMware Solution Exchange and My VMware.

Table 2-123. Design Decisions on Marketplace Integration in vRealize Suite Lifecycle Manager (continued)

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-010	Deploy the vRealize Log Insight content packs using the VMware Marketplace integration in vRealize Suite Lifecycle Manager.	Provides the ability to centrally deploy vRealize Log Insight content packs across the SDDC.	If My VMware and VMware Marketplace are inaccessible from the SDDC, you must manually download and install the content packs using the vRealize Log Insight administrative user interface. The VMware Marketplace integration and in-product Marketplace stores only the latest versions of the content packs. This validated design supports the latest versions of the packs that are available at the time this design is validated. When you deploy the validated design, a later version of a content pack might be available.
SDDC-OPS-LCM-011	Deploy manually vRealize Operations management packs.	The majority of vRealize Operations management packs in this design are pre-installed in the product and only some are available through the My VMware.	You must install the vRealize Operations management packs that are not included in the product directly using the vRealize Operations Manager administrative user interface.

Environments and Data Centers

vRealize Suite Lifecycle Manager solution supports the deployment and upgrade of vRealize Suite products in a logical environment grouping.

These products are as follows:

- vRealize Operations Manager

- vRealize Log Insight

- vRealize Automation (with embedded vRealize Orchestrator)

- vRealize Business

Environments are deployed to a data center object in vRealize Suite Lifecycle Manager. Each environment can contain only one instance of a vRealize Suite product. For example, only one vRealize Log Insight cluster can exist in an environment. However, you can use vRealize Lifecycle Manager to scale out this vRealize Log Insight cluster in the environment to the required number of nodes.

The data center object in vRealize Suite Lifecycle Manager represents a geographic location in the organization. Add the vCenter Server instances to each data center. Each vCenter Server instance is of one of the following types:

- Management

- Workload

- Consolidated Management and Workload

In this design, you create data centers and environments in vRealize Suite Lifecycle Manager to manage the lifecycle of the vRealize Suite components and to support the organic growth of the SDDC using a few operations.

You create the following data center and environment objects:

Table 2-124. Data Center to vCenter Server Instance Mapping in vRealize Suite Lifecycle Manager

Data Center	vCenter Server Instances
Cross-Region	- Management vCenter Server in Region A - Management vCenter Server in Region B
Region A	- Management vCenter Server in Region A
Region B	- Management vCenter Server in Region B

Table 2-125. Configuration of Data Center-Environment Pairs in vRealize Lifecycle Manager

Data Center-Environment Pair	Description
Cross-Region	Supports the deployment of cross-region components like vRealize Operations Manager, vRealize Automation and vRealize Business including collectors and proxy agents.
Region A	Supports the deployment of vRealize Log Insight in Region A. vRealize Log Insight has several instances across the SDDC, each instance specific to a region. You deploy each instance using a separate data center and environment.
Region B	Supports the deployment of vRealize Log Insight in Region B.

Table 2-126. Environment Layout in vRealize Suite Lifecycle Manager

Environment Name	Environment Type	Data Center	Product Components
Cross-Region	Production	Cross-Region	■ vRealize Operations Manager Analytics Cluster ■ vRealize Operations Manager Remote Collectors ■ vRealize Automation Appliances ■ vRealize Automation IaaS Managers ■ vRealize Automation IaaS Web Servers ■ vRealize Automation IaaS DEMs ■ vRealize Automation vSphere Proxy Agents ■ vRealize Business Server Appliances ■ vRealize Business Data Collectors
Region A	Production	Region A	vRealize Log Insight Cluster
Region B	Production	Region B	vRealize Log Insight Cluster

Table 2-127. Design Decisions on the Environment Configuration in vRealize Suite Lifecycle Manager

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-012	■ Create a data center object in vRealize Suite Lifecycle Manager for SDDC solutions that are managed across regions. ■ Assign the management vCenter Server instance in each region to the data center.	Allows you to deploy and manage the integrated vRealize Suite components across the SDDC as a group.	vRealize Suite Lifecycle Manager requires a geographic designation for each data center. The cross-region location is labeled as Region A, that is, San Francisco, only to satisfy the UI requirement.
SDDC-OPS-LCM-013	■ Create a data center object in vRealize Suite Lifecycle Manager for each region. ■ Assign each data center object the Management vCenter Server instance for the region.	Supports deployment and management of vRealize Suite components that are region-specific.	You must manage a separate data center object for the products that are specific to each region.

Table 2-127. Design Decisions on the Environment Configuration in vRealize Suite Lifecycle Manager (continued)

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-014	Create an environment in vRealize Suite Lifecycle Manager for SDDC solutions that are cross-region: ■ vRealize Operations Manager Analytics Nodes ■ vRealize Operations Remote Collectors ■ vRealize Automation Appliances ■ vRealize Automation IaaS Managers ■ vRealize Automation IaaS Web Servers ■ vRealize Automation IaaS DEMs ■ vRealize Automation vSphere Proxy Agents ■ vRealize Business for Cloud Serve Appliance ■ vRealize Business for Cloud Data Collectors Appliances	■ Supports deployment and management of the integrated vRealize Suite components across the SDDC regions as a group. ■ Enables the deployment of region-specific components, such as collectors and proxy agents, that provide data to primary components. In vRealize Lifecycle Manager, you can deploy and manage collectors and proxy agents only in an environment that contains the associated cross-region primary components.	You can management region-specific components, such as, collectors and proxy agents, only in an environment that is cross-region.
SDDC-OPS-LCM-015	Create an environment in vRealize Suite Lifecycle Manager for each region to deploy and manage the standalone vRealize Suite products that region-specific: ■ vRealize Log Insight Cluster	Supports the deployment of an instance of a management product in each region. Using vRealize Lifecycle Manager, you can deploy only one instance of a vRealize Suite product per environment. You use a separate environment for each region where you deploy a product instance.	You must maintain an environment for each the region to deploy and manage the standalone region-specific solutions.

Configuration Drift Intervals

You can use the configuration drift capability in vRealize Suite Lifecycle Manager to save a baseline of the vRealize Suite product configurations. You set the configuration drift interval in the vRealize Suite Lifecycle manager settings to compare the baseline with the current state configuration of each product and create a configuration drift report for review. If you discover that a product configuration has changed during the drift timeline, you can review the changes.

Table 2-128. Design Decisions on the Configuration Drift Interval in vRealize Suite Lifecycle Manager

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-016	Use the default configuration drift interval.	The default configuration drift monitors the changes in vRealize Suite product configurations over each 24-hour period.	Drift analysis occurs only every 24 hours.
SDDC-OPS-LCM-017	Create a baseline for each product in an environment post-deployment and post-upgrade.	You can view any changes to the current configuration of the product compared with the configuration drift baseline of the product.	You must save the configuration baseline after deployment or upgrades.

Information Security and Access Design in vRealize Suite Lifecycle Manager

You protect the vRealize Suite Lifecycle Manager deployment by configuring the authentication and secure communication with the other components in the SDDC. You dedicate a service account to the communication between vRealize Suite Lifecycle Manager and vCenter Server.

You use a custom role in vSphere with permissions to perform lifecycle operations on vRealize Suite components in the SDDC. A dedicated service account is assigned a custom role for communication between vRealize Suite Lifecycle Manager and the vCenter Server instances in the environment.

Encryption

Access to all vRealize Suite Lifecycle Manager endpoint interfaces requires an SSL connection. By default, vRealize Suite Lifecycle Manager uses a self-signed certificate for the appliance. To provide secure access to the vRealize Suite Lifecycle Manager and between SDDC endpoints, replace the default self-signed certificate with a CA-signed certificate.

Table 2-129. Design Decisions on vRealize Suite Lifecycle Manager Encryption

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-018	Replace the default self-signed certificate of the virtual appliance of vRealize Suite Lifecycle Manager with a CA-signed certificate.	Configuring a CA-signed certificate ensures that the communication to the externally facing Web UI and API for vRealize Suite Lifecycle Manager, and cross-product, is encrypted.	Replacing the default certificates with trusted CA-signed certificates from a certificate authority might increase the deployment preparation time as certificates requests are generated and delivered.

Authentication and Authorization

Users can authenticate to vRealize Suite Lifecycle Manager in the following ways:

- Local Administrator Account

- VMware Identity Manager

vRealize Suite Lifecycle Manager performs local authentication for the default administrator account only. You can also enable primary authentication by using VMware Identity Manager to ensure accountability on user access. You can grant both users and groups access to vRealize Suite Lifecycle Manager to perform tasks, and initiate orchestrated operations, such as, deployment and upgrade of vRealize Suite components and content.

Configure a service account for communication between vRealize Suite Lifecycle Manager and vCenter Server endpoint instances. You define a service account with only the minimum set of permissions to perform inventory data collection and lifecycle management operations for the instances defined in the data center.

Table 2-130. Design Decisions on Authentication and Authorization in vRealize Suite Lifecycle Manager

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-019	Use local authentication for vRealize Suite Lifecycle Manager.	vRealize Suite Lifecycle Manager supports only local authentication or authentication by using VMware Identity Manager. Although vRealize Suite Lifecycle Manager supports the use of VMware Identity Manager as an authentication source and access control, it is not used in this design.	■ The accountability in tracking user interactions between vRealize Suite Lifecycle Manager and the vRealize Suite components of the SDDC is limited. ■ You must control the access to the administrator account for vRealize Suite Lifecycle Manager.
SDDC-OPS-LCM-020	Define a custom vCenter Server role for vRealize Suite Lifecycle Manager that has the minimum privileges required to support the deployment and upgrade of vRealize Suite products in the design.	vRealize Suite Lifecycle Manager accesses vSphere with the minimum set of permissions that are required to support the deployment and upgrade of vRealize Suite products in the design.	You must maintain the permissions required by the custom role.

Table 2-130. Design Decisions on Authentication and Authorization in vRealize Suite Lifecycle Manager (continued)

ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LCM-021	Configure a service account svc-vrslcm-vsphere in vCenter Server for application-to-application communication from vRealize Suite Lifecycle Manager to vSphere.	Provides the following access control features: ■ vRealize Suite Lifecycle Manager accesses vSphere with the minimum set of required permissions. ■ 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 lifecycle and availability of the service account outside of the SDDC stack.
SDDC-OPS-LCM-022	Assign permissions for the vRealize Suite Lifecycle Manager service account svc-vrslcm-vsphere in vCenter Server using the custom role at the cluster level to the management cluster in the management domain for each region.	vRealize Suite Lifecycle Manager accesses vSphere with the minimum set of permissions that are required to support the deployment and upgrade of VMware vRealize Suite products in the design.	You must maintain the assignment of the service account and the custom role at a cluster level for each management cluster instead of using global permissions.

Disaster Recovery of vRealize Suite Lifecycle Manager

To retain operation when a disaster occurs, this design supports failing vRealize Suite Lifecycle Manager over between regions.

You place vRealize Suite Lifecycle Manager on the cross-region network, `Mgmt-xRegion01-VXLAN`. As a result, after a recovery, you continue using the same IP addresses, DNS records, and routing configuration. vRealize Automation, vRealize Business, and vRealize Operations Manager also use this network for their cross-region failover capabilities.

If a planned migration or disaster occurs, you use Site Recovery Manager and vSphere Replication for orchestrated recovery of the vRealize Suite Lifecycle Manager virtual appliance. vRealize Suite Lifecycle Manager continues to manage the deployment of the available environments post-recovery. See Recovery Plan for Site Recovery Manager and vSphere Replication.

vRealize Operations Manager Design

The deployment 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 two-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 which 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. You 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

   You provide isolation 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

   You protect the vRealize Operations Manager deployment by providing centralized role-based authentication and secure communication with the other components in the SDDC. You 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

   You use vRealize Operations Manager to monitor the state of the SDDC management components in the SDDC 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 sub-set of the components between regions. Disaster recovery covers only the analytics cluster components, including the primary, 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 which are presented through a number of dashboards and views.
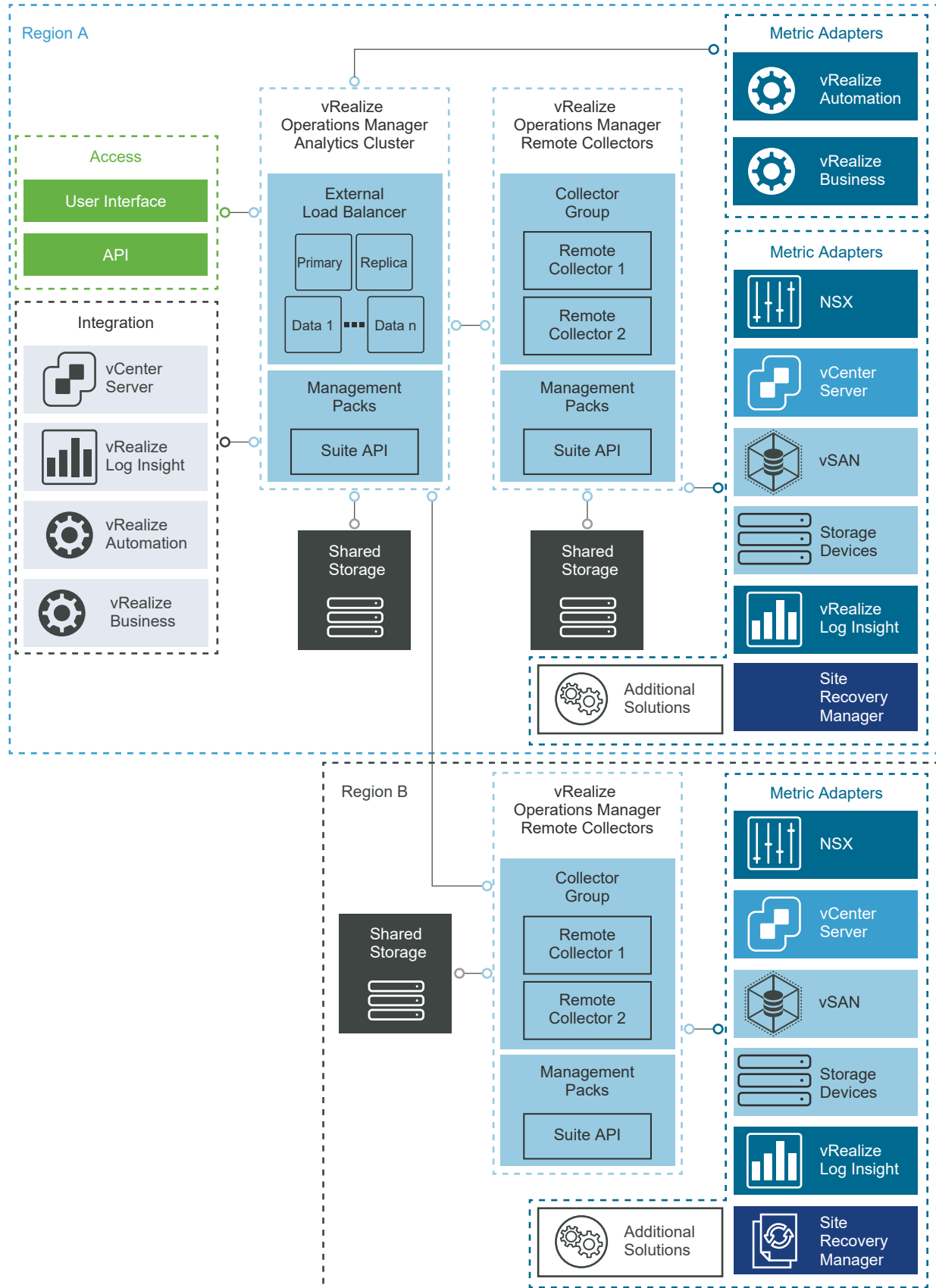
Logical Design

In a multi-region SDDC, you deploy a vRealize Operations Manager configuration that consists of the following entities.

- 3-node (medium-size) vRealize Operations Manager analytics cluster that is highly available (HA). This topology provides high availability, scale-out capacity up to sixteen nodes, and failover.

- 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. Use two remote collectors in each region for load balancing and fault tolerance.

Each region contains its own pair of remote collectors whose role is to ease scalability by performing the data collection from the applications that are not a 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 2-30. Logical Design of a Multi-Region Deployment of vRealize Operations Manager

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 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 IaaS Proxy Agents

    - Microsoft SQL Server

- vRealize Business for Cloud

    - vRealize Business Server

    - vRealize Business Data Collectors

- vRealize Log Insight

- 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. You 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 3-node vRealize Operations Manager analytics cluster in the cross-region application virtual network. The analytics cluster consists of one primary node, one primary replica node, and one data node to enable scale out and high availability.

Table 2-131. Design Decisions on the Node Configuration of vRealize Operations Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-001	Deploy vRealize Operations Manager as a cluster of three nodes: one primary, one primary replica, and one data node. Use this deployment to monitor Region A and Region B.	■ Provides the initial scale capacity required for monitoring up to 1,000 VMs. ■ Supports scale-up with more data nodes.	■ You must size identically all appliances which increases the resources requirements in the SDDC. ■ You must manually install the additional data nodes as per the data node scale guidelines.
SDDC-OPS-MON-002	Deploy two remote collector nodes per region.	Removes the load from the analytics cluster from collecting metrics from applications that do not fail over between regions.	You must assign a collector group when configuring the monitoring of a solution.
SDDC-OPS-MON-003	Apply vSphere Distributed Resource Scheduler (DRS) anti-affinity rules to the vRealize Operations Manager analytics cluster.	Using vSphere 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. ■ You can put in maintenance mode only a single ESXi host at a time in a management cluster of four ESXi hosts.
SDDC-OPS-MON-004	Apply vSphere Distributed Resource Scheduler (DRS) anti-affinity rules to the vRealize Operations Manager remote collector group.	Using vSphere 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. ■ You can put in maintenance mode only a single ESXi host at a time in a management cluster of four ESXi hosts.

Sizing Compute Resources for vRealize Operations Manager

You size compute resources for vRealize Operations Manager to provide enough resources for accommodating the analytics operations for monitoring the SDDC and the expected number of virtual machines in the SDDC.

Size the vRealize Operations Manager analytics cluster according to VMware Knowledge Base article 2093783. Further sizing recommendations can be obtained from http://vropssizer.vmware.com. vRealize Operations Manager is also sized so as 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 following two options:

Initial Setup	Scaled-Out Setup
4 vCenter Server Appliances	4 vCenter Server Appliances
4 NSX Managers	4 NSX Managers
6 NSX Controllers	6 NSX Controllers
50 ESXi hosts	100 ESXi hosts
4 vSAN datastores	4 vSAN datastores
1,000 virtual machines	10,000 virtual machines

Sizing Compute Resources for the Analytics Cluster Nodes

Deploying 3 medium-size virtual appliances satisfies the initial requirement for retention and for monitoring the expected number of objects and metrics for smaller environments up to 1,000 virtual machines. As the environment expands, deploy more data notes to accommodate the higher expected number of objects and metrics.

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

Table 2-132. Resources for a Medium-Size vRealize Operations Manager Virtual Appliance

Attribute	Specification
Appliance size	Medium
vCPU	8
Memory	32 GB
Single-Node Maximum Objects	8,500
Single-Node Maximum Collected Metrics (*)	2,500,000
Multi-Node Maximum Objects Per Node (**)	6,250
Multi-Node Maximum Collected Metrics Per Node (**)	1,875,000
Maximum number of End Point Operations Management agents per node	1,200
Maximum Objects for 16-Node Configuration	75,000
Maximum Metrics for 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 that is 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.

(**) The reduction in maximum metrics to permit some head room.

Table 2-133. Design Decisions on the Compute Size of the Analytics Cluster Nodes of vRealize Operations Manager

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 when at full capacity. If you use fewer large-size vRealize Operations Manager nodes, you must increase the minimum host memory size to handle the increased performance that is the result from stretching NUMA node boundaries.	ESXi hosts in the management cluster must have physical CPUs with a minimum of 8 cores per socket. In total, vRealize Operations Manager uses 24 vCPUs and 96 GB of memory in the management cluster.
SDDC-OPS-MON-006	Deploy initially 3 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 while having high availability in the analytics cluster enabled and collection of metrics about the following components. ■ vCenter Server and Platform Services Controller ■ ESXi hosts ■ NSX for vSphere components ■ Cloud Management Platform components ■ vRealize Log Insight components	The first 3 medium-size nodes take more resources per 1,000 virtual machines because they have to 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.	■ Ensures that the analytics cluster has enough capacity to meet the virtual machine object and metric growth up to 10,000 virtual machines. ■ Ensures that the management cluster always has enough physical capacity to take a host offline for maintenance or other reasons.	■ The capacity of the physical ESXi hosts must be enough to accommodate virtual machines that require 32 GB RAM without bridging NUMA node boundaries. ■ The management cluster must have enough ESXi hosts so that vRealize Operations Manager can run according to vSphere DRS anti-affinity rules. ■ The number of nodes must not exceed `number of ESXi hosts in the management cluster – 1`.

Table 2-133. Design Decisions on the Compute Size of the Analytics Cluster Nodes of vRealize Operations Manager (continued)

Decision ID	Design Decision	Design Justification	Design Implication
			For example, if the management cluster contains 6 ESXi hosts, you can deploy up to 5 vRealize Operations Manager nodes in the analytics cluster.

Sizing Compute Resources for the Remote Collector Nodes

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

Table 2-134. 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	6,000
Single-Node Maximum Collected Metrics (*)	1,200,000
Multi-Node Maximum Objects Per Node	N/A
Multi-Node Maximum Collected Metrics Per Node	N/A
Maximum number of End Point Operations Management agents per node	250
Maximum Objects for 16-Node Configuration	N/A
Maximum Metrics for 16-Node Configuration	N/A

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

Table 2-135. Design Decisions on the Compute Size of the Remote Collector Nodes of vRealize Operations Manager

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 4 vCPUs and 8 GB 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 that is defined in the objectives of this SDDC design.

This design uses medium-size nodes for the analytics cluster and standard-size nodes for the remote collector group. To collect the required number of metrics, you must add an additional virtual disk with the size of 1 TB 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, the need to add more data nodes to the analytics cluster will emerge. To plan the sizing requirements of your environment, refer to the vRealize Operations Manager sizing guidelines in VMware Knowledge Base article 2093783.

Table 2-136. Design Decision on the Storage Size of the Analytics Cluster of vRealize Operations Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-009	Add a virtual disk of 1 TB for each analytics cluster node.	Provides enough storage to meet the SDDC design objectives.	You must add the 1-TB 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 2-137. Design Decision on the Storage Size of the Remote Collector Nodes of vRealize Operations Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-010	Do not provide more 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 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, you deploy the vRealize Operations Manager analytics cluster in the shared cross-region application virtual network `Mgmt–xRegion01–VXLAN`, and the remote collector clusters in the shared local application isolated networks `Mgmt–RegionA01–VXLAN` and `Mgmt–RegionB01–VXLAN`.

Figure 2-31. Networking Design of the vRealize Operations Manager Deployment

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 can be failed over between regions after scaling out to a multi-region design. vRealize Automation and vRealize Business also share this network.

- All nodes have routed access to the vSphere management network through the NSX Universal Distributed Logical Router.

- Routing to the vSphere management network and other external networks is dynamic and is based on the Border Gateway Protocol (BGP).

For more information about the networking configuration of the application virtual network, see Virtualization Network Design and NSX Design .

Table 2-138. Design Decisions on the Application Virtual Network for vRealize Operations Manager

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.
SDDC-OPS-MON-012	Use the existing region-specific application virtual networks for vRealize Operations Manager remote collectors.	Ensures collection of metrics locally per region in the event of a cross-region network outage. It also co-locates metric collection with the region-specific applications 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 example subnets for each cluster in the vRealize Operations Manager deployment.

Table 2-139. IP Subnets in the Application Virtual Network for vRealize Operations Manager

vRealize Operations Manager Cluster Type	IP Subnet
Analytics cluster in Region A	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 2-140. Design Decision on the IP Subnets for vRealize Operations Manager

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.

FQDNs for vRealize Operations Manager

The FQDNs of the vRealize Operations Manager nodes follow a certain domain name resolution:

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

  From the public network, users access vRealize Operations Manager using the VIP address, the traffic to which is handled by an 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`.

- Name resolution for the IP addresses of the remote collector group nodes uses a root domain suffix, for example, `rainpole.local`

- The IP addresses of the remote collector group nodes are associated with names whose suffix is set to the root domain, for example, `rainpole.local`

Table 2-141. FQDNs for the vRealize Operations Manager Nodes

FQDN	Node Type	Region
vrops01svr01.rainpole.local	Virtual IP of the analytics cluster	Region A (failover to Region B)
vrops01svr01a.rainpole.local	Primary node in the analytics cluster	Region A (failover to Region B)
vrops01svr01b.rainpole.local	Primary 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 2-142. Design Decision on the DNS Names for vRealize Operations Manager

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 fully qualified domain names instead of by using IP addresses only.	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 the 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 2-143. Design Decisions on Networking Failover and Load Balancing for vRealize Operations Manager

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

Authentication and Authorization

Users can authenticate to 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 that have administration access in vCenter Server.

- Users that have one of the vRealize Operations Manager privileges, such as **PowerUser**, assigned to the account which appears at the root level in vCenter Server.

Create local user accounts in vRealize Operations Manager

vRealize Operations Manager performs local authentication using the account information stored in its global database.

Table 2-144. Design Decisions on Authorization and Authentication Management for vRealize Operations Manager

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. ■ Ensures that authentication is available even if vCenter Server becomes unavailable.	You must manually configure the 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, 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 ensure its availability.

Table 2-144. Design Decisions on Authorization and Authentication Management for vRealize Operations Manager (continued)

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 metric collection and topology mapping. ■ 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.	You must maintain the service account's life cycle outside of the SDDC stack to ensure 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, 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 life cycle outside of the SDDC stack to ensure its availability.
SDDC-OPS-MON-021	Configure a service account **svc-vrops-vsan** in vCenter Server for application-to-application communication from the vSAN 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 vSAN inventory 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.	You must maintain the service account's life cycle outside of the SDDC stack to ensure its availability.

Table 2-144. Design Decisions on Authorization and Authentication Management for vRealize Operations Manager (continued)

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, 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 life cycle outside of the SDDC stack to ensure its availability.
SDDC-OPS-MON-023	Use global permissions when you create the **svc-vrops-vsphere**, **svc-vrops-nsx**, **svc-vrops-vsan**, **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.

Table 2-144. Design Decisions on Authorization and Authentication Management for vRealize Operations Manager (continued)

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, 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 life cycle outside of the SDDC stack to ensure 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 metric collection and topology mapping. ■ 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.	You must maintain the service account's life cycle outside of the SDDC stack to ensure 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. To provide secure access to the vRealize Operations Manager user interface, replace the default self-signed certificates with a CA-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 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 single pane of glass overview.

Table 2-145. Design Decisions on Monitoring vRealize Operations Manager

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 2-146. vRealize Operations Manager Management Packs in VMware Validated Design

Management Pack	Installed by Default
Management Pack for VMware vCenter Server	X
Management Pack for NSX for vSphere	
Management Pack for vSAN	X
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 2-147. Design Decisions on the Management Packs for vRealize Operations Manager

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 vSAN ■ Management Pack for vRealize Automation ■ Management Pack for vRealize Business for Cloud	You must install and configure each non-default 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, though 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 ■ vSAN ■ 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, the design of vRealize Operations Manager supports failing over a sub-set of the components between regions. Disaster recovery covers only the analytics cluster components, including the primary, replica and data nodes. The region-specific remote collector nodes remain in the affected region.

When a disaster occurs, you 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 only collect data 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 for more details.

vRealize Log Insight Design

vRealize Log Insight design enables real-time logging for all components that build up the management capabilities of the SDDC.

- Logical Design and Data Sources of vRealize Log Insight

  vRealize Log Insight collects log events from all management components in each region of the SDDC.

- Node Configuration of vRealize Log Insight

- 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 properly.

- Networking Design of vRealize Log Insight

- 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.

- Alerting in vRealize Log Insight

  vRealize Log Insight supports alerts that trigger notifications about its health and about the health of monitored solutions.

- 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.

- 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.

- Collecting Logs in vRealize Log Insight

  As a part of vRealize Log Insight configuration, you configure syslog and vRealize Log Insight agents.

- Time Synchronization in vRealize Log Insight

  Time synchronization is important for the operation of vRealize Log Insight. By default, vRealize Log Insight synchronizes time with a pre-defined list of public NTP servers.

- Content Packs in vRealize Log Insight

  Use content packs to have the logs generated from the management components in the SDDC 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.

- Event Forwarding Between Regions with vRealize Log Insight

    vRealize Log Insight supports event forwarding to other clusters and standalone instances. Use log forwarding between SDDC regions to have access to all logs if a disaster occurs in a region.

- Disaster Recovery of vRealize Log Insight

    Each region is configured to forward log information to the vRealize Log Insight instance in the other region.

Logical Design and Data Sources of vRealize Log Insight

vRealize Log Insight collects log events from all management components in each region of the SDDC.

Logical Design

In a multi-region Software-Defined Data Center, deploy a vRealize Log Insight cluster that consists of three nodes in each region. This configuration provides continued availability and increased log ingestion rates.

Figure 2-32. Logical Design of vRealize Log Insight

Sources of Log Data

vRealize Log Insight collects logs as to provide monitoring information about the SDDC from a central location.

vRealize Log Insight 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 universal distributed logical router instances
  - NSX distributed firewall ESXi kernel module
- vRealize Suite Lifecycle Manager
- vRealize Automation
  - vRealize Automation Appliance
  - vRealize Orchestrator (embedded in the vRealize Automation Appliance)
  - vRealize IaaS Web Server
  - vRealize IaaS Management Server
  - vRealize IaaS DEM
  - vRealize IaaS Proxy Agents
  - Microsoft SQL Server
- vRealize Business
  - vRealize Business server
  - vRealize Business data collector
- vRealize Operations Manager
  - Analytics cluster nodes
  - Remote collectors
- 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 primary 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 need not reconfigure all log sources with a new destination address in a future scale-out. Using the ILB also guarantees that vRealize Log Insight accepts all incoming ingestion traffic.

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

A vRealize Log Insight cluster can scale out to 12 nodes, that is, 1 primary and 11 worker nodes.

Table 2-148. Design Decisions on the Node Configuration of vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-001	In each region, deploy vRealize Log Insight in a cluster configuration of 3 nodes with an integrated load balancer: one primary and two worker nodes.	- Provides high availability. - Using the integrated load balancer prevents a single point of failure. - Using the integrated load balancer simplifies the vRealize Log Insight deployment and subsequent integration. - Using the integrated load balancer simplifies the vRealize Log Insight scale-out operations reducing the need to reconfigure existing logging sources.	- You must deploy a minimum of 3 medium nodes. - You must size each node identically. - If the capacity of your vRealize Log Insight cluster must grow, identical capacity must be added to each node.
SDDC-OPS-LOG-002	Apply vSphere Distributed Resource Scheduler (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 in maintenance mode only a single ESXi host 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 properly.

By default, the vRealize Log Insight virtual appliance uses the predefined values for small configurations, which have 4 vCPUs, 8 GB of virtual memory, and 530.5 GB of disk space provisioned. vRealize Log Insight uses 100 GB of the disk space to store raw data, index, metadata, and other information.

Sizing Nodes

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

Table 2-149. Compute Resources for a vRealize Log Insight Medium-Size Node

Attribute	Specification
Appliance size	Medium
Number of CPUs	8
Memory	16 GB
Disk Capacity	530.5 GB (490 GB for event storage)
IOPS	1,000 IOPS
Amount of processed log data when using log ingestion	75 GB/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 usually based on IT organization requirements. However, this design provides calculations that are based on a single-region implementation, and is then implemented on a per-region basis. This sizing is calculated according to the following node configuration per region:

Table 2-150. Management Systems That Send Log Data to 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 ■ Universal distributed logical router ■ 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: ■ Universal distributed logical router ■ Distributed logical router ■ Two ESGs for north-south routing	4

Table 2-150. Management Systems That Send Log Data to vRealize Log Insight (continued)

Category	Logging Sources	Quantity
vRealize Suite Lifecycle Manager	vRealize Suite Lifecycle Manager Appliance	1
vRealize Automation	vRealize Automation Appliance with embedded vRealize Orchestrator	3
	vRealize IaaS Web Server	2
	vRealize IaaS Manager Server	2
	vRealize IaaS DEM	2
	vRealize Proxy 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 provide approximately 111 syslog and vRealize Log Insight Agent sources per region, or 225 sources in a cross-region configuration.

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

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

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

- The rate of 190 MB of logs per day is valid for Windows where 220 bytes 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)
```

In this validated design, to simplify calculation, all calculations have been done using the large 220 byte size which results in 190 MB of log data expected per-day per-source.

For 225 logging sources, at a basal rate of approximately 190 MB of logs that are ingested per-day per-source over 7 days, you need the following storage space:

Calculate the storage space required for a single day for log data using the following calculation:

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

Based on the amount of data stored in a day, to size the appliance for 7 days of log retention, 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.

Consider the following approaches when you must increase the Log Insight capacity:

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

    When you add storage so that you can increase the retention period, extend the storage for all virtual appliances. To increase the storage, add new virtual hard disks only. Do not extend existing retention virtual disks. To avoid data loss, once provisioned, do not reduce the size or remove virtual disks .

- If you must monitor more components by using log ingestion and exceed the number of syslog connections or ingestion limits defined in this design, you can do the following:

    - 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 out your environment. vRealize Log Insight can scale up to 12 nodes in an HA cluster.
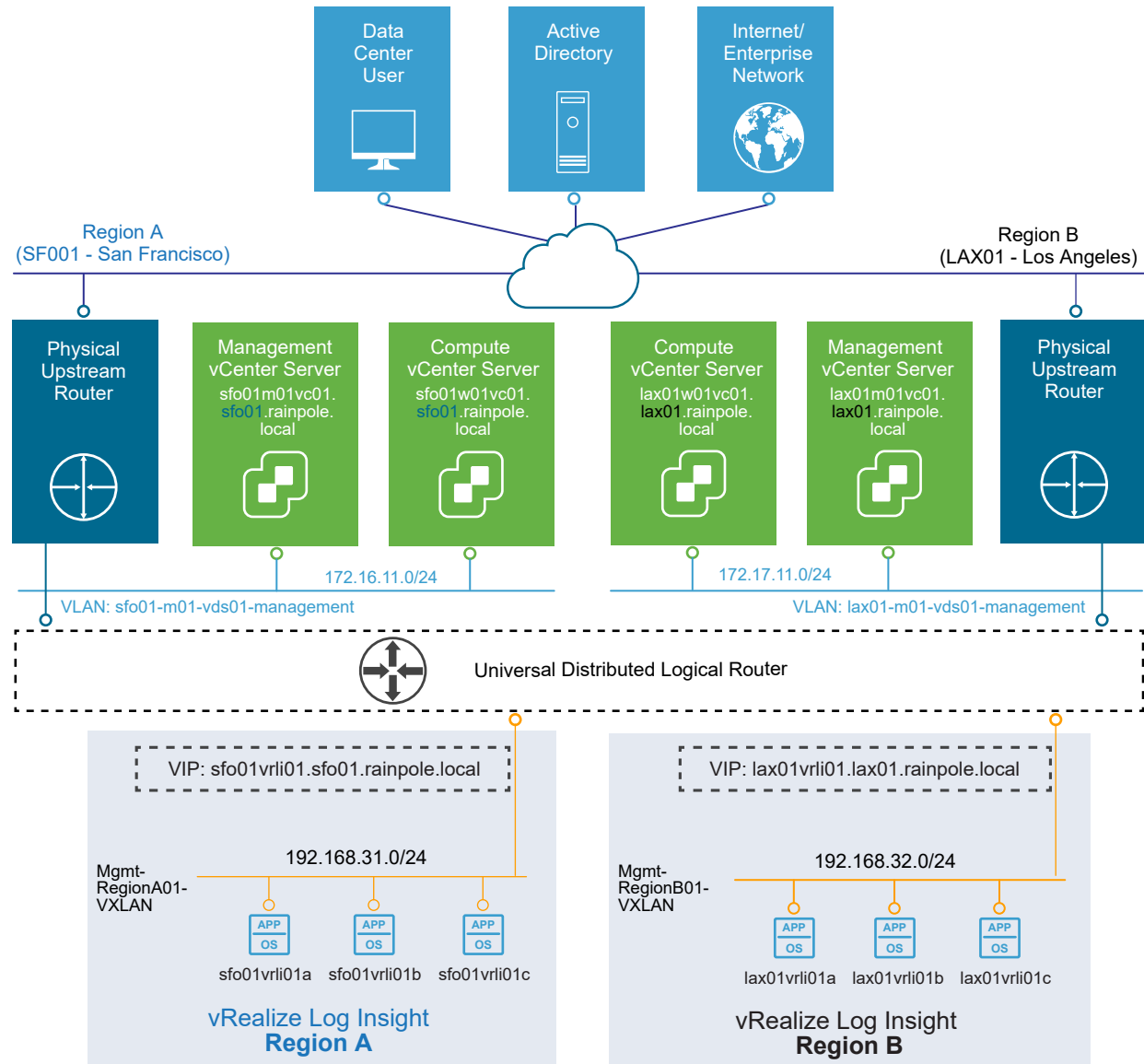
Table 2-151. Design Decisions on the Compute Resources for the vRealize Log Insight Nodes

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 and Compute vCenter Server, and 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 Suite Lifecycle Manager ■ vRealize Automation components ■ vRealize Business components ■ vRealize Operations Manager components ■ Cross- vRealize Log Insight cluster event forwarding. These components approximately generate about 225 syslog and vRealize Log Insight Agent sources. Using medium-size appliances ensures that the storage space for the vRealize Log Insight cluster is sufficient for 7 days of data retention.	You must increase the size of the nodes if you configure vRealize Log Insight to monitor additional syslog sources.

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` for isolation and co-location with logging sources. The networking design also supports public access to the vRealize Log Insight cluster.

Figure 2-33. Networking Design for the 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 universal distributed logical router (UDLR) for the management cluster for the home region.

- Routing to the vSphere management network and the external network is dynamic, and is based on the Border Gateway Protocol (BGP).

For more information about the networking configuration of the application virtual networks for vRealize Log Insight, see Application Virtual Network and Virtual Network Design Example .

Table 2-152. Design Decision on Networking for vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-004	Deploy vRealize Log Insight on the region-specific application virtual networks.	■ Ensures centralized access to log data per region if a cross-region network outage occurs. ■ Co-locates 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 impact event forwarding between the vRealize Log Insight clusters and 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 example subnets to the vRealize Log Insight deployment.

Table 2-153. 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

FQDNs for vRealize Log Insight

vRealize Log Insight node name resolution, including the load balancer virtual IP addresses (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 2-154. FQDNs of the vRealize Log Insight Nodes

FQDN	Role	Region
sfo01vrli01.sfo01.rainpole.local	Log Insight ILB VIP	Region A
sfo01vrli01a.sfo01.rainpole.local	Primary 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	Primary node	Region B
lax01vrli01b.lax01.rainpole.local	Worker node	Region B
lax01vrli01c.lax01.rainpole.local	Worker node	Region B
lax01vrli01x.lax01.rainpole.local	Additional worker nodes (not deployed)	Region B

Table 2-155. Design Decisions on FQDNs for vRealize Log Insight

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 fully qualified domain names instead of by using IP addresses only.	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 VIP address of vRealize Log Insight in Region A when you configure logging.	Lgging continues 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, you must 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 2-156. Virtual Disk Configuration in the vRealize Log Insight Virtual Appliance

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

Calculate the storage space that is available for log data 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 490 GB. If `/storage/core` is 490 GB, vRealize Log Insight can use 475 GB 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
```

You can 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 2-157. Design Decision on Retention Period in vRealize Log Insight

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

You 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.

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 an 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 will each use approximately 400 GB of shared storage calculated via the following:

```
(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
```

According to the business compliance regulations of your organization, these sizes might change.

Table 2-158. Design Decision on Log Archive Policy for vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-008	Provide 400 GB of NFS version 3 shared storage to each vRealize Log Insight cluster.	Accommodates log archiving from 225 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. ■ You must increase the size of the NFS shared storage if you configure vRealize Log Insight to monitor more logging sources or add more vRealize Log Insight workers are added. ■ 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, becomes available, or archiving is disabled. ■ When using two availability zones, ensure that the NFS share is available in both availability zones.

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, when the disk space is almost exhausted and 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.

vRealize Log Insight handles alerts in two ways:

- Send an e-mail over SMTP.

- Send to vRealize Operations Manager.

SMTP Notification

Enable e-mail notification for alerts in vRealize Log Insight.

Table 2-159. Design Decision on SMTP Alert Notification for vRealize Log Insight

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.

You can use the following integration points that 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 2-160. Design Decisions on Integration of vRealize Log Insight with vRealize Operations Manager

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 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 2-161. Design Decisions on Authorization and Authentication Management in vRealize Log Insight

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 the 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 life cycle outside of the SDDC stack to ensure 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 the 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, 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 life cycle outside of the SDDC stack to ensure its availability.

Encryption

To provide secure access to the vRealize Log Insight Web user interface, replace the default self-signed certificates with a CA-signed certificate.

Table 2-162. Design Decision on CA-Signed Certificates for vRealize Log Insight

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 ensures 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 using the syslog TCP, syslog TCP over TLS/SSL, or syslog UDP protocols

- 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 2-163. Design Decisions on Log Communication to vRealize Log Insight

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 virtual IP (VIP) address of the vRealize Log Insight integrated load balancer (ILB).	■ Allows for 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 integrated load balancer 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 using the default Ingestion API (cfapi), default disk buffer of 200 MB and non-default No SSL.	■ Supports multi-line message transmissions from logs. ■ Provides ability to add metadata to events generated from system. ■ Provides client-side compression, buffering, and throttling capabilities ensuring minimal to no message loss during intermittent connection issues ■ Provides server-side administration, metric collection, 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.
SDDC-OPS-LOG-020	Configure the vRealize Log Insight agent on the vRealize Suite Lifecycle Manager appliance.	Simplifies configuration of log sources in the SDDC that are pre-packaged 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-021	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-022	Configure the vRealize Log Insight agent on the vRealize Automation appliance.	Simplifies configuration of log sources in the SDDC that are pre-packaged 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 Business appliances including: ■ Server appliance ■ Data collectors	Simplifies configuration of log sources in the SDDC that are pre-packaged with the vRealize Log Insight agent.	You must configure the vRealize Log Insight agent to forward logs to the vRealize Log Insight VIP.

Table 2-163. Design Decisions on Log Communication to vRealize Log Insight (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-024	Configure the vRealize Log Insight agent for the vRealize Operations Manager appliances including: ■ Analytics nodes ■ Remote collectors	Simplifies configuration of log sources in the SDDC that are pre-packaged 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-025	Configure the NSX for vSphere components as direct syslog sources for vRealize Log Insight including: ■ NSX Manager ■ NSX Controllers ■ 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-026	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.
SDDC-OPS-LOG-027	Configure vRealize Log Insight to ingest events, tasks, and alarms from the Management vCenter Server and Compute vCenter Server instances .	Ensures that 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 events, tasks, and alarms pulling. ■ Configuring vSphere Integration within vRealize Log Insight does not capture events that occur on the Platform Services Controller.

Table 2-163. Design Decisions on Log Communication to vRealize Log Insight (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-028	Communicate with the syslog clients, such as ESXi, vCenter Server, NSX for vSphere, using the default syslog UDP protocol.	■ Using the default UDP syslog protocol simplifies configuration for all syslog sources ■ UDP syslog protocol is the most common logging protocol that is available across products. ■ UDP has a lower performance overhead compared to TCP.	■ If the network connection is interrupted, the syslog traffic is lost. ■ UDP syslog traffic is not secure. ■ UDP syslog protocol does not support reliability and retry mechanisms.
SDDC-OPS-LOG-029	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 ensures 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-030	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-031	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 maintain manually the vRealize Log Insight Agents on each of the SDDC components.

Time Synchronization in vRealize Log Insight

Time synchronization is important for the operation of vRealize Log Insight. By default, vRealize Log Insight synchronizes time with a pre-defined list of public NTP servers.

NTP Configuration

Configure consistent NTP sources on all systems that send log data (vCenter Server, ESXi, vRealize Operation Manager). See *Time Synchronization* in the *VMware Validated Design Planning and Preparation* documentation.

Table 2-164. Design Decision on Time Synchronization in vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-032	Configure consistent NTP sources on all virtual infrastructure and cloud management applications for correct log analysis in vRealize Log Insight.	Guarantees accurate log timestamps.	All applications must synchronize time to the same NTP time source.

Content Packs in vRealize Log Insight

Use content packs to have the logs generated from the management components in the SDDC 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 2-165. vRealize Log Insight Content Packs in VMware Validated Design

Content Pack	Installed in vRealize Log Insight by Default
General	X
VMware - vSphere	X
VMware - NSX for vSphere	
VMware - vSAN	X
VMware - vRops	X
VMware - vRA	
VMware - Orchestrator	
VMware - vRealize Business for Cloud	
Microsoft - SQL Server	
VMware - Linux	
VMware - SRM	

Table 2-166. Design Decisions on Content Packs for vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-033	Install the following content packs: ■ VMware - Linux ■ VMware - NSX-vSphere ■ VMware - Orchestrator ■ VMware - SRM ■ VMware - vRA ■ VMware - vRealize Business for Cloud ■ Microsoft - SQL Server	Provides additional granular monitoring on the virtual infrastructure. You skip the installation of the following content packs because they are installed by default in vRealize Log Insight: ■ General ■ VMware - vSphere ■ VMware - vSAN ■ VMware - vRops	Requires manual installation and configuration of each non-default content pack.
SDDC-OPS-LOG-034	Configure the following agent groups that are related to content packs: ■ vRealize Automation (Linux) ■ vRealize Automation (Windows) ■ VMware Virtual Appliances ■ vRealize Orchestrator ■ VMware Site Recover Manager ■ Microsoft SQL Server ■ Linux	■ Provides a standardized configuration that is pushed to the all vRealize Log Insight Agents in each of the groups. ■ Supports collection according to the context of the applications and parsing of the logs generated from the SDDC components by the vRealize Log Insight agent such as specific log directories, log files, and logging formats.	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. Use log forwarding between SDDC regions to have access to all logs if a disaster occurs in a region.

You forward syslog data in vRealize Log Insight by using the Ingestion API or a native syslog implementation. While forwarding events, the vRealize Log Insight instance still ingests, stores, and archives events locally.

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

■ Support for both structured and unstructured data, that is, multi-line messages

■ Metadata in the form of tags

■ Client-side compression

■ Configurable disk-backed queue to save events until the server acknowledges the ingestion

Table 2-167. Design Decisions on Event Forwarding Across Regions in vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-035	Forward log events to the other region by using the Ingestion API.	Supports the following operations: ■ Structured and unstructured data for client-side compression ■ Event throttling from one vRealize Log Insight cluster to the other. 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 introduces 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 identical size on both source and destination clusters.
SDDC-OPS-LOG-036	Configure log forwarding to use SSL.	Ensures that the log forward 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 that the Java keystore of the nodes before SSL can be used.
SDDC-OPS-LOG-037	Configure disk cache for event forwarding to 2,000 MB (2 GB).	Ensures 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 150 MB per day per 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 non-persistent 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 they were deployed.

Cloud Management Design

The Cloud Management Platform (CMP) is the cloud management component of the SDDC. You use the CMP to provision workload to tenants in an automated way by using a self-service portal.

The cloud management layer includes the following components and functionality:

Figure 2-34. The Cloud Management Layer in the Software-Defined Data Center

Service Catalog

> Provides a self-service portal where tenants can browse and request the IT services and resources they need, such a virtual machine on vSphere or on Amazon Web Services (AWS). You request a service catalog item to provision the item to the associated cloud environment.

Self-Service Portal

> Provides a unified interface for configuring and 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 tenants. You use the workflows to create and run automated, configurable processes to manage your SDDC infrastructure, and other VMware and third-party technologies.

vRealize Automation provides the self-service portal and the service catalog. The instance of vRealize Orchestrator that is embedded in the vRealize Automation appliance implements the Orchestration module.

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.

Logical Design of vRealize Automation

vRealize Automation provides several extensibility options to support various use cases and integrations. In addition, the Cloud Management Platform, of which vRealize Automation is a central component, enables a usage model that includes interactions between users, the Cloud Management Platform itself, and integrations with the supporting infrastructure.

Figure 2-35. Logical Design, Extensibility, and External Integrations of vRealize Automation

Fabric Endpoints

vRealize Automation can use existing and future infrastructure that represents multi-vendor, multi-cloud virtual, physical, and public cloud infrastructures. Each support type of 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 (ITSM) applications such as ServiceNow, PMG Digital Business Platform, and BMC Remedy.

vRealize Business

vRealize Business integrates with vRealize Automation to manage resource costs of provisioned workloads by displaying costing information in the following cases:

- At workload request

- On an ongoing basis with cost reporting by user, business group, or tenant

Pricing can be on blueprints, endpoints, reservations, and reservation policies for Compute Grouping Strategy. In addition, vRealize Business for Cloud supports the storage path and storage reservation policies for Storage Grouping Strategy.

vRealize Operations Manager

The vRealize Operations management pack for vRealize Automation provides performance and capacity metrics of a tenant's business groups and underlying cloud infrastructure.

Supporting Infrastructure

vRealize Automation integrates with the following supporting infrastructure:

- Microsoft SQL Server to store data relating to the vRealize Automation IaaS elements.

- NTP server for synchronization of the time between the vRealize Automation components

- Active Directory for tenant user authentication and authorization

- SMTP for sending and receiving notification emails for various actions that can be run in the vRealize Automation console.

NSX

NSX and vRealize Automation integration supports designing and authoring blueprints using the networking and security features of NSX. You can use all NSX network constructs such as logical switches, routers, and firewalls.

In a blueprint, you can place an on-demand load balancer, NAT network, routed network , and security groups. When a user requests the blueprint, vRealize Automation automatically provisions these constructs on NSX.

You can directly configure network provisioning as a part of the blueprint design instead of as a separate operation outside vRealize Automation.

Cloud Management Platform Usage Model

The Cloud Management Platform, of which vRealize Automation is a central component, enables a usage model that includes interaction between users, the platform itself, the supporting infrastructure, and the provisioning infrastructure.

Figure 2-36. vRealize Automation Usage Model

The usage model of vRealize Automation contains the following elements and components in them:

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 that can provision virtual machines and directly perform operations on them at the level of the operating system.
Tools and supporting infrastructure	VM templates and blueprints. VM templates are used to author the blueprints that tenants (end users) use to request cloud workloads.

Element	Components
Provisioning infrastructure	On-premise and off-premise resources which together form a hybrid cloud. **Internal Virtual Resources** Supported hypervisors and associated management tools. **External Cloud Resources** Supported cloud providers and associated APIs.
Cloud management portal	Self-service capabilities for users to administer, provision, and manage workloads. **vRealize Automation portal, Admin access** The root portal URL for the default tenant that is used to set up and administer tenants and global configuration options. **vRealize Automation portal, Tenant access** The portal URL for a custom tenant which you access by appending a tenant identifier.

Physical Design of vRealize Automation

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 the ability to provision to both Regions A and B.

To accomplish this design objective, you deploy or re-use the following components in Region A to create a cloud management portal of the SDDC.

- 3 vRealize Automation appliances

- 2 vRealize Automation IaaS web servers

- 2 vRealize Automation Manager Service nodes (including the DEM Orchestrator)

- 2 DEM Worker nodes

- 2 IaaS Proxy Agent nodes

- 1 vRealize Business server

- 1 vRealize Business data collector

- Supporting infrastructure such as Microsoft SQL Server, Active Directory, DNS, NTP, and SMTP.

You place the vRealize Automation components in several network units for isolation and failover.

Figure 2-37. vRealize Automation Design for Region A

Figure 2-38. vRealize Automation Design for Region B

- Deployment Considerations for vRealize Automation

- vRealize Automation Appliance

  The vRealize Automation appliance includes the cloud management Web portal, an embedded vRealize Orchestrator instance and database services. You consider the number of appliances, sizing for them and database replication mode according to the design objectives for the number of virtual machines that system must provision.

- IaaS Web Server and Model Manager of vRealize Automation

  vRealize Automation IaaS Web server provides a user interface in the vRealize Automation portal for consumption of IaaS components. The same virtual machine also runs the Model Manager that communicates with the other IaaS components. In the design of the IaaS Web server, consider the number of instances according to the design objectives and the sizing for them.

- IaaS Manager Service and DEM Orchestrator of vRealize Automation

  The IaaS Manager Service and Distributed Execution Manager (DEM) of vRealize Automation handle the business logic, and maintain the workflows for provisioning and lifecycle management of virtual machines. Consider the number of instances according to the design objectives and the sizing for these components.

- **IaaS DEM Worker of vRealize Automation**

  vRealize Automation IaaS DEM Workers are responsible for provisioning and deprovisioning tasks that are initiated by the vRealize Automation portal. DEM Workers also communicate with certain infrastructure endpoints.

- **IaaS Proxy Agent for vRealize Automation**

  The IaaS Proxy Agent in vRealize Automation communicates with endpoints of a certain type of infrastructure. In this design, you use a vSphere Proxy Agent to communicate with vCenter Server. Consider the number of instances according to the design objectives and the sizing for them.

- **Networking Design of vRealize Automation**

  As part of this design, use the application virtual network configuration to connect the vRealize Automation with the other management solutions in the SDDC. Use the load balancer in the cross-region application virtual network for high availability and request balancing of the vRealize Automation components.

- **Information Security and Access Control in vRealize Automation**

  For orchestrating and creating virtual objects in the SDDC, you use a service account for authentication and authorization of vRealize Automation to vCenter Server and vRealize Operations Manager.

Deployment Considerations for vRealize Automation

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 2-168. Design Decisions on the vRealize Automation Topology

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-001	Deploy 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. Because of the isolation of the vRealize Automation application over virtual networking, it is independent from any physical site locations or hardware.	You must size vRealize Automation to accommodate multi-region deployments.
SDDC-CMP-002	Deploy a large-scale configuration of vRealize Automation.	Deploying the three-node architecture satisfies the design objective of 10,000 virtual machines being provisioned and managed in the scope of the initial deployment architecture. This design enables future growth of up to 50,000 virtual machines after you expand the underlying virtual and physical infrastructure.	vRealize Automation components use more resources than the design objectives define.

Table 2-169. Design Decisions on Anti-Affinity Rules for vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-003	Apply vSphere Distributed Resource Scheduler (DRS) anti-affinity rules to all vRealize Automation components.	Using vSphere DRS prevents vRealize Automation nodes from residing 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. ■ You can put in maintenance mode only a single ESXi host at a time in a management cluster of four ESXi hosts.

vRealize Automation Appliance

The vRealize Automation appliance includes the cloud management Web portal, an embedded vRealize Orchestrator instance and database services. You consider the number of appliances, sizing for them and database replication mode according to the design objectives for the number of virtual machines that system must provision.

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 2-170. Resource Requirements for the vRealize Automation Appliance per Virtual Machine

Attribute	Specification
Number of vCPUs	4
Memory	18 GB
vRealize Automation function	Portal Web site, Application, vRealize Orchestrator, service catalog and Identity Manager.

Table 2-171. vRealize Automation Virtual Appliance Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-004	Deploy three instances of the vRealize Automation virtual appliance for redundancy. Each of these virtual appliances also runs an embedded vRealize Orchestrator instance.	Enables an active/active/active front-end portal for high availability.	None.
SDDC-CMP-005	Deploy three appliances that replicate data using the embedded PostgreSQL database.	Enables high database availability for the vRealize Automation appliances. The embedded vRealize Orchestrator instance also uses this database.	In this active/passive configuration, automatic failover between the three instances is enabled.

Table 2-171. vRealize Automation Virtual Appliance Design Decisions (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-006	During deployment, configure the vRealize Automation appliances with 18 GB RAM.	Supports deployment of vRealize Automation in environments with up to 95,000 Active Directory users.	For environments with more than 95,000 Active Directory users of vRealize Automation, you must increase virtual RAM to 30 GB.
SDDC-CMP-007	Enable synchronous replication mode for the vRealize Automation appliance cluster.	Supports automatic failover between the primary and secondary PostgreSQL databases.	The performance of vRealize Automation cluster is lower than when you use asynchronous replication.

IaaS Web Server and Model Manager of vRealize Automation

vRealize Automation IaaS Web server provides a user interface in the vRealize Automation portal for consumption of IaaS components. The same virtual machine also runs the Model Manager that communicates with the other IaaS components. In the design of the IaaS Web server, consider the number of instances according to the design objectives and the sizing for them.

The IaaS Web site provides infrastructure administration and service authoring capabilities to the vRealize Automation console. The Web site component communicates with the co-located Model Manager, which provides it with updates from the Distributed Execution Manager (DEM), proxy agents and database.

Table 2-172. Resource Requirements for the vRealize Automation IaaS Web Server per Virtual Machine

Attribute	Specification
vCPUs	2
Memory	8 GB
vNICs	1
Local drives	1
vRealize Automation functions	Model Manager (Web service)
Operating system	Microsoft Windows Server 2016

Table 2-173. Design Decisions on vRealize Automation IaaS Web Server

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-008	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.

IaaS Manager Service and DEM Orchestrator of vRealize Automation

The IaaS Manager Service and Distributed Execution Manager (DEM) of vRealize Automation handle the business logic, and maintain the workflows for provisioning and lifecycle management

of virtual machines. Consider the number of instances according to the design objectives and the sizing for these components.

The IaaS Manager Service and DEM Orchestrator are separate services, but you can install them on the same virtual machine.

The IaaS Manager Service of vRealize Automation has 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, interacting with the IaaS SQL server and with external systems as required.

Each DEM instance is either an orchestrator or a worker. 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 moves the workflow back to the queue.

The DEM Orchestrator performs the following operations:

- Manages the scheduled workflows by creating workflow instances at the scheduled time.

- Allows only one instance of a particular scheduled workflow to run at a given time.

- Pre-processes workflows before they are run. Preprocessing includes checking preconditions for workflows and creating the execution history of the workflow.

Table 2-174. Resource Requirements for the IaaS Model Manager and DEM Orchestrator per Virtual Machine

Attribute	Specification
vCPUs	2
Memory	8 GB
vNICs	1
Local drives	1
vRealize Automation functions	IaaS Manager Service, DEM Orchestrator
Operating system	Microsoft Windows Server 2016

Table 2-175. Design Decisions on vRealize Automation IaaS Model Manager and DEM Orchestrator Server

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-009	Deploy two virtual machines to run both the vRealize Automation IaaS Manager Service and the DEM Orchestrator services in a load-balanced pool.	vRealize Automation installer enables the automatic failover of IaaS Manager server in an outage of the first node. Automatic failover eliminates single point of failure for the Manager Service. The DEM Orchestrator must have strong network connectivity to the Model Manager at all the time.	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 later.

IaaS DEM Worker of vRealize Automation

vRealize Automation IaaS DEM Workers are responsible for provisioning and deprovisioning tasks that are initiated by the vRealize Automation portal. DEM Workers also communicate with certain infrastructure endpoints.

Table 2-176. Resource Requirements for the vRealize Automation DEM Worker per Virtual Machine

Attribute	Specification
vCPUs	2
Memory	8 GB
vNICs	1
Local drives	1
vRealize Automation functions	DEM Worker
Operating system	Microsoft Windows Server 2016

Table 2-177. Design Decisions on vRealize Automation IaaS DEM Worker

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-010	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 more DEM Workers on the DEM host.	If you add more DEM Workers, you must also provide additional resources to run them.

IaaS Proxy Agent for vRealize Automation

The IaaS Proxy Agent in vRealize Automation communicates with endpoints of a certain type of infrastructure. In this design, you use a vSphere Proxy Agent to communicate with vCenter Server. Consider the number of instances according to the design objectives and the sizing for them.

The IaaS Proxy Agent server provides the following functions:

- An IaaS proxy agent can interact with different types of infrastructure components.

- To provision and manage virtual machines, the proxy agent sends requests for resource virtualization to the hypervisor, that is, to vSphere.

  Proxy agents collect data, such as virtual machine data, from the hypervisors.

Table 2-178. Resource Requirements for vSphere Proxy Agent per Virtual Machine

Attribute	Specification
vCPUs	2
Memory	8 GB
vNICs	1
Local drives	1
vRealize Automation functions	vSphere Proxy Agent
Operating system	Microsoft Windows Server 2016

Table 2-179. Design Decisions on the vRealize Automation IaaS Proxy Agent

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-011	Deploy two vRealize Automation IaaS Proxy Agent virtual machines.	Provides redundancy for vSphere connectivity.	You must provide more resources because you deploy multiple virtual machines for this function.

Networking Design of vRealize Automation

As part of this design, use the application virtual network configuration to connect the vRealize Automation with the other management solutions in the SDDC. Use the load balancer in the cross-region application virtual network for high availability and request balancing of the vRealize Automation components.

This design uses NSX logical switches to abstract the vRealize Automation application and its supporting services. You can place the application in any region regardless of the underlying physical infrastructure such as network subnets, compute hardware, or storage types.

Application Virtual Networks

The vRealize Automation appliance and main IaaS components are installed in the cross-region shared application virtual network and the vSphere proxy agent nodes are installed in their region-specific shared application virtual networks.

This networking design has the following features:

- The vRealize Automation appliance and IaaS nodes are on the same network because they can be failed over between regions after scaling out to a multi-region design. vRealize Automation and vRealize Business also share this network.

- All nodes have routed access to the vSphere management network through the NSX Universal Distributed Logical Router.

- Routing to the vSphere management network and other external networks is dynamic and is based on the Border Gateway Protocol (BGP).

For more information about the networking configuration of the application virtual network, see Virtualization Network Design and NSX Design .

Table 2-180. Design Decisions on the Application Virtual Network for vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-012	Place the following components of vRealize Automation on the existing cross-region application virtual network: ■ vRealize Automation appliances ■ IaaS Web Servers ■ IaaS Manager Service instances ■ IaaS DEM Workers	Supports disaster recovery by isolating the vRealize Automation main components on the application virtual network `Mgmt-xRegion01-VXLAN`.	You must use an implementation in NSX to support this network configuration.
SDDC-CMP-013	Place the proxy agent virtual machines in both regions on the region-specific application virtual networks.	Ensures collection of metrics locally per region in the event of a cross-region network outage. It also co-locates metric collection with the region-specific applications using the virtual networks `Mgmt-RegionA01-VXLAN` and `Mgmt-RegionB01-VXLAN`.	You must use an implementation in NSX to support this network configuration

Load Balancer Configuration

By using session persistence on the load balancer, the same server can serve all requests after a session is established with that server. The 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 IaaS Manager Service because only one Manager Service is active at one time. The Manager Service can operate with the use of session persistence.

Consider the following load balancer characteristics for vRealize Automation.

Table 2-181. Specification of the Load Balancer Application Profiles

Server Role	Type	Enable SSL Pass-Through	Persistence	Expires in (Seconds)
vRealize Automation - Persistence	HTTPS (443)	Enabled	Source IP	1800
vRealize Automation	HTTPS (443)	Enabled	-	-

Table 2-182. Specification of the Load Balancer Service Monitoring

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	ProvisionService
vRealize Orchestrator	3	10	3	HTTPS		GET	/vco-controlcenter/docs	

Table 2-183. Specification of the Load Balancer Pools

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 2-184. Specification of the Load Balancer Virtual Servers

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

Table 2-185. Design Decisions on Load Balancing vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-014	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.
SDDC-CMP-015	Configure the load balancer for the vRealize Automation appliance, Remote Console Proxy, and IaaS Web Server to use Round-Robin algorithm with Source-IP based persistence with a 1800 second timeout.	▪ Round-robin provides a good balance of clients between both appliances, while the Source-IP ensures that individual clients remain connected to the same appliance. ▪ 1800-second timeout aligns with the vRealize Automation Appliance Server sessions timeout value. Sessions that transfer to a different vRealize Automation Appliance might result in a poor user experience.	None
SDDC-CMP-016	Configure the load balancer for vRealize IaaS Manager Service to use Round-Robin algorithm without persistence.	The Manager Service does not need session persistence.	None

Information Security and Access Control in vRealize Automation

For orchestrating and creating virtual objects in the SDDC, you use a service account for authentication and authorization of vRealize Automation to vCenter Server and vRealize Operations Manager.

Authentication and Authorization

Users can authenticate to vRealize Automation in the following ways:

Import users or user groups from an LDAP database

Users can use their LDAP credentials to log in to vRealize Automation.

Create local user accounts in vRealize Operations Manager

vRealize Automation performs local authentication using account information stored in its global database.

vRealize Automation also authenticates to the following systems:

▪ Compute vCenter Server and NSX for workload provisioning

▪ vRealize Operations Manager for workload reclamation

Table 2-186. Design Decisions on Authorization and Authentication Management for vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-017	Join the vRealize Automation IaaS VMs to Active Directory	Active Directory access is a hard requirement for vRealize Automation.	Active Directory access must be provided using dedicated service accounts.
SDDC-CMP-018	Configure a service account svc-vra in the Compute vCenter Server and NSX for application-to-application communication from vRealize Automation with vSphere and NSX.	Provides the following access control features: ■ The proxy agents in vRealize Automation access vSphere and NSX with the minimum set of permissions that are required to collect metrics about vSphere inventory 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.	You must maintain the service account's life cycle outside of the SDDC stack to ensure its availability.
SDDC-CMP-019	Use local permissions when you create the svc-vra service account in vCenter Server.	Only the Compute vCenter Server instances are valid and accessible endpoints from vRealize Automation.	If you deploy more Compute vCenter Server instances, you must ensure that the service account has been assigned local permissions in each vCenter Server so that this vCenter Server is a valid endpoint in vRealize Automation.
SDDC-CMP-020	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, 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 life cycle outside of the SDDC stack to ensure its availability.

Encryption

Access to all vRealize Automation Web applications requires an SSL connection. By default, vRealize Automation uses a self-signed certificate. To provide secure access to the vRealize Automation user interfaces and between the IaaS components interacting with each other by using Web applications, replace the default self-signed certificates with a CA-signed certificate.

Table 2-187. Design Decision on Using CA-Signed Certificates in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-021	Replace the default self-signed certificates with a CA-signed certificate.	Ensures that all communication to the externally facing Web UI and between the IaaS components is encrypted.	You must contact a certificate authority.

Database Design for vRealize Automation

To satisfy the requirements of this SDDC design, you configure third-party database servers for vRealize Automation.

- Microsoft SQL Server Database for vRealize Automation

  vRealize Automation uses a Microsoft SQL Server database to store information about the vRealize Automation IaaS elements and the machines that vRealize Automation manages.

- PostgreSQL Database Server for vRealize Automation

  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.

Microsoft SQL Server Database for vRealize Automation

vRealize Automation uses a Microsoft SQL Server database to store information about the vRealize Automation IaaS elements and the machines that vRealize Automation manages.

VMware, Inc.

269

Table 2-188. Design Decisions on the Microsoft SQL Server Database for vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-022	Set up a Microsoft SQL Server that supports the availability and I/O needs of vRealize Automation.	You can use a dedicated or shared Microsoft SQL server if it meets the requirements of vRealize Automation.	You must provide additional resources and licenses.
SDDC-CMP-023	Use the existing cross-region application virtual networks for the Microsoft SQL Server or set it up to have global failover available.	For failover of the vRealize Automation instance from one region to another, the Microsoft SQL Server must be running as a VM inside the cross-region application virtual network. If the environment uses a shared Microsoft SQL server, global failover ensures connectivity from both primary and secondary regions.	Adds additional overhead to managing Microsoft SQL services.
SDDC-CMP-024	Set up a Microsoft SQLServer instance with separate OS volumes for SQL Data, Transaction Logs, TempDB, and Backup.	While each organization might have their 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 must consult with the Microsoft SQL database administrators of your organization for guidance about production deployment in your environment.

Table 2-189. vRealize Automation SQL Database Server Resource Requirements per VM

Attribute	Specification
Number of vCPUs	8
Memory	16 GB
Number of vNIC ports	1
Number of local drives	1 40 GB (D:) (Application) 40 GB (E:) Database Data 20 GB (F:) Database Log 20 GB (G:) TempDB 80 GB (H:) Backup
vRealize Automation functions	Microsoft SQL Server Database
Microsoft SQL Version	SQL Server 2017
Microsoft SQL Database Version	SQL Server 2017 Standard or higher edition (64-bit)
Operating system	Microsoft Windows Server 2016

PostgreSQL Database Server for vRealize Automation

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 2-190. Design Decisions on the PostgreSQL Database in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-025	Use the embedded PostgreSQL database server 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 three vRealize Automation appliances.	None.

Notifications Design for vRealize Automation

System administrators configure default settings for both the outbound and inbound emails servers used to send system notifications. Systems 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.

vRealize Automation sends notification emails over 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.

You assign a global, inbound email server to handle inbound email notifications, such as approval responses.

The email server provides accounts that you can customize for each user. 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. vRealize Automation supports both the POP and the IMAP protocols, with or without SSL certificates.

Table 2-191. Design Decisions on Email Server Configuration for vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-026	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 emails (IMAP), and anonymous access can be used with outbound emails.

Cloud Tenant Design for vRealize Automation

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 Between Single-Tenant and Multi-Tenant Deployments of vRealize Automation**

  vRealize Automation supports deployments with a single tenant or multiple tenants. System-wide configuration is always performed using the default tenant, and can then be applied to one or more tenants. For example, system-wide configuration might specify defaults for branding and notification providers.

- **Tenant Design for vRealize Automation**

  This design deploys a single tenant containing two business groups.

Comparison Between Single-Tenant and Multi-Tenant Deployments of vRealize Automation

vRealize Automation supports deployments with a single tenant or multiple tenants. System-wide configuration is always performed using the default tenant, and can then be applied to one or more tenants. For example, system-wide 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. You organize your infrastructure resources, such as cloud or virtual compute resources, in fabric groups and assign a fabric administrator to manage those resources. Fabric administrators can allocate resources in their fabric group to business groups by creating 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 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 the tenant. Tenant-level configuration is isolated from the default tenant, although users with system-wide roles, such as system administrator and IaaS administrator, can view and manage both configurations. The IaaS administrator for the organization tenant creates fabric groups and assigns fabric administrators. Fabric administrators can create reservations for business groups in the organization tenant.

Multi-Tenant Deployment

In a multi-tenant deployment, the system administrator creates tenants for each organization that uses the same vRealize Automation instance. Tenant users log in to the vRealize Automation console at a URL specific to their tenant. Tenant-level configuration is isolated

from other tenants and from the default tenant, although users with system-wide roles can view and manage configuration across multiple tenants. The IaaS administrator for each tenant creates fabric groups and assigns 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 IaaS administrators or fabric administrators for each tenant.

Tenant Design for vRealize Automation

This design deploys a single tenant containing two business groups.

- The first business group is for production workloads provisioning.

- The second business group is for development workloads.

Tenant administrators manage users and groups, 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 per their role.

Figure 2-39. Rainpole Cloud Automation Tenant Design for Two Regions

Table 2-192. Design Decisions on the Tenant Configuration in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-027	Use vRealize Automation business groups for individual business units instead of separate tenants.	Provides 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.
SDDC-CMP-028	Create separate fabric groups for each deployment region. Each fabric group represents region-specific data center resources. Each of the business groups has reservations in 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-029	Let only the system administrator access the default tenant for managing tenants and modifying system-wide configurations.	Isolates the default tenant from individual tenant configurations.	Each tenant administrator is responsible for managing their own tenant configuration.
SDDC-CMP-030	Evaluate your internal organizational structure and workload needs. Configure business groups, reservations, service catalogs, and blueprints in the vRealize Automation instance according to your organization's needs.	The tenant configuration of vRealize Automation represents the needs of your organization. In this design, use the guidance for the Rainpole tenant as a starting point.	Partners and customers must evaluate their business needs.

Service Catalog in vRealize Automation

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. While the catalog does not enforce service-level agreements on services, the information about the service hours, support team, and change window is 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 Server 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 who should have access to it. For

example, some catalog items may be available only to a specific business group, while other catalog items may be shared between business groups using the same tenant. The administrator determines what catalog items are available to different users based on their job functions, departments, 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, such as virtual machine specifications or lease duration, or any additional information that the requester is prompted to provide when submitting the request.

Machine Blueprints in vRealize Automation

A machine blueprint is the complete specification for a virtual, cloud, or physical machine. A machine blueprint determines the machine's attributes, provisioning method, and its policy and management settings. You publish machine blueprints as catalog items in the service catalog.

Machine blueprints can be specific to a business group or shared among groups in a tenant in the following way:

Table 2-193. User Role Rights on Machine Blueprints

Type of Blueprints	Rights of the Tenant Administrator	Rights of the Business Group Manager
Shared blueprints	■ Create ■ Entitle to users in a business group in the tenant ■ Manage	■ Copy if configured by the tenant administrator
Group-specific blueprints	■ View and modify if a business group manager for the group	■ Create ■ Entitle to users in the business group

Blueprint Definitions

Define the services that provide basic workload provisioning to your tenants. This design introduces services for provisioning instances of Windows Server, Linux Server, or Windows Server with SQL Server installed.

Table 2-194. Single-Machine Blueprints in This Validated Design

Name	Description	Business Group
Base Windows Server (Production)	A standard operating environment for the Rainpole tenant for deployment of Windows Server.	Production
Base Linux (Production)	A standard operating environment for the Rainpole tenant for deployment of Linux.	Production
Windows Server with SQL Server (Production)	Base Windows Server with a silent SQL Server installation with custom properties.	Production

Table 2-195. Requirements and Standards for the Base Windows Server Blueprints

Service Name	Base Windows Server
Provisioning Method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter Server customizations.
Entitlement	Both Production and Development business group members.
Approval Process	No approval (pre-approval assumed based on approved access to platform).
Operating System and Version Details	Windows Server
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
Pre- and Post-Deployment Requirements	Email sent to manager confirming service request (include description details).

Table 2-196. Base Windows Blueprint Sizing

Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	4	60
Maximum	4	16	60

Table 2-197. Requirements and Standards for the Base Linux Server Blueprint

Service Name	Base Linux Server
Provisioning Method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter Server customizations.
Entitlement	Both Production and Development business group members.
Approval Process	No approval (pre-approval 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
Pre- and Post-Deployment Requirements	Email sent to manager confirming service request (include description details) .

Table 2-198. Base Linux Blueprint Sizing

Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	6	20
Maximum	4	12	20

Table 2-199. Requirements and Standards for the Blueprint for Base Windows Server with SQL Server

Service Name	Base Windows Server
Provisioning Method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter Server customizations.
Entitlement	Both Production and Development business group members
Approval Process	No approval (pre-approval assumed based on approved access to platform).
Operating System and Version Details	Windows Server Server
Configuration	Disk: Single disk drive Network: Standard vSphere Networks Silent Install: The Blueprint calls a silent script using the vRealize Automation Agent to install SQL 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- and Post-Deployment Requirements	Email sent to manager confirming service request (include description details)

Table 2-200. Sizing of the Base Windows with SQL Server Blueprint

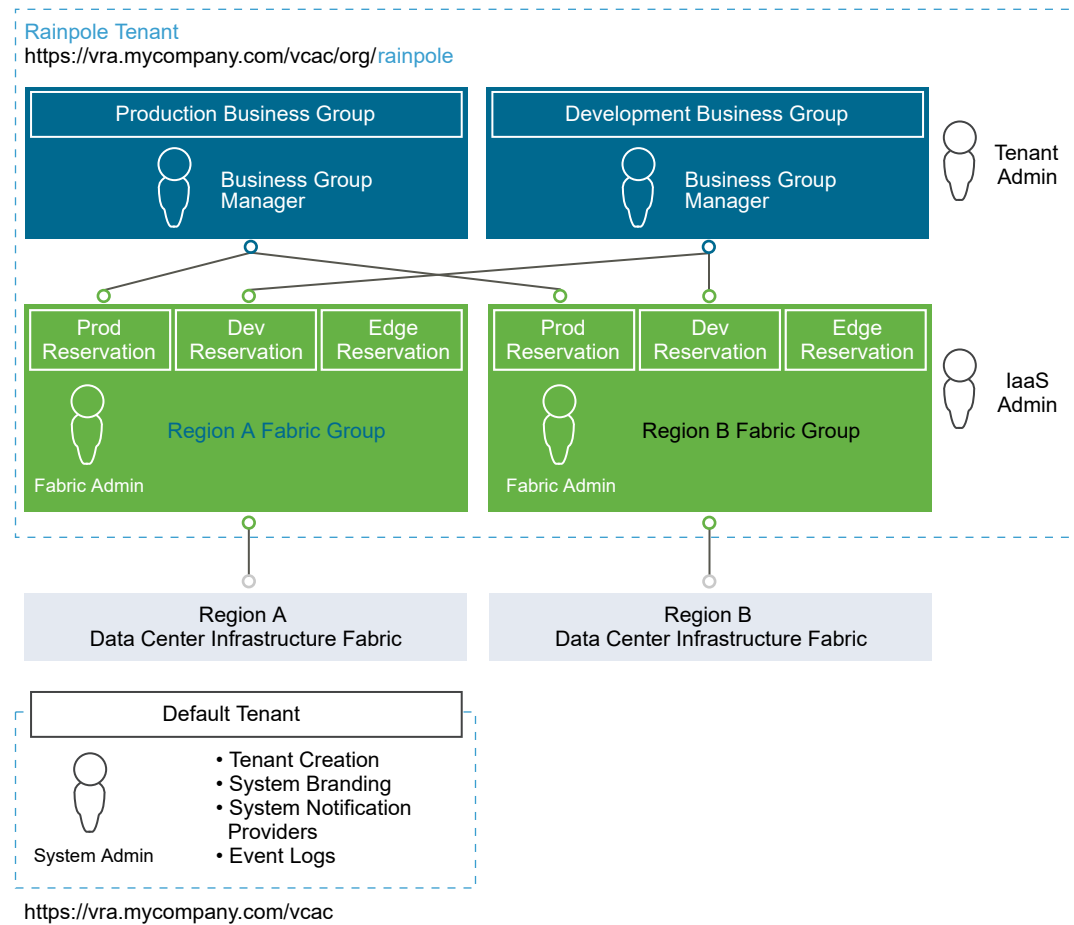| Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	8	100
Maximum	4	16	400

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 reconfigure branding for each tenant.

Infrastructure as a Service Design for vRealize Automation

Design the integration of vRealize Automation with vSphere resources to allocate resources to organizations in the tenant according to the requirements for provisioned workloads and resources policies.

Figure 2-40. Example Infrastructure as a Service Design for vRealize Automation in a Dual-Region Environment
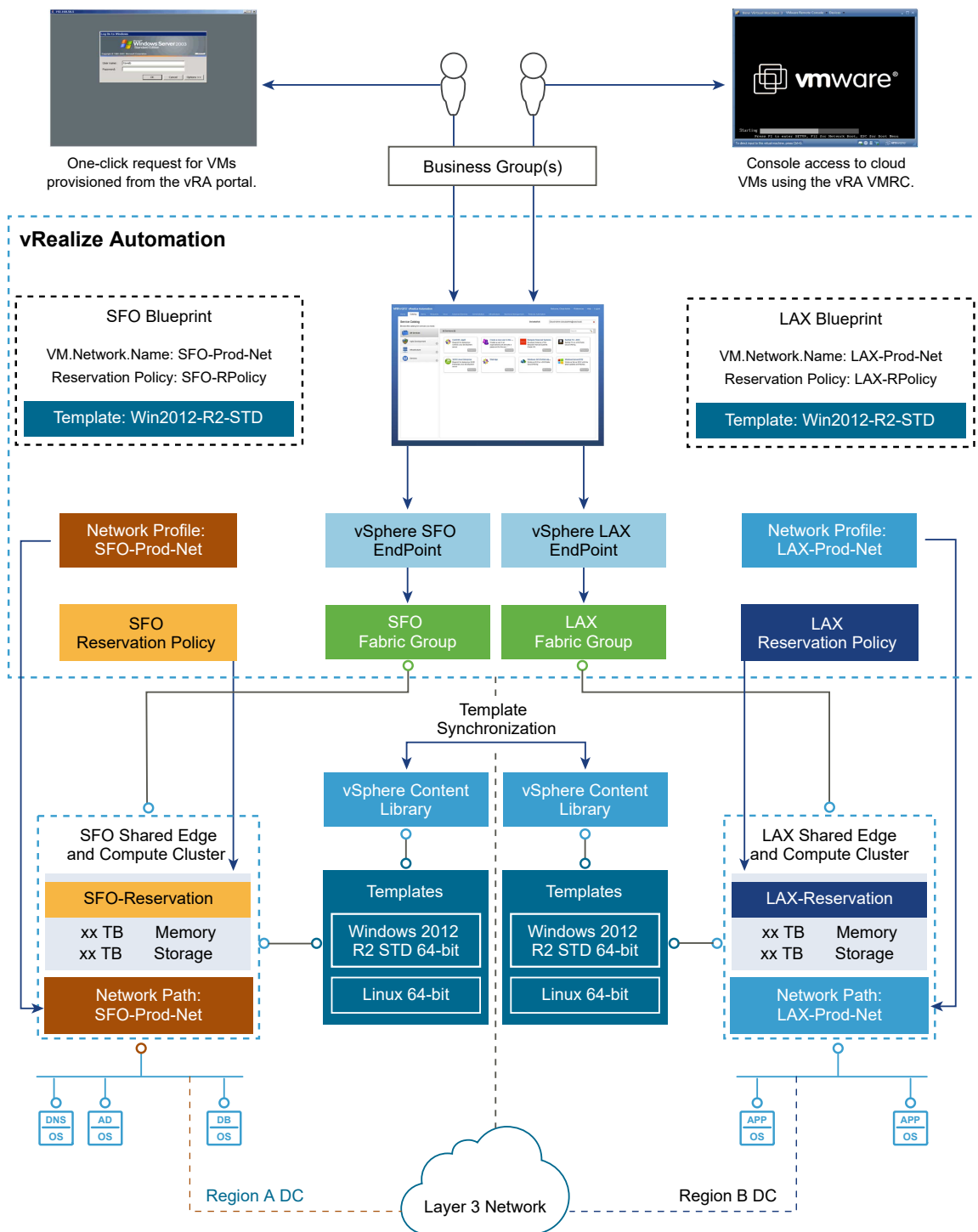

The following terms apply to vRealize Automation when integrated with vSphere. These terms and their meaning vary from the way they are used when referring only to vSphere.

Term	Definition
vSphere (vCenter Server) endpoint	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. Compute resources are CPU, memory, storage and networks. Datastores and datastore clusters are part of the overall storage resources.
Fabric groups	Organization of compute resources
Fabric administrators	Fabric administrators manage compute resources, which are organized into fabric groups.

Term	Definition
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, such as memory, storage or networks, in a cluster. This is different than 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, 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	A logical label or a pointer to the original reservation. vRealize Automation IaaS determines the reservation from which a particular virtual machine is provisioned. 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. The users of a business group use blueprints 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 2-41. vRealize Automation Integration with vSphere Endpoints

■ Infrastructure Source Endpoints in vRealize Automation

An infrastructure source endpoint is a connection to the infrastructure that provides a set or multiple sets of resources, which IaaS administrators can then be made available to end users. vRealize Automation IaaS regularly collects information about known endpoint resources and the virtual resources provisioned therein. Endpoint resources are referred to as compute resources (or as compute clusters, the terms are often used interchangeably).

■ Virtualization Compute Resources in vRealize Automation

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. Create a compute resources according to the cluster setup in vSphere.

Infrastructure Source Endpoints in vRealize Automation

An infrastructure source endpoint is a connection to the infrastructure that provides a set or multiple sets of resources, which IaaS administrators can then be made available to end users. vRealize Automation IaaS regularly collects information about known endpoint resources and the virtual resources provisioned therein. Endpoint resources are referred to as compute resources (or as compute clusters, the terms are often used interchangeably).

Proxy agents send infrastructure data at regular intervals about the compute resources on each infrastructure endpoint and the machines provisioned on each computer resource. They manage and communicate with the endpoint resources.

Table 2-201. Design Decisions on vRealize Automation Endpoints

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-031	Create two vSphere endpoints.	vSphere endpoints and the Compute vCenter Server instances in each region have one-to-one relationship. You use two endpoints for two regions.	As you add more regions online, you must deploy more vSphere endpoints.
SDDC-CMP-032	Create one vRealize Orchestrator endpoint to connect to the embedded vRealize Orchestrator instance.	vRealize Automation extensibility uses vRealize Orchestrator. The design uses one embedded vRealize Orchestrator cluster which requires the creation of a single endpoint.	You must additionally configure a vRealize Orchestrator endpoint.
SDDC-CMP-033	Create one NSX endpoint and associate it with the vSphere endpoint.	The NSX endpoint is required to connect to the NSX Manager instance and enable all the NSX-related operations supported in vRealize Automation blueprints.	None.

Virtualization Compute Resources in vRealize Automation

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. Create a compute resources according to the cluster setup in vSphere.

vRealize Automation regularly collects information about known compute resources and the virtual machines provisioned on them through the proxy agents.

Table 2-202. Design Decisions on the Compute Resource Configuration in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-034	Create at least one compute resource for each deployed region.	Each region has one shared edge and compute cluster and one compute resource is required for each such cluster.	If you add compute clusters, you must add them to the existing compute resource in their region, or you must use a new resource.

Note By default, compute resources are provisioned to the root of the compute cluster. In this design, the use of vSphere resource pools is mandatory.

Fabric Groups in vRealize Automation

A fabric group is a logical container of several compute resources, and can be managed by fabric administrators. Plan fabric groups according to the number of regions in your SDDC.

Table 2-203. Design Decisions on Fabric Groups in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-035	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.	If you add compute clusters to a region, you must add them to the fabric group.

Business Groups in vRealize Automation

A business group is a collection of machine consumers, often corresponding to a line of business, department, or other 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 in vRealize Automation

A reservation is a share of available memory, CPU and storage one compute resource that is 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, or reservations on multiple compute resources, or both. A reservation must include a vSphere resource pool.

Shared Edge and Compute Clusters and Resource Pools

While 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. Use vSphere resource pools to control the allocation of CPU and memory during time of resource contention on the underlying host. To fully use the mechanism of resource pools for provisioning of workloads, all VMs must be deployed on one of the following resource pools.

Table 2-204. Resource pool details

Resource Pool	Object Types
sfo01-w01rp-sddc-edge	NSX Edge components at the data center level. Place user workload in other resource pools.
sfo01-w01rp-user-edge	Statically or dynamically deployed NSX components such as NSX Edge gateways or load balancers which serve specific customer workloads
sfo01-w01rp-user-vm	Statically or dynamically deployed virtual machines such as Windows instances, Linux instances, databases, etc., which contain specific customer workloads

Table 2-205. Design Decisions on Reservations in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-036	Create at least one vRealize Automation reservation for each business group at each region.	In this design, each resource cluster has two reservations: one for production and one for development, so that you can provision both production and development workloads on the cluster.	Because production and development share compute resources, the development business group must be limited to a fixed amount of resources.
SDDC-CMP-037	Create at least one vRealize Automation reservation for edge resources in each region.	NSX can create edge services gateways on demand and place them on the shared edge and compute cluster.	The workload reservation must define the edge reservation in the network settings.
SDDC-CMP-038	Configure all vRealize Automation workloads to use the sfo01-w01rp-user-vm resource pool.	You introduce control over the resources allocated to the tenant application workloads. As a result, using another resource pool you can dedicate more compute resources to the NSX networking components that provide networking to the workloads. Workloads provisioned at the root resource pool level receive more resources than those in child resource pools. In contention situations, virtual machines might receive insufficient resources.	Cloud administrators must ensure that all workload reservations are configured with the right resource pool. You can have 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.

Table 2-205. Design Decisions on Reservations in vRealize Automation (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-039	Configure vRealize Automation reservations for dynamically provisioned NSX Edge components (routed gateway) to use the sfo01-w01rp-user-edge resource pool.	You dedicate compute resources to NSX networking components. You must assign a vSphere resource pool to end-user deployed NSX Edge components. Workloads provisioned at the root resource pool level receives more resources than child resource pools, which might starve those virtual machines in contention situations.	Cloud administrators must verify that all workload reservations are configured with the right resource pool.
SDDC-CMP-040	All vSphere resource pools 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 might result in unintended under- or over-allocation of resources during contention situations.	None.

Reservation Policies in vRealize Automation

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 2-206. Design Decisions on Reservation Policies in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-041	Create at least one workload reservation policy for each region.	Places a deployment on a specific set of reservations in each region. You also use reservation policies to place workloads into the right region, compute cluster and vSphere resource pool.	None.
SDDC-CMP-042	Create at least one reservation policy for placement of dynamically created edge services gateways into the shared edge and compute clusters.	Places the edge devices into the allocated shared edge and compute cluster 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.

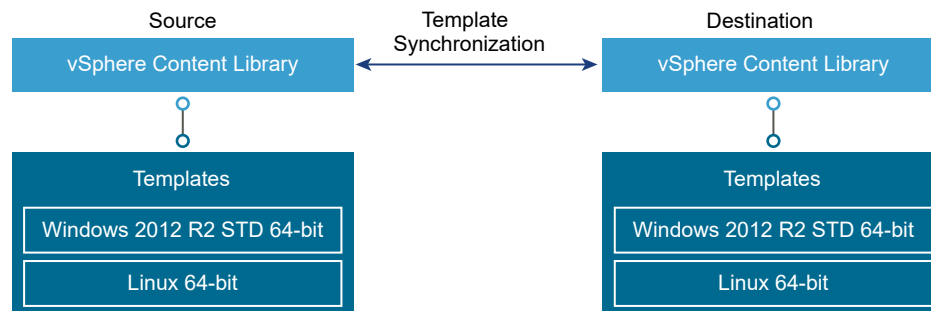Table 2-207. Design Decisions on Storage Reservation Policy in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-043	In this design, do not use storage tiers.	The underlying physical storage design does not use storage tiers.	■ Both business groups, Production and Development, have access to the same storage. ■ Tenants using multiple datastores with different storage capabilities must evaluate the usage of vRealize Automation storage reservation policies.

Template Synchronization in vRealize Automation

This dual-region design supports provisioning 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, by using vSphere Content Library or external services like vCloud Connector or vSphere Replication.

Table 2-208. Design Decision on Template Synchronization in vRealize Automation

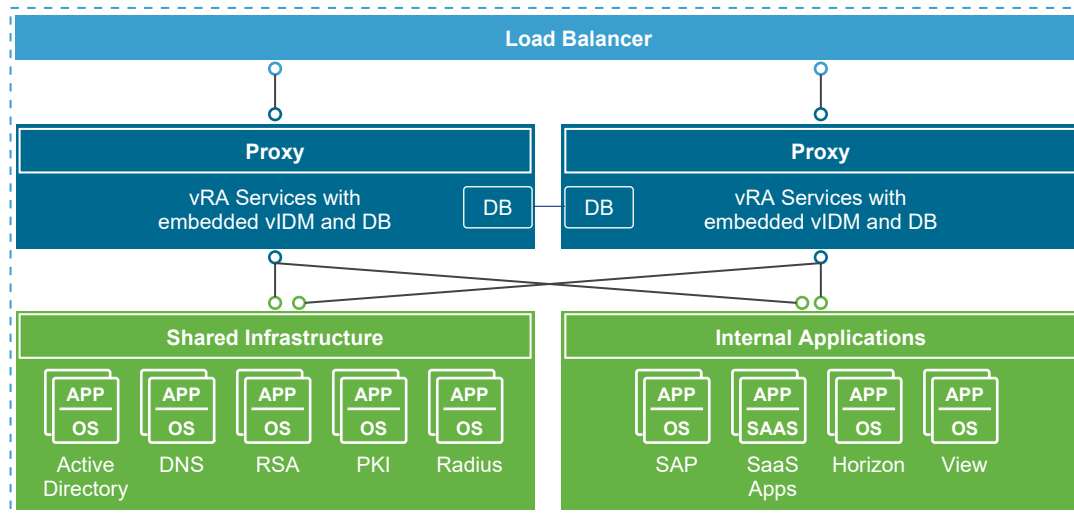| Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-044	This design uses vSphere Content Library to synchronize templates across regions.	The vSphere Content Library is built in to vSphere and meets all the requirements to synchronize templates.	■ You must provision storage space in each region. ■ vRealize Automation cannot directly consume templates from vSphere Content Library.

Figure 2-42. Template Synchronization

VMware Identity Management in vRealize Automation

VMware Identity Manager is integrated in 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 2-43. VMware Identity Manager Proxies Authentication Between Active Directory and vRealize Automation

Table 2-209. Design Decisions on Active Directory Authentication for Tenants in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-045	Use 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 multi-domain or multi-forest Active Directory environment.	The vRealize Automation appliances must be joined to the Active Directory domain.

By default, the vRealize Automation appliance is configured with 18 GB 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 (OU) have to be synchronized. An Active Directory environment with more than 25,000 users is considered large and needs additional memory and CPU. For more information on sizing your vRealize Automation deployment, see the vRealize Automation documentation.

The connector is a component of the vRealize Automation service and performs the synchronization of 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 2-210. Design Decisions on Connector Configuration in vRealize Automation

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-046	To support Directories Service high availability, configure second and third connectors that correspond to the second and third vRealize Automation appliances.	This design supports high availability by installing three vRealize Automation appliances and using load-balanced NSX Edge instances. Adding the second and third connectors to the first vRealize Automation appliance provides redundancy and improves performance by load balancing authentication requests.	You must use an implementation in NSX load balancing.

vRealize Business Design

vRealize Business for Cloud provides end-user transparency in the costs that are associated with operating workloads. vRealize Business shows provisioning costs both during a workload request and on a periodic basis, 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 request and on an ongoing basis with cost reporting by user, business group or tenant. Additionally, tenant administrators can create a wide range of custom reports according to the requirements of an organization. See Logical Design of vRealize Automation and Physical Design of vRealize Automation .

Table 2-211. Resource Requirements for vRealize Business per Virtual Machine

Attribute	Specification
Number of vCPUs	4
Memory	■ 8 GB for a server ■ 2 GB for a data collector
vRealize Business function	Server or data collector

Table 2-212. Design Decision on vRealize Business

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-047	Deploy vRealize Business as part of the Cloud Management Platform and integrate it with vRealize Automation.	You introduce tenant and workload costings.	You must deploy more appliances: for the vRealize Business server and for data collectors.
SDDC-CMP-048	Use the default vRealize Business appliance size of 8 GB. For the vRealize Business data collectors, select a reduced memory size of 2 GB.	The default appliance size of vRealize Business supports the design objective of 10,000 VMs. Data collectors do not run server service, and can run on 2 GB of RAM.	None.

Table 2-212. Design Decision on vRealize Business (continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-049	Use the default vRealize Business reference costing database.	The default reference costing is based on industry information and is periodically updated.	Default reference costing might not accurately represent actual customer costs. The vRealize Business server requires Internet access to periodically update the reference database.
SDDC-CMP-050	Deploy vRealize Business as a three-VM architecture with remote data collectors in Region A and Region B.	For best performance, the vRealize Business collectors must be regionally local to the resource which they are configured to collect data from. Because this design supports disaster recovery, the data collectors can reside in Region A and Region B.	If the environment does not have disaster recovery support, you must deploy an additional appliance for the remote data collector, although the vRealize Business server can handle the load on its own.
SDDC-CMP-051	Deploy the vRealize Business server VM in the cross-region logical network.	The vRealize Business deployment depends on vRealize Automation. During a disaster recovery event, vRealize Business migrates with vRealize Automation.	None.
SDDC-CMP-052	Deploy a vRealize Business remote data collector VM 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 extra Layer 3 hop through an NSX Edge device.

vRealize Orchestrator Design

VMware vRealize Orchestrator is a development and process automation platform that provides a library of extensible workflows to allow 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 VMware Validated Design, vRealize Automation uses the vRealize Orchestrator plug-in to connect to vCenter Server for customized virtual machine provisioning and post-provisioning actions.

▪ Physical Design of vRealize Orchestrator

This VMware Validated Design uses the vRealize Orchestrator instance that is embedded in the vRealize Automation Appliance, instead of using a dedicated or external vRealize Orchestrator instance. Using embedded vRealize Orchestrator simplifies the deployment model and improves the operational efficiency.

▪ Configuration of vRealize Orchestrator

The vRealize Orchestrator configuration includes guidance on client configuration, database configuration, SSL certificates, and plug-ins.

- Scalability of vRealize Orchestrator

  vRealize Orchestrator supports both scale-up and scale-out scalability.

Physical Design of vRealize Orchestrator

This VMware Validated Design uses the vRealize Orchestrator instance that is embedded in the vRealize Automation Appliance, instead of using a dedicated or external vRealize Orchestrator instance. Using embedded vRealize Orchestrator simplifies the deployment model and improves the operational efficiency.

Table 2-213. Design Decisions on vRealize Orchestrator

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-001	Use the internal vRealize Orchestrator instances that are embedded in the deployed vRealize Automation instances.	The use of embedded vRealize Orchestrator has the following advantages: ■ Provides faster time to value. ■ Reduces the number of appliances to manage. ■ Provides easier upgrade path and better support-ability. ■ Improves performance. ■ Removes the need for an external database. ■ Overall simplification of the design leading to a reduced number of appliances and enhanced support-ability. **Note** Using the embedded instance of vRealize Orchestrator is applicable in most use cases. However, refer to the product documentation to be aware of the cases where using the external vRealize Orchestrator is applicable.	None.

- Authentication to vRealize Orchestrator

  The embedded vRealize Orchestrator supports only vRealize Automation authentication method.

- Network Ports for vRealize Orchestrator

  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.

- Server Mode of vRealize Orchestrator

  vRealize Orchestrator supports standalone mode and cluster mode. In this design, vRealize Orchestrator is automatically in cluster mode because the environment contains three vRealize Automation appliances, each running an embedded vRealize Orchestrator.

- Load Balancer Configuration for vRealize Orchestrator

  Configure load balancing for the vRealize Orchestrator instances embedded within the two vRealize Automation instances to provision network access to the vRealize Orchestrator control center.

- Information Security and Access Control in vRealize Orchestrator

  You use a service account for authentication and authorization of vRealize Orchestrator to vCenter Server for orchestrating and creating virtual objects in the SDDC. You also have the objective to establish secure communication to vCenter Server by using CA-signed certificates.

Authentication to vRealize Orchestrator

The embedded vRealize Orchestrator supports only vRealize Automation authentication method.

Table 2-214. Design Decisions on the Directory Service of vRealize Orchestrator

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-002	Use vRealize Automation authentication with the embedded vRealize Orchestrator.	vRealize Automation authentication is the only authentication method available.	None.
SDDC-CMP-VRO-003	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 permissions.	End users who run vRealize Orchestrator workflows are required to have permissions on the vRealize Orchestrator server. Some plug-ins might not support vRealize Automation authentication.

Network Ports for vRealize Orchestrator

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.

Incoming Communication Ports

Set default network ports and configure your firewall to allow incoming TCP connections. Other ports may be required if you are using custom plug-ins.

Table 2-215. Default Inbound Ports of vRealize Orchestrator

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.
vRealize Appliance Orchestrator Control Center	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 2-216. Default External Communication Ports of vRealize Orchestrator

Port	Number	Protocol	Source	Target	Description
LDAP	389	TCP	vRealize Orchestrator server	LDAP server	Lookup port of your LDAP authentication server.
LDAP using SSL	636	TCP	vRealize Orchestrator server	LDAP server	Lookup port of 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
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.
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.
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 VM is running on.

Server Mode of vRealize Orchestrator

vRealize Orchestrator supports standalone mode and cluster mode. In this design, vRealize Orchestrator is automatically in cluster mode because the environment contains three vRealize Automation appliances, each running an embedded vRealize Orchestrator.

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, and 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 join the vRealize Automation appliances in a cluster, 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 as non-responsive, and one of the inactive instances takes over to resume all workflows from the point at which they are 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.

Load Balancer Configuration for vRealize Orchestrator

Configure load balancing for the vRealize Orchestrator instances embedded within the two vRealize Automation instances to provision network access to the vRealize Orchestrator control center.

Table 2-217. Design Decisions on vRealize Orchestrator Cluster

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-004	Configure the load balancer for the management applications for network access to the control center of the embedded vRealize Orchestrator instance.	The control center supports customization of vRealize Orchestrator, such as changing the tenant configuration and certificates.	None.

Information Security and Access Control in vRealize Orchestrator

You use a service account for authentication and authorization of vRealize Orchestrator to vCenter Server for orchestrating and creating virtual objects in the SDDC. You also have the objective to establish secure communication to vCenter Server by using CA-signed certificates.

Authentication and Authorization

Table 2-218. Design Decisions on Authorization and Authentication Management for vRealize Orchestrator

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-005	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 life cycle outside of the SDDC stack to ensure its availability.
SDDC-CMP-VRO-006	Use local permissions when you create the svc-vro service account in vCenter Server.	Ensures 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 assign the service account local permissions in each vCenter Server so that this vCenter Server is a viable endpoint in vRealize Orchestrator.

Encryption

The vRealize Orchestrator configuration interface uses a secure connection to communicate with vCenter Server, relational database management systems (RDBMS), LDAP, vCenter Single Sign-On, and other servers. You can import the required SSL certificate from a URL or file. You can import the vCenter Server SSL certificate from the SSL Trust Manager tab in the vRealize Orchestrator configuration interface.

Table 2-219. Design Decisions on Using CA-Signed Certificates in vRealize Orchestrator

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-007	Use the vRealize Automation appliance certificate.	Simplifies the configuration of the embedded vRealize Orchestrator instance.	None.

Configuration of vRealize Orchestrator

The 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.

vRealize Orchestrator Database

vRealize Orchestrator requires a database. This design uses the PostgreSQL database embedded within the vRealize Automation appliance.

Table 2-220. Design Decisions on the vRealize Orchestrator Database

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-008	Use the PostgreSQL database that is embedded in the vRealize Automation appliance.	Provides the following advantages: ■ Performance improvement ■ Design simplification	None.

vRealize Orchestrator Plug-Ins

You use plug-ins to use vRealize Orchestrator to access and control external technologies and applications. By exposing an external technology in a vRealize Orchestrator plug-in you can 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 for technologies as the vCenter Server API and email capabilities.

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 code 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 and running workflows on the objects in your vSphere inventory, you must configure the vCenter Server plug-in and connect vRealize Orchestrator to the vCenter Server instances you want to orchestrate.

Table 2-221. Design Decisions on the vCenter Server Plug-In of vRealize Orchestrator

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-009	Configure the vCenter Server plug-in to control communication with the vCenter Server instances.	Required for communication to vCenter Server instances, and as such required for workflows.	None.

Multi-Tenancy in vRealize Orchestrator

vRealize Orchestrator introduces a multi-tenant architecture where several vRealize Automation tenants can share a single external or embedded vRealize Automation instance. The Orchestrator multi-tenancy feature provides isolation between tenants.

After you enable multi-tenancy, the objects that the Orchestrator instance manages are assigned a system scope or a tenant-specific scope. The tenant-specific objects in Orchestrator are isolated between the vRealize Automation tenants and from the system-scope objects. Tenant users access their tenant-specific content in the Orchestrator client.

For backwards compatibility and simplified user experience with the product, the multi-tenancy feature in vRealize Orchestrator is disabled by default. Enabling multi-tenancy is an irreversible change.

Table 2-222. Design Decision on Multi-Tenancy in vRealize Orchestrator

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-010	If required by your organization, enable the multi-tenancy feature of vRealize Orchestrator .	Multiple vRealize Automation tenants can share the embedded vRealize Orchestrator cluster. To provide isolation between tenant objects, the objects that Orchestrator manages are assigned a system scope or a tenant-specific scope.	Enabling the multi-tenancy feature of vRealize Orchestrator is irreversible.

Scalability of vRealize Orchestrator

vRealize Orchestrator supports both scale-up and scale-out scalability.

Scale Up

A vRealize Orchestrator instance supports up to 300 concurrent workflow instances in the running state. Workflow instances that are in the waiting or waiting-event states do not count toward that number. You can design long running workflows that preserve resources by using the wait elements of the workflow palette.

A vRealize Orchestrator instance supports up to 35,000 managed virtual machines in its inventory. You can increase the memory and vCPU of 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 using a cluster of vRealize appliances that have the embedded vRealize Orchestrator appropriately configured using the same settings. By using a vRealize Orchestrator cluster, you can 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. Use 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.
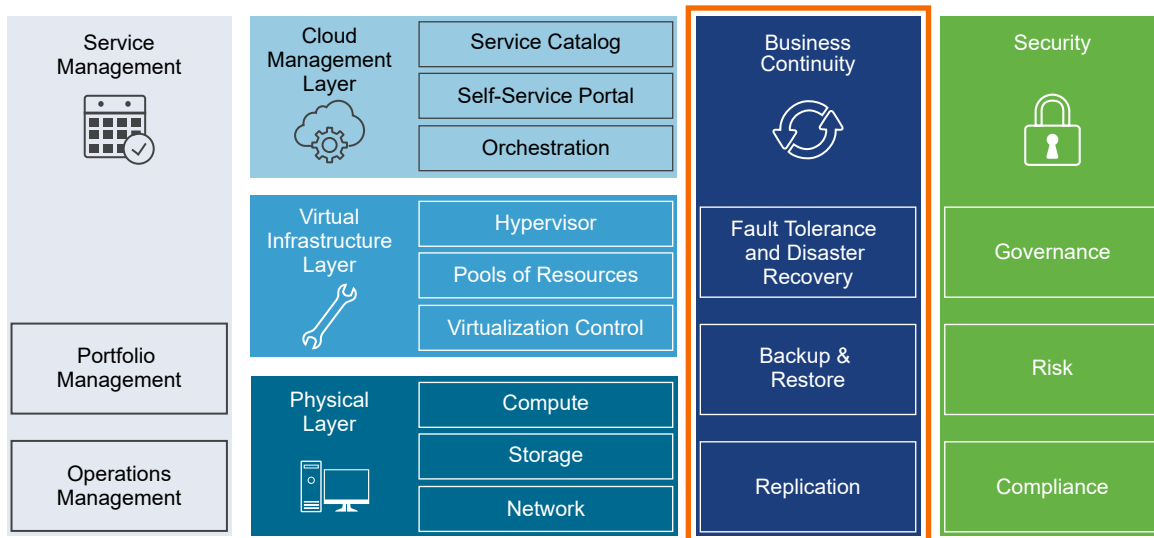
Table 2-223. Design Decisions on vRealize Orchestrator Scale-Out

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-011	Configure vRealize Orchestrator in an active-active cluster configuration.	vRealize Orchestrator servers equally balance workflow execution.	When you organize the vRealize Automation appliances in a cluster, the vRealize Orchestrator instances embedded in them are automatically clustered.

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 2-44. Business Continuity in the SDDC Layered Architecture

Data Protection and Backup Design

Design data protection of the management components in your environment for continuous operation of the SDDC if the data of a management application is compromised.

Backup protects the data of your organization against data loss, hardware failure, accidental deletion, or other fault for each region.

For consistent image-level backups, use backup software that is based on thevSphere Storage APIs - Data Protection (VADP). You can use any VADP-compatible backup solution. Adapt and apply the design decisions to the backup software you use.
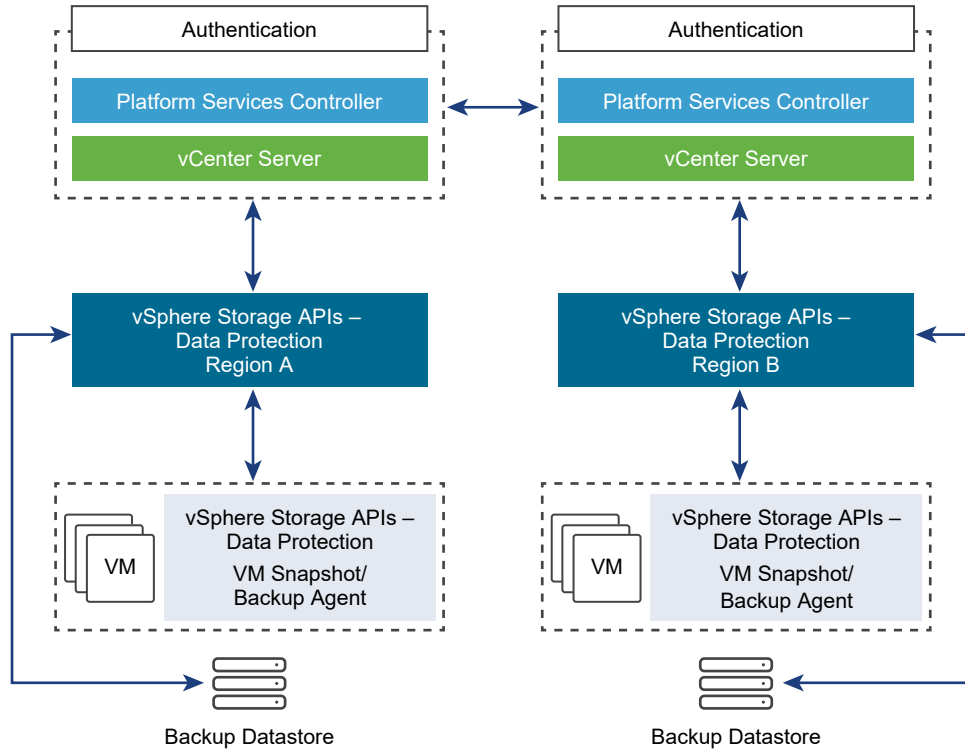
Table 2-224. Design Decisions on VADP-Compatible Backup Solution

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-001	Use a backup solution that is compatible with vSphere Storage APIs - Data Protection (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 vCenter Server level. 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 2-45. Data Protection Logical Design

Backup Datastore for Data Protection

The backup datastore stores all the data that is required to recover services according to a Recovery Point Objective (RPO). Determine the target location. It must meet 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 at least 12 TB of capacity without considering deduplication capabilities.

Table 2-225. Design Decisions on the Backup Datastore

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 the NFS Physical Design .	■ Emergency restore operations are possible even when the primary VMware vSAN datastore is not available because the VADP-compatible backup solution storage volume is separate from the primary vSAN datastore. ■ The amount of storage required for backups is greater than the amount of storage available in the vSAN datastore.	You must provide additional capacity using a storage array.
SDDC-OPS-BKP-004	Provide secondary storage with a capacity of at least 12 TB on-disk.	Secondary storage handles the backup of the management stack of a single region. The management stack consumes approximately 12 TB 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 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 multi-writer 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 2-226. Design Decisions on Virtual Machine Transport Mode

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 impact the vSphere management network.	All ESXi hosts must be connected to 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 the Changed Block Tracking technology to optimize the backup data, to avoid any business impact, do not use a backup window when the production storage is in high demand.

Table 2-227. Design Decisions on Backup Schedule

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 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.	▪ The non-peak time to complete backups might be limited. ▪ Backup duration also depends on storage I/O throughput.

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 2-228. Design Decisions on Backup Retention Policies

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.
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 lost 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 multi-region 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 2-229. Design Decisions on Authorization and Authentication Management for a VADP-Compatible Solution

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-011	Configure a service account in vCenter Server for application-to-application communication from 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, 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 life cycle outside of the SDDC stack to ensure its availability.
SDDC-OPS-BKP-012	Use global permissions when you create the 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 VM component backups as appropriate.

Table 2-230. Design Decision on Component Backup Jobs

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 damaging method to achieve a similar restoration of functionality.	You must provide space on an SFTP or FTP server to store the NSX configuration backups.

Site Recovery Manager and vSphere Replication Design

To support disaster recovery (DR) in the SDDC, you protect vRealize Operations Manager, vRealize Automation, vRealize Business for Cloud, and vRealize Suite Lifecycle Manager by using Site Recovery Manager and vSphere Replication. When failing over to a recovery region, these management applications continue the support of operations management and cloud management platform functionality.

The 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 VMware Validated Design is equivalent to the site construct in Site Recovery Manager.

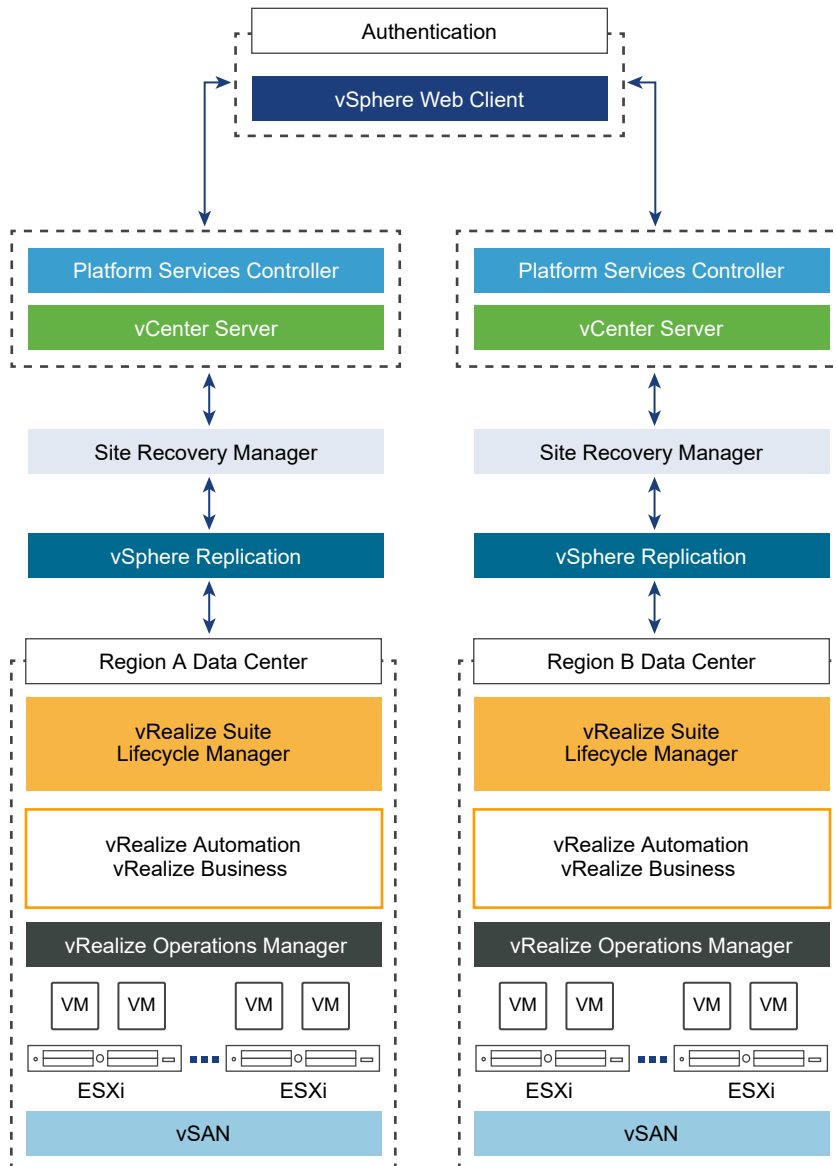## Logical Design for Site Recovery Manager and vSphere Replication

Critical SDDC management applications and services must be available in the event of a 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 following logical configuration of the SDDC management applications:

Table 2-231. 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 A might contain one availability zone or two availability zones by using a stretched vSAN cluster. ■ 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 within the region.
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 2-46. 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 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 2-232. Design Decisions on the Site Recovery Manager and vSphere Replication Deployment

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 2-233. Compute Resources for a Site Recovery Manager Node

Attribute	Specification
Number of vCPUs	2 (running at 2.0 GHz or higher)
Memory	4 GB
Number of virtual machine NIC ports	1
Number of disks	1
Disk size	40 GB
Operating system	Windows Server 2016

Sizing is usually done 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 following management node configuration per region:

Table 2-234. SDDC Nodes with Failover Support

Management Component	Node Type	Number of Nodes
Cloud Management Platform	vRealize Automation Appliance	3
	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 Suite Lifecycle Manager	vRealize Suite Lifecycle Manager Appliance	1
vRealize Operations Manager	vRealize Operations Manager Primary node	1
	vRealize Operations Manager Primary Replica node	1
	vRealize Operations Manager Data	1

You must protect a total of 15 virtual machines.

You use vSphere Replication as the replication solution between the Site Recovery Manager sites, and you distribute the virtual machines in two protection groups.

Table 2-235. Design Decisions on Compute Resources for the Site Recovery Manager Nodes

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 ■ 4 GB memory ■ 40 GB disk ■ 1 VMXNET 3 NIC	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 2-234. SDDC Nodes with Failover Support ■ Two protection groups ■ Two recovery plans	You must increase the size of the nodes if you add more protection groups, virtual machines to protect or recovery 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
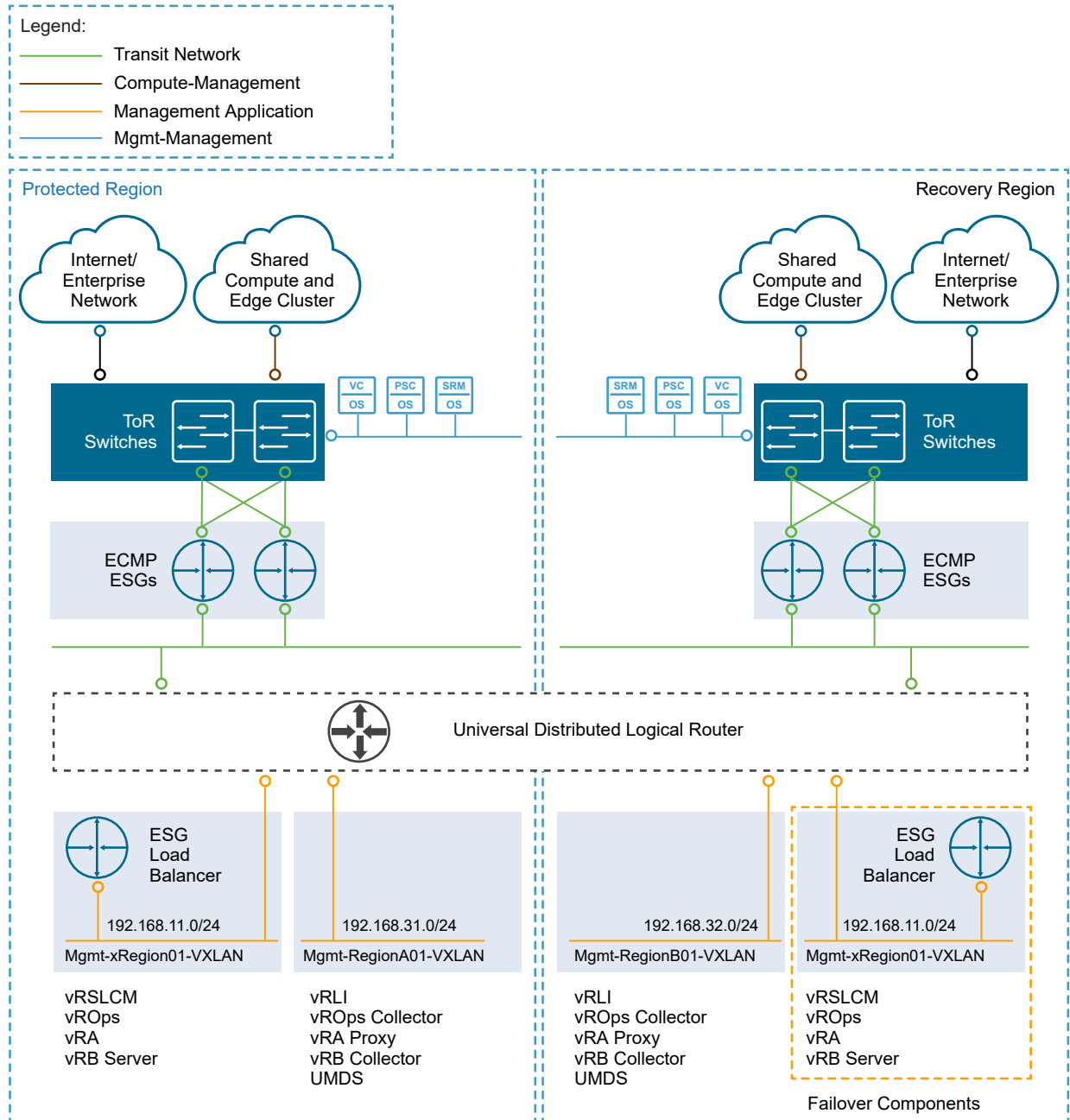
Moving an application physically from one region to another represents a networking challenge, especially if applications have hard-coded IP addresses. According to the requirements for the network address space and IP address assignment, you use either the same or a different IP address at the recovery 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 (AVNs). In AVNs, you can place workloads using a single IP network address space that spans across data centers. AVNs 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 IP address.

Figure 2-47. Logical Network Design for Cross-Region Deployment with Application Virtual Networks

The application virtual networks (orange networks) are routed across the SDDC. As a result, the nodes on these network segments are reachable from within the SDDC. The application virtual network Mgmt-xRegion01-VXLAN that contains the primary vRealize Suite components spans across regions.

NSX Edge devices provide the load balancing functionality. Each device fronts a network that contains the protected components of all management applications. In each region, you use the same configuration for the management applications and the relevant placeholder virtual machines. Active Directory and DNS services must be running in both the protected and recovery regions.

The virtual machines of Site Recovery Manager and vSphere Replication are on the VLAN backed management network in each 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 2-236. Design Decisions on Authorization and Authentication Management for Site Recovery Manager and vSphere Replication

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-005	Configure a service account 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, 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 life cycle outside of the SDDC stack to ensure its availability.
SDDC-OPS-DR-006	Configure a service account 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, 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 life cycle outside of the SDDC stack to ensure its availability.
SDDC-OPS-DR-007	Use global permissions when you create the Site Recovery Manager and vSphere Replication	■ Simplifies and standardizes the deployment of the service account across	All vCenter Server instances must be in the same vSphere domain.

Table 2-236. Design Decisions on Authorization and Authentication Management for Site Recovery Manager and vSphere Replication (continued)

Decision ID	Design Decision	Design Justification	Design Implication
	service accounts in vCenter Server.	all vCenter Server instances in the same vSphere domain. ■ Provides a consistent authorization layer. ■ If you deploy more Site Recovery Manager instances, reduces the efforts in connecting them to the vCenter Server instances.	

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 2-237. Design Decision on CA-Signed Certificates for Site Recovery Manager and vSphere Replication

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-008	Replace the default self-signed certificates in each Site Recovery Manager instance with a CA-signed certificate.	Ensures 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.
SDDC-OPS-DR-009	Replace the default self-signed certificates in each vSphere Replication instance with a CA-signed certificate.	Ensures that all communication 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 using Storage Replication Adapters (SRAs) with Site Recovery Manager

■ vSphere Replication with Site Recovery Manager.

Table 2-238. Design Decisions on Replication Technology

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-010	Use vSphere Replication in Site Recovery Manager as the protection method for virtual machine replication.	■ Allows for flexibility in storage usage and vendor selection between the two disaster recovery regions. ■ Minimizes administrative overhead required to maintain Storage Replication Adapter compatibility between 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. Within the VMware Validated Design, the vSphere Replication traffic has been isolated to its own dedicated port group and VLAN per region ensuring there is no impact other vSphere management traffic. For more information about the vSphere Replication traffic on the management ESXi hosts, see 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 2-239. Design Decisions on the Networking Design of vSphere Replication

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-011	Dedicate a distributed port group to vSphere Replication traffic.	Ensures that vSphere Replication traffic does not impact 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-012	When using two availability zones, dedicate a distributed port group to vSphere Replication traffic per availability zone.	■ VLANs ensure traffic isolation ■ vSphere Replication traffic between availability zones is routed which means an additional stretched VLAN is not required. ■ Limits extraneous networking use across the inter-site link (ISL).	■ Static routes on the ESXi hosts are required. ■ Enough VLANs are available within each cluster that should be used for traffic separation. ■ Host Profiles must be managed on a per-host level.
SDDC-OPS-DR-013	Dedicate a VMkernel adapter on the management ESXi hosts.	Ensures that the ESXi server replication traffic is redirected to the dedicated vSphere Replication VLAN.	None.
SDDC-OPS-DR-014	Attach a virtual machine NIC of the vSphere Replication VMs to the vSphere Replication port group.	Ensures 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.
SDDC-OPS-DR-015	When using two availability zones, deploy an additional virtual machine NIC on the vSphere Replication VM to the vSphere Replication port group for the second availability zone.	■ Ensures that in the event of a disaster in a single availability zone, replication traffic to and from Region B is not interrupted. ■ Because of the unique VLANs and port groups for vSphere Replication traffic per availability zone, a virtual adapter must be connected to each port group to receive data.	During a disaster recovery scenario, you must maintain the IP address of the incoming storage traffic according to the accessibility of the availability zones in Region A. By default, this is Availability Zone 1.

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 expands 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 while 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 2-240. Compute Resources for a vShere Replication Node with 4 vCPUs

Attribute	Specification
Number of vCPUs	4
Memory	4 GB
Disk Capacity	18
Environment	Up to 2000 replications between nodes

Sizing is done according to the 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 15 virtual machines. For information about the node configuration of the management components per region that is used in the calculations, see Table 2-234. SDDC Nodes with Failover Support.

Table 2-241. Design Decisions on vSphere Replication Deployment and Size

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-016	Deploy each vSphere Replication appliance in the management cluster.	All management components must be in the same cluster.	None
SDDC-OPS-DR-017	Deploy each vSphere Replication appliance using the 4 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 ■ vRealize Suite Lifecycle Manager	None.

Messages 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, but support pausing 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, and that the administrator must acknowledge the message before the recovery plan continues. Messages are specific to each IT organization.

Consider the following example messages and confirmation steps:

- Verify that IP address changes are made on the DNS server and that the changes are propagated.

- Verify that the 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 have a dependency 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 on 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 can impact a virtual machine.

If a script must run on the virtual machine, Site Recovery Manager does not run it directly, but 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 call only 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 recovery time objective (RTO) is the targeted duration of time and a 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 2-242. Design Decisions on the Configuration of Protected Management Components

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-018	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 ■ vRealize Suite Lifecycle Manager appliance	■ Provides an automated run book for the recovery of the management components in the event of a disaster. ■ Ensures that the recovery of management applications can be delivered in a recovery time objective (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 such 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 helps improve the reliability of recovering the virtual machine and its applications. However, any solution, including vSphere Replication, that quiesces an operating system and application might impact performance. For example, such an impact 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 recovery point objective (RPO) is configured per virtual machine. 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 has data corruption, data integrity or host OS infections, administrators can recover and revert to a recovery point before the compromising issue occurred.

Table 2-243. Design Decisions on the vSphere Replication Configuration

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-019	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.
SDDC-OPS-DR-020	Enable network compression on the management virtual machine policies in vSphere Replication.	■ Ensures 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, vSphere Replication VM might require more CPU resources on the source site as more virtual machines are protected.
SDDC-OPS-DR-021	Enable a recovery point objective (RPO) of 15 minutes on the management virtual machine policies in vSphere Replication.	■ Ensures 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 VMware Validated Design.	Any changes that are made up to 15 minutes before a disaster recovery event are lost.
SDDC-OPS-DR-022	Enable point-in-time (PIT) instances, keeping 3 copies over a 24-hour period on the management virtual machine policies in vSphere Replication.	Ensures 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 vSAN 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 also set the 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 2-244. Design Decisions on the Startup Order Configuration in Site Recovery Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-023	Use a prioritized startup order for vRealize Operations Manager and vRealize Suite Lifecycle Manager nodes.	- Ensures 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. - Ensures that the vRealize Suite Lifecycle Manager is started in such an order that the lifecycle management services are restored after a disaster. - Ensures that the vRealize Operations Manager and vRealize Suite Lifecycle Manager services are restored in the target of 4 hours.	- You must have VMware Tools running on each vRealize Operations Manager and vRealize Suite Lifecycle Manager node. - You must maintain the customized recovery plan if you increase the number of analytics nodes in the vRealize Operations Manager cluster.
SDDC-OPS-DR-024	Use a prioritized startup order for vRealize Automation and vRealize Business nodes.	- Ensures that the individual nodes within vRealize Automation and vRealize Business are started in such an order that cloud provisioning and cost management services are restored after a disaster. - Ensures 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. - You must maintain the customized recovery plan if you increase the number of nodes in vRealize Automation.

Recovery Plan Test Network

When you create a recovery plan, you must configure test network options as follows:

Isolated Network

Automatically created. 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 non-routable VLAN enabled. In this way, you isolate the network. It is not connected to other production networks.

Because the application virtual networks for failover are fronted by a load balancer, as recovery plan test network you can use the recovery plan production network and provides realistic verification of a recovered management application.

Table 2-245. Design Decisions on the Recovery Plan Test Network

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-025	Use the target recovery production network for testing.	The design of the application virtual networks supports their use as recovery plan test networks. This allows the re-use of existing networks.	During recovery testing, a management application is not reachable using its production FQDN. Access the application 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 mismatch.