Architecture and Design

vmware®

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

https://docs.vmware.com/

If you have comments about this documentation, submit your feedback to

docfeedback@vmware.com

VMware, Inc.
3401 Hillview Ave.
Palo Alto, CA 94304
www.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.

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
19 FEB 2018	The external load balancer node in Figure 2-29 is replaced with the integrated load balancer of vRealize Log Insight.
13 FEB 2018	Initial release.

Architecture Overview

The 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 more agility and predictability. Traditionally this has been referred to as IaaS, or Infrastructure as a Service, however the VMware Validated Design for Software-Defined Data Center extends the typical IaaS solution to include a broader and more complete IT solution.

The VMware Validated Design architecture is based on a number of layers and modules, which allows interchangeable components be part of the end solution or outcome such as the SDDC. If a particular component design does not fit a business or technical requirement for whatever reason, it should be possible for the component to be swapped out for another similar one. VMware Validated Designs are one way of putting an architecture together. They are rigorously tested to ensure stability, scalability and compatibility. Ultimately, the system is designed in such a way as to ensure the desired IT outcome will be achieved.

Figure 1‑1. Architecture Overview
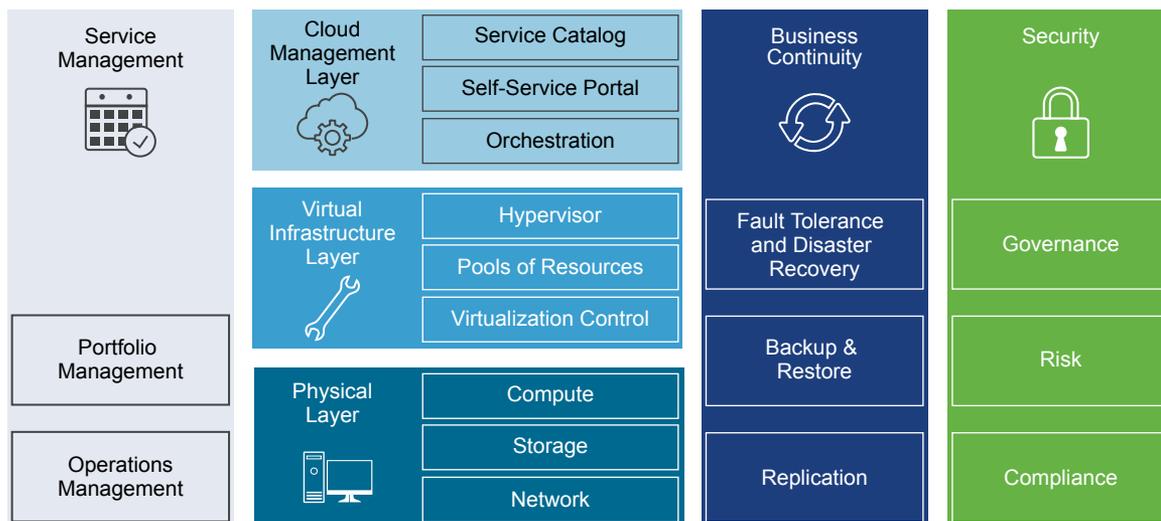
Physical Layer

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

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

Virtual Infrastructure Layer

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

Cloud Management Layer

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

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

Service Management Layer

When building any type of IT infrastructure, portfolio and operations management play key roles in continuous day-to-day service delivery. The Service Management area of this architecture mainly focuses on operations management, in particular 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. For the micro-segmentation use case, you can perform monitoring, logging with vRealize Log Insight.

Within the operations management layer, the underlying physical infrastructure and the virtual management and tenant workloads are monitored in real-time. Information is collected in the form of structured data (metrics) and unstructured data (logs). The operations management layer also knows about the SDDC topology, 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. When data loss occurs, the right elements must be in place to prevent permanent loss to the business. This design provides comprehensive guidance on how to operate backup and restore functions, and includes run books with detailed information on how to fail over components in the event of a disaster.

Security Layer

All systems need to be secure by design. A secure design reduces risk and increases compliance while providing a governance structure. The security layer outlines what is needed to ensure the entire SDDC 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 pods and the physical network topology.

Figure 1-2. Physical Infrastructure Design
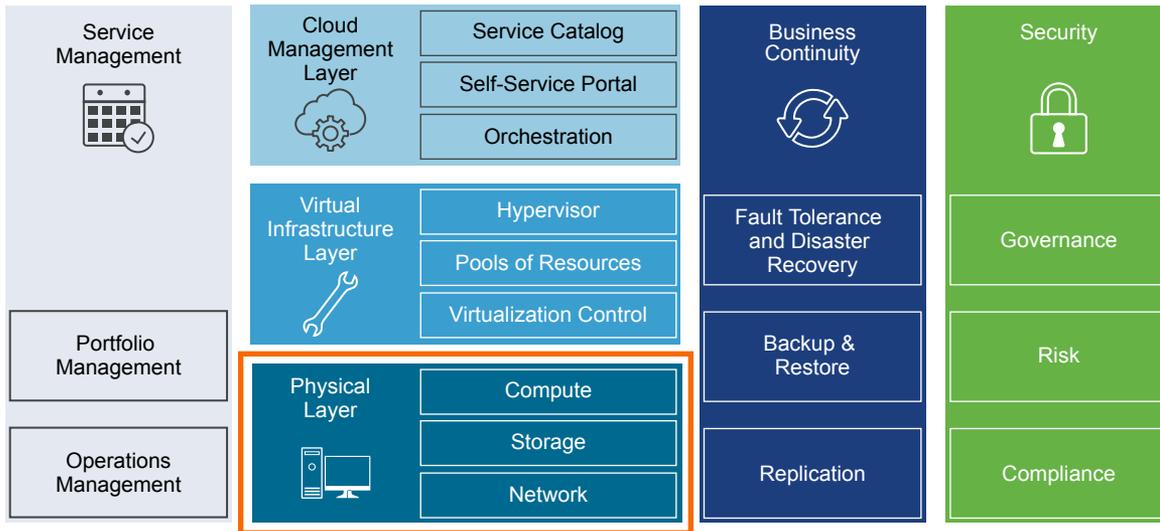
Workload Domain Architecture

VMware Validated Design for Software-Defined Data Center uses a set of common building blocks called workload domains.

Workload Domain Architecture Characteristics

Workload domains can include different combinations of servers, 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. The workload domain is not defined by any hard physical properties. It is a standard unit of connected elements within the SDDC.

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

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

Workload Domain to Rack Mapping

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

Note   When using a Layer 3 network fabric, the management and the shared edge and compute clusters cannot span racks. NSX Controller instances and other virtual machines rely on VLAN-backed networks. The physical network configuration terminates Layer 2 networks in each rack at the Top of 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**
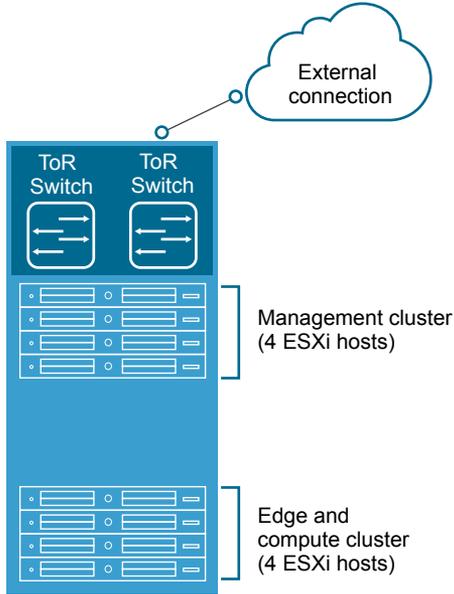
Figure 1-4. Clusters in Two Availability Zones in the SDDC
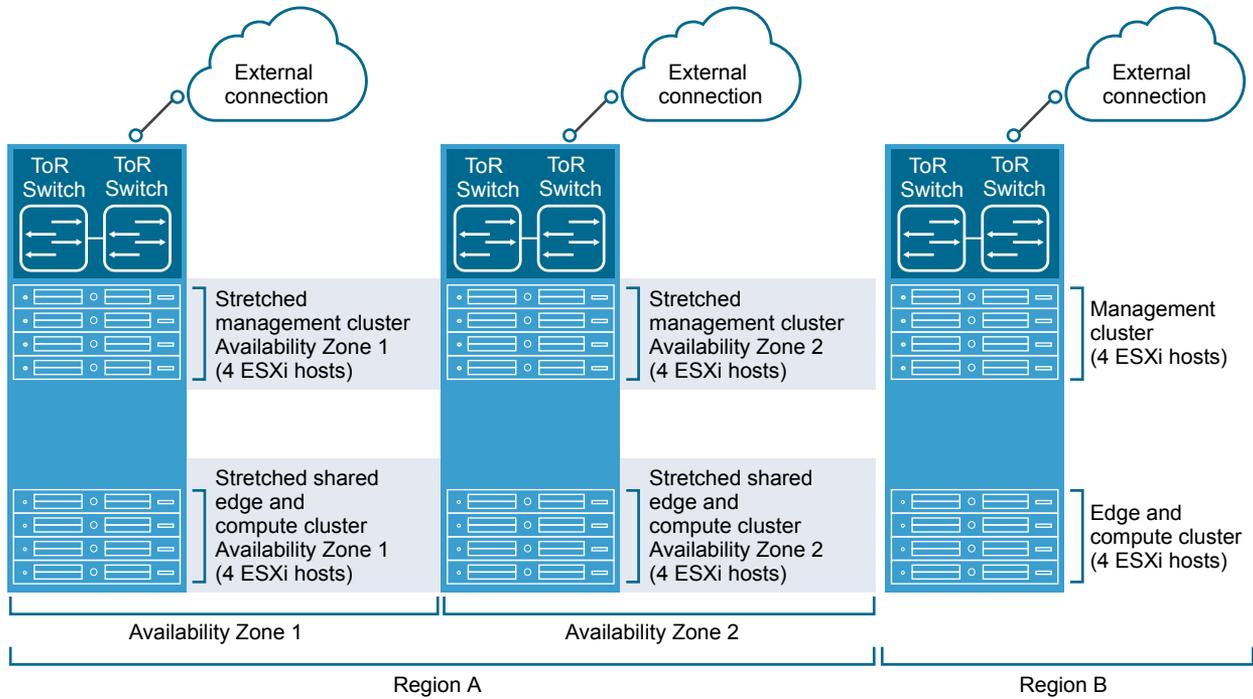
Management Cluster

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.

Network Transport

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

The VMware Validated Designs support both Layer 2 and Layer 3 transports. When deciding whether to use Layer 2 or Layer 3, consider the following :

■ 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 virtual machine traffic.

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

■ The following benefits and drawbacks for both layer 2 and layer 3 designs.

Benefits and Drawbacks for Layer 2 Transport

A design using Layer 2 transport requires 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 terminate 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:

- The benefit of this approach is more design freedom. You can span VLANs, which can be useful in some circumstances.

- The drawback is that the size of such a deployment is limited because the fabric elements have to share a limited number of VLANs. In addition, you may have to rely on a specialized data center switching fabric product from a single vendor.

Figure 1‑5. Example Layer 2 Transport
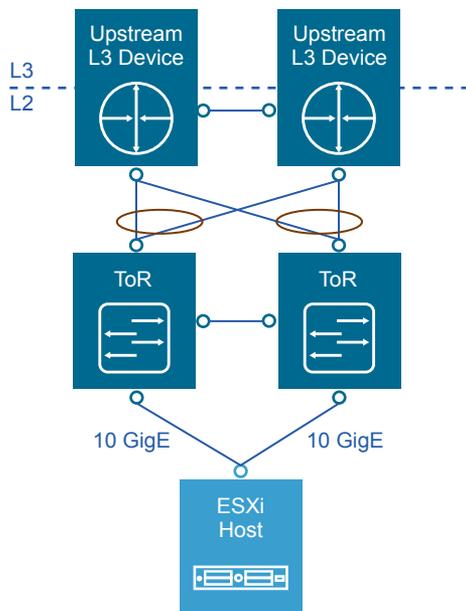
Benefits and Drawbacks for Layer 3 Transport

A design using Layer 3 transport requires these considerations:

- Layer 2 connectivity is limited within 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. VLAN trunking on the uplinks is not allowed.

- 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:

- The benefit is that you can choose from a wide array of Layer 3 capable switch products for the physical switching fabric. You can mix switches from different vendors because of general interoperability between implementation of OSPF, IS-IS or BGP. This approach is typically more cost effective because it makes use of only the basic functionality of the physical switches.

- A design restriction, and thereby a drawback of using Layer 3 routing, is that VLANs are restricted to a single rack. The restriction can affect vSphere Fault Tolerance, and storage networks. To overcame this limitation, use Layer 2 bridging in NSX.

Figure 1-6. Example Layer 3 Transport
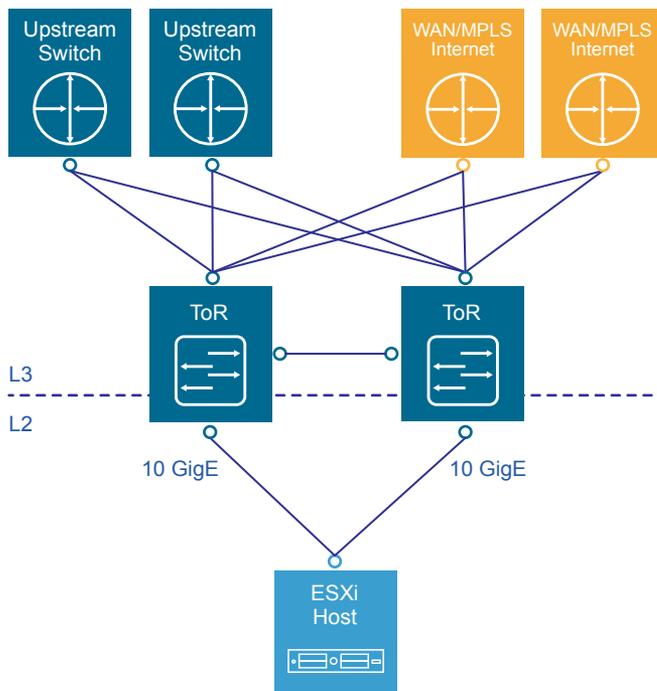
Infrastructure Network Architecture

A key goal of network virtualization is to provide a virtual-to-physical network abstraction.

To achieve this, the physical fabric must provide a robust IP transport with the following characteristics:

- Simplicity

- Scalability

- High bandwidth

- Fault-tolerant transport

- Support for different levels of quality of service (QoS)

Simplicity and Scalability

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

Simplicity

Switch configuration in a data center must be simple. General or global 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 the number of links vary to satisfy different bandwidth demands. You can vary them for each rack.

Quality of Service Differentiation

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

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

- IP storage traffic is typically high in volume and generally stays within a data center.

For virtualized environments, the hypervisor sets the QoS values for the different traffic types. The physical switching infrastructure 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
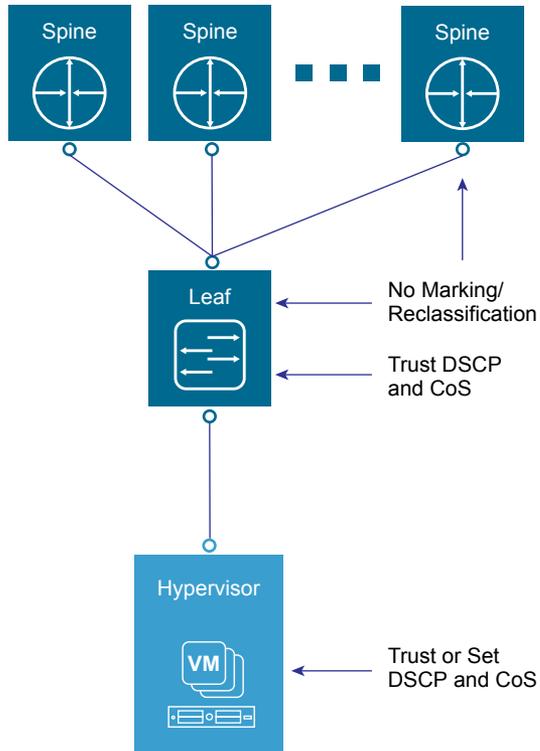
Two types of QoS configuration are supported in the physical switching infrastructure.

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

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

Physical Network Interfaces

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

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

Availability Zones and Regions

In an SDDC, availability zones are collections of infrastructure components. Regions support disaster recovery solutions and allow you to place workloads closer to your customers. Typically, 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 1-1. 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. Common points of failures within a physical data center, like generators and cooling equipment, should not be shared across availability zones. Additionally, these zones should be physically separate so that even uncommon disasters affect only a single availability zone. Availability zones are either two distinct data centers within metro distance or two safety/fire sectors (data halls) within 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) between the zones. This architecture allows the SDDC infrastructure in the availability zone to operate as a single virtual data center within a region.
	You can operate workloads across multiple availability zones in the same region as if they were part of a single virtual data center. This supports an architecture with high availability that is suitable for mission critical applications. When the distance between two locations of equipment becomes too large, these locations can no longer function as two availability zones within the same region and must be treated as separate regions.

Regions

Multiple regions support placing workloads closer to your customers, for example, by operating one region on the US east coast and one region on the US west coast, or operating a region in Europe and another region in the US.

Regions are helpful in several ways:

- Regions can support disaster recovery solutions: One region can be the primary site and another region can be the recovery site.

- You can use multiple regions to address data privacy laws and restrictions in certain countries by keeping tenant data within a region in the same country.

The distance between regions can be rather 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 an operational 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
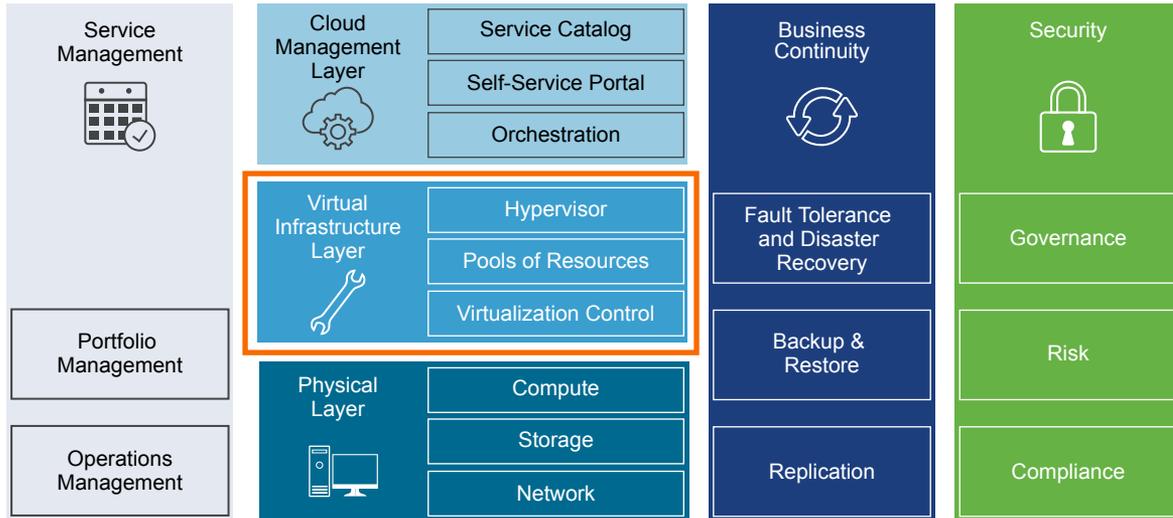
Virtual Infrastructure Overview

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 Single Availability Zone
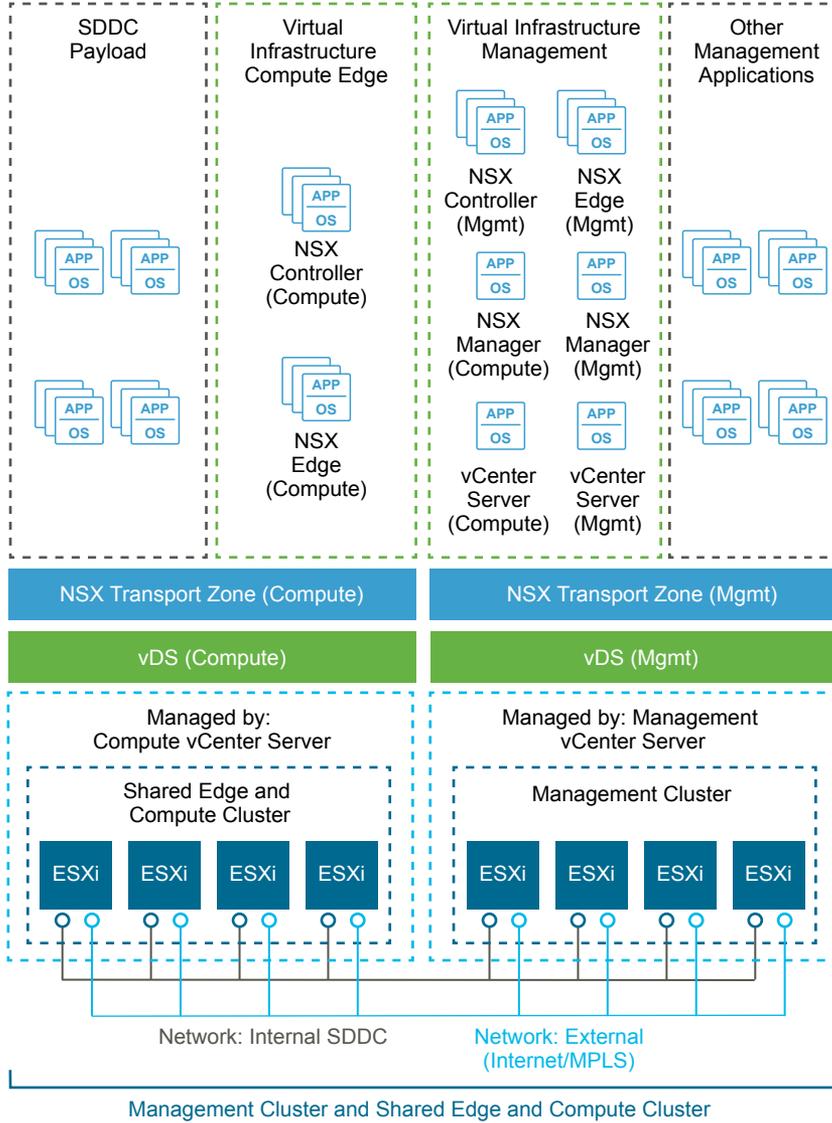
Figure 1-11. Logical Design for Two Availability Zones

Availability Zone 1 - Management and Shared Edge and Compute Clusters
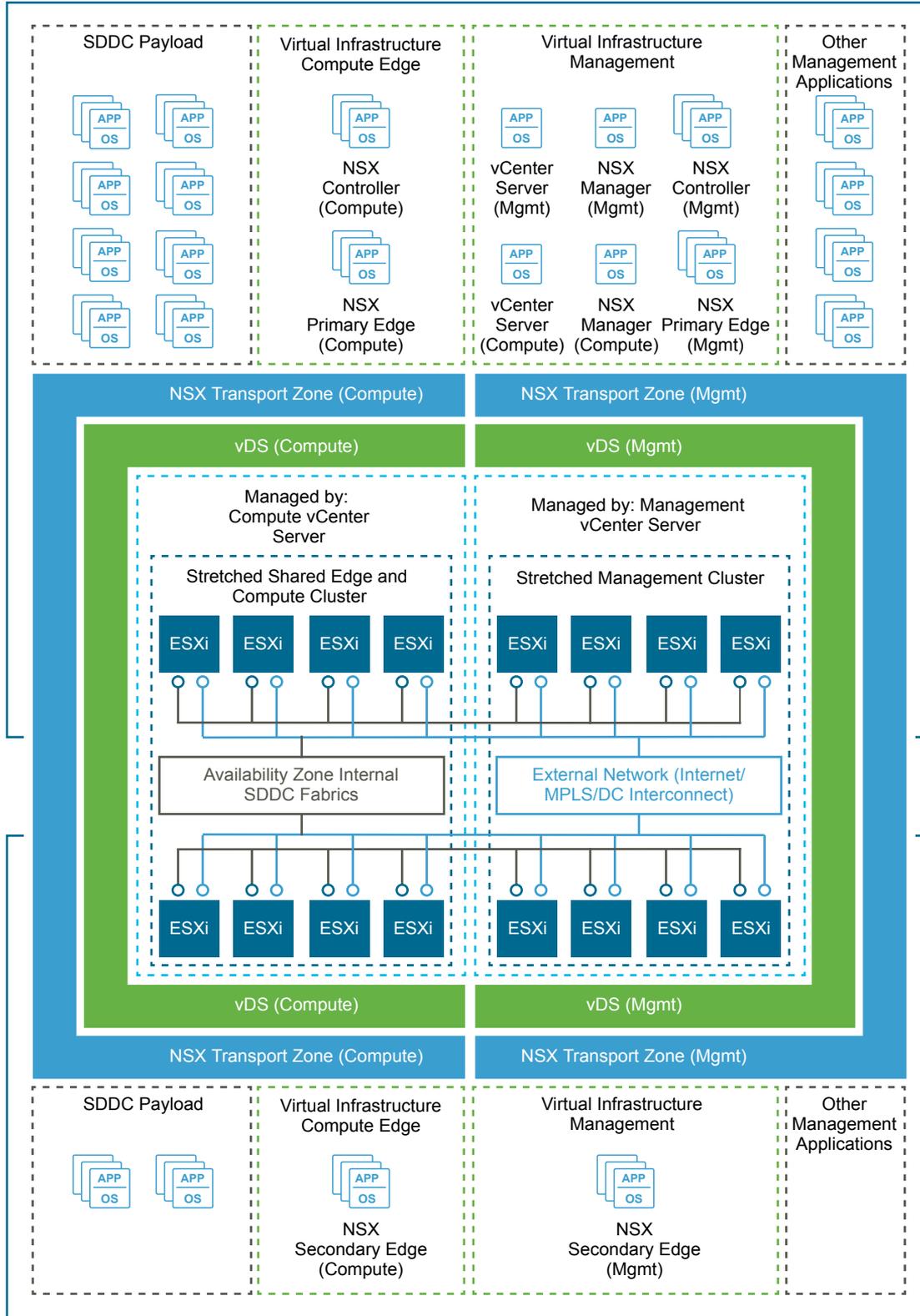
Availability Zone 2 (Secondary) - Management and Shared Edge and Compute Clusters

Management Cluster

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

NSX for vSphere Platform

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

- vCenter Server

- NSX Manager

- NSX Controllers

- NSX Virtual Switch

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

Data plane	Workload data is contained wholly within the data plane. NSX logical switches segregate unrelated workload data. The data is carried over designated transport networks in the physical network. The NSX Virtual Switch, distributed routing, and the distributed firewall are also implemented in the data plane.
Control plane	Network virtualization control messages are located in the control plane. Control plane communication should be carried on secure physical networks (VLANs) that are isolated from the transport networks used for the data plane. Control messages are used to set up networking attributes on NSX Virtual Switch instances, as well as to configure and manage disaster recovery and distributed firewall components on each ESXi host.
Management plane	The network virtualization orchestration happens in the management plane. In this layer, cloud management platforms such as VMware vRealize® Automation™ can request, consume, and destroy networking resources for virtual workloads. Communication is directed from the cloud management platform 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 for vSphere logical switches create logically abstracted segments to which tenant virtual machines can connect. A single logical switch is mapped to a unique VXLAN segment ID and is distributed across the ESXi hypervisors within a transport zone. This allows line-rate switching in the hypervisor without creating constraints of VLAN sprawl or spanning tree issues.

Universal Distributed Logical Router

The NSX for vSphere Universal Distributed Logical Router is optimized for forwarding in the virtualized space (between VMs, on VXLAN- or VLAN-backed port groups). Features include:

- 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.

The Universal Distributed Logical Router is installed in the kernel of every ESXi host, as such it requires a VM to provide the control plane. The universal distributed logical router Control VM 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. NSX for vSphere Universal Distributed Logical Router
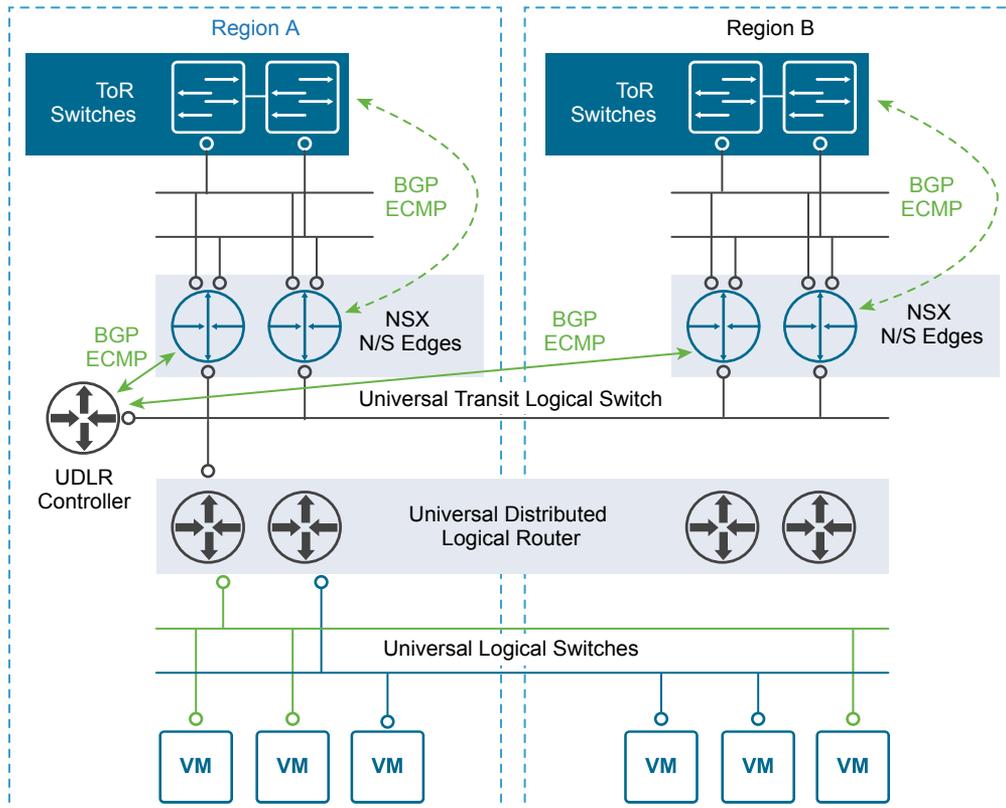
Designated Instance

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 Universal Logical Router provides VM to VM or east-west routing, the NSX Edge services gateway provides north-south connectivity, by peering with upstream Top of Rack switches, thereby enabling tenants to access public networks.

Logical Firewall

NSX for vSphere 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 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 key 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, maximizing throughput, minimizing response time, and avoiding overload. NSX Edge provides load balancing up to Layer 7.

Service Composer

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

NSX for vSphere 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
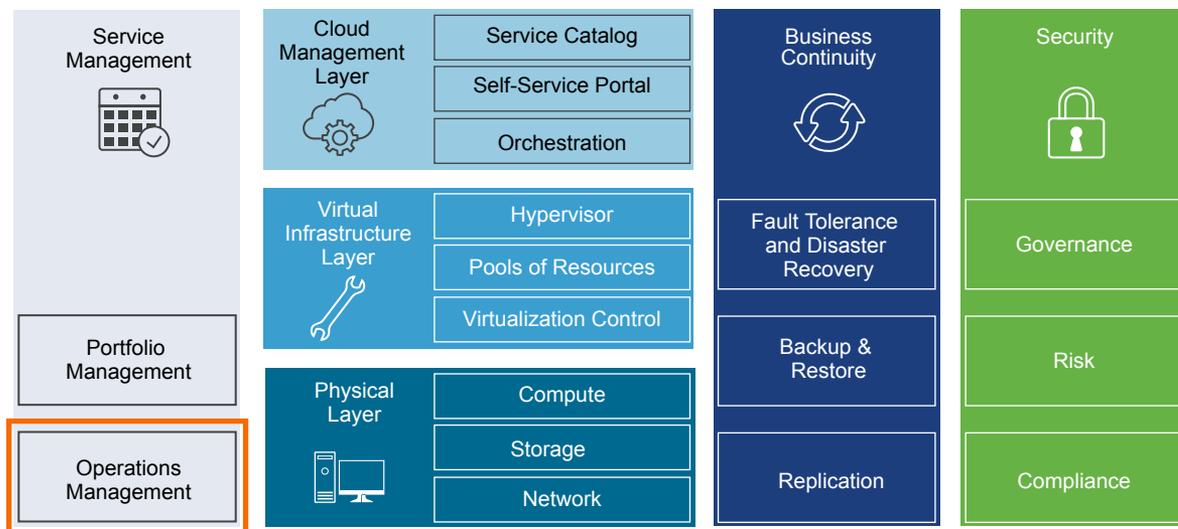
Monitoring Architecture

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 master and at least one data node, and optionally a group of remote collector nodes.

  You can establish high availability by using an external load balancer.

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

Architecture

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

Figure 1-14. vRealize Operations Manager Architecture
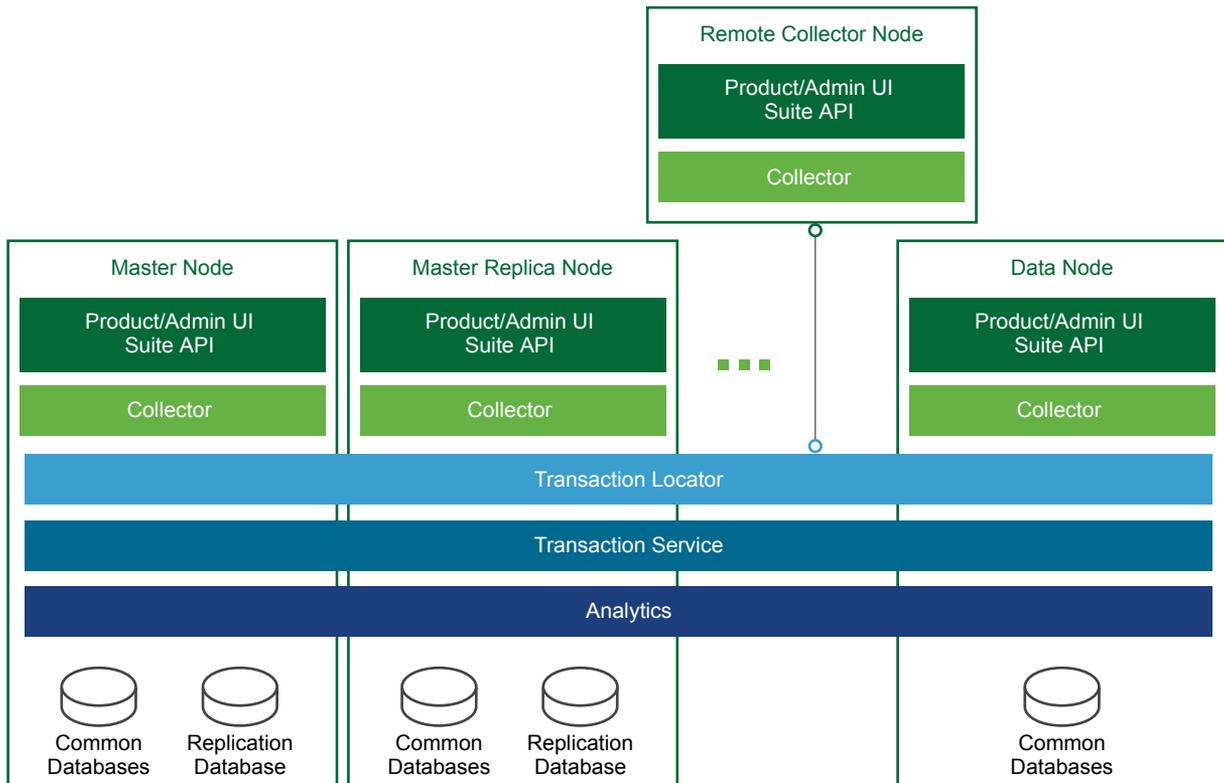
Types of Nodes

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.

Master Node	Required initial node in the cluster. In large-scale environments, manages all other nodes. In small-scale environments, the master node is the single standalone vRealize Operations Manager node.
Master Replica Node	Optional. Enables high availability of the master node.
Data Node	Optional. Enables scale-out of vRealize Operations Manager in larger environments. Data nodes have adapters installed to perform collection and analysis. Data nodes also host vRealize Operations Manager management packs.
Remote Collector Node	Overcomes data collection issues across the enterprise network, such as limited network performance. You can also use remote collector nodes to offload data collection from the other types of nodes.
	Remote collector nodes only gather statistics about inventory objects and forward collected data to the data nodes. Remote collector nodes do not store data or perform analysis.

The master and master replica nodes are data nodes that have extended capabilities.

Types of Node Groups

Analytics Cluster	Tracks, analyzes, and predicts the operation of monitored systems. Consists of a master node, data nodes, and optionally of a master replica node.
Remote Collector Group	Because it consists of remote collector nodes, only collects diagnostics data without storage or analysis. A vRealize Operations Manager deployment can contain several collector groups.
	Use collector groups to achieve adapter resiliency in 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 1‑15. Architecture of a vRealize Operations Manager Node
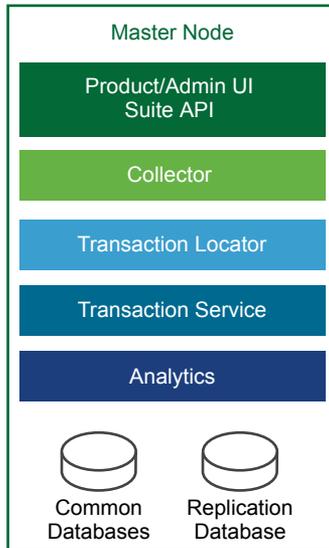
The components of vRealize Operations Manager node perform these tasks.

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 master, master replica, and remote collector nodes.

Transaction Service

The Transaction Service is responsible for caching, processing, and retrieving metrics for the analytics process.

Analytics

The analytics engine creates all associations and correlations between various 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 handle data coming from management solutions

In a multi-availability zone implementation, which is a sub-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 the 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 master and at least two worker nodes. You can establish high availability by using the integrated load balancer (ILB).

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

Architecture

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

Figure 1‑16. Architecture of vRealize Log Insight
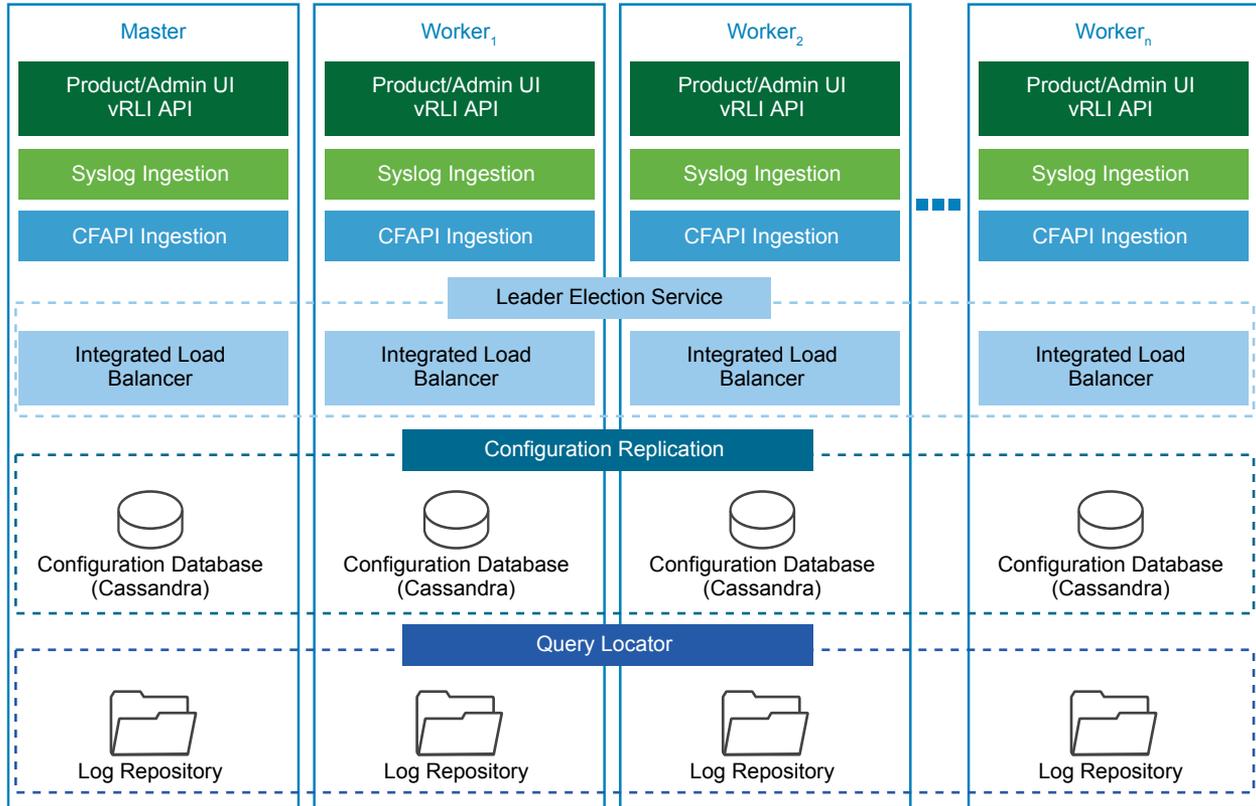
vRealize Log Insight clients connect to the ILB Virtual IP (VIP) address, and use the syslog or the Ingestion API via the vRealize Log Insight agent to send logs to vRealize Log Insight. Users and administrators interact with the ingested logs using the user interface or the API.

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:

Master Node	Required initial node in the cluster. In standalone mode, the master node is responsible for all activities, including queries and log ingestion. The master node also handles operations that are related to the lifecycle of a cluster, such as performing upgrades and addition or removal of worker nodes. In a scaled-out and highly available environment, the master 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 master node stores logs locally. If the master node is down, the logs stored on it become unavailable.
Worker Node	Optional. This component enables scale-out in larger environments. As you add and configure more worker nodes in a vRealize Log Insight cluster for high availability (HA), queries and log ingestion activities are delegated to all available nodes. You must have at least two worker nodes to form a cluster with the master node.
	The worker node stores logs locally. If any of the worker nodes is down, the logs on the worker become unavailable.
Integrated Load Balancer (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 master for the duration of the query. The query master queries all nodes, both master and worker nodes, for data and then sends the aggregated data back to the client.
	Use the ILB for administrative activities unless you are performing administrative activities on individual nodes. The Web user interface of the ILB presents data from the master and from the worker nodes in a scaled-out cluster in a unified display(single pane of glass).

Application Functional Components

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

- Analyze logging data that is ingested from the components of a data center

- Visualize the results in a Web browser, or support results query using API calls.

Figure 1-17. vRealize Log Insight Logical Node Architecture
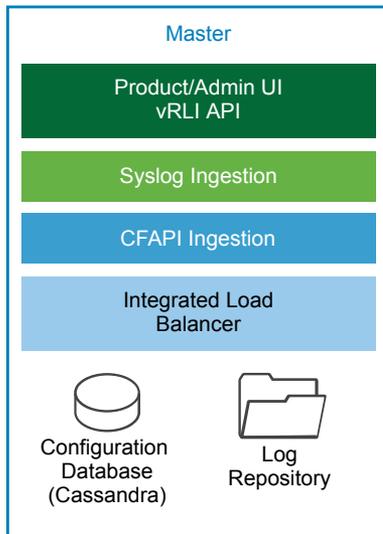
The vRealize Log Insight components perform these tasks:

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 master node. In a scaled-out

vRealize Log Insight cluster, this role can be inherited by any of the available nodes during the election process. The leader periodically performs health checks to determine whether a 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 1-18. Event Forwarding in vRealize Log Insight
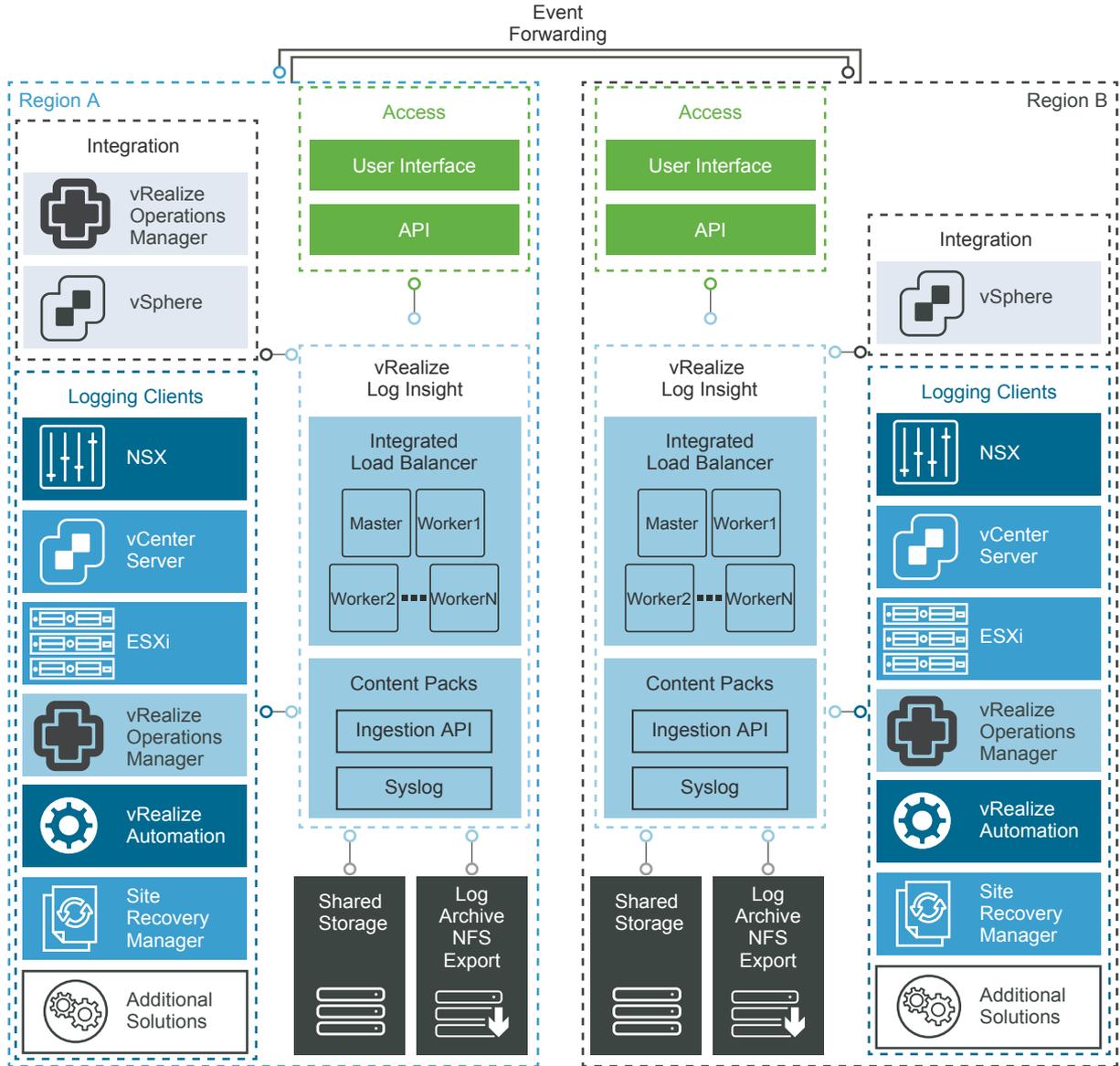
vSphere Update Manager Architecture

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

Overview

vSphere Update Manager registers with a single vCenter Server instance 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 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 1-2. Installation Models of vSphere Update Manager and Update Manager Download Service

Component	Installation Model	Description
vSphere Update Manager	Embedded in the vCenter Server Appliance	vSphere Update Manager is automatically registered with the container vCenter Server Appliance. You access vSphere Update Manager as a plug-in from the vSphere Web Client. Use virtual appliance deployment to easily deploy vCenter Server and vSphere Update Manager as an all-in-one package in which sizing and maintenance for the latter is dictated by the former.
	Windows installable package for installation against a Microsoft Windows vCenter Server	You must run the vSphere Update Manager installation on either vCenter Server itself or an external Microsoft Windows Server. After installation and registration with vCenter Server, you access vSphere Update Manager as a plug-in from the 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 (Host OS) that are detailed in VMware Knowledge Base Article 2091273. You cannot install UMDS on the same system as vSphere Update Manager.

Architecture

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

Figure 1-19. vSphere Update Manager and Update Manager Download Service Architecture
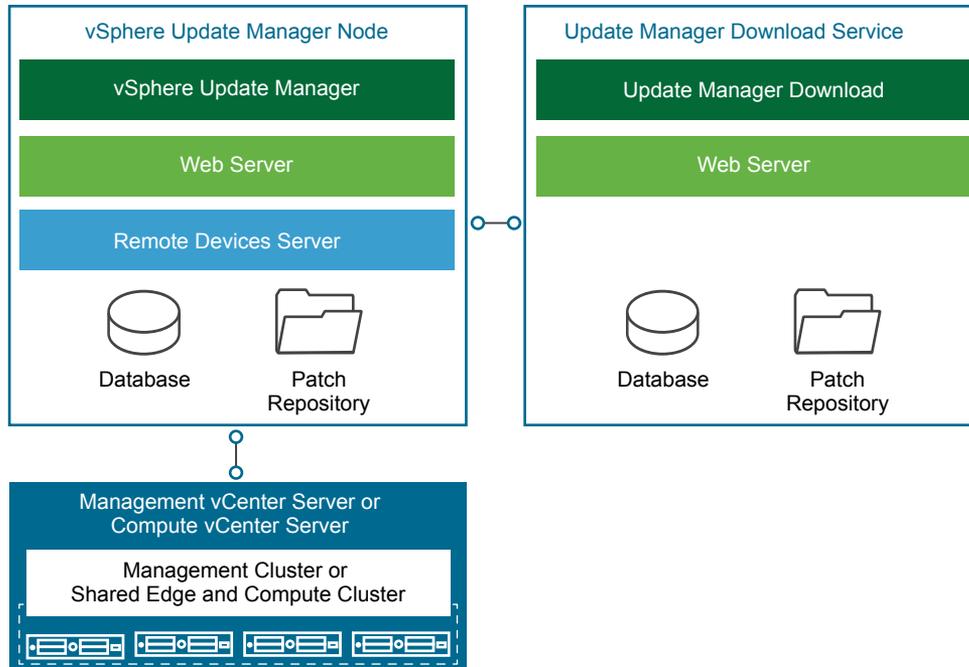
Types of Nodes

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

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

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 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, the 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 can deploy one UMDS instance per region. In this way, you have a central repository of aggregated patch binaries that are securely downloaded.

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

Figure 1-20. Dual-Region Interaction between vSphere Update Manager and Update Manager Download Service
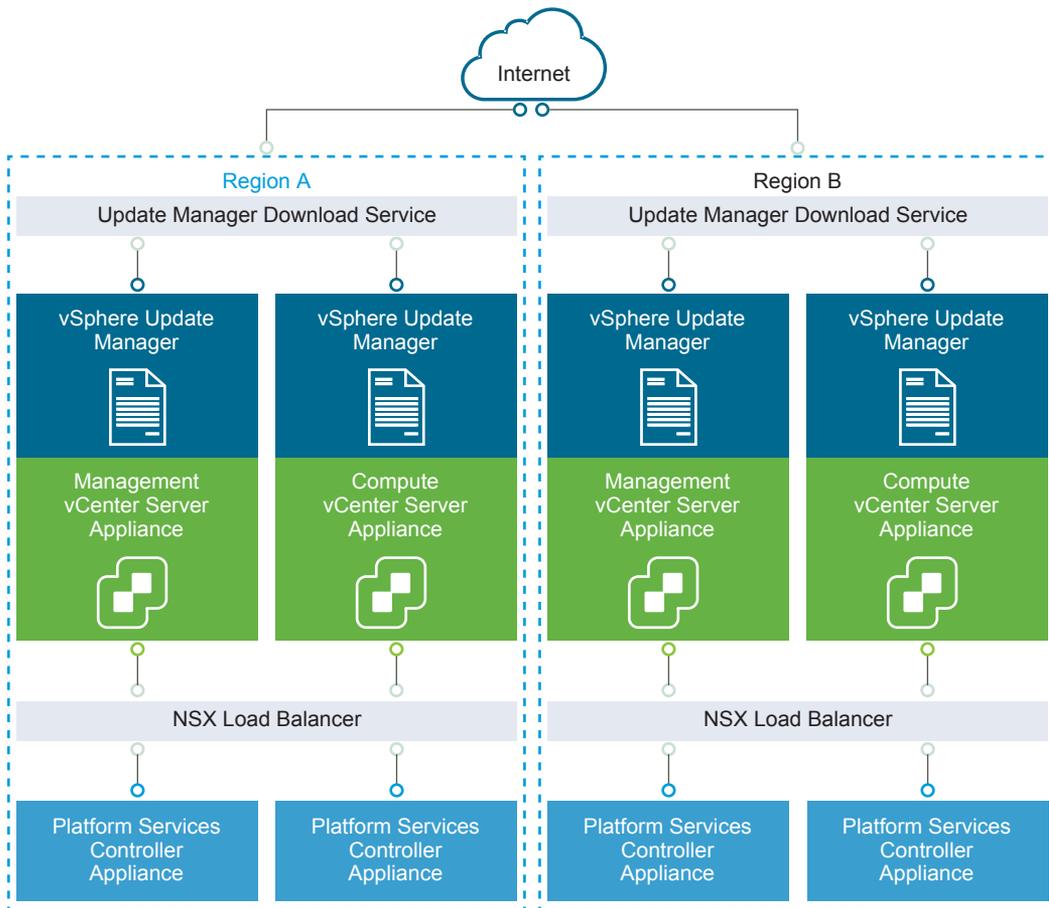
Cloud Management Architecture

The Cloud Management Platform (CMP) is the primary consumption portal for the entire Software-Defined Data Center (SDDC). Within the SDDC, you use vRealize Automation to author, administer, and consume VM templates and blueprints.

Figure 1-21. Cloud Management Platform Layer in the SDDC
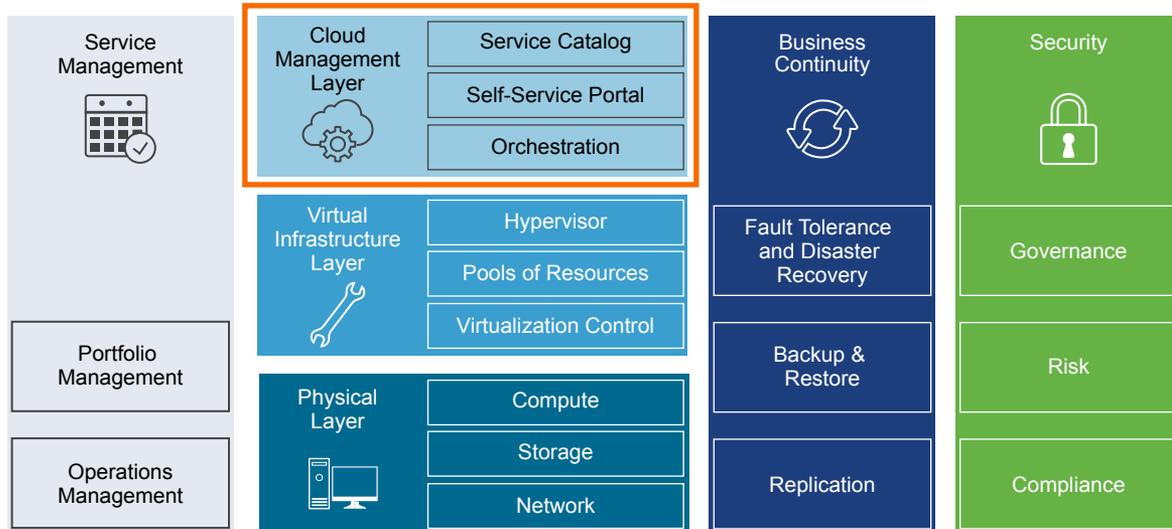
The Cloud Management Platform layer delivers the following multi-platform and multi-vendor cloud services.

- Comprehensive and purpose-built capabilities to provide standardized resources to global customers in a short time span.

- Multi-platform and multi-vendor delivery methods that integrate with existing enterprise management systems.

- Central user-centric and business-aware governance for all physical, virtual, private, and public cloud services.

- Architecture that meets customer and business needs, and is extensible.

vRealize Automation Architecture of the Cloud Management Platform

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, while ensuring compliance with business policies. Requests for IT service, including infrastructure, applications, desktops, and many others, are processed through a common service catalog to provide a consistent user experience.

vRealize Automation Installation Overview

Installing vRealize Automation requires deploying the vRealize Automation appliance, and the vRealize Automation Infrastructure as a Service IaaS components which need to be installed on one more Windows servers. To install, you deploy a vRealize Automation appliance and then complete the bulk of the installation 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 accepts input from an answer properties file.

- An installation REST API that accepts JSON formatted input.

vRealize Automation Architecture

vRealize Automation provides self-service provisioning, IT services delivery and life-cycle management of cloud services across a wide range of multivendor, virtual, physical and cloud platforms though a flexible and robust distributed architecture. The two main functional elements of the architecture are the vRealize automation server and the Infrastructure as a Service Components.

Figure 1-22. vRealize Automation Architecture
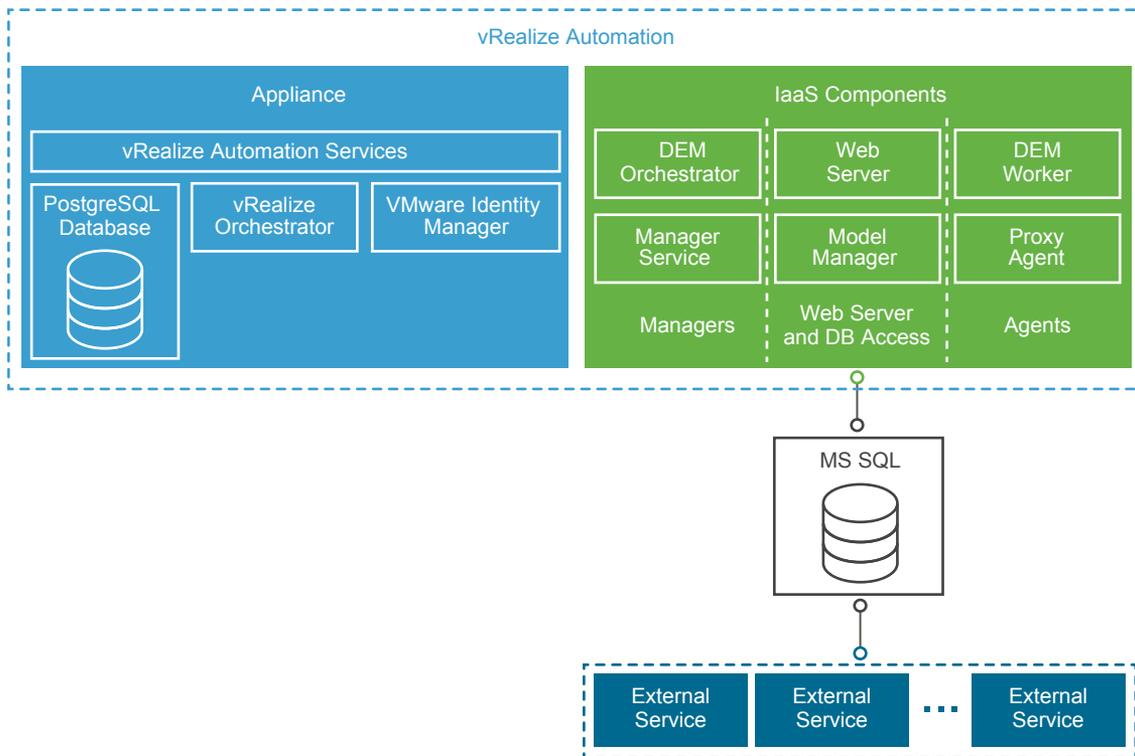
vRealize Automation Server Appliance	The vRealize Automation server is deployed as a preconfigured Linux virtual appliance. The vRealize Automation server appliance is delivered as an open virtualization file (.OVF) that you deploy on existing virtualized infrastructure such as vSphere. It performs the following functions:

- vRealize Automation product portal, where users log to access self-service provisioning and management of cloud services.

- Single sign-on (SSO) for user authorization and authentication.

- Management interface for vRealize Automation appliance settings.

Embedded vRealize Orchestrator	The vRealize Automation appliance contains a preconfigured instance of vRealize Orchestrator. vRealize Automation uses vRealize Orchestrator workflows and actions to extend its capabilities.
PostgreSQL Database	vRealize Server uses a preconfigured PostgreSQL database that is included in the vRealize Automation appliance. This database is also used by the instance of vRealize Orchestrator within the vRealize Automation appliance.
Infrastructure as a Service	vRealize Automation IaaS consists of one or more Microsoft Windows servers that work together to model and provision systems in private, public, or hybrid cloud infrastructures.
Model Manager	vRealize Automation uses models to facilitate integration with external systems and databases. The models implement business logic used by the Distributed Execution Manager (DEM).
	The Model Manager provides services and utilities for persisting, versioning, securing, and distributing model elements. Model Manager is hosted on one of the IaaS web servers and communicates with DEMs, the SQL Server database, and the product interface web site.
IaaS Web Server	The IaaS web server provides infrastructure administration and service authoring to the vRealize Automation product interface. The web server component communicates with the Manager Service, which provides updates from the DEM, SQL Server database, and agents.
Manager Service	Windows service that coordinates communication between IaaS DEMs, the SQL Server database, agents, and SMTP. The Manager Service communicates with the web server through the Model Manager, and must be run under a domain account with administrator privileges on all IaaS Windows servers.
Distributed Execution Manager Orchestrator	A Distributed Execution Manager (DEM) executes 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 vRealize Automation IaaS DEM Worker executes 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, vSphere proxy agent sends commands to and collects data from a vSphere ESX Server for the VMs provisioned by vRealize Automation.

VMware Identity Manager	VMware Identity Manager (vIDM) is the primary identity provider for vRealize Automation, and manages user authentication, roles, permissions, and overall access into vRealize Automation by means of federated identity brokering. The following authentication methods are supported in vRealize Automation using VMware Identity Manager:

- Username/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

VMware Validated Design Deployment Model

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

- 2 vRealize Automation Server Appliances behind a load balancer
- 2 vRealize Automation IaaS Web Servers behind a load balancer
- 2 vRealize Automation Manager Service nodes (including DEM Orchestrator) behind a load balancer
- 2 DEM Worker nodes
- 2 IaaS Proxy Agent nodes.

vRealize Business for Cloud Architecture

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

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

Figure 1‑23. vRealize Business for Cloud
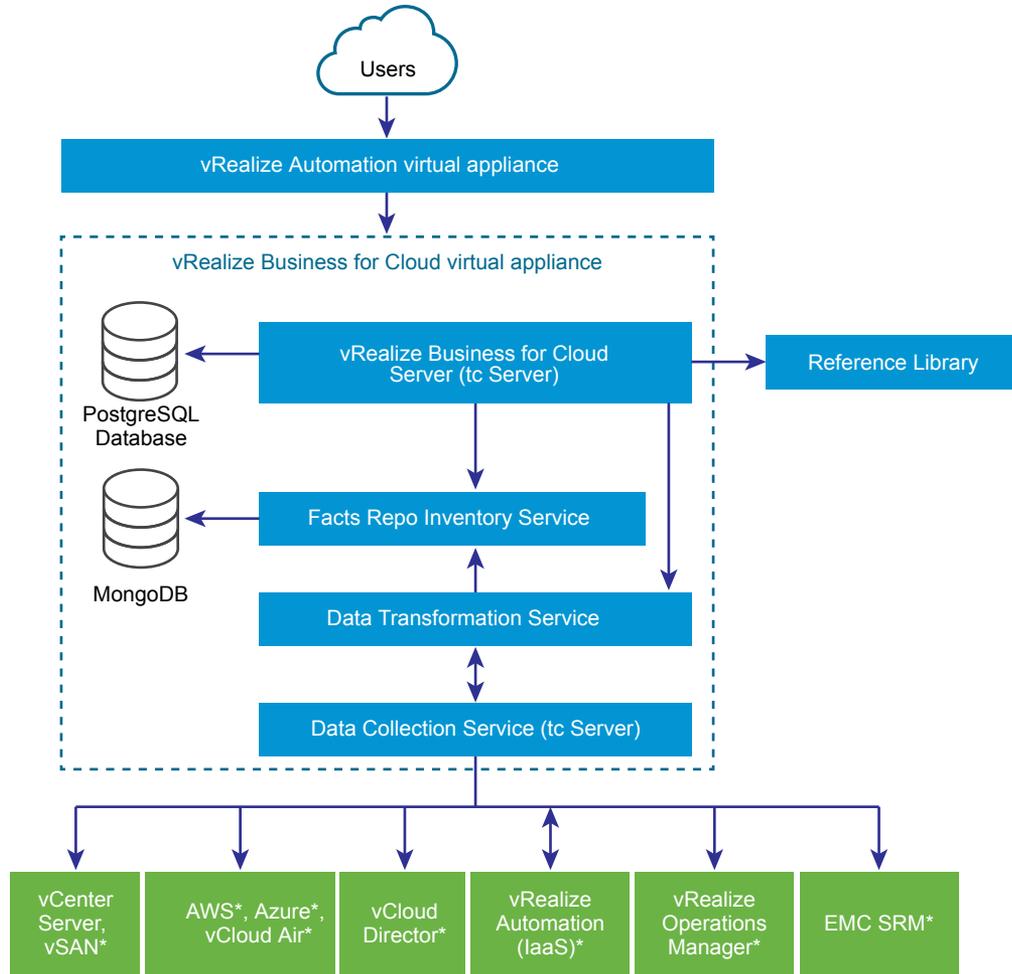
Data Collection Services

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

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

FactsRepo Inventory Service

An inventory service built on MongoDB to store the collected data that vRealize Business for Cloud uses for cost computation.

Data Transformation Service	Converts source specific data from the data collection services into data structures for consumption by the FactsRepo inventory service. The data transformation service serves as is a single point of aggregation of data from all data collectors.
vRealize Business for Cloud Server	A web application that runs on Pivotal tc Server. vRealize Business for Cloud has multiple data collection services that run periodically, collecting inventory information and statistics, which is in turn stored in a PostgreSQL database as the persistent data store. Data collected from the data collection services is used for cost calculations.
Reference Database	Responsible for providing default, out-of-the-box costs for each of the supported cost drivers. The reference database is updated automatically or manually, and you can download the latest data set and import it into vRealize Business for Cloud. 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. For more information, see Update the Reference Database for vRealize Business for Cloud.
Other Sources of Information	These information sources are optional, and are used only if installed and configured. The sources include vRealize Automation, vCloud Director, vRealize Operations Manager, Amazon Web Services (AWS), Microsoft Azure, and vCloud Air, and EMC Storage Resource Manager (SRM).
vRealize Business for Cloud Operational Model	vRealize Business for Cloud continuously collects data from external sources, and periodically updates the FactsRepo inventory service. You can view the collected data using the vRealize Business for Cloud dashboard or generate a report. The data synchronization and updates occur at regular intervals, however, you can manually trigger the data collection process when inventory changes occur. For example, in response to the initialization of the system, or addition of a private, public, or hybrid cloud account.
vRealize Business for Cloud Deployment Model	The scope of this VMware Validated Design uses a deployment model consisting of the following virtual machines: a single vRealize Business for Cloud Server appliance and a single vRealize Business for Cloud remote data collector for Region A, and a single vRealize Business for Cloud remote data collector for Region B.

Business Continuity Architecture

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

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

- Data protection

- Data replication

- Orchestrated disaster recovery

Figure 1-24. Business Continuity Layer of the SDDC
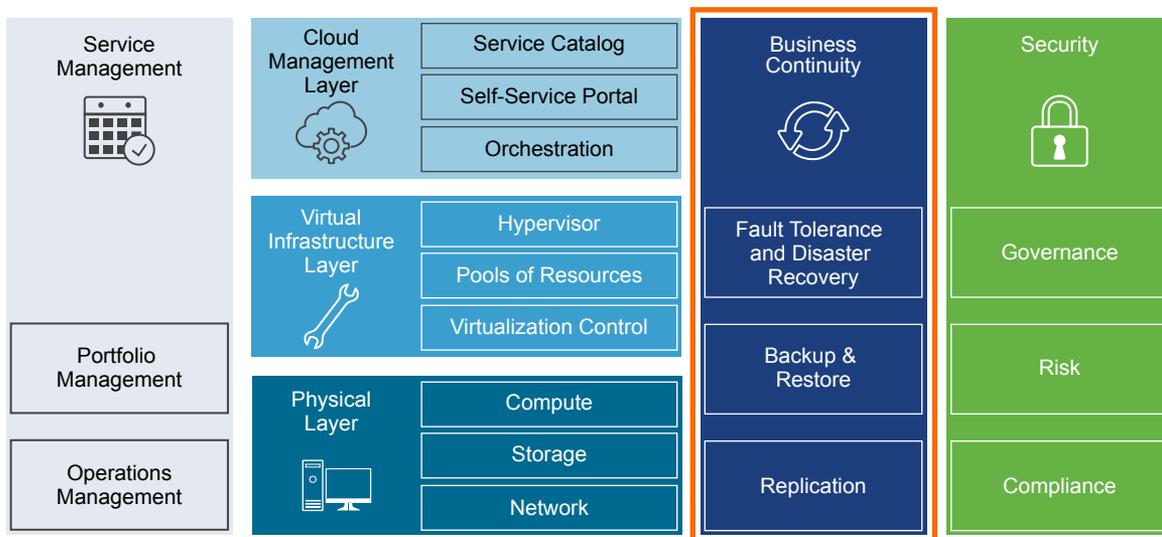
Data Protection and Backup Architecture

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 1-25. Dual-Region Data Protection Architecture
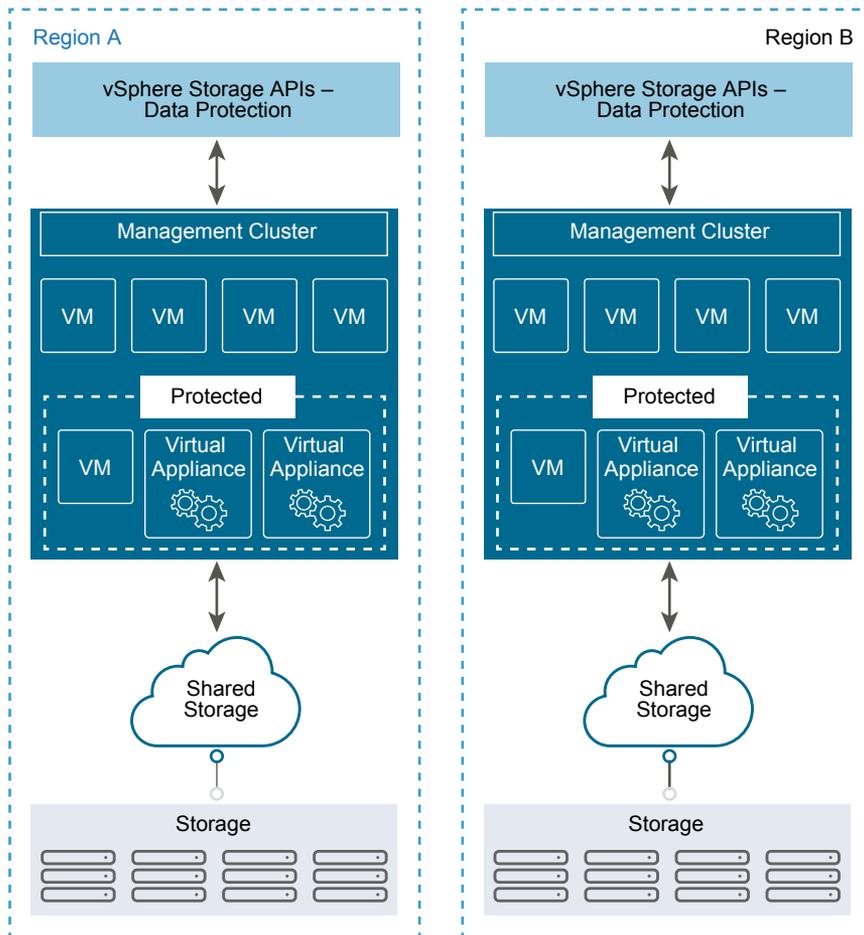
Multi-Region Data Protection Deployment

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

Multi-Region Data Protection Deployment

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

Consolidated vSphere Data Protection Deployment

Because of its scope, the VMware Validated Design for Workload and Management Consolidation deploys a single vSphere Data Protection appliance within the consolidated pod. The design contains recovery guidance about a number of SDDC management components.

vSphere Data Protection 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 1-26. Disaster Recovery Architecture
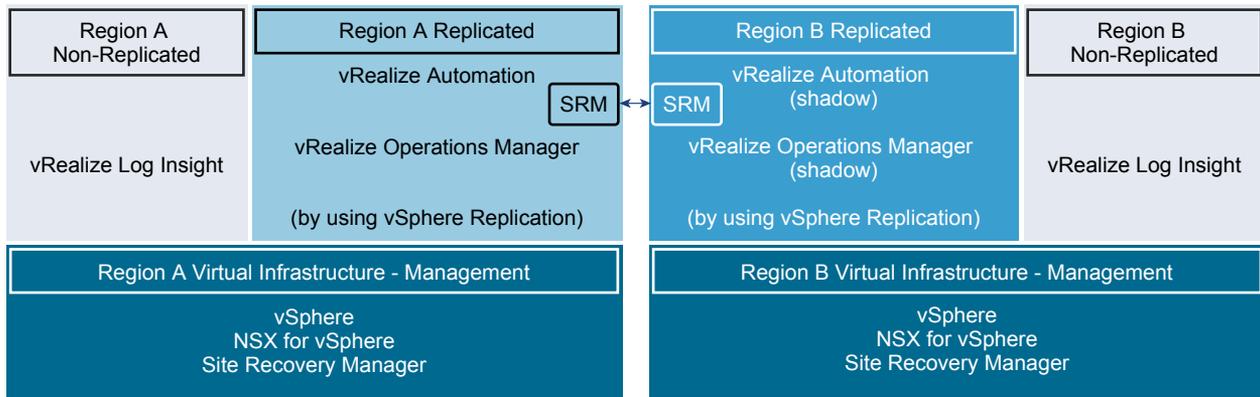
Multi-Region Site Recovery Manager Deployment

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 and vRealize Business Server

  - 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 2

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 the single availability zone setup and to multiple availability zone setup.

Physical Infrastructure Design	Focuses on the three main pillars 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.
Cloud Management Platform Design	Contains information on the consumption and orchestration layer of the SDDC stack, which uses vRealize Automation and vRealize Orchestrator. IT organizations can use the fully distributed and scalable architecture to streamline their provisioning and decommissioning operations.
Operations Infrastructure Design	Explains how to architect, install, and configure vRealize Operations Manager and vRealize Log Insight. You learn how to ensure that service management within the SDDC is comprehensive. This section ties directly into the *Operational Guidance* section.

This chapter includes the following topics:

- Physical Infrastructure Design

- Virtual Infrastructure Design

- Operations Management Design

- Cloud Management Platform Design

■ Business Continuity Design

Physical Infrastructure Design

The physical infrastructure design includes details on decisions for availability zones and regions and the cluster layout within datacenter racks.

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

Figure 2-1. Physical Infrastructure Design
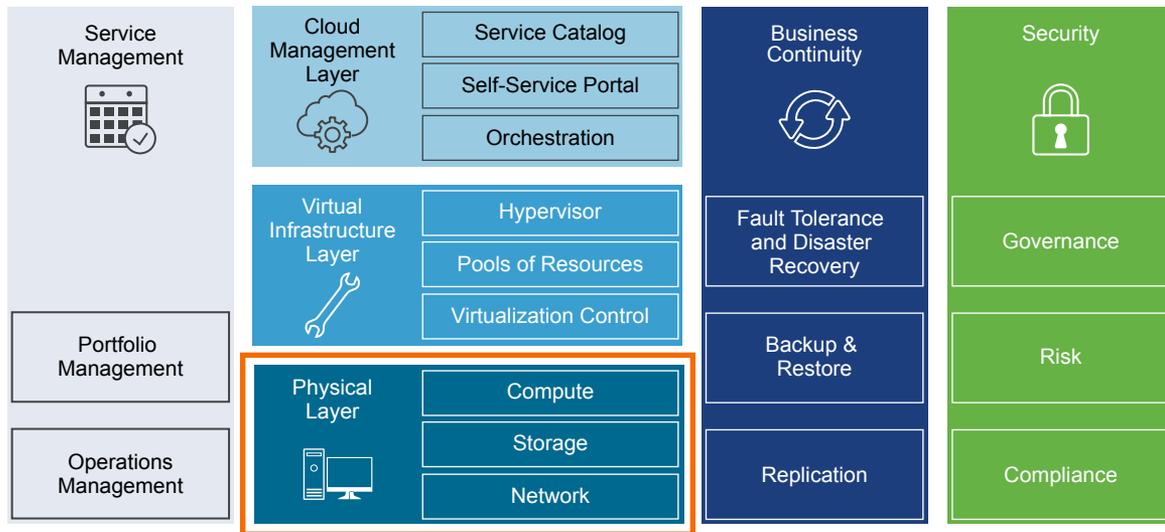
■ Physical Design Fundamentals

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

VMware Validated Design for Software-Defined Data Center can use most enterprise-grade physical network architectures.

■ Physical Storage Design

VMware Validated Design uses different types of storage, and the "Shared Storage Design" section explains where the SDDC uses each and gives background information. The focus of this section is physical storage 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.

Availability Zones and Regions

Availability zones and regions have different purposes. Availability zones protect against failures of individual hosts. Regions provide disaster recovery of the entire SDDC. This design deploys a protected region with two availability zones and recovery region with a single availability zone.

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

The SDDC in this validated design contains two regions. The design supports the option of one or two availability zones in the primary region and a single availability zone in the secondary region.

The identifiers follow United Nations Code for Trade and Transport Locations(UN/LOCODE) and also contain a numeric instance ID. Region identifiers might vary according to the locations in your deployment.

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

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

Table 2-1. **Availability Zones and Regions Design Decisions**

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-001	In Region A, deploy one or two availability zones to support all the SDDC management components and their SLAs.	A single availability zone can support all SDDC management and compute components for a region.	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 within a region.
		Two availability zones allows for stretched clusters and application aware failover for high availability between two physical locations.	Two availability zones increase the solution footprint and can complicate the operational procedures.

Table 2-1. Availability Zones and Regions Design Decisions (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 a single availability zone that can support disaster recovery of the SDDC management components.	A single availability zone can support all SDDC management and compute components for a region. You can later add another availability zone to extend and scale the management and compute capabilities of the SDDC.	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 within a region.

Clusters and Racks

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

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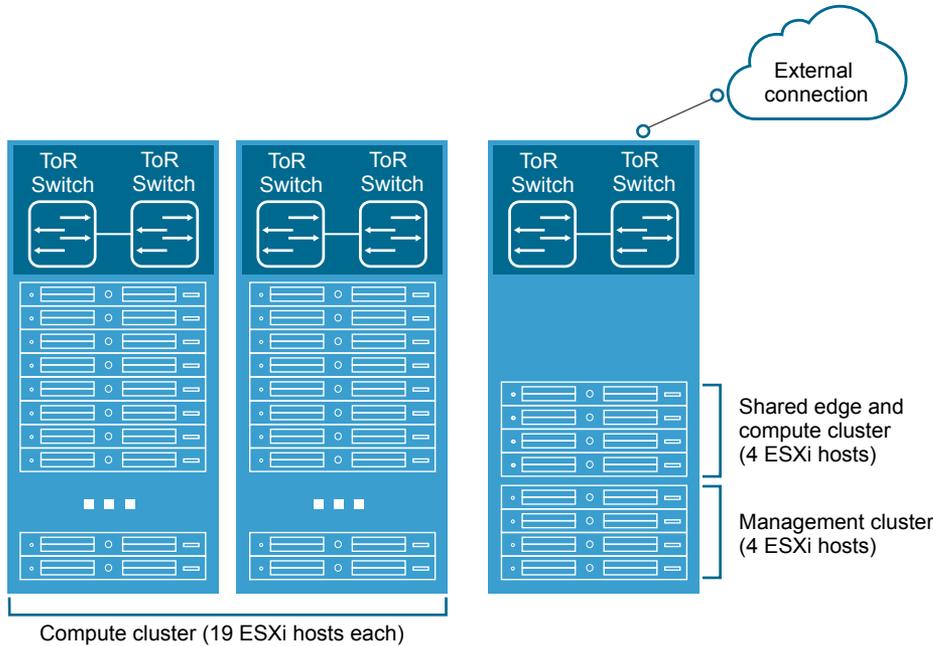
Figure 2-3. SDDC Cluster Architecture for Two Availability Zones

Table 2-2. Cluster and Racks Design Decisions

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

Table 2-2. Cluster and Racks Design Decisions (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 into a rack and within a rack, redundant power feeds prevent failure of equipment in an entire rack.	All equipment used must support two separate power feeds. The equipment must keep running if one power feed fails. If the equipment of an entire rack fails, the cause, such as flooding or an earthquake, also affects neighboring racks. Use a second region to reduce 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 list the characteristics of the ESXi hosts that were used during deployment and testing of 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. Standardizing the entire physical configuration of the ESXi hosts is critical to providing an easily manageable and supportable infrastructure because standardization eliminates variability. Deploy ESXi hosts with identical configuration, including identical storage, and networking configurations, across all cluster members. For example, consistent PCI card slot placement, especially for network controllers, is essential for accurate alignment of physical to virtual I/O resources. Identical configurations ensure an even balance of virtual machine storage components across storage and compute resources.

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

The sizing of the physical servers for the ESXi hosts for the management and edge clusters have special consideration as they use vSAN storage and vSAN Ready Nodes. See the VMware Virtual SAN Ready Nodes document for more information on vSAN Ready Nodes.

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

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

Table 2-3. ESXi Host Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-009	Use vSAN Ready Nodes with vSAN storage. If you do not use vSAN storage, verify that the system hardware complies with the ESXi requirements dictated by the *VMware Compatibility Guide*. This includes, but is not limited to: • System compatibility • I/O compatibility with network and host bus adapter (HBA) cards • Storage compatibility	Using a vSAN Ready Node ensures seamless compatibility with vSAN during the deployment. Using hardware, BIOS, and drivers from the VMware Compatibility Guide ensures the ESXi host hardware is supported.	Might limit hardware choices.
SDDC-PHY-010	You must ensure that all nodes have uniform configurations across a given cluster.	A balanced cluster delivers more predictable performance even during hardware failures. In addition, performance impact during 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 vary according to the workloads running in the cluster. When sizing memory for compute cluster hosts it is important to remember the admission control setting (n+1), which reserves one host resource for failover or maintenance.

Note See the *Administering VMware vSAN* for more information about disk groups, including design and sizing guidance. The number of disk groups and disks that an ESXi host manages determines memory requirements. 32 GB of RAM is required to support the maximum number of disk groups.

Table 2-4. Host Memory Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-011	Set up each ESXi host in the management cluster with a minimum of 192 GB RAM.	The management and edge VMs in this cluster require a total 424 GB RAM.	None.

ESXi Host Boot Device

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

Supported features depend on the version of vSAN:

- vSAN does not support stateless vSphere Auto Deploy

- vSAN 5.5 and later supports USB/SD embedded devices for ESXi boot device (4 GB or greater).

- vSAN 6.0 and later supports SATADOM as a boot device.

See the *VMware vSAN 6.5 Design and Sizing Guide* to choose the option that best fits your hardware.

Physical Networking Design

VMware Validated Design for Software-Defined Data Center can use most enterprise-grade physical network architectures.

Switch Types and Network Connectivity

Setup of the physical environment requires careful consideration. Follow best practices for physical switches, switch connectivity, 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 ssupport 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 found in a Cisco physical switch.

- Provide DHCP or DHCP Helper capabilities on all VLANs that are used by Management and VXLAN VMkernel ports. This setup simplifies the configuration by using DHCP to assign IP address based on the IP subnet in use.

- Configure jumbo frames on all switch ports, inter-switch link (ISL) and switched virtual interfaces (SVI's).

Top of Rack Connectivity and Network Settings

Each ESXi host is connected redundantly to the SDDC network fabric ToR switches by means of two 10 GbE ports. Configure the ToR switches to provide all necessary VLANs via an 802.1Q trunk. These redundant connections use features in the 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.

This Validated Design does not use hardware-based link aggregation; it is however a valid design option and is supported by VMware. If you use hardware-based link aggregation, check vendor firmware versions and VMware product documentation to verify support requirements. See VMware Cloud Foundation documentation for a VMware validated and supported design using hardware based link aggregation.

Figure 2-4. Host to ToR connectivity
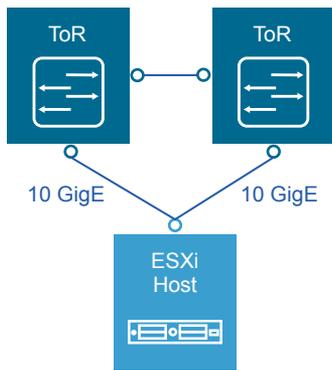
VLANs and Subnets

Each ESXi host uses VLANs and corresponding subnets.

Follow these guidelines:

- Use only /24 subnets to reduce confusion and mistakes when dealing with IPv4 subnetting.

- Use the IP address .253 as the (floating) interface with .251 and .252 for 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 2-5. 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
		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)	172.16.31.0/24
	Availability Zone 2	vMotion	1632	172.16.32.0/24

Table 2-5. Sample Values for VLANs and IP Ranges (Continued)

Cluster	Availability Zone	Function	Sample VLAN	Sample IP range
		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	4643	172.16.43.0/25

Note   The VXLAN VLAN uses the same VLAN ID in both Availability Zone 1 and Availability Zone 2. This is because NSX prepares the cluster. As such the VLAN ID must exist in both availability zones but use a different routable IP space so that NSX virtual tunnel end-points (VTEPs) can communicate across availability zones.

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 Spanning Tree Protocol, 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 (Management, vMotion, VXLAN and Storage) to 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.

Region Interconnectivity

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, etc. 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 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 region connection solution 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. Take performance requirements into consideration. Possible options are OSPF, BGP, and IS-IS. While 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 way of its IPv4 address. See the *VMware Validated Design Planning and Preparation* document for details on the DHCP server.

Table 2-6. Physical Network Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-NET-001	The physical network architecture must support the following requirements: ■ One 10-GbE port on each ToR switch for ESXi host uplinks ■ No ether-channel (LAG/vPC) configuration for ESXi host uplinks ■ Layer 3 device that supports BGP and IGMP	Guarantees availability during a switch failure. This design uses vSphere host profiles which are not compatible with link-aggregation technologies. BGP is used as the dynamic routing protocol in this design. NSX Hybrid mode replication requires IGMP.	Might limit the hardware choice. 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 two 10 GbE links to each server 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 2-7. Additional Network Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-NET-005	Assign static IP addresses to all management components in the SDDC infrastructure except for NSX VTEPs which DHCP assigns.	Avoids connection outages due to DHCP availability or misconfiguration.	Requires accurate IP address management.
SDDC-PHY-NET-006	Create DNS records for all management nodes to enable forward, reverse, short, and 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 2-8. Jumbo Frames Design Decisions

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

Physical Storage Design

VMware Validated Design uses different types of storage, and the "Shared Storage Design" section explains where the SDDC uses each and gives background information. The focus of this section is physical storage design.

All functional testing and validation of the designs is done using vSAN. Although the VMware Validated Design uses vSAN, in particular for the management clusters, you may use any supported storage solution.

If a storage solution other than vSAN is chosen, 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 VMware Validated Design, which for multiple availabilty zones includes vSAN stretched cluster.

vSAN Physical Design

This design uses VMware vSAN to implement software-defined storage for the management clusters. By using vSAN you have a high level of control on the storage subsystem.

Software-defined storage is a key technology in the SDDC. vSAN is a fully integrated hypervisor-converged storage software. vSAN creates a cluster of server hard disk drives and solid state drives, and presents a flash-optimized, highly-resilient, shared storage datastore to ESXi hosts and virtual machines. vSAN allows you to control capacity, performance, and availability on a per virtual machine basis through the use of storage policies.

Requirements and Dependencies

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

- Minimum of 3 ESXi hosts providing storage resources to the vSAN cluster.

- 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 vSphere 6.0 or later.

- Each ESXi host that provides storage resources to the cluster must meet the following requirements:
  - 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.

  - RAID controller compatible with vSAN.

  - 10 Gbps network for vSAN traffic.

  - vSphere High Availability host isolation response set to power off virtual machines. With this setting, no possibility of split-brain conditions in case of isolation or network partition exists. In a split-brain condition, the virtual machine might be powered on by two ESXi hosts by mistake. See design decision SDDC-VI-VC-012 for more details.

Hybrid Mode and All-Flash Mode

vSphere offers two different vSAN modes of operation, all-flash or 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) only as a write cache while other flash-based devices provide high endurance for capacity and data persistence.

Table 2-9. vSAN Mode Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-001	Configure vSAN in hybrid mode in the management cluster.	The VMs in the management cluster, which are hosted on vSAN, do not require the performance or expense of an all-flash vSAN configuration.	vSAN hybrid mode does not provide the potential performance or additional capabilities such as deduplication of an all-flash configuration.

Table 2-10. vSAN Physical Storage Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-002	Use one or more 200 GB or greater SSD and two or more traditional 1 TB or greater HDDs to create at least a single disk group in the management cluster.	Using a 200 GB SSD and two 1 TB HDDs provides enough capacity for the management VMs with a minimum of 10% flash-based caching.	When using only a single disk group, you limit the amount of striping (performance) capability and increase the size of the fault domain.

Hardware Considerations

You can build your own vSAN cluster or choose from a list of vSAN Ready Nodes.

Build Your Own   Be sure to use hardware from the VMware Compatibly Guide for the following vSAN components:

- Solid state disks (SSDs)

- Magnetic hard drives (HDDs)

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

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

As per design decision SDDC-PHY-009, the VMware Validated Design uses vSAN Ready Nodes.

Solid State Disk Characteristics

In a VMware vSAN configuration, the Solid State Disks (SSDs) are used for the vSAN caching layer for hybrid deployments and for the capacity layer for all flash.

- 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 holds many more write operations, thus elongating or extending the life of the SSD capacity-tier.

SSD Endurance

This VMware Validated Design uses class D endurance class 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 Petabytes Written (PBW) 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. TBW can be used for VMware vSAN 5.5, VMware vSAN 6.0, and VMware vSAN 6.5 and is reflected in the *VMware Compatibility Guide*.

The reasoning behind using TBW is that VMware provides the flexibility to use larger capacity drives with lower DWPD specifications.

If an SSD vendor uses Drive Writes Per Day as a measurement, you can calculate endurance in Terabytes Written (TBW) with the following equation.

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

For example, if a vendor specified DWPD = 10 for an 800 GB capacity SSD, you can compute TBW with the following equation.

```
TBW = 0.4TB X 10DWPD X 365days X 5yrs
TBW = 7300TBW
```

That means the SSD supports 7300 TB writes over 5 years (The higher the TBW number, the greater the endurance class.).

For SSDs that are designated for caching and all-flash capacity layers, the following table outlines which endurance class to use for hybrid and for all-flash VMware 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

Endurance Class	TBW	Hybrid Caching Tier	All-Flash Caching Tier	All-Flash Capacity Tier
Class C	>=3650	Yes	Yes	Yes
Class D	>=7300	Yes	Yes	Yes

Note This VMware Validated Design does not use All-Flash vSAN.

Table 2-11. SSD Endurance Class Design Decisions

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 designated for 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 might be more expensive than lower endurance classes.

SSD Performance Solid State Disk

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 class E or greater SSDs. 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 VMware vSAN. Before selecting a drive size, consider disk groups and sizing as well as expected future growth. VMware defines classes of performance in the VMware Compatibility Guide as follows.

Table 2-12. 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 anticipated size of the consumed HDD storage capacity, before failures to tolerate are considered. For example, select an SSD of at least 100 GB for 1 TB of HDD storage consumed in a 2 TB disk group.

Caching Algorithm

Both hybrid clusters and all-flash configurations adhere to the recommendation that 10% of consumed capacity for the flash cache layer. However, there are 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: very fast and durable write cache, and cost-effective capacity flash. Here cache is 100% allocated for writes, as read performance from capacity flash is more than sufficient.

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

Table 2-13. 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 the storage class that 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 2-14. SSD Performance Class Design Decisions

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.	The storage I/O performance requirements within the management cluster dictate the need for at least Class E SSDs.	Class E SSDs might be more expensive than lower class drives.

Magnetic Hard Disk Drive Characteristics

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 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). NL-SAS can be thought of 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 prioritized.

SAS and NL-SAS get you the best results. This VMware Validated Design uses 10,000 RPM drives to achieve 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 2-15. 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, VMware recommends that you 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 2‑16. HDD Selection Design Decisions

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 achieve a balance between performance and availability for the VMware vSAN configuration. The performance of 10,000 RPM HDDs avoids disk drain issues. In vSAN hybrid mode, the vSAN periodically flushes uncommitted writes to the capacity tier.	Slower and potentially cheaper HDDs are not available.

I/O Controllers

The I/O controllers are as important to a VMware 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.

- 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 normal day-to-day operations and increase of I/O due to Virtual Machine deployment operations or re-sync I/O activity as a result of automatic or manual fault remediation.

About SAS Expanders

SAS expanders are a storage technology that lets you maximize the storage capability of your SAS controller card. Like switches of an Ethernet network, SAS expanders enable you to connect a larger number of devices, that is, more SAS/SATA devices to a single SAS controller. Many SAS controllers support up to 128 or more hard drives.

Caution   VMware has not extensively tested SAS expanders, as a result performance and operational predictability are relatively unknown at this point. For this reason, you should avoid configurations with SAS expanders.

NFS Physical Storage Design

Network File System (NFS) is a distributed file system protocol that allows a user on a client computer to access files over a network much like local storage is accessed. In this case, the client computer is an ESXi host, and the storage is provided by an NFS-capable external storage array.

The management cluster uses vSAN for primary storage and when configured for a single availability zone NFS is introduced for secondary storage. The compute clusters are not restricted to any particular storage technology. For compute clusters, the decision on which technology to use is based on the performance, capacity, and capabilities (replication, deduplication, compression, etc.) required by the workloads that are running in the clusters.

Table 2‑17. NFS Usage Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-PHY-STO-006	When using a single availability zone, provide NFS storage.	Separate primary virtual machine storage from backup data in case of primary storage failure. vRealize Log Insight archiving requires a 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.	Store templates and ISO files 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 might increase.

Requirements

Your environment must meet the following requirements to use NFS storage in the VMware Validated Design.

■ All connections are made using 10 Gb Ethernet.

■ Jumbo frames are enabled.

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

Different disk speeds and disk types can be combined 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 2‑18. NFS Hardware Design Decision

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 requires 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.

Volumes

A volume consists of multiple disks in a storage array that RAID is applied to.

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 2-19. Volume Assignment Design Decisions

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, depending on 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 and that support the business continuity of the SDDC.

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

Figure 2-5. Virtual Infrastructure Layer in the SDDC
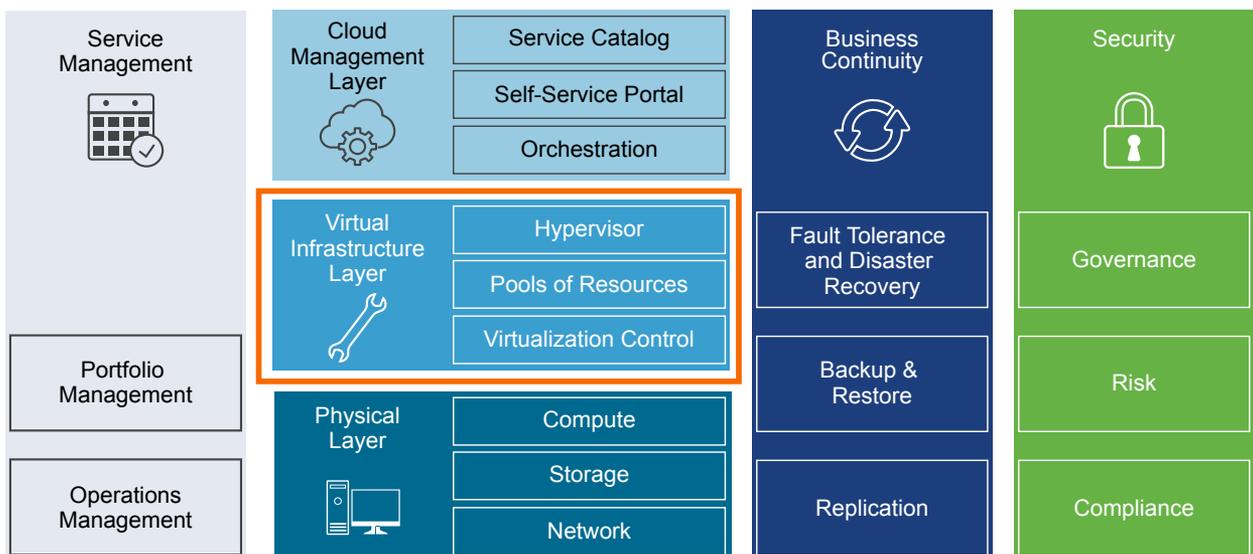
Virtual Infrastructure Design Overview

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**
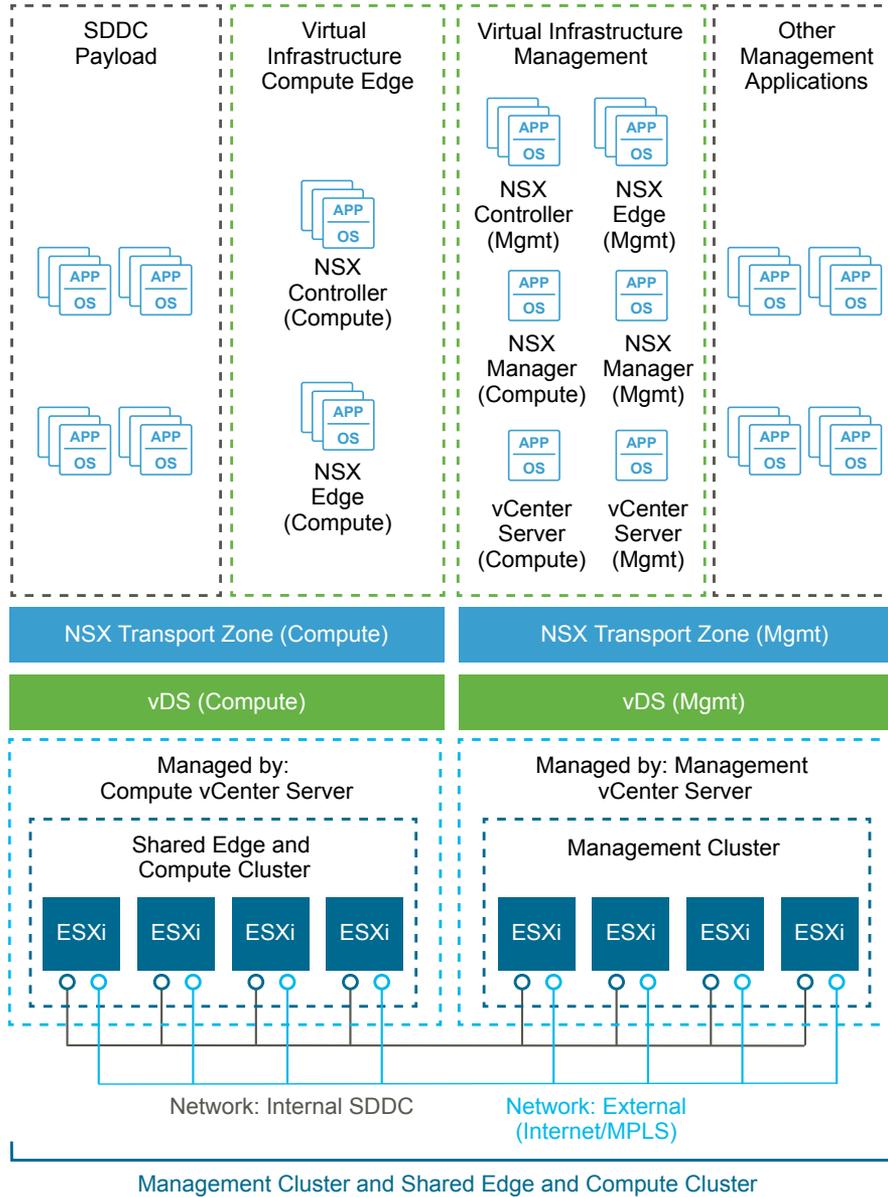
Figure 2-7. **SDDC Logical Design with Two Availablity Zones**

Availability Zone 1 - Management and Shared Edge and Compute Clusters
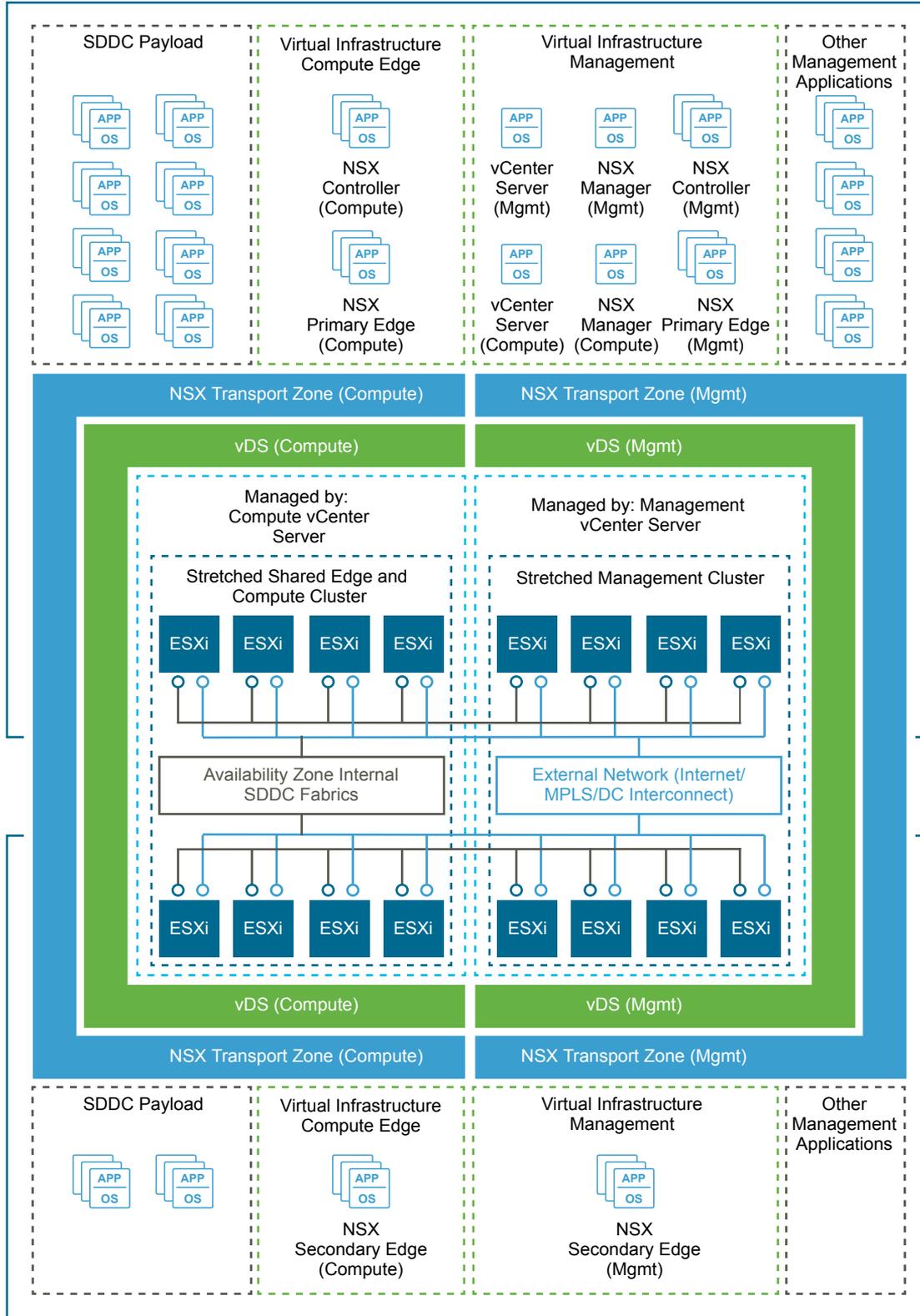
Availability Zone 2 (Secondary) - Management and Shared Edge and Compute Clusters

Management Cluster

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

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

ESXi Hardware Requirements

You can find the ESXi hardware requirements in Physical Design Fundamentals. The following design outlines the design of the ESXi configuration.

ESXi Manual Install and Boot Options

You can install or boot ESXi 6.5 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 installing ESXi.
SAN	Dedicated SAN disk on Fibre Channel or iSCSI.

USB devices	Supported for ESXi installation. 16 GB or larger SD card is recommended.
FCoE	(Software Fibre Channel over Ethernet)

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 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 USB media does not configure a default scratch partition. VMware recommends that you specify a scratch partition on a shared datastore and configure remote syslog logging for the ESXi host.

Table 2-20. ESXi Boot Disk Design Decision

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

ESXi Host Access

After installation, ESXi hosts are added to a VMware vCenter Server system and managed through that 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 three methods:

Direct Console User Interface (DCUI)	Graphical interface on the console. Allows basic administrative controls and troubleshooting options.
ESXi Shell	A Linux-style bash login on the ESXi console 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 is only used to manage single ESXi hosts. You use the VMware Host Client to conduct emergency management when vCenter Server is temporarily unavailable

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

ESXi User Access

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

Table 2-21. ESXi User Access Design Decisions

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

Virtual Machine Swap Configuration

When a virtual machine is powered on, the system creates a VMkernel swap file to serve as a backing store for the virtual machine's RAM contents. The default swap file is stored in the same location as the virtual machine's configuration file. This simplifies the configuration, however it can cause an excess of replication traffic that is not needed.

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

Table 2-22. Other ESXi Host Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-ESXi-004	Configure all ESXi hosts to synchronize time with the central NTP servers.	Required because the deployment of 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 have to allow NTP traffic on the required network ports.
SDDC-VI-ESXi-005	Enable Lockdown mode on all ESXi hosts.	To increase the security of ESXi hosts, by requiring that administrative operations be performed only from vCenter Server.	Lockdown mode settings are not part of Host Profiles and must be manually enabled on all hosts.

vCenter Server Design

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

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

- **vCenter Server Deployment**

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

- **vCenter Server Networking**

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

- **vCenter Server Redundancy**

  Protecting the vCenter Server system is important because it is the central point of management and monitoring for the SDDC. How you protect vCenter Server depends on maximum downtime tolerated and on whether failover automation is required.

- **vCenter Server Appliance Sizing**

  The following tables outline minimum hardware requirements for the management vCenter Server appliance and the compute vCenter Server appliance.

- **vSphere Cluster Design**

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

- **vCenter Server Customization**

  vCenter Server supports a rich set of customization options, including monitoring, virtual machine fault tolerance, and so on. For each feature, this VMware Validated Design specifies the design decisions.

- **Use of TLS Certificates**

  By default, vSphere 6.5 uses TLS/SSL certificates that are signed by VMCA (VMware Certificate Authority). These certificates are not trusted by end-user devices or browsers. It is a security best practice to 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 as VMCA-signed certificates.

vCenter Server Deployment

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

Table 2-23. vCenter Server Design Decisions

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 compute workloads.	Isolates vCenter Server failures to management or compute workloads. Isolates vCenter Server operations between management and compute. Supports a scalable cluster design where the management components may be re-used as additional compute needs to be added to the SDDC. Simplifies capacity planning for compute workloads by eliminating management workloads from consideration in the Compute vCenter Server. Improves the ability to upgrade the vSphere environment and related components by providing for explicit separation of maintenance windows: ■ Management workloads remain available while workloads in compute are being addressed ■ Compute workloads remain available while workloads in management are being addressed Ability to have clear separation of roles and responsibilities to ensure that only those administrators with proper authorization can attend to the management workloads. Facilitates quicker troubleshooting and problem resolution. Simplifies Disaster Recovery operations by supporting a clear demarcation between recovery of the management components and compute workloads. Enables the use of two NSX managers, one for the management cluster and the other for the shared edge and compute cluster. Network separation of the clusters in the SDDC allows for 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 2-24. vCenter Server Platform Design Decisions

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

Platform Services Controller Design Decision Background

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

- In an embedded deployment, vCenter Server and the Platform Services Controller run on the same virtual machine. Embedded deployments are recommended for standalone environments with only one vCenter Server system.

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

- If there is a need to replicate with other Platform Services Controller instances, or if the solution includes more than one vCenter Single Sign-On instance, you can deploy multiple external Platform Services Controller instances on separate virtual machines.

Table 2-25. Platform Service Controller Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-003	Deploy each vCenter Server with an external Platform Services Controller.	External Platform Services Controllers are required for replication between Platform Services Controller instances.	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 into a single vCenter Single Sign-On domain, they can share authentication and license data across all components and regions.	Only one Single Sign-On domain will exist.
SDDC-VI-VC-005	Create a ring topology for the Platform Service Controllers.	By default, Platform Service Controllers only replicate with one other Platform Services Controller, that creates a single point of failure for replication. A ring topology ensures each Platform Service Controller has two replication partners and eliminates any single point of failure.	Command-line interface commands must be used to configure the ring replication topology.
SDDC-VI-VC-006	Use an NSX Edge Services Gateway as a load balancer for the Platform Services Controllers.	Using a load balancer increases the availability of the PSC's for all applications.	Configuring the load balancer and repointing vCenter Server to the load balancers Virtual IP (VIP) creates administrative overhead.

Figure 2‑8. vCenter Server and Platform Services Controller Deployment Model
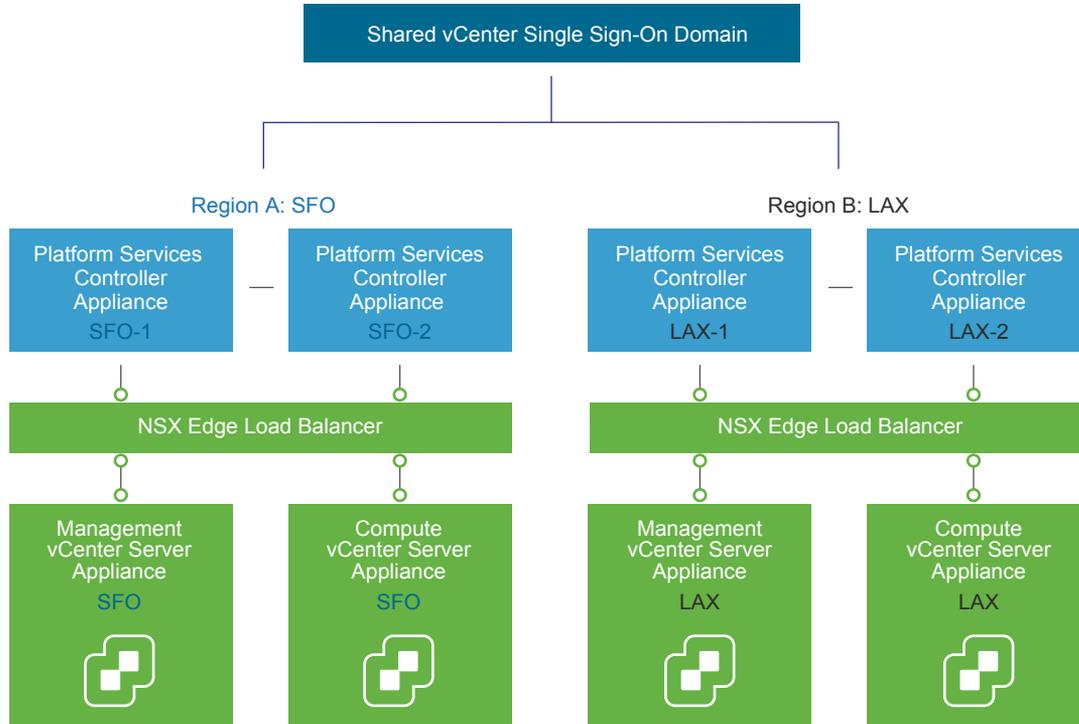
vCenter Server Networking

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:

- All VMware vSphere Client and vSphere Web Client user interfaces.
- 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. How you protect vCenter Server depends on maximum downtime tolerated and on whether failover automation is required.

The following table lists methods available for protecting the vCenter Server system and the vCenter Server Appliance.

Table 2-26. Methods for Protecting vCenter Server System and the vCenter Server Appliance

Redundancy Method	Protects vCenter Server system (Windows)	Protects Platform Services Controller (Windows)	Protects vCenter Server (Appliance)	Protects Platform Services Controller (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	Not Available

Table 2-27. vCenter Server Protection Design Decisions

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

vCenter Server Appliance Sizing

The following tables outline minimum hardware requirements for the management vCenter Server appliance and the compute vCenter Server appliance.

Table 2-28. Logical Specification for Management vCenter Server Appliance

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

Table 2-29. Logical Specification for Compute vCenter Server Appliance

Attribute	Specification
vCenter Server version	6.5 (vCenter Server Appliance)
Physical or virtual system	Virtual (appliance)
Appliance Size	Large (up to 1,000 hosts / 10,000 VMs)
Platform Services Controller	External
Number of CPUs	16

Table 2-29. Logical Specification for Compute vCenter Server Appliance (Continued)

Attribute	Specification
Memory	32 GB
Disk Space	640 GB

Table 2-30. vCenter Server Appliance Sizing Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-008	Configure the management vCenter Server Appliances with at least the small-size setting.	Based on the number of management VMs that are running, a vCenter Server Appliance installed with the small-size setting is sufficient.	If the size of the management environment changes, the vCenter Server Appliance size might need to be increased.
SDDC-VI-VC-009	Configure the compute vCenter Server Appliances with at least the large-size setting.	Based on the number of compute workloads and NSX Edge devices running, a vCenter Server Appliance installed with the large-size setting is preferable.	As the compute environment expands resizing to X-Large or adding additional vCenter Server instances may be required.

vSphere Cluster Design

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

vSphere Cluster Design Decision Background

The following heuristics help with cluster design decisions.

- Decide to use fewer, larger ESXi hosts or more, smaller ESXi hosts.

    - A scale-up cluster has fewer, larger ESXi hosts.

    - A scale-out cluster has more, smaller ESXi hosts.

- Compare the capital costs of purchasing fewer, larger ESXi hosts with the costs of purchasing more, smaller ESXi hosts. Costs vary between vendors and models.

- Evaluate the operational costs of managing a few ESXi hosts with the costs of managing more ESXi hosts.

- Consider the purpose of the cluster.

- Consider the total number of ESXi hosts and cluster limits.

Figure 2-9. vSphere Logical Cluster Layout with a Single Availablity Zone
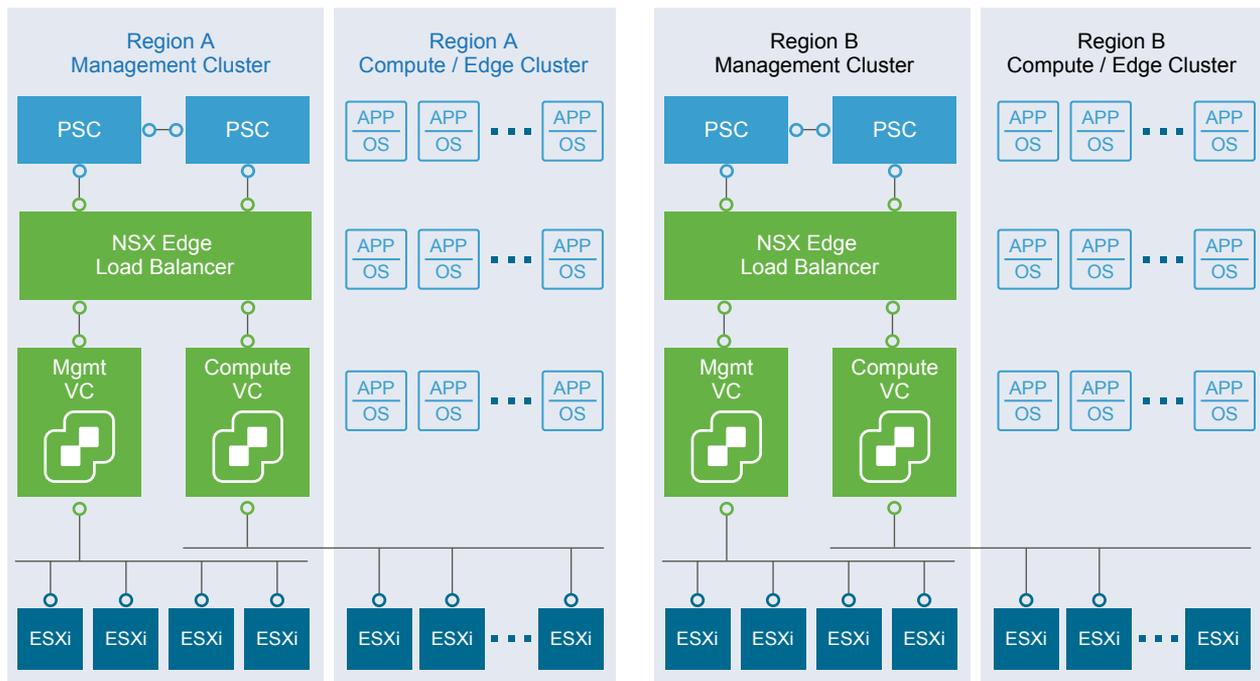
Figure 2-10. vSphere Logical Cluster Layout with Two Availability Zones
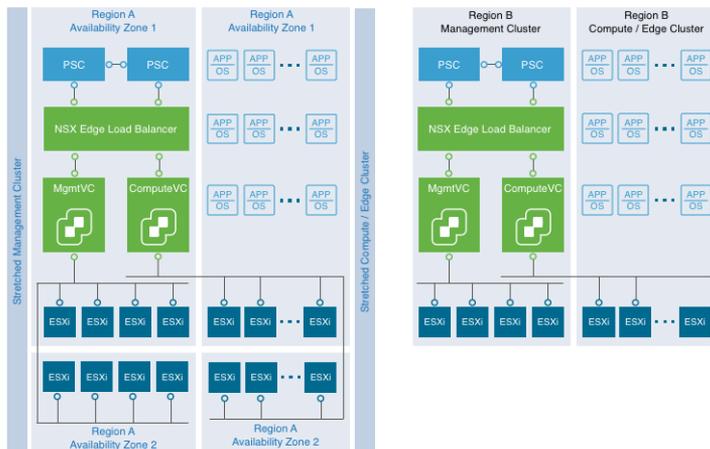
vSphere High Availability Design

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 HA Design Basics

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

The master ESXi host detects different types of failure:

- ESXi host failure, for example an unexpected power failure

- ESXi host network isolation or connectivity failure

- Loss of storage connectivity

- Problems with virtual machine OS availability

Table 2-31. vSphere HA Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-010	Use vSphere HA to protect all clusters against failures.	vSphere HA supports a robust level of protection for both ESXi host and virtual machine availability.	Sufficient resources on the remaining host are required to 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 judges' 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.
Percentage of cluster resources reserved. vSphere HA ensures that a specified percentage of aggregate CPU and memory resources are reserved 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

This design defines the following vSphere clusters and the workloads that they handle.

Table 2-32. vSphere Cluster Workload Design Decisions

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

Management Cluster Design

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

Table 2-33. Management Cluster Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-014	Create a management cluster in Region A with a minimum of 4 ESXi hosts for a single availability zone or a minimum of 8 ESXi hosts for two availability zones (minimum of 4 ESXi hosts in each availability zone).	Allocating four ESXi hosts provides full redundancy for each availability zone within the cluster. Having four 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 4 ESXi hosts.	Allocating four 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.

Table 2-33. Management Cluster Design Decisions (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
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 6.5 automatically calculates the reserved percentage based on ESXi host failures to tolerate and the number of ESXi hosts in the cluster.	In a four ESXi host management cluster only the resources of three ESXi hosts are available for use.
SDDC-VI-VC-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 there are enough resources for all VMs if an availability zone outage occurs.	In an eight-host management cluster, only the resources of four 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.
SDDC-VI-VC-020	When using a single availability zone create a host profile for the Management Cluster.	Using host profiles simplifies configuration of ESXi hosts and ensures settings are uniform across 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.
SDDC-VI-VC-021	When using two availability zones create a host profile for each availability zone in the Management Cluster.	Using host profiles simplifies configuration of ESXi hosts and ensures settings are uniform across the availability zones in the cluster.	Anytime an authorized change to a ESXi host is made the host profile must be updated to reflect the change or the status will show non-compliant. Due to configuration differences between availability zones two host profiles are required and must be applied at a per ESXi host level.

The following table summarizes the attributes of the management cluster logical design.

Table 2-34. Management Cluster Logical Design Background

Attribute	Specification
Number of ESXi hosts required to support management hosts with no over commitment .	2
Number of ESXi hosts recommended due to operational constraints (Ability to take a host offline without sacrificing High Availability capabilities) .	4
Number of ESXi hosts recommended due to 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. Due to the shared nature of the cluster, NSX Controllers and Edge devices run in this cluster. The design decisions determine the number of ESXi hosts and vSphere HA settings and several other characteristics of the shared edge and compute cluster.

Table 2-35. Shared Edge and Compute Cluster Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-022	Create a shared edge and compute cluster for the NSX Controllers and NSX Edge gateway devices.	NSX Manager requires a 1:1 relationship with a vCenter Server system.	Each time you provision a Compute vCenter Server system, a new NSX Manager is required. Set anti-affinity rules to keep each Controller on a separate ESXi host. A 4-node cluster allows 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.

Table 2-35. Shared Edge and Compute Cluster Design Decisions (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
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. As an example 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 there are enough resources for all VMs in an availability zone outage.	If additional ESXi hosts are added to the cluster, they will need to be added 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 a minimum of 8 ESXi hosts for two availability zones (minimum of 4 ESXi hosts in each availability zone).	Allocating four ESXi hosts provides full redundancy for each availability zone within the cluster. Having four 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.
SDDC-VI-VC-026	In Region B create a shared edge and compute cluster with a minimum of 4 hosts.	■ 3 NSX Controllers are required for sufficient redundancy and majority decisions. ■ One 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 Controllers 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 it is imperative that these virtual machines receive all the resources required.	During contention SDDC NSX components receive more resources than all other workloads as such monitoring and capacity management must be a proactive activity.
SDDC-VI-VC-029	Create a resource pool for all user NSX Edge devices with a CPU share value of Normal and a memory share value of Normal.	NSX 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.

Table 2-35. Shared Edge and Compute Cluster Design Decisions (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
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 above all other workloads as to not impact 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 user workload virtual machines might be starved for resources and experience poor performance. It is critical that monitoring and capacity management must be a proactive activity and that capacity is added or a dedicated edge cluster is created before contention occurs.
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.	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-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.	Using host profiles simplifies configuration of ESXi hosts and ensures settings are uniform across 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.
SDDC-VI-VC-034	When using two availability zones create a host profile for each availability zone in the cluster.	Using host profiles simplifies configuration of ESXi hosts and ensures 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. Due to configuration differences between availability zones two host profiles are required and must be applied on a per ESXi host level.

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

Table 2-36. Shared Edge and Compute Cluster Logical Design Background

Attribute	Specification
Minimum number of ESXi hosts required to support the shared edge and compute cluster	3
Number of ESXi hosts recommended due to operational constraints (Ability to take an ESXi host offline without sacrificing High Availability capabilities)	4
Number of ESXi hosts recommended due to 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, additional compute-only clusters can be configured. Tenant workloads run on the ESXi hosts in the compute cluster instances. Multiple compute clusters are managed by the Compute vCenter Server instance. The design determines vSphere HA settings for the compute cluster.

Table 2-37. Compute Cluster Design Decisions

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 to fail 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. As an example, 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 there are enough resources for all VMs in 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 rich set of customization options, including monitoring, virtual machine fault tolerance, and so on. For each feature, this VMware Validated Design specifies the design decisions.

Virtual Machine and Application Monitoring Service

When Virtual Machine and Application Monitoring is enabled, the Virtual Machine and Application Monitoring service, which uses VMware Tools, evaluates whether each virtual machine in the cluster is running. The service checks for regular heartbeats and I/O activity from the VMware Tools process running on guests. If the service receives no heartbeats or I/O activity, it is likely that the guest operating system has failed or that 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 Virtual Machine Monitoring for automatic restart of a failed virtual machine. The application or service running on the virtual machine must be capable of restarting successfully after a reboot or the virtual machine restart is not sufficient.

Table 2-38. Monitor Virtual Machines Design Decisions

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

VMware vSphere Distributed Resource Scheduling (DRS)

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

Manual

Recommendations are made but an administrator needs to 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 would provide a slight improvement in balancing will be executed.

When using two availability zones enable vSphere DRS to create VM/Host group affinity rules for initial placement of VMs and impacting read locality. This way you avoid unnecessary vSphere vMotion migration of VMs between Availability Zones. Because the 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 pin virtual machines to an availability zone. If a virtual machine, VM1, that resides in availability zone 1 moves freely across availability zones, VM1 could end up on availability zone 2. Because vSAN stretched clusters implement read locality, the cache on availability zone 1 is warm whereas the cache in availability zone 2 is cold. This may impact the performance of VM1 until the cache in availability zone 2 has been warmed.

Table 2-39. vSphere Distributed Resource Scheduling Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-VC-040	Enable DRS on all clusters and set it to Fully Automated, with the default setting (medium).	The default settings provide the best trade-off between load balancing and excessive migration with vMotion events.	In the event of a vCenter outage, 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 correct 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 Zone2 to it.	Ensures that virtual machines are located only in the assigned availability zone.	You must add VMs to the correct group manually to ensure they are not powered-on in or migrated to the wrong availability zone.

Enhanced vMotion Compatibility (EVC)

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 at a later date without disruption. You can use EVC for a rolling upgrade of all hardware with zero downtime.

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

Table 2-40. VMware Enhanced vMotion Compatibility Design Decisions

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

Use of TLS Certificates

By default, vSphere 6.5 uses TLS/SSL certificates that are signed by VMCA (VMware Certificate Authority). These certificates are not trusted by end-user devices or browsers. It is a security best practice to 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 as VMCA-signed certificates.

Table 2-41. vCenter Server TLS Certificate Design Decisions

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 3rd 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. Certificate warnings result with the default certificate.	Replacing and managing certificates is an operational overhead.
SDDC-VI-VC-047	Use an 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

  This VMware Validated Design follows high-level network design guidelines and networking best practices.

- Virtual Switches

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

- NIC Teaming

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

- Network I/O Control

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

- VXLAN

  VXLAN 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 vMotion and to assign a dedicated default gateway for vMotion traffic.

Virtual Network Design Guidelines

This 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 10 GbE NICs with VLANs might be more cost effective than configuring a dozen 1 GbE NICs on separate physical networks.

- Boost 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, typically by providing network redundancy.

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

- Enhance infrastructure functionality. You can configure the network to support vSphere features such as vSphere vMotion, vSphere High Availability, and vSphere Fault Tolerance.

Best Practices

Follow 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 migration with vMotion occurs, the contents of the guest operating system's memory is transmitted over the network. You can put vSphere vMotion on a separate network by using a dedicated vSphere vMotion VLAN.

- When using pass-through devices with Linux kernel version 2.6.20 or an earlier guest OS, avoid MSI and MSI-X modes. These modes have significant performance impact.

- For best performance, use VMXNET3 virtual 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. Separate networks are also required for access security.

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

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

Virtual Networks

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

- vSphere operational traffic.

  - Management

  - vMotion

  - vSAN

  - NFS Storage

  - vSphere Replication

  - VXLAN

- Traffic that supports the organization's services and applications.

Virtual Switches

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

Virtual Switch Design Background

A vSphere Distributed Switch (distributed switch) offers several enhancements over standard virtual switches.

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

Consider the following caveats for distributed switches.

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

Number of Virtual Switches

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

Table 2-42. Virtual Switch Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-001	Use vSphere Distributed Switches.	vSphere Distributed Switches simplify management.	Migration from a standard switch to a distributed switch requires a minimum of two physical NICs to maintain redundancy.
SDDC-VI-NET-002	Use a single vSphere Distributed Switch per cluster.	Reduces complexity of the network design. Reduces the size of the fault domain.	Increases the number of vSphere Distributed Switches that must be managed.
SDDC-VI-NET-003	Use an Ephemeral binding for the management port group.	Ephemeral binding allows for the recovery of the vCenter Server that is managing the vSphere Distributed Switch.	Port-level permissions and controls are lost across power cycles, so no historical context is saved.
SDDC-VI-NET-004	Use a Static binding for all 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.

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

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

- Mismatched virtual switch teaming policies for the physical switch port-channel settings.

Health check monitors VLAN, MTU, and teaming policies.

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

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

Table 2-43. Distribute Switch Health Check Design Decisions

Design ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-005	Enable vSphere Distributed Switch Health Check on all virtual distributed switches.	vSphere Distributed Switch Health Check ensures all VLANS are trunked to all ESXi hosts attached to the vSphere Distributed Switch and ensures 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 not all VLANs are 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 Switches

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

Table 2-44. 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) to enable ECMP	Enabled	2	9000

Table 2-45. vDS-MgmtPort Group Configuration Settings

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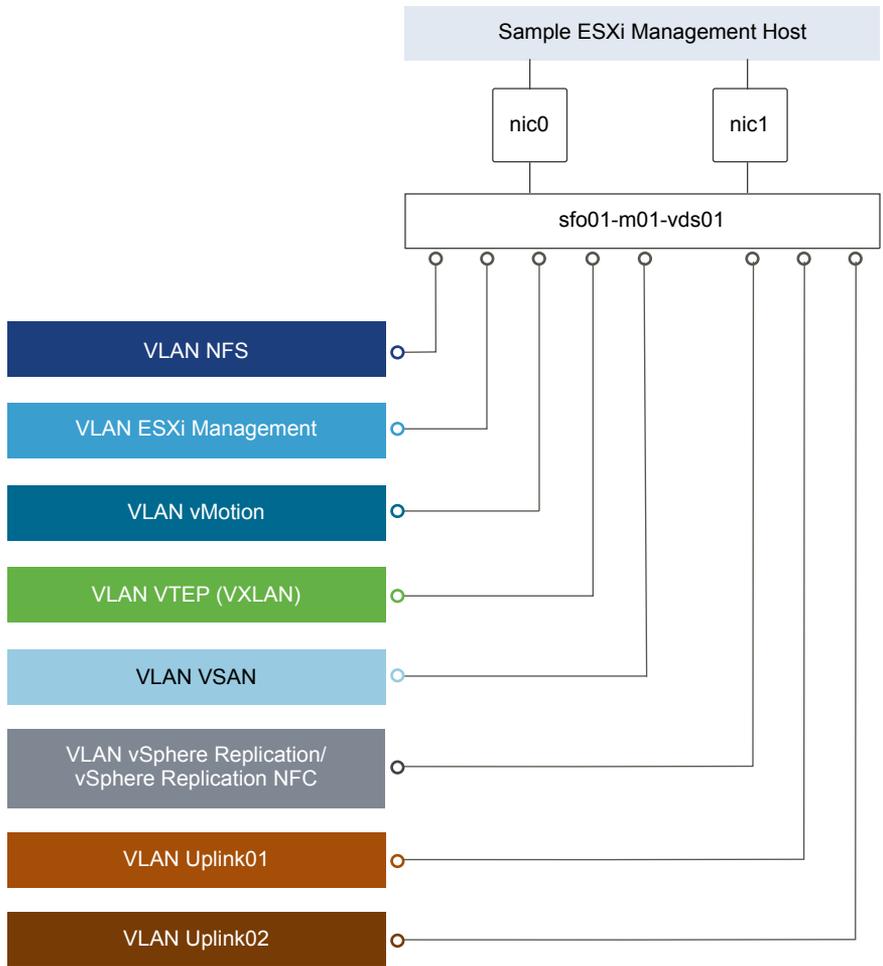
This section expands on the logical network design by providing details on the physical NIC layout and physical network attributes.

Table 2-46. Management Virtual Switches by Physical/Virtual NIC

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

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

Table 2-47. 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 2-47. 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 2-48. Management VMkernel Adapter

vSphere Distributed Switch	Network Label	Connected Port Group	Enabled Services	MTU
sfo01-m01-vds01	Management	sfo01-m01-vds01-management	Management Traffic	1500 (Default)
sfo01-m01-vds01	vMotion	sfo01-m01-vds01-vmotion	vMotion Traffic	9000
sfo01-m01-vds01	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
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 with the following configuration settings.

Table 2-49. 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 2-50. vDS-Comp01 Port Group Configuration Settings

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

Network Switch Design for Shared Edge and Compute ESXi Hosts

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

Figure 2-12. Network Switch Design for Shared Edge and Compute ESXi Hosts
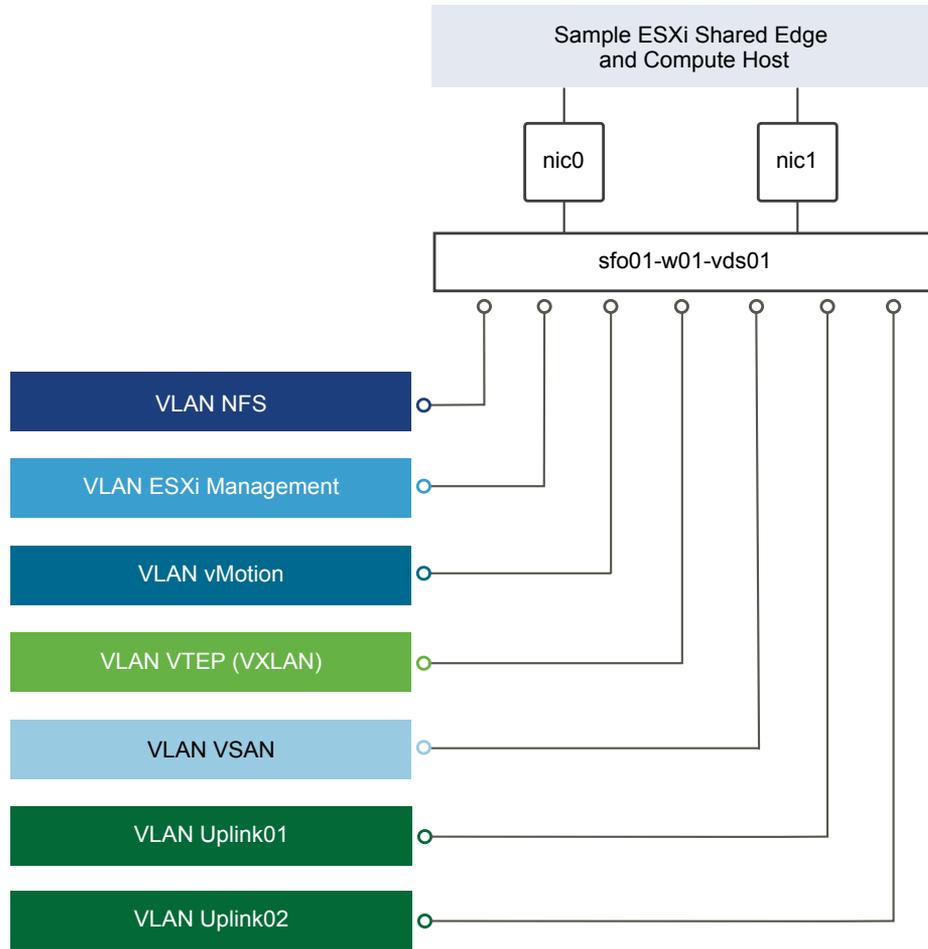
Table 2-51. Shared Edge and Compute Cluster Virtual Switches by Physical/Virtual NIC

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

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

Table 2-52. Shared Edge and Compute Cluster Virtual Switch Port Groups and VLANs

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
sfo01-w01-vds01	sfo01-w01-vds01-management	Route based on physical NIC load	1, 2	1631
sfo01-w01-vds01	sfo01-w01-vds01-vmotion	Route based on physical NIC load	1, 2	1632
sfo01-w01-vds01	sfo01-w01-vds01-vsan	Route based on physical NIC load	1, 2	1633
sfo01-w01-vds01	sfo01-w01-vds01-nfs	Route based on physical NIC load	1, 2	1615

Table 2-52. Shared Edge and Compute Cluster Virtual Switch Port Groups and VLANs (Continued)

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
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 2-53. Shared Edge and Compute Cluster VMkernel Adapter

vSphere Distributed Switch	Network Label	Connected Port Group	Enabled Services	MTU
sfo01-w01-vds01	Management	sfo01-w01-vds01-management	Management Traffic	1500 (Default)
sfo01-w01-vds01	vMotion	sfo01-w01-vds01-vmotion	vMotion Traffic	9000
sfo01-w01-vds01	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 vSphere Distributed Switch uses the following configuration settings.

Table 2-54. 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 2-55. vDS-Comp02 Port Group Configuration Settings

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

Network Switch Design for Compute ESXi Hosts

Figure 2-13. Network Switch Design for Compute ESXi Hosts
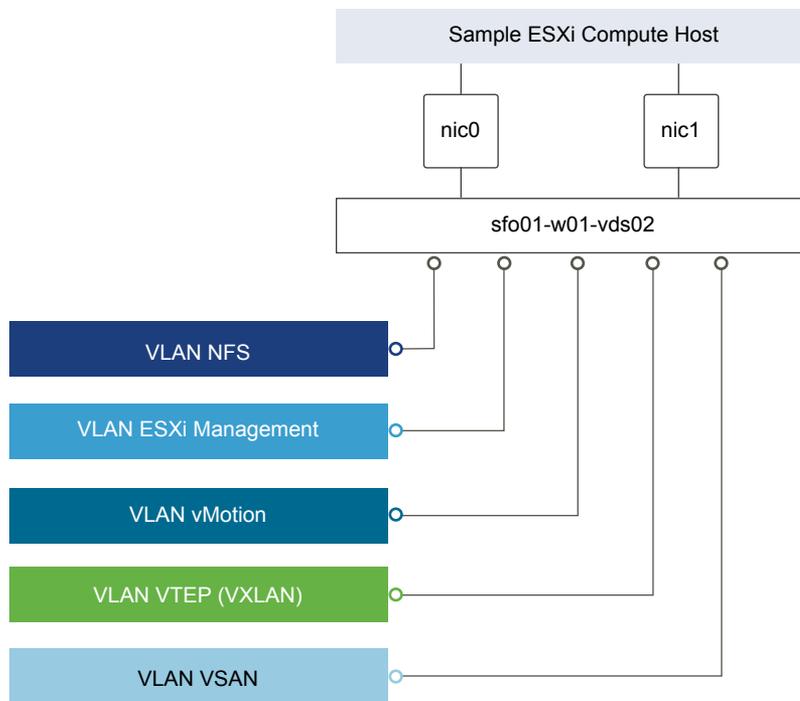
This section expands on the logical network design by providing details on the physical NIC layout and physical network attributes.

Table 2-56. Compute Cluster Virtual Switches by Physical/Virtual NIC

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

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

Table 2-57. Compute Cluster Virtual Switch Port Groups and VLANs

vSphere Distributed Switch	Port Group Name	Teaming Policy	Active Uplinks	VLAN ID
sfo01-w01-vds02	sfo01-w01-vds02-management	Route based on physical NIC load	1, 2	1621
sfo01-w01-vds02	sfo01-w01-vds02-vmotion	Route based on physical NIC load	1, 2	1622
sfo01-w01-vds02	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 2-58. 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 further reduce the risk of a single point of failure, build NIC teams by using ports from multiple NIC and motherboard interfaces.

Create a single virtual switch with teamed NICs across separate physical switches.

This VMware Validated Design uses an active-active configuration using the route based on physical NIC load algorithm for teaming. In this configuration, idle network cards do not wait for a failure to occur, and they aggregate bandwidth.

NIC Teaming Design Background

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

- An active-passive configuration that uses explicit failover when connected to two separate switches.

- An active-active configuration in which two or more physical NICs in the server are assigned the active role.

This validated design uses an active-active configuration.

Table 2-59. 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 2-60. NIC Teaming Design Decision

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 ones that carry VXLAN traffic. VTEP kernel ports and VXLAN traffic use route based on SRC-ID.	Reduce the complexity of the network design and increase 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 following system traffic types.

- Fault tolerance traffic
- iSCSI traffic
- vSphere vMotion traffic
- Management traffic
- VMware vSphere Replication traffic
- NFS traffic
- vSAN traffic

- Backup traffic

- Virtual machine traffic

How Network I/O Control Works

Network I/O Control enforces the share value specified for the different traffic types only when there is network contention. When contention occurs Network I/O Control applies the share values set to each traffic type. As a result, less important traffic, as defined by the share percentage, will be throttled, allowing more important traffic types to gain access to more network resources.

Network I/O Control also allows the 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 vCenter CPU and memory reservations.

Network I/O Control Heuristics

The following heuristics can help with design decisions.

Shares vs. Limits	When you use bandwidth allocation, consider using shares instead of limits. Limits impose hard limits on the amount of bandwidth used by a traffic flow even when network bandwidth is available.
Limits on Certain Resource Pools	Consider imposing limits on a given resource pool. For example, if you put a limit on vSphere vMotion traffic, you can benefit in situations where multiple vSphere vMotion data transfers, initiated on different ESXi hosts at the same time, result in oversubscription at the physical network level. By limiting the available bandwidth for vSphere vMotion at the ESXi host level, you can prevent performance degradation for other traffic.
Teaming Policy	When you use Network I/O Control, use Route based on physical NIC load teaming as a distributed switch teaming policy to maximize the networking capacity utilization. With load-based teaming, traffic might move among uplinks, and reordering of packets at the receiver can 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.

Network I/O Control Design Decisions

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

Table 2-61. Network I/O Control Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-007	Enable Network I/O Control on all distributed switches.	Increase resiliency and performance of the network.	If configured incorrectly, Network I/O Control might impact network performance for critical traffic types.
SDDC-VI-NET-008	Set the share value for vMotion traffic to Low.	During times of network contention, vMotion traffic is not as important as virtual machine or storage traffic.	During times of network contention, vMotion takes longer than usual to complete.
SDDC-VI-NET-009	Set the share value for vSphere Replication traffic to Low.	During times of network contention, vSphere Replication traffic is not as important as virtual machine or storage traffic.	During times of network contention, vSphere Replication takes longer and might violate the defined SLA.
SDDC-VI-NET-010	Set the share value for vSAN to High.	During times of network contention, vSAN traffic needs a guaranteed bandwidth so virtual machine performance does not suffer.	None.
SDDC-VI-NET-011	Set the share value for Management to Normal.	By keeping the default setting of Normal management traffic is prioritized higher than vMotion and vSphere Replication but lower than vSAN traffic. Management traffic is important as 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 are slower than usual.
SDDC-VI-NET-013	Set the share value for backup traffic to Low.	During times of network contention, it is more important that primary functions of the SDDC continue to have access to network resources over backup traffic.	During times of network contention, backups are slower than usual.
SDDC-VI-NET-014	Set the share value for virtual machines to High.	Virtual machines are the most important asset in the SDDC. Leaving the default setting of High ensures that they always have access to the network resources they need.	None.
SDDC-VI-NET-015	Set the share value for Fault Tolerance to Low.	Fault Tolerance is not used in this design therefore it can be set to the lowest priority.	None.
SDDC-VI-NET-016	Set the share value for iSCSI traffic to Low.	iSCSI is not used in this design therefore it can be set to 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 into logical networks that are directly attached to specific applications.

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

In the vSphere architecture, the encapsulation is performed between the virtual NIC of the guest 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 2-62. VXLAN Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-017	Use NSX for vSphere to introduce VXLANs for the use of virtual application networks and tenants networks.	Simplify the network configuration for each tenant using centralized virtual network management.	Requires additional compute and storage resources to deploy NSX components. Additional training may be needed on NSX.
SDDC-VI-NET-018	Use VXLAN with NSX Edge gateways, the Universal Distributed Logical Router (UDLR), and Distributed Logical Router (DLR) to provide customer/tenant network capabilities.	Create 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.	Leverage benefits of the network virtualization in the management cluster.	Requires installation and configuration of an NSX for vSphere instance in the management cluster.

vMotion TCP/IP Stack

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

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

- Route the traffic for the migration of virtual machines that are powered on or powered off by using a default gateway that is different from the gateway assigned to the default stack on the ESXi host.

- Assign a separate set of buffers and sockets.

- Avoid routing table conflicts that might otherwise appear when many features are using a common TCP/IP stack.

- Isolate traffic to improve security.

Table 2-63. vMotion TCP/IP Stack Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-NET-020	Use the vMotion TCP/IP stack for vMotion traffic.	By using the vMotion TCP/IP stack, vMotion traffic can use a default gateway on its own subnet, allowing for vMotion traffic to go over Layer 3 networks.	The vMotion TCP/IP stack is not available in the vDS VMkernel creation wizard, and as such the VMkernal 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 requires a decoupling of virtual networks from their physical counterparts. This, 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

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

Table 2-64. NSX for vSphere Design Decisions

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-001	Use two separate NSX instances per region. One instance is tied to the Management vCenter Server, and the other instance is tied to the Compute vCenter Server.	Software-defined Networking (SDN) capabilities offered by NSX, such as load balancing and firewalls, are crucial for the compute/edge layer to support 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
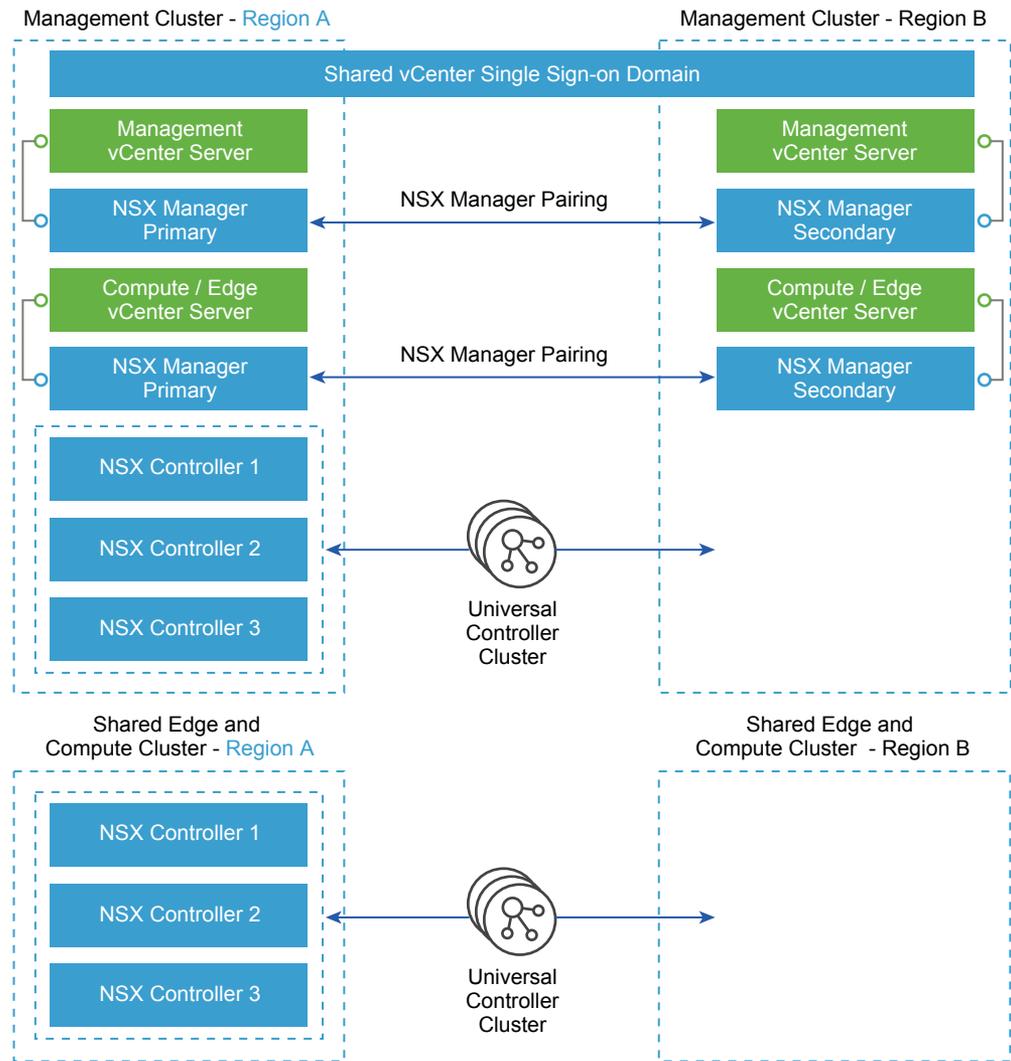
NSX Components

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 REST API and the vSphere Web Client.

Cloud Management Platform

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

API

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

- A client can read an object by making an HTTP GET request to the object's resource URL.

- A client can write (create or modify) an object with an HTTP PUT or POST request that includes a new or changed XML document for the object.

- A client can delete an object with an HTTP DELETE request.

vSphere Web Client

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

Table 2-65. Consumption Method Design Decisionsin vSphere Web Client

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.	Customers 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 2-66. NSX Controller Design Decision

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-005	Deploy NSX Controller instances in Universal Cluster mode with three members to provide high availability and scale. Provision these three nodes through the primary NSX Manager instance.	The high availability of NSX Controller reduces the downtime period in case of failure of one physical ESXi host.	The secondary NSX Manager 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, type 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. Hybrid replication mode is the recommended 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.	The NSX Manager requires NTP settings that synchronize it with the rest of the vSphere environment. Drift can cause problems with authentication. The 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

The following table lists the components involved in the NSX for vSphere solution and the requirements for installing and running them. The compute and storage requirements have been taken into account when sizing resources to support the NSX for vSphere solution.

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

VM	vCPU	Memory	Storage	Quantity per Stack Instance
NSX Manager	4	16 GB	60 GB	1
NSX Controller	4	4 GB	20 GB	3
NSX ESG	1 (Compact) 2 (Large) 4 (Quad Large) 6 (X-Large)	512 MB (Compact) 1 GB (Large) 1 GB (Quad Large) 8 GB (X-Large)	512 MB 512 MB 512 MB 4.5 GB (X-Large) (+4 GB with swap)	Optional component. Deployment of the NSX ESG varies per use case.
DLR control VM	1	512 MB	512 MB	Optional component. Varies with use case. Typically 2 per HA pair.
Guest introspection	2	1 GB	4 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 model is suitable for high performance firewall abilities and the X-Large 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, so the recommendation is to begin with the Large model and scale up if necessary. A Large NSX Edge service gateway is suitable for medium firewall performance but as detailed later, the NSX Edge service gateway does not perform the majority of firewall functions.

Note Edge service gateway throughput is influenced by the WAN circuit. An adaptable approach, that is, converting as necessary, is recommended.

Table 2-67. NSX Edge Service Gateway Sizing Design Decision

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

Network Virtualization Conceptual Design

This conceptual design 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
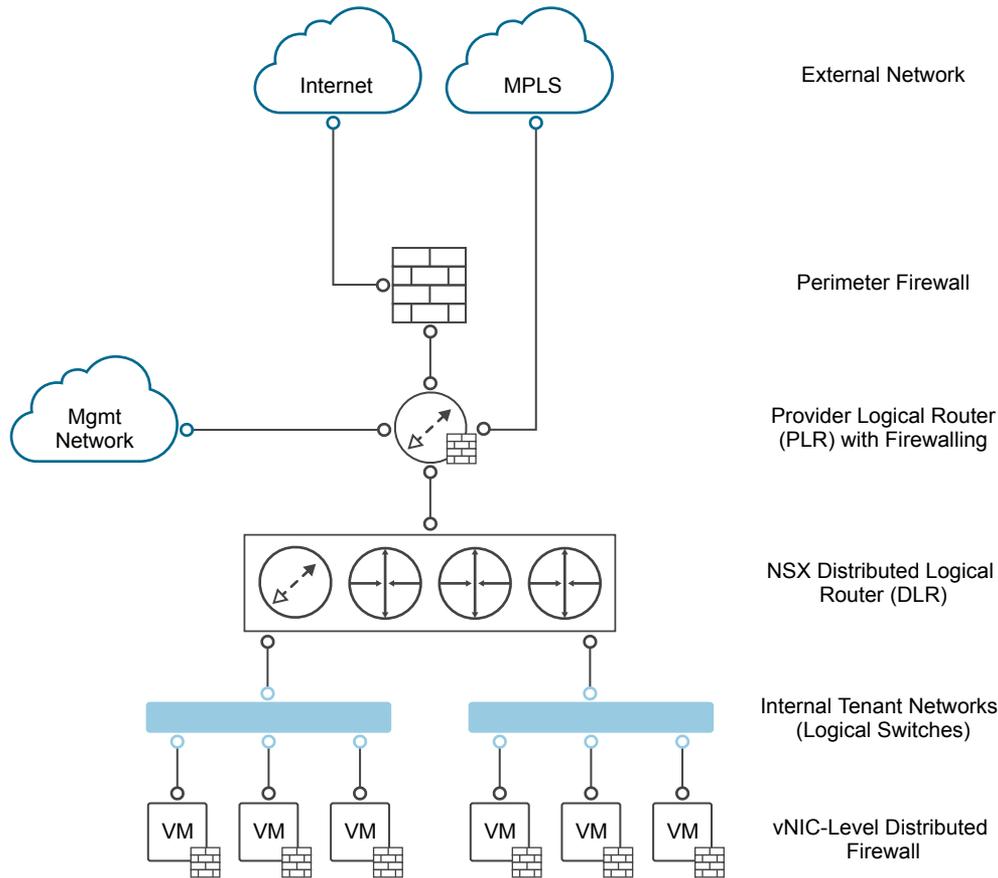
The conceptual design has the following key components.

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 for vSphere 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.

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.

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

Compute/Edge Stack

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

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

Table 2-68. vSphere Cluster Design Decisions

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-007	For the compute stack, do not use a dedicated edge cluster.	Simplifies configuration and minimizes the number of ESXi hosts required for initial deployment.	The NSX Controller instances, NSX Edge services gateways, and DLR control VMs of the compute stack are deployed in the shared edge and compute cluster. The shared nature of the cluster requires the cluster to be scaled out as compute workloads are added so as not to impact 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.

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

Figure 2-16. Cluster Design for NSX for vSphere with a single Availability Zone
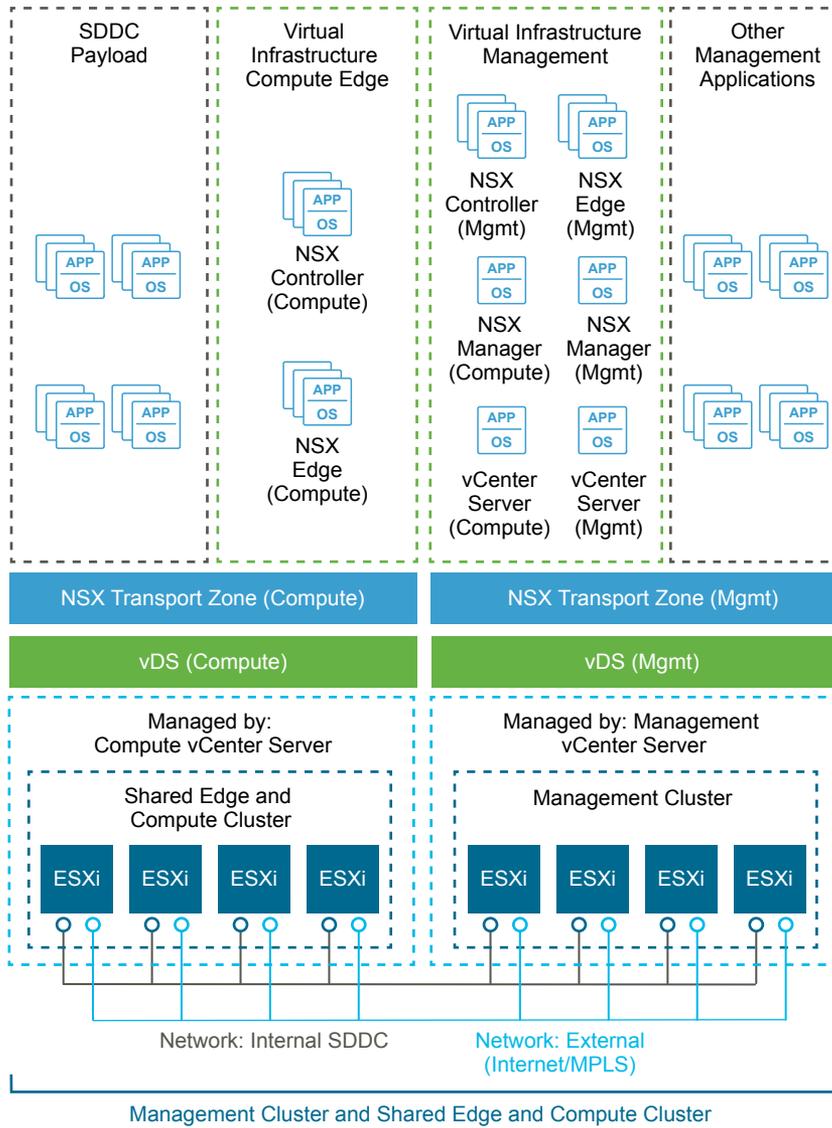
Figure 2-17. Cluster Design for NSX for vSphere with two Availability Zones

Region A - Management and Shared Edge and Compute Clusters

Region B - Management and Shared Edge and Compute Clusters
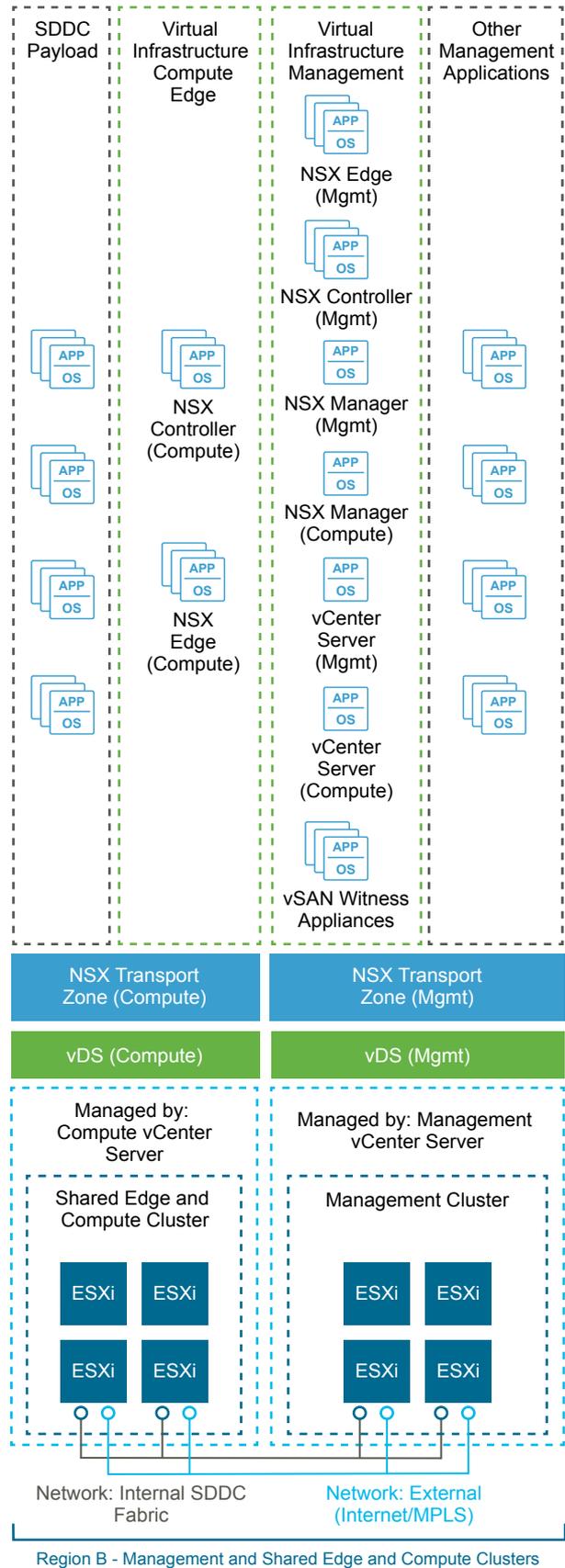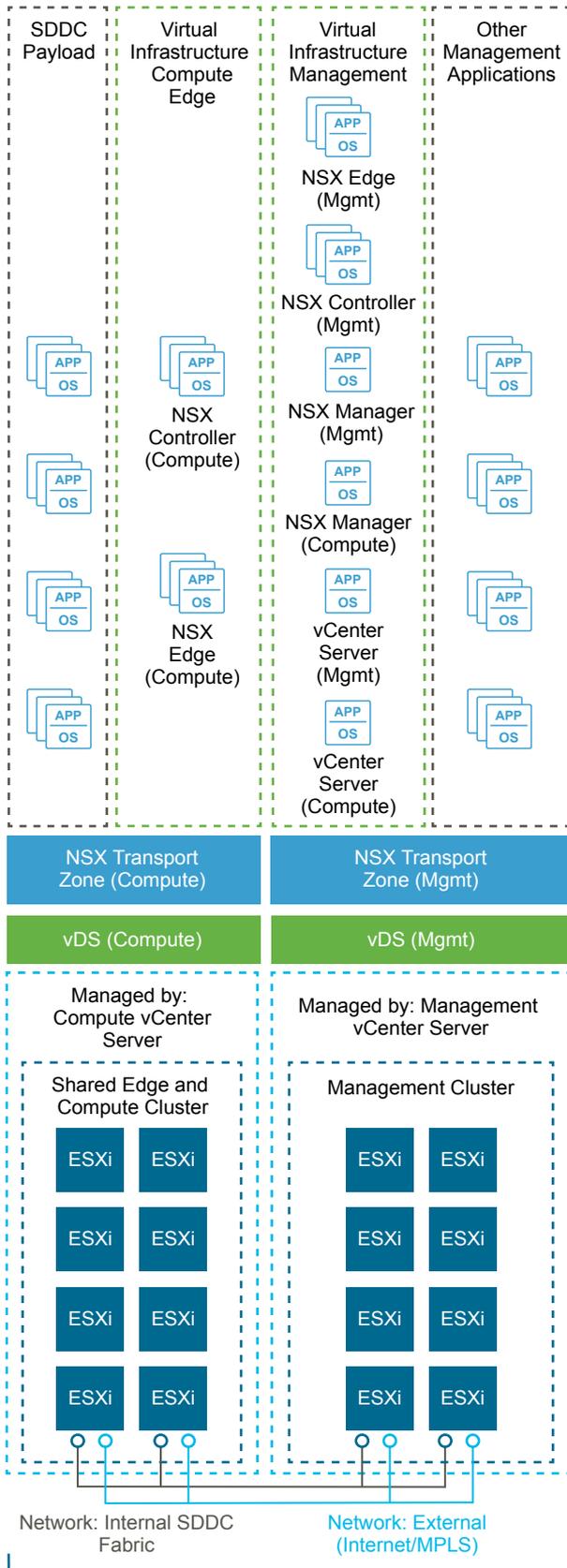
High Availability of NSX for vSphere Components

The NSX Manager instances of both stacks run on the management cluster. vSphere HA protects the NSX Manager instances 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 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 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 deployed for load balancing utilize 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 within 15 seconds. 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 utilizes two physical 10 Gb Ethernet adapters, associated with the uplinks on the vSphere Distributed Switches to which it is connected. Each uplink is connected to a different top-of-rack switch to mitigate the impact of a single top-of-rack switch failure and to provide two paths in and out of the SDDC.

Table 2-69. VTEP Teaming and Failover Configuration Design Decision

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.	Allows for the use of the two uplinks of the vDS resulting in better bandwidth utilization and faster recovery from network path failures.	None.

Logical Switch Control Plane Mode Design

The control plane decouples NSX for vSphere from the physical network and handles the broadcast, unknown unicast, and multicast (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
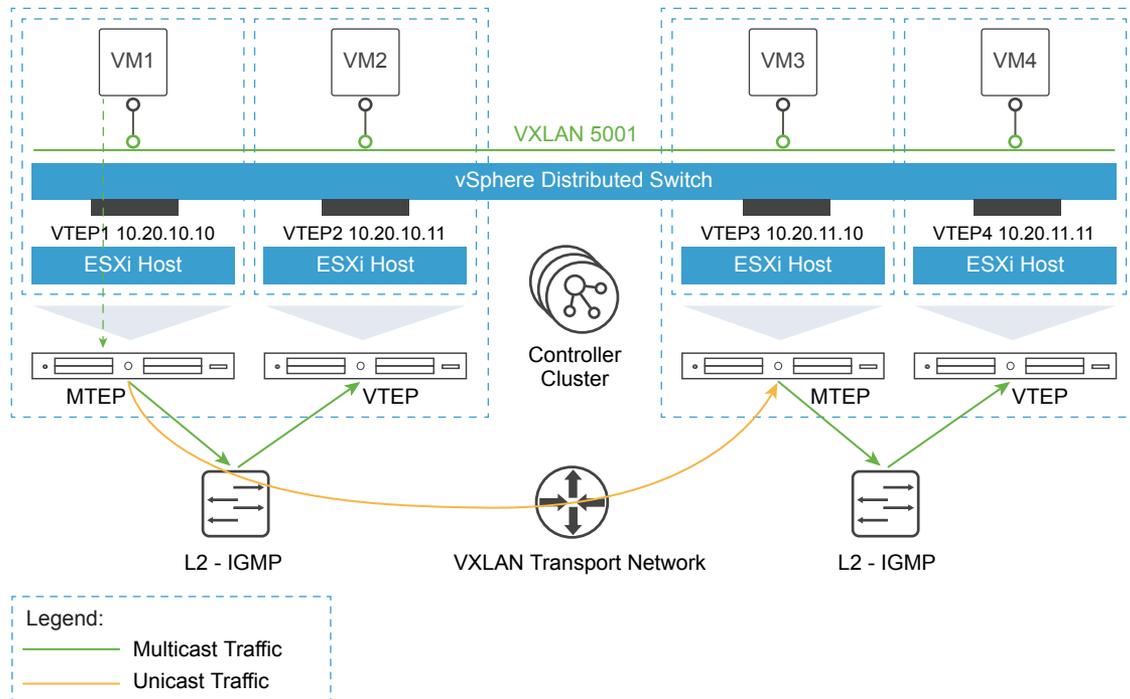
This design uses hybrid mode for control plane replication.

Table 2-70. Logical Switch Control Plane Mode Design Decision

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

Transport Zone Design

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

Table 2-71. Transport Zones Design Decisions

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-012	For the compute stack, use a universal transport zone that encompasses all shared edge and compute, 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. This allows seamless migration of applications across regions either by cross vCenter vMotion or by failover recovery with Site Recovery Manager.	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 managers with a role of secondary cannot deploy Universal objects. To allow 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. This allows 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 manager and new transport zone.

Table 2-71. Transport Zones Design Decisions (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 occurs.	Enabling CDO mode adds some 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 some 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 2-72. Routing Model Design Decisions

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-017	Deploy a minimum of two NSX Edge Services Gateways in an ECMP configuration for north/south routing in both management and shared edge and compute clusters. When using two availability zones deploy a minimum of two NSX Edge Services Gateways in an ECMP configuration into each availability zone.	The NSX ESG is the appropriate device for managing north/south traffic. Using ECMP provides multiple paths in and out of the SDDC. This results in faster failover times than deploying Edge service gateways in HA mode. Due to the upstream physical L3 devices residing in a single availability zone, ECMP edges are required in each availability zone to support north/south routing.	ECMP requires 2 VLANS in each availability zone and region for uplinks which adds an additional VLAN over traditional HA ESG configurations.
SDDC-VI-SDN-018	Deploy a single NSX UDLR for the management cluster to provide east/west routing across all regions.	Using the UDLR reduces the hop count between nodes attached to it to 1. This reduces latency and improves performance.	UDLRs are limited to 1,000 logical interfaces. When that limit is reached, a new UDLR must be deployed.

Table 2-72. Routing Model Design Decisions (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. This reduces latency and improves performance.	UDLRs are limited to 1,000 logical interfaces. When that limit is reached a new UDLR must be deployed.
SDDC-VI-SDN-020	Deploy a 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. This reduces latency and improves performance.	DLRs are limited to 1,000 logical interfaces. When that limit is reached a new DLR must be deployed.
SDDC-VI-SDN-021	Deploy all NSX UDLRs without the local egress option enabled.	When local egress is enabled, control of ingress traffic, is also necessary (for example using NAT). This becomes hard to manage for little to no benefit.	All north/south traffic is routed through Region A until those routes are no longer available. At that time, all traffic dynamically changes to Region B.
SDDC-VI-SDN-022	Use BGP as the dynamic routing protocol inside the SDDC.	Using BGP as opposed to OSPF eases the implementation of dynamic routing. There is no need to plan and design access to OSPF area 0 inside the SDDC. OSPF area 0 varies based on customer configuration.	BGP requires configuring each ESG and UDLR with the remote router that it exchanges routes with.
SDDC-VI-SDN-023	Configure BGP Keep Alive Timer to 1 and Hold Down Timer to 3 between the UDLR and all ESGs that provide north/south routing.	With Keep Alive and Hold Timers between the UDLR and ECMP ESGs set low, a failure is detected quicker, and the routing table is updated faster.	If an ESXi host becomes resource constrained, the ESG running on that ESXi host might no longer be used even though it is still up.
SDDC-VI-SDN-024	Configure BGP Keep Alive Timer to 4 and Hold Down Timer to 12 between the ToR switches and all ESGs providing north/south routing.	This provides a good balance between failure detection between the ToR switches and the ESGs and overburdening the ToRs with keep alive traffic.	By using longer timers to detect when a router is dead, a dead router stays in the routing table longer and continues to send traffic to a dead router.
SDDC-VI-SDN-025	Create one or more static routes on ECMP enabled edges for subnets behind the UDLR and DLR with a higher admin cost than the dynamically learned routes.	When the UDLR or DLR control VM fails over router adjacency is lost and routes from upstream devices such as ToR switches to subnets behind the UDLR are lost.	This requires each ECMP edge device be configured with static routes to the UDLR or DLR. If any new subnets are added behind the UDLR or DLR the routes must be updated on the ECMP edges.

Transit Network and Dynamic Routing

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

Table 2-73. Transit Network Design Decisions

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

Firewall Logical Design

The NSX Distributed Firewall is used to protect all management applications 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 2-74. Firewall Design Decisions

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-029	For all ESGs deployed as load balancers, set the default firewall rule to allow all traffic.	Restricting and granting access is handled by the distributed firewall. The default firewall rule does not have to do it.	Explicit rules to allow access to management applications must be defined in the distributed firewall.
SDDC-VI-SDN-030	For all ESGs deployed as ECMP north/south routers, disable the firewall.	Use of ECMP on the ESGs is 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-031	Configure the Distributed Firewall to limit access to administrative interfaces in the management cluster.	To ensure that only authorized administrators can access the administrative interfaces of management applications.	Maintaining firewall rules adds administrative overhead.

Load Balancer Design

The ESG implements load balancing within NSX for vSphere.

The ESG has both a Layer 4 and a Layer 7 engine that offer different features, which are summarized in the following table.

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
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 2-75. NSX for vSphere Load Balancer Design Decisions

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

Information Security and Access Control

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

Table 2-76. Authorization and Authentication Management Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-SDN-035	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-036	Use global permissions when you create the svc-nsxmanager service account in vCenter Server.	■ Simplifies and standardizes the deployment of the service account across all vCenter Server instances in the same vSphere domain. ■ Provides a consistent authorization layer.	All vCenter Server instances must be in the same vSphere domain.

Bridging Physical Workloads

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

Table 2-77. Virtual to Physical Interface Type Design Decision

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-037	Place all management and tenant virtual machines on VXLAN logical switches, unless you must satisfy an explicit requirement to use VLAN backed port groups for these virtual machines. Where VLAN backed port groups are used, configure routing from VXLAN to VLAN networks. If a Layer 2 adjacency between networks is a technical requirement, then connect VXLAN logical switches to VLAN backed port groups using NSX L2 Bridging.	Use NSX L2 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 L2 Bridging and Distributed Logical Routing are supported on the same VXLAN logical switch.	Network traffic from virtual machines on VXLAN logical switches generally is routed. Where bridging is required, the datapath occurs 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 will vary by customer 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 2-78. Inter-Site Connectivity Design Decisions

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-038	Provide a connection between regions that is capable of routing between each cluster.	When NSX is configured for cross-vCenter to enable universal objects, connectivity between NSX managers, ESXi host VTEPs, and NSX controllers to ESXi hosts management interface is required. To support cross-region authentication, the vCenter Server and Platform Services Controller design requires a single vCenter Single Sign-On domain. Portability of management and compute workloads requires connectivity between regions.	Jumbo frames are required across regions.
SDDC-VI-SDN-039	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 2-79. Isolated Management Applications Design Decisions**

Decision ID	Design Decision	Design Justification	Design Implications
SDDC-VI-SDN-040	Place the following management applications on an application virtual network. ■ vRealize Automation ■ vRealize Automation Proxy Agents ■ vRealize Business ■ vRealize Business collectors ■ vRealize Operations Manager ■ vRealize Operations Manager remote collectors ■ vRealize Log Insight ■ Update Manager Download Service	Access to the management applications is only through published access points.	The application virtual network is fronted by an NSX Edge device for load balancing and the distributed firewall to isolate applications from each other and external users. Direct access to application virtual networks is controlled by distributed firewall rules.
SDDC-VI-SDN-041	Create three application virtual networks. ■ Each region has a dedicated application virtual network for management applications in that region that 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 2-80. Portable Management Applications Design Decision

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

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

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

Figure 2-19. Virtual Application Network Components and Design

Legend:
- Transit Networks
- Compute-Management
- Management Application
- Mgmt-Management
- Ext-Management

Internet/ Enterprise Network

Shared Compute and Edge Cluster

VC OS   PSC OS   SRM OS

ToR Switches

ECMP ESGs

Universal Distributed Logical Router

ESG Load Balancer

192.168.11/24
Mgmt-xRegion01-VXLAN

vROps
vRA
vRB Server

192.168.31/24
Mgmt-RegionA01-VXLAN

vRLI
vROps Collector
vRA Proxy
vRB Collector
UMDS

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 for a management application virtual network.

Figure 2-20 shows an example for implementing a management application virtual network. The example service is vRealize Automation, but any other 3-tier application would look similar.

Figure 2-20. Detailed Example for vRealize Automation Networking
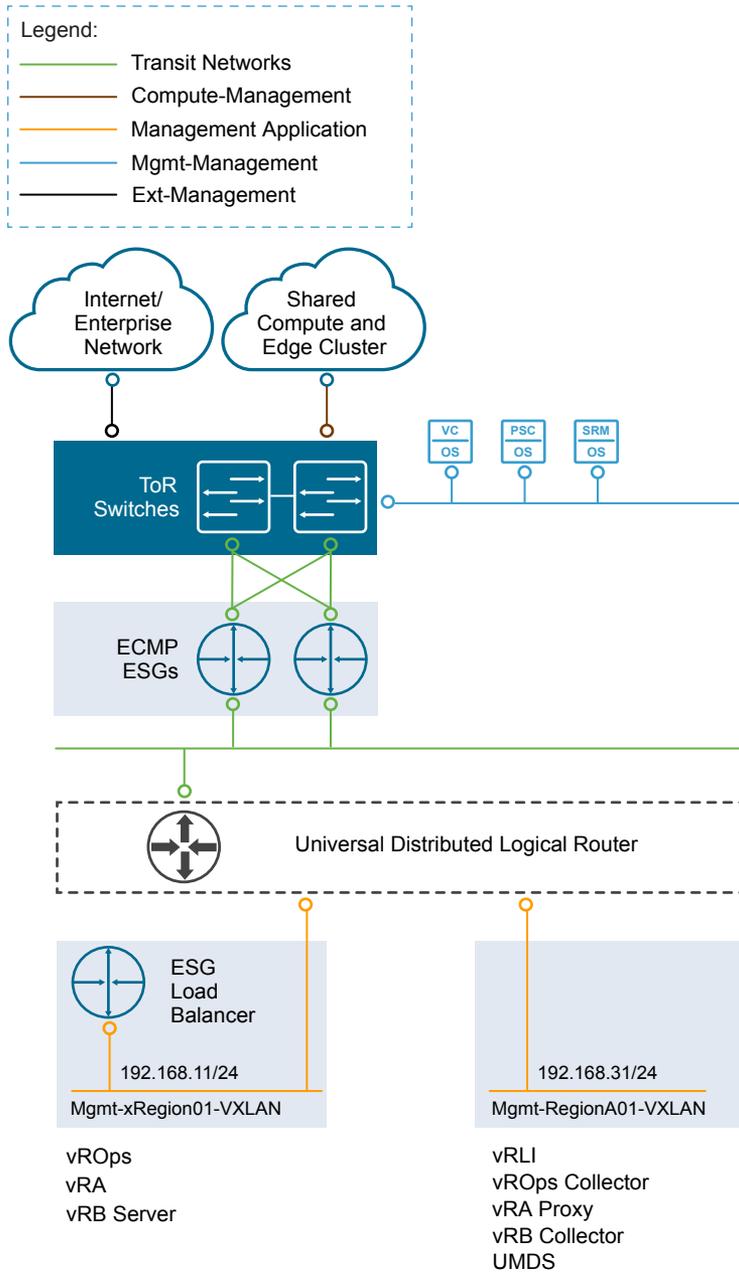
The example is set up as follows.

- 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).

- The network that is used by vRealize Automation connects to external networks through NSX for vSphere. NSX ESGs and the UDLR route traffic between the application virtual networks and the public network.

- Services such as a Web GUI, which must be available to the end users of vRealize Automation, are accessible via the NSX Edge load balancer.

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

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

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

Use of Secure Sockets Layer Certificates

By default, NSX Manager uses a self-signed 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).

Design ID	Design Decision	Design Justification	Design Implication
SDDC-VI-SDN-043	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 design decisions for vSAN storage 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 choose between 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 leverage 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. This solution makes software-defined storage a reality for VMware customers. 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 2-81. 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 2-82. vSphere Features Supported by Storage Type

Type	vSphere vMotion	Datastore	Raw Device Mapping (RDM)	Application or Block-level Clustering	HA/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 (via iSCSI Initiator)	Yes	Yes

Shared Storage Logical Design

The shared storage design selects the appropriate storage device 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**
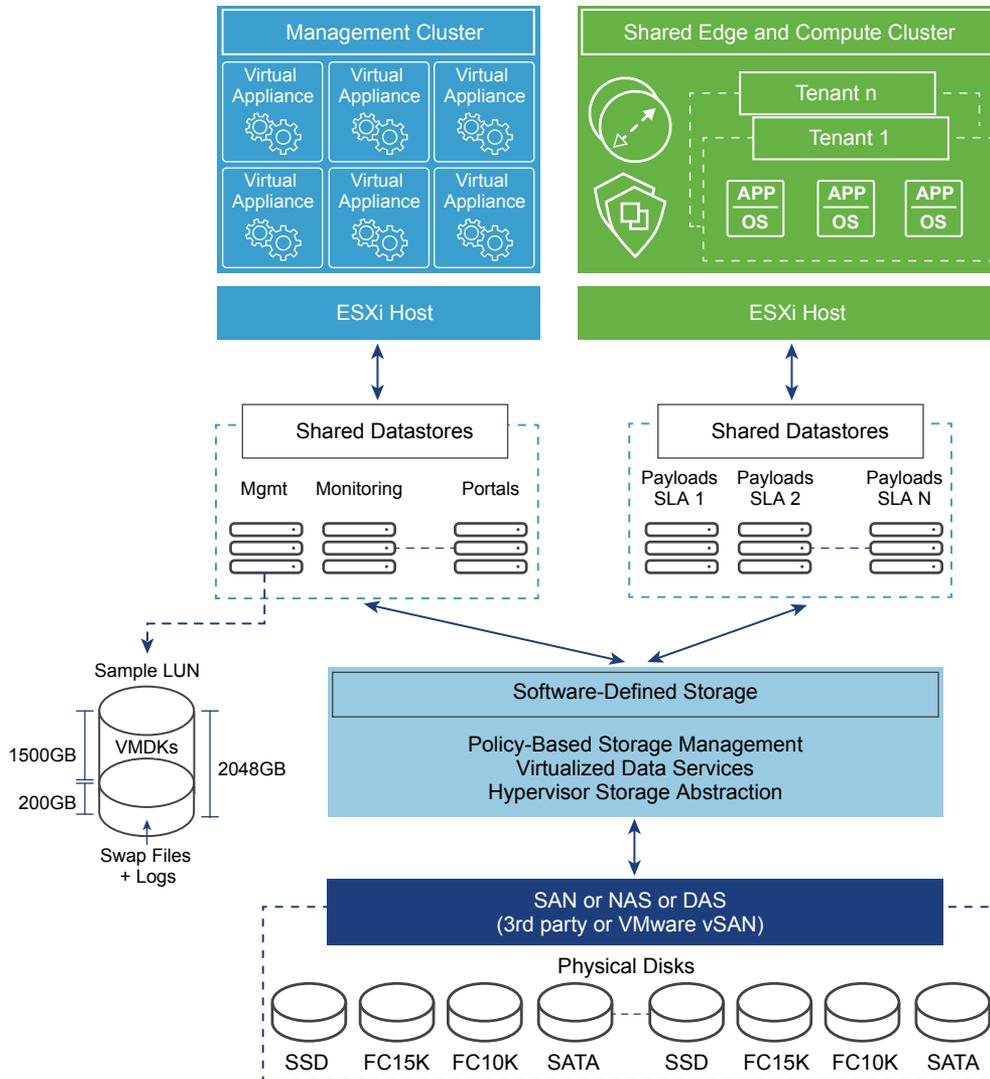
Table 2-83. Storage Type Design Decisions

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.	vSAN as the primary shared storage solution can take advantage of more cost-effective local storage. NFS is used primarily for archival and the need to maintain historical data. Using NFS provides large, low-cost volumes that have the flexibility to be expanded regularly depending on capacity needs.	The use of two different storage technologies increases the complexity and operational overhead. Multiple availability zones cannot be configured 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 the datastore runs out of free space, applications and services within the SDDC, including but not limited to the NSX Edge core network services, the provisioning portal, and VDP backups, fail. To prevent a failure, maintain adequate free space.	Monitoring and capacity management are critical, and must be proactively performed.

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.

VMware Hardware Acceleration API/CLI for Storage

The VMware Hardware Acceleration API/CLI for storage (previously known as vStorage APIs for Array Integration or 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.

In nearly all cases, hardware data movement performs significantly 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 2-84. vStorage APIs for Array Integration Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-003	Select an array that supports 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 VAAI arrays support VAAI over NFS. A plugin from the array vendor is required to enable this functionality.

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 vSAN uses storage policies to allow specification of the characteristics of virtual machines, so you can define the policy on an individual disk level rather than at the volume level for vSAN.

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

VMware vSphere API for Storage Awareness (VASA)	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.

You can associate a virtual machine with a virtual machine storage policy when you create, clone, or migrate that virtual machine. If a virtual machine is associated with a storage policy, the vSphere Web Client shows the datastores that are compatible with the policy. You can select a datastore or datastore cluster. If you select a datastore that does not match the virtual machine storage policy, the vSphere Web Client shows that the virtual machine is using non-compliant storage. See *Creating and Managing vSphere Storage Policies* in the vSphere 6.5 documentation.

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 2-85. Storage I/O Control Design Decisions

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 are throttled to allow other virtual machines access to 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 vStorage 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 in this VMware Validated Design includes conceptual design, logical design, network design, cluster and disk group design, and policy design.

VMware vSAN Conceptual Design and Logical Design

This VMware 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**
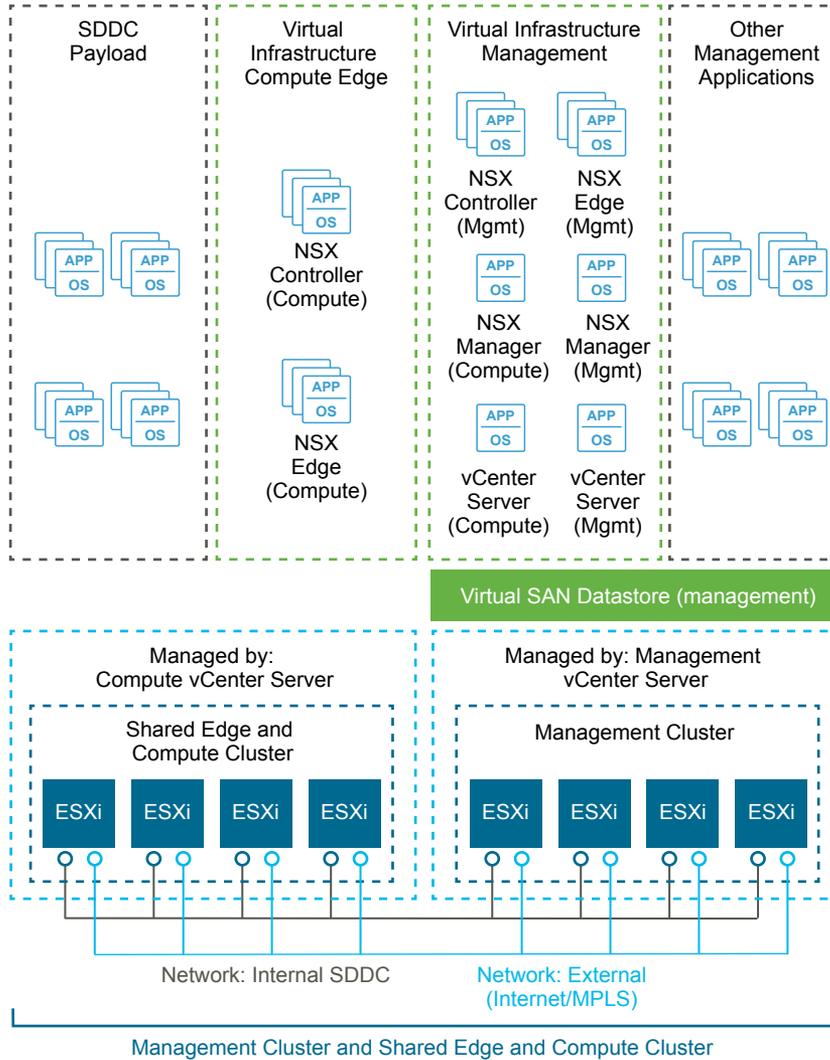
Figure 2-23. Conceptual vSAN Design with two Availiability Zones

Region A - Management and Shared Edge and Compute Clusters

Region B - Management and Shared Edge and Compute Clusters
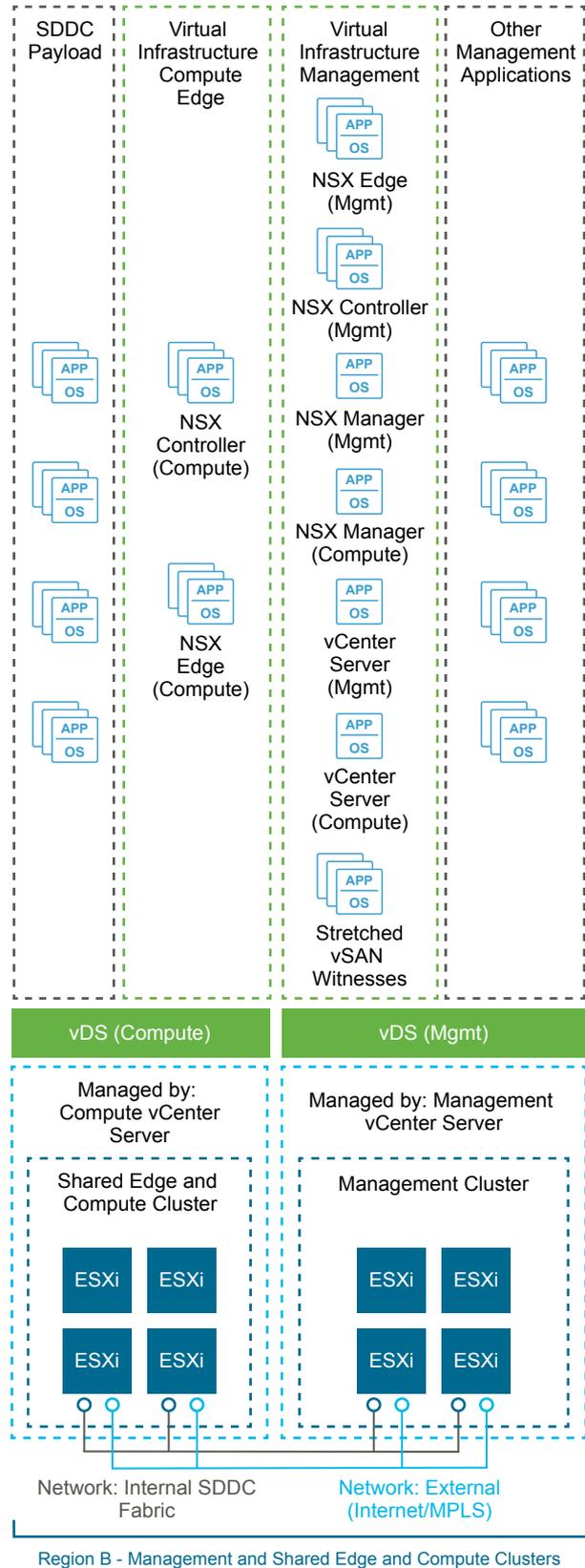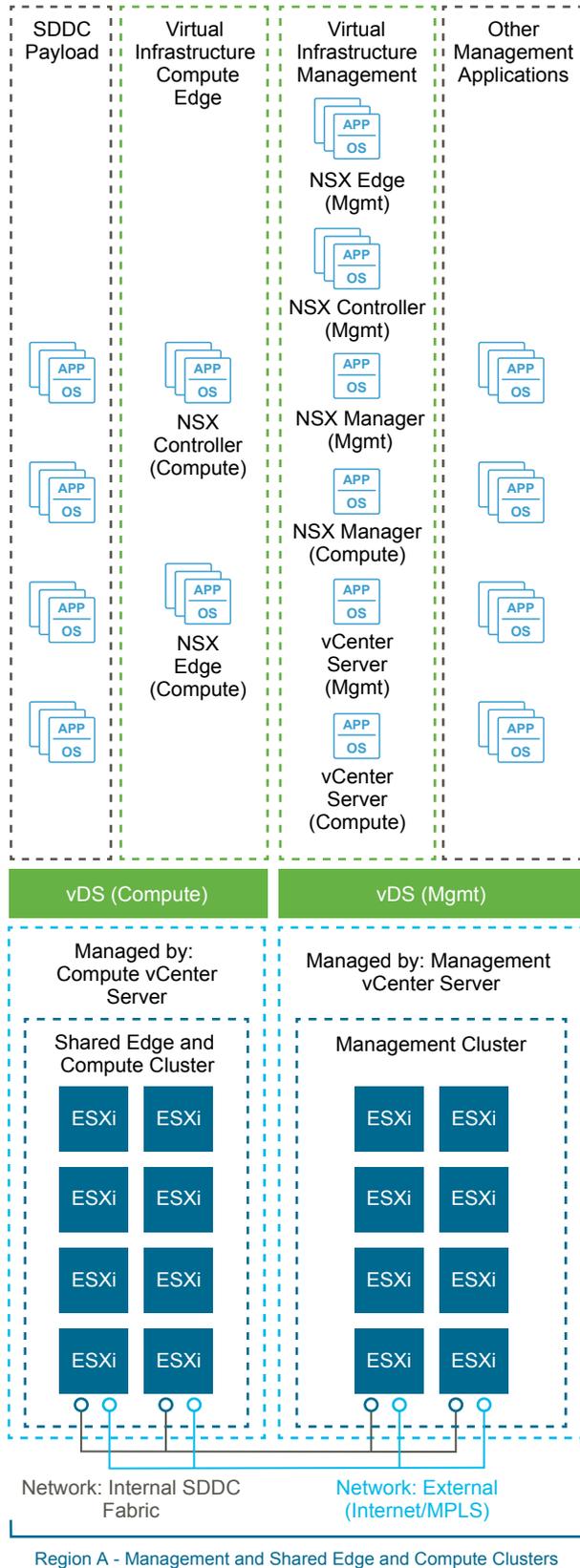
vSAN Logical Design

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.

VMware vSAN Network Design

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

vSAN Network Considerations

- Consider how much replication and communication traffic is running between ESXi hosts. With VMware 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.

- Isolate vSAN traffic on its own Layer 2 network segment. You can do this with dedicated switches or ports, or by using a VLAN.

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.

Availability Zones Network Considerations

When using two availability zones the management VLAN that vCenter Server and other VLAN backed management virtual machines utilize must be stretched across both availability zones. The technology used to stretch the VLAN is out of scope and will vary based on the customers' existing infrastructure.

The connectivity between Availability Zones must support jumbo frames, and ensure that latency is less than 5 ms.

The following diagrams illustrate the logical design of the network.

Figure 2-24. VMware vSAN Conceptual Network with a single Availability Zone
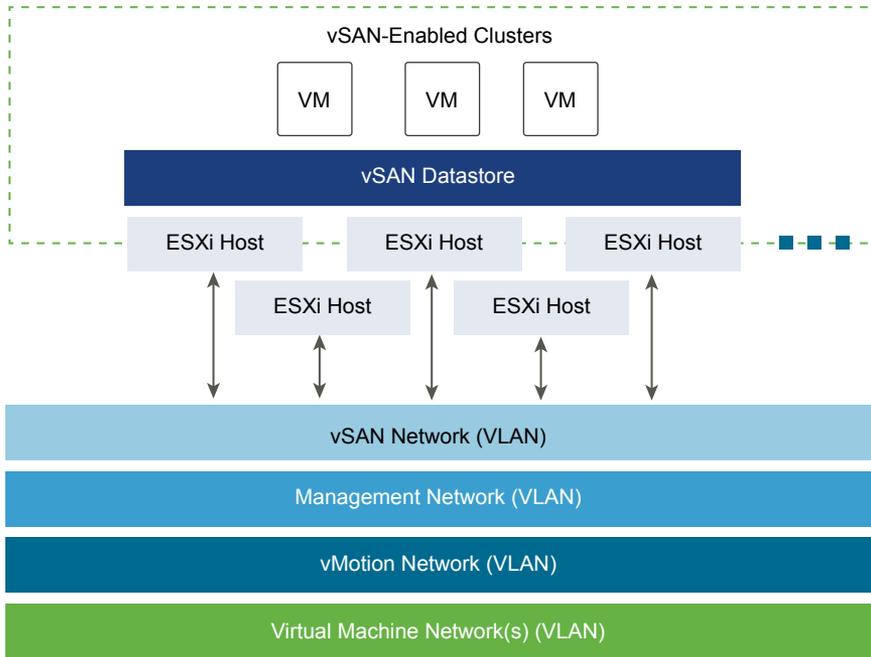
Figure 2-25. **VMware vSAN Conceptual Network with two Availability Zones**
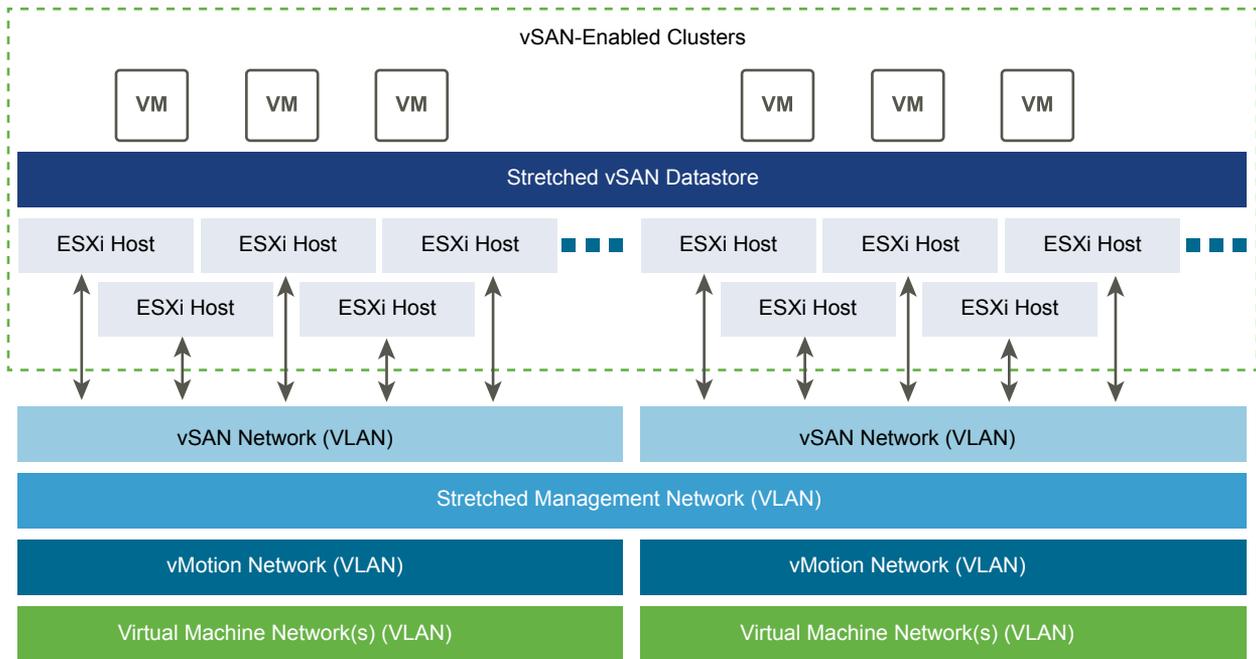
Network Bandwidth Requirements

VMware recommends that solutions use a 10 Gb Ethernet connection for use with vSAN to ensure the best and most predictable performance (IOPS) for the environment. Without it, a significant decrease in array performance results.

Note   vSAN all-flash configurations are supported only with 10 GbE.

Table 2-86. Network Speed Selection

Design Quality	1Gb	10Gb	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. This ensures that 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 2-87. Network Bandwidth Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-001	Use only 10 GbE for VMware vSAN traffic.	Performance with 10 GbE is optimal. Without it, a significant decrease in array performance results.	The physical network must support 10 Gb networking between every ESXi host in the vSAN clusters.

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 allows for prioritization of bandwidth in case of contention in an environment.

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 2-88. 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.

Table 2-88. Virtual Switch Types (Continued)

Design Quality	vSphere Standard Switch	vSphere Distributed Switch	Comments
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 2-89. Virtual Switch Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-002	Use the existing vSphere Distributed Switch instances in the management clusters.	Provide 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

VMware vSAN supports jumbo frames for vSAN traffic.

A VMware 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 2-90. Jumbo Frames Design Decision

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

VMware recommends isolating VMware 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 2-91. VLAN Design Decision

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 ensure traffic isolation.	VLANs span only a single cluster. A sufficient number of VLANs are available within each cluster and are to be used for traffic segregation.
SDDC-VI-Storage-SDS-005	When using two availability zones configure a dedicated VLAN in each availability zone for each vSAN enabled cluster.	VLANs ensure traffic isolation. vSAN traffic between availability zones is routed which means an additional stretched VLAN is not required.	A sufficient number of VLANs are available within each cluster and are to be used for traffic segregation. 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 ESXi hosts on either side of a stretched cluster.

This vSAN witness can be configured as a physical ESXi host or make use of the vSAN Witness Appliance.

Table 2-92. Witness Design Decision

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.

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.

VMware vSAN Datastore Size

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

Table 2-93. VMware vSAN Datastore Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-SDS-007	Provide the Management cluster with a minimum of 8 TB of raw capacity for vSAN.	Management cluster virtual machines that use vSAN require at least 8 TB of raw storage. NFS is used as secondary shared storage of some management components such as 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 2-94. Number of Hosts Per Cluster

Design Quality	3 ESXi Hosts	32 ESXi Hosts	64 ESXi Hosts	Comments
Availability	↓	↑	↑↑	The more ESXi hosts that are available in the cluster, the more failures the cluster can tolerate.
Manageability	↓	↑	↑	The more ESXi hosts in the cluster, the more virtual machines can be 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 2-95. Cluster Size Design Decision

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 underutilization of the cluster 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 clusters 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 2-96. Number of Disk Groups Per ESXi Host

Design Quality	1 Disk Group	3 Disk Groups	5 Disk Groups	Comments
Availability	↓	↑	↑↑	If more ESXi hosts are available in the cluster, the cluster tolerates more failures. This capability adds cost because additional hardware for the disk groups is required.
Manageability	o	o	o	If more ESXi hosts are 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, the vSAN can deliver increased performance and speed.

Table 2-96.   Number of Disk Groups Per ESXi Host (Continued)

Design Quality	1 Disk Group	3 Disk Groups	5 Disk Groups	Comments
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 2-97. Disk Groups Per ESXi Host Design Decision

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.

VMware 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 unless your environment requires policies with non-default behavior. If you configure a custom policy, vSAN will guarantee it. 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 2-98. VMware vSAN Policy Options

Capability	Use Case	Value	Comments
Number of failures to tolerate	Redundancy	Default 1 Max 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	Default 1 Max 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.
Flash read cache reservation (%)	Performance	Default 0 Max 100%	Flash capacity reserved as read cache for the storage is a percentage of the logical object size that will be 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.
Object space reservation (%)	Thick provisioning	Default 0 Max 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	Default: No	Force provisioning allows for 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 2-99. 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 2-99. Object Policy Defaults (Continued)

Object	Policy	Comments
Virtual disk(s)	User-Configured Storage Policy	Can be any storage policy configured on the system.
Virtual disk snapshot(s)	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, the default system policy of 1 failure to tolerate and 1 disk stripe is used for virtual disk(s) and virtual disk snapshot(s). Policy defaults for the VM namespace and swap are set statically and are not configurable to ensure appropriate protection for these critical virtual machine components. Policies must be configured based on the application's business requirements. Policies give VMware vSAN its power because it can adjust how a disk performs on the fly based on the policies configured.

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 2-100. Policy Design Decision

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.	The default vSAN storage policy provides the level of redundancy that is needed within the management cluster. Also, using unique vSAN storage policies are not required as the default policy provides the level of performance that is needed for the individual management components.	Additional policies might be needed if third party VMs are 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, edit the default vSAN Storage Policy settings to include: 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.	Additional policies might be needed if third-party VMs are hosted in these clusters because their performance or availability requirements might differ from what this VMware vSAN policy supports.
SDDC-VI-Storage-SDS-014	Configure the virtual machine swap file as a sparse object on VMware vSAN	Enabling this setting 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. The result can be significantly less space consumed on the vSAN datastore, provided virtual machines do not experience memory over commitment, requiring use of the virtual swap file.	Administrative overhead to enable the advanced setting on all ESXi hosts running VMware vSAN.

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 2-101. NFS Version Design Decision

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. Best practice is to use this feature to determine which ESXi hosts can mount the volumes that are being exported and have read/write access to those volumes. This 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 2-102. NFS Export Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-VI-Storage-NFS-002	Create 1 export to support the vRealize Log Insight Archive management components.	The storage requirements of these management components are separate from the primary storage.	You can add exports if you expand the design.
SDDC-VI-Storage-NFS-003	Place the VADP-based backup export on its own separate volume as per SDDC-PHY-STO-010	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 only the application VMs or hosts requiring the ability to mount the storage.	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.

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

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

Figure 2-26. Operations Management in the SDDC Layered Architecture
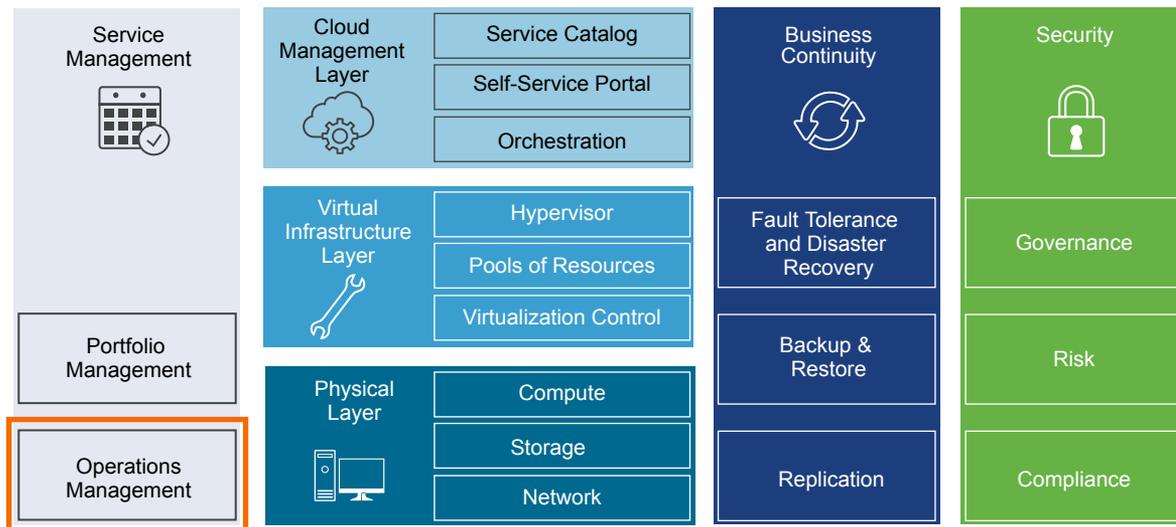
- **vRealize Operations Manager Design**

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

- **vRealize Log Insight Design**

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

- **vSphere Update Manager Design**

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

vRealize Operations Manager Design

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

- Logical and Physical Design of vRealize Operations Manager

  vRealize Operations Manager communicates with all management components in all regions of the SDDC to collect metrics 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 and failover of the vRealize Operations Manager nodes by placing them in several network segments. This networking design also supports public access to the analytics cluster nodes.

- Information Security and Access Control in vRealize Operations Manager

  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 and 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 master, replica and data nodes. The region-specific remote collector nodes remain in the affected region.

Logical and Physical Design of vRealize Operations Manager

vRealize Operations Manager communicates with all management components in all regions of the SDDC to collect metrics 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. For load balancing and fault tolerance, two remote collectors are deployed in each region.

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 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 2-27. Logical Design of vRealize Operations Manager Multi-Region Deployment
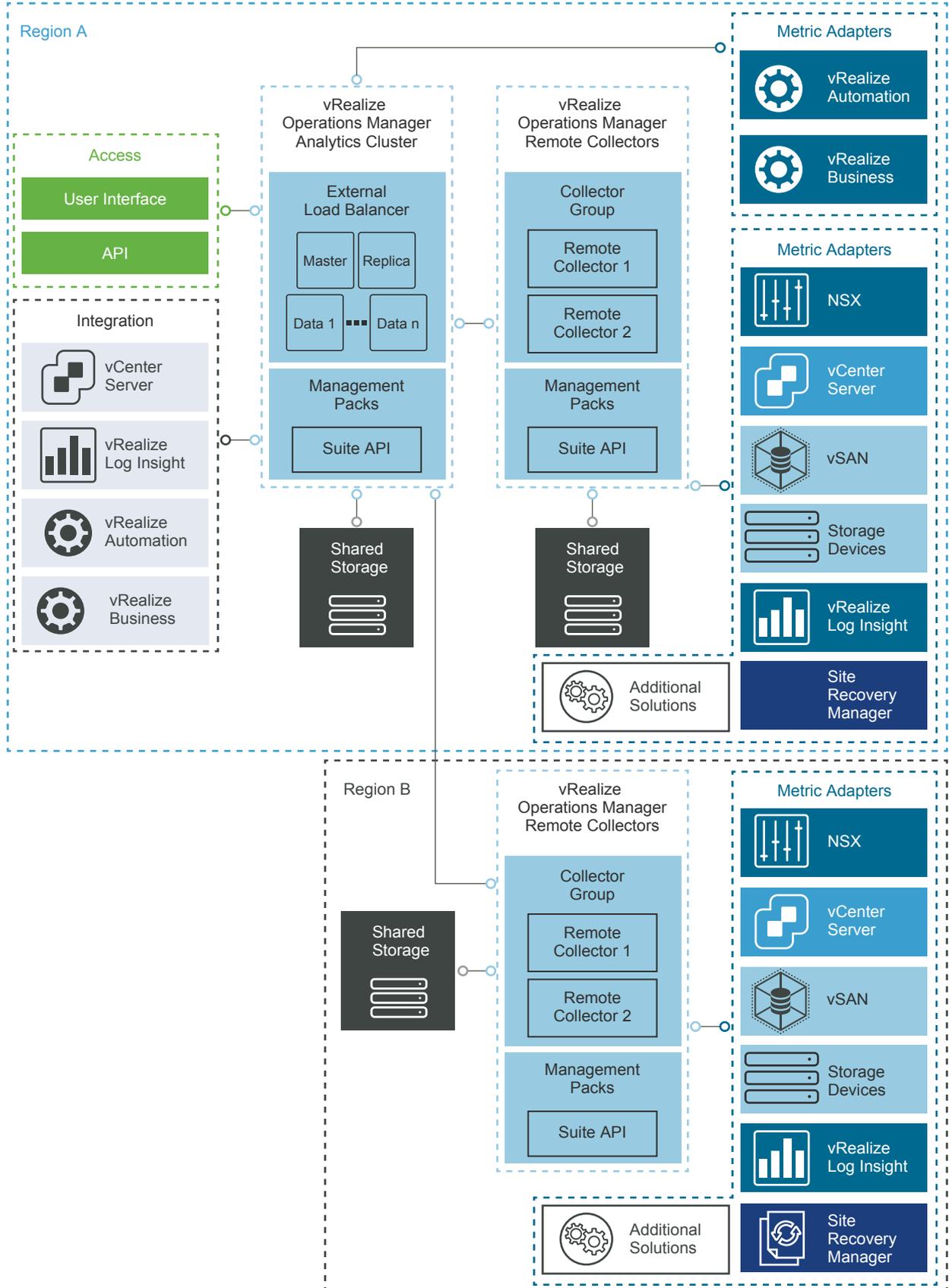
Physical Design

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

Data Sources

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

- Virtual Infrastructure

    - Platform Services Controller instances

    - vCenter Server instances

    - ESXi hosts

    - NSX Manager instances

    - NSX Controller instances

    - NSX Edge instances

    - Shared storage

- vRealize Automation

    - vRealize Automation Appliance

    - vRealize IaaS Web Server

    - vRealize IaaS Management Server

    - vRealize IaaS DEM

    - vRealize vSphere Proxy Agents

    - Microsoft SQL Server

- vRealize Business for Cloud

- vRealize Log Insight

- vRealize Operations Manager

- Site Recovery Manager

Node Configuration of vRealize Operations Manager

The analytics cluster of the vRealize Operations Manager deployment contains the nodes that analyze and store data from the monitored components. 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 master node, one master replica node, and one data node to enable scale out and high availability.

Table 2‑103. Design Decisions About the Node Configuration of vRealize Operations Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-001	Deploy initially vRealize Operations Manager as a cluster of three nodes: one master, one master replica and one data node.	Provides the initial scale capacity required for monitoring up to 1,000 VMs and can be scaled up with additional data nodes.	■ You must size 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 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 the 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 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 the 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. 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, you should deploy more data notes to accommodate the larger expected number of objects and metrics.

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

Table 2-104. 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. Th estimations in the specification table represent the sum of these metrics.

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

(**) The reduction in maximum metrics to permit some head room.

Table 2-105. Design Decisions About 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 a lower number of 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 enabled within the analytics cluster 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 metrics 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 large 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 without violating vSphere DRS anti-affinity rules.

Table 2-105. Design Decisions About the Compute Size of the Analytics Cluster Nodes of vRealize Operations Manager (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
			■ The number of nodes must not exceed `number of ESXi hosts in the management cluster – 1`. 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 2-106. Size of a Standard Remote Collector Virtual Appliance for vRealize Operations Manager

Attribute	Specification
Appliance size	Remote Collector - Standard
vCPU	2
Memory	4 GB
Single-node maximum Objects(*)	1,500
Single-Node Maximum Collected Metrics	600,000
Maximum number of End Point Operations Management Agents per Node	250

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

Table 2-107. Design Decisions About 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 and remote collector clusters. To collect the required number of metrics, you must increase Disk 2 to a 1-TB VMDK on each analytics cluster node.

Sizing Storage for the Analytics Cluster Nodes

The analytics cluster processes a large amount 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 2-108. Design Decision About the Storage Size of the Analytics Cluster of vRealize Operations Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-009	Increase Disk 2 in size to 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 2-109. Design Decision About 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 additional storage for remote collectors.	Remote collectors do not perform analytics operations or store data on disk.	None.

Networking Design of vRealize Operations Manager

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

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

Figure 2-28. Networking Design of the vRealize Operations Manager Deployment
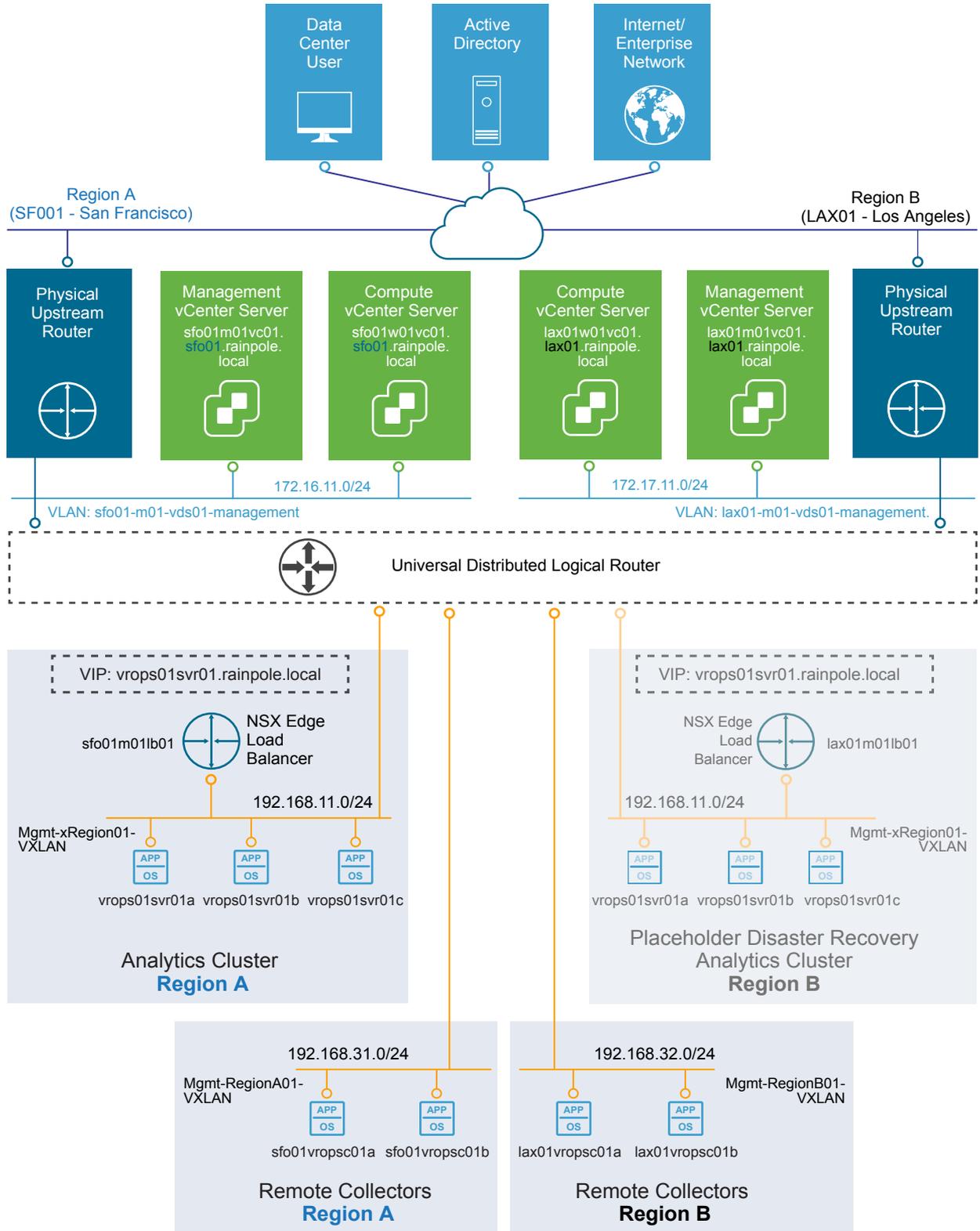
Application Virtual Network Design for vRealize Operations Manager

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 are failed over between regions. vRealize Automation also shares 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 2‑110. Design Decisions About 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 collections of metrics locally per region in the event of a cross-region network outage. Also, it co-localized metric collection to the per-region SDDC 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 2‑111. IP Subnets in the Application Virtual Network for vRealize Operations Manager

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

Table 2‑112. Design Decision About 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.

DNS Names for vRealize Operations Manager

The FQDNs of the vRealize Operations Manager nodes follow 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 a NSX Edge services gateway providing the load balancer function.

- Name resolution for the IP addresses of the remote collector group nodes uses a region-specific suffix, for example, `sfo01.rainpole.local` or `lax01.rainpole.local`.

- The IP addresses of the remote collector group nodes are associated with names whose suffix is set to the region-specific domain, for example, `sfo01.rainpole.local` or `lax01.rainpole.local`.

Table 2‑113. FQDNs for the vRealize Operations Manager Nodes

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

Table 2‑114. Design Decision About 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 an NSX load balancer located in the `Mgmt-xRegion01-VXLAN` application virtual network. This load balancer is configured to allow up to four connections per node. The load balancer must distribute the load evenly to all cluster nodes. In addition, configure the load balancer to redirect service requests from the UI on port 80 to port 443.

Load balancing for the remote collector nodes is not required.

Table 2-115. Design Decisions About 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

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

Import users or user groups from an LDAP database
Users can use their LDAP credentials to log in to vRealize Operations Manager.

Use vCenter Server user accounts
After a vCenter Server instance is registered with vRealize Operations Manager, the following vCenter Server users can log in to vRealize Operations Manager:

- Users 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 2-116. Design Decisions About 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 2‑116. Design Decisions About 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 metrics 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.
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.

Table 2-116. Design Decisions About Authorization and Authentication Management for vRealize Operations Manager (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
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.
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 metrics 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.

Table 2-117. Design Decision About CA-Signed Certificates in vRealize Operations Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-MON-026	Replace the default self-signed certificates with a CA-signed certificate.	Ensures that all communication to the externally facing Web UI is encrypted.	You must contact a certificate authority.

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 and receive alerts about issues that are related to its operational state.

vRealize Operations Manager display 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 2-118. Design Decisions About 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 2-119. 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 2-120. Design Decisions About 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 master, 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 in a dual-region setup.

Logical Design and Data Sources of vRealize Log Insight

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

Logical Design

In a multi-region Software-Defined Data Center (SDDC), 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 2-29. Logical Design of vRealize Log Insight
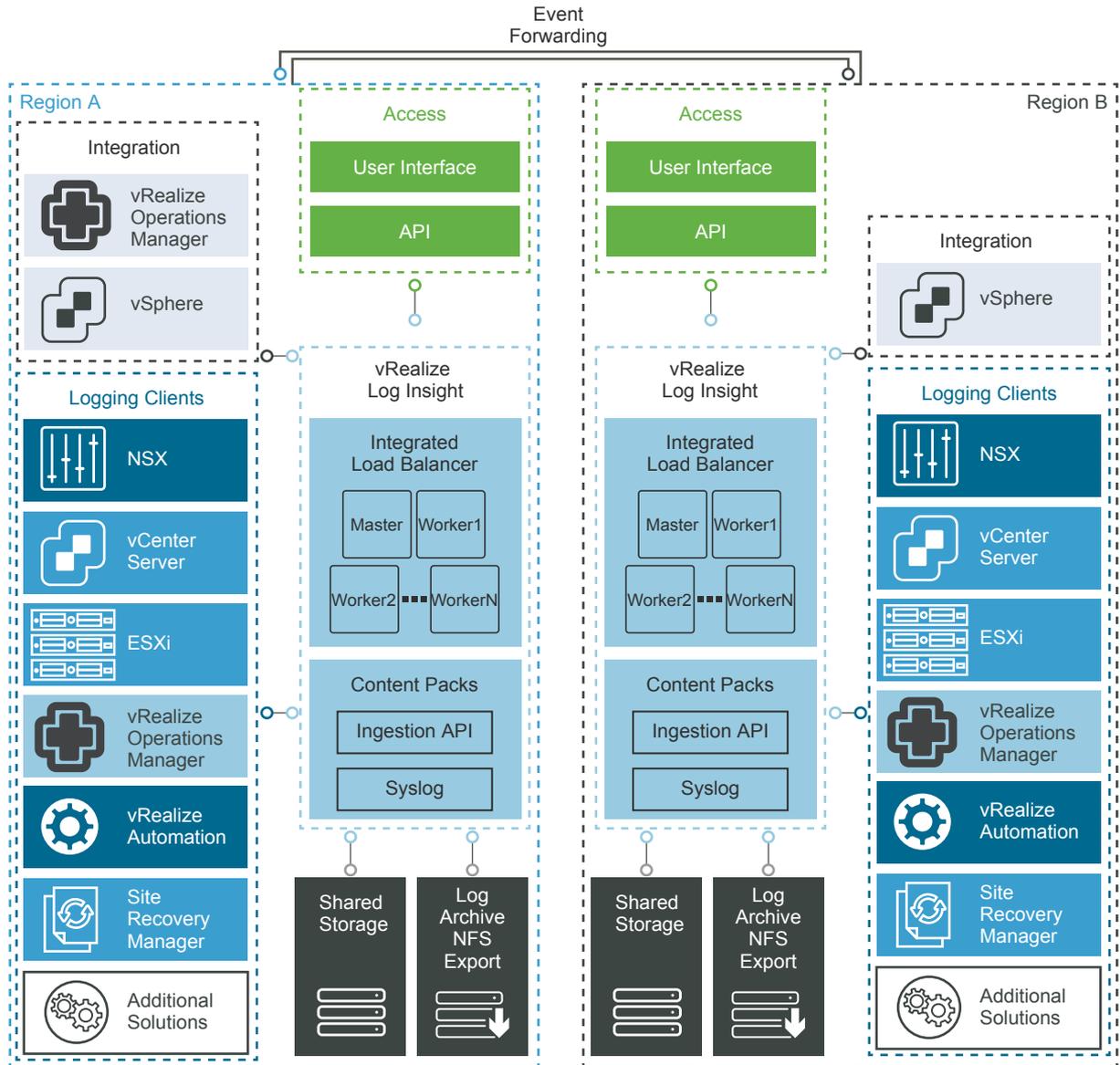
Sources of Log Data

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 Managers

  - 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 Automation

  - vRealize Automation Appliance

  - vRealize IaaS Web Server

  - vRealize IaaS Management Server

  - vRealize IaaS DEM

  - vRealize Agent Servers

  - vRealize Orchestrator (embedded in the vRealize Automation Appliance)

  - Microsoft SQL Server

- vRealize Business

  - vRealize Business server

  - vRealize Business data collectors

- vRealize Operations Manager

  - Analytics cluster nodes

  - Remote 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 master node and two worker nodes behind a load balancer.

You enable the integrated load balancer (ILB) on the 3-node cluster so that all log sources can address the cluster by its ILB. By using the ILB, you 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 via 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 master and 11 worker nodes.

Table 2-121. Design Decisions About Node Configuration for 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 master 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 Log Insight deployment and subsequent integration. - Using the integrated load balancer simplifies the 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 requirements for your vRealize Log Insight cluster expand, 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 2-122. 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 based on IT organization requirements, but this design provides calculations based on a single-region implementation, and is implemented on a per-region basis. This sizing is calculated according to the following node configuration per region:

Table 2-123. Management Systems Whose Log Data Is Stored by vRealize Log Insight

Category	Logging Sources	Quantity
Management cluster	Platform Services Controller	1
	vCenter Server	1
	Site Recovery Manager	1
	ESXi Hosts	4
Shared edge and compute cluster	Platform Services Controller	1
	vCenter Server	1
	ESXi Hosts	64
NSX for vSphere for the management cluster	NSX Manager	1
	NSX Controller Instances	3
	NSX Edge services gateway instances: ■ Two ESGs for north-south routing ■ 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 2-123. Management Systems Whose Log Data Is Stored by vRealize Log Insight (Continued)

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

These components aggregate to approximately 109 syslog and vRealize Log Insight Agent sources per region, or 220 sources with a cross-region configuration. Assuming that you want to retain 7 days of data, apply the following calculation:

vRealize Log Insight receives approximately 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 220 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:

```
220 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. Once provisioned, do not reduce the size or remove virtual disks to avoid data loss.

- 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 your environment out. vRealize Log Insight can scale up to 12 nodes in an HA cluster.

Table 2-124. Design Decisions About 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 Automation components ■ vRealize Business components ■ vRealize Operations Manager components ■ Cross- vRealize Log Insight cluster event forwarding. These sources approximately generate about 220 syslog and vRealize Log Insight Agent sources. Using a 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 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`. Each vRealize Log Insight instance is deployed within the shared management application isolated network.

Figure 2-30. Networking Design for the vRealize Log Insight Deployment
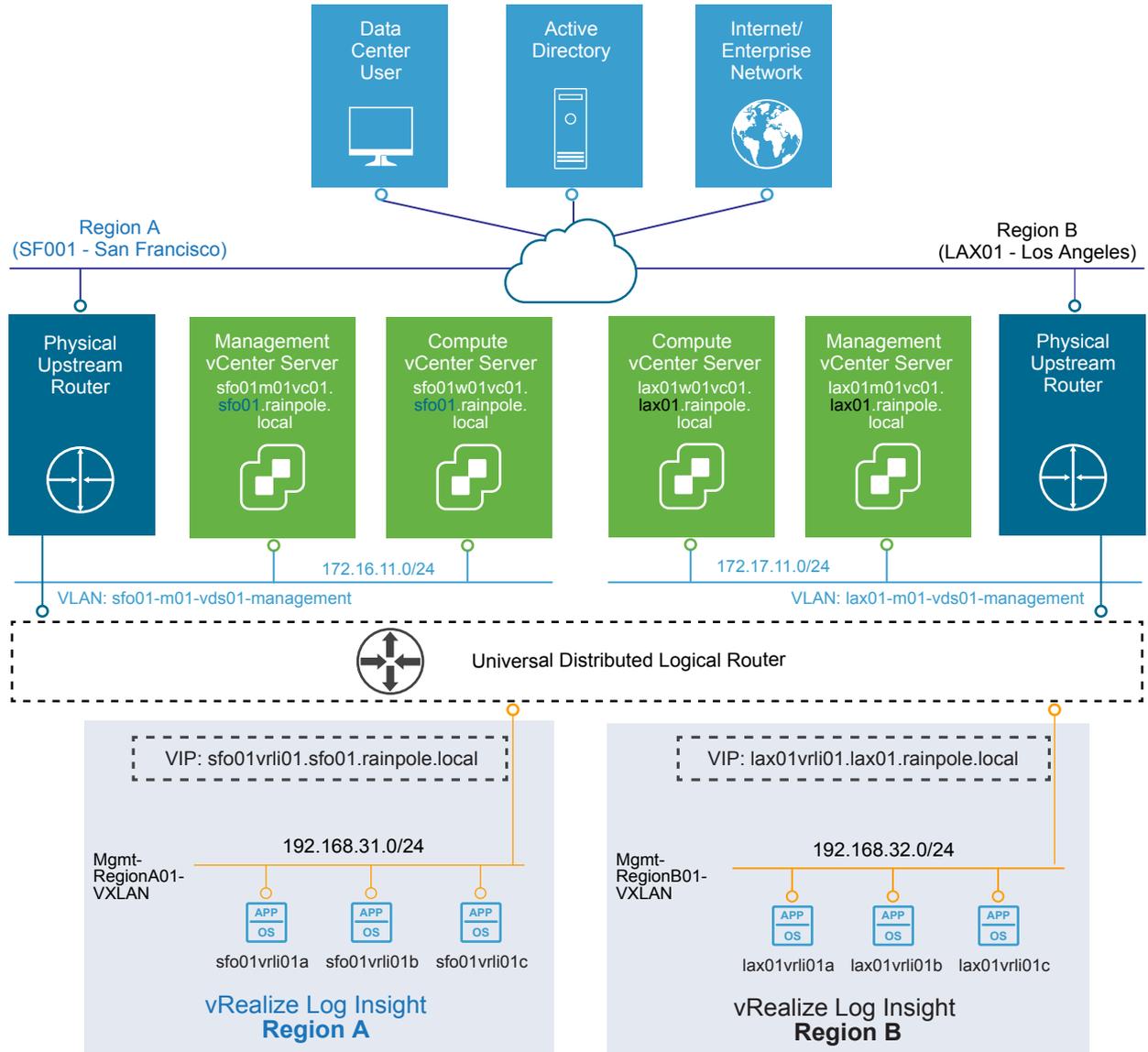
Application Network Design

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 2-125. Design Decision About 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 2-126. IP Subnets in the Application Isolated Networks of vRealize Log Insight

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

DNS Names for vRealize Log Insight

vRealize Log Insight node name resolution, including the load balancer 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 2-127. DNS Names of the vRealize Log Insight Nodes

DNS Name	Role	Region
sfo01vrli01.sfo01.rainpole.local	Log Insight ILB VIP	Region A
sfo01vrli01a.sfo01.rainpole.local	Master node	Region A
sfo01vrli01b.sfo01.rainpole.local	Worker node	Region A
sfo01vrli01c.sfo01.rainpole.local	Worker node	Region A
sfo01vrli01x.sfo01.rainpole.local	Additional worker nodes (not deployed)	Region A
lax01vrli01.lax01.rainpole.local	Log Insight ILB VIP	Region B
lax01vrli01a.lax01.rainpole.local	Master node	Region B
lax01vrli01b.lax01.rainpole.local	Worker node	Region B
lax01vrli01c.lax01.rainpole.local	Worker node	Region B
lax01vrli01x.lax01.rainpole.local	Additional worker nodes (not deployed)	Region B

Table 2-128. Design Decisions About DNS Names 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 vRealize Log Insight Region A VIP when you configure logging.	Support logging when not all management applications are failed over to Region B. For example, only one application is moved to Region B.	If vRealize Automation and vRealize Operations Manager are failed over to Region B and the vRealize Log Insight cluster is no longer available in Region A, update the A record on the child DNS server to point to the vRealize Log Insight cluster in Region B.

Retention and Archiving in vRealize Log Insight

Configure archive and retention parameters of vRealize Log Insight according to the company policy for compliance and governance.

Each vRealize Log Insight virtual appliance has three default virtual disks and can use more virtual disks for storage.

Table 2-129. 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 2-130. Design Decision About Retention Period for 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 2-131. Design Decision About 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 220 logging sources for 90 days.	■ You must manually maintain the vRealize Log Insight archive blobs stored on the NFS store, selectively cleaning the datastore as more space is required.
			■ 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 2-132. Design Decision About 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 2-133. Design Decisions About 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 2-134. Design Decisions About Authorization and Authentication Management for 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

Replace default self-signed certificates with a CA-signed certificate to provide secure access to the vRealize Log Insight Web user interface.

Table 2-135. Design Decision About 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 2-136. Design Decisions About Direct 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 within 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 within the SDDC.	■ Transmission traffic is not secure. ■ Agent presence increases the overall resources used on the system.

Table 2-136. Design Decisions About Direct Log Communication to vRealize Log Insight (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-020	Deploy and configure the vRealize Log Insight agent for the vRealize Automation Windows servers.	- Windows does not natively support syslog. - vRealize Automation requires the use of agents to collect all vRealize Automation logs.	You must manually install and configure the agents on several nodes.
SDDC-OPS-LOG-021	Configure the vRealize Log Insight agent on the vRealize Automation appliance.	Simplifies configuration of log sources within 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-022	Configure the vRealize Log Insight agent for the vRealize Business appliances including: - Server appliance - Data collectors	Simplifies configuration of log sources in the SDDC that are 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 Operation 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-024	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-025	Configure the vCenter Server Appliance and Platform Services Controller instances as direct syslog sources to send log data directly to vRealize Log Insight.	Simplifies configuration for log sources that are syslog-capable.	- You must manually configure syslog sources to forward logs to the vRealize Log Insight VIP. - Certain dashboards in vRealize Log Insight require the use of the vRealize Log Insight agent for proper ingestion. - Not all operating system level events are forwarded to vRealize Log Insight.

Table 2-136. Design Decisions About Direct Log Communication to vRealize Log Insight (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-026	Configure vRealize Log Insight to ingest events, tasks, and alarms from the Management vCenter Server and Compute vCenter Server instances .	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.
SDDC-OPS-LOG-027	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-028	Include the syslog configuration for vRealize Log Insight in the host profile for the following clusters: ■ Management ■ Shared edge and compute ■ Any additional compute	Simplifies the configuration of the hosts in the cluster and 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-029	Deploy and configure the vRealize Log Insight agent for the Site Recovery Manager Windows servers.	■ Windows does not natively support syslog. ■ VMware Site Recovery Manager requires the use of agents to collect all application-specific logs.	You must manually install and configure the agents on several nodes.
SDDC-OPS-LOG-030	Do not configure vRealize Log Insight to automatically update all deployed agents.	Manually install updated versions of the Log Insight Agents for each of the specified components in the SDDC for precise maintenance.	You must maintain manually the vRealize Log Insight Agents on each of the SDDC components.

Time Synchronization in vRealize Log Insight

Time synchronization is critical for the core functionality of vRealize Log Insight. By default, vRealize Log Insight synchronizes time with a 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 2-137. Design Decision About Time Synchronization for vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-031	Configure consistent NTP sources on all virtual infrastructure and cloud management applications for correct log analysis in vRealize Log Insight.	Guarantees accurate log timestamps.	All applications must synchronize time to the same NTP time source.

Content Packs in vRealize Log Insight

The SDDC contains several VMware products for networking, storage, and cloud management. Use content packs to have the logs generated from these components retrieved, extracted and parsed into a human-readable format. In this way, Log Insight saves log queries and alerts, and you can use dashboards for efficient monitoring.

Table 2-138. Design Decisions About Content Packs for vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-032	Install the following content packs: ■ VMware - Linux ■ VMware - NSX-vSphere ■ VMware - Orchestrator 7.0.1 ■ VMware - SRM ■ VMware - vRA 7 ■ Microsoft - SQL Server	Provides additional granular monitoring on the virtual infrastructure. You do not install the following content packs because they are installed by default in vRealize Log Insight: ■ General ■ VMware - vSphere ■ VMware - vSAN ■ VMware - vRops 6.x	Requires manual installation and configuration of each non-default content pack.
SDDC-OPS-LOG-033	Configure the following agent groups that are related to content packs: ■ vRealize Automation (Linux) ■ vRealize Automation (Windows) ■ VMware Virtual Appliances ■ vRealize Operations Manager ■ vRealize Orchestrator ■ VMware Site Recover Manager ■ Microsoft SQL Server	■ Provides a standardized configuration that is pushed to the all vRealize Log Insight Agents in each of the groups. ■ Supports application-contextualized collection 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. While forwarding events, the vRealize Log Insight instance still ingests, stores, and archives events locally.

You forward syslog data in vRealize Log Insight by using the Ingestion API or a native syslog implementation.

The vRealize Log Insight Ingestion API uses TCP communication. In contrast to syslog, the forwarding module supports the following features for the Ingestion API:

- Forwarding to other vRealize Log Insight instances

- Both structured and unstructured data, that is, 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 2-139. Design Decisions About Event Forwarding Across Regions in vRealize Log Insight

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-LOG-034	Forward log events to the other region by using the Ingestion API.	Using the forwarding protocol supports the following operations: ■ Structured and unstructured data for client-side compression ■ Event throttling from one vRealize Log Insight cluster to the other. Forwarding ensures that during a disaster recovery situation the administrator has access to all logs from the two regions although one region is offline.	■ You must configure each region to forward log data to the other. The configuration requires administrative overhead to prevent recursion of logging between regions using inclusion and exclusion tagging. ■ Log forwarding adds more load on each region. You must consider log forwarding in the sizing calculations for the vRealize Log Insight cluster in each region. ■ You must configure identical size on both source and destination clusters.
SDDC-OPS-LOG-035	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-036	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.

vSphere Update Manager Design

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

vSphere Update Manager can remediate the following objects over the network:

- VMware Tools and VMware virtual machine hardware upgrade operations for virtual machines

- ESXi host patching operations

- ESXi host upgrade operations

- Physical Design of vSphere Update Manager

  You use the vSphere Update Manager service on each vCenter Server Appliance and deploy a vSphere Update Manager Download Service (UMDS) in Region A and Region B to download and stage upgrade and patch data.

- Logical Design of vSphere Update Manager

  You configure vSphere Update Manager to apply updates on the management components of the SDDC according to the objectives of this design.

Physical Design of vSphere Update Manager

You use the vSphere Update Manager service on each vCenter Server Appliance and deploy a vSphere 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 2-31. vSphere Update Manager Logical and Networking Design
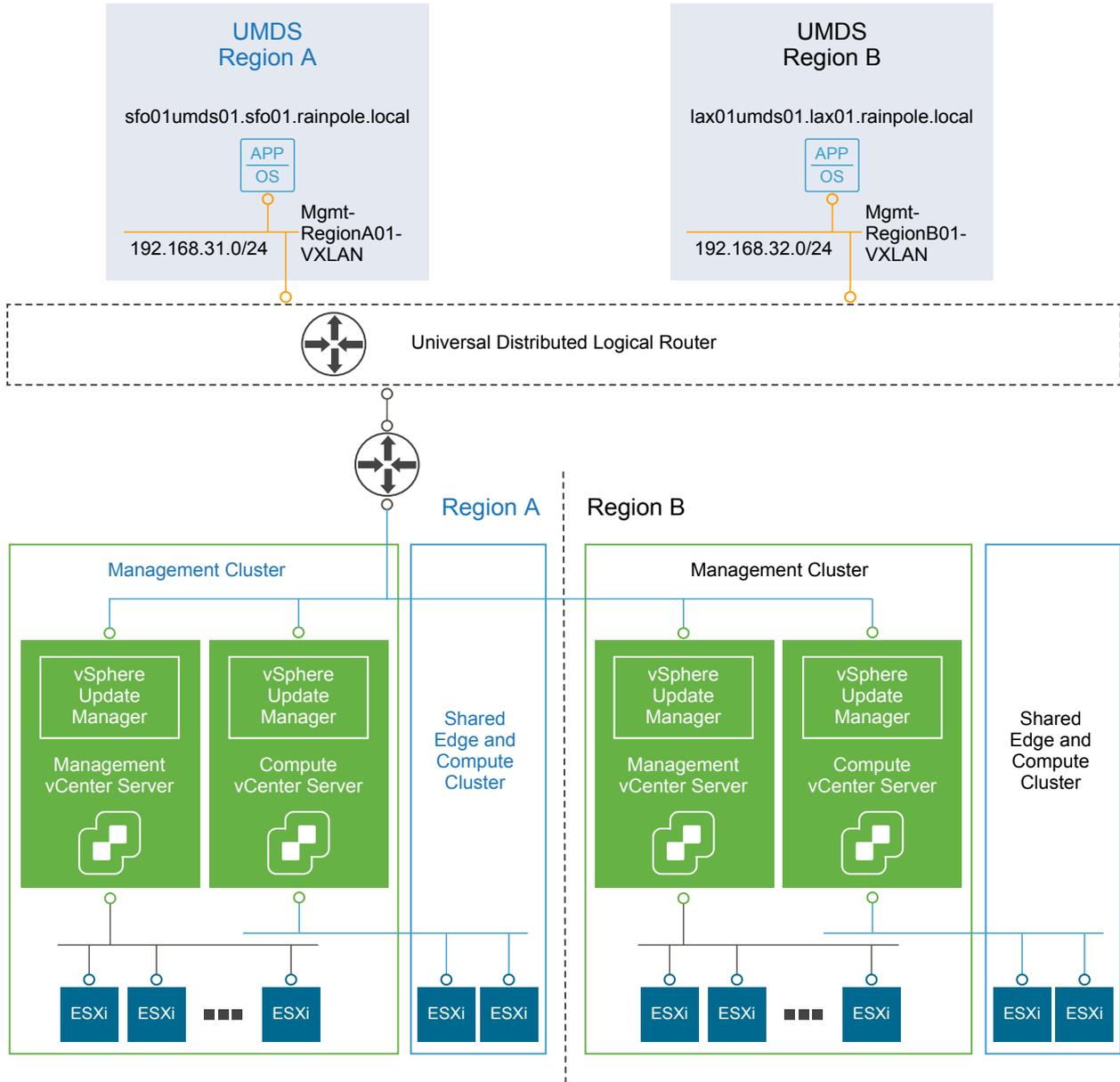
Deployment Model

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 2‑140. Update Manager Physical Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-001	Use the vSphere Update Manager service on each vCenter Server Appliance to provide a total of four vSphere Update Manager instances that you configure and use for patch management.	■ Reduces the number of management virtual machines that need to be deployed and maintained within the SDDC. ■ Enables centralized, automated patch and version management for VMware vSphere and offers support for VMware ESXi hosts, virtual machines, and virtual appliances managed by each vCenter Server.	■ All physical design decisions for vCenter Server determine the setup for vSphere Update Manager. ■ A one-to-one mapping of vCenter Server to vSphere Update Manager is required. Each Management vCenter Server or Compute vCenter Server instance in each region needs its own vSphere Update Manager.
SDDC-OPS-VUM-002	Use the embedded PostgreSQL of the vCenter Server Appliance for vSphere Update Manager.	■ Reduces both overhead and licensing cost for external enterprise database systems. ■ Avoids problems with upgrades.	The vCenter Server Appliance has limited database management tools for database administrators.
SDDC-OPS-VUM-003	Use the network settings of the vCenter Server Appliance for vSphere Update Manager.	Simplifies network configuration because of the one-to-one mapping between vCenter Server and vSphere Update Manager. You configure the network settings once for both vCenter Server and vSphere Update Manager.	None.

Table 2‑140. Update Manager Physical Design Decision (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-004	Deploy and configure UMDS virtual machines for each region.	Limits direct access to the Internet from vSphere Update Manager on multiple vCenter Server instances, and reduces storage requirements on each instance.	You must maintain the host operating system (OS) and the database 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 vSphere Update Manager repository data. ■ Avoids cross-region bandwidth usage for repository access. ■ Provides a consistent deployment model for management applications.	You must use NSX to support this network configuration.

Logical Design of vSphere Update Manager

You configure vSphere Update Manager to apply updates on the management components of the SDDC according to the objectives of this design.

UMDS Virtual Machine Specification

You allocate resources to and configure the virtual machines for UMDS according to the following specification:

Table 2‑141. UMDS Virtual Machine Specifications

Attribute	Specification
vSphere Update Manager Download Service	vSphere 6.5
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 2‑142. 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 other ESXi hosts within 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, set a vSAN policy where the number failures to tolerate is 1, as is the default case.

You can control the update operation by using a set of host and cluster settings in vSphere Update Manager.

Table 2‑143. 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 does not remediate clusters with active DPM, HA, and FT.
	Enable parallel remediation of hosts	vSphere Update Manager can remediate multiple hosts.
		Note Parallel remediation is not supported if you use vSAN, and 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 2-144. 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, the default baselines might not be allowed until patches are tested or verified on development or pre-production hosts. Baselines can be confirmed so that the tested patches are applied to hosts and only updated when appropriate.

Table 2-145. Baseline and Baseline Group Details

Baseline or Baseline Group Feature		Description
Baselines	Types	Four types of baselines exist:
		▪ 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 for dynamic selection of new items.
		▪ Critical host patches - Upgrades hosts with a collection of critical patches that are high priority as defined by 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 Tools version to match the host version.
		▪ VA Upgrade to Latest - Upgrades a virtual appliance to the latest version available.

Table 2-145. Baseline and Baseline Group Details (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. The image can then be upload to vSphere Update Manager.

There are two ways in which you can add packages to an ESXi image:

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 non-conflicting. Performed 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   Combine the extension baseline with the ESXi upgrade baseline in a baseline group so that you can use the hosts being upgraded in a software-defined networking setup.

vSphere Update Manager Logical Design Decisions

This design applies the following decisions on the logical design of vSphere Update Manager and update policy:

Table 2-146. vSphere Update Manager Logical Design Decisions

Design ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-006	Use the default patch repositories by VMware.	Simplifies the configuration because you do not configure additional sources.	None.
SDDC-OPS-VUM-007	Set the VM power state to Do Not Power Off.	Ensures highest uptime of management components and compute 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.	More resources unavailable 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 per business guidelines.	Aligns the remediation schedule with the business policies.	None.
SDDC-OPS-VUM-013	Use a baseline group to add NSX for vSphere software packages to the ESXi upgrade image.	■ Allows for parallel 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 for vSphere updates require periodic updates to Group Baseline.

**Table 2-146. vSphere Update Manager Logical Design Decisions (Continued)**

Design ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-VUM-014	Configure an HTTP Web server on each UMDS service that the connected vSphere Update Manager servers must use to download the patches from.	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 vSphere Update Manager integration with vSAN.	Enables the integration of vSphere Update Manager with the vSAN Hardware Compatibility List (HCL) for additional precision and optimization when patching ESXi hosts within a specific vSphere release that manage 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 myvmware.com.

Cloud Management Platform Design

The Cloud Management Platform (CMP) layer is the management component of the SDDC. The CMP layer allows you to deliver tenants with automated workload provisioning by using a self-service portal.

The CMP layer includes the following components and functionality.

Figure 2-32. The Cloud Management Platform Layer Within the Software-Defined Data Center
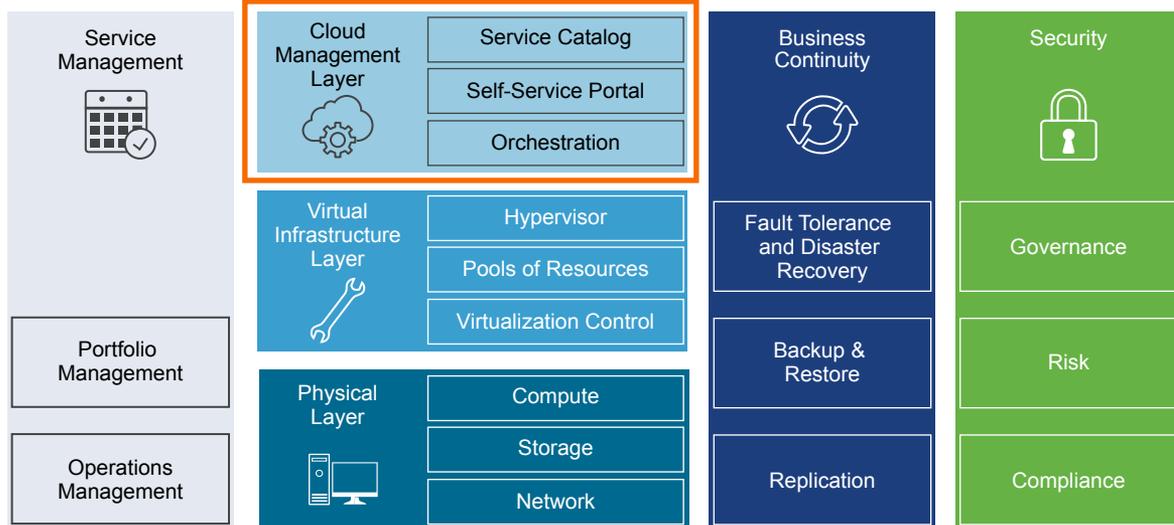
Service Catalog	A self-service portal where users can browse and request the IT services and resources they need, such a virtual machine or a machine on Amazon Web Services (AWS). When you request a service catalog item you provision the item to the designated cloud environment.
Self-Service Portal	Provides a unified interface for consuming IT services. Users can browse the service catalog to request IT services and resources, track their requests, and manage their provisioned items.
Orchestration	Provides automated workflows used to deploy service catalog items requested by users. You use the workflows to create and run automated, configurable processes to manage your SDDC infrastructure, as well as other VMware and third-party technologies.
	vRealize Automation provides the self-service portal and the service catalog. Orchestration is enabled by an instance of vRealize Orchestrator internal to vRealize Automation.

vRealize Automation Design

VMware vRealize Automation provides a service catalog from which tenants can deploy applications, and a portal that lets you deliver a personalized, self-service experience to end users.

vRealize Automation Logical Design

vRealize Automation provides several extensibility options designed to support a variety of 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.

The following diagram illustrates vRealize Automation internal components, and the integration of those components with other external components and the supporting infrastructure of the SDDC.

Figure 2-33. vRealize Automation Logical Architecture, Extensibility, and External Integrations
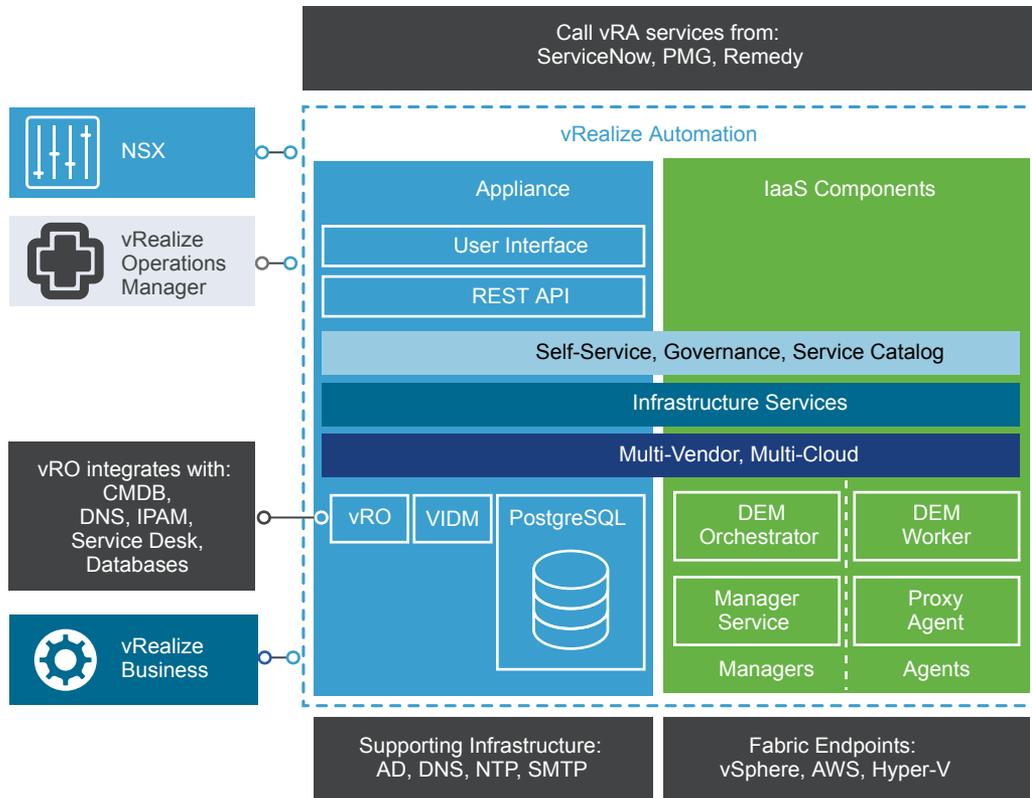
Fabric Endpoints	vRealize Automation can leverage existing and future infrastructure that represent multi-vendor, multi-cloud virtual, physical, and public cloud infrastructures. Each kind of infrastructure supported will be represented by a fabric endpoint.
Call vRealize Automation Services from Existing 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 for Cloud	vRealize Business for Cloud is tightly integrated with vRealize Automation to manage the vRealize automation resource costs by displaying costing information during workload request and on an ongoing basis with cost reporting by user, business group, or tenant. vRealize Business for Cloud

supports pricing based on blueprints, endpoints, reservations and reservation policies for Compute Grouping Strategy. In addition, vRealize Business for Cloud supports the storage path and storage reservation policies for Storage Grouping Strategy.

vRealize Operations Management

The vRealize Automation management pack for vRealize Operation Manager provides the comprehensive visibility into both performance and capacity metrics of a vRealize Automation tenant's business groups and underlying cloud infrastructure. By combining these new metrics with the custom dashboard capabilities of vRealize Operations, you gain a great level of flexibility and insight when monitoring these complex environments.

Supporting Infrastructure

vRealize Automation integrates with the following supporting infrastructure:

- Microsoft SQL Server to store data relating to the vRealize Automation IaaS elements.

- NTP server with which to synchronize the time between the vRealize Automation components

- Active Directory supports vRealize Automation tenant user authentication and authorization.

- SMTP sends and receives notification emails for various actions that can be executed within the vRealize Automation console.

NSX

NSX and vRealize Automation integration provides several options for designing and authoring blueprints with the networking and security features provided by NSX, and takes full advantage of all the NSX network constructs such as switches, routers, and firewalls. This integration allows you to use an on-demand load balancer, on-demand NAT network, on-demand routed network and on-demand security groups within a blueprint, which is automatically provisioned by vRealize Automation when the blueprint is requested. The integration with NSX eliminates the need for networking to be provisioned as a separate activity outside vRealize Automation.

Cloud Management Platform Usage Model

The Cloud Management Platform (CMP), of which vRealize Automation is a central component, enables a usage model that includes interaction between users, the CMP itself, the supporting infrastructure, and the provisioning infrastructure. The following diagram illustrates the usage model of the CMP in relation to these elements.

Figure 2-34. vRealize Automation Usage Model
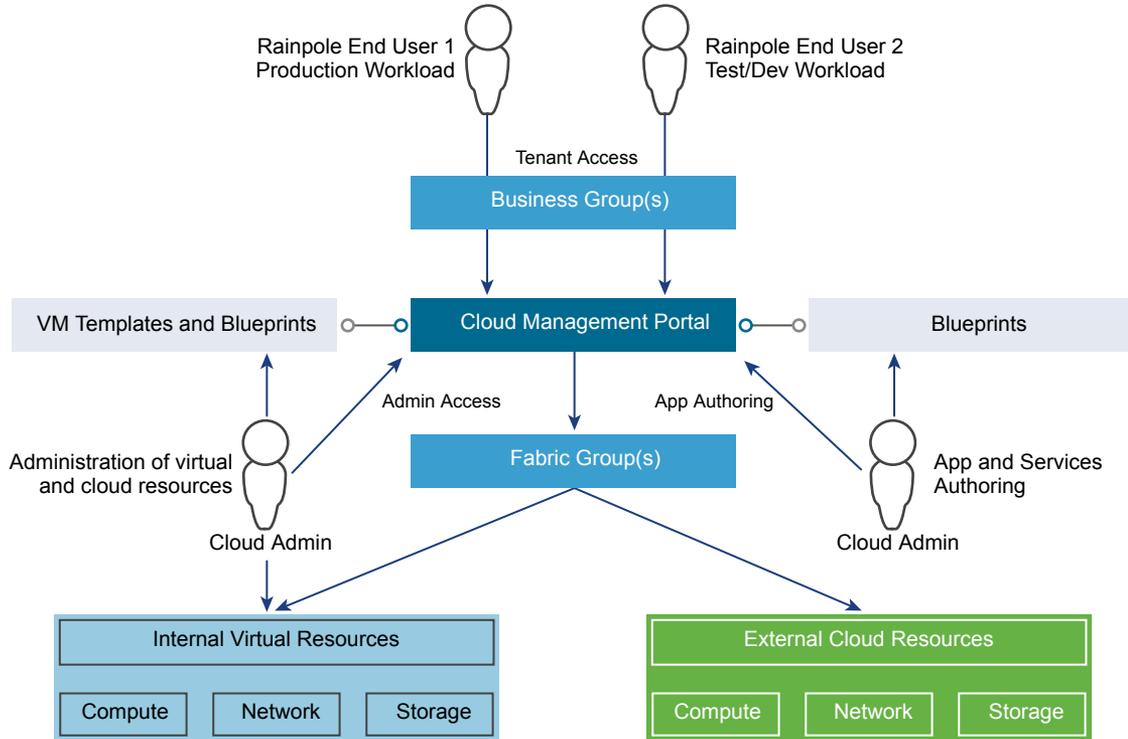
The following table lists the vRealize Automation elements, and the components that in turn comprise each of these elements.

Element	Components	
Users	**Cloud administrators**	Tenant, group, fabric, infrastructure, service, and other administrators as defined by business policies and organizational structure.
	Cloud (or tenant) users	Users within 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 are the building blocks that provide the foundation of the cloud. VM templates are used to author the blueprints that tenants (end users) use to provision their 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	A portal that provides self-service capabilities for users to administer, provision, and manage workloads.	
	Realize Automation portal, Admin access.	The default root tenant portal URL used to set-up and administer tenants and global configuration options.
	vRealize Automation portal, Tenant access.	Refers to a subtenant which is accessed using an appended tenant identifier.
	Attention A tenant portal might refer to the default tenant portal in some configurations. In this case the URLs match, and the user interface is contextually controlled by the role-based access control permissions assigned to the tenant.	

vRealize Automation Physical Design

The physical design consists of characteristics and decisions that support the logical design. The design objective is to deploy a fully functional Cloud Management Portal with High Availability and the ability to provision to both Regions A and B.

To accomplish this design objective, you deploy or leverage the following in Region A to create a cloud management portal of the SDDC.

- 2 vRealize Automation Server 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 for Cloud Server.

- 1 vRealize Business for Cloud Remote 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. All the components that make up the Cloud Management Portal, along with their network connectivity, are shown in the following diagrams.

Figure 2-35. vRealize Automation Design for Region A
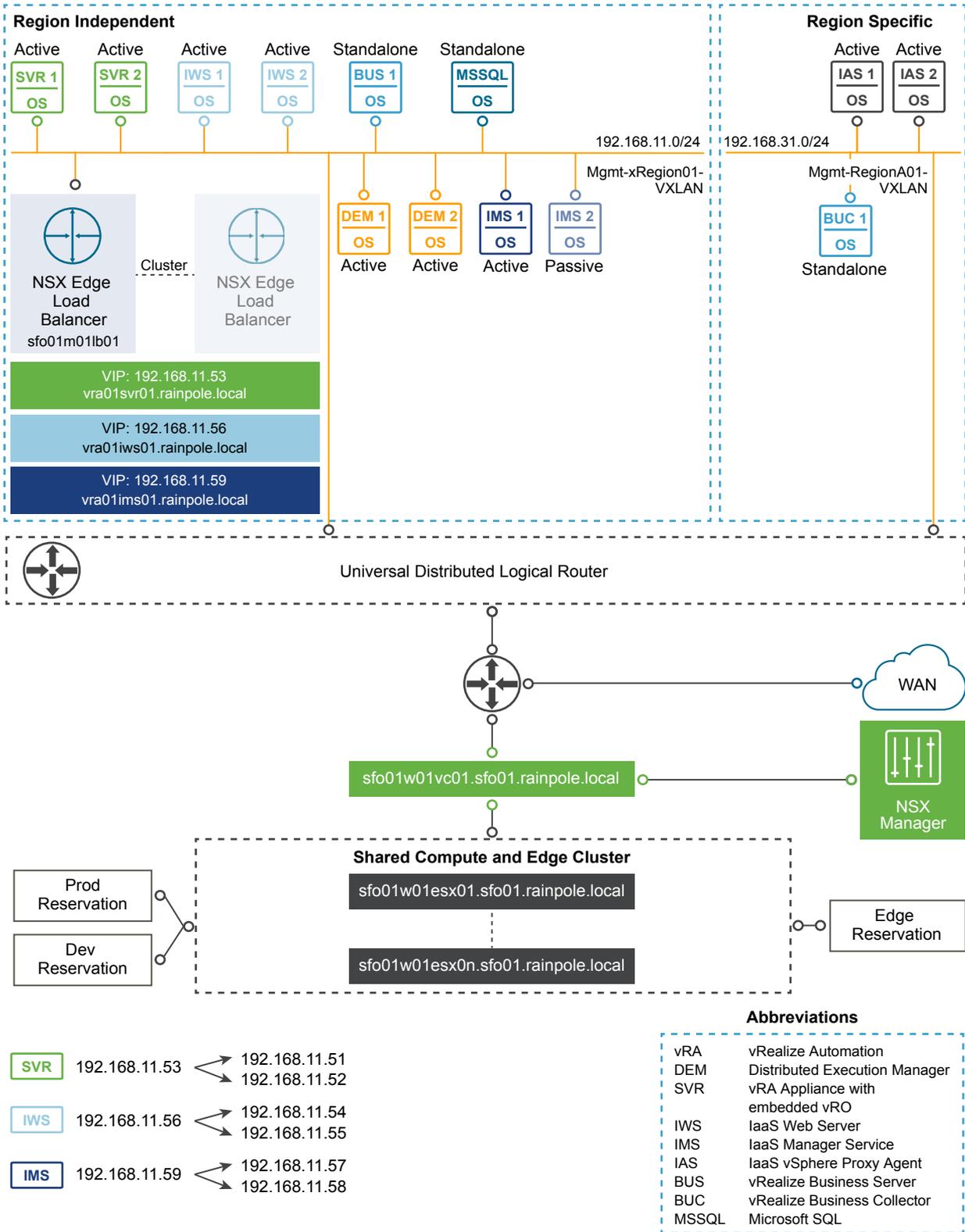
Figure 2-36. vRealize Automation Design for Region B
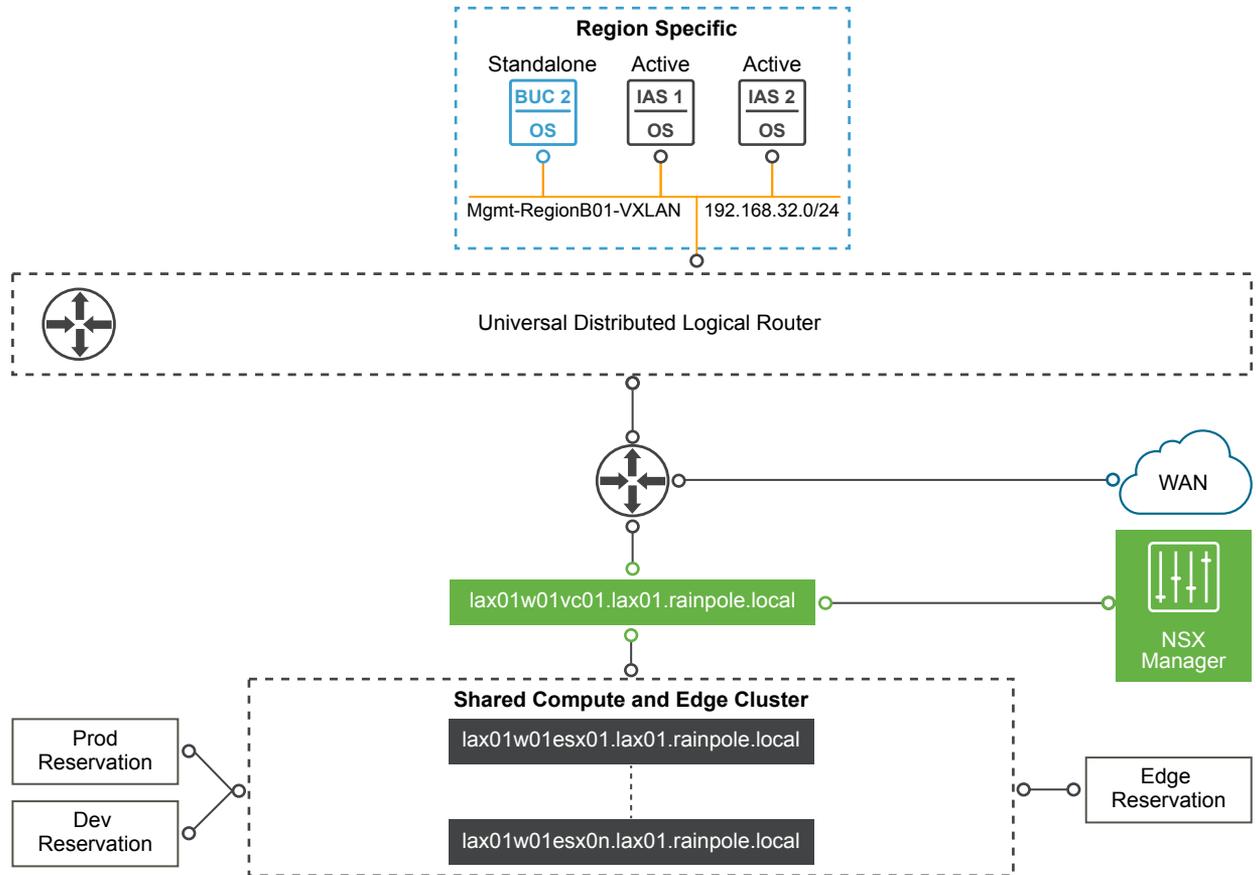
Deployment Considerations

This design uses NSX logical switches to abstract the vRealize Automation application and its supporting services. This abstraction allows the application to be hosted in any given region regardless of the underlying physical infrastructure such as network subnets, compute hardware, or storage types. This design places the vRealize Automation application and its supporting services in Region A. The same instance of the application manages workloads in both Region A and Region B.

Table 2-147. vRealize Automation Region Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-001	Use a single vRealize Automation installation to manage both Region A and Region B deployments from a single instance.	vRealize Automation can manage one or more regions and provides a single consumption portal regardless of region. The abstraction of the vRealize Automation application over virtual networking allows it to be independent from any physical site locations or hardware.	You must size vRealize Automation to accommodate multi-region deployments.

Table 2-148. vRealize Automation Anti-Affinity Rules

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-002	Apply vSphere Distributed Resource Scheduler (DRS) anti-affinity rules to the vRealize Automation components.	Using DRS prevents vRealize Automation nodes from residing on the same ESXi host and risking the cluster's high availability capability.	Additional configuration is required to set up anti-affinity rules. Only a single ESXi host in the management cluster, of the four ESXi hosts, can be put into maintenance mode at a time.

Table 2-149. vRealize Automation IaaS AD Requirement

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-003	vRealize Automation IaaS Machines are joined to Active Directory	Active Directory access is a hard requirement for vRealize Automation.	Active Directory access must be provided using dedicated service accounts

vRealize Automation Appliance

The vRealize Automation virtual appliance includes the cloud management Web portal, an embedded vRealize Orchestrator instance and database services. The vRealize Automation portal allows self-service provisioning and management of cloud services, as well as authoring blueprints, administration, and governance. The vRealize Automation virtual appliance uses an embedded PostgreSQL database for catalog persistence and database replication. The database is configured between two vRealize Automation appliances for high availability.

Table 2-150. vRealize Automation Virtual Appliance Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-004	Deploy two instances of the vRealize Automation virtual appliance to achieve redundancy. Each of these virtual appliances hosts an embedded vRealize Orchestrator instance.	Enable an active/active front-end portal for high availability.	None.
SDDC-CMP-005	Deploy two appliances that replicate data using the embedded PostgreSQL database.	Enable high availability for vRealize Automation. The embedded vRealize Orchestrator instance also uses this database.	In this active/passive configuration, manual failover between the two instances is required.
SDDC-CMP-006	During deployment, configure the vRealize Automation appliances with 18 GB vRAM.	Supports deployment of vRealize Automation in environments with up to 25,000 Active Directory users.	For environments with more than 25,000 Active Directory users of vRealize Automation, vRAM must be increased to 22 GB.

Table 2-151. vRealize Automation Virtual Appliance Resource Requirements per Virtual Machine

Attribute	Specification
Number of vCPUs	4
Memory	18 GB
vRealize Automation function	Portal web-site, Application, Orchestrator, service catalog and Identity Manager.

vRealize Automation IaaS Web Server

vRealize Automation IaaS web server provides a user interface within the vRealize Automation portal (a web site) for the administration and consumption of IaaS components.

The IaaS web-site provides infrastructure administration and service authoring capabilities to the vRealize Automation console. The web-site component communicates with the Model Manager, which provides it with updates from the Distributed Execution Manager (DEM), proxy agents and database.

The Model Manager communicates with the database, the DEMs, and the portal website. The Model Manager is divided into two separately installable components: the Model Manager web service and the Model Manager data component.

Note The vRealize Automation IaaS web server is a separate component from the vRealize Automation appliance.

Table 2-152. vRealize Automation IaaS Web Server Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-007	Install two vRealize Automation IaaS web servers.	vRealize Automation can support between 1,000 and 10,000 virtual machines. Two vRealize Automation IaaS web servers provide redundancy to the IaaS web server components.	Operational overhead increases as more servers are deployed.

Table 2-153. vRealize Automation IaaS Web Server Resource Requirements

Attribute	Specification
Number of vCPUs	2
Memory	4 GB
Number of vNIC ports	1
Number of local drives	1
vRealize Automation functions	Model Manager (web service)
Operating system	Microsoft Windows Server 2012 SP2 R2

vRealize Automation IaaS Manager Service and DEM Orchestrator Server

The vRealize Automation IaaS Manager Service and Distributed Execution Management (DEM) server are at the core of the vRealize Automation IaaS platform. The vRealize Automation IaaS Manager Service and DEM server supports several 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 Distributed Execution Manager (DEM) runs the business logic of custom models by interacting with other vRealize Automation component (repository) as required.

Each DEM instance acts in either an Orchestrator role or a Worker role. The DEM Orchestrator monitors the status of the DEM Workers. If a DEM worker stops or loses the connection to the Model Manager or repository, the DEM Orchestrator puts the workflow back in the queue. It manages the scheduled workflows by creating new workflow instances at the scheduled time and allows only one instance of a particular scheduled workflow to run at a given time. It also preprocesses workflows before execution. Preprocessing includes checking preconditions for workflows and creating the workflow's execution history.

Note   The vRealize Automation IaaS Manager Service and DEM Orchestrator service are separate services, but are installed on the same virtual machine.

Table 2-154. vRealize Automation IaaS Model Manager and DEM Orchestrator Server Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-008	Deploy two virtual machines to run 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 times.	You must provide more resources for these two virtual machines to accommodate the load of the two applications. If additional resources are required in the future, you can scale up these virtual machines at a later stage.

Table 2-155. vRealize Automation IaaS Model Manager and DEM Orchestrator Server Resource Requirements per Virtual Machine

Attribute	Specification
Number of vCPUs	2
Memory	4 GB
Number of vNIC ports	1
Number of local drives	1

Table 2-155. vRealize Automation IaaS Model Manager and DEM Orchestrator Server Resource Requirements per Virtual Machine (Continued)

Attribute	Specification
vRealize Automation functions	IaaS Manager Service, DEM Orchestrator
Operating system	Microsoft Windows Server 2012 SP2 R2

vRealize Automation IaaS DEM Worker Virtual Machine

vRealize Automation IaaS DEM Workers are responsible for the executing provisioning and deprovisioning tasks initiated by the vRealize Automation portal. DEM Workers are also utilized to communicate with specific infrastructure endpoints.

Table 2-156. vRealize Automation IaaS DEM Worker Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-009	Install three DEM Worker instances per DEM host.	Each DEM Worker can process up to 30 concurrent workflows. Beyond this limit, workflows are queued for execution. If the number of concurrent workflows is consistently above 90, you can add additional DEM Workers on the DEM host.	If you add more DEM Workers, you must also provide additional resources to run them.

Table 2-157. vRealize Automation DEM Worker Resource Requirements per Virtual Machine

Attribute	Specification
Number of vCPUs	2
Memory	6 GB
Number of vNIC ports	1
Number of local drives	1
vRealize Automation functions	DEM Worker
Operating system	Microsoft Windows Server 2012 SP2 R2

vRealize Automation IaaS Proxy Agent

The vRealize Automation IaaS Proxy Agent is a windows service used to communicate with specific infrastructure endpoints. In this design, the vSphere Proxy agent is utilized to communicate with vCenter.

The IaaS Proxy Agent server provides the following functions:

- vRealize Automation IaaS Proxy Agent can interact with different types of infrastructure components. For this design, only the vSphere Proxy agent is used.

- vRealize Automation does not itself virtualize resources, but works with vSphere to provision and manage the virtual machines. It uses vSphere Proxy agents to send commands to and collect data from vSphere.

Table 2-158. vRealize Automation IaaS Agent Server Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-010	Deploy two vRealize Automation vSphere Proxy Agent virtual machines.	Using two virtual machines provides redundancy for vSphere connectivity.	More resources are required because multiple virtual machines are deployed for this function.
SDDC-CMP-011	Abstract the proxy agent virtual machines on a separate virtual network for independent failover of the main vRealize Automation components across sites.	Allows the failover of the vRealize Automation instance from one site to another independently.	You must provision more application virtual networks and associated edge devices for the proxy agents.

Table 2-159. vRealize Automation IaaS Proxy Agent Resource Requirements per Virtual Machine

Attribute	Specification
Number of vCPUs	2
Memory	4 GB
Number of vNIC ports	1
Number of local drives	1
vRealize Automation functions	vSphere Proxy agent
Operating system	Microsoft Windows Server 2012 SP2 R2

Load Balancer

Session persistence of a load balancer allows the same server to serve all requests after a session is established with that server. 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 vRealize Automation Server (Manager Service) because only one Manager Service is active at any one time. Session persistence is not enabled because it is not a required component for the Manager Service.

Table 2‑160. Load Balancer Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC‑CMP‑012	Set up a load balancer to enable high availability for 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 balancers.
SDDC‑CMP‑013	Configure load balancer for vRealize Automation Server Appliance, Remote Console Proxy, and IaaS Web 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 may result in a poor user experience.	None
SDDC‑CMP‑014	Configure load balancer for vRealize IaaS Server and vRealize Orchestrator to use Round‑Robin algorithm without persistence.	Round‑robin provides a good balance of individual requests from the vRealize Server to the vRealize Orchestrator. Round‑robin distributes requests equally between the configured vRealize Orchestrator servers to allow the performance capacity of both to be best used. vRealize Automation IaaS Server is Active/stand‑by architecture and all requests go to a single node only.	None

Consider the following load balancer characteristics for vRealize Automation.

Table 2‑161. Load Balancer Application Profile Characteristics

Server Role	Type	Enable SSL Pass‑through	Persistence	Expires in (Seconds)
vRealize Automation - Persistence	HTTPS (443)	Enabled	Source IP	1800
vRealize Automation	HTTPS (443)	Enabled		

Table 2‑162. Load Balancer Service Monitoring Characteristics

Monitor	Interval	Timeout	Max Retries	Type	Expected	Method	URL	Receive
vRealize Automation Appliance	3	10	3	HTTPS	204	GET	/vcac/services/api/health	
vRealize Automation IaaS Web	3	10	3	HTTPS		GET	/wapi/api/status/web	REGISTERED
vRealize Automation IaaS Manager	3	10	3	HTTPS		GET	/VMPSProvision	ProvisionService
vRealize Orchestrator	3	10	3	HTTPS		GET	/vco-controlcenter/docs	

Table 2-163. Load Balancer Pool Characteristics

Server Role	Algorithm	Monitor	Members	Port	Monitor Port
vRealize Automation Appliance	Round Robin	vRealize Automation Appliance monitor	vRealize Automation Appliance nodes	443	
vRealize Automation Remote Console Proxy	Round Robin	vRealize Automation Appliance monitor	vRealize Automation Appliance nodes	8444	443
vRealize Automation IaaS Web	Round Robin	vRealize Automation IaaS Web monitor	IaaS web nodes	443	
vRealize Automation IaaS Manager	Round Robin	vRealize Automation IaaS Manager monitor	IaaS Manager nodes	443	
vRealize Automation Appliance	Round Robin	Embedded vRealize Automation Orchestrator Control Center monitor	vRealize Automation Appliance nodes	8283	

Table 2-164. Virtual Server Characteristics

Protocol	Port	Default Pool	Application Profile
HTTPS	443	vRealize Automation Appliance Pool	vRealize Automation - Persistence Profile
HTTPS	443	vRealize Automation IaaS Web Pool	vRealize Automation - Persistence Profile
HTTPS	443	vRealize Automation IaaS Manager Pool	vRealize Automation Profile
HTTPS	8283	Embedded vRealize Orchestrator Control Center Pool	vRealize Automation - Persistence Profile
HTTPS	8444	vRealize Automation Remote Console Proxy Pool	vRealize Automation - Persistence Profile

Information Security and Access Control in vRealize Automation

You use a service account for authentication and authorization of vRealize Automation to vCenter Server and vRealize Operations Manager for orchestrating and creating virtual objects in the SDDC.

Table 2-165. Authorization and Authentication Management Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-015	Configure a service account svc-vra in vCenter Server for application-to-application communication from vRealize Automation with vSphere.	You can introduce improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's life cycle outside of the SDDC stack to ensure its availability
SDDC-CMP-016	Use local permissions when you create the svc-vra service account in vCenter Server.	The use of local permissions ensures that only the Compute vCenter Server instances are valid and accessible endpoints from vRealize Automation.	If you deploy more Compute vCenter Server instances, you must ensure that the service account has been assigned local permissions in each vCenter Server so that this vCenter Server is a viable endpoint within vRealize Automation.
SDDC-CMP-017	Configure a service account svc-vra-vrops on vRealize Operations Manager for application-to-application communication from vRealize Automation for collecting health and resource metrics for tenant workload reclamation.	■ vRealize Automation accesses vRealize Operations Manager with the minimum set of permissions that are required for collecting metrics to determine the workloads that are potential candidates for reclamation. ■ In the event of a compromised account, 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.

vRealize Automation Supporting Infrastructure

To satisfy the requirements of this SDDC design, you configure additional components for vRealize Automation such as database servers for highly available database service and email server for notification.

Microsoft SQL Server Database

vRealize Automation uses a Microsoft SQL Server database to store information about the vRealize Automation IaaS elements and the machines that vRealize Automation manages.

Table 2-166. vRealize Automation SQL Database Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-018	Set up a Microsoft SQL server that supports the availability and I/O needs of vRealize Automation.	A dedicated or shared SQL server can be used so long as it meets the requirements of vRealize Automation.	Requires additional resources and licenses.
SDDC-CMP-019	Locate the Microsoft SQL server in the vRealize Automation virtual network or set it up to have global failover available.	For simple failover of the entire vRealize Automation instance from one region to another, the Microsoft SQL server must be running as a VM inside the vRealize Automation application virtual network. If the environment uses a shared SQL server, global failover ensures connectivity from both primary and secondary regions.	Adds additional overhead to managing Microsoft SQL services.
SDDC-CMP-020	Set up Microsoft SQL server with separate OS volumes for SQL Data, Transaction Logs, TempDB, and Backup.	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 might need to consult with the Microsoft SQL database administrators of your organization for guidance about production deployment in your environment.

Table 2-167. 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 2012
Microsoft SQL Database Version	SQL Server 2012 (110)
Operating system	Microsoft Windows Server 2012 R2

PostgreSQL Database Server

The vRealize Automation appliance uses a PostgreSQL database server to maintain the vRealize Automation portal elements and services, and the information about the catalog items that the appliance manages. The PostgreSQL is also used to host data pertaining to the embedded instance of vRealize Orchestrator.

Table 2‑168. vRealize Automation PostgreSQL Database Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-021	Use the embedded PostgreSQL database within each vRealize Automation appliance. This database is also used by the embedded vRealize Orchestrator.	Simplifies the design and enables replication of the database across the two vRealize Automation appliances.	None.
SDDC-CMP-022	Configure the embedded PostgreSQL database to use asynchronous replication.	Asynchronous replication offers a good balance between availability and performance.	Asynchronous replication provides a good level of availability in compliance with the design objectives.

Notification Email Server

vRealize Automation notification emails are sent using SMTP. These emails include notification of machine creation, expiration, and the notification of approvals received by users. vRealize Automation supports both anonymous connections to the SMTP server and connections using basic authentication. vRealize Automation also supports communication with or without SSL.

You create a global, inbound email server to handle inbound email notifications, such as approval responses. Only one, global inbound email server, which appears as the default for all tenants, is needed. The email server provides accounts that you can customize for each user, providing separate email accounts, usernames, and passwords. Each tenant can override these settings. If tenant administrators do not override these settings before enabling notifications, vRealize Automation uses the globally configured email server. The server supports both the POP and the IMAP protocol, with or without SSL certificates.

Notifications

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.

System administrators create a global outbound email server to process outbound email notifications, and a global inbound email server to process inbound email notifications, such as responses to approvals.

Table 2‑169. vRealize Automation Email Server Configuration

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-023	Configure vRealize Automation to use a global outbound email server to handle outbound email notifications and a global inbound email server to handle inbound email notifications, such as approval responses.	Requirement to integrate vRealize Automation approvals and system notifications through emails.	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.

vRealize Automation Cloud Tenant Design

A tenant is an organizational unit within a vRealize Automation deployment, and can represent a business unit within an enterprise, or a company that subscribes to cloud services from a service provider. Each tenant has its own dedicated configuration, although some system-level configuration is shared across tenants.

Comparison of Single-Tenant and Multi-Tenant Deployments

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. The infrastructure resources, such as cloud or virtual compute resources or physical machines, can be divided into fabric groups managed by fabric administrators. The resources in each fabric group can be allocated to business groups within each tenant by using reservations.

Default-Tenant Deployment	In a default-tenant deployment, all configuration occurs in the default tenant. Tenant administrators can manage users and groups, and configure tenant-specific branding, notifications, business policies, and catalog offerings. All users log in to the vRealize Automation console at the same URL, but the features available to them are determined by their roles.
Single-Tenant Deployment	In a single-tenant deployment, the system administrator creates a single new tenant for the organization that use 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 segregated from the default tenant, although users with system-wide roles can view and manage both configurations. The IaaS administrator for the organization tenant creates fabric groups and appoints fabric administrators. Fabric administrators can create reservations for business groups in the organization tenant.
Multi-Tenant Deployment	In a multi-tenant deployment, the system administrator creates new tenants for each organization that uses the same vRealize Automation instance. Tenant users log in to the vRealize Automation console at a URL specific to their tenant. Tenant-level configuration is segregated 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 appoints fabric administrators to their respective tenants. Although fabric administrators can create reservations for business groups in any tenant, in this scenario

they typically create and manage reservations within their own tenants. If the same identity store is configured in multiple tenants, the same users can be designated as IaaS administrators or fabric administrators for each tenant.

Tenant Design

This design deploys a single tenant containing two business groups.

- The first business group is designated for production workloads provisioning.

- The second business group is designated 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 by their roles.

**Figure 2-37. Rainpole Cloud Automation Tenant Design for Two Regions**
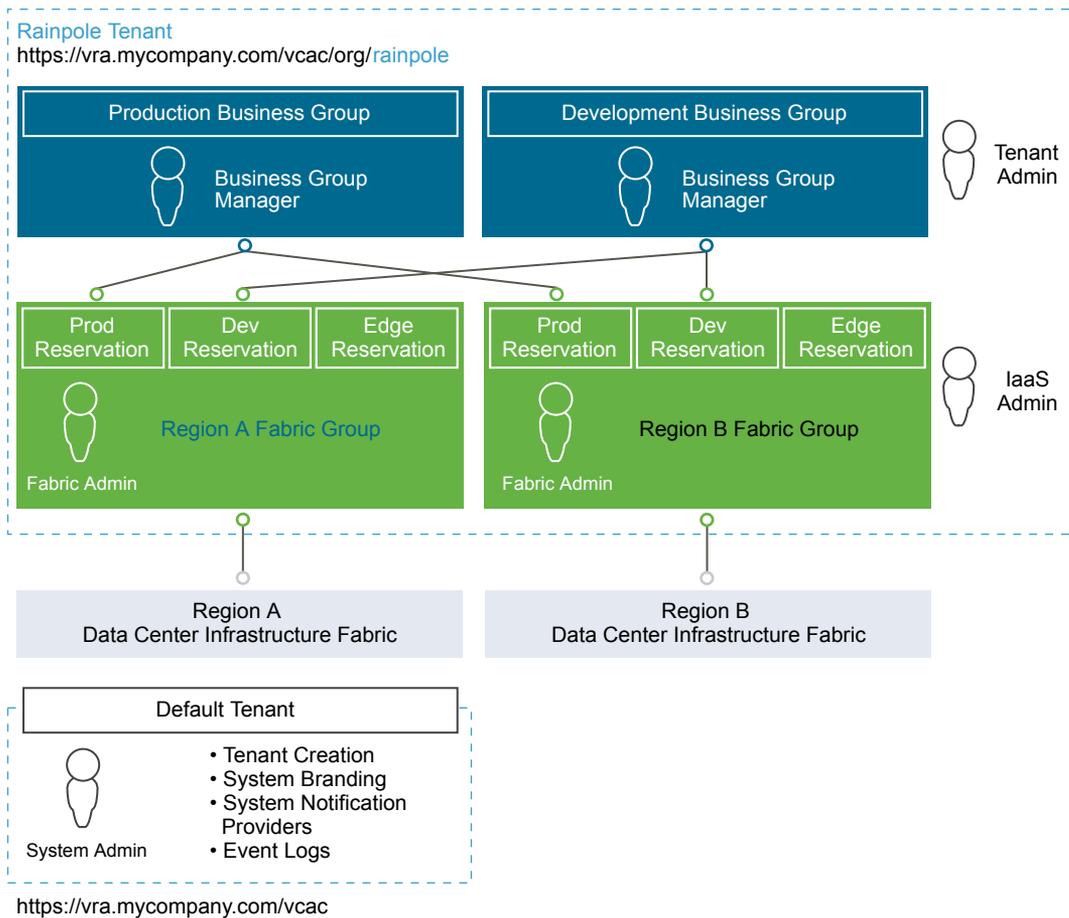
Table 2-170. Tenant Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-024	Uses vRealize Automation business groups for separate business units (instead of separate tenants).	Allows transparency across the environments and some level of sharing of resources and services such as blueprints.	Some elements, such as property groups, are visible to both business groups. The design does not provide full isolation for security or auditing.
SDDC-CMP-025	Create separate fabric groups for each deployment region. Each fabric group represent region-specific data center resources. Each of the business groups has reservations into each of the fabric groups.	Provides future isolation of fabric resources and potential delegation of duty to independent fabric administrators.	Initial deployment uses a single shared fabric that consists of one compute cluster.
SDDC-CMP-026	Allow access to the default tenant only by the system administrator and for the purposes of 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-027	Evaluate your internal organizational structure and workload needs. Configure Business Groups, Reservations, Service catalogs, and Blueprints in the vRealize Automation instance based on your organization's needs.	vRealize Automation integrates with your organization's needs. Within this design, guidance for Rainpole is provided as a starting point, but this guidance may not be appropriate for your specific business needs.	Partners and customers must evaluate their specific business needs.

Service Catalog

The service catalog provides a common interface for consumers of IT services to use to request and manage the services and resources they need.

A tenant administrator or service architect can specify information about the service catalog, such as the service hours, support team, and change window. While the catalog does not enforce service-level agreements on services, this service hours, support team, and change window information 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 2012 preinstalled, and then manage that virtual machine after it has been provisioned.

Tenant administrators define new catalog items and publish them to the service catalog. The tenant administrator can then manage the presentation of catalog items to the consumer and entitle new items to consumers. To make the catalog item available to users, a tenant administrator must entitle the item to the users and groups 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

A machine blueprint is the complete specification for a virtual, cloud or physical machine. A machine blueprint determines the machine's attributes, how it is provisioned, and its policy and management settings. Machine blueprints are published as catalog items in the service catalog.

Machine blueprints can be specific to a business group or shared among groups within a tenant. Tenant administrators can create shared blueprints that can be entitled to users in any business group within the tenant. Business group managers can create group blueprints that can only be entitled to users within a specific business group. A business group manager cannot modify or delete shared blueprints. Tenant administrators cannot view or modify group blueprints unless they also have the business group manager role for the appropriate group.

If a tenant administrator sets a shared blueprint's properties so that it can be copied, the business group manager can also copy the shared blueprint for use as a starting point to create a new group blueprint.

Table 2-171. Single Machine Blueprints

Name	Description
Base Windows Server (Development)	Standard Rainpole SOE deployment of Windows 2012 R2 available to the Development business group.
Base Windows Server (Production)	Standard Rainpole SOE deployment of Windows 2012 R2 available to the Production business group.
Base Linux (Development)	Standard Rainpole SOE deployment of Linux available to the Development business group.
Base Linux (Production)	Standard Rainpole SOE deployment of Linux available to the Production business group.
Windows Server + SQL Server (Production)	Base Windows 2012 R2 Server with silent SQL 2012 Server install with custom properties. This is available to the Production business group.
Windows Server + SQL Server (Development)	Base Windows 2012 R2 Server with silent SQL 2012 Server install with custom properties. This is available to the Development business group.

Blueprint Definitions

The following sections provide details of each service definition that has been included as part of the current phase of cloud platform deployment.

Table 2-172. Base Windows Server Requirements and Standards

Service Name	Base Windows Server
Provisioning Method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter customizations.
Entitlement	Both Production and Development business group members.
Approval Process	No approval (pre-approval assumed based on approved access to platform).
Operating System and Version Details	Windows Server 2012 R2

Table 2-172. Base Windows Server Requirements and Standards (Continued)

Service Name	Base 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 2-173. Base Windows Blueprint Sizing

Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	4	60
Maximum	4	16	60

Table 2-174. Base Linux Server Requirements and Standards

Service Name	Base Linux Server
Provisioning Method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter customizations.
Entitlement	Both Production and Development business group members.
Approval Process	No approval (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 2-175. Base Linux Blueprint Sizing

Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	6	20
Maximum	4	12	20

Table 2-176. Base Windows Server with SQL Server Install Requirements and Standards

Service Name	Base Windows Server
Provisioning Method	When users select this blueprint, vRealize Automation clones a vSphere virtual machine template with preconfigured vCenter customizations.
Entitlement	Both Production and Development business group members
Approval Process	No approval (pre-approval assumed based on approved access to platform).
Operating System and Version Details	Windows Server 2012 R2
Configuration	Disk: Single disk drive Network: Standard vSphere Networks Silent Install: The Blueprint calls a silent script using the vRealize Automation Agent to install SQL2012 Server with custom properties.
Lease and Archival Details	Lease: ■ Production Blueprints: No expiration date ■ Development Blueprints: Minimum 30 days – Maximum 270 days Archive: 15 days
Pre- and Post-Deployment Requirements	Email sent to manager confirming service request (include description details)

Table 2-177. Base Windows with SQL Server Blueprint Sizing

| Sizing	vCPU	Memory (GB)	Storage (GB)
Default	1	8	100
Maximum	4	16	400

Branding of the vRealize Automation Console

System administrators can change the appearance of the vRealize Automation console to meet site-specific branding guidelines by changing the logo, the background color, or information in the header and footer. System administrators control the default branding for tenants. Tenant administrators can use the default or reconfigure branding for each tenant.

vRealize Automation Infrastructure as a Service Design

This topic introduces the integration of vRealize Automation with vSphere resources used to create the Infrastructure as a Service design for use with the SDDC.

Figure 2-38 illustrates the logical design of the vRealize Automation groups and vSphere resources.

**Figure 2-38. vRealize Automation Logical Design**
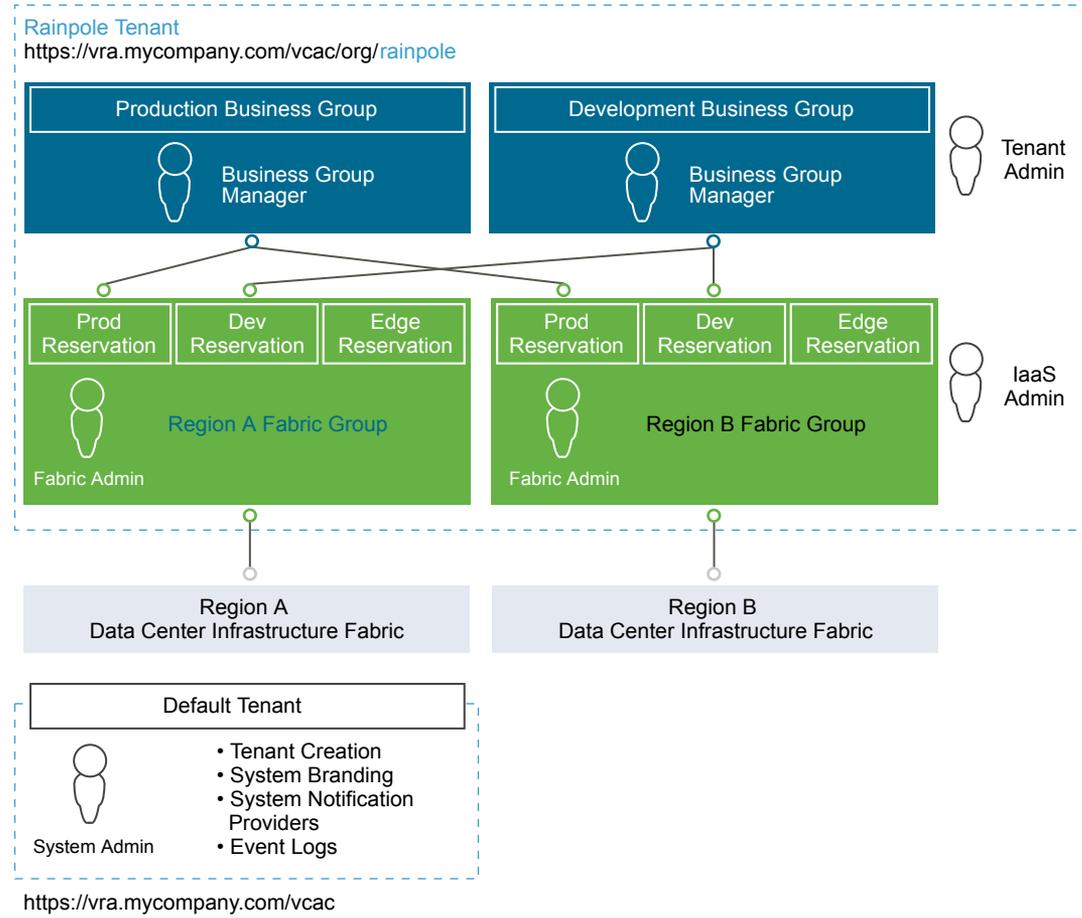
The following terms apply to vRealize Automation when integrated with vSphere. These terms and their meaning may vary from the way they are used when referring only to vSphere.

Term	Definition
vSphere (vCenter Server) endpoint	Provides information required by vRealize Automation IaaS to access vSphere compute resources.
Compute resource	Virtual object within vRealize Automation that represents a vCenter Server cluster or resource pool, and datastores or datastore clusters.
	Note Compute resources are CPU, memory, storage and networks. Datastores and datastore clusters are part of the overall storage resources.
Fabric groups	vRealize Automation IaaS organizes compute resources into fabric groups.
Fabric administrators	Fabric administrators manage compute resources, which are organized into fabric groups.
Compute reservation	A share of compute resources (vSphere cluster, resource pool, datastores, or datastore clusters), such as CPU and memory reserved for use by a particular business group for provisioning virtual machines.
	Note vRealize Automation uses the term reservation to define resources (be they 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.

Term	Definition
Storage reservation	Similar to compute reservation (see above), but pertaining only to a share of the available storage resources. In this context, you specify a storage reservation in terms of gigabytes from an existing LUN or Datastore.
Business groups	A collection of virtual machine consumers, usually corresponding to an organization's business units or departments. Only users in the business group can request virtual machines.
Reservation policy	vRealize Automation IaaS determines its reservation (also called virtual reservation) from which a particular virtual machine is provisioned. The reservation policy is a logical label or a pointer to the original reservation. Each virtual reservation can be added to one reservation policy.
Blueprint	The complete specification for a virtual machine, determining the machine attributes, the manner in which it is provisioned, and its policy and management settings.
	Blueprint allows the users of a business group to create virtual machines on a virtual reservation (compute resource) based on the reservation policy, and using platform and cloning types. It also lets you specify or add machine resources and build profiles.

Figure 2-39 shows the logical design constructs discussed in the previous section as they apply to a deployment of vRealize Automation integrated with vSphere in a cross data center provisioning.

Figure 2-39. vRealize Automation Integration with vSphere Endpoint
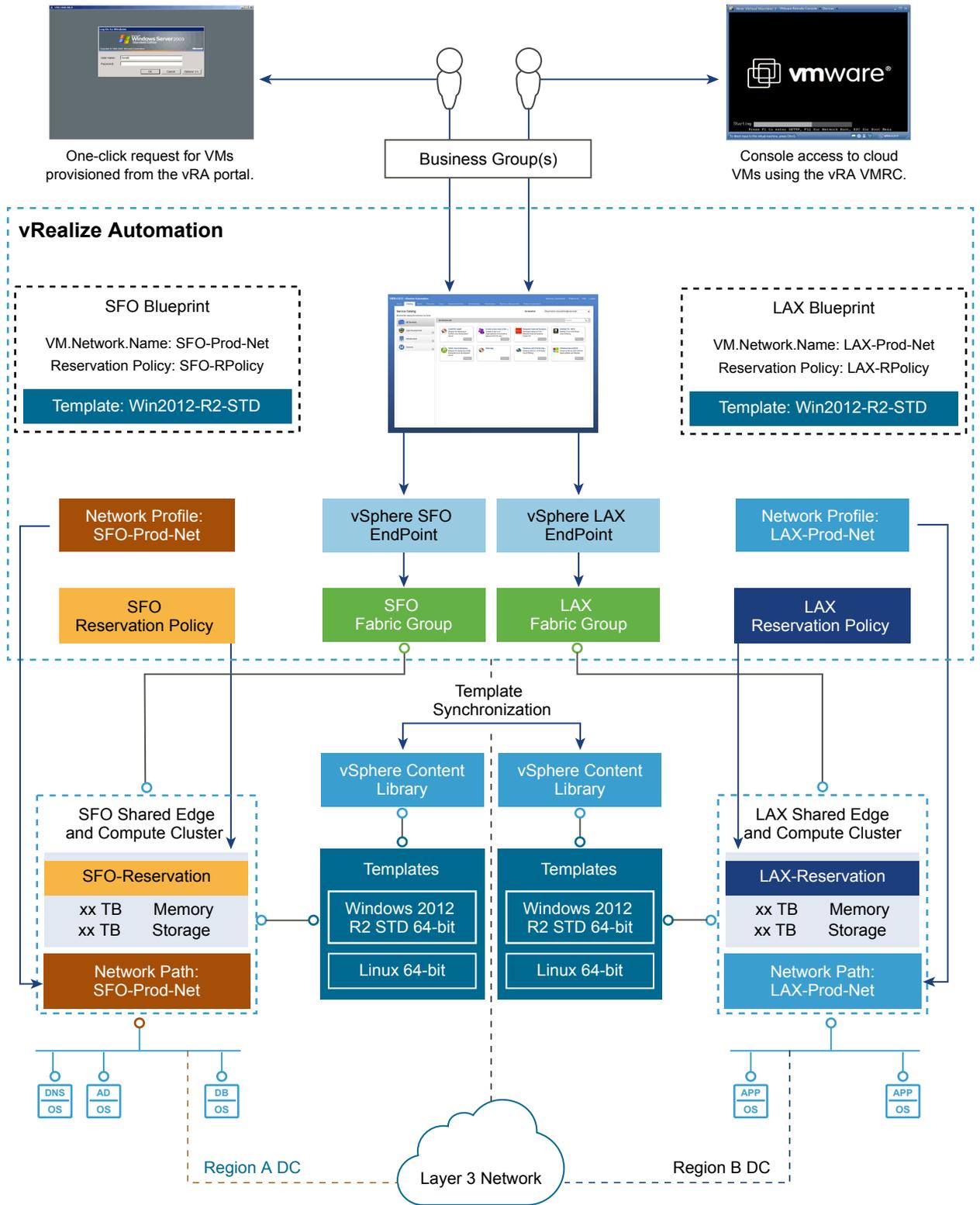
Infrastructure Source Endpoints

An infrastructure source endpoint is a connection to the infrastructure that provides a set (or multiple sets) of resources, which can then be made available by IaaS administrators for consumption by 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 pods, the terms are often used interchangeably).

Infrastructure data is collected through proxy agents that manage and communicate with the endpoint resources. This information about the compute resources on each infrastructure endpoint and the machines provisioned on each computer resource is collected at regular intervals.

During installation of the vRealize Automation IaaS components, you can configure the proxy agents and define their associated endpoints. Alternatively, you can configure the proxy agents and define their associated endpoints separately after the main vRealize Automation installation is complete.

Table 2‑178. Endpoint Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-028	Create two vSphere endpoints.	One vSphere endpoint is required to connect to each vCenter Server instance in each region. Two endpoints are needed for two regions.	As additional regions are brought online, you must deploy additional vSphere endpoints.
SDDC-CMP-029	Create one vRealize Orchestrator endpoint that is configured to connect to the embedded vRealize Orchestrator instance.	vRealize Automation extensibility uses vRealize Orchestrator. The design includes one embedded vRealize Orchestrator cluster exists which requires the creation of a single endpoint.	Requires configuration of a vRealize Orchestrator endpoint.
SDDC-CMP-030	Create one NSX endpoint and associate it with the vSphere endpoint.	The NSX endpoint is required to connect to the NSX Manager and enable all the NSX-related operations supported in vRealize Automation blueprints.	None.

Virtualization Compute Resources

A virtualization compute resource is a vRealize Automation object that represents an ESXi host or a cluster of ESXi hosts. When a group member requests a virtual machine, the virtual machine is provisioned on these compute resources. vRealize Automation regularly collects information about known compute resources and the virtual machines provisioned on them through the proxy agents.

Table 2‑179. Compute Resource Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-031	Create at least one compute resource for each deployed region.	Each region has one compute cluster, one compute resource is required for each cluster.	As additional compute clusters are created, you must add them to the existing compute resource in their region or to a new resource, which has to be created.

Note By default, compute resources are provisioned to the root of the compute cluster. In this design, use of vSphere resource pools is mandatory.

Fabric Groups

A fabric group is a logical container of several compute resources, and can be managed by fabric administrators.

Table 2-180. Fabric Group Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-032	Create a fabric group for each region and include all the compute resources and edge resources in that region.	To enable region-specific provisioning a fabric group in each region must be created.	As additional clusters are added in a region, they must be added to the fabric group.

Business Groups

A business group is a collection of machine consumers, 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

A reservation is a share of one compute resource's available memory, CPU and storage reserved for use by a particular fabric group. Each reservation is for one fabric group only but the relationship is many-to-many. A fabric group might have multiple reservations on one compute resource, or reservations on multiple compute resources, or both.

Converged Compute/Edge Clusters and Resource Pools

While reservations provide a method to allocate a portion of the cluster memory or storage within vRealize Automation, reservations do not control how CPU and memory are allocated during periods of contention on the underlying vSphere compute resources. vSphere Resource Pools are used to control the allocation of CPU and memory during time of resource contention on the underlying host. To fully use this, all VMs must be deployed into one of four resource pools: sfo01-w01rp-sddc-edge, sfo01-w01rp-sddc-mgmt, sfo01-w01rp-user-edge, and sfo01-w01rp-user-vm.

Resource pool details:

- sfo01-w01rp-sddc-edge is dedicated for data center level NSX Edge components and should not contain any user workloads.

- sfo01-w01rp-sddc-mgmt is dedicated for management VMs.

- sfo01-w01rp-user-edge is dedicated for any statically or dynamically deployed NSX components such as NSX Edge gateways or Load Balancers which serve specific customer workloads.

- sfo01-w01rp-user-vm is dedicated for any statically or dynamically deployed virtual machines such as Windows, Linux, databases, etc., which contain specific customer workloads.

Table 2-181. Reservation Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-033	Create at least one vRealize Automation reservation for each business group at each region.	In our example, each resource cluster has two reservations, one for production and one for development, allowing both production and development workloads to be provisioned.	Because production and development share compute resources, the development business group must be limited to a fixed amount of resources.
SDDC-CMP-034	Create at least one vRealize Automation reservation for edge resources in each region.	An edge reservation in each region allows NSX to create edge services gateways on demand and place them on the edge cluster.	The workload reservation must define the edge reservation in the network settings.
SDDC-CMP-035	Configure all vRealize Automation workloads to use the sfo01-w01rp-user-vm resource pool.	To ensure dedicated compute resources of NSX networking components, tenant deployed workloads must be assigned to a dedicated vSphere DRS resource pools. Workloads provisioned at the root resource pool level receive more resources than those at child resource pools. Such a configuration might starve those virtual machines in contention situations.	Cloud administrators must ensure that all workload reservations are configured with the appropriate resource pool. This may be a single resource pool for both production and development workloads, or two resource pools, one dedicated for the Development Business Group and one dedicated for the Production Business Group.
SDDC-CMP-036	Configure vRealize Automation reservations for dynamically provisioned NSX Edge components (routed gateway) to use the sfo01-w01rp-user-edge resource pool.	To ensure dedicated compute resources of NSX networking components, end-user deployed NSX Edge components must be assigned to a dedicated end-user network component vCenter Resource Pool. Workloads provisioned at the root resource pool level receives more resources than resource pools, which might starve those virtual machines in contention situations.	Cloud administrators must ensure that all workload reservations are configured with the appropriate resource pool.
SDDC-CMP-037	All vCenter resource pools used for Edge or Compute workloads must be created at the "root" level. Do not nest resource pools.	Nesting of resource pools can create administratively complex resource calculations that may result in unintended under or over allocation of resources during contention situations.	All resource pools must be created at the root resource pool level.

Reservation Policies

You can add each virtual reservation to one reservation policy. The reservation from which a particular virtual machine is provisioned is determined by vRealize Automation based on the reservation policy specified in the blueprint, if any, the priorities and current usage of the fabric group's reservations, and other custom properties.

Table 2-182. Reservation Policy Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-038	Create at least one workload reservation policy for each region.	Reservation policies are used to target a deployment to a specific set of reservations in each region. Reservation policies are also used to target workloads into their appropriate region, compute cluster and vSphere resource pool.	None
SDDC-CMP-039	Create at least one reservation policy for placement of dynamically created edge services gateways into the edge clusters.	Required to place the edge devices into their respective edge clusters and vSphere resource pools.	None

A storage reservation policy is a set of datastores that can be assigned to a machine blueprint to restrict disk provisioning to only those datastores. Storage reservation policies are created and associated with the appropriate datastores and assigned to reservations.

**Table 2-183.** Storage Reservation Policy Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-040	Within this design, storage tiers are not used.	The underlying physical storage design does not use storage tiers.	Both business groups have access to the same storage. Customers using multiple datastores with different storage capabilities must evaluate the usage of vRealize Automation Storage Reservation Policies.

Template Synchronization

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, vSphere Content Library or external services like vCloud Connector or vSphere Replication.

**Table 2-184.** Template Synchronization Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-041	This design uses vSphere Content Library to synchronize templates across regions.	The vSphere Content Library is built into the version of vSphere being used and meets all the requirements to synchronize templates.	Storage space must be provisioned in each region. vRealize Automation cannot directly consume templates from vSphere Content Library.

Figure 2-40. Template Synchronization
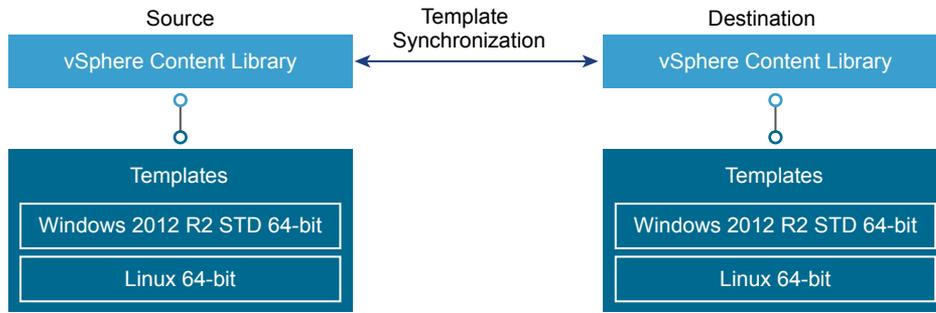
VMware Identity Management

VMware Identity Manager is integrated into the vRealize Automation appliance, and provides tenant identity management.

The VMware Identity Manager synchronizes with the Rainpole Active Directory domain. Important users and groups are synchronized with VMware Identity Manager. Authentication uses the Active Directory domain, but searches are made against the local Active Directory mirror on the vRealize Automation appliance.

Figure 2-41. VMware Identity Manager Proxies Authentication Between Active Directory and vRealize Automation
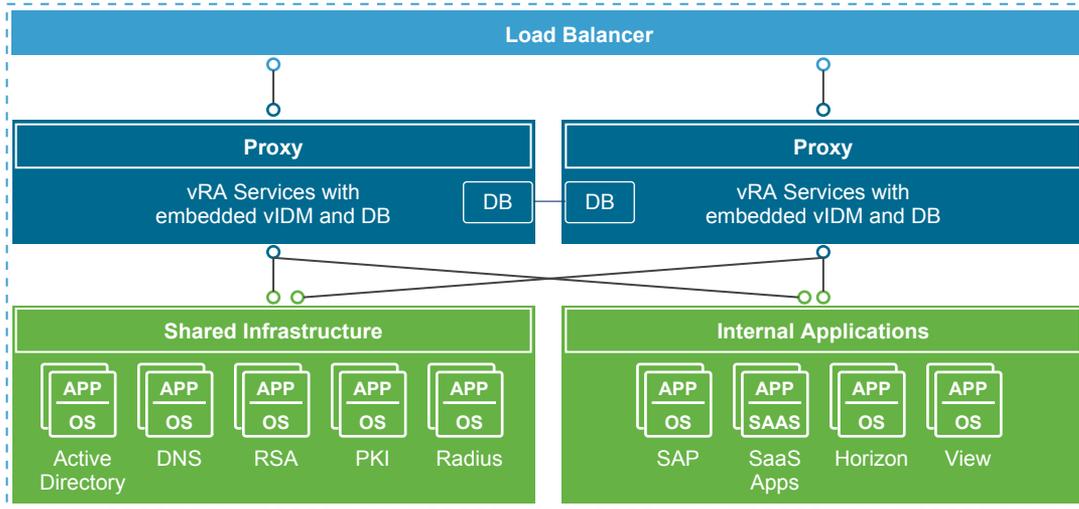
Table 2-185. Active Directory Authentication Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-042	Choose Active Directory with Integrated Windows Authentication as the Directory Service connection option.	Rainpole uses a single-forest, multiple-domain Active Directory environment. Integrated Windows Authentication supports establishing trust relationships in a multi-domain or multi-forest Active Directory environment.	Requires that the vRealize Automation appliances are joined to the Active Directory domain.

By default, the vRealize Automation appliance is configured with 18 GB of memory, which is enough to support a small Active Directory environment. An Active Directory environment is considered small if it 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 2-186. Connector Configuration Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-043	To support Directories Service high availability, configure a second connector that corresponds to the second vRealize Automation appliance.	This design supports high availability by installing two vRealize Automation appliances and using load-balanced NSX Edge instances. Adding the second connector to the second vRealize Automation appliance ensures redundancy and improves performance by load balancing authentication requests.	This design simplifies the deployment while using robust built-in HA capabilities. This design uses NSX for vSphere load balancing.

vRealize Business for Cloud Design

vRealize Business for Cloud provides end-user transparency in the costs that are associated with operating workloads. A system, such as vRealize Business, to gather and aggregate the financial cost of workload operations provides greater visibility 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 to meet the requirements of an organization.

Table 2-187. vRealize Business for Cloud Standard Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-044	Deploy vRealize Business for Cloud as part of the cloud management platform and integrate it with vRealize Automation.	Tenant and Workload costing is provided by vRealize Business for Cloud.	You must deploy more appliances - for vRealize Business for Cloud and for remote collectors.
SDDC-CMP-045	Use default vRealize Business for Cloud appliance size (8 GB). For vRealize Business for Cloud remote collector, use a reduced memory size of 2 GB.	Default vRealize Business for Cloud appliance size supports up to 10,000 VMs Remote Collectors do not run server service, and can run on 2 GB of RAM.	None.

Table 2‑187. vRealize Business for Cloud Standard Design Decision (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-046	Use default vRealize Business reference costing database.	Default reference costing is based on industry information and is periodically updated.	Default reference costing might not accurately represent actual customer costs. vRealize Business Appliance requires Internet access to periodically update the reference database.
SDDC-CMP-047	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 should be regionally local to the resource which they are configured to collect. Because this design supports disaster recovery, the CMP can reside in Region A or Region B.	In the case where the environment does not have disaster recovery support, you must deploy an additional appliance, the one for the remote data collector, although the vRealize Business server can handle the load on its own.
SDDC-CMP-048	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-049	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 additional Layer 3 hop through an NSX Edge device.

Table 2‑188. vRealize Business for Cloud Virtual Appliance Resource Requirements per Virtual Machine

Attribute	Specification
Number of vCPUs	4
Memory	8 GB for Server / 2 GB for Remote Collector
vRealize Business function	Server or remote collector

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 Administration uses the vRealize Orchestrator Plug-In to connect to vCenter Server for compute resource allocation.

vRealize Orchestrator Logical Design

This VMware Validated Design uses the vRealize Orchestrator instance that is embedded within the vRealize Automation appliance, instead of using a dedicated or external vRealize Orchestrator instance.

Table 2-189. vRealize Orchestrator Hardware Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-01	Use the internal vRealize Orchestrator instances that are embedded in the deployed vRealize Automation instances.	The use of 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.

vRealize Orchestrator Authentication

The embedded vRealize Orchestrator only supports the following authentication method:

■ vRealize Automation Authentication

Table 2-190. vRealize Orchestrator Directory Service Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-02	Embedded vRealize Orchestrator uses the vRealize Automation authentication.	Only authentication method available.	None.
SDDC-CMP-VRO-03	Configure vRealize Orchestrator to use the vRealize Automation customer tenant (rainpole) for authentication.	The vRealize Automation Default Tenant users are only administrative users. By connecting to the customer tenant, workflows running on vRealize Orchestrator may run with end-user granted permissions.	End users who run vRealize Orchestrator workflows are required to have permissions on the vRealize Orchestrator server. Some plug-ins may not function correctly using vRealize Automation Authentication.
SDDC-CMP-VRO-04	Each vRealize Orchestrator instance is associated with only one customer tenant.	To provide best security and segregation between potential tenants, vRealize Orchestrator instance is associated with a single tenant.	If additional vRealize Automation Tenants are configured, additional vRealize Orchestrator installations are needed.

Network Ports

vRealize Orchestrator uses specific network ports to communicate with other systems. The ports are configured with a default value, but you can change the defaults at any time. When you make changes, verify that all ports are available for use by your host. If necessary, open these ports on any firewalls through which network traffic for the relevant components flows. Verify that the required network ports are open before you deploy vRealize Orchestrator.

Default Communication Ports

Set default network ports and configure your firewall to allow incoming TCP connections. Other ports may be required if you are using custom plug-ins.

Table 2-191. vRealize Orchestrator Default Configuration Ports

Port	Number	Protocol	Source	Target	Description
HTTPS server port	443	TCP	End-user Web browser	Embedded vRealize Orchestrator server	The SSL secured HTTP protocol used to connect to the vRealize Orchestrator REST API.
Web configuration HTTPS access port	8283	TCP	End-user Web browser	vRealize Orchestrator configuration	The SSL access port for the control center Web UI for vRealize Orchestrator configuration.

External Communication Ports

Configure your firewall to allow outgoing connections using the external network ports so vRealize Orchestrator can communicate with external services.

Table 2-192. vRealize Orchestrator Default External Communication Ports

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.
SQL Server	1433	TCP	vRealize Orchestrator server	Microsoft SQL server	Port used to communicate with the Microsoft SQL Server or SQL Server Express instances that are configured as the vRealize Orchestrator database.
PostgreSQL	5432	TCP	vRealize Orchestrator server	PostgreSQL server	Port used to communicate with the PostgreSQL Server that is configured as the vRealize Orchestrator database.
Oracle	1521	TCP	vRealize Orchestrator server	Oracle DB server	Port used to communicate with the Oracle Database Server that is configured as the vRealize Orchestrator database.
SMTP Server port	25	TCP	vRealize Orchestrator server	SMTP Server	Port used for email notifications.
vCenter Server API port	443	TCP	vRealize Orchestrator server	VMware vCenter server	The vCenter Server API communication port used by vRealize Orchestrator to obtain virtual infrastructure and virtual machine information from the orchestrated vCenter Server instances.

Table 2-192. vRealize Orchestrator Default External Communication Ports (Continued)

Port	Number	Protocol	Source	Target	Description
vCenter Server	80	TCP	vRealize Orchestrator server	vCenter Server	Port used to tunnel HTTPS communication.
VMware ESXi	443	TCP	vRealize Orchestrator server	ESXi hosts	(Optional) Workflows using the vCenter Guest Operations API need direct connection between vRealize Orchestrator and the ESXi hosts the VM is running on.

vRealize Orchestrator Server Mode

vRealize Orchestrator supports standalone mode and cluster mode. This design uses cluster mode.

vRealize Orchestrator supports the following server modes.

Standalone mode
vRealize Orchestrator server runs as a standalone instance. This is the default mode of operation.

Cluster mode
To increase availability of the vRealize Orchestrator services, and to create a more highly available SDDC, you can configure vRealize Orchestrator to work in cluster mode, 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 cluster the vRealize Automation appliances, the vRealize Orchestrator instances embedded within them are automatically clustered.

All vRealize Orchestrator server instances communicate with each other by exchanging heartbeats at a certain time interval. Only active vRealize Orchestrator server instances respond to client requests and run workflows. If an active vRealize Orchestrator server instance fails to send heartbeats, it is considered to be non-responsive, and one of the inactive instances takes over to resume all workflows from the point at which they were interrupted. The heartbeat is implemented through the shared database, so there are no implications in the network design for a vRealize Orchestrator cluster. If you have more than one active vRealize Orchestrator node in a cluster, concurrency problems can occur if different users use the different vRealize Orchestrator nodes to modify the same resource.

vRealize Orchestrator Load Balancer Configuration

Configure load balancing for the vRealize Orchestrator instances embedded within the two vRealize Automation instances to provision network access to the vRealize Orchestrator control center.

Table 2-193. vRealize Orchestrator SDDC Cluster Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-05	Configure the load balancer to allow network access to the embedded vRealize Orchestrator control center.	The control center allows customization of vRealize Orchestrator, such as changing the tenant configuration and certificates.	None.

vRealize Orchestrator Information Security and Access Control

You use a service account for authentication and authorization of vRealize Orchestrator to vCenter Server for orchestrating and creating virtual objects in the SDDC.

Table 2-194. Authorization and Authentication Management Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-06	Configure a service account svc-vro in vCenter Server for application-to-application communication from vRealize Orchestrator with vSphere.	Introduces improved accountability in tracking request-response interactions between the components of the SDDC.	You must maintain the service account's life cycle outside of the SDDC stack to ensure its availability
SDDC-CMP-VRO-07	Use local permissions when you create the svc-vro service account in vCenter Server.	The use of local permissions 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 ensure that the service account has been assigned local permissions in each vCenter Server so that this vCenter Server is a viable endpoint in vRealize Orchestrator.

vRealize Orchestrator Configuration

vRealize Orchestrator configuration includes guidance on client configuration, database configuration, SSL certificates, and plug-ins.

vRealize Orchestrator Client

The vRealize Orchestrator client is a desktop application that lets you import packages, create, run, and schedule workflows, and manage user permissions.

You can install the standalone version of the vRealize Orchestrator Client on a desktop system. Download the vRealize Orchestrator Client installation files from the vRealize Orchestrator appliance page at `https://vRA_hostname/vco`. Alternatively, you can run the vRealize Orchestrator Client using Java WebStart directly from the homepage of the vRealize Automation appliance console.

SSL Certificates

The vRealize Orchestrator configuration interface uses a secure connection to communicate with vCenter Server, relational database management systems (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 2-195. vRealize Orchestrator SSL Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-08	The embedded vRealize Orchestrator instance uses the vRealize Automation appliance certificate.	Using the vRealize Automation certificate simplifies the configuration of the embedded vRealize Orchestrator instance.	None.

vRealize Orchestrator Database

vRealize Orchestrator requires a database. This design uses the PostgreSQL database embedded within the vRealize Automation appliance.

Table 2-196. vRealize Orchestrator Database Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-09	The embedded vRealize Orchestrator instance uses the PostgreSQL database embedded in the vRealize Automation appliance.	Using the embedded PostgreSQL database provides the following advantages: ■ Performance improvement ■ Design simplification	None.

vRealize Orchestrator Plug-Ins

Plug-ins allow you to use vRealize Orchestrator to access and control external technologies and applications. Exposing an external technology in a vRealize Orchestrator plug-in allows you to incorporate objects and functions in workflows that access the objects and functions of the external technology. The external technologies that you can access using plug-ins can include virtualization management tools, email systems, databases, directory services, and remote control interfaces. vRealize Orchestrator provides a set of standard plug-ins that allow you to incorporate such technologies as the vCenter Server API and email capabilities into workflows.

In addition, the vRealize Orchestrator open plug-in architecture allows you to develop plug-ins to access other applications. vRealize Orchestrator implements open standards to simplify integration with external systems. For information on developing custom content, see *Developing with VMware vRealize Orchestrator.*

vRealize Orchestrator and the vCenter Server Plug-In

You can use the vCenter Server plug-in to manage multiple vCenter Server instances. You can create workflows that use the vCenter Server plug-in API to automate tasks in your vCenter Server environment. The vCenter Server plug-in maps the vCenter Server API to the JavaScript that you can use in workflows. The plug-in also provides actions that perform individual vCenter Server tasks that you can include in workflows.

The vCenter Server plug-in provides a library of standard workflows that automate vCenter Server operations. For example, you can run workflows that create, clone, migrate, or delete virtual machines. Before managing the objects in your VMware vSphere inventory by using vRealize Orchestrator and to run workflows on the objects, you must configure the vCenter Server plug-in and define the connection parameters between vRealize Orchestrator and the vCenter Server instances you want to orchestrate.

You can configure the vCenter Server plug-in by using the vRealize Orchestrator configuration interface or by running the vCenter Server configuration workflows from the vRealize Orchestrator client. You can configure vRealize Orchestrator to connect to your vCenter Server instances for running workflows over the objects in your vSphere infrastructure.

To manage objects in your vSphere inventory using the vSphere Web Client, configure vRealize Orchestrator to work with the same vCenter Single Sign-On instance to which both vCenter Server and vSphere Web Client are pointing. Also, verify that vRealize Orchestrator is registered as a vCenter Server extension. You register vRealize Orchestrator as a vCenter Server extension when you specify a user (user name and password) who has the privileges to manage vCenter Server extensions.

Table 2-197. vRealize Orchestrator vCenter Server Plug-In Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-10	Configure the vCenter Server plug-in to control communication with the vCenter Servers.	Required for communication to vCenter Server instances, and as such required for workflows.	None.

vRealize Orchestrator Scalability

vRealize Orchestrator supports both scale-up and scale-out scalability.

Scale Up

A single vRealize Orchestrator instance allows up to 300 concurrent workflow instances in the running state. Workflow instances that are in the waiting or waiting-event 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 single 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. Using a vRealize Orchestrator cluster allows you to increase the number of concurrent running workflows, but not the number of managed inventory objects. When clustering a vRealize Orchestrator server, choose the following cluster type:

- An active-active cluster with up to five active nodes. VMware recommends a maximum of three active nodes in this configuration.

In a clustered vRealize Orchestrator environment you cannot change workflows while other vRealize Orchestrator instances are running. Stop all other vRealize Orchestrator instances before you connect the vRealize Orchestrator client and change or develop a new workflow.

Table 2-198. vRealize Orchestrator Scale Out Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-CMP-VRO-11	Configure vRealize Orchestrator in an active-active cluster configuration.	Active-active clusters allow both vRealize Orchestrator servers to equally balance workflow execution.	When you cluster the vRealize Automation appliances, the vRealize Orchestrator instances embedded within 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 2-42. Business Continuity in the SDDC Layered Architecture
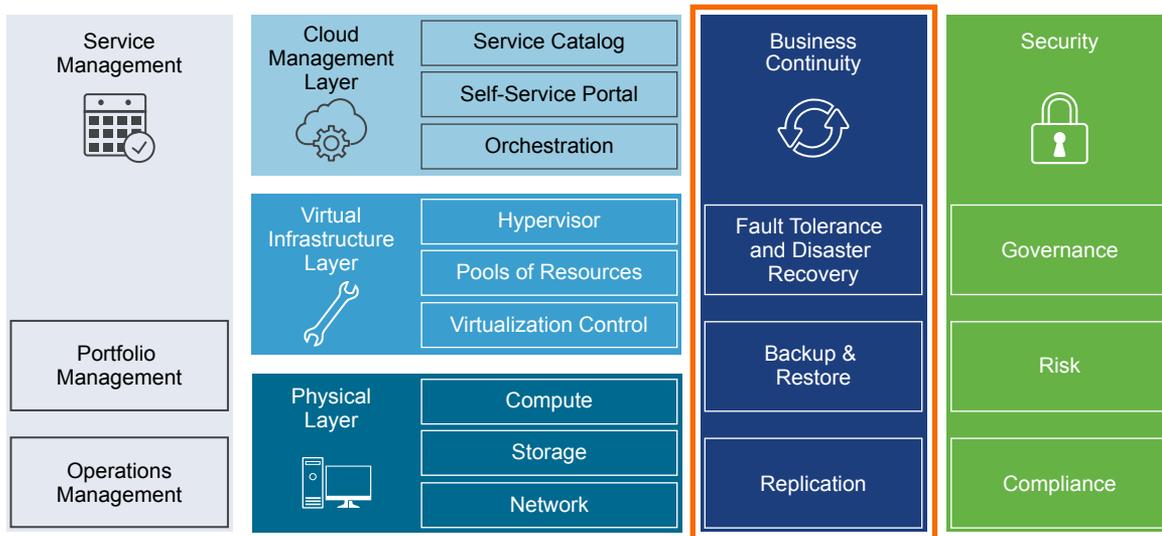
Data Protection and Backup Design

For continuous operation of the SDDC if the data of a management application is compromised, design data protection of the management components in your environment.

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 the vSphere Storage APIs for Data Protection (VADP). You can use any VADP-compatible backup solution. Adapt and apply the design decisions to the backup software you use.

Table 2-199. Design Decisions About 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 VMware vCenter Server layer. Because the VADP compatible backup solution is connected to the Management vCenter Server, it can access all management ESXi hosts, and can detect the virtual machines that require backups.

Figure 2-43. Data Protection Logical Design
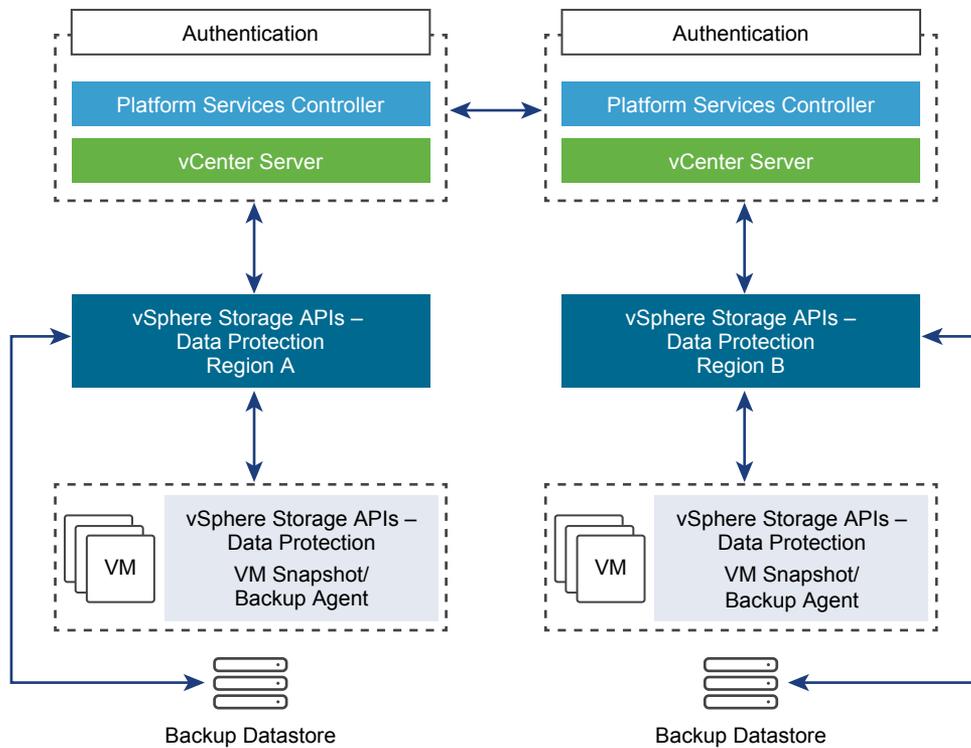
Backup Datastore for Data Protection

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 6 TB capacity without considering deduplication capabilities.

Table 2‑200. Backup Datastore Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-003	Allocate a dedicated datastore for the VADP-compatible backup solution and the backup data according to NFS Physical Storage Design.	■ Emergency restore operations are possible even when the primary 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 6 TB on-disk.	Secondary storage handles the backup of the management stack of a single region. The management stack consumes approximately 6 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 2-201. Design Decisions About 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 have the same visibility of 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 2-202. Backup Schedule Design Decisions

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-BKP-007	Schedule daily backups.	You can recover virtual machine data that is at most a day old.	You lose data that changed since the last backup 24 hours ago.
SDDC-OPS-BKP-008	Schedule backups outside the production peak times.	Backups occur when the system is under the lowest load. Make sure that backups are completed in the shortest time possible with the smallest risk of errors.	You must schedule backup to start between 8:00 PM and 8:00 AM or until the backup jobs are complete, whichever comes first.

Retention Policies

Retention policies are properties of a backup job. If you group virtual machines by business priority, you can set the retention requirements according to the business priority.

Table 2-203. Design Decision About 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 2-204. Design Decisions About 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 svc-bck-vcenter 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 svc-bck-vcenter service account in vCenter Server.	■ Simplifies and standardizes the deployment of the service account across all vCenter Server instances in the same vSphere domain. ■ Provides a consistent authorization layer.	All vCenter Server instances must be in the same vSphere domain.

Component Backup Jobs for Data Protection

You can configure backup for each SDDC management component separately. This design does not suggest a requirement to back up the entire SDDC.

Some products can perform internal configuration backups. Use those products in addition to the whole VM component backups as appropriate.

Table 2-205. Design Decision About 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 destructive method to achieve a similar restoration of functionality.	You must provide space on an SFT or FTP server to store the NSX configuration backups.

Site Recovery Manager and vSphere Replication Design

To support disaster recovery (DR) in the SDDC, you protect vRealize Operations Manager and vRealize Automation by using VMware Site Recovery Manager and VMware vSphere Replication. When failing over to a recovery region, these management applications continue 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 2-206. 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 2-44. Disaster Recovery Logical Design
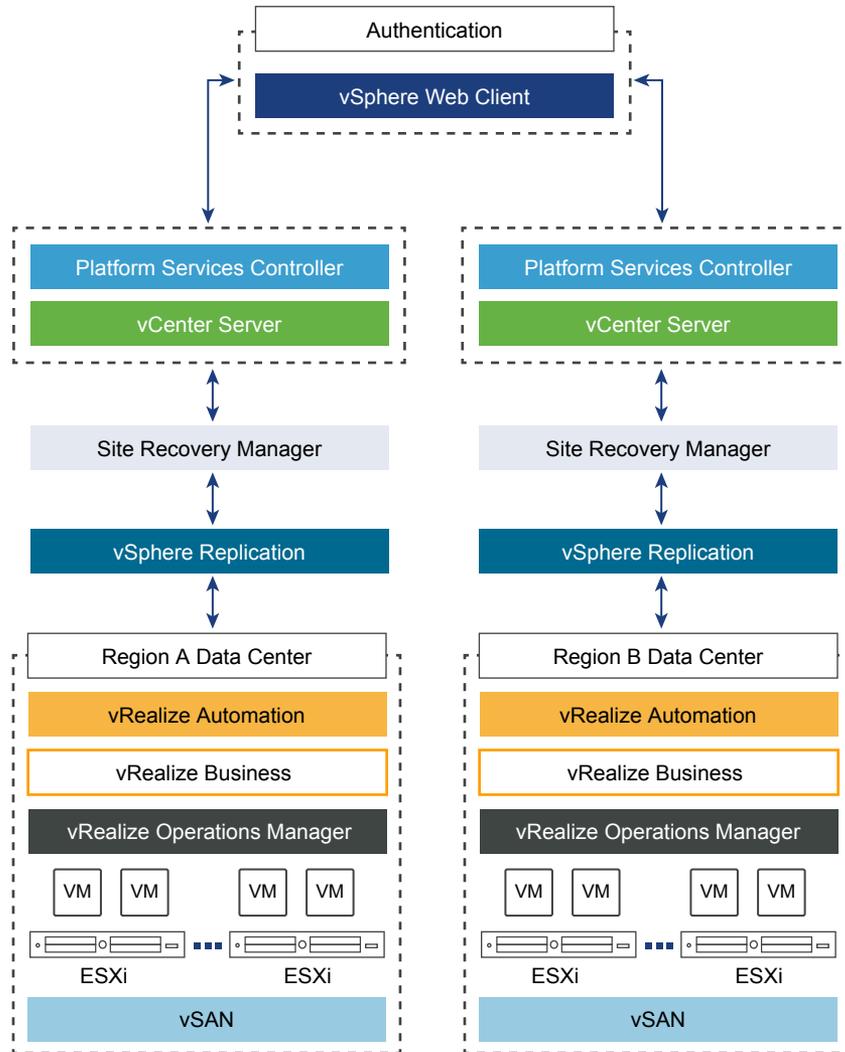
Deployment Design for Site Recovery Manager

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 2-207. Design Decisions About 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 2-208. 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 2012 R2

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 2-209. SDDC Nodes with Failover Support

Management Component	Node Type	Number of Nodes
Cloud Management Platform	vRealize Automation Appliance	2
	vRealize IaaS Web Server	2

Table 2‑209. SDDC Nodes with Failover Support (Continued)

Management Component	Node Type	Number of Nodes
	vRealize IaaS Management Server	2
	vRealize IaaS DEM	2
	Microsoft SQL Server	1
	vRealize Business for Cloud Appliance	1
vRealize Operations Manager	vRealize Operations Manager Master	1
	vRealize Operations Manager Master Replica	1
	vRealize Operations Manager Data	1

You must protect a total of 13 virtual machines.

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 2‑210. Design Decisions About 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 GbE	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 2-209 ■ 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 IPv4 address.

Figure 2-45. Logical Network Design for Cross-Region Deployment with Application Virtual Networks
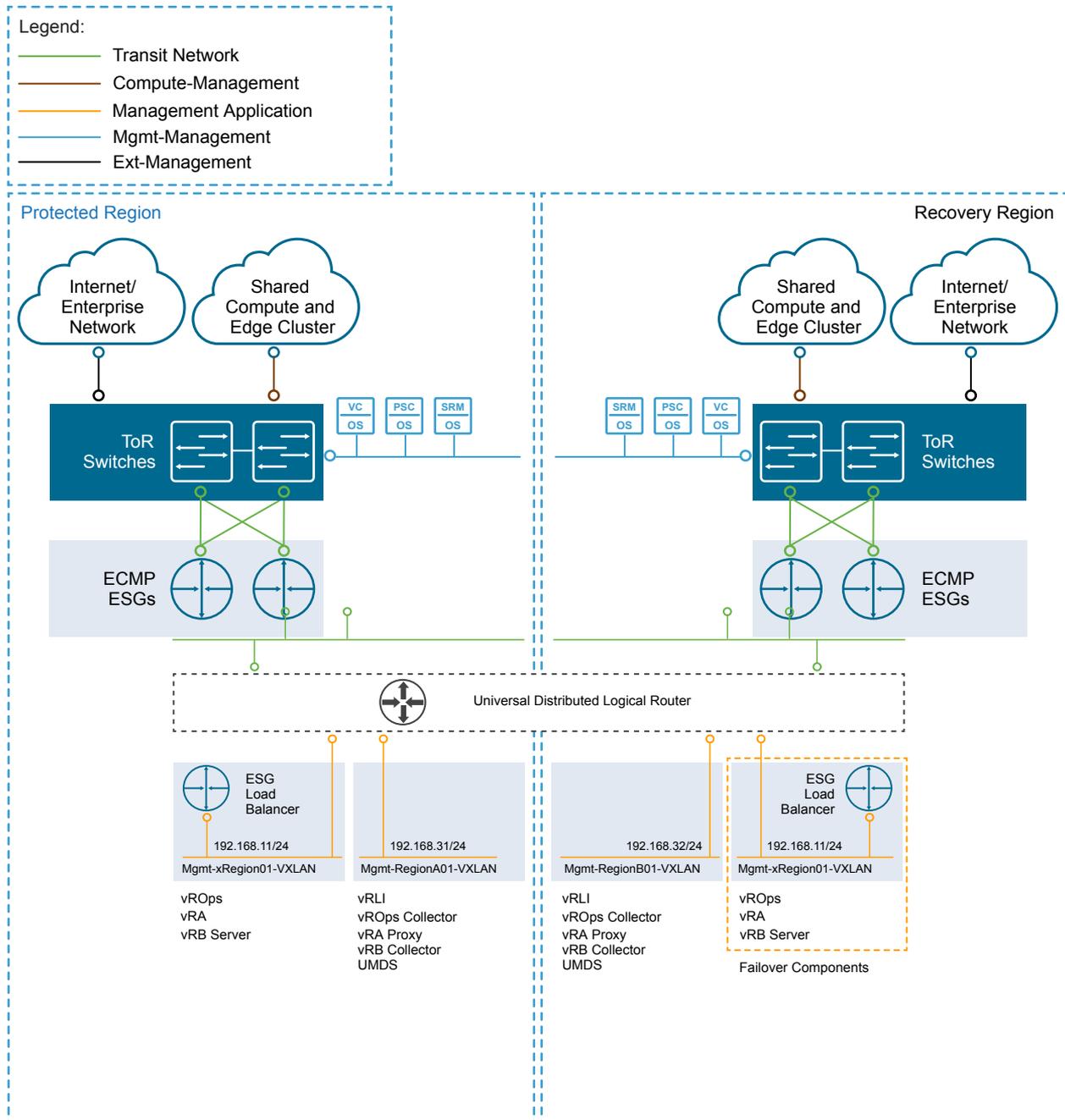
The IPv4 subnets (orange networks) are routed in the vSphere management network of each region. Nodes on these network segments are reachable from within the SDDC. IPv4 subnets, such as the subnet that contains the vRealize Automation primary components, overlap across a region.

Make sure that only the active IPv4 subnet is propagated in the region and beyond. The public facing Ext-Management network of both regions (gray networks) is reachable by SDDC users and provides connection to external resources, such as Active Directory or DNS. For information about the design of AVNs, see Application Virtual Network.

NSX Edge devices provide the load balancing functionality, each device fronting a network that contains the protected components of all management applications. In each region, you use the same configuration for the management applications and their Site Recovery Manager shadow. Active Directory and DNS services must be running in both the protected and recovery regions.

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 2‑211. Design Decisions About 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 svc-srm in vCenter Server for application-to-application communication from Site Recovery Manager with vSphere.	Provides the following access control features: ■ Site Recovery Manager accesses vSphere with the minimum set of permissions that are required to perform disaster recovery failover orchestration and site pairing. ■ In the event of a compromised account, 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 svc-vr in vCenter Server for application-to-application communication from vSphere Replication with vSphere	Provides the following access control features: ■ vSphere Replication accesses vSphere with the minimum set of permissions that are required to perform site to site replication of virtual machines. ■ In the event of a compromised account, 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 2‑211. Design Decisions About Authorization and Authentication Management for Site Recovery Manager and vSphere Replication (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-007	Use global permissions when you create the svc-srm service account in vCenter Server.	■ Simplifies and standardizes the deployment of the service account across all vCenter Server instances in the same vSphere domain. ■ Provides a consistent authorization layer. ■ If you deploy more Site Recovery Manager instances, reduces the efforts in connecting them to the vCenter Server instances.	All vCenter Server instances must be in the same vSphere domain.
SDDC-OPS-DR-008	Use global permissions when you create the svc-vr service account in vCenter Server.	■ Simplifies and standardizes the deployment of the service account across all vCenter Server instances in the same vSphere domain. ■ Provides a consistent authorization layer. ■ If you deploy more vSphere Replication instances, reduces the efforts in connecting them to the vCenter Server instances.	All vCenter Server instances must be in the same vSphere domain.

Encryption

Replace default self-signed certificates with a CA-signed certificate to provide secure access and communication for vSphere Replication and Site Recovery Manager.

Table 2‑212. Design Decision About CA-Signed Certificates for Site Recovery Manager and vSphere Replication

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-009	Replace the default self-signed certificates in each Site Recovery Manager instance with a CA-signed certificate.	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-010	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 via Storage Replication Adapters (SRAs) with Site Recovery Manager

■ vSphere Replication with Site Recovery Manager.

Table 2-213. Design Decisions About Replication Technology

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-011	Use vSphere Replication in Site Recovery Manager as the protection method for virtual machine replication.	- Allows 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 isolate to its own dedicated portgroup 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 2-214. Design Decisions About vSphere Replication Networking

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-012	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-13	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 utilization across the inter-site link (ISL).	- Static routes on the ESXi hosts are required. - A sufficient number of VLANs are available within each cluster and should be used for traffic segregation. - Host Profiles must be managed on a per-host level.

Table 2-214. Design Decisions About vSphere Replication Networking (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-014	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-015	Attach a vNIC 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-016	When using two availability zones, deploy an additional vNIC on the vSphere Replication VM to the second availability zone's vSphere Replication port group.	▪ 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 bound to each port group to receive data.	During a disaster recovery scenario, you must maintain the IP address of the Incoming Storage Traffic depending on 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 grows as test VMs write to their disks. Replication from the protected region to the recovery region continues during this time. The snapshots that are created 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 2-215. Compute Resources for a vShere Replication 4 vCPU Node

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 13 virtual machines. For information about the node configuration of the management components per region that is used in the calculations, see Table 2-209.

Table 2-216. Design Decisions About vSphere Replication Deployment and Size

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-017	Deploy each vSphere Replication appliance in the management cluster.	All management components must be in the same cluster.	None
SDDC-OPS-DR-018	Deploy each vSphere Replication appliance using the 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	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.

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-019	Use Site Recovery Manager and vSphere Replication together to automate the recovery of the following management components: ■ vRealize Operations analytics cluster ■ vRealize Automation Appliance instances ■ vRealize Automation IaaS components ■ vRealize Business Server	■ Provides an automated run book for the recovery of the management components in the event of a disaster. ■ 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 2‑217. Design Decisions About vSphere Replication

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-020	Do not enable guest OS quiescing in the policies for the management virtual machines in vSphere Replication.	Not all management virtual machines support the use of guest OS quiescing. Using the quiescing operation might result in an outage.	The replicas of the management virtual machines that are stored in the target region are crash-consistent rather than application-consistent.
SDDC-OPS-DR-021	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.

Table 2-217. Design Decisions About vSphere Replication (Continued)

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-022	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-023	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 virtual machine startup order.

▪ All priority 1 virtual machines are started before priority 2 virtual machines.

▪ All priority 2 virtual machines are started before priority 3 virtual machines.

▪ All priority 3 virtual machines are started before priority 4 virtual machines.

▪ All priority 4 virtual machines are started before priority 5 virtual machines.

▪ You can additionally set startup order of virtual machines within each priority group.

You can configure the following timeout parameters:

▪ Response time, which defines the time to wait after the first virtual machine powers on before proceeding to the next virtual machine in the plan.

▪ Maximum time to wait if the virtual machine fails to power on before proceeding to the next virtual machine.

You can adjust response time values as necessary during execution of the recovery plan test to determine the appropriate response time values.

Table 2-218. Startup Order Design Decisions for Site Recovery Manager

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-024	Use a prioritized startup order for vRealize Operations 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 Operations Manager services are restored in the target of 4 hours.	■ You must have VMware Tools running on each vRealize Operations 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-025	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 is 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 2-219. Recovery Plan Test Network Design Decision

Decision ID	Design Decision	Design Justification	Design Implication
SDDC-OPS-DR-026	Use the target recovery production network for testing.	The design of the application virtual networks supports their use as recovery plan test networks.	During recovery testing, a management application is not reachable 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.