vCloud NFV OpenStack Edition Reference Architecture

VMware vCloud NFV OpenStack Edition 3.1

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 vCloud NFV OpenStack Edition Reference Architecture

1

This reference architecture provides guidance for designing and creating a greenfield Network Functions Virtualization (NFV) platform by using VMware vCloud® NFV™ OpenStack Edition.

This document describes the high-level design principles and considerations when implementing an environment that is based on vCloud NFV OpenStack Edition. It also provides example scenarios to help you understand the platform capabilities.

Intended Audience

This document is intended for telecommunications and solution architects, sales engineers, field consultants, advanced services specialists, and customers who are responsible for the virtualized network services (VNFs) and the NFV environment on which they run.

Introduction to vCloud NFV OpenStack Edition

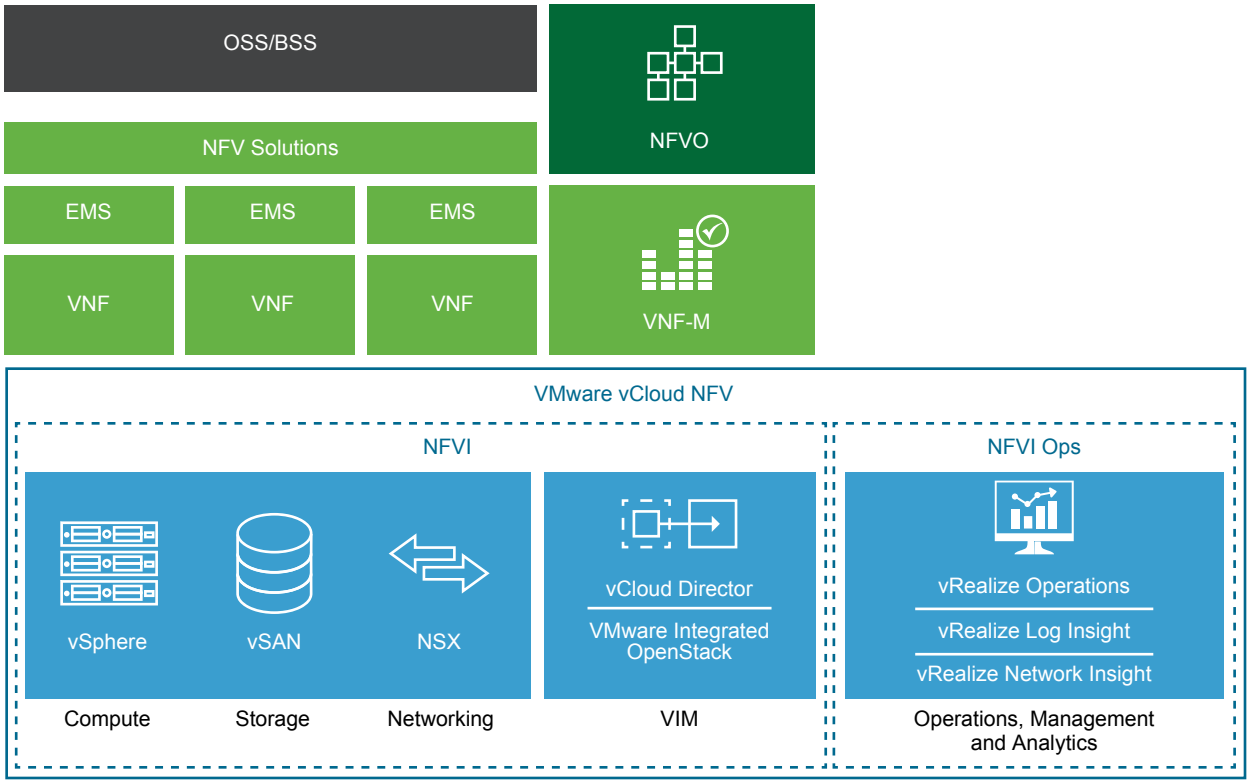
2

VMware vCloud NFV OpenStack Edition combines a carrier grade NFV infrastructure with VMware® Integrated OpenStack as the NFV Virtualized Infrastructure Manager (VIM). This version of the vCloud NFV OpenStack Edition platform combines the OpenStack API with stable and supportable vCloud NFV Infrastructure (NFVI). This way, vCloud NFV OpenStack Edition provides a platform to support Communication Service Providers (CSPs) in realizing the goal for network modernization and business transformation.

The vCloud NFV OpenStack Edition platform implements a modular design with abstractions that enable multi-vendor, multi-domain, and hybrid physical, and virtual execution environments. The IaaS layer that is exposed through the upstream OpenStack Queens release, provides a CI/CD environment for workload life cycle management. The platform also delivers an automation framework to interoperate with external functions for service orchestration and management.

In addition to the core NFV infrastructure components for compute, storage, networking, and VIM, the vCloud NFV OpenStack Edition platform includes a fully integrated suite for operational intelligence and monitoring. This suite can be used to further enhance the runtime environments with workflows for dynamic workload optimization and proactive issue avoidance.

Figure 2-1. vCloud NFV Components



The vCloud NFV OpenStack Edition components, their interactions with each other, and how they meet CSP requirements, are described in this reference architecture.

Acronyms and Definitions

3

vCloud NFV uses a specific set of acronyms that apply to the NFV technology and the telco industry.

Table 3-1. General Acronyms

Abbreviation	Description
BFD	Bidirectional Forwarding Detection, for failure detection on the transport links.
DPDK	Data Plane Development Kit, an Intel led packet processing acceleration technology.
MTTR	Mean Time to Repair.
MTTU	Mean Time to Understand.

Table 3-2. NFV Acronyms

Abbreviation	Description
CCP	Centralized Control Plane in the NSX-T Data Center architecture.
CNF	Container Network Function, executing within a Kubernetes environment.
LCP	Local Control Plane r in the NSX-T Data Center architecture.
MANO	Management and Orchestration components, a term originating from the ETSI NFV architecture framework.
NFVI	Network Functions Virtualization Infrastructure.
NFV-OI	NFV Operational Intelligence.
N-VDS (E)	Enhanced mode when using the NSX-T Data Center N-VDS logical switch. This mode enables DPDK for workload acceleration.
N-VDS (S)	Standard mode when using the NSX-T Data Center N-VDS logical switch.
VIM	Virtualized Infrastructure Manager.
VNF	Virtual Network Function, executing in a virtual machine.

Table 3-3. Telco Acronyms

Abbreviation	Description
HSS	Home Subscriber Server in the mobile evolved packet core 4G architecture.
MVNO	Mobile Virtual Network Operator.
PCRF	Policy, Charging and Rating Function, in the mobile evolved packet core 4G architecture.
PGW	Packet Gateway in the mobile evolved packet core 4G architecture.

Table 3-3. Telco Acronyms (Continued)

Abbreviation	Description
SGW	Service Gateway in the mobile evolved packet core 4G architecture.
SBC	Session Border Controller used in voice telephone for control and data plane communications between clients.
VPP	Vector Packet Processing

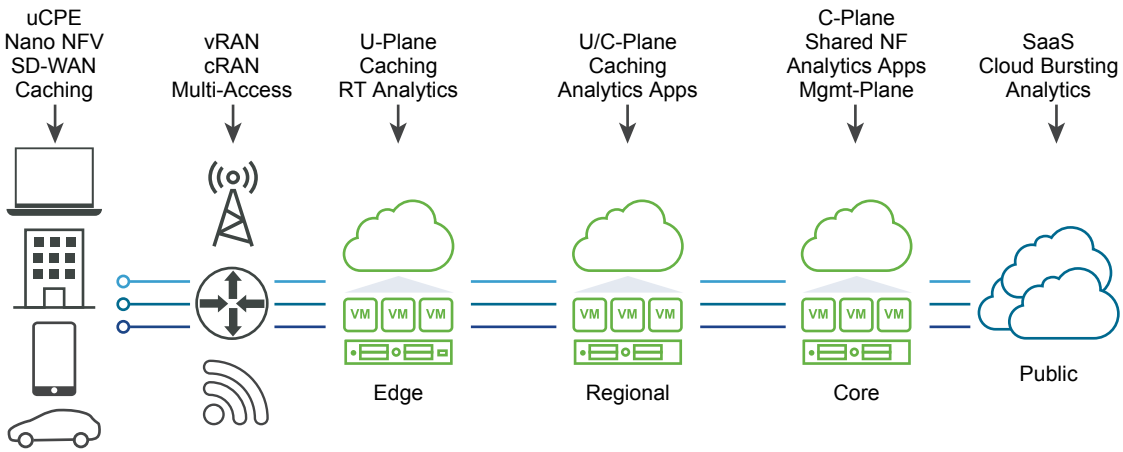
Reference Environment

4

5G services will require a mixture of low-latency, high throughput, and high user densities and concurrences. The distribution of functional components will require a more sophisticated service delivery model.

The network is transforming into a mixture of highly distributed functions together with centralized functions. This way, the network is moving away from the typical centralized models in service delivery. There is also an emerging paradigm shift with employing third-party IaaS, PaaS, and SaaS offerings from public cloud providers.

Figure 4-1. Reference Environment



The highly distributed topology of the network will support the next-generation service characteristics in distribution and composition. It will also require a new way of managing the network and infrastructure resources. The number of the services that are spanning the industry verticals is exploding exponentially. Today's end point ranging in fixed and mobile offers will grow into the billions with IoT connections. The highly distributed edge sites are projected to be in the thousands, regional sites in the 100's, core sites in the 10's, and a large variety of public cloud provider sites.

NFV and Software Defined Networking (SDN) transformations will introduce complex interconnections between end-points such as branches, small offices, connected cars, and IoT gateways to private data centers and public cloud providers. By definition, this reference environment and its transformation are not only a technical challenge but also impacts the business and operating processes.

Reference Environment Requirements

The reference environment places strict requirements for service placement and management to achieve optimal performance.

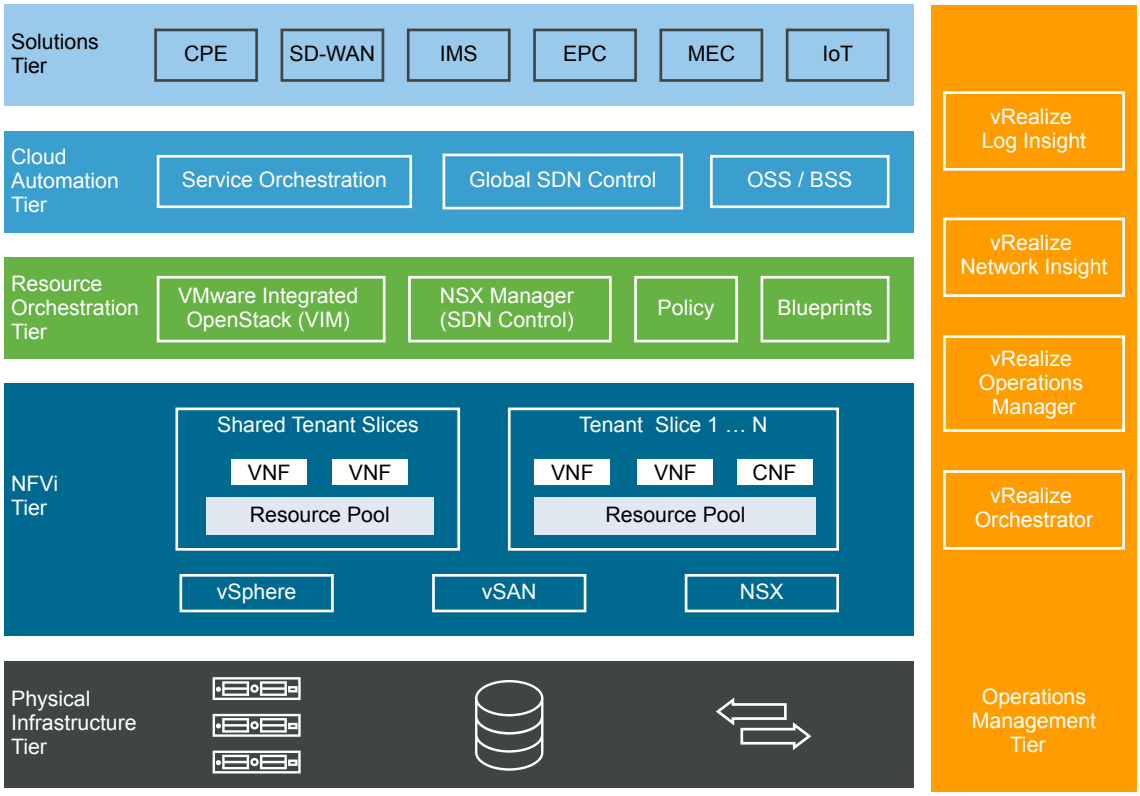
- Federation options. The reference environment topology offers a diverse set of federation options for end-points, private and public clouds, each with distinct ownership and management domains. Virtualized end-points provide better control and manageability, however they are not suitable for all types of use cases. Likewise, service functions can be distributed and managed across private and public clouds.
- Disaggregated functions. Services are highly disaggregated so that control, data, and management planes can be deployed across the distributed topology. Edge clouds offer the performance advantages of low latency and data plane intensive workloads. While control and management plane components can be centralized with a regional and global scope.
- Functional isolation. Network slicing provides network and service isolation across different tenancy models in the reference environment. However, resource management considerations need to be made for shared network functions such as DNS, policy, authentication, and so on.
- Service placement. The highly distributed topology allows for flexibility in the workload placement. Making decisions based on proximity, locality, latency, analytical intelligence, and other EPA criteria are critical to enable an intent-based placement model.
- Workload life cycle management. Each cloud is elastic with workload mobility and how applications are deployed, executed, and scaled. An integrated operations management solution can enable an efficient life cycle management to ensure service delivery and QoS.
- Carrier grade characteristics. Because CSPs deliver services that are often regulated by local governments, carrier grade aspects of these services, such as high availability and deterministic performance are also important.
- NFVI life cycle (patching and upgrades). The platform must be patched and upgraded by using optimized change management approaches for zero to minimal downtime.

NFV Reference Model

NFV is an architectural framework that is first developed by the ETSI NFV Industry Specification Group. The framework provides a reference model where network functions are delivered through software virtualization with commercial off-the-shelf (COTS) hardware. This way, NFV moves away from the proprietary, purpose-built hardware that is dedicated to a single service. The result is a network that is agile, resilient, and equipped to deliver high quality of services. The NFV framework defines functional abstractions and interactions between the building blocks. Some of these abstractions are already present in current deployments, while others must be added to support the virtualization process and operation.

The following diagram shows the reference model for an NFV environment with clear functional abstractions and interactions in a tiered approach.

Figure 4-2. Layered Abstractions of the NFV Environment



Physical Tier

Represents compute hardware, storage, and physical networking as the underlying pool of shared resources. In addition, there are numerous other physical network devices such as switches, routers, EMS, and so on, making the execution ecosystem a hybrid virtual and physical topology.

NFVI Tier

The lowest tier of the vCloud NFV OpenStack Edition platform. It delivers the virtualization run-time environment with network functions and resource isolation for VM and container-based workloads. In NFVI, virtualized compute, storage, and networking are delivered as an integrated solution through vSphere, vSAN, and NSX-T Data Center. Isolated resources and networks can be assigned to a tenant slice, which is a runtime isolated partition delivering services. Tenant slices can be dedicated to a tenant or shared across tenants. The NFVI is optimized and adjusted for telco-class workloads to enable the delivery of quality and resilient services. Infrastructure high availability, performance, and scale considerations are built into this tier for performance optimization.

Resource Orchestration Tier

It provides resource management capabilities to the NFVI tier. This way, the NFVI can deliver a flexible infrastructure-as-code for life cycle management of workloads, network management, and resource management. The resource orchestration tier is responsible for controlling, managing, and monitoring the NFVI compute, storage, and network hardware, the software

	for the virtualization layer, and the virtualized resources. The VIM module manages the allocation and release of virtual resources, and the association of virtual to physical resources, including resource optimization. VIM also maintains the inventory of NFVI, including the linkage and relationship between components as they relate to an instance of a VNF or CNF workload. This way, VIM allows for monitoring in the context of a single VNF.
Cloud Automation Tier	The service management and control functions which bridge the virtual resource orchestration and physical functions to deliver services and service chains. It is typically a centralized control and management function, including embedded automation and optimization capabilities.
Solutions Tier	The multi-domain ecosystem of software virtual functions as native VM or containerized functions. Such functions are composed in complex solutions to enable service offers and business models that CSP customers consume. Solutions can range from small branch office functions to a fully evolved packet core that is delivered as tenant slices across multiple clouds.
Operations Management Tier	An integrated operational intelligence for infrastructure day 0, 1, and 2 operations that spans across all other tiers. The functional components within the operations management tier provide topology discovery, health monitoring, alerting, issue isolation, and closed-loop automation.

Key Customer Objectives

The goal of network modernization is to drive greater classes of service innovation and timely enablement. Following are key objectives that CSPs are considering as they transform their networks and design for new business and operational models.

Fixed Mobile Convergence	As networks evolved through 2G and 3G generations, the voice and data network architectures were separated, particularly circuit-switched, and packet-switched networks. As networks evolved, the CSPs went towards an all IP network, therefore the convergence of Fixed and Mobile

	networking. The environments for voice mostly share the same core networking components with different access networks. The scale, performance, and management of such converged networks are more critical than before.
Data Intensive Workload Acceleration	The demand for throughput has increased exponentially with smart devices and immersive media services. The networking and compute expenditures continue to grow to meet such demands in traffic throughput. Acceleration technologies like DPDK, VPP, hardware offload, for example, are at the forefront to reduce OpEx for data intensive applications.
Cloud-Native and Hybrid Form-Factor Execution Environments	Cloud-Native approaches are dictating a new CI/CD paradigm and micro services application architectures. Container technology is the new light-weight execution environment for such micro services and delivery. While the fine-grained abstraction of applications might be a good fit for control plane functions in the next-generation architecture, the user plane functions are expected to execute as native VM functions. This requires the cloud infrastructure environment to be heterogeneous enabling such hybrid execution environments for native VM and containerized applications.
Distributed Clouds	To meet the increased bandwidth and low-latency requirements, network designs are expanding the centralized compute models to distributed edge computing models. Certain level of distribution already exists in regional and core data centers, however further edge distribution will be necessary to control traffic backhauling and to improve latencies. In conjunction, VNFs are disaggregating to distribute data plane functions at the edges of the network whereas control functions are centralized. Service distribution and elasticity will be vital part of the network design consideration.
Network Slicing	Network slicing is a way for cloud infrastructure to isolate resources and networking to control the performance and security for workloads that are executing on the same shared pool of physical infrastructure. With distributed topologies, the concept of network slicing furthermore stretches across multiple cloud infrastructures, including access, edge, and core virtual and physical infrastructures. Multi-tenancy leverages such resource isolation to deploy and optimize VNFs to meet customer SLAs.
Dynamic Operational Intelligence	The cloud infrastructures will have to become adaptive to meet the needs of workloads. Right-sizing the environment and dynamic workload optimizations, including initial placement, will be part of the continuous automation orchestration. The cloud infrastructure environments will require integrated operational intelligence to continuously monitor, report, and action in a timely manner with prescriptive and predictive analytics.

Policy-Based Consistency and Management	Model driven approaches will play a key role in the modern cloud infrastructures. Resource modeling, runtime operational policies, or security profiles, declarative policies and movement of policies with workloads, onboarding, and so on, will ensure consistency and ease of management.
Carrier Grade Platform	The cloud infrastructure environment will have to meet the strict requirements for availability, fault tolerance, scale, and performance. Security will be necessary across the transport, data, and workload dimensions. The mobility of workloads across distributed clouds introduces a new challenge for its authenticity and integrity.

Architectural Framework and Components

5

This section explores the overall framework for the vCloud NFV OpenStack Edition platform architecture, including the key stakeholders, conceptual architecture environment, logical architecture, and components of the vCloud NFV OpenStack Edition platform. The reference architecture design principles set the framing for the core and analytics-enabled designs that are discussed in this document.

This chapter includes the following topics:

Key Stakeholders

The reference architecture considers key stakeholders that are involved in the end-to-end service management, life cycle management, and operations.

Cloud provider	The CSP operations personnel who are responsible for provisioning and on-boarding all day 0 and day 1 functions to enable services for target customers and tenants.
Consumer	Consumer. The end user who is consuming the services that the tenants provide. For example, IoT devices, mobile handsets, API consumers, and MVNO.
Customer	The enterprise or entity who owns the business relationship with the CSP. The customer might be an internal line of business such as fixed line and mobile services and can also be an external enterprise.

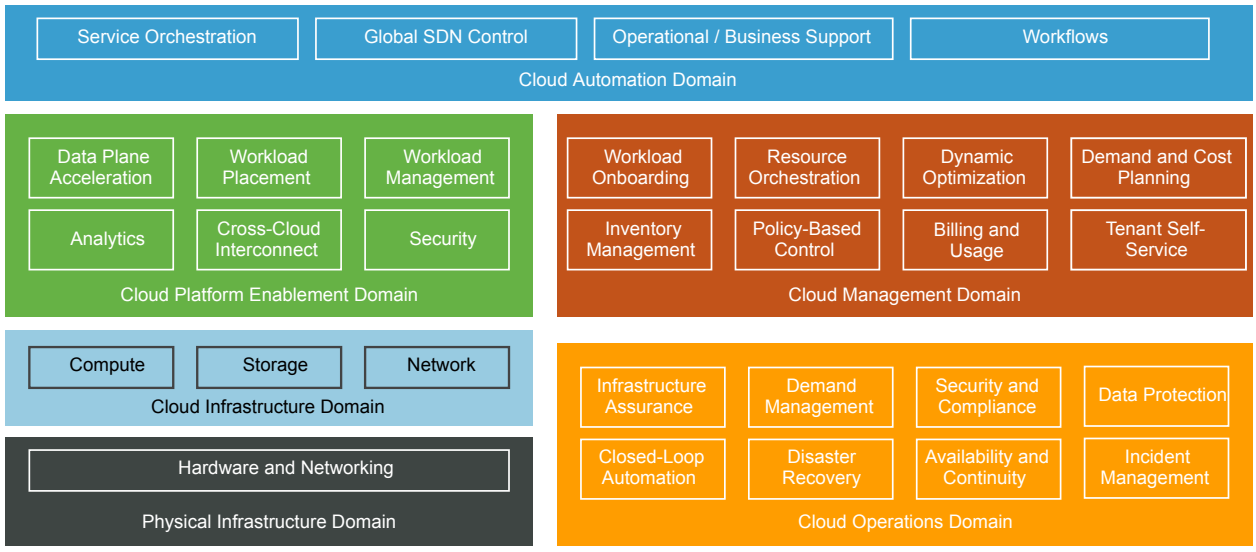
Tenant	The various classes of services (offers) that a customer provides to their consumers. Each tenant is represented as a resource slice, hence a customer can have one or more tenants. A mobile line of business can offer slices to a MVNO customer, for example a tenant for voice services and a tenant for data services.
Operations Support	The operations management process and team that ensure the services are operating to meet the promised stringent SLAs.
Network Planning	The operations management planning function that is responsible for the resource and VNF capacity and forecasting, and new data center designs.
Security Operations	Security operations function that is responsible for all aspects of security, network, application, and data.

Conceptual Architecture

Cloud infrastructure-based computing is the next-generation standard in modernizing the CSP networks as they evolve to 5G architectures, services, and agile delivery. The shared infrastructure with complete softwarization of network functions and applications provide greater advantages in cost, performance, and agility.

The modernization of the CSP infrastructure requires a complex ecosystem of solutions and functions delivering to a pre-set business and operating model. The cloud infrastructure modernization changes not only the business model in service agility and metered revenue models, but also challenges the silo operating model. The following figure shows the conceptual view of the various domains, capabilities, and their interactions that need consideration in the modernization of networks and business and operational models.

Figure 5-1. Conceptual Architecture



Cloud Automation Domain

Centralizes the overall service management functions such as service definitions, composition, onboarding, life cycle management, and support. This domain hosts functions such as an NFV-O that is responsible for service blueprinting, chaining, and orchestration across multiple cloud infrastructure environments. Next to NFV-O are the global SDN control functions that are responsible for stitching and managing physical and overlay networks for cross-site services. Real-time performance monitoring can be integrated into the SDN functions to dynamically optimize network configurations, routes, capacity, and so on.

The successful cloud automation strategy implies full programmability across other domains.

Cloud Platform Enablement Domain

Extends a set of platform capabilities from the cloud infrastructure that the automation domain, VNFs, their managers, and other core components can leverage. Example enablement capabilities include:

- Analytics to ingest VNF metrics that can be correlated with infrastructure metrics for smarter context and insights.
- Workload placement to determine the right location for a workload depending on available resources, class of resources, and feature capabilities such as data intensive acceleration.
- Workload acceleration using DPDK and SR-IOV for data intensive VNFs.
- Security for network, data, and workloads.

Cloud Management Domain

Plays a critical role across many different dimensions. More fundamentally, it provides a templated and prescriptive workload management set of capabilities that the automation layer can use to program and orchestrate service on-demand and with agility. Service onboarding models can be turned into fully zero-touch provisioning and exposed to tenants through a self-service portal. Business models such as metered billing can be enabled as a catalog of services and tariffs.

Once services and workloads are onboarded, the management domain also needs to ensure dynamic optimization such as workload rebalancing or capacity growth or shrink to maintain agreed SLAs. Such optimizations need to integrate with the cloud operations domain for real-time usage and performance intelligence. Polices, including platform awareness, NUMA affinity, host affinity, restart-sequences, are necessary for efficient optimization.

Cloud Operations Domain

Ensures that the operational policies and SLAs are being met by continuous data collection, correlation, and analytics. Infrastructure assurance is a key component of this domain. Intelligence can be tied into a closed-loop workflow that can be integrated in the automation domain for proactive issue avoidance, for example, triggering a trouble ticket incident management system.

In addition, other functions for day 2 operations such demand and capacity planning, security and compliance, high availability, and disaster recovery are necessary to ensure availability and integrity across the cloud infrastructure environments.

Cloud Infrastructure Domain

The core virtualization domain providing resource abstraction for compute, storage, and networking and their orchestration through a VIM to allocate, control, and isolate with full multi-tenancy and platform-awareness.

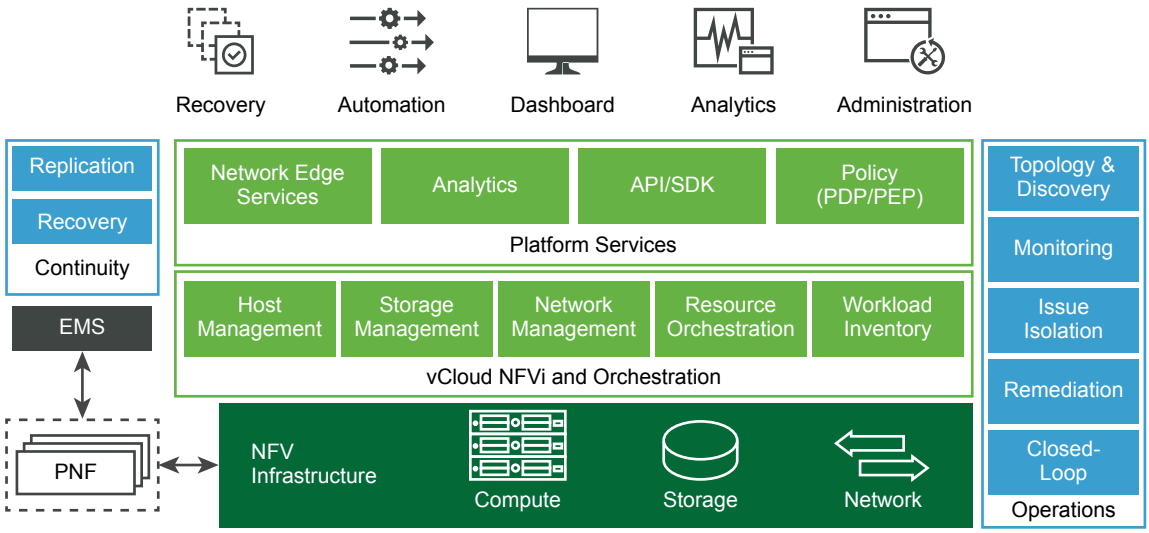
Logical Architecture and Components

The vCloud NFV OpenStack Edition platform implements the conceptual architecture that is outlined and defined at a high level through the logical building blocks and core components.

The VMware vCloud NFV OpenStack platform is an evolution of the VMware NFV solution, based on extensive customer deployment and the continued development of standards organizations such as the European Telecommunications Standards Institute (ETSI). The vCloud NFV OpenStack Edition platform provides a comprehensive, service-oriented solution, leveraging a cloud computing model that allows ubiquitous, programmatic, on-demand access to a shared pool of compute, network, and storage resources. The solution is integrated with holistic operations management and service assurance capabilities, empowering the CSP to rapidly deliver services while ensuring their quality. With a fully integrated VIM, the same vCloud NFV OpenStack Edition infrastructure delivers a myriad of telecommunications use cases and facilitates reusability of the service catalog based VNFs.

The following diagram maps the conceptual architecture to a logical view for the vCloud NFV OpenStack Edition platform.

Figure 5-2. Logical Architecture



The vCloud NFV OpenStack Edition platform delivers a complete integrated solution that has been rigorously tested to ensure compatibility, robustness, and functionality. The components that build the solution are currently deployed across many industries and scenarios. The vCloud NFV OpenStack Edition software components can be used in various ways to construct a comprehensive, end-to-end solution that meets the business goals of CSPs. This document discusses one way in which components can be used to create a vCloud NFV architecture.

Logical Architecture Components

The vCloud NFV platform consists of the three core domains of functions, core NFV infrastructure, infrastructure orchestration, and operations management. At the core infrastructure, ESXi is used to virtualize the compute resources, NSX-T Data Center to provide virtual networking, and vSAN for storage. The core NFV infrastructure virtualization layer provides the following functions:

- Physical Resource Abstraction. By using the software component layers between the physical hardware and the VNFs, physical resources are abstracted. This provides a standardized software-based platform for running workloads, regardless of the underlying hardware. As long as the CSP uses certified physical components, workloads can be deployed by the carrier at the point of presence (POP), distributed, or centralized data center.
- Physical Resource Pooling. Physical resource pooling occurs when vCloud NFV OpenStack Edition presents a logical virtualization layer to workloads, combining the physical resources into one or more resource pools. Resource pooling together with an intelligent scheduler facilitates optimal resource utilization, load distribution, high availability, and scalability. This allows for fine grained resource allocation and control of pooled resources based on specific workload requirements.
- Physical Resource Sharing. To truly benefit from cloud economies, the resources that are pooled and abstracted by the virtualization layer must be shared between various network functions. The virtualization layer provides the functionality that is required for VNFs to be scheduled on the same compute resources, collocated on shared storage, and to have network capacity divided among them. The virtualization layer also ensures fairness in resource utilization and usage policy enforcement.

vCloud NFV Infrastructure and Orchestration

The infrastructure and orchestration domain contain the various NFVI abstractions for compute, storage, and networking and the resource orchestration component, also known as the Virtual Infrastructure Manager (VIM).

Compute - VMware ESXi

ESXi is the hypervisor software that abstracts the physical x86 server resources from the VNFs. Each compute server is called a host in the virtual environment. ESXi hosts are the fundamental compute building blocks of vCloud NFV OpenStack Edition. ESXi host resources can be grouped to provide an aggregate set of resources in the virtual environment that is called a cluster. Clusters logically separate the management and VNF components and are discussed in details in the *Core Reference Architecture* section. ESXi hosts are managed by the VMware vCenter Server Appliance that is part of the VIM components.

Key new features introduced in ESXi:

- Single reboot upgrade. vSphere upgrades can now be completed with one single reboot.
- ESXi quick boot. Allows a system to reboot in less than two minutes as it does not reinitialize the physical server.
- Instant Clone. Enables a user to create powered-on VMs from the running state of another powered-on VM without losing its state.
- NVDIMM devices. Support for next generation of storage devices that use persistent DRAM memory.
- Enhanced vMotion Capability. Define minimum CPU preferred features per VM to ensure CPU aware migration.
- Hugepages. The Size of Hugepages has now been extended to 1GB, improving memory access performance due to lower TLB misses.
- Enhancements for NSX-T Data Center. Together with NSX-T Data Center, vSphere introduces a new N-VDS Enhanced mode for switching to deliver substantial switching performance.
- RDMA over Converged Ethernet. RDMA provides low latency and higher-throughput interconnects with CPU offloads between the end-points.
- Higher security enhancements for TLS 1.2 and FIPS 140-2 cryptography.
- Increases performance and availability.

Host Management - VMware vCenter Server

VMware vCenter Server® is the centralized management interface for compute and storage resources in the NFVI. It provides an inventory of allocated virtual to physical resources, manages inventory-related information, and maintains an overview of the virtual resource catalogs. vCenter Server collects data about the performance, capacity, and state of its inventory objects. It exposes APIs to other management components for fine-grained control, operation, and monitoring of the underlying virtual infrastructure.

Networking - VMware NSX-T Data Center

NSX-T Data Center is the successor to NSX for vSphere. It allows CSPs to programmatically create, delete, and manage software-based virtual networks. These networks are used for communication between VNF components, and to provide customers with dynamic control of their service environments. Dynamic control is provided through tight integration between the resource orchestration layer and NSX-T Data Center. Network multitenancy is implemented by using NSX-T Data Center, by assigning customers their own virtual networking components and providing different network segments. A two-tiered architecture is used in the NSX-T Data Center design to implement a provider and tenant separation of control across the logical switching and routing fabric. Logical switching is supported in two modes, N-VDS Standard with support for overlay and VLAN backed networks and N-VDS Enhanced for DPDK acceleration for overlay and VLAN backed networks. The fully distributed routing architecture enables routing functionality closest to the source. This structure gives both provider and tenant administrators complete control over their services and policies.

NSX-T Data Center also implements a separation of management, control, and data planes. The NSX Manager, Controller, and Edge are components of this architecture that are discussed in the sections to follow.

Storage - VMware vSAN

vSAN is the native vSphere storage component in the NFVI virtualization layer, providing a shared storage pool between hosts in the vSphere cluster. With vSAN, storage is shared by aggregating the local disks and flash drives that are attached to the host. Although third-party storage solutions with storage replication adapters that meet the VMware storage compatibility guidelines are also supported, this reference architecture discusses only the vSAN storage solution.

Resource Orchestration - VMware Integrated OpenStack

VMware Integrated OpenStack is the component that vCloud NFV OpenStack Edition exposes as the interface to the VNF services. It leverages the vCenter Server Appliance and NSX Manager to orchestrate compute, storage, network, and imaging infrastructure services from a single, programmable interface. The components that are used to enable the services include the Horizon, Keystone, Nova, Neutron, Cinder, Glance, and Heat OpenStack projects.

VMware Integrated OpenStack in addition extends the execution environment to deploy and maintain enterprise class Kubernetes clusters in an OpenStack environment. The implementation provides full heterogeneity and choice of native VM-based workloads and containerized micro-services. Kubernetes clusters are configured to use VMware Integrated OpenStack enterprise-grade services such as Keystone authentication for the cluster, Block Storage Cinder to provide persistent storage for stateful applications, and Neutron Load Balancing as a Service (LBaaS) for application services. Container networking is fully integrated into NSX-T Data Center by using the Container Network Interface (CNI) framework that can be configured in a consistent manner.

In addition, feature enhancements include support for elastic Tenant vDCs spanning multiple clusters, Keystone Federation to have unified identity management across multiple VMware Integrated OpenStack instances, Neutron QoS to shape bandwidth consumption per tenant, improved manageability, and API security using rate limiting.

Platform Services

The platform services domain represents the various capabilities that are enabled by the VMware Integrated OpenStack cloud infrastructure platform. VNFs, managers, and other components executing within the vCloud NFV OpenStack Edition platform can leverage these capabilities.

Edge Services - VMware NSX-T Data Center

NSX-T Data Center provides two classes of routing capabilities, Distributed Router (DR) and Service Router (SR). The service router capability enables services such as NAT, firewall, load balancer, and so on, at a Tier-0 and Tier-1 router that VNFs can employ with stateful and stateless options. The Edge services cannot be distributed and require a centralized pool of capacity with high availability and scalability. The appliances that host the centralized services or SR instances are called Edge Nodes. These nodes also provide connectivity to the physical infrastructure.

Analytics – VMware vRealize Operations Manager

The vCloud NFV OpenStack Edition platform is fully integrated with an operations management solution for day 1 and day 2 operations for health monitoring, issue avoidance, and closed-loop automation. This way, the platform provides infrastructure assurance out of the box. The analytics framework can be used by network function and application developers to ingest and correlate their services-specific data in vRealize Operations Manager and leverage its capabilities seamlessly. The framework provides various mechanisms through management and content packs for data management and closed-loop automation with workflows that are custom to their operations and planning needs.

Policy Consistency

The vCloud NFV OpenStack Edition platform components utilize policy-based frameworks that can be defined once and applied at runtime to maintain the desired end state. The decision and enforcement split makes the management and operation of the platform and its services highly flexible, which operations and applications can leverage. Policies in compute can be used for right-sizing the infrastructure to ensure capacity and performance. Workload placement and runtime optimization for DRS and vMotion can be prescribed with platform awareness. Policies in networking range in physical networking such as teaming, network management and control, and security for East-West and perimeter traffic control. Storage policies provide services such as availability levels, capacity consumption, and stripe widths for performance. Policies in operations can help to ensure that SLAs are met with configurable alerting, recommendation, and remediation framework.

Programmability

The VMware vCloud NFV OpenStack Edition solution is a fully open platform supporting flexible APIs and SDKs. The resource orchestration provides compliant upstream alignment with OpenStack APIs. Infrastructure management for compute, networking, and storage are fully programmable as well.

For more information on the VMware APIs, see the VMware API Explorer.

Continuity

This domain represents components for business continuity and disaster recovery solutions, which are an integral part of the vCloud NFV OpenStack Edition platform.

VMware Site Recovery Manager

Site Recovery Manager works with various storage replication solutions, including vSphere Replication, to automate the migration, recovery, testing, and failing back virtual machine workloads for disaster recovery across multiple sites.

VMware vSphere Replication

vSphere Replication is a virtual machine data protection and disaster recovery solution. It is fully integrated with vCenter Server and VMware vSphere® Web Client, providing host-based, asynchronous replication of virtual machines including their storage.

Operations Management

This domain represents the day 1 and day 2 functions to ensure that the infrastructure and service components are operating in a healthy state so that SLAs are met.

The operations management solution includes four components that together provide a holistic approach to the operations management for the NFVI of a CSP. Together vRealize Operations, vRealize Log Insight, and vRealize Network Insight monitor the health of the virtual environment, collect logs and alarms, correlate events across multiple data sources and components to predict future issues. These components and vRealize Orchestrator use the policy-based automation framework to conduct remediation and analyze data to help the operator with health prediction and issue avoidance.

The key tasks that the operations management components perform are the following:

- Topology and discovery. NFVI visibility is achieved by collecting essential performance and fault metrics from the virtualization layer, the physical devices, and the VIM components. The elastic and dynamic nature of the NFVI layer requires keeping track of objects and maintaining correlations to help with decision tree accuracy and intelligence for infrastructure assurance.
- Monitoring. The components of the Operations domain continuously collect and analyze health, SLA, and planning metrics to ensure that services are meeting stringent QoS and to help avoiding issues in a predictable and prescriptive cadence.
- Issue isolation. The components of the NFV environment in the physical infrastructure, the virtualization layer, or even the VNFs themselves, generate various log messages and alarms. vCloud NFV OpenStack Edition includes an integrated log collection system that correlates between alerts and log messages to quickly troubleshoot issues.
- Remediation. Ongoing management of performance and capacity across the NFVI is required for optimal and economic use of the platform. The performance management capability helps identify degraded performance before VNFs are affected. Issues pertaining to performance, capacity, congestion, and so on, can be proactively avoided, increasing the Mean Time To Failure (MTTF).
- Closed-loop optimization. The operations management components analyze the system usage and proactively provide optimization recommendations, including network topology modifications. Actions can be tied into orchestrated workflows and automation to trigger VNFM, service orchestrators, and others to ensure continuous optimization, balancing, and recover in the event of failures.

VMware vRealize Operations Manager

VMware vRealize® Operations Manager™ delivers operations management with full stack visibility across the physical and virtual infrastructure. Through performance and health monitoring functions, vRealize Operations Manager improves the system performance, avoids service disruption, and helps the CSP to provide proactive management of the NFVI. The key capabilities include predictive analytics, smart and configurable alerts, and guided remediation.

vRealize Operations Manager exposes the information it gathers through an API that MANO and other components can use.

With this release, vRealize Operations Manager enables intent-based business and operational policies for dynamic optimization and capacity targets. Cost analysis is embedded into the solution, as with improved capacity planning and forecasting engines.

VMware vRealize Log Insight

VMware vRealize Log Insight delivers heterogeneous log management with dashboards, analytics, and third-party extensibility. It provides deep operational visibility and troubleshooting across physical, virtual, and cloud environments. Its indexing and machine learning based grouping provides log searches that help with troubleshooting issues.

VMware vRealize Network Insight

vRealize Network Insight collects metrics, flow, network topology, and event data to provide a detailed view of the network configuration and its health. Information is collected on all NSX-T managed networks including East-West traffic between VNF components, and North-South traffic in and out of the NFV infrastructure. Broad layer 2 to layer 3 support means that vRealize Network Insight can visualize both the underlay and the overlay networks, providing the operator with a holistic view into all relevant network layers. By using this information for visibility and analytics across all virtual and physical elements, the operator can optimize network performance and increase its availability.

On the security aspects, vRealize Network Insight offers intelligence operations for SDN and security across virtual and physical infrastructure by using micro-segmentation planning and policy distribution into NSX-T. The solution can be scaled, providing early warning on security policy violations.

VMware vRealize Orchestrator

The vCloud NFV platform provides closed-loop automation workflows to enable self-healing across VMs, hosts, and datastores at infrastructure level. It also allows the CSP to create VNF-specific custom workflows for faster time to resolution. vRealize Operations Manager integrates with vRealize Orchestrator through a management pack that provides access to the Orchestrator workflow engine for more remediation actions and the ability to run Orchestrator workflows directly from the vRealize Operations Manager user interface.

NFV OpenStack Edition Components

The vCloud NFV OpenStack Edition bundle packages together the essential building blocks to deploy an NFVI and VIM platform, featuring the newest releases of VMware production proven solutions.

The reference architecture is separated into two sections comprising of the *Core Reference Architecture* and *Analytics Enabled Reference Architecture*. The *Core Reference Architecture* is the base building block for the vCloud NFV OpenStack Edition reference architecture and the *Analytics Enabled Reference Architecture* builds on top of it.

Table 5-1. vCloud NFV OpenStack Edition Components

Component	Included in the vCloud NFV OpenStack Edition Bundle	Required for the Core Reference Architecture	Required for the Analytics-Enabled Reference Architecture
VMware ESXi™	Yes	Yes	Yes
VMware vCenter® Server Appliance™	No	Yes	Yes
VMware vSphere® Replication™	Yes	Recommended	Recommended
VMware vSAN™ Standard Edition	Yes	Recommended	Recommended
VMware® Integrated OpenStack Carrier Edition	Yes	Yes	Yes
VMware NSX-T® Data Center	No	Yes	Yes
VMware Site Recovery Manager™	No	Recommended	Recommended
VMware vRealize® Operations™	Yes	No	Yes
VMware vRealize® Log Insight™	Yes	No	Yes
VMware vRealize® Orchestrator	Yes	No	Yes
VMware vRealize® Network Insight™	No	No	Recommended

Design Principles

The following design principles communicate a summary of the details represented in the core and analytics reference architecture.

Flexible Deployment Options

vCloud NFV OpenStack Edition can be deployed to support either a Three-Pod (Management, Resource, and Edge Pods) or a Two-Pod (Management and combined Resource and Edge Pods) configuration. The Three-Pod architecture provides the highest flexibility and performance, because Edge Nodes are dedicated to perform packet forwarding in and out of the virtual domain. CSPs can use the Two-Pod architecture for smaller starting deployments, where it is acceptable to combine the resource and Edge functionality in the same hosts.

Advanced Networking

To provide multitenant capabilities to distributed logical routers in the networking stack, vCloud NFV OpenStack Edition uses NSX-T Data Center to deploy multiple tiers of distributed routing through Tier-0 and Tier-1 routers. Providers can use Tier-0 routers, whereas tenants can use Tier-1 routers. Network virtualization capabilities that are enabled through Geneve encapsulation provide a flexible capability in line with industry standards. In addition, NSX-T Data Center performance enhancements for the N-VDS and NSX Edge Nodes offer advanced network capabilities.

Workload Acceleration

vCloud NFV Open Stack Edition includes several features to support workloads that require high performance. These features are delivered with the newly available N-VDS with Enhanced Data Path Mode that can be used for high-performance workloads. In addition, North-South traffic forwarding between logical and physical domains can benefit from bare metal NSX Edge Nodes. This high-performance capability is available through the Data Plane Development Kit (DPDK) based enhancements that come with NSX-T Data Center. Including optimizations through poll mode drivers, CPU affinity and optimization and buffer management, DPDK provides support for workloads requiring acceleration.

Hybrid Execution Environment

With the ability to support both VMs and VNFs delivered through VMs or containers, vCloud NFV Open Stack Edition provides a highly flexible environment for deployments that require VMs and containers. Through the integrated Kubernetes functionality, customers can deploy, manage, and scale container-based workloads and networking functions.

Integrated Operational Intelligence

By using a framework that continuously collects data from local and distributed agents, vCloud NFV OpenStack Edition also provides the capability to correlate, analyze, and enable day 2 operations. In addition, this analytical intelligence can be used with existing assurance engines for closed-loop remediation.

Core Reference Architecture

6

The VMware vCloud NFV OpenStack platform is an evolution of the VMware NFV solution, based on extensive customer deployment and the continued development of standards organizations such as the European Telecommunications Standards Institute (ETSI). The vCloud NFV OpenStack Edition platform provides a comprehensive, service-oriented solution, leveraging a cloud computing model that allows ubiquitous, programmatic, on-demand access to a shared pool of compute, network, and storage resources. The solution is integrated with holistic operations management and service assurance capabilities, empowering the operator to rapidly deliver services while ensuring their quality. With a fully integrated VIM, the same vCloud NFV infrastructure delivers a myriad of telecommunications use cases, and facilitates reusability of the service catalog based VNFs.

The vCloud NFV OpenStack Edition platform delivers a complete, integrated solution that has been rigorously tested to ensure compatibility, robustness, and functionality. Components used in creating the solution are currently deployed across many industries and scenarios. vCloud NFV OpenStack Edition software components can be used in a variety of ways to construct a comprehensive, end-to-end solution that meets the business goals of CSPs. This document discusses how components can be used to create a vCloud NFV OpenStack Edition architecture.

This chapter includes the following topics:

Core Building Blocks

Architecting vCloud NFV OpenStack Edition by using well-defined modules allows the CSP to accelerate the deployment of the platform and reliably expand it when needed. The platform components are grouped into three distinct containments. The VMware vCloud NFV OpenStack Edition platform uses the term Pods as a mean to streamline the NFV environment operations and delineate between different roles. For example, a cloud management team can easily operate the Management Pod, whereas a network management team is likely to oversee the Edge Pod. VNFs are always deployed in the Resource Pod.

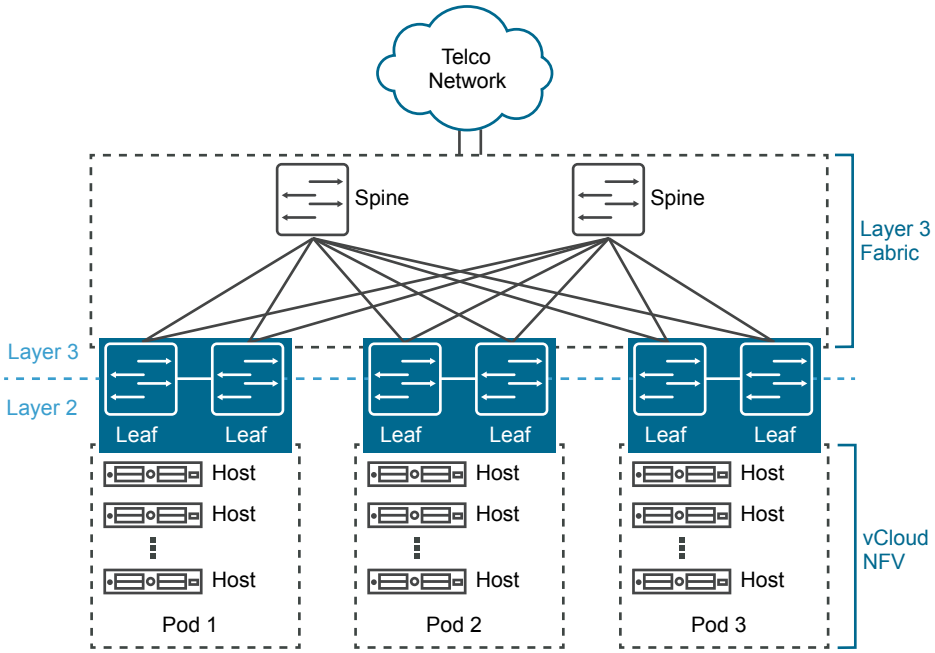
Each Pod is identified by its functional designation - Management Pod, Edge Pod, and Resource Pod. The Pod functions are the following:

Management Pod	Management functions are required to manage the NFV Infrastructure and the VNFs and their components. Management, orchestration, analytics functionality, and ancillary elements such as DNS, VNF-M, and NFV-O are grouped into this category. Resource orchestration such as vCenter Server Appliance, NSX Manager, NSX Controller, and VMware Integrated OpenStack are hosted in this Pod. The analytics components, which include vRealize Operations Manager, vRealize Network Insight, vRealize Log Insight, vRealize Orchestrator, and business continuity components such as Site Recovery Manager and vSphere Replication are all located in this pod. Other management-related components such as NFV Orchestrators run in the Management Pod. OSS/BSS can be very large in sizing, and therefore their placement depends on the system itself.
Edge Pod	The Edge functions provide a logical networking delineation between VNFs and external networks. Network traffic transitioning between the physical domain and the virtual domain is processed by these functions. NSX Edge is hosted in a VM or bare metal appliance in the Edge Pod and handles all connectivity to the physical domain in the architecture. The type of networking traffic that traverses the Edge Pod is called North-South traffic.
Resource Pod	VNFs are placed in the Resource Pod. The VNFs then form the virtual network service.

Physical Building Blocks

Traditional data center network fabrics are often designed with three tiers of switches that are core, aggregation, and access. Access switches connect to aggregation switches, which in turn connect to the core switches. The design topology of the physical layer can impact the efficiency and latencies.

Figure 6-1. Physical Network Design



Communication between two endpoints within the data center begins from the access switch, traverses the aggregation switch to the core switch, and travels from there to the remote endpoint. This traffic pattern results in inefficiencies and increased latency. While adequate for basic network traffic, this is not a reasonable traffic pattern for traffic that does not tolerate latency.

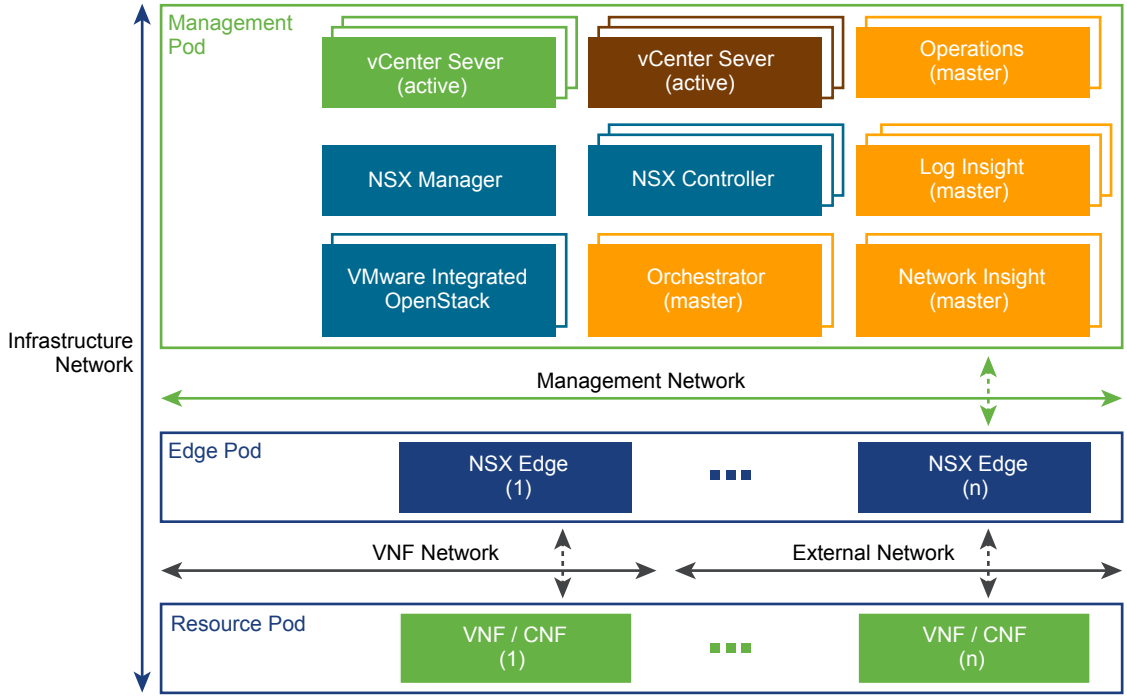
A two-tier leaf-and-spine network architecture is the more preferred approach for building newer data center infrastructure. The two-tier architecture uses an access switch, or leaf, which is connected to an aggregation switch, or spine. The leaf switch provides connectivity between endpoints in the data center, while the spine switch provides high-speed interconnectivity between leaf switches. The leaf-and-spine network is connected in a full mesh, providing predictable communication and latency between endpoints. Ethernet connectivity is used from the host to the leaf switch, and the broadcast domain terminates at the leaf. External Border Gateway Protocol (eBGP) is the control plane option for routing within the leaf-and-spine architecture.

Virtual Building Blocks

The virtual infrastructure design comprises the design of the software components that form the virtual infrastructure layer. This layer supports running telco workloads and workloads that maintain the business continuity of services. The virtual infrastructure components include the virtualization platform hypervisor, virtualization management, storage virtualization, network virtualization, and backup and disaster recovery components.

This section outlines the building blocks for the virtual infrastructure, their components, and the networking to tie all the components together.

Figure 6-2. Virtual Building Blocks



Storage Design

A shared storage design that is based on vSAN. vCloud NFV OpenStack Edition also supports certified third-party shared storage solutions, as listed in the VMware Compatibility Guide.

vSAN is a software feature built in the ESXi hypervisor that allows locally attached storage to be pooled and presented as a shared storage pool for all hosts in a vSphere cluster. This simplifies the storage configuration with a single datastore per cluster for management and VNF workloads. With vSAN, VM data is stored as objects and components. One object consists of multiple components, which are distributed across the vSAN cluster based on the policy that is assigned to the object. The policy for the object ensures a highly available storage backend for the cluster workload, with no single point of failure.

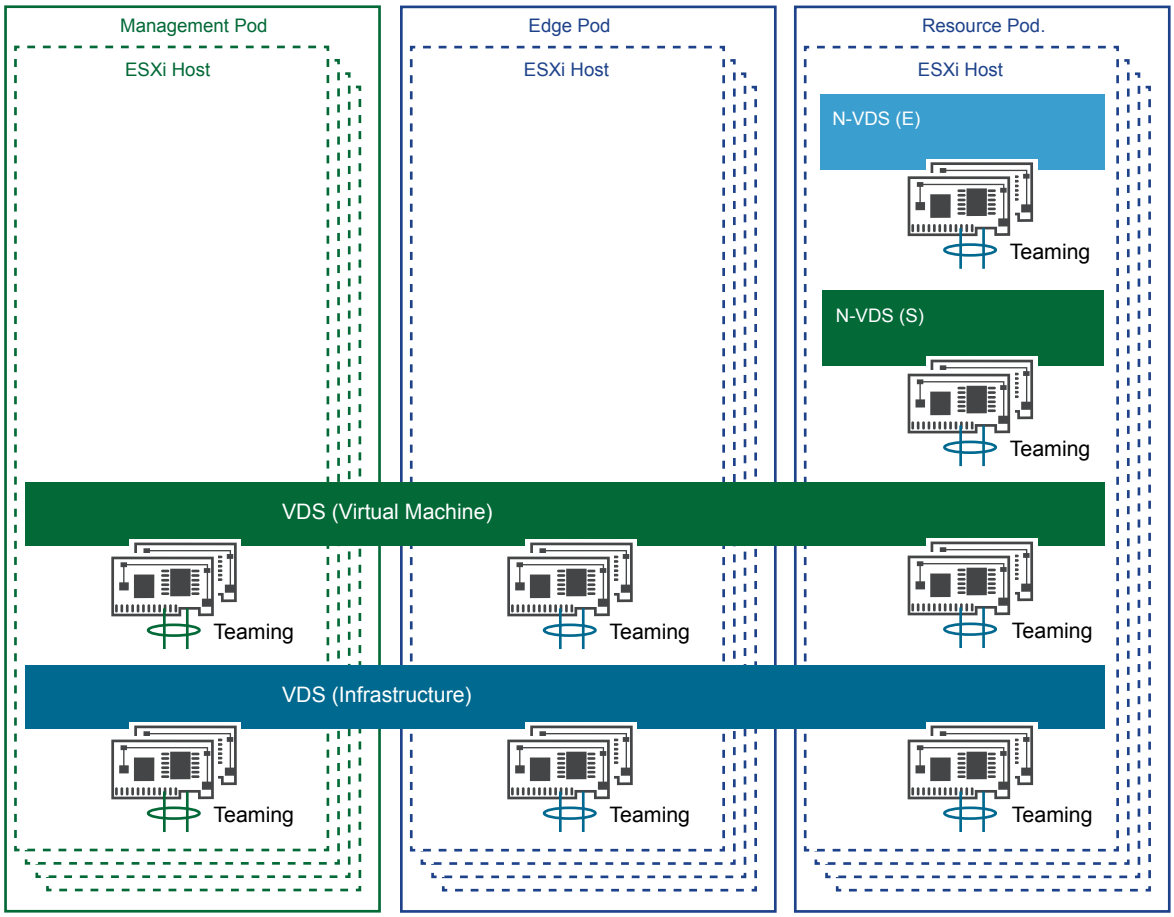
vSAN is a fully integrated hyperconverged storage software. Creating a cluster of server hard disk drives (HDDs) or solid-state drives (SSDs), vSAN presents a flash-optimized, highly resilient, shared storage datastore to ESXi hosts and virtual machines. This allows for the control of capacity, performance, and availability through storage policies, on a per VM basis.

Network Design

The vCloud NFV OpenStack platform consists of infrastructure networks and VM networks. Infrastructure networks are host level networks that connect hypervisors to physical networks. Each ESXi host has multiple port groups configured for each infrastructure network.

The hosts in each Pod are configured with VMware vSphere® Distributed Switch™ (vDS) devices that provide consistent network configuration across multiple hosts. One vSphere Distributed Switch is used for VM networks and the other one maintains the infrastructure networks. Also, the N-VDS switch is used as the transport for telco workload traffic.

Figure 6-3. Virtual Network Design



Infrastructure networks are used by the ESXi hypervisor for vMotion, VMware vSphere Replication, vSAN traffic, and management and backup. The Virtual Machine networks are used by VMs to communicate with each other. For each Pod, the separation between infrastructure and VM networks ensures security and provides network resources where needed. This separation is implemented by two vSphere Distributed Switches, one for infrastructure networks and another one for VM networks. Each distributed switch has separate uplink connectivity to the physical data center network, completely separating its traffic from other network traffic. The uplinks are mapped to a pair of physical NICs on each ESXi host, for optimal performance and resiliency.

VMs can be connected to each other over a VLAN or over Geneve-based overlay tunnels. Both networks are designed according to the requirements of the workloads that are hosted by a specific Pod. The infrastructure vSphere Distributed Switch and networks remain the same regardless of the Pod function. However, the VM networks depend on the networks that the specific Pod requires. The VM networks are created by NSX-T Data Center to provide enhanced networking services and performance to the Pod

workloads. The ESXi host's physical NICs are used as uplinks to connect the distributed switches to the physical network switches. All ESXi physical NICs connect to layer 2 or layer 3 managed switches on the physical network. It is common to use two switches for connecting to the host physical NICs for redundancy purposes.

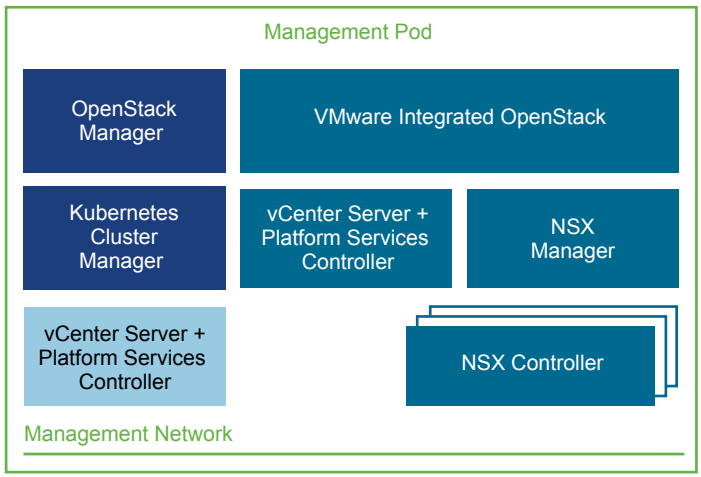
Following are the infrastructure networks used in the Pods:

- ESXi Management Network. The network for the ESXi host management traffic.
- vMotion Network. The network for VMware vSphere® vMotion® traffic.
- vSAN Network. The network for vSAN shared storage traffic.
- Backup Network. The network that is dedicated to offline storage such as NFS and used for workload backup and restore as needed.
- Replication Network. This is the network that is used for replicating data for data protection.

Management Pod

This section describes the design for the Virtualized Infrastructure Management (VIM) components that are vCenter Server Appliance, NSX Manager, VMware Integrated OpenStack, and VMware Integrated OpenStack with Kubernetes cluster.

Figure 6-4. Management Pod



In addition to these core components, the Management Pod also contains the operations management components. For more information , see the *Analytics-Based Reference Architecture* section.

Components

The Management Pod contains the components that manage the vCloud NFV OpenStack Edition runtime environment.

vCenter Server

The Management Pod is implemented as a cluster that is managed by the first vCenter Server instance. To form the foundation of a carrier grade virtualized infrastructure, the components of the Management Pod benefit from the cluster features such as resource management, high availability, and resiliency. A second vCenter Server is deployed in the Management Pod to oversee the Edge and Resource Pods.

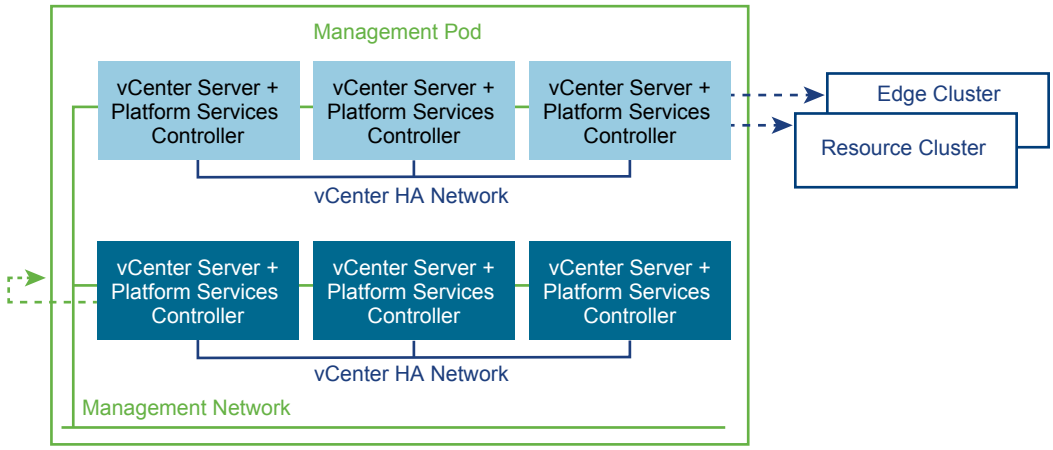
Each vCenter Server instance is a virtual appliance that is deployed with an embedded database. The vCenter® Server Appliance™ is preconfigured, hardened, and fast to deploy. The appliance allows for a simplified design, eases management, and reduces administrative efforts. vCenter Server Appliance availability is ensured by using a vCenter High Availability (vCenter HA) cluster, which is realized through three vCenter Server Appliance instances. The vCenter HA cluster consists of one active node that serves client requests, one passive node as a backup in the event of failure, and one quorum node that is called a witness node. Replication between nodes by using a dedicated vCenter HA network ensures that vCenter Server Appliance data is always synchronized and up-to-date.

The Platform Services Controller contains common infrastructure security services such as VMware vCenter® Single Sign-On, VMware Certificate Authority, licensing, service registration, and certificate management services. The Platform Services Controller handles identity management for administrators and applications that interact with the vSphere platform. The Platform Services Controller and its related services are embedded within the vCenter Server Appliance. This eliminates the need for separate Platform Services Controller VM instances and their corresponding load balancers, thus simplifying its deployment and administration and reducing the management components footprint.

Data backup and restore of each vCenter Server instance and its embedded Platform Services Controller is provided by using the native backup service that is built in the appliances. This backup is performed to a separate storage system by using network protocols such as SFTP, HTTPS, and SCP.

When vCenter HA is used with an embedded Platform Services Controller, the environment setup is as follows:

Figure 6-5. vCenter Server High Availability

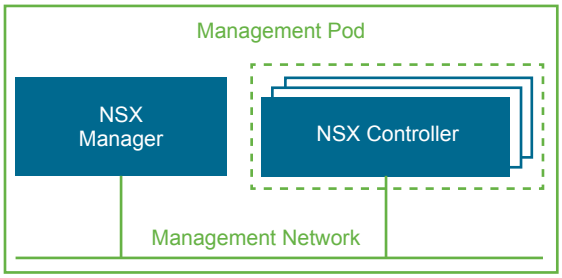


VMware NSX-T Data Center

NSX Manager. The management plane for the NSX-T system. It provides the ability to create, configure, and monitor NSX-T Data Center components, such as logical switches, and NSX Edge Nodes. NSX Manager provides an aggregated system view and is the centralized network management component of NSX-T Data Center. It provides a method for monitoring and troubleshooting workloads that are attached to the virtual networks that NSX-T Data Center creates. NSX-T Data Center provides configuration and orchestration of logical networking components such as logical switching and routing, networking services, Edge services, security services, and distributed firewall capabilities.

NSX Manager is deployed as a single node that uses vSphere HA for high availability. NSX Manager communicates with its controller and Edge clusters over a common management network. The management components of the vCloud NFV platform communicate over the same management network to request network services from NSX-T Data Manager.

Figure 6-6. NSX Manager and Components



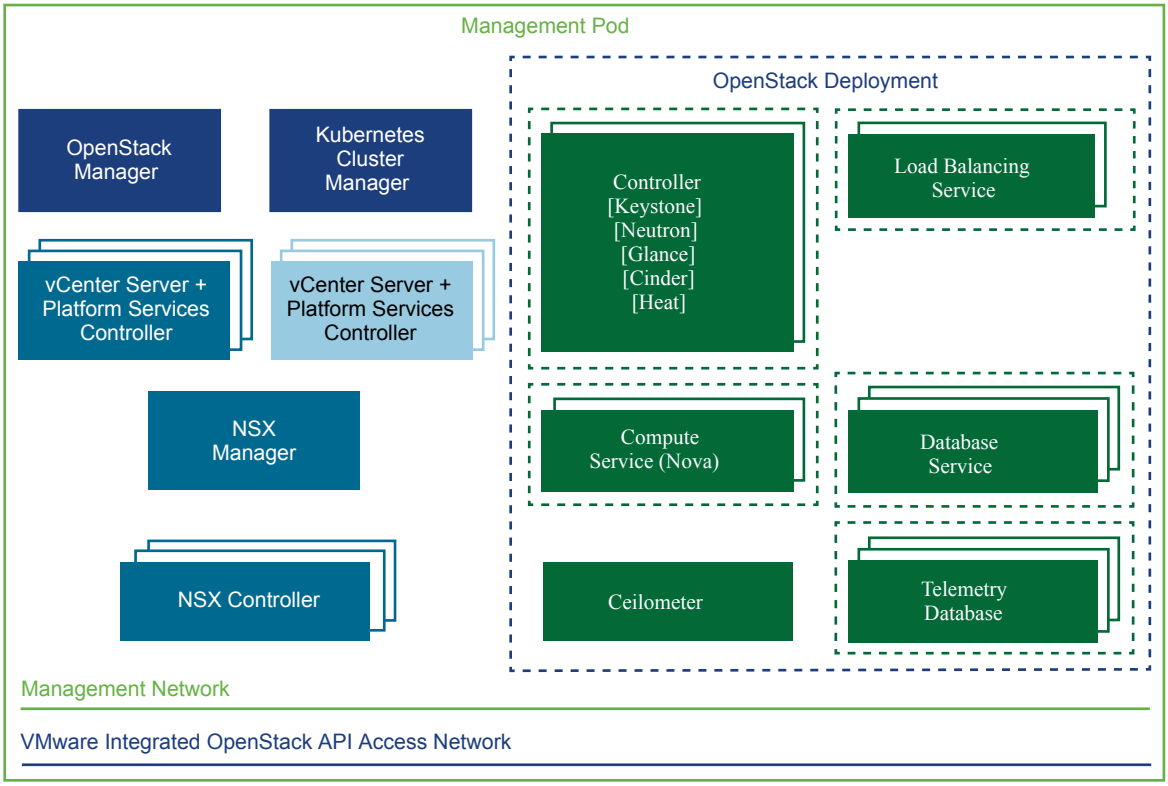
NSX Controller . An advanced distributed state management system that controls virtual networks and overlay transport tunnels. NSX Controller is deployed as a cluster of highly available virtual appliances that are responsible for the programmatic deployment of virtual networks across the entire NSX-T Data Center architecture. The control plane is split in two parts in NSX-T Data Center, the central control plane (CCP), which runs on the NSX Controller cluster nodes, and the local control plane (LCP), which runs on the transport nodes, adjacent to the data plane it controls. The Central Control Plane computes some ephemeral runtime state based on configuration from the management plane and disseminates information reported through the local control plane by the data plane elements. The Local Control Plane monitors local link status, computes most ephemeral runtime state based on updates from data plane and CCP, and pushes stateless configuration to forwarding engines. The LCP shares fate with the data plane element that hosts it. The NSX-T Data Center Central Control Plane (CCP) is logically separated from all data plane traffic, therefore any failure in the control plane does not affect the existing data plane operations. Traffic does not pass through the controller, instead the controller is responsible for providing configuration to other NSX Controller components such as the logical switches, logical routers, and edge configuration. Stability and reliability of data transport are central concerns in networking.

To further enhance the high availability and scalability, the NSX Controller is deployed in a cluster of three instances in the Edge cluster. Anti-affinity rules are configured to ensure that the controller instances reside on separate hosts to protect against host failures.

VMware Integrated OpenStack

OpenStack Manager. The VMware Integrated OpenStack Manager connects to the vCenter Server instance that manages the Management Pod. It uses a VM template to rapidly deploy, administer, and perform day 2 management operations of the VMware Integrated OpenStack management plane components that are deployed in the Management Pod. Once deployed, VMware Integrated OpenStack connects to the vCenter Server instance that manages the Edge and Resource Pods. This vCenter Server instance is responsible for storage and compute resources. VMware Integrated OpenStack also connects to the NSX Manager instance that is associated with tenant networking.

Figure 6-7. VMware Integrated OpenStack Management Components



The VMware Integrated OpenStack management plane is deployed with redundancy for all VMware Integrated OpenStack management components, ensuring that there is no single point of failure. Although this requires greater resource availability in the Management Pod, it offers the best configuration for high availability and is the recommended topology for production environments.

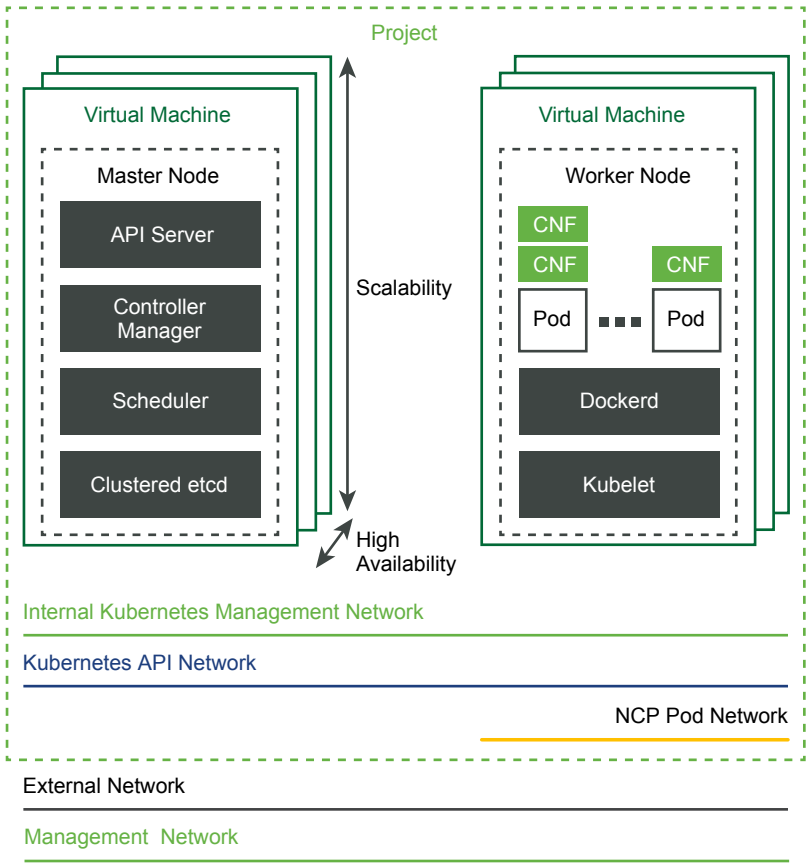
In a VMware Integrated OpenStack high availability deployment, all the components for a scalable and highly available VMware Integrated OpenStack full deployment, including clustered databases, controllers, and VMware Integrated OpenStack load balancers, can also be deployed by the Integrated OpenStack Manager. All management components have connectivity to each other through a dedicated management network.

VMware Integrated OpenStack is closely integrated with NSX-T Data Center, providing tenants with enhanced features and capabilities for managing their VNF networking needs by using the Horizon interface and APIs. Network services include firewalling, network NAT, static and dynamic routing, and load balancing. Tenants can provision Geneve-backed logical switches for East-West VNF component connectivity and deploy NSX Edges for North-South traffic as required when connecting to other tenants or to external networks.

It is a best practice that each cluster within vCloud NFV OpenStack Edition is configured to use a shared storage solution. When hosts in a cluster use shared storage, manageability and agility improve.

VMware Integrated OpenStack with Kubernetes. Kubernetes is an open-source platform for automating deployment, scaling, and operations of application containers across host clusters, providing container-centric infrastructure. By combining Kubernetes with VMware Integrated OpenStack, a common infrastructure management layer can be used to provision both VMs and containers. VMware Integrated OpenStack with Kubernetes builds highly available Kubernetes clusters that support scalability and multi-tenancy.

Figure 6-8. VMware Integrated OpenStack with Kubernetes Architecture



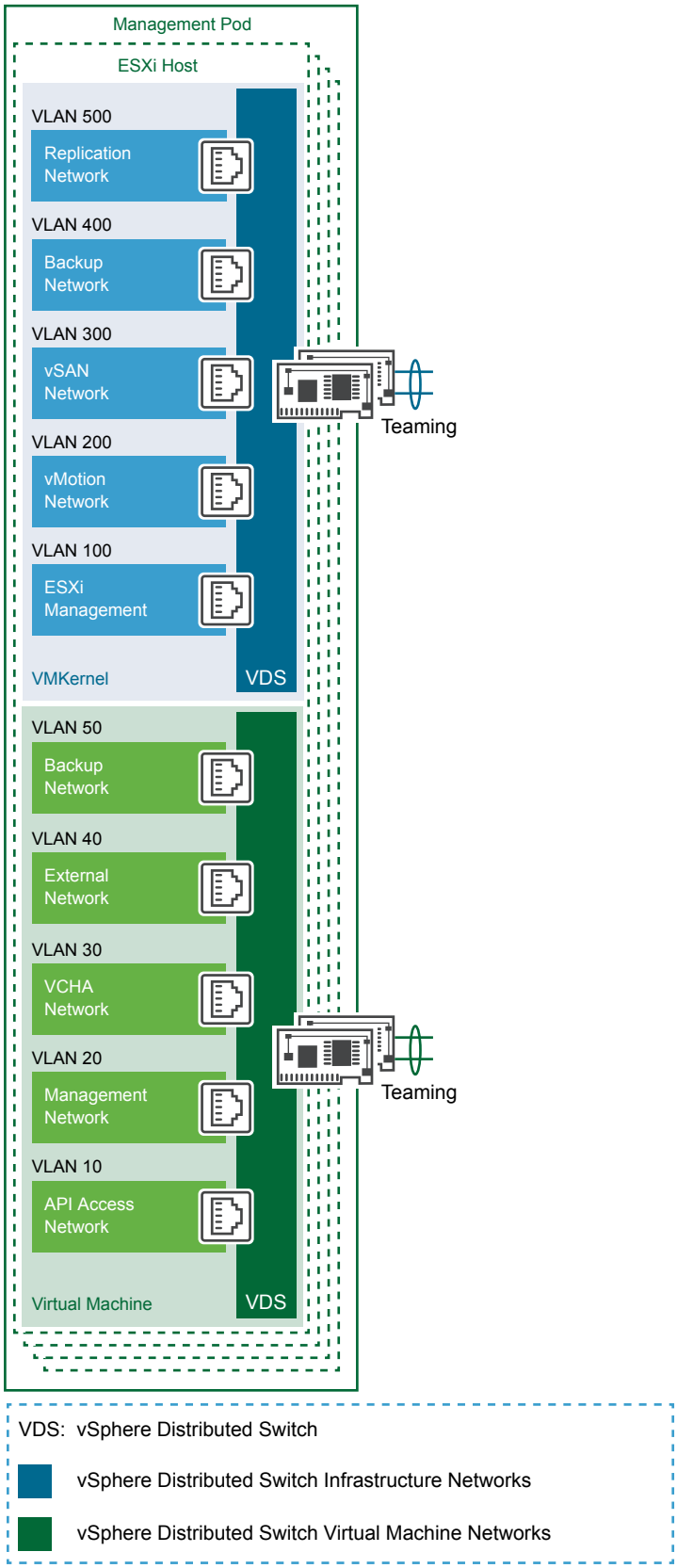
The highly available Kubernetes cluster consists of load-balanced master nodes, replicated API servers, and clustered services. In addition, the worker nodes in a Kubernetes cluster can be scaled in and out to meet changing demands for capacity.

For this reference architecture, a minimum of three master nodes are recommended and two worker nodes for high availability. vSphere anti affinity rules ensure that these nodes reside on separate ESXi hosts, so that if in case of ESXi host failure there is minimal impact to the Kubernetes cluster.

Networking

The Management Pod networking consists of the infrastructure and VM networks. The following diagram shows all the virtual switches and port groups of the Management Pod.

Figure 6-9. Management Pod Networking



Edge Pod

This section describes the components of the Edge Pod and their functions.

Components

The Edge Pod components provide the fabric for North-South connectivity to the provider networks. Multiple configurations can be used for performance, scale, and Edge services.

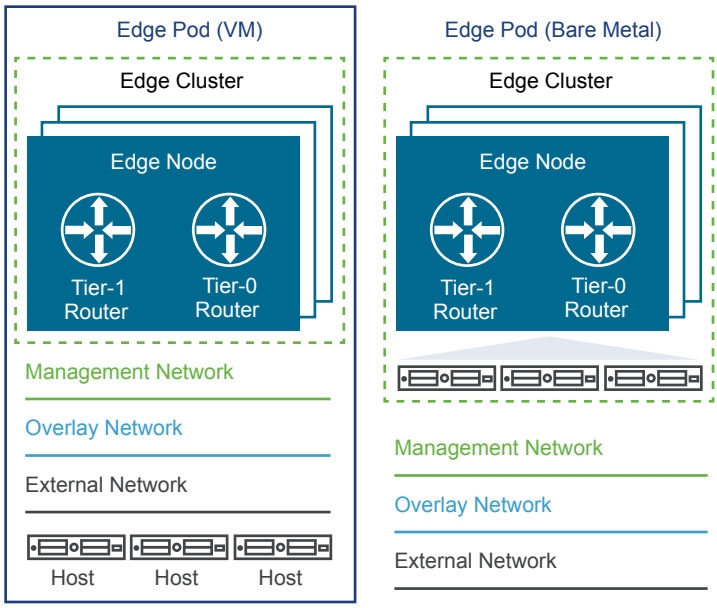
Edge Nodes

An Edge Node is the appliance that provides physical NICs to connect to the physical infrastructure and to the virtual domain. Edge Nodes serve as pools of capacity, dedicated to running network services that cannot be distributed to the hypervisors. The network functionality of the Edge node includes:

- Connectivity to physical infrastructure.
- Edge services such as NAT, DHCP, firewall, and load balancer.

Edge nodes are available in two form-factors, VM and bare metal. Both leverage the data plane development kit (DPDK) for faster packet processing and high performance.

Figure 6-10. Edge Pod Components



The NSX-T Data Center bare metal Edge runs on a physical server and is installed by using an ISO file or PXE boot. The bare metal Edge is recommended for production environments where services like NAT, firewall, and load balancer are needed in addition to Layer 3 unicast forwarding. A bare metal Edge differs from the VM form-factor Edge in terms of performance. It provides sub-second convergence, faster failover, and throughput greater than 10Gbps.

The NSX-T Data Center VM Edge in VM form-factor is installed by using an OVA, OVF, or ISO file. Depending on the required functionality, there are deployment-specific VM form-factors.

Edge Clusters

Edge nodes are deployed as pools of capacity (a cluster), dedicated to running network services that cannot be distributed to the hypervisors. An Edge cluster can either be all VM or all bare metal form-factors.

The Edge cluster provides scale out, redundant, and high-throughput gateway functionality for logical networks. Scale out from the logical networks to the Edge nodes is achieved by using ECMP. There is total flexibility in assigning logical routers to Edge nodes and clusters. Tier-0 and Tier-1 routers can be hosted on either same or different Edge clusters. Centralized services must be enabled for the Tier-1 logical router to coexist in the same cluster.

There can be only one Tier-0 logical router per Edge node, however multiple Tier-1 logical routers can be hosted on one Edge node.

In addition to providing distributed routing capabilities, the Edge cluster enables Edge services at a provider or tenant scope. As soon as one of these Edge services are configured or an uplink is defined on the logical router to connect to the physical infrastructure, a Service Router (SR) is instantiated on the Edge Node. The Edge Node is also a transport node just like compute nodes in NSX-T, and similar to compute node it can connect to more than one transport zone, one for overlay and other for N-S peering with external devices.

A maximum of eight Edge Nodes can be grouped in an Edge cluster. A Tier-0 logical router supports a maximum of eight equal cost paths, thus a maximum of eight Edge Nodes are supported for ECMP. Edge Nodes in an Edge cluster run Bidirectional Forwarding Detection (BFD) on both tunnel and management networks to detect Edge Node failure. The BFD protocol provides fast detection of failure for forwarding paths or forwarding engines, improving convergence. Bare metal form factors can support sub-second convergence.

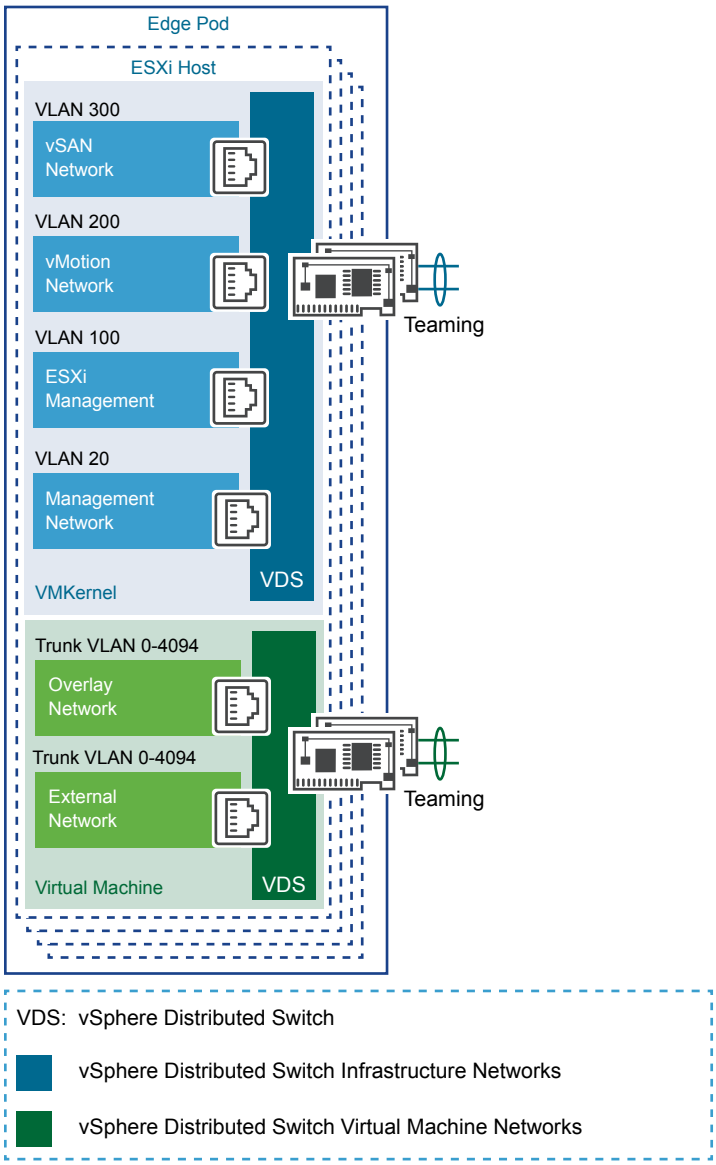
NSX-T Data Center supports static routing and the dynamic routing protocol BGP on Tier-0 logical routers on interfaces connecting to upstream routers. Tier-1 logical routers support static routes but do not support any dynamic routing protocols.

See the NSX Design Guide for more information.

Networking

The Edge Pod virtual network largely depends on the network topology that is required by the VNF workloads. In general, the Edge Pod has the infrastructure networks, networks for management and control plane connectivity, and networks for workloads.

Figure 6-11. Edge Pod Networking



Resource Pod

This section describes the components of the Resource Pod and their functions. VNFs are placed in the Resource Pod that forms the virtual network services.

Components

The Resource Pod provides the runtime environment for the network functions. The section covers the logical tenancy and networking components.

Projects

In VMware Integrated OpenStack, cloud administrators manage permissions through user, group, and project definitions. Projects in OpenStack are equal to tenants in vCloud NFV. A project is the administrative container where telco workloads are deployed and managed.

Tenant VDCs

A Tenant VDC allows creation of virtual data centers for tenants under different compute nodes that offer specific SLA levels for each telco workload. While quotas on projects set limits on the OpenStack resources, Tenant VDCs allow providing resource guarantees for tenants and avoid noisy neighbor scenarios in a multitenant environment.

VNFs

One or more VMs that are deployed in the tenancy to provide specific network functions or telco services.

Container Network Functions

Container Network Functions (CNFs) are containerized telco workloads similar to VNFs, with the difference that the VNF components (VNFCs) are deployed as microservices in a containerized environment.

For the vCloud NFV OpenStack Edition platform, a tenant can have an exclusive Kubernetes cluster that is deployed in the tenant's project. For high availability and to meet performance requirements, the master and worker nodes need to be spread across the ESXi hosts of the resource cluster.

At the same time, multiple tenants can share the resource cluster, because the underlying VMware Integrated OpenStack infrastructure ensures efficient resource pooling, partitioning, and allocation to the Kubernetes worker nodes of the tenants in the Resource Pod in a multi-tenant way.

As part of CNF onboarding, the CNF vendor also has to provide the specifications for the CNF infrastructure such as the number of master and worker nodes and any specific performance needs of the worker nodes. In terms of tenant provisioning, the project sizing can be done as though the Kubernetes cluster is just another set of tenant workload VMs.

Networking

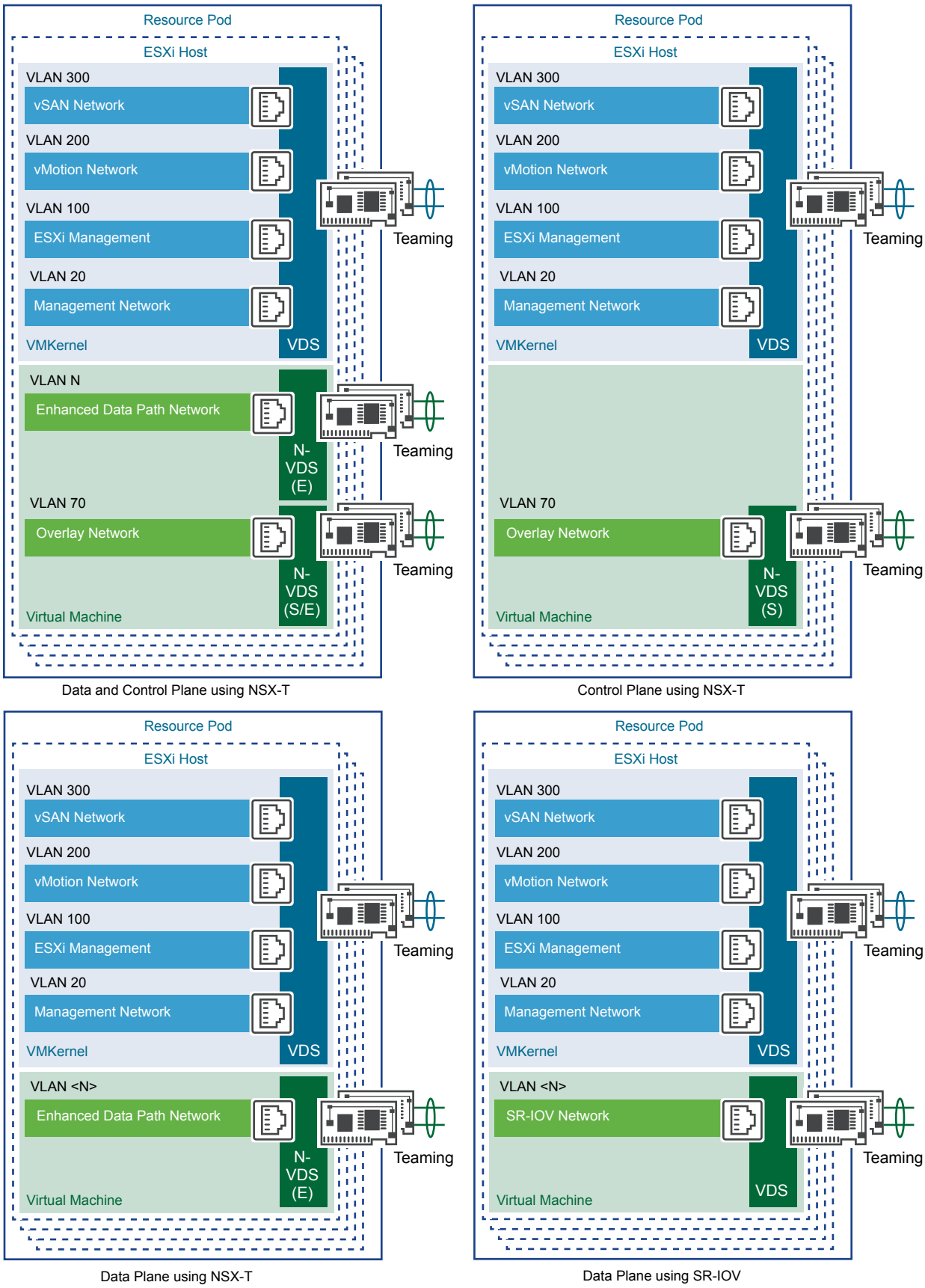
The networking of the Resource Pod is highly dependent on the network topology that is required by the telco workloads that are deployed by the tenant. This section describes the network building blocks as required by tenant workloads and is applicable to both VNF and CNF telco workloads.

Logical Switch

Logical switches are the layer 2 networks created by NSX-T Data Center to provide connectivity between its services and the VMs. Logical switches form the basis of the tenant networks in the vCloud NFV OpenStack Edition platform. The primary component in the data plane of the transport nodes is N-VDS. N-VDS forwards traffic between components running on the transport node (that is between VMs) or

between VMs and the physical network. In the latter case, N-VDS must own one or more physical interfaces (physical NICs) on the transport node. As with other virtual switches, an N-VDS cannot share a physical interface with another N-VDS. It can coexist with another N-VDS each using a separate set of physical NICs.

Figure 6-12. Resource Pod Networking



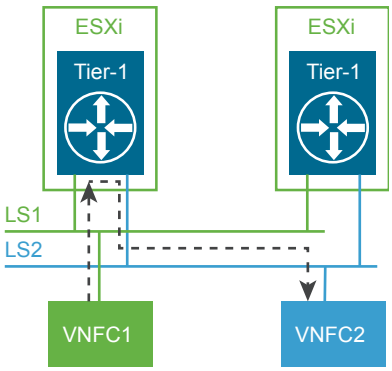
Logical Routing

The NSX-T Data Center platform provides the ability to interconnect both virtual and physical workloads that are deployed in different logical layer 2 networks. NSX-T enables the creation of network elements like switches and routers as software logical constructs and embeds them in the hypervisor layer, abstracted from the underlying physical hardware.

East-West Traffic

Configuring a logical router through the NSX Manager instantiates a logical router on each hypervisor. For the VNFs hosted on the same hypervisor, the East-West traffic does not leave the hypervisor for routing. The logical router is also responsible for routing East-West traffic between hypervisors. The logical router, also called the Tier-1 router is deployed and managed by the tenants of the vCloud NFV OpenStack Edition platform, for routing services between their respective tenant networks within their tenancy.

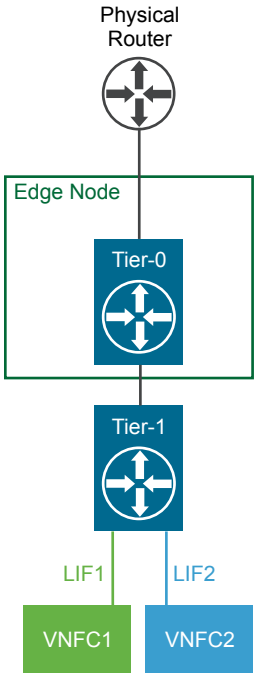
Figure 6-13. East-West Traffic



North-South Traffic

In addition to providing optimized distributed and centralized routing functions, NSX-T Data Center supports a multi-tiered routing model with logical separation between the provider routing function and the tenant routing function. This way, the concept of multitenancy is built in the routing model. The top-tier logical router is called a Tier-0 router, whereas the bottom-tier logical router is called a Tier-1 router. This structure provides both provider and tenant administrators a complete control over their services and policies. The provider administrator controls and configures Tier-0 routing and services and the tenant administrators control and configure Tier-1 routing services. Northbound, the Tier-0 logical router connects to one or more physical routers or layer 3 switches and serves as an on/off ramp to the physical infrastructure. Southbound, the Tier-0 logical router connects to one or more Tier-1 logical routers.

Figure 6-14. North-South Traffic



This model also eliminates the dependency on a physical infrastructure administrator to configure or change anything on the physical infrastructure when a new tenant is configured in the data center. For a new tenant, the Tier-0 logical router simply advertises the new tenant routes that are learned from the tenant Tier-1 logical router on the established routing adjacency with the physical infrastructure.

Deployment Options

7

The designs discussed in this reference architecture can be deployed in the data center to meet target design and scale objectives. Two design configurations are possible, a compact Two-Pod configuration and a Three-Pod configuration.

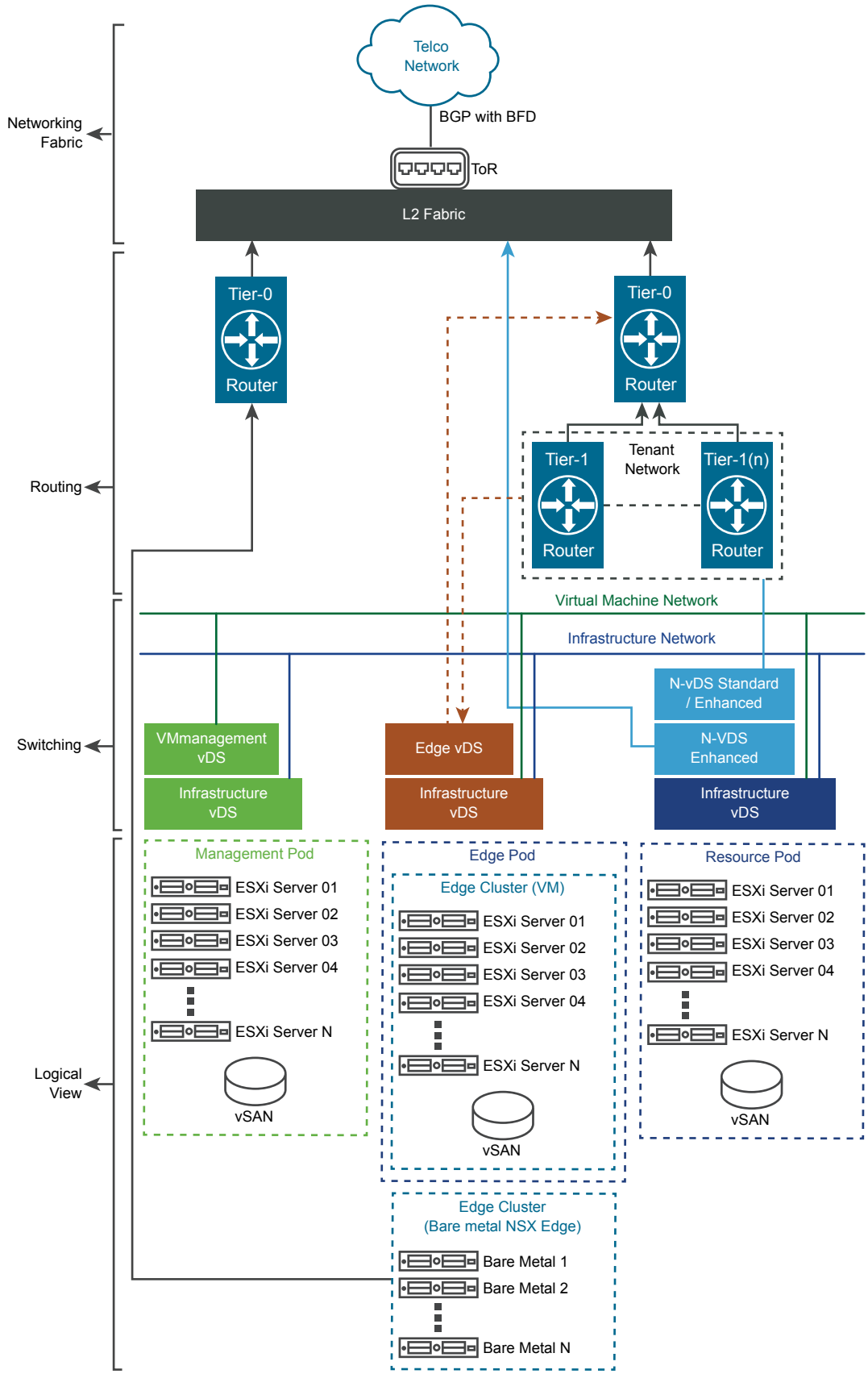
This chapter includes the following topics:

Three-Pod Configuration

The Three-Pod design completely separates the vCloud NFV functional blocks by using a Management Pod, Edge Pod, and Resource Pod for their functions. The initial deployment of a Three-Pod design consists of three vSphere clusters, respectively one cluster per Pod. Clusters are scaled up by adding ESXi hosts, whereas Pods are scaled up by adding clusters. The separation of management, Edge, and resource functions in individually scalable Pods allows the CSPs to plan capacity according to the needs of the specific function that each Pod hosts. This provides greater operational flexibility.

The following diagram depicts the physical representation of the compute and networking connectivity and the logical layers of the switching and routing fabric.

Figure 7-1. Three-Pod Conceptual Design



The initial deployment of a Three-Pod design is more hardware intensive than the initial deployment of a Two-Pod design. Each Pod in the design scales up independently from the others. A Three-Pod design consists of the same components as in a Two-Pod design, as the way functions are combined to form the solution is different in each design. Regardless of the Pod design that is used to create the NFVI, VNFs perform the same way.

Logical View

- Management Pod. Hosts all the NFV management components. Its functions include resource orchestration, analytics, BCDR, third-party management, NFV-O, and other ancillary management.
- Edge Pod. Hosts the NSX-T Data Center network components, which are the NSX Edge nodes. Edge nodes participate in East-West traffic forwarding and provide connectivity to the physical infrastructure for North-South traffic management and capabilities. Edge nodes can be deployed in a VM and bare metal form-factor to meet capacity and performance needs.
- Resource Pod. Provides the virtualized runtime environment, that is compute, network, and storage, to execute workloads.

Routing and Switching

Before deploying the Three-Pod configuration, a best practice is to consider a VLAN design to isolate the traffic for infrastructure, VMs, and VIM.

The vSphere Distributed Switch port groups have different requirements based on the Pod profile and networking requirements.

For example, two vSphere Distributed Switches can be used in the Management Pod, one for VMkernel traffic and another one for VM management traffic. While the switching design is standardized for the Management and Edge Pods, the Resource Pod offers flexibility with the two classes of NSX-T Data Center switches that are N-VDS Standard and N-VDS Enhanced. The N-VDS Standard switch offers overlay and VLAN networking and the N-VDS Enhanced switch offers acceleration by using DPDK to workloads.

The NSX-T Data Center two-tiered routing fabric provides the separation between the provider routers (Tier-0) and the tenant routers (Tier-1).

The Edge nodes provide the physical connectivity to the CSPs core and external networking. Both dynamic and static routing are possible at the Edge nodes.

Design Considerations

A Three-Pod configuration provides flexibility, performance, and VNF distribution design choice for telco workloads.

Footprint

The best practice when using vSAN as shared storage for all clusters, is to use a minimum of four hosts per cluster for the initial deployment, which sums up a total of 12 hosts. This creates balance between the implementation footprint and resiliency, while maintaining the operational requirements for each Pod. A mid-level class of servers can be used in this configuration, because the design maximizes the scale flexibility.

Scale Flexibility

The separation of Pods provides maximum scalability to the design. Individual Pods can be scaled to meet the target services, delivery, and topology objectives.

The Resource and Edge clusters are sized according to the VNFs and their respective networking requirements. CSPs must work with the VNF vendors to gather the requirements for the VNFs to be deployed. This information is typically available in deployment and sizing guides. As more tenants are provisioned, CSPs must provide additional resources to the Resource Pod to support the VNF growth. The increase in the VNF workloads in the Resource Pod can lead to an increase in the North-South network traffic, which in turn requires adding more compute resources to the Edge Pod so that to scale up the Edge Nodes. CSPs must closely monitor and manage the resource consumption and capacity that is available to the Edge Pod when the Edge Nodes are scaled. For higher performance, an Edge Pod can also comprise of bare metal Edges that are grouped in an Edge Cluster in NSX-T Data Center.

High level of resiliency is provided by using four-host clusters with vSAN storage in the initial deployment. Four-host clusters also ensure a highly available Management Pod design, because clustered management components such as vCenter Server active, standby, and witness nodes, can be placed on separate hosts. This same design principle is used for clustered OpenStack components such as database nodes.

The initial number and sizing of management components in the Management Pod should be planned in advance. As a result, the capacity that is required for the Management Pod can remain steady. When planning the storage capacity of the Management Pod, the CSP should consider the operational headroom for VNF files, snapshots, backups, VM templates, OS images, upgrade and log files.

When the Resource Pod is scaled up by adding hosts to the cluster, the newly added resources are automatically pooled resulting in added capacity to the compute cluster. New tenants can be provisioned to consume resources from the total pooled capacity that is available. Allocation settings for existing tenants must be modified before they can benefit from the increased resource availability. OpenStack compute nodes are added to scale out the Resource Pod. Additional compute nodes can be added by using the Integrated OpenStack Manager vSphere Web Client extension.

In the Edge Pod, additional Edge Nodes can be deployed in the cluster in conjunction with the physical leaf-spine fabric for higher capacity needs.

Performance

The platform is optimized for workload acceleration and the Three-Pod design offers maximum flexibility to achieve these goals. The configuration can support the needs of use cases with high bandwidth requirements. A dedicated DPDK accelerated Resource Pod or a hybrid standard and accelerated Resource Pod can be considered when designing for workload distribution. The separate Edge Pod provides not only the separation benefit but also alternatives with VM and bare metal options for the Edge Nodes.

Function Distribution

Network function designs are evolving to disaggregated control and data plane function. The separation also maximizes the distribution of data plane functions closer to the edge of the network with centralized control and management planes. A separated Resource or Edge Pod design allows for having different control plane versus data plane configurations and scale designs depending on the service offers.

Operations Management

The Management Pod provides centralized operation function across the deployment topology, be it a single or distributed data centers.

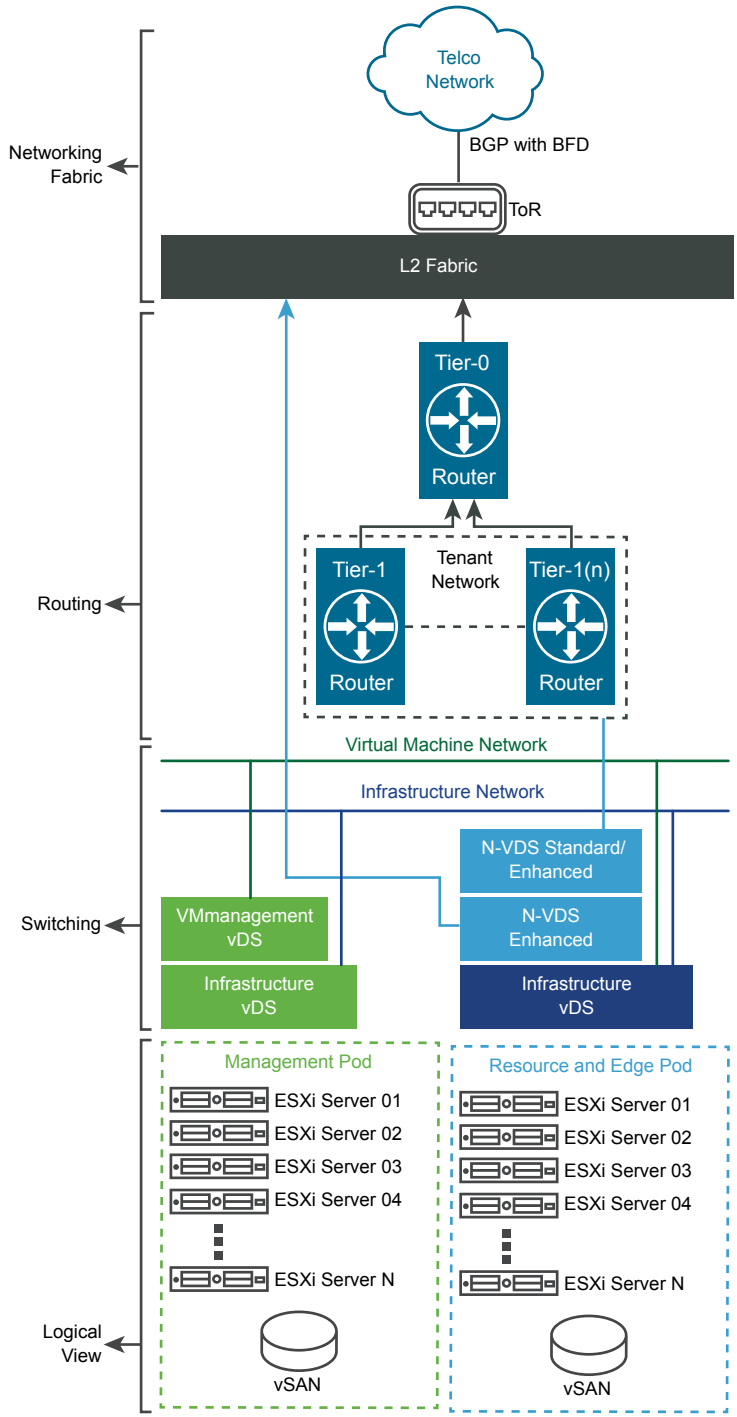
Two-Pod Configuration

The vCloud NFV OpenStack Edition facilitates combining the Edge and resource functionality in a single, collapsed Pod that provides a small footprint. CSPs can use a Two-Pod design to gain operational experience with the vCloud NFV OpenStack Edition platform. As demand grows, they can scale up and scale out within the Two-Pod construct.

An initial Two-Pod deployment consists of one cluster for the Management Pod and another cluster for the collapsed Edge & Resource Pod. Clusters are vSphere objects for pooling virtual domain resources and managing resource allocation. Clusters scale up as needed by adding ESXi hosts, whereas Pods scale up by adding new clusters to the existing Pods. This design ensures that management boundaries are clearly defined, capacity is managed, and resources are allocated based on the functionality that the Pod hosts. vCloud NFV OpenStack Edition VIM components allow for fine grained allocation and partitioning of resources to the workloads, regardless of the scaling method that is used.

The diagram shows a typical Two-Pod design with all management functions located centrally in the Management Pod. Edge and resource functions are combined in the collapsed Edge & Resource Pod. During the initial deployment, two clusters of ESXi hosts are used, one for the Management Pod, and another one for the collapsed Edge & Resource Pod. Additional clusters can be added to each Pod as the infrastructure is scaled up.

Figure 7-2. Two-Pod Conceptual Design



Logical View

- Management Pod. This Pod is responsible for hosting all NFV management components. These functions include resource orchestration, analytics, BCDR, third-party management, NFVO, and other ancillary management.

- Edge & Resource Pod. This Pod provides the virtualized runtime environment, that is compute, network, and storage, to execute workloads. It also consolidates the NSX Edge Node to participate in the East-West traffic and to provide connectivity to the physical infrastructure for North-South traffic management capabilities. Edge Nodes can be deployed in a VM form factor only.

Routing and Switching

Before deploying a Two-Pod configuration, a VLAN design needs to be considered as a best practice to isolate the traffic for infrastructure, VMs, and VIM.

The general design of the routing and switching fabric is similar to the Three-Pod design, with the resource and Edge fabric converged into a single pod.

Design Considerations

A Two-Pod configuration provides a compact footprint design choice for telco workloads.

Footprint

High level of resiliency is provided by using four-host clusters with vSAN storage in the initial deployment. Four-host clusters also ensure a highly available Management Pod design, because clustered management components such as vCenter Server active, standby, and witness nodes can be placed on separate hosts. The same design principle is used for clustered OpenStack components such as database nodes.

When using high-end class servers, this design maximizes the resource utilization and densification of workloads and Edge Nodes in the same Pod.

Scale Flexibility

The initial number and sizing of management components in the Management Pod should be planned in advance. As a result, the capacity that is requirement for the Management Pod can remain steady. When planning the storage capacity of the Management Pod, the CSP should consider the operational headroom for VNF files, snapshots, backups, VM templates, OS images, upgrade and log files.

The collapsed Edge & Resource cluster sizing will change according to the VNF and networking requirements. When planning the capacity of the Edge & Resource Pod, tenants must work with VNF vendors to gather requirements for the VNF to be deployed. Such information is typically available from the VNF vendors as deployment and sizing guidelines. These guidelines are directly related to the scale of the VNF service, for example to the number of subscribers to be supported. In addition, the capacity utilization of Edge Nodes must be taken into consideration, especially when more instances of Edge Nodes are deployed to scale up as the number of VNFs increases.

When scaling up the Edge & Resource Pod by adding hosts to the cluster, the newly-added resources are automatically pooled, resulting in added capacity to the compute cluster. New tenants can be provisioned to consume resources from the total pooled capacity that is available. Allocation settings for existing tenants must be modified before they can benefit from the increased resource availability.

Operations Management

The Management Pod provides centralized operation function across the deployment topology, be it a single or distributed data centers.

Next Generation Data Center Evolution

8

This section covers a set of solutions and use case scenarios to modernize the CSP cloud infrastructure environment with vCloud NFV OpenStack Edition.

This chapter includes the following topics:

Private Data Center NFV Transformation

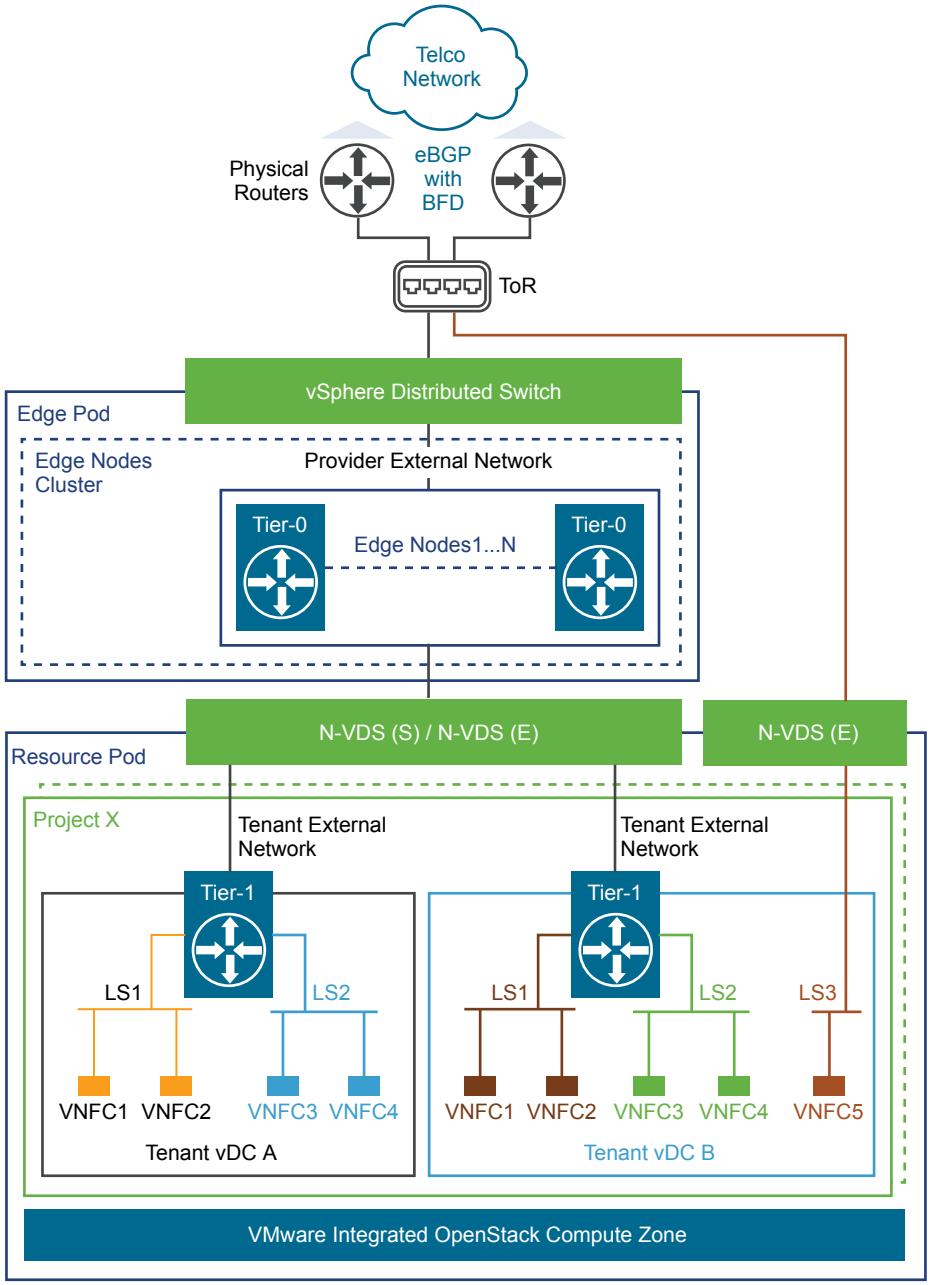
Transform the cloud infrastructure in the CSP private data centers with open source OpenStack IaaS layer together with the VMware core virtualization technologies.

Scope

This release of vCloud NFV OpenStack Edition introduces VMware Integrated OpenStack together with NSX-T Data Center, which is the next-generation advanced networking stack of the NFVI. vCloud NFV OpenStack Edition explores the transformation of the CSP private cloud environments to next-generation NFV networks by using multitenant design objectives.

Considering a Three-Pod design, the following diagram illustrates how the Resource Pod enables both accelerated and normal workloads with the NSX-T Data Center N-VDS switch in Enhanced Data Path (N-VDS (E)) and Standard (N-VDS (S)) modes. A separate Edge Pod provides external connectivity with different options for scale and performance.

Figure 8-1. Private Cloud with Multiple Tenants



The diagram illustrates how a fully-integrated VMware Integrated OpenStack Compute Zone, Project and Tenant VDC, as well as NSX-T logical switches and Tier-1 logical routers can be leveraged to provide a multitenant environment for deploying VNFs. The Edge Pod hosts the Tier-0 logical routers are used as provider routers for external network connectivity. The NSX Manager provisions the Tier-0 Edge routers, whereas other networking components can be created by using VMware Integrated OpenStack.

Design Objectives

Transform the private data center by using a common shared pool of resources with multitenant compute, storage, and network isolation. Use the NSX-T Data Center advanced networking and acceleration capabilities for control and data plane workloads.

Site Design

The Management, Resource, and Edge Pods are deployed in each data center that is part of a site. Pods that belong to a site can be scaled to meet the use cases and requirements of their data centers.

Resource Pods in each data center are sized to meet the scale and traffic requirements of the VNFs and other data plane functions.

Respectively, a pair of availability zones should be created in each site for redundancy and fault tolerance. A group of homogenous hosts, also called host aggregates should be mapped across the availability zones. A single tenant (Project and Tenant VDC) map to the host aggregates to deploy workloads for control and user plane functions.

Compute Resource Isolation

Resource pools represent a unit of CPU, memory, and storage that are portioned from the shared infrastructure pool. By default, when a Tenant VDC is created, a resource pool is also created with resource allocations already set. The resource pool is allocated to the Tenant VDC and can be scaled to meet future demand.

By using the Tenant VDC, virtual data centers for tenants can be created under different compute nodes (host aggregates) that provide specific SLAs for each telco workload.

Quotas on projects set limits on the OpenStack resources across multiple compute nodes or availability zones, but they do not guarantee resource reservation. Resource reservation is guaranteed by using a Tenant VDC to allocate CPU and memory for an OpenStack project or tenant on a compute node. In a multitenant environment, noisy neighbor scenarios can also be avoided in this way.

In vCloud NFV OpenStack Edition, workloads are mapped to Flavors with predefined memory, CPU, and storage capacity. In VMware Integrated OpenStack, a Flavor is a resource template that is configured to a workload when an instance is created.

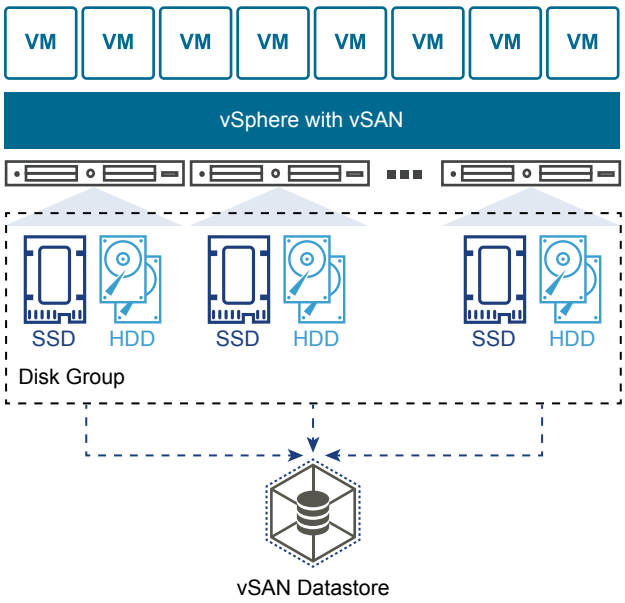
By modifying the metadata of the Flavor that is used to create the instance, QoS resource allocations such as limits, reservations, and shares, for CPU, memory, disk IOPS, and Virtual Network Interface (VIF) can be controlled. All instances that are created by using the Flavor inherit its metadata settings.

See *Multi-Tenancy with QoS* for more information about QoS shaping and control.

Shared Storage

In this design, VMware vSAN is used as a shared storage solution. Other storage options with NFS, iSCSI, and Fiber Channel are also supported. vSAN aggregates local or directly attached data storage devices to create a single storage pool that is shared across all hosts in the vSAN cluster. vSAN eliminates the need for external shared storage and simplifies the storage configuration and VM provisioning.

Figure 8-2. Shared Storage by Using vSAN



Using a shared datastore is recommend for high availability and fault tolerance. In case of host failure, vSphere HA can restart the VMs on another host.

A cluster in vCenter Server provides management of software-defined storage resources the same way as with compute resources. Instead of CPU or memory reservations, limits, and shares, storage policies can be defined and assigned to VMs. The policies specify the characteristics of the storage and can be changed as business requirements change.

Design for Tenants

The diagram in the *Scope* section of this use case shows how the fully-integrated VMware Integrated OpenStack Compute Node, Project and Tenant VDC, and NSX-T Data Center logical switches, and Tier-1 logical routers can be leveraged to provide a multitenant environment for VNF deployment. The Edge Pod hosts the NSX-T Data Center Tier-0 logical routers that are used as provider routers for external network connectivity. The NSX Manager is used to provision the Tier-0 Edge routers, while other networking components can be created by using VMware Integrated OpenStack.

The VMware Integrated OpenStack Projects and Tenant VDC represent tenants.

The CSP maps a vSphere compute cluster to a VMware Integrated OpenStack compute node, then for each tenant the CSP allocates and reserves resources by using the VMware Integrated OpenStack Tenant based constructs. Every Tenant VDC is associated with a resource pool within the compute cluster for resource guarantee.

The separation of network access between VMware Integrated OpenStack projects is important for multitenancy. VMware Integrated OpenStack integrates with the NSX Manager to create isolated layer 2 tenant networks. NSX-T Data Center Tier-1 logical routers allow tenants to route traffic between their tenant networks that are created by using the VMware Integrated OpenStack user interface.

The integrated design further helps the VMware Integrated OpenStack admin to create external IP networks for each project. Each tenant admin then attaches this network on a Tier-1 logical router by using SNAT to hide internal tenant networks. The external network on each project auto connects to a Tier-0 logical router on a separate network path then Tier-0 routers connect to the external physical network. This design choice restricts the access of the tenant admin to the Tier-1 router and creates easy connectivity to the provider router to carry tenant traffic without manual steps.

Dual-Mode Switching with NSX-T N-VDS

The tenant external network can carry network traffic to NSX-T Data Center switches that can be either N-VDS Standard or N-VDS Enhanced data path modes or both. This dual-mode N-VDS switching fabric can be used to design for accelerated and non-accelerated workloads within the same or separated compute clusters.

The N-VDS Standard switch can be used to carry the tenant Tier-1 to provider Tier-0 traffic, as well as East-West overlay Geneve traffic between VNFCs. VNFCs that require high data path traffic can use the N-VDS Enhanced DPDK fabric on a VLAN port group and connect to the ToR switch. This design can be leveraged by services such as the PGW in a mobile network for broadband connectivity, streaming applications for audio and video distribution, and SBC for IP voice peer-to-peer communications.

NSX-T N-VDS Standard and N-VDS Enhanced switches are used to carry tenant traffic to the provider router and external physical network respectively. The design allows for multiple tenant Tier-1 routers to use the same provider Tier-0 Edge Pod.

Service Provider Edge Cluster Configuration

Before establishing the VMware Integrated OpenStack configuration, the CSP has to create an Edge Node provider cluster. The NSX Edge Node cluster consists of logical Tier-0 routers. The Edge Node cluster can consist of either VM or bare metal form-factors. The bare metal Edge is installed on a physical server providing higher throughput data rates.

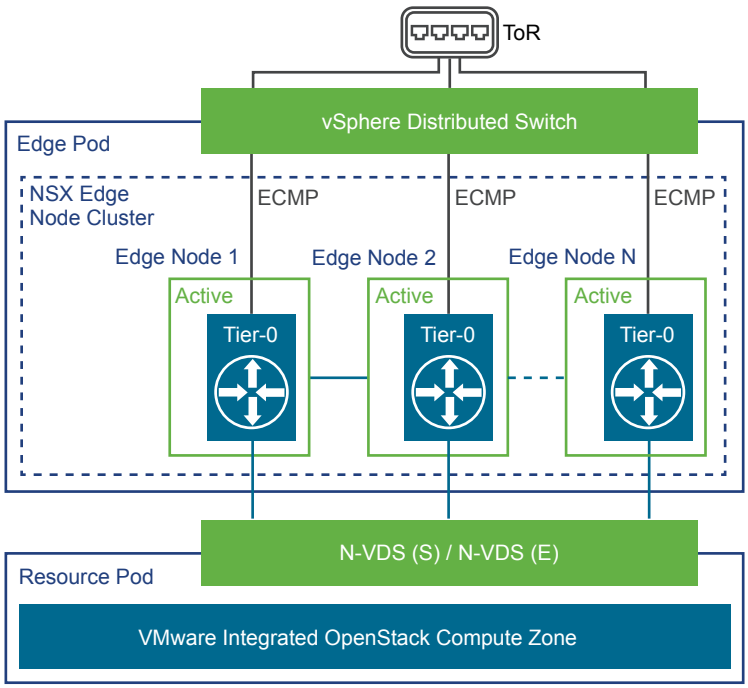
Table 8-1. Edge Node Options

Edge Node Type	Use
VM form-factor	■ Production deployment with centralized services like NAT, Edge firewall, and load balancer. ■ Workloads that can tolerate acceptable performance degradation loss with virtual edges. ■ Can tolerate lower failure convergence by using BFD (3 seconds). ■ Lower cost options instead of dedicated bare-metal nodes ■ Test proof of concept and trial setups.
Bare metal form-factor	■ Production deployment with centralized services like NAT, Edge firewall, and load balancer. ■ Higher throughput more than 10Gbps. ■ Faster failure convergence using BFD (less than 1 second).

Edge Node Active-Active

In an Edge Node Active-Active configuration, Tier-0 routers are hosted on more than one Edge Nodes at a time to provide high availability. In ECMP mode, the traffic is load balanced between the links to the external physical routers. A maximum of 8 Edge Nodes can be configured in ECMP mode to provide scalable throughput that spreads across the Edge Node physical uplinks to the provider network. Stateful services like NAT and Firewall cannot be used in this mode.

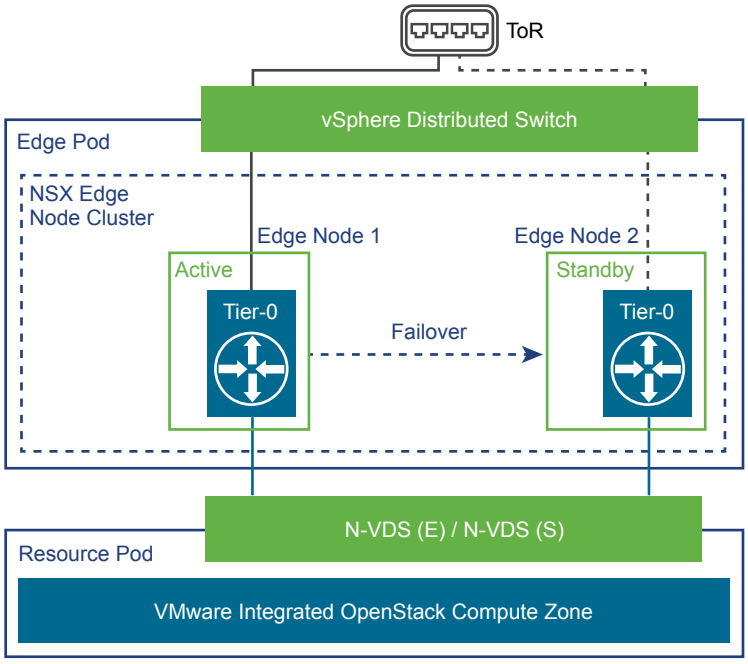
Figure 8-3. Edge Node Active-Active Design



Edge Node Active-Standby

A high availability configuration where a Tier-0 router is active on a single Edge Node at a time. This mode is required when stateful services like NAT and Firewall must remain in a constant state of synchronization between the active and standby Tier-0 routers on the Edge Node pair.

Figure 8-4. Edge Node Active-Standby Design



Dynamic Routing

Tier-0 routers can be connected to physical routers by using BGP or static routes. If static routes are used, every newly created external network has to be added manually to the Tier-0 router that peers with the physical routers.

The NSX-T Data Center Edge Node also supports fast failure recovery by using Bidirectional Forwarding Detection (BFD) that is integrated with BGP. VM form-factor edges support a minimum timeout of one second with three retries, providing a three second failure detection time between nodes. With bare-metal nodes, the detection or convergence timeout is less than one second.

For more information on NSX-T Data Center, see the NSX-T Reference design.

Workload Life Cycle Management

Once the compute, storage, and networking environment are set for the tenant, workloads can be onboarded to the Tenant VDC.

For more information on workload onboarding, see the *VNF Onboarding* section.

To enable dynamic optimization by using runtime operational intelligence, see the *Analytics-Enabled Reference Architecture* section.

Workload Acceleration

A key enhancement in this version of the vCloud NFV OpenStack Edition platform is the support for advanced packet processing acceleration. Both control or user plane workloads can take advantage of this capability and achieve higher throughput, reduced latency, and scale in performance.

Scope

VNF components that require high network throughput can achieve high data plane performance by using the advanced switching fabric that is introduced with the NSX-T Data Center N-VDS and Edge cluster acceleration in Three-Pod configurations.

Design Objectives

The VMware vCloud NFV OpenStack Edition platform includes NSX-T Data Center as the virtualized networking component. NSX-T Data Center leverages DPDK techniques to support data plane intensive VNFs efficiently. The NSX-T Data Center networking stack increases the CPU efficiency while preserving the existing functionality of the VMware NFV infrastructure. The new high-performance switching and enhanced platform awareness delivers the VMware advanced networking architecture that will be transparent to VNF providers and delivers kernel-based security, deterministic resource allocation, and linear scalability.

vSphere introduces support for a high-performance networking mode that is called N-VDS Enhanced Data Path, which works together with NSX-T Data Center. NSX-T Data Center N-VDS provides logical switching fabric that works in two modes, a Standard NSX-controlled Virtual Distributed Switch (N-VDS Standard) and Enhanced NSX-controlled Virtual Distributed Switch (N-VDS Enhanced). N-VDS Enhanced implements key DPDK features such as, Poll Mode Driver, Flow Cache, optimized packet copy, and provides better performance for both small and large packet sizes applicable for NFV workloads. N-VDS Enhanced mode provides three to five times faster performance compared to the vSphere Distributed Switch. VNF vendors can now use a high-performance virtual switch without sacrificing any of the operational benefits of virtualization such as vMotion and DRS.

Both modes of the N-VDS switch use dedicated physical NICs that are not shared with other switches. Because of this, they can exist on the same host and carry different traffic types.

Table 8-2. NSX-T Data Center Logical Switch Mode Options

Switch Mode	Use
N-VDS Enhanced	■ Recommended for data plane intensive workloads. ■ High transactional control plane VNFs. ■ Connectivity towards external networks can be overlay or VLAN backed.
N-VDS Standard	■ Suitable for control and management plane workloads. ■ VNFs that require overlay and VLAN backed connectivity. ■ VNFs that require stateful and stateless Edge services such as load balancer, firewall, and NAT

Acceleration with the N-VDS in Enhanced Datapath Mode

The Enhanced mode of the N-VDS switch uses DPDK and vertical NUMA alignment to accelerate workloads.

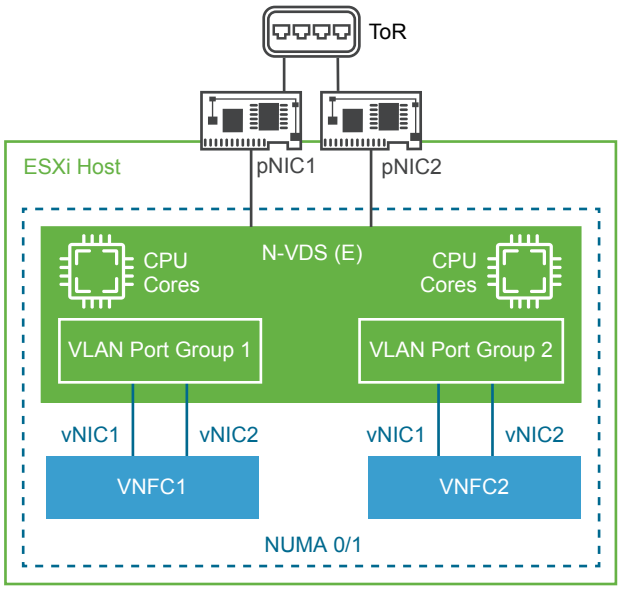
Hosts Preparation for Data Plane Intensive Workloads

Before data plane intensive workload can attach to an N-VDS Enhanced switch, the hosts in the Resource Pod need specific configuration.

- vNICs that are intended for use with the N-VDS Enhanced switch should be connected by using similar bandwidth capacity on all NUMA nodes.
- An N-VDS Enhanced switch with at least one dedicated physical NIC should be created on each host.
- The same number of cores from each NUMA node should be assigned to the N-VDS Enhanced switch.
- A Poll Mode Driver for the N-VDS Enhanced switch must be installed on the ESXi hosts for use of the physical NICs that are dedicated to the N-VDS Enhanced switch. You can download the driver from the VMware Compatibility Guide. Identify the correct drivers by looking for N-VDS Enhanced Data Path in the feature column of the VMware Compatibility Guide.

The host where the N-VDS Enhanced switch resides should be configured similarly to the topology depicted in the figure that follows. To ensure optimal performance, vertical alignment is established between the VNF-C, N-VDS Enhanced, and the CPU assigned to the NIC on the same NUMA node. The CPU scheduler ensures that the VNF-C's vCPU maps to physical CPU cores on the same NUMA node where the N-VDS Enhanced physical cores are mapped.

Figure 8-5. N-VDS Enhanced Switch Configuration

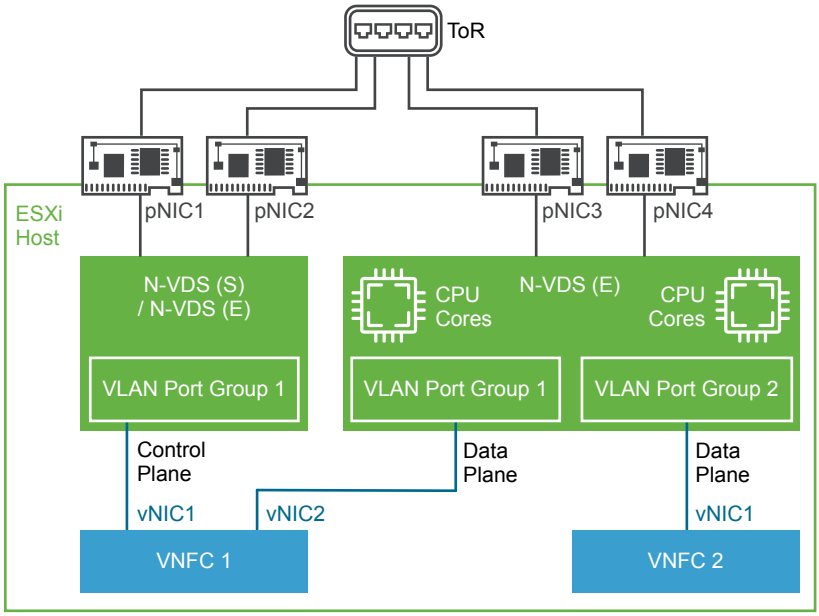


Workload Dual-Mode Connectivity

N-VDS Enhanced supports VLAN encapsulation and N-VDS Standard supports the overlay Geneve encapsulated traffic. If both VLAN and overlay traffic connectivity is required, both the N-VDS Enhanced and N-VDS Standard switches can be deployed on the same host with dedicated physical NICs to each switch. VNFs can have a VIF connected to the N-VDS Enhanced switch and another VIF connected to the N-VDS Standard switch for East-West connectivity.

Port groups are configured on both N-VDS Standard and N-VDS Enhanced switches enabling East-West communications between a data plane VNF-C and a control plane VNF-C.

Figure 8-6. N-VDS Dual-Mode Configuration



Acceleration by Using Bare Metal Edge

NSX Edge Nodes are also available in a bare metal form-factor with improved performance compared to the VM-based Edge Nodes.

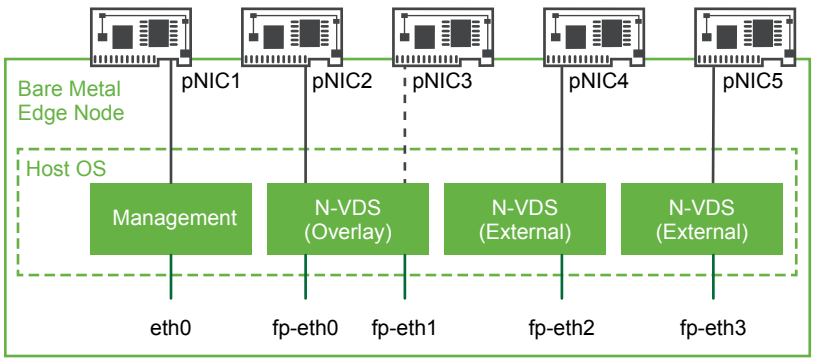
A Three-Pod configuration is required for a bare metal NSX Edge cluster deployment. The Edge cluster acts as an Edge router that connects to the CSP's physical router.

A bare metal NSX Edge delivers improved performance and higher throughput. The bare metal NSX Edge also provides sub-second BFD convergence and faster failover.

Physical Design

When a bare metal NSX Edge Node is installed, a dedicated interface is retained for management. Two physical NICs can be used for the management plane to provide high availability and redundancy.

Figure 8-7. Bare Metal Edge Physical View



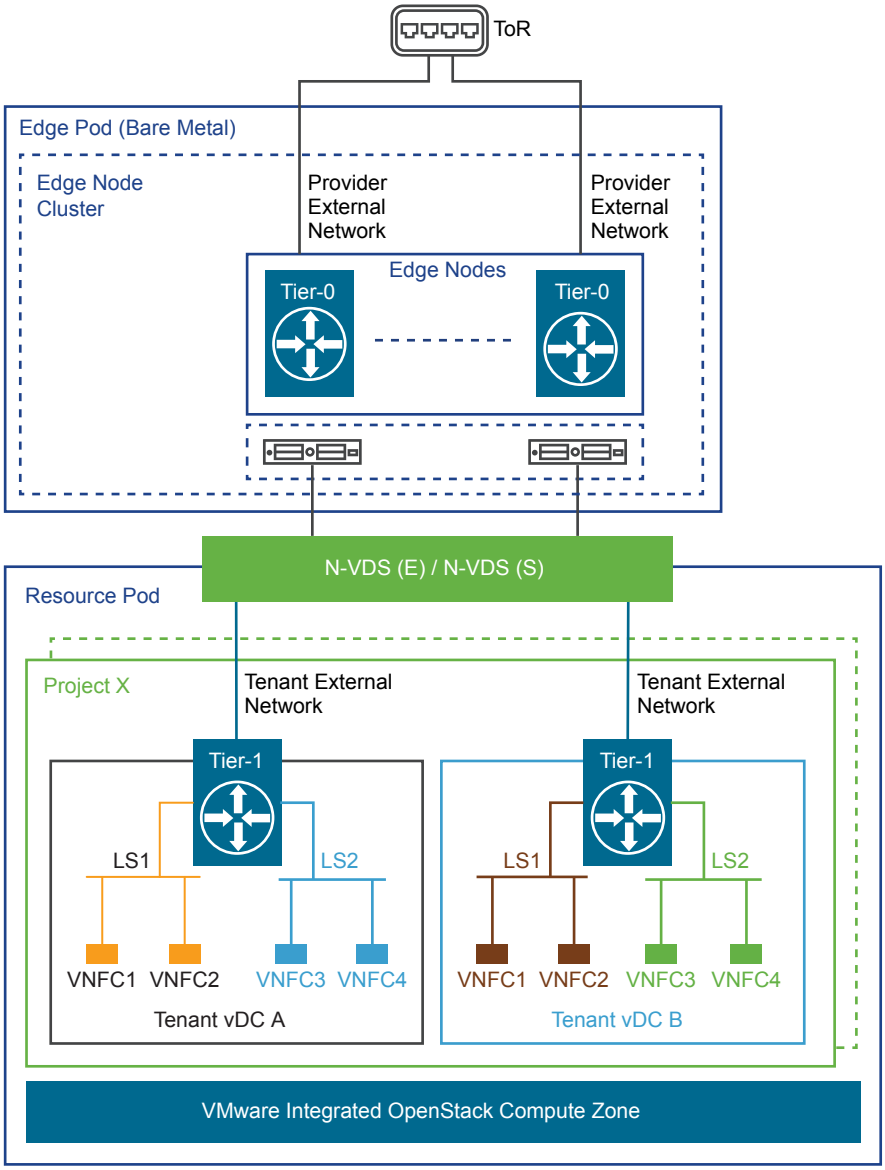
For each physical NIC on the server, an internal interface is created following the **`fp-ethX`** naming scheme. These internal interfaces are assigned to the DPDK FastPath and are allocated for overlay tunneling traffic or uplink connectivity to top-of-rack (ToR) switches. There is full flexibility in assigning `fp-eth` interfaces to physical NICs for overlay or uplink connectivity. Physical NICs can be assigned to each overlay or external connectivity for network redundancy. Because there are four `fp-eth` interfaces on the bare metal NSX Edge, a maximum of four physical NICs are supported for overlay and uplink traffic in addition to the primary interface for management.

The BFD protocol provides fast failure detection for forwarding paths or forwarding engines, improving loss of connectivity detection and therefore enabling quick response. The bare metal NSX Edges support BFD for both the interfaces towards the provider router and a BFD-like protocol operating between the NSX Edges and the resource hosts.

Logical Design

In a multitenant environment, VMware Integrated OpenStack is configured to use the N-VDS Standard switch together with a bare metal NSX Edge for higher performance. Regardless of the workload type, both control and data plane workloads can take advantage of this configuration. Tenant workloads can take advantage of bare metal router features like dynamic routing, firewall, NAT, and load balancer. With bare metal, there is no virtualization overhead, because it directly connects to physical NIC.

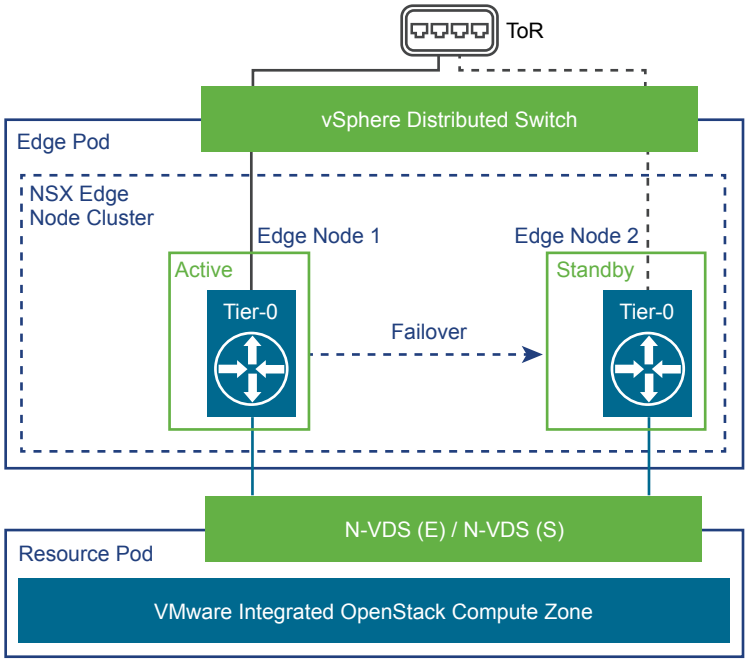
Figure 8-8. Bare Metal Edge Logical View



High Availability Options

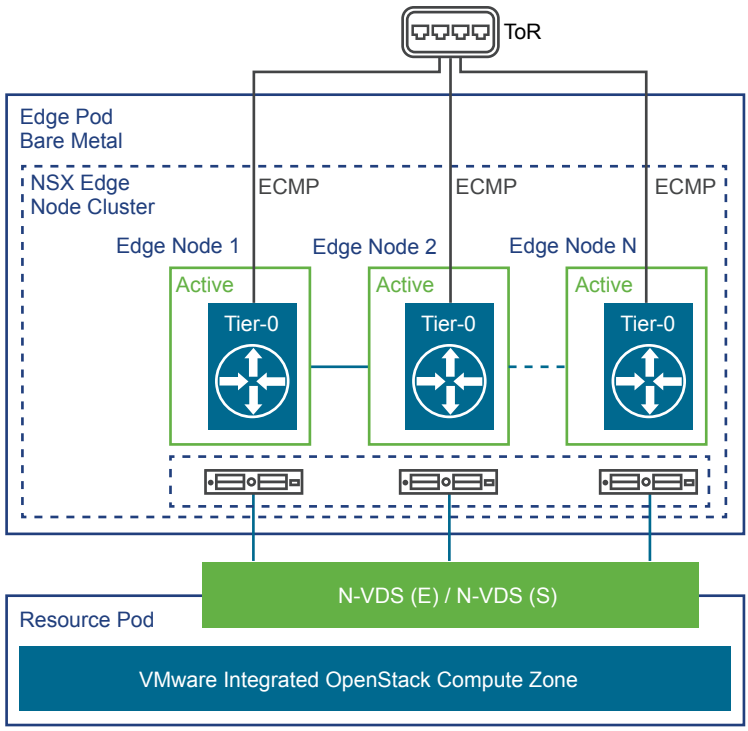
- Active-Standby mode. A high availability configuration that requires a minimum of two Edge Nodes hosting the Tier-0 routers. This mode also enables stateful services like NAT, firewall, and load balancer to be in constant state of sync between the active and standby Tier-0 on the Edge Nodes.

Figure 8-9. Bare Metal Edge Active-Standby Mode



- Active-Active Mode. Edge Node Active-Active configuration provides high availability mode as Tier-0 routers are hosted on more than one Edge Node at a time. In ECMP mode, traffic is load balanced between the links to external physical routers. A maximum of eight Edge Nodes can be configured in ECMP mode to provide scalable throughput that spreads across the Edge physical uplinks towards the provider network. Stateful services like NAT and firewall cannot be used in this mode.

Figure 8-10. Bare Metal Edge Active-Active Mode



Acceleration Using SR-IOV

SR-IOV is a specification that allows a single Peripheral Component Interconnect Express (PCIe) physical device under a single root port to appear as multiple separate physical devices to the hypervisor or the guest operating system.

SR-IOV uses physical functions (PFs) and virtual functions (VFs) to manage global functions for the SR-IOV devices. PFs are full PCIe functions that are capable of configuring and managing the SR-IOV functionality. VFs are lightweight PCIe functions that support data flowing but have a restricted set of configuration resources.

The number of virtual functions provided to the hypervisor or the guest operating system depends on the device. SR-IOV enabled PCIe devices require appropriate BIOS and hardware support, and SR-IOV support in the guest operating system driver or hypervisor instance.

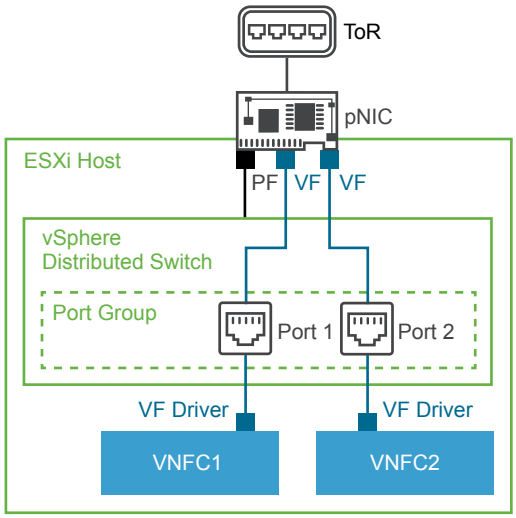
Prepare Hosts and VMs for SR-IOV

In vSphere, a VM can use an SR-IOV virtual function for networking. The VM and the physical adapter exchange data directly without using the VMkernel stack as an intermediary. Bypassing the VMkernel for networking reduces the latency and improves the CPU efficiency for higher data transfer performance.

vSphere supports SR-IOV in an environment with specific configuration only. Find a detailed support specification at the SR-IOV Support page.

In the topology below, the vSphere SR-IOV support relies on the interaction between the VFs and the PF of the physical NIC port for higher performance. VM network adapters directly communicate with the VFs that SR-IOV provides to transfer data. However, the ability to configure the VFs depends on the active policies for the vSphere Distributed Switch port group ports (VLAN IDs) on which the VMs reside. The VM handles incoming and outgoing external traffic through its virtual ports that reside on the host. The virtual ports are backed by physical NICs on the host and respectively the same physical NICs are also handling the traffic for the relevant port group to which they are configured. VLAN ID tags are inserted by each SR-IOV virtual function. For more information on configuring SR-IOV, see Configure a Virtual Machine to Use SR-IOV.

Figure 8-11. SR-IOV Virtual Function Configuration



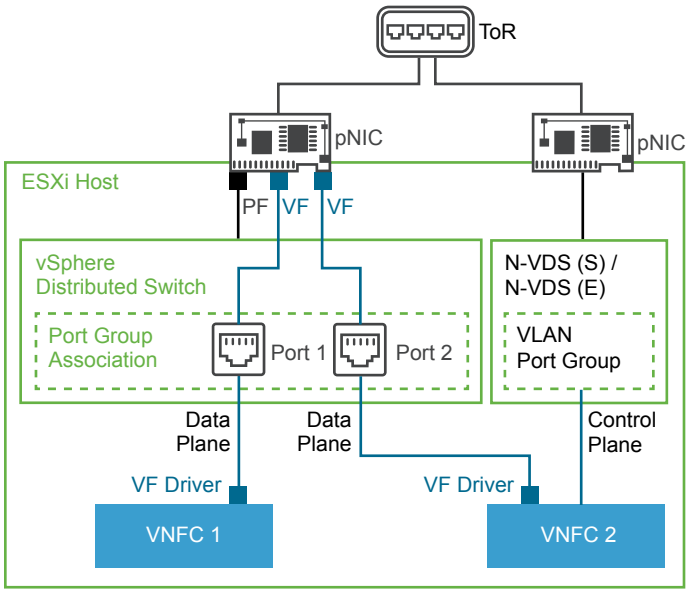
SR-IOV can be used for data intensive traffic, but it cannot use virtualization benefits such as vMotion, DRS and so on. Because of this reason, VNFs employing SR-IOV become static hosts. A special host aggregate can be configured for such workloads. The NSX-T Data Center fabric can be used to plumb interfaces into an N-VDS Standard switch as well as for VLAN and overlay connectivity.

SR-IOV Configuration by Using VMware Integrated OpenStack

VMware Integrated OpenStack supports a proxy TVD plug-in that must be configured to use the vSphere Distributed Switch and N-VDS networking. A port on a vSphere Distributed Switch port group can be configured from the VMware Integrated OpenStack API as an External network. Similarly, VMXNET3 interfaces can be plumbed into the VMs as well.

vSphere Distributed Switch and N-VDS Standard can be deployed on the same host with dedicated physical NICs to each switch. VNFs can have a VIF connected to both the vSphere Distributed Switch port group ports for North-South connectivity and N-VDS Standard switches for East-West connectivity.

Figure 8-12. SR-IOV Logical View



Hybrid Workload Execution Form-Factors

The objective of this use case is to provide seamless orchestration and management of VMs and Containers.

Scope

vCloud NFV OpenStack Edition allows seamless management of a hybrid execution environment consisting of VNFs that are implemented natively as VMs and containers. Through the integrated VMware Integrated OpenStack with Kubernetes, CSPs can easily deploy highly available Kubernetes clusters in a VMware Integrated OpenStack project for exclusive use by the respective cloud tenant. Once deployed, the container networking is fully integrated to NSX-T Data Center by using the NSX-T CNI framework for seamless connectivity within the tenancy.

Design Objectives

This version of the vCloud NFV OpenStack Edition platform supports the coexistence of container and VM workloads within the multitenant NFV environment. This design introduces the management plane, networking, and workload functions.

Container Management Plane

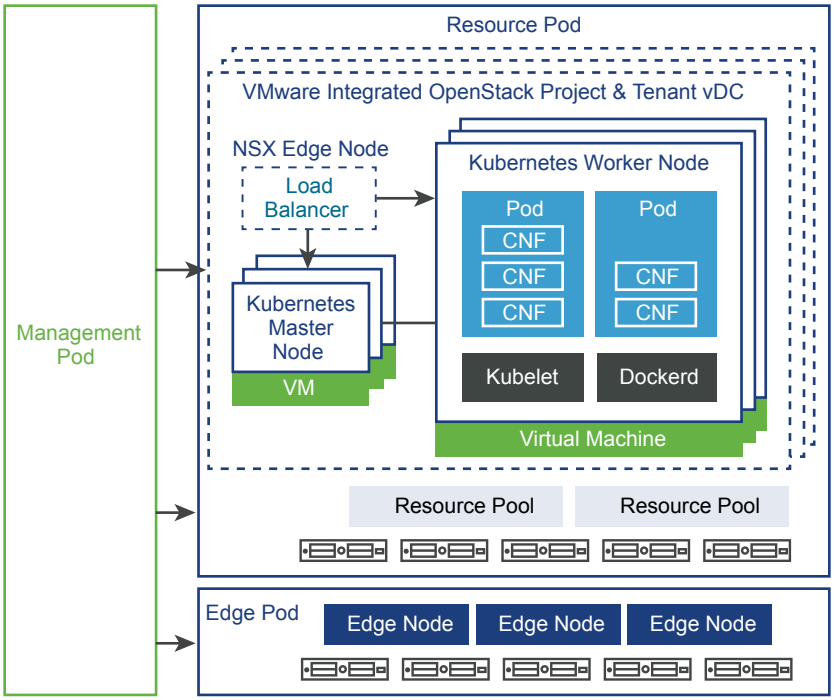
The container infrastructure is packaged as a separate appliance image that is called VMware Integrated OpenStack with Kubernetes. It is based on Docker runtime that is orchestrated by a Kubernetes cluster that is deployed within the tenancy and managed by the tenant admin.

VMware Integrated OpenStack Kubernetes Manager

The CSP can use the VMware Integrated OpenStack Kubernetes Manager to deploy and manage highly available Kubernetes clusters in an OpenStack environment. When deployed in the management cluster, the CSP can use the VMware Integrated OpenStack Kubernetes Managers to deploy Kubernetes clusters for tenants. During deployment, the Kubernetes manager allows the CSP to specify the VMware Integrated OpenStack project, the number of master and worker nodes to deploy the Kubernetes cluster. The cluster deployment creates the following nodes, as instances of each node are contained within separate VMs.

- Kubernetes master node. Hosts the replicated API servers, and clustered Etcd services. In addition, the worker nodes in a Kubernetes cluster can be scaled out or scale in to meet changing demands for capacity.
- Kubernetes worker node. Contains the Docker runtime and servers as the node where the container workloads are deployed. A minimum of two worker nodes are recommended for availability as additional nodes can be added depending on the capacity requirements.

Figure 8-13. VMware Integrated Open Stack Container Management Plane



VMware Integrated OpenStack Integration

The Kubernetes clusters are configured to use VMware Integrated OpenStack enterprise-grade services such as Keystone authentication for the cluster, Block Storage Cinder to provide persistent storage for stateful applications, and Neutron Load Balancing as a Service (LBaaS) for containerized application services. By combining Kubernetes with VMware Integrated OpenStack, tenant administrators can use a common infrastructure management layer to provision both VMs and containers.

Before a CNF can be onboarded, the CSP provisions a VMware Integrated OpenStack tenant by using the VMware Integrated OpenStack project construct to allocate resources to the tenant. The allocated resources are a sum of the resources that are required by the VNFs and the Kubernetes cluster that is deployed in the VMware Integrated OpenStack project. The sizing of the Kubernetes cluster dependents on the resource requirements of the CNF workloads that are planned to be deployed and the availability requirements of the Kubernetes cluster.

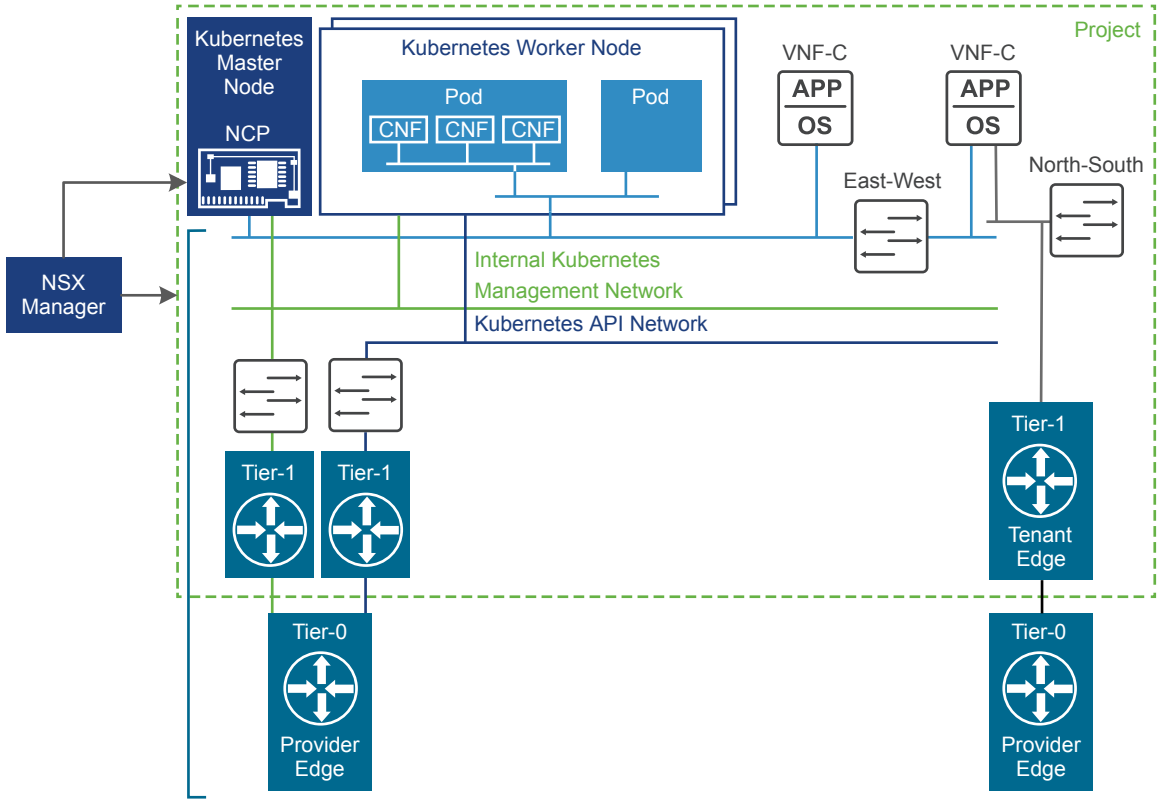
Tenant administrators assign quotas to container workloads from the resources available to the project. This provides tenant administrators with full control over VNFs and CNFs, while service providers simply monitor the total resource consumption of the project.

Container Networking

Container networking is integrated with NSX-T Data Center and provides a seamless fabric for both VM-based and container-based workloads.

NSX-T Data Center networking is used to enable the management, control, and data plane traffic within the multi-tenant logical constructs of VMware Integrated OpenStack.

Figure 8-14. NSX-T Container Networking Interfaces Plugin



Container Network Interface Plug-in

NSX-T provides seamless network virtualization for workloads running on either VMs or containers. This is achieved by using the NSX-T Container Network Interface Plug-in (NCP) that communicates with both NSX Manager and the Kubernetes API server on the Kubernetes Master nodes. NCP monitors changes to containers and other resources. It also manages networking resources such as logical ports, switches, routers, and security groups for the containers through the NSX API.

Once the Kubernetes cluster is deployed, a namespace is created and assigned to one or more tenant users. The NCP plug-in automatically creates an NSX-T Data Center logical Tier-1 router for the configured namespace and assigns it to the default Tier-0 provider router. Once the name space is created, pods and containers can be deployed and they can consume the NSX-T Data Center backed networks. VM-based workload connectivity to the containers is established by connecting containers to the NCP POD Network that is created when the namespace is provisioned.

The NSX CNI plug-in supports both the N-VDS Standard and Enhanced modes.

Switching and Routing

The NSX-T Data Center software driven container networking building blocks are identical to the building blocks that are used for building the East-West and North-South connectivity of the VM based VNF workloads. Since both are backed by the same NSX-T Data Center infrastructure, East-West connectivity can be provisioned between the hybrid workloads by using Logical Switches and Tier-1 routers.

When container networking building blocks are connected to the Tier-0 routers for North-South traffic, they provide the same level of functionality to the containerized workloads as they do to VM workloads.

Container Workloads

Container Network Functions (CNFs) are deployed by using standard Kubernetes development and integration methodologies.

The tenant container infrastructure that is deployed through VMware Integrated OpenStack with Kubernetes is identical to a typical standalone Kubernetes cluster in terms of the services and Kubernetes API endpoint. Container admins continue to use the same tools, clients and kubectl commands to manage their containerized network functions as they would in any other Kubernetes environment.

The Kubernetes cluster health is monitored by the Web-based Kubernetes Dashboard and the analytics solution discussed in the *Analytics-Enabled Reference Architecture* section.

Multi-Tenancy with QoS

The CSP can facilitate the NFV transformation on a shared resource infrastructure environment with multitenant consumption models. The design of those tenant environments is outlined in this section.

Scope

Multitenancy defines the isolation of resources and networks to deliver applications with quality. Because multiple tenants will share the same resource infrastructure, secure multitenancy can be enabled by using VMware Integrated OpenStack in a single cloud island and across distributed clouds. In addition to the built-in workload and resource optimization capabilities, predictive optimization can be enabled with analytics by using features like vSphere DRS.

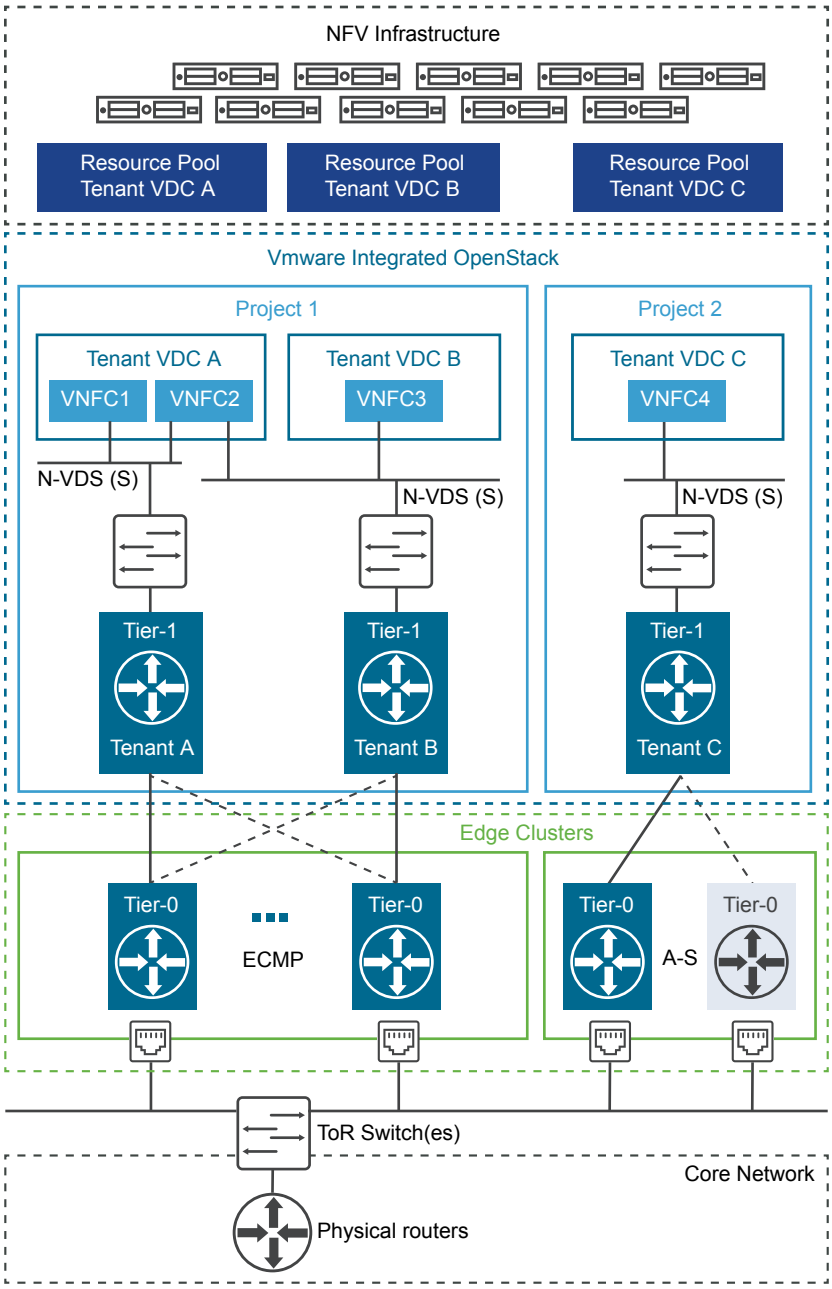
CSPs can converge their resource infrastructures across their IT and Network clouds enabling a multitenancy IaaS realization over it. Consumption models can serve both internal and external tenants over the common shared infrastructure to deploy and operate their respective workloads and services. Network, compute, and storage isolation, with QoS, are the design objective discussed in this section.

Design Objectives

A unit of tenancy is called a Tenant VDC within the scope of a Project. It is defined as a composition of dedicated compute, storage, and network resources and as workloads. The tenant is associated with a set of operational policies and SLAs. The Tenant VDC can be bound to a single tenant or shared across many tenants. Services such as HSS, DNS are examples of shared tenancy.

The design objectives include considerations for the compute and network resource isolation and automation.

Figure 8-15. Multi-Tenancy using VMware Integrated OpenStack



Management Plane

The management plane functions reside in the Management Pod. They are responsible for the orchestration of resources and operations. The management plane functions are local to each cloud instance providing the infrastructure management, network management, and operations management capabilities.

Resource isolation for compute and networking design are enabled together with vCenter Server, NSX Manager, and VMware Integrated OpenStack. Irrespective of the pod deployment configuration, VMware Integrated OpenStack provides the abstraction layers for multi-tenancy. vCenter Server provides the infrastructure for fine-grained allocation and partitioning of compute and storage resources, whereas NSX-T Data Center creates the network virtualization layer.

The concept of tenancy also introduces multiple administrative ownerships. A cloud provider, that is the CSP admin, can create a resource pool allocation for a tenant who in turn would manage the underlying infrastructure and overlay networking. In VMware Integrated OpenStack, multiple tenants can be defined with assigned RBAC privileges to manage compute and network resources as well as VNF onboarding.

Compute Isolation

Allocation of compute and storage resources ensures that there is an optimal footprint available to each tenant that is used to deploy workloads, with room for expansion to meet future demand.

Tenant VDCs, are provide a secured multitenant environment to deploy VNFs. Compute resources are defined as resource pools when a Tenant VDC is created. The resource pool is an allocation of memory and CPU from the available shared infrastructure, assignable to a Tenant VDC. More resources can be added to a pool as capacity needs to grow. The Tenant VDC can also stretch across multiple resource clusters residing in different physical racks.

Though the resource pools can be further sub-segmented in smaller resource pools, this is not a recommended design.

Network Isolation

The advanced networking model of NSX-T Data Center provides a fully-isolated and secure traffic paths across workloads and tenant switch or routing fabric. Advanced security policies and rules can be applied at the VM boundary to further control unwarranted traffic.

NSX-T Data Center introduces a two-tiered routing architecture which enables the management of networks at the provider (Tier-0) and tenant (Tier-1) tiers. The provider routing tier is attached to the physical network for North-South traffic, while the tenant routing context can connect to the provider Tier-0 and manage East-West communications. The Tier-0 will provide traffic termination to the cloud physical gateways and existing CSP underlay networks for inter-cloud traffic communication.

Each Tenant VDC will have a single Tier-1 distributed router that provides the intra-tenant routing capabilities. It can also be enabled for stateful services such as firewall, NAT, load balancer, and so on. VMs belonging to Tenant A can be plumbed to multiple logical interfaces for layer 2 and layer 3 connectivity.

By using VMware Integrated OpenStack as the IaaS layer, user profile and RBAC policies can be used to enable and restrict access to the networking fabric at the Tier-1 level.

QoS Resource Allocation

To avoid contention and starvation, compute, storage, and network isolation as well as QoS policies should be applied consistently to the workloads.

The CSP admin can allocate and reserves resources for tenants by using Tenant VDCs. Every Tenant VDC is associated with a resource pool across the Resource Pods. The resource settings of the resource pool are managed from VMware Integrated OpenStack. This ensures that every Tenant VDC allocates the resources to which it is entitled, without exceeding the infrastructure resource limits, such as CPU clock cycles, total memory, network bandwidth, and storage.

QoS policies can be applied to the VMs so that they receive a fair share of resources across the infrastructure pool. Each VM configuration is taken from a template that is called a Flavor. QoS can be configured by using Flavor metadata to allocate CPU (MHz), memory (MB), storage (IOPS), and virtual interfaces (Mbps).

QoS can be shaped by setting boundary parameters that control the elasticity and priority of the resources that are assigned to the VNF component executing within the VM.

- Reservation that is the minimum guarantee. Reservations ensure a minimum guarantee to each VM when it is launched.
- Limit that is the upper boundary. Should be used with caution in a production environment, because it restricts the VM from bursting utilization beyond the configured boundaries.
- Shares that are the distribution of resources under contention. Shares can be used to prioritize certain workloads over others in case of contention. In case resources are over-provisioned across VMs and there is resource contention, the VM with the higher shares will get the proportional resource assignment.

For control plane workload functions, a higher order elasticity can be acceptable and memory can be reserved based on the workload requirement. For data plane intensive workloads, both CPU and memory should be fully reserved. Storage IO and network throughput reservations need to be determined based on the VNF needs.

Automation

To meet the operational policies and SLAs for workloads, a closed-loop automation is necessary across the shared cloud infrastructure environment.

The vCloud NFV OpenStack Edition leverages the combination of vSphere DRS and Nova Scheduler to optimize the initial and runtime placement of workloads to ensure health and performance to the cloud infrastructure. Tenant VDCs and workloads are monitored to ensure resources are being tuned and balanced dynamically.

Capacity and demand planning, SLA violations, performance degradations, and issue isolation capabilities can be augmented with the analytics-enabled reference architecture. The analytics-enabled architecture provisions a workflow automation framework to provide closed-loop integrations with NFVO and VNFM for just-in-time optimizations.

For more information, see the *Analytics-Enabled Reference Architecture* section.

Distributed Clouds

The NFV transformations in the private cloud will span distributed topologies. User plane functions such as EPC gateways, IMS media for example, will move closer to the edge clouds for lower latency and compute proximity. Control plane functions can be still centralized.

Scope

Distributed clouds in this reference architecture are examined from a management and operations perspective.

Design Objectives

Along with the distribution of workload functions, user plane and control plane, the multiple cloud sites are also used for geographic redundancy for fault and maintenance domains. Multiple administrative ownerships also need to be considered across the topology and regions.

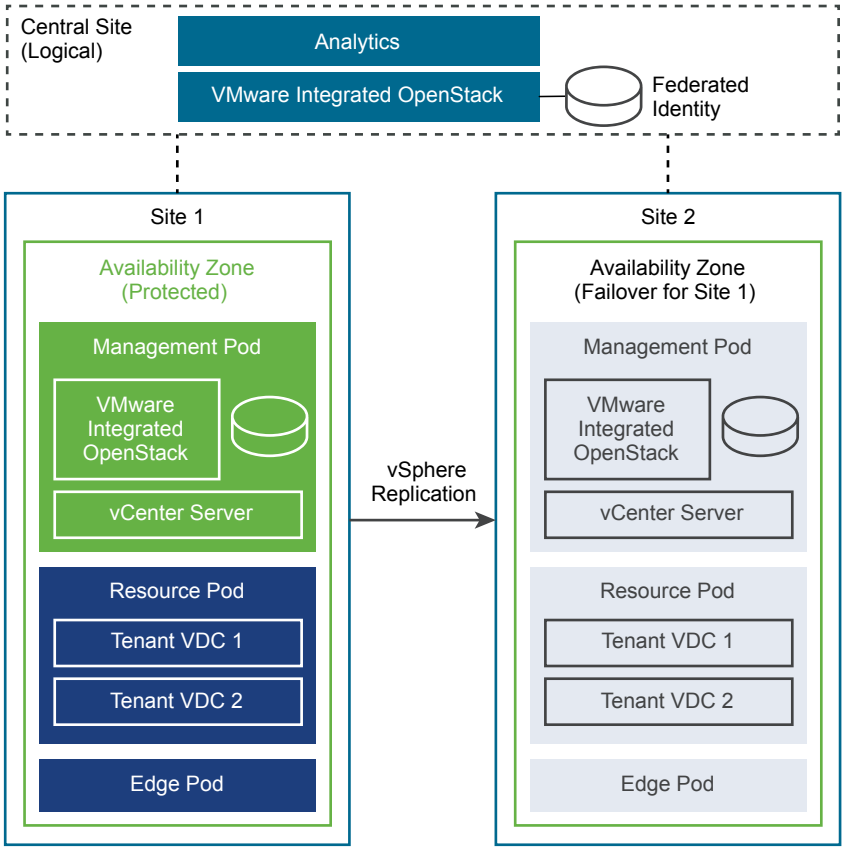
Site Design

In a distributed cloud topology, VNFs are stitched across the different cloud islands to delivery services. Each site can also have multiple instances of the NFV environment for redundancy and scale.

Each site is a standalone island with its own instantiation of the Management, Resource, and Edge Pods. Multi-site designs are typically used for load distribution and for higher availability in case of catastrophic site failures.

The design assumes an active-standby configuration for geographic redundancy. The geographic redundancy design can be implemented with proximal sites or with larger physical distances, as network latency between sites should be considered.

Figure 8-16. Distributed Clouds



Cloud Management Plane

The management plane functions reside in the Management Pod. The management plane functions are local to each site providing the virtual infrastructure and host management, network management, and operations management capabilities.

Resource Orchestration

The management of the virtual infrastructure by using VMware Integrated OpenStack functions is local to the individual cloud islands, however one of the instances can become the global context. Indicated as the central site in the diagram. The central VMware Integrated OpenStack federates identity across the various distributed Keystone instances within each of the distributed sites. The global VMware Integrated OpenStack instance can then be used as the centralized tenant authentication and delegation to the various distributed cloud instances.

Federated identity and the Horizon dashboard run in the central instance.

Operations Management

Operations management functions by using the VMware vRealize suite of components can be deployed in a single centralized site which can be used to collect and process analytics. The latency between sites should be considered in the design.

For more details on the operations management design, see the *Analytics-Enabled Reference Architecture* section.

Site Availability

For site disaster scenarios and maintenance needs, multiple sites are used to implement geographic redundancy, which ensures that a secondary site can be used to complement the primary site.

Redundancy Choice	Design Considerations
Singe Site	■ For non-critical services where longer downtime is acceptable. ■ Services are stateless and traffic can be routed to another instance or secondary site in case of failure. ■ Partial site loss which can be recovered by using vSphere HA and shared storage. ■ Backup-based approaches are acceptable for VM state recovery.
Active-Standby	■ Suitable for a recovery point objective with acceptable maximum of five minutes. ■ High speed and capacity links can be established between primary and secondary site ■ Higher capital costs required to maintain replicas in recovery site
Active-Active	■ Suitable for mission-critical services with almost no downtime ■ Service are deployed in active-active configuration across multiple sites ■ Any site should be able to handle the load of both active sites in case of failure ■ Higher capital cost is required in each site in case either site fails ■ This configuration is currently not available in this release

Availability Zones

Redundancy can be built in a single site to protect against power outages, for example. However, this design choice comes with duplicate hardware cost overhead. This design is out of scope of this reference architecture.

By design, multiple availability zones belong to a single region. The physical distance between the availability zones is short enough to offer low, single-digit latency and large bandwidth between the zones. This architecture allows the cloud infrastructure in the availability zone to operate as a single virtual data center within a region. Workloads can be operated 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 multi-site design.

Multi-Region

Multiple sites support placing workloads closer to the CSP's customers, for example, by operating one site in US Northern California and one site in US Southern California. In this multi-site design, a secondary site can become a recovery site for the primary site. When components in the primary site become compromised, all nodes are brought up in the failover site.

VMware vSphere Replication is used to replicate VM disks from the primary to the secondary site. In case of failure, the Resource Pod workloads can be brought up on the secondary site and routing updated to divert subscriber traffic to the secondary site.

Thought the state and database configuration for VMware Integrated OpenStack are replicated similarly, the control plane functions of VMware Integrated OpenStack need be to updated manually. For more information, see the Administering VMware Integrated OpenStackVMware Integrated OpenStack guide.

Workload On-Boarding

The VNF onboarding process is typically a collaborative effort between the VNF vendor and the CSP. A prescriptive set of steps need to be followed to package and deploy a VNF. The VMware Ready for NFV program is a good vehicle to pre-certify the VNFs and the onboarding compliance with the vCloud NFV OpenStack Edition platform to ensure smooth deployment in CSP's environment.

Scope

The workload life cycle involves provisioning of the cloud infrastructure, VNF packaging, resource assignment and configuration, deployment, and its placement.

Design Objectives

The workload onboarding process is aimed at VNFs that are deployed as native VM applications. This section covers the process of packaging, instantiation, to placement.

VMware Integrated OpenStack Conceptual Design

The conceptual design provides a high-level view of the roles, areas of responsibility, and tenant flow that are required to upload an image, onboard, and deploy it.

The VNF onboarding process is typically a collaborative effort between the VNF vendor and the CSP. Before a VNF is onboarded, the VNF vendor must provide the CSP with all the prerequisites for the successful onboarding of the VNF. This includes information such as the VNF format, number of the required networks, East-West and North-South network connectivity, routing policy, security policy, IP ranges, and performance requirements.

VNF Format and Packaging

VMware Integrated OpenStack supports ISO, VMDK, and OVA formats natively. Non-native formats such as RAW, QCOW2, VDI, and VHD are also supported after automatic conversion by the import process. The initial format of the VNF is taken into consideration for onboarding, as is the format of any additional components the VNF requires to function. Images are either able to be directly imported, or they can be converted. These formats can also be imported using the command line.

VNF Onboarding by Using VMware Integrated OpenStack

Once the initial VNF requirements, images, and formats are clarified, a project must be created so that to deploy the VNF in an operational environment. Projects are the VMware Integrated OpenStack constructs that map to tenants. Administrators create projects and assign users to each project. Permissions are managed through definitions for user, group, and project. Users have a further restricted set of rights and privileges. Users are limited to the projects to which they are assigned, although they can be assigned to more than one project. When a user logs in to a project, they are authenticated by Keystone. Once the user is authenticated, they can perform operations within the project.

Resource Allocation

When building a project for the VNF, the administrator must set the initial quota limits for the project. To guarantee resources to a VNF-C, a Tenant vDC can be used. A Tenant vDC provides resource isolation and guaranteed resource availability for each tenant. Quotas are the operational limits that configure the amount of system resources that are available per project. Quotas can be enforced at a project and user level. When a user logs in to a project, they see an overview of the project including the resources that are provided for them, the resources they have consumed, and the remaining resources. For fine-grained resource allocation and control, the quota of the resources that are available to a project can be further divided using Tenant vDCs.

VNF Networking

Based on specific VNF networking requirements, a tenant can provision East-West connectivity, security groups, firewalls, micro-segmentation, NAT, and LBaaS from by using the VMware Integrated OpenStack user interface or command line. VNF North-South connectivity is established by connecting tenant networks to external networks through NSX-T Data Center routers that are deployed in Edge Nodes. External networks are created by administrators and a variety of VNF routing scenarios are possible.

After the VNFs are deployed, their routing, switching, and security policies must be configured. There are many different infrastructure services available that can be configured in different ways, and in the coming sections of this document a couple of options are discussed.

Tenant networks are accessible by all Tenant vDCs within the project. Therefore, the implementation of East-West connectivity between VNF-Cs in the same Tenant vDC, and the connectivity between VNFs in two different Tenant vDCs belonging to the same project, is identical. Tenant networks are implemented as logical switches within the project. The North-South network is a tenant network that is connected to the telecommunications network through an N-VDS Enhanced for data-intensive workloads or by using N-VDS Standard through an NSX Edge Cluster.

VMware Integrated OpenStack exposes a rich set of API calls to provide automation. The deployment of VNFs can be automated by using a Heat template. With API calls, the upstream VNF-M and NFVO can automate all aspects of the VNF life cycle.

VNF Package for VMware Integrated OpenStack Consumption

The VMware Ready for NFV is a program where VMware and VNF vendors collaborate to ensure that the VNF is interoperable with VMware vCloud NFV OpenStack edition. Based on experience, VMware provides best practices guidelines to help VNF vendors in preparing their VNFs for consumption by VMware Integrated OpenStack.

VMware vCloud NFV OpenStack Edition customers expect their VNF supplier to deliver a package that is ready to be consumed by VMware Integrated OpenStack. The package must support:

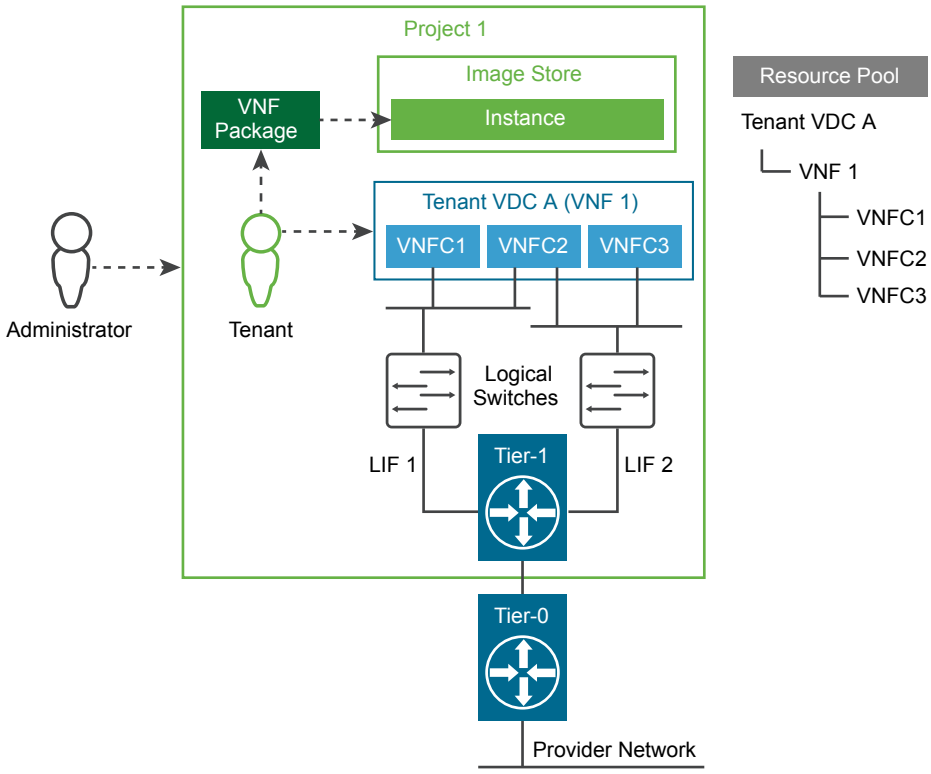
- Autonomous VNF Life Cycle Operations. Customers must be able to install, upgrade, and manage the VNF themselves.
- Compact. All components that are required to operate the VNF are packaged together. VNF Components (VNF-Cs) are clearly defined, use unique and specific names, and are included in the package.
- Unambiguous. The package provided to the customer must be verifiable. For example, an md5 hash could be used to verify that the package provided by the vendor is the same package that is installed by the customer.
- Holistic. The packaged VNF must include all configuration that is required to deploy a healthy VNF. For example, if the VNF is data plane intensive, all performance-related configuration such as CPU and NUMA affinity, CPU and memory reservations must be included in the package.

VNF Onboarding by Using VMware Integrated OpenStack

It is the VNF vendor's responsibility to provide onboarding instructions to the customers. This section provides general onboarding guidance assuming that the VNF is packaged in the format as stated in the *VNF package for VMware Integrated OpenStack* consumption section.

Before the onboarding processes can begin, the target cloud infrastructure should be designed and provisioned to meet the target VNF deployment. The following diagram illustrates the physical, virtualized, and logical layers of the stack upon which the VNFs are onboarded.

Figure 8-17. VNF Onboarding Conceptual Design



VMware Integrated OpenStack utilizes both UI (Horizon Dashboard) and CLI methods for VNF onboarding. The following high-level steps are required to onboard a VNF.

1 Create a project and import the VNF package into the image store.

2 Launch the service instance from Project Orchestration Stacks. The launched instances can be seen from Project Instances.

3 Create Tenant VDC based on the requirement.

4 Create a Resource Pool. Allocate CPU and memory for an OpenStack tenant on a compute node that provides the resource guarantee for tenants.

5 To manage the Tenant VDC, use the VMware Integrated OpenStack command line utility.

6 When the VM's are powered on, the system will apply configurations for each VNF-C using the packaged configuration files.

7 Verify the instances are running properly.

8 Apply any EPA configurations to the instances so that the VNF can be placed appropriately by the DRS, NUMA, and Nova Schedulers.

VNF Placement

Once the VNF is onboarded, its placement is determined based on the defined policies and the target host aggregates where to deploy the VNF.

Before determining how the workloads are placed in the Resource Pod, the following initial considerations need to be taken into account, based on the traffic profile and characteristics of the VNF.

- Is it an accelerated and non-accelerated workload? For the purposes of this discussion, assume that accelerated workloads are Data Plane, while non-accelerated are Control Plane functions.
- Does the control plane workload need a dedicated host infrastructure or can they be co-located with the data plane functions?

To optimize the resource consumption across the hosts in the resource pools and achieve performance across workloads, the shared infrastructure needs to be designed and configured appropriately.

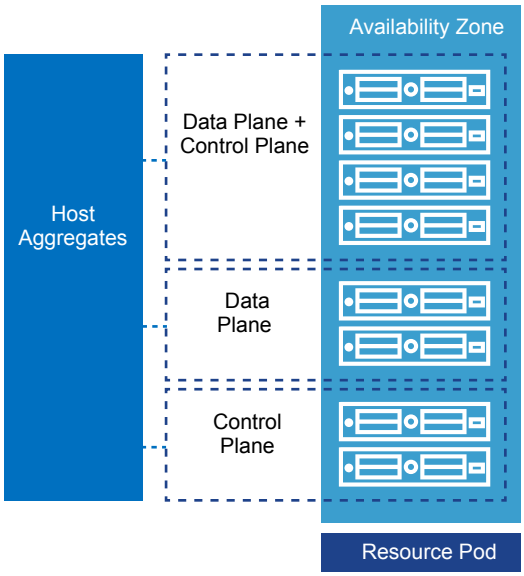
- Group compute nodes in Nova that are vCenter Server clusters in host aggregates with similar capability profiles.
- Define OpenStack Flavors with optimization policies.
- Define host affinity policies.
- Define DRS placement policies.
- Allow Nova schedulers to place initial workloads on a specific compute node and DRS to select a specific ESXi host.
- Allow DRS to optimize workloads at runtime.

Host Aggregate Definition

Host aggregates in the VMware Integrated OpenStack region have to be configured to meet the workload characteristics. The diagram from below shows the three potential options to consider.

- A host aggregate is a set of Nova compute nodes. A compute node in VMware Integrated OpenStack refers to vCenter Server clusters consisting of homogenous ESXi hosts that are grouped together by capability. A host aggregate with just one vCenter Server cluster can exist. Host Aggregate in VMware Integrated OpenStack can be considered as an elastic vCenter Server cluster that can offer a similar SLA.
- When using vSAN, a minimum of four hosts are recommended for each vCenter Server cluster. Operational factors need to be considered for each cluster when selecting the number of hosts.
- Configure the workload acceleration parameters for NUMA vertical alignment as discussed in the Workload Acceleration section.

Figure 8-18. Host Aggregates for Workload Placement



The DRS, NUMA, and Nova Schedulers ensure that the initial placement of the workload meets the target host aggregate and acceleration configurations defined in the policy. Dynamic workload balancing ensures the polices are respected when there is resource contention. The workload balancing can be manual, semi-supervised, and fully automated.

Once the host aggregates are defined and configured, policies for workload placement should be defined for the workload instances.

Flavor Specification for Instances

Flavors are templates with a predefined or custom resource specification that are used to instantiate workloads. A Flavor can be configured with additional Extra Specs metadata parameters for workload placement. The following parameters can be employed for workload acceleration by using the N-VDS Enhanced switch:

- Huge Pages. To provide important or required performance improvements for some workloads, OpenStack huge pages can be enabled for up to 1GB per page. The memory page size can be set for a huge page.
- CPU Pinning. VMware Integrated OpenStack supports virtual CPU pinning. When running latency-sensitive applications inside a VM, virtual CPU pinning could be used to eliminate the extra latency that is imposed by the virtualization. The flavor extra spec `hw:cpu_policy` can be set for CPU pinning.
- NUMA Affinity. ESXi has built-in smart NUMA scheduler that can load balance VMs and align resources from the same NUMA. This can be achieved by setting the latency sensitivity level to **`high`** or setting the flavor extra spec `hw:cpu_policy` to **`dedicated`**.

Host Affinity and Anti-Affinity Policy

The Nova scheduler provides filters that can be used to ensure that VMware Integrated OpenStack instances are automatically placed on the same host (affinity) or separate hosts (anti-affinity).

Affinity or anti-affinity filters can be applied as a policy to a server group. All instances that are members of the same group are a subject to the same filters. When an OpenStack instance in created, the server group to which the instance will belong is specified and therefore a filter is applied. These server groups policies are automatically realized as DRS VM-VM placement rules. DRS ensures that the affinity and anti-affinity policies are maintained both at initial placement and during runtime load balancing.

DRS Host Groups for Placement

In VMware Integrated OpenStack, Nova compute nodes map to vCenter Server clusters. The cloud administrator can use vSphere DRS settings to control how specific OpenStack instances are placed on hosts in the Compute cluster. In addition to the DRS configuration, the metadata of source images in OpenStack can be modified to ensure that instances generated from those images are correctly identified for placement.

vSphere Web Client provides options to create VM and host groups to contain and manage specific OpenStack instances and to create a DRS rule to manage the distribution of OpenStack instances in a VM group to a specific host group.

VMware Integrated OpenStack allows to modify the metadata of a source image to automatically place instances into VM groups. VM groups are configured in the vSphere Web Client and can be further used to apply DRS rules.

Availability and Disaster Recovery

Business continuity is supported by the vCloud NFV OpenStack Edition platform across its management, control, and data plane components. This section discusses the available designs and recovery considerations.

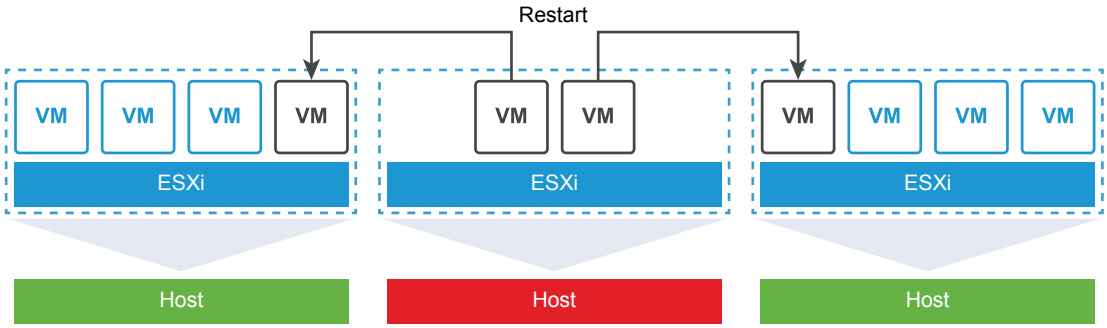
Availability

All vCloud NFV OpenStack Edition platform components implement a high availability design by default. In addition, VNF's can take advantage of platform capabilities to extend their availability in compute, storage, and networking.

vSphere High Availability

Redundancy with vSphere uses VM level replicas in conjunction with the VNF high availability architecture. vSphere HA can instantiate a new VM in the event of host failure thus providing redundancy for a VNF and VNFC pair. vSphere HA can be fully automated without the need for manual intervention for failure and recovery.

Figure 8-19. vSphere High Availability



VMware NSX-T Data Center Availability

NSX-T Data Center by default provides high availability to VNFCs in the overlay network. NIC Teaming and protocols such as Equal-Cost Multipath (ECMP), Graceful Restart, and Link Aggregation Group (LAG), provide redundant connectivity. The NSX-T architecture also separates management, control, and data plane traffic to further optimize service availability.

VMware vSAN

vSAN is fully integrated in vSphere and provides policy-driven fault handling with platform awareness such as chassis and rack affinity to store object replicas. The virtual storage is integrated with vRealize Operations so that in case of failure, the failed VM can be cloned and spun up automatically. Storage vMotion can be used to perform live migration of virtual machine disk files (VMDK) within and across storage arrays by maintaining continuous service availability and complete transaction integrity at the same time.

Disaster Recovery

In the event of a failure, the site is recovered by executing an automated recovery plan. Data replication across protected and failover zones is necessary to recover the state of the site.

vSphere Replication

vSphere Replication replicates virtual machine data between data center objects within a single site or across sites. vSphere Replication fully supports vSAN. It is deployed as a virtual appliance in the Management Pod to provide a Recovery Point Objective (RPO) of five minutes to 24 hours.

When executing a disaster recovery plan, RPO and Recovery Time Objective (RTO) are the most important aspects that must be considered. RPO is the duration of the acceptable data loss and it is fulfilled by the replication technology. RTO is a target duration with an attached SLA, during which the business process must be restored. It includes the time for the recovery and service readiness, in a state of normal business operation.

vSphere Replication provides the ability to set the RPO, however RTO is application-dependent.

Site Recovery Manager

Site Recovery Manager provides a solution for automating the recovery and execution of a disaster recovery plan in the event of a disaster in a data center. When a catastrophe occurs, components in the Management Pod must be available to recover and continue the healthy operations of the NFV-based services.

To ensure robust business continuity and disaster recovery, network connectivity between the protected and recovery sites is required, with enough bandwidth capacity to replicate the management components by using vSphere Replication. Each site must have an instance of vCenter Server that governs the Management Pod and its ESXi hosts, and a Site Recovery Manager server and vSphere Replication appliance to orchestrate the disaster recovery workflows and replicate content across the sites. The protected site provides business critical services, while the recovery site is an alternative infrastructure on which services are recovered in the event of a disaster.

Inventory Mappings

Elements in the vCenter Server inventory list can be mapped from the protected site to their vCenter Server inventory counterparts on the recovery site. Such elements include VM folders, clusters or resource pools, and networks. All items within a single data center on the protected site must map to a single data center on the recovery site.

These inventory mapping details are used across both the protected and recovery sites:

- Resource mapping maps cluster objects on the protected site to cluster objects on the recovery site.
- Folder mapping maps the folder structures like data centers or VM folders on the protected site to folder structures on the recovery site.
- Network mapping maps the management networks on the protected site to management networks on the recovery site.

Protection Groups

A protection group is a group of management components at the protected site that can failover together to the recovery site during testing and recovery. All protected management components are placed within a single protection group.

Recovery Plans

Recovery plans are the run books that are associated with a disaster recovery scenario. A recovery plan determines which management components are started, what needs to be powered down, which scripts to run, the startup order, and the overall automated execution of the failover.

A complete site failure is the only scenario that invokes a disaster recovery. There is no requirement for recovery plans to handle planned migrations or to move a single failed application within the management cluster. A single recovery plan is created for the automated failover of the primary site, and the placement of management components into priority groups ensures the correct startup order.

VNF Recovery Considerations

Every VNF vendor must provide a specific strategy for disaster recovery for any VNF managed directly by the VNF Managers.

NSX Data Center for vSphere Coexistence with NSX-T Data Center

As NSX evolves from NSX Data Center for vSphere to its successor NSX-T Data Center, CSPs can deploy the vCloud NFV Openstack Edition platform with two software-defined networking stacks simultaneously. CSPs can deploy NSX Data Center for vSphere in conjunction with NSX-T Data Center as part of vCloud NFV OpeStack Edition 3.1.

NSX Data Center for vSphere Interoperating with NSX-T Data Center in a Greenfield Deployment

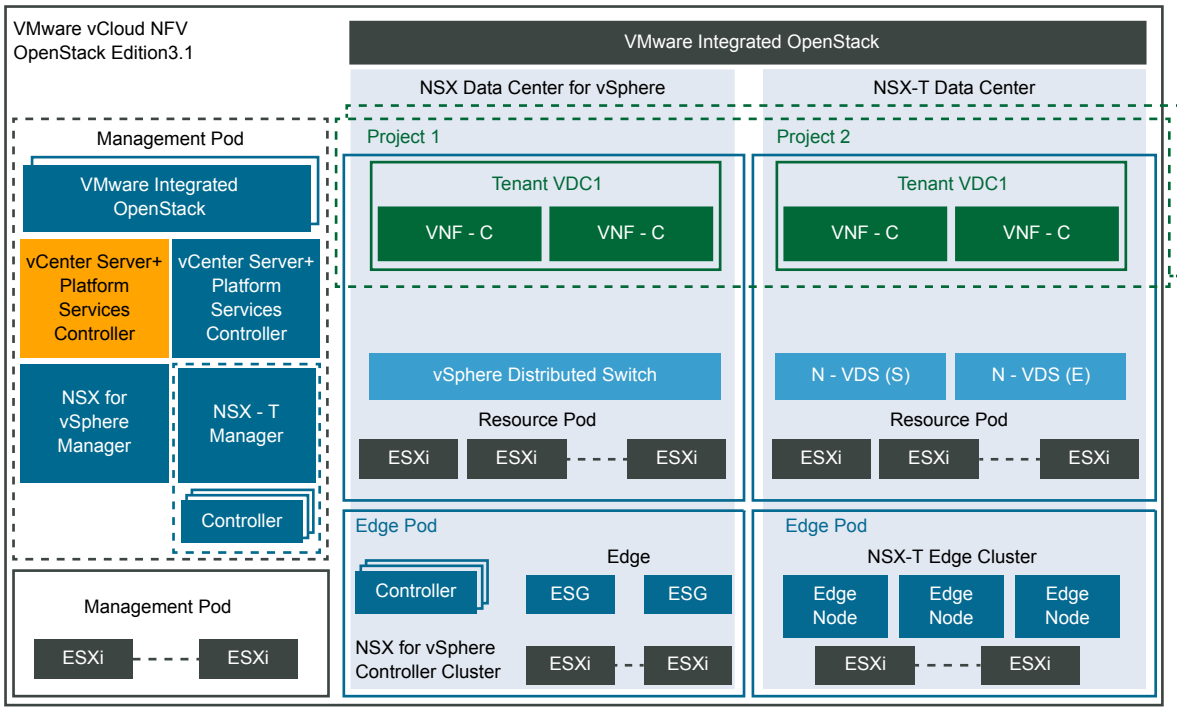
This deployment option allows CSPs to deploy the vCloud NFV OpenStack Edition 3.1 platform by using both NSX Data Center for vSphere and NSX-T Data Center networking stacks. The CSP can create a dedicated Resource Pod for each networking stack and map it to a VMware Integrated Openstack Project by using the TVD plug-in configurations in the Horizon Dashboard.

Table 8-3. Greenfield vCloud NFV OpenStack Edition 3.1 Deployment

Building Block	Design Objective
Management Pod	■ Single vCenter Server instance for both NSX Data Center for vSphere and NSX-T Data Center stacks. ■ Separate management and control planes for NSX Data Center for vSphere and NSX-T Data Center. ■ A VMware Integrated OpenStack instance for both NSX Data Center for vSphere and NSX-T Data Center, network and workload management.
Resource Pod	■ Dedicated NSX Data Center for vSphere and NSX-T Data Center resource pods. ■ Common vSphere versions across both stacks.
Edge Pod	■ Dedicated NSX Data Center for vSphere and NSX-T Data Center Edge pods.

Note To deploy platform instance based on NSX Data Center for vSphere, use the vCloud NFV Openstack Edition 2.0 Reference Architecture design guidelines and principles.

Figure 8-20. NSX-T Data Center and NSX Data Center for vSphere Coexistence in Greenfield Deployments



NFVI components such as vCenter Server treat both stacks as separate entities in terms of host clusters, resource pools, and virtual switches. At the top level VIM component, VMware Integrated Openstack allows Projects to consume the resources of each stack by configuring Tenant VDCs backed by the corresponding Resource Pod. Each stack has a dedicated Resource and Edge Pods for compute and North-South connectivity respectively.

Analytics-Enabled Reference Architecture

9

CSPs can enable the vCloud NFV OpenStack Edition platform for day 1 and day 2 operations after the platform is deployed in the cloud provider topology. The platform is integrated with an operations management suite that provides capabilities for health monitoring, issue isolation, security, and remediation of the NFVI and VNFs.

The NFVI operations management framework defines and packages a five-step approach to make day 1 and day 2 workflows operational.

- Onboard service operations.
- Service launch and monitoring.
- Dynamic optimizations.
- Issue isolation.
- Demand planning and expansion.

The integrated operational intelligence will adapt to the dynamic characteristics of the NFV infrastructure to ensure service quality and issue resolution. Some of the key characteristics include:

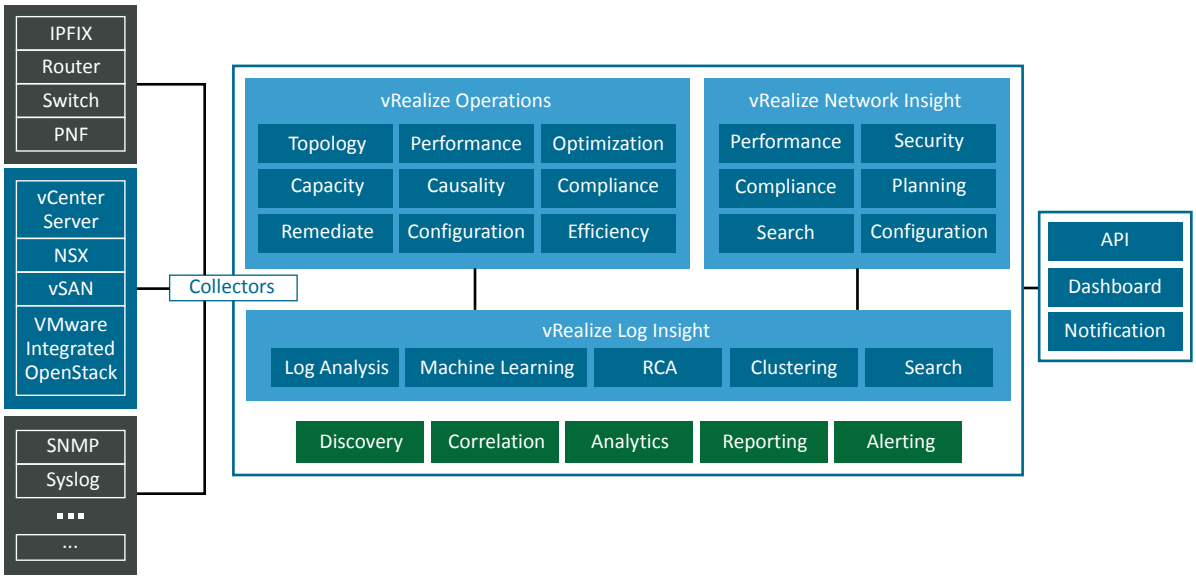
- Dynamic resource discovery. Distributed and complex topologies together with workloads in motion require dynamic resource and service discovery. The platform provides continuous visibility over service provisioning, workload migrations, auto-scaling, elastic networking, and network-sliced multitenancy that spans across VNFs, hosts, clusters, and sites.
- SLA management. Continuous operational intelligence and alert notifications enable proactive service optimizations, capacity scale-out or scale-in, SLA violations, configuration and compliance gaps, and security vulnerabilities.
- Remediation. Reduced MTTU and timely issue isolation for improved service reliability and availability. Prioritized alerting, recommendations, and advanced log searching enable isolation of service issues across physical and overlay networks.
- Security and policy controls. Multivendor services operating in a shared resource pool can create security risks within the virtual environment.
    - Ability to profile and monitor traffic segments, types, and destination to recommend security rules and policies for north-south and east-west traffic.

- Identification of security policy and configuration violations, performance impacts, and traffic routes.

- Capacity planning and forecasting. New business models and flexible networks demand efficient capacity planning and forecasting abilities in contrast to the traditional approach of over-provisioning that is costly and unrealistic.

The framework continuously collects data from local and distributed agents, correlating, analyzing and enabling day 2 operations. The analytical intelligence can be also queried and triggered by third-party components such as existing assurance engines, NMS, EMS, OSS/BSS, and VNF-M and NFV-O for closed loop remediation.

Figure 9-1. Analytics-Enabled Reference Architecture



CSPs can deploy the operations management components in the Management Pod and centralize them across the cloud topology, assuming that inter-site latency constraints are met.

- vRealize Operations Manager collects compute, storage, and networking data providing performance and fault visibility over hosts, hypervisors, virtual machines, clusters, and site.

- vRealize Log Insight captures unstructured data from the environment, providing log analysis and analytics for issue isolation. Platform component logs and events are ingested and tokenized, and mined for intelligence so that they can be searched, filtered, aggregated, and alerted.

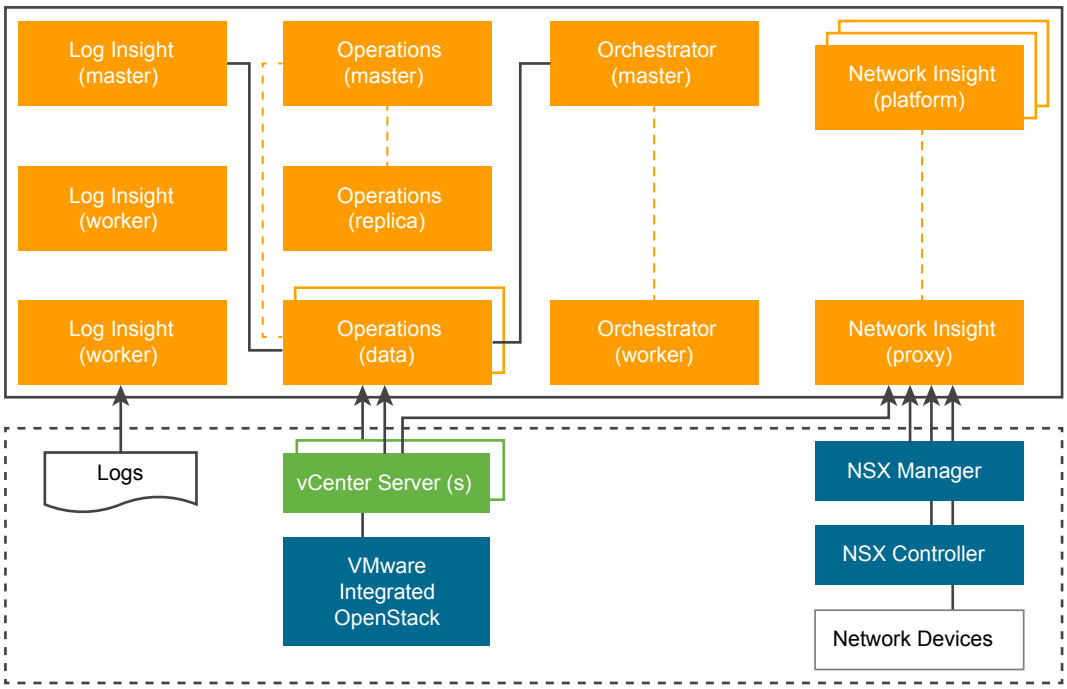
- vRealize Network Insight provides layer 2, 3, and 4 visibility into the virtual and physical networks and security policy gaps. The engine is integrated with the NFVI networking fabric, ingesting data that ranges in performance metrics, device and network configuration, IPFIX flow, and SNMP. It discovers gaps in the network traffic optimization, micro-segmentation, compliance, security violations, traffic routing, and performance.

This chapter includes the following topics:

Management Pod Extensions

The analytics-enhanced reference architecture enriches the Management Pod with the vRealize management components to provide infrastructure assurance capability.

Figure 9-2. Analytics Extension to Management Pod



Components

The operations management components are deployed as a centralized function that is capable of day 1 and day 2 operations spanning the CSP's deployment topology. The data collection architecture is specific to each operations management component with a centralized single pane for monitoring, reporting, troubleshooting, and closed-loop automation.

vRealize Operations Manager

The virtual environment relies on a monitoring solution that is able to collect data regarding its health, capacity, availability, and performance. vRealize Operations Manager provides a robust and integrated monitoring platform that resides at the center of the NFV environment. It monitors the virtual environment and collects data about its health, capacity, availability, and performance. vRealize Operations Manager serves as the single pane of glass to the NFV environment.

vRealize Operations Manager extends and collects information through management packs. The collected information is filtered for relevancy, analyzed, and presented in customizable dashboards. It exposes an API that retrieves performance and health data about NFVI and the virtual resources of the VNF instance.

The design of the vRealize Operations components is based on centralized management and collection, with optional remote collection for a distributed topology. vRealize Operations Manager supports HA across the various components. HA creates a master replica for the vRealize Operations Manager master node and protects the management functions. In smaller deployments, the master node can also act as a data node. In larger deployments, data nodes host adapters that are responsible for collecting data and can be scaled to meet additional capacity. To enable HA, at least one more data node must be deployed in addition to the master node. Anti-affinity rules should be used to keep nodes on specific hosts.

vRealize Operations Manager is installed in the Management Pod in both Two-Pod and Three-Pod designs. Depending on the number of metrics that are collected over time, additional storage capacity and compute capacity might be required. Adding more hosts to the management cluster or more storage is sufficient to address the growing storage needs of vRealize Operations Manager.

By default, VMware offers Extra Small, Small, Medium, Large, and Extra Large configurations during installation. The CSP can size the environment according to the existing infrastructure to be monitored. After the vRealize Operations Manager instance outgrows the existing size, the CSP must expand the cluster to add nodes of the same size. See the vRealize Operations Manager Sizing Guidelines .

vRealize Log Insight

CSPs can use vRealize Log Insight to collect log data from ESXi hosts and data about server events, tasks, and alarms from vCenter Server systems. vRealize Log Insight integrates with vRealize Operations Manager to send notification events. Because vRealize Log Insight collects real-time unstructured data, the CSP can configure all elements in the NFV environment to send their log data to vRealize Log Insight. This log aggregation provides a single log collector for the entire NFV environment.

vRealize Log Insight ingests syslog data from the physical and virtual NFVI components to deliver monitoring, search, and log analytics. It builds an index for analytics purposes by automatically identifying structure from machine-generated log data including application logs, network traces, configuration files, messages, performance data, and system state dumps. Coupled with a dashboard for stored queries, reports, and alerts, vRealize Log Insight assists the CSP in root cause analysis and reduction in MTTR. All NSX-T Data Center Manager syslog information, distributed firewall logs, and NSX-T Data Center Edge syslog information is sent to vRealize Log Insight.

The vRealize Log Insight API provides programmatic access to vRealize Log Insight functionality and to its datastore. As a result, the OSS/BSS systems or MANO components can integrate with vRealize Log Insight to gain further insight into the system events and logs.

vRealize Log Insight is deployed by using a single cluster configuration, which consists of a minimum of three nodes leveraging the Log Insight Integrated Load Balancer (ILB). A single log message is only present in one location within the cluster at a time. The cluster remains up and available to ingest data and serve queries during a temporary unavailability of a single node.

vRealize Log Insight provides preset VM sizes that the CSP can select from to meet the ingestion requirements of their environment, Extra Small, Small, Medium, Large, and Extra Large configurations. These presets are certified size combinations of compute and disk resources, though extra resources can be added afterward. See the VMware Documentation for sizing details.

vRealize Network Insight

vRealize Network Insight provides operations for software-defined networking and security across virtual and physical infrastructure with micro-segmentation planning that can be scaled to thousands of VNFs. vRealize Network Insight is installed in the Management Pod in both Two-Pod and Three-Pod designs.

The vRealize Network Insight architecture consists of a platform VM, a proxy VM, and data sources. The platform VM provides analytics, storage, and a user interface to the data. The proxy VM, or the collector, collects data by using various protocols such as HTTPS, SSH, CLI, and SNMP, depending on the source and the configuration. A variety of data sources are supported, including vCenter Server, NSX-T Data Center, firewalls, and various switch vendors.

The platform VM is deployed as a single cluster to provide high availability and scale. A minimum of three platform VMs are required in the cluster. The proxy VMs are used to collect data and can be deployed in a single data center or distributed across sites. Depending on the amount of data that will be collected, typically CSPs need one or more proxy VMs.

Ensure that the system meets the minimum hardware configurations to install vRealize Network Insight. See VMware Documentation for sizing details.

vRealize Orchestrator

VMware vRealize Orchestrator is a development-and process-automation platform that provides a library of extensible workflows. Orchestrator workflows run on objects that are exposed through plug-ins and custom scripting to interact with any component that is reachable through an API. By default, a VMware vCenter plug-in is provided to orchestrate tasks in the cloud infrastructure environment. The CSP can use Orchestrator workflows to define and run automated configurable processes thus creating a framework for closed-loop automation.

vRealize Orchestrator is integrated with vRealize Operations Manager through a management pack that provides workflows for automating the cloud infrastructure environment and for orchestration of third-party components as well as management functions.

The vRealize Orchestrator is highly available and configured as a single cluster of multiple virtual appliance instances. The appliance is registered with a vCenter Single Sign-On service by using the vSphere Authentication mode. The appliance also requires a shared database instance.

Ensure that the system meets the minimum hardware configurations to install vRealize Orchestrator. See the VMware Documentation for sizing details.

Enabling Analytics with vRealize

The vRealize components of the vCloud NFV OpenStack Edition platform are integrated in the platform and configured to report and trigger analytical intelligence. The various configurations and extensions that follow will help the CSP to get started with the analytics-enabled reference architecture.

vRealize Operations

vRealize Operations is configured with the adapters that are necessary to start collecting data from the infrastructure. The following solutions should be considered:

- vCenter Adapter. vSphere connects vRealize Operations Manager to one or more Resource and Management vCenter Server instances. The system collects data and metrics from those instances, monitors them, and runs actions in them.
- End-Point Operations Management. The End-Point Operations Management solution is used to gather operating system metrics and to monitor availability of remote platforms and applications.
- vSAN Adapter. In a production environment, operations management for vSAN can be provided by using dashboards to evaluate, manage, and optimize the performance of vSAN and vSAN-enabled objects in the vCenter Server system.
- OpenStack and NSX Adapter – The vRealize® Operations Management Pack™ for OpenStack allows the CSP to quickly view the health of the environment, including services running within the OpenStack infrastructure. The OpenStack management pack integrates with NSX-T Data Center, which allows for easy monitoring and management of the network infrastructure. vRealize Operations Management Pack for OpenStack includes dashboards to provide visibility into OpenStack deployments, such as OpenStack Management Services, OpenStack Compute Infrastructure, OpenStack Network Infrastructure, OpenStack Tenants, and OpenStack Storage.
- vRealize Orchestrator Solution. This management pack enables an adapter to communicate with the vRealize Orchestrator workflow automation engine.

vRealize Network Insight

vRealize Network Insight can be configured to monitor all networking-related components in the NFVI. vRealize Network Insight can connect to the VMware Integrated OpenStack NFV components that are related to networking and security and provide insights into the networking segments that are used to deploy and manage the vCloud NFV OpenStack Edition platform. The management VLANs, external VLANs, and Overlay segments are all available for monitoring and diagnostics.

vRealize Network Insight uses the following data sources:

- vCenter Server. Both Resource and Management vCenter Server instances.

- NSX Manager. The management plane of the NSX-T Data Center networking.
- IPFIX. For flow analysis, which can be enabled within vCenter Server.
- Network devices. Physical devices such as Dell switches, Cisco Nexus and Catalyst switches, Arista, Juniper Networks, Hewlett Packard Enterprise, Brocade, and Palo Alto Networks switches.

vRealize Orchestrator

vRealize Orchestrator is integrated into vRealize Operations Manager as a management pack to provide bidirectional communication with the Orchestrator workflow engine. The management pack is deployed in vRealize Operations Manager and configured to point and authenticate to the vRealize Orchestrator instance.

VMware vRealize Operations Management Pack

vRealize Operations Manager can be extended through management packs to monitor not just VNF virtual machines but even Kubernetes clusters giving visibility into the containers, pods, namespaces, replication sets, and so on, that provide the infrastructure for a tenant's containerized network functions.

VMware vRealize Operations Manager collects structure data from various vCloud NFV OpenStack Edition components, including gathering data from adapters that are connected to external components. For this mechanism to work, vRealize Operations Manager is configured to communicate with data sources by using an authorized user account for the respective components. If the user account has limited access to objects in the source server, it sees only the data for which the account has permissions. At a minimum, the user account must have read privileges across the objects from which it will collect data. A collection of management packs is available on VMware Solution Exchange.

To minimize the traffic between vCenter Server and the vRealize Operations Manager, the vCenter Server Adapter is installed with a five-minute collection interval.

Out-of-the-box, vRealize Operations Manager does not monitor VNF service availability or VNF internal key KPIs. The VNF Manager derives this information through direct interaction with the respective VNFs. To extend the operations management functionality to the VNFs, VNF vendors can create custom management packs for vRealize Operations Manager. Management pack development requires an understanding of the vRealize Operations Manager inventory model and the management functions that management packs implement. These include auto-discovery and monitoring. More information is available at Endpoint Operations Management Agent Plugin Development Kit .

VMware vRealize Log Insight Content Pack

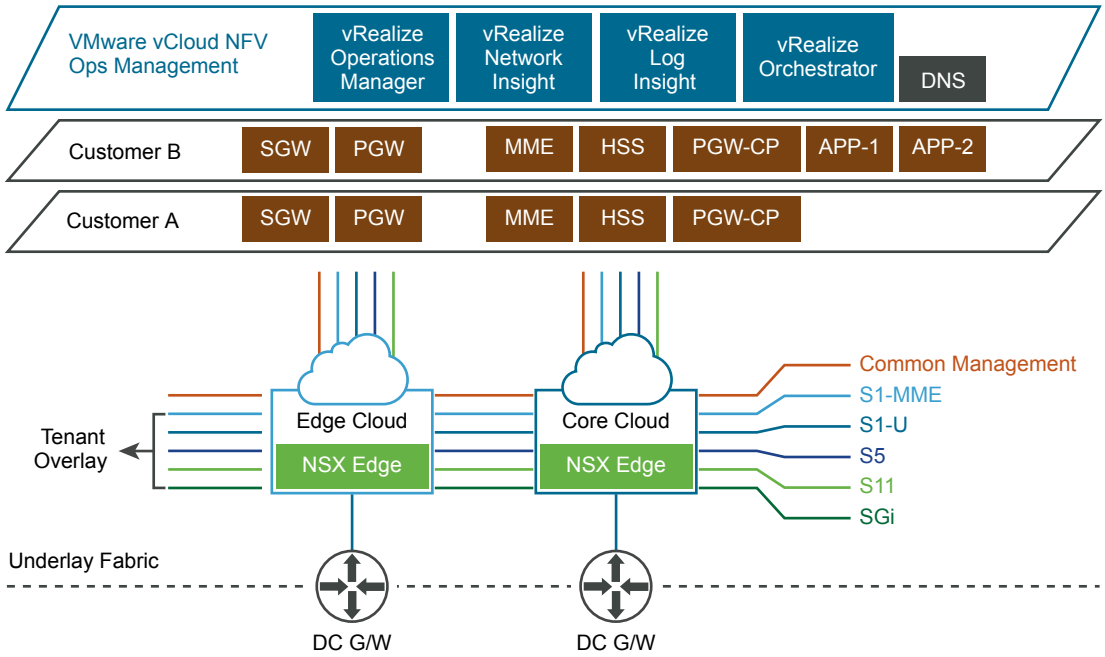
vRealize Log Insight gathers log events natively from multiple syslog data sources and through special content packs. Specific dashboards can be customized to perform log analytics and alerts. For additional information about vRealize Log Insight solutions, see VMware Solution Exchange.

Monitoring the Infrastructure

The vCloud NFV OpenStack Edition operations management components are pre-configured with default dashboards, alerts, actions, and profiles. To make day 2 operations effective, additional steps for onboarding service operations are required so that to enhance the monitoring and remediation of the cloud infrastructure environment.

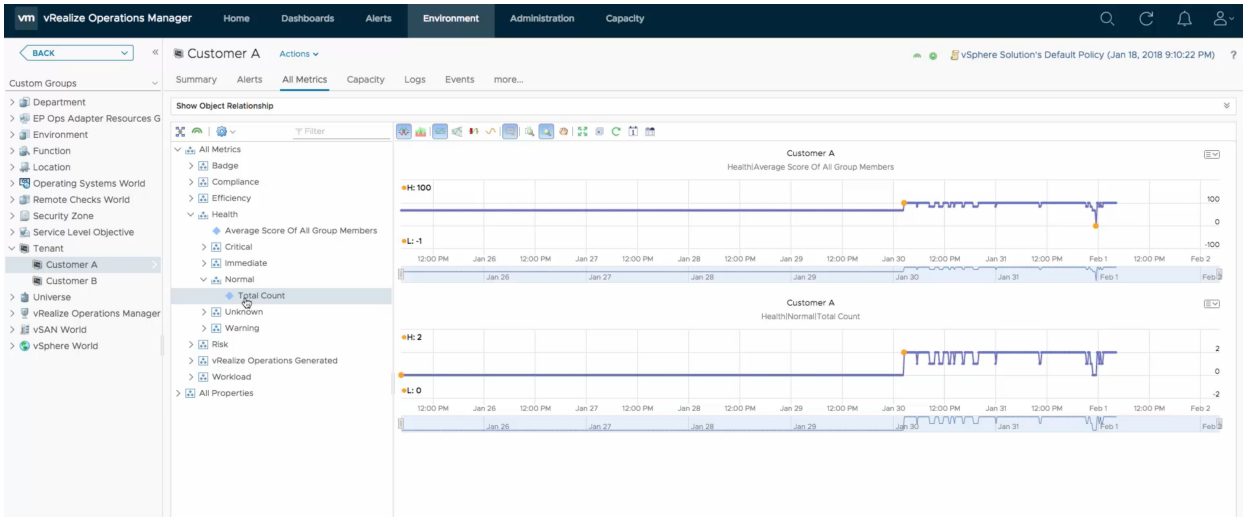
Consider the following topology as an illustration of a multi-tenant service offer of the mobile packet core:

Figure 9-3. Infrastructure Monitoring Reference Topology

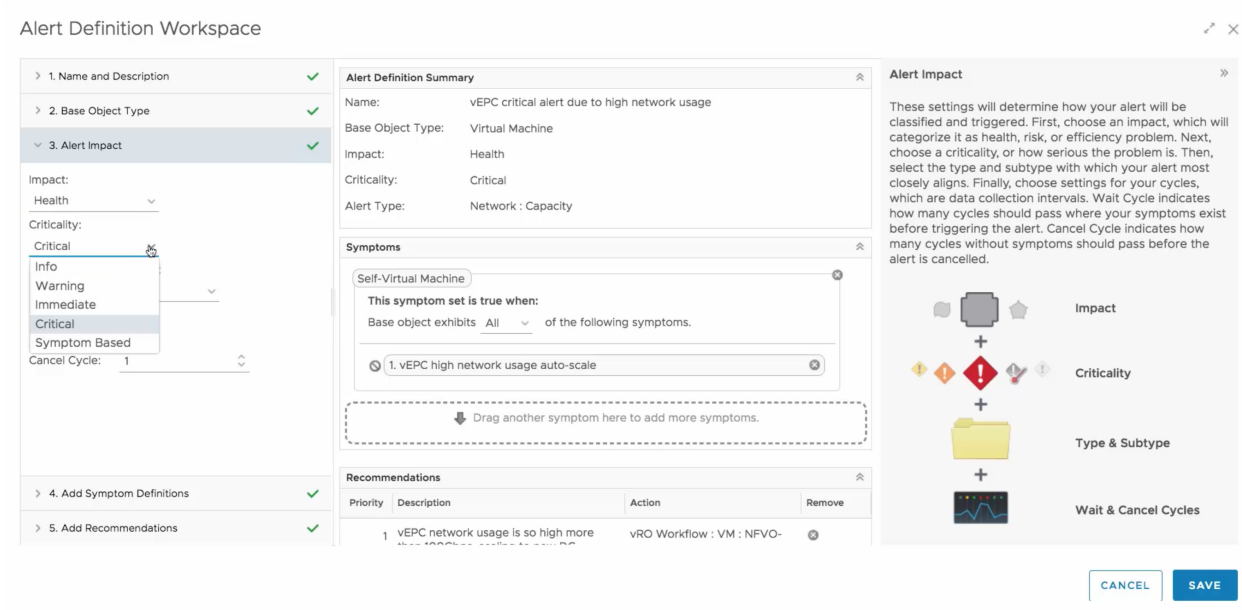


The example topology places the data plane components of the packet core in the edge cloud while control plane functions in the core cloud, including the operations management components.

- Access controls (RBAC).The CSP administrator should start by creating the user profiles, roles, and permissions within vRealize Operations Manager for their tenants. This way, tenant users are able to view dashboards and performance or fault information that the CSP administrator selects to expose.
- Service definition. Once the VNFs are onboarded for the tenant, the CSP administrator can create service definitions within vRealize Operations Manager. In the example, an evolve packet core service definition is created by using a Service Group construct. If the VMs for the various VNFs are placed in vCenter Server folders, the CSP administrator can map them to a service group or select individual VMs for that tenant. The underlying compute, networking, and storage infrastructure components are auto-discovered for each VM. This includes parent containers such as clusters, datastores, and data center. Once the CSP administrator creates the service definitions, they can monitor the various performance counters, health, risk, efficiency, anomalies, and faults for the delivered services. Similarly, the CSP administrator can create service definitions in vRealize Network Insight to create the networking and security views of the tenant services.



- Dashboards. All components of the operations management solution provide default and customizable dashboards. By using the library of widgets, the CSP administrator can create views for time-series, object-relationships, topologies, alerting, heat-maps, capacity, and so on.
- Default Alerts. A key component for effective monitoring and remediation or the virtual environment is the rich customizable alerting framework that is available in the platform. Along with the numerous out-of-the-box alerts in the platform, the CSP administrator can also create service-centric alerts by using complex rules and machine learning across multiple VNFs. The framework allows the CSP administrator to create alerts on target object types and specify the impact depending on the risk, efficiency, and health. They can also create complex symptom definitions and a recommendation. The recommendation can be an informational message on how to remediated the alert, which can be tied to a semi-supervised or automated action the system can take. Timely action is key to ensure that QoS and SLA guarantees are met.



Action workflows. Action workflows trigger remediation for issue avoidance. Workflows can be internal to the NFVI layer or trigger an external element such a VNF-M or NFV-O. A transient spike in the packet core utilization might trigger a rebalance of the Packet Gateway (PGW) to mitigate congestion on the host. A workflow trigger to the DRS can be leveraged in such a scenario. For external triggering, vRealize Operations Manager is integrated with vRealize Orchestrator. Custom workflows and scripts can be defined in vRealize Orchestrator. They appear as actions in vRealize Operations Manager that can then be mapped in the alerting framework.

Cloud-Native Ready

vRealize Operations Manager can be further enhanced with a management pack for Kubernetes clusters that are deployed together in the multi-tenant NFVI environment. Once installed, the management pack for Container Monitoring can be configured by pointing to the Master Node URL of the Kubernetes container and selecting the cAdvisor service. vRealize Operations Manager collects metrics for nodes, pods, and containers and applies the same class of assurance for discovery, monitoring, and analytics. For more information, see vRealize Operations Management Pack for Container Monitoring.

For more information on configuring and customizing the components, refer to the official documentation for vRealize Operations Manager, vRealize Log Insight, vRealize Network Insight, and vRealize Orchestrator.

Predictive Optimization

In vCloud NFV OpenStack Edition, resource contention is evaluated based on continuous monitoring. vRealize Operations Manager uses dynamic thresholds to monitor and analyze the behavior of VNFs. Statistical confidence bands are created for each metric and its time-series, indicating an upper and lower boundary for it. The boundaries provide visibility on resource consumption of various VNFs operating in the NFVI environment. The VNF intelligence is utilized by DRS to predict future contention and capacity exhaust and rebalance the workloads in a timely manner.

In addition to the dynamic utilization and forecasting intelligence data, policies can be defined for different classes of workloads, for example:

- Host affinity/anti-affinity rules, ensuring if tagged VMs are placed together or not on the same host.
- Latency Sensitivity is set to high for data plane intensive workloads, pinning CPU.
- NUMA affinity to ensure that the VM is assigned to a dedicate node.
- Current and target host CPU profile alignment to avoid mismatch (Enhanced vMotion Capability).

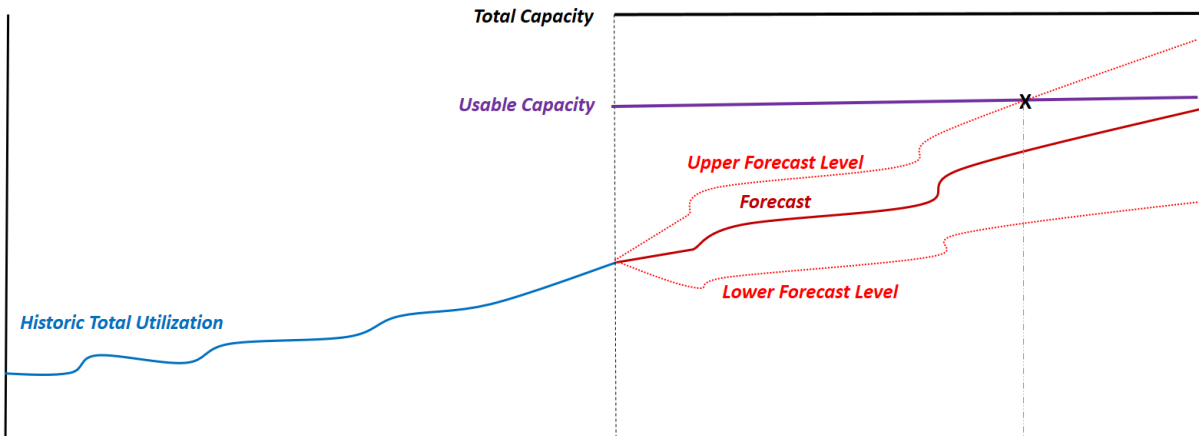
Distributed DRS can be used to balance VNF components based on their performance needs and capacity availability, eliminating resource contention that might occur. Forecasts are generated every five minutes, which allows DRS to predict and act on any possible imbalances. The optimization can be enabled in a supervised or non-supervised control.

Capacity and Demand Planning

To avoid the risk of capacity exhaust and unplanned infrastructure costs, capacity and demand planning are a critical part of the data center operations. vRealize Operations Manager provides a capacity management solution that continuously monitors and advises on the health of the NFVI environment.

Right-Sizing

Over-provisioning the infrastructure can be avoided by right-sizing the resources. Capacity risk tolerance can be set for a trial environment that is different from the production environment. You can plan for infrastructure procurement or migrations based on available capacity.

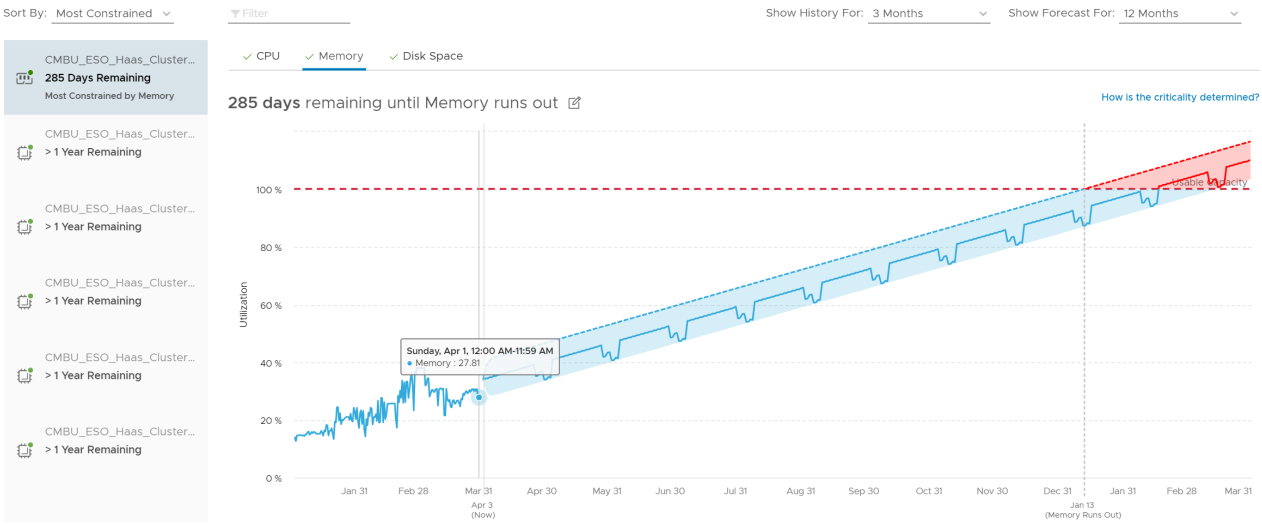


As the CSP considers the importance of their environment and workloads, they can set the capacity risk level to **`Conservative`** by using the upper bound projection of the forecast for high-performance workloads. In contrast, they can set the capacity risk to **`Aggressive`** policy by using the forecast for non-critical workload environment.

Head-room capacity can also be defined on the cluster to ensure that there is space for unexpected spikes in traffic and performance for critical workloads.

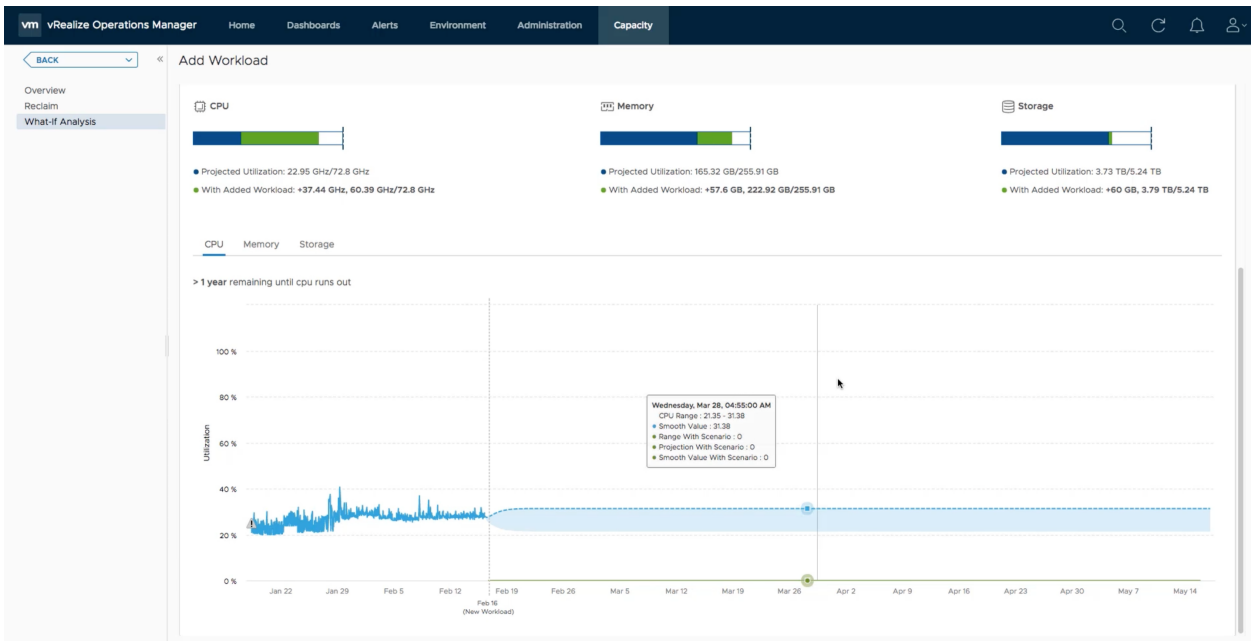
Forecasting

Demand planning might be challenging, especially if the CapEx spend should be balanced over time. The forecasting engine can be used to extrapolate existing utilization of workload CPU, memory , and storage across VMs, clusters, and data centers. The just-in-time capacity forecasts can be used to plan for new infrastructure or rebalance with predictable intelligence.



New Service Order

As an operations manager, when a new tenant order needs to be fulfilled you need to make decisions on where to place the new workloads without impacting current operations. Various approaches can be used in the data center clusters including making room for the new order by moving workloads, adding additional resources to the existing workloads, and adding new hosts to the cluster. You can use vRealize Operations What-If analysis tool to emulate such scenarios by using the existing workloads as the target profile. You can model additional CPU, memory, storage to existing workloads to calculate new capacity exhaust. You can emulate new instance of similar workloads to estimate if there is enough capacity in the cluster.



Cost Management

vRealize Operations Manager provides visibility into cost of operations providing business management visibility on workloads, resources, and data centers that are used to deliver the services. Operations and planning teams can better manage their resource cost and consumption based on network function to resource ratio's and effective ROI in service delivery.

- Reclaim Resources. Resource consumption is continuously monitored for utilization. This way VMs that are idle, underutilized, powered-off and so on, can be reclaimed to free up resource for cost avoidance.
- Densification. Capacity consolidation can be another way to recover cost offsets that are caused by over-provisioning. Certain class of non-performance workloads can be tagged, and policy enabled for higher consolidation, allowing the operator to utilize the resource cluster with a dense capacity risk.
- Optimize licenses. As with densification, VMs that have license restrictions that are tied to a cluster, can be tagged and enabled with policies. This way both the initial placement and on-going DRS balancing will respect the policy to ensure that such VMs are placed on the same cluster.

vRealize Operations Manager additionally provides an integrated cost management function for data centers. The cost of running workloads in a private data center can be compared to public cloud costs. Depending on the characteristics of the workload, you can consider the following options:

- Migrate non-mission critical workloads to a public cloud.
- Burst capacity into public clouds to meet transient spikes.
- Use public clouds for upgrades and other maintenance activities.

Closed Loop Automation for Higher Availability

Automation is key to the cloud infrastructure platform, ensuring that it is dynamic and adaptive to the workloads and consumption models. During operation, workloads experience an increase in traffic demand, congestion across resources and virtualization layers, therefore a proactive optimization approach is very effective.

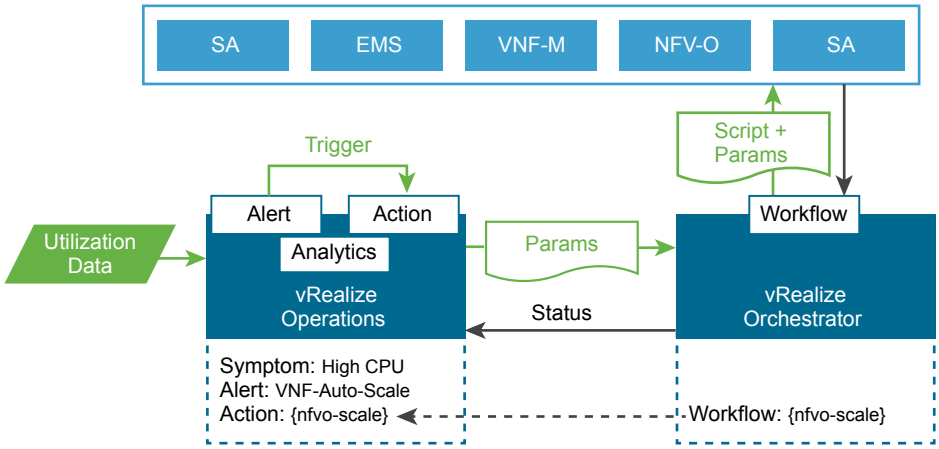
The vCloud NFV OpenStack Edition analytics framework provides key capabilities to develop closed-loop automation. This includes the ability to define intent-based operational policies and alerts as well as actions to enforce the alerts. You can use the framework to manage performance, resources and workload life cycle running within the cloud infrastructure.

In the next generation networks, the workloads will undergo changes in demand and utilization requiring the optimization of allocated resource pools. Infrastructure resources (CPU, memory, storage, network) might need to be added or reclaimed and their reservations modified. Workloads might need to be migrated to a less loaded host or cluster that has available resources. Workload life cycle management is

equally critical to ensure restart, including network function component boot sequence, scale, and heal operations in accordance to the operational policies that are defined. The closed loop automation for such challenges can be addressed at the NFV infrastructure layer and at the service orchestrator and VNF-M layers.

vRealize Orchestrator together with vRealize Operations Manager provides the framework for closed-loop automation. They separate the policy, alert definition and management from the decision point and enforcement. Actions can be triggered back to the NFV infrastructure or directed to third-party external systems.

Figure 9-4. Closed-Loop Automation and Workflows



When the alert is raised, the vRealize Orchestrator workflow is triggered, notifying the north-bound component recipient which will then carry out the remediation appropriate for that application it is managing.

Authors and Contributors

10

This section provides a brief description about the Authors and Contributors.

The following authors co-wrote this paper:

- Arunkumar Kodagi Ramachandra, Solutions Architect, NFV Solutions Engineering, VMware
- Indranil Bal, Solution Consultant, NFV Solutions, VMware
- Pradip Kadam, Solution Consultant, NFV Solutions, VMware
- Tuan Nguyen, Senior Solutions Engineer, NFV Solutions, VMware
- Sumit Verdi, Director Lighthouse Solutions, NFV Solutions, VMware
- T. Sridhar, Principal Engineer & Chief Architect, NFV Solutions, VMware

Many thanks for contributions from:

- Ramkumar Venketaramani, Director Product Management, NFV Solutions, VMware
- Jambi Ganbar, Sr. Technical Solutions Manager, NFV Solutions Engineering, VMware
- Andrea Li, Lead Solutions Test Architect, NFV Solutions Engineering, VMware
- Michelle Han, Director, Solutions Testing and Validation, NFV Solutions Engineering, VMware
- Danny Lin, Sr. Director Product Management and Architecture, NFV Solutions, VMware
- Dobrinka Boeva, Sr. Staff Technical Writer, Information Experience, VMware

Special thanks for their valuable feedback to:

- Frank Escaros-Buechsel, Mikael Brihed, Henrik Oberg, Mauricio Valdueza, Giridhar Jayavelu, Mark Voelker, Richard Boswell, Neil Moore, Mike Richer, Mustafa Bayramov, Christian Hasner, Gary Day, Mikael Brihed, Gary Kotton, Sudesh Tendulkar