Solutions Reference Architecture
September 2021

Building a Multicloud Analytics Solution with VMware Cloud Foundation

Deploy and manage data-intensive workloads from edge to cloud, taking advantage of high-performance 3rd Generation Intel® Xeon® Scalable processors and software optimized for Intel® architecture

Intel Cloud & Enterprise Solutions Group Authors:
Karol Brejna Cloud Solutions Architect
Marcin Gajzler Cloud Solutions Engineer
Piotr Grabuszynski Cloud Solutions Engineer
Marcin Hoffmann Cloud Solutions Engineer
Ewelina Kamyszek Cloud Solutions Engineer
Marek Malczak Cloud Solutions Engineer
Patryk Wolsza vExpert, Cloud Solutions Architect

VMware Authors:
Enrique Corro Fuentes Data Science Staff Engineer, Office of the CTO
Rick Walsworth VMware Cloud Senior Product Line Marketing Manager

Contributor: Tarun Viswanathan, Principal Engineer, Intel Network Platforms Group

Contents

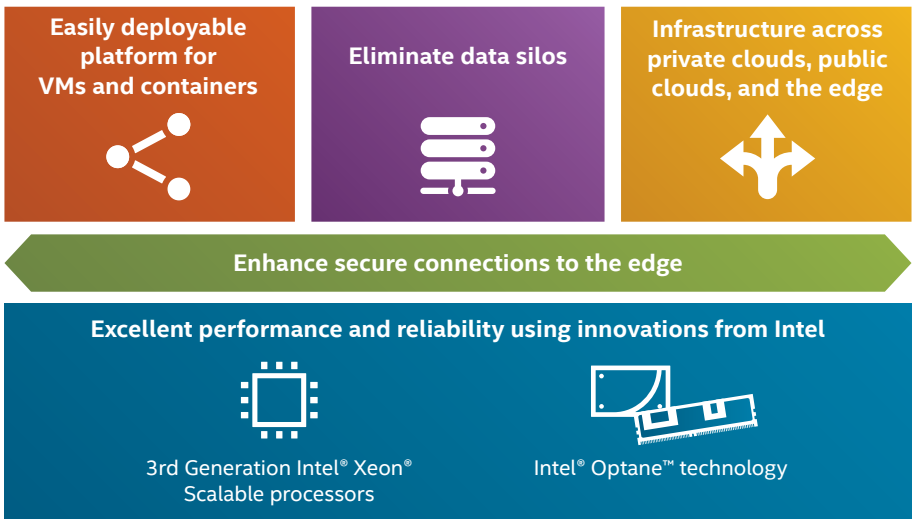
Executive Summary

A modern compute environment is key to remaining competitive. The traditional approach for deploying applications and services cannot deliver innovation at the pace today's businesses require. In addition, as data volumes grow, enterprises struggle to get more value out of their data. Data silos and cumbersome data management and analytics processes hinder discovering business insights that can drive competitive advantage. What's more, as applications move to the edge in industries such as retail, establishing secure connectivity between the core data center, the cloud, and the edge becomes crucial to success.

Addressing these challenges involves replacing legacy hardware and software with modern, multicloud-capable solutions that can accelerate and streamline the entire software and hardware provisioning, deployment, and maintenance lifecycle. Simultaneously, companies need a platform that natively supports containerization for efficient data-intensive workloads like AI and machine learning.

Intel's flexible Multicloud Analytics Solution, based on VMware Cloud Foundation, offers an easily deployable platform for managing virtual machines (VMs) and orchestrating containers. This solution helps eliminate data silos and provides security-enabled infrastructure, operations, and connectivity across private clouds, public clouds, and the edge. The solution delivers excellent performance and reliability using innovations from Intel like 3rd Generation Intel® Xeon® Scalable processors and Intel® Optane™ technology.

Multicloud Analytics Solution



Solution Brief

Business Challenge

Today's enterprises want the flexibility to run workloads where they make most sense—in the core data center, in one or more public clouds (multicloud), and/or at the edge. But to make this flexibility operationally feasible, there must be a way to efficiently manage all the workloads, wherever they reside. Without a single pane of glass, management costs rapidly spiral out of control, application development becomes inconsistent, and performance may suffer. Enterprises seek infrastructure that is characterized by reduced downtime, less setup time, easier maintenance, and lower overhead costs—without sacrificing performance. Legacy data centers cannot take advantage of the cost efficiencies and new technologies available in a multicloud environment. Nor can such data centers adapt to changing workload requirements quickly and nimbly.

For companies with outdated data center technologies, meeting these challenges involves replacing legacy hardware and software with modern, hybrid-cloud-capable solutions. These solutions can accelerate the entire software and hardware provisioning, deployment, and maintenance lifecycle along with application development, testing, and delivery. But, whether it's a machine-learning cluster or a remote branch office cluster, companies may find assembling and maintaining multicloud infrastructure daunting.

Solution Value

Intel and VMware have teamed up to offer the Multicloud Analytics Solution to help take the guesswork out of building multicloud and edge solutions. This solution combines VMware Cloud Foundation with innovative Intel technology to provide a unified Software-Defined Data Center (SDDC) platform for running and managing private cloud, multicloud, and edge containerized workloads.

VMware Cloud Foundation is a full-stack hyperconverged infrastructure (HCI) solution that simplifies the path to and helps accelerate adoption of hybrid/multicloud environments. It offers a complete set of software-defined services for compute, memory, storage, network, and security, along with application-focused cloud management capabilities. When combined with Intel technology, VMware Cloud Foundation provides consistently high performance, reduced data center footprint, and efficient operations management.

Enterprises can use the end-to-end Multicloud Analytics Solution to quickly launch database processing and AI, and scale workloads to accommodate future needs. The unified cloud solution presented in this solution brief can run containerized applications and traditional VMs that are located in an on-premises data center as well as in the public cloud, such as on Amazon Web Services (AWS) and Microsoft Azure.

In short, the Multicloud Analytics Solution is a simple, security-enabled, and agile cloud infrastructure for on-premises, as-a-service public cloud, and edge workloads.

Solution Benefits

- **Unified platform** for running, managing, and seamlessly connecting VMs and containers across private cloud, multicloud, and edge environments
- **Accelerated deployment** with a verified, end-to-end solution for a wide range of workloads
- Agile, scalable, and **security-enabled infrastructure** with excellent performance

Solution Architecture Highlights

The Multicloud Analytics Solution reference architecture from Intel includes several main VMware components: VMware vSphere with Kubernetes, VMware Secure Access Service Edge (SASE) with VMware Software-defined WAN (SD-WAN), VMware Tanzu Mission Control, VMware vSAN, VMware NSX-T, VMware SDDC Manager, and VMware vRealize Suite to provide infrastructure-as-a-service capabilities. It also includes VMware services on public clouds—VMware Cloud on AWS (VMC) and Azure VMware Solution (AVS). Container provisioning and lifecycle management are provided by VMware Tanzu Kubernetes Grid (TKG).

The hybrid/multicloud structure of the solution allows enterprises to extend available resources and easily distribute workloads between on-premises, public cloud, and the edge. VMware SD-WAN is used to provide reliable and secure network connectivity over public internet from any to any location (on-promises to the edge and to public cloud and vice versa).

Underlying the software components of VMware Cloud Foundation in the on-premises core data center are 3rd Generation Intel® Xeon® Scalable processors, Intel® Optane™ persistent memory (PMem), Intel Optane SSDs, Intel® SSD D7 and D5 Series, and Intel® Ethernet products (see Figure 1).

Enterprises can use Intel Optane technology to boost their VMware Cloud Foundation workload performance by placing data closer to the CPU. This technology is a class of non-volatile memory and storage media that fills the gap between high-performing volatile memory and lower-performing NAND storage and HDDs. By placing data closer to the CPU, Intel Optane technology helps architects to confidently deploy an agile, high-performing infrastructure that helps organizations create innovative services and optimize their infrastructure investments.

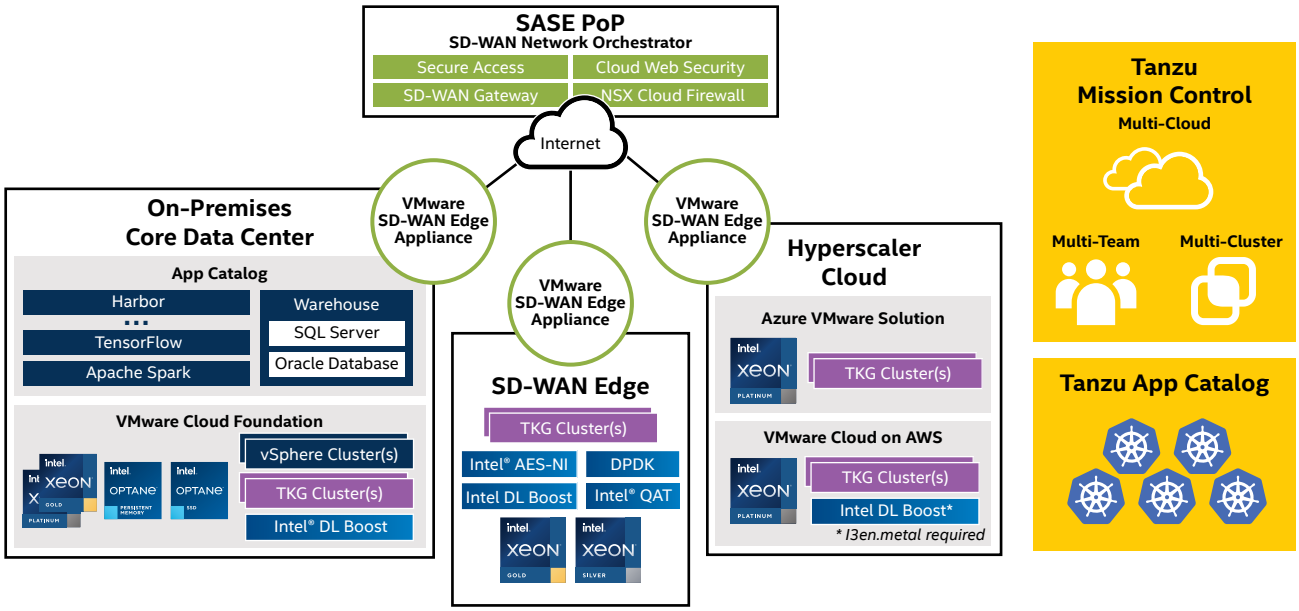


Figure 1. VMware and Intel provide the building blocks for the Multicloud Analytics Solution.

Intel Optane technology can be deployed in two different ways (see Figure 2):

- **Intel Optane PMem** gives enterprises the ability to extract more from larger datasets by combining more capacity and native persistence in a DIMM form factor. Data can be accessed, processed, and analyzed in near real time to deliver deep insights, improve operations, and create new revenue streams.
- **Intel Optane SSDs** help remove data bottlenecks to accelerate transactions and time to insights, so users get what they need, when they need it. With high quality of service and at least 6x faster performance than NAND SSDs at low queue depths, Intel Optane SSDs deliver fast, predictable performance—even in the most demanding environments.[1] For tiered storage like vSAN, it is recommended to use Intel Optane SSDs in the cache tier and Intel SSD D7 or D5 Series in the capacity tier.

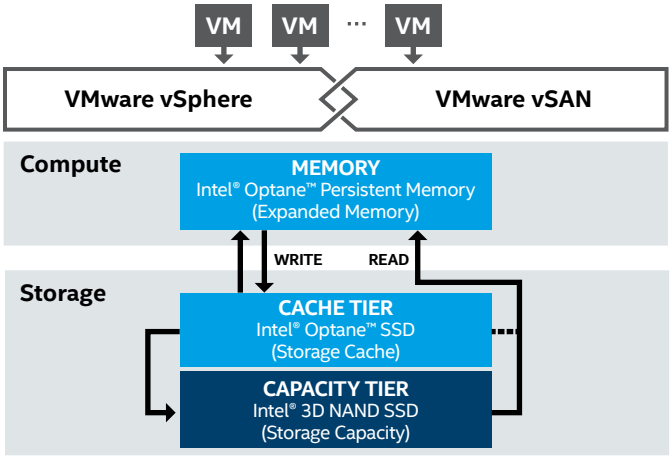


Figure 2. The placement of Intel® Optane™ SSDs and Intel® Optane™ persistent memory within the architecture.

A Closer Look at VMware Cloud Foundation 4.2

VMware Cloud Foundation 4.2 introduces several new features and enhancements that help customers deploy scalable, flexible infrastructure:

- **vSAN Data Persistence platform** enables customers to manage new S3-compatible object storage for unstructured data.
- **vSAN HCI Mesh** delivers a unique software-based approach for disaggregation of compute and storage resources, which enables customers to share capacity across vSAN clusters.
- **NSX-T 3.1 Federation** provides a cloud-like operating model for network administrators by simplifying the consumption of networking and security constructs. With NSX-T Federation, VMware Cloud Foundation customers can deploy stretched networks and unified security policies across multi-region deployments, providing workload mobility and simplified disaster recovery.

For more details about what's new in VMware Cloud Foundation 4.2, visit the release announcement.

Use Cases

The combination of VMware Cloud Foundation and Intel technology running on VMs or in containers can support a wide variety of use cases.

Machine-Learning Inference

Once a model is trained, it can run on new datasets to uncover hidden insights. Inference is compute-intensive and can benefit from innovations such as Intel® Deep Learning Boost (Intel® DL Boost) with Vector Neural Network Instructions (VNNI)—available starting with vSphere 7 and ESXi 7.0, which are foundational components of the VMware Cloud Foundation 4.2 platform.

Data Warehousing and Analytics

Data warehouses are considered one of the core components of business intelligence. They are a central location to store data from one or more disparate sources as well as current and historical data. The VMware hybrid/multicloud platform supports data warehousing, including industry-proven solutions based on Microsoft SQL Server 2019 or Oracle Database 19c.

Edge Computing

For retail stores, healthcare, and smart industry, running workloads closer to customers and closer to the sources of the data can improve performance, which can lead to increased customer satisfaction. VMware Cloud Foundation makes it easy to deploy and manage remote workloads, using the same technology that is used for public and private cloud workloads.

Learn More

- 3rd Gen Intel® Xeon® Scalable processors
- Intel® Ethernet products
- Intel® Optane™ persistent memory
- Intel® Optane™ SSDs
- VMware Cloud Foundation

Find the solution that is right for your organization. Contact your Intel representative or visit the **Intel and VMware Partnership website**.

Implementation Guide

Introduction

The previous pages provided a high-level discussion of the business value for Intel's Multicloud Analytics Solution, and the technologies used in the solution. In this section, more detail is provided about those technologies and how the benchmarks were conducted.

The Multicloud Analytics Solution is available in a Base, Plus, and Edge design; cloud-based deployments can run in either AWS or Microsoft Azure. The design choice depends on workload performance requirements.

Key Technologies

Several innovations from Intel contribute to high performance.

3rd Gen Intel Xeon Scalable Processors

Intel's latest processors for data center workloads are 3rd Gen Intel Xeon Scalable processors. They are packed with performance- and security-enhancing features, including:

- Enhanced per-core performance—up to 40 cores in a standard socket
- Enhanced memory performance, with support for up to 3200 MT/s DIMMs (2 DIMMs per channel)
- Database compression with Intel® Vector Byte Manipulation Instructions (Intel® VBMI)
- Increased memory capacity with up to eight channels
- Support for Intel Optane PMem 200 series
- Built-in AI acceleration with enhanced performance of Intel DL Boost
- Faster inter-node connections with three Intel® Ultra Path Interconnect (Intel® UPI) links at 11.2 GT/s
- More and faster I/O with PCI Express 4 and up to 64 lanes (per socket) at 16 GT/s
- Hardware-enhanced security of Intel® Crypto Acceleration

Intel SSD Data Center Family

- Intel® Optane™ SSD P5800X with next-generation Intel Optane storage media and advanced controller delivers "no-compromises" I/O performance—read or write. It also has high endurance, providing unprecedented value over legacy storage in the accelerating world of intelligent data. Intel Optane SSD P5800X delivers 4x greater random 4K mixed read/write IOPS and 67 percent higher endurance, compared to the previous-generation Intel Optane SSD P4800X, which uses PCIe gen 3.[2]
- Intel® SSD D7-P5500 and Intel® SSD D7-P5600 deliver predictably fast high performance at high capacities. Compared to previous-generation Intel SSDs, the D7-P5500 delivers up to 2x sequential performance and the performance-optimized D7-P5600 brings up to 44 percent higher random mixed-workload performance.[3]
- Intel® Optane™ SSD P1600X provides the right-sized capacity; high throughput and low latency; and excellent endurance that boot drives need. It offers greater than 4x better read/write performance compared to a SATA drive, and has an endurance rating of six drive writes per day and a reliability rating of 2 million mean time hours between failures.

Intel Optane PMem

Intel Optane PMem introduces a new memory class tier: persistent memory that aims to reduce latencies and optimize workloads that are memory-, capacity-, and cost-constrained. Intel Optane PMem 200 series is the second generation of Intel Optane PMem, and is available in 128 GB, 256 GB, and 512 GB modules. PMem modules can coexist with traditional DDR4 DIMMs, with up to 4 TB of memory per socket. Like the 100 series, the 200 series can be used in either Memory Mode or App Direct Mode (refer to the product brief for more information). The 200 series offers several enhancements compared to Intel Optane PMem 100 series:

- Increased maximum DDR-T speed from 2666 MT/S to 3200 MT/s (2 DIMMs per channel).
- An average of 32 percent higher memory bandwidth per channel.[4]
- Improved application performance by using extended asynchronous DRAM refresh (eADR) to avoid CPU cache flush commands at runtime.

Intel Ethernet 800 Series

Intel Ethernet 800 Series is the next evolution in Intel's line of Ethernet products. Compared to the Intel Ethernet 700 Series, the 800 Series offers higher bandwidth due to use of PCIe 4.0 and 50 Gb/s PAM4 SerDes. It also improves application efficiency with Application Device Queues and enhanced Dynamic Device Personalization. The 800 Series is versatile, offering 2x100/50/25/10 GbE, 4x25/10 GbE, or 8x10 GbE connectivity. It also supports RDMA for both iWARP and RoCE v2, which gives enterprises a choice when designing their hyperconverged networks.

Intel DL Boost

3rd Gen Intel Xeon Scalable processors offer something unique that is not available with any other processor on the market: Intel DL Boost with VNNI. This technology takes advantage of, and improves upon, Intel® Advanced Vector Extensions 512 (Intel® AVX-512). VNNI improves AI performance by combining three instructions into one, thereby optimizing compute resources, utilizing the cache more effectively, and avoiding potential bandwidth bottlenecks. In Intel benchmarks, VNNI speeds the delivery of inference results by up to 11x, compared to the previous-generation Intel Xeon Scalable processor.[5]

Intel® oneAPI Toolkit

Modern workload diversity necessitates architectural diversity; no single architecture is best for every workload. XPUs, including CPUs, GPUs, FPGAs, and other accelerators, are required to extract high performance. Intel® oneAPI products deliver the tools needed to deploy applications and solutions across these architectures. Its set of complementary toolkits—a base kit and specialty add-ons—simplify programming and help developers improve efficiency and innovation. The core Intel oneAPI DPC++ Compiler and libraries implement the oneAPI industry specifications available at oneapi.io.

Intel oneAPI Base Toolkit is a foundational kit that enables developers of all types to build, test, and deploy performance-driven, data-centric applications across a wide variety of architectures. In addition, there are domain-specific toolkits that can be used for specialized workloads that are powered by or based on the oneAPI Base Toolkit. Examples include:

- Intel® AI Analytics Toolkit for accelerating end-to-end machine-learning and data science pipelines:
  - Intel® Optimization for TensorFlow
  - PyTorch Optimized for Intel® Technology
  - Intel® Distribution for Python
  - Intel® Optimization of Modin
  - Model Zoo for Intel® Architecture
  - Intel® AI Quantization Tools for TensorFlow
- Intel® Distribution of OpenVINO™ Toolkit for deploying high-performance inference applications from device to cloud. This toolkit includes:
  - OpenCV: Optimized Functions for Intel® Accelerators
  - Intel® Deep Learning Deployment Toolkit
  - Inference Support
  - Deep Learning Workbench
- Intel oneAPI DL Framework Developer Toolkit for building deep-learning frameworks or customizing existing ones. This toolkit includes:
  - Intel oneAPI Collective Communications Library
  - Intel oneAPI Deep Neural Network Library

Multicloud Analytics Solution Designs

The following tables describe the required and recommended components needed to build Base, Plus, Edge, and cloud-based designs for the Multicloud Analytics Solution. The VMware Cloud Foundation Management Domain, required for on-premises and edge/remote workload deployments, can consist of up to 14 linked vCenter Servers to manage up to 14 Workload Domains and 1,000 ESXi nodes total in multiple clusters, per single SDDC. Base and Plus clusters permit up to 64 ESXi nodes each. The Edge design permits up to four nodes per VMware Cloud Foundation Remote Domain.

Table 1. Hardware Bill of Materials for Base and Plus On-premises Domains

Component	Base Design (per node)	Plus Design (per node)	Required or Recommended
CPU	2x Intel® Xeon® Gold 6342 processor (2.8 GHz, 24 cores)	2x Intel Xeon Platinum 8362 processor (2.8 GHz, 32 cores)	Required
Memory (DRAM)	512 GB (16x 32 GB DDR4 DRAM)	256 GB (16x 16 DDR4 DRAM)	Required
Memory (Intel® Optane™ PMem)	N/A	1024 GB (8 x 128 GB, PMem 200 series)[a]	Recommended
Boot Drive	1x Intel® Optane™ SSD P1600X 118 GB	1x Intel Optane SSD P1600X 118 GB	Recommended
Storage (Cache)	2x Intel Optane SSD DC P5800X 400 GB	2x Intel Optane SSD DC P5800X 800 GB	Required
Storage (Capacity)	4x Intel® SSD D7-P5510 3.84 TB	6x Intel SSD D7-P5510 3.84 TB	Required
Network Adapter	1x Intel® Ethernet Adapter E810-CQDA2 (100 GbE)	2x Intel Ethernet Adapter E810-CQDA2 (100 GbE)	Recommended
Top of the Rack (ToR) Switch	100 GbE per port *Switch capabilities: Jumbo Frames, BGP*	100 GbE per port *Switch capabilities: Jumbo Frames, BGP*	Recommended
SD-WAN Appliance	VMware SD-WAN Edge 3x00 Hardware Appliance *or* VMware SD-WAN Edge Virtual Appliance for ESXi (8-cores)[b]		Required (1) or Recommended (HA pair)

[a] Officially supported by VMware Cloud Foundation v4.3 and higher.

[b] For all SD-WAN Edge model performance and feature comparisons, refer to the VMware SD-WAN Edge platform specifications datasheet.

Table 2. Hardware Bill of Materials for Edge Remote Domain

Component	Verified Configuration (per node)	Alternative Configuration for Demanding Workloads[a]	Required for Validation
CPU	2x Intel® Xeon® Gold 5215 processor (2.5 GHz, 10 cores)	2x Intel Xeon Silver 4310 processor (2.1 GHz, 12 cores)	Required
Memory (DRAM)	192 GB (6x 32 GB DDR4 DRAM)	256 GB (8 x 32 GB DDR4 DRAM)	Required
Boot Drive	1x Intel® SSD D3-S4510 240 GB	1x Intel® Optane™ SSD P1600X 118 GB	Recommended
Storage (Cache)	1x Intel SSD DC P4610 1.6 TB	1x Intel SSD DC P4610 1.6 TB	Required
Storage (Capacity)	4x Intel SSD D3-S4510 1.92 TB	4x Intel SSD D3-S4510 1.92 TB	Recommended
Network Adapter	1x Intel® Ethernet Adapter X722-DA2 (2x 10GbE, SFP+)	1x Intel Ethernet Network Adapter E810-XXVDA2	Required
ToR Switch	25 GbE per port *(switch capabilities: Jumbo Frames, BGP)*		Required
SD-WAN Appliance	VMware SD-WAN Edge 6x0 Hardware Appliance		Required

[a] This configuration has not been verified by Intel as a reference architecture, but offers the latest technologies from Intel to support data-intensive and performance-sensitive edge workloads.

Table 3. Hardware Bill of Materials for Management Domain (required for on-premises and edge deployments)

Component	Verified Configuration (per node)	Alternative Configuration for Demanding Workloads[a]	Required or Recommended
CPU	2x Intel® Xeon® Gold 6248 processor (2.5 GHz, 24 cores)	2x Intel® Xeon® Gold 5318Y processor (2.1 GHz, 24 cores)	Required
Memory (DRAM)	128 GB (8x 16 GB DDR4 DRAM)	128 GB (4x 32 GB DDR4 DRAM)	Required
Memory (Intel® Optane™ PMem)	512 GB (4x 128 GB, Intel Optane PMem 100 series)	512 GB (4x 128 GB, Intel Optane PMem 200 series)	Recommended
Boot Drive	1x Intel® SSD D3-S4510 240 GB	1x Intel Optane SSD P1600X 118 GB	Recommended
Storage (Cache)	2x Intel Optane SSD DC P4800X 375 GB	2x Intel Optane SSD DC P5800X 800 GB	Required
Storage (Capacity)	6x Intel SSD P4510 4 TB	6x Intel SSD D7-P5510 Series 3.84 TB	Required
Network Adapter	1x Intel® Ethernet Adapter E810-CQDA2 (100 GbE)		Recommended
ToR Switch	100 GbE per port *(switch capabilities: Jumbo Frames, BGP)*		Recommended

[a] This configuration has not been verified by Intel as a reference architecture, but offers the latest technologies from Intel to support data-intensive and performance-sensitive edge workloads.

Table 4. Public CSP Details

Cloud Service Provider	Instance Type	Workload Characteristics
AWS	I3.metal I3en.metal	General-purpose clusters Data-intensive workloads
Microsoft Azure	AV36	Balanced configuration with all-flash storage

Table 5. Software Requirements (all required)

Software	Version	Build Number
VMware Cloud Foundation	4.2.1	18016307
Cloud Builder VM	4.2.1	18016307
VMware ESXi Hypervisor	7.0 Update 1d	17551050
VMware vSAN	7.0 Update 1d	Included in ESXi bundle
VMware vCenter Server Appliance	7.0.1.00301	17956102
VMware NSX-T Data Center	3.1.2	17883596
SDDC Manager	4.2.1	18016307
VMware vRealize Suite Lifecycle Manager	8.2.0	17513665
vSphere Kubernetes	v1.18.2	-

Table 6. Platform and Software Settings

Software	Base/Plus	Edge	Management	Required or Recommended
TPM 2.0 or Intel® PTT	Enabled	Enabled	Enabled	Recommended
Intel® Volume Management Device	Enabled	Enabled	Enabled	Recommended
Intel® Hyper-Threading Technology	Enabled	Enabled	Enabled	Required
Intel® Turbo Boost Technology	Enabled	Enabled	Enabled	Required
Uncore Frequency Scaling	Enabled	Enabled	Enabled	Recommended
Power Management Settings	Performance	Balanced	Performance	Recommended
Secure Boot	Enabled	Enabled	Enabled	Recommended
vSAN Disk Groups	2 per host	1 per host (min.)	2 per host	Required

Table 7. Firmware Versions (all required)

Ingredient	Base/Plus	Edge	Management	Required or Recommended
BIOS	SE5C6200.86B.0022.D64.2105220049	SE5C620.86B.02.01.0010.010620200716	SE5C620.86B.02.01.0010.010620200716	Required
BMC	2.81.76f13ccc	2.37.1f190479	2.37.1f190479	Required
ME	04.04.04.56	04.01.04.339	04.01.04.339	Required
SDR	0.35	1.98	1.98	Required
CPU microcode	0x0d0002b1	0x04002f01	0x05002f01	Required
Intel® Optane™ SSD DC P5800X firmware	L0310100 or later	N/A	L0310100 or later	Required
Intel® SSD D7-P5510 firmware	JCV10100 or later	N/A	JCV10100 or later	Required
Intel® SSD D3-S4510 firmware	N/A	XCV10132 or later	N/A	Required
Intel® SSD DC P4610 firmware	N/A	VDV10152 or later	N/A	Required
Intel® Network Adapter E810 firmware (NVM)	2.40 or later	2.40 or later	2.40 or later	Required
Intel® X722 Network Adapter firmware (NVM)	N/A	6.80 or later	N/A	Required

Security

For all Intel® architecture-based solutions, we recommend installing the Trusted Platform Module (TPM) and enabling Secure Boot, which allows administrators to secure platforms for a trusted (measured) boot with known trustworthy (measured) firmware and OS. The TPM also enables local and remote attestation by third parties to advertise such known good conditions (assuming the presence of Intel® Trusted Execution Technology).

Infrastructure Overview

VMware Cloud Foundation On-Premises

Cloud Builder

This is the virtual appliance (VM) used for automated deployment of the entire Management Domain software stack.

VMware SDDC Manager

SDDC Manager manages the bring-up of the VMware Cloud Foundation system, creates and manages Workload Domains, and performs lifecycle management to keep the software components up to date. SDDC Manager also monitors the logical and physical resources of VMware Cloud Foundation.

VMware vSphere with Tanzu

VMware vSphere extends virtualization to storage and network services and adds automated, policy-based provisioning and management. vSphere is the starting point for building an SDDC platform. VMware vSphere with Tanzu enables streamlined development, agile operations, and accelerated innovation for all enterprise applications. It consists of two core components: ESXi and vCenter Server. ESXi is the virtualization platform used to create and run VMs and virtual appliances, while vCenter Server manages multiple ESXi hosts as clusters, using shared pool resources.

VMware vSphere with Tanzu Workload Management enables the deployment and operation of compute, networking, and storage infrastructure for VMware TKG Service. It makes it possible to use vSphere as a platform for running Kubernetes workloads natively on the hypervisor layer.

VMware TKG Service for vSphere

This reference architecture uses the VMware TKG Service for vSphere offering, which is now integrated with vSphere 7.0 and is available starting from VMware Cloud Foundation 4.0. Kubernetes workloads may run directly on ESXi hosts, and upstream Kubernetes clusters can be created and operated within dedicated resource pools by using the TKG Service. Check the vSphere with Tanzu product webpage for a more high-level overview.

TKG is also available in other offerings (besides vSphere). These offerings can be used to provision and manage the lifecycle of Tanzu Kubernetes clusters, which are the proprietary installations of Kubernetes open-source software, built and supported by VMware. To learn more about TKG offerings, consult the VMware Tanzu Kubernetes Grid documentation main page.

VMware NSX-T Data Center

NSX-T Data Center (formerly NSX-T) is the network virtualization platform that enables a virtual cloud network with a software-defined approach. It works like a network hypervisor to reproduce a complete set of Layer 2 through Layer 7 networking services: routing, switching, access control, firewalls, QoS, and DHCP in software. All these components can be used in any combination to create isolated virtual networks on demand. The services can then be extended to a variety of endpoints within and across clouds. Starting with VMware Cloud Foundation 4.0, both Management and Workload Domain types support the NSX-T Data Center platform.

VMware vRealize Suite

VMware vRealize Suite is a multicloud management solution that provides IT organizations with a modern platform for infrastructure automation, consistent operations, and governance based on DevOps and machine-learning principles.

Multi-Cloud Offerings with VMware Cloud

VMC and AVS are the hybrid cloud solutions that allow easy extension, migration, and modernization of applications, and protection of applications in the public cloud. Both VMC and AVS infrastructures are delivered by the same vSphere-based SDDC stack that is used on-premises. The solutions take advantage of existing tools, processes, and familiar VMware technologies, along with native integration with AWS or Microsoft Azure services. This makes it easy to adopt, greatly reduces service disruption associated with migrating critical services to the cloud, and eliminates the need for rearchitecting the environment to suit a public cloud infrastructure.

The enterprise-grade infrastructure is delivered as a service, and has preconfigured vSAN storage, networking, compute, and security. VMC can also autoscale nodes as needed, depending on CPU, memory, and storage requirements. For AVS, extending the existing cluster or adding clusters is also supported.

Tanzu on VMware Cloud

VMware TKG is a multi-cloud Kubernetes footprint that you can run in the public cloud on Amazon AWS and Microsoft Azure, as a part of VMware Tanzu Standard or Advanced editions. Both editions are fully commercially supported by VMware when deployed to VMC and AVS (as well as to native public clouds such as AWS and Azure).

To operate a consistent Kubernetes distribution across each public cloud environments and enable centralized control across your entire Kubernetes estate, Tanzu Standard edition is the right choice. To run custom, containerized applications on Kubernetes at scale, Tanzu Advanced edition is a more suitable solution—it addresses the operational requirements for security, visibility, and manageability across clouds, while enabling development teams with self-service access to resources and automated functions, such as container builds.

Refer to the Compare VMware Tanzu Editions webpage for Standard versus Advanced feature comparison and FAQ.

VMware Cloud Foundation at the Edge

VMware Cloud Foundation Remote Workload Domains/Clusters

VMware Cloud Foundation Remote Clusters is a feature that enables the deployment of a Workload Domain or cluster at a remote site through SDDC Manager placed at the central location. This makes it possible to deploy and manage a full-stack lifecycle of the remote sites using a single SDDC Manager.

VMware SD-WAN and SASE Solutions

VMware SD-WAN is a cloud-delivered software-defined WAN that enables enterprises to support application growth, network agility, and simplified branch and end-point implementations. It also delivers high-performance, reliable access more securely to cloud services, private data centers, and software-as-a-service (SaaS)-based enterprise applications. The SD-WAN platform takes advantage of Intel® QuickAssist Technology (Intel® QAT), Data Plane Development Kit (DPDK), and Intel® Advanced Encryption Standard – New Instructions (Intel® AES-NI) features to deliver fast data-plane performance for virtualized SD-WAN, security, and other network functions. The ability to innovate and add features through updates to the SD-WAN software running on Intel architecture-based hardware helps organizations to meet evolving edge computing needs for application performance and reliability.

For more information about VMware SD-WAN, visit the product webpage.

With the shift to cloud adoption, enterprises have had to rethink the way security is enforced and security providers have had to evolve how they deliver their services. This is where SASE comes in. According to Gartner, "SASE capabilities are delivered as a service, based upon the identity of the entity, real-time context, enterprise security policies, and continuous assessment of risk and trust throughout the sessions."[6] The SASE model consolidates network and security into a cloud-delivered service that is fast, reliable, and software-defined—meaning SASE combines network as a service and security as a service.

VMware's SASE solution delivers secure, optimal, and automated access to applications and workloads in the cloud by extending software-defined networking and security to the doorstep of major IaaS and SaaS providers. Cloud-native VMware SASE protects users, apps, and distributed workloads against threats by harnessing the power of zero-trust network access (ZTNA), SD-WAN, and a next-generation secure web gateway (SWG). Global points of presence (PoPs), strategically distributed around the world, serve as an on-ramp to SaaS and other cloud services to easily scale organizations' SASE needs.[7]

For more details about VMware SASE, visit the product webpage.

Use Case: Retail at the Edge

The proposed use case showcases a solution that can increase retail customer engagement and improve the shopping experience. We include three scenarios:

- **Product recommendations.** When the client shows interest in a specific area or department, we can use a machine-learning algorithm to send personalized product recommendations. Based on people's similar choices and the customer's position in the store, the algorithm creates a list of the most relevant products. The customer is notified and can check the personalized recommendations using a mobile application. The process occurs every time the system discovers a new customer interest.
- **Presence detection.** We use deep-learning techniques and image recognition algorithms to detect customers in the Customer Service area. Cameras installed in the store send images to the deep-learning pipeline. When such an event occurs, the store staff is informed.
- **Hesitance detection.** When a customer is wandering around the store with no apparent purpose, without stopping, the business rules engine assumes the customer is looking for something, is lost, or may need assistance. A notification—including the customer's name, age, gender, and position in the store—is sent to the store staff so they can quickly find and identify a person in need.

Implementation Guidelines

To implement the given scenarios, we use a microservices architecture. Components have atomic roles and they are responsible for exchanging, processing, and storing data. Figure 3 shows a high-level overview of the system's architecture.

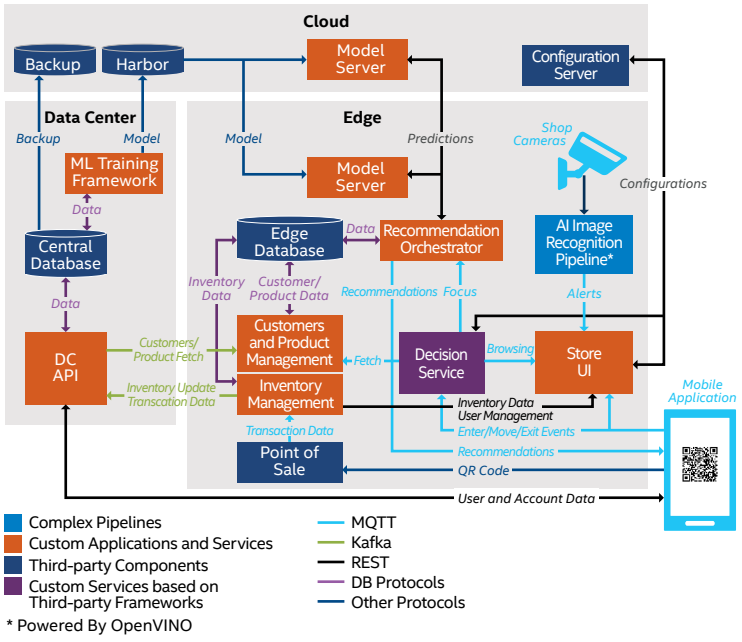


Figure 3. Microservices architecture with machine learning and deep learning.

To enable the architecture to support product recommendation, presence detection, and hesitance detection, we use both custom and third-party components. The system is built using frameworks and products such as FastAPI, Quarkus, Apache Kafka, and MQTT.

Model training for product recommendations is based on libraries and frameworks that are optimized for Intel architecture. A multicloud solution helps with high availability and reliability of the system. Because of a small message footprint, recommendations can be delivered both in the cloud and at the edge. Cloud applications are easily scalable. Depending on the number of stores and current traffic, the number of instances can change dynamically.

The image recognition service for presence detection uses the Intel Distribution of OpenVINO Toolkit and the OpenVINO Model Server. For this edge retail use case, a model from the OpenVINO Toolkit's Open Model Zoo repository can be used. It is also possible to create a new model or retrain an existing one. Deep-learning inference and image capturing are performed at the edge, so there is no additional network latency. The process is optimized and utilizes Intel DL Boost features.

For detecting user indecision, the system uses a rule-based engine, which in the absence of training data is a simple and fast solution. The system collects customers' behavior for further analysis and potential machine-learning usage.

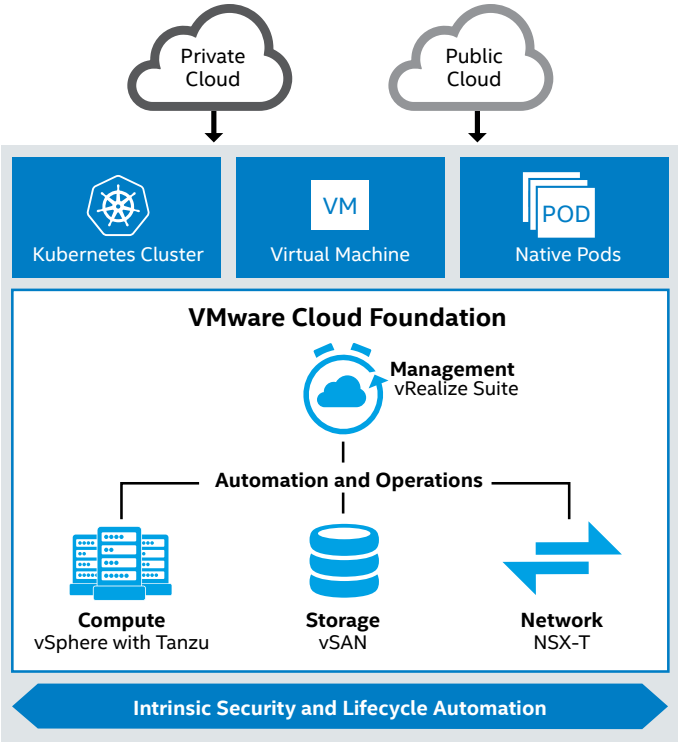


Figure 4. VMware Cloud Foundation 4.2 logical view. Source: docs.vmware.com/en/VMware-Cloud-Foundation/4.2/vcf-42-introducing/GUID-7EBCC024-9604-4064-90A1-4851A78F7641.html

Deployment Blocks

The goal of using solutions like VMware Cloud Foundation, NSX-T Data Center, TKG Service, and vSAN is to transform the legacy data center into an SDDC, where administrators can define, deploy, and manage clusters and resources based on actual demand from end users. Each of the mentioned components is a standalone product and may be used independently.

VMware Cloud Foundation

VMware Cloud Foundation is an integrated software platform that automates the deployment and lifecycle management of a complete SDDC on a standardized hyperconverged architecture. VMware Cloud Foundation consists of several core components (see Figure 4):

- VMware vSphere for compute virtualization
- VMware NSX-T Data Center for network virtualization
- VMware vSAN for storage virtualization
- Tanzu Mission Control for workload management
- VMware vRealize Suite for cloud monitoring

VMware Cloud Foundation allows organizations to build enterprise-ready cloud infrastructure for the private and public cloud. The standard architecture model for VMware Cloud Foundation includes a dedicated Management Domain (one per instance) for all management components and up to 14 virtual infrastructure Workload Domains created by the end user.

Management Domain

The Management Domain is a special-purpose Workload Domain that is used to host the infrastructure components needed to instantiate, manage, and monitor the VMware Cloud Foundation infrastructure. It is automatically created using the Cloud Builder on the first rack in the VMware Cloud Foundation system during bring-up. It contains management components such as SDDC Manager, vCenter Server, NSX-T Management Cluster, and optional components from VMware vRealize Suite.

The Management Domain uses vSAN as primary storage and requires a minimum of four nodes to work properly. When more racks are added to the system, the Management Domain automatically integrates those additional components.

Workload Domains

Workload Domains are a logical grouping of private cloud capacity; they are provisioned automatically by SDDC Manager. Each Workload Domain is administered and patched independently and has its own compute, storage, and network resources to consume. All the tasks related to the Workload Domains are performed using the SDDC Manager web interface. This includes the creation, expansion, and deletion of Workload Domains, along with physical-infrastructure monitoring and management. For more information, review the FAQ for VMware Cloud Foundation.

VMware vSAN

VMware vSAN is storage virtualization software—fully integrated with VMware vSphere—that joins all storage devices across a vSphere cluster into a shared data pool (see Figure 5). vSAN eliminates the need for external shared storage. Two vSAN cluster configurations are possible:

- A hybrid vSAN cluster uses two types of storage devices: flash devices for the cache tier and magnetic drives for the capacity tier.
- An all-flash vSAN cluster uses flash devices for both the cache and capacity tiers.

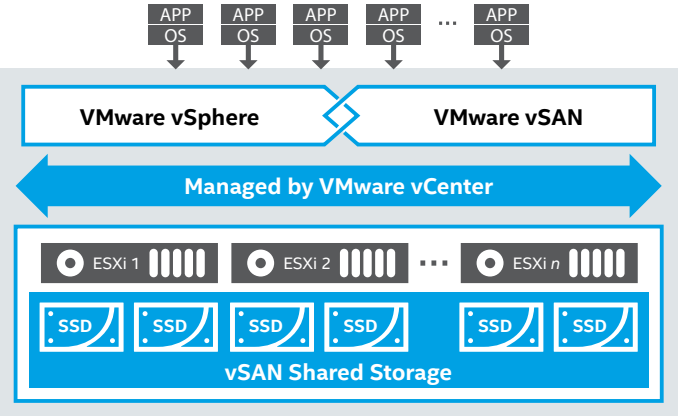


Figure 5. VMware vSAN combines storage devices into a shared data pool.

VMware NSX-T Data Center

NSX-T Data Center 3.1 includes a variety of new features for virtualized networking and security for private, public, and multicloud. These include NSX-T Federation enhancements, simplified migration from legacy NSX-V product to NSX-T, Distributed IPS enhancements, improved lifecycle management, and monitoring.

Starting from version 3.0, NSX-T Data Center can run on the vSphere Distributed Switch (VDS) version 7.0 on vSphere. The N-VDS NSX-T host switch that was used in the previous releases will be deprecated in a future release; however, it will still remain as a switch on KVM and for bare-metal workloads.

Another feature, beginning with the NSX-T Data Center 3.0 release, is support for Intel QAT provided on bare-metal servers. Intel QAT provides the hardware acceleration for various cryptography operations. NSX-T Edge takes advantage of Intel QAT to improve VPN performance.[8] For the list of changes introduced by the NSX-T Data Center 3.1 release, refer to the What's New section of the release notes.

NSX-T Data Center Components

The main components of VMware NSX-T Data Center are NSX Manager, NSX Controllers, and NSX Edge gateways:

- **NSX Manager** is a centralized component of NSX that is used for network management. This virtual appliance provides the GUI and the RESTful APIs for creating, configuring, orchestrating, and monitoring NSX-T Data Center components. NSX Manager is the management plane for the NSX-T Data Center ecosystem.
- **NSX Controllers** are a distributed state-management system used to overlay transport tunnels and control virtual networks, which can be deployed as VMs on VMware ESXi or KVM hypervisors. The NSX Controller manages all logical switches within the network, and it handles information about VMs, hosts, switches, and virtual segments. Using three controller nodes ensures data redundancy in case one NSX Controller node fails.
- **NSX Edge** is a gateway service that provides access to physical and virtual networks for VMs. It can be installed as a distributed virtual router or as a services gateway. The following services can be provided: dynamic routing, firewalls, network address translation (NAT), DHCP, VPNs, load balancing, and high availability. NSX Edge can connect to two transport zones—one for overlay and the other for north-south peering with external devices (see Figure 6).

Transport zones define which hosts and which VMs can participate in the use of a given network. There are two main transport zones that define the limits of logical network distribution on the NSX Edge:

- **The Overlay Transport Zone** provides east/west traffic in an overlay/tunnel between VMs, ESXi hosts, and NSX-T Edges using GENEVE encapsulation.
- **The VLAN Transport Zone** connects NSX Edge uplinks to the physical routers/switches. It provides north/south traffic between the overlay network and the external networks. It is sometimes referred as the uplink network.

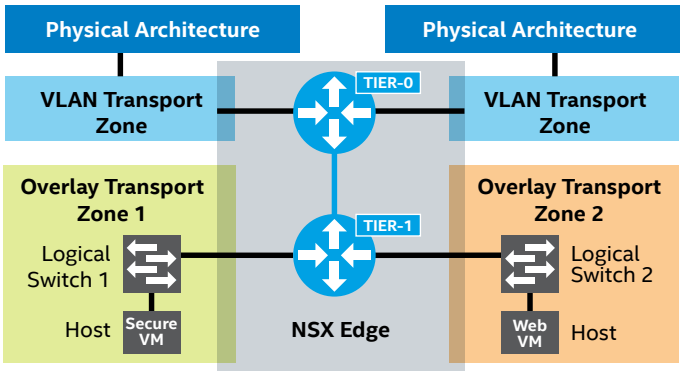


Figure 6. VLAN and Overlay Transport Zones using VMware NSX-T Data Center.
Source: docs.vmware.com/en/VMware-NSX-T-Data-Center/2.3/com.vmware.nsxt.install.doc/GUID-F47989B2-2B9D-4214-B3BA-5DDF66A1B0E6.html

vSphere with Tanzu

When vSphere with Tanzu is enabled on a vSphere cluster, it creates a Kubernetes control plane within the hypervisor layer. This layer provides objects that enable the capability to run Kubernetes workloads within ESXi. This control plane is called a Supervisor Cluster. It runs on top of an SDDC layer that consists of ESXi nodes for compute, NSX-T Data Center for networking, and vSAN for shared storage. The shared storage is used as persistent volumes for vSphere Pods, VMs running within the Supervisor Cluster, and pods within the Tanzu Kubernetes clusters. When a Supervisor Cluster is created, the vSphere administrator can create namespaces within it (Supervisor Namespaces) to provide access to DevOps engineers, who then can run workloads consisting of containers operating inside vSphere Pods or create Tanzu Kubernetes clusters.

vSphere Pods

A vSphere Pod is a new construct introduced by vSphere with Tanzu. It is an equivalent of a Kubernetes Pod. A vSphere Pod is a VM running one or more containers with a small footprint. Each vSphere Pod is an object in vCenter Server; for networking needs, vSphere Pods use the topology provided by NSX-T Data Center. In this Reference Architecture, we concentrate on using TKG clusters instead of vSphere Pods. For more information, read about vSphere Pods versus Tanzu Kubernetes clusters.

Supervisor Cluster Architecture

An overview of the Supervisor Cluster architecture is shown in Figure 7. Aside from the regular ESXi components, there are some new elements:

- The Spherelet process is a kubelet that is ported natively to ESXi and allows the ESXi host to become part of the Kubernetes cluster.
- Kubernetes control plane VMs are a total of three load-balanced machines for the Kubernetes control plane.
- Container Runtime Executive (CRX) includes a paravirtualized Linux kernel that works together with the hypervisor. CRX uses the same hardware virtualization techniques as VMs and has a VM boundary around it.
- Virtual Machine Service, Cluster API, and TKG Service are modules that run on the Supervisor Cluster and enable provisioning and management of Tanzu Kubernetes clusters.

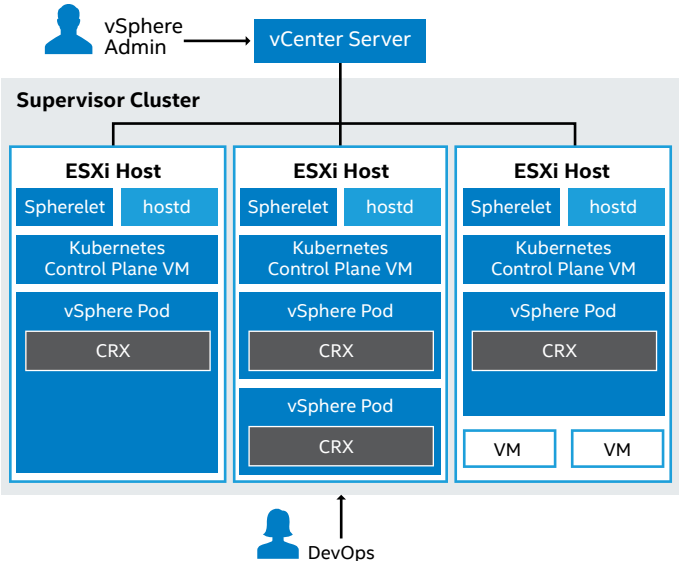


Figure 7. Supervisor Cluster architecture.
Source: docs.vmware.com/en/VMware-vSphere/7.0/vmware-vsphere-with-kubernetes/GUID-3E4E6039-BD24-4C40-8575-5AA0EECBBBEC.html

Supervisor Namespace

This namespace allows the vSphere administrator to define the resource boundaries where vSphere Pods and Tanzu Kubernetes clusters are created when using the TKG Service. The administrator can set limits for CPU, memory, and storage as well as the number of Kubernetes objects that can run within the namespace. A resource pool is created per each namespace in vSphere. User permissions can be set to users and groups to allow access to namespaces using an identity source that is associated with vCenter Single Sign-on. After the namespace is created, configured with resources, and set with access for users, the namespace can be accessed to run Kubernetes workloads and create Tanzu Kubernetes clusters by using the TKG Service.

Tanzu Kubernetes Clusters

A Tanzu Kubernetes cluster (see Figure 8) is a full distribution of the open-source Kubernetes software, signed and supported by VMware. You can use the TKG Service to provision Tanzu Kubernetes clusters on the Supervisor Cluster. The TKG Service API can be invoked by using kubectl and a YAML definition. Once deployed, Tanzu Kubernetes clusters can be accessed and used in the same way—and use the same tools—as a standard Kubernetes cluster. The entire Kubernetes environment may exist in parallel to any regular VMs in the cluster, as seen in Figure 8. Each namespace can be seen in the vSphere GUI as a Resource Pool.

From the logical overview, the Tanzu Kubernetes cluster exists within a Supervisor Cluster. vSphere with Tanzu consists of a single availability zone within a single geographic region. See Figure 9 for a general overview.

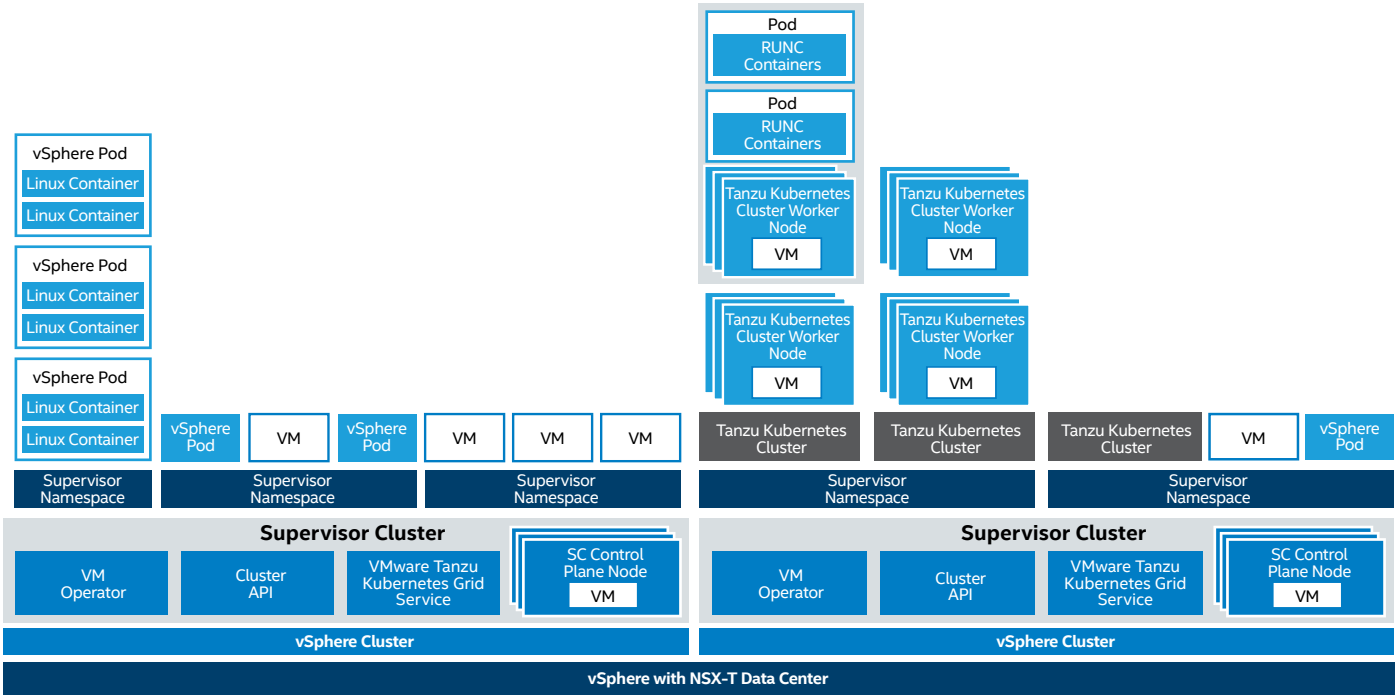


Figure 8. An example of vSphere with Tanzu for Tanzu Kubernetes clusters.
Source: docs.vmware.com/en/VMware-vSphere/7.0/vmware-vsphere-with-kubernetes/GUID-3E4E6039-BD24-4C40-8575-5AA0EECBBBEC.html

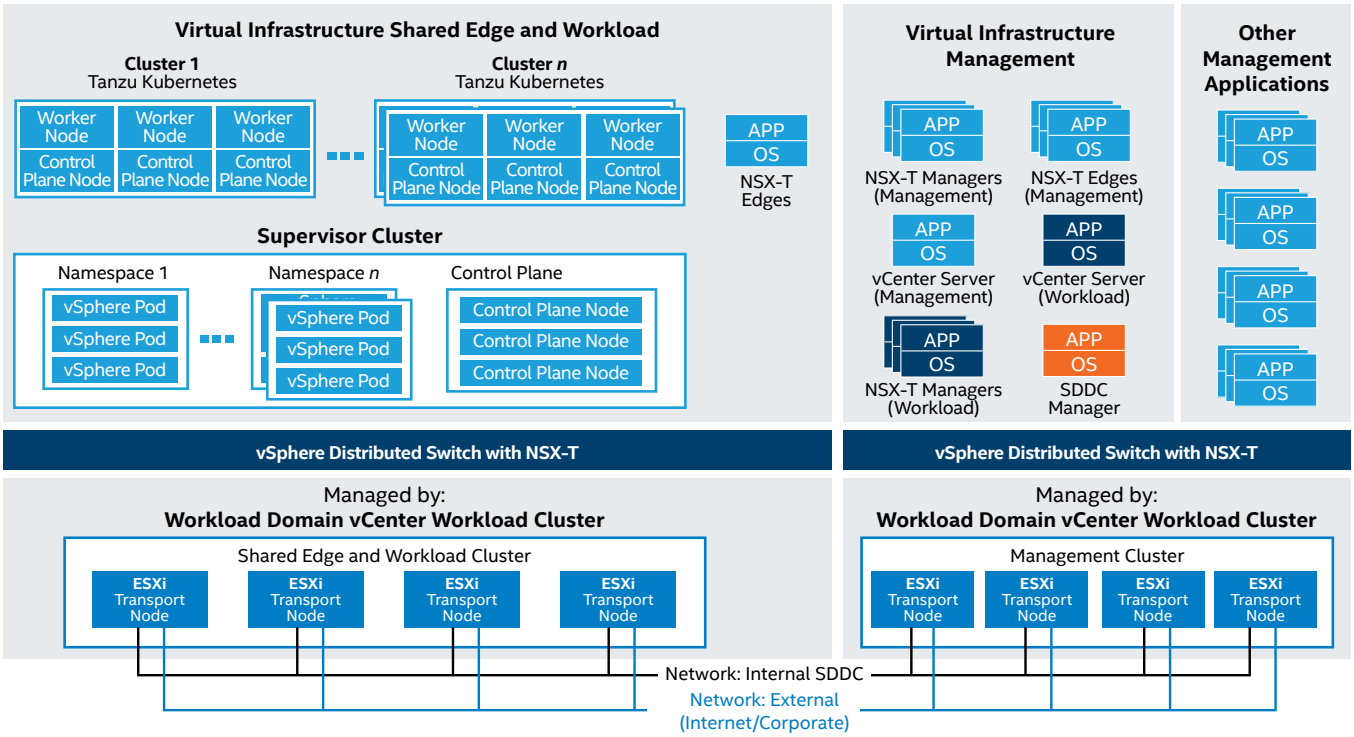


Figure 9. Tanzu Kubernetes and VMware Cloud Foundation 4.2 overview.
Source: docs.vmware.com/en/VMware-Validated-Design/6.0/sddc-architecture-and-design-for-a-vsphere-with-kubernetes-workload-domain/GUID-D72DB286-1907-4AF6-A644-42FBAB2BB7C7.html

VMware Cloud Foundation in the Public Cloud

VMware Cloud Foundation can be extended as a complete VMware solution in the public cloud (see Figure 10), such as on AWS or Azure. A cloud deployment consists of the same components as the on-premises environment—vSphere, vSAN, NSX-T Data Center, and vRealize—allowing rapid extension, migration, and protection of a regular VMware environment directly to the public cloud, along with seamless integration for deployment of Kubernetes. Optionally, with additional tools and add-ons (HCX and Hybrid Linked Mode), it provides methods of VM migration to and from the cloud. Cloud deployments have two distinctive pre-configured regions: one for management and one for the customer. VMware is responsible for the management portion and customers control the operational portion. Users have very limited access to the management resources and settings but can manage workloads from the compute resource pool.

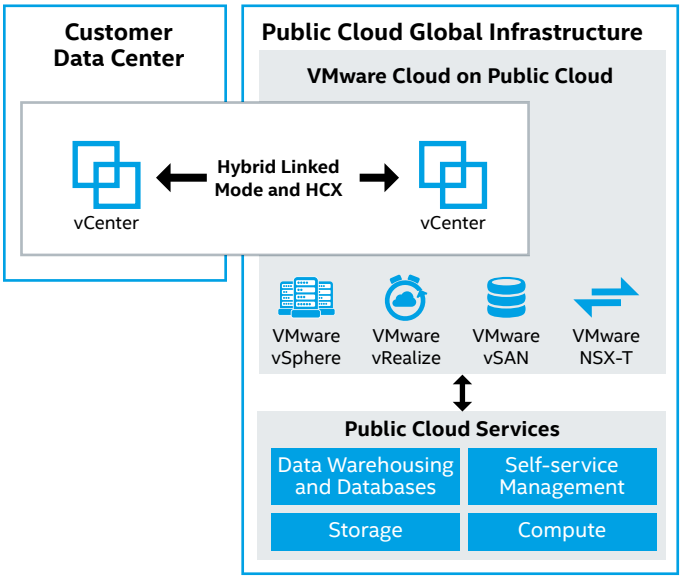


Figure 10. Components of VMware Cloud in the public cloud.
Source: youtube.com/watch?v=Ol7rNfkZT2c

On the network level, two gateways provide connectivity to VMware Cloud Foundation.

- The Management Gateway (MGW) enables users to connect to the management layer (vCenter, ESXi hosts, NSX-T Data Center, and optional SRM and HCX Managers), which uses a dedicated management subnet and restricted ports.
- The Compute Gateway (CGW) enables ingress and egress of workload VM and container network traffic traversing in and out of VMware Cloud.

The Distributed Firewall feature allows traffic to be filtered between VMs and/or containers on different segments within VMware Cloud. There are no restrictions on the CGW or Distributed Firewall, and users can configure firewall rules as they choose. The MGWs and CGWs use separate VMware NSX Edges. See Table 4 for details about which cloud instances are recommended for cloud-based VMware Cloud Foundation deployments.

Tanzu Mission Control

VMware Tanzu Mission Control (see Figure 11) is a centralized management platform for consistently operating and securing your Kubernetes infrastructure and modern applications across multiple clouds. It provides operators with a single control point ("pane of glass").

Tanzu Mission Control is available as SaaS in the VMware Cloud Services portfolio of products.

The key capabilities of Tanzu Mission Control include:

- **Attaching clusters.** Attach any conformant Kubernetes clusters running in other environments—either on-premises or in public clouds—to Tanzu Mission Control for centralized management.
- **Cluster lifecycle management.** Provision, scale, upgrade, and delete Kubernetes clusters via Tanzu Mission Control with the hosted TKG runtime.
- **Centralized policy management.** Apply consistent policies—such as access, network, and container registry policies—to a fleet of clusters and namespaces at scale.
- **Observability and diagnostics.** Gain global observability of the health of clusters and workloads across clouds for quick diagnostics and troubleshooting.
- **Data protection.** Back up and restore your clusters, namespaces, and even groups of resources using Kubernetes label selectors, using the built-in open-source Velero project.

For more details regarding Tanzu Mission Control, view the product web page.

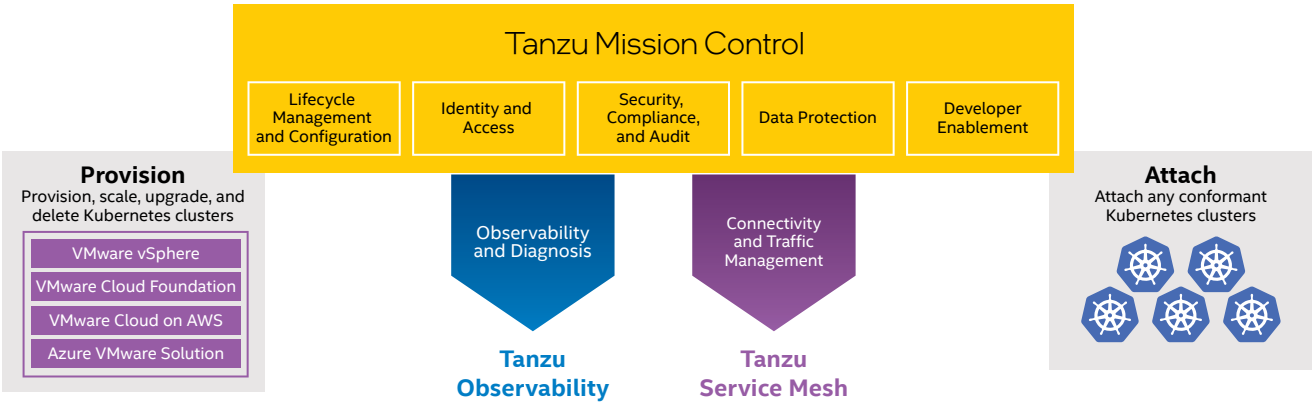


Figure 11. Tanzu Mission Control provides a centralized Kubernetes management platform for both ops and dev.

TKG in the Public Cloud

When a customer needs to have Kubernetes infrastructure on SDDC in the public cloud, VMware requires using Tanzu Standard and recommends Tanzu Advanced edition. These editions are not enabled by default on VMware's public cloud SDDC instances on either AWS or Azure; they are available as an additional subscription.

TKG as a part of Tanzu Standard or Advanced is infrastructure-agnostic; therefore, there is no dependency on vSphere 7 with Kubernetes add-on features, which is part of on-premises deployments of VMware Cloud Foundation. This also means that TKG for AWS or Azure does not rely on vSphere 7.0/ VMware Cloud Foundation 4.2 built-in Supervisor Cluster for Kubernetes Workload Management. Instead, it creates its own management cluster, running on VMs.

When initiating the bring-up of both management and workload TKG clusters on AWS or Azure, a user can choose between "Development" and "Production" types of deployments ("plans"). For Production deployments, control plane and worker nodes are deployed in groups of three VMs for redundancy. For more information about how to start deploying TKG to VMC and AVS or to native AWS and Azure, see the Prepare to Deploy Management Clusters webpage.

Remote Workload Domains for the Edge

VMware Cloud Foundation Remote Clusters enable administrators to create Workload Domains or clusters at the remote location, but also to expand an existing Workload Domain at the central site by adding a cluster at a remote location. It is also possible to run Tanzu Kubernetes clusters at remote sites.

Figure 12 shows the logical structure of the system and connections between the central SDDC Manager and remote sites when using the VMware Cloud Foundation Remote Clusters feature. vCenter Servers for all sites need to be placed at the central Management Domain. VMware Cloud Foundation Remote Clusters supports a minimum of three and maximum of four hosts. Servers at the remote site must be able to reach the management network at the central location. Redundant WAN connectivity is recommended to connect from the central site to each Cloud Foundation Remote Clusters site, hence VMware SD-WAN was used in the solution.

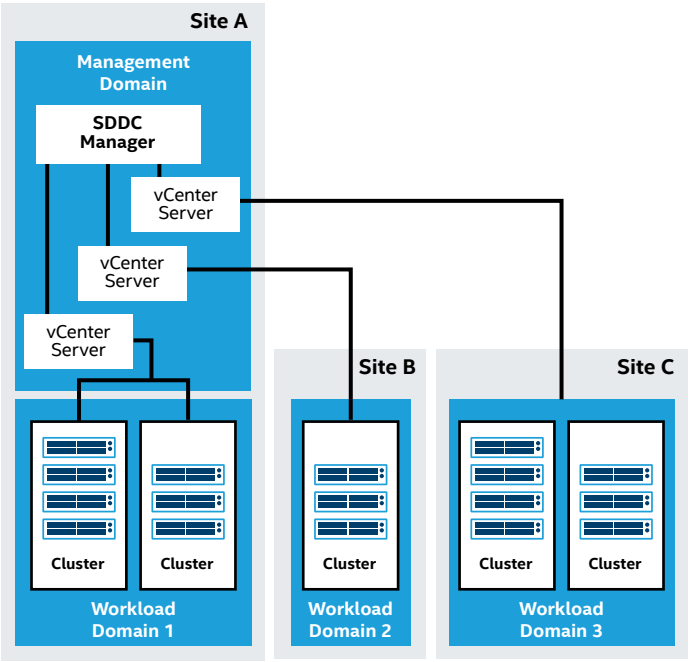


Figure 12. Logical connections between the central SDDC Manager and Workload Domains at remote sites.
Source: docs.vmware.com/en/VMware-Cloud-Foundation/4.2/vcf-admin/GUID-64B335F7-A67A-4287-AF72-0BBCDE003C2D.html

Environment Configuration and Deployment

Installation and Configuration of Intel Optane PMem Modules on VMware ESXi

To use Intel Optane PMem modules on an ESXi system, you must install and configure the modules on each server. The first step is installing them in the DIMM slots on the motherboard. Consult the platform manual for detailed rules and guidelines, because several restrictions exist on what configurations are possible. For example, one restriction is that the installed DIMM type and population configured to CPU1 must match CPU2. A detailed diagram for all possible population configurations is included in your platform manual.

Once the Intel Optane PMem modules are installed, choose between Memory Mode or App Direct Mode. Read Intel Optane Persistent Memory – BIOS settings for details. In the case of ESXi, the only requirement is to create a goal configuration for the region. If you intend to use App Direct Mode, set the Memory Mode [%] option to 0 (you still need to create the region). To benefit from additional memory with Memory Mode, change the Memory Mode [%] option to 100. You need to reboot the server each time you create a new goal configuration. Manual creation of namespaces is not necessary; ESXi will automatically create namespaces during boot if needed. After that, the system will be ready.

Important: If you want to change from Memory Mode to App Direct Mode (or vice versa), you must delete the namespaces created by ESXi after making the change in the BIOS. Follow the instructions in Delete a Namespace in the VMware Host Client.

Important: If you plan to use App Direct Mode, make sure to configure it on the ESXi hosts AFTER completing the Workload Management deployment. There is a high risk of false-positive validation issues during the Workload Management validation phase due to the way SDDC Manager lists and categorizes various types of storage on ESXi hosts that are within the cluster. Existence of PMem type of storage during that phase may cause problems.

VMware SD-WAN Configuration

To provide connectivity between the core data center, cloud instances, and remote sites, SD-WAN deployment and configuration needs to occur. The central point for managing and monitoring is the VMware SD-WAN Orchestrator. Orchestrator is available as SaaS. From there, the user can modify and push configurations for specific SD-WAN Edge instances (both physical and virtual) or groups of them. Orchestrator ensures that all the networks/subnets are propagated to the endpoints. The entire network traffic is directed by default through VMware SD-WAN Gateways, with a use of VeloCloud Multipath Protocol (VCMP) for tunneling and encapsulation. It is also possible to include non-SD-WAN remote network environments via IPsec tunneling.

Setting up Edges

Edge mass configuration is specified using Profiles (see Configure Profiles for details). Those templates are used to assign segments, create VLANs, and configure WAN and LAN interfaces, DNS settings, internet access policies, and firewall rules. Upon configuration of an Edge, a profile must be assigned to it to provide a set of global rules. Locally specific configuration (i.e., static IP addressing of its network interfaces, static routes, VLANs, and NTP) may be set independently per each Edge; consult the Edge Device Configurations for all possible overrides that can be made. Edge instances can be deployed as physical hardware or virtual appliances at the target sites. Edge hardware appliances must be ordered and physically delivered to each target site. An Edge virtual appliance (VM for ESXi or KVM) is available to download from the VMware website's product download webpage. For detailed provisioning and activation steps, follow the Provision an Edge documentation.

Non-SD-WAN Destinations

To add a Cloud instance (i.e., VMware Cloud), the user must configure a non-SD-WAN destination. Such configuration establishes an IPsec tunnel to a given endpoint, such as AWS VPN Gateway, Cisco ASA, Microsoft Azure Virtual Hub, and many more. For configuration details, consult the official Configure a Non-SD-WAN Destination documentation provided by VMware. In the case of our VMware Cloud instance, we used the Non-SD-WAN Destination via Gateway method for establishing the IPsec tunnel. It is important to note that this approach requires a specific configuration on the VMC side (policy-based VPN) to match the exact tunnel parameters that are provided by the Orchestrator in the IKE IPsec Template. Remote and local networks on both sides must be provided manually; the Local Auth ID on the Orchestrator side might need to be manually overwritten with the IP address of the SD-WAN Gateway.

Setting up Routing

After Edges and Non-SD-WAN Destinations are in place, routing needs to be added. The user may choose from Border Gateway Protocol (BGP), Open Shortest Path First (OSPF), or static route definitions, depending on the environment needs. The SD-WAN Orchestrator needs to be aware of any network that is expected to be reachable from the remote machines; in some cases, it may be required to add specific routes manually for the given Edge instances. See the Configure Static Route Settings for details.

Environment Provisioning

As mentioned in the Solution Architecture Highlights section, the complete on-premises environment consists of three main products: VMware Cloud Foundation, VMware NSX-T Data Center for vSphere, and Tanzu Kubernetes clusters (deployed using TKG Service for vSphere). The following sections describe how to provision these components.

Initial Preparation: Hardware and Software Requirements

Starting with VMware Cloud Foundation 4.0, the deployment process is more streamlined than for earlier releases. The main improvement is that NSX-T Data Center installation and configuration is now integrated into the automated deployment of the Management Domain and any Workload Domains that are added later. Each Workload Domain can have its own NSX-T Data Center Domain.

Because of this change, the engineer that is executing the deployment needs to provide very detailed network environment information and must prepare routing and naming prior to starting the actual deployment wizard. VMware provides a Planning and Preparation Workbook that provides detailed information about the software, tools, and external services that are required for VMware Cloud Foundation bring-up. This is a recommended starting point for the deployment. The Workbook is a pre-configured spreadsheet with multiple sheets that group particular types of data/environment variables that are needed to be taken care of before the deployment can start. The scope of the documents includes:

- Checking the hardware and network requirements (for example, 802.1Q tagging and Jumbo Frames MTU)
- Setting up IP addressing for all critical components
- Planning network subnets required by management, vSAN, vMotion, tunnel endpoints, and uplinks
- BGP routing information for Edge VM Tier 0 routing
- All required DNS entries
- Licensing info and NTP configuration
- Network profiles for the Edge VMs
- Naming of all availability zones, resource pools, port groups, and so on
- Usernames and passwords

The Workbook also provides guidance on the additional optional components for the VMware Cloud Foundation environment, such as vRealize Log Insight, vRealize Operations Manager, vRealize Automation, and VMware Workspace ONE Access.

While completing the Workbook is optional and is not a required step in the bring-up process, it makes the preparation of the environment easier and provides insight into the necessary steps that can be taken care of in advance.

Important: When configuring the Host and Edge Termination End Point (TEP) networks that handle the GENEVE encapsulated traffic, be aware that any misconfiguration of the MTU size will result in unexpected and difficult-to-troubleshoot network issues when using Tanzu Kubernetes clusters (or any other type of workload). TCP traffic is likely to be impacted, including missing packets. Also, any form of packet filtering on the Host and Edge TEP network may cause issues. For example, load balancers may not be able to communicate with the IPs in their own IP pool.

Initiate the VMware Cloud Foundation Bring-Up

When all the prerequisites are fulfilled, the bring-up process can be started. The detailed procedure is documented in the Deploying VMware Cloud Foundation document. It is split into several steps:

Step 1: Deploy the Cloud Builder VM

The Cloud Builder VM is available as an OVA file. Follow the Deploy VMware Cloud Builder Appliance document for the detailed deployment procedure.

The Cloud Builder VM deploys the VMware Cloud Foundation. It deploys and configures the Management Domain. After the deployment is completed, all further configuration and control of VMware Cloud Foundation is transferred to the SDDC Manager. The Cloud Builder VM appliance must be deployed on the same management network as the VMware Cloud Foundation servers to automate the deployment and validate the network information provided (DNS, VLANs, IPs, MTUs). It also requires access to all external services like DNS and NTP.

Step 2: Install ESXi Software on VMware Cloud Foundation Servers

VMware Cloud Foundation deployment requires having a specific ESXi hypervisor version installed on the servers that will be used during bring-up. The Cloud Builder VM includes the VMware Imaging Appliance (VIA), which can be used to install ESXi on the VMware Cloud Foundation servers. The detailed procedure can be found here. Using VIA has some advantages, as it not only installs ESXi, but it can also deploy any additional VIBs and configure standard passwords across all machines. However, use of VIA is optional. As an alternative, you may install ESXi manually on all nodes. For the exact supported ESXi version, consult the BOM section of the VMware Cloud Foundation Release Notes. In case of manual installation, be sure to also install any required or custom VIBs that your servers need. In most cases, those will be specific drivers for NICs or SSDs.

Step 3: Download and Complete the Deployment Parameter Sheet for VMware Cloud Foundation

The Deployment Parameter Sheet provides the Cloud Builder VM with all the information required for bring-up. It's a separate file that needs to be downloaded from vmware.com, from the same place where the Cloud Builder VM OVA file was located (see Step 1 above). You should use the information in the Planning and Preparation Workbook to fill out the Deployment Parameter Sheet. After completing all the required variables, you will import this file during the VMware Cloud Foundation bring-up process.

Full documentation of the Deployment Parameter Sheet is available in the VMware Cloud Foundation Architecture and Deployment Guide. The documentation details all the tabs and fields included in the Deployment Parameter Sheet.

Step 4: VMware Cloud Foundation Bring-up

When you have ESXi installed on all management nodes, added all needed custom VIBs, and completed the Deployment Parameter Sheet, you can begin the VMware Cloud Foundation bring-up.

The complete description of the VMware Cloud Foundation bring-up process is included in VMware's Deploy the Management Domain Using VMware Cloud Builder documentation.

The bring-up process deploys SDDC Manager, vCenter Server for the Management Domain, NSX-T Management Domain, NSX-T Edge VMs, and vSAN—that is, the complete Management Domain of VMware Cloud Foundation. The process takes about two hours. After the bring-up process is complete, you should see a notification with a link to the new SDDC Manager web interface, which is accessible through a standard web browser.

Important: When the deployment of the Management Domain is completed, be sure to log on to SDDC Manager and provide credentials to your VMware account in the Administration ➔ Repository Settings section, so that SDDC Manager can start syncing against the VMware repository for any available software bundles.

Bring-up Process Summary

The Management Domain is now created and contains all the components needed to manage the infrastructure. You should not deploy any user applications on this management cluster. Instead, create one or more Workload Domains that comprise separate vSphere clusters with vSAN and NSX-T Data Center preinstalled and configured, along with an additional, dedicated instance of vCenter Server for each such domain. Starting from VMware Cloud Foundation 4.0, when you create a Virtual Infrastructure (VI) Workload Domain, you can choose to either deploy a new NSX-T Manager cluster for the Workload Domain, or to share an existing NSX-T Manager cluster that was previously created for another VI Workload Domain. Any new instances of vCenter Server (one per each Workload Domain) and an NSX-T Manager cluster (if needed) will be deployed on the Management Domain.

VMware Cloud Foundation Workload Domain Deployment

Deployment and configuration of Workload Domains is performed using SDDC Manager. The process is split into three steps. The first step includes commissioning of the ESXi hosts that will be used for the cluster and configuring initial components (such as vSAN and the NSX-T Management Domain). In the second step, the NSX-T Edge cluster is deployed onto the created Workload Domain to enable two-tier routing for north-south traffic. The third and final step deploys the necessary Kubernetes services.

The detailed procedure for a complete Workload Domain setup is available in the Start the VI Configuration Wizard section of the VMware Cloud Foundation Operations and Administration Guide.

Similar to the Management Domain, creating the Workload Domain requires multiple network IPs, DNS entries, services, and BGP routing information to be available. The BGP neighbor configuration on the L3 switches, which will later have a BGP peering with the Workload Domain's T0 router, must be done prior to the Workload Domain deployment. See the detailed prerequisites here. However, you may share the existing Host TEP and Edge TEP subnet across Management and Workload Domains, which simplifies the configuration.

Step 1: Commissioning Hosts and Creating a VI Workload Domain

Creating a new Workload Domain on VMware Cloud Foundation is controlled and orchestrated by SDDC Manager, which installs and configures all the needed components (including vCenter Server, vSAN, and NSX-T Data Center). To deploy the new Workload Domain, a sufficient number of ESXi hosts must be available in SDDC Manager. Adding new hosts to SDDC Manager is called commissioning. See the Commission Hosts section of the official documentation for the detailed procedure.

Once hosts are commissioned, use the VI Configuration Wizard to start the Workload Domain deployment; see Start the VI Configuration Wizard for detailed instructions.

Step 2: Deploying an NSX-T Edge Cluster to the Workload Domain

By default, all new Workload Domains do not include any NSX-T Edge clusters and are isolated from a network perspective. An NSX-T Edge cluster needs to be deployed to provide routing and network services.

The procedure is quite complex and requires a similar level of data and preparation as the initial deployment of the Management Domain. You will need to configure the network layout similar to what can be seen in Figure 13 on the next page. Follow the official instructions in VMware's Deploying NSX-T Edge Clusters documentation. The deployment process is started on SDDC Manager and is fully automated after providing the complete data, so the risk of error is greatly reduced compared to the manual configuration process in VMware Cloud Foundation 3.9 and earlier.

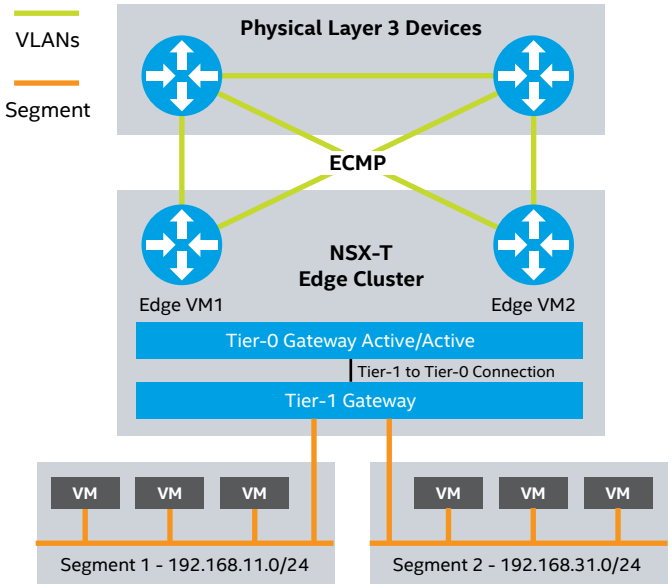


Figure 13. Sample two-node NSX-T Edge cluster in a single rack. All endpoints, ASNs, and DNS entries need to be provided by the user for the automatic deployment process.
Source: docs.vmware.com/en/VMware-Cloud-Foundation/4.2/vcf-admin/GUID-D17D0274-7764-43BD-8252-D9333CA7415A.html

Step 3: Deploying the Necessary Kubernetes Services

Step 3 of Workload Domain deployment is a three-step process in itself and is described in the next section.

Enabling vSphere with Tanzu (Creating the Supervisor Cluster)

To create a Kubernetes control plane inside the hypervisor layer, it is necessary to enable vSphere with Kubernetes on the Workload Domain cluster. This control plane is called the Supervisor Cluster and contains specific objects that enable the capability to run Kubernetes workloads within ESXi. Once this is done, the TKG Service (self-service lifecycle management of Tanzu Kubernetes cluster) is used to create and manage Tanzu Kubernetes clusters within the Supervisor Cluster. This is done in a declarative manner, similar to the standard Kubernetes process that operators and developers are familiar with.

To configure vSphere with Tanzu on a vSphere cluster, the environment must meet specific networking and infrastructure requirements. There are multiple areas that need to be addressed; consult the System Requirements and Topologies for Setting Up a Supervisor Cluster with NSX-T Data Center document for details. Because the Workload Domain cluster was created automatically as part of VMware Cloud Foundation, most requirements are already met.

With Kubernetes workload management, you validate the underlying infrastructure for vSphere with Tanzu and then complete the deployment in vSphere. The first step is performed using SDDC Manager, the second step is executed from vCenter. The entire process is described in the Enable Workload Management document.

Step 3.1: Kubernetes Workload Management Environment Validation

This step checks the environment compatibility for Kubernetes workload management. You must provide a cluster with a working NSX-T Edge cluster as well as DNS and NTP servers. Validations that will be performed include vCenter, Network, and Kubernetes workload management compatibility, including a check for licenses: a proper "vSphere with Tanzu" license must be applied to all hosts in the cluster. After a successful environment validation, you will be presented with a "Complete in vSphere" button that will redirect you to the Kubernetes Workload Management page in the vSphere user interface. By default, the vSphere Center for the VI Workload Domain where the Kubernetes workload management is to be deployed is selected.

Step 3.2: Deploy the Supervisor Cluster

Now that the environment is validated for Supervisor Cluster use, it can be deployed. This is accomplished from the vCenter ➔ Workload Management menu selection. Apart from the input data provided in the validation step, you must also provide several IP addresses and subnets: five IPs for the Kubernetes control plane VMs, non-routable Pod and Service subnets, and routable Ingress and Egress subnets.

Important: Despite the Pod and Service subnets being non-routable by design in Kubernetes, Intel's experience is that they are advertised by the T0 router using the default settings from the deployment procedure. Because of this, be sure to use different Pod and Service subnets for each Supervisor Cluster if you plan to have more than one (for example, having two Workload Domains, each with Kubernetes enabled).

You can also choose storage policies for placement of various components of the Kubernetes environment: control plane VMs, vSphere Pods, and cache for container images. The detailed procedure is available in the Enable Workload Management with NSX-T Data Center Networking document.

Step 3.3: Deploy the Tanzu Kubernetes Cluster

A Tanzu Kubernetes cluster is a distribution of the open-source Kubernetes container orchestration platform that is signed and supported by VMware. It can be provisioned on the Supervisor Cluster by using the TKG Service declarative API using the kubectl CLI and a cluster specification defined in YAML (consult the Configuration Parameters for Tanzu Kubernetes Clusters for a complete list of possible parameters). The complete documentation on how to deploy and use Tanzu Kubernetes clusters is available in the vSphere with Tanzu Configuration and Management document.

To use the TKG Service, Workload Management with NSX-T Data Center Networking needs to be enabled. Refer to the details provided in Prerequisites for Configuring vSphere with Tanzu on a Cluster and the Enable Workload Management with NSX-T Data Center Networking documents, followed by the Create and Configure a vSphere Namespace steps. Make sure to add a Content Library that will store Tanzu Kubernetes releases for use with Tanzu Kubernetes clusters. You will also

need to create Storage Policy prior to deploying the initial Workload Management. You may optionally enable a Private Image Registry (Harbor) that will be used by the Supervisor Cluster. You create it from the vSphere client. Follow the Enable the Embedded Harbor Registry on the Supervisor Cluster guide for details. You may deploy Pods, Services, and Deployments on the Supervisor Cluster; however, bear in mind that this cluster's intended usage is for administering the Kubernetes environment. It is advised that tenant workloads be run on guest Tanzu Kubernetes clusters.

Once the namespace is created, install Kubernetes CLI Tools for vSphere, connect to the Supervisor Cluster to define specifications for a Tanzu Kubernetes cluster, create configuration YAMLs (see the example files and complete list of parameters guides), and finally provision the Tanzu Kubernetes cluster.

Follow the entire workflow with details using the Provisioning Tanzu Kubernetes Clusters document. It includes step-by-step instructions for all the tasks mentioned above, along with the necessary links.

All Tanzu Kubernetes clusters that are provisioned by the TKG Service have the PodSecurityPolicy Admission Controller enabled. This means that a pod security policy is required to deploy workloads. There are two default pod security policies out-of-the-box, but you can create your own. To create Deployments, StatefulSets, and DaemonSets in the default namespace, a binding to one of the pod security policies needs to be created. Follow the Using Pod Security Policies with Tanzu Kubernetes Cluster document for the necessary steps.

To learn more about how to deploy workloads, read the Deploying Workloads to Tanzu Kubernetes Clusters guide.

This concludes the environment setup. If you need to deploy multiple Workload Domains with Kubernetes, repeat the above steps starting from the VMware Cloud Foundation Workload Domain Deployment section.

VMware Cloud on AWS Configuration

This section describes the components and steps needed to bring up VMC and connect it to the on-premises environment.

Creation of SDDC Manager

The first step to bring up the cloud environment is to deploy an SDDC from the VMware Cloud console. This process is simple and requires selecting an AWS region where the service should be located and choosing deployment options (one or more hosts within the SDDC with the option of a stretched cluster), host type, and name of the SDDC. You also must connect the newly created SDDC to an AWS account (within 14 days). The entire process is explained in detail in the Deploy an SDDC from the VMC Console document. After finishing the entire process, you will have a complete vCenter environment with a compute cluster ready. Be sure to add the necessary firewall rules in the VMware Cloud Console SDDC settings after deployment; by default, all traffic is blocked, and you won't be able to use your SDDC environment without changing these rules.

VPN Configuration between VMC and SD-WAN Orchestrator

A dedicated connection is needed to access an SDDC securely from the on-premises services. Achieve this by establishing a VPN connection. We configured an IPsec VPN between VMC and SD-WAN Orchestrator (see the Non-SD-WAN Destinations section). The process is relatively easy to configure from the SDDC side, but detailed configuration is needed to match the settings required by the Orchestrator side. Each environment is different and requires additional configuration to prepare the tunnel endpoint, routing, and firewall rules. The type of the on-premises tunnel endpoint defines the exact settings that need to be set for the tunnel to be established, and both ends of the tunnel must match. For the Non-SD-WAN Destination type of service selected on the Orchestrator side, we used a policy-based VPN on the VMC, but depending on your personal needs and environment, you can use a route-based VPN and BGP instead. End users can also connect to VMC without a VPN, but it is less secure than having a VPN or DX in place as a prerequisite to using some of the more advanced features that come with hybrid/multicloud.

For step-by-step VPN configuration information, read the VMware Cloud on AWS: Connecting with VPN article and the Configure a VPN Connection Between Your SDDC and On-Premises Data Center Procedure on VMware Docs.

Hybrid Linked Mode Enablement

For ease of manageability, the Hybrid Linked Mode can be configured. It enables administration of both on-premises and cloud vCenter with single sign-on. It also centralizes the management into one place: an additional on-premises VM appliance from which the entire infrastructure is visible, as if connected to a single platform service controller. Detailed deployment information is available in Configuring Hybrid Linked Mode on VMware Docs.

Important: You must configure the VPN before the Hybrid Linked Mode because it is required that the cloud vCenter is reachable using its internal (non-public) cloud IP address, which is possible only when the VPN is configured.

VMware Cloud on Azure Configuration

This section describes the components and steps needed to bring up AVS and connect it to the on-premises environment. For complete information, visit the Azure-VMware page.

Create SDDC Manager

The first step to bring up AVS is to register the AVS resource provider with a proper subscription. The administrator can create a new resource of AVS by selecting it in the Marketplace tab of the Azure portal. It is necessary to enter values used during the deployment process like resource name, location, size and number of hosts, and address block for the private cloud. After a successful deployment of the private cloud, the next step is to connect it to the Azure Virtual Network with ExpressRoute. Administrators can use an existing virtual network gateway or create a new one.

The entire deployment process is explained in detail in the Deploy and configure Azure VMware Solution document.

Connect to the On-premises Environment

The AVS private cloud can be easily connected to the on-premises environment. For this purpose, use the ExpressRoute Global Reach feature. Before using this feature, some prerequisites need to be satisfied. ExpressRoute circuits are specific for the Azure subscription type. For more information read the Enable connectivity between ExpressRoute circuits in different Azure subscriptions article. Administrators should also ensure that all gateways, including the ExpressRoute provider's service, support a 4-byte Autonomous System Number (ASN).

Before peering the private cloud to the on-premises environment, it is necessary to create an authorization key, which is used to connect virtual network gateways to the ExpressRoute circuit and can be used for only one connection. To establish a connection between environments it is also required to allow necessary routes to propagate by proper firewall settings on both sides. For step-by-step configuration information, read the Peer on-premises environments to Azure VMware Solution document.

VMware Remote Clusters Configuration

This section describes the steps needed to bring up the VMware Remote Clusters feature.

Prerequisites

To use the Cloud Foundation Remote Clusters feature, the remote site must meet the certain prerequisites.

Administrators need to provide a proper network configuration for the remote site that includes the configuration of separate VLANs for Management, vSAN, vMotion, and host overlay traffic. For NSX-T deployment it is also necessary to create private networks for Edge uplinks and overlay.

The most important step in Remote Cluster deployment is to ensure connectivity between the network at the Cloud Foundation Remote Clusters site and the management network at the central site. This connectivity can be achieved using static routing, but we recommend configuring BGP dynamic routing.

Adding a Remote Workload Domain

The steps required to create a remote Workload Domain are similar to adding a Workload Domain for a central site. For more details, see the VMware Cloud Foundation Workload Domain Deployment section.

VMware Tanzu Mission Control

This section provides insight on configuration of VMware Tanzu Mission Control. This service allows centralized management of all Tanzu Kubernetes clusters running in the environment.

Tanzu Mission Control is available as one of the VMware Cloud Services and requires no specific deployment steps—the user has access to the management interface once the service is enabled on VMC or AVS. Follow the steps in the Getting Started with VMware Tanzu Mission Control document. Once logged in to the Tanzu Mission Control console, the user should create at least one Cluster Group to organize the clusters; the Create a Cluster Group procedure describes this process.

With a Cluster Group in place, the user can start the process of attaching Kubernetes clusters. This process consists of two separate steps. The first step is to generate a custom YAML manifest file, specific to a given cluster. The second step is to use the autogenerated kubectl command directly on the cluster that is about to be attached. It runs a small set of extensions on the cluster to connect it with the cluster agent service. Once the extensions are up and running, the cluster becomes visible from the Tanzu Mission Control console. The entire procedure is described in the Attach an Existing Cluster document. This procedure needs to be performed for each Kubernetes cluster separately.

Revision History

Document Number	Revision Number	Description	Date
	1.0	Final version – VMware Cloud Foundation 4.2	9/24/2021

[1] Intel, https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/data-center-ssds/optane-ssd-p5800x-p5801x-brief.html

[2] Source: Claims [2] and [15] at Intel. "Intel Optane SSD P5800X Series – Performance Index." edc.intel.com/content/www/us/en/products/performance/benchmarks/intel-optane-ssd-p5800x-series

[3] Intel, https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/data-center-ssds/d7-p5600-p5500-series-brief.html

[4] Source: Claim [1] at https://edc.intel.com/content/www/us/en/products/performance/benchmarks/intel-optane-persistent-memory-200-series

[5] Source: Claim [118] at https://edc.intel.com/content/www/us/en/products/performance/benchmarks/3rd-generation-intel-xeon-scalable-processors

[6] "New SASE Report from Gartner." https://start.paloaltonetworks.com/gartner-report-roadmap-for-sase-convergence.html

[7] VMware, "What is SASE?" https://www.vmware.com/products/secure-access-service-edge-sase.html

[8] VMware, "Intel QAT Support for IPSec VPN Bulk Cryptography." docs.vmware.com/en/VMware-NSX-T-Data-Center/3.0/installation/GUID-F807EC0C-3417-4E89-BB90-F2F58AA34ECE.html