Dell EMC VxBlock™ Systems for VMware NSX-T Data Center

Architecture Overview

Document revision 1.0
May 2020

DELL EMC

Dell EMC
Hopkinton, Massachusetts 01748-9103
1-508-435-1000 In North America 1-866-464-7381
www.DellEMC.com

Contents

Revision history

Date	Document revision	Description of changes
May 2020	1.0	Initial release

Introduction

This document describes the high-level design of VMware NSX-T Data Center network virtualization technologies for stand-alone VxBlock Systems.

This document covers VMware NSX-T Data Center with VMware vSphere running on Cisco UCS C-Series Rack Servers for the physical edge host VMware vSphere cluster. See the *Release Certification Matrix* for more information about supported hardware and software with VMware NSX-T Data Center.

The target audience for this document includes Dell EMC sales engineers, field consultants, and advanced services specialists. Use this document to deploy a virtualized infrastructure using VMware NSX-T Data Center on VxBlock Systems.

This document is not meant to provide an exhaustive reference guide to the VMware NSX-T Data Center. The purpose is to educate the Dell EMC community and Dell EMC customers about the specifics of the VMware NSX-T Data Center design for VxBlock Systems. VMware has published a document titled *NSX-T Reference Design Guide Version 2.0* which provides information about how all aspects of VMware NSX-T Data Center work.

Dell EMC architected and engineered VMware NSX Data Center for VxBlock Systems (VMware NSX-T) where VMware NSX-V is not currently deployed. Although VMware NSX-T has many features and supports other hypervisors, the architecture in this guide is limited to VMware vSphere hypervisor only and is only for VxBlock Systems. The architecture from Dell EMC is currently for single site deployment of VMware NSX-T. For additional use cases (such as multi-site), contact Consulting Services. Dell EMC supports the VMware NSX-T architecture explained in this guide. Support for customizations and additional VMware NSX-T features is provided by VMware.

The Glossary provides related terms, definitions, and acronyms.

Related information

NSX-T Reference Design Guide Version 2.0

VMware NSX-T Data Center network virtualization

VMware NSX-T Data Center network virtualization is part of the software-defined data center that offers cloud computing across several platforms.

Platforms include VMware virtualization technologies, bare-metal workloads, Kubernetes-managed container-based workloads, and public cloud. VMware NSX-T Data Center expands on the NSX-V Data Center product by decoupling the management interface and network virtualization capabilities from VMware vCenter Server.

With VMware NSX-T Data Center, virtual networks are programmatically provisioned and managed independent of the underlying hardware. VMware NSX-T Data Center reproduces the entire network model in software, enabling a network topology to be created and provisioned in seconds. Network virtualization abstracts L2 switching and L3 routing operations from the underlying hardware, just as server virtualization does for processing power and operating systems.

Terminology changes

VMware NSX-T Data Center and VMware NSX-V use different terms for similar concepts.

The following table lists some of the terminology changes between the two products:

VMware NSX-T Data Center term	VMware NSX-V term
NSX-managed Virtual Distributed Switch (N-VDS)	Virtual Distributed Switch (VDS)
Segment	Port Group/VNI
Generic Network Virtualization Encapsulation (GENEVE)	VXLAN
Tier 0 Gateway	Edge Services Gateway (ESG)
Tier 1 Gateway	Distributed Logical Router (DLR)
Transport Node A transport node can also be a non-VMware vSphere ESXi operating system that participates in the N-VDS.	VMware vSphere ESXi Server

For a full explanation of the terms and concepts that are associated with VMware NSX-T Data Center, see the *NSX-T Reference Design Guide Versions 2.0*.

Related information

NSX-T Reference Design Guide Version 2.0

Cluster summary

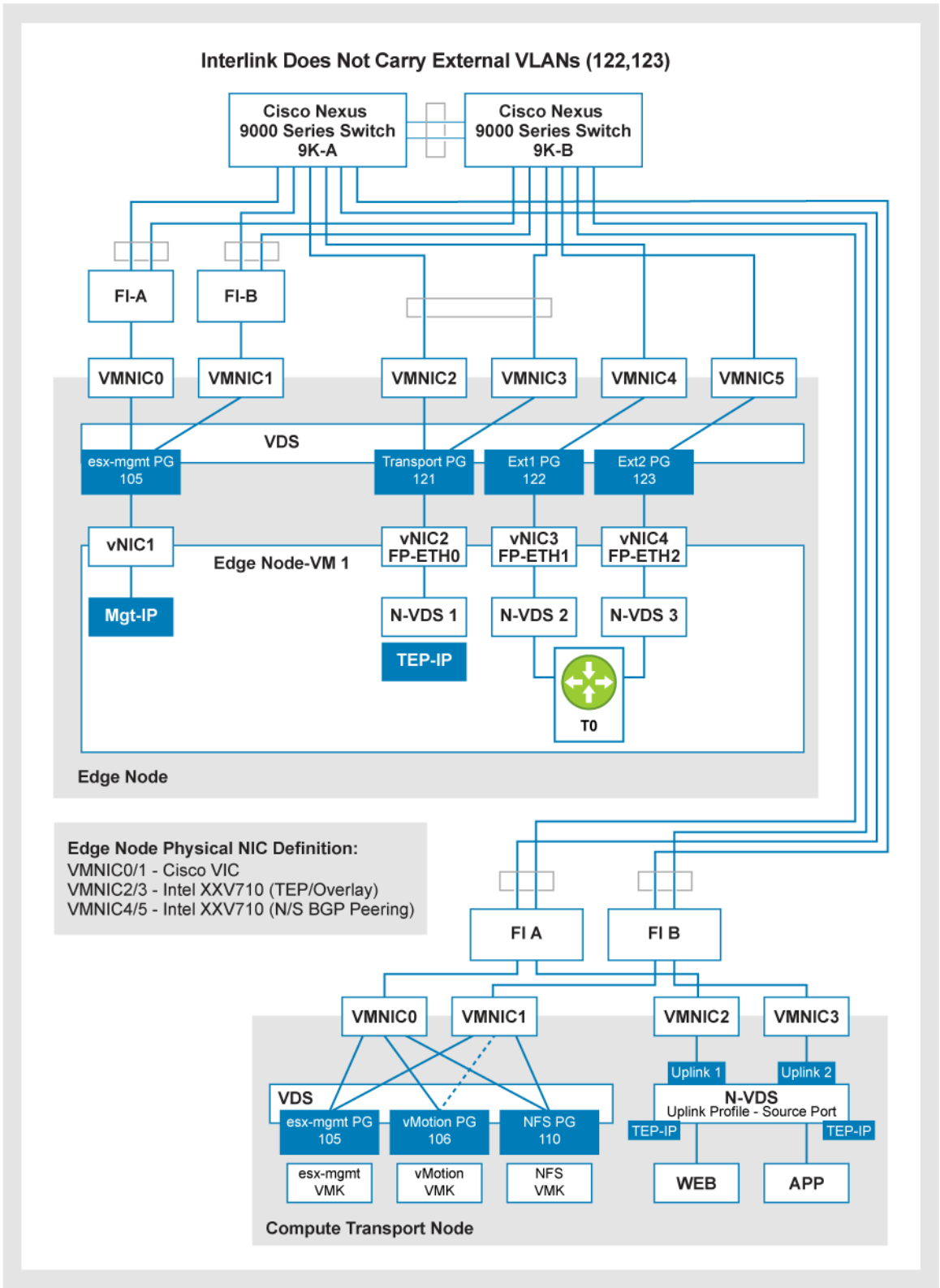
Provides a summary of VMware vSphere cluster information for VMware NSX-T Data Center.

The following table describes the VMware vSphere clusters for VMware NSX-T Data Center on the VxBlock Systems:

Cluster	Description
Management	Includes the VMware NSX-T Data Center manager appliance cluster that resides in the AMP The cluster includes AMP-VX, AMP-3S, or AMP-Central. The VMware NSX-T Data Center manager appliance cluster consists of three VMs that host the management plane and control plane components of the VMware NSX-T Data Center system. The management cluster also includes the VMware vCenter Server, which enables the VMware NSX-T Data Center manager appliance cluster to be deployed into the management cluster.
Physical Edge	Includes the physical edge Cisco UCS C220 servers and the edge node virtual machines that provide external connectivity to the physical network and various network services
Transport node	Includes production compute hosts and the production VMs There can be more than one transport node cluster.

VMware NSX-T Data Center logical topology

Use this figure to understand the logical topology of VMware NSX-T Data Center on a VxBlock System.



Cisco Nexus 9000 Series Switch requirements

When you deploy VMware NSX-T Data Center on a VxBlock System, some requirements apply to the ToR switches.

Ensure that the following requirements are met:

- Layer 3 licensing is present on the switches.
- The NSX-specific VLANs are deployed.
- The BGP routing protocol is enabled and configured.

VLANs specific to VMware NSX-T Data Center

VMware NSX-T Data Center on VxBlock Systems requires three additional VLANs for the ToR switch pair.

vcesys-nsx-transport

The vcesys-nsx-transport VLAN carries all east-west data flows for overlay traffic between transport nodes.

For VMware NSX-T Data Center on VxBlock System deployments, this network is defined as Layer 3 routable. If the use case requires transport nodes that are spread across multiple ToR switch pairs, the boundary between Layer 2 and Layer 3 is at the ToR switch pair. A VLAN cannot span that boundary. Examples include:

- A single-site multisystem deployment of VMware NSX-T Data Center
- A future multisite deployment of VMware NSX-T Data Center

In these use cases, each L2 domain (ToR switch pair) must provision a unique subnet to hold the TEP IP addresses for the connected transport nodes. The transport nodes for each transport network subnet need to reach the other transport network subnets in the VMware NSX-T Data Center deployment.

A routable transport network ensures that customer deployments are flexible and extensible for all current and future use cases.

VLAN	Description
vcesys-nsx-edge01	This VLAN enables BGP peering between the T0 Gateway in the edge VM and the Cisco Nexus 9000 Series ToR switch on the A side of the network fabric.
vcesys-nsx-edge02	This VLAN enables BGP peering between the T0 Gateway in the edge VM and the Cisco Nexus 9000 Series ToR switch on the B side of the network fabric.

BGP routing configuration

VMware NSX-T Data Center on VxBlock Systems requires that you enable the BGP on the ToR switches. Also, a BGP routing process must be created on the ToR switches and associated with the appropriate Autonomous System (AS) identifier.

VMware recommends the BGP dynamic routing protocol for peering T0 gateways to the physical network. If a different routing protocol such as OSPF is used for VxBlock System customer uplinks, the BGP routes from VMware NSX-T Data Center are redistributed into OSPF.

VMware NSX-T Data Center management cluster

The management cluster consists of the management and control planes for VMware NSX-T Data Center. The VMware NSX-T Data Center manager appliance handles both management plane and control plane functions.

VxBlock System management (AMP) cluster components

The VMware NSX-T Data Center management appliance cluster on AMP-VX, AMP-3S, or AMP-Central previously consisted of three VMware NSX-T Data Center controller appliances and a single VMware NSX-T Data Center management appliance. Beginning with VMware NSX-T Data Center version 2.4, there is no separate controller cluster. A unified manager appliance performs management and controller functions. Three unified appliances are deployed to ensure high availability for the control plane. A virtual IP address is assigned to the cluster upon deployment to ensure ease of access to the VMware NSX-T Data Center HTML5 user interface.

VxBlock System management (AMP) cluster specifications

NSX-V and NSX-T Data Center configuration values are compared.

Use the specifications in the following table for the VMware NSX-T Data Center manager appliance cluster.

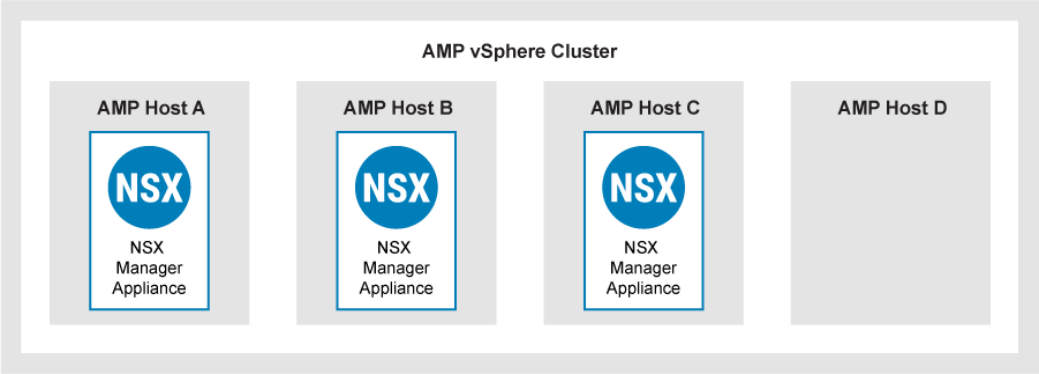
Specification	VMware NSX-V manager node	VMware NSX-T Data Center manager node
Quantity	One VM	Three VMs per VMware NSX-T Data Center manager cluster A single VMware NSX-T Data Center manager cluster can manage transport nodes that are connected to: • Up to eight VMware vCenter Servers • Other compute managers like KVM or Kubernetes
Location	Management cluster	Management cluster
Hardware	AMP-VX, AMP-3S, or AMP Central with a minimum of four servers	AMP-VX, AMP-3S, or AMP Central with a minimum of four servers
Size	Large Each appliance consumes: • 8 vCPU • 32 GB RAM • 200 GB disk	Large The three-node cluster consumes: • 24 vCPU • 96 GB RAM • 600 GB disk
Network	vcesys_esx_mgmt (VLAN 105)	vcesys_esx_mgmt (VLAN 105)
Availability	VMware HA	VMware HA
Distribution	The OVA is deployed using CLI commands or the VMware vSphere GUI.	The first node is deployed as part of the OVA deployment. Other nodes are deployed from the VMware NSX-T Data Center manager that runs on the first node.

AMP hardware requirements

VxBlock Systems support VMware NSX-T Data Center virtual networking with AMP-VX, AMP-3S, or AMP Central with a minimum of four servers.

When sizing the solution, ensure that the total AMP workload is compatible with the number of nodes being ordered. In some cases, four AMP nodes may not be able to handle the added workload for the VMware NSX-T Data Center management cluster. In this case, more nodes must be ordered. Ensure that there is sufficient storage available for the VMware NSX-T Data Center manager VMs. If necessary, upgrade the Dell EMC Unity storage array in the AMP to the larger option. No special cabling is required.

The following figure shows the layout of the VMware NSX-T Data Center manager cluster that is installed across a four-node AMP cluster:



The configuration requires at least one AMP node in the cluster that does not host a VMware NSX-T Data Center manager appliance under normal operation. This requirement ensures sufficient capacity to perform maintenance on the AMP cluster without degrading the VMware NSX-T Data Center manager appliance cluster.

VMware vSphere AMP cluster requirements

The management cluster requires VMware HA and VMware vSphere Distributed Resource Scheduler (DRS) for two reasons. The first is to provide VM protection against a VMware vSphere ESXi host failure. The second is to balance VM workloads in the cluster.

The following rules are applied to the DRS:

- Anti-affinity rules are applied to the management cluster to ensure that each VMware NSX-T Data Center manager appliance runs on its own host where possible.
- Too few hosts may prevent each VMware NSX-T Data Center manager appliance to run on its own host. If so, HA allows the VMware NSX-T Data Center manager appliances to co-exist on the same host.

AMP custom resource pool requirements

The VMware NSX-T Data Center management cluster does not require custom resource pools. However, for heavy workloads, create memory reservations for the VMware NSX-T Data Center manager. Configure the AMP with

sufficient resources so that there is no competition for resources among the workloads running on the VMware NSX-T Data Center manager appliance.

AMP storage requirements

The management cluster does not require a specific disk layout other than the standard disk layout of the AMP-VX, AMP-3S, or AMP-Central.

The VMware vSphere ESXi hosts that are connected to the management cluster use the AMP storage array. The VMware NSX-T Data Center manager appliances are deployed across separate data stores where possible to protect against LUN corruption and improve performance and resilience.

VMware NSX-T Data Center manager appliances require a disk latency that is less than 10 milliseconds.

AMP networking requirements

There are no special network requirements for the AMP-VX, AMP-3S, or AMP-Central. The VMware NSX-T Data Center management traffic, control plane traffic, and VMware vSphere ESXi management traffic share the same network segment to improve performance. A network latency value of less than 10 milliseconds is required between the VMware NSX-T Data Center manager appliances. All AMPs meet or exceed this requirement.

VMware NSX-T Data Center physical edge cluster

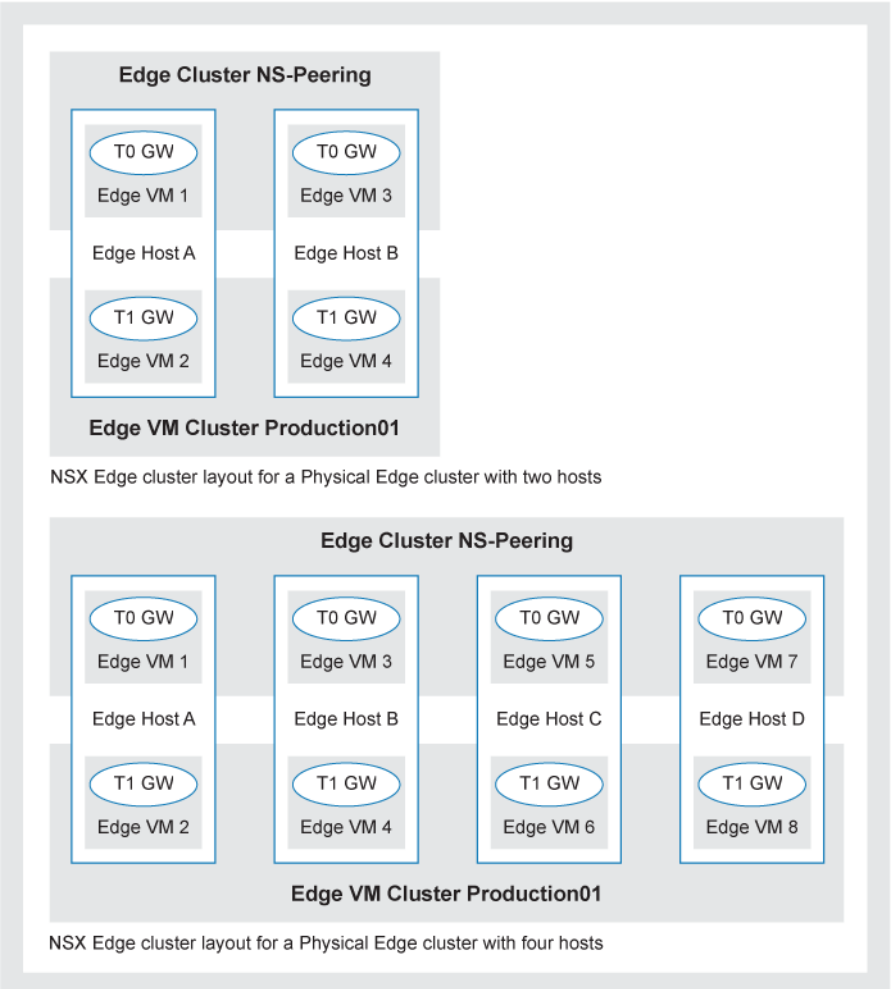
The VMware NSX-T Data Center physical edge cluster connects to the physical network. The cluster provides the physical platform for the network services the edge node virtual machines provide. These services include routing, bridging, and other network services. The Cisco UCS C-Series Rack Servers host the physical edge cluster.

VMware NSX-T Data Center physical edge cluster components

The VMware NSX-T Data Center physical edge cluster consists of an even number of physical servers to host edge VMs.

During initial deployment, the two VMware NSX-T Data Center logical edge clusters are built on the first two or four physical edge hosts in the cluster. Create host affinity rules in vSphere to ensure that this configuration is maintained in all situations except a component failure.

The following figure illustrates how the VMs are configured:



If the initial deployment includes more than four physical edge hosts, only the first four hosts are populated with edge VMs. Use the rest of the hosts as needed.

VMware NSX-T Data Center physical edge cluster specifications

NS-peering cluster and the Production01 cluster configuration values are compared.

The following table summarizes the VMware NSX-T Data Center edge cluster specifications:

Specification	Edge VMs NS-peering cluster	Edge VMs Production01 cluster
Quantity	Two or four VMs active/active with ECMP	Two or four VMs active/active with ECMP
Location	Physical edge host cluster	Physical edge host cluster
Hardware	Cisco UCS C-Series Rack Servers	Cisco UCS C-Series Rack Servers
Size	Medium 4 vCPUs 8 GB RAM 200 GB disk	Large 8 vCPUs 32 GB RAM 200 GB disk
Network	Four vmnic interfaces: • Appliance management on the esx_mgmt VLAN • Transport (East-West / TEP) traffic on the nsx_transport VLAN • Fabric A edge (North-South) traffic on nsx_edge_fabric_a VLAN • Fabric B edge (North-South) traffic on nsx_edge_fabric_b VLAN	Four vmnic interfaces: • Appliance management on the esx_mgmt VLAN • Transport (East-West / TEP) traffic on the nsx_transport VLAN • Fabric A edge (North-South) traffic on nsx_edge_fabric_a VLAN • Fabric B edge (North-South) traffic on nsx_edge_fabric_b VLAN
Availability	VMware HA or DRS in partially automated mode	VMware HA or DRS in partially automated mode
Distribution	Affinity rules are enabled to place the edge VMs to match the layout in the preceding figure.	Affinity rules are enabled to place the edge VMs to match the layout in the preceding figure.

Hardware requirements

This section provides the hardware requirements for the Cisco UCS C-Series Rack Servers.

VMware NSX-T Data Center hardware requirements

The selection and sizing of the hardware platform for the VMware NSX-T Data Center design considers the VMware system requirements for the various components.

See the VMware *System Requirements*.

The use cases for VMware NSX-T Data Center are broad and are customizable for many business use cases. Dell EMC provides two engineered sizes of physical edge hosts. The standard host has 96 GB of memory and supports most use cases. The extra large host has 192 GB of memory. It supports use cases that require many large edge node VMs such as SSL, VPN, and load-balancing today. The large host size also supports upcoming features that have larger resource requirements.

Related information

System Requirements

Cisco UCS C220 M5 component configuration overview

Component values for the Cisco UCS C220 M5 based edge node are presented.

See the following table for the component values for the Cisco UCS C220 M5 based edge node.

Component	Cisco UCS C220 M5 based edge node
CPU	Two Intel Xeon 5218 2.3 GHz, 16 core, 22 MB cache
Memory	96 GB (6 x 16 GB DDR4-2933) standard-size host or 192 GB (12 x 16 GB DDR4-2933) extra-large-size host
NIC	Two Intel XXV710-DA2, dual-port 25 Gbps PCIe adapter
VIC	Cisco VIC 1457 (4 x 10/25 Gbps SFP28 mLOM)
Storage	The Cisco UCS primary storage array provides the storage. Valid primary storage arrays are VMAX, PowerMax, Dell EMC Unity, and XtremIO X2.
Boot device	SD or boot from SAN

CPU

For VMware NSX-T Data Center edge workloads, the optimal configuration has enough CPU cores to prevent oversubscription. Having enough CPU cores provides maximum performance to the VMware NSX-T Data Center edge-node VMs that carry the workload on these nodes. The current VMware NSX-T Data Center design includes one large and one medium edge node VM on each of the first four physical edge hosts in the edge cluster.

The current design uses the Intel Xeon 5218 CPU, which has 16 cores running at 2.3 GHz. Two CPU sockets are populated per server. The Cisco UCS C220 M5 server requires a dual-socket configuration to enable the second PCIe slot that is used for the second Intel NIC. This configuration provides 32 physical cores per edge host.

The following table shows the vCPU consumption for a default configuration of the physical edge host:

Component	Quantity	CPU requirement	Notes
VMware NSX-T Data Center edge node, size medium	1	4 vCPU	This edge node is used for the NS peering edge cluster, which is used for the BGP peering between VMware NSX-T Data Center and the physical network.
VMware NSX-T Data Center edge node, size large	1	8 vCPU	This edge node is used for the Production01 edge cluster, providing T1 routing capabilities and services for customer applications.
Total used (no oversubscription)		12 cores	
Free		20 cores	Free represents the available capacity for customer-defined edge requirements.

This configuration consumes only 12 of the 32 available CPU cores in a physical edge host. The remaining capacity is available for customization. Use the additional CPU cores to deploy additional edge nodes or increase the size of the default edge nodes.

Memory

The Cisco UCS C220 M5 physical edge host uses 6 or 12 16 GB DDR4-2933 RDIMMs. These RDIMMs provide 96 GB or 192 GB of RAM.

The default deployment of a physical edge host allocates system memory as described in the following table:

Component	Quantity	Memory requirement (standard host)	Memory requirement (large host)	Notes
VMware NSX-T Data Center edge node VM, size medium	1	8 GB	8 GB	This edge node is used for the NS-Peering edge cluster, which is used for the BGP peering between VMware NSX-T Data Center and the physical network.
VMware NSX-T Data Center edge node VM, size large	1	32 GB	32 GB	This edge node is used for the Production01 edge cluster, providing T1 routing capabilities and services for customer applications.
VMware vSphere ESXi Hypervisor 6.5 or 6.7	1	4 GB	4 GB	This memory size represents the minimum requirement to run VMware vSphere ESXi.
Total used		44 GB	44 GB	
Free		52 GB	148 GB	This amount of memory is available for customer-defined edge requirements or for use for support or recovery from failures.

NICs

The choice of physical network interface cards in an VMware vSphere ESXi host is critical for maximizing performance. VMware NSX-T Data Center makes extensive use of protocol offloading to achieve line-rate performance.

The network interface card and its associated drivers for VMware vSphere ESXi must support the following:

- GENEVE offload
- Receive-side scaling (RSS)
- RX or TX queueing

The Intel XXV710-DA2 card provides dual-port SFP28 connectivity and can support both 10 Gbps and 25 Gbps transceivers. This card has a standard, eight-lane PCI Express interface. Because each lane has a bandwidth of 8 Gbps, the maximum PCIe throughput for the PCIe interface of a single card is 64 Gbps.

Cisco VIC

The Cisco VIC 1457 connects to FI using the Cisco SingleConnect technology.

Cisco SingleConnect provides the following services to the physical edge host:

- Enables Cisco UCS Manager to provide storage to the physical edge hosts from the VxBlock System primary storage arrays
- Provides connectivity to the VxBlock System VMware vSphere ESXi management and VMware vMotion networks
- Enables server configuration using Cisco UCS service profiles to simplify deployment

Connectivity model

This section provides information on the connectivity model for VMware NSX-T Data Center.

VIC to FI and FEX connectivity

A Cisco UCS VIC 1457 to FI connection uses either 10 Gbps direct-attach cables or SFP-10G-SR optics and LC cabling. Standardizing on 10 Gbps connectivity to the FI ensures compatibility with Cisco UCS Fourth-Generation domains. Connect ports 1 and 3 on the VIC to the FIs. Ports 2 and 4 on the VIC are unused. If two Cisco UCS FEXs are connected to the FI, FEX ports can be used to connect the VIC to the Cisco UCS infrastructure.

For physical edge hosts that are connected to a Cisco UCS Third-Generation FI model 6332-16UP, the hosts connect to QSFP-based server ports using optical breakouts. To configure the optical breakout, do the following:

- Connect a Cisco QSFP-40G-SR4 transceiver into a QSFP port on the FI.
- In Cisco UCS manager, configure the transceiver as a breakout port.

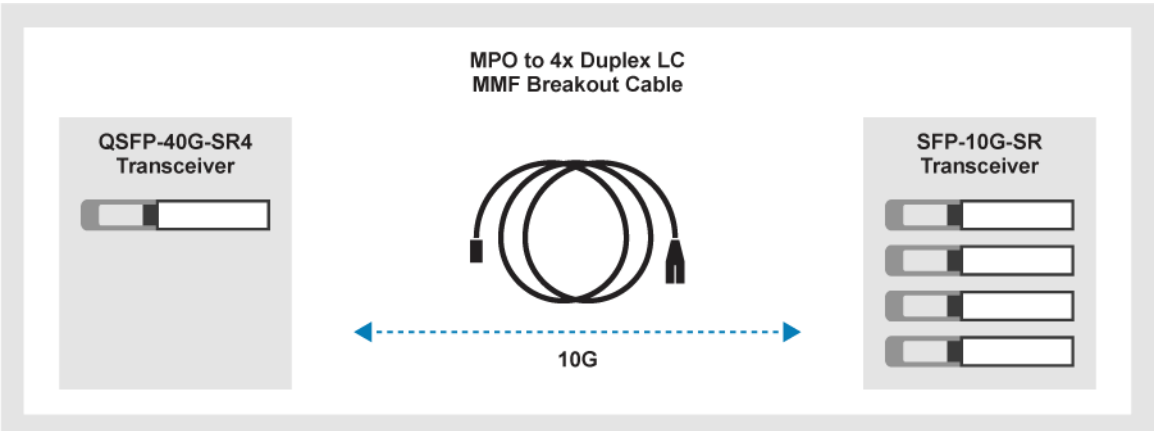
This configuration yields four 10 Gbps lanes that can be used for server connections. The number of QSFPs populated on the FI depends on the number of physical edge hosts in the cluster.

The VIC 1457 card in the server is populated with Cisco SFP-10G-SR transceivers. MPO hydra and trunk cabling connect the two together.

The following table shows the quantity of QSFP ports that are populated for various physical edge cluster sizes:

Cluster size (physical edge nodes)	QSFP ports populated on each Generation 3 FI
2 or 4	2
6 or 8	4
10 or 12	6
14 or 16	8

The 40 Gbps to 10 Gbps breakout connection is used when directly connecting a physical edge host to a UCS Generation 3 FI. The following figure shows the connection topology:

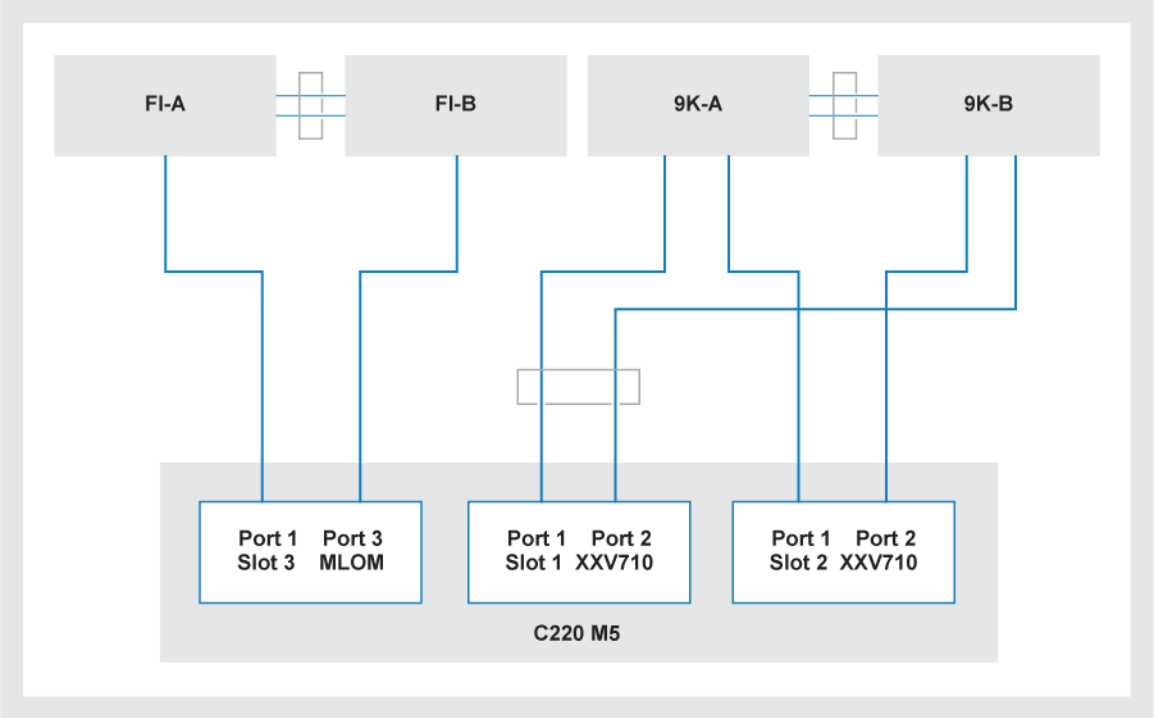


Intel NIC to ToR switch connectivity

Connect the Intel XXV710-DA2 card to a Cisco Nexus 9300 Series Switch with QSFP or QSFP28 ports. The Cisco PID for Cisco 25 Gbps SFP+ SR transceiver (SFP-25G-SR-S) is used in the Intel cards. Optical hydra cabling is used to connect to one of the four 25 Gbps channels on a Cisco QSFP-100G-SR4-S module in the switch.

Physical cabling between edge nodes and ToR switches

The following figure shows the physical topology from the Cisco UCS C220 M5 Server to Cisco FIs and to the Cisco Nexus 9000 Series ToR switches:

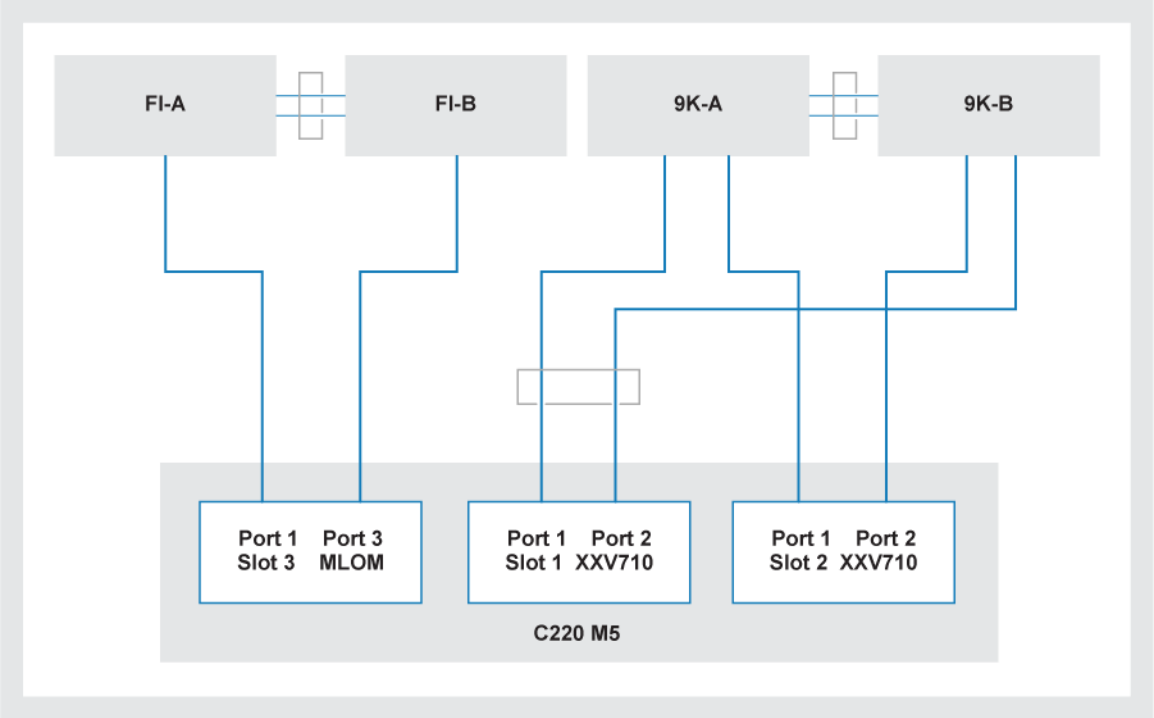


Intel NIC to ToR switch connectivity (using QSFP or QSFP28) for VxBlock Systems 340, 350, 540, and 740

Connect the Intel XXV710-DA2 card to a Cisco Nexus 93180YC-EX or 9396PX series ToR switch with SFP + ports. The Cisco PID for Cisco 10 Gb/s SFP + SR transceiver (SFP-10G-SR) is used in the Intel cards. LC Fiber cabling is used to connect to one of the 10 Gb/s interfaces on a Cisco SFP-10G-SR module in the switch.

Physical cabling between edge nodes and ToR switches

The following diagram shows the physical topology from the C220 M5 Server to Cisco fabric interconnects and to the VxBlock ToR switches:



Storage requirements

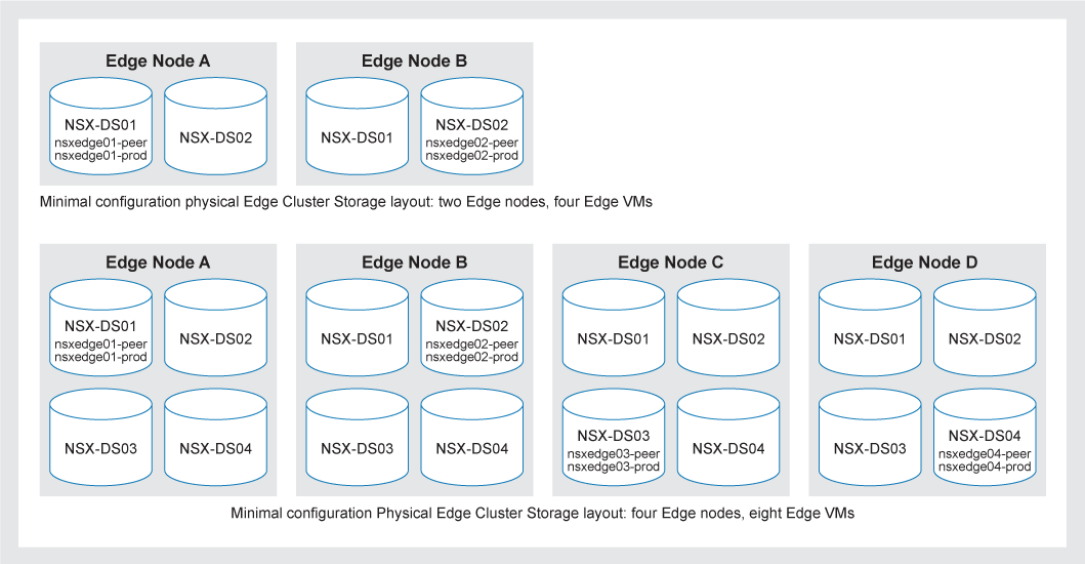
There are several requirements for the edge cluster.

The edge cluster has the following storage requirements:

- Data stores: The VMware vSphere ESXi hosts that are connected to the edge cluster use the XtremIO, Dell EMC Unity, VMAX, or PowerMax storage arrays. A 1.2 TB LUN is deployed for each physical edge server in the edge cluster. The edge node virtual machines that are associated with that host reside on this data store.
- All edge data stores are presented to all physical edge hosts in the cluster reducing recovery time after an HA event.
- The size of each edge node VM under VMware NSX-T Data Center is 200 GB, which is larger than the ESGs under NSX-V. A default deployment includes two edge node VMs on each VMware NSX-T Data Center physical edge host, which consumes 400 GB of storage. More storage space is left on the data store to support the deployment of additional edge node VMs to each edge node. The additional edge node VMs help to meet business requirements.
- Disk layout: No specific disk layout is necessary. VMware NSX-T Data Center supports all primary storage arrays that are available on the VxBlock System.

See *Network requirements for the Cisco UCS physical edge host servers*.

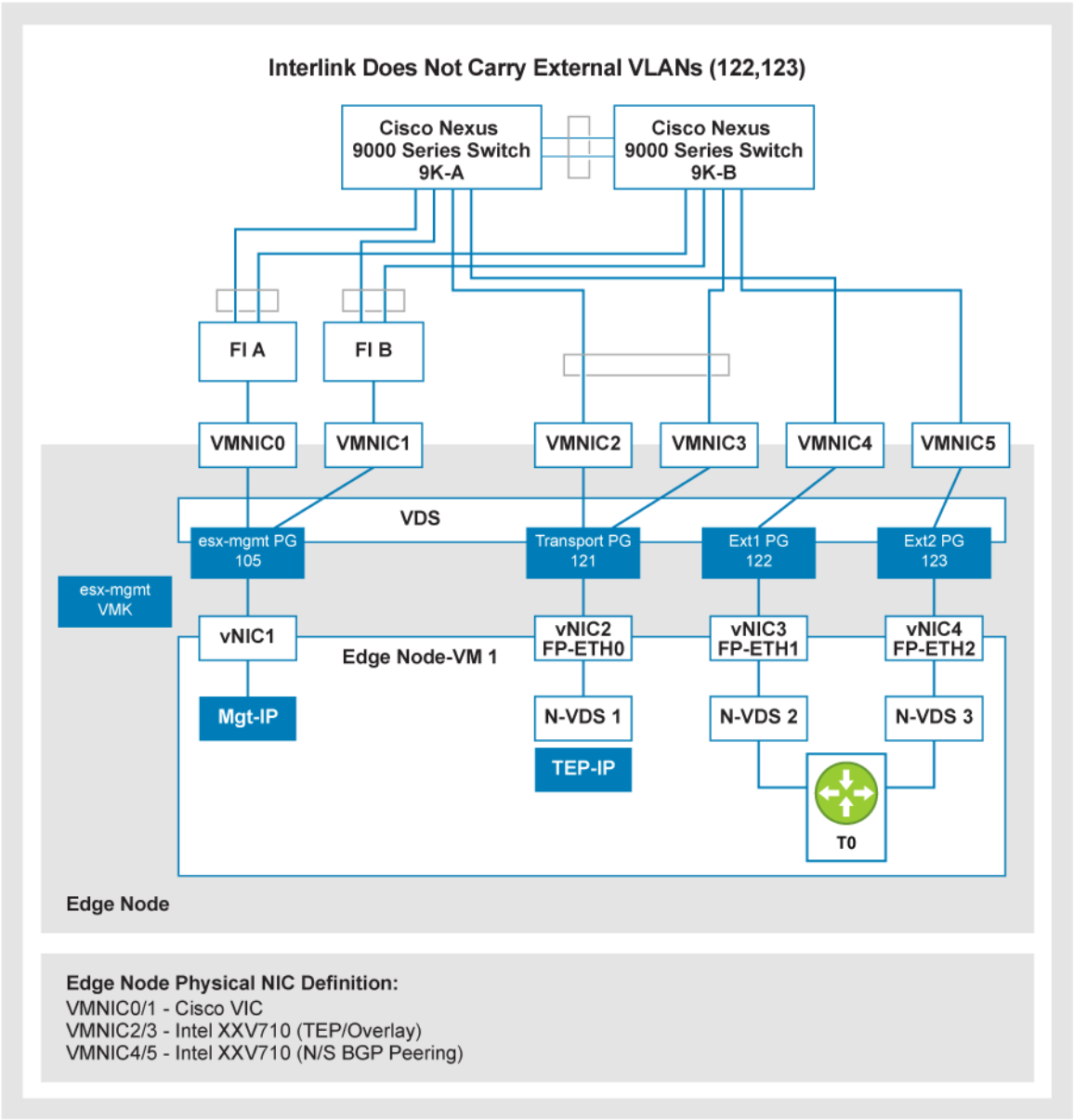
The figure below shows the physical edge hosts and their storage layout from the factory. The first configuration shows the layout for a minimal edge node configuration with two physical edge hosts. The second configuration shows the configuration for a four-node physical edge cluster. Any edge nodes above four are not configured in the factory. The customer must deploy the data stores for these hosts using the deployment model for the first four hosts.



Logical topology of the physical edge host

The physical edge host logical topology is presented.

The following figure shows the logical topology of the edge hosts in a VMware NSX-T Data Center configuration.



Network requirements for the Cisco UCS physical edge host servers

The network requirements for the Cisco UCS physical edge host servers are described.

The following table describes the network requirements for Cisco UCS C-Series Rack Servers.

Component	Cisco UCS C-Series Rack Servers (physical edge host)
VLAN IDs	The physical edge host cluster requires three VLAN SVIs on the Cisco Nexus 9300 Series Switches: • Two external edge VLAN SVIs are used for external traffic for North-South traffic flows. • One transport VLAN is used to pass overlay traffic for East-West traffic flows. With physical Edge host Cisco UCS C-Series Rack Servers, using the Cisco UCS Manager to create external edge traffic VLAN IDs is unnecessary. However, because the transport nodes pass overlay traffic, the transport VLAN ID must be added to the Cisco UCS Manager.

VMware virtual network for the physical edge hosts

For Cisco UCS C-Series Rack Server physical edge hosts, a single distributed virtual switch is created for the edge cluster.

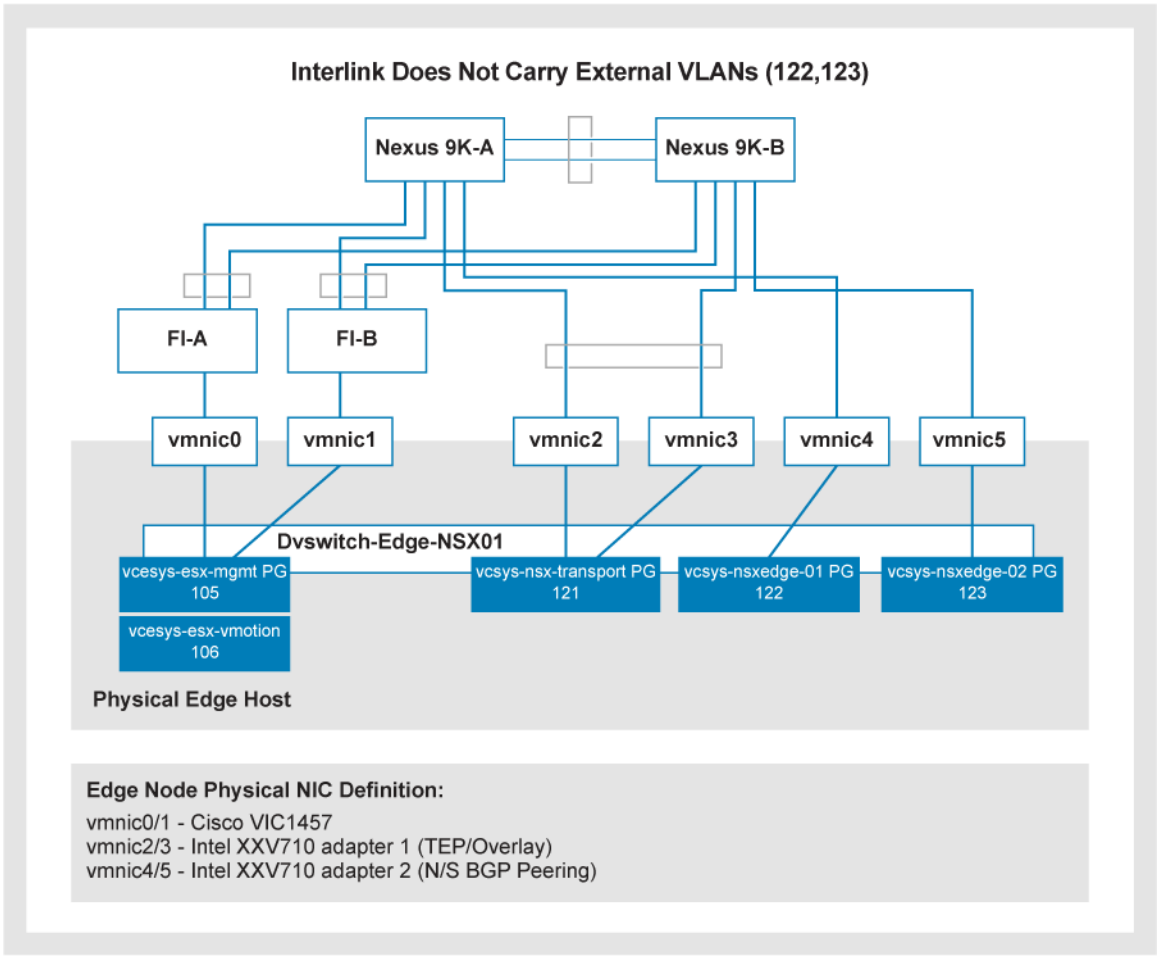
The VMware VDS uses the following uplinks:

- Two uplinks from the Cisco VIC 1457 mLOM adapter to the FI
- Four total uplinks, two from each of the two Intel XXV710-DA2 adapters to the ToR switches

The port groups are pinned to the appropriate uplinks for the traffic type in the VMware DVS. The following table describes the uplink pinning:

Distributed port group	VLAN ID	Description	Uplink pinning
vcesys-esx-mgmt	105 (default)	This VLAN provides VMware vSphere ESXi management VLAN access to the host.	Cisco VIC 1457 uplinks
vcesys-esx-vmotion	106 (default)	This VLAN provides the vSphere vMotion network to the host.	Cisco VIC 1457 uplinks
vcesys-nsx-transport	121 (default)	This VLAN carries transport traffic, which is the GENEVE-encapsulated traffic for all East-West traffic flows on overlay backed segment. On the physical edge cluster, this VLAN is used exclusively for traffic: • That arrives from a transport node destined to an edge service, or external network. That arrives from an edge service or external network destined to a transport node. • Must egress through the edge cluster to an endpoint outside of the VMware NSX-T Data Center deployment.	Intel XXV710-DA2 adapter 1 Uplinks 1 and 2 (VPC)
vcesys-nsx-edge01	122 (default)	This VLAN provides the network that is used on the A side of the network fabric for BGP peering northbound from the Edge node VM cluster to ToR Nexus 9000 series switch A in the VxBlock System. This VLAN is the egress point from VMware NSX-T Data Center to the rest of the customer network on the A side of the fabric.	Intel XXV710-DA2 adapter 2 Uplink 1
vcesys-nsx-edge02	123 (default)	This VLAN provides the network that is used on the B side of the network fabric for BGP peering northbound from the Edge node VM cluster to ToR Nexus 9000 series switch B in the VxBlock System. This VLAN is the egress point from VMware NSX-T Data Center to the rest of the customer network on the B side of the fabric.	Intel XXV710-DA2 adapter 2 Uplink 2

The following figure shows the uplink topology for a VMware NSX-T Data Center physical edge host:



Edge node cluster architecture standards

In VMware NSX-T Data Center, an edge node cluster is a group of edge nodes. The edge nodes can be deployed in a virtual (VM based) or physical (bare-metal) form factor. The initial VxBlock System 1000 implementation of VMware NSX-T Data Center supports only virtual edge clusters. This implementation provides some benefits for flexibility of deployment and serviceability over bare-metal edge nodes.

The VxBlock System VMware NSX-T Data Center design adopts the VMware recommended edge cluster design for service providers. This design includes two edge clusters:

- Edge-Cluster-NS-Peering is a dedicated edge cluster to host a Tier 0 router for BGP peering and north-south traffic flows. The traffic flows need to communicate from the VMware NSX-T Data Center environment to the physical environment and the outside world. The edge node VMs that make up this cluster are deployed as medium-sized appliances.
- Edge-Cluster-Production01 is a production edge cluster that tenants or business units can use. The edge node VMs that make up this cluster are deployed as large-sized appliances. A T1 gateway should be associated with this cluster only if the T1 gateway is hosting centralized services such as NAT or Edge Firewall.

This design provides role-based security. In a large IT shop or multi-tenant environment:

- One group can secure the NS-Peering edge cluster to VMware NSX-T Data Center administrators and support personnel.
- A different group can manage the Production01 edge cluster.

VMware NSX-T Data Center enables the definition of role-based access to these clusters independently, using the API.

This design also mitigates limitations in the BGP peering capabilities of the T0 gateway. The T0 gateway supports eight-way ECMP peering. The VxBlock System VMware NSX-T Data Center design connects a single T0 gateway in one of two ways in the edge-cluster-NS-peering cluster:

- Through two edge node VMs (4-way ECMP)
- Through four edge node VMs (8-way ECMP)

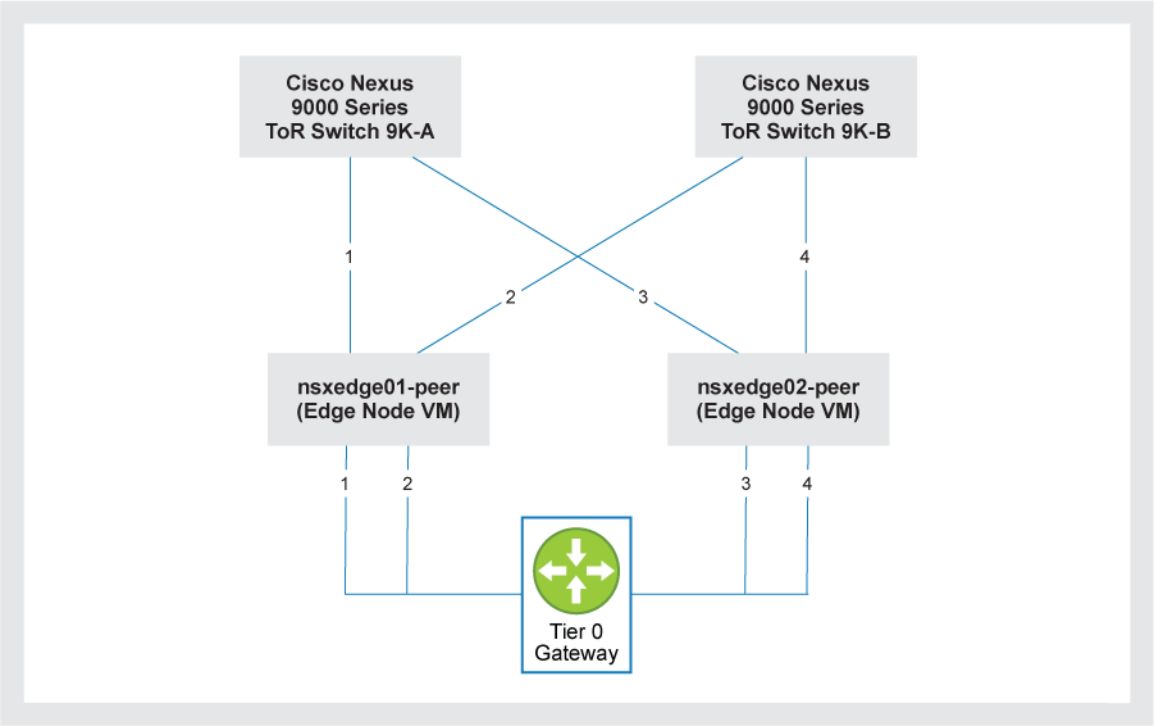
Depending on the need, deploy additional edge node VMs, add connectivity for the T0 gateway, or add a T0 gateway. Two ToR switches north of the edge cluster can support only four edge VMs peering north-bound to the ToR switches in a cluster. The Production01 edge cluster does not need to participate in ECMP peering with the ToR switches. If necessary, the cluster can contain more than four nodes.

Edge ECMP topology

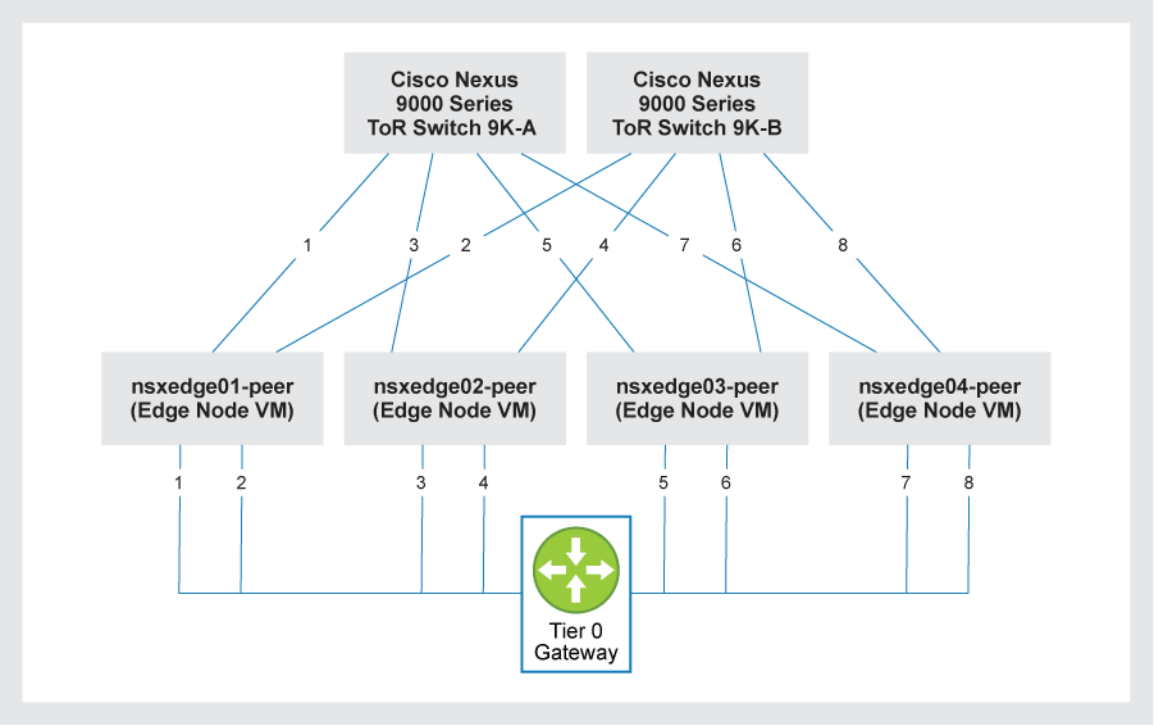
In VMware NSX-T Data Center for VxBlock Systems, ECMP provides load-balancing across multiple paths northbound from the T0s to the Cisco Nexus 9000 Series ToR switches. VMware NSX-T Data Center supports a maximum of eight paths per edge node cluster.

Each T0 gateway peers with each of the two Cisco Nexus 9000 Series ToR switches. This peering limits the maximum number of edge node VMs in the cluster to four.

The following figure shows a minimum edge cluster configuration with two edge node VMs peering with the ToR switches, consuming four ECMP paths:



The following figure shows a maximum edge node cluster configuration with four edge node VMs peering with the ToR switches, consuming eight ECMP paths:



This design peers each edge node with each ToR switch.

ECMP routing configuration

Each T0 gateway peers with each ToR switch using BGP, and traffic is distributed across the available edge node VMs using ECMP.

The ECMP implementation in VMware NSX-T Data Center limits the number of paths to eight.

In the VMware NSX-T Data Center for VxBlock Systems design, each Edge node VM peers with each of the two ToR switches. This pairing limits the maximum number of edge node VMs in the cluster to four. With two edge node VMs, the default configuration is a single T0 gateway with four interfaces. Two interfaces peer with Cisco Nexus 9000 Series Switch 9K-A and two interfaces peer with Cisco Nexus 9000 Series Switch 9K-B. This is the default configuration as delivered – each use case dictates the specific T0 configuration.

For more details on the ECMP design, see *VMware NSX-T Data Center physical edge cluster*.

Edge node VM resource usage and Data Plane Development Kit

The VMware NSX-T Data Center edge node VMs implement the Data Plane Development Kit (DPDK) standard to provide high-performance packet forwarding capabilities.

The DPDK reserves CPU cores to constantly poll the NIC for packets instead of waiting for them to be processed using an interrupt. The constant polling produces the following performance on edge mode VM vCPUs:

- Some of the vCPUs run at 50 percent of capacity under no load.
- Some of the vCPUs run at 100 percent of capacity under moderate load.

There may be a high CPU utilization level on edge node VMs that are operating properly.

NS-peering edge cluster

The engineered VxBlock System implementation of VMware NSX-T Data Center deploys medium-sized edge VMs for the NS-Peering edge cluster. This cluster is used exclusively for North-South traffic flows. It should not be running any services other than BGP peering and ECMP. The medium appliance meets these requirements.

Production01 edge cluster

The production01 edge cluster uses the large-sized edge VM. The production01 cluster hosts various services for the VMware NSX-T Data Center environment, including load balancing, VPN services, NAT, DHCP, and Edge Firewall. These services consume additional resources. Use of a large-sized edge VM is appropriate. This cluster should be used only in cases where VMware NSX-T Data Center Tier 1 centralized services are instantiated.

Edge cluster

The VMware NSX-T Data Center solution for VxBlock Systems includes two edge clusters. Optionally, deploy additional edge VM requirements after initial system deployment. All VMware supported configurations are allowed provided they do not interfere with operation of the original design. The physical edge nodes have sufficient resources to allow an additional large edge node or several small or medium edge nodes per host. Ensure that each of the hosts in the cluster has some reserve capacity. That way, maintenance and recovery operations can be completed as needed with no warnings or errors appearing in the VMware vCenter Server UI.

Segments

NSX-V included the concept of a VXLAN Network Identifier, or VNI. A VNI is the equivalent of a port group on a Distributed Virtual Switch (VDS) in VMware vSphere. However, it is located behind an NSX-V logical switch. VMware NSX-T Data Center uses GENEVE encapsulation instead of VXLAN, so the concept of a VNI is not relevant. Instead, VMware NSX-T Data Center uses a construct called an overlay backed segment.

VMware NSX-T Data Center also allows for VLAN backed segments, which are segments for which the traffic is not GENEVE encapsulated.

The VMware NSX-T Data Center for VxBlock System design uses segments for North-South and East-West traffic. The segments are used for uplinks to the ToR switches and overlay traffic for customer T1 gateway subnets. The N-VDS uses the overlay traffic segments to enable the hosts to pass the traffic through the overlay network.

- The 9k-a-seg carries the north-south traffic to the Cisco Nexus 9000 Series ToR 9K-A switch.
- The 9k-b-seg carries the north-south traffic to the Cisco Nexus 9000 Series ToR 9K-B switch.
- The T1-seg-1 carries the T1 gateway east-west traffic flows on the overlay network.
- The T1-seg-2 carries the T1 gateway east-west traffic flows on the overlay network.

The VMware NSX-T Data Center for VxBlock System design instantiates two segments on the T1 gateway. These segments enable testing of intersegment routing after the VMware NSX-T Data Center deployment is completed. If the segments are not needed, you can safely delete them.

Transport zones

Segments are created as part of a VMware NSX-T Data Center object called a transport zone. There are VLAN-backed transport zones and overlay-backed transport zones.

VLAN backed transport zones: These zones connect to the physical infrastructure usually for north-south connectivity. VLAN backed transit zones can also provide VMware NSX-T Data Center services such as microsegmentation to workloads that do not require the GENEVE overlay.

Overlay transport zones: These zones use the VMware NSX-T Data Center domain to route GENEVE-encapsulated traffic to external devices or networks and centralized services. An edge VM can support one overlay transport zone.

The VMware NSX-T Data Center for VxBlock System design includes transport zones for uplink and overlay segments. The transport zones create the NSX-managed Virtual Distributed Switch (N-VDS) that the segments are connected to.

The uplink transport zones are VLAN-backed segments which are defined as the following in the design:

- 9ka-tz for the VLAN traffic north-south to the Cisco Nexus 9000 Series 9K-A switch
- 9kb-tz for the VLAN traffic north-south to the Cisco Nexus 9000 Series 9K-B switch

An overlay transport zone carries the east-west Geneve traffic. The overly transport zone is:

- overlay-tz

The VMware NSX-T Data Center for VxBlock System design provides a minimal framework to support the following:

- East-west overlay-backed data flows
- North-south egress from overlay-backed segments to nonoverlay-backed endpoints on the physical network
- North-south ingress from the physical network destined to an endpoint behind the overlay

Customize or add to this design after the solution is delivered. However, modified designs must not interfere with the operation of the system as designed.

Profiles

An uplink profile is a template that defines how an NSX-managed Virtual Distributed Switch (N-VDS) connects to the physical network.

An uplink profile specifies:

- The format of the uplinks of an N-VDS
- The default teaming policy that is applied to those uplinks
- The transport VLAN used for overlay traffic (if relevant)
- The MTU of the uplinks
- The network I/O control profile

The VMware NSX-T Data Center for VxBlock System design includes two uplink profiles:

Profile	Description
Edge-9ka-uplink	This uplink profile carries north-south traffic to the Cisco Nexus 9000 Series 9K-A ToR switch. This profile has one uplink port that is assigned with an MTU of 9000. No VLAN is attached to this profile because tagging of this VLAN occurs at the port group of the DVS where the edge node is connected.
Edge-9kb-uplink	This uplink profile carries north-south traffic to the 9K-B ToR switch. This profile has one uplink port that is assigned with an MTU of 9000. No VLAN is attached to this profile because tagging of this VLAN occurs at the port group of the DVS where the edge node is connected.

For the compute workload environment, the VMware NSX-T Data Center for VxBlock System design includes the following profiles to connect workloads to the VMware NSX-T Data Center domain:

Profile	Description
nsx-edge-tep-profile	This profile defines the overlay network VLAN (121) the teaming policy (failover order) and uplink port (uplink-1) for which the workload traffic is routed. This profile also assigns the host/vm an overlay IP address based on the TEP pool for East-West traffic in the VMware NSX-T Data Center domain. The centralized Service Port Feature requires tagging to happen at the N-VDS. Trunk (not tag) in the VDS for the transport VLAN (121) on the edge nodes. The design, set the VLAN Type in the vcesys_nsx_transport port group to **VLAN Trunking**. Also, in the VMware NSX-T Data Center Manager, the design defines VLAN 121 in the uplink profile for nsx-edge-tep.
nsx-compute-transport	The transport hosts use this profile to connect to the VMware NSX-T Data Center domain using the following: • Teaming policy (load balance source) • Uplink ports (uplink-1,uplink-2) • VLAN 121, which carries the transport host traffic

Tier 0 gateway

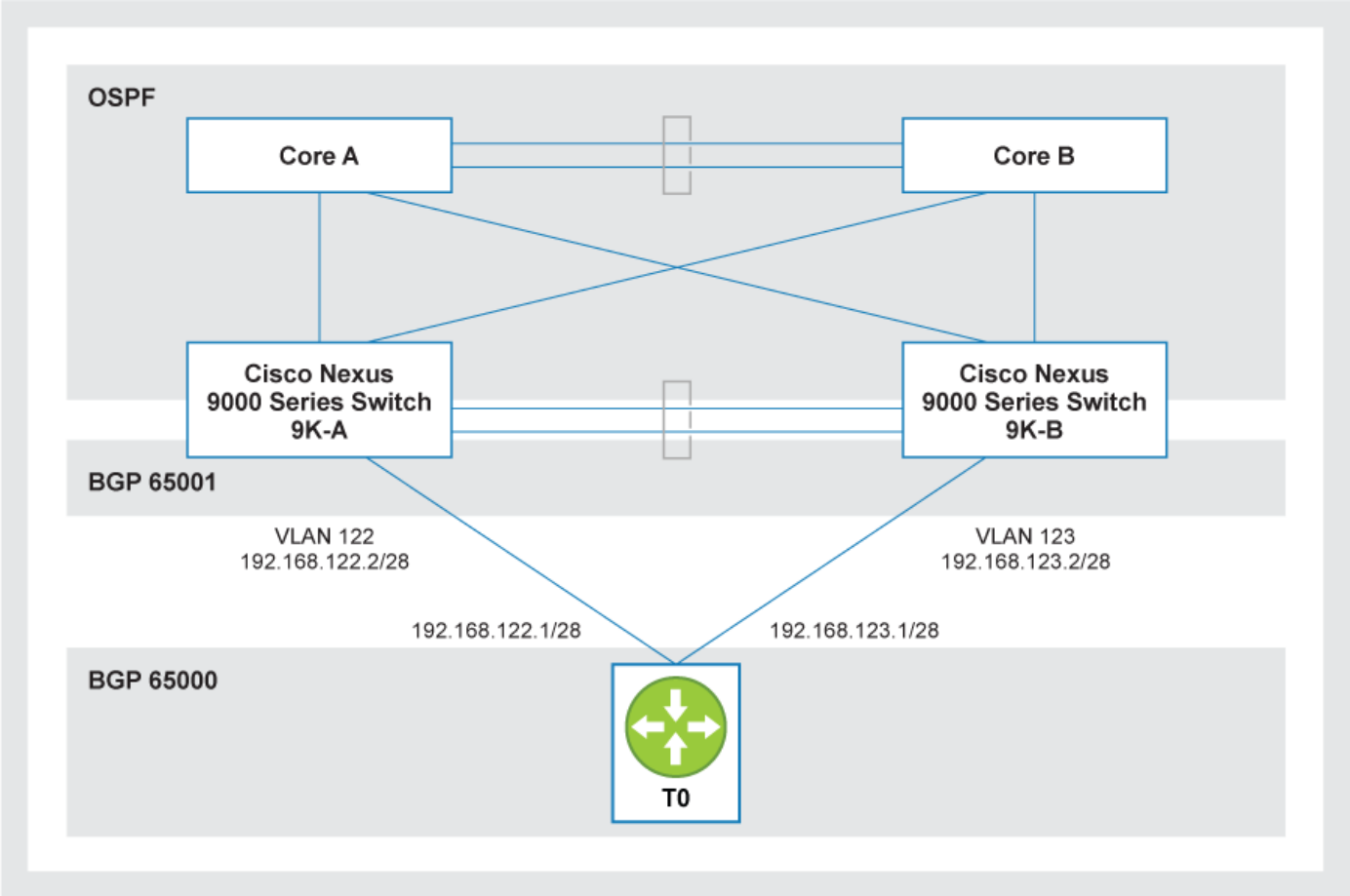
A Tier 0 gateway has downlink connections to Tier 1 gateways and uplink connections to physical networks. The Tier 0 gateway provides a gateway service between the logical and physical network. The Tier 0 gateway enables route redistribution, routing, and BGP, as well as optional T0 services.

The interfaces for this gateway connect the physical edge hosts to each ToR switch using the uplink segments described in the *Segments* topic.

In the VMware NSX-T Data Center for VxBlock System design, by default, there is one Tier 0 gateway to handle the T0-GW-1 traffic.

T0-GW-1 uses BGP and ECMP to link the nodes to the Cisco Nexus 9000 Series ToR switches.

The following figure shows the North-South routing for the Tier 0 gateway in the design:

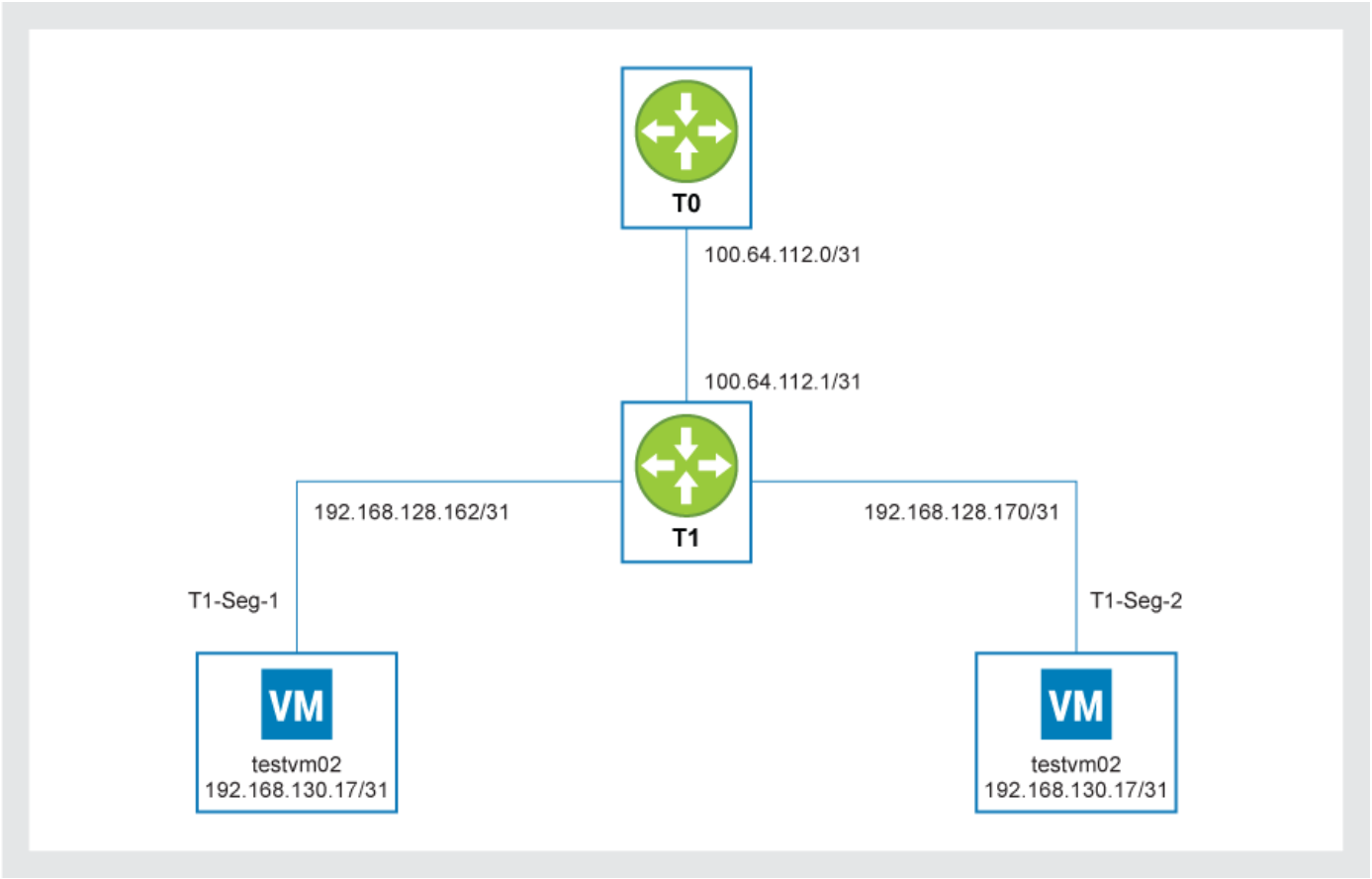


Tier 1 gateway

A Tier 1 gateway is typically connected to a Tier 0 gateway in the north-bound direction and to segments in the south-bound direction. Tier 1 gateways may also provide centralized services such as load balancing, VPN services, NAT, DHCP, and so on.

The VMware NSX-T Data Center for VxBlock System design includes, by default, a single Tier 1 gateway (T1-GW-1). This gateway provides a link between the two overlay-backed segments and the Tier 0 gateway that connects to the physical network.

The following figure shows the topology for the Tier 0 and Tier 1 gateways:


T0
100.64.112.0/31
100.64.112.1/31
T1
192.168.128.162/31
192.168.128.170/31
T1-Seg-1
T1-Seg-2
VM
testvm02
192.168.130.17/31
VM
testvm02
192.168.130.17/31

VMware NSX-T Data Center transport nodes

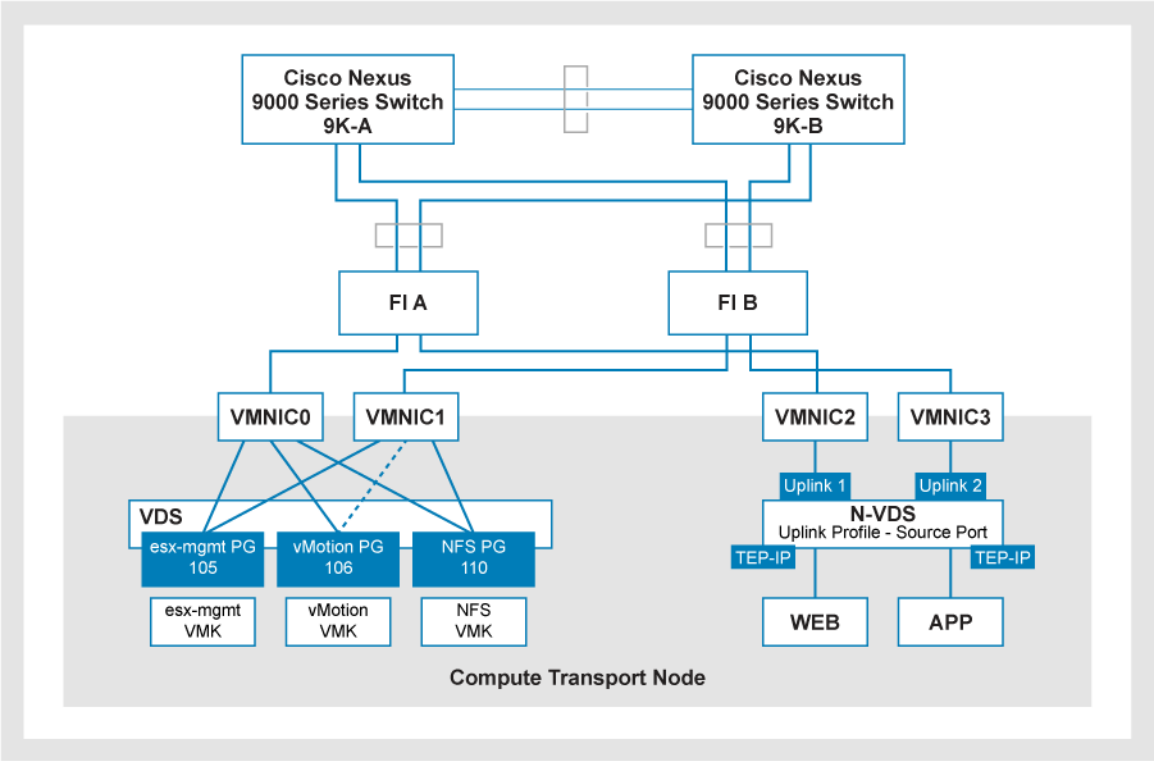
Hypervisor transport nodes are hypervisors that are prepared and configured for VMware NSX-T Data Center.

The VDS provides network services to the virtual machines that are running on those hypervisors. VMware NSX-T Data Center supports VMware vSphere ESXi and KVM hypervisors. The N-VDS that is implemented for KVM is based on the Open vSwitch (OVS) and is platform-independent. It can be ported to other hypervisors and serves as the foundation for the implementation of VMware NSX-T Data Center in other environments (for example, cloud and containers). For the VMware NSX-T Data Center on VxBlock Systems design, Dell EMC Sales Engineers deploy and support only VMware vSphere ESXi based transport nodes. The VMware support organization supports other types of transport nodes.

The VMware NSX-T Data Center for VxBlock System transport node design uses the following criteria:

- There is one N-VDS using vmnic0 and vmnic1 for VMware vSphere ESXi host functions including port groups and kernels for VMware vSphere ESXi management, vMotion, and NFS.
- There is one N-VDS using vmnic2 and vmnic3 for workload VMs.
  The N-VDS is used for East-West and North-South traffic with the use of TEPs to create an overlay network.
- Uplink teaming of source port is used on the N-VDS to ensure load-balancing.
- An MTU value of 9000 is important on these vNICs to allow for the overhead of GENEVE tunnel encapsulation.
- VLAN tagging is required at the uplink profile for TEP traffic. There is no VDS in front of the N-VDS.

The use of the VMware VDS and N-VDS enables the separation of the VMware vSphere ESXi host functions from VMware NSX-T Data Center traffic. If a failure occurs, it also makes troubleshooting and recovery of a host easier. The following figure illustrates the topology of this design:



In this design, the transport nodes are connected to FI A and B through vmnic0 and vmnic1. The N-VDS is connected through vmnic2 and vmnic3 to FI A and B.

VLAN 121 is added to the vNIC template for vNICs 2 and 3 and Cisco Nexus 9000 Series Switches (9K-A and 9K-B) trunk ports to FIs for the overlay network.

Adapter policy settings for transport nodes

To realize maximum performance from the VMware NSX-T Data Center transport nodes using Cisco VIC adapters, Cisco recommends modifying the Ethernet Adapter Policy from Cisco UCS manager.

Create an adapter policy to define these settings and apply it to vmnic2 and vmnic3 on each transport node. The settings are detailed in the following table:

Parameter	Value
Transmit queue	1
Ring size (transmit)	2048
Receive queues	8
Ring size (receive)	2048
Completion queues	9
Interrupts	11
Receive side scaling (RSS)	Enabled
VXLAN offload	Disabled

TEP IP address pool

To provide a mechanism to transmit GENEVE overlay-backed traffic over the physical VLAN backed network, VMware NSX-T Data Center requires a tunnel end point (TEP) IP address pool. This pool is a range of IP addresses on the VMware NSX-T Data Center transport VLAN that are reserved for TEPs.

Each edge node VM requires one IP address from this pool, and each transport node requires two IP addresses also from this pool. When sizing the pool, consider the current and future requirements for TEPs.

For VMware NSX-T Data Center deployments with more than one Layer 3 domain (for example, multisystem cross-VMware vCenter Server), you must deploy a separate TEP IP pool for each domain. Layer 3 routing must be configured and working between all TEP IP pools. This configuration enables transport traffic to traverse between transport nodes on different pools.

Licensing

Order either the Advanced or Enterprise Plus editions of VMware NSX-T Data Center for use on the VxBlock System. A detailed breakdown of each edition and the features that are supported is in the VMware NSX-T Data Center data sheet. There is an important difference in licensing between VMware NSX-V and VMware NSX-T Data Center. In VMware NSX-T Data Center, the physical Edge hosts do not need VMware NSX-T Data Center licensing. Since the physical Edge host is not prepared as a transport node, it does not consume any VMware NSX-T Data Center CPU licenses.

A customer can license any number of transport nodes (compute hosts). Consider the following:

- The licensed transport nodes can be a subset of the physical hosts in the VxBlock System.
- The licensed transport nodes could be more hosts than physically exist in the system.
  In this approach, the customer wants to stretch overlay-backed segments across multiple VxBlock Systems.
- All VMware NSX-T Data Center licensing must be purchased directly through VMware. Dell EMC is unable to resell the licensing for the product due to export compliance issues. See *Obtaining a VMware license*.

Cisco UCS licensing

The licensing that is required to connect the physical edge hosts to the UCS domain varies depending on the following:

- The type of Cisco UCS domain to which the hosts are connecting
- Whether there is a FEX connected to the FIs

The various connectivity models and their licensing requirements are detailed in the following table:

Connectivity model	Licensing details
Cisco UCS Third-Generation domain (6332-16UP FIs), no FEX	Each QSFP port on a Cisco UCS Third-Generation FI used for VMware NSX-T Data Center edge direct-connect should be licensed using the direct-connect SKU UCS-LIC-6300-40GC. Two edge hosts require two SKUs. Four, six, or eight hosts require four SKUs. More than 8 edge hosts require 8 SKUs.
Cisco UCS Fourth-Generation domain (64xxx FIs), no FEX	Each edge compute host that connects directly to a Cisco UCS Gen4 FI must be licensed to connect to the Cisco UCS domain. The license is the Cisco UCS C-series Rack Server only 25 Gbps SKU (UCS-LIC-6400-25GC), quantity two ports per server.
Any Cisco UCS domain with FEX	Any edge compute host that connects to a FEX and not to an FI does not need any additional port licensing.

VMware vSphere ESXi licensing

Each physical edge host consumes standard per-socket CPU licensing for VMware vSphere ESXi. In this configuration, each host has two physical CPU sockets that are populated, and consumes two VMware vSphere ESXi CPU licenses.

Related information

Obtaining a VMware license

Layer 3 licensing

For VMware NSX-T Data Center deployments on VxBlock Systems, the ToR switch pair must be licensed for Layer 3 routing services.

Dell EMC offers two license packages that enable this capability:

- Essentials
- Advantage

These packages include the licensing that is required to run VMware NSX-T Data Center on a VxBlock System.

For VMware NSX-T Data Center deployments on VxBlock 340, 350, 540, and 740 Systems, the switch pair must be licensed for Layer 3 routing services. License types are different based on the type of ToR switch that is installed in the system.

The VxBlock 340, 350, 540, and 740 Systems have a Cisco Nexus 93180YC-EX or a Cisco Nexus 9396PX ToR switches. One of the following licenses must be installed on the switch.

Switch	License type
Cisco Nexus 9396PX	LAN Enterprise
Cisco Nexus 93180YC-EX	Essentials Advantage

These packages include the licensing that is required to run VMware NSX-T Data Center on a VxBlock 340, 350, 540, or 740 System.

VxBlock Systems with Cisco Nexus 5000 series ToR switches are not supported with NSX-T.