v2.4 | April 2022 | BP-2074

BEST PRACTICES

VMware vSphere Networking

NUTANIX
YOUR ENTERPRISE CLOUD

Copyright

Contents

1. Executive Summary

With the VMware vSphere hypervisor, you can dynamically configure, balance, or share logical networking components across various traffic types. Therefore, it's imperative that you take virtual networking into consideration when you design and build a network solution for Nutanix hyperconverged appliances.

This best practice guide addresses the fundamental features of VMware's vSphere switching technology and details the configuration elements you need to run vSphere on Nutanix. The guidance in this document can help you quickly and easily develop an appropriate design strategy for deploying the Nutanix hyperconverged platform with VMware vSphere 6.x as the hypervisor.

Document Version History

Version Number	Published	Notes
1.0	March 2017	Original publication.
1.1	May 2017	Updated platform overview, virtual networking option diagrams, and Nutanix VLAN IDs.
1.2	February 2018	Updated platform overview and added network configuration best practices.
1.3	March 2018	Updated Jumbo Frames section.
1.4	July 2018	Updated Nutanix overview and vSphere 6.x: NIOC Version 3 section.

Version Number	Published	Notes
2.0	September 2018	Added Port Binding with vSphere Distributed Switches section and updated Jumbo Frames section.
2.1	September 2019	Updated Nutanix overview.
2.2	June 2020	Clarified NIOC recommendations, updated jumbo frame recommendations.
2.3	June 2021	Refreshed content.
2.4	April 2022	Updated Network Configuration Best Practices section.

2. Introduction to VMware vSphere Networking

VMware vSphere supports two main types of virtual switches: the vSphere Standard Switch (vSS) and the vSphere Distributed Switch (vDS). Both switches provide basic functionality, which includes the ability to forward layer 2 frames, provide 802.1q VLAN encapsulation, manage traffic shape, and use more than one uplink (NIC teaming). The main differences between these two virtual switch types pertain to switch creation, management, and the more advanced functionality they provide.

vSphere Standard Switch

The vSS is available in all versions of VMware vSphere and is the default method for connecting VMs, as well as management and storage traffic, to the external (or physical) network. The vSS is part of the standard vSphere licensing model and has the following advantages and disadvantages.

Advantages:

- Simple and easy to configure.
- No reliance on VirtualCenter (vCenter) availability.

Disadvantages:

- No centralized management.
- No support for Network I/O Control (NIOC).
- No automated network load balancing.
- No backup, restore, health check, or network visibility capabilities.

vSphere Distributed Switch

To address many of the vSS's shortcomings, the vDS separates the networking data plane from the control plane, enabling advanced networking features such as load balancing, traffic prioritization, backup and restore capabilities, and so on. These features are certainly compelling, but it's important to note the following disadvantages of the vDS:

- Requires Enterprise Plus licensing from VMware.
- Because VMware vCenter stores the control plane in this configuration, you must have vCenter server availability to configure and manage the vDS.
- Management configuration is complex.

The vDS has matured over the several iterations of the vSphere product, providing new functions as well as changes to existing components and behaviors. vSphere 5.x provided the following capabilities and enhancements:

- Inter-VM traffic visibility via the Netflow protocol.
- LLDP (link layer discovery protocol) support, to provide upstream switch visibility.
- Enhanced link aggregation support, including a variety of hashing algorithms.
- Single-root I/O virtualization (SR-IOV) support.
- Support for 40 GbE network interface cards (NICs).
- NIOC version 2.

In vSphere 6.x, VMware introduced improvements as well as new features and functions:

- NIOC version 3.
- Support for IGMP snooping for IPv4 packets and MLD snooping for IPv6 packets.
- Multiple TCP/IP stacks for vMotion.

The following table concisely compares the vSS and the vDS.

Table: Summary Comparison Between vSS and vDS

Switch Type	License	Mgmt. Method	Private VLANs	NIOC	LLDP	LACP
vSS	Standard	Host or vCenter	No	No	No	No
vDS	Enterprise Plus	vCenter	Yes	Yes	Yes	Yes

vSS teaming options:

- Originating port ID
- IP hash
- Source MAC hash
- Explicit failover order

vDS teaming options:

- Originating port ID
- IP hash
- Source MAC hash
- Explicit failover order
- Route based on physical NIC load

VMware NSX

Running VMware NSX for vSphere on Nutanix improves control and increases the agility of the compute, storage, virtualization, and network layers. Nutanix has evaluated and tested VMware NSX for vSphere and developed a technical guide for customers who want to move toward SDN architecture with Nutanix.

3. Discovery Protocols

From vSphere 5.0 onward, the vDS supports two discovery protocols: LLDP and CDP (Cisco discovery protocol). Discovery protocols give vSphere administrators visibility into connectivity from the virtual network to the physical network, which simplifies troubleshooting in the event of cabling issues, MTU (maximum transmission unit) or packet fragmentation, or other problems. The only disadvantage to using discovery protocols is a potential security issue, as they make both the topology of the network and switch-specific information visible.

VMware vSphere offers three configuration options or attributes for LLDP, presented in the following table. Each option changes the information that the selected discovery protocol sends and receives. CDP is either enabled or not—it doesn't have additional options.

Table: LLDP Discovery Options

Option	Description
Listen	In this mode, ESXi detects and displays information from the upstream switch, but doesn't advertise the vDS configuration to the upstream network device.
Advertise	In this mode, ESXi advertises information about the vDS to the switch administrator but doesn't detect or display information about the physical switch.
Both	This attribute listens and advertises information between the vDS and upstream switch provider.

Nutanix recommends the Both option, which ensures that ESXi collects and displays information from the physical switch and sends information about the vDS to the physical switch.

The following information is visible in the vSphere client when you enable discovery protocol data:

- The physical switch interface connected to the dvUplink.
- MTU size (in other words, whether you enabled jumbo frames).
- The switch management IP address.
- The switch name, description, software version, and capabilities.

The following table presents some general guidelines, but Nutanix recommends that all customers carefully consider the advantages and disadvantages of discovery protocols for their specific security and discovery requirements.

Table: Recommendations for Discovery Protocol Configuration

Option	Description
Type	Dependent on your switching infrastructure. Use CDP for Cisco-based environments and LLDP for non-Cisco environments.
Operation	Both CDP and LLDP are generally acceptable in most environments and provide maximum operational benefits. Configuring an attribute (listen, advertise, or both) in LLDP allows vSphere and the physical network to openly exchange information. You can also choose not to configure an attribute in LLDP if you don't want this exchange to occur.

4. Network I/O Control

NIOC is a vDS feature that has been available since the introduction of vSphere 4.1. The feature initially used network resource pools to determine the bandwidth that different network traffic types could receive in the event of contention. With vSphere 6.0 and beyond, we must also take shares, reservations, and limits into account. The following sections detail our recommendations.

NIOC Version 2

NIOC version 2 was introduced in vSphere 5.1 and became generally available in vSphere 5.5. Enabling NIOC divides vDS traffic into the following predefined network resource pools:

- Fault tolerance (FT)
- iSCSI
- vMotion
- Management
- vSphere Replication
- NFS
- VM

The physical adapter shares assigned to a network resource pool determine what portion of the total available bandwidth is guaranteed to the traffic associated with that network resource pool in the event of contention. It's important to understand that NIOC only impacts network traffic if there's contention. Thus, when the network is less than 100 percent utilized, NIOC has no advantage or disadvantage.

Comparing a given network resource pool's shares to the allotments for the other network resource pools determines its available bandwidth. You can apply limits to selected traffic types, but Nutanix recommends against it, as limits may inadvertently constrain performance for given workloads when there is available bandwidth. Using NIOC shares ensures that burst workloads such as vMotion (migrating a VM to a different ESXi host) can complete as quickly as possible when bandwidth is available, while also protecting other workloads from significant impact if bandwidth becomes limited.

The following table shows the Nutanix-recommended network resource pool share values.

None of the network resource pools configure artificial limits or reservations, so leave the host limit set to unlimited.

Table: Recommended Network Resource Pool Share Values

Network Resource Pool	Share Value	Physical Adapter Shares
Management traffic	25	Low
vMotion traffic	50	Normal
FT traffic	50	Normal
VM traffic	100	High
NFS traffic	100	High
iSCSI traffic	100	High
vSphere Replication traffic	50	Normal
vSphere Storage Area Network (SAN) traffic	50	Normal
Virtual SAN traffic	50	Normal

Note: The vSphere SAN traffic pool is exposed in vSphere 5.1, but as it has no function and no impact on NIOC, we can ignore it.

Note: The virtual SAN traffic pool is exposed in vSphere 5.5, but as Nutanix environments don't use it, it has no impact on NIOC here.

The following sections describe our rationale for setting the share values in this table.

Note: The calculations of minimum bandwidth per traffic type presented here assume that the system isn't using iSCSI, vSphere Replication, vSphere SAN traffic, or Virtual SAN traffic. Also, we assume that NFS traffic remains with the host for the default "NFS traffic" resource pool.

Management Traffic

Management traffic requires minimal bandwidth. A share value of 25 (low) and two 10 GbE interfaces ensure a minimum of approximately 1.5 Gbps for management traffic, which exceeds the minimum bandwidth requirement of 1 Gbps.

vMotion Traffic

vMotion is a burst-type workload that uses no bandwidth until the distributed resource scheduler (DRS), a vCenter service that actively rebalances VMs across hosts, invokes it or a vSphere administrator starts a vMotion or puts a host into maintenance mode. As such, it's unlikely to have any significant ongoing impact on the network traffic. A share value of 50 (normal) and two 10 GbE interfaces provide a minimum of approximately 3 Gbps, which is sufficient to complete vMotion in a timely manner without degrading VM or storage performance and well above the minimum bandwidth requirement of 1 Gbps.

Fault Tolerance Traffic

FT traffic depends on the number of fault-tolerant VMs you have per host (current maximum is four per host). FT is generally a consistent workload (as opposed to a burst-type workload like vMotion), as it needs to keep the primary and secondary VMs' CPU resources synchronized. You typically use FT for critical VMs, so to protect FT traffic (which is also sensitive to latency) from impact during periods of contention, use a share value of 50 (normal) and two 10 GbE interfaces. This setting provides a minimum of 3 Gbps, which is well above VMware's recommended minimum of 1 Gbps.

VM Traffic

VM traffic is the reason we have datacenters in the first place, so this traffic is always important, if not critical. As slow VM network connectivity can quickly impact end users and reduce productivity, you must ensure that this traffic has a significant share of the available bandwidth during periods of contention. With a share value of 100 (high) and two 10 GbE interfaces, VM traffic receives

a minimum of approximately 6 Gbps. This bandwidth is more than what is required in most environments and ensures a good amount of headroom in case of unexpected burst activity.

NFS Traffic

NFS traffic is always critical, as it's essential to distributed storage and to CVM and VM performance, so it must receive a significant share of the available bandwidth during periods of contention. However, as NFS traffic is serviced locally, it doesn't impact the physical network card unless the Nutanix CVM fails or is offline for maintenance. Thus, under normal circumstances, no NFS traffic crosses the physical NICs, so the NIOC share value has no impact on other traffic. As such, we have excluded it from our calculations.

In case of network contention, CVM maintenance, or failure, assign NFS traffic a share value of 100 (high) as a safety measure, ensuring a minimum of 6 Gbps of bandwidth.

Nutanix CVM Traffic

For Nutanix storage to function, it must have connectivity to the other CVMs in the cluster for tasks such as synchronously writing I/O across the cluster and Nutanix cluster management.

> Note: Under normal circumstances, minimal or no read I/O traffic crosses the physical NICs; however, write I/O always uses the physical network.

Since the CVM is a virtual machine, it inherits the VM traffic policy and shares discussed earlier, which means that in the event of contention, 100 shares are equally divided between all VMs. This method ensures that all applications and storage services continue to function with minimal network contention.

iSCSI Traffic

Like NFS traffic, iSCSI traffic isn't used by default; however, if you're using it, it may be critical to your environment. As such, give this traffic type a share value of 100.

> Note: NIOC doesn't cover in-guest iSCSI. If you're using in-guest iSCSI, we recommend that you create a dvPortGroup for in-guest iSCSI traffic and assign it to a custom network resource pool called In-Guest iSCSI. Assign this pool a share value of 100 (high).

vSphere Replication Traffic

In a non-Nutanix environment, vSphere Replication traffic may be critical to your environment if you choose to use vSphere Replication (with or without VMware Site Recovery Manager (SRM)). However, if you're using SRM, we strongly recommend using the Nutanix Storage Replication Adaptor (SRA) instead of vSphere Replication, because it's more efficient. If you use vSphere Replication without SRM, the default share value of 50 (normal) is suitable for most environments, ensuring approximately 3 Gbps of network bandwidth.

vSphere Storage Area Network and Virtual SAN Traffic

vSphere SAN traffic is visible as a parameter in vSphere 5.1, but it isn't used. Therefore, simply leave the default share value, as vSphere SAN traffic has no impact on NIOC.

Virtual SAN traffic is visible as a parameter from vSphere 5.5 onward, but Nutanix environments don't use it. Therefore, simply leave the default share value as it is, as virtual SAN traffic also has no impact on the environment.

vSphere 6.x: NIOC Version 3

VMware vSphere version 6.0 and later feature a newer iteration of NIOC: version 3. Depending on the version of vSphere 6, NIOC v3 allows an administrator to set not only shares, but also artificial limits and reservations on system traffic such as vMotion, management, NFS, VMs, and so on. Administrators can apply limits and reservations to a VM's network adapter using the same constructs they used when allocating CPU and memory resources to the VM. Admission control is now part of the vDS, which integrates with VMware High Availability and DRS to actively balance network resources across hosts.

Although setting artificial limits and reservations on various traffic types guarantees quality of service and SLAs, it prevents other workloads from using the total available bandwidth. Reserving a certain amount of bandwidth strictly for one workload or purpose constrains the network's ability to sustain an unreserved workload's short bursts of traffic.

vSphere 6.0.x: NIOC Version 3

> Note: For customers who've implemented vSphere 6.0.x and are considering NIOC v3, this combination of versions only works when you use artificial limits and reservations to prioritize traffic flows.

While testing and evaluating the functionality of NIOC v3 with vSphere 6.0.x, Nutanix has encountered various abnormalities in its usability and behavior as well as generally higher CPU overhead. Because of these results, we recommend that customers implement NIOC v2 instead of NIOC v3.

To implement NIOC v2, provision a vDS with 5.5.0 selected as the distributed switch version, following the proportional shares we described above for NIOC v2. For guidance on creating a new vDS on your cluster and configuring its settings, refer to VMware vSphere documentation.

vSphere 6.5 Update 1 and Later: NIOC Version 3

For customers who have implemented vSphere 6.5 update 1 and later, Nutanix fully supports the vDS operating at version 6.5 in combination with NIOC v3. Nutanix can fully support this version combination because VMware reverted to using the proportional share algorithm, which means you use shares, rather than limits and hard reservations, to prioritize traffic types only in the event of contention. Refer to the table Recommended Network Resource Pool Share Values for the share values we recommend. Nutanix testing shows only a minimal performance deviation between using vDS 5.5 with NIOC v2 and using vDS 6.5 with NIOC v3.

5. NIC Teaming and Failover

vSphere provides several NIC teaming and failover options. Each of these settings has different goals and requirements, and you need to understand and carefully consider them before you use them in your vSphere deployments.

The available options are:

1. Route based on originating virtual port.
2. Route based on physical NIC load (also called load-based teaming or LBT).
   - Only available with the vDS.
3. Route based on IP hash.
4. Route based on source MAC hash.
5. Use explicit failover order.

Recommendation for vSS: Route Based on Originating Virtual Port

The Route based on originating virtual port option is the default load balancing policy and doesn't require an advanced switching configuration such as LACP, Cisco EtherChannel, or HP teaming, making it simple to implement, maintain, and troubleshoot. It only requires 802.1q VLAN tagging for secure separation of traffic types. This option's main disadvantage is that it doesn't support load balancing based on network load, so the system always sends traffic from a single VM to the same physical NIC unless a NIC or upstream link failure causes a failover event.

Recommendation for vDS: Route Based on Physical NIC Load

The Route based on physical NIC load option (LBT) also doesn't require an advanced switching configuration such as LACP, Cisco Ether channel, or HP teaming. It offers fully automated load balancing, which takes effect when one or more NICs reach and sustain 75 percent utilization for a period of at least 30 seconds, based on egress and ingress traffic. LBT's only requirement is

802.1q VLAN tagging for secure separation of traffic types, so it is also simple to implement, maintain, and troubleshoot.

> Note: The vDS requires VMware vSphere Enterprise Plus licensing.

LBT is a simple and effective solution that works very well in Nutanix deployments.

Network Failover Detection

ESXi uses one of two network failover detection methods: beacon probing or link status. Beacon probing sends out and listens for beacon probes, which are made of broadcast frames. Because beacon probing must have three network connections to function, we don't recommend it for Nutanix solutions, which currently have two NICs of each speed (1 Gbps and 10 Gbps).

Link status depends on the state that the physical NIC reports for the link. Link status can detect failures such as a cable disconnection or a physical switch power failure, but it can't detect configuration errors or upstream failures that may also result in connectivity issues for VMs.

To avoid these link status limitations related to upstream failures, enable Link state tracking (if the physical switches support it) or an equivalent, which then enables the switch to pass upstream link state information back to ESXi, so the link status triggers a link down on ESXi where appropriate.

Notify Switches

The purpose of the notify switches policy setting is to enable or disable communication by ESXi with the physical switch in the event of a failover. Choosing Yes for this policy setting means that when a failover event routes a virtual Ethernet adapter's traffic over a different physical Ethernet adapter in the team, ESXi sends a notification to the physical switch to update the lookup tables. This configuration ensures that failover occurs in a timely manner and with minimal interruption to network connectivity.

Failback

For customers not using Enterprise Plus or the vDS with LBT, failback can help rebalance network traffic across the original NIC, which may improve network performance. The only significant disadvantage of setting failback to Yes is that, in the unlikely event of network instability (or flapping), having network traffic fail back to the original NIC may result in intermittent or degraded network connectivity.

Nutanix recommends setting failback to Yes when you use vSS and No when you use vDS.

Failover Order

Using failover order allows the vSphere administrator to specify the order NICs fail over in by assigning a physical NIC to one of three groups: active adapters, standby adapters, or unused adapters. For example, if all active adapters lose connectivity, the system uses the highest priority standby adapter, and so on. This feature is only necessary in a Nutanix environment when you perform multi-NIC vMotion.

When you configure multi-NIC vMotion, set the first dvPortGroup used for vMotion to have one dvUplink active and the other standby. Configure the reverse for the second dvPortGroup used for vMotion.

For more information, see Multiple-NIC vMotion in vSphere 5 (KB 2007467).

Summary of Recommendations for NIC Teaming and Failover

Table: Recommendations for NIC Teaming and Failover: vDS

Option	Recommended Setting
Load Balancing	Route based on physical NIC load (LBT)
Network Failover Detection	Link status only
Notify Switches	Yes
Failback	No

Table: Recommendations for NIC Teaming and Failover: vSS

Option	Recommended Setting
Load Balancing	Route based on originating virtual port
Network Failover Detection	Link status only
Notify Switches	Yes
Failback	Yes

Note: For network failover detection, enable link state tracking or the equivalent on physical switches.

6. Security

When configuring a vSS or vDS, there are three configurable options under security you can set to Accept or Reject: promiscuous mode, MAC address changes, and forged transmits.

In general, the most secure and appropriate setting for each of these options is Reject unless you have a specific requirement for Accept. Two examples of situations in which you might consider configuring Accept on forged transmits and MAC address changes are:

1. Microsoft load balancing in Unicast mode.
2. iSCSI deployments on select storage types.

The following table offers general guidelines, but Nutanix recommends that all customers carefully consider their requirements for specific applications.

Table: Recommendations for VMware Virtual Networking Security

Option	Recommended Setting
Promiscuous mode	Reject
MAC address changes	Reject
Forged transmits	Reject

7. Virtual Networking Configuration Options

The following two virtual networking options cover the Nutanix recommended configurations for both vSS and vDS solutions. For each option, we discuss the advantages, disadvantages, and example use cases. In both of the following options, Nutanix recommends setting all vNICs as active on the portgroup and dvPortGroup unless otherwise specified.

The following table defines our naming convention, portgroups, and corresponding VLANs as used in the next examples for various traffic types.

Table: Portgroups and Corresponding VLANs

Portgroup	VLAN	Description
MGMT_10	10	VM kernel interface for host management traffic
VMOT_20	20	VM kernel interface for vMotion traffic
FT_30	30	FT traffic
VM_40	40	VM traffic
VM_50	50	VM traffic
NTNX_10	10	Nutanix CVM to CVM cluster communication traffic (public interface)
Svm-iscsi-pg	N/A	Nutanix CVM to internal host traffic
VMK-svm-iscsi-pg	N/A	VM kernel port for CVM to hypervisor communication (internal)

All Nutanix deployments use an internal-only vSS for the NFS communication between the ESXi host and the Nutanix CVM. This vSS remains unmodified

regardless of the virtual networking configuration for ESXi management, VM traffic, vMotion, and so on.

> Note: Don't modify the internal-only vSS (vSS-Nutanix). vSS-Nutanix facilitates communication between the CVM and the internal hypervisor.

The configurations for vSS and vDS solutions look very similar—the main difference is the location of the control plane.

- With the default vSS, there is a control plane on each host, and each operates independently. This independence requires an administrator to configure, maintain, and operate each individual vSS separately.
- With the vDS, the control plane is located in vCenter. This centrality requires the system to keep vCenter online to ensure that configuration and maintenance activities propagate to the ESXi hosts that are part of the vDS instance.

Both options reduce cabling and switching requirements because they don't require 1 GbE ports. Moreover, they represent a simple solution that only requires 802.1q configured on the physical network. These options are suitable for environments where logical separation of management and VM traffic is acceptable.

Option 1: vSS with Nutanix vSS Deployment

Option 1 is the default configuration for a Nutanix deployment and suits most use cases. This option is a good fit for customers who don't have VMware vSphere Enterprise Plus licensing or prefer not to use the vDS.

Customers may wish to incorporate the 1 GbE ports into the vSS to provide additional redundancy; however, this configuration requires additional cabling and switch ports. If you connect the 1 GbE ports, set them to Standby and configure failback as noted.

Option 1 offers several benefits; the most important benefit is that the control plane is located locally on every ESXi host, so there is no dependency on vCenter.

A major disadvantage of the vSS lies in its load-balancing capabilities. Although both physical uplinks are active, traffic across these physical NICs doesn't actively rebalance unless the VM has gone through a power cycle or a failure has occurred in the networking stack. Furthermore, a standard vSS can't manage contention or prioritize network traffic across various traffic classes (for example, storage, vMotion, VM, and so on).

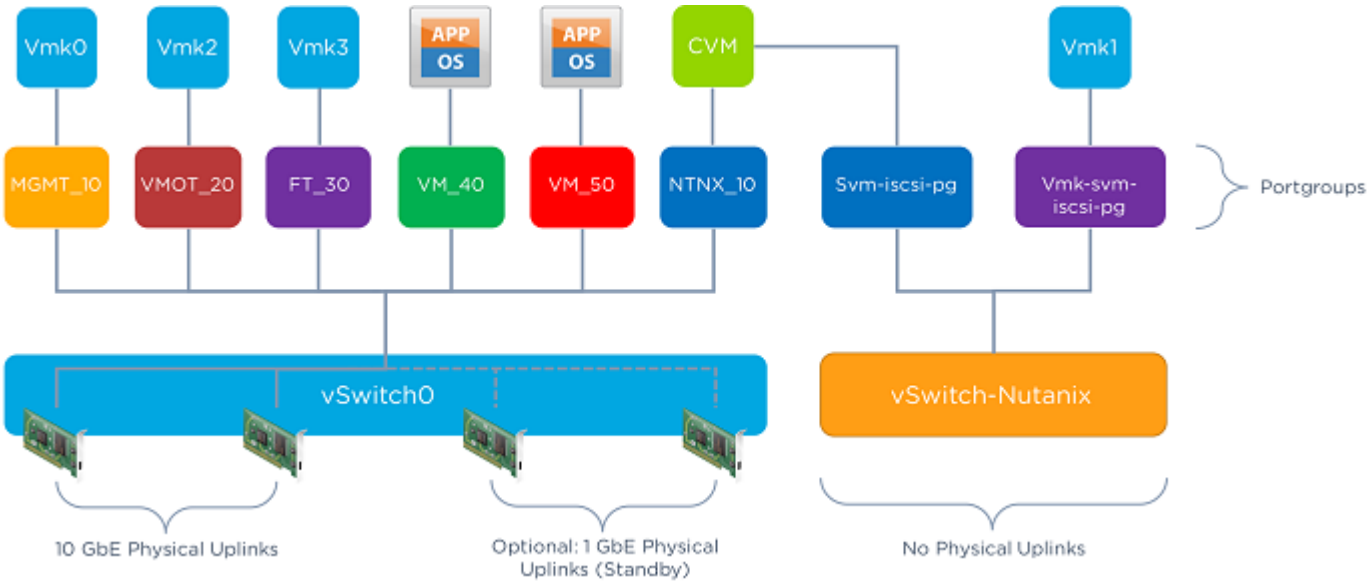

Figure 1: Virtual Networking Option 1: One vSS with vSS-Nutanix

Recommended use cases:

1. When vSphere licensing isn't Enterprise Plus.
2. When physical and virtual server deployments aren't large enough to warrant a vDS.

Option 2: vDS with Nutanix vSS Deployment

Option 2 is a good fit for customers using VMware vSphere Enterprise Plus licensing and offers several benefits:

- Advanced networking features, such as:
  - NIOC
  - Load-based teaming (route based on physical NIC load)
  - Actively rebalancing flows across physical NICs
- Ability for all traffic types to burst where required, up to 10 Gbps.
- Effective with both egress and ingress traffic.
- Lower overhead than IP hash load balancing.
- Reduced need for NIOC to intervene during traffic congestion.

Because the control plane is located in vCenter, vDS administration requires vCenter to be highly available and protected, as its corresponding database stores all configuration data. Particularly during disaster recovery operations, the requirement for vCenter availability can pose additional challenges; many administrators see this constraint as the primary disadvantage to using the vDS.

Recommended use cases:

1. When vSphere licensing is Enterprise Plus.
2. Where there is a requirement for corresponding VMware products or services.
3. Large physical and virtual server deployments.

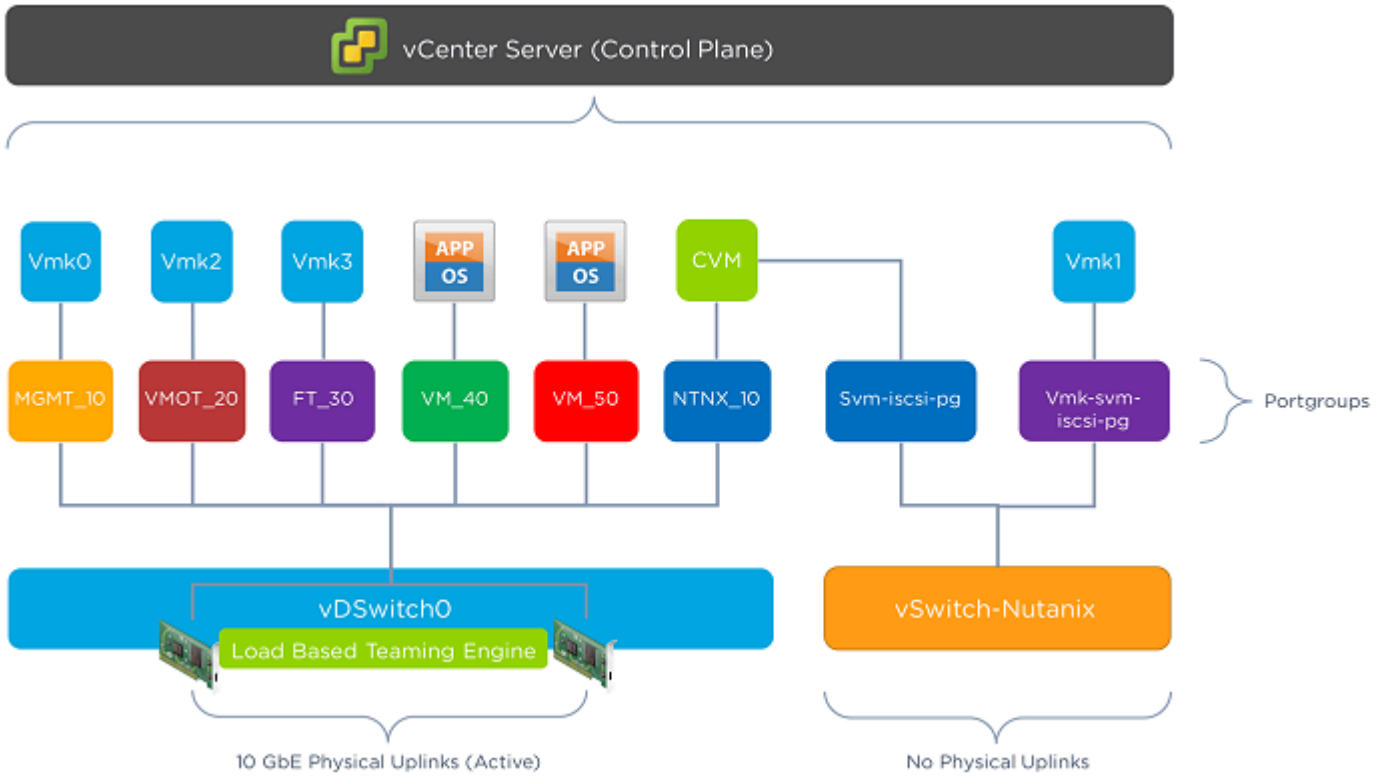

Figure 2: Virtual Networking Option 2: vDS with vSS-Nutanix

8. Jumbo Frames

The Nutanix CVM uses the standard Ethernet MTU (maximum transmission unit) of 1,500 bytes for all the network interfaces by default. The standard 1,500-byte MTU delivers excellent performance and stability. Nutanix doesn't support configuring the MTU on a CVM's network interfaces to higher values.

You can enable jumbo frames (MTU of 9,000 bytes) on the physical network interfaces of AHV, ESXi, or Hyper-V hosts and user VMs if the applications on your user VMs require them. If you choose to use jumbo frames on hypervisor hosts, enable them end to end in the desired network and consider both the physical and virtual network infrastructure impacted by the change.

9. Multicast Filtering in vSphere 6.0

When using vSphere 6.0 or later, administrators can configure advanced multicast filtering in the vDS. Prior to vSphere 6.0, VMs connected to a vSS encountered issues in forwarding and receiving their multicast traffic—VMs that may not have originally subscribed to the multicast group sending traffic received that traffic anyway. To overcome these traffic concerns, vSphere 6.0 introduced two configurable filtering modes:

1. Basic mode: Applies to the vSS and the vDS. This mode forwards multicast traffic for VMs according to the destination MAC address of the multicast group.
2. Multicast snooping: Applies only to the vDS and uses IGMP snooping to route multicast traffic only to VMs that have subscribed to its source. This mode listens for IGMP network traffic between VMs and hosts and maintains a map of the group's destination IP address and the VM's preferred source IP address (IGMPv3).

- When a VM doesn't renew its membership to a group within a certain time, the vDS removes the group's entry from the lookup records.

> Note: Enabling multicast snooping on a vDS with a Nutanix CVM attached affects its ability to discover and add new nodes in the cluster (using the Cluster Expand option in Prism and the Nutanix CLI).

If you need to implement multicast snooping, your standard operational procedures should include manual steps for adding additional nodes to a cluster. For more information, refer to Nutanix KB 3493 and VMware's vSphere documentation on multicast filtering modes.

10. Port Binding with vSphere Distributed Switches

When connecting VMs to a vDS, administrators have two options for defining how a virtual port on a vDS is assigned to a VM. For more information regarding port binding in vSphere, refer to VMware KB 1022312.

Static binding

Static binding is the default setting. Once a VM connects to the vDS, a port is immediately assigned and reserved for it. Because the port is only disconnected when you remove the VM from the port group, static binding guarantees virtual network connectivity. However, static binding also requires that you perform all connect and disconnect operations through vCenter Server.

Note: Because dynamic binding was deprecated in ESXi 5.0, customers on subsequent versions should use static binding.

Ephemeral binding

When you configure virtual ports with ephemeral binding, the host only creates a port and assigns it to a VM when the VM is turned on and its NIC is connected. Subsequently, when you turn off the VM or disconnect its NIC, the port is deleted.

Ephemeral binding doesn't rely on vCenter availability—either ESXi or vCenter can assign or remove ports—so administrators have increased flexibility in managing their environment, particularly in a disaster recovery situation. To preserve performance and scalability, only use ephemeral ports for disaster recovery, not for production workloads.

Note: Don't configure ephemeral binding for the Nutanix CVM's management port group. This configuration causes the CVM's upgrade process to fail or stop responding. The Nutanix Cluster Check (NCC) service notifies administrators when ephemeral binding is set up for the CVM management port group.

11. Network Configuration Best Practices

When you outline the logical networking attributes of any vSphere design, pay close attention to the physical networking requirements and to the impact of configuration decisions.

> Note: The following recommendations and considerations apply to all switch vendors; however, consult your switch vendor's implementation guides for specifics on how to enable these features and functionalities.

Physical Network Layout

- Use redundant top-of-rack switches in a leaf-spine architecture. This simple, flat network design is well suited for a highly distributed, shared-nothing compute and storage architecture.
  - If you need more east-west traffic capacity, add spine switches.
- Add all the nodes that belong to a given cluster to the same layer 2 network segment.
- Use redundant 40 Gbps (or faster) connections to ensure adequate bandwidth between upstream switches.
- When you implement a three-tier networking topology, pay close attention to the oversubscription ratios for uplinks while also taking into consideration that the spanning tree blocks ports to prevent network loops.

Upstream Physical Switch Specifications

- Connect the 10 GbE or faster uplink ports on the ESXi node to switch ports that are nonblocking, datacenter-class switches capable of providing line-rate traffic throughput.
- Use an Ethernet switch that has a low-latency, cut-through design and provides predictable, consistent traffic latency regardless of packet size,

traffic pattern, or the features enabled on the 10 GbE interfaces. Port-to-port latency should be no higher than two microseconds.

- Use fast-convergence technologies (such as Cisco PortFast) on switch ports connected to the ESXi host.
- To prevent packet loss from oversubscription, avoid switches that use a shared port-buffer architecture.
- Enable LLDP or CDP to assist with troubleshooting and operational verification.
- Implement a BPDU guard to ensure that devices connected to the STP boundary don't influence the topology or cause BPDU-related attacks.
  - Enable the BPDU filter on the ESXi hosts connecting to the physical switch.
  - Don't use this feature in deployments where you might wish to run software-based bridging functions in VMs across multiple vNICs. For more information, see VMware KB 2047822.

Link Aggregation

There are two possible ways to implement link aggregation: LAG (link aggregation group) and LACP (link aggregation control protocol). When you aggregate links, you bundle two or more physical links between a server and a switch or between two switches to increase the overall bandwidth and provide link redundancy. LAG is a static configuration, while LACP is a control protocol for the automatic negotiation of link aggregation. Both methods are functions of the vDS and support vSphere versions 5.1 and later.

Important notes regarding link aggregation:

- An ESXi host can support up to 32 LAGs. However, in real-world environments, the number of supported LAGs per host depends on the number of ports that can be members of an LACP port channel in the physical switch.
- When you configure LACP on the physical switch, the hashing algorithm must match what is configured in ESXi's vDS.

- You must configure all physical NICs connected to the LACP port channel with the same speed and duplex settings.
- Consult VMware's KB 1001938 for a complete explanation of LACP and its compatibility with vSphere versions.
- Set the mode for all members of the LACP group to Active. This setting ensures that both parties (ESXi host and switch) can actively send LACP data units and successfully negotiate the link aggregation.

Switch and Host VLANs

- Keep the CVM and ESXi host in the same VLAN. By default, the CVM and the hypervisor are assigned to VLAN 0, which effectively places them on the native untagged VLAN configured on the upstream physical switch.
- Configure switch ports connected to the ESXi host as VLAN trunk ports.
- Configure any dedicated native untagged VLAN other than 1 on switch ports facing ESXi hosts to carry only CVM and ESXi management traffic.

Guest VM VLANs

- Configure guest VM networks on their own VLANs.
- Use VLANs other than the dedicated CVM and ESXi management VLAN.
- Use VM NIC VLAN trunking (virtual guest tagging) only in cases where guest VMs require multiple VLANs on the same NIC. In all other cases, add a new VM NIC with a single VLAN in access mode to bring new VLANs to guest VMs.

CVM Network Configuration

- Don't remove the CVM from the internal vSS-Nutanix.
- If required for security, use the network segmentation feature to add a dedicated CVM backplane VLAN with a nonroutable subnet. This

segmentation separates CVM storage backplane traffic from CVM management traffic.

IP Address Management

- Coordinate the configuration of IP address pools to avoid address overlap with existing network DHCP pools.

IPMI Ports

- Don't allow multiple VLANs on switch ports that connect to the IPMI interface. For management simplicity, only configure the IPMI switch ports as access ports in a single VLAN.

12. Conclusion

Virtual networking in a vSphere 6.x environment plays an important role in infrastructure availability, scalability, performance, management, and security. With hyperconverged technology, customers need to evaluate their networking requirements carefully against the various configuration options, implications, and features in the VMware vSphere 6.x hypervisor.

In most deployments, Nutanix recommends vDS coupled with NIOC (version 2) and load-based teaming, as this combination provides simplicity, ensures traffic prioritization in the event of contention, and reduces operational management overhead.

With a strategically networked vSphere deployment on Nutanix, customers can be confident that their networking configuration conforms to field-tested best practices.

For feedback or questions, contact us using the Nutanix NEXT Community forums.

13. Appendix

References

1. VMware vSphere Documentation
2. VMware vSphere on Nutanix
3. Performance Evaluation of NIOC in VMware vSphere 6.0
4. Securing Traffic Through Network Segmentation

About Nutanix

Nutanix is a global leader in cloud software and a pioneer in hyperconverged infrastructure solutions, making clouds invisible and freeing customers to focus on their business outcomes. Organizations around the world use Nutanix software to leverage a single platform to manage any app at any location for their hybrid multicloud environments. Learn more at www.nutanix.com or follow us on Twitter @nutanix.

List of Figures