VMware® and Arista® Network Virtualization Reference Design Guide for VMware® vSphere Environments

Deploying VMware NSX with Arista's Software Defined Cloud Networking Infrastructure

Table of Contents

Executive Summary

Disruptive changes in server and network virtualization are revolutionizing modern data center capabilities, enabling greater flexibility in provisioning of workloads, higher efficiencies in use and placement of resources, and breakthroughs in mobility and portability of applications and services. Entire data center environments can now be defined in software and replicated as required to support specific tenants and workloads - a concept described here as the Software Defined Data Center (SDDC).

Where new applications and services place different and often dynamic demands on the compute, network and storage platforms contention can occur - leading to inefficient or unreliable services and poor end-user experience. Enhanced visibility can enable proactive responses to these conditions, and customizable automation of the underlay transport and overlay network can be used to ensure these new and complex topologies work efficiently and reliably. In the hyper-dynamic environment of the modern data center, the underlay transport network and the overlay network virtualization solutions are co-dependent actors in the delivery of optimal performance, reliability and scale.

To best utilize the new capabilities of a Software Defined Data Center while providing maximum transparency and performance, the underlying physical network must scale linearly and programmatically interface in a seamless manner with new network virtualization capabilities with very little contention and minimal end-to-end latency.

This white paper presents a design approach for implementing VMware's NSX network virtualization platform with Arista's Software Defined Cloud Networking (SDCN) infrastructure for optimal efficiency, reliability, scale and migration. The combined solution of Arista networking platforms with Arista's Extensible Operating System (EOS) and VMware NSX network virtualization platform provides a seamless path to an advanced SDDC.

Intended Audience for this Paper

This document is intended for virtualization and network architects interested in deploying VMware® NSX network virtualization solutions for vSphere with Arista® Networks' data center switches.

Enabling the Software Defined Data Center

VMware pioneered the Software Defined Data Center (SDDC) to transform data center economics, increase business agility and enable IT to become more efficient, and strategic to the organizations they serve. SDDC is defined by three pillars, virtualized compute, virtualized storage and virtualized network. Server and storage virtualization have already delivered tangible benefits including reduced operational complexity, fast provisioning of applications and dynamic repurposing of underlying resources, but the network has not kept pace until the arrival of network virtualization.

Network virtualization is the ability to separate, abstract and decouple the physical topology of networks from a ‘logical’ or ‘virtual’ topology by using encapsulated tunneling. This logical network topology is often referred to as an ‘Overlay Network’. VMware NSX provides network virtualization, the third critical pillar of the SDDC.

Similar to server virtualization, NSX network virtualization programmatically creates, snapshots, deletes, and restores software-based virtual networks. This transformative approach to networking delivers greater agility and economics while allowing for a vastly simplified operational model in the underlying physical network. NSX is a non-disruptive solution and can be deployed on any IP network, from existing traditional networking models to next generation fabric architectures. No physical network infrastructure changes are required to immediately get started with the Software Defined Data Center.

SDDC enables new infrastructure design options that have the potential to improve and scale the performance of applications. Figure 1 highlights the similarities between server and network virtualization.

With server virtualization, a software abstraction layer (i.e., server hypervisor) reproduces attributes of an x86 physical computer (e.g., CPU, RAM, disk, NIC) in software, allowing for programmatic assembly in any arbitrary combination to produce a unique virtual machine (VM) in a matter of seconds.

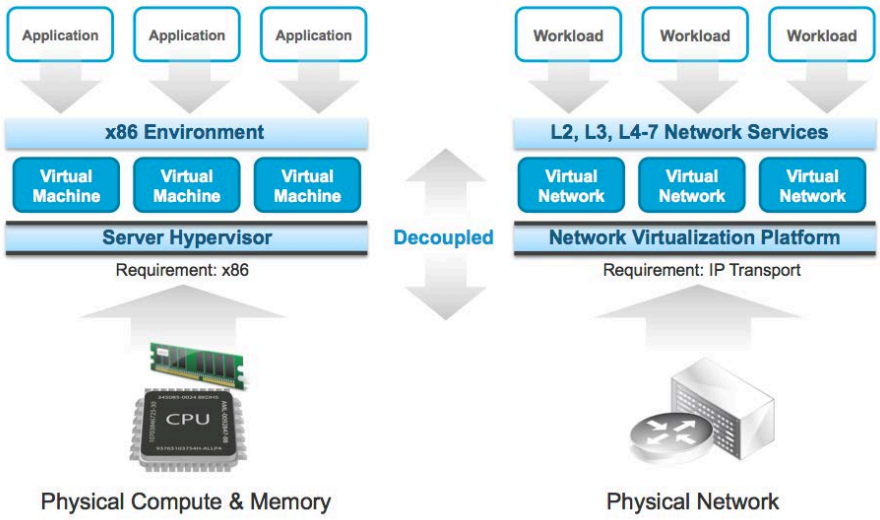

Figure 1. Server and Network Virtualization Analogy

With network virtualization, the functional equivalent of a "network hypervisor" reproduces the complete set of layer-2 to layer-7 networking services (e.g., switching, routing, access control, firewalling, QoS, load balancing, and visibility) in software. These services can be programmatically assembled to produce unique, isolated virtual networks.

Similar benefits are seen in both compute and network virtualization. Virtual machines are independent of the underlying x86 platform, allowing discrete physical hosts to be treated as a single pool of compute capacity. Virtual networks are independent of the underlying IP network, allowing the physical network to be treated as a single pool of transport capacity that can be consumed and repurposed on demand. Unlike legacy architectures, virtual networks can be provisioned, changed, stored, deleted and restored programmatically without reconfiguring the underlying physical hardware or topology. By matching the capabilities and benefits derived from familiar server and storage virtualization solutions, this flexible approach to networking unleashes the full potential of the Software Defined Data Center.

Combined solution using Arista EOS and VMware NSX

VMware and Arista Networks are aligned in their vision for the role of network virtualization for realizing the full potential of the Software Defined Data Center. VMware NSX works with any existing IP network, but the right coupling between NSX and the underlay network drives optimal data center benefits.

The combined Arista and VMware solution is based on Arista's data center class 10/40/100GbE networking portfolio with Arista EOS and VMware NSX Virtual Networking and Security platform.

At the core of the combined solution is the Arista Extensible Operating System (EOS) providing the industry's most advanced network operating platform. EOS combines modern-day software and O/S architectures, an open foundation for development with a standard Linux kernel, and a stateful publish/ subscribe in-memory database model to provide a real-time programmatic, seamless and automated model for cloud networking.

This paper will describe the use of VXLAN (Virtual extensible LAN) technology, an open multi-vendor standard that has been developed and adopted by industry leaders in network, switching, firewalling, load-balancing, WAN optimization and application delivery. VXLAN provides a consistent, multi-vendor connectivity model for implementing network virtualization.

Together, Arista EOS and VMware NSX provide the essential integration and programmatic capabilities to offer flexible workload placement and mobility for a true Software Defined Data Center.

The VMware NSX and Arista EOS combined solution offers the following benefits to deploying network virtualization within data centers built on the foundation of Arista's Software Defined Cloud Networking:

- Virtual and physical workloads can be connected on a common logical segment on-demand regardless of hypervisor, IP subnet or physical location
- Holistic views of the virtual and physical topology increase operational efficiency
- Network virtualization with NSX does not require IP multicast for learning or forwarding broadcast, unknown unicast or multicast packets
- A single point of management and control via NSX APIs and EOS APIs to configure the logical networks across hypervisors and the physical network fabric.

Arista EOS Extensibility

Core to successful implementation of Arista SDCN is the extensibility of Arista networking operating system. While the modularity, distributed scalability, and real-time database interaction capabilities of Arista EOS are mentioned throughout this document, there are other aspects to consider as well. These considerations include the ability to write scripts and load applications (such as third-party RPM Package Managers [RPMs]) directly onto the network operating system, and to run these applications as guest VMs. Arista provides a developer's site called "EOS Central" for customers that are interested in this hosting model.

Leveraging the extensibility model of Arista EOS several applications have been developed that dramatically enhance the availability, integration and transparency of the combined VMware NSX and Arista EOS solution:

Arista Smart System Upgrade

Arista Smart System Upgrade (SSU) is a series of patent-pending technologies that enable the network operator to seamlessly align one of the most challenging periods of network operations, the upgrade and change management operation, with the networks operational behaviors. The network, with SSU, is capable of gracefully exiting the topology, moving workloads off of directly-connected hosts, and aging out server load balancer Virtual IPs (VIPs) before any outage is ever seen. The multi-step, multi-hour process many network operators go through to achieve maximum system uptime becomes the default method of operations. SSU has demonstrated interoperability with F5 Load Balancers, VMware vSphere, OpenStack, and more.

Arista Network Telemetry

Lastly, Network Telemetry is all about accessing machine and wire data: generating, collecting, and distributing the telemetry data necessary to make well informed network decisions about where problems may be happening, thus ensuring the data is available and easily reachable and indexed so these hot spots, or problem areas, are rapidly fixed and troubleshooting is simple and quick. Network Telemetry programmatically interfaces with VMware vCenter Log Insight, Splunk, and several other log management and rotation/indexing tools and provides a rich source of operational visibility.

EOS: the Network Operating System designed to support Cloud Networking

An open modular network operating system with the ability to respond in real time to both internal and external control operations is required to support SDN, cloud and SDDC. Unfortunately, not all switch operating systems offer this capability because many of them were architected a decade or more ago, before

the need for hyper-dynamic cloud environments, and the interaction with external controllers was not envisioned.

These older operating systems typically interact internally through a proprietary message-passing protocols and externally with non real-time state information (or APIs). Many configuration, forwarding, race, and state problems arise when multitasking occurs in real time with multiple dynamic systems, as in the case of communicating with external controllers while trying to resolve topology changes. The message-passing architectures of these legacy switches prevent these operating systems from quickly and reliably multitasking with external controllers in dynamic cloud environments.

A modular network operating system designed with a real-time interaction database, and with API-level integration both internally and externally, is a better approach. The system can, therefore, integrate and scale more reliably. In order to build a scalable platform, a database that is used to read and write the state of the system is required. All processes, including bindings through APIs, can then transact through the database in real time, using a publish and subscribe message bus. Multiple systems, both internally and externally, can subscribe, listen, and publish to this message bus. A per-event notification scheme can allow the model to scale without causing any inter-process dependencies.

Closed network operating systems that are built on older design principles can, at best, offer one-off implementations and struggle to support the growing list of different SDN controller form factors. Arista, on the other hand, is in a unique leadership position—the industry award-winning modular Arista EOS software platform can interact with multiple virtualization and cloud orchestration systems concurrently, handling external controller updates and managing highly distributed switch forwarding states, both in real time. The Arista approach to Software Defined Cloud Networking offers the best of both worlds, providing service control to external controllers, while scaling with Leaf-Spine switching architectures for the most demanding enterprise and carrier-class software-defined cloud data centers.

VMware's NSX Network Virtualization Platform

VMware NSX is a network virtualization platform that delivers the operational model of a virtual machine for the network. Virtual networks reproduce the network model in software, allowing complex multi-tier network topologies to be created and provisioned programmatically in seconds. NSX includes a library of logical networking services – logical switches, logical routers, logical firewalls, logical load balancers, logical VPN, QoS, and distributed security.

A self-service interface allows users to create custom combinations of these services in isolated software-based virtual networks that support existing applications without modification, or that can deliver unique requirements for new application workloads on-demand. Similar to virtual machines in compute, virtual networks are programmatically provisioned and managed independent of networking hardware. Decoupling from hardware introduces agility, speed and operational efficiency that has the power to transform data center economics.

Components of the NSX Platform

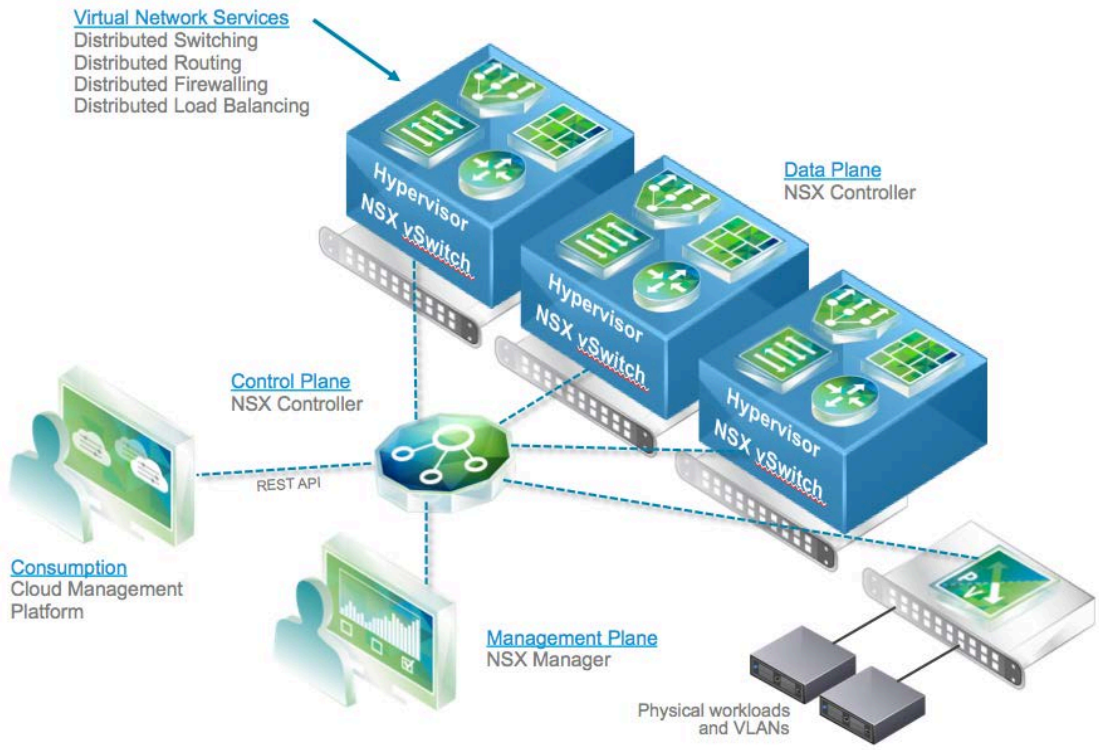

Figure 2. VMware Network Virtualization Platform Components

Consumption

Consumption of NSX is enabled directly via the NSX manager through the Web UI. In a vSphere environment this is available through the vSphere Web UI. Network virtualization is typically tied to a Cloud Management Platform (CMP) for application deployment. NSX provides a rich set of integration features to connect into virtually any CMP via the REST API. Out-of-the-box integration is also available through VMware vCloud Automation Center and vCloud Director.

Management Plane

The NSX manager builds the NSX management plane. The NSX manager provides the single point of configuration and REST API entry points in a vSphere environment for NSX.

Control Plane

The NSX control plane exists solely within the NSX controller. In a vSphere-optimized environment with the vSphere Distributed Switch (VDS), the controller enables multicast-free VXLAN and control plane programming for elements such as Logical Distributed Routers (LDRs).

The NSX controller is essentially contained within the control plane; no data plane traffic passes through it. Controller nodes are deployed in redundant and distributed clusters to enable high-availability and scale. With the NSX controller deployment model, any failure of a single controller node will not impact data plane traffic.

The NSX Control VM component performs Dynamic Routing Control plane functions, peering with NSX Edge Gateways and communicating Routing Protocol updates to the NSX Controller Cluster.

Data Plane

The NSX data plane is managed by the NSX vSwitch. The vSwitch in NSX for vSphere is based on the standard vSphere VDS with additional components that enable rich services. Add-on NSX components include kernel modules (VIBs) running within the hypervisor kernel, providing services that enable distributed routing, distributed firewall, and VXLAN bridging capabilities.

The NSX VDS vSwitch abstracts the physical network and provides access-level switching in the hypervisor. It is central to network virtualization, enabling logical networks that are independent of physical constructs such as VLANs. Benefits of the VDS vSwitch include:

- Support for overlay networking, leveraging the VXLAN protocol and centralized network configuration including -
  - ✓ Creation of a flexible logical layer-2 overlay over existing IP networks on existing physical infrastructure without the need to re-architect any of the data center networks
  - ✓ Provisioning of communication (east-west and north-south) while maintaining isolation between tenants
  - ✓ Operation of application workloads and virtual machines that are agnostic of the overlay network, as if they were connected to a physical layer-2 network
- Significant hypervisor scalability
- Multiple visibility and management features - including port mirroring, NetFlow/IPFIX, configuration backup and restore, network health check, QoS, and LACP – to provide a comprehensive toolkit for traffic management, monitoring, and troubleshooting within a virtual network.

The data plane also consists of network virtualization gateways, which provide layer-2 bridging from the logical networking space (VXLAN) to the physical network (VLAN). The gateway device is typically an NSX Edge virtual appliance, which offers services including layer-2, layer-3, perimeter firewall, load balancing, SSL VPN, and DHCP.

Functional Services of NSX for vSphere

This design guide details how the components described provide the following functional services:

- **Logical Layer-2 Connectivity.** Enabling extension of an layer-2 segment/IP subnet anywhere in the fabric irrespective of the physical network design.
- **Distributed Layer-3 Routing.** Routing between IP subnets can be done in a logical space without traffic touching the physical router. This routing is performed directly in the hypervisor kernel with minimal CPU/memory overhead. This functionality provides an optimal data path for routing traffic within the

virtual infrastructure. Similarly, the NSX Edge provides a mechanism for fully dynamic route peering using OSPF, BGP, or IS-IS with the physical network to enable seamless integration.

- **Distributed Firewall.** Security enforcement is done at the kernel and VNIC level, allowing firewall rule enforcement to scale in an efficient manner without creating bottlenecks on physical appliances. The firewall is distributed in kernel, creating minimal CPU overhead and allowing line-rate performance.
- **Logical Load-balancing.** Support for layer-4 to layer-7 load balancing with the ability to do SSL termination.
- **VPN Services.** SSL VPN services to enable layer-2 VPN services.

Why Deploy VMware NSX with Arista Network's Infrastructure

Arista's scale-out cloud network designs provide an ideal platform for deployment of NSX network virtualization, utilizing principles that have made both cloud computing and software-defined networking compelling. All Arista reference designs revolve around common central design goals.

Simplified Standards-based Architecture: Arista is an established leader in data center network architectures designed for consistent performance, deterministic latency, and easy troubleshooting regardless of workload and deployment sizes. Arista's Multi-Chassis Link Aggregation (MLAG) and Equal Cost Multipath (ECMP) routing are standards-based approaches used for scalable cloud networking designs and take the place of proprietary fabrics. These design fundamentals ensure effective use of all available network bandwidth, provide non-blocking active-active forwarding and redundancy, and enable excellent failover and resiliency characteristics. For any SDDC, MLAG and ECMP cover all of the important multipath deployment scenarios in a practical manner without introducing any complicated or proprietary lock-in.

Massively Scalable: Reference designs are based on open standards for building out horizontally scalable networks from the smallest of pod sizes to hyper-scale designs. Universal cloud networks may be built on layer-2 or layer-3 multi-pathing technologies (leveraging MLAG or ECMP routing) for a scalable, standards-based approach that does not compromise workload performance. Together these technologies cover all important multi-path deployment scenarios without introducing any proprietary protocols or design elements. Implementations in these reference designs can scale linearly from small enterprise deployments to large cloud provider networks.

Open and Programmable: Arista's EOS (Extensible Operating System), is a programmable network operating system based on a universal single image across all Arista products. Arista EOS provides extensibility at every level of the network. Central features include: a self-healing resilient in-memory state database; true on-switch access to a standard Linux operating system, advanced Event Manager (AEM) to trigger custom automations; custom Python scripting and programming environment; and direct JSON application interfaces via EOS Application Programming Interface (eAPI).

Consistent and Interoperable: All Arista switches use the same Arista EOS across the entire Arista product portfolio, allowing certification and tracking of a single software image for the entire network. Arista switches and designs use standard open protocols including spanning tree, Link Aggregation Control Protocol (LACP), Open Shortest Path First (OSPF), and Border Gateway Protocol (BGP) for interoperability with other networking systems.

Maximizing Throughput: Modern operating systems, network interface cards (NICs), and scale-out storage arrays make use of techniques such as TCP segmentation offload (TSO), generic segmentation offload (GSO) and large segment offload (LSO) to push more data onto the network. These techniques are fundamental to reducing CPU cycles on servers and storage arrays when sending large amounts of data. A side effect of these techniques is that systems that need to transmit large blocks of data will offload processing to it's NIC, which must slice the data into segments and put them on the wire as a burst of back-to-back frames at line-rate.

If more than one of these is destined to the same network destination, microburst congestion within the network could occur causing significantly reduced end-to-end throughput. This can be exceedingly difficult to troubleshoot. Common approaches to dealing with the impact of microbursts include over-provisioning and reducing traffic fan-in by design. An alternative approach is to deploy Arista switches with deep buffers at the Spine-layer to absorb the bursts that could otherwise result in frequent packet loss and inefficiency. Arista's deep buffer architectures are better at handling a variety of traffic loads and dynamically managing per-port packet memory allocation to avoid packet loss due to microbursts.

Support for Integration with Existing Legacy Infrastructure

In reality not all resources will be virtualized in the SDDC. This may be due to specific performance or latency-sensitive demands of specific applications, like databases or layer-4 to layer-7 services like load balancers or firewalls. In addition, during migration to the SDDC many existing storage and compute resources may need to be incorporated into the virtualization infrastructure. This is easily accomplished with Network virtualization gateways, mentioned earlier, which can provide VXLAN Tunnel Endpoint (VTEP) termination at a VLAN or physical port boundary. Gateways create an on-ramp for existing physical infrastructure components to tie into the virtualized NSX overlay network.

Either VMware's NSX Edge virtual appliance or Arista's EOS based switches can support network virtualization gateways based on software and/or hardware-based forwarding respectively. Increasing bandwidth demands that are driven as the SDDC implements more 10/40/100Gbps connectivity, will drive demand for scalable gateways that can provide terabits-per-second of aggregate bandwidth across many network segments. This is achievable in an edge switching platform which interfaces with NSX. Arista and VMware use the same set of industry standard protocol interfaces for programmatic interfacing between EOS and NSX to allow both software-based and hardware-based gateways to operate seamlessly and in concert.

VMware's NSX Manager front-ends the entire NSX network virtualization platform and acts as the single pane of glass for rolling out your SDDC network. In addition, NSX provides a northbound REST API to further integrate with multiple orchestration platforms. This speeds up service delivery and helps businesses better address their needs in a programmatic and automated way across data centers.

Physical Network Design Considerations

VMware NSX network virtualization can be deployed over existing data center networks. This section discusses how the logical networks using VXLAN technology, network virtualization gateways, and multi-faceted NSX integration can be deployed over most common data center network topologies. Topics covered include the requirements on the physical network, optimal network designs for effective network virtualization, scalability of resulting logical networks, and related services such as security and visibility.

Differences in customer environments will drive specific implementation decisions in physical network design. With network virtualization in place, compute and storage systems have very simple needs. They require non-blocking, deterministic access to any resource in the network regardless of underlying topologies. This simple requirement is forcing network architects to re-evaluate traditional network hardware designs while paying close attention to the cost-effective and powerful standardized offerings in the marketplace.

A good data center network design should:

- Provide optimal utilization of all hardware resources
- Provide consistent performance between any two points
- Simplify the provisioning of resources
- Clearly communicate the available capacity
- Limit the scope of failure domains
- Create an active/active infrastructure for east-west and north-south traffic
- Guarantee deterministic performance and latency
- Use open standards

The following section discusses reference designs for network virtualization:

- Layer-2 or Layer-3 Spline™ designs
- Layer-2 leaf-spine designs using MLAG
- Layer-3 leaf-spine designs using standard routing protocols and ECMP

Physical Network Design Choices

Network designs should not be limited to short term requirements but should be based on longer-term considerations for application infrastructure and workloads. Optimal network planning should target a design that meets future goals for port density at the network edge, while not exceeding a maximum level of oversubscription required to assure the availability of full bisectional bandwidth to applications. Ideally the network equipment deployed today should be able to fulfill the needs of scaling up the architecture to higher speeds and higher densities of time. For this reason, all Arista switches are suitable for placement in any position in the scenarios discussed in this section and have the same software support for both layer-2 and layer-3 services with a single system image across every model and design.

The following are some basic rules of thumb about how to make design decisions in a deployment strategy for the physical network that will avoid redesign or equipment substitution when scaling up:

- If long-term port-count requirements can be fulfilled in a single switch (or pair of switches in a high-availability MLAG design) then a single-tier Spline™ design may be used.

  Spline designs combining the roles of both a leaf and spine in a single high-density Arista switching platform such as the Arista 7300X. Modern single-tier designs can connect up to 2048 dual-attached physical servers and offer the lowest CAPEX and OPEX of all choices resulting from the use of a minimal number of network switches, less cabling and optics, and no wasted ports for interconnecting between tiers.

- Where a single-tier solution will not meet long-term requirements, because of the anticipated scale, a two-tier design will provide a close approximation of single-tier performance characteristics with much greater scale.

  For example, a combination of 16 Arista 7300X switches in a spine with 7150S layer-3 leaf switches scales the network to over 32K 10G Servers in a fully non-blocking, low-latency, two-stage network that provides predictable and consistent application performance. The flexibility of leaf-spine multipath design options combined with support for open standards provides maximum flexibility, scalability and network wide virtualization.

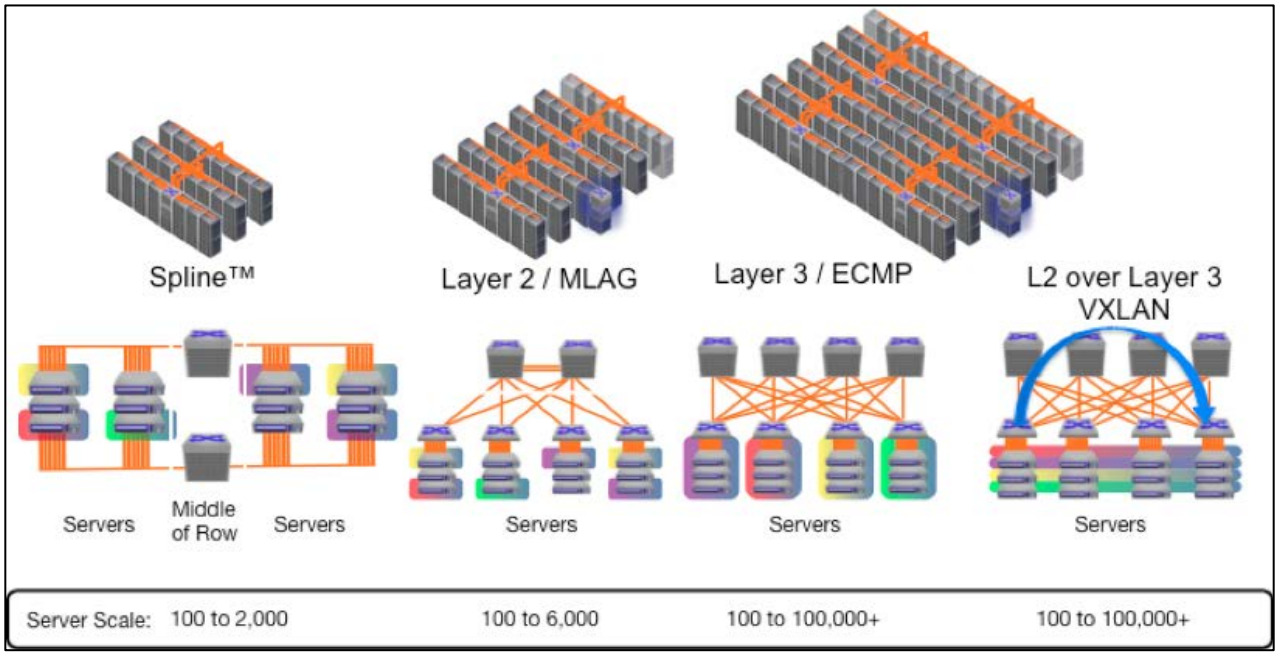

Figure 3. Arista SDCN Network Designs

Recommended Arista Platforms for Each Design Type

Any Arista modular or fixed configuration switch can be used as either the Leaf or Spine switch in a Spline or two-tier leaf-spine design.

- Spline designs usually depend upon the higher density modular switches like the Arista X-series 7300 platform with up to 2048 ports of 10GbE.
- In leaf-spine deployments, the choice of Spine switch port-density, link speeds, and oversubscription ratios determine the maximum scalability and performance of the design, so the best design choice is often the Arista E-series with support for dense 10/40/100GbE connectivity and large buffers to optimize aggregate performance.
- Leaf switches are typically selected based on the local connectivity requirements of equipment in the racks and may need to provide network virtualization VTEP gateway functionality in hardware. For Leaf deployments customer will typically select one of Arista's 7x50-series switches.

Regardless of these factors, any Arista switch can be deployed in either Spline, Leaf or Spine position in the network with either layer-3 or layer-2 forwarding topologies because they all support full wire-speed performance, full MLAG and ECMP features, and use a single consistent Arista EOS software image in all operating modes.

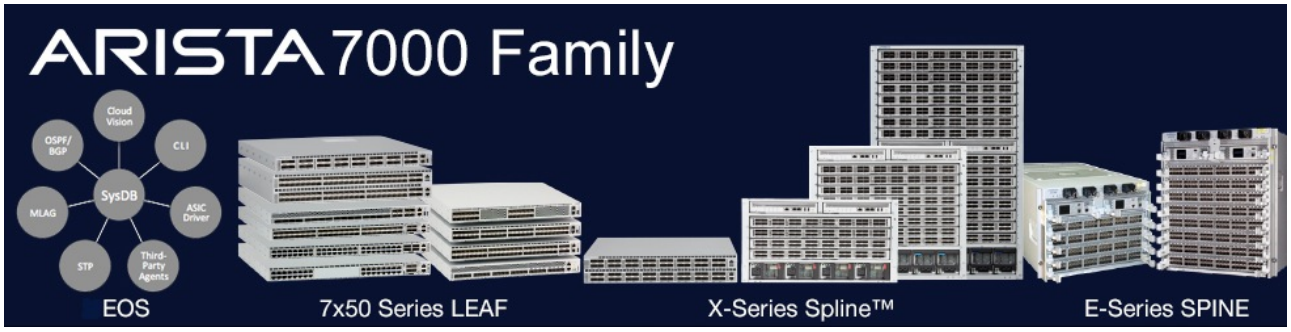

Figure 4. The Arista 7000 Family Product Line

Arista Spline™ Network Designs

Spline designs collapse what have historically been distinct spine and leaf tiers into a single pair of switches, thus providing a single tier, or single switch hop, between any two points on the network. In a Spline design, ports are not wasted for inter-device connectivity, so single tier Spline designs will offer the lowest CAPEX and OPEX per port and deliver the lowest latency in an inherently non-oversubscribed architecture with at most two management touch points.

Arista 7300 Series (4/8/16 slot modular chassis), Arista 7250X Series (64x40G to 256x10G 2U Fixed switch) and Arista 7050X Series (32x40G to 96x10G + 8x40G) switches are ideal for Spline network designs, providing from 96 to 2048 10G ports in a single switch. Spline designs cater to data centers with needs ranging from 3 to 49 racks. Flexible airflow options (front-to-rear or rear-to-front) on all modular Spline switches allow their deployment in server/compute racks in the data center with airflow that matches the thermal containment of the servers.

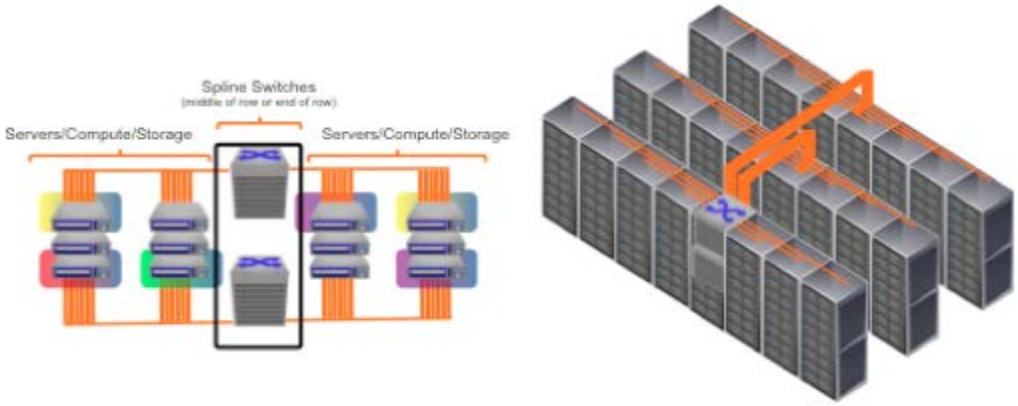

Figure 5. Spline™ Switch Placement

Arista Leaf-Spine Network Designs

For designs that do not fit within a single-tier Spline design, a two-tier leaf-spine design is the next step. Two-tier designs place all devices and services (e.g., servers, compute, storage, appliances) on a layer of leaf

switches. The leaf layer may consist of top-of-rack deployments or mid/end-of-row designs with each leaf uplinked to two or more spine switches.

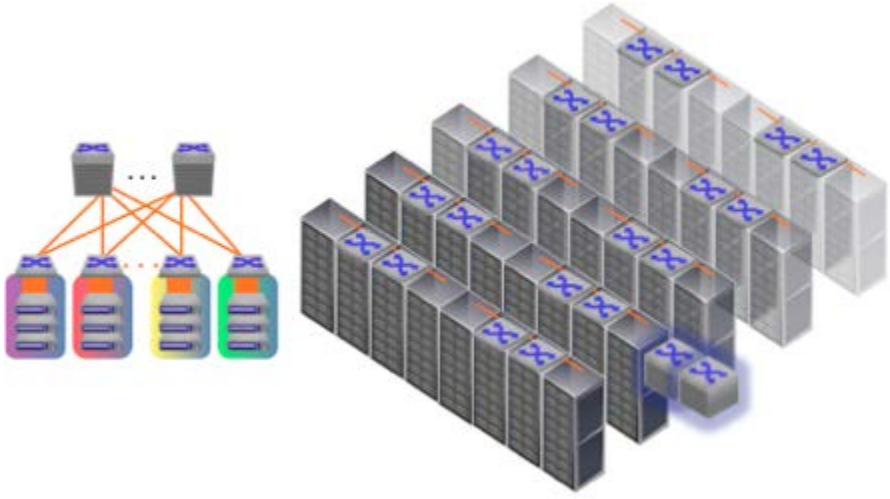

Figure 6. Two-tier Leaf-Spine Design

Scale out designs start with one pair of spine switches and some quantity of leaf switches. A two-tier leaf-spine network design at 3:1 oversubscription has 96x10G ports for servers/compute/storage and 8x40G uplinks per leaf switch (example: Arista 7050SX-128 – 96x10G : 8x40G uplinks = 3:1 oversubscribed).

Two-tier leaf-spine network designs enable horizontal scale-out with the number of spine switches growing linearly as the number of leaf switches grows over time. The maximum scale achievable is a function of the port density of the spine switches, the number of physical uplinks that can be supported from each leaf switch to the spine, and desired oversubscription ratio.

Choosing Between Layer-2 or Layer-3 Leaf-Spine Topologies

Two-tier leaf-spine networks can be built using either layer-2 or layer-3 forwarding topologies with the same Arista network equipment, because all Arista switches support wire-speed forwarding in all modes.

Layer-3 designs are able to scale substantially further than layer-2 MLAG designs through the deployment of a larger number of potential leaf and spine switches, larger tables, and the efficiencies of hierarchical layer-3 routing between the tiers.

> *To ensure consistent performance and latency, easier troubleshooting, and greater scale-out across deployments, Arista recommends a layer-3 leaf-spine (L3LS) design approach for most customers.*

Deploying Layer-3 Leaf-Spine Designs

Figure 7 presents an architecture consisting of a series of leaf switches residing at top of each rack connecting to a multi-switch spine layer. The spine layer acts as a high performance IP switching fabric for interconnecting the various leaf nodes.

This model provides flexibility to scale out horizontally. When the requirements for more ports or bandwidth arise, more leaf and spine switches may be added without redesigning the network. In addition, Arista

platforms can provide seamless migration from 10G to 40G/100G with technologies including the unique tri-speed 10/40/100G module on the 7500E-series of switches.

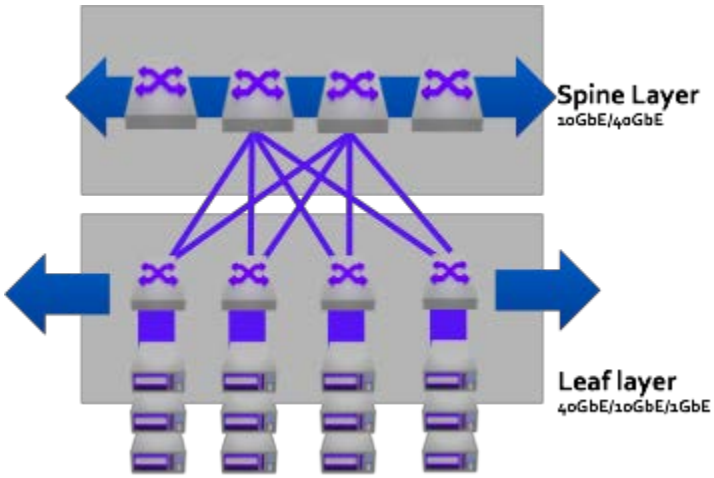

Figure 7. Layer-3 leaf-spine Architecture for Consistent Performance, Subscription, and Latency between Racks

With the overlay topology decoupled from the physical underlay topology, the layer-3 leaf-spine architecture can leverage mature, proven, and scalable layer-3 topology and routing protocols. Customers benefit from open standard-based layer-3 routing between the leaf and spine switches; equal cost multipathing (ECMP) provides traffic load balancing across the multi-switch spine and increases horizontal scale with support for up to 64 spine switches.

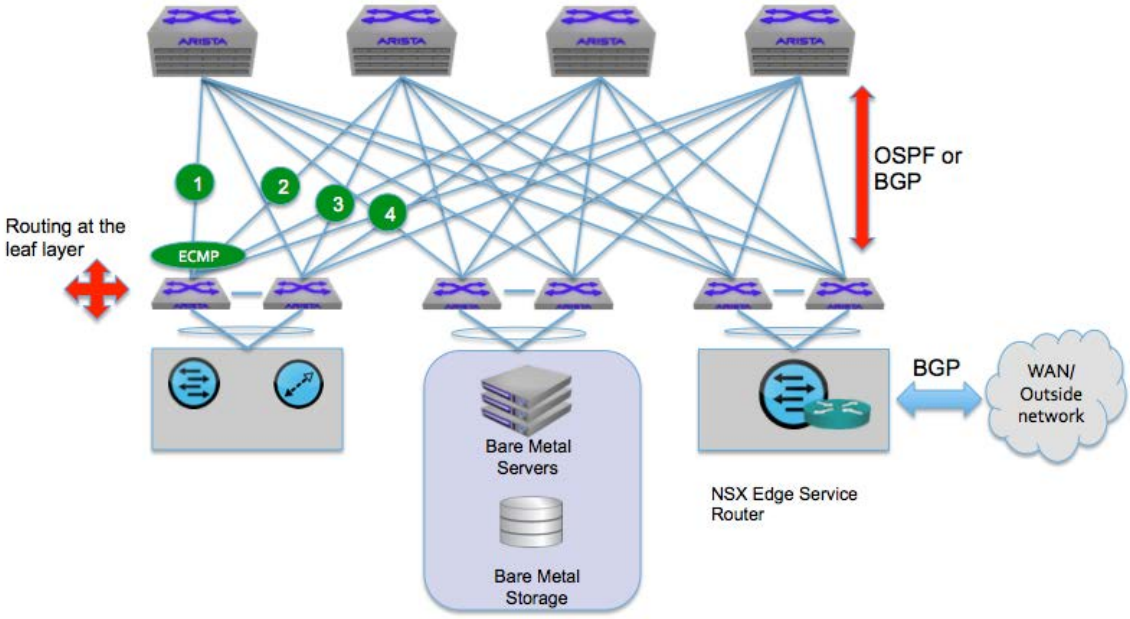

Figure 8. Open Standard Based Layer-3 leaf-spine for Horizontal Scale

Network Routing Protocol Considerations for Layer-3 Designs

The choice of network routing protocol has direct impact on performance. Open Shortest Path First (OSPF) is the most widely used link-state routing protocol in typical leaf-spine implementations. A single area design encompassing all of the leaf and spine nodes can scale to a relatively large environment. OSPF is a well-understood link-state protocol that is simple to configure and troubleshoot; however, there are a few key points to consider.

Control Plane Activity: The normal behavior of any link-state protocol is to flood the link-state updates. All neighboring nodes need to process these link-state updates and send acknowledgements. As the network grows, these link-state updates increase exponentially. It is important that designs using link-state routing protocols like OSPF consider the switching platforms capability for generating, absorbing, and re-calculating shortest path first (SPF) without compromising network stability.

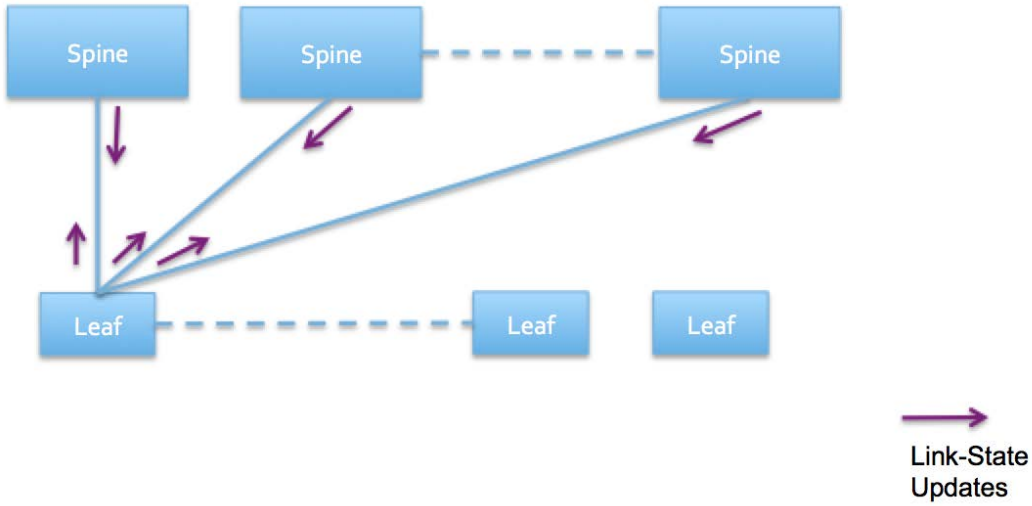

Figure 9. Link-Sate Updates in OSPF

Transient Leaf Transit Behavior: It is important to carefully implement link-state protocols like OSPF in large-scale networks. Bad implementation of link-state protocols can introduce momentary leaf transit scenarios during protocol state transitions. Such brief traffic patterns temporarily disrupt predefined oversubscription ratios and can be very hard to characterize given their short-lived nature. Tuning link-state timers and utilizing advanced features and functionality like Stub areas and summarization can help avoid such transient leaf behavior.

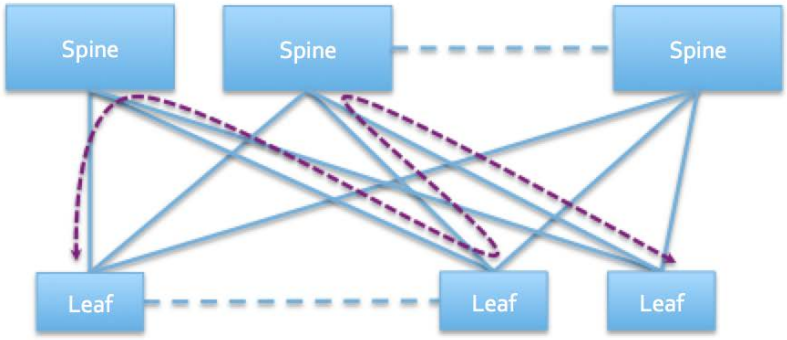

Figure 10. Intermittent leaf-spine Transit Behavior in Link-State Routing Protocol

Alternatively, BGP can be used as routing protocol for a leaf-spine network design. Consider a design where the spine nodes are configured to be in a single BGP autonomous system while each leaf-node is assigned a unique private autonomous system (AS) number. The private AS range of 64512-65535 is available for such designs, allowing up to 1023 AS numbers to be assigned within a data center. BGP's built-in loop suppression capabilities prevent unnecessary churn and decreases control plane overhead in large-scale networks. This design also eliminates the leaf transit scenario and significantly reduces the amount of control plane traffic that would otherwise result when nodes are rebooted, upgraded, or experience link flaps.

Key difference between any link state protocol and BGP is scale. Link state protocol are N^2 in nature and hence control plane overhead increases dramatically with scale. BGP is path vector protocol and hence control plane activity increases in order of N. On the other hand, link-state protocol like OSPF and IS-IS are easier to implement and troubleshoot.

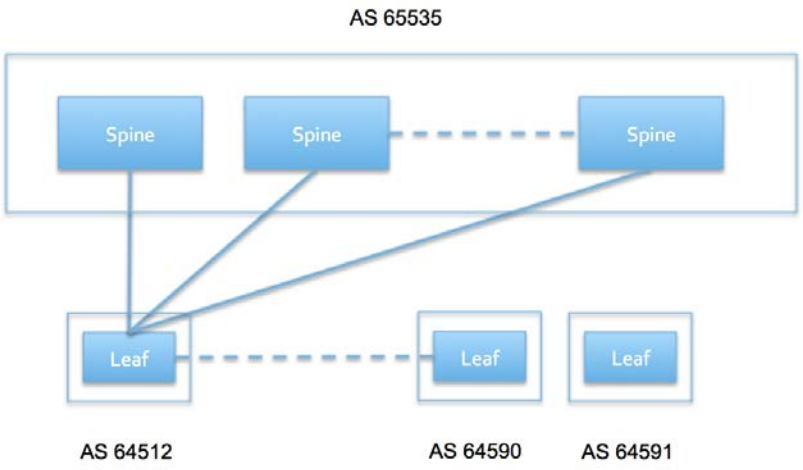

Figure 11. BGP Protocol in a leaf-spine Design

In a layer-3 leaf-spine physical network topology:

- Servers are typically dual-homed to a pair of top of rack (ToR) switches using standard LACP protocol. Arista switches provide active/active link aggregation to downstream nodes using MLAG.
- Each ToR switch acts as layer-3 gateway for the subnets assigned to the hypervisors while the tenant/application gateways reside in the logical space. Leaf switches provide IP connectivity to NSX Edge Gateways, compute hypervisors, and physical server infrastructure.
- ToR switches and spines can leverage open standard routing protocols to provide ECMP traffic load balancing across physical networks. VXLAN header also enables optimal load balancing of traffic across the ECMP spine infrastructure in hardware.
- A pair of leaf switches is also used to connect the NSX Edge Services Gateways and the NSX Logical Router Control VMs which provide necessary connectivity to outside networks.
- The spine switches on this network generally connect to a series of WAN routers, providing necessary connectivity to the NSX Edge Service Gateways.

Either modular or fixed configuration switches can be used as spine switches in a two-tier leaf-spine design. The choice of spine port-density determines the maximum scalability of these designs. Also, any Arista switch can be deployed in either leaf or spine position in the network with either layer-3 or layer-2.

Layer-2 MLAG Designs

Many customers choose to build large layer-2 networks, feeling they are easier to manage than layer-3 networks. A primary issue with creating a large layer-2 domains is the insufficient uplink bandwidth from each rack due to traditional spanning tree-based designs. Spanning tree typically blocks half the uplinks in order to avoid loops in network topology, thereby reducing the available bisectional bandwidth of the network by 50%. Arista's Multi-chassis Link Aggregation (MLAG) feature removes this inefficiency and allows the utilization of all interconnects in an active/active mode.

At the leaf of the network, servers with multiple interfaces connect to the switches using NIC bonding, or standard link aggregation. With two leaf switches in an MLAG pair, the server has active/active redundant connections to the network. The LAG groups can be formed using static link aggregation or LACP-based negotiation methods.

At the spine layer, two Arista switches can form an MLAG pair and aggregate all of the uplinks from the datacenter racks. This eliminates spanning tree blocked ports and allows use of interconnect bandwidth on all ports. MLAG at the spine layer significantly improves layer-2 scalability at no additional cost.

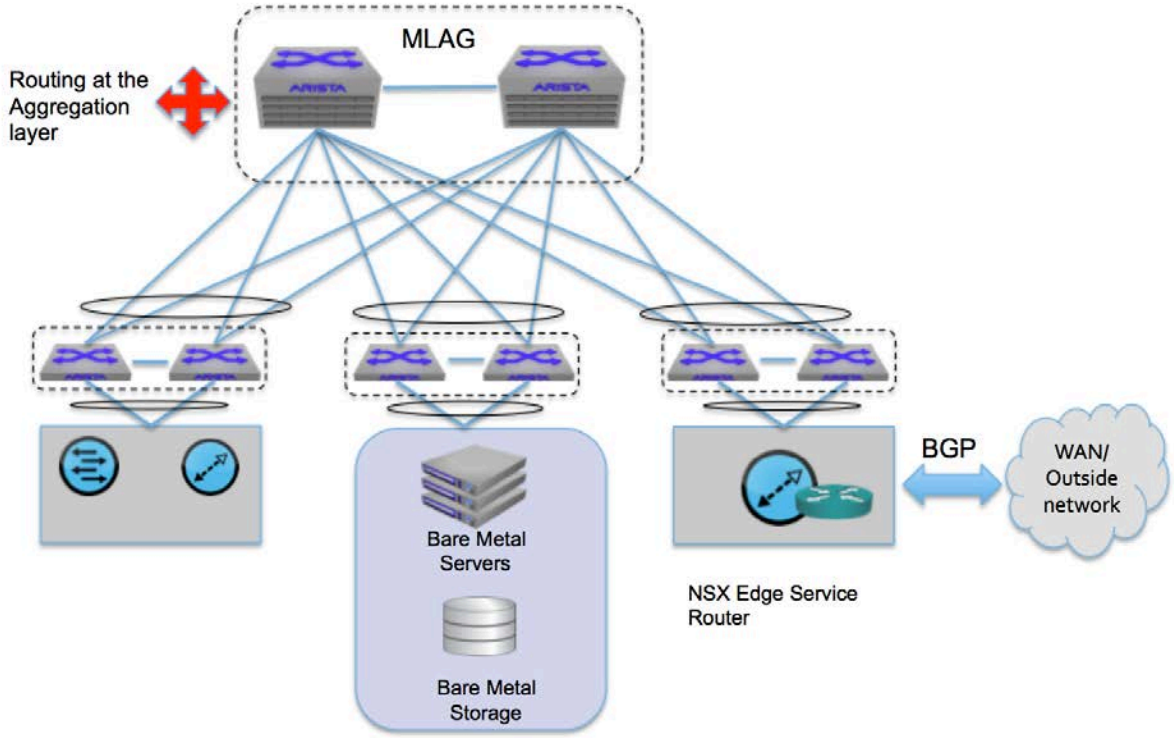

Figure 12. Layer 2 MLAG Topology for Layer 2 Design

In a layer-2 MLAG physical network topology:

- Each server forms an MLAG with a pair of ToR switches using standard LACP protocol. Arista switches provide active/active link aggregation to downstream servers using MLAG.
- The MLAG pair at the spine or aggregation layer acts as layer-3 gateway for hypervisors and provides IP connectivity to NSX gateways and controllers.
- The same leaf layer of Arista switches can serve as hardware gateways by means of translating between 802.1q and VNI tags. This allows physical hardware devices such as storage, databases, appliances, and non-VXLAN hypervisors to connect to the overlay with line rate hardware-based VXLAN encapsulation/de-encapsulation.
- A pair of leaf switches is also used to connect the NSX Edge Service Router VM. This Edge Service Router VM provides necessary connectivity to outside networks.
- This topology places all hypervisor and gateways one hop away.
- The MLAG pair at the spine connects to a series of edge routers and can provide necessary IGP or BGP connectivity to NSX Edge Service Router.
- Virtual Router Redundancy Protocol (VRRP) and Virtual ARP provide high availability layer-3 gateway services to hosts while allowing traffic to load balance across the redundant links.

Arista EOS Technologies used in the Designs

Arista's scale-out cloud network designs are based on a number of foundational features of Arista's award winning Extensible Operating System and are supported across all Arista switches.

Multi-Chassis Link Aggregation (MLAG)

MLAG enables devices to be attached to a pair of Arista switches (an MLAG pair) with all links running active-active and up to 4096 VLANs. MLAG eliminates bottlenecks, provides resiliency, and enables layer-2 links to operate concurrently via LACP or static configuration. This eliminates the usual 50% bandwidth loss seen with STP blocked links in layer-2 designs.

In MLAG designs, inter-VLAN traffic can be routed using a layer-3 anycast gateway located at the leaf in a Layer 3 Leaf-Spine topology or at the Spine in layer 2 MLAG topology. A layer-3 anycast gateway using virtual ARP (VARP) with MLAG enables the layer-3 gateway to operate in active-active and n-way active mode without the overhead of protocols like HSRP or VRRP. An MLAG pair of switches synchronizes forwarding state such that the failure of one node does not result in any protocol churn or state machine transition. MLAG's active-active mode ensures minimal network disruption.

Equal Cost Multipath Routing (ECMP)

Equal-cost multi-path routing (ECMP) is a standardized routing strategy where next-hop packet forwarding to a single destination can occur over multiple "best paths" which tie for top place in routing metric calculations. Multipath routing can be used in conjunction with most routing protocols, since it is a per-hop decision that is limited to a single router. In the data center it offers substantial increases in bandwidth by load-balancing traffic over multiple paths in a leaf-spine design. Arista's implementation of ECMP provides a solution for preserving and utilizing all bandwidth in a routed active-active layer-3 network design while allowing deployments to scale out to hundreds of thousands of nodes.

VXLAN

VXLAN is a standardized network overlay technology that enables large and distributed layer-2 networks to be built without the scaling and connectivity constraints that might be imposed by a physical network topology. VXLAN uses a MAC-in-IP encapsulation to transport layer-2 Ethernet frames within standard routable IP packets over any layer-2 or layer-3 network. From a hypervisor's perspective, VXLAN enables VMs that need

to maintain layer-2 adjacency to other resources to be deployed on any physical server in any location regardless of the IP subnet or lack of direct layer-2 connectivity of the physical resource.

VXLAN provides solutions to a number of underlying issues with layer-2 network scaling and VM mobility:

- Enables layer-2 connectivity across physical locations or pods without increasing the fault domain
- Scales beyond 4K layer-2 segments limitation of 802.1q based VLANs
- Localizes layer-2 flooding (unknown destination) and broadcast traffic to a single site
- Enables large layer-2 networks to be built without requiring every device to see all MAC address
- Provides demand-driven automation of layer-2 connectivity to enable seamless vMotion in large domains

VXLAN Network Virtualization Gateways (Hardware VTEP)

VXLAN has become the de-facto industry-standard method of supporting layer-2 network virtualization overlays across any layer-2 or layer-3 network. There are a variety of ways VXLAN can be deployed:

- as a software feature providing encapsulation and connectivity on hypervisor-resident virtual switches
- natively on firewall and load-balancing appliances to service virtualized network traffic
- with hardware or software gateways built into switches or appliance-based platforms to provide connectivity to within the overlay network to legacy storage and computing resources that do not natively support VXLAN

Arista's approach to VXLAN is to support scalable and non-blocking hardware-accelerated VXLAN gateway functionality across a range of switches including the Arista fixed configuration 7150S, 7050X, and 7250-series switches, and the Arista 7300 and 7500E-series modular switches.

Workload Portability using Arista's Open Workload Architecture

OpenWorkload is a network architecture enabling workload portability and automation through integration with leading virtualization and orchestration systems, and simplified troubleshooting through complete physical and virtual visibility. OpenWorkload highlights include:

- **Seamless scaling.** Full support for network virtualization, connecting to major SDN controllers
- **Integrated orchestration.** Interfaces to VMware NSX™ to simplify provisioning
- **Workload visibility.** Virtual Machine level visibility enabling definition consistent network policies by workload type, persistent monitoring of vMotion'd workloads, and rapid troubleshooting of cloud networks

Designed to interface with VMware ESX and NSX as well as other cloud platforms, Arista's Open Workload architecture allows for integration with any virtualization and orchestration system.

VMware NSX Network Design Considerations

Network virtualization consists of three major aspects; decouple, reproduce, and automate. All three functions are vital in achieving the desired efficiencies. This section focuses on decoupling, which is key to simplifying and scaling the physical infrastructure.

While the NSX network virtualization solution can be successfully deployed on top of different network topologies, the focus for this document is on the Arista routed access design where the leaf/access nodes provide full L3 functionality. In this model the network virtualization solution should not span VLANs beyond a single rack inside the switching infrastructure and provide the VM mobility with overlay network topology.

Designing for Scale and Future Growth

When designing a new environment, it is essential to choose an architecture that allows for future growth. The approach presented is intended for deployments that begin small with the expectation of growth to a larger scale while retaining the same overall architecture.

This network virtualization solution does not require spanning of VLANs beyond a single rack. Elimination of this requirement has a widespread impact on the design and scalability of the physical switching infrastructure.

Although this appears to be a simple requirement, it has widespread impact on how a physical switching infrastructure can be built and on how it scales.

Note the following three types of racks within the infrastructure:

- Compute
- Edge
- Infrastructure

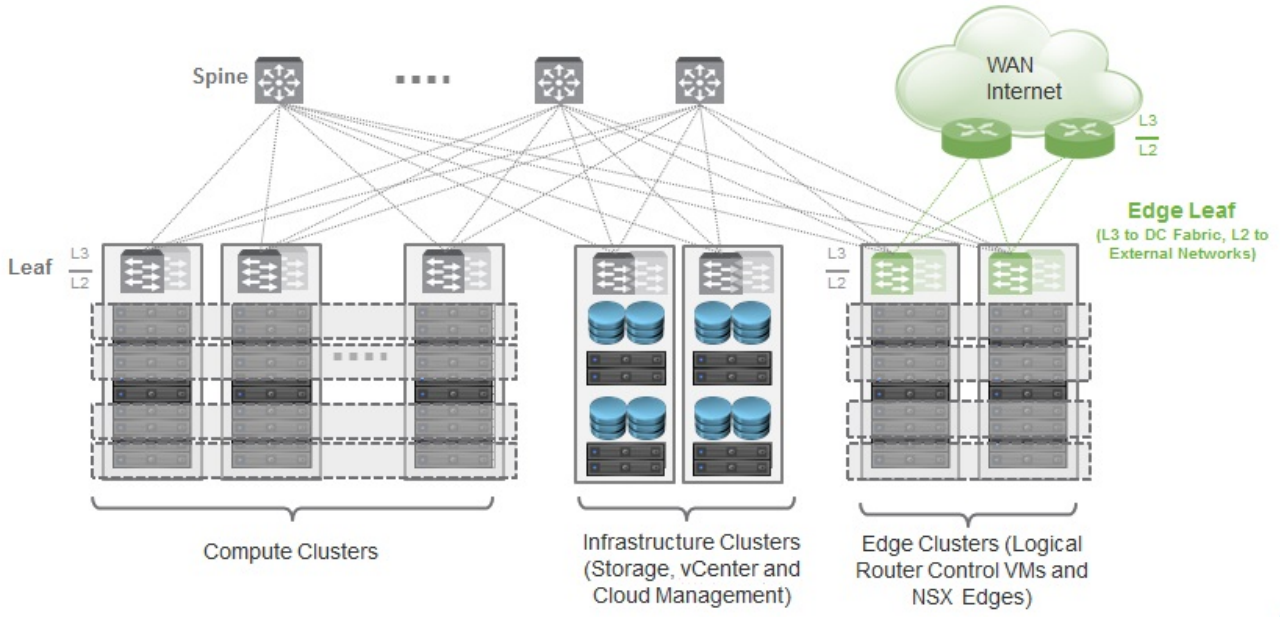

Figure 13. Data Center Design - layer-3 in Access Layer

In Figure 13, to increase the resiliency of the architecture, it is a best practice to deploy a pair of ToR switches in each rack and leverage technologies such as MLAG to dual connect them to all the servers which are part of the same rack.

Compute Racks

Compute racks are the section of the infrastructure where tenant virtual machines are hosted. Central design characteristics include:

- Interoperability with an existing network
- Repeatable rack design
- Connectivity for virtual machines without use of VLANs
- No requirement for VLANs to extend beyond a compute rack

A hypervisor typically sources three or more types of traffic. This example consists of VXLAN, management, vSphere vMotion, and storage traffic. The VXLAN traffic is a new traffic type that carries all the virtual machine communication, encapsulating it in the UDP frame. The following section will discuss how the hypervisors connect to the external network and how these different traffic types are commonly configured.

Connecting Hypervisors

The servers in the rack are connected to the access layer switch via a number of Gigabit Ethernet (1GbE) interfaces or a limited number of 10GbE interfaces. Physical server NICs are connected to the virtual switch on the other end. For best practices on how to connect the NICs to the virtual and physical switches, refer to the VMware vSphere Distributed Switch Best Practices technical white paper.

http://www.vmware.com/files/pdf/techpaper/vsphere-distributed-switch-best-practices.pdf

The connections between each server in the rack and the leaf switch are usually configured as an 802.1q trunks. A significant benefit of deploying VMware NSX network virtualization is the drastic reduction of the number of VLANs carried on those trunk connections.

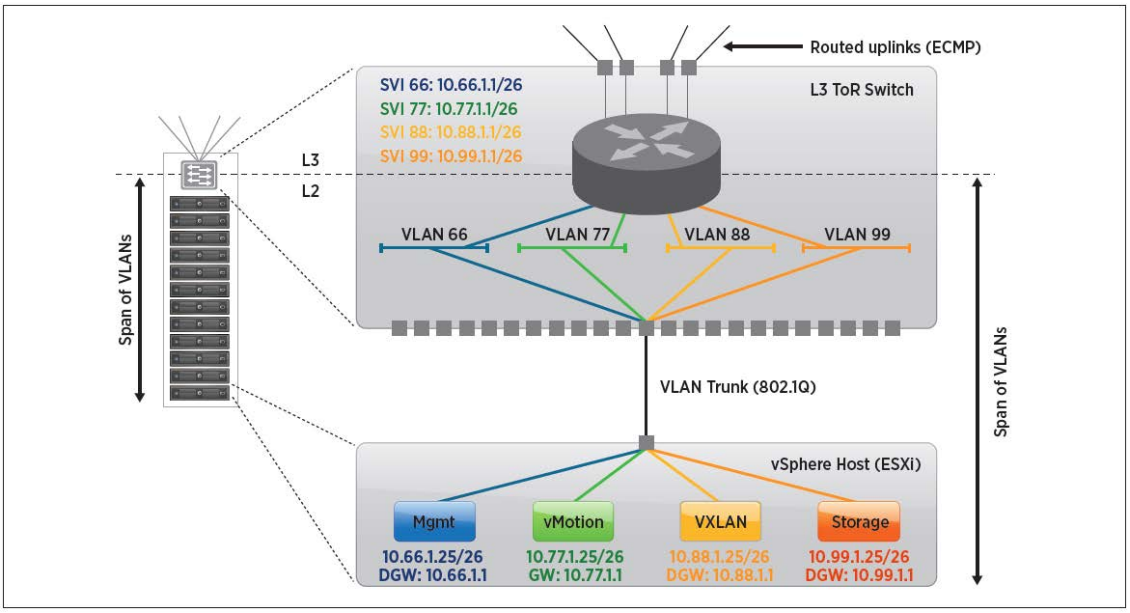

Figure 14. Example - Host and Leaf Switch Configuration in a Rack

In Figure 14, 802.1q trunks are now used for carrying few VLANs, each dedicated to a specific type of traffic (e.g., VXLAN tunnel, management, storage, VMware vSphere vMotion®). The leaf switch terminates and provides default gateway functionality for each VLAN; it has a switch virtual interface (SVI or RVI) for each VLAN. This enables logical isolation and clear separation from an IP addressing standpoint.

The hypervisor leverages multiple routed interfaces (VMkernel NICs) to source the different types of traffic. Please refer to the "VLAN Provisioning" section for additional configuration and deployment considerations of VMkernel interfaces.

VXLAN Traffic

After the vSphere hosts have been prepared for network virtualization using VXLAN, a new traffic type is enabled on the hosts. Virtual machines connected to one of the VXLAN-based logical layer-2 networks use this traffic type to communicate. The traffic from the virtual machine is encapsulated and sent out as VXLAN traffic. The external physical fabric never detects the virtual machine IP or MAC address. The virtual tunnel endpoint (VTEP) IP address is used to transport the frame across the fabric. In the case of VXLAN, the tunnels are initiated and terminated by a VTEP. Traffic that flows between virtual machines in the same data center is typically referred to as east–west traffic. For this type of traffic, both the source and destination VTEP are situated in hypervisors located in compute racks. Traffic leaving the data center will flow between a tenant virtual machine and an NSX Edge, and is referred to as north–south traffic.

VXLAN configuration requires an NSX VDS vSwitch. One requirement of a single-VDS–based design is that the same VLAN is defined for each hypervisor to source VXLAN encapsulated traffic (VLAN 88 in the example in Figure 14). Because a VDS can span hundreds of hypervisors, it can reach beyond a single leaf switch. This mandates that the host VTEPs—even if they are on the same VDS and source traffic on the same VLAN—must be able to reside in different subnets.

Management Traffic

Management traffic can be categorized into two types; one is sourced and terminated by the management VMkernel interface on the host, the other is involved with the communication between the various NSX components. The traffic that is carried over the management VMkernel interface of a host includes the communication between vCenter Server and hosts as well as communication with other management tools such as NSX Manager. The communication between the NSX components involves the heartbeat between active and standby edge appliances.

Management traffic stays inside the data center. A single VDS can span multiple hypervisors that are deployed beyond a single leaf switch. Because no VLANs can be extended beyond a leaf switch, the management interfaces of hypervisors participating in a common VDS and connected to separate leaf switches are in separate subnets.

vSphere vMotion Traffic

During the vSphere vMotion migration process, the running state of a virtual machine is transferred over the network to another host. The vSphere vMotion VMkernel interface on each host is used to move this virtual machine state. Each vSphere vMotion VMkernel interface on the host is assigned an IP address. The number of simultaneous vMotion migrations than can be performed is limited by the speed of the physical NIC. On a 10GbE NIC, eight simultaneous vSphere vMotion migrations are allowed.

Note: VMware has previously recommended deploying all the VMkernel interfaces used for vMotion as part of a common IP subnet. This is not possible when designing a network for network virtualization using layer-3 at the access layer, where it is mandatory to select different subnets in different racks for those VMkernel interfaces. Until VMware officially relaxes this restriction, it is recommended that customers requiring vMotion over NSX go through VMware's RPQ ("Request for Product Qualification") process so that the customer's design can be validated on a case-by-case basis.

Storage Traffic

A VMkernel interface is used to provide features such as shared or non-directly attached storage. Typically this is storage that can be attached via an IP connection (e.g., NAS, iSCSI) rather than FC or FCoE. The same rules that apply to management traffic apply to storage VMkernel interfaces for IP address assignment. The storage VMkernel interface of servers inside a rack (i.e., connected to a leaf switch) is part of the same subnet. This subnet cannot span beyond this leaf switch, therefore the storage VMkernel interface IP of a host in a different rack is in a different subnet. For an example of the IP address for these VMkernel interfaces, refer to the "VLAN Provisioning" section.

Edge Racks

Tighter interaction with the physical infrastructure occurs while bridging between the overlay world and the physical infrastructure. The main functions provided by an edge rack include:

- Providing on-ramp and off-ramp connectivity to physical networks
- Connecting with VLANs in the physical world
- Hosting centralized physical services

Tenant-specific addressing is exposed to the physical infrastructure where traffic is not encapsulated in VXLAN (e.g., NAT not used at the edge). In the case of an layer-3 edge, the IP addresses within the overlay are exposed to the physical fabric. The guiding principle in these cases is to separate VXLAN (overlay) traffic from the un-encapsulated (native) traffic. As shown in Figure 14, VXLAN traffic hits the data center internal Ethernet switching infrastructure. Native traffic traverses a dedicated switching and routing infrastructure facing the WAN or Internet and is completely decoupled from the data center internal network.

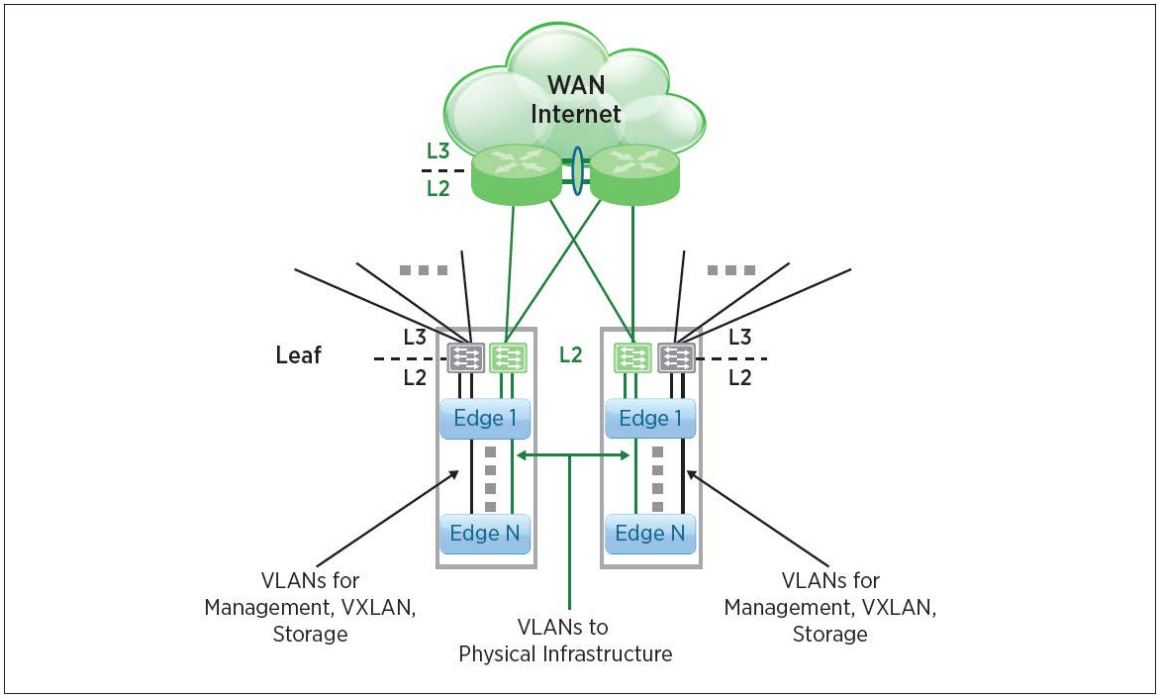

Figure 15. VXLAN Traffic and the Data Center Internal Ethernet Switching Infrastructure

To maintain the separation, NSX Edge virtual machines can be placed in NSX Edge racks, assuming the NSX Edge has at least one native interface. For routing and high availability, the two interface types—overlay and

native—must be examined individually. The failover mechanism is based on the active standby model, where the standby Edge takes over after detecting the failure of the active Edge.

Layer-3 NSX Edge Deployment Considerations

When deployed to provide layer-3 routing services, the NSX Edge terminates all logical networks and presents an layer-3 hop between the physical and the logical world. Depending on the use case, either NAT or static/dynamic routing may be used to provide connectivity to the external network.

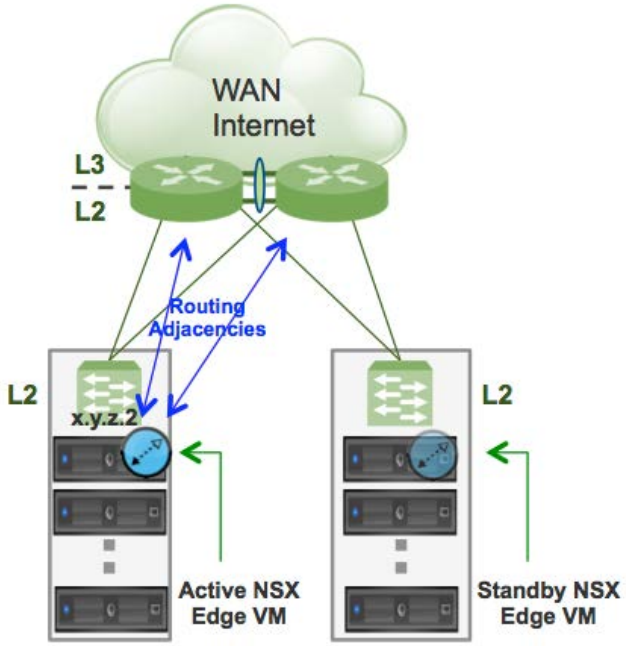

Figure 16. High Availability – Active Standby Edge Topology

In order to provide redundancy to the NSX Edge, each tenant should deploy an active/standby pair of NSX Edge devices. As highlighted in Figure 16, it is recommended to deploy those active/standby units in separate Edge racks, so that a failure scenario (e.g., power outage) affecting the first rack would cause the standby NSX Edge to take over and assume the outside IP address of the previously active Edge (x.y.z.2 in this example). To notify the upstream infrastructure, here layer-2 switches that potentially interconnect the Edge and the first physical router, a GARP message is sent out by the hypervisor where the NSX Edge is newly activated.

For this mechanism to work, two VLANs must be extended between the NSX Edge racks. The first VLAN is dedicated to the exchange of keepalive messages between the active and standby NSX Edge devices. The second VLAN is required to allow traffic to be sent from the active NSX Edge into the physical infrastructure. Tunnel interfaces connecting the VXLAN endpoints do not have to extend VLANs. Before the failover, traffic originating from the VTEPs of hypervisors in the compute racks is directed toward the VTEP of the hypervisor hosting the active NSX Edge. After failover, that traffic is sent to the VTEP of the hypervisor that hosts the newly activated NSX Edge. The VLANs used to transport VXLAN traffic to/from the separate edge racks can be chosen independently and do not need to be extended across these racks

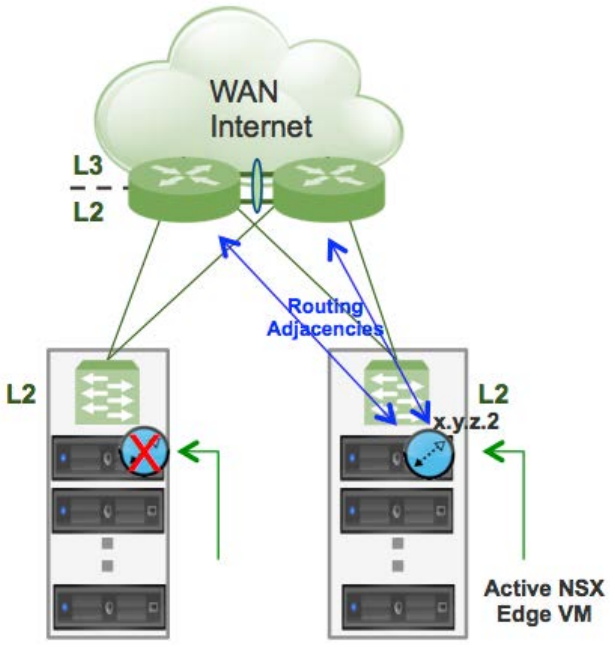

Figure 17. Failure of Active Edge

Infrastructure Racks

Infrastructure racks host the management components, including vCenter Server, NSX Manager, NSX Controller, CMP, and other shared IP storage–related components. It is key that this portion of the infrastructure does not have any tenant-specific addressing. If bandwidth-intense infrastructure services are placed in these racks—IP-based storage, for example—bandwidth of these racks can be dynamically scaled, as discussed in the "High Bandwidth" subsection of the "Data Center Fabric Attributes" section.

VLAN Provisioning

Every compute rack has four different subnets, each supporting a different traffic type; tenant (VXLAN), management, vSphere vMotion, and storage traffic. Provisioning of IP addresses to the VMkernel NICs of each traffic type is automated using vSphere host profiles. The host profile feature enables creation of a reference host with properties that are shared across the deployment. After this host has been identified and required sample configuration performed, a host profile can be created and applied across in the deployment. This allows quick configuration of a large numbers of hosts.

As shown in Figure 18, the same set of four VLANs—storage, vSphere vMotion, VXLAN, management—is provided in each rack. The following are among the configurations required per host:

- vmknic IP configuration per traffic type in the respective subnet or VLAN
- Static route configuration per subnet, to handle proper traffic routing to the respective gateways

Static routes are required because only two TCP/IP stacks are currently supported on the VMware ESXi™ hosts. This will limit the number of default gateway configurations to two; one to be used for the management traffic, the second for the VXLAN traffic.

For example, in rack 1, host 1 has the following vmknic configuration:

- Storage vmknic with IP address 10.66.1.10

- vSphere vMotion vmknic with IP address 10.77.1.10
- VXLAN vmknic with IP address 10.88.1.10
- Management vmknic with IP address 10.99.1.10

The two default gateway configurations on host 1 will be placed in the management vmknic subnet 10.99.1.0/26 and in the VXLAN vmknic subnet 10.88.1.0/26. To support proper routing for other subnets, the following static routes must be added as part of the host 1 preparation:

- Storage network route – esxcli network ip route ipv4 add -n 10.66.0.0/26 -g 10.66.1.1
- vSphere vMotion network route – esxcli network ip route ipv4 add -n 10.77.0.0/26 -g 10.77.1.1

After host 1 of rack 1 has been configured, a host profile is created and applied to other hosts in the rack. While applying the profile to the hosts, new vmknics are created and the static routes are added, simplifying the deployment.

In a vSphere Auto Deploy environment the PXE boot infrastructure, the Auto Deploy server, and vCenter Server support the host-booting process and help automate the deployment and upgrades of the ESXi hosts.

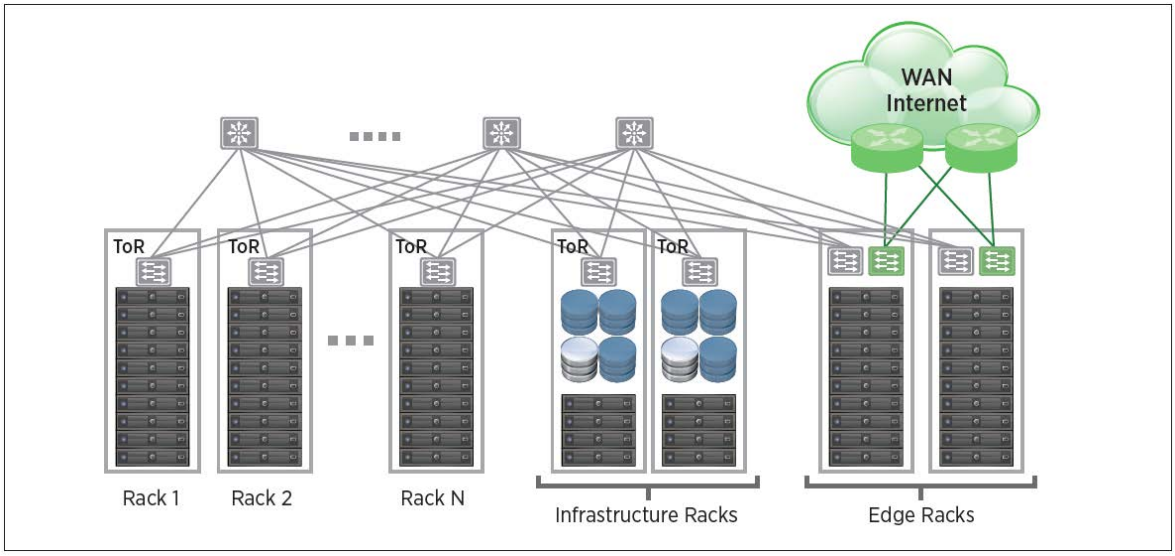

Figure 18. Host Infrastructure Traffic Types and IP Address Assignment

IP ADDRESS MANAGEMENT AND VLANs[1]		
Function	**Global VLAN ID**	**IP Address**
Storage	66	10.66.R_id.x/26
vMotion	77	10.77.R_id.x/26
VXLAN/VTEP	88	10.88.R_id.x/26
Management	99	10.99.R_id.x/26

[1] Values of VLANs, IP addresses, and masks are an example (not prescriptive to the design)

Table 1. IP Address Management and VLANs

Multi-tier Edges and Multi-tier Application Design Considerations

Classical multi-tier compute architectures have functions that are logically separated, where each function has different requirements for resource access, data segregation, and security. A classical three-tier compute architecture typically comprises a presentation tier, an application or data access tier, and a database tier. Communication between the application tier and the database tier should be allowed, while an external user has access to only the presentation tier, which is typically a web-based service.

The recommended solution to comply with data access policies is to deploy a two-tier edge design. The inner edge enables VXLAN-to-VXLAN east–west traffic among the presentation, database, and application tiers, represented by different logical networks. The outer edge connects the presentation tier with the outer world for on-ramp and off-ramp traffic. Communication within a specific logical network enables virtual machines to span across multiple racks to achieve optimal utilization of the compute rack infrastructure.

Note: At the current time, a logical network can span only a single vCenter domain. Figure 19 shows the placement of the logical elements of this architecture.

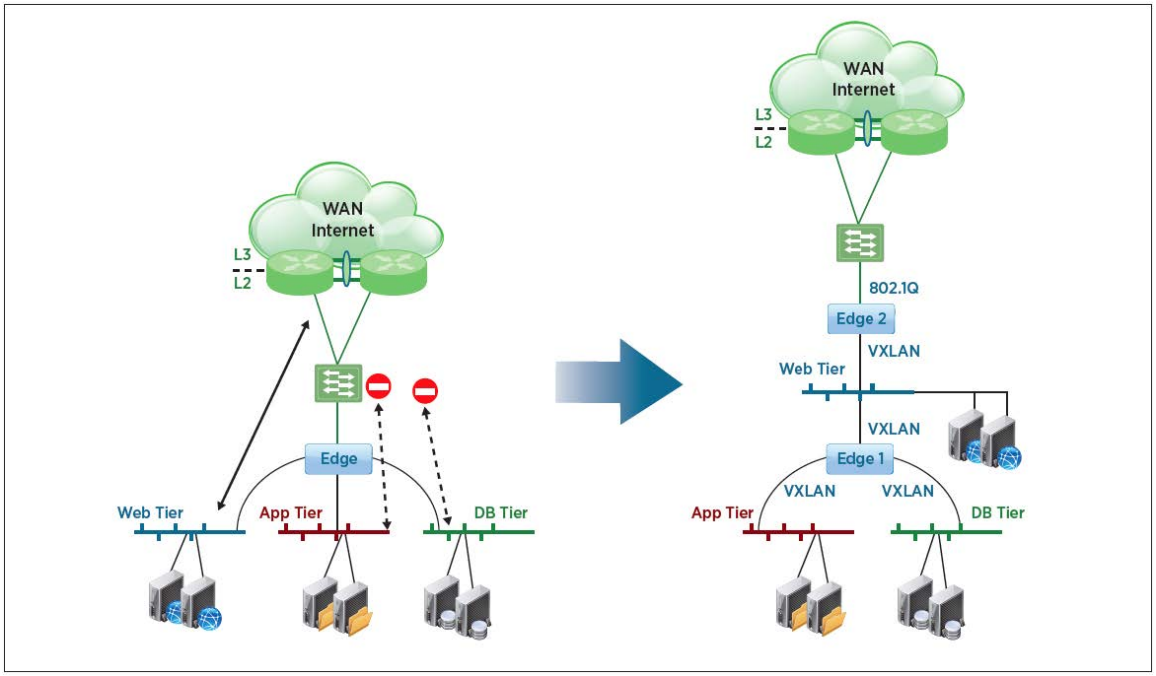

Figure 19. Two Options for Logical Element Placement in a Multitier Application

It is preferable that the outer edges be physically placed in the edge racks. Inner edges can be centralized in the Edge racks or distributed across the compute racks, where web and application compute resources are located.

Logical Switching

The logical switching capability in the NSX platform provides customers the ability to spin up isolated logical layer-2 networks with the same flexibility and agility as spinning up virtual machines. This section describes the various components that enable logical switching and the communication among those components.

Components

As shown in Figure 20, there are three main components that help decouple the underlying physical network fabric and provide network abstraction. This decoupling is achieved by encapsulating the virtual machine traffic using the VXLAN protocol.

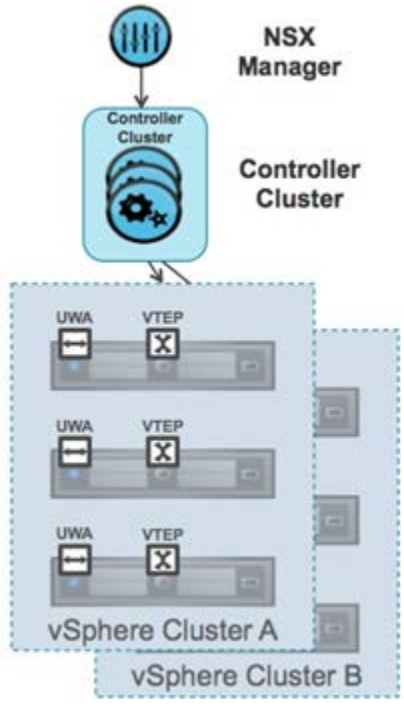

Figure 20. Logical Switching Components

NSX Manager

The NSX Manager is the management plane component responsible for configuring logical switches and connecting virtual machines. It also provides API interface, which automates deployment and management of these switches through a cloud management platform.

Controller Cluster

The Controller Cluster in the NSX platform is the control plane component responsible for managing the hypervisors' switching and routing modules. The Controller Cluster consists of controller nodes that manage specific logical switches. The controller manages the VXLAN configuration mode. Three modes are supported which are: unicast, multicast and hybrid. The recommendation and details of these modes are discussed in the section "Logical Switch Replication Mode". It is important to note that the data path (VM user traffic) does not go through controller even though controller is responsible for managing the VTEP configuration. In "Unicast mode" there is no need for multicast support from the physical network infrastructure. Consequently in "Unicast mode" there are no requirements to provision multicast group IP addresses or enable PIM routing or IGMP snooping features on physical switches or routers.

User World Agent (UWA) and VXLAN Tunnel Endpoint (VTEP)

There are two components on the hypervisor used to establish communication paths with the controller cluster and other hypervisors. The User World Agent establishes communication with the Controller Cluster while the VTEP component creates tunnels between hypervisors.

Transport Zone

As part of the host preparation process, the hypervisor modules are deployed and configured through the NSX Manager. After logical switching components are installed and configured, the next step is to define the span of logical switches by creating a transport zone. The transport zone consists of a set of clusters. For example, if there are 10 clusters in the data center, a transport zone can include some or all of those 10 clusters. In this scenario a logical switch can span the whole data center. Figure 21 shows a deployment after the NSX components are installed to provide logical switching. The Edge Services Router in the edge rack provides the logical switches access to the WAN and other network services.

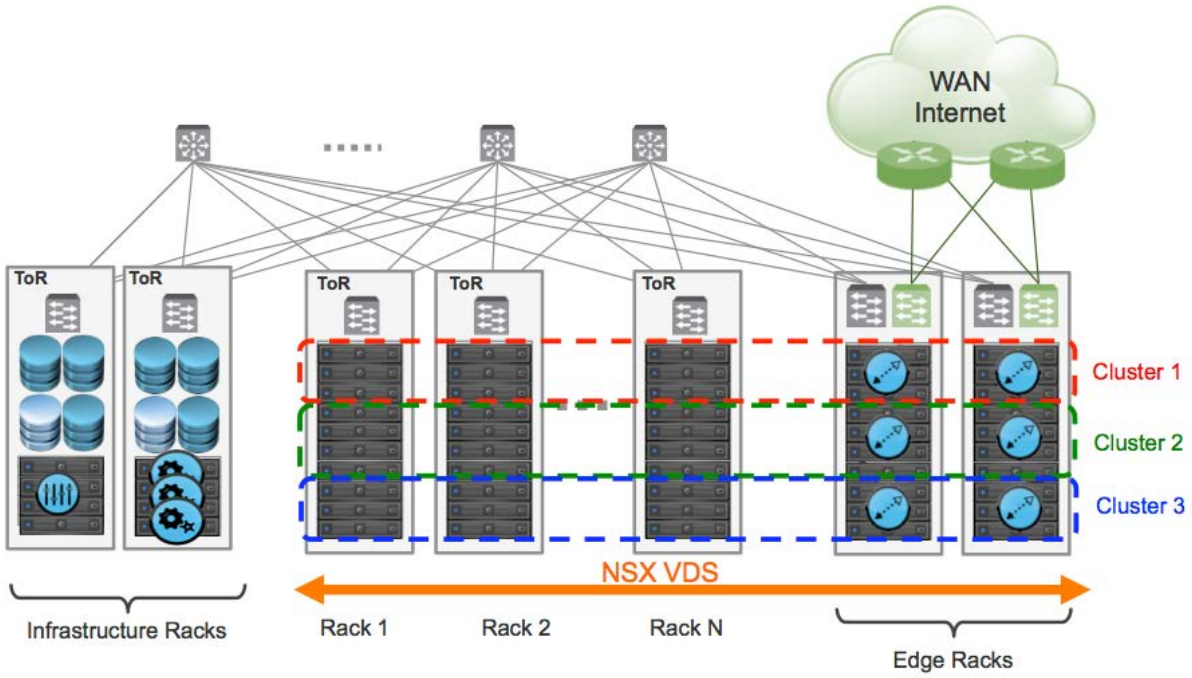

Figure 21. Logical Switching Components in the racks

Logical Switch Replication Modes

When two VMs connected to different ESXi hosts need to communicate directly, unicast VXLAN encapsulated traffic is exchanged between the VTEP IP addresses associated to the two hypervisors. Traffic originated by a VM may need to be sent to all the other VMs belonging to the same logical switch, specifically for three types of layer-2 traffic:

- Broadcast
- Unknown Unicast
- Multicast

Note: These types of multi-destination traffic types may be referred to using the acronym BUM (Broadcast, Unknown unicast, Multicast). In an NSX deployment with vSphere, there should never be a need to flood unknown unicast traffic on a given logical network since the NSX controller is made aware of the MAC addresses of any actively connected VM.

For these three scenarios, traffic originated by a given ESXi host must be replicated to multiple remote hosts (hosting other VMs part of the same logical network). NSX supports three different replications modes to enable multi-destination communication on VXLAN backed logical switches – unicast, hybrid and multicast. By default a logical switch inherits its replication mode from the transport zone, however this can be overridden.

Unicast Mode

For unicast mode replication, the ESXi hosts part of the NSX domain is divided in separate groups (segments) based on IP subnet addresses of VTEP interfaces. The NSX controller selects a specific ESXi host in each segment to serve as the Unicast Tunnel End Point (UTEP). The UTEP is responsible for replicating multi-destination traffic to all the ESXi hosts in its segment (i.e., whose VTEPs belong to the same subnet of the UTEP's VTEP interface) and to all the UTEPs belonging to different segments.

In order to optimize the replication behavior, every UTEP will replicate traffic only to ESXi hosts on the local segment that have at least one VM actively connected to the logical network where multi-destination traffic is destined. In addition, traffic will only be replicate to the remote UTEPs if there is at least one active VM connected to an ESXi host part of that remote segment. The NSX controller is responsible for providing to each ESXi host the updated list of VTEPs address for replication of multi-destination traffic.

Unicast mode replication requires no explicit configuration on the physical network to enable distribution of multi-destination VXLAN traffic. This mode is well suited for smaller deployments with fewer VTEPs per segment and few physical segments. It may not be suitable for extremely large scaled environments as the overhead of replication increases with the number of segments. Figure 22 illustrates a unicast mode logical switch. In this example there are 4 VMs on logical switch 5001. When VM1 sends a frame to all VMs on the logical switch, the source VTEP replicates the packet only to the other VTEPs belonging to the local segment and to the UTEPs part of remote segments (only one remote segment is shown in this example).

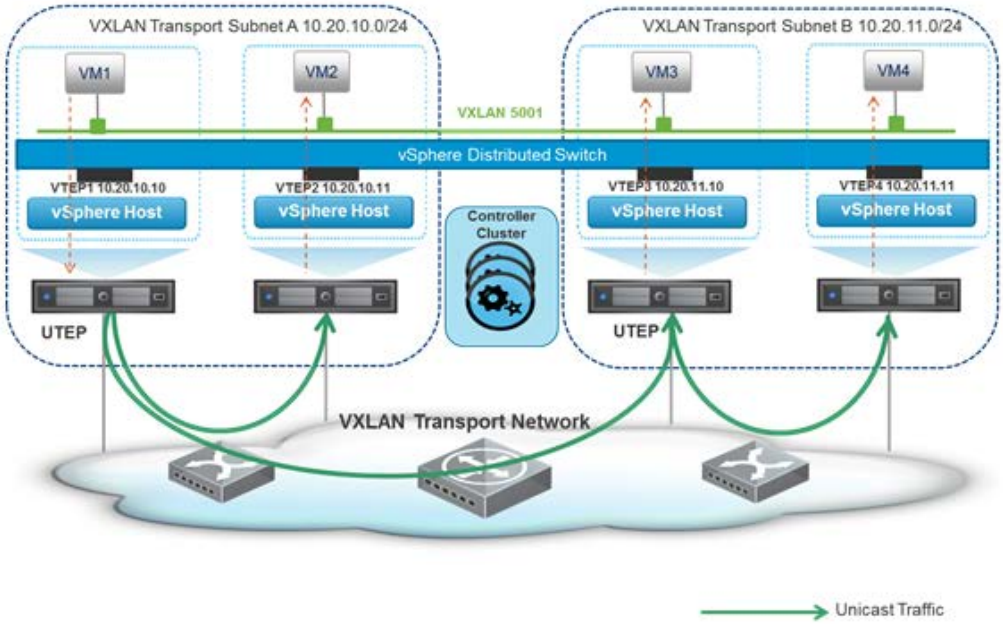

Figure 22. Unicast Mode Logical Switch

Multicast Mode

When Multicast mode is chosen for the logical switch, NSX relies on the layer-2 and layer-3 multicast capabilities of the physical network to ensure VXLAN traffic is sent to all the VTEPs. In this mode layer-2 multicast is used to replicate traffic to all VTEPs in the local segment (i.e., VTEP IP addresses are part of the same IP subnet). IGMP snooping must be configured on the physical switches. It is recommended to have an IGMP querier per VLAN. To ensure multicast traffic is delivered to VTEPs in a different subnet from the source VTEP, multicast routing and PIM must be enabled. Figure 23 shows a multicast mode logical switch. IGMP snooping enables the physical switch to replicate multicast traffic to all VTEPs in the segment and PIM allows multicast traffic to be delivered to VTEPs in remote segments. Using multicast mode eliminates additional overhead on the hypervisor as the environment scales.

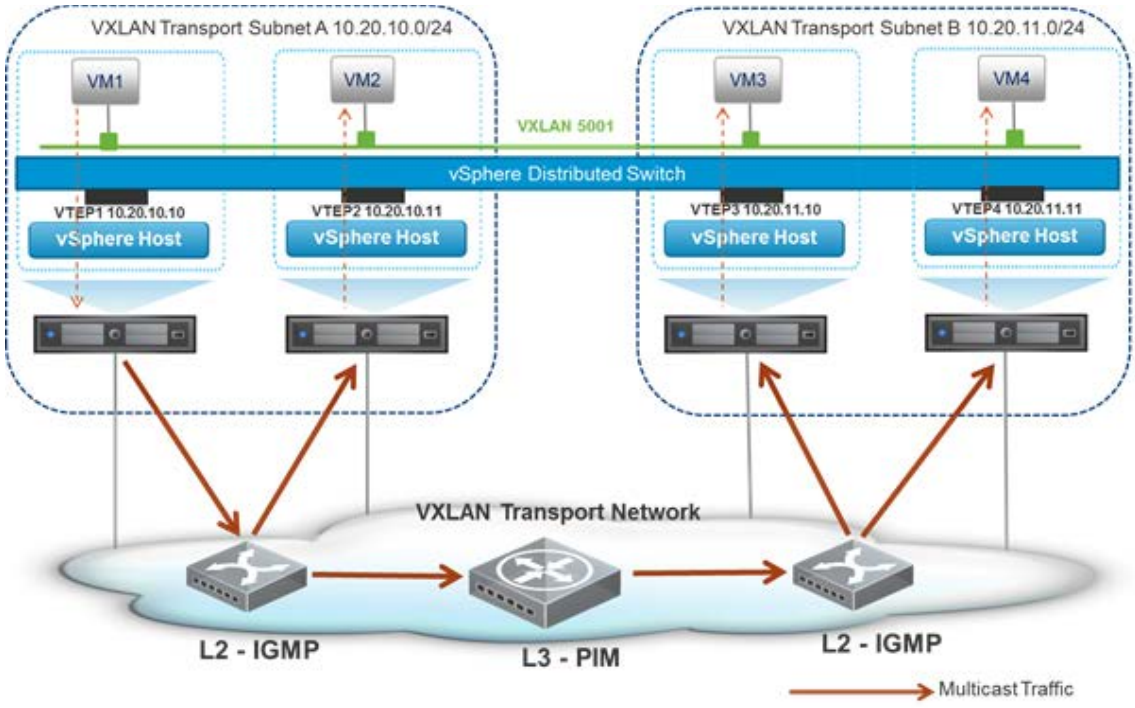

Figure 23. Multicast Mode Logical Switch

Hybrid Mode

Hybrid Mode offers the simplicity of unicast mode (i.e., no IP multicast routing configuration in physical network) while leveraging the layer-2 multicast capabilities of physical switches. In hybrid mode, the controller selects one VTEP per physical segment to function as a Multicast Tunnel End Point (MTEP). When a frame is sent over VXLAN to VTEPs in multiple segments, the MTEP creates one copy per MTEP and forwards the encapsulated frame to the IP address of the remote MTEPs. The source MTEP also encapsulates one copy of the original frame with an external destination IP address of the multicast address associated with the logical switch. This is then sent to the upstream physical switch. Layer-2 multicast configuration in the physical network is used to ensure that the VXLAN frame is delivered to all VTEPs in the local segment. This is illustrated in Figure 24 where the MTEP in segment 10.20.10.0/24 sends one copy to the MTEP in segment 10.20.11.0/24.

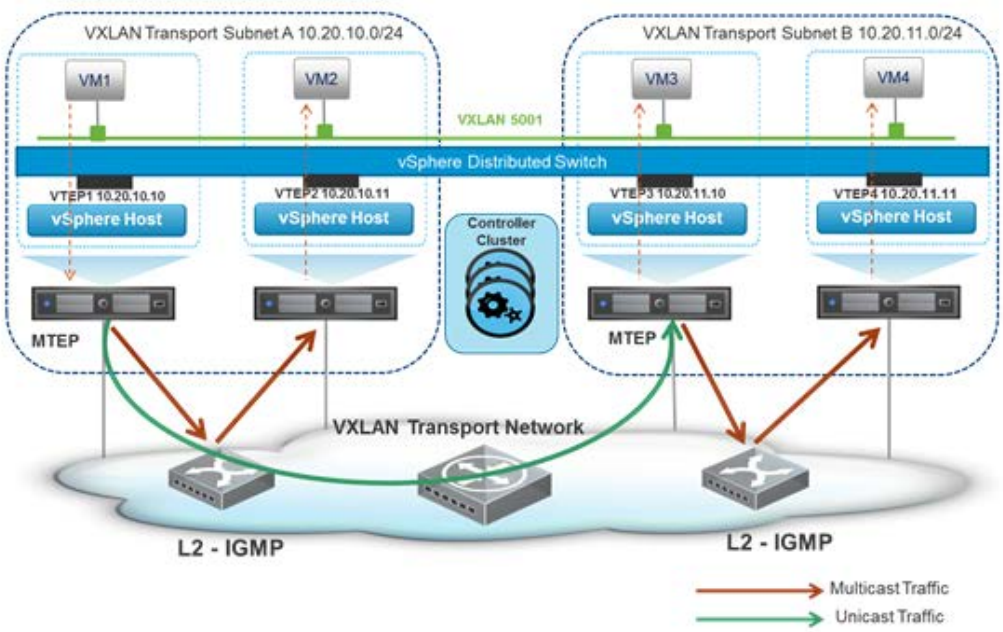

Figure 24. Hybrid Mode Logical Switch

Logical Switch Addressing

IP address management is a critical task in a large cloud environment with multiple tenants or big enterprises with multiple organizations and applications. This section focuses on IP address management of the virtual machines deployed on the logical switches. Each logical switch is a separate layer-2 broadcast domain that can be associated with a separate subnet using private or public IP space. Depending on whether private or public space is used for the assignment to the logical networks, users must choose either the NAT or non-NAT option on the NSX Edge Services Router. The IP address assignment depends on whether the virtual machine is connected to a logical switch through a NAT or a non-NAT configuration.

With Network Address Translation

In the deployments where organizations have limited IP address space, NAT is used to provide address translation from private IP space to the limited public IP addresses. An Edge Services Router can provide individual tenants with the ability to create distinct pools of private IP addresses, which may be mapped to the publicly routable external IP address of the external NSX Edge Services Router interface.

Figure 25 shows a three-tier app deployment with each tier virtual machine connected to separate logical switch. The web, app and DB logical switches are connected to the three internal interfaces of the NSX Edge Services Router; its external interface is connected to the Internet via an external data center router.

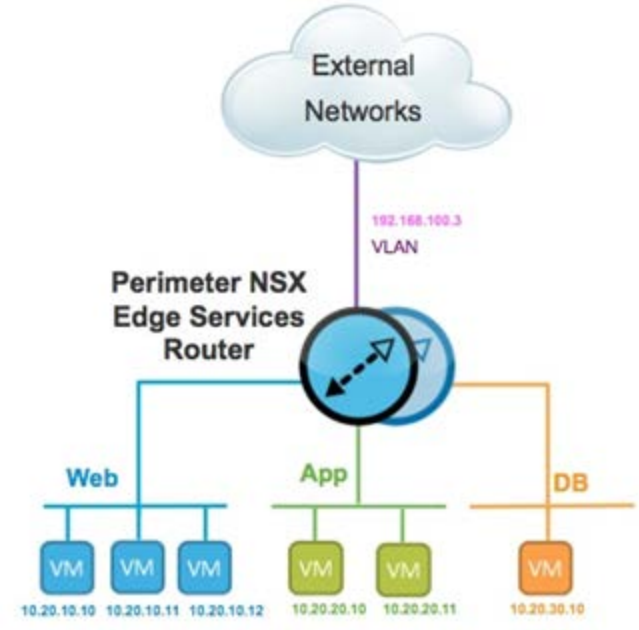

Figure 25. NAT and DHCP Configuration on NSX Edge Service Router

Configuration details of the NSX Edge Services Router include:

- Web, app and DB logical switches are connected to the Internal interfaces of the NSX Edge Services Router
- The NSX Edge Services Router uplink interface is connected to the VLAN port group in subnet 192.168.100.0/24
- Enable DHCP service on this internal interface of by providing a pool of IP addresses (e.g., 10.20.10.10 to 10.20.10.50)
- The NAT configuration on the external interface enables VMs on a logical switch to communicate with devices on the external network. This communication is allowed only when the requests are initiated by the VMs connected to the internal interface of the NSX Edge Services Router

In situations where overlapping IP and MAC address support is required, one NSX Edge Services Router per tenant is recommended. Figure 26 shows an overlapping IP address deployment with two tenants and two separate NSX Edge Services Routers.

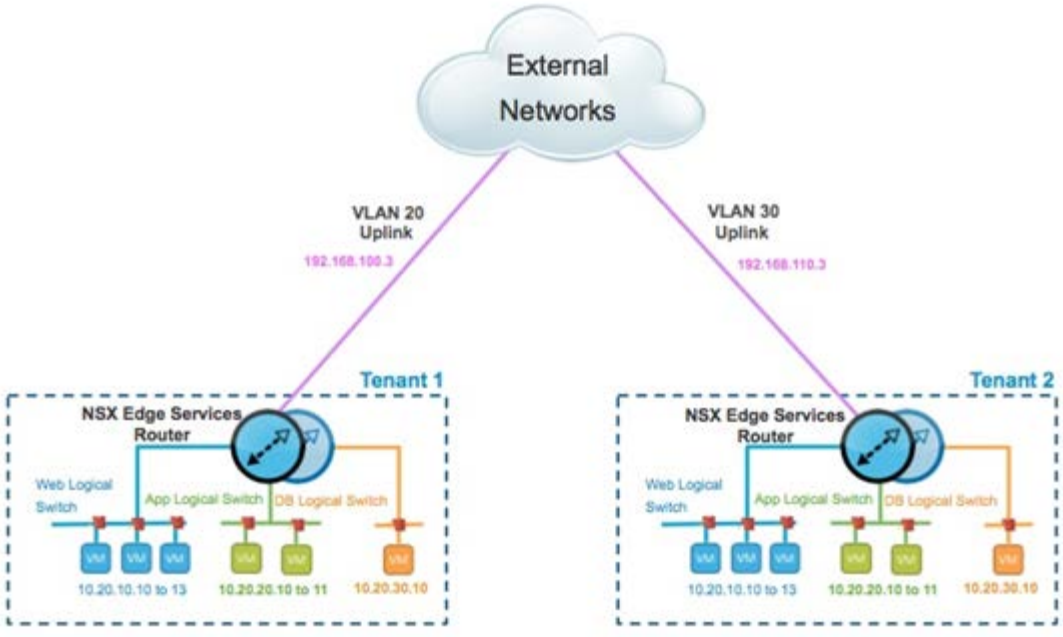

Figure 26. Overlapping IP and MAC Addresses

Without Network Address Translation

The static and dynamic routing features of the NSX platform are appropriate for organizations that are not limited by routable IP addresses, have VMs with public IP addresses, or do not want to deploy NAT.

Logical Routing

The NSX platform supports two different modes of logical routing, known as distributed routing and centralized routing. Distributed routing provides better throughput and performance for the east-west traffic while centralized routing handles north-south traffic. This section will provide more details on the two modes as well as describe common routing topologies. For the additional network services required for the applications in the datacenter please refer to the logical firewall and logical load balancer sections.

Distributed Routing

The distributed routing capability in the NSX platform provides an optimized and scalable way of handling east-west traffic within a data center. Communication between virtual machines or resources within the datacenter is referred to as east-west traffic. The amount of east-west traffic in the data center is growing. The new collaborative, distributed, and service oriented application architecture demands higher bandwidth for server-to-server communication.

If these servers are VMs running on a hypervisor and are connected to different subnets, the communication must go through a router. If a physical router is used to provide routing services, the VM communication must get to the physical router and return to the server after routing decision. This suboptimal traffic flow is referred to as “hairpinning”.

The distributed routing on the NSX platform prevents the hairpinning by providing hypervisor level routing functionality. Each hypervisor has a routing kernel module that performs routing between the logical interfaces (LIFs) defined on that distributed router instance. The components section below describes both the various modules in distributed routing and the communication between them.

Centralized Routing

The NSX Edge Services Router provides the traditional centralized routing support in the NSX platform. Along with the routing services the NSX Edge Services Router also supports other network services including DHCP, NAT, and load balancing.

Routing Components

Figure 26 and 27 show the multiple components for logical routing. Some of the components are related to distributed routing and some others to centralized routing.

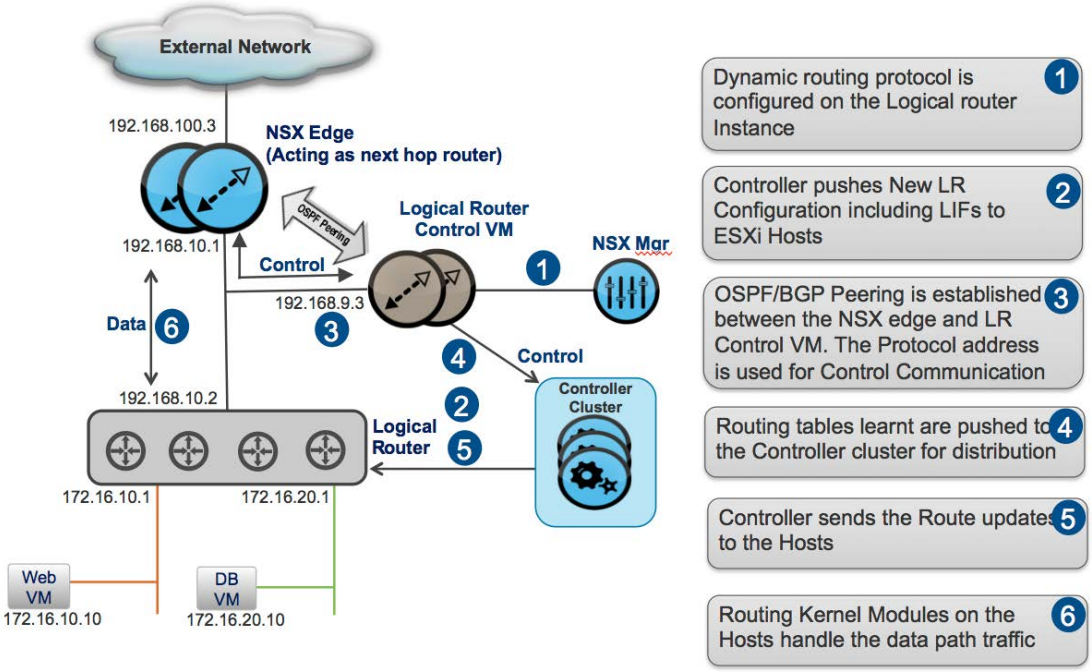

Figure 27. Logical Routing Components

NSX Manager

The NSX Manager helps configure and manage logical routing services. Deployment is possible as either a distributed or centralized logical router. If distributed router is selected, the NSX Manager deploys the logical router control VM and pushes the logical interface configurations to each host through the controller cluster. In the case of centralized routing, NSX Manager simply deploys the NSX Edge Services Router VM. The API interface of the NSX Manager helps automate deployment and management of these logical routers through a cloud management platform.

Logical Router Control VM

The logical router control VM is the control plane component of the routing process. It supports the dynamic routing protocols OSPF and BGP.

The logical router control VM communicates with the next hop router using the dynamic routing protocol and pushes the learned routes to the hypervisors through the controller cluster. High Availability (HA) may be configured while deploying the control VM. Two VMs are deployed in active-standby mode when HA mode is selected.

Logical Router Kernel Module

The logical router kernel module is configured as part of the preparation process through the NSX manager. The kernel modules are similar to the line cards in a modular chassis supporting layer-3 routing. The kernel modules have routing information base (RIB) that is pushed through the controller cluster. Data plane functionality of route and ARP entry lookup is performed by the kernel modules.

Controller Cluster

The Controller cluster is responsible for distributing routes learned from the control VM across the hypervisors. Each controller node in the cluster takes responsibility of distributing the information for a particular logical router instance. In a deployment where there are multiple logical router instance deployed, the load is distributed across the controller nodes.

NSX Edge Services Router

The NSX Edge Services Router is the centralized services router that provides support DHCP, NAT, firewall, load balancing, and VPN capabilities along with routing protocols OSPF, IS-IS, and BGP.

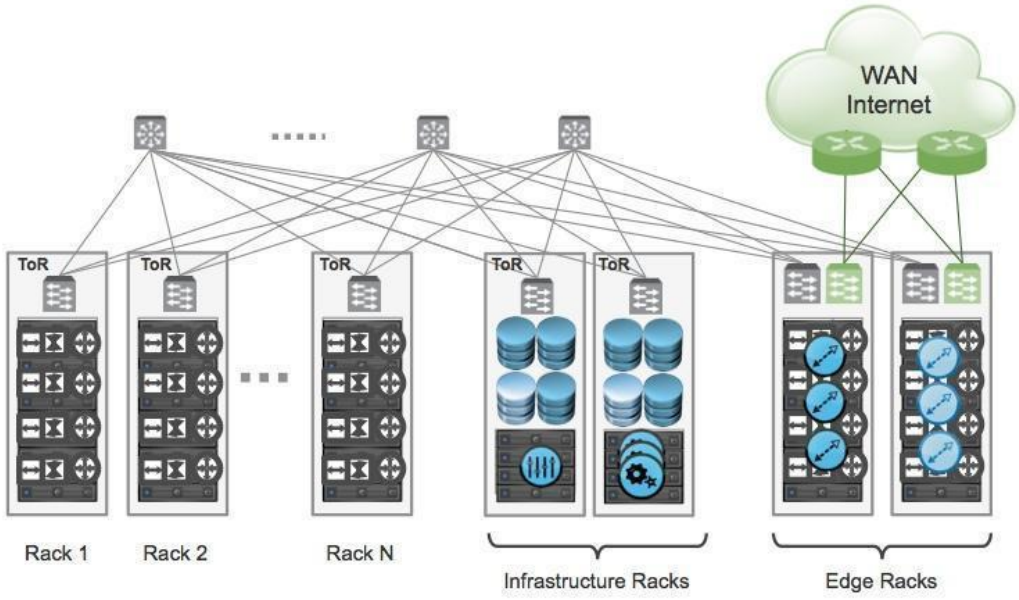

Figure 28. Logical Routing Components in the Racks

Logical Switching and Routing Deployments

Various topologies can be built using logical switching and logical routing features of the NSX platform. Examples for two routing topologies that utilizes both distributed and centralized logical routing capabilities are provided:

- Physical Router as Next Hop
- Edge Services Router as Next Hop

Physical Router as Next Hop

As shown in Figure 29, an organization is hosting multiple applications and wants to provide connectivity among the different tiers of the application as well as to the external network. Separate logical switches provide layer-2 network connectivity for the VMs in the particular tier. The distributed logical routing configuration allows the VMs on two different tiers to communicate with each other. Dynamic routing protocol support on the logical router enables the exchange of routes with the physical next hop router. This allows external users to access the applications connected to the logical switches in the data center.

In this topology the east-west and north-south routing decision happens at the hypervisor level in a distributed fashion.

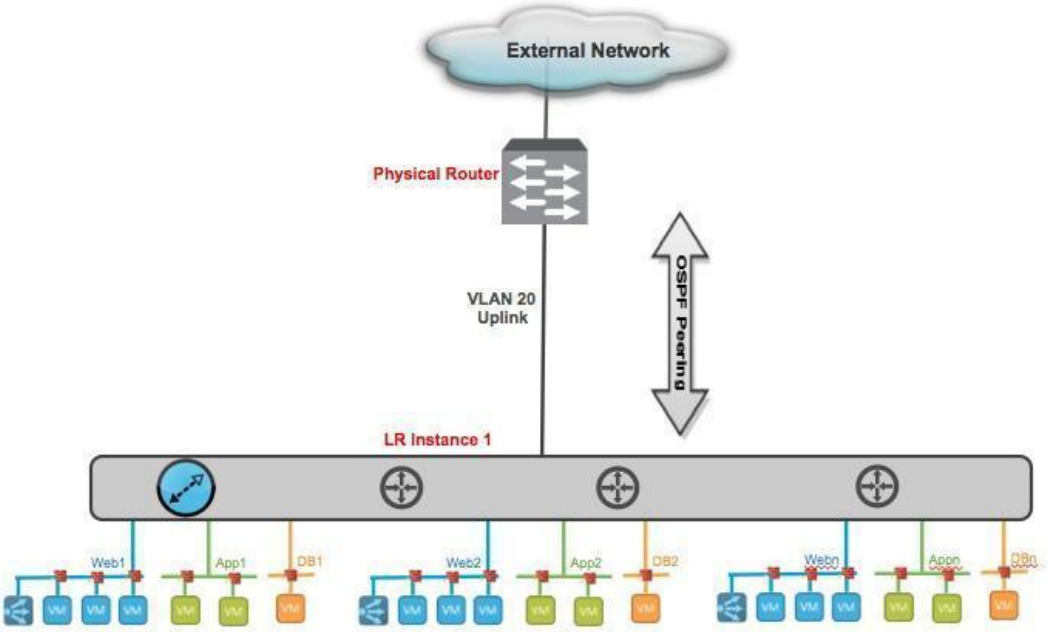

Figure 29. Physical Router as Next Hop

Edge Services Router as Next Hop

A service provider environment may have multiple tenants with each requiring a different number of isolated logical networks or network services such as load balancers, firewalls, and VPNs. In these environments, the NSX Edge Services Router provides network services capabilities along with dynamic routing protocol support.

Figure 29 shows two tenants connected to the external network through the NSX Edge Services Router. Each tenant has its logical router instance that provides routing within the tenant. The dynamic routing protocol configuration between the tenant logical router and the NSX Edge Services Router provides the connectivity from the tenant VMs to the external network.

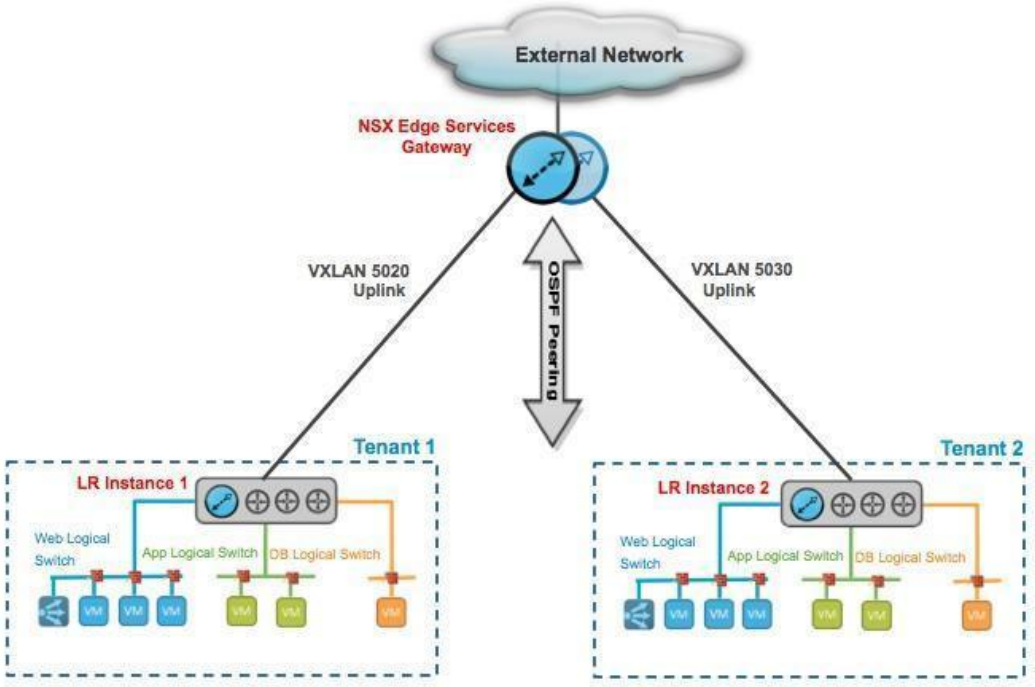

Figure 30. NSX Edge Services Router Providing Next Hop and Network Services

In this example the NSX Edge Services Gateway establishes a single routing adjacency with the routers on the physical infrastructure, independent from the number of the deployed logical router instances. The east-west traffic routing is handled by the distributed router in the hypervisor and the north-south traffic flows through the NSX Edge Services Router.

Scalable Topology

The service provider topology can be scaled out as shown in Figure 31. The diagram shows nine tenants served by NSX Edge on the left and the other nine by the Edge on the right. The service provider can easily provision another NSX Edge to serve additional tenants.

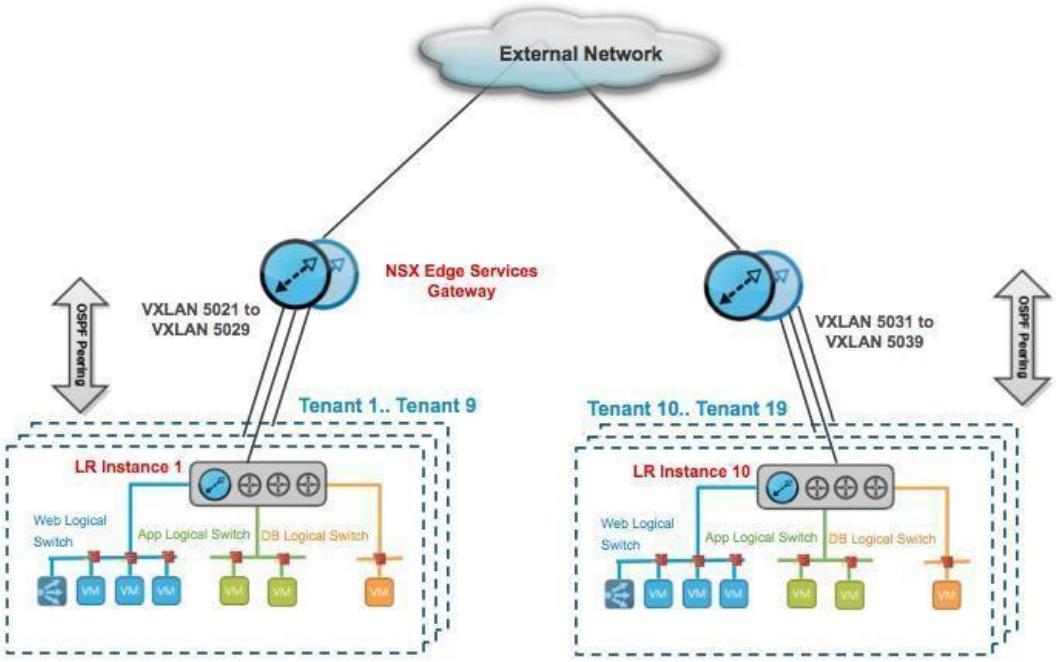

Figure 31. Scalable Topology

Logical Firewalling

The VMware NSX platform includes distributed kernel enabled firewalling with line rate performance, virtualization, and identity aware with activity monitoring. Other network security features native to network virtualization are also available.

Network Isolation

Isolation is the foundation of most network security, providing solutions for compliance, containment, or interaction of distinct environments. Access lists (ACLs) and firewall rules on physical devices have traditionally been used to enforce isolation policies.

Virtual networks are isolated from other virtual network as well as from the underlying physical network by default, delivering the security principle of least privilege. Virtual networks are created in isolation and remain isolated unless specifically connected together. No physical subnets, VLANs, ACLs, or firewall rules are required to enable this isolation.

An isolated virtual network can be made up of workloads distributed anywhere in the data center. Workloads in the same virtual network can reside on the same or separate hypervisors. Workloads in multiple isolated virtual networks can reside on the same hypervisor. Isolation between virtual networks allows for overlapping IP addresses. This makes it possible to have isolated development, test, and production virtual networks, each with different application versions, but with the same IP addresses, all operating at the same time on the same underlying physical infrastructure.

Virtual networks are also isolated from the underlying physical infrastructure. Because traffic between hypervisors is encapsulated, physical network devices operate in a distinct address space from the workloads connected to the virtual networks. A virtual network could support IPv6 application workloads on top of an

IPv4 physical network. This isolation protects the underlying physical infrastructure from any possible attack initiated by workloads in any virtual network and is independent from any VLANs, ACLs, or firewall rules that would traditionally be required.

Network Segmentation

Segmentation is easy with network virtualization. Segmentation is related to isolation but applied within a multi-tier virtual network. Network segmentation is traditionally a function of a physical firewall or router, designed to allow or deny traffic between network segments or tiers (e.g., segmenting traffic between a web tier, application tier, and database tier). Traditional processes for defining and configuring segmentation are time consuming and highly prone to human error, resulting in a large percentage of security breaches. Implementation requires deep and specific expertise in device configuration syntax, network addressing, application ports, and protocols.

Network segmentation, like isolation, is a core capability of VMware NSX network virtualization. A virtual network can support a multi-tier network environment. Examples include multiple layer-2 segments with layer-3 segmentation or micro-segmentation on a single layer-2 segment using distributed firewall rules. These could represent a web tier, application tier and database tier. Physical firewalls and access control lists deliver a proven segmentation function, trusted by network security teams and compliance auditors. Confidence manual processes continues to fall as attacks, breaches, and downtime attributed to human error rise

In a virtual network, network services (e.g., layer-2, layer-3, ACL, firewall, QoS) that are provisioned with a workload are programmatically created and distributed to the hypervisor vSwitch. Network services, including layer-3 segmentation and firewalling, are enforced at the virtual interface. Communication within a virtual network never leaves the virtual environment, removing the requirement for network segmentation to be configured and maintained in the physical network or firewall.

Taking Advantage of Abstraction

Network security has traditionally required the security team to have a deep understanding of network addressing, application ports, and protocols as they are bound to network hardware, workload location, and topology. Network virtualization abstracts application workload communication from the physical network hardware and topology, allowing network security to break free from these physical constraints and apply network security based on user, application, and business context.

Advanced Security Service Insertion, Chaining and Steering

The base VMware NSX network virtualization platform provides basic stateful firewalling features to deliver segmentation within virtual networks. In some environments, there is a requirement for more advanced network security capabilities. In these instances, customers can leverage VMware NSX to distribute, enable, and enforce advanced network security services in a virtualized network environment. NSX distributes network services into the vSwitch to form a logical pipeline of services applied to virtual network traffic. Third party network services can be inserted into this logical pipeline, allowing physical or virtual services to be directly consumed.

Network security teams are often challenged to coordinate network security services from multiple vendors in relationship to each other. A powerful benefit of the NSX approach is its ability to build policies that leverage NSX service insertion, chaining, and steering to drive service execution in the logical services pipeline, based on the result of other services. This makes it possible to coordinate otherwise completely unrelated network security services from multiple vendors.

Logical Load Balancing

Load balancing is another network service available within NSX. This service offers distribution workload across multiple servers as well as high-availability of applications:

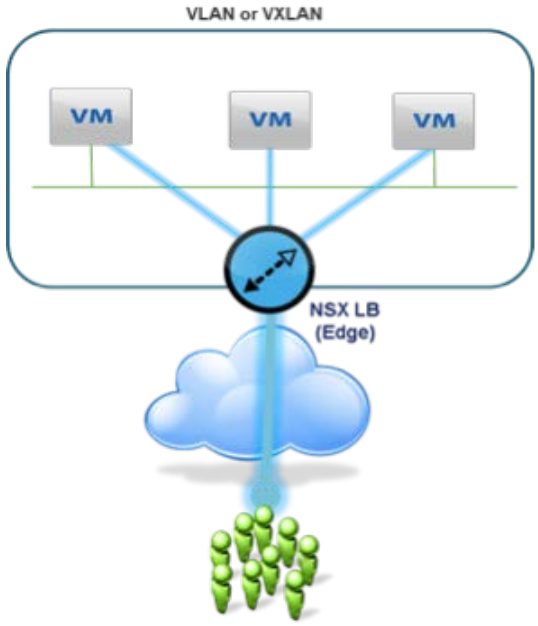

Figure 32. NSX Load Balancing

The NSX load balancing service is specially designed for cloud, being fully programmable via API and offering a common, central point of management and monitoring as other NSX network services.

The NSX load balancing service provides the following functionality:

- Multiple architecture support (one-armed/proxy mode or two-armed/transparent mode)
- Large feature set
- Broad TCP application support, including LDAP, FTP, HTTP, and HTTPS
- Multiple load balancing distribution algorithms; round robin, least connections, source IP hash, and URI
- Health checks for TCP, HTTP, and HTTPS including content inspection
- Persistence through source IP, MSRDP, cookie, and SSL session-id
- Throttling of maximum connections and connections/sec
- L7 manipulation including URL block, URL rewrite, and content rewrite
- Optimization of SSL offload

Each NSX Edge scales up to:

- Throughput: 9Gbps
- Concurrent connections: 1 million
- New connections per second: 131k

Figure 32 details examples of tenants with different applications with different load balancing needs. Each of these applications is hosted in the same cloud with the network services offered by NSX.

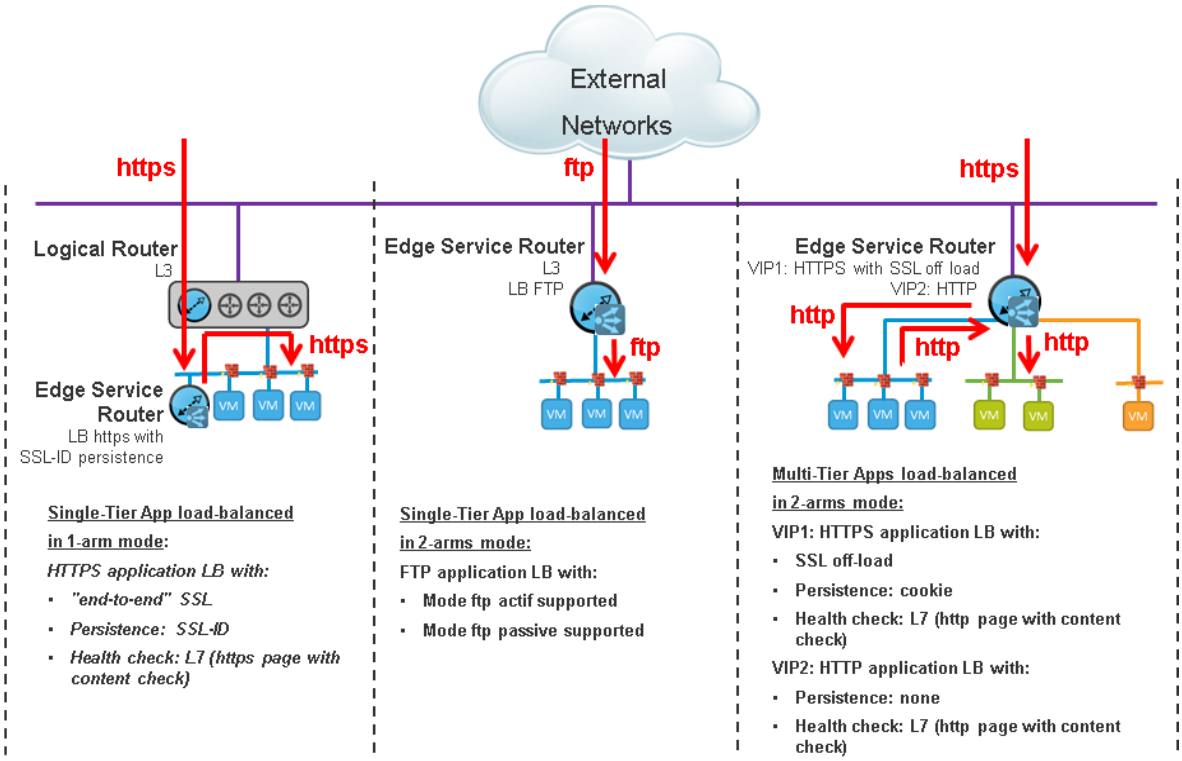

Figure 33. NSX Load Balancing

The NSX load balancing service is fully distributed. Multiple benefits from this architecture include:

- Each tenant has its own load balancer.
- Individual tenant configuration changes do not impact other tenants.
- A load increase on one tenant load balancer does not impact the scalability of other tenant's load balancers.
- Each tenant load balancing service can scale up to the maximum performance limits.

When utilizing load balancing services, other network services are still fully available. The same tenant can mix its load balancing service with other network services such as routing, firewalling, and VPN.

Integrating Visibility and Management with NSX and Arista

VMware NSX allows automated provisioning and context sharing across virtual and physical security platforms. Visibility and management services are traditionally deployed in a physical network environment with fixed access to network devices. When combined with traffic steering and policy enforcement at the virtual interface, they are easily provisioned and enforced in a virtual network environment. VMware NSX delivers a consistent model of visibility and security across applications residing on both physical and virtual workloads.

- **Existing tools and processes.** Dramatically increase provisioning speed, operational efficiency, and service quality while maintaining separation of duties between server, network, and security teams.
- **Control closer to the application, without downside.** This level of network security would traditionally have forced network and security teams to choose between performance and features. Leveraging the ability to distribute and enforce the advanced feature set at the application's virtual interface delivers the best of both.
- **Reduce human error in the equation.** The infrastructure maintains policy, allowing workloads to be placed and moved anywhere in the data center without any manual intervention. Pre-approved application security policies can be applied programmatically, enabling self-service deployment of even complex network security services.

Native NSX Features

As described earlier, NSX supports multiple visibility and management features - including VDS port mirroring, NetFlow/IPFIX, configuration backup and restore, network health check, QoS, and LACP – to provide a comprehensive toolkit for traffic management, monitoring, and troubleshooting within a virtual network.

Smart System Upgrade

Smart System Upgrade (SSU) reduces the burden of network upgrades, minimizing application downtime and reducing the risks taken during critical change controls across Arista switches and attached infrastructure. SSU provides a fully customizable suite of features that tightly couples data center infrastructure partners - such as Microsoft, F5, and Palo Alto Networks - with integration that allows devices to be seamlessly taken out of or put back into service. This helps systems stay current on the latest software releases without unnecessary downtime or systemic outages.

Network Telemetry

Network Telemetry is a new model for faster troubleshooting from fault detection to fault isolation within an Arista network. Network Telemetry agents in the Arista EOS platform stream data about network state, including both underlay and overlay network statistics, to applications from VMware, Splunk, ExtraHop, Corvil and Riverbed to provide real-time visibility into the state of the network as well as traffic running over the network.

Zero-Touch Provisioning (ZTP) with Zero-Touch Replacement (ZTR)

ZTP enables physical deployment of switches without configuration in a fully-automated, hands-free fashion out the box. With ZTP, a switch loads its image and configuration from a centralized location in the network. This simplifies deployment, enabling network-engineering resources to be used for more productive tasks. An extension to ZTP called ZTR enables switches to be physically replaced with the replacement switch picking up the same image and configuration as the switch it replaced.

Switch identity and configuration can be tied to the serial number, system MAC address, or more advanced location-based identifiers. ZTR dramatically reduces the time-to-restoration in the rare event of hardware failure and is not dependent on network engineer availability or physical site access.

Mobile Workload Management & Visibility using VM Tracer

Arista's VM Tracer feature for ESX and OpenStack, natively integrated into Arista EOS, automates discovery of directly connected virtual infrastructure, streamlining dynamic provisioning of related VLANs and port profiles on the network. Arista's switches utilize the VMware vCenter APIs to collect provisioning information. VM Tracer then combines this information with data from the switch's database to provide a clear and concise mapping of the virtual to physical network:

- Links Arista switches to VMware's vCenter configuration data and creates an adaptive infrastructure in which the network responds to changes in the virtual machine network
- Bridges the divide between the physical and virtual network making operational jobs easier by providing complete visibility into the virtual machine network and resources attached to it
- VM Tracer also supports the requirements of the virtualization and server administration teams by automating provisioning of VLANs in the physical infrastructure

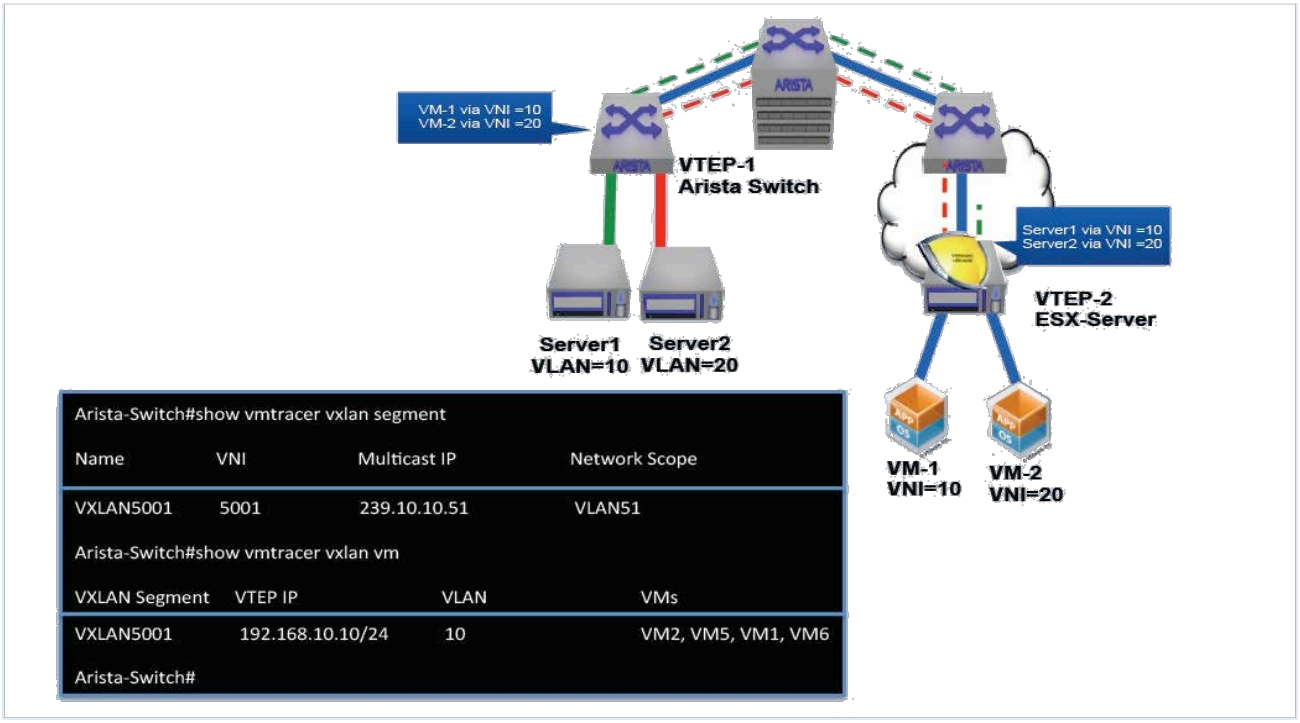

Figure 34. VM Tracer for VXLAN visibility exposes VMs and VXLAN information in VXLAN enabled network

VM Tracer utilizes vCenter API to collect provisioning information for virtual machines and provide visibility into where the VMs are created or moved to, what VXLANs they are part of, VLAN – VNI mapping for each VM amongst other useful information that this functionality provides (see Figure 34).

VM Tracer within the NSX VXLAN virtual overlay

Arista's VM Tracer revolutionizes how workloads are identified and tracked. VM Tracer has support for and awareness of VXLAN environments, with key functionality including: rapid identification of a virtual machine; direct control of policy bindings; and support for rapid auto-provisioning with Arista EOS. VM Tracer supports VMware vSphere virtualization environments.

Using LANZ

Arista Latency Analyzer (LANZ) enables tracking of network congestion in real time before it causes performance issues. Currently congestion detection and isolation are manual activities, often achieved through mirroring the problematic port to a packet capture device and waiting for the congestion problem reoccur.

With LANZ's proactive congestion detection and alerting capability, human administrators and integrated applications can:

- Preempt network conditions that induce latency or packet loss
- Adapt application behavior based on prevailing conditions
- Isolate potential bottlenecks early, enabling pro-active capacity planning
- Maintain forensic data for post-process correlation and back testing

Taking Advantage of EOS API

The Arista EOS API (eAPI) enables current and future applications and scripts to have complete programmatic control over EOS with a stable and easy to use syntax. eAPI exposes all state and all configuration commands for all features on Arista switches through a programmatic API.

Once eAPI is enabled, the switch accepts commands using Arista's CLI syntax and responds with machine-readable output and errors serialized in JSON. eAPI is language agnostic, can be integrated into any existing infrastructure and workflows, and can be utilized from scripts either on-box or off-box. All eAPI communication is served over HTTP or HTTPS.

Arista ensures that a command's structured output will remain forward compatible for multiple future versions of EOS, allowing end users to confidently develop critical applications without compromising their ability to upgrade to newer EOS releases and access new features.

Conclusions

The VMware network virtualization solution addresses current challenges with physical network and computing infrastructure, bringing flexibility, agility and scale to VXLAN-based logical networks. Along with the ability to create on-demand logical networks using VXLAN, the NSX Edge gateway helps users deploy various logical network services such as firewall, DHCP, NAT and load balancing. This is possible due to its ability to decouple the virtual network from the physical network and then reproduce the properties and services in the virtual environment.

Arista provides an ideal underlay network for VMWare NSX deployments for a number of reasons:

- Open Platform - Arista has a strong commitment to open standards and supports many NSX relevant protocols, including VXLAN and layer-3 ECMP.
- Proven scale out and well understood troubleshooting - Arista switches are deployed in 6 out of the 7 largest cloud data centers and a number of global financial institutions
- Fast provisioning - Zero Touch Provisioning provides a mechanism for rapid deployment of new switches or replacement of existing switches.
- Enhanced visibility - Arista has developed tools such as VM Tracer, LANZ (Latency Analyzer), and Network Telemetry, offering greater visibility into both physical and virtual networks.
- As the number of hosts and virtual machines continues to expand VMWare NSX and Arista will be able to serve the needs of the most demanding and scalable cloud networks.

In conclusion, Arista and VMware are delivering the industry's first scalable best-of-breed solution for network virtualization in the Software Defined Data Center. Cloud providers, enterprises and web 2.0 customers will be able to drastically speed business services, mitigate operational complexity, and reduce costs. All of this is

available now from a fully automated and programmatic SDDC solution that bridges the virtual and physical infrastructure.

References

[1] What's New in VMware vSphere 5.5

http://www.vmware.com/files/pdf/vsphere/VMware-vSphere-Platform-Whats-New.pdf

[2] vSphere 5.5 Configuration Maximums

http://www.vmware.com/pdf/vsphere5/r55/vsphere-55-configuration-maximums.pdf

[3] Arista solutions for Network Virtualization and SDCN

http://www.aristanetworks.com/en/solutions/network-virtualization

[4] VMware NSX Network Virtualization Platform whitepaper

https://www.vmware.com/products/nsx/resources.html

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com

Item No: VMW-NSX-NTWK-VIRT-DESN-GUIDE-V2-101