

Extending VMware NSX6.2 L2 Logical
networks to physical networks using DNOS
L2 Gateway Services (S6000/S4048)

Executive Summary

This paper outlines how DNOS9 running on Standard Networking switches like S6000/S4048 is used as a hardware VXLAN Layer2 Tunnel End Point (VTEP) to interconnect physical and virtual networks by using VMware NSX 6.2 (V) platform.

Virtual eXtensible LAN (VXLAN) is a standard-based Layer 2 overlay technology, defined in RFC 7348. VXLAN provides the same Ethernet Layer 2 network services as a VLAN, but with greater scalability, extensibility and flexibility. VXLAN provides multi-tenancy across the data centers by extending Layer 2 segments over Layer 3 boundaries. With VXLAN, up to 16M Layer 2 segments are possible in contrast to only 4K with a VLAN. VXLAN is suitable for large-scale deployments when a 4K Layer 2 segment is not enough. VXLAN is also used as an overlay solution to extend Layer 2 segments over. One of the common use case is to connect to virtualized applications with non-virtualized applications over an L2/L3 underlay Network domain.

VXLAN

VXLAN is an overlay technology based on RFC 7348. It uses UDP for transporting L2 MAC frames; it is a MAC-in-UDP encapsulation method. In VXLAN, the original Layer2 frame is encapsulated inside an IP-UDP packet by adding VXLAN header as shown in the figure.

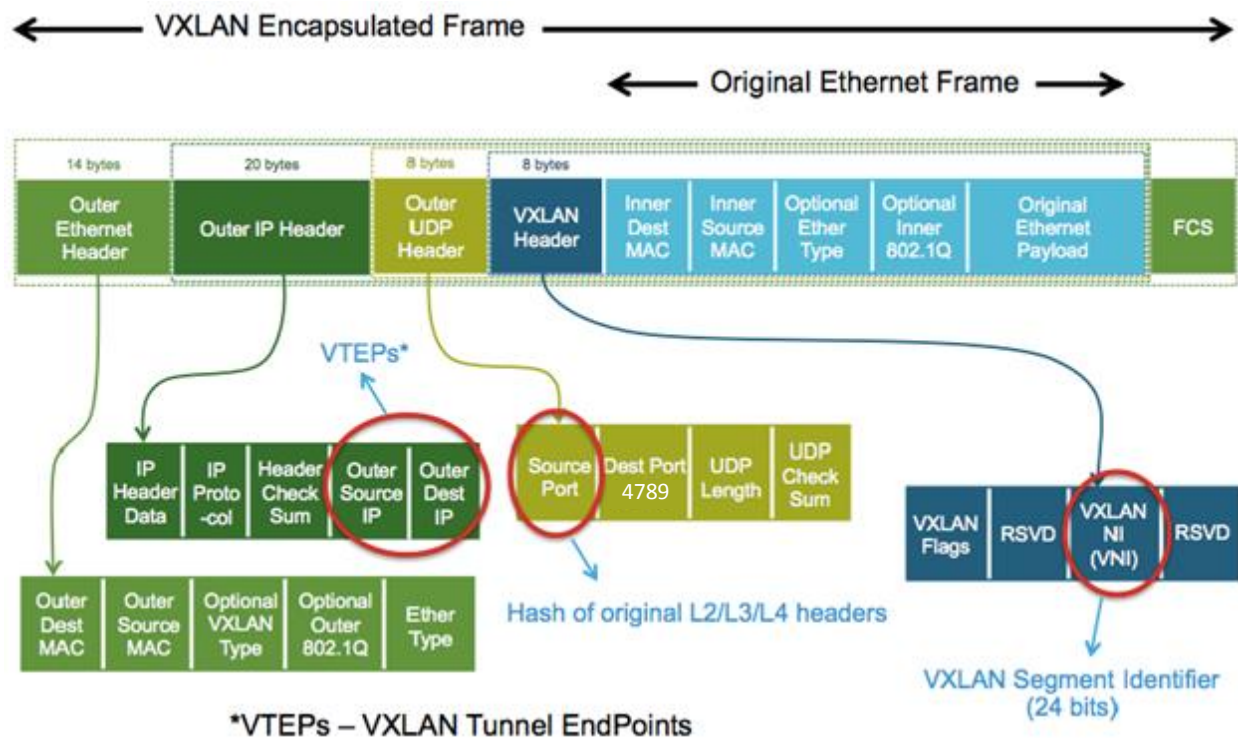


Figure 1 - VXLAN Header

By encapsulating the IP packet inside a UDP packet, 24-bit VXLAN ID helps solve number of problems in modern datacenters including 12-bit VLAN range limits, multi-tenancy in cloud computing environments, limits exponential MAC address table growth in TOR switches

VXLAN Tunnel Endpoint (VTEP)

VXLAN uses VTEP to perform underlay network to overlay network destination mapping for encapsulation and de-capsulation purposes. VTEPs primarily perform two important functions. In the underlay network, VTEPs provide end-to-end reachability to other VTEPs in the given NSX transport zone over L2 or L3 networks. In the overlay networks, VTEPs de-capsulate the outer header & VXLAN header and send the inner L2 packet to appropriate destination based on the VNI. In addition to forwarding the packets, VTEPs learn the MAC addresses of the VMs in the respective VNI and maintains a table with the NSX controller to limit the BUM traffic in the network.

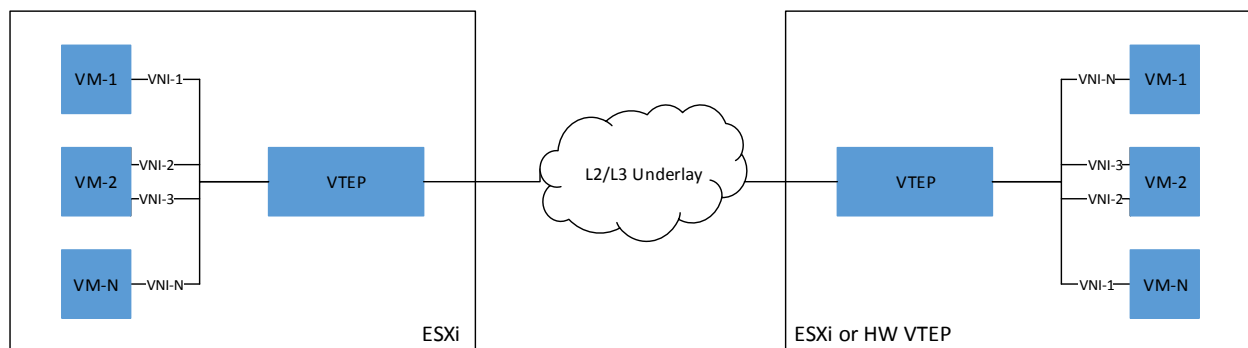


Figure 2 - Illustration of VTEPs

Software VTEP

A software VTEP typically is a physical server with hypervisor installed on it to host multiple VMs that could talk to NSX controllers over SSL authenticated TCP connection. VMware ESXi hypervisors running distributed vSwitch with NSX User World Agents or KVMs running Open vSwitch with Neutron-nicira plugin are a typical example of software VTEPs. An important challenge with software VTEPs is at some point the logical networks need to communicate with legacy physical networks. Because not all servers will be virtualized in a DC at the same time and virtualized application services need to talk to other applications in the physical networks as well. So to summarize, typical use cases include communication between,

1. VM in Logical network and baremetal servers
2. VM in Logical network and physical data centers

Hardware VTEP

The solution to solve the challenges with software VTEP is to develop a hardware VTEP device managed by NSX controller cluster. The switch will act as a L2 gateway on a TOR switch to connect physical servers/DCs to VMs in the logical networks. NSX will then manage physical ports and VLANs in the hardware VTEP to map them to logical networks. DNOS9 integrates with NSX and provides L2 gateway for terminating VXLAN tunnels controlled by the NSX controller. DNOS9 establishes a SSL encrypted TCP connection to communicate with the controllers via OVSDb protocol. Once the communication is established, NSX manages the DNOS9 switch to program the flow tables to provide connectivity between physical and logical networks.

The diagram illustrates a hybrid cloud network architecture. At the top, a green box labeled "Physical Servers" is connected to a blue box labeled "S4048 Hardware VTEP". This VTEP is also connected to a "Management Network" (represented by a vertical line) and an "SSL Cert" icon. Below the Hardware VTEP is a central switch structure consisting of four blue boxes: "Spine-1 S6000", "Spine-2 S6000", "Leaf-1 S6000", and "Leaf-2 S6000". These are interconnected in a mesh topology. At the bottom, a green box labeled "VMs" is connected to a blue box labeled "Software VTEP", which is also connected to the "Management Network" and an "SSL Cert" icon. A large green arrow on the right side points upwards, labeled "VXLAN Tunnel - L2 Logical Networks". A blue arrow on the right side points downwards, labeled "L3 Domain - OSPF".

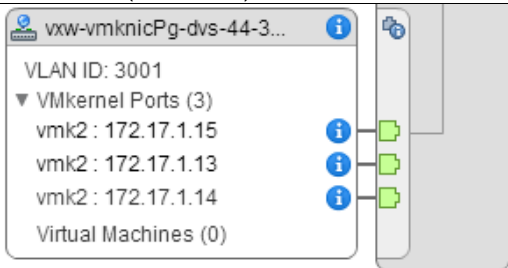
Figure 3 - DNOS L2 Gateway integration with VMware NSX

Topology Description

The topology consists of VMware NSX platform (i.e. vCenter registered with NSX manager and NSX controller cluster), VXLAN enabled ESXi hosts acting as software VTEPs, a DNOS9 installed Broadcom Trident2 based S6000/S4048 acting as hardware VTEP, any IP based L2 or L3 underlay network and a bare metal server. In this paper, by following the deployment steps listed below, we will establish a L2 overlay connection between VMs in logical networks with a bare metal server in physical network over a OSPF enabled L3 underlay network with the help of DNOS9 switch acting as a L2 Gateway to terminate the VTEP. As described earlier, the hardware VTEP could be part of a L2 underlay network as well to provide L2 gateway functionality.

1. Prepare L3 underlay networks

To create L2 connectivity between physical and logical networks it is necessary to enable end to end reachability between hardware VTEP and software VTEPs. We could use any L3 protocol to enable end to end reachability. As shown in the diagram, OSPF protocol is used here to ensure connectivity between VTEPs. In the DNOS9 switch, VTEP IP could be any valid IP in the system and for stability purpose we could use loopback IP as hardware VTEP IP

ESXi Host (SW VTEP)	S4048 (HW VTEP)
	<pre>interface Vlan 3003 ip address 172.17.3.4/24 tagged Port-channel 2 no shutdown ! interface Loopback 0 ip address 172.17.4.4/32 no shutdown ! router ospf 1 network 172.17.3.0/24 area 0 network 172.17.4.0/24 area 0</pre>
Leaf 1	Leaf 2
<pre>interface Vlan 3001 ip address 172.17.1.7/24 tagged Port-channel 20-28 no shutdown ! interface Vlan 3002 ip address 172.17.2.7/24 tagged Port-channel 1 no shutdown ! router ospf 1 network 172.17.2.0/24 area 0 network 172.17.1.0/24 area 0</pre>	<pre>interface Vlan 3001 ip address 172.17.1.8/24 tagged Port-channel 20-28 no shutdown ! interface Vlan 3002 ip address 172.17.2.8/24 tagged Port-channel 1 no shutdown ! router ospf 1 network 172.17.2.0/24 area 0 network 172.17.1.0/24 area 0</pre>
Spine 1	Spine 2
<pre>interface Vlan 3002 ip address 172.17.2.5/24 tagged Port-channel 1 no shutdown ! interface Vlan 3003 ip address 172.17.3.5/24 tagged Port-channel 2 no shutdown ! router ospf 1 network 172.17.2.0/24 area 0 network 172.17.3.0/24 area 0</pre>	<pre>interface Vlan 3002 ip address 172.17.2.6/24 tagged Port-channel 1 no shutdown ! interface Vlan 3003 ip address 172.17.3.6/24 tagged Port-channel 2 no shutdown ! router ospf 1 network 172.17.2.0/24 area 0 network 172.17.3.0/24 area 0</pre>

2. Create a Certificate file

Create a certificate file using '*crypto cert generate*' command. We could use the show file command to see the contents of the certificate file. The public key generated with this command is copied over to NSX service definition while adding the device to authenticate and establish a secure channel of communication between NSX and DNOS9 switch

```
DNOS#crypto cert generate cert-file flash://vtep-cert.pem key-file flash://vtep-privkey.pem
```

```
DNOS#show file vtep-cert.pem
```

```
-----BEGIN CERTIFICATE-----
MIIDkDCCAnigAwIBAgIBYzANBgkqhkiG9w0BAQUFADB6MQswCQYDVQQGEwJVUzEa
MBGGA1UEAwwRY29uZG9yLW1nbXQtczQwNDgxDTALBgNVBAoMBERlbGwxGDAWBgNV
BAsMD0RlbGwgTmV0d29ya2luZzERMA8GA1UEBwwIU0FOIEpvc2UxEzARBgNVBAGM
CkNhbgGlm3JuaWEwHhcNMTYwNDA3MDkyMjEyWhcNMjYwNDA1MDkyMjEyWjB6MQsw
CQYDVQQGEwJVUzEaMBGGA1UEAwwRY29uZG9yLW1nbXQtczQwNDgxDTALBgNVBAoM
BERlbGwxGDAWBgNVBAsMD0RlbGwgTmV0d29ya2luZzERMA8GA1UEBwwIU0FOIEpvc
2UxEzARBgNVBAGMCkNhbgGlm3JuaWEwggEiMA0GCSqGSIb3DQEBAQUAA4IBDwAw
ggEKAoIBAQDY5tjxtbykEA5NpIhA92cta0vA3/icFVWEKE80MQqU8u4h1KfrZ2of
GRTQj8apigoLhIYUIP5hhuFw6KXac7Nm3GSTFme1HYa4K+3df+XVPCPArFTigqf
m6IwZGvNIxgmz4fWPSSzKuCVscO+QtC8NoDf+Z23KgXlaCUh/+/eew9ir53ItGXS
iz23ZD9AB8tVH33+MZHDHtOQNbGQY2ShAgMBAAGjITAfMB0GA1UdDgQWBBTaOaPu
XmtLDTJVv++VYBiQr9gHCTANBgkqhkiG9w0BAQUFAAOCAQEAz8DtyPgZ2ozdUvcg
m9tIBGjLi6IAjBptSL9l2cSGNCeeJOmWXz+ZAqj/4kQpjyrgdymh086JrwF7N/Te
JavHnyOMYXPpENCjTfAqAkzPncnHZUG8335R1VPQ8VqR2k0PJdG1b5TuGTslHQUD
7qqybaK9/6vKRMbY8vMoQa14T0BFvCA7pr0sq4OrsTBKK3YGMESyOADDEpWPWrCq
PlU+JXzK6X30FToh+Kwvpl28FCbfA7pkRBNdhYOGKmpf777xZUutkJvgNZUPj5j0
YXyHpzWgszNxHrDTAHHIEs8V0An5HQ0UhdwXlcYmS2KDwZswlnWwQZ9MNg1Zjp8Q
Sy22Vw==
-----END CERTIFICATE-----
```

3. Add hardware L2 Gateway to Service definition

Navigate to NSX → Service Definitions → Hardware Devices to add the device and its certificate to the NSX service definition. To establish secure communication between NSX and DNOS9 switch, there should be reachability between NSX controllers and DNOS9 switch.

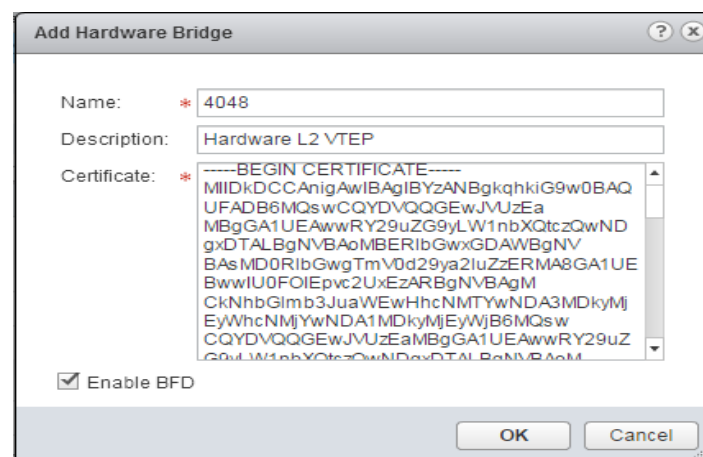


Figure 4 - Adding device certificate

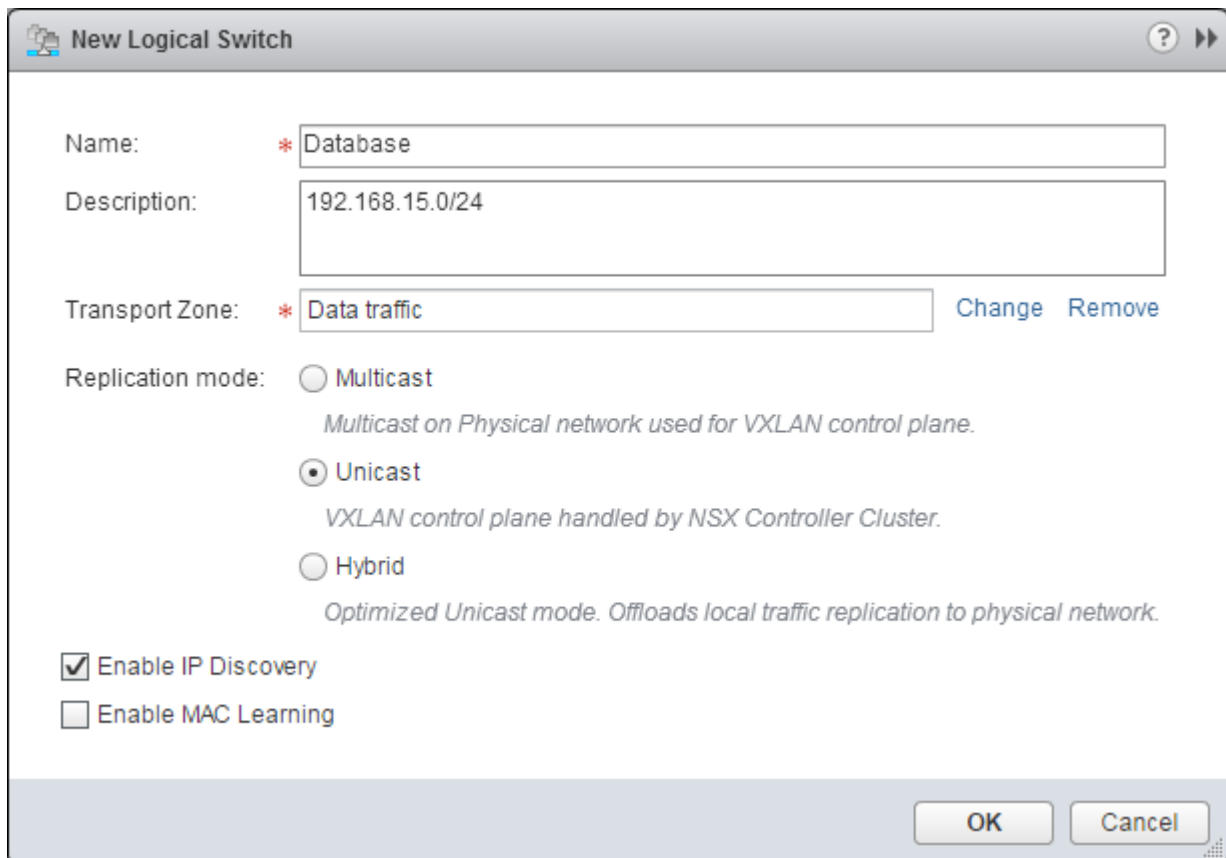
4. Configure VXLAN instance

To configure VXLAN in the DNOS9 switch, the VXLAN feature needs to be enabled using 'feature vxlan' command. Create a VXLAN configuration and configure the VTEP IP using the gateway command. NSX is installed typically with three controllers in a cluster. It is enough to give any one controller IP in the configuration. If the certificate copied to the NSX and switch certificate match, the authentication will succeed to create a secure OVSDB communication channel between NSX and DNOS9 switch.

```
feature vxlan          #Enable VXLAN feature
!
vxlan-instance 1      #Configure VXLAN Instance
 gateway-ip 172.17.4.4 #VXLAN Hardware VTEP IP
 fail-mode secure     #Specify mode on loss of connection
 controller 1 172.16.105.42 port 6640 ssl #Any one NSX controller IP
 no shutdown          #Enable the VXLAN Instance
```

5. Add HW VTEP ports to Logical switch

The logical switch can be created with appropriate Name/Description/Transport Zone and the replication mode.



The screenshot shows a 'New Logical Switch' dialog box with the following fields and options:

- Name:** * Database
- Description:** 192.168.15.0/24
- Transport Zone:** * Data traffic (with 'Change' and 'Remove' links)
- Replication mode:**
 - ☐ Multicast
Multicast on Physical network used for VXLAN control plane.
 - ☒ Unicast
VXLAN control plane handled by NSX Controller Cluster.
 - ☐ Hybrid
Optimized Unicast mode. Offloads local traffic replication to physical network.
- ☒ Enable IP Discovery
- ☐ Enable MAC Learning

At the bottom right are 'OK' and 'Cancel' buttons.

Figure 5 - Creating new logical network

Verify the logical switch is created in the correct transport zone. Navigate to Hardware tab under Service Definitions. Using Attach Logical Switch dialog box, select the appropriate Logical Switch and ports from the added device to create HW VTEP in the DNOS9 switch.

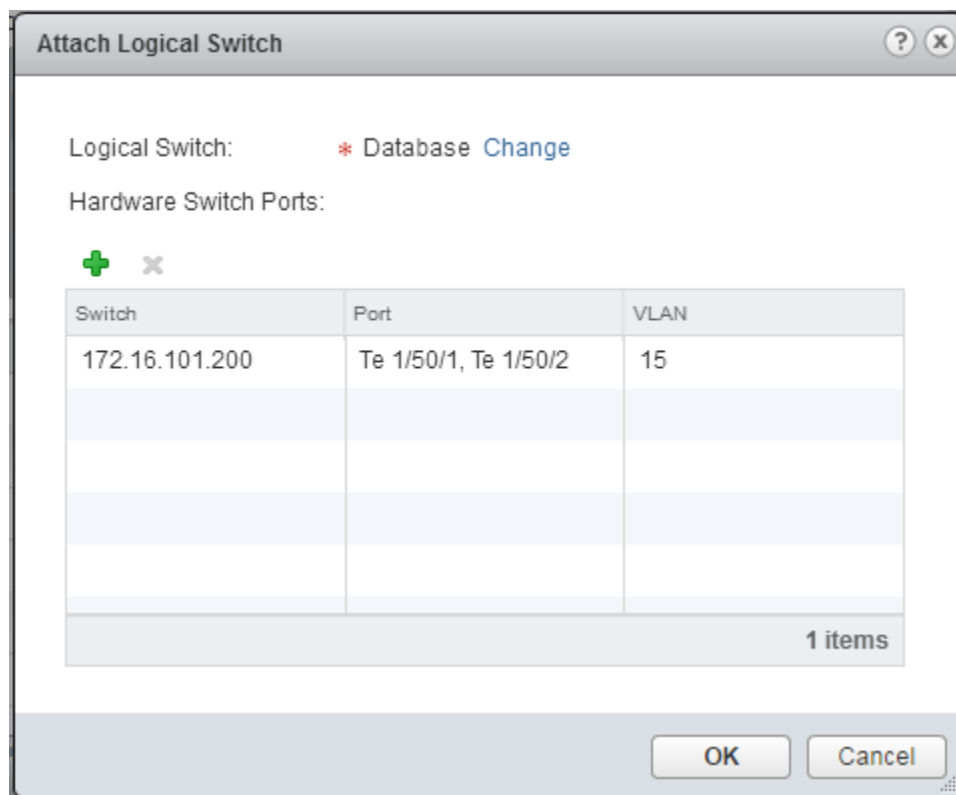


Figure 6 - Adding L2 Gateway ports to VNID

```
DNOS#show vxlan vxlan-instance 1 logical-network name e9470886-ecab-3743-
bb7e-a59420c245d2
Name          : e9470886-ecab-3743-bb7e-a59420c245d2
Description   :
Type          : ELAN
Tunnel Key    : 5009
VFI           : 28679
Port Vlan Bindings:
  Te 1/50/1: VLAN: 15 (0x80000015),
  Te 1/50/2: VLAN: 15 (0x80000014),
```

6. Configure replication cluster

Hardware VTEPs are not programmed to handle BUM traffic while performing L2 Gateway functionality. So it is important to configure replication cluster on the ESXi hosts prepared for NSX to handle BUM traffic. Under the hardware devices tab, edit the replication cluster to add the ESXi hosts. It is recommended to configure hosts in the compute cluster are configured as replication cluster. In the configured cluster only one host among the configured group of hosts handle the BUM traffic and multiple hosts are configured for redundancy purpose.

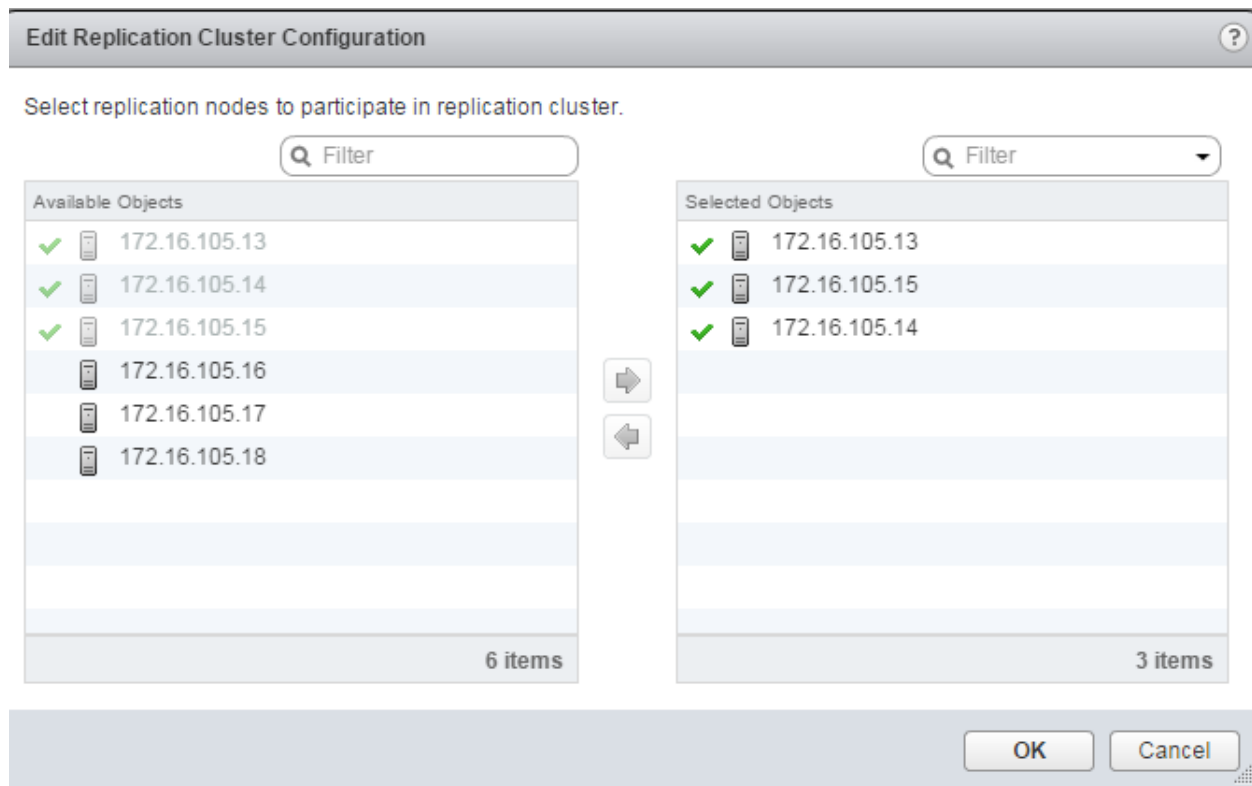


Figure 7 Adding Replication cluster

7. Verify end to end connectivity

As seen from the programmed logical network output, the VNID is 5009 and the VLAN tag 15. Associate the VM NIC in this VNID 5009. In the bare metal server, the incoming packet from the server should be tagged with VLAN 15. DNOS9 and NSX6.2 supports VLAN tagging, so multiple bare metal application servers could be aggregated at the TOR level with different VLAN IDs to enable end to end reachability. Initiate a ping request and check end to end L2 connectivity over L3 underlay. MAC addresses of bare metal servers as well as VMs could be verified with the show commands in the DNOS9 hardware L2 gateway with the following show commands.

MAC address learned in the Local ports

```
S4048#show vxlan vxlan-instance 1 unicast-mac-local
Total Local Mac Count:    3
VNI          MAC          PORT          VLAN
5009         00:50:56:a8:49:0f   Te 1/50/2    15
```

MAC address learned from the NSX controller

```
S4048#show vxlan vxlan-instance 1 unicast-mac-remote
Total Remote Mac Count:   3
VNI          MAC          TUNNEL
5009         00:50:56:92:ec:bf     172.17.1.15
```

Conclusion

VMware NSX6.2 provides a network virtualization platform using VXLAN technology to solve multi-tenancy problems in modern data centers. However in real data centers, it is very common to have mix of multi-tenant environment along with legacy servers. This paper describes how DNOS9 switches (S6000/S4048) are able to communicate with NSX controllers seamlessly and act like a NSX controlled L2 hardware gateway between logical networks and physical networks. By integrating with VMware NSX6.2, DNOS9 switches provides operational simplicity to traditional networks and help fully utilize the flexibility and automation capability provided by NSX to logical networks.