Nexus Validation Test Phase 3.5

1. Introduction

The Cisco Nexus line of data center products, hardware and software, must pass Cisco's comprehensive quality assurance process, which includes a multistage approach comprising extensive unit tests, feature tests, and system-level tests. Each successive stage in the process adds increasingly higher levels of complexity in a multidimensional mix of features and topologies.

This document describes the NVT Phase 3.5 network topologies, hardware and software configurations, and test procedures and findings.

NVT Phase 3.5 testing is performed on the following networks:

- Data Center 1 (DC1): This network focuses on building and operating a data center with the Nexus 7000 Sup1 as the core routing and switching component. It also covers interoperability with the Nexus 5000, Nexus 3000, Nexus 2000 and Catalyst 6500/4500 switches. This network uses virtual Port-Channel (vPC) and FabricPath to deliver highly available unicast and multicast services.

- Data Center 2 (DC2): This network focuses on building and operating a data center with the Nexus 7000 and 7700 Sup2E as the core routing and switching component. It also covers interoperability with the Nexus 6000, Nexus 5000, Nexus 3548, Nexus 2000 and Catalyst 6500/4500 switches. This network uses virtual Port-Channel (vPC) and FabricPath to deliver highly available unicast and multicast services.

- M1 vPC: This test bed focuses on scaling the virtual Port-Channel (vPC] with Nexus 7000. It also covers interoperability with Catalyst 6500. This network uses vPC and PVLAN to deliver high availability to servers connecting to data centers.

- Data Center 36 (DC36): This network focuses on building and operating a data center with the Nexus 3000 Series Switches. It covers interoperability with the Nexus 3048, Nexus 3064, and Catalyst 6500 switches. This network uses virtual Port-Channel (vPC) and ECMP to deliver highly available unicast and multicast services.

- Data Center 37 (DC37): This network focuses on building and operating a data center with the Nexus 3172 and its interoperability with the legacy Cisco platforms: Nexus 3048, Nexus 3548, Nexus 7000 and Catalyst 6500. This network uses virtual Port-Channel (vPC) and ECMP to deliver highly available unicast and multicast services.

- N9k GET: This network focuses on building and operating a data center with two Nexus 9058 as distribution switches and migrating core switches from Nexus 7000 Sup2E with Nexus 9508 as core routing and switching component. It also migrate a pair of TORs from Catalyst 4948 to Nexus 9396, and also cover its interoperability with Nexus 5548. This network uses vPC and PVLAN to deliver high availability to servers connecting to data centers.

- Operation: Network management including SNMP poling and inventory collection is performed through Data Center Network Management (DCNM) from Cisco and netMRI from Infoblox, TACAS+ authentication and syslog server. NetFlow is configured to export but the collector is not verified. NTP is synced to the server.

- In this phase, we have uplifted coverage for PVLAN, ACL and QoS based on Customer Found Defects (CFD) analysis. The coverage enhancement is performed on M1 and F1 line cards with Sup1 and Sup2. Corresponding profiles are DC1 and M1 vPC. The CFDs are:

    - QoS: CSCuo50598, CSCuo35180, CSCui58446, CSCuo71901, CSCuo71910, CSCuo93903, CSCuq06354, CSCue31241, CSCuq26934, CSCtg00407, CSCui69684, CSCtf39705, CSCur58946, CSCtr67673, CSCum20932
    - ACL: CSCuo71910, CSCuo79856, CSCun60847, CSCuo11751, CSCuo00001
    - PVLAN: CSCuo22348, CSCuo34849, CSCuo35180, CSCuo42047, CSCup02927, CSCur75014

2. NVT Topology Design Overview

2.1 DC1

2.1.1 Network Logical Topology Design Overview

The topologies and test cases validate high-available data center networks in order to provide unified fabric and computing services. This is achieved by using the Nexus 7010 and Nexus 5548 with features such as vPC and FabricPath.

2.1.1.1 Description of the Test Network

The data center site is built around the Nexus 7000 with Sup 1. This data center site is split into two halves:

- Nexus 7000 with vPC to Nexus 5000 and C4K/C6K for access
- Nexus 7000 with legacy ether-channel (trunk) with C4K
- Nexus 7000 with Nexus 2000 FEX
- Nexus 7000 (spine) with FabricPath to Nexus 5000 (leaf)
- Nexus 7000 with L3 to Nexus 3048

2.1.1.2 Hardware and Software Overview

	Model No.	**NVT 3.5**
N7K	**N7K SUP1 / F1 / M1**	6.2.12
N5K	**N5K-C5548UP-SUP**	7.0.1.N1.1
N3048	**N3K-C3048TP-1GE-SUP**	5.0.3.U5.1c
C6K	**VS-SUP720-10G**	151-1.SY
	WS-SUP720	122-33.SXJ4
C4K	**WS-X45-SUP7-E**	03.03.02.SG.151-1.SG2
	WS-C4948	150-2.SG6-6.10

2.1.1.3 Test Network Configuration

The following configurations are applied to the test network:

- Common system control, management and accounting: Common system features like SSH, TACACS+, Syslog, SNMP, NTP, SPAN, DNS and Management VRF are configured.
    - feature tacacs+ *# enabling the tacacs feature*
    - tacacs server host <ip address> key <0/7> *# configure the tacacs server to authenticate users*
    - aaa group server tacacs+ <group name> *# enable server groups for redundancy*
        - server <ip address>
        - use-vrf <vrf_name> *# use-vrf based on server reachability*
    - snmp-server user <user-name> <group-name> auth md5 <pass-phrase> priv <pass-phrase> localizedkey *# snmp v3 user with authentication enabled*
    - ntp server <ip address> *# enable ntp with server ip address*
    - ip domain-name <domain name> *# enable domain-name*
    - interface mgmt0 *# configure mgmt0*
        - vrf member management
        - ip address <ip_address >

- BGP: eBGP is configured between the core switches and the public cloud.
    - feature bgp *# enable bgp*
    - router bgp <autonomous-id> *# bgp autonomous -id*
    - router-id <router-id>
    - graceful-restart stalepath-time <120>
    - log-neighbor-changes
    - address-family ipv4 unicast
    - redistribute direct route-map <acl-name> *# route-map used for redistribution directly connected subnets*
    - redistribute ospf 1 route-map <acl-name> *# route-map used for redistribution OSPF routes*
    - maximum-paths <8>
    - maximum-paths ibgp <8>
    - neighbor <neighbor ip addres> remote-as 100090 *# BGP peer*
    - address-family ipv4 unicast
    - prefix-list NO_SELF in *# acl configured to restrict prefix import*

- OSPF: OSPF is the IGP running across the network. Each aggregation-access block is configured as a unique area with the core switches playing the role of the ABR.
    - feature ospf *# enable ospf for IPv4*
    - feature ospfv3 *# enable ospf for IPv6*
    - router ospf <instance-tag>
    - router-id <ip address>
    - redistribute bgp <as_no> route-map <acl-name> *# route-map used for redistribution for bgp routes*
    - log-adjacency-changes
    - timers throttle spf 100 200 500
    - timers throttle lsa 50 100 300

- auto-cost reference-bandwidth 1000000
- default-metric <1>

- PIM-SM: PIM Sparse Mode/PIM Any Source Multicast is deployed across the network to support multicast. Each aggregation-access block is configured with the RP for the locally sourced groups.
  - feature pim *# enable pim*
  - ip pim rp-address <rp-address> group-list <multicast-groups> *# configure static RP for a multicast group range*
  - ip pim send-rp-announce loopback2 prefix-list <multicast-groups> *# configure candidate auto-rp*
  - ip pim send-rp-discovery loopback2 *# configure auto-rp mapping-agent*
  - ip pim ssm range <> *# configure pim ssm for default range*
  - ip pim auto-rp forward listen *# enable auto-rp messages forwarding*

- MSDP Anycast RP: MSDP is deployed to exchange source information between Anycast RPs.
  - feature msdp *# enable msdp*
  - ip msdp originator-id <interface> *# configure source interface for msdp peering, generally loopback interface*
  - ip msdp peer <ip address> connect-source <interface> *# configure peer address*

- vPC: The vPC technology is deployed in the aggregation-access block DC1-Dist-N7k-101 as shown in Figure 1. In addition, dual-sided vPC is configured between the Nexus 7000 and Nexus 5000 switches
  - feature vpc *# enable vpc*
  - vpc domain <domain-id> *# configure vpc domain-id*
    - peer-switch *# enable peer-switch for faster STP convergence*
    - role priority 200 *# configure priority*
    - peer-keepalive destination <ip address> source <ip address> vrf vpc-keepalive *# configure keep-alive link*
    - peer-gateway *# enable peer-gateway to avoid vPC loop*
    - track <id> *# track the L3 core connectivity to avoid black-hole*
    - ip arp synchronize *# configure arp synchronize for faster convergence of address tables*

- FP: FabricPath is deployed in the aggregation block DC1-Dist-N7k-102. The spine layer is comprised of Nexus 7000 switches and the leaf switches are deployed using Nexus 5000 switches.
  - feature-set fabricpath *# configure feature-set fabricpath*
  - vlan <vlan-range>
    - mode fabricpath *# configure vlan-range in fabric path*

    - fabricpath switch-id <switch-id> *# configure switch-id*
    - vpc domain <domain-id> *# configure vpc domain-id*
        - fabricpath switch-id <vpc+_switch-id> *# configure virtual switch for peers present on the network*
    - interface port-channel <po> *# configure fabricpat interface*
        - switchport mode fabricpath *# configure fabricpath*

- STP: Rapid Spanning Tree Protocol is used to prevent Layer 2 loops in the aggregation-access blocks. The spanning tree root is placed on the aggregation level. BPDU Filter and PortFast Edge are configured on the access ports towards the hosts.
    - interface port-channel <po> *# configure port-channel*
        - switchport
        - switchport access vlan <vlan>
        - spanning-tree port type edge *# enable host*
        - spanning-tree bpdufilter enable *# configure bpdufilter*

- HSRP: HSRP is used as the first hop gateway protocol for hosts.
    - interface Vlan<id> *# configure svi*
    - ip access-group <acl> in *# enable access-list*
    - ip access-group <acl> out
    - no ip redirects
    - ip address <ip address>
    - hsrp version 2
    - hsrp 1
        - authentication md5 key-string cisco *# enable authentication*
        - preempt delay minimum 200
        - priority 200
        - ip <ip address> *# HSRP IP address*

- FEX: Fabric Extenders (Nexus 2000) are deployed on Nexus 7000

- IGMP: IGMP is used by hosts to join multicast groups of interest. IGMP snooping is enabled on all switches in the aggregation-access blocks to prevent flooding of multicast data traffic.
    - ip igmp snooping *# by default enabled on Nexus*

- LACP: LACP is used for link aggregation to form port-channels across the network.
    - feature lacp *# enable LACP, by default LACP is used on all port-channel*

- UDLD: UDLD aggressive mode is configured across the network to detect and prevent unidirectional links
    - feature udld *# enable feature udld*
    - udld aggressive *# udld aggressive mode is enabled to re-establish the connection with the neighbor*

- PVLAN: Isolated PVLAN configured between Nexus 7000 to Nexus 5000 over vPC at DC101
    - feature private-vlan *# enable feature private-vlan*
    - vlan <vlan-id> *# configure primary vlan*
        - private-vlan primary

- vlan <vlan-id>
  - private-vlan <isolated/community> *# configure secondary vlan*
  - private-vlan association <vlan-id> *# configure association with primary vlan*
- interface port-channel <port-channel *# configure port-channel*
  - switchport
  - switchport mode private-vlan trunk secondary
  - switchport private-vlan trunk allowed vlan 1 *# configure native vlan*
  - switchport private-vlan association trunk <primary> <secondary>

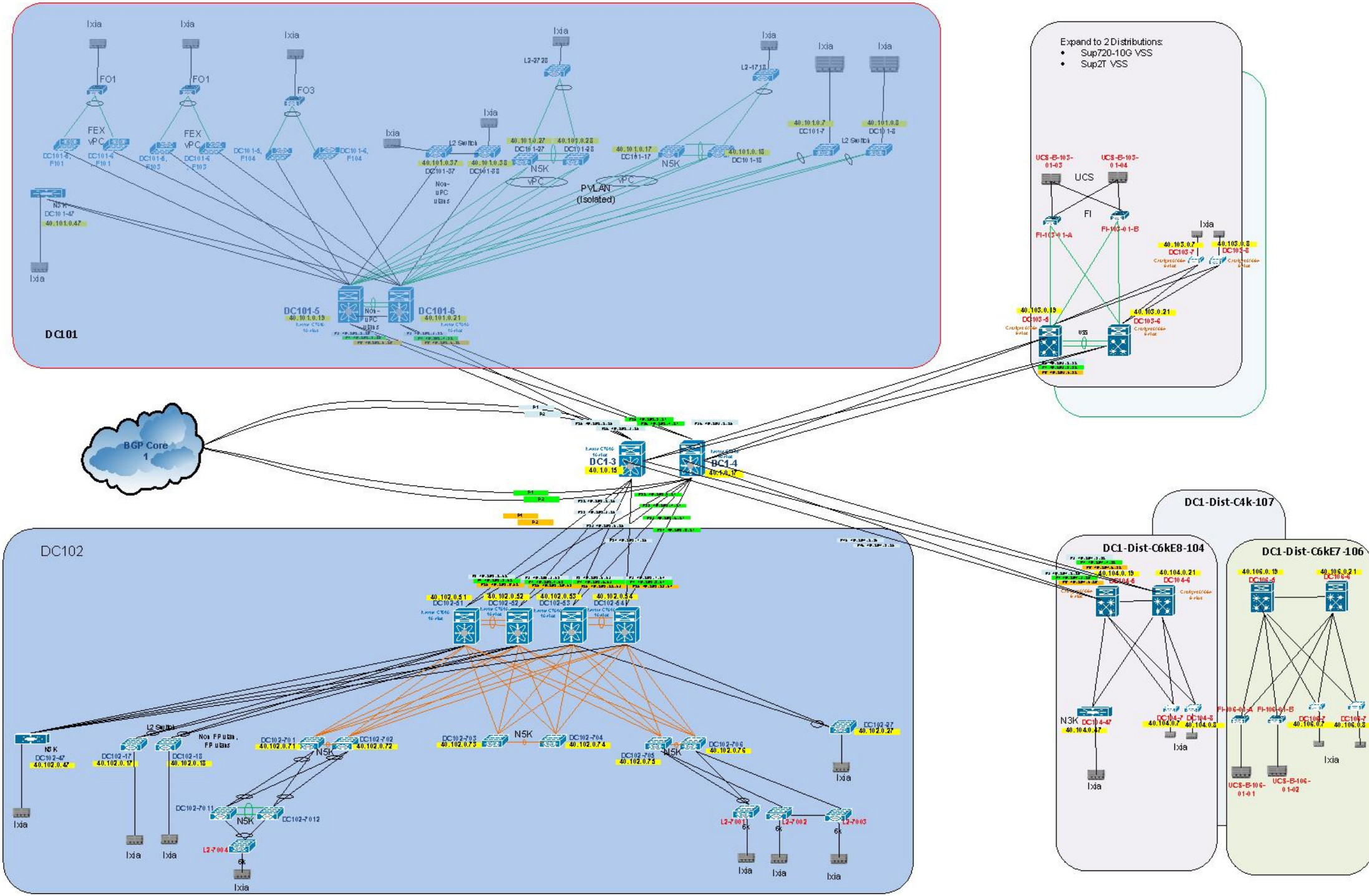


Figure 1: DC1 Topology

2.2 DC2

2.2.1 Network Logical Topology Design Overview

The topologies and test cases validate highly-available data center networks in order to provide unified fabric and computing services. This is achieved by using the Nexus 7000/Nexus 7700 at aggregation/core and Nexus 6000, Nexus 5000, Nexus 2000 and Nexus 3500 access switches.

2.2.1.1 Description of the Test Network

Figure 2 illustrates the test network topology of DC2 data center, which is built around Nexus 7000 with Sup 2E. This data center site is split into two halves:

- Nexus 7000 with vPC to Nexus 5000 for access.
- Nexus 7000 with FabricPath to Nexus 5000, Nexus 6000 and Nexus 7700. Nexus 2000 is connected to Nexus 7000 FabricPath spine and to FabricPath leaf's: Nexus 5000 and Nexus 6000.

While the majority of test cases focus on integrated solutions using Nexus switching, modular Catalyst switches are also included for interoperability between NX-OS and IOS.

2.2.1.2 Hardware and Software Overview

	Model No.	**NVT 3.2**
N7K	**N7K SUP2 / F3 10G**	6.2.12
N7700	**N7K SUP2**	6.2.12
N6000	**N6K-C6001-64P-SUP**	7.0(3)N1(1)
N5K	**N5K-C5548UP-SUP**	7.0.1.N1.1
N3548	**N3K-C3548P-10G-SUP**	6.0.2.A1.1e
C6K	**VS-SUP2T-10G**	150-1.SY3
	VS-SUP720-10G	122-33.SXJ4
	WS-SUP720	122-33.SXJ4
	WS-SUP32-GE	122-33.SXJ
C4K	**WS-X45-SUP7-E**	03.03.02.SG.151-1.SG2
	WS-C4948	150-2.SG6-6.9

2.2.1.3 Test Network Configuration

The following configurations are applied to the test network:

- Common system control, management and accounting: Common system features like SSH, TACACS+, Syslog, SNMP, NTP, SPAN, DNS and Management VRF are configured.
  - feature tacacs+ *# enabling the tacacs feature*
  - tacacs server host <ip address> key <0/7> *# configure the tacacs server to authenticate users*
  - aaa group server tacacs+ <group name> *# enable server groups for redundancy*
    - server <ip address>
    - use-vrf <vrf_name> *# use-vrf based on server reachability*
  - snmp-server user <user-name> <group-name> auth md5 <pass-phrase> priv <pass-phrase> localizedkey *# snmp v3 user with authentication enabled*
  - ntp server <ip address> *# enable ntp with server ip address*
  - ip domain-name <domain name> *# enable domain-name*
  - interface mgmt0 *# configure mgmt0*

        - vrf member management
        - ip address <ip_address >

- BGP: eBGP is configured between the core switches and the public cloud.
    - feature bgp *# enable bgp*
    - router bgp <autonomous-id> *# bgp autonomous -id*
    - router-id <router-id>
    - graceful-restart stalepath-time <120>
    - log-neighbor-changes
    - address-family ipv4 unicast
    - redistribute direct route-map <acl-name> *# route-map used for redistribution directly connected subnets*
    - redistribute ospf 1 route-map <acl-name> *# route-map used for redistribution OSPF routes*
    - maximum-paths <8>
    - maximum-paths ibgp <8>
    - neighbor <neighbor ip addres> remote-as 100090 *# BGP peer*
    - address-family ipv4 unicast
    - prefix-list NO_SELF in *# acl configured to restrict prefix import*

- OSPF: OSPF is the IGP running across the network. Each aggregation-access block is configured as a unique area with the core switches playing the role of the ABR.
    - feature ospf *# enable ospf for IPv4*
    - feature ospfv3 *# enable ospf for IPv6*
    - router ospf <instance-tag>
    - router-id <ip address>
    - redistribute bgp <as_no> route-map <acl-name> *# route-map used for redistribution for bgp routes*
    - log-adjacency-changes
    - timers throttle spf 100 200 500
    - timers throttle lsa 50 100 300
    - auto-cost reference-bandwidth 1000000
    - default-metric <1>

- PIM-SM: PIM Sparse Mode/PIM Any Source Multicast is deployed across the network to support multicast. Each aggregation-access block is configured with the RP for the locally sourced groups.
    - feature pim *# enable pim*
    - ip pim rp-address <rp-address> group-list <multicast-groups> *# configure static RP for a multicast group range*
    - ip pim send-rp-announce loopback2 prefix-list <multicast-groups> *# configure candidate auto-rp*
    - ip pim send-rp-discovery loopback2 *# configure auto-rp mapping-agent*
    - ip pim ssm range <> *# configure pim ssm for default range*
    - ip pim auto-rp forward listen *# enable auto-rp messages forwarding*

- MSDP Anycast RP: MSDP is deployed to exchange source information between Anycast RPs.
    - feature msdp *# enable msdp*
    - ip msdp originator-id <interface> *# configure source interface for msdp peering, generally loopback interface*
    - ip msdp peer <ip address> connect-source <interface> *# configure peer address*

- vPC: vPC technology is deployed in the aggregation-access block DC2-Dist-N7k-201. In addition, dual-sided vPC is configured between the Nexus 7000 and Nexus 5000 switches.
    - feature vpc *# enable vpc*
    - vpc domain <domain-id> *# configure vpc domain-id*
        - peer-switch *# enable peer-switch for faster STP convergence*
        - role priority 200 *# configure priority*
        - peer-keepalive destination <ip address> source <ip address> vrf vpc-keepalive *# configure keep-alive link*
        - peer-gateway *# enable peer-gateway to avoid vPC loop*
        - track <id> *# track the L3 core connectivity to avoid black-hole*
        - ip arp synchronize *# configure arp synchronize for faster convergence of address tables*

- FP: FabricPath is deployed in the aggregation blocks DC2-Dist-N7k-202. The spine layer is comprised of Nexus 7000 switches and the leaf switches are deployed using Nexus 5000, Nexus 6000 and Nexus 7700 switches.
    - feature-set fabricpath *# configure feature-set fabricpath*
    - vlan <vlan-range>
        - mode fabricpath *# configure vlan-range in fabric path*
    - fabricpath switch-id <switch-id> *# configure switch-id*
    - vpc domain <domain-id> *# configure vpc domain-id*
        - fabricpath switch-id <vpc+_switch-id> *# configure virtual switch for peers present on the network*
    - interface port-channel <po> *# configure fabricpat interface*
        - switchport mode fabricpath *# configure fabricpath*

- FP VLANs: On DC2-Dist-N7k-202, 2000 VLANs are deployed in mode FabricPath on all the spine and leaf devices..
- STP: Rapid Spanning Tree Protocol is used to prevent Layer 2 loops in the aggregation-access block DC-Dist-N7K-201. MSTP is enabled on DC-Dist-N7K-202 for the same purpose wherever applicable. The spanning tree root is placed on the aggregation level. BPDU Filter and PortFast Edge are configured on the access ports towards hosts.
    - interface port-channel <po> *# configure port-channel*
        - switchport
        - switchport access vlan <vlan>
        - spanning-tree port type edge *# enable host*
        - spanning-tree bpdufilter enable *# configure bpdufilter*

- SNMP: SNMP traps are enabled and SNMP scripts are used to collect system information and to monitor potential memory leaks.
- HSRP: HSRP is used as the first hop gateway protocol for hosts.
    - interface Vlan<id> *# configure svi*
    - ip access-group <acl> in *# enable access-list*
    - ip access-group <acl> out
    - no ip redirects
    - ip address <ip address>
    - hsrp version 2
    - hsrp 1
        - authentication md5 key-string cisco *# enable authentication*
        - preempt delay minimum 200

        - priority <priority>
        - ip <ip address> *# HSRP IP address*

- FEX: Multiple types of Fabric Extenders are deployed on Nexus 5000 parent switches.
- IGMP: IGMP is used by hosts to join multicast groups of interest. IGMP snooping is enabled on all switches in the aggregation-access blocks to prevent flooding of multicast data traffic.
    - ip igmp snooping *# by default enabled on Nexus*

- LACP: LACP is used for link aggregation to form port-channels across the network.
    - feature lacp *# enable LACP, by default LACP is used on all port-channel*

- UDLD: UDLD aggressive mode is configured across the network to detect and prevent unidirectional links
    - feature udld *# enable feature udld*
    - udld aggressive *# udld aggressive mode is enabled to re-establish the connection with the neighbor*

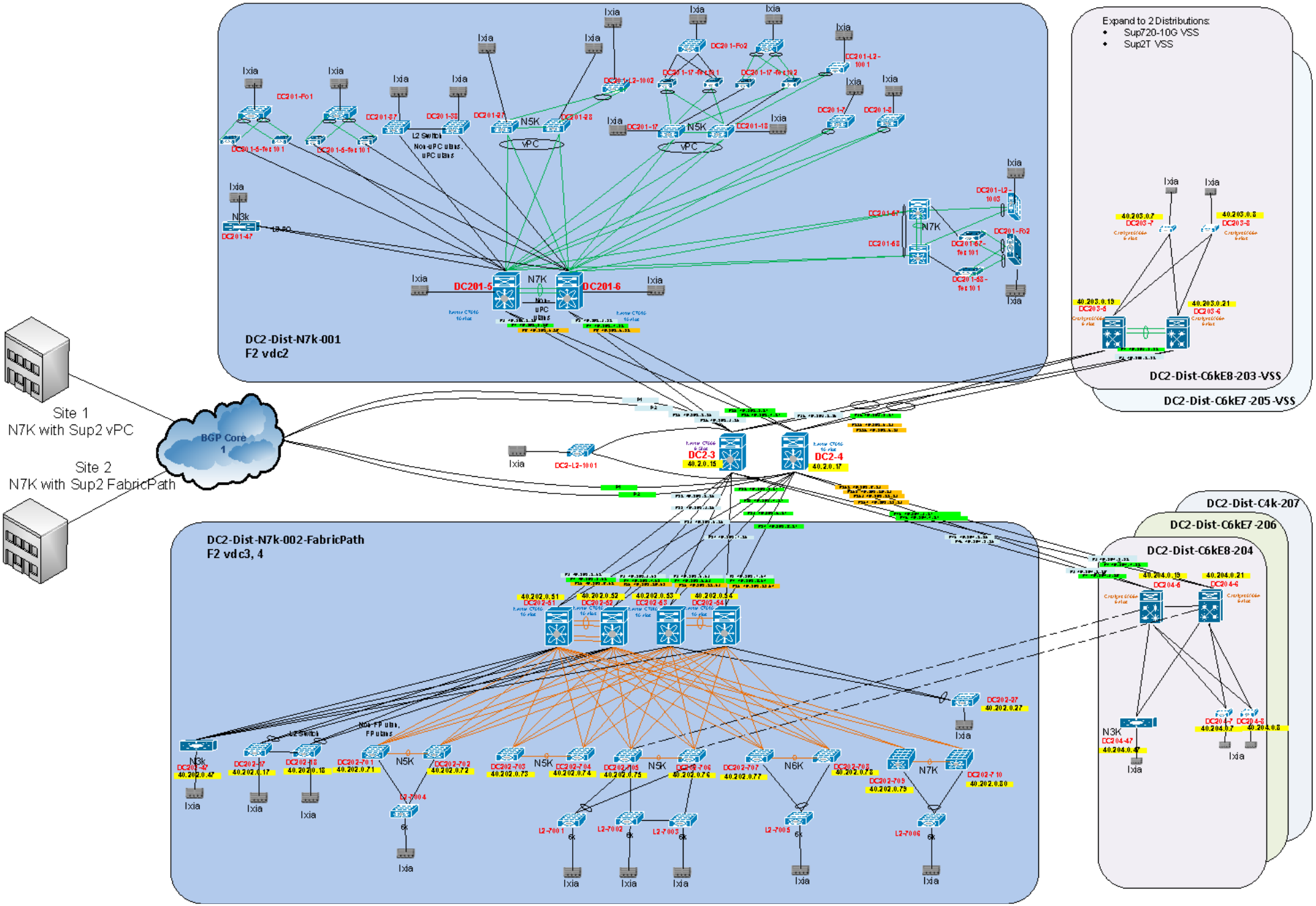


Figure 2: DC2 Topology

2.3 M1 vPC

2.3.1 Network Logical Topology Design Overview

The topology validates high-available networks that depict the various private VLAN feature implementations in order to provide an idea of the private VLAN scale numbers supported. This is achieved by using the Nexus 7000 and catalyst 6500 switches.

2.3.1.1 Network Logical Topology Design Overview

Figure 3 illustrates the network built around 2 Nexus 7000 switches with Sup2e and M1 modules. The topology contains:

- Nexus 7000 with VPC to Catalyst 6500 VSS switches for access
- Nexus 7000 connected to Catalyst 6500 switch with classical port-channels
- Nexus 7000 connected to Catalyst 6500 switches using orphan ports.

2.3.1.2 Hardware and Software Overview

	Model No.	**NVT 3.5**
N7K	**N7K SUP2E M1**	6.2.12
C6K	**VS-SUP720-10G**	122-33.SXJ1

2.3.1.3 Test Network Configuration

The following configurations are applied to the test network:

- Common system control, management and accounting: Common system features like SSH, Syslog, SNMP, NTP and Management VRF are configured.
    - snmp-server user <user-name> <group-name> auth md5 <pass-phrase> priv <pass-phrase> localizedkey *# snmp v3 user with authentication enabled*
    - ntp server <ip address> *# enable ntp with server ip address*
    - ip domain-name <domain name> *# enable domain-name*
    - interface mgmt0 *# configure mgmt0*
        - vrf member management
        - ip address <ip_address >

- PIM-SM: PIM Sparse Mode Multicast is deployed across the network to support multicast. The Nexus 7k are configured as RP.
    - feature pim *# enable pim*
    - ip pim rp-address <rp-address> group-list <multicast-groups> *# configure static RP for a multicast group range*

- Anycast RP: Anycast RP is deployed to exchange source information between RPs.
    - ip pim anycast-rp <rp-address> <ip-address-of-prospective-RP> *# configure loopback address as rp-address/ip-address-of-prospective-RP*

- vPC: vPC technology is deployed in the network between the N7k and the Catalyst VSS switches as shown in the figure 3.

    - feature vpc *# enable vpc*
    - vpc domain <domain-id> *# configure vpc domain-id*
        - peer-switch *# enable peer-switch for faster STP convergence*
        - role priority 200 *# configure priority*
        - peer-keepalive destination <ip address> source <ip address> vrf vpc-keepalive *# configure keep-alive link*
        - peer-gateway *# enable peer-gateway to avoid vPC loop*
        - track 10 *# track the L3 core connectivity to avoid black-hole*
        - ip arp synchronize *# configure arp synchronize for faster convergence of address tables*

- STP: Rapid Spanning Tree Protocol is used to prevent Layer 2 loops in the aggregation-access blocks. The spanning tree root is placed on the aggregation level. Root Guard is configured on the aggregation level to enforce root placement. BPDU Filter, BPDU Guard and PortFast Edge are configured on the access ports towards hosts.

    - interface port-channel <po> *# configure port-channel*
        - switchport
        - switchport access vlan <vlan>
        - spanning-tree port type edge *# enable host*
        - spanning-tree bpdufilter enable *# configure bpdufilter*

- LACP: LACP is used for link aggregation to form port-channels across the network

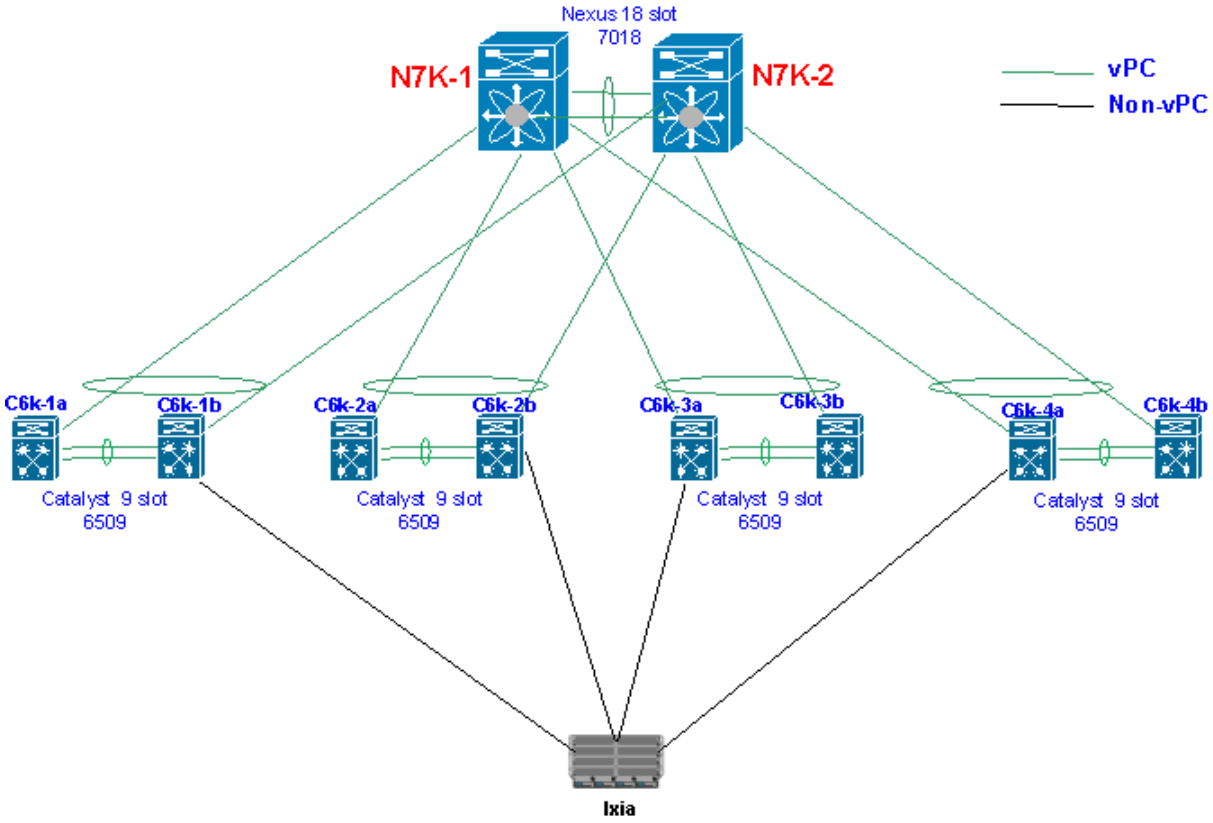
    - feature lacp *# enable LACP, by default LACP is used on all port-channel*

- PVLAN: PVLAN is configured in the network and is the main focus of testing. The following PVLAN components are covered in the network:
    - PVLAN primary and secondary VLAN(Community and Isolated)
    - Promiscuous access and promiscuous trunk on vPC
    - Private VLAN host on vPCnd orphan ports.
    - Private VLAN promiscuous trunk, secondary trunk on classic Port-channel

    - feature private-vlan *# enable feature private-vlan*
    - vlan <vlan-id> *# configure primary vlan*
        - private-vlan primary
    - vlan <vlan-id>
        - private-vlan <isolated/community> *# configure secondary vlan*
        - private-vlan association <vlan-id> *# configure association with primary vlan*
    - interface port-channel <port-channel> *# configure port-channel*
        - switchport
        - switchport mode private-vlan trunk secondary
        - switchport private-vlan trunk allowed vlan 1 *# configure native vlan*

        - switchport private-vlan association trunk <primary> <secondary>
    - interface port-channel <port-channel> *# configure port-channel*
        - switchport
        - switchport mode private-vlan trunk promiscuous
        - switchport private-vlan mapping trunk <primary secondary1, secondary2,…> *# configure primary to secondary mapping*
    - interface port-channel <port-channel> *# configure port-channel*
        - switchport
        - switchport mode private-vlan promiscuous *# configure promiscuous access*
        - switchport private-vlan mapping <primary secondary> *# configure primary to secondary mapping*
    - interface port-channel <port-channel> *# configure port-channel*
        - switchport
        - switchport mode private-vlan host *# configure host*
        - switchport private-vlan host-association <primary secondary> *# configure primary to secondary host-association*

Figure 3 M1vPC Topology



2.4 DC36

2.4.1 Network Logical Topology Design Overview

The topology analyzes and validates high-available networks of ECMP deployments for both unicast IPv4 and IPv6 traffic on the Nexus platforms in a typical spine/leaf structure.

2.4.1.1 Network Logical Topology Design Overview

Figure 4 illustrates the network built with N3K Series Switches. The topology contains:

- The spine layer: four Nexus 3048 switches and two Nexus 3064 switches
- The leaf layer for access: Nexus 3048, Nexus 3064 and Catalyst 6500 switches with ECMP connections to each of the six spine switches.

2.4.1.2 Hardware and Software Overview

Platform	**Model No.**	**NVT 3.5**
N3048	N3K-C3048TP-1GE-SUP	6.0(2)U5(1)
N3064	N3K-C3064PQ-10GE-SU	6.0(2)U5(1)
Cat 6500	WS-C6509-E	15.1(1)SY1
N7010	N7K-SUP2E	6.2(12)
N7010	N7K-SUP1	6.2(8a)

The following line cards are used on the Nexus 7000 devices:

- N7K-F248XP-25

The following line cards are used on the Catalyst 6500 device:

- WS-X6748-GE-TX
- WS-X6708-10GE

2.4.1.3 Test Network Configuration

The following configurations are applied to the test network:

- Common system control, management and accounting: Common system features like SSH, TACACS+, Syslog, SNMP, NTP, SPAN, DNS and Management VRF are configured.
    - feature tacacs+ *# enabling the tacacs feature*
    - tacacs server host <ip address> key <0/7> *# configure the tacacs server to authenticate users*
    - aaa group server tacacs+ <group name> *# enable server groups for redundancy*
        - server <ip address>
        - use-vrf <vrf_name> *# use-vrf based on server reachability*

- snmp-server user <user-name> <group-name> auth md5 <pass-phrase> priv <pass-phrase> localizedkey *# snmp v3 user with authentication enabled*
- ntp server <ip address> *# enable ntp with server ip address*
- ip domain-name <domain name> *# enable domain-name*
- interface mgmt0 *# configure mgmt0*
  - vrf member management
  - ip address <ip_address >

- BGP: eBGP is configured between the core switches and the public cloud.
  - feature bgp *# enable bgp*
  - router bgp <autonomous-id> *# bgp autonomous -id*
  - router-id <router-id>
  - graceful-restart-helper
  - log-neighbor-changes
  - address-family ipv4 unicast
  - maximum-paths <8>
  - maximum-paths ibgp <8>
  - neighbor <neighbor ip addres> remote-as 100090 *# BGP peer*
  - address-family ipv4 unicast
  - address-family ipv6 unicast

- OSPF/OSPFv3: OSPF is the IGP running across the network. Each aggregation-access block is configured as a unique area with the core switches playing the role of the ABR.
  - feature ospf *# enable ospf for IPv4*
  - feature ospfv3 *# enable ospf for IPv6*
  - router ospf <instance-tag>
  - router-id <ip address>
  - log-adjacency-changes
  - timers throttle spf 100 200 500
  - timers throttle lsa 50 100 300
  - auto-cost reference-bandwidth 1000000
  - default-metric <1>

- PIM-SM: PIM Sparse Mode/PIM Any Source Multicast is deployed across the network to support multicast. Each aggregation-access block is configured with the RP for the locally sourced groups.
  - feature pim *# enable pim*
  - ip pim rp-address <rp-address> group-list <multicast-groups> *# configure static RP for a multicast group range*
  - ip pim send-rp-announce loopback2 prefix-list <multicast-groups> *# configure candidate auto-rp*
  - ip pim send-rp-discovery loopback2 *# configure auto-rp mapping-agent*
  - ip pim ssm range <> *# configure pim ssm for default range*
  - ip pim auto-rp forward listen *# enable auto-rp messages forwarding*

- vPC: vPC technology is deployed on the leaf switch DC36-101 and DC36-102, DC36-105 and DC36-106 as shown in Figure 4.
  - feature vpc *# enable vpc*
  - vpc domain <domain-id> *# configure vpc domain-id*

        - peer-switch *# enable peer-switch for faster STP convergence*
        - role priority 200 *# configure priority*
        - peer-keepalive destination <ip address> source <ip address> vrf vpc-keepalive *# configure keep-alive link*
        - peer-gateway *# enable peer-gateway to avoid vPC loop*
        - track 10 *# track the L3 core connectivity to avoid black-hole*
        - ip arp synchronize *# configure arp synchronize for faster convergence of address tables*

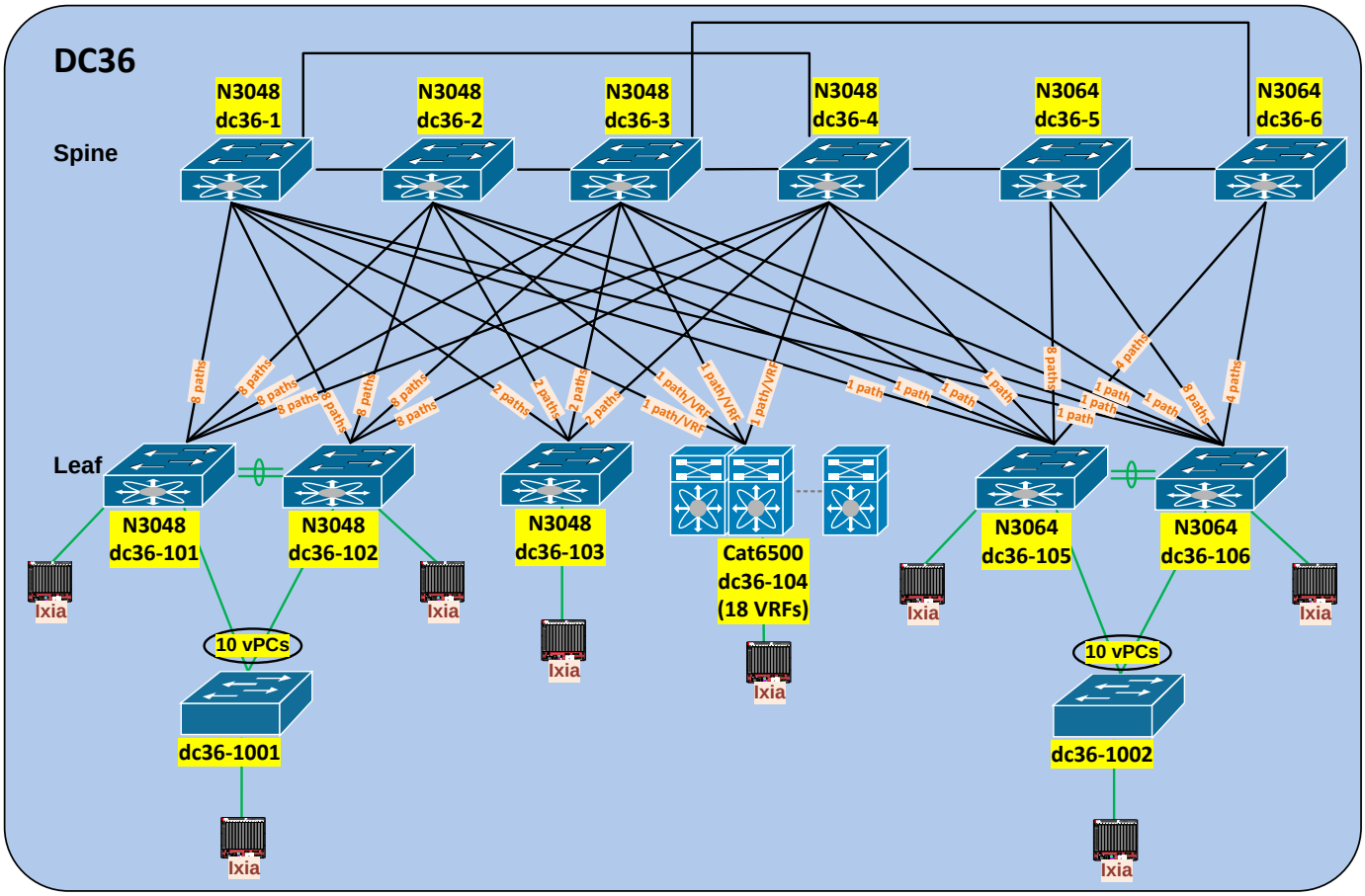
- HSRP: HSRP is used as the first hop gateway protocol for hosts.
    - interface Vlan11 *# configure svi*
    - no ip redirects
    - ip address <ip address>
    - ipv6 address <ipv6 address>
    - hsrp version 2
    - hsrp 1
        - authentication md5 key-string cisco *# enable authentication*
        - preempt delay minimum 200
        - priority 90 forwarding-threshold lower 1 upper 90
        - ip <ip address> *# HSRP IP address*
    - hsrp 101 ipv6
        - authentication md5 key-string cisco
        - preempt delay minimum 120
        - priority 90 forwarding-threshold lower 1 upper 90
        - ip <ipv6 address>

- IGMP: IGMP is used by hosts to join multicast groups of interest. IGMP snooping is enabled on all switches in the aggregation-access blocks to prevent flooding of multicast data traffic.
    - ip igmp snooping *# by default enabled on Nexus*

- LACP: LACP is used for link aggregation to form port-channels across the network.
    - feature lacp *# enable LACP, by default LACP is used on all port-channel*

- UDLD: UDLD aggressive mode is configured across the network to detect and prevent unidirectional links
    - feature udld *# enable feature udld*
    - udld aggressive *# udld aggressive mode is enabled to re-establish the connection with the neighbor*

Figure 4 DC36 Topology



2.5 DC37

2.5.1 Network Logical Topology Design Overview

The topology analyzes and validates high-available networks of ECMP deployments for unicast IPv4 traffic as well as multicast multipath traffic on the new Nexus 3172 and its interoperability with the pre-existing Cisco platforms: Nexus 3048, Nexus 3548 Nexus 7000 and Catalyst 6000.

2.5.1.1 Network Logical Topology Design Overview

Figure 5 illustrates the network built with N3K Series Switches. The topology contains:

- The spine layer: four Nexus 3172 switches
- The leaf layer for access: Nexus 3172, Nexus 3048, Nexus 3548 and Catalyst 6500 switches with ECMP connections to each of the four spine switches.

2.5.1.2 Hardware and Software Overview

Platform	Model No.	NVT 3.5
N3172	N3K-C3172PQ-10GE-SU	6.0(2)U5(1)
N3548	N3K-C3548P-10G-SUP	6.0(2)A4(1)
N3048	N3K-C3048TP-1GE-SUP	6.0(2)U4(1)
N7010	N7K-SUP2E	6.2(12)
N7010	N7K-SUP1	6.2(8a)
CAT6K	VS-SUP2T-10G	15.0(1)SY6

The following line cards are used on the Nexus 7000 devices:

- N7K-F248XP-25

The following line cards are used on the Catalyst 6509 devices:

- WS-X6704-10GE
- WS-X6716-10GE

2.5.1.3 Test Network Configuration

The following configurations are applied to the test network:

- Common system control, management and accounting: Common system features like SSH, TACACS+, Syslog, SNMP, NTP, SPAN, DNS and Management VRF are configured.
  - feature tacacs+ *# enabling the tacacs feature*
  - tacacs server host <ip address> key <0/7> *# configure the tacacs server to authenticate users*
  - aaa group server tacacs+ <group name> *# enable server groups for redundancy*
    - server <ip address>
    - use-vrf <vrf_name> *# use-vrf based on server reachability*
  - snmp-server user <user-name> <group-name> auth md5 <pass-phrase> priv <pass-phrase> localizedkey *# snmp v3 user with authentication enabled*
  - ntp server <ip address> *# enable ntp with server ip address*
  - ip domain-name <domain name> *# enable domain-name*
  - interface mgmt0 *# configure mgmt0*
    - vrf member management
    - ip address <ip_address >

- BGP: eBGP is configured between the core switches and the public cloud.
  - feature bgp *# enable bgp*
  - router bgp <autonomous-id> *# bgp autonomous -id*
  - router-id <router-id>
  - graceful-restart-helper
  - log-neighbor-changes
  - address-family ipv4 unicast

- maximum-paths <8>
- maximum-paths ibgp <8>
- neighbor <neighbor ip addres> remote-as 100090 *# BGP peer*
- address-family ipv4 unicast
- address-family ipv6 unicast

- OSPF: OSPF is the IGP running across the network. Each aggregation-access block is configured as a unique area with the core switches playing the role of the ABR.
  - feature ospf *# enable ospf for IPv4*
  - router ospf <instance-tag>
  - router-id <ip address>
  - log-adjacency-changes
  - timers throttle spf 100 200 500
  - timers throttle lsa 50 100 300
  - auto-cost reference-bandwidth 1000000
  - default-metric <1>

- PIM-SM: PIM Sparse Mode/PIM Any Source Multicast is deployed across the network to support multicast. Each aggregation-access block is configured with the RP for the locally sourced groups.
  - feature pim *# enable pim*
  - ip pim rp-address <rp-address> group-list <multicast-groups> *# configure static RP for a multicast group range*
  - ip pim ssm range <> *# configure pim ssm for default range*
  - ip pim auto-rp forward listen *# enable auto-rp messages forwarding*

- vPC: vPC technology is deployed on the leaf switch DC37-101 and DC37-102 as shown in Figure 5.
  - feature vpc *# enable vpc*
  - vpc domain <domain-id> *# configure vpc domain-id*
    - peer-switch *# enable peer-switch for faster STP convergence*
    - peer-keepalive destination <ip address> source <ip address> vrf vpc-keepalive *# configure keep-alive link*
    - peer-gateway *# enable peer-gateway to avoid vPC loop*
    - ip arp synchronize *# configure arp synchronize for faster convergence of address tables*

- HSRP: HSRP is used as the first hop gateway protocol for hosts.
  - interface Vlan11 *# configure svi*
  - no ip redirects
  - ip address <ip address>
  - ipv6 address <ipv6 address>
  - hsrp version 2
  - hsrp 1
    - authentication md5 key-string cisco *# enable authentication*
    - preempt delay minimum 200
    - priority 90 forwarding-threshold lower 1 upper 90
    - ip <ip address> *# HSRP IP address*
  - hsrp 101 ipv6

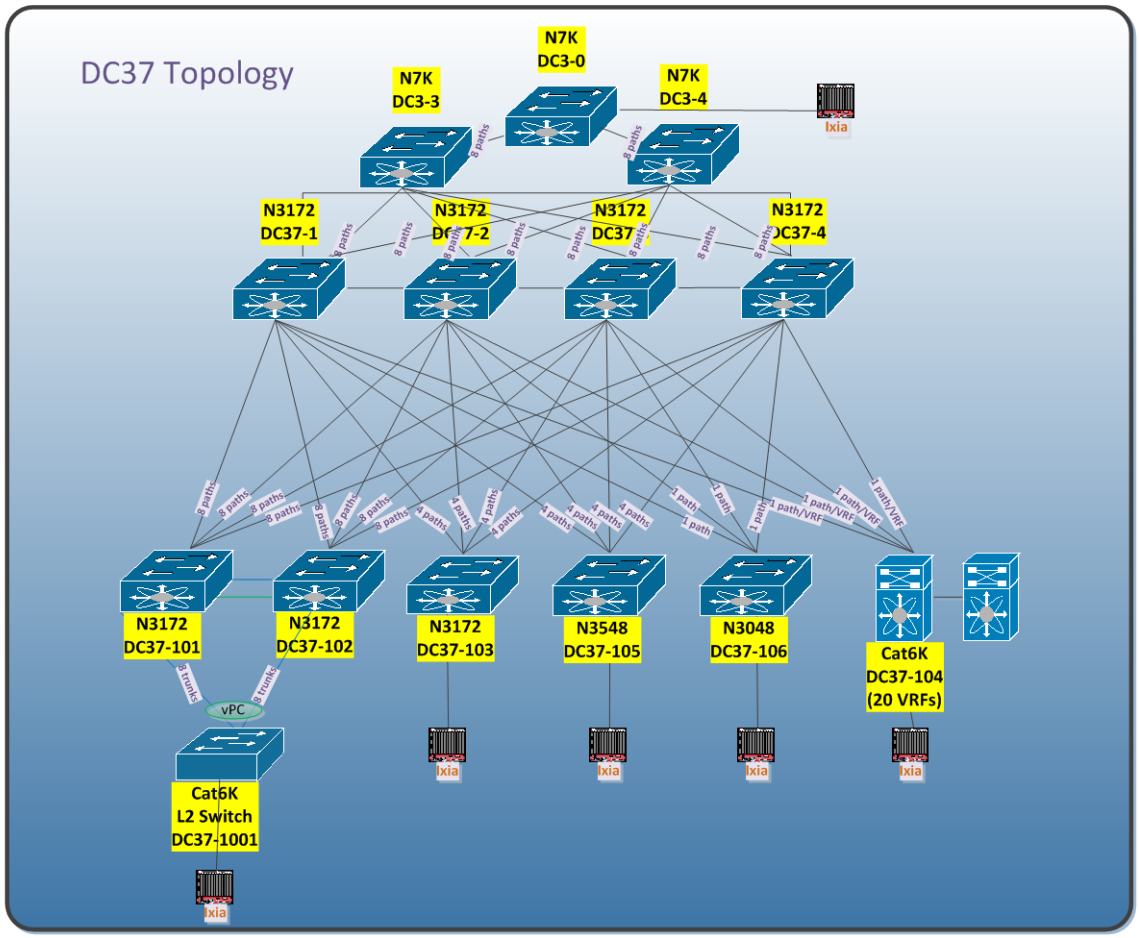
        - authentication md5 key-string cisco
        - preempt delay minimum 120
        - priority 90 forwarding-threshold lower 1 upper 90
        - ip <ipv6 address>

- IGMP: IGMP is used by hosts to join multicast groups of interest. IGMP snooping is enabled on all switches in the aggregation-access blocks to prevent flooding of multicast data traffic.
    - ip igmp snooping *# by default enabled on Nexus*

- LACP: LACP is used for link aggregation to form port-channels across the network.
    - feature lacp *# enable LACP, by default LACP is used on all port-channel*

- UDLD: UDLD aggressive mode is configured across the network to detect and prevent unidirectional links
    - feature udld *# enable feature udld*
    - udld aggressive *# udld aggressive mode is enabled to re-establish the connection with the neighbor*

Figure 5 DC37 Topologies



2.6 N9k GET

2.6.1 Network Logical Topology Design Overview

The topology validates high-available networks that deploy the Private VLAN (PVLAN) feature and validate PVLAN scale numbers supported. This is achieved by using the Nexus 9508 and Nexus 9396. The interoperability with the pre-existing Cisco platforms: Nexus 5548, Nexus 7000 and Catalyst 4948 also validated.

2.6.1.1 Network Logical Topology Design Overview

Figure 6 illustrates the network built around 2 Nexus 9508 as distribution switches. The topology contains:

- Nexus 9508 with vPV to a pair of TOR nexus 9396 switches for access
- Nexus 9508 with vPC to a pair of TOR nexus 5548 switches for access
- Nexus 9508 with vPC to catalyst 4948 switch for access
- Nexus 9508 with L3 routed port channel to core switches nexus 9508 and nexus 7000 with Sup2E

2.6.1.2 Hardware and Software Overview

Platform	Model No.	NVT 3.5
N9508	N9K-SUP-A	6.1(2)I3(4)
N9396	N9K-C9396PX	6.1(2)I3(4)
N5548	N5K-C5548UP-SUP	7.2(0)N1(1)
N7000	N7K-SUP2E	6.2(12)
Cat6k	VS-SUP2T-10G	12.2(50r)SYS2
CAT4948	WS-C4948-10GE	12.2(31)SG

The following line cards are used on the Nexus 9508 devices:

- N9K-X9636PQ
- N9K-X9464PX

The following line cards are used on the Nexus 7000 devices:

- N7K-F248XP-25E

The following line cards are used on the Catalyst 6504 devices:

- WS-X6704-10GE
- WS-X6716-10GE
- WS-X6708-10GE

2.6.1.3 Test Network Configuration

The following configurations are applied to the test network:

- Common system control, management and accounting: Common system features like SSH, TACACS+, Syslog, SNMP, NTP, SPAN, DNS and Management VRF are configured.
    - feature tacacs+ *# enabling the tacacs feature*
    - tacacs server host <ip address> key <0/7> *# configure the tacacs server to authenticate users*
    - aaa group server tacacs+ <group name> *# enable server groups for redundancy*
        - server <ip address>
        - use-vrf <vrf_name> *# use-vrf based on server reachability*
    - snmp-server user <user-name> <group-name> auth md5 <pass-phrase> priv <pass-phrase> localizedkey *# snmp v3 user with authentication enabled*
    - ntp server <ip address> *# enable ntp with server ip address*
    - ip domain-name <domain name> *# enable domain-name*
    - interface mgmt0 *# configure mgmt0*
        - vrf member management
        - ip address <ip_address >

- BGP: eBGP is configured between the core switches and the public cloud and between the distribution switches and the core switches.
    - feature bgp *# enable bgp*
    - router bgp <autonomous-id> *# bgp autonomous -id*
    - router-id <router-id>
    - graceful-restart-helper
    - log-neighbor-changes
    - address-family ipv4 unicast
    - maximum-paths <8>
    - maximum-paths ibgp <8>
    - neighbor <neighbor ip addres> remote-as 100090 *# BGP peer*
    - address-family ipv4 unicast
    - address-family ipv6 unicast

- OSPF: OSPF is the IGP running across the network. Each distribution-access block is configured as a unique area with the core switches playing the role of the ABR.
    - feature ospf *# enable ospf for IPv4*
    - router ospf <instance-tag>
    - router-id <ip address>
    - log-adjacency-changes
    - timers throttle spf 100 200 500
    - timers throttle lsa 50 100 300
    - auto-cost reference-bandwidth 1000000
    - default-metric <1>

- PIM-SM: PIM Sparse Mode/PIM Any Source Multicast is deployed across the network to support multicast. Each distribution-access block is configured with the RP for the locally sourced groups.
    - feature pim *# enable pim*
    - ip pim rp-address <rp-address> group-list <multicast-groups> *# configure static RP for a multicast group range*
    - ip pim ssm range <> *# configure pim ssm for default range*
    - ip pim auto-rp forward listen *# enable auto-rp messages forwarding*

- vPC: vPC technology is deployed on the distribution switches N9k-AGG-1 and N9k-AGG-2 as shown in Figure 6.
    - feature vpc *# enable vpc*
    - vpc domain <domain-id> *# configure vpc domain-id*
        - peer-switch *# enable peer-switch for faster STP convergence*
        - peer-keepalive destination <ip address> source <ip address> vrf vpc-keepalive *# configure keep-alive link*
        - peer-gateway *# enable peer-gateway to avoid vPC loop*
        - ip arp synchronize *# configure arp synchronize for faster convergence of address tables*

- HSRP: HSRP is used as the first hop gateway protocol for hosts.
    - interface Vlan11 *# configure svi*
    - no ip redirects
    - ip address <ip address>

- ipv6 address <ipv6 address>
- hsrp version 2
- hsrp 1
  - authentication md5 key-string cisco *# enable authentication*
  - preempt delay minimum 200
  - priority 90 forwarding-threshold lower 1 upper 90
  - ip <ip address> *# HSRP IP address*

- IGMP: IGMP is used by hosts to join multicast groups of interest. IGMP snooping is enabled on all switches in the distribution-access blocks to prevent flooding of multicast data traffic.
  - ip igmp snooping *# by default enabled on Nexus*

- LACP: LACP is used for link aggregation to form port-channels across the network.
  - feature lacp *# enable LACP, by default LACP is used on all port-channel*

- UDLD: UDLD aggressive mode is configured across the network to detect and prevent unidirectional links
  - feature udld *# enable feature udld*
  - udld aggressive *# udld aggressive mode is enabled to re-establish the connection with the neighbor*

- STP: Rapid Spanning Tree Protocol is used to prevent Layer 2 loops in the distribution-access blocks. The spanning tree root is placed on the aggregation level. Root Guard is configured on the aggregation level to enforce root placement. BPDU Filter, BPDU Guard and PortFast Edge are configured on the access ports towards hosts.

  - interface port-channel <po> *# configure port-channel*
    - switchport
    - switchport access vlan <vlan>
    - spanning-tree port type edge *# enable host*
    - spanning-tree bpdufilter enable *# configure bpdufilter*

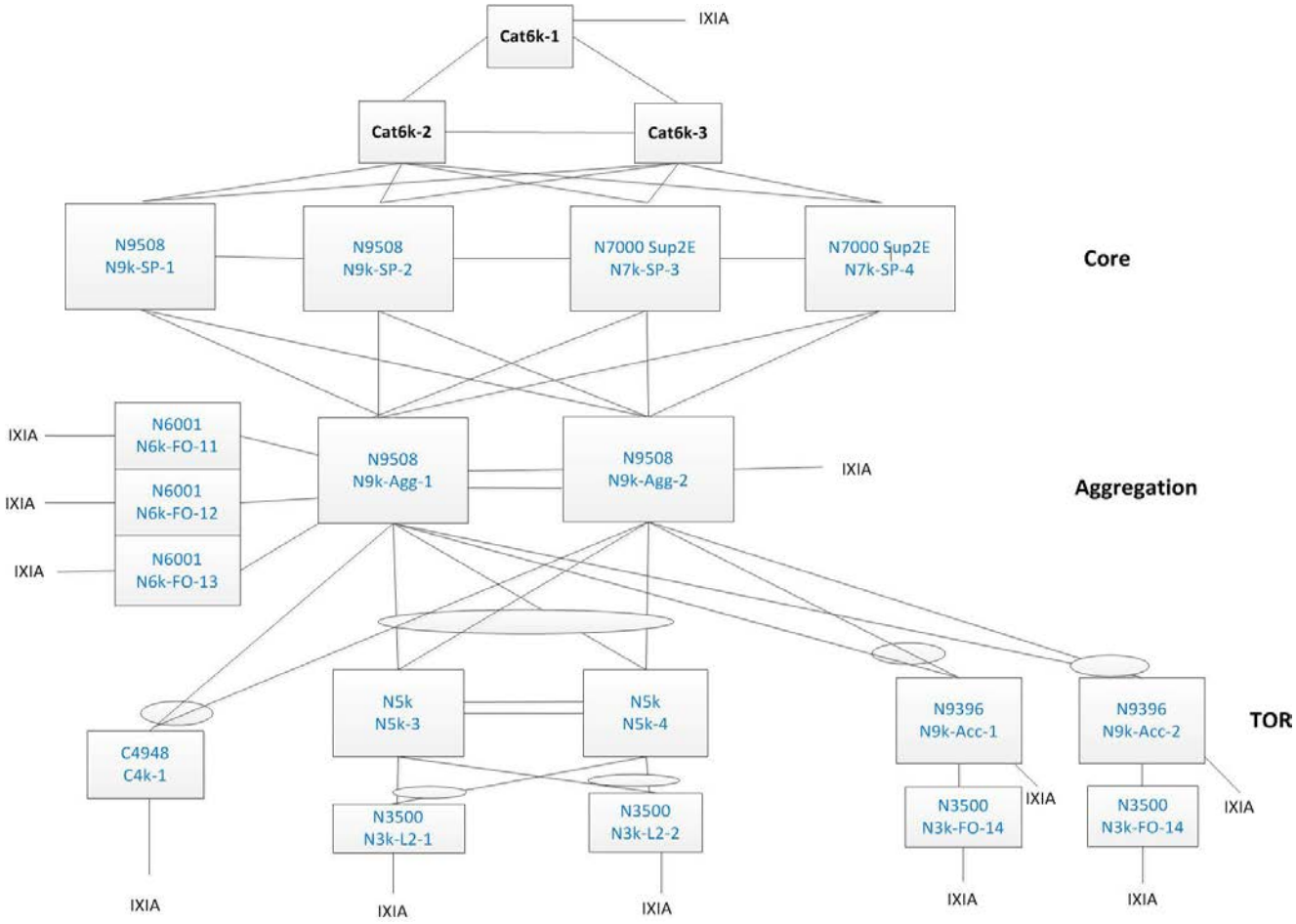
- DHCP Relay: DHCP reply is configured in distribution switches to forwards DHCP packets between host clients and DHCP server.
  - feature dhcp *# enable feature dhcp*
  - ip dhcp relay *# enable dhcp relay agent*
  - interface VLAN XXX *# configure dhcp server address*
    - ip dhcp relay address x.x.x.x

- PVLAN: PVLAN is configured in the network and is the main focus of testing for N9k GET. The following pvlan components are covered in the network:
  - PVLAN primary and secondary vlan(Isolated and community)
  - PVLAN promiscuous access port and promiscuous trunk port on distribution switches
  - PVLAN host ports on TOR switches
  - PVLAN Secondary trunk port on distribution switches and TOR switches
  - PVLAN primary vlan SVI on distribution switches

  - feature private-vlan *# enable feature private-vlan*

- vlan <vlan-id> *# configure primary vlan*
  - private-vlan primary
- vlan <vlan-id>
  - private-vlan <isolated/community> *# configure secondary vlan*
  - private-vlan association <vlan-id> *# configure association with primary vlan*
- interface Ethernet <x/y > *# configure secondary trunk port*
  - switchport
  - switchport mode private-vlan trunk secondary
  - switchport private-vlan trunk native vlan xxx *# configure native vlan*
  - switchport private-vlan trunk allowed vlan
  - switchport private-vlan association trunk <primary> <secondary>
- interface Ethernet <x/y > *# configure promiscuous trunk port*
  - switchport
  - switchport mode private-vlan trunk promiscuous
  - switchport private-vlan trunk native vlan xxx*# configure native vlan*
  - switchport private-vlan trunk allowed vlan
  - switchport private-vlan mapping trunk <primary secondary1, secondary2,…> *# configure primary to secondary mapping*
- interface Ethernet <x/y > *# configure promiscuous port*
  - switchport
  - switchport mode private-vlan promiscuous *# configure promiscuous access*
  - switchport private-vlan mapping <primary secondary> *# configure primary to secondary mapping*
- interface Ethernet <x/y > *# configure host port*
  - switchport
  - switchport mode private-vlan host *# configure host*
  - switchport private-vlan host-association <primary secondary> *# configure primary to secondary host-association*
- interface VLAN XXX *# configure primary VLAN SVI*
  - private-vlan mapping <secondary vlan>

Figure 6 N9k-GET Topologies



3 Scale Numbers tested by NVT

Feature/Parameter	DC1	DC2	M1vPC	DC36	DC37	N9k GET
VLAN	100	200		500	100	50
Fabric Extender	3	2		0	0	0
VLAN per FEX	20	20		0	0	0
MAC	7K	2K		2K	3K	4k
VPC	101	100		10	10	4
VLANS per VPC	20	20		10	10	
Fabric Path IS-IS adjacencies	50	50		0	0	0
Fabric Path Number of Switch ID's	50	50		0	0	0
HSRP V2	80	150		200	110	20
VRF	1					1
OSPF Peers	14	70		3	4	4
OSPF Routes	10K	10K		100	9	17
eBGP Sessions	2	2		2	2	4
Primary VLAN			25			10
Secondary VLAN			75			16

Physical ports used for PVLAN			200			121
Port-channel			100			10
Port-channel in PVLAN			32			0
VPC Scale with PVLAN			32			0
Host mode			20			40
Promiscuous Access			16			20
Promiscuous Trunk			150			101
Trunk Secondary			30			2

4 ISSU Matrix

ISSU/Cold boot	**DC1 (SUP1)**
ISSU 6.2.8a -> 6.2.12	Pass
ISSU 6.2.8b -> 6.2.12	Pass
ISSU 6.2.10 -> 6.2.12	Pass
Cold boot 6.2.10 -> 6.2.12	Pass

5 NVT Findings/Conclusion/Recommendations

Assigned/New	➔	*Still working on fixes and may be seen in CCO image*
Unreproducible	➔	*Not seen in CCO image may have fixed by other code fixes.*
Verified/Resolved	➔	*Fixed in CCO image*
Closed	➔	*System limitation and behavior will remain the same*

CSCus28111:

Symptom: PVLAN recreation resulted in Error / Ports in err disabled state
Conditions: In a sequence that deletes a range of pVLAN and recreate it, Some of the PVLAN resulted in error from vlan manager during the recreation and some ports get into error disabled state
Workaround: None.
Severity: Moderate
Status: Assigned
Platform Seen: Nexus 7000
Resolved Releases:
Applicable Releases: 6.2(12)

CSCus55931

Symptom: Traffic loss when the vPC peer link is shut for the receivers belonging to the private vlan secondary.
Conditions: Problem happens with private vlan configuration only. the receivers are present on the secondary vlans and the loss is seen for 60 seconds maximum for such receivers
Workaround: Possibly increasing the igmp query interval might help in recovering the reports quickly.

Severity: Moderate
Status: Assigned
Platform Seen: Nexus 7000
Resolved Releases:
Applicable Releases: 6.2(12)

CSCus76724
Symptom: On M1XL linecards, when some vlan config causes a private-vlan association to be non-operational , private-vlan trunk secondary port sees traffic loss. Similarly, when the trunk association is unconfigured and re-configured on private-vlan trunk-secondary port, the issue might be observed.
Conditions: This issue is seen on M1XL linecards. Will not be seen with M1 and F-series linecards
Workaround: Workaround is to do a shut no-shut on the port or PC or VPC leg where the issue is observed.
Severity: Severe
Status: Assigned
Platform Seen: Nexus 7000
Resolved Releases:
Applicable Releases: 6.2(12)

CSCus71454
Symptom: In a Private-Vlan VPC setup in private-vlan host mode, when peer link flaps, VPC leg in private-vlan host mode also flaps and comes back up in some time. There will be traffic loss from the VPC leg until the leg bringup happens again.
Conditions: The VPC legs have to be private-vlan host mode as follows: "switchport mode private-vlan host.
Workaround: None
Severity: Severe
Status: Assigned
Platform Seen: Nexus 7000
Resolved Releases:
Applicable Releases: 6.2(12)

CSCus13116
Symptom: N3K/Multipath: Mcast convergence degradation on link flap vs 6.0(2)U4(1)
Conditions: Upon link flap, between spine layer and leaf layer, multicast traffic convergence has degraded significantly when compared to 6.0(2)U4(1) multicast convergence numbers. The degradation is noticed ONLY with multipath, for both local and remote interface flaps.
Workaround:
Severity: Severe
Status: Closed
Platform Seen: Nexus 3048
Resolved Releases:
Applicable Releases: 6.0(2)U5(1)

Caveats for N9k-GET
CSCut15296/CSCut15002
Symptom: Can't remove private-vlan association after change primary VLAN to isolated VLAN
Conditions: On VLAN configuration mode “vlan xxx” for a primary PVLAN, if “no private-vlan primary” first, then “no vlan xxx”, then configure VLAN xxx as “private-vlan isolated”. At this

situation, “private-vlan association” under isolated VLAN can’t be removed from VLAN xxx anymore.

Workaround: None
Severity: Moderate
Status: Duplicated
Platform Seen: Nexus 9508, Nexus 9396
Resolved Releases:
Applicable Releases: 6.1(2)I3(4)

CSCus64028

Symptom: pvlan macs out of sync across vpc peers after flapping peer-link and keepalive link
Conditions: On vPC set up, when shut/no shut peer-link and keepalive link together, after two vPC peers are back to normal state, a few of macs are not in sync over vPC peer. The macs in secondary vPC peer are point to peer-link, while same macs in primary vPC peer are point to vPC leg PO.
Workaround: Clear mac
Severity: Moderate
Status: Assigned
Platform Seen: Nexus 9508
Resolved Releases:
Applicable Releases: 6.1(2)I3(4)

Appendix

Device Configuration


Archive.zip


write-up-config.zip