Lenovo

Configuring Lenovo Networking Switches in a Spine-Leaf Topology

Describes how to configure a spine-leaf topologies

Shows both Layer-2 and Layer-3 designs

Includes sample switch configuration texts

Includes configuration for routing protocols

Scott Lorditch

Andrey Naydenov

LENOVO PRESS

Abstract

This paper describes how to configure Lenovo® networking switches in a Spine-Leaf topology. It is a companion to the paper *Introduction to Spine-Leaf Networking Designs* which shows the rationale for using such an architecture and some of the metrics for different possible topologies.

https://lenovopress.com/lp0573-introduction-to-spine-leaf-networking-designs

The document includes sample configuration text for Layer 3 Spine-Leaf deployments (routed and switched) and Layer 2 Spine-Leaf deployments (switched only). Samples are provided for both spine and leaf switches.

The intended audience for this document is network engineers who are involved in or considering the deployment of Lenovo network switches in a Spine-Leaf topology.

At Lenovo Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at the Lenovo Press web site:

http://lenovopress.com

Do you have the latest version? We update our papers from time to time, so check whether you have the latest version of this document by clicking the **Check for Updates** button on the front page of the PDF. Pressing this button will take you to a web page that will tell you if you are reading the latest version of the document and give you a link to the latest if needed. While you're there, you can also sign up to get notified via email whenever we make an update.

Contents

Introduction

The traditional three-layer network topologies are losing momentum in the modern data center and are being supplanted by spine-leaf designs. This is even despite administrator familiarity with three-layer architecture and its benefits of scalability and ease of implementation.

This loss of momentum is because organizations are seeking to maximize the function and utilization of their data centers leading to architecture optimized for software defined and cloud solutions.

The spine-leaf architecture provides a strong base for the software defined data center optimizing the reliability and bandwidth available server communications.

New data centers are now being designed for cloud architectures with larger east-west traffic domains. This drives the need for a network architectures with an expanded flat east-west domain like spine-leaf as shown in Figure 1. Solutions like VMware NSX, OpenStack and others that distribute workloads to virtual machines running on many overlay networks running on top of a traditional underlay (physical) network require mobility across the flatter east-west domain.

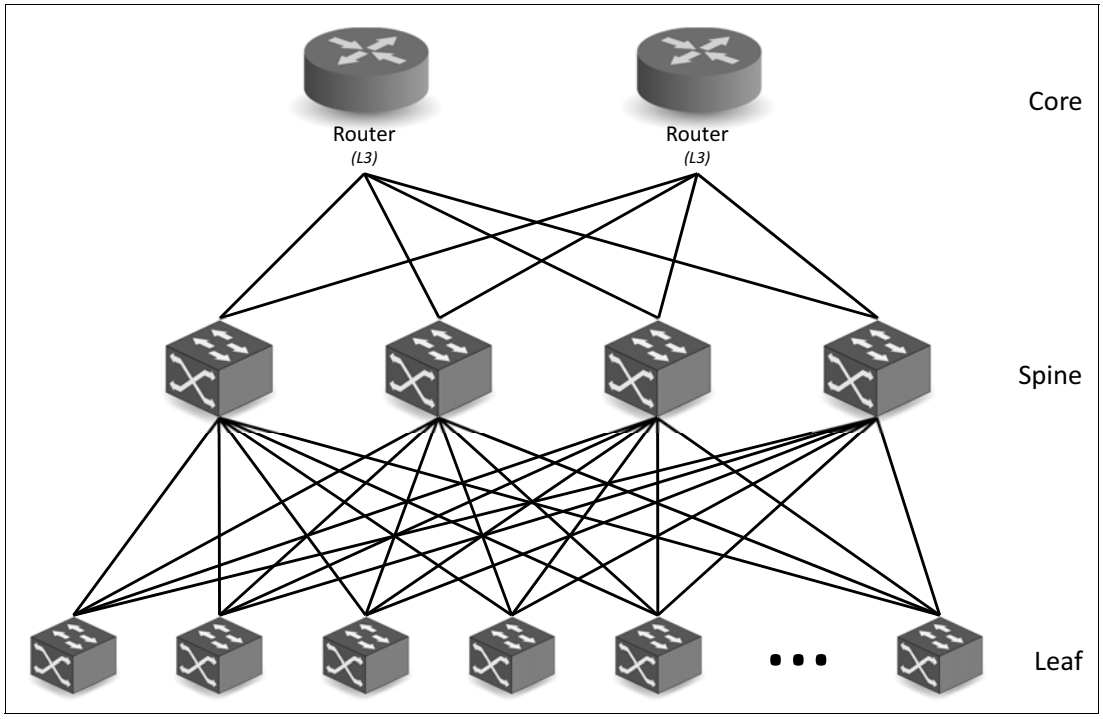

Figure 1 Spine-leaf Architecture

This paper describes the following:

- Layer 2 (switched) spine-leaf designs
- Layer 3 (routed) spine-leaf designs
- Configuration details for both designs

Layer 3 Spine-Leaf Design

This section provides a sample Layer 3 Spine-Leaf configuration, which includes the use of Border Gateway Protocol (BGP) and/or Open Shortest Path First (OSPF). Topics:

The Layer 3 Spine-Leaf topology is shown in Figure 2.

Lenovo ThinkSystem™ NE2572 switches are used at the leaf layer and Lenovo ThinkSystem NE10032 switches are used at the spine layer. Additional racks can be added to this design as long as there are available ports on the spine switches, and additional host servers beyond the one shown in the figure would be provisioned in each rack up to the limit of available ports on the leaf switches.

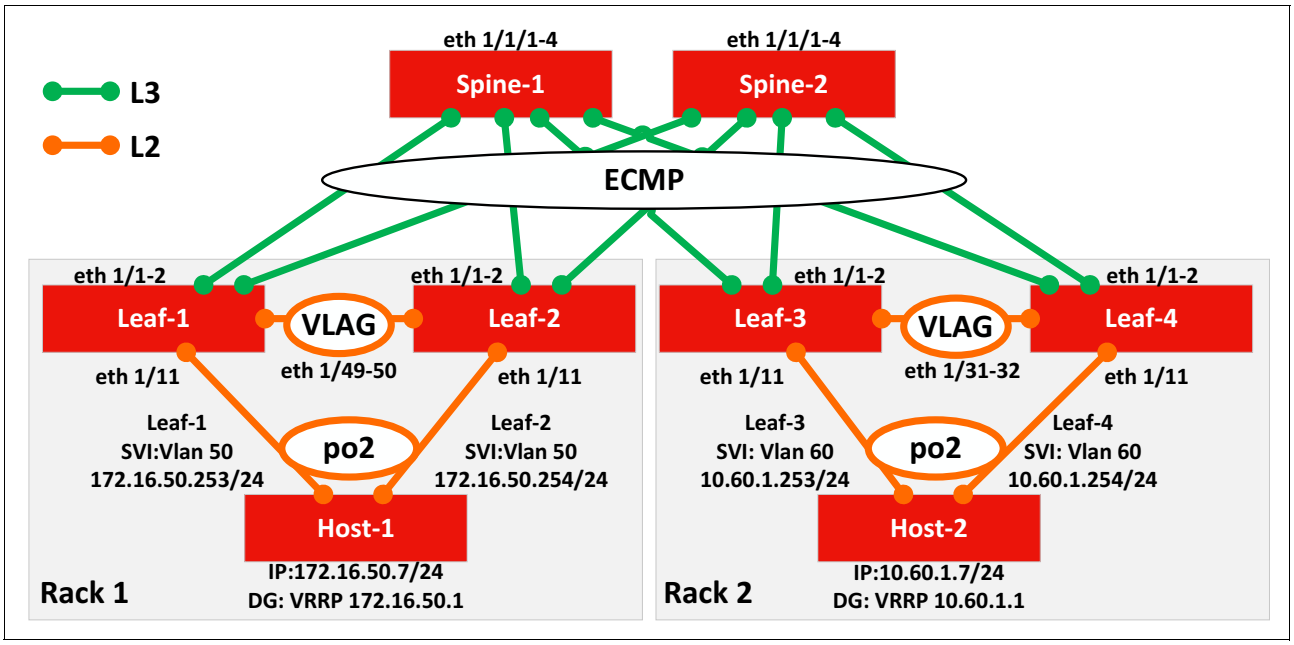

Figure 2 Layer 3 leaf-spine topology with Lenovo leaf and spine switches

In this topology, there is one broadcast domain in each rack.

In Rack 1, VLAN 50 is used and the 172.16.50.0/24 sub-network. With VRRP enabled between NE2572-Leaf1 and NE2572-Leaf2 in VLAN 50, Server 1 may utilize the VRRP IP as its default gateway. Traffic is load balanced across both leaves.

Rack 2 is configured identically, except VLAN 60 with the 10.60.1.0/24 sub-network. Server 2 may specify the VRRP IP in VLAN 60 as its default gateway.

Prerequisites

The configuration has the following prerequisites.

- BGP Configuration

  This example uses iBGP where a single Autonomous System Number (ASN) is assigned to every switch in the fabric.

  IP addressing is needed in various places in a Layer 3 design.

- Loopback Interfaces

  To identify the Layer 3 device within an AS (Autonomous System), a unique 32-bit router ID (RID) is utilized and loopback addresses may be used as router IDs when configuring routing protocols. As with ASNs (Autonomous System Numbers), loopback addresses should follow a logical pattern that will make it easier for engineers to manage the network and allow for growth.

  If multiple loopback or IP interfaces exist on a system, the interface with the highest numbered IP address is used as the router ID. To define manually specific router-ID use router-ID command under BGP or OSPF configuration menu.

  Figure 3 shows the loopback addresses used as well as router IDs in the BGP and OSPF examples in this document.

- Point-to-point addresses

  Point-to-point addresses are needed to enable BGP and also production traffic to find the appropriate route(s) between switches. These addresses are needed on the Layer 3 links between leaf and spine switches and also between spine switches if such links are in use.

  The point-to-point IP addresses used in this example are shown in Figure 3.

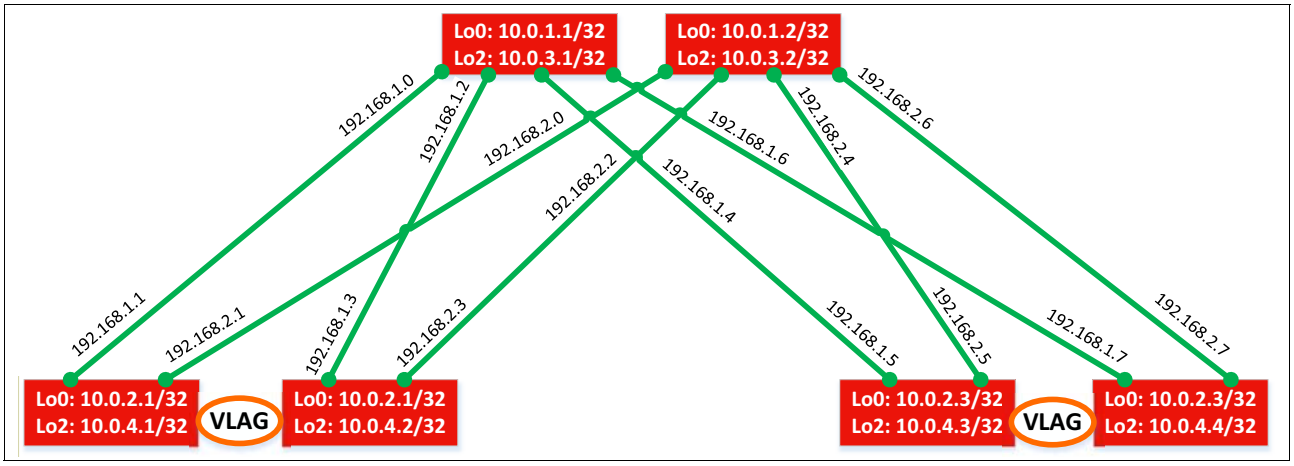

Figure 3 Sample topology with Layer 3 routing and point-to-point IP addresses

Steps to configure the Leaf switch

This section describes the steps that are necessary to deploy a leaf switch. These steps would be followed for each leaf switch, with each switch having its own unique parameters including hostname and IP addresses.

Basic configuration commands

These commands are necessary to make the switch identifiable and manageable over the network.

- Set the hostname. A suitable naming convention for these names in your organization should be available.
- Configure the out-of-band management interface and default gateway. This is typically `interface mgmt 0`.
- LLDP and SSH are required; they are enabled by default.

Configure spine-facing VLAG between each pair of leaf switches

Configuration of this feature is needed to enable redundant connections from the leaf switches to the spine switches as well as to the servers that each pair of leafs will support.

The specific interfaces mentioned are those in the sample topology and are their use is not required in a customer environment.

1. Assign interfaces ethernet 1/49-50 for use as the ISL for VLAG interconnect.

   The **`interface Ethernet 1/49-50`** command and the accompanying port-channel command accomplish this step.

2. Configure VLAG using a unique tier-id for each pair of switches. In addition, the port-channel created in Step 1 needs to be explicitly configured as the ISL (inter-switch link), and the “health check” that the switches use to maintain awareness of each other’s status must also be configured.

   The configuration commands needed here are:

   ```
   vlag tier-id <id>
   vlag ISL port-channel <number>
   vlag hlthchk peer-ip <partner's IP address> [vrf management (optional)]
   ```

3. Configure each downstream, server-facing interface with its own LACP port channel, even if only one physical port is used to connect to the server.

   The ports will have the **`channel group <x> mode active`** command in their configuration.

   A VLAG instance for the server is created using the `vlag instance <n> port-channel <x>` command. The port channel number must be the same for ports on the two peer switches that connect to the same server.

   In the sample configuration in Figure 4 on page 7, port channel 2 connects downstream to server 1 and is also configured as a PVRST edge port.

IP Addressing, VLANs, and VRRP

In our sample topology, each rack uses a unique VLAN and IP address range for the servers in that rack. In addition, IP addresses are required for the point-to-point connections between the leaf switches and the spine switches, and a loopback address is required for use by BGP and/or OSPF routing protocols.

VRRP is used to provide redundancy between the two switches in a rack to ensure reachability from elsewhere in the network. It will be configured to provide a *next hop* address for routes terminating in the rack which will be available even if one switch in the rack suffers an outage.

The following steps are required to configure VRRP so that each switch can successfully process traffic in the event its partner suffers a failure:

1. Create the assigned VLAN for the current rack. In our sample environment, this is VLAN 50 for rack 1, and VLAN 60 for rack 2. (The sample configuration text is for a rack 1 leaf switch). Use the following command:

   ```
   vlan 50
   ```

2. Assign all server-facing ports and port-channels to this VLAN. The interfaces' configuration will include the following commands:

   ```
   switchport trunk allowed vlan
   switchport trunk native vlan
   ```

3. Assign an IP address for this VLAN to `interface vlan <x>`. This address is unique to each switch.

4. Configure a VRRP address on the VLAN interface. The commands are:

   ```
   vrrp <instance number>
   ip address xxx.xxx.xxx.xxx
   ```

   No subnet mask is required since this address will inherit the subnet mask used in the interface being configured.

   The VRRP instance must be the same on the two peer switches serving a rack, and the shared address must be in the same subnet as the IP address assigned to the VLAN interface.

5. Configure IP addresses for each point-to-point link between a leaf switch and a spine switch.

   Additional IP addresses are needed on the point-to-point links between the leaf switches and the spines. These can be provisioned from small (/30) subnets. It is best if an addressing scheme is developed for these links so that their addresses are intuitively understandable; this will facilitate troubleshooting.

6. Configure and assign IP address for several loopback interfaces with following purposes:
   - As the router ID required when using BGP or OSPF.
   - As the Tunnel IP address for vxlan dynamic tunnels in high-availability mode. Such loopback interfaces must be identical on both vLAG peers.

To facilitate the use of BGP and/or OSPF, a route map and IP prefix list can be used to ensure that all of the addresses discussed above are known to the routing protocol(s).

The commands are in the in the sample configuration shown in Figure 4.

Figure 4 Leaf Switch Sample Base Configuration

```
enable
configure
!
line vty 0 0
exec-timeout 0 0
!
hostname NE2572-Leaf1
!
interface mgmt 0
ip address 192.168.90.30/24
!
vrf context management
ip route 0.0.0.0/0 192.168.90.1
```

```
!
interface Ethernet 1/49-50
description ISL
channel-group 1 mode active
!
interface port-channel 1
description ISL
switchport mode trunk
switchport trunk allowed vlan 50
shutdown
lacp suspend-individual
no shutdown
!
vlag enable
vlag tier-id 1
vlag isl port-channel 1
vlag hlthchk peer-ip 192.168.90.31 vrf management
!
interface ethernet 1/11
description Server 1
channel-group 2 mode active
!
interface port-channel 2
description Server 1
switchport mode trunk
switchport trunk allowed vlan 50
switchport trunk native vlan 50
spanning-tree port type edge
!
vlag instance 2 port-channel 2 enable
!
Vlan 50
!
interface Vlan 50
ip address 172.16.50.253/24
vrrp 1
address 172.16.50.1
!
interface ethernet 1/1
no switchport
description Spine-1
ip address 192.168.1.1/31
!
interface ethernet 1/2
no switchport
description Spine-2
ip address 192.168.2.1/31
!
interface loopback 0
description DCI_VTEP
ip address 10.0.2.1/32
!
interface loopback 2
description MP-BGP_src
ip address 10.0.4.1/32
```

```
!
!
route-map spine-leaf permit 10
match ip address prefix-list spine-leaf
!
ip prefix-list  spine-leaf description Redistribute loopback and leaf networks
ip prefix-list  spine-leaf seq 10 permit 10.0.0.0/8 ge 24
ip prefix-list  spine-leaf seq 20 permit 172.16.0.0/16 ge 24
```

Leaf BGP configuration

The section discusses the necessary steps to configure BGP. For the Layer 3 topology, either BGP or OSPF is required, but it is unlikely that both would be needed. The sample switch configuration commands for BGP is listed in Figure 5.

BGP will use the loopback IP address as its router ID if one is available; this address was set in step 6 on page 7 and is strongly recommended.

The BGP configuration in our sample environment uses iBGP, which means that single Autonomous System Numbers (ASNs) is used.

The steps to configure BGP are as follows:

1. Enable BGP with the `router bgp <ASN>` command. The ASNs should be private ASNs, and each switch should be assigned its own ASN number.
2. Use the `bgp bestpath as-path multipath-relax` command to enable ECMP (Equal Cost Multiple Path). This allows multiple paths from switch to switch to be used concurrently.
3. Use `maximum-paths ibgp 32` to specify the maximum number of parallel paths to a destination to add to the routing table. This number should be equal to or greater than the number of spines, up to 32.
4. Each BGP neighbor must be configured. Every spine switch will typically be a neighbor to a leaf switch; they will be identified by their configured loopback IP address. The `address family ipv4 unicast` command is required in the configuration of each neighbor; no diagnostic message will be issued if it is omitted, but BGP routing will not function. The `bfd` and `advertisement interval` commands are recommended to speed convergence of BGP in the event a link or switch suffers an outage.

The Leaf BGP configuration is shown in Figure 5.

Figure 5 BGP Configuration for Leaf Switch

```
router bgp 400
router-id 10.0.2.1
bestpath as-path multipath-relax
address-family ipv4 unicast
redistribute direct route-map spine-leaf
maximum-paths ibgp 32
!
neighbor 192.168.1.0 remote-as 400
address-family ipv4 unicast
bfd
advertisement-interval 1
!
neighbor 192.168.2.0 remote-as 400
address-family ipv4 unicast
bfd
advertisement-interval 1
!
Interface Ethernet 1/1-2
bfd interval 100 minrx 100 multiplier 3
!
```

Leaf OSPF configuration

This section outlines the steps needed to configure OSPF as the dynamic routing protocol in a Layer 3 Spine Leaf topology. The sample configuration commands are shown in Figure 6

OSPF uses the loopback IP address as its router ID if one is available; this is strongly recommended.

The steps to configure OSPF are as follows:

1. Enable OSPF with the `router ospf process-id` command. (At present, the only valid process-id is 0; this will change as features are added to CNOS firmware.)
2. Add the connected networks to OSPF area 0. This is accomplished for each interface which has an IP address with the `ip router ospf area 0` command; the point-to-point interfaces and the VLAN interfaces should have this in their configuration.
3. Enable ECMP with the `maximum-paths 32` command and specify the maximum number of parallel paths to a destination to add to the routing table. This number should be equal to or greater than the number of spines, up to 32.
4. Use BFD (bi-directional failure detection) to speed convergence in the event of a failure. Its settings are configured to 100 millisecond send/receive intervals. The multiplier is the number of packets that must be missed to declare a session down.
5. Use the `redistribute direct` command to ensure that the IP addresses for all point-to-point links are known to and propagated by OSPF.

The sample commands are listed in Figure 6.

Figure 6 OSPF configuration for leaf switch (L3)

```
maximum-paths 32
router ospf
router-id 10.0.2.1
```

```
bfd
log-adjacency-changes
redistribute direct route-map spine-leaf
!
interface ethernet 1/1
ip router ospf 0 area 0
!
interface ethernet 1/2
ip router ospf 0 area 0
!
Interface Ethernet 1/1-2
bfd interval 100 minrx 100 multiplier 3
```

Steps to configure the Spine switch

This section shows the configuration steps for a spine switch in a L3 topology. As was the case for the leaf switches, each spine switch is configured mostly identically with certain parameters such as hostname and IP addresses differing from one spine to the next.

Basic configuration commands

The commands outlined in this section make the switch reachable and manageable over the network. They also include commands to configure the point-to-point links to the leaf switches.

1. Set the hostname, OOB management interface and default gateway.
2. Configure the four point-to-point interfaces connected to leaf switches.

   These interfaces all must have IP addresses on the same subnet (even if /30) as the corresponding interface on a leaf switch. This is done with the `ip address` command for each interface.

   In our sample topology, there are four leafs but there can be more in a similar design if needed.
3. Configure and assign IP addresses for several loopback interfaces with following purposes:
   - As the router ID required when using BGP or OSPF.
   - As the Multi-Protocol BGP source IP address for VXLAN dynamic tunnels in high-availability mode.
4. If desired, you can break out any 100Gb ports as 4 x 25Gb using the `hardware profile portmode` command (this is not shown in Figure 7)

   If a breakout is configured, it changes interface naming – the four ports using physical port Ethernet 1/1 would be Ethernet 1/1/1, 1/1/2, and so on.

   However, if the entire 10Gb ports are used without breaking them out into multiple interfaces, then the leaf-facing interfaces would be ports such as Ethernet 1/1, 1/2, 1/3 and so on.
5. Configure a route map and IP prefix-list to redistribute all loopback addresses and leaf networks via BGP or OSPF.

   The command `seq 10 permit 10.0.0.0/8 ge 24` includes all addresses in the 10.0.0.0/8 address range with a mask greater than or equal to 24. This includes all loopback addresses used as router IDs in our sample.

The command `seq 20 permit 172.16.0.0/16 ge 24` includes the 172.16.50.0/24 network used on Leafs 1 and 2 as shown in the configuration.

Figure 7 shows the sample spine switch configuration.

Figure 7 Spine switch configuration for leaf ports – with IP addresses

```
Enable
configure
!
line vty 0 0
exec-timeout 0 0
!
hostname NE10032-Spine1
!
interface mgmt 0
ip address 192.168.90.28/24
!
vrf context management
ip route 0.0.0.0/0 192.168.90.1
!
!
interface ethernet 1/1/1
! (or just Ethernet 1/1...)
!
no switchport
description Leaf 1
ip address 192.168.1.0/31
!
interface ethernet 1/1/2
no switchport
description Leaf 2
ip address 192.168.1.2/31
!
interface ethernet 1/1/3
no switchport
description Leaf 3
ip address 192.168.1.4/31
!
interface ethernet 1/1/4
no switchport
description Leaf 4
ip address 192.168.1.6/31
!
interface loopback 0
description Router ID
ip address 10.0.1.1/32
!!
route-map spine-leaf permit 10
match ip address prefix-list spine-leaf
!
ip prefix-list  spine-leaf description Redistribute loopback and leaf networks
ip prefix-list  spine-leaf seq 10 permit 10.0.0.0/8 ge 24
ip prefix-list  spine-leaf seq 20 permit 172.16.0.0/16 ge 24
!
```

Spine Switch BGP configuration

The spine configuration for BGP mirrors that used on the leaf switches, and is shown in Figure 8. Each spine switch, like each leaf switch, will have same autonomous system number (ASN).

The steps to configure BGP on a spine switch in our sample topology are as follows:

1. Enable BGP with the `router bgp ASN` command.
2. Use the command `bgp bestpath as-path multipath-relax` to enable ECMP, and so allow multiple parallel paths to a destination to be used concurrently
3. Use the command `maximum-paths ibgp 32` to specify the maximum number of parallel paths to a destination to add to the routing table. In this topology, there are two equal cost best paths from a spine to a host, one to each leaf that the host is connected.
4. Configure each BGP neighbor. Every spine switch will typically be a neighbor to a leaf switch; they will be identified by their configured loopback IP address or router-id ID.

   The `address family ipv4 unicast` command is required in the configuration of each neighbor; no diagnostic message will be issued if it is omitted, but BGP routing will not

function. The `bfd` and `advertisement interval` commands are recommended to speed convergence of BGP in the event a link or switch suffers an outage.

Figure 8 Spine BGP configuration

```
!
enable
configure
!
router bgp 400
router-id 10.0.1.1
bestpath as-path multipath-relax
address-family ipv4 unicast
redistribute direct
maximum-paths ibgp 32
!
neighbor 192.168.1.1 remote-as 400
address-family ipv4 unicast
route-reflector-client
bfd
advertisement-interval 1
!
neighbor 192.168.1.3 remote-as 400
address-family ipv4 unicast
route-reflector-client
bfd
advertisement-interval 1
!
neighbor 192.168.1.5 remote-as 400
address-family ipv4 unicast
route-reflector-client
bfd
advertisement-interval 1
!
neighbor 192.168.1.7 remote-as 400
address-family ipv4 unicast
route-reflector-client
bfd
advertisement-interval 1
!
Interface Ethernet 1/1/1-4
bfd interval 100 minrx 100 multiplier 3
!
interface loopback 0
description Router ID
ip address 10.0.1.1/32
!
interface loopback 2
description Spine1_MP_BGP_source
ip address 10.0.3.1/32
!
```

Spine Switch OSPF configuration

The spine switch OSPF configuration mirrors that of the leaf switches and is listed in Figure 9 on page 15.

The steps to configure OSF on a spine switch in our sample topology are as follows:

1. Enable OSPF with the `router ospf process-id` command (at present, the process-id most be 0, however, this will change in forthcoming releases of the firmware).
2. Add the connected networks to OSPF area 0. This is accomplished for each interface which has an IP address with the `ip router ospf area 0` command; the point-to-point interfaces and the VLAN interfaces should have this in their configuration.
3. Enable ECMP with the command `maximum-paths 32` command and specify the maximum number of parallel paths to a destination to add to the routing table. In the sample topology, there are two equal cost paths from a spine switch to a host server.
4. Use BFD (bi-directional failure detection) to speed convergence in the event of a failure. Its settings are configured to 100 millisecond send/receive intervals. The multiplier is the number of packets that must be missed to declare a session down.
5. The `redistribute direct` command ensures that the IP addresses for all point-to-point links are known to and propagated by OSPF.

A sample configuration of OSPF for a spine switch is shown in Figure 9.

Figure 9 OSPF configuration for spine switch (L3)

```
router ospf
router-id 10.0.1.1
bfd
log-adjacency-changes
redistribute direct route-map spine-leaf
!
interface ethernet 1/1/1
! (or just Ethernet 1/1)
!
ip router ospf 0 area 0
!
interface ethernet 1/1/2
ip router ospf 0 area 0
!
interface ethernet 1/1/3
ip router ospf 0 area 0
!
interface ethernet 1/1/4
ip router ospf 0 area 0
!
maximum-paths 32
!
Interface Ethernet 1/1/1-4
bfd interval 100 minrx 100 multiplier 3
```

VXLAN configuration

This section presents a Layer 3 topology using VXLAN encapsulation. A design such as this enables virtualized Layer 2 environments far larger than can be implemented just using pure Layer 2 networking.

VXLAN is an extension to the VLAN protocol, designed to provide increased scalability in virtual networks. VXLAN is an Ethernet Layer 2 overlay protocol over a Layer 3 network. It uses an encapsulation method similar to VLAN that wraps MAC-based Ethernet Layer 2 frames with Layer 4 UDP packets, using destination UDP port 4789.

In typical physical networks, the number of VLANs is limited to 4094. VXLAN increases scalability up to 16 million logical networks and allows for Layer 2 adjacency across IP networks. This is achieved by adding a 24 bit segment ID to the VXLAN frame. The segment ID differentiates between individual logical networks, allowing millions of isolated Layer 2 VXLAN networks to coexist over the same Layer 3 infrastructure. Similar to VLANs, only virtual machines on the same VXLAN can exchange information with one another.

The virtualization of computing enables the mobility of virtual machines across different physical servers that exist in separate Layer 2 domains. This is done by tunneling virtual traffic over Layer 3 networks. Tunneling allows the dynamic distribution of resources within or across data centers without the limitations of Layer 2 boundaries or the necessity of creating large geographical Layer 2 domains. VXLAN uses Layer 2 over Layer 3 encapsulation. The Ethernet frame generated by a workload is wrapped within external VXLAN, UDP, IP, and Ethernet headers to ensure its transportation across the network infrastructure that connects the VXLAN endpoints together.

Multi-protocol Border Gateway Protocol (MP-BGP) Ethernet Virtual Private Network (EVPN) is a control plane protocol used to exchange information between VTEPs. It dynamically learns and updates VTEP, VNI, and MAC entries on the devices where the VTEPs are configured.

MP-BGP EVPN distributes Layer 2 reachability information for VXLAN overlay network end hosts. Each VTEP learns MAC information from its locally attached hosts and then distributes this information to remote VTEPs using MP-BGP EVPN. This decreases network flooding when learning about end hosts and offers a better control over the distribution of end host reachability information.

Leaf VXLAN configuration

The configuration examples shown in this section would be used in the leaf switches in a topology utilizing VXLAN. The necessary configuration for using VXLAN is included in each of them.

►

Figure 10 Configuration for NE2572-Leaf1

```
!
vlan 100
!
interface po2
switchport trunk allowed vlan add 100
!
interface po1
switchport trunk allowed vlan add 100
!
router bgp 400
```

```
address-family l2vpn evpn
redistribute host-info
!
neighbor 10.0.3.1 remote-as 400
update-source loopback 2
address-family l2vpn evpn
!
neighbor 10.0.3.2 remote-as 400
update-source loopback 2
address-family l2vpn evpn
!
nwv vxlan
vlan 100 virtual-network 100100
tunnel interface ip 10.0.2.1
nwv mode bgp-evpn ha
!
interface po2
vxlan enable
!
```

Figure 11 Configuration for NE2572-Leaf2

```
!
vlan 100
!
interface po2
switchport trunk allowed vlan add 100
!
interface po1
switchport trunk allowed vlan add 100
!
router bgp 400
address-family l2vpn evpn
redistribute host-info
!
neighbor 10.0.3.1 remote-as 400
update-source loopback 2
address-family l2vpn evpn
!
neighbor 10.0.3.2 remote-as 400
update-source loopback 2
address-family l2vpn evpn
!
```

Figure 12 Configuration for NE2572-Leaf3

```
!
vlan 100
!
interface po2
switchport trunk allowed vlan add 100
```

```
!
interface po1
switchport trunk allowed vlan add 100
!
router bgp 400
address-family l2vpn evpn
redistribute host-info
!
neighbor 10.0.3.1 remote-as 400
update-source loopback 2
address-family l2vpn evpn
!
neighbor 10.0.3.2 remote-as 400
update-source loopback 2
address-family l2vpn evpn
!
nwv vxlan
vlan 100 virtual-network 100100
tunnel interface ip 10.0.2.3
nwv mode bgp-evpn ha
!
interface po2
vxlan enable
!
```

Figure 13 Configuration for NE2572-Leaf4

```
!
vlan 100
!
interface po2
switchport trunk allowed vlan add 100
!
interface po1
switchport trunk allowed vlan add 100
!
router bgp 400
address-family l2vpn evpn
redistribute host-info
!
neighbor 10.0.3.1 remote-as 400
update-source loopback 2
address-family l2vpn evpn
!
neighbor 10.0.3.2 remote-as 400
update-source loopback 2
address-family l2vpn evpn
!
```

Spine VXLAN configuration

The configuration examples shown in this section would be used in the spine switches in a topology utilizing VXLAN. The necessary configuration for using VXLAN is included. Note that

because iBGP is used in this example, all of the neighbor configurations use the “route-reflector-client” specification, which avoids the need for a full mesh neighbor configuration.

Figure 14 Configuration for NE10032-Spine1

```
!
router bgp 400
address-family l2vpn evpn
!
neighbor 10.0.4.1 remote-as 400
update-source loopback 2
address-family l2vpn evpn
route-reflector-client
!
neighbor 10.0.4.2 remote-as 400
update-source loopback 2
address-family l2vpn evpn
route-reflector-client
!
neighbor 10.0.4.3 remote-as 400
update-source loopback 2
address-family l2vpn evpn
route-reflector-client
!
neighbor 10.0.4.4 remote-as 400
update-source loopback 2
address-family l2vpn evpn
route-reflector-client
!
```

Figure 15 Configuration for NE10032-Spine2

```
!
router bgp 400
address-family l2vpn evpn
!
neighbor 10.0.4.1 remote-as 400
update-source loopback 2
address-family l2vpn evpn
route-reflector-client
!
neighbor 10.0.4.2 remote-as 400
update-source loopback 2
address-family l2vpn evpn
route-reflector-client
!
neighbor 10.0.4.3 remote-as 400
update-source loopback 2
address-family l2vpn evpn
route-reflector-client
!
neighbor 10.0.4.4 remote-as 400
update-source loopback 2
```

```
address-family l2vpn evpn
route-reflector-client
!
```

Layer 3 example validation

In addition to sending traffic between hosts, the configuration shown in Figure 2 on page 4 can be validated with the commands shown in this section. For more information on commands and output, see the Application Guide for the exact switch model being used.

Download the Application Guide and other switch documentation from the Lenovo Information Center:

http://systemx.lenovofiles.com/help/index.jsp?topic=%2Fcom.lenovo.systemx.common.nav.doc%2Foverview_rack_switches.html&cp=0_4

Command and output examples are provided for one spine and one leaf. Command output on other switches is similar.

Commands and topics in this section:

show ip bgp summary

This command shows the status of all BGP connections. Each spine has four neighbors (the four leafs) and each leaf has two neighbors (the two spines).

Figure 16 Output of the show ip bgp summary commands

```
NE10032-Spine1#show ip bgp summary
BGP router identifier 10.0.1.1, local AS number 400
BGP table version is 4
1 BGP AS-PATH entries
0 BGP community entries

Neighbor            V     AS MsgRcv MsgSen TblVer InQ OutQ Up/Down State/PfxRcd
192.168.1.1         4    400    16     19      4    0    0 00:13:38       3
192.168.1.3         4    400    16     19      4    0    0 00:13:38       3
192.168.1.5         4    400    16     19      4    0    0 00:13:38       3
192.168.1.7         4    400    16     19      4    0    0 00:13:38       3

Total number of neighbors 4

Total number of Established sessions 4

NE2572-Leaf1#show ip bgp summary
BGP router identifier 10.0.2.1, local AS number 400
BGP table version is 45
1 BGP AS-PATH entries
0 BGP community entries
1  Configured ebgp ECMP multipath: Currently set at 1
2  Configured ibgp ECMP multipath: Currently set at 2

Neighbor            V     AS MsgRcv MsgSen TblVer InQ OutQ Up/Down State/PfxRcd
192.168.1.0         4    400    20     17     45    0    0 00:14:14      11
192.168.2.0         4    400    17     14     45    0    0 00:11:02      11

Total number of neighbors 2

Total number of Established sessions 2
```

show ip ospf neighbors

This command shows the state of all connected OSPF neighbors. In this configuration, each spine has four neighbors (the four leafs) and each leaf has two neighbors (the two spines).

Figure 17 Output of the show ip ospf neighbors commands

```
NE10032-Spine1#show ip ospf neighbor

Total number of full neighbors: 4

OSPF process 0 VRF(default):
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.0.4.1          1   Full/Backup     00:00:33    192.168.1.1     Ethernet1/1/1
10.0.4.2          1   Full/Backup     00:00:34    192.168.1.3     Ethernet1/1/2
10.0.4.3          1   Full/Backup     00:00:40    192.168.1.5     Ethernet1/1/3
10.0.4.4          1   Full/Backup     00:00:39    192.168.1.7     Ethernet1/1/4
NE2572-Leaf1#show ip ospf neighbor

Total number of full neighbors: 2

OSPF process 0 VRF(default):
Neighbor ID     Pri   State           Dead Time   Address         Interface
10.0.3.1          1   Full/DR        00:00:34    192.168.1.0     Ethernet1/1
10.0.3.2           1   Full/DR         00:00:31    192.168.2.0     Ethernet1/2
```

show ip route bgp

This command verifies the BGP entries in the Routing Information Base (RIB). Entries with multiple paths shown are used with ECMP. The two server networks in this example,

10.60.1.0 and 172.16.50.0, each have two paths from NE10032-Spine1, one through each leaf. The first set of routes with a subnet mask of /32 are the IPs configured for router IDs.

Figure 18 Output of the show ip route bgp commands

```
NE10032-Spine1#show ip route bgp
IP Route Table for VRF "default"
B       10.0.2.1/32 [200/0] via 192.168.1.1, Ethernet1/1/1, 00:15:16
B       10.0.2.3/32 [200/0] via 192.168.1.5, Ethernet1/1/3, 00:15:16
B       10.0.4.1/32 [200/0] via 192.168.1.1, Ethernet1/1/1, 00:15:16
B       10.0.4.2/32 [200/0] via 192.168.1.3, Ethernet1/1/2, 00:15:16
B       10.0.4.3/32 [200/0] via 192.168.1.5, Ethernet1/1/3, 00:15:16
B       10.0.4.4/32 [200/0] via 192.168.1.7, Ethernet1/1/4, 00:15:16
B       10.60.1.0/24 [200/0] via 192.168.1.5, Ethernet1/1/3, 00:15:16
B       172.16.50.0/24 [200/0] via 192.168.1.1, Ethernet1/1/1, 00:15:16

Gateway of last resort is not set

NE2572-Leaf1 has two paths to all other leafs and two paths to Server 2's network,
10.60.1.0. There is one path through each spine.

NE2572-Leaf1#show ip route bgp
IP Route Table for VRF "default"
B       10.0.1.1/32 [200/0] via 192.168.1.0, Ethernet1/1, 00:00:10
B       10.0.1.2/32 [200/0] via 192.168.2.0, Ethernet1/2, 00:00:10
B       10.0.2.3/32 [200/0] via 192.168.1.5 (recursive via 192.168.1.0 ), 00:12:32
                    [200/0] via 192.168.2.5 (recursive via 192.168.2.0 ), 00:12:32
B       10.0.3.1/32 [200/0] via 192.168.1.0, Ethernet1/1, 00:00:10
B       10.0.3.2/32 [200/0] via 192.168.2.0, Ethernet1/2, 00:00:10
B       10.0.4.2/32 [200/0] via 192.168.1.3 (recursive via 192.168.1.0 ), 00:12:31
                    [200/0] via 192.168.2.3 (recursive via 192.168.2.0 ), 00:12:31
B       10.0.4.3/32 [200/0] via 192.168.1.5 (recursive via 192.168.1.0 ), 00:12:32
                    [200/0] via 192.168.2.5 (recursive via 192.168.2.0 ), 00:12:32
B       10.0.4.4/32 [200/0] via 192.168.1.7 (recursive via 192.168.1.0 ), 00:12:32
                    [200/0] via 192.168.2.7 (recursive via 192.168.2.0 ), 00:12:32
B       10.60.1.0/24 [200/0] via 192.168.1.5 (recursive via 192.168.1.0 ), 00:12:32
                     [200/0] via 192.168.2.5 (recursive via 192.168.2.0 ), 00:12:32
B       192.168.1.2/31 [200/0] via 192.168.1.0, Ethernet1/1, 00:00:10
B       192.168.1.4/31 [200/0] via 192.168.1.0, Ethernet1/1, 00:00:10
B       192.168.1.6/31 [200/0] via 192.168.1.0, Ethernet1/1, 00:00:10
B       192.168.2.2/31 [200/0] via 192.168.2.0, Ethernet1/2, 00:00:10
B       192.168.2.4/31 [200/0] via 192.168.2.0, Ethernet1/2, 00:00:10
B       192.168.2.6/31 [200/0] via 192.168.2.0, Ethernet1/2, 00:00:10

Gateway of last resort is not set
```

show ip route ospf

This command verifies the OSPF entries in the Routing Information Base (RIB). Entries with multiple paths shown are used with ECMP. The two server networks in this example, 10.60.1.0 and 172.16.50.0, each have two paths from NE10032-Spine1, one through each leaf.

The first set of routes with a subnet mask of /32 are the IPs configured for router IDs.

Figure 19 Output of the show ip route ospf commands

```
NE10032-Spine1#show ip route ospf
IP Route Table for VRF "default"
O E2    10.0.1.2/32 [110/20] via 192.168.1.3, Ethernet1/1/2, 00:23:08
                    [110/20] via 192.168.1.5, Ethernet1/1/3, 00:23:08
                    [110/20] via 192.168.1.7, Ethernet1/1/4, 00:23:08
                    [110/20] via 192.168.1.1, Ethernet1/1/1, 00:23:08
O E2    10.0.2.1/32 [110/20] via 192.168.1.3, Ethernet1/1/2, 00:23:28
                    [110/20] via 192.168.1.1, Ethernet1/1/1, 00:23:28
O E2    10.0.2.3/32 [110/20] via 192.168.1.7, Ethernet1/1/4, 00:04:55
                    [110/20] via 192.168.1.5, Ethernet1/1/3, 00:04:55
O E2    10.0.3.2/32 [110/20] via 192.168.1.3, Ethernet1/1/2, 00:23:08
                    [110/20] via 192.168.1.5, Ethernet1/1/3, 00:23:08
                    [110/20] via 192.168.1.7, Ethernet1/1/4, 00:23:08
                    [110/20] via 192.168.1.1, Ethernet1/1/1, 00:23:08
O E2    10.0.4.1/32 [110/20] via 192.168.1.1, Ethernet1/1/1, 00:23:28
O E2    10.0.4.2/32 [110/20] via 192.168.1.3, Ethernet1/1/2, 00:23:28
O E2    10.0.4.3/32 [110/20] via 192.168.1.5, Ethernet1/1/3, 00:23:28
O E2    10.0.4.4/32 [110/20] via 192.168.1.7, Ethernet1/1/4, 00:23:28
O E2    10.60.1.0/24 [110/20] via 192.168.1.7, Ethernet1/1/4, 00:23:28
                     [110/20] via 192.168.1.5, Ethernet1/1/3, 00:23:28
O E2    172.16.50.0/24 [110/20] via 192.168.1.3, Ethernet1/1/2, 00:23:28
                       [110/20] via 192.168.1.1, Ethernet1/1/1, 00:23:28
O       192.168.2.0/31 [110/2] via 192.168.1.1, Ethernet1/1/1, 00:23:29
O       192.168.2.2/31 [110/2] via 192.168.1.3, Ethernet1/1/2, 00:23:29
O       192.168.2.4/31 [110/2] via 192.168.1.5, Ethernet1/1/3, 00:23:29
O       192.168.2.6/31 [110/2] via 192.168.1.7, Ethernet1/1/4, 00:23:29

Gateway of last resort is not set

NE2572-Leaf1#show ip route ospf
IP Route Table for VRF "default"
O E2    10.0.1.1/32 [110/20] via 192.168.1.0, Ethernet1/1, 00:24:39
O E2    10.0.1.2/32 [110/20] via 192.168.2.0, Ethernet1/2, 00:24:19
O E2    10.0.2.3/32 [110/20] via 192.168.2.0, Ethernet1/2, 00:24:19
                    [110/20] via 192.168.1.0, Ethernet1/1, 00:24:19
O E2    10.0.3.1/32 [110/20] via 192.168.1.0, Ethernet1/1, 00:24:39
O E2    10.0.3.2/32 [110/20] via 192.168.2.0, Ethernet1/2, 00:24:19
O E2    10.0.4.2/32 [110/20] via 192.168.2.0, Ethernet1/2, 00:24:19
                    [110/20] via 192.168.1.0, Ethernet1/1, 00:24:19
O E2    10.0.4.3/32 [110/20] via 192.168.2.0, Ethernet1/2, 00:24:19
                    [110/20] via 192.168.1.0, Ethernet1/1, 00:24:19
O E2    10.0.4.4/32 [110/20] via 192.168.2.0, Ethernet1/2, 00:24:19
                    [110/20] via 192.168.1.0, Ethernet1/1, 00:24:19
O E2    10.60.1.0/24 [110/20] via 192.168.2.0, Ethernet1/2, 00:24:19
                     [110/20] via 192.168.1.0, Ethernet1/1, 00:24:19
O       192.168.1.2/31 [110/2] via 192.168.1.0, Ethernet1/1, 00:24:40
O       192.168.1.4/31 [110/2] via 192.168.1.0, Ethernet1/1, 00:24:40
O       192.168.1.6/31 [110/2] via 192.168.1.0, Ethernet1/1, 00:24:40
O       192.168.2.2/31 [110/2] via 192.168.2.0, Ethernet1/2, 00:24:20
O       192.168.2.4/31 [110/2] via 192.168.2.0, Ethernet1/2, 00:24:20
O       192.168.2.6/31 [110/2] via 192.168.2.0, Ethernet1/2, 00:24:20

Gateway of last resort is not set
```

show bfd neighbors

This command verifies BFD is properly configured and sessions are established as UP indicated in the State column.

Figure 20 Output of the show bfd neighbors commands

```
NE10032-Spine1#show bfd neighbors
Codes: LD/RD         - Local Discriminator/Remote Discriminator
       RH/RS         - Remote Heard/Remote State
OurAddr                 NeighAddr                       LD/RD         RH/RS      Holdown(mult)
State      Interface       VRF
192.168.1.0             192.168.1.1                      5/1          UP             300( 3)
UP         Ethernet1/1/1   default
192.168.1.2             192.168.1.3                      6/1          UP             300( 3)
UP         Ethernet1/1/2   default
192.168.1.4             192.168.1.5                      7/1          UP             300( 3)
UP         Ethernet1/1/3   default
192.168.1.6             192.168.1.7                      8/1          UP             300( 3)
UP         Ethernet1/1/4   default

NE2572-Leaf1#show bfd neighbors
Codes: LD/RD         - Local Discriminator/Remote Discriminator
       RH/RS         - Remote Heard/Remote State
OurAddr                 NeighAddr                       LD/RD         RH/RS      Holdown(mult)
State      Interface       VRF
192.168.1.1             192.168.1.0                      1/5          UP             300( 3)
UP         Ethernet1/1     default
192.168.2.1             192.168.2.0                      2/5          UP             300( 3)
UP         Ethernet1/2     default
```

show VLAG info

This command validates VLAG configuration status on leaf switches in this topology. The Inter-Switch link (ISL) Status, Heart Beat Status and VLAG Peer Status must all be up. The role for one switch in the VLAG pair is primary and its peer switch (not shown) is assigned the secondary role.

Figure 21 Output of the show VLAG info command

```
NE2572-Leaf1#show VLAG info
  Global State    : enabled
  VRRP active     : enabled
  vLAG system MAC : 08:17:f4:c3:dd:00
  ISL Information:
    PCH     Ifindex     State       Previous State
    -------+-----------+-----------+--------------------------------
    1       100001      Active      Inactive

  Mis-Match Information:
                   Local                       Peer
    -------------+---------------------------+----------------------
    Match Result : Match                       Match
    Tier ID      : 1                           1
    System Type  : NE2572                      NE2572
    OS Version   : 10.8.x.x                    10.8.x.x
```

```
  Role Information:
                  Local                       Peer
    ------------+---------------------------+----------------------
    Admin Role   : Primary                    Secondary
    Oper Role    : Primary                    Secondary
    Priority     : 0                          0
    System MAC   : a4:8c:db:bb:c8:01          a4:8c:db:bb:d9:01

  Consistency Checking Information:
    State             : enabled
    Strict Mode       : disabled
    Final Result      : pass

  FDB refresh Information:
  FDB is doing refresh with below setting:
    FDB refresh is configured
    Bridge FDB aging timer is 1800 second(s)

  FDB synchronization Information:
  FDB is being synchronized.

 Auto Recovery Interval 300s (Finished)

 Startup Delay Interval 120s (Finished)

 Health Check Information:
  Health check Peer IP Address: 192.168.90.31
  Health check Local IP Address: 192.168.90.30
  Health check retry interval: 30 seconds
  Health check number of keepalive attempts: 3
  Health check keepalive interval: 5 seconds
  Health check status: UP

  Peer Gateway State    : disabled
```

show VLAG instance all info

This command shows the status of all VLAG LAGs (Port channel 2 in this example).

Figure 22 Output of the show VLAG instance all info command

```
NE2572-Leaf1#show VLAG instance all info
  VLAG instance 2 : enabled
  Instance Information
    PCH        ifindex     State          Previous State    Cons Res
    ----------+-----------+--------------+-----------------+--------
    2          100002      Formed         Local UP          pass
```

show VLAG config-consistency

This command highlights configuration issues between VLAG peers. Mismatch examples include incompatible VLAG configuration settings, VLAN differences, different switch operating system versions and spanning-tree inconsistencies. There should be no output to this command on any switch configured for VLAG. If there is, resolve the mismatch.

Figure 23 Output of the show VLAG config-consistency command

```
NE2572-Leaf1#show vlag config-consistency

   "N/A": Unavailable value
 "-"  : Digest value, detail value dump by detail show command
 item                    Prio  result local                remote
 ------------------------------------------------------------------------------
 sys mac learn           high  pass   enable               enable
 global tag native       high  pass   disable              disable
 ISL port mode           high  pass   trunk                trunk
 ISL access vlan         high  pass   1                    1
 ISL native vlan         high  pass   1                    1
 ISL allowed vlan        high  pass   -                    -
 ISL tag native          high  pass   none                 none
 ISL dot1q tunnel        high  pass   disable              disable
 ISL egress tagged vlans high  pass   -                    -
 stp mode                high  pass   rapid-pvst           rapid-pvst
 stp path cost           high  pass   short                short
 mst region name         high  pass   -                    -
 mst region version      high  pass   0                    0
 mst inst mapping        high  pass   -                    -
 mst max-age             low   pass   20                   20
 mst max-hops            low   pass   20                   20
 mst hello time          low   pass   2                    2
 mst forward time        low   pass   15                   15
```

show vxlan mac-address

When a data packet is received on an access port, its source address MAC address is mapped to a specific VNID and ingress VTEP tunnel IP address. The command shows an output of both local and remote address learning per vxlan tunel. Note the tunnel ID or IP address of the tunnel. This is the vtep tunnel on which the remote MAC address is learned.

Figure 24 Output of the show vxlan mac-address command

```
NE2572-Leaf1(config)#sh nwv vxlan mac-address
Local MAC Count: 1
VNID       MAC                Interface              Vlan
--------   ----------------   --------------------   -----
100100     A4:8C:DB:D9:E6:01  po2                    100

Remote MAC Count: 1
VNID       MAC                Tunnel
--------   ----------------   --------------------
100100     A4:8C:DB:D9:E3:01  10.0.2.3
```

show nwv vxlan information

This command displays the status of all of the VXLAN tunnels and the ports configured to access them, as well as the ports where traffic which will be forwarded to VXLAN will originate.

Figure 25 Output of the show nwv vxlan information command

```
NE2572-Leaf1(config)#show nwv vxlan information
Codes:  A - Access vPort
        N - Network vPort, M - Multicast Network vPort

Virtual Networks Count: 1
Tunnels Count: 1
Access vPorts Count: 2
Network vPorts Count: 2
Multicast vPorts Count: 2

Virtual Ports:
Interface           Mode     vPorts Count
---------           ----     ------------
po1                 A        1
po2                 A        1
Ethernet1/1         N/M      1
Ethernet1/2         N/M      1
```

show nwv vxlan tunnel

This command shows the status of configured tunnel endpoints, including the local one on the current switch.

Figure 26 Output of the command

```
NE2572-Leaf1(config)#show nwv vxlan tunnel
Tunnel Count: 2

Tunnel IP Address     Tunnel Type    Status
-----------------     -----------    ------
10.0.2.1              Local          UP
10.0.2.3              Remote         UP
```

show running-config nwv vxlan

This command shows the portion of the current running configuration that is for the vxlan and network-virtualization (nwv) features.

Figure 27 Output of the command

```
NE2572-Leaf1(config)#show running-config nwv vxlan
nwv vxlan
 tunnel interface ip 10.0.2.1
 vlan 100 virtual-network 100100
!
interface po2
 vxlan enable
!
nwv mode bgp-evpn ha
```

Layer 2 Spine-Leaf Design

This section provides configuration information to build the Layer 2 leaf-spine topology shown in Figure 28. Note that this example differs from the previous one in the absence of the routing protocols (BGP and OSPF).

Topics in this section:

- "Layer 2 Leaf configuration"
- "Steps to configure the Leaf switch"
- "Layer 2 Spine configuration" on page 31
- "Steps to configure the Spine switch" on page 32
- "Layer 2 example validation" on page 33

The topology for this example is shown in Figure 28.

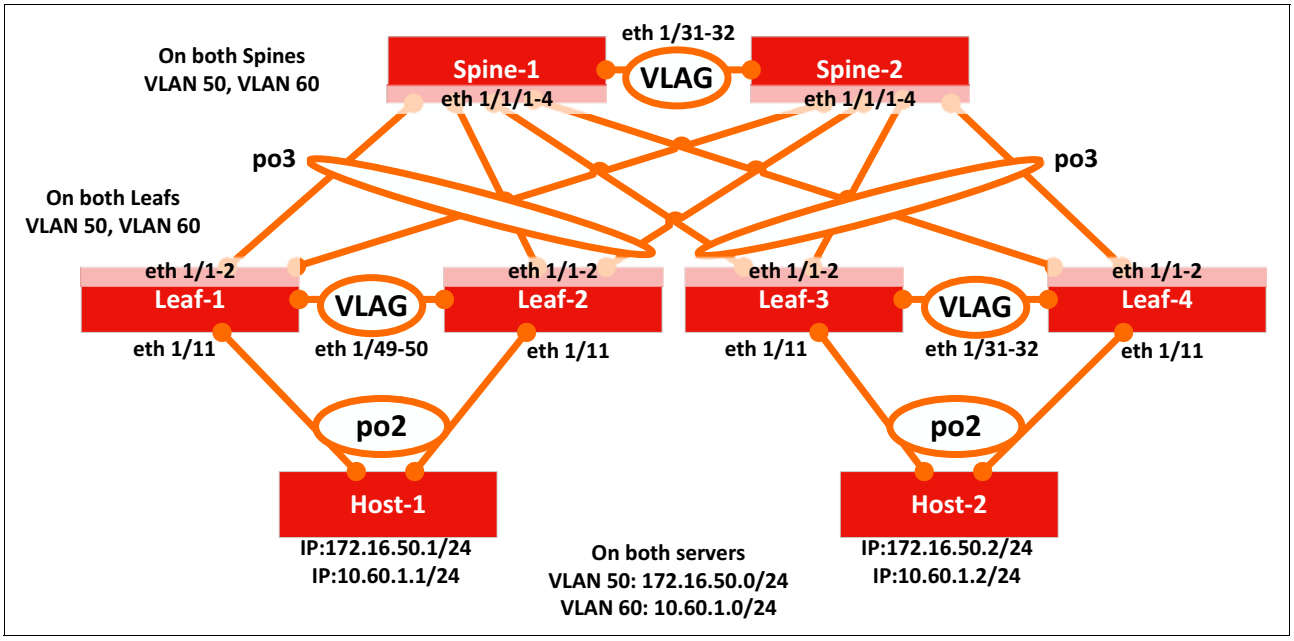

Figure 28 Layer 2 spine-leaf topology with Lenovo leaf and spine switches

Layer 2 Leaf configuration

A template leaf configuration is shown in Figure 29 on page 30. It is very similar to the leaf configuration in the L3 example, but the routing protocols are not needed and are not included.

Steps to configure the Leaf switch

Configure the Layer 2 Leaf switch as follows:

1. Set the switch hostname with the hostname command
2. Configure the OOB management interface and default gateway. This is typically done with the ip address command for `interface mgmt 0`, the management port.
3. LLDP and SSH are required and are enabled by default.

Configure VLAG to the servers and spine switches

Configuration of this feature is needed to enable redundant connections from the leaf switches to the spine switches as well as to the servers that each pair of leafs will support.

The specific interfaces mentioned are those in the sample topology and are their use is not required in a customer environment.

The following steps are performed to configure VLAG on the leaf switches for use in connecting to both the upstream spine switches and downstream servers:

1. Assign interfaces ethernet 1/49-50 for use as the ISL for VLAG interconnect.

   The `interface ethernet 1/49-50` command and the accompanying port-channel command accomplish this step.

2. Configure VLAG using a unique tier-id for each pair of switches. In addition, the port-channel created in Step 1 needs to be explicitly configured as the ISL (inter-switch link), and the “health check” that the switches use to maintain awareness of each other’s status must also be configured.

   The configuration commands needed here are:

   ```
   vlag tier-id <id>
   vlag ISL port-channel <number>
   vlag hlthchk peer-ip <partner's IP address> [vrf management (optional)]
   ```

3. Configure each downstream, server-facing interface with its own LACP port channel, even if only one physical port is used to connect to the server.

   The ports will have the `channel group <x> mode active` command in their configuration.

   A VLAG instance for the server is created using the `vlag instance <n> port-channel <x>` command. The port channel number must be the same for ports on the two peer switches that connect to the same server.

   In the sample configuration, port channel 2 connects downstream to server 1 and is also configured as a PVRST edge port.

4. Configure upstream L2 connections

   Two ports are used in a port-channel; one goes to each of two spine switches. The command to use is:

   ```
   interface ethernet 1/1-2; channel group 3 mode active
   ```

   This port channel is configured as a VLAG instance on each of the two peer leaf switches:

   ```
   vlag instance 3 port-channel 3
   ```

   This results on a four port mesh between two leaf switches and two spines.

   Because this configuration also uses VLAN 50 for one rack and VLAN 60 for the other, the upstream connections are configured to carry both VLANs.

   ```
   switchport trunk allowed vlan 50,60
   ```

The sample configuration is shown in Figure 29.

Figure 29 Leaf Layer 2 configuration

```
hostname NE2572-Leaf1
!
interface mgmt 0
ip address 192.168.90.30/24
!
vrf context management
```

```
ip route 0.0.0.0/0 192.168.90.1
!
Vlan 50,60
!
interface Ethernet 1/49-50
description ISL
channel-group 1 mode active
!
interface port-channel 1
description ISL
switchport mode trunk
switchport trunk allowed vlan 50,60
shutdown
lacp suspend-individual
no shutdown
!
vlag enable
vlag tier-id 1
vlag isl port-channel 1
vlag hlthchk peer-ip 192.168.90.31 vrf management
!
interface ethernet 1/11
description Server 1
channel-group 2 mode active
!
interface port-channel 2
description Server 1
switchport mode trunk
switchport trunk allowed vlan 50,60
spanning-tree port type edge
!
vlag instance 2 port-channel 2 enable
vlag instance 2 enable
!
interface ethernet 1/1-2
channel-group 3 mode active
!
interface port-channel 3
description Spines
switchport mode trunk
switchport trunk allowed vlan 50,60
!
vlag instance 3 port-channel 3
vlag instance 3 enable
```

Layer 2 Spine configuration

A template spine configuration is shown in Figure 30 on page 32. It is very similar to the spine configuration in the L3 example, but the routing protocols (BGP, OSPF) are not needed and are not included.

The connections between spine and leaf switches can use entire 100Gb physical ports (recommended) or one or more 25Gb ports broken out from the physical ports and connected

via a breakout cable. The ports are set in 100Gb more or 4x25Gb mode using the `portmode` command and a re-boot is required for a change to take effect.

Steps to configure the Spine switch

The steps needed to complete the configuration are outlined as follows:

1. Set the hostname using the command `hostname <name>`
2. Configure the OOB management interface and default gateway using these commands:

```
Interface mgmt 0
ip address xxx.xxx.xxx.xxx/xx
vrf context management
ip route 0.0.0.0/0 <gateway IP>
```

3. For spanning-tree, PVRST is enabled by default.

   Spine1 is configured as the primary PVRST root bridge using the `priority 0` command.

   The second spine switch should be configured as the secondary PVRST root bridge using the `priority 4096` command.

VLAG configuration

The steps to configure VLAG are as follows:

1. Configure the VLAG interconnect (ISL) between Spine1 and Spine2.

   In this example, add interfaces eth 1/31-32 to LACP port channel 1 for the VLAG interconnect

```
channel-group 1 mode active
```

2. Configure the basic VLAG parameters

```
vlag isl port-channel 1
vlag tier-id <ID>
vlag hlthchk peer-ip <management IP of peer switch>
vlag enable
```

   Note that the tier-id must be unique throughout the topology.

3. Configure leaf-facing ports

   Depending on whether full 100Gb physical ports are used or are broken out to 4x25Gb, the port names for these ports will vary. All port(s) going to a specific leaf switch will need to be in the same LACP port-channel. These port channels are configured for VLAG on both of the spine switches.

   Configuration commands are:

```
interface ethernet 1/<x-y>
   (or 1/<x>/<y-z> if port is used in breakout mode)
channel-group <x> mode active
vlag instance <z> port-channel <x> enable
```

The configuration is listed in Figure 30.

Figure 30 Layer 2 spine switch sample config

```
hostname NE10032-Spine1
!
interface mgmt 0
ip address 192.168.90.28/24
```

```
!
vrf context management
ip route 0.0.0.0/0 192.168.90.1
!
Vlan 50,60
!
spanning-tree vlan 50,60 priority 0
!
interface Ethernet1/31-32
channel-group 1 mode active
!
vlag tier-id 20
vlag isl port-channel 1
vlag hlthchk peer-ip 192.168.90.31 vrf management
vlag enable
!
interface Ethernet1/1/1-2 (or Ethernet 1/1-2)
channel-group 2 mode active
!
interface Ethernet1/1/3-4
channel-group 3 mode active
!
Interface port-channel 2
Switchport mode trunk
Switchport trunk allowed vlan 50,60
!
Interface port-channel 3
Switchport mode trunk
Switchport trunk allowed vlan 50,60
!
vlag instance 2 port-channel 2
vlag instance 2 enable
vlag instance 3 port-channel 3
vlag instance 3 enable
```

Layer 2 example validation

In addition to sending traffic between hosts, the configuration shown in Figure below can be validated with the commands shown in this section. For more information on commands and output, see the Application Guide for the exact switch

Download the Application Guide and other switch documentation from the Lenovo Information Center:

http://systemx.lenovofiles.com/help/index.jsp?topic=%2Fcom.lenovo.systemx.common.nav.doc%2Foverview_rack_switches.html&cp=0_4

Command and output examples are provided for one spine and one leaf. Command output on other switches is similar.

show VLAG info

The Inter-Switch link (ISL) Link Status, Heart Beat Status and VLAG Peer Status must all be up. The role for one switch in the VLAG pair is primary and its peer switch (not shown) is assigned the secondary role.

Figure 31 Output from the command

```
NE10032-Spine1(config)#show VLAG info

  Global State    : enabled
  VRRP active     : enabled
  vLAG system MAC : 08:17:f4:c3:dd:13
  ISL Information:
    PCH     Ifindex     State       Previous State
    -------+-----------+-----------+--------------------------------
    1       100001      Active      Inactive

  Mis-Match Information:
                   Local                       Peer
    -------------+---------------------------+-----------------------
    Match Result : Match                       Match
    Tier ID      : 20                          20
    System Type  : NE10032                     NE10032
    OS Version   : 10.8.x.x                    10.8.x.x

  Role Information:
                   Local                       Peer
    -------------+---------------------------+-----------------------
    Admin Role   : Primary                     Secondary
    Oper Role    : Primary                     Secondary
    Priority     : 0                           0
    System MAC   : a4:8c:db:bb:c8:01           a4:8c:db:bb:d9:01

  Consistency Checking Information:
    State             : enabled
    Strict Mode       : disabled
    Final Result      : pass

  FDB refresh Information:
  FDB is doing refresh with below setting:
    FDB refresh is configured
    Bridge FDB aging timer is 1800 second(s)

  FDB synchronization Information:
  FDB is being synchronized.

 Auto Recovery Interval 300s (Finished)

 Startup Delay Interval 120s (Finished)

 Health Check Information:
  Health check Peer IP Address: 192.168.90.29
  Health check Local IP Address: 192.168.90.28
  Health check retry interval: 30 seconds
  Health check number of keepalive attempts: 3
```

```
  Health check keepalive interval: 5 seconds
  Health check status: UP

  Peer Gateway State    : disabled

NE2572-Leaf1(config)#show VLAG info

  Global State    : enabled
  VRRP active     : enabled
  vLAG system MAC : 08:17:f4:c3:dd:00
  ISL Information:
    PCH     Ifindex     State       Previous State
    -------+-----------+-----------+--------------------------------
    1       100001      Active      Inactive

  Mis-Match Information:
                   Local                       Peer
    -------------+---------------------------+-----------------------
    Match Result : Match                       Match
    Tier ID      : 1                           1
    System Type  : NE2572                      NE2572
    OS Version   : 10.8.x.x                    10.8.x.x

  Role Information:
                   Local                       Peer
    -------------+---------------------------+-----------------------
    Admin Role   : Secondary                   Primary
    Oper Role    : Secondary                   Primary
    Priority     : 0                           0
    System MAC   : a4:8c:db:d9:dd:01           a4:8c:db:d9:c6:01

  Consistency Checking Information:
    State             : enabled
    Strict Mode       : disabled
    Final Result      : pass

  FDB refresh Information:
  FDB is doing refresh with below setting:
    FDB refresh is configured
    Bridge FDB aging timer is 1800 second(s)

  FDB synchronization Information:
  FDB is being synchronized.

 Auto Recovery Interval 300s (Finished)

 Startup Delay Interval 120s (Finished)

 Health Check Information:
  Health check Peer IP Address: 192.168.90.31
  Health check Local IP Address: 192.168.90.30
  Health check retry interval: 30 seconds
  Health check number of keepalive attempts: 3
  Health check keepalive interval: 5 seconds
  Health check status: UP
```

```
  Peer Gateway State    : disabled
```

show VLAG instance all info

This command shows the status and active VLANs of all VLAG LAGs (Port channels 2 and 3 in this example). The local and peer status must both be up.

Figure 32 Output from the command

```
NE10032-Spine1(config)#show VLAG instance all info

  VLAG instance 2 : enabled
  Instance Information
    PCH        ifindex     State          Previous State    Cons Res
    ----------+-----------+--------------+-----------------+--------
    2          100002      Formed         Local UP          pass

  VLAG instance 3 : enabled
  Instance Information
    PCH        ifindex     State          Previous State    Cons Res
    ----------+-----------+--------------+-----------------+--------
    3          100003      Formed         Local UP          pass
```

show VLAG config-consistency

This command highlights configuration issues between VLAG peers. Mismatch examples include incompatible VLAG configuration settings, VLAN differences, different switch operating system versions and spanning-tree inconsistencies. There should be no “fail” indications from this command on any switch configured for VLAG. If there is, resolve the mismatch.

Figure 33 Output from the command

```
NE10032-Spine1#show VLAG config-consistency

  "N/A": Unavailable value
  "-"  : Digest value, detail value dump by detail show command
  item                    Prio  result local                remote
  ------------------------------------------------------------------------------
  sys mac learn           high  pass   enable               enable
  global tag native       high  pass   disable              disable
  ISL port mode           high  pass   trunk                trunk
  ISL access vlan         high  pass   1                    1
  ISL native vlan         high  pass   1                    1
  ISL allowed vlan        high  pass   -                    -
  ISL tag native          high  pass   none                 none
  ISL dot1q tunnel        high  pass   disable              disable
  ISL egress tagged vlans high  pass   -                    -
  stp mode                high  pass   rapid-pvst           rapid-pvst
  stp path cost           high  pass   short                short
  mst region name         high  pass   -                    -
  mst region version      high  pass   0                    0
  mst inst mapping        high  pass   -                    -
  mst max-age             low   pass   20                   20
  mst max-hops            low   pass   20                   20
```

```
  mst hello time          low   pass   2                   2
  mst forward time        low   pass   15                  15


NE2572-Leaf1(config)#show VLAG config-consistency

  "N/A": Unavailable value
  "-"  : Digest value, detail value dump by detail show command
  item                    Prio  result local               remote
  ------------------------------------------------------------------------------
  sys mac learn           high  pass   enable              enable
  global tag native       high  pass   disable             disable
  ISL port mode           high  pass   trunk               trunk
  ISL access vlan         high  pass   1                   1
  ISL native vlan         high  pass   1                   1
  ISL allowed vlan        high  pass   -                   -
  ISL tag native          high  pass   none                none
  ISL dot1q tunnel        high  pass   disable             disable
  ISL egress tagged vlans high  pass   -                   -
  stp mode                high  pass   rapid-pvst          rapid-pvst
  stp path cost           high  pass   short               short
  mst region name         high  pass   -                   -
  mst region version      high  pass   0                   0
  mst inst mapping        high  pass   -                   -
  mst max-age             low   pass   20                  20
  mst max-hops            low   pass   20                  20
  mst hello time          low   pass   2                   2
  mst forward time        low   pass   15                  15
```

show spanning-tree vlan <XX> brief

This command validates spanning tree is enabled. All interfaces are forwarding (Sts column shows FWD) because VLAG is configured at the leaf and spine layers, eliminating the need for blocked ports. One of the spine switches (NE10032-Spine1 in this example) is the root bridge. Sever-facing interfaces on leaf switches (NE2572-Leaf1 interface eth 1/11 in this example) are edge ports.

Figure 34 Output from the command

```
NE10032-Spine1#show spanning-tree vlan 10 brief

VLAN0050
  spanning-tree enabled protocol rapid-pvst
  ROOT ID      priority    10
               address     a48c.dbbb.c800
               This bridge is the root
               Hello Time  2   Max age 20  Forward Delay 15

  BRIDGE ID    priority    10 (0 sys-id-ext 50)
               address     a48c.dbbb.c800
               Hello Time  2   Max age 20  Forward Delay 15

Interface        Role Sts cost      Prio.Nbr      Type
---------------- ---- --- --------- ------------- ----------------------
po1              Desg FWD 1          128.100001   point-to-point
po2              Desg FWD 1          128.100002   point-to-point
```

```
po3              Desg FWD 1          128.100003   point-to-point

NE2572-Leaf1#show spanning-tree vlan 50 brief

VLAN0050
  spanning-tree enabled protocol rapid-pvst
  ROOT ID      priority    10
               address     a48c.dbbb.c800
               Cost        1
               Port        100003 (po3)
               Hello Time  2   Max age 20  Forward Delay 15

  PREV ROOT    priority    32778
               address     a48c.dbd9.dd00
               Port        0

  BRIDGE ID    priority    32778 (32768 sys-id-ext 50)
               address     a48c.dbd9.dd00
               Hello Time  2   Max age 20  Forward Delay 15

Interface        Role Sts cost      Prio.Nbr      Type
---------------- ---- --- --------- ------------- ---------------------
po1              Root FWD 1          128.100001   point-to-point
po2              Desg FWD 1          128.100002   point-to-point
po3              Root FWD 1          128.100003   point-to-point
```

Validated hardware and operating systems

The following table includes the hardware and operating systems used to validate the examples in this paper.

Table 1 Switches and operating systems used in this guide

Switch	OS / Version
Lenovo Networking NE1032	CNOS 10.8
Lenovo Networking NE2572	CNOS 10.8
Lenovo Networking NE10032	CNOS 10.8

Technical support and resources

Lenovo Press is an online technical community where IT professionals have access to numerous resources for Lenovo software, hardware and services. The following product guides provide details product information about the switches:

- Lenovo ThinkSystem NE1032 RackSwitch

  https://lenovopress.com/lp0605-lenovo-thinksystem-ne1032-rackswitch
- Lenovo ThinkSystem NE2572 RackSwitch

  https://lenovopress.com/lp0608-lenovo-thinksystem-ne2572-rackswitch

- Lenovo ThinkSystem NE10032 RackSwitch

  https://lenovopress.com/lp0609-lenovo-thinksystem-ne10032-rackswitch

Authors

Scott Lorditch is a Consulting System Engineer for Lenovo. He performs network architecture assessments and develops designs and proposals for solutions that involve Lenovo Networking products. He also developed several training and lab sessions for technical and sales personnel. Scott joined IBM as part of the acquisition of Blade Network Technologies® and joined Lenovo as part of the System x® acquisition from IBM. Scott spent almost 20 years working on networking in various industries, as a senior network architect, a product manager for managed hosting services, and manager of electronic securities transfer projects. Scott holds a BS degree in Operations Research with a specialization in computer science from Cornell University.

Andrey Naydenov is a Consulting Systems Engineer for Lenovo and currently dedicated to Russia and the Eastern Europe and Central Asia (EECA) region. He has over 13 years of experience in the networking industry and has specialized in data center architecture, automation tools, compute platforms, and applications. Over the years he has held many certifications from IBM and Cisco for design, implementation and troubleshooting. Andrey is a frequent speaker at Lenovo conferences and conducts Networking Master Classses on the implementation of data center networking. Andrey holds a degree in Management and Informatics in Technical Systems from Moscow State Institute of Electronics and Mathematics.

Thanks to the following people for their contributions to this project:

- Bill Nelson
- David Watts

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document was created or updated on October 16, 2018.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp0930

Trademarks

Lenovo, the Lenovo logo, and For Those Who Do are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available on the Web at http://www.lenovo.com/legal/copytrade.html.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Blade Network Technologies®	Lenovo(logo)®	ThinkSystem™
Lenovo®	System x®	

The following terms are trademarks of other companies:

Other company, product, or service names may be trademarks or service marks of others.