Nokia Reference Design

Collapsed Spine EVPN/VXLAN Fabric

Authors: Aninda Chatterjee and Vivek V

3HE-21912-AAAA-TQZZA
Issue 1
July 2025

Nokia is committed to diversity and inclusion. We are continuously reviewing our customer documentation and consulting with standards bodies to ensure that terminology is inclusive and aligned with the industry. Our future customer documentation will be updated accordingly.

Contents

List of Tables

List of Figures

Preface for reference designs

The main distinction between the **Nokia Validated Designs (NVDs)** and **Nokia Reference Designs** is analogous to that of a supported product versus an open-source community version of the same product.

The validated designs (NVDs) are officially supported, tested on hardware, and recommended by Nokia. NVDs have a lifecycle management system, migration paths, official configuration recommendations, and support by Nokia support and services organizations.

Disclaimer: Reference designs are meant to be design guides that demonstrate alternate design and product capabilities of the Nokia portfolio but do not undergo solution-wide hardware testing and are not subjected to end-to-end lifecycle management and migration path testing like the NVDs.

With that disclaimer in place, the reference designs are also tested designs by deploying them as virtual fabrics using Containerlab. These are ensured to be functional designs and have deployable GitHub repositories that can be downloaded and customized for further use, testing, and exploration.

1 Reference architecture

This collapsed spine reference design demonstrates an architecture with enhanced scale-out requirements, using four spines instead of the typical two, while still following the general principles of a collapsed spine design. A high-level overview of the topology is shown in Figure 1.

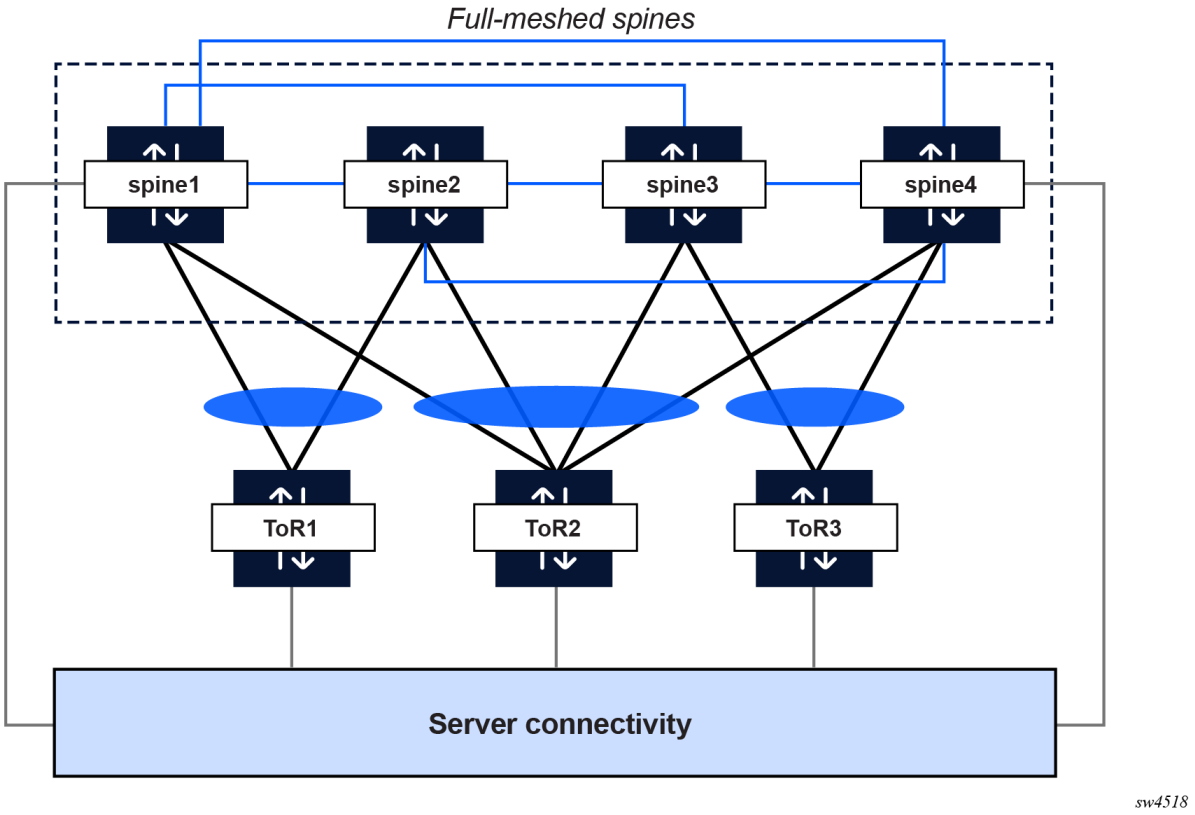

Figure 1 **4-way collapsed spine high-level design**

This design incorporates a full-mesh connectivity between all spines for alternate paths in the case of spine-to-spine link failures. Layer 2 Top-of-rack (ToR) switches are multihomed to the spines via Ethernet segments.

The 4-way collapsed spine design allows you to connect servers directly to the spines (either single-homed or multihomed) or to any of the Layer 2 ToR switches.

In general, a collapsed spine design provides the following advantages:

- Where scale-out and size of the data center are known to be limited and the main considerations become cost and power utilization, a design such as the collapsed spine design is a better choice since it reduces the number of devices and capacity of the devices required in the data center.

- Re-use of legacy Layer 2 switches (even Layer 2 switches of other vendors) as ToR switches in more modern data center designs. This allows you to connect legacy switches or switches with Layer 2 license only (thus minimizing investment in network infrastructure) while still moving into a modern architecture, which allows room to grow into a scaled-out 3-stage Clos design as and when the need arises.

Figure 2 depicts a low-level design for the 4-way collapsed spine architecture, deployed using Containerlab and SR Linux container.

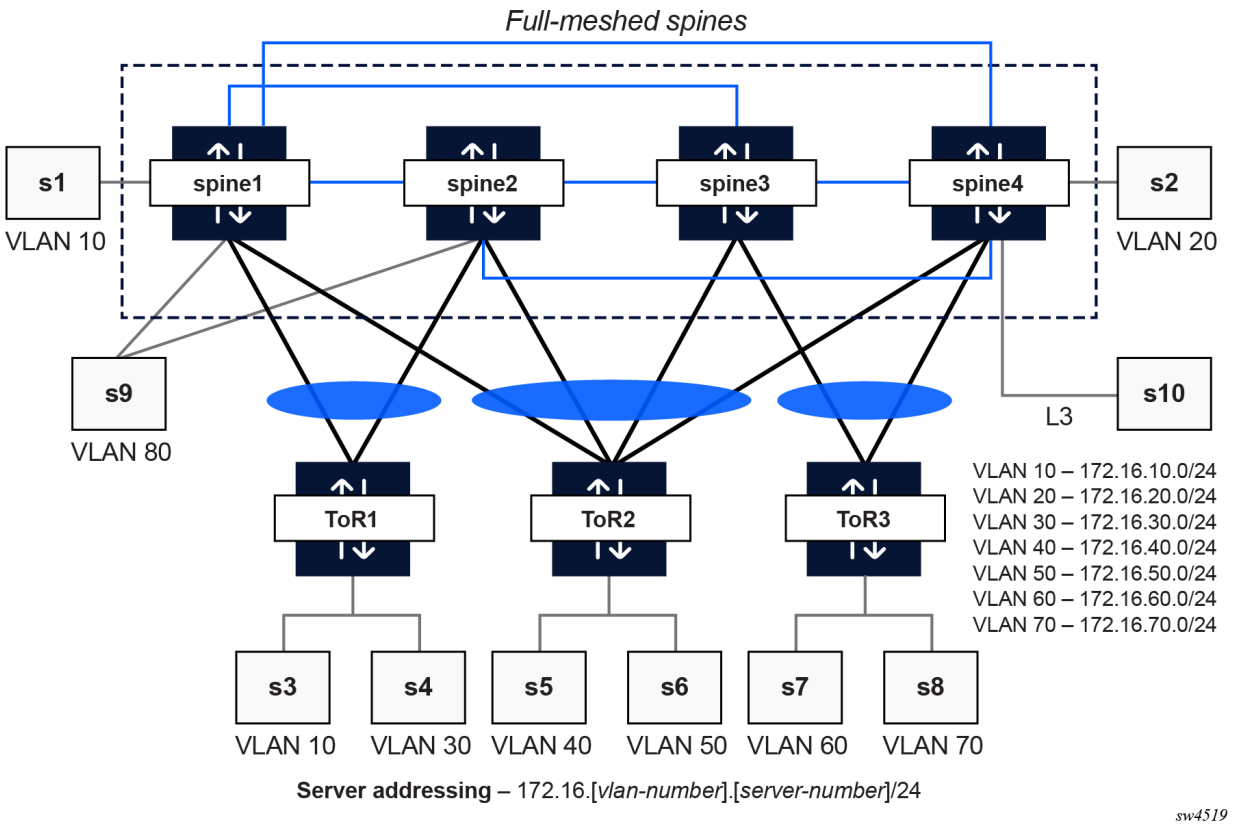

Figure 2 **4-way collapsed spine low-level design**

This design covers the server connectivity options:

- Layer 2 untagged
- Layer 2 tagged
- Layer 3 point-to-point with static routes on the leaf (for subnets behind the server) exported as Ethernet virtual private network (EVPN) Type-5 routes into the fabric
- 4-way ES-based link aggregation group (LAG) in all-active multihoming mode
- 2-way ES-based LAG in all-active multihoming mode

2 Design considerations

In a collapsed spine design, the collapsed spines (spine and leaf functionality on one node) are positioned as VXLAN tunnel endpoints (VTEPs), with the Layer 2 ToR switches connecting via Ethernet segments (and link aggregation groups on the ToR side). This design provides the following two variations in how the underlay can be implemented:

- External Border Gateway Protocol (eBGP) underlay using IPv6 link-local addressing and BGP dynamic neighbors
- Open Shortest Path First (OSPF) unnumbered

2.1 Overlay

Both design variants employ full mesh Internal Border Gateway Protocol (iBGP) peering between the spines. An eBGP peering (for the overlay) between the spines is not an appropriate design choice here since a single BGP update or withdraw, originated by any spine, will be advertised back and forth between the remaining spines, until it eventually stops at the originating spine. This creates a multiplier effect of BGP updates or withdraws between the spines, leading to unnecessary churn in the fabric. For this reason, iBGP is used instead, providing a stable overlay for the core.

2.2 Underlay

Each of these underlay variations have their own design considerations that must be carefully evaluated. Some of these are listed out in Table 1.

Table 1 **Underlay design choices for 4-way collapsed spine**

OSPF unnumbered	eBGP with IPv6 link-local
No point-to-point manual address assigning; uses a single IPv4 loopback for OSPF unnumbered peering	No point-to-point manual address assignment; uses IPv6 link-local addressing, IPv6 neighbor discovery (ND) and BGP auto-discovery
Works well with iBGP overlay since all spines are given a single Autonomous System Number (ASN)	BGP AS numbering must be altered to present a common ASN view for all spines. This can be problematic for policies that account for AS_PATH attribute since the underlay displays unique ASNs per spine and the overlay displays a single ASN across all spines

3 Feature configuration

3.1 OSPF unnumbered underlay

The point-to-point interfaces between the spines are not configured with any IPv4 or IPv6 addresses – instead, they are enabled as unnumbered interfaces, leveraging a loopback interface address. OSPF is then enabled over these unnumbered interfaces, advertising the system0 interface address for spine-to-spine reachability for the overlay.

The system0 interface, used as the VTEP address, is configured with a /32 address. These addresses are used as the source and destination addresses in the outer IP header for VXLAN tunnels. In this document, the IPv4 documentation range 192.0.2.0/24 is used for assignment.

```
// loopback address on spine1

A:admin@spine1# info interface lo0
    admin-state enable
    subinterface 0 {
        admin-state enable
        ipv4 {
            admin-state enable
            address 172.16.1.101/32 {
            }
        }
    }

// unnumbered p2p interface

A:admin@spine1# info interface ethernet-1/1
    admin-state enable
    subinterface 0 {
        admin-state enable
        ipv4 {
            admin-state enable
            unnumbered {
                admin-state enable
                interface lo0.0
            }
        }
    }

// system0 address

A:admin@spine1# info interface system0
    subinterface 0 {
        admin-state enable
        ipv4 {
            admin-state enable
            address 192.0.2.101/32 {
            }
        }
    }
```

Example 1 **Loopback interface and unnumbered point-to-point interface on spine1**

These interfaces are added in the default network-instance for the underlay, which is also enabled for OSPF.

```
// interfaces in default network-instance

A:admin@spine1# info network-instance default
    type default
    admin-state enable
    description "fabric: dc1 role: spine"
    router-id 192.0.2.101
    ip-forwarding {
        receive-ipv4-check false
    }
    interface ethernet-1/1.0 {
    }
    interface ethernet-1/2.0 {
    }
    interface ethernet-1/3.0 {
    }
    interface ethernet-1/4.0 {
    }
    interface ethernet-1/5.0 {
    }
    interface ethernet-1/6.0 {
    }
    interface lo0.0 {
    }
    interface system0.0 {
    }

*snip*

A:admin@spine1# info network-instance default protocols ospf
    instance 1 {
        admin-state enable
        version ospf-v2
        area 0.0.0.0 {
            interface ethernet-1/1.0 {
                admin-state enable
                interface-type point-to-point
                failure-detection {
                    enable-bfd true
                }
            }
            interface ethernet-1/2.0 {
                admin-state enable
                interface-type point-to-point
                failure-detection {
                    enable-bfd true
                }
            }
            interface ethernet-1/3.0 {
                admin-state enable
                interface-type point-to-point
                failure-detection {
                    enable-bfd true
                }
            }
            interface ethernet-1/4.0 {
                admin-state enable
                interface-type point-to-point
                failure-detection {
```

```
                    enable-bfd true
                }
            }
            interface ethernet-1/5.0 {
                admin-state enable
                interface-type point-to-point
                failure-detection {
                    enable-bfd true
                }
            }
            interface ethernet-1/6.0 {
                admin-state enable
                interface-type point-to-point
                failure-detection {
                    enable-bfd true
                }
            }
            interface system0.0 {
                passive true
            }
        }
    }
```

Example 2 **Default network-instance and OSPF configuration on spine1**

3.2 eBGP IPv6 link-local with BGP dynamic neighbors underlay

For an eBGP IPv6 link-local underlay, the point-to-point interfaces are first enabled with IPv6 (that will auto-generate a link-local address). Since we have IPv4 loopbacks for VXLAN tunnels and the point-to-point interfaces are enabled only with IPv6, the default network-instance must be configured to accept IPv4 packets on an IPv6 interface.

```
A:admin@spine1# info interface ethernet-1/1
    admin-state enable
    subinterface 0 {
        admin-state enable
        ipv6 {
            admin-state enable
            router-advertisement {
                router-role {
                    admin-state enable
                    max-advertisement-interval 10
                    min-advertisement-interval 4
                }
            }
        }
    }

A:admin@spine1# info interface system0
    subinterface 0 {
        admin-state enable
        ipv4 {
            admin-state enable
            address 192.0.2.101/32 {
            }
        }
    }

A:admin@spine1# info network-instance default
    type default
```

```
    admin-state enable
    router-id 192.0.2.101
    ip-forwarding {
        receive-ipv4-check false
    }
    interface ethernet-1/1.0 {
    }
    interface ethernet-1/2.0 {
    }
    interface ethernet-1/3.0 {
    }
    interface system0.0 {
    }

*snip*

A:admin@spine1# info routing-policy
    prefix-set prefixset-dc1 {
        prefix 192.0.2.0/24 mask-length-range 32..32 {
        }
    }
    policy allow-all {
        default-action {
            policy-result accept
        }
    }

A:admin@spine1# info network-instance default protocols bgp
    admin-state enable
    autonomous-system 65501
    router-id 192.0.2.11
    dynamic-neighbors {
        interface ethernet-1/1.0 {
            peer-group bgp-underlay
            allowed-peer-as [
                65502
            ]
        }
        interface ethernet-1/2.0 {
            peer-group bgp-underlay
            allowed-peer-as [
                65503
            ]
        }
        interface ethernet-1/3.0 {
            peer-group bgp-underlay
            allowed-peer-as [
                65504
            ]
        }
    }
    ebgp-default-policy {
        import-reject-all true
        export-reject-all true
    }
    afi-safi evpn {
        admin-state enable
        multipath {
            ibgp {
                maximum-paths 64
            }
        }
        evpn {
```

```
            rapid-update true
        }
    }
    afi-safi ipv4-unicast {
        admin-state enable
        multipath {
            allow-multiple-as true
            ebgp {
                maximum-paths 64
            }
        }
        ipv4-unicast {
            advertise-ipv6-next-hops true
            receive-ipv6-next-hops true
        }
        evpn {
            rapid-update true
        }
    }
    preference {
        ebgp 170
        ibgp 170
    }
    route-advertisement {
        rapid-withdrawal true
        wait-for-fib-install false
    }
    group bgp-underlay {
        admin-state enable
        export-policy [
            allow-all
        ]
        import-policy [
            allow-all
        ]
        failure-detection {
            enable-bfd true
            fast-failover true
        }
        afi-safi evpn {
            admin-state disable
        }
        afi-safi ipv4-unicast {
            admin-state enable
            ipv4-unicast {
                advertise-ipv6-next-hops true
                receive-ipv6-next-hops true
            }
        }
    }
```

Example 3 **Default network-instance, IPv6 link-local and BGP configuration for eBGP link-local underlay**

3.3 iBGP overlay with eBGP underlay

The overlay, as stated earlier, is a full mesh iBGP peering between all four spines with a common ASN – 65500. With an eBGP underlay, the ASNs on the spines must be altered to present a

common ASN view since for the eBGP underlay, each spine is configured with a unique ASN. As stated earlier, this is not an ideal design and can present operational challenges.

```
A:admin@spine1# info network-instance default protocols bgp
    admin-state enable
    autonomous-system 65501
    router-id 192.0.2.11
    dynamic-neighbors {
        interface ethernet-1/1.0 {
            peer-group bgp-underlay
            allowed-peer-as [
                65502
            ]
        }
        interface ethernet-1/2.0 {
            peer-group bgp-underlay
            allowed-peer-as [
                65503
            ]
        }
        interface ethernet-1/3.0 {
            peer-group bgp-underlay
            allowed-peer-as [
                65504
            ]
        }
    }
    ebgp-default-policy {
        import-reject-all true
        export-reject-all true
    }
    afi-safi evpn {
        admin-state enable
        multipath {
            ibgp {
                maximum-paths 64
            }
        }
        evpn {
            rapid-update true
        }
    }
    afi-safi ipv4-unicast {
        admin-state enable
        multipath {
            allow-multiple-as true
            ebgp {
                maximum-paths 64
            }
        }
        ipv4-unicast {
            advertise-ipv6-next-hops true
            receive-ipv6-next-hops true
        }
        evpn {
            rapid-update true
        }
    }
    preference {
        ebgp 170
        ibgp 170
    }
```

```
    route-advertisement {
        rapid-withdrawal true
        wait-for-fib-install false
    }
    group bgp-overlay {
        admin-state enable
        peer-as 65500
        export-policy [
            allow-all
        ]
        import-policy [
            allow-all
        ]
        failure-detection {
            enable-bfd true
            fast-failover true
        }
        multihop {
            admin-state enable
        }
        afi-safi evpn {
            admin-state enable
        }
        afi-safi ipv4-unicast {
            admin-state disable
        }
        local-as {
            as-number 65500
        }
        transport {
            local-address 192.0.2.101
        }
    }
    group bgp-underlay {
        admin-state enable
        export-policy [
            allow-all
        ]
        import-policy [
            allow-all
        ]
        failure-detection {
            enable-bfd true
            fast-failover true
        }
        afi-safi evpn {
            admin-state disable
        }
        afi-safi ipv4-unicast {
            admin-state enable
            ipv4-unicast {
                advertise-ipv6-next-hops true
                receive-ipv6-next-hops true
            }
        }
    }
    neighbor 192.0.2.102 {
        admin-state enable
        peer-group bgp-overlay
    }
    neighbor 192.0.2.103 {
        admin-state enable
        peer-group bgp-overlay
```

```
    }
    neighbor 192.0.2.104 {
        admin-state enable
        peer-group bgp-overlay
    }
```

Example 4 **iBGP configuration when using an eBGP underlay**

3.4 iBGP overlay with OSPF unnumbered underlay

For an iBGP overlay with OSPF unnumbered, all spines share a common ASN with BFD enabled for this iBGP peering for failure detection and fast failover.

```
A:admin@spine1# info network-instance default protocols bgp
    admin-state enable
    autonomous-system 65500
    router-id 192.0.2.101
    ebgp-default-policy {
        import-reject-all true
        export-reject-all true
    }
    afi-safi evpn {
        admin-state enable
        multipath {
            ibgp {
                maximum-paths 64
            }
        }
        evpn {
            rapid-update true
        }
    }
    afi-safi ipv4-unicast {
        admin-state enable
        multipath {
            allow-multiple-as true
            ebgp {
                maximum-paths 64
            }
        }
        ipv4-unicast {
            advertise-ipv6-next-hops true
            receive-ipv6-next-hops true
        }
        evpn {
            rapid-update true
        }
    }
    preference {
        ebgp 170
        ibgp 170
    }
    route-advertisement {
        rapid-withdrawal true
        wait-for-fib-install false
    }
    group bgp-overlay {
        admin-state enable
        peer-as 65500
        export-policy [
            allow-all
```

```
        ]
        import-policy [
            allow-all
        ]
        failure-detection {
            enable-bfd true
            fast-failover true
        }
        multihop {
            admin-state enable
        }
        afi-safi evpn {
            admin-state enable
        }
        afi-safi ipv4-unicast {
            admin-state disable
        }
        transport {
            local-address 192.0.2.101
        }
    }
    neighbor 192.0.2.102 {
        admin-state enable
        peer-group bgp-overlay
    }
    neighbor 192.0.2.103 {
        admin-state enable
        peer-group bgp-overlay
    }
    neighbor 192.0.2.104 {
        admin-state enable
        peer-group bgp-overlay
    }
```

Example 5 **iBGP configuration when using an OSPF unnumbered underlay**

3.5 BFD

For both underlay designs and the iBGP overlay design, BFD is enabled for failure detection and fast failure. First, BFD must be enabled on the point-to-point interfaces between the spines, and for the iBGP sessions, it must also be enabled on the system0 interface that is used as the local address (for the iBGP sessions).

```
A:admin@spine1# info bfd
    subinterface ethernet-1/1.0 {
        admin-state enable
        desired-minimum-transmit-interval 1000000
        required-minimum-receive 1000000
        detection-multiplier 3
        minimum-echo-receive-interval 1000000
    }
    subinterface ethernet-1/2.0 {
        admin-state enable
        desired-minimum-transmit-interval 1000000
        required-minimum-receive 1000000
        detection-multiplier 3
        minimum-echo-receive-interval 1000000
    }
    subinterface ethernet-1/3.0 {
        admin-state enable
```

```
        desired-minimum-transmit-interval 1000000
        required-minimum-receive 1000000
        detection-multiplier 3
        minimum-echo-receive-interval 1000000
    }
    subinterface system0.0 {
        admin-state enable
        desired-minimum-transmit-interval 1000000
        required-minimum-receive 1000000
        detection-multiplier 3
        minimum-echo-receive-interval 1000000
    }
```

Example 6 **BFD configuration**

3.6 Ethernet segments

The links to the ToR switches (from the spines) are configured as Ethernet segments in an all-active mode to allow for all links to be used.

```
A:admin@spine1# info interface ethernet-1/6
    description spine1-spine2-spine3-spine4-lag2
    admin-state enable
    ethernet {
        aggregate-id lag2
        lacp-port-priority 32768
        reload-delay 100
    }

A:admin@spine1# info system network-instance protocols evpn ethernet-segments bgp-
instance 1  ethernet-segment spine1-spine2-spine3-spine4-lag2
    admin-state enable
    esi 00:00:00:00:11:22:33:44:00:00
    multi-homing-mode all-active
    interface lag2 {
    }
    df-election {
        timers {
            activation-timer 0
        }
        algorithm {
            type default
        }
    }

A:admin@spine1# info interface lag2
    description spine1-spine2-spine3-spine4-lag2
    admin-state enable
    vlan-tagging true
    subinterface 40 {
        type bridged
        admin-state enable
        vlan {
            encap {
                single-tagged {
                    vlan-id 40
                }
            }
        }
    }
    subinterface 50 {
```

```
        type bridged
        admin-state enable
        vlan {
            encap {
                single-tagged {
                    vlan-id 50
                }
            }
        }
    }
    lag {
        lag-type lacp
        min-links 1
        lacp-fallback-mode static
        lacp-fallback-timeout 60
        lacp {
            interval FAST
            lacp-mode ACTIVE
            admin-key 2
            system-id-mac 00:00:11:22:33:44
            system-priority 32768
        }
    }
```

Example 7 **Ethernet segments configuration**

3.7 IRB interfaces

The Layer 3 gateway for all servers connected either directly to the spines or to the ToR switches is configured to be on the spines themselves using integrated routing and bridging (IRB) interfaces for the respective bridge domains (BD).

```
A:admin@spine1# info interface irb0 subinterface 10
    ip-mtu 1500
    ipv4 {
        admin-state enable
        address 172.16.10.254/24 {
            anycast-gw true
            primary
        }
        arp {
            timeout 250
            proxy-arp true
            evpn {
                advertise dynamic {
                }
            }
        }
    }
    anycast-gw {
        virtual-router-id 1
    }
```

Example 8 **IRB interfaces configuration**

3.8 MAC VRFs

Layer 2 connectivity (and Layer 2 attachment for servers) is provided via MAC VRF (Media Access Control Virtual Routing and Forwarding) instances on the spines with the appropriate import/export route targets configured for it. All expected Layer 2 interfaces are mapped to this MAC VRF; the IRB for this BD is also mapped to the MAC VRF, along with its corresponding Layer 2 VXLAN tunnel interface.

```
A:admin@spine1# info network-instance macvrf-v10
    type mac-vrf
    admin-state enable
    description macvrf-v10
    interface ethernet-1/4.4096 {
    }
    interface irb0.10 {
    }
    interface lag1.10 {
    }
    vxlan-interface vxlan0.510 {
    }
    protocols {
        bgp-evpn {
            bgp-instance 1 {
                vxlan-interface vxlan0.510
                evi 10
                ecmp 8
                routes {
                    bridge-table {
                        mac-ip {
                            advertise-arp-nd-only-with-mac-table-entry true
                        }
                    }
                }
            }
        }
        bgp-vpn {
            bgp-instance 1 {
                route-target {
                    export-rt target:10:10
                    import-rt target:10:10
                }
            }
        }
    }
```

Example 9 **MAC VRFs configuration**

3.9 IP VRFs

IP VRFs are used to provide Layer 3 segregation in the fabric, creating per-tenant networks over the common physical infrastructure. The IP VRF is also mapped to a corresponding Layer 3 VXLAN tunnel interface.

```
A:admin@spine1# info network-instance vrf1
    type ip-vrf
    admin-state enable
    description vrf1
```

```
    interface irb0.10 {
    }
    interface irb0.20 {
    }
    interface irb0.40 {
    }
    interface irb0.60 {
    }
    vxlan-interface vxlan0.500 {
    }
    protocols {
        bgp-evpn {
            bgp-instance 1 {
                vxlan-interface vxlan0.500
                evi 500
                ecmp 8
                routes {
                    route-table {
                        mac-ip {
                            advertise-gateway-mac true
                        }
                    }
                }
            }
        }
        bgp-vpn {
            bgp-instance 1 {
                route-target {
                    export-rt target:500:500
                    import-rt target:500:500
                }
            }
        }
    }
```

Example 10 **IP VRFs configuration**

4 Feature validation

4.1 Underlay validation

The underlay includes either OSPF unnumbered or eBGP IPv6 link-local (depending on what design is chosen) with BFD enabled for failure detection and fast failover. The OSPF unnumbered peering and BFD session for the same can be confirmed as follows:

```
A:admin@spine1# show network-instance default protocols ospf neighbor

=============================================================================================
=============================================================================================
=============================================================================

Net-Inst default OSPFv2 Instance 1 Neighbors
=============================================================================================
=============================================================================================
=============================================================================

+-------------------------------------------------------------------------------------------+
| Interface-Name         Rtr Id            State        Pri   RetxQ    Time Before Dead |
+===========================================================================================+
| ethernet-1/1.0         172.16.1.102      full         1     0        32                |
| ethernet-1/2.0         172.16.1.103      full         1     0        33                |
| ethernet-1/3.0         172.16.1.104      full         1     0        35                |
+-------------------------------------------------------------------------------------------+

---------------------------------------------------------------------------------------------
---------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------

No. of Neighbors: 3

=============================================================================================
=============================================================================================
=============================================================================


A:admin@spine1# info from state bfd network-instance default peer 16391
    oper-state up
    ipv4-unnumbered-interface ethernet-1/1.0
    local-address 172.16.1.101
    remote-address 172.16.1.102
    remote-discriminator 16391
    subscribed-protocols OSPF-0
    session-state UP
    remote-session-state UP
    last-state-transition "2025-06-19T05:45:34.535Z (49 minutes ago)"
    failure-transitions 0
    local-diagnostic-code NO_DIAGNOSTIC
    remote-diagnostic-code NO_DIAGNOSTIC
    remote-minimum-receive-interval 1000000
    remote-control-plane-independent false
    active-transmit-interval 1000000
    active-receive-interval 1000000
```

```
    remote-multiplier 3
    async {
        last-packet-transmitted "2025-06-19T06:35:12.450Z (now)"
        last-packet-received "2025-06-19T06:35:12.376Z (now)"
        transmitted-packets 3775
        received-packets 3777
        up-transitions 1
    }
```

Example 11 **OSPF underlay and BFD validation**

The goal of the underlay is to ensure that the system0 address is advertised and reachable via equal cost paths. The routing table can be confirmed as shown below, to display peer spine addresses learned via OSPF.

```
A:admin@spine1# show network-instance default route-table ipv4-unicast summary
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
IPv4 unicast route table of network instance default
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
+---------------------------------------+-------+------------+----------------------+----
------+----------+---------+------------+------------------------+----------------------
-+------------------------+-----------------------------------------------------+
|                Prefix                 |  ID   | Route Type |     Route Owner      |
Active  |  Origin  | Metric  |    Pref    |    Next-hop (Type)     |   Next-hop Interface
| Backup Next-hop (Type) |           Backup Next-hop Interface             |
|                                       |       |            |                      |
| Network  |         |            |                        |                      |
|                                          |
|                                       |       |            |                      |
| Instance |         |            |                        |                      |
|                                         |
+=======================================+=======+============+======================+====
======+==========+=========+============+========================+======================
=+========================+=====================================================+
| 172.16.1.101/32                       | 21    | host       | net_inst_mgr         |
True     | default  | 0       | 0          | None                   | None
|                        |                                                       |
| 192.0.2.101/32                        | 22    | host       | net_inst_mgr         |
True     | default  | 0       | 0          | None                   | None
|                        |                                                       |
| 192.0.2.102/32                        | 0     | ospfv2     | ospf_mgr             |
True     | default  | 4       | 10         | 169.254.0.1 (direct)   | ethernet-1/1.0
|                        |                                                       |
| 192.0.2.103/32                        | 0     | ospfv2     | ospf_mgr             |
True     | default  | 4       | 10         | 169.254.0.2 (direct)   | ethernet-1/2.0
|                        |                                                       |
| 192.0.2.104/32                        | 0     | ospfv2     | ospf_mgr             |
True     | default  | 4       | 10         | 169.254.0.3 (direct)   | ethernet-1/3.0
|                        |                                                       |
```

```
+-------------------------------------+-------+------------+----------------------+----
------+----------+---------+------------+----------------------+----------------------
-+-----------------------+-----------------------------------------------------+
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
---------------------------------------------------------------------------
IPv4 routes total                    : 5
IPv4 prefixes with active routes     : 5
IPv4 prefixes with active ECMP routes: 0
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
---------------------------------------------------------------------------
```

Example 12 **IPv4 routing table for the underlay with routes learned via OSPF**

If an eBGP IPv6 link-local design is implemented, this can be confirmed in Example 13, with the underlay routes (for reachability to peer system0 addresses) shown in the example:

```
A:admin@spine1# show network-instance default protocols bgp neighbor | as yaml
---
neighbor summary:

*snip*

      - Net-Inst: default
      - Peer: 'fe80::1827:dff:feff:2%ethernet-1/3.0'
        Group: bgp-underlay
        Flags: DB
        Peer-AS: 65504
        State: established
        Uptime: '0d:0h:3m:5s'
        AFI/SAFI: ipv4-unicast
        '[Rx/Active/Tx]': '[3/1/3]'
      - Net-Inst: default
      - Peer: 'fe80::183e:bff:feff:1%ethernet-1/1.0'
        Group: bgp-underlay
        Flags: DB
        Peer-AS: 65502
        State: established
        Uptime: '0d:0h:3m:4s'
        AFI/SAFI: ipv4-unicast
        '[Rx/Active/Tx]': '[3/1/3]'
      - Net-Inst: default
      - Peer: 'fe80::1862:cff:feff:3%ethernet-1/2.0'
        Group: bgp-underlay
        Flags: DB
        Peer-AS: 65503
        State: established
        Uptime: '0d:0h:3m:6s'
        AFI/SAFI: ipv4-unicast
        '[Rx/Active/Tx]': '[3/1/3]'
    config:
      - network-instance: default
        configured-peers: 3
        configured-up-peers: 3
        disabled-peers: 0
        dynamic-peers: 3

A:admin@spine1# show network-instance default route-table ipv4-unicast summary
```

```
--------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
--------------------------------------------------------------------------
IPv4 unicast route table of network instance default
--------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
--------------------------------------------------------------------------
+---------------------------------------+-------+------------+----------------------+----
------+----------+---------+------------+----------------------+----------------------
-+------------------------+-----------------------------------------------------+
|                Prefix                 |  ID   | Route Type |     Route Owner      |
Active  |  Origin  | Metric  |    Pref    |    Next-hop (Type)     |   Next-hop Interface
| Backup Next-hop (Type) |           Backup Next-hop Interface                 |
|                                       |       |            |                      |
| Network  |         |            |                        |                      |
|                                          |
|                                       |       |            |                      |
| Instance |         |            |                        |                      |
|                                          |
+=======================================+=======+============+======================+====
======+==========+=========+============+======================+======================
=+========================+=====================================================+
| 192.0.2.101/32                        | 21    | host       | net_inst_mgr         |
True     | default  | 0       | 0          | None                    | None
|                        |                                                        |
| 192.0.2.102/32                        | 0     | bgp        | bgp_mgr              |
True     | default  | 0       | 170        | fe80::188c:bff:feff:1   | ethernet-1/1.0
|                        |                                                        |
|                                       |       |            |                      |
|          |         |            | (direct)               |                      |
|                                          |
| 192.0.2.103/32                        | 0     | bgp        | bgp_mgr              |
True     | default  | 0       | 170        | fe80::18c7:cff:feff:3   | ethernet-1/2.0
|                        |                                                        |
|                                       |       |            |                      |
|          |         |            | (direct)               |                      |
|                                          |
| 192.0.2.104/32                        | 0     | bgp        | bgp_mgr              |
True     | default  | 0       | 170        | fe80::18c4:dff:feff:2   | ethernet-1/3.0
|                        |                                                        |
|                                       |       |            |                      |
|          |         |            | (direct)               |                      |
|                                          |
+---------------------------------------+-------+------------+----------------------+----
------+----------+---------+------------+----------------------+----------------------
-+------------------------+-----------------------------------------------------+
--------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
--------------------------------------------------------------------------
IPv4 routes total                    : 4
IPv4 prefixes with active routes     : 4
IPv4 prefixes with active ECMP routes: 0
--------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------------------
--------------------------------------------------------------------------
```

Example 13 **eBGP underlay validation**

4.2 Overlay validation

The outcome from iBGP configuration (regardless of whether it is an eBGP IPv6 link-local or OSPF unnumbered underlay) is the same: the overlay peering is established between the loopbacks of the spines and is only enabled for the EVPN AFI/SAFI (Address Family Identifier/Subsequent Address Family Identifier (SAFI); it is receiving EVPN routes from these overlay peers as well, confirmed in Example 14.

```
A:admin@spine1# show network-instance default protocols bgp neighbor
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------
BGP neighbor summary for network-instance "default"
Flags: S static, D dynamic, L discovered by LLDP, B BFD enabled, - disabled, * slow
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------
+--------------------------+-----------------------------------------+--------------------
--------+---------+--------------+----------------------+----------------------+---------
----------+-----------------------------------------+
|         Net-Inst         |                  Peer                   |           Group
|  Flags  |   Peer-AS    |        State         |        Uptime        |     AFI/SAFI
|            [Rx/Active/Tx]             |
+==========================+=========================================+====================
========+=========+==============+======================+======================+=========
==========+=========================================+
| default                  | 192.0.2.102                             | bgp-overlay
| SB      | 65500        | established          | 0d:0h:0m:27s         | evpn
| [26/26/28]                            |
| default                  | 192.0.2.103                             | bgp-overlay
| SB      | 65500        | established          | 0d:0h:0m:22s         | evpn
| [15/15/28]                            |
| default                  | 192.0.2.104                             | bgp-overlay
| SB      | 65500        | established          | 0d:0h:0m:23s         | evpn
| [18/18/28]                            |
+--------------------------+-----------------------------------------+--------------------
--------+---------+--------------+----------------------+----------------------+---------
----------+-----------------------------------------+
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------
Summary:
3 configured neighbors, 3 configured sessions are established, 0 disabled peers
0 dynamic peers

A:admin@spine1# info from state network-instance default bgp-rib afi-safi evpn evpn
local-rib
    ethernet-ad-route 192.0.2.101:10 esi 00:00:00:00:00:11:22:00:00:00 ethernet-tag-id 0
neighbor 0.0.0.0 path-id 0 {
        attr-id 27
        last-modified "2025-06-19T07:12:18.100Z (4 minutes ago)"
        used-route false
        valid-route true
        best-route true
        backup-route false
```

```
        stale-route false
        pending-delete false
        neighbor-as 0
        group-best true
        tie-break-reason none
        label {
            value 10010
            value-type vni
        }
    }
    ethernet-ad-route 192.0.2.101:10 esi 00:00:00:00:00:11:22:00:00:00 ethernet-tag-id
4294967295 neighbor 0.0.0.0 path-id 0 {
        attr-id 24
        last-modified "2025-06-19T07:12:18.100Z (4 minutes ago)"
        used-route false
        valid-route true
        best-route true
        backup-route false
        stale-route false
        pending-delete false
        neighbor-as 0
        group-best true
        tie-break-reason none
    }
    ethernet-ad-route 192.0.2.101:30 esi 00:00:00:00:00:11:22:00:00:00 ethernet-tag-id 0
neighbor 0.0.0.0 path-id 0 {
        attr-id 28
        last-modified "2025-06-19T07:12:18.100Z (4 minutes ago)"
        used-route false
        valid-route true
        best-route true
        backup-route false
        stale-route false
        pending-delete false
        neighbor-as 0
        group-best true
        tie-break-reason none
        label {
            value 10030
            value-type vni
        }
    }

*snip*
```

Example 14 **Overlay validation and EVPN routes**

4.3 MAC address and IP-MAC bindings on spines (single-homed and Ethernet segments)

MAC addresses are learned directly on the spines via single-homed interfaces or over Ethernet segments. Server s1 is single-homed to spine1, while server s3 is behind ToR1, which is connected to spine1 and spine2 via an Ethernet segment – the MAC address learning for both these servers is confirmed in Example 15.

```
A:admin@spine1# show network-instance macvrf-v10 bridge-table mac-table mac
AA:C1:AB:66:E4:E4
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Mac-table of network instance macvrf-v10
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Mac                     : AA:C1:AB:66:E4:E4
Destination             : ethernet-1/4.4096
Dest Index              : 9
Type                    : learnt
Programming Status      : Success
Aging                   : 284
Last Update             : 2025-06-19T07:33:47.000Z
Duplicate Detect time   : N/A
Hold down time remaining: N/A
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------

A:admin@spine1# show network-instance macvrf-v10 bridge-table mac-table mac
AA:C1:AB:78:A7:B3
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Mac-table of network instance macvrf-v10
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Mac                     : AA:C1:AB:78:A7:B3
Destination             : lag1.10
Dest Index              : 17
Type                    : evpn
Programming Status      : Success
Aging                   : N/A
Last Update             : 2025-06-19T07:36:38.000Z
Duplicate Detect time   : N/A
Hold down time remaining: N/A
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------

A:admin@spine1# show arpnd arp-entries interface irb0 subinterface 10
+-------------------------+-------------------------+-----------------+----------------
----------+-------------------------------------------------+---------------------------
-------------------------------------------------------------------+
|        Interface        |       Subinterface      |    Neighbor     |          Origin
|               Link layer address                  |
Expiry                                               |
+=========================+=========================+=================+================
==========+=================================================+===========================
===================================================================+
| irb0                    |                      10 |     172.16.10.1 |
dynamic | AA:C1:AB:66:E4:E4                                 | a minute from now
|
| irb0                    |                      10 |     172.16.10.3 |
dynamic | AA:C1:AB:78:A7:B3                                 | 2 minutes from now
|
+-------------------------+-------------------------+-----------------+----------------
----------+-------------------------------------------------+---------------------------
-------------------------------------------------------------------+
```

```
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
  Total entries : 2 (0 static, 2 dynamic)
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

Example 15 **MAC address state for server s1 and s3**

These are then advertised as EVPN Type-2 MAC+IP routes to BGP EVPN neighbors, as confirmed in Example 16:

```
A:admin@spine1# info from state network-instance default bgp-rib afi-safi evpn evpn rib-
in-out rib-out-post mac-ip-route 192.0.2.101:10 mac-length 48 mac-address
AA:C1:AB:66:E4:E4 ip-address 172.16.10.1 ethernet-tag-id 0 neighbor 192.0.2.102 path-id 0
    esi 00:00:00:00:00:00:00:00:00:00
    attr-id 52
    next-hop 192.0.2.101
    label1 {
        value 10010
        value-type vni
    }

A:admin@spine1# info from state network-instance default bgp-rib attr-sets attr-set 52
    origin igp
    atomic-aggregate false
    next-hop 192.0.2.101
    local-pref 100
    communities {
        ext-community [
            target:10:10
            bgp-tunnel-encap:VXLAN
        ]
    }
```

Example 16 **EVPN route originated for server s1's MAC + IP address**

4.4 Routes inserted into IP VRF via EVPN Type-5 routes

Server s10, attached via a Layer 3 interface to spine4, has several loopback addresses that must be reachable from within the fabric (that can be grouped into a 172.16.92.0/22 network). Spine4 has a static route configured for this, which is advertised as EVPN Type-5 routes into the fabric. Using spine1 as a reference, this route is shown below. This is eventually installed in the IP VRF routing table (and forwarding table) for proper forwarding to this subnet.

```
A:admin@spine1# info from state network-instance default bgp-rib afi-safi evpn evpn
local-rib ip-prefix-route 192.0.2.104:500 ethernet-tag-id 0 ip-prefix-length 22 ip-prefix
172.16.92.0/22 neighbor 192.0.2.104 path-id 0
    esi 00:00:00:00:00:00:00:00:00:00
    gateway-ip 0.0.0.0
    attr-id 117
    last-modified "2025-06-19T07:12:39.100Z (an hour ago)"
    used-route true
    valid-route true
    best-route true
    backup-route false
    stale-route false
    pending-delete false
```

```
    tie-break-reason none
    imported-network-instances [
        vrf1
    ]
    label {
        value 10500
        value-type vni
    }

A:admin@spine1# show network-instance vrf1 route-table ipv4-unicast prefix 172.16.92.0/22
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------
IPv4 unicast route table of network instance vrf1
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------
+--------------------------------------+-------+------------+----------------------+----
------+----------+---------+------------+------------------------+-----------------------
-+------------------------+-----------------------------------------------------+
|                Prefix                |  ID   | Route Type |     Route Owner      |
Active  |  Origin  | Metric  |    Pref    |    Next-hop (Type)     |   Next-hop Interface
| Backup Next-hop (Type) |            Backup Next-hop Interface             |
|                                      |       |            |                      |
| Network  |         |            |                        |                      |
|                                             |
|                                      |       |            |                      |
| Instance |         |            |                        |                      |
|                                            |
+======================================+=======+============+======================+====
======+==========+=========+============+========================+=======================
=+========================+=====================================================+
| 172.16.92.0/22                       | 0     | bgp-evpn   | bgp_evpn_mgr         |
True     | vrf1     | 4       | 170        | 192.0.2.104/32         |
|                          |                                                        |
|                                      |       |            |                      |
|          |         |            | (indirect/vxlan)       |                      |
|                                            |
+--------------------------------------+-------+------------+----------------------+----
------+----------+---------+------------+------------------------+-----------------------
-+------------------------+-----------------------------------------------------+
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------
```

Example 17 **EVPN Type-5 route originated for 172.16.92.0/22**

4.5 Ethernet segments

All ToR switches are connected via all-active Ethernet segments on the spines. These ToR switches are configured for LAG interfaces (on their spine-facing interfaces). The state of these Ethernet segments can be confirmed in Example 18 (for the two Ethernet segments that are configured on spine1):

```
A:admin@spine1# show system network-instance ethernet-segments spine1-spine2-lag1
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
spine1-spine2-lag1 is up, all-active
  ESI      : 00:00:00:00:00:11:22:00:00:00
  Alg      : default
  Peers    : 192.0.2.102
  Interface: lag1
  Next-hop : N/A
  evi      : N/A
  Network-instances:
     macvrf-v10
      Candidates : 192.0.2.101 (DF), 192.0.2.102
      Interface : lag1.10
  Network-instances:
     macvrf-v30
      Candidates : 192.0.2.101 (DF), 192.0.2.102
      Interface : lag1.30
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Summary
 1 Ethernet Segments Up
 0 Ethernet Segments Down
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------

A:admin@spine1# show system network-instance ethernet-segments spine1-spine2-spine3-
spine4-lag2
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
spine1-spine2-spine3-spine4-lag2 is up, all-active
  ESI      : 00:00:00:00:11:22:33:44:00:00
  Alg      : default
  Peers    : 192.0.2.104, 192.0.2.102, 192.0.2.103
  Interface: lag2
  Next-hop : N/A
  evi      : N/A
  Network-instances:
     macvrf-v40
      Candidates : 192.0.2.101 (DF), 192.0.2.102, 192.0.2.103, 192.0.2.104
      Interface : lag2.40
  Network-instances:
     macvrf-v50
      Candidates : 192.0.2.101, 192.0.2.102, 192.0.2.103 (DF), 192.0.2.104
      Interface : lag2.50
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Summary
 1 Ethernet Segments Up
 0 Ethernet Segments Down
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------
```

Example 18 **Ethernet segments' state on spine1**

5 Packet walks

5.1 Single-homed server to single-homed server

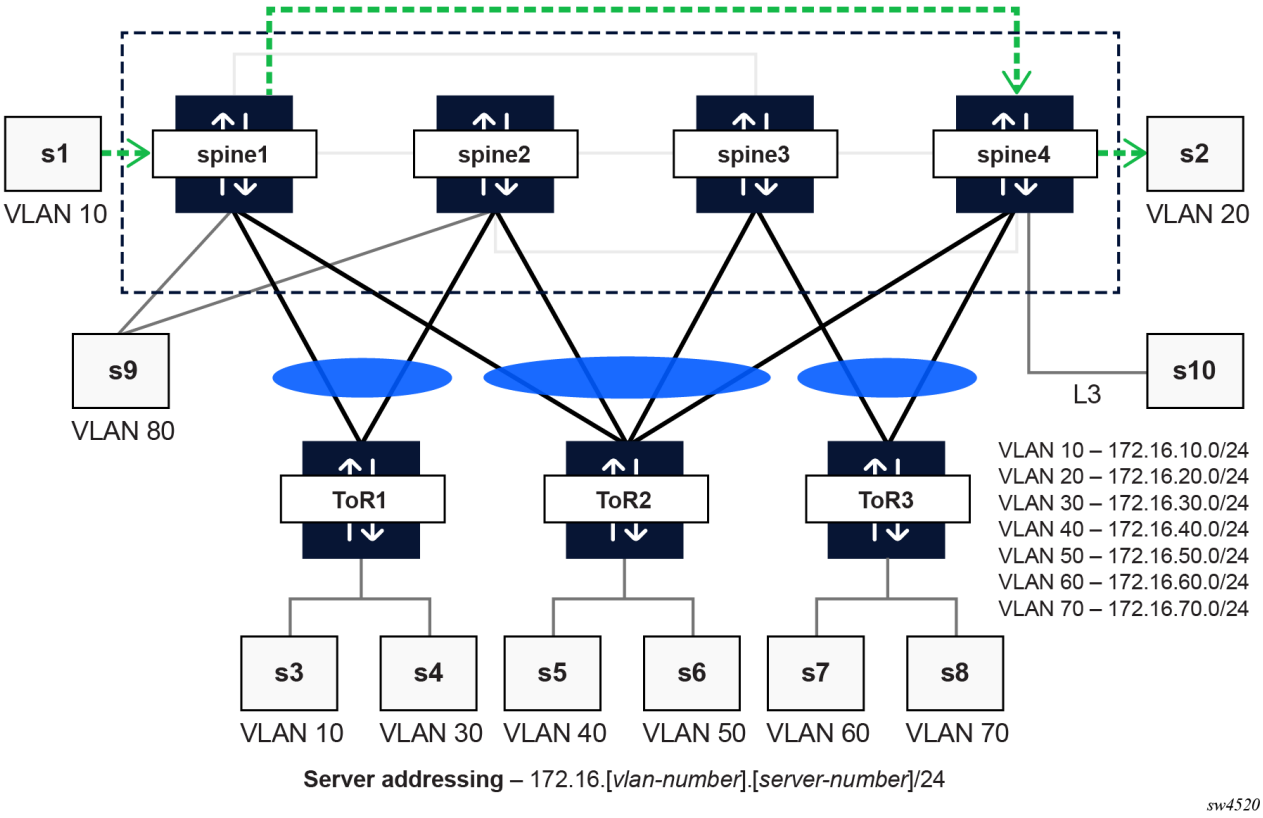

Figure 3 **Single-homed server to single-homed server**

5.2 Single-homed server to Layer 3-attached server

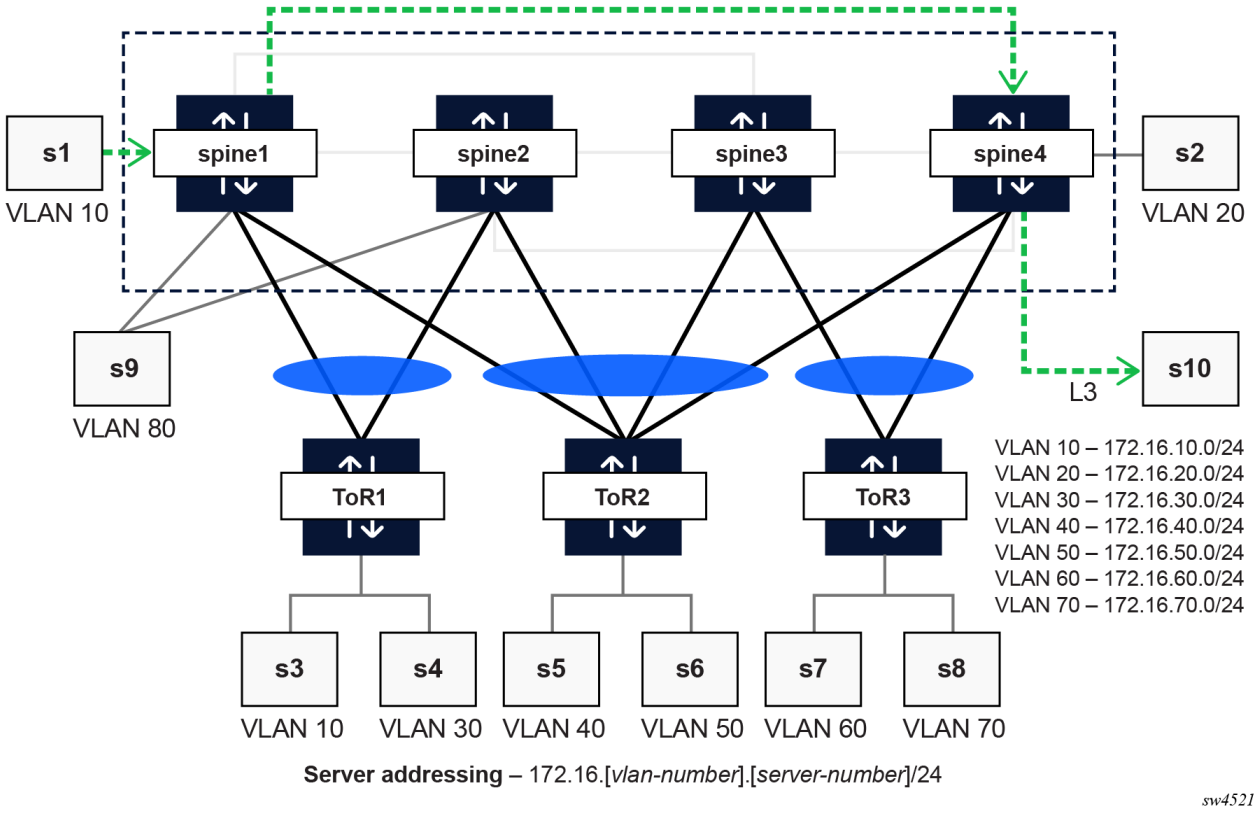

Figure 4 **Single-homed server to Layer 3-attached server**

5.3 Single-homed server to multihomed server (directly attached to spines)

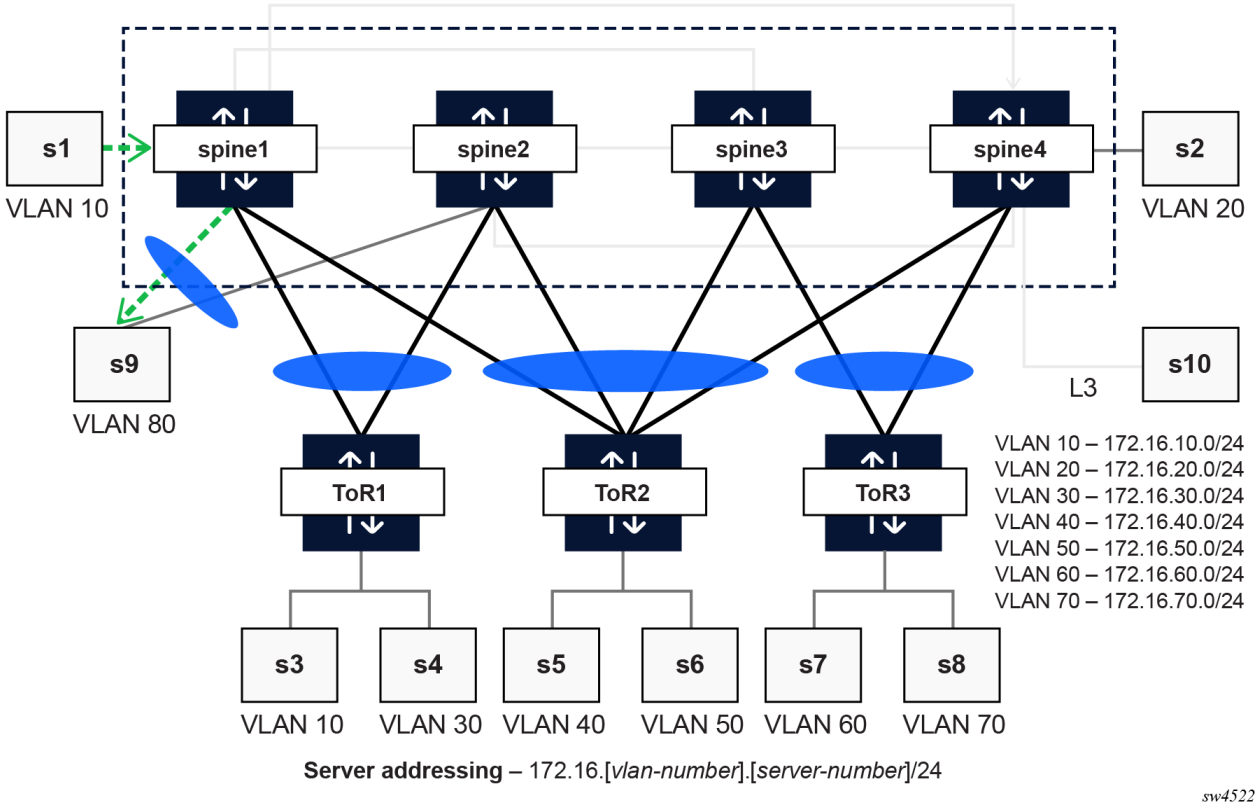

Figure 5 **Single-homed server to multihomed server (directly attached to spines)**

5.4 Single-homed server to multihomed server (behind ToR switch)

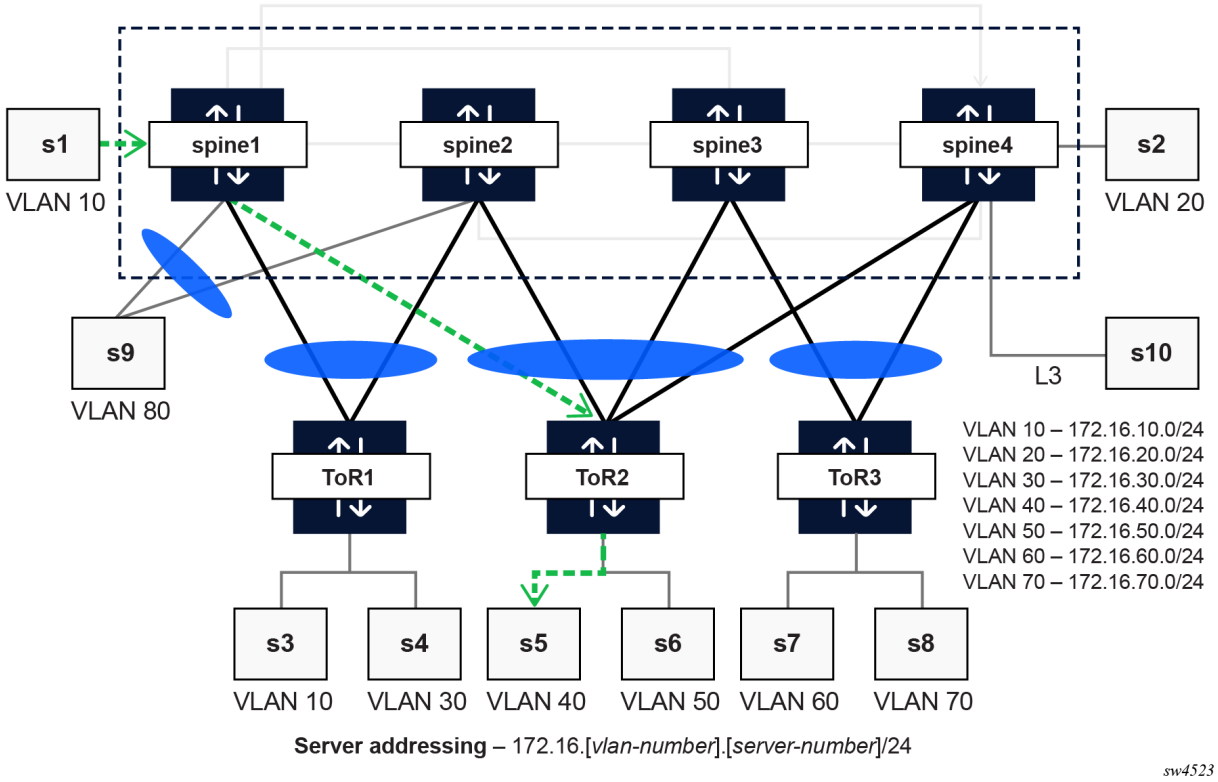

Figure 6 **Single-homed server to multihomed server (behind ToR switch)**

6 Digital twin with Containerlab

Digital twins are an integral part of Day-0 through Day-2 operations, providing the operations and deployment teams with the opportunity to continuously validate the look and feel of any deployment. These virtual fabrics also grant the ability to learn and play with technologies and designs – in this case, the collapsed spine EVPN VXLAN fabric reference design.

A digital twin of this reference design can be deployed using Containerlab and containerized SR Linux. The repository can be found here: https://github.com/nokia/nokia-validated-designs/tree/main/reference-designs/4-way-collapsed-spine