Configure ACI Multi-Site Deployment

Contents

Introduction

This document describes the steps to set up and configure Application Centric Infrastructure (ACI) multi-site fabric.

The ACI Multi-Site feature introduced in Release 3.0 allows you to interconnect separate Cisco ACI Application Policy Infrastructure Controller (APIC) cluster domains (fabrics). Each site represents a different availability zone. This helps to ensure multi-tenant Layer 2 and Layer 3 network connectivity across sites and it also extends the policy domain end-to-end across fabrics. You can create policies in the Multi-Site GUI and push them to all integrated sites or selected sites. Alternatively, you can import tenants and their policies from a single site and deploy them on other sites.

Prerequisites

Requirements

Cisco recommends that you:

- Complete the instructions in the [Cisco ACI Multi-Site Orchestrator Installation and Upgrade Guide](#) in order to set up the Multi-Site Controller (MSC).
- Ensure ACI fabrics have been fully discovered in two or more sites.
- Ensure the APIC clusters deployed in separate sites have the Out of Band (OOB) management connectivity to the MSC nodes.

Components Used

The information in this document is based on these software and hardware versions:

Site A

Hardware Device	Logical Name
N9K-C9504 w/ N9K-X9732C-EX	spine109
N9K-C93180YC-EX	leaf101
N9K-C93180YC-EX	leaf102
N9K-C9372PX-E	leaf103
APIC-SERVER-M2	apic1

Site B

Hardware Device	Logical Name
N9K-C9504 w/ N9K-X9732C-EX	spine209
N9K-C93180YC-EX	leaf201
N9K-C93180YC-EX	leaf202
N9K-C9372PX-E	leaf203
APIC-SERVER-M2	apic2

IP Network (IPN) N9K-C93180YC-EX

Hardware	Version
APIC	Version 3.1(2m)
MSC	Version: 1.2(2b)
IPN	NXOS: Version 7.0(3)I4(8a)

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.

Background Information

Note: The cross-site namespace normalization is performed by the connecting spine switches. This requires 2nd generation or later Cisco Nexus 9000 Series switches with "EX" or "FX" on the end of the product name. Alternatively, Nexus 9364C is supported in ACI Multi-Site Release 1.1(x) and later.

For more details on hardware requirements and compatibility information, see the [ACI Multi-Site Hardware Requirements Guide](#).

Configure

Logical Network Diagram
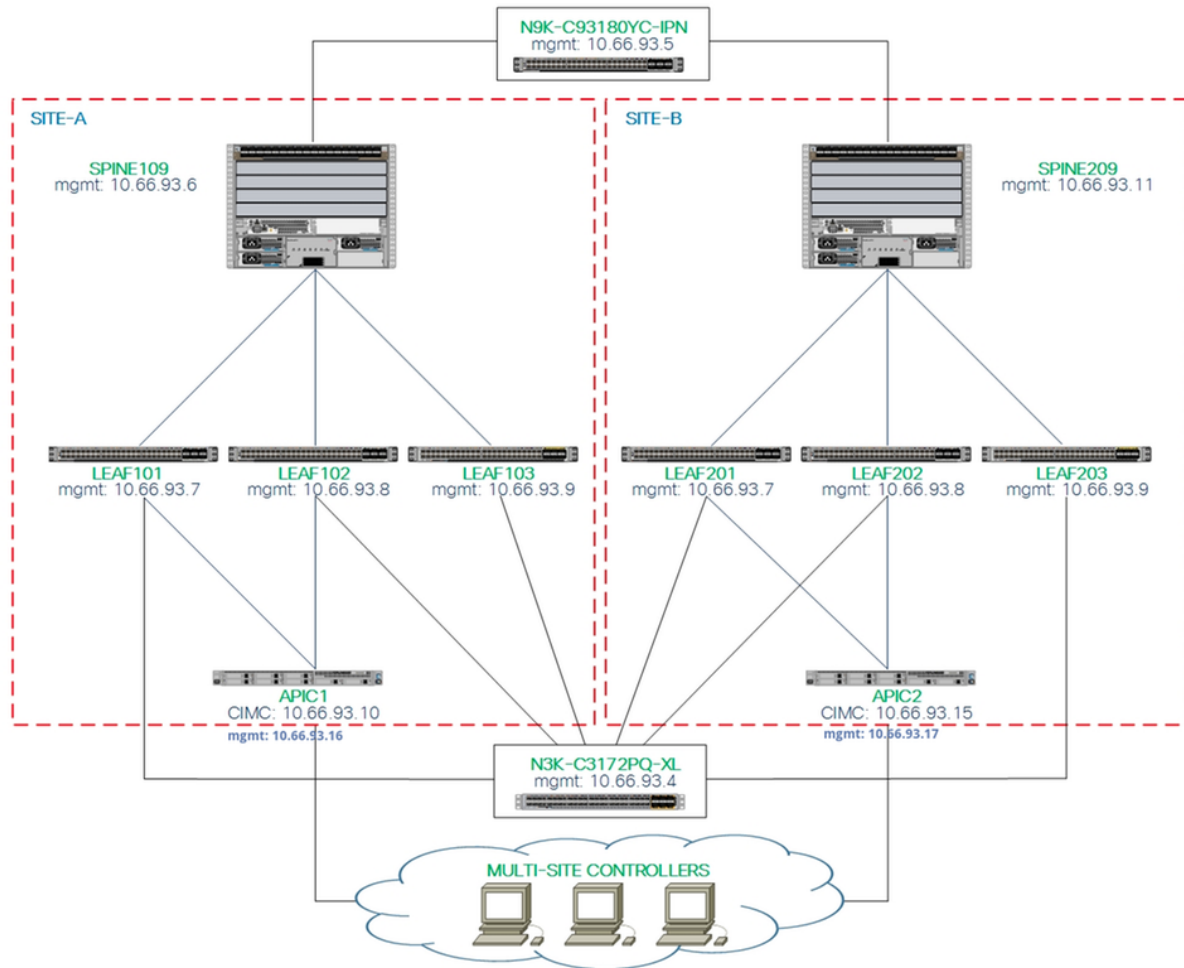
Configurations

This document mainly focuses on an ACI and MSC side configuration for Multi-Site deployment. IPN switch configuration details are not fully covered. However, a few important configurations from the IPN switch are listed for reference purposes.

IPN Switch Configuration

These configurations are used in the IPN device connected to the ACI spines.

```
vrf context intersite
  description VRF for Multi-Site lab

  feature ospf
  router ospf intersite
      vrf intersite
```

//Towards to Spine109 in Site-A   **// Towards to Spine209 in Site-B**

```
interface Ethernet1/49                    interface Ethernet1/50
  speed 100000                              speed 100000
  mtu 9216                                  mtu 9216
  no negotiate auto                         no negotiate auto
  no shutdown                               no shutdown

interface Ethernet1/49.4                   interface Ethernet1/50.4
  mtu 9150                                  mtu 9150
```

```
encapsulation dot1q 4                        encapsulation dot1q 4
vrf member intersite                         vrf member intersite
ip address 172.16.1.34/27                     ip address 172.16.2.34/27
ip ospf network point-to-point               ip ospf network point-to-point
ip router ospf intersite area 0.0.0.1        ip router ospf intersite area 0.0.0.1
no shutdown                                  no shutdown
```

Note: Maximum transmission unit (MTU) of Multiprotocol Border Gateway Protocol (MP-BGP) Ethernet Virtual Private Network (EVPN) control plane communication between spine nodes in different sites - By default, the spine nodes generate 9000-byte packets to exchange endpoint routing information. If that default value is not modified, the Inter Site Network (ISN) must support an MTU size of at least 9100 bytes. In order to tune the default value, modify the corresponding system settings in each APIC domain.

This example uses the default control plane MTU size (9000 bytes) on the spine nodes.

Required Configuration from APIC

1. Configure iBGP AS and Route Reflector for each site from the APIC GUI. Log in the site's APIC and configure internal Border Gateway Protocol (iBGP) Autonomous System Number and Route Reflector Nodes for each site's APIC cluster. Choose **APIC GUI > System > System Settings > BGP Route Reflector**. This is the default BGP Route Route Reflector Policy which will be used for the fabric pod profile.
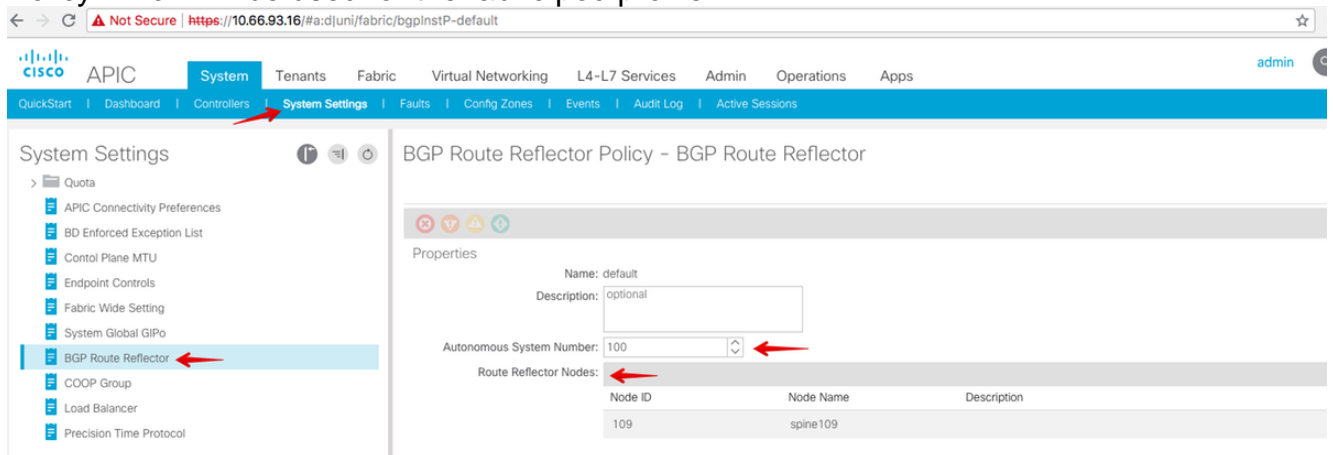

Configure the fabric pod profile for each site's APIC cluster. Choose **APIC GUI > Fabric > Fabric Policies > Pod Policies > Policy Groups**. Click the default Pod policy group. From the BGP Route Reflector Policy drop-down list, choose **default**.
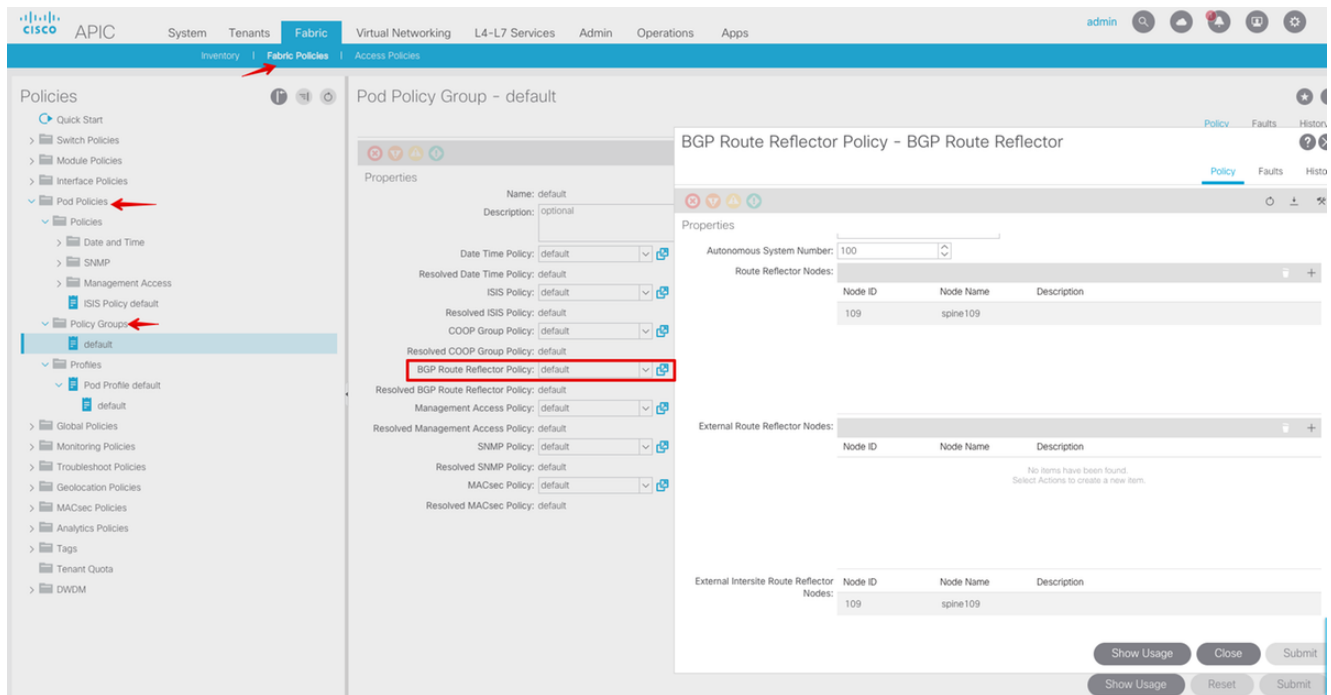
2. Configure spine access policies to include external routed domains for each site from the APIC GUI. Configure the spine access policies for spine uplink to the IPN switch with an Access Entity Profile (AEP) and Layer 3 routed domain (**APIC GUI > Fabric > Access Policies**). Create the switch profile.
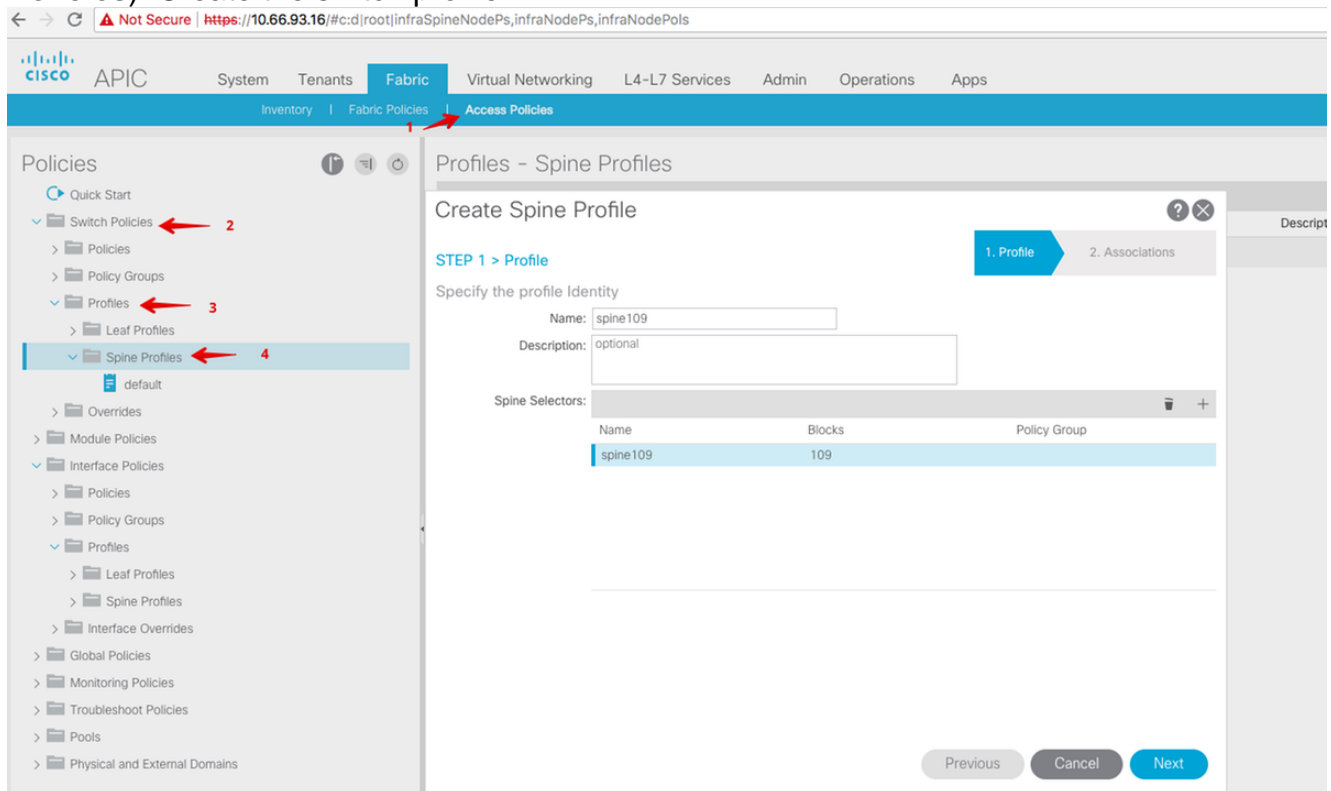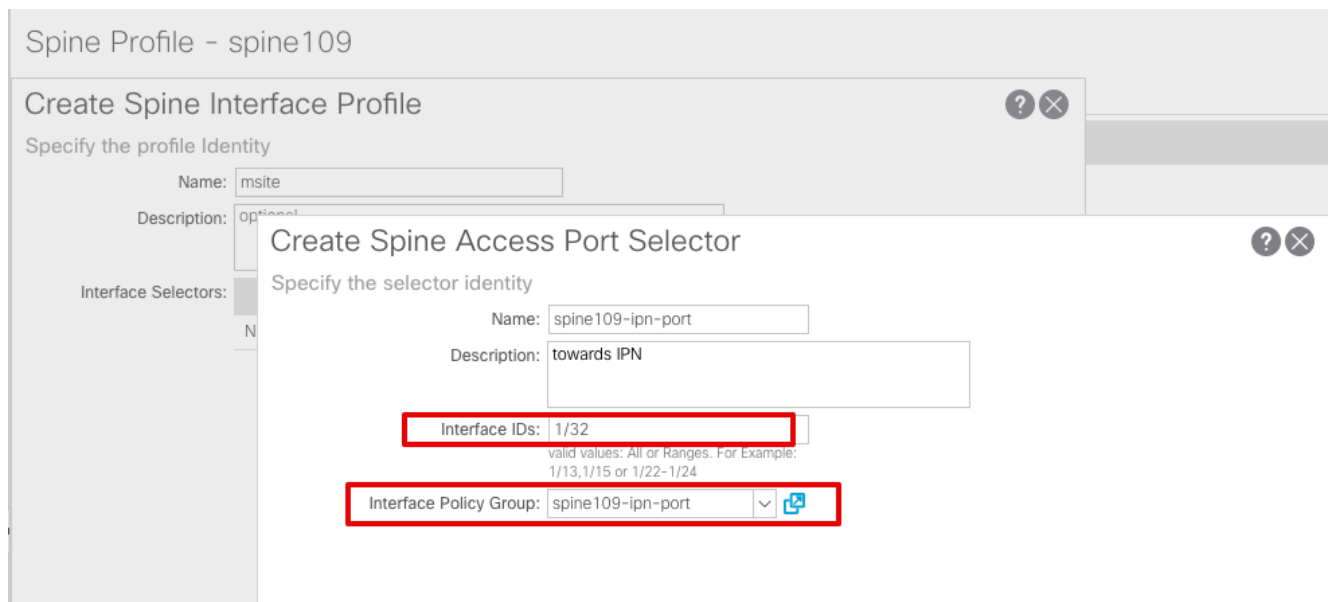

Create the Attachable Access Entity Profile (AAEP), Layer 3 Routed domain, and VLAN Pool.
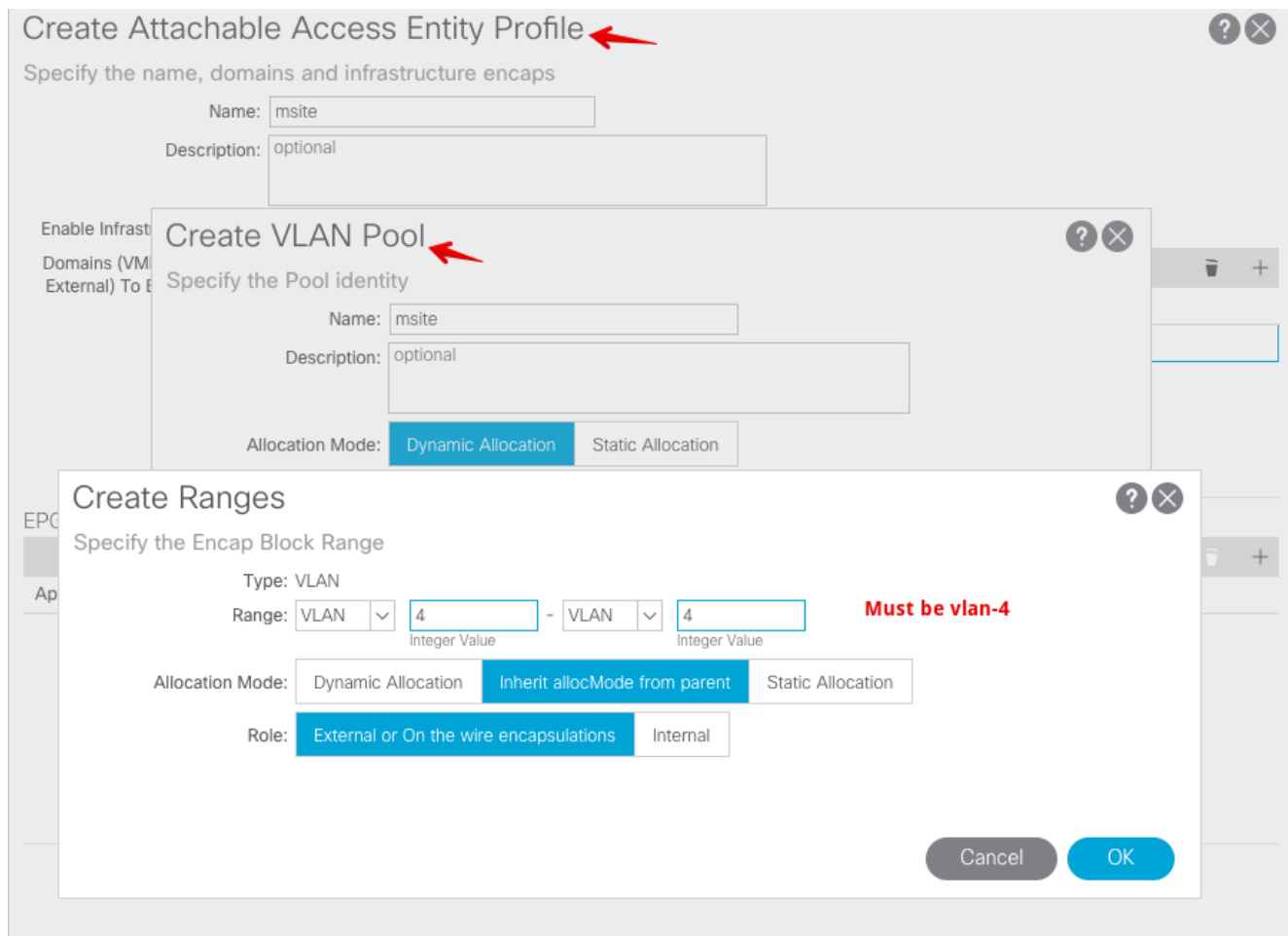
Create the Spine Access Port Policy Group. From the Attached Entity Profile drop-down list, choose **msite**.
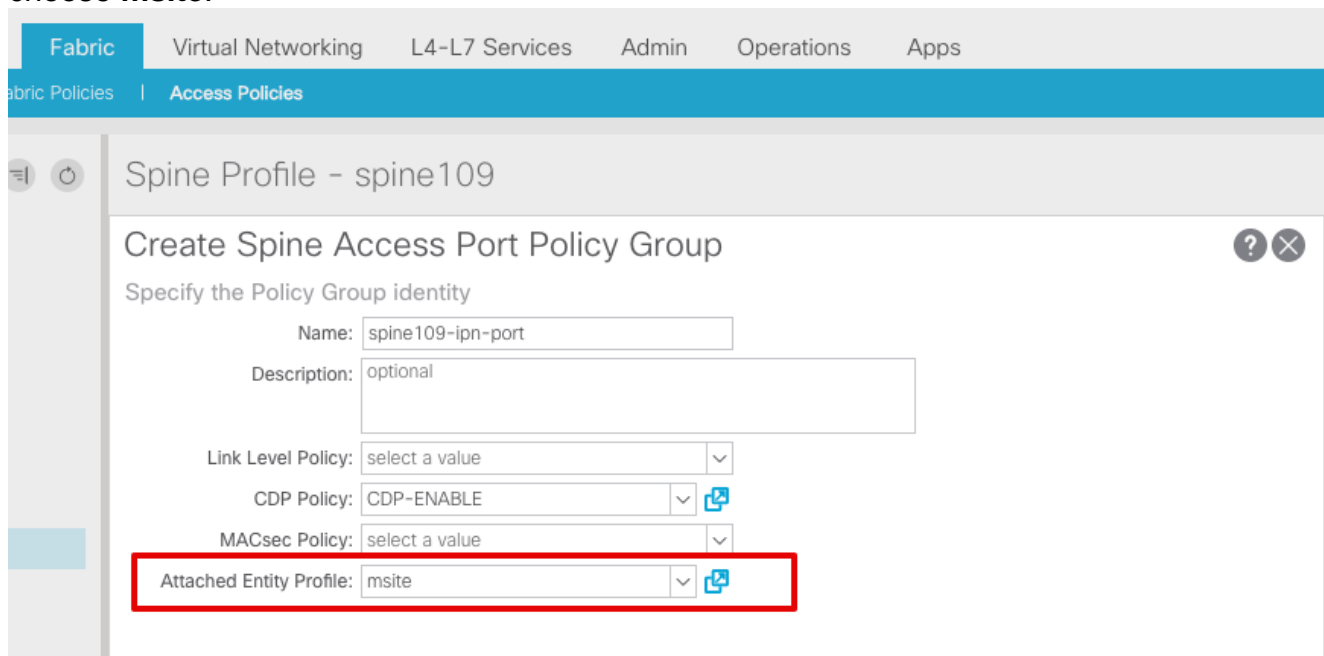

Create the Spine Interface Profile. Associate the IPN facing spine access port to the interface policy group created in the previous step.

Note: As for now, there is no need to configure L3Out of Open Shortest Path First (OSPF) under infra tenant from the APIC GUI. This will be configured via MSC and the configuration pushed to each site later.

3. Configure the external dataplane Tunnel End Point (TEP) per site from the APIC GUI. Choose **APIC GUI > Infra > Policies > Protocol > Fabric Ext Connection Policies**. Then create an intrasite/intersite profile.
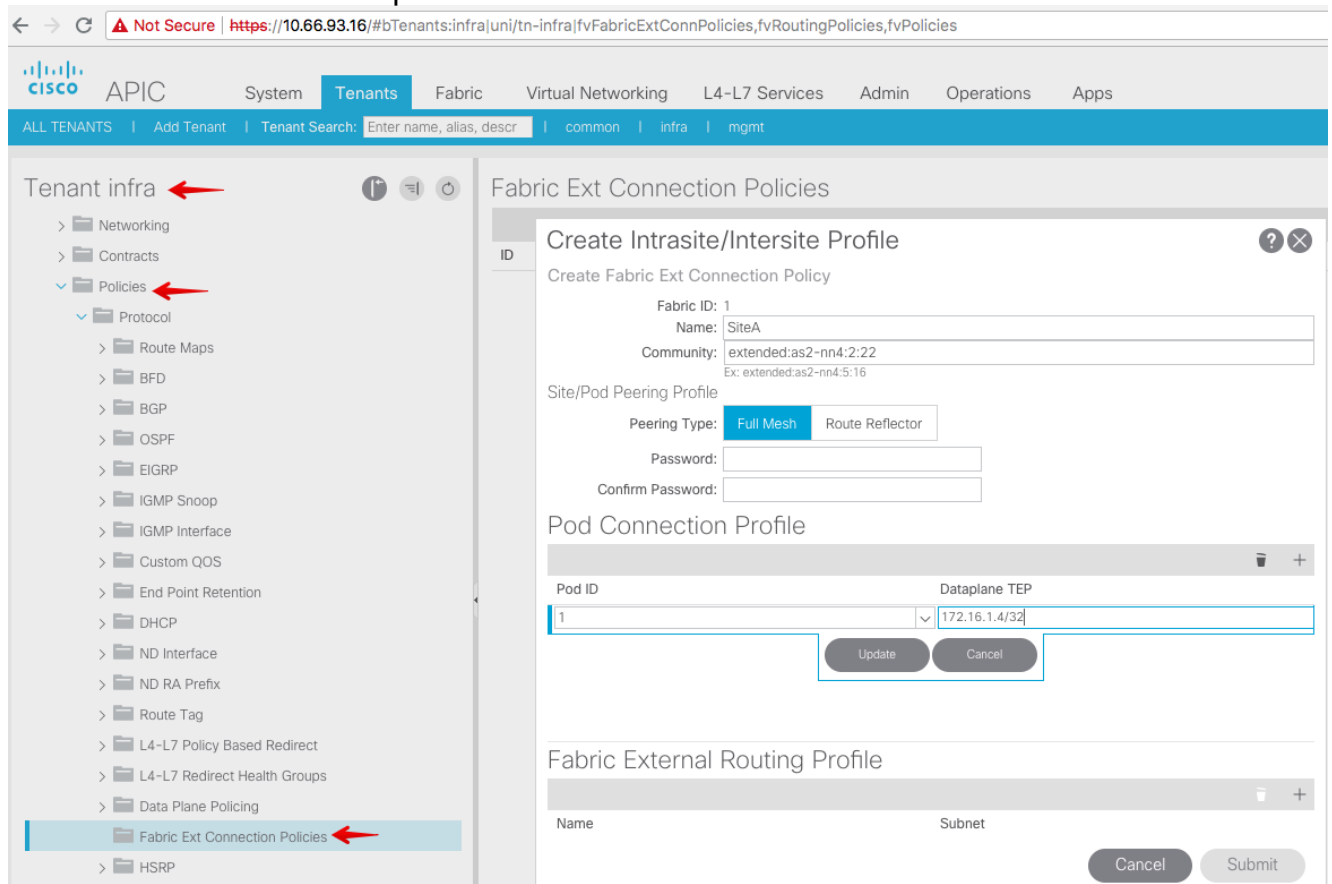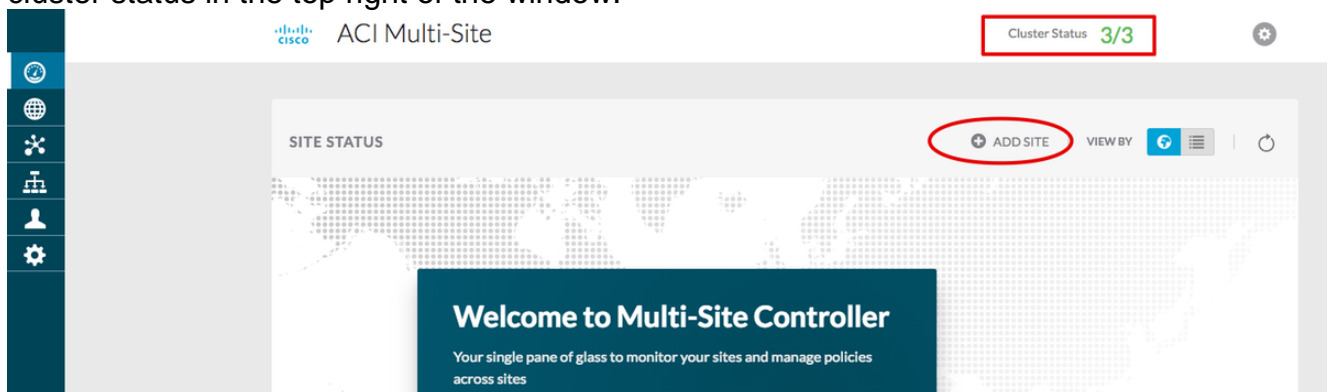

4. Repeat the previous steps in order to complete the APIC side configuration for SiteB ACI fabric.
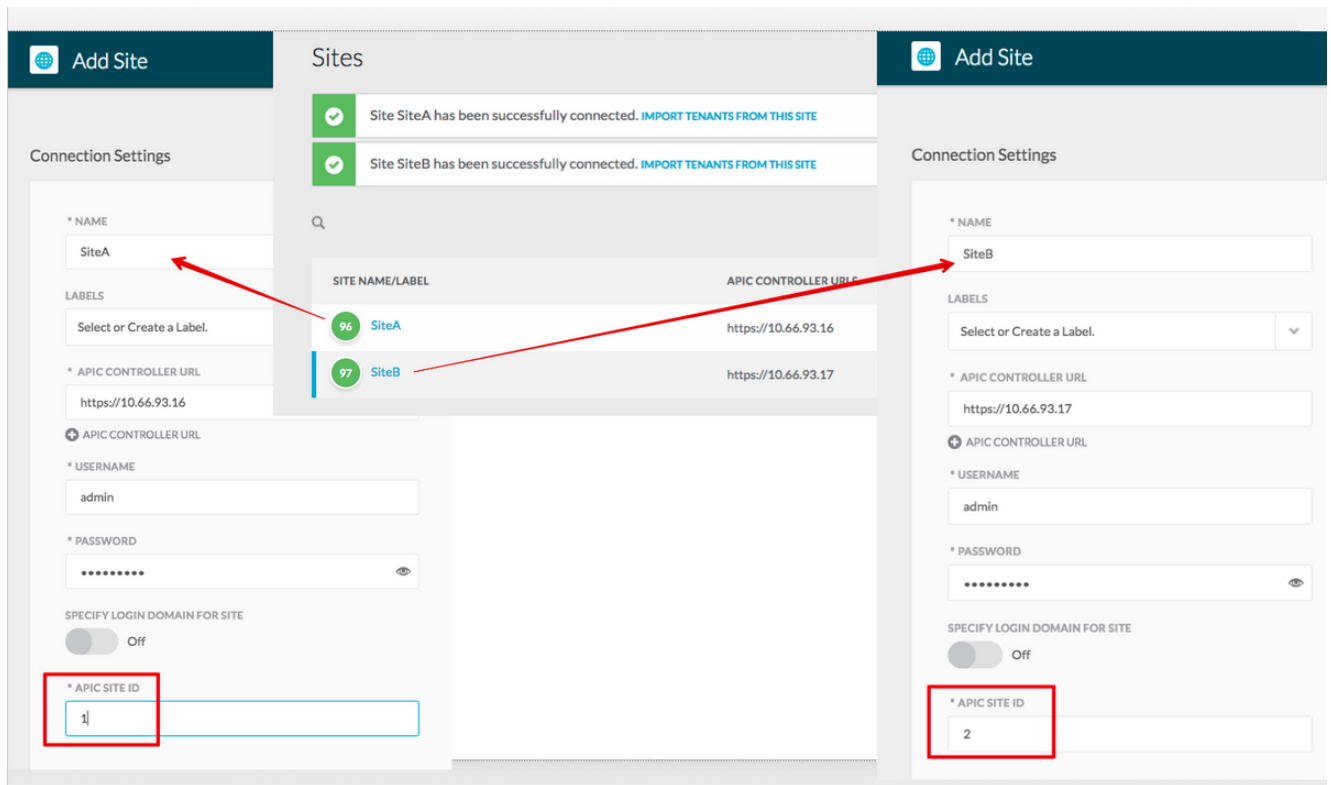
Multi-Site Controller Configuration

1. Add each site one by one in the MSC GUI. Connect and log in to the MSC GUI.
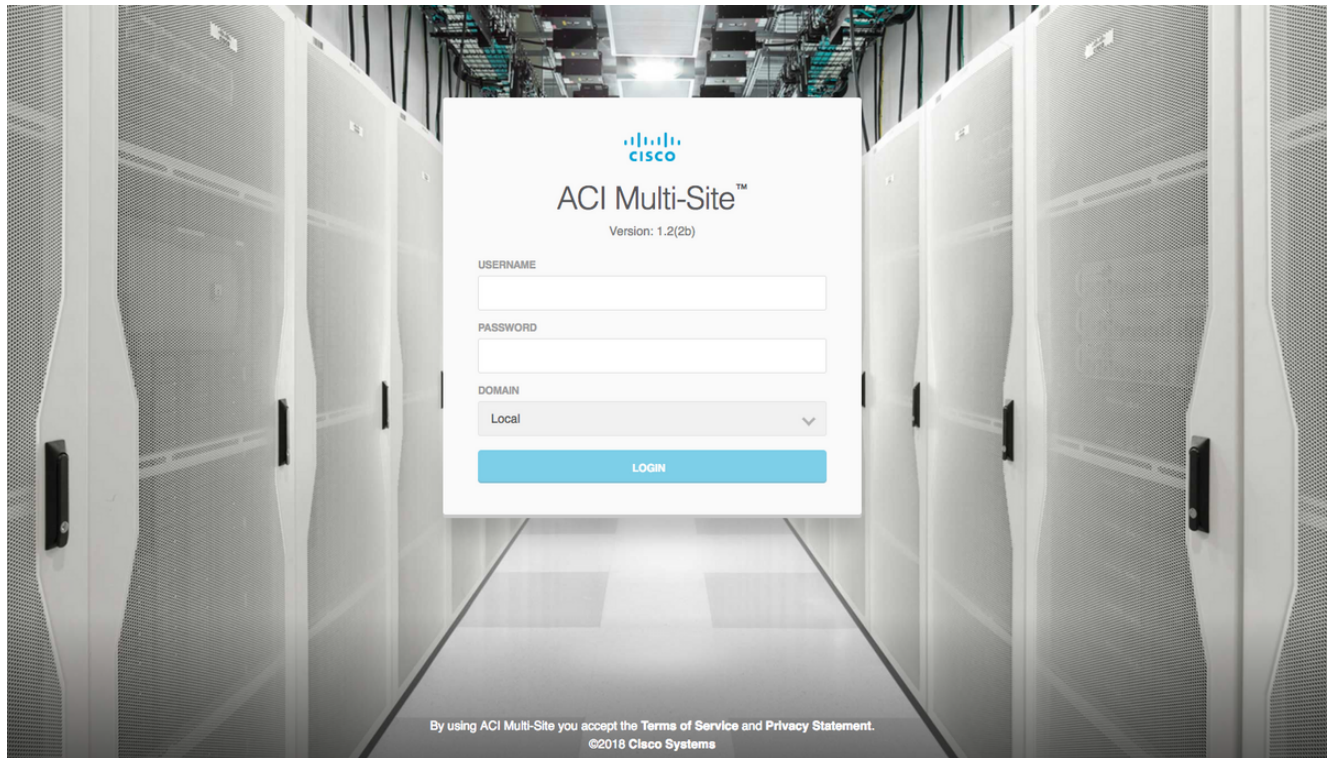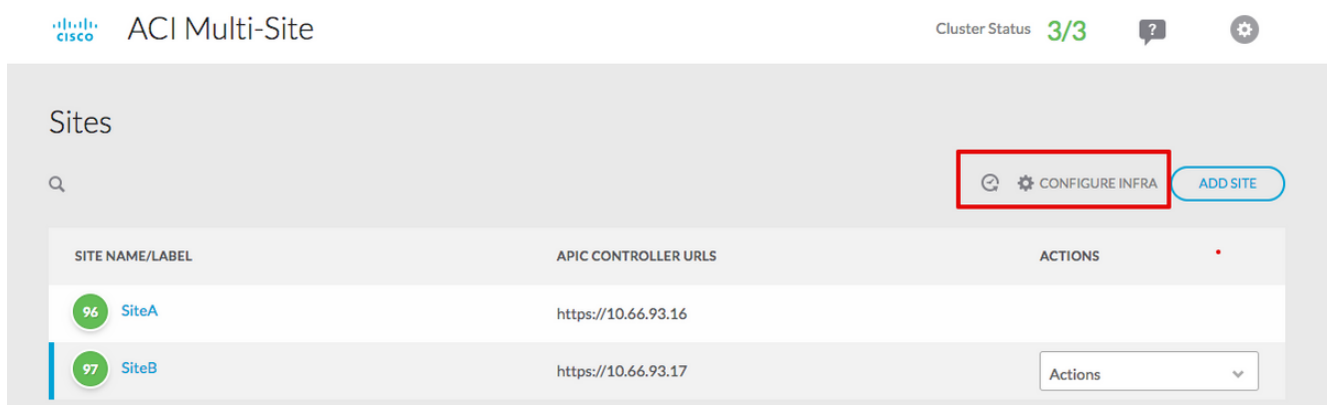
Click **ADD SITE** in order to register the sites one-by-one in MSC. You can also see the cluster status in the top right of the window.


Use one of the APIC's IP addresses and assign one **unique site ID** for each site. The valid range is 1-127.

2. Configure the infra policies per site in MSC. Log in to the MSC GUI. Choose **Sites** from the left pane and then click **CONFIGURE INFRA**.

Configure the Fabric Infra General settings. From the BGP Peering Type drop down list, choose **full-mesh** (full mesh - EBGP /route reflector - IBGP).
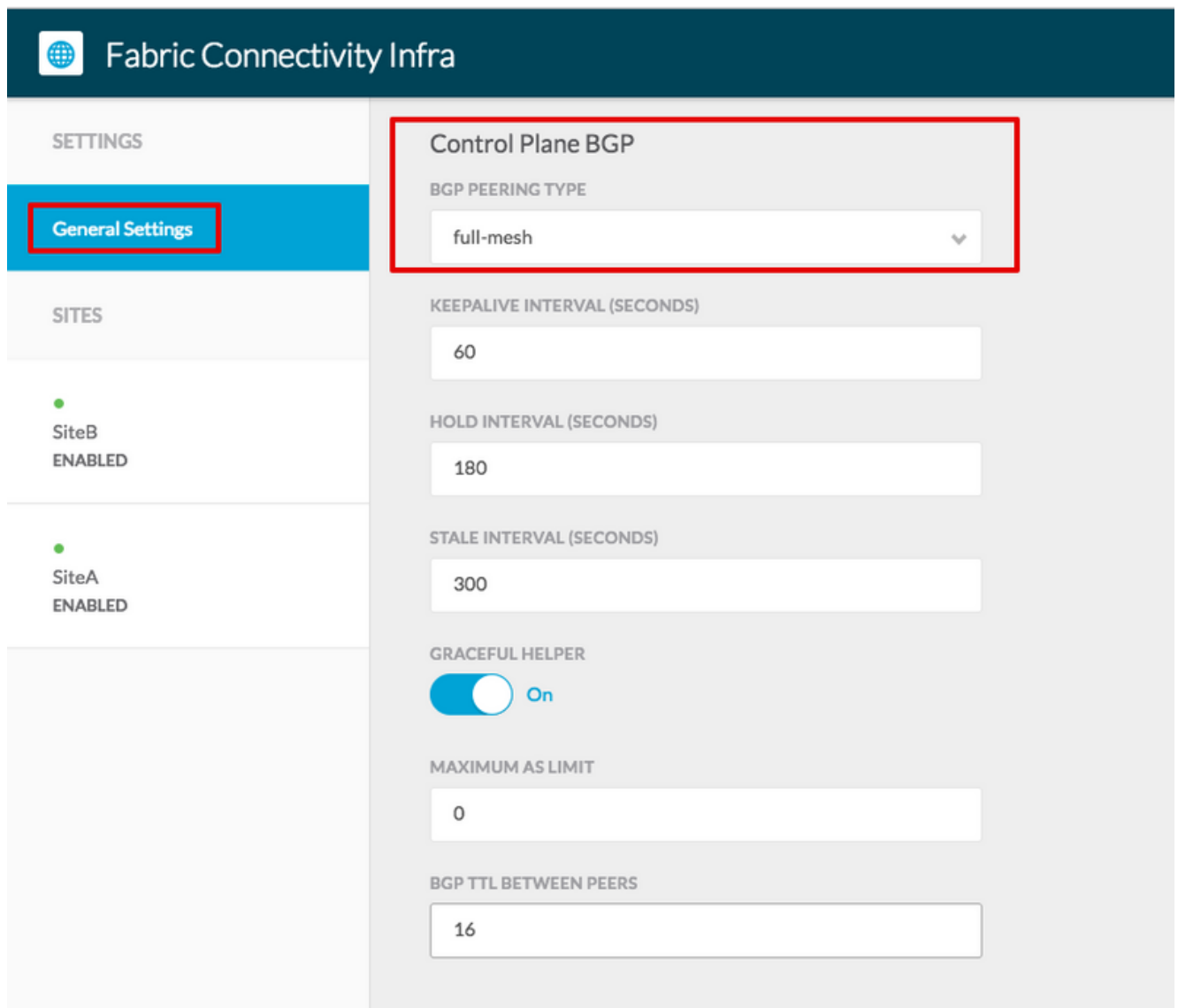
Once it is complete, choose one of the sites from the left pane. Then you will see site information in the middle pane. There are three different configuration levels. You can choose the Site level, the Pod level, or the Spine level. It will allow different settings on the configuration panel (right pane).
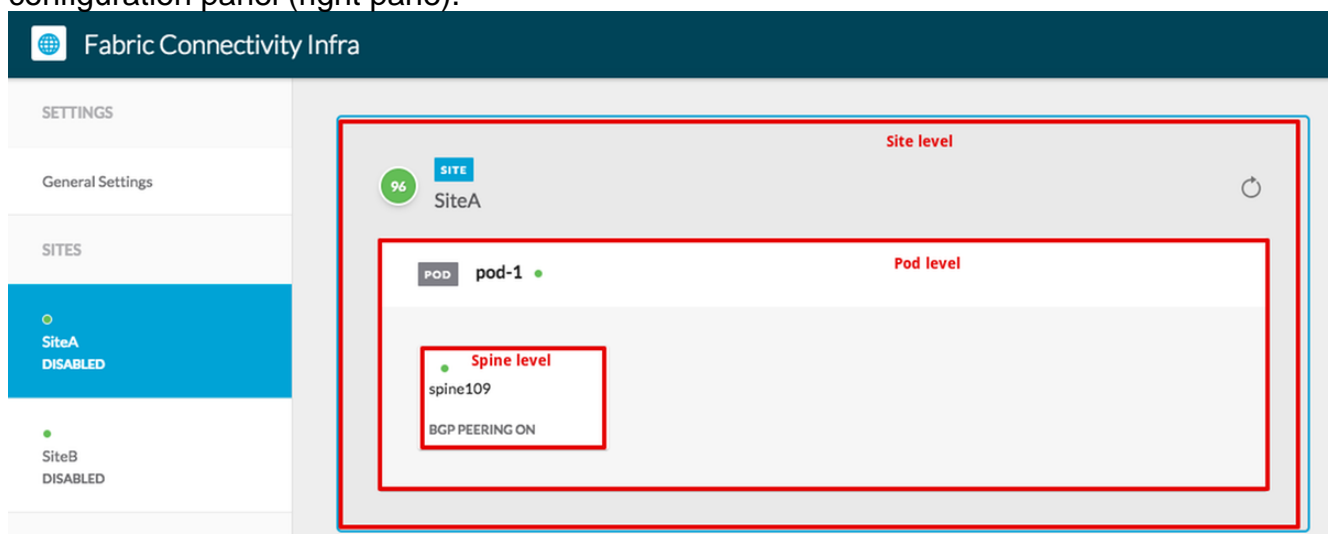

Once you click on the Site area, the site level configurations (Multi-Site Enable (On), Dataplane Multicast TEP, BGP ASN, BGP Community (for example, extended:as2-nn4:2:22) , OSPF Area ID, OSPF Area Type (stub prevent tep pool advertising), External Route Domain, and so on) will display in the right pane. Here, you can configure or modify:

Dataplane Multicast TEP (one loopback per site), used for Headend Replication (HREP)Border Gateway Protocol (BGP) Autonomous System (AS) (matching AS from the site configured in APIC)OSPF Area ID, OSPF Area Type, and OSPF Interface Policy (for spine interface towards IPN)External Routed DomainIn most cases, the attribute values would have already been retrieved automatically from APIC to MSC.
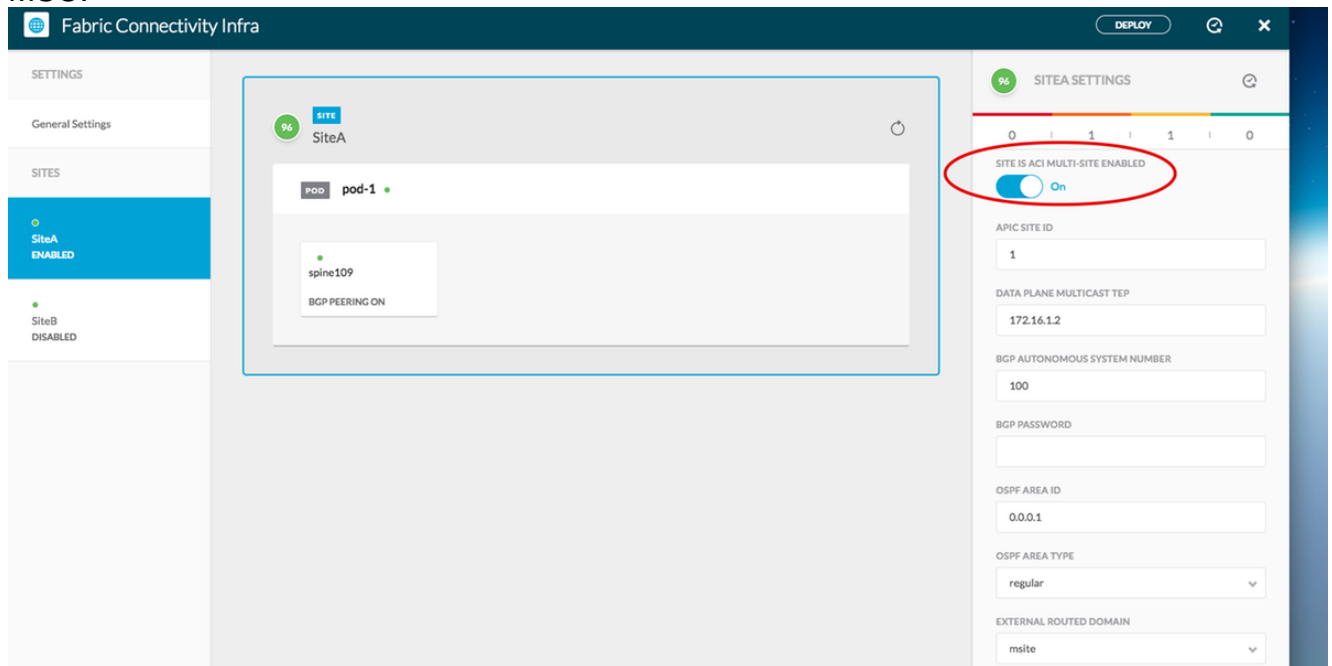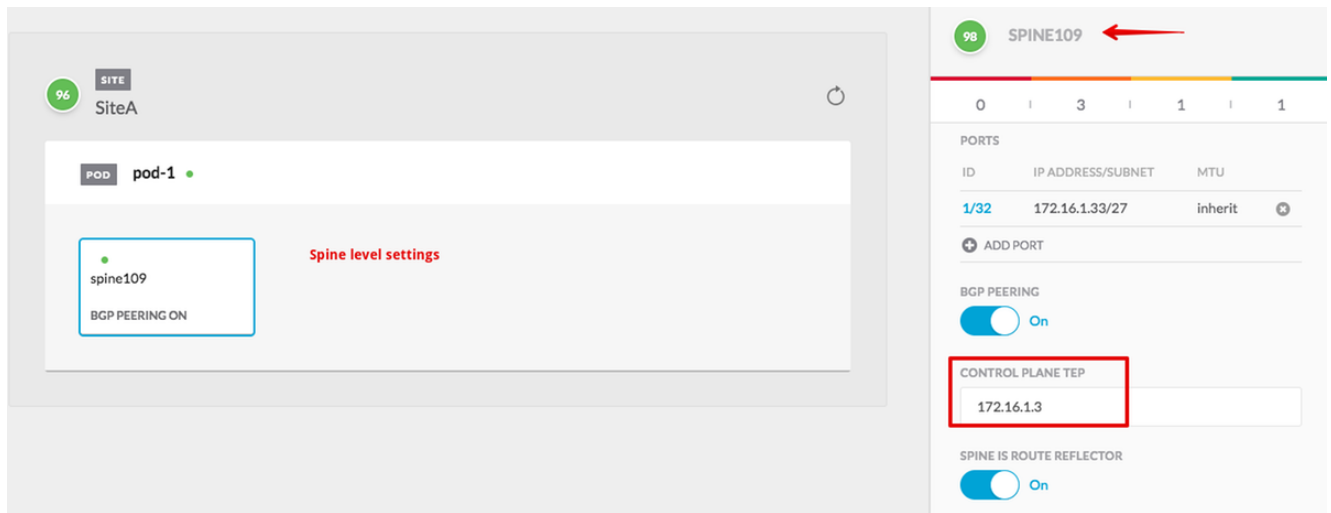

Click the Pod area and go to the POD level specific policies. Enter the Data Plane Unicast TEP.
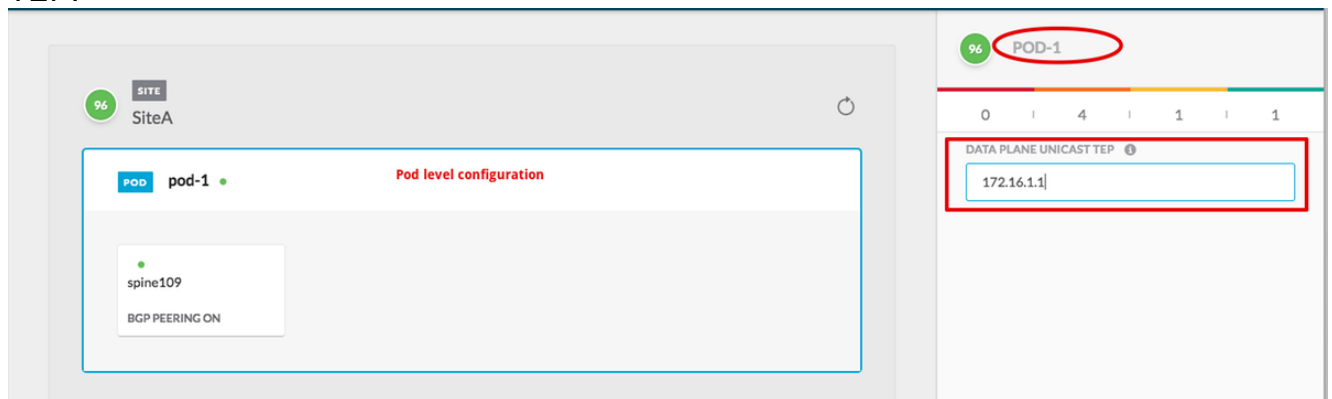
Click the Spine area and go to the spine specific infra settings. For each interface from the spine towards the IPN switch:

Set the IP address and mask
BGP Peering - On
Control Plane TEP - enter the Router IP address
Spine is Route Reflector -
On

Repeat these steps for other sites and complete the infra configuration in MSC.Click **DEPLOY**. This will save and push the infra configuration to APICs in both sites.
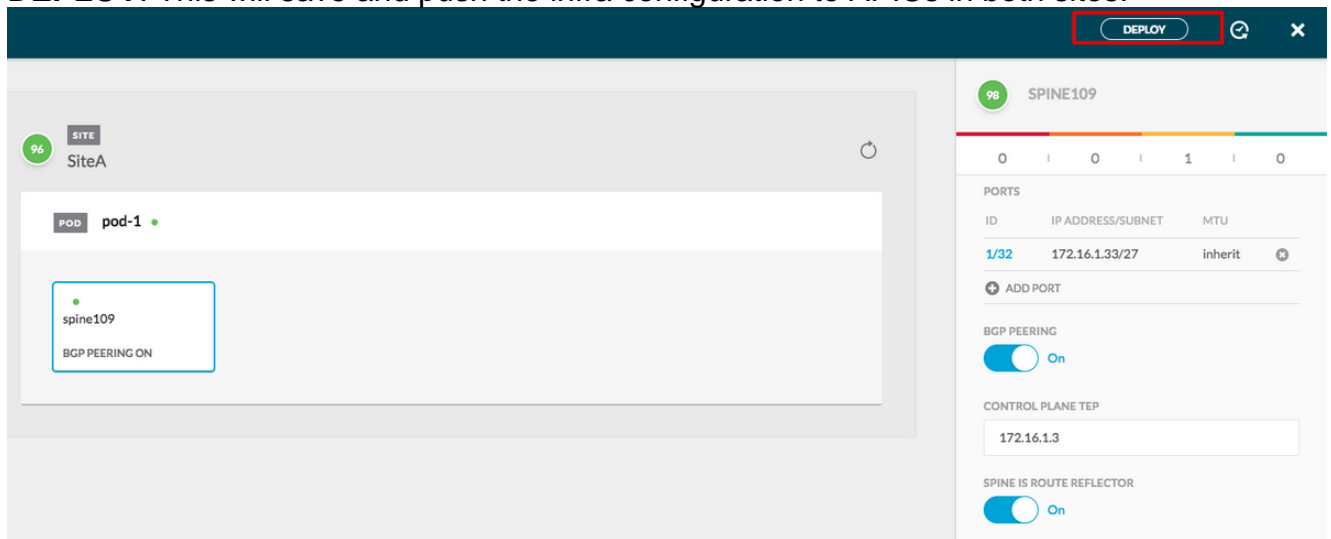
The initial integration between APIC clusters and MSC is complete and ready to use.

You should be able to configure stretched policies for tenants on MSC for different ACI sites.

Verify

Use this section in order to confirm that your configuration works properly.

1. Verify the infra configuration from the APIC GUI on each APIC cluster. Verify the Intrasite/Intersite profile was configured under infra tenant on each APIC cluster.Verify the infra L3Out (intersite), OSPF, and BGP was configured on each APIC cluster (APIC GUI).Log in the site's APIC and verify the Intrasite/Intersite Profile under **Tenant infra> Policies > Protocol > Fabric Ext Connection Policies** . The Intersite profile will look like this when the site is fully configured/managed by MSC.

Choose **APIC GUI > Tenant Infra > Networking > External Routed Networks**. Here the intersite L3Out profile should be created automatically under tenant infra in both sites.

Also, ensure the L3Out logical node and interface profile configuration is correctly set in VLAN
4.

Logical Node Profile - node-109-profile

Policy   Faults

⊗ ▽ △ ◔   ↻ ↓

Properties

Name: node-109-profile

Description: optional

Alias:

Target DSCP: Unspecified ∨

Nodes:

▲ Node ID	Router ID	Static Routes	Loopback Address
topology/pod-1/node-109	172.16.1.3		

BGP Infra Peer
Connectivity:

Peer IP Address	Time To Live
172.16.2.3	16

Logical Interface Profile - interface-109-1-32-profile   ⭐ ❓

Policy   Faults   History

General   Routed Sub-Interfaces   Routed Interfaces   SVI

⊗ ▽ △ ◔   ↻ ↓

Properties

Routed Sub-Interfaces: 🗑 +

▲ Path	IP Address	Secondary IP Address	MAC Address	MTU (bytes)	Encap
Pod-1/Node-109/eth1/32	172.16.1.33/27		00:22:BD:F8:19:FF	inherit	vlan-4

2. Verify the OSPF/BGP session from the Spine CLI on each APIC cluster. Verify OSPF is up on spine and gets routes from the IPN (Spine CLI).Verify the BGP session is up to the remote site (Spine CLI).Log in to the Spine CLI, verify the BGP L2VPN EVPN and OSPF is up on each spine. Also verify the node-role for BGP is msite-speaker.

```
spine109# show ip ospf neighbors vrf overlay-1
 OSPF Process ID default VRF overlay-1
 Total number of neighbors: 1
 Neighbor ID     Pri State           Up Time  Address        Interface
 172.16.1.34       1 FULL/ -         04:13:07 172.16.1.34    Eth1/32.32
spine109#

spine109# show bgp l2vpn evpn summary vrf overlay-1
BGP summary information for VRF overlay-1, address family L2VPN EVPN
BGP router identifier 172.16.1.3, local AS number 100
BGP table version is 235, L2VPN EVPN config peers 1, capable peers 1
0 network entries and 0 paths using 0 bytes of memory
BGP attribute entries [0/0], BGP AS path entries [0/0]
BGP community entries [0/0], BGP clusterlist entries [0/0]

Neighbor V AS MsgRcvd MsgSent TblVer InQ OutQ Up/Down State/PfxRcd
172.16.2.3 4 200 259 259 235 0 0 04:15:39 0

spine109#
spine109# vsh -c 'show bgp internal node-role'
Node role : : MSITE_SPEAKER


spine209# show ip ospf neighbors vrf overlay-1
 OSPF Process ID default VRF overlay-1
 Total number of neighbors: 1
 Neighbor ID     Pri State           Up Time  Address        Interface
 172.16.1.34       1 FULL/ -         04:20:36 172.16.2.34    Eth1/32.32
spine209#
spine209# show bgp l2vpn evpn summary vrf overlay-1
```

```
BGP summary information for VRF overlay-1, address family L2VPN EVPN
BGP router identifier 172.16.2.3, local AS number 200
BGP table version is 270, L2VPN EVPN config peers 1, capable peers 1
0 network entries and 0 paths using 0 bytes of memory
BGP attribute entries [0/0], BGP AS path entries [0/0]
BGP community entries [0/0], BGP clusterlist entries [0/0]

Neighbor        V    AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
172.16.1.3      4   100    264     264      270    0    0 04:20:40 0

spine209#
spine209# vsh -c 'show bgp internal node-role'
Node role :                        : MSITE_SPEAKER
```

3. Verify Overlay-1 interfaces from the Spine CLI on each APIC cluster. Log in to the Spine CLI to check and verify Overlay-1 interfaces.**ETEP (Multipod Dataplane TEP)**The Dataplane Tunnel Endpoint address used to route traffic between multiple Pods within the single ACI fabric.**DCI-UCAST (Intersite Dataplane unicast ETEP (anycast per site))**This anycast dataplane ETEP address is unique per site. It is assigned to all the spines connected to the IPN/ISN device and used to receive L2/L3 unicast traffic.**DCI-MCAST-HREP (Intersite Dataplane multicast TEP)**This anycast ETEP address is assigned to all the spines connected to the IPN/ISN device and used to receive L2 BUM (Broadcast, Unknown unicast and Multicast) traffic.**MSCP-ETEP (Multi-Site Control-plane ETEP)**This is the control plane ETEP address, which is also known as BGP Router ID on each spine for MP-BGP EVPN.

```
spine109# show ip int vrf overlay-1
<snip>
lo17, Interface status: protocol-up/link-up/admin-up, iod: 83, mode: etep
  IP address: 172.16.1.4, IP subnet: 172.16.1.4/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 1, tag: 0
lo18, Interface status: protocol-up/link-up/admin-up, iod: 84, mode: dci-ucast
  IP address: 172.16.1.1, IP subnet: 172.16.1.1/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 1, tag: 0
lo19, Interface status: protocol-up/link-up/admin-up, iod: 85, mode: dci-mcast-hrep
  IP address: 172.16.1.2, IP subnet: 172.16.1.2/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 1, tag: 0
lo20, Interface status: protocol-up/link-up/admin-up, iod: 87, mode: mscp-etep
  IP address: 172.16.1.3, IP subnet: 172.16.1.3/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 1, tag: 0


spine209# show ip int vrf overlay-1
<snip>
lo13, Interface status: protocol-up/link-up/admin-up, iod: 83, mode: etep
  IP address: 172.16.2.4, IP subnet: 172.16.2.4/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 1, tag: 0
lo14, Interface status: protocol-up/link-up/admin-up, iod: 84, mode: dci-ucast
  IP address: 172.16.2.1, IP subnet: 172.16.2.1/32
  IP broadcast address: 255.255.255.255
```

```
  IP primary address route-preference: 1, tag: 0
lo15, Interface status: protocol-up/link-up/admin-up, iod: 85, mode: dci-mcast-hrep
  IP address: 172.16.2.2, IP subnet: 172.16.2.2/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 1, tag: 0
lo16, Interface status: protocol-up/link-up/admin-up, iod: 87, mode: mscp-etep
  IP address: 172.16.2.3, IP subnet: 172.16.2.3/32
  IP broadcast address: 255.255.255.255
  IP primary address route-preference: 1, tag: 0
```

At the end, ensure no faults are seen from the MSC.**Troubleshoot**There is currently no specific troubleshooting information available for this configuration.**Related Information****[Cisco ACI Multi-Site Architecture White Paper](#)[Technical Support & Documentation - Cisco Systems](#)**