ACI Advanced Monitoring and Troubleshooting

Sadiq Memon, CCIE® No. 47508
Joseph Ristaino, CCIE® No. 41799
Carlo Schmidt, CCIE® No. 41842

Forewords written by **Yusuf Bhaiji,** Director of Certifications, Cisco Systems;
and **Ronak Desai,** VP of Engineering for the Data Center
Networking Business Unit, Cisco Systems

ciscopress.com

Contents

Chapter 6

VMM Integration

Cisco ACI virtual machine (VM) networking supports hypervisors from multiple vendors. It allows for multivendor hypervisors along with programmable and automated access to high-performance scalable virtualized data center infrastructure. In this chapter, you will learn about Virtual Machine Manager (VMM) and its integration into Cisco Application Centric Infrastructure (ACI) from the following virtualization-supported products and vendors:

- VMware
- Microsoft
- OpenStack
- Kubernetes
- OpenShift

You will also learn about VMM integration with ACI at multiple locations.

Virtual Machine Manager (VMM)

VMM integration enables the ACI fabric to extend network policies and policy group definitions into the virtualization switching layer on end hosts. This integration automates critical network plumbing steps that typically create delays in the deployment of overall virtual and compute resources in legacy network environments. VMM integration into ACI also provides value in getting visibility up to the virtualization layer of the application, which is a perpetually conflicting factor between network and server virtualization teams.

VMM Domain Policy Model

VMM domain profiles (vmmDomP) specify connectivity policies that enable virtual machine controllers to connect to the ACI fabric. Figure 6-1 shows the general hierarchy of VMM configuration.

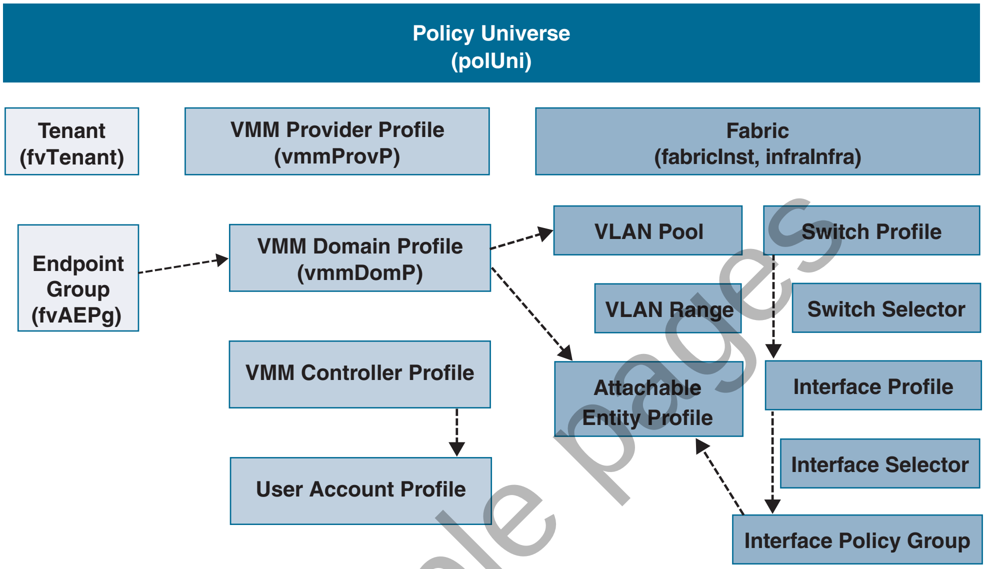

Figure 6-1 *VMM Policy Model*

VMM Domain Components

VMM domains enable an administrator to configure connectivity policies for virtual machine controllers in ACI. The essential components of an ACI VMM domain policy include the following:

- VMM domain
- VLAN pool association
- Attachable access entity profile association
- VMM domain endpoint group (EPG) association

VMM Domains

VMM domains make it possible to group VM controllers with similar networking policy requirements. For example, VM controllers can share VLAN pools and application EPGs. The Cisco Application Policy Infrastructure Controller (APIC) communicates with the VM controller to publish network configurations such as port groups, which are then applied to the virtual workloads. The VMM domain profile includes the following essential components:

- **Credential:** Associates a valid VM controller user credential with an APIC VMM domain.
- **Controller:** Specifies how to connect to a VM controller that is part of a policy enforcement domain. For example, the controller specifies the connection to a VMware vCenter instance that is part of a VMM domain.

> **Note** A single VMM domain can contain multiple instances of VM controllers, but they must be from the same vendor (for example, VMware, Microsoft).

An APIC VMM domain profile is a policy that defines a VMM domain. The VMM domain policy is created on an APIC and pushed into the leaf switches. Figure 6-2 illustrates VM controllers of the same vendor as part of the same VMM domain.

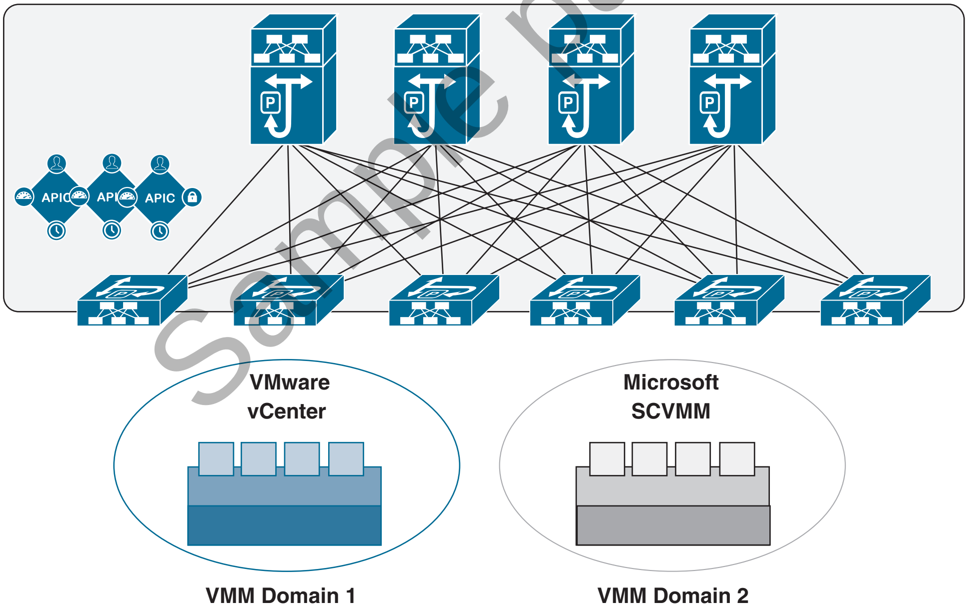

Figure 6-2 *VMM Domain Integration*

VMM domains provide the following:

- A common layer in the ACI fabric that enables scalable fault-tolerant support for multiple VM controller platforms.
- VMM support for multiple tenants within the ACI fabric.

VMM domains contain VM controllers such as VMware vCenter or Microsoft System Center Virtual Machine Manager (SCVMM) and the credentials required for the ACI API to interact with the VM controllers. A VMM domain enables VM mobility within the domain but not across domains. A single VMM domain can contain multiple instances of VM controllers, but they must be from the same vendor. For example, a VMM domain can contain many VMware vCenter instances managing multiple controllers, each running multiple VMs; however, it cannot contain Microsoft SCVMM instances. A VMM domain inventories controller elements (such as pNICs, vNICs, and VM names) and pushes policies into the controllers, creating port groups or VM networks and other necessary elements. The ACI VMM domain listens for controller events such as VM mobility events and responds accordingly.

VMM Domain VLAN Pool Association

A VLAN pool specifies a single VLAN ID or a range of VLAN IDs for VLAN encapsulation. It is a shared resource that can be consumed by multiple domains, such as physical, VMM, or external domains.

In ACI, you can create a VLAN pool with allocation type static or dynamic. With static allocation, the fabric administrator configures a VLAN; with dynamic allocation, the APIC assigns the VLAN to the domain dynamically. In ACI, only one VLAN or VXLAN pool can be assigned to a VMM domain.

A fabric administrator can assign a VLAN ID statically to an EPG. However, in this case, the VLAN ID must be included in the VLAN pool with the static allocation type, or the APIC will generate a fault. By default, the assignment of VLAN IDs to EPGs that are associated with the VMM domain is done dynamically by the APIC. The APIC provisions VMM domain VLAN IDs on leaf switch ports based on EPG events, either statically binding or based on VM events from controllers such as VMware vCenter or Microsoft SCVMM.

Attachable Access Entity Profile Association

An attachable access entity profile (AAEP) associates a VMM domain with the physical network infrastructure where the vSphere hosts are connected. The AAEP defines which VLANs will be permitted on a host-facing interface. When a domain is mapped to an endpoint group, the AAEP validates that the VLAN can be deployed on certain interfaces. An AAEP is a network interface template that enables the deployment of VM controller policies on a large set of leaf switch ports. An AAEP specifies which switches and ports are available and how they are configured. The AAEP can be created on-the-fly during the creation of the VMM domain itself.

VMM Domain EPG Association

Endpoint groups regulate connectivity and visibility among the endpoints within the scope of the VMM domain policy. VMM domain EPGs behave as follows:

- The APIC pushes these EPGs as port groups into the VM controller.
- An EPG can span multiple VMM domains, and a VMM domain can contain multiple EPGs.

The ACI fabric associates EPGs to VMM domains, either automatically through an orchestration component such as VMware vRealize suite (vRA/vRO) or Microsoft Azure, or when an APIC administrator creates such configurations. An EPG can span multiple VMM domains, and a VMM domain can contain multiple EPGs.

In Figure 6-3, endpoints (EPs) of the same color are part of the same EPG. For example, all the gray EPs are in the same EPG, even though they are in different VMM domains.

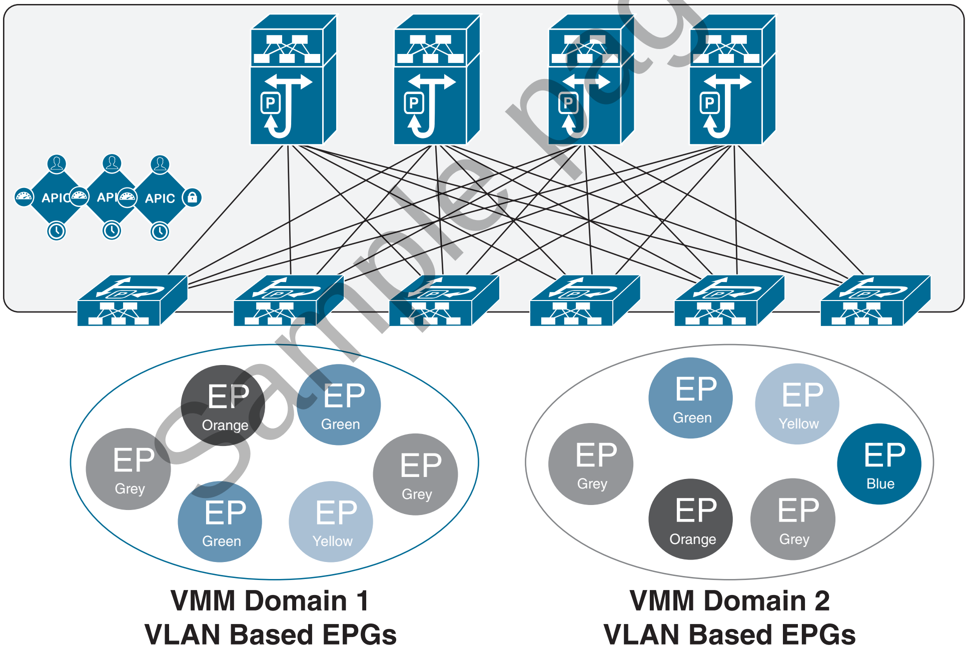

Figure 6-3 *VMM Domain EPG Association*

Note Refer to the latest *Verified Scalability Guide for Cisco ACI* at the Cisco website for virtual network and VMM domain EPG capacity information.

Figure 6-4 illustrates multiple VMM domains connecting to the same leaf switch if they do not have overlapping VLAN pools on the same port. Similarly, the same VLAN pools can be used across different domains if they do not use the same port of a leaf switch.

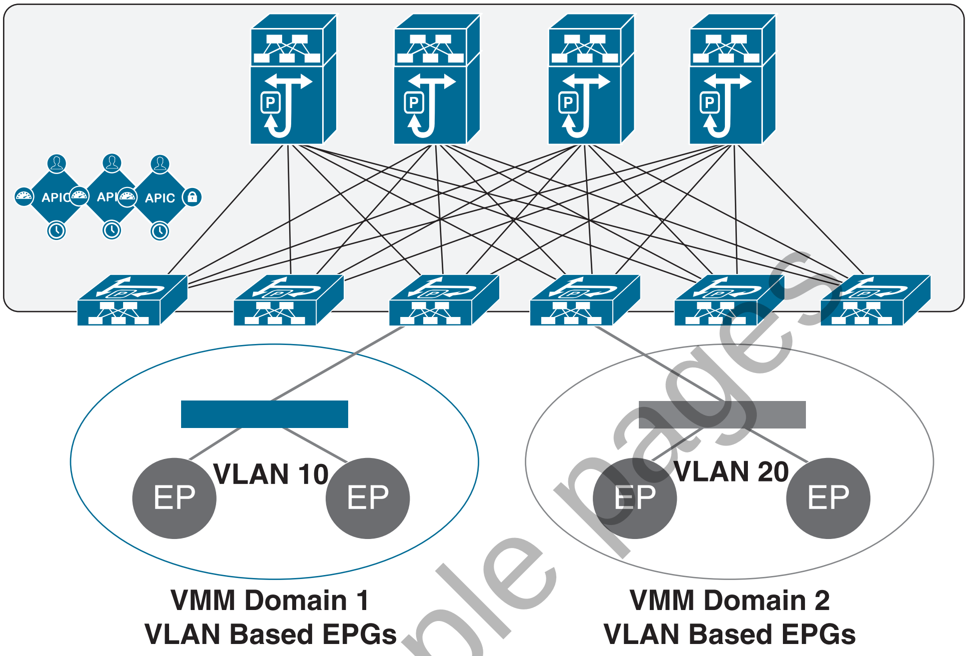

Figure 6-4 *VMM Domain EPG VLAN Consumption*

EPGs can use multiple VMM domains in the following ways:

- An EPG within a VMM domain is identified by an encapsulation identifier that is either automatically managed by the APIC or statically selected by the administrator. An example for a VLAN is a virtual network ID (VNID).
- An EPG can be mapped to multiple physical (for bare-metal servers) or virtual domains. It can use different VLAN or VNID encapsulations in each domain.

Note By default, an APIC dynamically manages the allocation of a VLAN for an EPG in a VMM integration. VMware vSphere Distributed Switch (VDS) administrators have the option of configuring a specific VLAN for an EPG. In that case, the VLAN is chosen from a static allocation block within the pool associated with the VMM domain.

Applications can be deployed across VMM domains, as illustrated in Figure 6-5. While live migration of VMs within a VMM domain is supported, live migration of VMs across VMM domains is not supported.

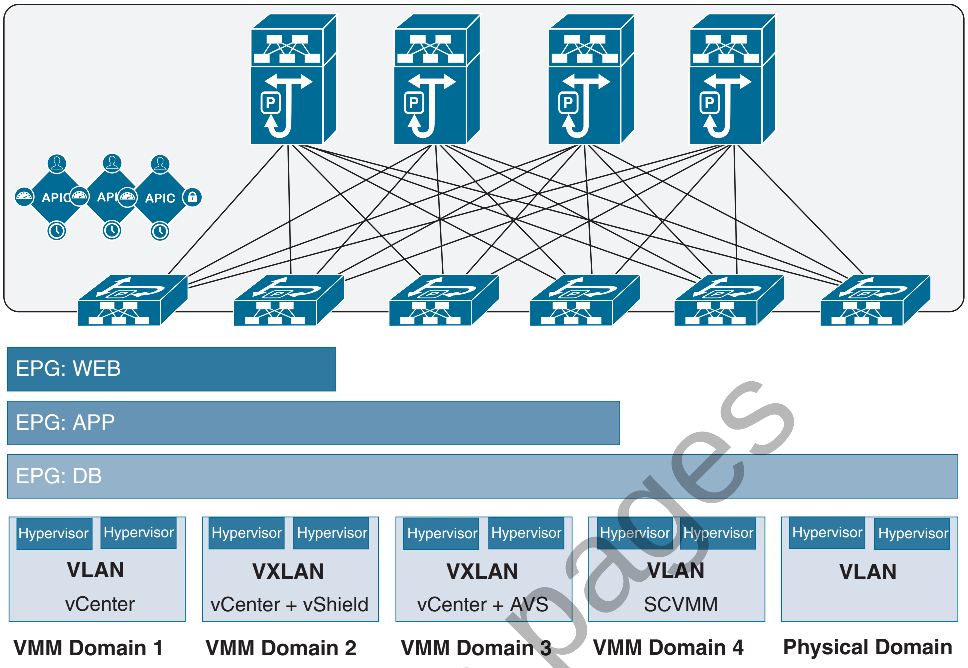

Figure 6-5 *Multiple VMM Domains and Scaling of EPGs in the ACI Fabric*

EPG Policy Resolution and Deployment Immediacy

Whenever an EPG associates to a VMM domain, the administrator can choose the policy resolution and deployment preferences to specify when it should be pushed and programmed into leaf switches. This approach provides efficient use of hardware resources because resources are consumed only when demanded. You should be aware of picking one option over the other, depending on the use case and scalability limits of your ACI infrastructure, as explained in the following sections.

Resolution Immediacy

The Resolution Immediacy option defines when policies are downloaded to the leaf software based on the following options:

- **Pre-provision:** This option specifies that a policy (such as VRF, VLAN, VXLAN binding, contracts, or filters) is downloaded to the associated leaf switch software even before a VM controller is attached to the distributed virtual switch (DVS), such as a VMware (VDS), defined by an APIC through the VMM domain.
    - This option helps when management traffic between hypervisors and VM controllers such as VMware vCenter is also using the APIC-defined virtual switch.
    - When you deploy a VMM policy such as VLAN or VXLAN on an ACI leaf switch, an APIC must collect CDP/LLDP information from hypervisors through

    the VM controller and ACI leaf switch to which the host is connected. However, if the VM controller is supposed to use the same VMM policy to communicate with its hypervisors or even an APIC, the CDP/LLDP information for hypervisors can never be collected because the required policy is not deployed yet.

    - With the Pre-provision immediacy option, policy is downloaded to the ACI leaf switch software, regardless of CDP/LLDP neighborship and even without a hypervisor host connected to the VMM domain-defined DVS.

- **Immediate:** This option specifies that a policy (such as VRF, VLAN, VXLAN binding, contracts, or filters) is downloaded to the associated leaf switch software upon ESXi host attachment to a DVS. LLDP or OpFlex permissions are used to resolve the VM controller to leaf switch attachments.
    - The policy is downloaded to a leaf when you add a host to the VMM domain-defined DVS. CDP/LLDP neighborship from host to leaf is required.
- **On Demand:** This option specifies that a policy (such as VRF, VLAN, VXLAN binding, contracts, or filters) is pushed to the leaf node only when a host running hypervisor is attached to a DVS and a VM is placed in the port group (EPG).
    - The policy is downloaded to a leaf when a host is added to the VMM domain-defined DVS and a virtual machine is placed in the port group (EPG). CDP/LLDP neighborship from host to leaf is required.

With both the Immediate and On Demand options for resolution immediacy, if the hypervisor running on the host and leaf lose LLDP/CDP neighborship, the policies are removed from the leaf switch software.

Deployment Immediacy

After the policies are downloaded to the leaf software through the Resolution Immediacy option, you can use Deployment Immediacy to specify when the policy is pushed to the hardware policy content-addressable memory (CAM). Two options are available:

- **Immediate:** This option specifies that the policy is programmed into the hardware policy CAM as soon as the policy is downloaded in the leaf software. You should be aware of your ACI infrastructure scalability limits when choosing this option.
- **On Demand:** This option specifies that the policy is programmed in the hardware policy CAM only when the first packet is received through the data path. This process helps optimize the hardware resources.

> **Note** When you use On Demand deployment immediacy with MAC-pinned VPCs, the EPG contracts are not pushed to the leaf ternary content-addressable memory (TCAM) until the first endpoint is learned in the EPG on each leaf. This can cause uneven TCAM utilization across VPC peers. (Normally, the contract would be pushed to both peers.)

VMware Integration

When integrating your VMware infrastructure into Cisco ACI, you have two options for deploying virtual networking:

- VMware vSphere Distributed Switch (VDS)
- Cisco Application Virtual Switch (AVS)

These two options provide similar basic virtual networking functionality; however, the AVS option provides additional capabilities, such as VXLAN and microsegmentation support.

Prerequisites for VMM Integration with AVS or VDS

The prerequisites for VMM integration with AVS or VDS are as follows:

- You need to decide whether to use VLAN or VXLAN encapsulation or multicast groups.
- A virtual machine manager must be already deployed, such as vCenter.
- The VMM must be accessible by the APIC through either the out-of-band or in-band management network.
- For Cisco AVS deployment, a vSphere Installation Bundle (VIB) must be installed on all hypervisor hosts to be added to the AVS.
- For a VXLAN deployment, you need to know whether intermediate devices have Internet Group Management Protocol (IGMP) snooping on or off by default.

Guidelines and Limitations for VMM Integration with AVS or VDS

The guidelines and limitations for VMM integration with AVS or VDS are as follows:

- When utilizing VLANs for VMM integration, whether with Cisco AVS or VMware VDS, the range of VLANs to be used for port groups must be manually allowed on any intermediate devices.
- For VMM integration with VLANs and the Resolution Immediacy setting On Demand or Immediate, there can be a maximum of one hop between a host and the compute node.
- For VMM integration with VXLAN, only the infrastructure VLAN needs to be allowed on all intermediate devices.
- For VMM integration with VXLAN, if the *Infra* bridge domain subnet is set as a querier, the intermediate devices must have IGMP snooping enabled for traffic to pass properly.

- To log in to the APIC GUI, choose Tenants > *Infra* > Networking > Bridge Domains > default > Subnets > 10.0.0.30/27.
- For VMM integration with VXLAN and UCS-B, IGMP snooping is enabled on the UCS-B by default. Therefore, you need to ensure that the querier IP address is enabled for the *Infra* bridge domain. The other option is to disable IGMP snooping on the UCS and disable the querier IP address on the *Infra* bridge domain.

ACI VMM Integration Workflow

Figure 6-6 illustrates the ACI VMM integration workflow steps.

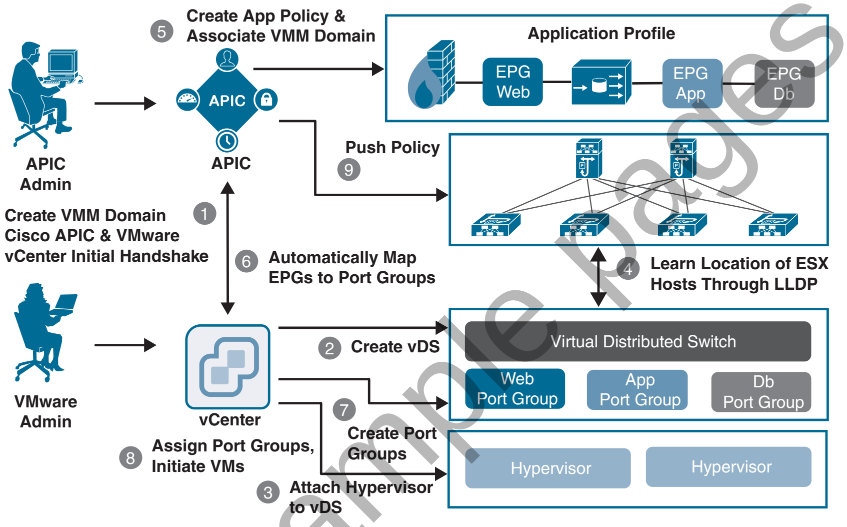
5 Create App Policy & Associate VMM Domain
Application Profile
EPG Web
EPG App
EPG Db
APIC
APIC Admin
APIC
Push Policy
9
Create VMM Domain Cisco APIC & VMware vCenter Initial Handshake
1
6 Automatically Map EPGs to Port Groups
4 Learn Location of ESX Hosts Through LLDP
2 Create vDS
Virtual Distributed Switch
Web Port Group
App Port Group
Db Port Group
VMware Admin
vCenter
7 Create Port Groups
8 Assign Port Groups, Initiate VMs
3 Attach Hypervisor to vDS
Hypervisor
Hypervisor

Figure 6-6 *ACI VMM Integration Workflow*

Publishing EPGs to a VMM Domain

This section details how to publish an existing EPG to a VMM domain. For an EPG to be pushed to a VMM domain, you must create a domain binding within the tenant EPG by following these steps:

Step 1. From the menu bar, choose Tenants > All Tenants.

Step 2. From the Work pane, choose the *Tenant_Name*.

Step 3. From the Navigation pane, choose *Tenant_Name* > Application Profiles > *Application_Profile_Name* > Application EPGs > *Application_EPG_Name* > Domains (VMs and bare-metal servers).

Step 4. From the Work pane, choose Actions > Add VM Domain Association.

Step 5. In the Add VM Domain Association dialog box, choose the VMM domain profile that you created previously. For Deployment and Resolution

Immediacy, Cisco recommends keeping the default option, On Demand. This provides the best resource usage in the fabric by deploying policies to leaf nodes only when endpoints assigned to this EPG are connected. There is no communication delay or traffic loss when you keep the default selections.

Step 6. Click Submit. The EPG is now available as a port group to your VMM.

Connecting Virtual Machines to the Endpoint Group Port Groups on vCenter

To connect virtual machines to the endpoint group port groups on vCenter, do the following:

Step 1. Connect to vCenter by using the VMware VI Client.

Step 2. From the Host and Clusters view, right-click on your virtual machine and choose Edit Settings.

Step 3. Click on the network adapter and from the Network Connection drop-down box, choose the port group that corresponds to your EPG. It should appear in the format of TENANT | *APPLICATION_PROFILE* | EPG | *VMM_DOMAIN_PROFILE.*

If you do not see your Cisco ACI EPG in the Network Connection list, it means one of the following:

- The VM is running on a host that is not attached to the distributed switch managed by the APIC.
- There may be a communication between your APIC and vCenter either through the OOB or the INB management network.

Verifying VMM Integration with the AVS or VDS

The following sections describe how to verify that the Cisco AVS has been installed on the VMware ESXi hypervisor.

Verifying the Virtual Switch Status

To verify the virtual switch status, follow these steps:

Step 1. Log in to the VMware vSphere client.

Step 2. Choose Networking.

Step 3. Open the folder for the data center and click the virtual switch.

Step 4. Click the Hosts tab. The VDS Status and Status fields display the virtual switch status. Ensure that the VDS status is Up, which indicates that OpFlex communication has been established.

Verifying the vNIC Status

To verify the vNIC status, follow these steps:

Step 1. In the VMware vSphere client, click the Home tab.

Step 2. Choose Hosts and Clusters.

Step 3. Click the host.

Step 4. In the Configuration tab, select the Hardware panel and choose Networking.

Step 5. In the View field, click the vSphere Distributed Switch button.

Step 6. Click Manage Virtual Adapters. The vmk1 displays as a virtual adapter with an IP address.

Step 7. Click the newly created vmk interface to display the vmknic status.

Note Allow approximately 20 seconds for the vmk to receive an IP address through DHCP.

Microsoft SCVMM Integration

Figure 6-7 shows a representative topology for a Microsoft SCVMM integration with Cisco ACI. Hyper-V clustering connectivity between SCVMM virtual machines and the APIC can run over the management network.

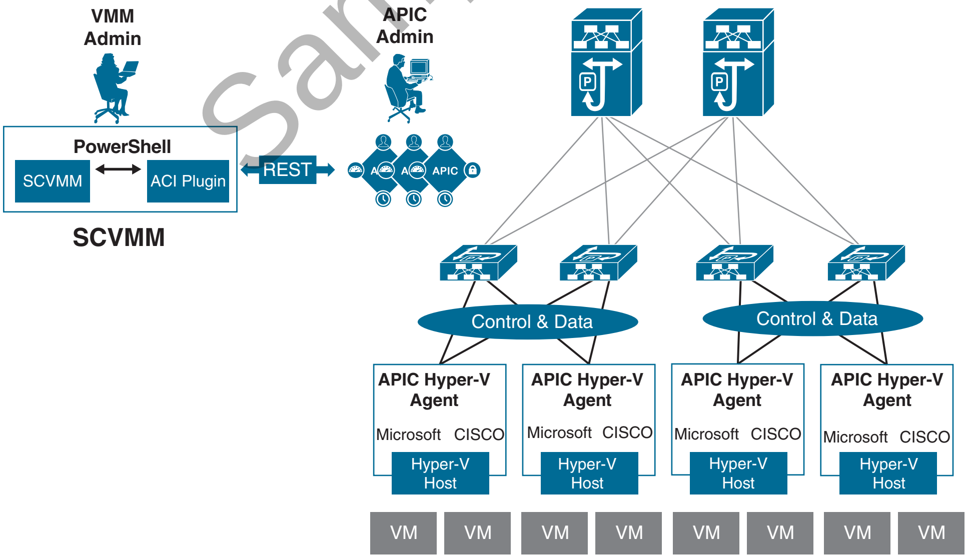

Figure 6-7 *Microsoft SCVMM Topology with ACI*

Figure 6-8 illustrates the workflow for integrating Microsoft SCVMM with Cisco ACI. The following sections describe the steps in this workflow.

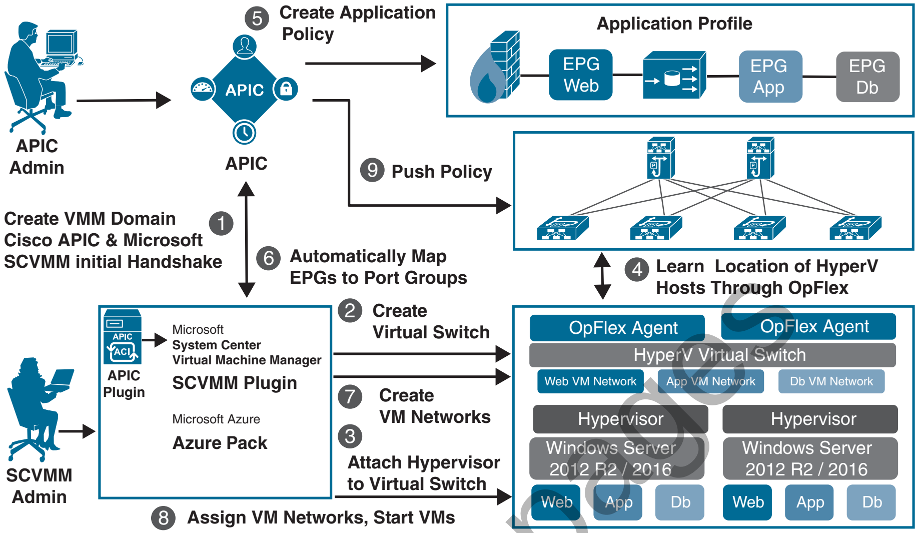
5 Create Application Policy
Application Profile
EPG Web
EPG App
EPG Db
APIC Admin
APIC
APIC
9 Push Policy
Create VMM Domain Cisco APIC & Microsoft SCVMM initial Handshake 1
6 Automatically Map EPGs to Port Groups
4 Learn Location of HyperV Hosts Through OpFlex
APIC ACI
APIC Plugin
Microsoft System Center Virtual Machine Manager
SCVMM Plugin
Microsoft Azure
Azure Pack
2 Create Virtual Switch
7 Create VM Networks
3 Attach Hypervisor to Virtual Switch
8 Assign VM Networks, Start VMs
SCVMM Admin
OpFlex Agent
OpFlex Agent
HyperV Virtual Switch
Web VM Network
App VM Network
Db VM Network
Hypervisor
Hypervisor
Windows Server 2012 R2 / 2016
Windows Server 2012 R2 / 2016
Web
App
Db
Web
App
Db

Figure 6-8 *Workflow for Integrating ACI and Microsoft SCVMM*

Mapping ACI and SCVMM Constructs

Figure 6-9 shows the mapping of Cisco ACI and the SCVMM constructs (SCVMM controller, cloud, and logical switches).

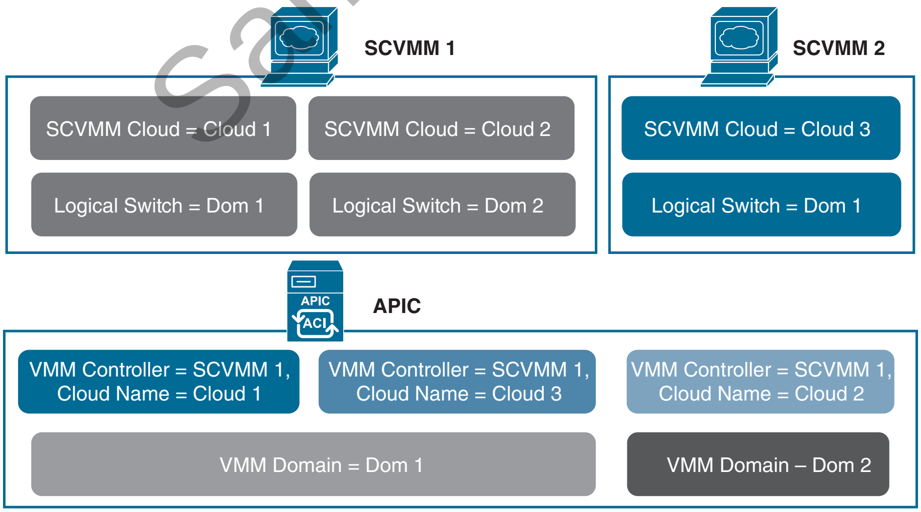

Figure 6-9 *Mapping ACI and SCVMM Constructs*

One VMM domain cannot map to the same SCVMM more than once. An APIC can be associated with up to five SCVMM controllers. For additional information on other limitations, see the *Verified Scalability Guide for Cisco ACI* on the Cisco website.

Mapping Multiple SCVMMs to an APIC

When multiple SCVMMs are associated with an APIC, the OpFlex certificate from the first SCVMM controller must be copied to the secondary controller and other controllers, as applicable. You use the **certlm.msc** command on the local SCVMM controller to import the certificate to the following location:

Certificates - Local Computer > Personal > Certificates

The same OpFlex certificate is deployed on the Hyper-V servers that are managed by this SCVMM controller. You use the **mmc** command to install the certificate on the Hyper-V servers.

Verifying That the OpFlex Certificate Is Deployed for a Connection from the SCVMM to the APIC

You can verify that the OpFlex certificate is deployed for a connection from the SCVMM to the APIC by viewing the Cisco_APIC_SCVMM_Service log file, which is located in the C:\Program Files (x86)\ApicVMMService\Logs\ directory. In this file, ensure that the correct certificate is used and also check to make sure there was a successful login to the APIC (see Example 6-1).

Example 6-1 *Viewing the Cisco_APIC_SCVMM_Service Log File*

```
4/15/2017 2:10:09 PM-1044-13||UpdateCredentials|| AdminSettingsController:
  UpdateCredentials.
4/15/2017 2:10:09 PM-1044-13||UpdateCredentials|| new: EndpointAddress:
  Called_from_SCVMMM_PS,
  Username ApicAddresses 10.10.10.1;10.10.10.2;10.10.10.3 CertName: OpflexAgent
4/15/2017 2:10:09 PM-1044-13||UpdateCredentials|| ########
4/15/2017 2:10:09 PM-1044-13||UpdateCredentials|| oldreg_apicAddresses is
4/15/2017 2:10:09 PM-1044-13||UpdateCredentials|| Verifying APIC address 10.10.10.1
4/15/2017 2:10:09 PM-1044-13||GetInfoFromApic|| Querying URL https://192.168.10.10/
  api/node/class/infraWiNode.xml
4/15/2017 2:10:09 PM-1044-13||GetInfoFromApic|| HostAddr 10.10.10.1
4/15/2017 2:10:09 PM-1044-13||PopulateCertsAndCookies|| URL:/api/node/class/
  infraWiNode.xml
4/15/2017 2:10:09 PM-1044-13||PopulateCertsAndCookies|| Searching Cached Store
  Name: My
4/15/2017 2:10:09 PM-1044-13||PopulateCertsAndCookies|| Using Certificate
  CN=OpflexAgent, C=USA, S=MI, O=CX, E=aci@lab.local in Cached Store Name:My
```

```
4/15/2017 2:10:09 PM-1044-13||PopulateCertsAndCookies|| Using the following CertDN:
  uni/userext/user-admin/usercert-OpFlexAgent
4/15/2017 2:10:09 PM-1044-13||GetInfoFromApic|| IFC returned OK to deployment query
4/15/2017 2:10:09 PM-1044-13||GetInfoFromApic|| Successfully deserialize deployment
  query response
4/15/2017 2:10:09 PM-1044-13||UpdateCredentials|| ApicClient.Login(addr 10.10.10.1)
  Success.
```

Verifying VMM Deployment from the APIC to the SCVMM

You can verify that the OpFlex certificate is deployed on the Hyper-V server by viewing log files in the C:\Program Files (x86)\ApicHyperAgent\Logs directory. In this file, ensure that the correct certificate is used and ensure that the connection with the Hyper-V servers on the fabric leafs is established. In addition, ensure that a VTEP virtual network adapter is added to the virtual switch and an IP address is assigned to the VTEP adapter.

In the SCVMM, check for the following:

- Under Fabric > Logical Switches, verify that apicVswitch_VMMdomainName is deployed from the APIC to the SCVMM.
- Under Fabric > Logical Networks, verify that apicLogicalNetwork_ VMMdomainName is deployed from the APIC to the SCVMM.
- Under Fabric > Port Profiles, verify that apicUplinkPortProfile_VMMdomainName is deployed. If it is not deployed, right-click the host under Servers and choose Properties. Go to Virtual Switches and ensure that the physical adapters are attached to the virtual switches.

Note In the APIC GUI, the Hyper-V servers and the virtual machines do not appear in the Microsoft SCVMM inventory until you ensure that these points for the SCVMM are satisfied.

OpenStack Integration

OpenStack defines a flexible software architecture for creating cloud-computing environments. The reference software-based implementation of OpenStack allows for multiple Layer 2 transports, including VLAN, GRE, and VXLAN. The Neutron project within OpenStack can also provide software-based Layer 3 forwarding. When OpenStack is used with ACI, the ACI fabric provides an integrated Layer 2/3 VXLAN-based overlay networking capability that can offload network encapsulation processing from the compute nodes to the top-of-rack or ACI leaf switches. This architecture provides the flexibility of software overlay networking in conjunction with the performance and operational benefits of hardware-based networking.

Extending OpFlex to the Compute Node

OpFlex is an open and extensible policy protocol designed to transfer declarative networking policies such as those used in Cisco ACI to other devices. By using OpFlex, you can extend the policy model native to ACI all the way down into the virtual switches running on OpenStack Nova compute hosts. This OpFlex extension to the compute host allows ACI to use Open vSwitch (OVS) to support common OpenStack features such as source Network Address Translation (SNAT) and floating IP addresses in a distributed manner.

The ACI OpenStack drivers support two distinct modes of deployment. The first approach is based on the Neutron API and Modular Layer 2 (ML2), which are designed to provide common constructs such as network, router, and security groups that are familiar to Neutron users. The second approach is native to the group-based policy abstractions for OpenStack, which are closely aligned with the declarative policy model used in Cisco ACI.

ACI with OpenStack Physical Architecture

A typical architecture for an ACI fabric with an OpenStack deployment consists of a Nexus 9000 spine/leaf topology, an APIC cluster, and a group of servers to run the various control and compute components of OpenStack. An ACI external routed network connection as a Layer 3 connection outside the fabric can be used to provide connectivity outside the OpenStack cloud. Figure 6-10 illustrates OpenStack infrastructure connectivity with ACI.

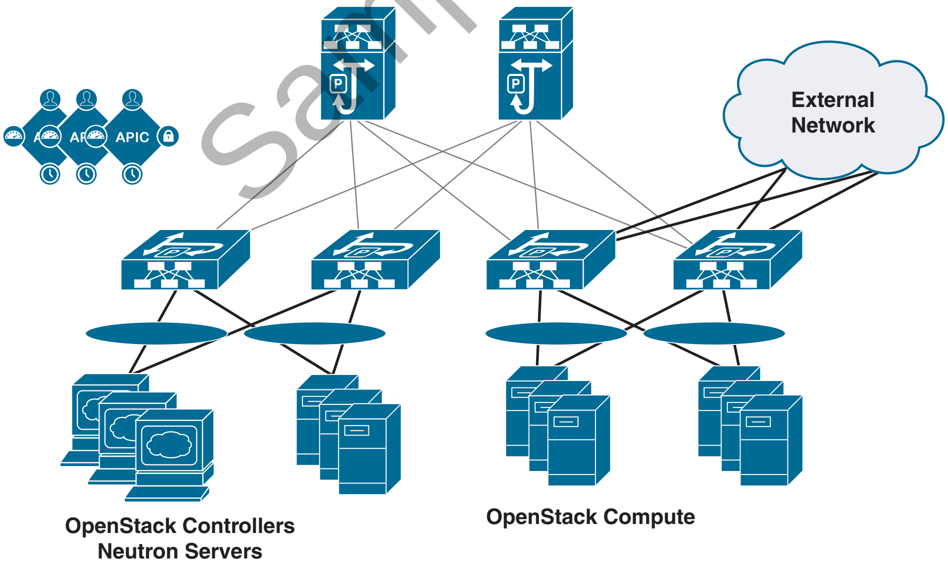

Figure 6-10 *OpenStack Physical Topology with ACI*

OpFlex Software Architecture

The ML2 framework in OpenStack enables the integration of networking services based on type drivers and mechanism drivers. Common networking type drivers include local, flat, VLAN, and VXLAN. OpFlex is added as a new network type through ML2, with an actual packet encapsulation of either VXLAN or VLAN on the host defined in the OpFlex configuration. A mechanism driver is enabled to communicate networking requirements from the Neutron servers to the Cisco APIC cluster. The APIC mechanism driver translates Neutron networking elements such as a network (segment), subnet, router, or external network into APIC constructs in the ACI policy model.

The OpFlex software stack also currently utilizes OVS and local software agents on each OpenStack compute host that communicates with the Neutron servers and OVS. An OpFlex proxy from the ACI leaf switch exchanges policy information with the agent OVS instance in each compute host, effectively extending the ACI switch fabric and policy model into the virtual switch. Figure 6-11 illustrates the OpenStack architecture with OpFlex in ACI.

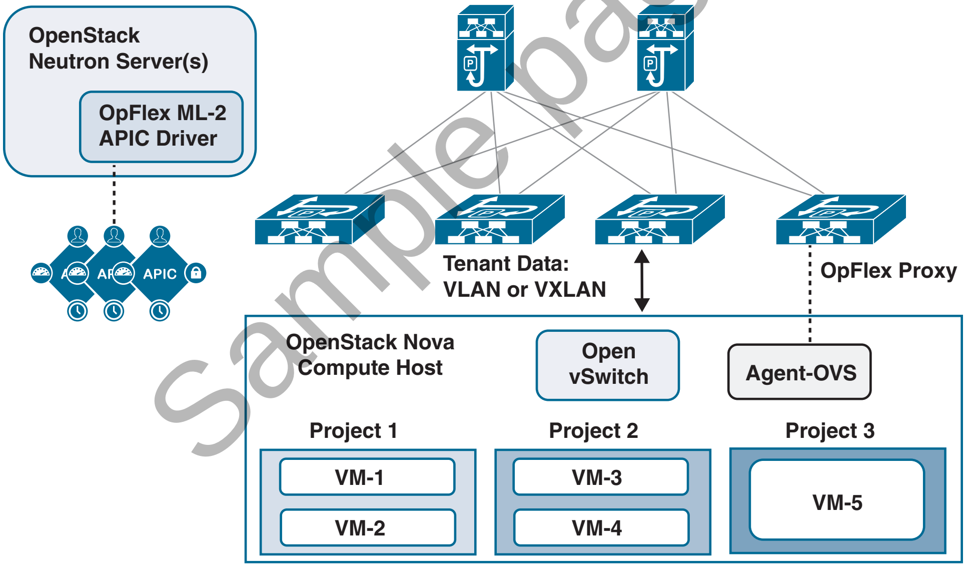

Figure 6-11 *OpenStack Architecture with OpFlex in ACI*

OpenStack Logical Topology

The logical topology diagram in Figure 6-12 illustrates the connections to OpenStack network segments from Neutron/controller servers and compute hosts, including the distributed Neutron services.

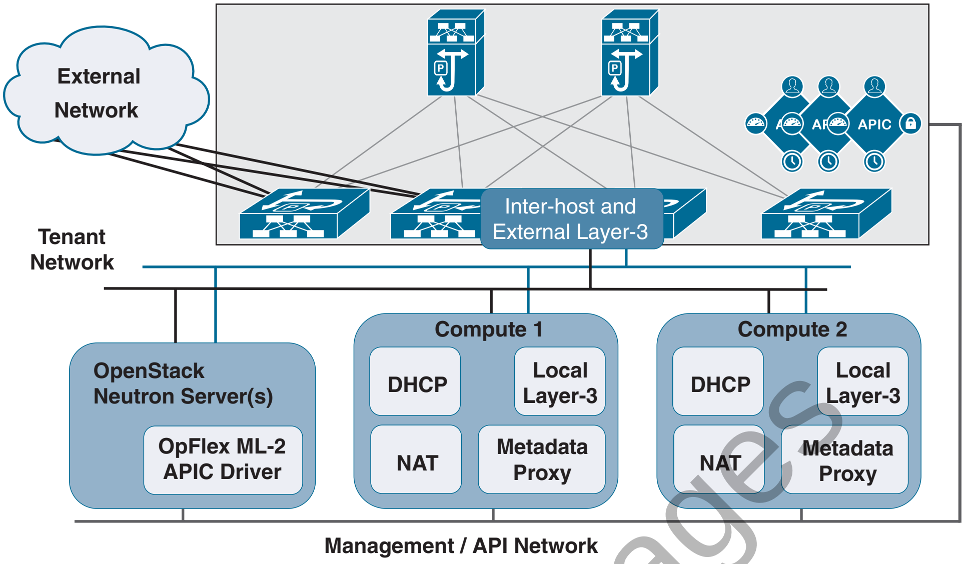

Figure 6-12 *OpenStack Logical Topology in ACI*

Note The management/API network for OpenStack can be connected to servers using an additional virtual NIC/subinterface on a common uplink with tenant networking to the ACI fabric, or by way of a separate physical interface.

Mapping OpenStack and ACI Constructs

Cisco ACI uses a policy model to enable network connectivity between endpoints attached to the fabric. OpenStack Neutron uses more traditional Layer 2 and Layer 3 networking concepts to define networking configuration. The OpFlex ML2 driver translates the Neutron networking requirements into the necessary ACI policy model constructs to achieve the desired connectivity. The OpenStack Group-Based Policy (GBP) networking model is quite similar to the Cisco ACI policy model. With the Cisco ACI unified plug-in for OpenStack, you can use both ML2 and GBP models on a single plug-in instance.

Note Only ML2 or GBP can be used for any given OpenStack project. A single project should not mix ML2 and GBP configurations.

Table 6-1 illustrates the OpenStack Neutron constructs and the corresponding APIC policy objects that are configured when they are created. In the case of GBP deployment, the policies have a direct mapping to the ACI policy model. Table 6-2 shows the OpenStack GBP objects and their corresponding ACI objects.

Table 6-1 *OpenStack Neutron Objects and Corresponding APIC Objects*

Neutron Object	APIC Object
(Neutron Instance)	VMM Domain
Project	Tenant + Application Network Profile
Network	EPG + Bridge Domain
Subnet	Subnet
Security Group + Rule	N/A (Iptables rules maintained per host)
Router	Contract
Network:external	L3Out/Outside EPG

Table 6-2 *OpenStack GBP Objects and Corresponding APIC Objects*

GBP Object	APIC Object
Policy Target	Endpoint
Policy Group	Endpoint Group (fvAEPg)
Policy Classifier	Filter (vzFilter)
Policy Action	--
Policy Rule	Subject (vzSubj)
Policy Ruleset	Contract (vzBrCP)
L2 Policy	Bridge Domain (fvBD)
L3 Policy	Context (fvCtx)

Prerequisites for OpenStack and Cisco ACI

Keep in mind the following prerequisites for OpenStack and Cisco ACI:

- **Target audience:** It is important to have working knowledge of Linux, the intended OpenStack distribution, the ACI policy model, and GUI-based APIC configuration.
- **ACI Fabric:** ACI fabric needs to be installed and initialized with a minimum APIC version 1.1(4e) and NX-OS version 11.1(4e). For basic guidelines on initializing a new ACI fabric, see the relevant documentation. For communication between multiple leaf pairs, the fabric must have a BGP route reflector enabled to use an OpenStack external network.
- **Compute:** You need to have a controller and servers connected to the fabric, preferably using NIC bonding and a VPC. In most cases the controller does not need to be connected to the fabric.

- **L3Out:** For external connectivity, one or more Layer 3 Outs (L3Outs) need to be configured on the ACI.
- **VLAN mode:** For VLAN mode, a non-overlapping VLAN pool of sufficient size should be allocated ahead of time.

Guidelines and Limitations for OpenStack and Cisco ACI

The following sections describes the guidelines and limitations for OpenStack and Cisco ACI.

Scalability Guidelines

There is a one-to-one correlation between the OpenStack tenant and the ACI tenant, and for each OpenStack tenant, the plug-in automatically creates ACI tenants named according to the following convention:

convention_*apic_system_id_openstack_tenant_name*

You should consider the scalability parameters for supporting the number of required tenants.

It is important to calculate the fabric scale limits for endpoint groups, bridge domains, tenants, and contracts before deployment. Doing so limits the number of tenant/project networks and routers that can be created in OpenStack. There are per-leaf and per-fabric limits. Make sure to check the scalability parameters for the deployed release before deployment. In the case of GBP deployment, it can take twice as many endpoint groups and bridge domains as with ML2 mode. Table 6-3 and Table 6-4 list the APIC resources that are needed for each OpenStack resource in GBP and ML2 configurations.

Table 6-3 *OpenStack GBP and ACI Resources*

GBP Resource	APIC Resources Consumed
L3 policy	One context
L2 policy	One bridge domain One endpoint group Two contracts
Policy group	One endpoint group
Ruleset	One contract
Classifier	Two filters (forward and reverse) Note: Five overhead classifiers are created

Table 6-4 *OpenStack ML2 and ACI Resources*

ML2 Resource	APIC Resources Consumed
Network	One bridge domain One endpoint group
Router	One contract
Security groups	N/A (no filters are used)

Availability Guidelines

For redundancy, you can use bonded interfaces (VPCs) by connecting two interfaces to two leaf switches and creating a VPC in ACI. You should deploy redundant OpenStack controller nodes to avoid a single point of failure. The external network should also be designed to avoid a single point of failure and service interruption.

NAT/External Network Operations

The OpFlex driver software can support external network connectivity and Network Address Translation (NAT) functions in a distributed manner using the local OVS instance on each OpenStack compute node. This distributed approach increases the availability of the overall solution and offloads the central processing of NAT from the Neutron server Layer 3 agent that is used in the reference implementation. You can also provide direct external connectivity without NAT or with a mix of NAT and non-NAT external connectivity.

Subnets Required for NAT

Unlike with the standard Neutron approach, three distinct IP subnets are required to take full advantage of external network functionality with the OpFlex driver:

- **Link subnet:** This subnet represents the actual physical connection to the external next-hop router outside of the fabric to be *assigned* to a routed interface, subinterface, or SVI.
- **Source NAT subnet:** This subnet is used for Port Address Translation (PAT), allowing multiple virtual machines to share an outside-routable IP address. A single IP address is assigned to each compute host, and Layer 4 port number manipulation is used to maintain unique session traffic.
- **Floating IP subnet:** With OpenStack, the term *floating IP* is used when a virtual machine instance is allowed to claim a distinct static NAT address to support inbound connections to the virtual machine from outside the cloud. The floating IP subnet is the subnet assigned within OpenStack to the Neutron external network entity.

Optimized DHCP and Metadata Proxy Operations

The OpFlex driver software stack provides optimized traffic flow and distributed processing to provide DHCP and metadata proxy services for virtual machine instances. These services are designed to keep processing and packet traffic local to the compute host as much as possible. The distributed elements communicate with centralized functions to ensure system consistency. You should enable optimized DHCP and metadata services when deploying the OpFlex plug-in for OpenStack.

Physical Interfaces

OpFlex uses the untagged fabric interface for an uplink trunk in VLAN mode. This means the fabric interface cannot be used for PXE because PXE usually requires an untagged interface. If you require PXE in a VLAN mode deployment, you must use a separate interface for PXE. This interface can be connected through ACI or an external switch. This issue is not present in VXLAN mode since tunnels are created using the tagged interface for an infrastructure VLAN.

Layer 4 to Layer 7 Services

Service insertion in OpenStack is done through a physical domain or device package. You should check customer requirements and the plug-in mode (GBP or ML2) to plan how service insertion/chaining will be done. The OpenStack Neutron project also defines Layer 4 to Layer 7 extension APIs, such as LBaaS, FWaaS, and VPNaaS. The availability of these extensions depends on the device vendor. Check the vendor for the availability of these extensions.

Blade Servers

When deploying on blade servers, you must make sure there is no intermediate switch between the fabric and the physical server interfaces. Check the OpenStack ACI plug-in release notes to make sure a particular configuration is supported. At this writing, there is limited support for B-Series blade servers, and the support is limited to VLAN mode only.

Verifying the OpenStack Configuration

Follow these steps to verify the OpenStack configuration:

Step 1. Verify that a VMM domain was created for the OpenStack system ID defined during installation. The nodes connected to the fabric that are running the OpFlex agent should be visible under Hypervisors. The virtual machines running on the hypervisor should be visible when you select that hypervisor. All networks created for this tenant should also be visible under the DVS submenu, and selecting the network should show you all endpoints connected to that network.

Step 2. Look at the health score and faults for the entity to verify correct operation. If the hypervisors are not visible or appear as being disconnected, check the OpFlex connectivity.