



ACI Advanced Monitoring and Troubleshooting

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

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

Contents

	Foreword by Yusuf Bhajji	xxviii
	Foreword by Ronak Desai	xxix
	Introduction	xxx
Part I	Introduction to ACI	
Chapter 1	Fundamental Functions and Components of Cisco ACI	1
	ACI Building Blocks	8
	Hardware Specifications	8
	<i>Nexus 9000 Platform</i>	9
	<i>APIC Controller</i>	12
	ACI Key Concepts	14
	Control Plane	15
	Data Plane	17
	VXLAN	17
	Tenant	18
	VRF	19
	Application Profile	20
	Endpoint Group	21
	Contracts	22
	Bridge Domain	24
	External Routed or Bridged Network	25
	Summary	26
	Review Key Topics	26
	Review Questions	27
Chapter 2	Introduction to the ACI Policy Model	31
	Key Characteristics of the Policy Model	32
	Management Information Tree (MIT)	33
	Benefits of a Policy Model	37
	Logical Constructs	37
	Tenant Objects	38
	VRF Objects	39
	Application Profile Objects	40
	Endpoint Group Objects	41

Bridge Domain and Subnet Objects	43
Bridge Domain Options	45
Contract Objects	46
Labels, Filters, and Aliases	48
Contract Inheritance	49
Contract Preferred Groups	49
vzAny	50
Outside Network Objects	51
Physical Construct	52
Access Policies	52
Switch Policies	53
Interface Policies	54
Global Policies	55
VLAN Pools	55
Domains	56
Attachable Access Entity Profile	56
Managed Object Relationships and Policy Resolution	57
Tags	58
Default Policies	58
How a Policy Model Helps in Diagnosis	60
Summary	63
Review Key Topics	63
Review Questions	64
Chapter 3 ACI Command-Line Interfaces	67
APIC CLIs	68
NX-OS–Style CLI	68
Bash CLI	74
ACI Fabric Switch CLIs	78
iBash CLI	78
VSH CLI	81
VSH_LC CLI	83
Summary	84
Reference	84

Chapter 4 ACI Fabric Design Options 85

Physical Design	85
Single- Versus Multiple-Fabric Design	87
<i>Dark Fiber</i>	90
<i>Dense Wavelength-Division Multiplexing (DWDM)</i>	92
<i>Ethernet over MPLS (EoMPLS) Pseudowire</i>	92
Multi-Pod	97
<i>ACI Multi-Pod Use Cases</i>	100
<i>ACI Multi-Pod Scalability</i>	103
<i>Inter-Pod Connectivity Deployment Considerations</i>	104
<i>APIC Cluster Deployment Considerations</i>	113
Multi-Site	116
<i>Cisco ACI Multi-Site Orchestrator</i>	120
<i>Cisco ACI Multi-Site Deployment Considerations</i>	122
<i>Migration Scenarios</i>	124
<i>Deployment Best Practices</i>	128
<i>General Best Practices for Cisco ACI Multi-Site Design</i>	129
Remote Leaf	131
Hardware and Software Support	134
<i>Recommended QOS Configuration for a Remote Leaf Solution</i>	134
<i>Discovery of a Remote Leaf</i>	136
<i>Remote Leaf Control Plane and Data Plane</i>	138
<i>Remote Leaf Design Considerations</i>	141
ACI Multi-Pod and Remote Leaf Integration	143
Logical Design	149
Design 1: Container-as-a-Service Using the OpenShift Platform and Calico CNI	149
<i>Business Case</i>	149
<i>Design Solution</i>	150
Design 2: Vendor-Based ERP/SAP Hana Design with ACI	165
<i>Business Case</i>	165
<i>Design Solution</i>	165
Design 3: vBrick Digital Media Engine Design with ACI	175
<i>Business Case</i>	176
<i>Design Solution</i>	176

Summary	180
Review Key Topics	181
Review Questions	181

Chapter 5 End Host and Network Connectivity 185

End Host Connectivity	185
VLAN Pool	186
Domain	186
Attachable Access Entity Profiles (AAEPs)	186
Switch Policies	187
<i>Switch Policy Groups</i>	187
<i>Switch Profiles</i>	187
Interface Policies	188
<i>Interface Policy Groups</i>	188
<i>Interface Profiles</i>	189
Virtual Port Channel (VPC)	191
<i>Configuring VPC</i>	192
<i>Defining the VPC Domain</i>	193
<i>Creating an Interface Policy</i>	195
<i>Creating a Switch Profile</i>	196
Port Channel	197
<i>Configuring a Port Channel</i>	198
Access Port	201
<i>Configuring an Access Port</i>	202
Best Practices in Configuring Access Policies	206
<i>Policy Best Practices</i>	206
<i>Domain Best Practices</i>	206
<i>AAEP Best Practices</i>	207
Compute and Storage Connectivity	207
<i>FEX Connectivity</i>	207
<i>Cisco Blade Chassis Servers UCS B-Series</i>	208
<i>Standalone Rack-Mount Servers</i>	209
<i>Connecting Storage in ACI</i>	209
L4/L7 Service Device Connectivity	210
<i>Connecting Firewalls</i>	211
<i>Connecting Load Balancers</i>	212

Network Connectivity	213
Connecting an External Bridge Network	213
<i>Extending EPGs Outside the ACI Fabric</i>	213
<i>Extending an ACI Bridge Domain Outside the Fabric</i>	216
Connecting an External Routed Network	218
<i>External Layer 3–Supported Routing Protocols</i>	220
<i>Configuring MP-BGP Spine Route Reflectors</i>	221
<i>Configuring External Routed Networks</i>	222
GOLF	227
<i>Network Connectivity Between Pods and Sites</i>	228
<i>IPN Connectivity Considerations for Remote Leafs</i>	237
Diagnosing Connectivity Problems	242
Summary	245
Review Questions	245
Chapter 6 VMM Integration	249
Virtual Machine Manager (VMM)	249
VMM Domain Policy Model	250
VMM Domain Components	250
VMM Domains	250
VMM Domain VLAN Pool Association	252
<i>Attachable Access Entity Profile Association</i>	252
<i>VMM Domain EPG Association</i>	253
<i>EPG Policy Resolution and Deployment Immediacy</i>	255
VMware Integration	257
Prerequisites for VMM Integration with AVS or VDS	257
Guidelines and Limitations for VMM Integration with AVS or VDS	257
ACI VMM Integration Workflow	258
Publishing EPGs to a VMM Domain	258
Connecting Virtual Machines to the Endpoint Group Port Groups on vCenter	259
Verifying VMM Integration with the AVS or VDS	259
<i>Verifying the Virtual Switch Status</i>	259
<i>Verifying the vNIC Status</i>	260
Microsoft SCVMM Integration	260
Mapping ACI and SCVMM Constructs	261

Mapping Multiple SCVMMs to an APIC	262
Verifying That the OpFlex Certificate Is Deployed for a Connection from the SCVMM to the APIC	262
Verifying VMM Deployment from the APIC to the SCVMM	263
OpenStack Integration	263
Extending OpFlex to the Compute Node	264
ACI with OpenStack Physical Architecture	264
OpFlex Software Architecture	265
OpenStack Logical Topology	265
Mapping OpenStack and ACI Constructs	266
<i>Prerequisites for OpenStack and Cisco ACI</i>	267
<i>Guidelines and Limitations for OpenStack and Cisco ACI</i>	268
<i>Verifying the OpenStack Configuration</i>	270
<i>Configuration Examples for OpenStack and Cisco ACI</i>	271
Kubernetes Integration	272
Planning for Kubernetes Integration	272
Prerequisites for Integrating Kubernetes with Cisco ACI	273
Provisioning Cisco ACI to Work with Kubernetes	274
Preparing the Kubernetes Nodes	277
Installing Kubernetes and Cisco ACI Containers	279
Verifying the Kubernetes Integration	280
OpenShift Integration	281
Planning for OpenShift Integration	282
Prerequisites for Integrating OpenShift with Cisco ACI	283
Provisioning Cisco ACI to Work with OpenShift	284
Preparing the OpenShift Nodes	287
Installing OpenShift and Cisco ACI Containers	290
Updating the OpenShift Router to Use the ACI Fabric	291
Verifying the OpenShift Integration	291
VMM Integration with ACI at Multiple Locations	292
Multi-Site	292
<i>Multiple Virtual Machine Managers Across Sites</i>	292
<i>Single Virtual Machine Manager Across Sites</i>	295
Remote Leaf	295
Summary	298

Chapter 7 L4/L7 Service Integration 299

Service Insertion	299
The Service Graph	300
Managed Mode Versus Un-Managed Mode	301
L4–L7 Integration Use Cases	302
How Contracts Work in ACI	303
The Shadow EPG	306
Configuring the Service Graph	307
<i>Step 1: Create an L4–L7 Device</i>	307
<i>Step 2: Create a Service Graph Template</i>	308
<i>Step 3: Deploy the Service Graph from the Template</i>	308
<i>Step 4: Configure the L4–L7 Parameters (Managed Mode Only)</i>	310
Verifying the Service Graph Configuration	310
Service Graph Design and Deployment Options	312
<i>Firewall as Default Gateway for Client and Server (Routed Mode)</i>	312
<i>Firewall Not the Default Gateway for Clients (Routed Mode)</i>	312
<i>Route Peering with a Firewall (Routed Mode)</i>	314
<i>Service Graph with Firewall (Transparent Mode)</i>	316
<i>Service Graph with ADC (One-Arm Mode with S-NAT)</i>	316
<i>Service Graph with ADC (Two-Arm Mode)</i>	316
<i>Service Graph with Two Service Nodes (Firewall with NAT and ADC in Two-Arm Mode)</i>	317
<i>Service Graph with Two Service Nodes (Firewall with No NAT and ADC in Two-Arm Mode)</i>	319
<i>Service Graph with Two Service Nodes (Firewall with No NAT and ADC in One-Arm Mode)</i>	319
<i>Service Graph with an Intrusion Prevention System (IPS)</i>	319
Policy-Based Redirect (PBR)	322
PBR Design Considerations	323
PBR Design Scenarios	324
<i>PBR Service Graph with an ADC (One-Arm Mode and No S-NAT)</i>	324
<i>PBR Service Graph with a Firewall (Two-Arm Mode and Routed)</i>	324
Configuring the PBR Service Graph	325

Service Node Health Check	326
<i>L4–L7 PBR Tracking</i>	326
<i>L4–L7 PBR Threshold</i>	326
<i>L4–L7 PBR Health Groups</i>	327
Common Issues in the PBR Service Graph	328
<i>Unnecessary Layer 2 Traffic Redirection Toward the Service Node</i>	328
<i>Inability to Ping the Consumer Connector</i>	329
<i>Routing on a Service Node</i>	330
L4/L7 Service Integration in Multi-Pod and Multi-Site	332
Multi-Pod	332
<i>Anycast Services in Multi-Pod</i>	334
Multi-Site	338
Review Questions	342

Chapter 8 Automation and Orchestration 343

The Difference Between Automation and Orchestration	343
Benefits of Automation and Orchestration	344
<i>Example 1</i>	345
<i>Example 2</i>	347
REST API	349
Automating Tasks Using the Native REST API: JSON and XML	351
API Inspector	351
Object (Save As)	353
Visore (Object Store Browser)	355
MOQuery	357
Automation Use Cases	364
Automating Tasks Using Ansible	372
Ansible Support in ACI	375
Installing Ansible and Ensuring a Secure Connection	378
APIC Authentication in Ansible	382
Automation Use Cases	384
<i>Use Case 1</i>	384
<i>Use Case 2</i>	388

Orchestration Through UCS Director	392
Management Through Cisco UCS Director	392
Automation and Orchestration with Cisco UCS Director	393
Automation Use Cases	395
Summary	402
Review Questions	402

Part II Monitoring and Management Best Practices

Chapter 9 Monitoring ACI Fabric 405

Importance of Monitoring	405
Faults and Health Scores	407
Faults	407
Health Scores	411
<i>Health Score Used in Proactive Monitoring</i>	413
<i>Health Score Used in Reactive Monitoring</i>	414
<i>Health Score with Interface Errors</i>	414
ACI Internal Monitoring Tools	415
SNMP	415
<i>Interface Failures Example</i>	418
Syslog	420
<i>Example: Leaf Membership Failure</i>	423
<i>Example: Spine/IPN Failure</i>	423
NetFlow	426
<i>Example: Network Visibility on a Border Leaf</i>	428
ACI External Monitoring Tools	430
Network Insights	430
<i>Network Insights for Resources (NIR)</i>	431
<i>Network Insights Advisor (NIA)</i>	432
<i>Example: Application Intermittent Disconnect Issue (Standalone Compute)</i>	433
<i>Example: Application Connectivity Issue (Virtual Compute)</i>	435
Network Assurance Engine	437
<i>NAE Installation</i>	439
<i>NAE Configuration and Initial Setup</i>	440
<i>Example: Subnet Reachability Issue</i>	450

Tetration	453
<i>Software Agents</i>	455
<i>Hardware Agents</i>	455
<i>Tetration Installation and Configuration</i>	455
<i>Tetration System Monitoring</i>	461
<i>Configuring Email Alerts</i>	463
<i>Enabling Syslog</i>	464
<i>Tetration Scopes</i>	465
<i>Tetration Applications</i>	465
<i>Tetration Code Upgrades</i>	467
<i>Tetration Patch Upgrade</i>	467
<i>Tetration Cluster Reboot</i>	469
<i>Tetration Cluster Shutdown</i>	469
<i>Example: Workload Security with Tetration</i>	470
Monitoring Through the REST API	473
Monitoring an APIC	475
<i>Monitoring CPU and Memory</i>	475
<i>Monitoring Disk Utilization</i>	477
<i>Monitoring Interfaces</i>	478
<i>Monitoring the APIC Cluster State</i>	481
Monitoring Leafs and Spines	482
<i>Monitoring CPU Utilization</i>	482
<i>Monitoring Memory Utilization</i>	485
<i>Monitoring Power Supply Unit (PSU) Status</i>	486
<i>Monitoring Fan Status</i>	488
<i>Monitoring Module Status</i>	489
<i>Monitoring Leaf/Spine Membership Status in a Fabric</i>	491
<i>Monitoring Interface Status</i>	496
Monitoring Applications	499
<i>Monitoring Application Traffic Status</i>	499
<i>Monitoring External Network Connectivity</i>	502
<i>Monitoring the PBR Service Graph</i>	504
Summary	505
Review Questions	506

Chapter 10 Network Management and Monitoring Configuration 509

Out-of-Band Management	509
Creating Static Management Addresses	510
Creating the Management Contract	510
Choosing the Node Management EPG	513
Creating an External Management Entity EPG	513
Verifying the OOB Management Configuration	515
In-Band Management	517
Creating a Management Contract	517
Creating Leaf Interface Access Policies for APIC INB Management	518
Creating Access Policies for the Border Leaf(s) Connected to L3Out	520
Creating INB Management External Routed Networks (L3Out)	522
Creating External Management EPGs	524
Creating an INB BD with a Subnet	527
Configuring the Node Management EPG	529
Creating Static Management Addresses	530
Verifying the INB Management Configuration	530
AAA	533
Configuring Cisco Secure ACS	533
Configuring Cisco ISE	542
Configuring AAA in ACI	547
Recovering with the Local Fallback User	550
Verifying the AAA Configuration	550
Syslog	551
Verifying the Syslog Configuration and Functionality	555
SNMP	556
Verifying the SNMP Configuration and Functionality	562
SPAN	566
Access SPAN	567
Fabric SPAN	571
Tenant SPAN	572
Ensuring Visibility and Troubleshooting SPAN	575
Verifying the SPAN Configuration and Functionality	576

NetFlow	577
NetFlow with Access Policies	580
NetFlow with Tenant Policies	582
Verifying the NetFlow Configuration and Functionality	585
Summary	587

Part III Advanced Forwarding and Troubleshooting Techniques

Chapter 11 ACI Topology 589

Physical Topology	589
APIC Initial Setup	593
Fabric Access Policies	595
Switch Profiles, Switch Policies, and Interface Profiles	595
Interface Policies and Policy Groups	596
Pools, Domains, and AAEPs	597
VMM Domain Configuration	601
VMM Topology	601
Hardware and Software Specifications	603
Logical Layout of EPGs, BDs, VRF Instances, and Contracts	605
L3Out Logical Layout	606
Summary	608
Review Key Topics	608
References	609

Chapter 12 Bits and Bytes of ACI Forwarding 611

Limitations of Traditional Networks and the Evolution of Overlay Networks	611
High-Level VXLAN Overview	613
IS-IS, TEP Addressing, and the ACI Underlay	615
IS-IS and TEP Addressing	615
FTags and the MDT	618
Endpoint Learning in ACI	626
Endpoint Learning in a Layer 2–Only Bridge Domain	627
<i>Council of Oracle Protocol (COOP)</i>	632
<i>Updating the Managed Object (MO) Tree</i>	634
Endpoint Learning in a Layer 3–Enabled Bridge Domain	635
Fabric Glean	640

Remote Endpoint Learning	641
Endpoint Mobility	645
Anycast Gateway	647
Virtual Port Channels in ACI	649
Routing in ACI	651
Static or Dynamic Routes	651
Learning External Routes in the ACI Fabric	656
Transit Routing	659
Policy Enforcement	661
Shared Services	664
L3Out Flags	668
Quality of Service (QoS) in ACI	669
Externally Set DSCP and CoS Markings	671
EPG QoS	671
Custom QoS Policy	671
Contract QoS	671
CoS Preservation in ACI	672
iTraceroute Class	672
QoS and Multi-Pod	672
DSCP Class-to-CoS Translation Policy	674
Multi-Pod	674
Multi-Site	680
Remote Leaf	684
Forwarding Scenarios	686
ARP Flooding	686
Layer 2 Known Unicast	688
ARP Optimization	690
Layer 2 Unknown Unicast Proxy	690
L3 Policy Enforcement When Going to L3Out	693
L3 Policy Enforcement for External Traffic Coming into the Fabric	695
Route Leaking/Shared Services	695
Consumer to Provider	695
Provider to Consumer	698

Multi-Pod Forwarding Examples	698
ARP Flooding	700
Layer 3 Proxy Flow	700
Multi-Site Forwarding Examples	703
ARP Flooding	703
Layer 3 Proxy Flow	705
Remote Leaf	707
ARP Flooding	707
Layer 3 Proxy Flow	710
Summary	713
Review Key Topics	713
References	714
Review Questions	714

Chapter 13 Troubleshooting Techniques 717

General Troubleshooting	717
Faults, Events, and Audits	718
moquery	722
iCurl	724
Visore	726
Infrastructure Troubleshooting	727
APIC Cluster Troubleshooting	727
Fabric Node Troubleshooting	734
How to Verify Physical- and Platform-Related Issues	737
Counters	737
CPU Packet Captures	743
ASIC	744
ASIC Interface	744
Application	745
SPAN	748
Troubleshooting Endpoint Connectivity	751
Endpoint Tracker and Log Files	752
Enhanced Endpoint Tracker (EPT) App	756
Rogue Endpoint Detection	758

Troubleshooting Contract-Related Issues	759
Verifying Policy Deny Drops	764
Embedded Logic Analyzer Module (ELAM)	765
Summary	769
Review Key Topics	769
Review Questions	769

Chapter 14 The ACI Visibility & Troubleshooting Tool 771

Visibility & Troubleshooting Tool Overview	771
Faults Tab	772
Drop/Stats Tab	773
Ingress/Egress Buffer Drop Packets	774
Ingress Error Drop Packets Periodic	774
Storm Control	774
Ingress Forward Drop Packets	775
Ingress Load Balancer Drop Packets	776
Contract Drops Tab	777
Contracts	777
Contract Considerations	778
Events and Audits Tab	779
Traceroute Tab	780
Atomic Counter Tab	782
Latency Tab	785
SPAN Tab	786
Network Insights Resources (NIR) Overview	787
Summary	790

Chapter 15 Troubleshooting Use Cases 791

Troubleshooting Fabric Discovery: Leaf Discovery	792
Solution	794
Troubleshooting APIC Controllers and Clusters: Clustering	795
Solution	798
Troubleshooting Management Access: Out-of-Band EPG	799
Solution	801
Troubleshooting Contracts: Traffic Not Traversing a Firewall as Expected	801
Solution	803

Troubleshooting Contracts: Contract Directionality	804
Solution	807
Troubleshooting End Host Connectivity: Layer 2 Traffic Flow Through ACI	807
Solution	810
Troubleshooting External Layer 2 Connectivity: Broken Layer 2 Traffic Flow Through ACI	812
Solution 1	813
Solution 2	813
Troubleshooting External Layer 3 Connectivity: Broken Layer 3 Traffic Flow Through ACI	814
Solution	816
Troubleshooting External Layer 3 Connectivity: Unexpected Layer 3 Traffic Flow Through ACI	816
Solution	820
Troubleshooting Leaf and Spine Connectivity: Leaf Issue	821
Solution	822
Troubleshooting VMM Domains: VMM Controller Offline	826
Solution 1	829
Solution 2	829
Troubleshooting VMM Domains: VM Connectivity Issue After Deploying the VMM Domain	829
Solution 1	830
Solution 2	831
Solution 3	831
Troubleshooting L4–L7: Deploying an L4–L7 Device	832
Solution	834
Troubleshooting L4–L7: Control Protocols Stop Working After Service Graph Deployment	834
Solution	836
Troubleshooting Multi-Pod: BUM Traffic Not Reaching Remote Pods	837
Solution 1	839
Solution 2	839
Troubleshooting Multi-Pod: Remote L3Out Not Reachable	839
Solution	841
Troubleshooting Multi-Site: Using Consistency Checker to Verify State at Each Site	841
Solution	842

Troubleshooting Programmability Issues: JSON Script Generates Error 844

Solution 844

Troubleshooting Multicast Issues: PIM Sparse Mode Any-Source Multicast (ASM) 846

Solution 847

Summary 860

Appendix A Answers to Chapter Review Questions 861

Index 873

Sample pages

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.

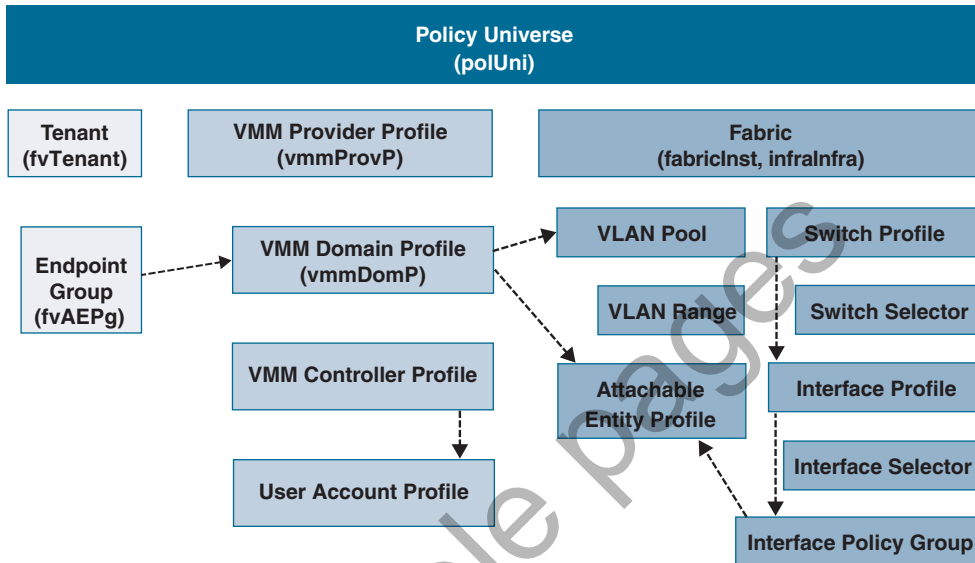


Figure 6-1 VMM Policy Model

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.

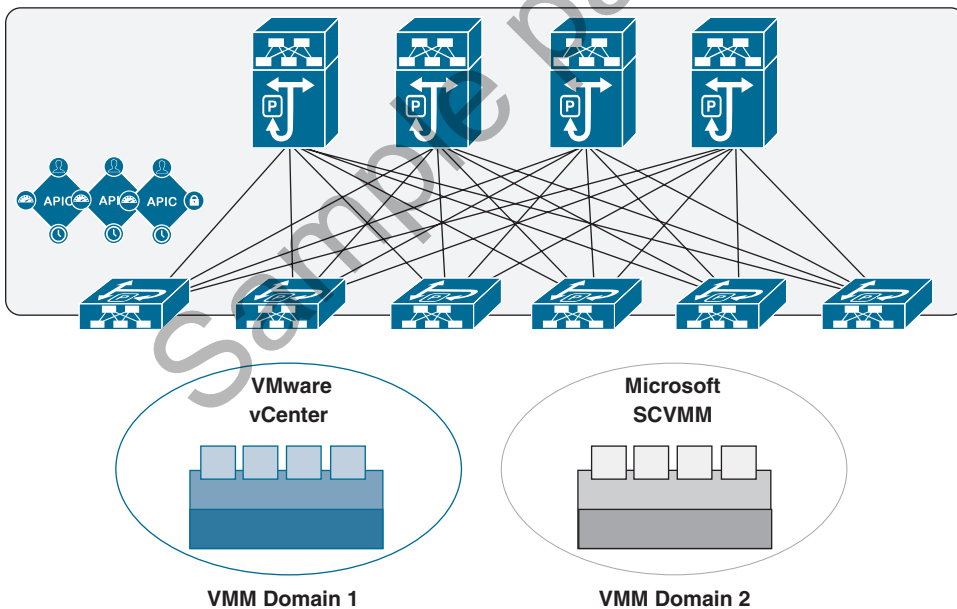


Figure 6-2 VMM Domain Integration

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 grey EPs are in the same EPG, even though they are in different VMM domains.

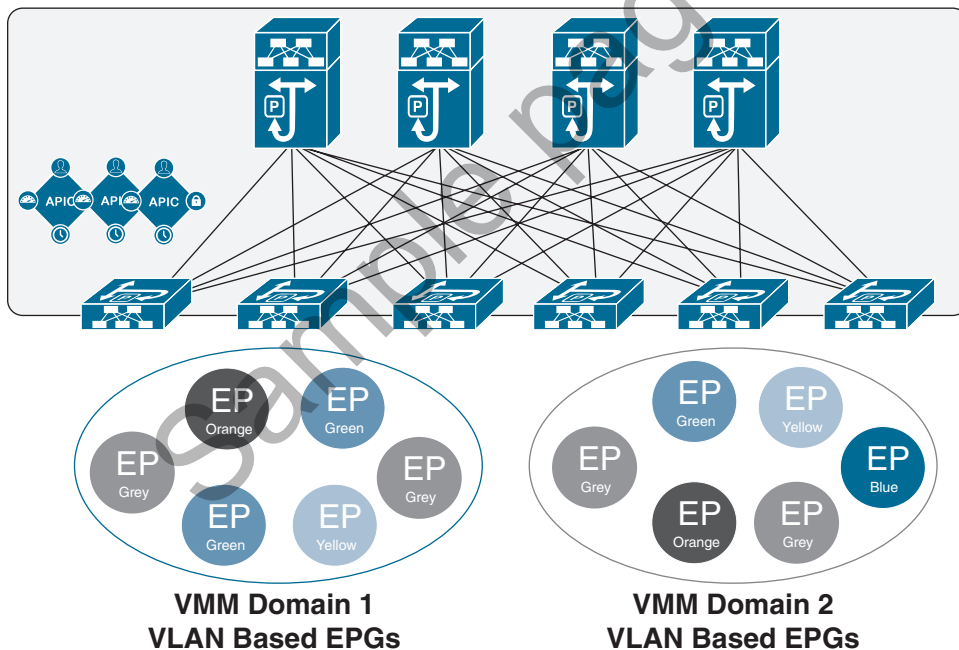


Figure 6-3 VMM Domain EPG Association

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.

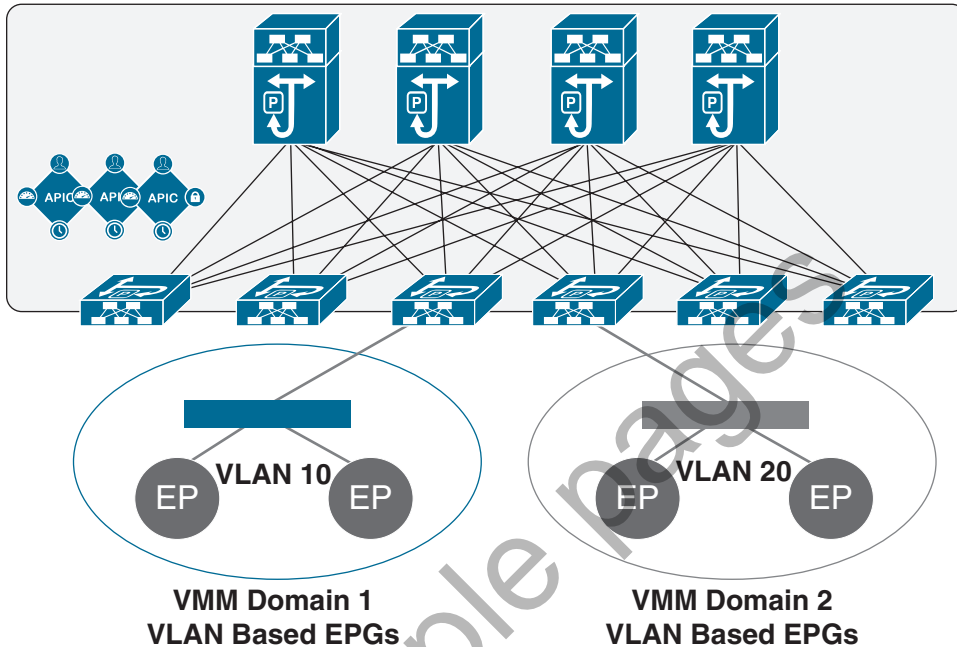


Figure 6-4 VMM Domain EPG VLAN Consumption

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.

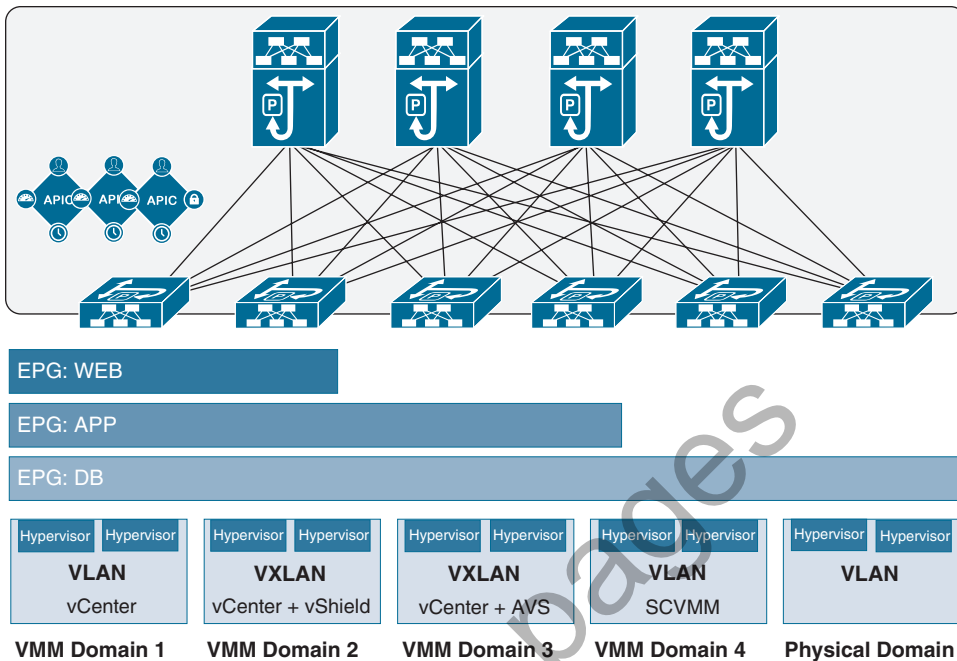


Figure 6-5 Multiple VMM Domains and Scaling of EPGs in the ACI Fabric

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 neighborhood 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 neighborhood 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 neighborhood 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 neighborhood, 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.

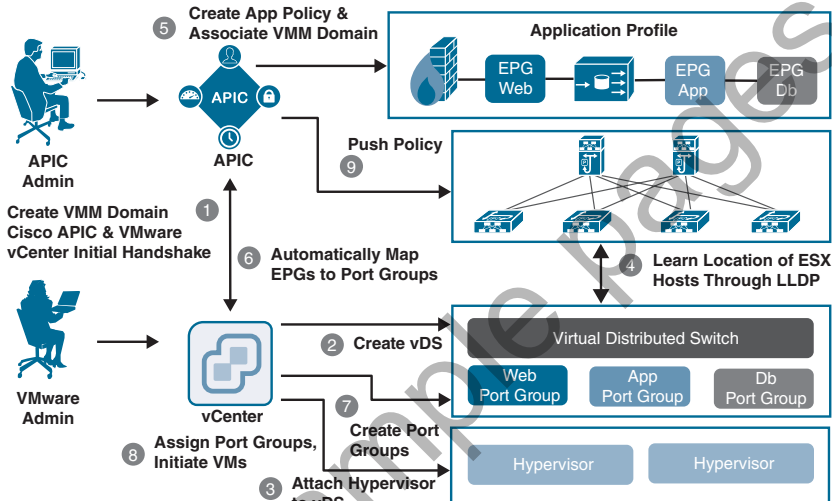


Figure 6-6 ACI VMM Integration Workflow

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.

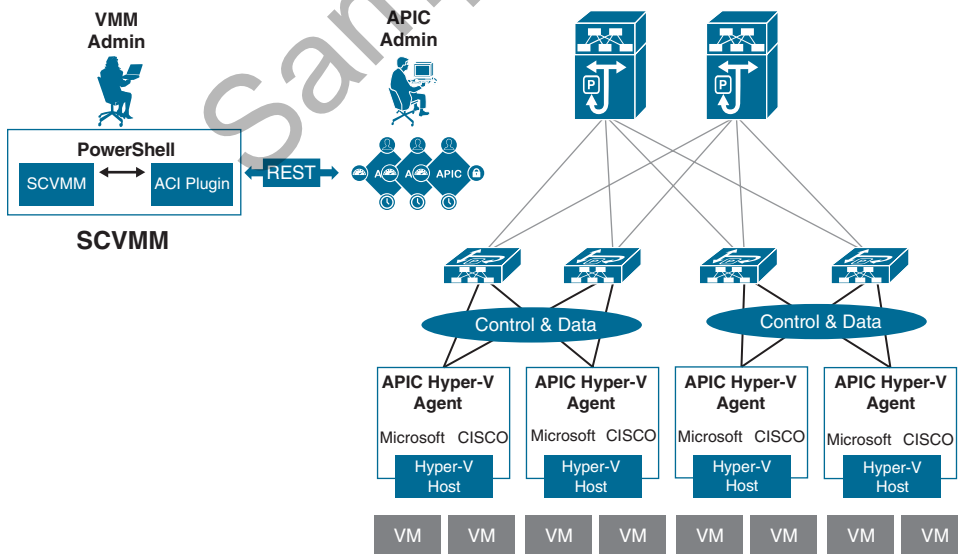


Figure 6-7 Microsoft SCVMM Topology with ACI

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

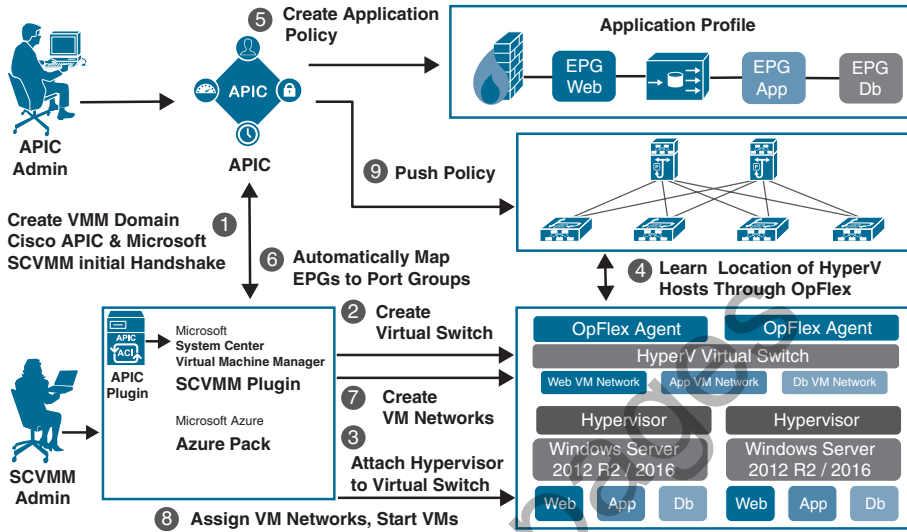


Figure 6-8 Workflow for Integrating ACI and Microsoft SCVMM

Mapping ACI and SCVMM Constructs

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

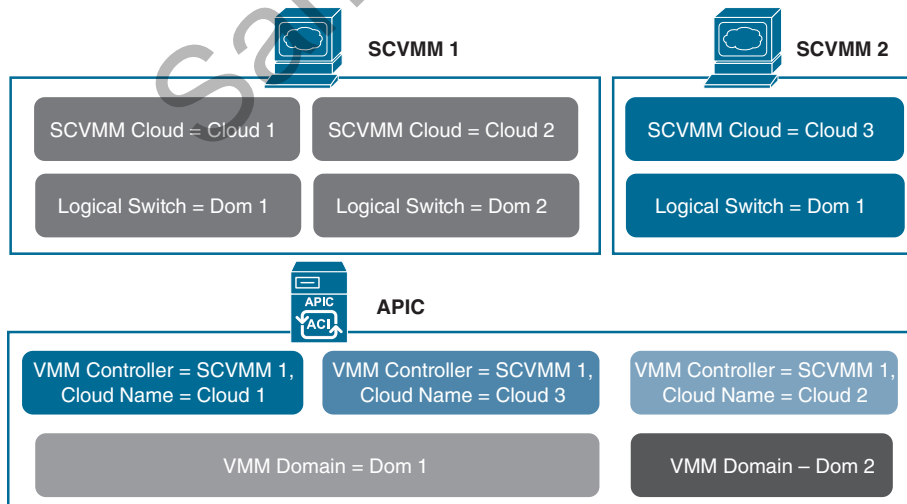


Figure 6-9 Mapping ACI and SCVMM Constructs

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_SCVMM_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 leaves 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.

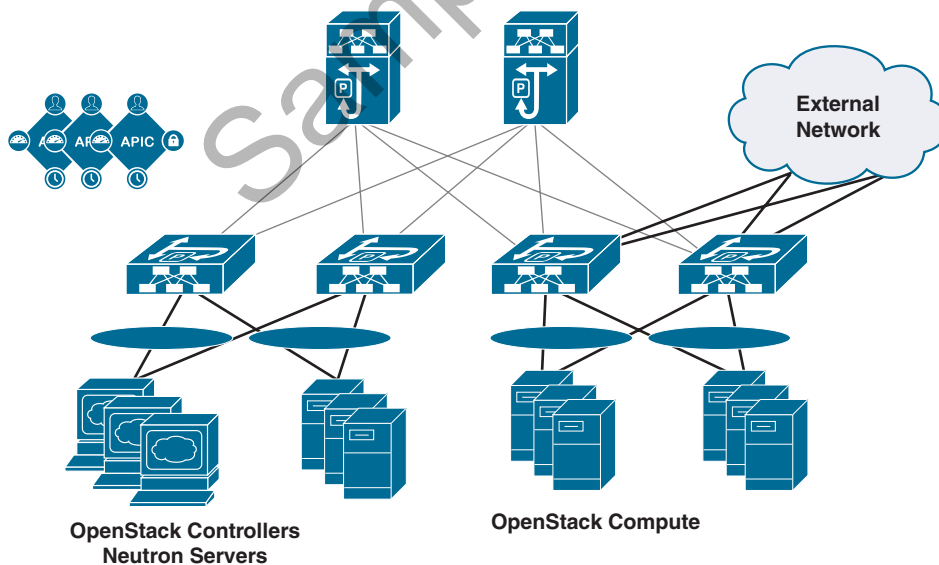


Figure 6-10 OpenStack Physical Topology with ACI

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.

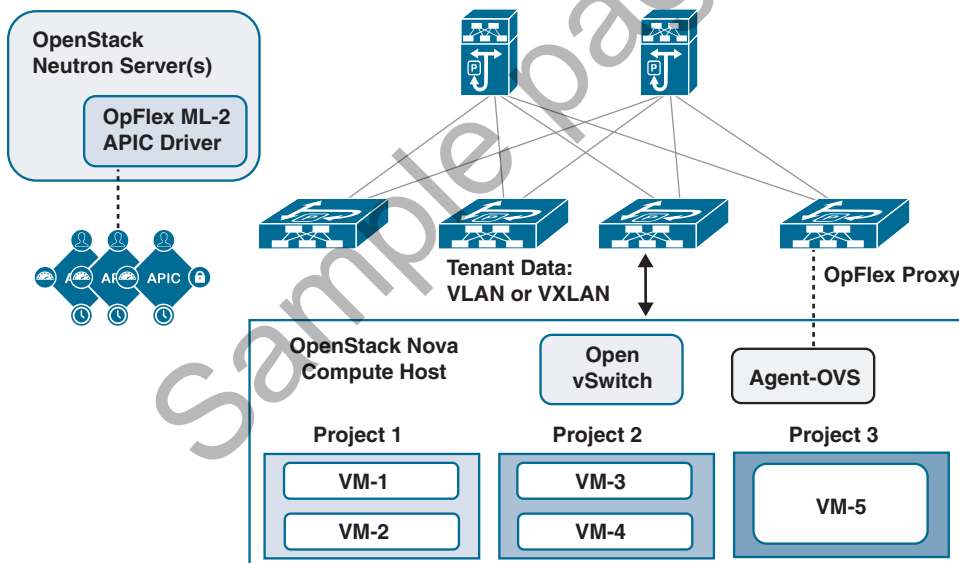


Figure 6-11 OpenStack Architecture with OpFlex in ACI

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.

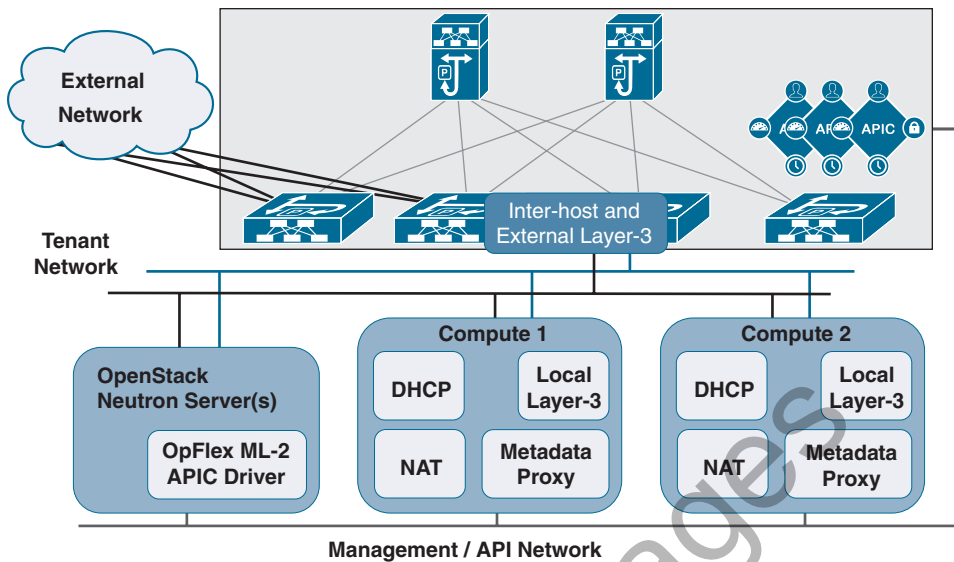


Figure 6-12 *OpenStack Logical Topology in ACI*

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 sub-menu, 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.