Video Training

Official Cert Guide

Advance your IT career with hands-on learning

CCNP Data Center Application Centric Infrastructure

DCACI 300-620

Ammar Ahmadi, CCIE® No. 50928

ciscopress.com

Contents

CHAPTER 2

Understanding ACI Hardware and Topologies

This chapter covers the following topics:

ACI Topologies and Components: This section describes the key hardware components and acceptable topologies for ACI fabrics.

APIC Clusters: This section covers available APIC hardware models and provides an understanding of APIC cluster sizes and failover implications.

Spine Hardware: This section addresses available spine hardware options.

Leaf Hardware: This section outlines the leaf platforms available for deployment in ACI fabrics.

This chapter covers the following exam topics:

- 1.1 Describe ACI topology and hardware
- 6.1 Describe Multi-Pod
- 6.2 Describe Multi-Site

ACI is designed to allow small and large enterprises and service providers to build massively scalable data centers using a relatively small number of very flexible topologies.

This chapter details the topologies with which an ACI fabric can be built or extended. Understanding supported ACI topologies helps guide decisions on target-state network architecture and hardware selection.

Each hardware component in an ACI fabric performs a specific set of functions. For example, leaf switches enforce security rules, and spine switches track all endpoints within a fabric in a local database.

But not all ACI switches are created equally. Nor are APICs created equally. This chapter therefore aims to provide a high-level understanding of some of the things to consider when selecting hardware.

"Do I Know This Already?" Quiz

The "Do I Know This Already?" quiz allows you to assess whether you should read this entire chapter thoroughly or jump to the "Exam Preparation Tasks" section. If you are in doubt about your answers to these questions or your own assessment of your knowledge of the topics, read the entire chapter. Table 2-1 lists the major headings in this chapter and their corresponding "Do I Know This Already?" quiz questions. You can find the answers in Appendix A, "Answers to the 'Do I Know This Already?' Questions."

Table 2-1 "Do I Know This Already?" Section-to-Question Mapping

Foundation Topics Section	Questions
ACI Topologies and Components	1–5
APIC Clusters	6
Spine Hardware	7, 8
Leaf Hardware	9, 10

CAUTION The goal of self-assessment is to gauge your mastery of the topics in this chapter. If you do not know the answer to a question or are only partially sure of the answer, you should mark that question as wrong for purposes of the self-assessment. Giving yourself credit for an answer you correctly guess skews your self-assessment results and might provide you with a false sense of security.

1. An ACI fabric is being extended to a secondary location to replace two top-of-rack switches and integrate a handful of servers into a corporate ACI environment. Which solution should ideally be deployed at the remote location if the deployment of new spines is considered cost-prohibitive and direct fiber links from the main data center cannot be dedicated to this function?

- **a.** ACI Multi-Site
- **b.** ACI Remote Leaf
- **c.** ACI Multi-Tier
- **d.** ACI Multi-Pod

2. Which of the following is a requirement for a Multi-Pod IPN that is not needed in an ACI Multi-Site ISN?

- **a.** Increased MTU support
- **b.** OSPF support on last-hop routers connecting to ACI spines
- **c.** End-to-end IP connectivity
- **d.** Multicast PIM-Bidir

3. Which of the following connections would ACI definitely block?

- **a.** APIC-to-leaf cabling
- **b.** Leaf-to-leaf cabling
- **c.** Spine-to-leaf cabling
- **d.** Spine-to-spine cabling

4. Which of the following are valid reasons for ACI Multi-Site requiring more specialized spine hardware? (Choose all that apply.)

- **a.** Ingress replication of BUM traffic
- **b.** IP fragmentation
- **c.** Namespace normalization
- **d.** Support for PIM-Bidir for multicast forwarding

5. Which of the following options best describes border leaf switches?
   a. Border leaf switches provide Layer 2 and 3 connectivity to outside networks.
   b. Border leaf switches connect to Layer 4–7 service appliances, such as firewalls and load balancers.
   c. Border leaf switches are ACI leaf switches that connect to servers.
   d. Border leaf switches serve as the border between server network traffic and FCoE storage traffic.
6. Which of the following statements is accurate?
   a. A three-node M3 cluster of APICs can scale up to 200 leaf switches.
   b. Sharding is a result of the evolution of what is called horizontal partitioning of databases.
   c. The number of shards distributed among APICs for a given attribute is directly correlated to the number of APICs deployed.
   d. A standby APIC actively synchronizes with active APICs and has a copy of all attributes within the APIC database at all times.
7. Out of the following switches, which are spine platforms that support ACI Multi-Site? (Choose all that apply.)
   a. Nexus 93180YC-EX
   b. Nexus 9364C
   c. Nexus 9736C-FX line card
   d. Nexus 9396PX
8. Which of the following is a valid reason for upgrading a pair of Nexus 9336PQ ACI switches to second-generation Nexus 9332C spine hardware? (Choose all that apply.)
   a. Namespace normalization for ACI Multi-Site support
   b. Support for 40 Gbps leaf-to-spine connectivity
   c. Support for CloudSec
   d. Support for ACI Multi-Pod
9. True or false: The Nexus 93180YC-FX leaf switch supports MACsec.
   a. True
   b. False
10. Which of the following platforms is a low-cost option for server CIMC and other low-bandwidth functions that rely on RJ-45 connectivity?
   a. Nexus 9336C-FX2
   b. Nexus 93180YC-FX
   c. Nexus 9332C
   d. Nexus 9348GC-FXP

Foundation Topics

ACI Topologies and Components

Like many other current data center fabrics, ACI fabrics conform to a Clos-based leaf-and-spine topology.

In ACI, leaf and spine switches are each responsible for different functions. Together, they create an architecture that is highly standardized across deployments. Cisco has introduced several new connectivity models and extensions for ACI fabrics over the years, but none of these changes break the core ACI topology that has been the standard from day one. Any topology modifications introduced in this section should therefore be seen as slight enhancements that help address specific use cases and not as deviations from the standard ACI topology.

Clos Topology

In his 1952 paper titled "A Study of Non-blocking Switching Networks," Bell Laboratories researcher Charles Clos formalized how multistage telephone switching systems could be built to forward traffic, regardless of the number of calls served by the overall system.

The mathematical principles proposed by Clos also help address the challenge of needing to build highly scalable data centers using relatively low-cost switches.

Figure 2-1 illustrates a three-stage Clos fabric consisting of one layer for ingress traffic, one layer for egress traffic, and a central layer for forwarding traffic between the layers. Multi-stage designs such as this can result in networks that are not oversubscribed or that are very close to not being oversubscribed.

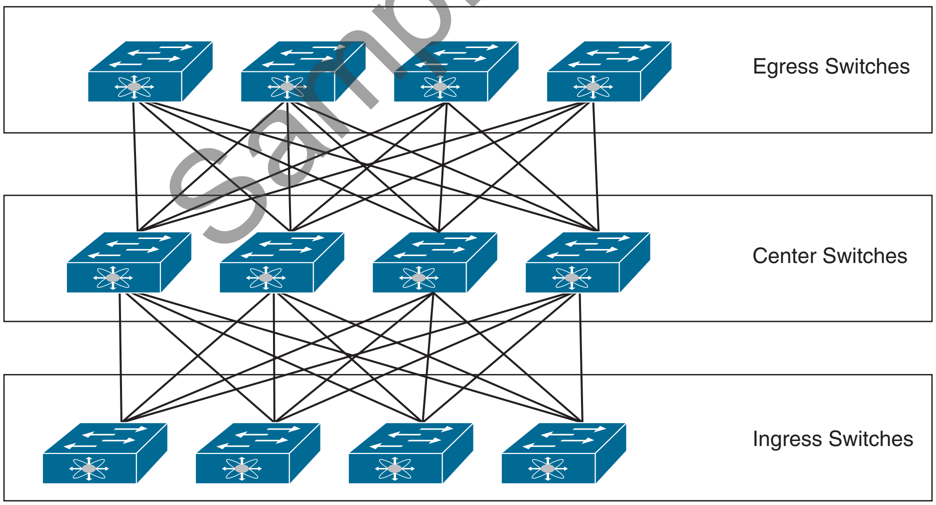

Figure 2-1 *Conceptual View of a Three-Stage Clos Topology*

Modern data center switches forward traffic at full duplex. Therefore, there is little reason to depict separate layers for ingress and egress traffic. It is possible to fold the top layer from the three-tier Clos topology in Figure 2-1 into the bottom layer to achieve what the industry refers to as a "folded" Clos topology, illustrated in Figure 2-2.

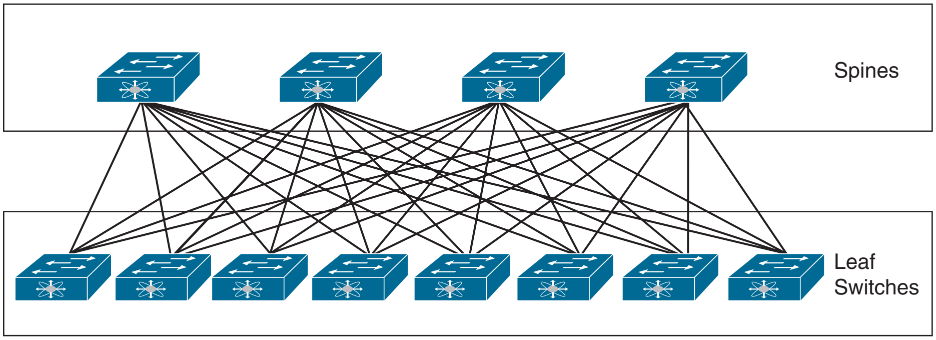

Figure 2-2 *Folded Clos Topology*

As indicated in Figure 2-2, a leaf switch is an ingress/egress switch. A spine switch is an intermediary switch whose most critical function is to perform rapid forwarding of traffic between leaf switches. Leaf switches connect to spine switches in a full-mesh topology.

NOTE At first glance, a three-tier Clos topology may appear to be similar to the traditional three-tier data center architecture. However, there are some subtle differences. First, there are no physical links between leaf switches in the Clos topology. Second, there are no physical links between spine switches. The elimination of cross-links within each layer simplifies network design and reduces control plane complexity.

Standard ACI Topology

An ACI fabric forms a Clos-based spine-and-leaf topology and is usually depicted using two rows of switches. Depending on the oversubscription and overall network throughput requirements, the number of spines and leaf switches will be different in each ACI fabric.

NOTE In the context of the Implementing Cisco Application Centric Infrastructure DCACI 300-620 exam, it does not matter whether you look at a given ACI fabric as a two-tiered Clos topology or as a three-tiered folded Clos topology. It is common for the standard ACI topology to be referred to as a two-tier spine-and-leaf topology.

Figure 2-3 shows the required components and cabling for an ACI fabric. Inheriting from its Clos roots, no cables should be connected between ACI leaf switches. Likewise, ACI spines being cross-cabled results in ACI disabling the cross-connected ports. While the topology shows a full mesh of cabling between the spine-and-leaf layers, a fabric can operate without a full mesh. However, a full mesh of cables between layers is still recommended.

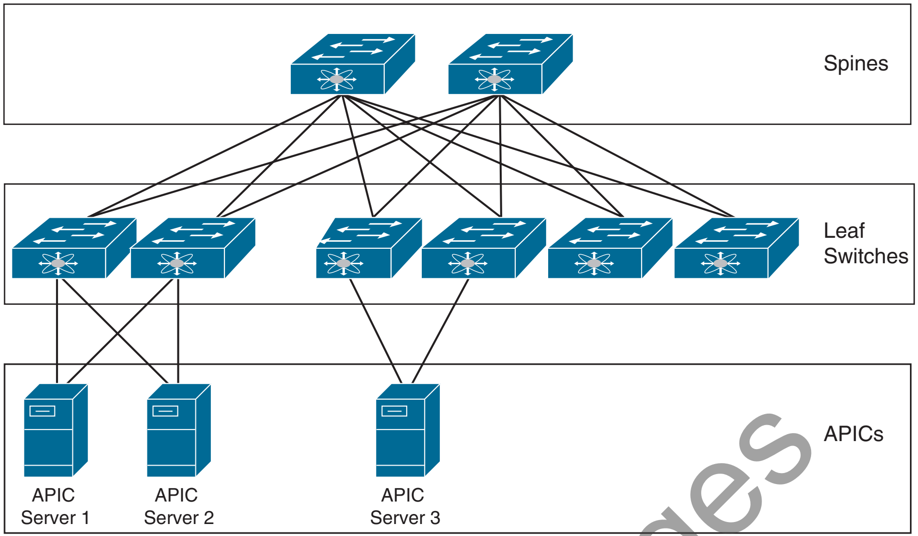

Figure 2-3 *Standard ACI Fabric Topology*

In addition to optics and cabling, the primary hardware components required to build an ACI fabric are as follows:

Key Topic

- **Application Policy Infrastructure Controllers (APICs):** The APICs are the brains of an ACI fabric and serve as the single source of truth for configuration within the fabric. A clustered set of (typically three) controllers attaches directly to leaf switches and provides management, policy programming, application deployment, and health monitoring for an ACI fabric. Note in Figure 2-3 that APICs are not in the data path or the forwarding topology. Therefore, the failure of one or more APICs does not halt packet forwarding. An ACI fabric requires a minimum of one APIC, but an ACI fabric with one APIC should be used only for lab purposes.

- **Spine switches:** ACI spine switches are Clos intermediary switches that have a number of key functions. They exchange routing updates with leaf switches via Intermediate System-to-Intermediate System (IS-IS) and perform rapid forwarding of packets between leaf switches. They provide endpoint lookup services to leaf switches through the Council of Oracle Protocol (COOP). They also handle route reflection to leaf switches using Multiprotocol BGP (MP-BGP), allowing external routes to be distributed across the fabric regardless of the number of tenants. (All three of these are control plane protocols and are covered in more detail in future chapters.) Spine switches also serve as roots for multicast trees within a fabric. By default, all spine switch interfaces besides the mgmt0 port are configured as fabric ports. ***Fabric ports*** are the interfaces that are used to interconnect spine and leaf switches within a fabric.

- **Leaf switches:** Leaf switches are the ingress/egress points for traffic into and out of an ACI fabric. As such, they are the connectivity points for endpoints, including servers and appliances, into the fabric. Layer 2 and 3 connectivity from the outside world into an ACI fabric is also typically established via leaf switches. ACI security policy enforcement occurs on leaf switches. Each leaf switch has a number of high-bandwidth uplink ports preconfigured as fabric ports.

In addition to the components mentioned previously, optional hardware components that can be deployed alongside an ACI fabric include fabric extenders (FEX). Use of FEX solutions in ACI is not ideal because leaf hardware models currently on the market are generally low cost and feature heavy compared to FEX technology.

FEX attachment to ACI is still supported to allow for migration of brownfield gear into ACI fabrics. The DCACI 300-620 exam does not cover specific FEX model support, so neither does this book.

NOTE There are ways to extend an ACI fabric into a virtualized environment by using ACI Virtual Edge (AVE) and Application Virtual Switch (AVS). These are software rather than hardware components and are beyond the scope of the DCACI 300-620 exam.

Engineers may sometimes dedicate two or more leaf switches to a particular function. Engineers typically evaluate the following categories of leaf switches as potential options for dedicating hardware:

Key Topic

- **Border Leaf:** *Border leaf* switches provide Layer 2 and 3 connectivity between an ACI fabric and the outside world. Border leaf switches are sometimes points of policy enforcement between internal and external endpoints.
- **Service Leaf:** *Service leaf* switches are leaf switches that connect to Layer 4–7 service appliances, such as firewalls and load balancers.
- **Compute Leaf:** *Compute leaf* switches are ACI leaf switches that connect to servers. Compute leaf switches are points of policy enforcement when traffic is being sent between local endpoints.
- **IP Storage Leaf:** *IP storage leaf* switches are ACI leaf switches that connect to IP storage systems. IP storage leaf switches can also be points of policy enforcement for traffic to and from local endpoints.

There are scalability benefits associated with dedicating leaf switches to particular functions, but if the size of the network does not justify dedicating leaf switches to a function, consider at least dedicating a pair of leaf switches as border leaf switches. Service leaf functionality can optionally be combined with border leaf functionality, resulting in the deployment of a pair (or more) of collapsed border/service leaf switches in smaller environments.

Cisco publishes a Verified Scalability Guide for each ACI code release. At the time of this writing, 500 is considered the maximum number of leaf switches that can be safely deployed in a single fabric that runs on the latest code.

ACI Stretched Fabric Topology

A *stretched ACI fabric* is a partially meshed design that connects ACI leaf and spine switches distributed in multiple locations. The stretched ACI fabric design helps lower deployment costs when full-mesh cable runs between all leaf and spine switches in a fabric tend to be cost-prohibitive.

Figure 2-4 shows a stretched ACI fabric across two sites.

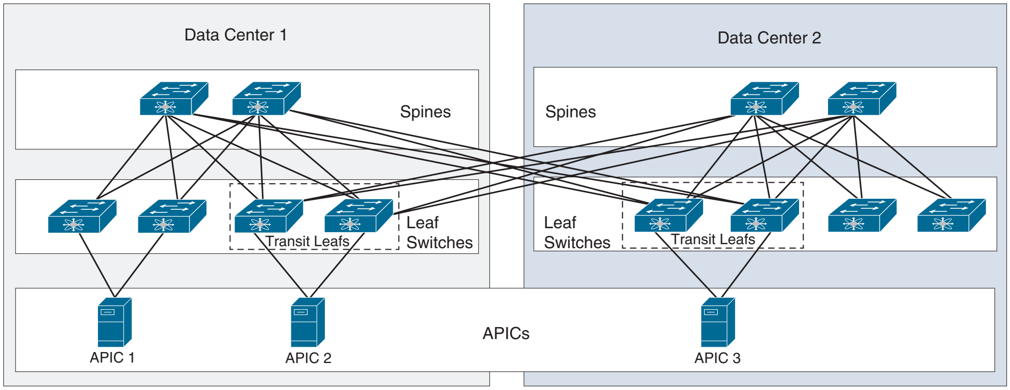

Figure 2-4 *ACI Stretched Fabric Topology*

A stretched fabric amounts to a single administrative domain and a single availability zone. Because APICs in a stretched fabric design tend to be spread across sites, cross-site latency is an important consideration. APIC clustering has been validated across distances of 800 kilometers between two sites.

A new term introduced in Figure 2-4 is *transit leaf*. A ***transit leaf*** is a leaf switch that provides connectivity between two sites in a stretched fabric design. Transit leaf switches connect to spine switches in both sites. No special configuration is required for transit leaf switches. At least one transit leaf switch must be provisioned in each site for redundancy reasons.

While stretched fabrics simplify extension of an ACI fabric, this design does not provide the benefits of newer topologies such as ACI Multi-Pod and ACI Multi-Site and stretched fabrics are therefore no longer commonly deployed or recommended.

ACI Multi-Pod Topology

Key Topic

The ***ACI Multi-Pod*** topology is a natural evolution of the ACI stretched fabric design in which spine and leaf switches are divided into pods, and different instances of IS-IS, COOP, and MP-BGP protocols run inside each pod to enable a level of control plane fault isolation.

Spine switches in each pod connect to an interpod network (IPN). Pods communicate with one another through the IPN. Figure 2-5 depicts an ACI Multi-Pod topology.

Key Topic

An ACI Multi-Pod IPN has certain requirements that include support for OSPF, end-to-end IP reachability, DHCP relay capabilities on the last-hop routers that connect to spines in each pod, and an increased maximum transmission unit (MTU). In addition, a Multi-Pod IPN needs to support forwarding of multicast traffic (PIM-Bidir) to allow the replication of broadcast, unknown unicast, and multicast (BUM) traffic across pods.

One of the most significant use cases for ACI Multi-Pod is active/active data center design. Although ACI Multi-Pod supports a maximum round-trip time latency of 50 milliseconds between pods, most Multi-Pod deployments are often built to achieve active/active functionality and therefore tend to have latencies of less than 5 milliseconds.

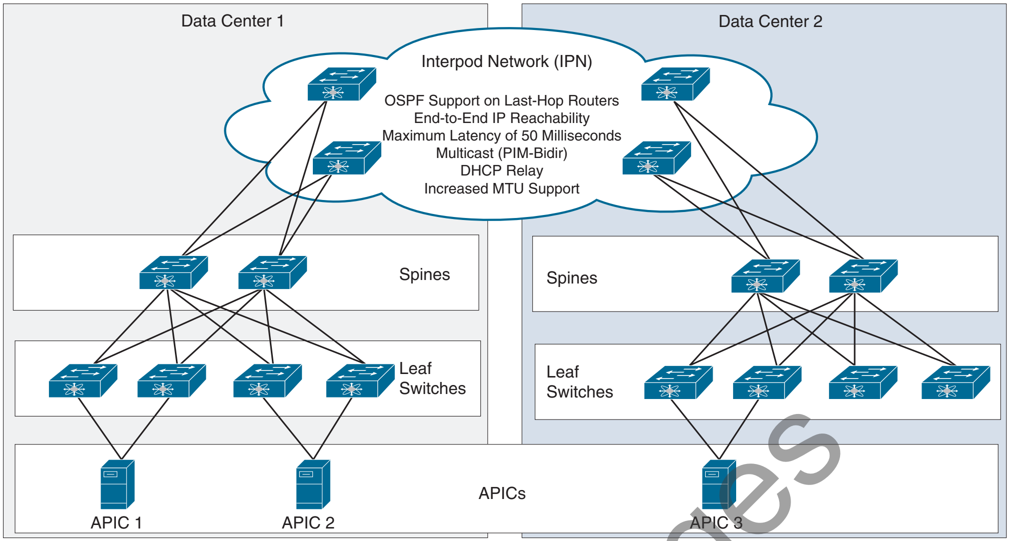

Figure 2-5 *ACI Multi-Pod Topology*

NOTE Another solution that falls under the umbrella of ACI Multi-Pod is Virtual Pod (vPod). ACI vPod is not a new topology per se. It is an extension of a Multi-Pod fabric in the form of a new pod at a remote location where at least two ESXi servers are available, and deployment of ACI hardware is not desirable. ACI vPod components needed at the remote site for this solution include virtual spine (vSpine) appliances, virtual leaf (vLeaf) appliances, and the Cisco ACI Virtual Edge. ACI vPod still requires a physical ACI footprint since vPod is managed by the overall Multi-Pod APIC cluster.

On the issue of scalability, it should be noted that as of the time of writing, 500 is the maximum number of leaf switches that can be safely deployed within a single ACI fabric. However, the Verified Scalability Guide for the latest code revisions specifies 400 as the absolute maximum number of leaf switches that can be safely deployed in each pod. Therefore, for a fabric to reach its maximum supported scale, leaf switches should be deployed across at least 2 pods within a Multi-Pod fabric. Each pod supports deployment of 6 spines, and each Multi-Pod fabric currently supports the deployment of up to 12 pods.

Chapter 16, "ACI Anywhere," covers ACI Multi-Pod in more detail. For now, understand that Multi-Pod is functionally a single fabric and a single availability zone, even though it does not represent a single network failure domain.

ACI Multi-Site Topology

Key Topic

ACI Multi-Site is a solution that interconnects multiple ACI fabrics for the purpose of homogenous policy deployment across ACI fabrics, homogenous security policy deployment across on-premises ACI fabrics and public clouds, and cross-site stretched subnet capabilities, among others.

Key Topic

In an ACI Multi-Site design, each ACI fabric has its own dedicated APIC cluster. A clustered set of three nodes called Multi-Site Orchestrator (MSO) establishes API calls to each fabric independently and can configure tenants within each fabric with desired policies.

NOTE Nodes forming an MSO cluster have traditionally been deployed as VMware ESXi virtual machines (VMs). Cisco has recently introduced the ability to deploy an MSO cluster as a distributed application (.aci format) on Cisco Application Services Engine (ASE). Cisco ASE is a container-based solution that provides a common platform for deploying and managing Cisco data center applications. ASE can be deployed in three form factors: a physical form factor consisting of bare-metal servers, a virtual machine form factor for on-premises deployments via ESXi or Linux KVM hypervisors, and a virtual machine form factor deployable within a specific Amazon Web Services (AWS) region.

Figure 2-6 shows an ACI Multi-Site topology that leverages a traditional VM-based MSO cluster.

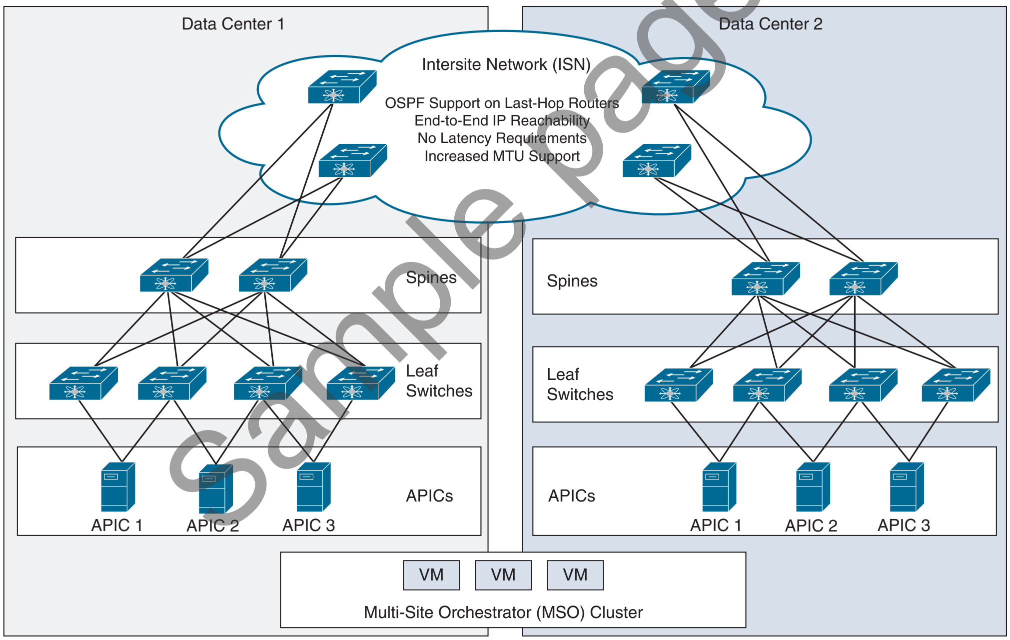

Figure 2-6 *ACI Multi-Site Topology*

Key Topic

As indicated in Figure 2-6, end-to-end communication between sites in an ACI Multi-Site design requires the use of an intersite network (ISN). An ACI Multi-Site ISN faces less stringent requirements compared to ACI Multi-Pod IPNs. In an ISN, end-to-end IP connectivity between spines across sites, OSPF on the last-hop routers connecting to the spines, and increased MTU support allowing VXLAN-in-IP encapsulation are all still required. However, ACI Multi-Site does not dictate any cross-site latency requirements, nor does it require support for multicast or DHCP relay within the ISN.

ACI Multi-Site does not impose multicast requirements on the ISN because ACI Multi-Site has been designed to accommodate larger-scale ACI deployments that may span the globe. It is not always feasible or expected for a company that has a global data center footprint to also have a multicast backbone spanning the globe and between all data centers.

Key Topic

Due to the introduction of new functionalities that were not required in earlier ACI fabrics, Cisco introduced a second generation of spine hardware. Each ACI fabric within an ACI Multi-Site design requires at least one second-generation or newer piece of spine hardware for the following reasons:

- **Ingress replication of BUM traffic:** To accommodate BUM traffic forwarding between ACI fabrics without the need to support multicast in the ISN, Multi-Site-enabled spines perform ingress replication of BUM traffic. This function is supported only on second-generation spine hardware.
- **Cross-fabric namespace normalization:** Each ACI fabric has an independent APIC cluster and therefore an independent brain. When policies and parameters are communicated between fabrics in VXLAN header information, spines receiving cross-site traffic need to have a way to swap remotely significant parameters, such as VXLAN network identifiers (VNIDs), with equivalent values for the local site. This function, which is handled in hardware and is called *namespace normalization*, requires second-generation or newer spines.

Note that in contrast to ACI Multi-Site, ACI Multi-Pod *can* be deployed using first-generation spine switches.

For ACI Multi-Site deployments, current verified scalability limits published by Cisco suggest that fabrics with stretched policy requirements that have up to 200 leaf switches can be safely incorporated into ACI Multi-Site. A single ACI Multi-Site deployment can incorporate up to 12 fabrics as long as the total number of leaf switches in the deployment does not surpass 1600.

Each fabric in an ACI Multi-Site design forms a separate network failure domain and a separate availability zone.

ACI Multi-Tier Architecture

Introduced in Release 4.1, ACI Multi-Tier provides the capability for vertical expansion of an ACI fabric by adding an extra layer or tier of leaf switches below the standard ACI leaf layer.

With the Multi-Tier enhancement, the standard ACI leaf layer can also be termed the Tier 1 leaf layer. The new layer of leaf switches that are added to vertically expand the fabric is called the Tier 2 leaf layer. Figure 2-7 shows these tiers. APICs, as indicated, can attach to either Tier 1 or Tier 2 leaf switches.

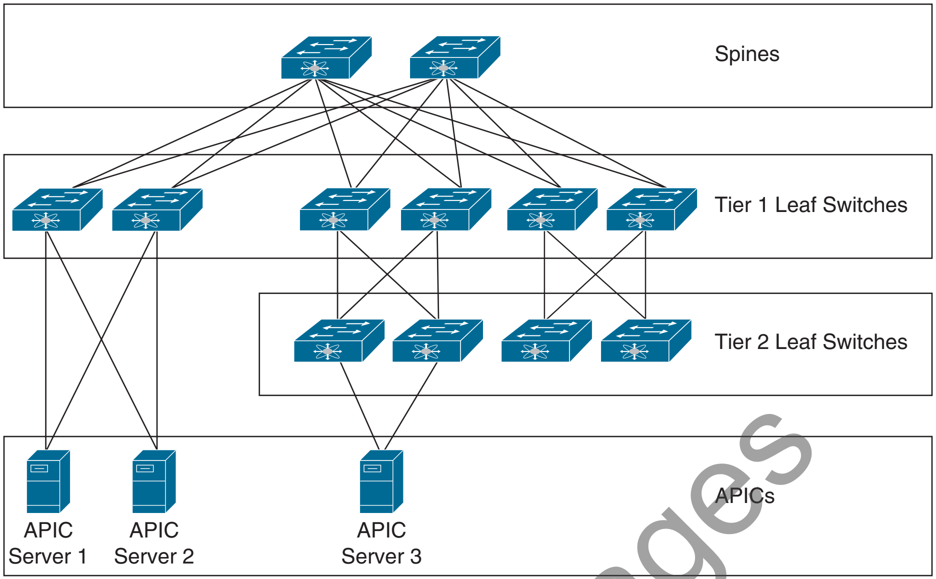

Figure 2-7 *ACI Multi-Tier Topology*

NOTE The topology shown in Figure 2-7 goes against the requirement outlined earlier in this chapter, in the section "Standard ACI Topology," *not* to cross-connect leaf switches. The ACI Multi-Tier architecture is an exception to this rule. Leaf switches within each tier, however, still should never be cross-connected.

An example of a use case for ACI Multi-Tier is the extension of an ACI fabric across data center halls or across buildings that are in relatively close proximity while minimizing long-distance cabling and optics requirements. Examine the diagram in Figure 2-8. Suppose that an enterprise data center has workloads in an alternate building. In this case, the company can deploy a pair of Tier 1 leaf switches in the new building and expand the ACI fabric to the extent needed within the building by using a Tier 2 leaf layer. Assuming that 6 leaf switches would have been required to accommodate the port requirements in the building, as Figure 2-8 suggests, directly cabling these 6 leaf switches to the spines as Tier 1 leaf switches would have necessitated 12 cross-building cables. However, the use of an ACI Multi-Tier design enables the deployment of the same number of switches using 4 long-distance cable runs.

ACI Multi-Tier can also be an effective solution for use within data centers in which the cable management strategy is to minimize inter-row cabling and relatively low-bandwidth requirements exist for top-of-rack switches. In such a scenario, Tier 1 leaf switches can be deployed end-of-row, and Tier 2 leaf switches can be deployed top-of-rack.

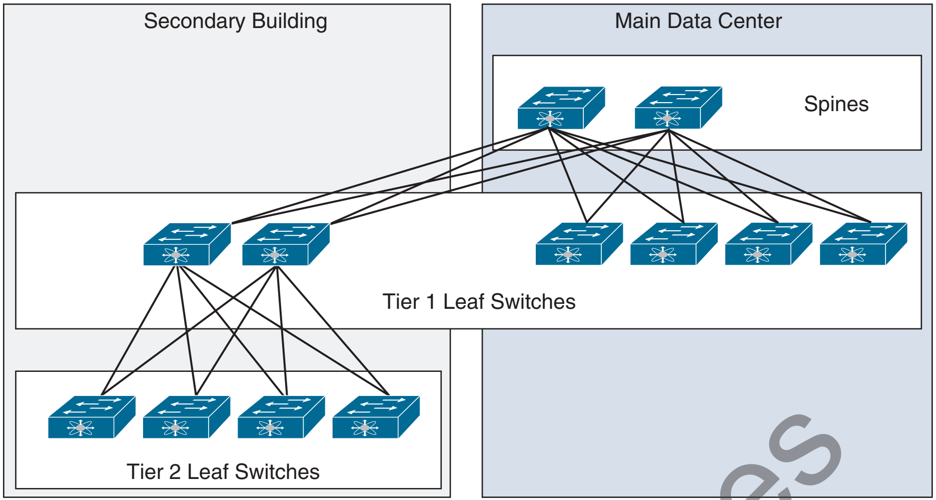

Figure 2-8 *Extending an ACI Fabric by Using ACI Multi-Tier in an Alternative Location*

NOTE ACI Multi-Tier *might not* be a suitable solution if the amount of bandwidth flowing upstream from Tier 2 leaf switches justifies the use of dedicated uplinks to spines.

Not all ACI switch platforms support Multi-Tier functionality.

Remote Leaf Topology

Key Topic

For remote sites in which data center endpoints may be deployed but their number and significance do not justify the deployment of an entirely new fabric or pod, the ACI *Remote Leaf* solution can be used to extend connectivity and ensure consistent policies between the main data center and the remote site. With such a solution, leaf switches housed at the remote site communicate with spines and APICs at the main data center over a generic IPN. Each Remote Leaf switch can be bound to a single pod.

There are three main use cases for Remote Leaf deployments:

- **Satellite/small colo data centers:** If a company has a small data center consisting of several top-of-rack switches and the data center may already have dependencies on a main data center, this satellite data center can be integrated into the main data center by using the Remote Leaf solution.
- **Data center extension and migrations:** Cross-data center migrations that have traditionally been done through Layer 2 extension can instead be performed by deploying a pair of Remote Leafs in the legacy data center. This approach often has cost benefits compared to alternative Layer 2 extension solutions if there is already an ACI fabric in the target state data center.
- **Telco 5G distributed data centers:** Telcom operators that are transitioning to more distributed mini data centers to bring services closer to customers but still desire centralized management and consistent policy deployment across sites can leverage Remote Leaf for these mini data centers.

In addition to these three main use cases, disaster recovery (DR) is sometimes considered a use case for Remote Leaf deployments, even though DR is a use case more closely aligned with ACI Multi-Site designs.

In a Remote Leaf solution, the APICs at the main data center deploy policy to the Remote Leaf switches as if they were locally connected.

Figure 2-9 illustrates a Remote Leaf solution.

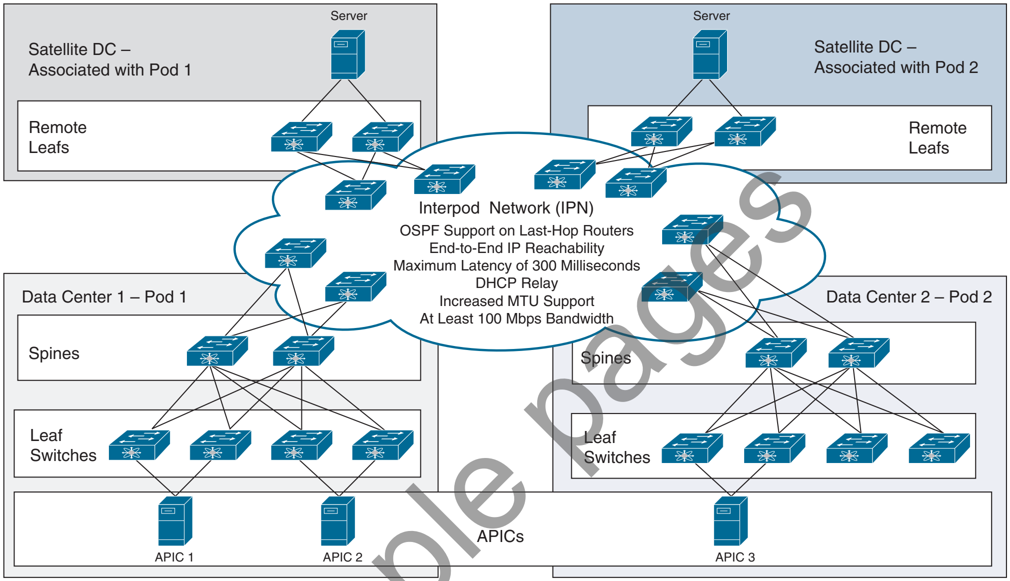

Figure 2-9 *Remote Leaf Topology and IPN Requirements*

IPN requirements for a Remote Leaf solution are as follows:

- **MTU:** The solution must support an end-to-end MTU that is at least 100 bytes higher than that of the endpoint source traffic. Assuming that 1500 bytes has been configured for data plane MTU, Remote Leaf can be deployed using a minimum MTU of 1600 bytes. An IPN MTU this low, however, necessitates that ACI administrators lower the ACI fabricwide control plane MTU, which is 9000 bytes by default.
- **Latency:** Up to 300 milliseconds latency between the main data center and remote location is acceptable.
- **Bandwidth:** Remote Leaf is supported with a minimum IPN bandwidth of 100 Mbps.
- **VTEP reachability:** A Remote Leaf switch logically associates with a single pod if integrated into a Multi-Pod solution. To make this association possible, the Remote Leaf should be able to route traffic over the IPN to the VTEP pool of the associated pod. Use of a dedicated VRF for IPN traffic is recommended where feasible.

- **APIC infra IP reachability:** A Remote Leaf switch needs IP connectivity with all APICs in a Multi-Pod cluster at the main data center. If an APIC has assigned itself IP addresses from a VTEP range different than the pod VTEP pool, the additional VTEP addresses need to also be advertised over the IPN.
- **OSPF support on upstream routers:** Routers northbound of both the Remote Leaf switches and the spine switches need to support OSPF and must be able to encapsulate traffic destined to directly attached ACI switches using VLAN 4. This requirement exists only for directly connected devices and does not extend end-to-end in the IPN.
- **DHCP relay:** The upstream router directly connected to Remote Leaf switches needs to enable DHCP relay to relay DHCP packets to the APIC IP addresses in the infra tenant. The DHCP relay configuration needs to be applied on the VLAN 4 subinterface or SVI.

Note that unlike a Multi-Pod IPN, a Remote Leaf IPN does not require Multicast PIM-Bidir support. This is because the Remote Leaf solution uses headend replication (HER) tunnels to forward BUM traffic between sites.

In a Remote Leaf design, traffic between known local endpoints at the remote site is switched directly, whether physically or virtually. Any traffic whose destination is in ACI but is unknown or not local to the remote site is forwarded to the main data center spines.

NOTE Chapter 16 details MTU requirements for IPN and ISN environments for ACI Multi-Pod and ACI Multi-Site. It also covers how to lower control plane and data plane MTU values within ACI if the IPN or ISN does not support high MTU values. Although it does not cover Remote Leaf, the same general IPN MTU concepts apply.

Not all ACI switches support Remote Leaf functionality. The current maximum verified scalability number for Remote Leaf switches is 100 per fabric.

APIC Clusters

The ultimate size of an APIC cluster should be directly proportionate to the size of the Cisco ACI deployment. From a management perspective, any active APIC controller in a cluster can service any user for any operation. Controllers can be transparently added to or removed from a cluster.

Key Topic

APICs can be purchased either as physical or virtual appliances. Physical APICs are 1 rack unit (RU) Cisco C-Series servers with ACI code installed and come in two different sizes: M for medium and L for large. In the context of APICs, "size" refers to the scale of the fabric and the number of endpoints. Virtual APICs are used in ACI mini deployments, which consist of fabrics with up to two spine switches and four leaf switches.

Key Topic

As hardware improves, Cisco releases new generations of APICs with updated specifications. At the time of this writing, Cisco has released three generations of APICs. The first generation of APICs (M1/L1) shipped as Cisco UCS C220 M3 servers. Second-generation APICs (M2/L2) were Cisco UCS C220 M4 servers. Third-generation APICs (M3/L3) are shipping as UCS C220 M5 servers.