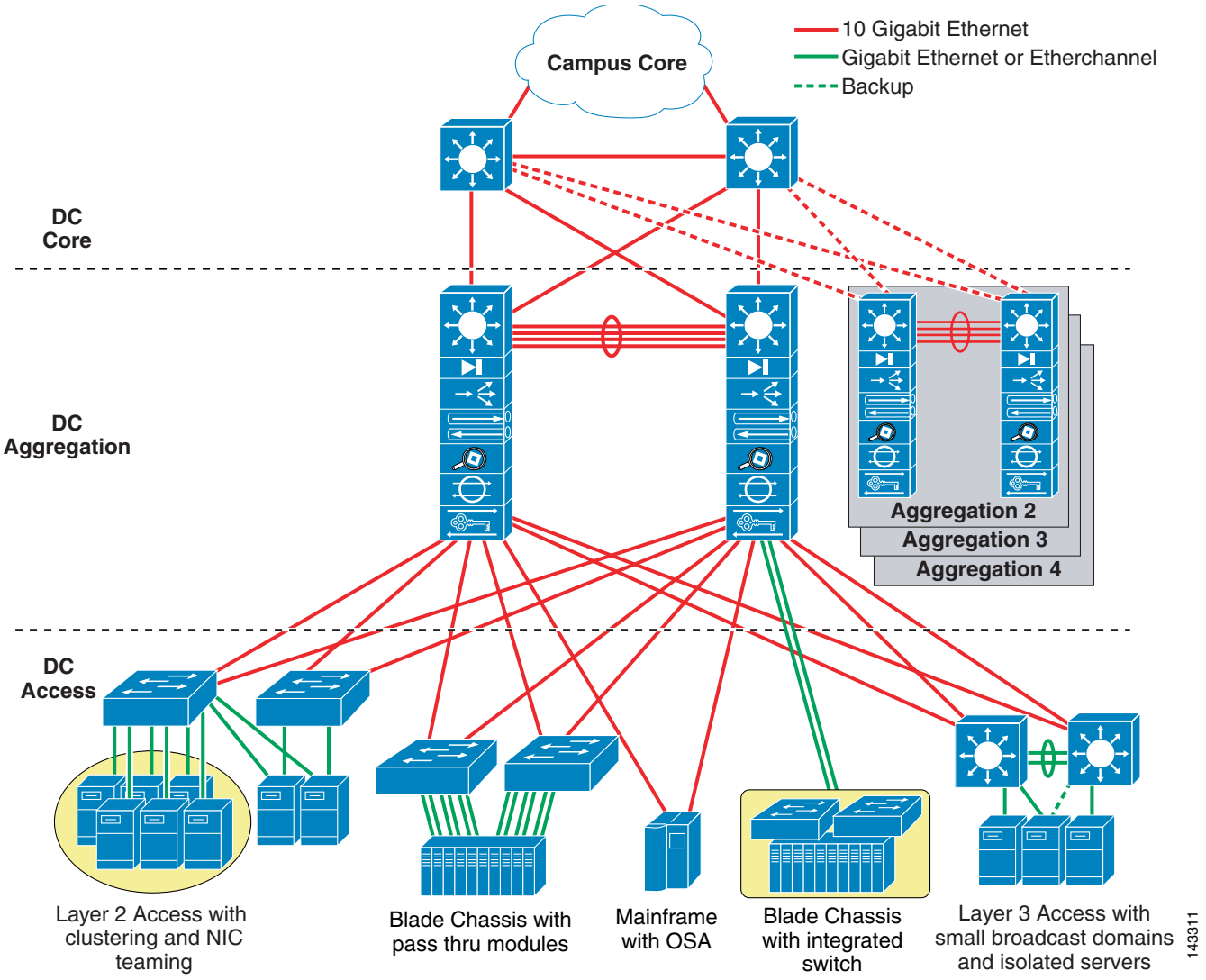
Blade Servers in the Data Center—Overview

Data Center Multi-tier Model Overview

The data center multi-tier model is a common enterprise design that defines logical tiers addressing web, application, and database functionality. The multi-tier model uses network services to provide application optimization and security.

Figure 1-1 shows a generic multi-tier data center architecture.

Figure 1-1 Data Center Multi-tier Model



The layers of the data center design are the *core, aggregation,* and *access* layers. These layers are referred to throughout this SRND and are briefly described as follows:

- Core layer—Provides the high-speed packet switching backplane for all flows going in and out of the data center. The core layer provides connectivity to multiple aggregation modules and provides a resilient Layer 3 routed fabric with no single point of failure. The core layer runs an interior routing protocol such as OSPF or EIGRP, and load balances traffic between the campus core and aggregation layers using Cisco Express Forwarding-based hashing algorithms.
- Aggregation layer modules—Provides important functions such as service module integration, Layer 2 domain definitions, spanning tree processing, and default gateway redundancy. Server-to-server multi-tier traffic flows through the aggregation layer and may use services such as firewall and server load balancing to optimize and secure applications. The smaller icons within the aggregation layer switch in Figure 1-1 represent the integrated service modules, which provide services that include content switching, firewall, SSL offload, intrusion detection, and network analysis.

- Access layer—Location where the servers physically attach to the network. The server components consist of 1RU servers, blade servers with integral switches, blade servers with pass-through cabling, clustered servers, and mainframes with OSA adapters. The access layer network infrastructure consists of modular switches, fixed configuration 1 or 2RU switches, and integral blade server switches. Switches provide both Layer 2 and Layer 3 topologies, fulfilling the various server broadcast domain or administrative requirements.

The multi-tier data center is a flexible, robust environment capable of providing high availability, scalability, and critical network services to data center applications with diverse requirements and physical platforms. This document focuses on the integration of blade servers into the multi-tier data center model. For more details on the Cisco Data Center infrastructure, see the *Data Center Infrastructure SRND 2.5* at the following URL: http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCI_SRND_2_5_book.html.

Blade Server Integration Options

Blade systems are the latest server platform emerging in the data center. Enterprise data centers seek the benefits that this new platform can provide in terms of power, cooling, and server consolidation that optimize the compute power per rack unit. Consequently, successfully incorporating these devices into the data center network architecture becomes a key consideration for network administrators.

The following section is an overview of the options available to integrate blade systems into the data center. The following topics are included:

- Integrated Switches
- Pass-Through Technology

Integrated Switches

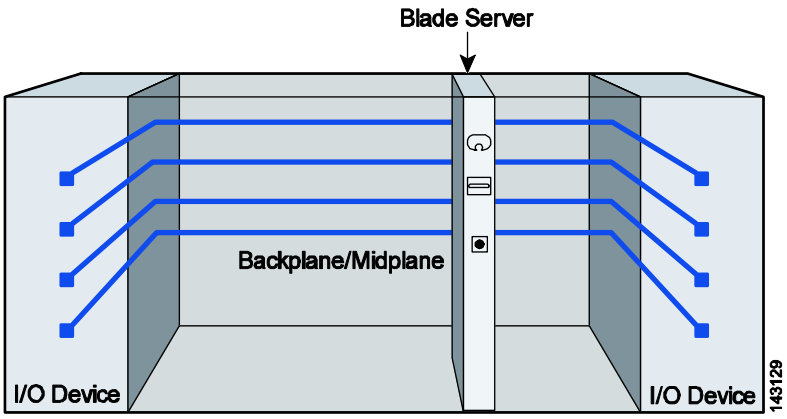
Blade systems allow built-in switches to control traffic flow between the blade servers within the chassis and the remaining enterprise network. Blade systems provide a variety of switch media types, including the following:

- Built-in Ethernet switches (such as the Cisco Ethernet Switch Modules)
- Infiniband switches (such as the Cisco Server Fabric Switch)
- Fibre Channel switches

Integrated switches are a passageway to the blade servers within the chassis and to the data center. As illustrated in Figure 1-2, each blade server connects to a backplane or a mid-plane that typically contains four dedicated signaling paths to redundant network devices housed in the chassis. This predefined physical structure reduces the number of cables required by each server and provides a level of resiliency via the physical redundancy of the network interface controllers (NICs) and I/O network devices.

Note The predefined connectivity of a blade system has NIC teaming implications. Therefore, network administrators must consider this when determining their blade server high availability strategy.

Figure 1-2 Sample Blade System Internal Connection



Example of a Blade System Backplane

Note The chassis illustrated in Figure 1-2 is for demonstration purposes. Chassis details differ between blade system vendors.

Introducing a blade server system that uses built-in Ethernet switches into the IP infrastructure of the data center presents many options to the network administrator, such as the following:

- Where is the most appropriate attachment point—the aggregation or access layer?
- What features are available on the switch, such as Layer 2 or trunk failover?
- What will the impact be to the Layer 2 and Layer 3 topologies?
- Will NIC teaming play a role in the high availability design?
- What will the management network look like?

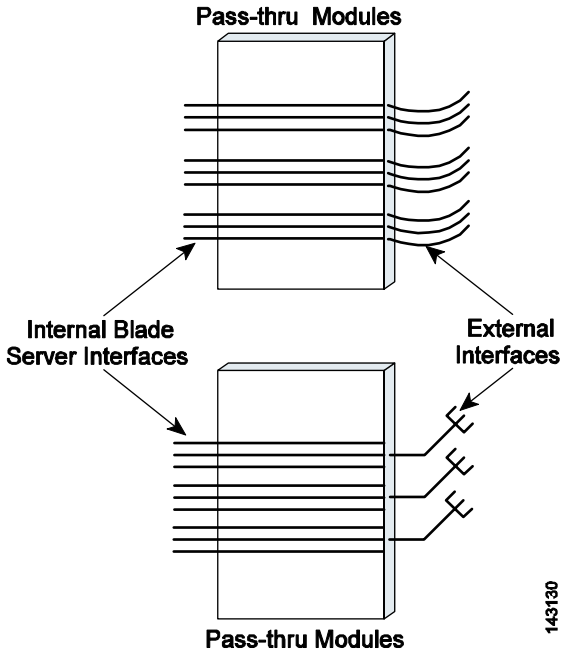
These topics are addressed in Chapter 2, "Integrated Switch Technology."

Pass-Through Technology

Pass-through technology is an alternative method of network connectivity that allows individual blade servers to communicate directly with external resources. Both copper and optical pass-through modules that provide access to the blade server controllers are available.

Figure 1-3 shows two common types of pass-through I/O devices. Each of these provides connectivity to the blade servers via the backplane or mid-plane of the chassis. There is a one-to-one relationship between the number of server interfaces and the number of external ports in the access layer that are necessary to support the blade system. Using an octopus cable changes the one-to-one ratio, as shown by the lower pass-through module in Figure 1-3.

Figure 1-3 Pass-Through Module Examples



Pass-through modules are passive devices that simply expose the blade server NIC to the external network. They do not require configuration by the network administrator. These I/O devices do not require configuration and do not extend the network Layer 2 or Layer 3 topologies. In addition, the blade servers may employ any of the NIC teaming configurations supported by their drivers.

The need to reduce the amount of cabling in the data center is a major influence driving the rapid adoption of blade servers. Pass-through modules do not allow the data center to take full advantage of the cable consolidation the blade platform offers. This lack of cable reduction in the rack, row, or facility often hinders the use of a pass-through based solution in the data center.

Pass-through technology issues are addressed in Chapter 3, "Pass-Through Technology."