Overview

- Overview, on page 1
- Cisco UCS Manager User CLI Documentation, on page 1
- Cisco Unified Computing System Overview, on page 2
- Unified Fabric, on page 4
- Multilayer Network Design, on page 5

Overview

This guide includes the following information:

- Configure/Enable Server Ports; Configure/Enable Uplink Ports; Configure/Enable FC Ports.
- Create LAN Pin Groups
- Create VLANs and VLAN groups
- Create Server Links
- Configure QoS System Class
- Configure Global Policies
- Monitor Network Health
- Traffic Monitoring

Cisco UCS Manager User CLI Documentation

Cisco UCS Manager offers you a new set of smaller, use-case based documentation described in the following table:

Guide	Description
Cisco UCS Manager Getting Started Guide	Discusses Cisco UCS architecture and Day 0 operations, including Cisco UCS Manager initial configuration, and configuration best practices.

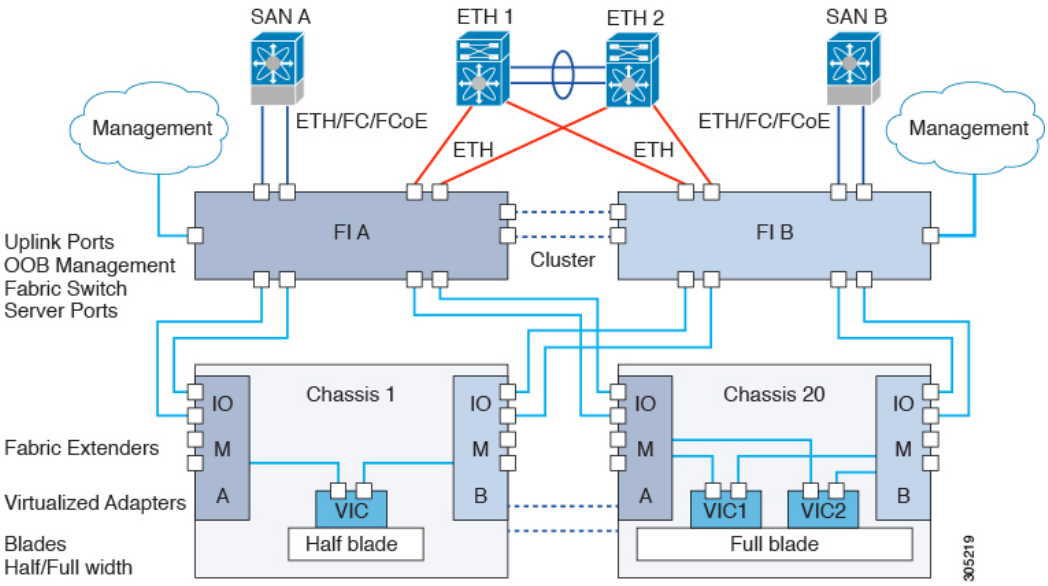
Guide	Description
Cisco UCS Manager Administration Guide	Discusses password management, role-based access configuration, remote authentication, communication services, CIMC session management, organizations, backup and restore, scheduling options, BIOS tokens and deferred deployments.
Cisco UCS Manager Infrastructure Management Guide	Discusses physical and virtual infrastructure components used and managed by Cisco UCS Manager.
Cisco UCS Manager Firmware Management Guide	Discusses downloading and managing firmware, upgrading through Auto Install, upgrading through service profiles, directly upgrading at endpoints using firmware auto sync, managing the capability catalog, deployment scenarios, and troubleshooting.
Cisco UCS Manager Server Management Guide	Discusses the new licenses, registering Cisco UCS domains with Cisco UCS Central, power capping, server boot, server profiles and server-related policies.
Cisco UCS Manager Storage Management Guide	Discusses all aspects of storage management such as SAN and VSAN in Cisco UCS Manager.
Cisco UCS Manager Network Management Guide	Discusses all aspects of network management such as LAN and VLAN connectivity in Cisco UCS Manager.
Cisco UCS Manager System Monitoring Guide	Discusses all aspects of system and health monitoring including system statistics in Cisco UCS Manager.
Cisco UCS S3260 Server Integration with Cisco UCS Manager	Discusses all aspects of management of UCS S-Series servers that are managed through Cisco UCS Manager.

Cisco Unified Computing System Overview

Cisco UCS has a unique architecture that integrates compute, data network access, and storage network access into a common set of components under a single-pane-of-glass management interface.

Cisco UCS fuses access layer networking and servers. This high-performance, next-generation server system provides a data center with a high degree of workload agility and scalability. The hardware and software components support Cisco's unified fabric, which runs multiple types of data center traffic over a single converged network adapter.

Figure 1: Cisco Unified Computing System Architecture



Architectural Simplification

The simplified architecture of Cisco UCS reduces the number of required devices and centralizes switching resources. By eliminating switching inside a chassis, network access-layer fragmentation is significantly reduced. Cisco UCS implements Cisco unified fabric within racks and groups of racks, supporting Ethernet and Fibre Channel protocols over 10/25/40 Gigabit Cisco Data Center Ethernet and Fibre Channel over Ethernet (FCoE) links. This radical simplification reduces the number of switches, cables, adapters, and management points by up to two-thirds. All devices in a Cisco UCS domain remain under a single management domain, which remains highly available through the use of redundant components.

High Availability

The management and data plane of Cisco UCS is designed for high availability and redundant access layer fabric interconnects. In addition, Cisco UCS supports existing high availability and disaster recovery solutions for the data center, such as data replication and application-level clustering technologies.

Scalability

A single Cisco UCS domain supports multiple chassis and their servers, all of which are administered through one Cisco UCS Manager. For more detailed information about the scalability, speak to your Cisco representative.

Flexibility

A Cisco UCS domain allows you to quickly align computing resources in the data center with rapidly changing business requirements. This built-in flexibility is determined by whether you choose to fully implement the stateless computing feature. Pools of servers and other system resources can be applied as necessary to respond to workload fluctuations, support new applications, scale existing software and business services, and accommodate both scheduled and unscheduled downtime. Server identity can be abstracted into a mobile service profile that can be moved from server to server with minimal downtime and no need for additional network configuration.

With this level of flexibility, you can quickly and easily scale server capacity without having to change the server identity or reconfigure the server, LAN, or SAN. During a maintenance window, you can quickly do the following:

- Deploy new servers to meet unexpected workload demand and rebalance resources and traffic.
- Shut down an application, such as a database management system, on one server and then boot it up again on another server with increased I/O capacity and memory resources.

Optimized for Server Virtualization

Cisco UCS has been optimized to implement VM-FEX technology. This technology provides improved support for server virtualization, including better policy-based configuration and security, conformance with a company's operational model, and accommodation for VMware's VMotion.

Unified Fabric

With unified fabric, multiple types of data center traffic can run over a single Data Center Ethernet (DCE) network. Instead of having a series of different host bus adapters (HBAs) and network interface cards (NICs) present in a server, unified fabric uses a single converged network adapter. This type of adapter can carry LAN and SAN traffic on the same cable.

Cisco UCS uses Fibre Channel over Ethernet (FCoE) to carry Fibre Channel and Ethernet traffic on the same physical Ethernet connection between the fabric interconnect and the server. This connection terminates at a converged network adapter on the server, and the unified fabric terminates on the uplink ports of the fabric interconnect. On the core network, the LAN and SAN traffic remains separated. Cisco UCS does not require that you implement unified fabric across the data center.

The converged network adapter presents an Ethernet interface and Fibre Channel interface to the operating system. At the server, the operating system is not aware of the FCoE encapsulation because it sees a standard Fibre Channel HBA.

At the fabric interconnect, the server-facing Ethernet port receives the Ethernet and Fibre Channel traffic. The fabric interconnect (using Ethertype to differentiate the frames) separates the two traffic types. Ethernet frames and Fibre Channel frames are switched to their respective uplink interfaces.

Fibre Channel over Ethernet

Cisco UCS leverages Fibre Channel over Ethernet (FCoE) standard protocol to deliver Fibre Channel. The upper Fibre Channel layers are unchanged, so the Fibre Channel operational model is maintained. FCoE network management and configuration is similar to a native Fibre Channel network.

FCoE encapsulates Fibre Channel traffic over a physical Ethernet link. FCoE is encapsulated over Ethernet with the use of a dedicated Ethertype, 0x8906, so that FCoE traffic and standard Ethernet traffic can be carried on the same link. FCoE has been standardized by the ANSI T11 Standards Committee.

Fibre Channel traffic requires a lossless transport layer. Instead of the buffer-to-buffer credit system used by native Fibre Channel, FCoE depends upon the Ethernet link to implement lossless service.

Ethernet links on the fabric interconnect provide two mechanisms to ensure lossless transport for FCoE traffic:

- Link-level flow control
- Priority flow control

Link-Level Flow Control

IEEE 802.3x link-level flow control allows a congested receiver to signal the endpoint to pause data transmission for a short time. This link-level flow control pauses all traffic on the link.

The transmit and receive directions are separately configurable. By default, link-level flow control is disabled for both directions.

On each Ethernet interface, the fabric interconnect can enable either priority flow control or link-level flow control (but not both).

When an interface on a Cisco UCS 6454 Fabric Interconnect has Priority Flow Control (PFC) admin configured as **auto** and Link-Level Flow Control (LLFC) admin configured as **on**, the PFC operation mode will be **off** and the LLFC operation mode will be **on**. On UCS 6300 Series and earlier Fabric Interconnects, the same configuration will result in the PFC operation mode being **on** and the LLFC operation mode being **off**.

Priority Flow Control

The priority flow control (PFC) feature applies pause functionality to specific classes of traffic on the Ethernet link. For example, PFC can provide lossless service for the FCoE traffic, and best-effort service for the standard Ethernet traffic. PFC can provide different levels of service to specific classes of Ethernet traffic (using IEEE 802.1p traffic classes).

PFC decides whether to apply pause based on the IEEE 802.1p CoS value. When the fabric interconnect enables PFC, it configures the connected adapter to apply the pause functionality to packets with specific CoS values.

By default, the fabric interconnect negotiates to enable the PFC capability. If the negotiation succeeds, PFC is enabled and link-level flow control remains disabled (regardless of its configuration settings). If the PFC negotiation fails, you can either force PFC to be enabled on the interface or you can enable IEEE 802.x link-level flow control.

Multilayer Network Design

When you design a data center using a modular approach, the network is divided into three functional layers: Core, Aggregation, and Access. These layers can be physical or logical, and you can add and remove them without redesigning the entire data center network.

Because of the hierarchical topology of a modular design, the addressing is also simplified within the data center network. Modularity implies isolation of building blocks, which are separated from each other and communicate through specific network connections between the blocks. Modular design provides easy control of traffic flow and improved security. In other words, these blocks are independent from each other; a change in one block does not affect other blocks. Modularity also enables faster moves, adds, and changes (MACs) and incremental changes in the network.

Modular networks are scalable. Scalability allows for the considerable growth or reduction in the size of a network without making drastic changes. Scalable data center network design is achieved by using the principle of hierarchy and modularity.

Keep a network as simple as possible. Modular designs are simple to design, configure, and troubleshoot.

- **Access Layer**—The access layer is the first point of entry into the network for edge devices, end stations, and servers. The Access layer grants user access to network devices and provides connectivity to servers. The switches in the access layer are connected to two separate distribution layer switches for redundancy. The data center access layer provides Layer 2, Layer 3, and mainframe connectivity. The design of the

access layer varies, depending on whether you use Layer 2 or Layer 3 access. The access layer in the data center is typically built at Layer 2, which allows better sharing of service devices across multiple servers. This design also enables the use of Layer 2 clustering, which requires the servers to be Layer 2 adjacent. With Layer 2 access, the default gateway, you can configure for the servers at the aggregation layer.

- **Aggregation Layer**—The aggregation (or distribution) layer aggregates the uplinks from the access layer to the data center core. This layer is the critical point for control and application services. Security and application service devices (such as load-balancing devices, SSL offloading devices, firewalls, and IPS devices) are often deployed as modules in the aggregation layer. The aggregation layer provides policy-based connectivity.
- **Core Layer**—Also known as backbone, the core layer provides high-speed packet switching, scalability and high availability, and fast convergence. Implementing a data center core is a best practice for large data centers. When you implement the core in an initial data center design, it eases network expansion and avoids disruption to the data center environment.

  Use the following criteria to determine whether a core solution is appropriate: The data center typically connects to the campus core using Layer 3 links. The data center network is summarized, and the core injects a default route into the data center network.

  - Ethernet bandwidth requirements
  - Port density
  - Administrative domains
  - Anticipated future development