Cisco Unified Computing System



© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Introduction

Cisco Unified Computing System™ (Cisco UCS®) is an integrated computing infrastructure with intent-based management to automate and accelerate deployment of all your applications, including virtualization and cloud computing, scale-out and bare-metal workloads, and in-memory analytics, as well as edge computing that supports remote and branch locations and massive amounts of data from the Internet of Things (IoT). This document provides an overview of the main components of Cisco UCS: unified fabric, unified management, and unified computing resources.

Since the system was first announced, we have joined the top tier of server vendors. The system has set more than 200 world performance records—all serving as testament to leadership and dedication to customer-centered innovation.

About this document

This document is designed to help you learn what you need to know about Cisco UCS to the level of depth that interests you.

- The challenge describes the stresses facing IT organizations today.
- The solution discusses the most important reasons why Cisco UCS is such a compelling solution for today's challenges.
- Cisco UCS anatomy overview gives a nuts-and-bolts description of Cisco UCS in terms of how our components—Cisco Unified Fabric, Cisco UCS management, and Cisco UCS servers—combine to create a single unified system.
- The remaining sections—Cisco UCS management features and tools, Cisco UCS management concepts, Unified I/O architecture, and Cisco UCS product details and best practices—take a deep dive into the technology behind Cisco UCS and present some best practices.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

The challenge

Digital business transformation is well underway and continues to create new demands on IT organizations:

- **Applications:** Modern applications are becoming less monolithic and more like organic entities that grow and shrink through the use of modular, distributed microservices. This reduces dependence on traditional IT infrastructure and places new demands on IT organizations in terms of the large number of endpoints to manage and flexible infrastructure needed to support them. With agile development and DevOps deployment approaches becoming the norm, developers and administrators demand the capability to program their own infrastructure. This is necessary in order to quickly roll out new applications and updates to existing ones.
- **Management:** The main issue for IT organizations is managing infrastructure at scale and being able to match resources to application requirements—yet manage all of the different types of infrastructure in a simple, holistic fashion. Furthermore, the IT organization's clients—line-of-business managers, application developers, and DevOps teams—are asserting increasing influence over the adoption, purchase, and deployment of technology. IT organizations saddled with ponderous management tools are hard pressed to compete with the fluidity of self-service, hybrid cloud environments that their clients can purchase on their own if the IT organization doesn't keep pace.
- **Location:** Once upon a time traditional IT organizations could support and manage their clients and applications safely within the glass walls of an on-premises data center. Now they must support clients, applications, and workloads in public, managed, edge, private cloud and public cloud environments—all while maintaining compliance with best practices, business and governmental regulations, and data sovereignty requirements.

As a modern IT organization, the challenge is to meet the needs of the digital business transformation while still supporting the traditional, monolithic applications that are the foundation of your business. You have to simultaneously embrace the traditional deployment model and the fluid nature of today's hybrid cloud deployments.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

The solution

The straightforward solution to today's infrastructure challenges is Cisco Unified Computing System (Cisco UCS). It's not a collection of servers. It's a fully self-aware, self-integrating system. Because it is 100 percent programmable, it has been the solution to computing challenges for more than a decade. The system is flexible, agile, and adaptable, and the portfolio of products supported by Cisco UCS includes rack and modular servers, converged and hyperconverged infrastructure, and solutions for AI workloads and the network edge.

There is a fundamental difference between vendors that sell servers and Cisco Unified Computing System. Servers arose as more powerful personal computers, taking many of their attributes, including time-consuming, manual, error-prone configuration of I/O, network, and storage subsystems. Traditional servers are monolithic, complex to deploy, and even more complex to adapt to new workload demands. In contrast, Cisco UCS is a single unified system, with six fundamental attributes that transformed the industry.

Fabric-centric

We blend all of the system's I/O traffic into a single shared active-active network that carries all modes of communication from servers to the outside world. Our low-latency, high-bandwidth network fabric is a shared resource so networking can be allocated to interfaces based on policies rather than physical interface configuration and hard-wired cabling. The result is that you can provision and balance resources to meet your workload needs easily.

Endpoint-aware

Cisco UCS was developed in the virtualization era, where the norm was multiple independent workloads running on the same server. Today the number of workloads has become practically unimaginable as containerized environments place hundreds of workloads in a single virtual machine. Whether you run virtualized, containerized, or bare-metal workloads, all I/O is virtualized. This gives you the capability to support a massive number of endpoints but with a level of control equivalent to each endpoint having its own dedicated (but virtual) cable to the outside world. This gives you the scale of virtualized, with the workload isolation and security of the physical world.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

100 percent programmable

From the very beginning, Cisco UCS was designed with the entire state of each server—identity, configuration, and connectivity—abstracted into software. This makes our system fully composable: adaptable through software to meet the varying requirements ranging from modern cloud-native and AI workloads to traditional monolithic business applications. With a completely programmable system, you can give your clients the level of control they need to manage their workloads. The Cisco Intersight® IT operations platform gives you complete role- and policy-based control over all of your resources regardless of where they reside. Fine-grained infrastructure management can be handled in Agile development and DevOps shops with scripting languages that provide access to the Cisco UCS unified API.

Intent-based

Intersight software helps you to more precisely align your infrastructure with the needs of your business and respond in real time to optimize application performance. It enables administrators to automate configurations or tasks based on specific requirements that are tied to business objectives and application performance. Rather than having to be concerned with every detail of system configuration, intent-based management enables you to describe what you want to accomplish, with the cloud-based automation we provide translating your intent into action.

Delivers business benefits

Our policy-based approach to management gives you the simplicity, automation, and capabilities you need to increase productivity and support a fast-paced business environment.

- **Improved staff productivity:** By aligning your infrastructure with applications and the way teams work together, you can create synergies that can't be achieved with other architectures.
- **Better use of IT staff:** The common, simplified management of servers, storage, and fabric establishes best practices and eliminates the need to understand the nuances of specific components. You can use your subject-matter experts to develop policies and use lower-level administrators or operations personnel to implement policies.
- **Effective communication:** Cisco UCS management improves communication between roles through cross-visibility and role-based administrative access.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

- **Faster time to value:** You can rapidly roll out new applications and business services at cloud-like speed and enhance the competitive strength of your enterprise.
- **Increased operational efficiency:** You can automate many routine tasks, improve resource utilization, and proactively prevent manual errors that typically keep your IT staff from working on high-value tasks.
- **Improved flexibility and agility:** The capability to automatically integrate additional resource capacity into larger, flexible pools helps ensure that your IT staff can achieve economies of scale and efficiency without greater complexity.

Analytics-powered

What if your infrastructure talked directly with your support organization? The recommendation engine built into Intersight software integrates with the Cisco® Technical Assistance Center (TAC) to help you easily detect problems and initiate support requests. As the Cisco Intersight recommendation engine gains intelligence, our vision is for it to provide suggestions and recommendations for you to optimize your configurations to gain the most from your investment.



Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco UCS anatomy overview

Cisco UCS is built using the hierarchy of components illustrated in Figure 1 and described in the sections that follow. Each Cisco UCS connectivity domain is established with a pair of Cisco UCS fabric interconnects, with a comprehensive set of options for connecting rack and modular servers to them either directly or indirectly.

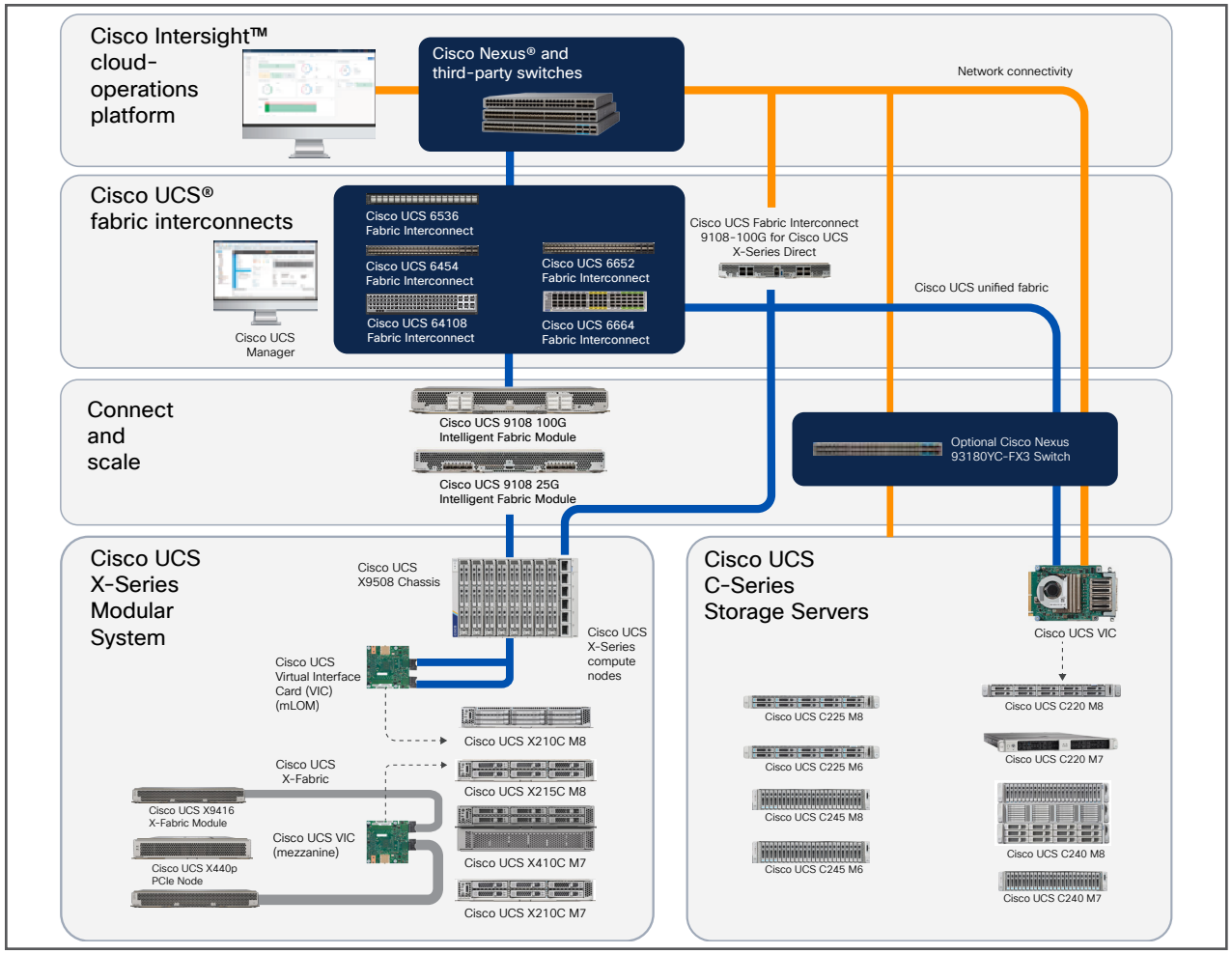


Figure 1 Cisco UCS component hierarchy

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco UCS management

Cisco UCS is based on the concept that infrastructure is code. Servers are designed to be stateless, with their identity, configuration, and connectivity extracted into variables that can be set through software. This enables management tools to help guarantee consistent, error-free, policy-based alignment of server personalities with workloads. Through automation, transforming the server and networking components of your infrastructure into a complete solution is fast and error-free because programmability eliminates the error-prone manual configuration of servers and integration into solutions. Server, network, and storage administrators are now free to focus on strategic initiatives rather than spending their time performing tedious tasks.

- **Cisco Intersight** is the premier management platform for Cisco UCS. It connects to the systems' fabric interconnects when they are configured for Intersight managed mode. Because it is cloud-based Software as a Service (SaaS), it has a broad scope that extends to complete infrastructure and application lifecycle management. It has limitless scale so you can manage all of your infrastructure from a single control point with role- and policy-based automation.

  Cisco Intersight can usher your IT operations into today's hybrid cloud world with your on-premises infrastructure acting as the foundation, with the capability to move and optimize workload performance across multiple clouds. Because it is SaaS, you don't have to support management servers or worry about updates: you are always accessing the latest software and you benefit from a continuous stream of new capabilities and features.

- **Cisco UCS Manager** is embedded in each fabric interconnect and its scope is a single connectivity domain defined as the devices that connect into a single pair of fabric interconnects. Running in a redundant, high-availability configuration, it creates a single, self-aware, self-integrating unified system that recognizes and integrates components as they are added to the system. It quickly and accurately configures computing, network, storage, and storage-access resources to reduce the chance of errors that can cause downtime. Its role- and policy-based approach helps to simplify aligning policies and configurations with workloads.

- **An entire ecosystem** of third-party management tools integrate with Cisco UCS through the Cisco UCS management API.

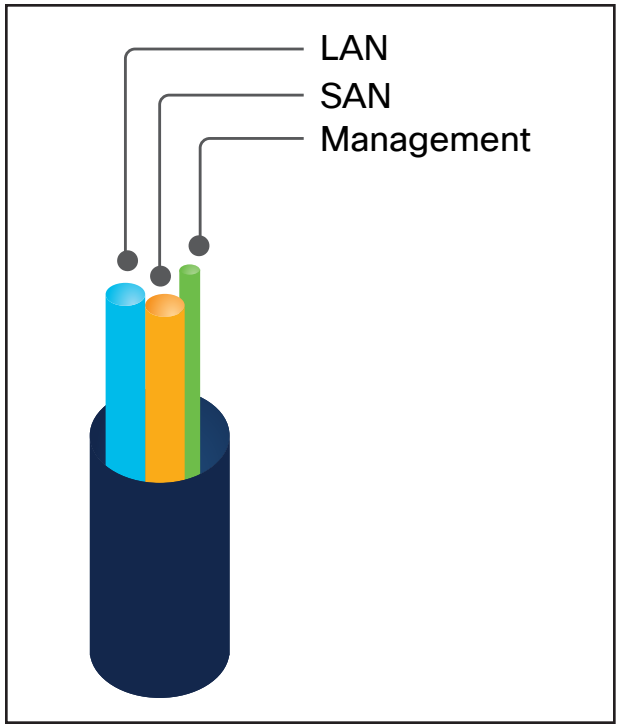
Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco SingleConnect technology

Cisco SingleConnect technology is based on a high-bandwidth, low-latency unified fabric that combines LAN, SAN, and management traffic on a single set of cables.



SingleConnect technology provides an exceptionally easy, intelligent, and efficient way to connect and manage computing in the data center. An exclusive Cisco innovation, SingleConnect technology dramatically simplifies the way that data centers connect to rack and blade servers; physical servers and virtual machines; and LAN, SAN, and management networks.

Cisco UCS fabric interconnects

Cisco UCS fabric interconnects provide a single point of connectivity and management for a system. Deployed as an active-active pair, the system's fabric interconnects integrate all components into a single, highly available connectivity domain. The fabric interconnects manage all I/O efficiently and securely at a single point, resulting in deterministic I/O latency regardless of a server or virtual machine's topological location in the system. Cisco fabric interconnects support low-latency, line-rate, lossless Ethernet and Fibre Channel over Ethernet (FCoE) connectivity. They also support up to 100 Gbps. For remote and edge locations, Cisco UCS 9108 100 Gbps Fabric Interconnects can be used to create a self-contained Cisco UCS X-Series Direct solution for branch offices and remote locations.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco fabric extender technology

Cisco fabric extender technology creates zero-management, low-cost, low-power-consuming devices that distribute the system's connectivity and management planes to rack servers and chassis to scale the system without adding complex switches or management points. This technology eliminates the need for top-of-rack switches and chassis-resident Ethernet and Fibre Channel switches and management modules. This flattens the network topology and dramatically reduces the infrastructure cost per server.

Cisco fabric extender technology extends network and management connectivity to Cisco servers. Rack-mount fabric extenders are optional, and can be used to aggregate network traffic from multiple rack servers. Regardless of connectivity method, all servers are integrated through single-wire management in which all network, storage-access, and management traffic is carried over a single set of cables.

Cisco virtual interface cards

Cisco Virtual Interface Cards (VICs) extend the network fabric directly to both servers and virtual switches so that a single connectivity mechanism can be used to connect both physical and virtual servers with the same level of visibility and control. Cisco VICs provide complete programmability of the Cisco UCS I/O infrastructure, with the number and type of I/O interfaces configurable on demand with a zero-touch model.

Cisco UCS servers

Delivering performance, versatility, and density in servers designed without compromise, Cisco UCS blade, modular, rack, multinode, and storage servers can power every workload, including workloads for:

- Artificial intelligence/machine learning
- Agile development environments requiring bare-metal servers
- Artificial intelligence and machine learning applications
- Big data
- Cloud-native applications
- Content delivery
- Cloud computing environments delivering containers, virtual machines, and bare-metal servers as a service
- Database management systems
- Gaming applications
- GPU-accelerated applications
- High-frequency trading

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

- Hyperconverged applications
- In-memory databases
- Internet infrastructure applications
- Mission-critical enterprise applications
- Mobile application back-end services
- Virtualized environments

Cisco UCS is designed so that every server in our product line can connect and be managed as if it were part of the same single, unified system. In this sense, Cisco UCS is form-factor neutral, giving you more flexibility to choose the server that best meets your needs without the penalty of having to use a different management approach for each type of server (Figure 2).

- Cisco UCS X-Series Compute Nodes are part of the Cisco UCS X-Series Modular System with Intersight, designed from the beginning to be managed from the cloud. The system helps you respond at the speed and scale of your business by shaping resources to workloads with Cisco Intersight and delivering intelligent visualization, optimization, and orchestration to all your applications and infrastructure. It is designed to quickly adapt to the needs of modern, cloud-native and AI applications with a platform that combines the density and efficiency of blade servers with the expandability of rack servers. It simplifies your data center with readiness to embrace emerging technology without requiring forklift upgrades. Best of all, Intersight management enables a stream of new capabilities delivered as software as a service.
- Cisco UCS C-Series Rack Servers can integrate into Cisco UCS through Cisco UCS fabric interconnects or be used as standalone servers with Cisco or third-party switches. With world-record-setting performance for 2- and 4-socket servers, Cisco rack and storage servers can integrate into Cisco UCS through a single set of cables. These servers provide a wide range of I/O, memory, internal disk, Solid-State Disk (SSD) drive and NVMe storage device capacity, enabling you to easily match servers to workloads.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

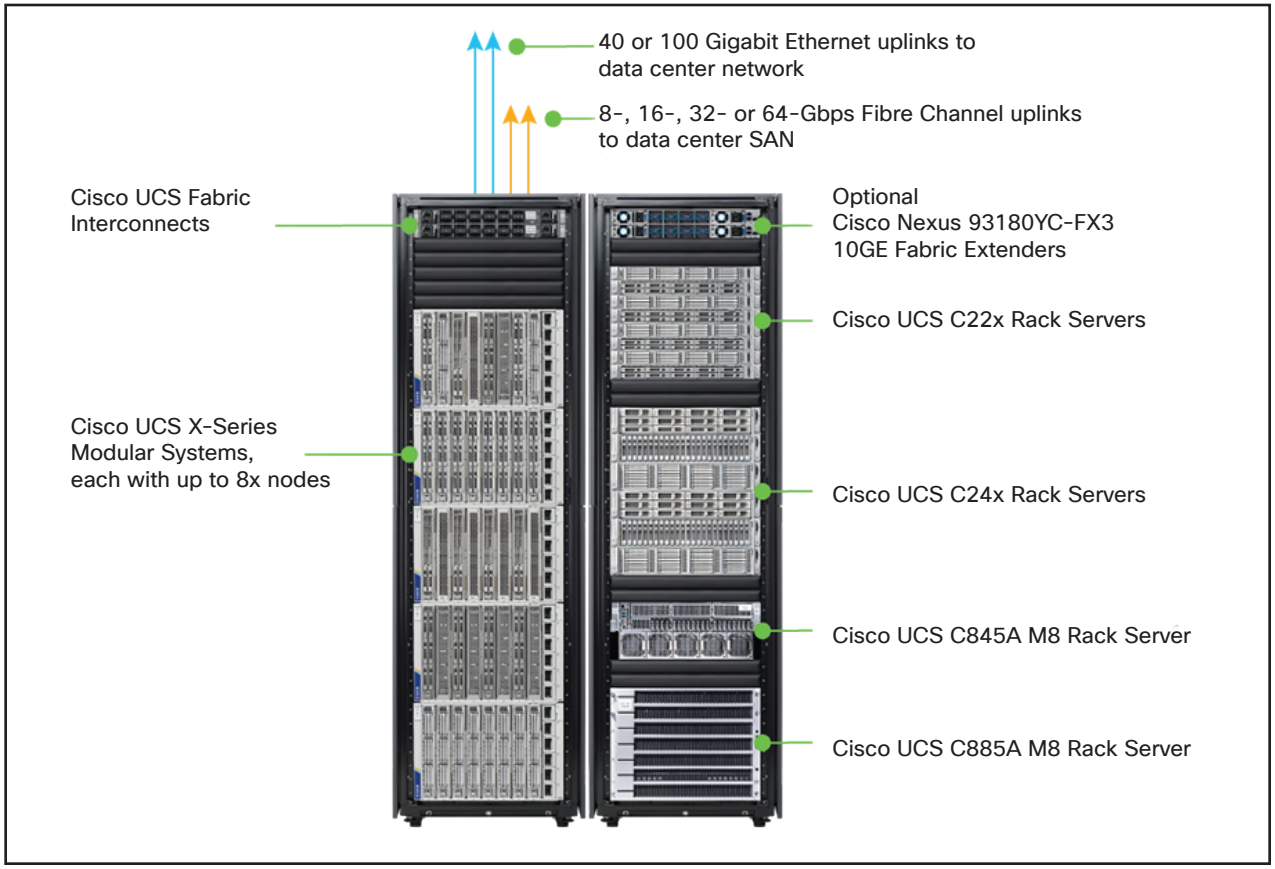


Figure 2. Cisco UCS supports rack and modular servers in a single domain of up to 160 servers

AI-ready infrastructure

As AI adoption skyrockets, we are committed to helping remove complexity and set you up for AI and machine learning success. Cisco AI PODs provide the expertise of Cisco Validated Designs for edge inference, Retrieval-Augmented Generation (RAG), and large-scale inferencing. They accelerate and simplify deployment and ease the IT burden with centralized management and automation. Cisco AI PODs have independent scalability at each layer of infrastructure and are perfect for data-center or edge AI deployments. They are available in multiple configurations that vary the amount of CPU and GPUs in the POD.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco Compute Hyperconverged with Nutanix

Cisco has teamed up with Nutanix to deliver the industry's most complete hyperconverged solution. The solution simplifies and accelerates infrastructure delivery and applications across your entire IT landscape using of best-in-class cloud-operating models, superior flexibility, and enhanced features for support and resiliency. Cisco Compute Hyperconverged with Nutanix allows you to:

- Easily automate HCI deployments
- Enhance resilience with natively integrated networking
- Enjoy more choice with Cisco UCS rack and modular server options
- Improve visibility and control with Cisco Intersight and Nutanix Prism Central Integration

Convergence solution partners

In addition to hyperconvergence solutions, Cisco has partnered with industry-leading converged solution providers to help you reduce energy consumption, remove complexity from hybrid infrastructure, and simplify management across your data-center, edge, and private- and public-cloud deployments. These solutions include:

- FlashStack with Pure Storage
- FlexPod with NetApp
- Hitachi Adaptive Solutions for Converged Infrastructure



Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco UCS management features and tools

The power of a unified API

The foundation of Cisco UCS management is the Cisco UCS unified API. The API enables our own management tools and also a rich ecosystem of third-party tools that can be used for everything from monitoring the system to automating processes through scripting languages.

The Cisco UCS unified API enables higher-level tools to turn the knobs and program the infrastructure by setting the abstracted state variables. This applies whether your Cisco UCS servers are deployed within a single UCS connectivity domain unified with fabric interconnects or standalone servers not connected to fabric interconnects. The Cisco UCS unified API is uniquely positioned to adopt evolving standards while continuing to provide a premier framework for automation.

While the API remains constant, how you access it depends on what management tools you use. In their default configuration, the Cisco UCS fabric interconnects host Cisco UCS Manager and make the API available through an XML interface. When the fabric interconnects are configured in Intersight managed mode, the fabric interconnects act as a proxy that makes all of the server configuration state variables available to Intersight. The fabric interconnects use Intersight Secure Connect technology to make the connection as secure as if an isolated management network were configured.

Programmable infrastructure

Cisco UCS is the only system designed from the beginning to abstract every aspect of system personality, configuration, and connectivity from the hardware so that it can be configured through software. Everything from firmware revisions to network profiles are abstracted into more than 125 configuration variables that fully specify each server. In this sense the system is stateless, and any of the management tools using the Cisco UCS unified API can set server state and thus automate the integration of servers into systems.

From one perspective, Cisco UCS does for physical servers what hypervisors do for virtual machines: it creates an environment in which any server can be configured through software to support almost any workload. This increases flexibility, utilization, and business agility.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco Intersight cloud operations platform

Intersight provides a single control point for organizations to help teams work securely, collaborate better, work smarter, and get the job done faster. Intersight is a unified, secure Software-as-a-Service (SaaS) platform with modules that provide support for everything from basic infrastructure configuration tasks to real-time workload optimization using visibility into how your applications are using the infrastructure on which they run. You don't have to worry about hosting management servers or making sure your management software is up to date. The SaaS nature of Intersight means that you have access to all of the latest features automatically.

Intersight supports complete lifecycle management, from Day 0 deployment using role- and policy-based server profiles, to Day N optimization and support. It supports your bare-metal servers, hypervisors, Kubernetes, and serverless deployments, helping you transform with AIOps to reach the scale and velocity your stakeholders demand. Intersight spans all of your Cisco UCS infrastructure, whether or not it is connected through Cisco UCS fabric interconnects, creating a single management domain with unlimited scale.

Intersight Infrastructure Service

The Intersight Infrastructure Service (Figure 3) enables you to configure the identity, personality, and connectivity of modular, blade, and rack servers. It provides access to automated deployment for entities such as hyperconverged clusters, and it can automate and support large-scale edge deployments. The service provides the following capabilities that are complementary to the basic deployment and configuration features:

- **Cisco Technical Assistance Center (Cisco TAC).** With Intersight's integration with Cisco TAC, you can quickly remediate problems because expertise and information can flow seamlessly between Intersight and your Cisco support center. The system can open cases and upload supporting documentation for fast resolution. It maintains the status of your contracts and licenses so that you can administer them from the same interface.
- **Recommendation engine.** This feature gives you recommendations on configurations and helps you implement best practices. Intersight has insight into your operating system and driver versions. It can use these to validate that your implementations are supported by Cisco's Hardware Configuration List (HCL).

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

- **Global dashboard and inventory.** When you manage your infrastructure with Cisco Intersight, you can view a global dashboard that gives you overall server status and enables you to drill down to view individual components (such as disk drives). With a global inventory of your devices, it's easy to track the location of each of your assets.

Information on the other services supported by Cisco Intersight are available at cisco.com/go/intersight.

Cisco UCS Manager

Cisco UCS Manager supports a single connectivity domain through an intuitive HTML 5-based GUI. Because Cisco UCS integrates computing and networking, it is aware of any device connected to it. When the system is first powered on, or when a new component is added, the system discovers the components and adds them to its internal model of the system configuration. Cisco UCS Manager is model based, allowing administrators to create models of desired server configurations and then configure them simply by associating a model with the physical resources. The system helps guarantee consistent, error-free, policy-based alignment of server personalities with workloads, increasing standards compliance. Server configuration is essentially guaranteed to be correct because Cisco UCS Manager automatically sequences configuration steps and backs them out if an error occurs. Role- and policy-based management preserves current administrator roles (server, storage, and network), helps administrators be more effective in their jobs, helping to reduce overall cost. After configuration, Cisco UCS Manager aggregates element monitoring so that every aspect of the system can be monitored from the GUI or through higher-level tools.

Cisco Integrated Management Controller

Cisco IMC gives you access to standalone rack server configuration. For more details see the "Cisco Integrated Management Controller" section.

DevOps and third-party ecosystem integration

The Cisco UCS API is extensible and fully accessible to automation, allowing you to integrate with Ansible and other DevOps and IT Operations Management (ITOM) tools, using Python and PowerShell Software Development Kits (SDKs). ServiceNow integration provides inventory and alerts to your IT Service Management (ITSM) platform to show inventory and configuration details.

Intersight Secure Connect technology provides a framework to scale and extend the universe of Intersight cloud-managed devices. It provides

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

enterprise-class, secure connectivity to Cisco UCS, converged infrastructure, and an expanding set of third-party hardware and software platforms including storage systems from NetApp, Pure Storage, and Hitachi. For an updated list of active SDKs and integrations, click here.

Cisco UCS management concepts

There is a set of fundamental concepts that are the same whether you use Cisco Intersight, Cisco UCS Manager, or any of the third-party and DevOps integrations.

Inventory and resource pools

When the system is powered on, or when new components are configured in the system, components are added to a hierarchical model that represents all the objects in the system. This model acts as the single source of truth for all connected components and their configuration. Physical components are configured by manipulating the model, not the actual devices.

You can classify servers into resource pools based on criteria including physical attributes (such as processor, memory, and disk capacity) and location (for example, blade chassis slot). Server pools can help automate configuration by identifying servers that can be configured to assume a particular role (such as web server or database server) and automatically configuring them when they are added to a pool.

Resource pools are collections of logical resources that can be accessed when configuring a server. These resources include Unique User IDs (UUIDs), MAC addresses, and World Wide Names (WWNs).

Role- and policy-based management

In typical use cases, subject-matter experts define the way different classes of systems are to be configured by creating resource pools and policies that cover their specific domains of expertise. For example, network administrators can create policies that determine every aspect of the way that a Microsoft Exchange Server should connect to the network. These policies can indicate that certain aspects of identity should be drawn from a specific resource pool—for example, a pool of MAC addresses dedicated to Microsoft Exchange Server NICs.

Management tools support typical divisions of responsibility in IT departments while giving each role visibility into the actions taken by other roles, enhancing communication and simplifying coordination between roles. Role- and policy-based management makes organizations more effective because administrators can configure policies once, and then administrators at any level of authority can use these policies to configure a server.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco UCS service profiles and templates

Administrators can select among policies to create a Cisco UCS service profile (known as a server profile in Intersight), which is the complete specification for how a system should be identified, configured, and connected to IP and storage networks (Figure 4). For example, an administrator might select server, network, and storage access policies designed to support Oracle database servers.

Cisco UCS service profiles also can include storage profiles. Disk groups specify a group of disks characterized by the number and type of disks, RAID level, and number of spares. From a disk group, administrators can create virtual disks that are connected to servers as if they were physical drives.

Whereas a Cisco UCS service profile dictates how to configure a single server, Cisco UCS Manager supports service profile templates that dictate how to create multiple service profiles. These templates can be used to create Cisco UCS service profiles to configure hundreds of servers as easily as you can configure one. Cisco UCS service profiles and templates allow Cisco UCS to be treated as a flexible, configurable pool of resources that can be configured rapidly and accurately to support changing workloads and business conditions:

- **Server configuration**, including changes in the number and type of I/O devices, is completely automated with a zero-touch model.
- **Applications** can quickly be scaled by adding new servers under the direction of service profiles, accelerating the movement of servers from the loading dock into production.
- **Servers** can be repurposed on demand to meet immediate workload requirements. For example, a server that supports a web server farm during the day can be repurposed to support a virtualization cluster at night simply by changing the service profile associated with the server.
- **Firmware** can be revised simply by changing the specification in a Cisco UCS service profile and applying it to a server. Changing versions in a Cisco UCS service profile template can cause all server configurations derived from it to be updated.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

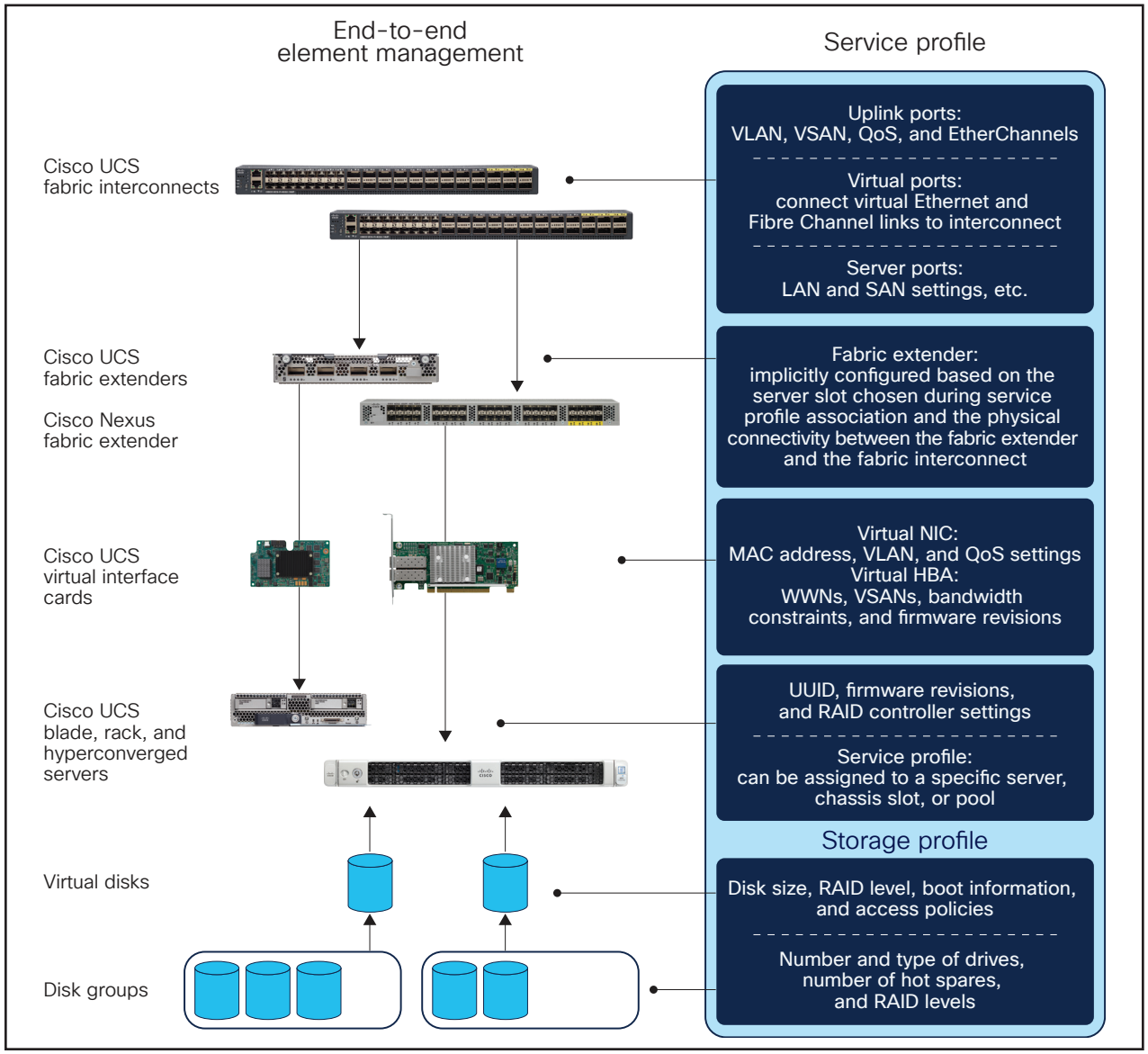


Figure 3. Cisco UCS service profiles fully specify a server's identity, configuration, storage, and connectivity

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

Figure 4 illustrates a workflow in which subject-matter experts create policies that contribute to the definition of a Cisco UCS service profile template. Cisco UCS service profile templates are then used to create multiple Cisco UCS service profiles. When applied to servers, the Cisco UCS service profiles completely specify the server personality, configuration, and connectivity.

Cisco UCS service profiles can be specified as "updating," so that changes to the template change the Cisco UCS service profiles derived from that template. Updating service profiles can allow you, for example, to change all firmware versions in a pool of servers at one time simply by changing the specification in the template.

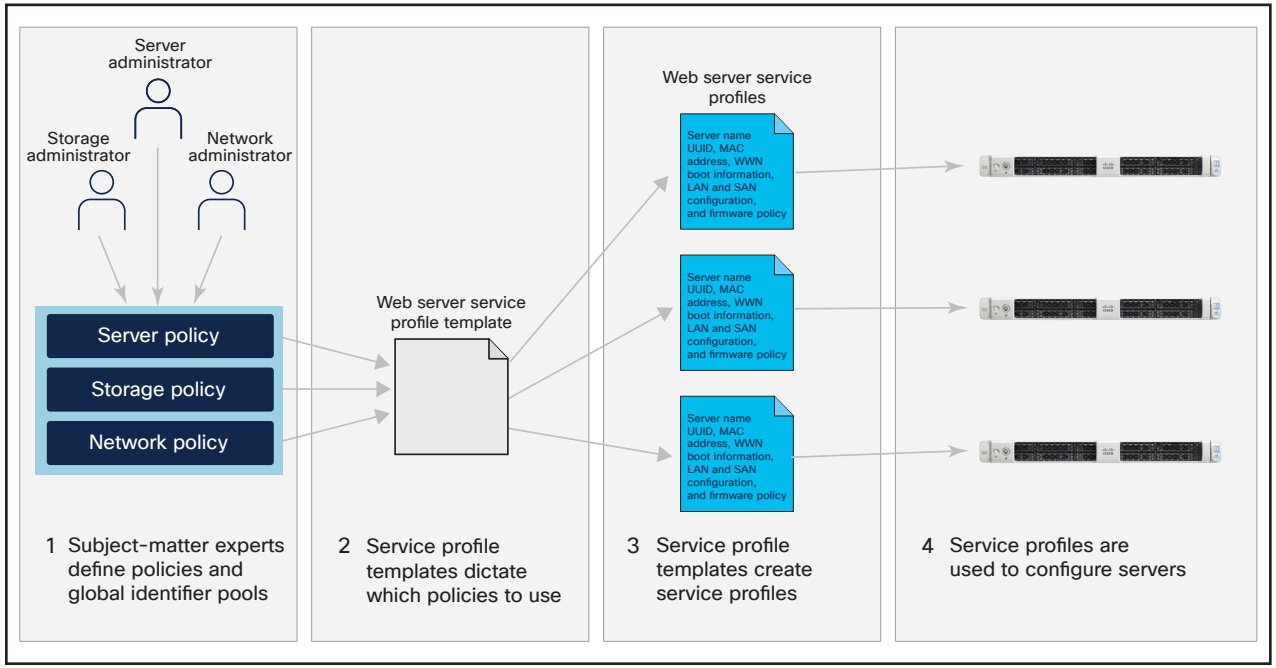


Figure 4. Workflow for role- and policy-based server configuration with Cisco UCS service profiles

© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Logging and audit capabilities

Cisco Intersight and Cisco UCS Manager can produce detailed logs that show how servers are configured and deployed. Because these logs are complete, accurate, and reflect all changes to any component in the entire system, they can be used to detect and automatically remediate any unauthorized change in server configuration, eliminating the configuration drift that can make a server become noncompliant and potentially cause downtime. In the event that a hardware failure does occur, the same information can be used to reboot the software that was running on a failed server onto a known good server. Backup servers can be recruited from a backup pool of servers.

Unified I/O architecture

Cisco UCS is organized around a low-latency, line-rate, lossless 10-, 25-, 40-, and 100-Gbps unified fabric that carries all I/O from servers (regardless of form factor) and virtual machines to the system's fabric interconnects. Using SingleConnect technology, the system is configured with A and B fabrics that are used in an active-active configuration to help provide high availability, increase resource utilization, and reduce costs.

Cisco SingleConnect technology, in combination with unified management, makes the network fabric a strategic asset. The approach in which a single fabric is used to interconnect multiple systems and types of I/O places Cisco UCS first in a class of solutions that Gartner Group refers to as fabric computing. The flexible, agile I/O infrastructure integrated into Cisco UCS lets you move more quickly with instant response to changing workload conditions and business priorities. Policy-based automation accelerates configuration and helps ensure consistency.

- **Wire-once model:** The unified fabric uses a wire-once model in which you configure Cisco UCS for the level of desired capacity at deployment time. After configuration, all I/O resource allocation within that capacity is controlled through software, resulting in zero-touch, instant server and I/O configuration.
- **Flexible pool of resources:** Intelligent networking brings the server and I/O resources of Cisco UCS together as a flexible pool of resources that can be applied on demand to meet any workload challenge. Workload silos are a thing of the past because server power and I/O connectivity can be allocated instantly and accurately through software. Bandwidth is shared among all types of I/O, so that bursts in one class of traffic can temporarily borrow bandwidth from other functions to achieve the best performance.

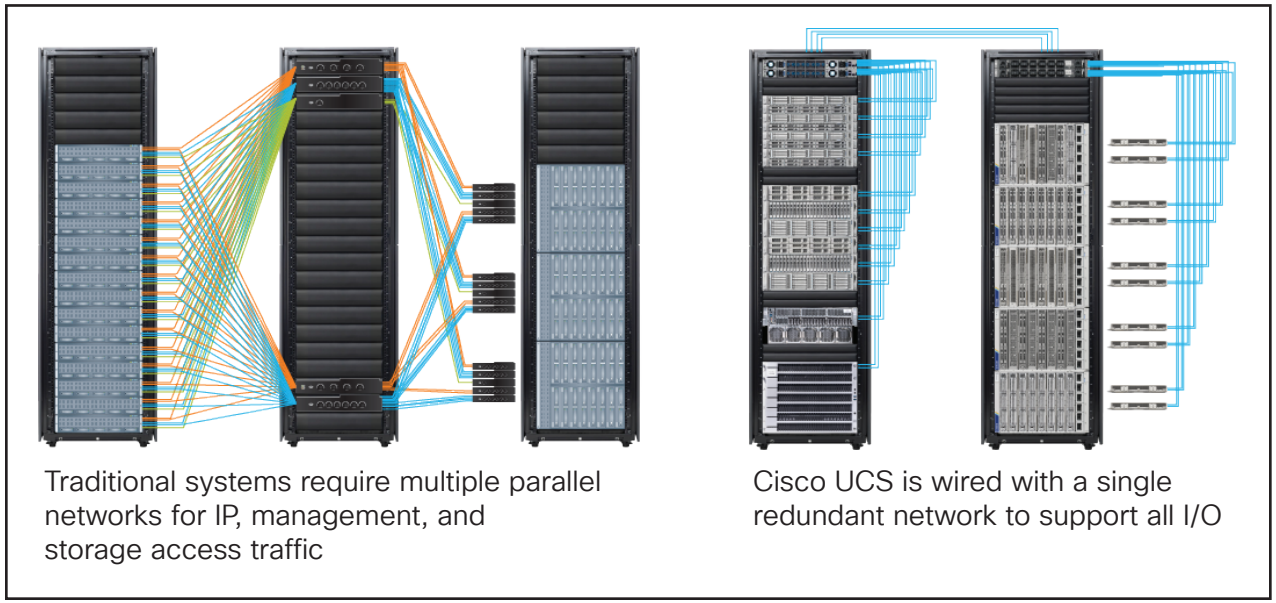
Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

- **Application-centric configuration:** In Cisco UCS, the system adapts to the needs of applications, in contrast to systems in which applications run only on the servers that have been designed and deployed to support them. In virtualized environments, SingleConnect technology can create multiple Fibre Channel Host Bus Adapters (HBAs) and separate Network Interface Cards (NICs) to accommodate shared storage, a management network, another network for virtual machine movement, and multiple networks for LAN traffic—all moments before the hypervisor is booted.

Condense multiple parallel networks

Traditional environments must implement multiple parallel networks to support management, IP networking, and storage access (Figure 5). The result is a proliferation of NICs and HBAs in each server, and upstream switch ports. Each separate network must be sized to handle workload bursts without the capability to share bandwidth between the multiple networks. The complexity of maintaining so much physical infrastructure can lead to cabling errors that can cause downtime or result in security vulnerabilities. Server airflow can be obstructed by the massive number of cables, increasing server temperature and reducing performance.



Traditional systems require multiple parallel networks for IP, management, and storage access traffic

Cisco UCS is wired with a single redundant network to support all I/O

Figure 5. Cisco Unified Fabric eliminates multiple parallel networks, simplifying modular- and rack-server environments

© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

In contrast, SingleConnect technology allows a single unified network to bring LAN, SAN, and management connectivity to each server in a Cisco UCS connectivity domain using Cisco Unified Fabric. All three networks are carried over a single set of cables that securely carries production data (Ethernet) traffic, Fibre Channel traffic through FCoE, and management traffic. Every server—whether modular or rack—has equal access to all network resources, eliminating the need to support three physical interfaces, each with its own NICs, HBAs, transceivers, cables, top-of-rack switches, and upstream switch ports.

Infrastructure silos are eliminated because software—not cabling—determines the way that each server connects to the network. Every server is ready to support any workload at a moment's notice through automated configuration. Instead of requiring separate physical networks to be sized for each traffic class, the shared I/O resources in the unified fabric enable more flexible resource allocation: bursts of traffic in one resource class can borrow bandwidth from other classes subject to Quality-of-Service (QoS) limitations.

Cisco Unified Fabric is based on open standards, some of which are based on Cisco innovations that were later approved by standards bodies:

- The lossless network fabric is implemented using the IEEE 802.3x PAUSE mechanism.
- Different classes of traffic are prioritized using the IEEE 802.1p Priority Flow Control capability; for example, management traffic has its own traffic class and is given the highest priority so that unified management can still function during even the most adverse traffic conditions.
- Network congestion is mitigated using the IEEE 802.1Qau Quantized Congestion Notification standard.
- Bandwidth management is handled through IEEE 802.1Qaz Enhanced Transmission Selection.
- FCoE is implemented following the International Committee for Information Technology Standards' T11 FC-BB-5 standard.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Reduce network layers

In traditional blade server environments, especially virtualized ones, the network access layer is fragmented into three layers, making visibility and control over network connectivity difficult to maintain. These layers add unnecessary and variable latency to virtual networks, they are complex because they usually have different feature sets, and they fragment access-layer management between network and server administrators.

A typical virtualized environment requires three network hops to reach the access-layer switch. As shown in Figure 6, one hop takes place between a virtual machine and the software switch, one between the software switch and the blade-chassis switch, and one between the blade-chassis switch and the access-layer switch.

SingleConnect technology brings the unified fabric directly to each server in Cisco UCS. This approach eliminates switches in modular systems and blade server chassis, reducing latency by one hop. Cisco Fabric Extender Technology, a prestandard implementation of the IEEE 802.1BR Bridge Port Extension standard, eliminates a switching layer with an architecture that is physically distributed but logically centralized. All network traffic passes through the system's fabric interconnects, establishing a single point of management and control. Fabric extenders are configured implicitly as if part of the parent fabric interconnect or switch.

- **In rack server environments**, Cisco Nexus® fabric extenders bring the system's unified fabric to the top of every rack, simplifying rack-level cabling.
- **In chassis-based environments**, Cisco fabric extender technology brings the system's unified fabric chassis. In the Cisco UCS X9508 Chassis, a pair of Cisco UCS 9108 Intelligent Fabric Modules bring up to 1.6 Tbps of bandwidth to each chassis.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

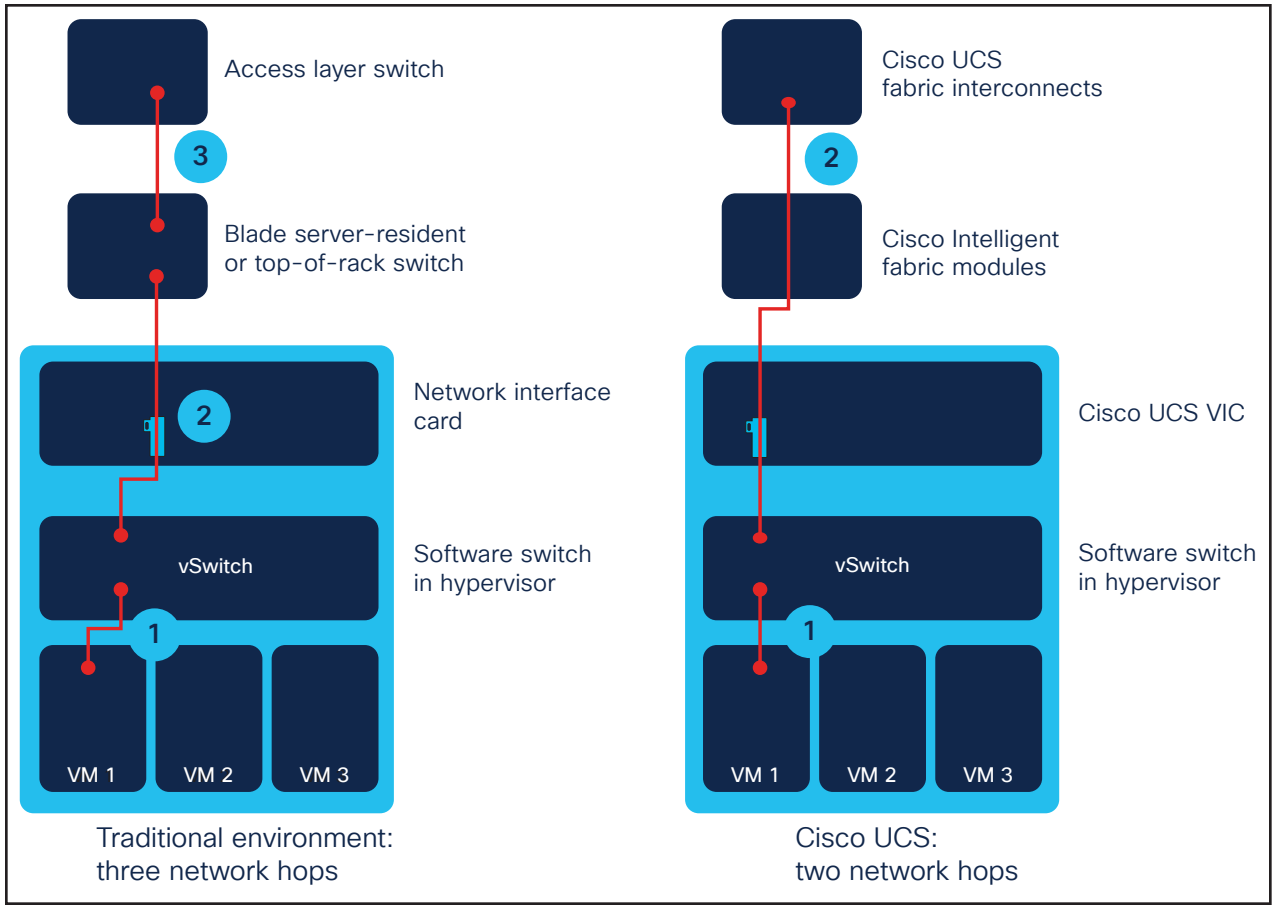


Figure 6. Cisco Unified Fabric eliminates multiple parallel networks, simplifying modular and rack server environments

The logical centralization of the I/O infrastructure means that after the system is established, it can scale without the need to reevaluate the infrastructure or configure additional switches to incorporate additional servers.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Virtualize networking

The unified fabric virtualizes I/O. Rather than requiring each server to be equipped with a set of physical I/O interfaces and cables for separate network functions, all I/O in the system is carried over a single set of cables and sent to separate physical networks at the system's fabric interconnects as necessary. For example, storage traffic destined for Fibre Channel storage systems is carried within the system using FCoE. At the fabric interconnects, storage traffic can transition to physical Fibre Channel networks through one or more of the fabric interconnect's unified ports.

I/O is further virtualized through the use of separate virtual network links for each class and each flow of traffic. For example, management, storage-access, and IP network traffic emanating from a server is carried to the system's fabric interconnects with the same level of secure isolation as if it were carried over separate physical cables (Figure 8). These virtual network links originate within the server's virtual interface cards and terminate at virtual ports within the system's fabric interconnects.

Virtual network links are managed exactly as if they were physical. The only characteristic that distinguishes physical from virtual networks within the fabric interconnects is how the ports are named. This approach has a multitude of benefits: changing the way that servers are configured makes servers flexible, adaptable resources that can be configured through software to meet any workload requirement at any time. Servers are no longer tied to a specific function for their lifetime because of their physical configuration. Physical configurations are adaptable through software settings. The concept of virtual network links brings immense power and flexibility to support almost any workload requirement through flexible network configurations that bring complete visibility and control.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

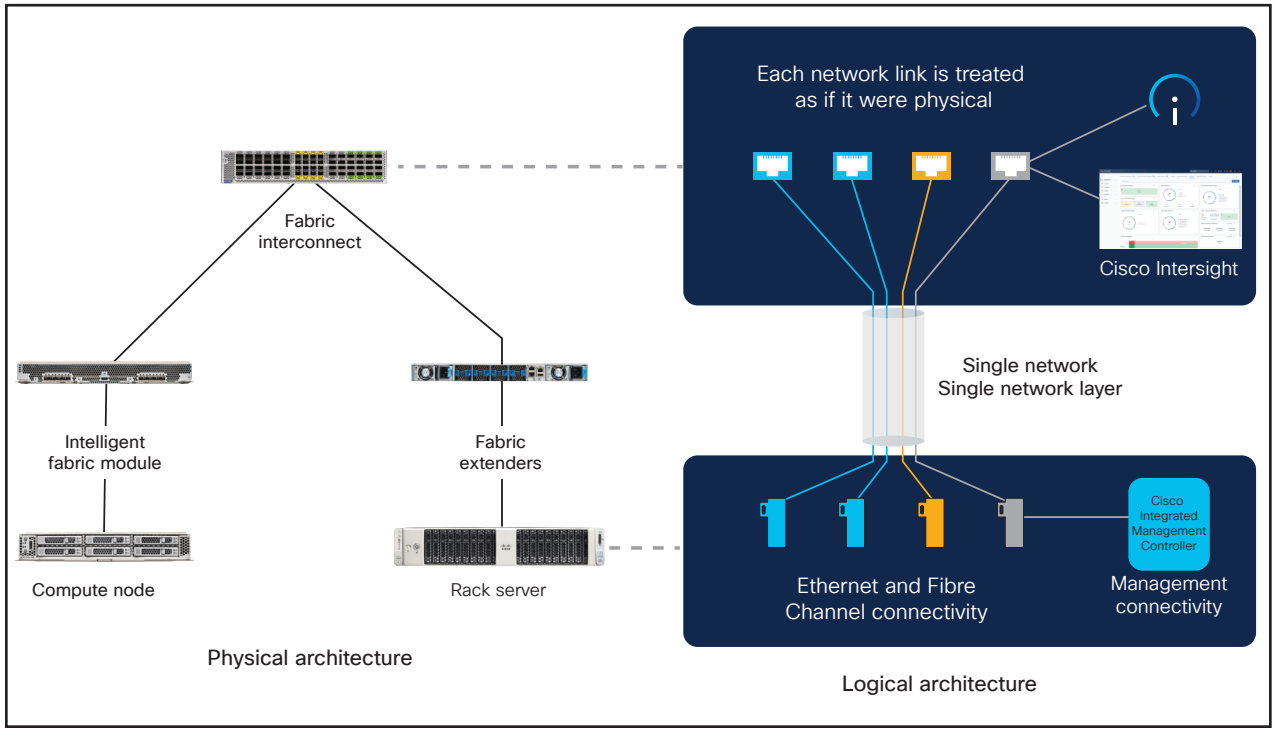


Figure 7. Cisco Unified Fabric eliminates multiple parallel networks, simplifying server environments

Virtualize I/O interfaces

Cisco VICs are PCIe-compliant interfaces that support 256 or more PCIe devices with dynamically configured type (NIC or HBA), identity (MAC address or WWN], fabric failover policy, bandwidth, and QoS policy settings. With Cisco VICs, server configuration—including I/O configuration—becomes configurable on demand, making servers stateless resources that can be deployed to meet any workload need at any time, without any physical reconfiguration or recabling required.

© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco UCS product details and best practices

Cisco UCS servers

Cisco UCS is based on industry-standard, x86-architecture servers with Cisco innovations. Although many vendors offer servers with the same processors, we integrate them into a system with a better balance of CPU, memory, and I/O resources. This balance brings processor power to life with more than 200 world-record-setting benchmark results that demonstrate our leadership in application areas including virtualization, cloud computing, enterprise applications, database management systems, enterprise middleware, high-performance computing, and basic CPU integer and floating-point performance metrics.

Match servers to workloads

The breadth of our server product line makes the process of matching servers to workloads straightforward, enabling you to achieve the best balance of CPU, memory, I/O, internal disk, and external storage-access resources using the modular or rack form factor that best meets your organization's data center requirements and preferred purchasing model. Our server family is featured in the sidebars on this and the following pages.

Powered by AMD EPYC processors

For workloads needing the highest number of cores in a 2-socket server, or the fastest integer and floating-point performance, we offer 4th and 5th Gen AMD EPYC processors in both rack and modular server form factors. Our modular and rack M8 servers are optimized to make more of the AMD EPYC processor's PCIe Gen5 connectivity available independently of whether you populate it with one or two CPUs. Many workloads don't need more than 160 cores, and using a single processor is an economical way to support them.

The AMD EPYC processor's high core density is matched with eight memory channels synchronized with the AMD Infinity Fabric to make compute-intensive applications run faster. The processor's 160 lanes of I/O capacity are harnessed to speed I/O from fifth-generation Cisco UCS VICs and direct connectivity to internal disk drives and NVMe storage. A silicon-embedded security processor manages encryption keys that enable individual virtual machines to be encrypted, helping to protect against a malicious hypervisor or virtual machine. Full memory encryption can be used without

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

any changes to your software for bare-metal workloads, helping repel cold-boot attacks against main memory. The security processor also helps verify the integrity of the boot process.

Powered by Intel Xeon Scalable processors

Cisco UCS rack servers are powered by 4th and 5th Gen Intel® Xeon® Scalable Processors and Intel Xeon 6 processors to deliver highly robust capabilities with outstanding performance, security, and agility. They offer up to 86 cores in 2-socket and up to 60 cores in 4-socket configurations for excellent performance and scalability. The CPUs provide excellent memory channel performance and include up to four Intel Ultra Path Interconnect (Intel UPI) links per CPU for scalability and intercore data flow. The processors also offer hardware-assisted security advancements that lower the performance overhead for data encryption and decryption, lowering the cost of securing data. These features further enhance the value of IT infrastructure in your enterprise.

Industry-leading bandwidth

Cisco virtual interface cards have dramatically simplified the deployment of servers for specific applications. By making the number and type of I/O devices programmable on demand, we enable organizations to deploy and repurpose server I/O configurations without ever touching the hardware.

- **Modular server bandwidth:** Cisco UCS X-Series compute nodes connect to up to 200 Gbps of bandwidth per server—with 100 Gbps per bidirectional port. When equipped with a mezzanine-form-factor Cisco VIC, the server can connect to future I/O devices connected through Cisco UCS X-Fabric technology.
- **Rack server bandwidth:** In rack servers, I/O connectivity can be configured as needed using PCIe-format adapters, dedicated mLOM adapters, or System I/O Controllers (SIOCs). Each adapter supports two 40-Gbps, four 25-Gbps, or two 100- Gbps unified fabric connections. Cisco UCS M7 and M8 rack servers offer high bandwidth capabilities, primarily through high-speed memory and flexible networking options. The Cisco UCS C845A/C885A rack servers are optimized for AI workloads and come with powerful GPU networks and DPUs to accelerate the most advanced Large Language Model (LLM) training, inference, and more.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

Consistent low latency

Consistent low latency is important for managing performance. Traditional systems exhibit different latency characteristics depending on workload placement, which places unnecessary constraints on applications.

With a single logically centralized yet physically distributed network, SingleConnect technology requires only a single network hop and offers consistent latency. With Cisco UCS, traffic between any two rack or blade servers requires only one network hop (see path A in Figure 9). With traditional environments, intrachassis communication likewise requires only one hop (path X), but communication between chassis or between blade and rack servers requires three hops (see path Y in Figure 8).

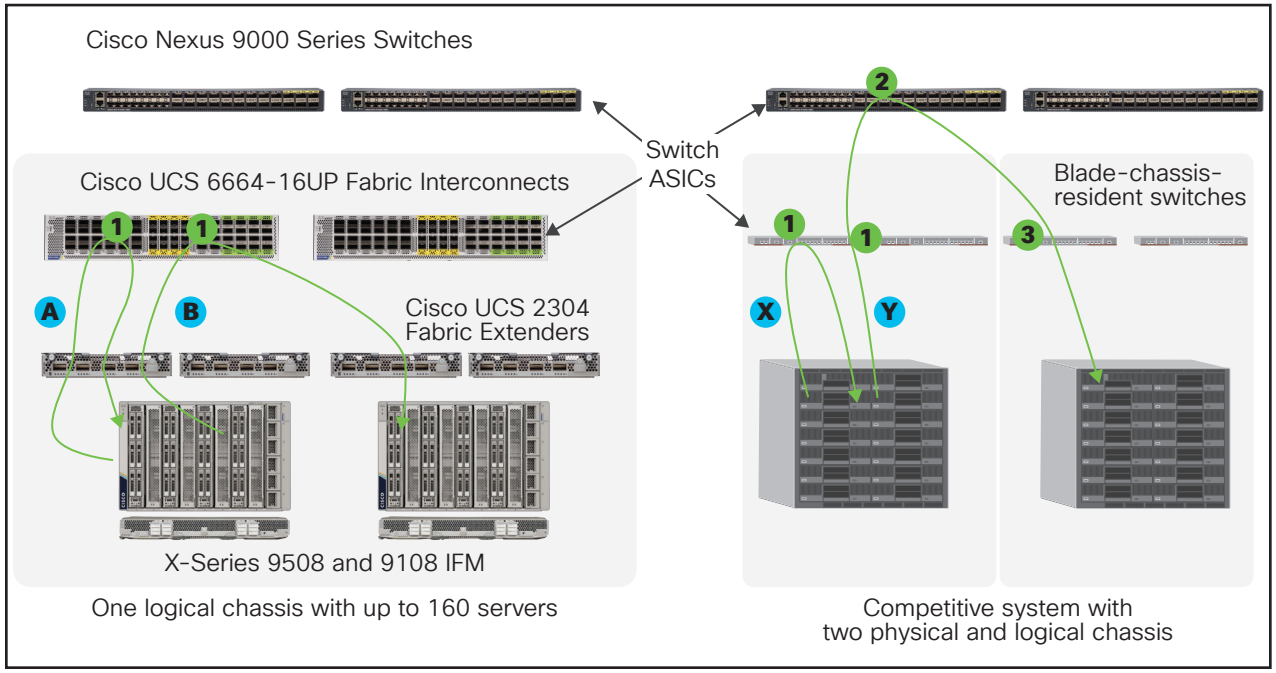


Figure 8. Every server is only one network hop away in Cisco UCS, compared to up to three hops in traditional blade server environments

© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Lower infrastructure cost

We designed Cisco UCS for lower infrastructure-cost per server, a choice that makes scaling fast, easy, and inexpensive in comparison to manually configured approaches. This choice is evident in the design of the Cisco UCS X9508 Chassis.

This chassis, designed to be low-cost, is little more than sheet metal, sensors, and slots for power supplies, fans, and servers. The intelligence is contained in the modular devices used to connect the compute nodes to the network fabric: the Cisco UCS 9108 Intelligent Fabric module for the Cisco UCS X-Series Modular System. These devices contain Cisco fabric-extender technology and are logically part of the Cisco UCS fabric interconnects that manage them.

Our rack servers are similarly integrated into Cisco UCS with lower infrastructure-cost per server. Instead of requiring up to five switch ports at the top of every rack (two Ethernet, two Fibre Channel, and one management network switch), Cisco UCS requires only two.

Rack server deployment flexibility

Cisco UCS C-Series Rack Servers are unique in the industry because they can be integrated with Cisco UCS connectivity and management or used as standalone servers.

Cisco Integrated Management Controller

The Cisco Integrated Management Controller (Cisco IMC) runs in the system's Baseboard Management Controller (BMC). When a Cisco UCS C-Series Rack Server is integrated into a Cisco UCS domain, the fabric interconnects interface with the Cisco IMC to make the server part of a single unified management domain. When a server is used as a standalone server, direct access to the Cisco IMC through the server's management port allows a range of software tools (including Cisco Intersight) to configure the server through its API. With either approach you have access to a wide range of management capabilities (Figure 9).

- Standalone operation with the Cisco IMC – As standalone servers, Cisco UCS C-Series Rack Servers are managed through the dedicated management network interface using any of the approaches illustrated in Figure 10. Cisco Intersight connects through SSL to this interface for secure management from the cloud or an optional locally-hosted management appliance.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

- Integrated operation with single-wire management – Single-wire management uses the unified fabric to carry management traffic between the server and the fabric interconnects. It is enabled with through Cisco VICs, which separate management traffic from production data and storage traffic, passing it to an internal switch that connects to the Cisco IMC. The internal switch also makes the controller accessible for standalone management through the server's network management ports (Figure 10).

When single-wire management is used, the unified fabric securely isolates management traffic by connecting the fabric interconnect's management network directly to the controller using the IEEE 802.1BR standard. To prevent any high-traffic condition on the network from impeding management traffic, Cisco UCS gives management traffic the highest priority using the IEEE 802.1Qbb Priority Flow Control standard.

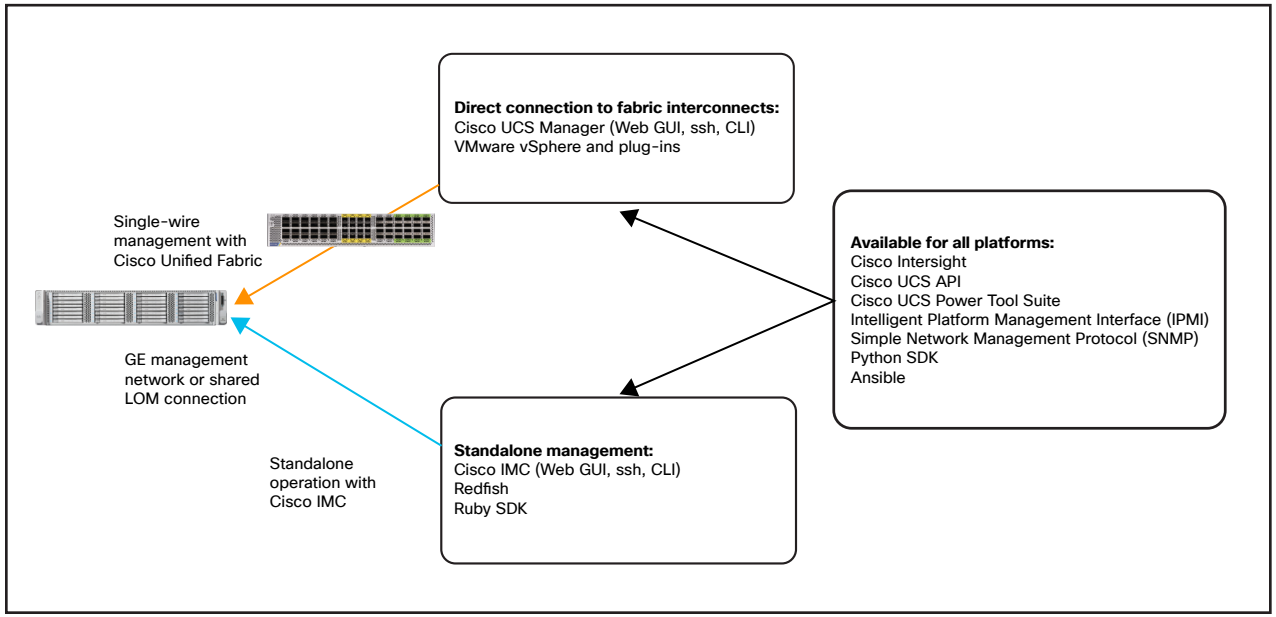


Figure 9. Whether integrated with Cisco UCS through fabric interconnects, or managed through Cisco IMC, a wide variety of tools can be used to manage the servers.

© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

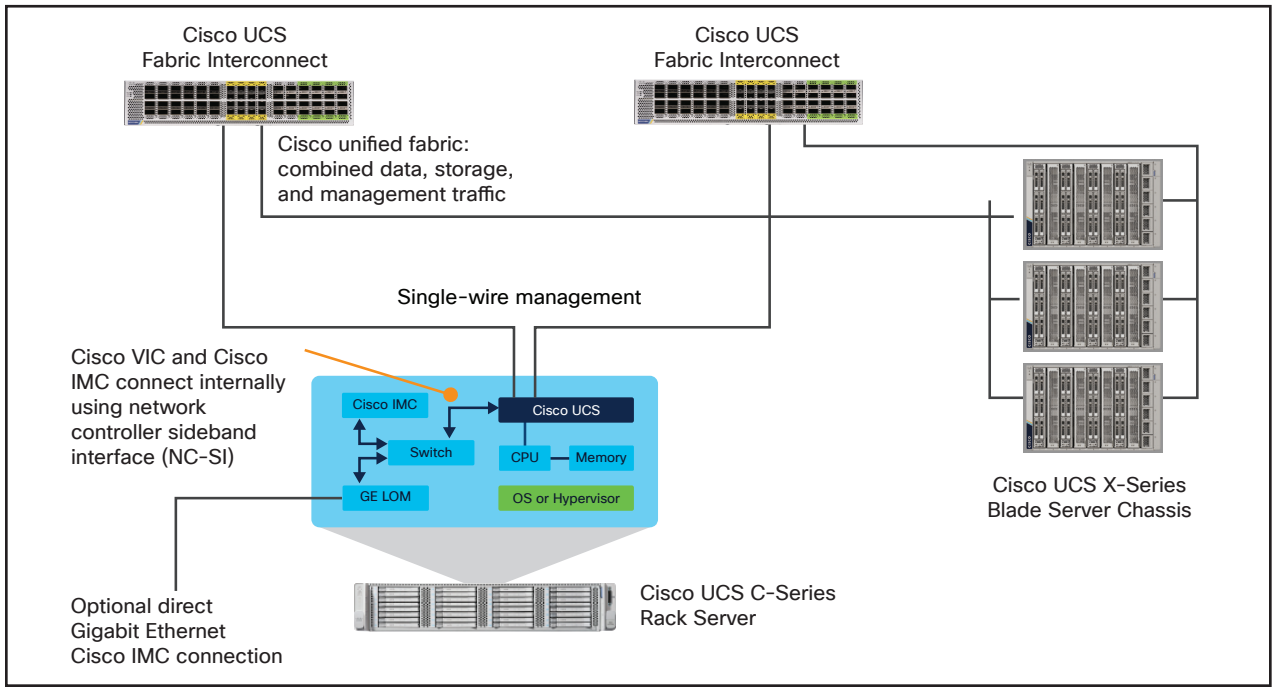


Figure 10. Rack servers integrate into Cisco UCS with single-wire management

Cisco UCS X-Series Modular System

Make your infrastructure future-ready with our modular compute nodes. Cisco UCS X-Series servers give you the choice of processors from both AMD and Intel:

- **More cores** to accelerate parallelized virtualized and bare-metal workloads.
- **Larger memory capacity** for better performance and larger in-memory databases.
- **Higher memory bandwidth** to accelerate the flow of information to and from the CPU.
- **Up to 16 GPU accelerators** for the most demanding machine-learning workloads, or a smooth user experience in virtual desktop environments.
- **Cloud-management ready**, and prepared to connect to the Cisco Intersight platform either directly or through Cisco UCS fabric interconnects.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco UCS X-Series nodes include:

- **Cisco UCS X210c M8 Compute Node** – offers flexible, high-density computing with up to two Intel Xeon 6 processors with up to 86 cores per processor and up to 336 MB of Level-3 cache per CPU.
- **Cisco UCS X215c M8 Compute Node** – 2-socket 4th or 5th Gen AMD EPYC processor–based compute node.
- **Cisco UCS X210c M7 Compute Node** – 2-socket modular server with 4th or 5th Gen Intel Xeon Scalable Processors.
- **Cisco UCS X410c M7 Compute Node** – 4-socket modular server with 4th Gen Intel Xeon Scalable Processors.
- **Cisco UCS X440p PCIe Node** – Up to 16 VDI class or two AI-class GPUs per chassis to accelerate your applications.
- **Cisco UCS X580p PCIe Node** – Up to four AI-class GPUs per chassis to accelerate your applications.
- **Cisco UCS X9508 Chassis** – 7RU, midplane-free chassis with eight front slots for modules, and rear slots for Ethernet and Cisco UCS X-Fabric Technology.

The Cisco UCS X-Series includes Cisco UCS X-Series Direct for customers who need smaller server environments. It offers performance, flexibility, and optimization for deployments in data centers, in the cloud, and at remote sites. It also offers one of the highest densities of compute, I/O, and storage per rack unit to reduce Operational Expenses (OpEx).



Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco UCS C-Series rack servers

Cisco UCS C-Series rack servers empower you to tackle any workload challenge with a flexible combination of processor, memory, I/O, and internal storage resources. We offer a choice of AMD EPYC and Intel Xeon Scalable processors in our 1RU and 2RU rack servers:

- **Cisco UCS C220 M8 Rack Server** – 1RU rack server with up to two Intel Xeon 6 processors and improved performance with AI-accelerated features.
- **Cisco UCS C240 M8 Rack Server** – 2RU rack server with up to two Intel Xeon 6 processors, AI-accelerated features, and increased storage capacity at higher speed.
- **Cisco UCS C225 M8 Rack Server** – 1-socket rack server with a 4th or 5th Gen AMD EPYC processor, up to 160 cores per server, and support for up to three GPUs.
- **Cisco UCS C245 M8 Rack Server** – 2-socket rack server with up to two 4th or 5th Gen AMD EPYC processors, up to 160 cores per server, and support for up to eight GPUs.
- **Cisco UCS C845A M8 Rack Server** – AI-optimized rack server with up to two 5th Gen AMD EPYC processors, and support for up to eight AMD or NVIDIA GPUs.
- **Cisco UCS C885A M8 Rack Server** – AI-optimized rack server with up to two 4th or 5th Gen AMD EPYC processors, NVIDIA HGX architecture, and support for up to eight GPUs.
- **Cisco UCS C220 M7 Rack Server** – high-density, 1RU, 2-socket rack server with up to two 4th or 5th Gen Intel Xeon Scalable Processors.
- **Cisco UCS C240 M7 Rack Server** – 2-socket, 2RU rack server with up to two 4th or 5th Gen Intel Xeon Scalable Processors.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

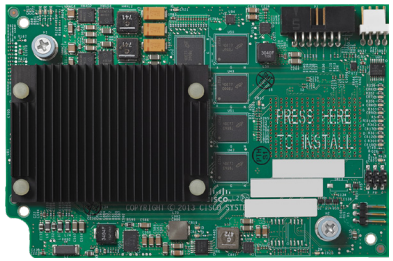
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

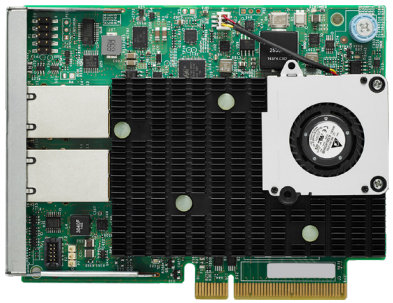
Cisco VICs

Cisco VICs present physical PCIe-compliant devices to operating systems and hypervisors, supporting them without the need for them to implement Single Root I/O Virtualization (SR-IOV). This approach allows almost any hypervisor or operating system to be supported without additional complexity. Cisco VICs are available in multiple form factors so that every Cisco UCS server can support virtualized I/O:

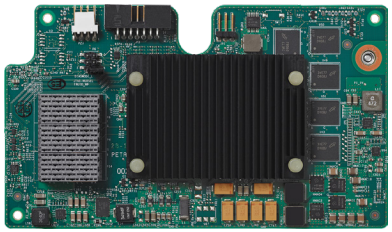
Modular LOM (blade servers)



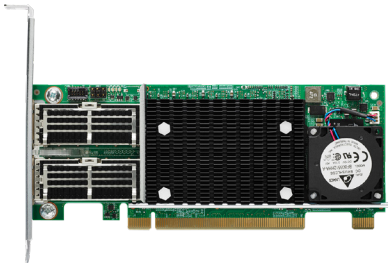
Modular LOM (rack servers)



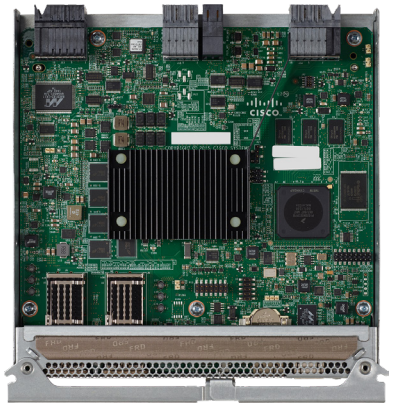
Mezzanine



PCIe



System I/O controller



Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco VICs support connectivity to 10-, 25-, 40- and 100-Gbps networks and are available in multiple form factors:

- **mLOM:** Modular LAN-on-motherboard adapters are available for both blade and rack servers. These have the benefit of not occupying a mezzanine slot (in blade servers) or a PCIe slot (for rack servers), leaving the space for other devices.
- **Mezzanine:** Standard Cisco VICs can be installed in any blade server or compute node's mezzanine slot: one for half-width blade servers, up to two for full-width blade servers, and one for Cisco UCS X-Series compute nodes. Each Cisco VIC supports up to 80 Gbps of connectivity per blade server.
- **PCIe:** These Cisco VICs integrate Cisco rack servers into Cisco UCS through circuitry that passes the unified fabric's management traffic to the server's management network, enabling single-wire, unified management of rack servers (see the "Cisco Integrated Management Controller" section.

Support best practices

Virtual interfaces can be used to support operating system requirements and best practices. Figure 12 shows devices supporting a hypervisor, with separate interfaces for the hypervisor console, the hypervisor management interface, virtual machine migration traffic, and two Fibre Channel HBAs to provide access to shared storage. Pairs of Ethernet NICs are shown connected to virtual switches in the hypervisor. Each interface connects to a virtual link that terminates at a virtual port in the fabric interconnect, giving the flexibility of virtual networking with the visibility and control of physical networks.

Integrate with data center networks and Cisco Application Centric Infrastructure

Condensing multiple network layers into a single physically distributed but logically centralized connectivity domain makes the process of integrating Cisco UCS into data center networks straightforward. Rather than appearing as an entire network with a collection of servers, Cisco UCS is integrated as a single system. This simplification is accomplished using both Ethernet and Fibre Channel end-host modes. It eliminates the need for Spanning Tree Protocol, and it pins the MAC addresses and WWNs for both physical and virtual servers at the uplink interfaces. This approach gives the fabric interconnects complete control over the unified fabric within Cisco UCS and allows greater utilization of uplink port bandwidth through the use of active-active Ethernet uplinks.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

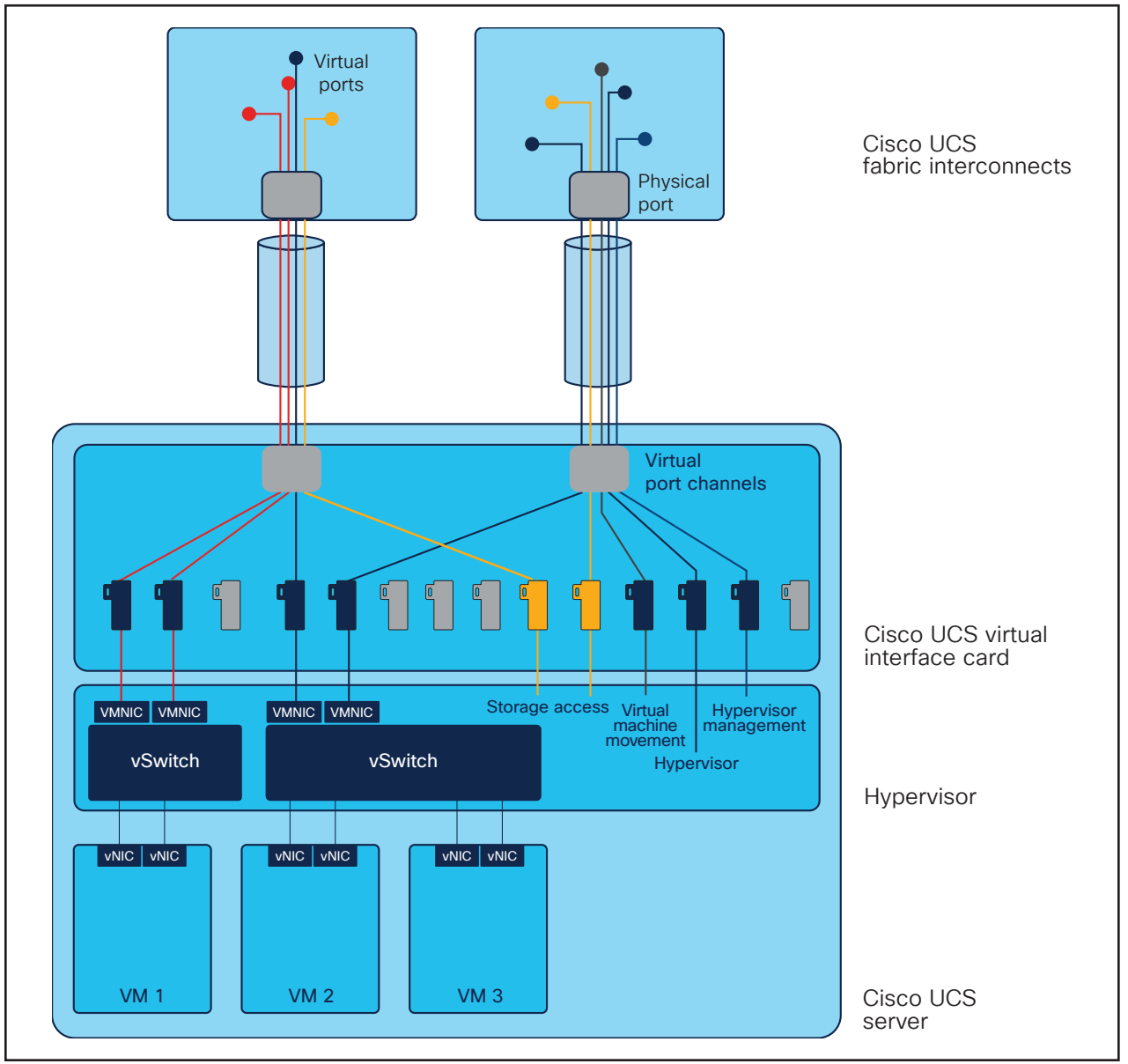


Figure 11. Cisco VIC with fabric failover enabled for Ethernet NICs and not for Fibre Channel HBAs, following best practices

© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

With the fabric interconnects in the default end-host mode, Cisco Application Centric Infrastructure (Cisco ACI™) leaf switches see their connections to fabric interconnects as if they were connected the servers themselves.

With Cisco UCS servers and virtual machines grouped into specific VLANs, and with each VLAN associated with a specific Cisco ACI Endpoint Group (EPG), any server appearing in that VLAN is automatically associated with the correct EPG. The Cisco ACI encapsulation normalization mechanism automatically wraps any upstream traffic from these servers in Virtual Extensible LAN (VXLAN) tunnels so that the traffic is securely isolated as if it were on its own physical network segment.

Fabric failover best practices

Cisco VICs support fabric failover at the device level, so losing one of the two fabrics can be transparent to the operating system or hypervisor. They can implement best practices for each type of network:

- **For LAN connections**, failover can be performed by the Cisco VIC so that the operating system or hypervisor can continue to operate without knowledge of the failure.
- **For SAN connections**, failover is usually handled by the operating system drivers, so fabric failover is typically not used.

Architecture for high availability

Cisco UCS is designed for high availability, with no single point of failure in its network infrastructure. The fabric interconnects are designed to work in an active-active model, with automated failover of network management in the event of a failure. The system is designed so that if either fabric A or fabric B fails, the remaining fabric will take on the traffic from the failed fabric. Cisco VICs support fabric failover by moving traffic from one fabric to the other according to failover policies established on a per-NIC basis. This eliminates complicated operating system or hypervisor NIC teaming configuration.

Fabric failover makes the failure of a fabric transparent to the operating system or hypervisor. Figure 12 illustrates that different types of network modalities have different best practices for failover, and Cisco VICs support both. Fabric failover is typically configured for Ethernet because it avoids complex operating system and hypervisor configuration. In the example, vNIC 1 connects primarily to fabric A but fails over to fabric B in the event of a fabric failure. Fibre Channel is typically cabled with two networks with the driver software handling failover, so HBAs 3 and 4 connect to fabrics A and B, respectively, without fabric failover configured.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

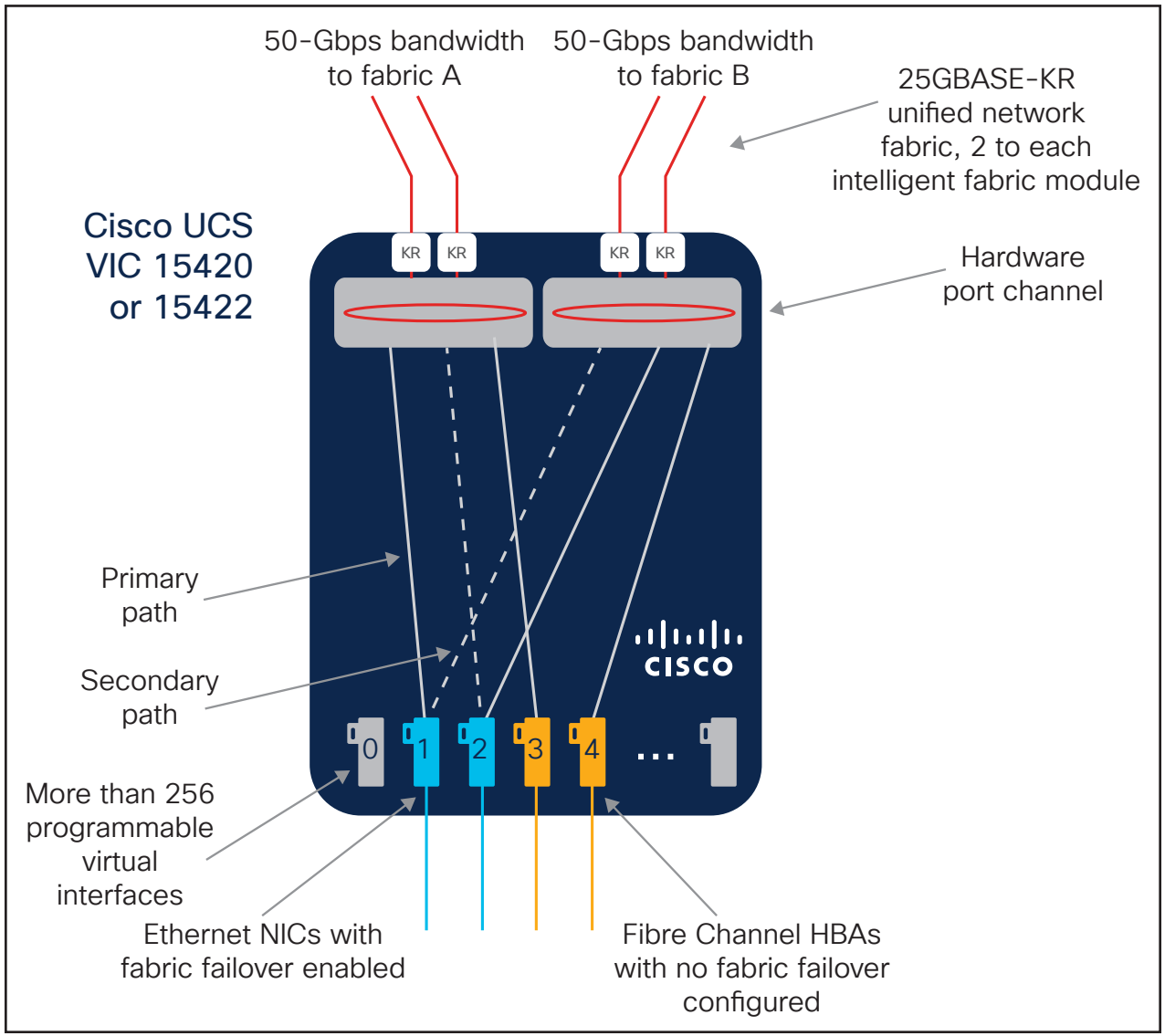


Figure 12. Cisco VIC with fabric failover enabled for Ethernet NICs and not for Fibre Channel HBAs, following best practices

© 2025 Cisco and/or its affiliates. All rights reserved.

Contents

Introduction

The challenge

The solution

Cisco UCS anatomy overview

Cisco UCS management features and tools

Cisco UCS management concepts

Cisco UCS service profiles and templates

Unified I/O architecture

Cisco UCS product details and best practices

Ready for modern practices

Power your applications with unified compute, networking, and storage

For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco UCS VIC models

- **Cisco UCS VIC 15000 Series**
  - 10/25/50 Gbps: 15425 (PCIe) and 15427 (mLOM) for Cisco UCS C-Series Rack Servers; 15420 (mLOM) and 15422 (Mezz) for Cisco UCS X-Series compute nodes
  - 40/100 Gbps: 15235 (PCIe) and 15237 (mLOM) for Cisco UCS C-Series Rack Servers 15230 (mLOM) for Cisco UCS X-Series compute nodes
- **Cisco UCS VIC 14000 Series for Cisco UCS X-Series compute nodes**
  - 25 Gbps: 14425 (mLOM), 14825 (Mezz)
- **UCS VIC 1400 Series for Cisco UCS C-Series Rack Servers**
  - 10/25 Gbps: 1455 (PCIe) 1457/1467 (mLOM)
  - 40/100 Gbps: 1477 (mLOM) 1495 (PCIe), 1497 (mLOM)

Cisco fabric interconnects and fabric extenders

Cisco UCS fabric interconnects and fabric extenders are part of the Cisco Unified Fabric solution for data center networks. Products include:

- **Cisco UCS 6664 Fabric Interconnect** – 64 ports, supports 10/25/50/100-Gigabit Ethernet
- **Cisco UCS 6536 Fabric Interconnect** – 36 ports, supports 1/10/25/40/100-Gigabit Ethernet or FCoE
- **Cisco UCS 6400 Series Fabric Interconnects** – 54 or 108 ports, supports 1/10/25/40/100-Gigabit Ethernet or FCoE
- **Cisco UCS 9108 100G Fabric Interconnect** – exclusively for Cisco UCS X-Series Direct, supports 1/10/25/40/100-Gigabit Ethernet or FCoE
- **Cisco UCS X9416 X-Fabric Module** – connects Cisco UCS X210c, Cisco UCS X215c, and Cisco UCS X410c compute nodes to Cisco UCS X440p PCIe Nodes
- **Cisco UCS X9516 X-Fabric Module** – connects Cisco UCS X210c and Cisco UCS X215c, compute nodes to Cisco UCS X580p PCIe Nodes
- **Cisco UCS 9108 25G Intelligent Fabric Module** – eight 25-Gbps uplink ports, and lossless and deterministic converged fabric
- **Cisco UCS 9108 100G Intelligent Fabric Module** – eight 100-Gbps uplink ports, and lossless and deterministic converged fabric

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

© 2025 Cisco and/or its affiliates. All rights reserved.

Ready for modern practices

Through a unified API, Cisco UCS adapts easily to support new practices. Cisco UCS infrastructure can be programmed on demand to support DevOps practices through a variety of tools and scripting languages.

The following tools, described in UCS Programmable Infrastructure, are designed to support programmable configuration of standalone rack servers:

- Power Shell
- Python Software Development Kit (SDK)
- Ansible SDK

Power your applications with unified compute, networking, and storage

The Cisco Unified Computing System is the first integrated data center platform that combines industry-standard, x86-architecture servers with networking and storage access into a single unified system. The system is intelligent infrastructure that uses integrated, model-based management to simplify and accelerate deployment of applications and services running in bare-metal, virtualized, and cloud-computing environments. Employing Cisco's innovative SingleConnect technology, the system's unified I/O infrastructure uses a unified fabric to support both network and storage I/O. The Cisco fabric extender architecture extends the fabric directly to servers and virtual machines for increased performance, security, and manageability.

Cisco UCS helps change the way that IT organizations do business, with benefits that include the following:

- **Increased IT staff productivity** and business agility through just-in-time provisioning and equal support for both virtualized and bare-metal environments.
- **Reduced TCO** at the platform, site, and organization levels through infrastructure consolidation.
- **A unified, integrated system** that is managed, serviced, and tested as a whole.
- **A comprehensive management ecosystem** that supports complete infrastructure provisioning and management that can make your Cisco UCS instance anything from a bare-metal enterprise application engine to a multlicloud containerized environment. Locally hosted tools give you a range of options and Intersight software as a service is emerging as the solution to help you manage all of your assets worldwide.

Contents

Introduction
The challenge
The solution
Cisco UCS anatomy overview
Cisco UCS management features and tools
Cisco UCS management concepts
Cisco UCS service profiles and templates
Unified I/O architecture
Cisco UCS product details and best practices
Ready for modern practices
Power your applications with unified compute, networking, and storage
For more information

- **Scalability** with the Intersight software-as-a-service platform that can manage all of your infrastructure wherever it resides. Many of the most tedious administration tasks, such as setting up clusters in edge locations, can be automated completely so that new locations can be rolled out quickly, easily, and accurately.
- **Open industry standards** supported by a partner ecosystem of industry leaders.
- **A system that evolves and scales** to meet future data center needs for computing power, memory footprint, and I/O bandwidth. Our current Cisco UCS X-Series Modular System is already in its third generation, with two generations of fabric—and many years of innovation to come.

More than 16 years of innovation

When we first entered the server market in 2009, the challenge was to prove to our customers that we could out-innovate our competitors, and that we were committed to the market for the long term. Today, we are still the only vendor to offer a system that eliminates the tedious, manual, error-prone assembly of components into systems, providing instead a system that is self-aware and self-integrating and that brings true automation to IT operations.

Our commitment to the marketplace has been demonstrated by a rise to join the top tier of server manufacturers. We have continued to innovate and demonstrate our commitment to customers and to the server market. With multiple generations of fabric technology supporting modular upgrades to the system's connectivity, and with even more generations of Intel Xeon and AMD EPYC processors incorporated into our products, we are demonstrating our strong support for customer investments and our readiness to take our customers into the future.

For more information

- Cisco UCS
- Cisco Intersight
- Cisco converged infrastructure solutions
- Cisco hyperconverged infrastructure solutions
- Cisco solutions for hybrid clouds
- Cisco UCS management

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

C22-744677-03 08/25