Cisco UCS Manager: End-to-End Management for the Cisco Unified Computing System

The Challenge

Deploying server resources in today's data centers has become a difficult, time-consuming, and error-prone operation that requires careful coordination among a number of subject-matter experts, including server, network, and storage administrators. Today, even in organizations striving to move toward methodologies based on Information Technology Infrastructure Library (ITIL) concepts, administrators often must use separate element managers to configure every component manually, including firmware updates, BIOS parameter settings, interface configurations, and access-layer interconnect configurations for the LAN and SAN.

Even when utmost care is taken to perform these tasks according to prescribed standards, the manual steps and the use of different, independent element managers by different subject-matter experts makes this an error-prone process. A configuration task takes the same amount of time per task whether it needs to be performed once or 100 times. All these factors together lengthen the time needed to deploy or redeploy a server resource and cause administrators to spend their time on tedious, time-consuming activities rather than on higher-level, business-critical issues. Additionally, this disparate process makes deriving a single, unified audit log that helps maintain standards compliance and keep track of all the low-level, but critical, device configuration tasks difficult to achieve.

Introducing Cisco UCS Manager

Cisco® UCS Manager overcomes these problems by integrating compute, network, storage access, and virtualization into a cohesive system designed to reduce total cost of ownership (TCO) and increase business agility. Cisco UCS Manager provides centralized management capabilities, creates a unified management domain, and serves as the central nervous system of the Cisco Unified Computing System. It manages the system from end to end as a single cohesive system using an intuitive GUI, with both command-line interface (CLI) and XML API options, speeding configuration and reconfiguration of resources so that tasks that used to require days or hours now can be accomplished in minutes (Figure 1).

Cisco UCS Manager shifts administration tasks away from isolated, individual-system configuration that lacks context and visibility toward role- and policy-based management that provides visibility across the entire system. One manager instance can manage two Cisco UCS 6100 Series Fabric Interconnects, 40 Cisco UCS 5100 Series Blade Server Chassis, 80 Cisco UCS 2100 Series Fabric Extenders, and 320 Cisco UCS B-Series Blade Servers in a single management domain. Policy- and role-based administration not only reduces the chance of error; it allows systems to be designed and defined once, based on the specifications of each subject-matter expert, and then deployed repeatedly in a simple operation through one of the manager's user interfaces. This approach separates the more difficult architecture and design processes from implementation, enabling a more agile, adaptable data center while reducing the chance of errors that can cause application downtime or otherwise affect service levels.

Figure 1. Cisco UCS Manager Provides an Intuitive GUI That Supports End-to-End System Management
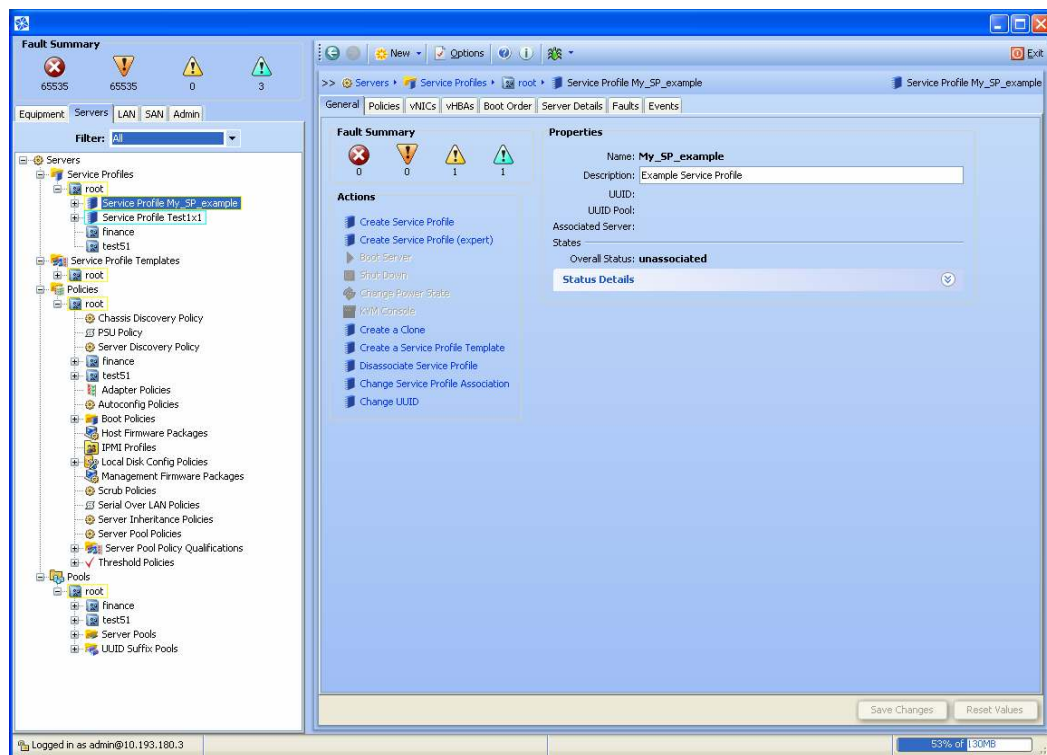
The manager's role-based design supports existing best practices, allowing server, network, and storage administrators to contribute their specific subject-matter expertise to a system design. Any administrator's role can be limited to a subset of the system's resources using *organizations* and *locales*, so that a Cisco Unified Computing System can be partitioned and shared among organizations using a multiple-tenant model.

Every server that is provisioned in the Cisco Unified Computing System is specified by a service profile. A *service profile* is a software definition of a server and its LAN and SAN network connectivity. Service profiles allow server resources to be treated as raw computing capacity that can be allocated and reallocated among application workloads, enabling a much more dynamic and efficient use of server capacity regardless of whether server virtualization is being used. Server deployment using service profiles takes minutes instead of days or weeks.

- **Service profiles** are created by server, network, and storage administrators and are stored in the Cisco UCS 6100 Series Fabric Interconnects. When a service profile is deployed on a server in the spare computing pool, the manager automatically configures the server, its network adapters, fabric extenders, and fabric interconnects to support the configuration specified in the service profile. This automatic device configuration reduces the number of manual steps required to configure servers, network interface cards (NICs), host bus adapters (HBAs), and LAN and SAN switches. The reduction in manual steps helps reduce human error, improves consistency, and reduces server deployment time. A service profile describes how all low-level firmware and device settings in the computing system need to be configured, including the RAID level of internal disk drives on the server, BIOS settings such as boot order, firmware revision levels for the BIOS and network adapters, NIC and HBA identities and settings, network settings for VLAN and VSAN membership and bandwidth quality of service (QoS), and uplink interface pinning.

- **Service profile templates** are also created by administrators, and they define policies for creating a specific kind of service profile. For example, they might specify the standard network connectivity for a web server and the pool from which its interface's MAC addresses can be obtained. Service profile templates can be used to provision many servers with the same simplicity as creating a single one.

Service profiles and their associated parameters can be created, modified, and monitored by external system management tools through a comprehensive XML API. The single API also facilitates integration with external configuration management databases (CMDBs) for inventory population and asset tracking, including tracking of granular configuration and state information. Service profiles configure resources that are logically below the operating system or hypervisor, so the manager can be used in conjunction with higher-level management tools that provision and patch operating systems, applications, and virtualization software.

In addition to managing the Cisco Unified Computing System, the manager supports standard monitoring and event systems, providing data to standard enterprise-management tools through a number of industry-standard APIs, including Simple Network Management Protocol (SNMP), Intelligent Platform Management Interface (IPMI), and Systems Management Architecture for Server Hardware Command Line Protocol (SMASH-CLP).
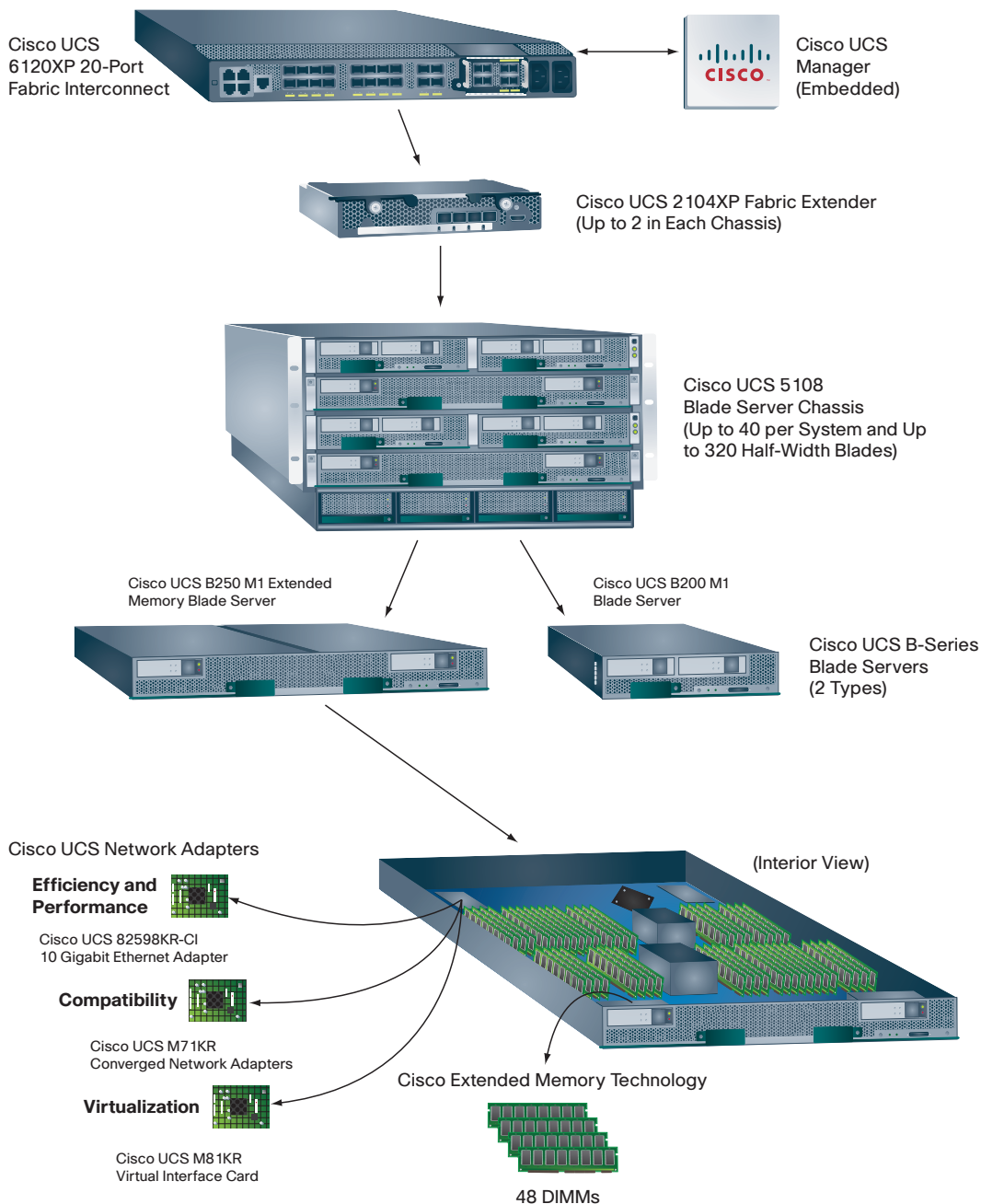
Cisco UCS Manager manages the elements in a Cisco Unified Computing System as a single cohesive system, and it does so without disrupting the current division of responsibilities and best practices of a data center. The remainder of this document describes Cisco UCS Manager.

Harnessing the Unified Computing System's Power

Cisco UCS Manager orchestrates the individual components of a Cisco Unified Computing System into servers that are networked together and connected to upstream aggregation-layer LAN and SAN switches. Every component of a Cisco Unified Computing System is designed to be flexible and malleable, with configuration, identity, and connectivity defined on demand to meet a specific business purpose. Components of a typical Cisco Unified Computing System include the following (Figure 2):

- A pair of Cisco UCS 6100 Series Fabric Interconnects with upstream LAN and SAN connectivity to the aggregation layer. Cisco UCS Manager resides on the interconnects, with the state synchronized between the two devices to increase the manager's availability so that it can continue to function through the loss of a single interconnect. The fabric interconnects support a unified network fabric that connects every server resource in the system: currently up to 40 blade system chassis and 320 blade servers.
- A pair of Cisco UCS 2100 Series Fabric Extenders bring the unified fabric into each blade chassis. The fabric extenders forward traffic using a cut-through architecture from blade-resident I/O adapters onto the unified fabric consisting of Cisco Data Center Ethernet (Cisco DCE™), Fibre Channel over Ethernet (FCoE), and additional standards-based extensions designed for greater network QoS management.
- Cisco UCS Network Adapters provide I/O interfaces to server resources in a mezzanine-card form factor. Three types of adapters offer a range of options to meet application requirements, including adapters optimized for virtualization, compatibility with existing driver stacks, and efficient, high-performance Ethernet. All network adapters have two unified fabric ports: one connected through the chassis midplane to each of the fabric extenders.

Figure 2.   Cisco Unified Computing System Consists of Fabric Interconnects, Fabric Extenders, Blade Server Chassis, Blade Servers, and Network Adapters

Cisco UCS
6120XP 20-Port
Fabric Interconnect

Cisco UCS
Manager
(Embedded)

Cisco UCS 2104XP Fabric Extender
(Up to 2 in Each Chassis)

Cisco UCS 5108
Blade Server Chassis
(Up to 40 per System and Up
to 320 Half-Width Blades)

Cisco UCS B250 M1 Extended
Memory Blade Server

Cisco UCS B200 M1
Blade Server

Cisco UCS B-Series
Blade Servers
(2 Types)

Cisco UCS Network Adapters

(Interior View)

**Efficiency and
Performance**

Cisco UCS 82598KR-CI
10 Gigabit Ethernet Adapter

Compatibility

Cisco UCS M71KR
Converged Network Adapters

Cisco Extended Memory Technology

Virtualization

Cisco UCS M81KR
Virtual Interface Card

48 DIMMs

- The Cisco UCS 82598KR-CI 10 Gigabit Ethernet Adapter is designed as an efficient, high-performance Ethernet interface.
- The Cisco UCS M71KR-E Emulex and M71KR-Q QLogic Converged Network Adapters (CNAs) present two 10 Gigabit Ethernet NICs and two 4-Gbps Fibre Channel HBAs to the operating system or hypervisor. Two models are available, with either Emulex or QLogic application-specific integrated circuits (ASICs), enabling the use of standard vendor-supplied drivers and management software, facilitating compatibility with existing SANs and storage arrays.

- The Cisco UCS M81KR Virtual Interface Card provides 128 virtual adapters whose type (Ethernet NIC or Fibre Channel HBA) and identity (MAC address and world-wide name [WWN]) are programmed on demand through software to match the adapter type and connectivity requirements of individual virtual machines. (Eight of the interfaces are reserved for use by the system.) This adapter, combined with the manager's integration with VMware VirtualCenter and ESX Server, implements Cisco VN-Link features that support network connections directly to individual virtual machines. Similar in operation to the Cisco Nexus™ 1000V Series Switches, the combination of adapter, manager, and VMware products provides policy-based virtual machine connectivity, mobility of network and security properties during virtual machine movement, and a nondisruptive operational model in which network administrators manage networks and server administrators manage servers.

- Up to 40 Cisco UCS 5100 Series Blade Server Chassis hold fabric extenders and blades. The chassis provides power and cooling, and its components and environmental characteristics are monitored by Cisco UCS Manager.

- The chassis house up to eight Cisco B-Series Blade Servers, for a total of up to 320 per system. The Cisco UCS B200 M1 Blade Server is a half-width, dual-socket server hosting up to two Intel Xeon 5500 series processors, up to 96 GB of main memory, up to two Small Form Factor (SFF) Serial Attached Small Computer System Interface (SAS) disk drives with onboard RAID 0 and 1, and a single network adapter slot. The Cisco UCS B250 M1 Extended Memory Blade Server is a full-width, dual-socket server hosting up to two Intel Xeon 5500 Series processors, up to 384 GB of main memory, up to two SFF disk drives, and two network adapter slots.

Wire-Once Model

The Cisco Unified Computing System is designed with a "wire once, walk away" model in which:

- Cabling and network infrastructure support a unified network fabric in which features such as FCoE can be enabled through Cisco UCS Manager as needed.

- Every element in the hierarchy is programmable and managed by Cisco UCS Manager using a just-in-time resource provisioning model.

- The manager can configure identity information including the universally unique identifier (UUID) of servers, MAC addresses, and WWNs of network adapters.

- It can install consistent sets of firmware throughout the system hierarchy, including each blade's baseboard management controller (BMC), RAID controller, network adapter firmware, and fabric extender firmware.

- It can configure the operational characteristics of every component in the hierarchy, from the hardware RAID level of onboard disk drives to uplink port configurations on the Cisco UCS 6100 Series Fabric Interconnects and everything in between.

- It can configure the types of I/O interfaces on Cisco UCS M81KR Virtual Interface Card adapters. The importance of this capability cannot be understated: when a server resource is configured with this mezzanine card, the number, type (HBA or NIC), and identities (WWNs and MAC addresses) of I/O interfaces can be programmed using just-in-time provisioning. This approach allows a server resource to support a traditional OS and application software stack with a pair of Ethernet NICs and Fibre Channel HBAs at one moment and then be rebooted to run a virtualized environment with a combination of up to 128 NICs and HBAs, with NICs connected directly to virtual machines through hypervisor pass-through technology.

Servers with Identity Programmed on Demand

The hierarchy of components comprising a Cisco Unified Computing System, combined with the capability to harness them using Cisco UCS Manager, means that organizations can deploy servers, and sets of servers, with configurations and identities determined with a just-in-time provisioning model. Cisco UCS Manager automates these operations and performs them at scale, helping increase not just efficiency, but also accuracy. This automation puts capital resources to work more quickly, and it allows administrators to spend more time focusing on business-level problems rather than administrative details. The capability to move UUID, WWN, and MAC identities from server to server enables a truly dynamic data center because the physical SAN and LAN infrastructure does not need to be changed. For example, consider a SAN in which zoning is performed based on the WWN of the HBA, and the WWN is moved from one blade to another based on a service profile. In this example, the SAN zoning and the programming in the storage array controller can remain unchanged because the interface's identity remains the same.

Automated Device Discovery

Cisco UCS Manager performs automated device discovery, recognizing any resource as soon as it is plugged into the Cisco Unified Computing System. After a resource is installed, the manager adds it and its characteristics to the system inventory, even preconfiguring it if directed to by an administrator-defined policy. Cisco UCS Manager can group resources into *pools* by type and allocate them to production use based on specific characteristics, such as number of CPUs, amount of memory, and type of network adapter installed. Figure 3 shows how Cisco UCS Manager presents the system inventory, information that can be exported to satisfy capital asset management and audit requirements.

Figure 3.   Cisco UCS Manager Discovers Resources When They Are Plugged into the System, Adding Them to Inventory or Provisioning Them Automatically
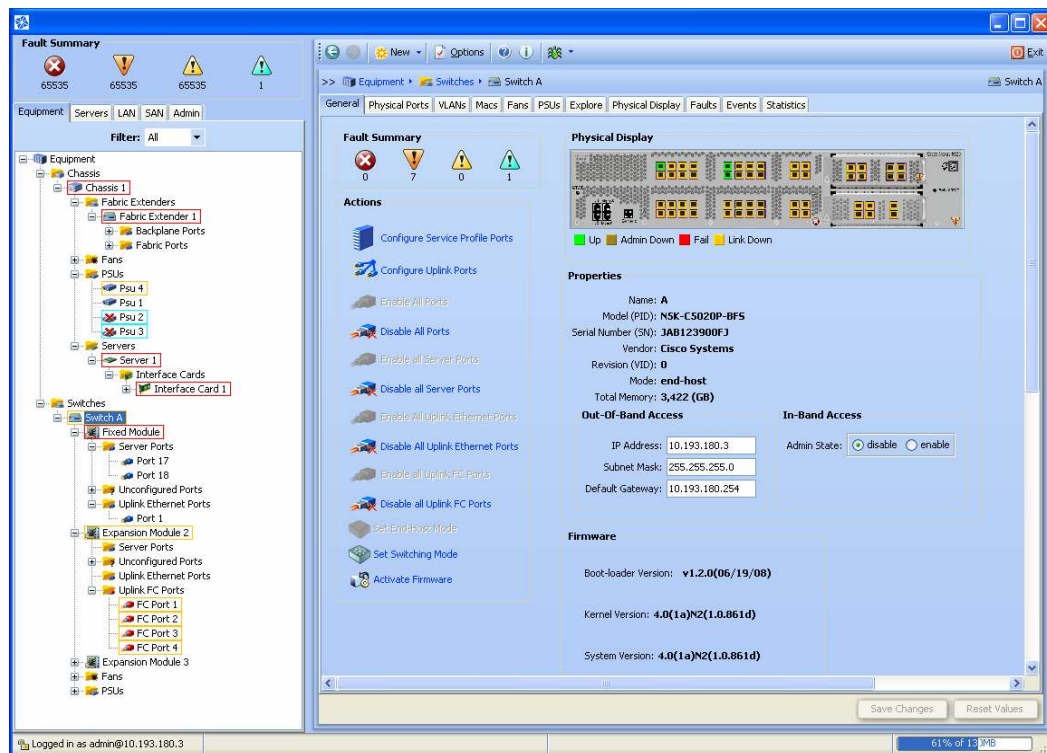
Application Examples

The capability to select stateless hardware resources from a pool and configure them using a just-in-time provisioning model dramatically simplifies a wide range of typical data center deployment scenarios:

- An application running on a server with 96 GB of memory can be scaled to run entirely in memory by moving it to a server with 384 GB of memory installed. Cisco UCS Manager moves the service profile to a new server with resources drawn from a pool matching the required memory characteristics. The server's identity and I/O interfaces are configured exactly as the existing server's to avoid any incompatibilities where these addresses have been configured or cached. The operating system and application software see only a simple shutdown and reboot process and then they are up and running on the new server.

- A dozen new web server instances are deployed so that service levels can be met during an unexpected workload spike. Cisco UCS Manager draws the servers from the appropriate pool and invokes a service template to provision each one with a service profile. Each server is provisioned with the same I/O configuration and network connectivity as the other front-end web servers used by the application.

- A VMware ESX Server cluster is expanded by adding a server with 384 GB of main memory. Cisco UCS Manager provisions the server so that its network connects to the correct VLAN and is part of the correct SAN zone so the hypervisor can access the shared virtual disk files of the cluster. This behavior allows VMware VirtualCenter software with VMware Dynamic Resource Scheduling to migrate running virtual machines to the new server instance with I/O configuration and network profiles following automatically without intervention by the administrator.

Unified, Cohesive System

Cisco UCS Manager orchestrates the resources in a Cisco Unified Computing System so that they work together as a single, cohesive system. Externally, the fabric interconnects pin server MAC addresses and WWNs to specific uplink interfaces for connectivity to standard, non-unified Ethernet and Fibre Channel aggregation-layer switches. Internally, Cisco UCS Manager handles all the element management necessary to provision servers and network them together with designated profiles that fully specify a configuration.

In its role as an element manager, Cisco UCS Manager automates element management from end to end, essentially eliminating the tedious and error-prone use of individual element managers typically required to provision a server, its interfaces, and its network interconnects. The elements managed by Cisco UCS Manager include:

- BMC firmware
- RAID controller firmware and settings
- BIOS firmware and settings, including server UUID and boot order
- Network adapter firmware and settings, including MAC and WWN addresses and SAN boot settings
- Virtual port groups used by virtual machines using Cisco VN-Link technology
- Interconnect configuration, including uplink and downlink definitions, MAC and WWN address pinning, VLANs, VSANs, QoS, bandwidth allocation, VN-Link settings, and EtherChannels to upstream LAN switches

Raising the Level of Abstraction

In addition to simplifying element management, Cisco UCS Manager raises the level of abstraction at which elements are managed. For example, the VN-Link feature of the Cisco Unified Computing System fabric interconnects link virtual NICs connected to virtual machines to virtual ports residing in the fabric interconnects. In virtualized environments, Cisco UCS Manager enables the synchronization of virtual port profiles between VMware VirtualCenter, ESX Server and the unified network fabric. After the initial configuration is complete, port profiles are synchronized with virtual machine movement though the control plane. The result is that NICs are created and connected to virtual machines and network profiles are set using a just-in-time model even as VMware VMotion moves virtual machines between servers. This process is automatic, with no intervention required on the part of administrators in either VMware VirtualCenter or Cisco UCS Manager tools.

Separate Management, Control, and Data Planes

The Cisco Unified Computing System is designed with a strict separation of management, control, and data planes. Once Cisco UCS Manager configures a system, the system runs without interaction with the management plane. The system accomplishes tasks -- such as moving network profiles along with virtual machines -- completely through control plane interactions. Even though Cisco UCS Manager operates in a high-availability configuration with synchronized state and failover between the two fabric interconnects, the manager can fail, or be taken out of service while its software is upgraded, without any impact on the system's overall operation. This also means the manager can be shut down and upgraded to a new version without impacting the data and control planes. This means the OS and application can still send and receive LAN and SAN traffic even while the manager is disabled. Some critical functions, such as fabric extender configuration and firmware upgrade, are initially set up by Cisco UCS Manager and then executed

via the control plane so that this critical firmware -- and the continuous network connectivity that it supports -- operate without any management plane interaction.

Service Profiles and Templates

Out of the box, the Cisco Unified Computing System is a stateless collection of computing resources that are harnessed into a set of interconnected servers by Cisco UCS Manager. Service profiles describe how the main elements, including server, networking, and storage, are configured and interconnected. Every server definition in the Cisco Unified Computing System begins with a service profile, so there is no way to subvert the process of creating a well-defined server element model before actually asking Cisco UCS Manager to provision the server.

Support for ITIL-Based Processes

The Cisco UCS Manager solution is good news for organizations striving to implement ITIL-based processes that are based on accurate configuration management databases. For a server to be provisioned from its raw resources and then connected to the network, its configuration must be completely specified through Cisco UCS Manager. The configuration is stored in the manager's internal database and can be exported to a data center's CMDB and to higher-level tools that can perform OS and application software provisioning. When a service profile is applied to provision a server, all the element-management operations at every level of the hardware stack are handled by Cisco UCS Manager, eliminating the tedious and error-prone manual operations required in discrete server environments.

Embodying Data Center Best Practices

A service profile embodies a data center's best practices and the coordinated efforts of server, network, and storage administrators. With the efforts of subject-matter experts focused on policies and service profiles from the very beginning, the use and reuse of service profiles become routine operations that can be performed by administrators with lower skill levels, thus freeing higher-level administrators to focus on business-domain-specific issues rather than on routine and tedious element management chores and so making more effective use of limited administrator time.

Service Profile Specifications

A service profile defines a single server and its networking characteristics. It includes or specifies four types of information:

- Server definitions that define the resources that are required to apply the profile: For example, a service profile may be constrained to apply to a specific server resource, a blade inserted into a specific chassis slot, or a server from a predefined pool of server resources with specified CPU, RAM, or other administrator-defined characteristics.
- Identity information that is used to create a server's identity: This information includes the UUID, MAC address for each virtual NIC (vNIC), and WWN specifications for each HBA.
- Firmware revision specifications required for a specific server: For example, the definition of an online transaction processing (OLTP) database server to be connected to a specific high-performance Fibre Channel storage system may require that the storage system vendor's approved and tested firmware revision be installed on the HBA portion of the CNA. A different version of HBA firmware may be required for a data warehousing system that uses a different vendor's storage array to access cheaper, high-capacity archival storage.
- Connectivity definition used to configure network adapters, fabric extenders, and parent interconnects: This information is abstract in that it does not include the details of how each

network component is configured. Instead, the connectivity information specifies the desired end result, and then Cisco UCS Manager takes the necessary actions at various levels of the hardware hierarchy to put the required connectivity into effect. An example of the use of this information is the VLAN configuration given to each vNIC. The manager configures the appropriate ports based on the server on which the service profile is selected to run. Cisco UCS Manager understands the system's internal topology so it finds and configures the port implicitly. This process is much simpler than the manual process of VLAN configuration based on the downlink port of a traditional access-layer switch.

Most of a service profile is a specification of which policies are to be used to define specific values at the moment when a service profile is actually used to provision a server. The late binding time of actual values to service profiles is part of the role- and policy-based design of Cisco UCS Manager. For example, a service profile may specify that an interface's MAC address is to be obtained from a specific pool, or that a VLAN number is to be determined by a policy defined for an organization's web servers.

The interaction of roles and policies is discussed in "Policy-Based Management" and "Role-Based Access Control" later in this document; for the purpose of this section, we assume that a service profile is bound to specific values.

Discrete Server Service Profile Example

Figure 4 shows how a service profile is used to harness a set of physical components and provision a server and its network interconnects. This figure illustrates a simple case of a discrete server; provisioning for virtualized environments adds a layer of configuration for virtual interfaces.

- At the left is the set of elements that are used: a Cisco UCS B-Series Blade Server, a Cisco UCS M71KR Converged Network Adapter, two Cisco UCS 2100 Series Fabric Extenders, and two Cisco UCS 6100 Series Fabric Interconnects.

- The middle of the figure shows how these elements are configured and interconnected based on the service profile. The CNA is connected to two parent fabric interconnects through a pair of fabric extenders. The CNA presents two Ethernet NICs and two Fibre Channel HBAs to the operating system. The virtual links between the NICs and HBAs and interconnect uplink ports are shown as blue dotted lines (Ethernet) and orange dotted lines (Fibre Channel). The actual network traffic is multiplexed onto the physical unified fabric connection shown in red.

- At the right is the service profile along with examples of the kinds of information used at each physical layer to provision a server. Note that the fabric extender configuration is specified indirectly by the interconnect settings. The fabric extender configuration and firmware are maintained by the interconnect through the control plane, essentially eliminating the risk of inconsistent firmware or settings that could result in a loss of connectivity. OS and application provisioning can be accomplished using higher-level tools that integrate with Cisco UCS Manager through its XML API.

Figure 4. Service Profiles Specify How Physical Components Are Used to Provision Servers and Their Interconnects
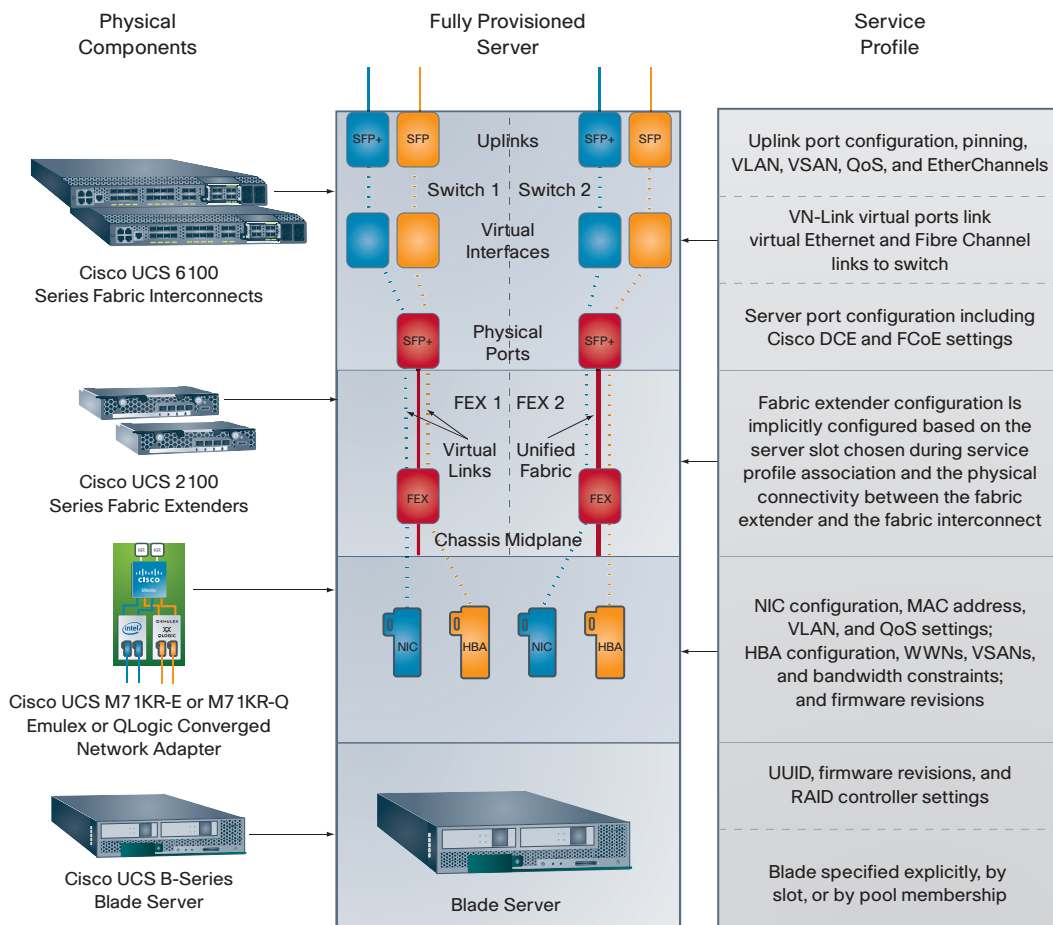
Virtual Machine Mobility In Operation

When servers with Cisco UCS M81KR Virtual Interface Cards are provisioned, virtual HBAs and NICs that are needed by the virtualization software are configured through the service profile so that the hypervisor layer can be provisioned and connected to both networks and storage.

The Cisco UCS M81KR Virtual Interface Card creates vNICs that connect directly to virtual machines through hypervisor pass-through switching or through hypervisor bypass technologies that eliminate the overhead and management complexity of adding another layer of switching. These vNICs are created and configured dynamically, not as part of the server provisioning, but as part of the implementation of the Cisco VN-Link technology. This approach allows virtual machines to be created as needed and moved between servers with VMware VMotion, or through cold migration, with no human interaction needed after initial configuration. The management plane keeps the network policies tied to the virtual machine, regardless of the physical server on which the virtual machine happens to be running.

The process begins with an administrator creating a named *port profile* (or *port group*) using Cisco UCS Manager. The port profile fully defines a set of vNIC properties and their connectivity. It specifies VLAN configuration, network QoS policies, and other network connectivity configurations. This information allows a virtual link to be established over the unified fabric connecting a vNIC to a virtual port in the parent fabric interconnect.

The names of port profiles are made available to VMware VirtualCenter so they become available as menu selections when vNICs in virtual machines are created. When VMware VirtualCenter directs VMware ESX Server to start the actual virtual machine, VMware ESX Server obtains the port profile name from the virtual machine definition, communicates the desired port profile's name to the virtual interface card, and sets up the desired pass-through connection. The virtual interface card retrieves the port profile details through the Cisco Unified Computing System's control plane, and it creates the required virtual ports and links with appropriate VLAN connectivity, network QoS settings, and other associated network parameters.

Note that after the port profile is set up by Cisco UCS Manager and communicated to VMware VirtualCenter and the vNIC in the virtual machine is created in VMware VirtualCenter and the port profile name is selected, all further network configuration is accomplished over the control plane with no further configuration management required by either Cisco UCS Manager or VMware VirtualCenter. This approach allows virtual machines to move between servers with no manual interaction. This approach also means that wide-open VLAN trunks are not needed to all servers in the VMware cluster since VLANs are allocated on per-vNIC and per–virtual machine bases, not on the basis of a virtual machine vNIC or physical NIC.

When a virtual machine is directed to move, the same port configuration process is performed on the destination VMware ESX Server instance. This is the mechanism that allows both NICs and network profiles to be configured dynamically. In the data center of the future, in which virtual machine movement will be more common than it is today, it will be important that the network configuration change dynamically based on virtual machine location, rather than relying on the static port-configuration model in common use today.

Full Automation with Service Profile Templates

While a service profile defines a single server and interconnections, a service profile template defines policies that can be applied any number of times to provision any number of servers. A service template can be derived from a service profile, with server and I/O interface identity information abstracted. Instead of specifying exact UUID, MAC address, and WWN values, a service template specifies where to get these values. Typically, a pool of UUID, MAC address, and WWN values is created in Cisco UCS Manager's inventory, and a service profile indicates the pools from which these values are obtained when it is applied.

Service profile templates help enable large-scale operations in which many servers are provisioned as easily as a single server. Operations that might ordinarily require hours or days of coordinated effort by subject-matter experts now can be performed in minutes. Server resources can be put into service more rapidly, making more effective use of capital resources. Administrator time is reduced by the manager's end-to-end configuration and provisioning capabilities. Most important is the capability to easily configure and reconfigure servers to more rapidly align IT resources with business objectives.

Policy-Based Management

Policy-based management is one of the primary features that helps IT organizations better define and implement their own best practices. Policies help ensure that only consistent, tested, and standards-compliant systems are provisioned, reducing the risk of problems caused by repetitive manual tasks. Future data centers will be increasingly dynamic, and technology for the definition, consumption, and resolution of policies is crucial for making infrastructure devices that can be reconfigured dynamically.

Building on Subject-Matter Expert Knowledge

Subject-matter experts create policies that establish menu options that can then be selected by lower-level administrative staff when provisioning resources with service profiles or templates. Policies are named so that their purpose is clear. For example, a storage administrator might create an adapter profile named "OLTP database server" to specify the firmware certified to work with the server's destination storage system and other adapter parameters such as QoS and WWNs. Similarly, a network administrator might create "database server" QoS and VLAN policies that specify QoS and VLAN parameters, respectively.

Constraining Choices with Pools

Pools similarly constrain the choices that can be used when provisioning servers. Administrators can establish pools of unique identifiers, such as MAC addresses and WWNs. Resources are consumed from pools as service profiles and templates are used to provision servers.

Server resources can be placed in pools based on various criteria. Service profiles and templates can require the use of servers from specific pools, thereby constraining the characteristics of the servers to which they apply. For example, a "large memory" pool of servers can be created by adding all servers with more than 128 GB of memory. A "small memory" pool of servers can be created by adding all servers with 16 GB or less of memory. A database server service template might require a server from the "large memory" pool, and a web server template might require a server from the "small memory" pool. Discovery policies can be created so that servers are automatically placed in an appropriate pool the moment they are inserted into a chassis.

Audit Logs and Compliance

In further support of a policy-based data center, audit logs can provide concrete evidence regarding how and when servers are configured and deployed, helping organizations meet their obligations under the rising number of governmental regulations to which they must comply. Cisco UCS Manager implements a fused management plane in which server, network, and storage administrators all use the same management system to configure servers and their interconnections. This approach makes audit logs complete and accurate, in contrast with the approach otherwise required to configure and deploy a server in a traditional environment, which relies on the difficult task of trying to merge audit logs from every element and element manager.

Role-Based Access Control

Role-based access control (RBAC) allows organizations to continue to use their existing organizational structure to manage a Cisco Unified Computing System with no changes required. This feature helps organizations use current staff skills without the need for retraining or cross-training. It also helps them to make more effective use of highly skilled administrators while empowering lower-level administrators to perform well-defined tasks.

Cisco UCS Manager supports roles that, at a fine-grained level, define the actions that can be taken by an administrator assuming a specific role. A super-administrator defines initial roles and specifies which administrators are allowed to assume what roles. The Cisco UCS Manager comes with server, network, and storage administrator roles predefined. These roles can be modified, merged, and deleted, and new roles can be created to fit the organization model in place. Coordination between roles is simplified on the Cisco Unified Computing System because, although roles are separated, an administrator assuming one role can view the actions taken by administrators having other roles. For example, a storage administrator can set up Fibre Channel configuration options to see the choices that a network administrator has made when setting up
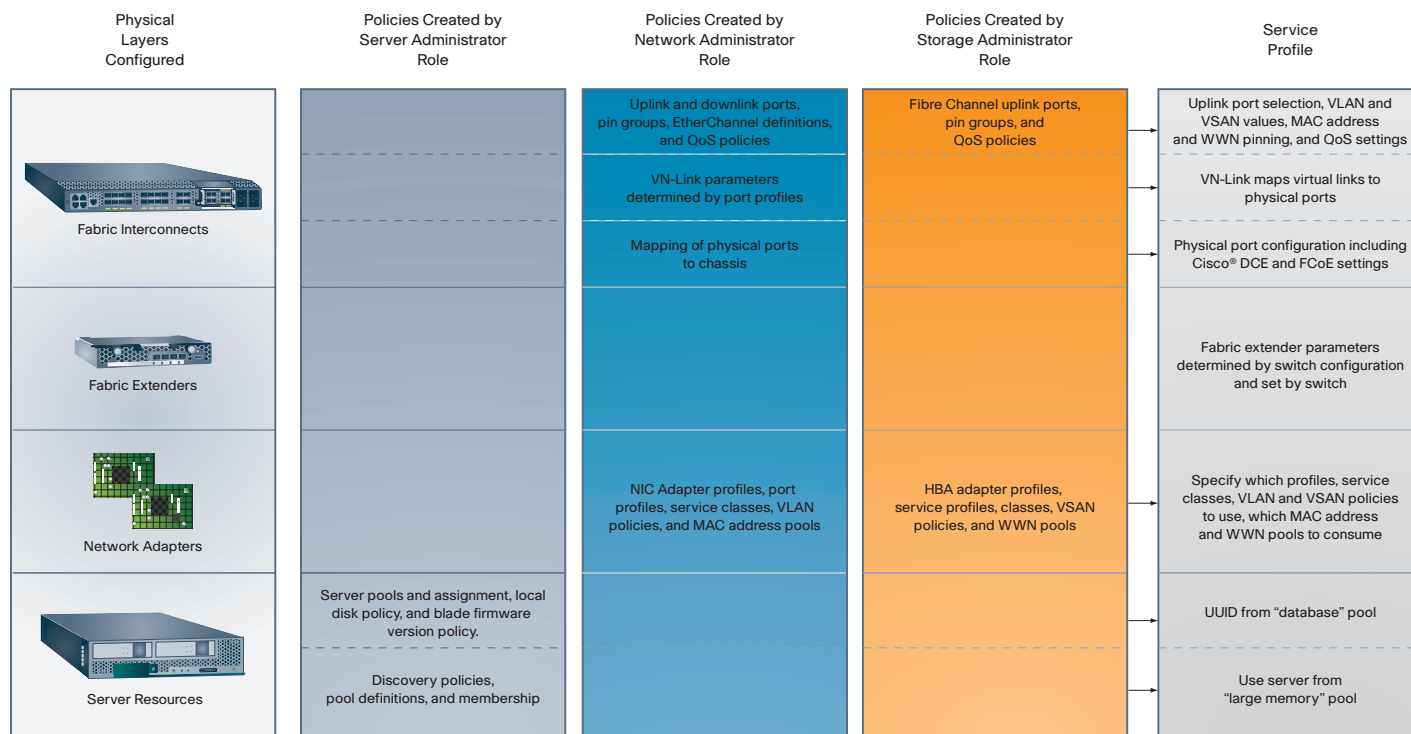
network options; the same is true for the settings applied by the server administrator. Visibility between roles helps eliminate ambiguity and reduce the chance of error due to miscommunication or lack of communication that may occur when administrators instead rely on phone calls, tickets, spreadsheets, or email.

Intersection of Policies and Roles

Depending on the size of the organization, various levels of administrator authority can be defined through roles. Creating a set of layered roles allows subject-matter experts to focus on high-level configuration issues and allows lower-level administrators to implement the configurations. For example:

- Server, network, and storage subject-matter expert roles might define a set of policies appropriate for their specific domains. For instance, each expert might define a set of domain-specific choices appropriate for provisioning each specific type of server, such as a web, database, or application server.
- The next level of administrator might be allowed by role to choose from the policies defined by the subject-matter experts to create service profiles or templates appropriate for specific server types. For instance, this level of administrator might be allowed to choose the database server network adapter profile and the database server host-bus adapter profile (along with other profiles) to create a database server template. This administrator might also be allowed to use the template to provision servers.
- A lower-level administrator might be allowed to use only existing templates to provision servers, with no choice of specific policies allowed.
- Figure 5 shows the resource layers in a Cisco Unified Computing System and the kinds of policies that might be created for each layer by each role. The resulting service profile or template at the right illustrates the result of choosing from the policies defined by the subject-matter experts.

Figure 5.   Types of Policies Defined in Cisco UCS Manager

Physical Layers Configured	Policies Created by Server Administrator Role	Policies Created by Network Administrator Role	Policies Created by Storage Administrator Role	Service Profile
Fabric Interconnects		Uplink and downlink ports, pin groups, EtherChannel definitions, and QoS policies	Fibre Channel uplink ports, pin groups, and QoS policies	Uplink port selection, VLAN and VSAN values, MAC address and WWN pinning, and QoS settings
		VN-Link parameters determined by port profiles		VN-Link maps virtual links to physical ports
		Mapping of physical ports to chassis		Physical port configuration including Cisco® DCE and FCoE settings
Fabric Extenders				Fabric extender parameters determined by switch configuration and set by switch
Network Adapters		NIC Adapter profiles, port profiles, service classes, VLAN policies, and MAC address pools	HBA adapter profiles, service profiles, classes, VSAN policies, and WWN pools	Specify which profiles, service classes, VLAN and VSAN policies to use, which MAC address and WWN pools to consume
Server Resources	Server pools and assignment, local disk policy, and blade firmware version policy.			UUID from "database" pool
	Discovery policies, pool definitions, and membership			Use server from "large memory" pool

System Monitoring

In addition to configuring server and I/O resources, Cisco UCS Manager monitors the entire system, providing fine-grained data on all resources from fan speeds to CPU die temperatures. Cisco UCS Manager integrates with existing network and enterprise management systems through SNMP, IPMI, and other industry-standard interfaces. Cisco UCS Manager integrates environmental data in the fabric interconnects, which in turn make the data available through SNMP. In addition, OS-based agents from system management tools will run without modification, behaving exactly as they do in traditional server environments.

Conclusion

Cisco UCS Manager integrates the components of a Cisco Unified Computing System to create a single, cohesive system. The manager helps increase return on investment (ROI) by helping data centers make more efficient use of their resources. It helps reduce application downtime by automating routine configuration operations and helping ensure that consistent, tested sets of policies are applied to system components when servers are configured. The just-in-time provisioning model speeds configuration and reconfiguration of system resources, turning server, network, and storage configuration into simple click-of-the-mouse operations. This model can dramatically reduce server deployment time, helping reduce the time required for activities that once took hours or days to mere minutes.

Service profiles and templates make effective use of subject-matter expertise, allowing server, network, and storage administrators to collaborate on server definitions that can be created once and then performed many times as routine operations. These operations can be performed by authorized administrators without detailed subject-matter expertise, freeing the experts to focus on business-critical activities and less on routine, day-to-day operations.

Compliance with existing standards, regulations, and data center best practices is enforced by the policy-based nature of Cisco UCS Manager. Collaboration and sharing is supported by the combination of RBAC and the capability for each subject-matter expert to view the configurations established by others. RBAC supports the continued use by data centers of their existing best practices, staff roles, and enterprise management systems with the Cisco Unified Computing System without disruption.

For organizations implementing ITIL-based processes, Cisco UCS Manager helps codify and enforce best practices. It maintains and conducts all operations based on an internal configuration database that can be exported to populate CMDBs and also to support higher-level software provisioning tools. The Cisco Unified Computing System advances the state of the art by putting the configuration management first and automating all element management as a side-effect of the highly structured service profile and template creation process.

Regardless of how focused your data center is on implementing ITIL-based processes, and regardless of whether you are using a traditional server deployment model or pools of virtualized resources, Cisco UCS Manager can automate and reduce the risk of manual element management processes, helping increase business agility, reduce TCO, and deliver the full benefits of the Cisco Unified Computing System to your data center.

For More Information

For more information about Cisco UCS Manager, please visit http://www.cisco.com/go/unifiedcomputing or contact your local Cisco representative.