Cisco UCS B200 M4 Blade Server Installation and Service Note

First Published: 2015-08-28

Last Modified: 2020-07-08

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

The following information is for FCC compliance of Class A devices: This equipment has been tested and found to comply with the limits for a Class A digital device, pursuant to part 15 of the FCC rules. These limits are designed to provide reasonable protection against harmful interference when the equipment is operated in a commercial environment. This equipment generates, uses, and can radiate radio-frequency energy and, if not installed and used in accordance with the instruction manual, may cause harmful interference to radio communications. Operation of this equipment in a residential area is likely to cause harmful interference, in which case users will be required to correct the interference at their own expense.

The following information is for FCC compliance of Class B devices: This equipment has been tested and found to comply with the limits for a Class B digital device, pursuant to part 15 of the FCC rules. These limits are designed to provide reasonable protection against harmful interference in a residential installation. This equipment generates, uses and can radiate radio frequency energy and, if not installed and used in accordance with the instructions, may cause harmful interference to radio communications. However, there is no guarantee that interference will not occur in a particular installation. If the equipment causes interference to radio or television reception, which can be determined by turning the equipment off and on, users are encouraged to try to correct the interference by using one or more of the following measures:

- Reorient or relocate the receiving antenna.
- Increase the separation between the equipment and receiver.
- Connect the equipment into an outlet on a circuit different from that to which the receiver is connected.
- Consult the dealer or an experienced radio/TV technician for help.

Modifications to this product not authorized by Cisco could void the FCC approval and negate your authority to operate the product.

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

All printed copies and duplicate soft copies of this document are considered uncontrolled. See the current online version for the latest version.

Cisco has more than 200 offices worldwide. Addresses and phone numbers are listed on the Cisco website at www.cisco.com/go/offices.

The documentation set for this product strives to use bias-free language. For purposes of this documentation set, bias-free is defined as language that does not imply discrimination based on age, disability, gender, racial identity, ethnic identity, sexual orientation, socioeconomic status, and intersectionality. Exceptions may be present in the documentation due to language that is hardcoded in the user interfaces of the product software, language used based on standards documentation, or language that is used by a referenced third-party product.

CONTENTS

PREFACE

Preface vii

Audience vii

Conventions vii

Related Cisco UCS Documentation ix

Obtaining Documentation and Submitting a Service Request ix

CHAPTER 1

Overview 1

Cisco UCS B200 M4 Blade Server 1

External Features Overview 2

LEDs 2

Buttons 3

Local Console Connection 3

Secure Digital Cards 4

Storage Module 5

CHAPTER 2

Installing a Blade Server 7

Installing a Half-width Blade Server 7

Server Configuration 8

Powering Off a Blade Server Using the Power Button 9

Removing a Blade Server 10

Server Troubleshooting 10

CHAPTER 3

Servicing a Blade Server 11

Replacing a Drive 11

Removing a Blade Server Hard Drive 12

Installing a Blade Server Drive 12

4K Sector Format SAS/SATA Drives Considerations 12
Setting Up UEFI Mode Booting in the UCS Manager Boot Policy 13
Removing a Blade Server Cover 14
Air Baffles 14
Internal Components 15
Diagnostics Button and LEDs 15
Installing a CMOS Battery 16
Installing the FlexStorage Module 17
Replacing the SuperCap Module 17
Removing the SuperCap Module 18
Installing the SuperCap Module 19
Upgrading to Intel Xeon E5-2600 v4 CPUs 20
Removing a Heat Sink and CPU 21
Installing a New CPU and Heat Sink 22
Installing Memory 25
Supported DIMMs 25
Memory Population 25
DIMMs and Channels 25
Memory Performance 27
Memory Mirroring and RAS 27
Installing a Virtual Interface Card Adapter 27
Installing an Adapter Card in Addition to the VIC mLOM Adapter 28
Installing the NVIDIA M6 GPU Adapter Card 30
Enabling the Trusted Platform Module 33
CHAPTER 4 Technical Specifications 35
Physical Specifications for the Cisco UCS B200 M4 Blade Server 35
APPENDIX A NVIDIA Licensing Information 37
NVIDIA GRID License Server Overview 37
Registering Your Product Activation Keys with NVIDIA 38
Downloading the GRID Software Suite 39
Installing NVIDIA GRID License Server Software 39
Platform Requirements for NVIDIA GRID License Server 39

Installing on Windows 40

Installing on Linux 40

Installing GRID Licenses From the NVIDIA Licensing Portal to the License Server 41

Accessing the GRID License Server Management Interface 41

Reading Your License Server's MAC Address 41

Installing Licenses From the Licensing Portal 42

Viewing Available Licenses 42

Viewing Current License Usage 42

Managing GRID Licenses 43

Acquiring a GRID License on Windows 43

Acquiring a GRID License on Linux 44

Installing Drivers to Support the NVIDIA GPU Cards 44

1. Updating the Server BIOS Firmware 45

2. Activating the Server BIOS Firmware 45

3. Updating the NVIDIA Drivers 46

Preface

- Audience, on page vii
- Conventions, on page vii
- Related Cisco UCS Documentation, on page ix
- Obtaining Documentation and Submitting a Service Request, on page ix

Audience

To use this installation guide, you must be familiar with electronic circuitry and wiring practices and preferably be an electronic or electromechanical technician who has experience with electronic and electromechanical equipment.

Only trained and qualified service personnel (as defined in IEC 60950-1 and AS/NZS60950) should install, replace, or service the equipment. Install the system in accordance with the U.S. National Electric Code if you are in the United States.

Conventions

Text Type	Indication
GUI elements	GUI elements such as tab titles, area names, and field labels appear in **this font**. Main titles such as window, dialog box, and wizard titles appear in **this font**.
Document titles	Document titles appear in *this font*.
TUI elements	In a Text-based User Interface, text the system displays appears in `this font`.
System output	Terminal sessions and information that the system displays appear in `this font`.
CLI commands	CLI command keywords appear in **this font**. Variables in a CLI command appear in *this font*.
[ ]	Elements in square brackets are optional.

Text Type	Indication
{x | y | z}	Required alternative keywords are grouped in braces and separated by vertical bars.
[x | y | z]	Optional alternative keywords are grouped in brackets and separated by vertical bars.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.
< >	Nonprinting characters such as passwords are in angle brackets.
[ ]	Default responses to system prompts are in square brackets.
!, #	An exclamation point (!) or a pound sign (#) at the beginning of a line of code indicates a comment line.

Note Means *reader take note*. Notes contain helpful suggestions or references to material not covered in the document.

Tip Means *the following information will help you solve a problem*. The tips information might not be troubleshooting or even an action, but could be useful information, similar to a Timesaver.

Timesaver Means *the described action saves time*. You can save time by performing the action described in the paragraph.

Caution Means *reader be careful*. In this situation, you might perform an action that could result in equipment damage or loss of data.

Warning IMPORTANT SAFETY INSTRUCTIONS

This warning symbol means danger. You are in a situation that could cause bodily injury. Before you work on any equipment, be aware of the hazards involved with electrical circuitry and be familiar with standard practices for preventing accidents. Use the statement number provided at the end of each warning to locate its translation in the translated safety warnings that accompanied this device.

SAVE THESE INSTRUCTIONS

Related Cisco UCS Documentation

Documentation Roadmaps

For a complete list of all B-Series documentation, see the *Cisco UCS B-Series Servers Documentation Roadmap* available at the following URL: https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/overview/guide/UCS_roadmap.html

For a complete list of all C-Series documentation, see the *Cisco UCS C-Series Servers Documentation Roadmap* available at the following URL: https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/overview/guide/ucs_rack_roadmap.html.

For information on supported firmware versions and supported UCS Manager versions for the rack servers that are integrated with the UCS Manager for management, refer to Release Bundle Contents for Cisco UCS Software.

Obtaining Documentation and Submitting a Service Request

For information on obtaining documentation, submitting a service request, and gathering additional information, see the monthly What's New in Cisco Product Documentation, which also lists all new and revised Cisco technical documentation.

Subscribe to the *What's New in Cisco Product Documentation* as a Really Simple Syndication (RSS) feed and set content to be delivered directly to your desktop using a reader application. The RSS feeds are a free service and Cisco currently supports RSS version 2.0.

Follow Cisco UCS Docs on Twitter to receive document update notifications.

CHAPTER 1

Overview

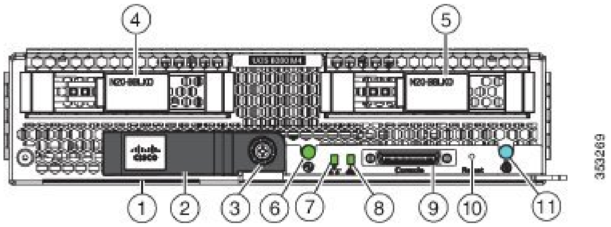
This chapter contains the following sections:

- Cisco UCS B200 M4 Blade Server, on page 1
- External Features Overview, on page 2
- Storage Module, on page 5

Cisco UCS B200 M4 Blade Server

The Cisco UCS B200 M4 is a density-optimized, half-width blade server that supports two CPU sockets for Intel E5-2600 v4 and v3 series CPUs and up to 24 DDR4 DIMMs. It supports one modular LOM (dedicated slot for Cisco's Virtual Interface Card) and one mezzanine adapter. In addition, the UCS B200 M4 supports an optional FlexStorage module that supports up to 2 SAS or SATA hard drives or solid state disks (SSDs). The UCS B200 M4 also supports PCIe NVMe SSDs. You can install up to eight UCS B200 blade servers in a UCS chassis, mixing with other models of Cisco UCS blade servers in the chassis if desired.

Figure 1: Cisco UCS B200 M4 Front Panel



1	Asset pull tag Each server has a blank plastic tag that pulls out of the front panel which is provided so that you can add your own asset tracking label without interfering with the intended air flow.	**2**	Blade ejector handle
3	Ejector captive screw	**4**	Hard drive bay 1
5	Hard drive bay 2	**6**	Power button and LED
7	Network link status LED	**8**	Blade health LED

9	Local console connector	**10**	Reset button access
11	Locator button and LED		

External Features Overview

The features of the blade server that are externally accessible are described in this section.

LEDs

Server LEDs indicate whether the blade server is in active or standby mode, the status of the network link, the overall health of the blade server, and whether the server is set to give a blinking blue locator light from the locator button.

The removable drives also have LEDs indicating hard disk access activity and disk health.

Table 1: Blade Server LEDs

LED	Color	Description
Power	Off	Power off.
	Green	Main power state. Power is supplied to all server components and the server is operating normally.
	Amber	Standby power state. Power is supplied only to the service processor of the server so that the server can still be managed. **Note** The front-panel power button is disabled by default. It can be re-enabled through Cisco UCS Manager. After it's enabled, if you press and release the front-panel power button, the server performs an orderly shutdown of the 12 V main power and goes to standby power state. You cannot shut down standby power from the front-panel power button. See the Cisco UCS Manager Configuration Guides for information about completely powering off the server from the software interface.
Link	Off	None of the network links are up.
	Green	At least one network link is up.

LED	Color	Description
Health	Off	Power off.
	Green	Normal operation.
	Amber	Minor error.
	Blinking Amber	Critical error.
Blue locator button and LED	Off	Blinking is not enabled.
	Blinking blue 1 Hz	Blinking to locate a selected blade—If the LED is not blinking, the blade is not selected. You can control the blinking in UCS Manager or by using the blue locator button/LED.
Activity (Disk Drive) ▯	Off	Inactive.
	Green	Outstanding I/O to disk drive.
Health (Disk Drive)	Off	Can mean either no fault detected or the drive is not installed.
	Flashing Amber 4 hz	Rebuild drive active. If the Activity LED is also flashing amber, a drive rebuild is in progress.
	Amber	Fault detected.

Buttons

The Reset button is recessed in the front panel of the server. You can press the button with the tip of a paper clip or a similar item. Hold the button down for five seconds, and then release it to restart the server if other methods of restarting do not work.

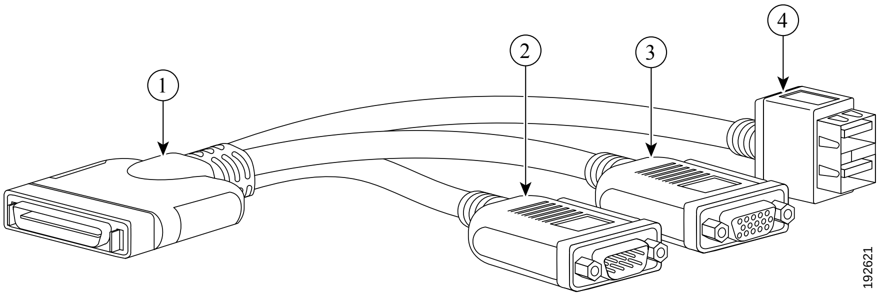
The locator function for an individual server may get turned on or off by pressing the locator button/LED.

The front-panel power button is disabled by default. It can re-enabled through Cisco UCS Manager. After it's enabled, The power button allows you to manually take a server temporarily out of service but leave it in a state where it can be restarted quickly. If the desired power state for a service profile associated with a blade server is set to "off," using the power button or Cisco UCS Manager to reset the server will cause the desired power state of the server to become out of sync with the actual power state and the server may unexpectedly shut down at a later time. To safely reboot a server from a power-down state, use the Boot Server action in Cisco UCS Manager.

Local Console Connection

The local console connector allows a direct connection to a blade server to allow operating system installation and other management tasks to be done directly rather than remotely. The port uses the KVM dongle cable that provides a connection into a Cisco UCS blade server; it has a DB9 serial connector, a VGA connector for a monitor, and dual USB ports for a keyboard and mouse. With this cable, you can create a direct connection to the operating system and the BIOS running on a blade server. A KVM cable ships standard with each blade chassis accessory kit.

Figure 2: KVM Cable for Blade Servers



1	Connector to blade server local console connection	**2**	DB9 serial connector
3	VGA connector for a monitor	**4**	2-port USB connector for a mouse and keyboard

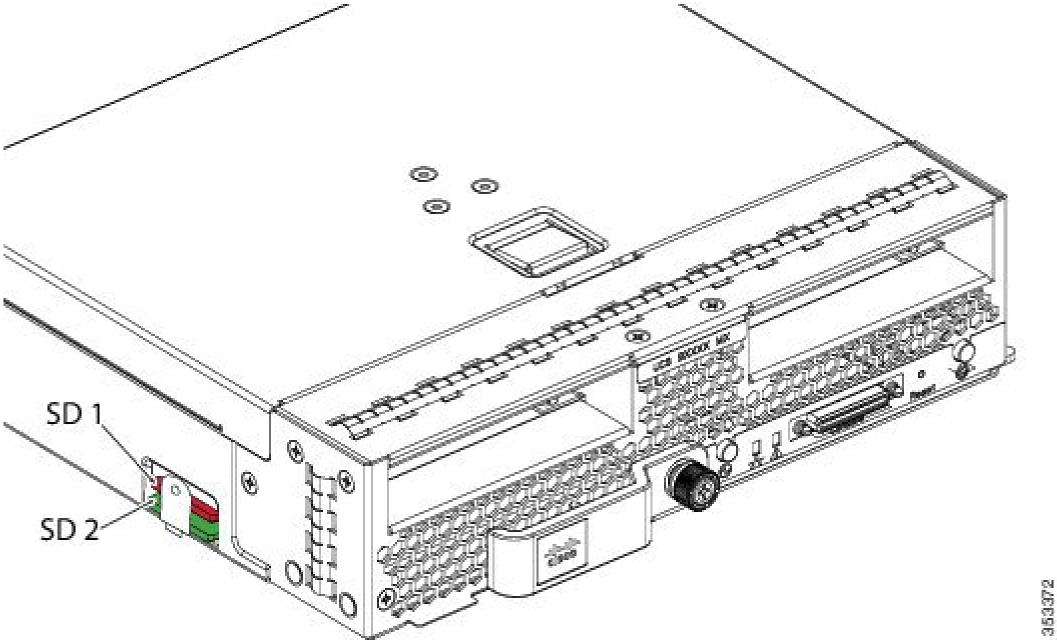
Secure Digital Cards

Secure Digital (SD) card slots are provided and one or more SD cards can be populated. If two SD cards are populated, they can be used in a mirrored mode.

Note Do no mix different capacity cards in the same server.

The SD cards can be uesd to store operating system boot images or other information. Once the server has been removed from the chassis, you can access the SD card slots by rotating the latch up so that it does not cover the slots. Remove or insert the SD cards as needed. Either or both slots may be used. Rotate the latch down to cover the slots before installing the server in the chassis.

Figure 3: SD Card Slots



Storage Module

The Cisco UCS B200 M4 blade server has an optional storage module that can be configured with SAS or SATA hard drives or solid state disks (SSDs). Because the UCS B200 M4 can be used without disk drives, it does not necessarily come with the storage module installed. A blanking panel (UCSB-LSTOR-BK) can be used to cover an empty drive bay. Order the same number of blanking panels as there are empty drive bays.

For information on installing the storage module, see Installing the FlexStorage Module, on page 17.

CHAPTER 2

Installing a Blade Server

This chapter contains the following sections:

- Installing a Half-width Blade Server, on page 7
- Server Configuration, on page 8
- Powering Off a Blade Server Using the Power Button, on page 9
- Removing a Blade Server, on page 10
- Server Troubleshooting, on page 10

Installing a Half-width Blade Server

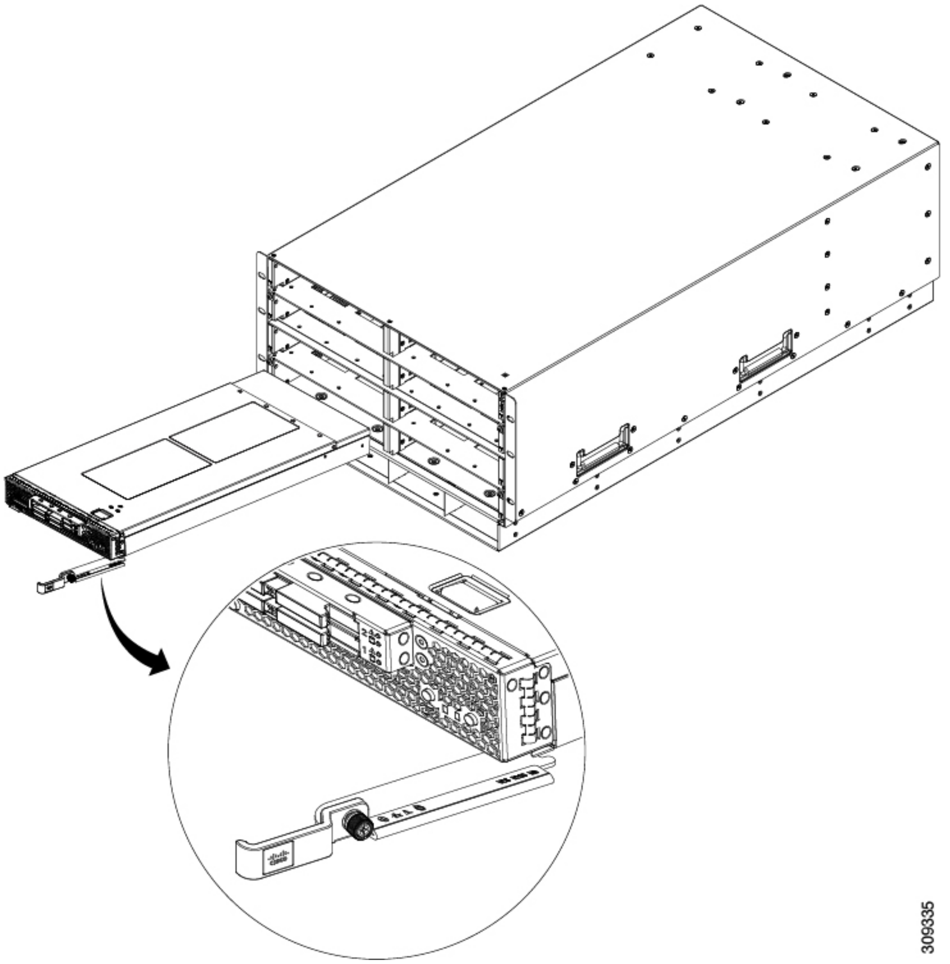
Before you begin

The blade server must have its cover installed before installing the server into the chassis to ensure adequate airflow.

Procedure

Step 1 Grasp the front of the blade server and place your other hand under the blade to support it.

Figure 4: Positioning a Blade Server in the Chassis



Step 2 Open the ejector lever in the front of the blade server.

Step 3 Gently slide the blade into the opening until you cannot push it any farther.

Step 4 Press the ejector so that it catches the edge of the chassis and presses the blade server all the way in.

Step 5 Tighten the captive screw on the front of the blade to no more than 3 in-lbs. Tightening only with bare fingers is unlikely to lead to stripped or damaged captive screws.

Assuming the server chassis is already discovered by UCS Manager, the blade will be auto discovered whenever it is inserted.

Server Configuration

Cisco UCS blade servers can be configured and managed using the following UCS management software interfaces.

Cisco Intersight Managed Mode

Cisco UCS blade servers can be configured and managed using the Cisco Intersight management platform in Intersight Managed Mode (Cisco Intersight Managed Mode). For details, see the *Cisco Intersight Managed Mode Configuration Guide*, which is available at the following URL: https://www.cisco.com/c/en/us/td/docs/unified_computing/Intersight/b_Intersight_Managed_Mode_Configuration_Guide.html

Cisco UCS Manager

Cisco UCS blade servers can be configured and managed using Cisco UCS Manager. For details, see the *Configuration Guide* for the version of Cisco UCS Manager that you are using. The configuration guides are available at the following URL:
http://www.cisco.com/en/US/products/ps10281/products_installation_and_configuration_guides_list.html

Powering Off a Blade Server Using the Power Button

Note The front panel power button is disabled by default to ensure that servers are decommissioned through the UCS management software interface before shutdown. If you prefer to shut down the server locally with the button, you can enable front power-button control in the UCS management software interface.

Tip You can also shut down servers remotely using the UCS management software interface. For details, see the configuration guide for the version the UCS management software interface that you are using. The configuration guides are available at the URLs documented in Server Configuration, on page 8.

Procedure

Step 1 If you are local to the server, check the color of the **Power Status** LED for each server in the chassis that you want to power off.

- Green indicates that the server is running and must be shut down before it can be safely powered off. Go to Step 2.
- Amber indicates that the server is already in standby mode and can be safely powered off. Go to Step 3.

Step 2 If you previously enabled front power-button control through the UCS management software interface, press and release the **Power** button, then wait until the **Power Status** LED changes to amber.

The operating system performs a graceful shutdown, and the server goes to standby mode.

Caution To avoid data loss or damage to your operating system, you should always invoke a graceful shutdown of the operating system.

Step 3 (Optional) Although not recommended, if you are shutting down all blade servers in a chassis, you can disconnect the power cords from the chassis to completely power off the servers.

Caution To avoid data loss or damage to your operating system, you should always invoke a graceful shutdown of the operating system.

The blade servers will power down. You can now perform additional tasks with the blades as needed, for example, replacing a blade.

Removing a Blade Server

Using the UCS management software interface, decommission the server before physically removing the server. To remove a blade server from the chassis, follow these steps:

Procedure

Step 1 Loosen the captive screw on the front of the blade.

Step 2 Remove the blade from the chassis by pulling the ejector lever on the blade until it unseats the blade server.

Step 3 Slide the blade partially out of the chassis and place your other hand under the blade to support its weight.

Step 4 Once completely removed, place the blade on an antistatic mat or antistatic foam if you are not immediately reinstalling it into another slot.

Step 5 If the slot is to remain empty, install a blank faceplate (N20-CBLKB1) to maintain proper thermal temperature and keep dust out of the chassis.

Server Troubleshooting

For general troubleshooting information, see the Cisco UCS Manager Troubleshooting Reference Guide.

CHAPTER 3

Servicing a Blade Server

This chapter contains the following sections:

- Replacing a Drive, on page 11
- Removing a Blade Server Cover, on page 14
- Air Baffles, on page 14
- Internal Components, on page 15
- Diagnostics Button and LEDs, on page 15
- Installing a CMOS Battery, on page 16
- Installing the FlexStorage Module, on page 17
- Replacing the SuperCap Module, on page 17
- Upgrading to Intel Xeon E5-2600 v4 CPUs, on page 20
- Removing a Heat Sink and CPU, on page 21
- Installing a New CPU and Heat Sink, on page 22
- Installing Memory, on page 25
- Installing a Virtual Interface Card Adapter, on page 27
- Installing an Adapter Card in Addition to the VIC mLOM Adapter, on page 28
- Installing the NVIDIA M6 GPU Adapter Card, on page 30
- Enabling the Trusted Platform Module, on page 33

Replacing a Drive

The Cisco UCS B200 M4 blade server uses an optional Cisco UCS FlexStorage modular storage subsystem that can provide support for two drive bays and RAID controller, or NVMe-based PCIe SSD support functionality. If you purchased the UCS B200 M4 blade server without the modular storage system configured as a part of the system, a pair of blanking panels may be in place. These panels should be removed before installing hard drives, but should remain in place to ensure proper cooling and ventilation if the drive bays are unused.

You can remove and install hard drives without removing the blade server from the chassis.

The drives supported in this blade server come with the hot-plug drive sled attached. Empty hot-plug drive sled carriers (containing no drives) are not sold separately from the drives. A list of currently supported drives is in the Cisco UCS B200 M4 Blade Server Specification Sheet.

Before upgrading or adding a drive to a running blade server, check in the service profile and make sure the new hardware configuration will be within the parameters allowed by the service profile.

Note See also 4K Sector Format SAS/SATA Drives Considerations, on page 12.

Removing a Blade Server Hard Drive

To remove a hard drive from a blade server, follow these steps:

Procedure

Step 1 Push the button to release the ejector, and then pull the hard drive from its slot.

Step 2 Place the hard drive on an antistatic mat or antistatic foam if you are not immediately reinstalling it in another server.

Step 3 Install a hard disk drive blank faceplate to keep dust out of the blade server if the slot will remain empty.

Installing a Blade Server Drive

To install a drive in a blade server, follow these steps:

Procedure

Step 1 Place the drive ejector into the open position by pushing the release button.

Step 2 Gently slide the drive into the opening in the blade server until it seats into place.

Step 3 Push the drive ejector into the closed position.

You can use Cisco UCS Manager to format and configure RAID services. For details, see the *Configuration Guide* for the version of Cisco UCS Manager that you are using. The configuration guides are available at the following URL:
http://www.cisco.com/en/US/products/ps10281/products_installation_and_configuration_guides_list.html

If you need to move a RAID cluster, see the Cisco UCS Manager Troubleshooting Reference Guide.

4K Sector Format SAS/SATA Drives Considerations

- You must boot 4K sector format drives in UEFI mode, not legacy mode. See the procedure in this section for setting UEFI boot mode in the boot policy.
- Do not configure 4K sector format and 512-byte sector format drives as part of the same RAID volume.
- Operating system support on 4K sector drives is as follows: Windows: Win2012 and Win2012R2; Linux: RHEL 6.5, 6.6, 6.7, 7.0, 7.2, 7.3; SLES 11 SP3, and SLES 12. ESXi/VMWare is not supported.

Setting Up UEFI Mode Booting in the UCS Manager Boot Policy

Procedure

Step 1 In the Navigation pane, click **Servers**.

Step 2 Expand **Servers** > **Policies**.

Step 3 Expand the node for the organization where you want to create the policy.

If the system does not include multitenancy, expand the root node.

Step 4 Right-click **Boot Policies** and select **Create Boot Policy**.

The Create Boot Policy wizard displays.

Step 5 Enter a unique name and description for the policy.

This name can be between 1 and 16 alphanumeric characters. You cannot use spaces or any special characters other than - (hyphen), _ (underscore), : (colon), and . (period). You cannot change this name after the object is saved.

Step 6 (Optional) After you make changes to the boot order, check the **Reboot on Boot Order Change** check box to reboot all servers that use this boot policy.

For boot policies applied to a server with a non-Cisco VIC adapter, even if the Reboot on Boot Order Change check box is not checked, when SAN devices are added, deleted, or their order is changed, the server always reboots when boot policy changes are saved.

Step 7 (Optional) If desired, check the **Enforce vNIC/vHBA/iSCSI Name** check box.

- If checked, Cisco UCS Manager displays a configuration error and reports whether one or more of the vNICs, vHBAs, or iSCSI vNICs listed in the Boot Order table match the server configuration in the service profile.
- If not checked, Cisco UCS Manager uses the vNICs or vHBAs (as appropriate for the boot option) from the service profile.

Step 8 In the **Boot Mode** field, choose the **UEFI** radio button.

Step 9 Check the Boot Security check box if you want to enable UEFI boot security.

Step 10 Configure one or more of the following boot options for the boot policy and set their boot order:

- Local Devices boot—To boot from local devices, such as local disks on the server, virtual media, or remote virtual disks, continue with *Configuring a Local Disk Boot for a Boot Policy* in the Cisco UCS Manager Server Management Guide for your release.
- SAN boot—To boot from an operating system image on the SAN, continue with *Configuring a SAN Boot for a Boot Policy* in the Cisco UCS Manager Server Management Guide for your release.

You can specify a primary and a secondary SAN boot. If the primary boot fails, the server attempts to boot from the secondary.

- LAN boot—To boot from a centralized provisioning server, continue with *Configuring a LAN Boot For a Boot Policy* in the Cisco UCS Manager Server Management Guide for your release.

- iSCSI boot—To boot from an iSCSI LUN, continue with *Creating an iSCSI Boot Policy* in the Cisco UCS Manager Server Management Guide for your release.

Removing a Blade Server Cover

Procedure

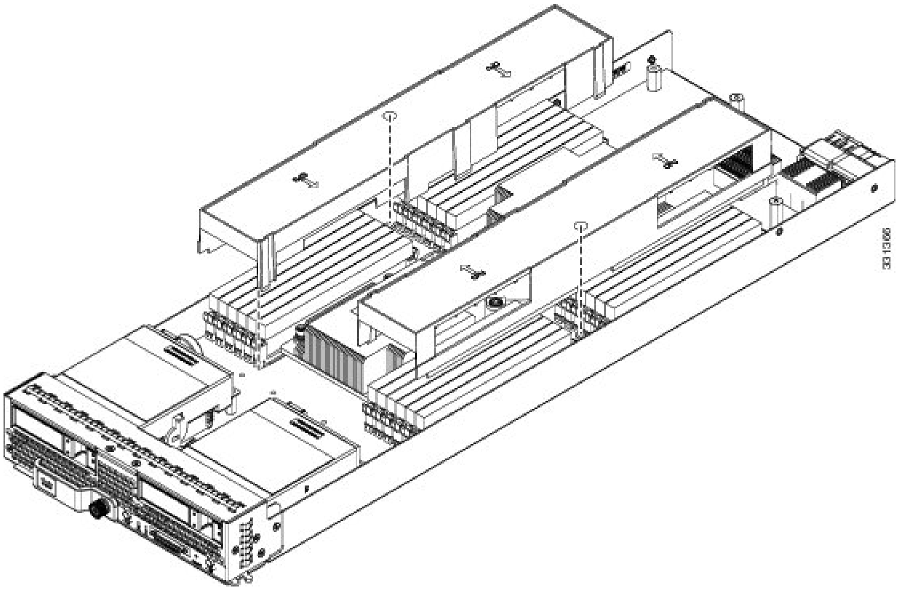
Step 1 Press and hold the button down as shown in the figure below.

Step 2 While holding the back end of the cover, pull the cover back and then up.

Air Baffles

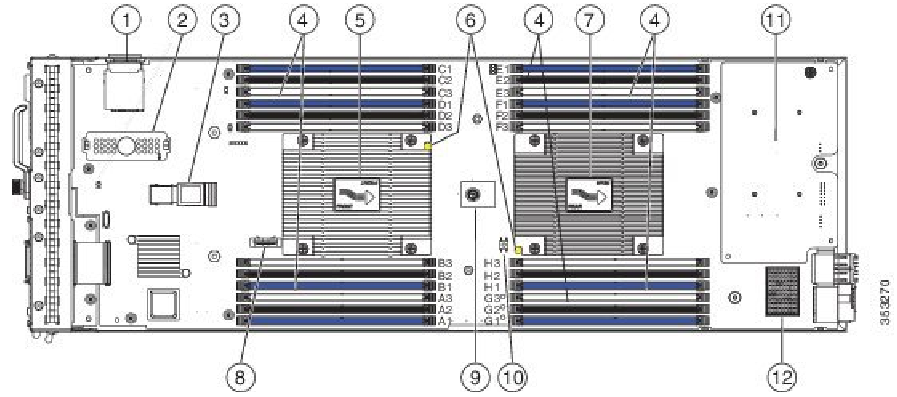
The air baffles direct and improve air flow for the server components. Two identical baffles ship with each B200 M4 server. No tools are necessary to install them, just place them over the DIMMs as shown, with the holes in the center of the baffles aligned with the corresponding motherboard standoffs.

Figure 5: Cisco UCS B200 M4 Air Baffles



Internal Components

Figure 6: Inside View of the UCS B200 M4 Blade Server



1	SD card slots	**2**	Modular storage subsystem connector
3	USB connector An internal USB 2.0 port is supported. A 16 GB USB drive (UCS-USBFLSHB-16GB) is available from Cisco. A clearance of 0.950 inches (24.1 mm) is required for the USB device to be inserted and removed.	**4**	DIMM slots
5	Front heat sink and CPU 1	**6**	CPU heat sink install guide pins
7	Rear heat sink and CPU 2	**8**	CMOS battery
9	Trusted Platform Module (TPM)	**10**	DIMM diagnostic LED button
11	Adapter slot 1	**12**	Adapter slot 2

Note When the storage module is installed, the USB connector is underneath it. Use the small cutout opening in the storage module to visually determine the location of the USB connector when you need to insert it.

Diagnostics Button and LEDs

At blade start-up, POST diagnostics test the CPUs, DIMMs, HDDs, and rear mezzanine modules, and any failure notifications are sent to Cisco UCS Manager. You can view these notifications in the Cisco UCS Manager System Error Log or in the output of the **show tech-support** command. If errors are found, an amber diagnostic LED also lights up next to the failed component. During run time, the blade BIOS and component drivers monitor for hardware faults and will light up the amber diagnostic LED as needed.

LED states are saved, and if you remove the blade from the chassis the LED values will persist for up to 10 minutes. Pressing the LED diagnostics button on the motherboard causes the LEDs that currently show a component fault to light for up to 30 seconds for easier component identification. LED fault values are reset when the blade is reinserted into the chassis and booted, and the process begins from its start.

If DIMM insertion errors are detected, they may cause the blade discovery process to fail and errors will be reported in the server POST information, which is viewable using the UCS Manager GUI or CLI. DIMMs must be populated according to specific rules. The rules depend on the blade server model. Refer to the documentation for a specific blade server for those rules.

Faults on the DIMMs or rear mezzanine modules also cause the server health LED to light solid amber for minor error conditions or blinking amber for critical error conditions.

Installing a CMOS Battery

All Cisco UCS blade servers use a CR2032 battery to preserve BIOS settings while the server is not installed in a powered-on chassis. Cisco supports the industry standard CR2032 battery that is available at most electronics stores.

Warning There is danger of explosion if the battery is replaced incorrectly. Replace the battery only with the same or equivalent type recommended by the manufacturer. Dispose of used batteries according to the manufacturer’s instructions.

To install or replace the battery, follow these steps:

Procedure

Step 1 Remove the existing battery:

a) Power off the blade, remove it from the chassis, and remove the top cover.
b) Push the battery socket retaining clip away from the battery.
c) Lift the battery from the socket. Use needle-nose pliers to grasp the battery if there is not enough clearance for your fingers.

Step 2 Install the replacement battery:

a) Push the battery socket retaining clip away from where the battery fits in the housing.
b) Insert the new battery into the socket with the battery’s positive (+) marking facing away from the retaining clip. Ensure that the retaining clip can click over the top of the battery to secure it in the housing.
c) Replace the top cover.
d) Replace the blade server in the chassis.

Installing the FlexStorage Module

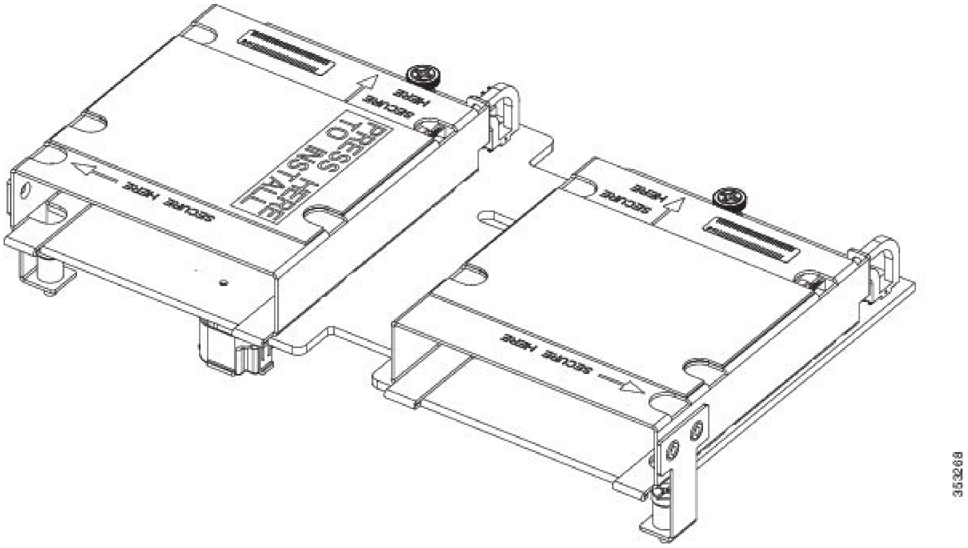
The Cisco UCS B200 M4 blade server uses an optional Cisco UCS FlexStoarge modular storage subsystem that can provide support for two drive bays and RAID controller or NVMe-based PCIe SSD support functionality.

Procedure

Step 1 Place the FlexStorage module over the two standoff posts on the motherboard at the front of the server.

Step 2 Press down on the drive bay cage where it is labeled "Press Here to Install" until the FlexStorage module clicks into place.

Figure 7: FlexStorage Module



Step 3 Using a Phillips-head screwdriver, tighten the four screws to secure the FlexStorage module. The locations of the screws are labeled "Secure Here."

Replacing the SuperCap Module

The SuperCap module is a battery bank which connects to the front mezzanine storage module board and provides power to the RAID controller if facility power is interrupted.

To replace the SuperCap module, use the following topics:

- Removing the SuperCap Module, on page 18
- Installing the SuperCap Module, on page 19

Removing the SuperCap Module

The SuperCap module sits in a plastic tray. The module connects to the board through a ribbon cable with one connector to the module and one connector to the board. The SuperCap replacement PID (UCSB-MRAID-SC=) contains the module only, so you must leave the ribbon cable in place on the board.

Caution When disconnecting the SuperCap module, disconnect the ribbon cable from the module only. Do not disconnect the cable from the board. The board connection and the tape that secures the cable must remain connected and undamaged.

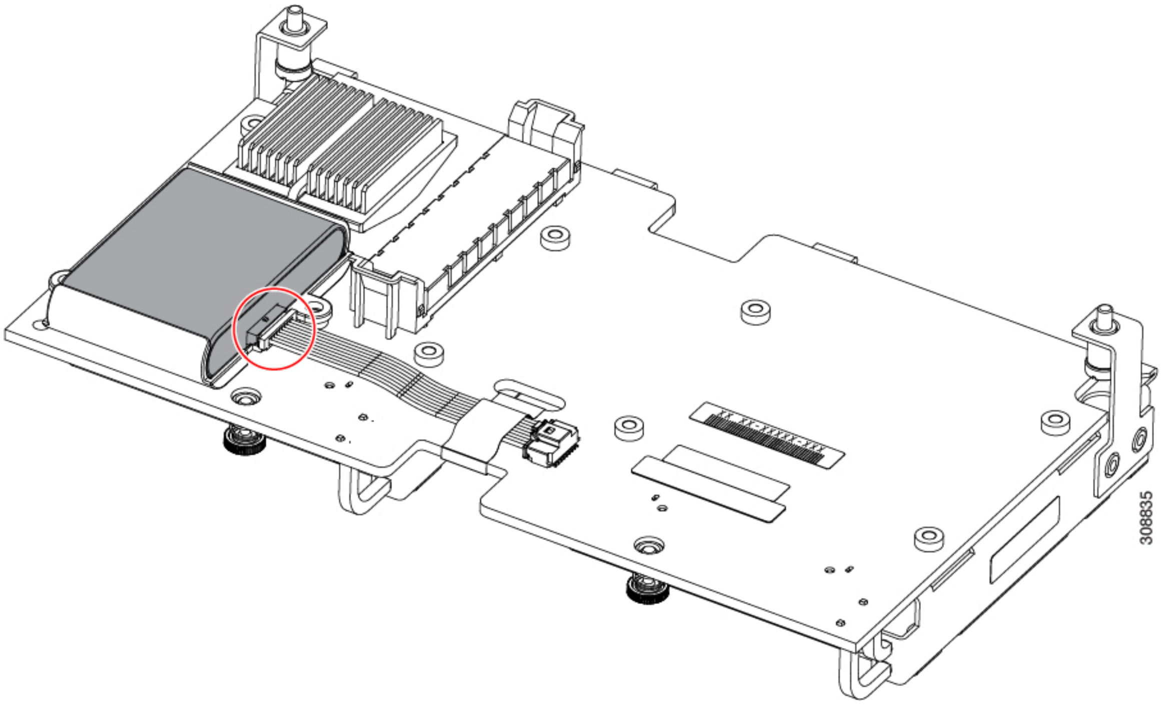
To replace the SuperCap module, follow these steps:

Procedure

Step 1 Grasp the cable connector at the SuperCap module and gently pull to disconnect the cable from the SuperCap module.

Do not grasp the cable itself, the tape, or the board connector.

Figure 8: Disconnecting the SuperCap Cable from the Module, Not the Board

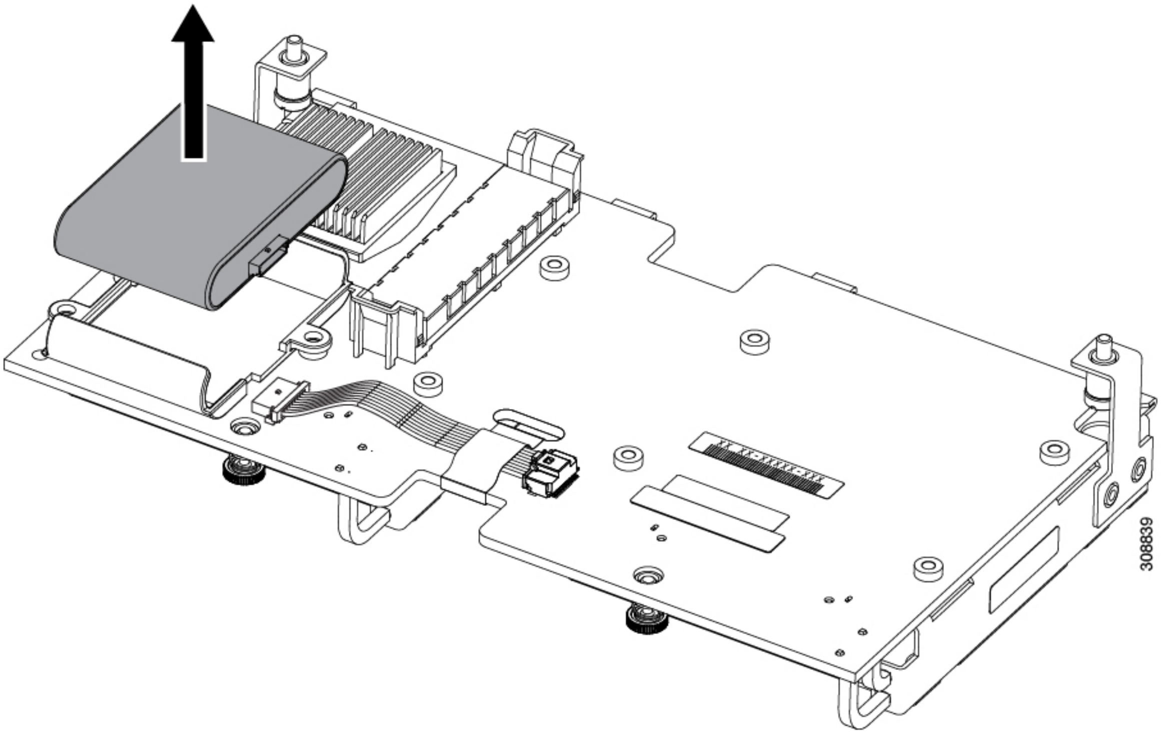


Step 2 Before removing the SuperCap module, note its orientation in the tray.

When correctly oriented, the connector is on the bottom half of the module and faces the cable. You will need to install the new SuperCap module with the same orientation.

Step 3 Grasp the sides of the SuperCap module, but not the connector, and lift the SuperCap module out of the tray.

Figure 9: Removing the SuperCap Module



You might feel some resistance because the tray is curved to secure the module.

Installing the SuperCap Module

To install a SuperCap module (UCSB-MRAID-SC=), use the following steps:

Procedure

Step 1 Orient the SuperCap module correctly, as shown (1).

When correctly oriented:

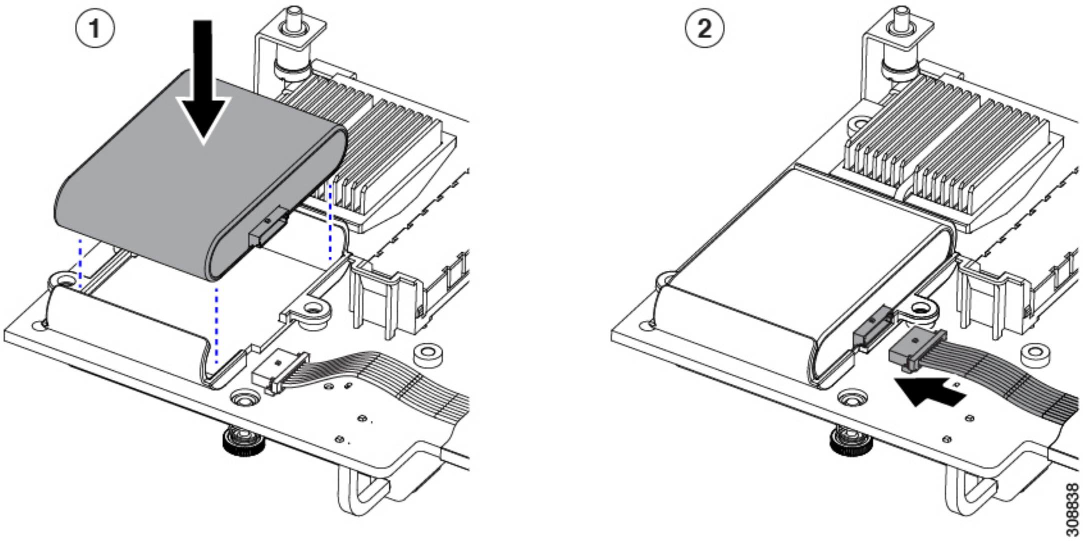
- The connector is on the bottom half of the module facing the cable.
- The connector will fit into the rectangular notch in the tray. This notch is specifically designed to accept the SuperCap module connector.

Caution Make sure the SuperCap module is properly oriented before proceeding. If the module is installed incorrectly, the ribbon cable can get snagged or damaged.

Step 2 When the module is correctly oriented, lower the module and press down until it clips into the tray.

You might feel some resistance while the module passes the curved clips at the top of the tray.

Figure 10: Orienting and Installing the SuperCap Module



Step 3 When the module is seated in the tray, reconnect the cable (2):

a) Grasp the cable connector and verify that the pins and sockets on the cable connector and module connector are correctly aligned.

b) When the cable connector and module connector are properly aligned, plug the cable into the SuperCap module.

What to do next

Reinstall the blade server. Go to Installing a Blade Server, on page 7.

Upgrading to Intel Xeon E5-2600 v4 CPUs

Before upgrading to Intel Xeon E5-2600 v4 Series CPUs, ensure that the server is running the required minimum software and firmware versions that support Intel E5-2600 v4 Series CPUs, as listed in the following table.

Software or Firmware	Minimum Version
Cisco UCS Manager	Release 3.1(1e) with 3.1(1g) ucs-catalog.3.1.1g.T.bin, or Release 2.2(7b) (See the following Note for additional supported versions.)
Cisco IMC	Release 3.1(1g) or Release 2.2(7b)
BIOS	Release 3.1(1g) or Release 2.2(7b)

Note Cisco UCS Manager Release 2.2(4) introduced a server pack feature that allows Intel E5-2600 v4 CPUs to run with Cisco UCS Manager Release 2.2(4) or later, provided that the Cisco IMC, BIOS and Capability Catalog are all running Release 2.2(7) or later.

Caution Ensure that the server is running the required software and firmware before installing the Intel E5-2600 v4 Series CPUs. Failure to do so can result in a non-bootable CPU.

Do one of the following actions:

- If the server software and firmware are already at the required minimum version as shown in the preceding table, replace the CPUs by using the procedure in the following section.
- If the server software or firmware is not at the required minimum version, follow the instructions in the Cisco UCS B200 M4 Server Upgrade Guide for E5-2600 v4 Series CPUs to upgrade it. Then replace the CPUs by using the procedure in the following section.

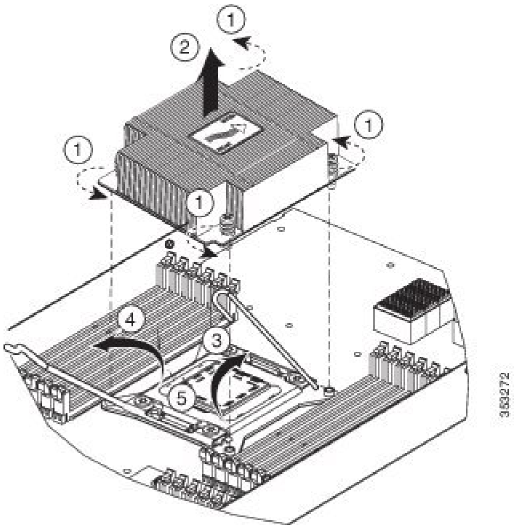
Removing a Heat Sink and CPU

Procedure

Step 1 Unscrew the four captive screws.

Step 2 Remove the heat sink.

Figure 11: Removing the Heat Sink and CPU

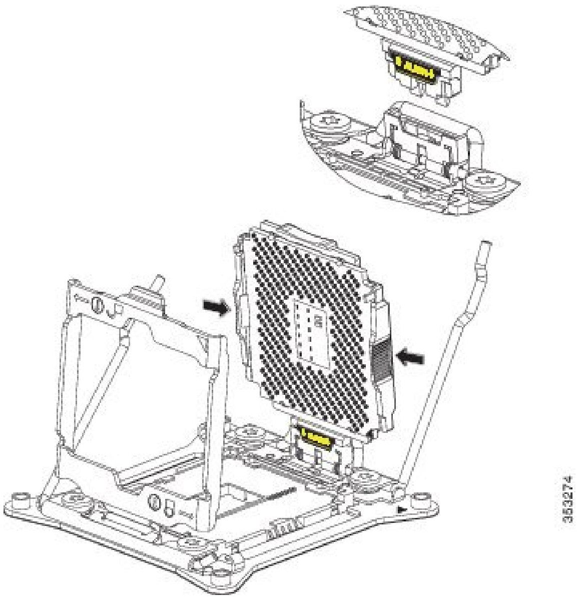


Step 3 Unhook the self-loading socket (SLS) lever that has the unlock icon 🔓.

Step 4 Unhook the SLS lever that has the lock icon 🔒.

Step 5 Grasp the sides of the CPU carrier (indicated by the arrows in the illustration) and swing it into a standing position in the SLS plug seat.

Figure 12: CPU Carrier and SLS Plug Seat



Step 6 Pull the CPU carrier up and out of the SLS plug seat.

Installing a New CPU and Heat Sink

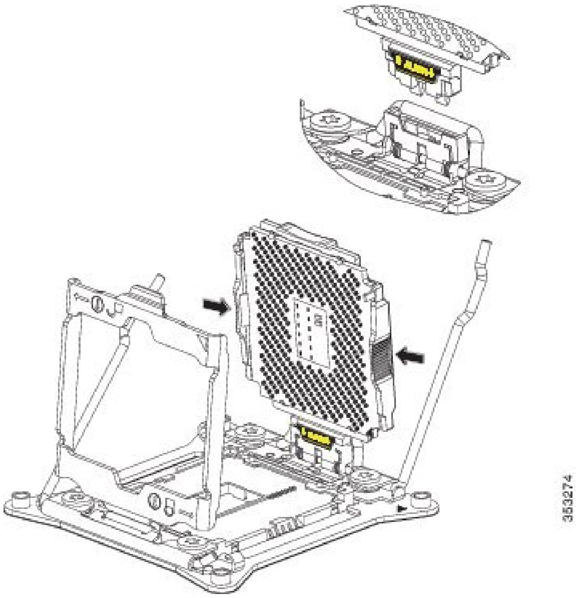
Before installing a new CPU in a server, verify the following:

- A BIOS update is available and installed that supports the CPU and the given server configuration.
- The service profile for this server in Cisco UCS Manager will recognize and allow the new CPU.

Procedure

Step 1 Hold the CPU carrier by its sides (indicated by the arrows). Insert and align the two CPU carrier pegs into the self-loading socket (SLS) plug seat. To ensure proper seating, verify that the horizontal yellow line below the word ALIGN is straight.

Figure 13: Inserting the CPU Carrier



Step 2 Press gently on the top of the CPU carrier from the exterior side until it snaps into place.

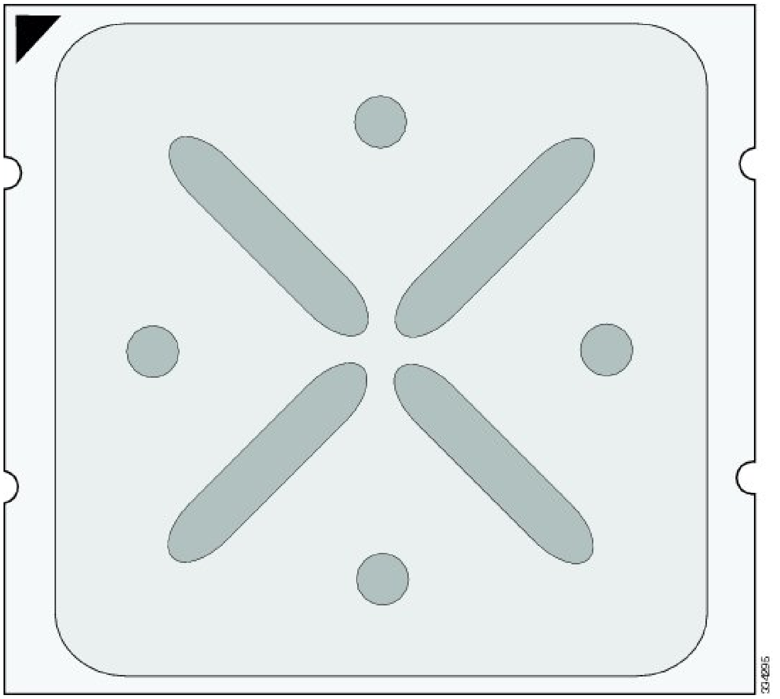
Step 3 Close the socket latch.

Step 4 Hook the self-loading socket (SLS) lever that has the lock icon 🔒.

Step 5 Hook the SLS lever that has the unlock icon 🔓.

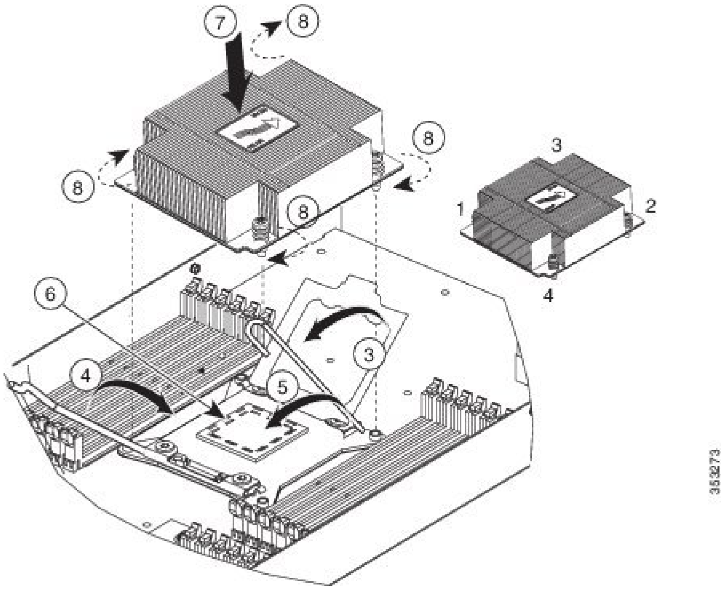
Step 6 Thermally bond the CPU and heat sink. Using the syringe of thermal grease provided with the replacement CPU, apply 2 cubic centimeters of thermal grease to the top of the CPU where it will contact the heat sink. Apply the grease in the pattern shown in the following figure, which should use approximately half the contents of the syringe.

Figure 14: Thermal Grease Application Pattern



Step 7 Replace the heat sink. The yellow CPU heat sink install guide pins that are attached to the motherboard must align with the cutout on the heat sink to ensure proper installation of the heat sink.

Figure 15: Replacing the Heat Sink



Step 8 Tighten the four captive screws in the order shown.

Installing Memory

To install a DIMM into the blade server, follow these steps:

Procedure

Step 1 Press the DIMM into its slot evenly on both ends until it clicks into place.

DIMMs are keyed. If a gentle force is not sufficient, make sure the notch on the DIMM is correctly aligned.

Note Be sure that the notch in the DIMM aligns with the slot. If the notch is misaligned you may damage the DIMM, the slot, or both.

Step 2 Press the DIMM connector latches inward slightly to seat them fully.

Supported DIMMs

Do not use any memory DIMMs other than those listed in the specification sheet. Doing so may irreparably damage the server and require down time.

Memory Population

The blade server contains 24 DIMM slots—12 for each CPU. Each set of 12 DIMM slots is arranged into four channels, where each channel has three DIMMs.

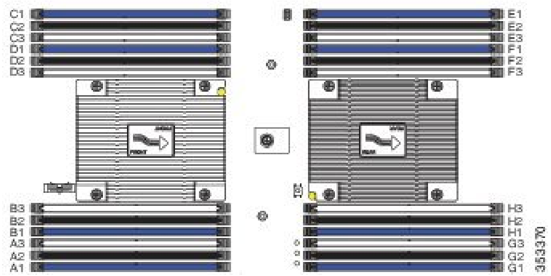
1	Channels A-D for CPU 1	**2**	Channels E-H for CPU 2

DIMMs and Channels

Each channel is identified by a letter—A, B, C, D for CPU 1, and E, F, G, H for CPU 2. Each DIMM slot is numbered 1, 2, or 3. Note that each DIMM slot 1 is blue, each slot 2 is black, and each slot 3 is off-white or beige.

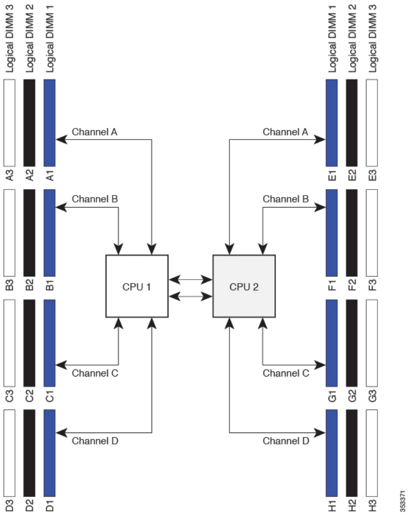
The figure below shows how DIMMs and channels are physically laid out on the blade server. The DIMM slots in the upper and lower right are associated with the second CPU (CPU shown on right in the diagram), while the DIMM slots in the upper and lower left are associated with the first CPU (CPU shown on left).

Figure 16: Physical Representation of DIMMs and Channels



The figure below shows a logical view of the DIMMs and channels.

Figure 17: Logical Representation of DIMMs and Channels



DIMMs can be used in the blade server in a one DIMM per Channel (1DPC) configuration, in a two DIMMs per Channel (2DPC) configuration, or a three DIMMs per Channel (3DPC) configuration.

The following tables show recommended DIMM population order for non-mirroring and mirroring configurations. For single-CPU configurations, read only the CPU 1 columns of the tables.

Table 2: Supported DIMM Population Order (Non-Mirroring)

DIMMs Per CPU	CPU 1 Installed Slots	CPU 2 Installed Slots
1	A1	E1
2	A1, B1	E1, F1
3	A1, B1, C1	E1, F1, G1
4	A1, B1, C1, D1	E1, F1, G1, H1
8	A1, B1, C1, D1, A2, B2, C2, D2	E1, F1, G1, H1, E2, F2, G2, H2
12	A1, B1, C1, D1, A2, B2, C2, D2, A3, B3, C3, D3	E1, F1, G1, H1, E2, F2, G2, H2, E3, F3, G3, H3

Note System performance is optimized when the DIMM type and quantity are equal for both CPUs, and when each populated channel is filled equally across the CPUs in the server.

Table 3: Supported DIMM Population Order (Mirroring)

DIMMs per CPU	CPU 1 Installed Slots	CPU 2 Installed Slots
2	A1, B1	E1, F1
4	A1, B1, C1, D1	E1, F1, G1, H1
8	A1, B1, C1, D1, A2, B2, C2, D2	E1, F1, G1, H1, E2, F2, G2, H2
8 (CPU1) and 4 (CPU2) Not recommended for performance reasons.	A1, B1, C1, D1, A2, B2, C2, D2	E1, F1, E2, F2
12	A1, B1, C1, D1, A2, B2, C2, D2, A3, B3, C3, D3	E1, F1, G1, H1, E2, F2, G2, H2, E3, F3, G3, H3

Memory Performance

When considering the memory configuration of the blade server, there are several things to consider. For example:

- When mixing DIMMs of different densities (capacities), the highest density DIMM goes in slot 1 then in descending density.
- Besides DIMM population and choice, the selected CPU(s) can have some effect on performance.
- DIMMs can be run in a 1DPC, a 2DPC, or a 3DPC configuration. 1DPC and 2DPC can provide the maximum rated speed that the CPU and DIMMs are rated for. 3DPC causes the DIMMs to run at a slower speed.

Memory Mirroring and RAS

The Intel CPUs within the blade server support memory mirroring only when an even number of **channels** are populated with DIMMs. Furthermore, if memory mirroring is used, DRAM size is reduced by 50 percent for reasons of reliability.

Installing a Virtual Interface Card Adapter

Note You must remove the adapter card to service it.

To install a Cisco VIC 1340 or VIC 1240 in the blade server, follow these steps:

Procedure

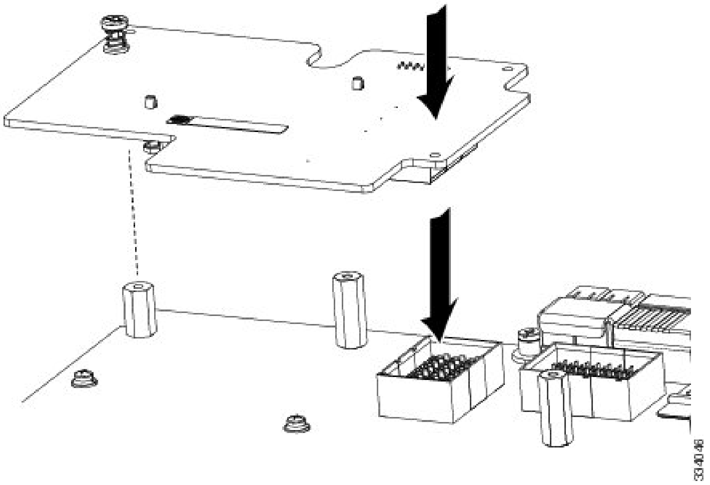
Step 1 Position the VIC board connector above the motherboard connector and align the captive screw to the standoff post on the motherboard.

Step 2 Firmly press the VIC board connector into the motherboard connector.

Step 3 Tighten the captive screw.

Tip To remove a VIC, reverse the above procedure. You might find it helpful when removing the connector from the motherboard to gently rock the board along the length of the connector until it loosens.

Figure 18: Installing a VIC mLOM Adapter



Installing an Adapter Card in Addition to the VIC mLOM Adapter

All supported adapter cards have a common installation process. A list of currently supported and available adapters for this server is in the Cisco UCS B200 M4 Blade Server Specification Sheet.

The UCS B200 M4 blade server has two adapter slots (Slots 1 [mLOM slot] and 2) that support the following VIC cards:

- VIC 1340 and VIC 1380
- VIC 1240 and VIC 1280

Slot 1 is for the VIC 1340 or VIC 1240 mLOM adapter cards. Slot 2 is for the VIC 1380 and VIC 1280 cards, and can also be used for the VIC port expander, the nVidia M6 GPU, the Intel Crypto accelerator card, and non-I/O mezzanine cards, such as Fusion ioMemory 3 Series.

Note When the Cisco Nexus 2104XP Fabric Extender (FEX) module is used, the VIC 1280 and the VIC port expander cards are ignored because there are no traces on the Cisco 2104XP to connect to any VIC or IO card installed in Slot 2.

The VIC 1340 and VIC 1380 require a Cisco UCS 6200 Series Fabric Interconnect or Cisco UCS 6300 Series Fabric Interconnect, and they support the Cisco Nexus 2208XP, 2204XP, 2348UPQ FEX modules.

The VIC 1240 and VIC 1280 support Cisco UCS 6200 and 6100 Series Fabric Interconnects, and they support the Cisco Nexus 2208XP, 2204XP, and 2104XP FEX modules. When a VIC 1240 or 1280 is used with a UCS 6100 Series Fabric Interconnect, the UCS B200 M4 blade server requires a maximum software release of 2.2(x) for Cisco UCS Manager.

If you are switching from one type of adapter card to another, before you physically perform the switch make sure that you download the latest device drivers and load them into the server's operating system. For more information, see the firmware management chapter of one of the Cisco UCS Manager software configuration guides.

Procedure

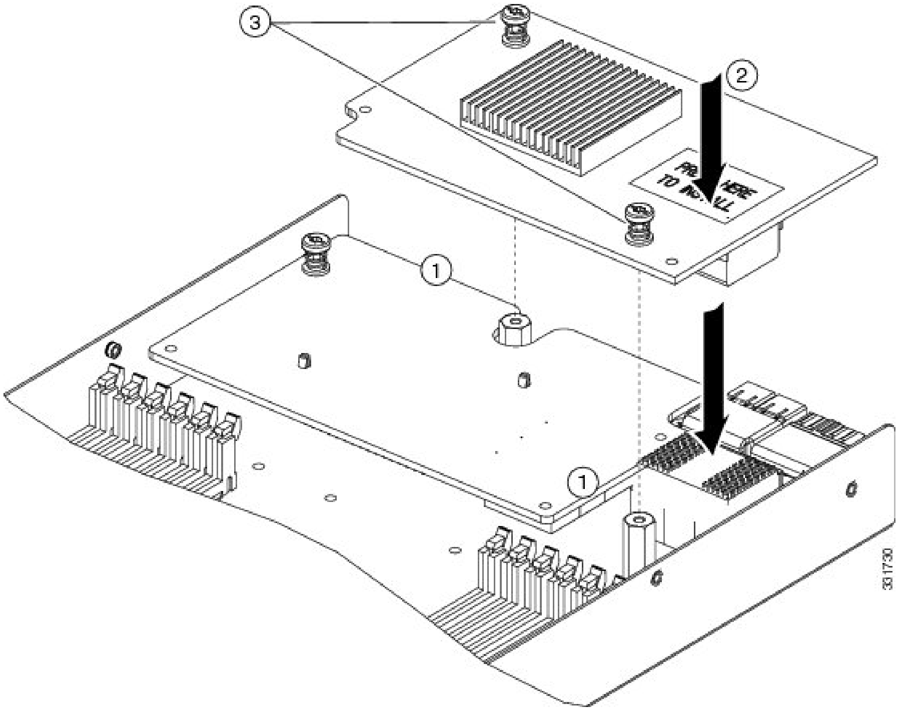
Step 1 Position the adapter board connector above the motherboard connector and align the two adapter captive screws to the standoff posts on the motherboard.

Step 2 Firmly press the adapter connector into the motherboard connector (callout 2).

Step 3 Tighten the two captive screws (callout 3).

Tip Removing an adapter card is the reverse of installing it. You might find it helpful when removing the connector from the motherboard to gently rock the board along the length of the connector until it loosens.

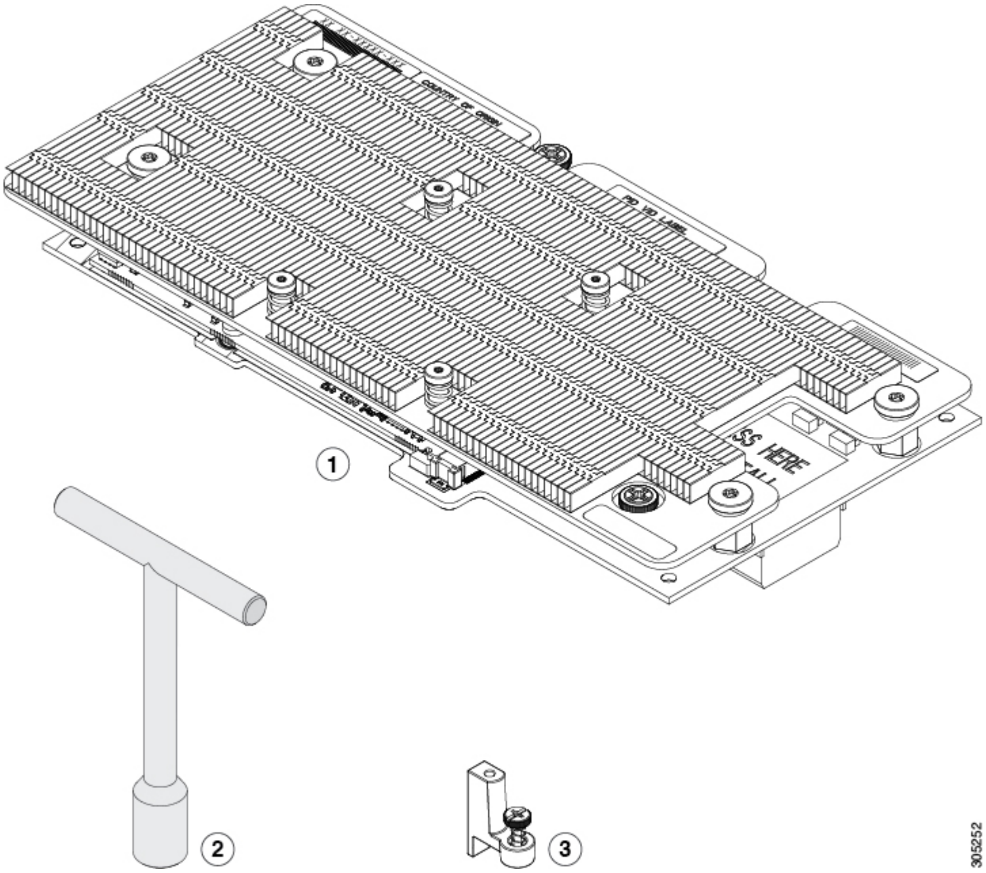
Figure 19: Installing an Adapter Card



Installing the NVIDIA M6 GPU Adapter Card

The NVIDIA M6 graphics processing unit (GPU) adapter card provides graphics and computing capabilities to the server. If you are installing the NVIDIA GPU to a B200 M4 in the field, the option kit comes with the GPU itself (CPU and heat sink), a T-shaped installation wrench, and a custom standoff to support and attach the GPU on the B200 M4 motherboard. See the three components of the option kit in the following figure:

Figure 20: NVIDIA M6 GPU Option Kit



1	NVIDIA M6 GPU (CPU and heat sink)	**2**	T-shaped wrench
3	Custom standoff		

Before you begin

Before installing the NVIDIA M6 GPU:

- Remove any adapter card, such as a VIC 1380, VIC 1280, or VIC port expander card from slot 2. You cannot use any other card in slot 2 when the NVIDIA M6 GPU is installed.
- Upgrade the Cisco UCS domain that the GPU will installed into to a version of Cisco UCS Manager that supports this card. Refer to the latest version of the *Release Notes for Cisco UCS Software* at the following URL for information about supported hardware: http://www.cisco.com/c/en/us/support/servers-unified-computing/ucs-manager/products-release-notes-list.html.

Procedure

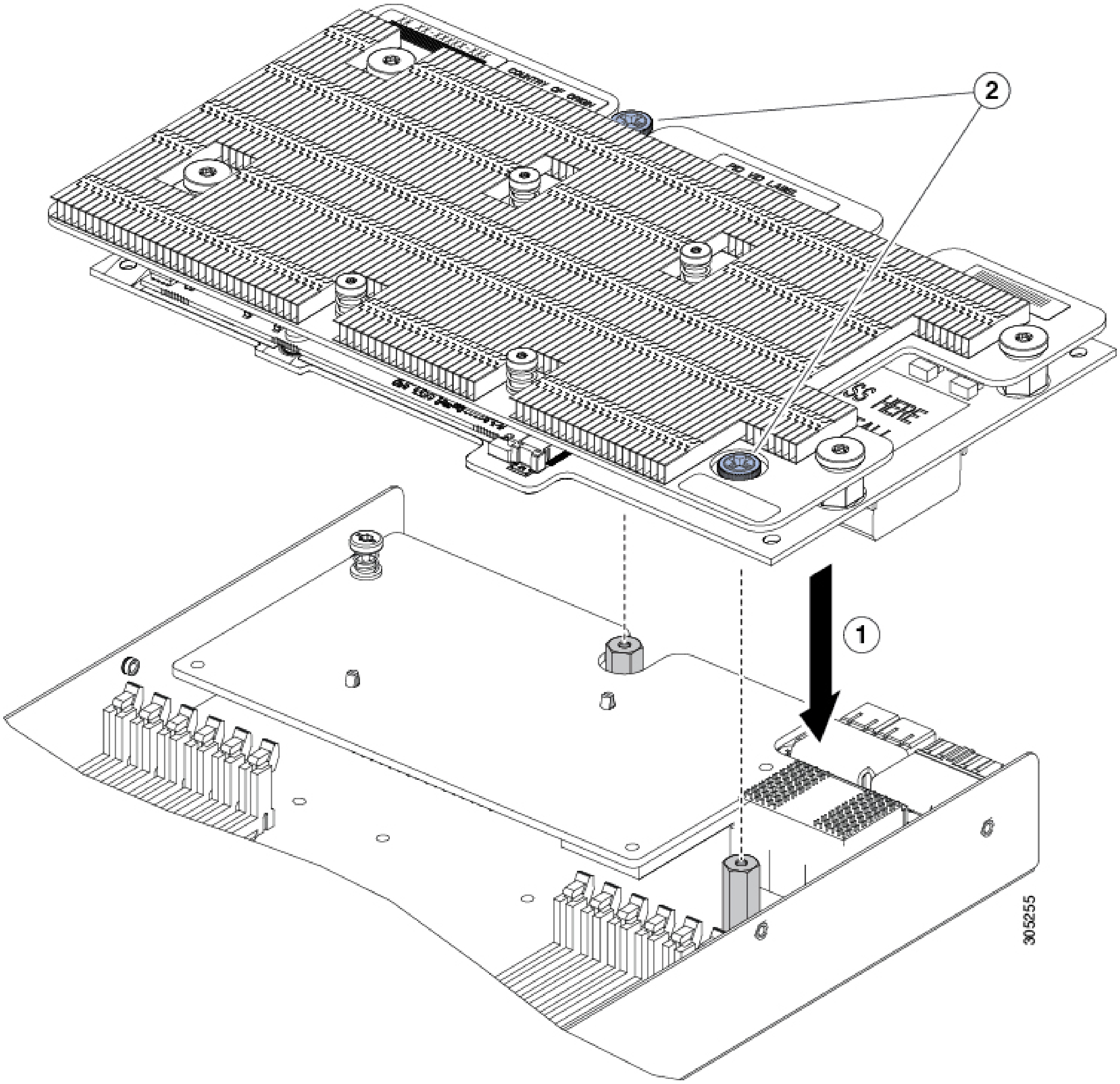
Step 1 Use the T-shaped wrench that comes with the GPU to remove the existing standoff at the back end of the motherboard.

Step 2 Install the custom standoff in the same location at the back end of the motherboard.

Step 3 Position the GPU over the connector on the motherboard and align all captive screws to the standoff posts (callout 1).

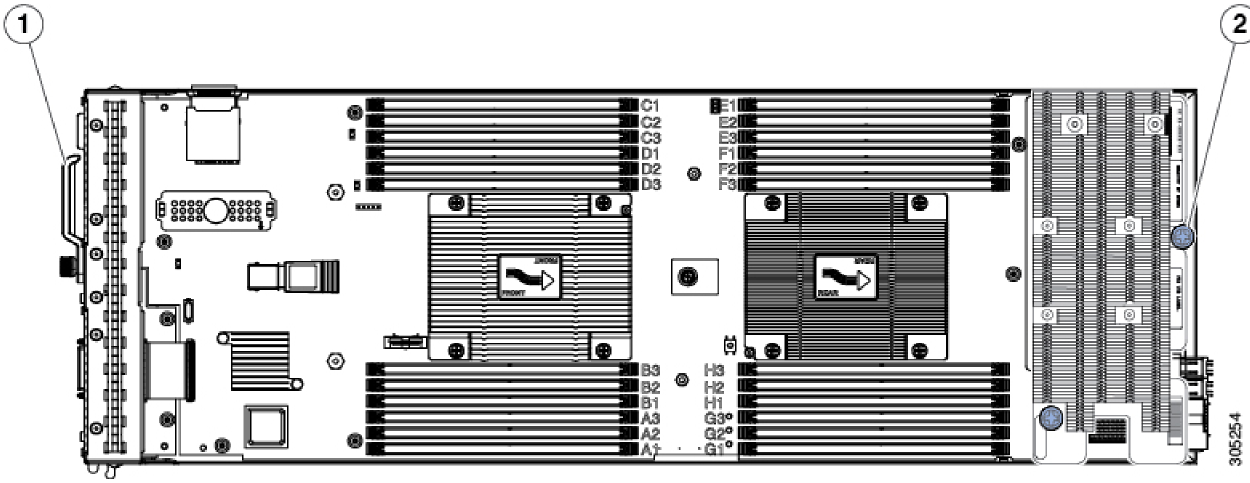
Step 4 Tighten the captive screws (callout 2).

Figure 21: Installing the NVIDIA M6 GPU



The following figure shows an NVIDIA M6 GPU installed in a Cisco UCS B200 M4 blade server.

Figure 22: Installed NVIDIA M6 GPU



1	Front of server	**2**	Custom standoff screw

What to do next

After you complete the installation of the NVIDIA M6 GPU, see NVIDIA Licensing Information, on page 37 for information on how to download NVIDIA software and acquire the necessary NVIDIA license. Follow the instructions to complete these steps in order:

1. Register your product activation keys with NVIDIA.
2. Download the GRID software suite.
3. Install the GRID License Server software to a host.
4. Generate licenses on the NVIDIA Licensing Portal and download them.
5. Manage your GRID licenses.
6. Decide whether to use the GPU in compute mode or graphics mode.

Enabling the Trusted Platform Module

The Trusted Platform Module (TPM) is a component that can securely store artifacts used to authenticate the server. These artifacts can include passwords, certificates, or encryption keys. A TPM can also be used to store platform measurements that help ensure that the platform remains trustworthy. Authentication (ensuring that the platform can prove that it is what it claims to be) and attestation (a process helping to prove that a platform is trustworthy and has not been breached) are necessary steps to ensure safer computing in all environments. It is a requirement for the Intel Trusted Execution Technology (TXT) security feature, which must be enabled in the BIOS settings for a server equipped with a TPM.

Procedure

Step 1 Install the TPM hardware.

a) Decommission and remove the blade server from the chassis.
b) Remove the blade server cover.
c) Install the TPM to the TPM socket on the server motherboard and secure it using the one-way screw that is provided. See the figure below for the location of the TPM socket.
d) Return the blade server to the chassis and allow it to be automatically reacknowledged, reassociated, and recommissioned.
e) Continue with enabling TPM support in the server BIOS in the next step.

Figure 23: TPM Socket Location

1	Front of server	**2**	TPM socket on motherboard

Step 2 Enable TPM Support in the BIOS.

If TPM support was disabled for any reason, use the following procedure to enable it.

a) In the Cisco UCS Manager Navigation pane, click the **Servers** tab.
b) On the Servers tab, expand **Servers** > **Policies**.
c) Expand the node for the organization where you want to configure the TPM.
d) Expand BIOS Policies and select the BIOS policy for which you want to configure the TPM.
e) In the Work pane, click the **Advanced** tab.
f) Click the **Trusted Platform** sub-tab.
g) To enable TPM support, click **Enable** or **Platform Default.**
h) Click **Save Changes**.
i) Continue with the next step.

Step 3 Enable TXT Support in the BIOS Policy.

Follow the procedures in the Cisco UCS Manager Configuration Guide for the release that is running on the server.

CHAPTER 4

Technical Specifications

This chapter contains the following section:

- Physical Specifications for the Cisco UCS B200 M4 Blade Server, on page 35

Physical Specifications for the Cisco UCS B200 M4 Blade Server

Specification	Value
Height	1.95 inches (50 mm)
Width	8.00 inches (203 mm)
Depth	24.4 inches (620 mm)
Weight	Base server weight = 9.51 lbs (4.31 kg) (no HDDs, no CPUs, no DIMMs, no mezzanine adapters or memory) Minimally configured server = 11.29 lbs (5.12 kg) (no HDDs, 1 CPU, 8 DIMMs, VIC 1340/1240 but no additional mezzanine adapter) Fully configured server = 15.98 lbs (7.25 kg) (2 HDDs, 2 CPUs, 24 DIMMs, VIC 1340/1240 and additional mezzanine adapter both populated)

APPENDIX A

NVIDIA Licensing Information

This chapter contains the following sections:

- NVIDIA GRID License Server Overview, on page 37
- Registering Your Product Activation Keys with NVIDIA, on page 38
- Downloading the GRID Software Suite, on page 39
- Installing NVIDIA GRID License Server Software, on page 39
- Installing GRID Licenses From the NVIDIA Licensing Portal to the License Server, on page 41
- Managing GRID Licenses, on page 43
- Installing Drivers to Support the NVIDIA GPU Cards, on page 44

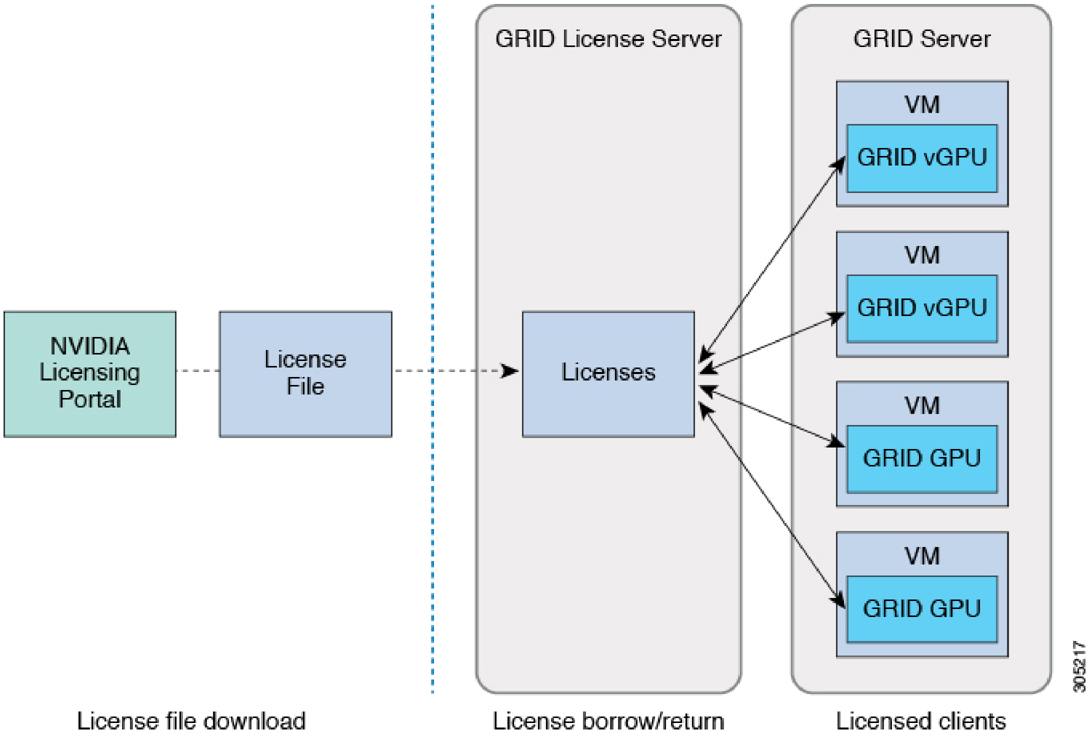
NVIDIA GRID License Server Overview

The NVIDIA Tesla P6 GPU combines Tesla and GRID functionality when you enable the licensed GRID features *GRID vGPU* and *GRID Virtual Workstation*. You enable these features during OS boot by borrowing a software license that is served over the network from the NVIDIA GRID License Server virtual appliance. The license is returned to the GRID License Server when the OS shuts down.

The NVIDIA Tesla P6 GPU has dual personality. It can work in Compute (Tesla) and GRID mode. Only GRID mode needs a license.

You obtain the licenses that are served by the GRID License Server from the NVIDIA Licensing Portal as downloadable license files, which you install into the GRID License Server via its management interface. See the following figure.

Figure 24: GRID Licensing Architecture



There are three editions of GRID licenses that enable three different classes of GRID features. The GRID software automatically selects the license edition based on the features that you are using. See the following table.

Table 4: GRID Licensing Editions

GRID License Edition	GRID Features
GRID Virtual GPU (vGPU)	Virtual GPUs for business desktop computing
GRID Virtual Workstation	Virtual GPUs for mid-range workstation computing
GRID Virtual Workstation - Extended	Virtual GPUs for high-end workstation computing Workstation graphics on GPU pass-through

Registering Your Product Activation Keys with NVIDIA

After your order is processed, NVIDIA sends you a Welcome email that contains your product activation keys (PAKs) and a list of the types and quantities of licenses that your purchased.

Procedure

Step 1 Select the **Log In** link, or the **Register** link if you do not already have an account.

The NVIDIA Software Licensing Center > License Key Registration dialog opens.

Step 2 Complete the License Key Registration form and then click **Submit My Registration Information**.
The NVIDIA Software Licensing Center > Product Information Software Dialog opens.

Step 3 If you have additional PAKs, click **Register Additional Keys**. For each additional key, complete the form on the License Key Registration dialog, and then click **Submit My Registration Information**.

Step 4 Agree to the terms and conditions and set a password when prompted.

Downloading the GRID Software Suite

Procedure

Step 1 Return to the NVIDIA Software Licensing Center > Product Information Software dialog box.

Step 2 Click **Current Releases**.

Step 3 Click the **NVIDIA GRID** link to access the **Product Download** dialog. This dialog includes download links for:

- The NVIDIA License Manager software
- The `gpumodeswitch` utility
- The host driver software

Step 4 Use the links to download the software.

Installing NVIDIA GRID License Server Software

For full instructions and troubleshooting information, see the *NVIDIA GRID License Server User Guide*. Also see the *NVIDIA GRID License Server Release Notes* for the latest information about your release. Both documents are available at the following URL:

http://www.nvidia.com

Platform Requirements for NVIDIA GRID License Server

- The hosting platform can be a physical or a virtual machine. NVIDIA recommends using a host that is dedicated to running only the License Server.
- The hosting platform must run a supported Windows OS.
- The hosting platform must have a constant IP address.
- The hosting platform must have at least one constant Ethernet MAC address.
- The hosting platform's date and time must be set accurately.

Installing on Windows

Before you begin

The License Server requires a Java Runtime Environment and an Apache Tomcat installation. Apache Tomcat is installed when you use the NVIDIA installation wizard for Windows.

Procedure

Step 1 Download and install the latest Java 32-bit Runtime Environment from https://www.oracle.com/downloads/index.html.

Note Install the 32-bit Java Runtime Environment, regardless of whether your platform is Windows 32-bit or 64-bit.

Step 2 Create a server interface.

a) In the **NVIDIA Software Licensing Center** dialog box, click **Grid Licensing** > **Create License Server**.

b) In the **Create Server** dialog box, fill in your desired server details.

c) Save the .bin file that is generated to your license server for installation.

Step 3 Unzip the NVIDIA License Server installer Zip file that you downloaded previously and run `setup.exe`.

Step 4 Accept the EULA for the NVIDIA License Server software and the Apache Tomcat software. Tomcat is installed automatically during the License Server installation.

Step 5 Use the installer wizard to step through the installation.

Note In the **Choose Firewall Options** dialog box, select the ports to be opened in the firewall. NVIDIA recommends that you use the default setting, which opens port 7070 but leaves port 8080 closed.

Step 6 To verify the installation, open a web browser on the License Server host and connect to the URL http://localhost:8080/licserver. If the installation was successful, you see the NVIDIA Licenses Client Manager interface.

Installing on Linux

Before you begin

The License Server requires a Java Runtime Environment and an Apache Tomcat installation. Use the following commands to install both separately before installing the License Server on Linux.

Procedure

Step 1 Verify that Java was installed with your Linux installation:

```
java -version
```

If a Java version does not display, use your Linux package manager to install Java:

```
sudo yum install java
```

Step 2 Use your Linux package manager to install the Tomcat and Tomcat webapp packages.

a) Install Tomcat:

```
sudo yum install tomcat
```

b) Enable the Tomcat service for automatic startup on boot:

```
sudo systemctl enable tomcat.service
```

c) Start the Tomcat service:

```
sudo systemctl start tomcat.service
```

d) To verify that the Tomcat service is operational, open a web browser on the License Server host and connect to the URL http://localhost:8080. If the installation was successful, you see the Tomcat webapp.

Step 3 Install the License Server.

a) Unpack the License Server tar file:

```
tar xfz NVIDIA-linux-2015.09-0001.tgz
```

b) Run the unpacked setup binary as root:

```
sudo./setup.bin
```

c) Accept the EULA and then continue with the installation wizard to finish the installation.

Note In the **Choose Firewall options** dialog, select the ports to be opened in the firewall. NVIDIA recommends that you use the default setting, which opens port 7070 but leaves port 8080 closed.

Step 4 To verify the installation, open a web browser on the License Server host and connect to the URL http://localhost:8080/licserver. If the installation was successful, you see the NVIDIA License Client Manager interface.

Installing GRID Licenses From the NVIDIA Licensing Portal to the License Server

Accessing the GRID License Server Management Interface

Open a web browser on the License Server host and access the URL http://localhost:8080/licserver.

If you configure the License Server host's firewall to permit remote access to the License Server, the management interface is accessible from remote machines at the URL http://localhost:8080/licserver.

Reading Your License Server's MAC Address

Your License Server's Ethernet MAC address is used as an identifier when registering the License Server with NVIDIA's Licensing Portal.

Procedure

Step 1 Access the GRID License Server Management Interface in a browser.

Step 2 In the left-side **License Server** panel, select **Configuration**.

Step 3 In the **License Server Configuration** panel, from the **Server host ID** pull-down menu, select your License Server's Ethernet MAC address.

Note It is important to use the same Ethernet ID consistently to identify the server when generating licenses on NVIDIA's Licensing Portal. NVIDIA recommends that you select one entry for a primary, non-removable Ethernet interface on the platform.

Installing Licenses From the Licensing Portal

Procedure

Step 1 Access the GRID License Server Management Interface in a browser.

Step 2 In the left-side **License Server** panel, select **Configuration**.

Step 3 From the **License Server Configuration** menu, click **Choose File**.

Step 4 Browse to the license .bin file that you generated earlier and want to install, and click **Open**.

Step 5 Click **Upload**.
The license file installs on your License Server. When the installation is complete, you see the confirmation message, "Successfully applied license file to license server."

Viewing Available Licenses

Use the following procedure to view the licenses that are installed and available and their properties.

Procedure

Step 1 Access the GRID License Server Management Interface in a browser.

Step 2 In the left-side **License Server** panel, select **Licensed Feature Usage**.

Step 3 Click a feature in the **Feature** column to see detailed information about the current usage of that feature.

Viewing Current License Usage

Use the following procedure to view information about that licenses that are currently in-use and borrowed from the server.

Procedure

Step 1 Access the GRID License Server Management Interface in a browser.

Step 2 In the left-side **License Server** panel, select **Licensed Clients**.

Step 3 To view detailed information about a single licensed client, click its **Client ID** in the list.

Managing GRID Licenses

Features that require GRID licensing run at reduced capability until a GRID license is acquired.

Acquiring a GRID License on Windows

To acquire a GRID license on Windows, use the following procedure.

Procedure

Step 1 Open the NVIDIA Control Panel using one of the following methods:

- Right-click the Windows desktop and select **NVIDIA Control Panel** from the menu.
- Open the Windows Control Panel and double-click the **NVIDIA Control Panel** icon.

Step 2 In the **NVIDIA Control Panel** left pane under **Licensing**, select **Manage License**.
The **Manage License** task pane opens and shows the current license edition being used. The GRID software automatically selects the license edition based on the feature that you are using. The default is **Tesla (unlicensed)**.

Step 3 If you want to acquire a license for GRID Virtual Workstation, under **License Edition**, select **GRID Virtual Workstation**.

Step 4 In the **License Server** field, enter the address of your local GRID License Server.

The address can be a domain name or an IP address.

Step 5 In the **Port Number** field, enter your port number or leave it set to the default used by the server, which is 7070.

Step 6 Select **Apply**.
The system requests the appropriate license edition from your configured License Server. After a license is successfully acquired, the features of that license edition are enabled.

Note After you configure licensing settings in the NVIDIA Control Panel, the settings persist across reboots.

Acquiring a GRID License on Linux

To acquire a GRID license on Linux, use the following procedure.

Procedure

Step 1 Edit the configuration file:

```
sudo vi /etc/nvidia/gridd.conf
```

Step 2 Edit the `ServerUrl` line with the address of your local GRID License Server.

The address can be a domain name or an IP address. See the Sample Configuration File.

Step 3 Append the port number (default **7070**) to the end of the address with a colon. See the Sample Configuration File.

Step 4 Edit the `FeatureType` line with an integer for the license type. See the Sample Configuration File.

- GRID vGPU = 1
- GRID Virtual Workstation = 2

Step 5 Restart the nvidia-gridd service:

```
sudo service nvidia-gridd restart
```

The service automatically acquires the license edition that you specified in the `FeatureType` line. You can confirm this in `/var/log/messages`.

Note After you configure licensing settings in `gridd.conf`, the settings persist across reboots.

Sample Configuration File

```
# /etc/nvidia/gridd.conf - Configuration file for NVIDIA Grid Daemon
# Description: Set License Server URL
# Data type: string
# Format: "<address>:<port>"
Server URL=10.31.20.45:7070

# Description: set Feature to be enabled
# Data type: integer
# Possible values:
# 1 => for GRID vGPU
# 2 => for GRID Virtual Workstation
FeatureType=1
```

Installing Drivers to Support the NVIDIA GPU Cards

After you install the hardware, you must update to the correct level of server BIOS, activate the BIOS firmware, and then install NVIDIA drivers and other software in this order:

1. Updating the Server BIOS Firmware

Install the latest Cisco server BIOS for your blade server by using Cisco UCS Manager.

Note You must do this procedure before you update the NVIDIA drivers.

Caution Do not remove the hardware that contains the endpoint or perform any maintenance on it until the update process completes. If the hardware is removed or otherwise unavailable due to maintenance, the firmware update fails. This failure might corrupt the backup partition. You cannot update the firmware on an endpoint with a corrupted backup partition.

Procedure

Step 1 In the **Navigation** pane, click **Equipment**.

Step 2 On the **Equipment** tab, expand **Equipment** > **Chassis** > ***Chassis Number*** > **Servers**.

Step 3 Click the **Name** of the server for which you want to update the BIOS firmware.

Step 4 On the **Properties** page in the **Inventory** tab, click **Motherboard**.

Step 5 In the **Actions** area, click **Update BIOS Firmware**.

Step 6 In the **Update Firmware** dialog box, do the following:

a) From the **Firmware Version** drop-down list, select the firmware version to which you want to update the endpoint.
b) Click **OK**.

Cisco UCS Manager copies the selected firmware package to the backup memory slot, where it remains until you activate it.

Step 7 (Optional) Monitor the status of the update in the **Update Status** field.

The update process can take several minutes. Do not activate the firmware until the firmware package you selected displays in the **Backup Version** field in the **BIOS** area of the **Inventory** tab.

What to do next

Activate the server BIOS firmware.

2. Activating the Server BIOS Firmware

Procedure

Step 1 In the **Navigation** pane, click **Equipment**.

Step 2 On the **Equipment** tab, expand **Equipment** > **Chassis** > ***Chassis Number*** > **Servers**.

Step 3 Click the **Name** of the server for which you want to activate the BIOS firmware.

Step 4 On the **Properties** page in the **Inventory** tab, click **Motherboard**.

Step 5 In the **Actions** area, click **Activate BIOS Firmware**.

Step 6 In the **Activate Firmware** dialog box, do the following:

a) Select the appropriate server BIOS version from the **Version To Be Activated** drop-down list.

b) If you want to set only the start-up version and not change the version running on the server, check **Set Startup Version Only**.

If you configure **Set Startup Version Only**, the activated firmware moves into the pending-next-reboot state and the server is not immediately rebooted. The activated firmware does not become the running version of firmware until the server is rebooted.

c) Click **OK**.

What to do next

Update the NVIDIA drivers.

3. Updating the NVIDIA Drivers

After you update the server BIOS, you can install NVIDIA drivers to your hypervisor virtual machine.

Procedure

Step 1 Install your hypervisor software on a computer. Refer to your hypervisor documentation for the installation instructions.

Step 2 Create a virtual machine in your hypervisor. Refer to your hypervisor documentation for instructions.

Step 3 Install the NVIDIA drivers to the virtual machine. Download the drivers:

- NVIDIA Enterprise Portal for GRID hypervisor downloads (requires NVIDIA login): https://nvidia.flexnetoperations.com/
- NVIDIA public driver area: http://www.nvidia.com/Download/index.aspx

Step 4 Restart the server.

Step 5 Check that the virtual machine is able to recognize the NVIDIA card. In Windows, use the **Device Manager** and look under **Display Adapters**.