

Cisco UCS X580p PCIe Node Installation and Service Guide

First Published: 2025-07-31

Last Modified: 2026-03-25

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

© 2025 –2026 Cisco Systems, Inc. All rights reserved.



CONTENTS

CHAPTER 1 **Overview** 1

Overview 1

Front Panel 1

Node Identification 2

LEDs 4

GPU Cage Options 5

Slot Numbering 5

Supported GPUs 9

Supported Network Adapters 9

Required Hardware 10

CHAPTER 2 **Installing the Cisco UCS X580p PCIe Node** 13

Installation Guides and Limitations 13

PCIe Node Companion Module 15

Cisco UCS X580p PCIe Node Mapping 16

Replacing the Cisco UCS X580P PCIe Node 17

Installing the PCIe Node 18

Removing the PCIe Node 19

PCIe Node Configuration 21

CHAPTER 3 **Servicing the Cisco UCS X580p PCIe Node** 23

Guidelines and Limitations 23

Serviceable Component Locations 25

Removing the PCIe Node Cover 26

Replacing the PCIe Node Air Baffle 27

Removing the PCIe Node Air Baffle 28

Installing the PCIe Node Air Baffle 30
Replacing the PCIe Node Front Cover 32
Removing the PCIe Node Front Cover 32
Installing the PCIe Node Front Cover 34
Replacing GPU Cages 35
Cable Reference 36
GPU Cage Population Guidelines 36
Removing a GPU Cage 37
Installing a GPU Cage 43
Replacing a GPU Card 48
GPU Replacement Guidelines and Limitations 49
Replacing a Legacy GPU Extender Bracket 49
Removing the Legacy Extender Bracket 50
Installing the Enhanced Extender Bracket 51
Removing a FHFL GPU 52
Installing a FHFL GPU 54
Replacing the NVIDIA NVLink GPU Interconnect 57
Removing the NVIDIA NVLink GPU Interconnect 57
Installing the NVIDIA NVLink GPU Interconnect 58
Installing the PCIe Node Cover 60

CHAPTER 4 **Recycling the PCB Assemblies** 61
Recycling the PCB Assemblies 61
Recycling the GPU Cage PCBA 61
Recycling the Node PCBA 63

APPENDIX A **PCIe Node Specifications** 73
Physical Specifications 73
Environmental Specifications 73
Power Specifications 74

APPENDIX B **Obtaining Hardware** 75
Obtaining Hardware 75



Preface

This preface contains the following topics:

- Bias-Free Documentation, on page v
- Full Cisco Trademarks with Hardware License, on page v
- Communications, Services, and Additional Information, on page vii

Bias-Free Documentation

Note 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.

Full Cisco Trademarks with Hardware License

THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS IN THIS MANUAL ARE SUBJECT TO CHANGE WITHOUT NOTICE. ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS MANUAL ARE BELIEVED TO BE ACCURATE BUT ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. USERS MUST TAKE FULL RESPONSIBILITY FOR THEIR APPLICATION OF ANY PRODUCTS.

THE SOFTWARE LICENSE AND LIMITED WARRANTY FOR THE ACCOMPANYING PRODUCT ARE SET FORTH IN THE INFORMATION PACKET THAT SHIPPED WITH THE PRODUCT AND ARE INCORPORATED HEREIN BY THIS REFERENCE. IF YOU ARE UNABLE TO LOCATE THE SOFTWARE LICENSE OR LIMITED WARRANTY, CONTACT YOUR CISCO REPRESENTATIVE FOR A COPY.

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.

NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED "AS IS" WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

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.

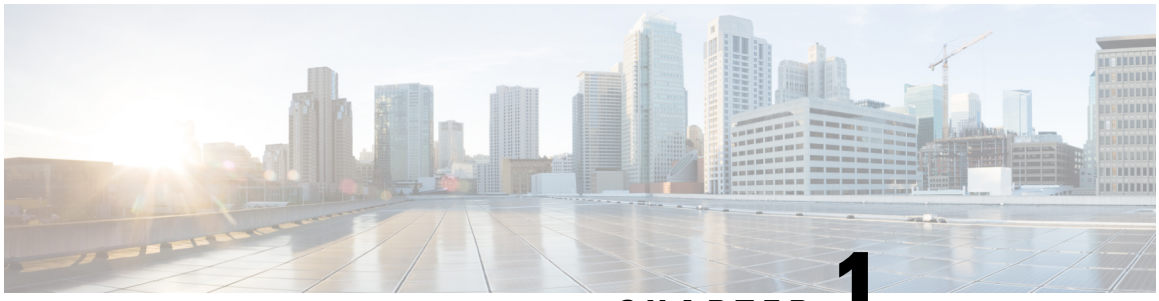
Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: https://www.cisco.com/c/en/us/about/legal/trademarks.html. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1721R)

Communications, Services, and Additional Information

- To receive timely, relevant information from Cisco, sign up at Cisco Profile Manager.
- To get the business results you're looking for with the technologies that matter, visit Cisco Services.
- To submit a service request, visit Cisco Support.
- To discover and browse secure, validated enterprise-class apps, products, solutions and services, visit Cisco DevNet.
- To obtain general networking, training, and certification titles, visit Cisco Press.
- To find warranty information for a specific product or product family, access Cisco Warranty Finder.

Cisco Bug Search Tool

Cisco Bug Search Tool (BST) is a web-based tool that acts as a gateway to the Cisco bug tracking system that maintains a comprehensive list of defects and vulnerabilities in Cisco products and software. BST provides you with detailed defect information about your products and software.



CHAPTER 1

Overview

This chapter contains the following topics:

- Overview, on page 1

Overview

The Cisco UCS X580p PCIe Node (UCSX-580P) is the second generation PCIe node that is part of the Cisco UCS X-Fabric system.

The PCIe node pairs with Cisco UCS X-Series M8 compute nodes and Cisco UCS X9516 X-Fabric Modules to support robust workloads that require GPU processing power. The node is designed to work with its related hardware to simplify adding, removing, or upgrading GPUs on compute nodes.

The PCIe node supports PCIe connectivity for a variety of GPUs.

- GPUs: Four x16 FHFL (full-height, full length) dual slot PCIe cards, zero, one, or two GPUs per GPU cage.

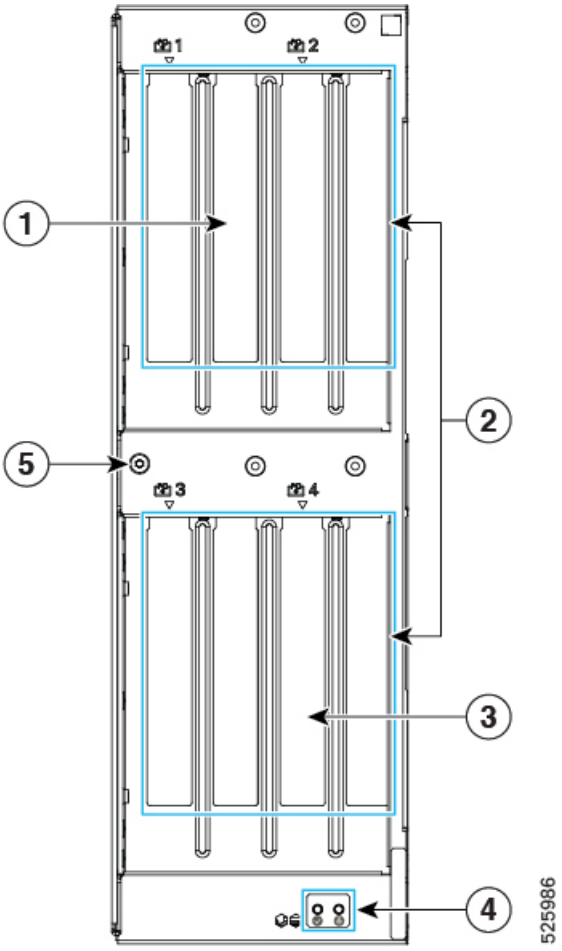
Note Each PCIe node supports GPUs through two GPU cages. Each PCIe node can be configured with different types of GPUs, but each GPU cage must be configured with the same type of GPU. For example, a PCIe node can support NVIDIA H200 GPUs and NVIDIA L40S GPUs, but each type must be installed in the same GPU cage. For supported GPUs, see Supported GPUs, on page 9.

- Host connection between the PCIe mezzanine (MEZZ) and the PCIe node is supported through a PCIe Gen 5 (2 x16) connector in the rear MEZZ slot.
- GPU cage option, a maximum of two GPU cages is supported in each PCIe node. Each GPU cage supports up to two FHFL dual-slot GPUs.

Front Panel

The Cisco UCS X580p PCIe Node is a single PCIe node that occupies two slots in the Cisco UCS X-Series server chassis. Each node is front loading, so it is inserted into, and removed from, the front of the Cisco UCS X9508 chassis.

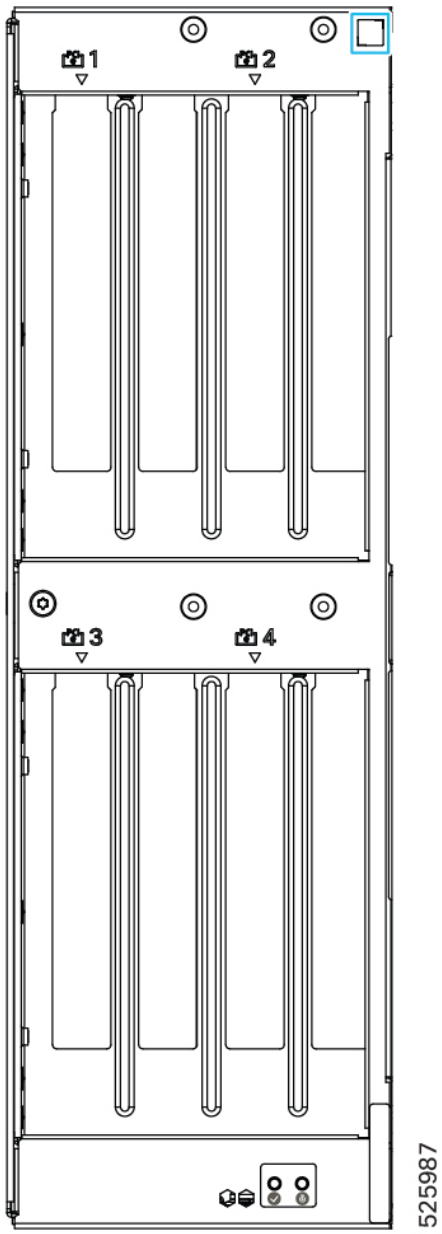
The following image shows the PCIe node front panel.



1	PCIe node slots 1 and 2	2	PCIe node Ejector handles, 2
3	PCIe node slots 3 and 4	4	PCIe Node Status LEDs. See LEDs, on page 4.
5	Mounting Hole for additional physical features.	-	

Node Identification

Each Cisco UCS X580p PCIe Node (UCSX-580P) features a product identification QR code at the upper right corner of the node.



The node identification QR code contains information that uniquely identifies the product, such as:

- The Cisco product identifier (PID) or virtual identifier (VID)
- The product serial number

The node identification QR code applies to the entire PCIe node, not the individual GPUs installed in it.

You will find it helpful to scan the QR code so that the information is available if you need to contact Cisco personnel.

LEDs

The PCIe node front panel has the following LEDs.

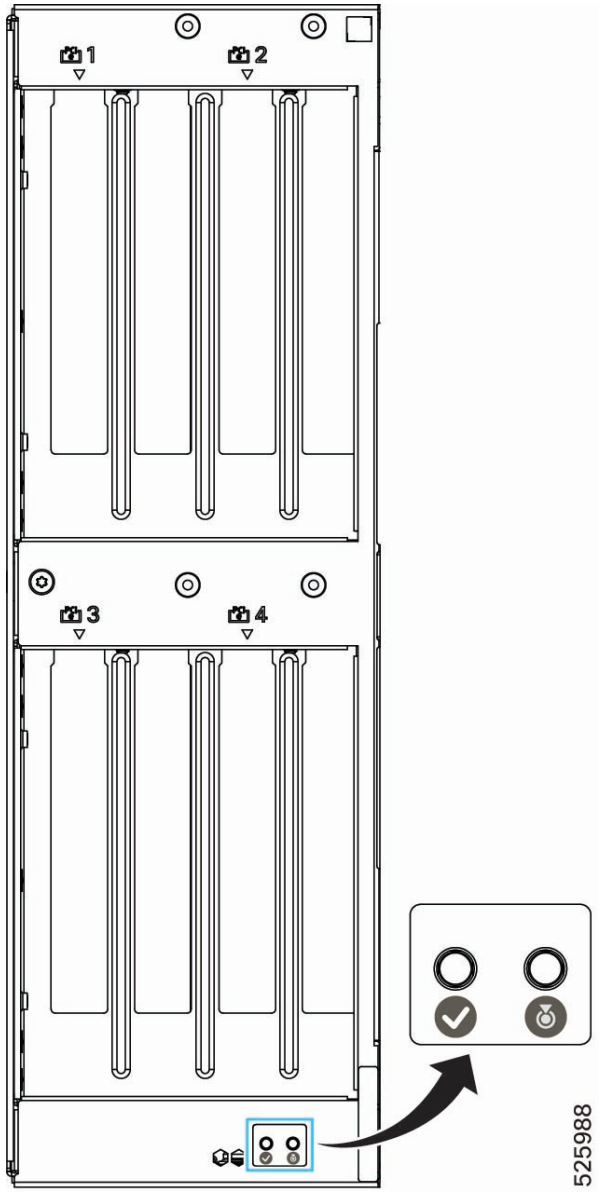


Table 1: PCIe Node LEDs

LED	Color	Description
PCIe Node Status 	Off	PCIe node is not powered on/initialized.
	Solid Green	Host is receiving power or the zone is active.
	Blinking Green (1 Hz)	Host is powered off. You can safely remove the PCIe node.
	Solid Amber	Warning or Fault condition exists, such as a configuration or system or device inventory problem.
	Blinking Amber (1 Hz)	Severe fault, such as an insufficient power condition.
PCIe Node ID/Locator LED 	Off	Locator not enabled.
	Blinking Blue (1 Hz)	Locates a selected PCIe node—If the LED is not blinking, the PCIe node is not selected. You can activate the LED in UCS Intersight, which enables toggling the LED on and off.

GPU Cage Options

Each PCIe node contains one type of GPU cage that accepts a maximum of two FHFL dual-slot GPUs through two x16 FHFL dual-slot PCIe connectors.

Each GPU cage mounts directly to the PCIe node's sheet metal. GPU cages are identified by their position as either `A` or `B` when mounted on the PCIe node. The node has labels, `A` or `B`, on the PCIe node under each GPU cage that identify the cages' locations.

Power and signaling are supported by cables that connect each GPU cage to the PCIe node rear mezzanine PCBA.

If your PCIe node is not fully populated, the empty GPU cage must be installed, and the GPU cage itself must have PCIe slot blanks installed when no GPUs are present. For example, if you have a PCIe node with two GPUs in Cage A, but no GPUs in Cage B, Cage B must contain two Cisco PCIe blanks, and it must be installed on the PCIe node even though Cage B is empty.

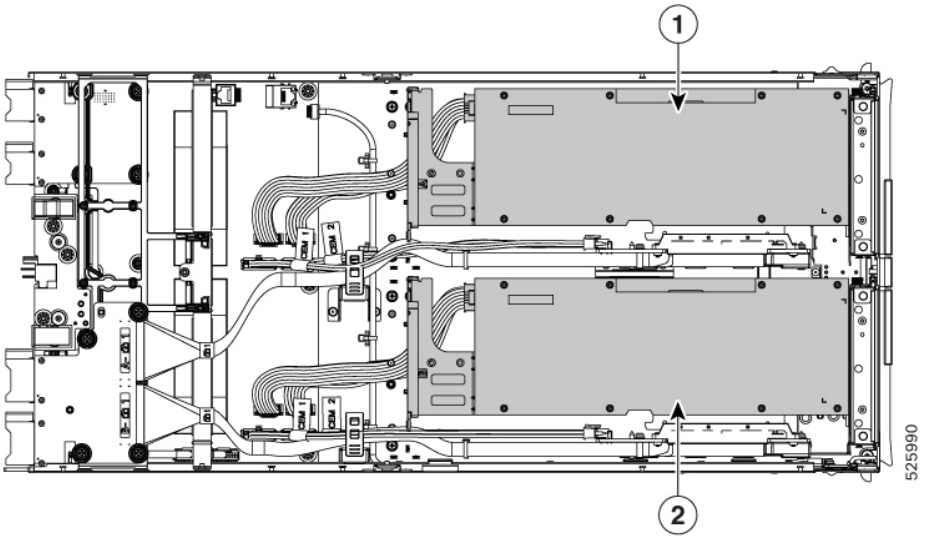
Slot Numbering

GPU cages and GPU slots have specific identifiers that determine their locations in the PCIe node.

- For hardware, slot numbering consists of `GPU_cage_Identifier/Slot`. So, for example, `A/1` indicates GPU cage A, slot 1.
- For Cisco management software, such as Cisco Intersight Managed Mode (IMM), the slot numbering consists of **Cage** *[identifier]*-**Slot** *[number]*. So, for example, `CageA-SLOT1` indicates GPU cage A, slot 1.

GPU Cage Numbering

Figure 1: GPU Cage Numbering



1	GPU cage A GPU cage A is controlled by CPU1 on the Cisco UCS M8 compute node paired with the PCIe node	2	GPU Cage B GPU cage B is controlled by CPU2 on the Cisco UCS M8 compute node paired with the PCIe node

GPU Slot Numbering

Figure 2: GPU Slot Numbering, Node View

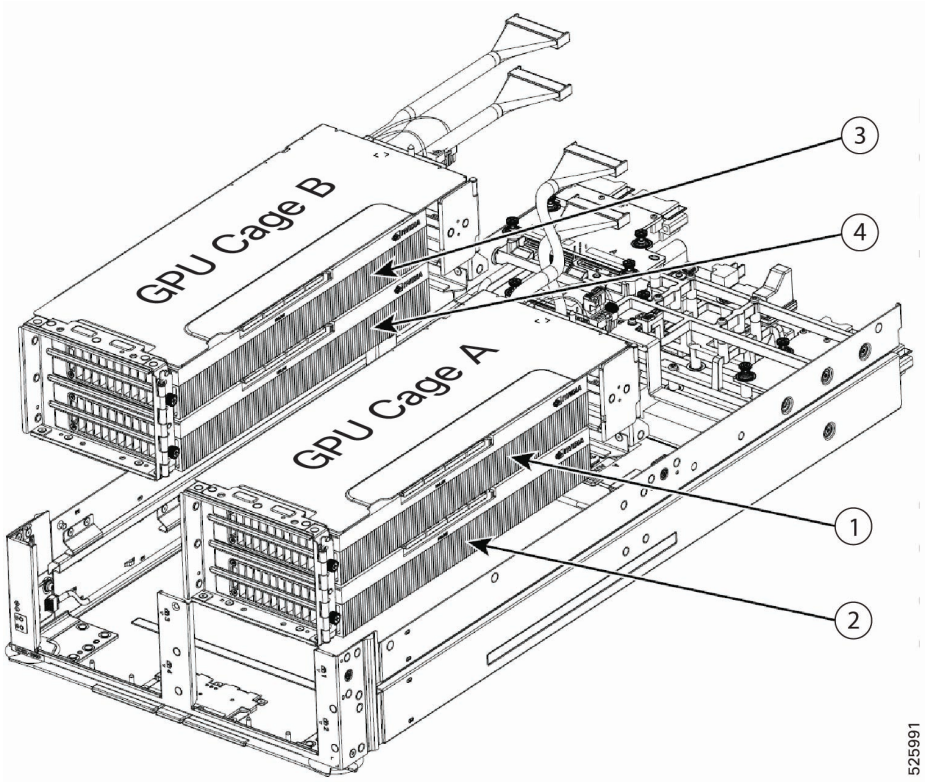
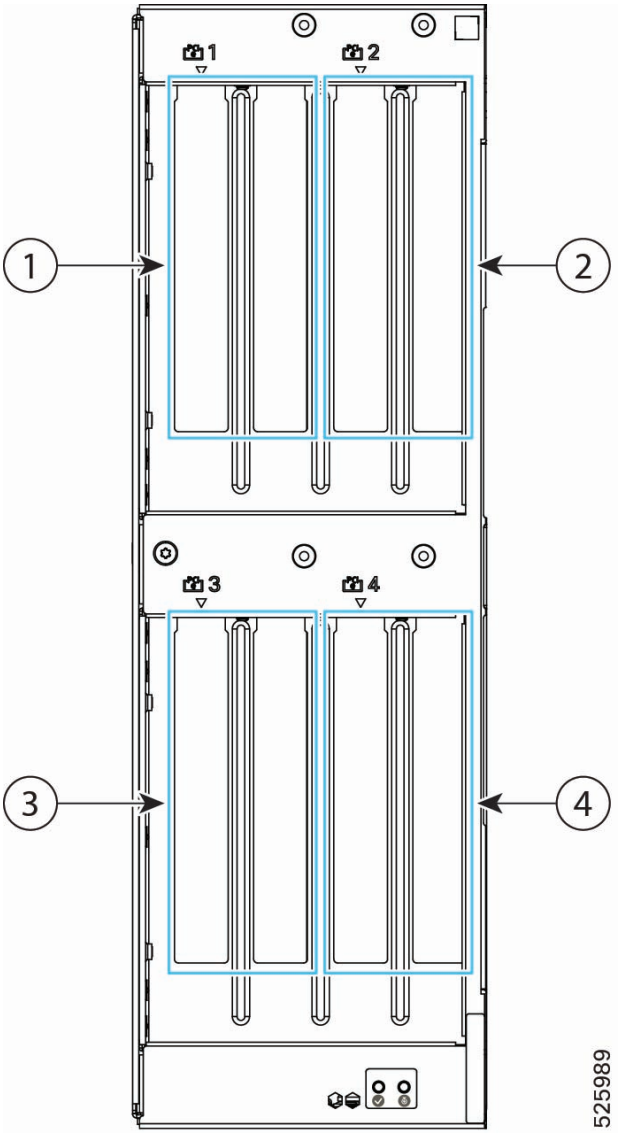


Figure 3: GPU Slot Numbering, Front Panel View



Note PCIe slot numbers are labeled onto the front of each X580p PCIe Node.

1	Slot 1 for FHFL dual-slot GPU GPU cage A (which contains slots 1 and 2) is controlled by CPU1 on the Cisco UCS M8 compute node paired with the PCIe node	2	Slot 2 for FHFL dual-slot GPU

3	Slot 3 for FHFL dual-slot GPU. GPU cage B (which contains slots 3 and 4) is controlled by CPU2 on the Cisco UCS M8 compute node paired with the PCIe node	4	Slot 4 for FHFL dual-slot GPU.

Supported GPUs

The following table shows the models and form factors of GPU supported by the Cisco UCS X580p PCIe Node. When the system is initially ordered, GPUs are installed at the factory, but additional GPUs can be ordered as spares as your deployment scales out. When additional GPUs are ordered, each GPU listed in the following table includes a power cable.

Any GPU cage can accept all of the supported GPUs. For example, Cage A is not "reserved" for only H200 GPUs.

However, be aware of the following GPU mixing consideration: GPUs can be mixed across the node, but not within GPU Cages. For example, H200 NVL GPUs can be installed in Cage A, and RTX PRO 6000 GPUs can be installed in Cage B. However, you cannot install a mix of H200 and RTX PRO 6000 GPUs in the same cage.

Table 2: Full-Height, Full Length (FHFL) Dual-Slot GPUs

GPU	Cisco PID
NVIDIA L40S FHFL PCIe Gen 4 350W, x16 lanes (64 Gbps bidirectional), 48 GB memory	UCSX-GPU-L40S=
NVIDIA H200 NVL Dual Slot, FHFL PCIe Gen 5 600W, x16, 5x 8, or 4x 16 lanes, 141 GB HBM3e memory Two-Way NVLink Bridge for point-to-point GPU Interconnect support	UCSX-GPU-H200-NVL=
NVIDIA RTX PRO 6000 FHFL PCIe Gen 5 600W, x16 lanes, 141 GB memory	UCSX-GPU-RTXP6000=

Supported Network Adapters

Network adapters provide logical connections, such as signaling and data, for the PCIe nodes.

It is important to understand that supported network adapters are not physically installed on the Cisco UCS X580p PCIe Node, but they are required for GPU-to-GPU communication across the nodes. Instead, the network adapters are installed in the UCS X9516 X Fabric Modules (XFMs) that are present in the same Cisco UCS X9508 chassis that contains the PCIe nodes. For more information about the Cisco UCS X9516 XFM, which accepts the network adapters, see X Fabric Modules in the *Cisco UCS X9508 Chassis Installation and Service Guide*.

The following network adapters are required for end-to-end communication between the PCIe Nodes and the rest of the network.

Supported Adapter	Cisco PID
NVIDIA ConnectX-7 Network Adapter, PCIe Gen 5 HHHL, 1x400 Gbps QSFP port, 66W Maximum of two per Cisco UCS X9516 X Fabric Module (XFM)	UCSX-P-N7S400GFO
NVIDIA ConnectX-7 Network Adapter, PCIe Gen 5 HHHL, 2x200 Gbps QSFP port, 66W Maximum of two per Cisco UCS X9516 X Fabric Module (XFM)	UCSX-P-N7D200GFO

Required Hardware

The Cisco UCS X580p PCIe Node is part of an integrated system to provide GPU acceleration for Cisco UCS M8 compute nodes. For a complete system, the PCIe node requires the following hardware components.

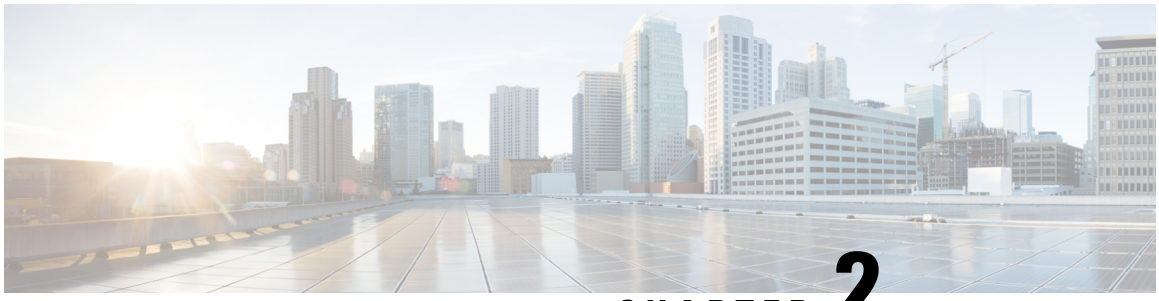
- An X-Series server chassis, such as the Cisco UCS X9508 Server Chassis, that contains the paired compute and PCIe Nodes and the Cisco X-Fabric Modules. For more information, go to the Cisco UCS X9508 Server Chassis Installation Guide.
- Two Cisco UCS X9516 X-Fabric PCIe Gen 5 Switch Modules (UCSX-FS-9516) for the Cisco UCS X9508 Chassis to connect Cisco UCS M8 compute nodes to the Cisco UCS X580p PCIe Node in the X-Series Server chassis. For more information, go to the Cisco UCS X9508 Server Chassis Installation Guide.
- Cisco UCS Compute Nodes that pair with the Cisco UCS X580p PCIe Node, such as the following:
  - Cisco UCS X210c M8 Compute Node with Intel® Xeon® 6 processors. For more information, go to the Cisco UCS X210c M8 Compute Node Installation and Service Guide.
  - Cisco UCS X215c M8 Compute Node with AMD® EPYC® processors. For more information, go to the Cisco UCS X215c M8 Compute Node Installation and Service Guide.
  - Cisco UCS X210c M7 Compute Node with Fifth Generation Intel® Xeon® Scalable Processors processors. For more information, go to the Cisco UCS X210c M7 Compute Node Installation and Service Guide.

Note Although the Cisco UCS X210c M7 Compute Node offers an option with Fourth Generation Xeon® processors, this version is not supported with the Cisco UCS X580p PCIe Node.

- Cisco UCS PCIe Gen 5 Mezz Card for X Fabric (UCSX-V5-PCIME), which is a rear mezzanine card required on each paired UCS M8 compute node.

- The Cisco UCS X580p PCIe Node itself, up to two per Cisco UCS X9508 chassis, to pair with the M8 compute nodes.
- GPU Cages, which are designed for full height, full length (FHFL) dual-slot GPUs. GPU cages are part of each PCIe Node. For more information, see GPU Cage Options, on page 5.
- GPUs, which provide the GPU processing offload and acceleration. For more information, see Supported GPUs, on page 9.

If you require hardware to support or expand your PCIe node support, see Obtaining Hardware, on page 75.



CHAPTER 2

Installing the Cisco UCS X580p PCIe Node

This chapter contains the following topics:

- Installation Guides and Limitations, on page 13
- PCIe Node Companion Module, on page 15
- Replacing the Cisco UCS X580P PCIe Node, on page 17
- PCIe Node Configuration, on page 21

Installation Guides and Limitations

Be aware of the following guidelines and limitations when installing the Cisco UCS X580p PCIe Node.

- The Cisco UCS X580p PCIe Node connects with the Cisco UCS X9516 X-Fabric Module to provide PCIe Gen5 connectivity. However, Cisco also offers PCIe Gen4 connectivity through the Cisco UCS X440p PCIe Node and Cisco UCS X9416 X-Fabric Modules. You cannot mix these products within the same Cisco UCS X9508 chassis. One X9508 chassis can support either the Gen5 or Gen4 products, but not both.
- The Cisco UCS X580p PCIe Node is supported in the Cisco UCS X9508 chassis only. Do not attempt to install the PCIe node in any other UCS server chassis.
- The companion compute node's host power should be off before attempting any removal or insertion procedures. Hot removal or insertion of the PCIe node while the host is on is not supported.
- When installing the Cisco UCS X580p PCIe Nodes and Cisco UCS X9516 X-Fabric Modules, make sure to install the XFMs before installing the PCIe nodes.

Recommended Configuration

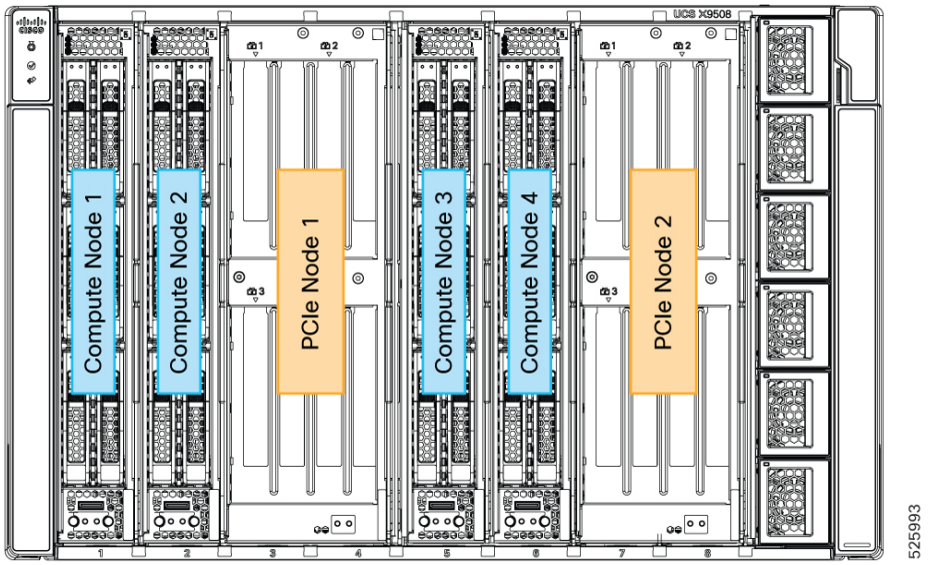
Note Each new Cisco UCS X9508 chassis will ship with the Cisco UCS X580p PCIe node(s) pre-installed. However, for existing deployments, the Cisco UCS X580p PCIe node can be shipped by itself. If you are installing PCIe nodes in an existing Cisco UCS X9508 deployment, make sure to install the PCIe nodes as documented below.

Each PCIe node must be paired with an X-Series M8 compute node, and therefore, has specific configurations based on the type of M8 compute node.

For single-slot M8 compute nodes (for example, the Cisco UCS X210c M8), the best practice is install PCIe and M8 compute nodes as shown.

In this configuration, the PCIe Nodes are installed in slots 3/4 and 7/8. The companion M8 compute nodes are installed in chassis slots 1/2 and 5/6. Be aware of these configurations, especially when removing and installing M8 compute nodes and PCIe nodes in your deployment.

Figure 4: Recommended Configuration



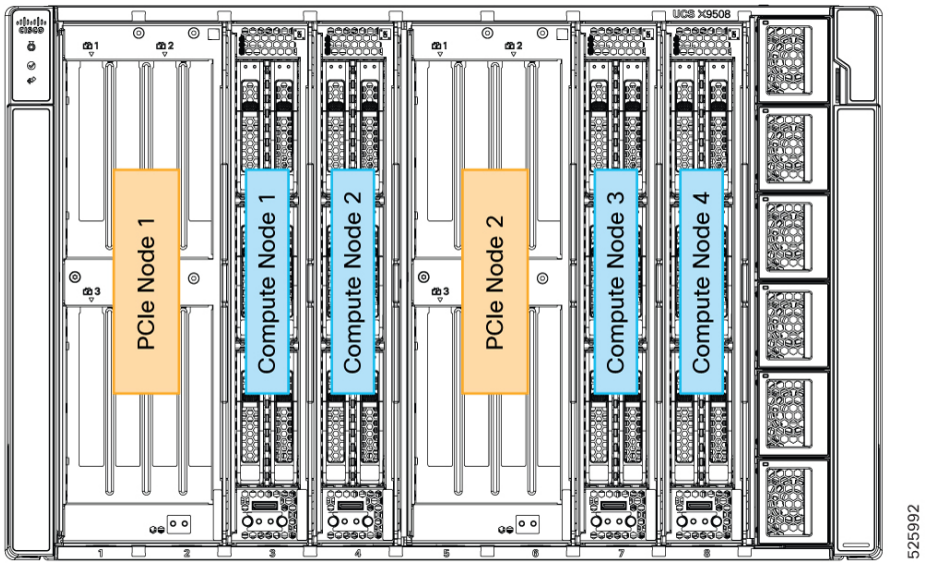
Supported Configuration

The following additional configuration is supported, but not recommended.

In this configurations, the PCIe node is installed in slots 1/2 and 5/6, and the M8 compute nodes are supported in slots 3/4 and 7/8.

Note Although this configuration is supported, the best practice is to install PCIe and M8 compute nodes as shown in the Recommended Configuration.

Figure 5: Supported Configuration



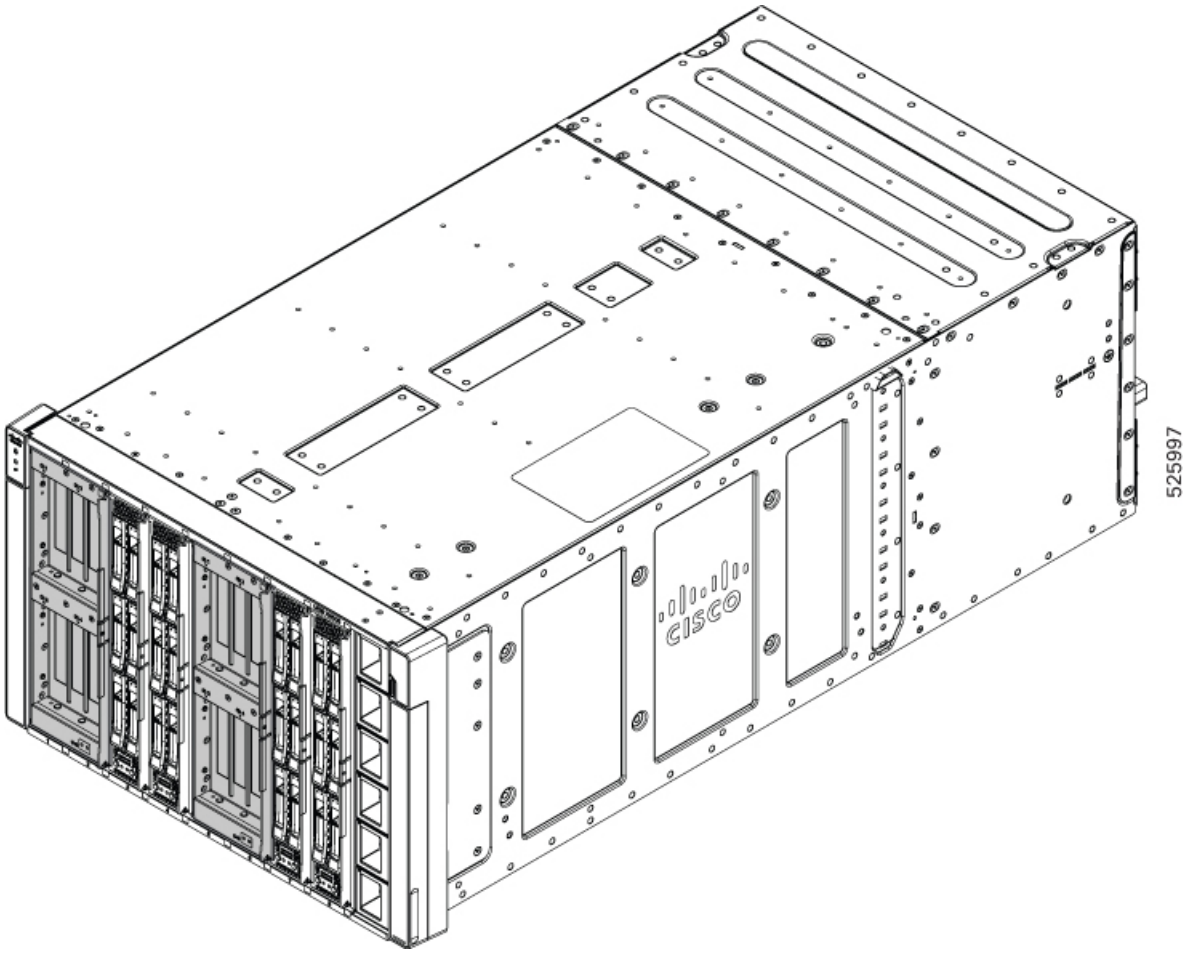
PCIe Node Companion Module

The PCIe node must be installed to pair with one or two M8 compute nodes. The PCIe node and M8 compute node(s) work together to provide compute and GPU processing acceleration.

Each pairing of PCIe node and M8 compute node connect to a UCS X9516 X Fabric switching module at the rear of the Cisco UCS X9508 chassis. Both XFMs serve all nodes in the chassis, but the following PCIe mappings are based on the specific XFM.

- XFM 1 (the top module) supports PCIe device mapping to CPU1 on the M8 compute node(s). On the PCIe node, PCIe slots 1 and 2 are supported by XFM 1 in the X9508 chassis and CPU1 on the companion M8 compute node.
- XFM 2 (the bottom module) supports PCIe device mapping to CPU2 on the M8 compute node(s). On the PCIe node, PCIe slots 3 and 4 are supported by XFM 2 in the X9508 chassis and CPU2 on the companion M8 compute node.

When you install, replace, or upgrade the PCIe node, make sure that you install it next to a supported companion modules, such as a Cisco UCS X210c M8 compute node.



Note The Cisco X-Fabric solution does not limit or restrict slot usage when PCIe nodes are not required. Cisco compute nodes can be installed in any slot without restriction when PCIe nodes are not used.

Cisco UCS X580p PCIe Node Mapping

The Cisco UCS X580p PCIe Node is connected to compute nodes in the Cisco UCS X9508 Server Chassis through an X-Fabric Module. This topic shows specific mappings for the X580p PCIe Node.

Physical Slot Mappings for the PCIe Node in the Cisco UCS X9508 Server Chassis

The X580p PCIe Node is a dual-slot node. In the Cisco UCS X9508 chassis, the X580p PCIe Node must be installed in specific slot pairs to maintain connectivity with its paired compute nodes (typically the Cisco UCS X210c M8 or Cisco UCS X215c M8).

Configuration Type	PCIe Node Slots	Paired Compute Node Slots
Recommended	Slots 3 and 4	Slots 1 and 2

Configuration Type	PCIe Node Slots	Paired Compute Node Slots
Recommended	Slots 7 and 8	Slots 5 and 6
Supported	Slots 1 and 2	Slots 3 and 4
Supported	Slots 5 and 6	Slots 7 and 8

Note Connectivity is established between nodes on the same side of the chassis where the vertical sheetmetal wall between slots 4 and 5 is the dividing line between the sides. For example, an X580p PCIe Node in the left half of the chassis (Slots 1-4) can map to a compute node in the left half of the chassis (Slots 1-4), but it cannot map to a compute node in the right half of the chassis (Slots 5-8).

GPU-to-CPU Logical Mappings

The X580p PCIe Node contains two GPU cages (Cage A and Cage B), each supporting up to two FHFL (Full-Height, Full-Length) dual-slot GPUs. The mapping of these GPUs to a compute node's CPUs is determined by the X-Fabric Modules (XFMs) that provide logical/data connection for the PCIe zone as shown in the following table.

Table 3: GPU Cage-to-CPU Mapping Through an X-Fabric Module

GPU Cage (Slots)	Target CPU on Compute Node	Fabric Module (XFM)
Cage A (Slots 1 & 2)	CPU 1	XFM 1 (Top Fabric Module)
Cage B (Slots 3 & 4)	CPU 2	XFM 2 (Bottom Fabric Module)

Mapping Scenarios

The X580p PCIe Node offers flexibility in how its four GPU slots are assigned to servers. The PCIe node can be assigned either 1:1 (one PCIe node to a single compute node) or 1:2 (one PCIe one to two compute nodes), as shown in the following table.

Table 4: X580p PCIe Node-to-Compute Node Mapping Scenarios

Mapping Type	Description	Example
1:1 Mapping (Single Server)	All 4 GPU slots in the X580p are assigned to a single dual-slot or single-slot compute node.	One X210c M8 in Slot 1/2 receives all 4 GPUs from the X580p in Slot 3/4.
1:2 Mapping (Shared Node)	The X580p is shared between two single-slot compute nodes.	Compute Node 1 (Slot 1) is assigned Cage A (2 GPUs), and Compute Node 2 (Slot 2) is assigned Cage B (2 GPUs).

Replacing the Cisco UCS X580P PCIe Node

The Cisco UCS X580p PCIe Node is tool-less, field replaceable hardware that can be installed into the Cisco UCS X9508 chassis. You can install or remove the node by using the node's integrated ejectors.

Caution Hot swap of the PCIe node is not supported while the host is on. The paired M8 compute node must be powered down first before removing or inserting the PCIe node.

For proper operation, if you remove a PCIe node and will not install a new one, you must cover the PCIe node slot with a filler panel. Do not operate the Cisco UCS X9508 chassis with any empty, uncovered module slot.

To replace the PCIe node, use the following tasks:

- Installation Guides and Limitations, on page 13.
- Removing the PCIe Node, on page 19
- Installing the PCIe Node, on page 18

Installing the PCIe Node

Use this task to install the PCIe node.

Caution Before inserting a PCIe node, make sure that its paired M8 compute node is powered off.

Before you begin

The PCIe node is a tools-free module that you install and secure to the chassis through ejectors.

Procedure

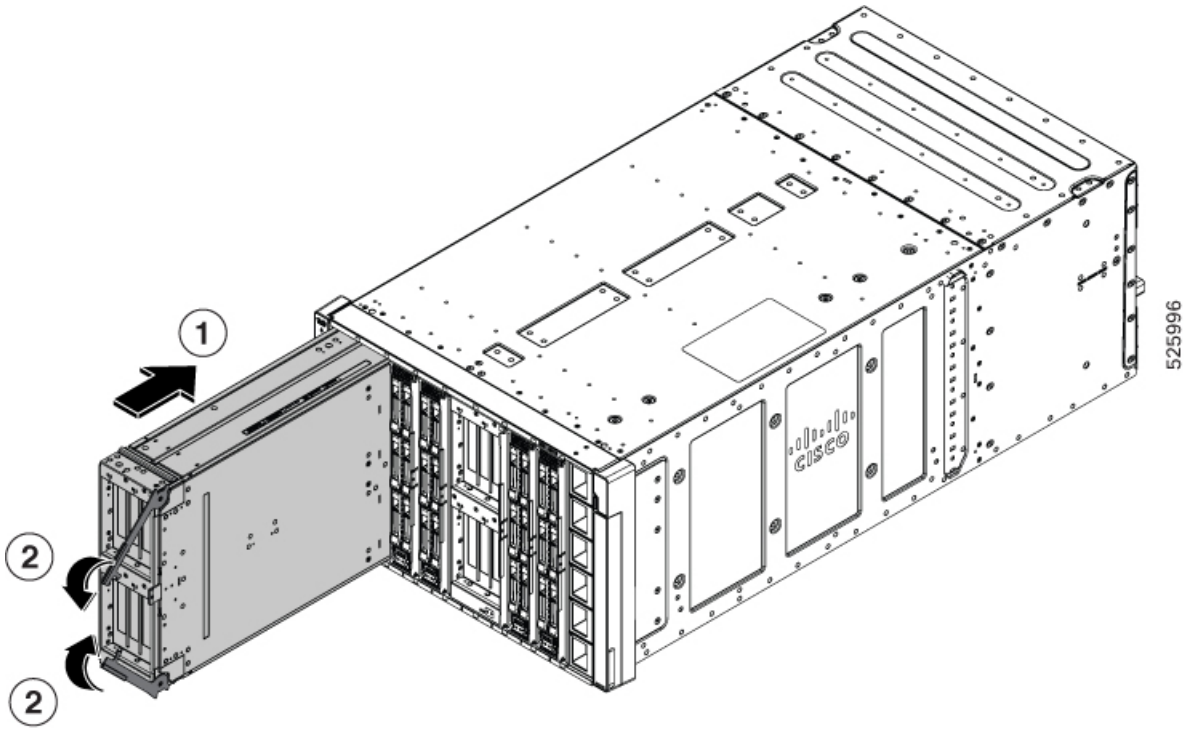
Step 1 Orient and align the PCIe node so that the ejectors are on the right vertical side of the slot.

Step 2 Using one hand to support the PCIe node, hold the node level and slide it into the chassis slot.

When the node is almost all the way in the slot, you might feel some resistance. This resistance is normal. It comes from the connector on the rear of the module meeting its socket on the inside of the chassis.

Step 3 Press on both ends of the faceplate to seat the module, then push the ejectors inward to secure the node in the slot.

The node is fully seated when the ejectors click into place and are flush with the node's faceplate.



Step 4 Using Cisco management software, such as Cisco Intersight, power on the PCIe node's paired M8 compute node.

Step 5 After the paired M8 compute node is completely powered on, use Cisco's management software, such as Cisco Intersight, to rediscover the PCIe node.

Removing the PCIe Node

After the M8 compute node and PCIe node are powered down, use this task to physically remove the PCIe node from the chassis.

Before you begin

You can remove each Cisco UCS X580p PCIe Node directly from the front of the chassis by using the node's ejectors. You do not need any tools to remove the node.

Note The PCIe node can be heavy. Use your other hand to support the module while you are removing it.

Caution The X580p PCIe Node itself is hot swappable, but the paired M8 compute node must be powered down first before hot swapping the PCIe node.

Before removing the PCIe node, you must properly decommission the M8 compute node(s) associated with the PCIe node. A brief summary follows, but make sure to follow the correct power off procedures as documented for your software management platform.

- Using a Cisco management tool, such as Cisco Intersight, gracefully power down M8 compute nodes that may have been connected to the PCIe node.
- Verify that all zones are deleted and associated M8 compute nodes are powered down. Doing so ensures that the GPUs are all powered off and can be safely removed and ensure no workloads are interrupted.
- Verify that the PCIe node's Health Status LED is blinking green, which indicates that the PCIe node is safe to remove.

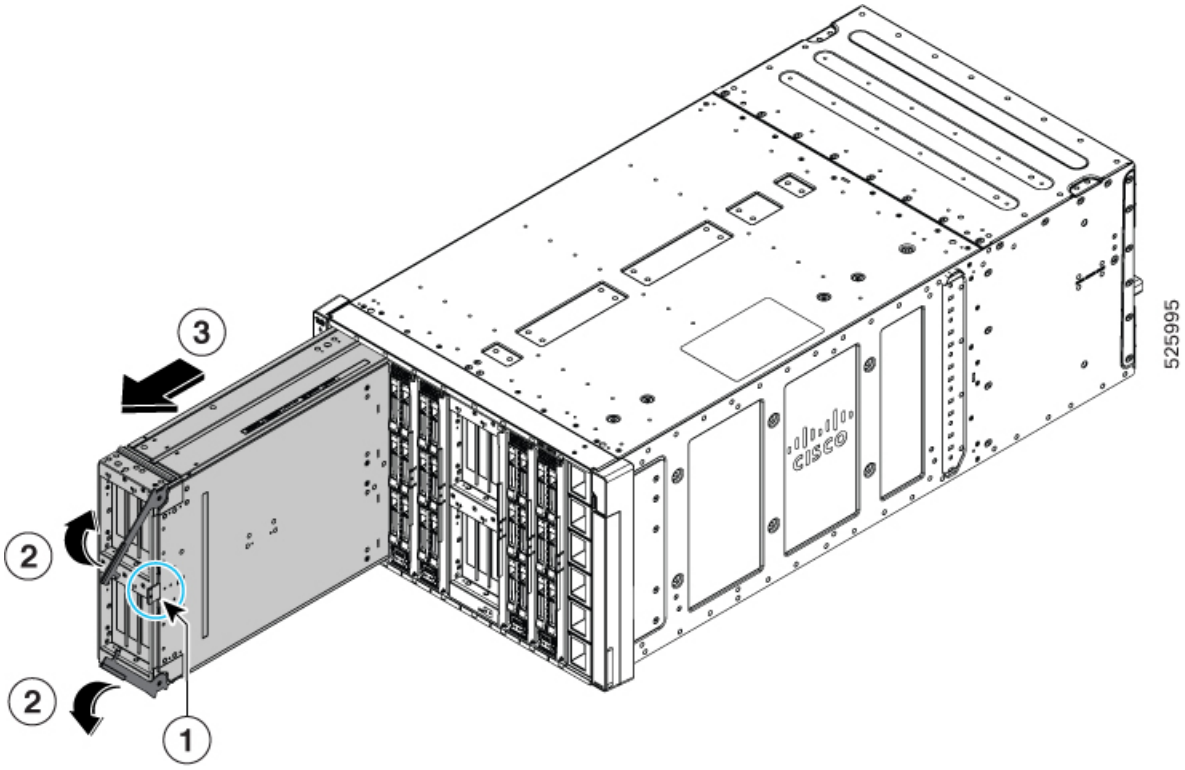
Procedure

Step 1 After the paired M8 compute node is completely powered off, push the PCIe node's ejector button (1) between the two ejectors to disconnect the ejectors from the sheet metal.

When you press the ejector button, the ejector handles will pop up slightly from the face of the node.

Step 2 Gently swing each ejector vertically to disconnect the PCIe node from the server (2).

Step 3 Using one hand to support the PCIe node, pull the node towards you to slide it out of the server chassis (3).

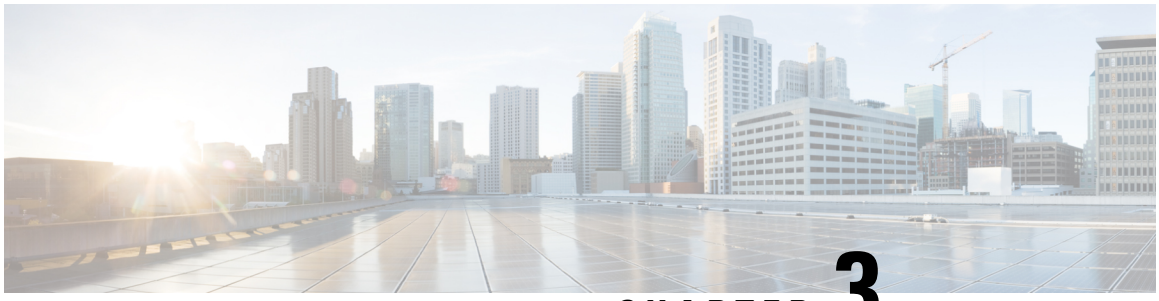


What to do next

Reinsert the PCIe node. See Installing the PCIe Node, on page 18.

PCIe Node Configuration

The Cisco UCS X580p PCIe Node can be configured and managed using the Cisco Intersight management platform. For details, see the *Cisco Intersight Managed Mode Configuration Guide*, which is available at: Cisco Intersight Managed Mode Configuration Guide.



CHAPTER 3

Servicing the Cisco UCS X580p PCIe Node

This chapter contains the following topics:

- Guidelines and Limitations, on page 23
- Serviceable Component Locations, on page 25
- Removing the PCIe Node Cover, on page 26
- Replacing the PCIe Node Air Baffle, on page 27
- Replacing the PCIe Node Front Cover, on page 32
- Replacing GPU Cages, on page 35
- Replacing a GPU Card, on page 48

Guidelines and Limitations

While handling or performing field-service procedures on the Cisco UCS X580p PCIe Node follow these general guidelines and limitations. Additional guidelines and limitations are documented throughout this document.

General Guidelines

Take note of the following general safety warnings:



Warning IMPORTANT SAFETY INSTRUCTIONS

Before you work on any equipment, be aware of the hazards involved with electrical circuitry and be familiar with standard practices for preventing accidents. Read the installation instructions before using, installing, or connecting the system to the power source. Use the statement number at the beginning of each warning statement to locate its translation in the translated safety warnings for this device.

SAVE THESE INSTRUCTIONS



Note You are strongly advised to read the safety instruction before using the product.

https://www.cisco.com/web/JP/techdoc/pldoc/pldoc.html

When installing the product, use the provided or designated connection cables/power cables/AC adapters.

〈製品使用における安全上の注意〉

www.cisco.com/web/JP/techdoc/index.html

接続ケーブル、電源コードセット、**AC**アダプタ、バッテリなどの部品は、必ず添付品または指定品をご使用ください。添付品・指定品以外をご使用になると故障や動作不良、火災の原因となります。また、電源コードセットは弊社が指定する製品以外の電気機器には使用できないためご注意ください。

Warning Blank faceplates and cover panels serve three important functions: they reduce the risk of electric shock and fire, they contain electromagnetic interference (EMI) that might disrupt other equipment, and they direct the flow of cooling air through the chassis. Do not operate the system unless all cards, faceplates, front covers, and rear covers are in place.

Warning To reduce risk of electric shock or fire, installation of the equipment must comply with local and national electrical codes.

Note An instructed person is someone who has been instructed and trained by a skilled person and takes the necessary precautions when working with equipment.

A skilled person or qualified personnel is someone who has training or experience in the equipment technology and understands potential hazards when working with equipment.

Warning Only a skilled person should be allowed to install, replace, or service this equipment. See statement 1089 for the definition of a skilled person.

Warning Ultimate disposal of this product should be handled according to all national laws and regulations.

PCIe Node Guidelines

- The Cisco UCS X580p PCIe Node is supported in the Cisco UCS X9508 chassis only. Do not attempt to install the PCIe node in any other UCS server chassis.

- Each Cisco UCS X580P PCIe Node must be paired with a Cisco X9516 X-Fabric Module, and therefore, has specific configurations based on the compute node.

  While it is possible that PCIe Gen 4 and Gen 5 hardware can interoperate in the same chassis, the devices will negotiate to the slower PCIe Gen 4 speeds, which will reduce performance.

- The Cisco UCS X580p PCIe Node supports 600W per GPU maximum.

- Hot removal or insertion of the PCIe node while the host is on is not supported. Before removing the PCIe node, you must properly decommission the M8 compute node(s) associated with the PCIe node.

  Make sure to follow the correct power off procedures as documented for your software management platform, but here is a brief example to help avoid interrupting any in-progress workloads served by the node:

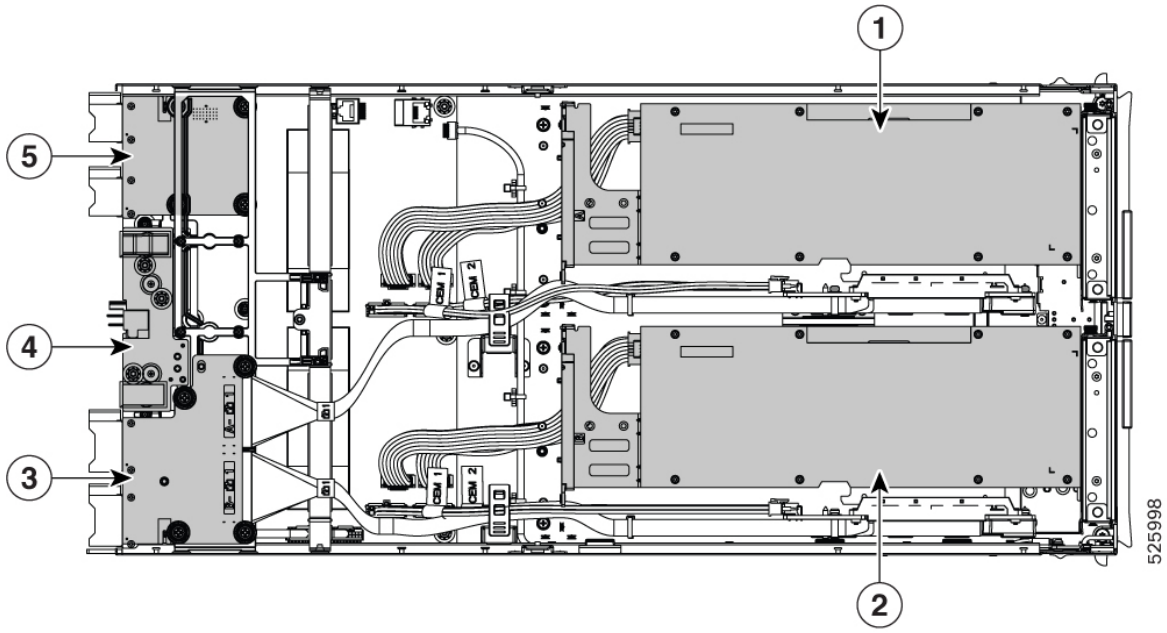
  - Using a Cisco management tool, such as Cisco Intersight, gracefully power down M8 compute nodes that may have been connected to the PCIe node.
  - Ensure that all zones are deleted and all connected nodes are powered down. Doing so ensures that the GPUs are all powered off and can be safely removed and ensure no workloads are interrupted.
  - Verify that the PCIe node's Health Status LED is blinking green, which indicates that the PCIe node is safe to remove. For information about the PCIe node's LEDs, see LEDs, on page 4.

Serviceable Component Locations

The following image shows the locations of serviceable components on the PCIe node.

Note The PCIe node has a heatsink next to the rear mezzanine (MEZZ) slot. This heatsink is not serviceable.

Serviceable Component Locations, PCIe Node



Note Items 3, 4, and 5 are shown for reference. Despite the presence of thumbscrews, none of these components are field-serviceable. Do not attempt to remove or replace them unless explicitly instructed by qualified Cisco personnel.

1	GPU Cage A, with FHFL dual-slot GPU shown	2	GPU Cage B, with FHFL dual-slot GPU shown
3	Rear Mezzanine Slot	4	Power Entry Board
5	mLOM slot	-	-

Removing the PCIe Node Cover

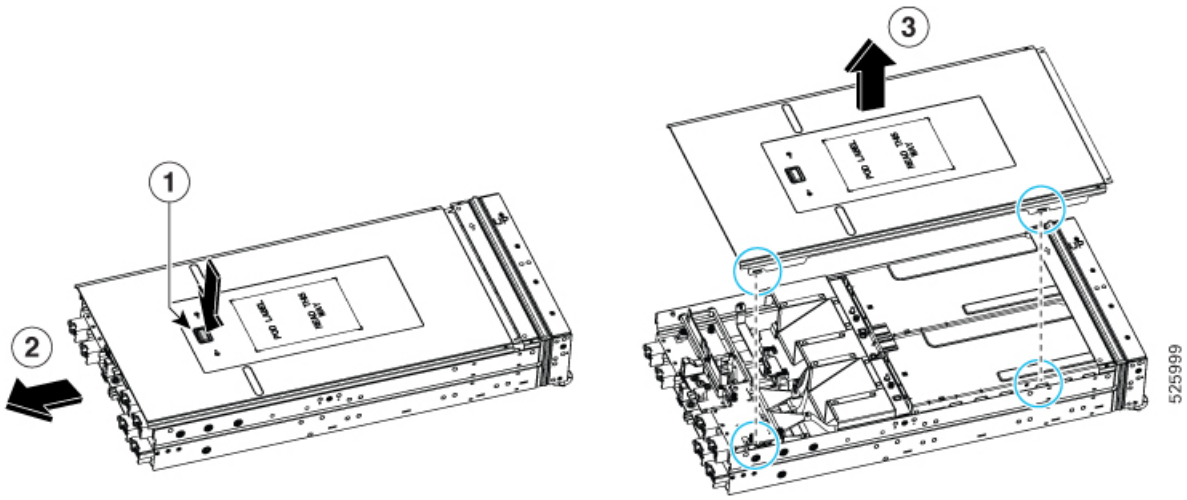
To remove the cover of the Cisco UCS X580p PCIe Node, follow these steps:

Procedure

Step 1 Press and hold the button down (1).

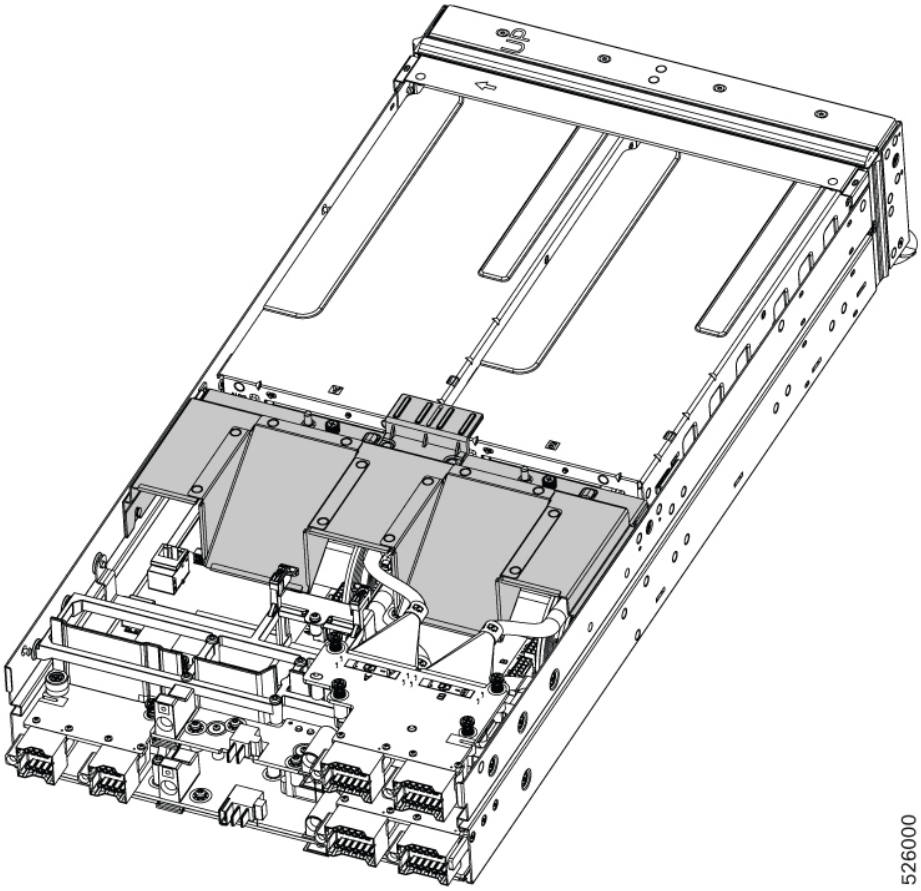
Step 2 While holding the back end of the cover, slide it back (2), then lift if off of the of node. (3).

By sliding the cover back, you enable the front edge to clear the metal lip on the rear of the PCIe node and enable the catch pins to release from the grooves on the top of the node's sidewalls.



Replacing the PCIe Node Air Baffle

The PCIe node has an air baffle that is located between the GPU cages and the node's rear mezzanine slots. The air baffle is one of the first components you see when you remove the PCIe node's top cover.



The node's air baffle optimizes the airflow from the cold aisle in the datacenter, across the node's components, and exhausts heated air into the hot aisle of the datacenter. When the PCIe node is operating, the air baffle must be installed and the node's top cover must be installed.

To replace the PCIe node's air baffle, use the following topics:

- Removing the PCIe Node Air Baffle, on page 28
- Installing the PCIe Node Air Baffle, on page 30

Removing the PCIe Node Air Baffle

The PCIe node's air baffle is a formed plastic component that rests behind the GPU cages. It optimizes airflow for node cooling and ventilation by channeling intake air from the cool aisle and exhausting hot air into the hot aisle.

The air baffle is installed by two captive thumbscrews that mount into threaded standoffs.

⚠

Caution The PCIe node air baffle must be in place, and the node's top cover must be installed for correct airflow. Do not operate the PCIe node without both components correctly installed.

Before you begin

Before attempting this procedure, gather the following tools:

- A #2 Phillips screwdriver.
- A T8 Torx screwdriver

Procedure

Step 1 If you have not already removed the node's top cover, do so now.

See Removing the PCIe Node Cover, on page 26.

Step 2 Loosen the air baffle's screws (1).

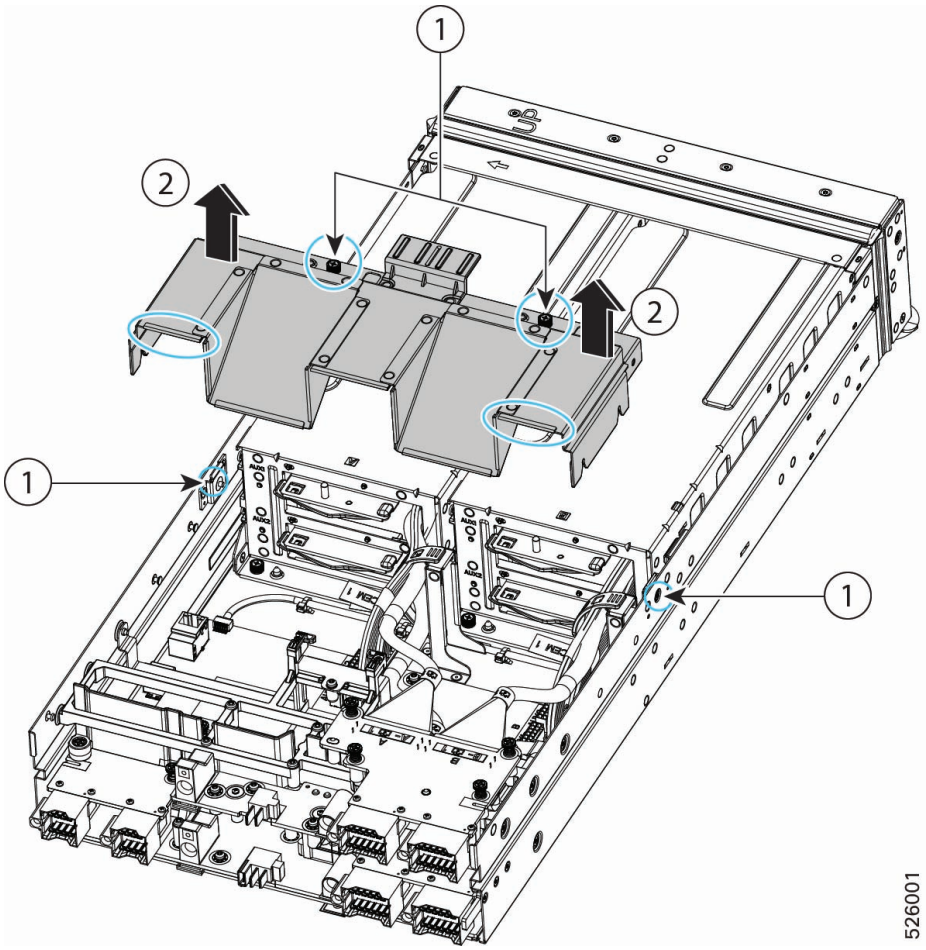
a) Using a T8 Torx screwdriver, remove the two screws, one on each side of the node.

The screws on the side of the node are not captive, so they can be fully removed, and possibly get lost. Put the screws somewhere safe. You will use them to reinstall the air baffle.

b) Using a #2 Phillips screwdriver, loosen the two captive thumbscrews on the top of the air baffle.

Step 3 Grasp the horizontal surfaces of the air baffle and slowly lift it straight up to remove it from the PCIe node (2).

Caution
Be careful when removing the air baffle! Cables are routed through it, and they might cause obstruction on the edges of the air baffle. If you feel any resistance while lifting the air baffle off of the node, check for and clear any obstruction.



Installing the PCIe Node Air Baffle

The PCIe node air baffle is required to ensure optimized air flow and ventilation across the node. You must replace the air baffle after completing any field-service tasks and before installing the node's top cover.

When the PCIe node is shipped, the air baffle is pre-installed at the factory. Use the following task to install the node's air baffle when needed.

Before you begin

Before attempting this procedure, gather the following tools:

- A #2 Phillips screwdriver
- A T8 Torx screwdriver

Procedure

Step 1 Orient the air baffle so that the captive screws are facing the front of the node.

Step 2 Notice the alignment features, which consist of threaded standoffs for thumbscrews and cutouts for the catch pins on the node (1).

Step 3 Grasp the air baffle by the horizontal edges.

Step 4 Install the air baffle.

a) Lower the air baffle onto the node, making sure to align the air baffle's thumbscrews with their threaded standoffs on the GPU cages.

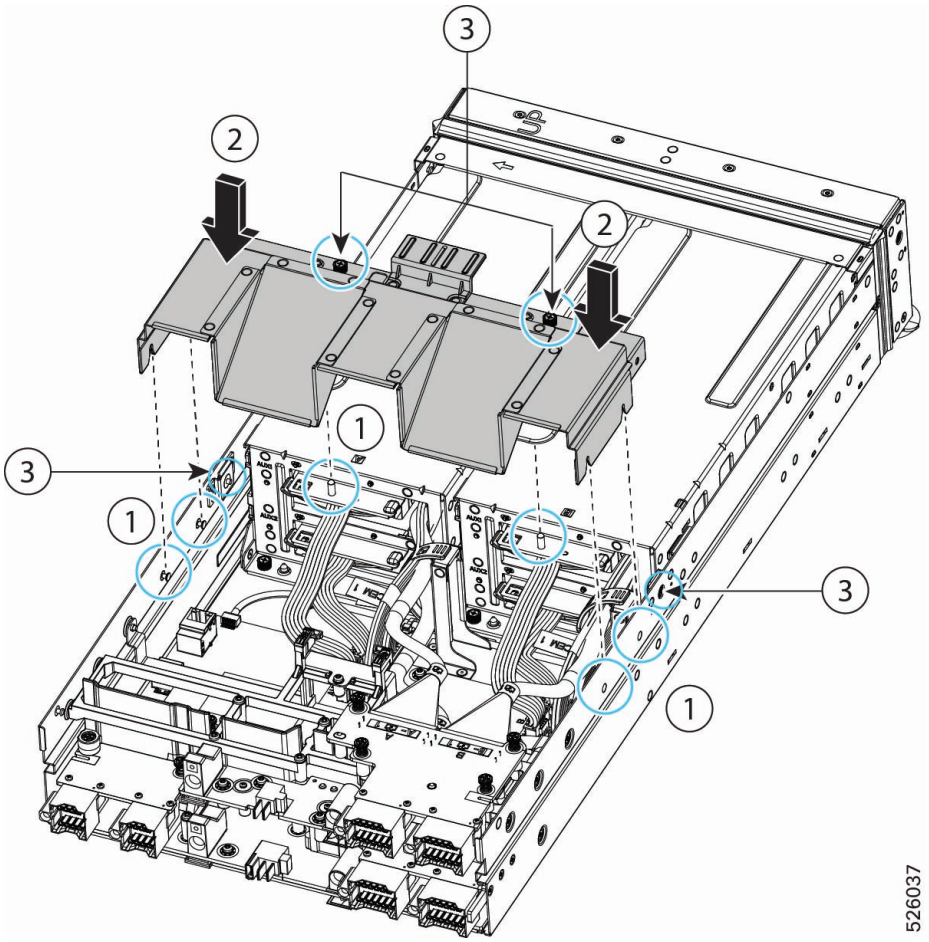
b) When the air baffle is in place, verify that the cutouts on the bottom sides of the baffle are correctly seated in the catch pins on the sheet metal sidewalls of the node.

Step 5 If the baffle is not properly seated, repeat this process until it is.

Step 6 Secure the air baffle to the node.

a) Using a #2 Phillips screwdriver, tighten the two captive screws on top of the air baffle (3).

b) Using a T8 Torx screwdriver, insert and tighten the remaining two screws, one on each of the node's sidewalls (3).



Replacing the PCIe Node Front Cover

The PCIe node front cover is a rectangular metal frame that rests on top of the node's faceplate and provides the beveled surface that the top cover slides into.

The front cover accepts seven mounting screws that join to, and attach, the node's faceplate. The front cover is accessible only when the top cover is removed. The node front cover must be removed as part of removing or replacing either of the GPU cages.

Use the following tasks to replace the PCIe node front cover.

- Removing the PCIe Node Front Cover, on page 32
- Installing the PCIe Node Front Cover, on page 34

Removing the PCIe Node Front Cover

The front cover is secured to the PCIe node by seven screws.

- Five T8 Torx screws, four across the top of the front cover and one at the top of the middle sheet metal stiffener on the node's faceplate.
- Two T8 Torx captive thumbscrews, one on each vertical side of the chassis near the top of the node.

Before you begin

Before attempting this procedure, gather a T8 Torx screwdriver.

Procedure

Step 1 If you have not already removed the node's top cover, do so now.

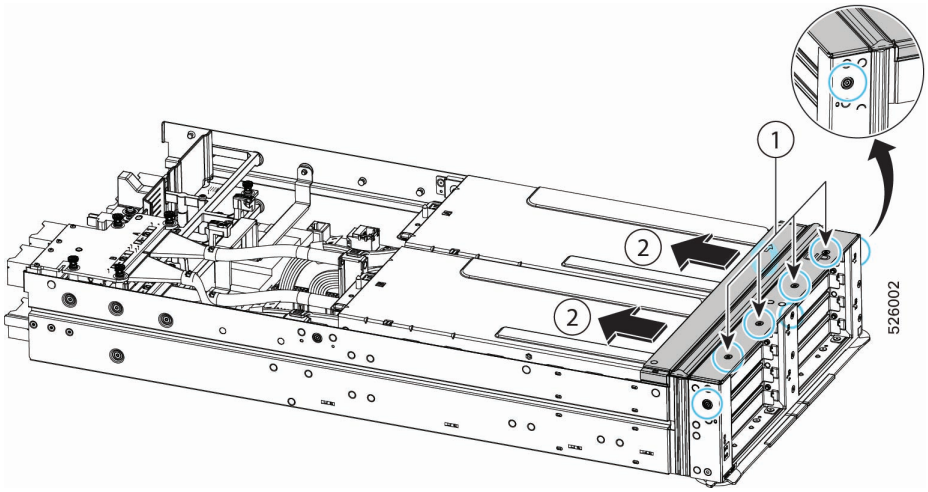
See Removing the PCIe Node Cover, on page 26.

Step 2 Disconnect the PCIe node.

a) Using the screwdriver, loosen all seven screws.

Note
Remove the five screws on the top and front before removing the two thumbscrews on the sides. This order of screw removal is not required, but it is a best practice.

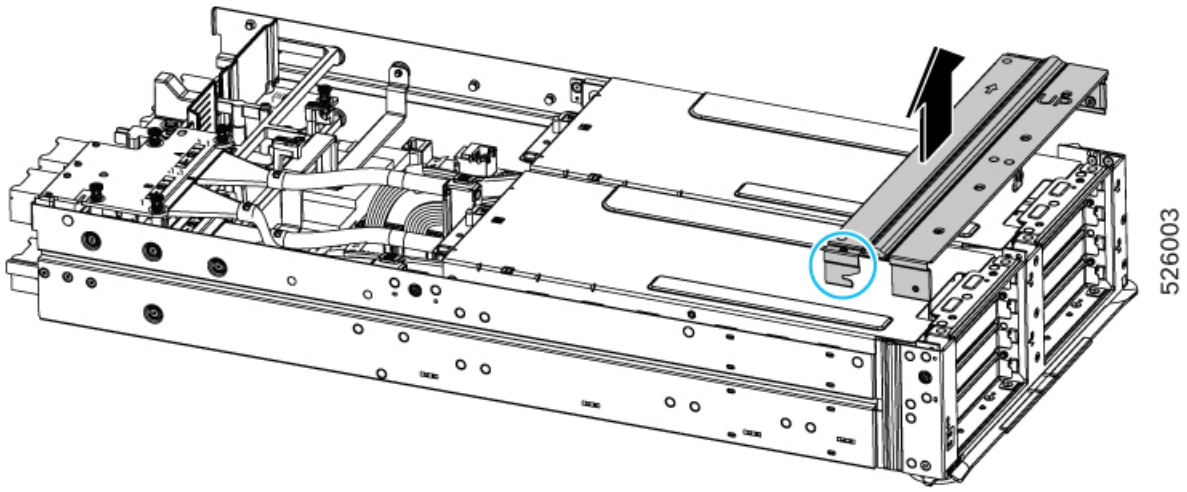
b) Keeping the front cover level, gently slide it toward the back of the node to disconnect it from the node.



Step 3 Detach the front cover from the node.

a) While sliding the front cover back, check each side to verify that notch on the front cover clears its catch pin on the node's side wall.

b) When the front cover has released from the catch pin, lift the front cover up to remove it from the node.



Installing the PCIe Node Front Cover

To install the node front cover, you will slide the front cover into place, then reattach the screws.

Before you begin

Before attempting this procedure, gather a T8 Torx screwdriver.

To facilitate proper orientation, the top of the front cover has the word UP printed on it.

Procedure

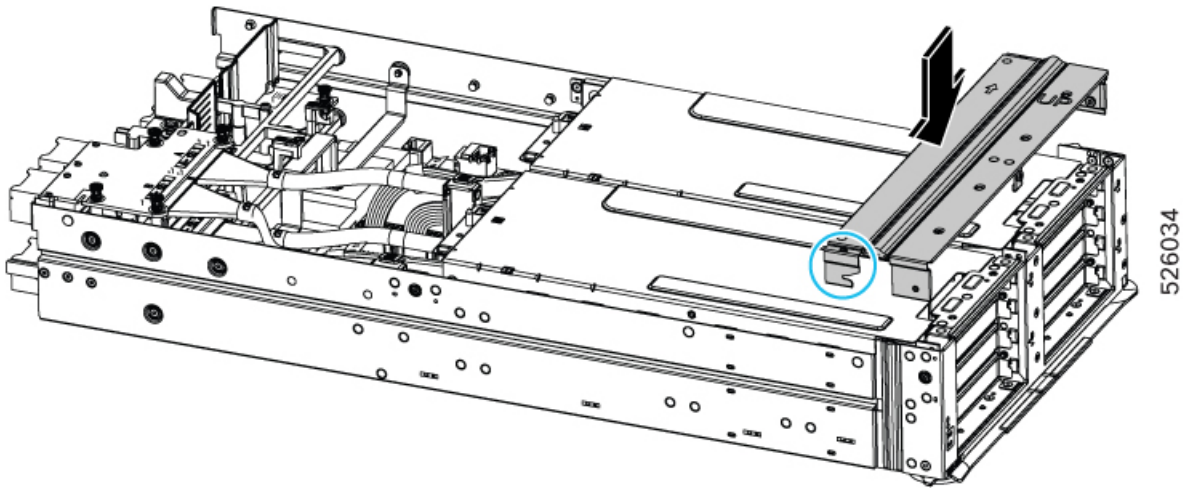
Step 1 Attach the front cover to the node.

a) Orient the front cover correctly.

Make sure the word UP on the front cover is facing up.

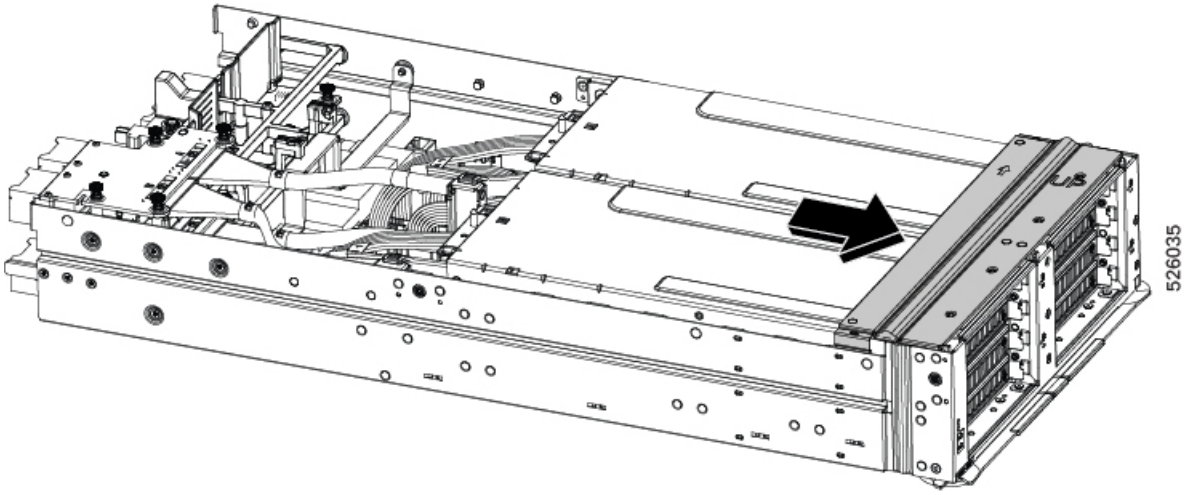
b) Note the location of the notches on each side of the front cover.

c) Lower the front cover onto the node and make sure that the cover is level.



Step 2 Connect the PCIe node.

a) Gently slide the front cover toward the front of the node, while checking each side to verify that the notch on the front cover engages its catch pin on the node's side walls.



b) Using the screwdriver, tighten all seven screws, four across the top, one on each side, and one in the front.

Note
Install the two thumbscrews on the sides before installing screws on the top and front. Although this order of screw installation is not required, it is a best practice.

Replacing GPU Cages

Each PCIe node has two GPU cages, which mount to the PCIe node sheet metal and contain the GPUs. One type of GPU cage is supported that accepts FHFL dual-slot GPUs. GPU cages rest directly on the PCIe node in two slots.

GPU cages are interchangeable, so you can install a GPU cage in either slot A or slot B. Even though GPU cages are interchangeable, it is a best practice to replace the cages in the same slots where they were originally located. Replacing the cages in the same location (cage A replaced in cage slot A and cage B replaced in cage slot B) facilitates reassembly especially related to cable routing.

Note The PCIe node's GPU cages support PCIe Gen 5 connectivity. Cisco offers similar GPU cages for Gen 4 connectivity through a similar product, the Cisco UCS X440p PCIe Node. Gen 4 and Gen 5 GPU cages are not interchangeable, so do not attempt to reuse the GPU cages across these two different products.

PCIe cages are not offered as a spare part. If you need a new PCI cage, contact Cisco to start an RMA for the node.

To replace GPU cages, use the following tasks:

- Removing a GPU Cage, on page 37
- Installing a GPU Cage, on page 43

Cable Reference

GPU cages, and the GPUs themselves, use the following cables. The following table summarizes which cables are used by the GPU cages or GPUs themselves as well as where the cables are connected.

Cables	Details
MCIO, 4 per node (two per GPU cage)	Each GPU cage has two MCIO cables. One cable connects the GPU cage to the upper mezzanine card, and one cable connects the GPU cage to the lower mezzanine card.
AUX, up to 4 (one per GPU)	Each GPU has one AUX cable that supplies power to the GPU. One end of the cable connects to the GPU itself, and the other end connects to either of the AUX1 or AUX2 connectors on the node's motherboard.
CEM, 4 per node (two per GPU cage)	Each GPU cage has two cables that supply power to the cage. Both cables connect to the GPU cage at one end, and the other ends connect to the CEM1 and CEM2 connectors on the node's motherboard.

GPU Cage Population Guidelines

Each PCIe node has two GPU cages: Cage A and Cage B. GPU cage identifiers are labeled on the PCIe node so that they are visible when the GPU cages are removed.

Each cage supports a maximum of two FHFL dual-slot GPUs for a total of four GPU slots. Slot identifiers are labeled on the front panel of the node.

When removing and installing GPU cages, be aware of the following:

- Always populate Cage A before Cage B. If only one cage needs to be populated, it must be Cage A.
- When populating GPU cages, always do so in ascending numerical order.
  - Populate Cage A slot 1 before Cage A slot 2

- Populate Cage B slot 3 before Cage B slot 4.

- Any GPU cage can accept any of the supported GPUs. For example, Cage A is not "reserved" for only H200 GPUs.
- Be aware of the following GPU mixing consideration: GPUs can be mixed across the node, but not within GPU Cages. For example, H200 NVL GPUs can be installed in Cage A, and RTX PRO 6000 GPUs can be installed in Cage B. But, you cannot install a mix of H200 and RTX PRO 6000 GPUs in the same cage.
- If your deployment requires multiple types of GPUs, install the same types in the same cage.

Removing a GPU Cage

Each PCIe node contains two PCIe GPU cages of the same type. Use this procedure to remove a GPU cage.

Procedure

Step 1 Remove the PCIe node from the server.

See Removing the PCIe Node, on page 19.

Step 2 Remove the top cover.

See Removing the PCIe Node Cover, on page 26.

Step 3 Remove the node's air baffle.

See Removing the PCIe Node Air Baffle, on page 28.

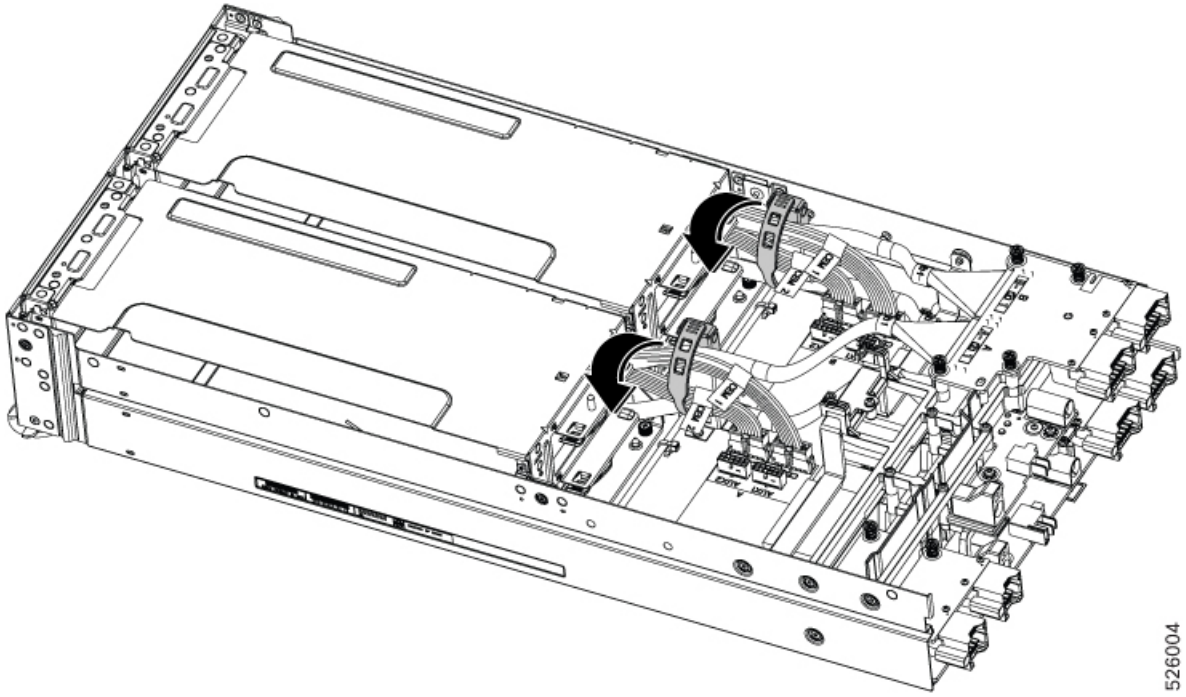
Step 4 Remove the front cover.

See Removing the PCIe Node Front Cover, on page 32

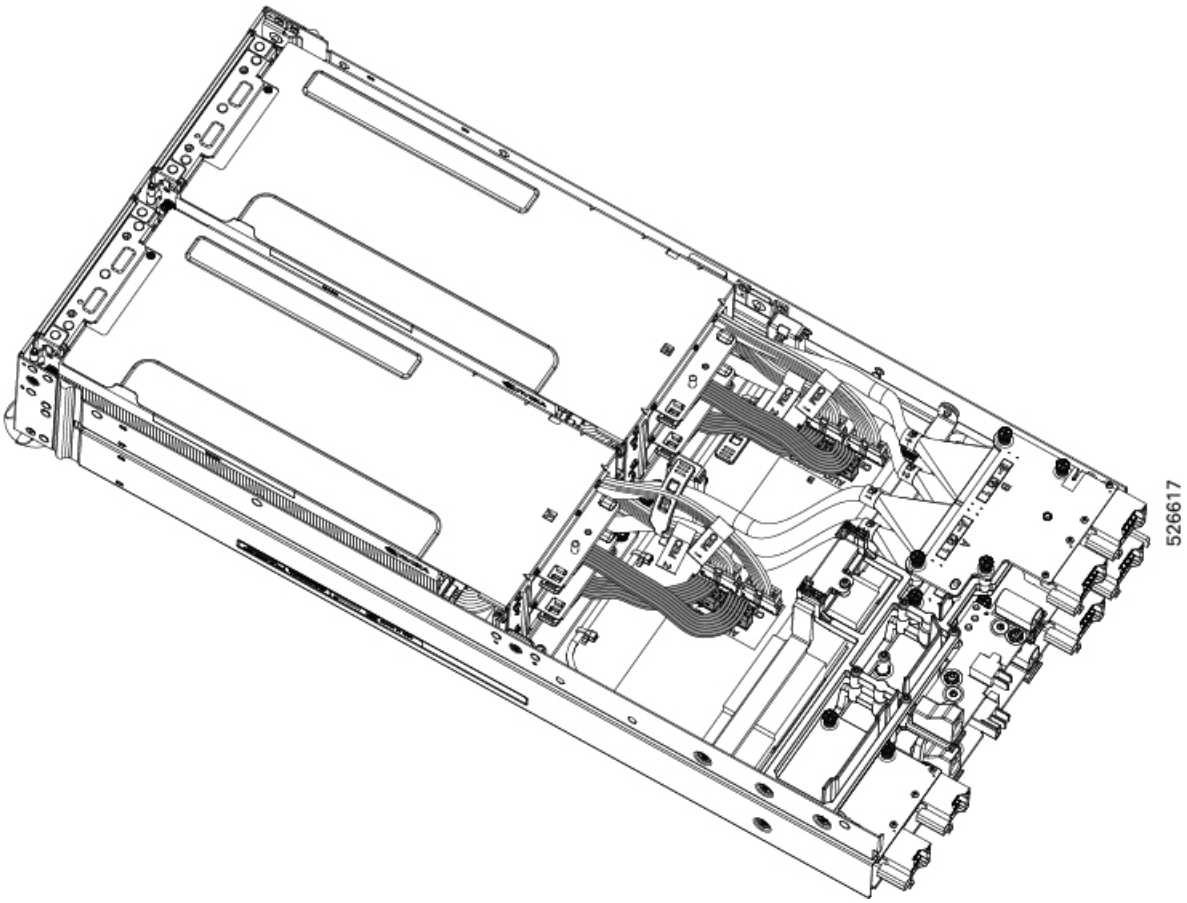
Step 5 Disconnect the cables.

a) Grasp the rubber cable retainers and unhook them from the metal tab that secures them.

When the cable retainers are unhooked, cables will become unbundled. You will find it helpful to flatten the retainers so that they lie flat on the motherboard to prevent obstruction or snagging on other parts.

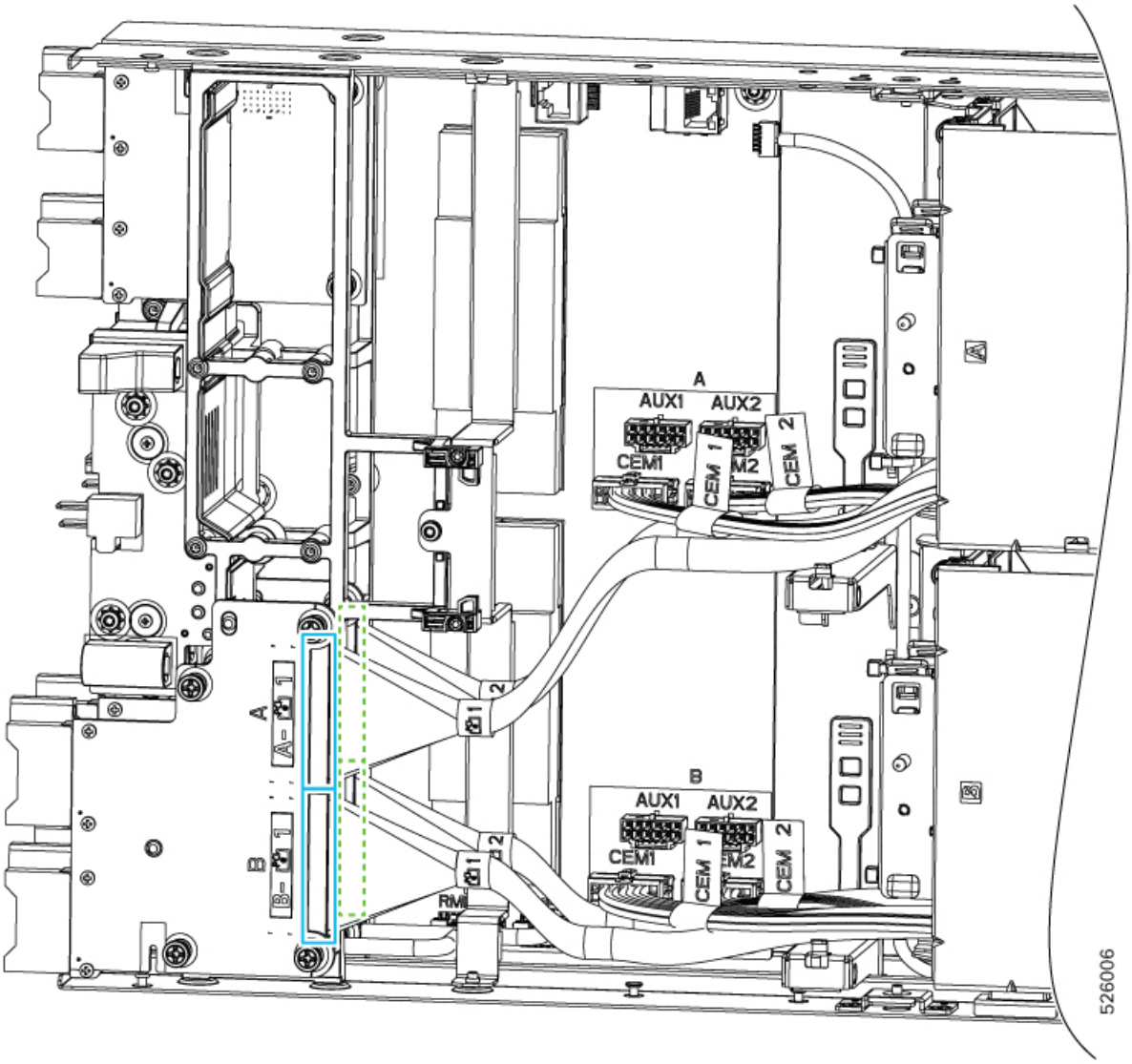


b) Grasp the AUX cables and disconnect them from the AUX1 and AUX2 connectors on the node motherboard.

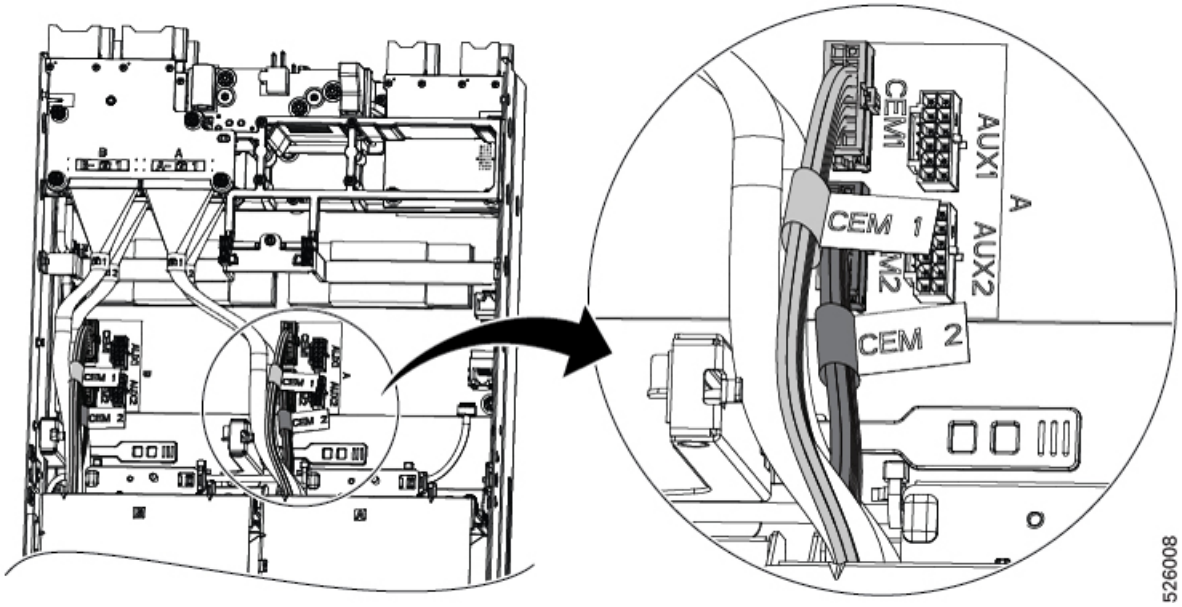


c) Disconnect the MCIO cables from the rear mezzanine (MEZZ) connector.

Note
MCIO cable connectors are stacked vertically on the rear mezz slot. If you're removing both GPU cages at the same time, you'll find it helpful to disconnect the top cable(s) first. Doing so allows easier access to the bottom cables.

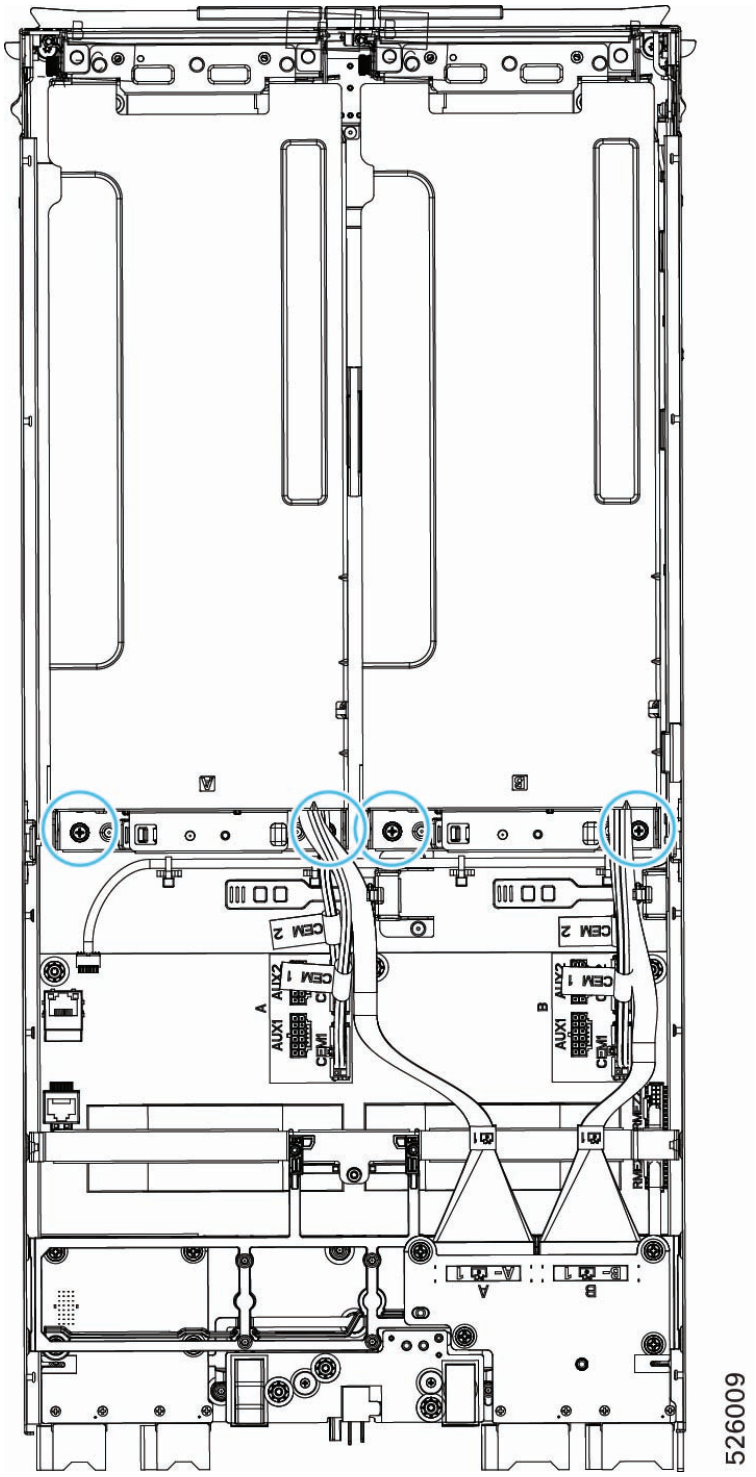


d) Disconnect the power cables from the node's CEM1 and CEM2 motherboard connectors.



Step 6 When MCIO and power cables are disconnected, detach the cages.

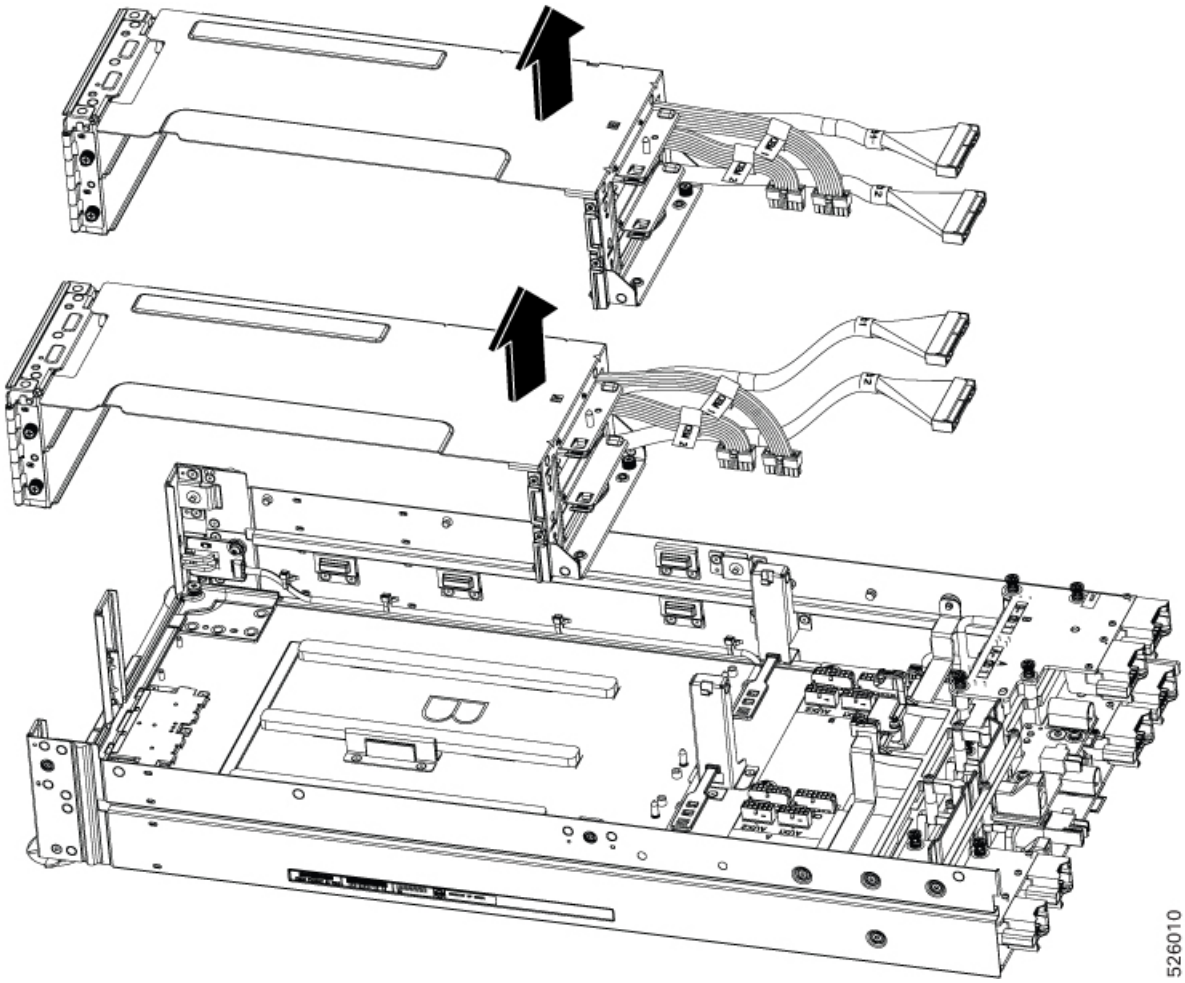
a) Using a #2 Phillips screwdriver, remove the captive screws that secure the cages to the PCIe node's base assembly.



b) When the thumbscrews are removed, grasp the cages and lift them off of the PCIe node.

Caution

If you feel any resistance while removing the cages, reseat the cage and verify that all screws are loose and check for obstructions, such as cables or cable retainers that might be snagged on other components. Proceed when the resistance is no longer present.



Installing a GPU Cage

Use this procedure to install a GPU cage onto the PCIe node.

Before you begin

Before attempting this procedure, gather a #2 Phillips screwdriver.

Procedure

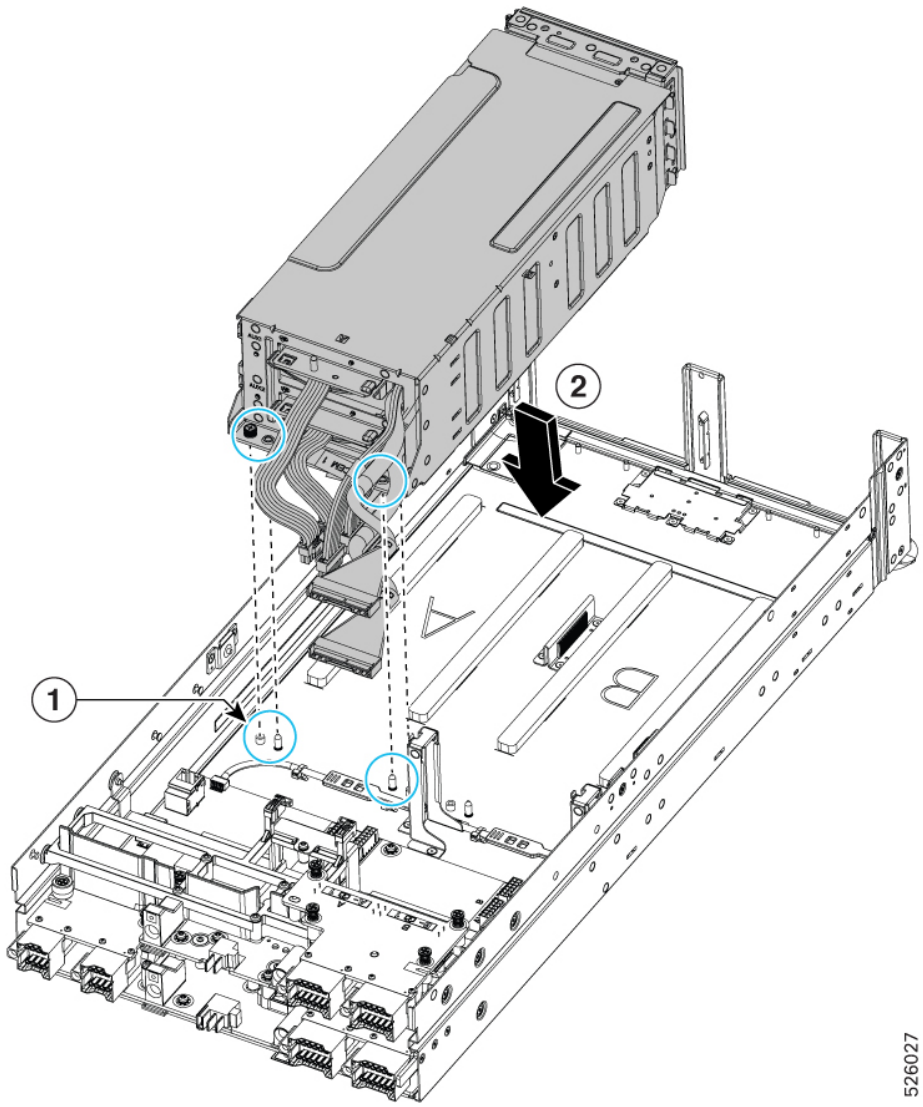
Step 1 Orient the GPU cage with the PCIe node so that the MCIO cables are facing the rear mezzanine module.

Step 2 Attach the GPU cage to the PCIe node.

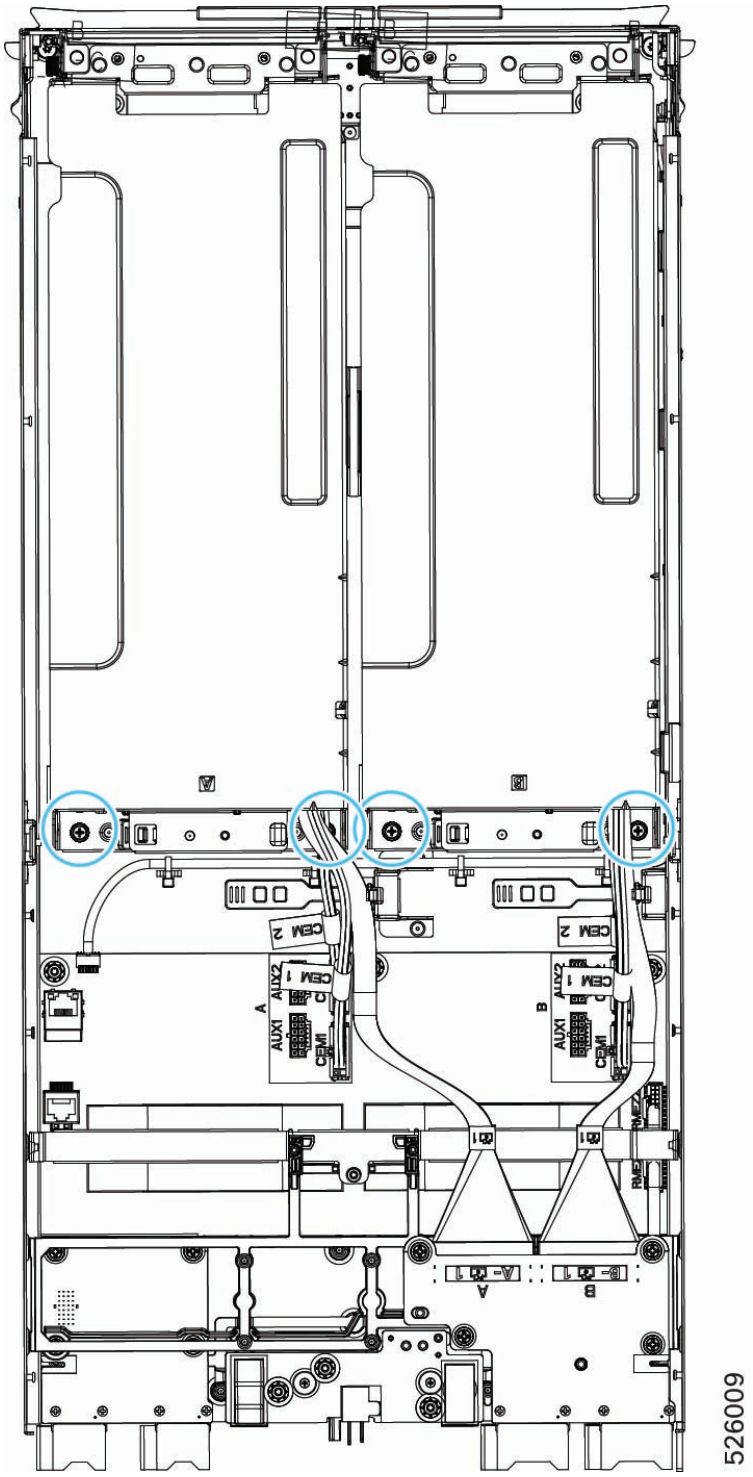
a) Before lowering the GPU cage onto the node, observe the two guide holes on the cage and their respective guide pins on the PCIe module and the thumbscrews on the cage and their screw holes on the node (1).

b) Lower the GPU cage onto the PCIe module (2), making sure that the guide holes on the cage align with the guide pins on the node, and the thumbscrews on the cage align with the screw holes on the node.

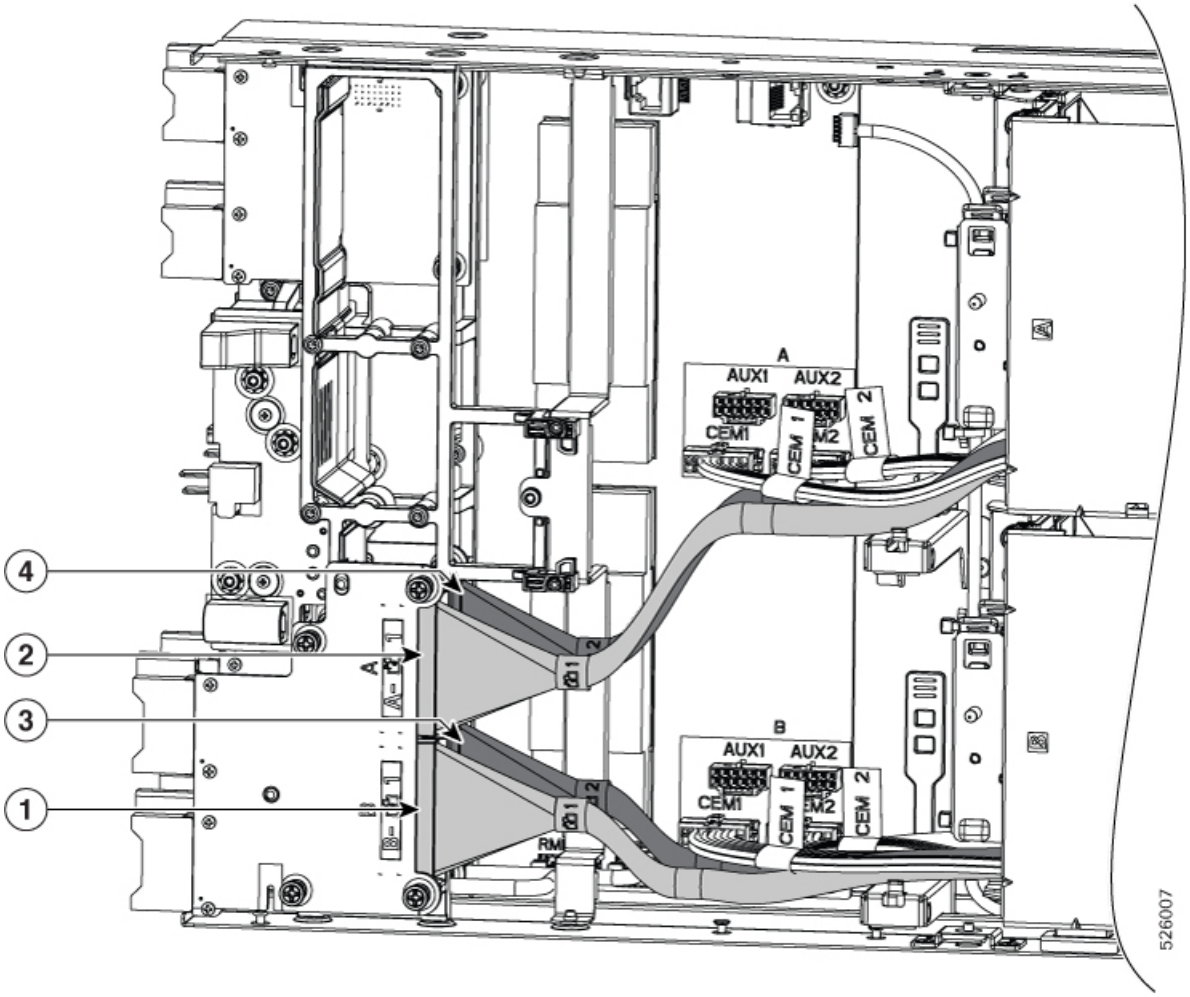
Caution
While installing the GPU cage, make sure that the cables are not pinched between the GPU cage and the node.



c) Using a #2 Phillips screwdriver, tighten the captive thumbscrews to secure the GPU cage to the node.

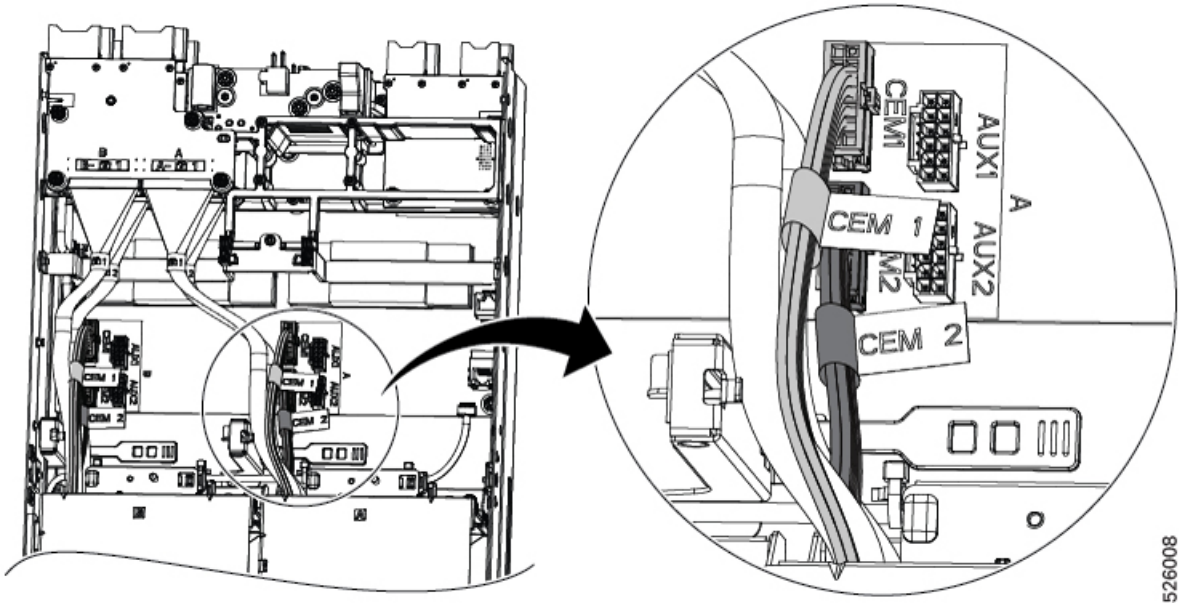


d) Connect the MCIO cables to the rear mezzanine (MEZZ) connector.

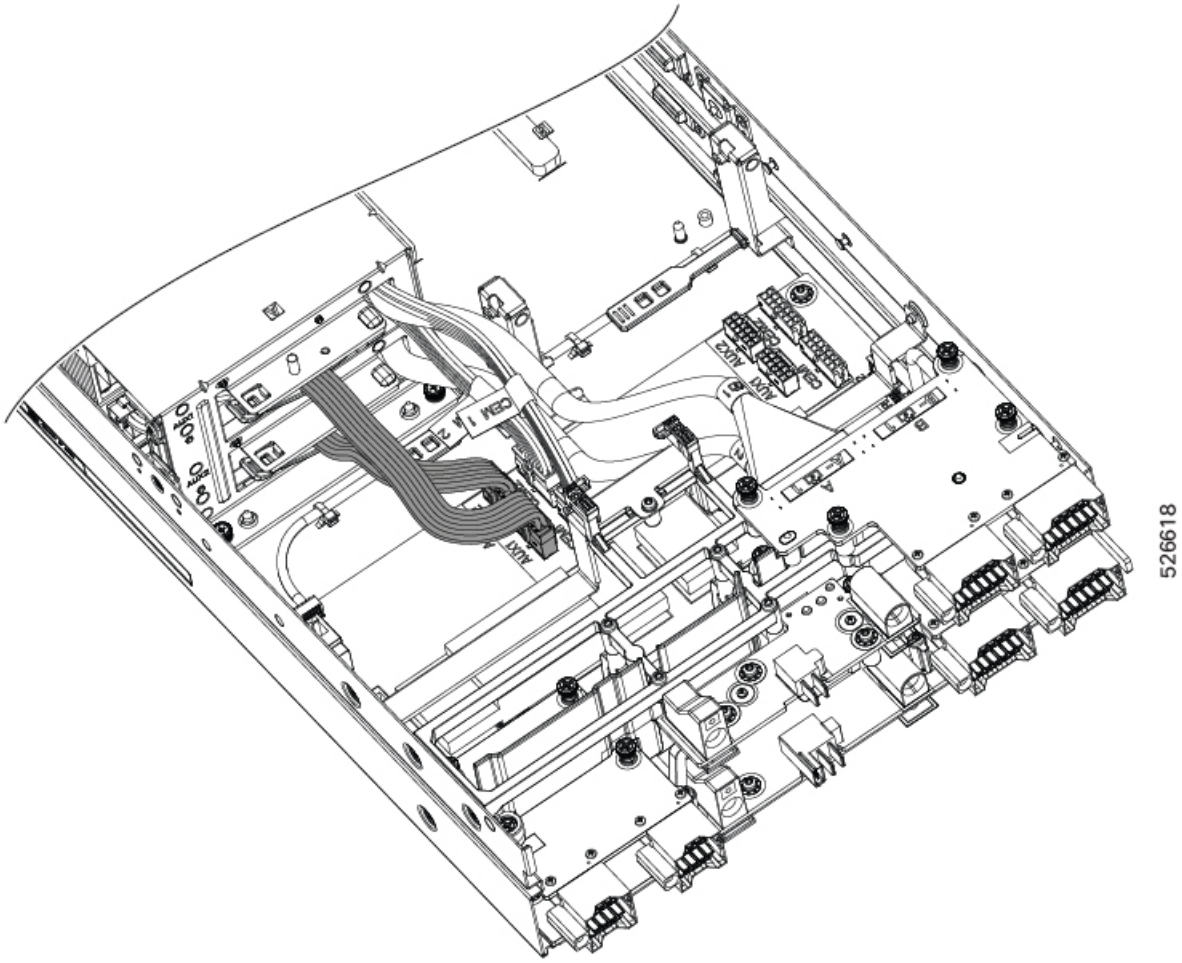


e) Connect the power cables to the node's CEM1 and CEM2 motherboard connectors.

Cables connect directly behind their cage, and cable are labeled to identify which cables connect to which plug on the motherboard.



f) Connect the AUX cables to the AUX1 and AUX2 connectors on the node motherboard.



Replacing a GPU Card

Supported GPU cards are contained in the GPU cage during normal operation. For more information, see GPU Cage Options, on page 5.

The PCIe node supports only full height, full length (FHFL) dual-slot GPUs. For more information, see Supported GPUs, on page 9.

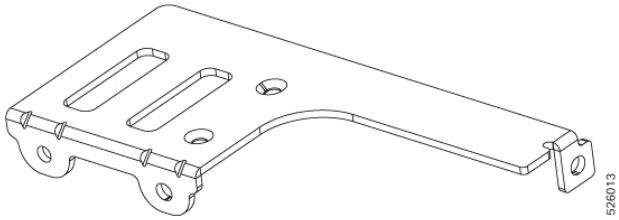
To replace a GPU card, use the following tasks:

- GPU Replacement Guidelines and Limitations, on page 49
- Removing a FHFL GPU, on page 52
- Installing a FHFL GPU, on page 54

GPU Replacement Guidelines and Limitations

Be aware of the following guidelines and limitations for replacing FHFL dual-slot GPUs.

- For installing GPUs into a GPU cage
  - It is a best practice to populate slots in ascending order, for example, slot 1 before slot 2 in the same GPU cage. For information about how slots are numbered in each GPU cage, see Slot Numbering, on page 5.
  - Completely populate each GPU cage before starting to populate another.
- Cisco offers Cisco PCIe blanks (UCSX-PCIF-GPU=) that fill unused GPU slots. If your PCIe node is not fully configured with GPUs, you must install the appropriate number of blanks. For PCIe nodes that are less than fully configured with GPUs, do not operate the PCIe node without a GPU filler blank.
- For NVIDIA H200 NVL GPUs, it is possible that existing GPUs have a legacy straight extender bracket that is currently incompatible with the GPU cages. In this case, Cisco offers an enhanced extender bracket (EXT-X580P-GPU-N=) to provide optimal physical support for the GPU.



You must replace the legacy extender bracket with the Cisco enhanced extender bracket. Bracket replacement must be completed before installing, or reinstalling, the H200 NVL GPU. For more information, see Replacing a Legacy GPU Extender Bracket, on page 49

Replacing a Legacy GPU Extender Bracket

All Cisco-provided GPUs have the correct GPU bracket installed. However, some deployments might use GPUs that were acquired from other vendors than Cisco.

For GPUs acquired from other sources, it is possible that the NVIDIA H200 NVL GPUs are assembled with a legacy straight extender bracket. This straight bracket must be replaced to use H200 NVL GPUs with the PCIe node.

To obtain an enhanced extender bracket (Cisco PID: EXT-X580P-GPU-N=) as the supported replacement part, contact Cisco and order the enhanced extender bracket by PID.

After receiving the enhanced extender bracket, use the following tasks to replace the legacy bracket with the new bracket.

The new bracket must be installed onto the H200 NVL GPU before the GPU is installed into the PCIe node's GPU cage.

- Removing the Legacy Extender Bracket, on page 50
- Installing the Enhanced Extender Bracket, on page 51

Removing the Legacy Extender Bracket

For NVIDIA H200 NVL GPUs only, this procedure enables the GPU to properly fit the GPU cages. For more information, see GPU Replacement Guidelines and Limitations, on page 49.

Use the following task to remove the legacy straight extender bracket so that you can replace it with the supported Cisco enhanced extender bracket.

Before you begin

Before attempting this procedure, gather a T15 Torx screwdriver.

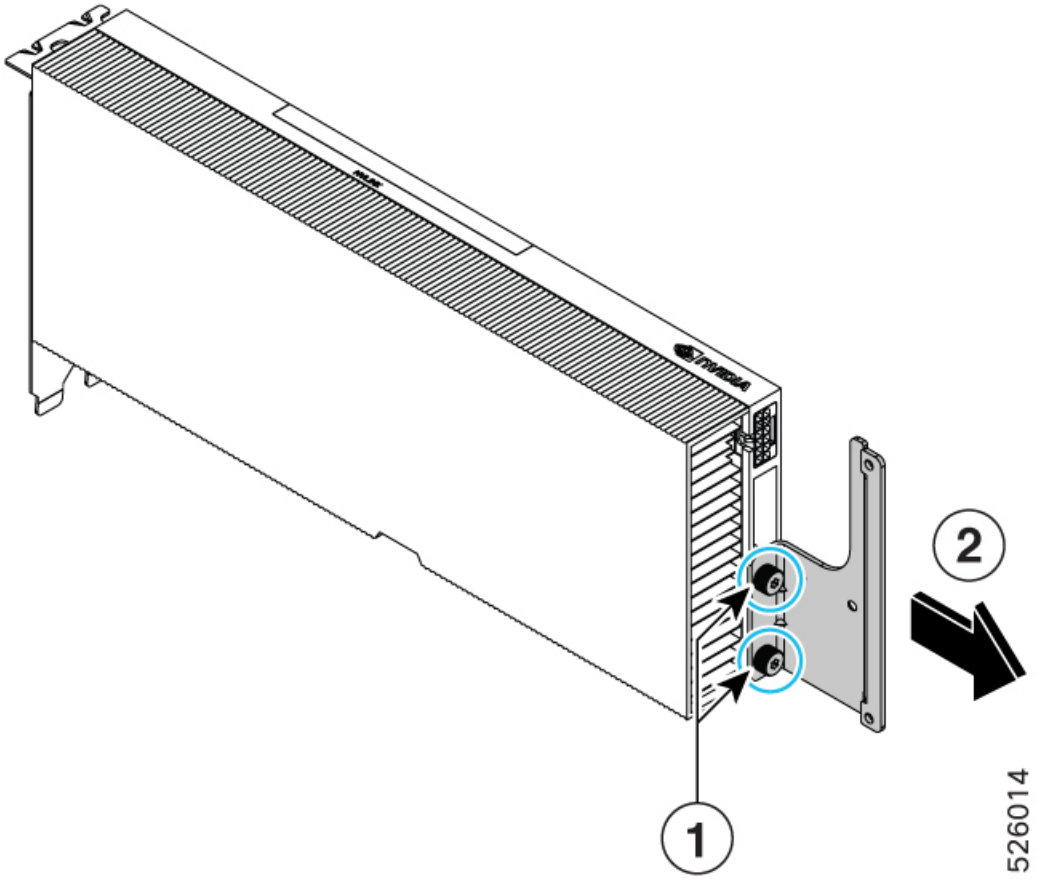
Procedure

Step 1 Remove the existing NVIDIA H200 NVL GPU.

See Removing a FHFL GPU, on page 52.

Step 2 Using the screwdriver, remove the two screws that secure the legacy straight extender bracket.

Step 3 Detach the bracket from the GPU.



What to do next

Connect the Cisco replacement bracket. See Installing the Enhanced Extender Bracket, on page 51.

Installing the Enhanced Extender Bracket

For NVIDIA H200 NVL GPUs only, this procedure enables the GPU to properly fit the PCIe node cages. For more information, see GPU Replacement Guidelines and Limitations, on page 49.

Use the following task to install the Cisco enhanced extender bracket before installing an H200 NVL GPU into the PCI node's GPU cages.

Before you begin

Before attempting this procedure, gather a T15 Torx screwdriver.

To obtain an enhanced extender bracket (Cisco PID: EXT-X580P-GPU-N=), contact Cisco and order the enhanced extender bracket by PID.

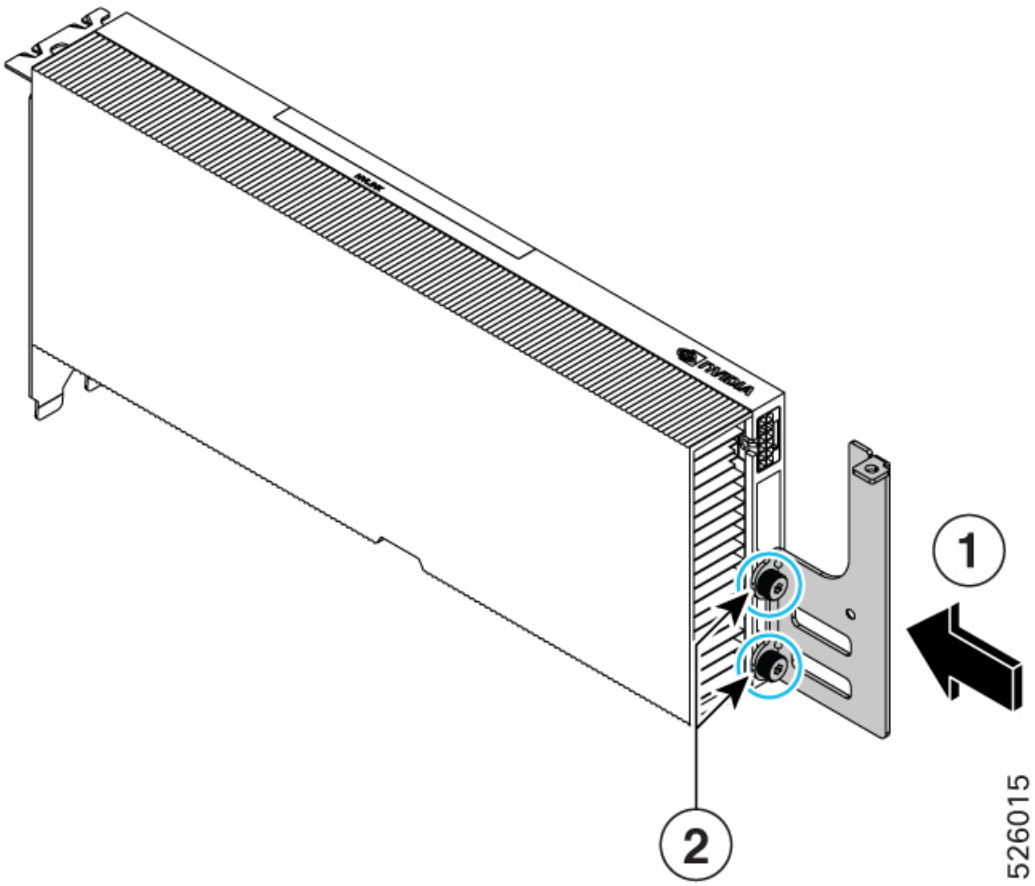
Procedure

Step 1 If you have not already removed the existing H200 NVL GPU, remove it now.

See Removing a FHFL GPU, on page 52.

Step 2 Attach the Cisco enhanced extender bracket to the GPU.

Step 3 Using the screwdriver, reinsert the two screws that you previously removed.



Step 4 Reinstall the GPU.

See Installing a FHFL GPU, on page 54.

Removing a FHFL GPU

Full height, full length (FHFL) dual-slot GPUs are supported in either GPU cage. Each GPU cage has two PCIe sockets, each of which can accept one GPU for a maximum of two GPUs in each cage.

Use this task to remove a FHFL GPU.

Before you begin

Before attempting this procedure, gather a #2 Phillips and a T8 Torx screwdriver.

Procedure

Step 1 If you have not already done so remove the GPU cage from the PCIe node.

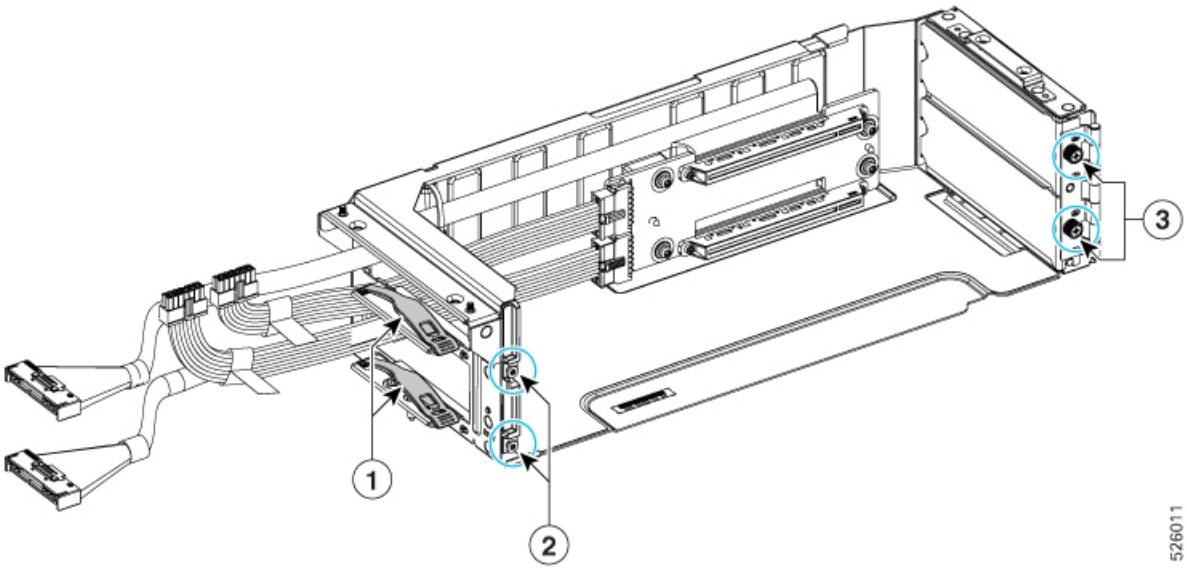
See Removing a GPU Cage, on page 37.

Step 2 Grasp the rubber cable retainer and unhook it so that the GPU cables are unbundled (1).

Step 3 Using the T8 screwdriver remove the securing screws that hold the GPU in the slot (2).

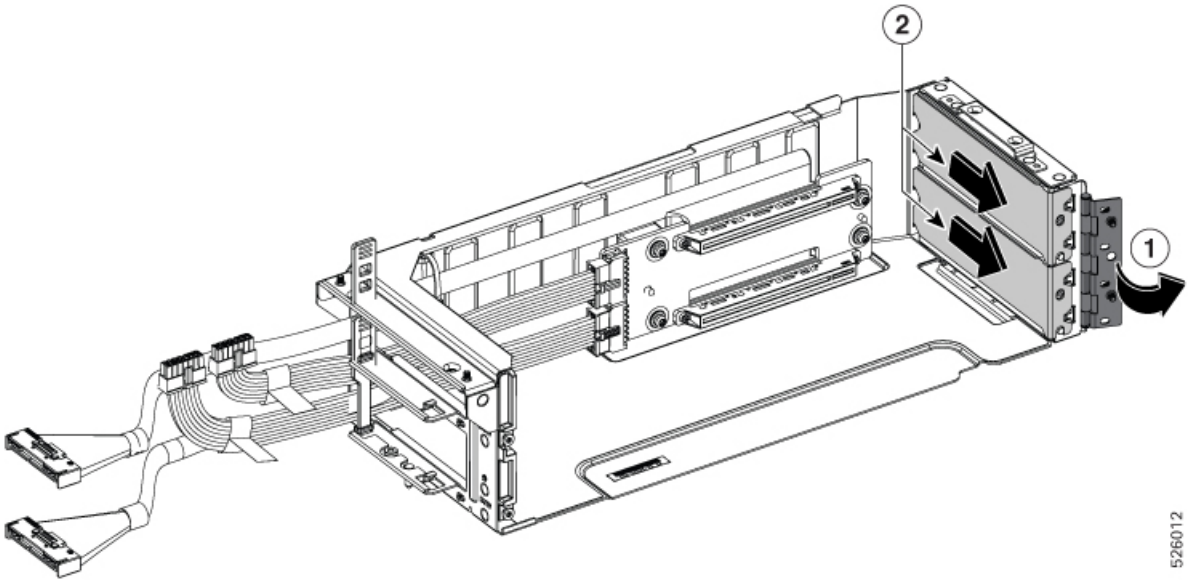
Step 4 Disconnect the GPU AUX power cable from the GPU.

Step 5 Using the #2 Phillips screwdriver, loosen the thumbscrews at the front of the GPU cage, which is the end opposite the cables (3).



Step 6 When the Phillips-head screws are loosened, swing the hinged door open so that the GPU can slide out of the slot (1).

Note
If you're installing a GPU for the first time, GPU filler blanks will be present. They must be removed (2).



Step 7 Grasp the GPU, and holding it level, pull it out so that it disconnects from the PCIe socket.

Step 8 Slide the GPU out until it clears the slot in the cage.

What to do next

Reinsert an FHFL GPU. See Installing a FHFL GPU, on page 54.

Installing a FHFL GPU

Full height, full length (FHFL) dual-slot GPUs install horizontally into the GPU cage so that the GPU's connector seats into the PCIe socket inside the GPU cage.

GPU cages have standard PCIe alignment features, such as a slot in the rear wall that accepts a tab on the GPU, and a horizontal notch on the front wall that accepts the edge of the GPU's extender bracket.

When installing the first GPU, install it in the lowest numbered slot in the GPU cage, slot 1 in GPU cage A or slot 3 in GPU cage B.

Use the following procedure to install a full height full length GPU into GPU cage.

Before you begin

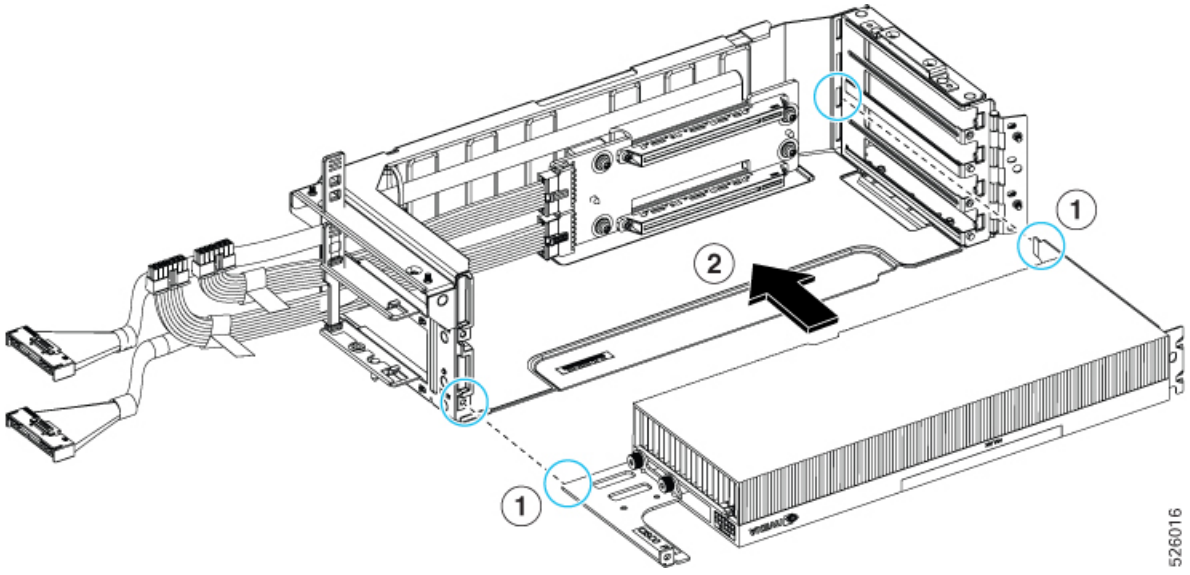
Review the GPU Replacement Guidelines and Limitations, on page 49. Before attempting this procedure, gather a #2 Phillips and a T8 Torx screwdriver.

Procedure

Step 1 Align the GPU with the slot so that the GPU's golden fingers will meet the PCIe socket.

Step 2 Holding the GPU level, align the front edge of the extender bracket with the notch on the front wall of the GPU cage (1).

Step 3 Slide the GPU into the slot (2), making sure that the tab on the GPU lines up with the correct slot in the rear wall (1).



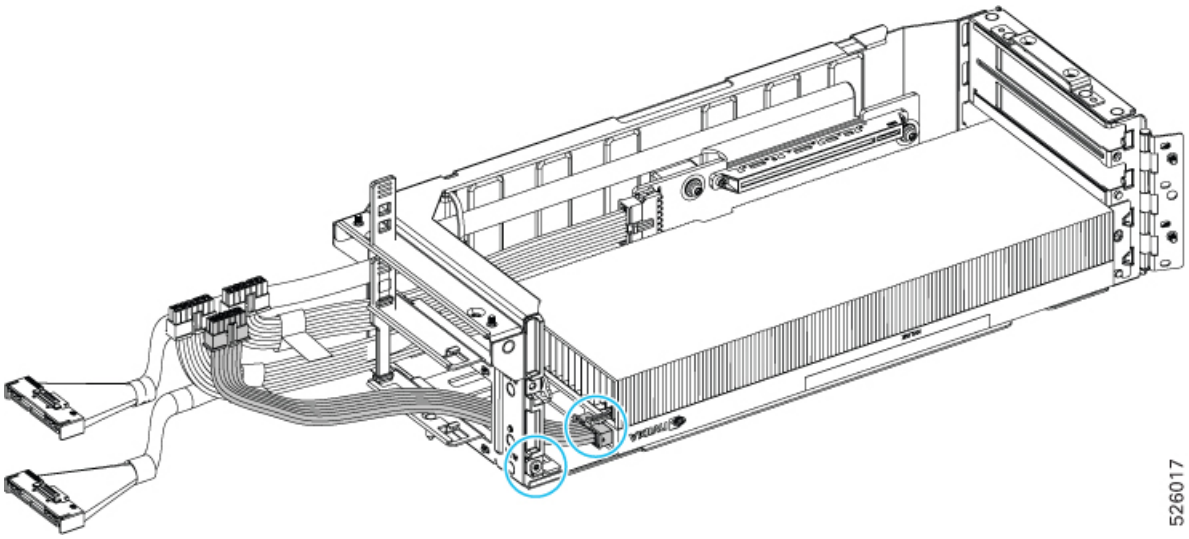
Step 4 When you feel the connector on the GPU meet the PCIe socket, press the GPU firmly into place to install it.

Step 5 Attach the GPU power cable to the node motherboard.

a) Route the GPU power cable through the cage and connect it to the GPU.

Note
The GPU power cable must be routed behind the corner of the GPU cage to prevent obstruction.

b) Using the T8 Torx screwdriver, reinsert the screw to secure the GPU into the slot.



c) Connect the GPU power cable to the AUX1 or AUX2 connector on the motherboard.

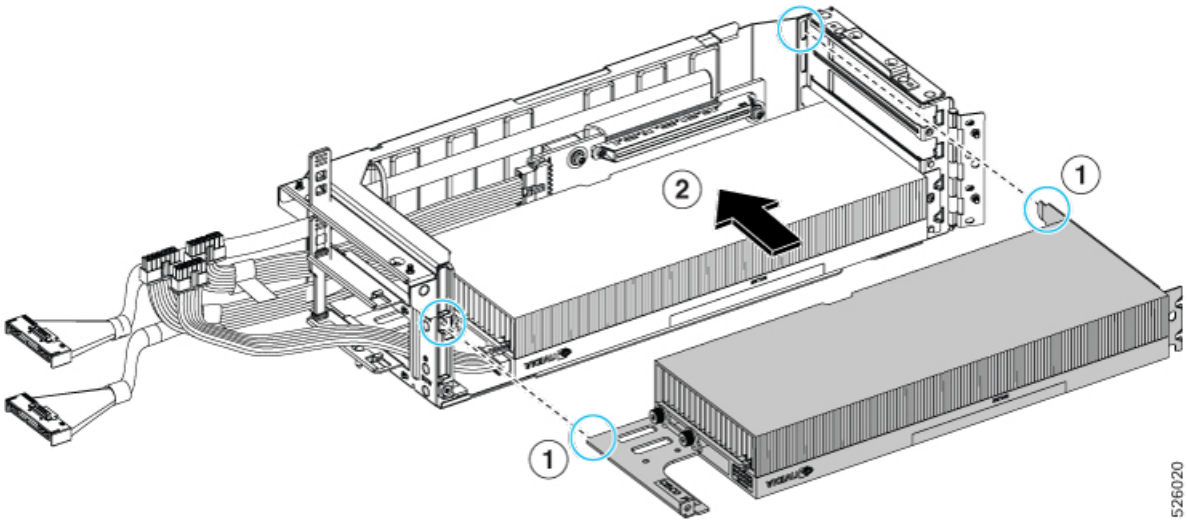
It is a best practice to connect the GPU to the corresponding AUX connector, for example, the GPU in slot 1 should be connected to the AUX1 connector, and the GPU in slot 2 should be connected to the AUX2 connector.

d) Using the T8 screwdriver, install the securing screw making sure to tighten the screw to a range of 4 to 6 in lbs. of torque.

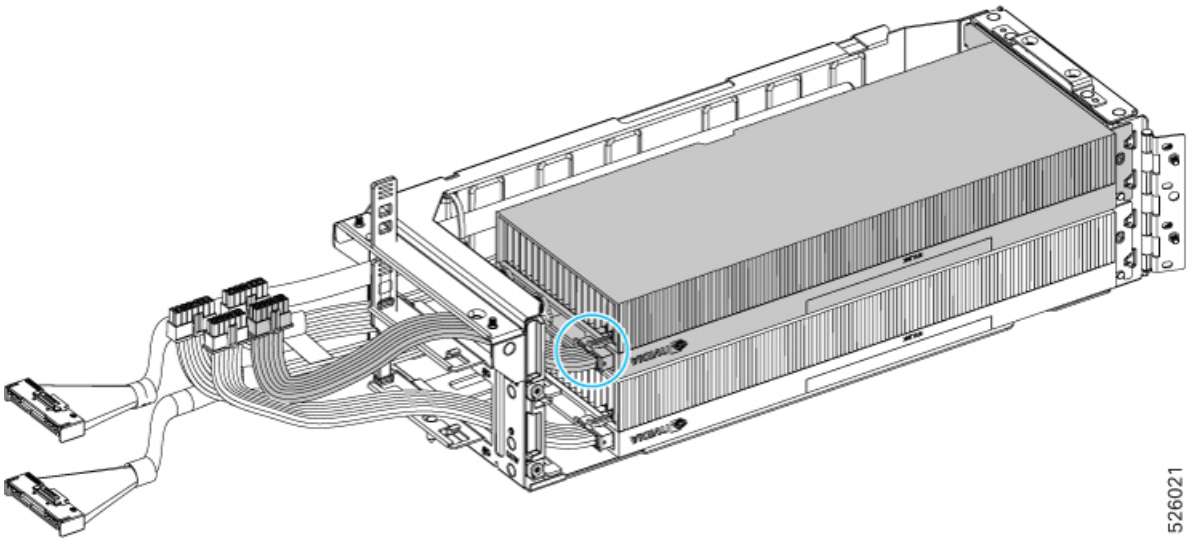
Caution
Do not overtighten! This screw is small and susceptible to shearing or stripping if overtightened. Do not exceed the documented torque specs.

Step 6 (Optional) To install a second GPU:

a) Align the GPU and slide into place until the GPU's connector is seated into the PCIe socket.



b) Route the GPU cable through the GPU cage and behind the front corner of the cage.



c) Connect the GPU power cable to the other AUX connector on the motherboard.
d) Using the T8 screwdriver, install the securing screw making sure to tighten the screw to a range of 4 to 6 in lbs. of torque.

Caution
Do not overtighten! This screw is small and susceptible to shearing or stripping if overtightened. Do not exceed the documented torque specs.

Step 7 Complete the GPU installation.

a) Verify that the securing screw is installed for each GPU to hold them into the PCIe slot.
b) Grasp the rubber cable retainers and reconnect them.
c) Swing the hinged doors closed.
d) Using the #2 Phillips screwdriver, tighten each of the thumbscrews to secure the GPU(s) into the slot(s).

What to do next

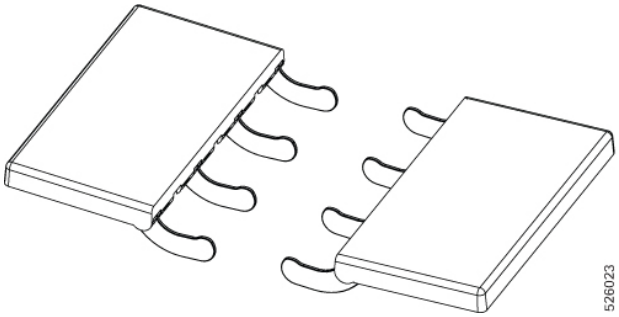
Install the GPU cage. See Installing a GPU Cage, on page 43.

Replacing the NVIDIA NVLink GPU Interconnect

The NVIDIA NVLink GPU Interconnect is a physical bridge card used to connect two adjacent NVIDIA H200 GPUs. When installed, the NVLink GPU Interconnect electrically connects the GPUs and increases GPU performance. For more information, see NVIDIA's website.

To use the NVIDIA NVLink GPU Interconnect, the GPUs must be installed in the same GPU cage (adjacent to each other) for the component to span across the GPUs.

The interconnect snaps onto connectors on the back of each GPU, so installation is tool-less. Removal of the interconnect, however, does require specific tools, which Cisco offers through UCSX-GPU-RKIT-NV=. When you order UCSX-GPU-RKIT-NV=, you will receive the NVLink removal tools, which are similar to the following.



To replace an NVIDIA NVLink GPU Interconnect, use the following procedures:

- Removing the NVIDIA NVLink GPU Interconnect, on page 57
- Installing the NVIDIA NVLink GPU Interconnect, on page 58

Removing the NVIDIA NVLink GPU Interconnect

This task explains how to remove the NVIDIA NVLink GPU Interconnect from NVIDIA H200 NVL GPUs. For additional information, see NVIDIA's website.

Caution Do not attempt this task without using the appropriate tools or you risk damaging GPUs!

Before you begin

Before attempting this procedure, you must have the NVLink removal tool, which is orderable from Cisco by requesting UCSX-GPU-RKIT-NV=. This tool has hooks that will be used to pry the GPU interconnect off of the GPUs without requiring you to remove the GPUs from the GPU cage.

Procedure

Step 1 If you have not ready removed the GPU cage, remove it now.

See Removing a GPU Cage, on page 37.

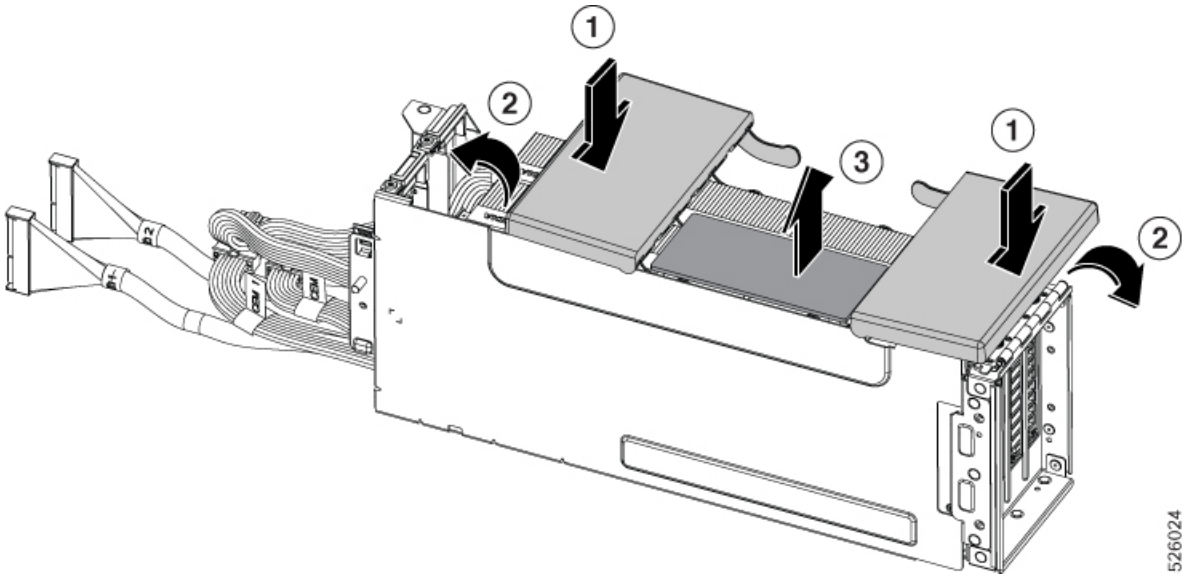
Step 2 For each tool, insert the hooks of the tool into the GPU slot under the GPU interconnect.

Caution
The hook ends of the tools must be between the GPUs and under the GPU interconnect. Make sure that both tools are installed the same way. Neither tool can be off center or further away from the GPU interconnect than the other tool.

Step 3 For each tool, make sure that the hooks are straddling the GPU, and that there is a firm pivot point to support the tool.

Step 4 When the tools are inserted, hold them in place so that they do not drift, and simultaneously push down on each tool with equal pressure.

If you have installed the tools and performed this step correctly, the tools will act like a lever to gently pry the GPU interconnect off of the GPUs.



Step 5 When the GPU interconnect disconnects from the GPUs, grasp it and lift it straight up to remove it.

Caution
To avoid damaging the golden fingers, do not tip, twist, or rotate the GPU interconnect while you are lifting it off of the GPUs. Keep the interconnect straight when you remove it.

Step 6 Remove the tools and replace the cover on the GPU interconnect golden fingers or replace the GPU interconnect, as needed.

Installing the NVIDIA NVLink GPU Interconnect

This task explains how to install the NVIDIA NVLink GPU interconnect on the NVIDIA H200 NVL GPUs. Installation is toolless.

The PCIe node supports two GPUs per GPU cage, so the supported NVLink GPU interconnect is a two-way interconnect.

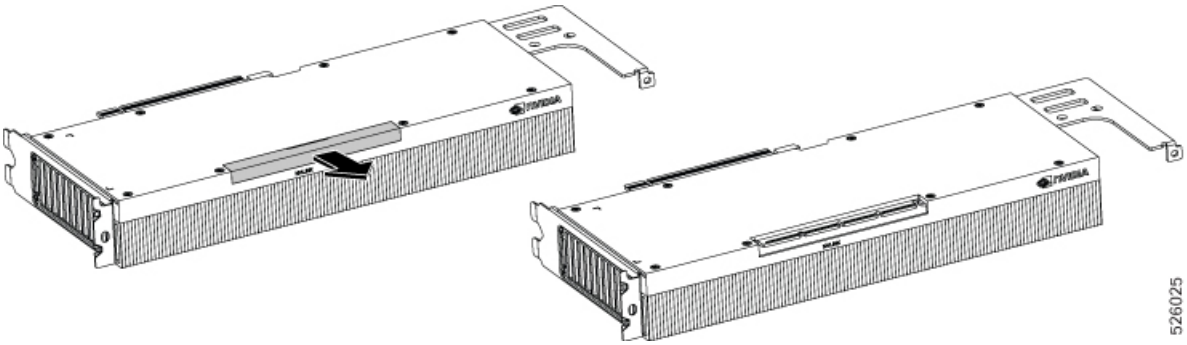
Note You will find it easier to install the GPU interconnect when the GPUs are already installed in a GPU cage.

Use this task to install the GPU interconnect on two adjacent H200 NVL GPUs.

Procedure

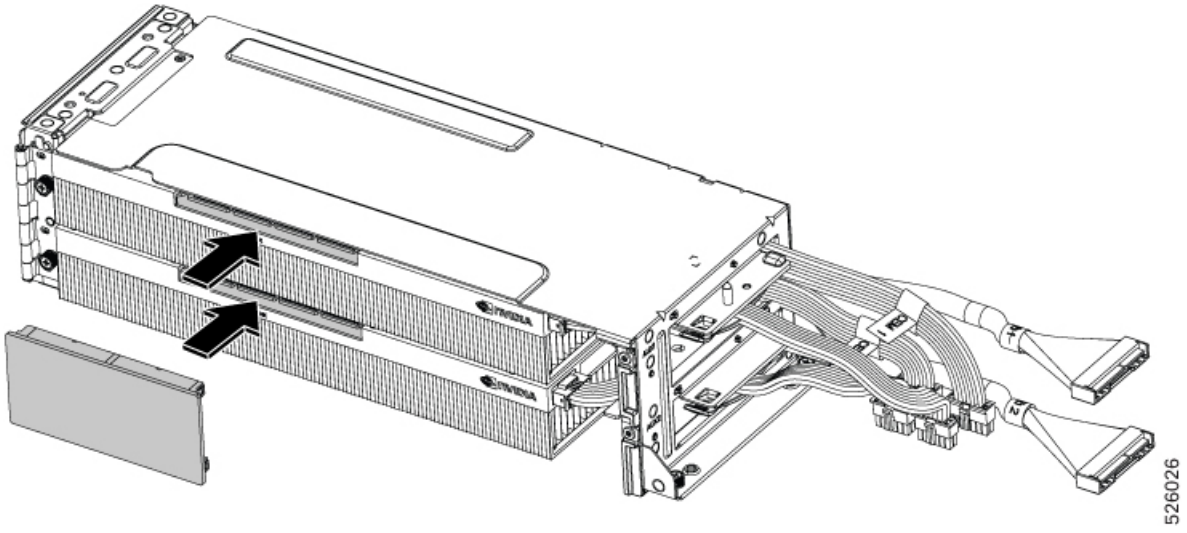
Step 1 If you have not already removed the GPU cage, remove it now.
See Removing a GPU Cage, on page 37.

Step 2 Remove the cover from both GPUs.



Step 3 Position the two-way GPU interconnect over the GPUs' connectors.

Step 4 Holding the GPU interconnect level, press the GPU interconnect onto the GPUs until it is completely seated onto both.



Installing the PCIe Node Cover

Use this task to install the PCIe node's top cover.

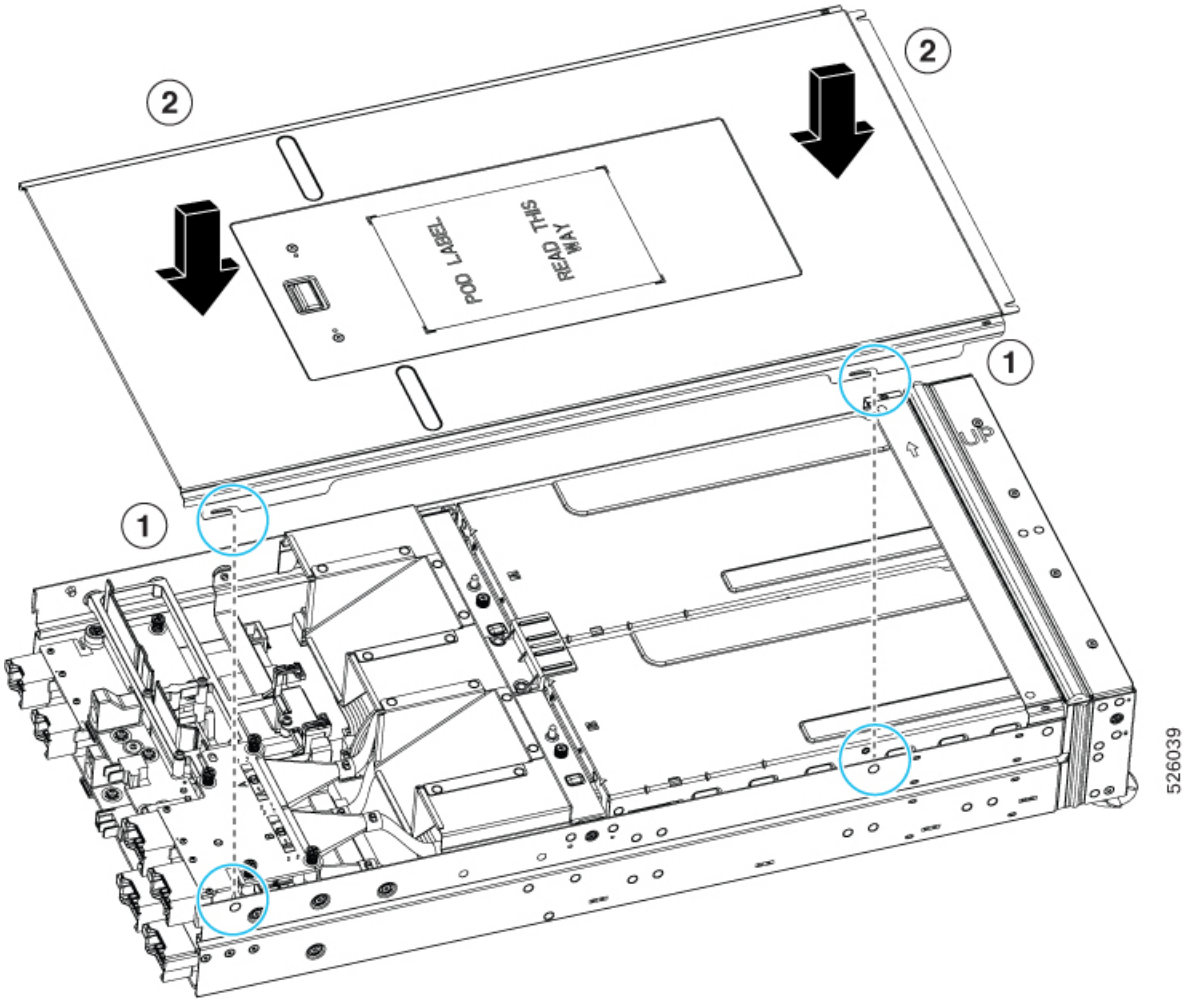
Procedure

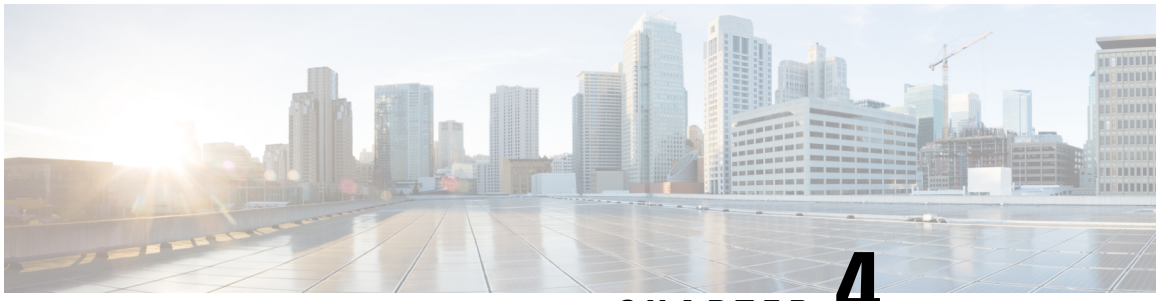
Step 1 Align the cutouts on the top cover's sidewalls with the guide pins on the PCIe node's sidewalls (1).

Step 2 Lower and install the cover at an angle so that the cutouts on the rear of the catch the guide pins on the inside walls of the node (2).

Step 3 When the top cover is flush with the PCIe node, keep the top cover flat, and slide it forward until the release button clicks.

Note
Make sure that the front edge of the top cover slides under the metal edge of the node front cover. If you feel resistance, these two edges might be making contact instead of sitting on top of each other.





CHAPTER 4

Recycling the PCB Assemblies

This chapter contains the following topics.

- Recycling the PCB Assemblies, on page 61

Recycling the PCB Assemblies

The Cisco UCS X580p PCIe Node has multiple different PCB assemblies (PCBAs), which consist of the PCB motherboard plus any components mounted on it.

For compliance with local e-waste and recycling laws, use the following procedures to disassemble the PCBAs.

- Recycling the GPU Cage PCBA, on page 61
- Recycling the Node PCBA, on page 63

Recycling the GPU Cage PCBA

Use the following task to disassemble the GPU cage for recycling.

Before you begin

Note **For Recyclers Only!** This procedure is not a standard field-service option. This procedure is for recyclers who will be reclaiming the electronics and sheet metal for proper disposal to comply with local eco design and e-waste regulations.

To remove a GPU cage PCBA, the node must be removed from the chassis.

You will find it helpful to gather a #2 Phillips screwdriver before beginning this procedure.

Procedure

Step 1 Remove the node's top cover.

Go to Removing the PCIe Node Cover, on page 26.

Step 2 Remove the node's air baffle.

Go to Removing the PCIe Node Air Baffle, on page 28.

Step 3 Remove the node's front panel.

Go to Removing the PCIe Node Front Cover, on page 32.

Step 4 If you have not already removed the GPU cages, remove them now.

Go to Removing a GPU Cage, on page 37.

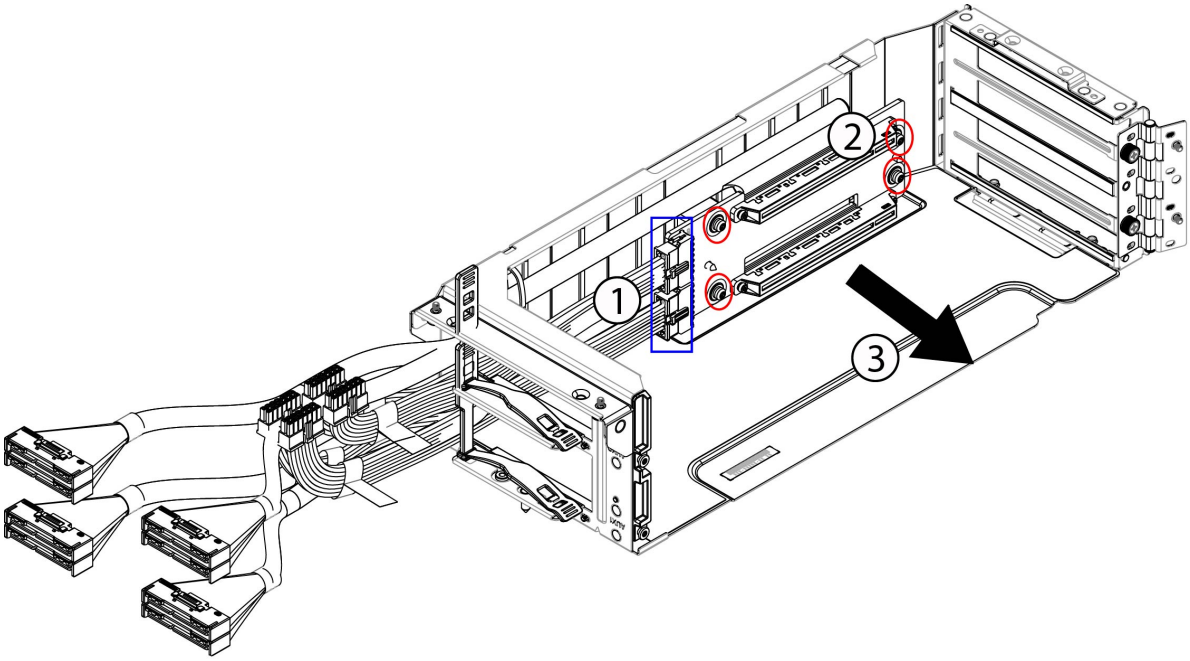
Step 5 If necessary, remove the NVIDIA NVLink GPU interconnect.

Removing the NVIDIA NVLink GPU Interconnect, on page 57

Step 6 Remove the GPU from the cage as documented in Removing a GPU Cage, on page 37.

Step 7 Remove the PCB from the cage.

a) Disconnect the cables from the cage's PCB (1).
b) Using a #2 Phillips screwdriver, remove the four screws that secure PCB to the cage (2).
c) Grasp the PCB and remove it (3).



493088

Step 8 Dispose of the PCB, components, and sheet metal in compliance with relevant e-waste and recycling laws.

Recycling the Node PCBA

Each Cisco UCS X580p PCIe Node has a PCBA (motherboard) that is connected to sheet metal. You must disconnect the PCBA from the chassis sheet metal to recycle the PCBA. You will need to disassemble and remove additional parts to gain access to the PCBA.

You will need to recycle the PCBA for each PCIe node.

Use the following procedure to recycle the Cisco UCS X580p PCIe Node's motherboard.

Note The following illustrations focus on the rear of the node, which is where the PCB is located. In these illustrations, dashed lines indicate the locations of fasteners present on the node, but not visible in the illustration.

Before you begin

Note **For Recyclers Only!** This procedure is not a standard field-service option. This procedure is for recyclers who will be reclaiming the electronics and sheet metal for proper disposal to comply with local eco design and e-waste regulations.

To remove the node's printed circuit board assembly (PCBA) which is the main PCB plus components mounted on it, the node must be removed from the chassis.

You will find it helpful to gather a #2 Phillips, T8 Torx, and T10 Torx screwdriver for this procedure.

The node's PCB is located at the rear of the node. You will find it helpful to remove the GPU cages before attempting this procedure. When the GPU cages are removed, you will have ample space to work on the node's PCB.

Procedure

Step 1 Remove the node's top cover.

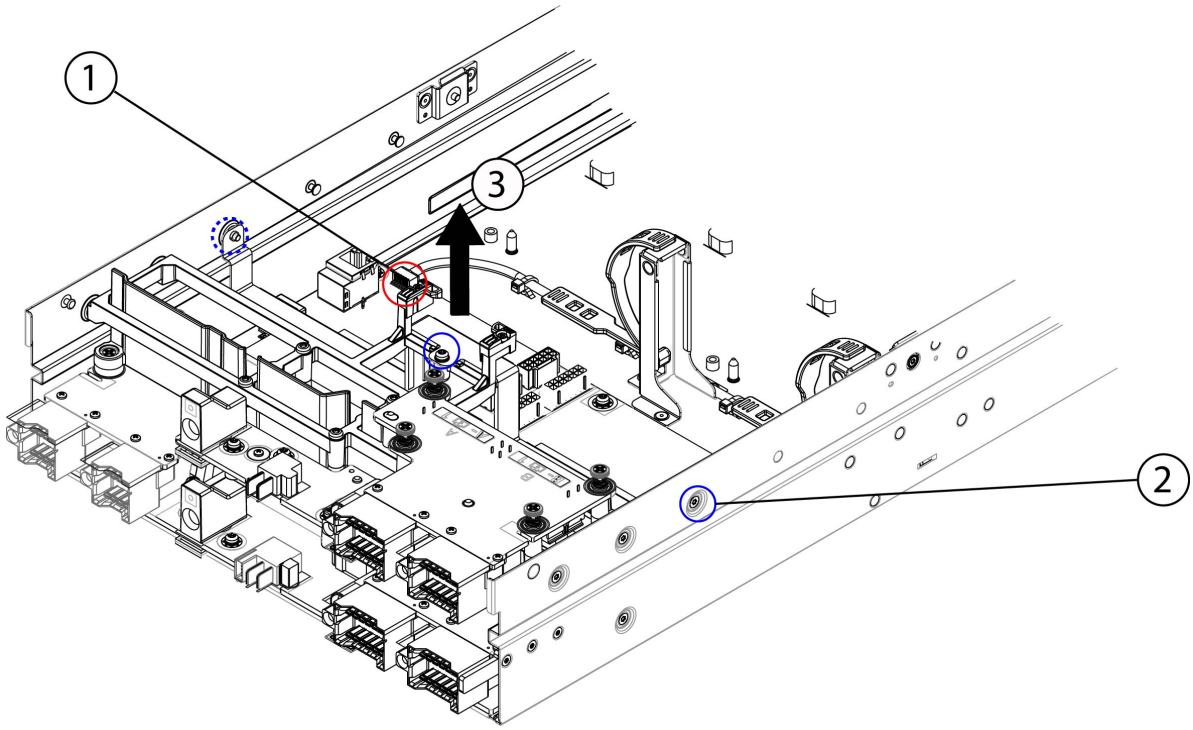
Go to Removing the PCIe Node Cover, on page 26.

Step 2 If you have not already removed the GPU cages, remove them now.

Go to Removing a GPU Cage, on page 37.

Step 3 Begin disconnecting the node PCB.

a) Disconnect the cable (1) from the node's PCB.
b) Using T8 and T10 Torx screwdriver, remove three mounting screws from the small mounting bracket (2), including the one screw on each exterior sidewall.
c) Grasp the small support bracket and remove it from the node (3).

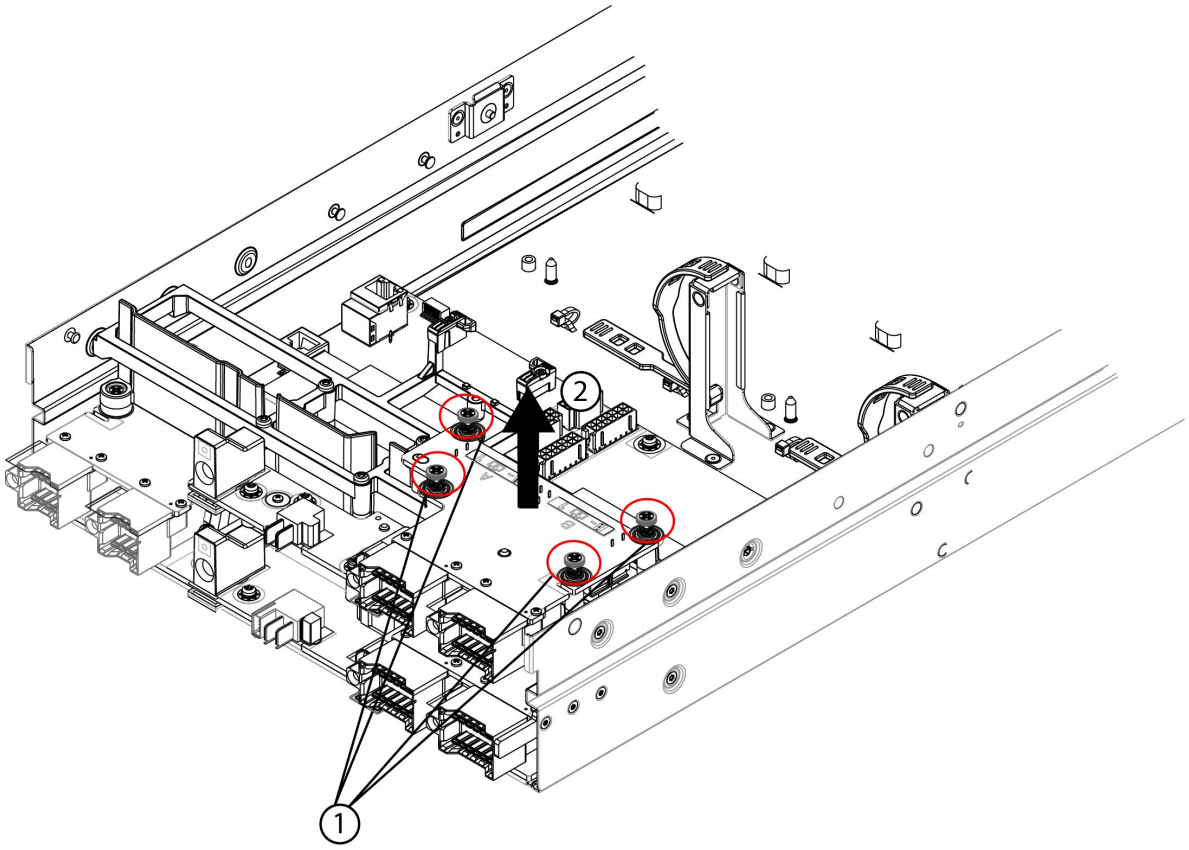


493089

Step 4 Remove the upper rear mezzanine card.

a) Using the #2 Phillips screwdriver, loosen the four captive thumbscrews (1).

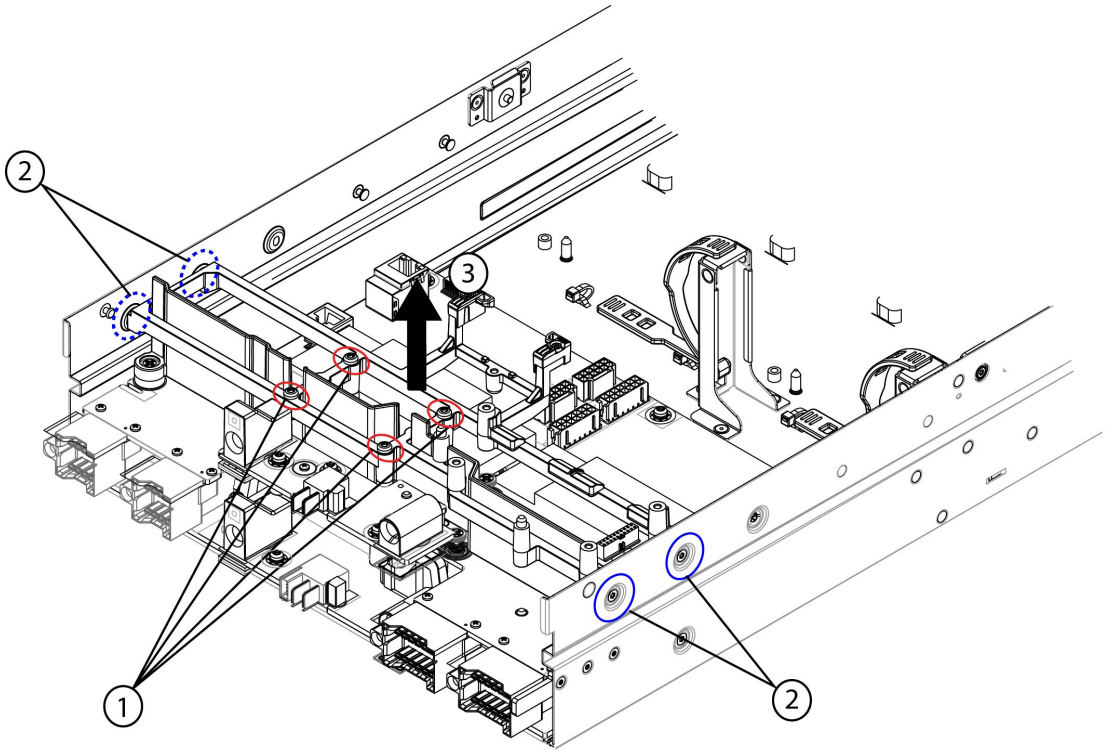
b) Grasp the top-layer rear mezzanine card, and remove it from the node (2).



493090

Step 5 Remove the upper support frame.

a) Using the T10 Torx screwdriver, remove the four screws on the top of the support frame (1).
b) Using the T8 Torx screwdriver, remove the four screws on the node's sidewalls, two screws per side (2).
c) Grasp the upper support frame, and remove it from the node (3).

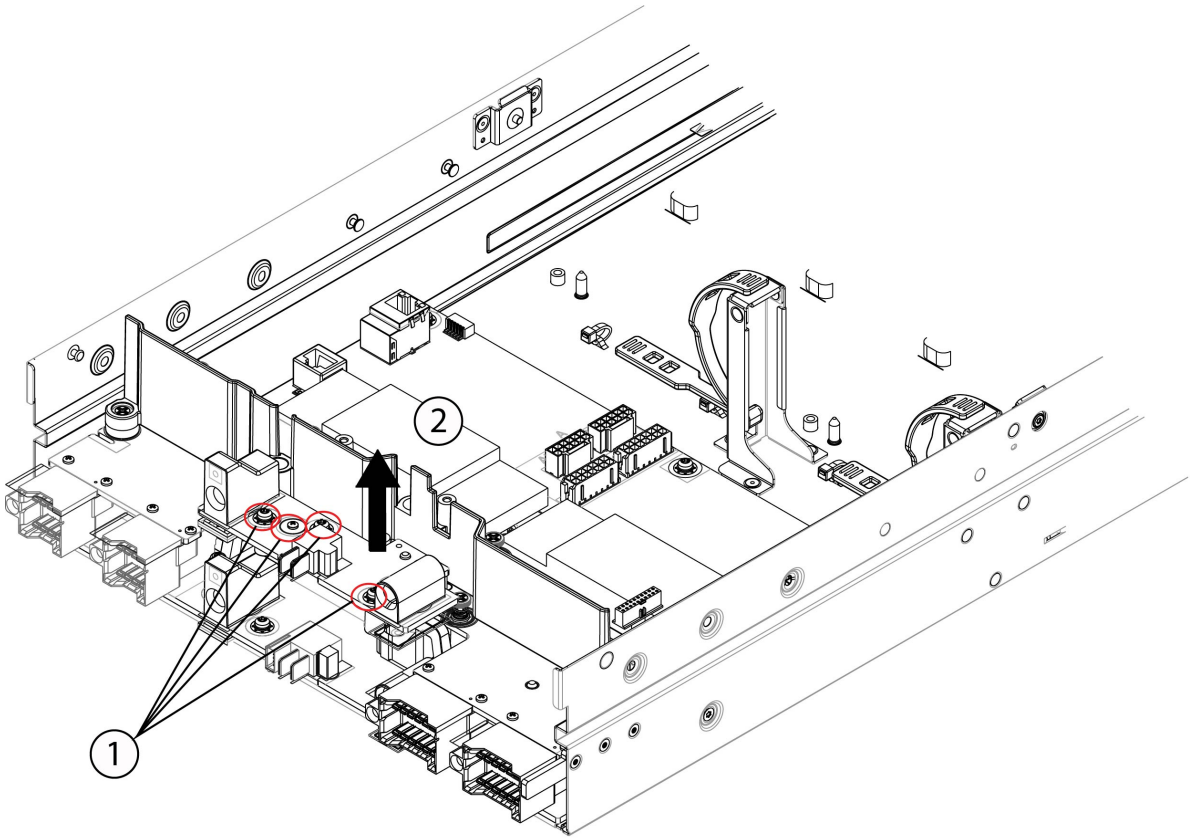


493091

Step 6 Remove the node power card.

a) Using the T10 Torx screwdriver, remove the four screws on top of the power card (1).

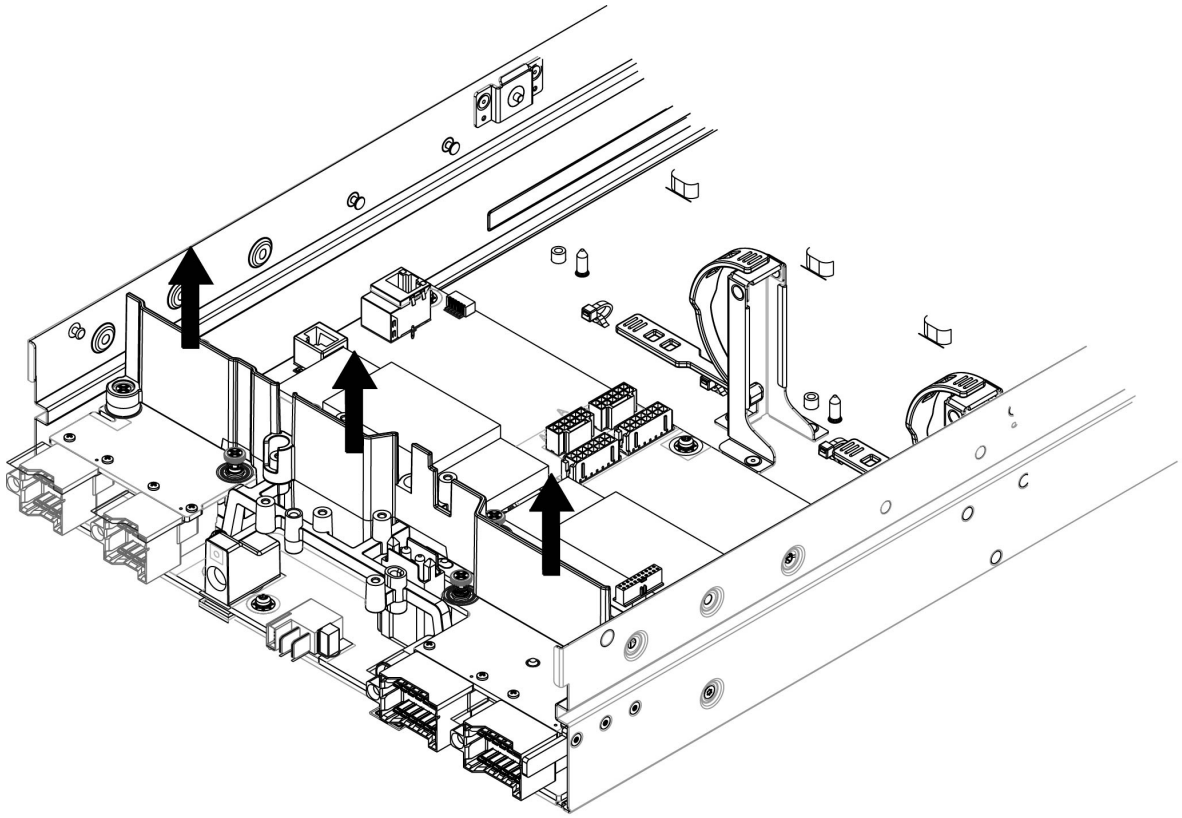
b) Grasp the power card and remove it from the node (2).



493092

Step 7 Grasp the node's air blocker and remove it.

Note
Each end of the air blocker has a cylinder that fits around a captive screw on the lower level rear mezzanine cards. If you feel some resistance while removing the air blocker, check the ends of the air blocker.

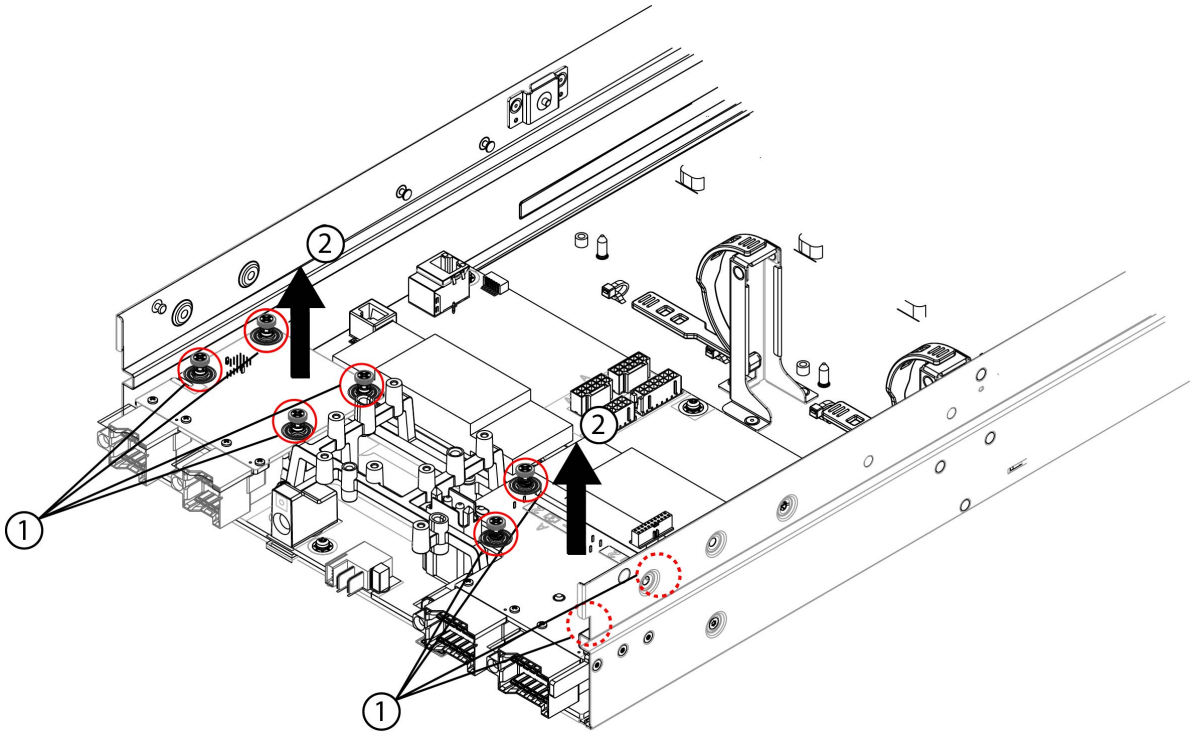


493093

Step 8 Remove both of the lower cards.

a) For each card, using the #2 Phillips screwdriver, loosen the four captive thumbscrews (1).

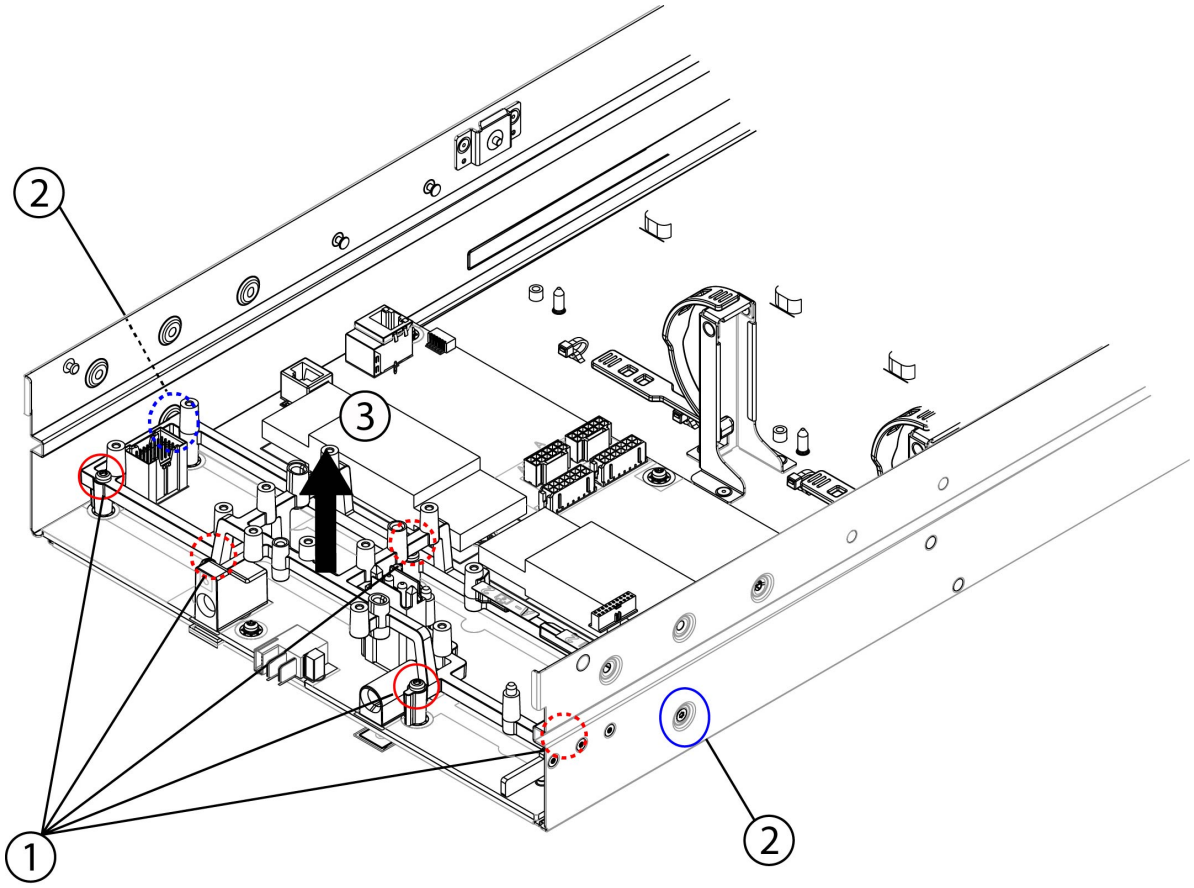
b) For each card, grasp the card, and remove it from the node (2).



493094

Step 9 Remove the lower support frame.

a) Using the T10 Torx screwdriver, loosen the five thumbscrews on top of the frame (1).
b) Using the T8 Torx screwdriver, loosen the two screws on the exterior of the node's sidewalls, one per side (2).
c) Grasp the lower support frame and remove it from the node (3).

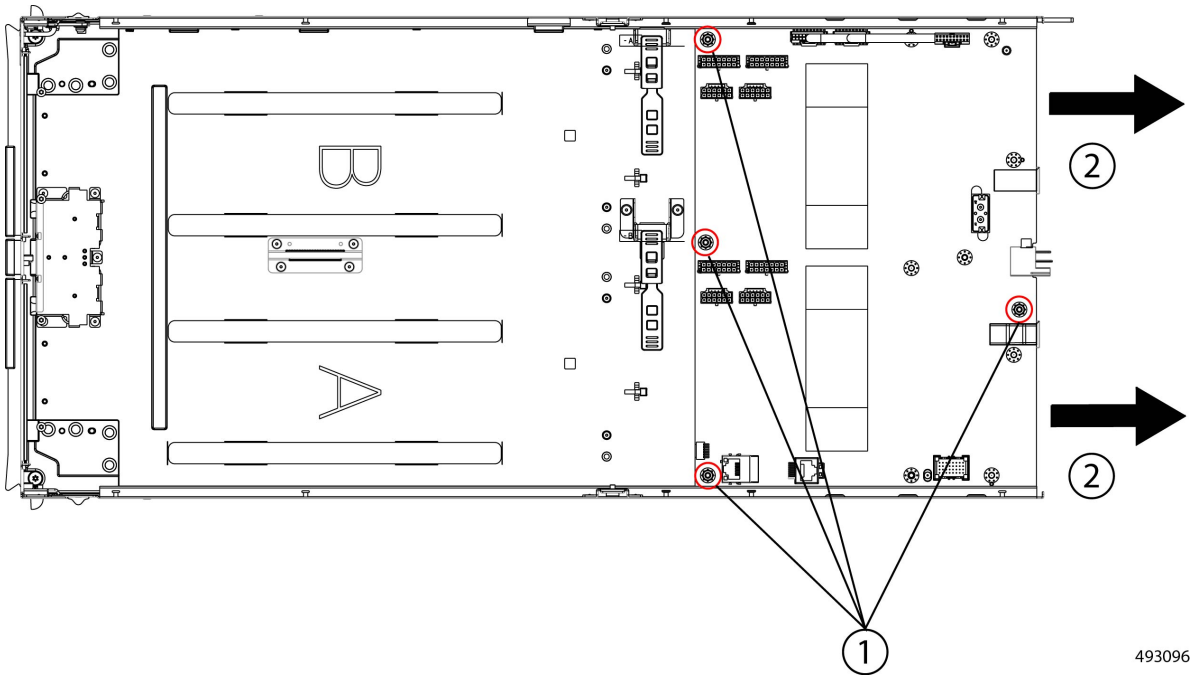


493095

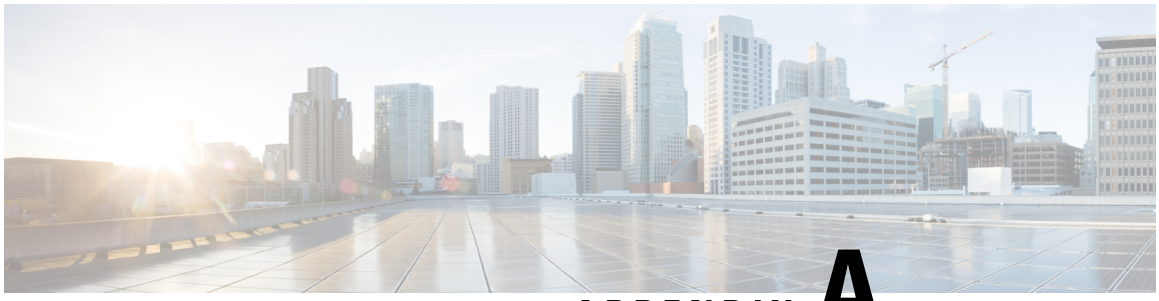
Step 10 Detach the node's PCB from the sheet metal tray.

a) Using a T10 Torx screwdriver, remove the four screws (1).

b) Grasp the node's PCB and remove it from the node (2).



Step 11 Dispose of the PCB, components, and sheet metal in compliance with relevant e-waste and recycling laws.



APPENDIX A

PCIe Node Specifications

This appendix contains the following topics:

- Physical Specifications, on page 73
- Environmental Specifications, on page 73
- Power Specifications, on page 74

Physical Specifications

The following table shows the physical specifications for the Cisco UCS X580p PCIe Node.

Parameter	Value
Height	3.70 in. (94 mm)
Width	11.28 in. (286.5 mm)
Depth	24 in. (602 mm)
Weight	• Minimally configured node weight = 18.51 lbs (8.4 kg) • Fully loaded PCIe Node with NVIDIA L40S GPU = 28.66 lbs (13 kg); minimum config with one of these GPUs = 21.05 lbs (9.55 kg) • Fully loaded PCIe Node with NVIDIA RTX PRO 6000 GPU = 29.98 lbs (13.6 kg); minimum config with one of these GPUs = 21.38 lbs (9.7 kg) • Fully loaded PCIe Node with NVIDIA H200 NVL GPU (including two NVL bridge cards) = 30.42 lbs (13.8 kg); minimum config with one of these GPUs = 21.43 lbs (9.72 kg)

Environmental Specifications

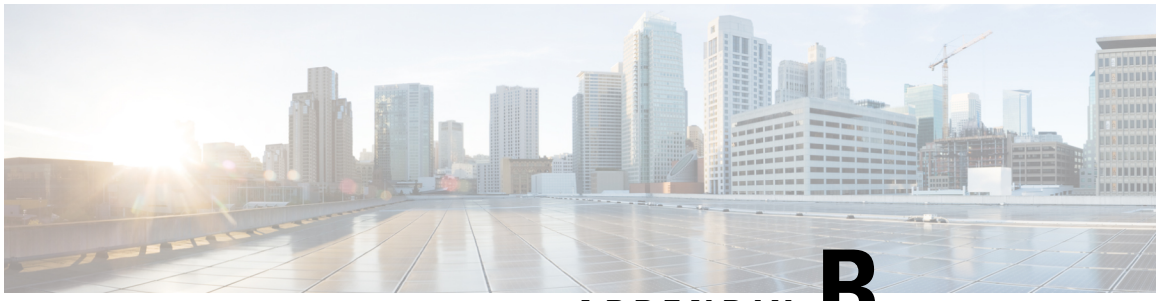
The following table shows the environmental specifications for the Cisco UCS X580p PCIe Node.

Parameter	Value
Operating Temperature	50° to 86° F (10° to 30° C)
Non-Operating Temperature	-40° to 149°F (–40° to 65° C)
Operating Humidity	5% to 90% non condensing
Non-Operating Humidity	5% to 93% non condensing
Operating Altitude	0 to 10,000 ft (0 to 3000m); maximum ambient temperature decreases by 1° C per 300m
Non-Operating Altitude	40,000 ft (12,000m)

Power Specifications

For configuration-specific power specifications, use the Cisco power calculator. Go to:

UCS Power Calculator



APPENDIX B

Obtaining Hardware

This appendix contains the following topic.

- Obtaining Hardware, on page 75

Obtaining Hardware

Each Cisco UCS X580p PCIe Node interoperates with other required hardware to provide GPU processing offload and acceleration for the Cisco UCS X-Series M8 compute node that is paired with it.

The following table shows the required hardware to support a fully functioning PCIe node. Use the following table when you want to order spares or additional PCIe nodes to scale out your Cisco UCS X-Fabric deployment.

Note Power cables are required and ship with the GPU cages.

Hardware Component	Description	Cisco PID
Chassis	UCS X-Series Server Chassis, for example the Cisco UCS X9508 Server Chassis	UCSX-9508
Compute Nodes	UCS X-Series M8 Compute Nodes, for example, the Cisco UCS X210c M8 Compute Node	UCSX-215C-M8 UCSX-210C-M8
	The UCS X210c M7 Compute Node	UCSX-210C-M7 (only with 5th Gen Intel® Xeon® Processor Scalable Family CPUs)
PCIe Nodes	UCS X-Series Gen5 PCIe node, for example the Cisco UCS X580p PCIe Node	UCSX-580P
GPU Cage Riser Blank (filler panel)	PCIe blank for UCS X-series 580p PCIe node	UCSX-PCIF-GPU

Hardware Component	Description	Cisco PID
PCIe Node GPUs, FHFL	NVIDIA RTX Pro 6000, PCIe, 600W, 96GB, 2-Slot, FHFL GPU	UCSX-GPU-RTXP6000 This PID comes with a power cable included.
	NVIDIA H200 NVL, PCIe, 600W, 141 GB, 2-Slot NVL2, FHFL GPU	UCSX-GPU-H200-NVL Required cable, 1 (CBL-X580p-GPU-N) is included with each GPU
	NVIDIA L40S 350W, 48GB, 2-Slot, FHFL GPU	UCSX-GPU-L40S This PID comes with a power cable included.
Cisco UCS X-Fabric Modules (XFMs)	UCS X9516 X-Fabric PCIe Gen5 switch module for 9508 chassis Must be deployed as a pair, so two are required per Cisco X-Series Server chassis	UCSX-FS-9516
Rear Mezzanine Card	UCS PCI Gen5 Mezz card for X-Fabric Connectivity Required for Cisco UCS X-Series M8 compute nodes (for example, the Cisco UCS X210c M8 Compute Node)	UCSX-V5-PCIME



INDEX

B

baffle 28, 30
bracket, H200 NVL 50–51

F

full height full length GPU, installing 54
full height full length GPU, removing 52

G

GPU bridge, H200 NVL 57–58
GPU cage 37, 43
GPU cage PCBA, recycling 61
GPU cage, installing 43
GPU cage, removing 37
GPU Interconnect, H200 NVL 57–58
GPU, FHFL 52, 54

I

installing PCIe node cover 60
installing, full height full length GPU 54
installing, GPU cage 43
installing, legacy H200 NVL bracket 51
installing, node air baffle 30
installing, node front cover 34
installing, NVIDIA NVLink GPU Interconnect 58
installing, X580p PCIe node 18

L

LED 5
    PCIe node health 5
LED *(continued)*
    PCIe node locator 5
legacy H200 NVL bracket, installing 51
legacy H200 NVL bracket, removing 50

N

node air baffle, installing 30
node air baffle, removing 28
node front cover 32, 34
node front cover, installing 34
node front cover, removing 32
node PCB, recycling 63
NVIDIA NVLink GPU Interconnect, installing 58
NVIDIA NVLink GPU Interconnect, removing 57

P

PCBA, GPU cage 61
PCIe node cover, installing 60
PCIe node cover, removing 26
PCIe node, installing 18
PCIe node, removing 19

R

recycling, GPU cage PCBA 61
recycling, node PCB 63
removing PCIe node cover 26
removing, full height full length GPU 52
removing, GPU cage 37
removing, legacy H200 NVL bracket 50
removing, node air baffle 28
removing, node front cover 32
removing, NVIDIA NVLink GPU Interconnect 57
removing, X580p PCIe node 19