Falcon Composable GPU Solution

— Falcon 5012

User Manual

Version 1.0

April 15th, 2026

Dear Users,

Thank you for choosing our product.

The excellent quality and performance make our products superior in the like
product. For you to have a good understanding of Falcon Composable GPU, Please read the replacement guide and operate according to the suggested steps.

If you have any questions when using our machine, please feel free to contact
us. We are more than happy to serve you constantly.

Technical Support: support@h3platform.com
FAQ: https://www.h3platform.com/

H3 Platform Inc. researches and develops PCIe switch-based technology and solutions.

---- H3 Platform Inc.

Revision History

Rev.	Date	Description
0.1		
0.2	2024.12.12	System modes and GUI functions updated
0.3	2025.01.20	System modes and GUI functions updated
0.4	2025.02.25	Retimer installation guide updated BMC GUI use guide updated
0.5	2025.03.17	Switch firmware update procedure
0.6	2025.04.18	Retimer specification and retimer firmware update procedures added
0.7	2025.06.18	User permissions revised
0.8	2025.06.24	Public template applied
0.9	2025.08.07	Fan replacement procedure updated and description of hot-swap revised SKU label elements updated
1.0	2026.04.15	1. Regulatory compliance notices added 2. web GUI guide updated 3. BMC GUI guide updated 4. System LED indicators updated 5. Host adapter firmware update settings updated

Table of Contents

Regulatory Compliance Notices

FCC Statement (Class A Information Technology Equipment)

This device complies with part 15 of the FCC Rules. Operation is subject to the following two conditions:

(1) This device may not cause harmful interference, and
(2) This device must accept any interference received, including interference that may cause undesired operation.

VCCI クラスAに関する注意事項

この装置は、クラスA機器です。この装置を住宅環境で使用すると電波妨害を引き起こすことがあります。この場合には使用者が適切な対策を講ずるよう要求されることがあります。

VCCI-A

VCCI Class A Statement (for Japan only)

This is a Class A equipment. Operation of this equipment in a residential environment could cause radio interference. In such a case, the user may be required to take corrective actions.

VCCI-A

Chapter 1.

Introduction to the PCIe Gen5 GPU Expansion System

Learn about the PCIe Gen5 GPU Expansion System (Falcon 5012), designed for flexible and scalable GPU deployment. This solution is built on a composable infrastructure, which separates GPUs from servers for increased flexibility and scalability. The system includes a GPU chassis, PCIe 5.0 switches, and centralized management software, providing efficient resource allocation and performance optimization (refer to Figure 1).

Figure 1. The PCIe Gen5 GPU Expansion System (Falcon 5012), featuring a high-density GPU enclosure optimized for modern computing environments.

This document provides essential information for users and administrators of the Falcon 5012 System.

1.1 System Overview

The PCIe Gen5 GPU Solution features a composable infrastructure, decoupling GPUs from host servers for flexible and scalable deployment. It includes external GPU chassis, PCIe 5.0 switches, and centralized management software.

To better understand the system's capabilities, refer to Table 1, which summarizes the key features of the PCIe Gen5 GPU Solution, including its resource disaggregation, high-bandwidth PCIe 5.0 connectivity, and dynamic scalability.

Table 1. Key Features of PCIe Gen5 GPU Solution

Feature	Description
Resource Disaggregation	GPUs are separated from servers and housed in an external chassis.
PCIe 5.0 Connectivity	High-bandwidth 1-meter CDFP cables connect servers and GPUs via switches.
Scalable Expansion	GPUs can be dynamically added or reassigned based on application needs.

The next section delves into the Falcon 5012 system architecture, highlighting its design and component integration for high-performance computing.

1.2 System Architecture

Figure 2 illustrates the high-level block diagram of the PCIe Gen5 GPU Solution, showing the relationship between disaggregated nodes, PoD management, and logic composable nodes.

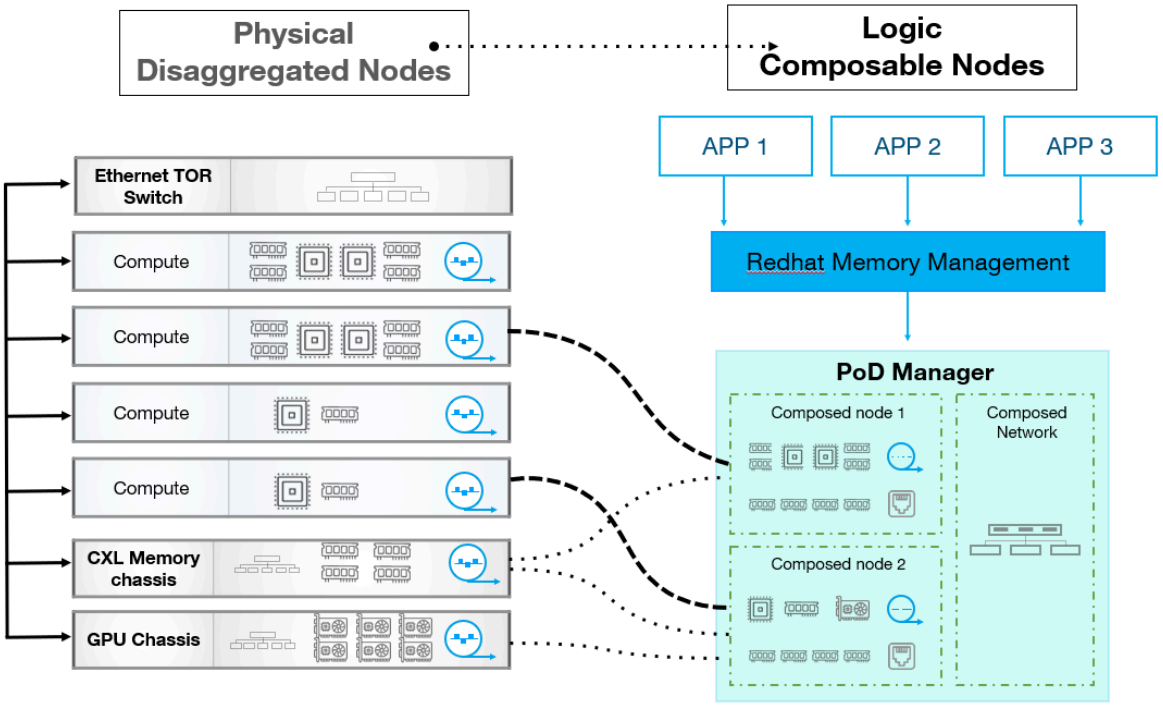


Figure 2. Block Diagram of PCIe Gen5 GPU Solution Architecture

The following process outlines how H3 Composable Management Software dynamically reorganizes GPU resources to enhance application execution, as shown in Figure 3:

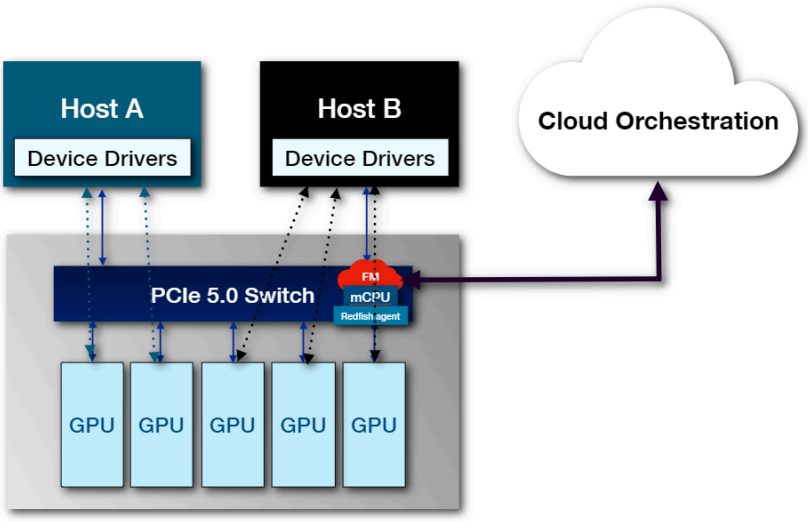


Figure 3. Block Diagram of System Operation Process

1. Command Transmission: Users send operational commands via Ethernet to the Redfish Agent on the mCPU.
2. Resource Mapping: The Redfish Agent uses the FM API to issue instructions to the PCIe 5.0 Switch.
3. GPU Allocation: The PCIe Switch dynamically allocates GPU resources to host servers

    - Two GPUs are assigned to Host A.
    - Three GPUs are assigned to Host B.

This process enables flexible GPU resource management and ensures efficient application execution.

This dynamic allocation process improves resource utilization, minimizes idle GPU rates, and reduces enterprises' Total Cost of Ownership (TCO).

The next section will provide an in-depth overview of the key features of the PCIe Gen5 GPU Solution.

1.3 Key Features

The Falcon 5012 GPU Expansion System offers advanced features designed to optimize GPU resource management and operational flexibility. These features are categorized into System Features, Fabric Manager Features, and System Management UI Features. The following tables provide an overview of each category:

System Features

Table 2 outlines the key system features, highlighting capabilities such as dynamic provisioning, multiple application scenarios, and real-time monitoring.

Table 2. System Features

Feature	Description
GPU Dynamic Provisioning	Supports flexible and dynamic allocation of GPU resources to meet workload demands.
Flexible Application scenarios for GPU and Network card configuration	Provides up to 10 dual-width slots for GPUs, or 8 dual-width GPU slots plus 4 single-width slots for NICs.
Two System Modes	Configurable in Standard Mode or Advanced Mode. Advanced Mode requires a Premium License and enables flexible, high-performance deployments.
Server Host Connections	Supports up to 4 x16 server host connections for enhanced data throughput.
GPU Peer-to-Peer (GPU P2P)	Enables high-speed communication directly between GPUs for optimized performance.
Host Connectivity	Utilizes 1m CDFP cables for reliable and high-bandwidth host connections.
Real-Time Monitoring	Provides real-time monitoring of GPU utilization, performance, and power consumption.

Fabric Manager Features

Table 3 details the Fabric Manager's advanced functionalities for GPU allocation, port configuration, and real-time performance monitoring.

Table 3. Fabric Manager Features

Feature	Description
GPU Composability and Dynamic Provisioning	Efficiently assign and unassign GPU resources as needed to match application demands.
GPU Fan-Out and Fabric Expansion Among Hosts	Scale GPU resources flexibly through fan-out under a single server or expand across multiple servers.
GPU Surprise Add and Remove	Seamlessly add or remove GPUs without causing system downtime.
Port Configuration	Configure ports for fabric, host, or device connectivity.
Inband Host GPU Information	Retrieve real-time downstream GPU data from host systems.
Host and GPU View	Provides detailed views of GPU assignments from both host and device perspectives.
Real-Time Performance and Error Counters	Monitor system performance in real-time and proactively identify potential issues.

System Management UI Features

Table 4 focuses on the System Management UI's tools for centralized management, real-time resource visualization, and role-based access control.

Table 4. System Management UI Features

Feature	Description
H3 Management Center	Provides a centralized interface for managing and monitoring the system.
Real-Time GPU Topology	Visualizes GPU clusters and their interconnections for clear resource mapping.
Dashboard for GPU Utilization and Performance	Monitors key metrics, GPU utilization, and overall system health in real-time.
Role-Based Authentication and Access Control	Secures system access with user-specific roles and permissions.

These features collectively enhance scalability, efficiency, and control, making the Falcon 5012 an adaptable solution for diverse workloads and environments. The following chapter will introduce the hardware specifications of the Falcon 5012.

Chapter 2.
Hardware Introduction

This chapter describes the Falcon 5012 system hardware, including its specifications, panel connections, LED indicators, and supported configurations. It also outlines GPU and network card deployment scenarios, compatibility requirements, and key accessories to ensure optimal system performance.

2.1 Hardware Overview

This section provides an overview of the Falcon 5012 system's hardware specifications. The system is designed to support high-performance workloads, leveraging advanced components such as an Intel ATOM X86 mCPU, Aspeed AST 2500 BMC, and Broadcom PCIe 5.0 CXL Switch. Key features include support for up to 700W GPUs, robust cooling with dual-rotor fans, and redundant hot-swappable power supplies. Detailed descriptions of these components and their configurations are provided in the following subsections.

2.1.1 Falcon 5012 Hardware Specifications

The Falcon 5012 integrates advanced components to support demanding workloads efficiently. Table 5 summarizes its key hardware, such as the Intel ATOM X86 mCPU, Broadcom PCIe 5.0 CXL Switch, and 700W GPU support. Features like redundant power supplies, conditioned hot-swappable fans, and high-speed 1m CDFP cables ensure reliability and performance. Refer to Table 5 for details.

Table 5. Falcon 5012 Hardware Specifications

Component	Description
1. mCPU	Intel ATOM X86 CPU
2. BMC	Aspeed AST 2500
3. PCIe Switch	Broadcom PEX 89144 PCIe 5.0 switch
4. Device	- Supports GPU, FPGA, network cards, and other PCIe add-in cards - Dual-slot width, 312mm length - Supports up to 700W GPU
5. Host Uplink	Flexible host port configuration
6. Slot Power	Four (4) pieces of 2100W or 3200W (2+2 or 3+1 redundant), back removable, hot-swap
7. Host Connection Cable	1m CDFP cable
8. Fan	- 8x 80x80mm dual rotor fans - Hot-swappable (conditioned)
9. Ethernet Ports	- RJ45 for mCPU GPU management - RJ45 for BMC chassis management
10. LED	- Power Stats LED - PCIe Link Speed LED (PCIe device status) - UID LED (Chassis location LED light) - System Status LED - AC Power LED - LAN Link/ Act LED - LAN Speed LED
11. Button	- 1x System Power Button - 1x UID Button
12. Operating Temperature	10°C to 35°C (50°F to 95°F)
13. Dimension	4U; 175(H) x 448(W) x 650(D) mm

The following sections introduce the functionalities and descriptions for the two application scenarios, starting with 2.1.2 Front Panel Connections and Controls, followed by 2.1.3 Rear Panel Connections and Controls.

2.1.2 Front Panel Connections and Controls

The front panel of the Falcon 5012 is equipped with multiple components to ensure efficient cooling, accessibility, and functionality. Figure 4 provides an overview of the front panel layout, highlighting key features such as the system fans, power and reset button, and UID controls.

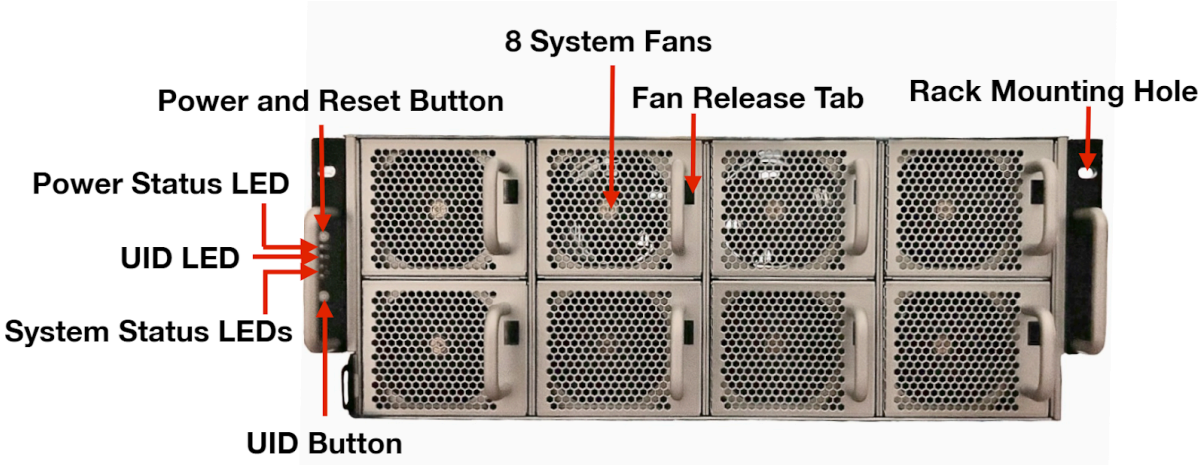


Figure 4. Image of Falcon 5012 front panel view

For a detailed breakdown of each component and its function, refer to Table 7, which summarizes the connections, indicators, and mechanisms essential for maintaining optimal system performance and ease of operation.

Table 7. Front Panel Connections and Controls

Control/Connection	Description
System Fans	8 cooling fans designed to maintain optimal operating temperatures by dissipating heat from the system.
Fans Release Tab	A tab mechanism used to easily release and remove the system fans for maintenance or replacement.
Rack Mounting Hole	Mechanism used for securely mounting the chassis into a server rack, ensuring stability and alignment.
Power and Reset Button	A button to power on/off the system or reset it when needed.
Power Status LED	Indicator to show the current power status of the system.
UID LED	Indicator to identify the Falcon 5012 location.
UID Button	A button to toggle the UID LED on or off, aiding in visually locating the device.
System Status LED	The LED provides visual feedback on the system's operational status: - Green: System is operating normally. - Black: System is not ready. - Red: Signals an error, such as fan failure or temperature issue.

The following section provides an introduction to the rear panel connections and controls, along with descriptions of each component.

2.1.3 Rear Panel Connections and Controls

The Falcon 5012 system offers versatile rear panel configurations tailored to different application requirements. The same chassis can be utilized in two distinct setups to address varying workloads:

---

10 GPUs Configuration:

This setup supports up to 10 dual-slot GPUs, ideal for compute-intensive scenarios requiring maximum GPU capacity. Each GPU slot utilizes PCIe 5.0 x16 for high-bandwidth connectivity. Refer to Figure 5 for an illustration of this configuration.

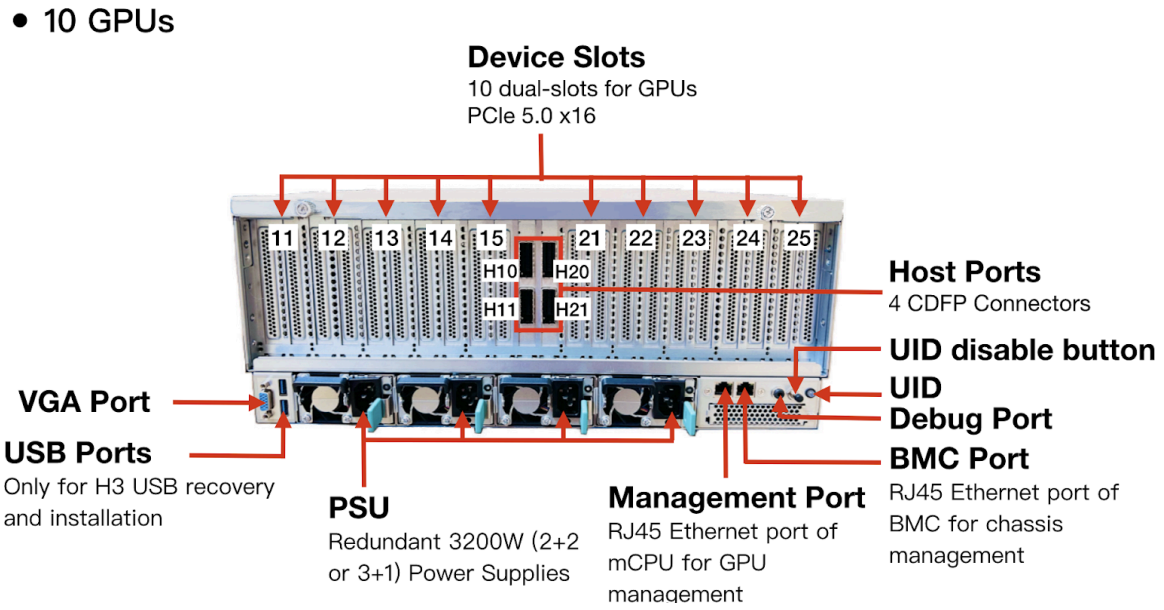


Figure 5. Image of Falcon 5012 rear panel view - 10 GPUs

---

8 GPUs + 4 NICs Configuration:

Designed for mixed compute and network demands, this configuration accommodates 8 Full Height Full Length (312mm) GPUs alongside 4 slots for

NICs. It provides a balanced solution for applications requiring both computational power and networking capabilities. Refer to Figure 6 for an illustration of this configuration.

- 8 GPUs + 4 NICs

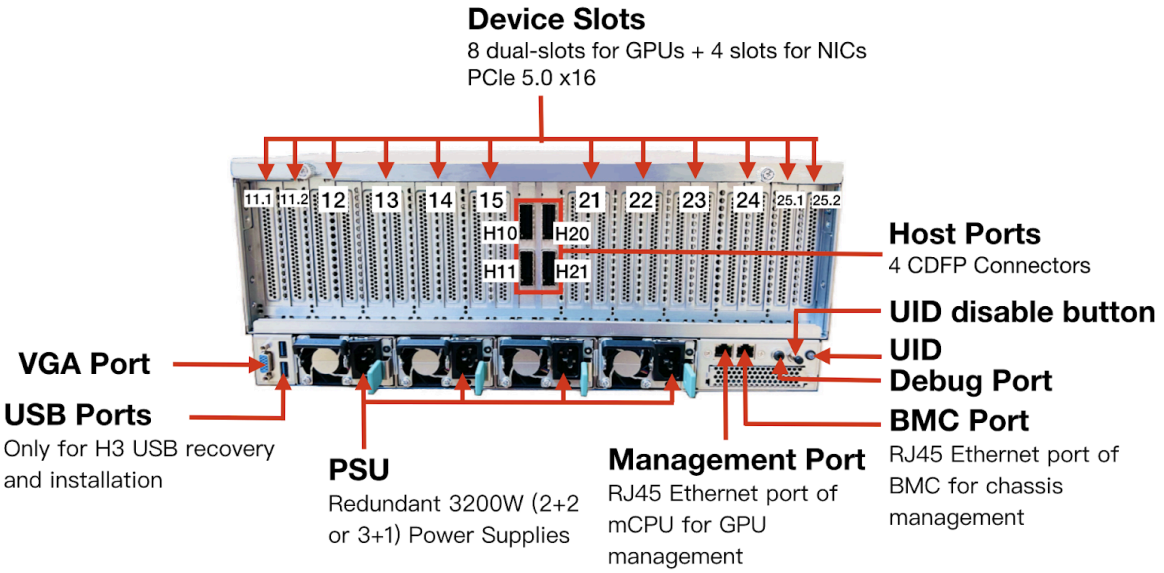


Figure 6. Image of Falcon 5012 rear panel view - 8 GPUs (FHFL 312mm) + 4 NICs

---

Both configurations utilize the Falcon 5012's advanced rear panel design, showcasing key features such as Host Ports, BMC Ports, and UID functionalities. Figures 4 and 5 illustrate the panel layout for 10 GPUs and 8 GPUs (FHFL 312mm) + 4 NICs configurations, respectively, highlighting their distinct slot arrangements and connection points. Table 6 provides a detailed breakdown of these rear panel components, explaining their roles in ensuring seamless system operation and management.

Table 6. Rear Panel Connections and Controls

Control/Connection	Description
Device Slots	Slots for installing GPUs, with two application scenarios: - 10 GPUs: 10 dual-slots for GPUs, supporting PCIe 5.0 x16. - 8 GPUs (FHFL 312mm) + 4 NICs: 8 dual-slots for FHFL 312mm GPUs, plus 4 slots for NICs, supporting PCIe 5.0 x16.
PCIe Link Speed LED	Indicator that displays the current PCIe link speed of the device.
Host Ports	4 CDFP connectors to connect hosts using CDFP cables (H10, H11, H20, H21).
UID Disable Button	Used to disable the UID function.
UID LED	The LED will glow blue to identify the Falcon 5012 location.
Debug Port	Reserved for diagnosing issues and troubleshooting hardware and firmware.
BMC Port	RJ45 Ethernet port of the Baseboard Management Controller (BMC) for chassis management.
Management Port	RJ45 Ethernet port to access the graphical user interface (GUI).
LAN Link ACT LED	Indicator that displays the current network condition.
LAN Speed LED	Indicator that displays the level of current LAN speed.
PSU	Redundant 2100W or 3200W power supplies with configurations of 2+2 or 3+1 for power redundancy.
USB Ports	For H3 USB recovery and system installation.
VGA Port	Connecting to a monitor to perform management system setup.

With the rear panel connections and controls covered, the following section provides a detailed overview of the Falcon 5012's LED indicators. These indicators are designed to convey critical system information, including power status, operational state, and link speed, ensuring users can monitor and troubleshoot the system efficiently.

2.1.4 LED Indicators

This section provides an overview of the LED indicators on the Falcon 5012, detailing their locations as shown in Figure 7 and explaining the significance of their colors and blinking patterns. For a comprehensive understanding of each indicator's status, refer to Table 8.

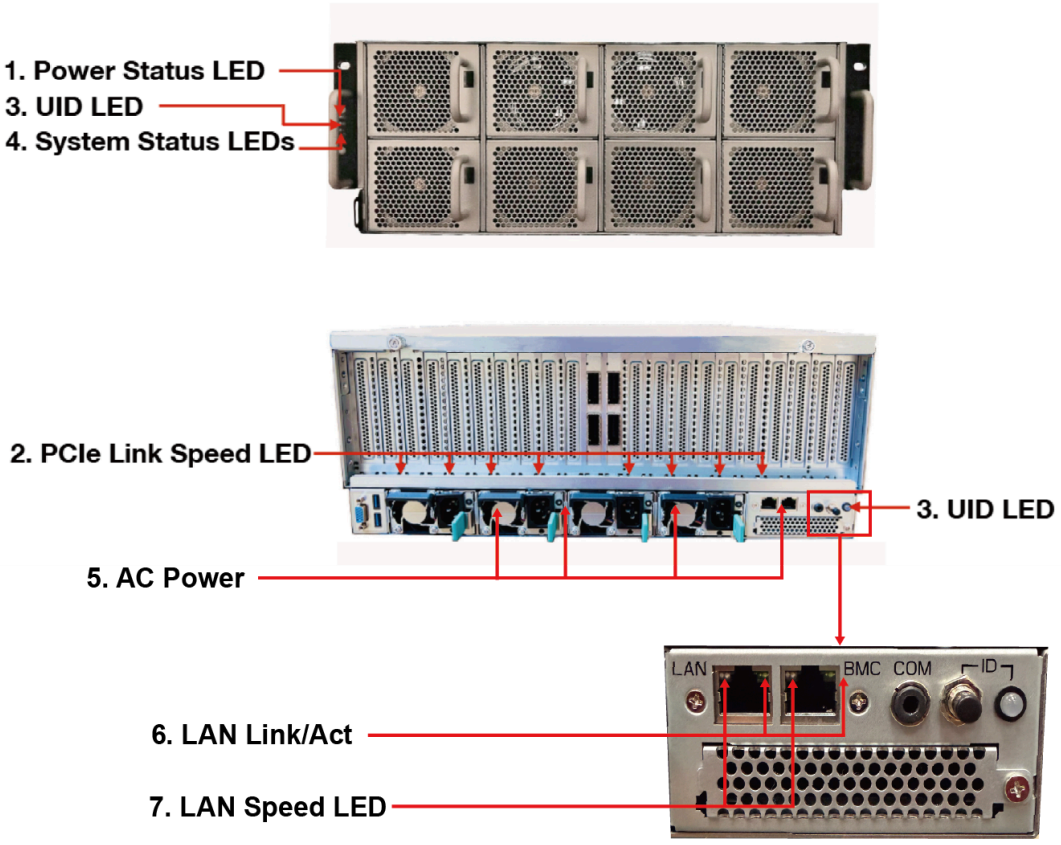


Figure 7. LED Indicators and Locations on the Falcon 5012

Table 8 below provides a detailed and comprehensive understanding of each indicator's status.

Table 8. System LED Indicators and Status

Item	Description	Status
1	Power Status LED	• Solid Green = Power, mCPU, and switch are ready • Off = Power is off

2	PCIe link Speed LED	• Solid Green = PCIe Gen 5 • Flashing Green at 1 Hz = PCIe Gen 4 • Flashing Green at 0.5 Hz = PCIe Gen 3 • Off = No link detected
3	UID LED	• Flashing blue at 0.5 Hz at 50% duty cycle = Activated • Off = Deactivated
4	System status LED	• Solid Green = System is operating normally • Flashing Green = The power is off • Solid Red= Error has occurred (e.g. fan failure, temperature issue) • Off = System is not ready
5	AC Power LED	• Solid Green = System is operating normally • Flashing Green = AC power is present or the system is in standby mode • Solid Amber = No power is detected from the AC cable
6	LAN Link/ Act LED	• Solid Green: System is connected to the network • Flashing Green: System is accessing the network • Off: System is not connected to the network
7	LAN Speed LED	• Solid Amber = System is operating at 1000BASE-T • Solid Green = System is operating at 100BASE-TX • Off = System is operating at 10BASE-T

2.1.5 PCBA Board Tray Components

This section provides an overview of the key components housed within the Falcon 5012 PCBA board tray. These components, which include GenZ connectors, PCIe switches, device slots, and the BMC and mCPU, form the backbone of the system's functionality, enabling efficient data processing, connectivity, and device management.

Refer to Figure 8 for a visual representation of the PCBA board tray and the locations of its primary components.

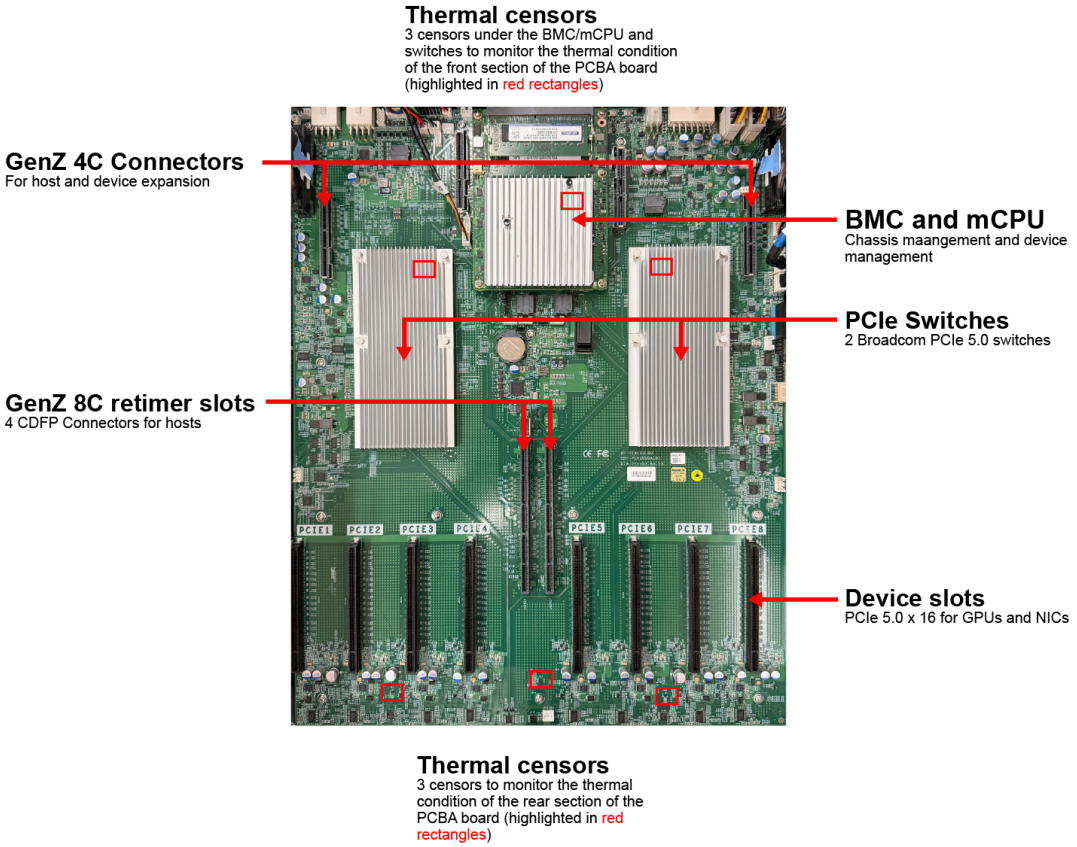


Figure 8. Image of Falcon 5012 PCBA board

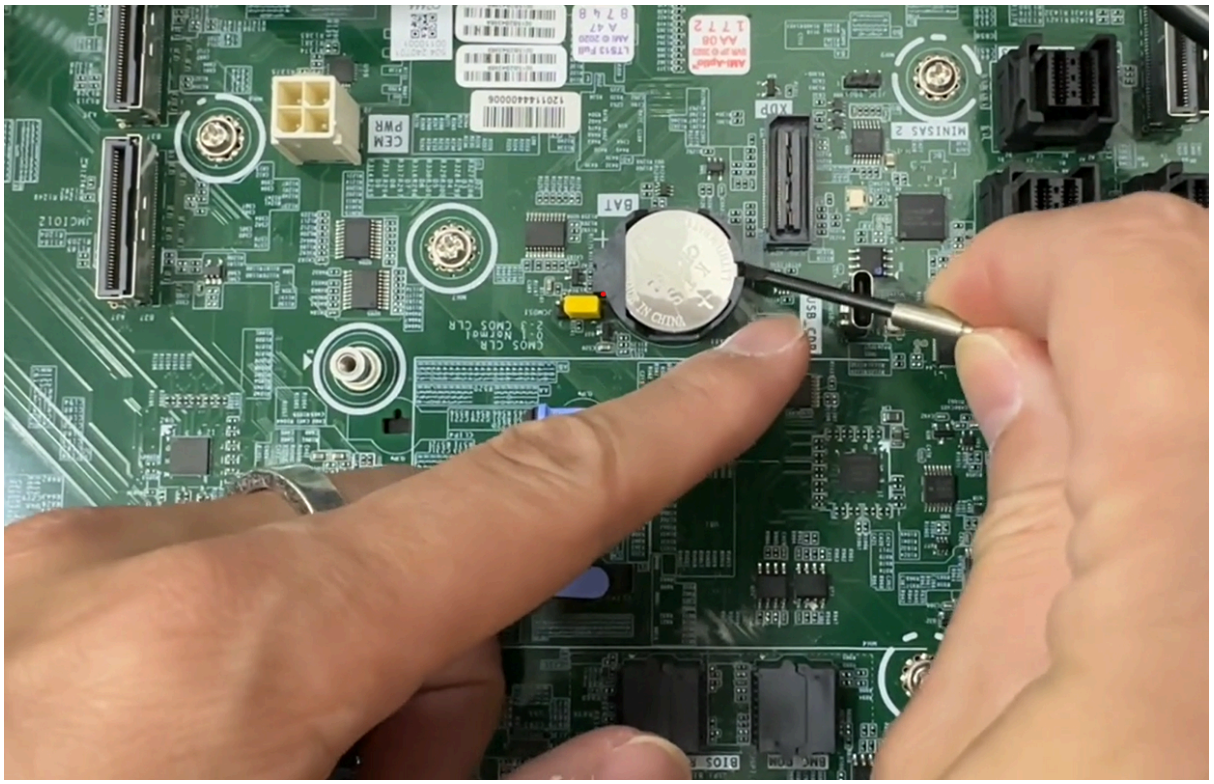
2.1.6 Replace Button Battery of the PCBA Board

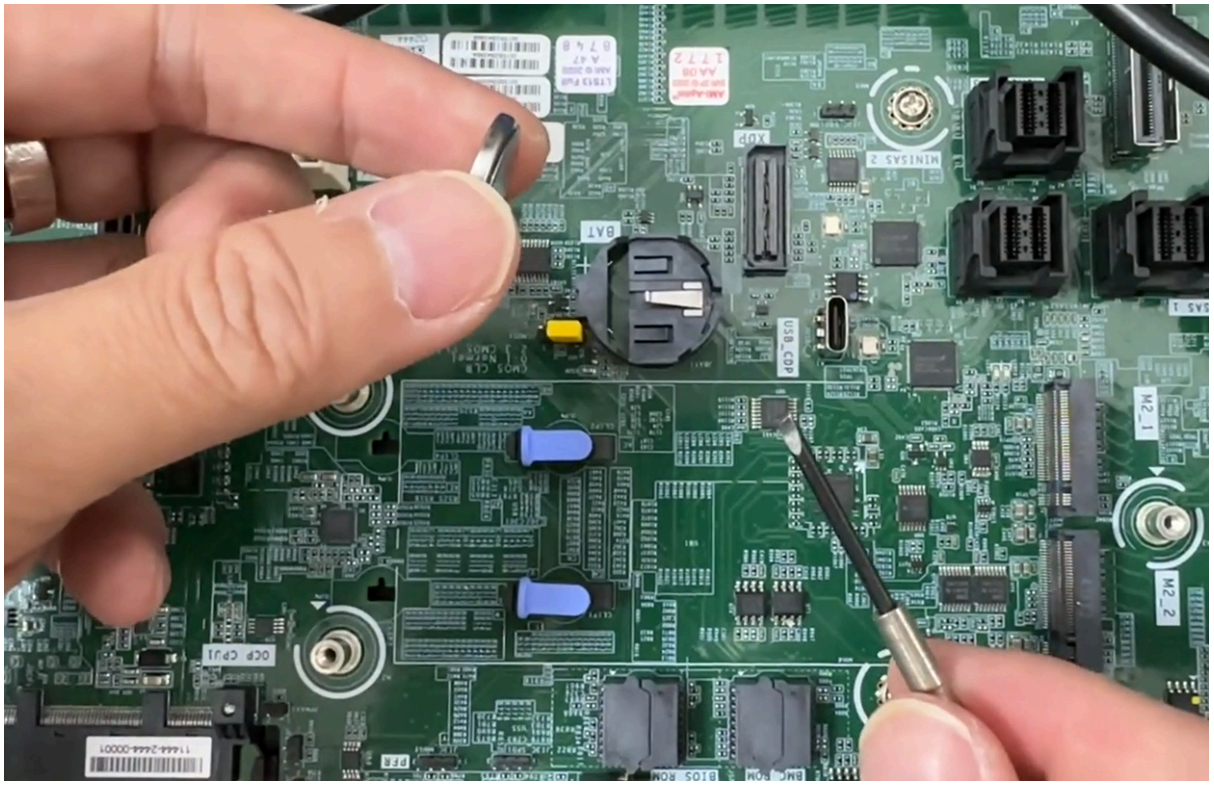
A button battery is installed on the PCBA board to provide backup power to low-power circuits such as the RTC and configuration memory. It preserves system time, event logs, and critical settings when the main power is removed.

The ideal temperature for operation of the battery is between **10 to 25 degrees** ℃. The suggested battery replacement intervals are as follows:

- **3 years** if the chassis remains disconnected from AC power (no power supply).
- **5 years** if the chassis remains connected to AC power in standby mode.

To replace the button battery, please use a screwdriver and refer to the images below. Please note that replacing the button battery will clear the time settings.





Section 2.1 reviewed the Falcon 5012 hardware, including its specifications, panel connections, LED indicators, and PCBA board tray components.
Section 2.2 will explore the System Topology, highlighting the interconnect architecture, data flow, and component integration driving its high-performance capabilities.

2.2 System Topology

The Falcon 5012 system supports the following configuration options to accommodate various deployment requirements:

- **10 GPUs**
- **8 GPUs + 4 Network Cards**

The system can be configured in two modes:

- **Standard Mode: Default mode providing essential functionality for general workloads.**
- **Advanced Mode: Unlocks additional flexibility and high-performance capabilities. A Premium License is required to enable Advanced Mode.**

Advanced Mode is ideal for demanding workloads that require optimized resource utilization and enhanced system performance.

---

1. 10 GPUs

Table 9 summarizes the 10 GPU deployment options. Standard Mode supports 2 Hosts + 10 GPUs, while Advanced Mode offers greater flexibility with 4 Hosts + 10 GPUs or 6 Hosts + 8 GPUs.

Table 9. Deployment Configurations for 10 GPUs in Standard and Advanced Modes

Application scenarios	Standard Mode	Advanced Mode
10 GPUs	a. 2 Hosts + 10 GPUs	b. 4 Hosts + 10 GPUs
		c. 6 Hosts + 8 GPUs

2. 8 GPUs+ 4 Network cards

Table 10 summarizes the 8 GPUs + 4 Network Cards deployment options. Standard Mode supports 2 Hosts + 8 GPUs + 4 Network Cards, while Advanced Mode provides more flexibility with configurations such as 4 Hosts + 8 GPUs + 4 Network Cards, 6 Hosts + 8 GPUs + 2 Network Cards, or 8 Hosts + 8 GPUs.

Table 10. Deployment Configurations for 8 GPUs+ 4 Network cards in Standard and Advanced Modes

Application scenarios	Standard Mode	Advanced Mode
8 GPUs + 4 Network cards	d. 2 Hosts + 8 GPUs + 4 Network cards	e. 4 Hosts + 8 GPUs + 4 Network cards
		f. 6 Hosts + 8 GPUs + 2 Network cards
		g. 8 Hosts + 8 GPUs

The following sections will be divided into 2.2.1: 10 GPUs and 2.2.2: 8 GPUs (FHFL 312mm) + 4 NICs, using diagrams to introduce each configuration.

2.2.1 10 GPUs

a. Standard Mode—2 Hosts + 10 GPUs

In the 10 GPUs scenario, each host independently accesses the GPUs connected to its respective switch under Standard Mode. As shown in Table 11, the chassis-wide configuration supports 2 hosts + 10 GPUs, while from the perspective of a single PCIe switch, the configuration is 1 host + 5 GPUs.

Table 11. 10 GPUs Standard Mode Topology: 2 Hosts + 10 GPUs

Configuration	Chassis-wide configuration	Per-switch configuration
Host port + Device port	2 Hosts + 10 GPUs	1 Host + 5 GPUs

Figure 9 shows the topology, with PCIe Switch 1 connecting to GPUs for H10 and PCIe Switch 2 for H20, ensuring isolated resources for optimized performance.

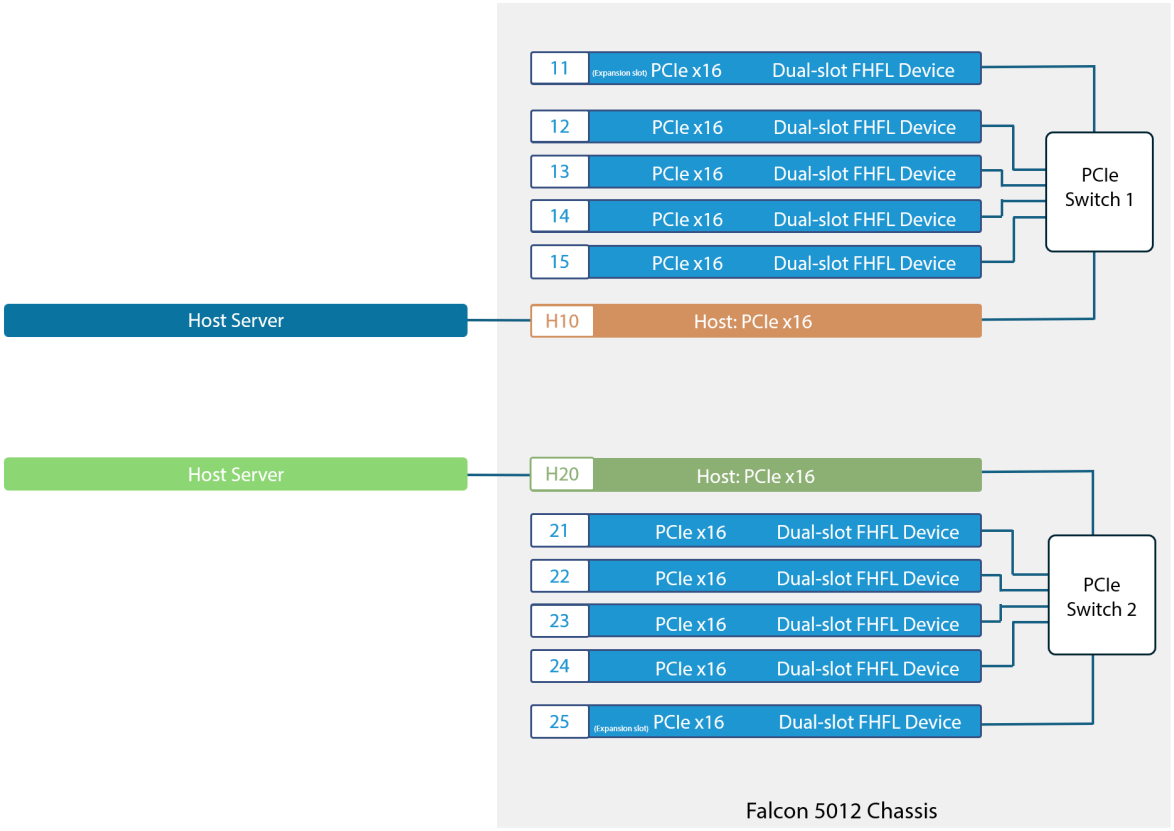


Figure 9. Block diagram of Standard Mode architecture topology with 2 x16 hosts and 10 GPUs

---

In the 10 GPUs application scenario under Advanced Mode, two configurations are available.

b. Advanced Mode — 4 Hosts + 10 GPUs

In 10 GPUs Advanced Mode, each PCIe switch can support at least 2 hosts and 5 GPUs, enabling the system to connect up to 4 hosts and 10 GPUs. Hosts can independently access GPUs on their respective switches and share GPUs within the same switch, allowing dynamic allocation based on host needs. Table 12 shows the chassis-wide configuration as 4 hosts + 10 GPUs, with each switch supporting 2 hosts + 5 GPUs.

Table 12. 10 GPUs Advanced Mode Topology: 4 Hosts + 10 GPUs

Configuration	Chassis-wide configuration	Per-switch configuration
Host port + Device port	4 Hosts + 10 GPUs	2 Host + 5 GPUs

Figure 10 illustrates the topology, where PCIe Switch 1 connects GPUs to H10 and H11, and PCIe Switch 2 connects GPUs to H20 and H21. This design ensures resource isolation while enabling dynamic GPU sharing among hosts for optimal performance.

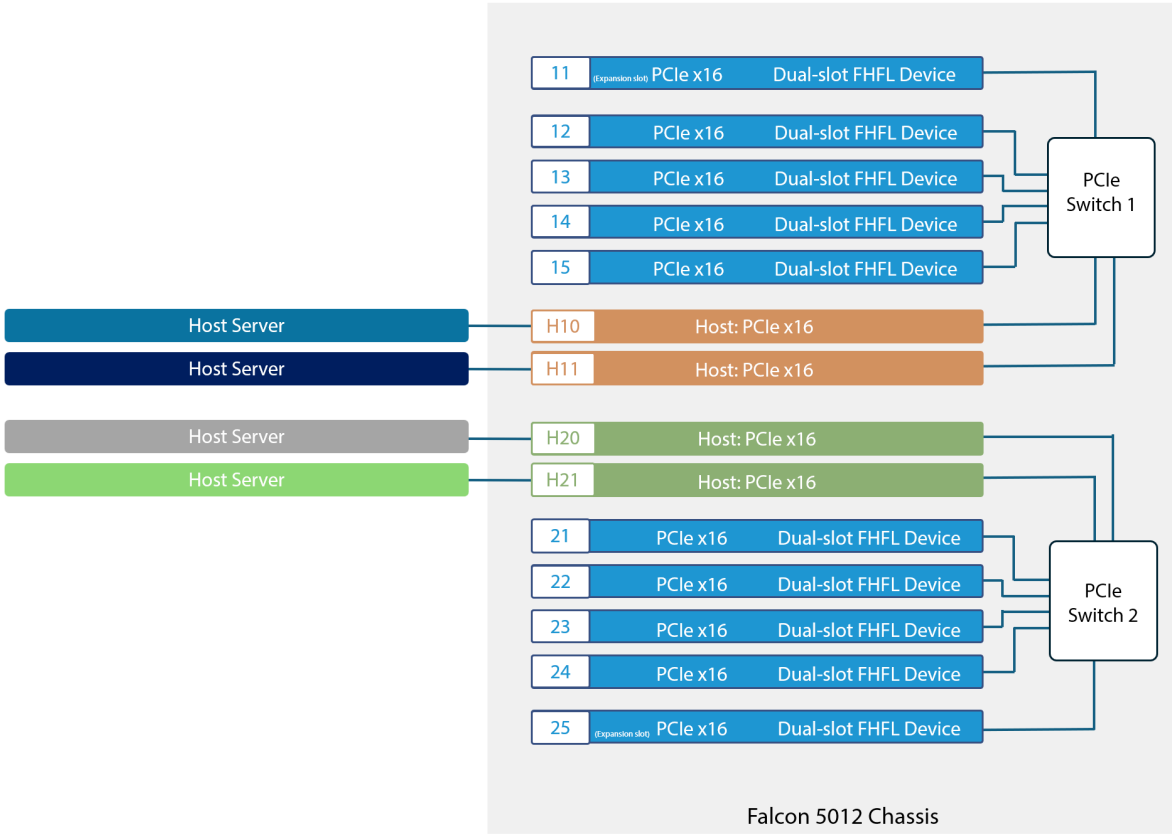


Figure 10. Block diagram of Standard Mode architecture topology with 4 x16 hosts and 10 GPUs

c. Advanced Mode — 6 Hosts + 8 GPUs

In 10 GPUs Advanced Mode, each PCIe switch can support 3 hosts and 4 GPUs, enabling the system to connect up to 6 hosts and 8 GPUs. Hosts can independently access GPUs on their respective switches and share GPUs within the same switch, allowing dynamic allocation based on host needs. Table 13 shows the chassis-wide configuration as 6 hosts + 8 GPUs, with each switch supporting 3 hosts + 4 GPUs.

Table 13. 10 GPUs Advanced Mode Topology: 6 Hosts + 8 GPUs

Configuration	Chassis-wide configuration	Per-switch configuration
Host port + Device port	6 Hosts + 8 GPUs	3 Host + 4 GPUs

Figure 11 illustrates the topology, where PCIe Switch 1 connects GPUs to H10 and H11 while converting an expansion cable into a host port to connect a third host. Similarly, PCIe Switch 2 connects GPUs to H20, H21, and an additional expanded host. This design ensures resource isolation while supporting multi-host GPU sharing and dynamic allocation for optimized performance.

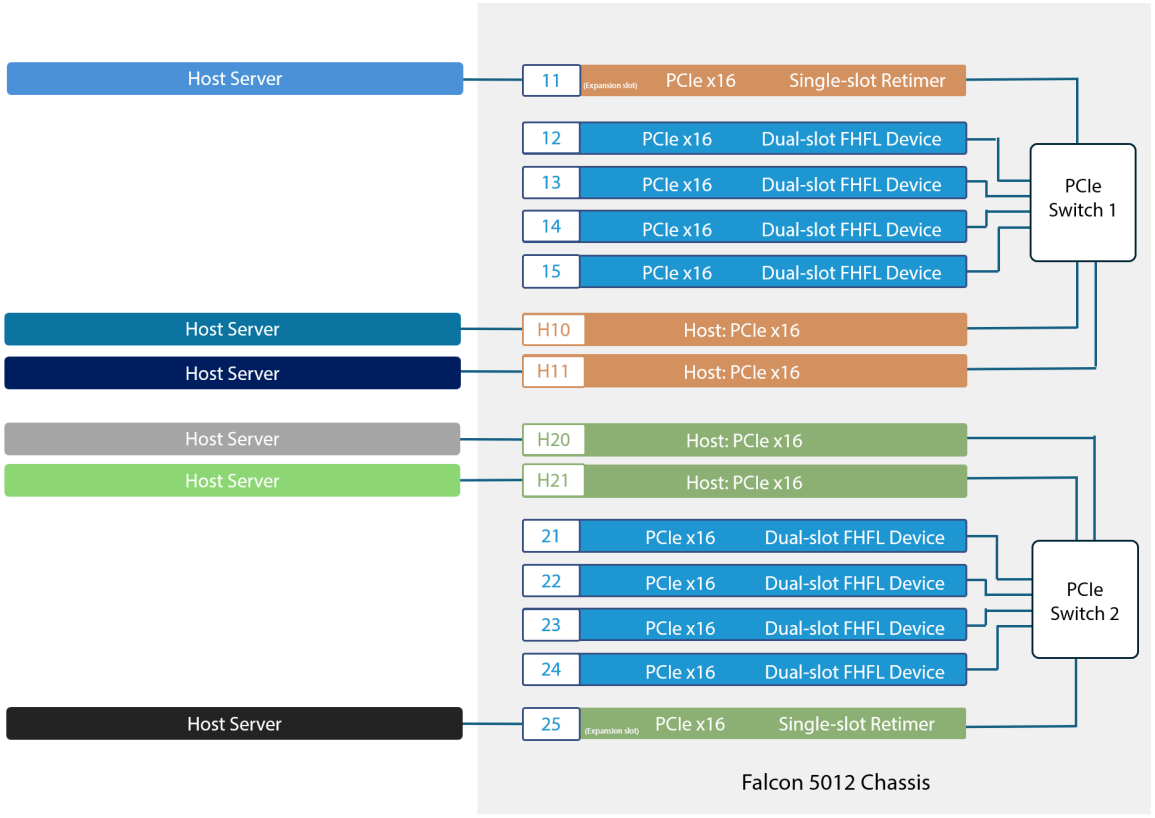


Figure 11. Block diagram of Advanced Mode architecture topology with 6 x16 hosts and 8 GPUs

---

2.2.2 8 GPUs (FHFL 312mm) + 4 Network cards

d. Standard Mode — 2 Hosts + 8 GPUs + 4 Network cards (NICs)

In the 8 GPUs + 4 Network cards scenario, each host independently accesses the GPUs and Network cards connected to its respective switch under Standard Mode. As shown in Table 14, the chassis-wide configuration supports 2 Hosts + 8 GPUs + 4 Network cards, while from the perspective of a single PCIe switch, the configuration is 1 Host + 4 GPUs + 2 Network cards.

Table 14. 8 GPUs + 4 NICs Advanced Mode Topology: 2 Hosts + 8 GPUs + 4 NICs.

Configuration	Chassis-wide configuration	Per-switch configuration
Host port + Device port	2 Hosts + 8 GPUs + 4 Network cards	1 Host + 4 GPUs + 2 Network cards

Figure 12 shows the topology, with PCIe Switch 1 connecting to 4 GPUs and 2 single-width Network cards (or GPUs) for H10 and PCIe Switch 2 for H20, ensuring isolated resources for optimized performance.

Figure 12. Block diagram of Standard Mode architecture topology with 2 x16 hosts and 8 GPUs + 4 NICs

---

In the 8 GPUs + 4 Network cards application scenario under Advanced Mode, three configurations are available.

e. Advanced Mode — 4 Hosts + 8 GPUs + 4 Network cards

In 8 GPUs + 4 Network cards Advanced Mode, each PCIe switch can support 2 hosts and 4 GPUs plus 2 Network cards, enabling the system to connect up to 4 hosts and 8 GPUs plus 4 Network cards. Hosts can independently access GPUs and Network cards on their respective switches and share resources within the same switch, allowing dynamic allocation based on host needs. Table 15 shows the chassis-wide configuration as 4 hosts + 8 GPUs + 4 Network cards, with each switch supporting 2 hosts + 4 GPUs + 2 Network cards.

Table 15. 8 GPUs + 4 NICs Advanced Mode Topology: 4 Hosts + 8 GPUs + 4 NICs.

Configuration	Chassis-wide configuration	Per-switch configuration
Host port + Device port	4 Hosts + 8 GPUs + 4 Network cards	2 Hosts + 4 GPUs + 2 Network cards

Figure 13 illustrates the topology, where PCIe Switch 1 connects GPUs and Network cards to H10 and H11, and PCIe Switch 2 connects GPUs and Network cards to H20 and H21. This design ensures resource isolation while enabling dynamic GPU and Network card sharing among hosts for optimal performance.

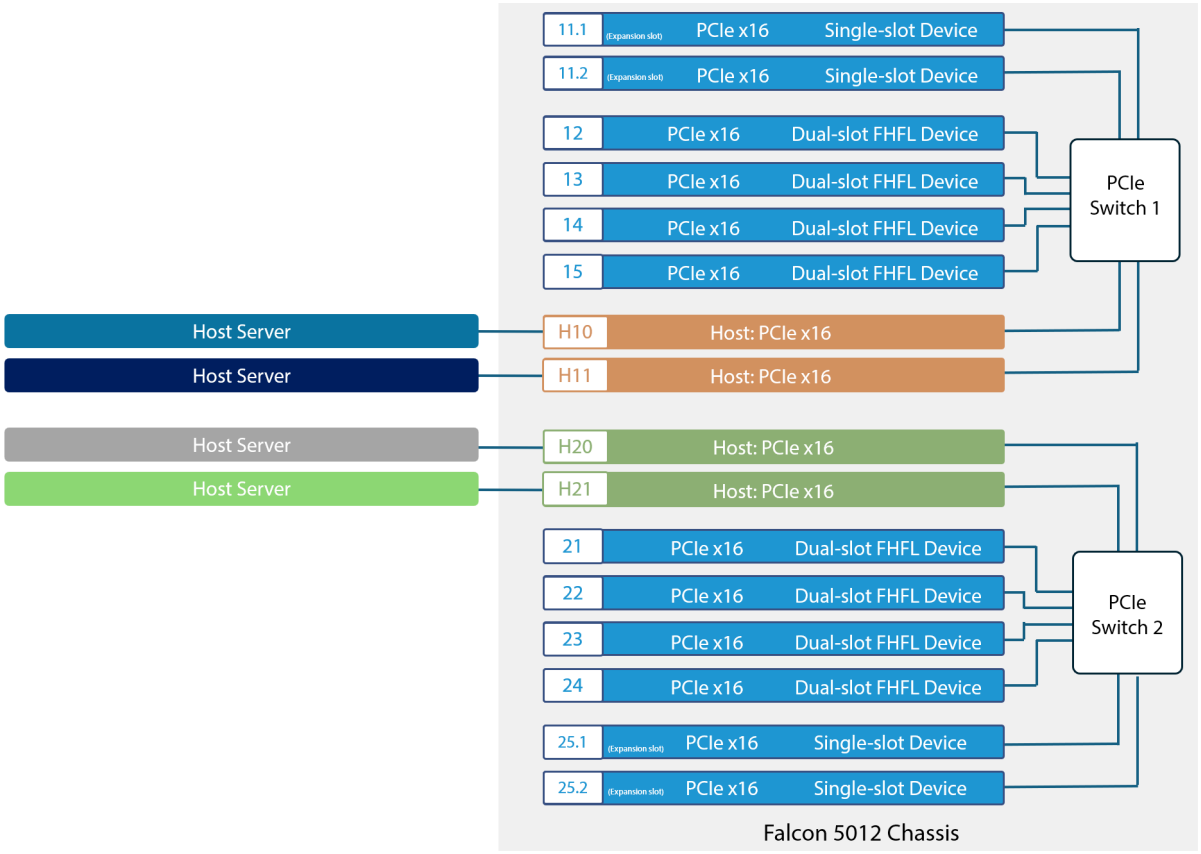


Figure 13. Block diagram of Advanced Mode architecture topology with 4 x16 hosts and 8 GPUs + 4 NICs

---

f. Advanced Mode — 6 Hosts + 8 GPUs + 2 Network cards

In 8 GPUs + 4 Network cards Advanced Mode, each PCIe switch can support 3 hosts and 4 GPUs plus 1 Network cards, enabling the system to connect up to 6 hosts and 8 GPUs plus 2 Network cards. Hosts can independently access GPUs and Network cards on their respective switches and share resources within the same switch, allowing dynamic allocation based on host needs. Table 16 shows the chassis-wide configuration as 6 hosts + 8 GPUs + 2 Network cards, with each switch supporting 3 hosts + 4 GPUs + 1 Network card.

Table 16. 8 GPUs + 4 NICs Advanced Mode Topology: 6 Hosts + 8 GPUs + 2 NICs.

Configuration	Chassis-wide configuration	Per-switch configuration
Host port + Device port	6 Hosts + 8 GPUs + 2 Network cards	3 Hosts + 4 GPUs + 1 Network card

Figure 14 illustrates the topology, where PCIe Switch 1 connects GPUs to H10 and H11 while converting an expansion cable into a host port to connect a third host. Similarly, PCIe Switch 2 connects GPUs to H20, H21, and an additional expanded host. This design ensures resource isolation while supporting multi-host GPU sharing and dynamic allocation for optimized performance.

Figure 14. Block diagram of Advanced Mode architecture topology with 6 x16 hosts and 8 GPUs + 2 NICs

g. Advanced Mode — 8 Hosts + 8 GPUs

In 8 GPUs + 4 Network cards Advanced Mode, each PCIe switch can support 4 hosts and 4 GPUs, enabling the system to connect up to 8 hosts and 8 GPUs. Hosts can independently access GPUs and on their respective switches and share resources within the same switch, allowing dynamic allocation based on host needs. Table 16 shows the chassis-wide configuration as 8 Hosts + 8 GPUs, with each switch supporting 4 Hosts + 4 GPUs.

Table 17. 8 GPUs + 4 NICs Advanced Mode Topology: 8 Hosts + 8 GPUs

Configuration	Chassis-wide configuration	Per-switch configuration
Host port + Device port	8 Hosts + 8 GPUs	4 Hosts + 4 GPUs

Figure 14 illustrates the topology, where PCIe Switch 1 connects GPUs to H10 and H11 while converting two expansion cables into host ports to connect two other hosts. Similarly, PCIe Switch 2 connects GPUs to H20, H21, and additional expanded hosts. This design ensures resource isolation while supporting multi-host GPU sharing and dynamic allocation for optimized performance.

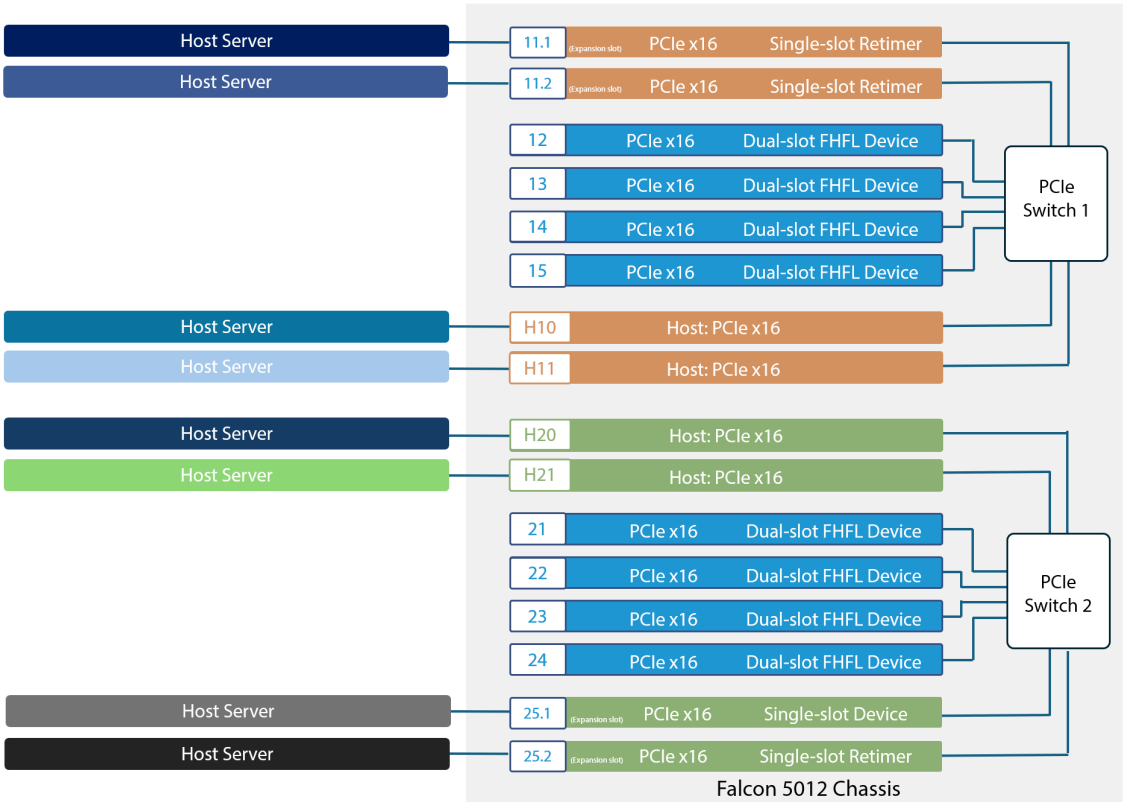


Figure 15. Block diagram of Advanced Mode architecture topology with 8 x16 hosts and 8 GPUs

---

For more information, please contact sales@h3platform.com for license purchase.

2.3 Compatibility List

2.3.1 GPU Compatibility List

Before installing any GPU on the Falcon 5012, ensure the GPU is listed in the compatibility list below (Table 18). These GPUs have been tested and verified by our team to ensure proper functionality and performance.

Note: Only GPUs listed in Table 18 are supported. Using unsupported GPUs may result in damage to the Falcon 5012 system. H3 Platform will not be responsible for any damage caused by using non-validated GPUs.

Table 18. GPU Compatibility List

GPU Module	Description
NVIDIA H100	Supported GPU Module
NVIDIA RTX 4080	Supported GPU Module
NVIDIA RTX 4090	Supported GPU Module
Xilinx Alveo™ U50	Supported GPU Module

The compatibility list is subject to updates. For the latest validated GPUs, please visit our company website.

Note: The Falcon 5012 system by default supports GPUs with a 16-pin power connector. If you need to use a GPU that requires an 8-pin power connector, please contact sales@h3platform.com for technical support.

2.3.2 Host CPU Compatibility List

The Falcon 5012 system supports the following host CPUs in Table 19. Use of unsupported CPUs may result in hardware damage. H3 Platform is not liable for any damage caused by incompatible CPUs.

Table 19. Host CPU Compatibility List

Host CPU	Requirements
Intel	5th Gen Xeon Scalable server processor or later (Code name: Emerald Rapids)
AMD	4th Generation AMD EPYC™ Processors OR LATER (Code name: Genoa)

2.3.3 Host OS Compatibility List

The Falcon 5012 system supports the following host operating systems in Table 20. Use of unsupported operating systems may result in hardware damage. H3 Platform is not liable for any damage caused by incompatible operating systems.

Table 20. Host OS Compatibility List

Host OS	Requirements
Ubuntu	16.04 LTS, 18.04 LTS, 20.04 LTS, 24.04.3 LTS
Windows	Build 1903–20H2
CentOS / RHEL	7.3–8.0

2.3.4 Host BIOS Setting

The 5012 retimer clock mode supports only SRNS (Separate Reference Clock with no SSC), a clocking scheme designed for high-performance and stable operations in PCIe environments. To ensure compatibility and proper operation of the 5012 retimer, both SSC (Spread Spectrum Clocking) and SRIS (Separate Reference Clock with Independent SSC) must be disabled in the Server BIOS. This guide provides detailed instructions for configuring BIOS settings on Intel EMR and AMD Turin platforms to meet these requirements.

About the Task

The objective of this task is to configure the Server BIOS to support the 5012 retimer clock mode. This involves:

- Disabling SSC and SRIS clocking schemes.
- Optionally configuring Port Clocking for specific platforms.

These changes ensure the system operates in compliance with the clocking requirements of the 5012 retimer.

Procedures

To configure the 5012 retimer clock mode properly, refer to Table 21 and follow these steps to disable the relevant clocking schemes:

Table 21. Clocking Configuration Requirements for Intel EMR and AMD Turin BIOS Settings

Clocking Configuration	Intel EMR BIOS Setting	AMD Turin BIOS Setting
SSC (Spread Spectrum Clocking)	Platform Configuration > Miscellaneous Configuration > SSC = Disabled	CBS > FCH > Miscellaneous > FCH Spread Spectrum = Disabled
SRIS (Separate Reference Clock with Independent SSC)	Socket Configuration > IIO Configuration > Socket Configuration > Port > SRIS = Disabled	CBS > NBIO > PCIe > SRIS = Disabled
Port Clocking (Optional)	Socket Configuration > IIO Configuration > Socket Configuration > Port > Port Clocking = Distinct	N/A

Detailed Instructions for Intel EMR Platforms

Step 1. Disabling SSC (Spread Spectrum Clocking):

A. Enter the BIOS setup by pressing the designated key (e.g., F2, DEL, or ESC) during system startup.
B. Navigate to: Platform Configuration > Miscellaneous Configuration.
C. Locate the SSC option and set it to Disabled.
D. Save changes and exit the BIOS.

Step 2. Disabling SRIS (Separate Reference Clock with Independent SSC):

A. Re-enter the BIOS setup.
B. Navigate to: Socket Configuration > IIO Configuration > Socket Configuration > Port.
C. Locate the SRIS option and set it to Disabled.
D. Save changes and exit the BIOS.

Step 3. Configuring Port Clocking (Optional):

A. If required, re-enter the BIOS setup.
B. Navigate to: Socket Configuration > IIO Configuration > Socket Configuration > Port.
C. Locate the Port Clocking option and set it to Distinct.
D. Save changes and exit the BIOS.

Detailed Instructions for AMD Turin Platforms

Step 1. Disabling SSC (Spread Spectrum Clocking):

A. Enter the BIOS setup by pressing the designated key (e.g., F2, DEL, or ESC) during system startup.
B. Navigate to: CBS > FCH > Miscellaneous.
C. Locate the FCH Spread Spectrum option and set it to Disabled.
D. Save changes and exit the BIOS.

Step 2. Disabling SRIS (Separate Reference Clock with Independent SSC):

A. Re-enter the BIOS setup.
B. Navigate to: CBS > NBIO > PCIe.
C. Locate the SRIS option and set it to Disabled.
D. Save changes and exit the BIOS.

Verification

After completing the BIOS configuration:

A. Boot the server and re-enter the BIOS to verify the changes.
B. Confirm that both SSC and SRIS are disabled.
C. If configured, ensure Port Clocking is set to Distinct.
D. Use diagnostic tools or the system's management interface to confirm proper operation.

Notes

- BIOS interfaces may vary slightly depending on the system manufacturer. Consult your hardware documentation if options are not found in the specified paths.
- Incorrect BIOS settings may cause system instability. Verify all changes carefully before saving.

For additional assistance, contact your hardware support team.

2.4 Accessory List

The following accessories are essential for the Falcon 5012 chassis in the PCIe Gen5 GPU solution. These components ensure seamless connectivity and optimal system performance:

- **CDFP Connection Cable:** Provides high-speed data transfer between the host system and the Falcon 5012 chassis.
- **Host Adapter Card (Retimer Card):** Enables integration with the host system, supporting PCIe Gen5 for enhanced bandwidth and performance.

For additional details, including specifications and images, refer to the sections below.

2.4.1 1m Connection Cables

The connection cable features a CDFP to CDFP interface, specifically designed to support PCIe Gen5 x16 bandwidth per cable. As shown in the Table 22, It is constructed from copper to ensure durability and optimal performance. Measuring 1 meter in length, the cable delivers reliable, high-speed connectivity between the chassis and host servers.

Table 22. CDFP Connection Cable Specifications

Specification	Details
Interface	CDFP
Connector	CDFP to CDFP
Bandwidth	PCIe Gen5 x16 (per cable)
Type	Copper
Length	1m

Figure 16 below provides a visual reference of the 1-meter CDFP connection cable, highlighting its design and connection points for ease of identification.

Figure 16. Picture of 1m CDFP cable

2.4.2 Host Adapter Card (Retimer Card)

The PCIe Gen5 GPU solution utilizes host adapter cards with integrated retimer functionality to support high-speed signaling between the host system and expansion components. Two variants of the host adapter card are available.

Host Adapter Card Overview

These adapter cards are designed to be installed into a standard PCIe Gen5 x16 slot on a host server.
Each card supports high-speed PCIe signaling over CDFP to connect the host server to downstream expansion chassis, enabling low-latency, high-bandwidth data exchange with external devices.

Specifications

The following table provides the mechanical and interface specifications of the host adapter cards.

Table 23. Host Adapter Card Specifications

Specification	Description
Form Factor	PCIe Gen5 x16
Connector	CDFP
Dimensions	160 mm (L) × 70 mm (H)

Note: These host adapter cards support full PCIe Gen5 signal integrity requirements and include on-board retimers for long-reach link extension.

Host Adapter Variants

Two versions of the host adapter card are available:

- **C01 Variant:** An earlier-generation design used in initial PCIe Gen5 GPU configurations.

- **D01 Variant:** A newer revision of the host adapter card, featuring updated board layout while maintaining identical electrical and functional characteristics.

Host Adapter Card Images

The following figures illustrate the top and bottom views of the C01 and D01 host adapter card variants.
The card model identifier is printed on the bottom side for reference.

Figure 17. C01 Host Adapter Card – Top View

Figure 18. C01 Host Adapter Card – Bottom View

Figure 19. D01 Host Adapter Card – Top View

Figure 20. D01 Host Adapter Card – Bottom View

Note: The physical layout differs slightly between the C01 and D01 revisions, but both cards share the same mechanical dimensions and connector interfaces.

Attention: Ensure proper cooling and airflow when installing these host adapter cards, as high-speed signaling components can generate significant heat under load.

Chapter 3.

Installation and Boot-Up Guide

3.1 Confirm Label and Product Match

This procedure provides the steps to verify the Falcon 5012 system's configuration to ensure it matches your requirements. The process includes identifying the SKU, confirming the PSU model, and verifying that the selected mode (Standard (std) or Advanced (adv)) aligns with your requested configuration and included accessories.

About this Task

To ensure the system is configured correctly, you will:

1. Identify the application scenarios:
    - Determine whether the system is configured for 10 GPUs (SKU A) or 8 GPUs + 4 NICs (SKU B).
2. Verify the mode:
    - Confirm whether the system is in Standard (std) or Advanced (adv) mode by checking the included accessories (Retimers and CDFP Cables).
3. Confirm the PSU model:
    - Verify that the installed PSU matches the power requirements (2100W or 3200W).

Notes: Ensure that all accessories match the configuration type indicated by the SKU before proceeding.

Tools and Materials

You will need:

- Access to the SKU label on the external packaging.
- Access to the system chassis for internal verification.

Procedures

Step 1. Identifying the SKU

A. Locate the SKU label on the external packaging (refer to Figure 21).

Figure 21. SKU Label on the Carton (Early & Current Version)

B. Use the following elements in Table 24 to interpret the SKU label:

Table 24. PN Elements and Interpretations for Falcon 5012

PN Elements		Interpretations
Early version	Current version	
“PRPF5012”	“PRG5”	Product - Falcon 5012 (PCIe Gen5 Chassis)
“10” or “08”	“2” or “4”	Application scenarios (SKU): - 10 GPUs (SKU A - 2 Cable CEMs) - 8 GPUs + 4 NICs (SKU B - 4 Cable CEMs)
“STD” or “ADV”	“2” or “4”	- Standard mode (2 host retimers and cables) - Advanced mode (4 host retimers and cables)
“2100” or “3200”	“2” or “3”	PSU rating 2100W or 3200W
N/A	“3” or later	Chassis version (This digit will change as the chassis version is updated)

Step 2. Verify SKU Configuration: Open the chassis and check the expansion cable set (Cable CEMs) count (refer to Figure 22):

- 10 GPU (SKU A): Requires 2 expansion cable sets.
- 8 GPUs + 4 NICs (SKU B): Requires 4 expansion cable sets.

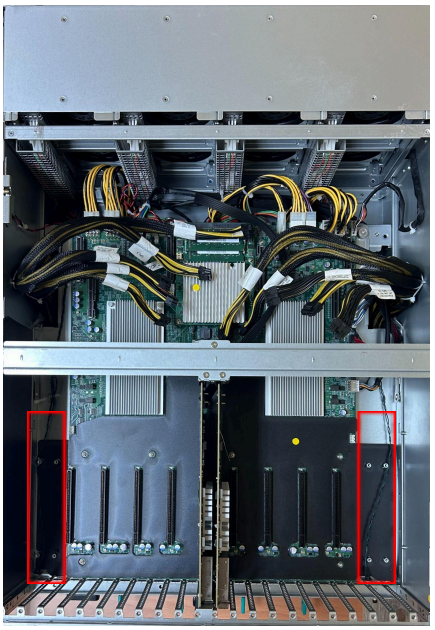


Figure 22. 10 GPUs with 2 expansion cable sets (cable CEMs)

Step 3. Confirming the Standard (std) and Advanced (adv) Mode Accessories

A. Check the included accessories according to Table 25 to ensure they match your configuration:

Table 25. Accessory Comparison Between Standard (std) and Advanced (adv) Modes

Configuration	Host Retimers	CDFP Cables
Standard (std)	2	2
Advanced (adv)	4	4

B. Verify the following:

- Standard Mode: Confirm the presence of 2 Retimers and 2 CDFP Cables.
- Advanced Mode: Confirm the presence of 4 Retimers and 4 CDFP Cables.

C. Compare the received accessories with the table above to confirm consistency.

Step 4. Confirming the PSU Model

A. Open the chassis and locate the installed PSU.
B. Verify that the PSU matches one of the supported models:
   - 2100W
   - 3200W

Results

- If the SKU, PSU model, and included accessories match your requested configuration, the system is ready for installation.
- If discrepancies are identified, contact technical support for further assistance.

3.2 Hardware Setup and Power-On Preparation

Set up the Falcon 5012 system by connecting peripherals and preparing the hardware for power-on.

About this task:

This task involves connecting necessary components and ensuring the system is ready for initialization.

Procedures:

Step 1. Connect the following peripherals (refer to Figure 23)

- Monitor via VGA port.
- Keyboard via USB port.
- LAN cables for GUI and BMC access.

> **Attention:**
> Ensure the LAN port indicator light is illuminated after connecting the network cable. This confirms successful communication.

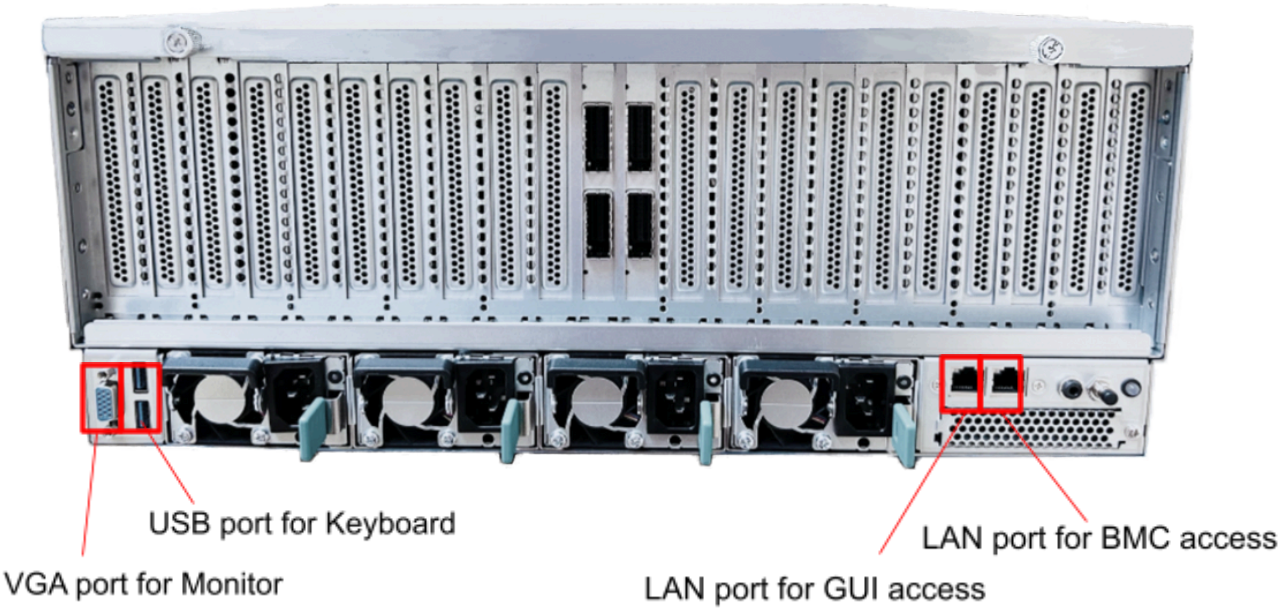


Figure 23. Rear Panel Connections for VGA, USB, and LAN Ports

Step 2. Plug in the Power Cord

Connect the power cord to the system (refer to Figure 24 and observe the UID light (refer to Figure 25):

A. The UID light should blink twice and then turn off.
B. Press the Power button to power on the system.

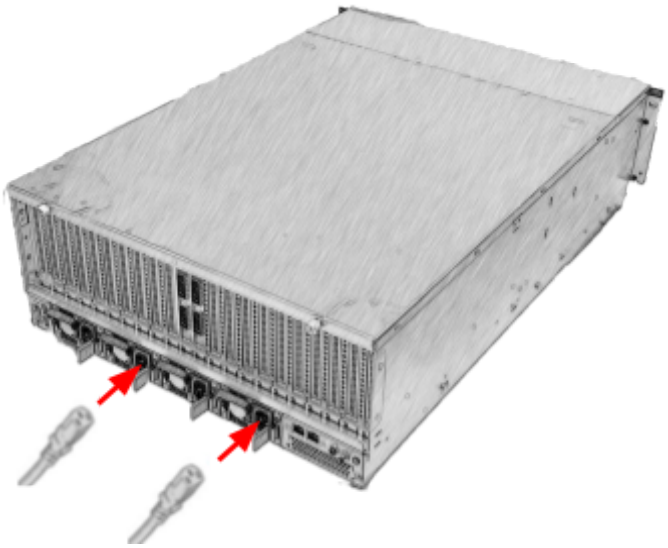


Figure 24. Plugging in Power Cords

Figure 25. UID Light Blinks Twice Before System Power-On

Attention: If the system does not power on (e.g., the screen remains unresponsive or black):

1. Perform an AC power cycle by unplugging and re-plugging the power cord.
2. Repeat Step 2 to confirm the system powers on.
3. If the issue persists, report the problem to the technical support team.

3.3 Complete Initial Setup for Falcon 5012 Management GUI

Access the Falcon 5012 Management GUI via the LAN port IP address to configure essential initial settings. These settings include creating an administrator account, configuring network settings, setting the system date and time, and ensuring the GUI is up-to-date.

About this task:

This task is crucial to prepare the system for operation and management. Completing the initial settings ensures seamless integration and functionality for subsequent resource management.

Procedures:

Step1. Access the Management GUI:

- Open a web browser on a device connected to the same network as the chassis.
- Enter the default IP address in the browser's address bar to access the system management interface.

GUI/API Default Address: http://169.254.100.200

Note:

- If two systems need to access the GUI for initial setup at the same time, make sure to complete the setup on one system before switching to the next. This prevents IP conflicts.
- Make sure your computer is on the same network as the device. Connect the network cable to the laptop, configure the network settings, verify GUI accessibility, and proceed with the initial setup via the GUI.

Step 2. Complete the Initial Setting Steps:

A. Set Administrator Account (refer to Figure 26):

1. Create and confirm a strong password for the administrator account.
2. Click Next to proceed to the next step of the setup.

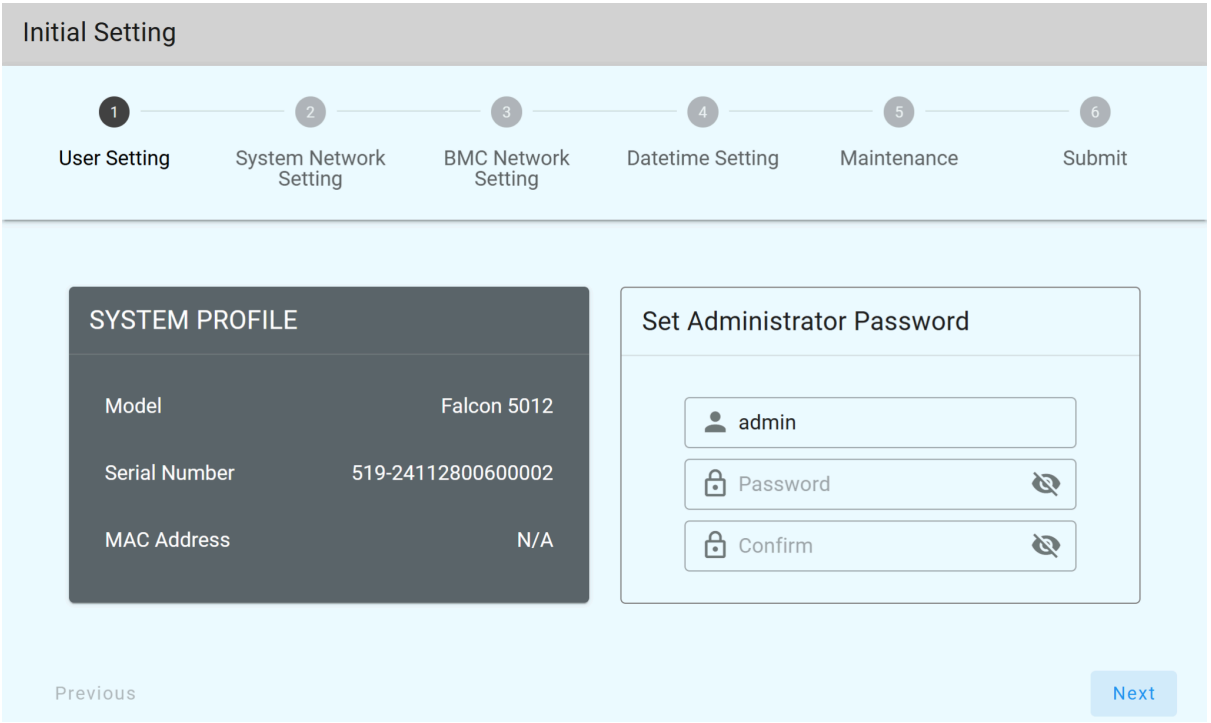


Figure 26. Setting the Administrator Password During Initial Setup

B. Configure System Network Settings (refer to Figure 27)

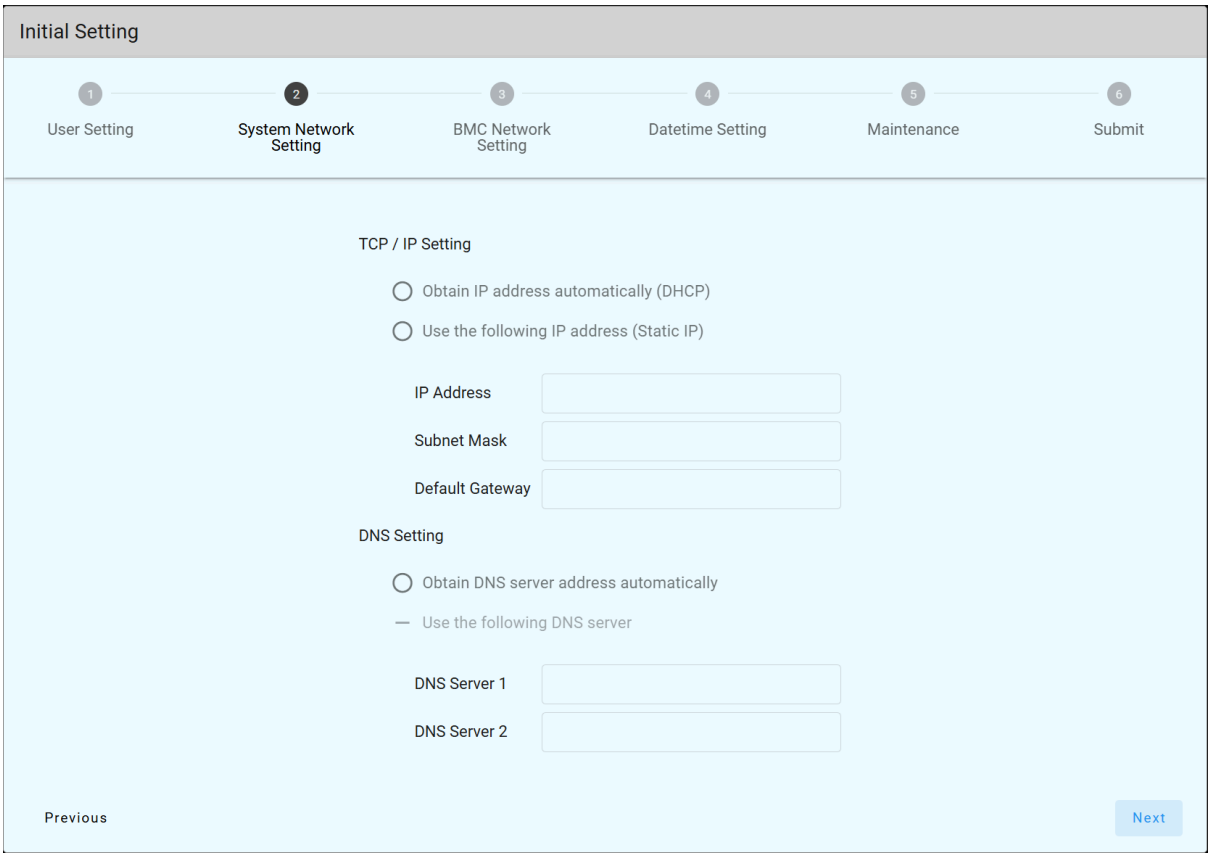


Figure 27. Configuring TCP/IP and DNS Settings in System Network Setting

C. Select IP Configuration Method:

Option 1: Select Obtain IP address automatically (DHCP) to use a dynamically assigned IP address from the network.

> **Important:** DHCP-assigned IP addresses are dynamic and may change. If the IP address expires or changes, you will need to return to the Linux OS to obtain the new DHCP IP address.

Option 2: Select Use the following IP address (Static IP) to manually configure a fixed IP address to avoid address changes.

Input Network Details (if using a Static IP):

a. IP Address: Enter the desired static IP address.
b. Subnet Mask: Input the appropriate subnet mask for your network.

c. Default Gateway: Provide the default gateway address for network communication

D. Select DNS Configuration Method:

Option 1: Select Obtain DNS server address automatically to use DNS settings provided by the network.

Option 2: Select Use the following DNS server and manually input:

a. DNS Server 1: Primary DNS server address.
b. DNS Server 2: Secondary DNS server address (optional).

Save and Confirm:

a. Review all entered information to ensure accuracy.
b. Click Next to save the settings and proceed to the next configuration step.

Note:

- DHCP IP addresses are suitable for temporary setups but may lead to connectivity issues if they change.
- Using a static IP is recommended for stable and long-term configurations to maintain consistent access to the GUI.
- Always ensure the selected IP address (static or dynamic) does not conflict with other devices on the network.

Step 3. Configure BMC Network Settings

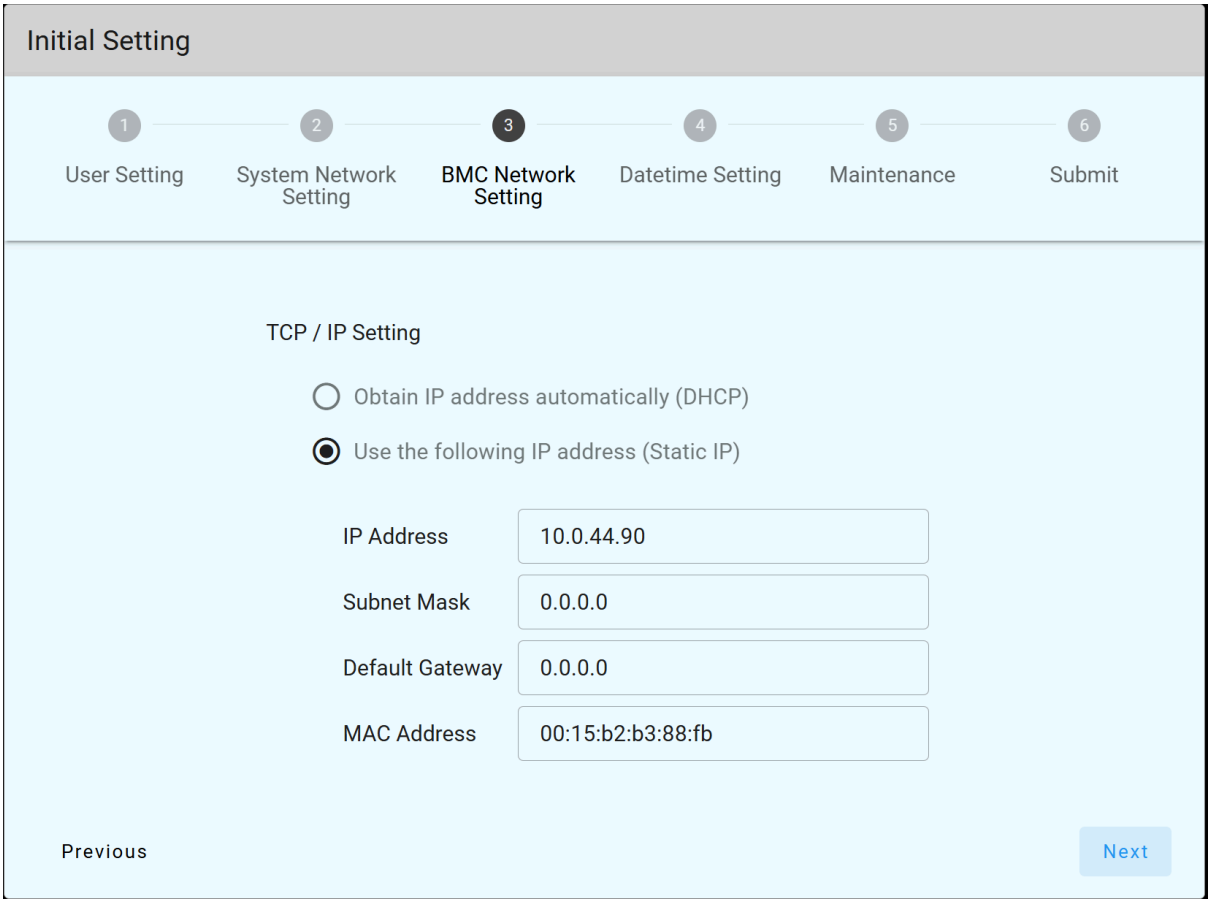


Figure 28. Configuring BMC Network Settings with TCP/IP Options

A. Select BMC IP Configuration Method:

Option 1: Dynamic IP Address (DHCP):

- Select Obtain IP address automatically (DHCP) to allow the network to assign an IP address dynamically.

> **Note:** DHCP-assigned IP addresses are dynamic and may change over time. If the BMC IP changes, you will need to return to the Linux OS to retrieve the new address.

Option 2: Static IP Address: Select Use the following IP address (Static IP) to manually assign a fixed IP address.

B. Input Static IP Details (if selected): Enter the following information:

- **IP Address:** The desired static IP address for the BMC.
- **Subnet Mask:** The appropriate subnet mask for your network.

- **Default Gateway:** The default gateway address for the BMC network.

C. Review and Save Settings:
   1. Verify the entered information for accuracy and alignment with your network configuration.
   2. Click Next to save the settings and proceed to the next configuration step.

> **Important Notes:**
> - For DHCP Users:
>   DHCP is suitable for temporary setups or dynamic environments. Be aware that IP addresses may change, requiring you to retrieve the new BMC IP address from the Linux OS if access is lost.
> - For Static IP Users:
>   A static IP ensures stable and consistent access to the BMC. Ensure the assigned IP does not conflict with other devices on the network.
> - General Precautions:
>   - Verify that the IP address, subnet mask, and gateway belong to the same subnet.
>   - Avoid leaving the gateway field blank, as it may cause connectivity issues.

Step 4. Set Datetime:

- Adjust the system date and time to ensure accuracy.

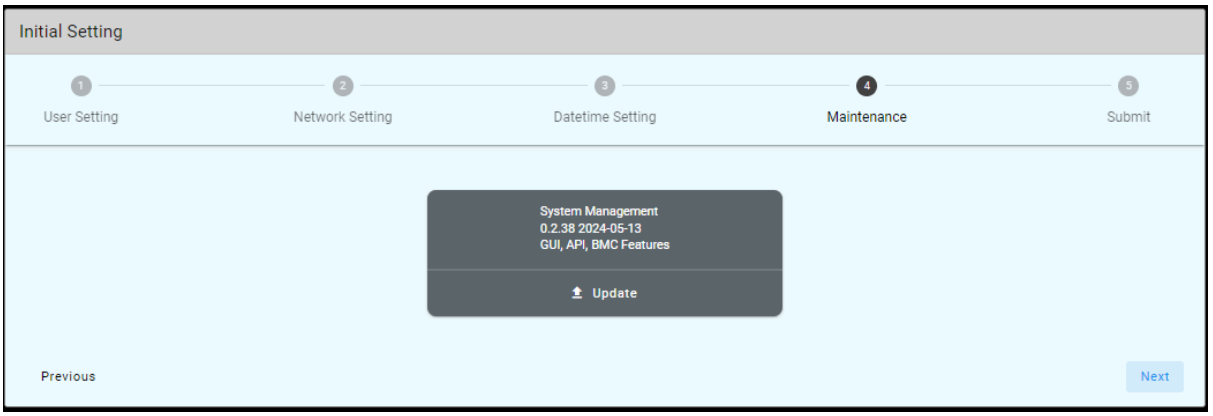
Step 5. Perform Maintenance Updates

A. Download the Latest Firmware:
   1. Access the H3 Platform support page:
      - Navigate to Support → Knowledge Base → Download on the H3 Platform website.
      - Alternatively, visit the following URL directly: https://www.h3platform.com/knowledge-base/document
   2. Select the appropriate options:
      - Product Type: Composable GPU Chassis.

- Model Type: Falcon 5012.

3. Download the latest System Management Firmware file.

B. Upload the Firmware to the System:



1. On the Maintenance screen: Click the Update button under the System Management section.
2. Select the downloaded firmware file: Confirm the file is correct and click Update.

Important: Uploading the firmware cannot be cancelled once you click the Update button.

3. Verify the upload process:
   a. Wait for the system to process the firmware update.
   b. A success message will be displayed:
      "System firmware update successful. Please proceed to the next step."

C. Complete the Maintenance Step:

1. After the firmware upload is complete: Click Submit to finalize the Maintenance step.
2. Proceed to the next steps in the Initial Setting process.

D. Restart the System (Post-Setup):

1. Complete all steps in the Initial Setting process.
2. After entering the GUI system, the system will prompt for a restart.
3. Restart the system to apply the firmware update.

Notes:

- File Compatibility: Ensure the downloaded firmware matches the Falcon 5012 model to prevent errors.

- Stable Upload Environment: Do not power off or interrupt the system during the firmware upload process.
- Firmware Application: The updated firmware will only take effect after completing the Initial Setting process and restarting the system.

Step 6. Submit the Configuration:

After completing all settings, click Submit to save the changes and proceed to the GUI dashboard for system management.

3.4 Device Installation in Falcon 5012 System

This task outlines installing GPUs and other PCIe devices into the Falcon 5012 system. It includes accessing the chassis, installing and securing GPUs, and connecting power cables. Follow the steps precisely to ensure system stability and performance.

About this task:
Proper GPU installation is critical to maintain system integrity and ensure optimal performance. This task includes safety precautions, required tools, and step-by-step instructions for installing dual-width and single-width PCIe devices.

Procedures:

Step 1. Preparation for GPU Installation:

A. Safety Precautions:
   1. Power off the system and disconnect it from all power sources.
   2. Use an ESD wrist strap to prevent electrostatic discharge damage.
   3. Verify the GPU's compatibility with the system's PCIe slot and power cables.

B. Required Tools:
   1. Screwdriver for stabilization plate and PCIe slot screws.
   2. GPU power cables (refer to Figure 34 for cable descriptions).

Important:
Always use an ESD wrist strap to protect the system from static damage. Ensure that the system is completely powered off before proceeding.

Step 2: Accessing the Chassis:

A. Opening the Chassis:
   1. Loosen the hand screws securing the top cover.

2. Slide the cover backward to expose internal components (refer to Figure 29).

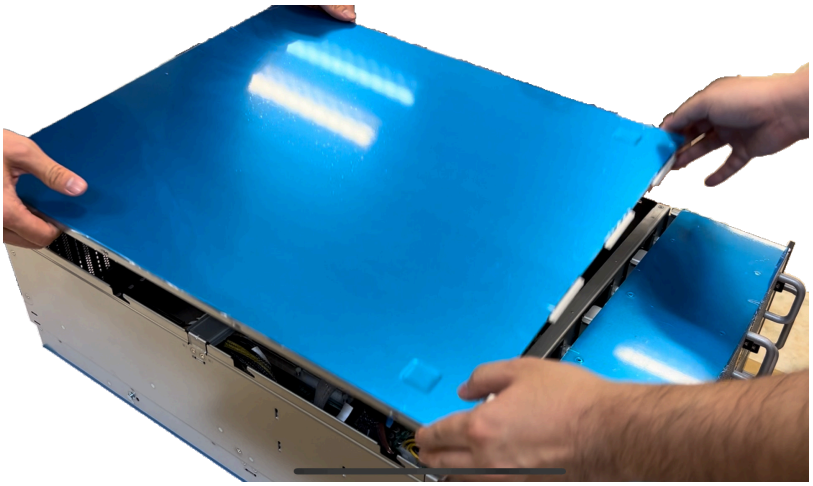


Figure 29. Removing the Top Cover to Access the Internal Components

B. Removing the Stabilization Plate:
    1. Unscrew the screws on both sides of the stabilization plate.
    2. Carefully lift and remove the plate (refer to Figure 30).

Figure 30. Removing the Stabilization Plate for Internal Component Access

---

Step 3: Installing Dual-Width GPUs:

A. Preparing the PCIe Slot: loosen and remove the two brackets adjacent to the target PCIe slot.
B. Inserting the GPU:

Important:
Be cautious of onboard components during the installation process to avoid accidental collisions.

1. Align the GPU connector with the PCIe slot.
2. Press the GPU vertically into the slot until it is securely seated (refer to Figure 31).

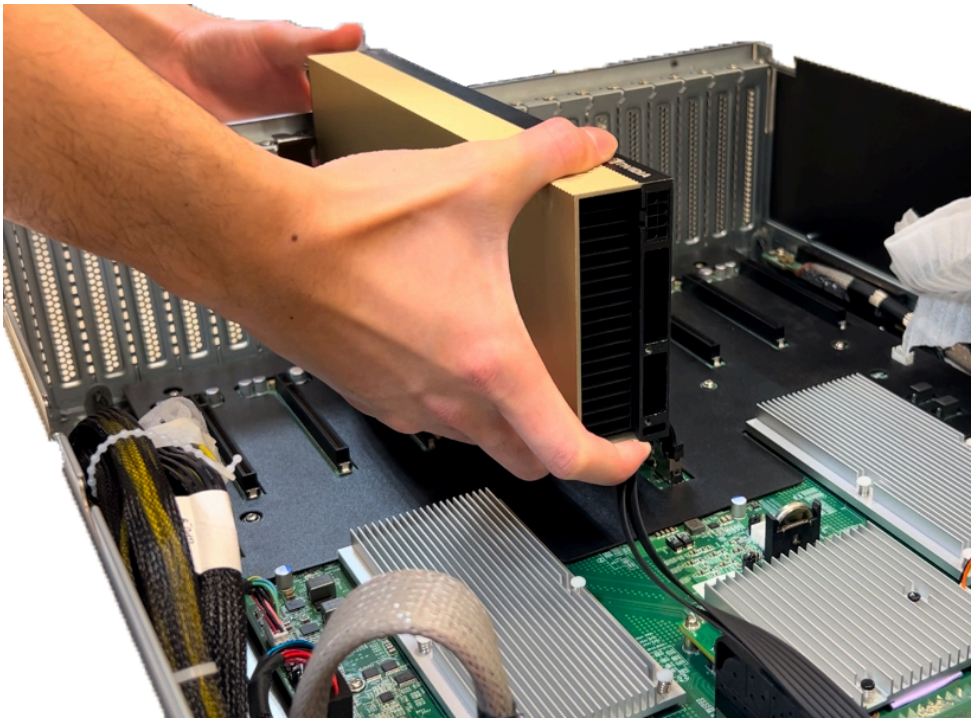


Figure 31. Inserting a Dual-Width GPU into the PCIe Slot

C. Securing the GPU: Use screws to secure the GPU in place.

Note:
Ensure the rear gold fingers are properly seated in the PCIe slot. Misalignment may cause system errors or physical damage.

Step 4: Connecting GPU Power Cables:

A. Power Cable Mapping: Use the appropriate GPU power cables and connect them according to the following mapping (refer to Figure 33):

- Left-side short cables (GPU_1, GPU_2, GPU_3): PCIe slots 11, 12, 13.
- Left-side long cables (GPU_4, GPU_5): PCIe slots 14, 15.
- Right-side long cables (GPU_6, GPU_7): PCIe slots 21, 22.
- Right-side short cables (GPU_8, GPU_9, GPU_10): PCIe slots 23, 24, 25.

Note: Connect the GPU power cables as indicated by the labels on the cables. Refer to Figure 32 for an example of the label format, which displays the corresponding GPU number (e.g., GPU_4, GPU_5) for accurate identification.

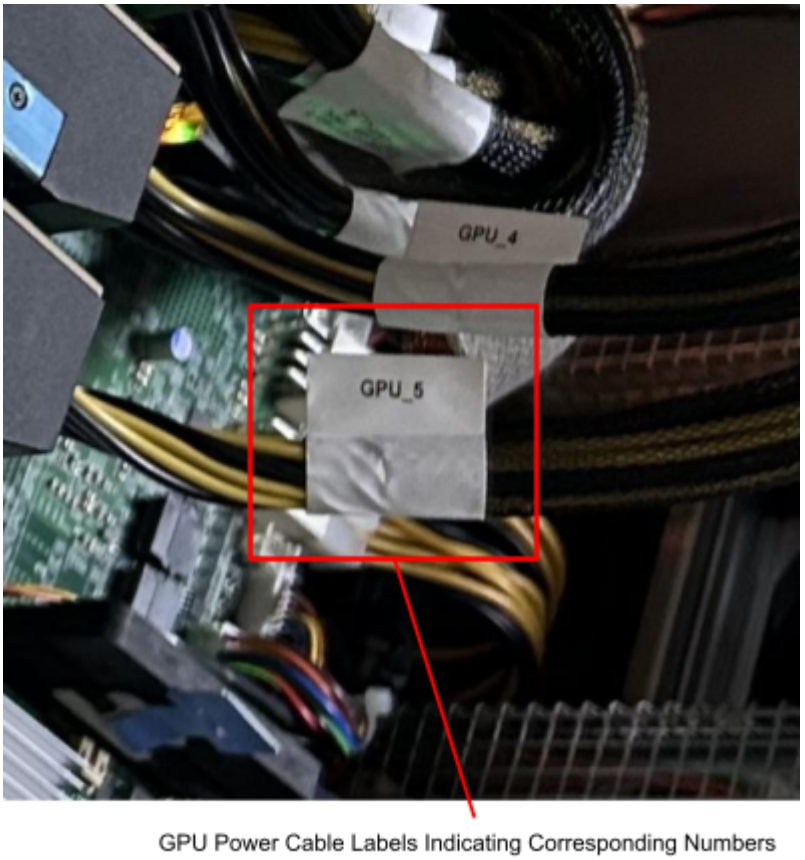


Figure 32. GPU power cables are labeled with corresponding numbers (e.g., GPU_4, GPU_5) to ensure accurate connections.

Important: Connect the GPU power cables to the corresponding PCIe slots as indicated. The figure below illustrates the slot numbers for each connection.

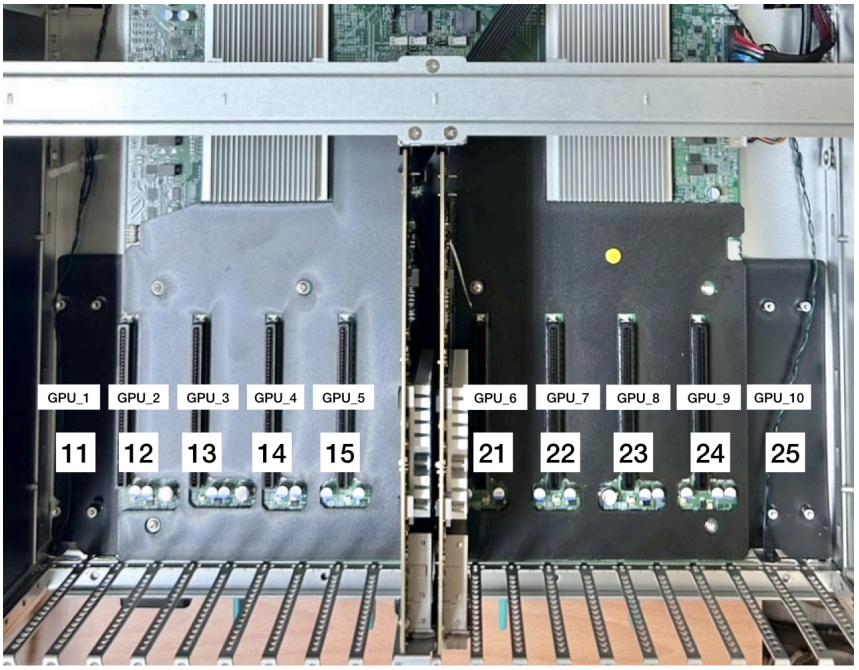


Figure 33. PCIe Slots with Corresponding Numbers

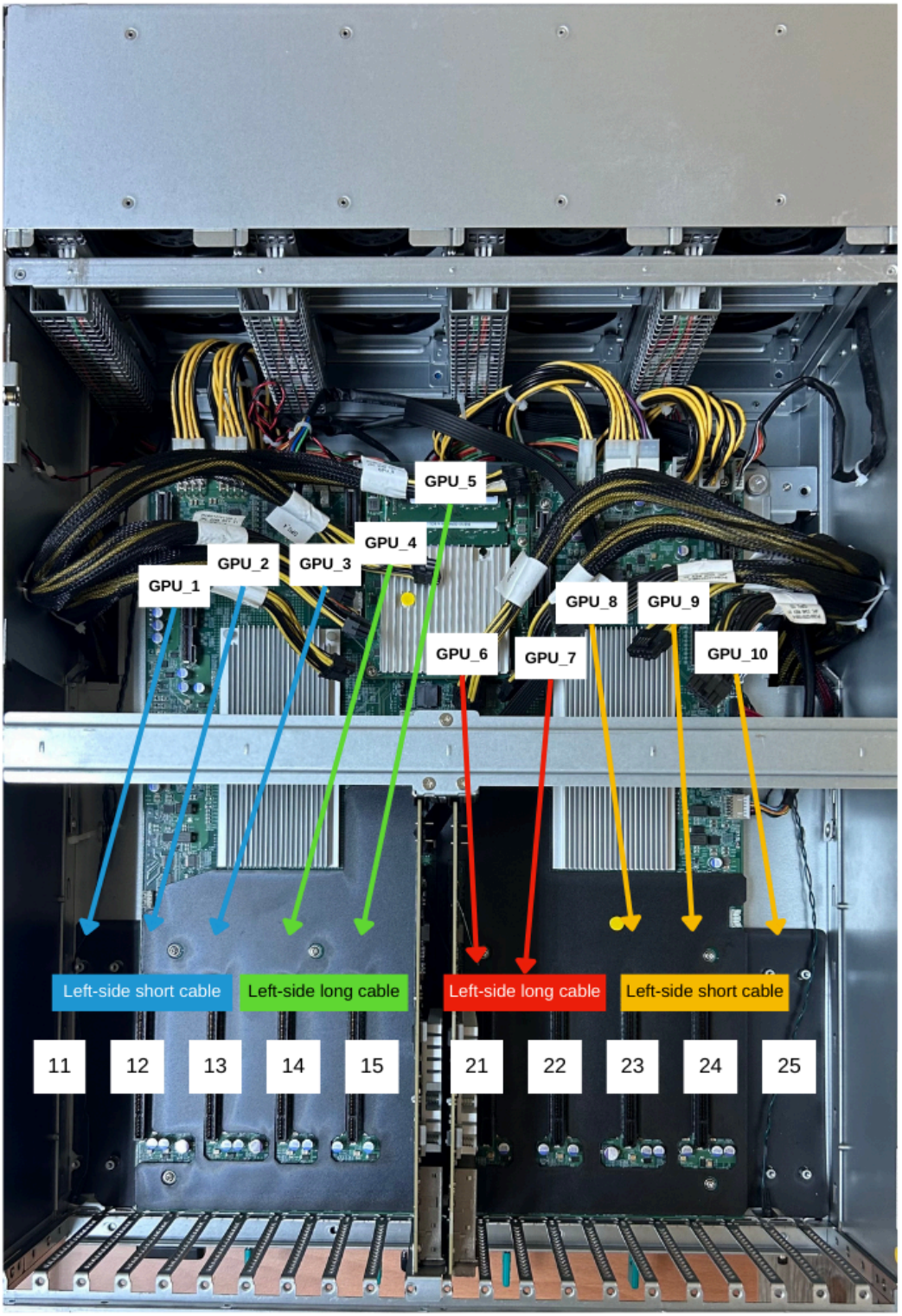


Figure 34. GPU Power Cable Connections and Label Mapping

B. Securing Connections: Ensure all power cables are securely connected.

Caution: Use only model-specific power cables to avoid damage. Ensure that each power cable is connected to its corresponding GPU for proper one-to-one mapping.

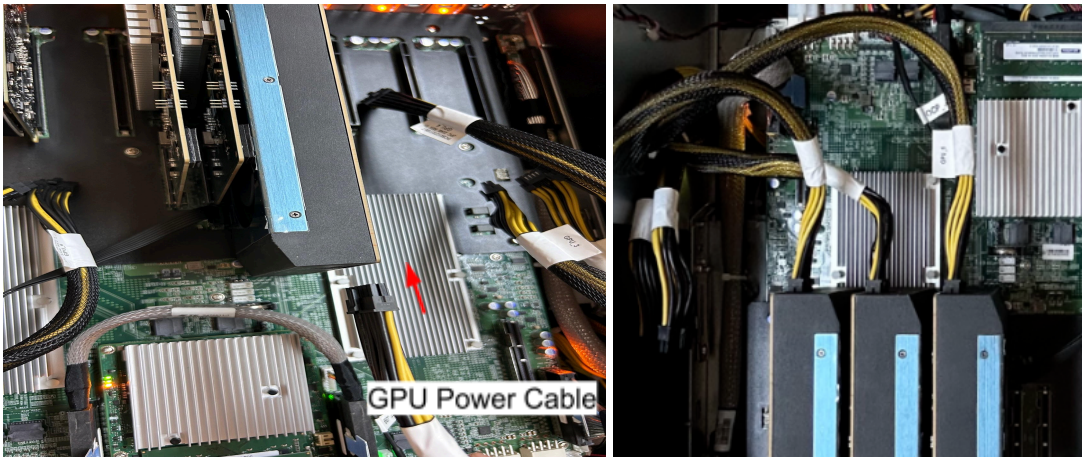


Figure 35. Securing GPU Power Cable Connections

Step 5. Installing Single-Width PCIe Devices:

Installation Steps:

Important: Be cautious of onboard components during the installation process to avoid accidental collisions.

A. Remove the bracket nearest to the target PCIe slot
B. Insert the single-width device (e.g., NIC card or single-width GPU) into the PCIe slot (refer to Figure 36)

Figure 36. Installing a Single-Width Device into the PCIe Slot

C. Secure the device with screws.

Important: Ensure PCIe slots can accommodate the device's width and follow installation guidelines to avoid connection errors.

3.5 Retimer Installation and Host Configuration

Retimer cards are required to maintain PCIe Gen5 signal integrity between the host server and the expansion chassis. This section focuses on installing and configuring retimer cards in host-side (server-side) deployments, where the card is installed directly into a PCIe Gen5 x16 slot on the server.

Two types of retimer cards are supported: C01 and D01. Each requires specific DIP switch settings to ensure proper link control and system compatibility when installed on the host.

Note: For advanced configurations where retimer cards are installed inline within the Falcon 5012 chassis-side expansion cabling, refer to Chapter 7 – Supplementary Notes for additional instructions and supported topologies.

About the Task

To configure a retimer card for host-side installation:

1. Identify the card variant:
    - C01: Single SW1 switch
    - D01: 3-bit DIP switch (CH1–CH3)
2. Set DIP switch positions:

   C01
    - SW1 = ON
    - SW2 = HLL
3. D01
    - SW1:
        - PERST = ON
        - I2C_SDA = OFF
        - I2C_SCL = OFF
    - SW2 = HLH
4. Install the card into the designated PCIe Gen5 x16 slot on the host server.
5. Connect the CDFP cable between the retimer and the expansion chassis.

Note: For chassis-side (jbox-side) installation and advanced multi-host configurations, refer to CH7 Supplementary Notes for complete guidance.

Prerequisites

Before proceeding with the installation, ensure the following:

- The host system has an available PCIe x16 slot with installation space of 150 mm length and 68.85 mm height (see Table 26).

Table 26. Retimer Installation Requirements

Requirement	Description
Installation Space	PCIe x16 socket, with 150 mm length and 68.85 mm height.

- The Retimer is ready for installation.
- Proper electrostatic discharge (ESD) precautions are taken to prevent damage to components.

Refer to Section 2.4.2 for the hardware information of the Retimer.

Procedures

Step 1. Configure Retimer Switches

Before installing the retimer card, configure the DIP switches properly. C01 and D01 variants have different switch layouts and behavior.

For C01 Host Adapter:

- Set SW1 = ON for server-side installation.
  Refer to Figure 37 for switch orientation.
- Set SW2 = HLL, which corresponds to I2C address 0xE8
  This is the default setting and typically does not require change.

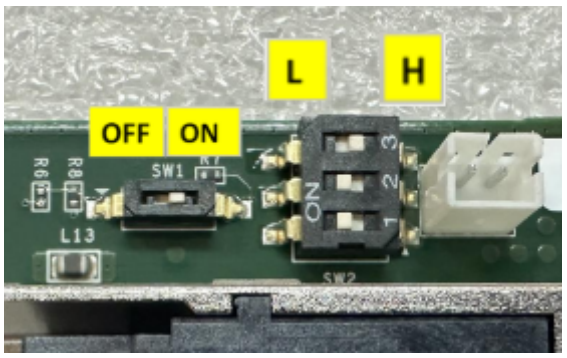


Figure 37. SW1 and SW2 DIP Switch Settings for C01 Host-Side Installation

---

For D01 Host Adapter:

- Configure SW1 (3-bit DIP switch) as follows for host-side use:
  - PERST switch = ON → Allow PERST from host to JBOX
  - I2C_SDA switch = OFF → Block I2C_SDA from host
  - I2C_SCL switch= OFF → Block I2C_SCL from host
  - Refer to Figure 39 for DIP switch layout.
- Set SW2 = HLH, which corresponds to I2C address 0xEA

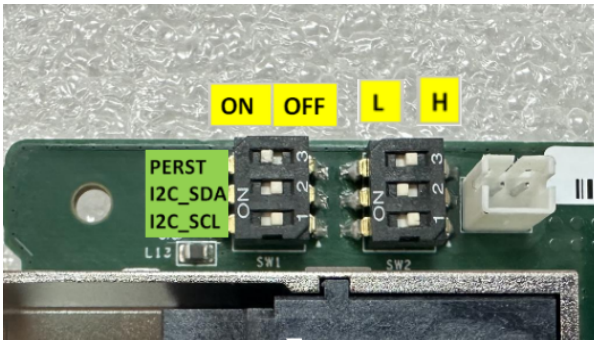


Figure 38. SW1 and SW2 DIP Switch Settings for D01 Host-Side Installation

---

Step 2. Install the Retimer Card

A. Insert the retimer card into the designated PCIe Gen5 x16 slot on the host server.
B. Verify that the card is fully seated and secured using the chassis retention mechanism.

Step 3. Connect the CDFP Cable

A. Use certified CDFP cables to connect the retimer card to the expansion chassis.
B. Ensure both ends are properly seated and locked to avoid signal loss.

Caution: Ensure the CDFP cable is securely connected. Failure to properly lock the connector can result in system malfunctions and potential damage to the Falcon 5012.

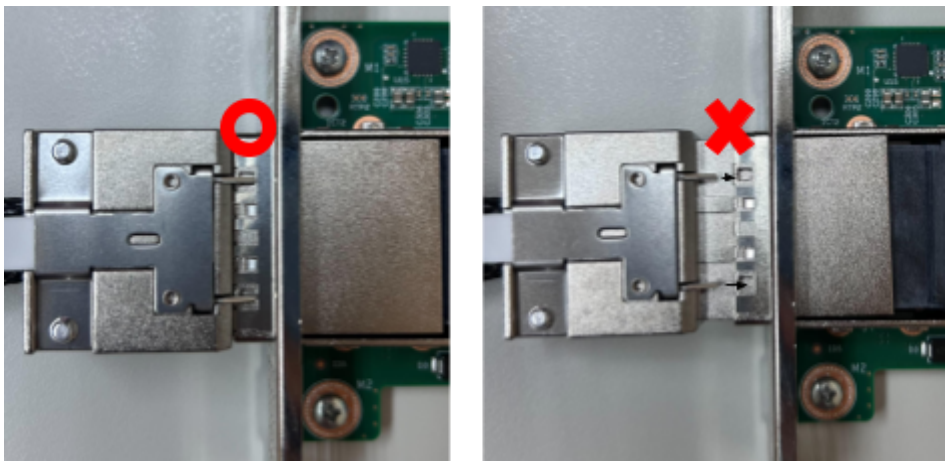


Figure 39. Picture of properly connecting the CDFP cable with the CDFP connector

Note: For JBox-side deployments, refer to Chapter 7 – Supplementary Notes for the appropriate DIP switch configuration.

3.6 GPU Assign and Unassign

This task explains how to assign or unassign GPUs to hosts using the Resource Management interface in the GUI.

Attention: The mode switch between Standard Mode and Advanced Mode must be synchronized across all PCIe switches.

- It is not possible to have mixed modes (e.g., some switches in Standard Mode and others in Advanced Mode).
- Ensure that all switches are set to the same mode to avoid operational conflicts.

About this task:

In Standard Mode, each PCIe switch is fixedly assigned to a single host, and the GPUs connected to that switch are automatically allocated to the host.

In Advanced Mode, one PCIe switch can connect to multiple hosts, enabling users to dynamically assign or unassign GPUs based on specific host requirements. This mode offers greater flexibility for resource allocation.

Procedures

Step 1. Assign GPUs in Advanced Mode:

A. Navigate to the Resource Management interface in the GUI.

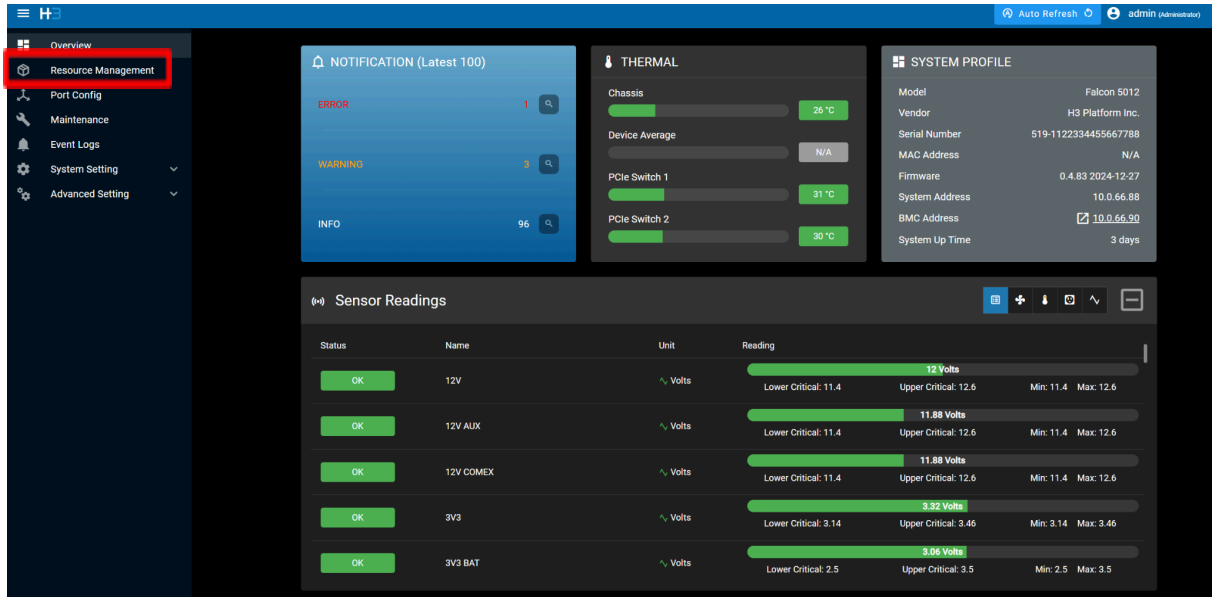


Figure 40. Accessing the Resource Management Interface on GUI

B. Select the host to which you want to assign GPUs.

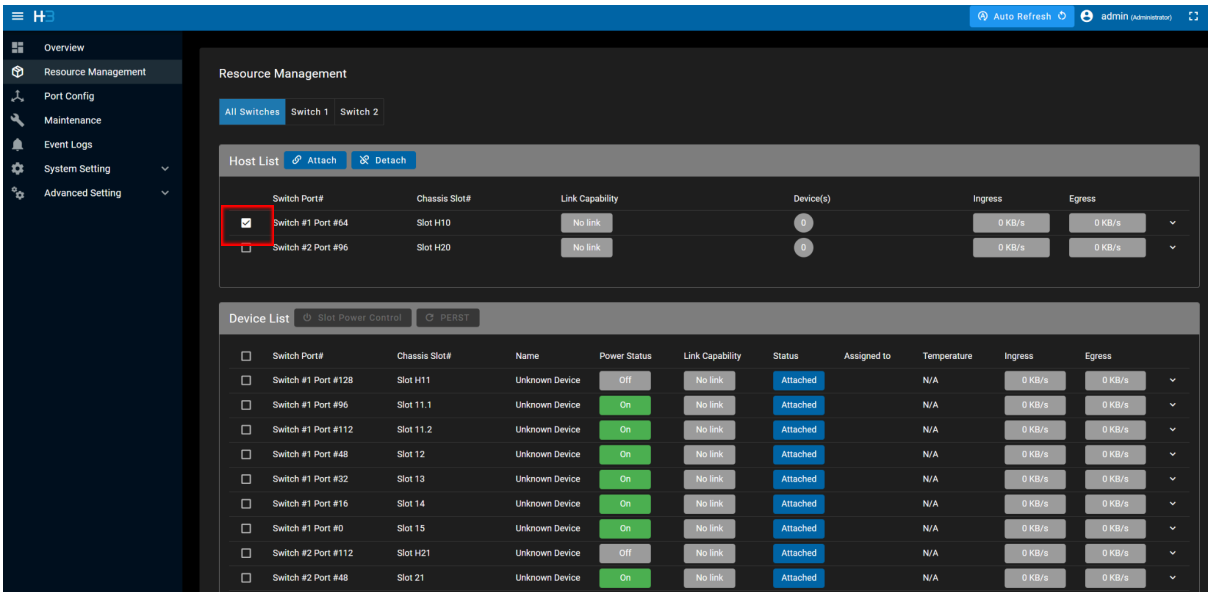


Figure 41. Selecting a Host for GPU Assignment in Resource Management

C. Click the Attach tab to access the GPU allocation page.

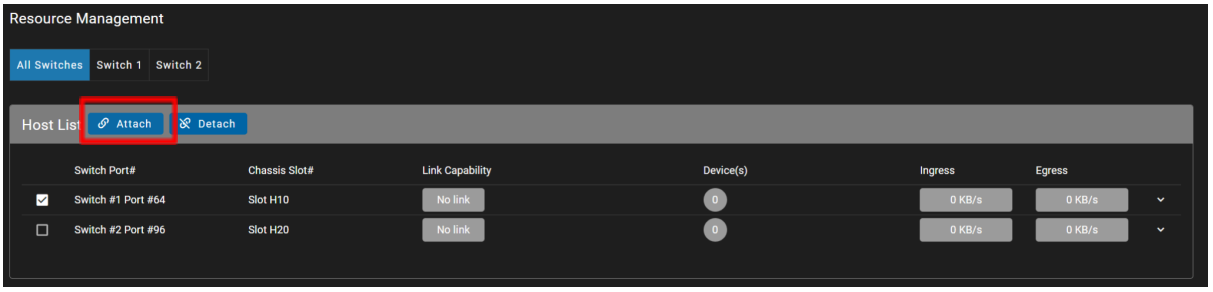


Figure 42. Accessing the GPU Allocation Page via the Attach Tab

D. Perform the following steps:

1. From the displayed GPU list, select the GPUs you want to assign.

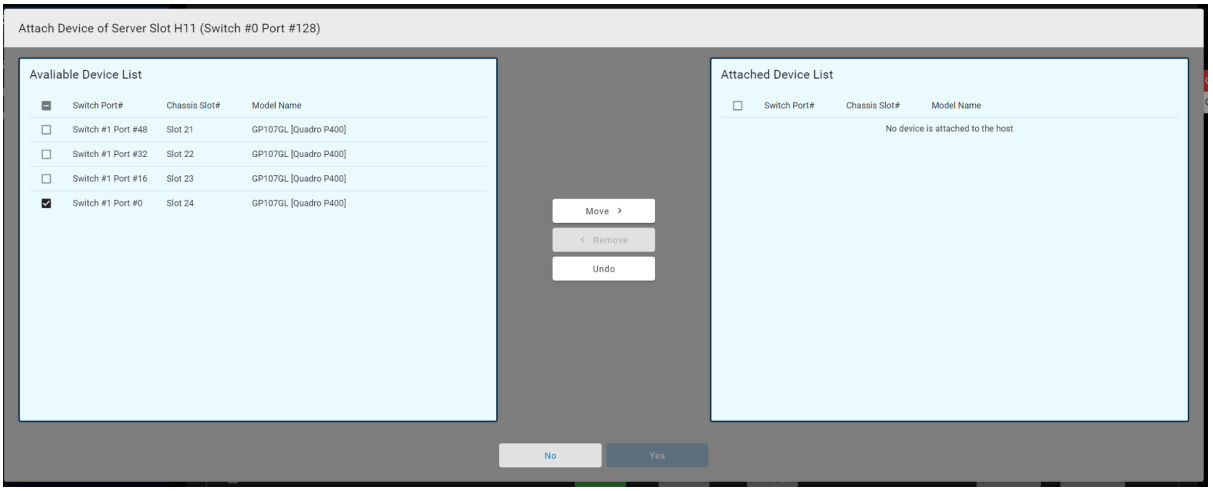


Figure 43. Selecting GPUs from the Available Device List

2. Click Move to transfer the selected GPUs to the Attached GPUs column.

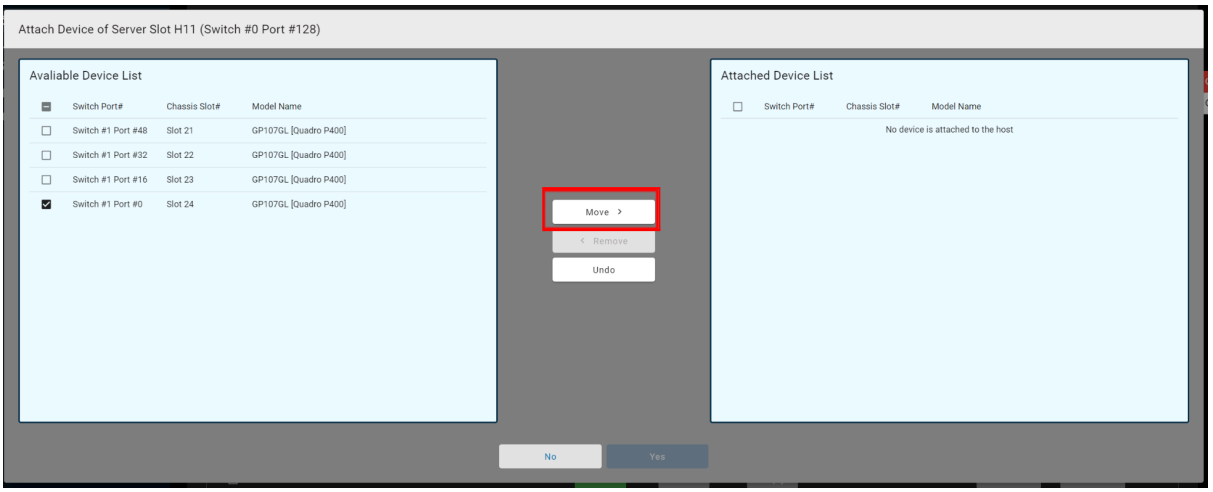


Figure 44. Moving GPUs from the Available Device List to the Attached Device List

E. Confirm the selection by clicking Yes to complete the GPU assignment.

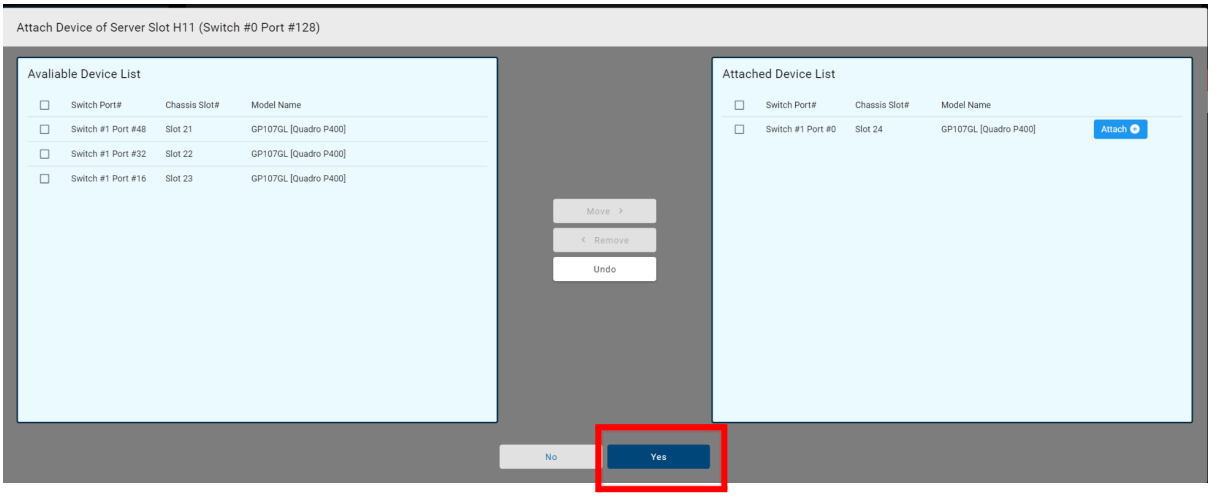


Figure 45. Confirming GPU Assignment by Clicking "Yes"

Step 2. Unassign GPUs in Advanced Mode (if needed):

A. Navigate to the Resource Management interface in the GUI.

B. Select the host from which you want to unassign GPUs.

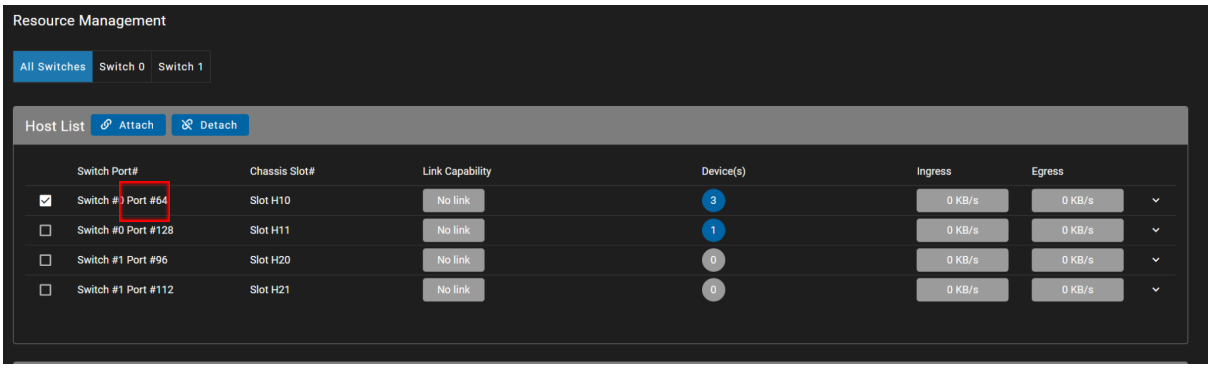


Figure 46. Selecting Host for GPU Unassignment on GUI

C. Click the Detach tab to access the unassignment page.

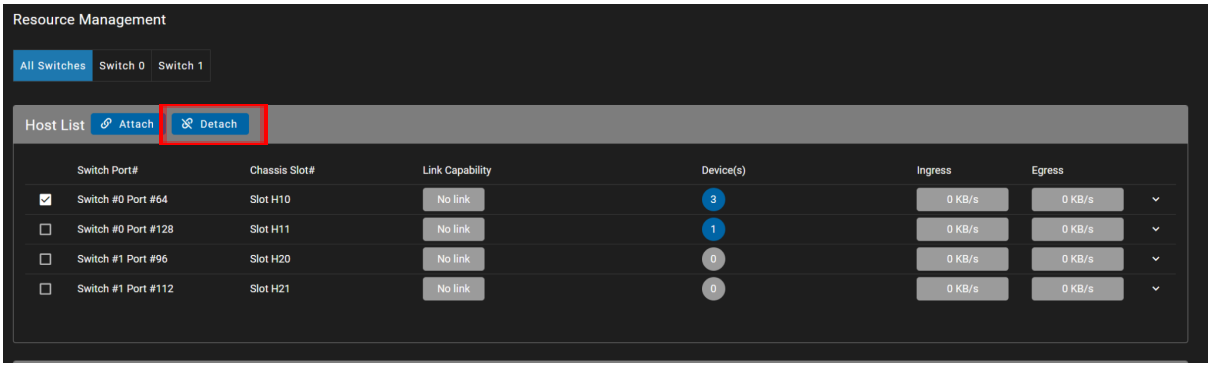


Figure 47. Accessing the Unassignment Page via the Detach Tab

D. Perform the following steps:

1. From the Attached GPUs column, check the GPUs you want to unassign.

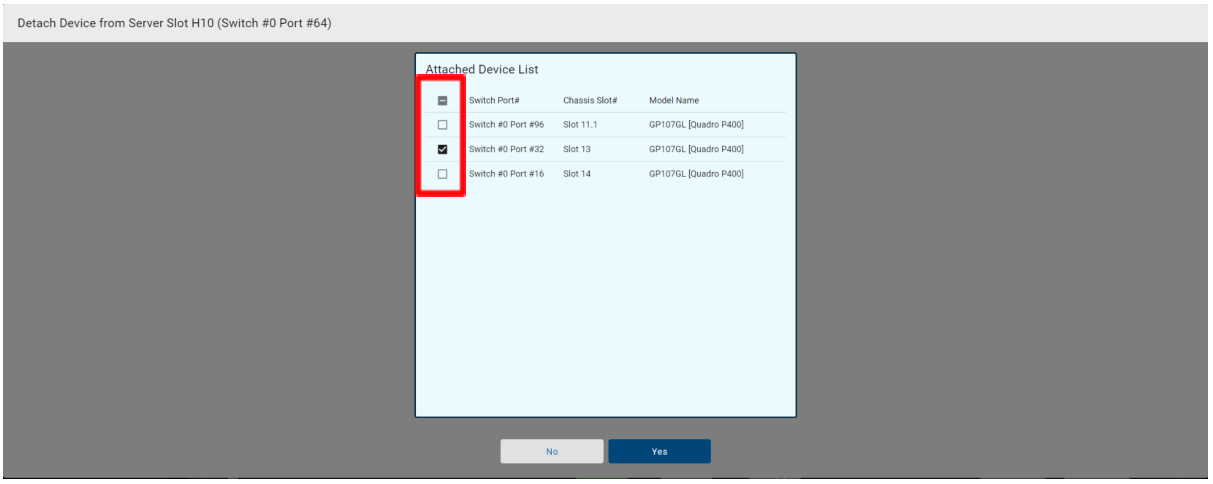


Figure 48. Checking the GPUs for Detachment

2. Click Yes to detach the selected GPUs back to the Available GPUs column.

Step 3. Verify Connections:

A. Ensure that the CDFP cable is securely connected between the chassis and host (refer to Figure 49). Connect the corresponding ports with CDFP cables as shown in Figure 50.

B. Confirm that GPUs are properly assigned to their respective hosts in the GUI.

C. After verifying all connections and assignments, proceed to power on the server.

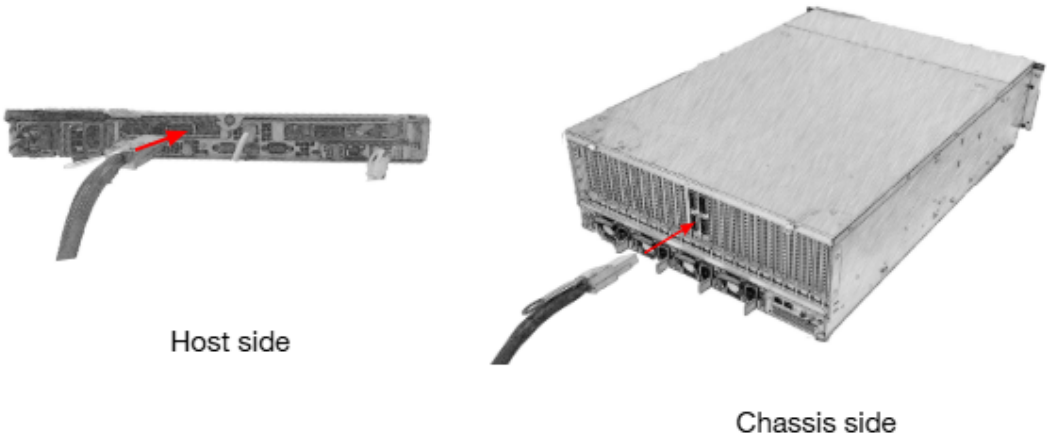


Figure 49. Host and Chassis Connections Using CDFP Cables

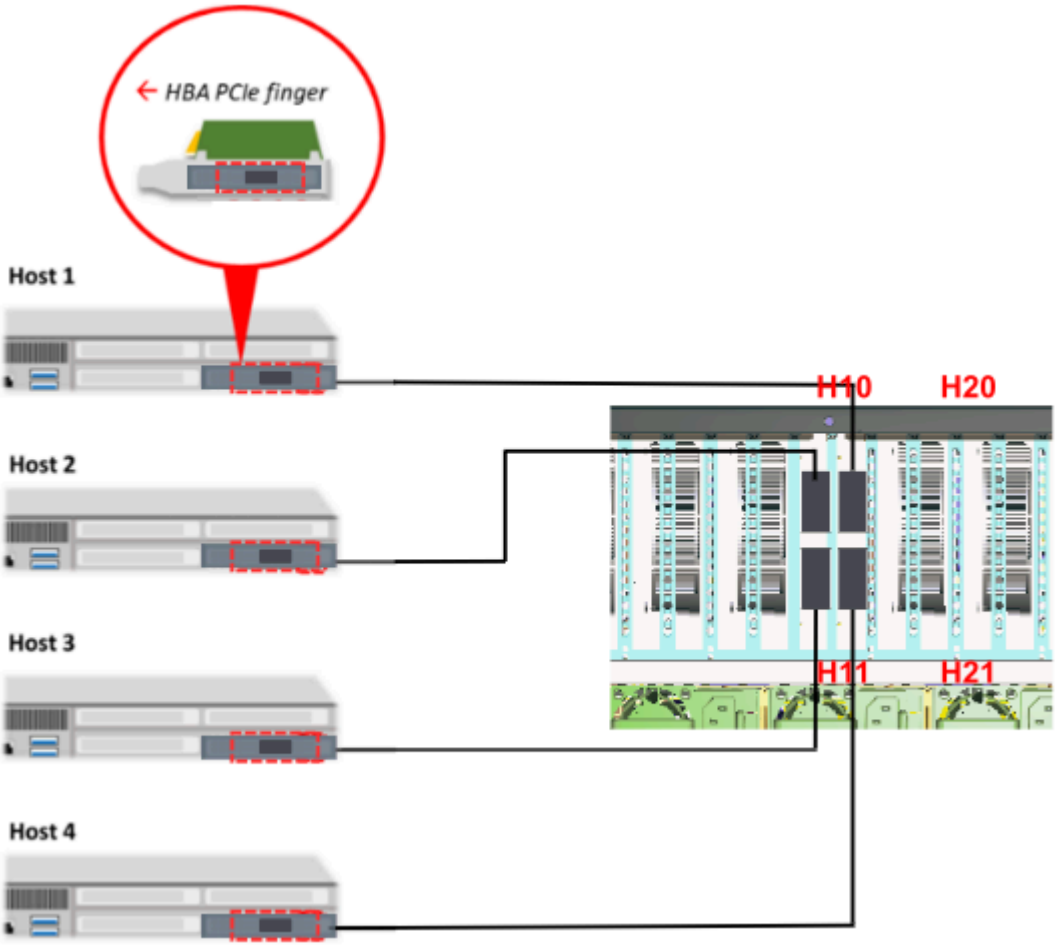


Figure 50. CDFP Cable Mapping Between Hosts and Chassis Ports

3.7 Verifying GPU Resource Assignment or Unassignment on the Host

This procedure outlines how to verify GPU assignment or unassignment on the host system, ensuring GPUs are correctly mapped, detected, and operational through server commands and GUI checks.

About this task

After assigning or unassigning GPU resources, the following verifications are necessary:

1. Ensure GPUs are properly mapped in the PCIe tree structure on the server.
2. Confirm the total number of GPUs matches the expected configuration.
3. Verify that the host GUI reflects the correct PCIe link capability and recognizes all assigned GPUs.

Procedures

Step 1. Verify PCIe Tree Structure on the Server

- On the server, open a terminal and run the following command lspci -vt:

```
None

lspci -vt
```

- Review the output to confirm GPU assignments.

Step 2. Verify Total GPUs on the Server

- Run the following command nvidia-smi to check the total number of NVIDIA GPUs on the server:

```
None

nvidia-smi
```

- Verify that the total GPU count matches the expected number.

Step3. Verify GPU Status on the Host via GUI

- Access the system's Web GUI and navigate to the Resource Management panel.
- Confirm the following:
  - Link Capability: Ensure PCIe links are displayed with the correct specifications (e.g., Gen4x16 or Gen5x16).
  - Device(s): Verify the assigned GPUs are listed and operational. For example, if the Device(s) field displays "4," it indicates that four GPUs are correctly assigned and connected.

Results

When the procedure is successfully completed:

- The server recognizes the assigned GPUs in the PCIe tree and displays the correct count.
- The host GUI confirms accurate PCIe link capability and GPU assignments.

If any discrepancies are identified, contact technical support for further diagnostics and solutions.

Chapter 4

Web GUI Introduction

After activation, the PCIe Gen5 GPU Expansion system is ready to use. Detailed guidance on operating the Falcon 5012 Web GUI and its functionalities is provided in subsequent chapters.

4.1 Log In

To access the GUI, users must log in using a valid Username and Password, ensuring only authorized personnel can access the system.

After the first login, users must update the default credentials or create a new account for security.

Note: The default login credentials are:

- Username: admin
- Password: admin

Refer to Figure 51 for the login page interface.

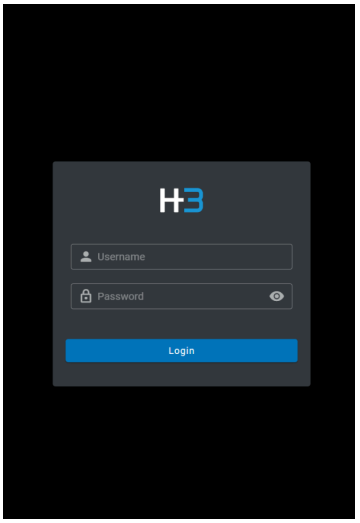


Figure 51. Login Page

4.2 Functions

4.2.1 Overview

The Overview page provides a centralized dashboard for comprehensive system analytics, enabling efficient monitoring and resource utilization. Key metrics in Table 27 such as notifications, thermal status, system profile, sensor readings, and PCIe link health are presented in a concise, user-friendly interface for quick insights into system health and performance.

Key Features:

- **Automatic Refresh:** The page automatically updates every 30 seconds to provide the latest system data.
- **Manual Refresh:** A Refresh button in the bottom-right corner allows users to update system information as needed manually.

Features of the Overview Page

Table 27. Overview Page Features

Feature	Description
1. Notifications	Displays system alerts and messages, categorized by severity: **Error** (Red), **Warning** (Amber), and Info (White). These indicators highlight critical issues requiring user attention.
2. Thermal Status	Shows average temperatures for components such as the chassis, device, and PCIe switches. A color-coded system (Green/Amber/Red) provides an at-a-glance status of temperature safety.
3. System Profile	Offers essential system details including chassis model, serial number, MAC address, firmware version, system IP address, and uptime.
4. Sensor Readings	Monitors key parameters like fan speed (RPM), power consumption (W), and voltage levels, ensuring the system operates within optimal thresholds.
5. PCIe Link Health	Displays the operational status of PCIe links, with visual indicators for healthy connections and error states, enabling proactive troubleshooting.
6. PCIe Switch 1 Throughput (MB/s)	Tracks bandwidth utilization (MB/s) for PCIe Switch 1 and Switch 2, including ingress and egress data for each slot.
7. Fan Speed (RPM)	Provides real-time monitoring of all system fans, including inlet and outlet speeds, to ensure effective cooling performance.
8. Chassis Temperature (°C)	Tracks temperatures across different chassis regions (e.g., front, rear, PCIe switches) for precise heat management.
9. Device Temperature	Displays individual slot temperatures to monitor thermal conditions for connected devices.
10. Power Consumption	Breaks down power usage by individual PCIe slots and PSUs, offering insights into system energy efficiency.

Detailed Feature Descriptions

Notifications

The Notifications section categorizes system events into:

- **Error**: Critical issues requiring immediate resolution to prevent damage or downtime.

- **Warning**: Moderate issues needing attention to avoid escalation.
- Info: Routine updates on system operations.

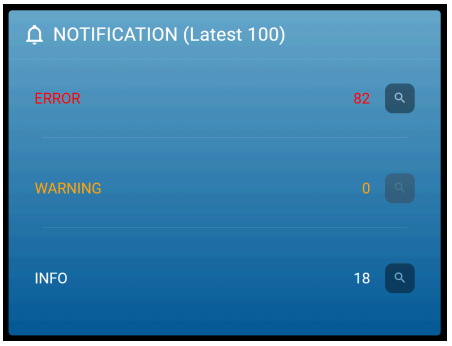


Figure 52. Notifications Overview

Example:
In the Falcon 5012 system, the notifications panel in Figure 52 displays the latest 100 messages, with 0 errors, 28 warnings, and 72 informational updates.

Thermal Status

The Thermal Monitor displays the average temperature of key components within the chassis, including the board, PCIe switch, and connected devices. The system uses a color-coded scheme to indicate the thermal status of each component:

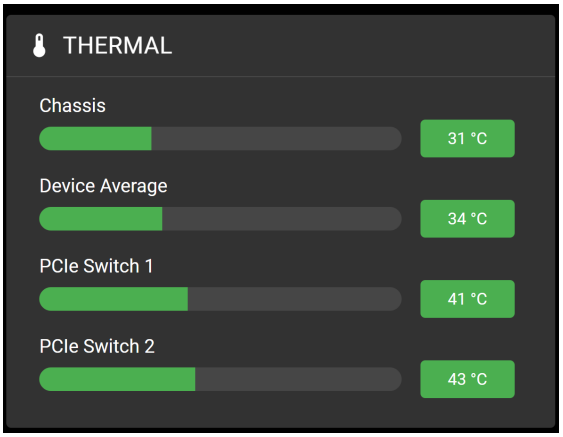


Figure 53. Thermal Monitoring Dashboard

- **Green:** Operating within normal temperature ranges.
- **Amber**: Temperature is nearing critical limits; action may be required.
- **Red**: Temperature exceeds safe operational thresholds and requires immediate intervention.

To protect system integrity, the Falcon 5012 system will automatically shut down if the temperature of any component exceeds 85°C for longer than 10 seconds. This automated protection mechanism ensures hardware reliability during intensive workloads.

System Profile

The System Profile provides a comprehensive summary of essential system information for quick identification and troubleshooting as shown in Figure 54. Table 28 displays the following critical attributes of the system:

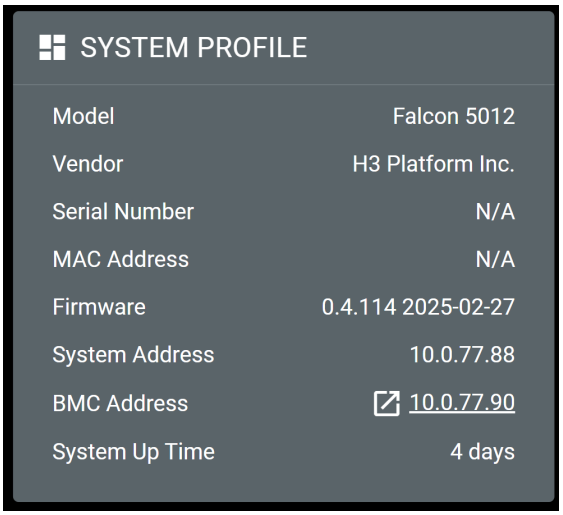


Figure 54. System Profile Overview

Table 28. System Profile Attributes

Model:	Displays the chassis model name.
Vendor:	Indicates the vendor name.
Serial number:	Shows the unique serial number of the machine.
Mac address:	Provides the MAC address of the system's network interface.
Firmware:	Specifies the mCPU firmware version installed.
System Address:	Lists the IP address assigned to the machine.
BMC Address:	Shows the Baseboard Management Controller (BMC) IP address.
System Up Time:	Displays the duration the system has been running without interruption.

Sensor Reading

The Sensor Reading section in the Falcon 5012 GPU Expansion System GUI offers real-time monitoring of key system parameters. This intuitive interface ensures users can easily assess the system's operational health and maintain optimal performance. The available sensor features indicated on Figure 55 are summarized in Table 29 below:

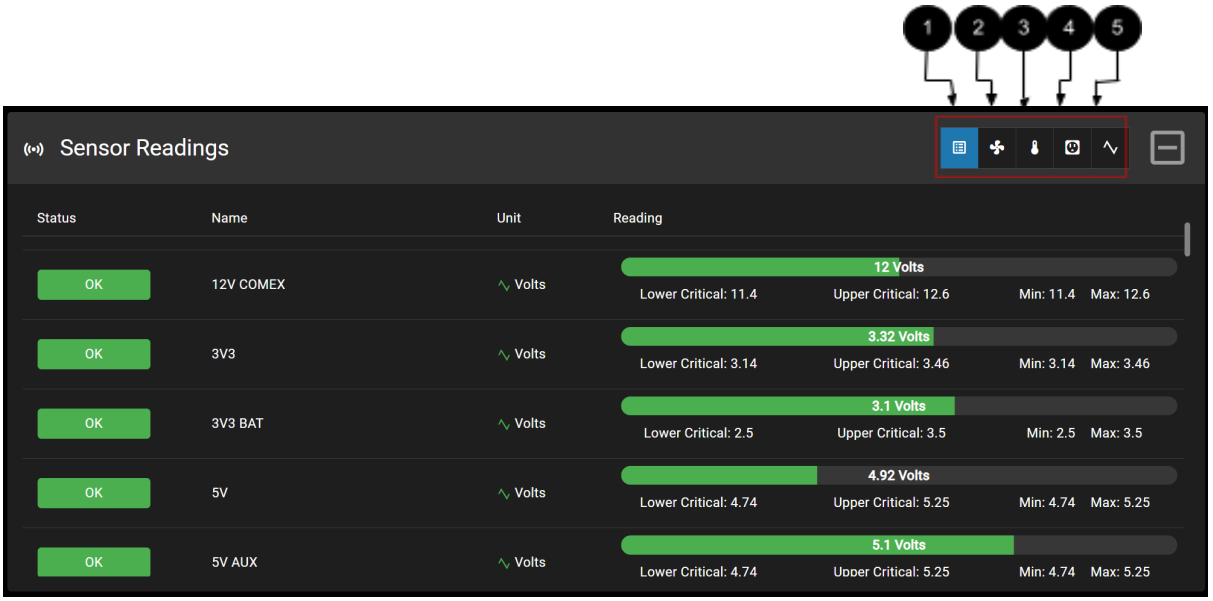


Figure 55. Sensor Readings Interface on Falcon 5012

Table 29. Features of the Sensor Reading Interface

Feature	Description	Purpose
1. Overview	Displays summarized data for fan speed, temperature, and power consumption in one view.	Enables efficient system status checks and quick identification of abnormal conditions.
2. Fan Speed Monitoring	Displays fan speed in RPM (Revolutions Per Minute).	Ensures proper cooling; alerts users to abnormal speeds to prevent overheating.
3. Temperature Monitoring	Tracks real-time chassis temperature in degrees Celsius (°C).	Maintains safe operating conditions; prevents thermal damage.
4. Power Consumption	Monitors power usage in Watts for the PSU and individual slots.	Identifies power delivery issues; ensures energy-efficient operation.
5. Voltage Monitoring	Tracks voltage levels across critical components in Volts.	Ensures stable power delivery; protects against surges or under-voltage instability.

Status Indicators

The sensor readings are represented by four key status indicators that provide real-time feedback on system health, using the following color codes:

- **Green** (OK): The system is operating within the normal, specified range.
- **Amber** (Non-critical): The system is slightly outside optimal parameters but still functional. Immediate attention is advised to prevent potential issues.
- **Red** (Critical/Non-Recoverable): A critical issue is present, and the system is operating beyond safe limits, requiring immediate corrective action to prevent permanent damage or failure.
- **Grey** (Not available/Not specified): The sensor value is either unavailable or not specified.

How to Use

- Each sensor reading is accompanied by a graphical bar with clear lower and upper threshold markers.
- When sensor readings exceed safe limits, the corresponding status indicator changes color, providing an intuitive visual alert.
- Users can view and interpret these indicators directly in the GUI for fast troubleshooting.

PCIe Link Health

The PCIe Link Health section in the Falcon 5012 GPU Expansion System GUI provides real-time monitoring of PCIe link statuses. This intuitive interface enables users to quickly assess link stability and identify potential issues affecting connectivity.

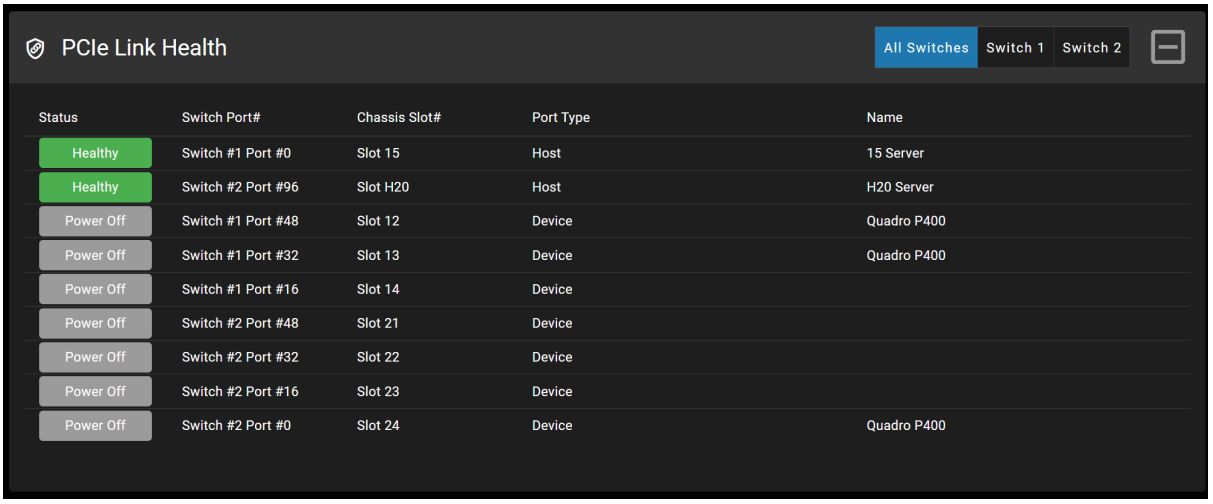


Figure 56. PCIe Link Health Interface on Falcon 5012

Status Indicators

- **Green** (Healthy): The PCIe link is active, and the system is operating within normal parameters.
- **Amber** (Non-Fatal Error/Init. Failed): Indicates a recoverable issue, such as a non-fatal error or an initialization failure. Immediate attention is recommended to avoid further complications.

- **Gray** (No Link/Power Off/Not Available): The PCIe link is inactive, unavailable, or the slot is powered off.

How to Use the Indicators

Each status indicator is displayed in the GUI alongside a graphical representation of the PCIe connection. To interpret and act on these indicators:

Step 1. Identify the Indicator Color:

- **Green**: No action required; the system is operating as expected.
- **Amber**: Review the error message or details provided in the GUI and follow the suggested resolution steps.
- **Gray**: Check hardware connections, power supply, or system configuration.

Step 2. Take Corrective Action:

- For amber indicators, resolve the reported issue promptly.
- For gray indicators, ensure the slot or device is properly connected and powered on.
- If the issue persists, follow these steps to diagnose whether it is a device or slot problem:
    1. Reinstall the device in the same slot and check if the issue persists.
    2. Move a known good device to the problematic slot to see if the issue can be reproduced.
    3. Move the potentially faulty device to a known good slot to see if the issue follows the device.

Step 3. Verify Status:

- Once the issue is resolved, confirm that the indicator changes to green, signaling normal operation.
- If the issue is determined to be with the slot, contact H3 technical support for further assistance.
- If the issue is determined to be with the device, contact the device's technical support for resolution.

PCIe Switch Throughput

The PCIe Switch Throughput chart allows administrators to monitor data transmission (in MB/s) across PCIe Switch 1 and PCIe Switch 2. It tracks ingress (data entering the PCIe device) and egress (data exiting the PCIe device) traffic for each slot, ensuring precise monitoring of system performance and connection stability.

Features

- **Ingress and Egress Data Tracking:** Monitors data flow into and out of each slot.
- **Multi-Slot Monitoring:** Uses distinct colors to represent each slot's data for easy identification.
- **Unified Display:** Consolidates throughput data for PCIe Switch 1 and PCIe Switch 2.
- **Customizable Time Period:** Allows monitoring over a user-defined duration, up to 24 hours.
- **Traffic View Options:** Lets users switch between total, ingress, and egress traffic views.
- **Interactive Data Points:** Hover over data points to view detailed throughput values.
- **Downloadable Data:** Enables data export for further analysis or record-keeping.

How to Use

Step 1. Locate the PCIe Throughput Chart:

  - Navigate to the monitoring interface to find the PCIe Switch Throughput chart (refer to Figure 57).

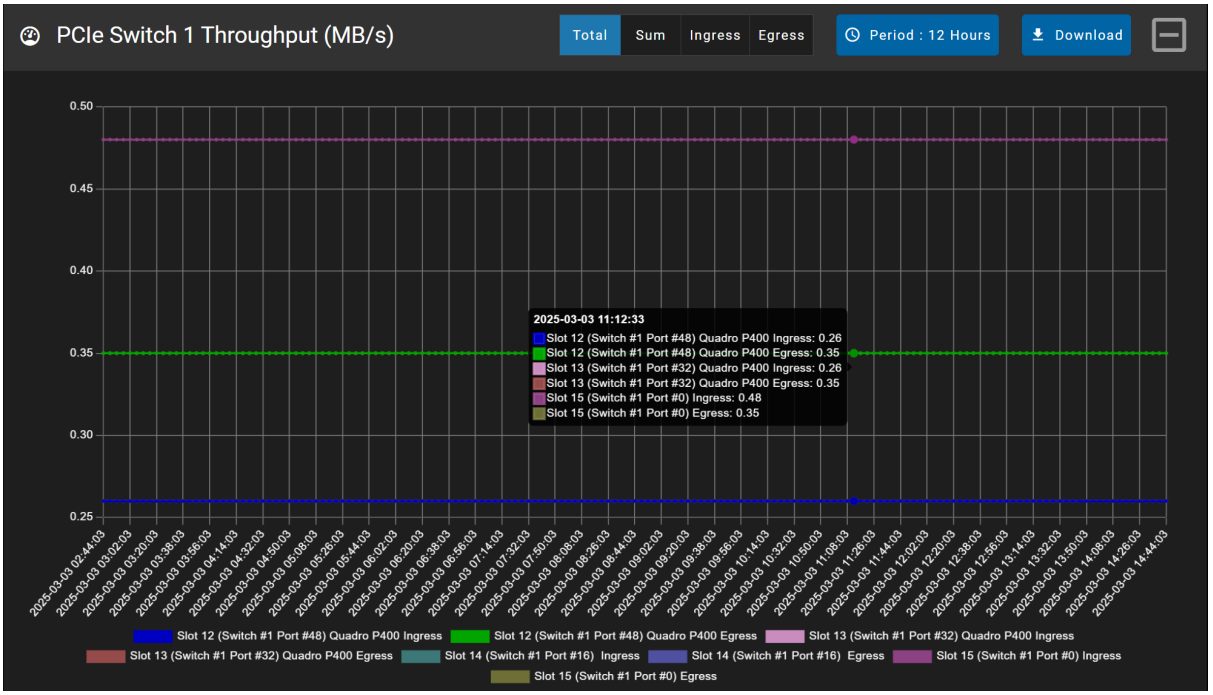


Figure 57. PCIe Switch Throughput Chart

Step 2. Select Monitoring Period:

- Use the "Period" dropdown to set a time range (up to 24 hours).

Step 3. Analyze Data:

- Review ingress and egress throughput for individual slots.
- Switch between total, ingress, and egress views for specific insights.
- Hover over data points for detailed values.

Step 4. Download Data:

- Click the "Download" button to save data for further analysis or documentation.

Fan Speed (RPM)

The Fan Speed (RPM) chart provides real-time monitoring of all system fans. It tracks inlet and outlet fan speeds at 30-second intervals over a user-defined time period, helping administrators evaluate the cooling system's performance.

Features

- **Real-Time Monitoring:** Provides continuous tracking of fan speeds for both inlet and outlet fans, updating every 30 seconds.
- **Distinct Fan Representation:** Each fan (e.g., Fan1 Inlet, Fan1 Outlet) is represented by a unique color for easy identification on the chart.
- **Interactive Data Points:** Hover over individual data points to display exact RPM values for each fan at a specific time.
- **Customizable Monitoring Period:** Allows users to set a specific time range for observation, with available options like 12 hours or 24 hours.
- **Separate Inlet and Outlet Views:** Users can toggle between Total, Inlet, and Outlet fan speeds for focused analysis.
- **Downloadable Data:** Provides the option to export fan performance data for further investigation or documentation.

How to Use

Step 1. Locate the Fan Speed Chart:

- Scroll down within the monitoring interface to find the Fan Speed (RPM) chart (refer to Figure 58).

Figure 58. Fan Speed (RPM) Chart

Step 2. Set Monitoring Period:

- Use the "Period" dropdown menu to define the time range for monitoring.

Step 3. Analyze Fan Performance:

- Observe the inlet and outlet fan speeds to ensure the cooling system operates effectively.
- Identify irregularities by hovering over specific data points for precise RPM readings.

Step 4. Download Data:

- Click the "Download" button to save the data for further analysis or documentation.

Chassis Temperature

The Chassis Temperature chart enables administrators to monitor temperature variations across key sections of the chassis in real time. This feature supports thermal management by providing precise data at 30-second intervals, ensuring optimal system performance and stability.

Features

- **Segmented Temperature Monitoring:**
    - Tracks temperatures in critical chassis sections: Front (Left, Middle, Right), Rear (Left, Middle, Right), PCIe Switches (1 and 2), and mCPU.
    - Each section is assigned a unique color for clear identification and analysis.
- **Real-Time Updates:** Temperature readings are updated every 30 seconds, allowing for timely detection of thermal changes.
- **Customizable Monitoring Period:** Users can set a specific time range, such as 12 or 24 hours, for detailed trend analysis.
- **Interactive Data Points:** Hovering over any data point displays the precise temperature value and timestamp for the corresponding chassis section.
- **Adjustable Data Granularity:**
    - Shorter monitoring periods provide highly detailed data points.
    - Longer monitoring periods display generalized trends, ensuring consistent visualization.
- **Downloadable Data:** Allows users to export temperature data for analysis or documentation purposes.

How to Use

Step 1. Locate the Chassis Temperature Chart:

- Navigate to the monitoring interface to find the Chassis Temperature chart (refer to Figure 59).

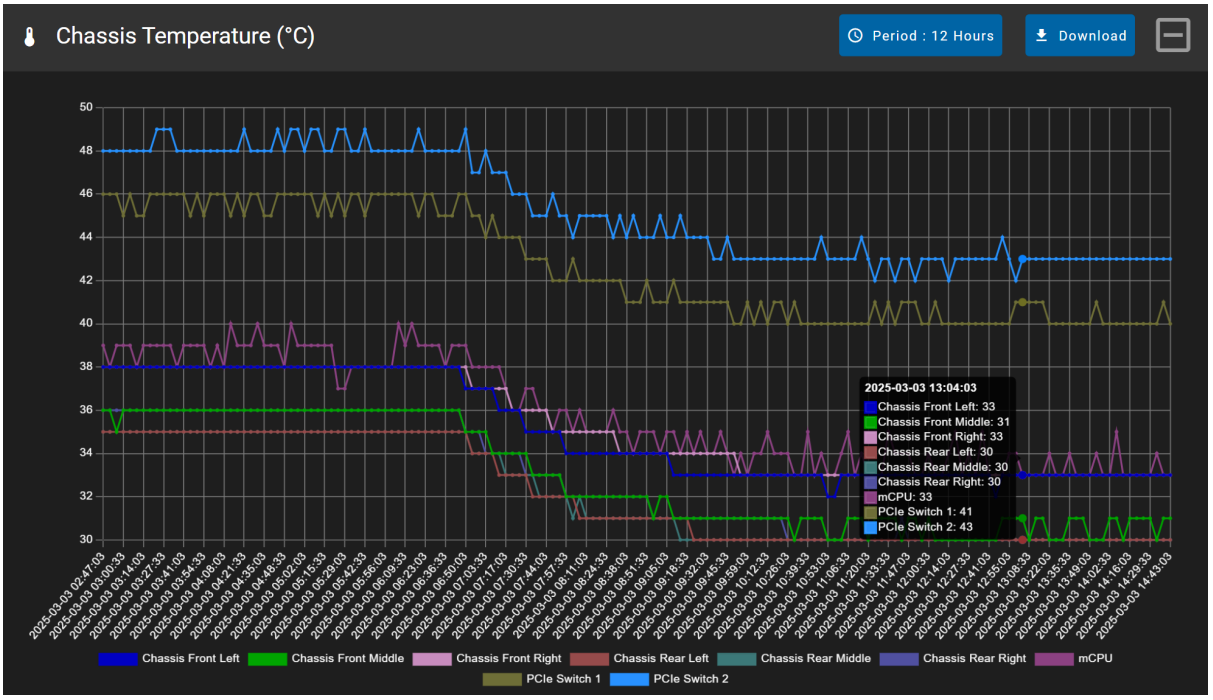


Figure 59. Chassis Temperature Chart

Step 2. Set Monitoring Period:

- Use the "Period" dropdown menu to select the desired time range for temperature tracking.

Step 3. Analyze Temperature Data:

- Observe temperature trends across chassis sections to identify potential issues.
- Hover over data points to view exact temperature readings at specific timestamps.

Step 4. Download Data:

- Click the "Download" button to save the temperature data for further analysis or reporting.

Device Temperature

The Device Temperature chart provides real-time monitoring of the thermal conditions for individual slots. It allows administrators to assess the temperature performance of devices across the system, enabling effective cooling and anomaly detection.

Features

- **Slot-Specific Monitoring:**
    - Tracks the temperature of each slot, with distinct colors for each slot to ensure clear identification.
    - Displays slot numbers and associated port numbers for precise monitoring.
- **Time-Based Temperature Trends:** Records temperature readings at regular intervals over the selected monitoring period.
- **Customizable Monitoring Period:** Allows users to select specific time ranges (e.g., 12 or 24 hours) to analyze real-time or historical temperature trends.
- **Interactive Data Points:** Hover over any data point to view exact temperature readings for a specific slot and time.
- **Downloadable Data:** Enables users to export temperature data for further analysis or documentation purposes.
- **Anomaly Detection:** Helps identify sudden temperature spikes or unusual patterns, enabling quick investigation of cooling issues or hardware malfunctions.

How to Use

Step 1. Locate the Device Temperature Chart:

- Scroll down in the monitoring interface to find the Device Temperature chart (refer to Figure 60).

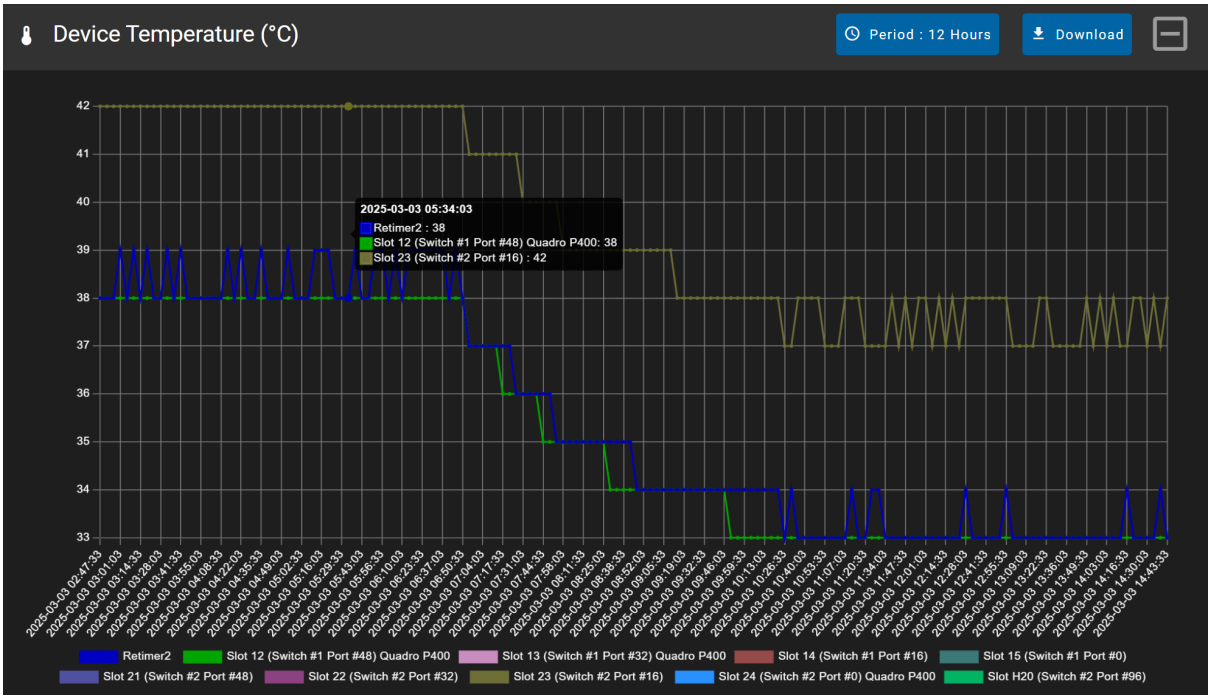


Figure 60. Device Temperature Chart

Step 2. Set Monitoring Period:

- Use the "Period" dropdown to select the desired time range for temperature monitoring.

Step 3. Analyze Device Temperatures:

- Review the temperature trends for individual slots to identify potential issues.
- Hover over specific data points to retrieve precise temperature readings.

Step 4. Download Data:

- Click the "Download" button to save temperature records for further review or reporting.

Power Consumption (Watt)

The Power Consumption chart provides real-time monitoring of the power usage of system components, including PCIe slots and PSUs (Power Supply Units). This chart enables administrators to assess energy distribution and detect potential inefficiencies.

Features

- **Power Consumption Breakdown:**
    - Tracks power usage for individual PCIe slots (e.g., Slot 11.1, Slot 13) and PSUs (e.g., PSU1, PSU2, PSU Total).
    - Distinct colors represent each component for clear identification.
- **Time-Based Power Trends:** Displays power consumption data at regular intervals over the selected monitoring period.
- **Customizable Monitoring Period:** Allows users to define a specific time range (e.g., 12 or 24 hours) to observe power trends.
- **Interactive Data Points:** Hover over data points to view exact power usage values for specific slots or PSUs at a given time.
- **Downloadable Data:** Enables users to export power consumption data for further analysis or documentation.
- **Anomaly Detection:** Highlights sudden spikes or consistent high usage, which may indicate inefficiencies or potential hardware issues

---

How to Use

Step 1. Locate the Power Consumption Chart:

- Scroll down in the monitoring interface to find the Power Consumption (Watt) chart (refer to Figure 61).

Figure 61. Power Consumption Chart

Step 2. Set Monitoring Period:

- Use the "Period" dropdown menu to select a predefined time range or customize the duration as needed.

Step 3. Analyze Power Consumption Data:

- Review power usage trends for individual slots and PSUs to identify anomalies or inefficiencies.
- Hover over specific data points to retrieve detailed power usage values.

Step 4. Download Data:

- Click the "Download" button to save power consumption records for reporting or further investigation.

4.2.2 Resource Management

The Resource Management section of the Falcon 5012 GPU Expansion system's GUI provides a detailed overview of the connected GPUs and their configurations. This section includes a list view and functionality for attaching and detaching GPUs.

Host List View

The Host List View provides users with a detailed overview of GPU resources connected to the system. It supports two viewing modes:

1. **All Switches:** Displays a consolidated summary of GPUs connected to each switch, offering an overall perspective.
2. **Individual Switches (Switch 1 or Switch 2):** Allows users to focus on GPUs connected to a specific switch for detailed monitoring and management.

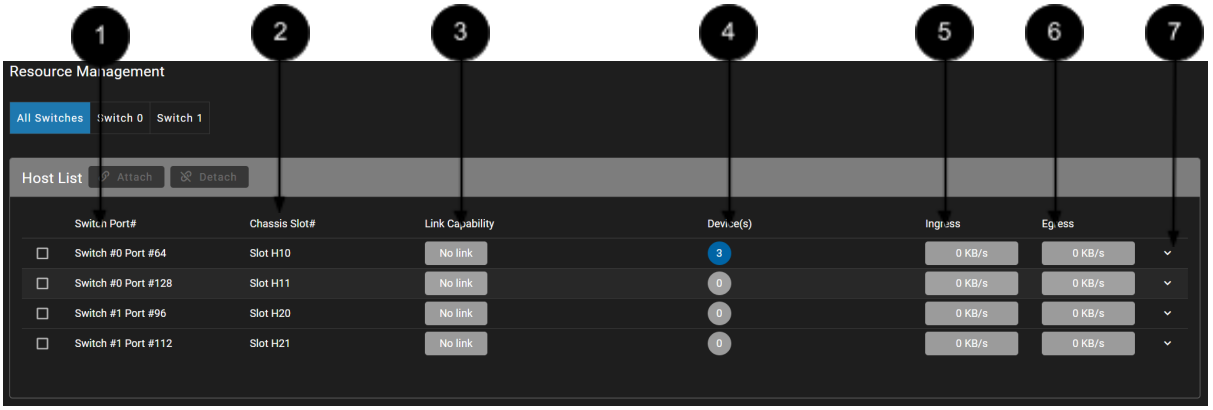


Figure 62. Picture of the list view

The attributes displayed in the Host List View are outlined in Table 30 below:

Table 30. Attributes Displayed in the Host List View

Attribute	Description
1. Switch Port#	Displays the PCIe switch and port number, e.g., Switch #1 Port #64.
2. Chassis Slot#	Shows the corresponding chassis slot number, e.g., Slot H10.
3. Link Capability	Displays the PCIe link capability, e.g., PCIe Gen5 x16.
4. Device(s)	Indicates the number of devices connected, e.g., "1" for one device, "0" otherwise.
5. Ingress	Displays the data rate entering the port, e.g., 0 KB/s.
6. Egress	Displays the data rate exiting the port, e.g., 0 KB/s.
7. Drop-down arrow	Expands to show detailed GPU information: Switch Port, Chassis Slot, Name, and Link Capability.

When the **drop-down arrow** next to the **Egress** column is clicked, it reveals additional details about the connected GPU(s) in a hierarchical format. These details are summarized in Figure 63 and Table 31 below:

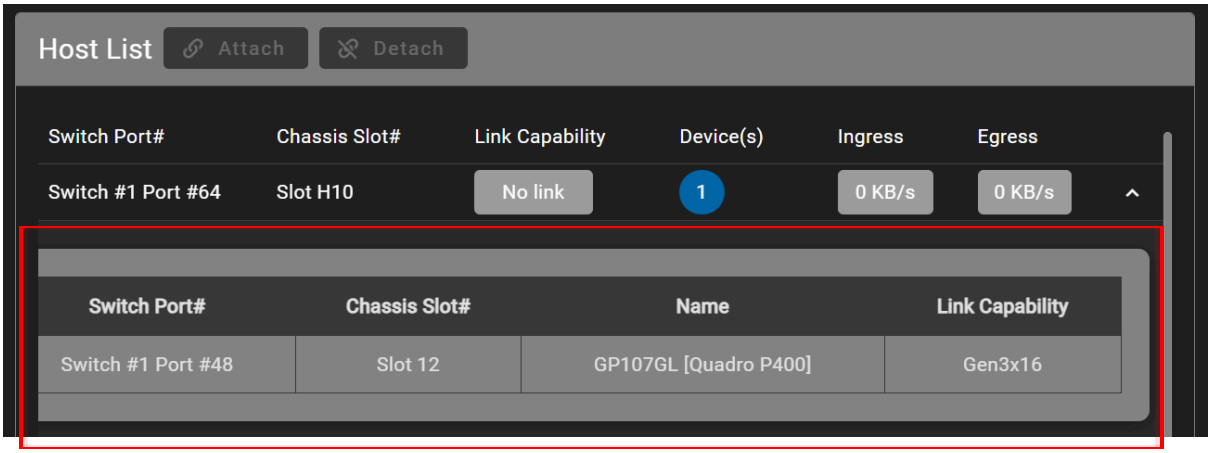


Figure 63. Host List with GPU Connection Details on Falcon 5012

Table 31. Attributes of GPU Connection Details

Attribute	Description
Switch Port#	Displays the port number of the switch to which the GPU is connected, e.g., Switch #1 Port #48.
Chassis Slot#	Indicates the physical slot location of the GPU, e.g., Slot 12.
Name	Displays the name of the connected GPU, e.g., GP107GL [Quadro P400].
Link Capability	Shows the PCIe link capability of the GPU, e.g., Gen3x16.

This intuitive layout allows users to easily access both high-level and detailed information about the GPUs connected to each host, enhancing resource management efficiency.

Device List View

The Device List View provides a detailed overview of devices connected to the system, including GPUs and other PCIe devices. Administrators can monitor and manage devices efficiently using the attributes listed in Table 32 below.

Figure 64. Overview of Device List View

Table 32. Device List Attributes:

Attribute	Description
1. Switch Port#	Displays the PCIe switch and port number to which the device is connected, e.g., Switch #1 Port #48.
2. Chassis Slot#	Indicates the physical chassis slot where the device is installed, e.g., Slot 12.
3. Name	Shows the name of the connected device, e.g., GP107GL [Quadro P400]. Displays "Empty" if unoccupied.
4. Power Status	Indicates whether the slot is powered on ("On") or off ("Off").
5. Link Capability	Displays the PCIe link capability of the slot, e.g., Gen3x16. Displays "No link" if no device is connected.
6. Status	Indicates the current status of the device: Available (ready for use), Attached (in use by a host), or Empty (no device installed).
7. Assigned to	Displays the host system to which the device is assigned, e.g., Slot H10. Displays "Empty" if unassigned.
8. Temperature	Displays the current temperature of the device, e.g., 28.00 °C.
9. Ingress	Shows the data rate entering the port, e.g., 269.70 KB/s.
10. Egress	Shows the data rate exiting the port, e.g., 359.60 KB/s.
11. Drop-down Arrow	Expands to show detailed device information, including product and vendor details.

Drop-down Details:

When the drop-down arrow is clicked, additional information about the selected device is displayed (refer to Figure 65), as outlined in Table 33 below:

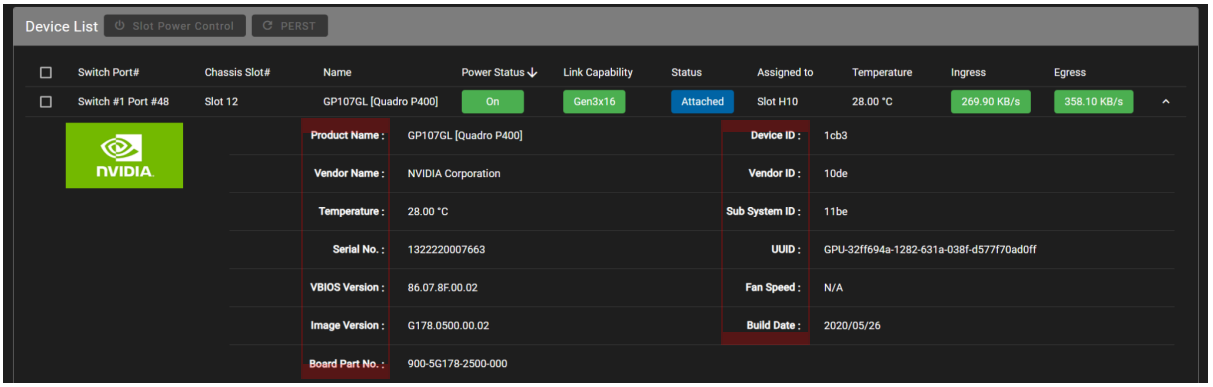


Figure 65. Detailed View of GPU Information

Table 33. Attributes of GPU Information in Dropdown Panel

Attribute	Description
Product Name	Displays the product name of the connected device, e.g., GP107GL [Quadro P400].
Vendor Name	Shows the name of the device vendor, e.g., NVIDIA Corporation.
Temperature	Displays the current temperature of the device, e.g., 28.00 °C.
Serial Number	Displays the serial number of the device, e.g., 1322220007663.
Device ID	Displays the device ID, e.g., 1cb3.
Vendor ID	Displays the vendor ID, e.g., 10de.
Sub System ID	Displays the sub-system ID, e.g., 11be.
UUID	Shows the unique identifier of the GPU, e.g., GPU-32ff694a-1282-631a-038f-d577f70ad0ff.
Fan Speed	Displays the fan speed if available, e.g., N/A if unavailable.
Build Date	Displays the manufacturing date of the device, e.g., 2020/05/26.
VBIOS Version	Displays the firmware version of the device, e.g., 86.07.8F.00.02.
Board Part Number	Displays the board part number, e.g., 900-5G178-2500-000.

GPU Assign and Unassign

This section explains the step-by-step procedures to assign (attach) or unassign (detach) GPUs to hosts using the Resource Management interface.

Step 1. Assign GPUs in Advanced Mode

A. Navigate to the Resource Management interface in the GUI(refer to Figure 66).

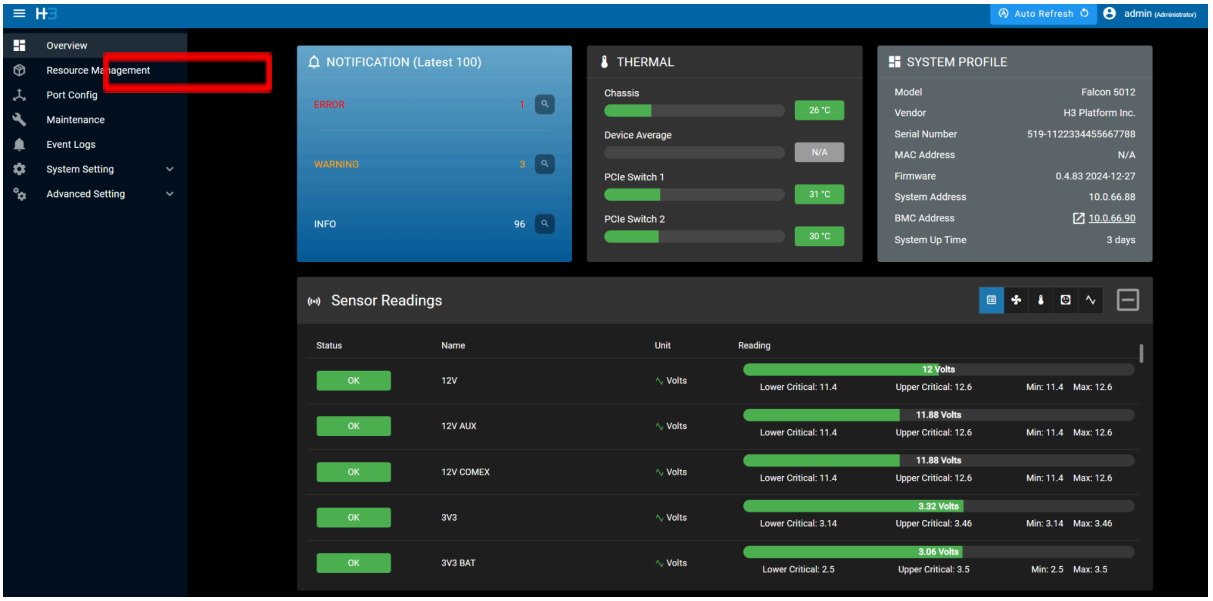


Figure 66. Accessing the Resource Management Interface on GUI

B. Select the host to which you want to assign GPUs (refer to Figure 67).

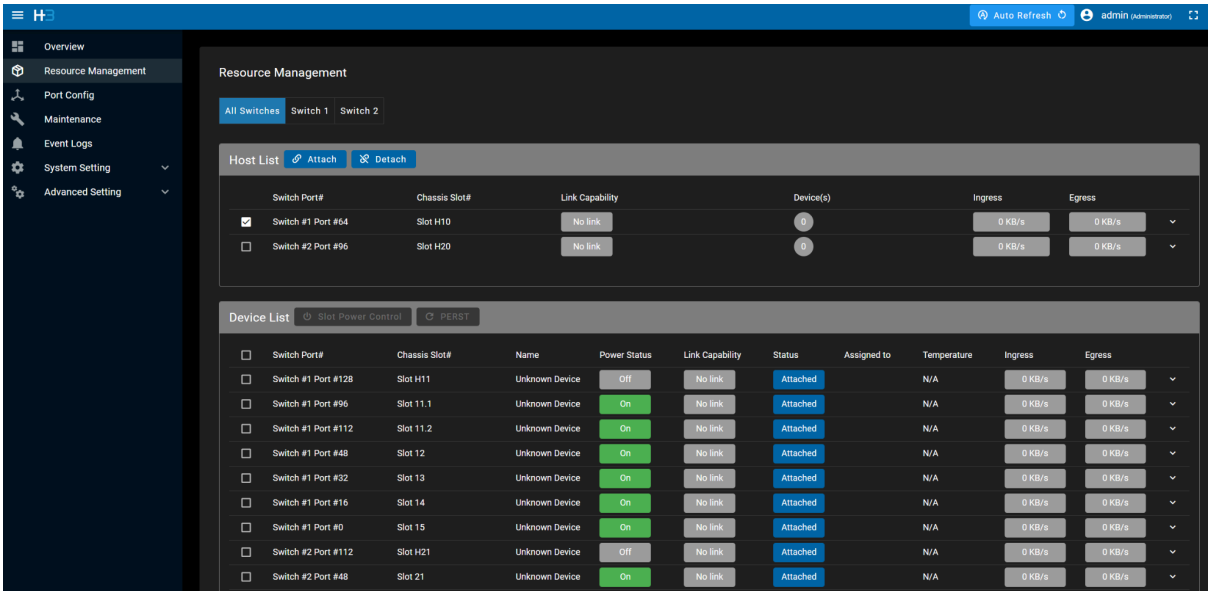


Figure 67. Selecting a Host for GPU Assignment in Resource Management

C. Click the Attach tab to access the GPU allocation page (refer to Figure 68).

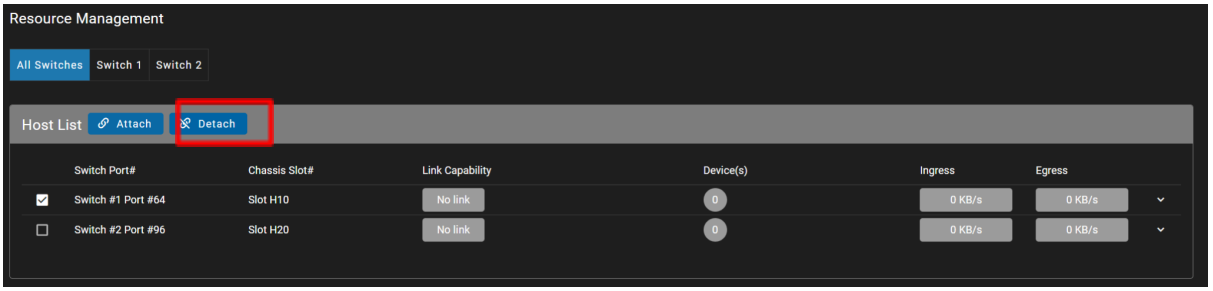


Figure 68. Accessing the GPU Allocation Page via the Attach Tab

D. From the displayed GPU list, select the GPUs you want to assign (refer to Figure 69).

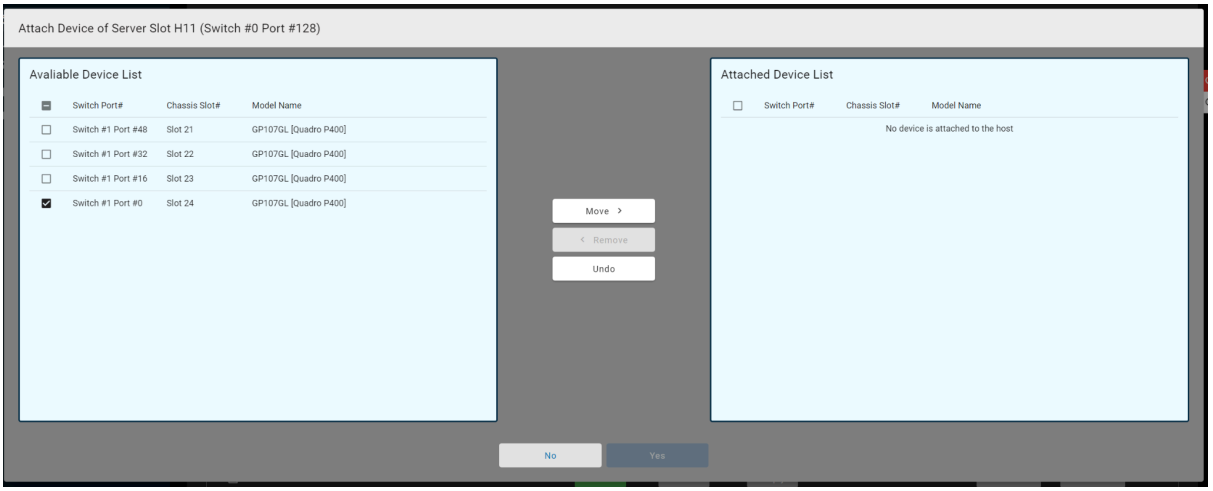


Figure 69. Selecting GPUs from the Available Device List

E. Click Move to transfer the selected GPUs to the Attached GPUs column (refer to Figure 70).

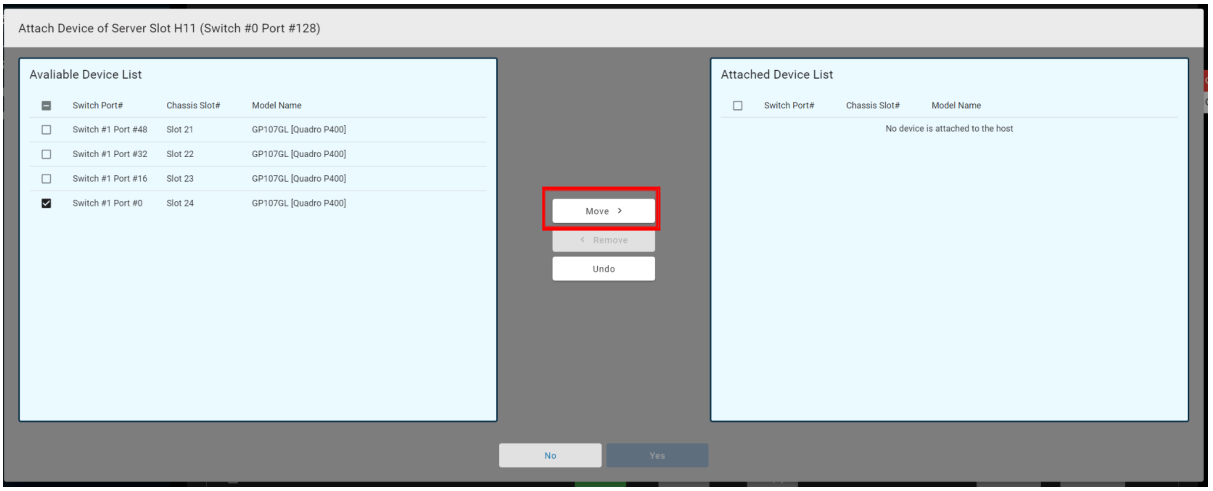


Figure 70. Moving GPUs from the Available Device List to the Attached Device List

F. Confirm the selection by clicking Yes to complete the GPU assignment (refer to Figure 71).

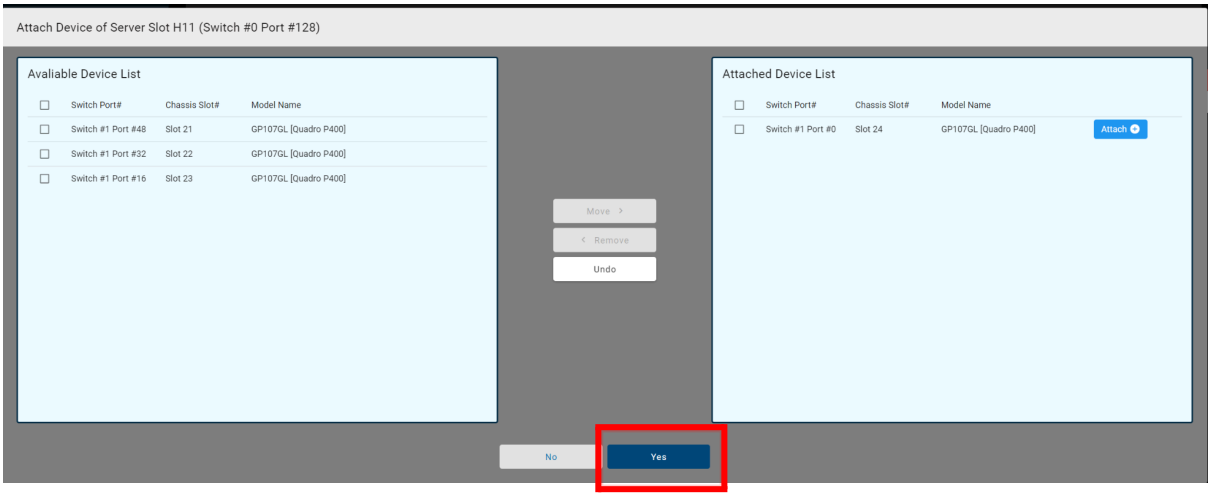


Figure 71. Confirming GPU Assignment by Clicking "Yes"

Step 2. Unassign GPUs in Advanced Mode

A. Navigate to the Resource Management interface in the GUI (refer to Figure 72).

B. Select the host from which you want to unassign GPUs.

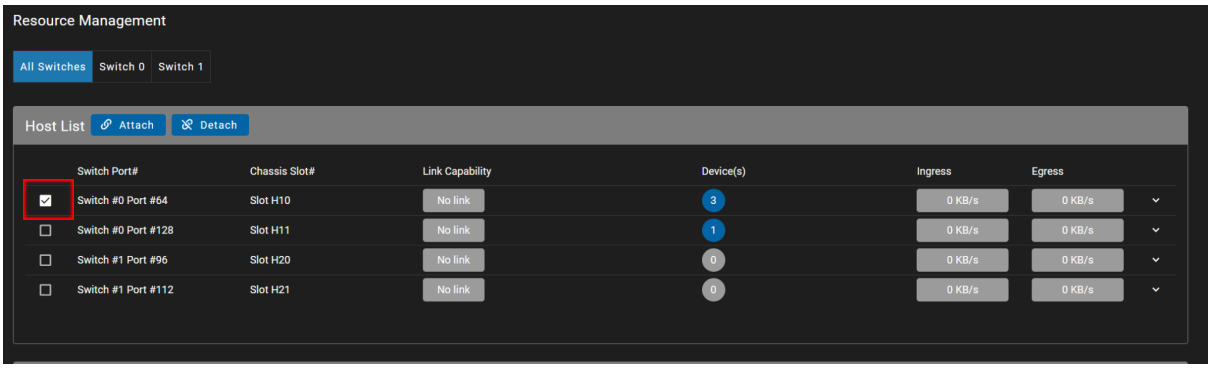


Figure 72. Selecting Host for GPU Unassignment on GUI

C. Click the Detach tab to access the unassignment page (refer to Figure 73).

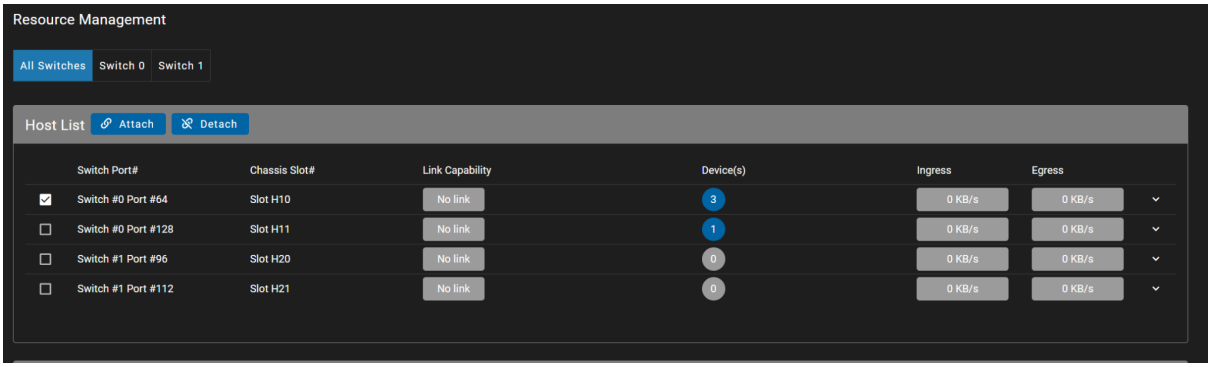


Figure 73. Accessing the Unassignment Page via the Detach Tab

D. From the Attached GPUs column, check the GPUs you want to unassign (refer to Figure 74).

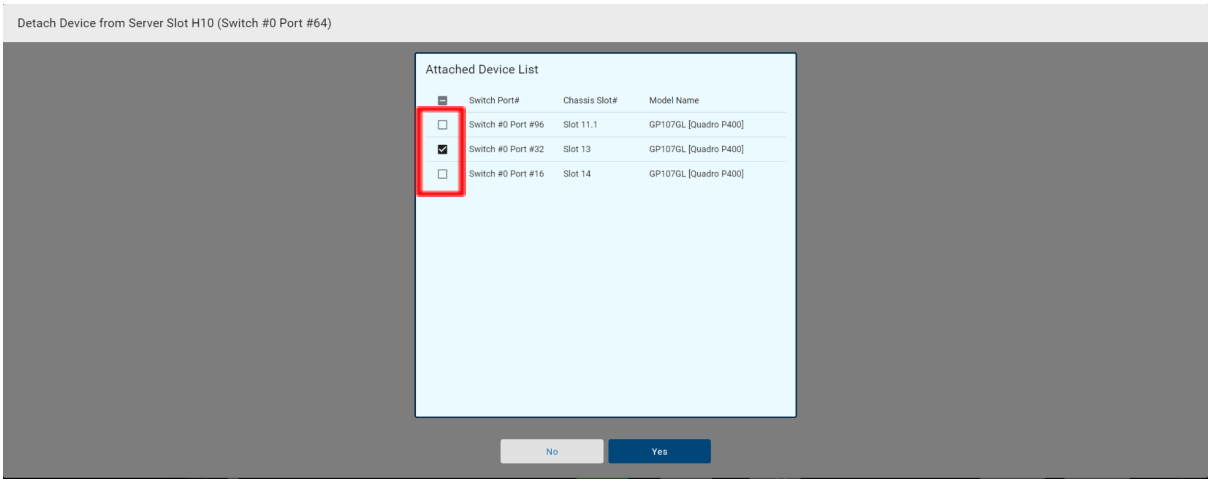


Figure 74. Checking the GPUs for Detachment

E. Click Yes to detach the selected GPUs back to the Available GPUs column.

Step 3. Verify Connections

A. Ensure that the CDFP cable is securely connected between the chassis and host (refer to Figure 75).

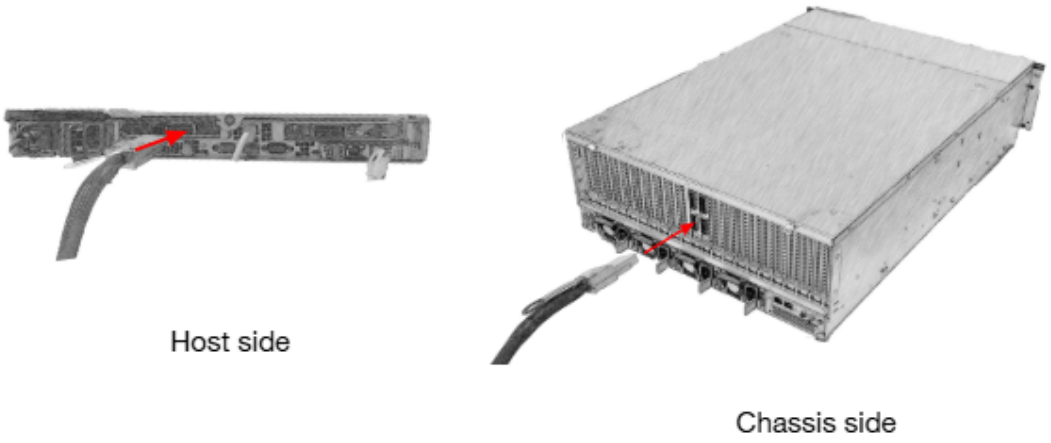


Figure 75. Host and Chassis Connections Using CDFP Cables

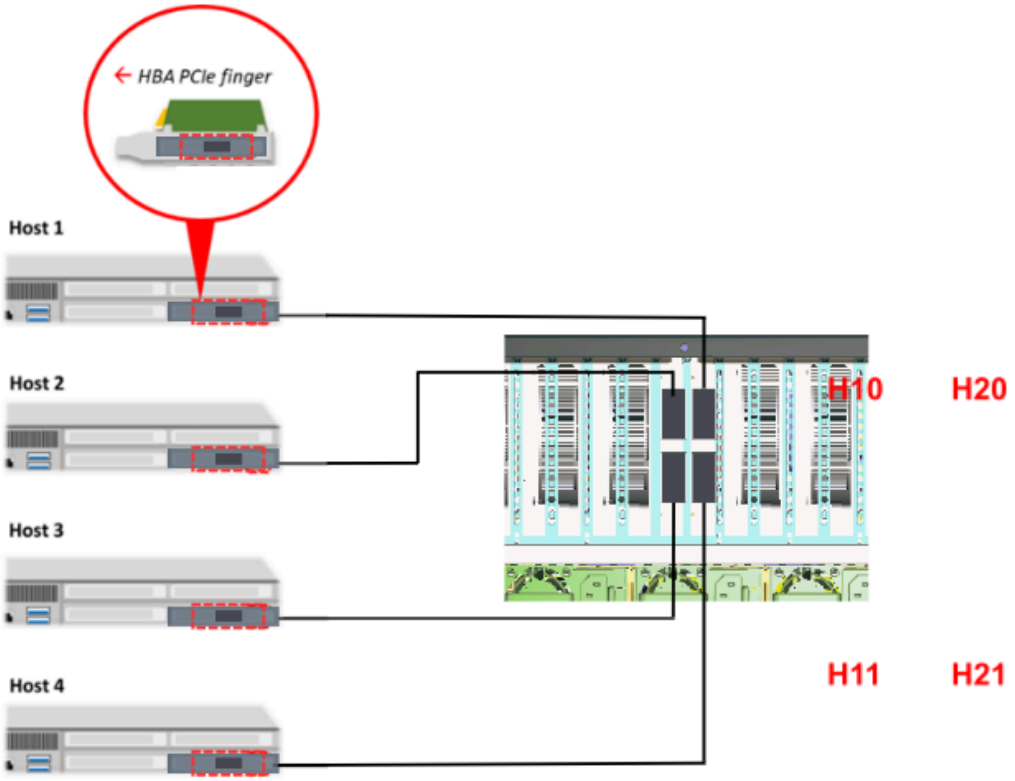


Figure 76. CDFP Cable Mapping Between Hosts and Chassis Ports

B. Confirm that GPUs are properly assigned to their respective hosts in the GUI.
C. Power on the server after verifying all connections.

Step 4. Verifying GPU Resource Assignment or Unassignment on the Host

This procedure outlines how to verify GPU assignment or unassignment on the host system, ensuring GPUs are correctly mapped, detected, and operational through server commands and GUI checks.

About this task:

After assigning or unassigning GPU resources, the following verifications are necessary:

- Ensure GPUs are properly mapped in the PCIe tree structure on the server.
- Confirm the total number of GPUs matches the expected configuration.
- Verify that the host GUI reflects the correct PCIe link capability and recognizes all assigned GPUs.

Procedures

A. Verify PCIe Tree Structure on the Server
   - On the server, open a terminal and run the following command:

```
None

lspci -vt
```

   - Review the output to confirm GPU assignments.

B. Verify Total GPUs on the Server
   - Run the following command to check the total number of NVIDIA GPUs on the server:

```
None

nvidia-smi
```

   - Verify that the total GPU count matches the expected number.

C. Verify GPU Status on the Host via GUI

- Access the system’s Web GUI and navigate to the Resource Management panel.
- Confirm the following:
  - Link Capability: Ensure PCIe links are displayed with the correct specifications (e.g., Gen4x16 or Gen5x16).
  - Device(s): Verify the assigned GPUs are listed and operational. For example, if the Device(s) field displays "4," it indicates that four GPUs are correctly assigned and connected.

Results

When the procedure is successfully completed:

- The server recognizes the assigned GPUs in the PCIe tree and displays the correct count.
- The host GUI confirms accurate PCIe link capability and GPU assignments.

If any discrepancies are identified, contact technical support for further diagnostics and solutions.

4.2.3 Port Configuration

The Port Configuration functionality in the Falcon 5012 GPU Expansion System's GUI allows users to configure switch ports as either Host Ports or Device Ports. This feature is essential for managing system roles and expanding host capacity when necessary.

This section provides detailed steps to access the Port Configuration interface, select and configure ports, and apply changes for both standard setups and host expansion scenarios.

About Port Configuration

Port Configuration varies based on the operational scenario:

- Without Host Expansion (Server Side): No additional configuration is required in the GUI.
- With Host Expansion (Chassis Side): Certain expansion cables must be reconfigured to enable host functionality, depending on the GPU count and system requirements.

> **Note:** Installing a retimer is required to enable host expansion. For detailed instructions on retimer installation, please refer to Section 3.5: Retimer Installation and Host Configuration

---

Procedures

Step 1. Configure Slots for Host Expansion (If Applicable)

For setups involving host expansion, reconfigure the following slots to enable host functionality:

- 10 GPUs: Configure Slot 11 and Slot 25.
- 8 GPUs + 4 NICs: Configure Slot 11.1, 11.2, 25.1, and 25.2.

Ensure that a retimer is installed in the specified slots before proceeding with the configuration. Refer to Section 3.5 for installation details.

Step 2. Access the Port Configuration Interface

Launch the GUI and navigate to the Port Configuration interface.
(Refer to Figure 77: Port Configuration Interface for Falcon 5012 System)

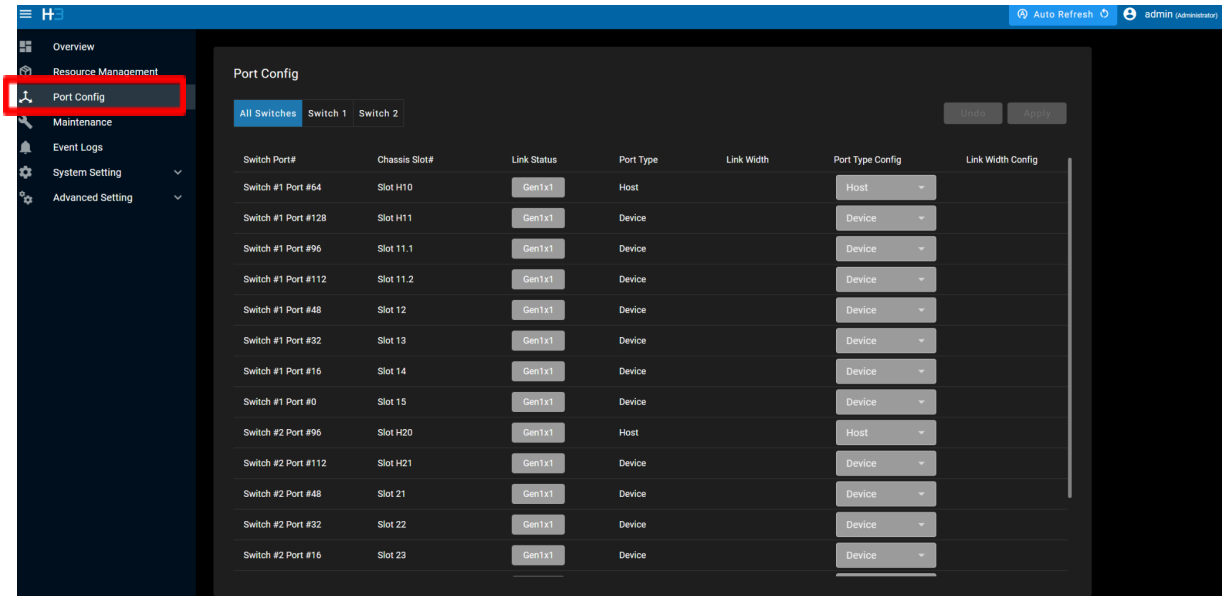


Figure 77. Port Configuration Interface for Falcon 5012 System

---

Step 3. Select the Port to Configure

A. From the Port Configuration List, select the target port to configure.
(Refer to Figure 78: Picture of Port Configuration List)

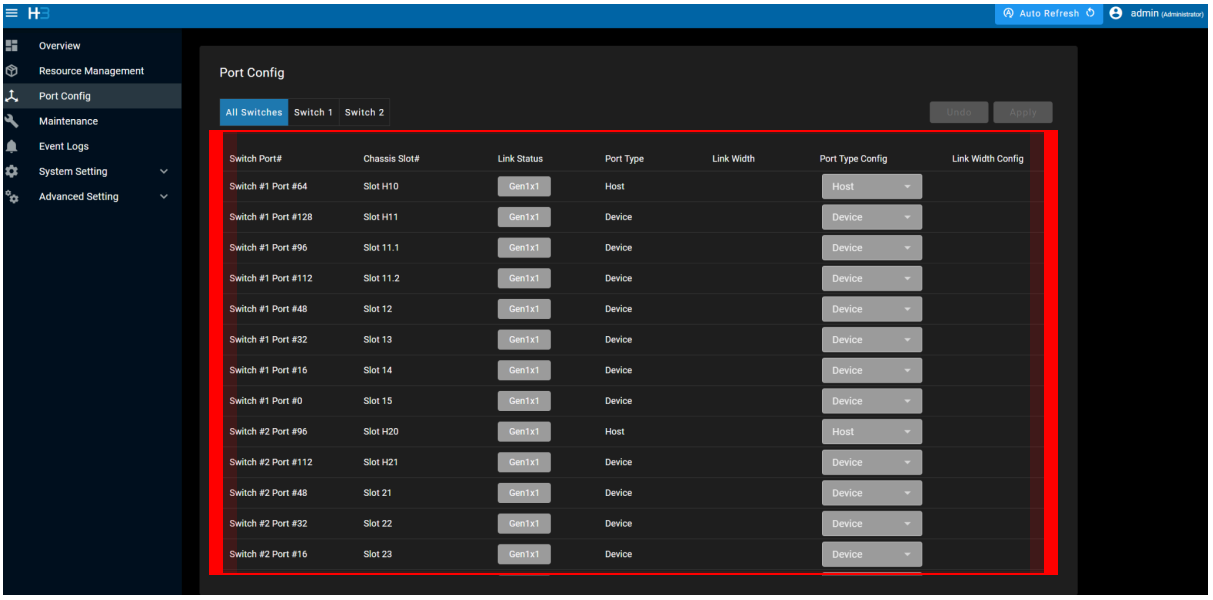
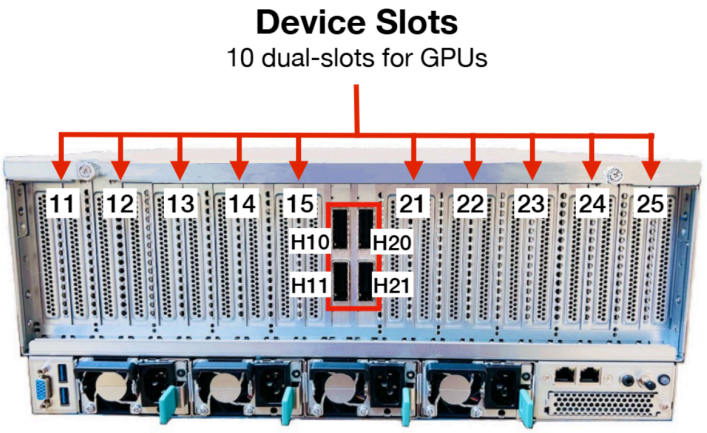


Figure 78. Picture of Port Configuration List

B. Identify the corresponding Slot Number on the chassis to ensure you are configuring the correct physical port (Refer to Figure 79: Pictures of Slot Number on the Chassis):

- 10 GPUs: Verify Slot 11 and Slot 25.



- 8 GPUs + 4 NICs: Verify Slot 11.1, 11.2, 25.1, and 25.2.

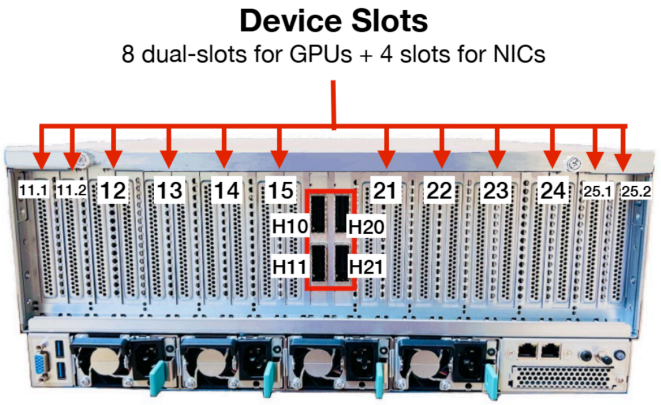


Figure 79. Pictures of Slot Number on the Chassis

Step 4. Select and Apply the Port Type

A. Choose the desired port type (refer to Figure 80):
   - Host: For ports connected to host systems.
   - Device: For ports connected to GPU devices or other peripherals.

B. Click Apply to initiate the port configuration.

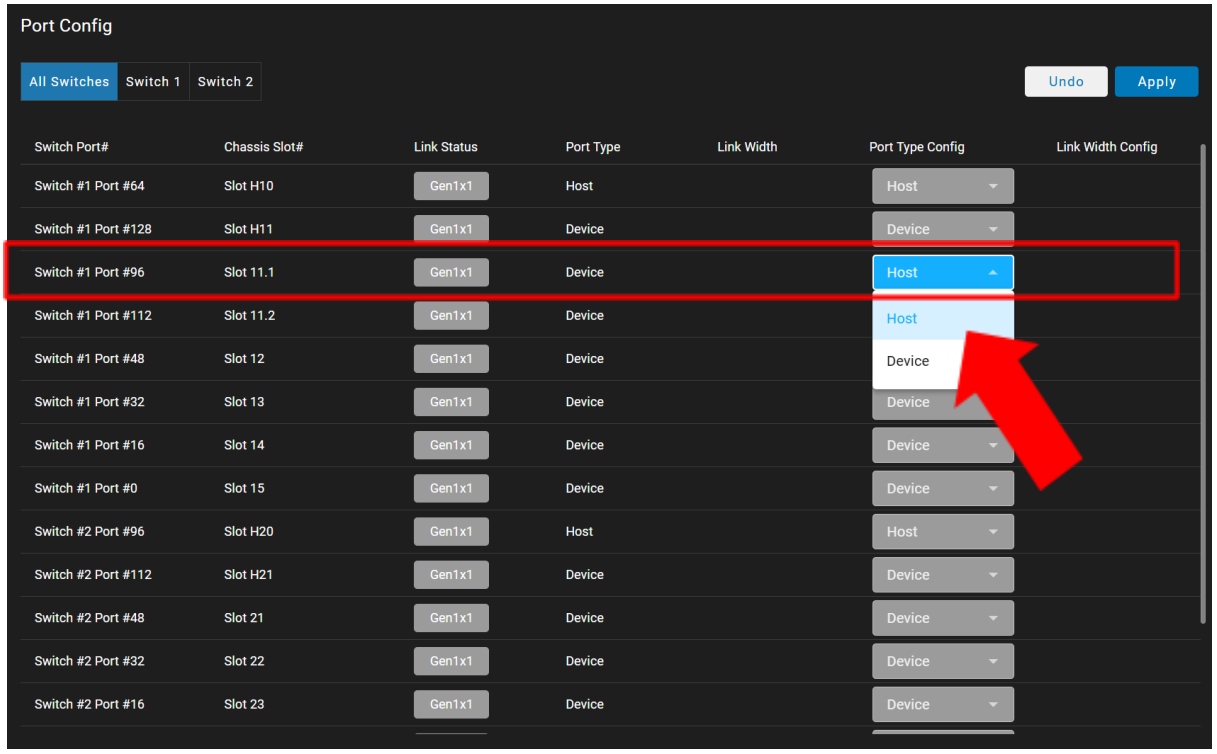


Figure 80. Changing Port Configuration from Device to Host in expansion cables

---

Step 5. Confirm and Finalize the Configuration

A. When prompted, confirm the port configuration:
   - Click Yes in the confirmation tab to proceed.
   - Click Close to finalize the process.

B. Verify the success message in the GUI to ensure the port configuration has been applied correctly.

Results

After completing the port configuration:

- The selected ports will function as Host Ports or Device Ports according to the applied settings.
- Host expansion setups will now support additional hosts, enabling flexible and scalable GPU resource allocation.

Note: A system reboot is required for the changes to take effect.
For additional instructions on system maintenance and updates, refer to the next section.

4.2.4 Maintenance

This procedure provides instructions for updating the firmware on Falcon 5012 using the GUI. The system automatically detects the type of firmware being updated, such as System Management firmware or PCIe switch firmware. Ensure all host machines are disconnected before proceeding.

About this Task

To update the firmware, you will:

1. Download the appropriate firmware file from the support team.
2. Access the Falcon 5012 GUI and navigate to the Maintenance section.
3. Upload the firmware file and initiate the update process.
4. Reboot the system to complete the update.

Tools and Materials

You will need:

- A PC with access to the Falcon 5012 GUI.
- The firmware file (.img format) provided by support@h3platform.com.

Procedures

Step 1. Obtain the Firmware File

A. Download the Firmware:
   - Access the H3 Platform support page:
     - Navigate to Support → Knowledge Base → Download on the H3 Platform website.
     - Alternatively, visit the following URL directly: https://www.h3platform.com/knowledge-base/document.
   - Select the appropriate options:
     - Product Type: Composable GPU Chassis.
     - Model Type: Falcon 5012.
   - Download the following firmware files:
     - PCIe Switch 1 Firmware (SwitchA)
     - PCIe Switch 2 Firmware (SwitchB)
     - System Management Firmware

B. Contact H3 Platform Support:
    - Email support@h3platform.com to request the latest firmware file.

Note: The system includes two PCIe switches, each requiring a separate firmware update. Ensure that the correct firmware is applied to each switch to maintain system stability and functionality.

- Switch 1 firmware is typically labeled as SwitchA and follows the naming convention:
  h3-pcifw-aicSwitchA_YYYYMMDDHHMM.fw.h3.signed
- Switch 2 firmware is labeled as SwitchB and follows the naming convention:
  h3-pcifw-aicSwitchB_YYYYMMDDHHMM.fw.h3.signed

Before proceeding with the update, verify the firmware file names to confirm they correspond to the correct switch. Applying incorrect firmware may result in system errors or unexpected behavior.

Step 2. Access the Maintenance Page

A. Log in to the Falcon 5012 GUI:
    - Enter the system's IP address in your web browser.
    - Log in using your credentials.
B. From the left navigation menu, click Maintenance to access the firmware update tools (refer to Figure 81).

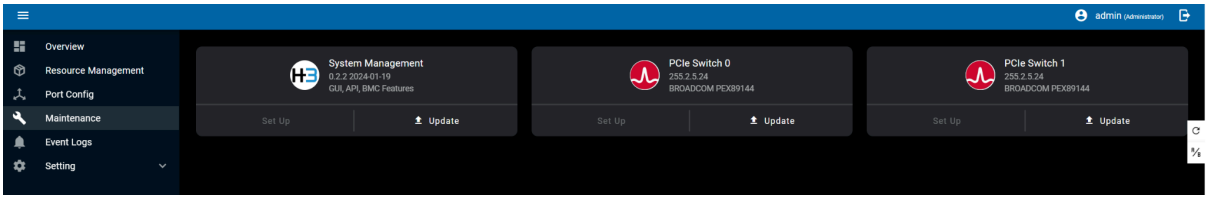


Figure 81. Firmware Update Tools on the Maintenance Page

Step 3. Upload the Firmware File

A. On the Maintenance page, click the Upload/Install button.
B. Select the system management firmware file (.img format) from your device and upload it (refer to Figure 82).

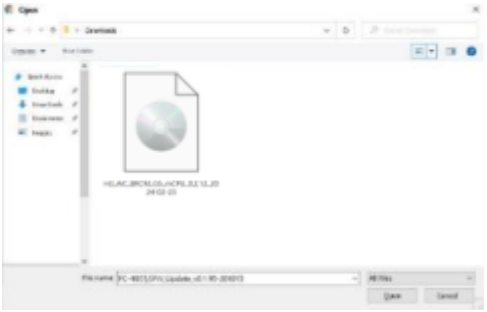


Figure 82. Selecting the Firmware File from the Download Location

C. A confirmation message will appear:
    - Verify the file name and file size of the firmware file you selected.
    - Click Yes to proceed with the update (refer to Figure 83).

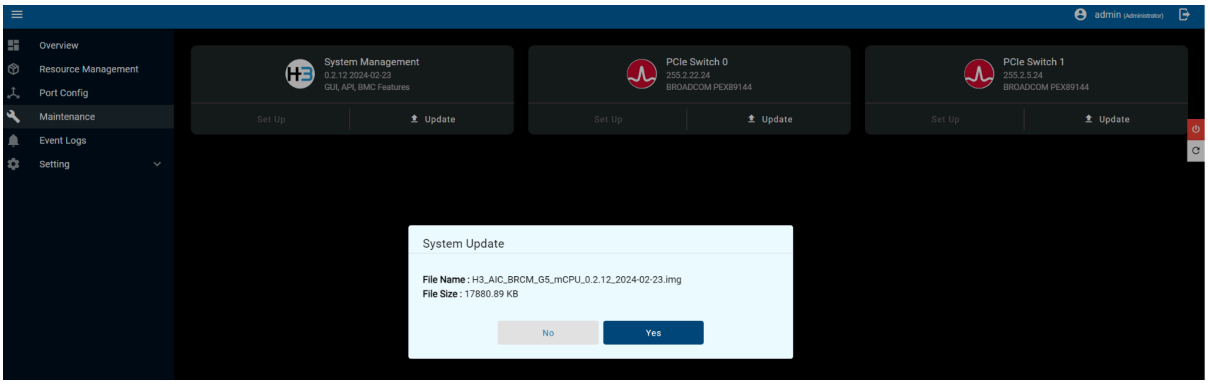


Figure 83. Confirmation Message for Firmware Update

Step 4. Firmware Update and System Reboot

A. Wait for the update to complete.
B. Before clicking Restart Now, ensure that all connected servers are powered off to proceed safely with the operation.
C. Once the update finishes, click Restart Now to reboot the system (refer to Figure 84).

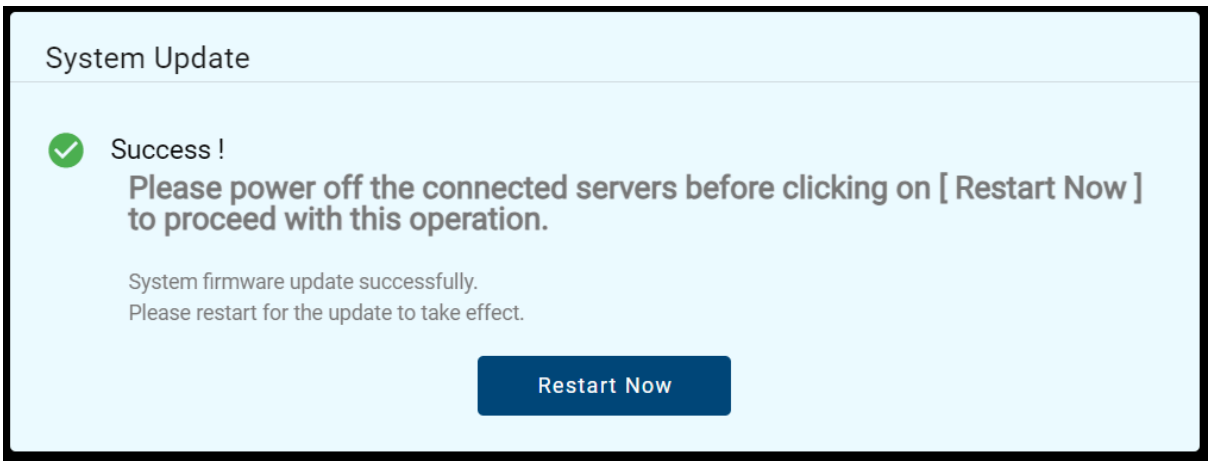


Figure 84. Firmware Update Success Prompt with Restart Now Option

---

Step 5. Post-Update Validation

1. After rebooting, the firmware update is complete.
2. If another user logs in or opens a new GUI session after the update, a reminder message will appear next to the Maintenance menu on every page (refer to Figure 85).

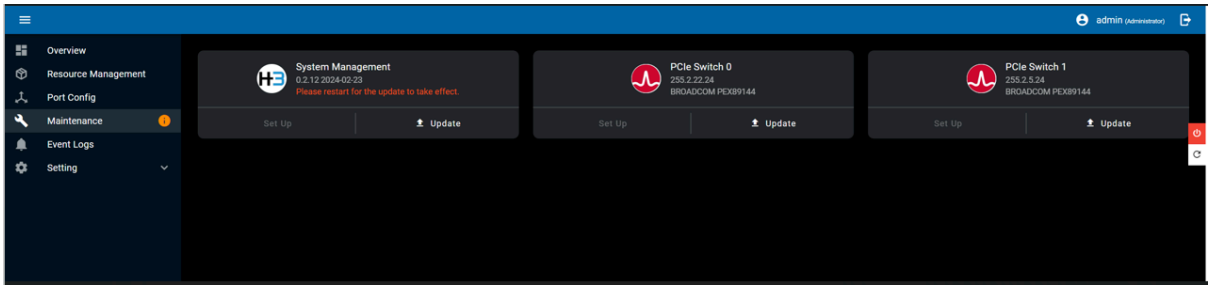


Figure 85. Post-Update Reminder Prompt in Maintenance Menu

3. To finalize the update, click the Restart button as prompted.

Notes:

- Ensure all host machines are disconnected during the update to prevent disruptions.
- If issues arise during the update, contact technical support at support@h3platform.com.

4.2.5 Event Logs

The Event Logs page provides users with a comprehensive view of system events. Logs are categorized by severity, enabling efficient management and quick issue identification. Additional functionalities, such as filtering and search, ensure precise log handling.

Refer to Figure 86 for a visual overview of the Event Logs page layout and its key features. Detailed descriptions of these features, including log categories, search functionality, and page navigation, are provided in Table 34. These tools allow users to effectively manage and analyze system events.

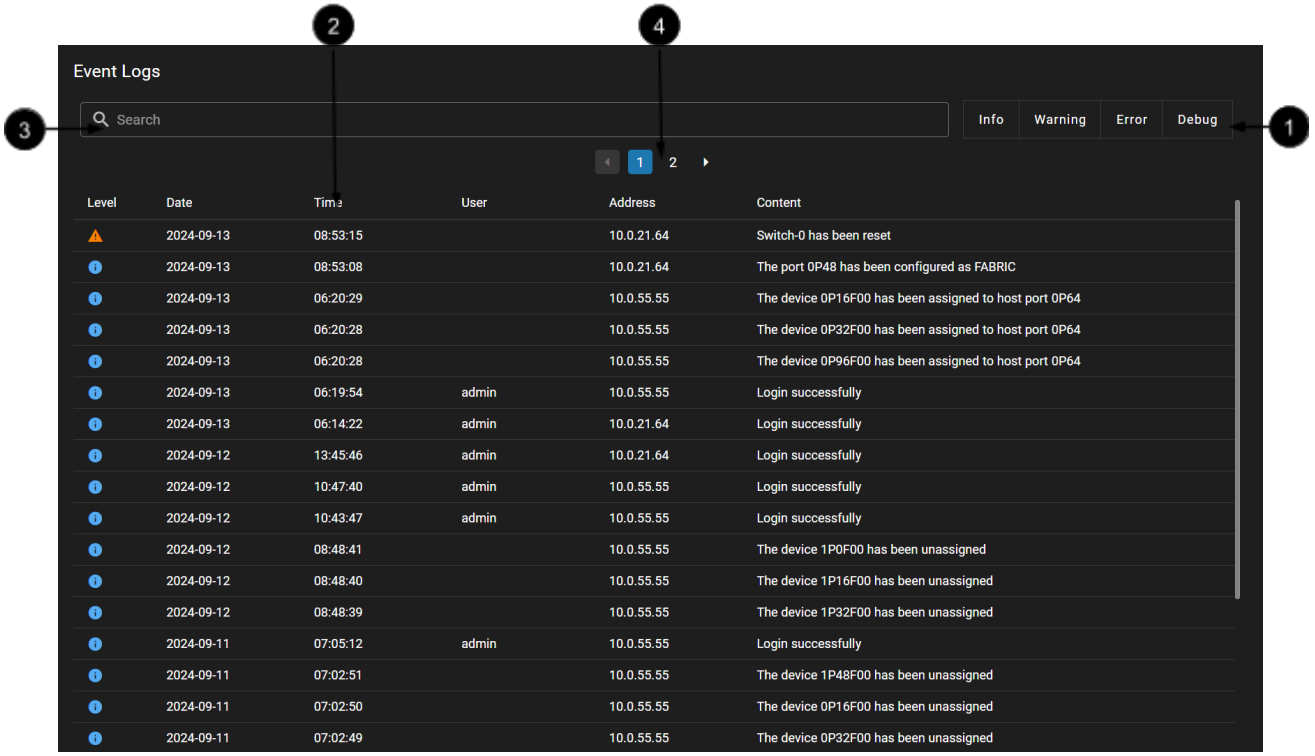


Figure 86. Picture of event log page

Table 34. Features of the Event Logs Page

1. Log Categories:	Filter logs by categories.
2. Logs:	Event logs are ordered from newest to oldest.
3. Search bar:	Search for specific logs.
4. Page:	Select pages of logs.

Note 8: Logs in bold text are unread logs.
Note 9:
Info: regular system events
Warning: moderate severity, events that require attention
Error: highest severity, events that may damage the system
Log-in: user account-related activities

Event Log List and Countermeasures

Event Type	Event Message	Level Defined	Countermeasure
General	Logged in Successfully	Info.	N/A
General	Network set up successfully	Info.	N/A
General	System restarted	Info.	N/A
Firmware	System firmware updated	Info.	N/A
Firmware	Switch-<switch_id> firmware updated	Info.	N/A
Device	Device <dev_id> assigned to host port <host_id>	Info.	N/A
Device	Device <dev_id> unassigned	Info.	N/A
Port	Port <port_id> configured as <PortType>	Info.	N/A
Port	Port <port_id> configured to <PortConfig>	Info.	N/A
Slot	Slot <h3_slot_number> powered <pw_status>	Info.	N/A
Slot	Slot <h3_slot_number> reset	Info.	N/A

Switch	Switch-<switch_id> reset	Warning	This log is triggered by the API. No further measure is required.
Switch	Error running command: <cmd> on switch <switch_id>	Error	Web GUI: Switch power cycle
Switch	Switch-<switch_id> not responding	Error	Web GUI: Switch power cycle
Device	Failed to assign device <dev_id> to host port <host_id>	Error	Web GUI: (1) Download diagnostic data to record logs when a failure appears (2) Slot power cycle
Device	Failed to unassign device <dev_id>	Error	WebGUI: (1) Download diagnostic data to record logs when a failure appears (2) Slot power cycle
Slot	Failed to power <pw_status> slot <h3_slot_number>	Error	BMC GUI: Chassis power cycle
Slot	Failed to reset slot <h3_slot_number>	Error	BMC GUI: Chassis power cycle

The following section outlines the procedures for configuring key system settings, including date and time, network parameters, and user management.

4.2.6 System Setting

The Setting section of the Falcon 5012 GPU Expansion system's GUI allows users to configure various system parameters, including date-time setting, network setting, user management (refer to Figure 87). This section provides the necessary

tools to ensure the system operates following user preferences and organizational requirements.

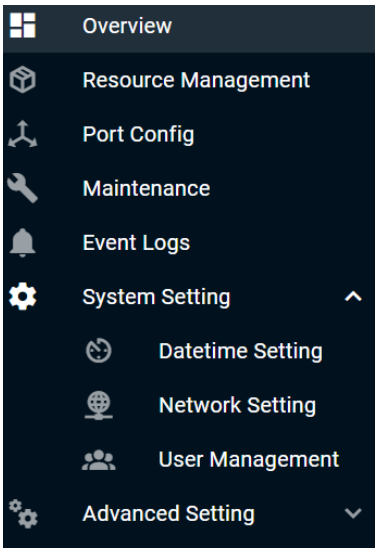


Figure 87. Picture of System Setting Function on the menu list

Datetime Setting

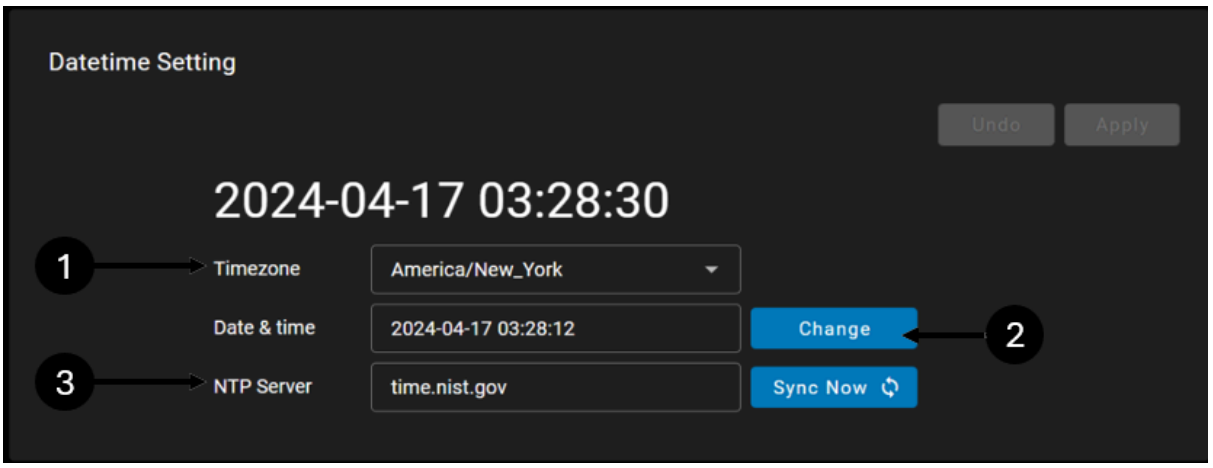


Figure 88. Datetime Setting Interface

The Datetime Setting page (refer to Figure 88) allows users to adjust the system's date and time settings to ensure accurate log timestamps and network synchronization. The interface includes options for setting the time zone, manually adjusting date and time, and synchronizing with an NTP server.
Table 35 summarizes the key functionalities of the Datetime Setting interface:

Table 35. Key functionality of Datetime Setting Interface

1. Time zone:	Set/modify system time zone.
2. Change:	Manually set/modify the date and time with the calendar tool.
3. Sync. with NTP server:	Sync the system with an NTP server. (Requires NT server IP address)

Procedures:

Access the Datetime Setting Page

Navigate to the Datetime Setting section in the system's GUI. The interface includes fields for Timezone, Date & Time, and NTP Server (refer to the figure above).

Step 1. Configure the Time Zone

A. Select the desired time zone from the dropdown menu under Timezone.
B. Verify the selected time zone matches your location.

Step 2. Modify Date and Time Manually

A. Click the Change button next to the Date & Time field.
B. Use the provided tool to set the correct date and time.
C. Click Apply to confirm changes.

Step 3. Synchronize with an NTP Server

A. Input the NTP server address into the NTP Server field (e.g., time.nist.gov).
B. Click Sync Now to synchronize the system clock with the NTP server immediately.
C. Ensure synchronization is successful before proceeding.

Step 4. Apply and Save Changes

A. Review the updated settings.
B. Click Apply to save all changes.

Network Setting

This section explains how to configure the System Network Settings, including TCP/IP and DNS settings. Proper network configuration ensures stable connectivity and reliable access to the system.

Procedures:

Step 1. Access the Network Setting Interface

A. Navigate to the Network Setting page in the system GUI.
B. Refer to Figure 89 for the interface layout.

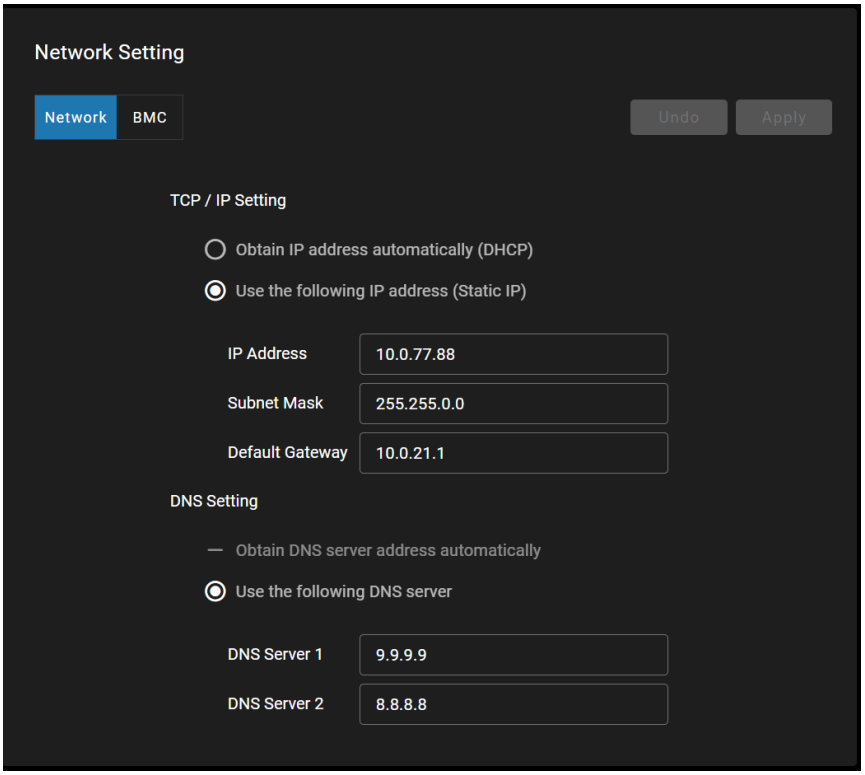


Figure 89. Configuring TCP/IP and DNS Settings in System Network Setting

Table 36 summarizes the key functionalities of the Network Setting interface:

Table 36. Key functionality of Network Setting Interface

1. TCP/IP Setting:	▪ Obtain an IP address automatically. ▪ Use a static IP address. (Requires IP address, subnet

	mask, and default gateway)
2. DNS Setting:	▪ Obtain DNS server address automatically. ▪ Use a custom DNS server. (Requires DNS server address)
3. Apply:	Apply the new settings.

Step 2. Select IP Configuration Method

Option 1: Obtain IP address automatically (DHCP): Select this option to use a dynamically assigned IP address from the network's DHCP server.

Important: DHCP-assigned IP addresses are dynamic and may change. If the IP address expires or changes, you must return to the Linux OS to retrieve the new IP address.

Option 2: Use the following IP address (Static IP):

a. Select this option to manually assign a fixed IP address for stable and consistent connectivity.
b. Input the following information:
    - IP Address: Enter the desired static IP address.
    - Subnet Mask: Input the appropriate subnet mask for your network.
    - Default Gateway: Provide the default gateway address for network communication.

Step 3. Select DNS Configuration Method

Option 1: Obtain DNS server address automatically: Select this option to allow the network to automatically configure DNS server addresses.

Option 2: Use the following DNS server:

a. Select this option to manually input DNS server addresses.
b. Enter the following details:
    - DNS Server 1: Provide the primary DNS server address.

- DNS Server 2: Provide the secondary DNS server address (optional).

Step 4. Save and Confirm Settings

A. Review all entered information to ensure accuracy.
B. Click Next to save the settings and proceed to the next configuration step.

Notes:

- DHCP Usage:
  - DHCP is ideal for temporary setups but may cause connectivity issues if the assigned IP address changes.
  - If using DHCP, be prepared to check the system for the new IP address when changes occur.
- Static IP Recommendation:
  - A static IP address is recommended for stable and long-term configurations, ensuring consistent GUI access.
  - Ensure the static IP address does not conflict with other devices on the network.
- Subnet and Gateway Validation:
  - Verify that the IP address, subnet mask, and default gateway belong to the same subnet.
  - Avoid leaving the gateway field blank to prevent connectivity issues.

BMC Setting

This section provides instructions for configuring the BMC (Baseboard Management Controller) Network Settings, enabling effective communication and remote management capabilities.

Accessing the BMC Network Setting Interface:

1. Navigate to the Network Setting page in the system GUI.
2. Switch to the BMC tab to access the BMC-specific settings (see Figure 90 below).

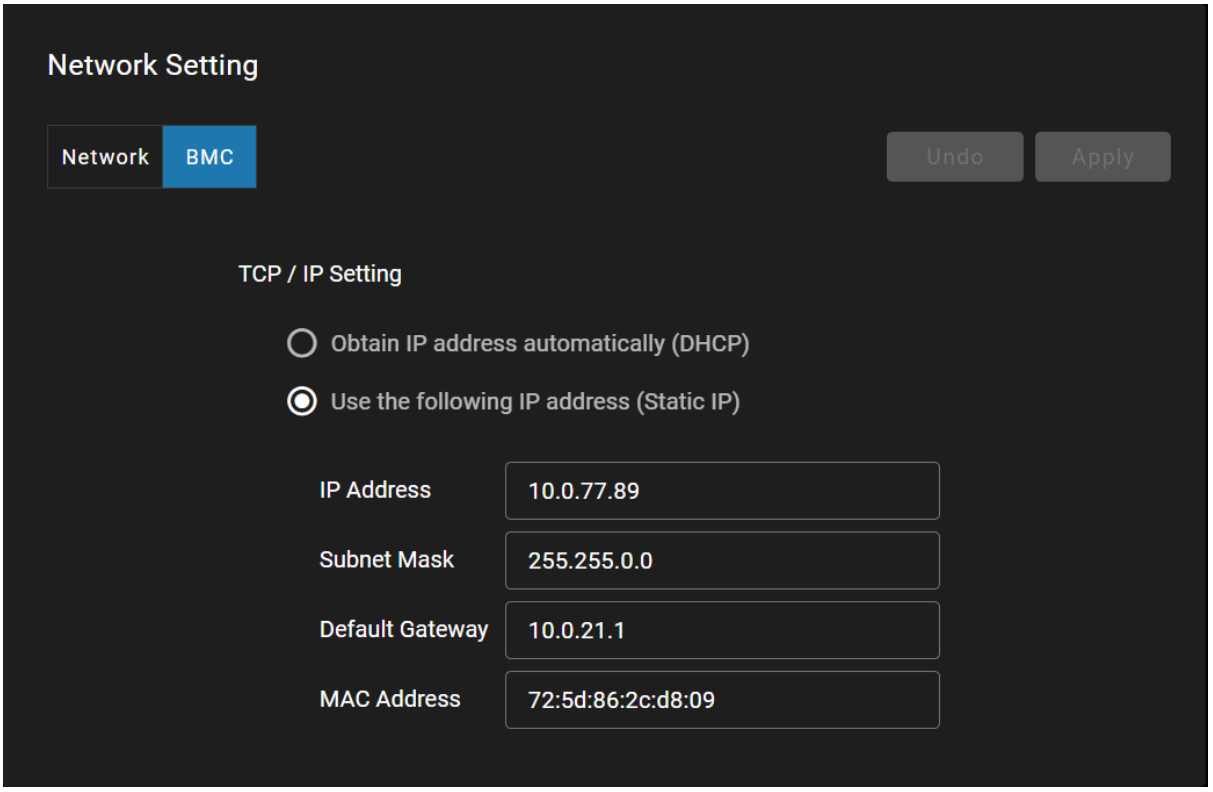


Figure 90. BMC Network Setting Interface

The BMC network configuration includes options for TCP/IP settings, allowing users to configure the IP address and subnet details.

Configuring BMC TCP/IP Settings

Procedures

Step 1. Select the IP Configuration Method:

Option 1: Obtain IP address automatically (DHCP): Select this option to allow the BMC to dynamically acquire an IP address from the network's DHCP server.

Important: DHCP-assigned IP addresses are dynamic and may change periodically. If the address changes, you must access the system locally to retrieve the new IP.

Option 2: Use the following IP address (Static IP):

a. Select this option to manually assign a fixed IP address for the BMC.
b. Input the following details:

- IP Address: Enter the desired static IP address for the BMC (e.g., 10.0.77.89).
- Subnet Mask: Provide the subnet mask for the network (e.g., 255.255.0.0).
- Default Gateway: Specify the gateway address for network communication (e.g., 10.0.21.1).
- MAC Address: This field is automatically populated with the BMC's hardware address (e.g., 72:5d:86:2c:d8:09).

Step 2. Saving and Confirming BMC Network Settings

A. Verify all entered information for accuracy and ensure it aligns with your network setup.
B. Click Apply to save the settings.
C. Perform a connectivity test to confirm that the BMC is accessible via the configured IP address.

Notes

- DHCP Usage:
  - Ideal for temporary or dynamic environments.
  - Be aware that changes to the DHCP-assigned IP address may disrupt remote access to the BMC.
- Static IP Recommendation:
  - A static IP address ensures consistent and reliable access to the BMC for management purposes.
  - Ensure the selected static IP does not conflict with other devices on the network.
- Subnet and Gateway Validation:
  - Verify that the IP address, subnet mask, and gateway belong to the same subnet.
  - Leaving the gateway field blank may result in connectivity issues.

User Management

The User Management feature allows administrators to manage system user accounts, including searching, editing, deleting, and creating accounts. Proper management ensures that users have appropriate access levels and system security is maintained.

Accessing the User Management Interface:

1. Navigate to the User Management page via the system GUI.
2. Refer to Figure 91 for an overview of the interface layout.

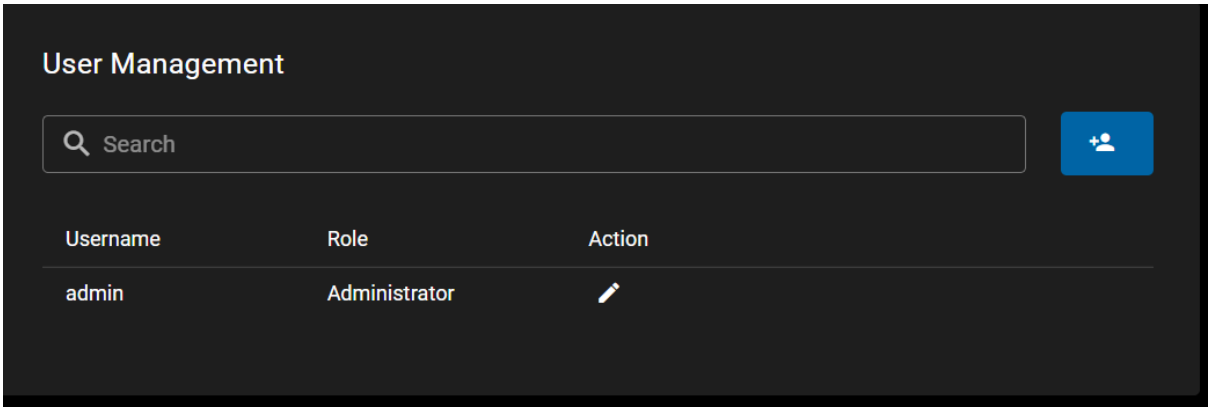


Figure 91. User Management Interface

Procedures

Step 1. Search for a Specific User

A. Use the Search Bar at the top of the page to locate a specific user account.
B. Enter the username or part of it to dynamically filter the user list.

Step 2. View and Manage User Accounts

The User Accounts Table displays existing user accounts with the following details:

- **Username**: The name of the user account.
- **Role**: The assigned role (e.g., Administrator, User, Guest).
- **Action**: Available actions for managing the account.
    - **Edit Account**:
        - Click the 🖍️ icon to edit account details such as the assigned role.
        - Save changes after editing.
    - **Delete Account**:
        - Click the 🗑️ icon to delete a user account permanently.
        - Confirm the deletion in the dialog box.

Note: The "admin" account cannot be deleted.

Step 3. Create a New User Account

A. Access the Create Account Interface
   - Click the blue Create Account button at the top-right of the page.
   - The Create Account dialog box will appear (see Figure 92).

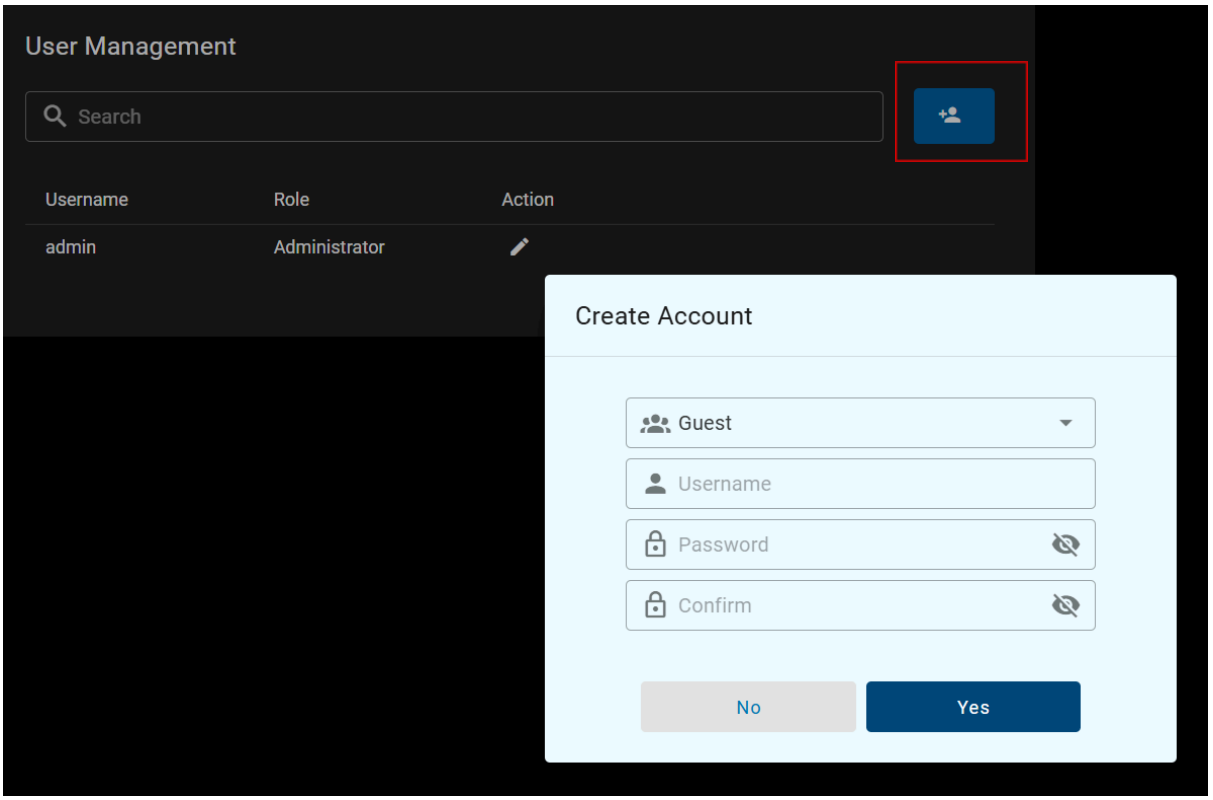


Figure 92. Create Account Dialog Box

B. Input User Information
   - Role: Select the role for the new user from the dropdown menu. Options include:
     - Guest: Limited access.
     - User: Standard access.
   - Username: Enter a unique username for the new user.
   - Password: Enter a secure password.
   - Confirm Password: Re-enter the password to ensure it matches.

C. Save or Cancel
   - Click Yes to save the new account.
   - Click No to cancel the operation.

User Roles and Permissions

Each role determines the access level and permissions granted to a user. Refer to Table 37 below for details:

Table 37. User Roles and Associated Permissions

Permission	Administrator	User	Guest
Read PCIe Resource	✓	✓	✓
Read Chassis info	✓	✓	✓
Read System Logs	✓	✓	
Manage PCIe Resource	✓		
User Account Management	✓		
Change Password	✓		
Read Maintenance info	✓		
Maintenance Operation	✓		
Read System Setting	✓		
Modify System Setting	✓		
Read Advanced Setting	✓		
Modify Advanced Setting	✓		

Notes

- The admin account is mandatory and cannot be deleted.
- When creating new accounts, ensure usernames are unique and roles are assigned according to user responsibilities.
- Passwords should meet security standards to maintain system integrity.

Change Web GUI Passwords

A. Change the password for Administrator
    a. Log in as the administrator. Update the password in the User Management page.
B. Change the password for User or Guest
    a. Log in as the administrator. Update the user/guest's password in the User Management page.

Reset mCPU/Web GUI to Default Settings

Long-press the UID button on the chassis for 10 seconds to reset the IP address and the administrator password to the default settings. Please note that the firmware will remain the latest version that was installed before the reset.

4.2.7 Advanced Setting

Mode Switch

The Mode Switch feature of the Falcon 5012 chassis, accessible through the System Management GUI's Advanced Settings, allows users to configure the system into Standard Mode or Advanced Mode. This functionality is critical for adapting system performance and resource allocation to meet workload requirements.

The system can be set to one of the following modes based on user needs:

- Standard Mode: Provides essential functionality, with fixed GPU allocation for general workloads. In this mode, each PCIe switch is assigned to a single host.
- Advanced Mode: Unlocks additional flexibility and high-performance capabilities. Advanced Mode enables dynamic GPU allocation across multiple hosts, making it ideal for demanding workloads. A Premium License is required to enable this mode.

Attention: Switching between Standard Mode and Advanced Mode must be synchronized across all PCIe switches to avoid operational conflicts. Mixed-mode configurations are not supported.

Procedures

Step 1: Accessing the Mode Switch Interface

A. Navigate to System Management GUI > Advanced Settings > Mode Switch (refer to Figure 93).

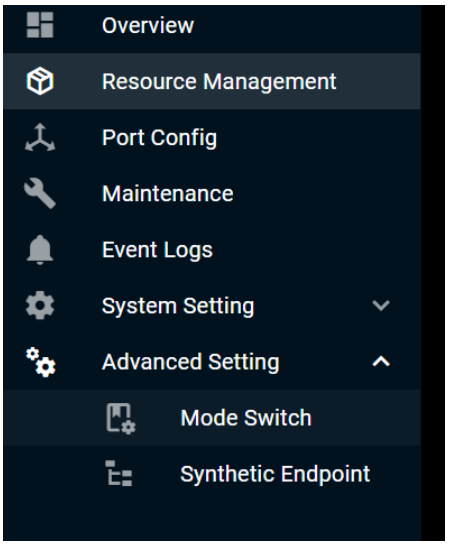


Figure 93. Mode Switch in System Management GUI

Step 2: Selecting and Applying the Desired Mode

A. Select the desired mode from the available options:
   - Standard Mode: For fixed GPU allocation.
   - Advanced Mode: For dynamic GPU allocation across multiple hosts.

B. Refer to Figure 94 for selecting the PCIe Mode.

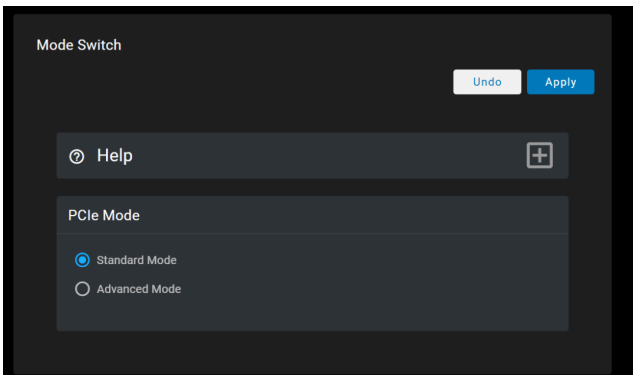


Figure 94. Selecting Standard or Advanced Mode

C. Click Apply to confirm the selection.
   - A confirmation dialog will appear, as shown in Figure 95.
   - Verify that all PCIe switches are synchronized and click Yes to proceed.

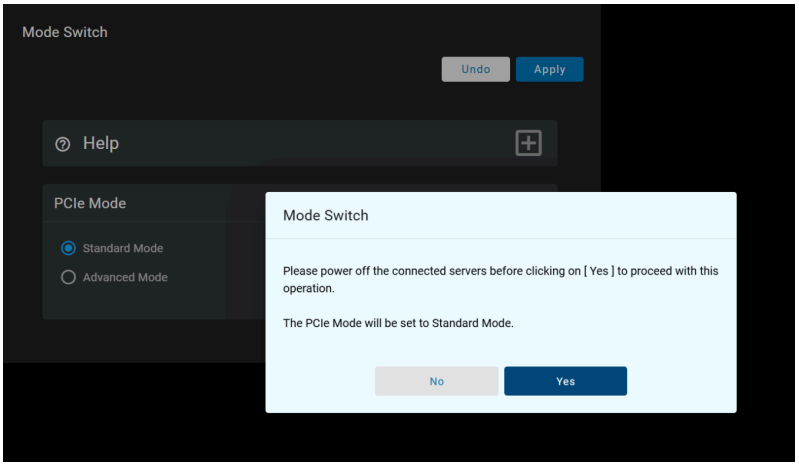


Figure 95. Confirmation Dialog

Step 3: Verifying the Mode Switch

A. After the system completes the configuration, a success message will appear, as shown in Figure 96.

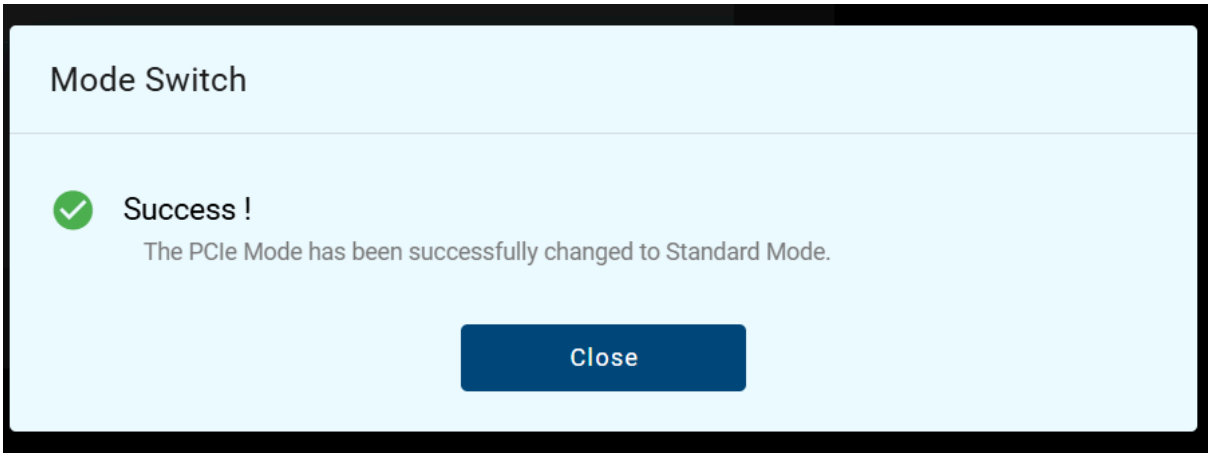


Figure 96. Success Confirmation

B. Return to the Resource Management page in the System Management GUI to verify the selected mode has been successfully applied.

Notes

- Advanced Mode Requirement: Ensure a valid Premium License is installed to enable Advanced Mode.
- Synchronization: All PCIe switches must be set to the same mode to avoid operational conflicts.
- Dynamic Resource Allocation: Advanced Mode offers greater flexibility by allowing users to dynamically assign or unassign GPUs to hosts.

Synthetic Endpoint Configuration

The Synthetic Endpoint feature in the Falcon GPU chassis provides the ability to pre-allocate PCIe MMIO resources at system boot, enabling efficient hot-plugging of PCIe devices without requiring host machine reboots. This feature is accessible in the System Management GUI under Advanced Config > Synthetic Endpoint.

When enabled, the Synthetic Endpoint reserves PCIe MMIO resources equal to the designated BAR size for devices (including placeholders) during the boot phase. This eliminates the need for manual PCIe resource adjustments or reboots after software-wise hot-plugging devices.

Key configurations include:

- Synthetic Endpoint Count: Sets the number of endpoints to reserve for device hot-plugging.
- Reserved PCI MMIO size: Defines the memory size reserved for each device, adjustable for both 32-bit and 64-bit systems.

Default Settings:

- Synthetic Endpoint Count: 6
- Reserved PCI MMIO size for 32-bit: 32 MB
- Reserved PCI MMIO size for 64-bit: 128 GB

Procedures

Step 1. Accessing the Synthetic Endpoint Page

A. Log in to the System Management GUI.
B. Navigate to Advanced Config > Synthetic Endpoint (refer to Figure 97).

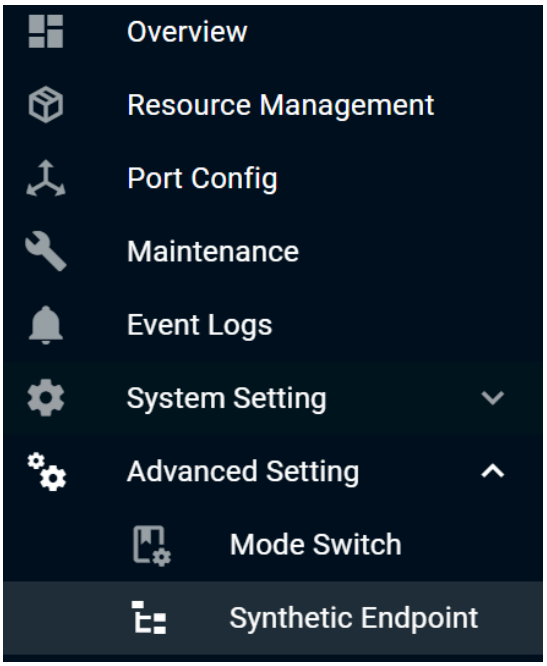


Figure 97. Accessing Synthetic Endpoint Configuration

Step 2. Configuring the Synthetic Endpoint

A. Enable Synthetic Endpoint: Select the Enabled option under the PCIe Switch you want to configure.
B. Set Synthetic Endpoint Count: Choose the desired endpoint count from the dropdown menu.
C. Configure PCI MMIO Sizes:
    - Adjust the reserved PCI MMIO sizes for 32-bit and 64-bit systems as needed.
    - Select sizes from the dropdown menus (refer to Figure 98).

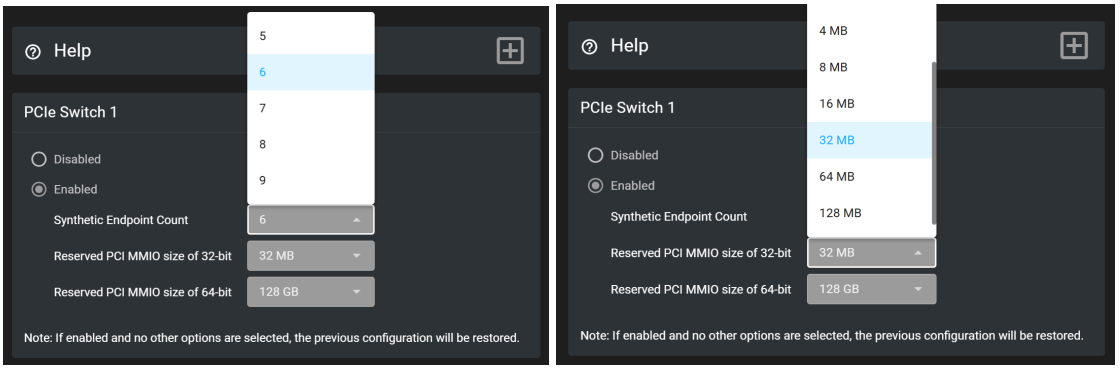


Figure 98. Configuring Endpoint Count and PCI MMIO Sizes

Step 3. Applying the Configuration

A. After selecting all options, click Apply in the top-right corner of the page.
B. Wait for the system to process the settings. A success message will confirm the changes.

Step 4. Verifying the Configuration

A. Return to the Synthetic Endpoint page to ensure the new settings are correctly displayed.
B. Optionally, verify the PCIe MMIO allocation in the system's resource management tools to confirm successful configuration.

Notes

- This feature only applies in Advanced Mode. Ensure the system is set to Advanced Mode before enabling synthetic endpoints.
- MMIO size must be equal to or larger than the memory size of the PCIe device to ensure functionality.
- If this feature is disabled, the system will revert to its previous configuration.

Chapter 5

BMC GUI Usage and Firmware Updates

The Baseboard Management Controller (BMC) GUI provides a comprehensive interface for system monitoring, remote management, and firmware updates. This chapter outlines key procedures, including logging into the BMC, remotely controlling power states, and updating firmware components such as BMC, CPLD, and retimer firmware.

For detailed guidance on BMC operations, refer to The Falcon 5012 BMC User Manual available on the H3 Platform website:

- Navigate to Product → GPU Chassis → Falcon 5012 → Resource
- Alternatively, navigate to Support → Knowledge Base → Download on the H3 Platform website.
    - Alternatively, visit the following URL directly: https://www.h3platform.com/knowledge-base/document
- Select the appropriate options:
    - Product Type: Composable GPU Chassis.
    - Model Type: Falcon 5012.
- Download the latest BMC user manual.

By following the instructions in this chapter, users can ensure secure access, efficient power management, and proper firmware maintenance for optimal system performance.

5.1 BMC GUI Log-in

To access the BMC GUI, follow these steps:

1. Open a supported web browser.
2. Enter the BMC IP address in the URL bar. (Refer to the previous section 3.3 Retrieve GUI IP for System Boot-Up for instructions on retrieving the BMC IP address.)
3. When prompted, log in using the default login credentials (see Figure 99).

Default Login Credentials:

- **Username:** admin
- **Password:** 12345678

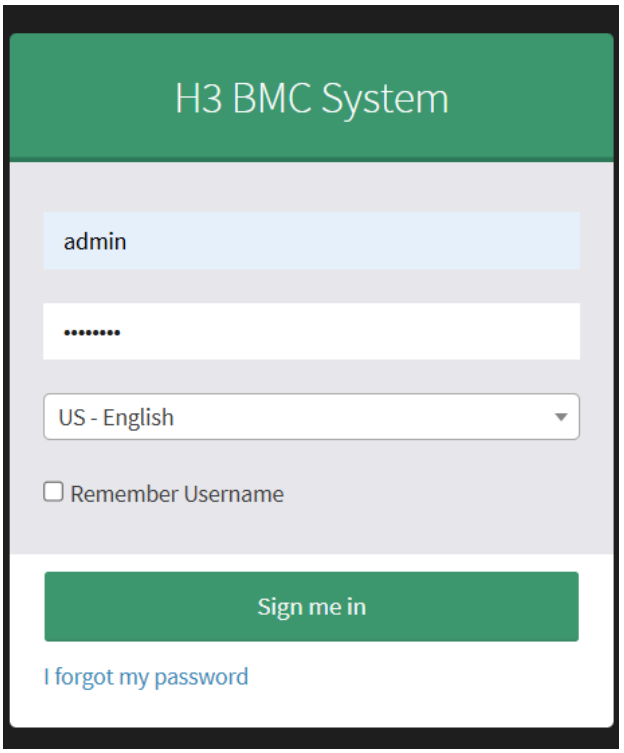


Figure 99. BMC System Log-in page

For security purposes, it is recommended to change the default password after the first login.

Note: If the above password does not work, try the alternative password: admin.

Change BMC GUI Password:

To change the BMC GUI password, you can follow one of the methods below:

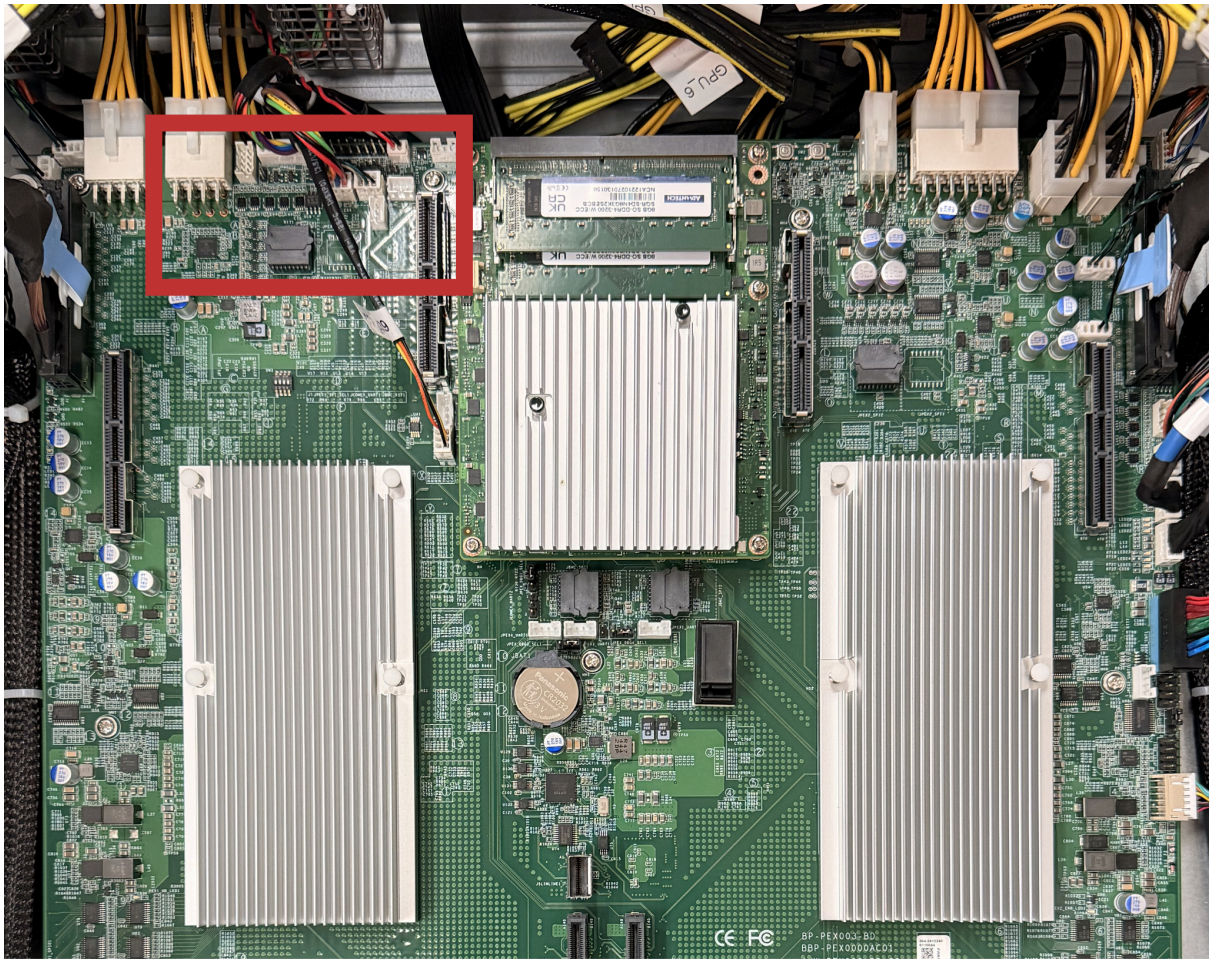
A. Reset on the BMC GUI interface
   a. Click “I forgot my password” on the BMC GUI. A new password will be mailed to the email address previously configured in the BMC GUI “Settings” → “User Management”.

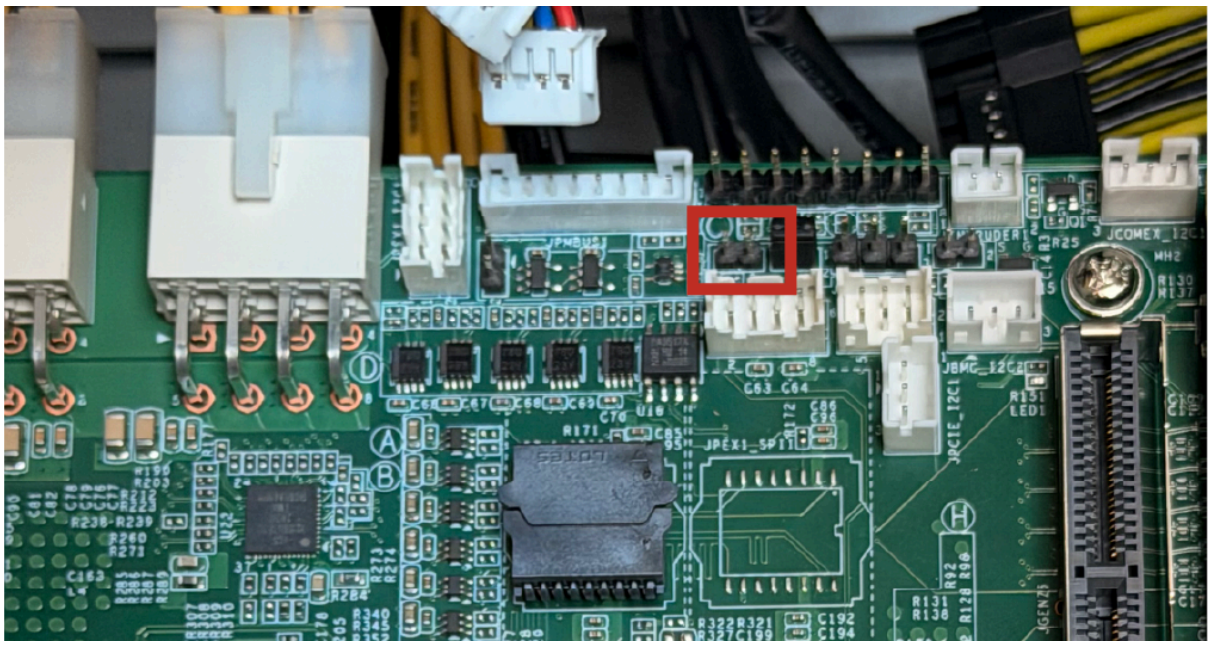
B. Reset on the hardware

   Step 1. Power off the AC. Install a jumper cap across the pin J1 (marked in the image below).

   Step 2. Power on AC.

   This action resets all BMC configurations including the passwords and logs.





Reset BMC to Default Settings

Please refer to page 77 (title: Restore Factory Defaults) of the BMC user manual. URL: https://www.h3platform.com/knowledge-base/document

---

5.2 Remote Powering On/Off Using BMC

This procedure provides instructions for using the BMC (Baseboard Management Controller) GUI to remotely manage the power states of the system. By accessing the Remote Control section and launching the H5Viewer, users can utilize the Power menu to execute actions such as resetting, shutting down, or powering on the server.

About this Task

To remotely manage the system's power state, you will:

1. Access the BMC GUI and navigate to the Remote Control section.
2. Launch the H5Viewer to open the KVM interface.
3. Use the Power menu to execute desired actions.

Procedures

Step 1. Navigate to Remote Control

A. After logging in, locate the Remote Control tab in the left navigation bar (refer to Figure 100).

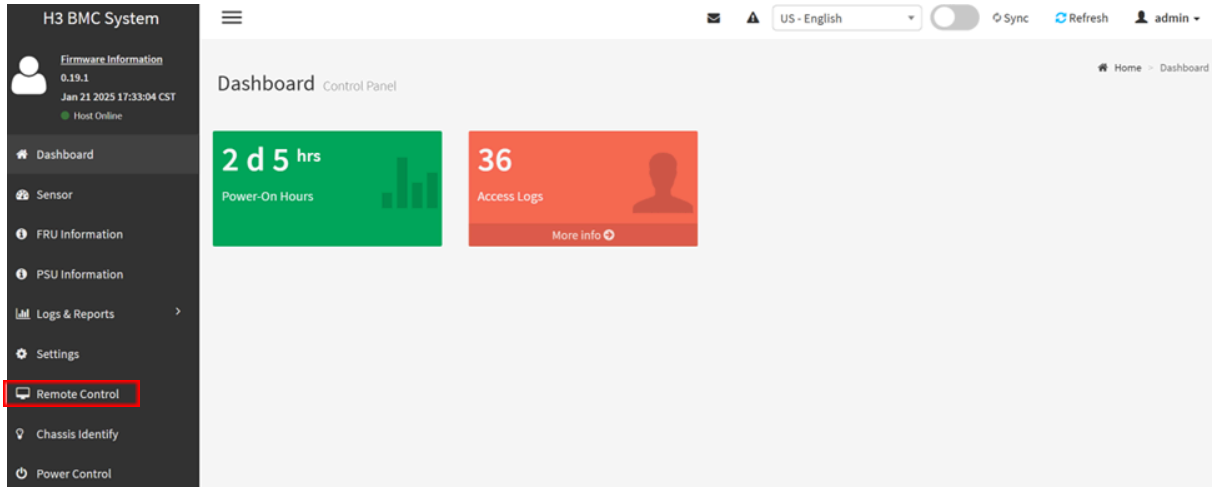


Figure 100. Locating the Remote Control Tab in the BMC GUI Navigation Bar

B. Click on Launch H5Viewer to access the KVM interface (refer to Figure 101).

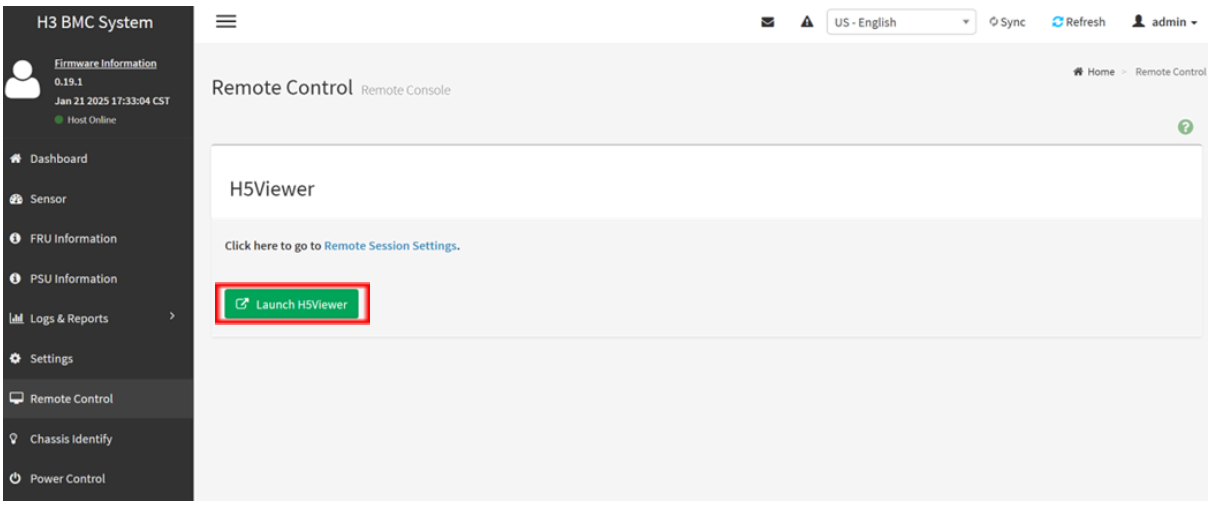


Figure 101. Launching H5Vi1ewer from the Remote Control Tab

Step 2. Use the Power Menu

A. In the KVM interface, locate the Power menu in the top navigation bar (refer to Figure 102).

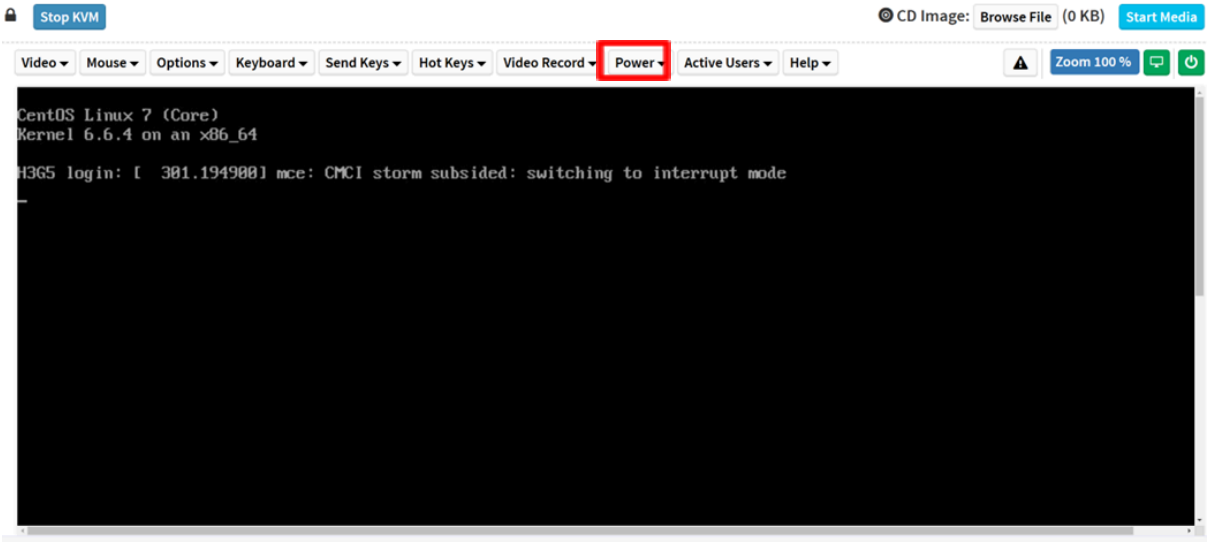


Figure 102. Locating the Power Menu in the KVM Interface Navigation Bar

B. From the Power menu (refer to Figure 103), select the desired action:

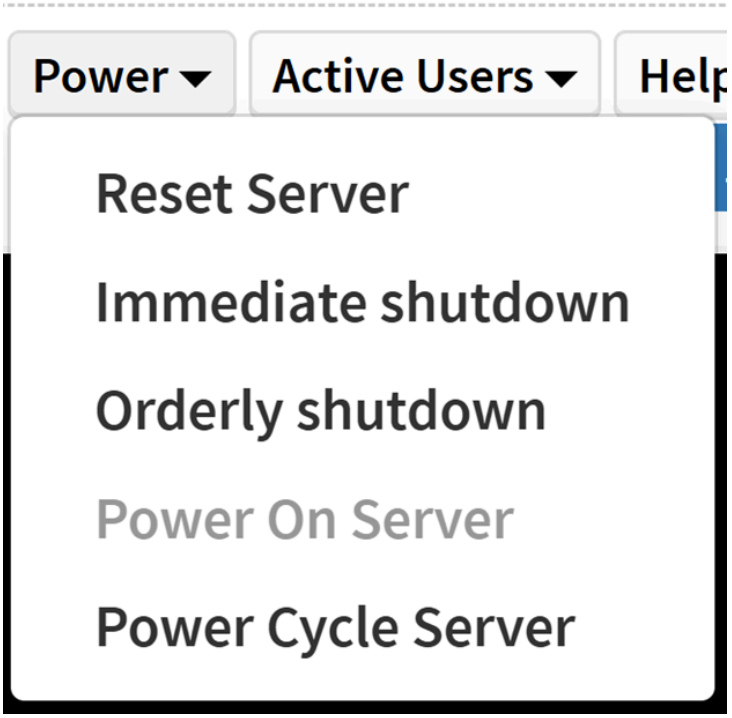


Figure 103. Power Menu Options in the KVM Interface

- Reset Server: Reboot the system.
- Immediate Shutdown: Instantly powers off the system.
- Orderly Shutdown: Initiates a graceful system shutdown.
- Power On Server: Turns the server on remotely.
- Power Cycle Server: Restarts the system by cutting and restoring power.

C. Confirm your selection to execute the chosen power action.

Results

- The selected power action is executed successfully.
- The system reflects the updated power state.

Notes

- Ensure proper network connectivity between the BMC and your management machine.
- Use Immediate Shutdown only when necessary, as it may not allow the system to save its state.
- For routine operations, prefer Orderly Shutdown for a graceful power-off process.

5.3 Updating BMC Firmware

The BMC firmware update is an essential maintenance task designed to ensure optimal system performance, enhanced security, and the availability of new features for the Falcon 5012 GPU chassis. This process involves updating the Baseboard Management Controller (BMC), which serves as the core for system management functions such as hardware monitoring and remote administration.

Users are advised to refer to the Falcon 5012 BMC User Manual for a comprehensive step-by-step guide. The manual provides detailed instructions on firmware preparation, upload, validation, and finalization procedures. Follow these guidelines to ensure a successful update and to minimize system downtime.

About the Task

The BMC firmware update requires the following key considerations:

- Purpose: To enhance system functionality by addressing known firmware issues, improving compatibility, and incorporating the latest security patches.
- Preconditions: Ensure the system is stable, critical data is backed up, and all necessary update files are available.
- Key Steps:
    1. Access the Falcon 5012 BMC interface via a secure connection.
    2. Upload the firmware package, which must be obtained from a verified source.
    3. Follow the firmware validation and installation instructions provided in the manual.
    4. Restart and verify the BMC firmware version after the update process is complete.

Reference Guide

For a detailed walkthrough of the BMC firmware update process, refer to the Falcon 5012 BMC User Manual:

- Path:
  Navigate to the H3 Platform website using the following steps:
  Product → GPU Chassis → Falcon 5012 → Resource → Falcon 5012 → BMC User Manual.
- Direct URL:
  https://drive.google.com/file/d/1y47HJo2CnhJFhwqODxwvuWUHyaiDresN/view

Ensure to follow the safety and preparatory instructions outlined in the manual.
For additional assistance, contact H3 Platform Support at support@h3platform.com.

5.4 Updating CPLD Firmware Using BMC

This procedure provides step-by-step instructions to update the CPLD Firmware using the BMC interface. The process includes selecting the firmware file, initiating the update, and completing an AC power cycle.

About this Task

To update the CPLD firmware, you will:

1. Access the CPLD Firmware Utility from the Maintenance section of the BMC GUI.
2. Select the appropriate CPLD firmware file for the update.
3. Execute the firmware update and complete the process with a power cycle.

Tools and Materials

You will need:

- The CPLD firmware file (.jed format) stored on your computer.
- Network access to the BMC.
- Supported web browser (e.g., Firefox or Chrome).

Procedures

Step 1. Access CPLD Firmware Utility

A. Open a supported browser and enter the BMC IP address in the URL bar.
B. Log in to the BMC GUI using your credentials:
    - Default Username: admin
    - Default Password: 12345678

Note: If one password does not work, try the alternative password provided.

C. From the navigation bar on the left, select Maintenance (refer to Figure 104).

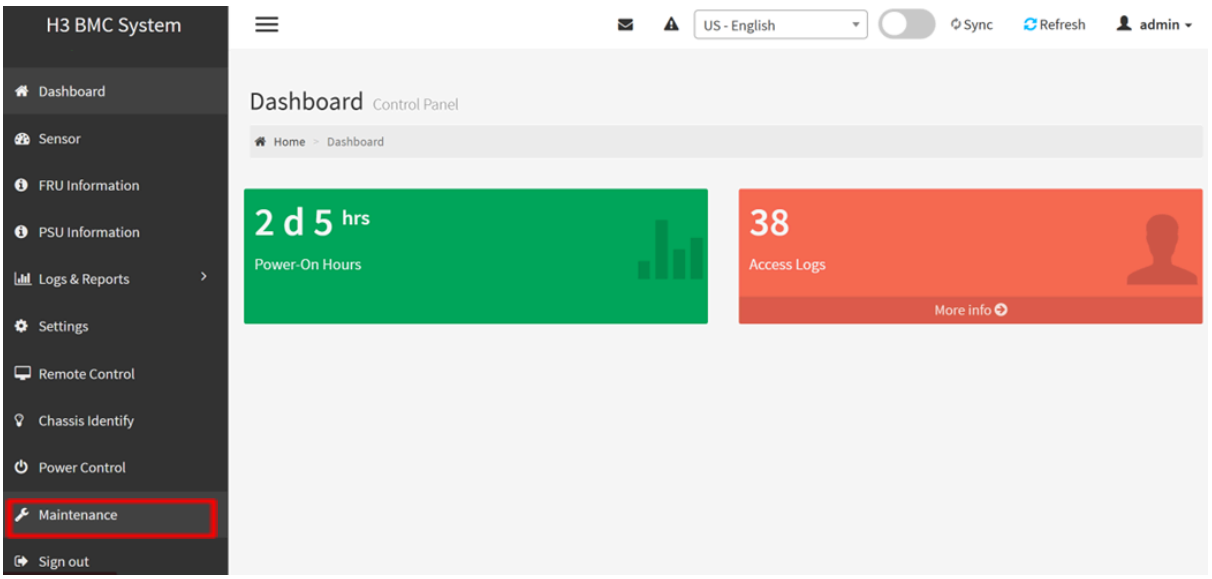


Figure 104. Navigating to the **Maintenance** section.

Step 2. Select the CPLD Firmware Utility

A. Under the Maintenance section, click on CPLD Firmware Utility (refer to Figure 105).

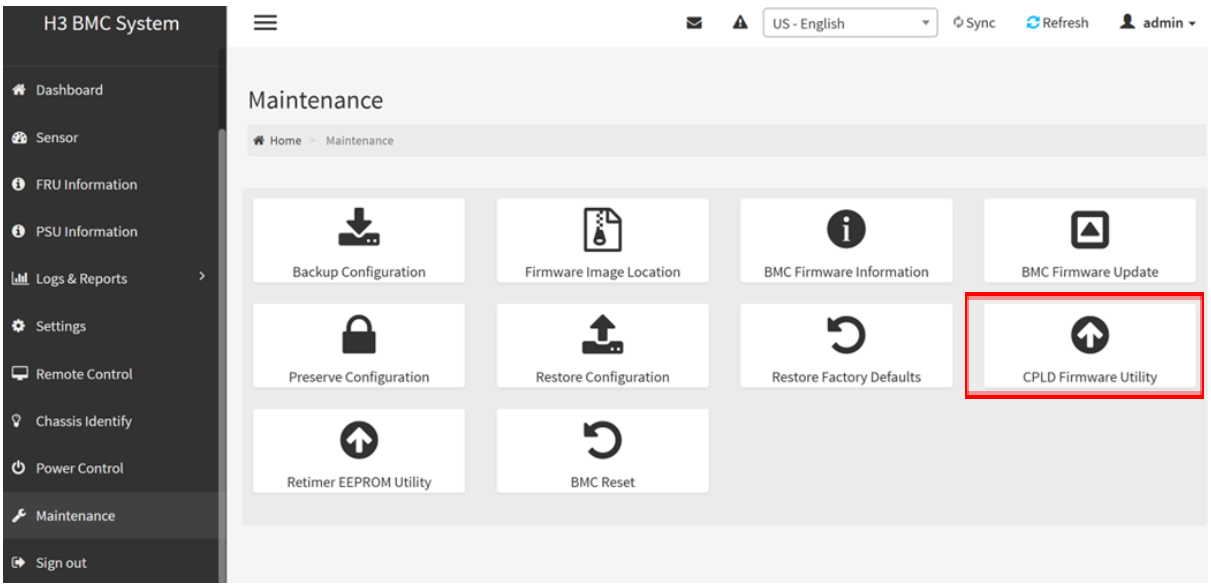


Figure 105. Accessing the CPLD Firmware Utility page.

B. On the CPLD Firmware Utility page, review the current CPLD IDs, Bus, and Firmware Version.

Step 3. Perform a Remote System Shutdown Using BMC

A. Refer to the Remote Power On/Off Using BMC section for detailed instructions.
B. Access the Power Control option under the Remote Control section in the BMC interface.
C. Select Immediate Shut Down to safely shut down the system remotely.
D. The screen will suddenly black out, indicating that the system has successfully powered off.

Step 4. Update CPLD Firmware

A. Click the Select CPLD firmware file button and upload the firmware file (.jed format) from your computer.
B. Click the Start Update button to initiate the update process (refer to Figure 106)

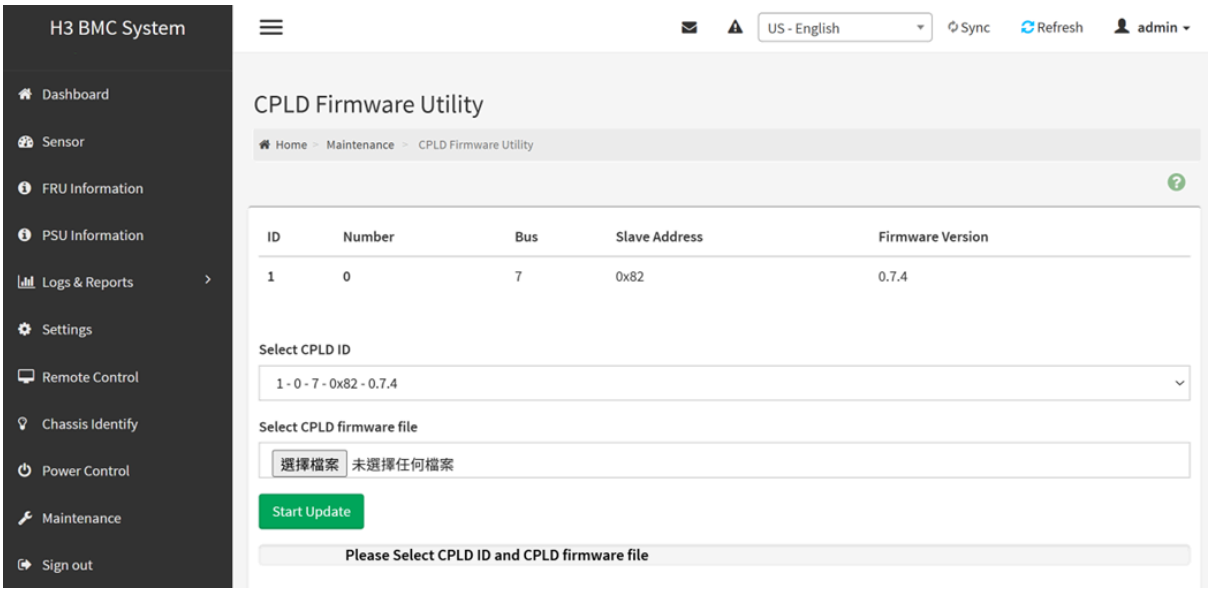


Figure 106. Selecting and starting the CPLD Firmware update.

C. During the update, a Processing message will appear, indicating the progress percentage (refer to Figure 107).

Figure 107. Processing During CPLD Firmware Update

Step 5. Confirm Update Completion

A. After the update, a notification will appear: Update Completed (Please unplug and replug the power cable) (refer to Figure 108).

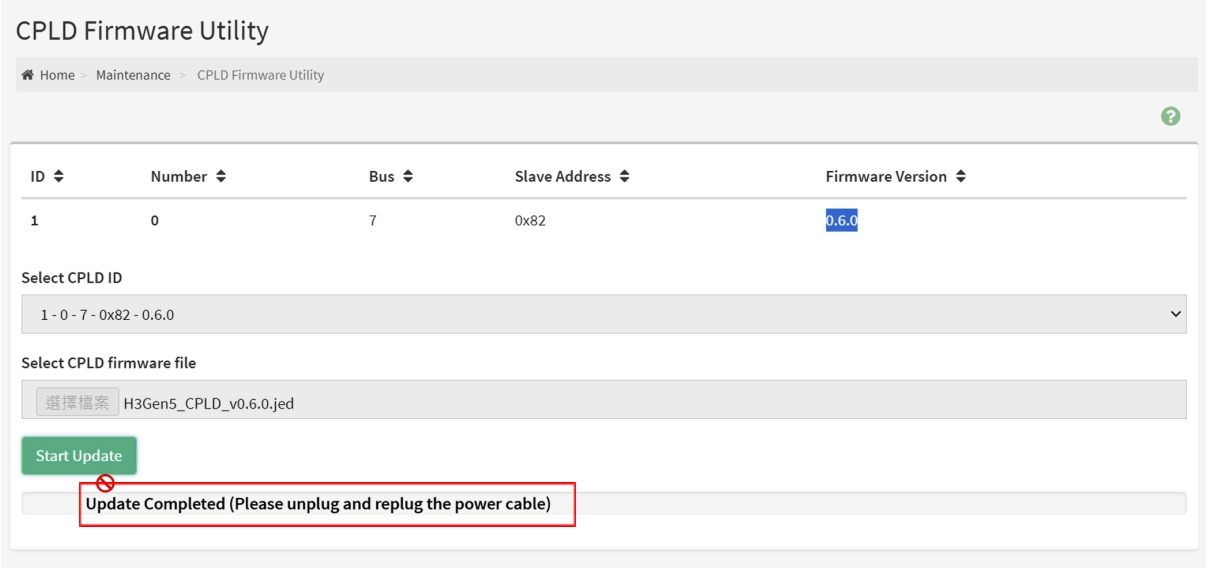


Figure 108. Firmware update completion notification.

B. Disconnect the power cable to perform an AC power cycle.
C. Access the BMC interface to verify whether the CPLD version has been successfully updated.

Notes

- Ensure the correct CPLD firmware file is selected to avoid compatibility issues.
- Performing the AC power cycle is mandatory for the firmware update to take effect.

5.5 Updating Host Adapter (Retimer) Firmware

This section provides instructions for updating the firmware on the Host Adapter (Retimer) installed in Falcon 5012that supports PCIe expansion. The update procedure uses the BMC Retimer EEPROM Utility. Host Adapter (Retimer) modules C01 and D01 are supported.

About this task

This procedure describes how to update the firmware on the Host Adapter (Retimer) using the BMC Retimer EEPROM Utility. This utility provides an interface to upload firmware images, select individual Host Adapter (Retimer), and perform controlled updates. Follow all instructions precisely to avoid incorrect flashing.

To update the firmware:

- Identify the Host Adapter (Retimer) module type.
    -
- Power on the system and log in to the BMC Web Interface.
- Download the correct firmware version from the H3 Platform website.
- Use the Retimer EEPROM Utility to upload and apply the firmware image.
- Verify the version after a full power cycle.

Procedures

Step 1. Preparing for Host Adapter (Retimer) Firmware Update

A. Power off the server system.
B. Host adapters must be installed into Falcon 5012 on the Slots 12~15, and 21~24 (except the Slots 11 and 25) to perform firmware updates.

    - C01 DIP switch settings must be configured as follows (see Figure 109):
        -
        - SW1: ON (Right)
        - SW2: HLL (Right, Left, Left)

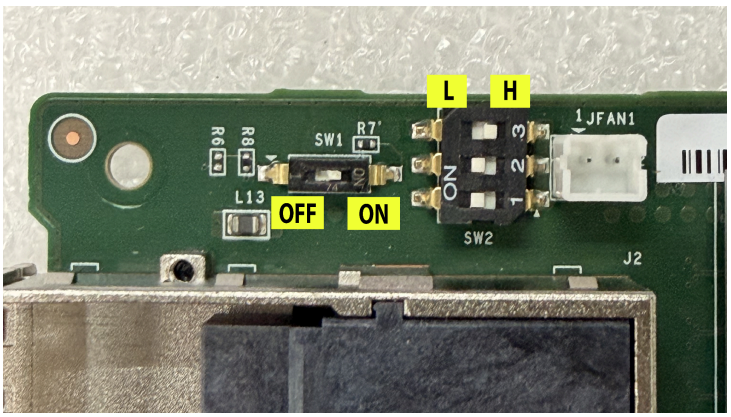


Figure 109. Host Adapter (C01) DIP Switch Settings for Firmware Update at Falcon side

- D01 DIP switch settings must be configured as follows (see Figure 110):
  -
  - SW1: ON, ON, ON (Left, Left, Left)
  - SW2: HLH (Right, Left, Right)

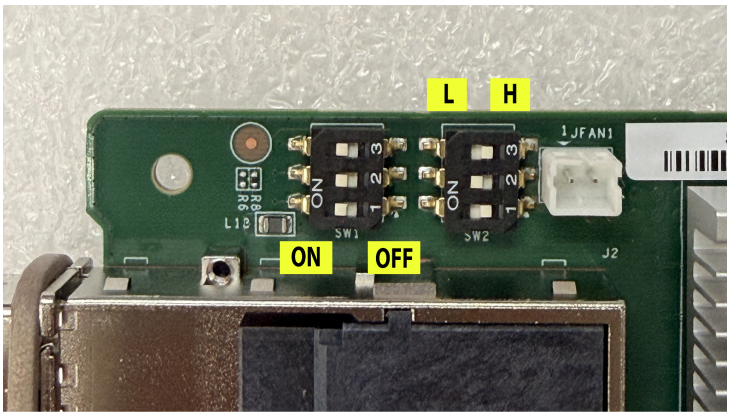


Figure 110. Host Adapter (D01) DIP Switch Settings for Firmware Update at Falcon side

---

Step 2. Accessing the BMC Interface

A. Power on the system.
B. Log in to the BMC Web Interface.
C. Navigate to the Maintenance tab (see Figure 111).
D. Select the Retimer EEPROM Utility option (see Figure 112).

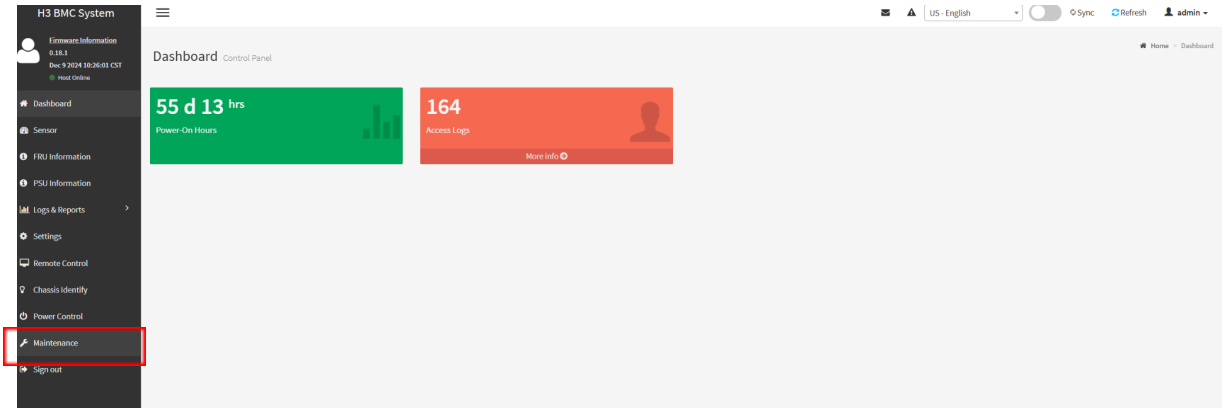


Figure 111. Navigating to the **Maintenance** section.

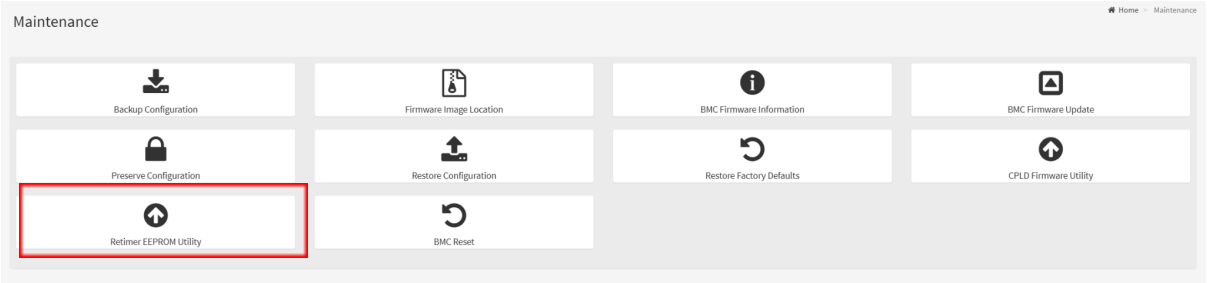


Figure 112. Navigating to the Retimer EEPROM Utility section.

Step 3. Downloading the Host Adapter (Retimer) Firmware File

A. Go to the H3 Platform support site: https://www.h3platform.com/knowledge-base/document
B. Navigate to: **Support → Knowledge Base → Download.**
C. Select:
    - **Product Type:** Composable GPU Chassis
    - **Model Type:** Falcon 5012

D. Download the latest Host Adapter (Retimer) firmware file.

Step 4. Performing the Update

The Retimer EEPROM Utility lists all Host Adapter (Retimer) modules detected in the system (see Figure 113). Displayed details include:

- Slot ID: Logical ID mapped to the physical chassis or host slot.

- Link Width: PCIe channel configuration (e.g., x1, x4).
- Bus Address / Device ID: Indicates I2C control interface.
- EEPROM Version: Displays current firmware revision.

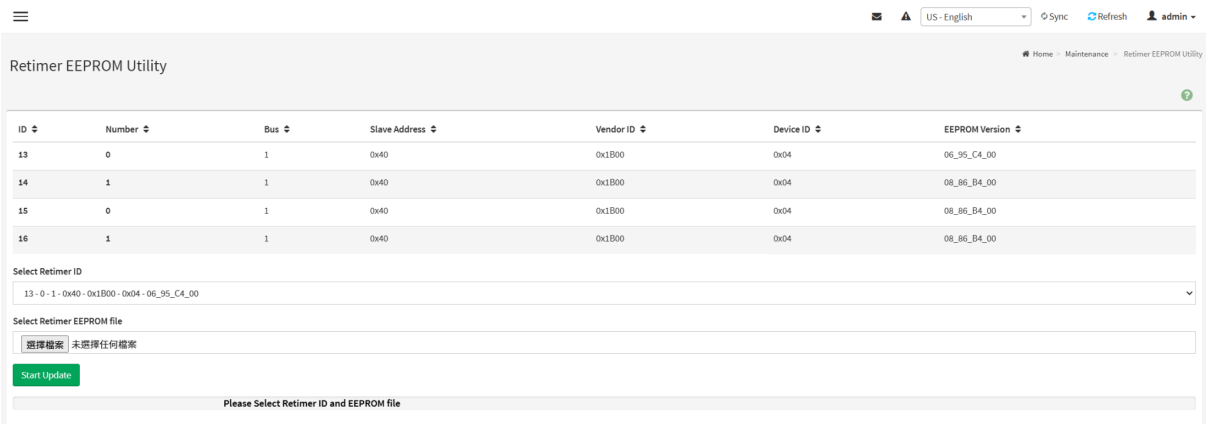
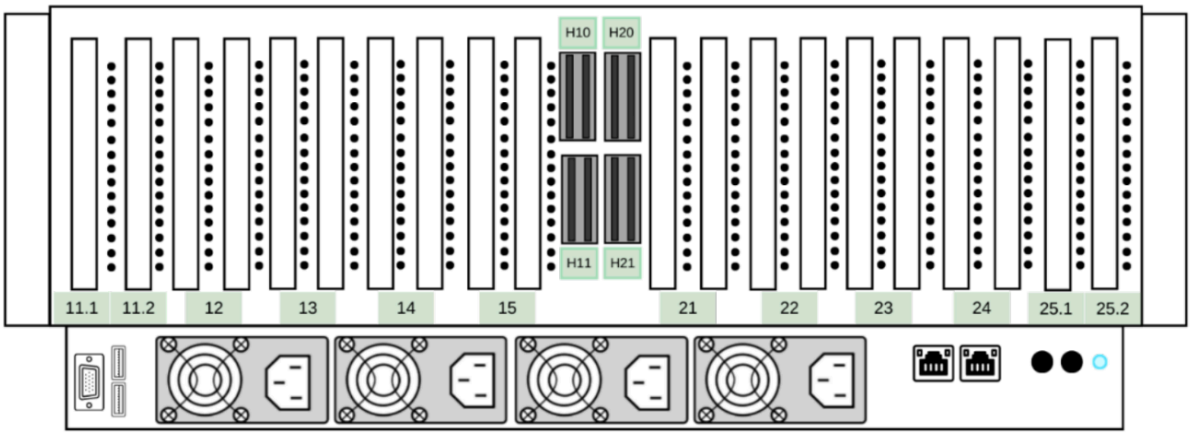


Figure 113**.** Retimer EEPROM Utility Interface

The table in Figure 114 is a reference for host ports and slot IDs . Please use the Slots 12~15, and 21~24 (except the Slots 11 and 25)



Slot	ID
H11	13
H10	14
H21	15
H20	16

Figure 114. Physical slot to Slot ID mapping for chassis modules

To begin the update:

1. Select the correct Slot ID. Verify it corresponds to the installed Host Adapter (Retimer) to prevent misflashing.
2. Click Select Retimer EEPROM File and choose the firmware file.
3. Click Start Update.
4. Monitor the status via the progress bar (see Figure 115).

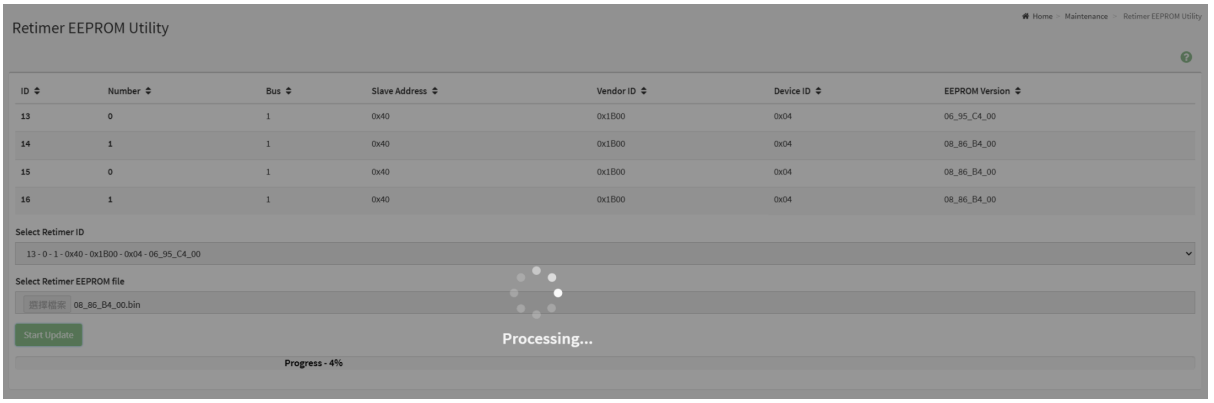


Figure 115. Firmware update progress status

Step 5. Verifying the Update

A. The firmware version may not immediately appear in the EEPROM Utility after update.
B. Perform a full power cycle by unplugging and reconnecting the system's power cord.
C. Reconnect to the BMC Web Interface to verify the updated version.
D. If the updated version is still not shown, contact H3 Technical Support for assistance.

Chapter 6

Part Replacement

This chapter provides comprehensive guidelines for identifying and replacing faulty components in the Falcon 5012 GPU Solution. It focuses on ensuring system reliability, minimizing downtime, and maintaining warranty compliance during part replacement.

To ensure compatibility and avoid voiding the warranty, it is strongly recommended to order replacement parts directly from H3 Platform. For further details, visit the H3 Platform website.

The chapter includes step-by-step instructions for identifying defective components through the system GUI and performing replacements for the following key parts:

- Fans
- Power Supply Units (PSUs)

These procedures are designed to assist users in maintaining optimal system performance and ensuring proper functionality after component replacement.

6.1 Fans

The fans in the Falcon 5012 system are conditioned hot-pluggable and located on the front panel of the chassis. They can be replaced directly, ensuring minimal downtime during maintenance.

Identifying a Faulty Fan:

1. Use the system GUI to determine which fan requires replacement.
2. In the Overview Sensor Readings panel, a Red status indicates critical or non-recoverable conditions, signaling the need for replacement.

Replacement Procedures:

1. Set Fan Speed to 30%:
    - Adjust fan speed (PWM duty-cycle) to 30% in the settings by BMC web-GUI or by using the IPMI command: #ipmitool raw 0x36 0xFA 0x00 0x20 0x20 0x20 0x20 0x20 0x20 0x20 0x20
2. Locate the Faulty Fan:
    - Identify the faulty fan using the numbering system shown in the figure below. (Refer to Figure 116: Picture of the Fans from the Front View)

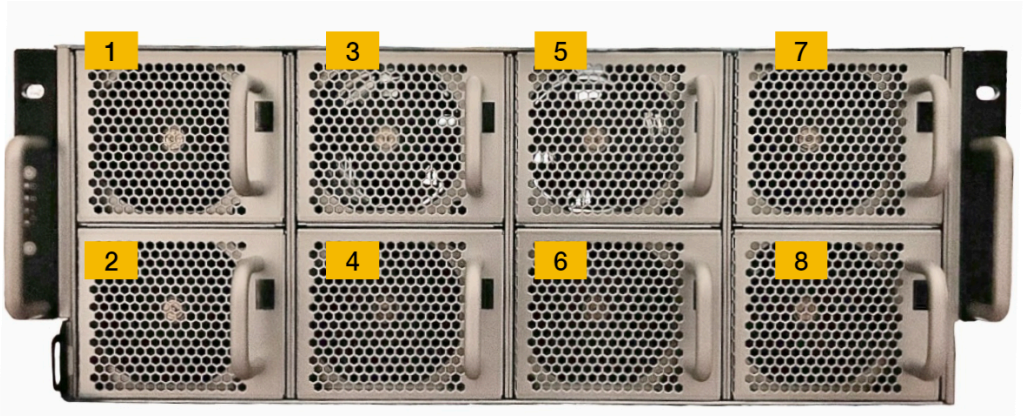


Figure 116. Picture of the Fans from the Front View

3. Remove the Faulty Fan:
    - Use the fan's built-in handle to pull it out of the slot.
4. Install a New Fan:
    - Align the new fan module with the slot and slide it in until it clicks securely into place.
5. Restore Fan Speed to Automatic Control:
    - Restore fan speed to automatic control in the settings by BMC web-GUI or by using the IPMI command: #ipmitool raw 0x36 0xF9
6. Verify the Replacement Effect:
    - Use the GUI to confirm that the new fan is operational and functioning correctly.

Attention:

- Use only compatible fans provided by H3 Platform to prevent damages and maintain warranty coverage.
- Step 1 is critical for a controlled hot-swap procedure. It ensures that the airflow within the chassis does not create reverse pressure that could prevent the newly inserted fan from spinning up properly.

6.2 Power Supply Unit (PSU)

The Power Supply Unit (PSU) in the Falcon 5012 system is designed for modular hot-swap replacement, allowing users to replace faulty PSUs without shutting down the system in most cases.

Identifying a Faulty PSU:

1. Use the system GUI to identify the faulty PSU.
2. In the Overview Sensor Readings panel, a Red status indicates a PSU failure, such as critical or non-recoverable conditions.

Replacement Procedures

1. Identify the Faulty PSU:
    - Use the system GUI to determine which PSU requires replacement.
    - Refer to the PSU numbering system in the back view diagram for precise identification.
      (Refer to Figure 117: Picture of the PSUs from the rear View)

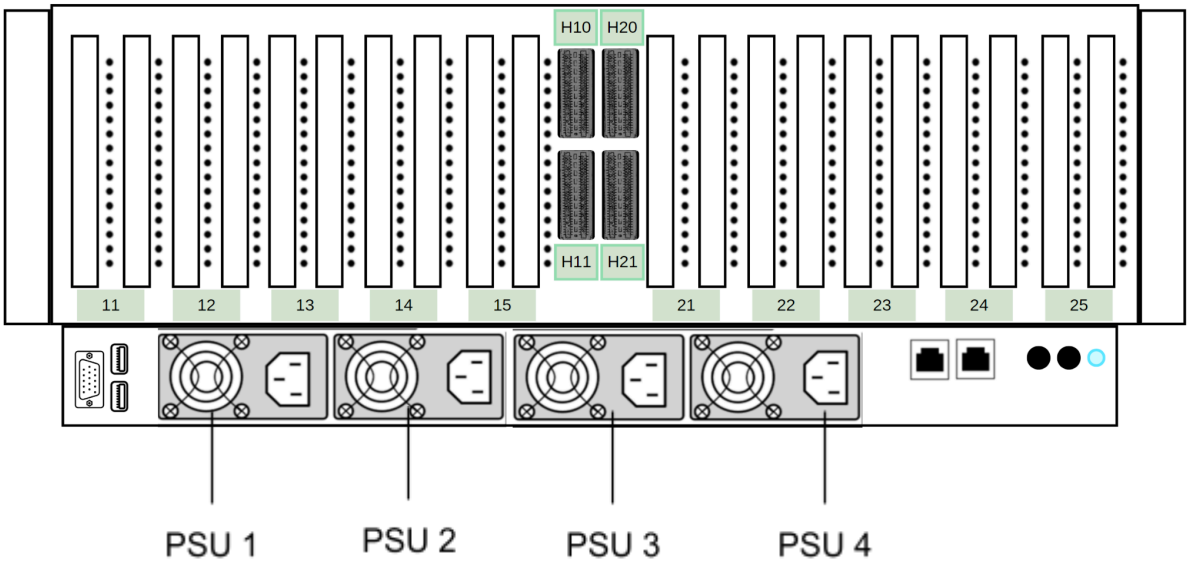


Figure 117. Picture of the PSUs from the rear view

2. Evaluate System Status:

- If only one PSU is faulty but the system remains operational, replacement can be scheduled at a convenient time. There is no immediate need to contact technical support.
- If multiple PSU failures occur, rendering the system non-operational, contact H3 Platform Technical Support for immediate assistance.

3. Unlock the PSU:
   - Locate the PSU slot on the rear panel.
   - Lift the handle of the PSU to expose the release button.
   - Press the release button while gently pulling the PSU outward.
     (Refer to Figure 118: Picture of the PSU Unlock Process)

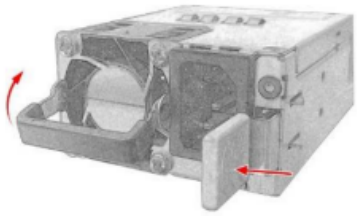


Figure 118. Picture of the PSU unlock process

4. Remove the Faulty PSU:
   - Use the handle to completely slide the PSU out of its slot.
   - Ensure the PSU is fully removed before proceeding.
5. Install the Replacement PSU:
   - Align the replacement PSU with the empty slot.
   - Slide the PSU into the slot until it locks into place with a secure click.
6. Verify the Replacement:
   - Use the GUI to confirm that the new PSU is recognized and functioning properly.
   - Ensure the system displays a stable power status with no errors.

Attention:
- Please refer to the PSU numbering system to ensure the correct unit is replaced.
- Use only H3 Platform-approved PSUs to guarantee compatibility and maintain warranty coverage.
- Contact H3 Platform Technical Support only if multiple PSU failures render the system non-operational. A single PSU failure usually does not affect normal system operations.

Chapter 7
Supplementary Notes

7.1 Host Expansion with Chassis-Side Retimer Installation

This chapter provides information on chassis-side installation of Retimer modules to support host expansion scenarios. This configuration enables multiple hosts to share GPU resources housed within a single chassis. Proper installation, including DIP switch settings and slot allocation, is essential to ensure signal integrity and system stability.

For complete operation, a matched set of Retimers is required on both the chassis side and the host side. If additional Retimers are needed, contact H3 Technical Support. For details on host-side installation, see Section 3.6.

About this task

The installation process involves mounting the Retimer into the specified expansion slots, setting DIP switches based on module type (C01 or D01), and securely connecting power and CDFP data cables. Before powering on the system, ensure all switches are correctly configured. Failure to do so may result in communication issues or hardware damage.

Prerequisites: Chassis-Side DIP Switch Settings

C01 Module:

- SW1 Configuration:
  - SW1 = OFF
  - See Figure 119 to confirm switch orientation

- SW2 Configuration:
  - A2 = H
  - A1 = L
  - A0 = L
  - This sets the I2C address to 0xE8 (HLL)
  - This is the default setting and typically does not require changes

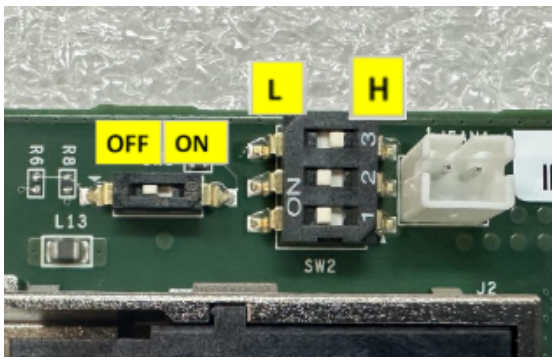


Figure 119. DIP switch configuration for C01 module (chassis side)

D01 Module:

- SW1 Configuration:
  - PERST switch = OFF – Blocks PERST signal from JBOX to host
  - I2C_SDA switch = ON – Allows I2C_SDA from JBOX to the gold finger
  - I2C_SCL switch = ON – Allows I2C_SCL from JBOX to the gold finger

- SW2 Configuration:
  - A2 = H
  - A1 = L
  - A0 = H

  - This sets the I2C address to 0xEA (HLH)
  - This is the default setting and typically does not require changes

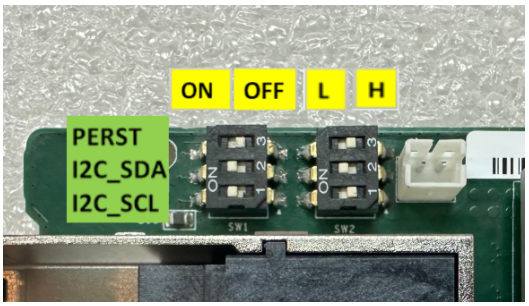


Figure 120. DIP switch configuration for D01 module (chassis side)

Note: Be sure to complete all DIP switch settings before powering on the system to avoid communication issues.

Installation Procedures

Step 1. Install the Retimer

Insert the Retimer into the designated chassis-side expansion cable slot:

- 10-GPU Configuration:
  - Use Slot 11 and Slot 25
  - *See Figure 121*

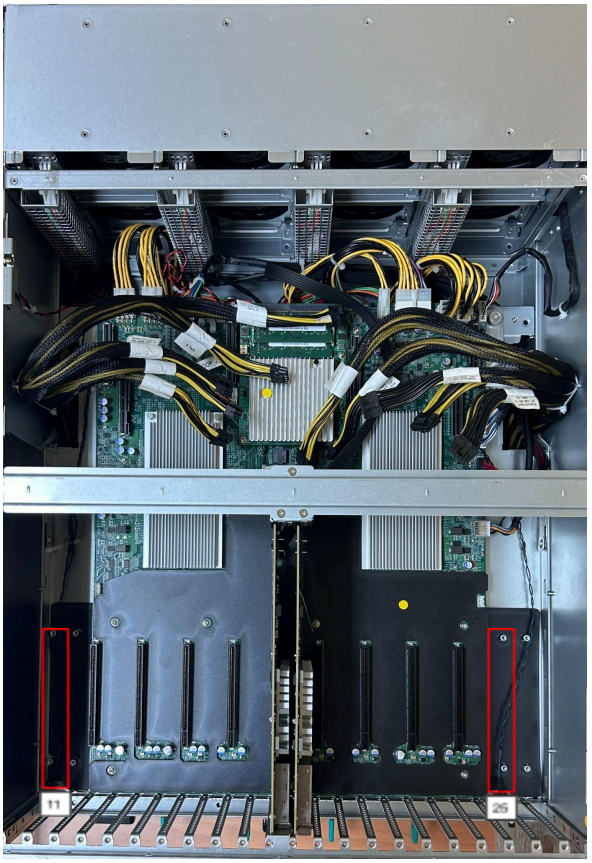


Figure 121. Slot 11 and Slot 25 Designated as expansion cables for Retimer Installation (10-GPU)

- 8-GPU Configuration:
  - Use Slot 11.1, 11.2, 25.1, and 25.2
  - *See Figure 122*

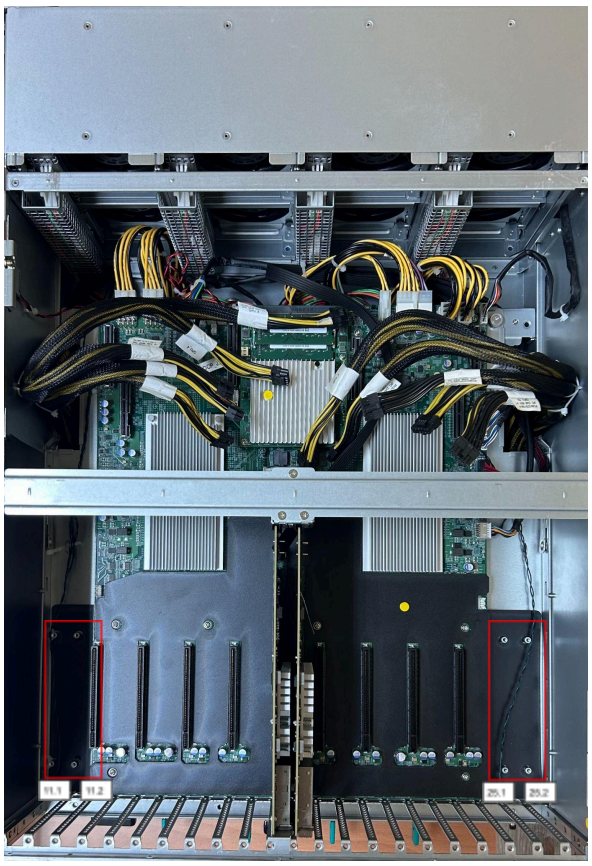


Figure 122. Slot 11.1, 11.2 and Slot 25.1, 25.2 Designated as expansion cables for Retimer Installation (8-GPU)

Step 2. Power Off the System

A. Ensure the system is completely powered down.
B. Disconnect all power cables to prevent any damage during installation.

Step 3. Verify Retimer Installation

A. Verify that the Retimer is securely installed in the designated expansion cable slot.
B. Check that the DIP switch settings are correctly configured according to the module type (C01 or D01).

Step 4. Connect the CDFP Cable

A. Use CDFP cables to connect the server and chassis.
B. Ensure that all cables are securely locked to maintain stable signal integrity.

Caution: Ensure the CDFP cable is securely connected. Failure to properly lock the connector can result in system malfunctions and potential damage to the Falcon 5012 (refer to Figure 123).

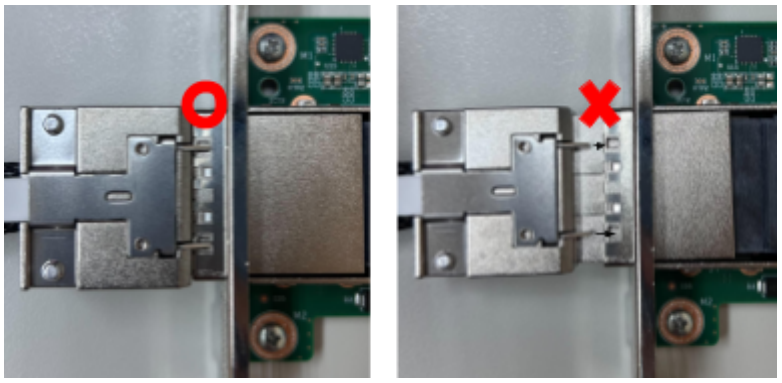


Figure 123. Picture of properly connecting the CDFP cable with the CDFP connector

---

Step 5. Configure System Ports

A. Access the system GUI and navigate to the Port Config interface (refer to Figure 124).

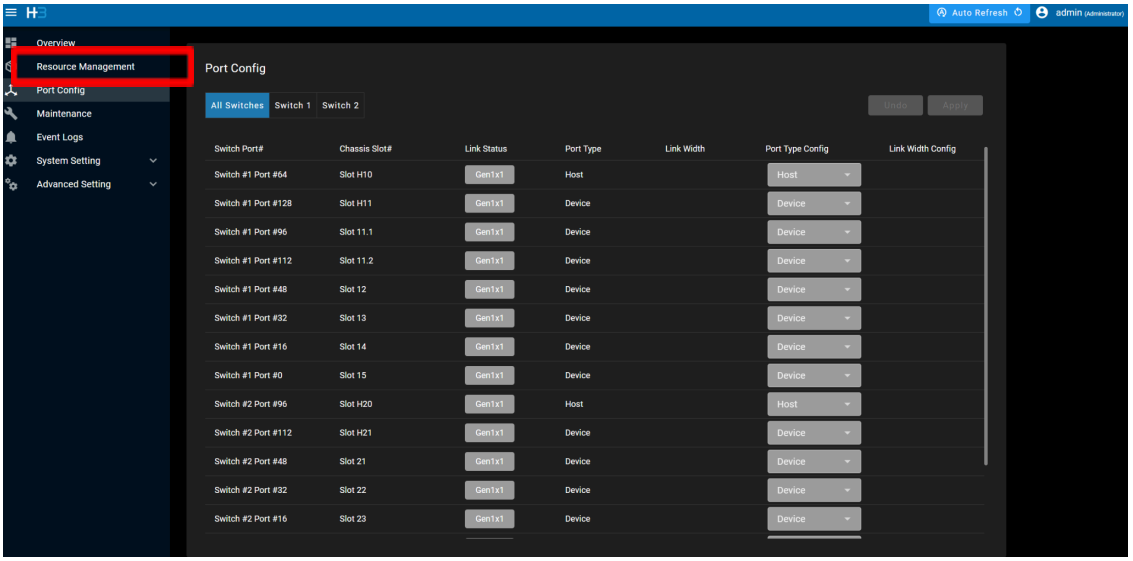


Figure 124. Port Configuration Interface for Falcon 5012 System

B. Adjust expansion cable settings based on Retimer installation:
    - 10-GPU configuration: Configure Slot 11 and Slot 25.
    - 8-GPU configuration: Configure Slot 11.1, 11.2, 25.1, and 25.2.
C. Change the slot role from Device to Host to enable host functionality.(refer to Figure 125)

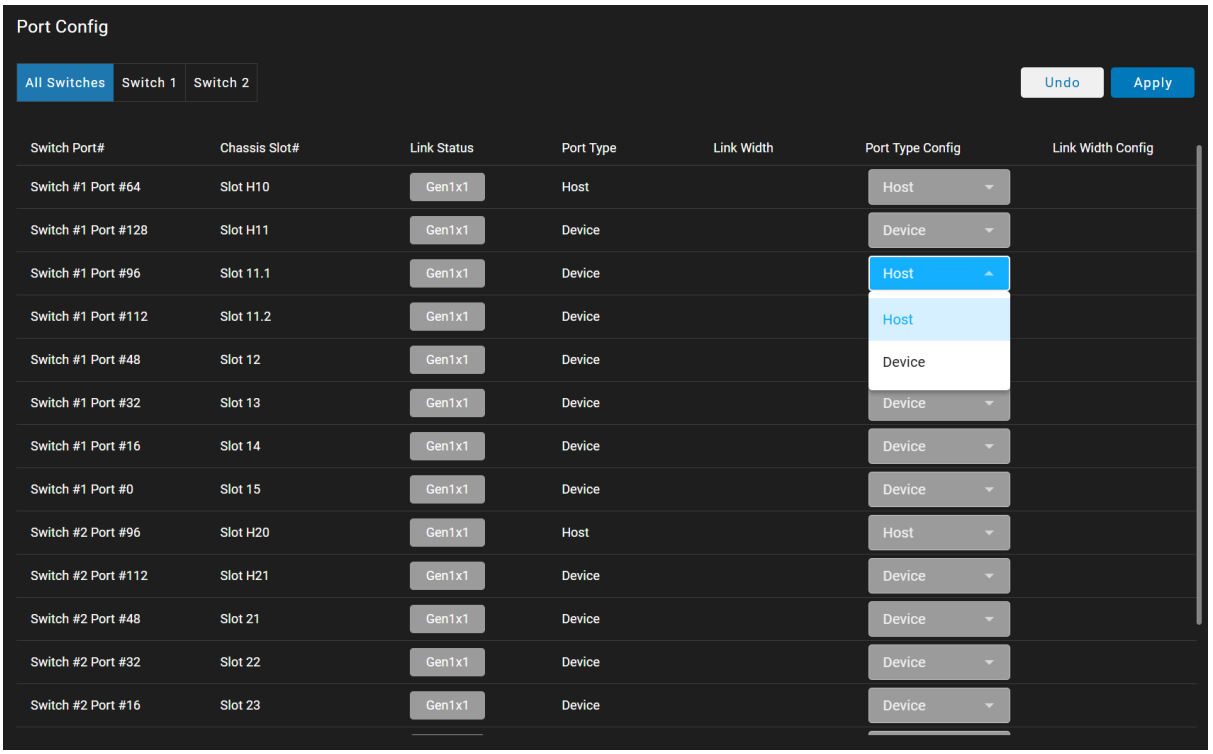


Figure 125. Changing Port Configuration from Device to Host in expansion cables

Following these steps ensures a proper setup for multi-host GPU resource sharing.

Chapter 8

Troubleshooting

8.1 No link between Falcon 5012 and the host

1. Please check the following hardware settings:
   1.1. Host ports usage priority:
   First: H10 and H20; these two are enabled by default in Standard Mode.
   Secondly: H11 and H21; these two are enabled in Advanced Mode (See Chapter 2.2 for more details).
   1.2. DIP-switch settings of the host retimer card (See Chapter 3.5 for details).
      - C01: SW1 = ON, SW2 = HLL
      - D01: SW1 = ON, OFF, OFF, SW2 = HLH
   1.3. CDFP connectors are securely seated and locked between Falcon 5012 and the host side (See Chapter 3.5 for details).
2. Please check the following software requirements:
   2.1. Host BIOS settings (See Chapter 2.3.4 for details)
   2.2. Change the PCIe configuration of the server slot, where the host retimer card is installed, from “auto” to “1 x16”.
   2.3. Update the firmware of the host retimer card to the latest version. Please note that the retimer must be installed in Falcon 5012 to perform the firmware update (See Chapter 5.5 for details).

8.2 When a GPU/PCIe device is not seen from host side

In Standard Mode, a GPU/PCIe device is assigned automatically to the host port. If you are using Advanced Mode, please make sure to manually assign the GPUs/PCIe devices to one of the host ports.

8.3 When the link of a GPU/PCIe device is not established or not stable

1. Please check the following hardware requirements (See Chapter 3.4):
   1.1. Whether the GPU/PCIe device is installed properly?
   1.2. Whether the aux power cable is connected properly?
   1.3. Whether the bracket screw is well screwed?
2. Please perform a power cycle on the PCIe slot to re-train the PCIe link.

8.4 When the name of a GPU/PCIe device is shown as an “unknown device” in the Device List of the web GUI

Due to access limitations between the PCIe switch and the device, sometimes a GPU/PCIe device may operate normally while its device name is not detected in Standard Mode. In this case, the device will appear as an “unknown device”. However, additional PCIe information is still available through the drop-down menu in the web GUI.

8.5 When the GPUs/PCIe devices shut down unexpectedly during application execution

Please check whether the brackets are installed on all unused slots. Missing brackets may cause airflow leakage through empty slots, causing overheating of the whole chassis.
Under normal operation, a PCIe expansion chassis is designed for proper installation of brackets on unused slots to maintain a balanced and consistent thermal environment across the entire chassis.

Warranty

H3 Platform offers free 2-year limited warranty service on our products. If the standard warranty period is not enough, H3 Platform offers optional purchase extended warranty service coverage that will provide a warranty for 1, 2, or 3 additional years.

Limited Warranty

H3 Platform warrants its products against defects in material and workmanship. Under normal use and service, every hardware portion of the products will be free from physical defects in material and workmanship during the warranty period, or the product will be repaired or replaced as determined solely by H3 Platform. H3 Platform provides a limited warranty for its products only to the person or entity that originally purchased the product from H3 Platform
or its authorized distributor or retailer. H3 Platform's products are PCI switches, and H3 Platform makes no warranty on the devices installed, or warranty on the compatibility of all PCIe devices. H3 Platform will not be liable in any way for the loss of data stored on H3 Platform products and any damage caused by this.

An extra warranty period might be provided by H3 Platform's worldwide dealers/distributors. Please contact dealers/distributors to purchase the extended warranty.

If a product does not operate as warranted above during the applicable warranty period, H3 Platform shall, at its option and expense (except for shipping cost), repair the defective product or part, and deliver to the customer an equivalent product or part to replace the defective item. All products that are replaced will become the property of H3 Platform. Replacement products may be new or reconditioned.

Warranty does not apply, if:

1. The warranty period is expired,
2. The warranty label is broken or removed,
3. The serial number label is missing or unrecognizable,
4. The product has been modified or repaired by any unauthorized service center or personnel.

5. The defect was subject to abuse, improper use not conforming to product manual instructions, or environmental conditions more severe than those specified in the manual and specification.
6. The defect was subject to Force Majeure, such as acts of God, flood, lighting, earthquake, war, vandalism, theft, brownouts, or sags (damage due to low voltage disturbances)
7. Consumables, such as batteries
8. Routine cleaning, or normal cosmetic and mechanical wear.
9. Damage caused by misuse, abuse, or neglect
10. Damage caused by parts that were not manufactured or sold by H3 Platform
11. Damage caused by installing devices not on the compatible list
12. Damage caused when warranted parts were repaired or replaced by an organization other than H3 Platform or by a service provider not authorized by H3 Platform.

If the customer's product is not covered under warranty, H3 Platform may offer Repair Services under the customer's payment.

Software Limited Warranty

Except as provided in the applicable software end-user license or program license agreement, including any software products or the operating system preinstalled by H3 Platform are provided “as is” and with all faults, and H3 Platform hereby disclaims all other warranties and conditions, either express, implied, or statutory, including, but not limited
to, warranties of title and non-infringement, any implied warranties, duties or conditions of merchantability, of fitness for a particular purpose, and lack of viruses. Some states/jurisdictions do not allow the exclusion of implied warranties or limitations on the duration of implied warranties, so the above disclaimer may not apply to you in its entirety. To the maximum extent permitted by applicable law, in no event shall H3 Platform or its suppliers be liable for any special, incidental, indirect, or consequential damages whatsoever (including, but not limited to, damages for loss of profits or confidential or other information, for business interruption, for personal injury, for loss of privacy arising out of or in any way related to the use of or inability to use the software product, even if H3 Platform or any supplier has been advised of the possibility of such damages and

even if the remedy fails of its essential purpose. H3 Platform is not responsible for any interoperability or compatibility issues. Some states/jurisdictions do not allow the exclusion or limitation of incidental or consequential damages, so the above limitation or exclusion may not apply to you.

H3 Platform warrants all pre-installed H3 Platform software will execute its programming instructions when all files are properly installed. H3 Platform does not warrant that the software will be uninterrupted or error-free. In the event H3 Platform software fails to execute its programming instructions during the warranty period, the customer's remedy will be either (1) replacement of the H3 Platform software or (2) a refund upon return of the product and all copies of the software as well as installation instructions.

Compatibility

We have tested and verified the compatibility of third-party products and peripherals on the H3 Platform Compatibility List. Please note that the tests conducted in H3 Platform labs may not have covered every aspect and that changes to firmware or hardware may affect device compatibility and stability. H3 Platform does not guarantee compatibility with third-party products and peripherals on the List and reserves the right to update the List at any time without prior notice.

Repair Service

Warranty and out-of-warranty service should be obtained by contacting the system integrator/dealer/retailer/e-retailer or distributor where the customer purchased the product. When requesting service, the proof of purchase and the product serial number must be provided. The return of the defective product should be strictly through the original route of purchase, and the customers shall pack the product appropriately to prevent the returned product from suffering in transportation.

There are no user-serviceable parts inside the product. Do not allow unauthorized service centers or personnel to repair or modify the product. If the original channel is no longer in business or unavailable, the customer may contact H3 Platform Technical Support for international RMA services. In this case, the customer will be charged a handling fee, of $50 (USD), and all fees incurred, including two-way freight, duties, taxes, and brokerage fees.

It is the customer's sole responsibility to back up his/ her data. Before allowing any service from H3 Platform or its service provider, including remote login check and repairing service, the customer must back up the data and remove any of the customer's confidential, proprietary, or personal information. Neither H3 Platform nor its service provider will be liable for any damage, loss, or exposure of confidential or private information or data in any
product, hardware, software, or media.

Notice of Warranty Claims

All warranty claims must be made during the applicable warranty period. Any claim made after that time will not be eligible for warranty service but may be serviced under a separate H3 Platform support contract covering the product.

DISCLAIMER OF WARRANTY

H3 Platform does not warrant uninterrupted or error-free operation of an H3 Platform product. If the customer transfers an H3 Platform product to another user, warranty service is NOT available to that user for the remainder of the warranty period. Consumables, such as batteries and PCIe power cables have no warranty. The above warranties
will not apply to products from which serial numbers have been removed or to defects resulting from misuse (including operation of an H3 Platform product without covers and incorrect input voltage), unauthorized modification, operation, or storage outside the environmental specifications for the product, in-transit damage, improper maintenance, or defects resulting from use of third-party software, accessories, media, supplies, consumables, or such items not designed for use with the product. The H3 Platform Software Product Limited Warranty does not apply to software used with products not covered by any of the H3 Platform limited warranties offered as part of your purchase of an H3 Platform product.

TO THE EXTENT ALLOWED BY LOCAL LAW, THE ABOVE WARRANTIES ARE EXCLUSIVE, AND NO OTHER WARRANTY OR CONDITION, WHETHER WRITTEN OR ORAL, IS EXPRESSED OR IMPLIED AND H3 PLATFORM SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OR CONDITIONS OF MERCHANTABILITY, SATISFACTORY QUALITY, AND FITNESS FOR A PARTICULAR PURPOSE.

Some Countries, states, or provinces do not allow limitations on the duration of an implied warranty, so the above limitation or exclusion might not apply to you. The warranty gives you specific legal rights and you might also have other rights that vary from country to country, state to state, or province to province.

Limitation of Liability and Remedies.

In no event will H3 Platform or its affiliates, subcontractors, or suppliers be liable for any of the following:

1. Damages for loss of data, or software restoration;
2. Damages relating to the Customer's procurement of substitute products or services (i.e., "cost of cover");
3. Incidental, special, or consequential damages (including lost profits, loss of use, Customer downtime, cost of data recovery/re-creation), even if H3 Platform is informed of their possibility;
4. Third-party claims against Customers for losses or damages.
5. H3 Platform PCIe Switches are not specifically designed, manufactured, or intended for sale as parts, components, or assemblies for the planning, construction, maintenance, or direct operation of a nuclear facility. Customer is solely liable if Products or Support purchased by Customer are used for these applications.

TO THE EXTENT ALLOWED BY LOCAL LAW, THE REMEDIES IN THIS WARRANTY STATEMENT ARE THE CUSTOMER'S SOLE AND EXCLUSIVE REMEDIES. EXCEPT AS INDICATED ABOVE, IN NO EVENT WILL H3 PLATFORM OR ITS SUPPLIERS BE LIABLE FOR LOSS OF DATA OR FOR DIRECT, SPECIAL, INCIDENTAL, CONSEQUENTIAL (INCLUDING LOST PROFIT OR DATA), OR OTHER DAMAGE, WHETHER BASED IN CONTRACT, TORT, OR OTHERWISE.

Some countries, states or provinces do not allow the exclusion or limitation of incidental or consequential damages, so the above limitation or exclusion may not apply to you.

THE WARRANTY TERMS CONTAINED HERE, EXCEPT TO THE EXTENT LAWFULLY PERMITTED, DO NOT EXCLUDE, RESTRICT OR MODIFY AND ARE IN ADDITION TO

THE MANDATORY STATUTORY RIGHTS APPLICABLE TO THE SALE OF THIS PRODUCT TO YOU.