NF5468M6
White Paper

Document Version: V2.2
Release Date: February 7, 2023

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent.

Environmental Protection

Please dispose of product packaging by recycling at a local recycling center for a greener planet.

Trademarks

All the trademarks or registered trademarks mentioned herein may be the property of their respective holders. This document does not mark any product or brand with symbol (® or ™).

Security Statement

We are intensely focused on server product safety and have placed a high priority on this. For a better understanding of our server products, carefully read through the following security risk statements.

- When servers are to be repurposed or retired, it is recommended to restore their firmware factory settings, delete information and clear logs from BIOS and BMC to protect data privacy. Meanwhile, we recommend you to wipe the drive data thoroughly and securely with trusted erasing tools. You can use ISQP. Contact us for specific server models to which ISQP is applicable.
- For server open source software statements, please contact us.
- Some interfaces and commands for production, assembly and return-to-depot, and advanced commands for locating faults, if used improperly, may cause equipment abnormality or business interruption. This is not described herein. Please contact us for such information.
- External ports of our servers do not use private protocols for communication.
- Our products will not initiatively obtain or use your personal data. Only when you consent to use certain functions or services, some personal data such as IP address and email address for alerts may be obtained or used during business operation or fault location. We have implemented necessary measures on product functions to ensure personal data security throughout the data lifecycle, including but not limited to data collection, storage, use, transmission, and destruction. Meanwhile, you are obligated to establish necessary user privacy policies in accordance with applicable national/regional laws and regulations to fully protect user personal data.
- Committed to product data security, we have implemented necessary measures on

product functions to protect system operation and security data throughout its lifecycle in strict accordance with relevant laws, regulations and supervision requirements. As the owner of system operation and security data, you are obligated to establish necessary data security policies and take adequate measures in accordance with applicable national/regional laws and regulations to fully protect system operation and security data.

- We will remain committed to the safety of our products and solutions to achieve better customer satisfaction. We have established emergency response procedures and action plans for security vulnerabilities, so that product safety issues can be dealt with in a timely manner. Please contact us for any safety problems found or necessary support on security vulnerabilities when using our products.

Disclaimer

The purchased products, services and features shall be bound by the contract made between us and the customer. All or part of the products, services and features described herein may not be within your purchase or usage scope. Unless otherwise agreed in the contract, we make no express or implied statement or warranty on the contents herein. Images provided herein are for reference only and may contain information or features that do not apply to your purchased model. This document is only used as a guide. We shall not be liable for any damage, including but not limited to loss of profits, loss of information, interruption of business, personal injury, or any consequential damage incurred before, during, or after the use of our products. We assume you have sufficient knowledge of servers and are well trained in protecting yourself from personal injury or preventing product damages during operation and maintenance. The information in this document is subject to change without notice. We shall not be liable for technical or editorial errors or omissions contained in this document.

Support

Email	Technical Support	serversupport@aivres.com
	RMA/ARMA Support	serversupportusa@aivres.com
Web	Official Website	www.aivres.com
	Service Portal	www.service.aivres.com

Preface

Abstract

This white paper describes the NF5468M6 server's appearance, features, performance parameters, and software and hardware compatibility, providing in-depth information of NF5468M6.

Intended Audience

This white paper is intended for pre-sales engineers.

Symbol Conventions

The symbols that may be found in this document are defined as follows.

Symbol	Description
DANGER	A potential for serious injury, or even death if not properly handled
WARNING	A potential for minor or moderate injury if not properly handled
CAUTION	A potential loss of data or damage to equipment if not properly handled
IMPORTANT	Operations or information that requires special attention to ensure successful installation or configuration
NOTE	Supplementary description of document information

Revision History

Version	Date	Description of Changes
V1.0	2021/03/30	Initial release
V1.1	2021/09/30	Changed the overall format
V2.0	2021/11/22	Modified product features, exploded views, system architecture, software and hardware compatibility, power consumption calculation, and new technology description
V2.1	2022/04/11	Updated the format
V2.2	2023/02/07	Updated some descriptions

Table of Contents

1 Product Overview 8
2 Features 9
2.1 Super Computing Power 9
2.2 Increased Communication Performance 9
2.3 Multiple Topologies Supported 9
2.4 Large Local Storage Capacity 9
2.5 Multi-Host Technology 9
2.6 Security 10
3 System Parts Breakdown 11
4 Logical Structure 12
4.1 System Topology Overview 12
4.2 PCIe Topologies of NF5468M6 - P Configuration 14
4.3 PCIe Topologies of NF5468M6 - T Configuration 23
4.4 PCIe Topologies of NF5468M6 - V Configuration 24
4.5 Topology Switching Settings for NF5468M6 - P Configuration 25
4.5.1 Getting GPU Topology Features 26
4.5.2 Setting GPU Topology Switching 27
5 Hardware Description 29
5.1 Front Panel 29
5.1.1 Front View 29
5.1.2 Buttons and LEDs 30
5.1.3 Ports 32
5.2 Rear Panel 33
5.2.1 Rear View 33
5.2.2 Buttons and LEDs 34
5.2.3 Ports 36
5.3 Processors 37
5.4 Memory 38
5.4.1 DDR4 DIMMs 38
5.5 Storage 44
5.5.1 Drive Configurations 44
5.5.2 Drive Numbering 45
5.5.3 Drive LEDs 48
5.5.4 RAID Controller Cards 49
5.6 Network 49
5.7 I/O Expansion 49

5.7.1 PCIe Cards 49
5.7.2 PCIe Slots 50
5.8 PSUs 53
5.9 Fans 54
5.10 Boards 55
5.10.1 Motherboard 55
5.10.2 GPU Boards 56
5.10.3 Drive Backplanes 58
6 Product Specifications 61
6.1 Technical Specifications 61
6.2 Environmental Specifications 63
6.3 Physical Specifications 65
7 Operating System and Hardware Compatibility 67
7.1 Supported Operating Systems 68
7.2 Hardware Compatibility 69
7.2.1 CPU Specifications 69
7.2.2 DIMM Specifications 69
7.2.3 Drive Specifications 70
7.2.4 SAS/RAID Controller Card Specifications 71
7.2.5 NIC Specifications 71
7.2.6 HCA Card Specifications 72
7.2.7 GPU Specifications 72
7.2.8 PSU Specifications 72
8 Regulatory Information 74
8.1 Safety 74
8.1.1 General 74
8.1.2 Personal Safety 74
8.1.3 Equipment Safety 76
8.1.4 Transportation Precautions 77
8.1.5 Manual Handling Weight Limits 77
8.2 Warranty 78
9 System Management 79
9.1 Intelligent Management System ISBMC 79
9.2 ISPIM 81
9.3 ISIB 83
10 Certifications 85
11 Appendix A 86
11.1 Operating Temperature Specification Limits 86

11.2 Model 86
11.3 RAS Features 86
11.4 Sensor List 86
12 Appendix B - Acronyms and Abbreviations 91

1 Product Overview

The NF5468M6 is an industry-leading AI cloud infrastructure server designed for mainstream AI computing scenarios such as AI public cloud, enterprise-level AI cloud platform, smart security, and video encoding/decoding. It supports PCIe 4.0, twice as fast as PCIe 3.0. In addition, the NF5468M6 is compatible with accelerator cards in different PCIe form factors and supports topology switch between Balance, Common, and Cascade, meeting customers' performance requirements for different AI application scenarios. With up to twenty-four 2.5-inch drives or twelve 3.5-inch drives at the front, the server provides a high storage capacity.

Figure 1-1 NF5468M6

2 Features

2.1 Super Computing Power

- Up to two 3rd Gen Intel Xeon Scalable processors (code-named “Ice Lake”) with a TDP of up to 270 W in a standard 4U space.
- 3 UPI links per CPU.
- 32 DDR4 DIMMs.
- 4 or 8 dual-slot FHFL PCIe GPU cards or 16 Tesla T4 graphics cards.
- With wide compatibility to support various GPU cards.

2.2 Increased Communication Performance

- Supports PCIe 4.0 x16 accelerator cards with a bandwidth of up to 64 GB/s.
- Backward compatible with PCIe 3.0 accelerator cards.
- Supports NVLink Bridges, further enhancing the P2P performance.

2.3 Multiple Topologies Supported

The NF5468M6 supports Balance, Common, and Cascade topologies. Topology switching can be achieved by BMC commands to meet performance requirements for different AI application scenarios in the most convenient way.

2.4 Large Local Storage Capacity

The NF5468M6 supports twenty-four 2.5-inch drives or twelve 3.5-inch drives at the front, with up to 8 U.2 NVMe drives supported, delivering efficient data access. Besides, it supports 2 internal M.2 SSDs, further expanding the system storage capacity.

2.5 Multi-Host Technology

Multi-Host technology enables direct connectivity from multiple compute or storage hosts to a single network controller, reducing capital and operating expenses by reducing the quantity of switches, network cards, and cables. Besides, Multi-Host technology can balance the performance of network communication among the cores of multiple processors, providing an efficient network communication environment.

2.6 Security

- Supports Trusted Platform Module (TPM 2.0) and Trusted Cryptography Module (TCM) that provide advanced encryption functions.
- Supports Intel Trusted Execution Technology that provides hardware-based mechanisms to help protect against software-based attacks.
- Supports Intel Software Guard Extensions (SGX) technology that allows applications to run in enclaves to help prevent malicious theft and modification of critical codes and data.
- Supports the firmware update mechanism based on digital signatures to prevent unauthorized firmware updates.
- Supports UEFI Secure Boot to protect the system from malicious boot loaders.
- Supports hierarchical password protection in BIOS to ensure system boot and management security.
- Supports BIOS Secure Flash and BIOS Lock Enable (BLE) to reduce attacks from malicious software on the BIOS flash regions.
- Supports dual-image mechanism for BMC and BIOS to recover firmware upon the detection of corrupted firmware.
- Supports BMC Secure Boot to protect BMC from malicious tampering.
- Supports flexible BMC access control policies to improve BMC management security.
- Supports chassis intrusion detection to enhance physical security.

3 System Parts Breakdown

Figure 3-1 NF5468M6 Exploded View (12 × 3.5-inch Drive Configuration)

Item	Feature	Item	Feature
1	Front Bezel	9	PSUs
2	Drives	10	M.2 SSDs
3	Drive Backplane	11	Air Duct
4	Top Cover	12	Motherboard
5	GPUs	13	Fans
6	PCIe Expansion Cards	14	Chassis
7	GPU Board and Tray	15	RAID Controller Card Module
8	OCP 3.0 Card		

4 Logical Structure

4.1 System Topology Overview

The NF5468M6 supports two 3rd Gen Intel Xeon Scalable processors (code-named “Ice Lake”) and 32 DDR4 DIMMs. The 2 processors are interconnected by 3 UPI links at up to 11.2 GT/s. Each processor is connected to 16 Slimline x8 connectors on the motherboard via the PCIe bus to provide PCIe signal.

The server supports 1 RAID riser card which supports up to 2 RAID controller cards and is connected to Slimline x8 connectors of CPU0 via the PCIe cable. The RAID controller cards can be connected to different drive backplanes via SAS/SATA signal cables to provide multiple local storage solutions. NVMe SSDs can be connected to a tri-mode RAID controller card, supporting hardware RAID.

There are 3 Slimline x4 connectors directly routed from the PCH that can be connected to 12 SATA drives on the drive backplane via SATA cables, providing a low-cost local storage solution. Mixing drives directly attached to a PCH with drives directly attached to a RAID controller card is not supported, because wrong drive sequence issue will occur if storage controllers from different manufacturers are used in the same system. The motherboard supports up to 8 NVMe SSDs directly connected to CPUs, via connecting the drive backplane to Slimline x8 connectors directly routed from the CPU on the motherboard.

The OCP 3.0 riser card supports an OCP 3.0 card in Single-Host or Multi-Host mode. In Single-Host mode, the OCP 3.0 riser card is connected to CPU0 via their Slimline x8 connectors. In Multi-Host mode, the OCP 3.0 riser card is connected to CPU0 and CPU1 via connecting its 2 Slimline x8 connectors with the Slimline x8 connectors of CPU0 and CPU1 respectively.

The NF5468M6 supports 8 GPUs in P configuration, 4 GPUs in T configuration, and 16 GPUs in V configuration. The GPU board is connected to the motherboard via a Slimline x8 cable.

Figure 4-1 Motherboard Logic Block Diagram

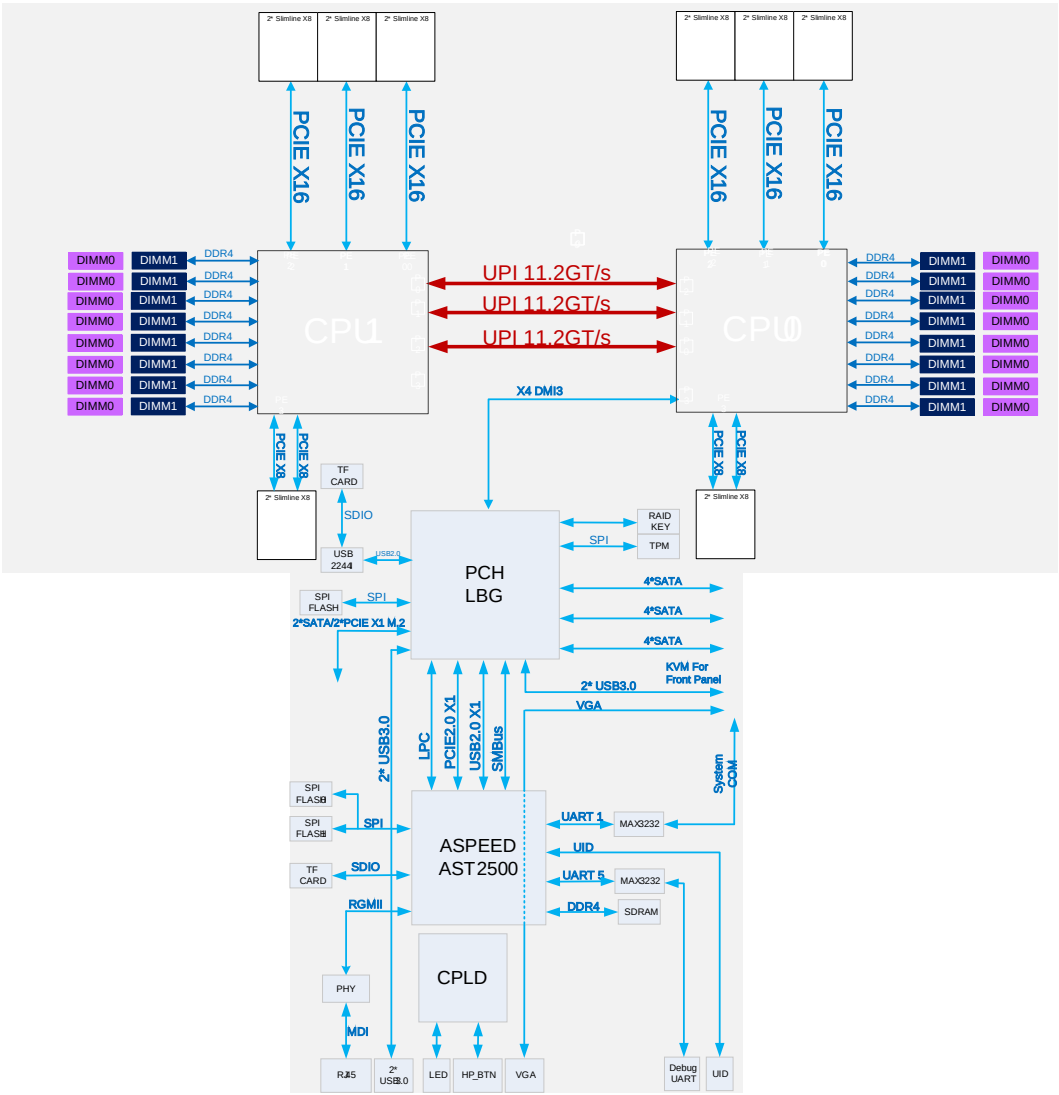
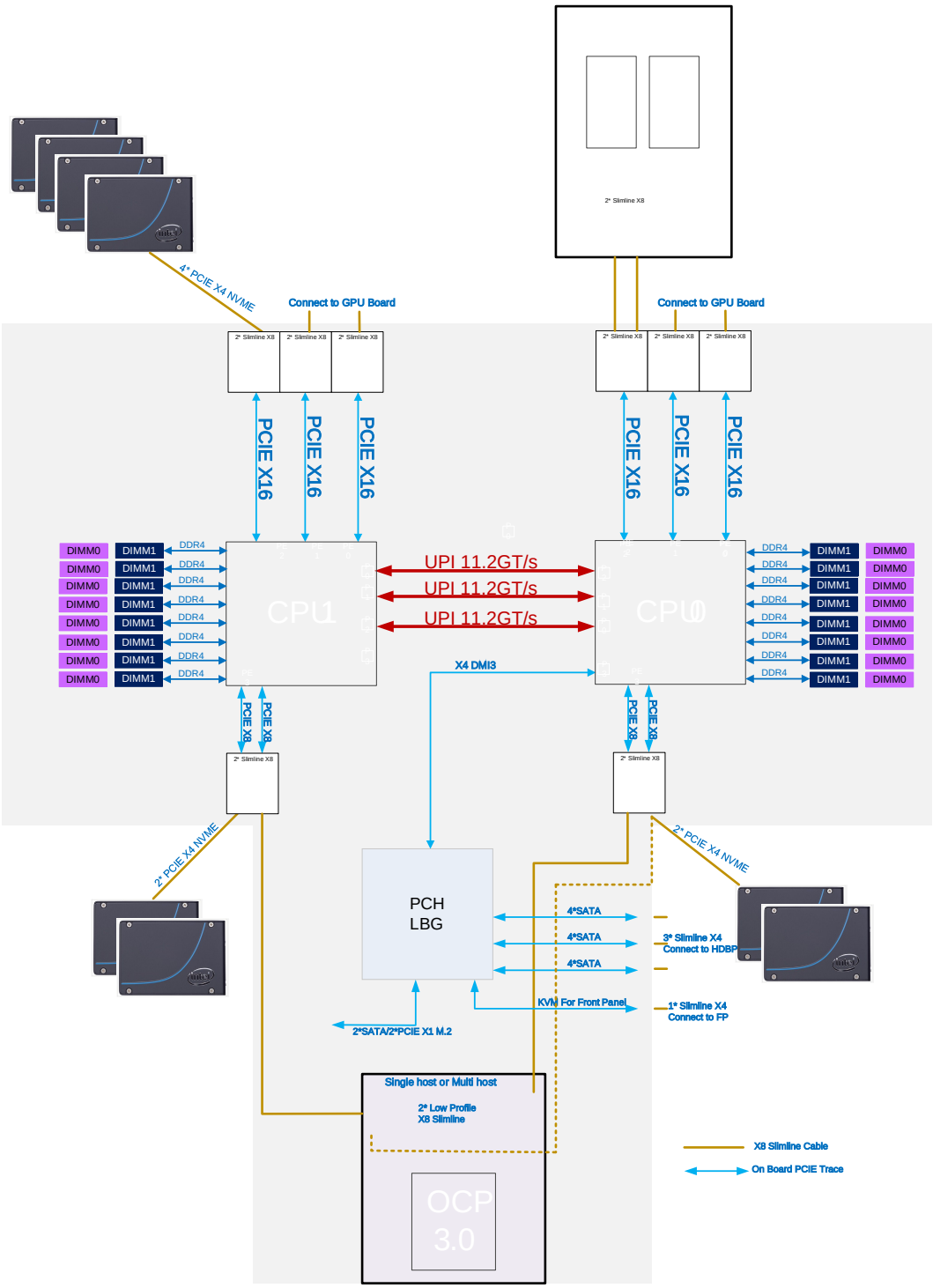


Figure 4-2 PCIe Topology Logic Block Diagram

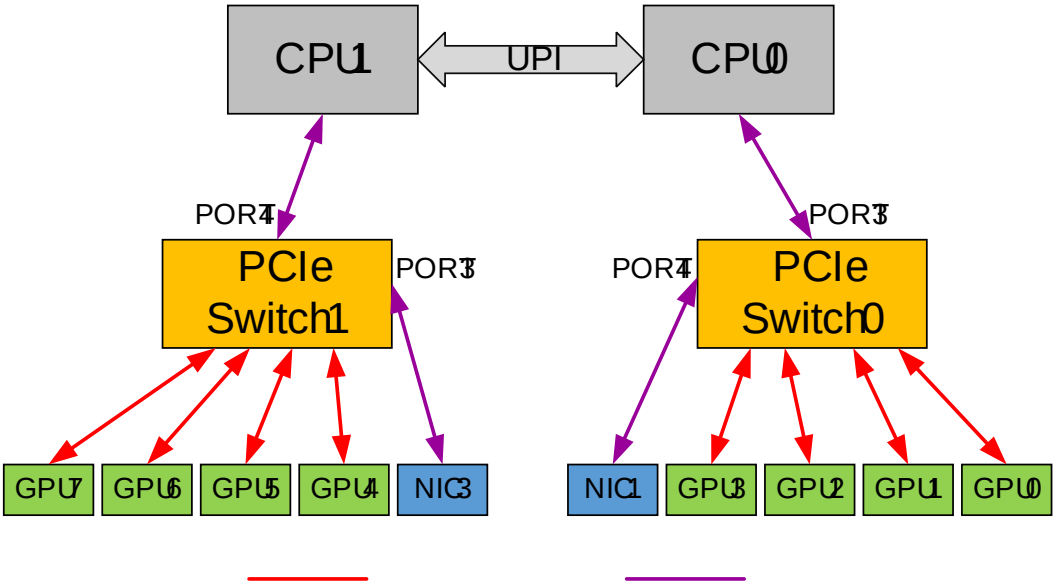


4.2 PCIe Topologies of NF5468M6 - P Configuration

NF5468M6 - P configuration supports remote topology switching between Balance,

Common, and Cascade modes by BMC commands, thereby meeting the needs of AI computing in different business scenarios.

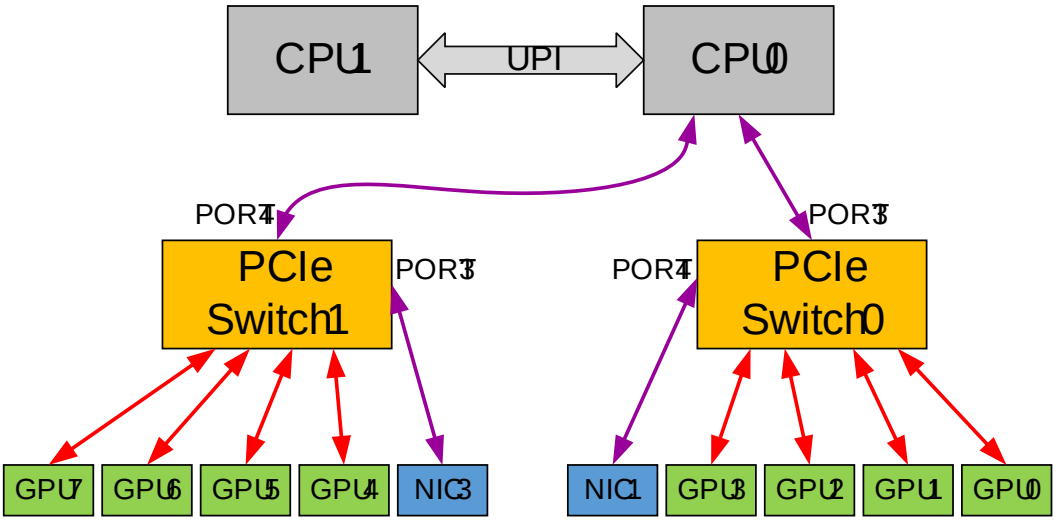
Figure 4-3 Balance Topology



Features: Each CPU has a PCIe switch, which is then connected to 4 GPUs. This topology maximizes CPU utilization and provides a greater upstream bandwidth. But in this topology, the P2P communication between 2 GPUs connected to different CPUs is limited by the UPI link speed.

Applicable business scenarios: As the mainstream topology of NF5468M6 servers, it is applicable to high performance computing (HPC), virtual desktop infrastructure (VDI), public cloud, AI training and other scenarios.

Figure 4-4 Common Topology

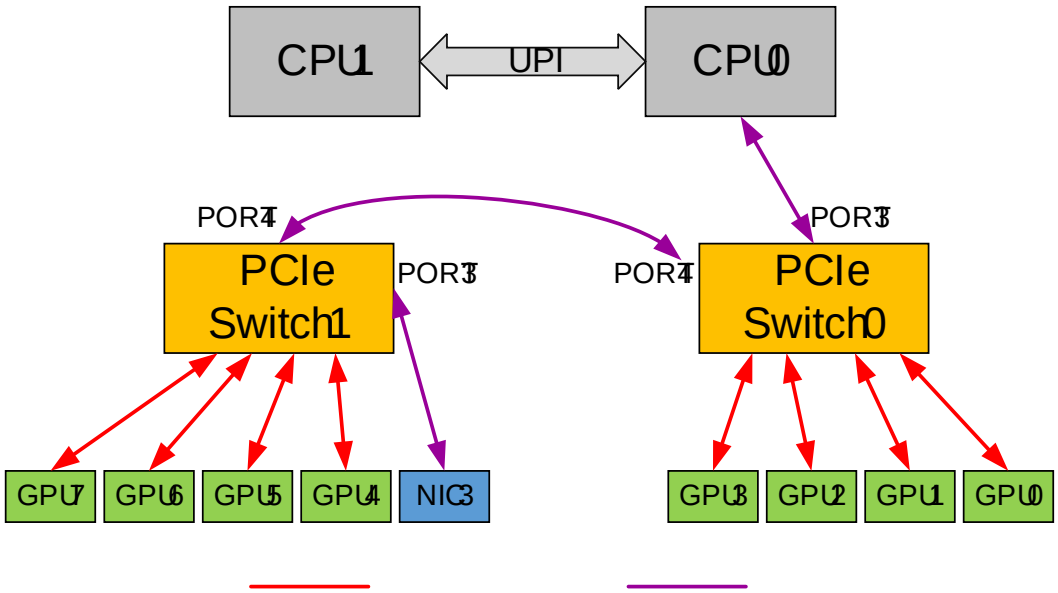


Features: CPU0 is connected to 2 PCIe switches. Each PCIe switch is connected to 4

GPUs. The communication between 2 GPUs connected to different PCIe switches can be accomplished without cross-CPU communication, delivering better P2P communication and higher data throughput.

Applicable business scenarios: It is applicable to the P2P communication-intensive training algorithm models where CPUs are involved in many tasks, for example, ResNet101/50.

Figure 4-5 Cascade Topology



Features: Only CPU0 is connected to a PCIe switch and this PCIe switch interconnects with another one. Each PCIe switch is connected to 4 GPUs. The PCIe switch interconnection provides more powerful GPU P2P communication but low data throughput from CPU to GPU.

Applicable business scenarios: It is applicable to the P2P parameter-intensive training algorithm models where CPUs are involved in few tasks, for example, VGG-16.

Based on the Balance, Common and Cascade topologies, there are a total of 14 PCIe topology options for the NF5468M6 - P configuration, which are as follows:

Figure 4-6 PCIe Topology Logic Block Diagram (Balance or Cascade GPU Topology/Single-Host OCP NIC 3.0)

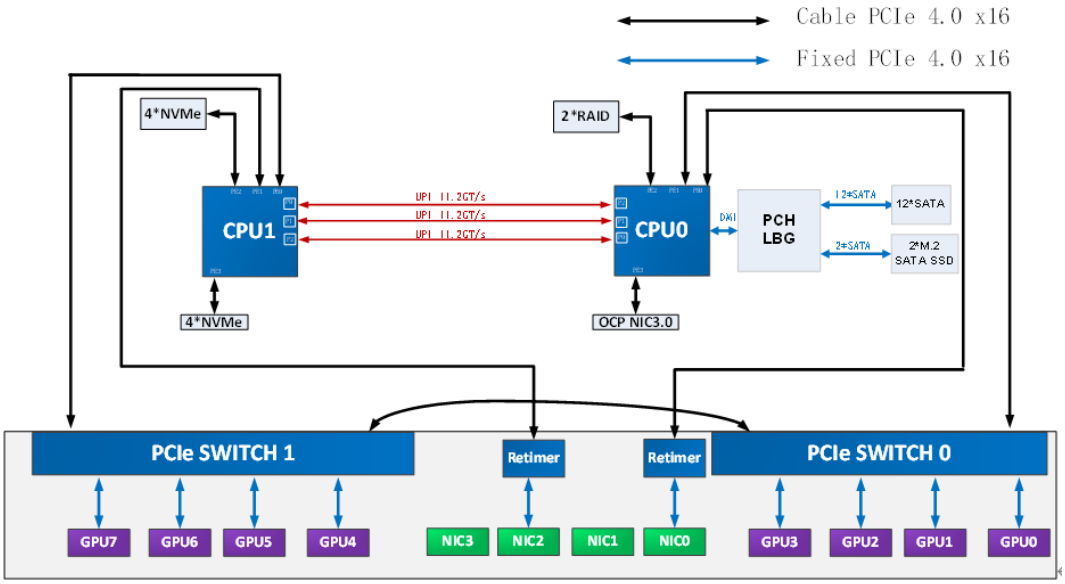


Figure 4-7 PCIe Topology Logic Block Diagram (Balance or Cascade GPU Topology/Multi-Host OCP NIC 3.0)

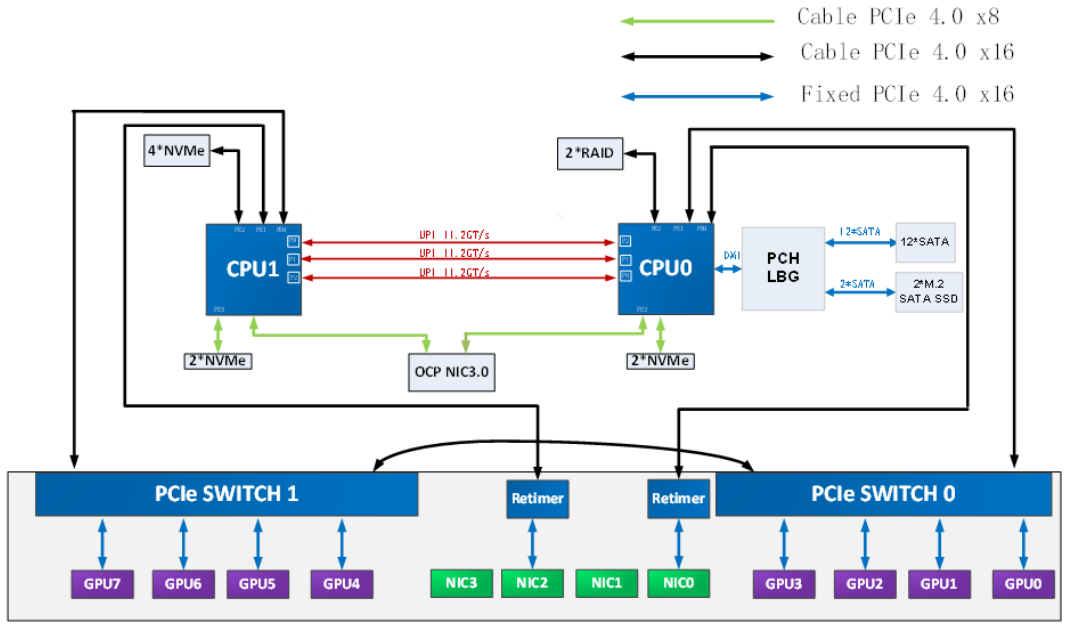


Figure 4-8 PCIe Topology Logic Block Diagram (Balance or Common GPU Topology/Single-Host OCP NIC 3.0)

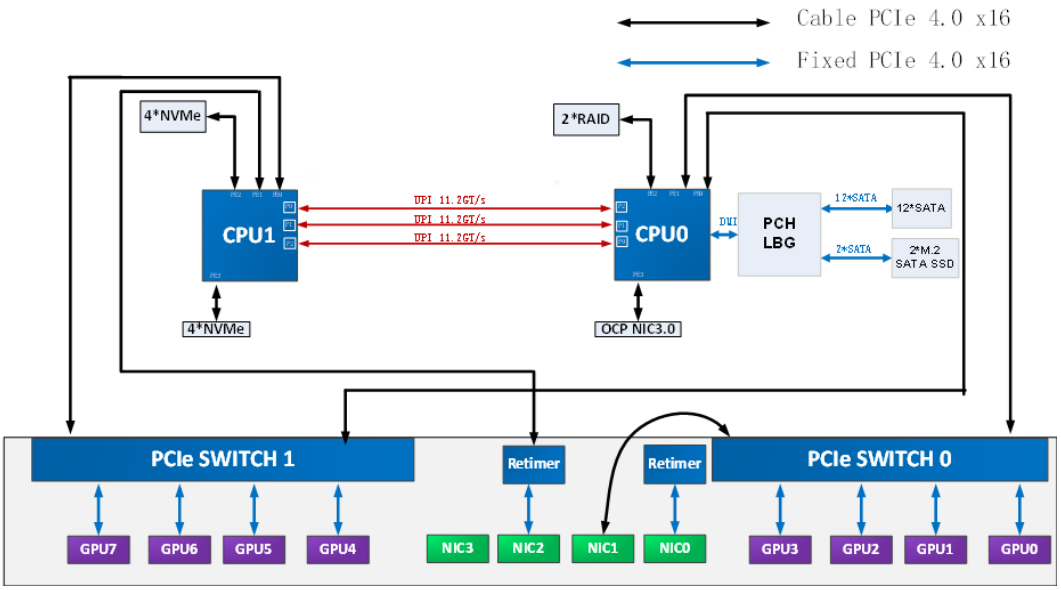
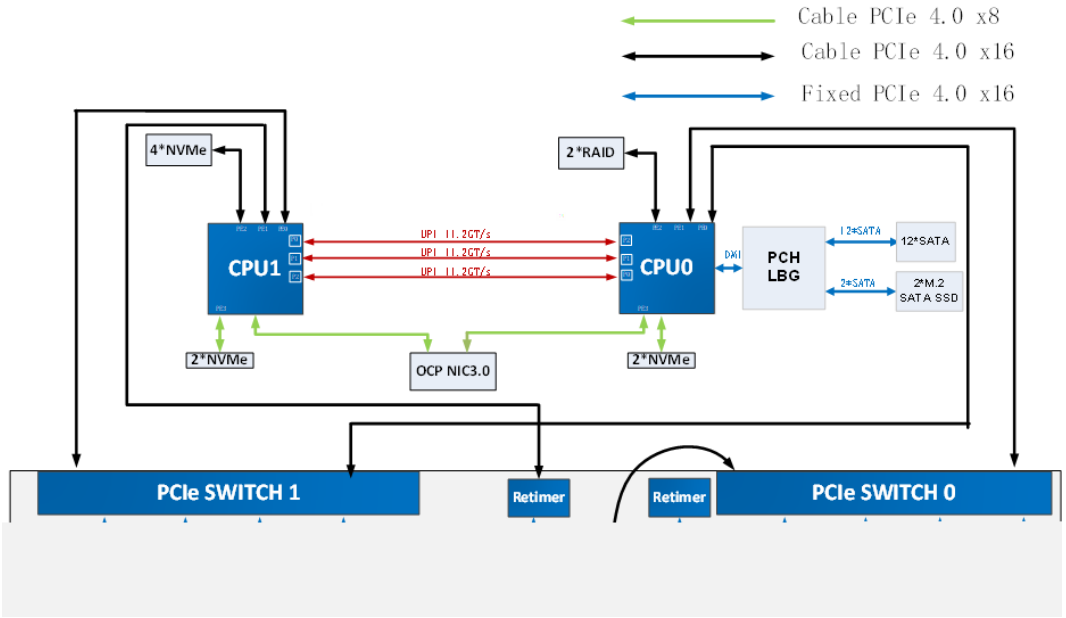



Figure 4-9 PCIe Topology Logic Block Diagram (Balance or Common GPU Topology/Multi-Host OCP NIC 3.0)