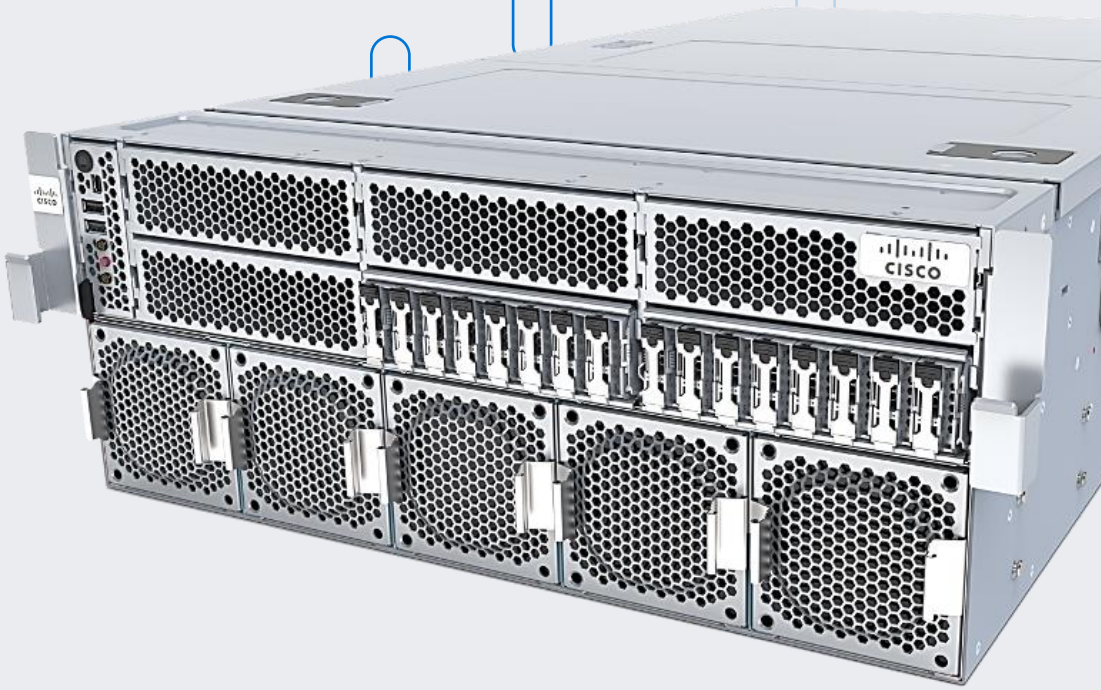
Cisco **Compute**

TECHNICAL MARKETING

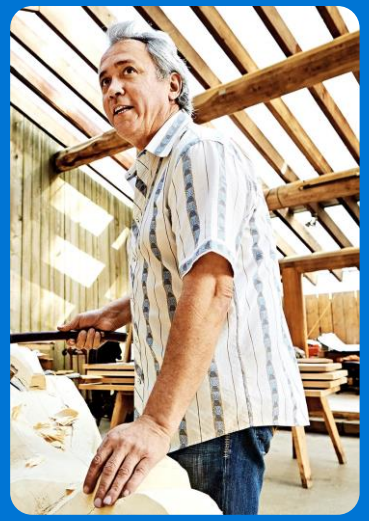


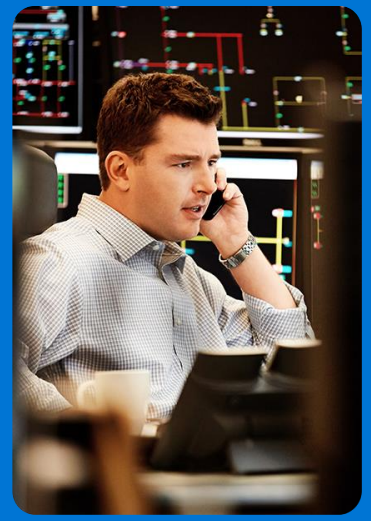
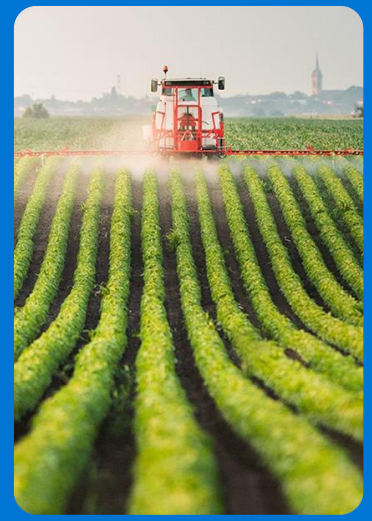
UCS C845A M8

Technical Decision Maker Deck

May. 2025

Every organization's AI approach and needs are different


Build the model
Training

Optimize the model
Fine-tuning and RAG

Use the model
Inferencing

Bringing Scale Up and Scale Out capabilities to the Cisco UCS family and to Cisco's AI solution portfolio

- Training: large/small models
- Fine-tuning
- Large Model Inferencing
- Retrieval-Augmented Generation (RAG)

UCS Accelerated

UCS C845A M8

Nvidia MGX with 2/4/6/8 PCI Nvidia H100 NVL, H200 NVL or L40S, AMD MI210

2 AMD 5th Gen EPYC™ Processors

UCS C845A M8

Positioning

Compute AI Portfolio

Address AI workloads with visibility, consistency, and control

Validated solutions for AI with compute, network, storage, and software

Build the model
Training

Optimize the model
Fine-tuning and RAG

Use the model
Inferencing

UCS Dense GPU Servers

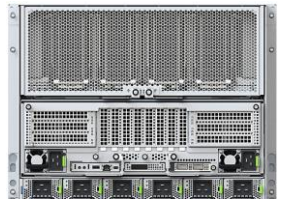
UCS Blade (w/GPU Expansion) and Rack Servers

Enterprise AI Edge

Dense compute for demanding AI

Full stack AI with compute and networking

Cisco Rack Server for AI Workloads



	C240 M7	C245 M8	C240 M8	C885A M8 HGX	C845A M8 MGX
Form Factor	2RU	2RU	2RU	8RU	4RU
Processor Family	(2) Intel Xeon 4th and 5th gen processor	(2) AMD 4th and 5th gen EPYC processor	(2) Intel Xeon 6th Gen processor	(2) AMD 5th Gen EPYC processor	(2) AMD 5th Gen EPYC processor
Expansion Slots	(8) PCIe 5 with (3) Riser slots Max power 350W	(8) PCIe 5 with (3) Riser slots Max power 400W	3 Riser slots Max power 450W (500W engineering max)	(13) PCIe 5 slots (1) OCP	(5) PCIe Gen5 slots (1) OCP slot
Number and connection type of GPUs	**(2)** GPUs NVIDIA H100 NVL *400W* H100, L40S *350W* **(3)** GPUs NVIDIA L40 *300W* Intel Flex 170 *150W*	**(2)** NVIDIA H100 NVL *400W*, H100, L40S *350W* **(3)** NVIDIA L40 *300W*, A16 *250W* AMD Instinct M10-210* Dec24	**(3)** GPUs NVIDIA H100 NVL *400W*, L40S 350W Future GPU Models	**(8)** GPUs NVIDIA HGX H100, H200 SXM5 *700W* AMD MI300X OAM *750W*	**(2-8)** GPUs NVIDIA *H200 NVL 600W* H100 NVL *400W* *L40S 350W* *AMD MI210 300W*
GPU Fabric	PCIe	PCIe	PCIe	NVLINK AMD Infinity Fabric	NVLINK or Infinity Fabric Bridges
GPU Placement	Rear of server PCIe riser	Rear of server PCIe riser	Rear of server PCIe riser	Top of Chassis front facing	Rear of server PCIe riser
PSU count and max power	(2) 2300W PSUs	(2) 2300W PSUs	(2) 2300W PSUs	(6) 3000W and (2) 2700W PSUs	(4) 3200W PSUs
Cooling	Air-cooled	Air-cooled	Air-cooled	Air-cooled	Air-cooled

Cisco UCS C845A Versatile AI Server

Optimized for GenAI

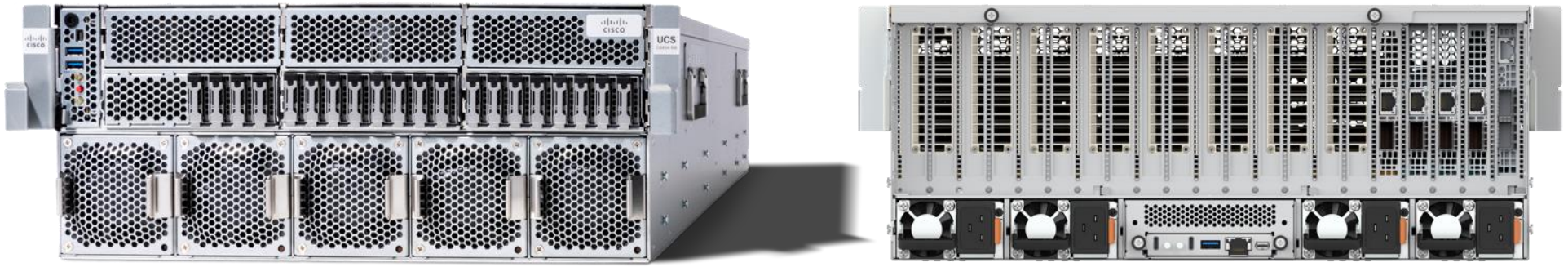
- Address a variety of AI use cases
- Modular NVIDIA MGX™ architecture enables configuration flexibility
- Built on air-cooled enterprise rack design that easily fits in your data center

Scalable AI performance

- Choose the number of GPUs that match your use case
- Increase the number of GPUs when your workloads demand it
- Easily create clusters of AI servers to scale out and meet increasing needs

Consistent management

- Cisco Intersight delivers a unified management paradigm across all your UCS servers
- Manage your AI server with the same tool as your traditional servers



C845A Business Use Cases

High-performance computing (HPC)

- Complex calculations required in simulations and large-scale data processing

Data analytics—visualization

- Leverage advanced analytics tools to extract insights from large datasets, facilitating AI-driven decision making
- Apply simulation technologies to design innovative products and optimize operational workflows

AI

- GenAI training and fine-tuning
- Language processing
- Conversational AI

Traditional workloads

- Acceleration of graphics and rendering to generate complex visual images
- Virtual desktop for graphic-intensive applications such as CAD and 3D modeling

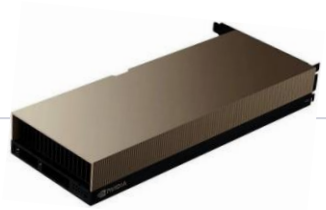
C845A Use Cases By GPU

NVIDIA H100 NVL

- Large-Scale HPC & AI Workloads
- AI Research and high-performance computing
- Large language model generative AI Inference
  - AI Training an HPC

NVIDIA L40S

- Language Processing
- Conversational AI
- Recommenders
- Graphics & Rendering
- Omniverse
- Virtual Desktops



NVIDIA H200 NVL

- High-performance LLM Inferencing
- Gen AI Training and Fine-Tuning

AMD MI210

- Excels at double-precision (FP64) accelerated compute
- Training large-scale machine learning models
- AI model training and inference

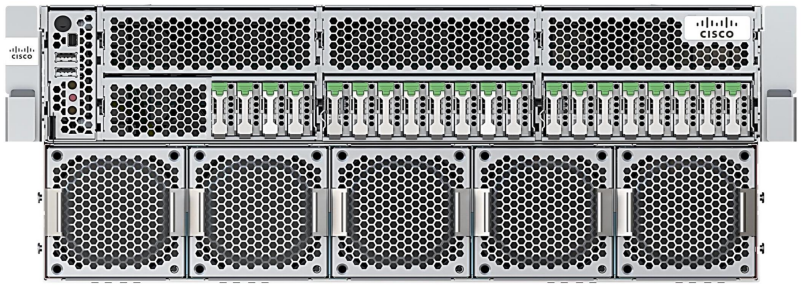
Cisco C845A M8 (MGX) vs. C885A M8 (HGX)

Feature/Aspect	UCS C845A M8	UCS C885A M8
Purpose	Modular GPU Expansion for diverse workloads	High-Performance Computing and AI workloads
Architecture	Modular, flexible design	Integrated, high-density design
Scalability	Highly scalable with modular components	Scalable, but within a more integrated framework
Target Use Cases	Data centers, edge computing, enterprise	Supercomputing, large-scale AI, data analytics
Networking	Support for advanced networking	Integrated high-speed networking capabilities
Deployment Flexibility	Supports diverse configurations	Optimized for specific configurations
Performance Optimization	Customizable for specific workloads	Optimized for maximum performance in HPC and AI
Security Features	Standard enterprise security features	Enhanced security for HPC environments

UCS C845A M8

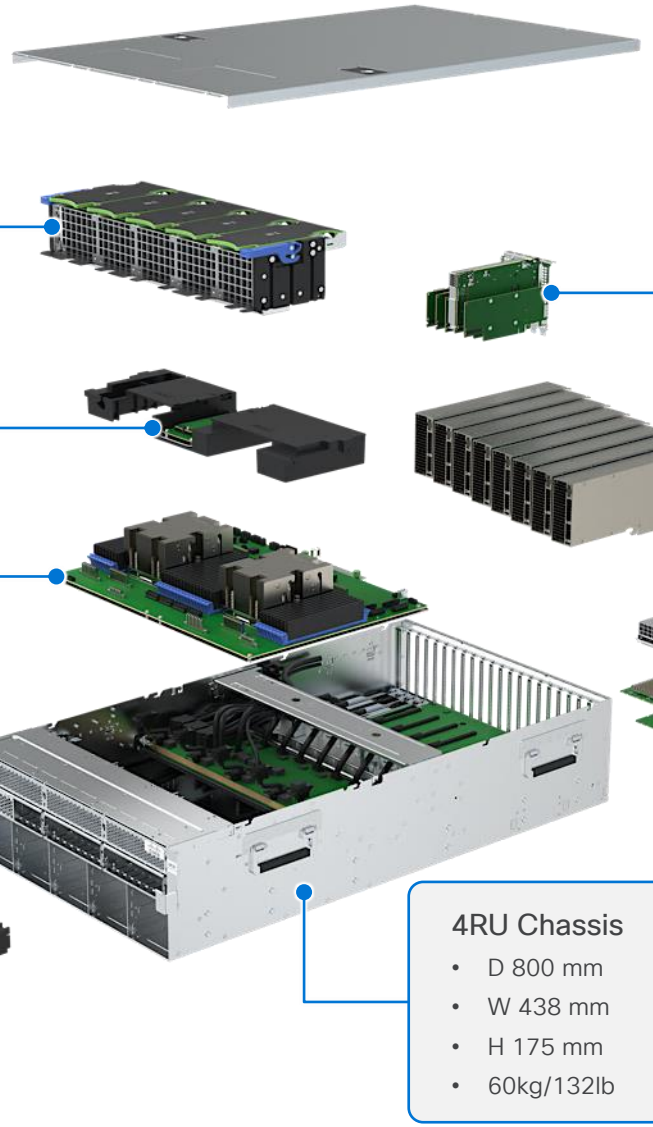
Architecture

UCS C845A M8 Dense GPU Server Specifications



Product Specifications	
Form Factor	• MGX 4RU 19” Rack Server
Compute + Memory	• 2x 5th Gen AMD EPYC CPUs (400W, up to 96 cores, up to 5GHz per core) • Up to 32x DDR5 RDIMM slots (5200 MT/s, 1 DPC - 4400 MT/s, 2 DPC) • 8 Memory Channels per CPU
Storage	• Up to 2 M.2 SATA SSDs for boot (hardware RAID Option Available) • Up to 20 E1.S NVMe SSDs for cache and data
GPUs	• Up to 8x NVIDIA or AMD PCIe GPUs
Network Cards	• 4 PCIe Gen5x16 FHHL for East-West (EW) NIC ConnectX-7 (1x400G) or BF3 B3140H SuperNIC • 1 PCIe Gen5x16 FHHL for North-South (NS) NIC ConnectX-7 (2x200G) , BF3 B3220 • OCP 3.0 X710-DA2 (dual port 10GBaseT) for x86 management
Cooling	• 10 hot-swappable fans for front-to-rear cooling
Front IO	• 2 USB 3.0, 1 ID Button, 1 Power Button, 1 Reset Button, 1 miniDP port
Rear IO	• 1 USB 3.0 A, 1 USB 3.0 C, 1 mDP, 1 ID Button, 1 Power Button, 1 USB 2.0 C (for debugging), 1 RJ45 (mgmt.)
Power Supply	• Mandatory 4x 3200W (AC), Titanium hot-swappable power supplies (N+1)

UCS C845A M8 Modular Design



Upper Cooling/Hot Swap
- 5x 8080 Fan

M.2

Host Processor Module

Storage
- 20x E1.S

Lower Cooling/Hot Swap
- 5x 8080 Fan

PCIe E-W Option
- 4x CX7 (1x400G)
- 4x BF3 B3140H

PCIe N-S Option
- 1x BF3 B3220 (2x200G)
- CX7 (2x200G)
- 1x BF3 B3240 (2x400G) (Double Wide)

GPU Cards
- 8x H100 NVL
- 8x H200 NVL
- 8x L40s
- 8x MI210

Power Supply

OCP 3.0 NIC (RJ45)
- Intel X710-DA2 (2x10GbE)

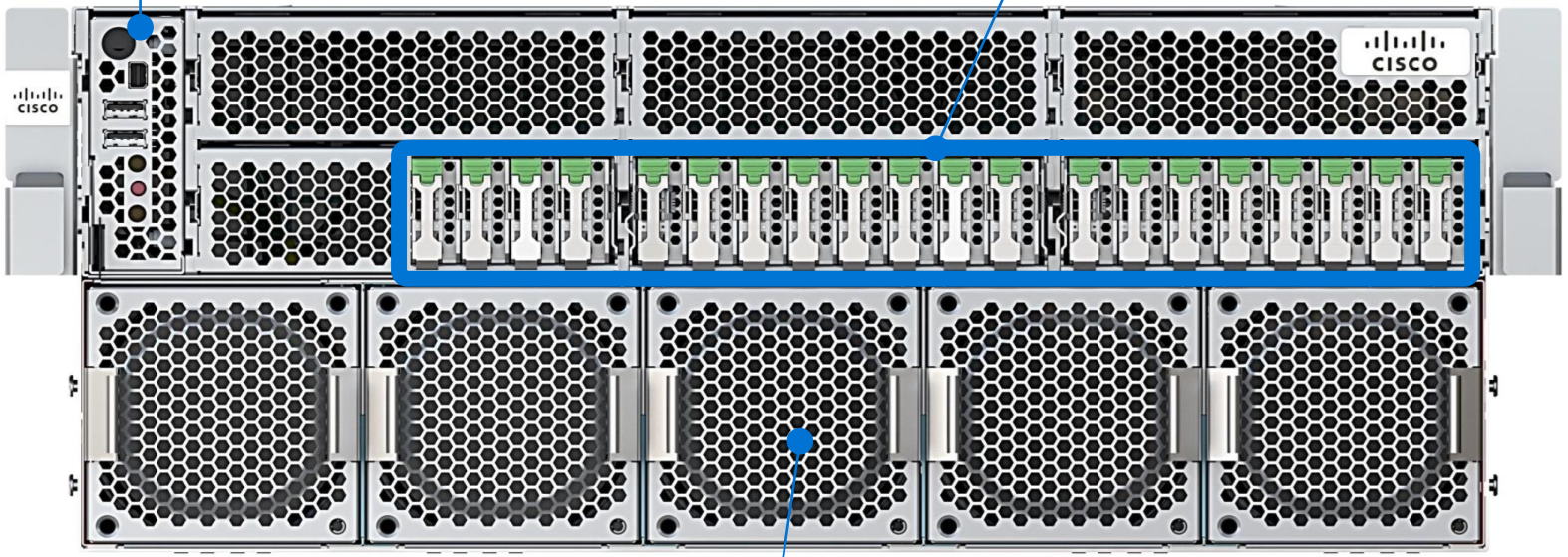
DC-Secure Control Module (BMC)

4RU Chassis
- D 800 mm
- W 438 mm
- H 175 mm
- 60kg/132lb

UCS C845A Front View

Control Panel

20x
Hot-Pluggable E1.S Drives



Hot-Pluggable Fans

CPU

2x

AMD 5th Gen CPUs

Up to 96 cores & up to 5GHz

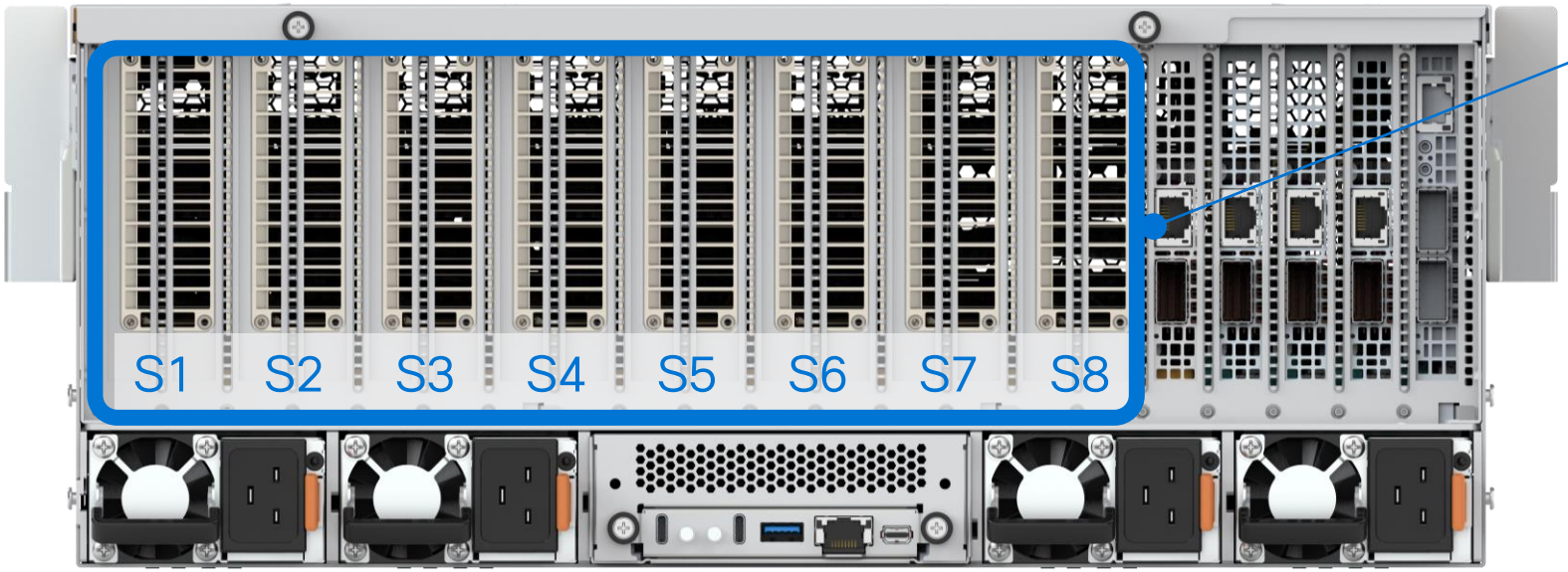
400W/CPU

Memory

Up to 32x

64GB, 96GB or 128GB DDR5 RDIMMs

UCS C845A Rear View



2/4/6/8 x

NVIDIA

Nvidia
H100 NVL,
H200 NVL or L40S GPUs
350W/400W/600W GPUs

or

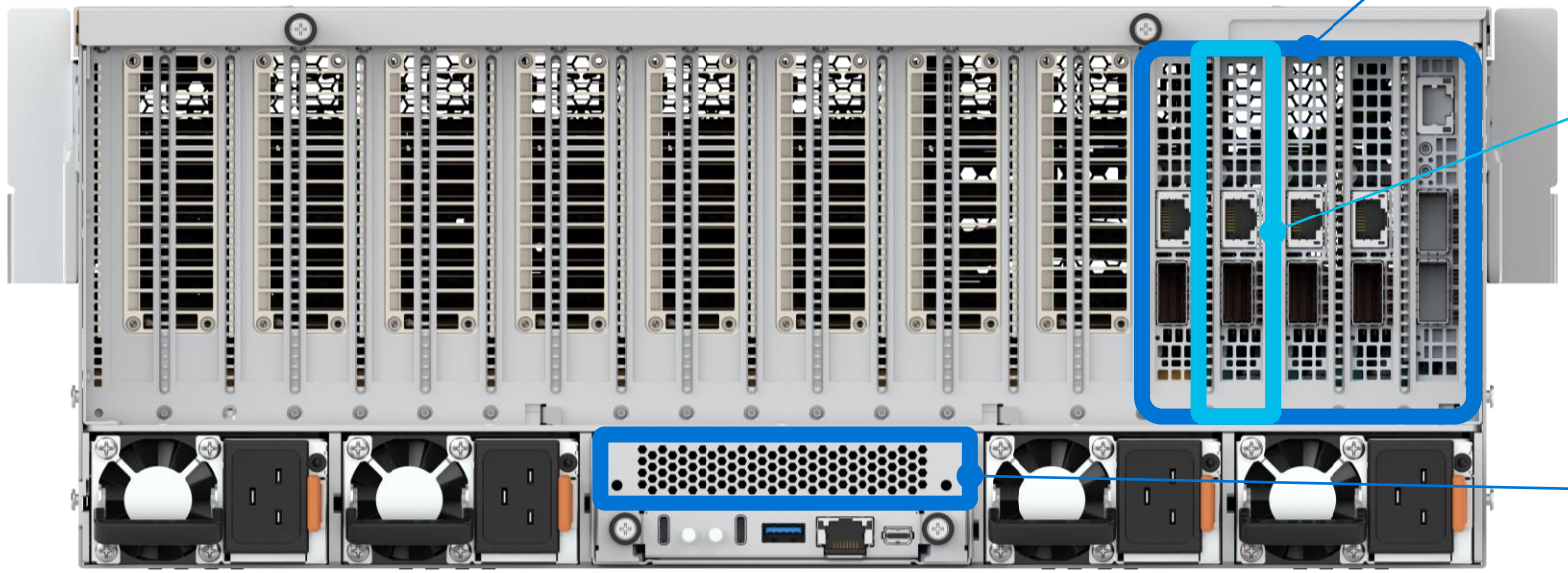
2/4/6/8 x

AMD

AMD
MI210 GPUs
300W/GPU

Note: GPUs cannot be mixed!

UCS C845A Rear View



4x

Nvidia CX-7 or
BlueField-3 3140H

for east-west GPU traffic

1x

Nvidia CX-7 or
BlueField-3 3220

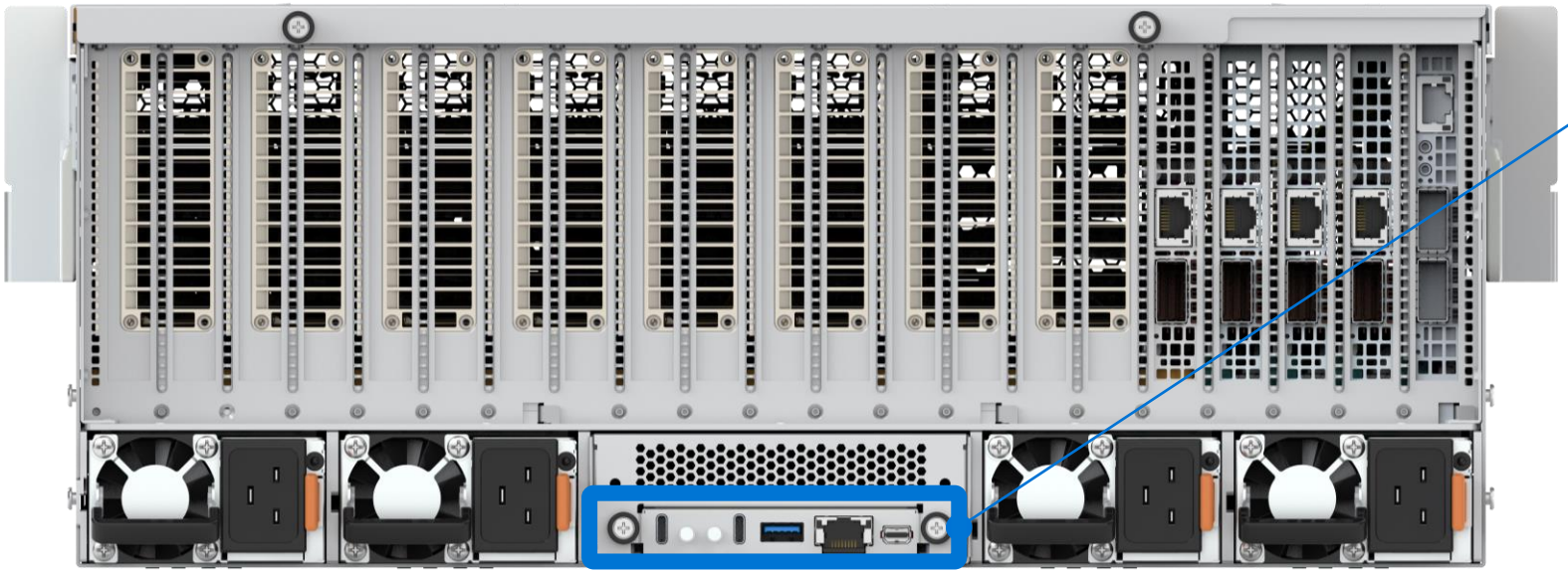
for north-south front-end traffic

1x OCP 3.0

Intel X710

for north-south host management traffic

UCS C845A Rear View

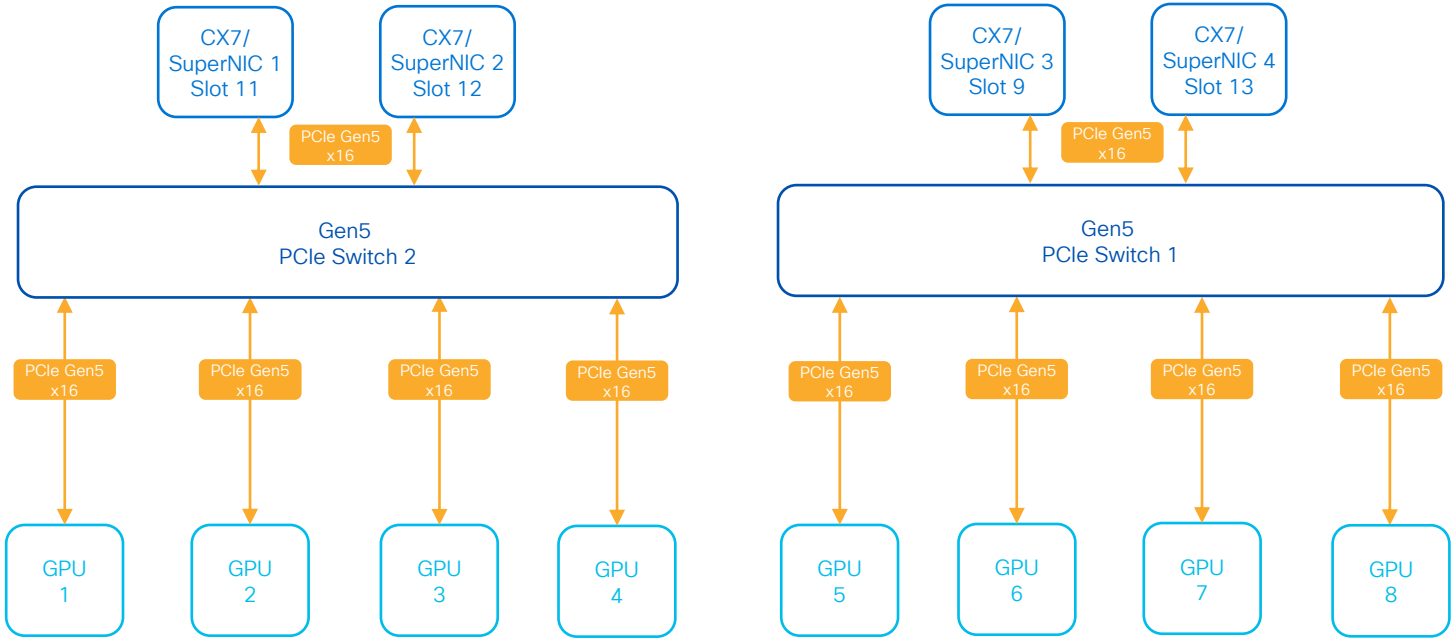


DC-SCM

BMC

for out-of-band server management

UCS C845A – GPU Connectivity



- All GPU cards must be procured from Cisco (unique SBIOS ID required by CIMC)
- Max 8 GPUs, even number, minimum 2, GPUs cannot be mixed
  - 2-Way Bridge
    - Minimum 2 adjacent GPUs per bridge
    - If selected, each 2-way bridge must be fully populated with GPUs
  - 4-Way Bridge
    - Minimum 4 adjacent GPUs per bridge
    - If selected, each 4-way bridge must fully populated with GPUs
- Start with slot 8 for initial GPU is recommended
- Place any additional GPU pairs in adjacent slots
- Communication between GPUs on different PCI Switches will go through CPUs

UCS C845A – PCIe and NVLink Topology

Best NVLink Topology (Recommended):

- Bridge two GPUs under the same CPU or PCIe switch
- GPU count in a system should be in powers of two (1, 2, 4, 8, and so on)
- Locate the same (even) number of GPUs under each CPU socket
- Maintain a balanced configuration: same count of CPU:GPU:NIC for each grouping

Good NVLink Topology:

- Bridge two GPUs under different PCIe switches but under the same CPU
- Same number of GPUs and NICs under each CPU socket, but not powers of two

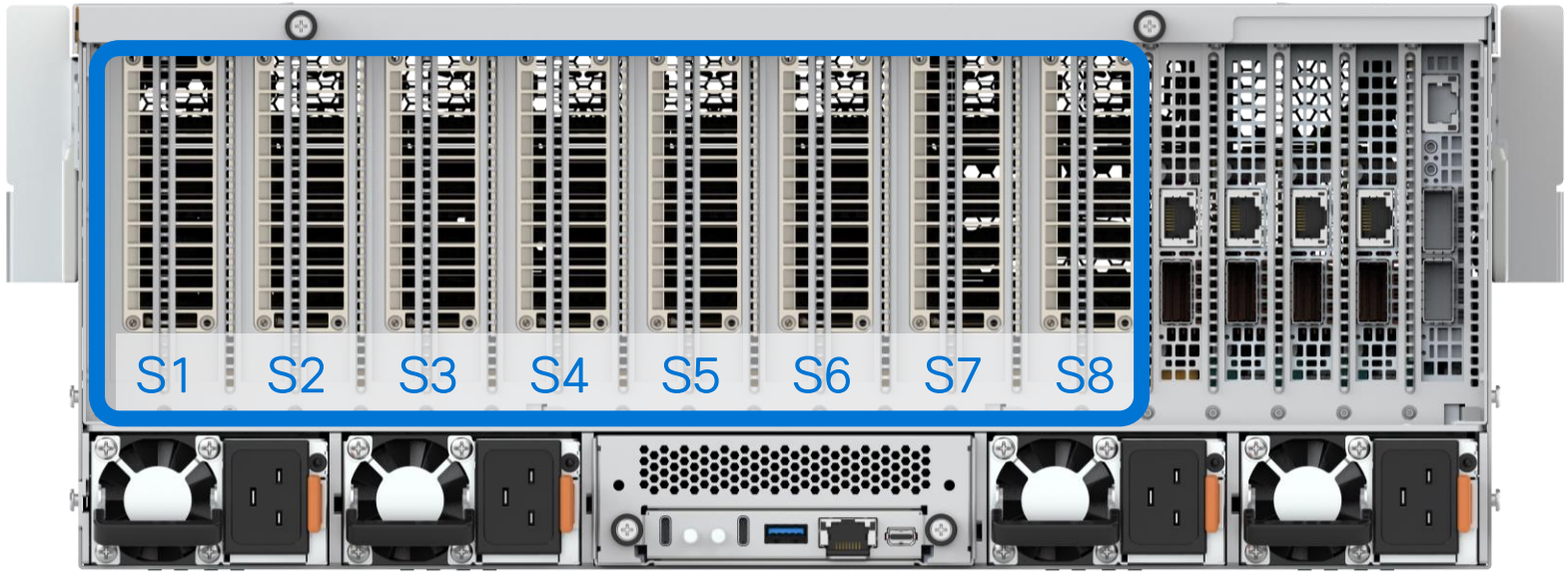
Allowed but Not Recommended:

- Bridge two GPUs under two different CPUs
- Odd number of GPUs under each CPU
- Unbalanced configurations: Different ratios of CPU:GPU:NIC for each grouping

Source: https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/h100/PB-11773-001_v01.pdf

UCS C845A – GPU Population Rules

# of GPUs	Recommended Population (slot #)	Recommended Population
2	(S8, S7)	2 GPUs at GPU Block1
4	(S8, S7, S6, S5)	4 GPUs at GPU Block1
	(S8, S7) (S4, S3)	2 GPUs at GPU Block1; 2 GPUs at GPU Block2
6	(S8, S7, S6, S5) (S4, S3)	4 GPUs at GPU Block1; 2 GPUs at GPU Block2
8	(S8, S7, S6, S5) (S4, S3, S2, S1)	4 GPUs at GPU Block1; 4 GPUs at GPU Block2



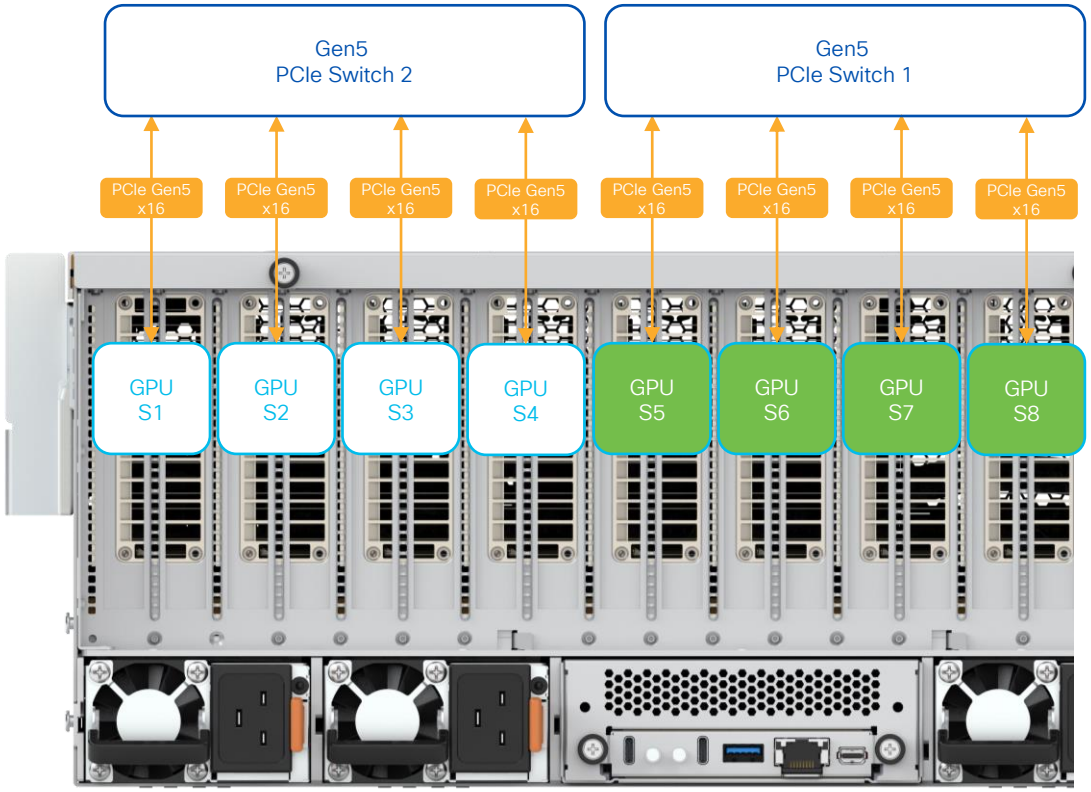
UCS C845A – GPU Population Rules

When NVL Bridge or Infinity Bridge Selected

# of GPUs	Recommended Population (slot #)	NVL Bridge	Infinity Bridge	Recommended Population
2	(S8, S7)	1x NVL2	1x 2-way	1x 2-way bridge for GPU Block1
4	(S8, S7, S6, S5)	2x NVL2	2x 2-way	2x 2-way bridge for GPU Block1
		1x NVL4	1x 4-way	1x 4-way bridge for GPU Block1
	(S8, S7) (S4, S3)	2x NVL2	2x 2-way	1x 2-way bridge for GPU Block1 and 1x 2-way bridge for GPU Block2
6	(S8, S7, S6, S5) (S4, S3)	3x NVL2	3x 2-way	2x 2-way bridge for GPU Block1 and 1x 2-way bridge for GPU Block2
		1x NVL2 + 1x NVL4	1x 2-way + 1x 4-way	1x 4-way bridge for GPU Block1 and 1x 2-way bridge for GPU Block2
8	(S8, S7, S6, S5) (S4, S3, S2, S1)	4x NVL2	4x 2-way	2x 2-way bridge for GPU Block1 and 2x 2-way bridge for GPU Block2
		2x NVL4	2x 4-way	1x 4-way bridge for GPU Block1 and 1x 4-way bridge for GPU Block2

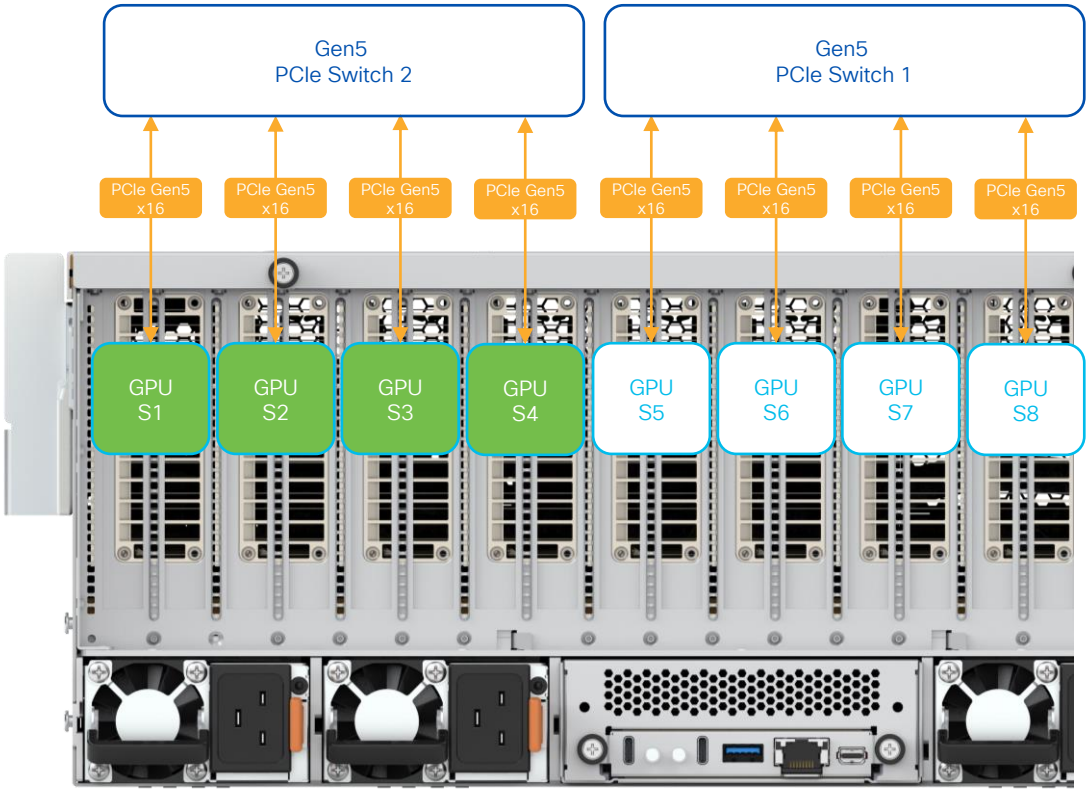
Note: MI210 supports both 2-way and 4-way infinity fabric bridge.
H100 supports NVL2 only. H200 supports NVL2 or NVL4. L40s does not support NVL

UCS C845A – GPU Population Examples – H100



Recommended

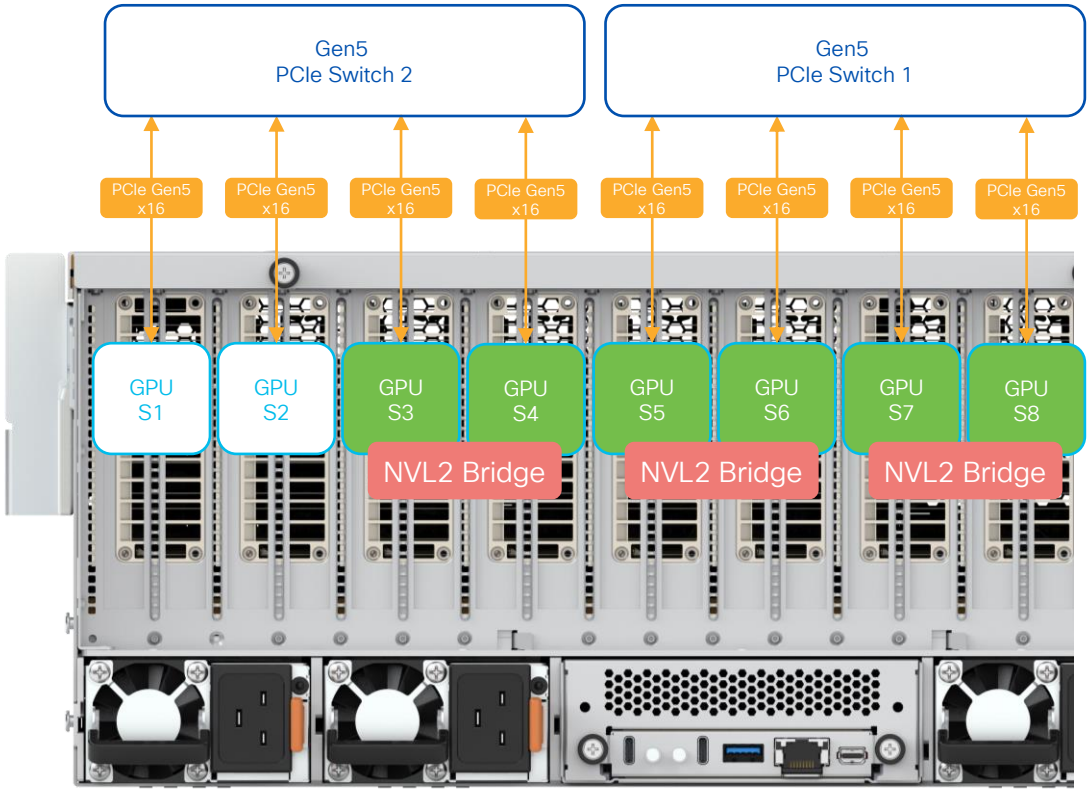
(Note: S8, S7, S4, S3 is also ok but comms among GPUs on different PCI Switches will be going through CPU)



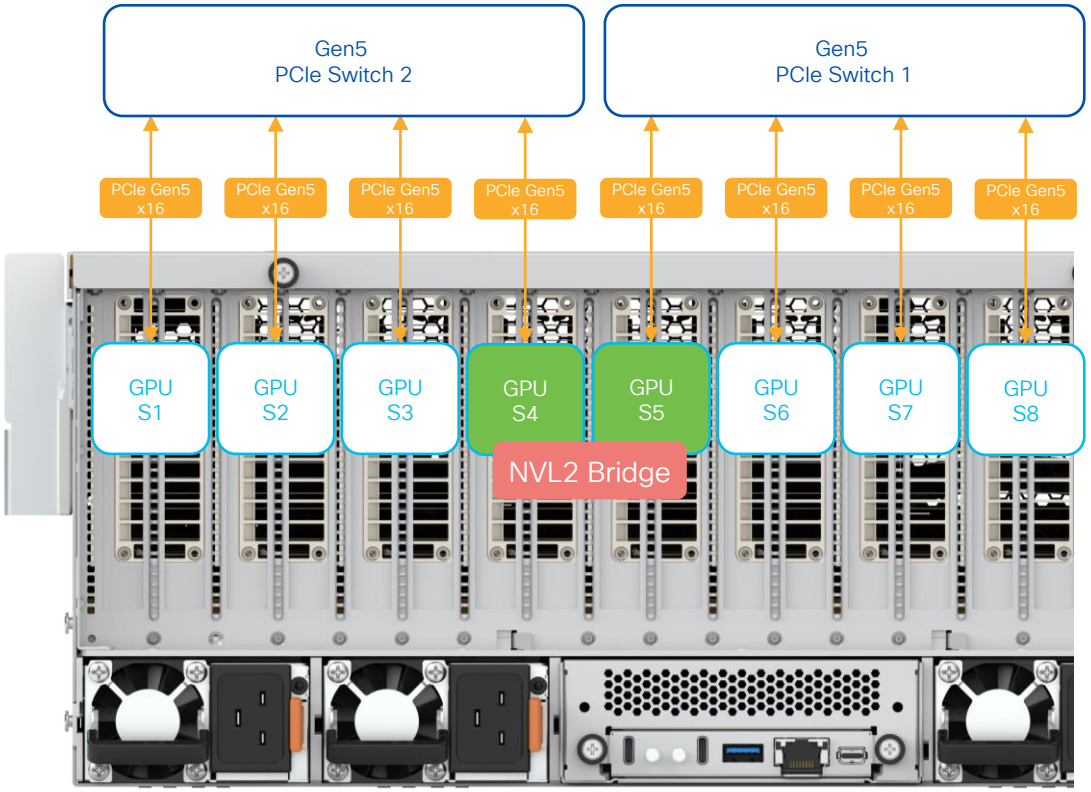
Not Recommended

Reason: should start from S8

UCS C845A – GPU Population Examples – H100



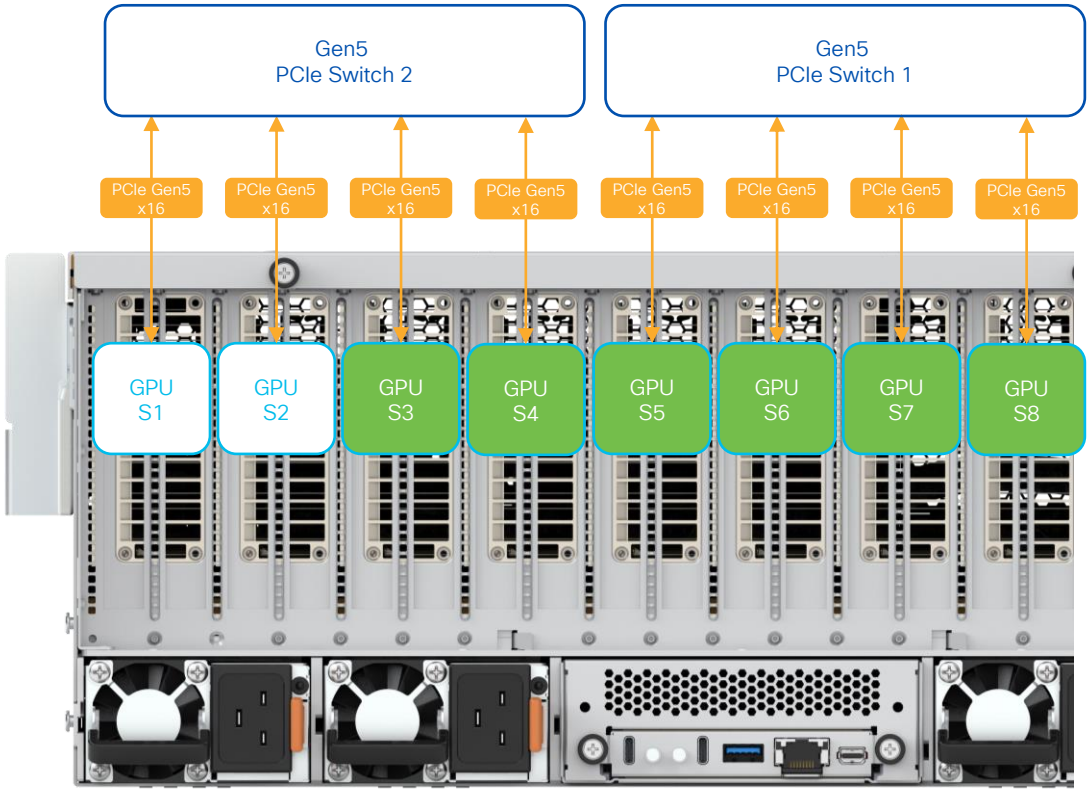
Recommended



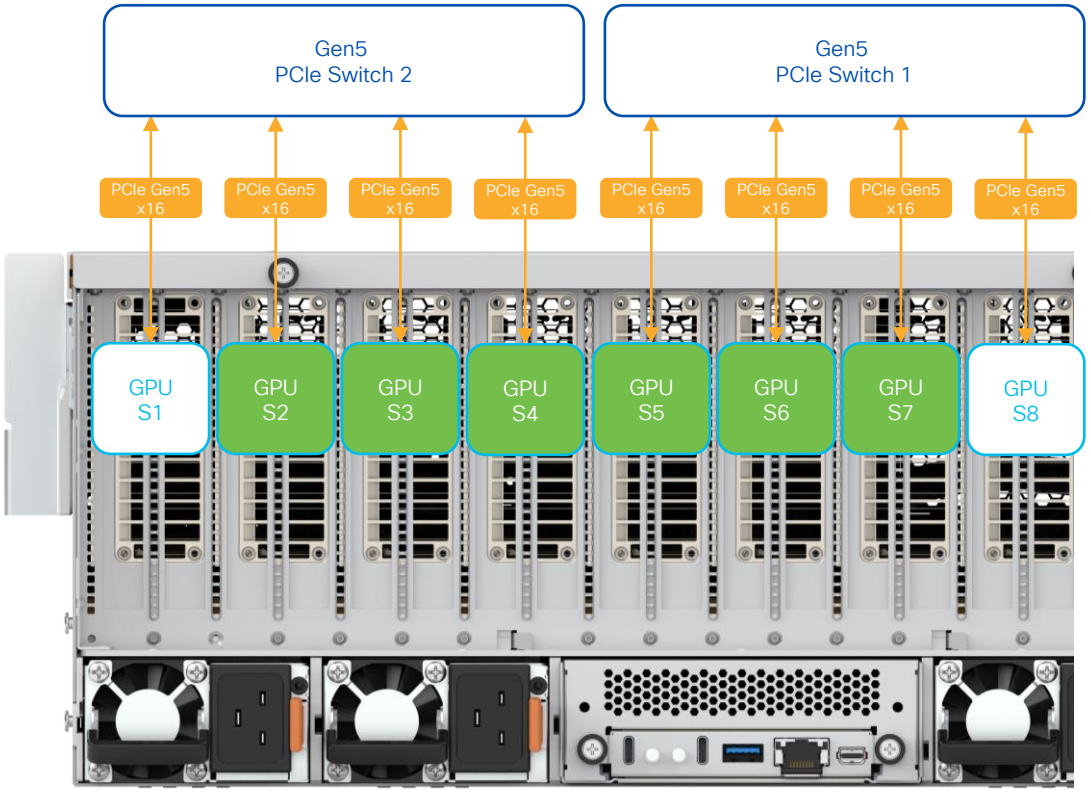
Not Recommended

Reason: bridging GPUs under two different CPUs

UCS C845A – GPU Population Examples – H200



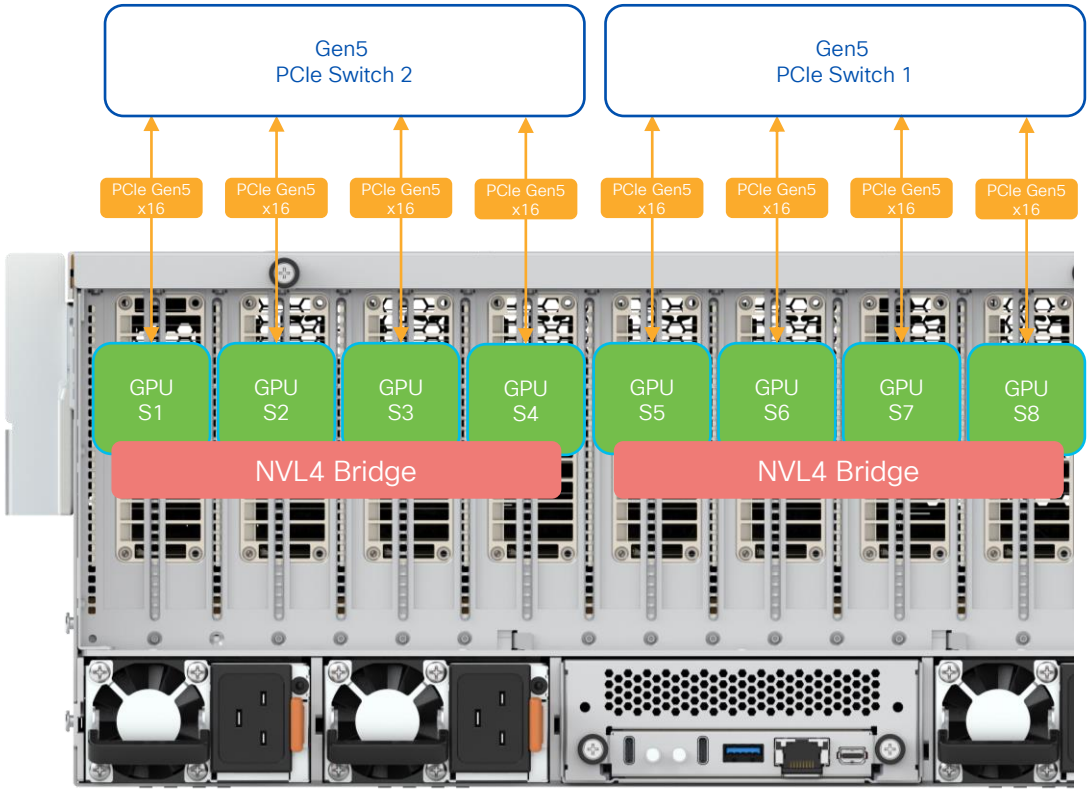
Recommended



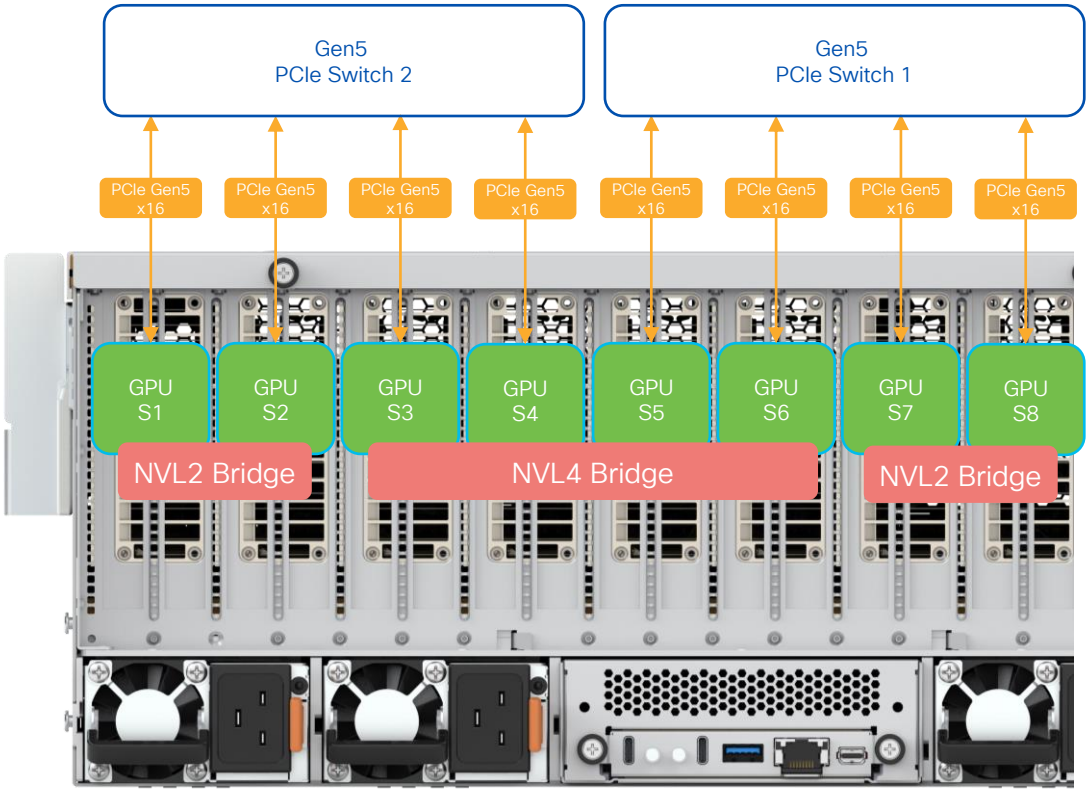
Not Recommended

Reason: should start from S8, odd number of GPUs per block

UCS C845A – GPU Population Examples – H200



Recommended

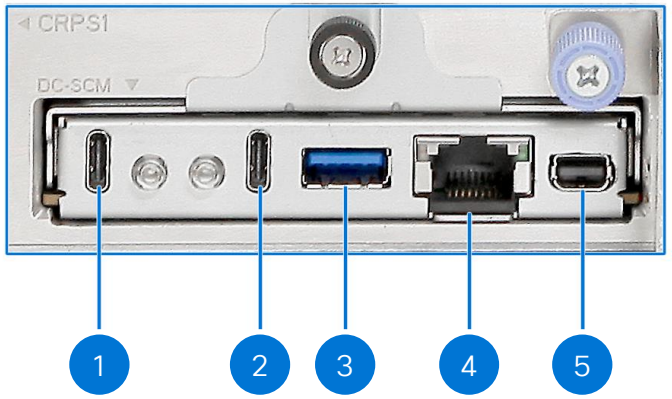


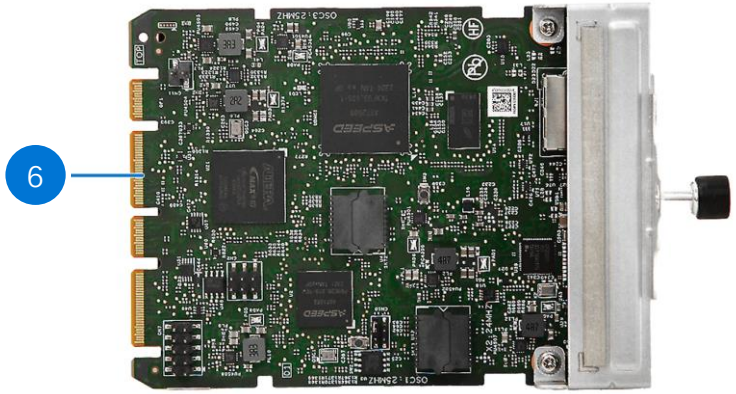
Not Recommended

Reason: bridging GPUs under different CPUs

UCS C845A M8 Data Center Secure Control Module

DC-SCM*





1: 1x USB-C debug port

2: 1x USB-C 3.2 port

3: 1x USB-A 3.2 port

4: 1x RJ45 Management ethernet port

5: 1x Mini display port

6: DC-SCM top view

*The DC-SCM is where the BMC and RoT reside

Note: Recommended AST DP with native DP monitor for optimum performance. If an external adaptor is needed, use active type adaptor only. This module does not support DP++ type adapter

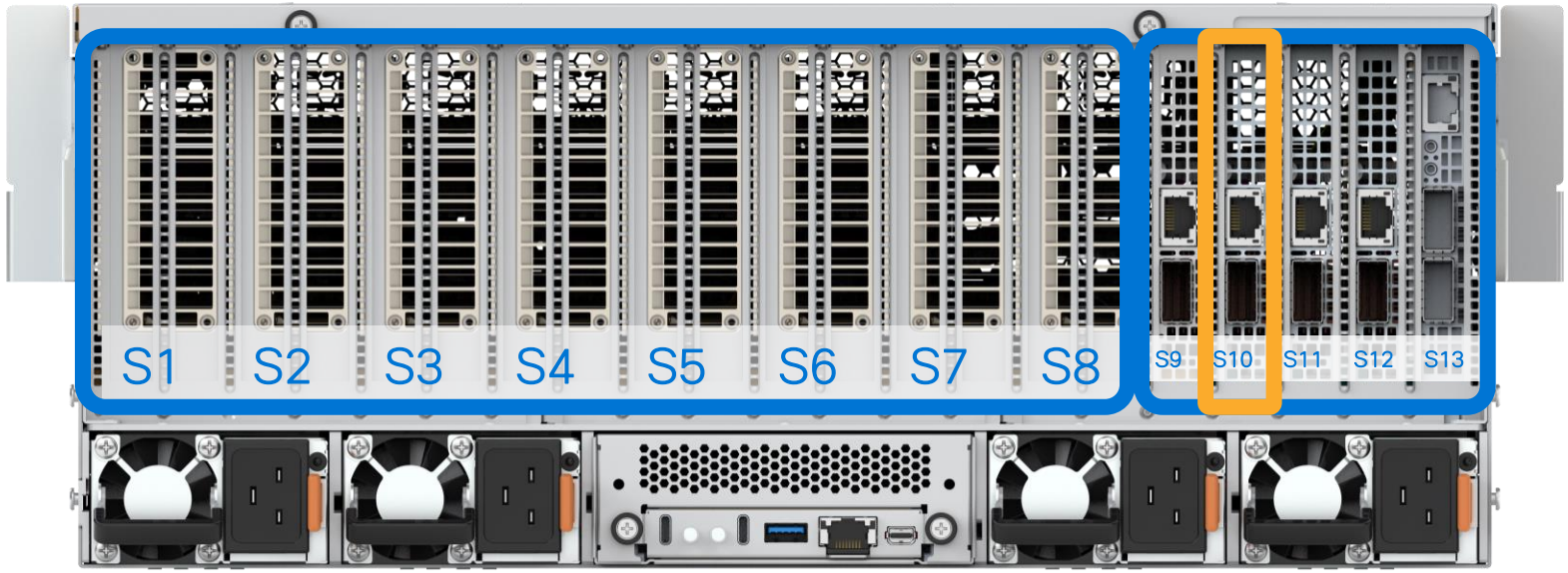
UCS C885A M8

Networking

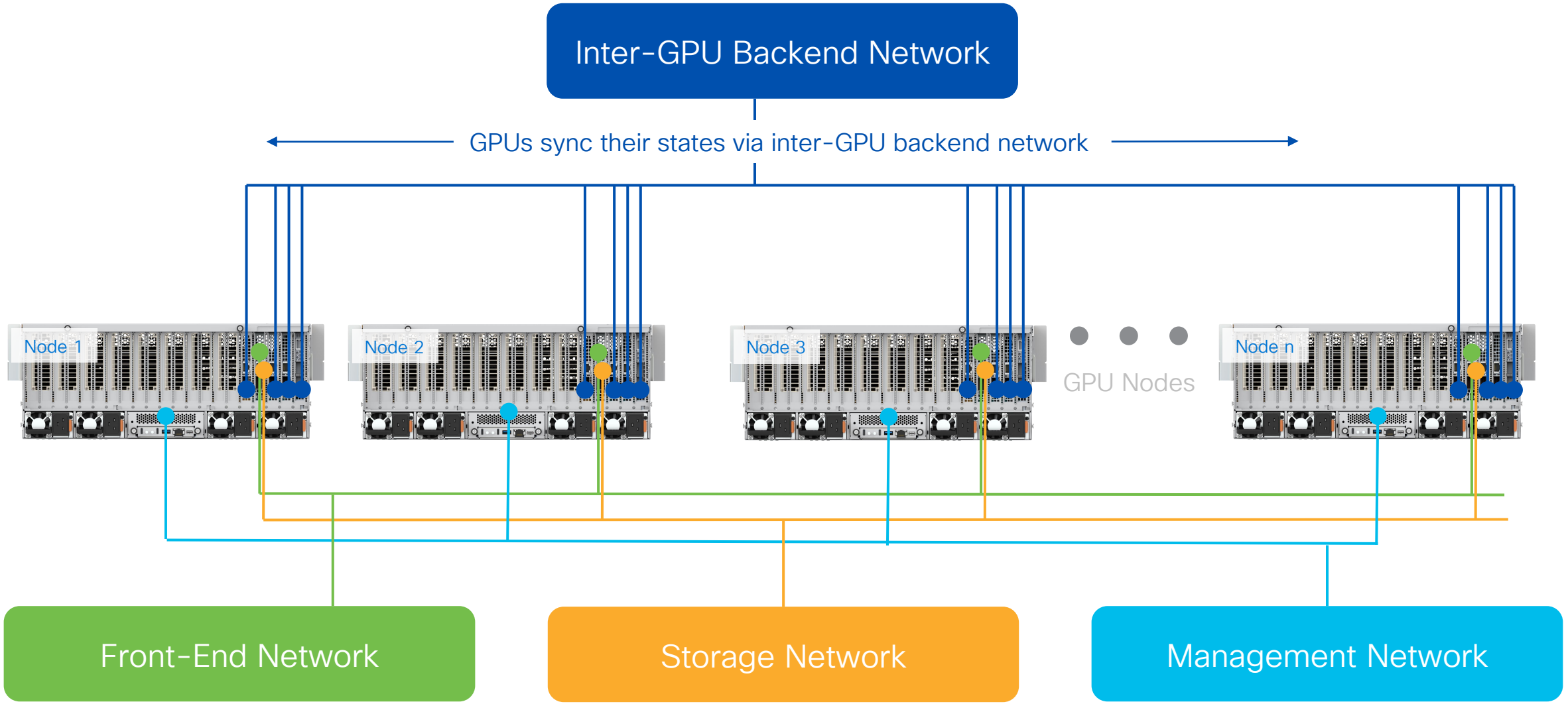
UCS C845A – NIC Population – Single Wide NS

B3220 / CX7

# of GPUs	Recommended GPU Pop.	# E/W NICs	Recommended E/W Pop.	Recommended NS Population
2	(S8, S7)	1	S9 or S13	S10
4	(S8, S7, S6, S5)	2	S9 and S13	S10
	(S8, S7) (S4, S3)	2	S9 and S12	S10
6	(S8, S7, S6, S5) (S4, S3)	3	S9, S13 and S12	S10
8	(S8, S7, S6, S5) (S4, S3, S2, S1)	4	S9, S13, S11 and S12	S10



UCS C845 M8 Networking Blueprint



UCS C845 M8 Network Definitions

Multiple networks of an AI/ML Infrastructure…

- Inter-GPU backend network: An Inter-GPU backend network connects the dedicated GPU ports for running distributed training. This network is also known as the back-end network, compute fabric, or scale-out network.

- Front-end network: A front-end network connects the GPU nodes to the data center network for inferencing, logging, managing in-band devices, and so on.

- Storage network: A storage network connects the GPU nodes to the shared storage devices providing parallel file system access to all the nodes for loading (reading) the data sets for training, and checkpointing (writing) the model parameters as they are learned. Some users may share the front-end network to connect storage devices, eliminating a dedicated storage network.

- Management network: A management network provides out-of-band connectivity to the devices of the AI/ML infrastructure, such as GPU nodes, network switches, and storage devices.

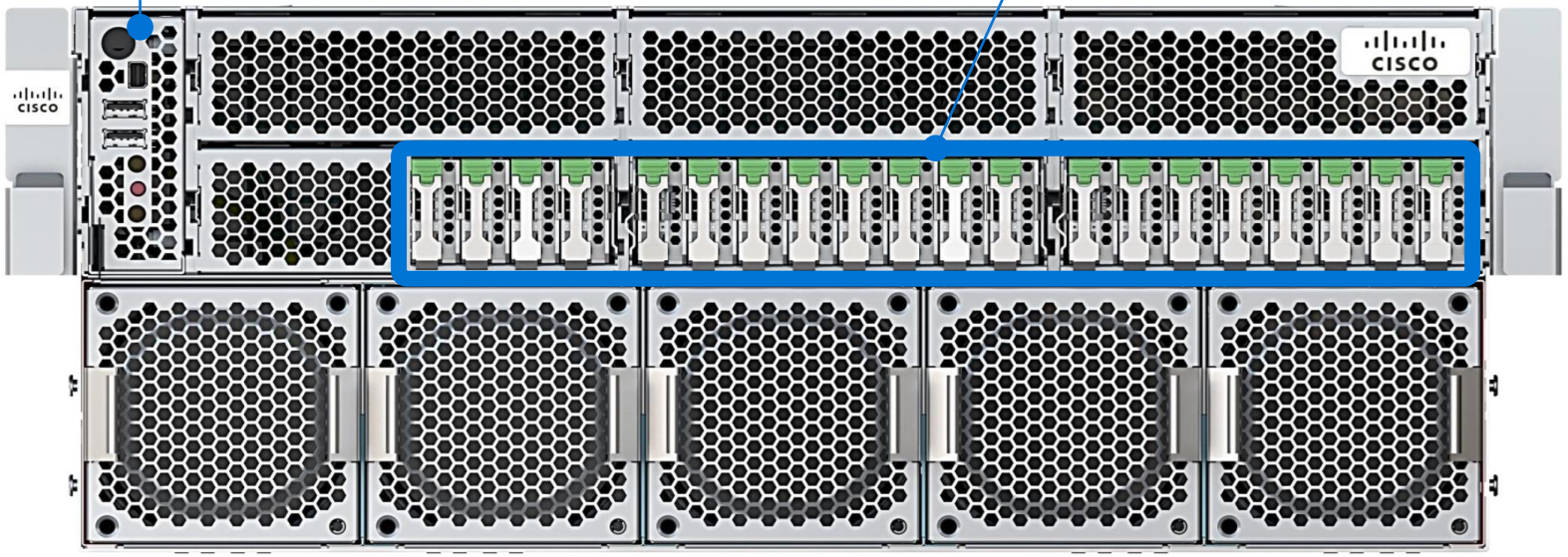
UCS C885A M8

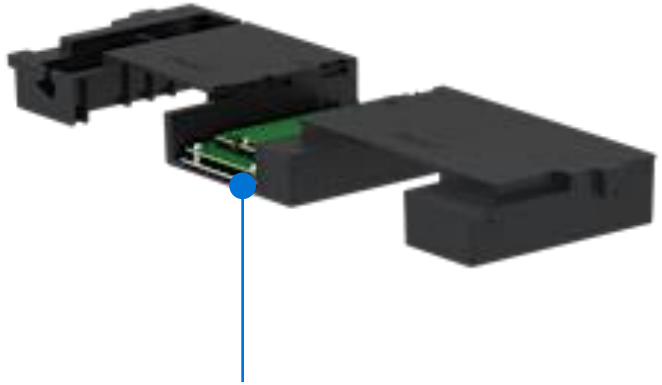
Storage

UCS C845A Local Storage

Control Panel

20x
Hot-Pluggable
E1.S Drives

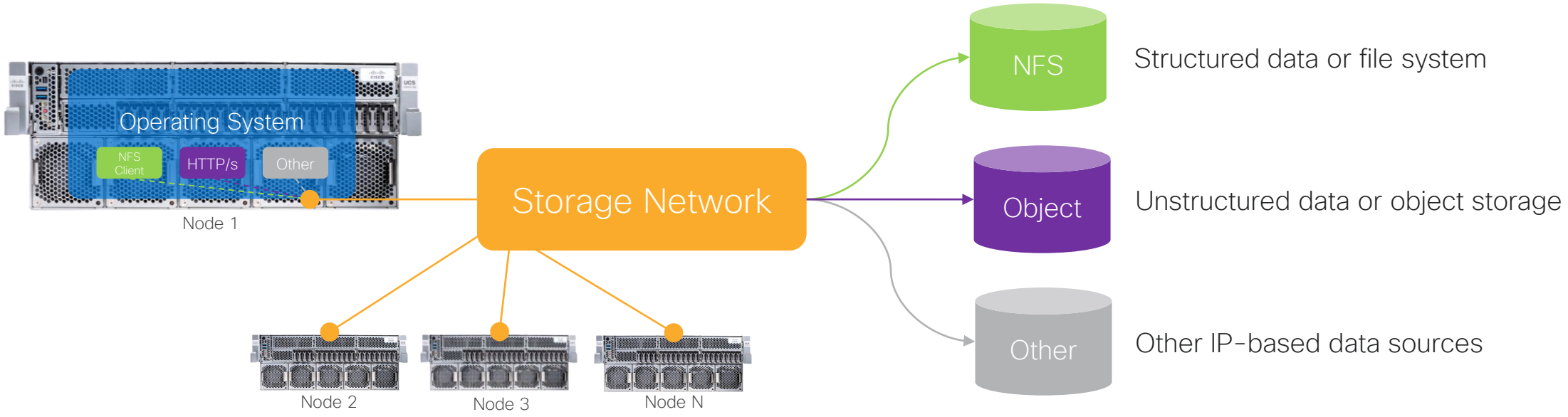




Up to 2x
Boot M.2 SATA SSDs
Optional RAID Controller
240GB and 960GB
(UEFI mode only)

UCS C845A M8 External Storage

- Support for standard IP-based storage protocols leveraging OS/software clients
- Common examples for AI/ML training use-cases include NFS and object storage



Note: Most customers leverage a dedicated storage network, but storage traffic can be aggregated with other north-south networks as well

UCS C885A M8

Power Usage

C845A M8 Power Consumption Examples

This slide provides rough estimates that might be shared verbally with a customer as of May 7 2025.
Read before you share!

GPU	GPU Qty.	E/W NICs	NS NICs	SSDs	Power Estimate
H100	2	1	1	16	5977 W
H100	8	4	1	20	8854 W
H200	2	1	0	6	5610 W
H200	4	2	1	6	7154 W
H200	8	4	1	20	10276 W
L40S	4	0	0	16	6433 W
L40S	8	4	1	16	8298 W
MI210	2	1	1	20	4928 W
MI210	4	2	1	20	5628 W
MI210	8	4	1	20	7027 W

Note: Rough estimates as there are too many possible parts combinations and use cases.
Official info will come on the power calculator.

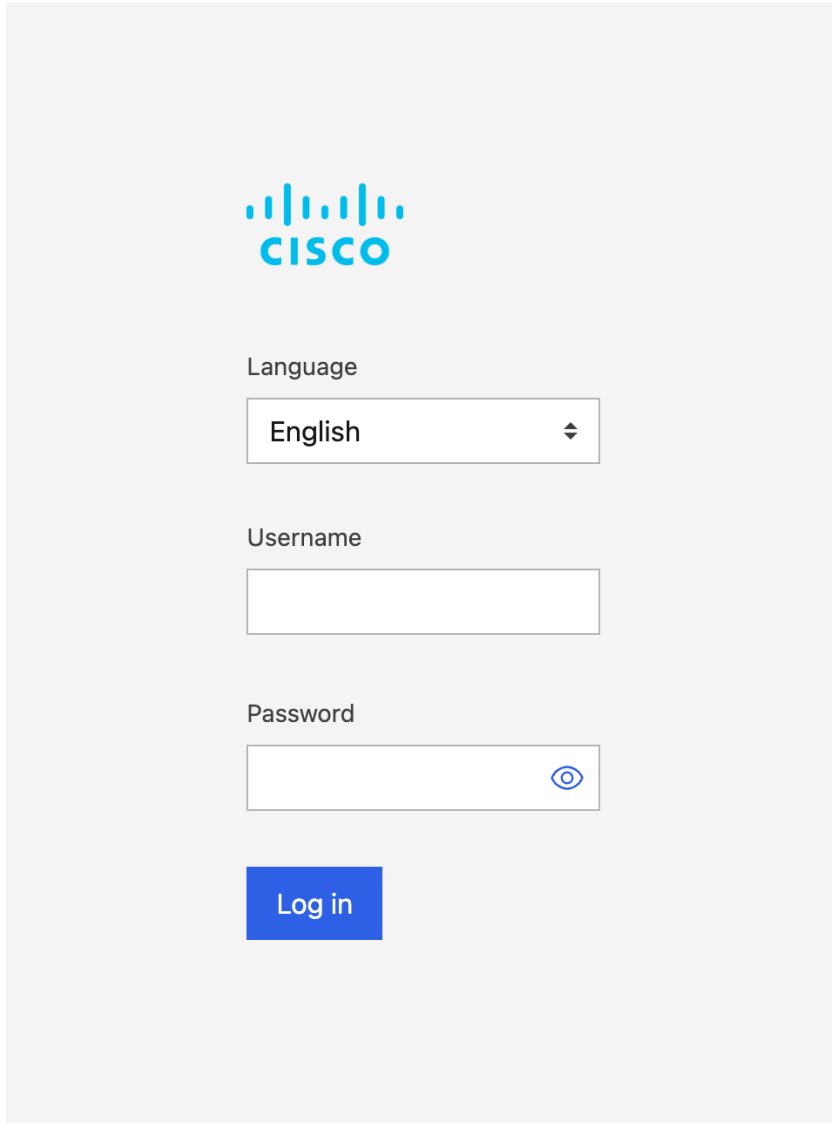
UCS C885A M8

Management

UCS C845A M8 Management from Local BMC

Local BMC Server Operations

- Power Operations
- System Inventory
- Event Logs
- KVM Launch and KVM Session management
- Monitoring
- Boot order configuration
- System Firmware management
- User management and certificates



UCS C845A M8 Intersight Support

At FCS (Q3 CY25)

- Claim UCS C845A M8 Server in Intersight via the Device Connector on the BMC.
- Inventory Display: Show hardware details in Intersight including – CPU, GPU, Memory, NVMe Drives, and Network Cards (Intel, NVIDIA CX-7 and Bluefield 3)
- Health and Alerts Management: Monitor health, alerts, alarms, and enable Online Insert Removal (OIR).
- KVM Cross Launch
- Power Management: Power On/Off, Power Cycle, Hard Reset
- OS Installation via native KVM
- Connected TAC Support: Ensure support for technical assistance.
- Monitoring: Limited hardware metrics

Post FCS – Target Q4 CY25

- Real time Inventory Updates
- Tunneled KVM & Session Management
- Firmware Management including GPU from Intersight
- Advanced Hardware Monitoring capacities and metrics including GPU monitoring
- Hardware Compatibility List (HCL)
- Advisories
- Automated OS Installation
- Policies for Boot, BIOS, Network Connectivity
- Server Profile Management

Certification of Supported Operating Systems

- Ubuntu Server 22.04
- RedHat Enterprise Linux CoreOS 4.16
- Rocky Linux 9.5
- RedHat Enterprise Linux 9.4
- Ubuntu Core 24.04

UCS C885A M8

Orderability

Cisco UCS C845A M8 FCS Options

	Component	PID
Chassis	**UCS C845A M8 base server** - w/o CPU, mem, drives	CAI-845A-M8
CPUs (Select 2/No CPU Mixing)	**AMD 9575F** 3.3GHz 400W 64C/256MB Cache DDR5 6000MT/s	CAI-CPU-A9575F
	AMD 9475F 3.65GHz 400W 48C/256MB Cache DDR5 6000MT/s	CAI-CPU-A9475F
	AMD 9375F 3.85GHz 320W 32C/256MB Cache DDR5 6000MT/s	CAI-CPU-A9375F
	AMD 9655 2.6GHz 400W 96C/384MB Cache DDR5 6000MT/s	CAI-CPU-A9655
	AMD 9555 3.2GHz 360W 64C/256MB Cache DDR5 6000MT/s	CAI-CPU-A9555
	AMD 9455 3.15GHz 300W 48C/256MB Cache DDR5 6000MT/s	CAI-CPU-A9455
	AMD 9355 3.55GHz 280W 32C/256MB Cache DDR5 6000MT/s	CAI-CPU-A9355
Memory (no DIMM size mixing/Select up to 32)	**64GB** DDR5-6400 RDIMM 2Rx4 (16Gb)	CAI-MRx64G2RE5
	96GB DDR5-6400 RDIMM 2Rx4 (24Gb)	CAI-MRx96G2RF5
	128GB DDR5-6400 RDIMM 2Rx4 (32Gb)	CAI-MR128G2RG5
GPUs (Select 2, 4, 6, 8 GPUs)/ NVL/Infinity Bridges	NVIDIA **L40S**: 350W, 48GB, 2-slot FHFL GPU	CAI-GPU-L40S
	NVIDIA **H100 NVL**:400W, 94GB, 2-slot FHFL GPU	CAI-GPU-H100-NVL
	NVIDIA **NVL-2 Way Bridge for H100 GPU**	CAI-NVL2-H100
	NVIDIA OEM **H200-NVL** GPU 600W, 141GB, 2-slot FHFL	CAI-GPU-H200-NVL
	NVIDIA **NVL-2 Way Bridge for H200 GPU**	CAI-NVL2-H200
	NVIDIA **NVL-4 Way Bridge for H200 GPU**	CAI-NVL4-H200
	AMD Instinct **MI210**: 300W, 64GB, 2-slot FHFL GPU	CAI-GPU-MI210
	AMD Infinity 2-Way Bridge for the AMD MI210 GPU	CAI-INF2-MI210
	AMD Infinity 4-Way Bridge for the AMD MI210 GPU	CAI-INF4-MI210
E1.S Drives (Select 0-20)	**1.9TB E1.S** 15mm Kioxia XD7P Hg Perf Med End **Gen4** 1X NVMe	CAI-NVES1T9K1V
	3.8TB E1.S 15mm Kioxia XD7P Hg Perf Med End **Gen4** 1X NVMe	CAI-NVES3T8K1V
	7.6TB E1.S 15mm Kioxia XD7P Hg Perf Med End **Gen4** 1X NVMe	CAI-NVES7T6K1V
M.2 Boot Drives (0/2)/HW RAID for Boot (0/1)	**240GB M.2**	CAI-M2-240G
	960GB M.2	CAI-M2-960G
	Cisco Boot optimized **M.2 HW Raid Controller**	CAI-M2-HWRAID
NICs/DPUs (Select 1 for N/S traffic) and/or Intel X710-DA2	NVIDIA **CX-7 (2x200)** (N/S Traffic)	CAI-P-N7D200GFO
	NVIDIA **BF-3 B3220** (2x200G) (N/S Traffic)	CAI-P-N3220
	Intel X710-DA2 (2x10GbE) (RJ45 OCP 3.0)	CAI-O-ID10GC
NICS/DPUs (Select 0 - 4 for E/W Traffic) No mixing	NVIDIA **CX-7** (1x400) E/W Traffic	CAI-P-N7S400GFO
	NVIDIA **BF-3 B3140H** (E/W Traffic)	CAI-P-N3140H
Accessories (Select 0 or 1)	Reversible CMA for C845A M8 ball bearing rail kit	CAI-845A-CMA
PSU (Must Select 4)	UCS C845A M8 3200W AC Power Supply Titanium	CAI-845A-PSU

Cisco UCS C845A M8 Spare Parts

Component	PID
Front Facing Fan	CAI-845A-FAN=
Internal Fan	CAI-845A-FANI=
CS-SCM (BMC)	CAI-845A-DCSCM=
Reversible CMA for C845A M8 ball bearing rail kit	CAI-845A-CMA=
UCS C845A Ball Bearing Rail Kit	CAI-845A-RAIL=
UCS C845A GPU Accessories	CAI-BRK-GPU=
UCS C845A Bluefield 3 Power Cable N-S NIC	CAI-CBL-BF3-N-S=
UCS C845A AMD MI210 GPU Power Cable	CAI-CBL-GPU-A=
UCS C845A NVIDIA GPU Power Cable	CAI-CBL-GPU-N=
UCS C845A E1.S Blanking Panel	CAI-E1S-BLANK=
UCS C845A-GPU/NVLINK Bridge Removal Tool Kit AMD	CAI-GPU-RKIT-AMD=
UCS C845A-GPU/NVLINK Bridge Removal Tool Kit NVIDIA	CAI-GPU-RKIT-NV=
UCS C845A Heat Sink	CAI-HS-C845A=
AMD Infinity-2way Bridge for MI210 GPU	CAI-INF2-MI210=
AMD Infinity-4way Bridge for MI210 GPU	CAI-INF4-MI210=

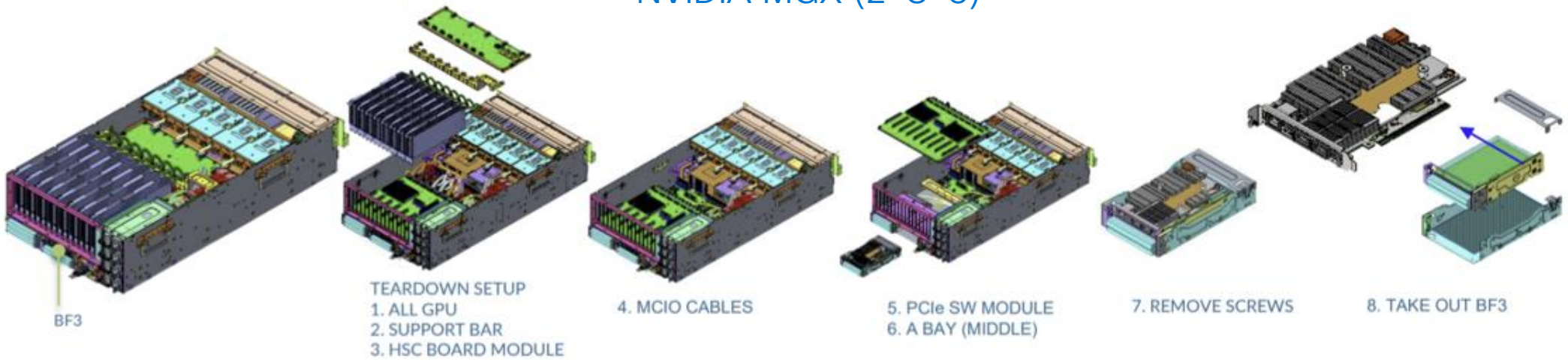
UCS C845A M8

Support and Professional Services

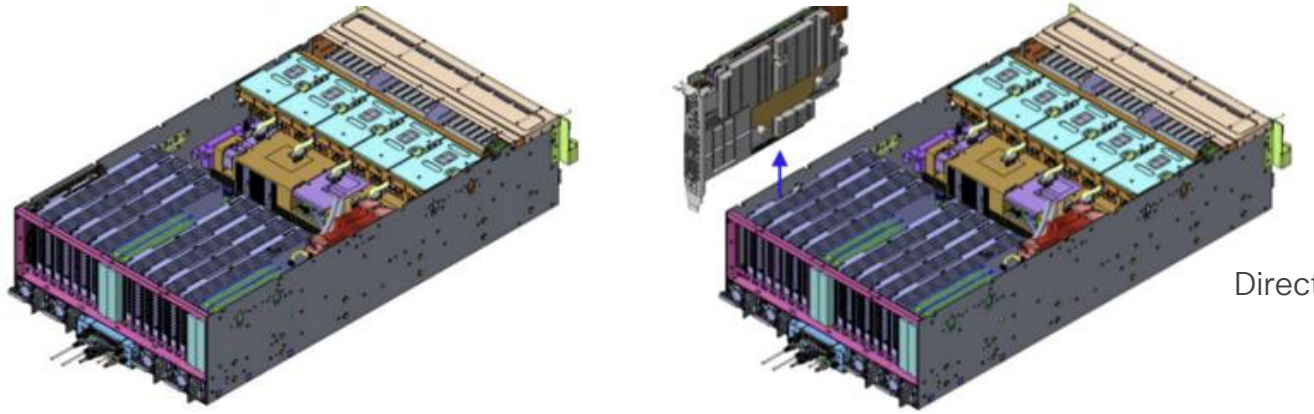
The UCS C845A M8 – A Better Overall Design

PCIe Card Serviceability Example

NVIDIA MGX (2-8-5)



Cisco UCS C845A M8 (2-8-5)



Directly Remove BF3

UCS C845A M8 Technical Support Service

Key Support Services Elements

- UCS C845A M8 **Mandatory Support required**
  - *FCS customers are able to 'opt out'*
- UCS C845A M8 **Success Tracks (ST) L1 and L2** support options **ONLY**
  - *Full suite of RMA **options** (inclusive of 24x7x3) available*
  - *Drive Retention is also available*
- UCS C845A M8 **FE Onsite Support** options **ONLY**
  - *'No Onsite' options will **NOT** be offered*

For more detailed information on the ST deliverables, please refer to the Legal Service Description at https://www.cisco.com/c/dam/en_us/about/doing_business/legal/service_descriptions/docs/cisco-success-tracks.pdf.

Cisco Solution Attached Services for AI Optimized UCS Servers

Accelerating your Cisco AI Optimized UCS Server adoption →

Solution Design and Architecture

- Conduct a design workshop with your AI and infrastructure teams to understand the design requirements for Cisco AI Optimized UCS Servers to meet your training or inferencing needs.

Deployment and Configuration

- Configure AI Optimized UCS Server management
- Set up the AI Optimized UCS Server infrastructure with NVIDIA or AMD GPUs and update firmware
- Install the operating systems, GPU drivers, AI software stack (Python, PyTorch, etc), and integrate with backend storage
- Configure server OS, network connectivity, and the AI stack software as required
- Use automation tools like Ansible for efficient operations
- Connect Server(s) to the upstream network

Testing and Validation

- Validate AI Optimized UCS Infrastructure stack including GPU drivers, and backend storage integration
- Validate inter-GPU network bandwidth performance (if applicable)
- Ensure compliance with Cisco and partner best practices

Optimization and Support

- Guide/recommend AI Optimized UCS Server infrastructure fine-tuning (operating systems and drivers)

Knowledge Transfer

- Provide knowledge transfer sessions on AI Optimized UCS Server infrastructure implementation

UCS C885A M8

Supplemental Material

UCS C845A M8 Supplemental Material

AI-Ready Data Center Hub on SalesConnect for the UCS C845A M8 **BDM Presentation**, **At-A-Glance**, **Data Sheet**:
SalesConnect

Cisco UCS C845A M8 Spec Sheet:
Cisco UCS C845A M8 AI Server Spec Sheet

Cisco UCS C845A M8 AI Server Memory Spec Sheet:
Cisco UCS C845A M8 AI Server Memory Spec Sheet

Cisco UCS C885A M8 Ordering Guide:
Ordering Guide

Blog Post:
Blog Post

Press Release at CiscoLive Amsterdam:
Cisco Showcases Simplicity, Security and AI Readiness at Cisco Live Amsterdam

BDM Deck: C845A BDM Deck

Partner BDM Deck: C845A Partner BDM Deck


CISCO