Check-M8!

Changing the AI Game with the latest UCS Innovations

Greg Wilkinson
Technical Marketing Engineering
Technical Leader

Bryan Hilton
Technical Marketing Engineering
Technical Leader

BRKCOM-1047

Cisco Webex App

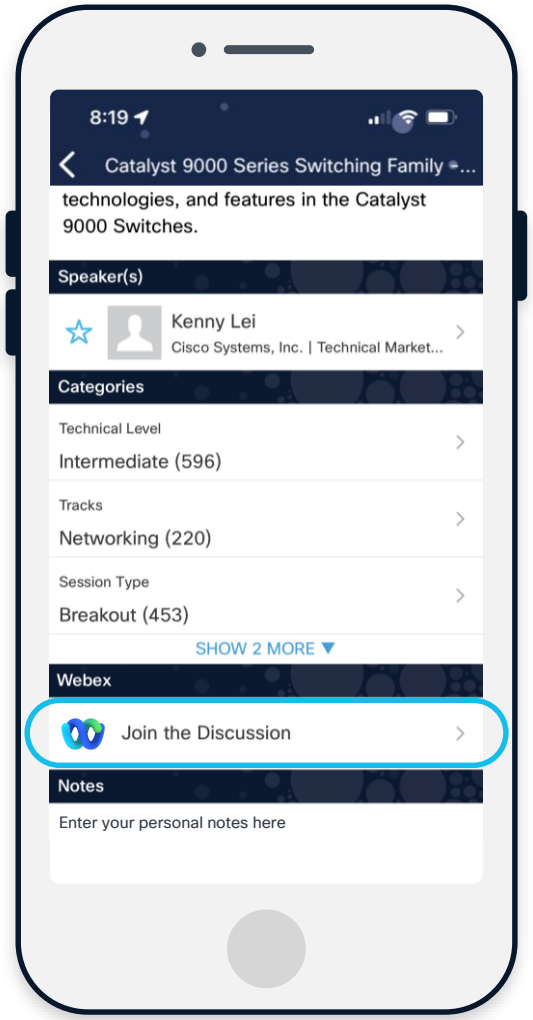
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 13, 2025.



Agenda

01 What's Changed and Why?

02 Rack Portfolio Updates

03 AI Server Updates

04 Modular Platform Updates

05 AI Platform Considerations

What's Changed and Why?

AI Use Cases

$15.7T
Potential contribution to global economy by 2030

$300B
Global spending on AI by 2026

75%
Of large enterprises will rely on AI-infused processes by 2026

Healthcare and Life Sciences
Diagnosis
Drug discovery
Personalized medicine

Financial Services
Fraud detection
Risk assessment
Trading

Retail
Personalization
Inventory optimization
Virtual agents

Manufacturing
Predictive maintenance
Quality control
Demand forecasting

Agriculture
Yield optimization
Automated irrigation
Pest prediction & prevention

Transportation
Route optimization
Autonomous vehicles
Predictive maintenance

Energy
Distribution optimization
Fault prediction
Demand forecasting

Public Sector
Smart cities
Security
Services improvement

AI Readiness

85%

say they have less than 18 months to deploy an AI strategy, or they will see negative business effects.[1]

13%

IT organizations with infrastructure prepared for AI today[2]

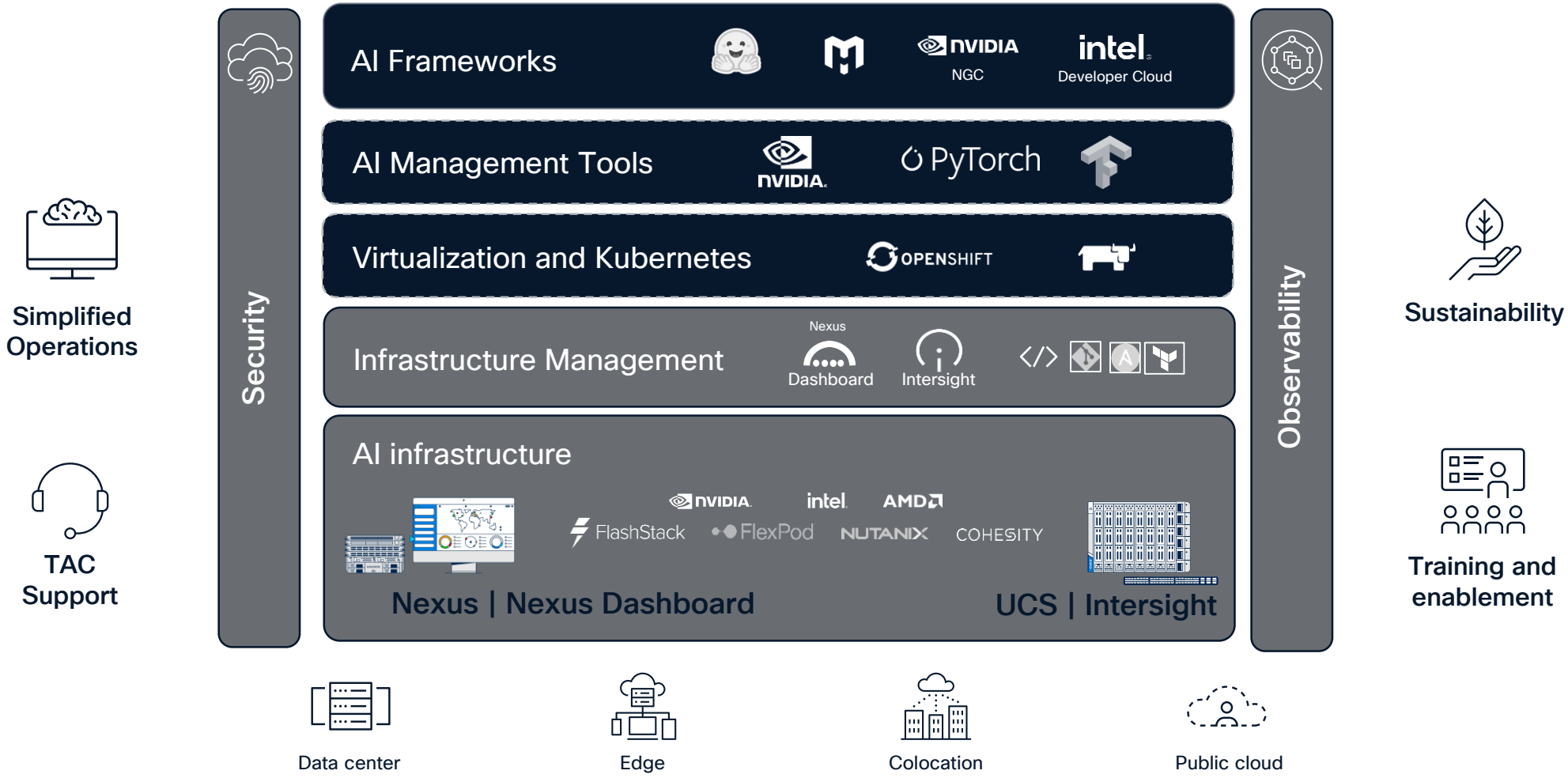
59%

believe AI's impact on their business will surpass expectations after 5 years. [1]

1. Cisco Global AI Readiness Index, 2024
2. Cisco Global AI Readiness Index, December 2023

AI-Ready Infrastructure Stack

Easy



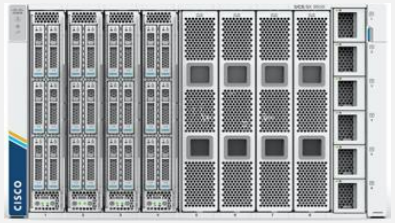
Rack Portfolio Updates

Cisco UCS Compute Portfolio

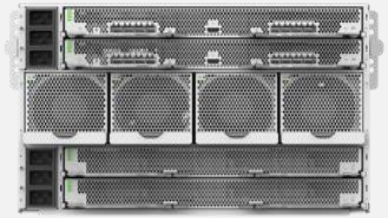
MAINSTREAM ENTERPRISE SERVERS

UCS X-Series X9508 Chassis

IFM Module



UCS X-Series Direct



UCS X210c M8 — New

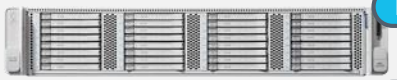
UCS X215c M8 — New

UCS X410c M7

UCS X210c M7

UCS B200 M6

UCS C240 M8E3S
36 EDSFF E3.S1T — New

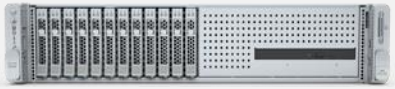


UCS C240 M8SX
28 HDD/SDD/NVMe — New

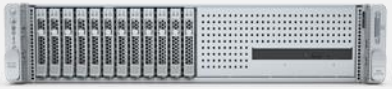
UCS C240 M8L
16 LFF + 4 SFF — New

UCS C240 M7SN
28 NVMe

UCS C240 M6S
14 SSD/HDD Media drive



UCS C240 M6N
14 NVMe Media Drive



UCS C220 M8E3S
16 EDSFF E3.S1T — New

UCS C220 M8S
10 HDD/SSD/NVMe — New

UCS C220 M7N
10 NVMe

UCS C245 M8SX
28 HDD/SDD — New

UCS C225 M8S
10 HDD/SSD — New

UCS C225 M8N
10 NVMe — New

AI SERVERS

UCS C885A M8
8RU Dense GPU Server — New

UCS C845A M8
4RU MGX Server — New

Cisco UCS C220 M8 Rack Server

Dense 1RU form factor for a wide range of workloads, including virtualization, web, collaboration, cloud, and bare-metal applications

C220 M8 SFF SKU
Up to 2x Intel Xeon 6 Processors
Up to 86C / 350W per CPU

Hybrid drive backplane supports up to 10x 2.5in HDD/SSD/NVMe (up to 8x direct attach NVMe SSDs

C220 M8 EDSFF SKU
Up to 2x Intel Xeon 6 Processors
Up to 86C / 350W per CPU

Up to 16x E3.S direct attach NVMe SSDs (Gen5 x4 each)

Cisco Intersight and IMM support

3x HH PCIe 5.0 or 2x FH PCIe 5.0
MLOM / OCP 3.0 Support
VIC 15000 Series

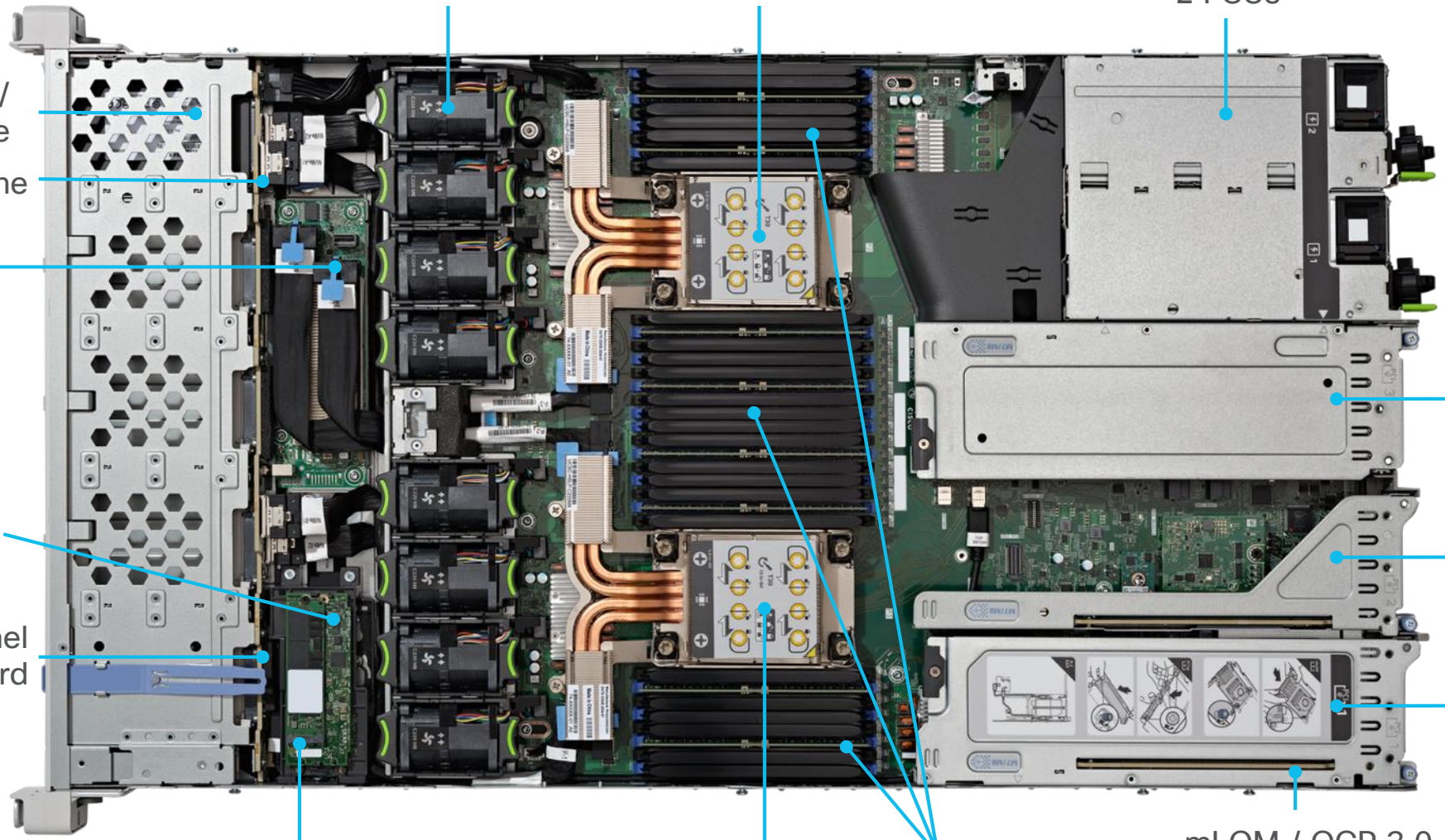
Dual M.2 SATA RAID for boot (hot-swappable M.2 available)

DDR5 6400 MT/s DIMMs (running at 6400 MT/s 1DPC, 5200 MTS 2DPC)
DDR5 8000 MT/s MRDIMMs

CPUs >330W limited to 30C Ambient

UCS C220 M8 – System Placement

Top View – 3 HHHL



Cisco UCS C240 M8 Rack Server

Exceptional performance for enterprise workloads, including big data analytics, collaboration, databases, virtualization, and high-performance applications

C240 M8 SFF SKU
Up to 2x Intel Xeon 6 Processors
Up to 86C / 350W per CPU

Hybrid drive backplane supports up to 28 2.5in HDD/SSD/NVMe (up to 12x direct attach NVMe SSDs)

C240 M8 EDSFF SKU
Up to 2x Intel Xeon 6 Processors
Up to 86C / 350W per CPU

Up to x 36 E3.S direct attach NVMe SSDs
(Up to 32 Gen5 x4 each)

Cisco Intersight
and
IMM support

8x FH PCIe 5.0
MLOM / OCP 3.0 Support
VIC 15000 Series

Dual M.2 SATA RAID for boot
(hot-swappable M.2 available)

DDR5 6400 MT/s DIMMs
(running at 6400 MT/s 1DPC, 5200 MTS 2DPC)
DDR5 8000 MT/s MR DIMMs

Cisco UCS C240 M8 Rack Server

Exceptional performance for enterprise workloads, including big data analytics, collaboration, databases, virtualization, and high-performance applications

C240 M8 LFF SKU
Up to 2x Intel Xeon 6 Processors
Up to 64C / 330W per CPU

3.5in LFF backplane supports up to 16 3.5in SAS HDDs (plus up to 4x rear SFF SAS/SATA/NVMe SSDs

Dual M.2 SATA RAID for boot
(hot-swappable M.2 available)

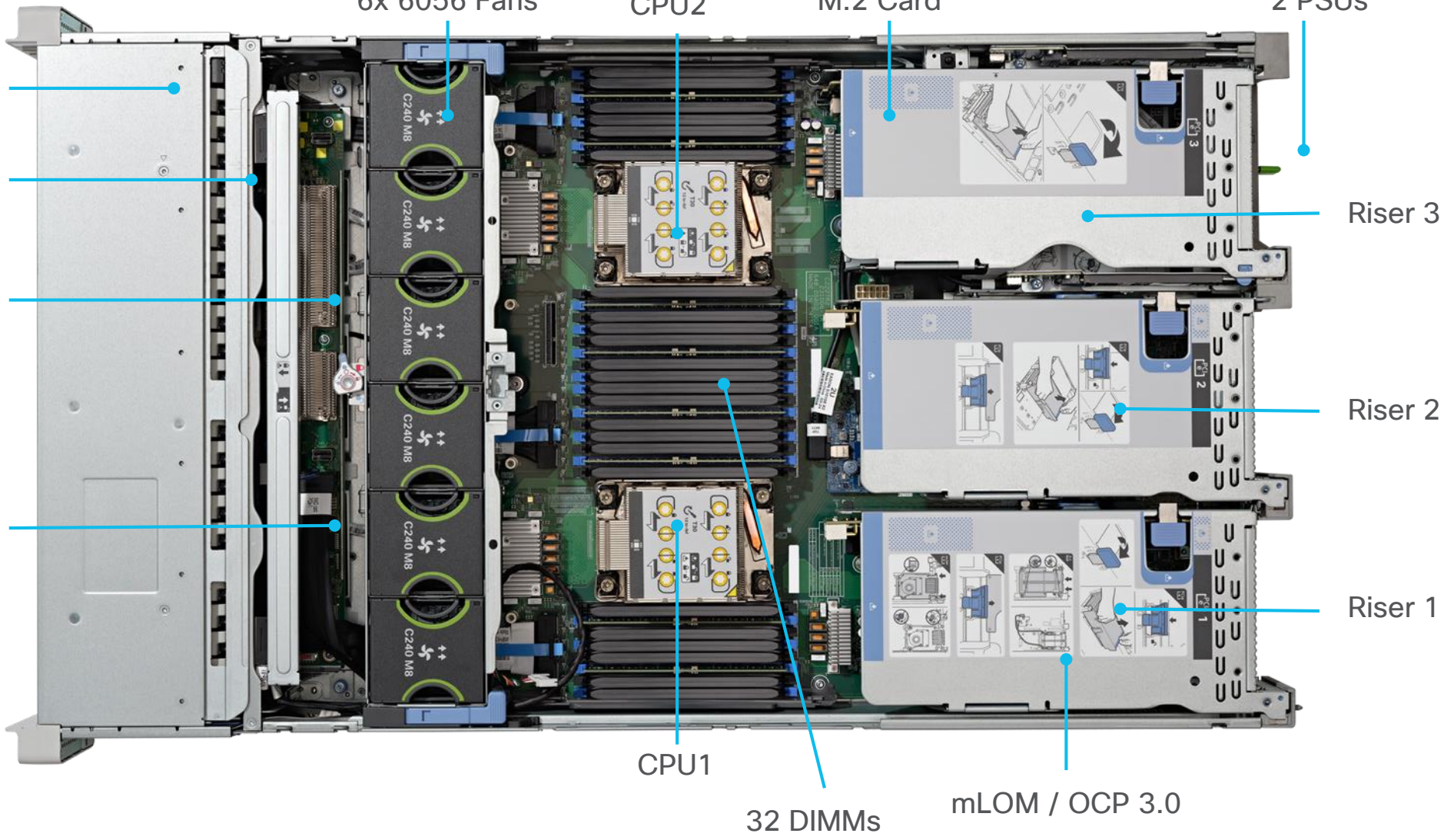
Cisco Intersight
and
IMM support

5x FH PCIe 5.0
MLOM / OCP 3.0 Support
VIC 15000 Series

DDR5 6400 MT/s DIMMs
(running at 6400 MT/s 1DPC, 5200 MTS 2DPC)
DDR5 8000 MT/s MR DIMMs

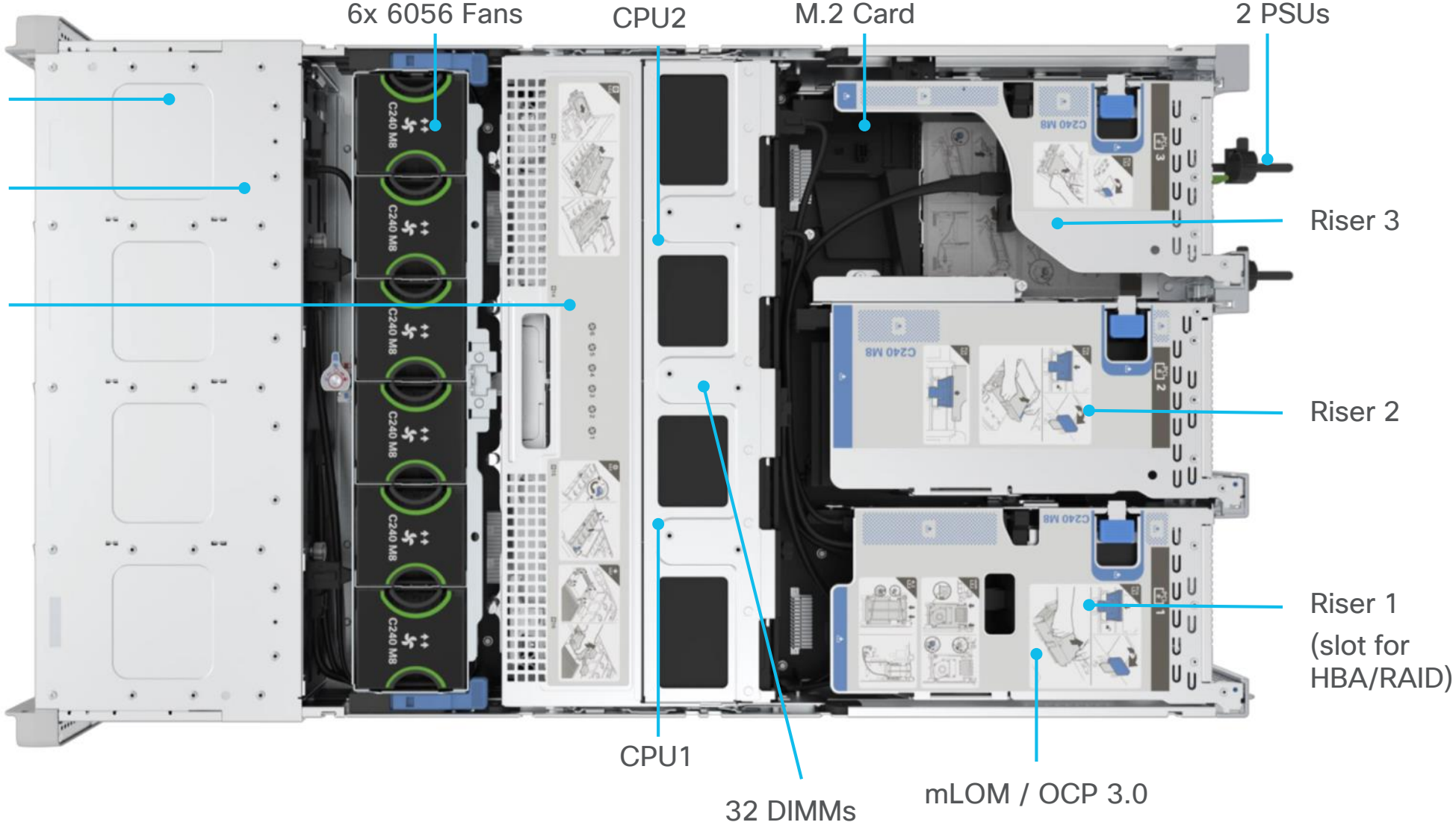
UCSC-C240-M8SX/M8E3S - System Placement

Top View



UCSC-C240-M8L - System Placement

Top View



Cisco UCS C225 M8 Rack Server

Dense 1RU form factor for a wide range of workloads, including EDA, SDS, Big Data, Edge Centric application and other common 1U server use cases

Single socket optimized

All the IO is derived from CPU1 for maximum flexibility in a single socket footprint

Hybrid drive backplane suppots up to 10 2.5in HDD/SSD/NVMe (up to 4x direct attach NVMe SSDs)

Direct Attach All NVMe SKU

Up to 10 NVMe SSD drives directly connected to CPU1

All NVMe Backplane supports 10 direct attach SFF NVMe SSDs

Cisco Intersight
and
IMM support

PCIe 5.0
MLOM / OCP 3.0 Support
VIC 15000 Series

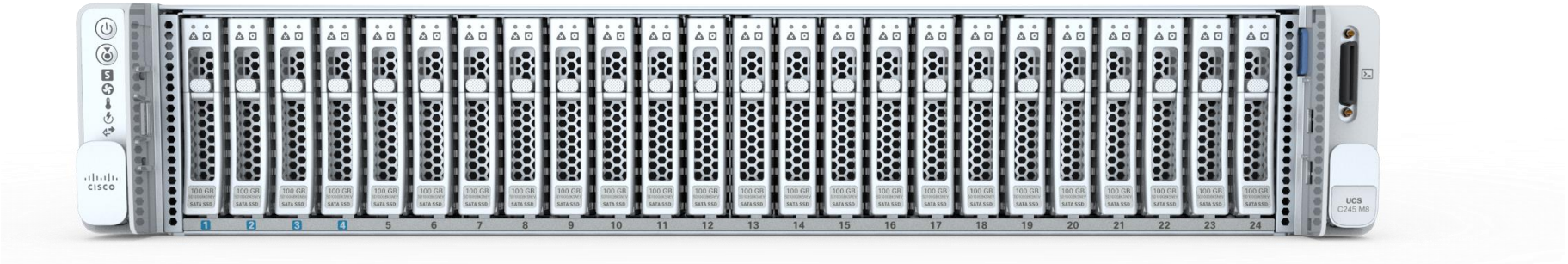
Support for all 4th and 5th Gen EPYC CPU SKUs (up to 360W TDP)
Up to 160 Cores

DDR5 6400 MT/s DIMMs
12 DIMM slots
Up to 3TB of Memory

Cisco UCS C245 M8 Rack Server

Exceptional performance for enterprise workloads, including big data analytics, collaboration, databases, virtualization, and high-performance applications

C245 M8 SFF SKU
Supports 2x 4th and 5th Gen EPYC CPU SKUs
(up to 160C / proc)



Hybrid drive backplane supports up to 28 2.5in HDD/SSD/NVMe (up to 4x front / 4x rear direct attach NVMe SSDs

Dual M.2 SATA RAID for boot
(hot-swappable M.2 available)

Cisco Intersight
and
IMM support

PCIe 5.0
MLOM / OCP 3.0 Support
VIC 15000 Series

DDR5 6400 MT/s DIMMs
24 DIMM slots
Up to 6TB of Memory

New Storage Options for M8 Rack Servers

Flexible Storage options to align with a variety of workloads and applications

- **Cisco 24G Modular Tri-Mode RAID Controller + 4GB Cache** provides support for up to 14 drives
  - Provides RAID 0/1/5/6/10/50/60
  - UCSC-RAID-M1L16
  - Supported on C220M8S, C225M8S, C240M8SX (10 drives on 1U and 14 on 2U)
  - Up to 2x controllers on 2U
- **Cisco 24G Modular Tri-Mode RAID Controller + 4GB Cache** provides support for up to 28 drives
  - Provides RAID 0/1/5/6/10/50/60
  - UCSC-RAID-MP1L32
  - For C240M8SX and C245M8SX
- **Cisco 24G Modular Tri-Mode RAID Controller + 8GB Cache** provides support for up to 20 drives
  - Provides RAID 0/1/5/6/10/50/60
  - UCSC-RAID-MP1LL32
  - For C240M8L only

New Storage Options for M8 Rack Servers

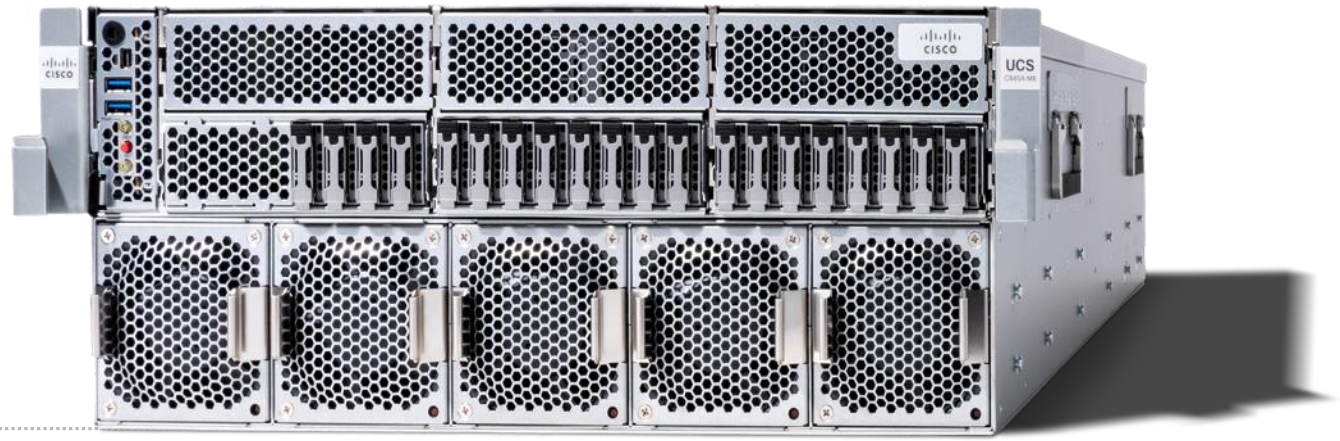
Flexible Storage options to align with a variety of workloads and applications

- **Cisco 24G Modular Tri-Mode HBA** for pass-through support for up to 14 drives
  - No RAID
  - UCSC-HBA-M1L16
  - For C220M8S, C225M8S, C240M8SX and C245M8SX (10 drives on 1U and 14 on 2U)
  - Up to 2x controllers on 2U
- **Cisco 24G Modular Tri-Mode HBA** for pass-through support for up to 20 drives
  - No RAID
  - UCSC-HBAMP1LL32
  - For C240M8L only

AI Rack Server Updates

Introducing the newest member to Cisco's AI server family

Flexible, modular design for organizations looking to "start small and scale up" with AI



Announced February 2025, Orderable NOW

UCS Accelerated Cisco UCS C845A M8

NVIDIA MGX reference design

Configurations with NVIDIA or AMD GPUs

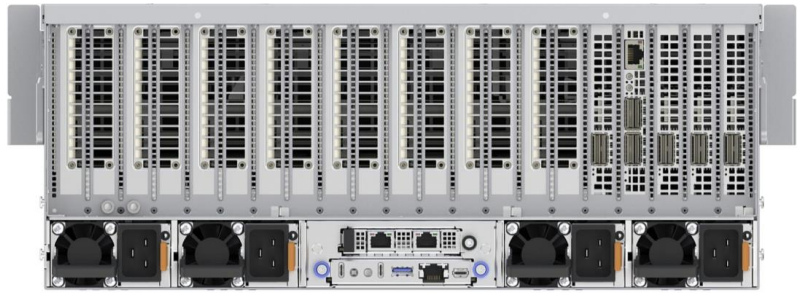
Included as an option in Cisco Hyperfabric

High performance in a compact form factor

Enhanced power delivery, fewer PCBs, and better cable routing for optimal airflow and thermal management

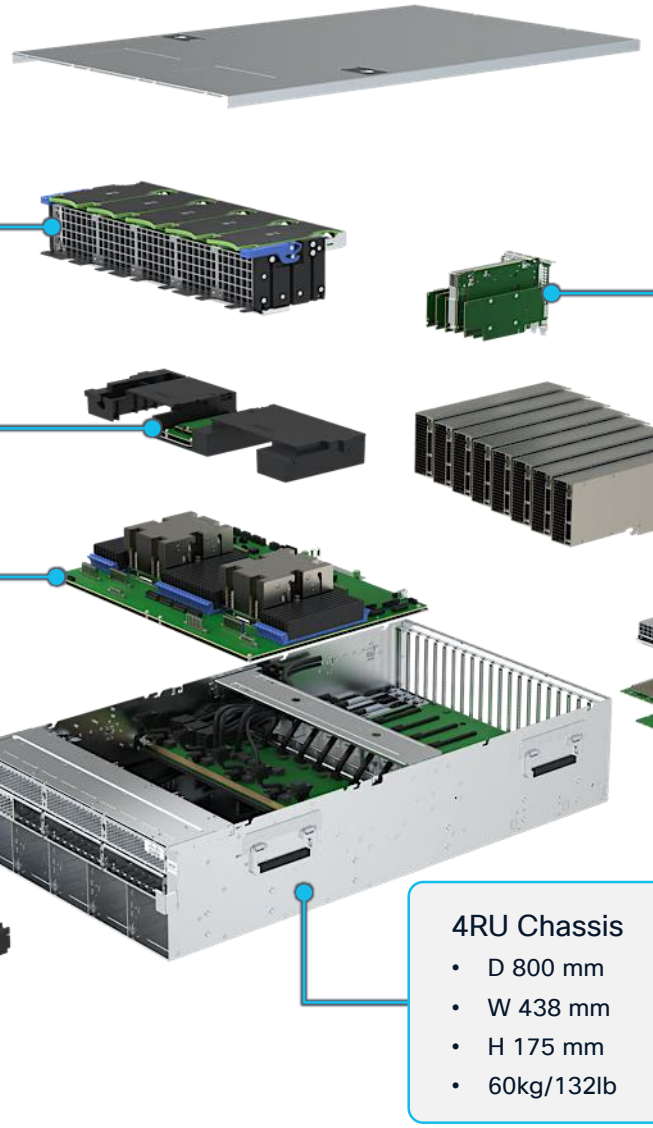
Introducing Cisco UCS C845A M8

UCS C845A 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 exploded drawing



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 RTX Pro 6000
- 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

Cisco UCS C845A M8 Versatile AI Server

Optimized for GenAI

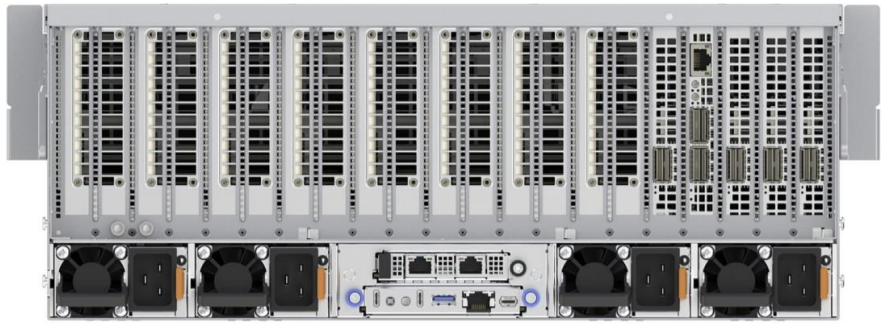
- Address a variety of AI use cases
- Modular NVIDIA MGX™ architecture enables configuration flexibility
- Built on air-cooled enterprise rack design that easily fits in your data center

Scalable AI performance

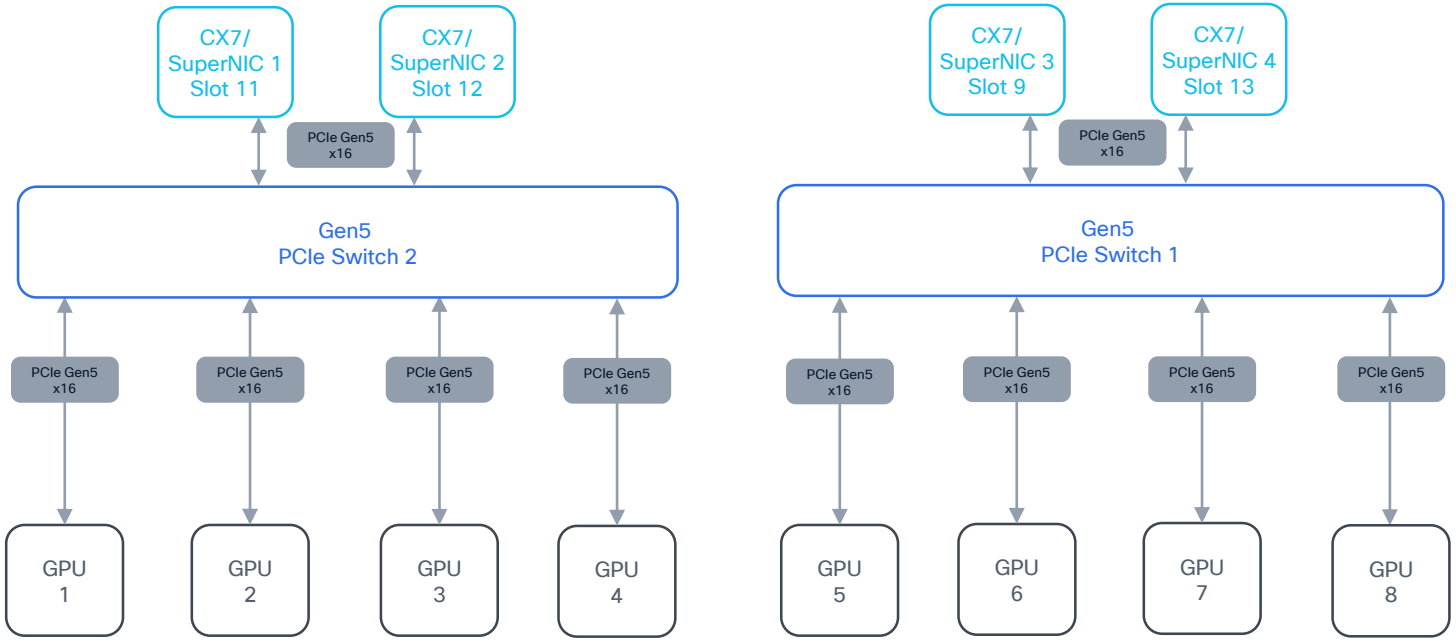
- Choose the number of GPUs that match your use case
- Increase the number of GPUs when your workloads demand it
- Easily create clusters of AI servers to scale out and meet increasing needs

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

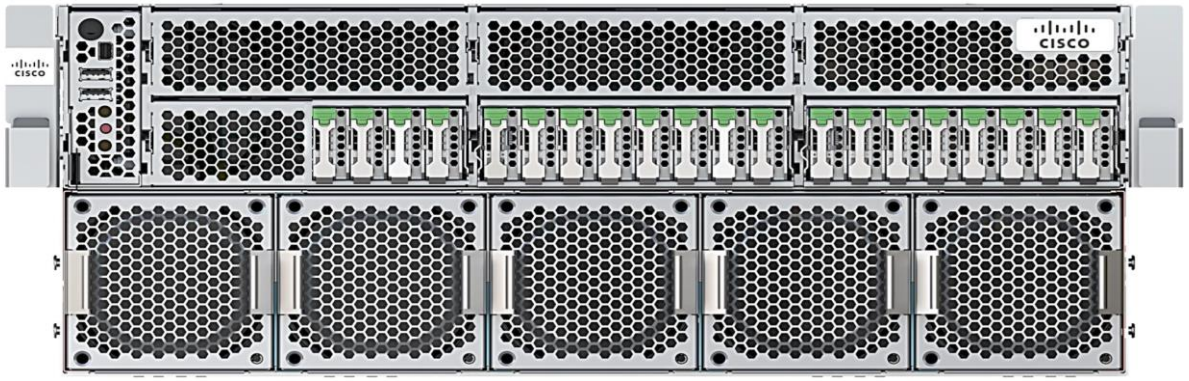


UCS C845A – GPU Connectivity



- 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 be 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

The UCS C845A M8—a better overall AI server improving upon the MGX reference design



UCS C845A M8 MGX Server (2-8-5) Mechanicals

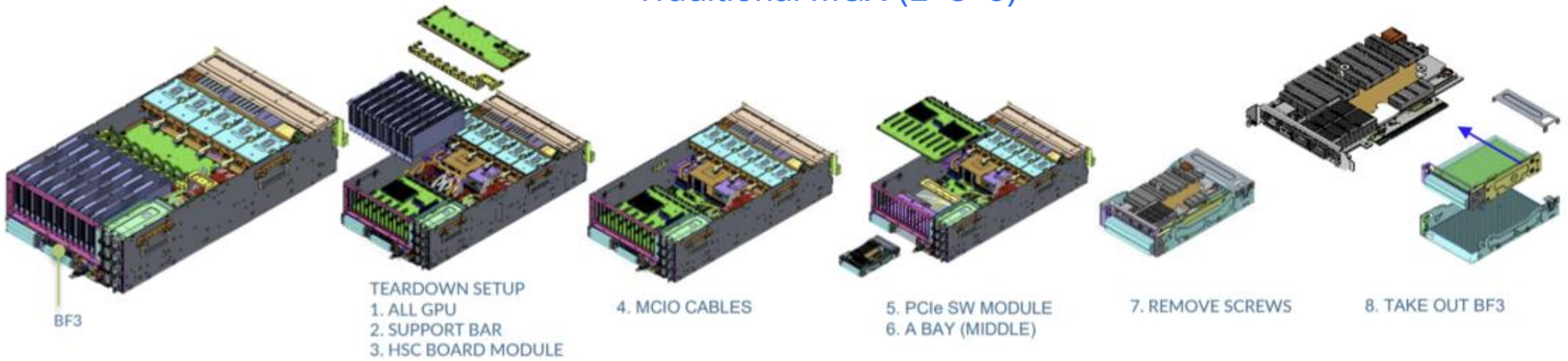
- Improved power delivery
- Better design (fewer PCB's (17 PCB -> 11 PCB)
- Improved cable routing in manufacturing and serviceability
  - Total decrease of 19 cables
- Leverages MGX building block infrastructure

*From an electrical, system and thermal design perspective, C845A M8 is **100% compliant to the MGX Architecture***

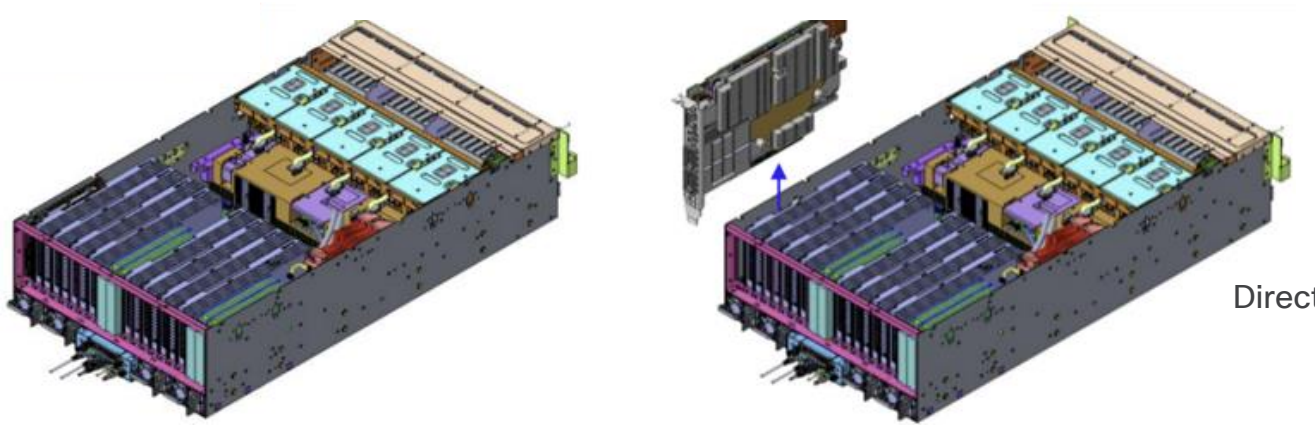
The UCS C845A M8 – A Better Overall Design

PCIe Card Serviceability Example

Traditional MGX (2-8-5)



Cisco UCS C845A M8 (2-8-5)



Directly Remove BF3

Cisco UCS C885A M8 High Density Rack Server

Bringing high-density GPU servers to the Cisco UCS® family and to Cisco's AI solution portfolio

Purpose-built for data-intensive use cases like model training and deep learning

Training—large/small models
Fine-tuning
Large model inferencing
Retrieval-Augmented Generation (RAG)

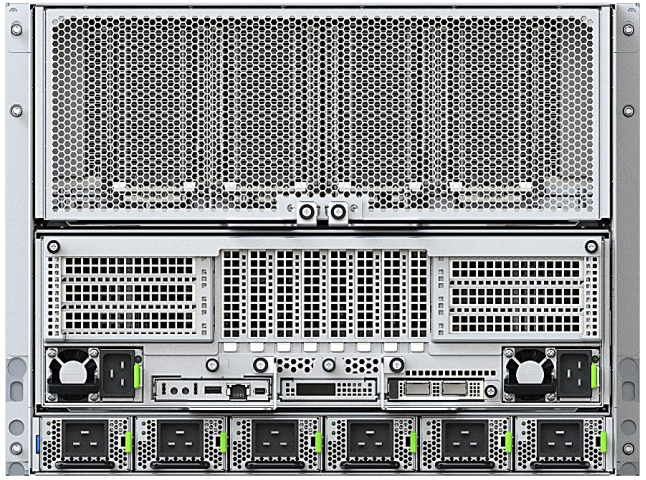
Cisco UCS C885A M8

NVIDIA HGX with
8 NVIDIA H100/H200 GPUs or

8 AMD MI300X GPUs

2 AMD 4th or 5th Gen EPYC™ Processors

UCS C885A M8 Dense GPU Server Specifications

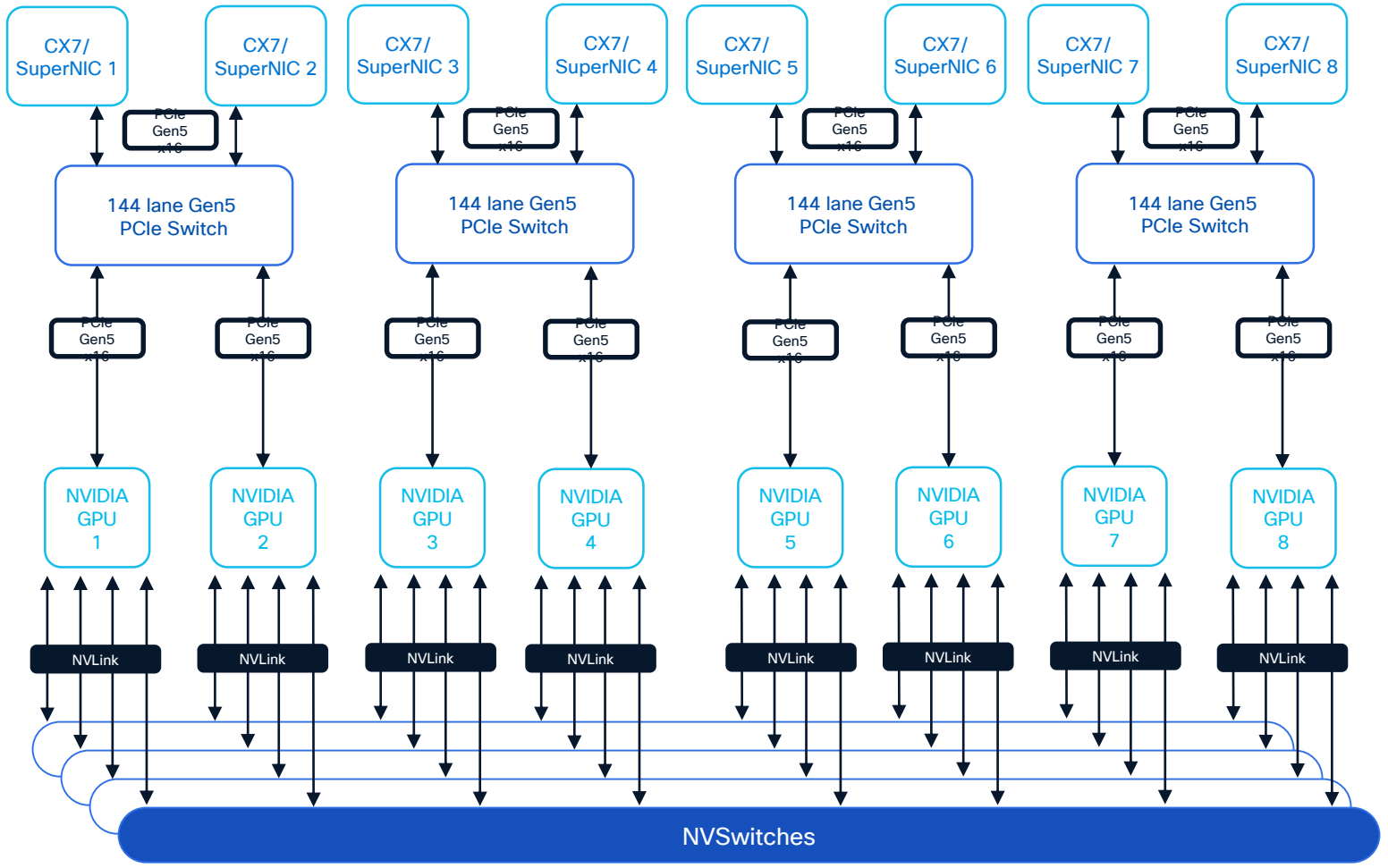


Product Specifications	
Form Factor	• HGX 8RU 19” Rack Server
Compute + Memory	• 2x 4th Gen AMD EPYC 9554 (360W, 64 core, up to 3.75GHz) or • 2x 5th Gen AMD EPYC 9575F (400W, 64 core, up to 5GHz) • 2x 5th Gen AMD EPYC 9535 (300W, 64 core, up to 3.5GHz) • 24 DDR5 RDIMMs • Up to 6,000 MT/S
Storage	• 1 PCIe3 x4 M.2 NVMe (Boot Device), optional 2x M.2 with HW Raid • Up to 16 PCIe5 x4 2.5” U.2 NVMe SSD (Data Cache)
GPUs	• NVIDIA: 8 x H100 (700W) or 8 x H200 (700W) • AMD: 8 x MI300X (750W)
Network Cards	• 8 PCIe x16 HHHL for East-West NIC ConnectX-7 or BF3 B3140H • Up to 2 PCIe x16 FHHL for North-South NIC ConnectX-7 or BF3 B3220 • 1 OCP 3.0 X710-T2L for North-South or host management
Cooling	• 12 Hot swappable (N+1) fans for system cooling • 4 fans for SSD cooling
Front IO	• 2 USB 2.0, 1 ID Button, 1 Power Button
Rear IO	• 1 USB 3.0 A, 1 USB 3.0 C, mDP, 1 ID Button, 1 Power Button, 1 USB 2.0 C (for debugging), 1 RJ45 (mgmt.)
Power Supply	• Up to 6 54V 3kW and 2 12V 2.7kW MCRPS/CRPS, N+1 redundancy

UCS C885A M8 Modular Sled Design

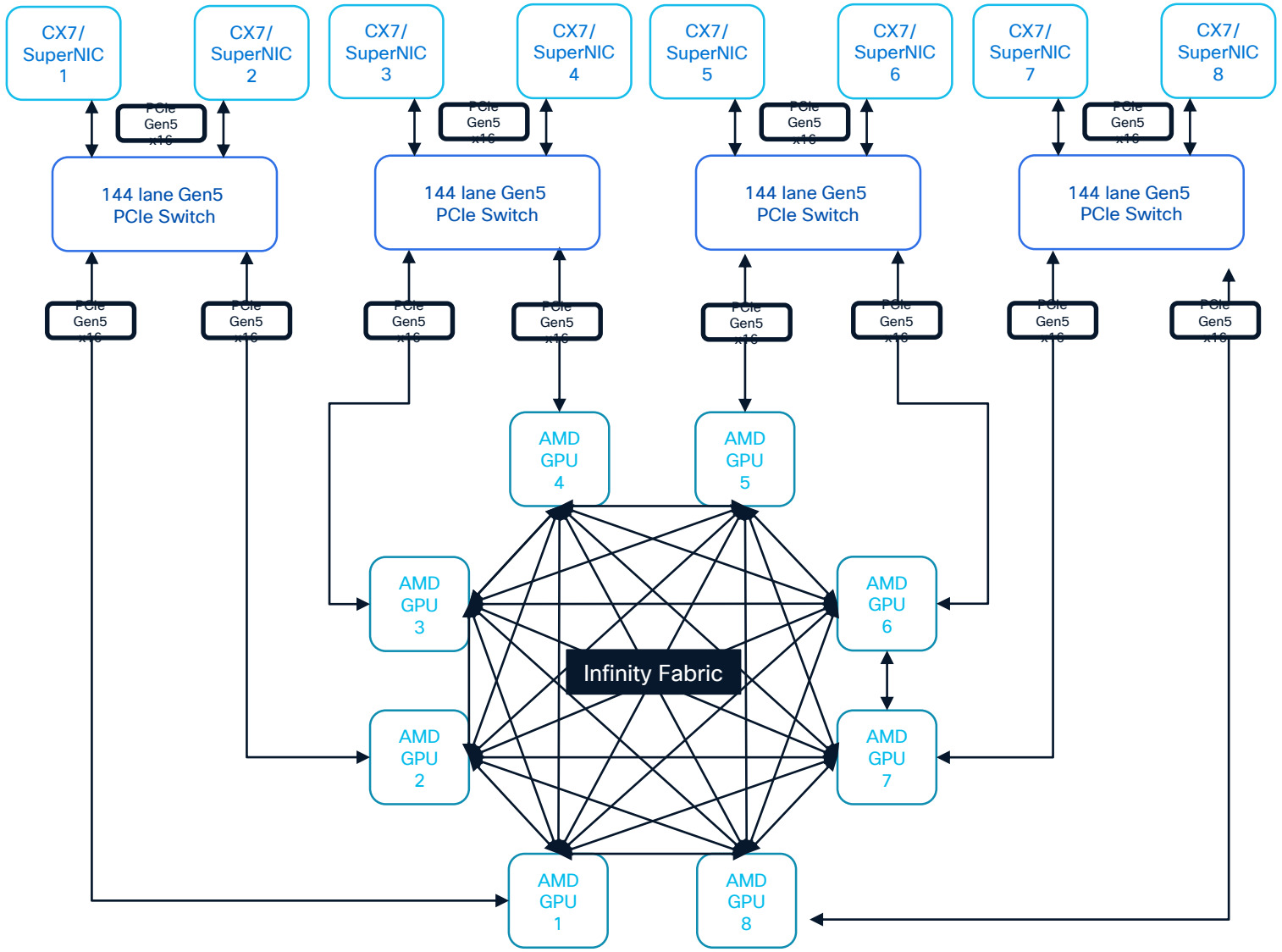
NVIDIA HGX Architecture

UCS C885A M8 – NVIDIA GPU Connectivity



- 8x NVIDIA H100/H200 SXM5 Tensor Core GPUs
- Each H100/H200 GPU has multiple NVLink ports and connects to all four NVSwitches
- 4 x fully non-blocking NVSwitches that connect all 8 GPUs
- Each H100/H200 GPU also has a dedicated NIC/SuperNIC connected via PCIe Gen5 x16 for GPU-to-GPU connectivity across nodes

UCS C885A M8 – AMD GPU Connectivity

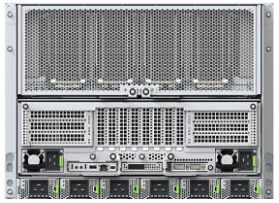


- 8x AMD MI300X OAM GPUs
- Each MI300X GPU has direct access to all other GPUs in full mesh topology over AMD Infinity Fabric mesh
- Each MI300X GPU also has a dedicated NIC/SuperNIC connected via PCIe Gen5 x16 for GPU-to-GPU connectivity across nodes

NVIDIA’s C845A M8 (MGX) vs. C885A M8 (HGX)

Feature/Aspect	NVIDIA MGX	NVIDIA HGX
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

Cisco Rack Server for AI Workloads



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

Modular Platform Updates

X-Series Portfolio

Fabric

4th and 5th Gen FI

- 25/100G ports
- unified ports – up to 16x 32G FC ports (6536)
- Supports VIC 1400, 14000 and 15000 series

25/100G IFM

8 x 25/100G connectivity

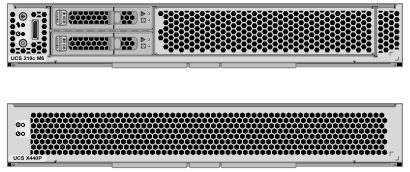
4th and 5th Gen VIC

25/100G connectivity for both blades and racks.

X-Fabric and PCIe Node

X-Fabric

- Based on native PCIe Gen. 4
- Provides GPU acceleration to enterprise application
- No backplane or cables = Easy upgrades



GPU Node and Front Mezz GPUs

- Various NVIDIA GPU options with X210c and X215c.
- AMD GPU option with X215c M8 compute.

Compute

X210c Compute Node

- 2- Socket, single slot servers
- Three Generations: M6, M7, M8
- Intel Xeon CPUs

X410c Compute Node

- 4- Socket, dual slot servers
- Intel 4th Gen Xeon CPU
- Up to 64 DDR5 DIMMs

X215c Compute Node

- 2- Socket, single slot servers
- M8 with AMD 4th and 5th Gen EPYC CPU

UCS X9508 System Chassis

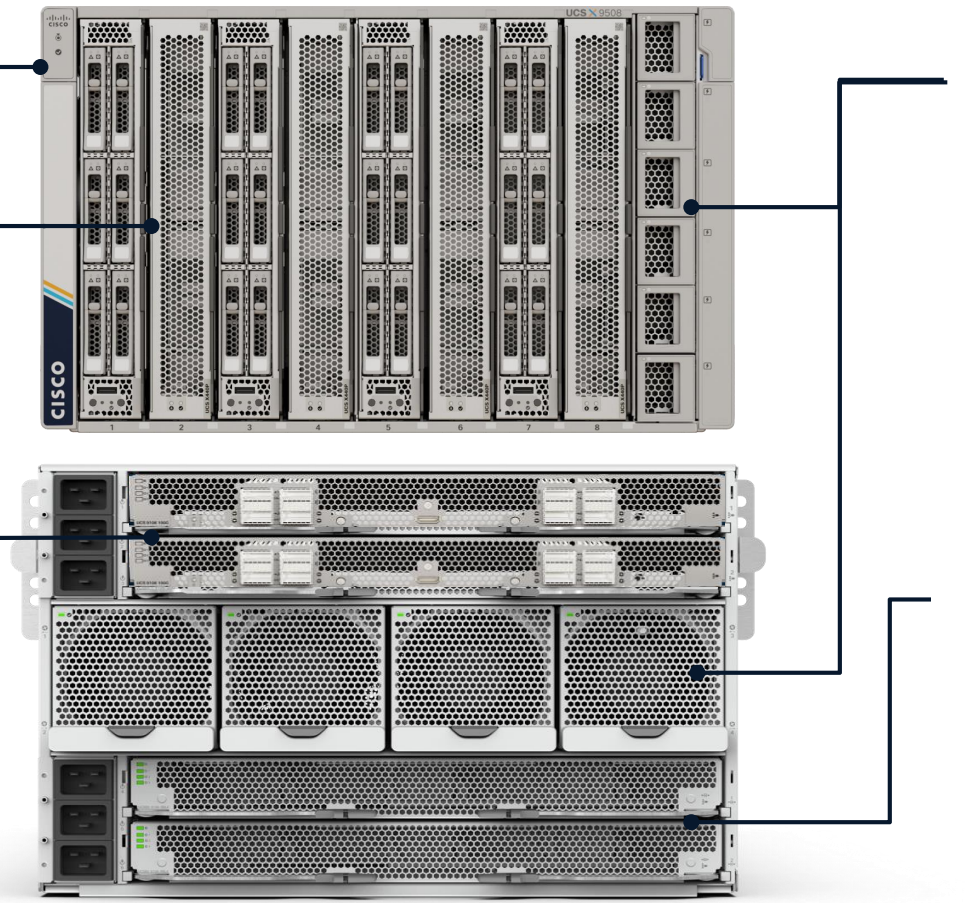
Chassis

7RU IO direct connect

8 flexible slots

Optical ready

Liquid cooling ready



Power and cooling

6x 2800W PSU

54V power distribution

4x 100mm Dual rotor fan

Ethernet fabric

Two Ethernet Modular fabrics

2 TB/s throughput

X-Fabric modules

Two flexible IO modules

For storage and GPU nodes connectivity

UCS X210c M8 X-Series Server Specs Overview

Use Cases	Enterprise-class mainstream blade server with high performance for compute-intensive workloads.	
	• Virtualization • Hyperconverged • Mixed Workload Standardization	• Database and Analytics • Virtual Desktop Infrastructure • AI/ML
Core Platform	**Dual Socket Platform – Up to 172 cores per node, 1376 cores per chassis** Support for Intel Xeon6 6500P and 6700P processors Support for 32xDDR5 6400 DIMMs (1DPC/2DPC, 8 channel) up to 8TB (32 * 256GB*) **PCIe Gen 5 support and CXL 1.1+ support**	
X Fabric/PCIe Node	**Connect to PCIe nodes,** NVIDIA GPUs	
VICS	**200G aggregate/100G per fabric, 100G5th gen VIC mLOM or 25G 5th gen VIC mLOM, 25G 5th Gen VIC Mezz**	
Drives/GPU	**Front mezz options | 6 SAS/SATA U.2 with HW RAID, 6 NVMe PCIe Gen5 x4, 9 E3.S NVMe, up to 2x GPUs** **Internal | 2x M.2 SATA HW RAID1, 2x M.2 NVMe**	
Management	**Intersight Management and UCSM**	

UCS X215c M8 X-Series Server Specs Overview

Use Cases	Enterprise-class mainstream blade server with high performance for compute-intensive workloads.	
	• Virtualization • Hyperconverged • Mixed Workload Standardization	• Database and Analytics • Virtual Desktop Infrastructure • AI/ML
Core Platform	**Dual Socket Platform – Up to 256 cores per node, 2048 cores per chassis** Support for Genoa (-X)/Bergamo/Turin Support for 24xDDR5 DIMMs (1DPC, 12 channel) up to 6TB (24 * 256GB) **PCIe Gen 5 support and CXL 1.1+ support**	
X Fabric/PCIe Node	**Connect to PCIe nodes,** NVIDIA/AMD GPUs	
VICS	**200G aggregate/100G per fabric, 100G5th gen VIC mLOM or 25G 5th gen VIC mLOM, 25G 5th Gen VIC Mezz**	
Drives/GPU	**Front mezz options | 6 SAS/SATA with HW RAID, 6 NVMe PCIe Gen5 x4, up to 2x GPUs** **Internal | 2x M.2 SATA HW RAID1, 2x M.2 NVMe**	
Management	**Intersight Management and UCSM**	

New Storage Option for X210c M8

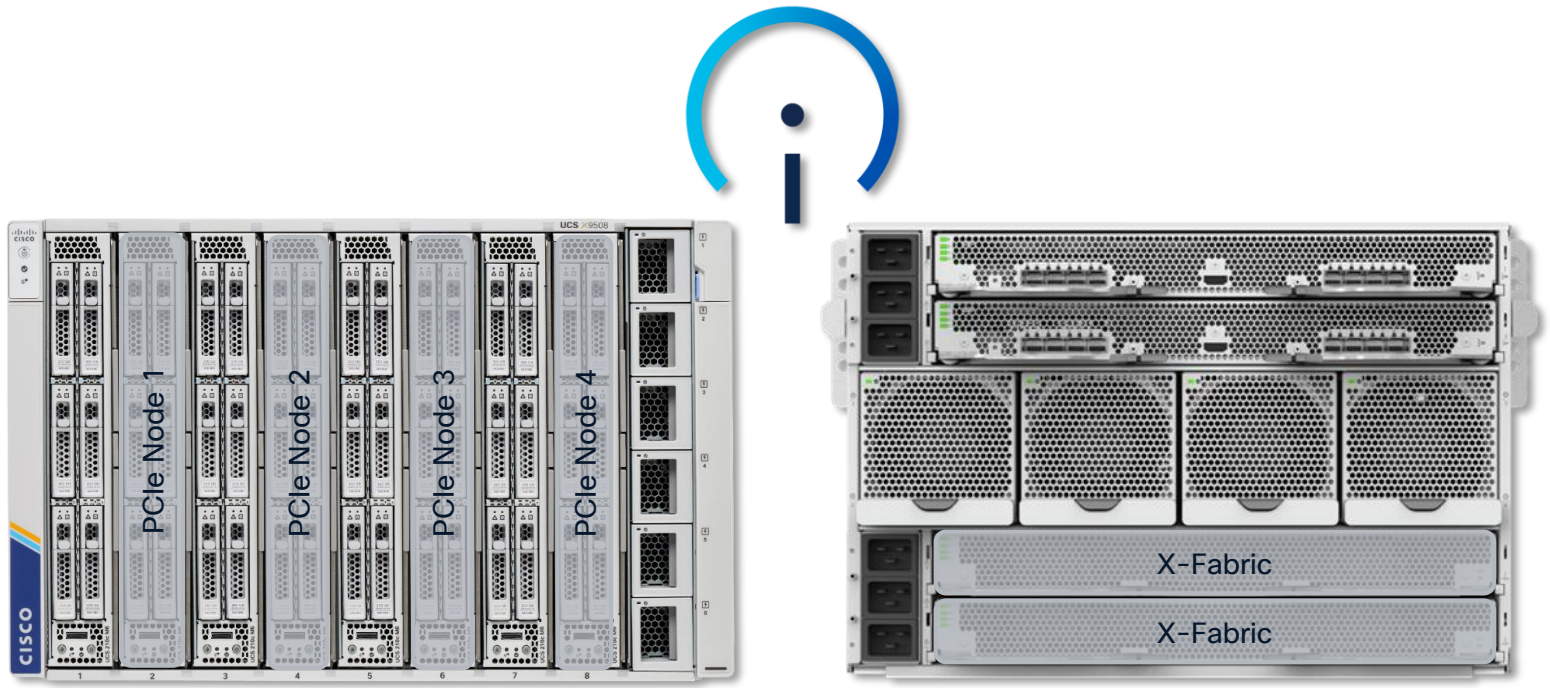
E3.S

- 9x E3.S
- Support of NVMe only
- Pass-Through PCIe Gen 5 only – Up to ~16 GB/s per drive
- Up to 15.3TB drive capacity
- Use cases: vSAN, NTNX, OpenShift, SDS, Cohesity, etc.

UCS X-Fabric Technology, PCIe Nodes with GPU

When paired with M8, PCIe node supports up to:

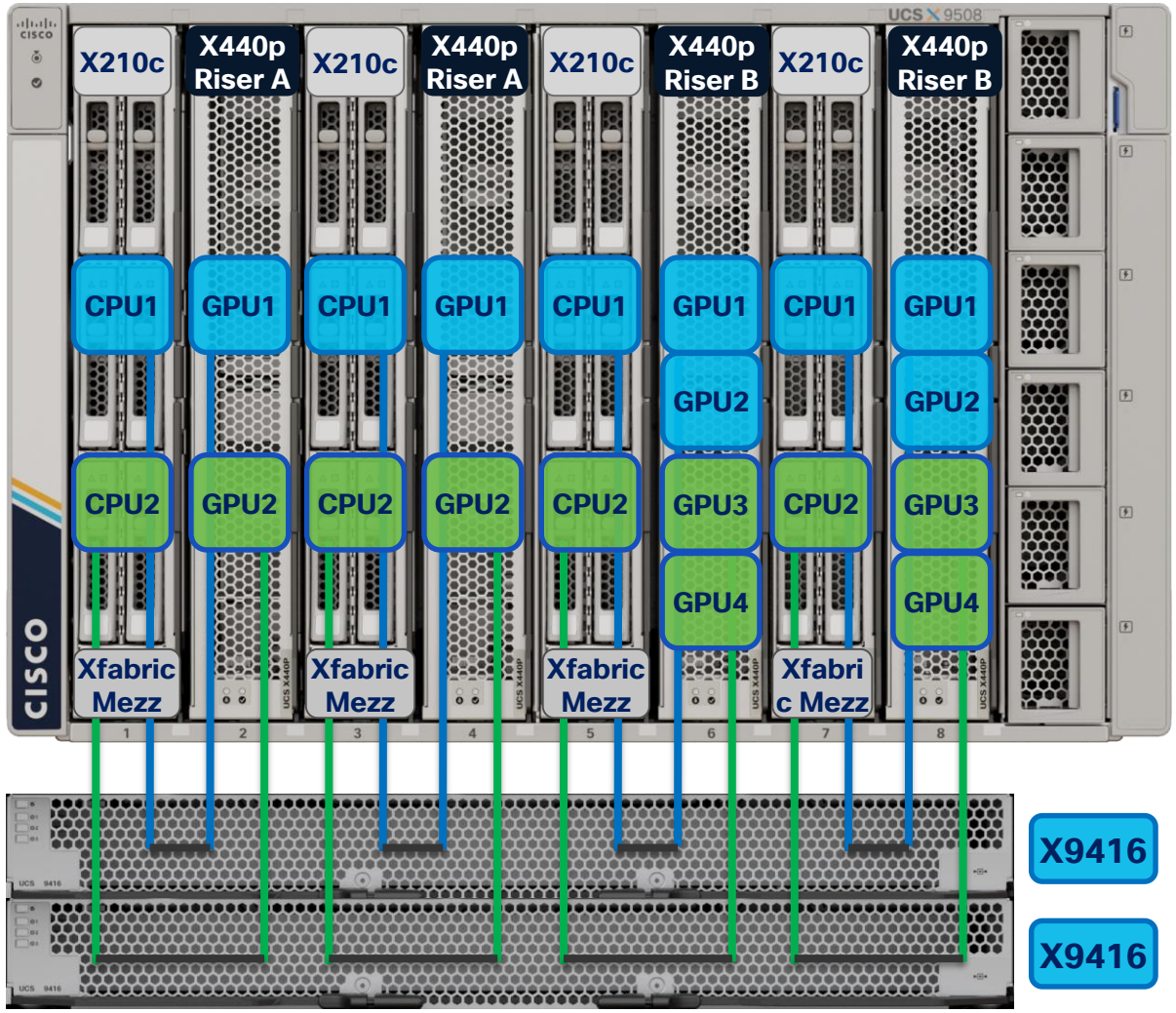
- 2x NVIDIA A16
- 4x NVIDIA L4
- 2x NVIDIA L40S
- 2x NVIDIA H100-NVL
- 2x AMD MI210



- ✓ Based on native PCIe Gen. 4
- ✓ Provides GPU acceleration to enterprise application
- ✓ No backplane or cables = Easily upgrades

UCS X-Fabric - X210c and X440p PCIe Node

- An X-Fabric Mezz card (VIC or Pass-through) on the compute node connects it to the XFM pair
- One Gen4 x16 link per CPU is routed to the XFM pair (CPU1 to XFM1 and CPU2 to XFM2)
- The X440p connects the XFM1 link to Riser 1 and XFM2 to Riser 2
- GPU firmware is updated with the attached server firmware upgrade



AI Platform Considerations

Enterprise GPU server strategy

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**

Cisco UCS® GPU-dense servers

Cisco UCS blade (with GPU extensions) and rack servers

Enterprise AI edge

Dense compute for demanding AI

Full-stack AI with compute and networking

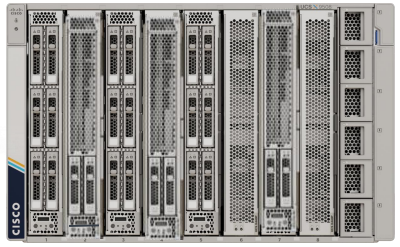
AI Platform Considerations: UCS M8 GPU Options

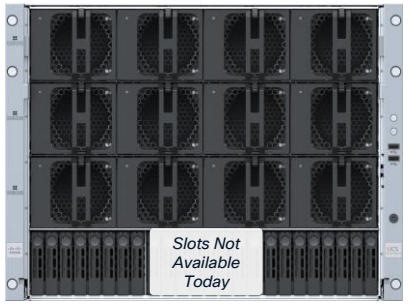
Potential Workload Type

- Entry Tier & Edge
- Universal AI, Text to Image/Video
- AI Training / Inferencing (PCIe)
- AI Training, HPC, Data Analytics

GPUs are subject to change as new peripherals are added to the portfolio

	UCS C220/C225	UCS C240/C245	UCS X-Series	UCS C845A	UCS C885A
	NVIDIA L4*	AMD MI210*		NVIDIA H200 NVL*	AMD MI300X OAM
		NVIDIA H100 NVL*			NVIDIA H100 / H200 SXM
		NVIDIA L40S*			
Max GPUs	3	2-8*	2-8*	2-8	8





** NOTE: GPU Form Factor and GPU model support may vary between AMD and Intel Platforms (i.e. c220/c225, c240/c245, and x210c/x215c). Check the spec sheet for each platform to determine maximum GPU support based on GPU selection*

Game

Complete your session evaluations

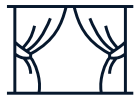
Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to win 1 of 5 full conference passes to Cisco Live 2026.

Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.

Level up and earn exclusive prizes!

Complete your surveys in the Cisco Live mobile app.

Continue your education



Visit the Cisco Showcase for related demos

Book your one-on-one Meet the Engineer meeting

Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs

Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact us at: Via Webex | https://ciscolive.ciscoevents.com/ciscolivebot/**#BRKCOM-1047**

Thank you

cisco Live !


CISCO