GIGA COMPUTING

SIE event 2026

Manufacturing Sites

Eindhoven

600M²
5.4K Server

Ning-Bo

60,000M²
650K Mobo / 30K Server

Tu-Cheng

1,660M²
28.8K Server / 4K Rack

Los Angelas

3,800M²
28.8K Server / 5.4K Rack

Nan-Ping

45,000M²
400K Mobo / 50K Server

São Paulo

145M²
1.2K Server

Dong-Guan

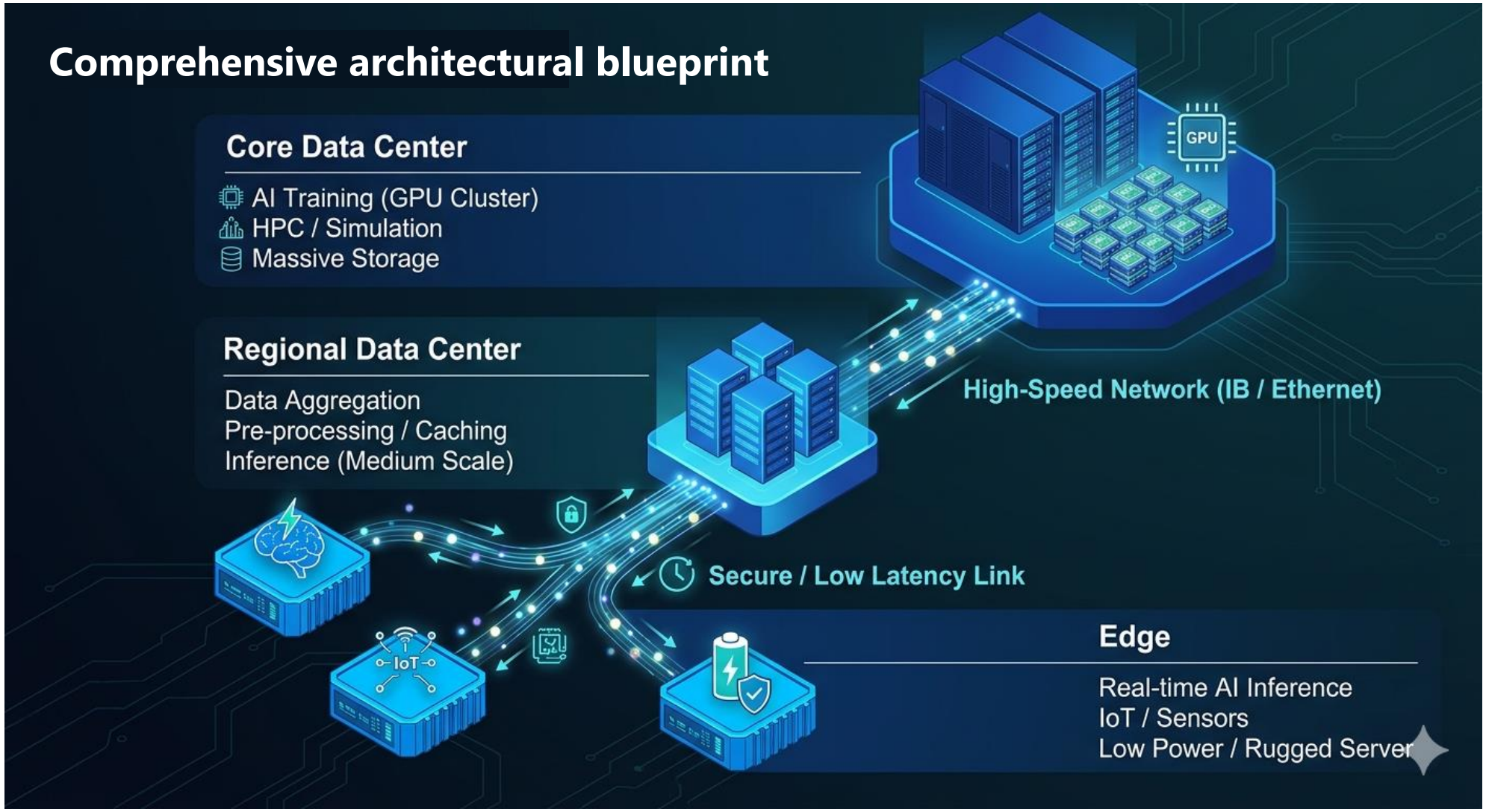
38,000M²
680K Mobo / 20K Server

Johor Bahru

1,200M²
18K Sever / 1.8K Rack

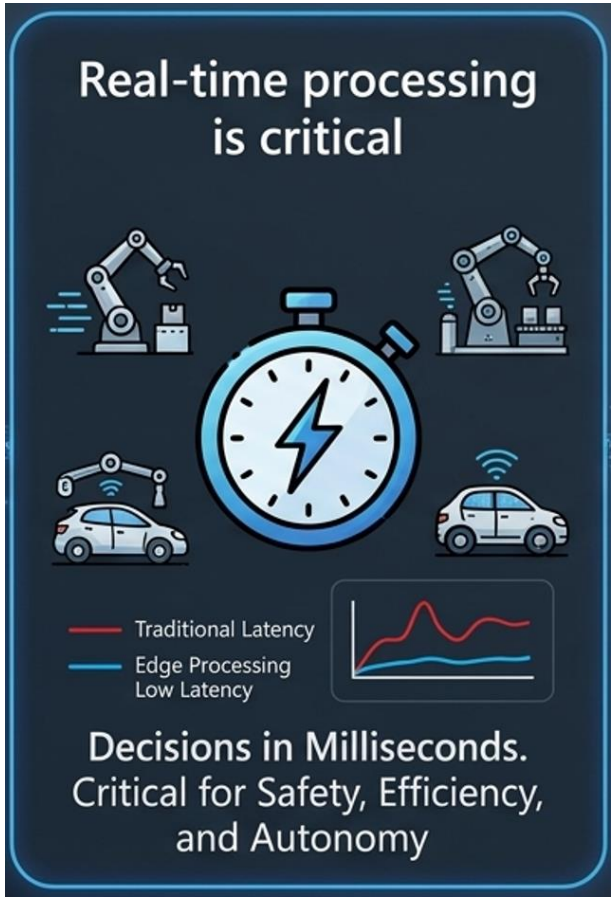
"End-to-end Solutions from Data Center to Edge."

The challenge for enterprises is no longer just about hardware procurement. It's about building a, high-efficiency data ecosystem—**from the Core to the Edge**



Why This Matters?

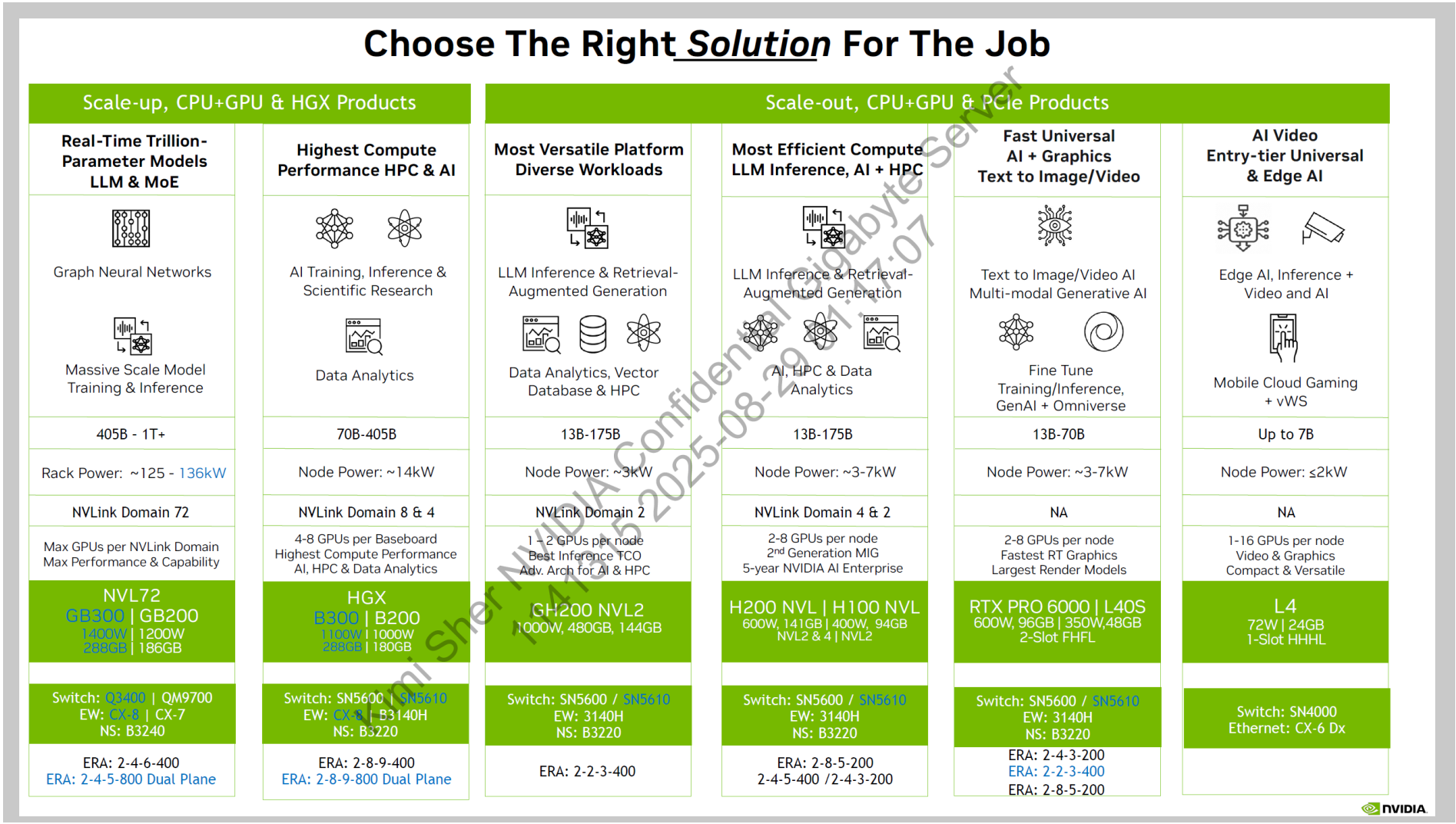
- Data is growing everywhere
- Real-time processing is critical
- AI is moving beyond the data center



Low Latency = High Efficiency

Architecture Overview

- **"How do we choose the right solution for the job?"**



NVIDIA Confidential Gigabyte Server
Ami Sher 2025-08-29 01:17:07

Complete Enterprise Portfolio

From edge to cloud, compute portfolio built for all levels of scale

Public and Cloud Data Centers

Public and Private Networks/ Edge Computing

	Cloud	Regional Data Center	Base Station	Far Edge
Footprint	Large	Large	Small	Smallest
Power Consumption	High	High	Medium	Low

GIGA COMPUTING

Product Portfolio

R-Series

Affordable and expandable rackmount servers, offering ease-of-use, low power consumption and quiet operation

H-Series

Compact and scalable systems providing higher density computing power in a smaller footprint for cloud and other scale-out computing applications

B-Series

Blade servers feature high-density, resource-sharing design with hot-swappable modules, offering efficient, powerful, and cost-effective data center performance.

G-Series

Versatile and scalable high-performance computing with leading efficiency and performance. Ideal for datacenters

S-Series

Storage optimized servers that offer a high data density design, configuration flexibility, and HA and reliability features for data integrity

W-Series

A range of tower servers suitable for an office environment, from entry level to high end computing and HPC

E-Series

Edge computing minimizes latency by relocating computation and storage near data generation points.

OCP-Series

A datacenter solution simple in design, but also highly efficient in power consumption, computing power and configuration

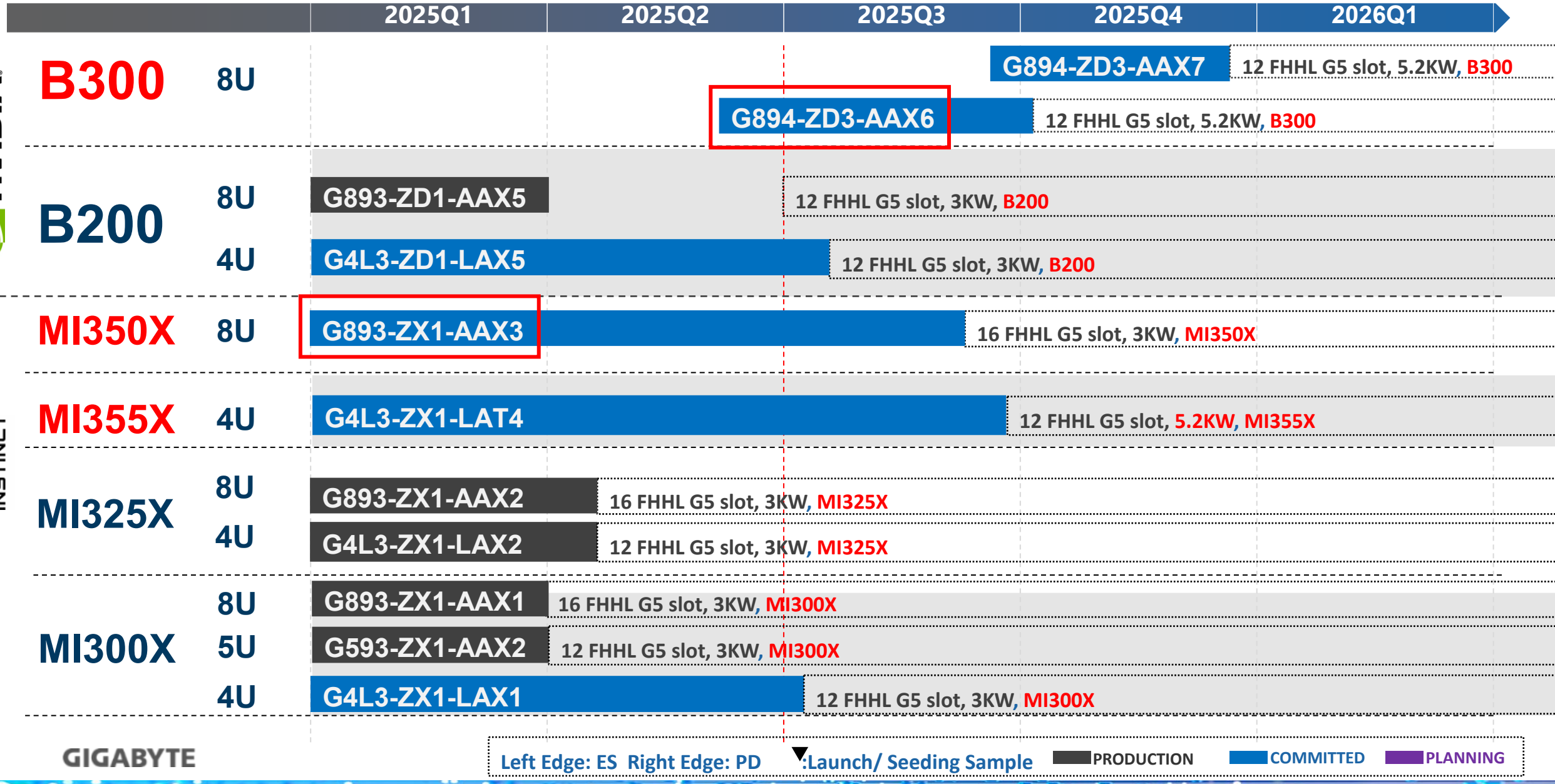
I-Series

Immersion cooling dissipates heat across all system components, supported by specialized enclosure and tray designs that ensure consistently low operating temperatures for efficient computation.

X-Series

Following the NVIDIA MGX modular design. It features a single NVIDIA architecture for accelerated, giant-scale AI and HPC applications.

GPU Module Server Roadmap - AMD Platform



G894-ZD3-AAX6 Series Roadmap

NVIDIA HGX 8-GPU solution (NVIDIA B300)
8UDP, AMD 9005 & 9004 Series, 12 FHHL Gen5 slots,
8 x SXM5 modules on HGX

- **Dimensions: 447 x 351 x 900mm**
- **MB:** MZB3-PE0
- **Memory**
  - 24 x DIMM slots support 24 channel DDR5
  - 4800 MT/s (1DPC) for 9004 Series
  - ***6400 MT/s (1DPC) for 9005 Series***
- **Expansion Slots**
  - 8 x SXM5 OAM GPUs
  - 4 x FHHL Slot (@Gen 5 x16) dual-slot
  - 8 x FHHL Slot (@Gen 5 x16) single-slot
- **Network**
  - Intel X710-AT2 2 x 10GbE ports
- **Storage**
  - 8 x 2.5" U.2 (@Gen5)
  - 2 x M.2
- **12 x 3000W 80 PLUS Titanium redundant power supplies**

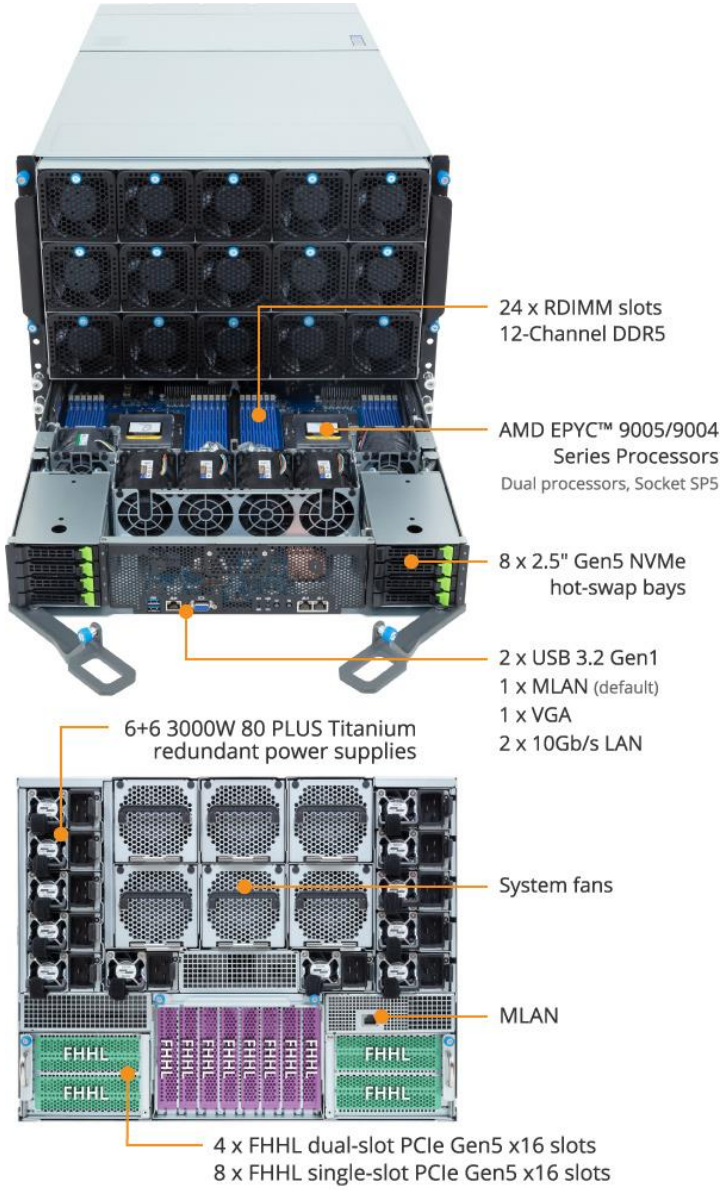
Standard SKUs	MB	GPU	Storage	Expansion Slot	PSU
G894-ZD3-AAX6	MZB3-PE0	**B200**	8 x 2.5" U.2	(4) FHHL (dual)+ (8) FHHL (single)	12 x 3KW
G894-ZD3-AAX7	MZB4-PE1	**B300**	8 x 2.5" U.2	(12) FHHL G5 x16	12 x 3KW

GIGABYTE

HGX B300 vs HGX B200 Comparison

Features	HGX B300 (Blackwell Ultra)	HGX B200 (Blackwell)
TDP	1100W	1000W
GPU Memory / Bandwidth	288GB HBM3E 8.0TB/s	180GB HBM3E 7.7TB/s
System NIC Option	CX8 800G	CX7 400G

G893-ZX1-AAX3 Series Roadmap



AMD Instinct GPU Module Server

- **8UDP, AMD 9005 & 9004 Series, 12 FHHL Gen5 slots, 8 x MI350X GPUs**
- **Dimensions: 447 x 351 x 923mm**
- **MB:** MZB3-PE0
- **Memory**
  - 24 x DIMM slots support 24 channel DDR5
  - 4800 MT/s (1DPC) for 9004 Series
  - 6400 MT/s (1DPC) for 9005 Series
- **Expansion Slots**
  - 8 x AMD Instinct™ MI350X OAM GPUs
  - 4 x FHHL Slot (@Gen 5 x16) dual-slot
  - 8 x FHHL Slot (@Gen 5 x16) single-slot
- **Network**
  - Intel X710-AT2 2 x 10GbE ports
- **Storage**
  - 8 x 2.5” U.2 (@Gen5)
  - 2 x M.2 Gen3 x4 & x1
- **12 x 3000W 80 PLUS Titanium redundant power supplies**

Standard SKUs	GPU	Storage	Expansion Slot	PSU
G893-ZX1-AAX1	MI300X	8 x 2.5” U.2	(12) FHHL G5 x16	8 x 3KW
G893-ZX1-AAX2	MI325X	8 x 2.5” U.2	(12) FHHL G5 x16	12 x 3KW
G893-ZX1-AAX3	MI350X	8 x 2.5” U.2	(12) FHHL G5 x16	12 x 3KW

NVIDIA B300 / AMD MI350X– Target Markets & Applications

Aim for AI Training, Inference & LLM Development

Target Market	Potential Users	Applications
Hyperscalers / CSPs	AWS, Azure, Google Cloud, Oracle	Large-scale AI training clusters, cloud AI inference services, GPU cloud offerings.
AI Labs / Research Institutes	OpenAI, Anthropic, Meta FAIR, DeepSeek	Training of foundation models (LLMs, multimodal AI), experimental model scaling
Large Enterprises	Finance, Pharma, Energy, Manufacturing	High-performance simulations, drug discovery, quantitative trading, digital twins.
SaaS / Internet Companies	Search & Ads, Social Media, Streaming, Gaming	Cloud-scale AI inference, content generation, recommendation engines.
Government & HPC Centers	➢ USA - Oak Ridge National Laboratory ➢ Germany - Jülich Supercomputing Centre (JSC) ➢ Japan - RIKEN Center for Computational Science (R-CCS)	Scientific computing, climate modeling, national AI infrastructure.

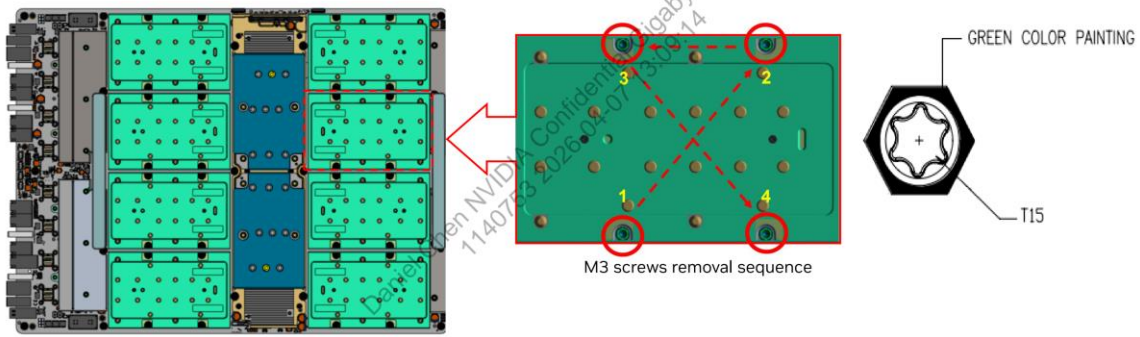
Key Advantages

1. **High Computing Performance** – **+50%** inference improvement than HGX B200.
2. **High Bandwidth & Memory Capacity** – **288GB** HBM3E **with 8TB/s** bandwidth, for long-context workloads.
3. **Networking** – **CX8 (800Gbps)** per GPU, enabling low-latency for training and inference.

HGX Single GPU Removal and Installation Steps

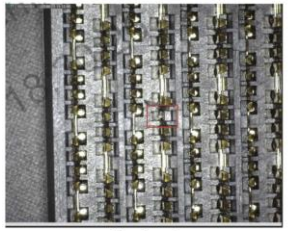
3. Install single SXM6 GPU

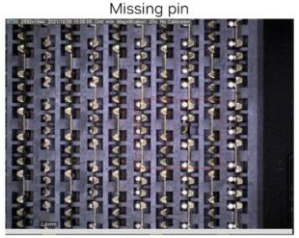
- Torque setting : 8.26 +- 4% kgf-com (0.81Nm)
- Screws in diagonally pattern

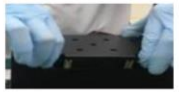


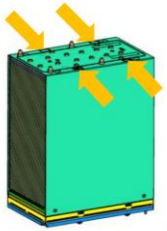
4. Any damage

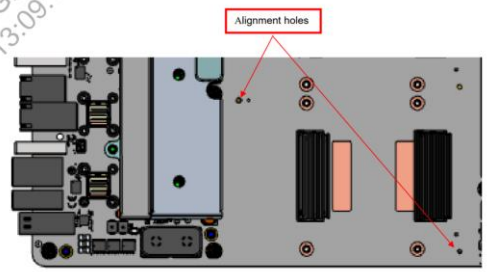
Deformed housing


Missing pin


Bent pin


Grasp the indented area





Caution: A damaged connector on the baseboard will damage its mating connector on the GPU module, and vice versa.

2U Rack Server Roadmap - AMD 9005/9004 Series

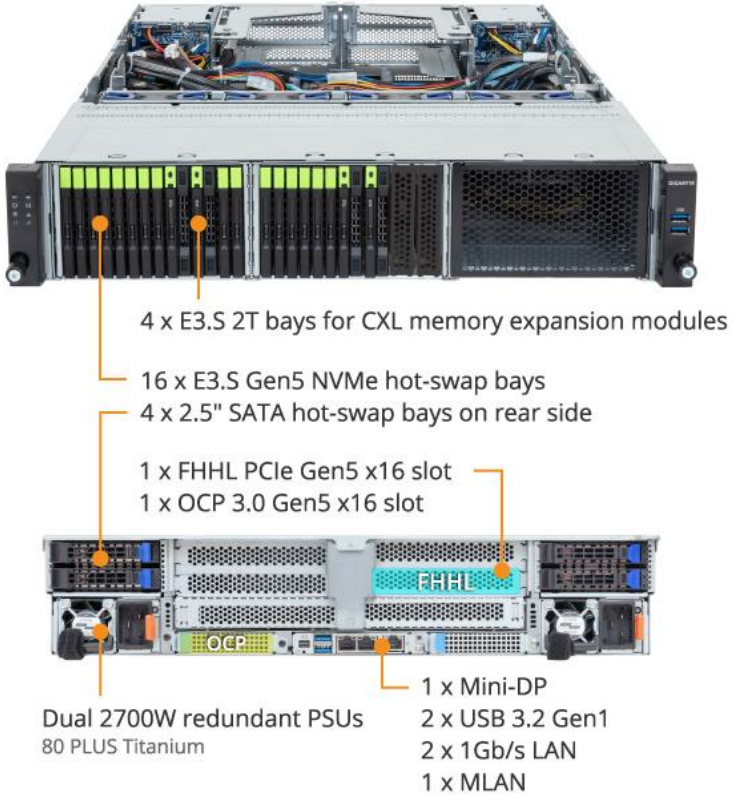
AMD

		2025Q4	2026Q1	2026Q2	2026Q3	2026Q4
DP	Performance Large Memory	R283-ZK0	2 OCP3 & 4 G5 slots, 2.7KW, 12 x 2.5” Bay, 48 DIMM			
	GPU	R283-ZF0	1 OCP3 & 1 G5 slots, 2.7KW, 12 x 3.5” Bay, 4 GPU			
		R283-ZF1	1 OCP3 & 1 G5 slots, 2.7KW, 24 x 2.5” Bay, 4 GPU			
		R283-Z93/ Z97	2 OCP3 & 2 G5 slots, 2.4-2.7KW, 12 x 3.5” Bay, 2 GPU			
	Mainstream Storage	R283-Z98	1 OCP3 & 1 G5 slots, 2.7KW, 16 x E3.S + 4 x CXL			
		R283-Z92	1-2 OCP3 & 2-4 G5 slots, 2KW, 16-24x 2.5” Bay + Expander			
	Cost Effective Cost	R283-Z91/ Z95	2 OCP3 & 5-6 G5 slots, 1.6KW, 12 x 3.5” Bay + Expander			
		R283-Z90/ Z94	1-2 OCP3 & 3-6 G5 slots, 1.6KW, 12 x 3.5” Bay			
UP	Performance GPU	R263-ZG0	1 OCP3 slots, 2.7KW, 12 x 3.5” Bay, 4 GPU			
		R263-ZG2	1 OCP3 slots, 2-2.7KW, 24 x 2.5” Bay, 4 GPU			
		R263-ZG5	1 OCP3 & 1 G5 slots, 2.7KW, 20 x E3.S Bay, 4-8 CXL			
	Mainstream Storage	R263-Z39	1 OCP3 & 1 G5 slots, 2KW, 20 x E3.S + 4 x CXL			
		R263-Z30	2 OCP3 & 2 G5 slots, 1.6KW, 12 x 3.5” Bay			
	Cost Effective Cost	R263-Z33	2 OCP3 slots, 2.4KW, 12 x 3.5” Bay, 3 GPU			

GIGABYTE

Left Edge: ES Right Edge: PD ▼:Launch/ Seeding Sample PRODUCTION COMMITTED PLANNING

R283-Z98-AAL1 Roadmap



CXL & E3.S support system | 2UDP

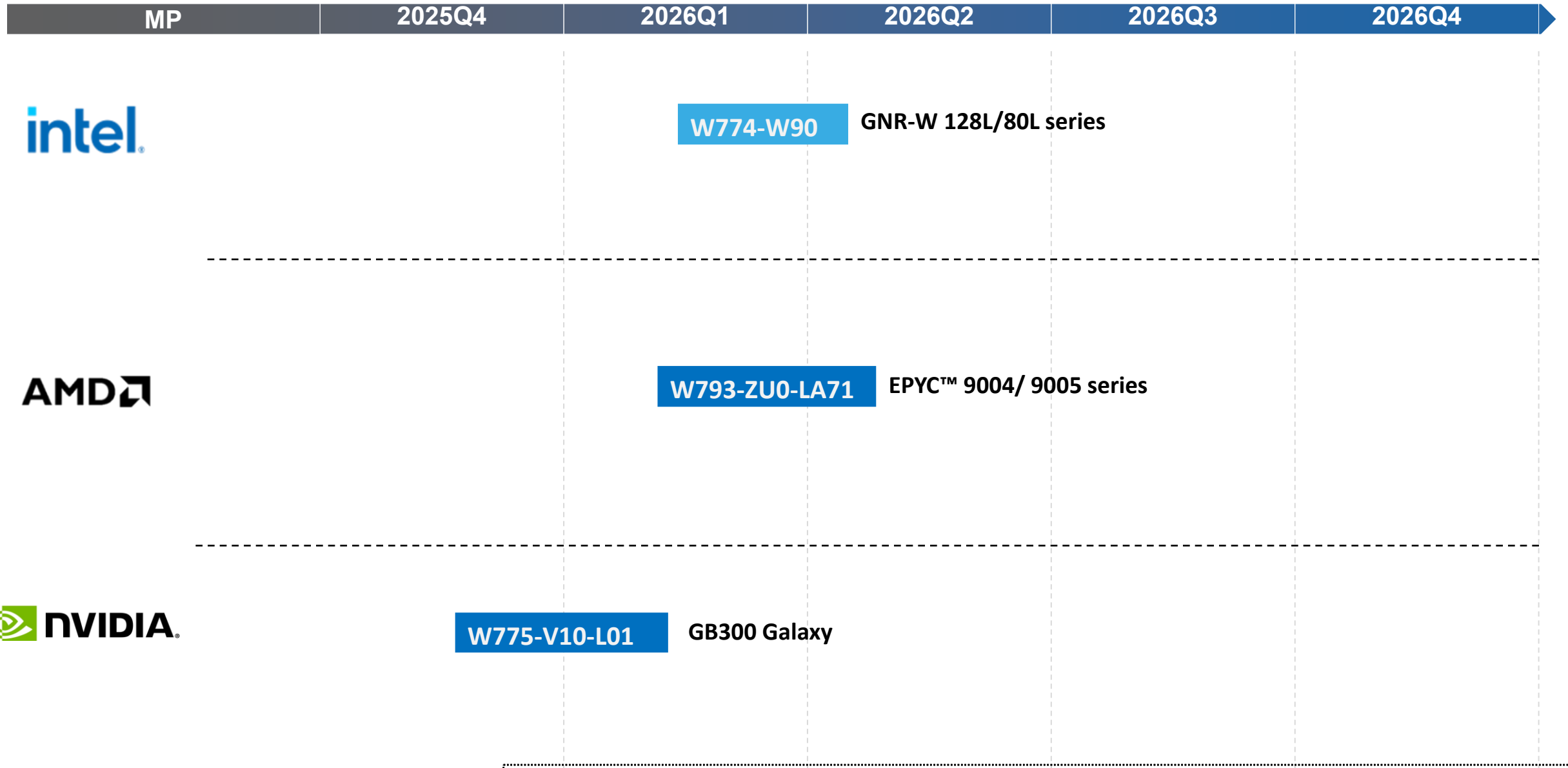
EPYC™ 9005/9004 Series, 1 FHHL G5 slots & OCP3.0

- **Dimensions: 447 x 87.5 x 815mm**
- **MB:** MZ93-FS1
- **Memory**
  - 24 x DIMM slots support 12 channel DDR5
  - 6400 MT/s (1DPC)
- **Expansion Slots**
  - 1 x FHHL Slot (@Gen5 x16)
  - 1 x OCP3.0 Slot (@Gen5 x16)
- **Network**
  - Intel i350 2 x 1GbE ports
- **Storage**
  - 16 x E3.S (@Gen5)
  - 4 x CXL
  - Rear 4 x 2.5" SAS/ SATA Bay
  - 2 x 22110 M.2 slots (@Gen3 x4)
  - 1 x 22110 M.2 slots (@Gen3 x2)
- **PSU**
  - (1+1) x 2700W 80 PLUS Titanium redundant power supplies

Std SKUs	GPU	Storage	Expansion Slot	PSU
R283-Z97-AAL1/ AAF1	2	8 x 3.5" U.2(@Gen5) + 4 x 3.5" SAS + 4 x 2.5" SAS(rear)	2 FHHL(@Gen5 x16) + 2 x OCP (@Gen5 x16)	2 x 2.4KW
R283-Z98-AAL1	-	16 x E3.S + 4 x CXL + 4 x 2.5" SAS + 2 x M.2	1 x FHHL(@Gen5 x16) + 1 x OCP3 (@Gen5 x16)	2 x 2.7KW

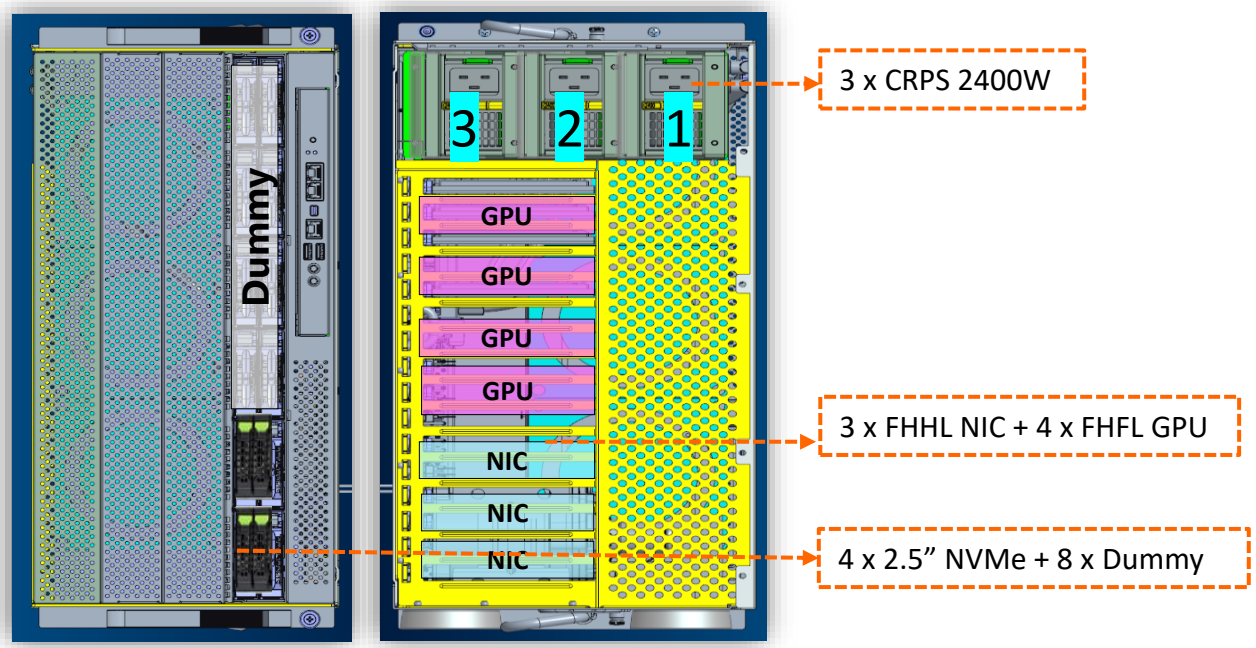
Std SKUs	Storage	Expansion Slot	PSU
R283-Z94-AAV1/ AAD1	(@Gen5) 12 x 3.5" U.2/ 4 x 2.5" SAS	2 x FHHL (@Gen5 x16) + 2 x OCP3	2 x 1.6KW
R283-Z94-AAV2/ AAD2	(@Gen5) 4 x 3.5" U.2 + 8 x 3.5" SAS/ 2 x 2.5" SAS	4 x FHHL (@Gen5 x16) + 2 x OCP3	2 x 1.6KW
R283-Z95-AAV1/ AAD1	(@Gen5) 4 x 3.5" U.2 + 8 x 3.5" SAS/ 2 x 2.5" SAS + Expander	5 x FHHL (@Gen5 x16) + 2 x OCP3	2 x 1.6KW
R283-Z96-AAJ1/ AAE1	(@Gen5) 24 x 2.5" U.2/ 4 x SAS + 2 x M.2	1 x FHHL (@Gen5 x16) + 1 x OCP3	2 x 2KW
R283-Z96-AAJ2/ AAE2	(@Gen5) 4 x 2.5" U.2 + 20 x SAS/ 4 x SAS + Expander	4 x FHHL (@Gen5 x16) + 1 x OCP3	2 x 2KW
R283-Z96-AAJ3/ AAE3	(@Gen5) 8 x 2.5" U.2 + 16 x SAS	3 x FHHL (@Gen5 x16) + 1 x OCP3	2 x 2KW

WS Roadmap

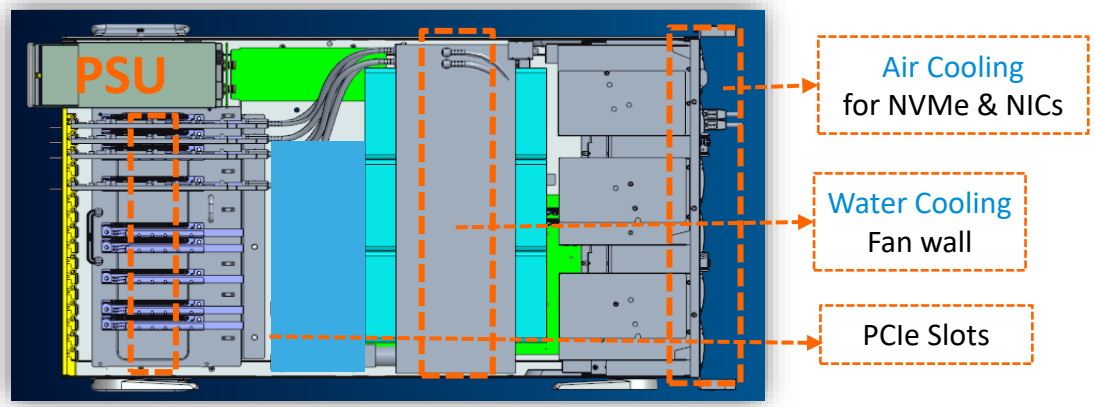

MP
2025Q4
2026Q1
2026Q2
2026Q3
2026Q4
intel
W774-W90
GNR-W 128L/80L series
AMD
W793-ZU0-LA71
EPYC™ 9004/ 9005 series
NVIDIA
W775-V10-L01
GB300 Galaxy
Left Edge: ES Right Edge: PD
:Launch/ Seeding Sample
PRODUCTION
COMMITTED
PLANNING

W793-ZU0-LA71 Roadmap

Front **Rear**



Side



UP, EPYC™ 9004/ 9005 series, Close Loop Liquid Cooling Solution

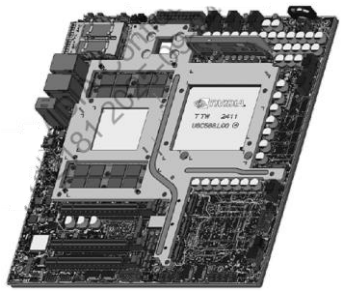
- **Dimensions (WxHxD): 449 x 218 x 780 mm**
- **Motherboard:**
  - **MZU3-MG0 (Micro MGX)**
- **Memory**
  - 12 x DIMM slots support 12 channel (1DPC) DDR5
  - DDR5 6000 MT/s (1DPC)
- **Expansion Slots**
  - 4 x PCIe (Gen5 x16) for GPU Dual Slot, for R9700S or MI250
  - 3 x PCIe (Gen5 x16) for LAN Dual Slot
- **Network**
  - 2 x 10 GbE ports (Intel X550-AT2)
  - 1 x Management LAN 10/100/1G
- **Storage**
  - 4 x 2.5” U.2
- **PSU:**
  - **3 x 2400W 80+ Titanium Redundant Power Supplies**

Standard SKUs	GPU	Storage	Expansion Slot	PSU
W793-ZU0-LA71	4	4 x 2.5” U.2	4 x PCIe (Gen5 x16) Dual for GPU 3 x PCIe (Gen5 x16) Dual for LAN	3 x 2400W

W775-V10-L01 Roadmap

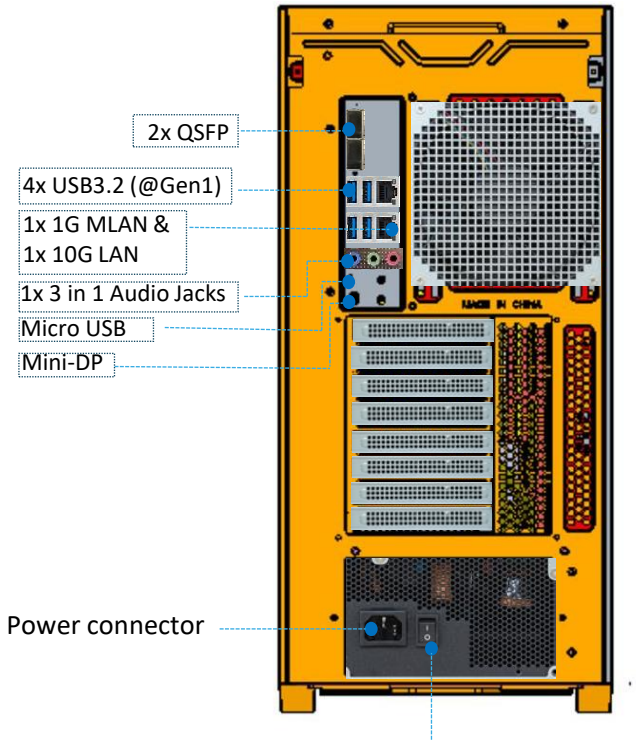
Nvidia GB300 Galaxy Workstation Close Loop Liquid Cooling Solution

Front View



GB300

Rear View



- **Dimensions (WxHxD): 245 x 500.4 x 531 mm**
- **Motherboard:**
  - **NVDIA GB300 Galaxy HPM (CEB Like)**
- **Memory**
  - 4 X 128GB SOCAMM LPDDR5X Modules (496GB total system memory)
- **Expansion Slots**
  - 1 x PCIe (Gen5 x16) for GPU Slot
  - 2 x PCIe (Gen5 x8) Slot
- **GPU Support (Option Part)**
  - RTX PRO 4000 Blackwell SFF Edition (ATX BKT)/ (LP BKT)
  - RTX PRO 2000 Blackwell (ATX BKT)/ (LP BKT)
- **Network**
  - 2 x 400GbE QSFP ports (from CX8)
  - 1 x 10GbE RJ45 port (Marvel AQC113)
  - 1 x Management LAN 10/100/1G RJ45
- **Storage**
  - 2 x M.2 (2242/2280) Gen5 x4 (OS RAID 1) from CPU (Option Part)
  - 2 x M.2 (2242/2280) Gen6 x4 (Training data SSD) from CX8 (Option Part)
- **Rear I/O Connector**
  - 1x Mini DP, 2x QSFP, 1x 1G MLAN, 1x 10G LAN, 4x USB3.2 (@Gen1)
  - 1x Micro USB (serial over), 1x 3 in 1 Audio Jacks
- **PSU:**
  - ATX 1600W

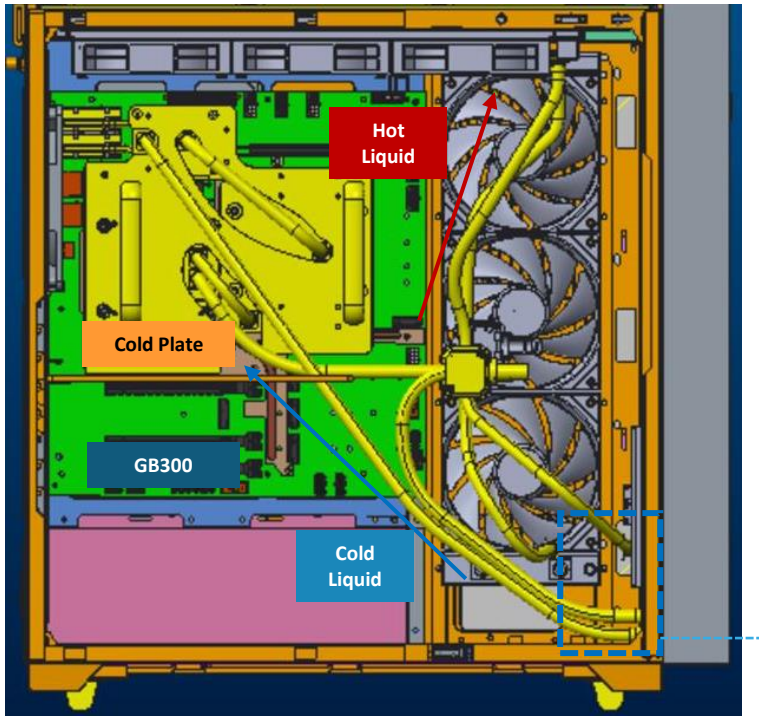
  (This system requires utility AC power within the 115V-240V range)

Standard SKUs	Storage	Expansion Slot	PSU
W775-V10-L01	No extra support	1 x PCIe (Gen5 x16) for GPU + 2 x PCIe (Gen5 x8)	ATX 1600W

Notes: 因Nvidia 尚未launch, 如要提供此規格給客戶,請先和PL Canon & Garry說明並取得同意。

W775-V10-L01 Liquid Cooling Solution

Side View



Liquid Cooling Solution

- A fully liquid-cooled architecture, including MB, LAN etc.
- Close Loop design

Noise Limitation

- **Due to a liquid-cooled architecture,** the WS can operate **below 50 dB**, suitable for a quiet office environment.

W775-V10-L01 Key Features

Design Purpose

- Specifically designed **for NVIDIA GB300 Galaxy**, to show a **fully NVIDIA-based system** and with the commitment to a **long-term collaboration with NVIDIA** .

Target Market

- Enable **SMBs** as well as **advanced consumers** to access and utilize **NVIDIA GB300 Galaxy systems** without requiring enterprise-scale servers.

Operating System

- **Linux**

S463-Z30-AAB1 Series Roadmap

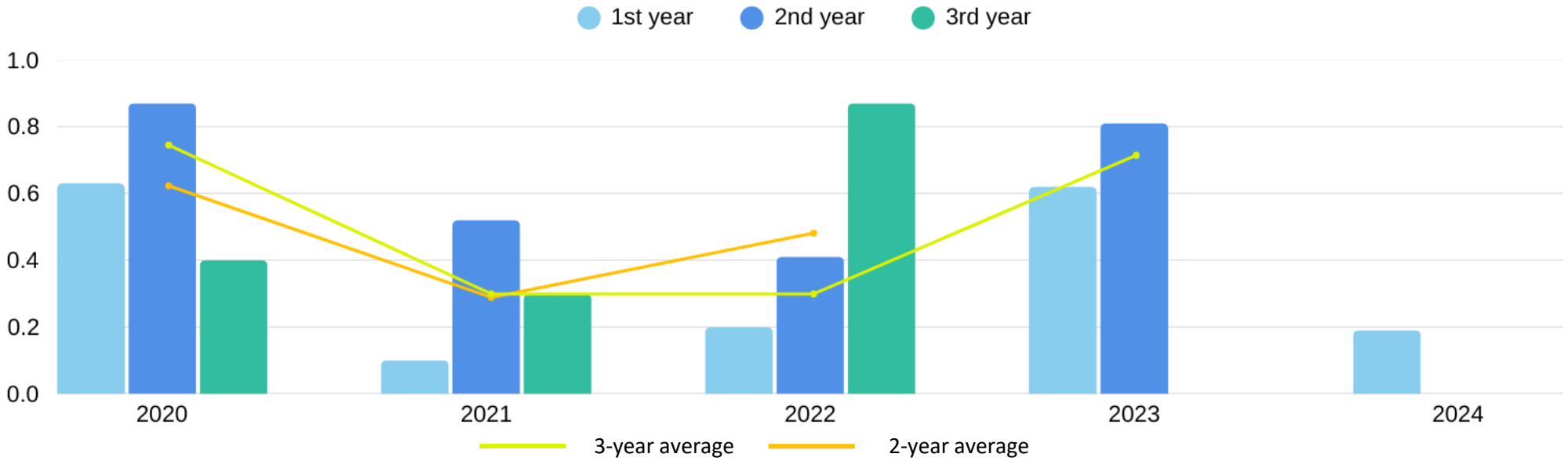
Side-accessed High density Storage 60 Bay server | 4U UP

EPYC™ 9005/ 9004-series, 4 LP Slot, 60 Bay

- **Dimensions: 434 x 176 x 853mm**
- **MB:** MZ33-AR1
- **Memory**
  - 24 x DIMM slots
  - DDR5 memory supported
  - 12-Channel memory per processor
  - **AMD EPYC™ 9005:**
    - RDIMM: Up to 4800 MT/s (1DPC)
    - RDIMM: Up to 4000 MT/s (1R 2DPC), 3600 MT/s (2R 2DPC)
  - **AMD EPYC™ 9004:**
    - RDIMM: Up to 4800 MT/s (1DPC), 3600 MT/s (2DPC)
- **Expansion Slots**
  - 4 x FHHL Slot (@Gen5 x16)
- **Network**
  - 2 x 10Gb/s LAN (BCM57416)
- **System FAN**
  - 4 x 8038 Hot-swap fans
- **Storage**
  - 60 x 3.5” SATA
  - 4 x 2.5” U.2 (@Gen5)
  - 2 x 2.5” SATA
  - 1 x M.2 (@Gen4 x4)
- **PSU**
  - (2+2) x 800W 80 PLUS Platinum redundant power supplies

Standard SKUs	Storage	Expansion Slot	PSU
S463-Z30-AAB1	60 x 3.5” + 4 x 2.5” U.2 + 2 x 2.5 SATA	4 x FHHL (@Gen5 x16)	2 x 800W

Track Record of Quality



	2020	2021	2022	2023	2024
1~365 days	0.63%	0.10%	0.20%	0.62%	0.19%
366~730 days	0.87%	0.52%	0.41%	0.81%	-
731~1095 days	0.40%	0.30%	0.87%	-	-
1-3 years return rate	1.9%	0.91%	1.48%	1.43%	0.19%

Three-year Return Rate: < 2%

- Product repair rate in 1st year: 0.1-0.63%
- Product repair rate in 2nd year: 0.65%
- Product repair rate in 3rd year: 0.52%
- In the industry, the general three-year repair rate is about 3% to 4.5%.

Thank you

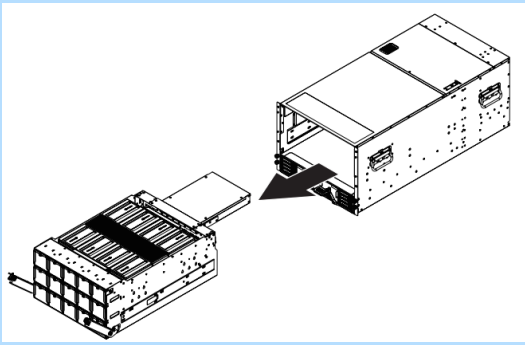
GCT HGX B200 to HGX B300 Networking Transition

HGX B200 ConnectX-7

Features

- Supports **1 x 400 Gbps** connection / SXM GPU
- HGX H100/ H200/ B200 has **NO** networking on the GPU baseboard.

HGX B300 ConnectX-8 SuperNIC



Features

- Highest-performance **800G** networking, designed for massive-scale AI.
- HGX B300 **HAS** networking on the GPU baseboard.
- High bandwidth, low latency, and improved GPU-to-GPU training performance.

Specifications

Supported Network	• InfiniBand • Ethernet
InfiniBand Speeds	800 Gbps
Ethernet Speeds	2 x 400 GbE
Per SXM GPU	• Physically presented as a single **OSFP port** • Logically presented as **2 x 400GbE** connections / GPU

W775-V10-L01 Configurations Notes

M.2 Config. Options:

Category	Description	Partner	Part Number
M.2	Gen 5 NVMe SSD	Samsung	9M9E1 (Gen5) MZVLC2T0HBLD-00B07 (2TB)
			PM9E3 (Gen5) MZVLC1T9HBLD-00B07 (2TB)
		Sandisk	SN8050S (Gen5) SDFQNJL-2T00 (2TB)
			SN8050S (Gen5) SDFQNJL-1T92(2TB)
		Micron	4600 (Gen5) MTFDLBA2T0THJ-1BP15ABYY (2TB)

GPU Config. Options:

RTX PRO 4000 Blackwell SDD Edition – (ATX BKT) / (LP BKT
RTX PRO 2000 Blackwell (ATX BKT)/ (LP BKT)

Wifi Options:

Category	Description	Partner	Part Number
M.2 Wifi/BT	ARM Colmpatible	Intel	AX210
		Azurewave	AW-EB600 (w/ MT7925-Wifi7)

Options QSFP Transceivers: 2 x QSFP Transceiver per HPM

Market Trends

- AI workloads are exploding
- Edge computing is growing rapidly
- Hybrid infrastructure is the new standard

Core Data Center

- AI training clusters
- HPC workloads
- High-density GPU servers

Connectivity

- High-speed networking
- RDMA / GPUDirect
- Efficient data movement

Software Stack

- Kubernetes
- AI frameworks
- Optimized GPU software stack

Security & Management

- Remote management (BMC/Redfish)
- Secure boot & TPM
- Monitoring tools

Use Case

- Smart factory: edge inference + core training
- AI data center deployment

Our Value

- Complete portfolio
- Strong ecosystem partnerships
- Customization capability