GIGABYTE | NVIDIA

NVIDIA-based Enterprise Solutions

AI Optimized Solutions for Accelerated Computing

2026 1H

NVIDIA Vera Rubin NVL72

Six New Chips, One AI Supercomputer

NVIDIA Vera Rubin NVL72 unifies leading-edge technologies from NVIDIA: 72 Rubin GPUs, 36 Vera CPUs, ConnectX®-9 SuperNICs, and BlueField®-4 DPUs. It scales up intelligence in a rack-scale platform with the NVIDIA NVLink™ 6 switch and scales out with NVIDIA Quantum-X800 InfiniBand and Spectrum-X™ Ethernet to power the AI industrial revolution at scale.

Built on the third-generation NVIDIA MGX™ NVL72 rack design, Vera Rubin NVL72 offers a seamless transition from prior generations. It delivers AI training with one-fourth the GPUs and AI inference at one-tenth the cost per million tokens versus NVIDIA Blackwell. Featuring cable-free modular tray designs and support from over 80 MGX ecosystem partners, the rack-scale AI supercomputer delivers world-class performance with rapid deployment.

	NVIDIA Vera Rubin NVL72	NVIDIA Vera Rubin Superchip	NVIDIA Rubin GPU
Configuration	72 NVIDIA Rubin GPUs | 36 NVIDIA Vera CPUs	2 NVIDIA Rubin GPUs | 1 NVIDIA Vera CPU	1 NVIDIA Rubin GPU
NVFP4 Inference	3,600 PFLOPS	100 PFLOPS	50 PFLOPS
NVFP4 Training[2]	2,520 PFLOPS	70 PFLOPS	35 PFLOPS
FP8/FP6 Training[2]	1,260 PFLOPS	35 PFLOPS	17.5 PFLOPS
INT8[2]	18 POPS	0.5 POPS	0.25 POPS
FP16/BF16[2]	288 PFLOPS	8 PFLOPS	4 PFLOPS
TF32[2]	144 PFLOPS	4 TFLOPS	2 TFLOPS
FP32	9,360 TFLOPS	260 TFLOPS	130 TFLOPS
FP64	2,400 TFLOPS	67 TFLOPS	33 TFLOPS
FP32 SGEMM[3]	28,800 TFLOPS	800 TFLOPS	400 TFLOPS
FP64 DGEMM[3]	14,400 TFLOPS	400 TFLOPS	200 TFLOPS
GPU Memory | Bandwidth	20.7 TB HBM4 | 1,580 TB/s	576 GB HBM4 | 44 TB/s	288 GB HBM4 | 22 TB/s
NVLink Bandwidth	260 TB/s	7.2 TB/s	3.6 TB/s
NVLink-C2C Bandwidth	65 TB/s	1.8 TB/s	-
CPU Core Count	3,168 custom NVIDIA Olympus cores (Arm® compatible)	88 custom NVIDIA Olympus cores (Arm® compatible)	-
CPU Memory	54 TB LPDDR5X	1.5 TB LPDDR5X	-
Total NVIDIA + HBM4 Chips	1,296	30	12

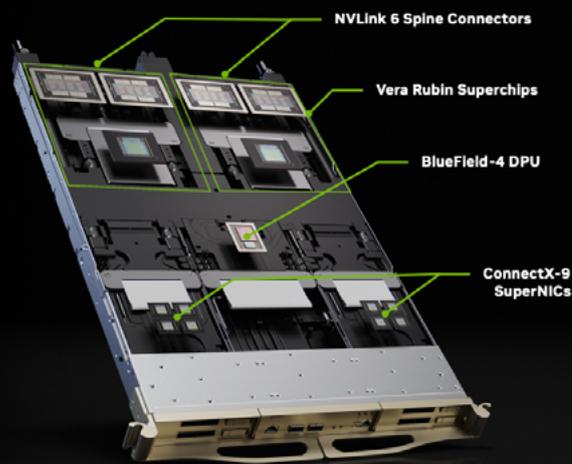1. Preliminary information. All values are up to and subject to change.
2. Dense specification.
3. Peak performance using Tensor Core-based emulation algorithms.

NVIDIA Vera Rubin NVL72

NVFP4 Inference	3.6 EFLOPS	**5x Blackwell**
NVFP4 Training	2.5 EFLOPS	**3.5x**
LPDDR5X Capacity	54 TB	**2.5x**
HBM4 Capacity	20.7 TB	**1.5x**
HBM4 Bandwidth	1.6 PB/s	**2.8x**
Scale-Up Bandwidth	260 TB/s	**2x**

NVLink 6 Spine Connectors
Vera Rubin Superchips
BlueField-4 DPU
ConnectX-9 SuperNICs

NVIDIA HGX™ Rubin NVL8

GPU	8x NVIDIA Rubin GPUs
Total GPU Memory | Bandwidth	2.3 TB | 160 TB/s
Performance	NVFP4 Inference: 400 PF NVFP4 Training: 280 PF FP8/FP6 Training: 140 PF
CPU	2x Intel® Xeon® 6 processors
NVIDIA NVLink Switch System	4x
NVIDIA NVLink Bandwidth	28.8 TB/s total bandwidth
System Power Usage	~24 kW
Networking	8x OSFP ports serving 8x single-port NVIDIA® ConnectX®-9 VPI > up to 800 Gb/s NVIDIA InfiniBand and Ethernet 2x 400G QSP112 NVIDIA BlueField®-4 DPUs > up to 800 Gb/s NVIDIA InfiniBand and Ethernet
Software	Ubuntu, Red Hat Enterprise Linux, Rocky

NVIDIA GB300 NVL72

DLB2-CB3

Built for the Age of AI Reasoning

A fully liquid-cooled, rack-scale design optimized for test-time scaling inference, delivering up to 50× higher output for reasoning model inference compared to the NVIDIA Hopper™ platform, based on NVIDIA Quantum-X800 InfiniBand or Spectrum™-X Ethernet paired with ConnectX®-8 SuperNIC™.
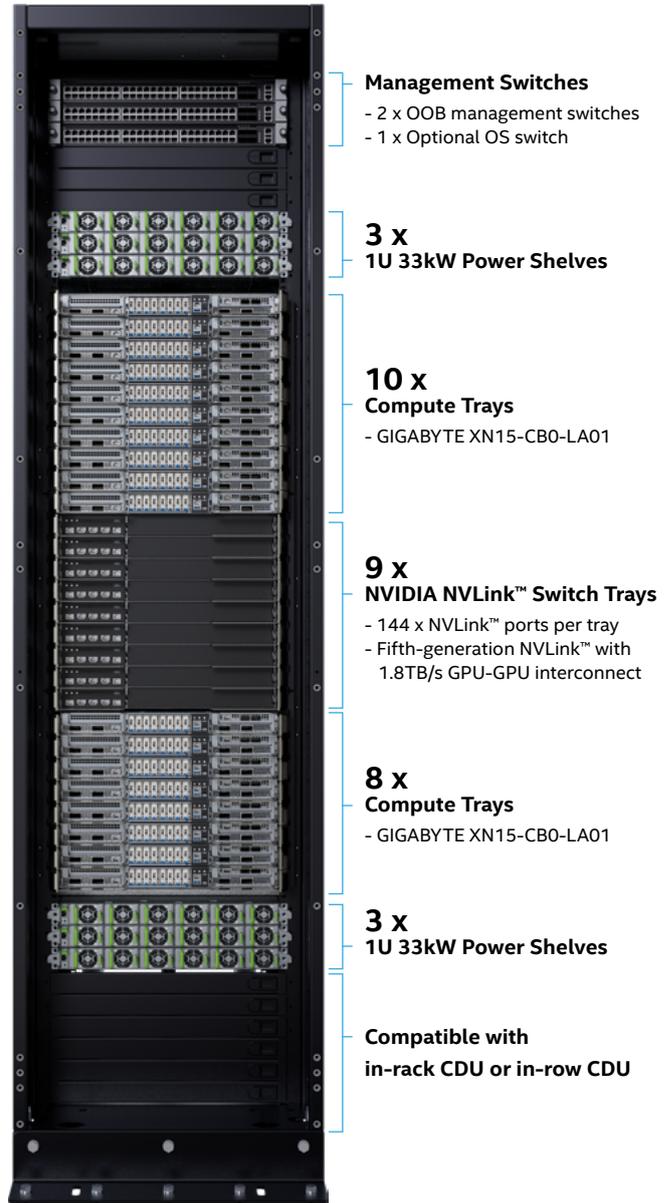
Fast Memory	**FP4 Inference**
60X	**70X**
vs. NVIDIA HGX H100	vs. NVIDIA HGX H100

HBM Bandwidth	**Networking Bandwidth**
20X	**18X**
vs. NVIDIA HGX H100	vs. NVIDIA HGX H100

- 36 NVIDIA Grace™ CPUs
- 72 NVIDIA Blackwell Ultra GPUs
- 17 TB of LPDDR5X ECC memory with up to 14 TB/s bandwidth
- 20 TB of HBM3E with up to 576 TB/s bandwidth
- Up to 37 TB of fast-access memory
- NVIDIA NVLink™ domain: 130 TB/s of low-latency GPU communication

Management Switches
- 2 x OOB management switches
- 1 x Optional OS switch

3 x
1U 33kW Power Shelves

10 x
Compute Trays
- GIGABYTE XN15-CB0-LA01

9 x
NVIDIA NVLink™ Switch Trays
- 144 x NVLink™ ports per tray
- Fifth-generation NVLink™ with 1.8TB/s GPU-GPU interconnect

8 x
Compute Trays
- GIGABYTE XN15-CB0-LA01

3 x
1U 33kW Power Shelves

Compatible with
in-rack CDU or in-row CDU

XN15-CB0-LA01
Compute Tray

Form Factor	1U Liquid-cooled server node
CPU	2x NVIDIA Grace™ CPU (72 Arm Neoverse V2 cores)
GPU	4 x NVIDIA Blackwell Ultra GPUs
CPU Memory	480GB LPDDR5X per CPU (960GB per tray)
GPU Memory	279GB HBM3E per GPU (1116GB per tray)
Storage	8 x E1.S Gen5 NVMe drive bays 1 x M.2 (2280/22110; PCIe 5.0 x4)
Networking	4 x 800Gb/s OSFP (NVIDIA ConnectX®-8 SuperNIC™) 1 x NVIDIA® BlueField®-3 DPU 1 x 1Gb/s LAN (Intel® I210-AT)
Front IO	1 x USB 3.2 Gen1, 1 x Mini-DP, 1 x RJ45, 1 x MLAN
Rear IO	4 x NVIDIA NVLink™ Switch connectors
OS Support	Ubuntu 22.04.3 arm64 Red Hat Enterprise Linux Server 9.3 aarch64 SUSE Linux Enterprise Server 15 SP5 aarch64
System Cooling	8 x 40x40x56mm fans 2 x Superchip cold plate loops

NVIDIA GB200 NVL72

DLB2-CB1

Powering the Era of Accelerated Computing

A fully liquid-cooled rack-scale design that interconnects all nodes with NVIDIA NVLink™ technology and delivers the performance of "one big GPU," surpassing previous-generation GPU platforms with exceptional interconnect bandwidth and energy efficiency for AI and HPC workloads.
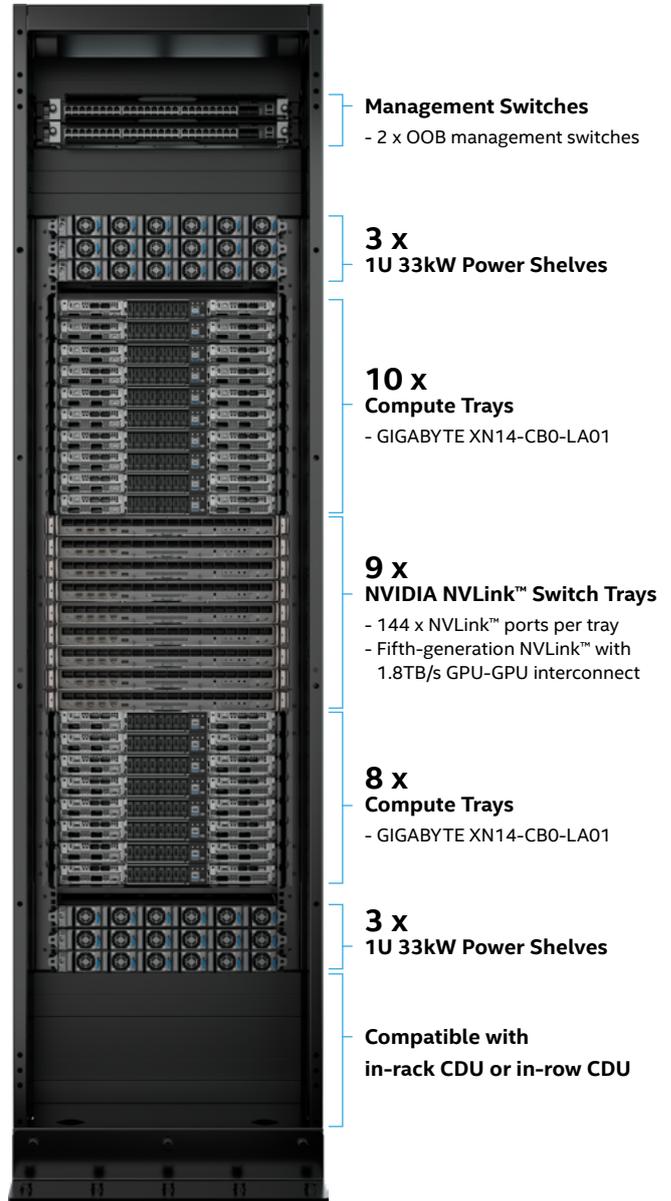
LLM Inference	LLM Training
30X	**4X**
vs. NVIDIA HGX H100	vs. NVIDIA HGX H100

Energy Efficiency	Data Processing
25X	**18X**
vs. NVIDIA HGX H100	vs. x86 CPU

- 36 NVIDIA Grace™ CPUs
- 72 NVIDIA Blackwell GPUs
- Up to 17TB of LPDDR5X memory with ECC
- Supports up to 13.5 TB of HBM3E
- Up to 30.5TB of fast-access memory
- 5th Gen NVIDIA NVLink™ technology
- NVIDIA NVLink™ domain: 130 TB/s of low-latency GPU communication

Management Switches
- 2 x OOB management switches

3 x
1U 33kW Power Shelves

10 x
Compute Trays
- GIGABYTE XN14-CB0-LA01

9 x
NVIDIA NVLink™ Switch Trays
- 144 x NVLink™ ports per tray
- Fifth-generation NVLink™ with 1.8TB/s GPU-GPU interconnect

8 x
Compute Trays
- GIGABYTE XN14-CB0-LA01

3 x
1U 33kW Power Shelves

Compatible with in-rack CDU or in-row CDU

XN14-CB0-LA01
Compute Tray

Form Factor	1U Liquid-cooled server node
CPU	2x NVIDIA Grace™ CPU (72 Arm Neoverse V2 cores)
GPU	4 x NVIDIA Blackwell GPUs
CPU Memory	480GB LPDDR5X per CPU (960GB per tray)
GPU Memory	192GB HBM3E per GPU (768GB per tray)
Storage	8 x E1.S Gen5 NVMe drive bays 1 x M.2 (2280/22110; PCIe 5.0 x4)
Networking	4 x 400Gb/s OSFP (NVIDIA ConnectX®-7 NIC) 2 x NVIDIA® BlueField®-3 DPUs 1 x 1Gb/s LAN (Intel® I210-AT)
Front IO	1 x USB 3.2 Gen1, 1 x Mini-DP, 1 x RJ45, 1 x MLAN
Rear IO	4 x NVIDIA NVLink™ Switch connectors
OS Support	Ubuntu 22.04.3 arm64 Red Hat Enterprise Linux Server 9.3 aarch64 SUSE Linux Enterprise Server 15 SP5 aarch64
System Cooling	8 x 40x40x56mm fans 2 x Superchip cold plate loops

NVIDIA GB300 Grace Blackwell Ultra Desktop Superchip

W775-V10-L01

The Ultimate Deskside AI Supercomputer

To meet the demands of large-scale AI models, this platform features the NVIDIA GB300 Grace Blackwell Ultra Desktop Superchip and a massive 748 GB of coherent memory. This breakthrough architecture delivers an unprecedented amount of compute performance, enabling AI development, research, and data science teams to develop and run large-scale AI training and inferencing workloads directly from the deskside.

Form Factor	Pedestal (W218 x H519.5 x D726.3 mm)
CPU	1 x NVIDIA Grace™ CPU
GPU	1 x NVIDIA Blackwell Ultra
CPU Memory	496GB LPDDR5X with bandwidth up to 396GB/s
GPU Memory	252GB HBM3E with bandwidth up to 7.1TB/s
Networking	2 x 400Gb/s QSFP ports via NVIDIA ConnectX®-8 SuperNIC™
Power Supply	Single 1600W ATX 80+ Gold PSU

NVIDIA GB200 NVL4

XN24-VC0-LA61

Form Factor	2U Liquid-cooled server (W438 x H87 x D900 mm)
CPU	1 x NVIDIA Grace™ CPU
GPU	2 x NVIDIA Blackwell
CPU Memory	480GB LPDDR5X with ECC per CPU Memory bandwidth up to 512GB/s
GPU Memory	186GB HBM3E per GPU Memory bandwidth up to 8TB/s
Storage	8 x 2.5" Gen5 NVMe 4 x 2.5" Gen5 NVMe via optional NVIDIA® BlueField®-3 DPU 1 x M.2 (2242/2260/2280/22110), PCIe 5.0 x4
Networking	4 x 800Gb/s OSFP (NVIDIA ConnectX®-8 SuperNIC™) 1 x 1Gb/s LAN (Intel® I210-AT) 1 x MLAN
Rear IO	4 x OSFP, 2 x USB 3.2 Gen1, 1 x Mini-DP, 1 x RJ45, 1 x MLAN

NVIDIA Grace™ CPU Superchip

Model	XV23-VC0-AAJ1/AAJ2	H263-V60-AAW1	H263-V60-LAW1
Motherboard	MVC3-MG0	MV63-HD0	MV63-HD0
Form Factor	2U MGX server (W438 x H87 x D900 mm)	2U 4-node rear access server (W440 x H87.5 x D850 mm)	2U 4-node rear access DLC server (W440 x H87.5 x D850 mm)
Superchip	NVIDIA Grace™ CPU Superchip: - 2 x NVIDIA Grace™ CPUs - Connected with NVIDIA NVLink™-C2C - TDP up to 500W (CPU + memory)	NVIDIA Grace™ CPU Superchip: - 2 x NVIDIA Grace™ CPUs - Connected with NVIDIA NVLink™-C2C - TDP up to 500W (CPU + memory)	NVIDIA Grace™ CPU Superchip: - 2 x NVIDIA Grace™ CPUs - Connected with NVIDIA NVLink™-C2C - TDP up to 500W (CPU + memory)

Memory	**[AAJ1]** 480GB of LPDDR5X with ECC
	    Memory bandwidth up to 1TB/s
	[AAJ2] 960GB of LPDDR5X with ECC
	    Memory bandwidth up to 768GB/s
		Per node:
		480GB of LPDDR5X memory with ECC*
		Memory bandwidth up to 1TB/s*
			Per node:
			480GB of LPDDR5X memory with ECC*
			Memory bandwidth up to 1TB/s*

Networking	1 x MLAN	1 x MLAN 1 x CMC	1 x MLAN 1 x CMC
Storage	2 x 2.5" Gen5 NVMe 2 x M.2 (PCIe 5.0 x4)	16 x 2.5" Gen5 NVMe 8 x M.2 (PCIe 5.0 x4)	16 x 2.5" Gen5 NVMe 8 x M.2 (PCIe 5.0 x4)
Expansion Slots	4 x FHFL PCIe 5.0 x16 for GPUs 2 x FHFL PCIe 5.0 x16 for add-in cards - Compatible with NVIDIA® BlueField®-3 DPUs	8 x LP PCIe 5.0 x16 4 x OCP NIC 3.0 (PCIe 5.0 x16)	4 x FHHL PCIe 5.0 x16 4 x OCP NIC 3.0 (PCIe 5.0 x16)
I/O Ports	Front: 2 x USB 3.2 Gen1 1 x Mini-DP 1 x MLAN	Rear: 8 x USB 3.2 Gen1 4 x VGA 4 x MLAN 1 x CMC	Rear: 8 x USB 3.2 Gen1 4 x VGA 4 x MLAN 1 x CMC
Security	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012
Power Supply	2+2 redundant PSUs 2000W 80+ Titanium	2+1 redundant PSUs 3000W 80+ Titanium	2+1 redundant PSUs 3000W 80+ Titanium
System Management	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC ASPEED® AST2520 CMC GIGABYTE Management Console	ASPEED® AST2600 BMC ASPEED® AST2520 CMC GIGABYTE Management Console
Other Features	Supports 4 x dual-slot Gen5 GPUs	-	Direct liquid cooling with leak detection

* Modules with 960GB of CPU memory and 768GB/s memory bandwidth are also available. Please contact our sales representatives for more details.

NVIDIA Grace Hopper™ Superchip

Model	XH23-VG0-AAJ1	H223-V10-AAW1	H263-V11-LAW1
Motherboard	MVG3-MG0	MV13-HD0	MV13-HD0
Form Factor	2U MGX server (W438 x H87 x D900 mm)	2U 2-node rear access server (W440 x H87.5 x D850 mm)	2U 4-node rear access server (W440 x H87.5 x D850 mm)
Superchip	NVIDIA GH200 Grace Hopper™ Superchip: - 1 x NVIDIA Grace™ CPU - 1 x NVIDIA Hopper™ GPU - Connected with NVIDIA NVLink™-C2C - TDP up to 1000W (CPU + GPU + memory)	NVIDIA GH200 Grace Hopper™ Superchip: - 1 x NVIDIA Grace™ CPU - 1 x NVIDIA Hopper™ GPU - Connected with NVIDIA NVLink™-C2C - TDP up to 1000W (CPU + GPU + memory)	NVIDIA GH200 Grace Hopper™ Superchip: - 1 x NVIDIA Grace™ CPU - 1 x NVIDIA Hopper™ GPU - Connected with NVIDIA NVLink™-C2C - TDP up to 1000W (CPU + GPU + memory)
Memory	NVIDIA Grace™ CPU: - 480GB of LPDDR5X memory with ECC - Memory bandwidth up to 512GB/s NVIDIA Hopper™ GPU: - 96GB HBM3* - Memory bandwidth up to 4TB/s*	NVIDIA Grace™ CPU: - 480GB of LPDDR5X memory with ECC - Memory bandwidth up to 512GB/s NVIDIA Hopper™ GPU: - 96GB HBM3* - Memory bandwidth up to 4TB/s*	NVIDIA Grace™ CPU: - 480GB of LPDDR5X memory with ECC - Memory bandwidth up to 512GB/s NVIDIA Hopper™ GPU: - 96GB HBM3* - Memory bandwidth up to 4TB/s*
Networking	2 x 10Gb/s LAN 1 x MLAN	4 x 10Gb/s LAN 2 x MLAN 1 x CMC	8 x 10Gb/s LAN 4 x MLAN 1 x CMC
Storage	4 x 2.5" Gen5 NVMe 2 x M.2 (PCIe 5.0 x4)	8 x 2.5" Gen5 NVMe Optional 4 x NVMe via BlueField-3 DPUs** 4 x M.2 (PCIe 5.0 x4)	16 x 2.5" Gen5 NVMe Optional 8 x NVMe via BlueField-3 DPUs** 8 x M.2 (PCIe 5.0 x4)
Expansion Slots	3 x FHFL PCIe 5.0 x16 - Compatible with NVIDIA® BlueField®-3 DPUs	4 x FHHL PCIe 5.0 x16 2 x OCP NIC 3.0 (PCIe 5.0 x16)	4 x FHHL PCIe Gen5 x16 4 x OCP NIC 3.0 (PCIe 5.0 x16)
I/O Ports	Front: 2 x USB 3.2 Gen1 1 x Mini-DP 2 x RJ45 1 x MLAN	Rear: 4 x USB 3.2 Gen1 2 x Mini-DP 4 x RJ45 2 x MLAN 1 x CMC	Rear: 8 x USB 3.2 Gen1 4 x Mini-DP 8 x RJ45 4 x MLAN 1 x CMC
Security	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012
Power Supply	1+1 redundant PSUs 2000W 80+ Titanium	2+1 redundant PSUs 3000W 80+ Titanium	Triple redundant PSU 3000W 80+ Titanium
System Management	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC ASPEED® AST2520 CMC GIGABYTE Management Console	ASPEED® AST2600 BMC ASPEED® AST2520 CMC GIGABYTE Management Console
Other Features	NVIDIA-Certified Systems™	-	Direct liquid cooling with leak detection

* Modules with 144GB of HBM3E GPU memory and 4.9TB/s memory bandwidth are also available.
  Please contact our sales representatives for more details.
** Requires optional NVIDIA® BlueField®-3 DPUs, with each DPU supporting two drives.

NVIDIA HGX™

Supercharging AI and High-Performance Computing for Every Data Center

AI, complex simulations, and massive datasets require multiple GPUs with extremely fast interconnections and a fully accelerated software stack. The NVIDIA HGX™ platform brings together the full power of NVIDIA GPUs, NVIDIA NVLink™, NVIDIA networking, and fully optimized AI and high-performance computing (HPC) software stacks to provide the highest application performance and drive the fastest time to insights for every data center.

To meet the need for robust thermal solutions driven by unprecedented performance and increased heat generation, GIGABYTE provides versatile server designs. These include 8U air-cooled and 4U liquid-cooled options, ensuring compatibility with diverse infrastructure requirements.

GIGABYTE servers are equipped with features designed to enhance both performance and usability, including:

1. Support for full-height add-in cards, accommodating DPUs and SuperNICs.
2. An easily removable PCIe cage design and front-access motherboard/GPU trays for streamlined maintenance.
3. Hot-swappable, fully redundant PSUs with multiple connector ptions for enhanced flexibility.

Other features include incorporating a PCIe switch to overcome PCIe lane limits and to enable fast GPU-GPU communication, ensuring GIGABYTE servers deliver reliable performance and efficiency for demanding AI and HPC workloads.
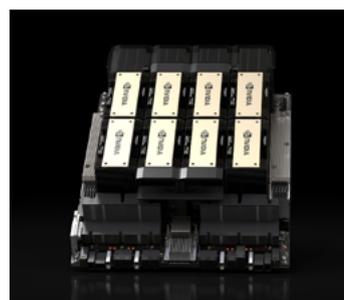
Model	HGX B300	HGX B200	HGX H200
Form Factor	8 x NVIDIA Blackwell Ultra SXM	8 x NVIDIA Blackwell SXM	8 x NVIDIA H200 SXM
FP4 Tensor Core**	144 PFLOPS | 108 PFLOPS	144 PFLOPS | 72 PFLOPS	-
FP8/FP6 Tensor Core*	72 PFLOPS	72 PFLOPS	32 PFLOPS
INT8 Tensor Core*	3 POPS	72 POPS	32 POPS
FP16/BF16 Tensor Core*	36 PFLOPS	36 PFLOPS	16 PFLOPS
TF32 Tensor Core*	18 PFLOPS	18 PFLOPS	8 PFLOPS
FP32	600 TFLOPS	600 TFLOPS	540 TFLOPS
FP64/FP64 Tensor Core	10 TFLOPS	296 TFLOPS	270 TFLOPS / 540 TFLOPS
Total Memory	2.1 TB	1.4 TB	1.1 TB
NVIDIA NVLink™	Fifth generation	Fifth generation	Fourth generation
NVIDIA NVLink Switch™	NVLink 5 Switch	NVLink 5 Switch	NVLink 4 Switch
GPU-to-GPU Bandwidth	1.8 TB/s	1.8 TB/s	900 GB/s
Total NVLink Bandwidth	14.4 TB/s	14.4 TB/s	7.2 TB/s

* With sparsity
** With sparsity | without sparsity

NVIDIA HGX™ B300

Air Cooling

Model	**G894-AD3-AAX7**	**G894-SD3-AAX7**	**G894-ZD3-AAX7**
Motherboard	MAB4-PE2	MSB4-PE2	MZB3-PE2
Form Factor	8U (W447 x H351 x D923 mm)	8U (W447 x H351 x D923 mm)	8U (W447 x H351 x D923 mm)
GPU	NVIDIA Blackwell Ultra 8-GPU	NVIDIA Blackwell Ultra 8-GPU	NVIDIA Blackwell Ultra 8-GPU
CPU	Intel® Xeon® 6900-Series Dual processor, TDP up to 500W	Intel® Xeon® 6700/6500-Series Dual processor, TDP up to 350W	AMD EPYC™ 9005/9004 series Dual processor, cTDP up to 500W
Chipset	System on Chip	System on Chip	System on Chip
Memory	12-Channel RDIMM/MRDIMM [1] 24 x DDR5 DIMMs	8-Channel RDIMM/MRDIMM [1] 32 x DDR5 DIMMs	12-Channel RDIMM 24 x DDR5 DIMMs
Networking	8 x 800 Gb/s OSFP InfiniBand XDR / Dual 400 Gb/s Ethernet [2] 2 x 10Gb/s LAN 1 x MLAN	8 x 800 Gb/s OSFP InfiniBand XDR / Dual 400 Gb/s Ethernet [2] 2 x 10Gb/s LAN (Intel® X710-AT2) 1 x MLAN	8 x 800 Gb/s OSFP InfiniBand XDR / Dual 400 Gb/s Ethernet [2] 2 x 10Gb/s LAN 1 x MLAN
Storage	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 5.0 x4) 1 x M.2 (PCIe 5.0 x2)	Front hot-swap: 8 x 2.5"Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 5.0 x4) 1 x M.2 (PCIe 5.0 x2)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 3.0 x4) 1 x M.2 (PCIe 3.0 x1)
Expansion Slots	4 x FHHL PCIe 5.0 x16 (dual-slot)	4 x FHHL PCIe 5.0 x16 (dual-slot)	4 x FHHL PCIe 5.0 x16 (dual-slot)
I/O Ports	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN 8 x OSFP	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN 8 x OSFP	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN 8 x OSFP
Security	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 module: CTM012	Optional TPM 2.0 module: CTM012
Power Supply	6+6 redundant PSUs 3000W 80+ Titanium	6+6 redundant PSUs 3000W 80+ Titanium	6+6 redundant PSUs 3000W 80+ Titanium
System Management	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console
Other Features	Dual ROM architecture	Dual ROM architecture	Dual ROM architecture

[1] MRDIMMs are only supported with Intel® Xeon® 6 Processors with P-cores and in a 1DPC configuration
[2] GPU networking ports via onboard NVIDIA ConnectX®-8 SuperNIC™

NVIDIA HGX™ B300

Direct Liquid Cooling

Model	**G4L4-SD3-LAX7**	**G4L4-ZD3-LAX7**
Motherboard	MSB4-PE3	MZB4-PE3
Form Factor	4U (W447 x H175.3 x D900 mm)	4U (W447 x H175.3 x D900 mm)
GPU	Liquid-cooled NVIDIA Blackwell Ultra 8-GPU	Liquid-cooled NVIDIA Blackwell Ultra 8-GPU
CPU	Intel® Xeon® 6700/6500-Series Dual processor, TDP up to 350W	AMD EPYC™ 9005/9004 series Dual processor, TDP up to 500W
Chipset	System on Chip	System on Chip
Memory	8-Channel RDIMM/MRDIMM [1] 32 x DDR5 DIMMs	12-Channel RDIMM 24 x DDR5 DIMMs
Networking	8 x 800 Gb/s OSFP InfiniBand XDR / Dual 400 Gb/s Ethernet [2] 2 x 10Gb/s LAN 1 x MLAN	8 x 800 Gb/s OSFP InfiniBand XDR / Dual 400 Gb/s Ethernet [2] 2 x 10Gb/s LAN 1 x MLAN
Storage	8 x 2.5" Gen5 NVMe 2 x M.2 (PCIe 5.0 x4 & x2)	8 x 2.5" Gen5 NVMe 2 x M.2 (PCIe 5.0 x4 & x2)
Expansion Slots	3 x FHHL PCIe 5.0 x16 (dual-slot) 1 x FHHL PCIe 5.0 x16 (single-slot)	3 x FHHL PCIe 5.0 x16 (dual-slot) 1 x FHHL PCIe 5.0 x16 (single-slot)
I/O Ports	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 8 x OSFP 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 8 x OSFP 1 x MLAN
Security	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012
Power Supply	5+5 redundant PSUs 3000W 80+ Titanium	5+5 redundant PSUs 3000W 80+ Titanium
System Management	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console
Other Features	Dual ROM architecture	Dual ROM architecture

[1] MRDIMMs are only supported with Intel® Xeon® 6 Processors with P-cores and in a 1DPC configuration
[2] GPU networking ports via onboard NVIDIA ConnectX®-8 SuperNIC™

NVIDIA HGX™ B200

Air Cooling

Model	G894-AD1-AAX5	G894-SD1-AAX5	G893-SD1-AAX5	G893-ZD1-AAX5
Motherboard	MAB4-PE0	MSB4-PE0	MSB3-PE0	MZB3-PE0
Form Factor	8U (W447 x H351 x D923 mm)	8U (W447 x H351 x D923 mm)	8U (W447 x H351 x D923 mm)	8U (W447 x H351 x D923 mm)
GPU	NVIDIA Blackwell GPU	NVIDIA Blackwell GPU	NVIDIA Blackwell GPU	NVIDIA Blackwell GPU
CPU	Intel® Xeon® 6900-Series Dual processor TDP up to 500W	Intel® Xeon® 6700/6500-Series Dual processor TDP up to 350W	5th/4th Gen Intel® Xeon® Scalable Dual processor TDP up to 350W	AMD EPYC™ 9005/9004 series Dual processor cTDP up to 500W
Chipset	System on Chip	System on Chip	Intel® C741	System on Chip
Memory	12-Channel RDIMM/MRDIMM [1] 24 x DDR5 DIMMs	8-Channel RDIMM/MRDIMM [1] 32 x DDR5 DIMMs	8-Channel RDIMM 32 x DDR5 DIMMs	12-Channel RDIMM 24 x DDR5 DIMMs
Networking	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN
Storage	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 5.0 x4) 1 x M.2 (PCIe 5.0 x2)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 5.0 x4) 1 x M.2 (PCIe 5.0 x2)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 3.0 x2) 1 x M.2 (PCIe 3.0 x1)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 3.0 x4) 1 x M.2 (PCIe 3.0 x1)
Expansion Slots	4 x FHHL dual-slot PCIe 5.0 x16 8 x FHHL single-slot PCIe 5.0 x16	4 x FHHL dual-slot PCIe 5.0 x16 8 x FHHL single-slot PCIe 5.0 x16	4 x FHHL dual-slot PCIe 5.0 x16 8 x FHHL single-slot PCIe 5.0 x16	4 x FHHL dual-slot PCIe 5.0 x16 8 x FHHL single-slot PCIe 5.0 x16
I/O Ports	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN
Security	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012
Power Supply	6+6 redundant PSUs 3000W 80+ Titanium	6+6 redundant PSUs 3000W 80+ Titanium	6+6 redundant PSUs 3000W 80+ Titanium	6+6 redundant PSUs 3000W 80+ Titanium
System Management	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console

[1] MRDIMMs are only supported with Intel® Xeon® 6 Processors with P-cores and in a 1DPC configuration

NVIDIA HGX™ B200

Direct Liquid Cooling

Model	G4L4-AD1-LAX5	G4L4-SD1-LAX5	G4L3-SD1-LAX5	G4L3-ZD1-LAX5
Motherboard	MAB4-PE0	MSB4-PE0	MSB3-PE0	MZB3-PE0
Form Factor	4U (W447 x H175.5 x D901 mm)	4U (W447 x H175.5 x D901 mm)	4U (W447 x H175.5 x D901 mm)	4U (W447 x H175.5 x D901 mm)
GPU	NVIDIA Blackwell GPU	NVIDIA Blackwell GPU	NVIDIA Blackwell GPU	NVIDIA Blackwell GPU
CPU	Intel® Xeon® 6900-Series Dual processor TDP up to 500W	Intel® Xeon® 6700/6500-Series Dual processor TDP up to 350W	5th/4th Gen Intel® Xeon® Scalable Dual processor TDP up to 385W	AMD EPYC™ 9005/9004 series Dual processor cTDP up to 500W
Chipset	System on Chip	System on Chip	Intel® C741	System on Chip
Memory	12-Channel RDIMM/MRDIMM [1] 24 x DDR5 DIMMs	8-Channel RDIMM/MRDIMM [1] 32 x DDR5 DIMMs	8-Channel RDIMM 32 x DDR5 DIMMs	12-Channel RDIMM 24 x DDR5 DIMMs
Networking	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN
Storage	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 5.0 x4) 1 x M.2 (PCIe 5.0 x2)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 5.0 x4) 1 x M.2 (PCIe 5.0 x2)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 3.0 x2) 1 x M.2 (PCIe 3.0 x1)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 3.0 x4) 1 x M.2 (PCIe 3.0 x1)
Expansion Slots	4 x FHHL dual-slot PCIe 5.0 x16 8 x FHHL single-slot PCIe 5.0 x16	4 x FHHL dual-slot PCIe 5.0 x16 8 x FHHL single-slot PCIe 5.0 x16	4 x FHHL dual-slot PCIe 5.0 x16 8 x FHHL single-slot PCIe 5.0 x16	4 x FHHL dual-slot PCIe 5.0 x16 8 x FHHL single-slot PCIe 5.0 x16
I/O Ports	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN
Security	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012
Power Supply	4+4 redundant PSUs 3000W 80+ Titanium	4+4 redundant PSUs 3000W 80+ Titanium	4+4 redundant PSUs 3000W 80+ Titanium	4+4 redundant PSUs 3000W 80+ Titanium
System Management	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console

[1] MRDIMMs are only supported with Intel® Xeon® 6 Processors with P-cores and in a 1DPC configuration

NVIDIA HGX™ H200

Air Cooling

Model	G893-SD1-AAX3	G893-ZD1-AAX3	G593-SD1-AAX3	G593-ZD1-AAX3
Form Factor	8U (W447 x H351 x D923 mm)	8U (W447 x H351 x D923 mm)	5U (W447 x H219.7 x D945 mm)	5U (W447 x H219.7 x D945 mm)
GPU	NVIDIA H200 8-GPU	NVIDIA H200 8-GPU	NVIDIA H200 8-GPU	NVIDIA H200 8-GPU
CPU	5th/4th Gen Intel® Xeon® Scalable Intel® Xeon® CPU Max Series, Dual processor, TDP up to 350W	AMD EPYC™ 9005/9004 series Dual processor cTDP up to 500W at ambient 30°C	5th/4th Gen Intel® Xeon® Scalable Intel® Xeon® CPU Max Series, Dual processor, TDP up to 350W	AMD EPYC™ 9005/9004 series Dual processor cTDP up to 500W at ambient 25°C
Memory	8-Channel RDIMM 32 x DDR5 DIMMs	12-Channel RDIMM 24 x DDR5 DIMMs	8-Channel RDIMM 32 x DDR5 DIMMs	12-Channel RDIMM 24 x DDR5 DIMMs
Networking	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN
Storage	Front hot-swap: 8 x 2.5" Gen5 NVMe/SATA Internal M.2: 1 x M.2 (PCIe 3.0 x2) 1 x M.2 (PCIe 3.0 x1)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 3.0 x4) 1 x M.2 (PCIe 3.0 x1)	Front hot-swap: 8 x 2.5"Gen5 NVMe/SATA/SAS4*	Front hot-swap: 8 x 2.5" Gen5 NVMe/SATA/SAS4** Internal M.2: 1 x M.2 (PCIe 3.0 x4) 1 x M.2 (PCIe 3.0 x1)
Expansion Slots	4 x FHHL PCIe 5.0 x16 (dual-slot) 8 x FHHL PCIe 5.0 x16 (single-slot)	4 x FHHL PCIe 5.0 x16 (dual-slot) 8 x FHHL PCIe 5.0 x16 (single-slot)	4 x FHHL PCIe 5.0 x16 8 x LP PCIe 5.0 x16	4 x FHHL PCIe 5.0 x16 8 x LP PCIe 5.0 x16
I/O Ports	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN
Security	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 module: CTM012	Optional TPM 2.0 module: CTM010	Optional TPM 2.0 module: CTM010
Power Supply	4+4 redundant PSUs 3000W 80+ Titanium	4+4 redundant PSUs 3000W 80+ Titanium	4+2 redundant PSUs 3000W 80+ Titanium	5+1 redundant PSUs 3000W 80+ Titanium
System Management	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console
Other Features	Dual ROM architecture	Dual ROM architecture	Dual ROM architecture	Dual ROM architecture

* SAS card is required to support SAS drives.
** Storage card is required to support SATA and SAS drives.

NVIDIA HGX™ H200

Direct Liquid Cooling

Model	G593-SD1-LAX3	G593-ZD1-LAX3	G4L3-SD1-LAX3	G4L3-ZD1-LAX3
Form Factor	5U (W447 x H219.7 x D945 mm)	5U (W447 x H219.7 x D945 mm)	4U (W447 x H175.5 x D901 mm)	4U (W447 x H175.5 x D901 mm)
GPU	NVIDIA H200 8-GPU	NVIDIA H200 8-GPU	NVIDIA H200 8-GPU	NVIDIA H200 8-GPU
CPU	5th/4th Gen Intel® Xeon® Scalable Intel® Xeon® CPU Max Series, Dual processor, TDP up to 385W	AMD EPYC™ 9005/9004 series Dual processor TDP up to 500W	5th/4th Gen Intel® Xeon® Scalable Intel® Xeon® CPU Max Series, Dual processor, TDP up to 385W	AMD EPYC™ 9005/9004 series Dual processor cTDP up to 500W
Memory	8-Channel RDIMM 32 x DDR5 DIMMs	12-Channel RDIMM 24 x DDR5 DIMMs	8-Channel RDIMM 32 x DDR5 DIMMs	12-Channel RDIMM 24 x DDR5 DIMMs
Networking	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN	2 x 10Gb/s LAN 1 x MLAN
Storage	Front hot-swap: 8 x 2.5" Gen5 NVMe/SATA/SAS4*	Front hot-swap: 8 x 2.5" Gen5 NVMe/SATA/SAS4** Internal M.2: 1 x M.2 (PCIe 5.0 x4) 1 x M.2 (PCIe 5.0 x2)	Front hot-swap: 8 x 2.5" Gen5 NVMe/SATA Internal M.2: 1 x M.2 (PCIe 3.0 x2) 1 x M.2 (PCIe 3.0 x1)	Front hot-swap: 8 x 2.5" Gen5 NVMe Internal M.2: 1 x M.2 (PCIe 3.0 x4) 1 x M.2 (PCIe 3.0 x1)
Expansion Slots	4 x FHHL PCIe 5.0 x16 8 x LP PCIe 5.0 x16	4 x FHHL PCIe 5.0 x16 8 x LP PCIe 5.0 x16	4 x FHHL PCIe 5.0 x16 (dual-slot) 8 x FHHL PCIe 5.0 x16 (single-slot)	4 x FHHL PCIe 5.0 x16 (dual-slot) 8 x FHHL PCIe 5.0 x16 (single-slot)
I/O Ports	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN	Front: 2 x USB 3.2 Gen1 1 x VGA 2 x RJ45 1 x MLAN (default) Rear: 1 x MLAN
Security	Optional TPM 2.0 kit: CTM010	Optional TPM 2.0 kit: CTM010	Optional TPM 2.0 kit: CTM012	Optional TPM 2.0 kit: CTM012
Power Supply	4+2 redundant PSUs 3000W 80+ Titanium	4+4 redundant PSUs 3000W 80+ Titanium	4+4 redundant PSUs 3000W 80+ Titanium	4+4 redundant PSUs 3000W 80+ Titanium
System Management	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console	ASPEED® AST2600 BMC GIGABYTE Management Console
Other Features	Dual ROM architecture	Dual ROM architecture	Dual ROM architecture	Dual ROM architecture

* SAS card is required to support SAS drives.
** Storage card is required to support SATA and SAS drives.
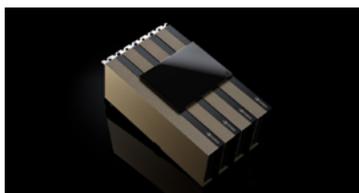
NVIDIA PCIe GPUs

Broad Accelerator Support

GIGABYTE servers have been tested and validated across a range of workloads with various computing accelerators designed to deliver high levels of performance for 3D rendering, deep learning and high-performance computing, which enables content creators, data scientists, researchers, and engineers to tackle challenges that were once impossible.

GPU Models	Qualified Model			
NVIDIA H200 NVL	XV23-VC0-AAJ	G293-S42-AAP1	G293-Z42-AAP1	
	XV23-ZU0-AAJ1	G293-S43-AAP1	G294-A21-AAP2	
	XV23-ZX0-AAJ1	G293-S45-AAP1	G294-A22-AAP2	
	XV24-AX0-AAJ1	G293-S47-AAP1	G294-S41-AAP2	
	XV24-SU0-AAJ1	G293-Z22-AAP1	G294-S42-AAP2	
	XV24-SX0-AAJ1	G293-Z23-AAM1	G294-S45-AAP2	
	G293-S40-AAP1	G293-Z40-AAP1	XL44-SX0-AAS2	
	G293-S41-AAP1	G293-Z41-AAP1	XL43-ZX0-AAS2	
NVIDIA L40S GPU	XV23-ZU0-AAJ1	G294-S41-AAP2	G494-SB4-AAP2	G493-SB3-AAP1
	XV23-ZX0-AAJ1	G294-S42-AAP2	G494-ZB0-AAP2	G493-SB4-AAP1
	XV24-AX0-AAJ1	G294-Z21-AAP2	G494-ZB4-AAP2	G493-ZB0-AAP1
	XV24-SU0-AAJ1	G294-Z22-AAP2	G294-Z22-AAP2	G493-ZB2-AAP1
	XV24-SX0-AAJ1	G294-Z41-AAP2	G294-Z42-AAP2	G493-ZB3-AAP1
	XL44-SX0-AAS2	G294-Z42-AAP2	G493-SB0-AAP1	G493-ZB4-AAP1
	G294-A21-AAP2	G494-SB0-AAP2	G493-SB1-AAP1	G494-SB1-AAP2
	G294-A22-AAP2	G494-SB1-AAP2	G493-SB2-AAP1	G494-SB4-AAP2
NVIDIA RTX PRO™ 6000 Blackwell Server Edition	XV23-ZU0-AAJ1	G494-SB4-AAP2		
	XV23-ZX0-AAJ1	G494-ZB0-AAP2		
	XV24-SU0-AAJ1	G494-ZB4-AAP2		
	XL44-SX2-AAS1			
	G294-A21-AAP2			
	G294-A22-AAP2			
	G294-S42-AAP2			
	G294-Z42-AAP2			
NVIDIA RTX PRO™ 6000 Blackwell Workstation	W773-H5D-AA01			
	W773-W80			
	W774-W90-AAF1			
	W775-V10-L01			
NVIDIA RTX PRO™ 6000 Blackwell Max-Q Workstation	W773-H5D-AA01			
	W773-W80			
	W774-W90-AAF1			
	W775-V10-L01			

NVIDIA RTX PRO™ Server

Model	XLS4-SX2-LAS1	XL44-SX2-AAS1	XL43-ZX2-AAS1
Form Factor	6U (W438 x H262.3 x D802.5 mm)	4U (W438 x H176.6 x D802.5 mm)	4U (W438 x H176.6 x D802.5 mm)
GPU	8 x Liquid-cooled dual-slot PCIe GPUs* - Configured with 8 x NVIDIA RTX PRO™ 6000 Blackwell Server Edition GPUs & 1 x NVIDIA® BlueField®-3 DPU	8 x dual-slot PCIe GPUs* - Configured with 8 x NVIDIA RTX PRO™ 6000 Blackwell Server Edition GPUs & 1 x NVIDIA® BlueField®-3 DPU	8 x dual-slot PCIe GPUs* - Configured with 8 x NVIDIA RTX PRO™ 6000 Blackwell Server Edition GPUs & 1 x NVIDIA® BlueField®-3 DPU
CPU	Intel® Xeon® 6700/6500-Series Dual processor, TDP up to 350W	Intel® Xeon® 6700/6500-series Dual processor, TDP up to 350W	AMD EPYC™ 9005 series Dual processor, cTDP up to 500W
Memory	8-Channel RDIMM/MRDIMM [1] 32 x DIMMs	8-Channel RDIMM/MRDIMM [1] 32 x DIMMs	12-Channel RDIMM 24 x DIMMs
Networking	8 x 400Gb/s QSFP** 2 x 10Gb/s LAN, 1 x MLAN	8 x 400Gb/s QSFP** 2 x 10Gb/s LAN, 1 x MLAN	8 x 400Gb/s QSFP** 2 x 10Gb/s LAN, 1 x MLAN
Storage	8 x E1.S Gen5 NVMe, 2 x M.2 (PCIe 5.0 x4)	8 x 2.5" Gen5 NVMe, 2 x M.2 (PCIe 4.0 x2)	8 x 2.5" Gen5 NVMe, 2 x M.2 (PCIe 3.0 x2/x1)
Expansion Slots	8 x Liquid-cooled PCIe 5.0 x16 FHFL, for GPUs 1 x Liquid-cooled PCIe 5.0 x16 FHHL, for DPU	8 x PCIe 5.0 x16 FHFL for GPUs 1 x PCIe 5.0 x16 FHFL for DPU	8 x PCIe 5.0 x16 FHFL for GPUs 1 x PCIe 5.0 x16 FHFL for DPU
Power Supply	3+1 3200W 80+ Titanium PSUs	3+1 3200W 80+ TitaniumPSUs	4+0 3200W 80+ TitaniumPSUs

Model	XL44-SX0-AAS2	XL43-ZX0-AAS2
Form Factor	4U (W438 x H176.6 x D802.5 mm)	4U (W438 x H176.6 x D802.5 mm)
GPU	8 x dual-slot PCIe GPUs - Supports NVIDIA H200 NVL - Supports 8 x NVIDIA RTX PRO™6000 Blackwell Server Edition	8 x dual-slot PCIe GPUs - Supports NVIDIA H200 NVL - Supports 8 x NVIDIA RTX PRO™ 6000 Blackwell Server Edition
CPU	Intel® Xeon® 6700/6500-series Dual processor,TDP up to 350W	AMD EPYC™ 9005 series Dual processor, cTDP up to 500W
Memory	8-Channel RDIMM/MRDIMM [1] 32 x DIMMs	12-Channel RDIMM 24 x DIMMs
Networking	2 x 10Gb/s LAN 1 x MLAN (front), 1 x MLAN (rear)	2 x 10Gb/s LAN 1 x MLAN (front), 1 x MLAN (rear)
Storage	4 x 2.5" Gen5 NVMe, 2 x M.2 (PCIe 4.0 x2)	4 x 2.5" Gen5 NVMe, 2 x M.2 (PCIe 3.0 x2/x1)
Expansion Slots	8 x PCIe 5.0 x16 FHFL for GPUs 1 x PCIe 5.0 x16 FHFL for DPU	8 x PCIe 5.0 x16 FHFL for GPUs 1 x PCIe 5.0 x16 FHFL for DPU
Power Supply	3+1 3200W 80+ Titanium PSU	4+0 3200W 80+ Titanium PSU

[1] MRDIMMs are only supported with Intel® Xeon® 6 Processors with P-cores and in a 1DPC configuration
* Directly to the GPUs via NVIDIA ConnectX®-8 SuperNIC™ switch board.
** Onboard 400Gb/s InfiniBand/Ethernet QSFP ports with PCIe Gen6 switching for peak GPU-to-GPU performance.

NVIDIA OVX™ Server

Model	**XV24-AX0-AAJ1**	**XV24-SX0-AAJ1**	**XV24-SU0-AAJ1**
Form Factor	2U (W438 x H87 x D900 mm)	2U (W438 x H87 x D900 mm)	2U (W438 x H87 x D900 mm)
GPU	4 x dual-slot PCIe GPUs - Supports NVIDIA H200 NVL - Supports NVIDIA L40S GPU	4 x dual-slot PCIe GPUs - Supports NVIDIA H200 NVL - Supports NVIDIA L40S GPU	4 x dual-slot PCIe GPUs - Supports NVIDIA H200 NVL - Supports NVIDIA L40S GPU
CPU	Intel® Xeon® 6900-series Dual processor, TDP up to 500W	Intel® Xeon® 6700/6500-series Dual processor, TDP up to 350W	Intel® Xeon® 6700/6500-series Singlel processor, TDP up to 350W
Memory	12-Channel RDIMM/MRDIMM [1] 24 x DIMMs	8-Channel RDIMM/MRDIMM [1] 32 x DIMMs	8-Channel RDIMM/MRDIMM [1] 16 x DIMMs
Networking	2 x 10Gb/s LAN, 1 x MLAN	2 x 10Gb/s LAN, 1 x MLAN	2 x 10Gb/s LAN, 1 x MLAN
Storage	6 x 2.5" Gen5 NVMe/SATA/SAS4* 2 x M.2 (PCIe 5.0 x2)	4 x 2.5" Gen5 NVMe/SATA/SAS4* 2 x M.2 (PCIe 5.0 x2)	2 x 2.5" Gen5 NVMe 2 x M.2 (PCIe 5.0 x2)
Expansion Slots	4 x PCIe 5.0 x16 FHFL for GPUs 2 x PCIe 5.0 x16 FHFL for NICs 1 x PCIe 5.0 x16 FHFL for DPU	4 x PCIe 5.0 x16 FHFL for GPUs 2 x PCIe 5.0 x16 FHFL for NICs 1 x PCIe 5.0 x16 FHFL for DPU	4 x PCIe 5.0 x16 FHFL for GPUs 2 x PCIe 5.0 x16 FHFL for NICs 1 x PCIe 5.0 x16 FHFL for DPU
Power Supply	3+1 2000W 80+ Titanium PSUs	3+1 2000W 80+ Titanium PSUs	3+1 2000W 80+ Titanium PSUs

Model	**XV23-ZX0-AAJ1**	**XV23-ZU0-AAJ1**
Form Factor	2U (W438 x H87 x D900 mm)	2U (W438 x H87 x D900 mm)
GPU	4 x dual-slot PCIe GPUs - Supports NVIDIA H200 NVL - Supports NVIDIA L40S GPU	4 x dual-slot PCIe GPUs - Supports NVIDIA H200 NVL - Supports NVIDIA L40S GPU
CPU	AMD EPYC™ 9005/9004 series Dual processor, cTDP up to 500W	AMD EPYC™ 9005/9004 series Singlel processor, cTDP up to 500W
Memory	12-Channel RDIMM 24 x DIMMs	12-Channel RDIMM 12 x DIMMs
Networking	2 x 10Gb/s LAN, 1 x MLAN	2 x 10Gb/s LAN, 1 x MLAN
Storage	6 x 2.5" Gen5 NVMe/SATA/SAS4* 2 x M.2 (PCIe 3.0 x2) 1 x M.2 (PCIe 3.0 x1)	2 x 2.5" Gen5 NVMe 2 x M.2 (PCIe 5.0 x4)
Expansion Slots	4 x PCIe 5.0 x16 FHFL for GPUs 2 x PCIe 5.0 x16 FHFL for NICs 1 x PCIe 5.0 x16 FHFL for DPU	4 x PCIe 5.0 x16 FHFL for GPUs 2 x PCIe 5.0 x16 FHFL for NICs 1 x PCIe 5.0 x16 FHFL for DPU
Power Supply	3+1 2000W 80+ Titanium PSUs	3+1 2000W 80+ Titanium PSUs

[1] MRDIMMs are only supported with Intel® Xeon® 6 Processors with P-cores and in a 1DPC configuration
* Storage card is required to support SATA and SAS drives.

STAY INFORMED STAY AHEAD

Official Website

Subscribe to Newsletter

Designed by

GIGA COMPUTING