GIGABYTE | NVIDIA

DLB2-CB3

NVIDIA GB300 NVL72: Built for the Age of AI Reasoning

The NVIDIA GB300 NVL72 features a fully liquid-cooled, rack-scale design that unifies 72 NVIDIA Blackwell Ultra GPUs and 36 Arm®-based NVIDIA Grace™ CPUs in a single platform optimized for test-time scaling inference.

AI factories powered with the GB300 NVL72 using NVIDIA Quantum-X800 InfiniBand or Spectrum™-X Ethernet paired with ConnectX®-8 SuperNIC™ provide a 50x higher output for reasoning model inference compared to the NVIDIA Hopper™ platform.

Fast Memory	HBM Bandwidth	Networking Bandwidth
60X	**20X**	**18X**
vs. NVIDIA HGX H100	vs. NVIDIA HGX H100	vs. NVIDIA HGX H100

Key Features

- Built on NVIDIA GB300 Grace™ Blackwell Ultra Superchips
- 72 NVIDIA Blackwell Ultra GPUs and 36 NVIDIA Grace™ CPUs
- CPU and GPU connected by fifth-generation NVIDIA NVLink™
- Up to 17 TB LPDDR5X memory with Error-correction code (ECC)
- Up to 20 TB HBM3E GPU memory with 576 TB/s bandwidth
- Up to 37 TB of fast-access memory
- Integrated NVIDIA ConnectX®-8 SuperNIC™

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

3 x 1U 33kW Power Shelves

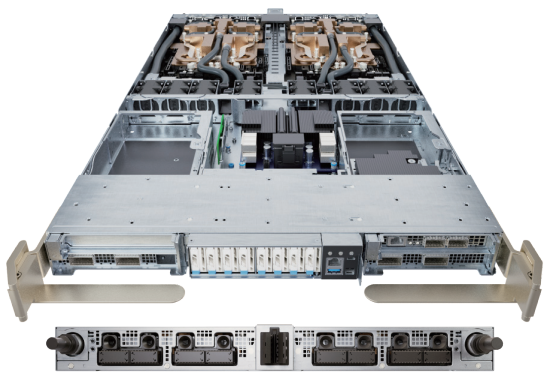
10 x Compute Trays
- 1U XN15-CB0-LA01

9 x NVIDIA NVLink™ Switch Trays
- 1U NVLink™ Switch tray
- 144 x NVLink™ ports per tray
- Fifth-generation NVLink™ with 1.8TB/s GPU-GPU interconnect

8 x Compute Trays
- 1U XN15-CB0-LA01

3 x 1U 33kW Power Shelves

Compatible with in-rack CDU or in-row CDU



XN15-CB0-LA01 Compute Tray

- 2 x NVIDIA GB300 Grace™ Blackwell Ultra Superchip
- 4 x 279GB HBM3E GPU memory
- 2 x 480GB LPDDR5X CPU memory
- 8 x E1.S Gen5 NVMe drive bays
- 4 x NVIDIA ConnectX®-8 800Gb/s OSFP ports
- 1 x NVIDIA® BlueField®-3 DPU

Specifications

System Comparison: NVIDIA GB300 NVL72 vs. GB200 NVL72

	GB300 NVL72	GB200 NVL72
Configuration	36 Grace CPUs, 72 Blackwell Ultra GPUs	36 Grace CPUs, 72 Blackwell GPUs
FP4 Tensor Core	1,440 PFLOPS | 1,100 PFLOPS (dense)	1,440 PFLOPS | 720 PFLOPS (dense)
FP8/FP6 Tensor Core	720 PFLOPS	720 PFLOPS
INT8 Tensor Core	23 PFLOPS	720 POPS
FP16/BF16 Tensor Core	360 PFLOPS	360 PFLOPS
TF32 Tensor Core	180 PFLOPS	180 PFLOPS
FP32	5,760 TFLOPS	5,760 TFLOPS
FP64 / FP64 Tensor Core	100 TFLOPS	2,880 TFLOPS
NVLink Bandwidth	130 TB/s	130 TB/s
Fast Memory	Up to 37 TB	Up to 31 TB
GPU Memory | Bandwidth	Up to 20 TB HBM3E | Up to 576 TB/s	Up to 13.5 TB HBM3E | 576 TB/s
CPU Memory | Bandwidth	Up to 17 TB LPDDR5X | Up to 18.4 TB/s	Up to 17 TB LPDDR5X | Up to 18.4 TB/s
CPU Cores	2,592 Arm Neoverse V2 cores	2,592 Arm Neoverse V2 cores

*All Tensor Core specifications are with sparsity unless otherwise noted.

DLB2-CB3 Rack Specs

Dimensions	1,068L x 600W x 2,299H mm
GPU	72 x NVIDIA Blackwell Ultra GPUs
CPU	36 x NVIDIA Grace™ CPUs
Compute Tray	18 x GIGABYTE XN15-CB0-LA01
NVLink Switch Tray	9 x NVIDIA NVLink™ Switches
Management Switch	2 x OOB management switches + 1 x Optional OS switch for rack-to-rack connection
Power Shelf	6 x 1U 33kW power shelves
Bus Bar	1 x 54V DC 1400A Bus Bar
Cable Cartridge	4 x NVIDIA NVLink™ cable cartridges
CDU	Compatible with in-row or in-rack CDU
Part Numbers	9NDLB2CB3MR000

XN15-CB0-LA01 Specs

Form Factor	1U Liquid-cooled server node
Superchip	2 x NVIDIA GB300 Grace Blackwell Ultra Superchip
Storage	8 x E1.S Gen5 NVMe drive bays 1 x M.2 (2280/22110; PCIe 5.0 x4)
Networking	4 x NVIDIA ConnectX®-8 800Gb/s OSFP ports 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

Learn more at **https://www.GigaComputing.com/en**

* All specifications are subject to change without notice. Please visit our website for the latest information.
* All trademarks and logos are the property of their respective owners.

© 2025 Giga Computing Technology Co., Ltd. All rights reserved.

Designed by GIGA COMPUTING