Lenovo NVIDIA GB300 NVL72

Advancing Enterprise AI: Optimized for Reliable, Rack Scale Model Training Performance

Revolutionary Rack Scale Design for Enterprise AI

The Lenovo NVIDIA GB300 NVL72 is engineered as a turnkey, rack scale AI platform built around the NVIDIA Blackwell architecture. Each rack delivers 72 NVIDIA Blackwell GPUs interconnected through the 5th generation NVIDIA NVLink™ technology, forming a single high performance compute domain ideal for LLM training, multimodal reasoning, and inference workloads. The 1U compute trays, NVLink switch trays, and power infrastructure are tightly integrated to maximize compute density, serviceability, and energy efficiency while reducing datacenter footprint.

Optimized to Handle Trillion Parameter Models

With dual NVIDIA Grace™ CPUs and four Blackwell GPUs per tray, the GB300 NVL72 delivers exceptional performance across diverse AI workloads. The Grace–Blackwell coherent memory architecture provides high bandwidth, low-latency data movement between CPU and GPU, improving model throughput and reducing training times. This unified platform enables enterprises to handle trillion parameter models, real-time inference pipelines, and memory intensive vector operations with consistent and predictable performance.

Scalable NVLink Fabric for Unified 72 GPU Performance

The GB300 NVL72's integrated NVLink switch enables every GPU in the rack to communicate as part of one unified fabric.

This reduces communication overhead for largescale training jobs and eliminates traditional multi-server bottlenecks. The NVLink domain supports GPU-to-GPU 130 TB/s bandwidth - orders of magnitude higher than PCIe-based architectures, enabling seamless scaling for transformers, diffusion models, and complex simulation workloads.

Efficient Hybrid Cooling for Sustainable AI Growth

To support its exceptionally high compute density, the GB300 NVL72 uses a hybrid cooling design that relies primarily on warm water liquid cooling across CPUs, GPUs, NVLink switches, and power subsystems, with supplemental airflow managing the remaining heat. This approach removes the majority of the thermal load without requiring chilled water, reducing datacenter energy usage. Advanced blind mate manifolds allow trays to be serviced easily without draining coolant, while internal leak detection ensures operational resilience in demanding environments.

Power, Reliability, and Infrastructure Built for Enterprise Scale

The rack integrates up to eight 33kW power shelves and a 48V DC busbar for stable, efficient power delivery. A modular backend design enables smooth scaling from a single rack to multi-rack clusters, with support for both high speed InfiniBand and Ethernet fabrics. These capabilities make the GB300 NVL72 suitable for enterprise AI factories, R&D centers, cloud environments, and on-prem HPC deployments requiring predictable uptime and long-term reliability.

Lenovo

NVIDIA

Specifications

Form Factor	• 1U Grace–Blackwell compute tray • 18 trays in a 48U MGX rack • 9 1U NVLink switch trays
Processors	2x NVIDIA Grace CPUs with high bandwidth LPDDR5X memory and cache coherent GPU integration
GPU Accelerators	• 4x NVIDIA Blackwell GPUs per tray • HBM3e memory architecture designed for large model training and inference
CPU Memory	High performance LPDDR5X memory for fast data throughput
GPU Memory	High bandwidth HBM3e memory optimized for large AI models
Storage	Up to 8x E1.S NVMe SSDs per compute tray
Connectivity	Integrated 800 Gb/s CX8 networking with dual 25GbE LOM
Network Interfaces	Rack scale NVLink domain across 72 GPUs via NVLink switch trays
Cooling	• Hybrid cooling system (liquid + air) using PG25 or DI water • Warm water operation (18–45°C)
Power & Power Management	• Up to 8x 33kW power shelves • Balanced 48V busbar distribution
Systems Management	NVIDIA OpenBMC based management
OS Support	• Ubuntu 22.04 (HWE)
Warranty	• 3-year onsite NBD service standard*

* NBD with spare nodes on-site availability

About Lenovo

Lenovo (HKSE: 992) (ADR: LNVGY) is a US$62 billion revenue global technology powerhouse, ranked #171 in the Fortune Global 500, employing 77,000 people around the world, and serving millions of customers every day in 180 markets. Focused on a bold vision to deliver smarter technology for all, Lenovo is expanding into new growth areas of infrastructure, mobile, solutions and services. This transformation is building a more inclusive, trustworthy, and sustainable digital society for everyone, everywhere.

For More Information

To learn more about the Lenovo NVIDIA GB300 NVL72, contact your Lenovo representative or Business Partner or visit www.lenovo.com/thinksystem. For detailed specifications, consult the Product guide.