



Cisco AI Rack for NVIDIA GB300 NVL72

(Supermicro SRS-GB300-NVL72-M0 and SRS-GB300-NVL72-M1)

Contents

Product overview	2
Features and benefits.....	3
Prominent feature.....	4
Platform support	4
Product sustainability	5
Product specifications.....	6
System requirements	7
Ordering information	7
Warranty information.....	8
Cisco and Partner Services	8
Cisco Capital.....	8
Learn more.....	8

A complete, factory-integrated NVIDIA GB300 NVL72 rack with 72 Blackwell Ultra GPUs and 36 Grace CPUs in a single NVLink domain, sold and supported by Cisco.

Product overview

This is rack-scale AI infrastructure ordered as one product. It is Supermicro's NVIDIA GB300 NVL72 system, each rack contains: 18 1U compute trays, 9 NVLink switch trays, 8 power shelves, and in-rack management networking, integrated and tested as a complete rack before delivery. Cisco offers it under two part numbers that differ only in cooling scope: SRS-GB300-NVL72-M0 is the rack without any coolant distribution unit, for customers whose facility already provides one; SRS-GB300-NVL72-M1 includes in-rack CDUs and support in the offer. Cisco handles the selling motion, support through Cisco CX Services and Supermicro for liquid cooling. Management is done through Cisco Nexus One (Hyperfabric).

Each compute tray is packaged with two NVIDIA GB300 Grace Blackwell Ultra superchips, offering 72 Blackwell Ultra GPUs and 36 Grace CPUs connected through fifth-generation NVLink at 1.8 TB/s per GPU in a single rack. All 72 GPUs operate as one accelerator with roughly 21 TB of pooled HBM3e and about 38 TB of total unified memory including LPDDR5X. Target workloads are frontier-scale training and reasoning-model inference, where a single NVLink domain of this size replaces multi-node clusters. Liquid cooling is integrated into the platform; for information on any liquid cooling components, contact your Cisco sales specialist.



Figure 1. Cisco AI Rack for NVIDIA GB300 NVL72 (Supermicro SRS-GB300-NVL72-M1)



Features and benefits

Table 1 summarizes the primary features and benefits of the platform.

Table 1. Features and benefits

Feature	Benefit
72x NVIDIA Blackwell Ultra GPUs in one NVLink domain	The whole rack behaves as a single accelerator for training and long-context inference
36x NVIDIA Grace CPUs (72-core Arm Neoverse V2)	CPU and GPU share coherent memory over NVLink-C2C, removing PCIe bottlenecks
Approx. 38 TB unified memory (approx. 21 TB HBM3e)	Frontier-scale models and KV caches fit inside one rack
9x NVLink switch trays at 1.8 TB/s per GPU	Non-blocking all-to-all GPU communication without external switching
4x NVIDIA ConnectX-8 SuperNICs per tray at 800 Gb/s	Scale-out fabric bandwidth to grow beyond one rack when needed
NVIDIA BlueField-3 DPUs (up to 2 per tray)	Offloads north-south networking, storage, and security from the hosts
Factory-integrated, liquid-cooled rack	Arrives built and tested; direct-to-chip cooling handles the majority of the compute heat load
Two ordering options: without any CDU (SRS-GB300-NVL72-M0) or with in-rack CDU (SRS-GB300-NVL72-M1)	Match the offer to the facility instead of retrofitting the facility to the offer
Cisco CX Services support	Cisco support contract for a rack-scale, along with Supermicro support



Prominent feature

The rack is the computer: 72 GPUs delivered as one integrated system.

Most AI infrastructure is bought as servers and assembled into clusters on site. GB300 NVL72 inverts that: the rack is the unit of compute. Seventy-two Blackwell Ultra GPUs and thirty-six Grace CPUs arrive interconnected through nine NVLink switch trays at 1.8 TB/s per GPU, powered by eight 33 kW power shelves, and plumbed for direct-to-chip liquid cooling. What would otherwise be months of integration work is done before the truck arrives.

Cisco removes the remaining friction at the facility boundary. SRS-GB300-NVL72-M0 is for data centers that already operate a compatible coolant distribution unit and just need the rack. SRS-GB300-NVL72-M1 integrates an in-rack CDU and support into the same order. In both cases the liquid cooling scope is specified with your Cisco sales specialist, who can advise on in-rack, in-row, and liquid-to-air sidecar options and provide quotes for liquid cooling products. Cisco and Supermicro support contracts then covers hardware, and support.

Platform support

Cisco offers this Supermicro platform in a defined rack configuration under two part numbers, listed in Table 2.

Table 2. Cisco selected configuration, NVIDIA GB300 NVL72 rack

Item	Selected configuration
PID, without CDU	SRS-GB300-NVL72-M0 (Supermicro-built GB300 NVL72 rack; facility provides the coolant distribution unit)
PID, with CDU	SRS-GB300-NVL72-M1 (Supermicro-built GB300 NVL72 rack with CDU support included in the offer)
Rack base	1x NVIDIA GB300 NVL72 rack: 18x 1U compute trays, 9x NVLink switch trays, 8x 1U 33 kW power shelves
GPU/CPU	72x NVIDIA Blackwell Ultra GPUs and 36x NVIDIA Grace CPUs (4x B300 and 2x Grace per tray)
Memory	Approx. 38 TB unified memory: approx. 21 TB HBM3e plus LPDDR5X
M.2 boot drives	1x 1.9 TB NVMe M.2 per tray
NVMe storage	Qty (4) 3.84TB SED Drives, Up to 8x E1.S NVMe per tray
East-west networking	4x NVIDIA ConnectX-8 SuperNICs at 800 Gb/s per tray



Item	Selected configuration
North-south networking	1x NVIDIA BlueField-3 DPUs per tray
Management switching	(optional, purchased separately) 2x Cisco Nexus N93108TC-FX3 switches in rack
Liquid cooling options	In-rack CDU up to 250 kW, in-row CDU up to 1.8 MW, or liquid-to-air sidecar up to 200 kW; contact your Cisco sales specialist for liquid cooling component information and quotes
Support	Cisco CX Services and Supermicro Services
Availability	October 2026

Product sustainability

Information about Cisco’s Environmental, Social and Governance (ESG) initiatives and performance is provided in Cisco’s CSR and sustainability [reporting](#).

Table 3. Cisco Environmental Sustainability Information

Sustainability Topic		Reference
General	Information on product-material-content laws and regulations	Materials
	Information on electronic waste laws and regulations, including our products, batteries and packaging	WEEE Compliance
	Information on product takeback and resuse program	Cisco Takeback and Reuse Program
	Sustainability Inquiries	Contact: csr_inquiries@cisco.com
Material	Product packaging weight and materials	Contact: environment@cisco.com



Product specifications

Table 4 reflects the Supermicro data sheet for the NVIDIA GB300 NVL72 rack. Product specifications may change without notice.

Table 4. Specifications, Supermicro NVIDIA GB300 NVL72 rack

Category	Specification
Form factor	48U rack-scale system; approx. 2236 x 600 x 1068 mm (88 x 23.6 x 42 in)
Compute trays	18x 1U trays; each with 2x NVIDIA GB300 Grace Blackwell Ultra superchips (4x B300 GPUs, 2x Grace CPUs)
GPU	72x NVIDIA Blackwell Ultra (B300) GPUs total
CPU	36x NVIDIA Grace CPUs; 72-core Arm Neoverse V2 each
Memory	Per tray approx. 1.15 TB HBM3e plus 960 GB LPDDR5X; rack total approx. 21 TB HBM3e, approx. 38 TB unified
GPU interconnect	NVIDIA NVLink 5 via 9x NVLink switch trays; 1.8 TB/s per GPU, all 72 GPUs in one domain
Scale-out networking	4x NVIDIA ConnectX-8 SuperNICs at 800 Gb/s per tray
In-band/out-of-band	Ethernet leaf switches in rack for in-band management; 1G/10G IPMI switch for out-of-band
North-south/storage	1x NVIDIA BlueField-3 DPUs per tray
Storage	2x 1.9 TB M.2 NVMe boot per tray; up to 8x E1.S PCIe 5.0 NVMe per tray
Cooling	Direct-to-chip liquid cooling across compute and switch trays; CDU options: in-rack up to 250 kW in-row up to 1.8 MW (multiple racks), or liquid-to-air sidecar up to 200 kW with no facility water required
Power	8x 1U 33 kW power shelves (6x 5.5 kW PSUs each); operating power approx. 132 to 140 kW



System requirements

Table 5 lists what must be in place before installation. This is a rack-scale liquid-cooled system; for information on any liquid cooling components, or for a quote on liquid cooling products, contact your Cisco sales specialist.

Table 5. System requirements

Requirement	Description
Facility space	Full-rack footprint approx. 2236 x 600 x 1068 mm; verify delivery path, floor loading, and service clearance for an integrated rack
Power	Facility feeds sized for approx. 132 to 140 kW operating power per rack, delivered through 8x 33 kW power shelves
Liquid cooling	SRS-GB300-NVL72-M0: facility-provided coolant distribution unit and water loop. SRS-GB300-NVL72-M1: in-rack CDU and support included; options scoped with your Cisco sales specialist
Scale-out fabric	Cisco N9100 Switches with NVIDIA Spectrum-X Ethernet infrastructure for multi-rack deployments

Ordering information

The rack is a Supermicro-built system offered through Cisco under the two part numbers below. Certain liquid cooling components, such as in-row CDUs and liquid-to-air sidecards, are quoted separately: contact your Cisco sales specialist for information on any liquid cooling components or for a quote on liquid cooling products specifically. To order, contact your Cisco account representative or visit the Cisco Ordering Home Page.

Table 6. Ordering information

Part number	Product description
SRS-GB300-NVL72-M0	Supermicro 48U liquid-cooled NVIDIA GB300 NVL72 rack, no In-Rack CDU
SRS-GB300-NVL72-M1	Supermicro 48U liquid-cooled NVIDIA GB300 NVL72 rack with In-Rack CDU
LCS-SCDU-1K3LR001	1.8MW In-Row CDU, L2L, N+1 Pumps
LCM-SMLCPM3-3Y	Liquid Cooling Maintenance Service, SMLCPM3, 3-Year
LCS-SCDU-200AR001	200kW L2A Sidecar CDU, N+1 Pumps

Part number	Product description
SFT-ODLCRSUPSTD-5Y	BMC Software Support, Standard, 5Yr
SVC-NVSTDSUP-3Y	NVIDIA Enterprise Support Entitlement, 3Yr

Warranty information

Warranty terms for this Cisco-offered Supermicro platform are standard Supermicro warranty terms: <https://www.supermicro.com/en/support/warranty>. Support is delivered through Cisco CX Services under the applicable service contract. Please contact your Cisco sales specialist for more details.

Cisco and Partner Services

Cisco CX and Supermicro Services cover this platform from planning through deployment and ongoing operations, with dedicated support contracts. For more information, visit <https://www.cisco.com/go/services>.

Cisco Capital

Flexible payment solutions to help you achieve your objectives

Cisco Capital makes it easier to get the right technology to achieve your objectives, enable business transformation and help you stay competitive. We can help you reduce the total cost of ownership, conserve capital, and accelerate growth. In more than 100 countries, our flexible payment solutions can help you acquire hardware, software, services and complementary third-party equipment in easy, predictable payments. [Learn more](#).

Learn more

For more on Cisco AI infrastructure, visit <https://www.cisco.com/go/ai>. Full platform specifications are available on the Supermicro product page for the NVIDIA GB300 NVL72 system.

Learn more about Cisco and Supermicro partnership:

<https://www.cisco.com/site/us/en/solutions/global-partners/supermicro/index.html>.