NVIDIA DGX GB300 NVL72

72x B300 | Blackwell Ultra | rack | 2025

OVERVIEW

360	20.7 TB	576.0 TB/s	—
FP8 PFLOPS Dense	Aggregate Memory	Aggregate Bandwidth	Aggregate Power

PERFORMANCE METRICS

Precision	Peak (PFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP32	6.0	83.3	–
TF32	90.0	1250.0	–
FP16	180.0	2500.0	–
BF16	180.0	2500.0	–
FP8	360.0	5000.0	–
FP4	1080.0	15000.0	–

All figures are dense. Vendors commonly headline the sparse number, which is twice the dense one.

CONFIGURATION

Accelerator	B300 SXM 288GB
Count	72
Memory per GPU	288 GB
Bandwidth per GPU	8.0 TB/s

POWER & EFFICIENCY

Aggregate Power	—
TDP per GPU	1400 W
FP8 Efficiency	—
Aggregate Bandwidth	576.0 TB/s

SYSTEM DETAILS

Vendor: NVIDIA
Family: NVL
Architecture: Blackwell Ultra
Form Factor: rack
Release Year: 2025
Process Node: 4nm

INTERCONNECT

NVLink + NVSwitch

NOTES

36 GB300 superchips (72 B300 GPUs + 36 Grace CPUs) in liquid-cooled rack