

NVIDIA HGX B300

8x B300

Blackwell Ultra

baseboard

2025

OVERVIEW

36.0 FP8 PFLOPS Dense	2.1 TB Aggregate Memory	64.0 TB/s Aggregate Bandwidth	— Aggregate Power
---	---	---	---

PERFORMANCE METRICS

Precision	Peak (PFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP64	0.0	1.3	—
FP32	0.6	75.0	—
TF32	9.0	1125.0	—
FP16	18.0	2250.0	—
BF16	18.0	2250.0	—
FP8	36.0	4500.0	—
FP4	108.0	13500.0	—
FP6	36.0	4500.0	—

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

CONFIGURATION

Accelerator	B300 SXM 288GB
Count	8
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	64.0 TB/s

SYSTEM DETAILS

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

INTERCONNECT

NVLink 5.0 / NVSwitch 4.0

NOTES

The eight-GPU Blackwell Ultra baseboard OEMs build around. NVIDIA publishes both a dense and a sparse figure for every tensor precision here, and for FP4 the two are not a factor of two apart: dense rose by half over the B200, from 72 to 108 PFLOPS, while sparse stayed at 144. Total memory

is NVIDIA's own 2.1 TB rather than the 2.304 TB that eight 288 GB GPUs would suggest. Memory bandwidth of 64 TB/s is the one derived number, being eight times the B300's published 8 TB/s, because NVIDIA gives no board-level HBM figure. No INT8 row is recorded: NVIDIA's table gives 3 POPS sparse for this board while our B300 chip entry implies 36 POPS dense, the two cannot both be right, and the question is open rather than guessed at. NVIDIA publishes no baseboard power figure, so total power is blank.