

NVIDIA DGX B300

8x B300

Blackwell Ultra

rack

2025

OVERVIEW

36.0 FP8 PFLOPS Dense	2.1 TB Aggregate Memory	64.0 TB/s Aggregate Bandwidth	14.0 kW 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	14.0 kW
TDP per GPU	1400 W
FP8 Efficiency	2.57 TFLOPS/W
Aggregate Bandwidth	64.0 TB/s

SYSTEM DETAILS

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

INTERCONNECT

NVSwitch + NVLink

NOTES

Eight Blackwell Ultra SXM GPUs on the HGX B300 baseboard with Intel Xeon 6776P processors, in a 10U chassis. NVIDIA quotes 144 PFLOPS of FP4 inference performance for this machine, which is the

sparse figure; the dense equivalent is 108 PFLOPS, and both are recorded. Power is given by NVIDIA as approximately 14 kW.