

NVIDIA DGX SuperPOD GB200

2304x B200

pod

2025

OVERVIEW

11,520 FP8 PFLOPS Dense	— Aggregate Memory	— Aggregate Bandwidth	3.8 MW Aggregate Power
-----------------------------------	-----------------------	--------------------------	----------------------------------

PERFORMANCE METRICS

Precision	Peak (PFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP64	85.2	37.0	27.00
FP32	172.8	75.0	54.00
TF32	2534.4	1100.0	1500.00
FP16	5760.0	2500.0	2700.00
BF16	5760.0	2500.0	2700.00
FP8	11520.0	5000.0	5400.00
FP4	23040.0	10000.0	10800.00
INT8	10368.0	4500.0	5400.00

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

CONFIGURATION

Accelerator	B200 SXM 180GB
Count	2304
Memory per GPU	180 GB
Bandwidth per GPU	7.7 TB/s

POWER & EFFICIENCY

Aggregate Power	3.8 MW
TDP per GPU	1000 W
FP8 Efficiency	5400.00 TFLOPS/W
Aggregate Bandwidth	—

SYSTEM DETAILS

Vendor: NVIDIA	Family: SuperPOD	Architecture: —
Form Factor: pod	Release Year: 2025	Process Node: 4nm

INTERCONNECT

NVLink, NVSwitch, InfiniBand

NOTES

32-rack SuperPOD configuration with GB200 NVL72 systems

