



NVIDIA HGX H200

8x H200

Hopper

baseboard

2024

OVERVIEW

15.8 FP8 PFLOPS Dense	1.1 TB Aggregate Memory	38.4 TB/s Aggregate Bandwidth	8.0 kW Aggregate Power
---------------------------------	-----------------------------------	---	----------------------------------

PERFORMANCE METRICS

Precision	Peak (PFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP64	0.3	34.0	68.00
FP32	0.5	67.0	134.00
TF32	4.0	494.5	989.00
FP16	7.9	989.5	1979.00
BF16	7.9	989.5	1979.00
FP8	15.8	1979.0	3958.00
INT8	15.8	1979.0	3958.00

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

CONFIGURATION

Accelerator	H200 SXM 141GB
Count	8
Memory per GPU	141 GB
Bandwidth per GPU	4.8 TB/s

POWER & EFFICIENCY

Aggregate Power	8.0 kW
TDP per GPU	700 W
FP8 Efficiency	3958.00 TFLOPS/W
Aggregate Bandwidth	38.4 TB/s

SYSTEM DETAILS

Vendor: NVIDIA	Family: HGX	Architecture: Hopper
Form Factor: baseboard	Release Year: 2024	Process Node: 4nm

INTERCONNECT

NVSwitch + NVLink

NOTES

HGX platform for cloud providers with H200 GPUs