

HPE ProLiant Compute XD685 (HGX H200)

8x H200

Hopper

rack

2025

OVERVIEW

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

PERFORMANCE METRICS

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

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	—
TDP per GPU	700 W
FP8 Efficiency	—
Aggregate Bandwidth	38.4 TB/s

SYSTEM DETAILS

Vendor: HPE	Family: ProLiant Compute XD	Architecture: Hopper
Form Factor: rack	Release Year: 2025	Process Node: 4nm

INTERCONNECT

NVLink 4.0 / NVSwitch 3.0 across the HGX H200 baseboard; 4x FHHL PCIe Gen5 x16 slots for fabric, NIC or DPU

NOTES

The eight-way XD685 in its NVIDIA HGX H200 configuration, the only one of the four HPE offers in both air-cooled and liquid-cooled form: 6U for air with accelerator kit S3P89A, 5U for direct liquid cooling with S4X53A. Two fifth-generation AMD EPYC processors drive it, from 32 to 96 cores and 280 to 400 W, across 24 DIMM slots with twelve channels per socket and up to 6 TB of DDR5-6400.

Storage is eight EDSFF NVMe E3.S drives, or twelve with a storage controller, plus RAID-1 M.2 boot. No chassis part number is recorded because two apply depending on cooling, S4Q28A for air and S4R96A for liquid. HPE publishes no system power draw for the XD685, only twelve 3,000 W Titanium supplies with ten deployable, so nothing is summed here; the compute figures are NVIDIA's for the HGX H200 baseboard rather than an HPE claim.