HPE Compute XD690

8x B300 | Blackwell Ultra | rack | 2025

OVERVIEW

36.0	2.1 TB	64.0 TB/s	14.3 kW
FP8 PFLOPS Dense	Aggregate Memory	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	14.3 kW
TDP per GPU	1400 W
FP8 Efficiency	2.52 TFLOPS/W
Aggregate Bandwidth	64.0 TB/s

SYSTEM DETAILS

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

INTERCONNECT

NVLink 5.0 / NVSwitch 4.0 across the HGX B300 baseboard; 8x NVIDIA ConnectX-8 for InfiniBand XDR or Spectrum-X Ethernet scale-out

NOTES

A 10U air-cooled single-node chassis carrying eight NVIDIA Blackwell Ultra SXM6 GPUs on the HGX B300 baseboard, which HPE names directly in the orderable part S6A37A. Two Intel Xeon 6

processors with P-cores drive it, either the 6776P or the 6767P, both 64-core at 350 W, with up to 32 DDR5 DIMMs and a minimum system memory of 2,048 GB running at 5,200 MT/s in the two-DIMM-per-channel configuration HPE recommends for performance. Scale-out is eight ConnectX-8 adapters supporting InfiniBand XDR or Spectrum-X Ethernet, with storage of eight E1.S drives and two M.2. Power comes from twelve fixed 3,200 W Titanium supplies in an N+N redundant arrangement, and HPE rates the machine at 14.3 kW continuous and 16.4 kW peak. Air cooling a Blackwell Ultra system of this density is unusual and it shows in the mechanicals: the chassis weighs 226 kg, ships pre-racked by default, and HPE requires a server lift on site before its engineers will service it. HPE describes it as delivering 50 percent higher FP4 performance than the prior Blackwell generation but publishes no FLOPS figure of its own, so the compute shown is NVIDIA's specification for the HGX B300 board rather than an HPE claim.
