



QCT QuantaGrid D75F-7U

8x H200 Hopper rack

OVERVIEW

—	1.1 TB	—	—
Peak Dense	Aggregate Memory	Aggregate Bandwidth	Aggregate Power

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
Efficiency	—
Aggregate Bandwidth	—

SYSTEM DETAILS

Vendor: QCT	Family: QuantaGrid	Architecture: Hopper
Form Factor: rack	Release Year: —	Process Node: 4nm

INTERCONNECT

One NVIDIA HGX H200 baseboard connecting the eight GPUs at 900 GB/s over NVLink, which QCT notes is seven times the bandwidth of PCIe Gen 5; two BlueField-3 DPUs handle networking

NOTES

QCT's 7U Hopper machine, with two fifth or fourth generation Intel Xeon Scalable processors and eight NVIDIA HGX H200 GPUs linked at 900 GB/s over NVLink. Storage is eighteen small-form-factor all-NVMe bays for GPU-Direct Storage plus a boot drive, and two NVIDIA BlueField-3 DPUs offload networking. QCT gives 1.1 TB of aggregate high-bandwidth memory. It also quotes over 32 petaflops of FP8, which is not shown here: QCT does not say whether that is dense or sparse, it is close to double the dense figure NVIDIA published for this baseboard, and NVIDIA no longer lists the H200 on its HGX page, so there is no way to settle which column it belongs in without guessing. Power is four-plus-two 4,000 W Titanium supplies, which is capacity rather than draw.