



QCT QuantaGrid S74G-2U

1x GH200 Grace Hopper rack

OVERVIEW

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

CONFIGURATION

Accelerator	GH200 96GB HBM3
Count	1
Memory per GPU	96 GB
Bandwidth per GPU	4.0 TB/s

POWER & EFFICIENCY

Aggregate Power	—
TDP per GPU	1000 W
Efficiency	—
Aggregate Bandwidth	—

SYSTEM DETAILS

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

INTERCONNECT

NVLink-C2C joins the Grace CPU and Hopper GPU into one coherent memory pool; three PCIe 5.0 x16 full-height full-length dual-width slots for expansion

NOTES

QCT's Grace Hopper machine, and its entry in the first generation of NVIDIA's MGX modular architecture, which QCT says it partnered with NVIDIA to introduce. A single GH200 Superchip pairs 72 Arm Neoverse cores with a Hopper GPU over NVLink-C2C, giving one coherent memory pool across CPU and GPU rather than a host and a device that copy between them. QCT lists up to 480 GB of embedded LPDDR5X alongside 96 GB of HBM3 GPU memory, four E1.S NVMe drives by default, and three full-height full-length dual-width PCIe 5.0 x16 slots in a 2U chassis 900 mm deep. The memory figure here counts the HBM only, since the LPDDR5X is the Grace side of the pair. Power is one-plus-one 2,000 W Titanium supplies, which is capacity rather than draw, and QCT publishes no throughput figure.