# QCT Cutting-Edge Infrastructures and Solutions Powered by NVIDIA

Accelerated computing performance
for AI and HPC applications

Order Now

## NVIDIA Grace Hopper Superchip

### Grace-Hopper Superchip MGX Server

**QuantaGrid S74G-2U** is the NVIDIA Grace-Hopper Superchip server which combines ARM-based Grace CPU (72 cores) and Hopper GPU (H100) with NVLink chip-to-chip interconnection and supports 96GB of HBM3 and 480GB of LPDDR5 memory.

Powered by the revolutionary architecture of the NVIDIA Grace-Hopper, QCT QuantaGrid S74G-2U empowers the breakthrough computing performance to the AI and HPC applications.

The NVIDIA Grace Hopper™ Superchip combines the Grace™ CPU and Hopper™ GPU with NVLink®-C2C interconnection empowering breakthrough computing performance to various AI and HPC applications.

## NVIDIA MGX

### Bring Accelerated Computing to AI and HPC users with modular design

GPU CPU DPU

A single MGX architecture enables

**QuantaGrid S74G-2U**

AI | HPC+Data Analytics | Digital Twins

Cloud Services | Cloud Gaming | 5G

## Accelerating AI & HPC Workloads

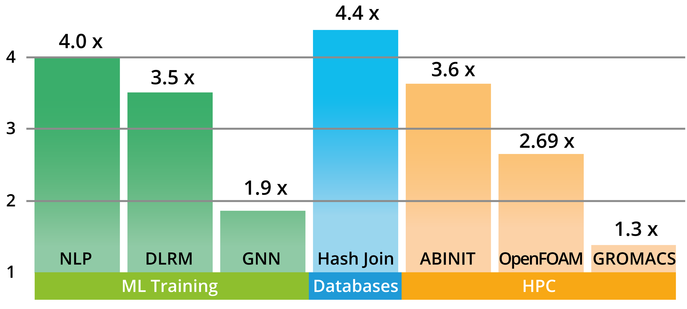
### Accelerated Computing Workloads

- Machine Learning and Inference workloads. e.g., NLP, DLRM
- Database workloads. e.g., Hash Join
- HPC workloads. e.g., OpenFOAM, GROMACS

### True Heterogenous AI/HPC System

- Future compatible with next-gen NVIDIA CPUs, GPUs and DPUs
- Easy-to-deploy with full support for an NVIDIA software stack

### Grace Hopper vs. Hopper Performance Results



# QCT Rack-scale Architecture Designed to Handle Terabyte-Class AI Models

First Architecture with Rack-level NVLink

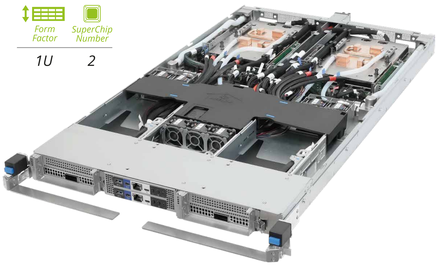
## Grace™ Hopper™ Superchip Rack-scale Architecture

This new class of rack-scale architecture is capable of interconnecting 32 NVIDIA Grace Hopper Superchips via NVIDIA NVLink, offering linear scalability, massive shared-memory space across GPUs and exceptional power efficiency with liquid-cooled NVIDIA Grace Hopper Superchips.

Compared to traditional architectures, this rack-scale architecture eliminates complexity, speeds up deployments and simplifies operations, unlocking enormous potential in the era of generative AI.

### QunataGrid D74S-1U (MGX™ Architecture)

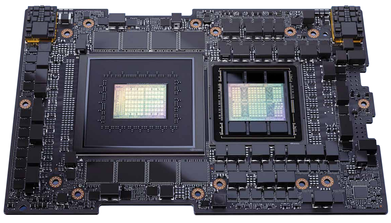
The **QuantaGrid D74S-1U** is an innovative QCT server abiding by the NVIDIA® MGX™ architecture. Populating QCT's Grace™ Hopper™ Superchip Rack-scale Architecture, the QuantaGrid D74S-1U is equipped with QCT's advanced liquid cooling technology to support 2 NVIDIA® GH200 Grace Hopper™ Superchips with 1000W TDP in a 1U form factor.

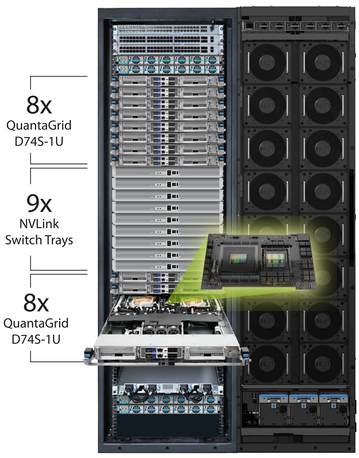



| | |
|---|---|
| **Processor** | (2) NVIDIA® GH200 144GB Grace Hopper™ Superchips , 1000W TDP |
| **Memory** | CPU: Up to 480GB LPDDRX embedded per Superchip<br>GPU: 144GB HBM3e embedded per Superchip |
| **Networking** | Dedicated (1) GbE management port |
| **Storage** | (4) E1.S 15mm PCIe SSD (CPU direct attached)<br>(2) PCIe M.2 22110/2280 |
| **Expansion Slot** | (2) PCIe 5.0 x16 FHFL dual width slots<br>(2) PCIe 5.0 x16 HHHL slots |
| **Cooling** | CPU/GPU: Liquid cooling cold plate<br>Peripheral: (8) 4056 dual rotor fans (N+2 redundant) |
| **Power** | 48-54V DC bus bar clip |
| **Dimension** | (W)438 x (H)43.6 x (D)766mm |

## NVIDIA Grace Hopper Superchip

The QuantaGrid D74S-1U is an innovative QCT server abiding by the NVIDIA® MGX™ architecture. Populating QCT's Grace™ Hopper™ Superchip Rack-scale Architecture, the QuantaGrid D74S-1U is equipped with QCT's advanced liquid cooling technology to support 2 NVIDIA® GH200 Grace Hopper™ Superchips with 1000W TDP in a 1U form factor.






**7X** **Faster** than PCIe Gen5 with the NVIDIA NVLink C2C CPU-GPU coherent interface

**90X** **More area-efficient** utilizing advanced packaging

**624 GB** Shared memory for fast-access memory, utilizing HBM3 and HBM3e memory

# QCT AI x 5G Infrastructures Powered by NVIDIA

## NVIDIA Grace Hopper Superchip

Grace-Hopper Superchip MGX Server

**QuantaEdge EGX77GE-2U** is an NVIDIA Grace-Hopper Superchip server which combines ARM-based Grace CPU (72 cores) and Hopper GPU (H100) with NVLink chip-to-chip interconnection.

Powered by the revolutionary architecture of the NVIDIA Grace Hopper, QuantaEdge EGX77GE-2U empowers the breakthrough computing performance of AI and HPC applications.

## NVIDIA MGX

Bring Accelerated Computing from AI to 5G vRAN with a Modular Design

**QuantaEdge EGX77GE-2U**

## NVIDIA Grace Hopper Superchip

- **Arm Neoverse V2 CPU + H100 GPU**
- Featuring **NVLink-C2C** (High-bandwidth and low-latency interconnect)
- **Coherent memory**