

QCT Showcases Server Platforms Accelerated by NVIDIA AI at SuperComputing 2024

*QCT Servers for Generative AI Use Cases
Built on NVIDIA AI Full-Stack Platform for
Every Data Center*

ATLANTA, GA, UNITED STATES,
November 21, 2024 /

EINPresswire.com/ -- Quanta Cloud
Technology (QCT), a leading provider of
data center solutions, is attending
SuperComputing 2024 (SC24) at the
Georgia World Congress Center in
Atlanta, Georgia, from November 17-

22. Under the slogan "Transforming Tomorrow with HPC & AI," QCT is showcasing a comprehensive portfolio of accelerated computing solutions built on the NVIDIA Blackwell platform, including those adopting NVIDIA Hopper™ GPUs, NVIDIA NVLink™-C2C interconnect technology, the NVIDIA Spectrum-X™ Ethernet platform, NVIDIA BlueField®-3 networking, and the [NVIDIA AI Enterprise](#) software platform, which includes NVIDIA NIM microservices, at SC24. The QCT servers are built to enable generative AI use cases for every data center with new levels of performance, energy efficiency, and scalability. QCT's product showcase at SC24 addresses different AI and HPC needs with solutions based on different NVIDIA technologies.

"Generative AI has become a key driver of digital transformation in virtually every industry, creating an almost insatiable demand for AI infrastructures," said Mike Yang, President of QCT. "QCT as an experienced hyperscaler infrastructure provider is ready to help every data center and enterprise expedite generative AI deployments with performance-optimized, energy-efficient accelerated computing platforms powered by NVIDIA technologies."

"Using NVIDIA's full-stack accelerated computing platform, QCT is helping meet the growing demand for advanced AI infrastructure in data centers worldwide, empowering enterprises to maximize performance, optimize costs and accelerate time to value for AI-driven innovations," said Kaustubh Sanghani, vice president of GPU product management at NVIDIA.

QCT accelerated computing systems showcased at Booth #1013 include:



□ NVIDIA MGX Architecture-based QuantaGrid Servers

Based on NVIDIA MGX modular reference designs, the QuantaGrid D75B-1U and QuantaGrid D75E-4U deliver flexibility, faster time to market and multi-gen compatibility. Accelerated by the latest NVIDIA Blackwell GPUs and NVIDIA PCIe GPUs such as the [NVIDIA H200 NVL](#), these systems are easy to deploy and scale out, supporting mainstream AI inference and LLM workloads.

□ NVIDIA HGX with Blackwell or Hopper 8-GPU Servers

The QuantaGrid D75F-9U, QuantaGrid D75H-7U, and QuantaGrid D75L-5U are 8-way NVIDIA HGX accelerator servers meticulously engineered to confront the most intricate AI and HPC tasks, including large language models, recommender systems, and chatbots. They are enhanced with 8 NVIDIA Blackwell or NVIDIA Hopper GPUs interconnected via NVIDIA NVLink, and feature flexible I/O slots to support NVIDIA BlueField-3 SuperNICs and DPUs.

□ Rack-level NVIDIA GB200 NVL72 Solution

An AI factory built for real-time trillion-parameter inference and training. Comprising 18x QCT QuantaGrid D75B-1U and 9x fifth-generation NVIDIA NVLink switch trays, this liquid-cooled rack-level solution delivers powerful performance, impressive energy efficiency and exceptionally low TCO by interconnecting 72 Blackwell GPUs as one.

In addition, QCT offers a broad portfolio of NVIDIA-accelerated hardware systems supporting NVIDIA GB200 NVL72, NVIDIA H200 NVL with HBM3e memory, and NVIDIA HGX H200 4- and 8-GPUs. The latest NVIDIA H200 NVL unleashes AI acceleration for mainstream enterprise servers with up to 1.8X faster large language model (LLM) inference and 1.3X more performance on HPC applications over the NVIDIA H100 NVL GPU. QCT is also exhibiting its integrated solutions and AI/HPC use cases at SC24, including QCT Platform on Demand (QCT POD), an HPC and AI converged system that supports both cloud-native and bare metal environments, and is integrated with the NVIDIA NGC catalog to streamline AI and HPC workflows.

Visit QCT Booth #1013 during SC24 and talk to QCT experts to learn more about QCT's innovative products and solutions adopting NVIDIA technology.

About QCT

Quanta Cloud Technology (QCT) designs, manufactures, integrates, and services cutting-edge offerings for 5G Telco/Edge, AI/HPC, Cloud, and Enterprise infrastructure via its global network. Product lines include hyper-converged and software-defined data center solutions as well as servers, storage, and network switches from 1U to entire racks with a diverse ecosystem of hardware components and software partners to fit a variety of business verticals and workload parameters. www.qct.io

All other brands, names, and trademarks are the property of their respective owners.

Jean Ko

QCT

+886 912 025 348

jean_ko@quantatw.com

Visit us on social media:

[Facebook](#)

[X](#)

[LinkedIn](#)

[YouTube](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/762491863>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2024 Newsmatics Inc. All Right Reserved.