

QCT Enables Heterogeneous Inference on Intel Xeon 6-Powered Servers at Computex 2026

Demonstrating End-to-End Intel-Powered Infrastructures from Core Data Centers to AI and the Edge

TAIPEI, TAIWAN, June 2, 2026

[/EINPresswire.com/](#) -- [Quanta Cloud Technology](#) (QCT), a global data center and 5G solution provider, is attending Computex 2026, where the company is showcasing a comprehensive portfolio of servers and solutions, powered by Intel Xeon 6, designed to help enterprises and service providers address evolving workloads across AI, cloud, edge, and enterprise data centers.



“As more and more companies deploy AI, inference is being driven by the processor of those

“

“By combining QCT’s system design expertise with Intel’s technologies and our GPU-accelerated systems, we are enabling customers to build resilient, GPU-accelerated heterogeneous systems.”

Mike Yang, President of QCT

agentic AI and reasoning systems,” said Mike Yang, President of QCT. “By combining QCT’s system design expertise with Intel’s technologies and our GPU-accelerated systems, we are enabling customers to build resilient, GPU-accelerated heterogeneous systems that provide optimized performance while reducing total cost of ownership.”

At this year’s COMPUTEX, QCT is highlighting its QCT Heterogeneous AI Solution, a rack-scale heterogeneous system architecture built on Intel Xeon 6 processors,

integrating multiple accelerated systems and AI infrastructure software stacks to support diverse AI scenarios. As AI workloads shift from large scale training toward distributed inference, AI performance is increasingly defined not only by GPU throughput but also by CPU-led system performance.□

The new QCT Heterogeneous System integrates Intel Xeon 6-powered QuantaGrid D75E-4U

server supporting Intel Arc Pro B70—with an upgrade path to Intel’s next generation data center GPU, codenamed Crescent Island, alongside QuantaGrid D55X-1U, QuantaGrid D55Q-2U, QuantaGrid S55J-2U QuantaPlex S45Z-2U servers, and the QuantaMesh TA064-IXM Ethernet switch to form a heterogeneous cluster unifying compute, management, storage and networking resources.□

Leveraging the Intel Xeon 6 processor with fast memory support, industry-leading PCIe connectivity, and advanced power efficiency, QCT enables seamless integration of GPUs and accelerators for distributed and disaggregated inference, while delivering strong CPU performance for scheduling, data movement, and workload orchestration. By leveraging open-source distributed inference frameworks such as [NVIDIA Dynamo](#) and [llm-d](#), QCT and Intel maximize throughput and resource efficiency to power scalable agentic AI workloads, while supporting techniques like KV cache reuse to accelerate time to first token and reduce total costs.

To showcase the practical advantages of this architecture, QCT will demonstrate a multi-agent AI scenario at its COMPUTEX booth. In the demonstration, AI agents collaborate to process user requests, while distributed inference frameworks assign workloads to the most efficient compute resources based on model size and task requirements.□

The QCT Heterogeneous AI solution supports demanding AI inference, HPC, and data analytics workloads, delivering balanced performance and efficiency for a distributed computing environment. This architecture ensures maximum resource utilization while driving down infrastructure Capex.

At Computex 2026, QCT is also showcasing the following Intel-powered edge solutions:

- QCT OmniPOD Private 5G Solution: An end-to-end, Intel Xeon powered solution that integrates a standalone 5G core, 5G RAN, and network management software. Fully compliant with 3GPP Release 16, it delivers cloud native resilience, network slicing, multi cell RAN, and URLLC capabilities, alongside enhanced security, energy efficiency, and manageability. Together, these capabilities enable a wide range of on premises applications that help customers maximize their return on investment.
- QCT Public Network Solutions: From cloud to edge, spanning 5G core and ORAN-compliant virtualized RAN (vRAN) infrastructures, QCT delivers full-stack, cloud-native solutions strengthened by its in-house cloud and data center management tools to enhance and simplify modern cloud environments. Optimized with Intel FlexRAN reference architecture and Intel Xeon 6 SoC with Intel vRAN Boost and built-in AI via Intel Advanced Matrix Extensions and Advanced Vector Extensions, these solutions deliver seamless orchestration from core to edge to support hybrid and telco cloud deployments, with QuantaEdge EGX74I-1U and QuantaEdge EGX88D-1U systems extending high-performance compute to the edge.□
- QCT Application-Ready Solution (ARS): A portfolio of pre-validated solutions combining

optimized software stacks with QCT's high-performance infrastructure to accelerate application deployment across industries. Designed to streamline integration from infrastructure to applications, QCT ARS enables enterprises to rapidly adopt scalable and efficient solutions for diverse vertical use cases while reducing deployment complexity and shortening time-to-market. □

QCT's solutions featured at COMPUTEX 2026 and powered by Intel Xeon 6 processors reflect a shared commitment by QCT and Intel to advance open ecosystems and industry standards. Through close collaboration, the two companies continue to address critical challenges such as energy efficiency and scalability, as organizations modernize their infrastructure to support next-generation workloads. To learn more details, visit QCT Booth #G0042 on the 3rd floor of Hall 1, in the Nangang Exhibition Center, or visit the QCT website at www.QCT.io □

© Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries.

Other names and brands may be claimed as the property of others.

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.

QCT Marketing Communication

QCT

marcom@qct.io

Visit us on social media:

[LinkedIn](#)

[Facebook](#)

[X](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/916484786>

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.

© 1995-2026 Newsmatics Inc. All Right Reserved.