



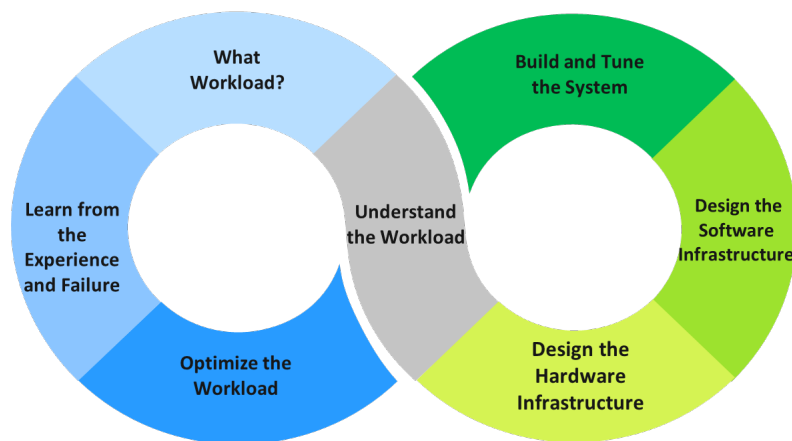
An Adaptive Platform for Converged HPC/AI workloads

Introduction

High-performance computing (HPC) uses a significant amount of computing power and parallel processing techniques for solving complex computational problems and performing research through computer modeling, simulation, and analysis. Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI) use algorithms and techniques that iteratively learn from data to allow computers to find hidden insights without being explicitly programmed where to look. Generally speaking, AI and HPC systems exist in separate environments. However, Quanta Cloud Technology (QCT), through years of extensive experience with numerous customers, has found that a converged HPC and AI environment can benefit customers and while remaining flexible and meeting their workload demands.

Converged HPC and AI environment

Experience with AI and HPC has enabled QCT to design a converged environment with a flexible infrastructure for customers running different workloads. HPC is associated with compute-intensive workflows, while AI tasks are more data-intensive. These workflows require different environment settings, and QCT finds mutual benefits and brings them together to resolve customers' complex issues. QCT improves AI training through HPC paralleled architecture. Based on experience with customers and continuous improvement, QCT developed QCT Platform on Demand (QCT POD), a concept that's ideal for running diverse workloads on one system infrastructure, including QCT's optimized management and monitoring system. Below is a diagram of how QCT works with customers to continuously deliver new solutions. With a deep understanding of the customer's workload, QCT can accurately judge the best-fitting hardware and software infrastructure. Through system tuning and optimization of the workload's aspects, QCT aims to build a workload-driven design environment for customers.



QCT Platform on Demand (QCT POD)

QCT's proven approach to getting customers' HPC and AI systems up and running quickly is to deliver all the components integrated and tested. The components should be selected based on the known and anticipated workloads and optimized for these requirements.

QCT Platform on Demand (QCT POD) is based on the concept of an on-premises, workload driven, integrated design system. QCT will provide various solutions with optimized hardware and software integration for specific workloads. QCT POD's system architecture includes a set of diverse common building blocks, which have greater flexibility and scalability, fulfilling customers' various demands. QCT POD also provides management tools to simplify the deployment and management process. With a pre-validated and pre-configured rack-level system design, QCT POD solutions also include all necessary power and network accessories, decreasing implementation time for customers.

QCT POD Building Blocks

QCT POD comes with all building components installed, configured, and ready to go. The components have been selected for the highest performing workload requirements. QCT uses Intel® Next Generation Server technologies, including Intel® Xeon Scalable Processors. The Intel® Xeon Platinum 8280, combined with QCT servers, delivers excellent performance on HPC and AI scenarios and benchmarks. Performance is further heightened when the system is bundled with Intel® Optane persistent memory and combined with the Intel® compiler.

The software stack is implemented with open source software and specific pre-validated commercial software, selected for maximum performance, reliability, and flexibility. QCT POD components and building blocks reflect its inherently flexible design and ability to be catered to customer needs. Different CPUs or GPUs can be integrated, various storage options can be selected, and the network fabric can also be tailored to customers' demands. QCT POD's components are demonstrated by Figure A.

QCT POD delivers first-class performance for a variety of workloads, utilizing flexible building blocks and QCT's engineering expertise. QCT POD's system design allows customers to realize continuous success.

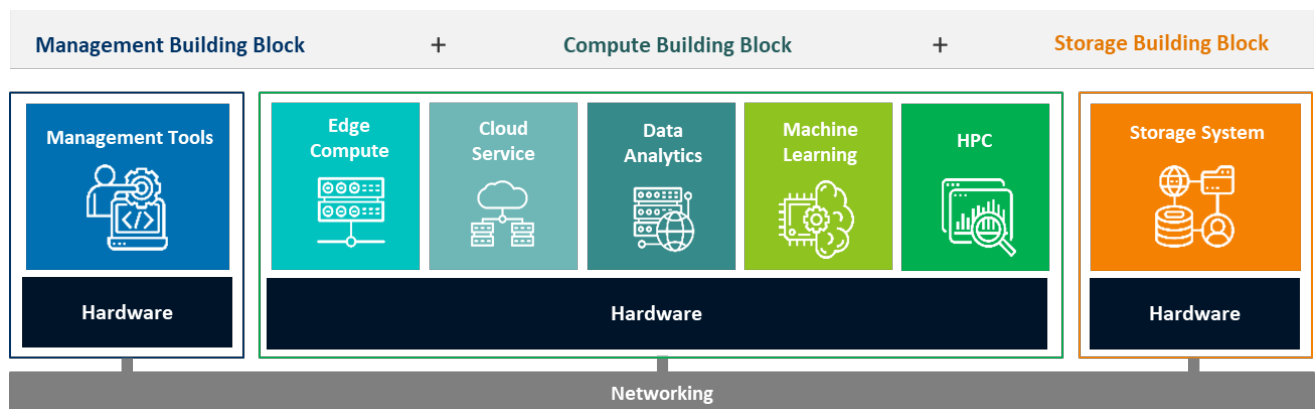


Figure A: QCT Platform on Demand Building Blocks

About QCT

Quanta Cloud Technology (QCT) is a global data center solution provider. We combine the efficiency of hyperscale hardware with infrastructure software from a diversity of industry leaders to solve next-generation data center design and operation challenges. QCT serves cloud service providers, telecoms and enterprises running public, hybrid and private clouds.

Product lines include hyper-converged and software-defined data center solutions as well as servers, storages, switches, integrated racks with a diverse ecosystem of hardware components and software partners. QCT designs, manufactures, integrates and services cutting-edge offerings via its global network. The parent of QCT is Quanta Computer, Inc., a Fortune Global 500 corporation.

To learn more about the QCT, please visit us at: [here](#)

