QCT POD MODERN CONVERGED PLATFORM FOR HPC/AI WORKLOADS

Produced by Tabor Custom Publishing
in conjunction with:

HPC wire

QCT | **intel.**

The massive growth of data and the need to process, store, and analyze the data is requiring many industries to revise their infrastructure and environments. Industries such as finance, manufacturing, transportation, medical as well as Higher Education and Research (HER) now use Artificial intelligence (AI) to help analyze their data. Life science organizations and researchers use High Performance Computing (HPC) and AI in areas such as Next-Generation Sequencing (NGS), medical imaging, and Numeric Weather Prediction (NWP) research. These industries are increasingly moving to adopt infrastructure capable of running HPC&AI workloads. The convergence of HPC and AI use in industry is driven by their workloads.

However, building an integrated workload driven system with optimized performance is time consuming and requires many resources. Traditionally, separate systems with different hardware and software were required for running HPC and AI workloads. Organizations are looking for an integrated system which is capable of running both HPC and AI workloads to meet their demands for enhanced workload processing, storage, and analysis efficiency.

This paper describes how to develop the QCT POD concept which is a Platform on Demand solution with common building blocks used to design on-premises, rack-level systems with hardware and software integration capable of running both HPC and AI workloads on one system infrastructure. QCT POD systems powered by Intel contain pre-validated and pre-configured custom system settings designed to meet various workload requirements. The paper provides examples of how the QCT POD for Medical and QCT POD for Higher Education & Research (HER) and solutions are used in NGS and in weather prediction research.

QCT is a leading cloud datacenter solution provider with extensive experience in developing HPC and AI solutions for companies in a variety of fields. The QCT POD solution provides organizations with building blocks designed to meet business demands from various workloads. With rack-level integration, QCT enables an organization's ability to reach time and resource efficiency for quick deployment and management of their system.

Introducing QCT POD—
Building a Workload Driven Solution

The QCT POD solution provides an on-premises system with best practice hardware and software integration for specific workloads designed for HPC and AI. With the common building blocks, it has greater flexibility and scalability to meet business demands from different industries.

The QCT POD solution helps shorten the time required for demand pattern analysis, system architecting, deployment, development, tuning and installation from months to days as shown by the steps in Figure 1.

Figure 1. End-to-end Journey of Building a Workload-driven Solution

QCT delivers a QCT POD system with pre-validated and pre-configurated system settings, pre-compiled libraries, compilers, and environment modules which make it easy to launch commonly used tools. QCT develops a unique workload driven solution for each industry which provides integration flexibility, a reference system approach and is easy to deploy.

QCT POD Infrastructure Meets Workload Demands

QCT POD provides development environments and configurations which can be used to fine-tune workloads and streamline workflows. Workload solutions are included for QCT POD such as Next-Generation Sequencing, Quantum Chemistry, Computational Fluid Dynamics, Finite Element Analysis, Electronic Design Automation, Computer-Aided Engineering, and Weather/Climate Modeling.

QCT POD architecture is composed of a set of building blocks to ensure a high degree of flexibility and scalability. The building blocks include the Management Building Block, Compute Building Block, and Storage Building Block that are each connected by a network fabric that can be customized to fit specific user workload demands.

Figure 2 shows the architecture of the QCT POD solution that provides dedicated management services capable of meeting diverse workload demands and processing needs. QCT storage provides a storage building block which is comprehensive and flexible. The QCT infrastructure supports edge computing, cloud services, data analytics, machine learning, and HPC resource and workload manager.
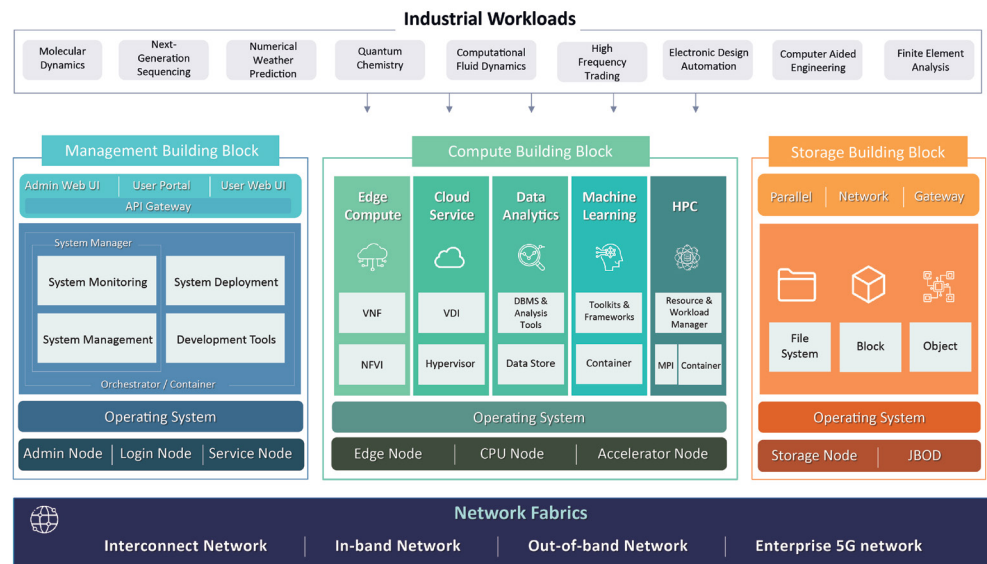

Figure 2. QCT POD System Overview

QCT POD Building Blocks for Administrators

Automated System Deployment

It is a challenge for administrators to deploy and manage a cluster with converged infrastructure and diverse services. The QCT POD solution contains a one-key deployment tool, which administrators can use to simplify their deployment process and help alleviate IT staff stress. The QCT POD solution can shorten the offline system deployment process from days to hours to set up a new cluster with hardware, container, and storage environments. It allows administrators to provision their software environment automatically and also provides the latest BIOS and BMC firmware, and Graphic Processing Unit (GPU) accelerators firmware and drivers.

- Features automated system deployment to simplify the originally complicated and error-prone configuration setting process. Users can complete overall system deployment by inputting the information required via a dedicated interface. With consistent system configurations, the system not only saves time, but also prevents failures and ensures normal system operations.
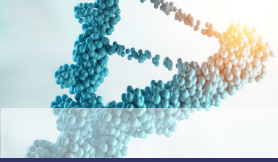
- Supports offline installation of bare-metal operation systems, virtual machine environments, and container orchestration services with built-in software packages and container image repositories, saving software downloading time. During deployment, the system automatically provides optimized firmware and driver packages for the hardware platforms to enable the system to start running in the shortest period of time with the best configurations.

Real-time System Monitoring & Simplified Cluster System Management

QCT POD provides real-time monitoring support including web-based dashboards and system alerting with useful administration tools to simplify management and ease administrators' operational burden. It enables administrators to monitor and interpret the system status easily for both container and bare metal environments.

As shown in Figure 3, administrators can check the resource utilization for Central Processing Units (CPUs), memory, disk, and GPUs. Management tools also simplify cluster system management for log, resource, containerized applications, event alerts, and account management. An integrated job scheduling system helps resolve job conflict issues because the system performs automatic job queuing. It makes QCT POD become highly reliable, high available, and scalable.

- Offers popular performance testing and evaluation tools for computing platforms, network equipment and storage systems to ensure the performance meets the designated service level.

- Consolidates user account and group management functions, and supports multiple authentication mechanisms.

- Supports high-availability standby resources for the system's core services.

- Supports load balancing for web services.

- Supports multiple software-defined storage systems, and provides automated configuration management and system integration functions.

- Features automated resource management and job scheduler functions to allow multiple users to share system resources and execute applications interactively or in a batch manner.

- Features an integrated container orchestration platform that enables automated container deployment, scale-out and management.
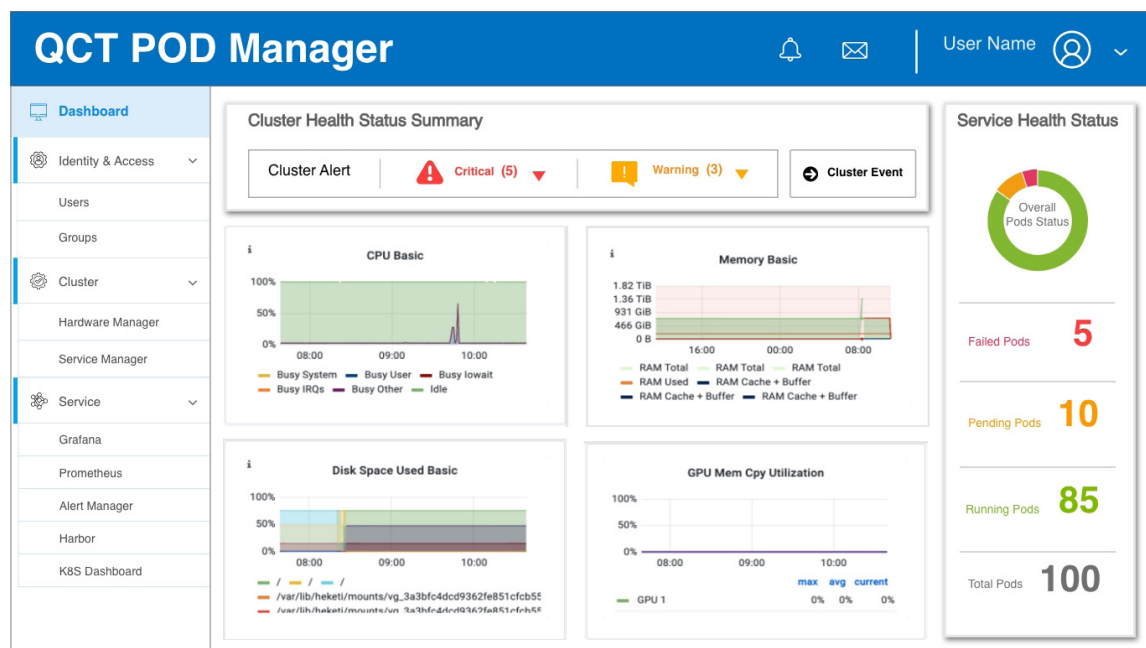
Figure 3. QCT Web-based Dashboard

QCT POD Building Blocks for Developers & End Users

Robust Software Development Environment

The QCT POD system includes a robust development environment that makes it easier for developers and users to launch their tools and helps streamline AI workflow and increase HPC workload performance. It is composed of pre-validated workload packages as well as pre-installed diverse development tools composed of compilers, libraries, application frameworks, and data visualization tools.

The system allows users to automatically provision a Jupyter notebook environment and multiple developers can access the web-based computational platform to run code and do data analysis interactively. The QCT POD platform makes work easier for developers because they can focus on their area of expertise and reduce their efforts for environment preparation.

- Offers pre-defined application environment management modules, which enables dynamic setting loading and deleting of system environment variables. It also supports dynamic switching among multiple versions to allow users and applications to select the right system environment variables.

- Provides mainstream compliers (e.g., C, C++, and Fortran), libraries, debugging tools and various application frameworks to meet the development needs of a wide range of applications.
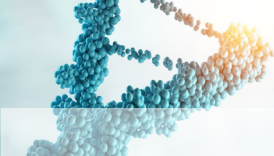
- Features a virtual system console, with which users and program developers can log in either in the text or graphics mode via a web browser or a Secure Shell (SSH) Client. Users and program developers can also run X11 applications and Jupyter Notebook for program development or data analysis.

- Offers free and popular data visualization tools to help users analyze data or generate charts and graphs.

- Supports both bare metal and container service environments, and allows multiple users to develop, test and execute containerized applications with automated resource management and a job scheduler.

Fine-tuned Application Workflows & Workloads

QCT POD provides customized, fine-tuned execution binary files for open-source and community versions of AI & HPC applications commonly used in research fields in Medical and Higher Education and Research. This not only saves time and efforts spent on application compilation, but also simplifies the operation process and provides the best application performance with pre-defined environment management modules selected based on the systems' platforms.

QCT POD Hardware Solutions
Use the Latest Intel Components

QCT POD systems include all system components installed, configured and ready to use. QCT QuantaGrid D53X, QuantaGrid D53XQ, and QuantaGrid T43Z servers available as part of the QCT POD solution are optimized for HPC/AI. The server includes 3rd Gen Intel® Xeon® Ice Lake Scalable processors that are designed to efficiently use processor cores which can increase processors speeds.

The Intel® Advanced Vector Extensions 512 (Intel® AVX-512) provides HPC support while the while Intel® Vector Neural Network Instructions (VNNI) provides an AI inference instruction set. Intel® Optane™ Persistent Memory streamlines computing by combining storage capacity and intelligent system acceleration.

Introducing QCT POD for Medical

QCT developed QCT POD for Medical with common building blocks designed to meet different precision medical demands. The solution includes dedicated tools for rapid deployment, and easy to use management system tools. The solution supports diverse workloads with resources and tools for data processing, analytics, HPC and AI.
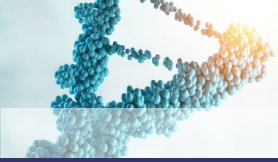
QCT POD for Medical is designed to improve the data flow of medical workloads. QCT POD for Medical building blocks include environments which aid in easily using tools for NGS, molecular dynamics research for protein analysis, and medical image recognition. QCT ensures the quality and serviceability of infrastructure, which could accelerate time to value on cancer genes detection, drug design for medical researchers as well as aid in medical imaging diagnosis.

QCT POD for Medical—Meeting the Needs of Next Generation Sequencing (NGS) Research

An important tool in predictive medicine and genomic analysis is the use of NGS, a form of DNA sequencing that analyzes individual strands of DNA simultaneously. Processing and storage of NGS data are major pain points for organizations because legacy compute, storage and file systems are often based on old designs which cannot meet the demands of modern NGS workflows. NGS workflow generates massive amounts of data during the primary, secondary, and tertiary gene analysis phases. Organizations need an adaptive HPC infrastructure that can handle NGS AI workloads.

The QCT POD for Medical solution powered by Intel provides pre-configured hardware optimized for performance in processing NGS data. In addition, QCT POD for Medical contains pre-complied and pre-built workloads packages, pre-installed diverse development tools, compilers, libraries, application framework, and data visualization tools. The Burrows-Wheeler Aligner (BWA) and GATK tools and environments commonly used in gene data analysis are included as part of the QCT POD for Medical building blocks. Using the QCT POD for Medical solution allows researchers to focus on their research without spending so much time on environment and configuration settings.

QCT POD for HER— Meeting the Needs of Higher Education and Research

QCT POD for Higher Education and Research (HER) is a pre-defined end-to-end HPC/AI converged platform purposely built for HPC Research in the Higher Education and Academia field. QCT POD for HER is composed of cluster management software, built-in application frameworks, compute-optimized hardware, storage technologies and low-latency networking seamlessly integrated and validated to accelerate the HPC research process.

QCT POD for HER is designed with "simplicity" in mind to streamline the deployment and management of the system for administrators and reduce time-to-simulation for developers and researchers with built-in ready-to-use tools and frameworks. QCT POD for HER is built to be a converged platform to support both traditional HPC and AI workloads to eliminate the complexity of managing siloed systems to reduce overall Total Cost of Ownership (TCO).

QCT leverages experience in building supercomputing systems and domain knowledge in the HPC research field to ensure the reliability, serviceability, and optimized performance of the solution. QCT POD for HER can be used to assist research and academic institutions in solving complex problem in different research fields such as weather forecasting, imitating an atom's movement and interaction or understanding electronic structure, astronomy research, new material development and drug discovery.

How QCT POD for HER Accelerate Time to Simulation

For academic researchers, how to solve more complex problem in less time is what they care about the most when they are running their HPC workloads. Educational and academic HPC workloads are very diverse and each of them are not optimized in the same way. This creates challenges for the researchers. To successfully run their simulation, researchers have to go through the time-consuming process of managing the software dependency and also compile and configure the application.

Commonly used HPC workload applications are included in QCT POD's HPC Workload Package, which is a selected package of pre-compiled and pre-configured applications with relevant libraries, compilers, and Message Passing Interfaces (MPIs) that is delivered in the form of environment modules to allow researchers and developers to focus on their research instead of spending time fine-tuning their workloads. The QCT POD HPC workload package covers the most commonly used applications in the academic researcher field such as Numeric Weather Prediction (NWP), and Quantum Chemistry and Molecular Dynamics (MD). Figure 4 shows how pre-configured and pre-compiled tools simplify the process for researchers to run a simulation.
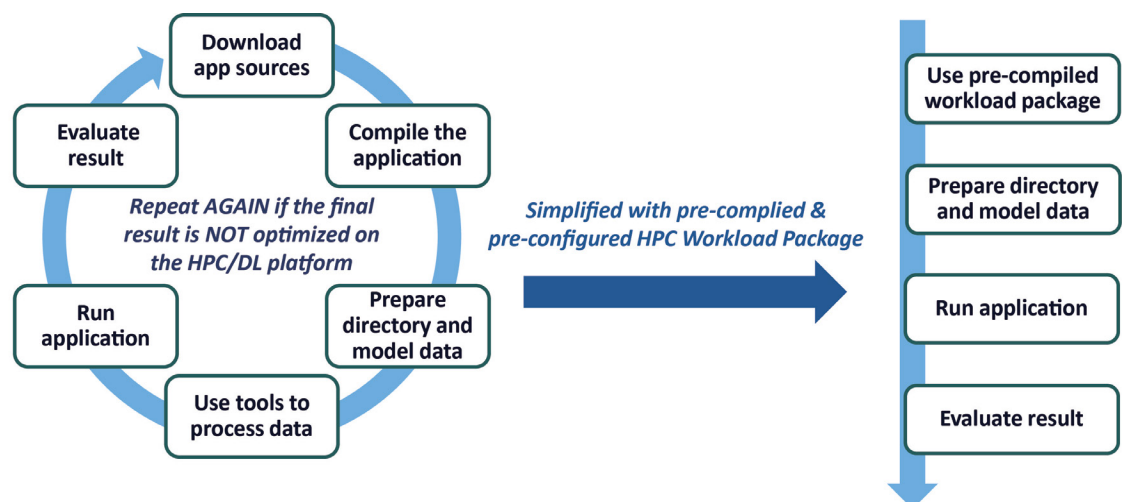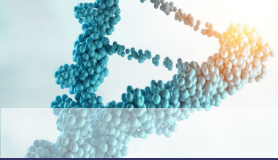

Figure 4. Simplified Process to Run Simulation

Summary

Industries including manufacturing, transportation, life sciences, medical, as well as higher education and research have large workloads that require HPC infrastructure to meet processing, storage, and analysis needs. In addition, HPC and AI are increasingly being used across these industries to gain insight from their massive amounts of data. Many organizations have an infrastructure that is not capable of HPC processing or meeting their AI needs.

Today's workloads drive the need for an integrated system capable of running both HPC and AI workloads. However, it is time consuming and requires many resources to build an integrated system. QCT has expertise on end-to-end HPC and AI solutions for companies in a wide variety of fields such as medicine, higher education and research. QCT developed a platform called QCT POD that is specifically designed to meet the needs of life science and medical companies doing NGS research as well as higher education and research users doing research in areas such as weather prediction.

As described in this paper, QCT POD powered by Intel can provide an integrated solution to meet HPC and AI workloads needs. The QCT infrastructure and QCT POD solution provides pre-validated and pre-configured custom system hardware and management tools which optimize system performance and can help organizations increase their Return on Investment (ROI).

—

For more information on QCT and how QCT POD can help your organization, see:

QCT POD: https://go.qct.io/qct-pod/

QCT POD for Medical: https://go.qct.io/qct-pod/qctpod-for-medical/

QCT POD for Higher Education and Research:
https://go.qct.io/qct-pod/qctpod-for-higher-education-and-research/

—

Intel, the Intel logo, Optane, and Xeon Inside are trademarks or registered trademarks of Intel Corporation in the U.S. and/or other countries. All trademarks and logos are the properties of their respective holders.