A PURPOSE-BUILT HPC/AI SOLUTION FOR HIGHER EDUCATION AND RESEARCH

Produced by Tabor Custom Publishing
in conjunction with:

HPCwire

QCT | intel

Executive Summary

Researchers and scientists seek to use high-performance computing (HPC) to solve some of the longest-standing mysteries of the universe and the greatest problems facing humankind today. In many cases, it is in a nation's interest to grow local and regional higher education HPC capacity.

From the origins of the universe to new therapeutic treatments to the impact of climate change, many traditional lab-based science and research efforts now all require HPC to conduct simulations, models, analysis, and to visualize vast amounts of data.

These new approaches to scientific research are driving the demand for HPC resources that can process larger data sets per run. The HPC on-premises market is expected to reach nearly $19 billion in 2024, with a five-year CAGR of 6.8% from 2019 to 2024, according to Hyperion research.

Common HER HPC Challenges Encountered

HPC workloads across Higher Education and Research (HER), which includes university and academic institutions or government institutions, are getting increasingly diverse and more demanding. Users and organizations want to run progressively more sophisticated models, simulations, and analytic routines in order to get results and insights in a shorter time frame which is critical to all areas when making new discoveries.

An additional factor that must be considered is that many more institutions are using HPC in comparison to just a few years ago. Many individuals, departments, or groups in HER who never used HPC in the past, now see its benefits. Numerous academic research efforts that were exclusively lab-based operations now require HPC to model and analyze data or derive insights from vast volumes of data.

As such, there is a need for easy-to-use systems that let data scientists, academics, researchers, and others running the workloads to tend to their areas of expertise without saddling them with IT or systems management chores.

Why a Purpose-Built Solution is Needed for HER

There is an ever-growing need for HPC resources to accelerate the wide variety of academic computing and research workloads. Fortunately, there are technologies such as processors, accelerators, high-speed storage, and interconnects, HPC cloud instances, etc., that can help.

However, this wealth of compute, storage, and interconnect options complicates matters. An institute must match the right combination of technologies with the workloads that they need to run. The challenge is finding the right combination, optimizing it, and keeping the solution running over time. And then, that solution must be integrated with the numerous other technologies that comprise a complete workflow from pre-processing, modeling, simulation, analysis, and visualization.

Further complicating matters is that home-built solutions are quite complex from a software point of view with many moving parts. And things continue to get more complex over time. Distributed computing, schedulers, MPI, OFED, and multiple software layers make it complex to develop a working solution that keeps on running over time.

Such complexity should be considered in determining how HPC capabilities are acquired. Certainly, an institution can select discrete elements for its end-to-end workflow. But in many cases, such a do-it-yourself (DIY) approach is not the best choice, as integrating various components and optimizing them consumes vast amounts of time that could be spent on core issues.

In many cases, an organization may not have the expertise to carry out a DIY effort. Therefore, they need to build up their own experiences in component selection, assembly, and the fine-tuning of performance for a particular workload. Then, even if the expertise exists, most organizations prefer that their scientists, researchers, or engineers aren't spending their own valuable time assembling and managing their own HPC systems.

As a result, buying a complete solution is a better choice when selecting a system for modern higher education and research workloads.

QCT POD for Higher Education and Research

QCT has developed such a system powered by Intel® Xeon Scalable Processors. QCT POD is a purpose-built system designed and pre-configured for Higher Education and Research. QCT POD, which is a platform on-demand, provides the infrastructure and system management needed to meet HER processing and storage requirements.

The QCT POD for Higher Education and Research solution offers an innovative technology system with building blocks designed to meet different research demands, including those found in numerical weather prediction (NWP), quantum chemistry, and computational physics.

It offers an optimized run-time environment for HPC/AI workload applications for a streamlined HPC/AI development environment and workflow. Additionally, the system includes cluster management software, built-in application frameworks, compute-optimized hardware, storage technologies, and low-latency networking, all seamlessly integrated and validated to accelerate the HPC research process.

Compared to assembling a home-grown solution, using the QCT POD simplifies the journey from system preparation to deployment phase and system management and monitoring. And once it is installed, the support continues. QCT ensures the quality and serviceability of the infrastructure, with their own QCT HPC experts being available to offer system configuration consultation and benchmarking and tuning services.

Using QCT POD provides several benefits to different stakeholders. For example:

- End users achieve faster productivity and time-to-results, getting deeper insights faster.
- System administrators get a solution with built-in system management and monitoring, removing the burden of the many chores they would otherwise need to conduct to keep the system running.
- Developers are provided with a pre-built and comprehensive environment to do their work, eliminating the need to assemble a system every time a new application is created.

Taking a Deeper Look at the QCT POD for HER

The QCT POD for HER is a pre-configured and pre-validated system that integrates hardware and software to deliver its promised performance. It is an HPC and AI converged platform to support both types of workloads with a built-in development environment and framework.

Diagram 1 demonstrates the building blocks of QCT POD for HER. The system consists of compute, storage, management, and networking blocks. It also includes the QCT HPC Workload Package to speed up a researcher's time to simulation and the QCT POD Manager to provide a user-friendly dashboard and to ease monitoring and management.

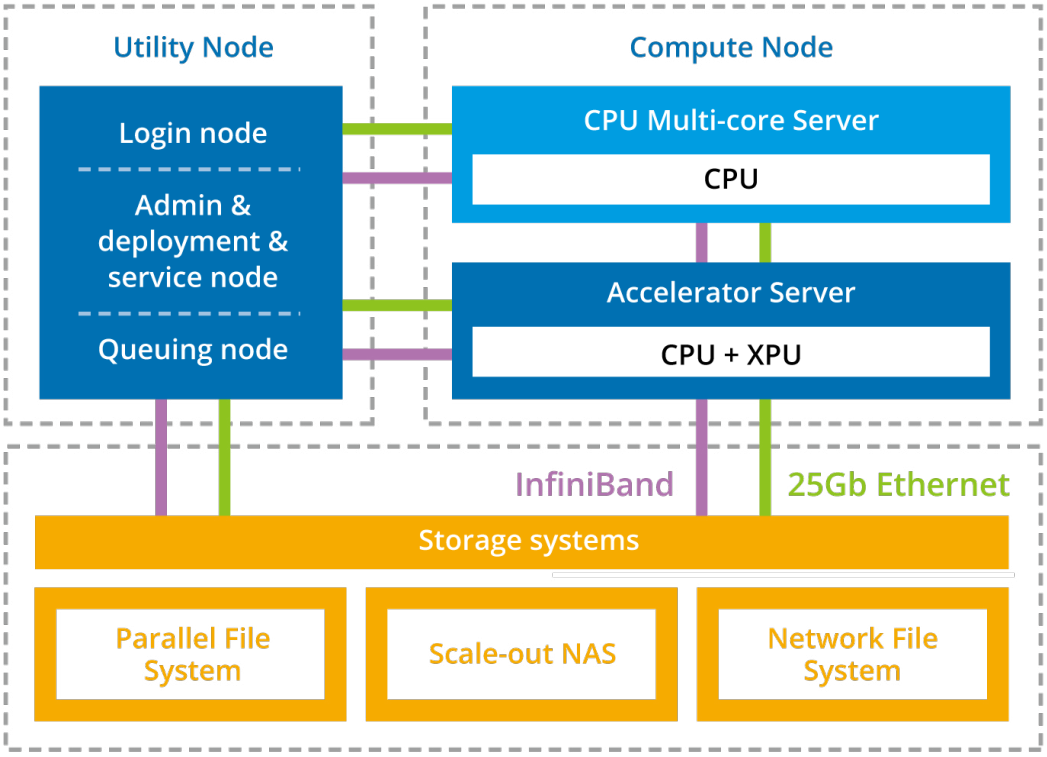

Diagram 1.

Taking a deeper look at the QCT POD for HER suite in diagram 2, QCT defines four levels to meet different researchers' demands and workloads. QCT POD for HER will deliver QCT's best-in-class hardware with its integrated software and system environment to allow researchers to achieve "simplicity" and "efficiency" all the way from system planning & preparation, implementation, operation, and maintenance.

Building Blocks			Starter		Standard		Advance		Premium	
			Qty	Specification	Qty	Specification	Qty	Specification	Qty	Specification
Utility Node (Login, admin, service and queuing)	All in one utility& storage node		1	QuantaGrid D53XQ-2U	1	QuantaGrid D53XQ-2U				
	All in one Utility Node						1	QuantaGrid D53X-1U	1	QuantaGrid D53X-1U
Computing Node	CPU Node		4-6	QuantaGrid D53X-1U	6-8	QuantaGrid D53X-1U	6-8	QuantaGrid D53X-1U	6-10	QuantaGrid D53X-1U
Storage System BeeGFS Filesystem	Metadata Server						1	QuantaGrid D53XQ-2U	1	QuantaGrid D53XQ-2U
	Object Storage Server						2	QuantaGrid D53XQ-2U	2	QuantaGrid D53XQ-2U
	Storage Enclosure for Object Storage Target						1	QuantaVault JB4602	1	QuantaVault JB4602
Networking	Data& Management Network	25G Ethernet Switch	1	QuantaMesh T4048-IX8D	1	QuantaMesh T4048-IX8D	1	QuantaMesh T4048-IX8D	1	QuantaMesh T4048-IX8D
	Management	Management Switch			1	QuantaMesh T1048-LB9M			1	QuantaMesh T1048-LB9M
	Data Network	Infiniband Switch			1	HDR Infiniband Switch			1	HDR Infiniband Switch

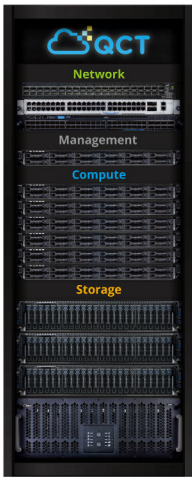

Diagram 2.

With regard to compute and storage capabilities, QCT adopts in-house hardware technologies as listed below:

QuantaGrid D53X-1U

A balanced 1U architecture compute server each equipped with dual 3rd Generation Intel® Xeon® Scalable Processors, which is optimized for AI, HPC, data analytics workloads due to advanced CPU Instruction set such as Intel® AVX512, VNNI, DDIO. The model also supports 2nd Gen Intel® Optane™ PMem 200 Series Barlow Pass on.

QuantaGrid D53XQ-2U

Designed with scalable and flexible NVMe, each is equipped with dual 3rd Generation Intel® Xeon® Scalable Processors, to support for HPC/AI storage nodes and deliver high performance for a parallel file system or scale-out NAS. QuantaGrid D53XQ also can support 2nd Gen Intel® Optane™ PMem 200 Series Barlow Pass to boost memory capacity.

Storage System

QCT offers different storage systems based on workloads and user demands, as shown in diagram 1. In terms of the file system, QCT offers a BeeGFS Parallel file system to enable easy-to-use features and is a great fit for I/O intensive HPC workloads. QCT has a team with expertise using the BeeGFS Parallel File System. Through such experiences, QCT offers a system that is:

- **Easy to deploy:** the setup process of BeeGFS has been streamlined and automated through QCT Rapid Deployment Kits.
- **Optimal architecture design:** the most optimal BeeGFS design in terms of hardware design, software architecture, and configurations to deliver optimal performance.
- **Unified management:** a GUI dashboard to perform administrative management tasks and to monitor the state of the file system and its components.

Deployment and Management Made Easy

The QCT POD for HER is simple to deploy. The solution features an automatic offline system deployment and scaling of bare-metal environment, automatic software provisioning and configuration, and a built-in software repository for a bare-metal and containerized environment. Managing these aspects of the system frees up the IT staff to these tasks that they would normally perform.

Once a system is installed, additional built-in features ensure that it keeps running with minimal IT intervention. The QCT POD Manager offers real-time system monitoring & simplified cluster system management to streamline administration workflow.

Ready to Use HPC Applications

The QCT POD includes the QCT HPC Workload Package, which is a package of pre-compiled and pre-configured HPC applications that are delivered in the form of environment modules to allow researchers and developers to run simulations instantly.

The QCT HPC Workload Package provides various types of HPC applications that are commonly used in different research fields.

Using their HPC expertise and HER domain knowledge, QCT pre-compiles and pre-configures the HPC applications with optimized and best-practice configurations. Managing these tasks means researchers and developers don't have to compile and configure these applications from zero and go through a trial-and-error process to find optimized configurations.

QCT also manages software dependencies. In a typical HER HPC site, there are multiple researchers and users working on different research projects. Each of these projects runs different applications that are dependent on different compilers, libraries, and MPIs to work. Researchers and developers often spend quite a bit of time to ensure the software dependencies are set properly before actually running simulations. For admins who have to get these applications ready for users, installing the application and managing the dependencies is a complex and time-consuming process.

The QCT engineering team's experiences and domain knowledge in these applications, and their extensive work to manage the complex software dependencies, is used to built the QCT HPC Workload Package for users and admins, allowing them to use the software without worrying about dependency issues.

Real-World Proof Point

QCT accumulated numerous experiences in building the Taiwania 2 and 3 supercomputers for HPC/AI research fields to ensure the reliability, serviceability, and optimized performance of their solutions to assist a national laboratory and academic institutions solve complex problems in different research fields such as weather forecasting, simulating atom's movement and interaction, understanding electronic structures, astronomy research, new material development, and drug discovery.

QCT built its Taiwania 2 supercomputer for a national laboratory in Taiwan. During its development, QCT established an AI cloud computing platform, which could provide real-time and convenient computing services to industries, universities, and higher institutional circles. QCT delivered an AI cluster that includes x252 compute servers which ranked in the Top 500 list of the world's most powerful supercomputers, where each server is equipped with 2 CPUs and 8 GPUs. Furthermore, the system adopted direct-to-chip liquid cooling to enable superior energy efficiency to place it concurrently on the Green 500 list.

QCT built another Taiwania 3 supercomputer, adopting ultra-high-density 2U4N compute nodes. With 900 compute nodes in the cluster, it also ranked on the Top 500 list. For the large number of compute nodes, QCT introduced QCT Rapid Deployment Kits to simplify the system deployment journey and provide customers with built-in software environments to accelerate time-to-market.

Working with a Technology Partner

QCT develops its own solution centers at its headquarters across hundreds of nodes to streamline systems through configuration, benchmarking, and the certification process to industrial standards to ensure key workload performance.

QCT has HPC experts in molecular dynamics, chemistry, weather and climate, and next-generation sequencing. With their years of experience with cloud service providers, and various industries, QCT also provides benchmarking and tuning services for applications in customized areas upon request.

To learn more about QCT POD for HER, visit:
https://go.qct.io/qct-pod/qctpod-for-higher-education-and-research/

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 datacenter 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, storage, switches, integrated racks with a diverse ecosystem of hardware components and software partners. QCT designs, manufactures, integrates, and services cutting-edge offerings via its own global network. The parent of QCT is Quanta Computer, Inc., a Fortune Global 500 corporation.

http://www.QCT.io

—