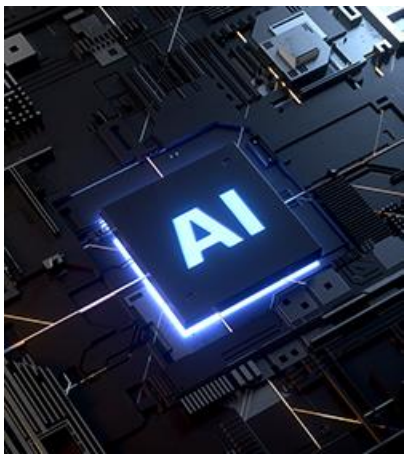
www.pipelinepub.com
Volume 17, Issue 3

Advancing Peak Performance in HPC Environments

By: George Clement



Whether deployed on premises, in the cloud, or at the edge, high performance computing (HPC) solutions are now transforming a wide range of industries, including financial services, healthcare, manufacturing, oil and gas, and research and educational institutions.

Often integrating artificial intelligence (AI) and machine learning technologies, HPC is able to process data and perform complex calculations at speeds up to quadrillions per second. Today, the data science applications, large-scale analytics, and recommendation engines driven by HPC are helping to detect credit card fraud, make faster and more accurate patient diagnoses, and assist scientists to find sources of renewable energy.

Indeed, HPC is more vital than ever as both public and private organizations seek faster paths to solving their toughest technical and research challenges. According to Grand View Research, the global HPC market is expected to increase at a compound annual growth rate (CAGR) of 6.5 percent from 2020 to 2027 to reach $53.6 billion, up from $39.1 billion in 2019. There was a time when HPC systems were primarily used by the aerospace and navigation industries. However, the growing adoption of cloud computing, continuous developments in AI, and the rising need for business analytics have prompted a broad expanse of industries to adopt HPC systems.

Game-changing technology—with challenges

While HPC is undoubtedly a technology that is powering game-changing projects, in the data center, HPC environments present several formidable challenges. Chief among these are power

requirements, which can mean high energy costs. HPC also requires denser banks of computer resources to increase capacity and reduce latency while minimizing floorspace. An HPC cluster can comprise hundreds or even thousands of compute servers, and in order to avert unplanned downtime, special consideration must be given to future-proofing as it relates to power availability.

Running a high-power-density HPC deployment also generates significant heat. In fact, due to the size of the compute workloads and density, servers in an HPC environment can run approximately 30 percent hotter than traditional computing platforms. Especially in an older data center where the cooling systems were designed to accommodate significantly lower power densities, HPC will expose ineffective or insufficient cooling capacity.

While large cloud and hyperscale data centers utilize ultra-efficient cooling systems, at legacy facilities, cooling alone can account for 30 to 40 percent of the power costs of the data center. To offset this liability, the common response is to overcool a facility. However, this leads to wasted electrical energy and an expansion of a data center's carbon footprint.

Real-time data and alerts

Now that we understand some of the challenges that HPC presents to the data center, let's take a look at how one world-renowned research institute was able to gain real-time power, thermal, and utilization analysis in its high-performance computing environment, thereby improving server utilization and uptime.

The Institute for Health Metrics and Evaluation (IHME) is an independent global health research center at the University of Washington. The IHME provides rigorous measurement of the world's most pressing health problems and evaluates the strategies used to address them. IHME makes this information freely available so that policymakers have the evidence they need to make informed decisions about how to allocate resources to best improve population health.

The IHME IT staff deployed Intel® Data Center Manager (Intel® DCM) to monitor more than 600 servers in its HPC data center environment at the university's colocation facility. Intel® DCM is a software solution that collects and analyzes the real-time health, power, and thermals of a variety of devices in data centers, providing the clarity needed to improve data center reliability and efficiency.

Staff installed Intel® DCM to gain greater insight into power demand, thermal efficiency, server utilization, and capacity planning. Intel® DCM does not require the installation of any software agents on managed nodes—and therefore does not impact performance.

The IHME IT team was immediately impressed by Intel® DCM's short learning curve, ease of use, and simplicity of deployment. Within hours, they were able to compile and aggregate actionable, real-time data from its collection of servers. Intel DCM eliminates the need for complex, device-specific configuration, setup, or customization.

The institute's IT team was able to use Intel® DCM to create the needed power statistics for every rack and server model with no additional hardware or software. Intel® DCM also enabled IHME IT staff to implement a power consumption policy, and the solution's health monitoring feature allowed the team to receive alerts based on custom power and thermal events, which will further ensure uptime.

"We learned a lot from other products about realistic maximum power consumption, but it was only relevant for historical data, it couldn't provide us the real-time alerting," said Vern Harbers, Technical Project Manager, Infrastructure IHME, at the University of Washington. "If something goes wrong in the data center, right now, other products couldn't tell us that. Intel DCM was easy to plug in, and easy to get the data and analysis from our machines immediately. The alerts and power limitations were set up within a day."

Increasing efficiency, utilization and uptime

Diving deeper into IHME IT staff's deployment of Intel® DCM in its HPC data center environment, we find that the solution enabled the team to quickly detect and analyze underutilized systems by monitoring CPU utilization and power consumption over time. Typically, the lack of sufficient workload performance monitoring leads IT administrators to purchase more hardware.

Efficient space and power capacity management is certainly an essential part of operating data centers. However, this becomes increasingly difficult when data centers grow in density and complexity, and with no easy way to get granular power consumption details. Using Intel® DCM, the IHME IT team was able to leverage a single solution for power management across all devices in the data center, supporting the multiple proprietary power measurement and control protocols required by different OEMs.

IHME IT staff members were able to use Intel® DCM to create the needed power statistics for every rack and server model with no additional hardware or software. Hence, they were able to better plan and manage capacity and utilization in racks, safely increase rack densities, and delay adding new racks. Additionally, Intel® DCM enabled the IHME IT team to maintain group power capping while dynamically adapting to changing server loads.

Furthermore, without the control and insight provided by Intel® DCM, it is difficult to gain an integrated view of a server pool and incomplete data sets offer at best limited visibility. A recent study sponsored by Intel® revealed that as many as 43 percent of data centers rely on manual research. However, Intel® DCM analysis allowed IHME to identify and redeploy long-term, low-utilization servers.

Today, HPC is being used to train autonomous robots, assist scientists to find new approaches to clean electricity, and help researchers to design next-generation aircraft, automobiles and shipping vessels that are safer, faster, and more fuel-efficient. Meanwhile, we are happy that Intel® DCM is being used to great advantage at IHME's HPC data center environment at the University of Washington's colocation facility and foresee a tremendous value for this technology solution across other HPC data center environments that are striving to increase efficiency, server utilization, and uptime.