Precision Medicine Speeds Ahead with High Performance Computing Clusters

Scalable, standards-based technology is fueling a Big Data turning point in processing genomes.

By Jeff Durst, Dedicated Computing

In genomics, huge data files are generated, which are then broken down into thousands of smaller fragments for effective analytics. Today this kind of precision medicine is advancing because clinicians can access and interpret these data files—patient information in the form of genetic sequences—faster than ever. This is High Performance Computing (HPC) cluster technology at work, delivering remarkable benefits for individual patients and helping advance the field of genomics decades faster than originally anticipated. Where precision medicine was thought to have global impact by 2050, the new target is 2020. 'One size fits all' is no longer the standard of care, and HPC technology is poised to deliver higher value at lower costs in the widest range of research and treatment environments.

Tapping into Big Data Faster

It once took weeks to benefit from healthcare data—collaborating and sharing data from labs, medical records, pharmacies, insurance companies, and specialists, and then turning that information into well informed treatment plans. HPC cluster architecture has cut this timeline down to a matter of hours. As exciting as that is, the real value is that this scalable, standards-based technology is accessible to hospitals, clinics, and labs of all sizes and resources, instead of just government labs and super-computing environments. On

a global basis, the spectrum of healthcare innovators can tap into HPC to increase the value of data while significantly lowering their costs. Reactions to treatments are faster and more responsive, saving lives in all types of treatment settings.

> "The real-world result is that a child in treatment for cancer could have genomic sequencing done in four hours instead of seven weeks."

It's data-driven healthcare, and it enables treatments tailored to an individual's DNA. The real-world result is that a child in treatment for cancer could have genomic sequencing done in four hours instead of seven weeks. Sequencing can even be done multiple times during treatment, monitoring the patient for drug resistance or relapse. Ineffective treatments can be modified more quickly, minimizing unintended damage to patient health, and fast-tracking the search for treatments with the best results.

Cluster Technology at Work

The "cluster" represents a group of computers all working in parallel, splitting the load of analyzing large data files with each computer solving its own part of the bigger problem. HPC recognizes that the solution to processing huge data files is to not only have more computing power, but also more computers that are purpose built to work together.

HPC assigns specific processes to different nodes in the system as a means of accelerating each individual process. This kind of targeted processing creates a much larger volume of data analysis and interpretation and enables scaling—unlike traditional computing systems. Adding resources, such as more nodes to handle more processes, does not add significant cost. Future applications and resources are protected, as nodes can easily be added for greater data analysis, interpretation, and storage needs.

Simplifying implementation is important to the growth of cluster computing, illustrated by the Intel® Cluster Ready program. Intel Cluster Ready architecture defines application and component interoperability, alleviating the challenges of integrating a collection of independent hardware and software
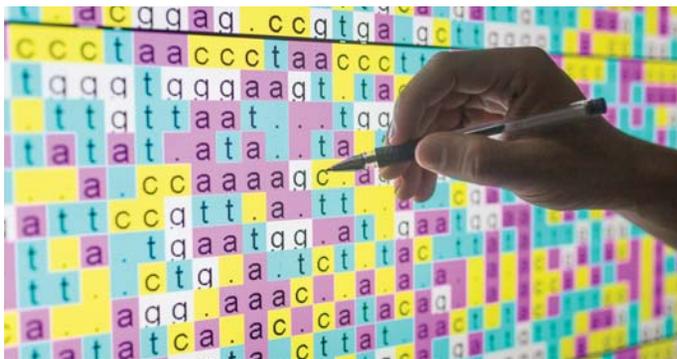

Figure 1: Data drives healthcare—and care improves when accurate, timely information is shared effectively. That is essentially the guiding principal of genomics, which creates highly personalized treatment plans based on an individual's genetic makeup. Fast data analysis is making a difference here, with technology at the heart of ambitious goals like improving outcomes and reducing the long-term cost of care for a worldwide population.

components for cluster performance. Interoperability is pre-validated, removing risk and complexity as well as reducing costs over the cluster's lifetime. For example, system security is already layered in. This enables ready compliance with industry regulations like HIPAA and access to sophisticated hardware-assisted security for accelerated encryption, anti-theft, identity protection, and malware detection.

Scaling Cluster Technology Across All Research Settings

Powered by Intel Xeon® processors and built upon Intel Cluster Ready architecture, a fully integrated and optimized cluster solution enables clinicians to simulate, analyze, and visualize complex, data-intensive models. These standards-based systems drop into existing business systems within a health network, scaling as needed and minimizing the burden of secure data integration with other systems such as electronic medical records (EMR) or laboratory information systems (LIS). It's this scalability that ensures affordability and industrywide access to sophisticated tools. For example,

Two Examples of Faster Insight

Data-driven healthcare is at a turning point, fueled by computing advances that both speed and simplify processes, adding new medical value with the immediacy of sophisticated genetic data. For example, according to use cases provided by Intel, one pediatric oncology center successfully reduced its genomics analysis time from seven days to four hours. With this speed of data access, clinicians can sequence patients multiple times during treatment, monitoring for signs of relapse or drug resistance and adjusting treatment strategies accordingly. Another research team provisioned a large cloud-based computing cluster to screen 10 million compounds in just 11 hours, defining three compounds to continue to next stage testing. Using traditional research methods, this is the equivalent of 39 years of science at a cost of $44MM; by capitalizing on HPC clusters, it was achieved at a fraction of the time and cost[1].

an entry-level network might feature 20 Terabytes and four multicore homogeneous nodes, ideal for a smaller research institution. At the same time, a super lab could deploy the same technology but incorporate multiple Petabytes of object storage and hundreds of nodes.

Genomic data requires object storage (rather than file storage), which is typically included in the head node and blended with a cluster management application to centralize coordination of the entire system. The head node itself is connected to a LAN or other network, linking the cluster to a human-machine interface (HMI) or local storage components. Additional nodes or "clients" connect via a secure fabric interface such as Ethernet. Workloads are executed in parallel, providing the basis for scalable high-performance computing.

Turning Data into Insight and Optimized Healthcare

Integrated cluster management tools further ensure ease of use. As a result, all types of precision medicine settings—from smaller medical groups to large research organizations—can access the value of HPC and easily manage their own unique performance requirements. Smaller labs are well-positioned to capitalize on hybrid systems, managing data with on-site high performance and connectivity to cloud-based analytics providers; larger installations such as government research centers can scale up infrastructure to handle sequencing and analysis of data entirely on-site.

Scientists and clinicians also benefit from Intel Cluster Checker, a tool that helps ensure cluster performance remains within required certifications over time. With system health managed remotely, medical professionals don't have to focus on system IT and management. No specialized support skills are required, yet their systems have access to optimal preventive maintenance and management tools. All these factors are helping advance genomics quickly—and when scientists spend less time on system development, maintenance, and monitoring, they can spend more time improving patient outcomes through lab research and innovation.

Jeff Durst is Director of Product Management and Solutions Architect, Dedicated Computing. Durst ensures that Dedicated Computing's product roadmap aligns with company vision and converts customer requirements into the architecture and design of the solutions. He taps into more than 30 years of technical and business leadership experience to guide Dedicated's product strategies in healthcare, communications, military, aerospace, industrial and scientific computing. Connect with Durst via LinkedIn or email at jeff.durst@dedicatedcomputing.com.

1. Source: "High-Performance Computing in Personalized Medicine" video
https://www.intel.sg/content/www/xa/en/healthcare-it/solutions/genomicscode.html
Intel, 2014. Accessed March 26, 2018.