Case Study
Intel® Xeon® Processor E5 and E7 Families
Intel® True Scale Fabric
High-Performance Computing
Big Data Analytics
Life Sciences/Biotechnology

Heterogeneous Cluster with Intel® Technologies Powers Biomedical Computing

Intel® Xeon® processors and Intel® True Scale Fabric help give Virginia Bioinformatics Institute (VBI) a threefold speedup for data-intensive workloads

VirginiaTech
Virginia Bioinformatics Institute

> "The amazing thing is that we see almost a three times performance increase on 48 nodes compared to 56 nodes of the previous generation, even though the processors are slightly slower clock speed. The Intel® QuickPath Interconnect and Intel® True Scale Fabric have had a big impact."
>
> *– Dr. Kevin Shinpaugh, Director of IT and HPC, Virginia Bioinformatics Institute*

Big data is transforming biomedicine, and VBI is on the forefront. A world leader in bioinformatics and medical informatics, the institute has expanded its Shadowfax* high-performance computing (HPC) cluster with server, storage, interconnect, and network technologies from Intel. The additional capacity provides performance and throughput for bioscience computing, enabling progress in areas ranging from population health to personalized cancer treatments.

Challenges

- **Scalable infrastructure for rapid data growth.** Output from gene sequencing machines and other biomedical data sources is increasing more than tenfold annually, vastly raising the demands on server, storage, and network infrastructure.
- **Varied applications.** Researchers' applications at VBI span a range of requirements, often within a single workflow.

Solutions

- **High performance.** Dell PowerEdge* R910 servers with the Intel® Xeon® processor E7-4860 and Dell PowerEdge C6220 servers with the Intel Xeon processor E5-2690, both running SUSE Linux Enterprise*.
- **Low latency.** Intel® True Scale 12000 Switch Series with Intel® True Scale Host Channel Adapters based on quad data rate (QDR) InfiniBand*.
- **Intelligent storage.** SGI InfiniteStorage* IS16000 platform powered by the Intel Xeon processor E5 family.

Technology Results

- **Application speedups.** VBI technologists say they're seeing 300 percent speedups moving from the previous-generation processors, which they used with another vendor's InfiniBand product, to the Intel processor and network technologies—despite running on 16 percent fewer nodes.
- **Rapid time-to-value for HPC investments.** Choosing widely used, well-supported technologies, VBI created flexible systems and brought them online within days.

Business Value

- **Lower costs, added capacity.** With affordable Intel True Scale Fabric technology, VBI saved enough to purchase additional computing nodes, giving VBI scientists even more capacity to apply to their research problems.
- **Faster progress.** The quick deployment, expanded capacity, and high performance are helping VBI accelerate progress on some of biomedicine's toughest research challenges.

Flexibility and Performance for Data-Intensive Applications

A world-class research institute based at Virginia Polytechnic Institute and State University (Virginia Tech), VBI has over 250 faculty and staff members dedicated to the biological sciences. VBI researchers develop advanced informatics algorithms and apply their expertise to key challenges in the biomedical, environmental, and agricultural sciences. Their innovative work with rapidly growing databases may one day influence the discovery of new vaccines and medical treatments, help population scientists manage or forestall a global pandemic, and even improve information security.

HPC is at the heart of VBI's mission, and many of the institute's research programs pool their funding to maximize the value of their collective investments. Choosing infrastructure for them requires meeting a wide range of user and application requirements.

The cluster performs millions of calculations on multi-petabyte databases

"We have a diverse set of users, and biology has a very diverse set of problems and scopes," explains Dr. Kevin Shinpaugh, director of information technology and high-performance computing at VBI. Workloads can involve both structured and unstructured data, with databases that can easily reach multiple petabytes. Some workloads demand massive throughput. Others require high memory bandwidth and large memory capacity so that in-memory database applications can natively run workloads in memory. A single genome can require a terabyte of memory to assemble and a terabyte to store.

Heterogeneous Cluster with Intel Technologies

To support such diverse workloads, VBI established a heterogeneous cluster powered by Intel technologies. "This is not just a normal homogeneous cluster," Shinpaugh says. "We've tried to make the machine very flexible."

The VBI cluster includes symmetric multiprocessing (SMP) nodes that benefit from the high core counts and large memory capacity of the Intel Xeon processor E7 family, as well as massively parallel processing (MPP) nodes powered by the Intel Xeon processor E5 family. VBI configures the SMP nodes with 1 TB of RAM, and the MPP nodes with 64 GB.

VBI balances the high-performing processors with the Intel True Scale Fabric host adapters and Intel True Scale QDR InfiniBand switches. To store the massive data volumes, VBI has 50 PB of tape storage capacity for raw data, and close to 600 TB of usable spinning disk capacity on the cluster. The SGI InfiniteStorage IS16000 platform uses the Intel Xeon processor E5 family to provide the intelligence for massively parallel, high-performance storage.

"The Intel Xeon processors give us excellent performance and significant increases in internal bandwidth," says Shinpaugh. "We deployed them with the Intel True Scale InfiniBand QDR Fabric and use that technology for some of our storage systems as well. It is pretty much a complete Intel solution."

The solution has dramatically increased performance for VBI's data-intensive workloads. Shinpaugh's team observed that 48 MPP nodes based on the Intel Xeon processor E5 family and Intel True Scale Fabric provide nearly triple the performance of 56 nodes based on previous-generation Intel Xeon processors and used with another vendor's InfiniBand product. In addition, using the IS16000 platform and the Intel True Scale Fabric, VBI moves data through the storage systems at 2 GB per second.

Production-Ready Technologies

As one of the lead users of the cluster, Dr. Harold "Skip" Garner praises another benefit of the Intel technologies: rapid deployment. Garner, a professor of biological sciences, computer science, and medicine at VBI, is an expert in analyzing repetitive DNA—what scientists used to call "junk DNA" until they began to realize its potential value.

"Our choice of technologies is driven by price and performance, of course, but you also have to think about the probability that the system is actually going to work when you put everything together," Garner says. "We've seen systems that get installed and still aren't in production nine months later. Shadowfax was a production machine from day one."

For the next addition to Shadowfax, VBI is eyeing the Intel® Xeon Phi™ coprocessor. "If we're doing 2 million tiny calculations to rebuild a genome, that's something that could really benefit from the coprocessor," says Garner.

Lessons Learned

Standardize on a flexible architecture and grow the system incrementally. "If you spend all your time designing the ultimate machine and making one purchase, then the entire machine is old in three years," Garner says. "It's better to get something working and in production as fast as you can, then grow the machine with newer technologies as they become available—not just computational power, but storage and networking as well. That way, you start getting value from your investment quickly, and you grow things in synchrony with the kinds of jobs you're doing as opposed to trying to put it all into the initial system."

"With the Intel coprocessor, we don't have to do a lot of new programming to take advantage of it."

Meanwhile, Garner looks for continued data growth and says biomedical computing is just warming up. "There is a fundamental shift underway, with more and more biomedical calculations—and we're on the verge of personalized medicine, where everybody gets their own genome done," he says. "That's when the real data is going to start coming. We haven't even started yet."

Find the solution that's right for your organization. Contact your Intel representative, visit Intel's **Business Success Stories for IT Managers,** or explore the **Intel.com IT Center.**

Printed in USA 0913/LJ/TDA/XX/PDF

327923-001US