



Biogemma accelerates genomic research with Lenovo

Lenovo™



Overview

Biotech firm Biogemma needed to ramp up its computational power to enable faster analysis of larger and more complex genomic data sets. By deploying an HPC cluster based on Lenovo NeXtScale nx360 M5, and a large-memory system based on a Lenovo System x3850 X6 server, the firm significantly increased performance—cutting run times from hours to minutes; typically over a 90 percent improvement—and eliminated the need to queue for computational resources.

Founded in 1997 by seed companies and field crop producers, Biogemma is a plant biotechnology company. The company's mission is to discover or acquire genes and technologies which control the expression of commercially valuable crop traits such as drought tolerance, nitrogen-use efficiency, disease resistance and high yield.

Olivier Dugas, CEO at Biogemma, explains the company's need for high-performance computing (HPC) resources: "Our bioinformatics research generates and processes huge volumes of genetic data, so our computational requirements are constantly rising. The genomes we analyze are complex and often polyploid, which can make them significantly larger than the human genome. Alongside that, we are doing more work around the characterization of genetic material, in which we compare a reference genome with a huge number of sequencing data points—which requires parallel computations across multiple cores simultaneously."

A growing challenge

Understanding how crop genomes are expressed in physical traits such as resistance to disease is of enormous financial and societal value. Biogemma's work will help ensure hardier, higher-yield crops to feed a growing global population.

"We have seen a very clear increase in the speed of computation with our Lenovo NeXtScale System cluster."

—**Nathalie Rivière**,
Upstream Genomics Coordinator /
Bioinformatics Manager,
Biogemma



To determine which genetic sequences result in which desirable traits, the company must analyze and cross-reference huge volumes of sequence data from complex genomes.

“In de novo genome assembly, we may be comparing a million sequences with another million sequences, which is very computationally complex and demands a system with large memory resources,” comments Nathalie Rivière, Upstream Genomics Coordinator / Bioinformatics Manager at Biogemma. “The characterization of genetic material is a different class of problem. Here, we need to parallelize statistical calculations to get results faster, which demands a computational cluster.”

Since deploying its previous 200-core cluster, Biogemma had seen bioinformatics workloads increase dramatically: the number of samples in a typical analysis had grown from ten to 400. On the statistical side, the previous cluster could not cope with growth in terms of both the number of analyses and the complexity of the models.

Sowing the seeds of change

After an internal requirements-gathering process, Biogemma contacted several major HPC vendors for architectural recommendations. “We didn’t provide a precise specification,” recalls Nathalie Rivière. “Rather, we explained our requirements for both large-memory and parallelized computations, and we supplied a set of representative use-cases for benchmarking purposes.”

She adds, “We selected the solution proposed by Lenovo and its partner, Serviware, based partly on its price-performance ratio and partly on the expertise both parties showed during the benchmarking process, which helped us to optimize the solution and tune our applications. Their expertise also helped us significantly in improving the scheduling and prioritization of compute jobs.”

For bioinformatics workloads requiring large amounts of memory per core, Biogemma deployed a Lenovo System x3850 X6 server featuring Intel® Xeon® E7 Series processors with 48 cores and 1.5 TB of memory. For the highly parallelized statistical workloads, Biogemma deployed a 36-node Lenovo NeXtScale nx360 M5 cluster, each node having two 12-core Intel Xeon E5 processors, for a total of 864 cores and 34.5 Tflops of performance. High-speed InfiniBand interconnects link the nodes and connect them to an IBM Storwize V3700 storage server with 120 TB capacity. The environment uses IBM Spectrum Scale (IBM GPFS) for high-performance parallel file access.

Solution components

Hardware

Lenovo NeXtScale nx360 M5

Lenovo System x3850 X6

IBM Storwize V3700

Software

IBM Spectrum Scale (IBM GPFS)



Reaping the rewards

The Lenovo cluster and large-memory environment have delivered significant performance gains at Biogemma, helping the company meet fast-growing internal demands around the number and complexity of computational analyses. “We have seen a very clear increase in the speed of computation with our Lenovo NeXtScale System cluster, which provides many more cores running at higher clock speeds,” says Nathalie Rivière. “A job that used to take a day to run now completes in a matter of hours or even minutes – typically over a 90 percent improvement”

Increased performance also improves the overall throughput of research at Biogemma, as Nathalie Rivière explains: “In the past, when resources were limited, we had to plan our computation sessions, creating a queue of projects. Today, we can go ahead with analysis as soon as the data is available, working more efficiently and getting the results faster. We can also handle many more simultaneous analyses, and can tackle much more ambitious studies. Our Lenovo HPC solution therefore helps Biogemma to keep its place at the forefront of the global biotech industry.”

For more information

To learn more about Lenovo Enterprise Systems contact your Lenovo Sales Representative or Lenovo Business Partner, or visit: lenovo.com/systems

To learn more about Biogemma, visit: www.biogemma.com

“Today, we can go ahead with analysis as soon as data is available, working more efficiently and getting the results faster.”

— **Nathalie Rivière**,
Upstream Genomics Coordinator /
Bioinformatics Manager,
Biogemma



© 2016 Lenovo. All rights reserved.

Availability: Offers, prices, specifications and availability may change without notice. Lenovo is not responsible for photographic or typographic errors. **Warranty:** For a copy of applicable warranties, write to: Warranty Information, 500 Park Offices Drive, RTP, NC, 27709, Attn: Dept. ZPYA/B600. Lenovo makes no representation or warranty regarding third-party products or services. **Trademarks:** Lenovo, the Lenovo logo, NeXtScale, and System x are trademarks or registered trademarks of Lenovo. Intel, the Intel logo, Xeon and Xeon Inside are registered trademarks of Intel Corporation in the U.S. and other countries. Other company, product, and service names may be trademarks or service marks of others.

Visit <http://www.lenovo.com/lenovo/us/en/safecomp.html> periodically for the latest information on safe and effective computing.

