MALAYSIA GENOME INSTITUTE

SUCCESS STORY

Malaysia Genome Institute Advances Bioinformatics Research with SGI® Technology

"SGI delivered flexible computing with hybrid architecture that integrates high-throughput, high-performance and FPGA-based solutions for all of our bioinformatics applications — a solution that seamlessly runs applications in the environment best suited for optimal performance, allowing scientists to focus on science. This enables us to broaden the scope of our research into areas we could never have considered before.

– Mohd Noor Mat Isa, Laboratory Manager, Malaysia Genome Institute

The Malaysia Genome Institute (MGI) is a network-based organization, headquartered in the state of Selangor, dedicated to discovery research on tropical bio-resources through projects in genome sequencing, comparable and functional genomics, and structural biology. Officially launched by the Prime Minister of Malaysia in 2005, the Institute is a full realization of the vision of the biotechnology cooperative centers of the National Biotechnology Directorate and the Ministry of Science, Technology and Innovation to create state-of-the-art facilities for genomics and molecular biology research. To further that vision, MGI selected SGI to create an optimal workflow that encompasses the diverse computing requirements of bioinformatics research. In April 2007, SGI designed and installed a unique SGI® InfiniteStorage NAS/SAN environment.

Because bioinformatics utilizes a variety of software at various stages in the workflow, neither HPC nor clusters provided an environment for the efficient use of each application. MGI looked to SGI to design a seamless, cohesive workflow for the entire process that would integrate into existing infrastructure. SGI delivered the SGI BioCluster solution with SGI InfiniteStorage CXFS shared filesystem, which makes the entire workflow process transparent to the user regardless of which software is employed or what stage of the process scientists are performing.

"Bioinformatics brings together an avalanche of systematic biological data with the analytic theory and practical tools of mathematics and computer science," says Mohd Noor Mat Isa, Laboratory Manager, Malaysia Genome Institute. "SGI delivered flexible computing with hybrid architecture that integrates high-

"Technologies acquired from structural biology-based projects will be applied to structural genomics studies on the data sets generated by the other genomics-based projects," says Dr. Noor. "The application will realize one of the ultimate objectives of the genomics projects,i.e. to understand how biomolecules act and work together in living organisms."

– Mohd Noor Mat Isa, Laboratory Manager, Malaysia Genome Institute

throughput, high-performance and FPGA-based solutions for all of our bioinformatics applications — a solution that seamlessly runs applications in the environment best suited for optimal performance, allowing scientists to focus on science. This enables us to broaden the scope of our research into areas we could never have considered before."

At MGI, also known as GENOMalaysia, SGI Altix high performance systems and SGI Altix XE clusters, optimized as the SGI BioCluster solution, delivers a cohesive NAS/SAN workflow environment for all stages of data processing – solving the problem of storage capacity and processing capacity. The SGI system is used to support all genomics and proteomics projects conducted at the Institute and as a shared resource available to all researchers throughout the country. MGI is also a partner to all public universities for postgraduate programs in molecular biology.

Bioinformatics Research Projects at MGI

Several major research projects are currently underway at MGI. The Microbial Genomics Research for Gene and Natural Product Discovery project is focusing on two model organisms: Burkholderia pseudomallei, a locally important soil pathogen, and Eimeria tenella, a local avian protozoan. The project harnesses the SGI systems to dissect the organisms' genomes to gain new insights on virulence, antibiotic resistance, host specificity and infectivity.

In Marine Biology, the Genomic Approaches to Seaweed Genes and Natural Product Discovery project is using the SGI BioCluster to run combinatorial chemistry, genomics and proteomics applications in order to understand the genetic basis of adaptation to different environmental conditions of two economically important species of red algae. Researchers are studying the algae's relationship with gene expression and secondary metabolite production.

"Technologies acquired from structural biology-based projects will be applied to structural genomics studies on the data sets generated by the other genomics-based projects," says Dr. Noor. "The application will realize one of the ultimate objectives of the genomics projects, i.e. to understand how biomolecules act and work together in living organisms."

"Although the above projects are largely aimed at biotechnology development, they are actually fundamental research on the molecules and organisms concerned," adds Dr. Noor. "The projects are seeking answers to fundamental questions in developmental biology, biological diversity and adaptation, molecular-molecular interactions, control of gene expression, cellular networks, mechanisms of host-parasite interactions, and so forth. With the guiding principle of 'molecular biology is the lifeline of modern biotechnology,' there is no longer a separation between fundamental and applied research."

International Bioinformatics Projects

MGI collaborates with other research facilities and universities throughout the world, including the Wellcome Trust Sanger Institute in Cambridge, U.K. on comparative genomics, genome sequencing and analysis. MGI also has a joint project on host-pathogen interaction study with the Department of Genetics at Stanford University, adjacent to Palo Alto, Calif.

MGI, working with the Antarctica Research Group, University of Tasmania, will harness the high performance attributes of the SGI InfiniteStorage NAS/SAN for Whole Genome Shotgun (WGS) sequencing of an Antarctic fungus and an alkaline polar bacterium. Polar aquatic conditions are thought to be similar to the early history of the planet Mars; current Antarctic life forms would probably look and behave in similar ways to their early Martian counterparts.

Malaysia Genome Institute

Genomic sequencing operations, such as listed above, typically utilize specific applications each with specific compute requirements necessitating multiple systems. The SGI BioCluster solution includes a software layer that provides transparent workload scheduling through which users can submit jobs without thinking about hardware. The SGI solution's administrative software intelligently cues submissions to run on the appropriate platform, providing best-in-class performance and optimal load balancing. Integrated cluster management tools and dynamic software controls can change the total system from a single meta-cluster to semi-autonomous sub-clusters as needs dictate. For time-critical computing challenges, the SGI BioCluster system enables administrators to bring all available resources to bear on a single job.

In Conclusion

Malaysia Genome Institute needed a single, cohesive computing platform that leveraged the best of two distinct computing environments, HPC and high-throughput cluster computing. SGI worked with channel partner Quantum Beez Sdn Bhd (formerly known as Open Source Systems) to design and develop the SGI BioCluster solution for MGI, which includes an SGI Altix 4700 system with 144GB RAM, and 48 Dual Core Intel Itanium 2 processors running Novel SUSE 9 Linux Environment, an SGI Altix XE 1200 Cluster consisting of one SGI Altix XE 210 head node and 42 SGI Altix 210 compute nodes, each with two Dual Core Intel Xeon processors. The NAS/SAN is backed by an SGI InfiniteStorage 350 system with 21 500GB SATA drives.

SGI produced one solution with two architectures that delivers maximum return on investment for heterogeneous or multi-vendor based workflows.

"The projects are seeking answers to fundamental questions in developmental biology, biological diversity and adaptation, molecular-molecular interactions, control of gene expression, cellular networks, mechanisms of host-parasite interactions, and so forth. With the guiding principle of 'molecular biology is the lifeline of modern biotechnology,' there is no longer a separation between fundamental and applied research."

– Mohd Noor Mat Isa, Laboratory Manager, Malaysia Genome Institute

Corporate Office
1140 E. Arques Avenue
Sunnyvale, CA 94085
(650) 960-1980
www.sgi.com

North America +1 800.800.7441
Latin America +55 11.5185.2860
Europe +44 118.912.7500
Japan +81 3.5488.1811
Asia Pacific +61 2.9448.1463

4107 [5.2008] J15376