

# MALAYSIA GENOME INSTITUTE

## Discovering gains in national genomic projects



### ESSENTIALS

#### Industry

Science and technology

#### Company Size

42 employees

#### Business Challenges

- Massive amounts of data generated through NGS analysis
- Demand to archive all “omics” data
- Data management was impacting the efficiency of project delivery

#### Results

- Reduced data processing times by 30 percent
- Provided researchers and partners with direct access to data for fast analysis
- Balanced costs with continual growth of the storage capacity

### DEVELOPING BIOTECHNOLOGY INNOVATION

After its establishment in 2005, Malaysia Genome Institute (MGI) has grown to become a leading network-based organization carrying out discovery research on tropical bioresources. The institute supports national genomic projects and collaborates with research organizations and universities throughout Malaysia.

Highly scalable storage is a core requirement for next-generation sequencing (NGS) analysis. MGI uses the Illumina HiSeq 2000 and Illumina MiSeq sequencing platforms for DNA sequencing, whole genome sequencing, whole transcriptome sequencing, and targeted resequencing. The platforms generate massive amounts of data, with a single run generating 13 terabytes of data at an average of 1.5 terabytes per day. Internal storage requirements associated with such processes mean that MGI will double its current capacity within the next three to five years.

Research institutes like MGI are faced with meeting the phenomenal growth in genome sequencing and matching this growth with an efficient means of storing, transmitting, and analyzing the data. MGI needed a storage system that could capture data from the run and meet the demands of analyzing the information quickly to deliver results to MGI’s partners and collaborators involved in the next stage of the process.

As the data store capacity increased, the complexity of storage management was also increasing, which slowed data processing. Previously, researchers had to retain their data on storage directly attached to the HPC cluster. Users required access to the HPC cluster and data, which had to be pulled manually to individual user home directories for analysis.

In addressing these challenges, MGI partnered with Integriti Padu to deploy a single, scalable solution featuring Dell EMC® Isilon® scale-out storage within its next-generation sequencing architecture.

With centralized control over storage that can be scaled on demand, and with data access provided directly to users, MGI can ensure that there are no bottlenecks within its workflows. Faster analysis means problems in the genome sequencing are identified earlier and data can be delivered more quickly to its partners, improving the overall efficiency of each project.

### ENVIRONMENT

MGI deployed a single cluster comprising three Dell EMC Isilon S200 nodes providing a high-performance tier. Dell EMC Isilon SmartPools™ automatically tiers data to six Dell EMC Isilon NL400 nodes for its capacity tier, after an allocated time period. The solution connects to the institute’s existing Dell EMC Isilon cluster. Dell EMC Isilon SmartQuotas™ controls and manages user storage allocation, while Dell EMC Isilon SmartConnect™ enables load balancing across the storage nodes to provide MGI with optimal utilization of the cluster resources.



## Solutions

- Dell EMC Isilon S200
- Dell EMC Isilon NL400
- Dell EMC Isilon SmartPools
- Dell EMC Isilon SmartQuotas
- Dell EMC Isilon SmartConnect
- Dell EMC Isilon InsightIQ

Dell EMC Isilon InsightIQ® provides the organization with performance monitoring and reporting tools, so MGI can maximize the performance of the storage solution. Illumina sequencers are also connected directly to the Dell EMC Isilon cluster, enabling real-time transfer of sequencing data to the storage.

## DELIVERING PROJECT EFFICIENCIES

MGI has improved the efficiency of its workflow that delivers data to users for analysis by eliminating manual processes and providing direct access to data for its researchers. Making its research teams more independent means reducing IT administration overhead and enabling a greater focus on project efficiency. As a result, the institute can work strategically to speed analysis and enhance the downstream processes provided by its partners. "Moving to network-attached storage was critical to reducing our IT administration and improving efficiency. Users can access data directly on the Dell EMC storage from any workstation instead of having data pulled to individual directories," says Mohd. Noor Mat Isa, Head of Genome Technology and Innovation at Malaysia Genome Institute.

In moving from storage within its HPC cluster to single file system storage, MGI has reduced the dependence on its small IT team. With Dell EMC Isilon OneFS®, MGI can now respond in real time to demands for performance and capacity. "With a small team working across multiple areas of the institute, the web interface makes it easy to manage the system without requiring specialist IT skills," says Mohd. Noor.

**"We have made gains of close to 30 percent in our processing speed after deploying the Dell EMC Isilon storage."**

Mohd. Noor Mat Isa, Head of Genome Technology and Innovation, Malaysia Genome Institute

## MEETING THE DEMANDS OF BIG DATA

MGI is meeting the demands of the I/Os per second (IOPS)-intensive processes involved in genome sequencing. With the sequencing machines delivering 1 gigabyte per second of throughput, the institute wanted a solution that could balance IOPS and performance. MGI can grow a single file system while delivering the data throughput necessary for genome sequencing. "The way we analyze Big Data can require millions of inputs at the same time. This involves transferring data back and forth between the storage and the HPC cluster. Dell EMC can comfortably handle the high throughput required within the analysis," says Mohd. Noor.

## ENABLES 30 PERCENT FASTER DATA PROCESSING

MGI has realized an almost 30 percent reduction in data processing times as a result of moving to the Dell EMC storage platform. "In working with Big Data, the time to process data decreases once the IOPS increases. We have made gains of close to 30 percent in our processing speed after deploying the Dell EMC Isilon storage," says Mohd. Noor.

MGI works with a number of collaborators to complete genome projects. The ability to share data through access to the Dell EMC storage supports knowledge generation. This faster processing by the institute impacts the sequencing workflow by providing results more quickly to its partners involved in subsequent processes. Mohd. Noor says, "The entire workflow is connected. Quick access to data means we can complete the analysis faster, and this means we can also identify any errors or problems early. This ultimately improves the quality of analysis the institute can deliver within the genome sequencing."

## Company Overview

The Malaysia Genome Institute engages in national and international collaborative projects in comparative genomics and genetics, structural and synthetic biology, computational and systems biology, and metabolic engineering.

Launched in 2005, the institute is designed to partner with public universities and other research institutes throughout the country in the areas of biotechnology and molecular biology.

In 2014, the institute transformed to a not-for-profit company, and become part of the consortium of the National Institutes of Biotechnology Malaysia.

## BALANCING COSTS WITH SCALABILITY

As a government research institute, an ongoing challenge for Mohd. Noor and his team is expanding storage capacity. Internal policy requires that data be archived indefinitely, which means storage demands will continue to grow. "We store data for several years as our users might extract the data again when an analysis comes out. We need data to remain highly available, and that's a financial burden on the institute," says Mohd. Noor. MGI is addressing the issue of cost-efficient archival through its single volume storage solution based on preconfigured nodes that can be added as required. The storage platform now provides 500 terabytes of capacity, and MGI has plans to grow this to 1 petabyte within the next three to five years. Mohd. Noor says, "Our storage capacity scales based on demand, and we simply add nodes as our user numbers increase based on our projects. Our focus remains on developing our sequencing services to drive Malaysia to become a leader in biotechnology."

## CONTACT US

To learn more, contact your local representative or authorized reseller.



Copyright © 2017 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners. Published in the USA. 1/17, Customer Profile, H14351

Dell EMC believes the information in this document is accurate as of its publication date. The information is subject to change without notice.