National University of Singapore

Taming the Big Data Avalanche from Instrumentation

Key Facts

Organization:
NUS: CBIS and MBI

Location:
Singapore

Application:
Life Sciences

While exponential growth in data is not an unusual story these days, Alan Davis, from the National University of Singapore (NUS) is striving to keep ahead of the avalanche of data produced by instruments at the Centre for Bio Imaging Sciences (CBIS) and the Mechanical Biology Institute (MBI). Mr. Davis is the Senior Facility Manager for Information Technology working across these two cutting edge reseach institutions. Davis is not new to this area of research technology, having worked previously with very large memory systems in the Massachusetts Institute of Technology (MIT) Bio Imaging Laboratory.

The Problem

Mr. Davis outlines the problem simply. “A few years ago, common digital cameras had 1000x1000 resolutions with a maximum acquisition rate of 100 MB/s. Now cameras are approaching 10,000x10,000 resolutions with acquisition rates exceeding 1 GB/s. Individual data sets currently range from a few GB to TB in size and up to 200,000 files, with future data sets to approach 10 TB and over 1 million files each.”

Together, CBIS and MBI, founded in 2010 and 2009 respectively, have a primary focus on basic research into the mechanisms and functions of biological processes at the molecular, cellular and tissue levels. Davis says, “We aim to identify, measure and quantitatively describe how physical and mechanical forces are detected, expressed and responded to by biological organisms.”

The infrastructure at the institutions supports over 300 researchers with strong collaboration of many faculties at NUS to local institutions such as ASTAR, Nanyang Technical University, and includes institutions in Japan, China, India and the USA.

Computational Intensive Work

At CBIS, Biological imaging is computationally intensive work requiring the ability to accumulate, store, process, analyze, and visualize terabyte-level data sets. The solution CBIS sought was a high-bandwidth/ high-capacity Bio Imaging Network especially suited to the rigorous demands of image processing and analysis. This network consists of acquisition instruments and high-density computer processors connected by fiber channel to high-performance data storage. CBIS is able to move data at speeds of several gigabytes per second for processing, analysis, and visualization locally or via remote access.

> “CBIS and MBI aim to identify, measure and quantitatively describe how physical and mechanical forces are detected, expressed and responded to by biological organisms. SGI provides the tools that enable us to answer our critical biological research questions in a timely and efficient manner.”

Alan Davis

Senior Facility Manager for IT,
National University of Singapore

More specifically, CBIS and MBI were supplied more than 500TB of SGI® InfiniteStorage™ high performance storage area network (SAN). Utilizing industry leading SGI® CXFS™ file systems to provide the necessary performance to handle the growth rates of 5-10TB per month currently observed. MBI has recently added 150TB of SGI® InfiniteStorage™ 5500 to deal with the ever growing demand.

"Our data sets are so large that Windows-based servers cannot adequately handle the size or provide high-speed access to the files," said Davis. "SGI's storage servers have historically provided us large capacity, high bandwidth transfer rates and 64-bit file systems providing a complete solution for storing very large data sets at an affordable cost."

Big Data Instruments

Instrumentation at CBIS and MBI generates massive amounts of data that must be cost effectively stored. To enable this, SGI provided the SGI® InfiniteStorage™ Data Migration Facility to entrust the management and storage of data that is accessed infrequently but also must meet strenuous delivery response to drive results when researchers require.

"We wouldn't be able to do what we do without SGI systems," says Prof Matsudera, Professor and Head, Department of Biological Sciences.

The Process

Bio Imaging is much more than simply taking interesting pictures. In fact, image acquisition is but one stage (and not the first) in a logical, linear process. NUS defines the art and science of bio molecular imaging as a four-stage strategy:

1. Manipulate: in this stage the biological question is asked, prompting experiment in which molecules, cells, or tissues are modified in some defined manner
2. Measure: this is the imaging stage, which may involve a range of microscopic techniques
3. Mine: the data is processed, quantified, and analyzed to produce quantitative results
4. Model: the results are represented through computational, mechanical, chemical, and other models

Challenges

Storage, archive and retrieval of data is not the only challenge the infrastructure design must meet. The strategy includes a number of computationally intensive processes. One common computational method of processing involves deconvolution, which increases the resolution of light microscopy data and removes the noise and blur that often accompany high rates of data acquisition.

Once an image has been acquired, it must be analyzed to identify features of interest, calculate 3D structures, track objects with time, and correlate objects in time and space. After image data is processed and analyzed, the next challenge involves developing algorithms to model the data and assemble the knowledge gained from the imaging initiative. The quality of the original data determines the accuracy and reproducibility of the resultant model, and a mistake or oversight at any stage will corrupt the ultimate findings.

To answer these needs, NUS asked SGI to supply large memory (more than 1TB each) SGI® UV100™ and SGI® UV2000™ shared memory compute platforms. Utilizing more than 224 cores of Intel processor technology combined with the power of Intel® Xeon Phi™ Coprocessors (adding 5 TFLOPS), CBIS and MBI are able to provide the speed and reliability needed by their large research community.

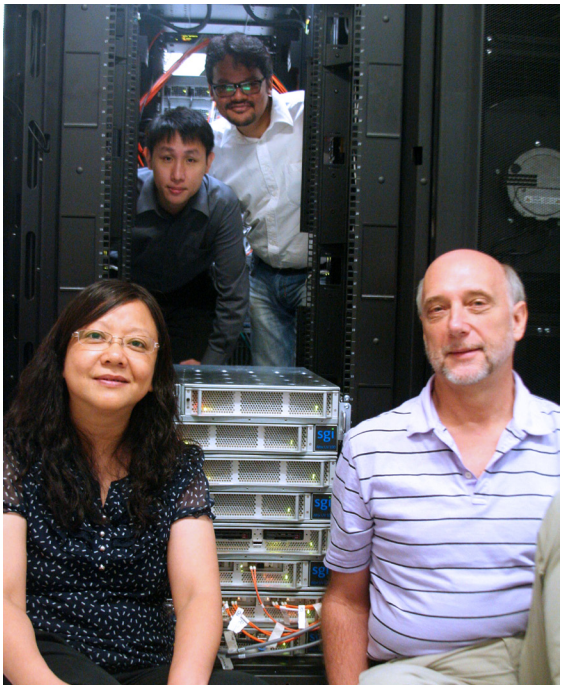
sgi

The Future

"The main area of improvement needed for CBIS and MBI is in the area of a data management solution. Many of our projects involve the collaboration of multiple researchers performing many experiments, using different techniques and instrumentation, and the collection and analysis of data sets over long periods of time. We need a better mechanism for cataloging and sharing these data sets," said Davis. "We are looking at LiveArc, from SGI, as a solution to this need."

"SGI provides the tools that enable us to answer our critical biological research questions in a timely and efficient manner," said Davis. "Without this computational infrastructure it would have been difficult to collect and analyze the data from our research experiments, leaving us unable to answer many of our scientific questions and greatly limiting our rate of progress."

National University of Singapore: www.nus.edu.sg



Alan Davis and Team

Global Sales and Support: sgi.com/global