SGI® Accelerates ICR's Drug Development Pipeline

Key Facts

Organization:
The Institute of Cancer Research

Locations:
Chelsea and Sutton, England

Application:
Process and store a massive volume of cancer research-related data

Founded in 1909 as a small research department at the Royal Marsden Hospital, the Institute of Cancer Research (ICR) is at the forefront of international cancer research and has since then contributed significantly to the considerable advances in reducing mortality for a wide range of cancers. Based in Chelsea and Sutton, the organisation continues to grow and now has approximately 1,100 internationally renowned staff. Requiring high-performance big memory supercomputers, superclusters, multi-tiered storage design and applications expertise to process its growing data requirements, the ICR has chosen SGI to provide a comprehensive server and storage solution.

Business Challenge

The ICR focuses on three main themes – molecular pathology, therapeutic development and genetic epidemiology. One of the research sections, the Integrative Network Biology initiative, forms part of the Cell and Molecular Biology division, and studies basic molecular and cell biology in order to identify new strategies for cancer therapeutics. These studies are then integrated with other investigations on other cellsignalling molecules in attempts to identify new targets for drug development. The increasing availability and advances in technology within this area, such as massspectrometry, high content microscopy and functional imaging has created the challenge of processing massive volumes of cancer-relevant data. The ICR has chosen a complete SGI® server and storage solution, enabling it to quickly process and correlate large, diverse sets of medical and biological data.

Technology Solution

The ICR currently benefits from SGI UV shared memory supercomputers, several SGI ICE clusters and a fully integrated, multi-tiered storage design. The SGI UV solution uniquely enables extremely large and heterogeneous data to be integrated swiftly, which allows researchers to correlate data from patient to cancer cells on an unprecedented scale. This research is leading to network-based cancer models that can be used in drug development for network medicine. In addition to massive parallelization of the compute solution in use, shared memory is also critical, as the assembly algorithms that build the final sequences are all memory limited. The SGI UV supports up to 16 terabytes of global shared memory in a single system image, thus ensuring that it remains highly efficient at scale for applications from in-memory databases to a diverse set of data- and compute-intensive HPC platforms. As a result of using the SGI UV, the ICR is now able to integrate extremely large data sets, providing significant support in the areas of deep sequencing, phenotyping and MRI imaging.
"The rapidly expanding data requirements within systems biology has made large, shared memory essential," said Professor Chris Marshall, Director of Research at The Institute of Cancer Research. "Images and sequence data results are increasingly converging, indicating a growing need for cross-industry collaborations. The SGI solution facilitates this by drawing on large memory capacity, enabling us to manage data at a higher speed and with greater focus."

The Institute also relies on two generations of the SGI ICE bladed cluster platform - SGI Altix® ICE 8200 and SGI ICE 8400 - and three generations of Intel® Xeon® processors for a range of purposes including high-throughput image analysis (e.g., image processing, recognition and feature extractions) and implementation of machine-learning algorithms on very high dimensional datasets.

Another key use of the SGI ICE is for processing raw data such as searching MS data against databases as well as for quantification. ICR is also using its SGI ICE for developing database structures for data integration.

As there are several systems stationed between both the Chelsea and Sutton sites in support of The Institute's extensive research, they were also in need of a comprehensive, mirrored storage solution from SGI.

The SGI CXFS™ shared file system and DMF tier virtualisation products are used at the ICR headquarters to ensure that storage capacity utilisation is optimised and all cancer-relevant data is easily accessible. Dedicated Parallel Data Movers move data at both of the ICR sites to and from tier 2 SGI (COPAN™) and tier 3 (LTO 5 tape), depending on how critical and timely the data is that is being processed. In the instance that a data mover at one of the sites fails, the central DMF server will use the remaining data mover, providing both reliability and security. All of this is a transparent process and occurs without any input from the ICR researchers

The SGI DMF management system also gives the ICR the option of scaling its existing infrastructure in the future, to keep pace with the continued data growth.

By continuously monitoring and migrating the most critical or timely data to higher performance storage media, and less critical data to less expensive, lower performance media, the ICR achieves considerable time and cost savings.

Some of the SGI solutions utilised by the ICR (pictured left to right): SGI UV, SGI ICE, SGI InfiniteStorage 5000 and SGI COPAN.

Business Results

The ICR's server and data requirements highlight the growing demand for scalability and power within the scientific research field. Aside from the need to manage and analyse hundreds of terabytes of data, the ICR needs to be able to perform rapid calculations across a broad range of research in a way that is cost effective and requires as little human input and intervention as possible.

SGI UV itself provides a flexible system architecture for accelerating productivity within an HPC environment. It offers advances in data efficiency, unmatched scalable processing, memory and I/O capabilities, as well as the ability to support system-wide message acceleration. This is managed within a framework that has offered a lower cost of ownership by utilizing the most advanced x86 processor technology, a Linux® environment that currently scales to 2,048 processor cores, 16TB of shared memory and scalable I/O capabilities for outstanding performance capabilities, while minimizing administration and operating at industry-leading rack level power efficiency.

Overall, the entire SGI solution enables the Institute to systematically process and correlate its growing data requirements, providing a system that is essential to the future of integrative biology and the subsequent development of human disease therapies.

For additional information, please contact Simon Appleby at sappleby@sgi.com

For press enquiries, please contact Jason Coari at jcoari@sgi.com

Solutions

- SGI UV Shared Memory Systems
- SGI ICE Bladed Cluster Systems (three generations of Intel® processors)
- SGI LiveArc™ Servers
- SGI DMF Software
- SGI CXFS™ Shared File System
- SGI InfiniteStorage 220
- SGI InfiniteStorage 4100
- SGI InfiniteStorage 4600
- SGI InfiniteStorage 5000
- SGI XFS® File System
- SGI COPAN™ Data Storage Systems
- SGI Rackable™ "Web Service" Servers (AMD®/Intel®)
- SGI Multi-Site Tiered Storage Design

About SGI www.sgi.com

SGI, the trusted leader in technical computing, is focused on helping customers solve their most demanding business and technology challenges. Visit www.sgi.com for more information.

Corporate Headquarters
46600 Landing Parkway
Fremont, CA 94538
tel 510.933.8300
fax 408.321.0293

Global Sales and Support
North America +1 800.800.7441
Latin America +55 11.5185.2860
Europe +44 118.927.8000
Asia Pacific +61 2.9448.1463