



The National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at Guy's and St Thomas' NHS Foundation Trust and King's College London, is one of only five NIHR comprehensive BRC's in the UK. Since being formed in 2006 it has made significant progress in translational research across a number of important disease areas. It aims to develop new diagnostic tests and medical treatments that will impact management and treatment of diseases.



High performance data processing, management and storage

KCL Uses New High Performance Server Cluster and Storage System to Understand Links Between Genetics and Disease

Business Impact Summary

Challenge:

The Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London has a genomics facility equipped for genomic sequencing, genotyping and gene expression studies. The centre needed significant processing power to analyse the genomics data it generated and an adequate storage, back-up and archiving system for the processed data.

Solution:

OCF used technology from its key partners, including IBM, Panasas and BLADE Network Technologies, to design a bespoke server cluster and storage system for KCL. Effective SLA driven support and maintenance from OCF ensures the system remains fully operational at all times.

Result:

The server cluster's IBM iDataPlex servers can reduce the time necessary to analyse genome sequencing data by 20 fold or more, reducing the time scales for analysis from days to hours. The storage system enables administrators to take control and deliver effective storage, back-up and archiving of data.

"The genomics facility houses three 'Sequencers' which collectively generate up to 50 billion base pairs of useable DNA sequence data every 10 days."

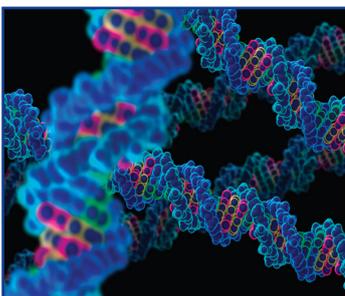
Dr. Rebecca Oakey, Reader in Epigenetics, KCL

Demanding Computing Requirements

In December 2006, Guy's and St Thomas' NHS Foundation Trust, along with its academic partner King's College London, was awarded the right to become one of five new National Institutes for Health Research (NIHR) comprehensive Biomedical Research Centres. The Centres have a strong focus on 'translational research', which take advances in medical research out of the lab and into the clinical setting.

"The sequence of the human genome has been known for ten years now so we are using new sequencing technologies to sequence specific regions of the genome in large numbers of people in order to help understand the contributory factors to a variety of common complex disorders and developmental defects such as bone diseases and cancer," says Dr. Rebecca Oakey, Reader in Epigenetics, Department of Medical & Molecular Genetics, School of Medicine, King's College London.

The new BRC Centre at Guy's Hospital in London had a genomics facility equipped with instruments for genomic sequencing, genotyping and gene expression studies. At that time, the facility housed three Illumina Genome Analyser Iix (GA Iix) Sequencers, which collectively generated up to 50 billion base pairs (A, T, G, C's) of useable DNA sequence data (the equivalent of 17 human genomes) every 10 days. This is around 400GB of useable data (the equivalent of 585 CDs).



Challenge

The genomics facility was using an Illumina Pipeline Analysis Server to provide the processing power to analyse the genomics data generated by the three Illumina sequencing machines. The facility also made use of multiple local PCs and servers spread across different research groups. As demands for server power continued to increase across all research groups, the current disparate, siloed architecture meant researchers were constrained by the processing power available to them. In addition, administration of storage, back up and archiving of analysed data was logistically very difficult. At times, identifying what data was stored and then backed up, how, how often, where, i.e. memory stick or PC, or by whom was a challenge.

Solution

In April 2010, medical researchers working across all research groups at the facility took ownership of a central, shared server cluster and storage system to replace the Illumina Pipeline Analysis Server and the multiple servers and local PCs. The system's bespoke design, rapid implementation and configuration is handled by data processing, data management and data storage provider, OCF.

HPC server cluster

The server cluster and storage system's design includes technology from a number of OCF's preferred partners:

- IBM's iDataPlex server hardware, which includes ultra-low latency, 10Gb Ethernet switching modules from BLADE Network Technologies
- Panasas 'plug and play' ActiveStor Series 8 clustered storage using its built-in Panasas ActiveScale distributed parallel file system. Current limits enable storage of up to 180 TB of raw data
- An IBM TS3310 Tape Library Unit with Tivoli Storage Manager to enable long-term, secure, off-site data back-up

Support and maintenance

OCF also provides ongoing remote and onsite SLA driven support, plus user training. Don Lokuadassuriyage, Biomedical Research Centre System Administrator, adds: "OCF is always willing to help. We did have a small server error, but OCF had replacements shipped, installed and operational in a matter of days."

Benefits

As the BRC cluster and storage began to prove itself amongst the research teams, usage of it by different groups began to grow. There are now around 10-15 research groups using the sequencing hardware and the cluster. Some research groups are just starting into next generation sequencing and some are more experienced. In 2011, the BRC increased its count of sequencing machines to five. The two, new additional sequencers (Illumina HiSeq 2000) creates five times more data (longer read length with higher read quality) than the existing GAllx sequencers, around 1Tb of data per sequencing run, every 10 days. This means research projects are being completed with greater accuracy, and also creates a need for a larger on site storage capacity. The server and storage cluster enables researchers to more quickly analyse data generated during their quest to understand the role of genetics in a range of common health issues. The system can reduce the time necessary to analyse sequencing data by 20 fold or more. What used to take days to analyse on the Illumina Pipeline Analysis Server, now takes just hours.

Julian Fielden, managing director, OCF, says: "The comprehensive Biomedical Research Centre is a great example of an organisation that acknowledges data on its own delivers little or no value; organisations must analyse and take value from their data as quickly as possible so that the findings can be translated to improved patient care at the earliest opportunity. In many cases, this analysis is best performed using a server cluster."

For the administration team, the new system enables them to take control and deliver effective storage, back-up and archiving of data. Don says: "We are currently using around 50% of our 180Tb of storage. We can now automate data back-up and are doing so daily. Due to user error, we have had to restore some files. This was easy and painless. We can also manually archive data in accordance with our policies. We have two separate archives, one on-site and one off-site which reduces risk of data loss from system malfunction, disaster or other unplanned event. We will 'run out' of storage space within the next 4-5 months, so we're actively looking to increase our storage capacity with additional secondary (slower), storage (around 500TBs).

"We can now automate data back-up and we can manually archive data in accordance with our policies."

Don Lokuadassuriyage,
Biomedical Research Centre
System Administrator

To learn more about the topics covered here visit: <http://www.biomedicalresearchcentre.org/> and to see how OCF can help you visit: www.ocf.co.uk

