



HPC NEWS

HPC has a growing role to play in genomic sequencing

8 July 2014

Tweet 0

The growing importance of high-performance computing (HPC) in the life sciences was highlighted by two announcements this week of genome research centres upgrading their HPC infrastructures.

Public Health England (PHE) and The Centro Nacional de Análisis Genómico (National Centre for Genome Analysis) (CNAG) in Barcelona have upgraded their HPC resources to enable faster and more effective analysis of genome sequences, with the eventual goal of developing genomic science to the point where it can be used in personalised medicine.

Genome research generates massive amounts of data, which requires large computer processing and storage capacity. By using HPC, researchers can eliminate the bottlenecks that occur in data analysis and can store that data more efficiently.

Ivo G Gut, Director of CNAG said: ‘The challenge in genomics is such that it cannot be met with traditional computing. The solution is not to increase the number of sequencers. The key lies in the balance between sequencing and HPC. It is not only about increasing sequencing capacity by acquiring new hardware, but designing an appropriate computing infrastructure. It is also essential to choose a flexible infrastructure that can grow to keep pace with genome projects.’

Public Health England (PHE), an executive agency of the Department of Health in the UK, is upgrading its HPC system and has also recently started using a DDN based storage system. The planned upgrade will see the addition of another 16 compute nodes to the existing IBM system, further increasing the potential to parallelise the analysis of genomic data.

PHE is also expanding its archiving storage capacity with 250 Tb of DDN WOS cloud storage and implementing the open-source data grid software iRODS to help organise, share, and protect scientific data. The system uses DDN, HP and IBM hardware, Open Source software including Linux and xCAT and commercial software and technical support from Univa and Red Hat.

Similarly Bull and the National Centre for Genome Analysis (CNAG) in Barcelona have been developing the resources at the CNAG facility to accommodate big data requirements of next generation sequencing. CNAG's capacity of more than 800 sequencing Gbases per day is equivalent to sequencing eight full human genomes a day.

In its current iteration, the HPC system at CNAG features 1,200 processing cores and 2.7 Petabytes of storage based on Lustre file system that uses DDN storage arrays. The facility also has an internal 10 Gbps network and multiple 10 Gbps direct physical connections to the Barcelona Supercomputing Centre which has over 10,000 compute cores.

The UK system was implemented and configured by the British-based integrator, OCF. Julian Fielden, OCF's Managing Director, said: 'The team at PHE realises the benefits of HPC and big data storage and is using both to set the standards for the rest of the world to follow. PHE is pioneering use of DNA bacterial sequence data to provide a public service. It's the first project of its type in the world.'

To support performance and data growth requirements, OCF has now installed DDN SFA storage and EXAScaler appliance with a Lustre File System with a capacity of 300TB and 2.5Gbps performance. PHE keeps data for around 3-4 months enabling researchers to analysis data sets simultaneously. The data is then tiered off to a DDN storage archive and also made available for sharing with clinical partners and other research organisations.

These resources have been upgraded in line with UK Prime Minister David Cameron's announcement in 2012 of the '100,000 Genome Project' where the personal DNA code of up to 100,000 patients, or infections in patients, will be decoded. The Department of Health prioritised a number of areas with infectious disease sequencing undertaken by PHE.

This project supports both the PHE's goal of being a leader in the adoption of genomics in clinical microbiology and takes steps towards achieving the goals of the 100,000 Genome Project.

'The full implementation of genomics in public health systems involves resolving certain specific scientific and technological challenges: these latest are the ones that Bull is determined to meet using its supercomputing capabilities,' said Natalia Jiménez, Senior Health and Life Sciences Adviser at Bull. Currently, the CNAG centre in Barcelona makes a significant contribution to three major international initiatives: the International Cancer Genome Consortium (ICGC), the International Rare Diseases Research Consortium (IRDiRC) and the International Human Epigenome Consortium (IHEC).

Related internet links

IBM

Bull

DDN