CASE STUDY: EMED LAB

MRC eMedLab is a collaboration between Europe's leading biomedical research institutions. It consists of University College London, Queen Mary University of London, London School of Hygiene & Tropical Medicine, the Francis Crick Institute, the Wellcome Trust Sanger Institute, the EMBL European Bioinformatics Institute and King's College London.

Funded by the Medical Research Council, the vision of MRC eMedLab has been to maximise the gains for patients and for medical research that will come from the explosion in human health data.

HPC Cloud Environment Supports Cancer, Cardio-Vascular and Rare Diseases Research

Impact Summary

Challenge:

To accumulate medical and biological data on an unprecedented scale and complexity, to coordinate it, to store it safely and securely, and to make it readily available to interested researchers.

Solution:

A private cloud HPC environment with significant data storage capacity and very fast internal networking designed specifically for the types of computing jobs used in biomedical research.

Result:

Hundreds of researchers studying various aspects of life sciences are supported by a shared HPC resource that everybody can share efficiently, whilst sharing the learning, technology and the science at MRC eMedLab.

Challenge

Researchers are hampered in their work to generate new medical insights because of the fragmented accessibility of fundamental clinical and research data, and the lack of a high-performance computing (HPC) facility in which to analyse them. Lots of biomedical projects want to access the same datasets. However, it's simply not practical – from a data transfer and data storage perspective - to have scientists replicating the same core datasets across their own, separate physical high performance computing resources.

To realise this potential, the consortium needed to accumulate medical and biological data on an unprecedented scale and complexity, to coordinate it, to store it safely and securely, and to make it readily available to interested researchers.

"Bioinformatics is a very, very data intensive discipline," says Jacky Pallas, Director of Research Platforms, University College London. "We want to study a lot of de-identified, anonymous human data. It's not practical – from data transfer and data storage perspectives - to have scientists replicating the same datasets across their own, separate physical HPC resources." MRC eMedLab wanted to adopt a cloud-like virtual environment to give it the flexibility to accommodate lots of different types of projects in a shared infrastructure resource.

Solution

The ground-breaking facility is focused on the needs of the biomedical community and will revolutionise the way data sets are shared between leading scientific institutions internationally. The private cloud HPC environment was built using a Red Hat OpenStack Platform with Lenovo Flex System hardware to create virtual HPC clusters bespoke to individual researchers' requirements.

The Lenovo x Flex system comprises 252 hypervisor nodes and a 10Gb Mellanox network with a 40Gb/56Gb core. Its five tiers of storage are managed by IBM Spectrum Scale. OCF worked closely with MRC eMedlab's research technologists and its partners Red Hat, Lenovo, IBM, Mellanox Technologies, and Jisc. The whole system has been designed, integrated and configured by OCF.

"We needed to create a single store for up to six Petabytes of data and a shared HPC environment within which researchers can build their own virtual clusters to support their work," says Jacky Pallas.

The Red Hat OpenStack Platform enables scientists to create and use virtual clusters bespoke to their needs and enables different institutions and research groups to securely co-exist on the same hardware, and share data when appropriate. Researchers are able access up to 6,000 cores of processing power.

Mellanox's networking Ethernet solution has enabled the HPC cloud infrastructure to optimise its performance and to accelerate big data analytics.

The HPC environment is hosted at a shared data centre for education and research, offered by digital technologies charity Jisc. The facility has the capacity to accommodate multiple and varied research projects concurrently in a highly collaborative environment. The cloud infrastructure was live in April 2015 and research use was scaled up during 2016, with new projects coming on board throughout the year.

Benefits

MRC eMedLab has become a key infrastructure resource for the MRC, which has funded six of these projects. Its success has been attributed to MRC eMedLab's concept of partnership working where everybody is using one shared resource. This means not just sharing the HPC resource and sharing it efficiently, but also sharing the learning, the technology and the science at MRC eMedLab.

"From the beginning there was an excellent partnership between the MRC eMedLab operations team and the technical specialists at OCF"

Dr Jacky Pallas, Director of Research Platforms, University College London

In total, there are over 20 different projects running on the MRC eMedLab infrastructure. Projects range from the analysis of virus genome sequences to enable the modelling and monitoring of infectious flu type epidemic to the deep mining of cancer genomics data to understand how cancer tumours evolve.

Professor Taane Clark, Professor of Genomics and Global Health, London School of Hygiene and Tropical Medicine says, "The processing power of the MRC eMedLab computing resource has improved our ability to analyse human and pathogen genomic data, and is assisting us with providing insights into infectious disease genomics, especially in malaria host susceptibility, tuberculosis drug resistance and determining host-pathogen interactions."

Jacky Pallas, says "From the beginning there was an excellent partnership between the MRC eMedLab operations team and the technical specialists at OCF, working together to solve the issues which inevitably arise when building and testing a novel compute and data storage system."

Professor Charles Swanton, Professor of Translational Cancer Therapeutics, The Francis Crick Institute and UCL Cancer Institute: "Understanding cancer evolution over space and time is a complex task reliant on advanced computational technologies. We are incredibly fortunate to be working with MRC eMedLab that allows us access to a high performance cluster in the cloud. MRC eMedLab has been critical to the developments in our understanding of how cancers evolve and adapt, the processes that foster cancer cell variation that allow natural selection to function and ultimately how we might go about slowing down tumour evolution to improve patient survival outcomes."

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