

Overview

- The UK-based Sanger Institute delivered over one-third of the work for the Human Genome Project.
- Platform LSF enabled improved workload management, with the scalability to extend to the UK Grid.

Challenge

- Improve researchers' efficiency and time-to-results for the Human Genome Project.
- Increase computing power to 400 nodes to run up to half a million sequence matching jobs per day.

Solution

- Platform LSF® software

Results

- Platform LSF enables faster, easier and more reliable running of each sequencing job, which is paramount to the success of the Human Genome Project.
- Platform's ability to provide phenomenal scalability will enable the Sanger Institute to implement blade technologies and heighten its involvement in the UK Grid Project.

Founded in 1992, the Wellcome Trust Sanger Institute is the focus for the UK sequencing effort of the Human Genome Project. The finished sequence, which was announced in 2003, is essential for growth of research worldwide to produce further medical advances. The Sanger Institute carried out nearly one-third of the work, making it the biggest contributor to the project.

To fulfill its target of producing the Institute's quota for the human genome sequence, the Sanger Institute needed to increase its computing power to 400 nodes to run up to half a million sequence matching jobs a day. The Sanger Institute depends on its workload management system to handle the load and distribute jobs evenly across its infrastructure.

Business Challenge

The Sanger Institute has worked very closely with Platform over the years to manage its workload distribution across its Compaq Alpha DSL10s to support the human genome sequencing required to fill the gaps in the DNA string. An increase of up to 500,000 jobs per day meant a shift from ATM to Gigabit networking on the Institute's compute farm. The farm is currently using a Cabinet Area Network (CAN), where each rack in each cabinet is seen as an individual component, with 400 nodes (cabinets) in use.

The Sanger Institute is actively involved in the UK Grid Project, developed for high-energy physics, including genome research. Once the UK Grid is complete, it will become an easy way to access a huge data stockpile from individual desktops around the world. Currently, the Sanger Institute has to synchronise its data on tape and similar data replication in other global institutions.

At the Sanger Institute there is currently a dual 34Mb connection into the site and an internal Grid is in operation. Once the UK Grid Project is under way, they could require 2.5Gb links to access the data, ultimately aiming at a 2.5Gb WAN.

"The Human Genome Project has used enormous, scalable compute power to make the data available throughout the project. The Project is as much an exercise in IT and systems needs as in lab science and we and our partners will finish the genome years ahead of schedule partly because of our investment in flexible systems."

Phil Butcher

Head of IT at the Sanger Institute



Solution

When the Sanger Institute started the Human Genome Project, they used 25 Platform LSF licenses. They now run between 600 and 700 Platform LSF licenses to manage the project.

Platform LSF works to increase productivity by increasing the throughput of batch and interactive jobs and the effective and timely use of specific Sanger Institute applications, ultimately improving the researchers' efficiency and time-to-results. By automating job scheduling across the Institute's compute farm, Platform LSF provides the team with knowledge they could not hope to gain otherwise.

At any given time, Platform LSF identifies what resources are important to individual jobs, where unused resources are, who gets to use the resources and when. This facilitates the faster, easier and more reliable running of each job, which is paramount to the success of the Human Genome Project.

Platform is experienced at working on large-scale projects and it is vital that the Sanger Institute uses a distributed computing solution that can extend its capabilities well into the future. Once a higher level of computing is required, the scalability of Platform LSF will enable the Sanger Institute to use in excess of 1000 nodes at any one time.

Research into additional resources for the future has included a look at blade technologies to add another 256 CPUs to the existing facility. The use of blade server allows low-powered CPUs to be housed in densely populated nodes, each containing between 300 and 400 CPUs. The compute farm at the Sanger Institute has the space to continue its growth in excess of 1000 nodes.

Customer Site

Sanger Institute (www.sanger.ac.uk)

The Wellcome Trust Sanger Institute was founded in 1992 as the focus for the UK sequencing effort of the human and mouse genomes. The Institute is responsible for the completion of the sequence of approximately one-third of the human genome and one-fifth of the mouse. The Institute is also a major contributor to the mapping and sequencing of the zebrafish genome and genomes of a range of pathogens. The Human Genome Project is an international collaboration to map and decipher the information contained within the human genome. Twenty research groups from six countries (Britain, China, France, Germany, Japan and USA) are involved.



About Platform Computing

Platform Computing's intelligent, practical enterprise grid software solutions help organizations optimize IT resources to Accelerate Intelligence™. We plan, build, run and manage grids that link IT to core business objectives, and help our customers improve service levels, reduce costs and enhance business performance. With industry-leading partnerships and a strong commitment to standards, we are at the forefront of grid software development, propelling over 1,600 clients toward powerful insights that create real, tangible business value. For more please visit www.platform.com.



World Headquarters

Platform Computing Inc.
3760 14th Avenue
Markham, Ontario
L3R 3T7 Canada
Tel: 905 948 8448
Fax: 905 948 9975
Toll-free tel: 877 528 3676
info@platform.com

United States

Boston: 781 685 4966
Detroit: 248 359 7820
Reston: 703 251 4850
Newport Beach: 949 798 5654
New York: 212 672 1770
San Jose: 408 392 4900

Asia-Pacific

Beijing: +86 10 62381125
Hong Kong: +852 2869 5687
Tokyo: +81 3 5326 3105

Europe

Düsseldorf: +49 (0) 2102 610390
London: +44 (0) 1256 370500
Paris: +33 (0) 1 41 10 09 20

For more information, visit www.platform.com/contactus

Copyright © 2005 Platform Computing Corporation. ® TM Trademarks of Platform Computing Corporation.
All other logo and product names are the trademarks of their respective owners, errors and omissions excepted. Printed in Canada. Platform and Platform Computing refer to Platform Computing Inc. and each of its subsidiaries.