UK's TGAC Uses Coherent Shared Memory Supercomputer for Genome Analysis

Professional Services Organization's Implementation Sets Vendor Apart

Key Facts

Organization:
The Genome Analysis Centre (TGAC)

Location:
Norwich, UK

Application:
Life Sciences

The Genome Analysis Centre (TGAC) specializes in genomics and bioinformatics with a focus on analysis and interpretation of plant, animal and microbial genomes. Launched in July 2009, TGAC has steadily grown its team to just over 50 members, with half of the institute working in a bioinformatics capacity to interpret, assemble and analyze datasets generated from the sequencers housed in their lab.

Located at the Norwich Research Park in Norwich, UK, TGAC receives strategic funding from the UK's Biotechnology and Biological Sciences Research Council (BBSRC).

SGI UV 100 Used for Over a Year

Mario Caccamo is Head of Bioinformatics at The Genome Analysis Centre, and Paul Fretter is the Science Computing Team Leader, Norwich Bioscience Institutes. They have been using an SGI UV 100 since early 2011.

"We use the SGI UV 100 for next generation sequencing assembly and analysis with our own cutting edge software and algorithms," states Fretter. "The main benefit of using such a system is the ability to assemble and analyze large and complex genome sequences in memory."

Caccamo adds, "This purchase was important for our work. We were experiencing memory limitations in standard X86 hardware and thus had difficulty in assembling large and complex genomes."

"The purchase was driven by the existing infrastructure of a Beowolf cluster which did not fit the hardware profile required by cutting edge sequence production and analysis software," he continues.

The original installation also included a flash storage solution as well as an NVIDIA® Tesla™ S2050, all integrated by SGI Professional Services.

An additional UV 100 was purchased with switch-on scheduled for June 2012. The machines will be configured for defined applications to ensure the right balance throughput and latency.

The installation, in totality, also includes SGI InfiniteStorage™ consisting of 54TB worth of SAS drives. The entire UV network shares the same CXFS™ file system.

SGI Chosen for Superior Performance and Scalability

"SGI was chosen because of the product's superior performance and scalability, and especially the company's support and experience in high end, high performance computing," states Fretter.

"The ability to effectively process large amounts of genomics data will provide scientists with the missing pieces required to understand the genetic basis of some important traits such as yield and salt resistance in wheat and other important crops" states Caccamo.

Installation Enhanced by SGI Services Organization

SGI Professional Services was involved in helping the Centre drive the implementation of fusion_IO SSD and the integration of the NVIDIA component. At the time of installation, this was the world's largest Red Hat® Enterprise Linux® 6 system.

"We were quite pleased with the Professional Services team's highly effective project management, enthusiasm and willingness to get involved all the way down to the application level," says Caccamo.

"We are very pleased with the versatility of the system in adapting to different types of bioinformatics analysis. For example, the system is good for both divide-and-conquer approaches as well as big memory tasks," Fretter explains. "Also, using an open source operating system on X86 hardware is key in bioinformatics today," he adds.

Caccamo explains that "streaming bioinformatics information" is the future trend, and that the SGI platform is definitely a good choice. As costs drop and sequencing technologies improve, we will see that this kind of approach will increasingly be used in diagnostics and disease surveillance where in order to meet the performance requirements the data will need to be processed on demand as it emerges from the instruments."

About TGAC

TGAC is a specialist in genomics, including DNA sequencing and whole genome scale analysis, and in bioinformatics for analysis and interpretation of sequencing data, particularly in plant, animal and microbial science. Their work drives and enables advances in bioscience applications, particularly focused on food, health and the environment.

Opened in July 2009, TGAC has already established itself as an expert partner in high throughput sequencing, with projects including de novo sequencing of a rubber tree genome in collaboration with an industrial client and partnering as a member of the international wheat genome project consortium. Its team includes genomics scientists, technologists and bioinformaticians.

Please visit www.tgac.ac.uk

Global Sales and Support: sgi.com/global