A global leader in genome research

Wellcome Sanger Institute helps the world understand and fight COVID-19 and other deadly diseases.

Customer profile

wellcome sanger institute

Scientific Research | United Kingdom

"At its core, the Sanger Institute is performing genomic research for the betterment of human health worldwide."

Dr. Peter Clapham
Team leader, Informatics Support Group,
Wellcome Sanger Institute

Business needs

Scientists working with the Sanger Institute need leading-edge high performance computing systems to power genome research.

Business results

- Accelerating genome research
- Unlocking the secrets of deadly diseases
- Driving scientific discoveries and innovations
- Improving the lives of people around the world

Solutions at a glance

- HPC systems with OpenStack®
- Dell EMC PowerEdge C6420 and R740 servers wiith NVIDIA T4 GPUs
- Dell EMC PowerScale Isilon scale-out NAS
- NVIDIA© DGX™ Systems
- NVIDIA Clara™ Parabricks software

Improving life with genome sequences

The Wellcome Sanger Institute is a world leader in genome research that delivers groundbreaking insights into the biology of humans, pathogens and species across the tree of life. Scientists at the UK institute tackle some of the most difficult challenges in genomic research to fuel scientific discoveries and push the boundaries of our understanding of biology in ever new and exciting ways.

Studies conducted through the Sanger Institute and its scientific programs provide vital insights into human, parasite and microbe evolution, cellular growth and activity, and the processes that underlie mutation and tumor formation. In addition, the Institute's Tree of Life program explores the diversity of complex organisms found in the UK through sequencing and cellular technologies.

This is more than theoretical research. With its collaborative approach to research and innovations, the Sanger Institute generates insights that improve people's lives around the world on a daily basis.

"With every new generation of hardware, whether it be IT or sequencing, there are new opportunities that our scientists can take advantage of."

Dr. Peter Clapham

Informatics Support Group
Wellcome Sanger Institute

And this is very much the case today with the Institute's work to track the spread and evolution of the SARS-CoV-2 virus, which causes the deadly COVID-19 disease.

Using powerful NVIDIA DGX systems, researchers from the Wellcome Sanger Institute — a world leader in genomics — and the University of California San Diego, collaborated with NVIDIA developers to achieve more than 30x acceleration when running their machine learning software SigProfiler. Read more.

Researching a deadly virus

At the start of the COVID-19 pandemic, the Sanger Institute, together with UK Public Health Agencies, academic partners and NHS organizations across the country, formed the COVID-19 Genomics UK Consortium (COG-UK) to sequence and analyze virus genomes.

One of the aims of the COG-UK initiative is to use sequence data to trace the SARS-CoV-2 virus as it evolves and spreads. Tracking the virus within a hospital, town, country or global environment is possible because genomes mutate — which means that letters in the genome sequence change as organisms replicate. Researchers can use this data to determine the relatedness of different viruses and help identify chains of transmission, super-spreading events and fast-growing variants.

"At the moment, the Sanger Institute is very heavily involved in the UK COVID response," says Dr. Peter Clapham, team leader for the Institute's Informatics Support Group, which provides high performance computing environments for Sanger's scientific research teams. "Our scientists are performing a lot of genome sequencing onsite and conducting the primary analysis to turn the sequences into something useful that we can then share with other sites to perform downstream analysis."

Today, the Sanger Institute's large-scale genome sequencing capabilities and the data they generate are being used, in real time, to inform public health measures and to help save lives, according to Dr. Clapham.

As of May 2021, the Sanger Institute had sequenced over 300,000 coronavirus genomes. That's a sign of the enormous scale of the research taking place at the Institute — and the extreme demands placed on the Institute's HPC resources.

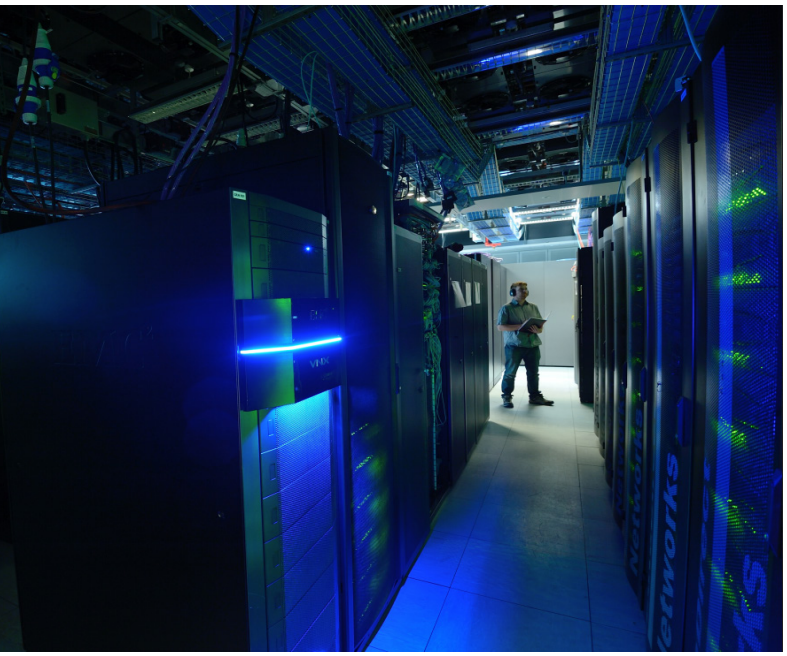
The power of HPC

To do science on the scale of that conducted at the Sanger Institute, researchers need access to massive amounts of data processing power — and this is where Dr. Clapham's Informatics Support Group shines. The ISG team provides support, architecture design and development services for the Sanger Institute's traditional HPC environment and an expansive OpenStack private cloud compute infrastructure, among other HPC resources.

The traditional HPC cluster provides about 38,000 cores of computational resource, 23.5PB of Lustre clustered file systems and a 30PB virtualized iRODS managed storage repository, all tied together with a 160Gb/sec network backbone.

The OpenStack private cloud encompasses about 12,000 cores, more than 100 TB of memory and 5.5 PB of usable NVMe index accelerated Ceph storage, 1 TB or 2 TB SSD per compute node for high-speed ephemeral storage, and 100GB/sec point-to-point software-defined networking.

As those system details show, the ISG team delivers HPC on a huge scale to meet the needs of the Sanger Institute's research community. The ISG typically supports more than 1,000 onsite scientists and their off-site collaborators.



"With every new generation of hardware, whether it be IT or sequencing, there are new opportunities that our scientists can take advantage of, and Sanger's ability to take on these huge-scale research projects is one of the things that really sets it aside," Dr. Clapham says. "The Institute has the ability to do at scale things that a lot of other institutes just don't have the opportunity to do."

Accelerating storage performance

Ever-faster and more scalable data storage is one of the ongoing success factors for the Sanger Institute's HPC environment. With that thought in mind, the IT leaders at the Sanger Institute made the decision to move ahead to Dell EMC PowerScale scale-out storage. The Institute now uses PowerScale to store genomic sequencing data for computational analysis and to make the vast amounts of data available to researchers around the world quickly and reliably. Read the case study.

In addition, with PowerScale storage, the Sanger Institute has been able to scale and upgrade its storage without sacrificing simplicity. The migration-free design of PowerScale with its auto-discover capabilities allows new nodes to be added quickly while legacy nodes can be decommissioned with no downtime. All the while, auto balancing helps ensure that as storage scales out there are no "hot spots" that can constrict processing.

Genomics comes of age

As he reflects on the devastating impact of the COVID-19 pandemic, Dr. Clapham sees hope in the saving power of science. Thanks in large part to the advance of genome sequencing, humanity is fighting back admirably against the SARS-CoV-2 virus.

"It has been our greater understanding of genomics, and the informatics that goes with it, that has accelerated the development of vaccines and our ability to respond to this disease in a way that's never been possible before," he says. "We can now track the variations and the mutations of the virus as they arise, and develop vaccines and treatments at unprecedented speeds."

Along the way, the world is witnessing firsthand the amazing power of genomic science, as millions of people flock into vaccination centers to arm their bodies to fight off the virus.

"I think people are now seeing, for the first time, the true benefits of all the genomics work that's taking place behind the scenes," Dr. Clapham says. "They are seeing the evidence of genomics coming of age."

> “I think people are now seeing, for the first time, the true benefits of all the genomics work that’s taking place behind the scenes.”

Dr. Peter Clapham

Team leader, Informatics Support Group, Sanger Institute