The Wellcome Trust Sanger Institute selects Arista for innovative cloud-based infrastructure as a service platform to aid scientific breakthroughs

Highlights

Challenge

The Wellcome Trust Sanger Institute required an open and flexible switching architecture to underpin its new OpenStack cloud and S3-compatible storage architecture.

Solutions

- Arista 7060X Series Switches
- Arista EOS®
- CloudVision® workload orchestration and automation

Results

- Arista switches deliver high performance with low latency and ability to scale to 100GbE
- CloudVision offers enhanced automation reducing management overheads
- New cloud delivers better job queuing and resource allocation speeding up analytics tasks

The Wellcome Trust Sanger Institute is one of the world's pre-eminent genome research centres and was the single, largest contributor to the global Human Genome Project. As part of an ongoing strategy to enhance the agility of its information technology systems, the Institute has developed a standards based cloud using OpenStack software and Arista networking technology. With its new cloud-based infrastructure as a service platform, the Institute is helping its scientists and research projects across the world gain better access to critical informatics resources with the ability to scale in line with advances in technology.

Project Background

The Wellcome Trust Sanger Institute is one of the premier centres of genomic discovery and understanding in the world. It leads ambitious collaborations across the globe to provide the foundations for further research and transformative healthcare innovations. Its success is founded on the expertise and knowledge of its people and the Institute seeks to share its discoveries and techniques with the next generation of genomics scientists and researchers worldwide. More than a 1000 people work at the Institute organised into five scientific Programmes, each defining a major area of research with a particular biological, disease or analytic focus. In all cases, the studies provide insights into the phenotypic and biological consequences of genome variation and the processes which cause mutations, including in humans, animals, pathogens, and cellular evolution.

Challenge

Since its foundation in 1993, The Sanger Institute has provided the ability to conduct research at scale and engage in bold and long-term exploratory projects that are designed to influence and empower medical science globally. Institute research findings, generated through its own research programmes and through its leading role in international consortia, are being used to develop new diagnostics and treatments for human disease across the world.

Many of the projects rely on genome research which in turn means the Institute is at the leading-edge of genomics technology development and implementation. This has led to Innovation in aggregation, analysis and interpretation of large quantities of genomic data as part of both local and global collaborative research initiatives with international partners.

With vast quantities of scientific research data within its systems and ongoing projects generating tens of terabytes of new data each day, its information technology infrastructure is a vital element of the Institute's ability to deliver new scientific breakthroughs.

"To give it a sense of scale, a single Genomic sample can generate up to 160 Gigabytes of data," explains Dr Peter Clapham, Informatics Support Group Team Leader, "Just one of our projects is sequencing around 3000 samples at the moment and this data is then processed using various analysis techniques and made available for other researchers across the world."

The Institute runs one of the largest computing resources in the UK with over 20,000 CPU cores and 40 petabytes of storage, both of which are expected to increase by a third over the next few years. However, one of the challenges it faces is the ability to efficiently allocate and scale resources for projects as needed. The Institute had grappled with this problem in the past when looking at how it could provide IT resources for organisations that had spun out of the Institute and created a small private cloud. Yet scaling the previously successful design for the much larger central IT function required a design rethink.

Solution

"One of our major requirements is the ability to scale," explains Dr Clapham, "The sequencing process is becoming faster and our data sets are continually growing plus advances in processing technologies allows us to conduct more complex research in less time."

For the new service based architecture, the Institute recognised the need for a standards based cloud approach and chose OpenStack along with its Neutron component to deliver a set of application program interfaces (APIs) to enable interoperability and orchestration of network devices and technologies. "The network is critical in delivering a robust infrastructure-as-a-service (IaaS) environment," explains Dr Clapham, "Low latency and scalable performance were a given, but we also needed maturity on the SDN side."

"One of the considerations for a switching technology was one offering an open approach with the ability to work seamlessly with OpenStack and Neutron as well as bespoke hardware, for example FPGA's or even bare metal," he adds.

Dr Clapham and his team examined and tested a number of networking solutions engaging in a multi-month process including getting in test kit and talking to developers. "We weren't just looking for a piece of hardware, we were looking for a partner," he explains, "The impact of this project will affect not just us but research projects around the world and we were embarking on a bit of a journey of discovery."

Following a multi-month implementation project working closely with technical experts from Arista Networks, the first iteration of the cloud architecture has been deployed with some of its early adopters. The cloud uses Arista 7060X switches throughout a design that has Layer 3 Leaf at 25GbE and Spine at 100GbE. The switch also uses a large shared packet buffer and delivers a maximum I/O rate of 6.4Tbps at a low latency of 450ns. The Institute also uses Arista's CloudVision® portal software for centralized workload orchestration and automation as well as zero-touch provisioning.

Conclusion

"We are at the last stages of deployment using pre-production code while working with a few internal customers and so far the results are good," explains Dr Clapham, "One of our scientists says that a project that would normally require queuing that would lead to months of waiting was completed in two days as the system was able to allocate and dynamically provision resources as they became available to complete the job faster."

The improvements have also been felt by the ICT team. As Jon Nicholson, Networking Specialist for the Institute explains, "The major benefit is that there is relatively little to do on the networking side as CloudVision gives us a single access point for change control and management. The biggest shift for us is that we are now able to manage the infrastructure as code and this opens up more opportunities for automation."

As the team move into its full rollout and plans the next iteration of its cloud, which will add support for S3-compatible object storage to aid data sharing, Dr Clapham believes that the project and working with Arista has been a success. "It has been delivered as expected into a production environment as part of an iterative process. Over the project life cycle, software versions evolved and we worked together to deliver a solution – it has been a journey with less surprises and ultimately a positive one."

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office 1390
Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office
Global Tech Park, Tower A & B, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office
10 Tara Boulevard
Nashua, NH 03062