Success Story

Accelerating Genomic Medicine Research with Shared Computational Resources

Life Sciences
Japan

東北メディカル・メガバンク機構
TOHOKU MEDICAL MEGABANK ORGANIZATION

- Multi-user, multi-protocol data sharing
- GPU Accelerated Computing
- Life Science Research
- Artificial Intelligence and Deep Learning

Solution

System overhaul for improved performance and build capabilities for simulations, data analysis and sharing.

Professor Kengo Kinoshita, Ph. D.

Deputy Executive Director of ToMMo and Director, the Center for Genome Platform Projects

Tohoku University Tohoku Medical Megabank Organization (ToMMo)

A More Robust Biobank, Through Storage Expansion: Accelerating genomic medicine research by providing shared computational resources to researchers and research organizations.

Tohoku University Tohoku Medical Megabank Organization (ToMMo) was established in February of 2012 to begin building the advanced medical system. Created at the heart of the disaster area hit by the 2011 Great East Japan Earthquake and Tsunami, it also sought to help rebuild those areas.

Its supercomputer system, launched in July of 2014, underpins a significant part of its operations and has produced a great deal of results. The system consolidates analytical data about biospecimens, including health survey and genome sequence data; that data can then be shared with researchers across Japan through a registration and review process. ToMMo is also developing special educational programs in leading-edge medical fields, like genetic research.

In the four years since the launch of this system, it was the largest of its kind in northeastern Japan in the life sciences field. With the support of the Japan Agency for Medical Research and Development, the system was overhauled in 2018, improving the organization's competitiveness internationally and providing and sharing data and computational/ analytical functionality with researchers and research organizations struggling to secure such resources. Professor Kengo Kinoshita, Ph.D., Deputy Executive Director of ToMMo and Director, the Center for Genome Platform Projects, manages this system and discussed its history and outlook.

The Challenge

- Keeping costs down while updating a system under a tight budget.
- Enabling external access for data analysis and sharing, and building a foundation for a national-level initiative across all of Japan.
- An increased need for the system's ability to run simulations, and boost storage capacity.
- Moving large amounts of data without system outages.

According to Dr. Kinoshita, "Our operations, built in large part to help with disaster recovery, have entered the phase where we return surveyed results to people in the disaster areas. This has raised the bar for operating our system. However, by bringing in new technologies we didn't have in the beginning - AI, GPUs, next-generation CPUs - we have boosted our analytical capacity and our sample sizes, giving us a major leg up in terms of the accuracy of our data analyses. Therefore, by being able to deal with a much larger dataset, methods that had previously only existed in theory have now become viable for our use."

The number of whole-genome sequence data from samples supplied by residents in the disaster area went from 1,000 in autumn of 2013 to around 4,000 as of January 2019, and is expected to reach around 5,000 by year's end. The number of total participants in the organization's cohort studies has reached 150,000, and the samples and data collected as part of these studies has been not only put to use by ToMMo, but also by a large number of outside researchers to accelerate and advance a great deal of research projects.

Dr. Kinoshita believes that "the era where we defend our ownership of data to the death is over."

"Sustainability in medicine is building a system that can maintain each individual's health data for fifty years. When a child born today falls ill fifty years from now, being able to trace their entire medical history would be an incredible achievement. I want to build a system where the data when that person was three years old can be called up in an instant. That vision is what led us to expand our total data storage pool to 29 petabytes. I can tell you that moving large amounts of data into this new storage system is presently one of our biggest challenges."

Dr. Kinoshita concluded by saying, "We're coming out of the 'our computer' era and moving into the 'everyone's computer' era. It's going to take a new kind of initiative built on this new sharing mindset to build a mechanism that can support that."

Performance

System enables fast data analysis for GPU-based simulations

Scale

Expanded capacity, improving functionality

Flexibility

Enabled external data access for new data analysis use cases and research sharing

Experience

Migration complete within two days and no interruptions to research activity

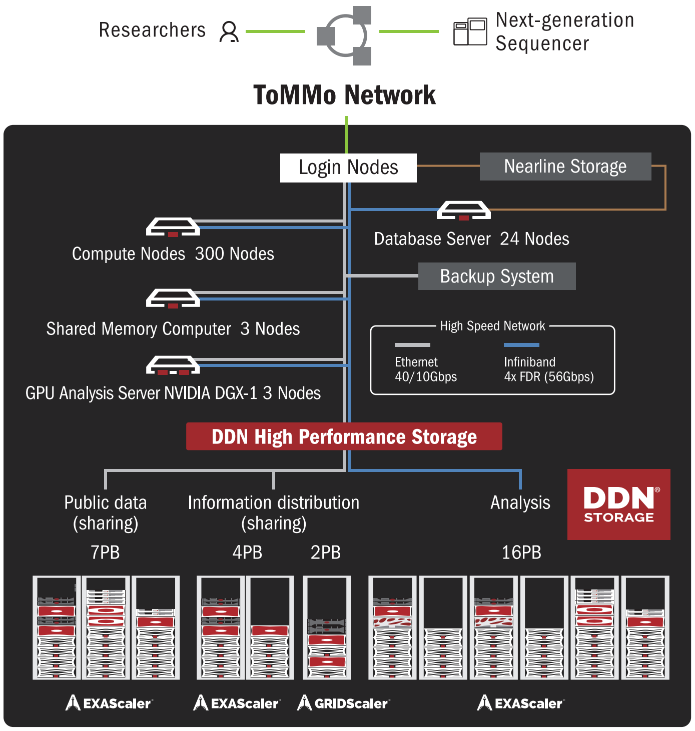
"By bringing in new technologies we didn't have in the beginning - AI, GPUs, next-generation CPUs - we have boosted our analytical capacity and our sample sizes, giving us a major leg up in terms of the accuracy of our data analyses. Therefore, by being able to deal with a much larger dataset, methods that had previously only existed in theory have now become viable for our use."

Professor Kengo Kinoshita, Ph. D.

Deputy Executive Director of ToMMo and Director, the Center for Genome Platform Projects

The Solution

ToMMo Supercomputer System Overview



The Benefits

- Enabled system overhaul under limited budgets, combining Infiniband and 40/10 Gbps Ethernet with a DDN Storage solution to improve performance and build capabilities appropriate for the different needs for each of the three computational units.
- Greatly expanded storage capacity to 29 PB, improving data biobank functionality and supporting Japan's genomic medicine research, all while limiting cost outlays.
- DDN Storage is connected to a NVIDIA DGX-1 GPU-based analysis server running Parabricks (genomic analysis software); the combination of these technologies results in superior performance.
- Smoothly migrated an enormous amount of existing data (approx. 6 PB) to the new system within only two days without interrupting research activities.

Future Challenges

DDN was expected to proactively propose solutions leveraging the newest technology to accelerate research. Compared to when the system was initially implemented, the organization has become a more critical player and has a greater presence as a datacenter, which has translated to a heavier burden of social responsibility as well. Though the system's original intended use was for analysis, it has become a system upon which many researchers have built their projects.

Going forward, there will be more of a need to make contributions not only as a provider of superior technology, but also as a partner who can respond to users' growing emphasis on reliability.

This article was drafted based on interviews conducted at the Tohoku University Tohoku Medical Megabank Organization on January 31, 2019.

About DDN

DataDirect Networks (DDN) is the world's leading big data storage supplier to data-intensive, global organizations. DDN has designed, developed, deployed, and optimized systems, software, and solutions that enable enterprises, service providers, research facilities, and government agencies to generate more value and to accelerate time to insight from their data and information, on premise and in the cloud.

v2 (4/20)