Hewlett Packard Enterprise

大学共同利用機関法人 情報・システム研究機構
国立遺伝学研究所

Industry
Life Sciences

Objective
- Refresh supercomputing system for genome analysis and genome database development.
- Secure resources to meet growing demand for analysis.
- Meet diverse workload demands with the right resources in the right place.

Approach
- Deployed optimal resources to meet diverse analytical needs by selecting different types of platforms such as memory-shared, cluster-type, and GPU-mounted systems.
- Developed an environment to reduce researchers' burdens such as advance preparation.

IT matters
- Adopted the right HPE platform in the right place selecting large memory-sharing systems, large cluster-type systems, and GPU systems to create the genome analysis infrastructure with the right mix.
- Achieved high performance on the assembly processing that restores genome sequences by deploying it on HPE Superdome Flex that allows the use of 12 TB memory from a single operating system instance.
- Facilitated the management and development of OS, job schedulers, AI-related frameworks, libraries, etc. with cluster management tool Bright Cluster Manager.

Business matters
- Maximized cost efficiencies of the supercomputer system by allocating computational resources optimal for a wide range of analytic demands unique to the genome field.
- Eliminated researchers' tasks related to the preparation for analysis and let them focus on research operations.
- Established analytical techniques that use AI, with expected progress in bioinformatics research.

NATIONAL INSTITUTE OF GENETICS ACCELERATES LIFE SCIENCE AND GENOMIC RESEARCH BY BUILDING 5TH GENERATION SUPERCOMPUTER SYSTEM

Achieves ~1.1 PFLOPS total theoretical performance, 2.2X versus prior system, by widely adopting HPE HPC portfolio

In March 2019, NIG reinvented its supercomputer system and achieved total theoretical operation performance of approximately 1.1 PFLOPS, about 2.2 times compared to the previous system. This 5th Generation NIG Supercomputer is used for the International Nucleotide Sequence Database (INSD), in which researchers around the world register base sequence data, and as an environment for performing various types of analytic processing related to genomics. NIG is positioned as the core institution for Japan's life science and genome medical research. HPE HPC solution was adopted widely, spanning a large memory-sharing system, cluster-type system, GPU system, and an integrated operation management environment.

Case study
National Institute of
Genetics

Industry
Life Sciences

Page 2

"A large memory space that exceeds 10 TB is very effective in assembly processing, in which an extremely large amount of genomic fragments are connected to restore sequences."

– Ogasawara Osamu, PhD, Project Associate Professor, Head of HPC Division, DDBJ Center, National Institute of Genetics

Ogasawara Osamu
PhD (physical science)
Project Associate Professor

Head of HPC Division
DDBJ Center
National Institute of Genetics

CHALLENGE

Supercomputer system refreshed by Japan's core institution for genomic medical research

The National Institute of Genetics (NIG) is the core institution for Japan's life science and genomic medical research, engaging in cutting-edge genetics research, preservation and use of genetic resources, development and use of genetic information databases, and advanced education and human resource development based on genetics. As a member of the Research Organization of Information and Systems (ROIS), an Inter-University Research Institute Corporation, NIG provides researchers in wide-ranging fields with opportunities to share the supercomputer system and conduct joint research.

Dr. Ogasawara Osamu, a project associate professor in physical science, who serves as the Head of High Performance Computing Division of DDBJ Center, explains: "The DNA Data Bank of Japan (DDBJ) Center has been operating this supercomputer system (NIG Supercomputer) since 1996 with two key missions. One is to build the International Nucleotide Sequence Database (INSD) with the cooperation of the U.S.-based National Center for Biotechnology Information (NCBI) and the European Bioinformatics Institute (EBI). The other is to provide life science and genomic medical researchers with powerful computational resources."

Genome researchers around the world are required to register their research achievements in one of the databases, including the DDBJ, NCBI, and EBI. Data in the U.S., Europe, and Japan databases is exchanged among them to enable operations as a globally shared database of base sequences. Data registered in the INSD is preserved permanently as scientific information and available for use by anyone without any limitation. It is an asset to be shared by all humans. In addition to the INSD development, the NIG Supercomputer

has been used over the years for analytic processing of a vast amount of diverse data, ranging from personal genome to metagenome and from the molecular level to the population level.

"The use of next-generation sequencers for the human genome analysis and clinical application has resulted in an explosive increase in the amount of data. We significantly improved the architecture of the NIG Supercomputer in 2012 and achieved 15 times greater performance than the previous system, while controlling costs by shifting it to open-source software. The first platform that we used for this system was a then HP (now HPE) product," says Dr. Ogasawara.

Since then, the cost-efficient HPE platforms have powered the NIG Supercomputer. Subsequently in March 2019, the 5th Generation NIG supercomputer operation, consisting of the latest large memory-shared system and large cluster-type system, started.

SOLUTION

Providing optimal computational resources to meet diverse demands in genome analysis

The 5th Generation NIG Supercomputer achieves a total theoretical operation performance of approximately 1.1 PFLOPS, which is about 2.2 times compared to the previous system. Its key features include providing optimal analysis resources to meet each of the researchers' demands, facilitating the researchers' analysis environment setup using container virtualization technology, and facilitating the management of cluster environments with varying specifications.

"We improved the analytic performance and raised storage capacity while maintaining the basic architecture. Analysis resources are delivered by a fat compute node, which provides 12 TB large shared memory, medium compute nodes equipped with

Case study
National Institute of
Genetics

Industry
Life Sciences

Page 3

National Institute of Genetics (NIG) in Japan accelerates life science and genomic medical research by building 5th Generation NIG Supercomputer System with the right mix of HPE HPC solutions.
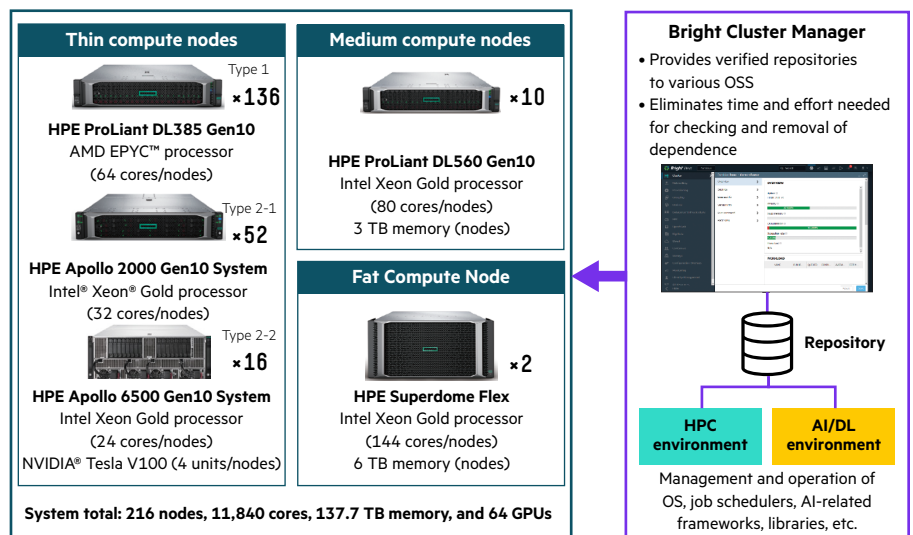
[f] [in] [twitter]

3 TB memory, and a thin compute node that structures clusters using a total of 204 servers. We have also introduced the latest GPU platform for AI to perform genome analysis," says Dr. Ogasawara.

The server selected for the Fat Compute Nodes is HPE Superdome Flex. It is the industry's largest SMP system allowing the use of a single memory space of up to 48 TB on a single operating system instance. Its unique crossbar fabric composed of crossbar chips arranged in a mesh can minimize the latency in the connection between

nodes and maximize the benefit of large in-memory processing in the HPC domain.

"A large memory space that exceeds 10 TB is very effective in assembly processing, in which an extremely large amount of genomic fragments are connected to restore sequences. The ability of NIG, as an Inter-University Research Institute Corporation, to provide a large analysis system that is difficult to be placed in ordinary research institutions, has a great significance," remarks Dr. Ogasawara.

The 5th Generation NIG Supercomputer System

Thin compute nodes

Type 1
×136
HPE ProLiant DL385 Gen10
AMD EPYC™ processor
(64 cores/nodes)

Type 2-1
×52
HPE Apollo 2000 Gen10 System
Intel® Xeon® Gold processor
(32 cores/nodes)

Type 2-2
×16
HPE Apollo 6500 Gen10 System
Intel Xeon Gold processor
(24 cores/nodes)
NVIDIA® Tesla V100 (4 units/nodes)

Medium compute nodes

×10
HPE ProLiant DL560 Gen10
Intel Xeon Gold processor
(80 cores/nodes)
3 TB memory (nodes)

Fat Compute Node

×2
HPE Superdome Flex
Intel Xeon Gold processor
(144 cores/nodes)
6 TB memory (nodes)

System total: 216 nodes, 11,840 cores, 137.7 TB memory, and 64 GPUs

Bright Cluster Manager
• Provides verified repositories to various OSS
• Eliminates time and effort needed for checking and removal of dependence

Repository

HPC environment | AI/DL environment

Management and operation of OS, job schedulers, AI-related frameworks, libraries, etc.

For the Type 1 system, consisting of 136 nodes, the largest number among thin compute nodes, NIG selected the HPE ProLiant DL385 Gen10 server equipped with a 32-core AMD EPYC processor. "There are a number of applications in which high integer operation performance such as mapping processing of reference genome sequences becomes advantageous. In Type 1 thin compute nodes, we considered that being able to use the multiple cores and wide memory band of AMD EPYC processor offered a significant benefit," says Dr. Ogasawara.

HPE Apollo 6500 Gen10 System is categorized as Type 2-2 thin compute node. Dr. Ogasawara explains the objectives of

using the GPU server that can mount up to eight units of the latest version of NVIDIA Tesla V100 as "one is the application to parallel processing that uses the many cores characteristic of GPUs and the other is the development of a deep learning model for promoting AI-based genome analysis."

More than 2,300 types of genome analysis software are registered in the 5th Generation NIG Supercomputer System. This system achieves the use of resources in the "right mix" (the right resource in the right place) using a set of HPE platforms with different characteristics in order to meet various analytic demands with the optimal computational resources.

Case study
National Institute of
Genetics

Industry
Life Sciences

Page 4

"The 5th Generation NIG Supercomputer has been built to contribute to the activities of researchers engaging in life science and genomic medical research by building and sharing the INSD and providing powerful analysis resources. We not only stepped up the computational resources, but significantly improved the usability for researchers by, for instance, reducing advance preparation. We expect that a number of new achievements in bioinformatics research will be made from now on."

– Ogasawara Osamu, PhD, Project Associate Professor, Head of HPC Division, DDBJ Center, National Institute of Genetics

Use of Bright Cluster Manager for managing a complex cluster environment

Measures to manage a large and complex environment in an integrated manner and meet researchers' demands promptly and efficiently are taken in the 5th Generation NIG Supercomputer. This important role is played by Bright Cluster Manager, which is a cluster management tool equipped with an intuitive GUI and provides comprehensive functions such as provisioning, image management, and cluster monitoring. It has an extensive track record both in the AI/deep learning space and also in HPC systems. "We Adopted Bright Cluster Manager, to which we had been paying attention, in order to simultaneously meet the conventional demands in genome analysis and new analytic approaches using AI/deep learning. This makes it easier to manage and operate an OS, job scheduler, AI-related frameworks, library, and others. Being able to use Singularity containers is also a major benefit," says Dr. Ogasawara.

NIG uses Singularity containers to help researchers install genome analysis software and ensure the reproducibility of an analysis environment. This packages an application, library, and data as container images and creates an analysis environment that is highly reproducible, portable, and available. "Complex installation tasks and examination of combinations that would work correctly in testing middleware for container management such as Kubernetes and Mesos gave us a serious stress. We found Bright Cluster Manager while searching for tools to automate these tasks. This proved to be an ideal solution for us, as we must consider and examine new technologies at all times," reflects Dr. Ogasawara.

The repositories provided by Bright Cluster Manager are delivered in combinations in which various OSS used in AI/deep learning work correctly. It is also increasingly used in Japan as the benefit of reducing the time and effort needed for checking and removal of dependences is recognized. As the authorized distributor, Hewlett Packard Enterprise Japan is the only company in Japan providing local language support for Bright Cluster Manager. The company has also improved the installation service through HPE Pointnext Services.

BENEFITS

Use of Silicon Root of Trust as part of defense in depth

A section for personal genome analysis was newly established in the 5th Generation NIG Supercomputer.

Dr. Ogasawara explains the purpose of this. "Demand for personal genome analysis has been doubling every year, which substantially exceeds the growth of other areas. A requirement strikingly different from others is strict security. Researchers around the world also recognize that the establishment of techniques to safely share personal genome data is one of the most important tasks."

The personal genome analysis section uses strict authorization and access control and a system to detect unauthorized access and suspicious behavior. Dr. Ogasawara has taken additional measures to ensure safety and security. "Security of the HPE ProLiant DL385 Gen10 server used in the personal genome analysis section has been strengthened based on the hardware, which allows detection and correction of

Customer at a glance

Hardware

- HPE Superdome Flex

- HPE ProLiant DL560 Gen10

- HPE Apollo 6500 Gen10 System

- HPE Apollo 2000 Gen10 System

- HPE ProLiant DL385 Gen10

Software

- Bright Cluster Manager

misconduct such as firmware tampering. We make use of this as part of multilayer defense. We consider that HPE's iLO 5 an outstanding feature among the commoditized x86 servers."

HPE Gen10 server platforms such as HPE ProLiant DL385 and HPE Apollo 6500 System include ASIC HPE Integrated Lights Out 5 (iLO 5). It is equipped with Silicon Root of Trust (silicon level security), corresponding to data encryption at the most powerful level with the commercialization through Commercial National Security Algorithm Suite (CNSA), and conforms to Special Publication 800-53 issued by the U.S. National Institute of Standards and Technology (NIST).

Dr. Ogasawara reflected on the project and summarized it as follows: "The 5th Generation NIG supercomputer has been built to contribute to the activities of researchers engaging in life science and genomic medical research by building and sharing the International Nucleotide Sequence Database (INSD) and providing powerful analysis resources. We not only stepped up the computational resources, but significantly improved the usability for researchers by, for instance, reducing advance preparation. We expect that a number of new achievements in bioinformatics research will be made from now on. We expect that Hewlett Packard Enterprise Japan will provide us with full-portfolio support from cutting-edge HPC solutions to consulting services."

LEARN MORE AT

hpe.com/info/hpc

**Make the right purchase decision.
Contact our presales specialists.**

Chat **Email** **Call**

Share now

Get updates

**Hewlett Packard
Enterprise**