NAIST drives forward with multidimensional Big Data analysis

HPE ProLiant DL385 Gen10 Server equipped with AMD EPYC enhances machine learning and cloud sourcing capabilities

Industry
Education

Objective
To generate new findings through multidimensional Big Data analysis. To create a system environment for data mining, machine learning, and analysis to support the data analysis and pursue high-speed, high-precision processing.

Approach
Select a server capable of installing large capacity memory that leads to enhanced performance of data mining, machine learning, and analysis with importance on the bandwidth of the interconnection between CPU and memory.

IT matters
• Used HPE ProLiant DL385 Gen10 Server equipped with an AMD EPYC™ processor to build a multidimensional Big Data analysis environment that enables use of up to 4 TB of memory.

• Built a virtual server environment that offers both high density and superior cost performance to integrate multiple systems.

• Reduced the process time frame from requesting a quote and, determining specifications to delivery, by discussing specifications via HPE DirectPlus.

Business matters
• Contributed to the establishment of a classification model that combines machine learning and cloud sourcing.

• Promoted the verification test of the service to provide real-time distribution of tourism information for foreigners who visit Japan in Kyoto Prefecture.

• Provided a multidimensional Big Data analysis environment with an excellent return on investment that is expected to contribute to generating new knowledge.

Affiliate Associate Professor Dr. Yu Suzuki and his research group at the Division of Information Sciences, Graduate School of Science and Technology of Nara Institute of Science and Technology (NAIST) are working on generating new findings employing multidimensional Big Data analysis. The research group has built a system environment for data mining, machine learning, and analysis, using HPE ProLiant DL385 Gen10 Server equipped with an AMD EPYC processor. Taking advantage of the superior architecture that connects CPU and memory at top speed has helped achieve high-speed, high-precision processing. The equipment was procured smoothly in a short time using the HPE online store, HPE DirectPlus.

"Combining structured data managed by businesses and organizations with unstructured data from a completely different perspective should offer a wide-ranging potential for creating new knowledge. The approach of merging machine learning and cloud sourcing opens a new path to making it possible to categorize unstructured data while sequentially creating teaching data."

– Dr. Yu Suzuki, Affiliate Associate Professor, Division of Information Sciences, Graduate School of Science and Technology, NAIST Data Science Center, Big Data Project, Augmented Human Communication Laboratory

Challenge

Contribute to the society and create value with multidimensional Big Data analysis

NAIST is a Japanese national graduate university that specializes in the three areas of information science, biological science, and materials science. Since its founding in 1991, the institution has been highly regarded in various fields for its world-class research activities and fostering of human resources. In order to enhance interdisciplinary education and learning of these three areas, NAIST was transformed into a single research institution called **Graduate School of Science and Technology** in 2018.

Yu Suzuki, Doctor of Engineering, Affiliate Associate Professor in the Division of Information Sciences, introduces the new system, "Seeking to realize more significant research results and contribute to the real society, a system to organically link researchers of the three fields was established. There is an increasing number of opportunities to work on a single research theme across all areas. The Division of Information Sciences studies methods to analyze test data more effectively together with the researchers of biological science and materials science, and leads the way to taking each study to a new stage based on the accumulated knowledge."

The **Augmented Human Communication Laboratory** to which Suzuki belongs handles an assortment of technologies that support communication among people and between people and IT. Beginning with the building of a prototype system that puts emphasis on feasibility to implementation into a service system and evaluation, the laboratory has also been responsible for collaboration within the industry. One of the initiatives that Dr. Suzuki and his research group are currently engaged in is leveraging multidimensional Big Data analysis to generate new findings.

"We aim to make practical use of the service to provide real-time distribution of tourism information for foreigners who visit Japan, leveraging multidimensional Big Data analysis. This service is intended to provide real-time information such as places of interest in famous tourist areas, weather, and crowd control, connecting sensor information placed in tourist areas with information posted on Twitter. The major aim of the service would be to closely monitor ever-changing circumstances and to provide appropriate information to tourists in real-time, rather than fixed information," says Dr. Suzuki.

A more proactive environment to utilize data mining, machine learning, and cloud sourcing was built in February 2018. The HPE ProLiant DL385 Gen10 Server equipped with an AMD EPYC processor was adopted.

Solutions

Adopted the HPE ProLiant DL385 Gen10 Server equipped with an AMD EPYC processor.

The HPE ProLiant DL385 Gen10 Server offers enhanced security on the hardware level and comes with an HPE Gen10 Server platform provided with silicon-level reliability (Silicon Root of Trust). Equipped with an AMD EPYC processor, it enables up to 64 cores and 4 TB of memory to be installed despite being a 2-socket/2U server.

"We handle a diverse assortment of lifestyle/social, biological, and material-related data that is used for analytics and data mining. We see great potential for multidimensional Big Data analysis that combines multiple types of data such as sensor data, social network data, and life log data," points out Dr. Suzuki.

> "You can also encrypt the physical memory for each virtual OS by changing the encryption key for each virtual OS. This process is referred to as Secure Encrypted Virtualization, or SEV. Even if SEV is used, impact on performance is held to about 1%."
>
> – Mr. Sekine of AMD

Multidimensional Big Data analysis has a wide scope of application, such as verifying the reliability of Wikipedia information, development of multimedia information search systems, development of behavioral history search systems/meta search engines, and quality control for cloud sourcing. Dr. Suzuki mentions one of the projects that he is currently involved with, "The HPE ProLiant DL385 Gen10 Server equipped with an AMD EPYC processor attracted our attention because it enables you to mount larger capacity memory than other servers of the same class. We also felt the architecture that connects the CPU and memory in the broadband was superior to that of other servers."

Masato Sekine, Manager Sales Engineering Enterprise Solution, Sales Division, AMD Japan Co. Ltd says, "The key feature of the AMD EPYC processor is the SoC (system-on-a-chip) that integrates the CPU core, memory channel, PCI Express, security chip, etc. Equipped with an 8-channel DDR4 Memory and 128-lane PCI Express, it directly connects the CPU with the memory at 170 GB/s without a chipset. You could also say it is equipped with the ideal capacity for the requirements of this project."

HPE ProLiant DL385 Gen10 Servers are carefully designed to offer high performance and low cost through a close cooperation between Hewlett Packard Enterprise and AMD at the development stage and are perfected as a product that changes the generally accepted perception of cost performance for 2-socket servers. The AMD EPYC processor has attained two world records for high performance—257 for SPECrate 2017 FP base and 270 for SPECrate 2017 FP peak— discrediting what has been commonly known about 2-socket servers now obsolete.

In order to take advantage of large capacity memories, it is important to maximize performance of the memory itself. This test once again confirmed the effectiveness of large capacity memories for building a classification model that combines

machine learning and cloud sourcing. However, a requirement for developing it into a service was also revealed.

"Each time learning is added, the memory consumed by the classifier is increased. Memories with larger capacity will be needed for providing the service in real-time. This will likely serve to further show the superiority of HPE ProLiant DL385 Gen10 servers that enable memories to be extended to a maximum of 4 TB," says Mr. Sekine.

The fact that HPE ProLiant DL385 Gen10 servers are equipped with a security feature "adds a significant amount of reassurance," as Dr. Suzuki points out, when selecting a server.

Build a classification model that combines machine learning and cloud sourcing
The project with Kyoto Prefecture included an attempt to identify and classify only tweets pertaining to Kyoto tourism from among more than 100,000 tweets.

"Tweets classified by supervised machine learning were assessed by human beings by means of cloud sourcing and imported as pertinent or not pertinent learning data in real-time. The precision of the classification is enhanced by applying it to the classifier over and again. This large capacity memory exhibits a powerful impact on building a classification model that combines machine learning and cloud sourcing. The superior memory architecture takes maximum advantage of the performance of the AMD EPYC processor," says Dr. Suzuki.

"There are more and more occasions where sensitive information is shared for joint projects with local governments and businesses, so it will be increasingly important for the server itself to be equipped with advanced security features."

HPE Gen10 Servers, which are the world's most secure industry standard servers,[1] ensure perfection of the system on the hardware level with HPE Integrated Lights-Out 5 (iLO 5) management processor as the Silicon Root

[1] Based on external firm conducting cybersecurity penetration testing of a range of server products from a range of manufacturers, May 2017.

"Large capacity memory exhibits powerful impact for building a classification model that combines machine learning and cloud sourcing."

– Dr. Yu Suzuki, PhD (Engineering), Affiliate Associate Professor, Division of Information Sciences, Graduate School of Science and Technology, Nara Institute of Science and Technology (NAIST), National University Corporation
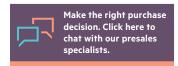
Solutions overview

Installation and deployment services

HPE ProLiant DL385 Gen10 Server equipped with an AMD EPYC processor along with HPE Integrated Lights Out 5

- AMD EPYC processor (max. 2 CPUs/ 64 cores)
- 8-channel DDR4 memory (max. 4 TB)
- 128-lane PCI Express
- Directly connects the CPU with memory at 170 GB/s without a chipset

Learn more at
hpe.com/us/en/servers/ gen10-servers.html

Make the right purchase decision. Click here to chat with our presales specialists.

✉ **Share now**

🖥 **Get updates**

of Trust. HPE ProLiant DL385 Gen10 servers add assurance of memory protection of the AMD EPYC processor to protection of firmware/hardware by iLO 5.

An AMD Secure processor is built into the AMD EPYC processor. The processor runs independently of the OS and can manage the encryption key. Secure Memory Encryption (SME) holds the overhead to approximately 1% by means of the encryption key and AES engine, and encrypts data in the memory by page table units so as to protect against attacks to the physical memory.

Joint project with Kyoto Prefecture for service to provide real-time distribution of tourism information for foreigners who visit Japan.

Cloud sourcing
Real-time learning by assessing information concerning food, transportation, weather, crowd control, scenery, and more (More than 5000 people in total).

Tweet classification that combines machine learning and cloud sourcing:

- Classification of tweets through supervised machine learning
- Assessment by more than 5000 people in total using cloud sourcing
- Learning in real-time, whether or not data is pertinent
- Precision of classification is enhanced by applying it to the classifier over and again

Benefits

HPE DirectPlus assisted Dr. Suzuki with his specifications enabling a smooth introduction in a short period of time.

Dr. Suzuki purchased his HPE ProLiant DL385 Gen10 Server using the HPE DirectPlus online store. With HPE DirectPlus, the user can

decide the configuration and order online. If preferred, a consultant can help you with your configuration by telephone or email. Dr. Suzuki related his experience as follows, "I first decided the memory capacity and then went into the details of the server configuration. I have used HPE DirectPlus from before, and this time again, it was quick; I discussed the specifications on the telephone, received a quote the following day, and was able to implement the server smoothly in a short period time."

With HPE DirectPlus, you receive a quote by the next business day and the configured server including CPU, memory, and disk is delivered in five business days from the official order (does not apply in certain areas). You can also take the quote and configuration ID provided by HPE DirectPlus and purchase at a sales outlet.

Dr. Suzuki says happily, "We received our HPE ProLiant DL385 Gen10 Server configured exactly to our specifications in less than two weeks after making the initial inquiry, including the time we spent thinking it over on campus. HPE DirectPlus facilitates placing orders and speeds up delivery, so it really helps procurements at busy times such as at fiscal year ends."

Dr. Suzuki and his research group are thinking of expanding the scope of applying the multidimensional Big Data analysis. Dr. Suzuki concludes, "Combining structured data managed by businesses and organizations with unstructured data from a completely different perspective should offer a wide-ranging potential for creating new knowledge. The approach of merging machine learning and cloud sourcing opens a new path to making it possible to categorize unstructured data while sequentially creating teaching data. I anticipate HPE Server products to achieve even higher computation power with great cost performance."