



CASE STUDY



Japan Science and Technology Agency selects Fusion ioMemory™ PCIe Cards to build knowledge infrastructure

Solution Focus

- Big Data
- Public sector agency

Summary of Benefits

- High-speed processing of a large amount of data
- Quicker response time to users
- Improved economics for setting up services

Product

- 2.4TB* Fusion ioMemory™ ioDrive®2 Duo® PCIe

Summary

Established in 1996, Japan Science and Technology Agency (JST) is a core national institution responsible for creating knowledge and sharing the outcome of research activities with society. The Agency's main activities include undertaking strategy research and development for creating innovation and building knowledge infrastructure. J-GLOBAL—a comprehensive portal for science and technology—supports the linking of diverse information, such as papers, patents, research projects and researchers. This service offered by JST enables creative thinking, the search for knowledge within different fields, and idea exploration activities of researchers. To develop an enhanced version of J-GLOBAL known as J-GLOBAL knowledge, JST selected the HP ProLiant DL980 G7 server and Fusion ioMemory PCIe application accelerators from SanDisk®.

Background

Japan Science and Technology Agency's (JST) goal is to realize an affluent and happy life for Japanese citizens through science and technology. JST has been acting as a bridge for transferring the outcomes of research activities to industry, and it has continuously worked on the initiatives for supporting the information and people required for pursuing research activities. Amid that, J-GLOBAL, which was officially launched in September 2012, has been extensively used by several researchers and institutions as the database of basic information related to ten categories of science and technology, including papers, patents, research projects, and researchers. For example a search on the organic compound "nitrobenzene" will return the names of 74 researchers, nearly 16,000 cases, and 790 patents.

Mr. Takahiro Kimura, Chief Supervisor of Knowledge Infrastructure and Mr. Katsuji Matsumura, Deputy Manager of Knowledge Infrastructure—both of the Office of Information Analysis, Department of Information Planning at the Japan Science and Technology Agency—worked on the further advancement of J-GLOBAL. Mr. Kimura explained the background of JST's knowledge infrastructure development. "At JST, we have collected the data of papers in Japan for more than 50 years. Even at present, we collect more than one million data items every year. The total number of papers that we have collected so far exceeds 36 million. We have converted this data into a database and built a system where searches can be performed. A part of this data has also been made available to researchers as J-GLOBAL. In 2011, to achieve further advancement of J-GLOBAL and to offer a system where more multifaceted searches can be performed, we started a project of developing J-GLOBAL knowledge."

“For high-speed processing of a large amount of data, large amount of memory and high-speed storage I/O were required. It was the combination of HP ProLiant DL980 G7 and Fusion ioMemory that met our requirements..”

**Mr. Takahiro Kimura, Chief Supervisor
In Charge of Knowledge Infrastructure,
Office of Information Analysis,
Department of Information Planning,
Japan Science and Technology**

The Challenge

J-GLOBAL knowledge converts the science and technology information accumulated by the JST knowledge infrastructure into the Linked Data RDF format compliant with the Semantic Web, so that multifaceted searches can be performed on the data. The Semantic Web can mechanically determine the meaning of a web page using metadata models such as the RDF format. In the RDF format, relational information of the target resource is expressed with three elements of subject, predicate, and object. Therefore, with respect to normal data, the amount of information that has to be stored and processed becomes massive.

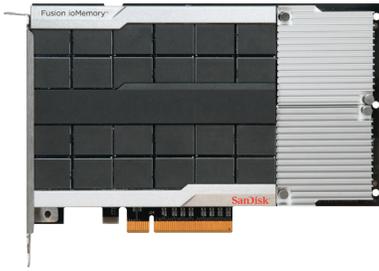
Mr. Matsumura told us about the significance of converting the data into the RDF format. “JST is an institution that contributes to increasing the international competitiveness of Japan as a leading science and technology country. Therefore, in accordance with the science and technology policies of the country, it is necessary to find information—with even higher accuracy—that has been derived from a large amount of data accumulated in the past and that may be useful in the future. Our initiative for the Semantic Web based on the RDF format is an important policy measure contributing to this.”

JST’s Office of Information Analysis, which was working on converting data to the RDF format for building J-GLOBAL knowledge, estimated the amount of data to be handled for developing the system. A specialized thinktank analyzed the requirements of the system needed for processing this data. “RDF creates data groups expressed with three elements known as a triple. Upon calculating the number of triples, it came to light that the number of triples would reach 15.5 billion,” explained Mr. Matsumura. “As of 2013, data published as open data all over the world is about 62 billion triples, which means that J-GLOBAL knowledge alone would have about one-fourth of this data. For high-speed processing of such a massive amount of data, large memory and high-speed storage I/O were essential.”

In 2011 when JST made the trial calculations for processing RDF data of 15.5 billion triples, the Agency required a system equipped with a few terabytes of memory that is also capable of high-speed reading and writing of data. While there were hardware solutions available that met this requirement, most of them were very expensive. “Although we were working on building J-GLOBAL knowledge, which would contribute to Japan’s national strategy, our budget was limited,” Mr. Kimura told us. “Therefore, the important point was to source a hardware solution that would meet the thinktank requirements and that offered excellent cost performance. The proposal that met our requirements was the combination of an HP ProLiant DL980 server and Fusion ioMemory PCIe cards.”

The Solution

The HP ProLiant DL980 G7 is a rack-mounted server of 8 sockets, 80 cores, and 160 threads equipped with the E7 family of Intel Xeon processors. It was supporting a large amount of memory and was also compatible with Fusion ioMemory PCIe cards—high-speed flash storage connected through PCI Expression—which would resolve bottlenecks of the storage I/O. Fusion ioMemory PCIe cards have read/write access latency of less than 19 microseconds, which is two to three digits lower compared to conventional storage. Its reading bandwidth reaches 2.7 gigabytes per second and its maximum writing speed is 2.2 gigabytes per second. Fusion ioMemory PCIe cards also offer maximum writing tolerance of 64 PBW (petabyte write amount) and a 10 to 20 uncorrectable bit-error rate (UBER), which is the lowest in the industry. The cards are also known for a long lifespan and high reliability, including a self-healing function. This performance and durability won the approval of JST and HP ProLiant DL980 G7 with a foolproof backup structure based on 7.2TB and three 2.4TB Fusion ioMemory ioDrive2 Duo PCIe cards. A 4-tiered disk configuration was used as the system foundation of J-GLOBAL knowledge.



2.4TB Fusion ioMemory ioDrive2 PCIe

The Result

Because of this new configuration leveraging the HP/Fusion ioMemory architecture, the entire procedure of data search has been streamlined and processing performance has been accelerated. The team has measured reads as fast as 2.8GB per second and writes as fast as 1.8GB per second.

“Starting in May 2015, we started publishing data of about 3 million items. There was some anxiety, but the performance has been greater than expected.” Focusing specifically on the knowledge of chemicals, Mr. Matsumura commented about the significance of J-GLOBAL knowledge. “Internally, the Office of Information Analysis will act as a pivot and work on data analysis using all data accumulated in J-GLOBAL knowledge. With the RDF format, we can relate data in a multifaceted manner from various angles in order to search and analyze data. Therefore, we would like to find scientific technologies and information that were hidden so far, and help many researchers and industries.”

In 2011, the Japanese government had formulated the 4th Science and Technology Basic Plan of Japan. The plan outlined a system of reform for pursuing science and technology innovation, strengthening the network of knowledge among government, industry, and academia and for the development of an architecture platform for collaboration. “The 4th Basic Plan set the national goal of building the knowledge infrastructure for information related to science and technology. Going forward, the 5th Basic Plan will mostly likely talk about developing the environment where this knowledge infrastructure can be more actively used,” said Mr. Kimura regarding future prospects. Even for responding to this national strategy, it is necessary to achieve high-speed processing of the massive amount of data accumulated in J-GLOBAL knowledge and offer it as knowledge for the country and the industry.”

Contact information

enterprise-sales-jp@sandisk.com

Western Digital Technologies, Inc.

951 SanDisk Drive
Milpitas, CA 95035-7933, USA
T: 1-800-578-6007

Western Digital Technologies, Inc. is the seller of record and licensee in the Americas of SanDisk® products.

SanDisk Europe, Middle East, Africa

Unit 100, Airside Business Park
Swords, County Dublin, Ireland
T: 1-800-578-6007

SanDisk Asia Pacific

Suite C, D, E, 23/F, No. 918 Middle
Huahai Road, Jiu Shi Renaissance Building
Shanghai, 20031, P.R. China
T: 1-800-578-6007

For more information, please visit:

<http://www.sandisk.co.jp/enterprise>

SanDisk®

a Western Digital brand

At SanDisk, we're expanding the possibilities of data storage. For more than 25 years, SanDisk's ideas have helped transform the industry, delivering next generation storage solutions for consumers and businesses around the globe.

* 1 TB = 1,000,000,000,000 bytes. Actual user capacity less.

The performance results and cost savings discussed herein are based on internal testing and use of Fusion ioMemory products. Results and performance may vary according to configurations and systems, including drive capacity, system architecture and applications.

©2016 Western Digital Corporation or its affiliates. All rights reserved. SanDisk is a trademark of Western Digital Corporation or its affiliates, registered in the United States and other Countries. Fusion ioMemory, ioDrive, and others are trademarks of Western Digital Corporation or its affiliates. Other brand names mentioned herein are for identification purposes only and may be the trademarks of their holder(s). JST_CS_SanDisk_v4 06/03/16 5020EN