

“As a persistent store, we estimate performance is ten times better on a node-per-node basis. If a parallel file system is a truck, this Fujitsu HPC platform is a race car. One is designed for capacity, the other for speed.”

Tiago Quintino
Senior Analyst
ECMWF

Co-creation project tackles I/O challenges

At a glance

Country: United Kingdom

Industry: Education

Founded: 2015

Website: nextgenio.eu

Challenge

Multiple HPC systems, that communicate simultaneously with slow disk storage, limit how fast the system can read and write data, creating bottlenecks. A solution was needed to eliminate these bottlenecks for frustrated customers.

Solution

NEXTGenIO was formed to explore a new approach to HPC and Fujitsu was invited to help develop the prototype. Fujitsu designed a 34-node computing cluster using Intel® Optane™ DCPMM. Each of the 34 compute nodes is equipped with two second-generation Intel® Xeon® Scalable Family processor CPUs and 3TB of Intel® Optane™ DC persistent memory.

Benefit

- Application performance has increased up to tenfold
- Customers' time to results for data-intensive workloads dramatically reduced
- Cost per capacity is lower than traditional DRAM deployments
- Increased memory capacity compared to DDR4 configurations, offering flexible usage models
- NEXTGenIO prototype technology is now market-ready in PRIMERGY and PRIMEQUEST models

Customer

NEXTGenIO started in 2015 by a consortium consisting of partners EPCC, Intel, Fujitsu, Technische Universität Dresden, Barcelona Supercomputing Center, the European Centre for Medium-Range Weather Forecasts, Arm (formerly Allinea) and Arctur. Its research has bridged the gap between memory and storage, using Intel's revolutionary Optane DC Persistent Memory, which is technically close to DRAM speed and significantly better than disk storage. NEXTGenIO has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement no. 671951.

Products and Services

- FUJITSU Intel® Optane™ DC persistent memory (DCPMM) modules for use in: PRIMERGY TX2550 M5, RX2530 M5, RX2540 M5, RX4770 M5, CX2560 M5, CX2550 M5, as well as the PRIMEQUEST PQ3800E2 and PQ3800B2

Eliminating I/O bottlenecks in HPC

High-performance computing (HPC) is a vital tool in industry and research, however, these advanced systems typically use data storage that is separate from the compute nodes. That means the I/O often struggles with multiple systems communicating simultaneously with slow disk storage so, although systems are capable of processing data quickly, speeds are limited by how fast the system can read and write data.

The Next Generation I/O for the Exascale project (NEXTGenIO), which runs under the European Union's Horizon 2020 Research and Innovation programme, aimed to resolve this I/O bottleneck and remove the roadblock to achieving Exascale computing, which is up to 1000 times faster than current Petascale systems. The consortium is led by EPCC, the supercomputing centre at the University of Edinburgh, and includes Fujitsu, the European Centre for Medium-Range Weather Forecasts (ECMWF) as well as commercial partner Arctur.

EPCC was investigating Intel's new DC persistent memory modules (DCPMM) to tackle bottlenecks and wanted to gather a consortium of partners with broad use cases and technologies. Fujitsu, with its broad experience in user centric and innovative data centre technologies, was invited to lead the definition of the NEXTGenIO hardware architecture and implement the prototype system.

Fellow member the European Centre for Medium-Range Weather Forecasts (ECMWF) runs critical weather forecasts four times each day for the whole globe – an incredibly complex workflow with literally tens of thousands of HPC tasks running concurrently, which presents many challenges. Every forecast consists of around 20TB of data and 70% of that data is read within minutes of being written so it puts high stress on the I/O systems. This new approach promised to provide a better way of eliminating those pain points.

Arctur is a small company, which provides HPC resources and consulting for clients, such as Pipistrel, which designs and produces ultra-light aircraft. Typically, modelling two seconds of simulated time can take up to three days on its existing HPC platform, so the company was eager to see how this radical approach could speed the process through the elimination of bottlenecks.

Flexible co-creation

Fujitsu designed, manufactured and validated a prototype 34-node computing cluster to close the gap between memory and storage by leveraging Intel® Optane™ DCPMM. Each of the 34 compute nodes is equipped with two second-generation Intel® Xeon® Scalable Family processor CPUs and 3TB of Intel® Optane™ DC persistent memory and includes a software stack to seamlessly support I/O and memory intensive workloads.



EPCC, Arctur and ECMWF were involved from the design stage, providing requirements and then co-creating the new solution with Fujitsu and the other partners, which all had their own use cases to bring to the table. For many consortium members, it was the first time being so closely involved with hardware development and it was essential that Fujitsu listened to their needs and collaborated openly.

"Fujitsu took account of the specific use cases and used them to design a requirement-driven, flexible platform that can run in either 'fast storage' or 'memory' modes," adds Dr. Michèle Weiland, Senior Research Fellow at EPCC. "The I/O performance on the nodes is really impressive because writing data locally means there is no slow down."

Faster performance at lower costs

The new Intel DCPMM technology, first used in the NEXTGenIO project, is now available in a range of Fujitsu PRIMERGY and PRIMEQUEST servers, and speeds customer's time to results for data-intensive workloads while having a lower cost than traditional DRAM deployments. This makes it ideal for HPC environments with heavy data demands, such as ECMWF and Arctur.

"We handle hypercubes of data in six dimensions, which are expected to grow from 20TB to 100TB in a few years. What this NEXTGenIO project has enabled us to do is store a whole day's worth of forecasts and analyse it in any direction without hitting I/O bottlenecks," remarks Dr. Tiago Quintino, Senior Analyst, ECMWF. "It's a gamechanger that frees our scientists from the constraints of data access. As a persistent store, we estimate performance is ten times better on a node-per-node basis. If a parallel file system is a truck, this Fujitsu HPC platform is a race car. One is designed for capacity, the other for speed."

"We loved the ease of use with the Fujitsu hardware – there was no need to read a manual or have training because there is no change to the underlying code and no special compilations," says Tomislav Subic, HPC Research Engineer, Arctur. "And it has dramatically reduced our development cycle from three days to less than one. The Fujitsu platform is at least twice as fast as more traditional methods."

"It has been a genuine pleasure to partner with Fujitsu – the system it delivered was so professionally assembled that we couldn't tell it was a prototype," concludes Weiland. "The entire experience has been tremendously educational and has fulfilled our vision for a new approach to HPC."

FUJITSU

Email: AskFujitsu@uk.fujitsu.com

Tel: +44 (0)1235 797711

IN COLLABORATION WITH



© 2019 Fujitsu and the Fujitsu logo are trademarks or registered trademarks of Fujitsu Limited in Japan and other countries. Other company, product and service names may be trademarks or registered trademarks of their respective owners. Technical data subject to modification and delivery subject to availability. Any liability that the data and illustrations are complete, actual or correct is excluded. Designations may be trademarks and/or copyrights of the respective manufacturer, the use of which by third parties for their own purposes may infringe the rights of such owner.