Case Study

High Performance Computing (HPC)
2nd Gen Intel® Xeon® Scalable Processors
Intel® Optane™ DC Persistent Memory

NEXTGenIO Research Cluster Helps Solve Exascale I/O Challenges at Edinburgh Parallel Computing Center (EPCC)

EPCC Uses Intel® Optane™ Persistent Memory with 2nd Gen Intel® Xeon® Scalable Processors to Accelerate HPC Workloads and Improve Cluster Efficiencies

NEXTGenIO research cluster highlights

- Intel® Xeon® Platinum 8260 processors in a cluster of 34 NEXTGenIO prototype nodes developed and manufactured by Fujitsu
- 3 TBs Intel® Optane™ persistent memory per node
- Cornelius Network fabric for compute nodes
- Supports internationally funded, multi-partner NEXTGenIO project

Executive Summary

EPCC is a hotbed for scientific research using High Performance Computing (HPC). The center has gained an international reputation for the most advanced capability in all aspects of HPC, High Performance Data Analytics (HPDA), and novel computing. EPCC hosts a range of leading HPC systems, including Tesseract, the United Kingdom's Extreme Scaling service of the DiRAC (Distributed Research utilizing Advanced Computing) facility. Tesseract is a 20,000-plus core hypercube-based supercomputer built on Intel® Xeon® Scalable processors and Cornelis Network's fabric. But among its other resources is a smaller, yet as important, cluster that enabled breakthrough research on the NEXTGenIO project. Researchers use this cluster, built by Fujitsu on 2nd Generation Intel Xeon Scalable processors and Intel® Optane™ persistent memory, to modify and optimize codes that can deliver the benefits of Intel Optane persistent memory for high-throughput parallel workloads.

Challenge

To achieve Exascale computing requires addressing many challenges, including those with HPC I/O, which has lagged behind other HPC advances. EPCC led the NEXTGenIO project, funded by the European Commission, with research into how to leverage byte-addressable persistent memory (B-APM) for large parallel computing workloads. Partners in NEXTGenIO include EPCC, Intel, Fujitsu, Barcelona Supercomputing Center (BSC), Technische Universitat Dresden, ARM/Allinea, European Center for Medium-range Weather Forecasting (ECMWF), and ARCTUR.

Intel Optane persistent memory along with 2nd Generation Intel Xeon Scalable processors deliver a flexible B-APM architecture for servers using Intel Optane persistent memory.

For the NEXTGenIO project, work at EPCC focused on the overall technical and architectural needs to maximize the potential of Intel Optane persistent memory in its several different usage modes. Working with industry and academia across Europe, researchers implemented new filesystems, designed and developed data-aware schedulers, investigated check-pointing software, and integrated I/O and communication libraries for the necessary modifications and optimizations to software that can benefit from Intel Optane persistent memory.

Solution

Exploring new memory hierarchies for HPC architectures in order to accelerate performance and increase system efficiencies has been a focus of supercomputing for many years. From burst buffers to SSDs in each node for local data storage, and now B-APM, system architects are moving faster, more efficient technologies closer to the CPUs.

Intel Optane persistent memory combines the traits of storage and memory into a single high-capacity module that fits into a server DRAM slot. Most 2nd Generation Intel Xeon Scalable processors recognize this technology, providing near-DRAM performance with up to 3 TB of capacity per socket. Intel Optane persistent memory offers different usage modes to optimize the technology for various types of workloads, whether they need massive volatile memory capacity, persistent data storage with near-DRAM performance, or a combination of both.

EPCC's new research cluster built by Fujitsu provided computing resources for NEXTGenIO research. In an intensive co-creation process, Fujitsu analyzed NEXTGenIO partners' application I/O and compute requirements and designed and manufactured the NEXTGenIO system focused to overcome existing bottlenecks. The cluster houses 34 nodes of dual-socket 2nd Generation Intel Xeon Platinum 8260 processors, 3 TBs of Intel Optane persistent memory, and 192 GB of DRAM per node. A 100 Gbps Cornelis Network fabric connects the compute nodes, while a 56 Gbps InfiniBand network attaches to an external Lustre parallel filesystem. The architecture allowed researchers to learn how to take advantage of high-capacity persistent memory across many different large parallel codes, such as OpenFOAM, a computational fluid dynamics (CFD) code used in the ECMWF's IFS forecasting software.

Intel Optane persistent memory along with 2nd Generation Intel Xeon Scalable processors deliver a flexible B-APM architecture for servers.

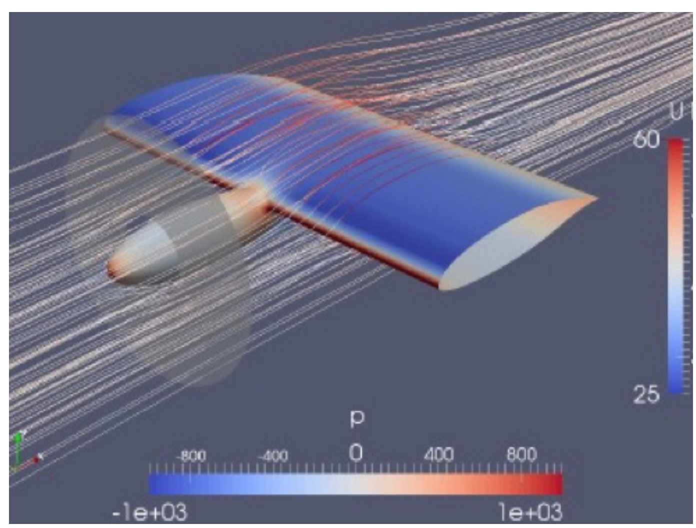
"We are glad Fujitsu could contribute to the success of the NEXTGenIO project," said Olivier Delachapelle, Fujitsu Head of Category Management-Products Sales Europe. "Solving the I/O bottleneck opens the door to significantly increased performance and scalability towards HPC Exascale. Fujitsu successfully integrated the DCPMM technology to our PRIMERGY and PRIMEQUEST product lines. This breakthrough in I/O performance therefore will also reduce time to results for many of our customers applications in a broad range of business sectors."

Results

EPCC's new Intel Optane persistent memory cluster delivered promising benefits for the NEXTGenIO project. Benchmarks and large-scale applications were used to measure performance of the cluster with Intel Optane persistent memory.

Working with the ECMWF, project partners were able develop novel software to take advantage of the new cluster's technologies. The ECMWF ensemble forecasting software runs 56 forecasts every day to provide nine to 16-day weather forecasts, which are made available to member organizations. The software was modified to use Intel Optane persistent memory in app-direct mode, leveraging the technology as a persistent data store.

"With an OpenFOAM code with high amounts of I/O," explained Adrian Jackson, Senior Research Fellow at EPCC, "we used Intel Optane persistent memory as a filesystem on the node. We used some of our tools to get data into and out of the node and used the persistent memory as the data store.



Using Intel Optane Persistent Memory, an OpenFOAM code with high amounts of I/O ran faster compared to a Lustre filesystem.[1]

The solver ran over 8x faster—12 percent of the original run time—compared to going to the Lustre filesystem for data."[1]

Other codes took advantage of the system's two-level memory architecture, using the large memory capacity of Intel Optane persistent memory for data and the DRAM as cache (memory mode).

In another study, while a large-scale simulation required 20 nodes or more to run without persistent memory, the same code ran on a single node with Intel Optane persistent memory, according to Jackson. "While the simulation did not run faster, it ran more efficiently on a single node," added Jackson.[1] Such consolidation potentially promises benefits for scaling out large applications on smaller clusters.

Several other studies were completed, including synthetic workflows and IOR benchmarks. The work is described in a University of Edinburgh Research Explorer report.2 The work is ongoing to further optimize I/O for large-scale codes on HPC machines with Intel Optane persistent memory and 2nd Generation Intel Xeon Scalable processors.

Solution Summary

- Intel Xeon Platinum 8260 processors in a cluster of 34 NEXTGenIO prototype nodes developed and manufactured by Fujitsu
- Intel Optane persistent memory (3 TBs/node)

Where to Get More Information

Learn more about Intel Xeon Scalable Processors.

Get details on Intel Optane Persistent Memory.

Learn more about the EPCC.

Learn more about Fujitsu's NextGENIO system.

[1] Results provided by EPCC
[2] https://www.research.ed.ac.uk/portal/files/104519942/BAPM_HPC_IODC19.pdf

052022/RJMJ/RL/PDF ♻ Please Recycle 341599-002US