Home » Feed aggregator » Sources » HPC Wire

HPC Wire

Since 1987 - Covering the Fastest Computers in the World and the People Who Run Them
URL: https://www.hpcwire.com
Updated: 11 hours 1 min ago

Knights Landing Processor with Omni-Path Makes Cloud Debut

Tue, 04/18/2017 - 10:15

HPC cloud specialist Rescale is partnering with Intel and HPC resource provider R Systems to offer first-ever cloud access to Xeon Phi "Knights Landing" processors. The infrastructure is based on the 68-core Intel Knights Landing processor with integrated Omni-Path fabric (the 7250F Xeon Phi).

"It is a three-way joint initiative," Tyler Smith, head of partnerships at Rescale told *HPCwire*. R Systems is hosting and managing the Intel technology at its HPC datacenters in Champaign, Illinois, with access to the cluster available through Rescale's "big compute" ScaleX Pro platform.

R Systems Principal Brian Kucic said his company is committed to offering advanced, bare metal technologies to the HPC research community through its partnership with Intel and Rescale. At Intel, General Manager of the Accelerated Workgroup Barry Davis emphasized that the partnership is intended to support increased adoption of HPC applications to drive discovery and innovation.

Starting today, the Xeon Phi cloud is being made available at no cost for four weeks on a first-come-first-serve basis. Via a single queue, end users can submit one job at a time for up to six single-socket Knights Landing nodes. Running follow-up jobs will entail getting back in the queue. Rescale has plans to offer a premium, priority queue that will allow users to run multiple jobs simultaneously. Initially, however, Smith said the effort should be viewed as a test lab that brings next-generation computation power to the cloud.

After the complimentary period expires, the cost will be $0.03 / core / hour, so for the 68-core platform, the price will be about $2 per CPU.

While the new Intel gear is initially targeting dev and test environments, at least two Rescale customers have expressed interest in running larger workloads, according to the company. "We can reserve a certain number of nodes for a specific customer to do a POC if they'd like but I think that's all going to come with time," said Smith.

The KNL cloud offering is part of new effort called ScaleX Labs, and Rescale expects this one to be the first of many, anticipating future opportunities both with Intel and with other hardware providers. "The [larger] initiative is reflective of what we've done [here] with Intel, to provide customers with access to pre-release or early-release hardware or be the first to offer a product in the cloud," Smith explained. "Depending on the hardware partner, the lab can be used in a few different ways: it can hopefully create demand from software partners, to optimize their code, or create demand from the public cloud provider. And it allows Rescale to provide a platform in which our customers can test the latest and greatest hardware."

Rescale offers the cloud labs through its ScaleX Pro platform, which features an intuitive GUI for job management and collaboration and secure SSL data transfer and encryption at rest, according to the company. ScaleX Labs users will also receive beta access to ScaleX Developer, Rescale's product that allows software developers to integrate and deploy their own software on the ScaleX platform. ScaleX Developer is currently on track for general availability in Q3.

"I think one thing that attracted Intel to partner with Rescale on this initiative was our ability to enable developers to have beta access to our ScaleX Developer product so they can deploy their applications directly to the ScaleX platform and then end users could bring proprietary code in a bring-your-own-software model and run it on the KNL," said Smith.

A single-socket 68-core Knights Landing CPU running at 1.60 GHz, with two AVX-512 vector units (512-bit) per core, provides around 3 teraflops of double-precision (peak) performance. The bootable Knights Landing chips, based on the second generation of Intel's Many Integrated Core (MIC) architecture, are largely competing against Nvidia's Pascal-generation P100 GPUs released last year.

Rescale doesn't own or operate any of its infrastructure. Instead the San Francisco company partners with cloud providers, both major IaaS purveyors and others that are more regional in nature. And with ScaleX Labs, Rescale also has a path to work directly with hardware suppliers.

"Once we do the API integration, that's where our secret sauce is, where we can quickly add our software application portfolio to that cloud provider and within a very short period of time make those software applications available in that cloud provider," said Smith.

He also emphasized the advantages of a multi-cloud strategy. "[The only way to have] access to an "infinite" amount of resources, [is to] partner with multiple cloud providers. Google had access to the Skylake processor first; AWS came out with a bigger GPU offering. If you're locked into one public cloud provider, you're limited. It's a competitive world out there and folks are worried about time-to-market and performance is highly sensitive. It's also about scalability. If you're ever bound or constrained by computing you're not really a cloud provider; it kind of goes against the ethos of it, at least in my opinion."

Although Rescale supports 220 software applications, the ScaleX Labs with Intel Knights Landing will launch with only a few software packages at the outset. The aim is to have it grow organically through customer demand. "We want to offer it within the parameters that Intel has set out to target applications that have been optimized for KNLs and will see a performance boost or will see a performance boost from the Omni-Path fabric," said Smith. "There's also a lot of proprietary code out there and this will be a way for end users to test out how it performs on KNL. We'll expect to grow the software as more commercial software vendors support KNL, which is coming, and secondly, as other customers request it, which our sales guys are already starting to see happen."

AI and deep learning frameworks won't be there at launch, but according to Smith, a proactive effort is underway to build up this support over the next couple of months. Given Intel's AI ambitions and proclamations for Knights Landing's machine learning chops, it makes sense that this would be a priority.

The post Knights Landing Processor with Omni-Path Makes Cloud Debut appeared first on HPCwire.

PRACE 14th Call for Proposals Awards Nearly 2000M Core Hours
Tue, 04/18/2017 - 08:53

April 18, 2017 — The 14th PRACE Call for Proposals yielded 113 eligible proposals of which 59 were awarded a total of close to 2 thousand million core hours. This brings the total number of projects awarded by PRACE to 524. Taking into account the 3 multi-year projects from the 12th Call that were renewed and the 10 million core hours reserved for Centres of Excellence, the total amount of core hours awarded by PRACE to more than 14 thousand million.

The 59 newly awarded projects are led by principal investigators from 15 different European countries. In addition, two projects are led by PIs from New Zealand and the USA.

This Call was the first under the recently ratified PRACE 2 Programme (http://www.prace-ri.eu/prace2-council-ratification/). As a new feature of this programme, 3 project proposals are tagged Candidate Tier-0 and will receive tailored assistance from a High-Level Support Team (HLST) to push their excellent research towards more advanced use of world-class high performance computers.

Seven scientific domains are represented: 8 projects are linked to the fields of Biochemistry, Bioinformatics and Life Sciences; 22 to Chemistry; 4 to Earth System Sciences; 4 to Engineering; 9 to Fundamental Constituents of Matter; and 12 to Universe Sciences.

The 14th PRACE Call for the first time included the Piz Daint system from CSCS, Switzerland, PRACE's newest Hosting Member. 8 projects were awarded a total of close to 6 million node hours (401 million core hours) on this system. Hosting Members BSC (Spain) and CINECA (Italy) have recently upgraded their systems (MareNostrum and Marconi respectively) and these are now also available for PRACE Calls.

Amongst the renewed multi-year projects from Call 12, Charge and Spin Hall Kubo Conductivity by Order N Real Space Methods led by Dr. Stephan Roche is linked to the FET Graphene Flagship program. The project received 20 million core hours on MareNostrum @ BSC, Spain.

One PRACE-awarded project is linked to the FET Human Brain Flagship program:CBNR – CereBellar Network Reconstruction led by Prof. Egidio D'Angelo. The project received 32 million core hours on Juqueen @ GCS @ JSC, Germany.

One PRACE-awarded project is linked to the Autism Research and Technology Initiative: iHART – Characterization of genetic risk variants in ASD families using a reference-free approach led by Dr. Daniel Geschwind. Taking into account its potential societal impact, it was awarded 5.5 million core hours on MareNostrum @ BSC, Spain.

Two projects are led by PIs from industry: EDF (France, energy-related R&D in the field of materials) and Cenaero (Belgium, public/private R&D centre in aeronautics/combustion), and another involves a Co-PI from industry: Termo Fluids S.L. (Spain, SME in fluid mechanics).

The PRACE 14th Call awarded resources to 9 ERC and 2 Marie Sklodowska Curie funded projects, 2 H2020 funded projects, 2 EC FET Flagship programmes, 2 projects with links to the MaX Centre of Excellence and 2 projects with links to EUROfusion and the European Strategic Energy Technology Plan (https://ec.europa.eu/energy/en/topics/technology-and-innovation/strategic-energy-technology-plan).

All information and the abstracts of the projects awarded under the 14th PRACE Call for Proposals can be found here: http://www.prace-ri.eu/14th-project-call/ (going live soon).

The **14th Call** for Proposals for PRACE Project Access (Tier-0) was open from 10 October until 21 November 2016. Selected proposals will receive allocations to PRACE resources from 3 April 2017 to 31 March 2018.

About PRACE

The Partnership for Advanced Computing in Europe (PRACE) is an international non-profit association with its seat in Brussels. The PRACE Research Infrastructure provides a persistent world-class high performance computing service for scientists and researchers from academia and industry in Europe.

The computer systems and their operations accessible through PRACE are provided by 5 PRACE members (BSC representing Spain, CINECA representing Italy, CSCS representing Switzerland, GCS representing Germany and GENCI representing France). The Implementation Phase of PRACE receives funding from the EU's Seventh Framework Programme (FP7/2007-2013) under grant agreement RI-312763 and from the EU's Horizon 2020 research and innovation programme (2014-2020) under grant agreements 653838 and 730913. For more information, see www.prace-ri.eu.

Source: PRACE

The post PRACE 14th Call for Proposals Awards Nearly 2000M Core Hours appeared first on HPCwire.

Mellanox 25Gb/s Ethernet Adapters Chosen By Major ODMs
Tue, 04/18/2017 - 08:08

SUNNYVALE, Calif. & YOKNEAM, Israel, April 18, 2017 — Mellanox Technologies, Ltd. (NASDAQ: MLNX), a supplier of high-performance, end-to-end interconnect solutions for data center servers and storage systems, today announced that the ConnectX-4 Lx 25Gb/s OCP and PCIe Ethernet adapters have been adopted by major ODMs. The ConnectX-4 Lx 25 gigabit Ethernet adapters provide 2.5 times the data throughput at lowest latency needed for data center applications while utilizing the same infrastructure as 10 gigabit Ethernet, thus maximizing the data center return on investment. The company is currently shipping hundreds of thousands of Ethernet adapters every quarter, reflecting a growing demand for Mellanox Ethernet solutions.

"We are proud to see our 25GbE solutions being deployed with our partner ODMs' top-class servers, enabling the most cost-effective Ethernet connectivity for Hyperscale and cloud infrastructures," said Gilad Shainer, vice president of marketing at Mellanox Technologies. "The state-of-the-art 25GbE speed provides superior price/performance advantages compared to 10GbE fabric, maximizing the data center return on investment."

Since 2016, Wiwynn Corporation, a leading cloud infrastructure provider of high quality computing and storage products, has shipped its OCP server SV7221G2 product family with the Mellanox 25GbE ConnectX-4 Lx OCP Mezzanine NICs and PCIe cards to major Internet service providers.

"With Mellanox ConnectX-4 Lx cards, our OCP server SV7221G2 product family can provide the advanced 25GbE features our customers need," said Steven Lu, chief of product marketing at Wiwynn. "We see the market evolving into the 25Gb/s era and we are proud to be at the leading edge, helping to advance the market."

"Inventec has qualified ConnectX-4 Lx 25GbE cards for TB800G4, Balder and K800G3 platforms to be supplied to major Cloud and Web2.0 providers in China," said Evan Chien, China business line director, Inventec. "We are pleased to adopt the Mellanox ConnectX-4 Lx 25GbE card to provide hyperscale data center customers with the best fit for advanced applications and performance."

Acer Inc., a Taiwanese multinational hardware and electronics corporation specializing in advanced electronics technology has also qualified ConnectX-4 Lx PCIe adapters and will soon offer their servers Altos R380 F3, R360 F3 and AW2000h F3, to the market.

Since 2016, Mitac-TYAN has been shipping ConnectX-3 Pro 40GbE OCP mezzanine cards and recently added the ConnectX-4 Lx 25GbE OCP mezzanine cards to its GT86A-B7083 server offering.

"TYAN has been successfully shipping the ConnectX-3 Pro 40GbE OCP mezzanine cards and now is proud to add the 25GbE ConnectX-4 Lx OCP cards to our servers' offerings," said Mr. Danny Hsu, vice president of MiTAC Computing Technology Corporation's TYAN Business Unit. "We see increasing demand for 25GbE in the market, and working with Mellanox to deliver state-of-the-art network cards that benefit our customer's deployments tremendously."

ConnectX-4 Lx, the industry's most efficient 10, 25, 40, 50Gb/s Ethernet intelligent adapter, enables datacenters to migrate from 10G to 25G and from 40G to 50G speeds with similar power consumption, cost, and infrastructure needs. Together with RDMA over Converged Ethernet (RoCE), ConnectX-4 Lx dramatically improves storage and compute platform efficiency. With ConnectX-4 Lx, IT and applications managers can leverage greater data speeds of 25G and 50G to handle the growing demands for data analytics today.

About Mellanox

Mellanox Technologies (NASDAQ: MLNX) is a leading supplier of end-to-end Ethernet and InfiniBand intelligent interconnect solutions and services for servers, storage, and hyper-converged infrastructure. Mellanox intelligent interconnect solutions increase data center efficiency by providing the highest throughput and lowest latency, delivering data faster to applications and unlocking system performance. Mellanox offers a choice of high performance solutions: network and multicore processors, network adapters, switches, cables, software and silicon, that accelerate application runtime and maximize business results for a wide range of markets including high performance computing, enterprise data centers, Web 2.0, cloud, storage, network security, telecom and financial services. More information is available at: www.mellanox.com.

Source: Mellanox

The post Mellanox 25Gb/s Ethernet Adapters Chosen By Major ODMs appeared first on HPCwire.

OpenPOWER Foundation Announces Developer Congress focused on AI
Mon, 04/17/2017 - 21:03

SAN FRANCISCO, Calif., April 17, 2017 — On the wave of strong momentum around machine learning and AI in 2017, the OpenPOWER Foundation will put these innovative technologies center stage at the upcoming OpenPOWER Foundation Developer Congress, May 22-25, at the Palace Hotel in San Francisco. The conference will focus on continuing to foster the collaboration within the foundation to satisfy the performance demands of today's computing market.

Developers will have the opportunity to learn and gain first-hand insights from the creators of some of the most advanced technology currently driving Deep Learning, AI and Machine Learning. Key themes will include:

- Deep Learning, Machine Learning and Artificial Intelligence through GPU Acceleration and OpenACC. Learn the latest techniques on how to design, train and deploy neural network-powered machine learning in your applications.
- Deploy a fully optimized and supported platform for machine learning with IBM's PowerAI that supports the most popular machine learning frameworks — Anaconda, Caffe, Chainer, TensorFlow and Theano.
- Custom Acceleration for AI through FPGAs
- Databases & Data Analytics
- Porting, Optimization, Developer Tools and Techniques
- Firmware & OpenBMC

The Developer Congress is supported by the newly formed OpenPOWER Machine Learning Work Group (OPMLWG), an addition to the OpenPOWER Foundation community. The new group — which includes Canonical, Cineca, Google and Mellanox, among others — provides a forum for collaboration that will help define frameworks for the productive development and deployment of machine learning solutions using the IBM POWER architecture and OpenPOWER ecosystem technology.

As part of the ecosystem, the OPMLWG plays a crucial role in expanding the OpenPOWER mission. It focuses on addressing the challenges machine learning project developers are continuously facing by identifying use cases, defining requirements and extracting workflows, to better understand processes with similar needs and pain points. The working group will also identify and develop technologies for the effective execution on machine learning applications by enabling hardware (HW), software (SW) and acceleration across the OpenPOWER ecosystem.

The OPMLWG group and Developer Congress come soon after the OpenPOWER Foundation surpassed a 300-member milestone, with large players joining the fold that have developed new processes and technologies based on the OpenPOWER architecture. Some recent additions include:

- Red Hat, which joined as a Platinum member and part of the board, adding open source leadership and expertise around community driven software innovation
- Kinetica, offers a high-performance analytics database that harnesses the power of GPUs for unprecedented performance to ingest, explore and visualize data in motion and at rest
- Bitfusion, leaders in end to end application lifecycle management and developer automation for deep learning, AI and GPUs.
- MapD Technologies, which offers a fast database and visual analytics platform that leverages the parallel processing power of GPUs

"Open standards are a critical component of modern enterprise IT, and for OpenPOWER having a common set of guidelines for integration, implementation and enhanced IT security are key," said Scott Herold, senior manager, Multi-Architecture product strategy Red Hat. "Red Hat is a strong proponent of open standards across the technology stack and we are pleased to work with the OpenPOWER Foundation's various work groups in driving these standards to further enterprise choice as it relates to computing architecture."

All OpenPOWER Members can join and work on:

- Collection and description of use cases
- Porting, tuning and optimization of important Open Source Library / Frameworks
- Creating a ML/DL Sandbox for quick start, including example use cases, data sets and tools
- Recommending platform features for machine learning

"OpenPOWER was founded with the goal of granting the marketplace more technology choice and the ability to rethink the approach to data centers. Today, we see the growing application of machine-learning and cognitive technology, the OpenPOWER foundation is actively supporting technical initiatives and solution development in these areas to help drive innovation and industry growth," said John Zannos, Chairman of The OpenPOWER Foundation. "The Machine Learning Work Group will focus on addressing this need for innovation, allowing technology developers and users to collaborate as they search for the solutions to the computational challenges being posed by machine learning and artificial intelligence."

About The OpenPOWER Foundation

OpenPOWER Foundation was founded in 2013 as an open technical membership organization enabling data centers to rethink their approach to technology. Member companies are empowered to customize POWER CPU processors and system platforms for optimization and innovation for their business needs. At the heart of the efforts, are member offerings and solutions that can further OpenPOWER adoption, developer community engagement and a continuous effort to foster innovation in and outside the data center.

OpenPOWER members are actively pursuing innovation and all are welcome to join in moving the state of the art of OpenPOWER systems design forward. Learn more through the OpenPOWER Intro Video and read more about OpenPOWER Ready products here.

Source: OpenPOWER Foundation

The post OpenPOWER Foundation Announces Developer Congress focused on AI appeared first on HPCwire.

IBM Coils Anaconda Around Power Processor
Mon, 04/17/2017 - 12:40

IBM, which recently extended support for the Anaconda data science platform to its open source mainframe, takes another step this week by offering the data platform on its Cognitive Systems platform in collaboration with Anaconda developer Continuum Analytics.

IBM also announced on Monday (April 17) the formation of a machine learning work group within the OpenPOWER Foundation. The new group, which includes Google, will help define machine-learning frameworks for both OpenPOWER and IBM's Power architecture.

Meanwhile, the closer coupling of Anaconda with IBM's cognitive platform will include integration with IBM's PowerAI software for machine and deep learning. The combination is touted as leveraging the IBM processor along with GPU acceleration for cognitive workloads. The company also noted that the collaboration with Continuum Analytics would help data scientists and developers to deploy and scale deep learning applications.

Among the goals of the collaboration is moving Anaconda deeper into enterprises and spurring adoption of machine and deep learning frameworks used to develop cognitive applications. IBM also is using the collaboration to promote its PowerAI software as a way for enterprises to deploy open source frameworks based on its Power architecture. Those frameworks can be "tuned for high performance," IBM noted, allowing the cognitive platform to handle commercial as well as hyper-scale workloads.

Optimizing Anaconda on the IBM architecture also gives developers access to libraries in the PowerAI platform for deploying the enterprise version of Anaconda, Travis Oliphant, co-founder and chief data scientist at Continuum Analytics, noted in a statement.

The cognitive platform is based on IBM's POWER8 architecture that includes a high-speed interface to Nvidia's Tesla Pascal P100 GPU accelerators. The high-bandwidth chip connections are designed to boost the performance of predictive analytics and deep learning applications, IBM said.

In February, IBM announced it was working with Anaconda developer Continuum Analytics, Austin, Texas, and Rocket Software, Waltham, Mass., to host the open source analytics platform on IBM z/OS mainframes. Last year, IBM and several partners collaborated to bring Apache Spark to it z System mainframe, allowing the open-source analytics framework to run natively on the company's mainframe operating system.

In announcing formation of the machine learning work group, IBM noted a roster of new OpenPOWER Foundation members, including Red Hat (NYSE: RHT), database specialist Kinetica, application management and developer automation vendor Bitfusion along with MapD Technologies. The database and analytics vendors are all using GPUs deep learning, AI and other analytics applications and platforms.

IBM's broader commitment to Anaconda illustrates how enterprise technology vendors are embracing open source platforms as they seek to move processing power closer to data sources. IBM is the latest tech giant to embrace the Anaconda stack. Continuum Analytics said last year that Anaconda also supports Intel Corp.'s Math Kernel Library and Microsoft Corp.'s R Open for statistical analysis.

The post IBM Coils Anaconda Around Power Processor appeared first on HPCwire.

Baidu Advances AI in the Cloud with Latest NVIDIA Pascal GPUs
Mon, 04/17/2017 - 09:06

SANTA CLARA, Calif., April 17, 2017 — NVIDIA (NASDAQ: NVDA) today announced that its deep learning platform is now available as part of Baidu Cloud's deep learning service, giving enterprise customers instant access to the world's most adopted AI tools.

The new Baidu Cloud offers the latest GPU computing technology, including Pascal architecture-based NVIDIA Tesla P40 GPUs and NVIDIA deep learning software. It provides both training and inference acceleration for open-source deep learning frameworks, such as TensorFlow and PaddlePaddle.

"Baidu and NVIDIA are long-time partners in advancing the state of the art in AI," said Ian Buck, general manager of Accelerated Computing at NVIDIA. "Baidu understands that enterprises need GPU computing to process the massive volumes of data needed for deep learning. Through Baidu Cloud, companies can quickly convert data into insights that lead to breakthrough products and services."

"Our partnership with NVIDIA has long provided Baidu with a competitive advantage," said Shiming Yin, vice president and general manager of Baidu Cloud Computing. "Baidu Cloud Service powered by NVIDIA's deep learning software and Pascal GPUs will help our customers accelerate their deep learning training and inference, resulting in faster time to market for a new generation of intelligent products and applications."

World's Most Adopted AI Platform

NVIDIA's deep learning platform is the world's most adopted platform for building AI services. All key deep learning frameworks are accelerated on NVIDIA's platform, which is available from leading cloud service providers worldwide, including Alibaba, Amazon, Google, IBM and Microsoft. Organizations ranging from startups to leading multinationals are taking advantage of GPUs in the cloud to achieve faster results without massive capital expenditures or complexity of managing the infrastructure.

Organizations are increasingly turning to GPU computing to develop advanced applications in areas such as natural language processing, traffic analysis, intelligent customer service, personalized recommendations and understanding video.

The massively efficient parallel processing capabilities of GPUs make NVIDIA's platform highly effective at accelerating a host of other data-intensive workloads, from AI and deep learning to advanced analytics to high performance computing.

Baidu Cloud's deep learning service is available today.

About NVIDIA

NVIDIA's (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined modern computer graphics and revolutionized parallel computing. More recently, GPU deep learning ignited modern AI — the next era of computing — with the GPU acting as the brain of computers, robots and self-driving cars that can perceive and understand the world. More information at http://nvidianews.nvidia.com/.

Source: NVIDIA

The post Baidu Advances AI in the Cloud with Latest NVIDIA Pascal GPUs appeared first on HPCwire.

DOE's INCITE Program Seeks Advanced Computational Research Proposals for 2018

Mon, 04/17/2017 - 08:01

ARGONNE, Ill., April 17, 2017 — The Department of Energy's (DOE's) *Innovative and Novel Computational Impact on Theory and Experiment (INCITE)* program will be accepting proposals for high-impact, computationally intensive research campaigns in a broad array of science, engineering, and computer science domains. DOE's Office of Science plans to award over 6 billion supercomputer processor-hours at Argonne National Laboratory and at Oak Ridge National Laboratory.

From April 17 to June 23, INCITE's open call provides an opportunity for researchers to make transformational advances in science and technology through large allocations of computer time and supporting resources at the Leadership Computing Facility (LCF) centers located at Argonne and Oak Ridge national laboratories. ALCF and OLCF are DOE Office of Science User Facilities.

The winning proposals will receive large awards of time on two primary systems: Mira, a 10-petaflops IBM Blue Gene/Q system at Argonne, and Titan, a 27-petaflops Cray XK7 at Oak Ridge. In addition, certain 2018 INCITE awards will receive time on Argonne's new Intel/Cray system, a 9.65-petaflops system called Theta.

The INCITE program will host open instructional proposal writing webinars on April 19, May 18, and June 6, 2017. Staff from both LCFs will participate in all three sessions. In addition, the ALCF is hosting a Computational Performance Workshop, May 2-5, 2017, to train INCITE users and others on ways to boost their code performance on ALCF's manycore systems.

Proposals will be accepted until the call deadline of 8:00 p.m. EDT on Friday, June 23, 2017. Awards are expected to be announced in November 2017.

To submit an application or for additional details about the proposal requirements, visit the 2018 INCITE Call for Proposals webpage.

For more information on the INCITE program and a list of previous awards, visit the INCITE program website.

About the U.S. Department of Energy's Office of Science

The U.S. Department of Energy's Office of Science is the single largest supporter of basic research in the physical sciences in the United States and is working to address some of the most pressing challenges of our time. For more information, visit the Office of Science website.

Source: Argonne National Laboratory

The post DOE's INCITE Program Seeks Advanced Computational Research Proposals for 2018 appeared first on HPCwire.

DDN Names Jessica Popp General Manager of IME Business Unit

Mon, 04/17/2017 - 07:48

SANTA CLARA, Calif., April 17, 2017 — DataDirect Networks (DDN) today announced Jessica Popp as the company's general manager (GM) of the newly created Infinite Memory Engine (IME) business unit. In this new role, Popp will leverage more than 20 years of experience in storage software and engineering management to build a world-class development organization for DDN's IME business. In addition, she will oversee the growth of the IME development and quality assurance teams, advance the IME feature set, and accelerate customer acquisition and revenue growth.

"Jessica Popp is an execution-focused engineering executive with a proven track record in managing highly specialized development organizations that build complex, system-level software products," said Robert Triendl, senior vice president, global sales, marketing, and field services, DDN. "As customer requirements evolve in both the technical computing and the enterprise big data markets, elastic storage solutions built from the ground up for non-volatile memory will become the norm for enterprise and cloud data centers. Jessica's leadership skills, her experience in managing distributed development teams, and her relentless focus on execution will be extremely valuable to DDN as we advance the IME product to meet broader market requirements."

Prior to joining DDN, Popp was the engineering director of the High Performance Data Division at Intel Corporation, where she managed the development and support of the Lustre* parallel file system and related products. In this role, she led a globally-distributed engineering division that provided end-to-end R&D, quality assurance, production support and product management for storage software products. Popp joined Intel in 2012 through its acquisition of Whamcloud, the main development arm for the open-source Lustre file system, where she played a vital role in the development and growth of the Lustre open-source file system software business.

DDN has maintained its commitment to lead innovation in storage performance and scalability for the technical computing and enterprise big data markets. The company's newly created IME business unit reflects this dedication and will drive the development, evolution and maturity of DDN's IME software-defined storage product. Available as software-only or as an appliance server, IME is a flash-native data cache that scales out to solve I/O scale and bottleneck challenges cost-effectively. It provides predictable, fast application performance with one-tenth to one-hundredth the storage hardware as compared with conventional storage solutions.

"With the growing challenge of I/O bottlenecks and achieving high performance at scale, DDN's IME scale-out, flash-native data caching solution, while still in its early stages, will undoubtedly play a major role for technical computing and enterprise big data environments in the future – and drive significant growth for DDN," Jessica Popp said. "I am excited to work with the high-caliber engineering team at DDN, to position the organization to seize the exciting market challenges and opportunities of the future, and to advance DDN's IME technology radically to meet and exceed the demanding requirements of our worldwide customers and partners, now and well into the future."

About DDN

DataDirect Networks (DDN) is the world's leading big data storage supplier to data-intensive, global organizations. For more than 18 years, DDN has designed, developed, deployed and optimized systems, software and storage solutions that enable enterprises, service providers, universities and government agencies to generate more value and to accelerate time to insight from their data and information, on premise and in the cloud. Organizations leverage the power of DDN storage technology and the deep technical expertise of its team to capture, store, process, analyze, collaborate and distribute data, information and content at the largest scale in the most efficient, reliable and cost-effective manner. DDN customers include many of the world's leading financial services firms and banks, healthcare and life science organizations, manufacturing and energy companies, government and research facilities, and web and cloud service providers. For more information, go to www.ddn.com or call 1-800-837-2298.

Source: DDN

The post DDN Names Jessica Popp General Manager of IME Business Unit appeared first on HPCwire.

Expanded use of New Microscopy Technology Requires an Innovative Approach to Storage

Mon, 04/17/2017 - 01:01

Life sciences organizations have been grappling with large and growing volumes of data for years by conducting research using the latest generation of sequencers, microscopes, and imaging systems.

But what happens when new equipment is expected to require a four-fold increase in storage? That was the situation the Van Andel Research Institute (VARI) recently found itself in. Already a leading life sciences research institute, VARI enhanced its capabilities with a new state-of-the-art cryo-electron microscopy (cryo-EM) facility. The centerpiece of the facility is an FEI Titan Krios from Thermo Fisher Scientific, the world's highest-resolution, commercially available cryo-EM.

A new phase in life sciences research

VARI, which is part of Van Andel Institute, is at the forefront of what industry experts see as a next step in research that explores the origins and treatment of diseases. While cryo-EM technology has been used for decades, the newest instruments can quickly create high-resolution models of molecules, something that was not attainable with other techniques before. This gives researchers a powerful new tool to more quickly and more precisely see some of the smallest yet most important biological components in their natural state.

Managing and analyzing such microscopy imaging data has moved life sciences computing beyond traditional genomics and bioinformatics and gets into phenotyping and correlation and structural biology. All of this work requires more computational power and much more storage capacity. To that point, cryo-EMs can generate up to 13 TBs of data per day. This represented a storage and data management challenge for the institute. Other organizations that move into this type of research can benefit from the lessons VARI learned.

Keeping the HPC workflows running

VARI has a 20-year history of conducting biomedical research and providing scientific education. Its focus is on improving health and enhancing the lives of current and future generations. Using state-of-the-art technologies and instrumentation, the institute's scientists, educators, and staff work to translate discoveries into highly innovative and effective diagnostics and treatments. The institute needed a powerful HPC and storage environment to serve teams of scientists with diverse research demands and aggressive project timelines.

The addition of the powerful cryo-EMs built on this tradition. The cryo-EMs enable VARI scientists to see the structure of molecules that are one-ten-thousandth the width of a human hair, and is expected to quadruple VARI's storage requirements. The institute's scientists also are conducting data- and storage-intensive trailblazing molecular dynamics simulations and large-scale sequencing projects in the search for new ways to diagnose and treat cancer, Parkinson's, and many other diseases.

What was needed was an infrastructure that would allow the institute to elevate the standard of protection, increase compliance, and push the boundaries of science on a single, highly scalable storage platform.

After determining that an IBM® Spectrum Scale™ (formerly known as GPFS) parallel file system was the way to meet current and future storage needs, VARI chose DataDirect Network's (DDN's) GS7K® parallel file system appliance with enterprise-class features, including snapshots, rollbacks, and replication. The institute also selected DDN's WOS® storage to provide an active archive for greater global data sharing and research collaboration.

Most important, the solution simplified data tiering between the GS7K and WOS, providing a single storage system that enables instrument and other research data to be ingested, analyzed, and shared in a manner that addresses both performance and cost-efficiency. In addition, DDN's OpenStack® driver significantly streamlined storage integration with VARI's hybrid on-premises and cloud computing environment.

Benefits abound

The new storage solution is helping in a number of ways.

The end-to-end solution replaced fragmented data silos with powerful, scalable centralized storage for up to 2PB of instrument and research data. According to Zachary Ramjan, research computing architect for Van Andel Research Institute, consolidating primary data storage for both state-of-the-art scientific instruments and research computing offers better protection for irreplaceable data while reducing infrastructure costs considerably.

"We've saved hundreds of thousands of dollars by centralizing the storage of our data-intensive research and a dozen data-hungry scientific instruments on DDN," said Ramjan.

With the highly scalable storage solution, the institute is prepared to accommodate the expected 13TB a day of data generated using its cryo-EM technology. The solution employs a tiered storage approach. New data goes straight into the high performance DDN GS7K tier.

As the data "cools" and investigators move to new projects, the institute may still have to retain the data due to obligations or the user wants to keep it around. At this point, the data is automatically moved to a lower performance and more economical tier. This is the WOS controlled tier. It's where much of the cryo-EM data will end up after initial processing.

Data movement is controlled by policy capabilities in the file system. Automating data flow in this way greatly reduces steps and admin requirements.

The result is that researches get simple, fast access to petabytes of storage for research and instrument data that has the high performance of a well tune parallel file system but the easy expandability of an object storage solution all-in-one. And the solution meets the institute's exponential storage growth and active archive requirements of that data. The end-to-end DDN solution thus provides the scalable storage capacity VARI needs to keep pace with the increased use of cryo-EM and next-generation sequencing technologies.

http://www.ddn.com/customers/van-andel-research-institute/

The post Expanded use of New Microscopy Technology Requires an Innovative Approach to Storage appeared first on HPCwire.

Supermicro to Share Third Quarter Fiscal 2017 Financial Results
Fri, 04/14/2017 - 07:56

SAN JOSE, Calif., April 14, 2017 — Super Micro Computer, Inc. (NASDAQ: SMCI) has announced that it will release third quarter fiscal 2017 financial results on Thursday, April 27, 2017, immediately after the close of regular trading, followed by a teleconference beginning at 2:00 p.m. (Pacific Time).

Conference Call/Webcast Information for April 27, 2017

Supermicro will hold a teleconference to announce its third quarter fiscal 2017 financial results on Thursday, April 27, 2017, beginning at 2:00 p.m. (Pacific Time). Those wishing to participate in the conference call should dial 1-888-715-1389 (International callers dial 1-913-312-1383) a few minutes prior to the call's start to register. The conference ID is 7820076. A replay of the call will be available through 11:59 p.m. (Eastern Time) on Thursday, May 11, 2017, by dialing 1-844-512-2921 (International callers dial 1-412-317-6671) and entering replay PIN 7820076.

Those wishing to access the live or archived webcast via the Internet should go to the Investor Relations tab of the Supermicro website at www.Supermicro.com.

About Super Micro Computer, Inc.

Supermicro, a provider of high-performance, high-efficiency server technology and innovation is a provider of end-to-end green computing solutions for Data Center, Cloud Computing, Enterprise IT, Hadoop/Big Data, HPC and Embedded Systems worldwide. Supermicro's advanced Server Building Block Solutions offer a vast array of components for building energy-efficient, application-optimized, computing solutions. Architecture innovations include Twin, TwinPro, Big Twin, FatTwin, Ultra Series, MicroCloud, MicroBlade, SuperBlade, Double-sided Storage, Battery Backup Power (BBP) modules and WIO/UIO.

Products include servers, blades, GPU systems, workstations, motherboards, chassis, power supplies, storage, networking, server management software and SuperRack cabinets/accessories delivering unrivaled performance and value.

Founded in 1993 and headquartered in San Jose, California, Supermicro is committed to protecting the environment through its "We Keep IT Green®" initiative. The Company has global logistics and operations centers in Silicon Valley (USA), the Netherlands (Europe) and its Science & Technology Park in Taiwan (Asia).

Source: Supermicro

The post Supermicro to Share Third Quarter Fiscal 2017 Financial Results appeared first on HPCwire.

DOE Supercomputer Achieves Record 45-Qubit Quantum Simulation
Thu, 04/13/2017 - 18:21

In order to simulate larger and larger quantum systems and usher in an age of "quantum supremacy," researchers are stretching the limits of today's most advanced supercomputers. Just as classical computing systems have been instrumental in advancing their own forward progression, today's fastest machines are helping pave the way for quantum computing breakthroughs, which will be revolutionary for applications in quantum chemistry, material science, machine learning, and cryptography.

A research team from ETH Zurich in Switzerland recently succeeded in simulating the largest quantum device yet — a 45-qubit circuit — using the Knights Landing-based Cori II machine at Lawrence Berkeley National Laboratory. Cori II has 9,304 compute nodes each outfitted with one 68-core Intel "Knights Landing" Xeon Phi 7250 processor, linked with the Cray Aries interconnect, with a total peak performance of 29.1 petaflops and 1 PB of aggregate memory.

Thomas Häner and Damian S. Steiger of the Institute for Theoretical Physics at ETH Zurich performed simulations of low-depth random quantum circuits, which were proposed by Google to demonstrate quantum supremacy. In the paper describing the work, the authors specify that "the execution of low-depth random quantum circuits is not scientifically useful on its own" but "running such circuits is of great use to calibrate, validate, and benchmark near-term quantum devices."

"In order to make use of the full potential of systems featuring multi- and many-core processors, we use automatic code generation and optimization of compute kernels, which also enables performance portability," they write.

Summary of all simulation results carried out on Cori II. (Source)
To simulate the 45-qubit quantum circuit, Häner and Steiger used 8,192 Cori II nodes and a total of 0.5PB of memory, achieving an average 0.428 petaflops. In explaining the low performance in relation to peak FLOPS, the team refer to 1) heavy communication overhead (75 percent of circuit simulation time spent in communication) and state 2) kernel performance suffers where few kqubit gates are applied before a global-to-local swap needs to be performed (see section 4.1.2 for further analysis and discussion).

The research team says to the best of their knowledge, this 45-qubit quantum circuit simulation is the largest ever conducted. "Our highly-tuned kernels in combination with the reduced communication requirements allow an improvement in time-to-solution over state-of-the-art simulators by more than an order of magnitude at every scale," they write.

The next step for the ETH Zurich team is to add more qubits to their simulation. Although 49-qubits is widely-held as the point at which quantum devices surpass the most capable traditional supercomputers and thwart larger simulations, the researchers have a plan to reach this threshold.

"While we do not carry out a classical simulation of 49 qubits, we provide numerical evidence that this may be possible," the research team states. "Our optimizations allow reducing the number of communication steps required to simulate the entire circuit to just two all-to-alls, making it possible to use, e.g., solid-state drives if the available memory is less than the 8 petabytes required."

Cori is the flagship resource of the DOE National Energy Research Scientific Computing Center (NERSC). The system was named in honor of the American biochemist Gerty Cori, the first American woman to win a Nobel Prize in science. Cori II is currently ranked number five on the (November 2016) Top500 list. The full system is comprised of two partitions: 2,004 Intel Xeon "Haswell" processor nodes and 9,300 Intel Xeon Phi "Knight's Landing" nodes. According to its Top500 submission, the KNL partition (Cori II) has a peak performance of 27.9 petaflops and a measured Linpack score of 14 petaflops.

The 11-page paper, "[0.5 Petabyte Simulation of a 45-Qubit Quantum Circuit](...)," is published on arXiv.org. There's also a [writeup of the research](...) at *MIT Technology Review.*

The post [DOE Supercomputer Achieves Record 45-Qubit Quantum Simulation](...) appeared first on [HPCwire](...).

Larry Smarr Talks Machine Intelligence at Jackson State

Thu, 04/13/2017 - 14:01

JACKSON, Miss., April 13, 2017 — Dr. Larry Smarr, physicist and Big Data thought leader, practices what he preaches.

When asked about the auto-pilot option packaged in the Tesla, he proudly says: "I own one. I have the model X. It drives me from UC San Diego up to UC Irvine and back every time."

It's Monday, April 10, the day before Smarr, founding director of California Institute of Telecommunications and Information Technology (Calit2), is to appear in Jackson State University's CSET Engineering auditorium to present his theory on the mounting growth of machine artificial intelligence.

Big Data thought leader Larry Smarr, holder of the Harry E. Gruber professorship in UCSD's Department of Computer Science and Engineering lectured and facilitated an open discussion in the auditorium of the College of Science, Engineering and Technology. Professor Smart speculated on the exponentially growing machine intelligence and how it will increasingly inter-operate with human intelligence. (Charles A. Smith/University Communications)

JSU and Calit2 are partners on National Science Foundation-funded grants, like SCOPE — Scalable Omnipresent Environment — a visual metaphor for a combined microscope and telescope that enables users to explore data from the nano to the micro to the macro to the mega scale. Another is already implemented at Jackson State – a camera system called SENSEI, a/k/a the virtual wall, capable of capturing 3D stereo and still images for viewing in nationally networked, virtual reality systems.

Seemingly unfazed by a fatal 2016 accident that involved a Tesla set on auto pilot, Smarr points out that the U.S. National Highway Traffic Safety Administration cleared the automaker of any fault in the incident.

"I'm much less concerned that my Tesla will have an accident than some stupid human driving a car is going to have an accident," he quips then smiles.

Smarr, of course, backs his opinion with science, saying, "There are ultrasound, radar and optical sensors on each of the Teslas. As they're driving, the data stream of how they're interacting with the world is sent wirelessly over the Internet to the Tesla cloud."

What happens next, according to Smarr, is that a "hive mind" develops where the experiences of all the Teslas are shared with each other in the cloud, and a machine-learning algorithm goes in and improves the way that all Teslas drive.

"So, the more Teslas there are and the more miles they drive, the safer everybody becomes," he says matter of factly.

The Harvard Junior Fellow adds that the average car driver's knowledge base is dependent upon whatever the driver recalls and not 10,000 similar vehicles second to second sharing data.

To call Smarr a big deal is an understatement. His bio contains phrases like "theoretical observational" and "computational astrophysics," a field he has pioneered for 25 years. He is credited with galvanizing the early development of foundational components of U.S. global cyberinfrastructure and most recently became a leader of the quantified self-movement or "life logging."

In addition to being a member of the National Academy of Engineering, Smarr holds memberships with several distinguished science organizations. He has also served on the NASA Advisory council to four NASA administrators as well as other esteemed positions.

If who and what Smarr is does not stimulate the brain even slightly, then his description of the work conducted at the Calit2 Pattern Recognition Laboratory should be an eye-opener.

"So there are new kinds of computer chips emerging that are specialized for recognizing patterns in images, for instance, a hearing aid. What a hearing aid is trying to do is recognize the pattern of a voice in a noisy environment," he says.

Through specialized processors like hearing aids or voice authentication technology, Calit2 develops machine learning that can identify patterns and collects this information into a lab that faculty and students can access in order to conduct experiments to understand how, particularly, Big Data can be analyzed more efficiently.

"So, it's a pattern that's in a bunch of data, which as a human, if you just look at all of the data then you can't figure anything out. But if you have a specialized computer that can do it then maybe you can," he says.

Smarr and Calit2 are working with everyone from startups to big name companies like IBM that may be in the infancy stages of new technology and need assistance in determining how it may be best used.

"So, by having a lot of faculty and particularly students who may have a lot of clever ideas that faculty may not think of," he chuckles, "they get access to the lab so they can conduct experiments."

In addition to determining uses for new technology, Smarr adds that identifying market niches is also an outcome of Big Data research.

Testing science in more ways than cruising in his Tesla, Smarr is engaged in a computer-aided study of his own body – lifelogging. Every month, the noted director relinquishes 5-6 vials of his blood for analysis and every quarter he hands over up to 20 vials of the life-giving substance.

Smarr says his venture into the "quantified self" movement stems from his 25 years of experience as an astrophysicist. Over that period, he has conducted studies like taking measurements over time (a time series) to figure out how an unusual point came to be in the sky and then determine what that point is doing at various stages.

"I began to think about the body as a dynamic, multi-component, nonlinear system. And I thought, well, surely, that's the way people are trying to figure out what's going on with you before it gets to the point that you have a symptom," he says.

But to his confoundment, Smarr discovered his theory was not the way medicine is conducted.

"So, I said why don't I make myself a lab animal and turn my body into an observatory and see if we can begin to understand the dynamics of some fundamental things like inflammation, glucose and insulin and things like cholesterol. All that kind of stuff," he says.

Expounding on his quest to become a human guinea pig, Smarr discloses one of his goals: "Can we figure out unintended consequences of some of our activities? Whether it's medical treatments or whether it's things like our nutrition, exercise, sleep or other things that people are arbitrary about what they do as if there are not going to be any consequences and there are."

Pulling out his smartphone, he reveals an image of what he describes as "70 different biomarkers in me" taken over 20 years with each marker representing things like glucose, cholesterol, sodium and potassium levels. Through a color-coordinated effort, Smarr is able to see when his body is healthy and when it exceeds the healthy range by a certain percentage.

"I thought I was healthy," he laughs as he glances at the rainbow-colored images.

Smarr has data compiled from over 20 years but explains that he's been doing the blood work and intensively studying microbiome – the general composite material present in or on the human body – for about five years.He declares that if others were able to see clear indicators that revealed the onslaught of Alzheimer's or diabetes, then people could possibly make some behavioral changes and choices to stop diseases before they develop."By bringing in the big data that can be read out of your body," Smarr hopes to improve various areas of medicine.

"I just lifelog a little more intensely than most people. And I'm a scientist, and I have a big laboratory, so I can afford to do it. He explains, "I'm not saying you should be like me. What I'm trying to be is the patient of the future, so in the future, the cost will go way down, and it will be much less evasive to measure these things."

Smarr readily admits that it's an experiment, and "it's a little weird being me, but somebody has to do it."

Experimentation appears to be working for the scientist whose research and curiosity is taking him to heights he had not imagined if that's easy to believe after reading his bio.

"I never in a million years thought I would end up working with surgeons and medical doctors. I didn't have any training on biomedical things," says Smarr, referring to his recent surgery to remove a portion of his colon.

In 2012, after an MRI, Smarr was less than inspired by the 2D black and white splices he was shown by his radiologist. He then requested the images and the 3D data that is routinely included in individual medical evaluations.

Together with his "virtual reality guy," Jurgen Schulze, he created a 3D image of his abdomen.

After being diagnosed with colonic Crohn's disease late last year, he offered the 3D image to his surgeon, Dr. Sonia Ramamoorthy, several days before the removal of the inflamed area.

In short, Ramamoorthy's review of the detailed image led her to change her original plan for the incisions she would make to Smarr's abdomen and reduced his operating time by nearly 30 minutes.

Smarr again pulls out his phone and this time he shows a video of himself lying unconscious on the operating table while Ramamoorthy works simultaneously with the surgical robot and the 3D imagery.

"It was like having a 3D Google map," he jokes.

A week after his procedure, he recalls Ramamoorthy saying, "Oh my God, I had a patient today. It would've been so helpful if we had this."

Smarr's innovative thinking has the potential to not only benefit the medical industry but society as a whole. "The fact that you can be under anesthesia less means that recovery chances would be improved." He says, "So, we're making a program at UC San Diego, in the medical school, to then simplify the software to make it more integrated with the robot and then train and do another 10 patients."

After learning that a lot of Smarr's esteemed accomplishments started with having an organic knack for investigation, it is easy to imagine him as a kid making a Tesla coil and blacking out all the televisions and radios in his neighborhood as he once said he did.

When asked what type of skills or characteristics a student would need to become a Larry Smarr, he says: "Well, I think you have to be intensely curious, and then I guess you can do what President Reagan said about the Russians – 'Trust but verify.'"

"Think" is the operative term Smarr wants kids to learn. "Doing your Facebook updates a billion times a day, well, I'm sure there is some good that comes from that – it is not thinking. And the other thing is reading is so important."

He conveys that while playing video games results in some positives, like the ability to multitask, an inadvertent outcome is that kids' attention spans have grown shorter. He becomes quiet as if searching for the right words to describe his thoughts; then he adds that the younger generation is no longer reading.

"When you think about the way books are written – the very best books, like the very best art and the very best music – you are having a personal mentor on how to organize your thoughts, how to communicate, how to convince someone of something or how to paint a word picture," he says.

"You don't know how to do that to start with, but as you read, your brain is remembering patterns like that's how you write, that's how you phrase your words, that's how you communicate. So this is not generally talked about – why you need to read. But if you don't, then you dumb down the population because an incredible number of things are going to happen in your lifetime and how are you going to be prepared to know how to react?

About Jackson State University: Challenging Minds, Changing Lives

Jackson State University, founded in 1877, is a historically black, high research activity university located in Jackson, the capital city of Mississippi. Jackson State's nurturing academic environment challenges individuals to change lives through teaching, research and service. Officially designated as Mississippi's Urban University, Jackson State continues to enhance the state, nation and world through comprehensive economic development, health-care, technological and educational initiatives. The only public university in the Jackson metropolitan area, Jackson State is located near downtown, with five satellite locations throughout the area. For more information, visit www.jsums.edu or call 601-979-2121.

Source: Rachel James-Terry, Jackson State University

The post Larry Smarr Talks Machine Intelligence at Jackson State appeared first on HPCwire.

CERN openlab Explores New CPU/FPGA Processing Solutions
Thu, 04/13/2017 - 13:00

At CERN, the European Organization for Nuclear Research, physicists and engineers are probing the fundamental structure of the universe. The Large Hadron Collider (LHC), which began working in 2008, is the world's largest and most powerful particle accelerator; it is housed in an underground tunnel at CERN. Niko Neufeld is a deputy project leader at CERN who works on the Large Hadron Collider beauty (LHCb) experiment, which explores what happened after the Big Bang that allowed matter to survive and build the Universe we inhabit today.

"CERN experiments produce an enormous amount of data with forty million proton collisions every second, which leads to primary data rates of terabits per second," says Neufeld when speaking on a recent FPGA vs. CPU panel. "This is an enormous amount of data and there are a number of technical challenges in our work. We use a number of processing solutions including central processing units (CPUs), field-programmable gate arrays (FPGAs), and graphic processing units (GPUs), but each of these solutions have some limitations. We are collaborating with Intel in experimenting with a co-packaged Intel Xeon processor plus FPGA Quick Path Interconnect (QPI) processor in our LHCb research to try to determine which technology provides the best results."

CERN collaborates with leading ICT companies and other research institutes through a unique public-private partnership known as 'CERN openlab'. Its goal is to accelerate the development of cutting-edge solutions for the worldwide LHC community and wider scientific research. Through a CERN openlab project known as the 'High-Throughput Computing Collaboration,' researchers are investigating the use of various Intel technologies in data filtering and data acquisition systems.

Figure 1. CERN researchers shown in the Large Hadron Collider tunnel in front of the LHCb detector. Courtesy of CERN (courtesy CERN).

Introducing the co-packaged Intel CPU / FPGA Processor

Today the CPU and FPGA are used as discrete chips in a solution – with an Intel Xeon processor and an FPGA which is typically attached via a PCIe interconnect to the CPU. And the development environment is also discrete using independent development tools from Intel and tools such as OpenCL and C++. Intel is working toward a common workflow and development flow to better integrate FPGAs.

"FPGA typically uses a higher level machine abstraction language (such as Verilog and VHDL) which have a painful low-level hardware programming model for most people. As a next step, Intel has a solution that co-packages the CPU and FPGA in the same Multichip Chip Product (MCP) package to deliver higher performance and lower latency than a discrete solution," states Bill Jenkins, Intel Senior AI Marketing Manager. The Intel MCP is supported by a cross-platform development framework like OpenCL that can be used to develop applications for both the CPU and FPGA. The Intel solution includes a fully unified intellectual property (IP) and development suite, including languages, libraries and development environments. The roadmap to a unified development flow leverages common tools and libraries to support both FPGA and Intel Xeon processor + FPGA systems along with an expansive ecosystem network of Intel and vendors working on independent development tools for demanding workloads such as HPC, imaging identification, security and big data.

Abstracting away FPGA Coding

Intel is building an abstraction layer (as part of the product containing the Intel CPU and FPGA in the same MCP package), called the Orchestration Software layer. This layer and the higher level IP and software models help make development less complex so that developers don't need to code specifically to the FPGA. The FPGA-enabled Orchestration software layer abstracts away the API to communicate with the FPGA as shown in the following example.

Figure 2. Example of Intel implementation of user IP implemented into FPGA via an abstraction Orchestration software layer

There is a cloud-based library of functions and end-user IP that have been pre-compiled and built that is loaded into the FPGA at runtime. The user first launches a workload from the host and it goes into the Orchestration software which pushes a function into the FPGA. This produces a bitstream that is pre-compiled on the FPGA to bring the data in—it is almost like a fixed architecture I/O interface.

In the example scenario, users simply download the image from the abstraction Orchestration software layer to the FPGA and it is ready to run without compilation. "With the abstraction Orchestration software layer," Jenkins explained, "Intel is abstracting away all the difficulties of FPGA programming using machine language tools while enabling all the higher level Intel frameworks including the Intel Trusted Analytics Platform (TAP) and Intel Scalable System Framework (SSI) and tying the FPGA into the frameworks. Intel is developing this approach for a variety of markets including visual understanding, analytics, enterprise, Network Function Virtualization (NFV), VPN, genomics, HPC and storage."

Large Hadron Collider High-Energy Physics Research at CERN

Neufeld indicates that the experiments at CERN — through what they refer to as 'online computing' — require a first-level data-filtering to reduce the data to an amount that can be stored and processed on more traditional processing units such as Intel Xeon processors. Figure 3 shows a schematic view of the future LHCb readout system. At the top level, there is a detector and optical fiber links, which transfer data out of the detector. CERN uses FPGAs to acquire data from the detector. There are also large switching fabrics, as well as clusters of processing elements including CPUs, FPGAs, and GPUs to reduce the amount of data. One of the questions the CERN team is testing is "Which technologies should we use and which provide the best performance and lowest energy usage results?"

Figure 3. Schematic diagram showing future LHCb first-level data-filtering system. Courtesy of CERN.

CERN Tests Complex Cherenkov Angle Reconstruction Calculation

CERN has extensive experience using FPGAs in their research work. "We typically use FPGAs in our research to run algorithms looking for simple integer signatures, or for other less complicated calculations. When we heard about the Intel Xeon / FPGA combined processor, we chose a test using a complex algorithm to do a Cherenkov angle reconstruction of light emission in a particle detector, which is not typically performed on an FPGA. This involves tracing a light particle — photon — through a complex arrangement of optical reflection and deflection systems. Our test case used a rich PID algorithm to calculate the Cherenkov angle for each track and detection point. This is a complex mathematical calculation that involves hyperbolic functions, roots, square roots, etc., as shown in Figure 4. It is one of the most costly calculations done in online reconstruction," states Neufeld.

Figure 4. Test case running Rich PID algorithm to calculate Cherenkov angle. Courtesy of CERN.

Coding the Cherenkov Angle Reconstruction in Verilog versus OpenCL

The CERN team first implemented the Cherenkov angle reconstruction by coding it in the Verilog HDL. The team wrote a 748 clock-cycle long pipeline in Verilog, along with additional blocks developed for the test including: cubic root, complex square root, rotational matrix, and cross/scalar product. It was a lengthy task doing this coding in Verilog with 3,400 lines of code. With all test benches, the implementation took 2.5 months.

Next, the team recoded the Cherenkov angle code using the OpenCL and the BSP (board support package) designed to work across a variety of hardware platforms. Because OpenCL is an abstraction language, it required only 250 lines of code and took two weeks of coding. Not only was coding in OpenCL much faster but the performance results were similar. Figure 5 shows the results of the Verilog versus OpenCL implementation.

Figure 5. Result of Verilog (CQRT) versus OpenCL (RICH) code and performance. Courtesy of CERN.

CERN Compares Co-packaged Intel Xeon – FPGA Processor against Nallatech PCIe Stratix V FPGA Board

To test performance of the Verilog code, the CERN team used a commercially available Stratix V GXA7 FPGA board / Nallatech 385 board for testing. They achieved an acceleration of a factor up to six with the Stratix – Nallatech PCIe board. However, they found a bottleneck in data transfer—they could not keep the pipeline busy because the PCIe card was limited to an eight-lane interface. Next, the CERN team did tests with the Cherenkov angle code comparing a Nallatech FPGA Board with the co-packaged Intel Xeon/FPGA QPI processor.

Finally, the CERN team tested an Intel Xeon CPU, PCIe Stratix V FPGA and Intel Xeon processor/Stratix V QPI (where only the interconnect was different). As shown in Figure 6, there was a factor of 9 speed up for the PCIe Stratix V FPGA and a 26 factor speed up for the Intel Xeon processor/Stratix V QPI with the faster interconnect.

Figure 6. Test results from the CERN team comparing Intel Xeon CPU, PCIe Stratix V FPGA and Intel Xeon processor/FPGA QPI. Courtesy of CERN.

CERN Plans to do Future Testing using co-packaged Intel Xeon/ Intel Arria10 FPGA Processor

"Our CERN team found the results of using the co-packaged Intel Xeon processor/Stratix V QPI processors to be very encouraging. In addition, we find the programming model with OpenCL attractive and it will be mandatory for the High-Energy Physics (HEP) field. Intel will be launching a co-packaged Intel Xeon processor / Intel Arria 10 FPGA processor in the future. We want to do other experiments with the co-packaged Intel Xeon processor/ Arria 10 FPGA. We expect that the high-bandwidth interconnect and modern Arria 10 FPGA card will provide high performance and performance per Joule for HEP algorithms," states Neufeld.

Linda Barney is the founder and owner of Barney and Associates, a technical/marketing writing, training and web design firm in Beaverton, OR.

The post CERN openlab Explores New CPU/FPGA Processing Solutions appeared first on HPCwire.

Intel Promotes Three Executives

Thu, 04/13/2017 - 08:43

SANTA CLARA, Calif., April 13, 2017 — Intel Corporation has announced that Intel executives Diane Bryant, Murthy Renduchintala and Stacy Smith have been promoted to Group Presidents and Renduchintala has been appointed chief engineering officer.

Diane Bryant, Murthy Renduchintala and Stacy Smith

"The new leadership appointments reflect the scope of Diane, Murthy and Stacy's responsibilities," said Intel CEO Brian Krzanich. "Their groups are a significant part of Intel's business and are instrumental to driving Intel's growth strategy and flawless execution going forward."

Bryant has been named Group President of the Data Center Group (DCG). She has been leading DCG since 2012 and is responsible for the group's P&L, strategy and product development. The organization consists of the technology for server, storage and network infrastructure to serve the cloud service providers, communications service providers, enterprise and government organizations. She joined Intel in 1985.

Renduchintala has been named Group President, Client and Internet of Things Business & Systems Architecture Group (CISA), and appointed chief engineering officer. The CISA umbrella division is responsible for aligning technology, engineering, product design and business direction across the client device and IoT segments. He joined Intel in November 2015.

Smith has been named Group President of Manufacturing, Operations and Sales. He has overseen the company's global Technology and Manufacturing Group and its worldwide sales organization since last fall. Previously, Smith served nine years as Intel's chief financial officer. He joined Intel in 1988.

About Intel

Intel (NASDAQ: INTC) expands the boundaries of technology to make the most amazing experiences possible. Information about Intel can be found at newsroom.intel.com and intel.com.

Source: Intel

The post Intel Promotes Three Executives appeared first on HPCwire.

Engility to Pursue NASA Advanced Computing Services Opportunity

Thu, 04/13/2017 - 08:08

CHANTILLY, Va., April 13, 2017 — Engility Holdings, Inc. (NYSE: EGL), announced today that the company will bring its world-class high performance computing (HPC) capabilities to bear as it competes to win NASA's Advanced Computing Services contract.

"HPC is a strategic, enabling capability for NASA," said Lynn Dugle, CEO of Engility. "Engility's cadre of renowned computational scientists and HPC experts, coupled with our proven high performance data analytics solutions, will help increase NASA's science and engineering capabilities."

Engility subject matter experts are on the forefront of developing systems and solutions that leverage integrated, multi-scaled scientific and engineering models. Engility applies sophisticated visualization, numerical algorithms and computational frameworks to complex scientific and engineering problems. These can be executed on multi-peta scale computing systems to develop HPC solutions, simulations and tools. These solutions, simulations and tools enable warfighters, academics and policymakers to better understand complex phenomena, reduce time to action, enhance their productivity and move complex technologies from the laboratory to the operational mission while saving and improving lives.

Engility's legacy of HPC success stretches back a quarter century. In 1993, the company reduced U.S. Air Force weather model time by 75 percent. Engility also delivered the National Oceanic and Atmospheric Administration's (NOAA's) first commodity-based 64-bit cluster in 2004. In addition, the company established the advanced computing research program in support of the Army's HPC Research Center, partnering with universities and government labs such as NASA Ames Research Center in 2007. Just last year, Engility won a $112 million Food and Drug Administration HPC contract that improves scientific computing. For instance, bioinformatics insights provide faster, better-informed regulatory decisions and enhance collaboration among scientists and the worldwide scientific community.

Engility's HPC team:

- Includes some of the best computational and supercomputing scientists in the industry, enabling advanced HPC users to work across all scientific and engineering domains. Engility scientists worked with NOAA to enable the design and acquisition of their first large research and development HPC systems.
- Secures data. Engility delivers in-depth knowledge, support and direction for all IT security-related activities at NOAA's Geophysical Fluid Dynamics Laboratory's Research and Development HPC Program.
- Helps customers strategize and prioritize investments in HPC. Engility developed a tool that enabled NOAA to make strategic investments in HPC technology and acquisition that resulted in an accelerated path to next-generation weather and climate modeling.

For more information about Engility's high performance computing expertise, please visit http://www.engilitycorp.com/services/hpc/.

About Engility

Engility (NYSE: EGL) is engineered to make a difference. Built on six decades of heritage, Engility is a leading provider of integrated solutions and services, supporting U.S. government customers in the defense, federal civilian, intelligence and space communities. Our innovative, highly technical solutions and engineering capabilities address diverse client missions. We draw upon our team's intimate understanding of customer needs, deep domain expertise and technical skills to help solve our nation's toughest challenges. Headquartered in Chantilly, Virginia, and with offices around the world, Engility's array of specialized technical service offerings include high-performance computing, cyber security, enterprise modernization and systems engineering. To learn more about Engility, please visit www.engilitycorp.com and connect with us on Facebook, LinkedIn and Twitter.

Source: Engility

The post Engility to Pursue NASA Advanced Computing Services Opportunity appeared first on HPCwire.

Tutorials Schedule Announced for PEARC17

Wed, 04/12/2017 - 14:52

NEW ORLEANS, April 12, 2017 — PEARC17 organizers today announced 20 full-day and half-day tutorials for attendees at the Practice & Experience in Advanced Research Computing 2017 conference, July 9-13, in New Orleans.

All tutorials will be held Monday, July 10. Of two full-day tutorials, one focuses on enabling and advancing on-campus research computing and the other on manycore programming. Half-day tutorials are scheduled on topics ranging from using science gateways, cloud computing, virtual clusters, programming, and managing scientific data to visualization, data sharing, and building data portals.

When registering for tutorials day, registrants will be asked to select the tutorials they are likely to attend. While attendees may ultimately choose to attend different tutorials on site, tutorials selected by fewer than five attendees during the registration period may be cancelled.

Anyone interested in advanced research computing challenges is encouraged to review the schedule and register soon.

About PEARC

The PEARC (Practice & Experience in Advanced Research Computing) conference series is being ushered in with support from many organizations and will build upon earlier conferences' success and core audiences to serve the broader community. In addition to XSEDE, organizations supporting the new conference include the Advancing Research Computing on Campuses: Best Practices Workshop (ARCC), the Science Gateways Community Institute (SGCI), the Campus Research Computing Consortium (CaRC), the ACI-REF consortium, the Blue Waters project, ESnet, Open Science Grid, Compute Canada, the EGI Foundation, the Coalition for Academic Scientific Computation (CASC), and Internet2.

Source: PEARC

The post Tutorials Schedule Announced for PEARC17 appeared first on HPCwire.

Hyperscalers Emerging From 'Hype Phase'
Wed, 04/12/2017 - 14:27

The number of companies fitting the description of "hyperscale" now accounts for 68 percent of the cloud infrastructure services market, a researcher found.

Synergy Research Group said this week is has identified 24 companies that meet its definition of hyperscale operators. Those companies offer infrastructure, platform and private hosted cloud services. An additional 59 percent offer software services. In 2012, the market research found that hyperscale operators accounted for only 47 percent of each of those markets.

The survey defined hyperscale operators as running hundreds of thousands of servers in their datacenters. The largest cloud services vendors such as Amazon Web Services, Google and Microsoft have millions of servers.

The market researcher reported that 45 percent of hyperscale datacenters are located in the U.S.; China is a distant second with 8 percent of high-end datacenters. "Country distribution of data centers reflects the U.S. dominance of cloud and internet technologies," the survey noted.

The 24 companies identified as hyperscalers operate an estimated 320 datacenters. Market leaders AWS, Microsoft and IBM each have more than 45 datacenters with at least two of the four regions covered by the survey (North America, Asian Pacific, Europe, Middle East, Africa and Latin America).

Synergy Research, Reno, Nev., did not identify the 24 companies it designated as hyperscale operators, but did update its rankings of the leading public cloud vendors. The researcher said AWS maintained its dominant 40 percent share of public cloud revenues as of the fourth quarter of 2016.

Cloud rivals Microsoft Azure, Google Cloud Platform and IBM SoftLayer together accounted for about 24 percent of the public cloud market at the end of last year, up 5 percent over the previous year. That trend reflects the growing enterprise shift to a multi-cloud strategy designed to avoid overreliance on a single cloud provider.

Market share for other cloud providers declined slightly over the last year, the survey found.

Microsoft continues to lead the enterprise software services market along with a handful of hyperscale operators. Those emerging vendors lead in market segments such as database search, electronic commerce and mobile apps. Another key market for emerging hyperscalers is social media, where Facebook dominates and helps drive open-source infrastructure development.

As the number of hyperscale operators grows, the market researcher predicts that IT spending will continue to shift away from on-premise infrastructure to public cloud and software services as the shift to hybrid clouds accelerates. Hence, the cloud services market is "booming," the market researcher said, with quarterly revenues soaring.

Over the next five years, Synergy Research projects that growth rates for infrastructure, platform and software services along with hosted private clouds will continue to grow "at impressive rates." Hence, hyperscale providers will continue to increase their share of service markets, overall IT budgets and spending for datacenter infrastructure, it added.

The new hyperscalers are "radically changing IT spending patterns within enterprises, and causing major disruptions among infrastructure technology vendors," noted John Dinsdale, research director at Synergy Research Group.

The post Hyperscalers Emerging From 'Hype Phase' appeared first on HPCwire.

Penguin Takes a Run at the Big Cloud Providers
Wed, 04/12/2017 - 10:57

Fighting for sway among the big HPC-in-the-cloud providers – AWS, Azure, etc. – is challenging for smaller players. The giants own the mindshare. HPC specialist Penguin Computing recently re-ran benchmarks from a study of its larger brethren and says the results show its 'public cloud' – Penguin on Demand (POD) – is among the leaders in cost and performance. The natural caveat here is Penguin ran the tests. The company says the tests accurately recreate those in the study and that it is willing to let prospects or third parties run the tests for themselves at no charge.

A marketing exercise? Of course. The Penguin effort was in response to a paper – *Comparative benchmarking of cloud computing vendors with High Performance Linpack* – authored by Exabyte.io and posted on arXiv.org last month. But besides simply being an effort by Penguin to gain user attention and push its way more forcefully into the land of giants such as Microsoft, Google, AWS, and IBM/SoftLayer, and Rackspace, all examined in the paper, the effort also reveals the growing competitive zeal among many providers to offer HPC in the cloud.

Where Microsoft Azure was the top performer in the original study, Penguin reports that POD comes out ahead in the new benchmarking it performed on both price and performance measures. Penguin, according to its measurements, takes the lead in terms of Linpack gigaflops-per-core, Linpack gigaflops-per-node and has it has the most favorable speedup at scale.

To some extent one would expect Penguin to outperform public cloud providers. POD is best characterized as an on-demand HPC cluster rather than a typical public cloud. It is a tightly coupled environment (with the option of OmniPath 100Gb/s interconnect, which was used for this testing). Definitional nuances aside, the bigger takeaway, if accurate, may be the POD is cheaper for HPC workloads than the major cloud players.

HPCwire covered the earlier paper (see, <u>Azure Edges AWS in Linpack Benchmark Study</u>). In it, performances varied with Azure besting the pack, which included a NERSC run. The primary purpose of the study was to demonstrate whether or not HPC workloads can be reliably and cost-effectively run in the cloud. Using the high performance Linpack as the performance metric and a cost comparison with traditional infrastructure, the authors' answer is an unqualified yes.

"We benchmarked the performance of the best available computing hardware from public cloud providers with high performance Linpack. We optimized the benchmark for each computing environment and evaluated the relative performance for distributed memory calculations…Based on our findings we suggest that the concept of high performance computing in the cloud is ready for a widespread adoption and can provide a viable and cost-efficient alternative to capital-intensive on-premises hardware deployments," write the authors, Mohammad Mohammadi, Timur Bazhirov of Exabyte.io.

Penguin, of course, is well known for its HPC expertise and Tundra servers. In November the company had <u>seven systems</u> on the Top500. Not surprisingly, the POD offering is HPC-centric. "We recognize the need for high performance nodes with modern processors, bare metal, non-blocking fabric, the kind of infrastructure you would design if you were building a large scale HPC cluster in house," says Victor Gregorio, Sr. VP of Cloud Services.

POD provides two locations for cloud computing that can be accessed through a single POD Portal, MT1 and MT2. Each location has localized storage, and high-speed interconnects between the sites to facilitate easy migration of data from one location to another.

Login nodes and storage volumes (user home directories) are local to each location, but usernames are global to all of POD's locations. The POD Portal's usage reports are global to all POD locations.

The tests were run at the MT2 datacenter on POD's B30 nodes, which feature Intel Broadwell E5-2680 v4 processors (2.4Ghz, 14 cores, 16 double precision Flops/cycle) and deliver 1.07 Tflops peak per node. The MT2 location uses Intel OmniPath interconnect. The other clouds in the study do not have the current highest-end networking, although the Azure nodes (AZ-A and AZ-H) and NERSC Edison do employ high-speed interconnects (40Gbs InfiniBand and Cray Aries respectively) — details of the various node architectures/processors are best gleaned from the Exabyte.io paper.

As shown in the table and figures below, Penguin argues POD performs on par with the best of the big providers and that its costs are less, at least as evaluated in the study.

By this comparison, POD consistently performed at or near the top with performance scaling well with increased node/core count. The cost comparison is even more disparate, despite POD having a significantly higher initial rate per node, $2.80 per node hour versus the highest cost of the majors, $1.90/node/hour for the Azure IB-A node. Buyer beware: Penguin compared publicly disclosed rates and did not factor in potential discounts.

Source: Penguin Computing

Source: Penguin Computing
The cost reduction drivers, according to Penguin, include higher performance, reduced wall clock time, fine grain metering down to roughly three seconds versus rounding up to the nearest hour used, and the lack of many common additional charges such as data transfer, bandwidth, set-up, etc.

"You are not billed for idle time, you are only billed for compute time. But the biggest thing is the other cloud vendors will round up to the hour," he says. A job that ran 90 minutes would cost the same as one that ran 61 minutes or 62 minutes in billing schemes that round up. Most jobs, he notes, are not nicely divisible by hours. Penguin meters down to roughly three seconds of use. "When you look at what I feel is a real-world situation we are clearly more cost-effective than other providers for HPC workloads," Gregorio says.

Currently, there are five different queues on POD, each pointing to a cluster with different resources and capabilities. Lower performing environments are lower cost. That said, all of them have been designed for HPC workloads according to Gregorio.

Penguin doesn't say much about the size of the POD customer base. Sid Mair, SVP, Federal Systems, emphasized POD is the fastest growing part of the company and that its customers are not drawn only from existing Penguin customers for on-premises equipment and services. One POD customer is running multi-thousand core jobs, he says while there's a university with on the order of 300 students submitting "tiny" jobs daily. Weather forecasting, automotive, and traditional engineering disciplines are all represented. Gregorio considers a 4,000-core job to be a big one.

"Almost every HPC application on the market place runs on POD and many of them are already loaded in the environment we use. We have many relationships where customers transfer their corporate licenses right into POD rather than have to worry about managing them," says Gregorio, adding <u>the list</u> of applications on the web-site slightly lags their ongoing efforts add more applications and tools. Most of the familiar names are already there such as ANSYS, Dassult Systems, and MathWorks.

Perhaps surprisingly, the GPU node offering (NVIDIA K40) is modest when compared to recent aggressive adoption of K80 and P100 devices by the major cloud players. Gregorio says Penguin is demand-driven and able to scale as the need arises.

"As we see demand in the market place for things like deep learning, we'll adopt it. We are currently working with a number of customers for on-premises deep learning/machine learning environments and we're using that experience to understand the needs for optimal HPC environment in the cloud. We aren't ready to publicly disclose our plans yet," says Gregorio.

The post <u>Penguin Takes a Run at the Big Cloud Providers</u> appeared first on <u>HPCwire</u>.

Molex Becomes Annual Member of Open Networking Users Group

Wed, 04/12/2017 - 07:50

LISLE, Ill., April 12, 2017 — Molex, a leading global manufacturer of electronic solutions, has announced its membership in the 2017 Open Networking Users Group (ONUG). A coalition of industry leaders, IT end users and open source professionals, ONUG members leverage their collective expertise, engineering, manufacturing and procurement power to influence the pace and deployment of open networking solutions.

ONUG hosts an annual Spring and Fall Conference, Working Groups and Open IT Framework Industry Initiatives. Molex plans to attend the upcoming ONUG Spring 2017 Conference being held April 25th and 26th at the UCSF Mission Bay Conference Center in San Francisco.

The ONUG Spring 2017 Conference agenda features keynote addresses, panel discussions, real-world case studies, concept presentations, and a technology showcase designed to highlight interoperability demonstrations. Conference themes include building hybrid cloud infrastructure to address security, retooling operational teams, and restructuring the IT culture for new paradigms in cloud-based infrastructure.

"Layer 1 connectivity technology must keep pace as data rates, cloud deployments and data center growth continue to accelerate," says Ben Cammarata, regional industry manager, data centers, Molex. "We are pleased to join the ONUG community. Promoting open standards and interoperability drives value for customers by providing the right solutions in a timely manner."

Through innovation and industry partnerships, Molex strives to increase the availability of higher density, higher speed connectivity solutions that support next-generation networking, computing and storage architectures to 100G today with a roadmap to 400G and beyond. The company is an industry leader in R&D investment.

Developing solutions to meet tomorrow's demands, Molex supports a wide range of IT applications, including hyper scale and enterprise data centers, Web 2.0, big data, high performance computing (HPC) and embedded markets. Molex solutions are designed to provide high levels of reliability, quality and scalability, benefiting customers in the areas of computing and network performance, density, thermal management and power efficiency. To learn more, please visit www.molex.com/ind/datacenter.html.

About Molex

Molex brings together innovation and technology to deliver electronic solutions to customers worldwide. With a presence in more than 40 countries, Molex offers a full suite of solutions and services for many markets, including data communications, consumer electronics, medical, industrial, automotive, and commercial vehicle. For more information, please visit www.molex.com.

Source: Molex

The post Molex Becomes Annual Member of Open Networking Users Group appeared first on HPCwire.

DDN Claims Faster Object Storage 'Cross-Region' S3 Connectivity than AWS

Tue, 04/11/2017 - 16:37

Advanced-scale storage specialist DataDirect Networks (DDN), whose solutions populate the upper-rung of the Top500 supercomputing list, today announced new "cross-region" features to its WOS object storage platform that support multi-site AWS connectivity as well as new data protection and disaster recovery capabilities.

According to DDN, Extended ObjectAssure offers low-cost data protection along with control for local-only rebuilds the company said lowers risk of data loss during a rebuild and reduces drive failure impacts. In addition, DDN has expanded its AWS S3 multi-site capabilities in support of collaboration, disaster recovery and content distribution for customers that share data across multiple sites. DDN said its S3 interface is a faster and cheaper alternative to public clouds.

"We've done skunkworks-type research about Amazon S3's own replication," Michael King, DDN senior director of marketing, said, finding DDN has a 3X performance edge. "Important for us is to make this replication work well in a timely enough fashion that it provides the redundancy, resiliency and availability that's required for our customers."

Typically, King said, the use case for object storage is low cost storage on disk. But DDN customers are getting "real-world" WOS performance of up to 2.2Gb/second, he said, "so the idea of being a performant object storage system is where we're headed." Added Laura Shepard, DDN senior director of product marketing, "Object storage is really coming into its own with surging adoption across key use cases like large active archives, multi-site collaboration, content distribution, disk backup and disaster recovery."

DDN said the WOS architecture allows organizations to start with a single WOS appliance in four standard rack units and scale in single-node increments. It can deliver up to a quarter-million drives in a single, shared namespace and provide a single view of files and objects, supporting high-performance storage for active archive and collaboration environments.

"Most of our success with object storage is not in that cheap-and-deep kind of storage, but rather in an active archive (environment) that is so performant that it can help supplement Tier 1 storage," King said. "It (WOS) can pull data back so quickly to supplement semi-warm data that you didn't know if you'd need in an active archive; it pulls it back quickly enough that it makes everything work out. You can reduce the size of your Tier 1 and save money in the process."

DDN's object storage suite includes WOS as software-only along with several appliances. Extended ObjectAssure provides additional erasure coding and replication combinations, enabling WOS to be configured to rebuild failed drives within a single node. These features reduce rebuild time and risk of data

loss – capabilities the company said are increasingly important as drive sizes grow.

WOS Data Protection schemes (Source: DDN)
"The new digitized world demands infrastructure solutions that are extremely scalable and flexible in terms of delivery performance and connectivity models and that offer full data protection with a wide choice of user-selected protection methods," said Amita Potnis, research manager in IDC's Storage team. "With continual innovation in object storage technologies and features, DDN's object storage platform is leading end users along this path of digitization."

Users recently added to DDN's WOS customer list include Cold Springs Harbor, Van Andel and Children's Mercy Hospital of Kansas City. In addition, the company said, the University of Edinburgh's Parallel Computing Centre is evaluating WOS to support an infinitely scalable storage pool and to manage what will become a large active archive of research data, the goal being to offer a scalable, lower cost alternative to file storage to the university's industry users and Tier 3 HPC centers around the UK.

DDN said ObjectAssure is available this quarter across the WOS product line.

The post DDN Claims Faster Object Storage 'Cross-Region' S3 Connectivity than AWS appeared first on HPCwire.