



Penguin Computing Upgrades Corona with latest AMD Radeon Instinct GPU Technology for Enhanced ML and AI Capabilities

November 18, 2019

Fremont, CA., November 18, 2019 - [Penguin Computing](#), a leader in high-performance computing (HPC), artificial intelligence (AI), and enterprise data center solutions and services, today announced that *Corona*, an HPC cluster first delivered to Lawrence Livermore National Lab (LLNL) in late 2018, has been upgraded with the newest AMD Radeon Instinct™ MI60 accelerators, based on Vega which, per AMD, is the World's 1st 7nm GPU architecture that brings PCIe® 4.0 support. This upgrade is the latest example of Penguin Computing and LLNL's ongoing collaboration aimed at providing additional capabilities to the LLNL user community.

As previously released, the cluster consists of 170 two-socket nodes with 24-core AMD EPYCTM 7401 processors and a PCIe 1.6 Terabyte (TB) nonvolatile (solid-state) memory device. Each Corona compute node is GPU-ready with half of those nodes today utilizing four AMD Radeon Instinct MI25 accelerators per node, delivering 4.2 petaFLOPS of FP32 peak performance. With the MI60 upgrade, the cluster increases its potential PFLOPS peak performance to 9.45 petaFLOPS of FP32 peak performance. This brings significantly greater performance and AI capabilities to the research communities.

"The Penguin Computing DOE team continues our collaborative venture with our vendor partners AMD and Mellanox to ensure the Livermore Corona GPU enhancements expand the capabilities to continue their mission outreach within various machine learning communities," said Ken Gudenrath, Director of Federal Systems at Penguin Computing.

Corona is being made available to industry through LLNL's High Performance Computing Innovation Center (HPCIC). Funded through the Commodity Technology Systems (CTS-1) contract with the National Nuclear Security Administration (NNSA), the upgrade will help LLNL researchers and their industry partners improve capabilities in scalable deep learning, big data analytics and data science, while enhancing NNSA's ability to assess future architectures and meet the needs of the NNSA's Advanced Simulation & Computing program. It will also provide a higher level of performance for researching cognitive computing and developing predictive simulations for applications such as inertial confinement fusion and molecular dynamics simulations for precision medicine.

"This upgrade significantly increases the capability available on Corona," said Bronis R. de Supinski, Chief Technical Officer for Livermore Computing. "The new Vega GPUs offer substantial double-precision performance, in addition to much more single-precision performance. LLNL scientists will use the combination to understand the potential of mixed-precision algorithms for a variety of domains."

AMD's Radeon Instinct MI60 accelerators bring many new features that improve performance, including the Vega 7nm GPU architecture and the AMD Infinity Fabric™ Link technology, a peer-to-peer GPU communications technology that delivers up to 184 GB/s transfer speeds between GPUs – which is 5.75X faster than PCIe Gen 3, and full-chip Error-correcting code (ECC)11 and Reliability, Accessibility and Serviceability (RAS) 12 technologies. The new accelerators also utilize the latest ROCm open source software stack, which is now integrated into leading frameworks like TensorFlow and PyTorch and maps workloads to the heterogeneous compute resources of the underlying hardware.

"AMD is pleased to continue collaboration with LLNL and the NNSA in advancing open accelerator solutions. Access to systems like Corona enable next generation scientific discovery as we move to the exascale era," said Ogi Brkic, Corporate Vice President and General Manager of the Data Center GPU Business Unit at AMD.

For more information about AMD Radeon Instinct, please visit <https://www.amd.com/en/graphics/servers-radeon-instinct-mi>

About Penguin Computing

For 20 years, the Penguin Computing team of artificial intelligence (AI), engineering, and computer science experts has reimagined how startups, Fortune 500, government, and academic organizations solve complex technology challenges and achieve their organizational goals. Penguin Computing is focused on open platforms, including Open Compute Project (OCP) systems. We specialize in innovative on-premise high-performance computing (HPC), bare metal HPC in the cloud, AI, and storage technologies coupled with leading-edge design, implementation, hosting, and managed services including sys-admin and storage-as-a-service, and highly rated customer support. More information at www.penguincomputing.com.

Penguin Computing, Relion and FrostByte are trademarks or registered trademarks of Penguin Computing, Inc. Penguin Computing is a subsidiary of SMART Global Holdings Inc., (NASDAQ: SGH)

PCIe is a registered trademark of PCI-SIG.

AMD, the AMD logo, EPYC, Infinity Fabric, Radeon Instinct, and combinations thereof are trademarks or registered trademarks of Advanced Micro Devices, Inc.

Penguin Computing Media Contact

Karbo Communications

Madeline Kalicka

penguin@karbocom.com