NEWS RELEASE

Penguin Computing Announces Support for New 24GB NVIDIA Tesla M40 GPU Accelerators and Introduces Latest OpenPOWER-based Magna Servers

2016-04-05

Significant Activity for GPU Technology Conference and OpenPOWER Summit
SAN JOSE, CA - NVIDIA GPU Technology Conference/OpenPOWER Summit, April 4, 2016…Penguin Computing, a provider of high-performance computing, enterprise data center and cloud solutions, announced Open Compute Project (OCP)-based systems that reinforce both its continued collaboration with NVIDIA and new options in Penguin Computing's Magna family of OpenPOWER-based servers.

"Customers benefit when we partner with exceptional organizations like NVIDIA, the OpenPOWER Foundation and Open Compute Foundation in developing our systems," said Jussi Kukkonen, Director Product Management, Penguin Computing. "An essential part of our mission is to provide customers with form factor flexibility, choice of architecture and peak performance, which are all hallmarks of Penguin Computing."

Penguin Computing introduced the company's latest systems based on OpenPOWER architecture at the OpenPOWER Summit. The Penguin Magna 2002 combines the dual processor OpenPOWER platform with the NVIDIA® Tesla® Accelerated Computing Platform in a conventional EIA form factor. This new architecture option is a demonstration of the company's continuing commitment and investment in accelerated computing and customer choice.

NVIDIA's Tesla M40 GPU, the most powerful accelerator designed for training deep neural networks, now provides 24GB RAM of GDDR5 memory. It is being validated on all Penguin Computing GPU host platforms, including both Intel x86 and OpenPOWER host architectures. Penguin Computing provides optimized systems for accelerated computing, ranging from 1:1, 1:2 and 1:4 ratio of CPUs to GPUs.

Penguin Computing also announced support for the NVIDIA Tesla M4 GPU accelerator in its OCP-based Tundra ES 1930g open compute server. The Tesla M4 GPU is a low-power, small form-factor accelerator for deep learning inference, as well as streaming image and video processing.

“Our hyperscale accelerator line enables developers to drive deep learning development in large data centers and create new classes of applications for artificial intelligence,” said Roy Kim, group product manager of Accelerated Computing at NVIDIA. “Penguin Computing offers rich deployment options for NVIDIA GPU technologies, including high-density, low TCO platforms supporting the Tesla M4 GPU, and systems with memory and I/O subsystem scalability designed for developing deep neural networks with our Tesla M40 GPUs.”

Visit www.penguincomputing.com or contact your local Penguin Computing Representative for more information and product availability on these Penguin Computing systems. Visit Penguin Computing’s booth #510 at the NVIDIA GPU Technology Conference (GTC) and booth #1409 at the co-located OpenPOWER Summit.