FOR IMMEDIATE RELEASE

Contact:
Mike LaPan
Cirrascale Corporation
(858) 874-3800
mike.lapan@cirrascale.com

CIRRASCALE® EXPANDS MULTI-GPU DEEP LEARNING SOLUTIONS WITH SUPPORT FOR LATEST ADDITION TO NVIDIA® TESLA® ACCELERATED COMPUTING PLATFORM

The Cirrascale GX Series Rackmount and GB5600 Blade Server support up to eight NVIDIA Tesla M40 GPU accelerators with 24GB GDDR5 memory, the world's fastest deep learning training accelerator

San Diego, CA -- April 1, 2016 -- Cirrascale Corporation®, a premier developer of blade and rackmount solutions enabling GPU-driven deep learning infrastructure, today announced it will offer the new NVIDIA® Tesla® M40 GPU accelerators with 24GB GDDR5 memory throughout its high-performance, deep learning product lines.

Utilizing the company's proprietary 96-lane Gen3 PCIe switch-enabled risers, the GX Series and GB5600 Series product lines can attach up to eight discrete NVIDIA Tesla M40 GPUs on the same PCIe root complex in a single rackmount or blade server chassis, with additional room for other Gen3 PCIe x16 devices such as InfiniBand® or NVMe cards.

"Our concentration with current customers often involves helping to build their deep learning training infrastructure utilizing both our hardware and cloud services because of our unique and powerful multi-GPU peering abilities," said PJ Go, president, Cirrascale Corporation. "The Tesla M40 GPU is the world's fastest deep learning training accelerator, and now with double the memory, it will enable us to create some of the fastest training solutions for larger, more accurate deep neural networks."

Extending the capabilities of these accelerators, the Cirrascale SR3615 PCIe switch riser enables up to 10 PCIe Gen3 x16 compatible devices -- such as GPU accelerators, InfiniBand network or NVMe storage cards -- to communicate directly with each other on the same PCI root complex. This eliminates the need for host CPU intervention by allowing the accelerators to share a single memory address space and make use of DMA to control data movement. When used in conjunction with NVIDIA GPUDirect™ technology, compatible PCIe Gen3 x16 devices can directly read and write CUDA host and device memory, including memory owned by network and storage devices. By doing so, it eliminates unnecessary memory copies, dramatically lowers CPU overhead, and reduces latency resulting in significant performance improvements in data transfer times.

"Today's larger, more sophisticated deep neural networks require far more GPU memory to handle ever-expanding volumes of training data," said Roy Kim, product manager of Accelerated Computing at NVIDIA. "Cirrascale's purpose-built solutions, with support for the new Tesla M40 accelerators with double the memory, will help researchers and data scientists develop better, more accurate artificial intelligence applications for image and video object recognition, natural language processing, and more."

The Cirrascale GX Series rackmount and GB5600 Series blade servers supporting the NVIDIA Tesla M40 GPU accelerators -- as well as the Cirrascale proprietary PCIe switch-enabled riser -- are immediately available to order and are shipping to customers now. Licensing opportunities for these technologies are also available immediately to both customers and partners.

About Cirrascale Corporation

Cirrascale Corporation is a premier developer of hardware and cloud-based solutions enabling GPU-driven deep learning infrastructure. Cirrascale leverages its patented Vertical Cooling Technology and proprietary PCIe switch riser technology to provide the industry's densest rackmount and blade-based peered multi-GPU platforms. The company sells hardware solutions to large-scale deep learning infrastructure operators, hosting and cloud service providers, and HPC users. Cirrascale also licenses its award winning technology to partners globally. To learn more about Cirrascale and its unique multi-GPU infrastructure solutions, please visit http://www.cirrascale.com or call (888) 942-3800.

# # #