



FOR IMMEDIATE RELEASE

Contact:

Mike LaPan
Cirrascale Cloud Services
(888) 942-3800
mike.lapan@cirrascale.com

CIRRASCALE CLOUD SERVICES BROADENS DEEP LEARNING CLOUD OFFERINGS WITH WORLD'S MOST POWERFUL GPU FOR AI SUPERCOMPUTING

Company adds the NVIDIA A100 80GB and A30 GPUs to its burgeoning deep learning cloud for development, training, and inference workloads.

San Diego, CA -- June 3, 2021 -- Cirrascale Cloud Services®, a premier cloud services provider of deep learning infrastructure solutions for autonomous vehicles, natural language processing, and computer vision workflows, today announced its dedicated, multi-GPU deep learning cloud servers support the NVIDIA® A100 80GB and A30 Tensor Core GPUs. With record-setting performance across every category on the latest release of MLPerf, these latest offerings provide enterprise customers with mainstream options for a broad range of AI inference, training, graphics, and traditional enterprise compute workloads.

"Model sizes and datasets in general are growing fast and our customers are searching for the best solutions to increase overall performance and memory bandwidth to tackle their workloads in record time," said Mike LaPan, vice president, Cirrascale Cloud Services. "The NVIDIA A100 80GB Tensor Core GPU delivers this and more. Along with the new A30 Tensor Core GPU with 24GB HBM2 memory, these GPUs enable today's elastic data center and deliver maximum value for enterprises."

The NVIDIA A100 80GB Tensor Core GPU introduces groundbreaking features to optimize inference workloads. It accelerates a full range of precision, from FP32 to INT4. Multi-Instance GPU (MIG) technology enables up to 7 instances with up to 10GB of memory to operate simultaneously on a single A100 for optimal utilization of compute resources. Structural sparsity support delivers up to 2X more performance on top of the A100 GPU's other inference performance gains. A100 provides up to 20x higher performance over the NVIDIA Volta® and on modern conversational AI models like BERT Large, A100 accelerates inference throughput by 100x over CPUs.

Also available through Cirrascale Cloud Services is the NVIDIA A30 Tensor Core GPU, which delivers versatile performance supporting a broad range of AI inference and mainstream enterprise compute workloads, such as recommender systems, conversational AI and computer vision. The A30 also supports MIG technology, delivering superior price/performance with up to 4 instances containing 6GB of memory, perfectly suited to handle entry-level applications. Cirrascale's accelerated cloud server solutions with NVIDIA A30 GPUs provide the needed compute power -- along with large HBM2 memory, 933GB/sec of memory bandwidth, and scalability with NVIDIA NVLink® interconnect technology -- to tackle massive datasets and turn them into valuable insights.

"Customers deploying the world's most powerful GPUs within Cirrascale Cloud Services can accelerate their compute-intensive machine learning and AI workflows better than ever," said Paresh Kharya, senior director of Product Management, Data Center Computing at NVIDIA.

Customers are already using the NVIDIA A100 80GB GPUs on the Cirrascale Cloud Services platform and the A30 will be available for use by end of Q2 2021. Interested customers and partners can visit <https://cirrascale.com> or call (888) 942-3800 to sign up for service.

About Cirrascale Cloud Services

Cirrascale Cloud Services is a premier provider of public and private dedicated cloud solutions enabling deep learning workflows. The company offers cloud-based infrastructure solutions for large-scale deep learning operators, service providers, as well as HPC users. To learn more about Cirrascale Cloud Services and its unique cloud offerings, please visit <https://cirrascale.com> or call (888) 942-3800.

Cirrascale Cloud Services, Cirrascale and the Cirrascale Cloud Services logo are trademarks or registered trademarks of Cirrascale Cloud Services LLC. NVIDIA, the NVIDIA logo, and NVLink are trademarks or registered trademarks of NVIDIA Corporation. All other names or marks are property of their respective owners.