Denvr Dataworks expands NVIDIA H100 GPU cluster, enabling enhanced levels of AI performance, scalability, and innovation

Denvr Cloud expansion enables businesses and researchers to unlock the power of Generative AI and Large Language Models


Denvr Cloud Nvidia H100 Cluster

NEW YORK, NY, UNITED STATES, October 11, 2024 /EINPresswire.com/ -- Denvr Dataworks a leading provider of solutions for the development and operations of AI, today announced the recent expansion of its NVIDIA H100 Tensor Core GPU cluster for customers on Denvr Cloud, scaling its cloud services capacity for AI workloads. This expanded offering enables AI developers, operators, and integrators to achieve enhanced levels of AI performance, scalability, and innovation.

Designed to meet the growing demand for generative AI model training and inference, the NVIDIA H100 cluster on Denvr Cloud features NVIDIA HGX H100 servers with 1,024 Hopper architecture-based SXM5 GPUs, connected with NVIDIA Quantum-2 InfiniBand networking. The expanded NVIDIA H100 capacity is optimized for developing, training, and inferencing AI workloads:

- Generative AI (GenAI)
- Deep learning training
- AI-driven simulations and digital twins
- Multi-modal AI workloads

Qualified customers can experience the capabilities of NVIDIA H100 on Denvr Cloud firsthand through an exclusive 1-week test-drive program. This program allows businesses to explore and validate NVIDIA H100 GPUs on Denvr Cloud, test their workloads, and uncover new growth

opportunities.

“Denvr Dataworks is dedicated to empowering our customers to unlock their AI innovation,” said Geoff Gordon, CEO of Denvr Dataworks. “We continue our close collaboration with NVIDIA and Supermicro through the added capacity of NVIDIA H100 GPUs on Denvr Cloud.”

Key benefits of the NVIDIA H100 cluster on Denvr Cloud include:

- High computing power for large-scale AI model training
- Exceptional real-time AI inference performance
- Scalability to meet growing workload demands
- Single-click deployment and easy management
- Access to NVIDIA AI software stack
- Expertise and support from Denvr Dataworks

Key Features:

 4th-generation Tensor Cores: Designed for AI workloads, the NVIDIA H100 has enhanced precision support (FP8) and can handle large-scale matrix multiplications required by AI models.
 Higher memory bandwidth: The NVIDIA H100 has up to 3 TB/s memory bandwidth for working with large datasets and models.
 Provides high-speed NVIDIA NVLink interconnect technology for scaling across multiple GPUs within a node.
 Provides up to 3200 Gb/s of aggregate bandwidth per node with NVIDIA Quantum-2 InfiniBand networking for scaling beyond a node.
 FP8, FP16, and INT8 support: Enabling faster computation without significant loss in accuracy.

Don’t miss this opportunity to revolutionize your AI capabilities. Apply for the 1-week test-drive program today and experience the transformative power of the NVIDIA H100 cluster on Denvr Cloud.

About Denvr Dataworks

Denvr Dataworks provides cloud services for the development and operations of AI technologies, including Inference-as-a-Service offerings, with options for accessing on-demand or dedicated compute instances, or hardware, that are designed for AI workloads.

Marc Kronewitt
Denvr Dataworks
+1 780-288-0143
marc@denvrdata.com
Visit us on social media:

X
LinkedIn