



NVIDIA



GENESIS CLOUD SELECTS SUPERMICRO'S POWERFUL AI SERVERS FOR NEW OFFERINGS

Genesis Cloud Offers AI and ML Services Using Supermicro Advanced AI Training System Featuring NVIDIA HGX H100 GPUs



INDUSTRY

Cloud Service Provider

Introduction

Providing cloud services that accelerate specific workloads at scale and more optimally than previously available allows customers to both explore new capabilities for their enterprises and quickly put them into production environments. Genesis Cloud enables customers to accelerate HPC and AI workloads with the latest generation of GPUs from NVIDIA, the HGX H100 GPUs. Facilitating fast AI training and machine learning for customers results in more effective use of servers, lower costs, and a greater understanding of data.

CHALLENGES

- Adding Large Scale AI Training to Offerings
- Increased ML Abilities
- Enterprise Class Servers and AI Hardware

Challenges

Genesis Cloud has been a leader in offering high performance servers, storage, and networking for a range of customers who require the latest technologies. As part of a business expansion that addresses customers' requirements to accelerate AI and Machine Learning, Genesis Cloud investigated a number of AI server solutions. The requirement was fast processing of ML benchmarks in an air cooled server. In addition, the servers needed to be expandable in processing power, which required a design that allowed multiple AI-specific processors to be connected to work outside a single server. This results in applications being able to address up to hundreds of AI processors to complete enormous tasks in multi-node environments.

SOLUTION

Supermicro GPU Servers

SYS-821GE-TNHR

- Dual 4th Gen Intel Xeon Scalable Processors
- NVIDIA HGX H100 8 GPUs
- 2TB Memory per server
- NVIDIA BlueField-3 DPU

Solution

After significant benchmarking and system comparison, Genesis Cloud decided that the Supermicro AI Training SuperServer SYS-821GE-TNHR includes dual 4th Gen Intel Xeon Scalable processors and the NVIDIA HGX H100 GPUs. The systems Genesis Cloud selected each contain eight NVIDIA HGX H100 GPUs networked together with the NVIDIA NVSwitch™.



The NVIDIA HGX H100 GPUs increases training and inference performance compared to any other GPU in the market. The Supermicro SYS-821GE-TNHR that Genesis Cloud is using for its AI cloud includes:

- Dual 4th Gen Intel Xeon Scalable processors, the Intel Xeon 8480+
- 2TB Memory DDR5-4800MHz Memory
- NVIDIA HGX H100 8-GPU
- NVIDIA BlueField-3 DPUs (specific systems)
- 30TB NVMe Storage
- Dual Titanium Power Supplies

BENEFITS

- New Offerings
- Large Scale Generative AI
- High End Workloads
- Fastest GPU Available Reducing Time to Completion

NVIDIA HGX H100 combines H100 Tensor Core GPUs with high-speed interconnects to form the world's most powerful servers. With up to eight HGX H100 GPUs, the HGX H100 has up to 640 gigabytes (GB) of GPU memory and 24 terabytes per second (TB/s) of aggregate memory bandwidth for unprecedented acceleration. NVIDIA HGX H100 delivers a staggering 32 petaFLOPS, forming the world's most powerful accelerated scale-up server platform for AI and HPC

Genesis Cloud opted for three configurations of the Supermicro SYS-821GE-TNRT servers for general use, bare metal use, and heavy networking use. All of the systems contain the NVIDIA HGX H100 8-GPU GPUs.

Benefits

Genesis Cloud was immediately able to offer customers a high end AI/ML server to perform intensive workloads. These include higher performance in the areas of:

- General ML
- Foundational Models / Generative AI
- Image segmentation
- Computer vision (home, agriculture, construction)
- Large Language Models
- Pseudo Quantum Computing

"Genesis Cloud, using the most powerful Supermicro AI servers available, continues to offer customers the most advanced AI Training servers available today. Our data centers are powered by green energy, allowing users fast, low latency access to the Supermicro GPU servers with the latest NVIDIA HGX H100 8-GPU hardware. We continue to work closely with Supermicro to offer our customers the most advanced GPU servers in the industry".

Dr. Stefan Schiefer, CEO, of Genesis Cloud

SUPERMICRO

Supermicro is a global leader in high performance, green computing server technology and innovation. We provide our global customers with application-optimized servers and workstations customized with blade, storage, and GPU solutions. Our products offer proven reliability, superior design, and one of the industry's broadest array of product configurations, to fit all computational need.

For more information, visit
<https://www.supermicro.com>

GENESIS CLOUD

Genesis Cloud is an Accelerated Cloud Service Provider offering the latest NVIDIA reference architecture. We make cutting-edge accelerated computing more affordable and secure at enterprise scale. We offer best-in-class infrastructure via our network of green HPC data centers, where cutting-edge technology meets sustainability. Our data centers In Europe and North America boast the latest AI accelerators and GPUs and a robust network and storage solution, specifically designed to supercharge GenAI, Machine Learning, Simulations, Rendering, and other high-performance computing workloads.

To learn more, visit the company's website:
<https://www.genesiscloud.com>