



APPLIED DIGITAL BUILDS MASSIVE AI CLOUD WITH SUPERMICRO GPU SERVERS

Applied Digital Offers Users the Latest In Scalable AI and HPC Infrastructure For AI Training and HPC Simulations with Supermicro High Performance Servers



Supermicro SYS-821GE-TNHR

Introduction

Applied Digital Corporation (Nasdaq: APLD) is a designer, builder, and operator of next-generation digital infrastructure designed for Artificial intelligence (AI) and High-Performance Computing (HPC) applications, cloud services, and data center hosting. With many data centers located across North America, Applied Digital is a leading cloud service provider that works closely with its customers to offer the latest in server technology with the latest GPUs. Applied Digital’s Cloud Services, offered through its wholly owned subsidiary Sai Computing, provide bare metal and supercomputing-as-a-service for workloads requiring no-compromise performance at a large scale.

In many cases, Applied Digital locates its data centers close to renewable energy, which reduces the carbon footprint associated with high end computing. In addition, Applied Digital is committed to operational efficiency to keep costs low, using renewable energy when available, and providing white glove service when working with customers to solve their unique needs.

Challenges

The Applied Digital HPC team has deep experience in the architecture, deployment, and optimization of parallel computing clusters for the latest server technology for a range of applications. With expertise in offering hardware and services for the most demanding compute intensive applications, Applied Digital was in a unique position to offer the AI/ML market the additional computing capacity that is needed, which continues to be scarce.

INDUSTRY

AI Cloud Service Provider

Data Center Colocation Services

CHALLENGES

Adding Large Scale AI Training to Offerings

Increased AI and ML Abilities

Enterprise Class Servers and AI Hardware

Exceeding performance, value, and service over traditional cloud HPC offerings

SOLUTION

Supermicro Servers

SYS-821GE-TNHR

- Dual 4th Gen Intel Xeon Scalable Processors
- 8x NVIDIA HGX H100 8-GPUs
- 2TB Memory
- 19.2TB Storage
- 3.2Tbps Infiniband per node

AS -4124GS-019

- Dual AMD EPYC 7513 CPUs
- 1TB Memory
- 8x NVIDIA A100 80GB GPUs

SYS-111E-WR

- Single 4th Gen Intel Xeon Scalable Processor (5415+)
- 64GB Memory

AS -2014TP-HTR

- Single AMD EPYC 7534P CPU
- 1TB Memory

Applied Digital needed to evaluate which servers were available that would allow their customers to train their AI models and run HPC applications that were available in quantity in short timeframes. The requirements for the new servers were to have the most performant GPUs with fast interconnects between the GPUs. This architecture allows for quick training of AI models without having the CPU involved at every step. These servers are subsequently scaled over a 3.2Tbps NDR GPU-Direct RDMA fabric, allowing thousands of GPUs to operate in parallel.

To give customers an environment where they can execute their AI and HPC applications, Applied Digital needed to partner with a systems vendor to acquire the latest technology that could supply the best AI servers on the market.

Solution

Applied Digital decided the most optimal system for their users would be the Supermicro SYS-821GE-TNHR with dual 4th Gen Intel® Xeon® platinum processors, the 8462Y+. These servers use NVIDIA HGX H100 GPUs, each with 80GB of memory. The NVIDIA H100 delivers 67 teraflops of FP64 Tensor Core computing for HPC, and AI-fused HPC applications can leverage H100's TF32 precision to achieve one petaflop of throughput for single-precision matrix-multiply operations. The system hosts eight H100 Tensor Core GPUs and 900GB/s NVSwitch for GPU-to-GPU communication inside the compute node. Applied Digital opted for 2TB of system RAM to stage workloads prior to moving to GPU memory. For networking, Applied Digital uses 100GbE for in-band management and object storage and NDR fabric for GPU Direct and converged flash file system traffic. Leveraging the NVIDIA DGX reference architecture, Applied Digital scales to thousands of H100 GPUs working in a single parallel compute cluster.

In addition to the Supermicro GPU servers that Applied Digital acquired, several other Supermicro systems are part of the infrastructure. These include:

- AS -2014TP-HTR (cluster infrastructure)
- SYS-111E-WR for UFM (cluster infrastructure)
- AS -4124GS-TNR with PCIE GPU (non-clustered general purpose GPU cloud)



Switching is also a critical piece of an AI infrastructure solution. Applied Digital also acquired the following Supermicro switches, which are matched to the workloads that Applied Digital customers are performing:

- SSE-SN3700-CS2FC (32x port 100GbE switch)
- SSE-F3548SR (48x port 10/25GbE switch)
- 3SSE-G3748R-SMIS (48x port 1GbE switch)

Benefits

Applied Digital immediately realized the benefits of using the Supermicro 8 GPU servers. With the AI training market exploding, Applied Digital was able to offer new services for AI training and HPC simulations. By giving customers access to the most advanced and powerful computing capability available today, HPC and AI applications run faster, and more complex AI pipelines can be used. Besides offering hardware-based solutions, Applied Digital experts can work with customers to determine the software that is most optimized for their particular workloads.

“Applied Digital takes pride in providing some of the most advanced AI clusters to the market for AI and HPC workloads. We work closely with Supermicro to deliver a range of systems to give our customers the most advanced and performant solutions to speed up their demanding workloads.”

Michael Maniscalco, CTO, Applied Digital

BENEFITS

- Additional AI Offerings
- Reduced Time for Executions of AI Workloads

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>

APPLIED DIGITAL

Applied Digital (Nasdaq: APLD) designs, develops, and operates next-generation data centers across North America to provide digital infrastructure solutions to the rapidly growing high-performance computing (HPC) industry.

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