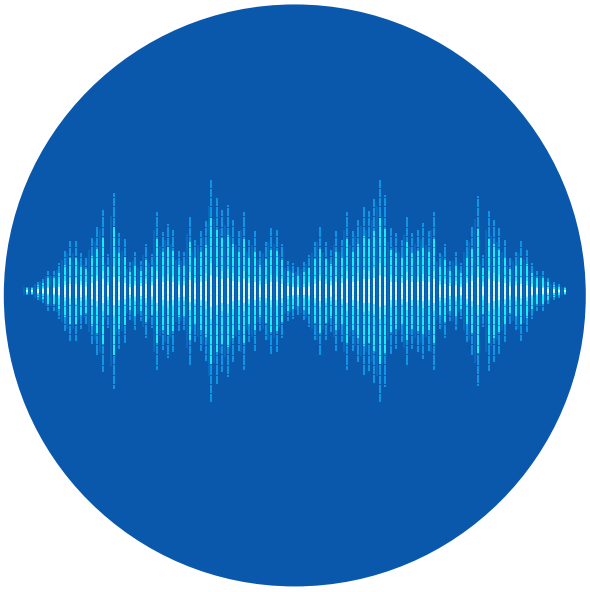
Case Study

Leading GenAI Developer **Trains Advanced Voice Models** on AMAX Deployed NVIDIA DGX SuperPOD



Company
Confidential

Use Case
Generative AI / LLMs

Deployment
NVIDIA DGX SuperPOD™ with NVIDIA DGX™ B200

Services
Data Center Planning, Rack & Stack Deployment, Networking, Cabling, Bring-up, Testing, System Validation

Performance
4.6 exaflops AI training
9.2 exaflops inference

Background

A fast-growing generative AI company developing voice synthesis and multimodal content tools needed infrastructure that could keep pace with their expanding model complexity. As their team moved to scale training and inference for larger models, the company required a high-density, high-bandwidth solution that could grow with them.

AMAX worked closely with the customer's engineering and operations teams to assess compute requirements, model growth patterns, and long-term infrastructure goals. Based on performance targets and deployment timelines, AMAX recommended the NVIDIA DGX SuperPOD™ platform, built with NVIDIA DGX™ B200 systems, as the most scalable solution.

Challenge

The customer needed more than raw performance. They required a platform that could handle training and fine-tuning at scale while keeping workflows running smoothly. This meant:

- Support the rapid training of new voice and content generation models
- Scale as they introduced more features and parameters
- Deliver predictable performance for both training and inference

Solution

AMAX delivered a complete NVIDIA DGX SuperPOD deployment consisting of 64 DGX B200 systems with 512 NVIDIA Blackwell GPUs. The systems were interconnected using NVIDIA Quantum-2 InfiniBand to provide 400 gigabit per second networking across the entire cluster.

AMAX designed the rack layout and power distribution plan in line with NVIDIA's DGX SuperPOD reference architecture. The team managed the full physical build, including integration and structured cabling. Once online, every system was benchmarked and tested under load to ensure reliability from day one. NVIDIA AI Enterprise was provisioned to support model training and orchestration.

> "With AMAX's expertise and experience in deploying NVIDIA DGX SuperPOD, organizations can gain the computational power and scalability needed to power the full lifecycle of AI, from training to inference, helping businesses deliver the ROI of AI at unprecedented speed."

Tony Paikeday
Senior Director
of AI systems at NVIDIA

Outcome

The new DGX SuperPOD provides the customer with 4.6 exaflops of AI training and 9.2 exaflops of inference performance. Working with AMAX, the customer gained a high-performance foundation for training large-scale models, with the flexibility to support future growth as their AI workloads evolve.

Resources

The Infrastructure Behind NVIDIA DGX Systems
Press Release

AMAX | 1565 Reliance Way, Fremont, CA 94539 | 1 (800) 800-6328 | www.amax.com | sales@amax.com