

AI at Any Scale Portfolio

Scale Smarter with AMAX and NVIDIA

NVIDIA DGX™ System

For fine-tuning and inference

Enterprise rackmount node ideal for running LLMs in production or refining pretrained models.

Up to 144 PFLOPS



NVIDIA DGX™ Spark

For AI prototyping

Desktop system for testing, debugging, and early-stage model development.

Up to 2 PFLOPS (1 PFLOP per system)



NVIDIA DGX BasePOD™

For scaling teams and shared training

Multi-node platform for collaborative development and mid-scale training jobs.

Up to 1152 PFLOPS

NVIDIA DGX SuperPOD™

For training massive models

Rack-scale system built for full-scale training, foundation models, and AI infrastructure at enterprise scale.

Up to 18,432 PFLOPS



Performance values are based on FP4 precision. NVIDIA DGX System, DGX BasePOD, and DGX SuperPOD metrics reflect NVIDIA DGX B200 systems.