IEI® AI Server

NF5688G7: Hyperscale Training Platform

NF5688G7, the 6U hyperscale training platform equipped with dual 4th Gen Intel Xeon Scalable Processors or AMD EPYC™ 9004 Series Processors and 8x NVIDIA latest GPUs, features industry-leading performance, ultimate I/O expansion, and ultrahigh energy efficiency. The precisely optimized system architecture with 4x CPU to GPU bandwidth, up to 4.8Tbps networking bandwidth, 8TB system memory, and 300TB massive local storage can fully satisfy the communication and capacity demands of multi-dimensional parallelism training for giant-scale models. 12 PCIe expansions can be flexibly configured with CX7, OCP3.0, and multiple SmartNICs, making it an ideal solution for both on-premises and cloud deployment. It is built to handle the most demanding AI computing tasks like trillion-parameter Transformer model training, massive recommender systems, AIGC, and Metaverse workloads.

Benefits

■ Unprecedented performance

- ✓ Powered by 8* NVIDIA latest GPUs in a 6U chassis, TDP up to 700W
- ✓ Support 2x 4th Gen Intel Xeon Scalable Processors or AMD EPYC™ 9004 Series Processors
- ✓ Industry-leading performance with 16PFlops AI performance by 3 times enhancement

■ Leading architecture design

- ✓ Lightning-fast intra-node connectivity with 4x CPU to GPU bandwidth improvement
- ✓ Ultra-high scalable inter-node networking with up to 4.8Tbps non-blocking bandwidth
- ✓ Cluster-level optimized architecture, GPU : Compute Network : Storage Network = 8:8:2

■ Optimized energy efficiency

- ✓ Extremely low air-cooled heat dissipation overhead, less fan, higher power efficiency
- ✓ 54V, 12V separated power supply with N+N redundancy reducing power conversion loss
- ✓ Direct liquid cooling design with more than 80% cold plate coverage, PUE≤1.15

■ Multi-scenarios adaptation

- ✓ Full modular design and extremely flexible configurations satisfying both on-premises and cloud deployment
- ✓ Easily harness large-scale model training, such as GPT-3, MT-NLG, stable diffusion and Alphafold.
- ✓ Diversified SuperPod solutions accelerating the most cutting-edge innovation including AIGC, AI4Science and Metaverse.

Product Specifications

Item	NF5688-M7-A0-R0-00	NF5688-A7-A0-R0-00
GPU	NVIDIA HGX Hopper/Ampere 8-GPU, TDP up to 700W per GPU	
Processor	2*4th Gen Intel Xeon Scalable Processors, TDP 350W	2* AMD EPYC™ 9004 Series Processors, TDP 400W
Memory	32 * DDR5 DIMMs, 4800MT/s	24 * DDR5 DIMMs, 4800MT/s
Storage	24 * 2.5’ SSD, up to 16 * NVMe U.2	
M.2	2* NVMe/SATA M.2	2* NVMe M.2
PCIe Slot	Up to 12 * PCIe x16 Gen5 slots Support Bluefield-3, CX7, and various SmartNICs	
Front I/O	1*USB 3.0, 1*USB 2.0, 1*VGA	
Rear I/O	1*USB3.0, 1*RJ45, 1* MicroUSB, 1* VGA	
OCP	1*OCP 3.0, Support NCSI	
Management	DC-SCM BMC management module with Aspeed 2600	
TPM	TPM 2.0	
Fan	6 hot-swap fans with N+1 redundancy	
Power	2* 12V 3200W and 6* 54V 2700W, Platinum/Titanium PSU with N+N redundancy	
Size	6U chassis, 447mm*263mm*860mm (W × H × D)	
Environmental Parameters	Working temperature：10℃~35℃； Storage temperature：-40℃~70℃	
	Working humidity：10%~80% R.H.； Storage humidity：10%~93% R.H.	
	The working temperature is 0 ~ 40 ℃ at 0 to 1000 meters (3300 feet); Operating temperature 5 ~ 32 ℃ from 1000 to 3050 m (10000 feet)	