![AiVRES]

KR6288V2

6U AI Server with HGX H100 8-GPU & AMD EPYC™ 9004

KR6288V2 with AMD EPYC™ 9004 processors is an advanced AI system supporting NVIDIA HGX H100 8-GPUs. This server delivers industry-leading 16PFlops of AI performance and lightning-fast CPU-to-GPU interconnect bandwidth, with the H100 Transformer Engine supercharging training speeds for GPT large language models. Its optimized power efficiency and modular design with flexible configuration makes it ideal for the most demanding AI tasks in various scenarios like hyperscale data centers, AI model training, and metaverse workloads.

- Powered by NVIDIA HGX H100 8-GPUs in 6U
- 2x AMD EPYC™ 9004
- Delivers 16 PFlops industry-leading AI performance
- Direct liquid cooling design available with over 80% cold plate coverage

AMD

Unprecedented AI Performance

- Powered by 8 NVIDIA latest GPUs in 6U, TDP up to 700W
- 2x AMD EPYC™ 9004 processors
- 16PFlops of industry-leading AI performance
- H100 Transformer Engine delivers supercharged training speed for GPT large language models

Optimized Energy Efficiency

- Low air-cooled heat dissipation overhead and high power efficiency
- 54V, 12V separated power supply with N+N redundancy reducing power conversion loss
- Direct liquid cooling design with over 80% cold plate coverage keeps PUE ≤1.15

Leading Architecture Design

- Lightning-fast CPU-to-GPU interconnect bandwidth
- Ultra-high scalable inter-node networking with up to 4.0 Tbps non-blocking bandwidth
- Optimized cluster-level architecture with 8:8:2 ratio of GPU to compute network to storage network

Flexible Configurations for AI Scenarios

- Fully modular design and flexible configurations satisfy 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

Technical Specifications

Model	KR6288-E2-A0-R0-00
Form Factor	6U rack server
GPU	NVIDIA HGX H100 8-GPUs, TDP up to 700W
CPU	2x AMD EPYC™ 9004 processors, TDP 400W
Memory	24x DDR5 DIMMs, up to 4800MT/s
I/O	**Front**: 1x USB 3.0, 1x USB 2.0, 1x VGA **Rear**: 2x USB 3.0, 1x RJ45, 1x MicroUSB, 1x VGA
OCP	Supports 1x OCP 3.0 (optional), supports NCSI
Management	DC-SCM BMC management module with Aspeed 2600
Security	TPM 2.0 (Trusted Platform Module)

Technical Specifications Continued

Storage	24x 2.5" SSD, up to 16x NVMe U.2; 2x Onboard NVMe M.2 (optional)
PSU	2x 12V 3200W and 6x 54V 2700W, Platinum/Titanium PSU with N+N redundancy
Dimensions	447mm (W) x 263mm (H) x 860mm (D)

Ver 1.0