AiVRES

KR6298-X2

6U Extreme AI Server with Intel Habana® Gaudi®2 8-OAM

KR6298-X2 is an advanced 6U AI system supporting 8 Habana® Gaudi®2 accelerators and 2 Intel® Xeon® processors. This server delivers the massive memory capacity and computational power necessary for demanding AI workloads and HPC applications, as well as large-scale AI deployments in cloud environments and data centers.

- 8x fully-connected Habana® Gaudi®2
- 2x 4th Gen Intel® Xeon® Scalable processors
- Up to 32x DDR5, 8TB maximum system memory
- Unprecedented performance to support extra-large-scale training models, gen-AI, mainstream AI frameworks, and development tools

Extreme Memory, High Performance

- **Habana® Gaudi®2:** 96GB HBM3 memory capacity, 4.2TB/s converged P2P bandwidth
- **AMD Instinct™ MI300X:** 192GB HBM3 memory capacity, 5.3 TB/s peak memory bandwidth, TDP 750W

Supports AI & HPC Workloads

- LLM training, fine-tuning and inference; generative AI for text, content, image, video
- Scientific simulations, weather forecasting

Open Ecosystem

- Supports mainstream AI frameworks and AI development tools
- Mature open ecosystem enables cost-efficient AI compute solution development

Technical Specifications

Model	KR6298-X2
Form Factor	6U rack server
OAM	8x Habana® Gaudi®2 HL225B, 600W
Processor	2x 4th Gen Intel® Xeon® Scalable Processors, up to 350W
Memory	32x DDR5 DIMMs, up to 4800MT/s
Storage	24x 2.5" SSD, up to 16x NVMe U.2
M.2	2x NVMe/SATA M.2
PCIe	8x low-profile x16 + 2x FHHL x16 slots
OCP	Supports 1x OCP 3.0, supports NCSI
I/O	**Front:** 1x USB 3.0, 1x USB 2.0, 1x VGA **Rear:** 2x USB3.0, 1x RJ45, 1x MicroUSB, 1x VGA
Security	TPM 2.0 (Trusted Platform Module)
PSU	2x 12V 3200W and 6x 54V 2700W, Platinum/Titanium PSU with N+N redundancy
Dimensions	482mm (W) x 263mm (H) x 860mm (D)