kerno

AI Server Portfolio 2026

kerno

AI Server

UQLEAD K03-8220 Ultra

SXM GPUs with NVLink™ & NVSwitch™

CPU & GPU compute

- Dual Intel® Xeon® 6700/6500 Series processors
- NVIDIA HGX™ B300
- 8 x 800 Gb/s OSFP InfiniBand XDR or dual 400 Gb/s Ethernet GPU networking ports via onboard NVIDIA ConnectX®-8 SuperNIC™
- 1.8TB/s GPU-to-GPU bandwidth with NVIDIA NVLink™ and NVSwitch™
- Compatible with NVIDIA® BlueField®-3 DPUs

Memory support

- Up to 32 DDR5 DIMMs (8 channels per CPU)
- Up to 6400 MT/s
- Dual ROM architecture

Power supply

- 12 x 3000W 80 PLUS Titanium redundant power supplies

Storage

- 8 x 2.5" Gen5 NVMe hot-swap bays
- 2 x M.2 slots with PCIe Gen5 x4 and x2 interface

Flexible I/O

- 4 x FHHL dual-slot PCIe Gen5 x16
- 8 x FHHL single-slot PCIe Gen5 x16
- 2 x 10Gb/s LAN via Intel® X710-AT2

TARGET WORKLOADS

AI training

Designed for large-scale AI training workloads; enabling efficient deep learning with high performance and scalability of 8 NVIDIA HGX™ B300 GPUs.

AI inference

Optimized for low-latency AI inference and real-time model deployment; supporting fast and reliable AI-driven applications.

HPC

Ideal for compute-intensive HPC workloads; delivering high throughput and efficient parallel processing for complex simulations and analytics.

AI Server

UQLEAD K03-8220

SXM GPUs with NVLink™ & NVSwitch™

CPU & GPU compute

- Dual Intel® Xeon® 6700/6500 Series processors
- NVIDIA HGX™ B200
- 1.8TB/s GPU-to-GPU bandwidth with NVIDIA NVLink™ & NVSwitch™
- Compatible with NVIDIA® BlueField®-3 DPUs and NVIDIA ConnectX®-7 NICs

Memory support

- Up to 32 DDR5 DIMMs (8 channels per CPU)
- Up to 6400 MT/s
- Dual ROM architecture

Power supply

- 6+6 3000W 80 PLUS Titanium redundant power supplies

Storage

- 8 x 2.5" Gen5 NVMe hot-swap bays
- 2 x M.2 slots with PCIe Gen5 x4 and x2 interface

Flexible I/O

- 4 x FHHL dual-slot PCIe Gen5 x16
- 8 x FHHL single-slot PCIe Gen5 x16
- 2 x 10Gb/s LAN via Intel® X710-AT2

TARGET WORKLOADS

AI training

Designed for large-scale AI training workloads; enabling efficient deep learning with high performance and scalability of 8 NVIDIA HGX™ B200 GPUs.

AI inference

Optimized for low-latency AI inference and real-time model deployment; supporting fast and reliable AI-driven applications.

HPC

Ideal for compute-intensive HPC workloads; delivering high throughput and efficient parallel processing for complex simulations and analytics.

kerno

AI Server

UQLEAD K03-4220

4U, up to 8 x Dual-Slot PCIe Gen5 GPUs

CPU & GPU compute

- Dual Intel® Xeon® 6700/6500 Series processors
- 8 x NVIDIA H200 NVL PCIe GPUs with 4-way NVLink™ bridges or 8 x NVIDIA RTX™ PRO 6000 Blackwell Server Edition

Memory support

- Up to 32 DDR5 DIMMs (8 channels per CPU)
- Up to 6400 MT/s
- Dual ROM architecture

Power supply

- Quad 3000W 80 PLUS Titanium redundant power supply

Storage

- 12 x 2.5" Gen5 NVMe / SATA / SAS hot-swappable bays

Flexible I/O

- 8 x FHFL PCIe Gen5 x16 for GPUs
- 2 x FHHL PCIe Gen5 x16
- 3 x LP PCIe Gen5 x16
- 2 x FHHL PCIe Gen5 x16 (front)

TARGET WORKLOADS

AI training & AI inference

Optimized for AI training and inference; delivering high computational performance and exceptional memory bandwidth.

Virtual computing & simulations

Optimized for visual computing with support for high-density professional NVIDIA RTX™ GPUs; delivering efficient real-time visualization, 3D rendering, and data-intensive simulation workloads.

HPC

Suits a wide range of HPC workloads with dual Intel® Xeon® 6th Generation CPUs, 32 DDR5 DIMMs for massive memory bandwidth, and up to 8 Gen5 GPUs for extreme parallel processing.

kerno

AI Server

UQLEAD K03-42A0

4U, up to 8 x Dual-Slot PCIe Gen5 GPUs

CPU & GPU compute

- Dual AMD EPYC™ 9005/9004 Series processors
- 8 x NVIDIA H200 NVL PCIe GPUs with 4-way NVLink™ bridges or 8 x NVIDIA RTX™ PRO 6000 Blackwell Server Edition

Memory support

- Up to 48 DDR5 DIMMs (12 channels per CPU)
- Up to 5200 MT/s
- Dual ROM architecture

Power supply

- Quad 3000W 80 PLUS Titanium redundant power supply

Storage

- 12 x 2.5" Gen5 NVMe / SATA / SAS hot-swappable bays
- 2 x M.2 (2280/22110), PCIe Gen3 x4

Flexible I/O

- 8 x FHFL PCIe Gen5 x16 for GPUs
- 2 x FHHL PCIe Gen5 x16
- 3 x LP PCIe Gen5 x16
- 2 x FHHL PCIe Gen5 x16 (front)

TARGET WORKLOADS

AI training & AI inference

Optimized for AI training and inference; delivering high computational performance and exceptional memory bandwidth.

Virtual computing & simulations

Optimized for visual computing with support for high-density professional NVIDIA RTX™ GPUs; delivering efficient real-time visualization, 3D rendering, and data-intensive simulation workloads.

HPC

Suits a wide range of HPC workloads with dual AMD EPYC™, 48 DDR5 DIMMs for massive memory bandwidth, and up to 8 Gen5 GPUs for extreme parallel processing.

kerno

Explore how our servers can help you achieve your technology goals

www.kerno.ae • info@kerno.ae • +971 434-66-216