R985 8U Flagship AI Server

The R985 is a high-end 8U rack-mounted server based on 4th/5th generation Intel® Xeon® Scalable processors. Equipped with NVLink professional GPU AI acceleration cards, it effectively meets the demands of ultra-high-density AI computing applications and is well-suited for large-scale cluster deployments, delivering stable and reliable performance for intensive AI workloads.



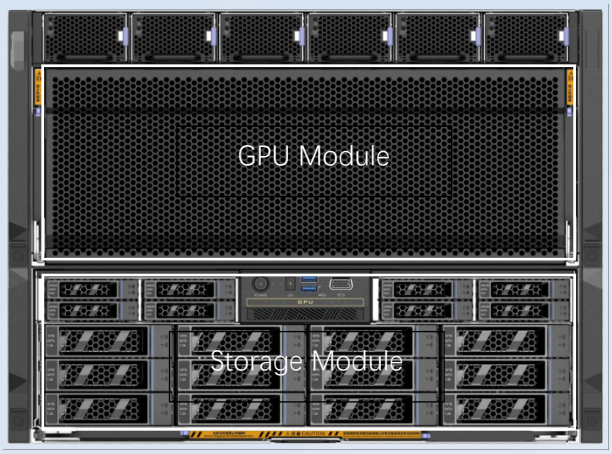
Superior Performance AI Server

Specifications

Model	R985
Processor	Supports 2*4th/5th Gen Intel Xeon® Scalable Processors, up to 350W
Memory	Up to 32*DDR5 slots, up to 5600MHz
RAID	Supports BCM 95 series RAID cards.
Storage	Supports up to 12*3.5 inches hot-swappable drives Supports 2.5-inch U.2 hard drives; supports 4 or 8 hard drives, supports hot-swappable Built-in: 2*SATA M.2
Network Interfaces	1*OCP3.0 network card with NCSI Optional OCP 3.0 interfaces: 2*GE, 2*10GE, 2*25GE, 2*100GE, 2*200GE
PCIe Expansion	9*PCIe 5.0 slots + 1*OCP slot
GPU	Supports 8*NVLink GPU

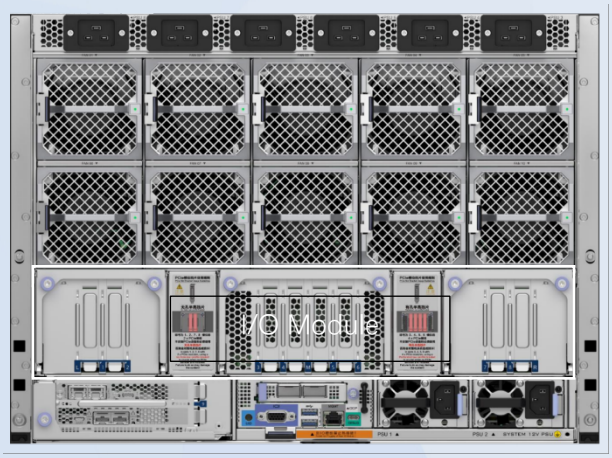
Additional Interfaces	Rear: 1*RJ-45 management port, 2*USB 3.0, 1*VGA Front: 2*USB 3.0+1*VGA
Power Supply	Optional 3300W 54V rear in and out power supply module, support 3+3 redundancy Supports 2700W CRPS Platinum PSU (hot-swappable, 1+1 redundant) Voltage: 100~240V AC (refer to nameplate)
OS Support	Windows Server, RHEL, SLES, CentOS, Ubuntu, VMware ESXi, etc.
Operating Temperature	5°C ~ 35°C

Product View



Front View

- 6*54V power supplies
- The GPU module can be extracted, 4U height
- The storage module can be pulled out, 3U height



Rear View

- 10*hot-swappable fans on the rear;
- The rear I/O module supports up to 8*PCIe cards, note that these 8 cards need to be the same type, and the number of options is 0,1, 2, 4, 8.

Product Features

Ultimate Performance and Flexible Scalability

Unleash Unmatched Hardware Power with Premium Configuration. Powered by up to 2*4th/5th generation Intel® Xeon® Scalable processors (up to 64*cores per CPU), it delivers exceptional parallel computing performance to tackle the most demanding AI, HPC, and cloud workloads. Boasting 32*DDR5 memory slots with a maximum frequency of 5600 MHz per DPC, it achieves a 75% boost in memory bandwidth, ensuring lightning-fast data retrieval and processing. With support for up to 12*3.5-inch and 8*2.5-inch hard drives, it offers flexible, massive storage capacity to accommodate large-scale datasets. Fully integrated with the PCIe 5.0 protocol, it doubles data transfer rates, breaking through I/O bottlenecks and unlocking the full potential of your hardware.

Streamlined Operations and Intelligent Management

It features streamlined operations and intelligent management capabilities to reduce operational complexity and improve maintenance efficiency. The out-of-band visualization management enables remote and precise location tracking of physical devices, further optimizing operational efficiency. It supports automatic crash logging, with online viewing and download capabilities for crash logs, enabling precise fault diagnosis and reducing troubleshooting time. Additionally, it provides real-time health monitoring and reporting for critical components, delivering comprehensive intelligent services throughout the entire product lifecycle.

Extremely Quiet and Energy-Efficient

It features extremely quiet operation and excellent energy efficiency, balancing high performance with low energy consumption and low noise pollution. It supports high-performance online inference with 8*GPUs, realizing seamless interconnection of 8*professional-grade GPU cards to ensure stable and efficient data transmission. Meanwhile, it optimizes GPU Direct efficiency specifically for deep learning training scenarios, effectively reducing data transmission latency and enhancing the overall training efficiency of deep learning models.