Workstation

Many customers may not have enough space to install a rack-mount server. In such an environment, a workstation server is the best choice.

TEAMMAX *System Integration is led by an experienced management team committed to providing comprehensive solutions.*

Key Features

- AMD CPU
- Large language model platform
- Dual Gigabit Ethernet
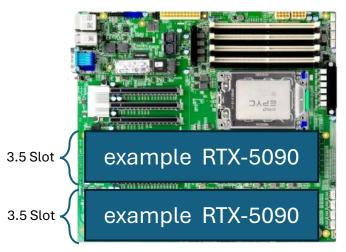- 8-slot graphics card support

Applicable places

- Hospitals
- Schools
- Research units
- Banks
- Small businesses

Applications

- LLM
- GenAI
- RAG
- AI reasoning
- small model training

Item	Description
CPU	AMD EPYC 7003 series (Milan) AMD EPYC 7002 series (Rome) AMD EPYC 7001 series (Naples)
Memory	Each DIMM supports up to 256GB
Graphic	8-slot graphics card
PCIe slot	PCIe4.0 high-speed expansion slot *10
Power Supply	2000W *2(Option Redundant / ATX power)

Actual test ： With 2 RTX-5090 graphics cards to run the deepseek-r1:70b AI model generates about 30 tokens per second, which is smoother than using 3 RTX A5000 graphics cards. If you need to execute a larger AI model, you can consider upgrading to a graphics card with larger VRAM.

GeForce GPU support	Slot
NVIDIA RTX-5090 (32G)	3.5 Slot
NVIDIA RTX-5080 (16G)	3 Slot
NVIDIA RTX-4090 (24G)	3 Slot

TESLA GPU support	Slot
NVIDIA TESLA H200 (141G)	2 Slot
NVIDIA TESLA H100 (80G)	2 Slot

Quadro GPU support	Slot
NVIDIA RTX-PRO6000 (96G)	2 Slot
NVIDIA RTX-6000ADA (48G)	2 Slot
NVIDIA RTX-5000ADA (32G)	2 Slot
NVIDIA RTX-4500ADA (24G)	2 Slot
NVIDIA RTX-4000ADA (20G)	2 Slot

3.5 Slot { example RTX-5090

3.5 Slot { example RTX-5090

2 Slot { example RTX-6000ADA

2 Slot { example RTX-6000ADA

2 Slot { example RTX-6000ADA

2 Slot { example RTX-6000ADA