

TG780V3

LARGE MODEL INFERENCE AI SERVER

The TG678V3 is a versatile 8-GPU server supporting 4th or 5th Gen Intel® Xeon® Scalable Processors and up to 8x 600W GPUs. It delivers exceptional performance, high scalability, and robust reliability, making it suitable for various application scenarios such as Artificial Intelligence (AI), Large Model Inference, Graphics Rendering, and Cloud Gaming.



NEW PLATFORM, EXCEPTIONAL PERFORMANCE

- Utilizes the Intel 4th/5th Gen Xeon Scalable processor platform, featuring comprehensive upgrades in CPU compute power, PCIe speed, and memory bandwidth to fully unleash GPU performance.
- Supports 8x 600W GPUs; each card utilizes PCIe 5.0 x16, delivering efficient heterogeneous computing power.
- CPU-GPU direct-connect design ensures high efficiency and low latency, significantly improving data transfer efficiency compared to PCIe Switch architectures.

COMPATIBLE WITH FAN AND BLOWER CARDS

- Superior compatibility, optionally supports either fan-cooled (active) or blower-style (turbine) GPU cards.
- Supports 8x four-slot fan-cooled cards or 8x dual-slot blower-style cards.

FLEXIBLE CONFIGURATION & ON-DEMAND SELECTION

- Supports up to 11 standard PCIe 5.0 slots; multiple PCIe configurations available.
- Optional 1 x OCP 3.0 NIC with multiple speed options available.
- Supports 8 x 3.5"/2.5" SAS/SATA HDDs/SSDs; optionally supports 2/4 x U.2 NVMe SSDs, balancing high capacity and high-performance local storage.

STABLE, RELIABLE & INTELLIGENT MANAGEMENT

- Key system components feature redundant and hot-swappable designs, complemented by tool-less maintenance, enhancing fault recovery efficiency and system availability.
- Integrated intelligent management chip provides an open management platform supporting multiple protocols like IPMI 2.0, Redfish, and SNMP.
- Supports various management functions including KVM, Virtual Media, status monitoring of key components, anomaly alerts, and more, offering comprehensive remote system-level intelligent management capabilities.

TECHNICAL SPECIFICATIONS

Category	Specification Details
Form Factor	Standard 7U Rackmount
CPU	Supports 2 x 4th or 5th Gen Intel® Xeon® Scalable Processors, TDP up to 350W
Memory	32 x DDR5 DIMM slots, up to 5600MHz, supports RDIMM/3DS RDIMM
RAID Controller	Integrated onboard SATA; Optional 12Gb/s SAS HBA or RAID card
GPU Support	Supports 8 x 600W GPUs; Optional support for fan-cooled or blower-style cards
PCIe Expansion	Up to 11 x standard PCIe 5.0 slots
OCP NIC	Optional OCP 3.0 NIC (PCIe 5.0 x8), options include 4x1GbE / 2x10Gb SFP+ / 2x25Gb SFP28
Local Storage	8 x 3.5"/2.5" SAS/SATA bays (Optional support for 2 or 4 x U.2 NVMe SSDs)
Onboard Storage	2 x onboard M.2 slots (SATA 3.0 / PCIe 4.0 x4, 2280/22110)
I/O Ports Front	2 x USB 3.0, 1 x VGA
I/O Ports Rear	1 x Serial port, 2 x USB 3.0, 1 x VGA, 1 x RJ45 Dedicated Management port
System Cooling	Optimized thermal design with isolated cooling zones for CPU and GPU modules. Hot-pluggable fans support intelligent speed control.
Power Supply	Option 1: 4 x CRPS power modules, supports 2+2 / 3+1 redundancy Option 2: 6 x CRPS power modules, supports 3+3 redundancy Options: 2000W / 2700W / 3200W / 3600W modules; All modules are hot-pluggable
Management	Integrated BMC (AST2600), supports IPMI 2.0, Redfish, SOL, KVM, Virtual Media, etc. 1x dedicated 1Gbps RJ45 management port.
Security Features	Optional TPM/TCM security module; Chassis intrusion detection; Lockable, tool-less chassis cover; Supports BIOS/BMC dual flash redundancy.
Chassis Dimensions	Width 448mm x Height 306mm (7U) x Depth 896mm

Temperature	Operating: 5°C to 35°C Non-operating: -40°C to 65°C
Humidity	Operating: 8% to 90% RH (non-condensing) Non-operating: 5% to 95% RH (non-condensing)
OS Support	Microsoft Windows Server, Red Hat Enterprise Linux, CentOS, Ubuntu, and other mainstream operating systems.

DATA CENTER SOLUTIONS BASED ON BUSINESS OBJECTIVES

- Maximize the value of your IT and enterprise solutions with a business-focused approach that helps reduce IT complexity, lower costs, and improve efficiency. While leveraging your existing technologies, we seamlessly integrate innovative, customized solutions designed to meet your specific needs. Based on your business applications, we deliver an efficient, agile, open, and intelligent end-to-end data center solution that simplifies IT, addresses business challenges, and continuously creates value for your organization. For more information, please contact your sales representative.

FOR MORE INFORMATION, PLEASE VISIT WWW.AETHLUMIS.COM

- Other trademarks and trade names that maybe used in this document refer to entities or their products that own these trademarks and names. Aethlumis declares that it does not have any exclusive rights and interests in the trademarks and names of other entities. The content of this article is for reference only. Aethlumis reserves the right to make changes to any of the products described in this article without prior notice. The content of this article is provided as is, without any form of express or implied warranty.