

混合算力引擎

TuringStack LZ4412A-G8



卓越性能，为AI业务提供混合算力

- 混合算力，灵活适配多种AI加速卡，集训练、推理能力于一身，满足不同AI负载对算力的需求

灵活拓扑，适配不同应用场景需求

- 4种GPU拓扑，灵活适配客户不同应用场景，智能调配资源，驱动算力的高效运转

算存一体，提供经济高效的算力引擎

- 最多可支持10个双宽GPU及400TB海量存储设计，满足人工智能数据对于存储空间的需求

产品概述

TuringStack LZ4412A-G8是图灵智新推出的新一代4U两路的人工智能服务器产品，在CPU和GPU的规格上相对于上一代有显著的提升，该产品具有卓越性能、重塑架构、扩展性强和可靠性高等特点，可应用于深度学习模型训练、深度学习推理、高性能计算、数据分析等多种应用场景，易于管理和部署，算存一体的设计，使其成为智能时代下混合算力引擎。

产品优势

TuringStack LZ4412A-G8可支持多种类型人工智能加速卡，应对人工智能不同场景下对异构计算算力的需求。CPU和GPU挂载比可支持1:4、1:8，并可支持直通等多种拓扑配置，用户可以基于应用模型按需配置/灵活（或一键）切换，并且支持丰富的AI加速卡和智能网卡，从而适应不同人工智能场景的需求。

TuringStack LZ4412A-G8支持多达2颗AMD EPYC处理器，可实现90%性能提升和50%的内核数量提升，配合12通道内存，内存容量可达6TB。LZ4412A-G8支持多达24块2.5/3.5英寸硬盘，其中多达8块支持NVMe，可满足人工智能海量数据对于存储空间的需求。LZ4412A-G8可支持多达12个PCIe 5.0插槽、内置1个专用阵列卡槽位和1个OCP 3.0专用网卡插槽，支持400GE网卡，满足人工智能对网络带宽的高速需求。

TuringStack LZ4412A-G8

GPU计算模块	8 GPU模块	4卡直通模块
GPU支持	支持多达8张双宽GPU,采用PCIe 5.0 Switch解决方案	支持多达6张双宽GPU
处理器	支持2颗AMD EPYC 9004系列处理器, 单颗多达96个核心, 最大功率400W	
内存	支持多达24个DDR5内存插槽, 速率最高支持4800MT/s, 支持RDIMM或LRDIMM, 内存容量高达6TB; 支持CXL 1.1*	
存储控制器	可选配高性能RAID控制器,支持RAID0/1/10/5/6/50/60/Tri-Mode*	
PCIe插槽	8个双宽 ×16插槽PCIe5.0, 4个单宽 ×16插槽PCIe5.0, 1个OCP 3.0网卡专用插槽和1个阵列卡专用插槽	6个双宽x16插槽PCIe5.0, 1个OCP 3.0网卡专用插槽 和1个阵列卡专用插槽
前部硬盘扩展	系统SAS/SATA/NVMe硬盘配置: <ul style="list-style-type: none"> • 24块2.5/3.5英寸SAS/SATA/NVMe (可支持8块NVMe PCIe4.0 x4) , • 支持2块板载M.2 12块2.5/3.5英寸SAS/SATA/NVMe (可支持8块NVMe PCIe4.0 x4) , 支持2块板M.2 	
网络	可选配OCP 3.0网卡, 可选配1/10/25/50/100/200/400GE PCIe标准网卡	
风扇	支持8个8056风扇, N+1热插拔冗余风扇	
管理软件	HDM无代理管理工具 (带独立管理端口), 支持LCD可触摸智能管理模块, 支持64M本地显存*可选U-Center数据中心管理平台	
安全性	支持机箱入侵检测, TCM/TPM安全模块和CPU内置专用安全处理器;	
接口	2个USB 3.0+1个VGA+1个TypeC Debug口 (前面板) , 1个USB 3.0 (内部) , 2个USB 3.0+1个VGA+1个1Gbps专用管理网口 (后面板) ,可选1个管理串口*	
电源	支持4个2000W/2400W/2700W白金电源, 可选钛金电源*, 支持N+N冗余	
系统支持	支持Microsoft Windows Server2019及以上版本、CentOS7.9及以上版本、Ubuntu22.04 LTS及以上版本、银河麒麟V10等	
外形/机箱尺寸	4U机箱; 174.8mm (高) ×447mm (宽) ×828mm (深) (不含安全面板) 174.8mm (高) ×447mm (宽) ×856mm (深) (含安全面板)	
保修	三年5*9, 下一工作日响应(NBD)	

[注]*具体当前可配置信息以详细产品手册为准。

广东图灵智新技术有限公司