



UNIS银河大模型一体机 产品介绍



产品介绍



UNIS Server R3935 G6 Pro
海光系列



UNIS Server R3910 G6
飞腾系列

随着DeepSeek大模型加速AI对各行业的重塑，紫光恒越推出了UNIS银河大模型一体机成为重要支撑。一体机已全面适配 DeepSeek 全系列及其他主流模型，深度融合瀚博、海光、天数、沐曦、昆仑芯等国产化GPU，夯实自主算力底座。通过推理加速优化，强化性能表现。智能体工作站支持MCP智能体协议适配，整合先进智能体框架与范式，精准匹配行业需求推出合同审核、财务审计、设备运维监控、写作助手等一体机，搭配 UNIS 银河大模型平台软件，整合大模型算法、资产、开发工具，为客户构建自身“专属大模型”及典型场景应用提供了可视化的开发和管理能力。



UNIS银河大模型一体机集成全国产化、训推一体、多智能体、高可靠和强扩展性等优点，可广泛应用于办公辅助、科研设计、仿真推演、网络安全、情报分析及生产制造等领域。安全防护体系引入三员管理明确权责分工，相互制约监督，规范操作行为。同时，通过软硬件国产化自主可控、数据闭环运行等手段，支持大模型的二次训练和私有化部署，实现“数据不出营区、模型不离阵地”，为领域客户构建“专属大模型”，为行业数据安全与智能升级筑牢防线。





产品规格

UNIS银河大模型一体机-海光系列

银河大模型一体机配置		轻量版	标准版 (推理)	专业版 (高性能推理)	集群版 (训推一体)
算力硬件	大模型版本	DeepSeek-R1 32B	DeepSeek-R1 70B	DeepSeek-R1/V3 671B	DeepSeek-R1/V3 671B
	算力服务器	UNIS R3830 G6 Pro	UNIS R3935 G6 Pro	UNIS R3935 G6 Pro	UNIS R3935 G6 Pro*4台
	CPU	7470 48C*2	7490 64C*2	7490 64C*2	7490 64C*2
	GPU	K100 AI 64G显存 FP16@196TFLOPS*1卡	K100 AI 64G显存 FP16@196TFLOPS*2卡	K100 AI 64G显存 FP16@196TFLOPS*8卡	K100 AI 64G显存 FP16@196TFLOPS*8卡
	内存	DDR5 256G	DDR5 512G	DDR5 1.5T	DDR5 1.5T
	系统盘	480G SATA SSD*2+RAID	480G SATA SSD*2+RAID	480G SATA SSD*2+RAID	480G SATA SSD*2+RAID
	数据盘	960G NVMe*2	1.92T NVMe*2	3.84T NVMe*4	3.84T NVMe*4
大模型软件	应用	RAG问答检索	RAG问答检索、 报告生成、图像识别	RAG问答检索、报告生成、 图像识别、文生图/视频、 数据分析、公文写作、智能体	RAG问答检索、报告生成、 图像识别、文生图/视频、 数据分析、公文写作、智能体、 智能体工作站、方案助手
	工具链	知识库	知识库、应用编排、 40+基础组件 (文件解析、图检索等)	知识库、应用编排、 40+基础组件，10个高级组件 (报表、多维检索等)	知识库、应用编排、 40+基础组件，20个高级组件 (路径规划、评估反馈等)、 多智能体框架
	模型平台	系统管理	系统管理、模型库	系统管理、模型库、 推理加速、MCP协议	系统管理、模型库、推理加速、 数据集管理、模型训练、 微调、强化学习、多机多卡 调度、负载均衡、多LoRA推理、 MCP协议、数据预处理、数据安全
并发用户数		10-20人	20-40人	80-160人	200-500人



产品规格

UNIS银河大模型一体机-飞腾系列

银河大模型一体机配置		轻量版	标准版 (推理)	专业版 (高性能推理)	集群版 (训推一体)
算力硬件	大模型版本	DeepSeek-R1 32B	DeepSeek-R1 70B	DeepSeek-R1/V3 671B	DeepSeek-R1/V3 671B
	算力服务器	UNIS R3810 G6	UNIS R3910 G6	UNIS R3910 G6	UNIS R3910 G6 * 4台
	CPU	S5000-64C*2	S5000-64C*2	S5000-64C*2	S5000-64C*2
	GPU	天数智铠100 32G显存 FP16@96TFLOPS *2卡	天数智铠100 32G显存 FP16@96TFLOPS *4卡	天数天垓150 64G 显存 FP16@192TFLOPS *8卡	天数天垓150 64G 显存 FP16@192TFLOPS *8卡
	内存	DDR5 256G	DDR5 512G	DDR5 1T	DDR5 1T
	系统盘	480G SATA SSD*2+RAID	480G SATA SSD*2+RAID	480G SATA SSD*2+RAID	480G SATA SSD*2+RAID
	数据盘	960G NVMe*2	1.92T NVMe*2	3.84T NVMe*4	3.84T NVMe*4
大模型软件	应用	RAG问答检索	RAG问答检索、 报告生成、图像识别	RAG问答检索、报告生成、 图像识别、文生图/视频、 数据分析、公文写作、智能体	RAG问答检索、报告生成、 图像识别、文生图/视频、 数据分析、公文写作、智能体、 智能体工作站、方案助手
	工具链	知识库	知识库、应用编排、 40+基础组件 (文件解析、图检索等)	知识库、应用编排、 40+基础组件，10个高级组件 (报表、多维检索等)	知识库、应用编排、 40+基础组件，20个高级组件 (路径规划、评估反馈等)、 多智能体框架
	模型平台	系统管理	系统管理、模型库	系统管理、模型库、推理加速、 MCP协议	系统管理、模型库、推理加速、 数据集管理、模型训练、 微调、强化学习、多机多卡调度、 负载均衡、多LoRA推理、 MCP协议、数据预处理、数据安全
并发用户数		10-20人	20-40人	80-160人	200-500人

紫光恒越技术有限公司

公司地址：
北京市海淀区中关村东路1号院2号楼402室
邮编：100084

咨询与服务：
服务热线：400-910-9998

www.unisyue.com



紫光恒越官方微信



紫光恒越官方网站

Copyright © 2025紫光恒越 保留一切权利

免责声明：虽然紫光恒越试图在本资料中提供准确的信息，但不保证本资料的内容不含有技术性误差或印刷性错误，为此紫光恒越对本资料中信息的准确性不承担任何责任。
紫光恒越保留在没有任何通知或提示的情况下对本资料的内容进行修改的权利。

[注] *具体配置以实际下单为准，具体认证实际周期以获得证书时间为准。