



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

随着生成式AI加速驱动产业智能化变革，大模型一体机已成为构建自主AI能力的核心基础设施。紫光恒越UNIS银河大模型一体机在软硬协同、开放生态与场景落地方面进行了全面升级。一体机深度融合海光、天数、沐曦等国产高性能GPU，构建了自主可控的异构算力底座，实现对DeepSeek、Qwen等主流模型的动态优化适配。通过推理加速技术，显著提升计算与显存效率，保障高并发、低延迟的模型服务。产品集成智能体工作站，兼容MCP等协议，内置智能体编排框架与行业范式。结合UNIS银河大模型平台，提供从模型精调、评估到应用编排的全生命周期可视化管理，助力用户高效构建与运营“专属大模型”。产品聚焦高频刚需，提供开箱即用的解决方案，如桌面智能体、合同审核、财务审计、设备预测性维护、公文写作，并支持通过低代码 workflow 快速定制，精准对接业务需求。



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	海光/天垓/沐曦*1卡	海光/天垓/沐曦*2卡	海光/天垓/沐曦*8卡	海光/天垓/沐曦*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、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	海光/天垓/沐曦*1卡	海光/天垓/沐曦*2卡	海光/天垓/沐曦*8卡	海光/天垓/沐曦*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、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。