



Ai WorkStation

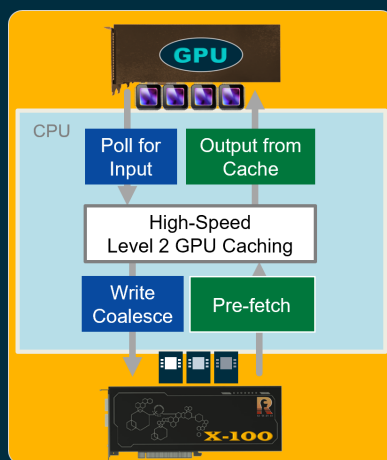
免主板、模块化的智算工作站可提供2颗GPU与4颗GPU扩展能力
具备从CPU到GPU全国产化设计体系、从而让智算底座更加自主可控

晓软合作研发的创新型体系结构 GPU WorkStation、
通过默认配置FPGA+ASIC制式的通算模块X-100、
可让智算工作站具备完整的通用算力环境、
即：具备安装国产操作系统的能力。

具备2颗GPU扩展能力AWS-P2与4颗GPU
扩展AWS-P4。

已适配广泛的国产GPU厂商、让企业在GPU选型
谈判方面掌握更多的主动权。

我们致力于开放性的体系结构、
这不仅体现在客户可自由的选择GPU品牌、
还包括：客户可在AWS-P2、AWS-P4桌面型
算力设备选择多种处理器体系，甚至日益强
大的RISC-V生态。



自主研发的X-100加速模块可有效的卸载GPU
对于数据的处理；
通常情况下、GPU显存因为昂贵的容量
在处理大尺寸模型加载时会产生瓶颈。
X-100预载数据到本地NVMe持久存储系统、
通过高带宽的互联协议供GPU近线使用、
降低延迟的同时、做为二级显存扩展GPU容量。

Ai 智能体的最佳算力载体

Artificial Intelligence



GpuDirect Storage

便捷的灵活性



能耗低，超静音



全模块化设计



模型框架已然就绪

AIOS-vLLM
Workflow RAG

提供完善的交付能力，从国产GPU驱动程序、算子适配、框架优化与调优、一站式交付、
解决国产GPU市场化推广难题的最后一公里！