

深度计算处理器 DCU K100 (AI 版)

释放澎湃计算力量，实现智能计算价值



生态兼容



通用性好



Tensor 加速



稳定量产

产品介绍

DCU K100 系列标卡，是基于通用 GPGPU 架构设计的全精度加速卡，性能可对标 NVIDIA 产品，应用生态完善，覆盖全精度场景。迁移成本低，基于 PyTorch、TensorFlow 等主流框架实现的代码无需转码，可直接使用。是构建 AI 算力和 HPC 算力的不二之选。

◆ 通用加速卡架构

通用图形处理器架构设计，指令集丰富，可编程性更灵活，适用范围更广。

◆ 开放软件生态

支持主流深度学习框架和开源软件开发环境，应用移植高效便捷。

◆ 超强计算性能

支持 TF32、BF16 大模型训练推理主力算力精度，性能领先，可满足新一代智能计算中心一体化算力部署。

◆ 系统定制服务

提供一体化整机定制服务。

◆ 丰富的量产经验

20 万片大规模量产、部署经验。

◆ 快速交付

芯片供货周期短，系统建设经验丰富。

技术参数

规格参数		K100_AI
芯片	计算单元	120CU
性能指标	FP32	24
	FP16	48
	INT8	96
	FP32（Tensor）	49
	TF32（Tensor）	98
	FP16（Tensor）	196
	BF16（Tensor）	196
	INT8（Tensor）	393
显存规格	显存类型	GDDR6
	显存容量	64GB
	显存带宽	896GB/s
接口类型	PCIe 接口	PCIe 4.0 x16
散热	散热类型	风冷
最大功耗	TDP	400W
尺寸规格	尺寸	PCIe 双宽全高全长