共创“MXAIE”解决方案

方案通过与金融、医疗健康、能源、教科研、交通、大文娱等场景的深度融合，并积极探索拓展具身智能、低空经济等“X”新兴领域，构建基于沐曦 GPU 硬件和红帽OpenShift 容器云平台的全栈开源、云原生 AI 计算平台。

方案展示

- 提供本地化的开源 GPU 计算AI生态
- 利用 OpenShift 实现统一的容器编排、算力调度与多租户管理
- 支撑本地化芯片的 AI云基础设施


金融


医疗健康


能源


交通


教科研


大文娱


X（其它）

Red Hat OpenShift AI

Agentic App Development

Red Hat AI Inference Server

MIM
MetaX Inference Microservice

Optimized Profiles

Prompts & Responses

Model Monitoring & Observability

Metrics | Telemetry | Accelerator Profiles | Logging | Alerting

MetaX GPU Operator+Network Operator+MetaXLink+MIM Operator

Red Hat OpenShift

Enterprise Kubernetes | Enterprise Linux | Container & VM Management | DevOps Tooling | Security | Ecosystem

方案优势

- 本地化芯片适配
- 开源与可控
- 云原生加速能力
- 多租户与统一管理
- 性能优化与弹性伸缩
- MLOps 一体化生态
- 生态延展性强
- MIM应用生态

*图片由大模型生成

沐曦集成电路（上海）股份有限公司
地址：上海市浦东新区海科路999弄C8栋
网址：www.metax-tech.com
邮箱：business@metax-tech.com

欢迎关注沐曦

沐曦开发者社区

沐曦&Red Hat共创“MXAIE”本地化AI应用典型应用场景

单文本语言模型 | 混合专家 / MoE 模型 | 多模态 / 视觉 — 文本结合模型 | Pooling / Embedding 模型

Red Hat AI Inference Server

Red Hat × 沐曦 METAX

Red Hat OpenShift AI

AI4Weather | AI4Biomedical | PaddleScience | AI4Materials

公司介绍

沐曦致力于为异构计算提供安全可靠的GPU芯片及解决方案，打造全栈GPU芯片产品，推出曦思®N系列GPU用于智算推理，曦云®C系列GPU用于通用计算、以及曦彩,G系列GPU用于图形渲染、满足“高能效”及“高通用性”的算力需求。沐曦产品均采用完全自主研发的GPU IP，拥有完全自主的指令集和架构，配以兼容主流GPU生态的完整软件栈(MXMACA®)，具备高能效和高通用性的天然优势，能够为客户构建软硬件一体的全面生态解决方案，是“双碳”背景下推动数字经济建设和产业数字化、智能化转型升级的算力基石。

*图片由大模型生成

沐曦集成电路（上海）股份有限公司
地址：上海市浦东新区海科路999弄C8栋
网址：www.metax-tech.com
邮箱：business@metax-tech.com

沐曦应用生态