



GOTC

全球开源技术峰会

THE GLOBAL OPENSOURCE TECHNOLOGY CONFERENCE

OPEN SOURCE , OPEN WORLD

基础技术开源开放 激发AI创新动能

郑墨 | 第四范式技术副总裁、基础技术负责人

2021年7月31日

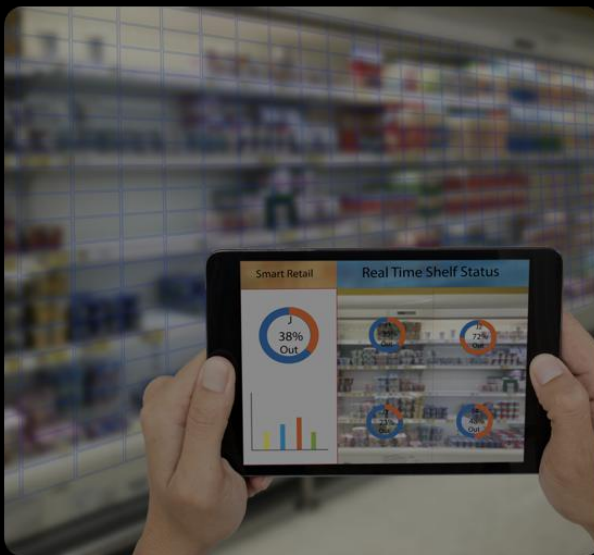


郑墨 | 第四范式技术副总裁、基础技术负责人

- 在大规模机器学习系统、个性化推荐、搜索、工程技术团队管理等领域拥有丰富经验。曾任Google 展示广告架构团队架构师、Pinterest 个性化推荐与搜索团队技术负责人
- 现任第四范式技术副总裁、基础技术负责人，主持设计与开发国内首个开源机器学习数据库 OpenMLDB、开源AI操作系统内核 OpenAIOS。带领团队打造出国内首款软件定义算力软硬一体化AI集成系统 SageOne Appliance
- 热爱 coding, 曾获 ACM 世界冠军。研发并开源国内首个推荐算法工具 SVDFeature 获得 KDD Cup 2011 季军、2012 冠军

AI 进入所有人的工作和生活，进入各个行业

GOTC



全球开源技术峰会

AI 快速步入行业的背后，是人类科学发展的四个范式

GOTC



第一范式：实验科学

人类记录现象
(钻木取火、摩擦起电)
重复记录的现象



第二范式：理论科学

人类总结理论
诠释自然现象



第三范式：计算科学

通过计算机推演理论
模拟现象



第四范式：数据科学

计算机从海量数据中
发现规律、形成理论诠释
自然现象

科学规律数量井喷

机器寻找海量规律



产业全面进入科学时代

机器在海量领域寻找规律

全球开源技术峰会

“

The Fourth Paradigm: Data-Intensive Scientific
Discovery” 2007《科学发展的四个范式》 -
—— 1998年图灵奖获得者 Jim Gray

”

人 vs 机器

有限理性

数百条规则
抓大放小，无法适应动态变化

极限理性

超千万规则
面面俱到，可以形成闭环动态更新

事后响应

事后诸葛亮
无法实时响应

实时响应

毫秒级响应，模型、特征高频更新
对突发事件实现细粒度反应

无法复用

科学家资源供不应求
专家经验、沉淀无法复用

规模复制

复用性强
可大规模复制至多类业务场景

AI挖掘海量数据中的科学规律，带来应用效果提升

GOTC

> 99%

某银行
反洗钱效率提升

千条规则->亿级规则

> 96%

某保险公司
保险索赔案风险审核准确率

数千条规则->千万级规则

99.33%

某石油企业
化工产品价格预测1-7天平均准确率

数百条规则->千万级规则

50%

某零售企业
销售预测准确率提升

百条规则->千万级规则

~ 3倍

糖尿病预警准确率
同专业医师协会相比提升

十数条规则->百万级规则

1,900%

某国内最大单机游戏媒体
点击率提升

百级规则->亿级规则

全球开源技术峰会

成功案例做到什么：基础技术链接数据、算力、应用形成合力

GOTC



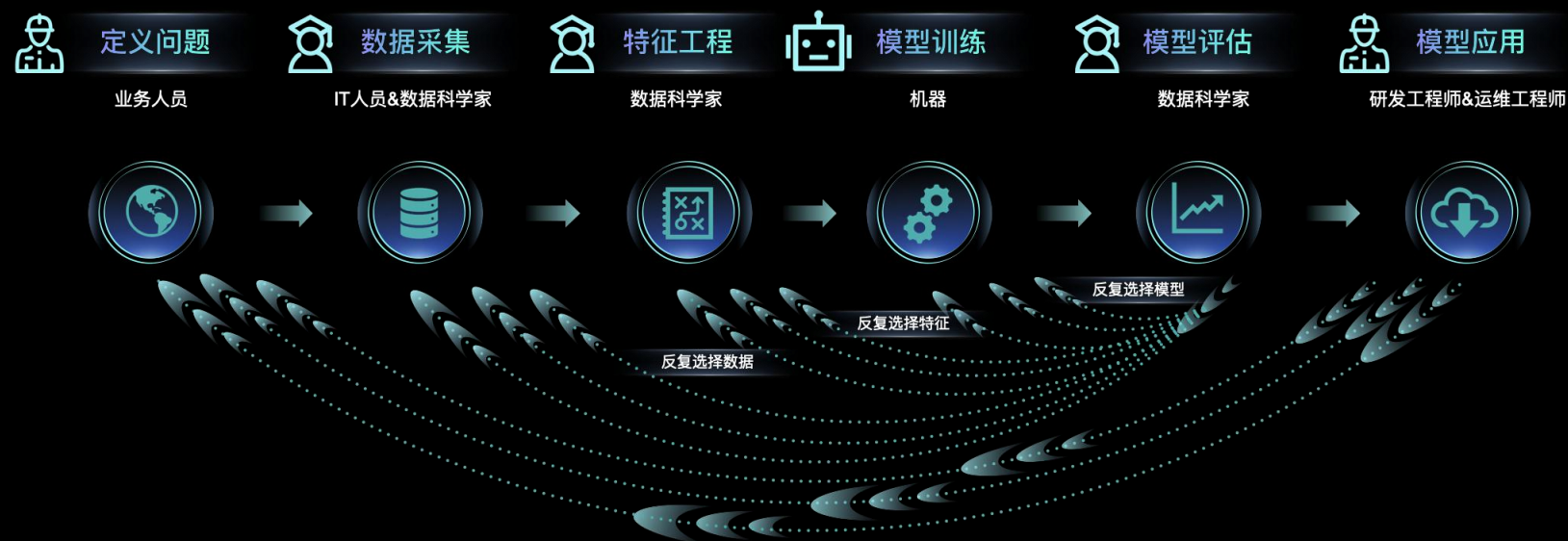
全球开源技术峰会

应用侧受制于各环节开发者反复的工作

GOTC

6个月

单个AI应用平均交付时间



全球开源技术峰会

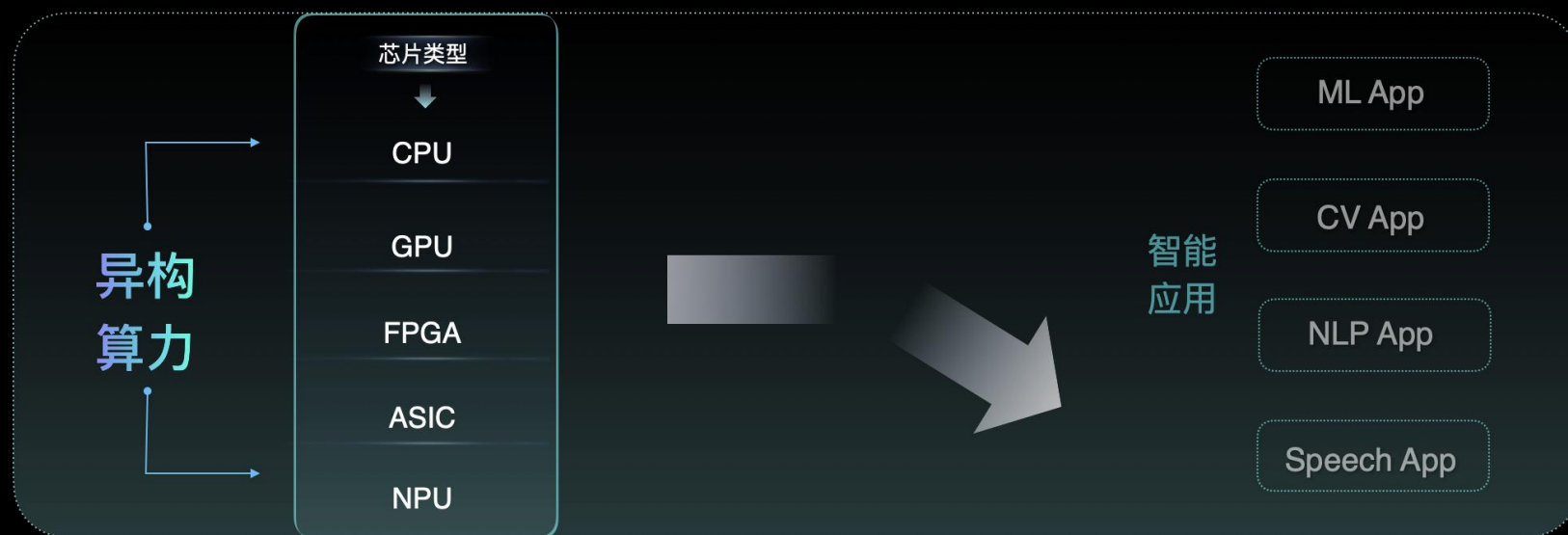
全栈AutoML算法，全场景、全流程AI自动化

GOTC



全球开源技术峰会

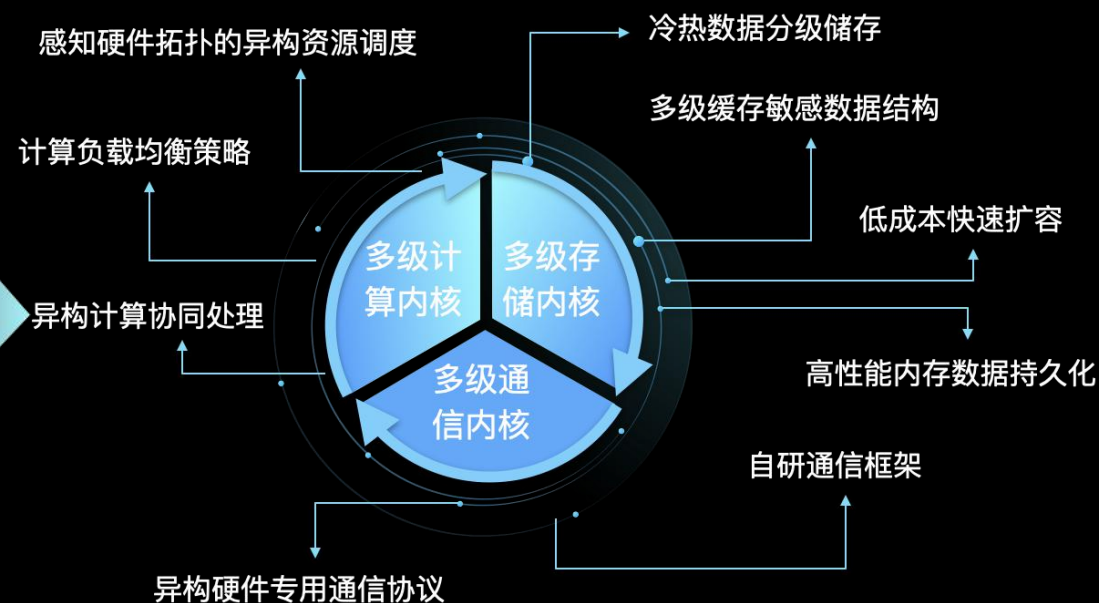
算力侧利用率低下，异构硬件使用门槛高



- 业务与算力适配难
- 现有调度系统落后
- 资源调度的门槛高
- 集群管理操作复杂

OpenAIOS 实现对异构设备的统一调度管理

GOTC



全球开源技术峰会

数据侧的技术演进为人机协同决策提供了可能

GOTC

人手工录入数据并分析数据

系统自动收集数据，人分析数据



数据记全、记对

~百G 级别数据
~数百数据处理秒级响应

瞬时新信息快速获取，全量数据计算

~PB 级别数据
~百万数据处理秒级响应

超大规模并发和数据量，云化基础设施

~数百 PB 级别数据
~百万数据处理毫秒级响应

全球开源技术峰会

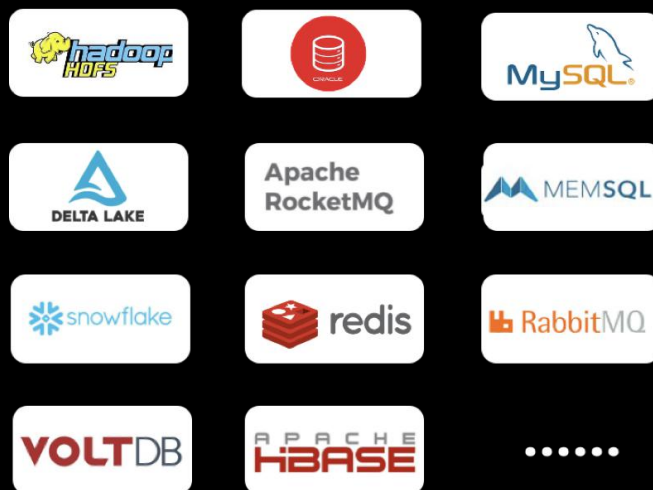
正确、高效的AI数据供给成为数据侧的新挑战

GOTC

95%

时间精力花费在数据上

数据来源: Gartner How to Operationalize Machine Learning and Data Science Projects Published July 2018



AI

全球开源技术峰会

OpenMLDB, 供给正确的机器学习数据

GOTC



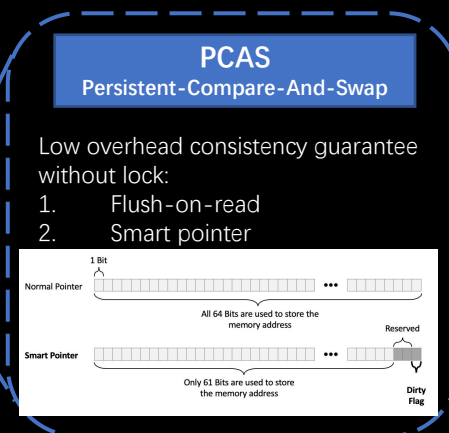
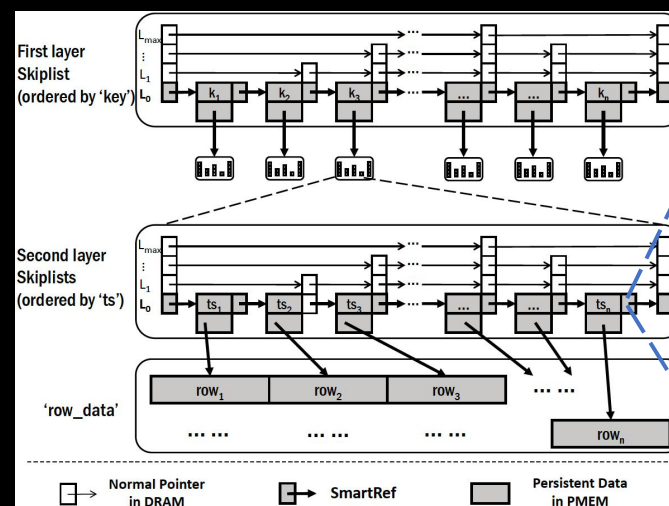
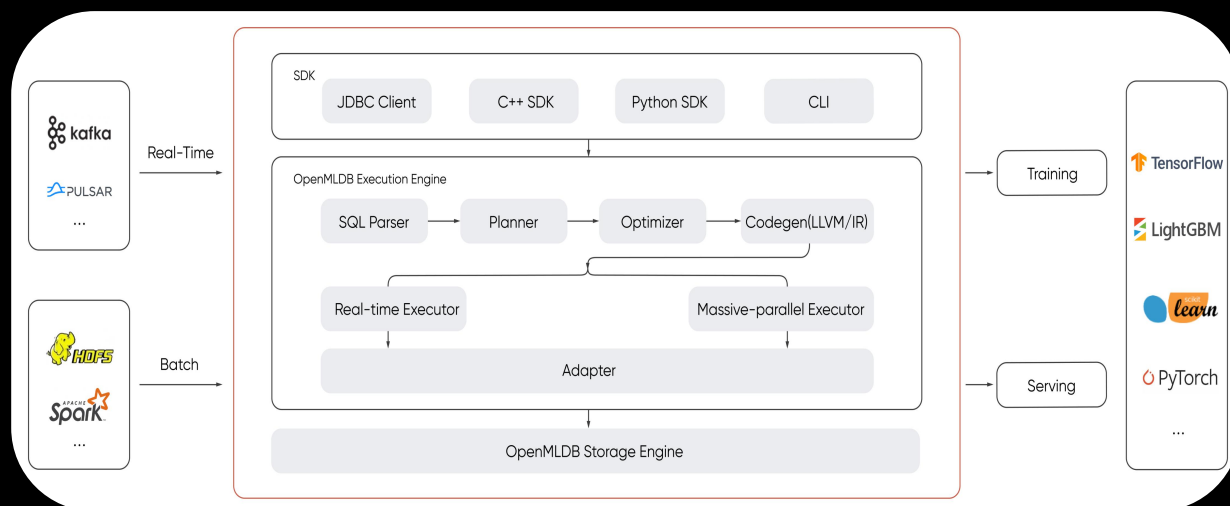
全球开源技术峰会

OpenMLDB, 供给高效的机器学习数据

GOTC

相比商用 DBMS 实现

- > 50% TCO 节省
- > 90% 灾备恢复速度
- > 30x 提速



全球开源技术峰会

OpenMLDB, 正确数据保障的同时降低运维复杂性

GOTC



全球开源技术峰会

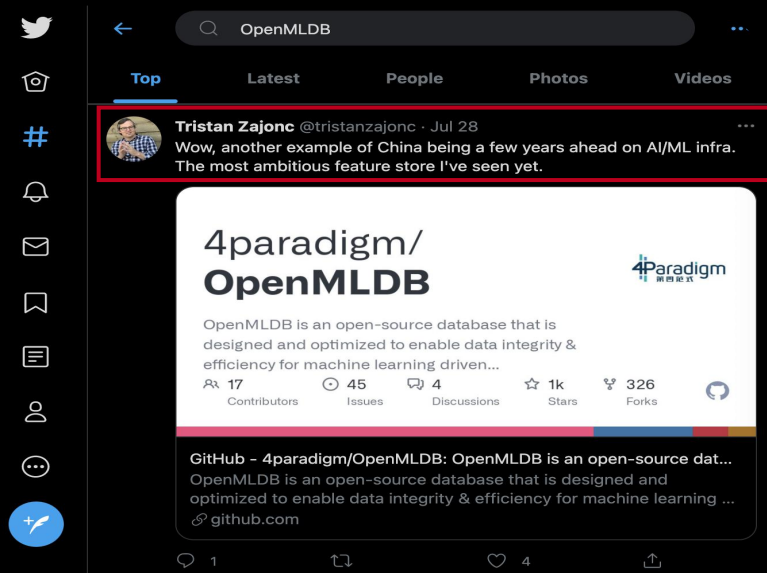
OpenMLDB 正式发布月余，获得了社区小伙伴的关注

GOTC

50万行代码

1.5k GitHub stars

全球开源技术峰会

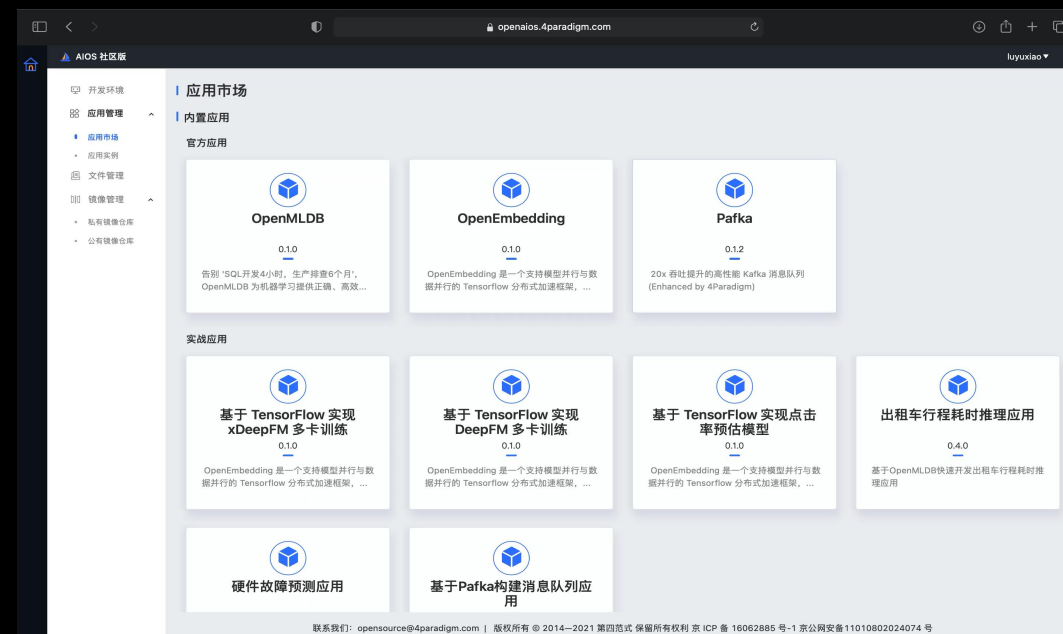


- Former CTO for Machine Learning at Cloudera

AIOS社区版，提供免费算力和开发环境，供开发者体验和学习

GOTC

AIOS社区版—开源AI核心技术



全球开源技术峰会

► 欢迎扫码下载/体验开源组件产品

GOTC

OpenMLDB



<https://github.com/4paradigm/OpenMLDB>

AIOS社区版



<https://openaios.4paradigm.com>

全球开源技术峰会

THANKS