

Apache Doris 在中国邮政 储蓄银行金融反欺诈领域的 应用探索

狄潇然

中国邮储银行 软件研发中心 西安分中心
技术专家

目录

1. 业务背景介绍
2. 应用实践及价值
3. 未来展望

1 业务背景介绍

企业简介

邮储银行

中国邮政储蓄银行拥有近**4万个**营业网点，服务个人客户超**6.5亿户**，定位于服务“**三农**”、**城乡居民和中小企业**，依托“自营+代理”的独特模式和资源禀赋，致力于为中国经济转型中最具活力的客户群体提供服务，加速向数据驱动、渠道协同、批零联动、运营高效的新零售银行转型。

中国邮政储蓄银行拥有优良的资产质量和显著的成长潜力，是中国**领先**的大型零售银行。

反欺诈平台

反欺诈平台作为邮储银行**智能风控体系的屏障**，以科技助力业务发展为理念，前沿技术为支撑，以规则模型策略为基础，打造**事前检测、事中阻断、事后分析**的全方位一体风险防控体系。

实现对全行**网络金融、个人金融、信用卡、消费信贷**等业务交易的全面风险监控。

日均拦截风险交易**数万笔**，有效阻止客户损失**数千万**，**毫秒间**即可完成风险识别。

实时分析诉求

案件溯源

支持对已生成案件进行**即席深入追踪溯源分析**，包括：欺诈行为的源头、路径、涉及人员、设备和其他相关信息。

运营监测

支持对交易数据、行为数据、模型策略、风险特征运营情况进行**实时监测**，一旦发生风险事件，能够进行**风险预警和干预**。

高时效性

分析报表**实时性高且精准**，能够实时捕获各类风险变化，具备**高效、准确、快速响应**的能力。

技术愿景

多表关联 性能优异

内部分级建模，风险运营分析**涉及多表，数据量大、分析维度不固定**且逻辑复杂。

并发查询 性能优异

计划向网点业务人员开放查询权限，查询**并发量大、时间集中、案件溯源逻辑复杂**。

实时写入 性能优异

功能特点强调实时性，对接入数据**实时性有极高的要求**，以期迅速捕获风控特征变化。

事务&数 据一致性

某些运营指标的统计精度要求高，需要做到端到端的**“精准一次消费”**语义。

运维 难度低

在监控、灾备、扩缩容、数据迁移等方面有自动化工具，**运维成本低，风险小**。

2 应用实践及价值

技术选型

大数据分析领域现状

- **架构复杂**：多套数据组件，带来高昂的使用成本和极大的运维、管控压力。
- **性能不足**：每个场景一个组件，单一组件难以满足业务对数据分析效率的需求，时效性差。

Doris VS ClickHouse

- 对比分析了 ClickHouse、GreenPlum、Cassandra、Kylin 等多个列式存储组件。
- 着重对 **Doris** 和 **ClickHouse** 进行对比分析。

技术选型

评测过程

评测结论

TPC-H 数据集

89个测试案例

Apache Doris在实时写入性能方面表现优异，现有硬件资源下，实时写入性能是Clickhouse的6倍，且支持事务，支持“精准一次消费”语义。

5轮安装部署

在单表查询场景下，ClickHouse的表现相对更优，适合查询模式固化的宽表分析；Doris的单表查询性能良好，能够满足当前业务场景中单表查询性能要求。

SSB 数据集

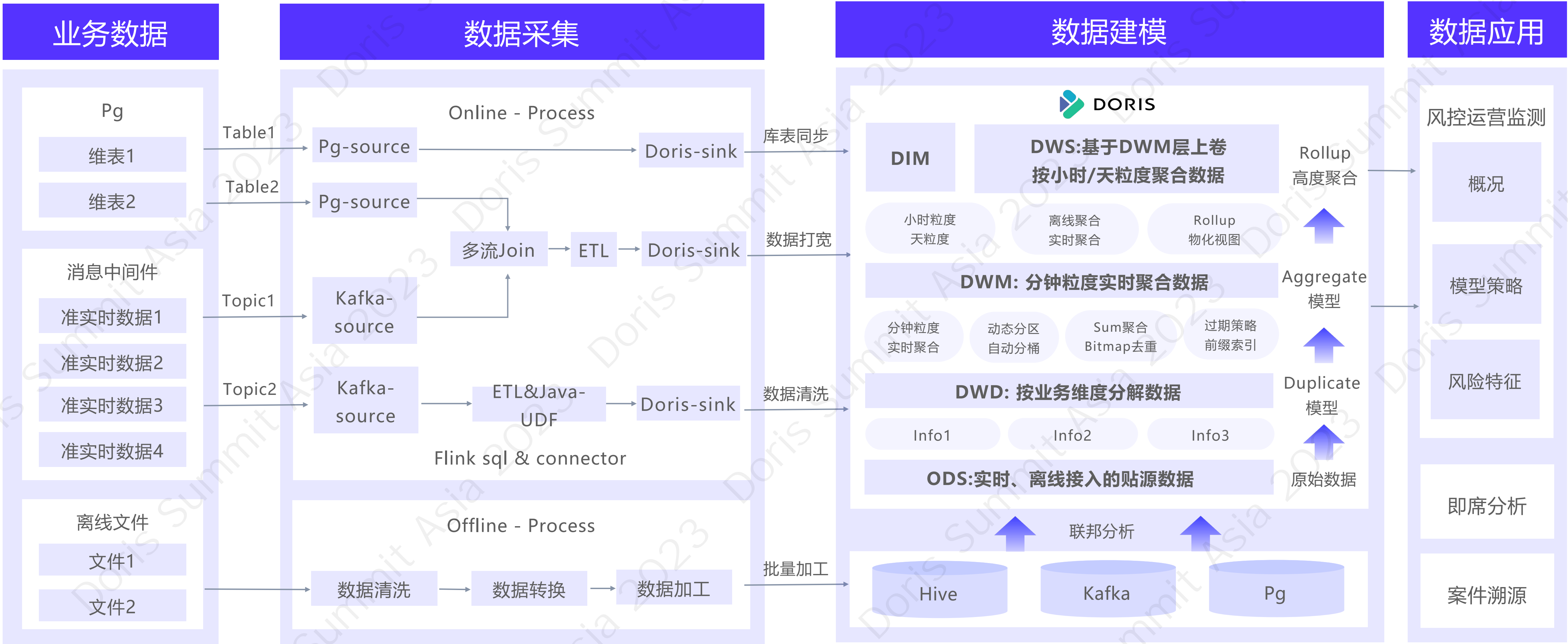
11条修改意见，7次修改迭代

Apache Doris在多表Join、复杂聚合及高并发场景下的查询性能具有明显优势。

10个文件，156页，5W多字

Apache Doris具备完善的高可用方案，可自动完成切换，无须人工干预。

解决方案（脱敏）



场景应用



监控报表

Doris+Flink 流计算引擎联合，完成基于规则、模型设备、地址等维度报表的实时统计。



即席分析

基于近一年交易记录，提供包含10多种复杂条件的实时统计分析。



工单标记

支持工单信息与交易信息进行关联分析，快速定位阻断详情。



运营预警

实时接入系统告警信息，作业运行数据、模型策略运行数据。

应用价值

业务量

- 1.支持对**1000万+**客户、**3万+**柜员、**1万+**支行网点、**1千+**产品的在线统计分析。
- 2.支持**1万+**笔工单的数据追踪、详情分析。
- 3.支持动态检测**近百种**不同类别的告警规则，实现系统级别预警提醒。

业务效果

- 1.核心报表数据实时性从1-2天延迟骤减至**5s**内。
- 2.**80%**即席分析可在**2s**内返回结果，**95%**的即席分析可在**5s**内返回结果。
- 3.受益于Apache Doris极致的存储压缩比，存储成本降低**70%**。

应用价值

近两月，触发规则模型运营告警

超 **100** 次

告警有效率

95%+

告警延迟

5s内

3 未来展望

未来展望

日志检索

高并发点查

2.0 升级

现状

使用ELK、Kafka、Flink、Hive等组件进行日志采集，涉及组件较多，不易管理，开发运维成本高，且占用了较多服务器硬件资源。

优化

统一采用Kafka+Doris的轻量级日志采集框架，实现日志统一管控，以SQL的方式便捷查询，降低开发运维难度，在不影响原业务基础上节省服务器资源。

现状

Doris基于列式存储引擎构建，对于大宽表高并发查询整行数据的场景下，会导致CPU开销较高，进而影响其他正常业务的资源使用。

优化

实时数仓在面向不同分行、支行时，能够承受住高并发下行式数据的检索分析，提供大数量级下明细数据的统计分析。

当前版本无法支持以下业务场景

- (1) 在Unique模型中使用数组类型进行数据存储。
- (2) 日志关键字模糊查询、等值查询和范围查询场景。



获取更多社区动态与最佳实践

Apache Doris 官方平台:

- Apache Doris 官网: doris.apache.org
- Apache Doris GitHub: github.com/apache/doris/

获取更多峰会资料:

- Doris Summit 峰会官网: doris-summit.org.cn
- Doris Summit 峰会回放: <https://space.bilibili.com/1196172099/channel/collectiondetail?sid=1824324>