

Apache Flink在中国农业银行的 探索和实践——神荼平台

侯鑫磊 · 中国农业银行 高级研发工程师

Kafka X Flink Meetup 深圳 - 2019年08月31日

Contents

目录 >>

1 背景介绍

3 应用效果

2 实践分享

4 思考展望

PART 01

背景介绍

站在数字化转型的门口

商业银行赖以生存的环境正发生着深刻的变化，过去单纯依靠信贷资产规模扩张的“外延式”高速发展和高盈利的时代已基本结束，信贷领域发展已进入下半场，亟需将数字科技嵌入到金融服务，开启数字化革命的大门。

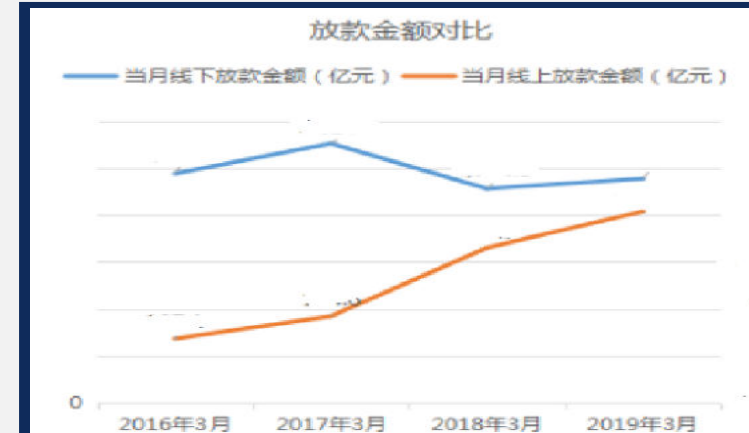


外部环境变化

- 宏观经济增速放缓
- 利率市场化改革与强监管
- 互联网金融冲击
- 客户消费习惯趋于数字化
- 科技推动商业变革

内部环境变化

- 银行利润增速放缓
- 信贷资产质量承压
- 信贷产品供需矛盾凸显
- 信贷风险防控压力加大
- 业务发展的线上化趋势



农行信贷业务驶入数字化转型快车道

农业银行承担着大行责任，为服务乡村振兴，服务普惠金融，服务消费升级，聚焦线下、线上两个信贷主战场，以支持分行(A)创新、深化银企(B)协同、提升客户(C)体验、发力银政(G)合作等主要任务为抓手，在惠农e贷、网捷贷、微捷贷、数据网贷和场景贷等方面持续发力，搭建” To A B C G” 四种服务方式全构建的平台雏形，累计服务百万农户、百万个人客户、近5万小微客户，助力信贷领域数字化转型。

普惠金融提升收益的三大要素

01

下沉的客户群体

普惠金融客户群体数量大，覆盖面大，长尾客户数量多，客户需求丰富，是银行提升收益的来源。

02

多样化的产品设计

普惠金融客户的需求复杂，对成本考量多，产品的计价、多样性、匹配度非常重要。

03

更丰富的获客渠道

基于银行成本和客户便利性的原因，线上渠道、尤其是移动端的流量日渐增长，随着开放银行的发展，Anything、Anytime、Anywhere成为方向。



对应的风控问题

01

下沉的客户群体

普惠金融客户与银行既有客户群体相比是一个相对下沉的群体，资信水平难免降低，欺诈和违约风险难免提高，需要更有效的风控手段。

02

多样化的产品设计

多样化的产品设计给系统设计和开发都带来很大挑战，风控的难度更高。

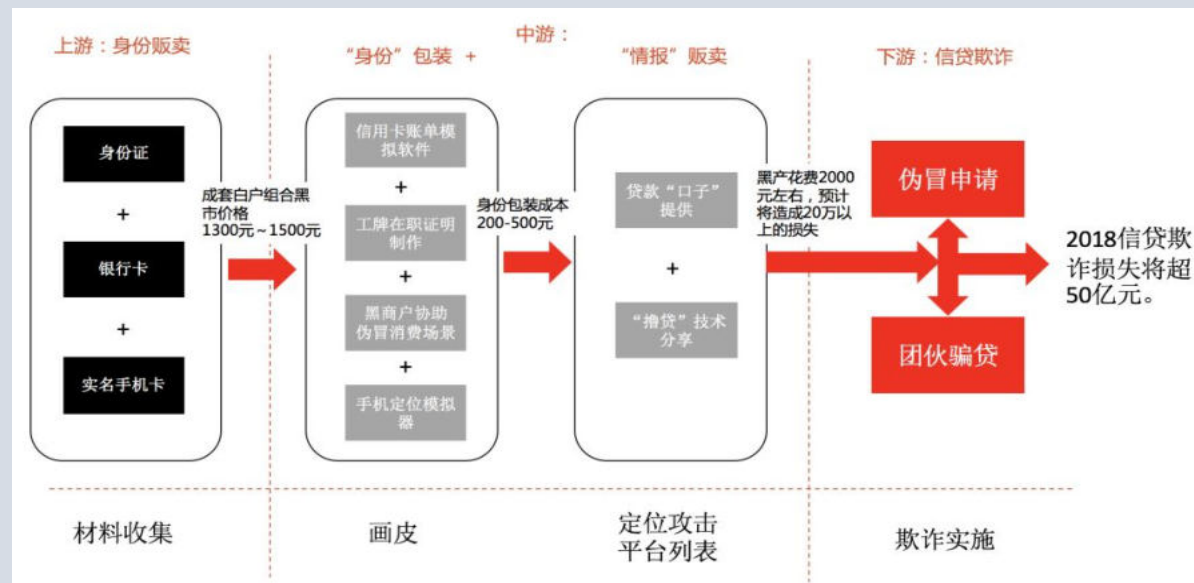
03

更丰富的获客渠道

日益多样的获客渠道，面临更复杂的反欺诈调整，欺诈更具隐蔽性，欺诈模式更新速度也随之提升，风控需要覆盖全渠道，也需要随时应对不断升级的猫鼠有戏。

风控形势日趋严峻

目前，我国有160万的黑产从业人员，1000万张游离的身份证，1.3亿运营商黑卡。国内每年因黑产欺诈导致的损失超千亿。全球超90%的金融企业在2017年遭受过欺诈问题的困扰。互联网信贷行业，50%的逾期来源于欺诈。随着客户下沉，欺诈风险渐渐凸显。



团伙化

专业化

线上化

智能化

- 羊毛党
- 中介
- 冒名贷款
- 个体欺诈
- 团伙攻击



黑产从业人员



生物特征



设备



身份



其他技术.....

- 伪造设备
- 代理IP
- 修改LBS
- 批量养号
- 模拟器
- 伪造资质



经验为主

业务趋势由大额低频转变为小额高频，依靠客户经理难以顾全大量线上信贷客户。



客户体验

无法根据客户特点进行针对性风控措施，加强风控往往需要牺牲客户体验。



实时性差

针对线上产品310的特点，往往采取事后分析的手段，缺少实时风险防控的能力。



模型失灵

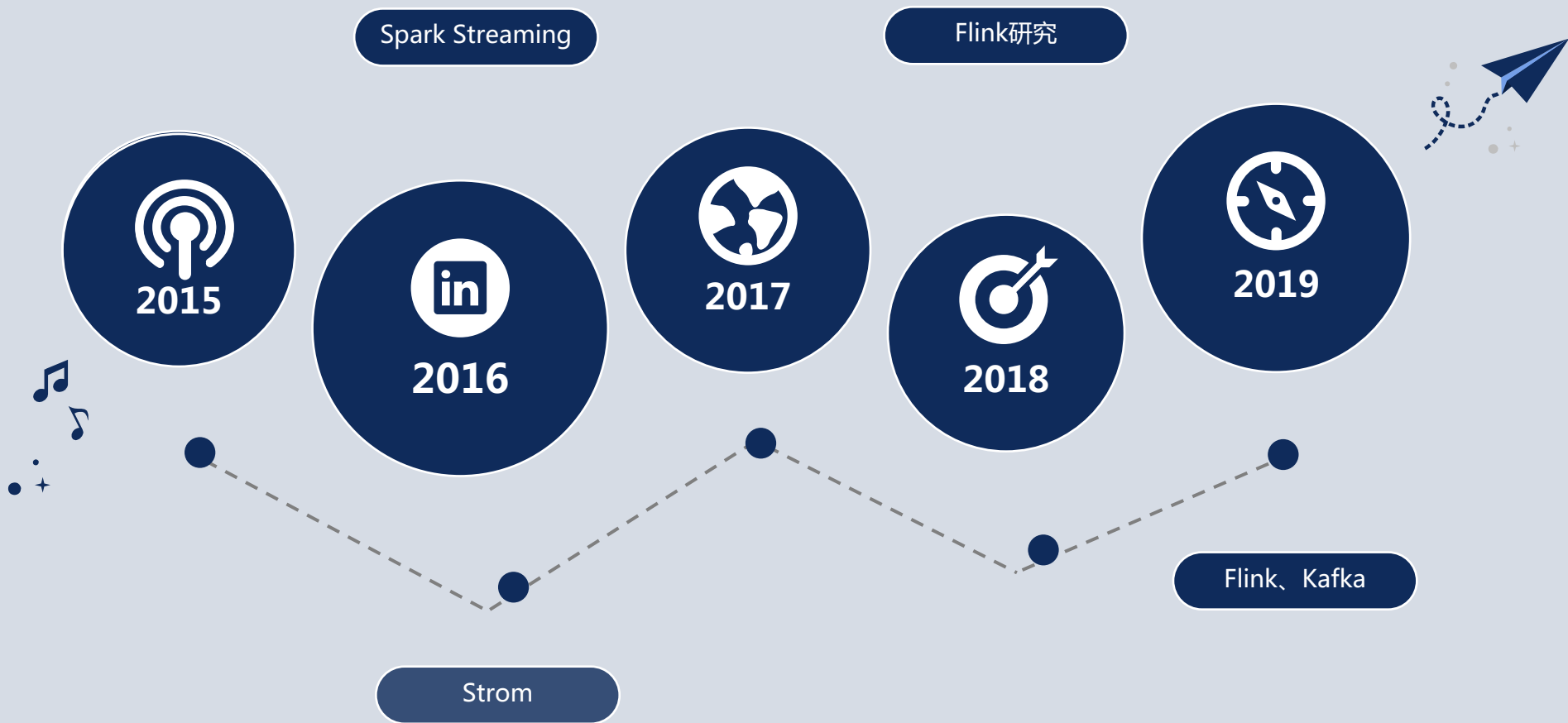
信用风险转变为欺诈风险，传统风控模型不能防范黑产、黑户、黑团、多头、网络欺诈、贷款中介等等。



1.7 选型—实时计算在农业银行的演进



Apache Flink





实时性

线上贷款秒审秒批秒贷
客户全流程自助操作
风控的实时性成为难点



准确性

只计算在时间窗口内发生的事件
时间的指定方式：事件时间、处理时间
发生故障后仍需保持准确



吞吐率

日均消息数百万
用户查询需毫秒级反馈
高吞吐低延时如何保证

	Storm	Spark Streaming	Flink
架构组成	原生流计框架	微批模式	原生流计框架
流处理方式	每次传入一个事件，ms级延时。	一个窗口内的所有事件，秒级延时	单条事件处理，ms级延时
容错机制	ACK机制	WAL和RDD血统机制	基于快照方式容错
可靠性保证	At last once，在Trident模式下支持exactly once语义	支持exactly once	支持exactly once
高级API	应用根据Storm定义规则编写	能对接Spark栈的相关高级API	提供图计算、机器学习等高级API
易用性	不支持SQL Streaming	支持SQL Streaming，批量和Steaming采用同一的编程框架	支持SQL Streaming，批量和Steaming采用同一的编程框架
成熟度	相对比较早的流系统，比较稳定	较成熟	新型框架，处于发展阶段

PART 02

实践分享

业务流程

登录

贷款申请

合同签约

贷款放款

风险类型

黑产骗贷

中介代办

产品漏洞

团伙攻击

反欺诈能力

风险名单

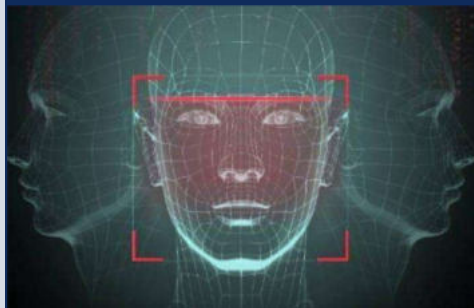
异常行为检测

团伙扫描

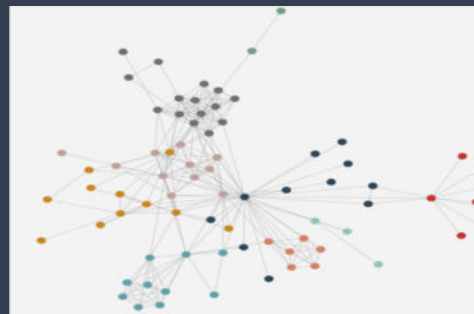
流量预警



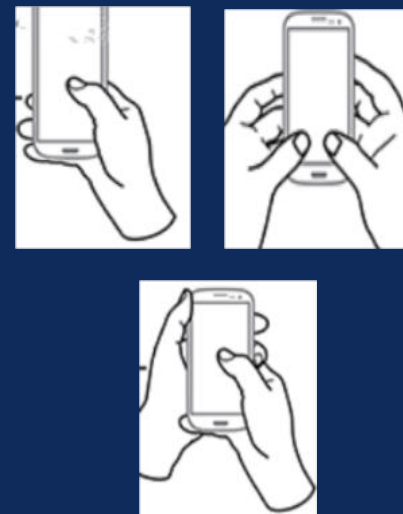
设备指纹



活体检测

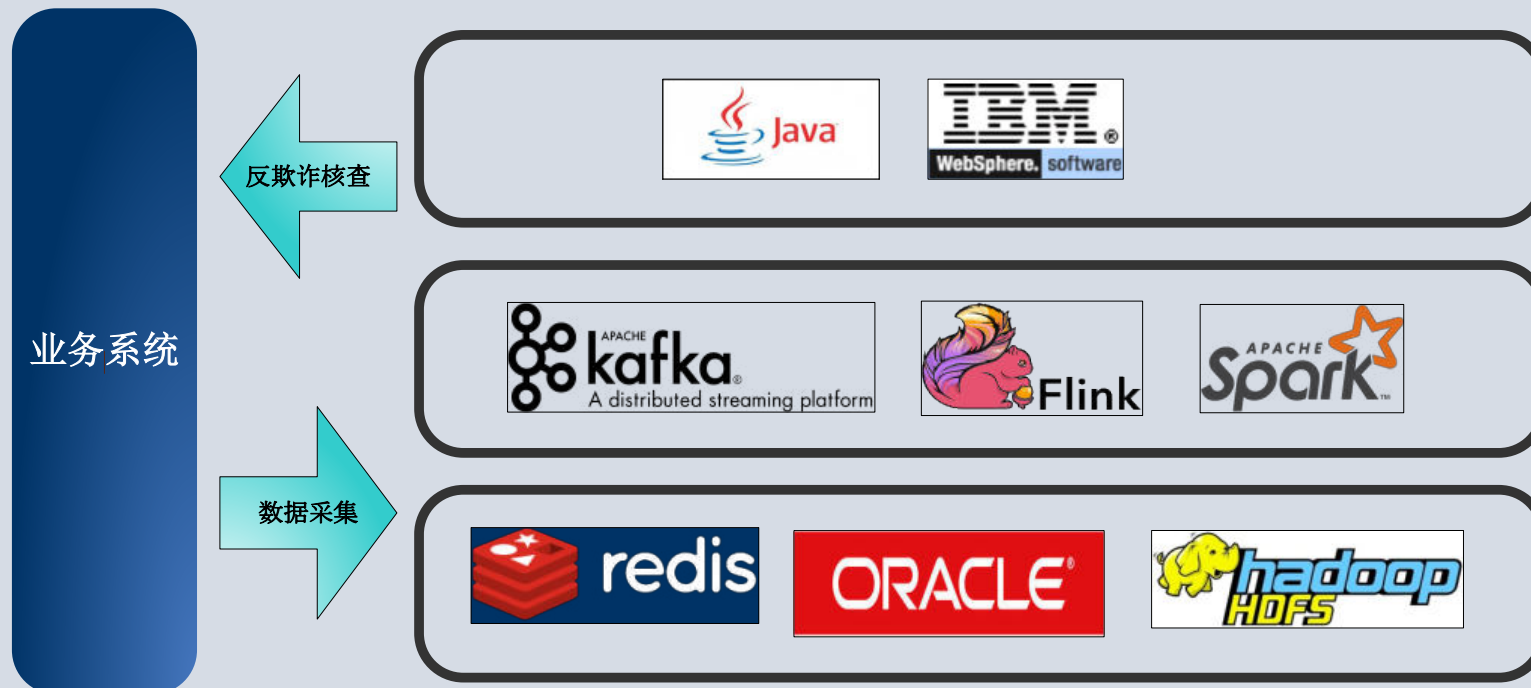


关联关系



行为分析





数据应用层

反欺诈核查服务

风险线索处置

风险大盘展示

系统运行管理

数据服务层

Redis+Oracle+Hbase

数据计算层

实时规则计算
Flink

批量规则计算
Spark

图计算

YARN

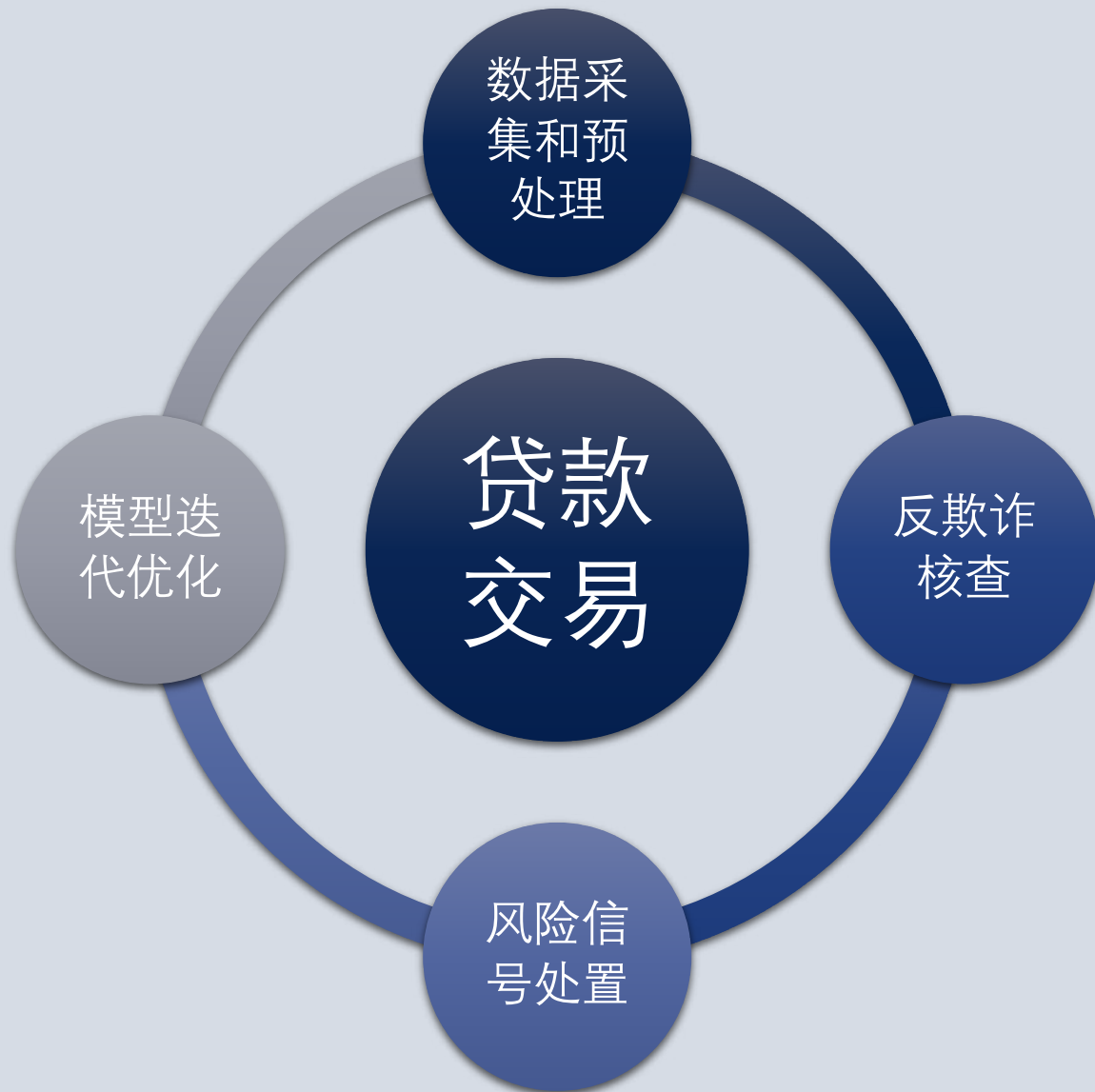
数据源数据层

实时交易数据

交易旁路数据

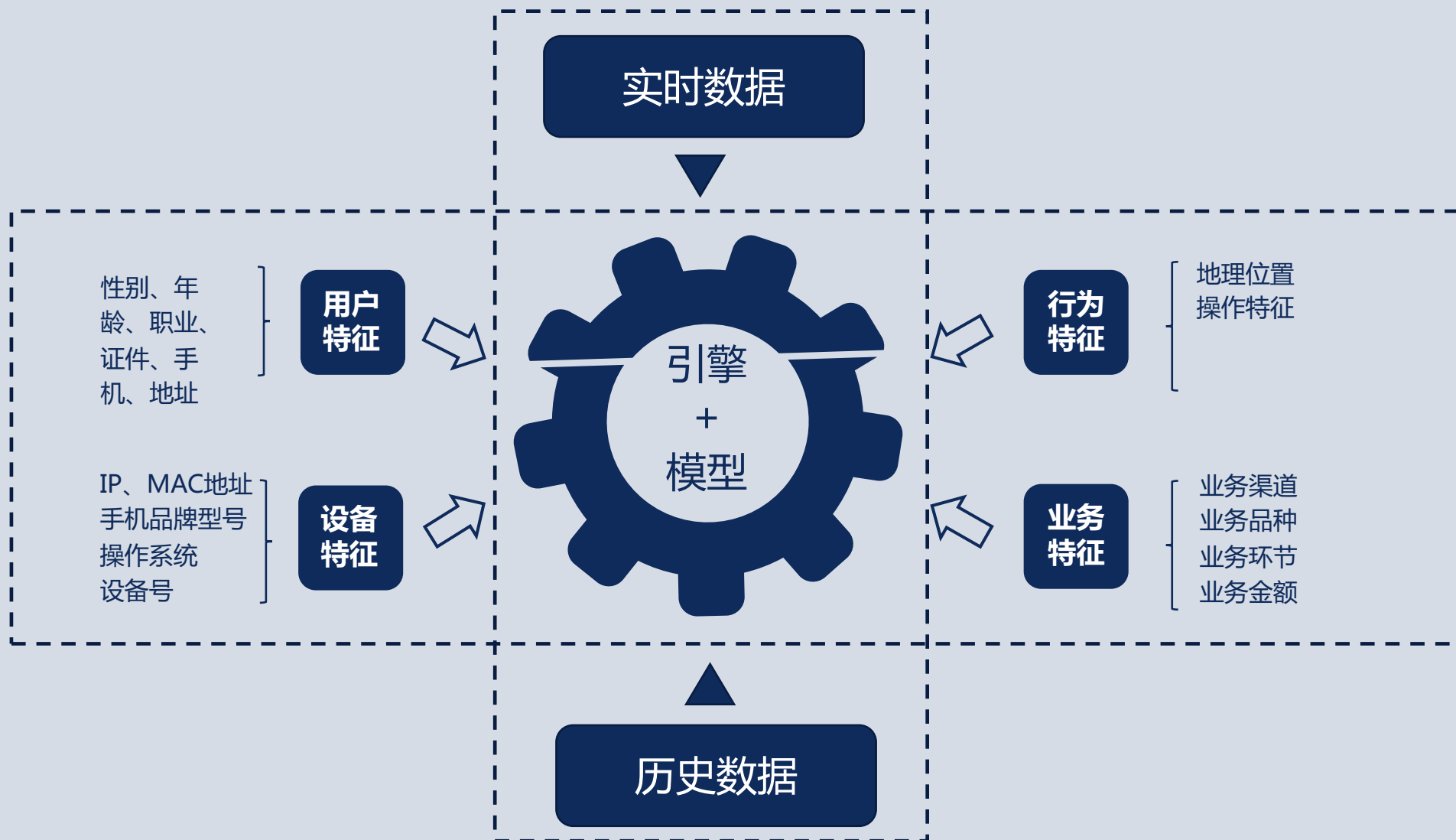
日志数据

离线业务数据



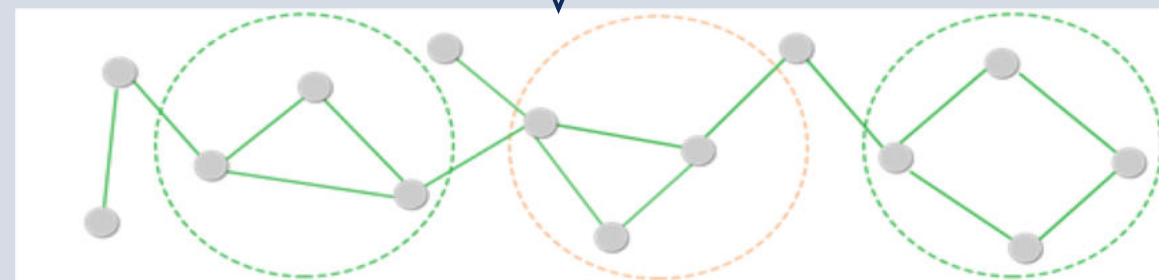
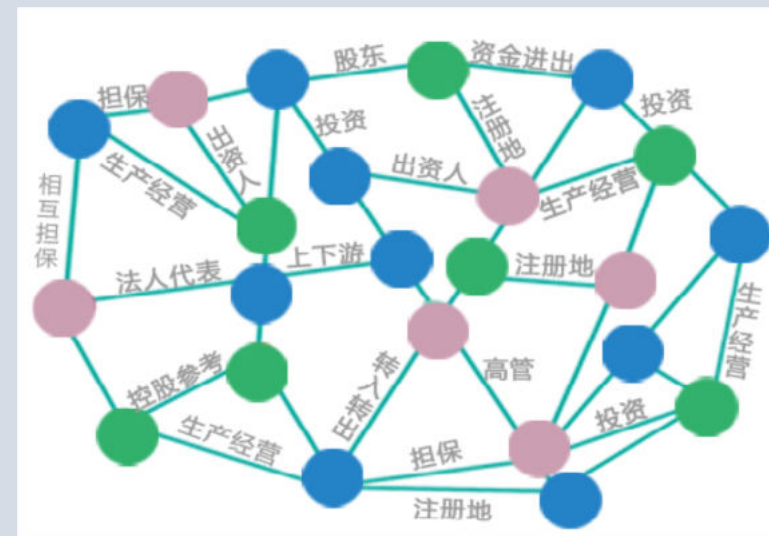
反欺诈流程

- 对贷款交易的实时消息、旁路数据和离线数据进行采集整合，应用反欺诈规则和模型加工预警信号和决策指标
- 贷款交易实时调用反欺诈核查服务，毫秒级返回核查结果：通过、加强认证、阻断
- 对发现的风险信号进行自动处置和人工处置，并进行归档消息日志、核查日志、核查结果、处置结果
- 对发现的风险信号进行自动处置和人工处置，并进行归档消息日志、核查日志、核查结果、处置结果



知识图谱以结构化的形式描述客观世界中概念、实体及其之间的关系，提供一种更好地组织、管理和理解海量信息的能力。

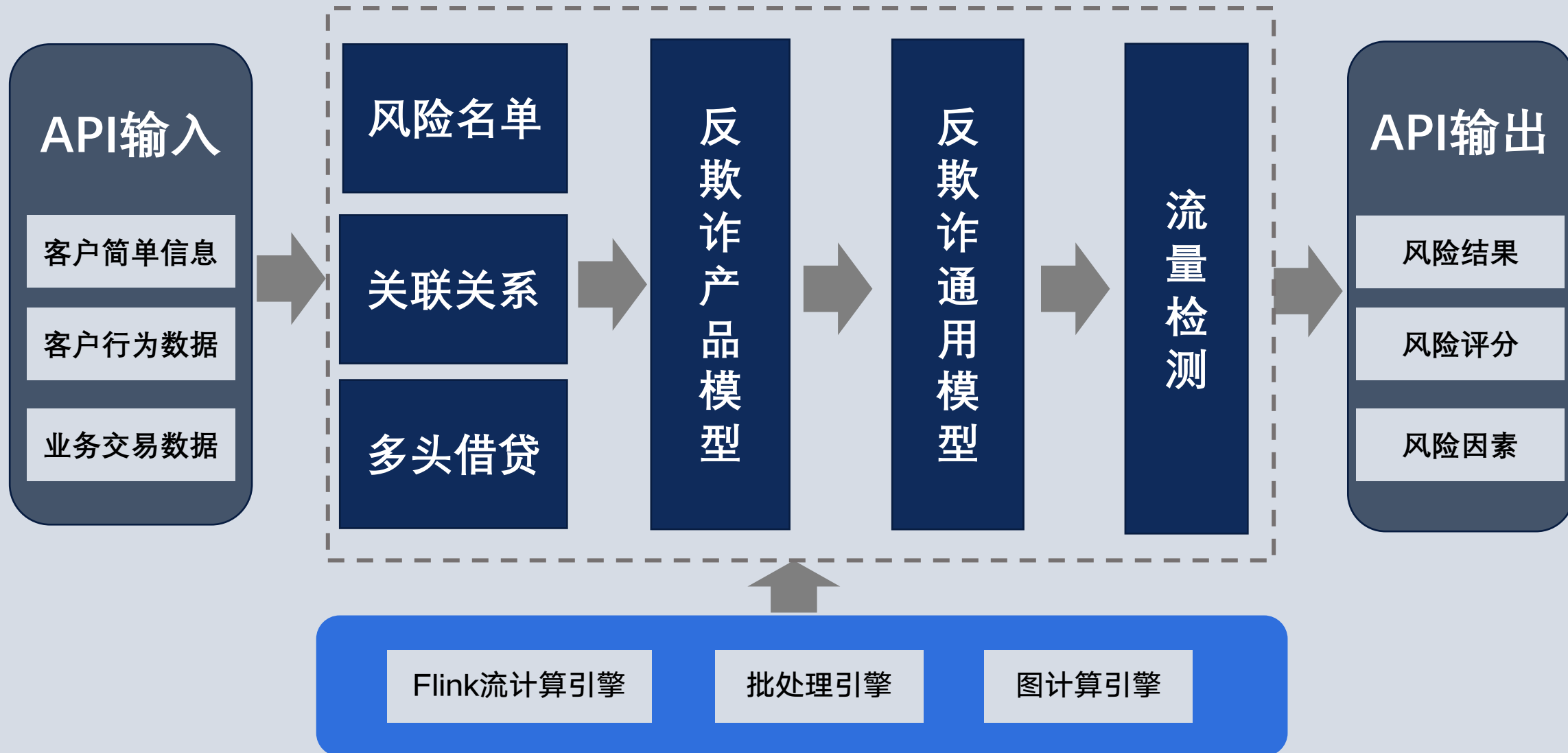
神荼整合了黑产关联、资金流向、担保、投资、亲属关联、行为关联等多种关系数据，构建了反欺诈知识图谱，涵盖法人个人全量信贷客户。节点规模千万级，关系上亿条。通过模块度、标签传播等算法，分析羊毛党、贷款中介、黑产团伙等欺诈行为。



羊毛党

黑产嫌疑

贷款中介





实时数据高效处置

基于Flink实时流计算构建高性能风险监控云图，实现秒级风险预警及处置数据展示，提升风险发现与处置效率。

多元数据接入

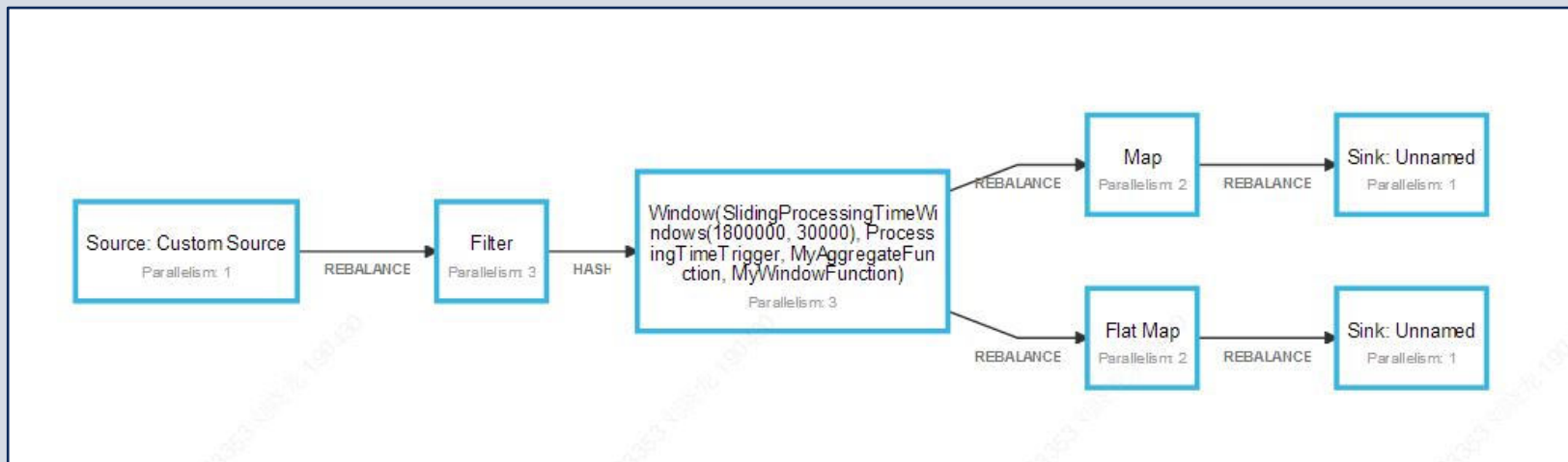
多元数据接入云图，灵活展示法人业务、个人业务、三农业务等多个产品风险数据，全面统览风险大盘。

统计维度丰富全面

将信贷风险以过检量、预警量、拦截量、发生地区、触发规则、每日趋势等维度统计，构建全行信贷风险实时“画像”。

使用的Flink规则类型

结合实际业务场景和需求，我们使用了四类Flink job来进行实时消息的反欺诈规则计算和加工：无容器的state类规则、带容器的state类规则、窗口类规则、CEP类规则。



无容器state类规则

进行简单的次数累加或者金额累加的场景

带容器的state类规则

需要进行去重、排序或其他处理并记数的场景

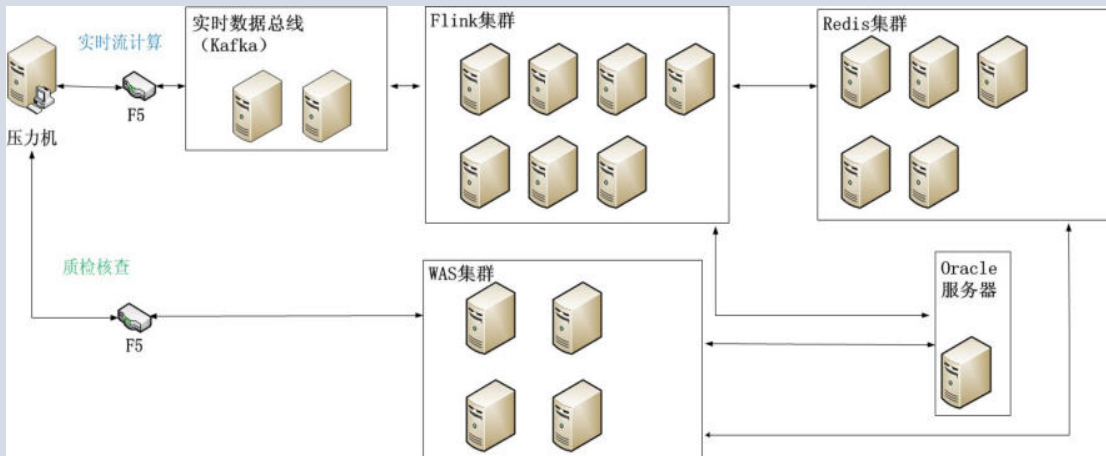
窗口类规则

对时间段内交易进行统计计算的场景

CEP类规则

对客户复杂交易模式进行筛选的场景

设备编号	软件配置	服务器配置	用途
PC Server 7台	Suse Linux 12 SP2Flink 1.6.0	CPU数量：32C，内存：256G实际分配给Flink的资源：CPU数量：8C，内存：24G	Flink流计算服务器集群
PC Server 3台	Suse Linux 11 SP3Redis 3.2.11	CPU数量：4C，内存：8G	Redis数据库集群
PC Server 1台	Suse Linux 12 SP3Oracle 12.1.0.2	CPU数量：4C，内存：8G	Oracle数据库服务器

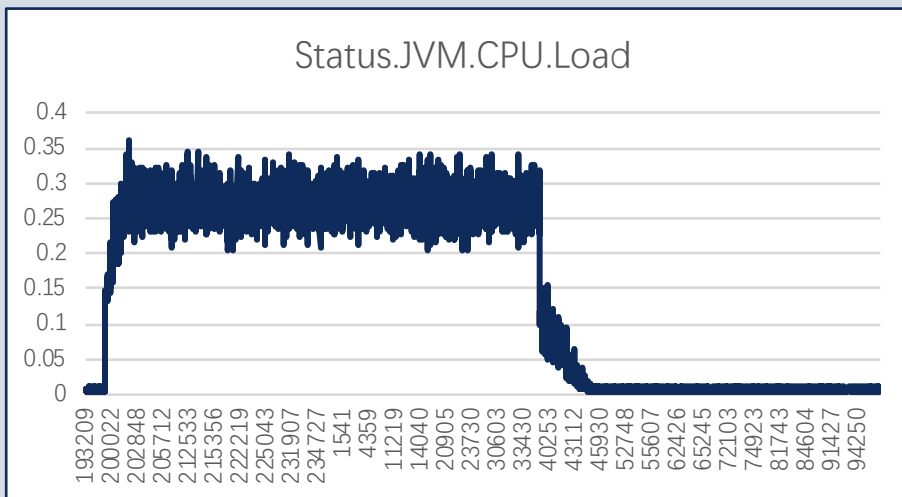


性能测试场景和方法

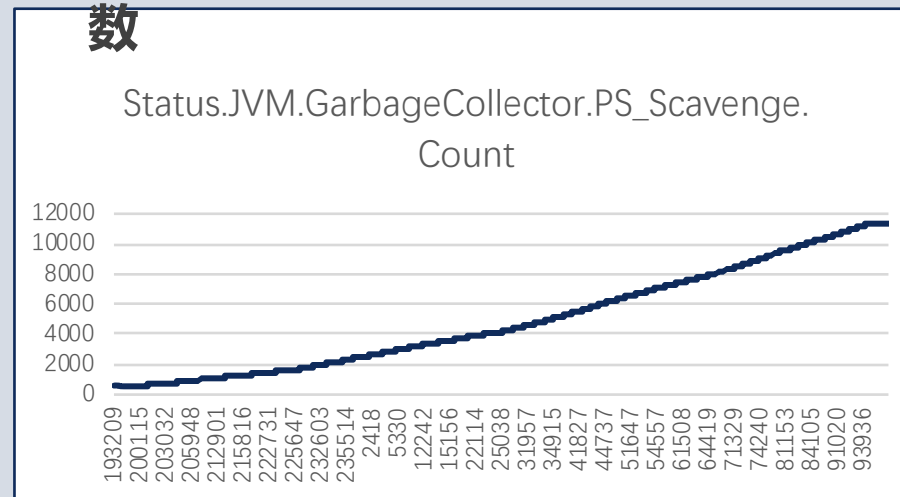
- 由于每次触发预警时，Flink都需要向Oracle和Redis写入数据，因此消息触发预警的概率对Flink的性能影响非我们进行了100%触发和1%触发情况下的性能测试；100%触发下是Flink性能最差场景；1%触发是模拟真实生产场景压力递增——提高发送消息速率到目标吞吐量，寻找性能拐点（极限吞吐量）
- 进行了单规则单slot基准压力测试、单规则多slot目标吞吐量压力测试、混合疲劳测试
- 关注指标：每条Flink规则的吞吐量、端到端延时，TaskManager的CPU、内存使用情况



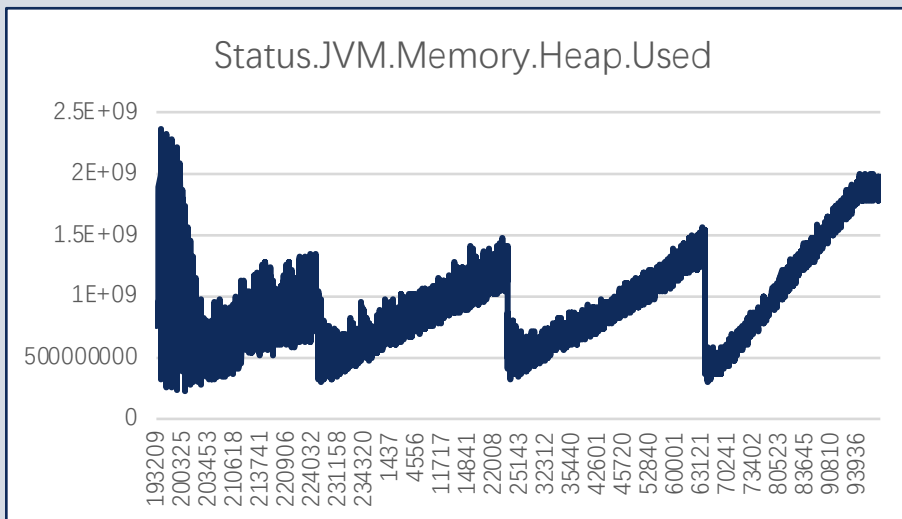
TaskManager的CPU消耗情况



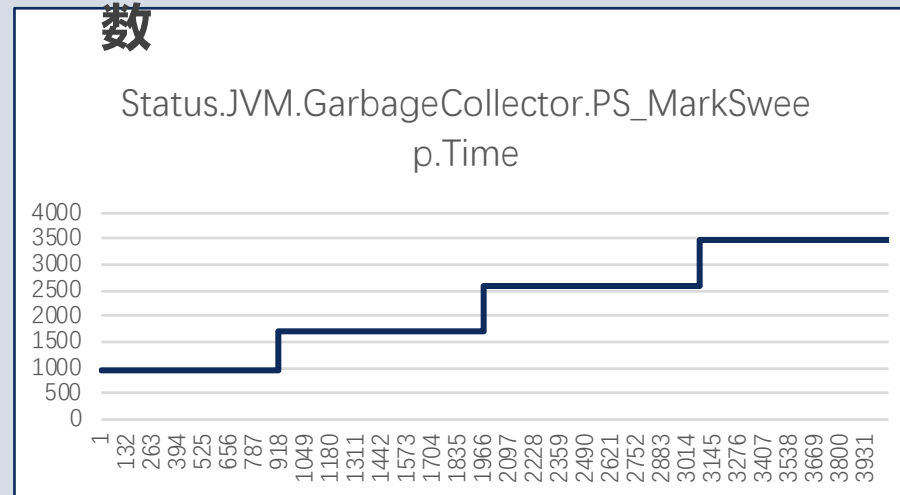
TaskManager调用YGC次数



TaskManager的堆内存使用情况



TaskManager调用FGC次数



PART 03

应用效果



日API调用量
100万

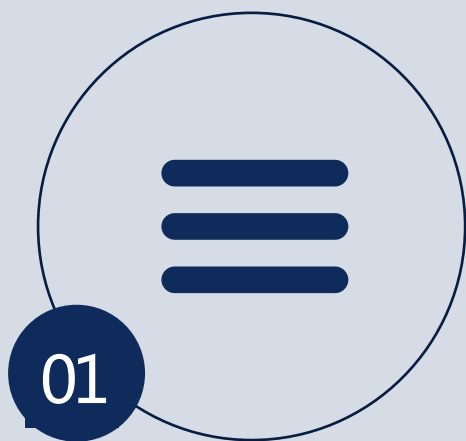
响应时间
10ms

拦截率
交易0.45%

不良率显著下降

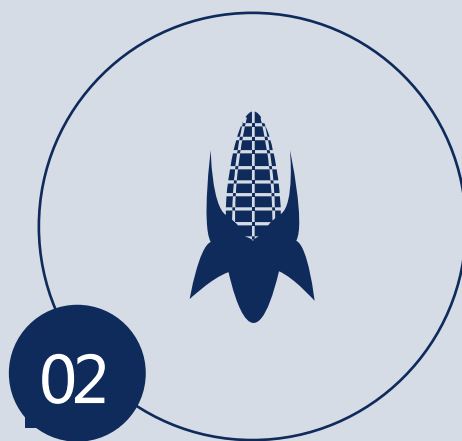
PART 04

思考展望



性能优化

Flink规则优化
参数调优



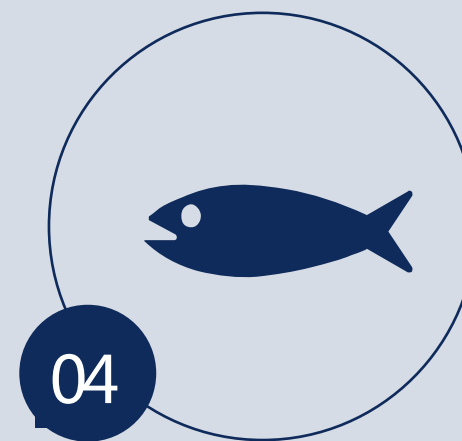
平台化

向更高版本的Flink迁移
批流一体计算平台



机器学习

更多的时序、聚类算法
反欺诈模型
规则优化



智能运行

限流算法
异常检测算法
资源管理
自动模型迭代



THANKS

Kafka X Flink Meetup

SHENZHEN