

# 流处理技术在事中反欺诈领域的应用

关键词：流处理技术 超低延时 非线性 超高并发 流立方



杭州邦盛金融信息技术有限公司CTO王新根

如何把当前最流行的流处理技术应用在反欺诈领域？

从整个金融行业的反欺诈技术路线来看，一般分为三种模式：

第一种是P处理的反欺诈识别，以反洗钱最为流行，反洗钱是其中最为常见的一种处理模式。第二种是准实时处理，准实时和批量分析本质上都是事后模式，准实时处理模式的时效性能够做到几秒钟或者说几分钟或者几小时的情况。第三种是实时处理。从实际场景来看，实时处理以事中处理模式最为有效。与准实时不一样，用户交易都需要实时通过风控引擎，由我们的风控决策来预判这个交易有没有风险，这种场景化应用能提高我们的客户体验。

事中反欺诈三大技术指标是我们攻克的主

要问题，即超低延时、非线性、超高并发。

我们怎样做反欺诈？

第一步对用户交易进行拦截，第二步基于拦截的交易，从数据库回溯交易中相关要素，过去一段历史时间内的流水，从数据库取回来，取回来后会进行事中预算。这些指标，比如当前登录的用户是A，A客户在过去24小时过去3个月内在哪些银行登录过，他的频次是什么，他的交易是什么，他过去在什么地方交易过，诸如此类的指标有几百个，第三步需要做的则是针对这些指标的预算。得到这些指标之后与当前的交易数据合在一起，再基于这些既定的规则进行比对（这个规则不一定是真实的规则，也有一种机器学习模型），还要进行决策，通过一个触发的规则列表给出一个最终风险定义。计算性能和延时如果能够解决掉，我们的事中反欺诈基本上也得到了完美的解决。

OLAP怎么发展？

最开始是数据库的批处理技术。我们的数据仓库运行模型一般需要几十个小时，经过几十个小时才能给出一个完整答案。谷歌推出一个理念，基本上解决了批处理的时效性问题，把原来几十个小时压缩在几小时、几十分钟甚至几分钟之内。批处理技术讲究的理念是先将原始数据累积下来，用的时候及时得到查询结果。往往数据的回溯计算耗时较长，最多延时几十毫秒为极限。

后来出现了流处理技术，与批处理技术相

比, 最本质的区别在于逻辑先行, 计算所有数据中每一用户过去24小时累积交易额逻辑, 这个逻辑是增量运营的。流处理中所有处理结果是动态的, 随着数据的进入不断地进行变化。跟传统的数据仓库不太一样, 流处理讲究分布式处理, 理论上我们可以做到在1毫秒以内解决事中反欺诈的场景。

流处理技术会解决很多问题, 第一是指标的存储问题。所有计算中, 结果怎么存储? 这些结果非常大, 要解决所有数据的每一个维度指标, 包括所有的IP、所有设备、所有仓库、所有地区的各种各样的维度, 一般而言不能放在一个机器上。此外, 特别是在金融领域, 要求我们存储数据是高可靠、高可用和高保障。

第二是算法需增量计算。所有的流处理算法一定要增量, 需要把我们原来在大数据处理空间上的应用转化到时间上, 这里的简单算法很好做, 比如求和、平均。在金融领域, 特别是在反欺诈里面经常用到的算法, 连续的统计某一个用户过去连续的、最大的失败交易指数是多少, 这个用户在他过去24小时连续的递减交易, 这类在流动性检测、市场风险检测波动的算法, 包括一些排序、趋同的计算, 这些都需要流处理实现。

第三是事件序列识别, 对定位风险的时候非常有帮助。在判定当前用户有没有欺诈嫌疑时, 我的对比是跟他类似所有用户的结果, 所有用户是个大维度, 这个数据量非常密级, 原来是30天, 后来发现不行, 得具备统计30年的长周期能力, 这是要解决的一些难点。

流立方就是在这样一个背景下兴起的云处理技术, 这个平台本质上是流处理平台, 所有数据流过即处理, 处理以立方体的形式保存。立方体细分为三个维度: 数据维度、计算指标和时间窗口。

我们要让数据动起来, 系统时间在移动的时候, 在我们立方体里的数据也在变化。对于某一个用户过去24小时的值在发生变化, 刚好有几笔交易在一两秒钟之内发生, 24小时精准定位需要剔除, 窗口数据已经到来, 我们的数据是动态的。后面的数据加进来, 前面流失的数据也需要解决。数据的时间窗口要进行任意调整, 比当前时间往前推24小时, 这是非常重要的几点。

流立方具备基于时间窗口移动的动态数据快速处理技术(时序处理) 以及基于事件驱动的模式识别技术(CEP), 支持计数、求和、平均、最大、最小、方差、标准差、K阶中心矩、连续、递增/递减、最大连续递增/递减、唯一性判别、采集、过滤、排序等多种分布式实时计算模型。

流立方平台本质上是流式处理, 高存速、时序存储、复杂事件是我们自有的特色, 流立方有三倍效率的提升度。

我们的流立方平台应用反欺诈是怎么做的呢?

通过设备发起一笔交易, 这笔交易流入电子银行业务和信息系统会被拦截, 并进行识别。在识别过程中, 用到历史数据得到的指标是流立方平台实时计算出来的, 计算完后告知当前这笔交易有没有风险, 这笔交易没有风险就会流入核心电子平台业务系统, 流入业务系统后, 这笔流水会在流立方的计算引擎中进行指标计算, 并存储到立方体里, 效率非常高。

此外, 流处理技术在实时报表、仪表盘方面有很大的价值, 通过采集业务系统的数据进行流处理, 把指标数据在图表上进行展现, 可以取到一个非常好的结果。这样的图表在高并发、低延时方面都有先天的特性。● (根据演讲内容整理, 未经本人审核)