

高性能的多维分析解决方案

背景

以某国内最大综合类证券公司为例，介绍 KAP 的高性能多维分析解决方案。该证券公司注册资本 60 亿元以上，经营业绩稳居业内前三，经营管理、风险控制、合规体系、信息技术等水平领先，是国内最早开展各类创新业务的券商之一。凭借全方位的业务创新、服务创新和管理创新，成就了一系列市场第一。

伴随着大量的线上交易操作的完成，该证券企业的业务系统已经积累了丰富的后台数据和信息。如何了解客户的交易习惯，掌握理财产品的销售状况，以及完善客户画像，识别潜在客户成为了证券行业急切依赖大数据去解决和发现的问题。

需求

该券商的大部分交易数据都存储在传统的数据仓库中，Hadoop 平台存储了海量的 APP 日志数据和第三方数据，但是由于缺少适合的分析工具，基于 Hadoop 平台的交互式分析应用相对较少。随着数据量爆炸性增长，在传统的数仓中查询分析响应速度越来越慢，且不易于扩展。特别是针对非重复计数等指标的实时计算需求，在大数据量的场景下更是无法满足。

该券商希望基于互联网客户的行为分析和经营分析，通过使用大数据分析平台及技术，使得超大规模数据集上的多维分析、交互式分析能力开放给各运营部人员，从集中式、专家式数据分析演进到分布式、全民数据分析式的时代，人人成为数据分析师，满足公司在大数据场景下的超大数据量、多维度、高安全隔离、高并发条件下的数据分析和数据服务能力，为即将到来的更具挑战的实时分析、数据挖掘、个性化营销等场景打下坚实的基础。

痛点：

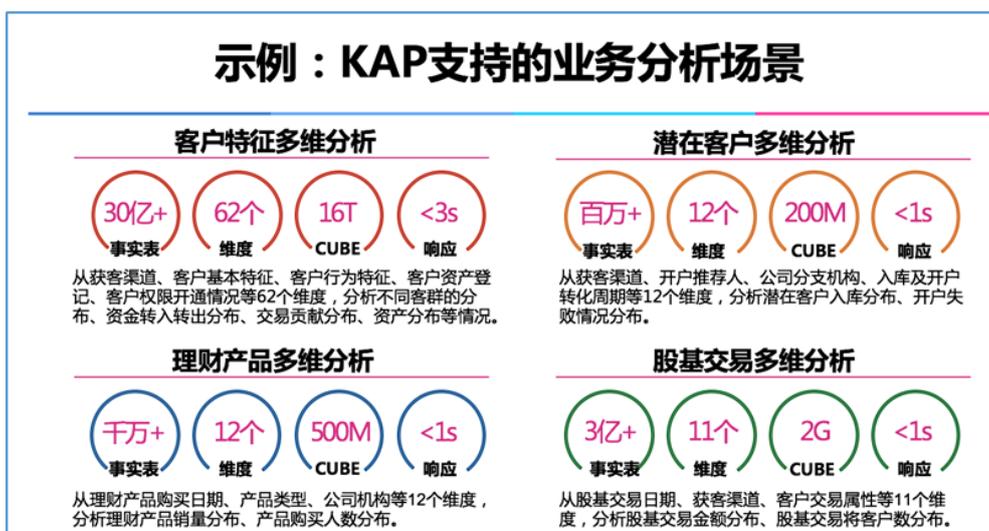
- 最大的 Cube 每天构建时间超过 20 小时，业务部门只能看到 2 天以前的数据
- 基于 Cube 的单次多维分析需要 1 小时才能返回结果，严重影响业务部门的工作效率和正常分析流程
- 将 Cube 数据固化成报表文件来提升查询性能，带来大量额外维护成本的同时，彻底丧失了通过 Cube 进行 OLAP 多维分析的灵活性
- 由于权限管理问题，同一类 Cube 要为全国 30 多个省分行建立副本，导致 Cube 数量达到 2000+，大大增加了 IT 部门的运维难度和工作量

解决方案

通过将 OLAP on Hadoop 的大数据分析引擎 KAP 搭建在企业已有的 Hadoop 大数据平台之上，作为 Hadoop 底层数据与用户前端查询工具之间的桥梁，很好地减轻了 IT 人员数据准备的工作，提升了平台查询性能，并满足了业务人员自助分析的需求。

此次项目主要包含潜在客户分析、理财产品交易分析、客户特征分析等业务主题。采用增量构建 cube 的方式来接入每天新增的交易业务数据，并进行自动合并，所有的 cube 构建过程在 1 小时内全部完成。其中最大的一个 cube 包含 30 亿以上的总记录数，以及 60 多个维度，整个 cube 的容量为 20T 左右，通过一定地聚合优化设置，平均查询响应时间低于

3s。



业务人员通过使用 KyAnalyzer 或 Tableau 等 BI 工具连接至 KAP 数据引擎，秒级获取汇总或明细查询结果，自助完成业务报表的创建与分享，极大地提高了工作效率，并减轻了 IT 人员对数据处理的繁重工作。另外，在安全方面，由于各营业部的信息需要互相隔离，同时对于不同业务范围的用户也需要进行严格的权限控制。KAP 的可配置权限管理功能能够很好地在 Cube 级别，以及单一 Cube 内实现数据行、列级别的权限控制，充分满足了企业对数据管控的需求，以保证证券数据的高度安全性。

客户价值

通过使用 KAP 来搭建企业的大数据平台数据仓库层，IT 部门能够通过可视化的界面来更方便地开发与管理 CUBE 模型，并根据业务分析主题来设置维度与度量属性，每一个 CUBE 都可以满足用户在同一分析主题下对各种维度进行任意组合的多种报表需求，极大简化了后台针对各种特定需求而二次开发的方式。

用户通过拖放的操作便可以快速开发出自己需要的报表，不再需要通过 IT 人员反复修改脚本来生成报表数据。大大缩短了数据到用户的应用过程，并降低了业务人员在大数据平台上使用数据的门槛，使最终用户可以直接面对授权内的数据进行自助探索与分析。

数据分析在大数据平台上的响应效率也得到了极大的改善。对于百亿条基础数据的分析需求，90%的查询能通过 KAP 在 3 秒内返回结果，其中包括对千万数量级记录的非重复计数，比在 HIVE 上查询快百倍。

客户反馈

“KAP 具有良好的兼容性，可以快速部署到我司 hadoop 集群中，还通过提供 SQL 的数据访问方式，可以很方便地作为查询引擎来使用，并给下游和前端提供数据，极大地提升了移动 app 客户行为分析、互联网活动运营、客户运营及经营分析等金融行业常用分析场景的数据分析效率。”

——信息技术部 大数据平台总监