

新一代交易分析

概览

挑战

- 以近乎实时的方式将 6 TB 的数据构建为 40 个 OLAP 多维数据集
- 处理模式更改并优化多维数据集数据的质量
- 支持用户对不断变化的多维数据集执行高级分析
- 客户尝试了 10 种解决方案，包括 RedShift、Snowflake 和 Exadata，期望能满足性能需求；但全都未能满足要求

解决方案

- 将工作负载从三个 Oracle Exadata 环境迁移到 Xcalar 虚拟数据仓库
- 通过插入/修改/删除 (IMD) 创建基于 CDC 的微批次，以便近乎实时地更新多维数据集
- 利用可视化数据流促进高级分析

价值

- Xcalar 将数据分析时间从 24 小时缩短到 10 分钟左右
- 现在，面向客户的分析师拥有了更新的市场数据，从而使他们能够为客户提供更好的建议

一家领先的财富管理公司需要创建新一代交易数据分析平台。主要的平台要求是，近乎实时地执行复杂的数据转换，进而更好地帮助交易者和交易经理做出数据驱动的决策，并为交易者提供洞察信息。客户还需要相应的支持，以确保持续的数据质量和一致性，并进行高级分析。通过与 Xcalar 合作，他们创建了一个新的分析平台，实现了所有目标，既能以近乎实时的速度更快实现数据可用性，又能将数据准备时间从几个月/几周缩短到几天/几小时。此外，这一变化为开发人员创造了应用高级分析技术的机会。

挑战

将 6 TB 不断变化的数据构建为 40 个 OLAP 多维数据集是一个耗时且容易出错的过程。哪怕是源数据中的一个模式更改，就会破坏大多数数据仓库平台进程，而评估数据质量根本不属于数据仓库技术的功能。对于这家领先的财富管理公司的 Oracle Exadata 环境，用 6 TB 数据构建 40 个多维数据集需要 24 个小时。对于他们的下一个开发阶段而言，必须要向多维数据集数据应用简单的分析操作，而应用高级分析是他们最终的目的。

解决方案

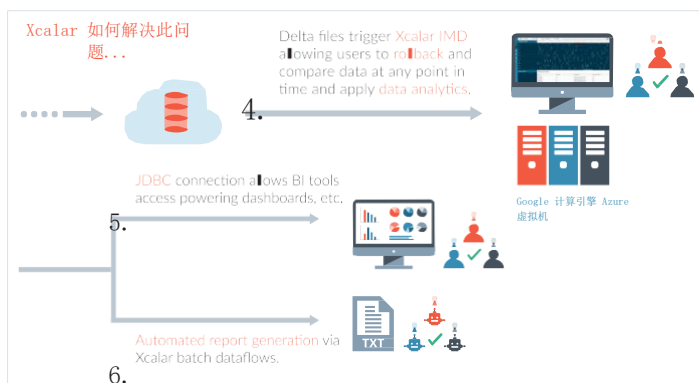
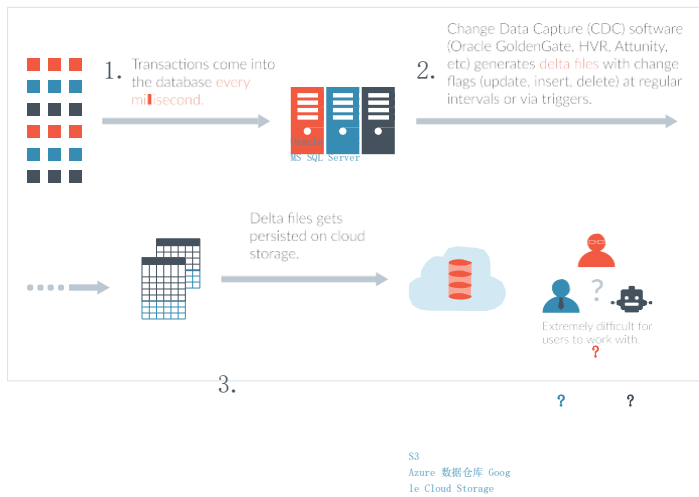
解决方案是构建一个 Xcalar 集群，负责创建 OLAP 多维数据集。用户通过使用 Xcalar Design 的 SQL 运算符运行以前存储的 SQL 过程来创建其 OLAP 多维数据集。Xcalar 计算引擎具有在多个包含数亿行内容的数据集上对数据流建模的规模和性能。

Xcalar 与客户合作开发了新的产品功能，包括实时微批插入/修改/删除 (IMD) 功能，以及用于时间序列窗口化的交互式横向扩展转换。此 IMD 功能可实现亚秒级的数据插入、修改和记录删除，同时确保以原子方式更新所有表，从而解决了当前解决方案的问题。向 OLAP 多维数据集应用 IMD 后，用户随后即可通过相同的 Xcalar 平台结合使用 SQL、编程自定义代码或以直观方式与其多维数据集数据进行交互。在此阶段，他们新获得了评估和优化数据质量的能力，甚至能向他们的数据应用高级分析技术，从而创建新的数据流。

此外，Xcalar 计算引擎可针对更大的数据集或根据源数据的特定时间戳运行这些数据流。Xcalar 将最复杂转换的处理时间从超过 10 分钟缩短到 27 秒。Xcalar 虚拟数据仓库跟上了数据生成的速度，使最终用户能使用最新的数据，而不是旧数据。这就使客户无需再不断管理相同数据的不同副本。Xcalar 成功证明，Xcalar 虚拟数据仓库的速度、简便性和可扩展性大大提高了数据管道的效率，同时降低了数据移动开销和解决方案的复杂性。

Xcalar 的主要优势:

- **无缝集成:** 客户通过提供一组丰富的 API 将 Xcalar 集成到了他们现有的数据管道中。通过在现有的 Python 程序中仅另行添加 10 行代码, 客户的变更数据捕获 (CDC) 软件就能启动 Xcalar 数据流。只需点击几下, Xcalar 就能利用 AWS 的 CloudFormation 模板部署 Xcalar 集群。
- **双时态数据分析:** 数据进入系统后, Xcalar 可跟踪数据批次, 使用户能够回滚到任何时间点。
- **无摩擦数据导出:** Xcalar 可将数据导出到任何目标软件或存储。这就让客户能够将最终的表导出到 Amazon S3 存储, 将元数据导出到数据编目软件, 使其可用于 Tableau 的查询。
- **亚秒级插入、更新、删除操作:** 借助 Xcalar, 系统可在一秒钟内以原子方式将变更应用于涉及的所有表, 并且最终用户可以立即看到这些变更。用户无需再等待其仪表板更新。
- **多表事务边界:** 这位 Xcalar 客户使用的是规范化表。这就要求必须一次性将所有事务应用到多个表。Xcalar 可确保在任何表上都不会发生部分更新, 并且所有表更新都以原子方式进行。



主要特性、产品和服务

主要特性

- 通过时间点回滚对虚拟表进行事务性插入/更新/删除 (IMD) 操作
- 高性能 Tableau 连接器
- JDBC/ODBC 连接
- 可使用可视化编程、SQL 和结构化编程进行灵活开发
- 机器学习集成
- 从节点故障中快速恢复

产品

- Xcalar 数据平台高级版
- Xcalar Design 企业版
- Xcalar 企业管理器

服务

- 产品培训
- 解决方案架构和设计
- 设置、配置和监控基础架构与自定义 CDC 解决方案集成
- 用户自定义函数数据导入/导出
- 转型设计与实施
- 数据流设计和实施
- 集群大小调整和性能优化

关于 Xcalar

Xcalar 是一个可扩展的开放性分析平台, 适用于完整的分析管道, 包括数据质量、虚拟数据仓库、数据科学和运维。用户可使用可视化编程、SQL 和结构化编程以交互方式构建数据流, 并对非结构化、结构化和半结构化数据以 PB 级别执行这些数据流。Xcalar 的企业级软件可扩展到数百个节点和数千个用户, 既可进行云部署也可进行本地部署。凭借其专利技术, 用户能轻松、快速、大规模地获得可操作的见解。