

Building a Next Generation Trading Analytics Platform

CHALLENGES

A leading wealth management firm faced challenges in creating a next-generation trading data analytics platform that would need to:

- Build 6 TB of data into 40 OLAP cubes in near-real-time
- Handle schema changes smoothly
- Allow users to perform advanced analytics on dynamic data cubes

SOLUTION

Xcalar assisted the firm with moving its workload from three Oracle Exadata servers to Xcalar Data Platform, creating cubes that are updated in near-real-time by CDC-based micro-batches with insert/modify/delete (IMD) capabilities, and facilitating advanced analytics using visual dataflows.

BENEFITS WITH XCALAR

- Overall time to prepare new data for analysis is reduced from 24 hours to 27 seconds
- With dashboards updated in near-real-time, customer-facing analysts can make better recommendations to customers
- **Bitemporal Data Analysis:** Comprehensive point-in-time rollback capabilities helps with risk management and regulatory compliance
- **Multi-table Transactional Boundaries:** Robust transactional consistency across normalized tables reduces memory and storage needs

Introduction

A leading wealth management firm needed to create its next-generation trading analytics platform. The main platform requirement was to perform complex data transformations in near real-time and therefore, to better serve traders and trading managers in data-driven decision making. The client also required support for continuous data consistency and data management, as well as advanced analytics. Through working with Xcalar, they created a new analytics platform that achieved all objectives, both by accelerating data availability to near real-time and reducing data preparation time from months/weeks to days/hours. In addition, the new platform created opportunities for developers to apply advanced analytics techniques.

Challenges

In the firm's Oracle Exadata environment, building 6 TB of constantly changing data into 40 OLAP cubes was a 24-hour-long process, that was not capable of providing analysts with more recent data to work with. The process was highly error-prone and any schema changes in the source data would break most workloads. They had an increasingly critical business need to apply advanced analytics operations on their cube data. Data quality management was also a significant challenge that needed to be addressed.

Xcalar Solution

Using the intuitive and versatile Xcalar Design studio and IDE integrated into Xcalar Data Platform, legacy stored procedures were easily modeled as visual Xcalar dataflows that created OLAP cubes. Xcalar Data Platform provided the scale and performance to model dataflows on multiple diverse datasets with hundreds of millions of rows and then efficiently operationalized them for optimal performance.

Xcalar's real-time micro-batch Insert/Modify/Delete (IMD) capability allowed scale-out transformations for time series windowing. This IMD feature addressed the issues with their current solution by seamlessly integrating with HVR, their chosen change data capture (CDC) solution. Other CDC solutions such as Oracle GoldenGate, Attunity, etc., could also have been used. Xcalar ingests the deltas that are constantly sent by the CDC solution and applies the necessary sub-second insert, modify or delete operations on rows of data, while ensuring that updates are transactionally consistent across all tables.

The firm's analysts use data visualization tools, both Tableau and Qlik, to access near-real-time data. Xcalar Design users work interactively with the constantly updated OLAP cubes using a combination of SQL, Python code, and visual programming within the Xcalar Design interface. They have the ability to model and operationalize new dataflows to assess and refine data quality, as well as to apply advanced analytics algorithms to their data.

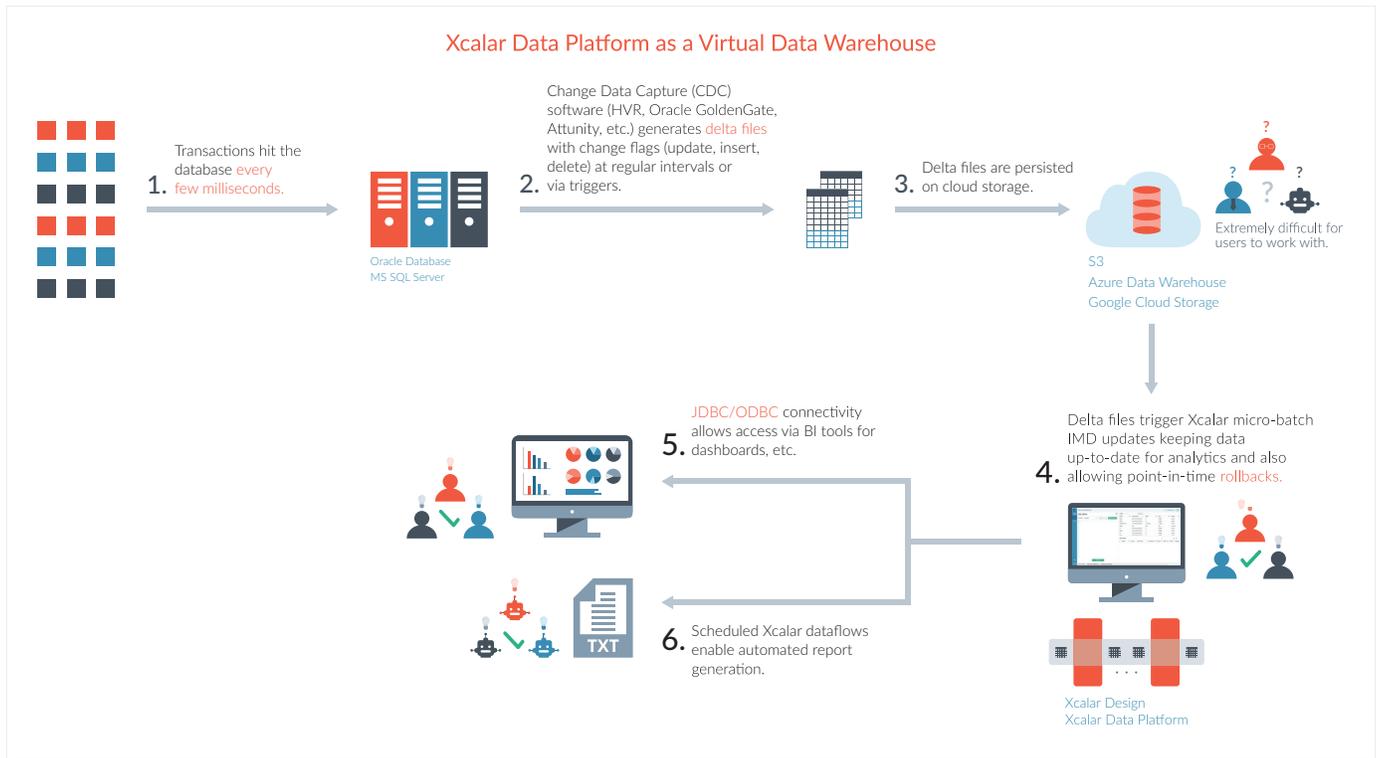


Figure 1: Xcalar enables BI tools to query near-real-time data

In addition, Xcalar Data Platform automatically operationalized these dataflows on larger datasets or according to specific timestamps in their source data. Xcalar’s comprehensive point-in-time rollback capability supports their bitemporal data model to facilitate risk management and regulatory compliance. The processing time of their most complex transformations was reduced from 24 hours to 27 seconds. With Xcalar Data Platform keeping pace with the speed of data generation, end users can now work with the current data instead of the old data. This eliminates the need to constantly manage different copies of the same data. Xcalar has demonstrated that the speed, simplicity, and scalability of the Xcalar Data Platform dramatically improves the efficiency of their data pipeline while reducing data movement overhead and solution complexity.

KEY FEATURES

- Transactional Insert/Update/Delete (IMD) operations to virtual tables with point-in-time rollback
- High-performance Tableau interface via JDBC/ODBC connectivity
- Visual programming, SQL, and structured programming development flexibility
- Integration with Machine Learning libraries such as TensorFlow, SparkML and H2O.ai
- Checkpointing for rapid recovery from node failures

About Xcalar

Xcalar is a scale-out platform for data processing applications and operationalizing ML. The platform is open and extensible, and suitable for developing and operationalizing business logic. Xcalar’s use cases include virtual data warehousing to enable BI tools to query real-time transactionally consistent data, operationalizing ML algorithms at cloud-scale, as well as simplifying data transformation and quality processes. Users use a versatile IDE to interactively build dataflows using SQL, visual programming, and structured programming, and execute them at petabyte scale. Xcalar’s enterprise-grade software scales linearly to hundreds of nodes and thousands of users for public/private cloud and hybrid deployments.