

Next Generation Trading Analytics

AT A GLANCE

Challenges

- Building 6 TB of data into 40 OLAP cubes in near real-time
- Handling schema changes and refining the quality of cube data
- Allowing users to perform advanced analytics on changing cubes
- Client tried 10 solutions, including RedShift, Snowflake, and Exadata, to meet performance needs; none met requirements

Solution

- Lifted and shifted workloads from three Oracle Exadata environments to Xcalar Virtual Data Warehouse
- Created CDC-based micro-batches with insert/modify/delete (IMD) to update cubes in near real-time
- Facilitated advanced analytics using visual dataflows

Value

- Xcalar reduced the overall time to analyze new data from 24 hours to 27 seconds
- Customer-facing analysts now have fresher market data, enabling them to make better recommendations for customers

A leading wealth management firm needed to create their next generation Trading Data 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 and trader insight. The client also required support for continuous data quality and consistency, as well as for 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 change created opportunities for developers to apply advanced analytic techniques.

Challenges

Building 6 TB of constantly changing data into 40 OLAP cubes is a time consuming and error prone process. A single schema change in source data breaks most data warehousing platform processes, and assessing data quality is simply not a capability of data warehouse technologies. For this leading wealth management firm's Oracle Exadata environment, crafting 40 cubes out of 6 TB of data was a 24-hour process. Applying simple analytics operations to cube data was simply necessary for their next stage of development, and applying advanced analytics was where they wanted to go.

Solution

The solution was to build a Xcalar cluster that would operationalize the burden of OLAP cube creation. Users created their OLAP cubes by running their legacy stored SQL procedures using Xcalar Design's SQL operator. Xcalar Compute Engine provides the scale and performance to model dataflows on multiple datasets with hundreds of millions of rows.

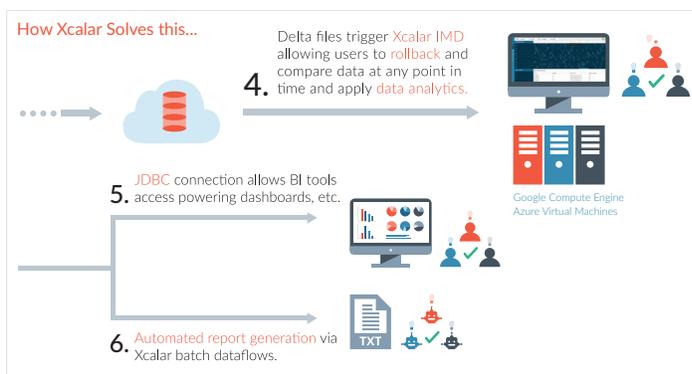
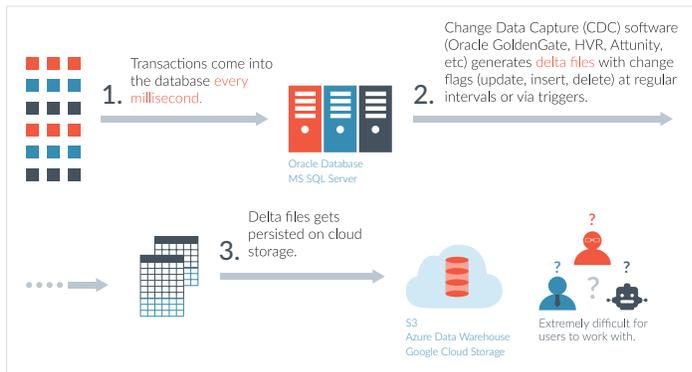
Xcalar worked with the client to develop new product features, including a real-time microbatch Insert/Modify/Delete (IMD) capability with interactive scale-out transformations for time series windowing. This IMD feature addressed the issues with their current solution by enabling sub-second data inserts, modifies, and deletes of records, while ensuring that the updates are atomic across all tables. After applying IMD to their OLAP cubes, then users interacted with their cube data using a combination of SQL, programming custom code, or visually using the same Xcalar platform. At this phase, they were newly able to assess and refine data quality, and even apply advanced analytics techniques to their data, creating new dataflows.

In addition, Xcalar Compute Engine operationalized these dataflows on larger datasets, or according to specific timestamps in their source data. Xcalar reduced the processing time of their most complex transformations from 24 hours to 27 seconds. Xcalar Virtual Data Warehouse kept pace with the speed of data generation, allowing end users to work with current data

instead of old data. This eliminated the need for the client to constantly manage different copies of the same data. Xcalar demonstrated that the speed, simplicity, and scalability of the Xcalar Virtual Data Warehouse dramatically improves the efficiency of their data pipeline, while reducing data movement overhead and solution complexity.

Key Xcalar Benefits:

- **Seamless Integration:** The client integrated Xcalar into their existing data pipeline by providing a rich set of APIs. By adding only 10 additional lines of code to their existing Python program, the client change data capture (CDC) software initiated Xcalar's dataflows. Xcalar is able to utilize AWS's cloud formation templates to deploy Xcalar clusters with a few clicks.
- **Bitemporal Data Analysis:** Xcalar tracks data batches as they come into the system and enables users to rollback to any point-in-time.
- **Frictionless Data Export:** Xcalar exports into any destination software or storage. This allowed the client to export the final tables to Amazon S3 storage, the metadata to Data Cataloguing software, and made it available for queries from Tableau.
- **Sub-second Insert, Update, Delete Operations:** With Xcalar, changes are applied atomically to all tables involved in under a second, and changes are immediately visible to the end user. Users no longer wait for their dashboards to update.
- **Multi-table Transactional Boundaries:** This Xcalar client uses normalized tables. This mandates that all transactions must be applied to several tables at once. Xcalar ensures both that no partial update happens on any table, and that all table updates happen atomically.



KEY FEATURES, PRODUCTS, AND SERVICES

Key Features

- Transactional Insert/Update/Delete (IMD) operations to virtual tables with point-in-time rollback
- High performance Tableau connector
- JDBC/ODBC connectivity
- Visual programming, SQL, and structured programming development flexibility
- Machine Learning integration
- Rapid recovery from node failures

Products

- Xcalar Virtual Data Warehouse
- Xcalar Design Enterprise Edition
- Xcalar Enterprise Manager

Services

- Product training
- Solution architecture and design
- Infrastructure setup, configuration, and monitoring
- Integration with custom CDC solution
- User-Defined Function Data import/export
- Transformation design and implementation
- Data flow design and implementation
- Cluster sizing and performance tuning

About Xcalar

Xcalar is an open and extensible analytics platform for the complete analytics pipeline that includes data quality, virtual data warehousing, data science, and workload operationalization. Users interactively build dataflows using visual design, SQL, and structured programming, and execute them at petabyte scale on unstructured, structured, and semi-structured data. Xcalar's enterprise-grade software scales to hundreds of nodes and thousands of users for both cloud and on-premises deployments. Its patented technologies deliver actionable insights with simplicity, speed, and scale.