

CASE STUDY

Transforming Data Management with Databricks for a Health & Fitness Club



OVERVIEW

The client is a prominent health and fitness club franchise operating across 50 countries in over 5,000 locations with more than 3,000,000 members worldwide.

CHALLENGES

The client's increasing data volume led to concurrency challenges during data read and write operations, affecting data consistency and operational efficiency. Complex ETL processes hindered data processing efficiency, requiring a robust solution for smooth aggregation in the Gold layer to streamline analytics and reporting. Delays in data availability impacted report generation and incorporating new data sources was cumbersome due to rigid schemas, necessitating a unified platform for analytics, business intelligence (BI), and machine learning for flexibility.

INDUSTRY

- Health & Fitness

SERVICES

- Lakehouse
- Reporting
- Machine Learning

TECHNOLOGY

- Databricks Lakehouse
- Delta Lake Architecture
- Databricks SQL Warehouse
- Delta Sharing



SOLUTION

Contata addressed the challenges by leveraging Databricks. We implemented Delta Lake as part of its Lakehouse architecture, combining the best features of data lakes and data warehouses.

Delta Lakehouse

The implementation of Delta Lakehouse ensured schema integrity on write, addressing the concurrency issue during data read and write operations. The adoption of the open Delta Parquet file format facilitated efficient CRUD operations and time travel features, enhancing the gold layer's aggregation process. Optimization features like data compaction and caching significantly improved query performance, contributing to a more efficient storage system.

Databricks Lakehouse

Utilizing Azure Data Lake Storage Gen2, Databricks Lakehouse provided a key solution for cost-effective and scalable storage, addressing challenges related to data accessibility and volume. The unified analytics platform accommodated structured and semi-structured/unstructured data, serving as a single platform for analytics, business intelligence, and machine learning on top of Delta Lake. The schema-on-read approach ensured enhanced flexibility, while robust data governance features addressed stringent security standards, ensuring data integrity.

Unity Catalog

The integration of Unity Catalog simplified metadata management, aiding in the process of data discovery and maintaining consistent governance practices. Unity Catalog's integration with Delta Sharing enhanced secure data sharing, fostering collaboration across organizations and platforms. Data masking techniques were also implemented to ensure data security, protecting sensitive information and complying with privacy regulations.



BENEFITS

- **Simplified Implementation:** Faster data load and pipeline creation due to the streamlined processes enabled by Databricks and Delta Lake.
- **Seamless Migration:** Reporting table code seamlessly migrated from SQL to Databricks Notebook, ensuring a smooth transition and continuity in reporting.
- **Cost-Effective Solution:** Databricks provided a cost-effective and faster solution compared to ADF + SQL, contributing to financial savings for the client.
- **Versatile Data Storage:** Increased ability to store various data types without concerns about cutting table sizes, enhancing data storage flexibility and adaptability.

About Contata

Contata Solutions is a trusted leader in technology and digital innovation. Through our work in data engineering, data analytics, machine learning, marketing automation and app development, we deliver solutions that address complex problems in ways that are simple, insightful and impactful.

Our promise and value proposition to our customers is simple: we leverage our deep technical expertise and global presence to bring software products and data-driven decision capabilities to life.

Founded in 2000, Contata is a privately-held company headquartered in Minneapolis that serves clients globally from offices in the United States and India.