



Data Transformation Solution for BFSI

A scalable solution for managing and transforming large financial data volumes. It enhances pipeline performance, ensures compliance, and improves risk prediction accuracy.

Data Transformation Solution for BFSI

Background

In the banking sector, managing and transforming large volumes of user and retail data is crucial for optimizing operations, ensuring compliance, and making data-driven decisions. A major financial institution needed to streamline its data handling processes and enhance pipeline performance to improve efficiency and accuracy in data transformation and risk prediction.

Key Challenges

Complex data handling, managing vast amounts of user and retail data across multiple platforms and systems. Pipeline Performance optimization to ensure efficient data transformation and reduce processing time. Automation is needed to reduce manual intervention in managing database schema updates across multiple tables. Scalability & storage

to handle large-scale data ingestion and transformation while maintaining high performance. Financial risk predictions by utilizing cloud-based solutions to enhance predictive analytics capabilities.

Our Solution

Developed a solution to streamline column addition across all data tables, eliminating the need for manual updates. Optimized Data Warehousing by implementing Slowly Changing Dimension Type logic to efficiently load data, also ensured historical data tracking. Built and maintained pipelines for large-scale data ingestion, transformation, and loading to enhance efficiency and reduce latency. Cloud-Based Enhancements for denormalized dataset management, allowing real-time analytics and enhanced financial risk predictions.

Tech Stack

Data Warehousing & Storage: Snowflake, Hive, MySQL, Greenplum, AWS S3

Data Processing & Transformation: Python, Apache Spark, Hadoop

Cloud Infrastructure: AWS

Value Delivered

Automation of schema updates led to a significant reduction in manual effort and potential human errors.

Adoption of SCD-2 logic ensured consistent and accurate tracking of historical data.

Spark-based ETL pipelines improved data ingestion speed and overall system performance.

Integration with AWS and Snowflake facilitated real-time analytics, empowering the bank with more accurate financial risk predictions.