

Data Lake Integration project delivers multi-million dollar ROI for Luxury Travel Retailer



Summary

Our customer is a leading luxury travel retailer with stores across four continents. We worked with them to integrate their existing enterprise data warehouse with a Hadoop data lake, resulting in vastly improved data availability for their end users. Additionally, the project delivered an ROI of over two million dollars by reducing data warehousing costs and enabling more precise demand forecasting.

Industry

Luxury Travel Retail

Technologies

Hadoop, Netezza, Talend, DataStage, Spark, HortonWorks Platform

Users

General Managers, AGM Operations, Operations Support Managers, LBS Managers, Operations and Control Managers, Cashier Lead

Team

9 InfoCeptions and 3 Customer Associates

→ The Challenge

Our customer, a leading luxury travel retailer, owns 400+ boutiques across four continents. The incoming data from these outlets is voluminous. Their existing data warehouse was unable to handle large scale processing loads and could not ensure the availability of consumable data. Consequently, they decided to upgrade to a state-of-the-art business analytics system.

After studying their needs in more detail, the decision was made to integrate their enterprise data warehouse with a Hadoop data lake. However, this upgrade would come with its own set of challenges, such as:

- Maintaining all major functionalities (like backroom metadata, audit, duplicate checks, null check, and referential integrity check) of ETL stages even after an ETL offload. The offload is required to transfer large amounts of ETL processing from IBM DataStage to Hadoop, reducing the load on the ETL product. This ETL offload will increase data warehouse availability for the end user.
- Establishing a unified framework that reuses existing code with minimal changes from Netezza SQLs to Hadoop/Hive SQLs.
- Ensuring that the data lake's framework uses backroom metadata and a logging functionality similar to their data warehouse. The similarity ensures quick tracking and reconciliation.
- Configuring the most efficient mechanism to push end-state data & tables from data lake to data warehouse for increased productivity.

→ The Solution

We collaborated with our customer's team to import raw data into the Hadoop data lake. Next, we used the data lake's staging area to check and improve data quality & referential integrity and create a surrogate key. These functions enabled us to synchronize data between the on-premise Netezza warehouse with the Hadoop database.

Additionally, we developed a working prototype of the Hadoop data lake. The prototype was used to verify the actual data lake's ability to maintain all the major processes of ETL stages. The verification enabled us to analyze the customer's IBM DataStage ETL framework and ensure its reusability with minimal effort.

Finally, we used the Hadoop data lake's better processing power to provide quick availability of all data required by end users. The data lake integration enabled orchestration, business rule processing, and cluster processing for faster data output.

→ The Results

Our solution successfully delivered the following benefits:

- Increased ROI – An estimated \$2.45 million of ROI facilitated by the Hadoop data lake's near real-time supply of data. Real-time reports offer the potential to reduce costs related to inventory, travel dynamics, weather, and customer interests through better demand forecasting.
- Rapid data access – Users no longer have to wait overnight for the ETL process to complete. It enables our customer to develop strategies faster, plan effective promotions, improve customer service and evaluate product category performance.
- Enhanced IT Initiatives – The data lake integration provides a foundation for future-proofing the customer's business. It enables users to access dormant data without help from the IT department. IT teams now have much more time to focus on more challenging and productive tasks.