# Impala to Databricks Transformation

The client is a Swiss multinational investment bank and financial services company that also provides wealth management, asset management, and investment banking services for private, corporate, and institutional clients with international service. They wanted to improve and consolidate their refinery and central data store for analytics on a unified data platform including business logic, as part of their data transformation strategy on cloud. The client wanted a complete platform modernization and enhance its performance while making it feasible for all dependent data applications as well as customers to be able to adopt the new unified platform at ease. In order to drive end to end transformation, Wipro developed an automated BOT as part of IntelliProc to migrate the data to Azure Data Lake, migrate existing script to Spark SQL/API and deploy it on Databricks platform. This helped the client to modernize their existing big data applications at scale and leverage the benefits of a unified data platform – Databricks, on Azure

## At-a-glance:

**Country**: United States

**Industry**: Investment Banking

**Products and Services:** Microsoft Azure

## Customer challenges

The client wanted to modernize their current big data platform at scale, and had the following technical challenges:

- All the data transformation/processing logic on Impala/Hive (1000+ scripts) needed conversion to Spark, in order to complete the modernization effort and switch to Azure.
- The client wanted to get same outputs feeds using the new platform, with no logic change and better or similar performance. Also, all dependent data applications/ consumers should be able to adopt the new unified data platform without any disruption.

## Partner Solution

Wipro developed an automated BOT as part of IntelliProc to migrate the data to Azure Data Lake, convert the HQL/Impala scripts to Spark SQL/API and deploy it on Databricks platform, in order to drive the end-to-end Databricks transformation journey at scale on Microsoft Azure.

The process involved creating the DDL scripts from Impala/Hive and establishing the Database and Tables in Databricks. The data was then transported from Impala to Cloud - HDFS to Databricks - ADLS. Lastly, scripts and UDFs were segregated based on complexity and spark custom code was provided to handle wrapper invocation (as-is) or custom function invocation.

## Customer Benefits

Wipro's solution has allowed the client to modernize their existing big data applications at scale & start leveraging the benefits of a unified data platform – Databricks on Azure. Here are some of the key benefits provided to the client as part of the overall solution.

Higher level of Automation and Data mapping validation coverage achieved ensuring data integrity and accuracy.

Automated test case generation, SQL scripting, email notification and test execution delivered effort reduction of more than 20-25% and faster time to market.

Impala script execution with timelines – SLA used as "Baseline" to measure the performance of the Spark/Databricks transformation.


Contact Us:

- satheesh.balan@wipro.com
- abhishek.dey@wipro.com


Learn More

https://wipro.com