




Customer Objective

- Leading French personal care company.
- Customer was receiving data from multiple sources – CSV Files and Rest API
- Customer wanted to automate the ETL and validation processes
- Customer wanted to validate the ETL process.
- Customer wants to load the data to Cassandra and Hive
- Customer wants to apply transformations on top of source data before loading to the Target.

Business challenges

- Data types differences between the source and target resulted in ETL errors or truncated data.
- Performance issues loading and extracting large volumes of data.
- Inability to apply transformations on the data.
- Cumbersome DataQuality checks and constraint validations.
- ETL issues result in bad insights to the business user.
- Manual testing is limited to a small data set and is very time consuming.

Technologies

-  RestAPI and File Datasources
-   **HIVE**
- DataOps Suite 2020.1.0.0

Solution

- Customer implemented DataOpsSuite to help automate the ETL process for Data Migration to Cassandra and Validation.
- Was able to migrate large volumes of data from multiple sources leveraging parallel processing capability of Spark.
- Was able to apply pre built and custom transformations on the source data leveraging the components of DataOps Suite.
- Was able to manage different dataflows and pipelines for ETL and Validation processes and reuse the same.
- Customer was able to compare millions of records and achieve 100% test coverage.
- Define and maintain Data Quality and Referential Integrity rules.
- Monitoring the production ETL runs on a daily basis.
- Automated the execution of dataflows and pipelines in the Dev, Test and Prod environments (Continuous Integration)

Benefits Realized

Reduction in migration testing time by

45-60%

Reduction in data quality testing time by

30-40%

Reduction in migration time by

25-35%

Overall reduction of TCO by

30%

Faster ETL Validation with reusability & scheduling capabilities.

Ease of Data Transformation with No Code required.

Automated testing for **100%** of the migrated data.