

International Retail Analytics Modernization

Challenge An international retail and fast food company needed help migrating its on-premises Greenplum cluster to AWS, and its data science environment to Databricks. The company also needed help translating existing use cases to work more efficiently with Spark, converting Greenplum queries into SparkSQL DSL, operationalizing the Spark cluster, and training its data science team on Spark.

Solution Cloudwick conducted a cost analysis comparing the Greenplum solution with a new AWS environment and then provided the professional services to conduct the migration, translated the business cases to work with Spark, operationalized the Spark Cluster and trained the company's team on how to use Spark on Databricks.

Benefits

- The company saves 50% over the Greenplum environment and enjoys flexibility, agility and scalability.
- Data science is moved into the cloud in accordance with the company's overarching goal.
- It can leverage Spark for advanced data science applications.
- With Tableau, the company has easy access to powerful drag-and-drop dashboards and queries.

Global Goals

An international retail and fast food company operating more than 1,200 corporate and 270 franchise outlets in 16 countries with revenue over a billion dollars needed to migrate its on-premises data center to Amazon Web Services (AWS) and its data science environment to Databricks. This objective was in accordance with the company's overall strategy of migrating its on-premises data center to the cloud in order to increase agility, scalability, flexibility and cost savings as the company was reaching further around the globe. As with most enterprises, the company lacked the specific technology expertise in-house, so it turned to Cloudwick, the data and analytics modernization experts, for help.

Before starting the migration, Cloudwick conducted a cost analysis comparing the Greenplum cluster with a new hybrid AWS/Databricks environment. After the company chose to move forward with AWS/Databricks, Cloudwick's recommendation, Cloudwick provided professional services to conduct the migration of Greenplum SQL server queries to Databricks Spark notebooks.

S3 was used as primary data store with Spark leveraging EBS backed for cluster operations (storage and processing); post processed results were also stored in S3 for reliability. The company also chose Cloudwick's recommendations of AWS Direct Connect, AWS Hardware VPN and Talend for the bulk import from the data warehouse to ensure security and to reduce risk.

Security and Analytics

AWS EC2 is used for the Spark cluster and AWS Identity and Access Management (IAM) enables the organization to **manage users and user permissions** in the EC2 environment. Cloudwick then worked with the data science team to **translate the existing business use cases to work more efficiently** with Spark.

The first two weeks also consisted of Cloudwick optimizing the existing analytics to **work more efficiently** with Spark, translating the Greenplum queries into SparkSQL DSL, **operationalizing the Spark cluster**, and **training the data science team** to use Apache Spark on the Databricks platform. Finally, Cloudwick has implemented Tableau so the company can create its own dashboards and analyze data with **drag-and-drop simplicity**.