CASE STUDY

Tech Mahindra

Modernizing Data and Analytics Platform on Cloud

Overview

The customer is a US multinational package delivery supply chain management, and a global logistics company. They wanted to modernize on-premises data platform to highly scalable, reliable, cost effective, and maintenance free cloud solution using Google Cloud Platform.

Client Background and Challenge

The customer employs approximately 444,000 employees, 362,000 in the U.S. and 82,000 internationally. In terms of revenue, it is the largest company in the world.

The challenges faced by the customer were:

- Keeping the data in multiple locations was expensive and was using lot of space.
- Their existing architecture was difficult to scale up for increasing number of users and data sources.
- Moving from HDFS to cloud storage.
- Consolidating the data from different sources and locations into one place was a big challenge as it required massive development effort and infrastructure scale up. As the data grows with time, infrastructure scaling will be a big issue.
- Showcasing and viewing of data in forms of reports and other means was difficult. The customer wanted to move to Google Cloud BigQuery tables to show the data in a synchronized manner.

Our Approach and Solution

- Understanding the existing sources of mainframe, log files from Linux, and Oracle DWH.
- Analyze the scope and complexity of data sources.
- Around 150 functional areas which were sending files to HDFS and Datameer have been pushed into BQ tables by the shell scripts that uses BQ commands. Identifying bad records in log files and cleansing them before pushing into GCP.
- Automating the process of fetching data from tables into HDFS using Sqoop tool and then migrating the data into GCP BigQuery tables is done through Shell scripting.
- GCP Cloud monitoring to analyze load metrics and logging information.
- GCP ML will provision predictive analytics in the future. Vertex AI will take predictive analytics to the next level.

Technologies:

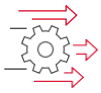
- SQL Server 2014, Oracle, Hadoop, Linux, Mainframe Cobol
- Datameer, Google Cloud Storage and BigQuery

Volumetric

- Target Tables: 1400
- Data volume migrated : 150TB

Business and Community Impact

Combined multiple processes/jobs into one single platform for better manageability as compared to various flows in legacy systems.



Google Cloud helps in optimizing space and loading constraints.

Automation of migration process saved time and effort of the existing migration into HDFS.

Users to view or access data from BigQuery tables in Google Cloud instead of viewing data from individual files.

Process improvement and migration to BigQuery helped reduced query execution time reduced from 12 hours to 5 mins.

To know more, reach us at **vbuoilandgas@techmahindra.com**