

# Implementation of a Modern Data Platform for a Leading Transportation and Logistics Provider



## Overview

The customer is a US based leading provider of transportation and logistics services in the world. They wanted to move from on-premises conventional data warehouse to data Lake on Google Cloud Platform (GCP). This helped address all existing pain areas including performance bottlenecks, scalability, and maintenance.

## Client Background and Challenge

The customer is an American freight transportation company that provides less-than-truckload and truck brokerage services in multiple countries. It employs around 100000 employees.

The challenges faced by the customer were:



The existing on-premises data warehouse became a challenge to the customer with respect to timely report generation, collating various sources of data, and scaling the infrastructure.



The existing data warehouse in Netezza had performance issues and was not consistent. Increasing number of users and data sources required frequent upgrade of infrastructure, which was difficult, expensive, and time consuming. Along with that, maintenance was high and so were the costs. The number of connections made to Netezza could not be scaled up to the desired level.



The customer wanted to move out of Netezza to a high performance, low maintenance, and low-cost solution, which is why they narrowed down to GCP Big Query warehouse.

## Our Approach and Solution

- ▶ Pre-Phase: Tech Mahindra set up the GCP and Informatica environment to test the migration approach.
- ▶ The solution is being co-developed in phased manner through Customer – Tech Mahindra partnership.
- ▶ Phase 1: Migrated the schema to BigQuery and cloud storage. Data model migration uses lift and shift approach, where we used Informatica to load data from Netezza to cloud storage and BigQuery. Developed automation tools to create the Informatica code and do data validation between Netezza and BigQuery.
- ▶ Phase 2: Migration of near real time workloads to BigQuery environment. Informatica PWX CDC was used for near real time streaming.
- ▶ Phase 3: Involved building of batch data integration pipelines to integrate/transform data from the source to cloud data lake and then to the GCP data warehouse. The existing Informatica code will be reused and repointed, to populate the target in BigQuery. This phase is in progress.






### Technologies:

- ▶ Data lake/warehouse: Netezza – BigQuery
- ▶ Near real time CDC: Informatica PWX CDC
- ▶ ETL tool: Informatica PowerCenter
- ▶ Reporting: MicroStrategy - Looker + Data Studio from GCP for quick reporting validation between Netezza and BigQuery.

### Volumetric

- ▶ Target Tables: 2100
- ▶ Intraday data volume: Range from 10000 - 50 million rows / table

## Business and Community Impact

-  Move out of Netezza and be business compliant.
-  Use of inbuilt GCP reporting tools.
-  Improved performance with lesser cost.
-  Utilize the advantages of Looker, a new reporting tool by Google.
-  Scalable data warehouse on BQ and data consolidation through cloud data lake.

To know more, reach us at [vbuoilandgas@techmahindra.com](mailto:vbuoilandgas@techmahindra.com)



**Tech  
Mahindra**