

Enabling Advanced Analytics Data Platform for Semiconductor OSAT



Overview

The customer is a US based outsourced semiconductor assembling and testing (OSAT) company and wanted to modernize data analytics to speed up time to insights, accelerate innovation, and promote data sharing and collaboration. To achieve the same, they wanted to implement a modern cloud ecosystem that enables them to drive analytics with ease and quality.

Client Background and Challenge

The customer is a semiconductor product packaging and test services provider. The company has been headquartered in Tempe, Arizona, it has approximately 29,300 employees worldwide and a reported \$4.19 billion in sales.

The challenges faced by the customer were:



The customer wanted to modernize their legacy big data storage to be more scalable along with a cost-effective data warehouse that enables business agility



The legacy reporting solution had limited features that could generate canned excel reports without "on the fly" reporting. This was a huge challenge while scaling the existing reporting solution.



Enable advanced analytics (predictive, prescriptive).










It was difficult for the engineering teams to trace back any defect from historical data required to do a root cause analysis

Our Approach and Solution

- ▶ Migrated over 10 TB of data from multiple data sources like Oracle, MySQL, log files from equipment's to cloud storage and then to BigQuery. BigQuery offers real time analytics and helped OSAT to analyze latest business data as soon as it is compiled.
- ▶ Implemented an extraction load and transformation(ELT) approach to load data to GCP BigQuery. To schedule the analytics workflow, Airflow was used.
- ▶ To orchestrate ingest data from on-premise source to target data store in BigQuery, Apache NiFi was used.
- ▶ The data was loaded to the staging area, post transformation it was loaded to processed area from which reporting tables were built.
- ▶ TIBCO Spotfire was connected to BigQuery which helped in generating reports..



Business and Community Impact

-  Ability to scale by leveraging GCP BigQuery scalable architecture.
-  GCP solution is serverless, no-ops and no hardware provisioning is required.
-  Developed the capability to enable "on the Fly" and self-service reporting.
-  Provision to integrate with advanced analytics tools using AI/ML.
-  Significant reduction in development effort to create a new report from few weeks to few days.
-  Effort to enhance any existing report reduced from few days to few hours.
-  Unstructured data analytics was enabled since GCP BigQuery APIs allow python integration for structuring data.

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



**Tech
Mahindra**