Case Study

Global Water & Waste Water Treatment Major builds Enterprise Information Catalog on Google Cloud Platform with LTI's Mosaic Catalog

Business benefits at a glance

This project is unique as it involved several firsts for LTI. It was the first instance of Mosaic Catalog being installed and configured on the Google Cloud Platform. Similarly, for the first time Mosaic Catalog scanned metadata from Google BigQuery Data Lake on GCP and created a "Dynamic Query Cost Calculator" as an aid for controlling the recurring cost of usage on GCP, which was much appreciated by the client. Additional business benefits involved:

100+
coverage of metadata scan on top of BigQuery Data lake

Completed deployment in just **four weeks**

95%
accuracy in prediction of cost of running queries on GCP

Centralized business glossary **available to everyone**

This story is about a global leader in water and wastewater treatment, providing a complete range of services required to design, build, maintain and upgrade water and wastewater treatment facilities and systems for industrial clients and public authorities across the globe.

As part of its innovation agenda, the client had been an early adopter of Industrial Internet of Things (IIoT) and deployed multiple sensors in their machineries and equipment. In order to tame the huge volume and velocity of data generated by these IIoT-enabled devices, they had decided to create a Big Data Lake on Google Cloud Platform (GCP), using Google's BigQuery technology. Although implemented successfully, it was feeling challenged in extracting insights from the ingested data. They were also sceptical about the subscription cost they would incur if any runway query was executed on the BigQuery Data Lake, as Google would charge them based on the compute requirement to execute such queries.

The company's CIO was looking for a solution that could:

- Offer a metadata portal which would sit on top of their existing data infrastructure, that would scan metadata from both on-premise databases, as well as the BigQuery Data lake, and curate them in a central repository
- Allow business users to discover, explore data sets and data elements, look at their profile and lineage information
- Provide Data-as-a-Service (DaaS) capabilities to ingest and consume data on-demand using a governed workflow.

This is where LTI stepped in with its Mosaic Catalog and own IP on the Google Cloud Platform, to offer the following solutions:

Centralized Business Glossary – created and managed business definition glossary, which in turn enabled indexing/ tagging data as per business definitions.

Discovery & Searchability of data – enabled semantic & faceted search with lightning fast results covering scanned metadata from BigQuery Data Lake on GCP as well as databases on-premise.

Data Profiling & Lineage – offered intuitive interface to understand data definitions, attributes, content summary, structure, data lineage, relationships, dependencies, issues, inconsistency and anomalies for all data assets, both at the dataset level, as well as the individual data element level.

Data Virtualization – integrated with Opensource Presto to offer self-service federated queries joining datasets from disparate sources, without moving the data, that to without the help of IT Engineers.

Dynamic Query Cost Predictor – created a simple predictor of Google usage cost for executing any query based on the number of records the query is likely to fetch without querying the dataset. This was used as a control measure to govern the execution of runaway queries on BigQuery Data lake by the user community.

Stepping towards becoming a breakaway enterprise, the client successfully ensured that their innovation agenda, to derive value from their data, is boosted and ensured the anticipated high ROI on their investment to create a Big Data Lake on GCP is achieved.

Visit us at **https://www.lti-mosaic.com/index.html** to know more on our innovative Mosaic product suite