GCP Project to Explore BigQuery using Python

Business Overview

Google Cloud is a collection of physical assets, such as computers and hard disk drives, and virtual resources, such as virtual machines (VMs), housed in Google data centers worldwide. This resource distribution has various advantages, including redundancy in a failure and decreased latency by putting resources closer to customers. This release also presents some guidelines for combining resources.

GCP offers a web-based graphical user interface for managing Google Cloud projects and resources. If a user prefers to work at the command line, the G-Cloud command-line tool can handle most Google Cloud activities.

This is the second project in the GCP Roadmap project series, the first project utilizes services such as PubSub, Compute Engine, and Cloud storage. In this project, we will explore GCP BigQuery:

Tech Stack

➔ Language: Python3
➔ Services: Cloud Storage, BigQuery, G-Cloud SDK

Cloud Storage

Cloud Storage is a service that allows users to store their data on the Google Cloud. An object is an immutable piece of data that consists of a file in any format. Objects can be stored in containers known as buckets. All buckets are related to a project, and the user may organize their projects into organizations. After starting a project, users may create Cloud Storage buckets, upload things to the buckets, and get objects. Users can also give rights to make data accessible to certain domains or for specific use cases such as establishing a website.

BigQuery

Google Bigquery is a Cloud Datawarehouse powered by Google, which is Serverless, highly scalable, and cost-effectively designed for making data driven business decisions quickly. It offers both the batch and streaming insertion capabilities and is integrated with Tensorflow as well to perform machine learning using SQL like dialects.

.

Key Takeaways

- Introduction to Google Cloud BigQuery
- Understanding Managed Tables
- Understanding ExternalTables
- Introduction to Partitioning in BigQuery
- Introduction to Clustering in BigQuery
- Understanding different supported File Formats
- Using BQ CLI commands
- Creating External BigQuery Table using GCS Bucket
- Using Client API to load BigQuery table
- Defining schema programmatically using Python