Google BigQuery

© Elephant Scale

June 20, 2021

Overview

- BigQuery (BQ) is a fully-managed, serverless data warehouse that enables scalable analysis over petabytes of data. It is a Platform as a Service that supports querying using ANSI SQL. It also has built-in machine learning capabilities. ## Benefits

- After taking the course, participants will be able to

    - Point out BQ use cases
    - Implement BQ project on Google Cloud
    - Be fluent in best practices for BQ and integration

Duration:

- 2 days

Audience:

- Data scientists
- Software developers
- Project managers

Prerequisites

- Knowledge of SQL
- Familiarity with GCP

Lab environment

- Ideally, students come with their own GCP account
- Qwiklabs access for 30 days will be provided for every student

Course Outline:

- Deep dive partition strategy, performance tuning, query visualizer
- Resource Monitoring, Telemetry – Information Schema jobs table, stackdriver, etc, Looker dashboards
    - TBD on Telemetry
- Advanced features (Nested tables) Star schemas
- DDL and DML including External and Internal tables
- Priority workloads – Reservations API
- Storage API and Spark Integration
    - Cross GCP projects (or VPC peering)
- Google project provisioning
- IAM, Roles, Permissions, ServiceAccounts, etc
- Data Transfer API, Scheduling Queries and Airflow integration
- VPC and Data Exfiltration prevention
    - TBD

- BigQuery Roadmap – BQ Omni on Azure, DataPlex, etc