Course Outline

Data to Insights with Google Cloud Platform

Course Summary

Description

Explore ways to derive insights from data at scale using BigQuery, Google Cloud's serverless, highly scalable, and cost-effective cloud data warehouse. This course uses lectures, demos, and hands-on labs to teach you the fundamentals of BigQuery, including how to create a data transformation pipeline, build a BI dashboard, ingest new datasets, and design schemas at scale.

Objectives

At the end of the course, students will be able to:

- Derive insights from data using the analysis and visualization tools on Google.
- Load, clean, and transform data at scale with Dataprep.
- Explore and visualize data using Google Data Studio.
- Troubleshoot, optimize, and write high performance queries.
- Practice with pre-built ML APIs for image and text understanding.
- Train classification and forecasting ML models using SQL with BigQuery ML.

Topics

- Introduction to Data on Google Cloud
- Analyzing Large Datasets with BigQuery
- Exploring your Public Dataset with SQL
- Cleaning and Transforming your Data with Dataprep
- Visualizing Insights and Creating Scheduled Queries
- Storing and Ingesting New Datasets
- Enriching your Data Warehouse with JOINs
- Advanced Features and Partitioning your queries and Tables for Advanced Insights
- Designing Schemas that Scale: Arrays and Structs in BigQuery
- Optimizing Queries for Performance
- Controlling Access with Data Security
- Predicting Visitor Return Purchase with BigQuery ML
- Deriving Insights from Unstructured Data Using Machine Learning

Audience

This course is designed for Data Analysts, Business Analysts, Business Intelligence professionals, and Cloud Data Engineers who will be partnering with Data Analysts to build scalable data solutions on Google Cloud Platform.

Prerequisites

Before taking this course, student should have basic proficiency with ANSI SQL.

Duration

Three days

Data to Insights with Google Cloud Platform

Course Outline

I. Introduction to Data on Google Cloud
- A. Analytics Challenges Faced by Data Analysts
- B. Big Data On-premise Versus on the Cloud
- C. Real-world Use Cases of Companies Transformed Through Analytics on the Cloud
- D. Google Cloud Project Basics

II. Analyzing Large Datasets with BigQuery
- A. Data Analyst Tasks, Challenges, and Google Cloud Data Tools
- B. Fundamental BigQuery Features
- C. Google Cloud Tools for Analysts, Data Scientists, and Data Engineers

III. Exploring your Public Dataset with SQL
- A. Common Data Exploration Techniques
- B. Use SQL to Query Public Datasets

IV. Cleaning and Transforming your Data with Dataprep
- A. 5 Principles of Dataset Integrity
- B. Dataset Shape and Skew
- C. Clean and Transform Data using SQL
- D. Introducing Dataprep by Trifecta

V. Visualizing Insights and Creating Scheduled Queries
- A. Data Visualization Principles
- B. Common Data Visualization Pitfalls
- C. Google Data Studio

VI. Storing and Ingesting New Datasets
- A. Permanent Versus Temporary Data Tables
- B. Ingesting New Datasets

VII. Enriching your Data Warehouse with JOINs
- A. Merge Historical Data Tables with UNION
- B. Introduce Table Wildcards for Easy Merges
- C. Review Data Schemas: Linking Data Across Multiple Tables
- D. JOIN Examples and Pitfalls

VIII. Advanced Features and Partitioning your queries and Tables for Advanced Insights
- A. Advanced Functions (statistical, Analytic, User-defined)
- B. Date-Partitioned Tables

IX. Designing Schemas that Scale: Arrays and Structs in BigQuery
- A. BigQuery Versus Traditional Relational Data Architecture
- B. ARRAY and STRUCT Syntax
- C. BigQuery Architecture

X. Optimizing Queries for Performance
- A. BigQuery Performance Pitfalls
- B. Prevent Data Hotspots
- C. Diagnose Performance Issues with the Query Explanation Map

XI. Controlling Access with Data Security
- A. Hashing Columns
- B. Authorized Views
- C. IAM and BigQuery Dataset Roles
- D. Access Pitfalls

XII. Predicting Visitor Return Purchase with BigQuery ML
- A. Machine Learning on Structured Data
- B. Scenario: Predicting Customer Lifetime Value
- C. Choosing the Right Model Type
- D. Creating ML models with SQL

XIII. Deriving Insights from Unstructured Data Using Machine Learning
- A. ML Drives Business Value
- B. How does ML on unstructured data work?
- C. Choosing the Right ML Approach
- D. Pre-build AI Building Blocks
- E. Customizing Pre-build Models with AutoML
- F. Building a Custom Model