



CPB200 - Version: 1
01 May 2021

From Data to Insights with Google Cloud Platform



From Data to Insights with Google Cloud Platform

CPB200 - Version: 1

 **2 days Course**

Description:

Want to know how to query and process petabytes of data in seconds? Curious about data analysis that scales automatically as your data grows? Welcome to the Data Insights course!

This two-day instructor-led class teaches course participants how to derive insights through data analysis and visualization using the Google Cloud Platform. The course features interactive scenarios and hands-on labs where participants explore, mine, load, visualize, and extract insights from diverse Google BigQuery datasets. The course covers data loading, querying, schema modeling, optimizing performance, query pricing, and data visualization.

Intended audience:

Data Analysts, Business Analysts, Business Intelligence professionals
Cloud Data Engineers who will be partnering with Data Analysts to build scalable data solutions on Google Cloud Platform

Prerequisites:

Basic proficiency with ANSI SQL

Objectives:

Derive insights from data using the analysis and visualization tools on Google Cloud Platform
Interactively query datasets using Google BigQuery
Load, clean, and transform data at scale

Visualize data using Google Data Studio and other third-party platforms
Distinguish between exploratory and explanatory analytics and when to use each approach
Explore new datasets and uncover hidden insights quickly and effectively
Optimizing data models and queries for price and performance

Topics:

Module 1: Introduction to Data on the Google Cloud Platform

- Highlight Analytics Challenges Faced by Data Analysts
- Compare Big Data On-Premises vs on the Cloud
- Learn from Real-World Use Cases of Companies Transformed through Analytics on the Cloud
- Navigate Google Cloud Platform Project Basics
- Lab: Getting started with Google Cloud Platform

Module 2: Big Data Tools Overview

- Walkthrough Data Analyst Tasks, Challenges, and Introduce Google Cloud Platform Data Tools
- Demo: Analyze 10 Billion Records with Google BigQuery
- Explore 9 Fundamental Google BigQuery Features
- Compare GCP Tools for Analysts, Data Scientists, and Data Engineers
- Lab: Exploring Datasets with Google BigQuery

Module 3: Exploring your Data with SQL

- Compare Common Data Exploration Techniques
- Learn How to Code High Quality Standard SQL
- Explore Google BigQuery Public Datasets
- Visualization Preview: Google Data Studio
- Lab: Troubleshoot Common SQL Errors

Module 4: Google BigQuery Pricing



- Walkthrough of a BigQuery Job
- Calculate BigQuery Pricing: Storage, Querying, and Streaming Costs
- Optimize Queries for Cost
- Lab: Calculate Google BigQuery Pricing

Module 5: Cleaning and Transforming your Data

- Examine the 5 Principles of Dataset Integrity
- Characterize Dataset Shape and Skew
- Clean and Transform Data using SQL
- Clean and Transform Data using a new UI: Introducing Cloud Dataprep
- Lab: Explore and Shape Data with Cloud Dataprep

Module 6: Storing and Exporting Data

- Compare Permanent vs Temporary Tables
- Save and Export Query Results
- Performance Preview: Query Cache
- Lab: Creating new Permanent Tables

Module 7: Ingesting New Datasets into Google BigQuery

- Query from External Data Sources
- Avoid Data Ingesting Pitfalls
- Ingest New Data into Permanent Tables
- Discuss Streaming Inserts
- Lab: Ingesting and Querying New Datasets

Module 8: Data Visualization

- Overview of Data Visualization Principles
- Exploratory vs Explanatory Analysis Approaches
- Demo: Google Data Studio UI
- Connect Google Data Studio to Google BigQuery
- Lab: Exploring a Dataset in Google Data Studio

Module 9: Joining and Merging Datasets

- Merge Historical Data Tables with UNION
- Introduce Table Wildcards for Easy Merges
- Review Data Schemas: Linking Data Across Multiple Tables
- Walkthrough JOIN Examples and Pitfalls
- Lab: Join and Union Data from Multiple Tables

Module 10: Advanced Functions and Clauses

- Review SQL Case Statements
- Introduce Analytical Window Functions
- Safeguard Data with One-Way Field Encryption
- Discuss Effective Sub-query and CTE design
- Compare SQL and Javascript UDFs
- Lab: Deriving Insights with Advanced SQL Functions

Module 11: Schema Design and Nested Data Structures

- Compare Google BigQuery vs Traditional RDBMS Data Architecture
- Normalization vs Denormalization: Performance Tradeoffs
- Schema Review: The Good, The Bad, and The Ugly
- Arrays and Nested Data in Google BigQuery
- Lab: Querying Nested and Repeated Data

Module 12: More Visualization with Google Data Studio

- Create Case Statements and Calculated Fields
- Avoid Performance Pitfalls with Cache considerations
- Share Dashboards and Discuss Data Access considerations

Module 13: Optimizing for Performance

- Avoid Google BigQuery Performance Pitfalls



- Prevent Hotspots in your Data
- Diagnose Performance Issues with the Query Explanation map
- Lab: Optimizing and Troubleshooting Query Performance

Module 14: Advanced Insights

- Introducing Cloud Datalab
- Cloud Datalab Notebooks and Cells
- Benefits of Cloud Datalab

Module 15: Data Access

- Compare IAM and BigQuery Dataset Roles
- Avoid Access Pitfalls
- Review Members, Roles, Organizations, Account Administration, and Service Accounts