

From Data to Insights with Google Cloud Platform

Course Code: GCPDI

Duration: 3 days

Instructor-led Training (ILT) | Virtual Instructor-led Training (VILT)

OVERVIEW

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.

SKILLS COVERED

- Derive insights from data using the analysis and visualization tools on Google Cloud Platform
- Load, clean, and transform data at scale with Google Cloud 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 BQML

WHO SHOULD ATTEND

This class is intended for the following participants: Data analysts, business analysts, business intelligence professionals.

PREREQUISITES

Basic proficiency with ANSI SQL

MODULES

Module 1: Introduction to Data on the Google Cloud Platform

- Highlight Analytics Challenges Faced by Data Analysts.
- Compare Big Data On-Premise 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: Data Access

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

Module 15: Notebooks in the Cloud

- Cloud Datalab.
- Compute Engine and Cloud Storage.

Lab: Rent-a-VM to process earthquakes data.

- Data Analysis with BigQuery.

Module 16: How Google does Machine Learning

- Introduction to Machine Learning for analysts.

- Practice with Pretrained ML APIs for image and text understanding.

Lab: Pretrained ML APIs.**Module 17: Applying Machine Learning to your Datasets (BQML)**

- Building Machine Learning datasets and analyzing features.
- Creating classification and forecasting models with BQML.

Lab: Predict Visitor Purchases with a Classification Model in BQML.**Lab: Predict Taxi Fare with a BigQuery ML Forecasting Model.**