CPB200: Google BigQuery for Data Analysts

Course Description

This 3 day instructor-led class introduces participants to Google BigQuery. Through a combination of instructor-led presentations, demonstrations, and hands-on labs, students learn how to store, transform, analyze, and visualize data using Google BigQuery.

Audience

This class is intended for data analysts and data scientists responsible for: analyzing and visualizing big data, implementing cloud-based big data solutions, deploying or migrating big data applications to the public cloud, implementing and maintaining large-scale data storage environments, and transforming/processing big data.

Prerequisites

Before attending this course, participants should have:
- Attended CP100A - Google Cloud Platform Fundamentals OR CPB100 - Google Cloud Platform Big Data & Machine Learning Fundamentals (or equivalent experience)
- Experience using a SQL-like query language to analyze data

Duration

3 days (24 hours)

Delivery Method

Instructor-led, Instructor-led online

Objectives

At the end of this course, participants will be able to:
- Understand the purpose of and use cases for Google BigQuery
- Describe ways in which customers have used Google BigQuery to improve their businesses
- Understand the architecture of BigQuery and how queries are processed
- Interact with BigQuery using the web UI and command-line interface
- Identify the purpose and structure of BigQuery schemas and data types
- Understand the purpose of and advantages of BigQuery destinations tables and caching
- Use BigQuery jobs
- Transform and load data into BigQuery
- Export data from BigQuery

- Store query results in a destination table
- Create a federated query
- Export log data to BigQuery and query it
- Understand the BigQuery pricing structure and evaluate mechanisms for controlling query and storage costs
- Identify best practices for optimizing query performance
- Troubleshoot common errors in BigQuery
- Use various BigQuery functions
- Use external tools such as spreadsheets to interact with BigQuery
- Visualize BigQuery data
- Use access controls to restrict access to BigQuery data
- Query Google Analytics Premium data exported to BigQuery

Modules

Module 1: Introducing Google BigQuery
Learning Objectives

- Understand the purpose of and use cases for Google BigQuery
- Describe ways in which customers have used Google BigQuery to improve their businesses

Lab: Sign Up for the Free Trial and Create a Project
Learning objectives
- Register for the GCP free trial
- Create a project using the Cloud Platform Console

Module 2: BigQuery Functional Overview
Learning Objectives
- Describe the components of a BigQuery project
- Identify how BigQuery stores data and list the advantages of the storage model
- Understand the architecture of BigQuery and how queries are processed
- Describe the methods of interacting with BigQuery

Lab: Explore BigQuery Interfaces
Learning objectives
- Explore features of the BigQuery web UI
- Learn how to use the bq shell
- Execute queries using the BigQuery CLI in Cloud Shell

Module 3: BigQuery Fundamentals
Learning Objectives
- Describe the purpose of denormalizing data

- Identify the purpose and structure of BigQuery schemas and data types
- Explain the types of actions available in BigQuery jobs
- Understand the purpose of and advantages of BigQuery destinations tables and caching

Lab: BigQuery Components and Jobs
Learning Objectives
- Explore how data is organized in BigQuery
- Learn about the two types of table schemas
- Learn about jobs, and how to cancel them
- Investigate caching and destination tables

Module 4: Ingesting, Transforming, and Storing Data
Learning Objectives
- Describe the methods for ingesting data, transforming data, and storing data using BigQuery
- Explain the function of BigQuery federated queries

Lab 4, Part I: Loading Data into BigQuery and Using Federated Queries
Learning Objectives
- Load a CSV file into a BigQuery table using the web UI
- Load a JSON file into a BigQuery table using the CLI
- Transform data and join tables using the web UI
- Store query results in a destination table
- Query a destination table using the web UI to confirm your data was transformed and loaded correctly
- Export query results from a destination table to Google Cloud Storage
- Create a federated query that queries data in Cloud Storage

Lab 4, Part II: Exporting App Engine Logs to BigQuery
Learning Objectives
- Set up Google Cloud Logging to export App Engine log data from the Guestbook application
- Use the BigQuery web UI to query the log data

Module 5: Pricing and Quotas
Learning Objectives
- Explain the advantages of the BigQuery pricing model
- Use the pricing calculator to calculate storage and query costs
- Identify the quotas that apply to BigQuery projects

Lab: BigQuery Pricing
Learning Objectives
- Evaluate the size of a query within BigQuery using the BigQuery web UI

- Use the Pricing Calculator and the total size of the query to estimate the query cost
- Examine how changing a query affects query cost

Module 6: Clauses and Functions
Learning Objectives
- Explain the differences between BigQuery SQL and ANSI SQL
- Identify the purpose of and use cases for user-defined functions
- Explain the purpose of various BigQuery functions

Lab: BigQuery Clauses and Functions
Learning Objectives
- Create and run a query using a wildcard function
- Create and run a query using a window function
- Create and run a query using a user-defined function

Module 7: Nested and Repeated Fields
Learning Objectives
- Identify the purpose and structure of BigQuery nested, repeated, and nested repeated fields
- Describe the use cases for nested, repeated, and nested repeated fields

Lab: Nested Fields
Learning Objectives
- Create a BigQuery table using nested data
- Run queries to explore the structure of the nested data

Lab: Repeated Fields
Learning Objectives
- Create a BigQuery table using repeated data
- Run queries to explore the structure of the repeated data

Lab: Nested Repeated Fields
Learning Objectives
- Create a BigQuery table using nested repeated data
- Run queries to explore the structure of the nested repeated data

Module 8: Query Performance
Learning Objectives
- Explain the impact of the following in query performance: JOIN and GROUP BY, table wildcards, and table decorators
- Identify various best practices for optimizing query performance

Lab: BigQuery Best Practices and Optimization Techniques

Learning Objectives
- Use denormalization to improve query performance
- Use subselects to improve the performance of queries with JOIN clauses
- Use destination tables to lower costs when running multiple, similar queries
- Use table decorators and table wildcards to improve query performance and to reduce costs

Module 9: Troubleshooting Errors
Learning Objectives
- Describe how to handle the most common BigQuery errors: request encoding errors, resource errors, and HTTP errors

Lab: Handling Errors
Learning Objectives
- Correct queries that produce syntax-related error messages
- Correct an error involving the order of a JOIN clause
- Correct an error involving an invalid table name
- Modify queries that exceed resource constraints

Module 10: Access Control
Learning Objectives
- Describe the purpose of access control lists in BigQuery
- List and explain the project and dataset roles available in BigQuery
- Apply views for row-level security

Lab: Access Control
Learning Objectives
- Manage access to datasets using project-level ACLs
- Manage access to datasets using dataset-level ACLs
- Set row-level permissions using views

Module 11: Exporting Data
Learning Objectives
- List the methods of exporting data from BigQuery and the data formats available
- Describe the process of creating a job to export data from BigQuery
- Explain the purpose of wildcard exports to partition export data

Lab: Exporting Data
Learning Objectives
- Export data from BigQuery using the web UI and CLI
- Export large tables using wildcard URIs

Module 12: Interfacing with External Tools

Learning Objectives
- ● Describe how to use external tools to interface with BigQuery, including: spreadsheets, ODBC and JDBC drivers, the BigQuery encrypted client, and R

Lab: Interfacing with External Tools
Learning Objectives
- ● Set up the BigQuery Reports add-on for Google Sheets
- ● Use the Reports add-on to query BigQuery data

Module 13: Working with Google Analytics Premium Data
Learning Objectives
- ● Describe the schema of the Google Analytics Premium and AdSense data exported to BigQuery

Lab: Working with Google Analytics Premium Data
Learning Objectives
- ● Build queries to analyze data from Google Analytics Premium

Module 14: Data Visualization
Learning Objectives
- ● Describe the options available for visualizing BigQuery data

Lab: Visualizing Data
Learning Objectives
- ● Use Google Cloud Datalab to visualize data