Score
—/10

GCP Fundamentals: Getting Started with BigQuery

30 minutes   Free   ★★★★★ Rate Lab

Overview

In this lab, you load a web server log into a BigQuery table. After loading the data, you query it using the BigQuery web user interface and the BigQuery CLI.

BigQuery helps you perform interactive analysis of petabyte-scale databases, and it enables near real time analysis of massive datasets. It offers a familiar SQL 2011 query language and functions.

Data stored in BigQuery is highly durable. Google stores your data in a replicated manner by default and at no additional charge for replicas. With BigQuery, you pay only for the resources you use. Data storage in BigQuery is inexpensive. Queries incur charges based on the amount of data they process: when you submit a query, you pay for the compute nodes only for the duration of that query. You don't have to pay to keep a compute cluster up and running.

Using BigQuery involves interacting with a number of Google Cloud Platform resources, including projects (covered elsewhere in this course), datasets, tables, and jobs. This lab introduces you to some of these resources, and this brief introduction summarizes their role in interacting with BigQuery.

Datasets: A dataset is a grouping mechanism that holds zero or more tables. A dataset is the lowest level unit of access control. Datasets are owned by GCP projects. Each dataset can be shared with individual users.

Tables: A table is a row-column structure that contains actual data. Each table has a schema that describes strongly typed columns of values. Each table belongs to a dataset.

Objectives

In this lab, you learn how to perform the following tasks:

- Load data from Cloud Storage into BigQuery.

- Perform a query on the data in BigQuery.

Task 1: Sign in to the Google Cloud Platform (GCP) Console

What you'll need

To complete this lab, you'll need:

- Access to a standard internet browser (Chrome browser recommended).

- Time. Note the lab's **Completion** time in Qwiklabs. This is an estimate of the time it should take to complete all steps. Plan your schedule so you have time to complete the lab. Once you start the lab, you will not be able to pause and return later (you begin at step 1 every time you start a lab).

- The lab's **Access** time is how long your lab resources will be available. If you finish your lab with access time still available, you will be able to explore the Google Cloud Platform or work on any section of the lab that was marked "if you have time". Once the Access time runs out, your lab will end and all resources will terminate.

- You **DO NOT** need a Google Cloud Platform account or project. An account, project and associated resources are provided to you as part of this lab.

- If you already have your own GCP account, make sure you do not use it for this lab.

- If your lab prompts you to log into the console, **use only the student account provided to you by the lab**. This prevents you from incurring charges for lab activities in your personal GCP account.

Start your lab

When you are ready, click **Start Lab**. You can track your lab's progress with the status bar at the top of your screen.

> **Important** What is happening during this time? Your lab is spinning up GCP resources for you behind the scenes, including an account, a project, resources within the project, and permission for you to control the resources needed to run the lab. This means that instead of spending time manually setting up a project and building resources from scratch as part of your lab, you can begin learning more quickly.

Find Your Lab's GCP Username and Password

To access the resources and console for this lab, locate the Connection Details panel in Qwiklabs. Here you will find the account ID and password for the account you will use to log in to the Google Cloud Platform:

CONNECTION DETAILS

[OPEN GOOGLE CONSOLE]

googlexxx3-student@qwiklabs.net

YzJeX87dk

If your lab provides other resource identifiers or connection related information, it will appear on this panel as well.

Make a note of whether your assigned region is closer to the United States or to Europe.

Task 2: Load data from Cloud Storage into BigQuery

1. In the Console, on the **Navigation menu** (☰) click **BigQuery** then click **Done**.

2. Create a new dataset within your project by selecting your project in the Resources section, then clicking on CREATE DATASET on the right.

3. In the **Create Dataset** dialog, for **Dataset ID**, type **logdata**.

4. For **Data location**, select the continent closest to the region, click **Create dataset**.

5. Create a new table in the **logdata** to store the data from the CSV file.

6. Click on **Create Table**. On the **Create Table** page, in the **Source** section:

- For **Create table from**, choose select **Google Cloud Storage**, and in the field, type `gs://cloud-training/gcpfci/access_log.csv`.
- Verify **File format** is set to **CSV**.

Note: When you have created a table previously, the Create from Previous Job option allows you to quickly use your settings to create similar tables.

7. In the **Destination** section:

- For **Dataset name**, leave **logdata** selected.
- For **Table name**, type **accesslog**.
- For **Table type**, **Native table** should be selected and unchangeable.

8. Under **Schema** section, for **Auto detect** check the **Schema and input Parameters**.

9. Accept the remaining default values and click **Create Table**.

BigQuery creates a load job to create the table and upload data into the table (this may take a few seconds).

10. (Optional) To track job progress, click **Job History**.

11. When the load job is complete, click **logdata** > **accesslog**.

12. On the **Table Details** page, click **Details** to view the table properties, and then click **Preview** to view the table data.

Each row in this table logs a hit on a web server. The first field, **string_field_0**, is the IP address of the client. The fourth through ninth fields log the day, month, year, hour, minute, and second at which the hit occurred. In this activity, you will learn about the daily pattern of load on this web server.

Click Check my progress to verify the objective.

○ Load data from Cloud Storage into BigQuery
   [Check my progress]

Task 3: Perform a query on the data using the BigQuery web UI

In this section of the lab, you use the BigQuery web UI to query the **accesslog** table you created previously.

1. In the **Query editor** window, type (or copy-and-paste) the following query:

```
select int64_field_6 as hour, count(*) as hitcount from logdata.accesslog group by hour order by hour
```

2. Because you told BigQuery to automatically discover the schema when you loaded the data, the hour of the day during which each web hit arrived is in a field called **int_field_6**.

Notice that the Query Validator tells you that the query syntax is valid (indicated by the green check mark) and indicates how much data the query will process. The amount of data processed allows you to determine the price of the query using the Cloud Platform Pricing Calculator.

3. Click **Run** and examine the results. At what time of day is the website busiest? When is it least busy?

Task 4: Perform a query on the data using the bq command

In this section of the lab, you use the bq command in Cloud Shell to query the **accesslog** table you created previously.

1. On the **Google Cloud Platform** menu, click **Activate Cloud Shell** [>_]. If a dialog box appears, click **Start Cloud Shell**.

2. At the Cloud Shell prompt, enter this command:

```
bq query "select int64_field_6 as hour, count(*) as requestcount from logdata.accesslog group by hour order by requestcount desc"
```

The first time you use the bq command, it caches your Google Cloud Platform credentials, and then asks you to choose your default project. Choose the project that Qwiklabs assigned you to. Its name will look like qwiklabs-gcp- followed by a hexadecimal number.

The bq command then performs the action requested on its command line. What URL offered by this web server was most popular? Which was least popular?

Congratulations!

In this lab, you loaded data stored in Cloud Storage into a table hosted by Google BigQuery. You then queried the data to discover patterns.

End your lab

When you have completed your lab, click **End Lab**. Qwiklabs removes the resources you've used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

- 1 star = Very dissatisfied
- 2 stars = Dissatisfied
- 3 stars = Neutral
- 4 stars = Satisfied
- 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support** tab.

Manual Last Updated: April 01, 2019
Lab Last Tested: April 01, 2019

©2019 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.

More resources

Read the Google Cloud Platform documentation on BigQuery.