www.cloudskillsboost.google /course_sessions/3751705/labs/383335

Exploring a BigQuery Public Dataset | Google Cloud Skills Boost

Qwiklabs : 9-11 minutes

Overview

Storing and querying massive datasets can be time consuming and expensive without the right hardware and infrastructure. BigQuery is an enterprise data warehouse that solves this problem by enabling super-fast SQL queries using the processing power of Google's infrastructure. Simply move your data into BigQuery and let us handle the hard work. You can control access to both the project and your data based on your business needs, such as giving others the ability to view or query your data.

You access BigQuery through the Cloud Console, the command-line tool, or by making calls to the BigQuery REST API using a variety of client libraries such as Java, .NET, or Python. There are also a variety of third-party tools that you can use to interact with BigQuery, such as visualizing the data or loading the data. In this lab, you access BigQuery using the web UI.

You can use the BigQuery web UI in the Cloud Console as a visual interface to complete tasks like running queries, loading data, and exporting data. This hands-on lab shows you how to query tables in a public dataset and how to load sample data into BigQuery through the Cloud Console.

Objectives

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

- Query a public dataset

- Create a custom table

- Load data into a table

- Query a table

Set up your environments

Lab setup

For each lab, you get a new Google Cloud project and set of resources for a fixed time at no cost.

1. Sign in to Qwiklabs using an **incognito window**.

2. Note the lab's access time (for example, `1:15:00`), and make sure you can finish within that time.
   There is no pause feature. You can restart if needed, but you have to start at the beginning.

3. When ready, click **Start lab**.

4. Note your lab credentials (**Username** and **Password**). You will use them to sign in to the Google Cloud Console.

5. Click **Open Google Console**.

6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.
   If you use other credentials, you'll receive errors or **incur charges**.

7. Accept the terms and skip the recovery resource page.

Open BigQuery Console

1. In the Google Cloud Console, select **Navigation menu** > **BigQuery**.

The **Welcome to BigQuery in the Cloud Console** message box opens. This message box provides a link to the quickstart guide and lists UI updates.

2. Click **Done**.

Task 1. Query a public dataset

In this task, you load a public dataset, USA Names, into BigQuery, then query the dataset to determine the most common names in the US between 1910 and 2013.

Load the USA Names dataset

1. In the Explorer pane, in **Type to search**, type **usa_names**, and then press **Enter**.

2. Click **Broaden search to all**.

3. In the Explorer pane, hover the pointer over `bigquery-public-data,` and then click ☆ **Star**.

4. In **Type to search** field, type `bigquery-public-data`. This display all the datasets in the project.

Note: If the new project `bigquery-public-data` doesn't appear to the Explorer pane, then click on **+ ADD DATA** > **Star a project by name** > **Star a project** (bigquery-public-data) and **STAR**.

5. Click **Expand node** for `bigquery-public-data`.

6. Scroll down the list of public datasets, click **More Results** until you find **usa_names**.

7. Click **usa_names** to expand the dataset.

8. Click **usa_1910_2013** to open that table.

Query the USA Names dataset

Query `bigquery-public-data.usa_names.usa_1910_2013` for the name and gender of the babies in this dataset, and then list the top 10 names in descending order.

1. Click **Query, and then** select **In new tab**.

2. Copy and paste the following query into the **Query editor** text area, replacing the existing query:

SELECT name, gender, SUM(number) AS total FROM `bigquery-public-data.usa_names.usa_1910_2013`
GROUP BY name, gender ORDER BY total DESC LIMIT 10

3. In the upper right of the window, view the query validator.
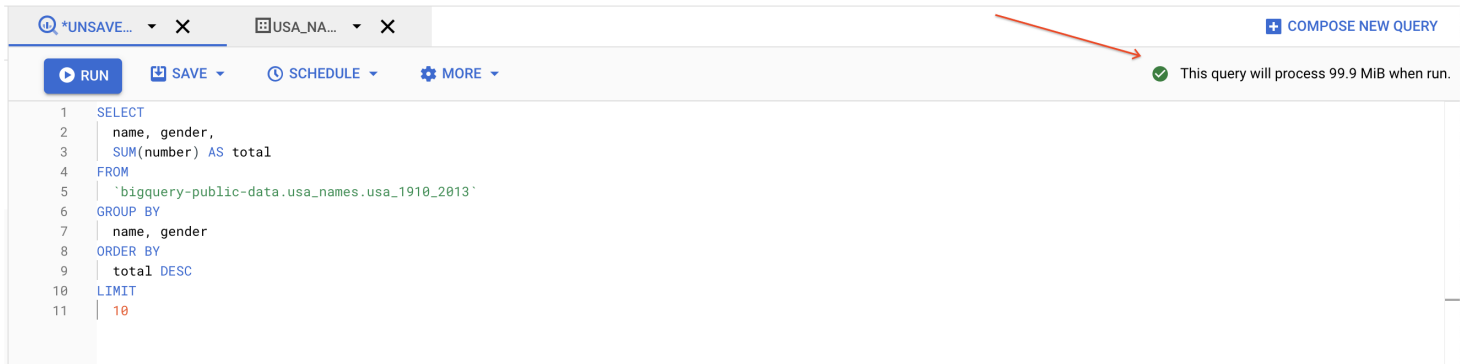
BigQuery displays a green check mark icon if the query is valid. If the query is invalid, a red exclamation point icon is displayed. When the query is valid, the validator also shows the amount of data the query processes when you run it. This helps to determine the cost of running the query.

4. Click **Run**.

The query results open below the Query editor. At the top of the Query results section, BigQuery displays the time elapsed and the data processed by the query. Below the time is the table that displays the query results. The header row contains the name of the column as specified in `GROUP BY` in the query.

Task 2. Create a custom table

In this task, you create a custom table, load data into it, and then run a query against the table.

Download the data to your local computer

The file you're downloading contains approximately 7 MB of data about popular baby names, and it is provided by the US Social Security Administration.

1. Download the baby names zip file to your local computer.

   Note: If this download link fails please copy the baby names zip file from the student resources on the left pane of the instruction guide.

2. Unzip the file onto your computer.

3. The zip file contains a `NationalReadMe.pdf` file that describes the dataset. Learn more about the dataset.

4. Open the file named `yob2014.txt` to see what the data looks like. The file is a comma-separated value (CSV) file with the following three columns: name, sex (`M` or `F`), and number of children with that name. The file has no header row.

5. Note the location of the `yob2014.txt` file so that you can find it later.

Task 3. Create a dataset

In this task, you create a dataset to hold your table, add data to your project, then make the data table you'll query against.

Datasets help you control access to tables and views in a project. This lab uses only one table, but you still need a dataset to hold the table.

1. Return to the the Cloud Console. In the **Explorer** pane, clear **bigquery-public-data** from the Type to search box.

Note: If you used the method to **Star a project by name**, then scroll back up to the top of the search results.

2. Click your Project ID (it will start with qwiklabs).
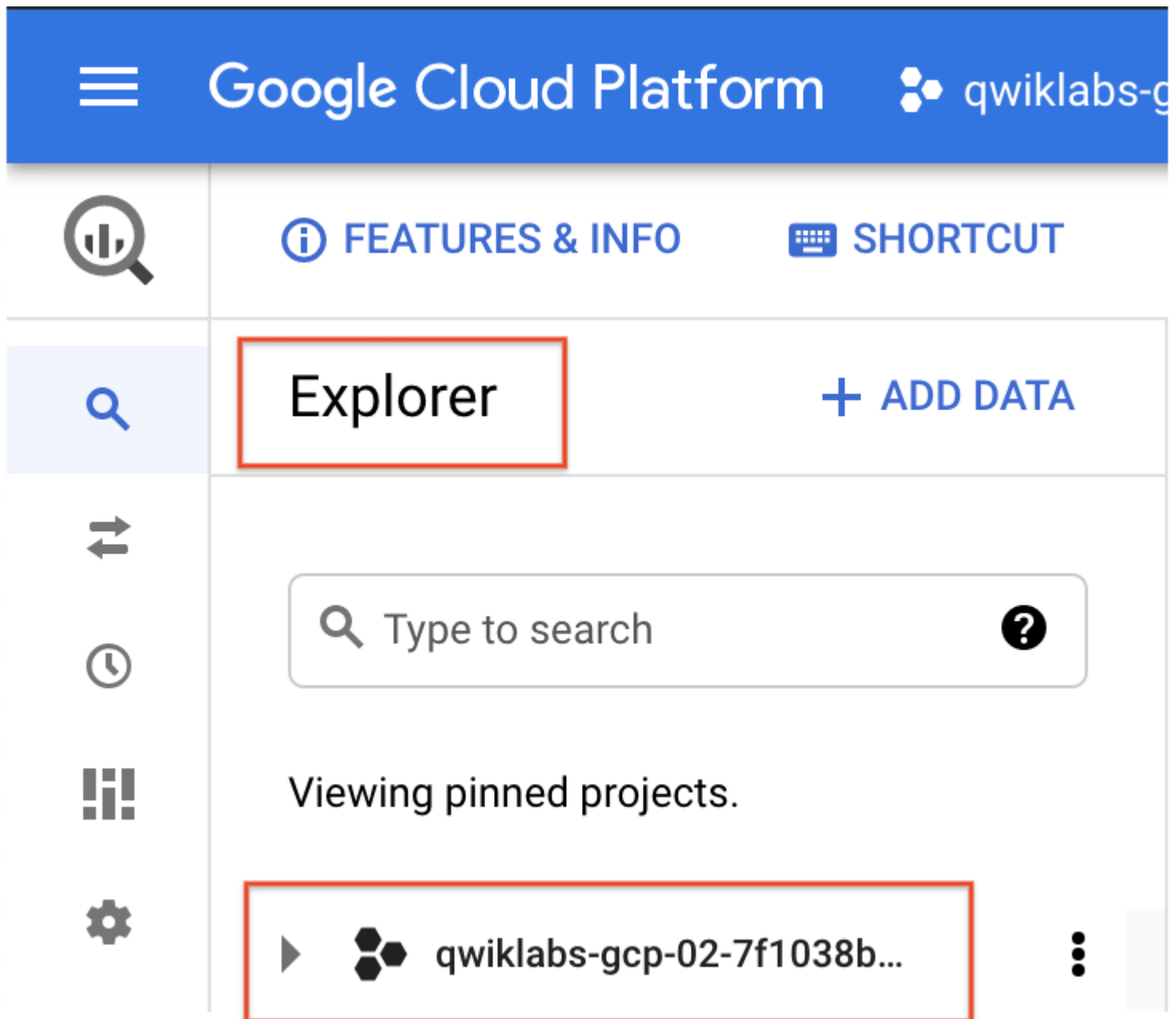
3. Click on the three dots next to your project ID and then click **Create dataset**.

4. On the **Create dataset** page:

   - For **Dataset ID**, enter babynames.
   - For **Data location**, choose **us (multiple regions in United States)**.
   - For **Default table expiration**, leave the default value.
   - For **Encryption**, leave the default value.

5. Click **Create dataset** at the bottom of the pane.

Task 4. Load the data into a new table

In this task, you load data into the table you made.

1. In the Explorer pane, expand your project ID dataset.

2. Click on the three dots next to **babynames**, and then click **Create table**.

Use the default values for all settings unless otherwise indicated.

3. On the **Create table** page:

   - For **Source**, choose **Upload** from the `Create table from:` dropdown menu.
   - For **Select file**, click **Browse**, navigate to the `yob2014.`txt file and click **Open**.
   - For **File format**, choose **CSV** from the dropdown menu.
   - For **Table name**, enter `names_2014`.
   - In the **Schema** section, click the **Edit as text** toggle and paste the following schema definition in the text box.

name:string,gender:string,count:integer

4. Click **Create table** (at the bottom of the window).

Note: If you see an import error, your data should still have been imported. To clear the error click **Close** , then click **Cancel** to exit the import dialog, and finally click **Yes, quit** in response to the warning that your changes will not be saved.

Preview the table

1. In the Explorer pane, select **babynames > names_2014**.
2. In the details pane, click the **Preview** tab.

Task 5. Query the table

Now that you've loaded data into your table, you can run queries against it. The process is identical to the previous example, except that this time, you're querying your table instead of a public table.

1. In the Query editor, click **Compose new query**.
2. Copy and paste the following query into the **Query editor**. This query retrieves the top 5 baby names for US males in 2014.

Note: Inside **"** is case sensitive, so be sure to align exactly the names of the dataset and the table you created.
SELECT name, count FROM `babynames.names_2014` WHERE gender = 'M' ORDER BY count DESC LIMIT 5

3. Click **Run**. The results are displayed below the query window.

Congratulations!

You queried a public dataset, then created a custom table, loaded data into it, and then ran a query against that table.

End your lab

When you have completed your lab, click **End Lab**. Google Cloud Skills Boost 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.