Using BigQuery in the Google Cloud Console

GSP406

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 Cloud Console.
Using BigQuery in the Cloud Console will give you 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.

What you'll do

In this lab you:

- Query a public dataset

- Create a custom table

- Load data into a table

- Query a table

Setup and requirements

Qwiklabs setup

Before you click the Start Lab button

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources will be made available to you.

This Qwiklabs hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access Google Cloud for the duration of the lab.

What you need

To complete this lab, you need:

- Access to a standard internet browser (Chrome browser recommended).
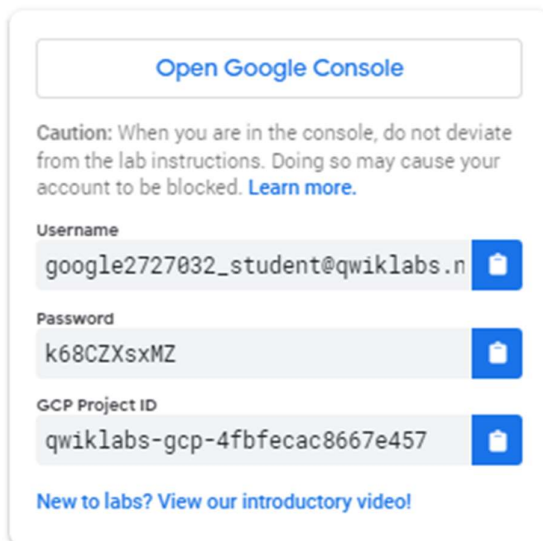- Time to complete the lab.
  Note: If you already have your own personal Google Cloud account or project, do not use it for this lab.

Note: If you are using a Pixelbook, open an Incognito window to run this lab.

Cloud Console

How to start your lab and sign in to the Google Cloud Console

1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is a panel populated with the temporary credentials that you must use for this lab.

2. Copy the username, and then click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Sign in** page.

Tip: Open the tabs in separate windows, side-by-side.

If you see the **Choose an account** page, click **Use Another**
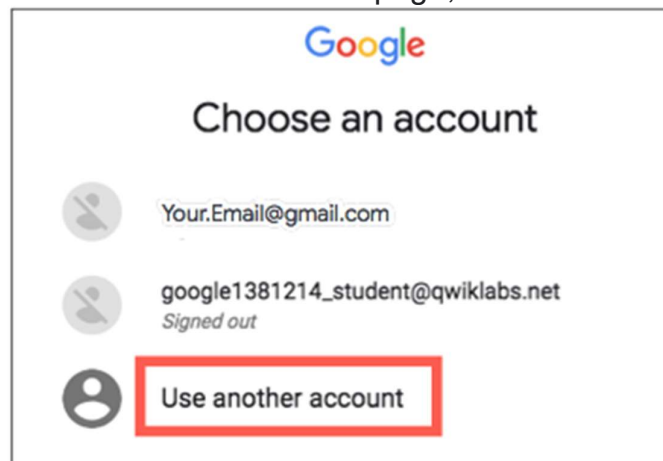
Account.

3. In the **Sign in** page, paste the username that you copied from the Connection Details panel. Then copy and paste the password.

   Important: You must use the credentials from the Connection Details panel. Do not use your Qwiklabs credentials. If you have your own Google Cloud account, do not use it for this lab (avoids incurring charges).

4. Click through the subsequent pages:

   - Accept the terms and conditions.
   - Do not add recovery options or two-factor authentication (because this is a temporary account).
   - Do not sign up for free trials.

After a few moments, the Cloud Console opens in this tab.

Note: You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-
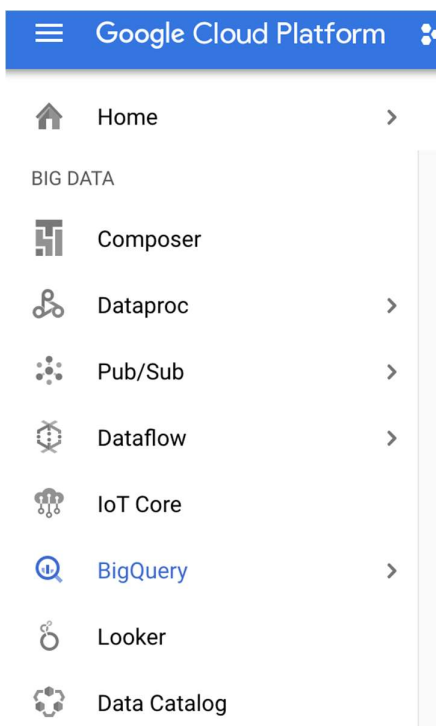
left.
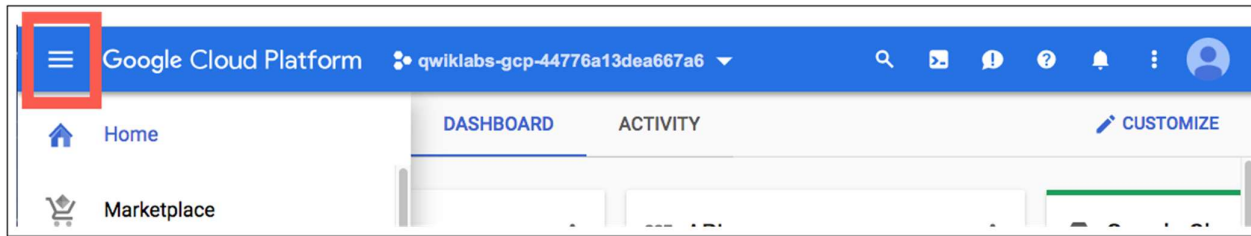
Open BigQuery

The BigQuery console provides an interface to query tables, including [public datasets](#) offered by BigQuery.
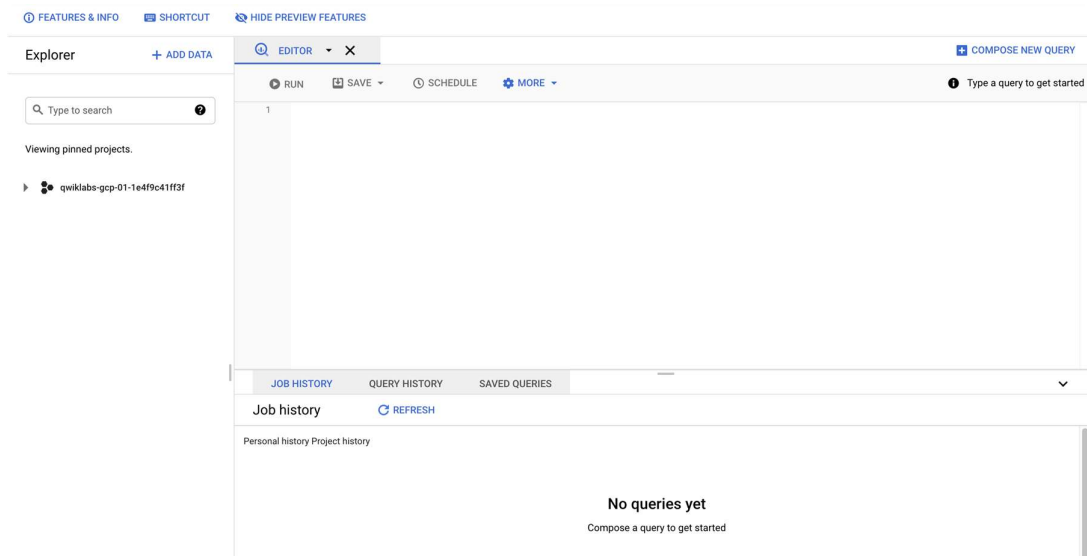
Open BigQuery Console

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 the release notes.
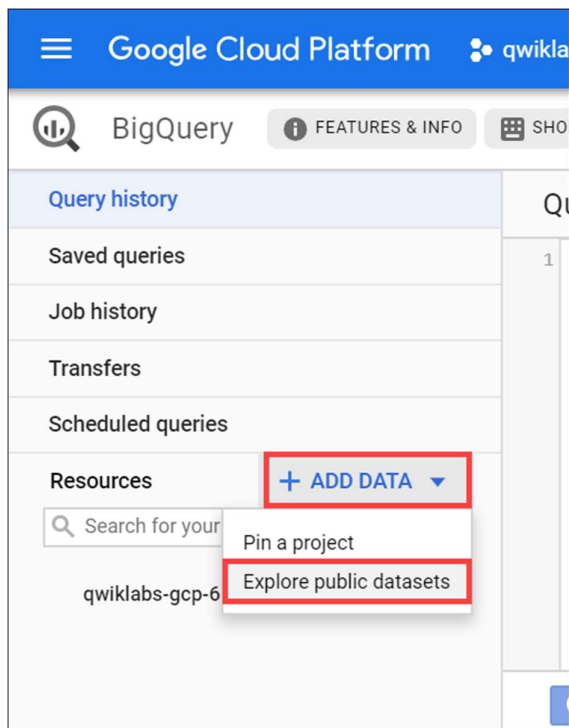
Click **Done**.

The BigQuery console opens.

Query a public dataset

In this section, 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.
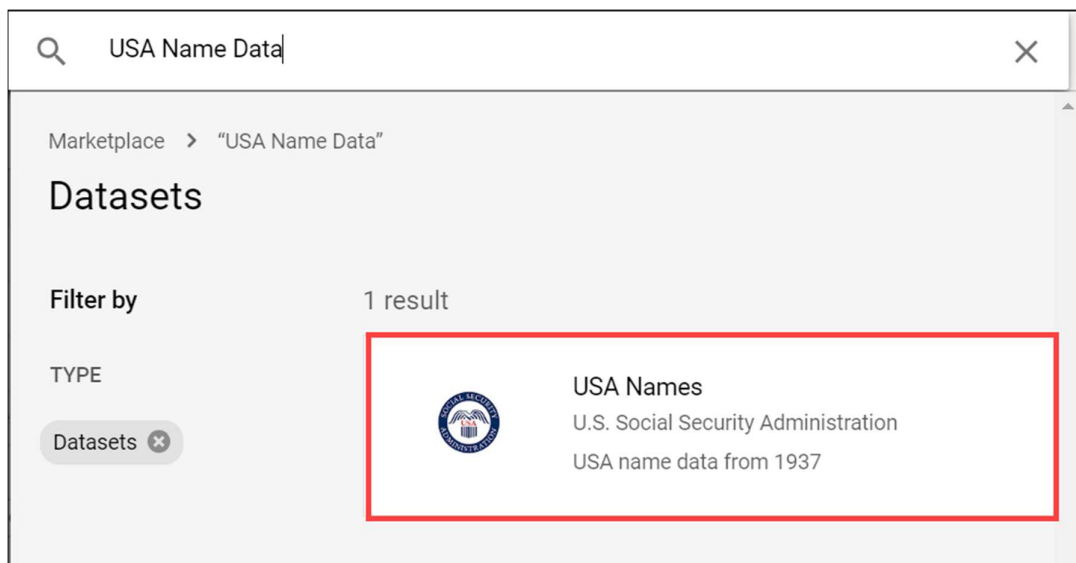
Load USA Name dataset

1. In the left pane, click **ADD DATA** > **Explore public datasets**.

The Datasets window opens.

2. In the searchbox, type "USA Names" then enter.

3. Click on the **USA Names** tile you see in the search results.

4. Click **VIEW DATASET**.

BigQuery opens in a new browser tab. The project bigquery-public-data is added to your resources and you see the dataset usa_names listed in the left pane in your Resources tree.
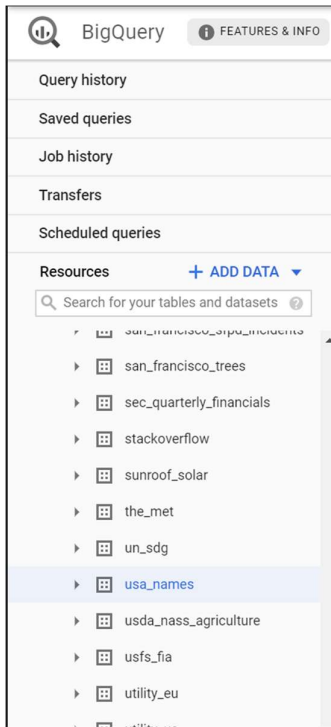
Query the USA Name dataset

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. Copy and paste the following query into the **Query editor** text area:

```
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
```

2. In the lower 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.

3. Click **Run**.

The query results opens 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.

Click **Check my progress** to verify the objective.

Create a custom table

In this section, 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.
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.

Create a dataset

In this section, 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. Back in the console, in the left pane, in the **Resources** section, click your Project ID (it will start with qwiklabs).
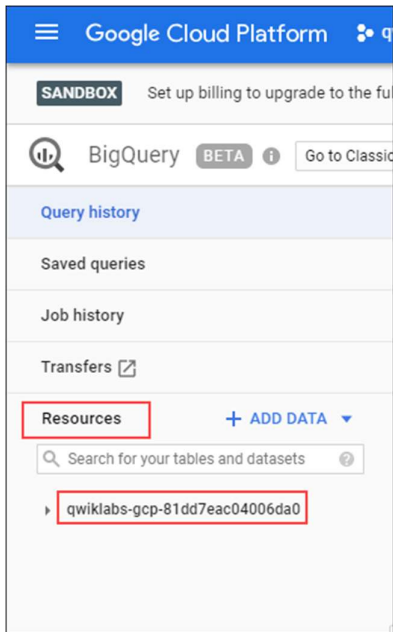
Your project opens under the Query editor.

    2. On the right side in the project section, click **CREATE DATASET**.

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

- For **Dataset ID**, enter `babynames`.
- For **Data location**, choose **United States (US)**.
- For **Default table expiration**, leave the default value.
  Currently, the public datasets are stored in the US multi-region [location](#). For simplicity, place your dataset in the same location.

    4. Click **Create dataset** at the bottom of the panel.
Click **Check my progress** to verify the objective.

Load the data into a new table

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

1. Click **babynames** found in the left pane in the **Resources** section, and then click **Create table**.

Use the default values for all settings unless otherwise indicated.

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

- For **Source**, choose **Upload** from the 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
```


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

4. Wait for BigQuery to create the table and load the data. While BigQuery loads the data, a **(1 running)** string displays beside the **Job history** in the left pane. The string disappears after the data is loaded.

Preview the table

1. In the left pane, select **babynames** > **names_2014** in the navigation panel.
2. In the details panel, click the **Preview** tab.

names_2014   🔍 QUERY TABLE   📄 COPY TABLE   🗑 DELETE TABLE   ⬆ EXPORT ▾

Schema   Details   Preview

Row	name	gender	count
1	Emma	F	20924
2	Olivia	F	19791
3	Sophia	F	18598
4	Isabella	F	17068
5	Ava	F	15688
6	Mia	F	13506
7	Emily	F	12642
8	Abigail	F	12076
9	Madison	F	10315
10	Charlotte	F	10111
11	Harper	F	9606
12	Sofia	F	9591
13	Avery	F	9563

Quick quiz. You need a table to hold the dataset
False

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.

```
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.

Query results   ⬇ SAVE AS ▾   📊 EXPLORE IN DATA STUDIO

Query complete (0.945 sec elapsed, 621.82 KB processed)

Job information   Results   JSON   Execution details

Row	name	count
1	Noah	19263
2	Liam	18440
3	Mason	17177
4	Jacob	16842
5	William	16798

Click **Check my progress** to verify the objective.

Query new dataset

Check my progress

Congratulations!

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

Finish Your Quest

This self-paced lab is part of the Qwiklabs [NCAA® March Madness®: Bracketology with Google Cloud](#), and [Data Catalog Fundamentals](#) Quests. A Quest is a series of related labs that form a learning path. Completing this Quest earns you the badge above, to recognize your achievement. You can make your badge (or badges) public and link to them in your online resume or social media account. Enroll in a Quest and get immediate completion credit if you've taken this lab. See other available Qwiklabs [Quests](#).

Take Your Next Lab

Continue your quest with [Exploring NCAA Data with Big Query](#) or check out these suggestions:
- [Bracketology with Google Machine Leanring](#)
- [BigQuery: Qwik Start - Command Line](#)

Next steps/learn more

For more information about BigQuery, see [BigQuery Documentation](#) and [BigQuery Public Datasets](#).

Google Cloud Training & Certification

...helps you make the most of Google Cloud technologies. [Our classes](#) include technical skills and best practices to help you get up to speed quickly and continue your learning

journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. [Certifications](#) help you validate and prove your skill and expertise in Google Cloud technologies.

Manual Last Updated June 23, 2020

Lab Last Tested May 24, 2019