Explore a BigQuery Public Dataset

1 hour 1 Credit  Rate Lab

Overview

Storing and querying massive datasets can be time consuming and expensive without the right hardware and infrastructure. Google 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 GCP 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 GCP 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 GCP 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 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 the Google Cloud Platform for the duration of the lab.

What you need

To complete this lab, you need:

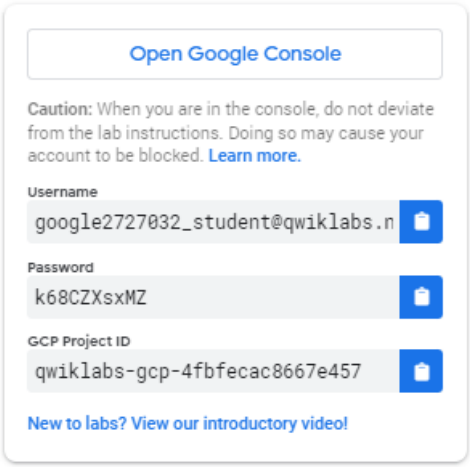
- Access to a standard internet browser (Chrome browser recommended).
- Time to complete the lab.

Note: If you already have your own personal GCP account or project, do not use it for this lab.

Google Cloud Platform Console

How to start your lab and sign in to the Console

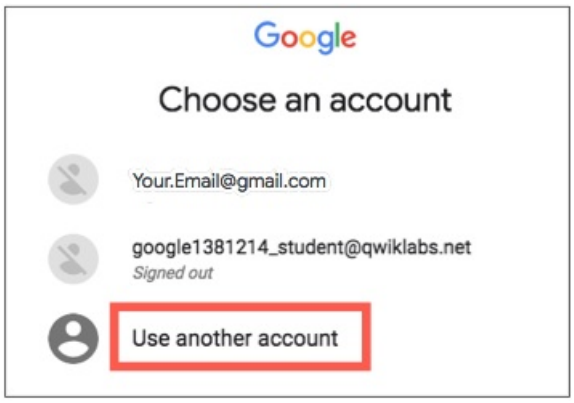
1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is a panel populated with the temporary credentials that you must use for this lab.



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

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

3. On the Choose an account page, click **Use Another Account**.



4. The Sign in page opens. Paste the username that you copied from the Connection Details panel. Then copy and paste the password.

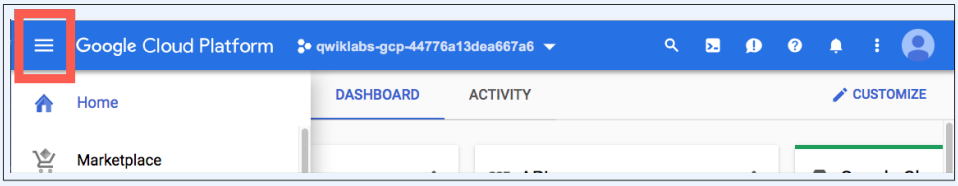
   Important: You must use the credentials from the Connection Details panel. Do not use your Qwiklabs credentials. If you have your own GCP account, do not use it for this lab (avoids incurring charges).

5. 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.

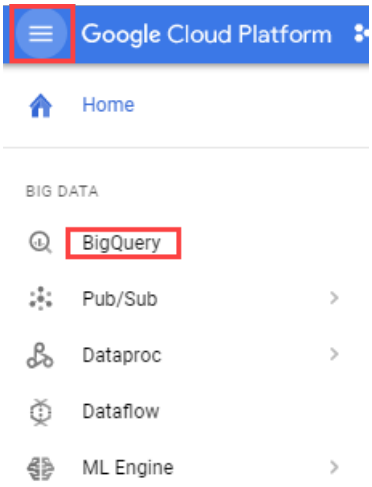
   After a few moments, the GCP console opens in this tab.

Note: You can view the menu with a list of GCP Products and Services by clicking the **Navigation menu** at the top-left, next to "Google Cloud Platform".



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 lists UI updates.

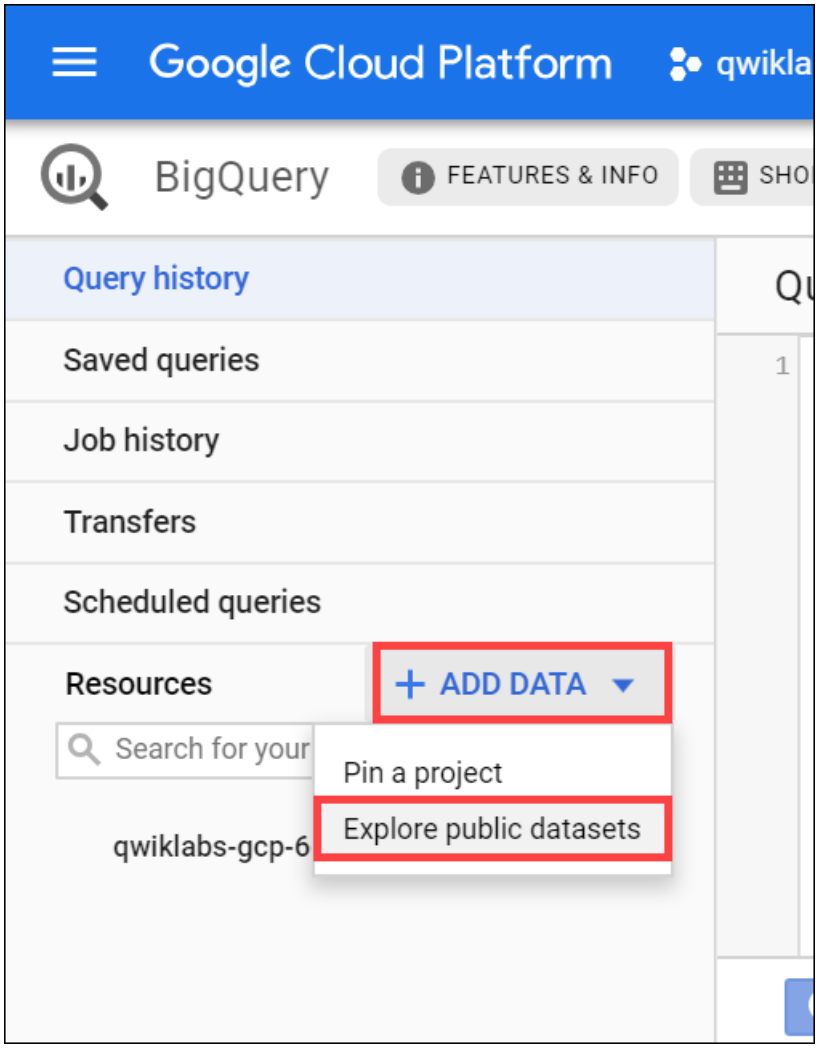
Click **Done**.

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

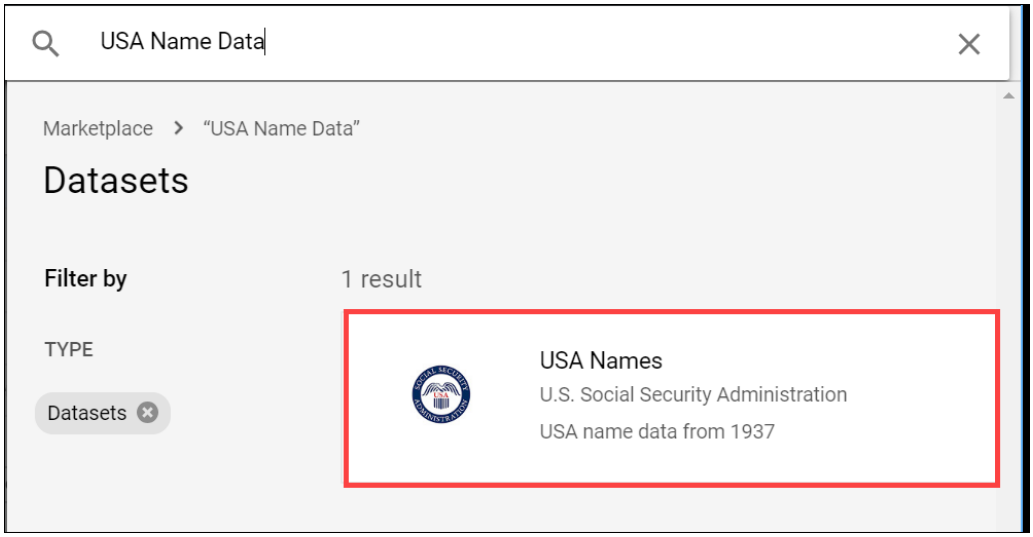
1. In the left pane, click **ADD DATA** > **Explore public datasets**.



The Datasets window opens.

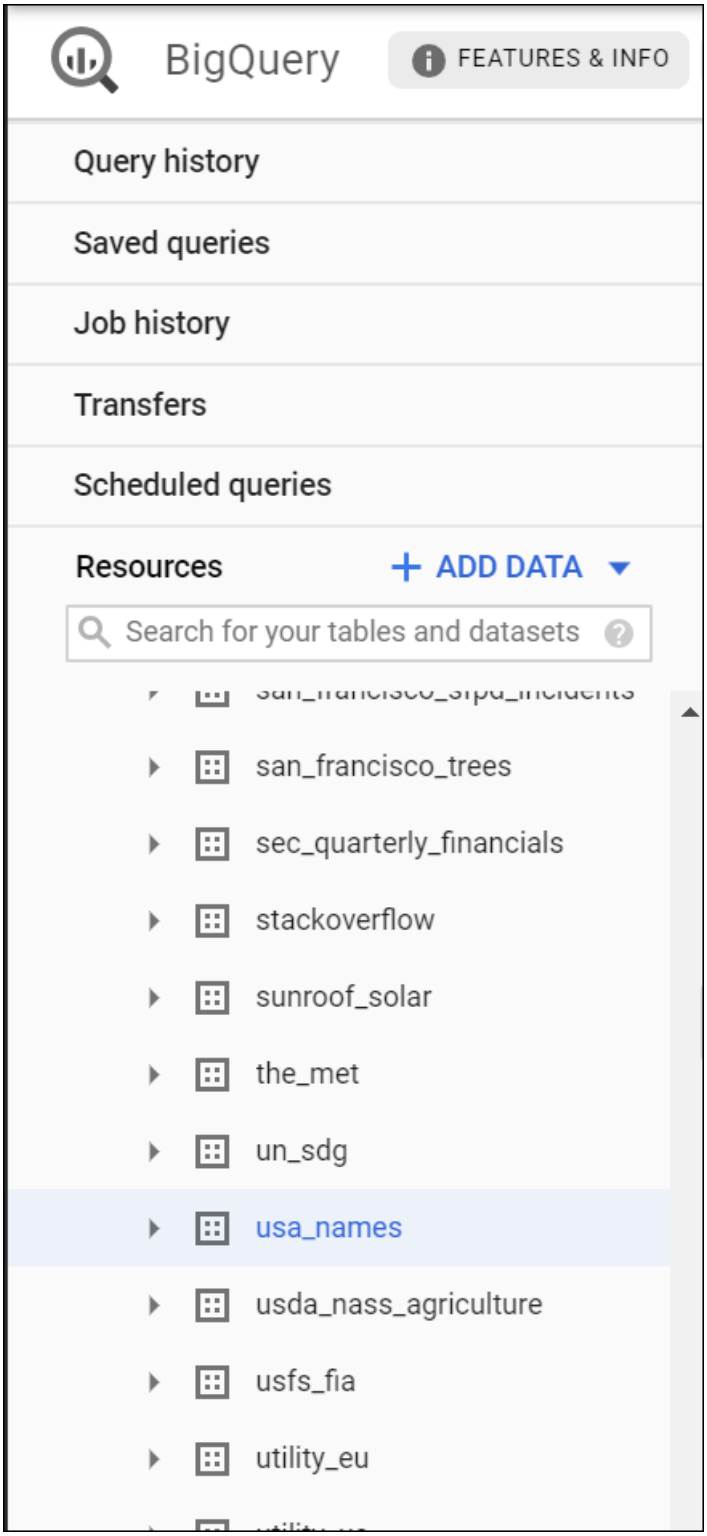
2. In the searchbox, type `USA Names` then Enter.

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.



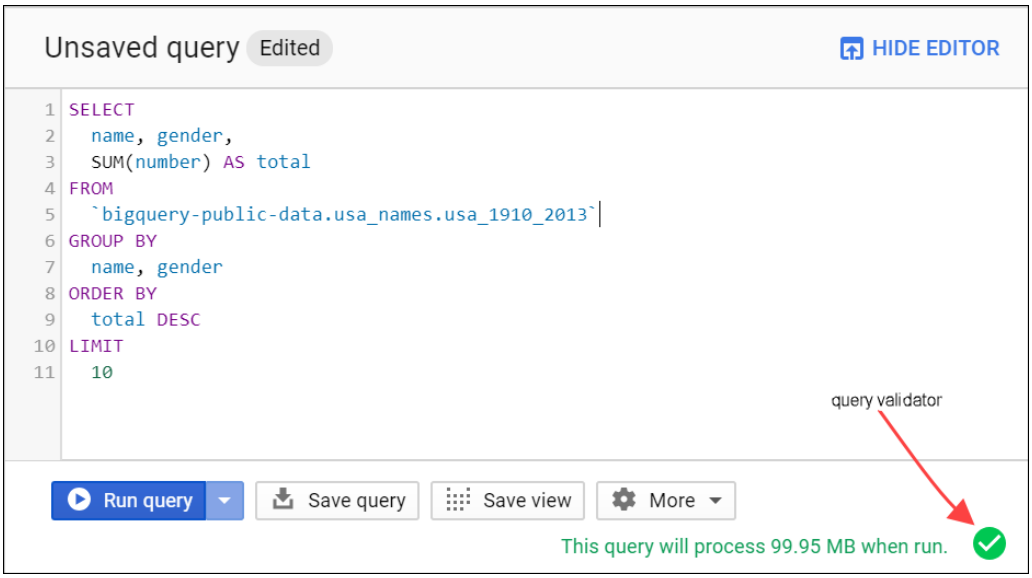
Query the USA Name dataset

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
```

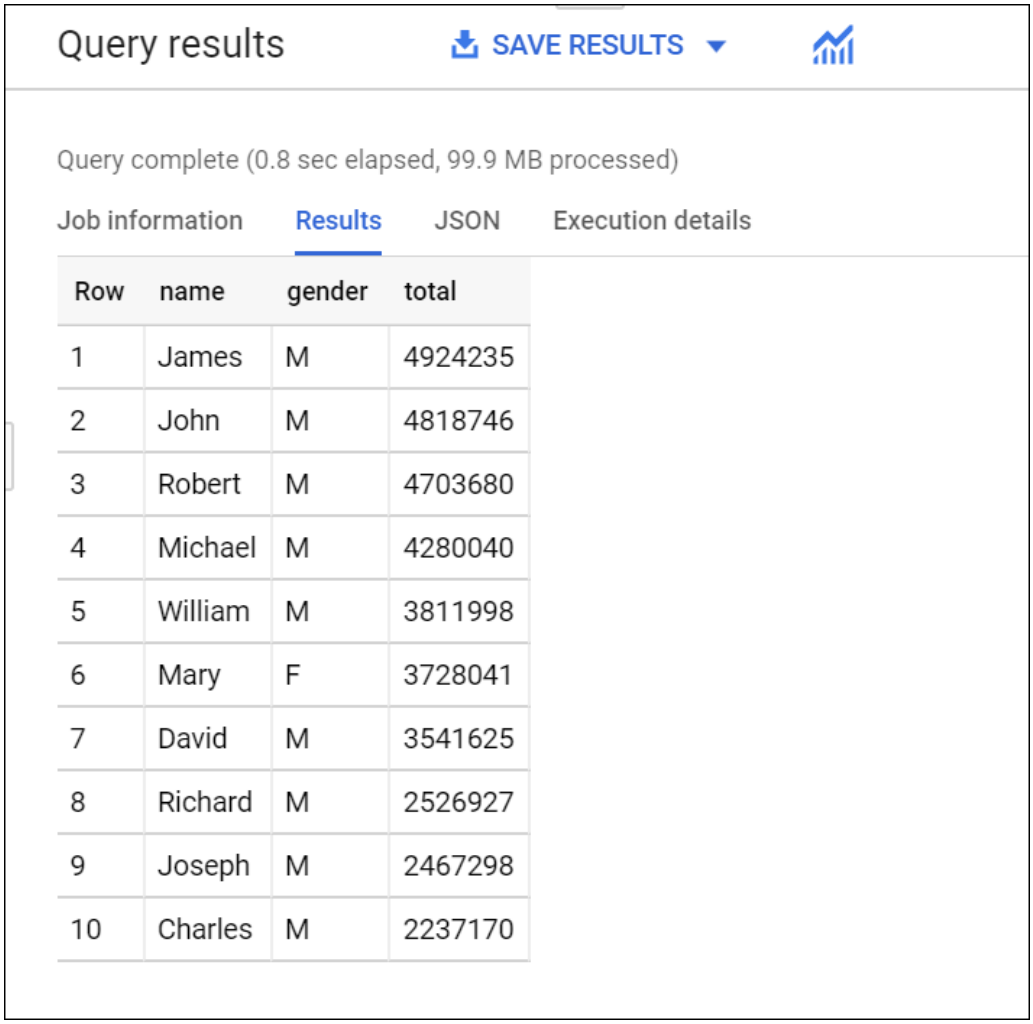
2. In the lower right of the window, view the query validator.



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.



Row	name	gender	total
1	James	M	4924235
2	John	M	4818746
3	Robert	M	4703680
4	Michael	M	4280040
5	William	M	3811998
6	Mary	F	3728041
7	David	M	3541625
8	Richard	M	2526927
9	Joseph	M	2467298
10	Charles	M	2237170

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.

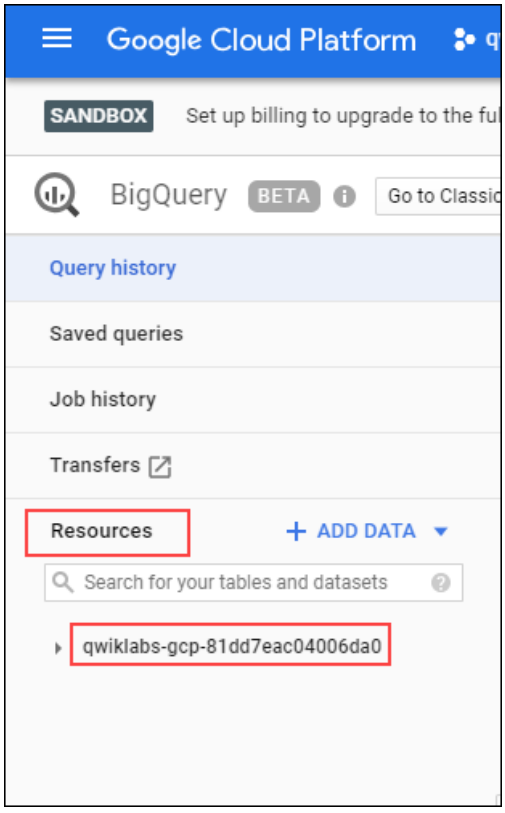
5. Note the location of the `yob2014.txt` file so that you can find it later.

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 GCP Project ID (it will start with qwiklabs).



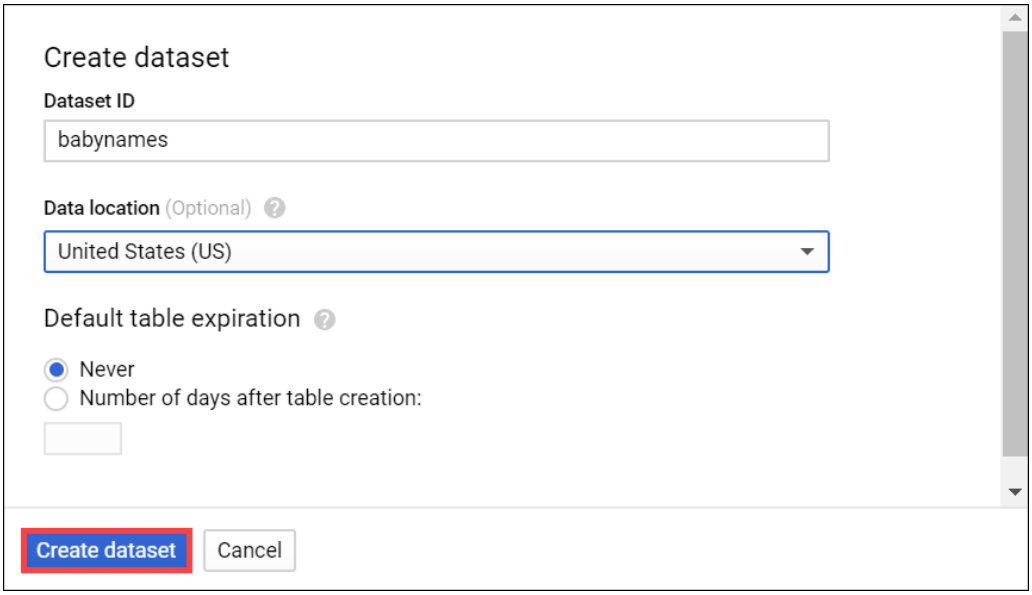
Your project opens under the Query editor.

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.

Load the data into a new table

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

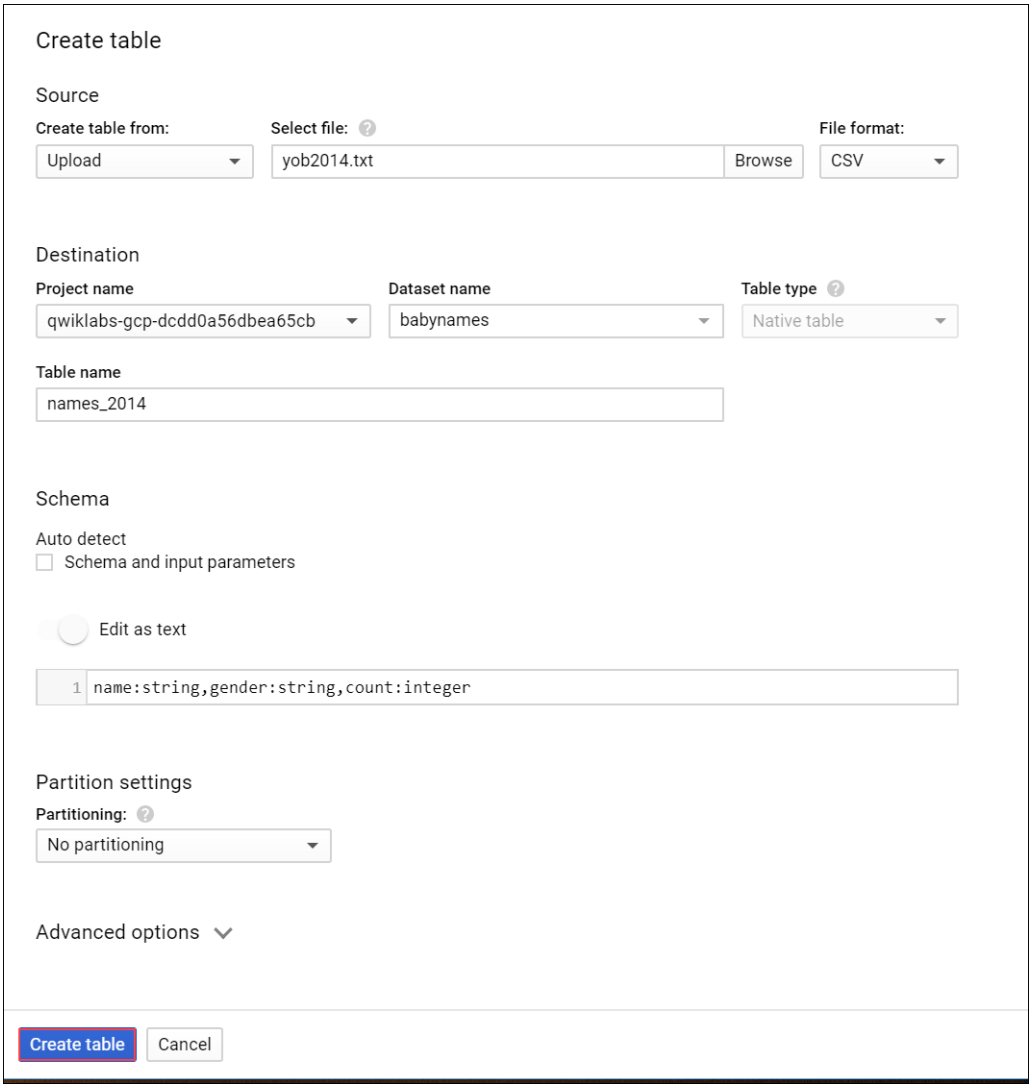
1. Click **babynames** found in the left pane in the **Resources** section, and then click **Create table**.

   Use the default values for all settings unless otherwise indicated.

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

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



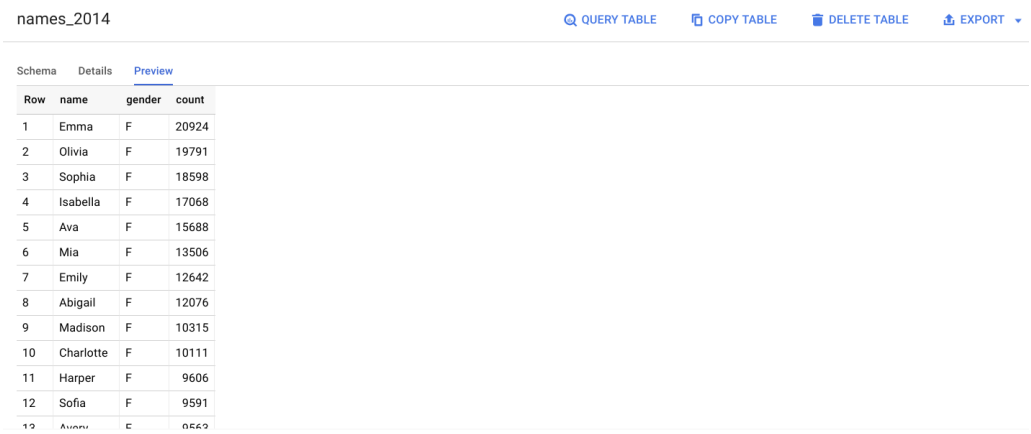
3. Click **Create table** (at the bottom of the window).

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.



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

Query the table

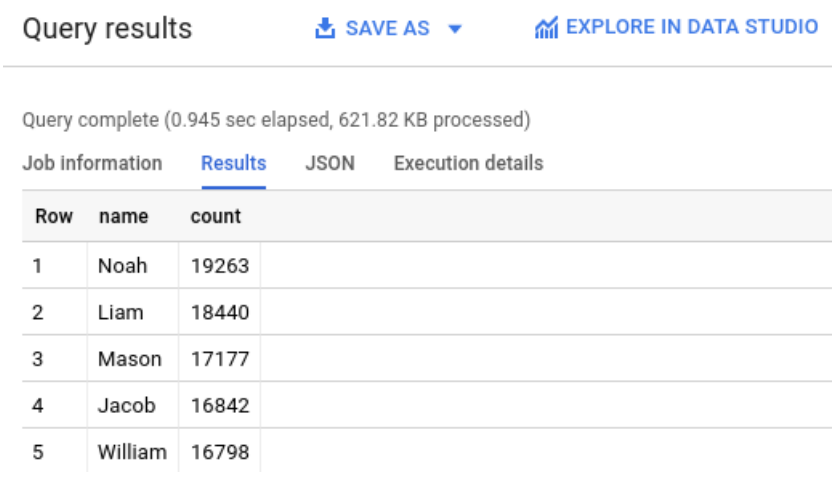
Now that you've loaded data into your table, you can run queries against it. The process is identical to the previous example, except that this time, you're querying your table instead of a public table.

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.



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

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**. 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.