

Big Query Experiment :

Exploring a BigQuery Public Dataset

1 hourFree
[Rate Lab](#)

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:

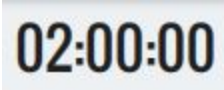
- Query a public dataset
- Create a custom table
- Load data into a table
- Query a table

Set up your environments

Qwiklabs setup

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

1. Make sure you signed into Qwiklabs using an **incognito window**.

2. Note the lab's access time (for example,  and make sure you can finish in that time block.


There is no pause feature. You can restart if needed, but you have to start at the beginning.


3. When ready, click  .


4. Note your lab credentials. You will use them to sign in to Cloud Platform Console.

Open Google Console

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. [Learn more.](#)

Username
google2876526_student@qwiklabs.n 

Password
TG959yrKDX 

GCP Project ID
qwiklabs-gcp-0855e773352d3560 

[New to labs? View our introductory video!](#)

5. Click **Open Google Console**.
6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.

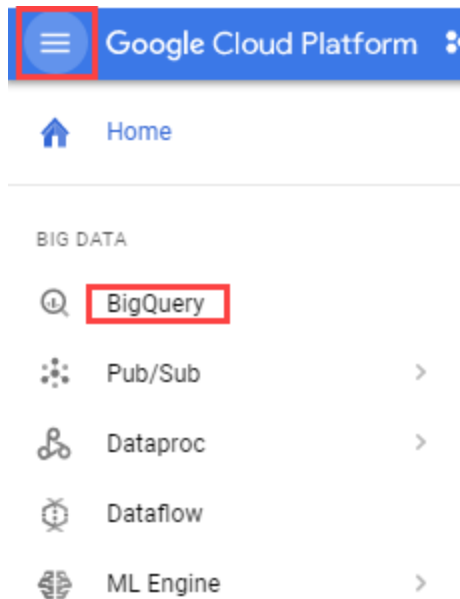
If you use other credentials, you'll get errors or **incur charges**.

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

Do not click **End Lab** unless you are finished with the lab or want to restart it. This clears your work and removes the project.

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.

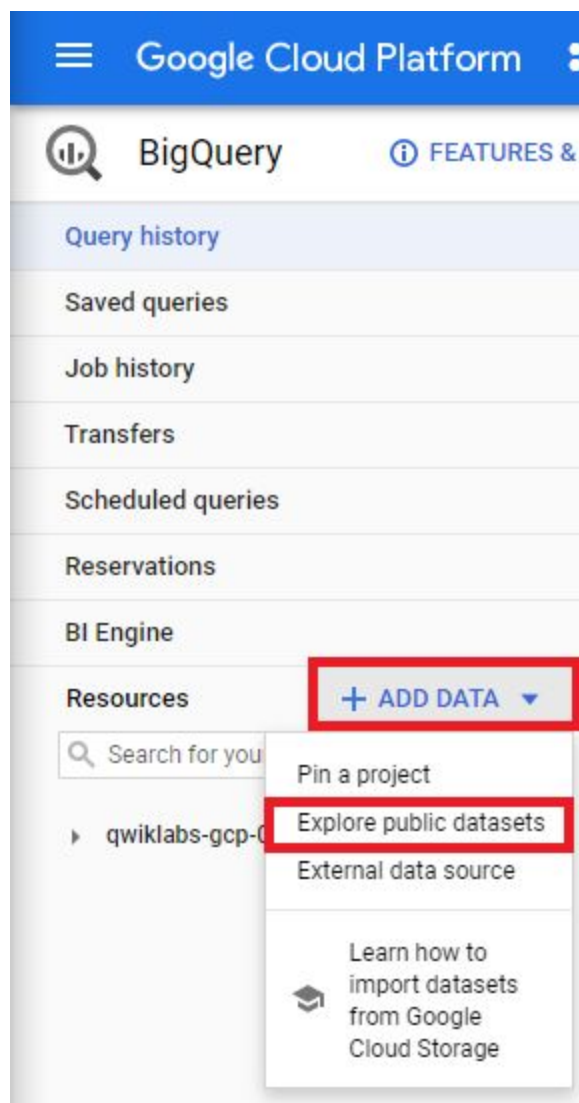
Click **Done**.

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.

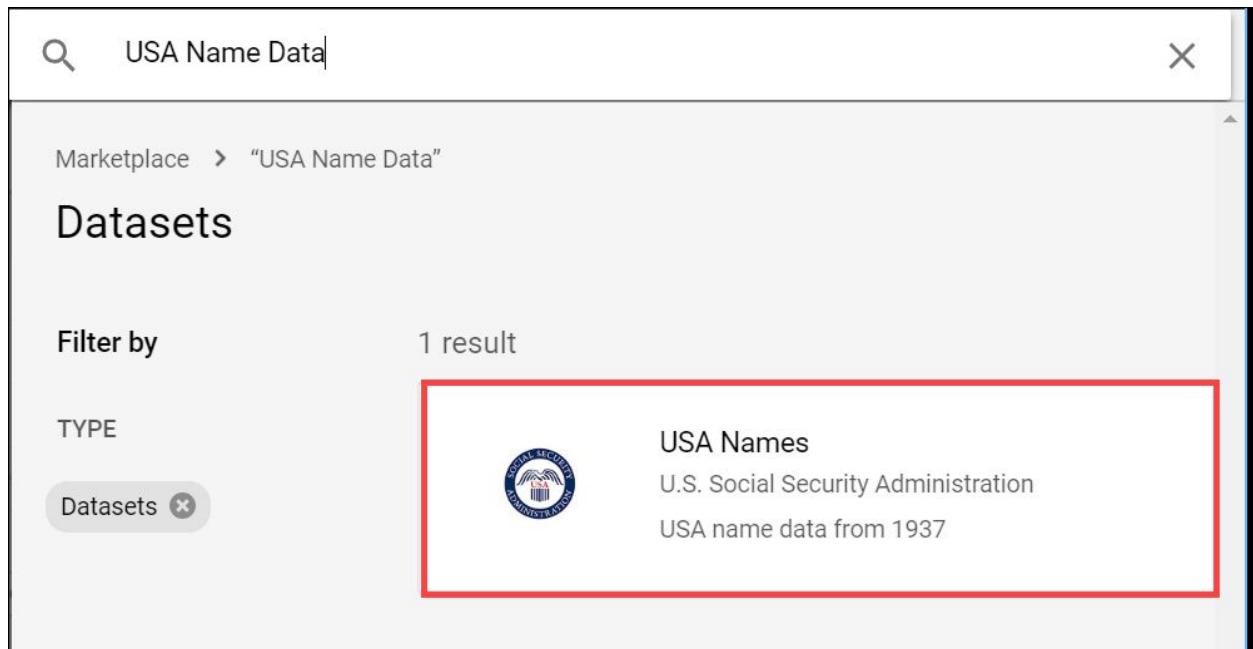
Load the USA Names dataset

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



The Datasets window opens.

2. In the searchbox, type USA Names then press 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.

Google Cloud Platform

BigQuery

FEATURES

Query history

Saved queries

Job history

Transfers

Scheduled queries

Reservations

BI Engine

Resources

+ ADD DATA

Search for your tables and datasets

usa_names

the_met

umiami_lincs

un_sdg

usa_names

usda_nass_agriculture

usfs_fia

utility_eu

utility_us

Query the USA Names dataset

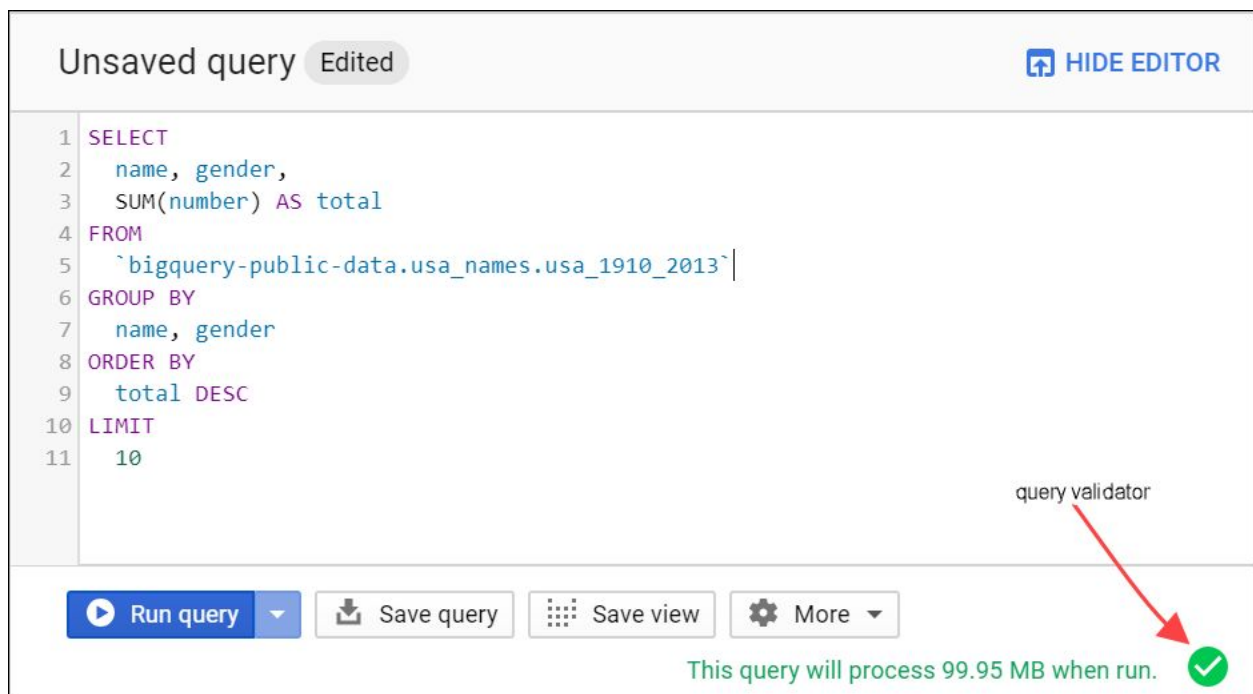
Query `bigquery-public-data.usa_names.usa_1910_2013` for the name and gender of the babies in this dataset, and then list the top 10 names in descending order.

1. 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
```

content_copy

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



The screenshot shows the BigQuery Query Editor interface. At the top, it says "Unsaved query" and "Edited". There is a "HIDE EDITOR" button. The query text is as follows:




```
1 SELECT
2   name, gender,
3   SUM(number) AS total
4 FROM
5   `bigquery-public-data.usa_names.usa_1910_2013`
6 GROUP BY
7   name, gender
8 ORDER BY
9   total DESC
10 LIMIT
11   10
```

At the bottom, there are buttons for "Run query", "Save query", "Save view", and "More". Below these buttons, a status message says "This query will process 99.95 MB when run." followed by a green checkmark icon. A red arrow points from the text "query validator" to the green checkmark icon.

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.

Query results				 SAVE RESULTS 	
Query complete (0.8 sec elapsed, 99.9 MB processed)					
Job information Results JSON Execution details					
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		

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.

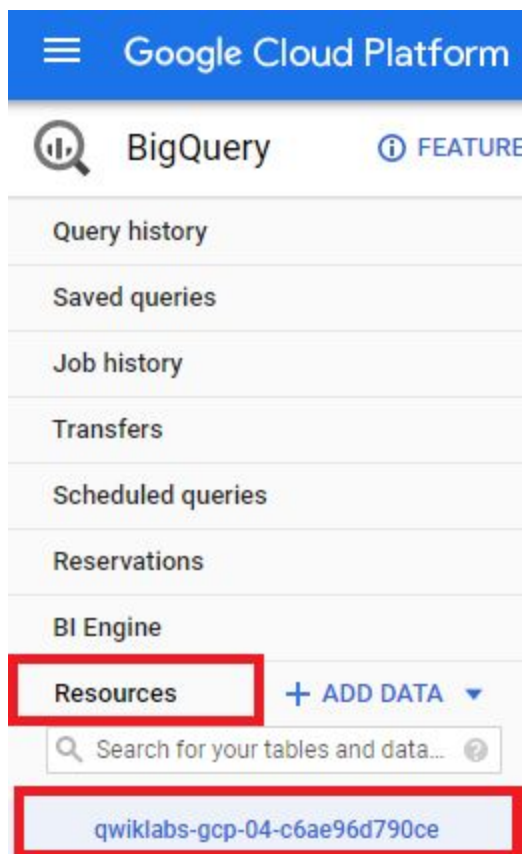
1. Download the [baby names zip file](#) to your local computer.
2. **Note:** If this download link fails please copy the baby names zip file from the student resources on the left pane of the instruction guide.
3. Unzip the file onto your computer.
4. The zip file contains a `NationalReadMe.pdf` file that describes the dataset.
[Learn more about the dataset.](#)
5. 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.
6. 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. Back in the Cloud 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.
- For **Encryption**, 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.

Create dataset

Dataset ID

Data location (Optional) ?

United States (US) ▼

Default table expiration ?

☒ Never

☐ Number of days after table creation:

Encryption
Data is encrypted automatically. Select an encryption key management solution.

☒ Google-managed key
No configuration required

☐ Customer-managed key
Manage via Google Cloud Key Management Service

Create dataset

Cancel

4. 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 navigation pane, click **babynames** from 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
```

```
content_copy
```

Create table

Source

Create table from:

Upload

 Select file:

yob2014.txt

Browse

 File format:

CSV

Destination

Project name

qwiklabs-gcp-dcdd0a56dbea65cb

 Dataset name

babynames

 Table type

Native table

Table name

names_2014

Schema

Auto detect
☐ Schema and input parameters

Edit as text

1

 name:string,gender:string,count:integer

Partition settings

Partitioning:

No partitioning

Advanced options

Create table

Cancel

- Click **Create table** (at the bottom of the window).
- 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 pane.
2. In the details pane, 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	8562

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

True

False

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.

```
SELECT
  name, count
FROM
  `babynames.names_2014`
WHERE
  gender = 'M'
ORDER BY count DESC LIMIT 5
```

content_copy

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	

Quick quiz. Check which you can use to access BigQuery.

Web UI

Third-party tools

Make calls to BigQuery REST API

Command-line tool

Submit

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.

Copyright 2020 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.