DA 320 Project 6 M. Blanco

**Part 1**: Follow the tutorial “Quickstart Using the Web UI”. Include all command lines/queries and output/query results.

Quickstart Using the Web UI

You can use the BigQuery web UI as a visual interface to complete tasks like running queries, loading data, and exporting data. This quickstart shows you how to query public tables and load sample data into BigQuery.

# Before you begin

1. You may select or create a Cloud Platform project.

[GO TO THE MANAGE RESOURCES PAGE](https://console.cloud.google.com/cloud-resource-manager)

1. Enable billing for your project.

[ENABLE BILLING](https://support.google.com/cloud/answer/6293499#enable-billing)

1. BigQuery is automatically enabled in new projects. To activate BigQuery in a preexisting project, go to Enable the BigQuery API.

[ENABLE THE API](https://console.cloud.google.com/flows/enableapi?apiid=bigquery)

# Query a public dataset

The BigQuery web UI provides an interface to query tables, including [public datasets](https://cloud.google.com/bigquery/public-data) offered by BigQuery, e.g., <https://www.reddit.com/r/bigquery/wiki/datasets>

1. Go to the BigQuery web UI.

[GO TO THE BIGQUERY WEB UI](https://bigquery.cloud.google.com/)

1. Click the **Compose query** button.
2. Copy and paste the following query into the **New Query** text area:

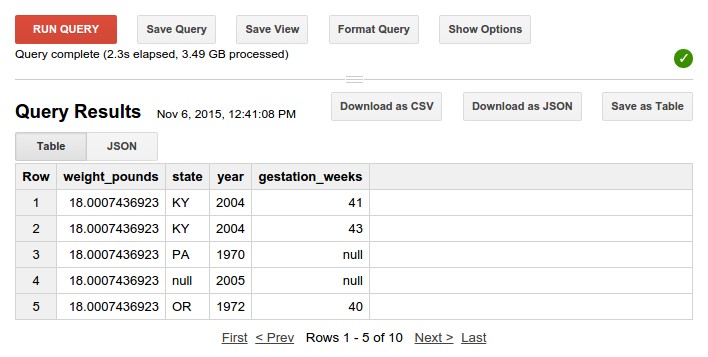
|  |
| --- |
| #standardSQL SELECT  weight\_pounds, state, year, gestation\_weeks  FROM  `bigquery-public-data.samples.natality`  ORDER BY weight\_pounds DESC LIMIT 10; |

1. Click the circular icon to activate the query validator.



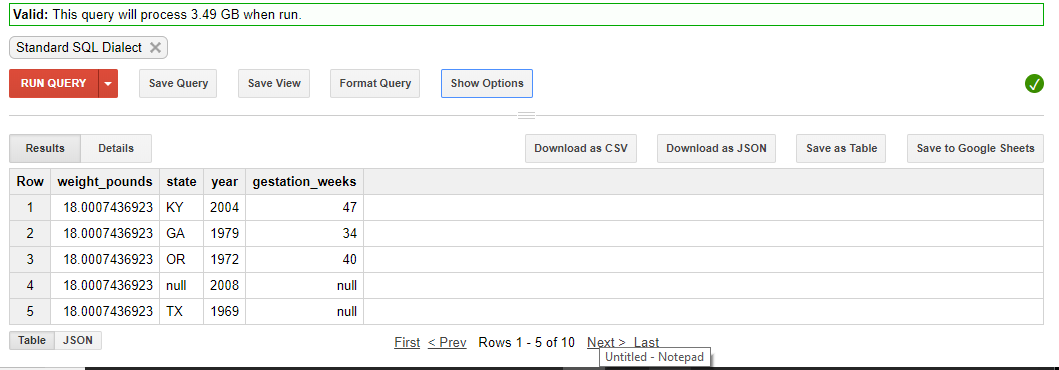
A green or red section displays above the buttons depending on whether the query is valid or invalid. If the query is valid, the validator also describes the amount of data to be processed after you run the query. This information is helpful for determining the cost to run a query.

1. Click the **Run query** button. The query results display below the buttons.



The above query accesses a table from a public dataset that BigQuery provides.

You can browse the schema of other public datasets by clicking [**bigquery-publicdata:samples**](https://bigquery.cloud.google.com/dataset/bigquery-public-data:samples) in the left-hand navigation of the web UI. The expanded list of titles are all of the public tables you can query against.



# Load data into a table

Next, we'll load custom data into a table and run a query against it.

As a reminder, billing must be enabled to run this part of the quickstart. For more information, see [Before you begin.](https://cloud.google.com/bigquery/quickstart-web-ui#before-you-begin)

## Download custom data

The custom data contains approximately 7 MB of data about popular baby names, and it is provided by the US Social Security Administration.

1. Download the [babynames zip file.](http://www.ssa.gov/OACT/babynames/names.zip)

2. Unzip the file onto your hard drive.

The zip file contains a read me file that describes the dataset schema. [Learn more about the dataset.](http://www.ssa.gov/OACT/babynames/background.html)

1. Open the file named yob2014.txt to see what it looks like. The file is a commaseparated 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.
2. Note the location of the yob2014.txt file so that you can find it later.

## Create a dataset

Next, create a dataset in the web UI to hold the data.

1. Go to the BigQuery web UI.

[GO TO THE BIGQUERY WEB UI](https://bigquery.cloud.google.com/)

1. Click the down arrow icon  next to your project name in the navigation, then click **Create new dataset**.
2. Input the following name for the dataset ID.

|  |
| --- |
| babynames |
| Dataset IDs are unique on a per-project basis, so if **babynames** is already listed under your project name in the navigation, append a number to the name to make it unique. Click the question mark icon to see ID limitations. |

1. Leave all of the other default settings in place and click **OK**.

## Load the data into a new table

Next, load the data into a new table.

1. In the navigation, hover on the **babynames** dataset ID that you just created.
2. Click the down arrow icon next to the ID and click **Create new table**.

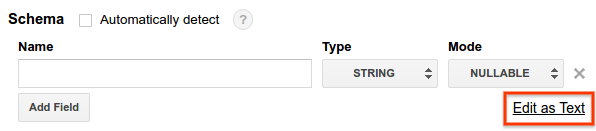


Use the default values for all settings unless otherwise indicated.

1. Under **Source Data**, click the **Choose file** button. Navigate to the data you unzipped in the step above, and select the yob2014.txt file.
2. Under **Destination Table**, enter the following value for the destination table name.

names\_2014

1. In the **Schema** section, click the **Edit as Text** link.



Then replace the contents of the **Schema** input area with the following schema:

name:string,gender:string,count:integer

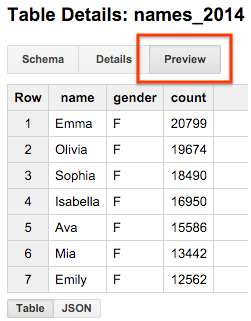
1. Click the **Create Table** button.
2. Wait for BigQuery to create the table and load the data. While BigQuery loads the data, a **(loading)** string displays after your table name in the navigation. The string disappears after the data has been fully loaded.

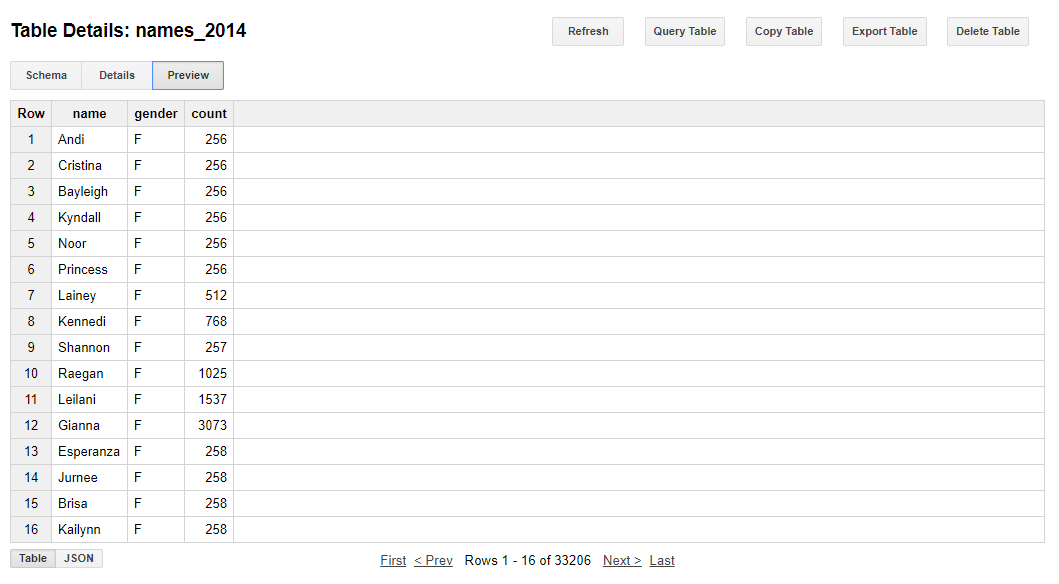


## Preview the table

After the **(loading)** string disappears, you can access the table. To preview the first few rows of the data:

1. Select **names\_2014** in the navigation.
2. Click **Preview** in the **Table Details: names\_2014** section.





## Query the table

Now that you've loaded custom data into a table, you can run queries against it. The process is identical to the [query a public dataset](https://cloud.google.com/bigquery/quickstart-web-ui#query_a_public_dataset) example above, except that this time, you're querying your custom table instead of a public table.

1. Click the **Compose query** button.
2. Copy and paste the following query into the **New Query** text area.

|  |
| --- |
| #standardSQL SELECT name, count  FROM  `babynames.names\_2014` WHERE gender = 'M'  ORDER BY count DESC LIMIT 5; |

1. Click the **Run query** button. The query displays the top 5 men's names for the year of data you loaded into the table.



# Clean up

To avoid incurring charges to your Google Cloud Platform account for the resources used in this quickstart:

1. Go to the BigQuery web UI.

[GO TO THE BIGQUERY WEB UI](https://bigquery.cloud.google.com/)

1. In the navigation, hover on the **babynames** dataset you created.
2. Click the down arrow icon  next to your dataset name in the navigation, then click **Delete dataset**.
3. In the **Delete dataset** dialog box, confirm the delete command by typing the name of your dataset (babynames) and clicking **OK**.

**Part 2**: Follow the tutorial “Quickstart Using the bq Command”. Include all command lines/queries and output/query results.

Quickstart Using the bq Command-Line Tool

This page explains how to use the bq command-line tool to run queries, load data, and export data.

# Before you begin

Before you begin this quickstart, use the Google Cloud Platform Console to create or select a project and enable billing. You also need to install the Google Cloud SDK.

1. Select or create a Cloud Platform project.

[GO TO THE MANAGE RESOURCES PAGE](https://console.cloud.google.com/cloud-resource-manager)

1. Enable billing for your project.

[ENABLE BILLING](https://support.google.com/cloud/answer/6293499#enable-billing)

1. [Install and initialize the Cloud SDK.](https://cloud.google.com/sdk/docs/)
2. BigQuery is automatically enabled in new projects. To activate BigQuery in a preexisting project, go to Enable the BigQuery API.

[ENABLE THE API](https://console.cloud.google.com/flows/enableapi?apiid=bigquery)

# Examine a table

BigQuery offers a number of [sample tables](https://cloud.google.com/bigquery/docs/sample-tables) that you can run queries against. In this quickstart, you'll run some queries against the Shakespeare table, which contains an entry for every word in every play.

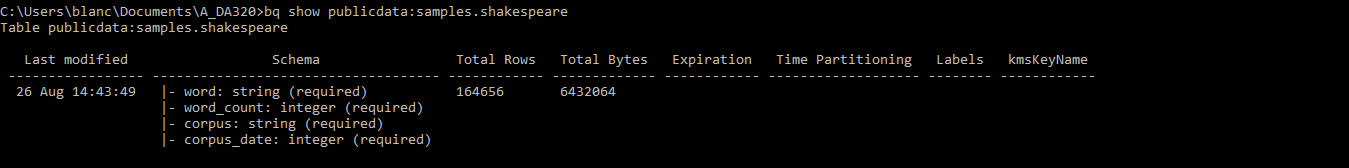
To examine the schema of a specific table, run

bq show *projectId*:*datasetId*.*tableId*

where the project and dataset IDs can be omitted if they are the default values for your bq tool. The following example examines the shakespeare table in the samples dataset:

bq show publicdata:samples.shakespeare

Output:



# Run the help command

Use bq help to get detailed information about the bq command-line tool.

bq help

Include a command name to get information about a specific command. For example, the following call to bq help retrieves information about the query command.

bq help query

# Run a query

Run a query to see how many times the substring "raisin" appears in Shakespeare's works.

To run a query, run the command bq query "*sql\_statement*".

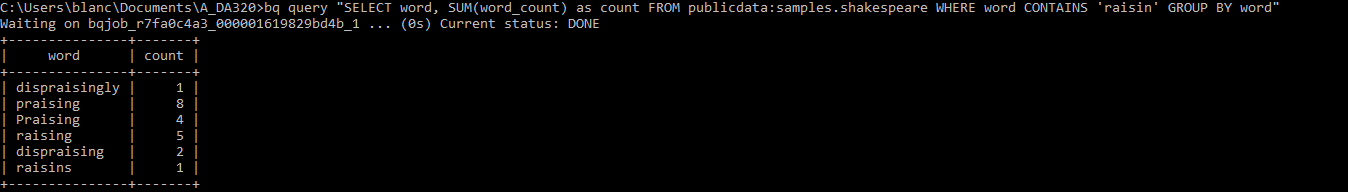
* Escape any quotation marks inside the sql\_statement with a \ mark, or
* Use a different quotation mark type than the surrounding marks (" versus').

The following example does a word count for the number of times that the substring "raisin" appears in all of Shakespeare's works. The query shown is case-sensitive; string comparisons are case-sensitive, unless you use the SQL feature IGNORE CASE.

bq query "SELECT word, SUM(word\_count) as count FROM

publicdata:samples.shakespeare WHERE word CONTAINS 'raisin' GROUP BY word"

Output:



If you search for a word that isn't in Shakespeare's works, no results are returned. For example, the following search for "huzzah" returns no matches.

bq query "SELECT word FROM publicdata:samples.shakespeare WHERE word = 'huzzah' IGNORE CASE"

Output:



# Create a new table

Now create your own table. Every table must live inside a dataset, which is simply a group of tables. A dataset is assigned to a single project.

## Step 1: Download custom data

The custom data contains approximately 7 MB of data about popular baby names, and is provided by the US Social Security Administration.

1. [DOWNLOAD THE BABY NAMES ZIP FILE](http://www.ssa.gov/OACT/babynames/names.zip)

2. Unzip the file onto your hard drive.

The zip file contains a read me file that describes the dataset schema. [Learn more about the dataset.](http://www.ssa.gov/OACT/babynames/background.html)

1. Open the file named yob2010.txt to see what it 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.
2. Copy or move the yob2010.txt file into the directory you are using to run bq commands.

## Step 2: Create a new dataset

1. Use the bq ls command to see whether your default project has any existing datasets.

bq ls

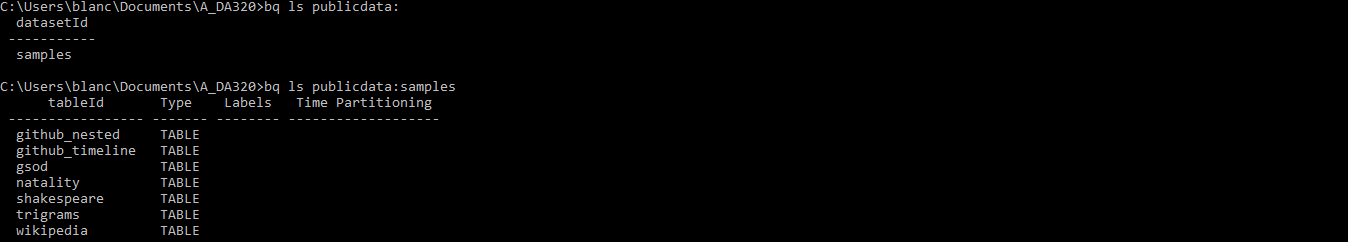
Output:



1. Run bq ls again to list the datasets in a specific project by including the project ID followed by a colon (:). The following example lists the datasets in the publicdata project.

bq ls publicdata:

Output:



1. Use the bq mk command to create a new dataset named babynames in your default project. A dataset name can be up to 1,024 characters long, and consist of A-Z, a-z, 0-

9, and the underscore, but it cannot start with a number or underscore, or have spaces.

bq mk babynames

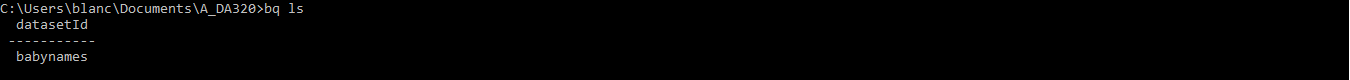
Output:



1. Run bq ls to confirm that the dataset now appears as part of the default project:

bq ls

Output:



## Step 3: Upload the table

The bq load command creates or updates a table and loads data in a single step.

1. Run the bq load command to load your source file into a new table called names2010 in the babynames dataset you created above. By default, this runs synchronously, and will take a few seconds to complete.

bq load babynames.names2010 yob2010.txt name:string,gender:string,count:integer

The bq load command arguments:

* datasetID: babynames
* tableID: names2010
* source: yob2010.txt
* schema: name:string,gender:string,count:integer

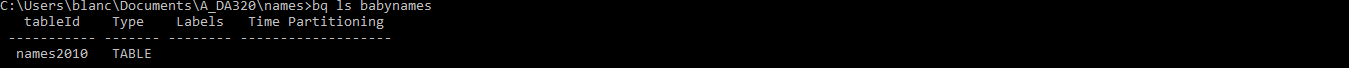
Output:



1. Run bq ls to confirm that the table now appears in the dataset:

bq ls babynames

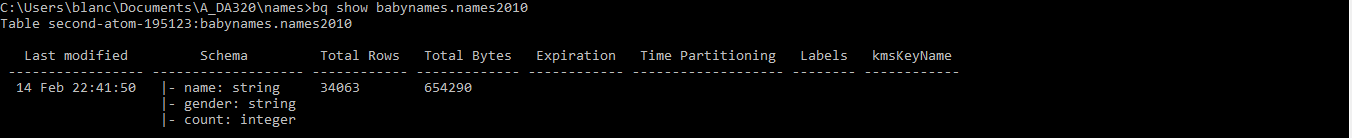
Output:



1. Run bq show to see the schema:

bq show babynames.names2010

Output:



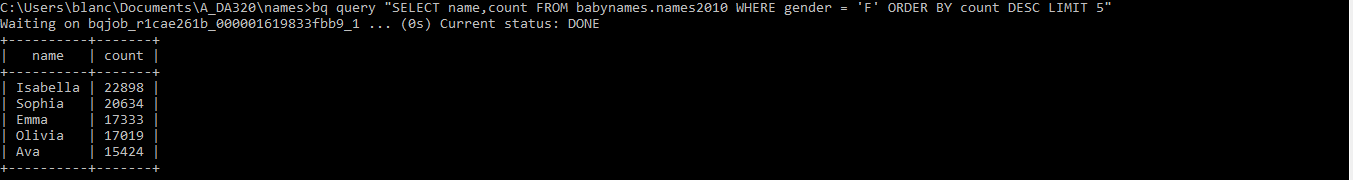
By default, when you load data, BigQuery expects UTF-8 encoded data. If you have data that is in ISO-8859-1 (or Latin-1) encoding and are having problems with your loaded data, you can tell BigQuery to treat your data as Latin-1 explicitly, using the E flag. For more information, see [Character Encodings.](https://cloud.google.com/bigquery/bq-command-line-tool#characterencodings)

## Step 4: Run queries

1. Run the following command to return the most popular girls' names:

bq query "SELECT name,count FROM babynames.names2010 WHERE gender = 'F' ORDER BY count DESC LIMIT 5"

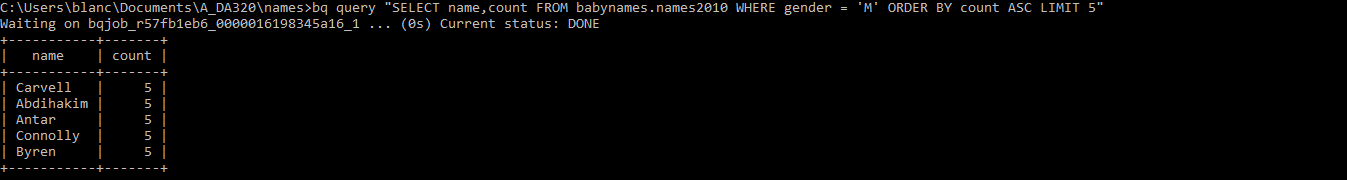
Output:



1. Run the following command to see the most unusual boys' names. The minimum count is 5 because the source data omits names with fewer than 5 occurrences.

bq query "SELECT name,count FROM babynames.names2010 WHERE gender = 'M' ORDER BY count ASC LIMIT 5"

Output:



# Clean up

To avoid incurring charges to your Google Cloud Platform account for the resources used in this quickstart:

1. Run the bq rm command to remove the babynames dataset. Use the -r flag to delete all tables in the dataset, include the names2010 table.

bq rm -r babynames

1. Confirm the delete command by typing y.



**Part 3**: **(Optional)**

Follow the BigQuery API Quickstart exercise found at <https://cloud.google.com/bigquery/create-simple-app-api>Choose any one of languages (C#, GO, JAVA, NODE.JS, PHP, PYTHON, or RUBY), and use it to build an application with BigQuery API. Include all command lines/queries and output/query results.