

Lab Title:

**Introduction to DBMS and databases**

Submitted to:

**Ma’am Darakhshan**

Submitted by:

**SABA**

Course:

**CS-363L Database Systems**

Semester:

**6th**

Date:

**22ndJanuary, 2022**

**Department of Computer Engineering**

**University of Engineering and Technology, Lahore**

**BigQuery**

**What is BigQuery?**

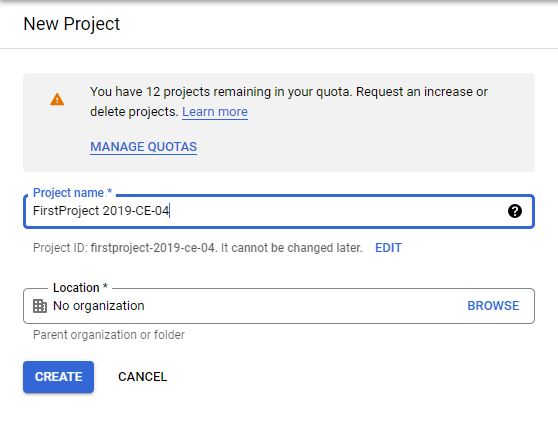
BigQuery is a highly scalable corporate data warehouse that comes with built-in tools like machine learning, geospatial analysis, and business intelligence to help you organize and analyze your data. The server less design of BigQuery allows you to utilize SQL queries to solve your organization's most pressing problems while requiring no infrastructure administration. The scalable, distributed analytical engine in BigQuery allows you to query terabytes of data in seconds and petabytes of data in minutes.

**BigQuery Sandbox**

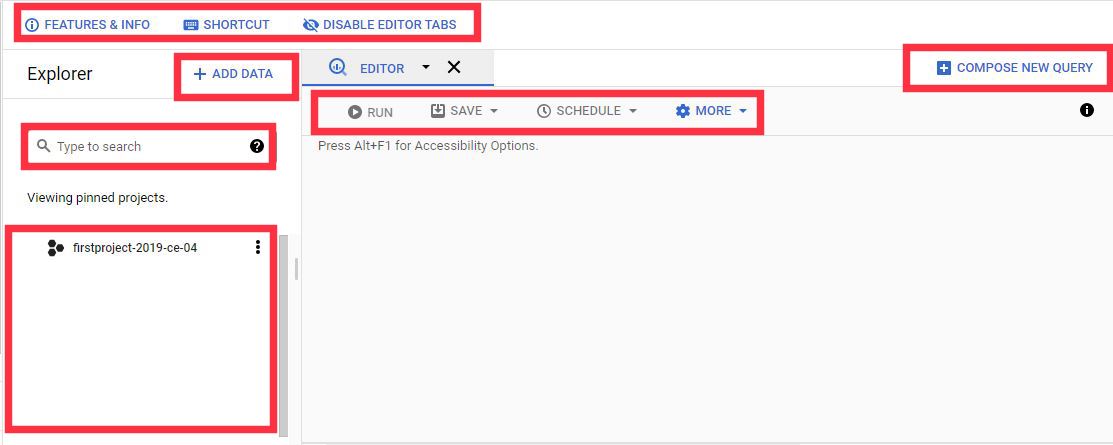
The BigQuery Sandbox is a graphical user interface for creating and managing BigQuery resources as well as running SQL queries. The sandbox gives us 10 GB of active storage and 1 TB of processed query data per month.

* **Creating a new project**

Before exploring BigQuery, we must log in to Cloud Console and create a project. I have created a new project called ‘FirstProject 2019-CE-04’.



* There are multiple options in BigQuery sandbox. I have explained one by one.



* **Search**

In search box, we can search through our resources by name or label. Includes partial matches.

* **Project Pane**

In the bottom left side, we have a pane where we can see our project. We can also pinned our projects.

* **Add Data**

In this dropdown we have three options.

* + Pin a Project— To pin a project by searching or entering project name.
  + Explore public datasets— To explore public datasets
  + External data source— To create connections with external data.
* **Features & Info**

We can see new features and Information of BigQuery on clicking Features & Info.

* **Shortcut**

Here we can see all the keyboard shortcuts of the BigQuery sandbox.

* **Compose New Query**

This will open a new tab for new query.

* **Run Save Schedule**

These options are used for running saving and scheduling queries.

* **More Settings**

In this dropdown we have three options.

* + Format Query— To format our queries.
  + Query Settings— To change our query settings.
  + Enable SQL Translation— To enable SQL translation.

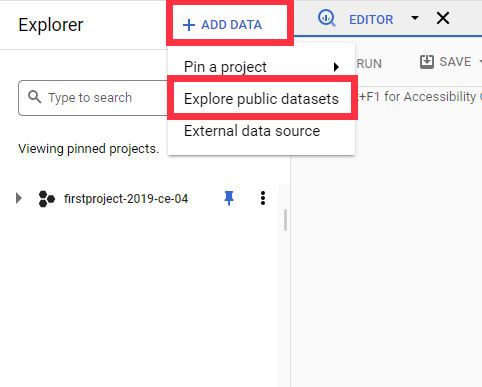
**Exploring Public Datasets**

BigQuery makes public datasets available for us to use and integrate into our applications. Google pays for the storage of these datasets and makes the data available to the public through a project.

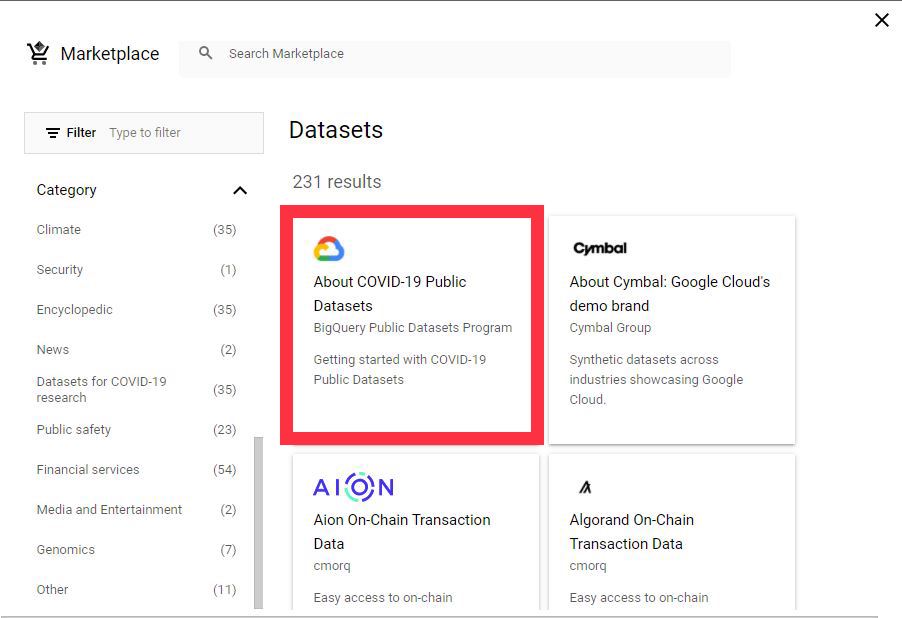
* **Accessing Public datasets**

There are two ways to access public datasets.

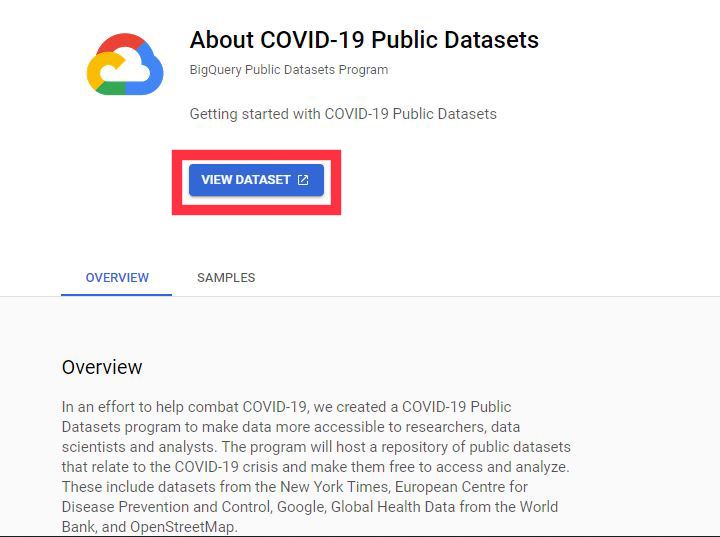
1. **First Method**
2. Click on Add Data and then click on Explore public datasets.



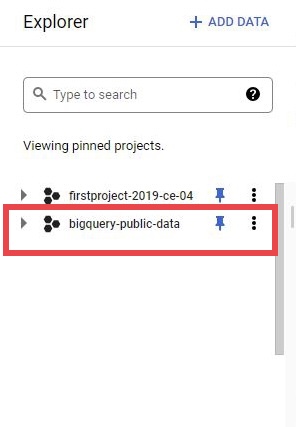
1. All public datasets will be appear. Now click on any dataset.



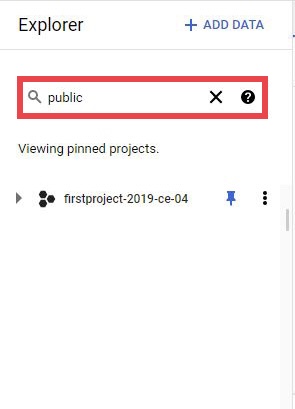
1. Click on View Dataset.



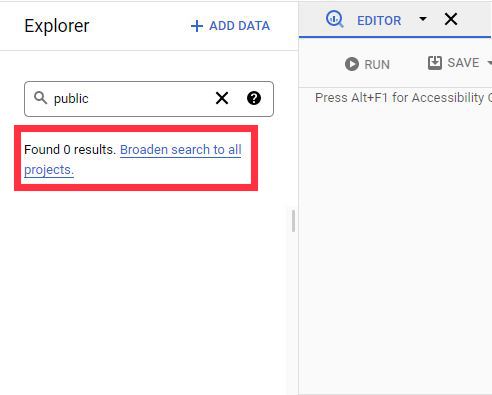
1. Now, we can see all public datasets here. We can also pin this.



1. **Second Method**
2. Write public in search box and Enter.

****

1. Click on Broaden search to all projects.

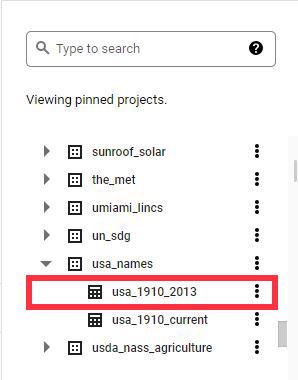
****

1. Now, we can see all public datasets here. We can also pin this.

****

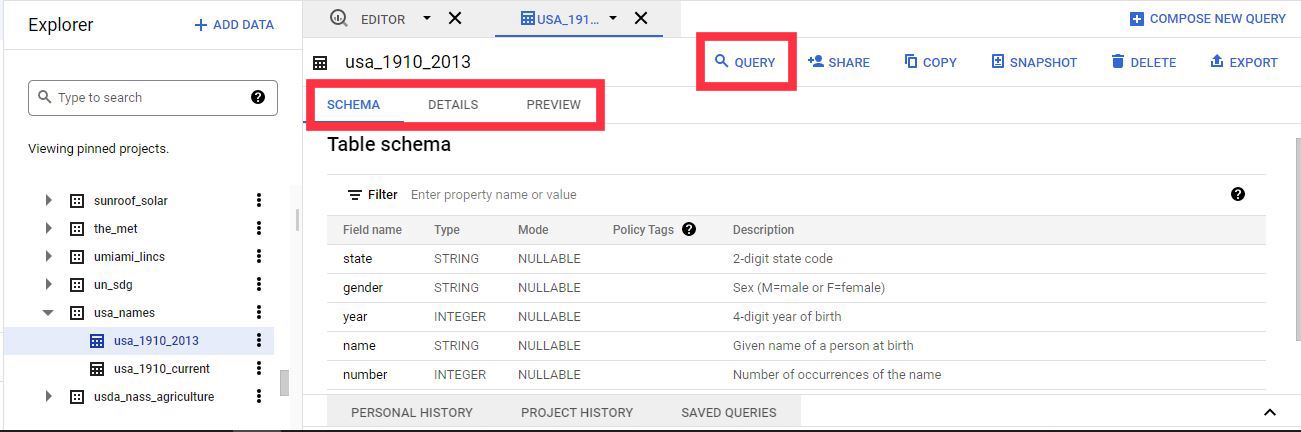
* **Querying Public Datasets**

For querying public datasets, we first search our desired dataset and click on this. I have query the USA Names Data public dataset to determine the most common names in the US between 1910 and 2013.



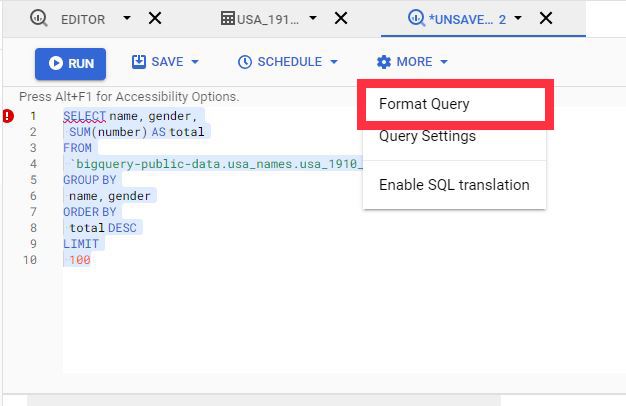
After selecting we can see all the details of the dataset.The schema option tells us what the table structure is. For example field names, tyes, mode and description. The Details option tells

us complete detail of dataset table like Table ID, Table size, storage size, number of rows, creation date, last modified, table expiration, data location and description. The preview option shows the sample data which is available in this particular table.

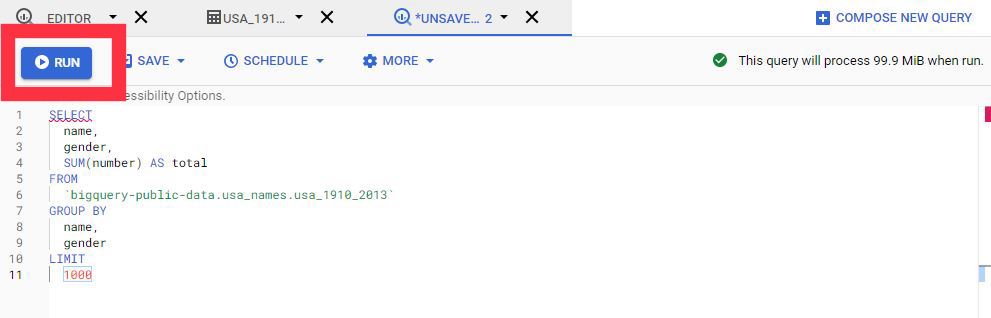


Now, click on Query. We have two options, one is “In new tab” and second is “In split tab”. I

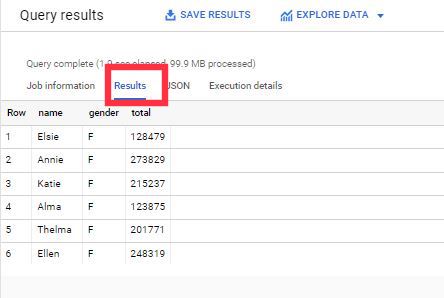
have selected “In new tab”. A new query window appears. Then I write my query to determine the most common names in the US between 1910 and 2013. After this I have formatted my query so that it looks finer.



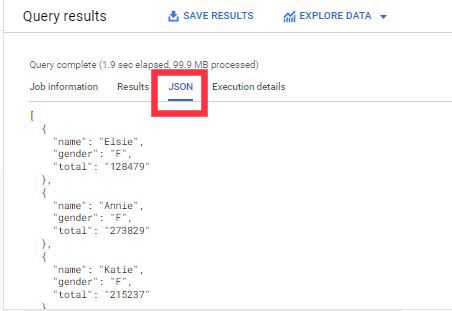
Afer formatting, we can see that it looks finer.



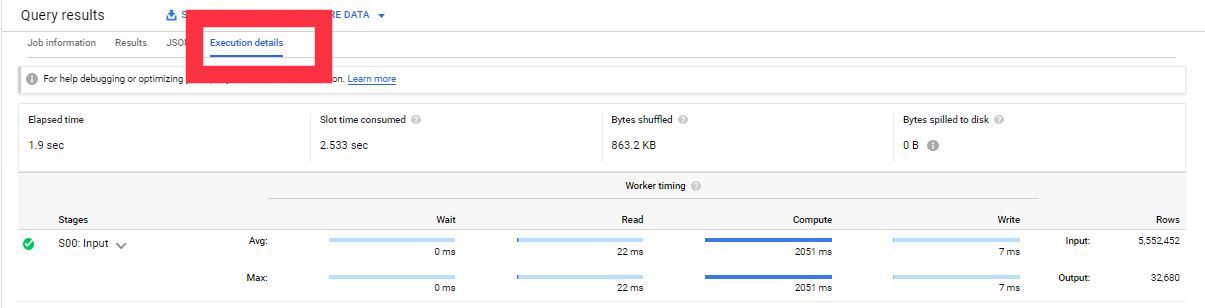
Now, run this query to see results. We can see our result in tabular form. We have three columns name, gender and total.



We can also see our result in JSON format. In this, we have three field-value pairs. Three fields are name, gender and total.



We can also see all the execution details of the data. It shows the execution time of iput and output.

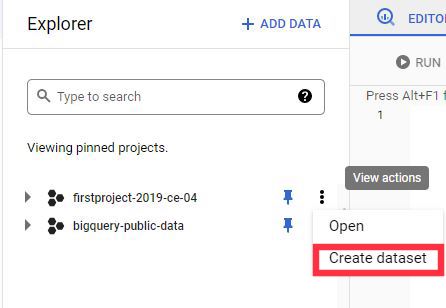


“Each public dataset is stored in a specific location like US or EU. Currently, the BigQuery sample tables are stored in the US multi-region location. When you query a sample table, supply the --location=US flag on the command line, choose US as the processing location in the Cloud Console. Because the sample tables are stored in the US, you cannot write sample table query results to a table in another region, and you cannot join sample tables with tables in another region.”

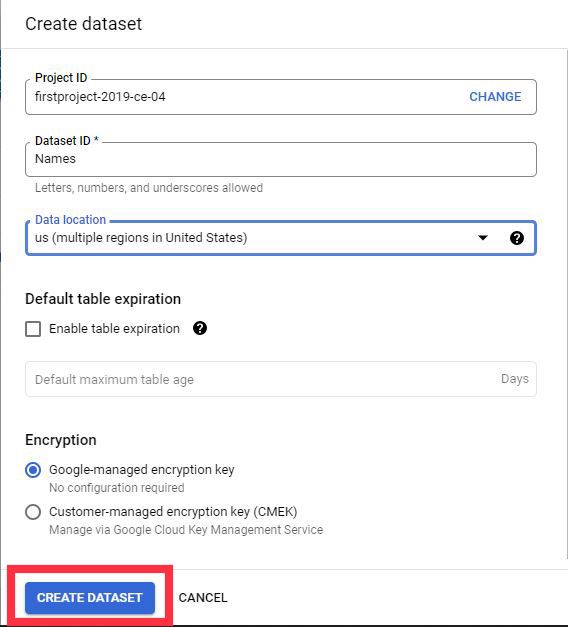
**Exploring GitHub Dataset**

Download dataset from internet or from GitHub. Create dataset in a project and then querying.

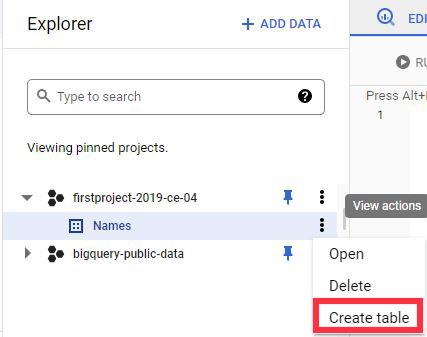
In the project that I have created earlier, click on create dataset.



For creating dataset, we have to fill dataset id and data location. After filling these fields, click on create dataset.



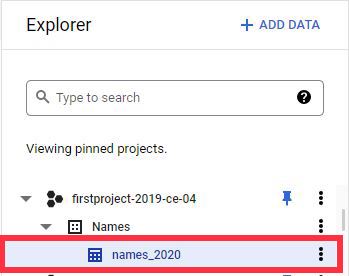
We can see that our dataset created. Click on Open if you want to view the details. Now click on create table for adding table.



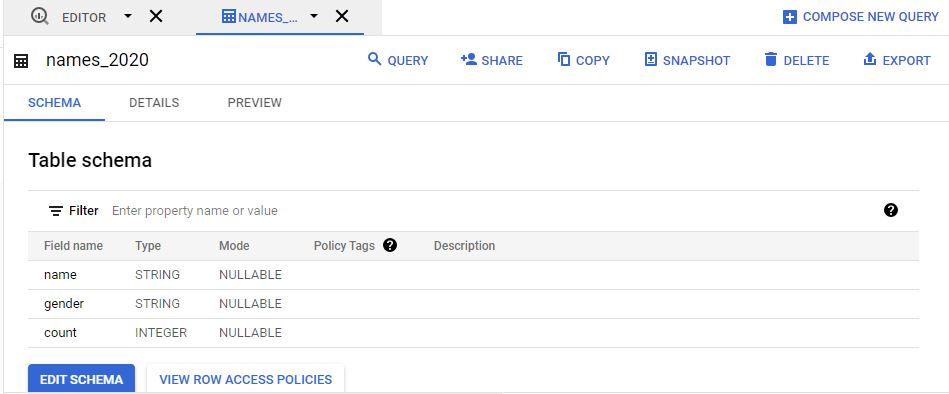
Now, we have to fill these fields to import our dataset. After filling these fields, click on create table.



We can see that our table has been successfully created. This dataset includes the names of both male and female in a year 2020.

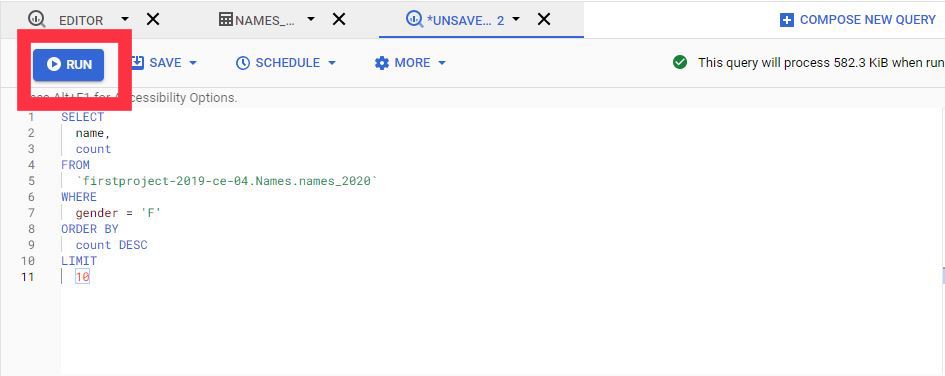


On clicking this table name, we can see the tables fileds. i.e. name, gender and count. We can see the schema, details and preview of the table.

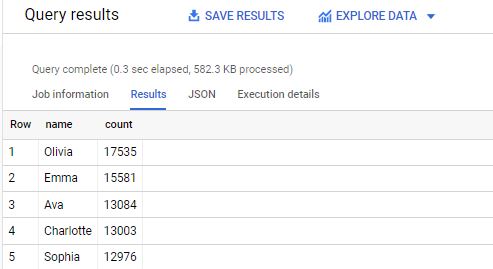


For querying, click on query and then in new tab. This will open the new tab for querying.

I have written this query to show the names of females in the year 2020 and their count. I have set the limit to 10. This shows only top 10 names.



Click on Run to rum this query.



We can see that our desired output. We can also see this result in JSON format or also the execution details of this query.

**Conclusion**

BigQuery is a Cost-effective and highly scalable warehousing solution. We can easily analyze public datasets or datasets that we have created by ourselves. I have successfully running queries of public dataset or dataset that I have downloaded from internet.