# BigQuery - Load and query data with the bq tool

Very Easy to complete

Execute On CLI and can be observed on GUI also

Prerequisites:

Ensure bigquery API enabled

Select project

Check billing for project

Enable cloudshell

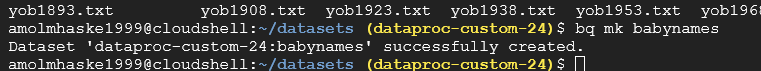
Overview:

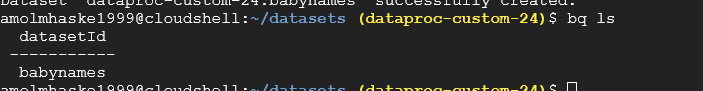
Bq is a tool

Step1 : Prepare the Dataset for processing

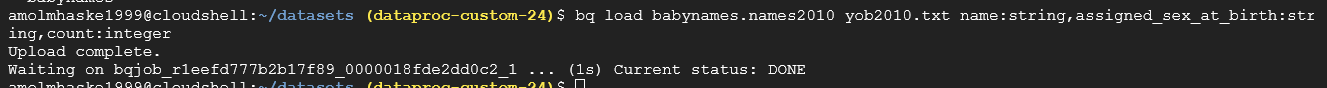
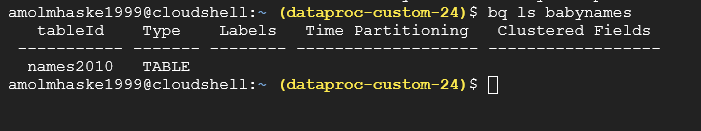
* Create a dir(of your choice) : sudo mkdir –p <dir-path>
* Chang present working dir : cd <dir-path>
* Download the dataset archive “names.zip” on your local machin/where you’re using cloudshell.
* Extract/unzip this file : unzip names.zip
* Open yob2010.txt file. It’s a CSV file
* The file has no header row.
* If u didn’t do above using cloudshell, upload yob2010.txt file to cloudshell using “more(vertical\_three\_dots) -> **Choose Files -> (select** yob2010.txt **file) -> Upload**.” Options flow. Since we’ll be working here.

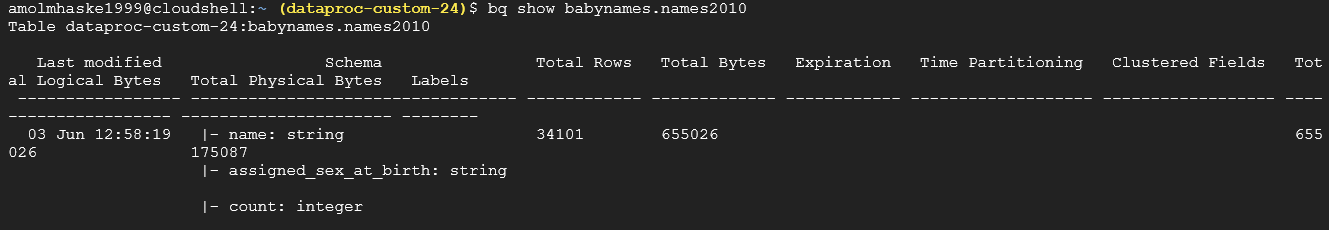
Step 2 – Create the Dataset

* Create dataset using bq tool : bq mk babynames
* 
* Verufy creted dset : bq ls

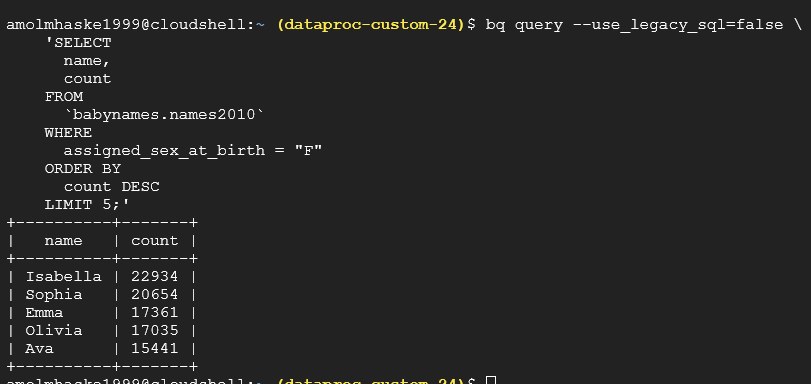


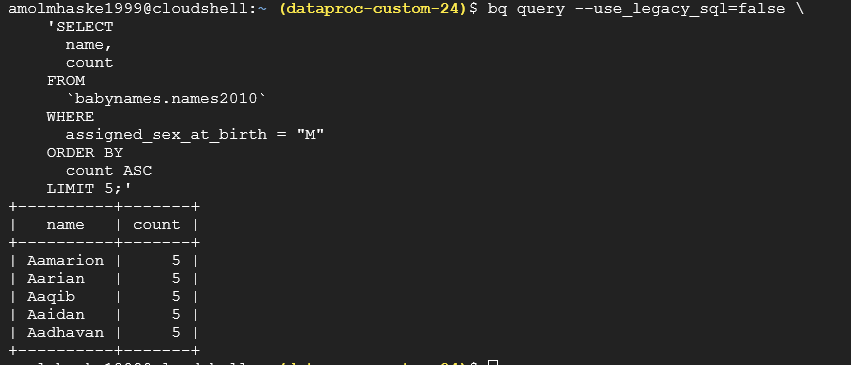
Step 3 – Load the dataset into bigquery table

* We’re creating table from the above csv file
* Make sure to be in the dir where yob2010.txt file exists and bq tool is operable
* bq load babynames.names2010 yob2010.txt name:string,assigned\_sex\_at\_birth:string,count:integer
* 
* Check the loaded table
* bq ls babynames
* 
* Check table schema : bq show babynames.names2010



## Step 4 - Query table data

* Determine the most popular girls' names in the data
* bq query --use\_legacy\_sql=false \  
      'SELECT  
        name,  
        count  
      FROM  
        `babynames.names2010`  
      WHERE  
        assigned\_sex\_at\_birth = "F"  
      ORDER BY  
        count DESC  
      LIMIT 5;'
* 
* Determine the least popular boys' names in the data:
* bq query --use\_legacy\_sql=false \  
      'SELECT  
        name,  
        count  
      FROM  
        `babynames.names2010`  
      WHERE  
        assigned\_sex\_at\_birth = "M"  
      ORDER BY  
        count ASC  
      LIMIT 5;'



Clean UP process [avoid excessive charges]:

* Delete the babynames dataset: bq rm --recursive=true babynames
* C:\Users\HP\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Screenshot 2024-06-06 at 15-56-51 Welcome – dataproc-custom-24 – Google Cloud console.png