Build a Data Warehouse with BigQuery: Challenge Lab

Task 1

Create a table partitioned by date

The starting point for the machine learning model will be the **oxford_policy_tracker** table in the COVID 19 Government Response public dataset which contains details of different actions taken by governments to curb the spread of Covid-19 in their jurisdictions.

Given the fact that there will be models based on a range of time periods, you have to create a dataset and then create a date partitioned version of the **oxford_policy_tracker** table in your newly created dataset, with an expiry time set to **1445** days.

While creating a table, you have also been instructed to exclude the United Kingdom (alpha_3_code=**GBR**), Brazil (alpha_3_code=**BRA**), Canada (alpha_3_code=**CAN**) & the United States of America (alpha_3_code=**USA**) as these will be subject to more in-depth analysis through nation and state specific analysis.

1. Create a new dataset **covid** and create a table **oxford_policy_tracker** in that dataset partitioned by date, with an expiry of **1445** days. The table should initially use the schema defined for the **oxford_policy_tracker** table in the COVID 19 Government Response public dataset .
2. You must also populate the table with the data from the source table for all countries and exclude the United Kingdom (**GBR**), Brazil (**BRA**), Canada (**CAN**) and the United States (**USA**) as instructed above.

1. Open Cloud Shell

Make sure you're in the correct Qwiklabs project and open the Cloud Shell.

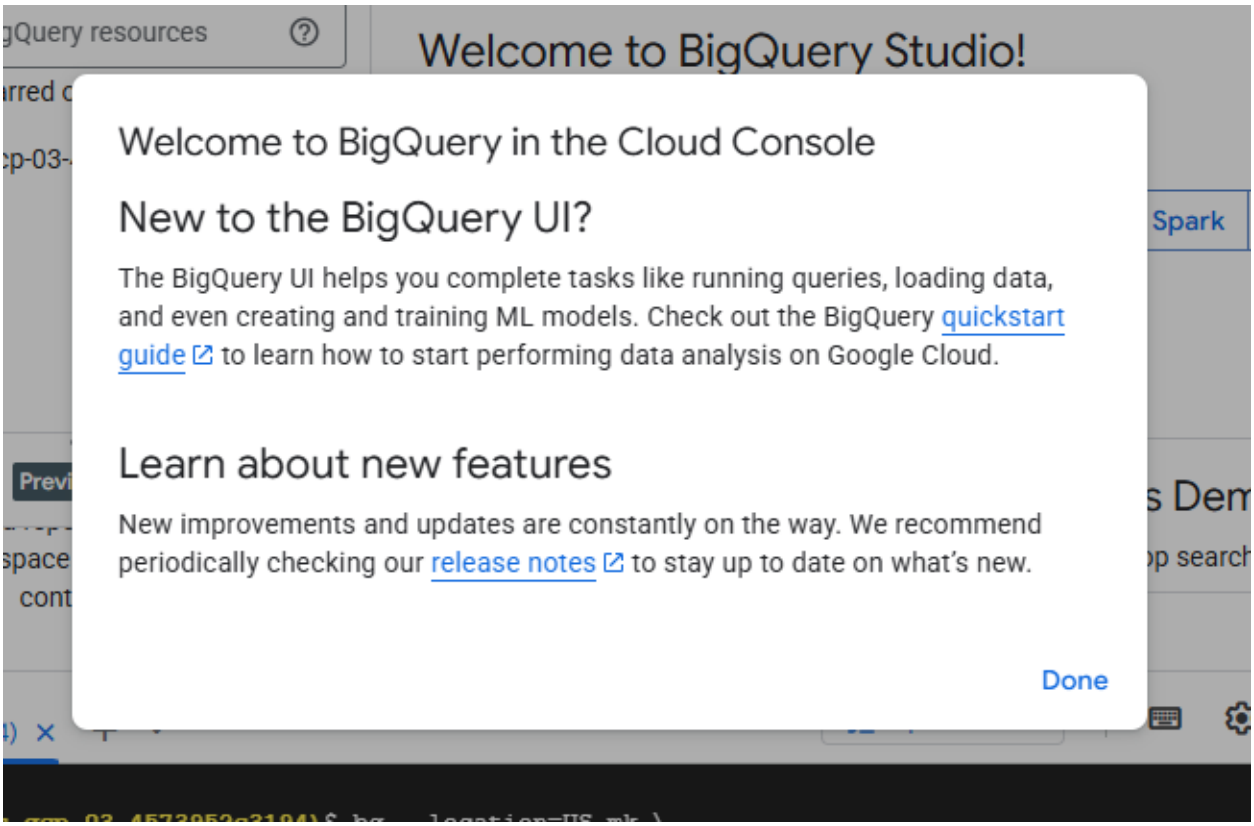
2. Set variables (replace with your dataset/project ID if needed)

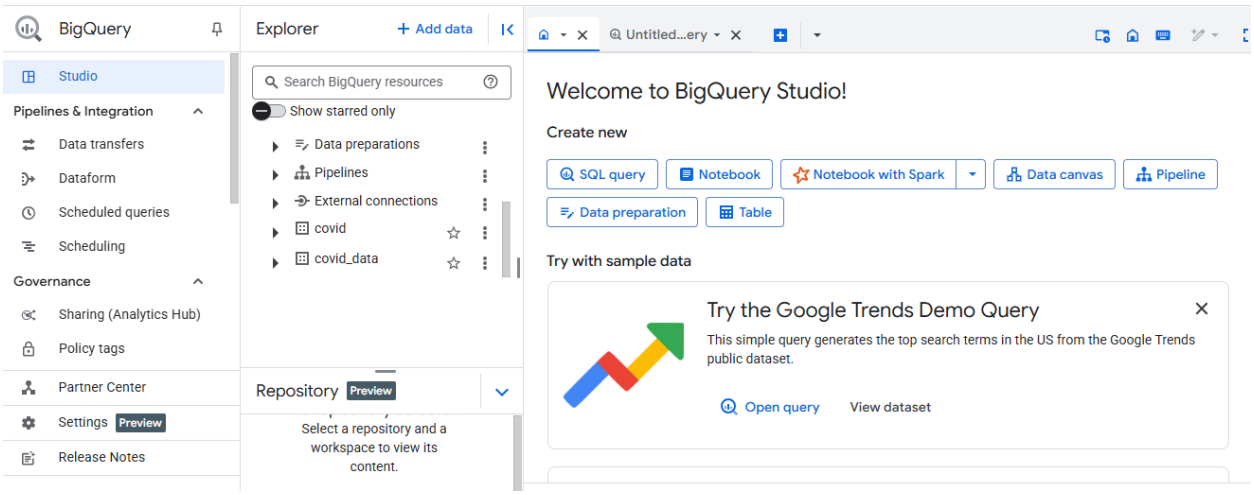
PROJECT_ID=$(gcloud config get-value project) bq --location=US mk --dataset --default_table_expiration 124800000 covid_data

```
student_00_35793e38f044@cloudshell:~ (qwiklabs-gcp-03-4573952c3194)$ bq --location=US mk \
--default_table_expiration 124848000 \
--description "COVID dataset with expiration" \
covid
Dataset 'qwiklabs-gcp-03-4573952c3194:covid' successfully created.
student_00_35793e38f044@cloudshell:~ (qwiklabs-gcp-03-4573952c3194)$
```

3. Create a partitioned table from the public dataset (excluding GBR, BRA, CAN, USA)

bq query --use_legacy_sql=false ' CREATE OR REPLACE TABLE covid_data.oxford_policy_tracker PARTITION BY date OPTIONS (expiration_timestamp = TIMESTAMP_ADD(CURRENT_TIMESTAMP(), INTERVAL 1445 DAY)) AS SELECT * FROM `bigquery-public-data.covid19_govt_response.oxford_policy_tracker` WHERE alpha_3_code NOT IN ("GBR", "BRA", "CAN", "USA")'





Task 2

Create a new table for country area data

In this step, you need to create a copy of **country_names_area** table from **Census Bureau International public dataset** into your dataset provided in the task description.

1. Create a new table '**country_area_data**' within the dataset named as '**covid_data**'. The table should initially use the schema defined for the **country_names_area** table data from the Census Bureau International public dataset.
2. Add the country area data to the '**country_area_data**' table with **country_names_area** table data from the Census Bureau International public dataset.

Copy country_names_area to covid_data.country_area_data

bq query --use_legacy_sql=false ' CREATE TABLE covid_data.country_area_data AS SELECT * FROM `bigquery-public-data.census_bureau_international.country_names_area`'

Task 3

Create a new table for mobility record data

In this step, you need to create a copy of **mobility_report** table from **Google COVID 19 Mobility public dataset** into your dataset provided in the task description.

1. Create a new table '**mobility_data**' within the dataset named as '**covid_data**'. The table should initially use the schema defined for the **mobility_report** table data from the Google COVID 19 Mobility public dataset.
2. Add the mobility record data to the '**mobility_data**' table with data from the Google COVID 19 Mobility public dataset.

Task 3: Copy mobility_report to mobility_data

bq query --use_legacy_sql=false ' CREATE TABLE covid_data.mobility_data AS SELECT * FROM `bigquery-public-data.covid19_google_mobility.mobility_report`'

```
student_00_35793e38f044@cloudshell:~ (qwiklabs-gcp-03-4573952c3194)$ bq query --use_legacy_sql=false '
CREATE TABLE covid_data.mobility_data AS
SELECT * FROM `bigquery-public-data.covid19_google_mobility.mobility_report`'
Waiting on bqjob_r7611adc9b168b48c_00000197817f3e73_1 ... (2s) Current status: DONE
Created qwiklabs-gcp-03-4573952c3194.covid_data.mobility_data

student_00_35793e38f044@cloudshell:~ (qwiklabs-gcp-03-4573952c3194)$
```

Task 4

Delete Null population and country area data from oxford_policy_tracker_by_countries table

In this step, you will have to delete data for countries that do not have population information and country area information.

1. Delete data from the **oxford_policy_tracker_by_countries** table in the **covid_data** dataset where the **population** value is null.
2. Now delete data from the **oxford_policy_tracker_by_countries** table in the **covid_data** dataset where the **country_area** value is null.

Task 4: Delete null population and country_area values

Make sure the table covid_data.oxford_policy_tracker_by_countries already exists (probably from a previous join step or lab-provided). Then run:

bq query --use_legacy_sql=false ' DELETE FROM covid_data.oxford_policy_tracker_by_countries WHERE population IS NULL'

Then:

bq query --use_legacy_sql=false ' DELETE FROM
covid_data.oxford_policy_tracker_by_countries WHERE country_area IS NULL'