Loading data into BigQuery

1 hour 30 minutes No cost

Overview

BigQuery is Google's fully managed, NoOps, low cost analytics database. With BigQuery you can query terabytes and terabytes of data without having any infrastructure to manage or needing a database administrator. BigQuery uses SQL and can take advantage of the pay-as-you-go model. BigQuery allows you to focus on analyzing data to find meaningful insights.

In this lab you will ingest subsets of the NYC taxi trips data into tables inside of BigQuery.

What you'll learn

- Loading data into BigQuery from various sources
- Loading data into BigQuery using the CLI and Console
- Using DDL to create tables

Setup

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

1. Sign in to Qwiklabs using an **incognito window**.
2. Note the lab's access time (for example, `1:15:00`), and make sure you can finish within that time.
   There is no pause feature. You can restart if needed, but you have to start at the beginning.
3. When ready, click **Start lab**.
4. Note your lab credentials (**Username** and **Password**). You will use them to sign in to the Google Cloud Console.
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 receive errors or **incur charges**.
7. Accept the terms and skip the recovery resource page.

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

Open BigQuery Console

1. 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.

2. Click **Done**.

Task 1. Create a new dataset to store tables

1. To create a dataset, click on the **View actions** icon (the three vertical dots) next to your project ID and select **Create dataset**.
2. Next, name your Dataset ID **nyctaxi** and leave all other options at their default values, and then click **Create dataset**.

You'll now see the nyctaxi dataset under your project name.

Click **Check my progress** to verify the objective. Creating a dataset to store new tables

Task 2. Ingest a new dataset from a CSV

In this section, you will load a local CSV into a BigQuery table.

1. Download a subset of the NYC taxi 2018 trips data locally onto your computer from this link.
2. In the BigQuery Console, Select the **nyctaxi** dataset then click **Create Table**

Specify the below table options:

Source:

- Create table from: **Upload**
- Choose File: **select the file you downloaded locally earlier**
- File format: **CSV**

Destination:

- Table name: **2018trips** Leave all other settings at default.

Schema:

- Check **Auto Detect** (**tip**: Not seeing the checkbox? Ensure the file format is CSV and not Avro)

Advanced Options

- Leave at default values

Click **Create Table**.

3. You should now see the **2018trips** table below the nyctaxi dataset.

Select the 2018trips table and view **details**:

4. Select **Preview** and confirm all columns have been loaded (sampled below):

You have successfully loaded a CSV file into a new BigQuery table.

Running SQL Queries

Next, practice with a basic query on the 2018trips table.

1. In the Query Editor, write a query to list the top 5 most expensive trips of the year:

#standardSQL SELECT * FROM nyctaxi.2018trips ORDER BY fare_amount DESC LIMIT 5

Click **Check my progress** to verify the objective. Ingest a new Dataset from a CSV

Task 3. Ingest a new dataset from Google Cloud Storage

Now, let's try to load another subset of the same 2018 trip data that is available on Cloud Storage. And this time, let's use the CLI tool to do it.

1. In your Cloud Shell, run the following command :

bq load \ --source_format=CSV \ --autodetect \ --noreplace \ nyctaxi.2018trips \ gs://cloud-training/OCBL013/nyc_tlc_yellow_trips_2018_subset_2.csv **Note:** With the above load job, you are specifying that this subset is to be appended to the existing 2018trips table that you created above.

2. When the load job is complete, you will get a confirmation on the screen.
3. Back on your BigQuery console, select the 2018trips table and view **details**. Confirm that the row count has now almost doubled.
4. You may want to run the same query like earlier to see if the top 5 most expensive trips have changed.

Click **Check my progress** to verify the objective. Ingest a dataset from google cloud storage

Task 4. Create tables from other tables with DDL

The 2018trips table now has trips from throughout the year. What if you were only interested in January trips? For the purpose of this lab, we will keep it simple and focus only on pickup date and time. Let's use DDL to extract this data and store it in another table

1. In the Query Editor, run the following CREATE TABLE command :

#standardSQL CREATE TABLE nyctaxi.january_trips AS SELECT * FROM nyctaxi.2018trips WHERE EXTRACT(Month FROM pickup_datetime)=1;

2. Now run the below query in your Query Editor find the longest distance traveled in the month of January:

#standardSQL SELECT * FROM nyctaxi.january_trips ORDER BY trip_distance DESC LIMIT 1

Click **Check my progress** to verify the objective. Create tables from other tables with DDL

Congratulations!

You've successfully created a new dataset and ingested data into BigQuery from CSV, Google Cloud Storage, and other BigQuery tables.

End your lab

When you have completed your lab, click **End Lab**. Google Cloud Skills Boost 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.

Manual Last Updated: July 27, 2022

Lab Last Tested: July 15, 2022

- Overview
- Setup
- Task 1. Create a new dataset to store tables
- Task 2. Ingest a new dataset from a CSV
- Task 3. Ingest a new dataset from Google Cloud Storage
- Task 4. Create tables from other tables with DDL
- Congratulations!
- End your lab