QLabs Overview

BA770 Lab Session

Questrom School of Business, Boston University

August 7, 2019

QLabs Overview

- You may follow the bullet points in this slides to review a QLab after you complete it.
- Key takeaways are marked with an asterisk before index.

Lecture1, QLabA
A Tour of Qwiklabs and the Google Cloud Platform

a) Gain an overall understanding of Qwiklabs platform and identify key features of a lab environment.

b) Access the GCP console with (temporary) specific credentials.

⋆ c) Know the definition of GCP projects.

⋆ d) Use the GCP navigation menu to identify types of GCP services.

e) Learn about primitive roles and use the Cloud IAM service to inspect actions available to specific users.

⋆ f) Learn basic Cloud Shell commands; run commands like touch, nano, and cat to create, edit, and output the content of files.

g) Learn about the API library and examine its chief features.

Lecture1, QLabA, Tips

a) Make sure you're using the temporary account to log in!

b) If you find something wrong or you get stuck in a session, don't be nervous - end the current session and open a new one. Redoing a lab will not influence your QLab grades.

c) Be careful with all keyboard and/or mouse operations when you are on the lab page. Do not click the End Lab button until you have completed all the tasks.

d) Learn more about applications provided by GCP, including virtual machine, storage, database, etc.

Lecture1, QLabB
Getting Started with Cloud Shell & gcloud

a) Gain an understanding of Google Cloud Shell, including definition, function, feature, etc.

⋆ b) Be familiar with **cd** command and **vi** editor.

c) Use gcloud commands to view configurations.

⋆ d) Use gsutil commands to manage Cloud Storage resources; know how to create a bucket and copy an existing file to a bucket.

Lecture1, QLabB, Tips

a) Bucket names are universally unique, so avoid using names like 'my_bucket' 'test_bucket'; otherwise you may receive 'ServiceException: 409 Bucket xxx already exists'.

b) You may use '**Ctrl+C**' to exit the current command.

c) Find **vi** documentation here.

d) Refer to documentation (you could just google it) to know more about **gcloud** and **gsutil** commands.

Lecture1, QLabC*
Creating a Virtual Machine

⋆ a) Create a virtual machine with 1) the GCP Console, 2) **gcloud** command line.

b) Check existing instances in Navigation menu - Compute Engine - VM instances.

⋆ c) Access the virtual machine by 1) launching a SSH client directly from browser, or 2) SSH'ing into the instance using gcloud.

Lecture1, QLabC*, Tips

a) Be aware of the window you're typing in:
   - When you are SSH'ing into an instance with a prompt window, make sure you execute the commands in this window.
   - If you are using **gcloud** commands, make sure the commands are running in the Cloud Shell.

b) If you get root access to your instance by mistake, press '**Ctrl+D**' to exit root user.

c) Learn more about SSH here.

d) Learn more about **sudo** command here.

e) Learn more about commonly used commands in GCP here.

Lecture2, QLabD
Introduction to SQL for BigQuery and Cloud SQL

- ⋆ a) Understand the relationship among project, database, and table.
- ⋆ b) Get familiar with BigQuery console, e.g. loading databases and tables into BigQuery.
- ⋆ c) Use SELECT, FROM, WHERE, COUNT, GROUP BY, AS, and ORDER BY keywords to fetch meaningful data from datasets.
- d) Export query results to local repository and upload them to Cloud Storage bucket.
- e) Create a new Cloud SQL instance and load files from Cloud Storage bucket as new tables.
- ⋆ f) Know basic knowledge of keywords CREATE, DELETE, INSERT INTO, and UNION. Be able to run queries in Cloud SQL using Cloud Shell.

Lecture2, QLabD, Tips

a) Make sure you have switched to the temporary account, and you are under the temporary project. Double check from time to time.

b) Leave enough time for file importing, instance creation and connecting, etc. You are strongly advised to read all instructions before starting the lab.

c) In Query editor, you could press '**Ctrl+Enter**' instead of clicking '**run**'.

d) Remember to rename the two csv documents.

e) After you enter '**gcloud sql connect qwiklabs-demo –user=root**' in Cloud Shell and wait for a while, you will see '**Connecting to database with SQL user [root]. Enter password:**'. Be aware that there's no flash cursor when you are typing. Enter your password and press 'Enter'.

f) When you are running SQL queries using Cloud Shell Command Line, remember to end each query with a semicolon '**;**'.

g) When importing files to tables in SQL instance, click Browse, and then double click the bucket name to find your files.

Lecture2, QLabD, SQL Keywords/Functions Summary

- **SELECT**: Specify the fields you want to pull from the dataset.
- **FROM**: Specify what table or tables to pull data from.
- **GROUP BY**: Aggregate result-set rows that share common criteria and return all of the unique entries found for such criteria.
- **COUNT**: Return the number of rows that share the same criteria.
- **AS**: Create an alias of a table or column.
- **ORDER BY**: Sort the returned data from a query in ascending or descending order based on a specified criteria or column value.
- **CREATE**: Create new databases or tables.
- **DELETE**: Delete existing databases or tables.
- **INSERT INTO**: Insert values into tables.
- **UNION**: Combine the output of two or more queries into a result set.

Lecture2, QLabE*
Using BigQuery in the GCP Console

⋆ a) Query a public dataset using aggregation functions.

⋆ b) Create a dataset in your project and load the data into a table.

c) Query your own table using GROUP BY and ORDER BY keywords.

Lecture2, QLabE*, Tips

a) After you click '**VIEW DATASET**', make sure you are still using the temporary google account and under the temporary project.

b) '**CREATE DATASET**' button is in blue font, right under the green check mark icon if your Query editor is not hidden. '**CREATE TABLE**' button is in a similar place.

Lecture3, QLabF
Analyzing Natality Data Using Datalab and BigQuery

- ⋆ a) Launch Cloud Datalab with Cloud Shell command.
- ⋆ b) Enable BigQuery client in Datalab notebook, execute SQL queries, and generate pandas dataframes.
- c) Create charts in Datalab.

Lecture3, QLabF, Tips

a) Be patient when waiting for the Datalab instance to launch. Remember to check the progress and to press Enter when responses are required.

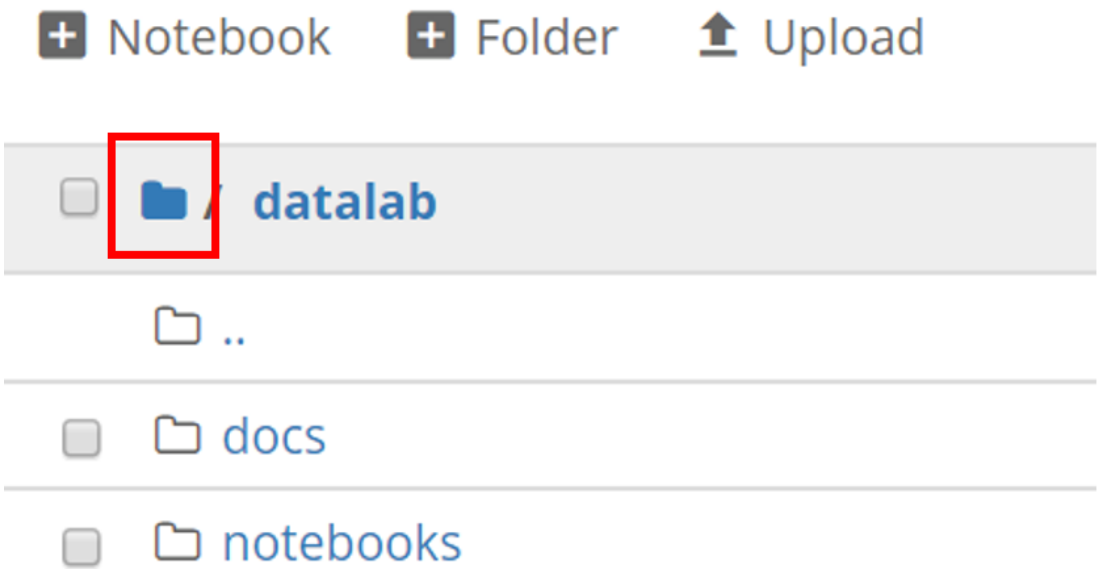
b) BigQuery result that you may refer to:

Row	plurality	num_babies	ave_weight
1	2	507706	5.16662858551256
2	3	27697	3.7188113817178317
3	5	325	2.625698693765901
4	1	15736332	7.336915793502335
5	4	1846	2.842509406912898

Lecture3, QLabF, Tips

c) A convenient way to open Datalab is to click the link 'http://localhost:8081/'. You can find the link in the output, right after 'Waiting for Datalab to be reachable at'. In this way you don't need to change port manually.

d) If you meet an error saying 'Creating Notebook Failed', click the solid folder icon to enter folder 'datalab' again, and then try starting a new notebook again.



Lecture3, QLabF, Tips

e) If you happen to lose connection to your Datalab instance (for example the error below):

Error: Could not connect to Cloud Shell on port 8081.

Ensure your server is listening on port 8081 and try again.

Dont worry, enter the following command in Cloud Shell:

datalab connect babyweight

Press Enter and wait for a short moment, and then youll be able to open Datalab again.

f) In case you meet errors when running cells in the notebook, shutdown the notebook by clicking Running Sessions in the upper right corner and open it again.

Lecture3, QLabF, Tips

g) You may use the following command in Cloud Shell to remove the passphrase of the SSH key:

rm .ssh/google_compute_engine*

You may press '**Tab**' to autocomplete the command.

h) Here is a cheat sheet of Notebook shortcuts. You may use these shortcuts to improve efficiency.

i) Download the notebooks in your datalab instance to your laptop from time to time, in case you may lose the instance by accident.

j) Use CAST, CONCAT and EXTRACT functions to create and manipulate SQL timestamps.

Lecture3, QLabG
Weather Data in BigQuery

a) Get a sense of GCP's great benefits for big data processing.

⋆ b) Combine and run analytics on multiple datasets.

⋆ c) Understand nested query statements with multiple functions; get familiar with table and column alias.

Lecture3, QLabG, Tips

a) COUNT(1): equivalent to COUNT(*), which returns all rows whether they are null or not null.
COUNT(Column): return all non-null rows.

b) Default JOIN type in BigQuery: INNER JOIN

c) Documentation for reference:
   1) JOIN
   2) Aggregate functions (e.g. AVG)
   3) Date functions (e.g. DATE)
   4) Mathematical functions (e.g. ABS)
   5) Statistical Aggregate Functions (e.g. CORR)
   6) Subquery

Lecture3, QLabH*
Exploring NCAA Data with BigQuery

a) Review how to view a table's basic information with Schema, Details, Preview panels.

b) Write and execute queries using WHERE, GROUP BY, and ORDER BY to answer specific questions.