Introduction to BigQuery (A GCP data warehouse)

Sayak Paul (Data Science Instructor at DataCamp)
GDG Kolkata Cloud Study Jam at Academy of Technology
12/04/2019

Outline

- ★ What is BigQuery and why you might need it?
- ★ Using BigQuery in various forms
- ★ Exploring the BigQuery Web UI
- ★ Adding a public dataset in BigQuery Web UI
- ★ Running simple queries (but at scale) in BigQuery
- ★ Q & A

About me

Courtesy: Snehangshu

- ➔ Data Science Instructor at DataCamp
- ➔ An avid Data Science enthusiast
- ➔ LinkedIn: https://www.linkedin.com/in/sayak-paul/
- ➔ GitHub: https://github.com/sayakpaul

BigQuery: Food or what!?

Here's what Google says -

"A *fast*, *highly scalable*, *cost-effective*, and fully managed enterprise data warehouse for analytics." *

* Source: https://cloud.google.com/bigquery/

Why BigQuery

- It is blazingly *fast*!

```
SELECT * FROM `bigquery-public-data.london_bicycles.cycle_hire`;
```

Query complete (24.4 sec elapsed, 2.6 GB processed)

Continued

Why BigQuery

- It is highly *scalable* (and it is no joke)!

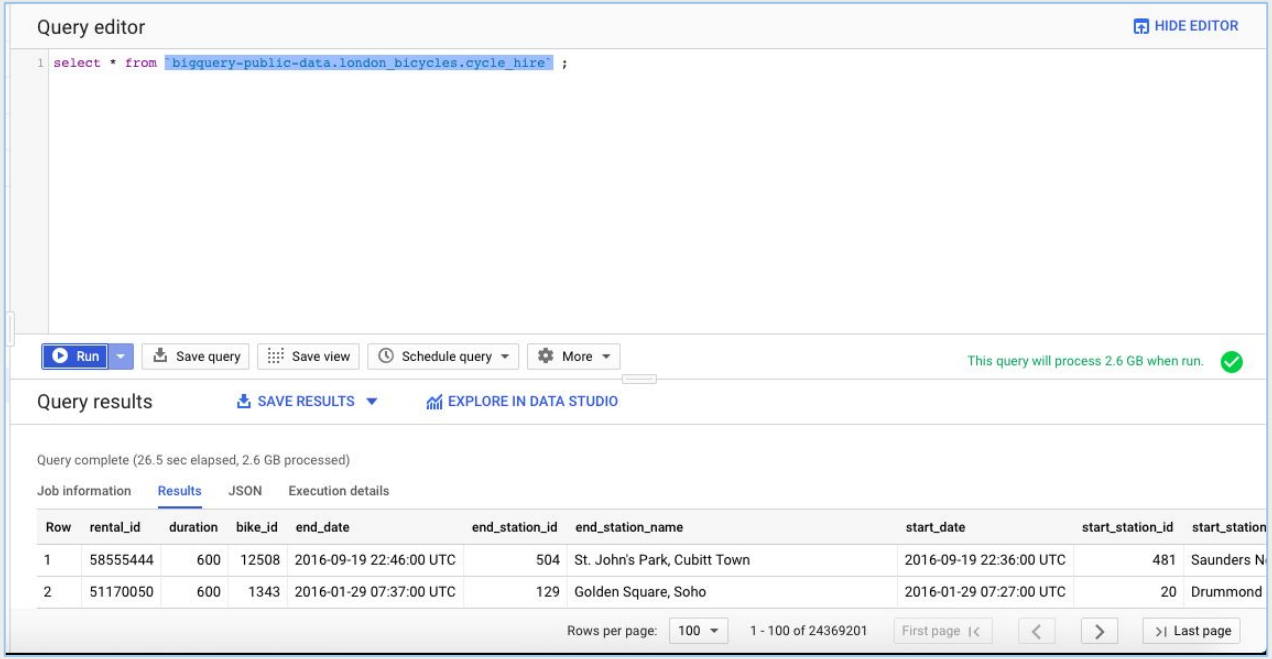
Not just GBs, easily extendible to PBs!

Continued

Why BigQuery □

- It is ridiculously *cost-effective* (and it is no joke)!
- Well segregated cost model providing support for enterprises of varied sizes.
- Refer https://cloud.google.com/bigquery/pricing for a detailed overview.

BigQuery: One product, multiple options (1)



A classic Web UI in form a query editor

Continued

BigQuery: One product, multiple options (2)

```
bq query "SELECT word FROM publicdata:samples.shakespeare WHERE word = 'huzzah' IGNORE CASE"
```

```
Waiting on job_e19 ... (4s) Current status: DONE
$
```

BigQuery command-line interface (bq)

Images taken from https://cloud.google.com/bigquery/docs/quickstarts/quickstart-command-line

Continued

BigQuery: One product, multiple options 💪 (3) [Unbelievable one]

```python
 # from google.cloud import bigquery
# client = bigquery.Client()

query = (
    "SELECT name FROM `bigquery-public-data.usa_names.usa_1910_2013` "
    'WHERE state = "TX" '
    "LIMIT 100"
)
query_job = client.query(
    query,
    # Location must match that of the dataset(s) referenced in the query.
    location="US",
)  # API request - starts the query

for row in query_job:  # API request - fetches results
    # Row values can be accessed by field name or index
    assert row[0] == row.name == row["name"]
    print(row)
```

Calling the BigQuery REST API from Python code

Code taken from https://cloud.google.com/bigquery/docs/reference/libraries#client-libraries-resources-python

BigQuery: DIY (C'mon, it will be fun!)

QwikLabs
waiting!

- Exploring the BigQuery Web UI
- Adding a public dataset in BigQuery Web UI
- Running simple queries (but at scale) in BigQuery Web UI

Questions?

References

Google BigQuery documentation

Introduction to SQL for BigQuery and Cloud SQL (a lab by QwikLabs)

```
SELECT substring('Thank you!' from 1 for
10);
```