



# Introduction to BigQuery (A GCP data warehouse)

Sayak Paul (Data Science Instructor at DataCamp)

Google Cloud Next '19 Extended - Kolkata

20/04/2019

# 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



- Data Science Instructor at [DataCamp](#)
- An avid Data Science enthusiast
- LinkedIn: [www.linkedin.com/in/sayak-paul/](https://www.linkedin.com/in/sayak-paul/)
- GitHub: <https://github.com/sayakpaul>

Courtesy: [Snehangshu](#)



# 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)!

Query complete (24.4 sec elapsed, 2.6 GB processed)

Not just GBs, easily extendible to PBs!

# 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 🐼

Query editor HIDE EDITOR

```
1 select * from bigquery-public-data.london_bicycles.cycle_hire ;
```

Run Save query Save view Schedule query More This query will process 2.6 GB when run. ✓

Query results SAVE RESULTS EXPLORE IN DATA STUDIO

Query complete (26.5 sec elapsed, 2.6 GB processed)

Job information Results JSON Execution details

Row	rental_id	duration	bike_id	end_date	end_station_id	end_station_name	start_date	start_station_id	start_station
1	58555444	600	12508	2016-09-19 22:46:00 UTC	504	St. John's Park, Cubitt Town	2016-09-19 22:36:00 UTC	481	Saunders N
2	51170050	600	1343	2016-01-29 07:37:00 UTC	129	Golden Square, Soho	2016-01-29 07:27:00 UTC	20	Drummond

Rows per page: 100 1 - 100 of 24369201 First page < > Last page

A classic Web UI in form a query editor

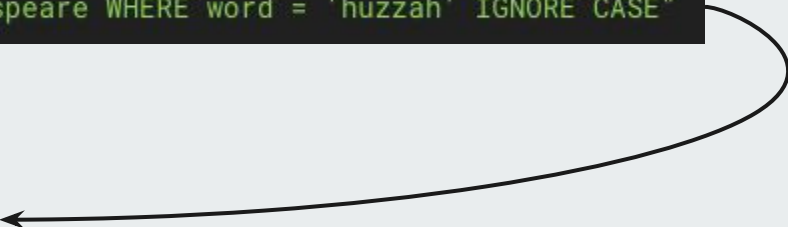
Continued



# BigQuery: One product, multiple options 🐙

```
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)

# BigQuery: One product, multiple options 🐼

```
# 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

Head over to  
<https://bit.ly/BigQueryNext>

# 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);



Connect with me via [LinkedIn](#)