

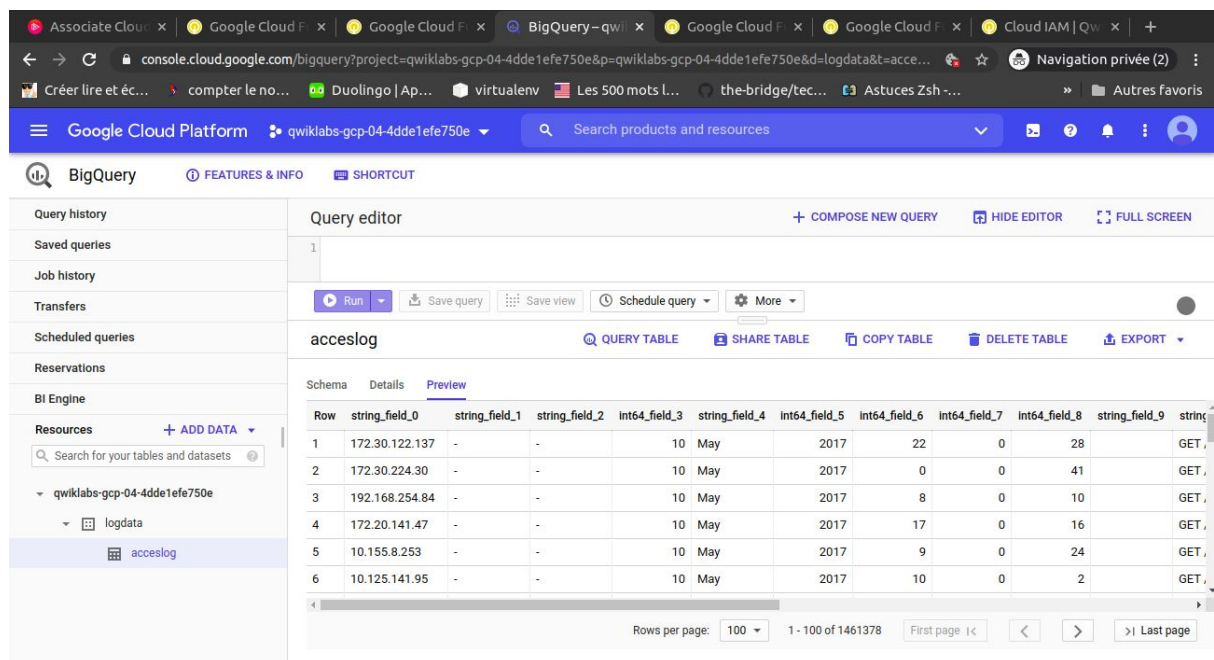
Getting Started with BigQuery

Lab Objectives:

- Load data from Cloud Storage into BigQuery.
- Perform a query on the data in BigQuery.

This lab was done in two parts:

First, we created a Dataset in the BigQuery of our project. This is very important because all queries must be done on a dataset. In our dataset, we created a table and loaded data from cloud storage into it.



The screenshot shows the Google Cloud Platform BigQuery console. The left sidebar contains navigation links for Query history, Saved queries, Job history, Transfers, Scheduled queries, Reservations, and BI Engine. The main area displays the 'acceslog' table in the 'logdata' dataset. The table has 10 columns: string_field_0, string_field_1, string_field_2, int64_field_3, string_field_4, int64_field_5, int64_field_6, int64_field_7, int64_field_8, and string_field_9. The first six rows of data are visible, showing IP addresses, timestamps, and various numerical and string values.

Row	string_field_0	string_field_1	string_field_2	int64_field_3	string_field_4	int64_field_5	int64_field_6	int64_field_7	int64_field_8	string_field_9
1	172.30.122.137	-	-	10	May	2017	22	0	28	GET
2	172.30.224.30	-	-	10	May	2017	0	0	41	GET
3	192.168.254.84	-	-	10	May	2017	8	0	10	GET
4	172.20.141.47	-	-	10	May	2017	17	0	16	GET
5	10.155.8.253	-	-	10	May	2017	9	0	24	GET
6	10.125.141.95	-	-	10	May	2017	10	0	2	GET

[Dataset]

Second, we made queries on our loaded data. We made query from the GCP console:

The screenshot shows the Google Cloud Platform BigQuery console. On the left is a sidebar with navigation links like 'Query history', 'Saved queries', 'Job history', etc. The main area is split into a 'Query editor' and 'Query results'.

Query editor: Contains a SQL query:


```
1 select int64_field_6 as hour, count(*) as hitcount from logdata2.accesslog
2 group by hour
3 order by hour
```

 Below the query is a 'Valid.' status and buttons for 'Run', 'Save query', 'Save view', 'Schedule query', and 'More'. A green message states: 'This query will process 11.1 MB when run.'

Query results: Shows the execution status 'Query complete (0.4 sec elapsed, 11.1 MB processed)'. It has tabs for 'Job information', 'Results' (selected), 'JSON', and 'Execution details'. Below is a table with 4 rows of results:

Row	hour	hitcount
1	0	26983
2	1	12287
3	2	8824
4	3	6607

At the bottom right, it shows 'Rows per page: 100' and '1 - 24 of 24'.

[First query]

This query shows the number of queries made per hour (0h to 23h). We were able to deduce from this request that the largest request is made at 9 a.m.(219769 requests) and a low number of requests is noted at 3 a.m.(6607 requests)

The second query, made from the cloud shell, showed the type of requests carried out and their number. Thanks to this request, we deduced that the most performed request is **GET /store HTTP/1.0** (337293 requests) and the least performed one is **GET /favicon.ico HTTP/1.0** (55845).

The screenshot shows a Google Cloud Shell terminal window. The prompt is '(wikilabs-gcp-04-4dde1efe750e)'. The user has entered a BigQuery query:


```
student_04_ae58a6f12dc4@cloudshell:~ (wikilabs-gcp-04-4dde1efe750e)$ bq query "select string_field_10 as request, count(*) as requestcount from logdata.accesslog group by request order by requestcount desc"
```

 The output shows the status 'Waiting on bqjob_r5ddfc24ec3906fec_000001747c472bf1_1 ... (0s) Current status: DONE' followed by a table of results:

request	requestcount
GET /store HTTP/1.0	337293
GET /index.html HTTP/1.0	336193
GET /products HTTP/1.0	280937
GET /services HTTP/1.0	169090
GET /products/desserttoppings HTTP/1.0	56589
GET /products/floorwaxes HTTP/1.0	56451
GET /careers HTTP/1.0	56412
GET /services/turnipwinding HTTP/1.0	56401
GET /services/spacetravel HTTP/1.0	56176
GET /favicon.ico HTTP/1.0	55845

The terminal prompt is now '(wikilabs-gcp-04-4dde1efe750e)\$'.

[Query2]