

BigQuery Assignment

1. Screenshot of query validator costs for each of the two queries.

Query 1

The screenshot shows the Google Cloud Platform BigQuery interface for the 'Vrushali Banda trail Project'. The left sidebar contains navigation links: Query history, Saved queries, Job history, Transfers, Scheduled queries, Reservations, BI Engine, and Resources. The Resources section is expanded, showing the 'vrushali-banda-trail-project' and its datasets: 'BAN623_Vrushali' (containing 'Listings' and 'sales' tables) and 'gdelt-bq'. The main area is the 'Query editor', which displays a SQL query:

```
1 select sum(qtysold)
2 from vrushali-banda-trail-project.BAN623_Vrushali.sales as S1, vrushali-banda-trail-project.BAN623_Vrushali.Listings as L1
3 where S1.listid = L1.listid
4 and L1.listtime > '2008-12-26';
5
6
```

Below the query editor, a green checkmark indicates the query is 'Valid'. A 'Run' button is visible, along with options to 'Save query', 'Save view', 'Schedule query', and 'More'. A status message at the bottom right states: 'This query will process 4.2 MB when run.' with a green checkmark.

The 'Query results' section shows the query is complete (0.4 sec elapsed, 4.2 MB processed). The 'Job information' tab is selected, displaying the following details:

Job ID	vrushali-banda-trail-project:US.bqjob_35b67ff3_17186bd8577
User	vbanda2@horizon.csueastbay.edu
Location	United States (US)
Creation time	Apr 16, 2020, 11:10:46 PM
Start time	Apr 16, 2020, 11:10:46 PM
End time	Apr 16, 2020, 11:10:47 PM
Duration	0.4 sec
Bytes processed	4.18 MB
Bytes billed	20 MB
Job priority	INTERACTIVE
Destination table	Temporary table
Use legacy SQL	false

Query 2

The screenshot shows the Google Cloud Platform BigQuery interface for the 'Vrushali Banda trail Project'. The left sidebar is identical to the first screenshot. The main area is the 'Query editor', which displays a SQL query:

```
1 select sum(qtysold)
2 from vrushali-banda-trail-project.BAN623_Vrushali.sales as S1
3 where S1.saletime > '2008-12-25'
4 and S1.listid in (select listid from vrushali-banda-trail-project.BAN623_Vrushali.Listings as L1
5 where L1.listtime > '2008-12-26');
6
```

Below the query editor, a green checkmark indicates the query is 'Valid'. A 'Run' button is visible, along with options to 'Save query', 'Save view', 'Schedule query', and 'More'. A status message at the bottom right states: 'This query will process 5.2 MB when run.' with a green checkmark.

The 'Query results' section shows the query is complete (0.6 sec elapsed, 5.2 MB processed). The 'Job information' tab is selected, displaying the following details:

Job ID	vrushali-banda-trail-project:US.bqjob_5e57f395_1718695a43a
User	vbanda2@horizon.csueastbay.edu
Location	United States (US)
Creation time	Apr 16, 2020, 11:04:58 PM
Start time	Apr 16, 2020, 11:04:59 PM
End time	Apr 16, 2020, 11:05:00 PM
Duration	0.6 sec
Bytes processed	5.21 MB
Bytes billed	20 MB
Job priority	INTERACTIVE
Destination table	Temporary table
Use legacy SQL	false

2. A brief discussion comparing the two queries in terms of their expected output and query costs in terms of data processed (GB or TB).

Query 1: This query calculates the total number of quantities sold from the Sales and Listings table.

We have the join on the “listid” from both the tables. Also, the query says that Listtime should be later than '2008-12-26'

Query 1 Output: 46

Query 2: This query calculates the total number of quantities sold from the Sales table where the Saletime is later than “2008-12-25” and the listid is later than ‘2008-12-26’

Query 2 Output: 46

- Outputs of both the queries are the same.
- Cost of Query 1 is 4.2 MB
- Cost of Query 2 is 5.2 MB
- Cost of Query 1 is less as it selects the total number of quantities sold initially where the Listid is the same from two tables and then filters the record. We run the filter on less data.
- In the query 2 , we run the filter twice on both the tables. We run filter on a lot more data than the first one so the cost is more.
- Query 1 will cost less than Query 2.

If your query processes less than 1 TB, the estimate is \$0 because BigQuery provides 1 TB of on- demand query processing free per month.

In normal scenario: (On Demand Pricing)

Query 1		Units
Bytes Processed	4.2	MB
Bytes Billed	20	MB
Pricing	\$5	
Total Query Price	\$0.001	
Table Size	8.53	MB
Per GB Storage Cost	\$0.02	
Storage Pricing	\$0.00017	
Total Price	\$0.001170	

Query 2		Units
Bytes Processed	5.2	MB
Bytes Billed	20	MB
Pricing	\$5	
Total Query Price	\$0.001	
Table Size	10.24	MB
Per GB Storage Cost	\$0.02	
Storage Pricing	\$0.000205	
Total Price	\$0.001205	