

Google Cloud Platform

BigQuery Pricing

BigQuery for Data Analysts

V1.2

Approximate timing: 30 minutes

Agenda

1

Pricing Model

2

Calculating Cost

3

Quotas

4

Quiz & Lab

Pricing Model (1 of 2)

- Traditional data warehouse costs include:
 - Hardware
 - Licensing
 - Maintenance
- BigQuery is a pure consumption-based pricing model
 - Available as a fully-managed “NoOps” service
 - You save on hardware, software, maintenance costs

Pricing Model (2 of 2)

- BigQuery pricing has three categories
 - Free operations
 - Storage pricing
 - Query pricing
- Free operations include:
 - Loading and exporting data (streaming data has cost)
 - Metadata operations
- Information on free operations is found [here](#)

Pricing Model - Storage

- Storage is based on amount of data in tables
 - Calculated based on type of data stored
 - Prorated per MB per second
- Streaming data currently billed per 200 MB block ingested
 - Individual rows calculated using a 1 KB minimum size
- Current storage pricing is found [here](#)
- Further reduce storage costs by setting `expirationTime` to delete table

Notes:

There are two charges in BigQuery - one for data storage in datasets and the other for processing of data. Ingress and egress of data are free of charge with the exception of streaming inserts. Google bills monthly but only charges for the time that the data is physically stored in BigQuery.

Pricing Model - Long Term Storage

- Automatic discount for data stored in BigQuery longer than 90 days
 - Storage price drops from .02 per GB/month to .01 per GB/month
 - Query pricing remains the same
 - Discount counted on per-table, per-partition basis
 - If table data modified, 90-day counter resets
- No need to delete or archive old data
 - Equivalent to cost of Cloud Storage Nearline

Notes:

Long term storage pricing means not having to delete old data or architect a data archival process. Once this data remains in BigQuery, another added benefit is the ability to query older data using the same interface, at the same cost levels, with the same performance characteristics:

- This discount is automatic
- There's no degradation of performance, durability or functionality whatsoever
- Query cost is the same as for Standard Storage
- Discount is counted on a per-table, per-partition basis
- If you modify the data in the table, the 90-day counter resets

If your table is partitioned by time (e.g., daily), partitions older than 90 days will benefit from long term storage pricing — even as you keep creating new partitions. And, as you can see below, this process is becoming automated, which will make it even easier to benefit from long term storage pricing.

Example: Storage Pricing

*Based on storage pricing as of February 20, 2016

Storage - \$0.02 per GB, per month

Streaming Inserts - \$0.01 per 200 MB

Storage pricing is prorated per MB, per second. For example, if you store:

Total Size of Tables Stored	Cost
100 MB for half a month	You pay \$0.001
500 GB for half a month	You pay \$5
1 TB for a full month	You pay \$20

Pricing Model - Queries

- On-demand pricing (standard)
 - Queries use a shared pool of resources across users
 - First 1 TB of processed data per month is free of charge
 - Not charged for queries that:
 - Use cached results
 - Return an error
- Cancelling a running query may incur charges
- Current query pricing is found [here](#)

Notes:

Successful query execution is bill to the nearest MB with a 10 MB minimum referenced and processed. The 1 TB of data processed per billing account is free of charge. The following queries do not incur charges – queries that return an error and queries that return cached results. Cancelling running queries may result in charges up to the full cost of the query as if it were allowed to run to full completion.

Pricing Model - Slots

- Slots are the amount of total query *throughput*
 - Guarantee resources, regardless of demand on the overall multi-tenant pool
 - Minimizes variability in query performance
 - More concurrent queries without slowing down
 - Larger queries may run faster
- The more you use BigQuery, the more slots you get automatically
- Use cases: SaaS platforms, ETL tasks, and reporting
- Contact support or local sales rep for pricing

Notes:

Slots are for customers that require guaranteed set of resources regardless of the overall demand on the multi-tenant pool. This is an option for customers that use BigQuery for mission-critical applications. Customers should contact support or their local sales rep to discuss workload demand and pricing if slots are needed.

Pricing Model - High Compute Queries

- High-compute queries require large amount of processing resources per byte of data
 - Very large number of JOIN, CROSS JOIN clauses or complex user-defined functions
- Must opt-in to run [high-compute queries](#)
- If query too expensive to complete at standard \$5 per TB pricing tier, `billingTierLimitExceeded` error returned with estimate of cost
- To run query at higher pricing tier, pass new value for `maximumBillingTier` with query
 - No upper limit to the value - Use custom quotas to control costs
 - Request project-wide default value using [BigQuery high-compute queries form](#)

Agenda

1

Pricing Model

2

Calculating Cost

3

Quotas

4

Quiz & Lab

Calculating On-Demand Query Prices

- Queries priced based on data type size times columns returned
 - Column pricing is for total amount of data in the column
 - Repeated field is calculated as the sum of values in the field
 - Record field is calculated as sum of contained fields (per entry)
 - Null values for any data type count as 0 bytes -- not charged
 - BigQuery automatically calculates string length
- Data type sizes can be found [here](#)
- Calculation examples can be found [here](#)
- Using an explicit LIMIT on results does not affect price
 - LIMIT applies at root server

Notes:

Instructors should take some time to talk through the sample query calculations in the link.

On-demand pricing is calculated based on the table columns that are referenced in the query. The cost is calculated for all of the values in the column not the values returned from the query. The exception is sparse column values. Null values are not charged, so columns with high sparsity cost less than columns with low sparsity having the same data type and row count.

Unlike other DBMS's the string data type is not pre-defined meaning that you cannot limit the size of the field to a specified amount. The string size varies from zero bytes (NULL) to 2 MB in length. BigQuery will calculate the string length of every string field accessed for billing. For example, if you had a table using a string column that had 1000 rows and 100 of those row have a value 5 characters long, the size would be $((5 + 2) \times 100) + (900 \times 0)$ or 700 bytes.

Null values are not billed, either in storage or query costs.

Calculating Cost - Web UI and CLI

- Web UI shows amount of data processed in validator

```
1 SELECT
2   year,
3   month,
4   COUNT(mean_temp) AS num_hot_days
5 FROM
6   [publicdata:samples.gsod]
7 WHERE
8   station_number = 071490
9   AND mean_temp > 80
10 GROUP BY
11  year,
```

Valid: This query will process 3.41 GB when run.

- For CLI, use `--dry_run` flag

bq query --dry_run "select title from publicdata:samples.wikipedia"

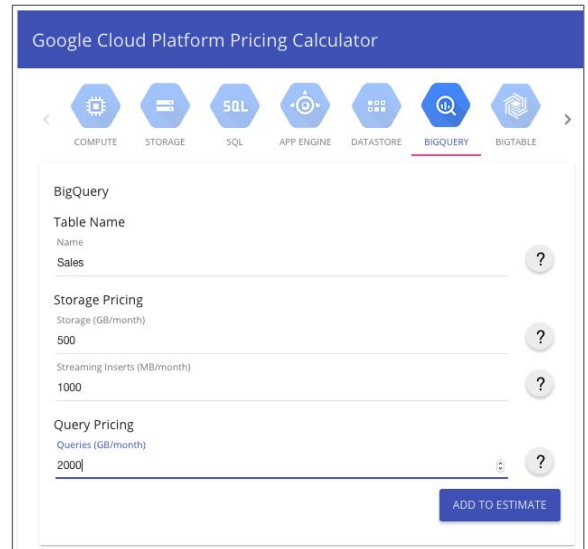
Query successfully validated. Assuming the tables are not modified, running this query will process 7294285723 bytes of data.

Notes:

Both the web UI and the CLI will return the number of bytes processed in the query. In the web UI use the query validator to get this information. In the CLI, use the `--dry_run` flag to return the information. Neither of these operations incur charges.

Demo - Pricing Calculator

- Estimate costs using the Pricing Calculator
<https://cloud.google.com/products/calculator/>



The screenshot shows the Google Cloud Platform Pricing Calculator interface. At the top, there's a navigation bar with icons for COMPUTE, STORAGE, SQL, APP ENGINE, DATASTORE, BIGQUERY (which is selected), and BIGTABLE. Below the navigation bar, the 'BigQuery' section is active. It contains the following fields:

- Table Name:** Name (Sales) with a help icon (?)
- Storage Pricing:**
 - Storage (GB/month): 500 with a help icon (?)
 - Streaming Inserts (MB/month): 1000 with a help icon (?)
- Query Pricing:**
 - Queries (GB/month): 2000 with a help icon (?)

At the bottom right of the BigQuery section, there is a blue button labeled 'ADD TO ESTIMATE'.

Notes:

Show students how cost effective using BigQuery can be. Enter 2 or 3 different examples to show how storage and queries affect pricing.

See: <https://cloud.google.com/bigquery/pricing#samplecosts>.

Agenda

1

Pricing Model

2

Calculating Cost

3

Quotas

4

Quiz & Lab

Quotas (1 of 2)

- Quotas are designed to protect all tenants
- BigQuery limits maximum rate of incoming requests
- Quotas are enforced on a per-project basis
- Specific policies vary on resource availability, user profile, service usage history, and other factors

Notes:

Policies are subject to change without notice.

Quotas (2 of 2)

- The following are subject to quotas:
 - Queries
<https://cloud.google.com/bigquery/quota-policy#queries>
 - Load jobs
<https://cloud.google.com/bigquery/quota-policy#import>
 - Export requests
<https://cloud.google.com/bigquery/quota-policy#export>
 - Streaming inserts
<https://cloud.google.com/bigquery/quota-policy#streaminginserts>
 - API requests
<https://cloud.google.com/bigquery/quota-policy#apirequests>

Custom Quotas

- Provide a more granular query cost control mechanism
- Can specify a limit on the number of bytes processed per day
- In 10TB increments
- Can set limits at project level or user level
 - See: <https://cloud.google.com/bigquery/cost-controls>

Agenda

1

Pricing Model

2

Calculating Cost

3

Quotas

4

Quiz & Lab

Module Review (1 of 2)

Which of the following statements regarding BigQuery pricing are true?
(select **3** of the available options)

- ☐ Query pricing is based on the amount of data returned in a results set
- ☐ Table storage is prorated per MB, per second
- ☐ Queries which return errors have costs
- ☐ Queries perform a full table scan on queried columns and are billed accordingly
- ☐ An INTEGER and a FLOAT are billed the same in BigQuery

Module Review (2 of 2)

Which of the following statements regarding quotas are true?
(select **3** of the available options)

- ☐ The daily query limit is 100,000 queries including cached results
- ☐ The daily limit for load jobs is 10,000 per table per day (including failures)
- ☐ The maximum query length is ~ 256 KB
- ☐ The maximum row size for a streaming insert is 1 MB

Lab

Determine query costs and explore the pricing calculator

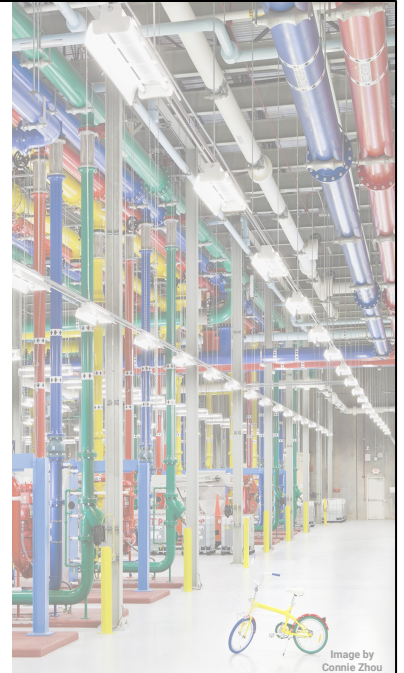


Image by
Connie Zhou

Resources

- BigQuery pricing
<https://cloud.google.com/pricing/>
- Pricing philosophy
<https://cloud.google.com/pricing/philosophy/>
- Pricing Calculator
<https://cloud.google.com/products/calculator/>

Module Review Answers (1 of 2)

Which of the following statements regarding BigQuery pricing are true?
(select **3** of the available options)

- ☐ Query pricing is based on the amount of data returned in a results set
- ✓ Table storage is prorated per MB, per second
- ☐ Queries which return errors have costs
- ✓ Queries perform a full table scan on queried columns and are billed accordingly
- ✓ An INTEGER and a FLOAT are billed the same in BigQuery

Module Review Answers (2 of 2)

Which of the following statements regarding quotas are true?
(select **3** of the available options)

- ✓ The daily query limit is 100,000 queries including cached results
- ☐ The daily limit for load jobs is 10,000 per table per day (including failures)
- ✓ The maximum query length is ~ 256 KB
- ✓ The maximum row size for a streaming insert is 1 MB

