Google Cloud

BigQuery Pricing

Evan Jones

As you've learned earlier, BigQuery has a pay for what you use pricing model. You don't pay for any setup or infrastructure. All that is managed fully behind the scenes.

Now let's talk a bit more about the pricing specifics and how you can optimize your queries for cost.

BigQuery pricing

01 BigQuery jobs

02 Sizing and cost management

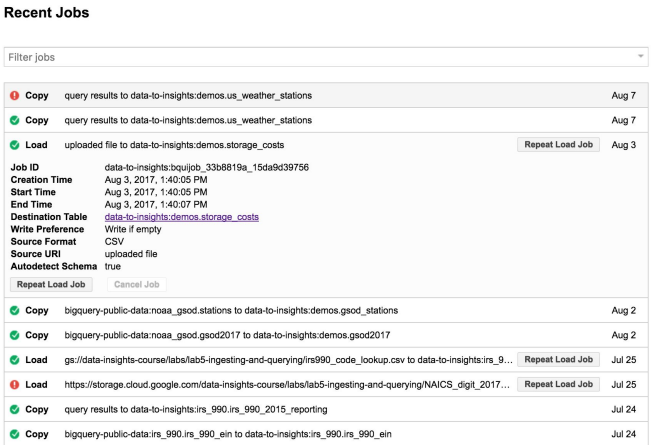
03 Optimize queries for cost

Okay, to understand pricing for BigQuery, we have to understand what a unit of work for BigQuery is. So BigQuery does work on the order of jobs.

The unit of work in BigQuery is called a Job

Each Job:

- Given a Unique ID by the web UI
- Can run concurrently
- Perform tasks
- History stored for 6 months

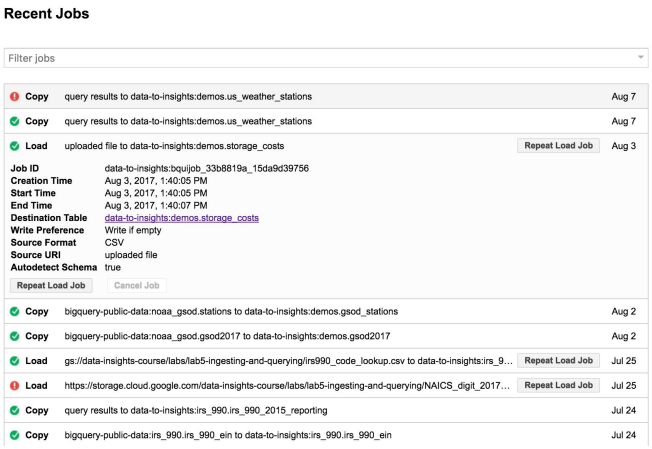


So a job is a task, so it can be a querying task, much like you're running a SQL query, but also it could be loading data, exporting data. All these things take time and resources for BigQuery to execute. Now you can do multiple jobs at once. So for example if, for your Google Cloud account, you had 20 different users, everyone's executing jobs inside of BigQuery and maybe you have a team that's loading data in and you have ten analysts that are querying data. You could have multiple, jobs that are running concurrently at the same time, asynchronously. So the jobs perform tasks, and if you needed to check back on those jobs, the history of them is stored for six months. So you can potentially repeat an import job or take a look at some of the other things that you've done. So let's talk how that relates to pricing.

The four types of BigQuery Jobs

Job types:

- **Query**
- **Load data** into a table
- **Extract** to Cloud Storage
- **Copy** existing table



So four jobs, you have querying data, loading data, extracting data, as an example, to Google Cloud Storage, or copying tables. For example, if you wanted to say hey, let's mirror our production data in our development environment, we're just going to copy that table on over. So for pricing purposes, great news for you guys is …

You only incur Query Job processing costs

Job types:

- Query - charged by bytes processed
- Load data - free
- Extract - free
- Copy - free

Note: Storing data in BigQuery is a separate cost.

…that three out of the four jobs are free. So loading or adjusting data into big queries is free, extracting it to the Google Cloud Storage is free, and copying tables are free. So the two main components of cost for BigQuery, again, consumption-based, is you're charged for the amount of data that you process. Or the bytes, like the amount of gigabytes or megabytes or terabytes that you process. And for the storage cost of storing your data permanently on disk. Now, largely, you've seen we're querying these public data sets. Obviously, you're not charged for public datasets. The BigQuery team kind of manage and hosts those. But in the next course, where you're actually creating your very own dataset, you'll actually be having to use these permanent tables that are stored within BigQuery. And we'll get into how much those cost in just a few slides.

BigQuery pricing

01 BigQuery jobs

02 Sizing and cost management

03 Optimize queries for cost

Organizations always want to know how much things will cost so let's take a look at how to determine and control the cost of both storage and analytics.

Storage pricing

Active storage pricing

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

- 100 MB for half a month, you pay $0.001 (a tenth of a cent)
- 500 GB for half a month, you pay $5
- 1 TB for a full month, you pay $20

Long-term storage pricing

- Partition or non-partitioned table not edited 90+ consecutive days
- Pricing drops ≈ 50%
- No degradation (performance, durability, availability, functionality)
- Applies to BigQuery storage only

Once your data is loaded into BigQuery, you are charged for storing it.

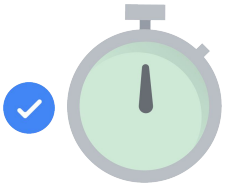
Storage pricing is based on the amount of data stored in tables when it's uncompressed. The size of the data is calculated based on the data types of the individual columns. Active storage pricing is prorated per megabyte, per second.

If a partition or non-partitioned table isn't edited for 90 consecutive days it's considered long-term storage. The price of storage for that table automatically drops by approximately 50 percent. There's no degradation of performance, durability, availability, or any other functionality.

Different actions reset the timer

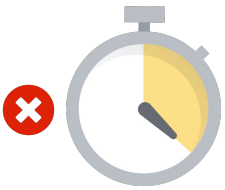
Activities that reset the timer

- Loading, copying, or streaming data into a partition or non-partitioned table
- Writing query results to a partition or non-partitioned table
- Using Data Definition Language



Activities that don't reset the timer

- Querying a table
- Creating a view that queries a table
- Exporting data from a table
- Copying a table to another destination table
- Patching or updating a table resource



If the table is edited, the price reverts back to the regular storage pricing, and the 90-day timer starts counting from zero. This is another reason for considering using partitioned tables. Even if it doesn't improve query performance you can save money on storage.

Any actions on a table that do not change the data or the structure of the table don't affect the timer.

Which BigQuery pricing model to pick?

On-demand pricing

- $5/TB of data processed
- Quota limit of 2,000 slots
- Slots shared amongst all on demand users
- 1st TB of data processed is free each month

Flat-rate pricing

- Fixed rate pricing is $10k* per 500 slots per month
- Slots are dedicated 24/7
- Starting at 500 slots
- Unlimited use of slots

Analytics engine cost is separate from storage costs. The default pricing is “on-demand”, but organizations who prefer a consistent bill amount may opt for flat-rate pricing. Capacity is sold in increments of 500 slots with a current minimum of 500 slots.

Slots are available in the following commitment plans:

- Flex slots, where you commit to an initial 60 seconds.
- Monthly, where you commit to an initial 30 days, and
- Annual, where you commit to 365 days.

Google offers a 25% discount for organizations choosing a term length of a least 1-year. Use the Google Cloud Pricing Calculator to find the exact price for the number of slots you want in a region.

Let’s compare the on-demand and flat-rate options in a little more detail.

On-demand pricing is flexible and efficient

- On-demand pricing is based solely on usage
- It doesn't matter where the data is stored
- You can set user-level and project-level custom cost controls
- You can set the maximum bytes billed by query

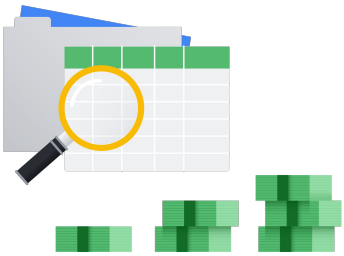
Under on-demand pricing, BigQuery charges for queries by using one metric: the number of bytes processed, also referred to as bytes read.

You're charged for the number of bytes processed whether the data is stored in BigQuery or in an external data source such as Cloud Storage, Google Drive, or Bigtable.

To help control costs you can set user-level and project-level custom quotas, and

You can set the maximum bytes billed by query. If the query will read bytes beyond the limit, the query fails with a "Query exceeded limit for bytes billed" error without incurring any charges.

Query charges



- Charged according to the total data processed in the columns selected and type of data
- No charge for errors or results from the cache
- Rounded to nearest MB, min 10 MB
- Cancelled queries may incur charges
- Data processed in columns selected supersedes LIMIT set
- Use partitioning and clustering to reduce data processed
- On-demand pricing = analysis pricing

There are a number of things to note regarding query charges.

- BigQuery uses a columnar data structure. YoU are charged according to the total data processed in the columns you select, and the total data per column is calculated based on the types of data in the column.
- You aren't charged for queries that return an error, or for queries that retrieve results from the cache. Using results from cache is on by default.
- Queries are rounded to the nearest MB, with a minimum 10 MB data processed per table referenced by the query, and with a minimum 10 MB data processed per query.
- Cancelling a running query job may incur charges up to the full cost for the query were it allowed to run to completion.
- When you run a query, you're charged according to the data processed in the columns you select, even if you set an explicit LIMIT on the results. That’s because LIMIT is processed AFTER the query completes.
- Partitioning and clustering tables can help reduce the amount of data processed by queries. As a best practice, use partitioning and clustering whenever possible.
- On-demand pricing is referred to as analysis pricing on the Google Cloud SKUs page. This is useful if you’re looking at billing reports.

Flat-rate pricing

- Applies to query costs, including DML and DDL statements
- Doesn't apply to storage costs
- Purchased as a regional resource
- Available in monthly and annual commitments
- Raise per-project concurrency quotas by contacting Google Cloud Support

Now, let's have a look at flat-rate pricing.

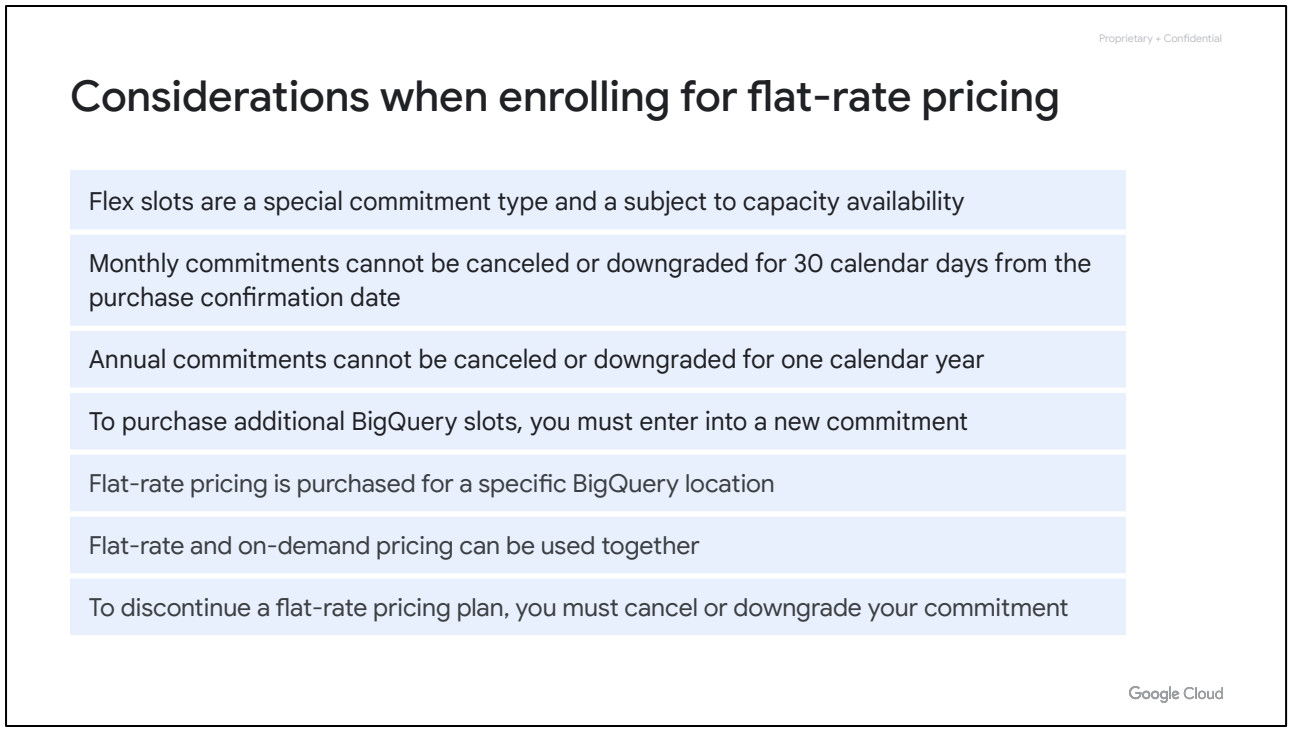
Flat-rate pricing limits the number of slots used at any time. Slot usage includes resources required for both DML and DDL. If your queries exceed your flat-rate capacity, your queries are queued until your flat-rate resources become available.

As with fixed-rate pricing, storage costs are calculated independently.

The flat-rate cost depends on both the number of slots and the region in which the slots are purchased.

By default, flat-rate pricing is calculated based on a monthly commitment, but you can get a discounted price for an annual commitment.

BigQuery limits the rate of incoming requests on a per-project basis. You can contact Google Cloud support to increase quotas, however, additional charges may occur.



There are a number of considerations for flat-rate pricing.

Flex slots are a special commitment type.

The commitment duration is only 60 seconds and you can cancel Flex slots any time thereafter. You are also charged only for the seconds your commitment was deployed.

Flex slots are subject to capacity availability. When you attempt to purchase Flex slots, the success of the purchase isn't guaranteed. However, once your commitment purchase is successful, your capacity is guaranteed until you cancel it.

Monthly commitments cannot be canceled for 30 days after your commitment is active. After the first 30 calendar days, you can cancel or downgrade at any time. If you cancel or downgrade, charges are prorated per-second at the monthly rate. For example:

- You cannot cancel on day 29,
- If you cancel during the first second of day 31, you're charged for 30 days and 1 second, and
- If you cancel at the midpoint of the third month, you're charged 50% of your monthly rate for that month.

Prior to the anniversary of your commitment date, you can choose to renew for another year, or convert it to a monthly or flex commitment. If you move to the monthly

rate, you can cancel any time, and you're charged per-second at the monthly rate. For example:

- If you renew for another year after your annual commitment date, you enter into a new annual commitment, and you continue to be charged the yearly commitment rate.
- Also, if you don't renew for another year after your annual commitment date, you can cancel at any time, and your charges are prorated per-second at the monthly rate.

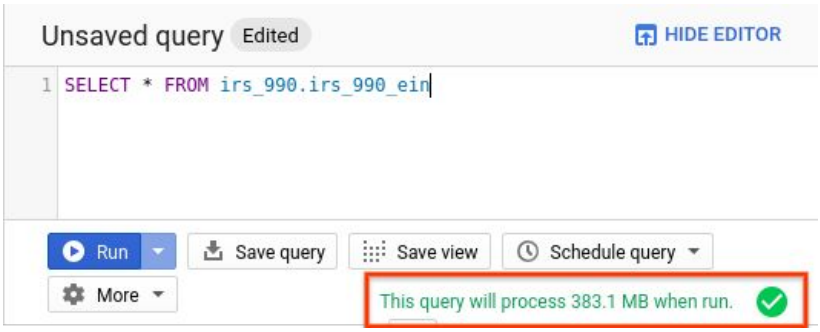
If you determine you need more BigQuery slots, you can purchase additional increments of 500. However, doing so will create a new commitment.

When you purchase at a flat-rate plan, you specify the allocation of slots by location. To use slots in multiple locations, you must purchase slots in each location.

A project can use either flat-rate or on-demand pricing. If you have multiple projects in a given location, you can choose which projects use flat-rate pricing and which projects use on-demand pricing.

Lastly, to discontinue a flat-rate pricing plan, you must cancel or downgrade your commitment, but only AFTER the initial commitment period - 30 days or 1 year

Estimating query costs: Determine the query size through the Cloud Console



On-demand queries are charged based on the number of bytes read. If your query processes a small amount of data, you might need to convert the bytes that are processed from kilobytes to megabytes. A megabyte is the smallest measure that is used by the Google Cloud Pricing Calculator.

When you enter a query in the Cloud Console, the query validator verifies the query syntax and provides an estimate of the number of bytes to be read. You can use this estimate to calculate query cost in the Pricing Calculator.

Estimating query costs: Determine the query size through the command-line

```
bq --location=[LOCATION] query --use_legacy_sql=false --dry_run
'SELECT [FIELD1], [FIELD2], [FIELD3] FROM
`[PROJECT].[DATASET].[TABLE]` LIMIT 1000'
```

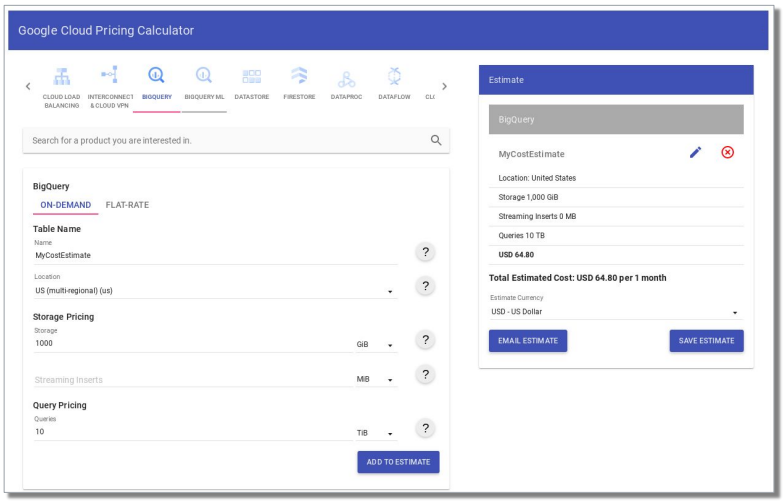
```
Query successfully validated. Assuming the tables are not
modified, running this query will process 10918 bytes of data.
```


When you run a query in the command line interface, you can use the `--dry_run` flag to estimate the number of bytes read. You can use this estimate to calculate query cost in the Pricing Calculator.

When you run the command with dry_run set, the query is not executed, and the response contains the estimated bytes.

You can also perform dry-run estimates by the API or using language libraries for Go, Python, and so on.

Estimate on-demand costs

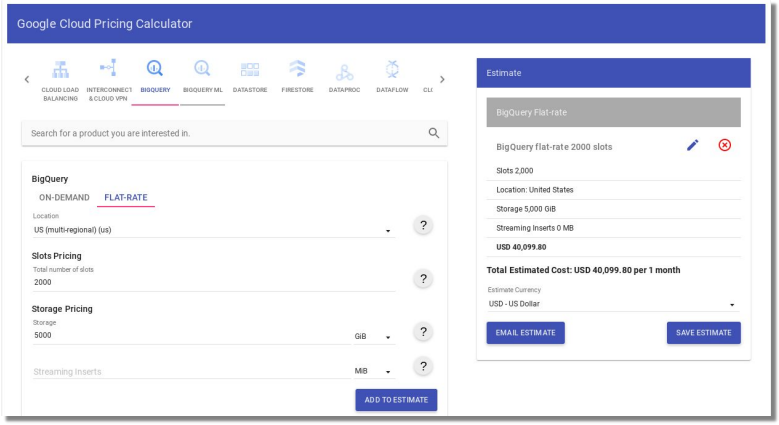


The Google Cloud Pricing Calculator can be used to estimate the on-demand cost of queries or storage, or both.

To estimate on-demand **query** costs, enter the number of bytes that are processed by the query as megabyte, gigabyte, terabyte, or petabyte. If your query processes less than 1 terabyte, the estimate is $0 because BigQuery provides 1 terabyte of on-demand processing free per month.

To estimate on-demand **storage** costs, enter the number of bytes that are stored as megabyte, gigabyte, terabyte, or petabyte. BigQuery provides 10 gigabytes of storage free per month.

Estimate flat-rate costs



https://cloud.google.com/products/calculator/

If flat-rate pricing is applied to a billing account, you can click the Flat-Rate tab, choose the location and number of slots, then add storage costs to the estimate.

A BigQuery Slot is a proprietary measure of capacity. You can choose a number of slots between 500 and two hundred thousand. As discussed, a slot is a combination of CPU, memory, and networking resources. It also includes a number of supporting technologies and sub-services.

Refer to the link in the Course Resources to access the Google Cloud Pricing Calculator.

[https://cloud.google.com/products/calculator]

BigQuery pricing

01 BigQuery jobs

02 Sizing and cost management

03 Optimize queries for cost

Okay, let's talk a little bit about what you can do to optimize your queries for cost.

Apply cost-optimization principles when writing your queries

- Only include the columns and rows you need (filter early)
- Use cached results when possible
  - (i.e. Permanent Tables instead of Views)
  - Views are saved Queries -- covered later
- Limit the use of User-Defined Functions
  - UDFs covered later

So as we covered, filter early and filter often. Use those cache results when possible. So if you're constantly executing the same query or if you can, try to do that. And one of the things that you can do there is make use of permanent tables instead of re-running your SQL query every time through review, and what a view is a saved SQL query that sits on top of a table. But again, views can take advantage of that cache as well.

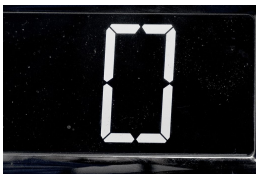
So views are saved queries, we're going to talk a little bit more about them in the next course on creating new data sets. And the difference between permanent tables, temporary tables, and those views, which are the saved queries.

So we haven't introduced them yet, but I'm already warning you about them, which are UDFs or User Defined Functions. So say, you can't find a SQL function that does what you really need it to do which is like, add three, divide by six, look up this value and then use some kind of string matching JavaScript. You could write that all in a custom function and operate on every single role in your data set. But, that has performance implications and it uses potentially more resources as well, which means higher cost for you. So there's some situations where UDFs can't be avoided if you want to do some really cool things, but just keep that in mind as the pricing and performance hit.

Summary: Calculate costs and optimize your queries

BigQuery jobs include query, load, extract, copy.

BigQuery charges for what data you consume in your queries (bytes processed).



1TB / month of free data processing. No charges for queries using cache.

Consume only the rows and columns of data you need.

Let's wrap up our discussion in BigQuery pricing by reviewing a few of the key points we covered. A unit of work in BigQuery is called a job. And the only job that actually cost money is the amount of data consumed, or bytes processed by your SQL queries, which you can actually see before you run your queries by clicking on that validator. A big benefit here is that cache data itself is free to query. And, you have one terabyte of free processing per month.

Now to save in the amount of data you consume, reduce the columns and filter the rows of your data set, so you only use the ones you actually need to query.

We'll cover more optimization techniques when we discuss BigQuery and performance as part of the Achieving Advanced Insights course.

Now, let's practice a bit with that query validator and pricing calculator in our next lab.