Module 4

# Google BigQuery Pricing

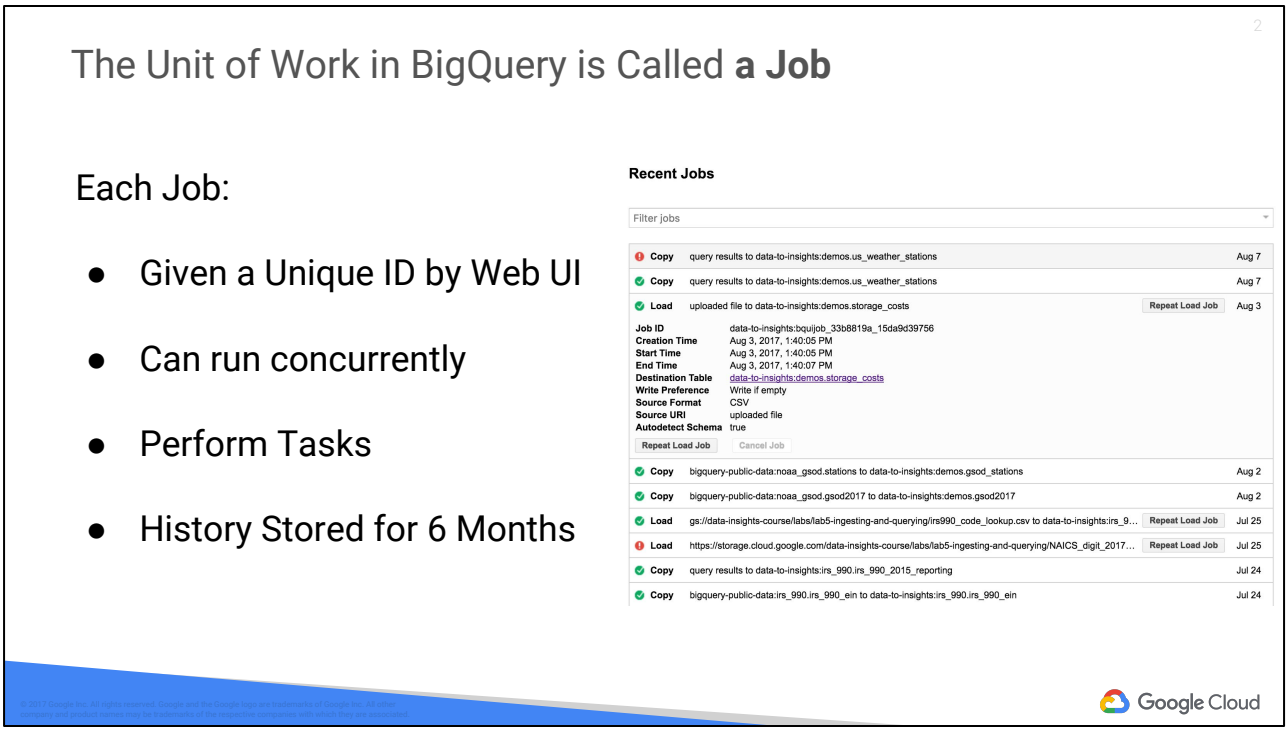
*In this module we will:*

- **Walkthrough of a BigQuery Job**
- Calculate BigQuery Pricing: Storage, Querying, and Streaming Costs
- Optimize Queries for Cost



Google Cloud

As you have learned earlier, BigQuery is a pay-for-what-you-use pricing model. You don’t pay for infrastructure.




Jobs are objects that manage asynchronous tasks such as running queries, loading data, and exporting data. You can run multiple jobs concurrently in BigQuery, and completed jobs will be listed in the Jobs collection.

Documentation:
https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs

Job History stored for 6 months (e.g. Query History). Be sure to Save and Name the queries you want to keep.

Note: If you're using the **API** and not the Web UI, you will need to specify a BigQuery Job ID in your code. Then you can poll out to BigQuery and request the status of the job

# Running jobs

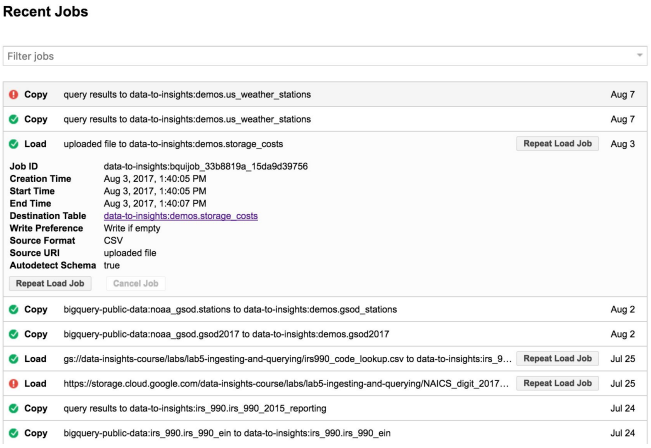
In order to run a job, you must have `bigquery.jobs.create` permissions. The following predefined IAM roles include `bigquery.jobs.create` permissions:

- `bigquery.user`
- `bigquery.jobUser`
- `bigquery.admin`

# The **Four Types** of BigQuery Jobs

Job Types:

- **Query**
- **Load Data** into a Table
- **Extract** to GCS
- **Copy** Existing Table






Google Cloud

Jobs are objects that manage asynchronous tasks such as running queries, loading data, and exporting data. You can run multiple jobs concurrently in BigQuery, and completed jobs will be listed in the Jobs collection.

https://cloud.google.com/bigquery/docs/reference/rest/v2/jobs

Job History stored for 6 months (e.g. Query History). Be sure to Save and Name the queries you want to keep.

## You Only Incur **Query Job Processing** Costs

Job Types:

- **Query** - charged by bytes processed
- **Load Data** - free
- **Extract** - free
- **Copy** - free

Note that storing data in BigQuery is a separate cost



Google Cloud

Traditional data warehouse costs include:
Hardware
Licensing
Maintenance

BigQuery:
Available as a fully-managed “NoOps” service
You save on hardware, software, maintenance costs

Module 4

# Google BigQuery Pricing

*In this module we will:*

- Walkthrough of a BigQuery Job
- **Calculate BigQuery Pricing: Storage, Querying, and Streaming Costs**
- Optimize Queries for Cost



Google Cloud

# Three categories of BigQuery pricing

**Storage**

- Amount of data in table
- Ingest rate of streaming data
- Automatic discount for old data

**Processing**

- On-demand OR Flat-rate plans
- On-demand based on amount of data processed
- 1 TB/month free
- Have to opt-in to run high-compute queries

**Free**

- Loading
- Exporting
- Queries on metadata
- Cached queries
- Queries with errors



Google Cloud

BigQuery pricing tiers:
https://cloud.google.com/bigquery/pricing

Image credits:
https://pixabay.com/en/storage-papers-office-cabinet-1209059/ (cc0)
https://pixabay.com/en/jet-engine-turbine-jet-airplane-371412/ (cc0)
https://pixabay.com/en/skydiving-jump-falling-parachuting-678168/ (cc0)

## On-demand Query and Storage Pricing (as of August 2017)

**Query** - charged by bytes processed
- $5 per TB (first TB each month is free)

**Storage** - charged by GB stored per month
- $0.02 per GB, per month

**Streaming Inserts**
- $0.05 per GB

Latest Pricing
https://cloud.google.com/bigquery/pricing



Google Cloud

Storage costs are cut in half for tables who have not been modified in 90 days.

Image:
https://pixabay.com/en/piggy-bank-save-piglet-economical-1595992/

## Storage Pricing is Prorated

| 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 |



Google Cloud

Automatic discount for data stored >90 days
Storage price drops from .02 per GB/month to .01 per GB/month

No need to delete or archive old data
Equivalent to cost of Cloud Storage Nearline

## Demo: How much is 5TB of monthly storage and 5TB querying?

Access the price calculator:

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

Latest Bigquery Pricing
https://cloud.google.com/bigquery/pricing

Google Cloud

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

Click on BigQuery

Change Storage to TB and type 5

Type 5 TB for Querying and add to estimate

Should be $122.40 per month (as of late 2017)

## Pricing Model - **Reserved** 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

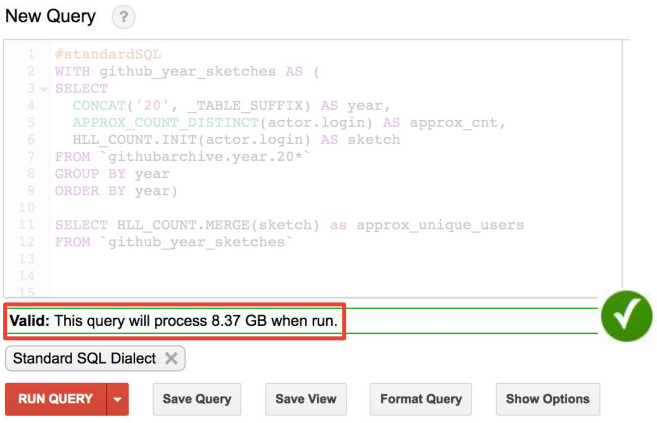


Google Cloud

Dedicated 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.

Use cases: SaaS platforms, ETL tasks, and reporting
Contact support or local sales rep for pricing

## Use the validator in the web UI to help calculate cost

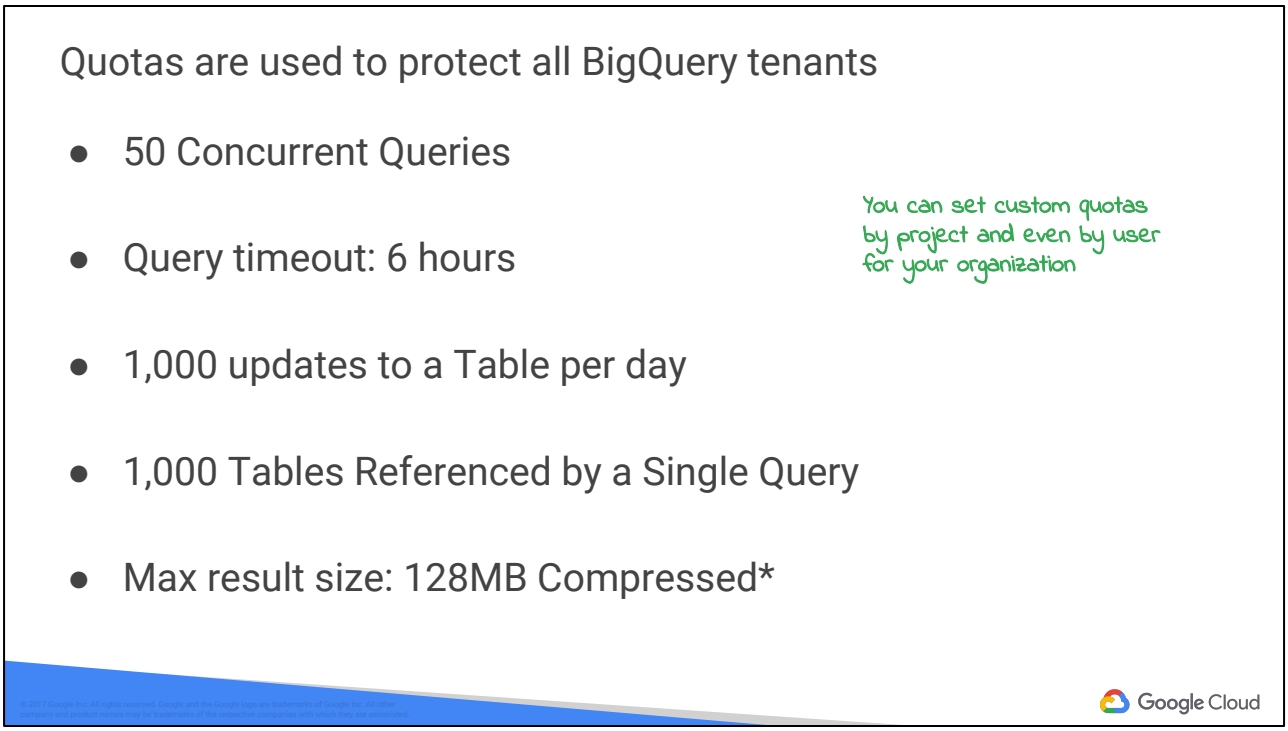





Google Cloud

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.

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.




More on quota restrictions:
https://cloud.google.com/bigquery/quota-policy

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

- **Concurrent rate limit for interactive queries under on-demand pricing:** 50 concurrent queries. Queries that return cached results, or queries configured using the dryRun property, **do not** count against this limit.
- **Concurrent rate limit for queries that contain user-defined functions (UDFs):** 6 concurrent queries, including both interactive and batch queries. Interactive queries that contain UDFs count toward the concurrent rate limit for interactive queries.
- **Daily query size limit:** unlimited by default, but you may specify limits using custom quotas.
- **Daily update limit:** 1,000 updates per table per day; applies only to the destination table in a query.
- **Query execution time limit:** 6 hours
- **Maximum concurrent slots per BigQuery account for on-demand pricing:**

- 2,000
- **Maximum number of tables referenced per query:** 1,000
- **Maximum number of authorized views per dataset:** 1,000
- **Maximum query length:** 256 KB
- **Maximum response size:** 128 MB compressed1 (unlimited when returning large query results)

**Pricing Pitfalls:** Only query what you actually need

- `SELECT column WHERE column = 123 LIMIT 10` will still need to process all rows to satisfy the filter condition
- SELECT * with very **long string length columns** means more bytes processed

Module 4

# Google BigQuery Pricing

*In this module we will:*

- Walkthrough of a BigQuery Job
- Calculate BigQuery Pricing: Storage, Querying, and Streaming Costs
- **Optimize Queries for Cost**

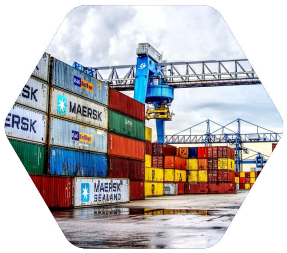


Google Cloud

Apply cost optimizing principles when writing your queries

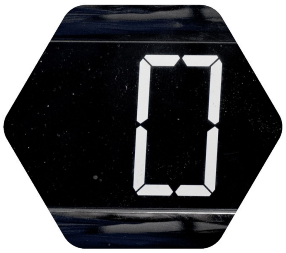
- Only include the columns and rows you need (filter early)
- Use cached results when possible
  - (i.e. Permanent Tables instead of Views)
  - Views are saved Queries -- covered later
- Limit the use of User-Defined Functions
  - UDFs covered later

## 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



Google Cloud

To wrap up our discussion on BigQuery pricing let's review a few key points. A unit of work in BigQuery is called a job. The only job that costs money is the amount of data consumed by your SQL queries which you can see by clicking on the validator. The caveat here is that cached data is free and you have 1 terabyte of free processing per month. To save on the amount of data you consume, reduce the columns and filter the rows to only those that you need. We'll cover more optimization techniques when we discuss BigQuery and performance as part the Achieving Advanced Insights with BigQuery course in this specialization.

Let's practice a bit with the query validator and pricing calculator in our next lab.

Links:
https://cloud.google.com/bigquery/pricing

Image (container ship) cc0:
https://pixabay.com/en/container-port-loading-stacked-2687008/
Image (consume cereal) cc0:
https://pixabay.com/en/cereal-spoonful-strawberry-spoon-556786/
Image (zero) cc0: https://pixabay.com/en/zero-electronic-digit-black-one-1207806/
Image (coin columns) cc0:
https://pixabay.com/en/money-money-tower-coins-euro-2180330/

# Lab 3
# **Calculate Google BigQuery Pricing**



Google Cloud

Lab 3 in Qwiklabs

## Calculate BigQuery Pricing

In this lab, you will explore Google BigQuery pricing and how to estimate query and storage costs. Additionally, you will see how modifying a query can affect the cost.

BigQuery is a pay-for-what-you-use tool.
You don't pay for infrastructure costs.



Image (bug) cc0: https://pixabay.com/en/money-business-book-parts-stock-767778/