

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

The Unit of Work in BigQuery is Called a **Job**

Each Job:

- Given a Unique ID by Web UI
- Can run concurrently
- Perform Tasks
- History Stored for 6 Months

Recent Jobs

Filter jobs		
❗ Copy	query results to data-to-insights:demos.us_weather_stations	Aug 7
✅ Copy	query results to data-to-insights:demos.us_weather_stations	Aug 7
✅ Load	uploaded file to data-to-insights:demos.storage_costs	Aug 3
<div> <div>Job ID</div> <div>data-to-insights:bqjob_33b8819a_15da9d39756</div> </div> <div> <div>Creation Time</div> <div>Aug 3, 2017, 1:40:05 PM</div> </div> <div> <div>Start Time</div> <div>Aug 3, 2017, 1:40:05 PM</div> </div> <div> <div>End Time</div> <div>Aug 3, 2017, 1:40:07 PM</div> </div> <div> <div>Destination Table</div> <div>data-to-insights:demos.storage_costs</div> </div> <div> <div>Write Preference</div> <div>Write if empty</div> </div> <div> <div>Source Format</div> <div>CSV</div> </div> <div> <div>Source URI</div> <div>uploaded file</div> </div> <div> <div>Autodetect Schema</div> <div>true</div> </div> <div> <div>Repeat Load Job</div> <div>Cancel Job</div> </div>		
✅ Copy	bigquery-public-data:noaa_gsod.stations to data-to-insights:demos.gsod_stations	Aug 2
✅ Copy	bigquery-public-data:noaa_gsod.gsod2017 to data-to-insights:demos.gsod2017	Aug 2
✅ Load	gs://data-insights-course/labs/lab5-ingesting-and-querying/irs990_code_lookup.csv to data-to-insights:irs_9...	Jul 25
❗ Load	https://storage.cloud.google.com/data-insights-course/labs/lab5-ingesting-and-querying/NAICS_digit_2017...	Jul 25
✅ Copy	query results to data-to-insights:irs_990.irs_990_2015_reporting	Jul 24
✅ Copy	bigquery-public-data:irs_990.irs_990_ein to data-to-insights:irs_990.irs_990_ein	Jul 24

The Four Types of BigQuery Jobs

Job Types:

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

Recent Jobs

Filter jobs		
❌ Copy	query results to data-to-insights:demos.us_weather_stations	Aug 7
✅ Copy	query results to data-to-insights:demos.us_weather_stations	Aug 7
✅ Load	uploaded file to data-to-insights:demos.storage_costs	<div>Repeat Load Job</div> Aug 3
Job ID data-to-insights:bqjob_33b8819a_15da9d39756 Creation Time Aug 3, 2017, 1:40:05 PM Start Time Aug 3, 2017, 1:40:05 PM End Time Aug 3, 2017, 1:40:07 PM Destination Table data-to-insights:demos.storage_costs Write Preference Write if empty Source Format CSV Source URI uploaded file Autodetect Schema true <div> <div>Repeat Load Job</div> <div>Cancel Job</div> </div>		
✅ Copy	bigquery-public-data:noaa_gsod.stations to data-to-insights:demos.gsod_stations	Aug 2
✅ Copy	bigquery-public-data:noaa_gsod.gsod2017 to data-to-insights:demos.gsod2017	Aug 2
✅ Load	gs://data-insights-course/labs/lab5-ingesting-and-querying/irs990_code_lookup.csv to data-to-insights:irs_9...	<div>Repeat Load Job</div> Jul 25
❌ Load	https://storage.cloud.google.com/data-insights-course/labs/lab5-ingesting-and-querying/NAICS_digit_2017...	<div>Repeat Load Job</div> Jul 25
✅ Copy	query results to data-to-insights:irs_990.irs_990_2015_reporting	Jul 24
✅ Copy	bigquery-public-data:irs_990.irs_990_ein to data-to-insights:irs_990.irs_990_ein	Jul 24

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

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

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

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>

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

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>

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

Use the validator in the web UI to help calculate cost

New Query ?

```
1 #standardSQL
2 WITH github_year_sketches AS (
3 SELECT
4   CONCAT('20', _TABLE_SUFFIX) AS year,
5   APPROX_COUNT_DISTINCT(actor.login) AS approx_cnt,
6   HLL_COUNT.INIT(actor.login) AS sketch
7 FROM `githubarchive.year.20*`
8 GROUP BY year
9 ORDER BY year)
10
11 SELECT HLL_COUNT.MERGE(sketch) as approx_unique_users
12 FROM `github_year_sketches`
13
14
15
```

Valid: This query will process 8.37 GB when run.



Standard SQL Dialect X

RUN QUERY

Save Query

Save View

Format Query

Show Options

Quotas are used to protect all BigQuery tenants

- 50 Concurrent Queries
- Query timeout: 6 hours
- 1,000 updates to a Table per day
- 1,000 Tables Referenced by a Single Query
- Max result size: 128MB Compressed*

You can set custom quotas
by project and even by user
for your organization

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**

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

Lab 3

Calculate Google BigQuery Pricing

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.