Examining Billing data with BigQuery

1 hourFree

**Overview**

In this lab, you learn how to use BigQuery to analyze billing data.

**Objectives**

In this lab, you learn how to perform the following tasks:

* Sign in to BigQuery from the Cloud Console
* Create a dataset
* Create a table
* Import data from a billing CSV file stored in a bucket
* Run complex queries on a larger dataset

Qwiklabs setup

For each lab, you get a new GCP project and set of resources for a fixed time at no cost.

1. Make sure you signed into Qwiklabs using an **incognito window**.
2. Note the lab's access time (for example,  and make sure you can finish in that time block.

There is no pause feature. You can restart if needed, but you have to start at the beginning.

1. When ready, click A green rectangle with white text

   Description automatically generated with medium confidence.
2. Note your lab credentials. You will use them to sign in to Cloud Platform Console. Graphical user interface, text, application

   Description automatically generated
3. Click **Open Google Console**.
4. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.

If you use other credentials, you'll get errors or **incur charges**.

1. Accept the terms and skip the recovery resource page.

Do not click **End Lab** unless you are finished with the lab or want to restart it. This clears your work and removes the project.

**Task 1: Use BigQuery to import data**

**Sign in to BigQuery and create a dataset**

1. In the Cloud Console, on the **Navigation menu** ( ), click **BigQuery**.
2. If prompted, click **Done**.
3. click on to your **Project ID** (starts with qwiklabs-gcp) and click **Create Dataset**.

You can export billing data directly to BigQuery as outlined [here](https://cloud.google.com/billing/docs/how-to/export-data-bigquery). However, for the purposes of this lab, a sample CSV billing file has been prepared for you. It is located in a Cloud Storage bucket where it is accessible to your student account. You will import this billing information into a BigQuery table and examine it.

1. Specify the following:

|  |  |
| --- | --- |
| **Property** | **Value (type value or select option as specified)** |
| **Dataset ID:** | **imported\_billing\_data** |
| **Data location:** | **US** |
| **Default table expiration > Number of days after table creation:** | In **1** days |

1. Click **Create Dataset**. You should see **imported\_billing\_data** in the left pane.

**Create a table and import**

1. Point to **imported\_billing\_data**, and then click **Create Table** to create a new table.
2. For **Source**, specify the following, and leave the remaining settings as their defaults:

|  |  |
| --- | --- |
| **Property** | **Value (type value or select option as specified)** |
| **Create table from:** | **Google Cloud Storage** |
| **Select file from GCS bucket** | gs://cloud-training/archinfra/export-billing-example.csv |
| **File format** | **CSV** |

1. For **Destination**, specify the following, and leave the remaining settings as their defaults:

|  |  |
| --- | --- |
| **Property** | **Value (type value or select option as specified)** |
| **Table name** | **sampleinfotable** |
| **Table type** | **Native table** |

1. Under **Schema** for **Auto detect** click **Schema and input parameters** .
2. Open **Advanced options**
3. Under **Header rows to skip** specify 1
4. Click **Create Table**. After the job is completed, the table appears below the dataset in the left pane.

Click *Check my progress* to verify the objective.

Use BigQuery to import data

Check my progress

**Task 2: Examine the table**

1. Click **sampleinfotable**.

This displays the schema that BigQuery automatically created based on the data it found in the imported CSV file. Notice that there are strings, integers, timestamps, and floating values.

1. Click **Details**. As you can see in **Number of Rows**, this is a relatively small table with 44 rows.
2. Click **Preview**.
3. Locate the row that has the **Description:** Network Internet Ingress from EMEA to Americas.

What was the total consumption and units consumed?



9,738 bytes

check9,738,199 bytes



9 bytes



9,738,199,000 bytes

Submit

1. Scroll to the **Cost** column.

The cost was 0.0, so with an ingress of 9.7 Mbytes, traffic from EMEA to the Americas had no charge.

1. Locate the row that has the **Description:** Network Internet Egress from Americas to China.

Can you interpret the information?

check5,542 bytes exited the Americas and was transferred to China at a charge of 1e-06.



5,542,000 bytes exited the Americas and was transferred to China at a charge of 1e-06.



5,542 bytes exited China and was transferred to the Americas at a charge of 1e-06.

Submit

**Task 3: Compose a simple query**

When you reference a table in a query, both the dataset ID and table ID must be specified; the project ID is optional.

If the project ID is not specified, BigQuery will default to the current project.

All the information you need is available in the BigQuery interface. In the column on the left, you see the dataset ID (imported\_billing\_data) and table ID (sampleinfotable).

Recall that clicking on the table name brings up the **Schema** with all of the field names.

Now construct a simple query based on the **Cost** field.

1. click **Compose New Query**.
2. Paste the following in Query Editor:

SELECT \* FROM `imported\_billing\_data.sampleinfotable`

WHERE Cost > 0content\_copy

1. Click **Run**.

How many rows had cost greater than 0?



104 rows



10 rows

check20 rows



44 rows

Submit

How many rows involved non-zero charges?

The table shows 20 rows and they all have non-zero charges.

Click *Check my progress* to verify the objective.

Compose a simple query

Check my progress

**Task 4: Analyze a large billing dataset with SQL**

In the next activity, you use BigQuery to analyze a sample dataset with 22,537 lines of billing data.

The **cloud-training-prod-bucket.arch\_infra.billing\_data** dataset used in this task is shared with the public. For more information on public datasets and how to share datasets with the public, refer to the [documentation](https://cloud.google.com/bigquery/public-data/#sharing_a_dataset_with_the_public).

1. For New Query, paste the following in Query Editor:

SELECT

product,

resource\_type,

start\_time,

end\_time,

cost,

project\_id,

project\_name,

project\_labels\_key,

currency,

currency\_conversion\_rate,

usage\_amount,

usage\_unit

FROM

`cloud-training-prod-bucket.arch\_infra.billing\_data`content\_copy

1. Click **Run**. Verify that the resulting table has 22,537 lines of billing data.
2. To find the latest 100 records where there were charges (cost > 0), for New Query, paste the following in Query Editor:

SELECT

product,

resource\_type,

start\_time,

end\_time,

cost,

project\_id,

project\_name,

project\_labels\_key,

currency,

currency\_conversion\_rate,

usage\_amount,

usage\_unit

FROM

`cloud-training-prod-bucket.arch\_infra.billing\_data`

WHERE

Cost > 0

ORDER BY end\_time DESC

LIMIT

100content\_copy

1. Click **Run**.
2. To find all charges that were more than 3 dollars, for Compose New Query, paste the following in Query Editor:

SELECT

product,

resource\_type,

start\_time,

end\_time,

cost,

project\_id,

project\_name,

project\_labels\_key,

currency,

currency\_conversion\_rate,

usage\_amount,

usage\_unit

FROM

`cloud-training-prod-bucket.arch\_infra.billing\_data`

WHERE

cost > 3content\_copy

1. Click **Run**.
2. To find the product with the most records in the billing data, for New Query, paste the following in Query Editor:

SELECT

product,

COUNT(\*) AS billing\_records

FROM

`cloud-training-prod-bucket.arch\_infra.billing\_data`

GROUP BY

product

ORDER BY billing\_records DESCcontent\_copy

1. Click **Run**.

Which product had the most billing records?



Cloud SQL has 10,271 records

checkCloud Pub/Sub has 10,271 records



Stackdriver has 11,271 records

Submit

1. To find the most frequently used product costing more than 1 dollar, for New Query, paste the following in Query Editor:

SELECT

product,

COUNT(\*) AS billing\_records

FROM

`cloud-training-prod-bucket.arch\_infra.billing\_data`

WHERE

cost > 1

GROUP BY

product

ORDER BY

billing\_records DESCcontent\_copy

1. Click **Run**.

Which product had the most billing records of over $1

checkCompute Engine has 17 charges costing more than 1 dollar.



Kubernetes Engine has 7 charges costing more than 1 dollar.



Cloud SQL has 15 charges costing more than 1 dollar.

Submit

1. To find the most commonly charged unit of measure, for Compose New Query, paste the following in Query Editor:

SELECT

usage\_unit,

COUNT(\*) AS billing\_records

FROM

`cloud-training-prod-bucket.arch\_infra.billing\_data`

WHERE cost > 0

GROUP BY

usage\_unit

ORDER BY

billing\_records DESCcontent\_copy

1. Click **Run**.

What was the most commonly charged unit of measure?



Requests were the most commonly charged unit of measure with 6,539 requests.



Requests were the most commonly charged unit of measure with 504 requests.

checkByte-seconds were the most commonly charged unit of measure with 2,937 requests.

Submit

1. To find the product with the highest aggregate cost, for New Query, paste the following in Query Editor:

SELECT

product,

ROUND(SUM(cost),2) AS total\_cost

FROM

`cloud-training-prod-bucket.arch\_infra.billing\_data`

GROUP BY

product

ORDER BY

total\_cost DESCcontent\_copy

1. Click **Run**.

Which product has the highest total cost?

checkCompute Engine has an aggregate cost of $112.02.



Cloud SQL has an aggregate cost of $47.37



BigQuery has an aggregate cost of $114.02

Submit

**Task 5: Review**

In this lab, you imported billing data into BigQuery that had been generated as a CSV file. You ran a simple query on the file. Then you accessed a shared dataset containing more than 22,000 records of billing information. You ran a variety of queries on that data to explore how you can use BigQuery to ask and answer questions by running queries.

**End your lab**

When you have completed your lab, click **End Lab**. Qwiklabs removes the resources you’ve used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

* 1 star = Very dissatisfied
* 2 stars = Dissatisfied
* 3 stars = Neutral
* 4 stars = Satisfied
* 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support** tab.

Copyright 2020 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.