

ONDEMAND- Analyzing Billing Data with BigQuery

40 minutes1 Credit

Overview

Introduction

[BigQuery](#) is Google's serverless, highly scalable enterprise data warehouse that is designed to make data analysts more productive with unmatched price-performance.

In this lab, you will learn how to use BigQuery to examine sample Cloud Billing records. After you gain some familiarity with the tool and the dataset, you will run SQL queries to gain insights from your billing data.

Objectives

In this lab, you will learn how to do the following:

- Sign in to BigQuery from the Google Cloud Console.
- Examine the sample dataset and table.
- Compose and run simple queries on the billing data.
- Run queries on the data and answer pertinent billing questions.

Setup and Requirements

Qwiklabs setup

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

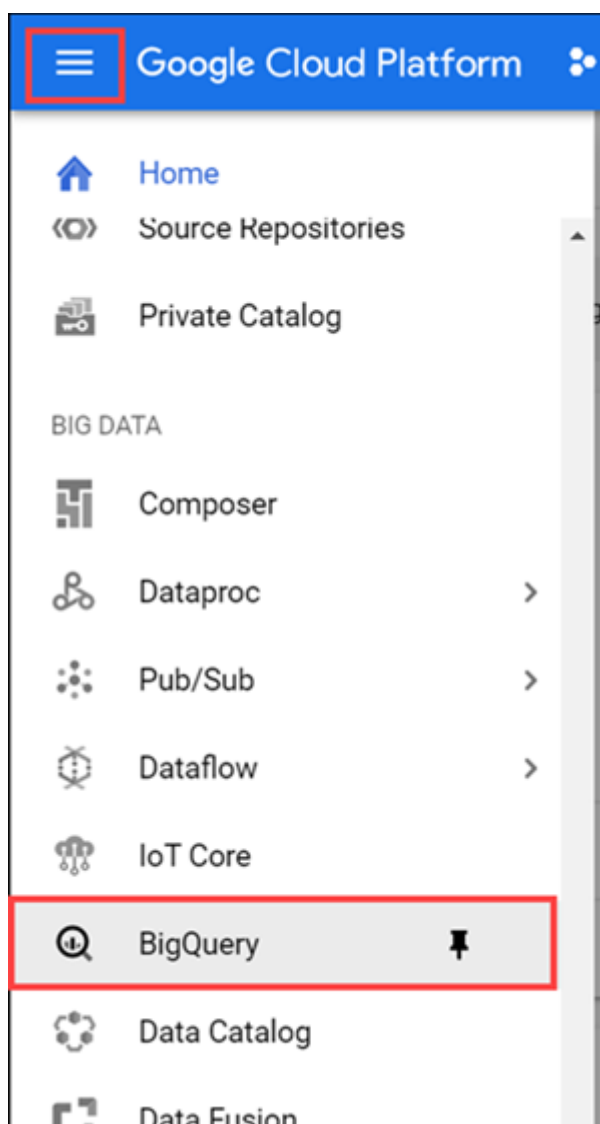
1. Sign in to Qwiklabs using an **incognito window**.
2. Note the lab's access time (for example, 02:00:00), and make sure you can finish within that time. There is no pause feature. You can restart if needed, but you have to start at the beginning.
3. When ready, click **Start lab**.
4. Note your lab credentials (**Username** and **Password**). You will use them to sign in to the Google Cloud Console.
5. Click **Open Google Console**.
6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts. If you use other credentials, you'll receive errors or **incur charges**.
7. Accept the terms and skip the recovery resource page.

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

Locate your dataset and table in BigQuery

To examine your billing data, you must first locate your dataset in the Cloud Console. In this lab, you use BigQuery to look at billing data associated with your project.


1. In the Cloud Console, open BigQuery:
 - Click the **Navigation menu**.
 - Scroll down and click **BigQuery**.




2. If prompted with the following, click **Done**:

Welcome to BigQuery in the Cloud Console

New to the BigQuery UI?

The BigQuery UI helps you complete tasks like running queries, loading data, and even creating and training ML models. Check out the BigQuery [quickstart guide](#)  to learn how to start performing data analysis on Google Cloud.

Learn about new features

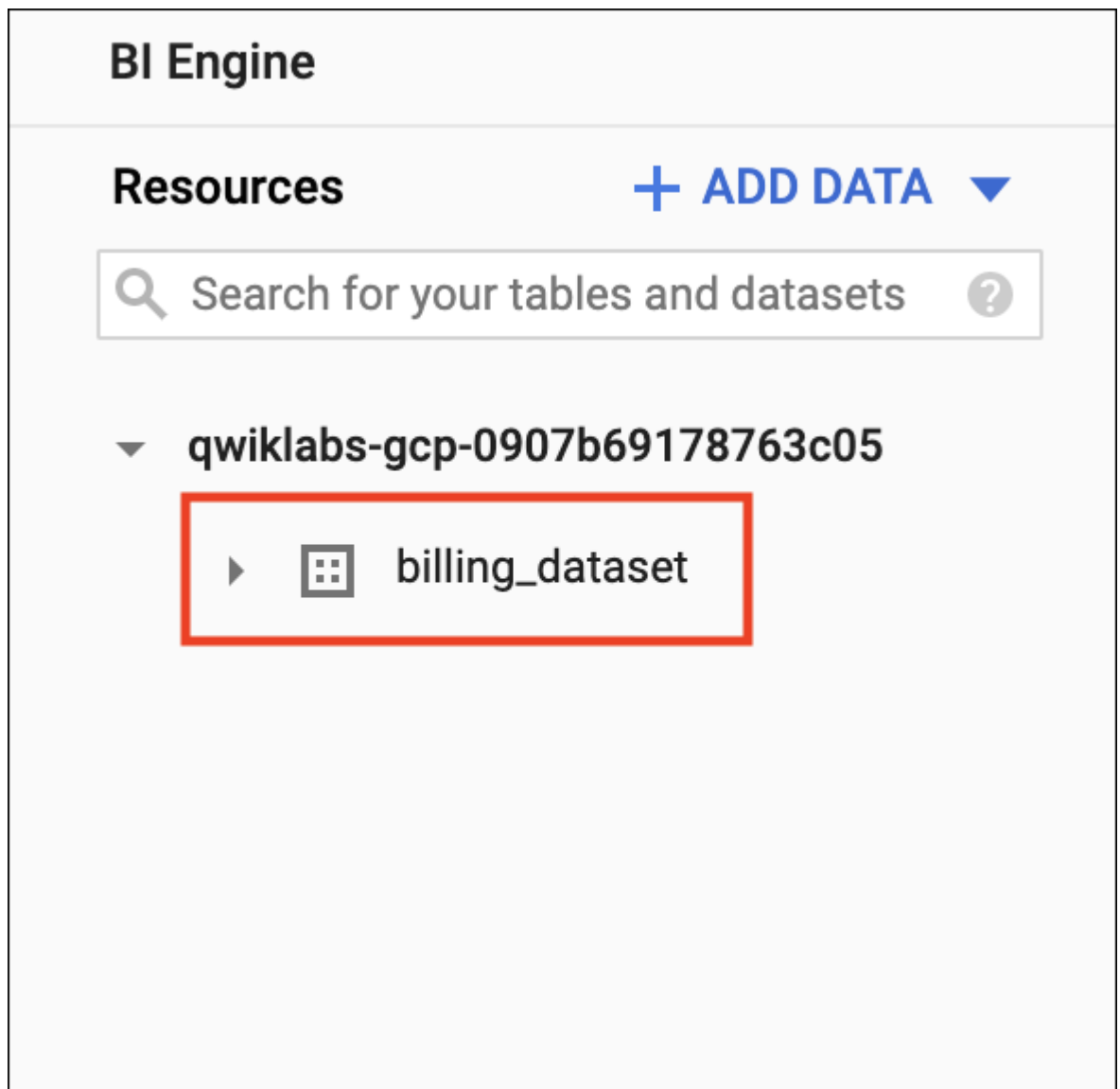
New improvements and updates are constantly on the way. We recommend periodically checking our [release notes](#)  to stay up to date on what's new.

DONE

3. To view `billing_dataset`:

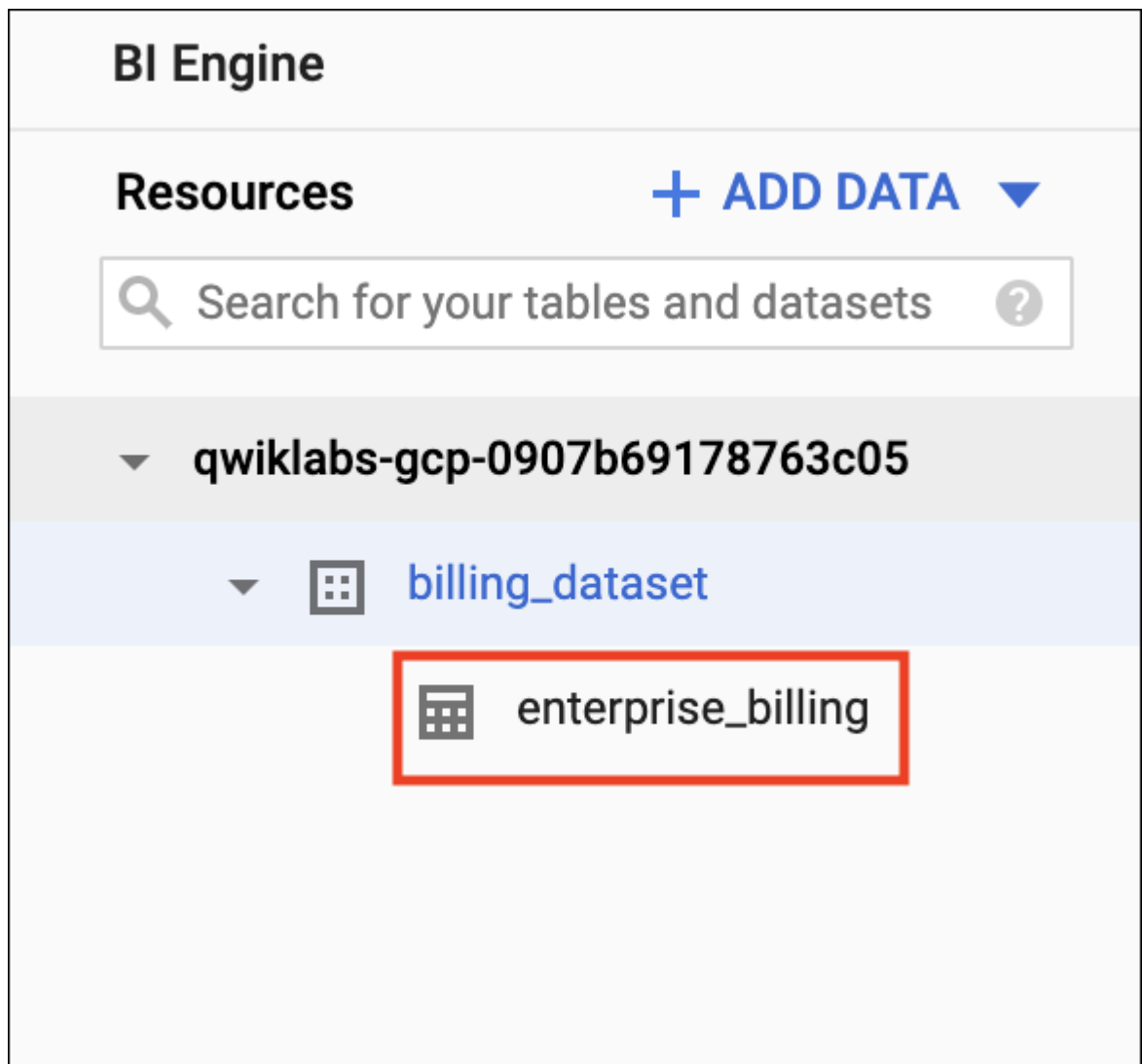
- Under **Resources** in the left panel, locate your Project ID. This is your BigQuery project.
- Click the triangle to view datasets in your project.

You should see `billing_dataset`:



4. Similarly, click the triangle next to `billing_dataset` to view tables and views in the dataset.

You should see the `enterprise_billing` table under `billing_dataset`:



Examine the billing data

You've located your billing data. In this section you examine the billing data in BigQuery.

1. In the left panel, click the `enterprise_billing` table. This displays three tabs that provide information on the `enterprise_billing` table. The tabs are **Schema**, **Details**, and **Preview**. The **Schema** tab is open by default.

Resources
+ ADD DATA

Search for your tables and datasets

qwiklabs-gcp-01-75457b763d8a

billing_dataset

enterprise_billing

Run
Save query
Save view
Schedule

enterprise_billing
QUERY TABLE

Schema
Details
Preview

Field name	Type	Mode	Policy
billing_account_id	STRING	NULLABLE	
service	RECORD	NULLABLE	
service.id	STRING	NULLABLE	
service.description	STRING	NULLABLE	
sku	RECORD	NULLABLE	
sku.id	STRING	NULLABLE	
sku.description	STRING	NULLABLE	
usage_start_time	TIMESTAMP	NULLABLE	
usage_end_time	TIMESTAMP	NULLABLE	
project	RECORD	NULLABLE	
project.id	STRING	NULLABLE	
project.name	STRING	NULLABLE	

You see the schema that BigQuery automatically created based on the sample Cloud Billing records. Notice that there are strings, integers, timestamps, and floating values.

- Click the **Details** tab for table information.
You see a table with 415,602 rows.
- Click the **Preview** tab to view the table.
- Look at the header row of the table to see what information the data provides, then answer the following question:

Which of the following lists some of the information provided by this table?



The account charges are billed to; the service provided; usage cost; and the invoice on which this charge appears



The company/person charges are billed to; the billing address; the service provided; and the usage cost



The project using the provided service; the service provided; the region in which the service is used; and where a backup copy of any data consumed is stored



The account charges are billed to; the contact person for the account; the project that uses the provided service, when the project starts and stops using the provided service; the invoice on which a charge appears

Submit

6. Look at the bottom right and find the Rows per page field. You see you can set the number of rows you see per page from 10 - 200.

The first 200 entries show that Cloud Pub/Sub service was provided to the same billing account and within the same country.



True



False

Before going on to the next section, click around to explore what other information this data provides.

Analyze data using SQL queries

It's not enough to view your billing data for information. For meaningful information, you must analyze your data by asking then answering questions, such as which services were used and what were their associated costs; which project(s) incurred the most cost; and are the costs as expected?

In BigQuery, you use SQL queries to pull and process data from a table to answer your questions. To reference a table in a query, you specify the dataset and table; the project is optional.

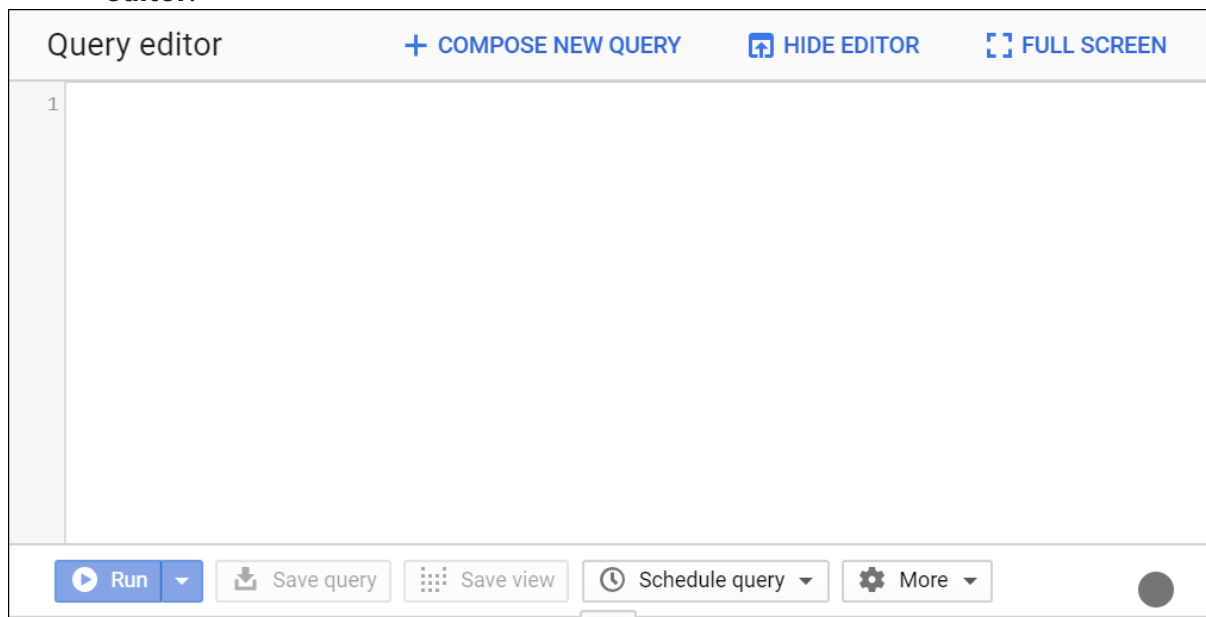
If you do not specify the Project, BigQuery defaults to the current project. You can access all the information you need in the BigQuery interface. In the left panel, under **Resources**, you see the dataset (`billing_dataset`) and the table (`enterprise_billing`).

Recall that clicking on the table name in the left panel brings up the **Schema** tab on the right, which lists the field names.

Query 1: Analyze your data based on costs

You've looked at your data, now do some analysis based on costs. Construct a simple query based on the **Cost** field.

1. Locate the **Query editor**. You enter and run your SQL queries in the **Query editor**.



2. Type or paste the following into the **Query editor**:

```
SELECT * FROM `billing_dataset.enterprise_billing` WHERE Cost > 0
```

Copied!

content_copy

This script queries data in the `enterprise_billing` table for records with a `Cost` of greater than zero.

3. Click **Run**. The following output is returned:

<div> <div>Run</div> <div>Save query</div> <div>Save view</div> <div>Schedule query</div> <div>More</div> </div> <div>This query will process 147.2 MB when run. </div>							
<div>Query results</div> <div> SAVE RESULTS EXPLORE DATA </div>							
Query complete (1.9 sec elapsed, 147.2 MB processed)							
<div> Job information Results JSON Execution details </div>							
Row	billing_account_id	service.id	service.description	sku.id	sku.description	usage_start_time	usage_end_time
1	01150A-B8F62B-47D999	95FF-2EF5-5EA1	Cloud Storage	0D5D-6E23-4250	Standard Storage US Multi-region	2020-05-14 19:00:00 UTC	2020-05-14 20:00:00 UTC
2	01150A-B8F62B-47D999	95FF-2EF5-5EA1	Cloud Storage	0D5D-6E23-4250	Multi-Regional Storage US	2020-02-23 20:00:00 UTC	2020-02-23 21:00:00 UTC
3	01150A-B8F62B-47D999	95FF-2EF5-5EA1	Cloud Storage	0D5D-6E23-4250	Standard Storage US Multi-region	2020-03-06 20:00:00 UTC	2020-03-06 21:00:00 UTC
4	01150A-B8F62B-47D999	95FF-2EF5-5EA1	Cloud Storage	0D5D-6E23-4250	Standard Storage US Multi-region	2020-07-18 19:00:00 UTC	2020-07-18 20:00:00 UTC
5	01150A-B8F62B-47D999	95FF-2EF5-5EA1	Cloud Storage	0D5D-6E23-4250	Standard Storage US Multi-region	2020-06-22 19:00:00 UTC	2020-06-22 20:00:00 UTC
6	01150A-B8F62B-47D999	95FF-2EF5-5EA1	Cloud Storage	0D5D-6E23-4250	Standard Storage US Multi-region	2020-03-09 19:00:00 UTC	2020-03-09 20:00:00 UTC
7	01150A-B8F62B-47D999	95FF-2EF5-5EA1	Cloud Storage	0D5D-6E23-4250	Multi-Regional Storage US	2020-01-22 20:00:00 UTC	2020-01-22 21:00:00 UTC

While this shows you how to run a query in BigQuery, the resulting table is not clear or helpful. For a more useful query, run the following script to see how much was spent for services.

1. Click **COMPOSE NEW QUERY** to clear the **Query editor**.
2. Type or paste the following script into the **Query editor**:

```
SELECT
  project.name as Project_Name,
  service.description as Service,
  location.country as Country,
  cost as Cost
FROM `billing_dataset.enterprise_billing`;
```

Copied!

content_copy

Notice you've reduced the number of columns by selecting what information (project.name, service.description, location.country, and cost) you want to see.

The following output is returned :

<div> <div>Run</div> <div>Save query</div> <div>Save view</div> <div>Schedule query</div> <div>More</div> </div>					
<div> <div>Query results</div> <div>SAVE RESULTS</div> <div>EXPLORE DATA</div> </div>					
Query complete (0.0 sec elapsed, cached)					
<div> <div>Job information</div> <div>Results</div> <div>JSON</div> <div>Execution details</div> </div>					
Row	Project_Name	Service	Country	Cost	
1	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
2	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
3	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
4	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
5	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
6	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
7	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
8	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
9	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
10	CTG - Sandbox	Cloud Pub/Sub	US	0.0	
					<div> <div>Rows per page:</div> <div>100</div> <div>1 - 100 of 415602</div> </div>

Complete the next steps to check your progress to verify an objective and then answer a question using the SQL query output.

- Click the `enterprise_billing` table in the left pane, and then click the Schema tab in the **enterprise_billing** section on the right to answer this question.

To see when the service was used, what field would you add to the query - feel free to test in BigQuery?



service.id



system_labels.value



what_is_it.time



usage_start_time

Submit

- Click **Check my progress** to verify you've completed the objective. Construct a simple query based on the Cost field

Check my progress

Query 2: Examine key information

In the previous section, you queried for specific information, also known as key information, to reduce the amount of data in the table. You used parameters to identify key information. In this section, you'll list key information.

For this example, the number of unique services that are available is the key information you want. Run a query that combines the service description and the sku description, then lists that as line items.

1. Click **COMPOSE NEW QUERY** to clear the **Query editor**.
2. Type or paste the following into the **Query editor**:

```
SELECT CONCAT(service.description, ' : ',sku.description) as Line_Item  
FROM `billing_dataset.enterprise_billing` GROUP BY 1
```

Copied!

content_copy

Note that `GROUP BY 1` means to group the list by the first column.

3. Click **Run**.
You should receive the following output:

Query results

[SAVE RESULTS](#)[EXPLORE DATA](#) ▼

Query complete (0.4 sec elapsed, 20.7 MB processed)

[Job information](#)[Results](#)[JSON](#)[Execution details](#)

Row	Line_Item
1	Cloud Pub/Sub : Message Delivery Basic
2	Cloud Storage : Standard Storage US Multi-region
3	Cloud Storage : Multi-Regional Storage US
4	Compute Engine : Network Google Ingress from Americas to Americas
5	Compute Engine : Network Intra Zone Egress
6	Compute Engine : Licensing Fee for Debian 9 Stretch (RAM cost)
7	Compute Engine : Network Inter Region Ingress from Montreal to Americas
8	Cloud Storage : Coldline Data Retrieval
9	BigQuery : Analysis
10	Compute Engine : Network Inter Region Ingress from Americas to Americas
11	Compute Engine : Network Internet Ingress from China to Americas
12	Cloud Storage : Download Worldwide Destinations (excluding Asia & Australia)
13	Cloud Storage : Coldline Storage Asia Multi-region
14	Cloud Storage : Multi-Regional Coldline Storage Asia
15	Compute Engine : N1 Predefined Instance Core running in Americas

What did you just do?



Determined the number of unique services available and listed the service name and description



Determined and listed the number of services by region



Determined and listed the services used in the US for the last 60 days



Determined and listed how many unique services were used in this billing cycle

Submit

How many different line items does the sample Cloud Billing data cover?



160 services

☐ 12 services

☐ 68 services

☐ 5 services

Submit

4. Click **Check my progress** to verify you've completed the objective.

List unique services available from the sample bill

Check my progress

Query 3: Analyze service usage

You want to look at service usage. Find out the number of times a resource utilized a service/SKU?

1. Click **COMPOSE NEW QUERY** to clear the **Query editor**.
2. Type or paste the following into the **Query editor**:

```
SELECT CONCAT(service.description, ' : ',sku.description) as Line_Item,  
Count(*) as NUM FROM `billing_dataset.enterprise_billing` GROUP BY  
CONCAT(service.description, ' : ',sku.description)
```

Copied!

content_copy

3. Click **Run**.

You should receive the following output:

Query results

[SAVE RESULTS](#)
[EXPLORE DATA](#)

Query complete (0.1 sec elapsed, cached)

[Job information](#)
[Results](#)
[JSON](#)
[Execution details](#)

Row	Line_Item	NUM
1	Cloud Pub/Sub : Message Delivery Basic	4093
2	Cloud Storage : Standard Storage US Multi-region	205
3	Cloud Storage : Multi-Regional Storage US	233
4	Compute Engine : Network Google Ingress from Americas to Americas	119
5	Compute Engine : Network Intra Zone Egress	42
6	Compute Engine : Licensing Fee for Debian 9 Stretch (RAM cost)	34538
7	Compute Engine : Network Inter Region Ingress from Montreal to Americas	10
8	Cloud Storage : Coldline Data Retrieval	65
9	BigQuery : Analysis	59
10	Compute Engine : Network Inter Region Ingress from Americas to Americas	56
11	Compute Engine : Network Internet Ingress from China to Americas	116
12	Cloud Storage : Download Worldwide Destinations (excluding Asia & Australia)	4
13	Cloud Storage : Coldline Storage Asia Multi-region	29
14	Cloud Storage : Multi-Regional Coldline Storage Asia	33
15	Compute Engine : N1 Predefined Instance Core running in Americas	30816

How many services were used 6 times in the billing record?

☐

7

☐

4

☐

3

☐

0

Submit

What services produced 3349 logs in your billing record?

☐

BigQuery : Active Storage

☐

Cloud Functions : Network Egress from us-central1



Cloud Pub/Sub : Inter-region data delivery from North America to North America



Compute Engine : Network Internet Standard Tier Egress from Belgium

Submit

4. Click **Check my progress** to verify you've completed the objective.

Get the count of logs generated for each service from the sample bill

Check my progress

Query 4: Find which project has the most records

Find the Google Cloud project with the most records.

1. Click **COMPOSE NEW QUERY** to clear the **Query editor**.
2. Type or paste the following into the **Query editor**:

```
SELECT project.id, count(*) as count from  
`billing_dataset.enterprise_billing` GROUP BY project.id
```

Copied!


content_copy

This query counts how many times a `project.id` appears in a record, and groups the results by `project.id`.

3. Click **Run**.

You should receive the following output:

Query results

 [SAVE RESULTS](#)

Query complete (0.2 sec elapsed, 6.4 MB processed)

[Job information](#) [Results](#) [JSON](#) [Execution details](#)

Row	id	count	
1	ctg-sandbox-242206	23886	
2	ctg-storage	39877	
3	ctg-prod-241521	87771	
4	ctg-dev-241406	264068	

Which Google Cloud Project has the most billing records?



ctg-storage



ctg-dev-241406



ctg-prod-241521



ctg-sandbox-242206

Submit

4. Click **Check my progress** to verify you've completed the objective.

Find the Google Cloud project with the most records in the billing data

Check my progress

Query 5: Find the cost per project

Find the cost breakdown for each project:

1. Click **COMPOSE NEW QUERY** to clear the **Query editor**.
2. Type or paste the following into the **Query editor**:

```
SELECT ROUND(SUM(cost),2) as Cost, project.name from  
`billing_dataset.enterprise_billing` GROUP BY project.name
```

Copied!

content_copy

This query adds the cost per project.name, then returns the results grouped by project.name.

3. Click **Run**.

You should receive the following output:

Query results

 [SAVE RESULTS](#)

Query complete (0.3 sec elapsed, 7.9 MB processed)

[Job information](#) [Results](#) [JSON](#) [Execution details](#)

Row	Cost	name	
1	0.0	CTG - Sandbox	
2	241.46	CTG - Storage	
3	1000.13	CTG - Prod	
4	2459.35	CTG - Dev	

Which project generates the largest cost?

☐

CTG - Storage

☐

CTG - Prod

☐

CTG - Dev

☐

CTG - Sandbox

Submit

4. Click **Check my progress** to verify you've completed the objective.

Find the cost breakdown per project

[Check my progress](#)

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.

©2021 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.