Analyzing Billing Data with BigQuery

GSP621

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 familiarity with the tool and 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 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

Before you click the Start Lab button

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources will be made available to you.

This Qwiklabs hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access Google Cloud for the duration of the lab.

What you need

To complete this lab, you need:

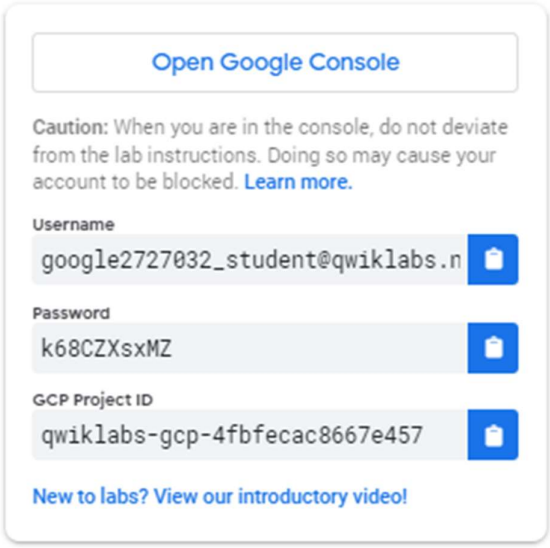
- Access to a standard internet browser (Chrome browser recommended).
- Time to complete the lab.
  Note: If you already have your own personal Google Cloud account or project, do not use it for this lab.

Note: If you are using a Pixelbook, open an Incognito window to run this lab.

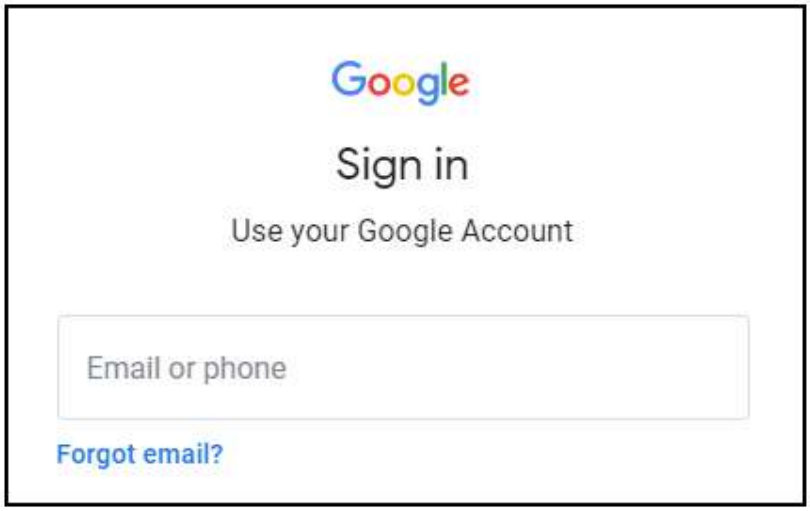
Cloud Console

How to start your lab and sign in to the Google Cloud Console

1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is a panel populated with the temporary credentials that you must use for this lab.

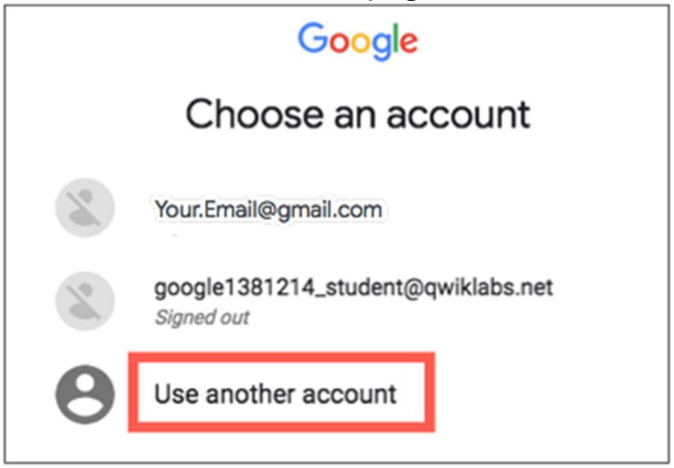


2. Copy the username, and then click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Sign in** page.



Tip: Open the tabs in separate windows, side-by-side.

If you see the **Choose an account** page, click **Use Another Account**.



3. In the **Sign in** page, paste the username that you copied from the Connection Details panel. Then copy and paste the password.

   Important: You must use the credentials from the Connection Details panel. Do not use your Qwiklabs credentials. If you have your own Google Cloud account, do not use it for this lab (avoids incurring charges).

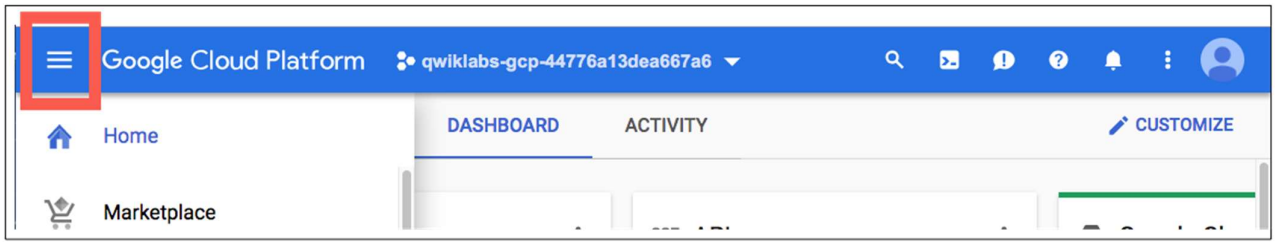
4. Click through the subsequent pages:

   - Accept the terms and conditions.
   - Do not add recovery options or two-factor authentication (because this is a temporary account).
   - Do not sign up for free trials.

After a few moments, the Cloud Console opens in this tab.

Note: You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-

left.

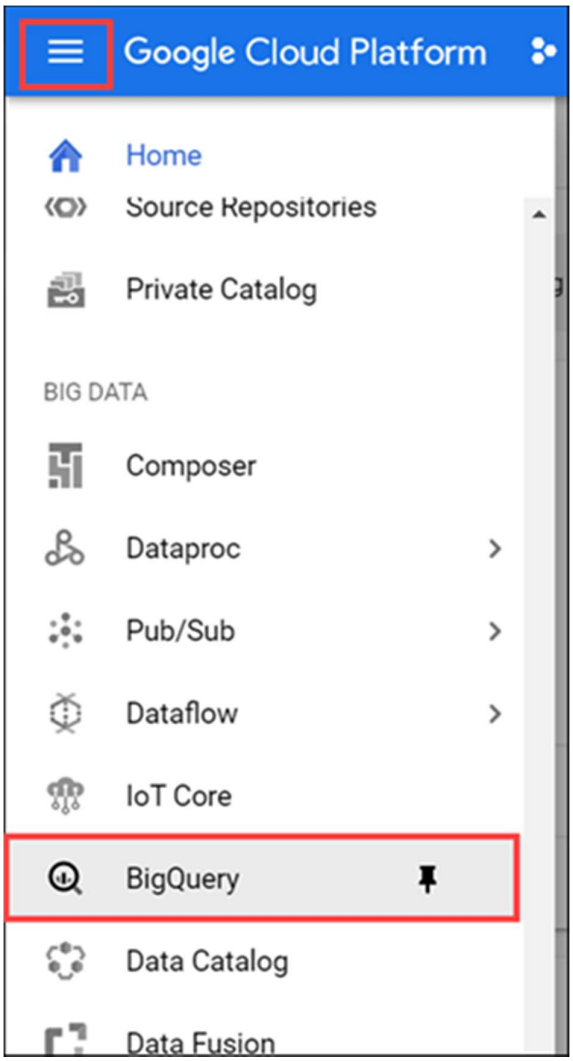

Google Cloud Platform
qwiklabs-gcp-44776a13dea667a6
Home
Marketplace
DASHBOARD
ACTIVITY
CUSTOMIZE

Locate your dataset and table in BigQuery

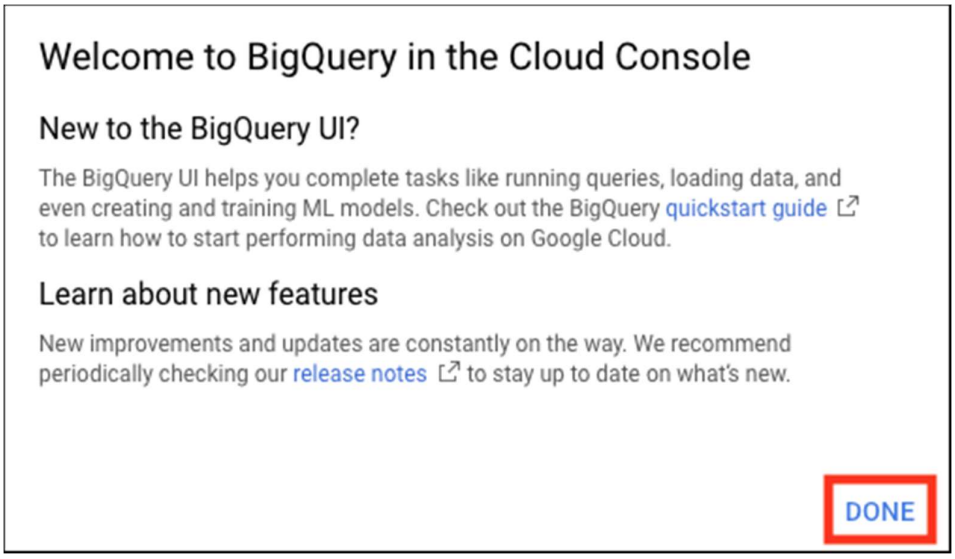
To examine your billing data, you must first locate your dataset in the Cloud Console. In this lab, you use BigQuery to look at billing data associated with your project.

1. In the Cloud Console, open BigQuery:

- Click the **Navigation menu**.
- Scroll down and click **BigQuery**.



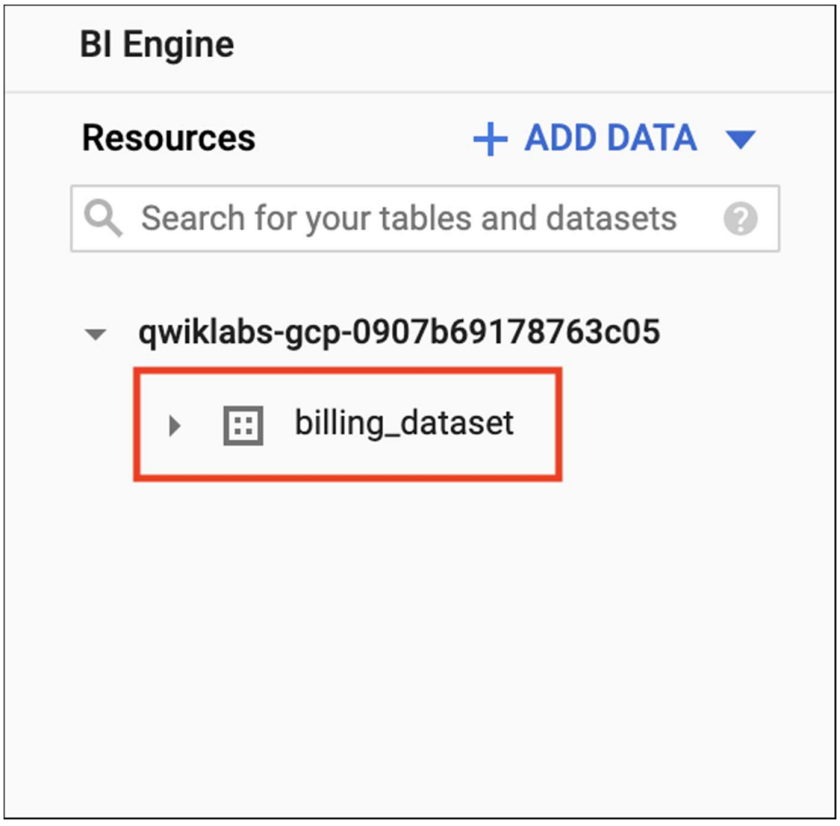
2. If prompted with the following, click **Done**:



3. To view `billing_dataset`:

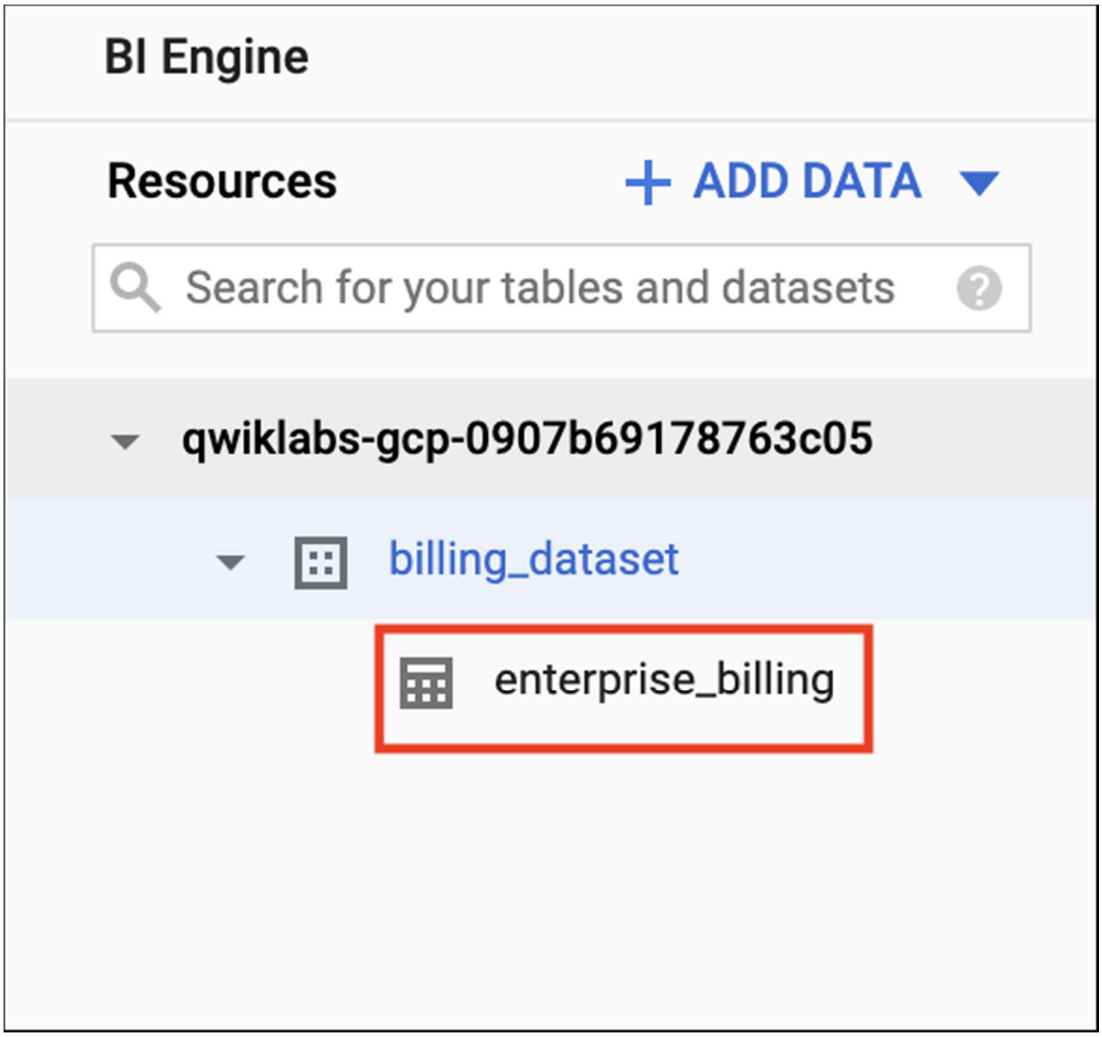
- Under **Resources** in the left panel, locate your Project ID. This is your BigQuery project.
- Click the triangle to view datasets in your project.

  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`:

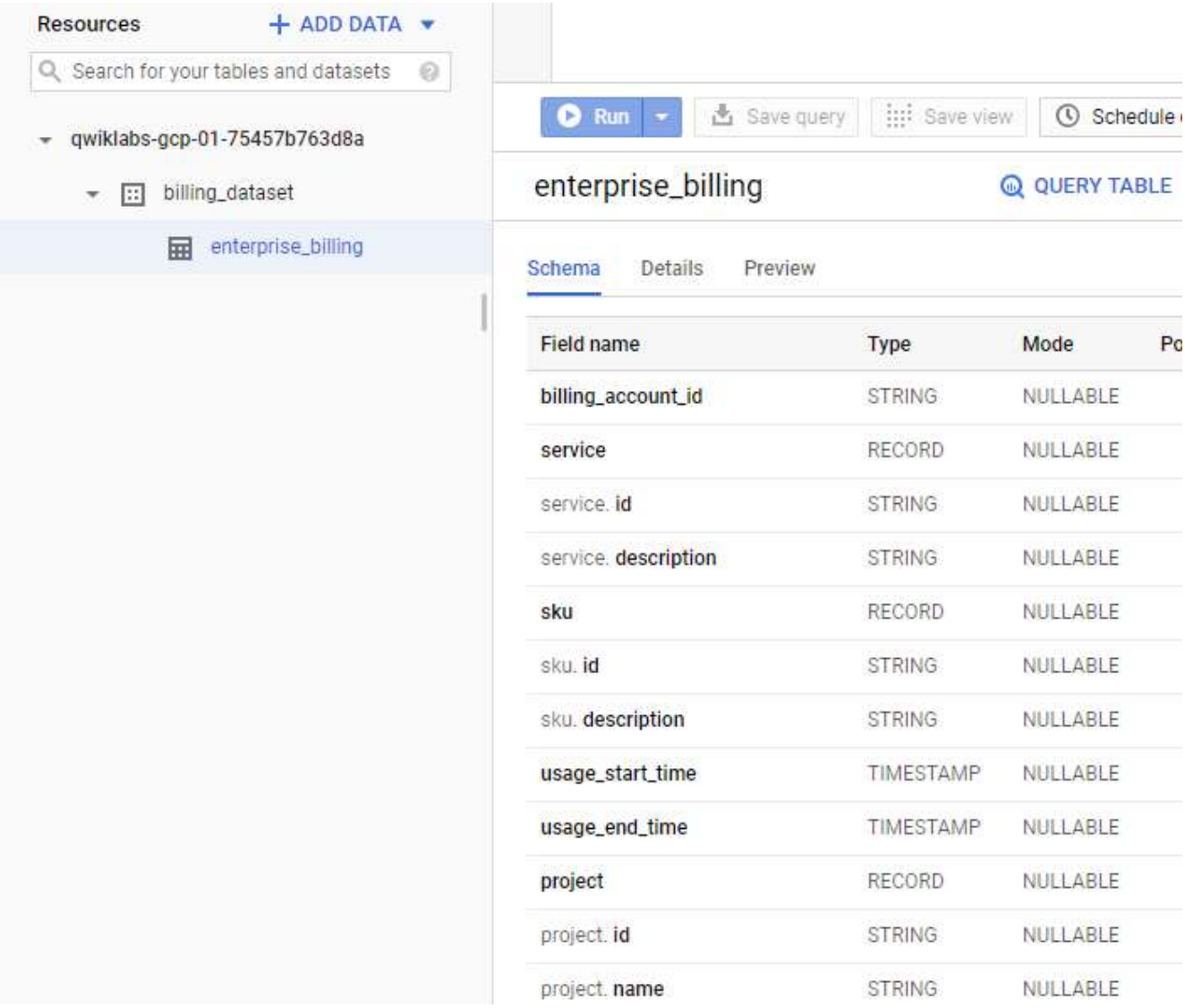


Examine the billing data

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.



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.

2. Click the **Details** tab for table information.

You see a table with 415,602 rows.

3. Click the **Preview** tab to view the table.

4. 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

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.

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.

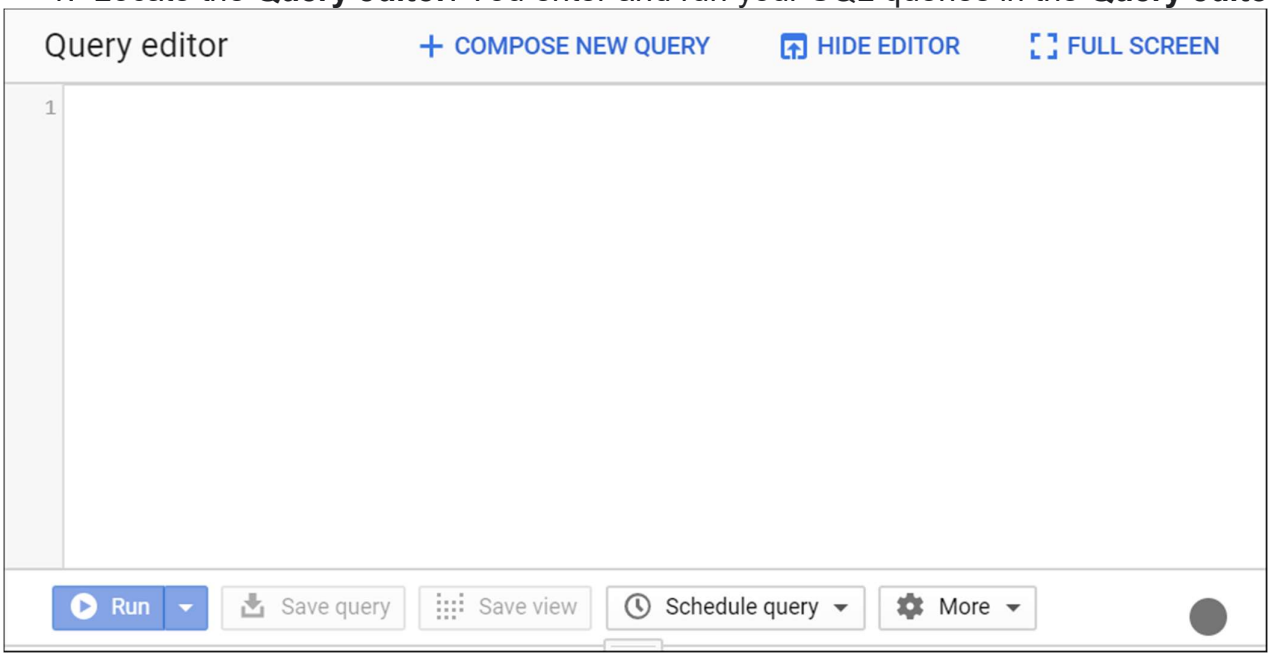
If you do not specify the Project, BigQuery defaults to the current project.
You can access all the information you need in the BigQuery interface. In the left panel, under **Resources**, you see the dataset (`billing_dataset`) and the table (`enterprise_billing`).

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**.

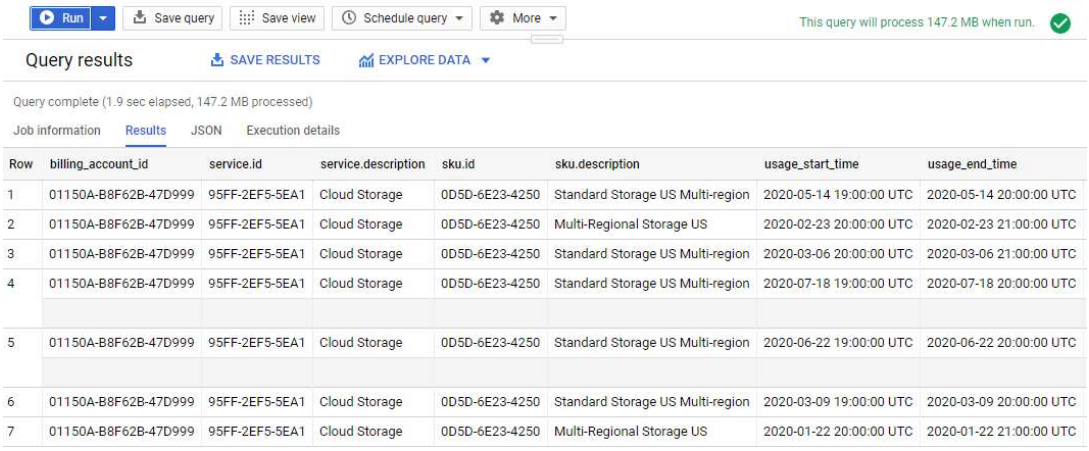


2. Type or paste the following into the **Query editor**:

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



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.

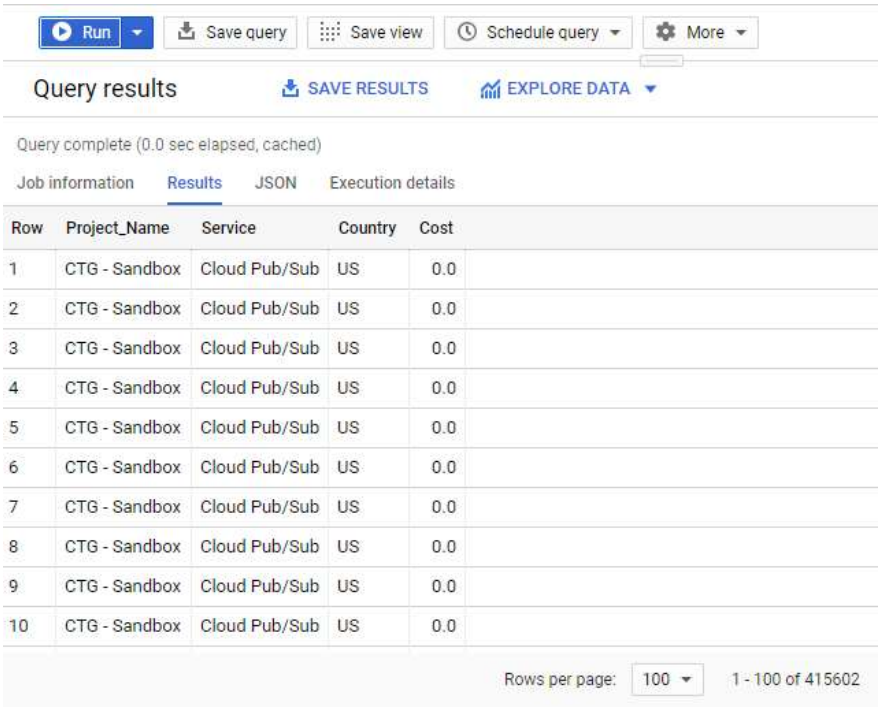
1. Click **COMPOSE NEW QUERY** to clear the **Query editor**.

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`;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 :



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

3. 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?
close~~service.id~~
close~~system_labels.value~~
close~~what_is_it.time~~
checkusage_start_time
Submit

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

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**.

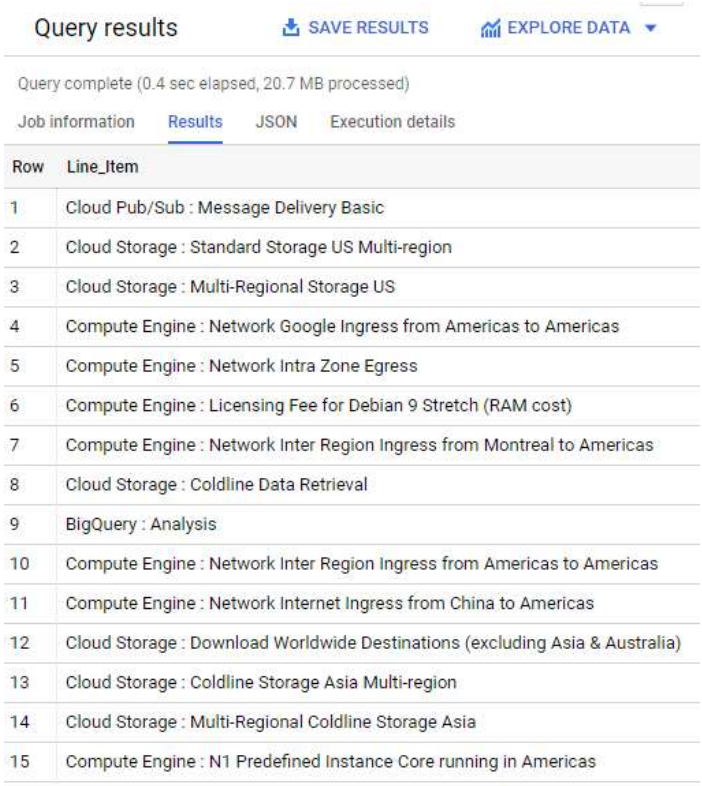
2. Type or paste the following into the **Query editor**:

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

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

3. Click **Run**.

You should receive the following output:



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

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

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

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)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?
0
What services produced 3349 logs in your billing record?
Cloud Functions : Network Egress from us-central1

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

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**.

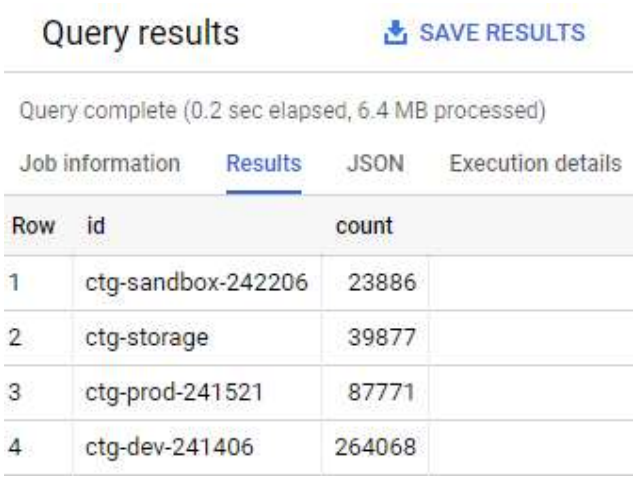
2. Type or paste the following into the **Query editor**:

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



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-dev-241406

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

Query 5: Find the cost per project

Find the cost breakdown for each project:

1. Click **COMPOSE NEW QUERY** to clear the **Query editor**.

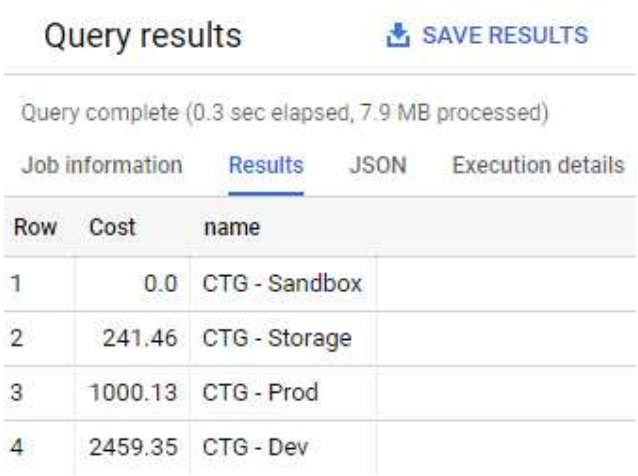
2. Type or paste the following into the **Query editor**:

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



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 - Dev

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

Congratulations!

In this lab, you explored a sample of Cloud Billing records in BigQuery. After examining the sample dataset and table, you composed and ran queries on the billing data. You then used those queries to answer pertinent billing questions. You are now ready to take more labs in this series.

Google Cloud Training & Certification

...helps you make the most of Google Cloud technologies. Our classes include technical skills and best practices to help you get up to speed quickly and continue your learning journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. Certifications help you validate and prove your skill and expertise in Google Cloud technologies.

Manual Last Updated October 07, 2020
Lab Last Tested October 07, 2020