

Ingesting New Datasets into BigQuery v1.5 | Google Cloud Skills Boost

Qwiklabs : 9-11 minutes

Overview

BigQuery is Google's fully managed, NoOps, low cost analytics database. With BigQuery you can query terabytes and terabytes of data without having any infrastructure to manage or needing a database administrator. BigQuery uses SQL and can take advantage of the pay-as-you-go model. BigQuery allows you to focus on analyzing data to find meaningful insights.

The dataset you'll use is an **ecommerce dataset** that has millions of Google Analytics records for the **Google Merchandise Store** loaded into BigQuery. You have a copy of that dataset for this lab and will explore the available fields and row for insights.

In this lab you will ingest several types of datasets into tables inside of BigQuery.

Setup and requirements

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, 1:15:00), 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.

Open BigQuery Console

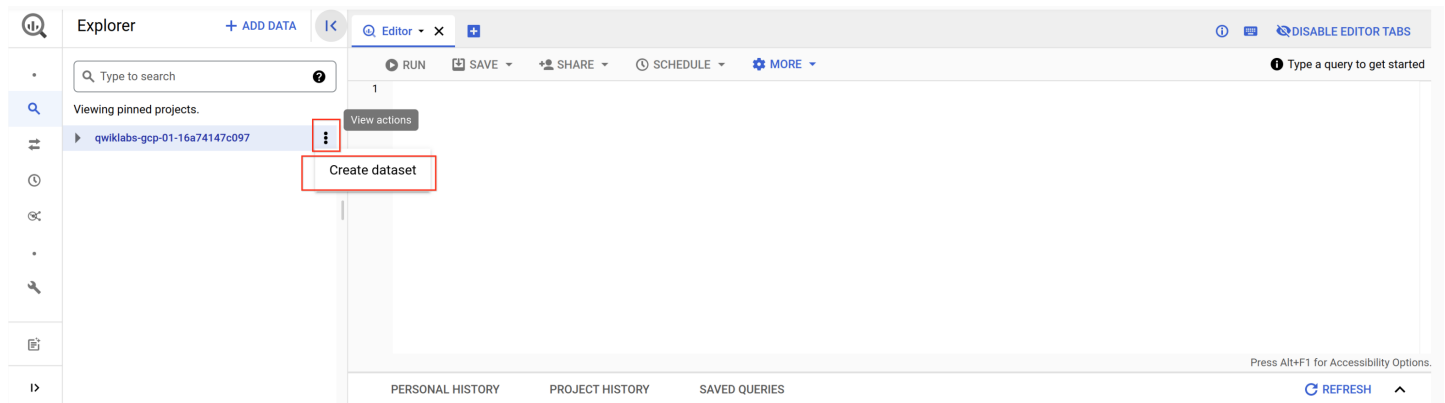
1. In the Google Cloud Console, select **Navigation menu > BigQuery**.

The **Welcome to BigQuery in the Cloud Console** message box opens. This message box provides a link to the quickstart guide and lists UI updates.

2. Click **Done**.

Task 1. Create a new dataset to store tables

1. In the BigQuery console, click on the 3 vertical dots next to your project, then click **Create Dataset**.



2. Set the *Dataset ID* to **ecommerce**. Leave the other fields at their default values.
3. Click **Create dataset**.

You'll now see the ecommerce dataset under your project name.

Task 2. Ingest a new dataset from a CSV

Scenario: Your marketing team is looking to you to help guide them with what products should be up for promotions based on inventory stock levels. They have also asked how each product is trending in customer sentiment based on the product reviews.

Your existing ecommerce transactional dataset does not have inventory stock levels or product review data in it, but your operations and marketing teams have provided you with new datasets for you to analyze.

Here is how you get started:

1. Download the [product stock level dataset](#) locally onto your computer.
2. Select the **ecommerce** dataset; then click **Create Table**.

The screenshot shows the Google Cloud BigQuery interface. On the left, the Explorer panel displays the project 'qwiklabs-gcp-04-bde0aae96b92' and the dataset 'ecommerce'. The 'ecommerce' dataset is highlighted with a red box and labeled '1.'. The main panel shows the details for the 'ecommerce' dataset, including its ID 'qwiklabs-gcp-04-bde0aae96b92:ecommerce', creation date 'Apr 6, 2021, 7:29:49 PM', and location 'US'. A red box labeled '2.' highlights the '+ CREATE TABLE' button in the top right corner of the dataset view.

3. Specify the below table options:

Source:

- Create table from: **Upload**
- Select file: select the file you downloaded locally earlier
- File format: **CSV**

Destination:

- Table name: **products**
- Leave other settings at their default value.

Schema:

- Check **Auto Detect** for Schema and input parameters

Tip: Not seeing the checkbox? Ensure the file format is CSV and not Avro.

- Partition and Cluster settings: Leave at default settings

Advanced Options:

- Leave at default settings

4. Click **Create Table**.

You should now see the **products** table below the ecommerce dataset.

5. Select the products table and view **details**:

The screenshot shows the Google Cloud BigQuery interface. On the left, the Explorer pane shows the 'products' table under the 'ecommerce' dataset, highlighted with a red box and labeled '1.'. The main Editor pane shows the 'products' table with a 'Details' tab selected, highlighted with a red box and labeled '2.'. The Table info section shows details for the 'products' table:

Table ID	Table size	Number of rows	Created	Table expiration	Last modified	Data location
qwklabs-gcp-04-bde0aae96b92:ecommerce.products	77.04 KB	1,090	Apr 6, 2021, 7:34:01 PM	Never	Apr 6, 2021, 7:34:01 PM	US

6. Select **Preview** and confirm all columns have been loaded (sampled below):

SKU	name	orderedQuantity	stockLevel	restockingLeadTime
GGOEGDHQ014899	20 oz Stainless Steel Insulated Tumbler	499	652	2
GGOEGOAB022499	Satin Black Ballpoint Pen	403	477	2
GGOEYHPB072210	Twill Cap	1429	1997	2
GGOEGEVB071799	Pocket Bluetooth Speaker	214	246	2

You have successfully loaded in a CSV file into a new BigQuery table.

Exploring newly loaded data with SQL

Next, practice with a basic query to gain insights from the new products table.

- In the **Query editor**, write a query to list the top 5 products with the highest stockLevel:

```
#standardSQL SELECT * FROM ecommerce.products ORDER BY stockLevel DESC LIMIT 5
```

Task 3. Ingest data from Google Cloud Storage

1. Select the ecommerce dataset and click **Create Table**.

2. Specify the below table options:

Source:

- Create table from: **Google Cloud Storage**
- Select file from GCS bucket: **data-insights-course/exports/products.csv**
- File format: **CSV**

Destination:

- Table name: **products**
- Leave all other settings as default.

Schema:

- Check **Auto Detect** for Schema and input parameters.

Advanced Options:

- Leave at default settings

3. Click **Create Table**.

Does it work? **No**

4. Click **Close** to close the message, then click **Cancel** in the **Create table** dialog.

5. In the left menu, click **Project history** and select the error message.

The screenshot shows the Google Cloud BigQuery interface. On the left, the 'Explorer' panel shows the 'ecommerce' dataset under the 'products' folder. The main panel displays the 'Dataset info' for 'ecommerce'. Below this, the 'PROJECT HISTORY' tab is selected, showing a table of jobs. The first job, 'bquxjob_49b1973e_180f62bdcdb', is highlighted with a red box and has a red error icon in the 'Job ID' column. The other two jobs, 'bquxjob_9f7eb75_180f629898c' and 'bquxjob_659ac92b_180f627ba14', are marked with green checkmarks.

Job ID	Creation time	Owner	Type	Summary	Session ID	Actions
bquxjob_49b1973e_180f62bdcdb	May 24, 2022, 6:35:21 PM	student-01-5f4da99ba88e...	LOAD			⋮
bquxjob_9f7eb75_180f629898c	May 24, 2022, 6:32:40 PM	student-01-5f4da99ba88e...	QUERY	#standardSQL SELECT * FROM ecommerce.pr...		⋮
bquxjob_659ac92b_180f627ba14	May 24, 2022, 6:30:42 PM	student-01-5f4da99ba88e...	LOAD			⋮

6. Click the **Repeat load job** button.

7. In the Create table form, click on **Advanced Options** and in the **Write Preference** dropdown menu, select **Overwrite table**.

8. Now click **Create Table**.

9. Confirm the table was executed successfully.

Task 4. Ingest a new dataset from a Google Spreadsheet

1. Select **Compose New Query**.
2. Execute this next query to show which products are in the greatest restocking need based on inventory turnover and how quickly they can be resupplied:

```
#standardSQL SELECT *, SAFE_DIVIDE(orderedQuantity,stockLevel) AS ratio FROM ecommerce.products WHERE # include products that have been ordered and # are 80% through their inventory orderedQuantity > 0 AND SAFE_DIVIDE(orderedQuantity,stockLevel) >= .8 ORDER BY restockingLeadTime DESC
```

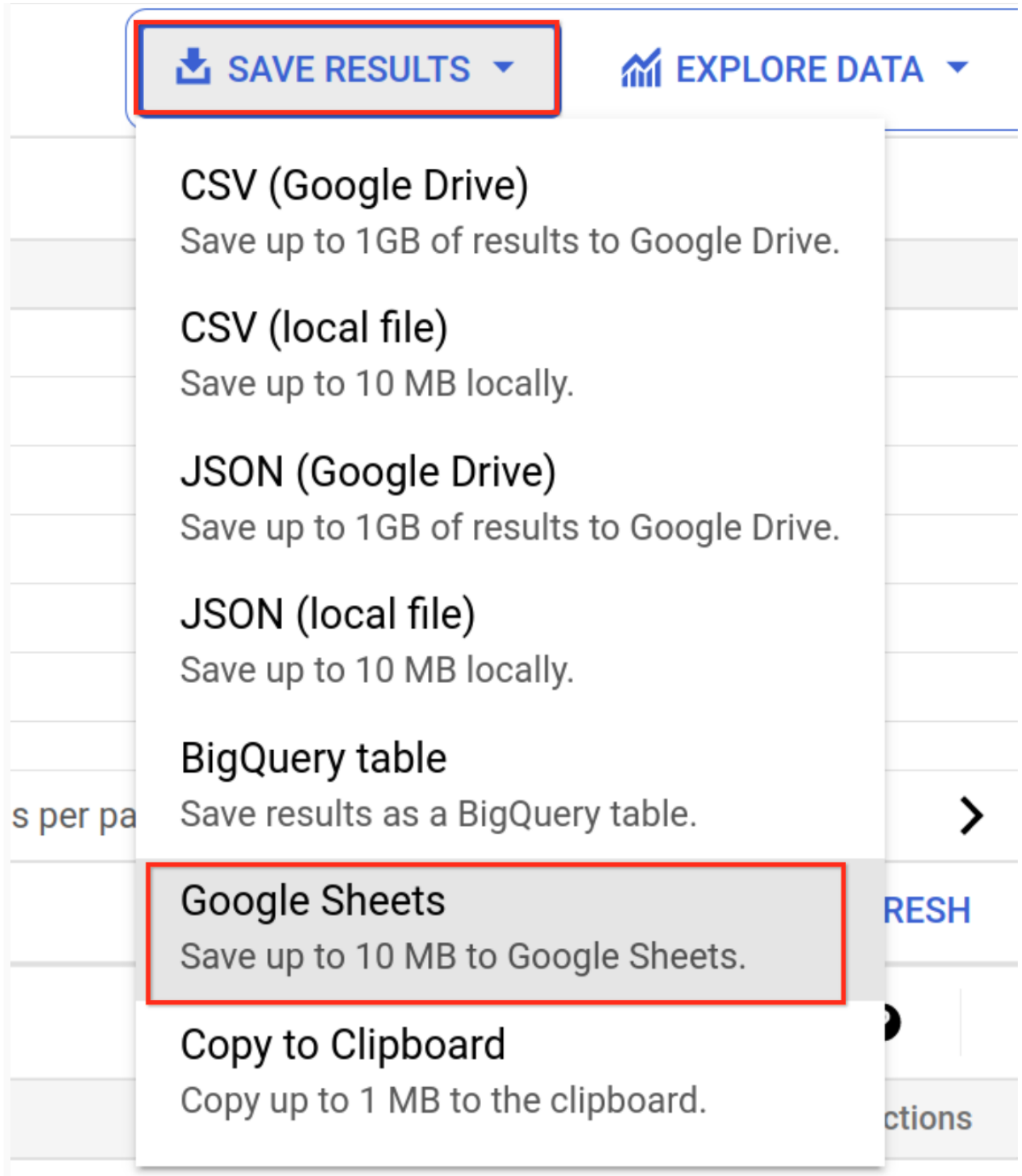
Note: If you specify a relative project name path like `ecommerce.products` instead of `project_id.ecommerce.products`, BigQuery will assume the current project.

Task 5. Saving Data to Google Sheets

Scenario: You want to provide your supply chain management team with a way to notate whether or not they have contacted the supplier to reorder inventory, and to make any notes on the items. You decide on using a Google Spreadsheet for a quick survey.

Now you'll create it:

1. In Query Results, select **Save Results > Google Sheets**.



The screenshot shows the 'SAVE RESULTS' dropdown menu in the BigQuery interface. The menu is open, displaying several options. The 'Google Sheets' option is highlighted with a red border. The 'SAVE RESULTS' button at the top of the menu is also highlighted with a red border. The background shows a table with columns and rows, and a 'REFRESH' button is visible on the right.

SAVE RESULTS ▼

EXPLORE DATA ▼

- CSV (Google Drive)**
Save up to 1GB of results to Google Drive.
- CSV (local file)**
Save up to 10 MB locally.
- JSON (Google Drive)**
Save up to 1GB of results to Google Drive.
- JSON (local file)**
Save up to 10 MB locally.
- BigQuery table**
Save results as a BigQuery table.
- Google Sheets**
Save up to 10 MB to Google Sheets.
- Copy to Clipboard**
Copy up to 1 MB to the clipboard.

2. A popup will appear with a link to Open the spreadsheet, select **Open**.
3. In your spreadsheet, in column G add a new field titled **comments** and for the first product row type new shipment on the way.

results-20180708-162222.csv ☆

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

100% \$ % .0 .00 123 Arial 10 B I S A

fx new shipment on the way

	A	B	C	D	E	F	G	H
1	SKU	name	orderedQuantity	stockLevel	restockingLeadTime	ratio	1	comments
2	GGOENEBB078	Cam Indoor Sec	2139	2615	42	0.81	2	new shipment on the way
3	GGOEGCBQ016	SPF-15 Slim & S	3682	4069	31	0.9048906365		
4	GGOEGOAQ020	Four Color Retra	2002	2450	22	0.8171428571		

4. In Google Sheets, select **Share** and **Get Shareable Link** then copy the link.

5. Return to your BigQuery tab.

6. Click on the **ecommerce** dataset, then **Create Table**.

7. Specify the these table options:

Source:

- Create table from: **Drive**
- Select Drive URI: put-your-spreadsheet-url-here
- File format: **Google Sheet**

Destination:

- Table type: Leave as default (External table)
- Table name: **products_comments**

Schema:

- Check **Auto Detect** for Schema and input parameters

Create table

Source

Create table from:

Drive

Select Drive URI:

https://drive.google.com/open?id=1cuRCDgG-8m5YPrp7lwQlewA€

File format:

Google Sh...

Destination

Project name

qwiklabs-gcp-d9ccd750c92585c3

Dataset name

ecommerce

Table type

External table

Table name

products_comments

Schema

Auto detect

☒ Schema and input parameters

i

 Schema will be automatically generated.

Advanced options

▼

Advanced options:

- Set **Header rows to skip:** to **1**.
- Leave all options at their default.

8. Click **Create Table**.

Note: There is no load job for creating an externally-linked table since the data is not ingested.

Query data from an external spreadsheet

1. Click **Compose New Query**.
2. Add the below query then **Run**:

```
#standardSQL SELECT * FROM ecommerce.products_comments WHERE comments IS NOT NULL
```

Wait for the query to execute. You will see that the new **comments** field is now returned.

SKU	name	orderedQuantity	stockLevel	restockingLeadTime	ratio	comments
GGOENEBB078899	Cam Indoor	2139	2615	42	0.8179732314	new shipment

Security

Camera

- USA

on the way

3. Navigate back to your Google Spreadsheet tab.
4. Type in more comments in the Comments field.
5. Navigate back to BigQuery and execute the query again by clicking **Run**.
6. Confirm the new data properly shows in the results.

You have successfully created an external table connection into BigQuery from Google Spreadsheets.

Task 6. External table performance and data quality considerations

Linking external tables to BigQuery (e.g. Google Spreadsheets or directly from Google Cloud Storage) has several [limitations](#). Two of the most significant are:

- Data consistency is not guaranteed if the data values in the source are changed while querying.
- Data sources stored outside of BigQuery lose the performance benefits of having BigQuery manage your data storage (including but not limited to auto-optimization of your query execution path, certain wildcard functions are disabled, etc.).

Congratulations!

You've successfully created a new dataset and ingested new external data sources into BigQuery from CSV, Google Cloud Storage, and Google Drive.

End your lab

When you have completed your lab, click **End Lab**. Google Cloud Skills Boost 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 2022 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.