

Building Data Warehouse Using Google Big Query

By Aditya Virgiansyah

Overview

- Data Warehouse Working
- Getting Started with Google Big Query
- Advantages of using Google Big Query
- Big Query Use Case Scenarios
- Building Data Warehouse in Big Query for E-commerce Scenario



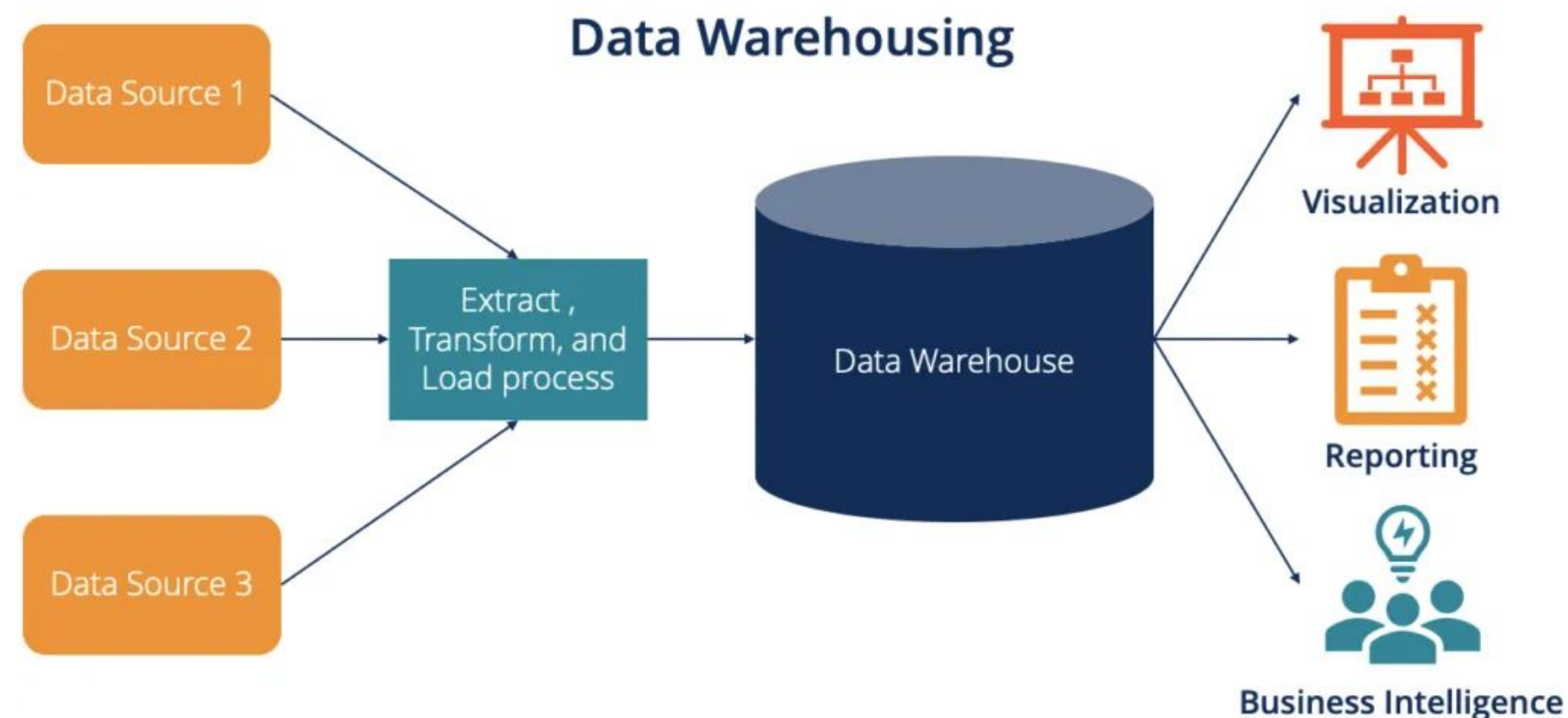
Data Warehouse Working

Briefly of the data warehouse



A data warehouse becomes your central hub, gathering info from various sources to create a reliable space for business insights. Picture this: data flows in regularly, maybe every hour or day, from different apps and systems. Once it's in, things get processed to match the existing data format in the warehouse.

The processed data is stored in the data warehouses and later used by the organizations for effective decision-making. The data processing procedure depends upon the data format. The data could be unstructured, semi-structured, or structured based on the source from which it is coming such as transaction processing systems, healthcare, supply chain, etc.



Source:

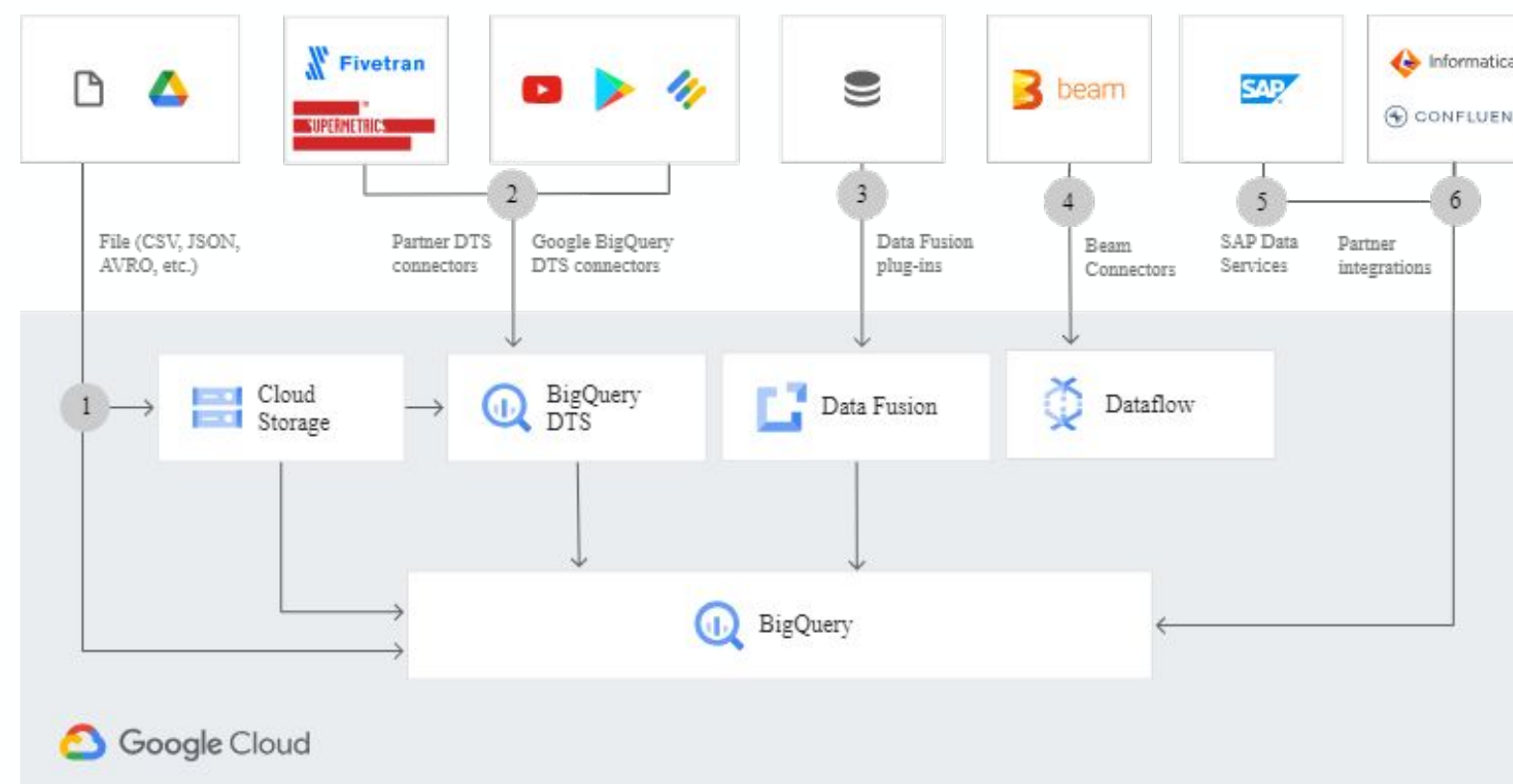
<https://corporatefinanceinstitute.com/resources/business-intelligence/data-warehousing/>

Getting Started with Google Big Query

In the realm of Google Cloud Platform, you're equipped with the dynamic capabilities of BigQuery. This fully managed, serverless Platform as a Service (PaaS) is not just a data warehouse; it's a centralized hub that effortlessly pulls in data from diverse sources like Cloud Storage, Cloud SQL, Amazon S3, Azure Blob Storage, and more.

With features like federated queries, you can directly analyze external data within BigQuery without the need for importing, while its robust SQL query engine and built-in machine learning capabilities make data analysis a seamless experience. BigQuery isn't just about creating a data warehouse; it's about crafting a highly scalable solution that offers prescriptive and descriptive analysis, serves as a powerful data storage solution, and acts as a centralized resource for managing both your data and computing needs efficiently. Discover the possibilities and scalability of Google BigQuery for a transformative data management experience.

using Google Cloud, think of Identity and Access Management (IAM) as your security guard. It helps keep your data safe and in check. In BigQuery, the jobs that do things like exporting or loading data follow specific rules (roles) to make sure only the right folks can do certain tasks. Plus, in BigQuery, storage and computing are like two best friends who work together but have their own space. This makes it super flexible when you're setting up a big data warehouse. It's all about keeping things secure, organized, and easy to manage!



Source: <https://cloud.google.com/bigquery>

Advantages of using Google Big Query



01

Enhanced business
intelligence capabilities

02

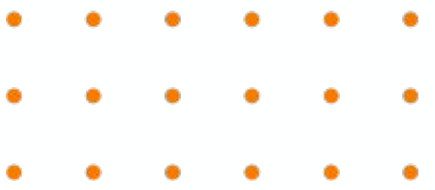
Data consistency and
enhanced data quality

03

Security

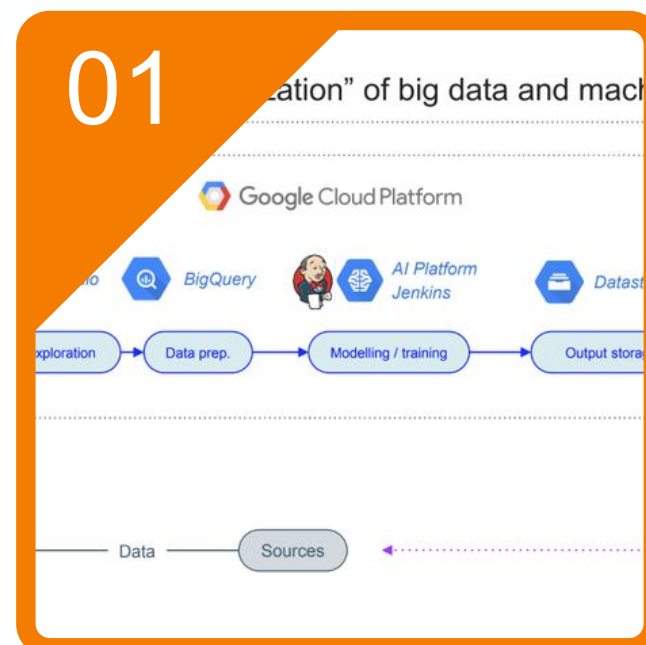
04

Disaster recovery & fault
tolerance



Big Query Use Case Scenarios

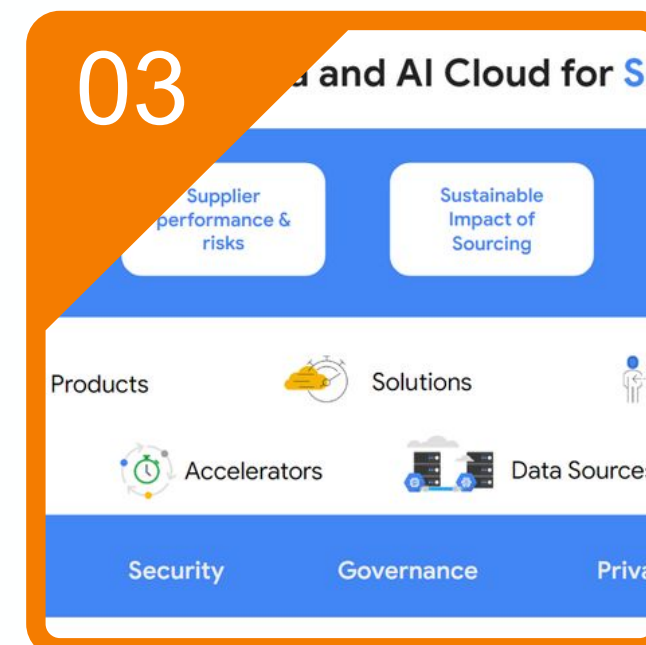
Big Query Use Case Scenarios



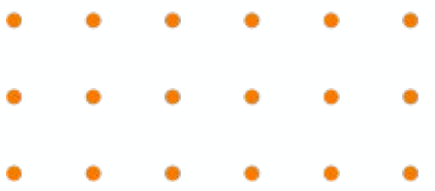
E-commerce
recommendation system



Migrating data
warehouses



Supply chain analytics



Building Data Warehouse in Big Query for E-commerce Scenario



Case

We are assuming that we have an e-commerce company that stores data from two different sources, i.e., mobile data and mobile brands scraped data from Flipkart, based on the type of details the company needs to store for performing peer performance analysis.

link dataset:

<https://www.kaggle.com/datasets/devsubhash/flipkart-mobiles-dataset>

Free trial status: Rp4,781,551.00 credit and 69 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

Google Cloud

My First Project

Search (/) for resources, docs, products, and more

Explorer

+ ADD

<

Type to search

?

Viewing resources.

SHOW STARRED ONLY

plucky-hue-391709

External connections

mydataset

WELCOME TO BIGQUERY STUDIO!

Get started

COMPOSE A NEW QUERY

ADD

Recently accessed

zip_codes

US : cylistic-405703

OPEN

avocados_cl...

US : plucky-hue-391709

OPEN

avocados_p...

US : plucky-hue-391709

OPEN

avocado_ba...

US : plucky-hue-391709

OPEN

avocados

US : plucky-

OPEN

Try with sample data

Try the Google Trends Demo Query

This simple query generates the top search terms in the US from the Google Trends public dataset.

OPEN THIS QUERY

VIEW DATASET

Add your own data

Local file

Upload a local file

LAUNCH THIS GUIDE

Google Drive

Google storage service

LAUNCH THIS GUIDE

Google Cloud Storage

Google object storage service

LAUNCH THIS GUIDE

Job history

CREATE DATASET

CANCEL

Create dataset

Project ID

plucky-hue-391709

CHANGE

Dataset ID *

flipkart_ecommerce

Letters, numbers, and underscores allowed

Location type

Region

Specify a region to colocate your datasets with other Google Cloud services.

Multi-region

Allow BigQuery to select a region within a group to achieve higher quota limits.

Multi-region *

US (multiple regions in United States)

Default table expiration

Enable table expiration

Default maximum table age

Days

Advanced options

Create new project in bigquery (GCP)

Free trial status: Rp4,781,551.00 credit and 69 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DISMISSACTIVATE

Google CloudMy First ProjectSearch (/) for resources, docs, products, and moreSearch

Explorer+ ADD<

•
Type to search
?
Viewing resources.
SHOW STARRED ONLY
▼ plucky-hue-391709
▶ External connections
▶ flipkart_ecommerce
▶ mydataset
☆
☆
☆
Open
Open in
Create table
Share
Copy ID
Refresh contents
Delete

SUMMARY
Nothing currently selected
Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM, and creating Python notebooks with Colab.
ENABLENOWDISMISS

Welcome to BigQuery Studio!

Get started
+ COMPOSE A NEW QUERY+ ADD

Recently accessed
des ☆
istic-405703
avocados_cl... ☆
US : plucky-hue-391709
OPEN
avocados_p... ☆
US : plucky-hue-391709
OPEN
avocado_ba... ☆
US : plucky-hue-391709
OPEN
avocados ☆
US : plucky-hue-391709
OPEN

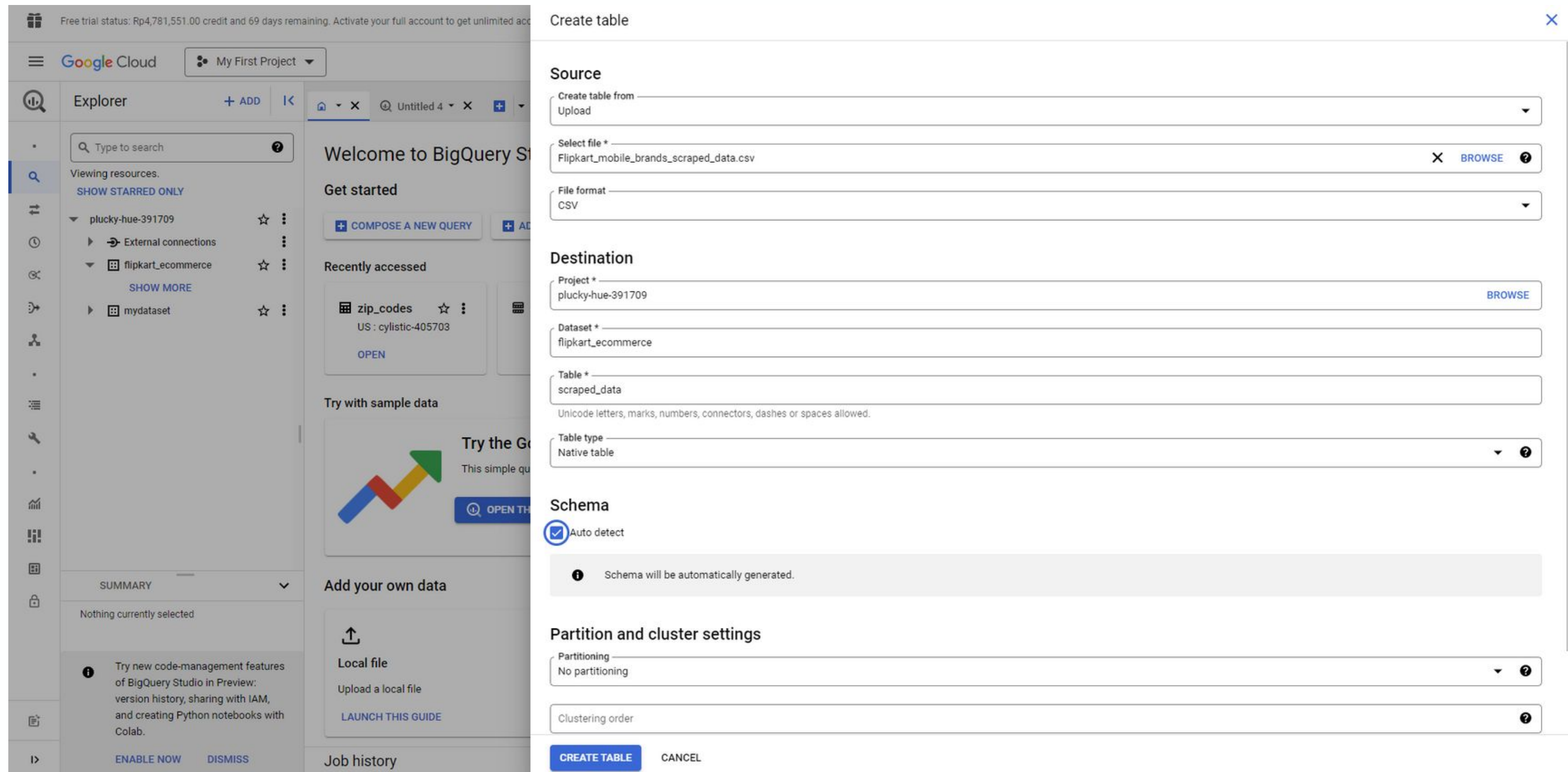
Try the Google Trends Demo Query
This simple query generates the top search terms in the US from the Google Trends public dataset.
OPEN THIS QUERYVIEW DATASET

Add your own data
Local file
Upload a local file
LAUNCH THIS GUIDE
Google Drive
Google storage service
LAUNCH THIS GUIDE
Google Cloud Storage
Google object storage service
LAUNCH THIS GUIDE

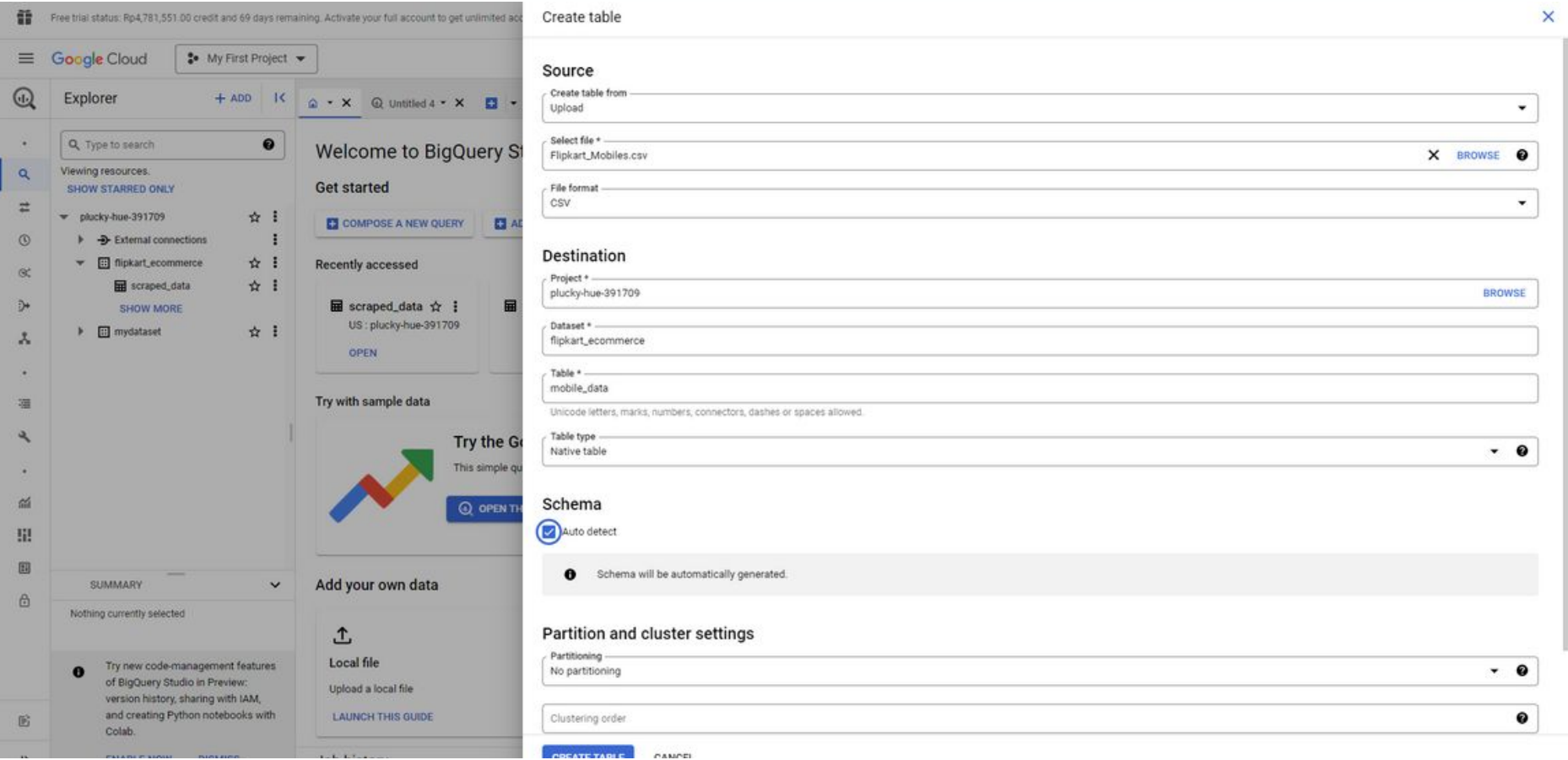
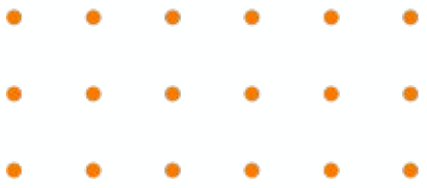
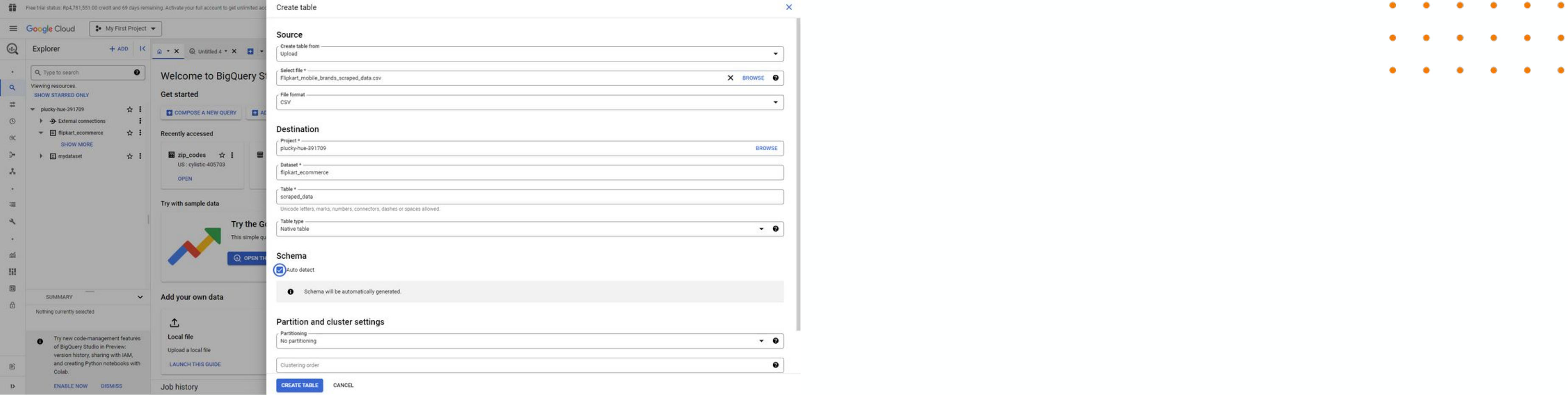
Job history

REFRESH^

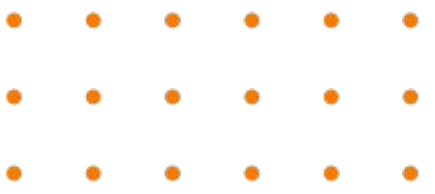
Create new dataset in bigquery (GCP)



Click Resource-> Big Query-> ProjectName
-> Create dataset.



Creating tables inside the dataset in Big Query



Analyzing data format
and schema of
imported tables

Free trial status: Rp4,781,551.00 credit and 69 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DISMISSACTIVATE

Google CloudMy First Project

Search (/) for resources, docs, products, and moreSearch

Explorer

+ ADD

I<

mobile_data

scraped_data

*Untitled 4

+ +

home

info

chat

lightbulb

full screen

Untitled 4

RUN

SAVE

DOWNLOAD

SHARE

SCHEDULE

MORE

This query will process 202.05 KB when run.

1 SELECT * FROM `plucky-hue-391709.flipkart_ecommerce.mobile_data` LIMIT 100;

Press Alt+F1 for Accessibility Options.

Query results

SAVE RESULTSEXPLORE DATA

JOB INFORMATIONRESULTSCHARTPREVIEWJSONEXECUTION DETAILSEXECUTION GRAPH

Row	Brand	Model	Color	Memory	Storage	Rating	Selling_Price	Original_Price
1	realme	1	Diamond Black	6 GB	null	4.3	10999	14990
2	realme	1	Moonlight Silver	3 GB	null	null	10999	12990
3	realme	1	Black	4 GB	null	4.4	10990	14990
4	realme	1	Moonlight Silver	6 GB	null	4.3	10990	14990
5	realme	1	Diamond Black	3 GB	null	null	11990	12990
6	realme	1	Diamond Red	3 GB	null	null	11900	12990
7	realme	2	Diamond Black	3 GB	null	4.5	9990	9990
8	realme	2	Diamond Red	3 GB	null	4.5	9990	9990
9	realme	3	Diamond Red	3 GB	null	4.5	8499	10999
10	realme	6	Comet White	6 GB	null	4.4	14999	17999

Results per page: 501 – 50 of 100

REFRESH

SUMMARY

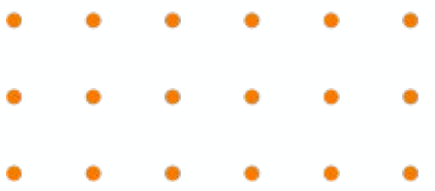
Nothing currently selected

Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM, and creating Python notebooks with Colab.

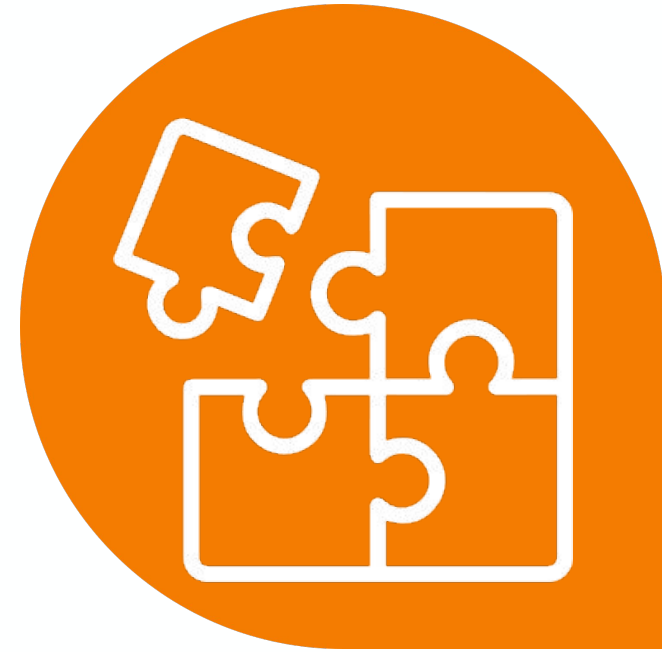
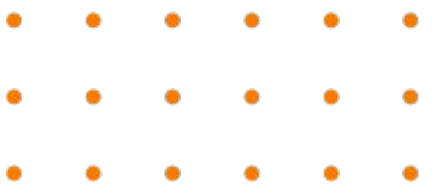
ENABLE NOWDISMISS

Job history

See 100 records from the mobile_data table



From the schema, we can see that the tables `mobile_data` and `scraped_data` have the same data schema, i.e., both have the fields `Brand`, `Model`, `Color`, `Memory`, `Storage`, `Rating`, `Selling_Price`, and `Original_Price`



Aggregating from
imported tables

Free trial status: Rp4,781,551.00 credit and 69 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DISMISSACTIVATE

Google CloudMy First Project

Search (/) for resources, docs, products, and more

Search

🏠 ⓘ 🗨️ 💡 ⌵ 🗲

Explorer

+ ADD<

🔍 Type to search

Viewing resources.

SHOW STARRED ONLY

plucky-hue-391709

External connections

flipkart_ecommerce

mobile

mobile_data

scraped_data

SHOW MORE

mydataset

SUMMARY

Nothing currently selected

Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM, and creating Python notebooks with Colab.

ENABLE NOWDISMISS

mobile_data

scraped_data

*Untitled 4

🔍 RUNSAVEDOWNLOADSHARESCHEDULEMORE

Query completed.

1 # Aggregating data from the imported tables to create a new table in Big Query

2 CREATE TABLE `flipkart_ecommerce.mobile` AS SELECT * FROM `plucky-hue-391709.flipkart_ecommerce.mobile_data`

3 UNION ALL

4 SELECT * FROM `plucky-hue-391709.flipkart_ecommerce.scraped_data`;

Press Alt+F1 for Accessibility Options.

Query results

SAVE RESULTSEXPLORE DATA

JOB INFORMATIONRESULTSEXECUTION DETAILSEXECUTION GRAPH

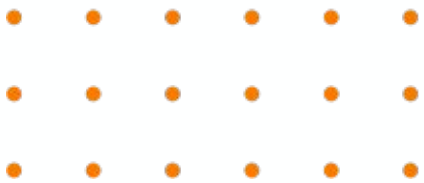
This statement created a new table named mobile.

GO TO TABLE

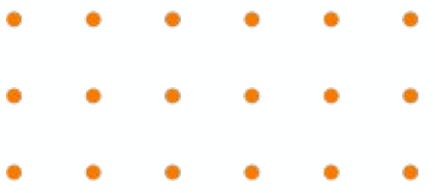
Job history

REFRESH

Perform data aggregation by creating a new table by performing the *UNION operation*



The screenshot shows the BigQuery Studio interface. On the left is a sidebar with navigation icons: a list icon, a key icon, a small square icon, a bar chart icon, a table icon, a grid icon, and a lock icon. Below these is a 'Code' icon and a terminal icon labeled 'I>'. The main panel has a 'SUMMARY' tab selected. Below the tab, it says 'Nothing currently selected'. A notification banner at the bottom right contains an information icon, the text 'Try new code-management of BigQuery Studio in Preview', and two buttons: 'ENABLE NOW' and 'DISMISS'.



Getting useful insights

The screenshot shows the BigQuery Studio interface. On the left is a sidebar with various navigation icons. The main panel displays a 'SUMMARY' tab with the text 'Nothing currently selected'. Below this, a notification banner informs the user about new code-management features of BigQuery Studio in Preview, including version history, sharing with IAM, and creating Python notebooks with Colab. At the bottom of the notification are two buttons: 'ENABLE NOW' and 'DISMISS'.

The screenshot shows the Google Cloud BigQuery Studio interface. On the left is a vertical sidebar with icons for Explorer, Recent, Search, Tables, Views, Jobs, Datasets, Tables, Views, Jobs, and a lock icon. The main panel displays a 'SUMMARY' section with the text 'Nothing currently selected'. Below this, a message box states: 'Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM, and creating Python notebooks with Colab.' At the bottom of the message box are two buttons: 'ENABLE NOW' and 'DISMISS'.

The screenshot shows the BigQuery Studio interface. On the left is a sidebar with navigation icons: a person icon, a list icon, a key icon, a bar chart icon, a table icon, a document icon, and a terminal icon. The main panel displays a 'SUMMARY' tab with a dropdown arrow. Below the tab, it says 'Nothing currently selected'. At the bottom of the main panel, there is a notification box with an information icon and the text: 'Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM, and creating Python notebooks with Colab.' Below the notification are two buttons: 'ENABLE NOW' and 'DISMISS'.

Free trial status: Rp4,781,551.00 credit and 69 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DISMISS

ACTIVATE

Google Cloud

My First Project

Search (/) for resources, docs, products, and more

Search

Explorer

+ ADD

<

Type to search

?

Viewing resources.

SHOW STARRED ONLY

plucky-hue-391709

External connections

flipkart_ecommerce

mobile

SHOW MORE

mydataset

SUMMARY

Nothing currently selected

Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM, and creating Python notebooks with Colab.

ENABLE NOW

DISMISS

Untitled 4

RUN

SAVE

DOWNLOAD

SHARE

SCHEDULE

MORE

Query completed.

1 # Analyze data

2 # mobile phones which have selling price same as that of the original price

3 SELECT * FROM 'plucky-hue-391709.flipkart_ecommerce.mobile' WHERE Selling_Price = Original_Price;

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION

RESULTS

CHART

PREVIEW

JSON

EXECUTION DETAILS

EXECUTION GRAPH

Row	Brand	Model	Color	Memory	Storage	Rating	Selling_Price	Original_Price
1	realme	2	Diamond Red	3 GB	null	4.5	9990	9990
2	realme	2	Diamond Black	3 GB	null	4.5	9990	9990
3	realme	X	Space Blue	4 GB	null	4.5	17999	17999
4	realme	X	Space Blue	8 GB	null	4.5	20999	20999
5	realme	5i	Aqua Blue	4 GB	null	4.5	10999	10999
6	realme	X2	Pearl White	4 GB	null	4.5	17999	17999
7	realme	X2	Pearl Green	8 GB	null	4.5	20999	20999
8	SAMSUNG	Galaxy M20	Ocean Blue	32 GB	null	4.2	9990	9990
9	Apple	iPhone 6s Plus	Rose Gold	2 GB	null	4.4	34900	34900
10	Annie	iPhone 6s Plus	Gold	2 GB	null	4.4	49999	49999

Results per page: 50

1 - 50 of 1920

<

>

REFRESH

Job history

mobile phones who has selling price same
as original price

Free trial status: Rp4,781,551.00 credit and 69 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DISMISS

ACTIVATE

Google Cloud

My First Project

Search (/) for resources, docs, products, and more

Search

🏠 ⓘ 🗨️ 💡 🔍

Explorer

+ ADD

⏪

🔍

Type to search

?

Viewing resources.

SHOW STARRED ONLY

plucky-hue-391709

External connections

flipkart_ecommerce

mobile

SHOW MORE

mydataset

🔍

Untitled 4

RUN

SAVE

DOWNLOAD

SHARE

SCHEDULE

MORE

This query will process 360.13 KB when run.

1 # Analyze data

2 # Finding the details of mobile phones which have Memory greater than 6 GB

3 SELECT

4 *

5 FROM

6 `plucky-hue-391709.flipkart_ecommerce.mobile`

7 WHERE

8 ((CAST(SUBSTRING(Memory,1,length(Memory)-2) AS FLOAT64)>6))

9 AND (RIGHT(Memory,2)="GB");

Query results

SAVE RESULTS

EXPLORE DATA

⌵

JOB INFORMATION

RESULTS

CHART

PREVIEW

JSON

EXECUTION DETAILS

EXECUTION GRAPH

Row	Brand	Model	Color	Memory	Storage	Rating	Selling_Price	Original_Price
1	realme	X	Space Blue	8 GB	null	4.5	20999	20999
2	realme	X	Space Blue	8 GB	null	4.5	20999	null
3	realme	X2	Pearl Green	8 GB	null	4.5	20999	null
4	realme	X2	Pearl Green	8 GB	null	4.5	20999	20999
5	realme	8 Pro	Illuminating Yellow	8 GB	null	4.4	19999	21999
6	realme	8 Pro	Illuminating Yellow	8 GB	null	4.4	19999	21999
7	GIONEE	Ctrl V5	White	8 GB	null	3.8	4790	null
8	SAMSUNG	Galaxy M20	Ocean Blue	32 GB	null	4.2	9990	null
9	SAMSUNG	Galaxy M20	Ocean Blue	32 GB	null	4.2	9990	9990
10	realme	X3 SuperZoom	Arctic White	8 GB	null	4.3	29999	31999

Results per page: 50

1 – 50 of 726

⏪ ⏩

Job history

REFRESH

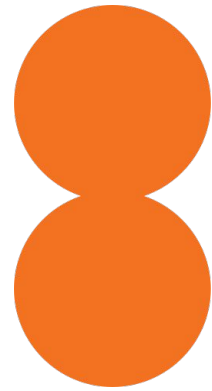
⏶

Try new code-management features of BigQuery Studio in Preview: version history, sharing with IAM, and creating Python notebooks with Colab.

ENABLE NOW

DISMISS

mobile phones who has memory more than
6GB



GOT QUESTIONS?

Reach out.



085722450160



adityavirgiansyah3@gmail.com



<https://www.linkedin.com/in/adityavirgiansyah/>