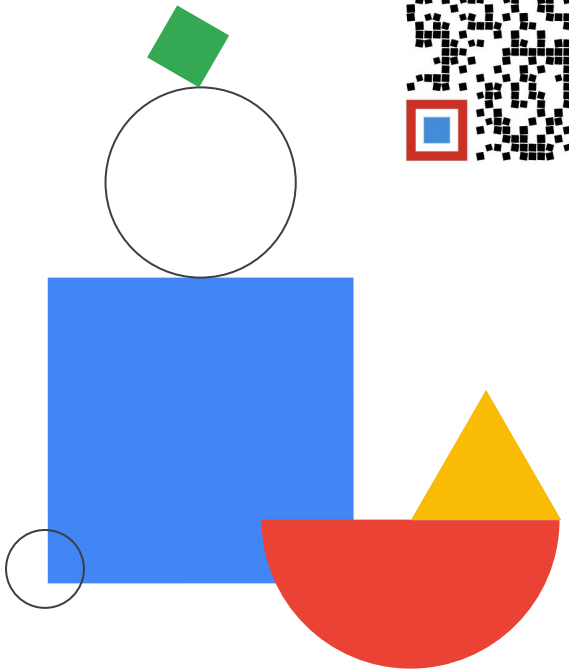
Google Cloud

SCAN to REGISTER your Attendance

BigQuery Demo



Trainer: Paggie Chen

Agenda

01

How to Create a Table

Native table
From public dataset
External table (through biglake connection or not)

02

Core Concepts

Project, Dataset, Table
Partitioned & Clustered table
Table & View
Authorized view (column-level access control)
Row-level access control
Policy Tag
Data Lineage

03

Build Pipelines

Data Canvas
Dataform

04

Augment with BigQuery ML

Scenario 1: [Prebuilt API] Document AI with BQML
Scenario 2: [Vertex AI] Image Data Summarization with BQML
Scenario 3: [Native Model] Custom BQML Model

05

Data Governance

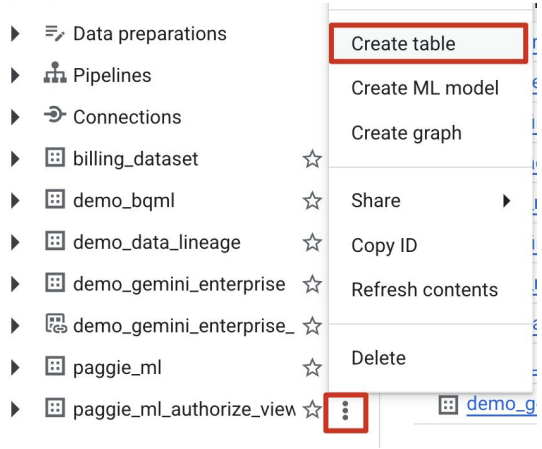
Data Profile, Data Quality Scan

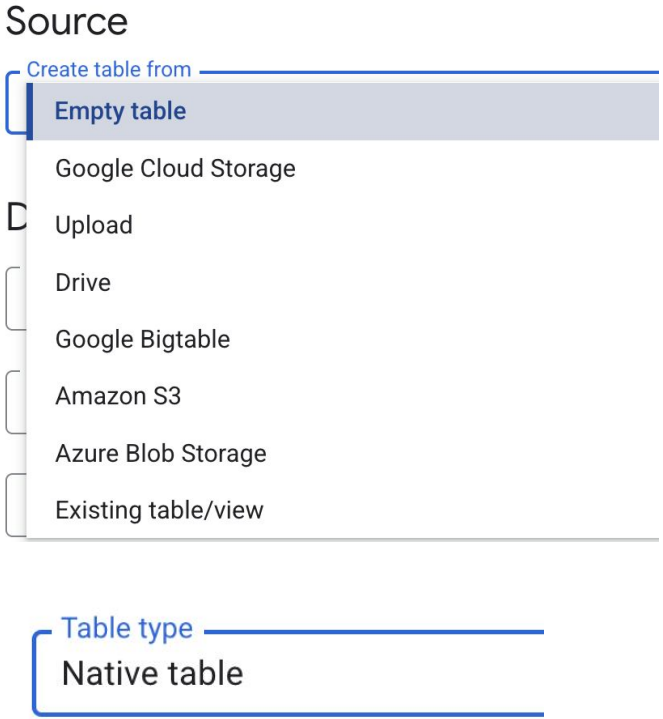
How to Create a Table

Native table

From public dataset

External table (through a connection or not)



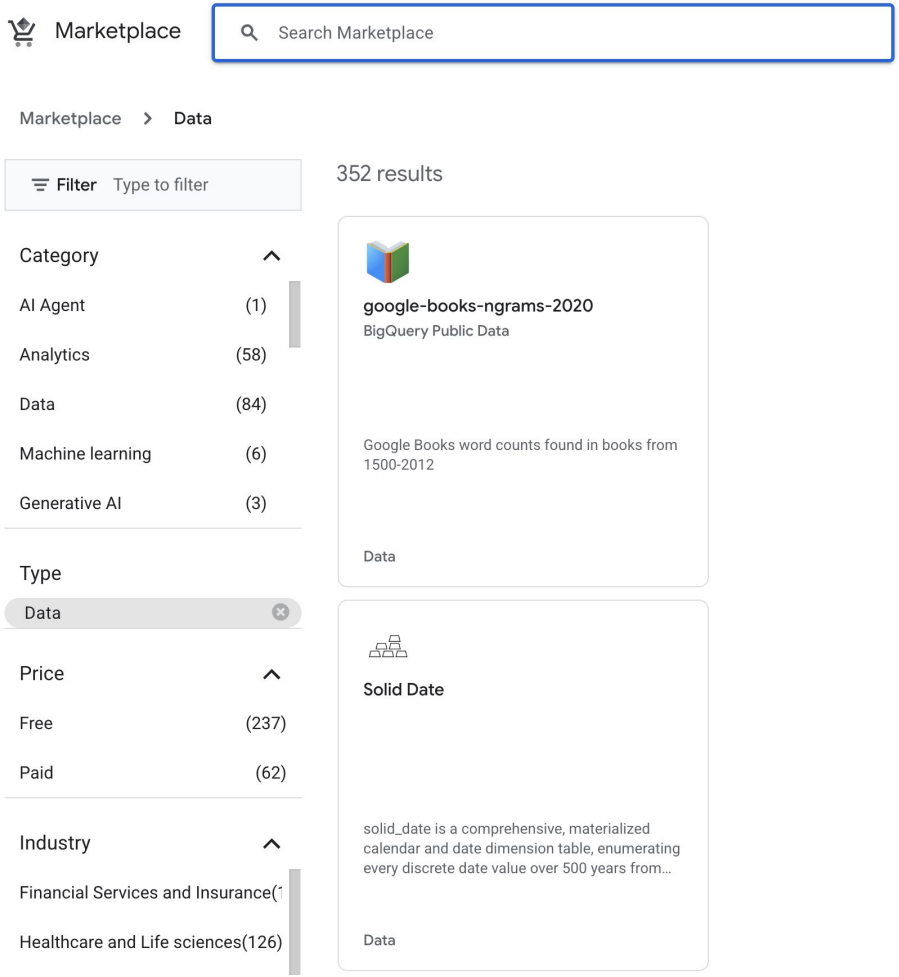


How to Create a Table

Native table

From public dataset

External table (through a connection or not)



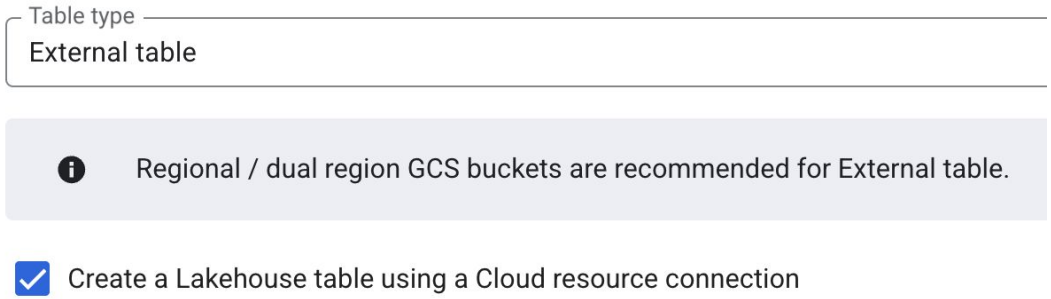
How to Create a Table

Native table

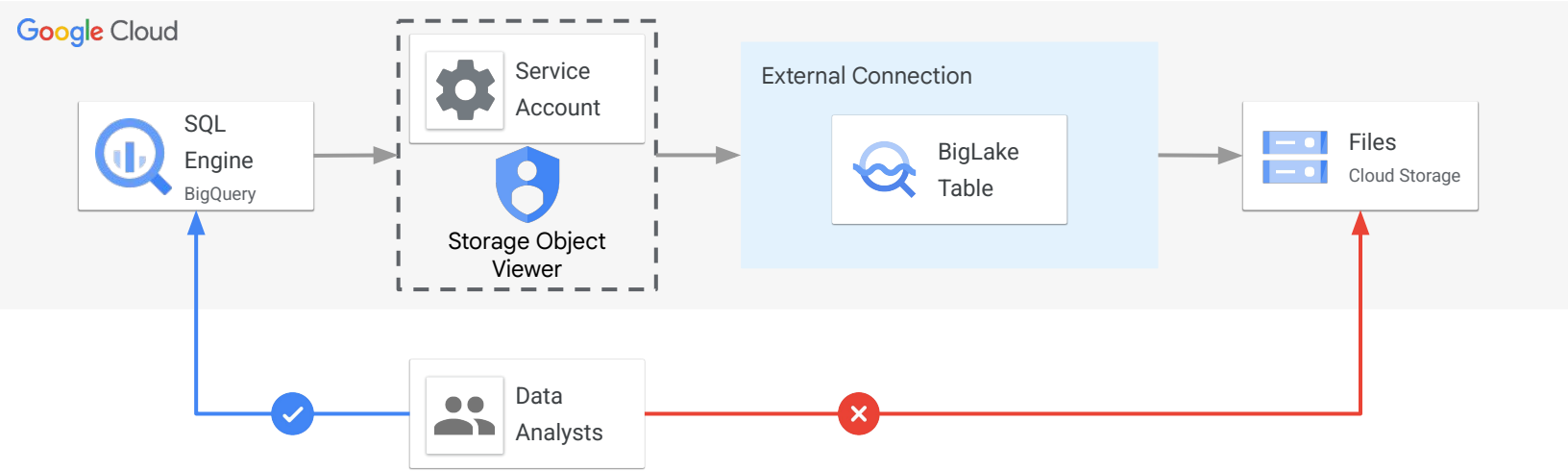
From public dataset

External table

(through a connection or not)



Access delegation enforces fine-grained security



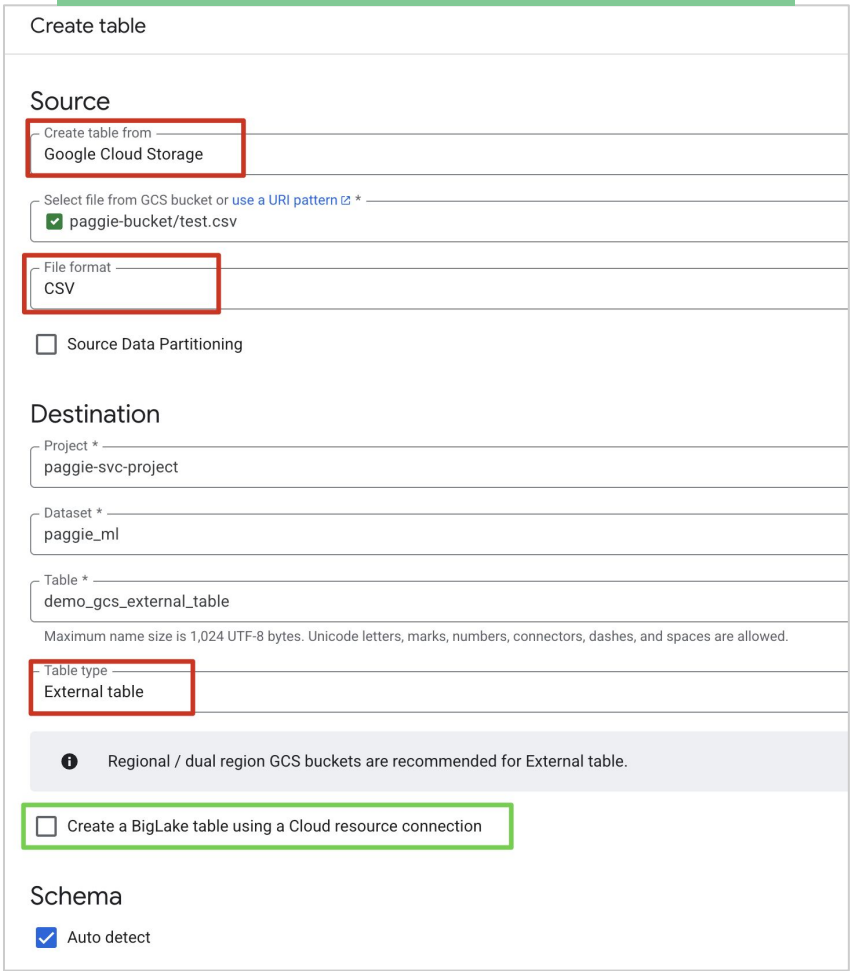
Access delegation

- Decouples access to BigLake table from underlying data store, fine-grained security at table level.
- An external connection associated with a service account is used to connect to the data store.

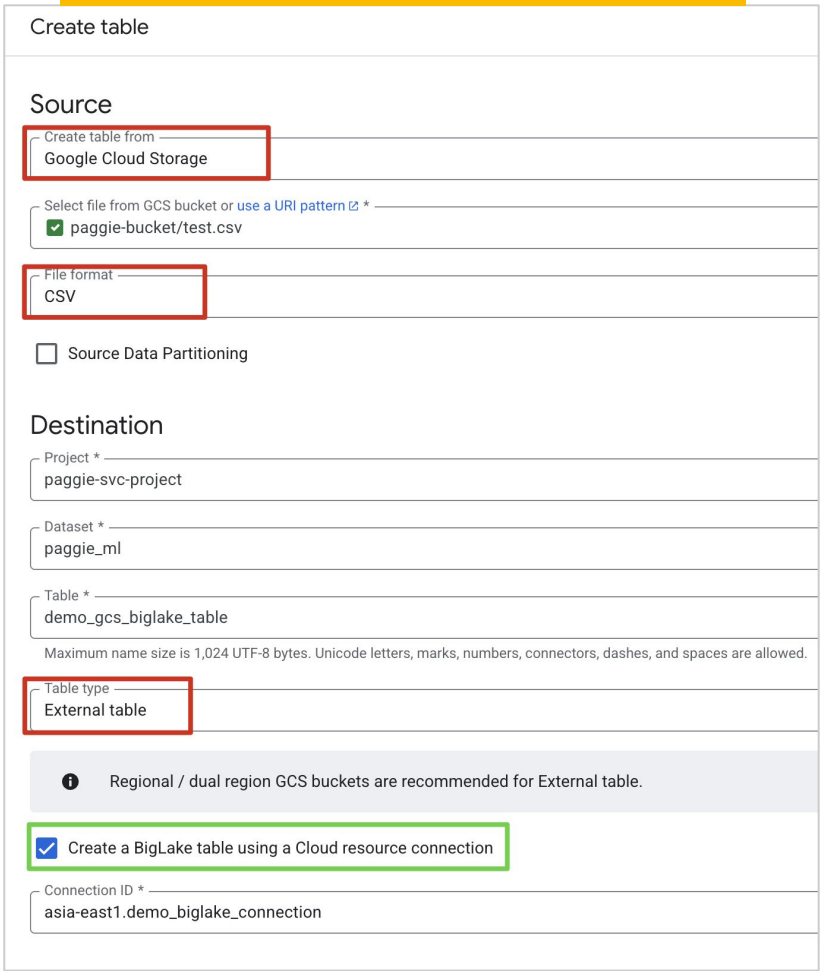
Caution

- Users should not have access to the objects directly in Cloud Storage.
- Direct file access lets users bypass governance policies.

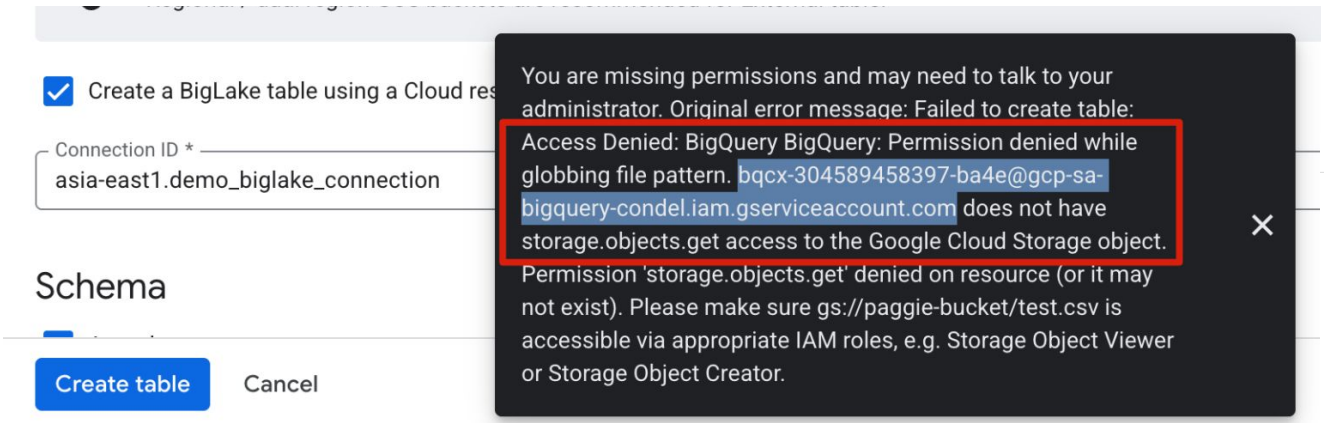
Non-BigLake Table (Creation)



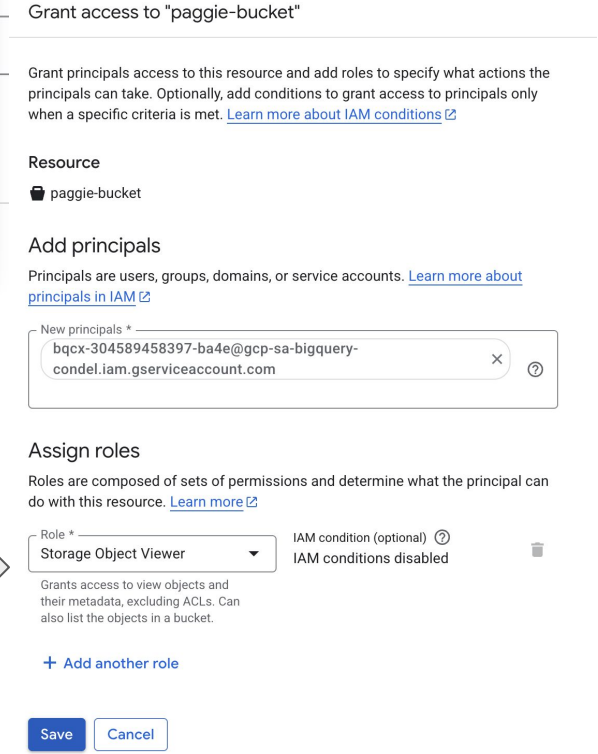
BigLake Table (Creation)



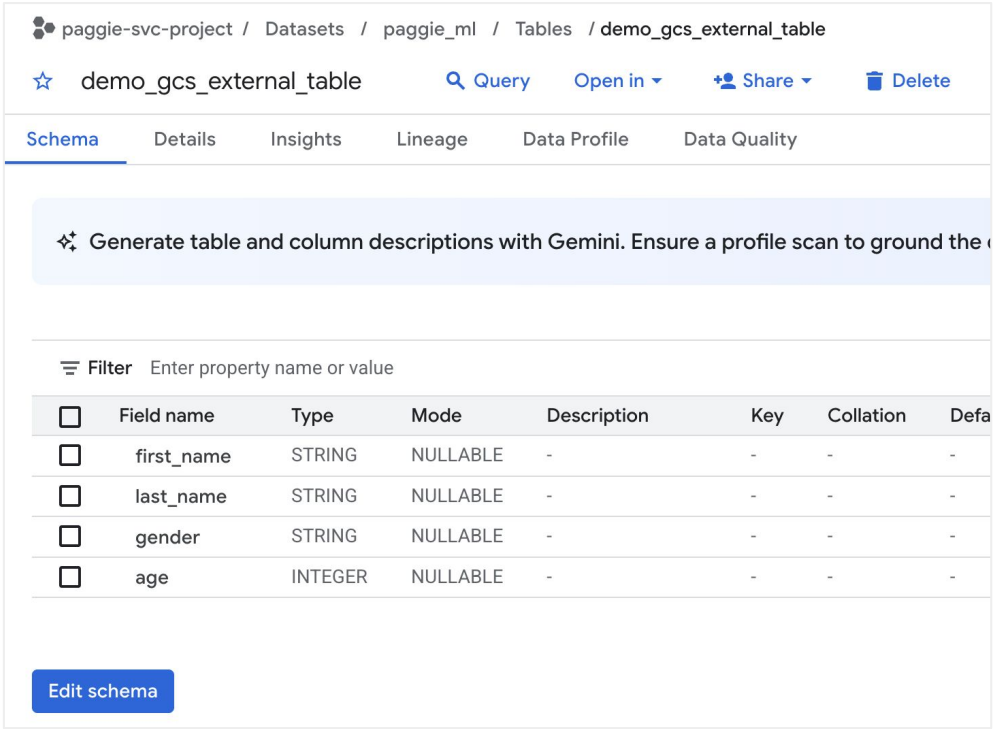
BigLake Table (Creation)



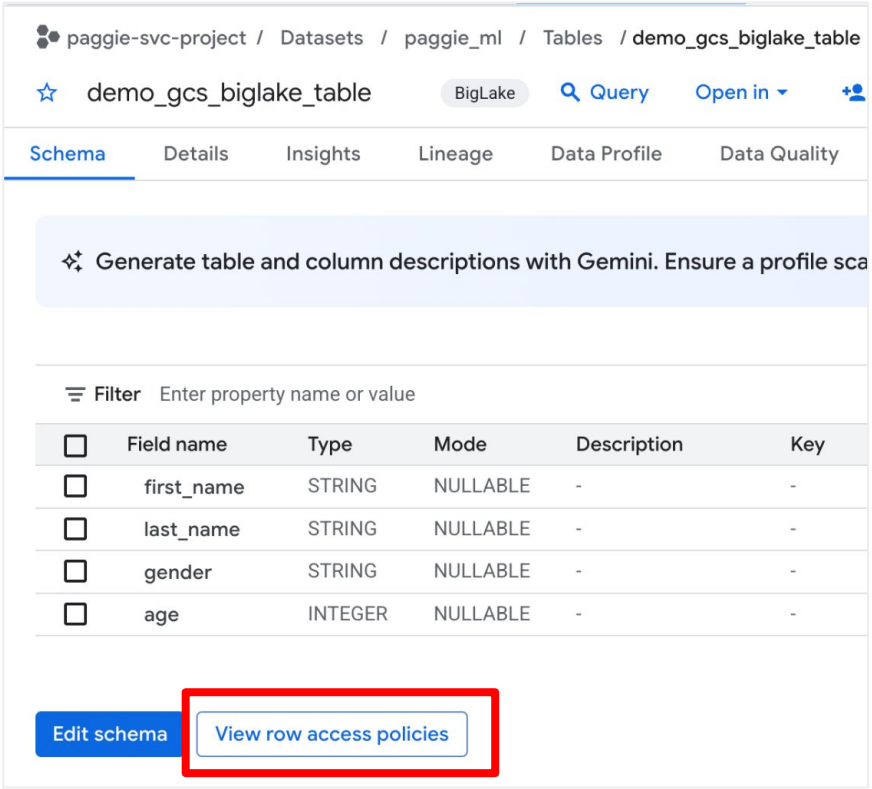
Grant BQ Service Account access in GCS



Non-BigLake Table



BigLake Table



Core Concepts

Project, Dataset, Table

Partitioned & Clustered table

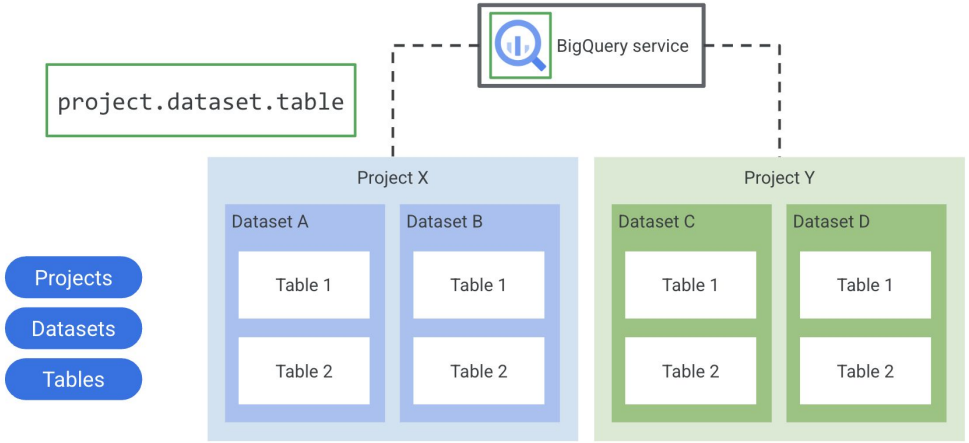
Table & View

Authorized view (column-level access control)

Row-level access control

Policy Tag

Data Lineage



Core Concepts

Project, Dataset, Table

Partitioned & Clustered table

Table & View

Authorized view (column-level access control)

Row-level access control

Policy Tag

Data Lineage

Clustered and Partitioned Tables

Orders table
Not Clustered; Not partitioned

Order_Date	Country	Status
2022-08-02	US	Shipped
2022-08-04	JP	Shipped
2022-08-05	UK	Canceled
2022-08-06	KE	Shipped
2022-08-02	KE	Canceled
2022-08-05	US	Processing
2022-08-04	JP	Processing
2022-08-04	KE	Shipped
2022-08-06	UK	Canceled
2022-08-02	UK	Processing
2022-08-05	JP	Canceled
2022-08-06	UK	Processing
2022-08-05	US	Shipped
2022-08-06	JP	Processing
2022-08-02	KE	Shipped
2022-08-04	US	Shipped

Orders table
Clustered by Country; Not partitioned

Order_Date	Country	Status
2022-08-04	JP	Shipped
2022-08-04	JP	Processing
2022-08-05	JP	Canceled
2022-08-06	JP	Processing
2022-08-06	KE	Shipped
2022-08-02	KE	Canceled
2022-08-04	KE	Shipped
2022-08-02	KE	Shipped
2022-08-05	UK	Processing
2022-08-06	UK	Canceled
2022-08-02	UK	Canceled
2022-08-06	UK	Processing
2022-08-02	US	Shipped
2022-08-05	US	Processing
2022-08-05	US	Shipped
2022-08-04	US	Shipped

Orders table
Clustered by Country; Partitioned by Order_Date (Daily)

	Order_Date	Country	Status
Partition: 2022-08-02	2022-08-02	KE	Shipped
	2022-08-02	KE	Canceled
Clusters: Country	2022-08-02	UK	Processing
	2022-08-02	US	Shipped

	Order_Date	Country	Status
Partition: 2022-08-04	2022-08-04	JP	Shipped
	2022-08-04	JP	Processing
Cluster: Country	2022-08-04	KE	Shipped
	2022-08-04	US	Shipped

	Order_Date	Country	Status
Partition: 2022-08-05	2022-08-05	JP	Canceled
	2022-08-05	UK	Canceled
Cluster: Country	2022-08-05	US	Shipped
	2022-08-05	US	Processing

	Order_Date	Country	Status
Partition: 2022-08-06	2022-08-06	JP	Processing
	2022-08-06	KE	Shipped
Cluster: Country	2022-08-06	UK	Canceled
	2022-08-06	UK	Processing

Core Concepts

Project, Dataset, Table

Partitioned & Clustered table

Table & View

Authorized view (column-level access control)

Row-level access control

Policy Tag

Data Lineage

Clustered Tables

Table 1
Not clustered

Order_Date	Country	Status
2022-08-02	US	Shipped
2022-08-04	JP	Shipped
2022-08-05	UK	Canceled
2022-08-06	KE	Shipped
2022-08-02	KE	Canceled
2022-08-05	US	Processing
2022-08-04	JP	Processing
2022-08-04	KE	Shipped
2022-08-06	UK	Canceled
2022-08-02	UK	Processing
2022-08-05	JP	Canceled
2022-08-06	UK	Processing
2022-08-05	US	Shipped
2022-08-06	JP	Processing
2022-08-02	KE	Shipped
2022-08-04	US	Shipped

Table 2
Clustered by country

Order_Date	Country	Status
2022-08-04	JP	Shipped
2022-08-04	JP	Processing
2022-08-05	JP	Canceled
2022-08-06	JP	Processing
2022-08-06	KE	Shipped
2022-08-02	KE	Canceled
2022-08-04	KE	Shipped
2022-08-02	KE	Shipped
2022-08-05	UK	Canceled
2022-08-06	UK	Canceled
2022-08-02	UK	Processing
2022-08-06	UK	Processing
2022-08-02	US	Shipped
2022-08-05	US	Processing
2022-08-05	US	Shipped
2022-08-04	US	Shipped

Table 3
Clustered by country and status

Order_Date	Country	Status
2022-08-05	JP	Canceled
2022-08-04	JP	Processing
2022-08-06	JP	Processing
2022-08-04	JP	Shipped
2022-08-02	KE	Canceled
2022-08-06	KE	Shipped
2022-08-04	KE	Shipped
2022-08-02	KE	Shipped
2022-08-05	UK	Canceled
2022-08-06	UK	Canceled
2022-08-02	UK	Processing
2022-08-06	UK	Processing
2022-08-05	US	Processing
2022-08-02	US	Shipped
2022-08-05	US	Shipped
2022-08-04	US	Shipped

Optimizing Queries on Clustered Tables

Orders table
Clustered by Order_Date, Country and Status

Order_Date	Country	Status
2022-08-02	KE	Canceled
2022-08-02	KE	Shipped
2022-08-02	UK	Processing
2022-08-02	US	Shipped
2022-08-04	JP	Processing
2022-08-04	JP	Shipped
2022-08-04	KE	Shipped
2022-08-04	US	Shipped
2022-08-05	JP	Canceled
2022-08-05	UK	Canceled
2022-08-05	US	Processing
2022-08-05	US	Shipped
2022-08-06	JP	Processing
2022-08-06	KE	Shipped
2022-08-06	UK	Canceled
2022-08-06	UK	Processing

Example 1: Query filters on Order_Date and Country
✔ - Query optimized for clustering

Order_Date	Country	Status

Example 2: Query filters on Country and Status
✘ - Not optimized for clustering

Order_Date	Country	Status

Core Concepts

Project, Dataset, Table

Partitioned & Clustered table

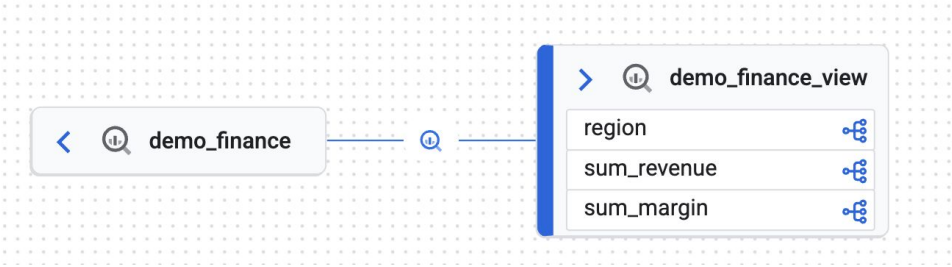
Table & View

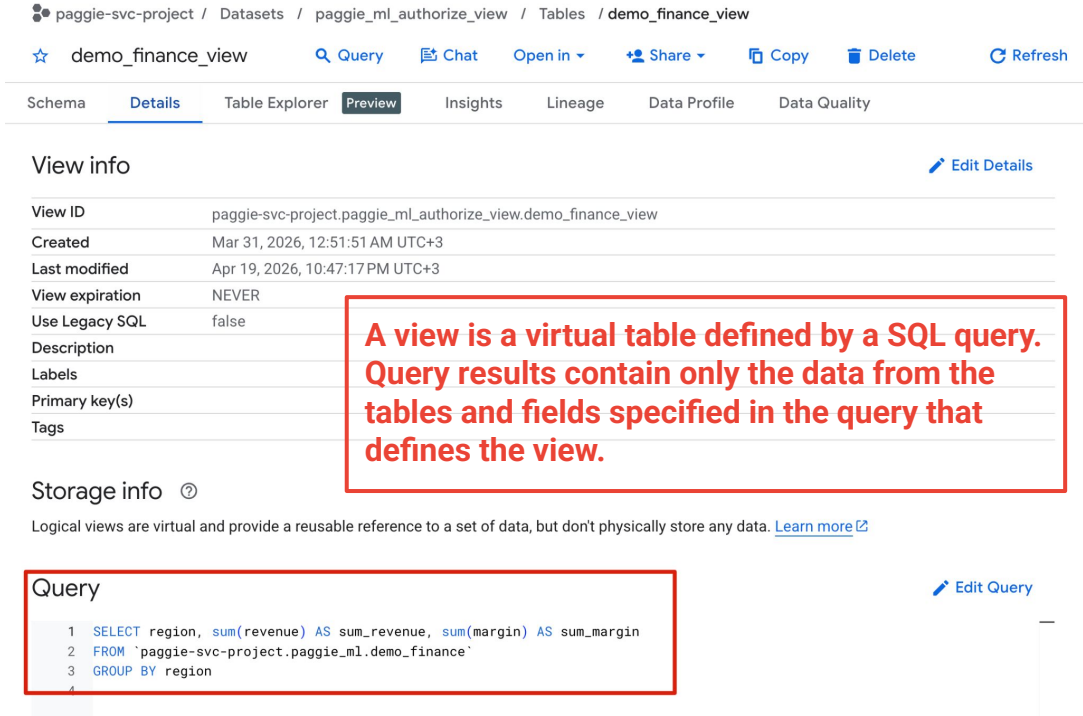
Authorized view (column-level access control)

Row-level access control

Policy Tag

Data Lineage





Core Concepts

Project, Dataset, Table

Partitioned & Clustered table

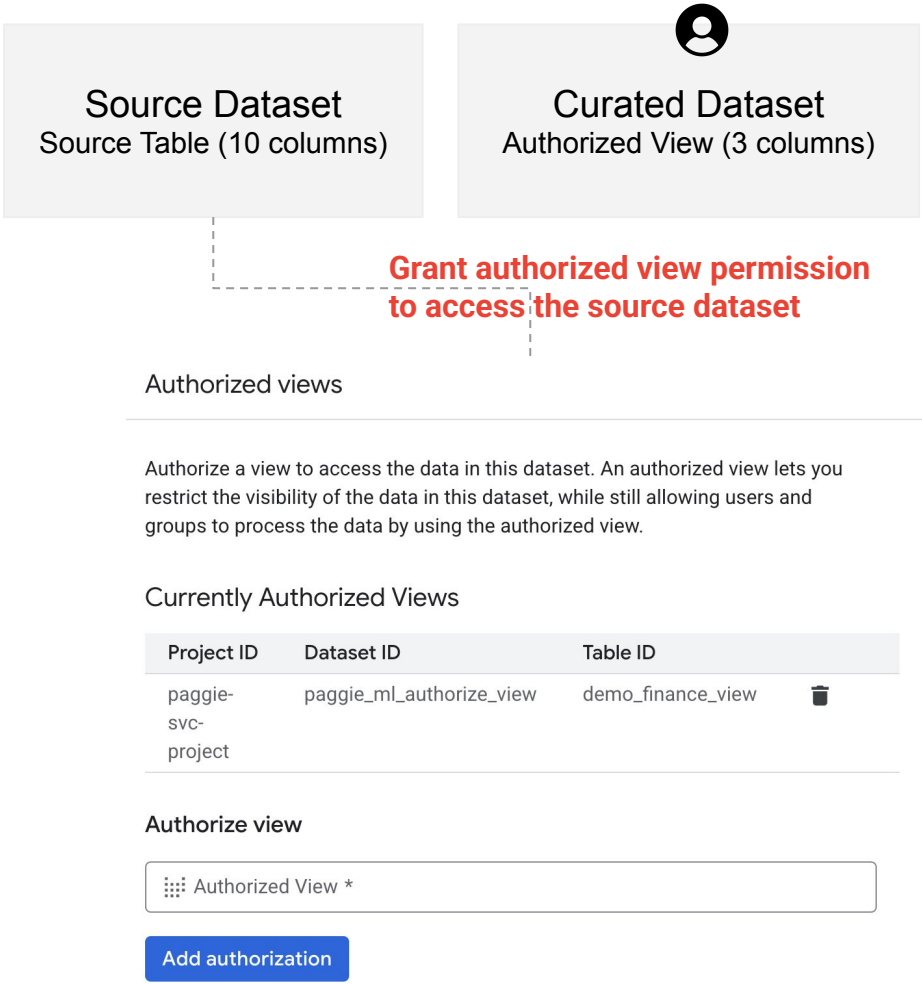
Table & View

Authorized view (column-level access control)

Row-level access control

Policy Tag

Data Lineage



Core Concepts

Project, Dataset, Table

Partitioned & Clustered table

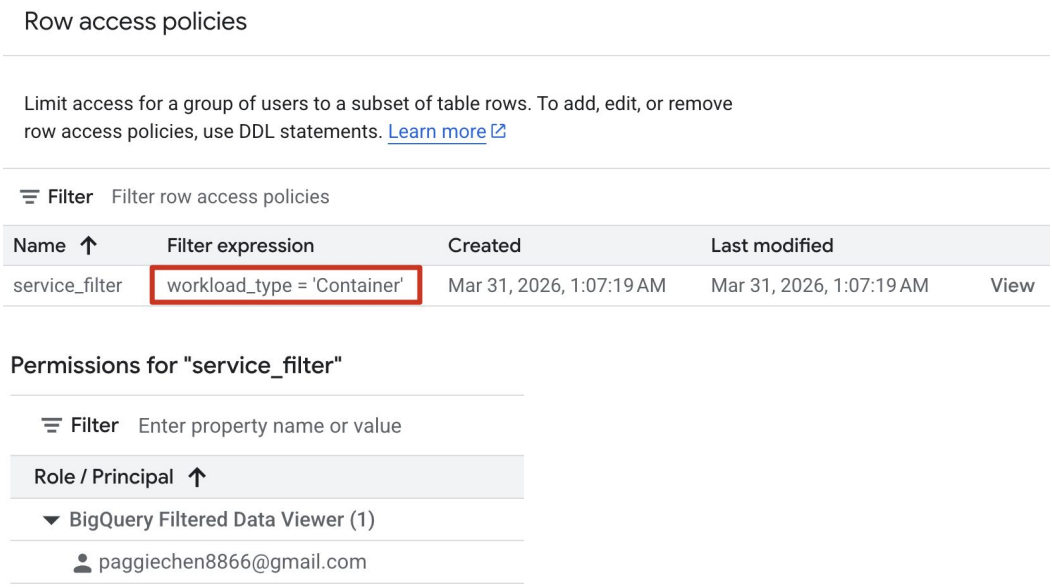
Table & View

Authorized view (column-level access control)

Row-level access control

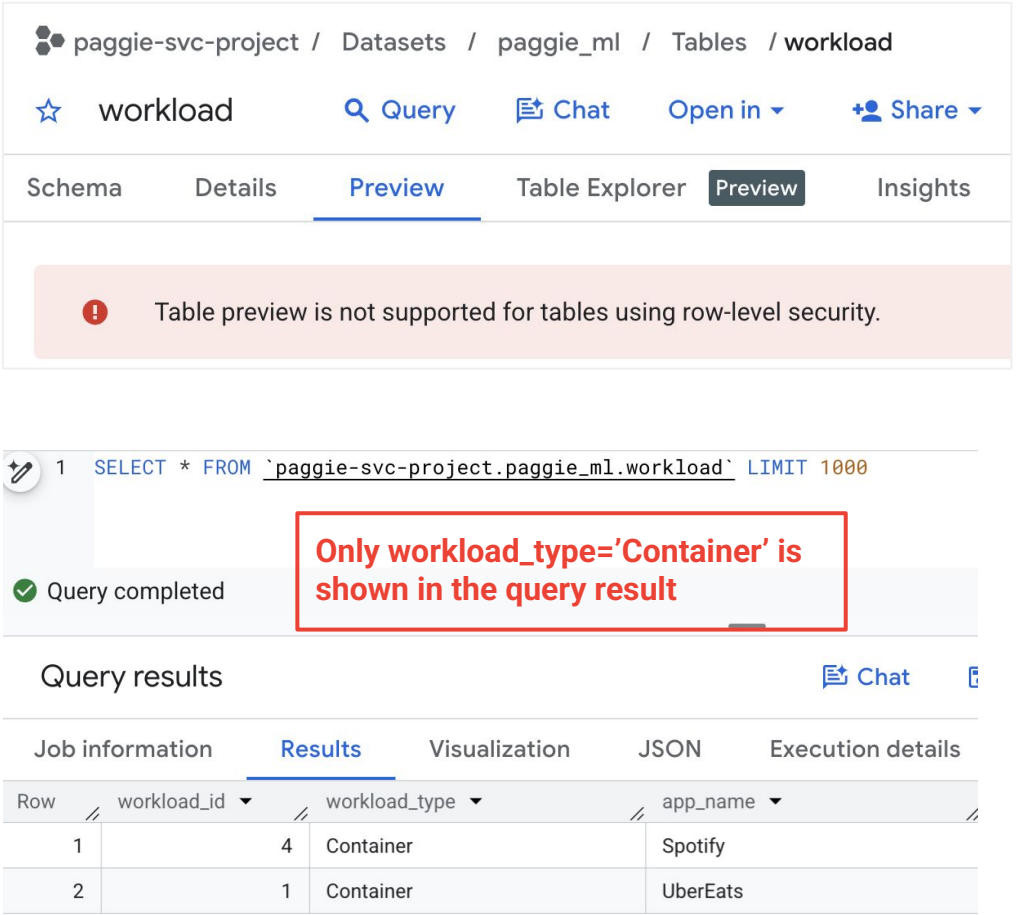
Policy Tag

Data Lineage



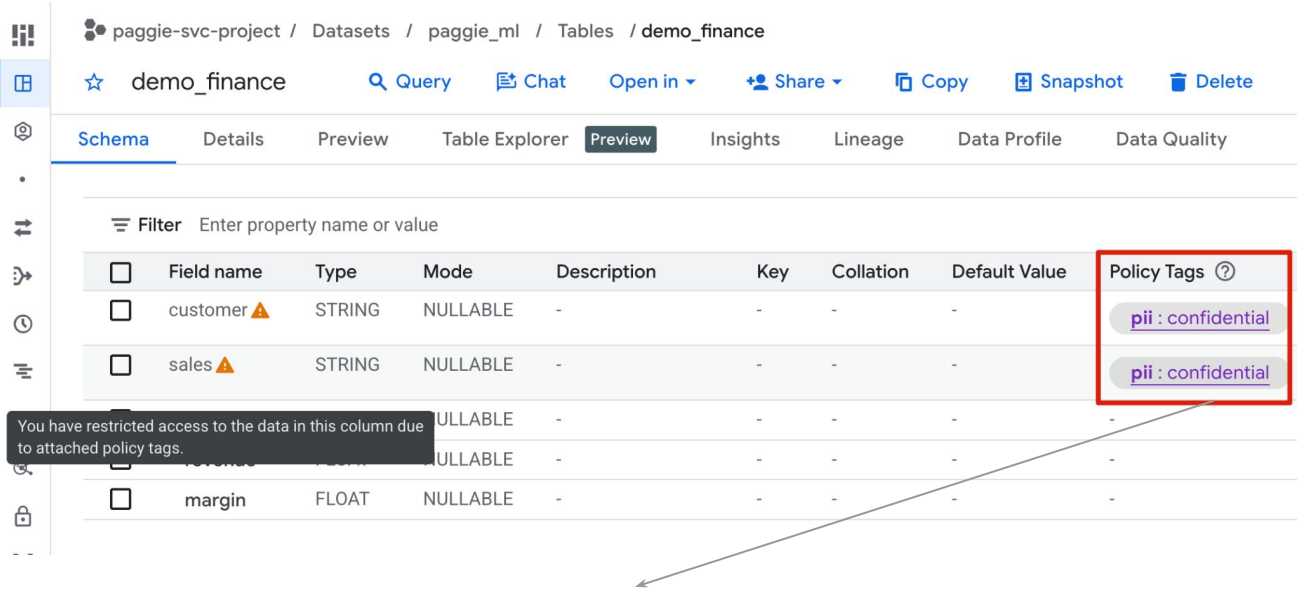
Core Concepts

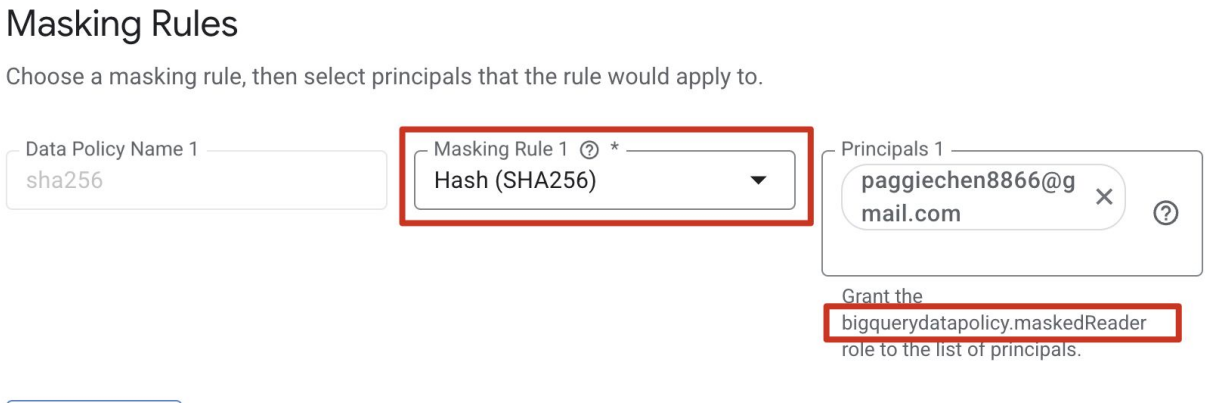
Project, Dataset, Table
Partitioned & Clustered table
Table & View
Authorized view (column-level access control)
Row-level access control
Policy Tag
Data Lineage



Core Concepts

Project, Dataset, Table
Partitioned & Clustered table
Table & View
Authorized view
Row-level access control
Policy Tag
Data Lineage





Core Concepts

Project, Dataset, Table

Partitioned & Clustered table

Table & View

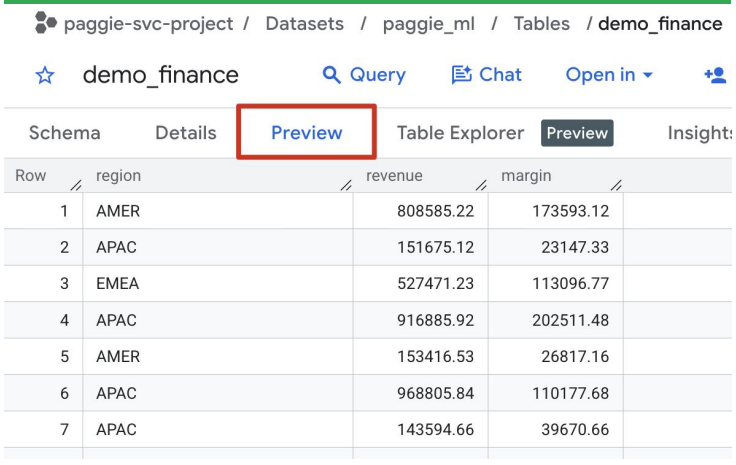
Authorized view

Row-level access control

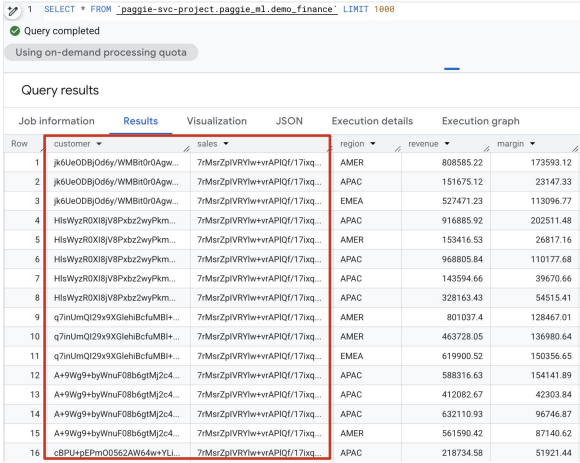
Policy Tag

Data Lineage

Masked Reader Role (Grant on Policy Tag)

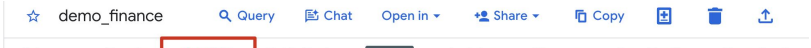


Row	region	revenue	margin
1	AMER	808585.22	173593.12
2	APAC	151675.12	23147.33
3	EMEA	527471.23	113096.77
4	APAC	916885.92	202511.48
5	AMER	153416.53	26817.16
6	APAC	968805.84	110177.68
7	APAC	143594.66	39670.66



Row	customer	sales	region	revenue	margin
1	jk6UeODBjOd6y/WMBit0r0Agw...	7rMsrZplVRYlw+vrAPlQf/17ixq...	AMER	808585.22	173593.12
2	jk6UeODBjOd6y/WMBit0r0Agw...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	151675.12	23147.33
3	jk6UeODBjOd6y/WMBit0r0Agw...	7rMsrZplVRYlw+vrAPlQf/17ixq...	EMEA	527471.23	113096.77
4	HlsWyzR0XI8jV8Pxbz2wyPkm...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	916885.92	202511.48
5	HlsWyzR0XI8jV8Pxbz2wyPkm...	7rMsrZplVRYlw+vrAPlQf/17ixq...	AMER	153416.53	26817.16
6	HlsWyzR0XI8jV8Pxbz2wyPkm...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	968805.84	110177.68
7	HlsWyzR0XI8jV8Pxbz2wyPkm...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	143594.66	39670.66
8	HlsWyzR0XI8jV8Pxbz2wyPkm...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	328163.43	54515.41
9	q7inUmQI29x9XGlehiBcfuMBI+...	7rMsrZplVRYlw+vrAPlQf/17ixq...	AMER	801037.4	128467.01
10	q7inUmQI29x9XGlehiBcfuMBI+...	7rMsrZplVRYlw+vrAPlQf/17ixq...	AMER	463728.05	136980.64
11	q7inUmQI29x9XGlehiBcfuMBI+...	7rMsrZplVRYlw+vrAPlQf/17ixq...	EMEA	619900.52	150356.65
12	A+9Wg9+byWnuF08b6gtMj2c4...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	588316.63	154141.89
13	A+9Wg9+byWnuF08b6gtMj2c4...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	412082.67	42303.84
14	A+9Wg9+byWnuF08b6gtMj2c4...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	632110.93	96746.87
15	A+9Wg9+byWnuF08b6gtMj2c4...	7rMsrZplVRYlw+vrAPlQf/17ixq...	AMER	561590.42	87140.62
16	cBPU+pEPmO0562AW64w+YLi...	7rMsrZplVRYlw+vrAPlQf/17ixq...	APAC	218734.58	51921.44

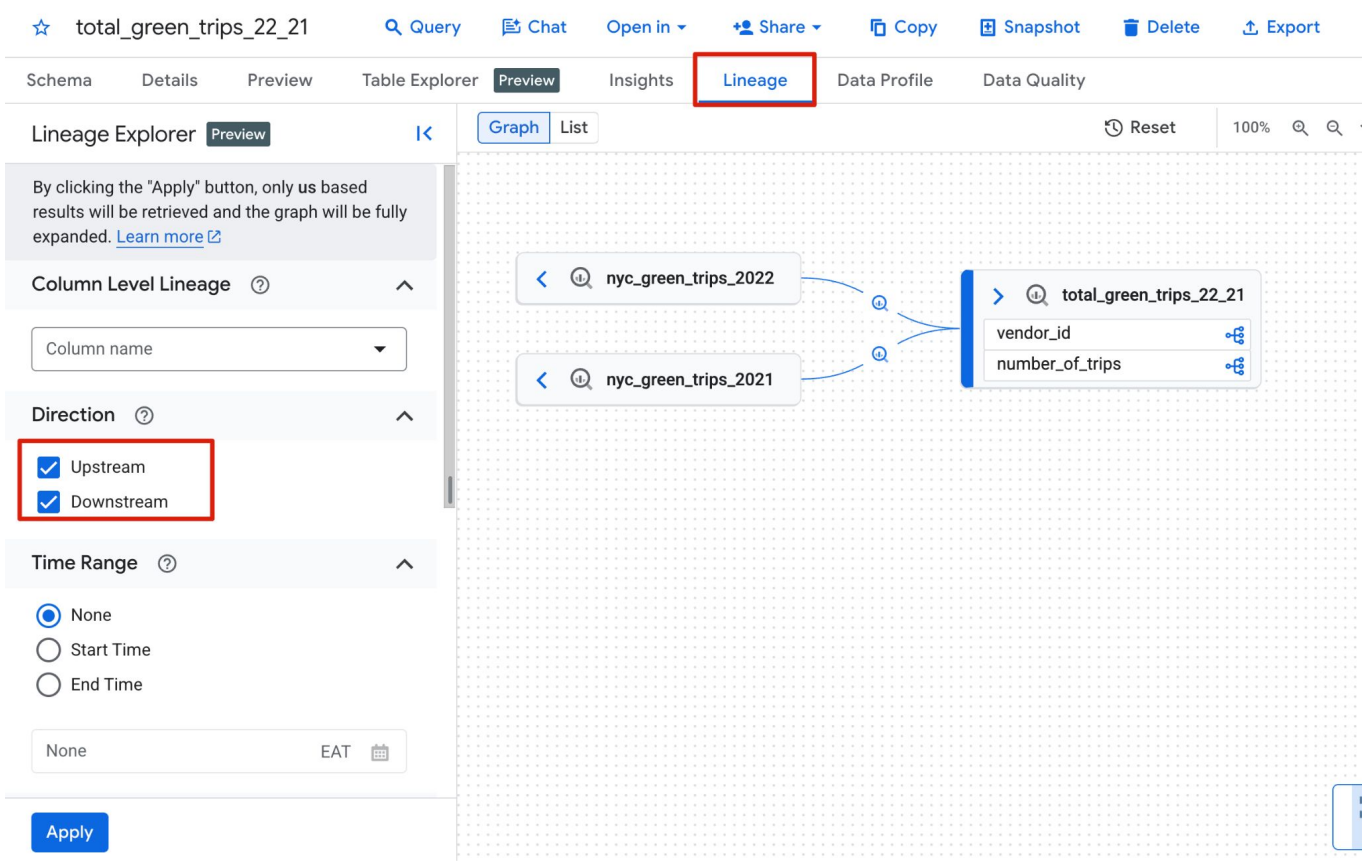
Fine-grained Reader Role (Grant on IAM page)



Row	customer	sales	region	revenue	margin
1	Alpine Logistics	David Chen	AMER	808585.22	173593.12
2	Alpine Logistics	David Chen	APAC	151675.12	23147.33
3	Alpine Logistics	David Chen	EMEA	527471.23	113096.77
4	Desert Oasis Ltd	David Chen	APAC	916885.92	202511.48
5	Desert Oasis Ltd	David Chen	AMER	153416.53	26817.16
6	Desert Oasis Ltd	David Chen	APAC	968805.84	110177.68
7	Desert Oasis Ltd	David Chen	APAC	143594.66	39670.66
8	Desert Oasis Ltd	David Chen	APAC	328163.43	54515.41
9	Doha Dynamics	David Chen	AMER	801037.4	128467.01
10	Doha Dynamics	David Chen	AMER	463728.05	136980.64
11	Doha Dynamics	David Chen	EMEA	619900.52	150356.65

Core Concepts

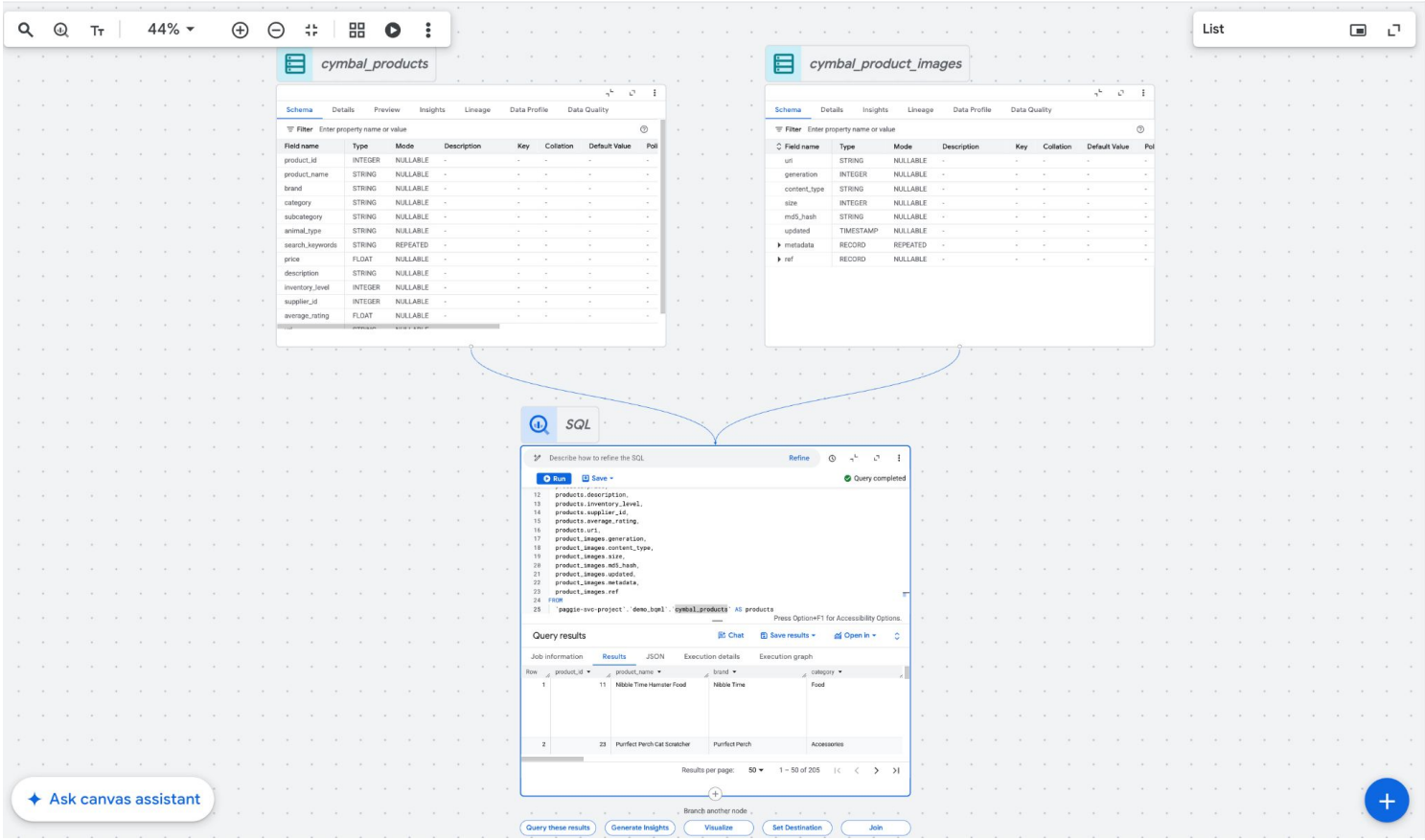
Project, Dataset, Table
Partitioned & Clustered table
Table & View
Authorized view
Row-level access control
Policy Tag
Data Lineage



Build Pipelines

Data Canvas

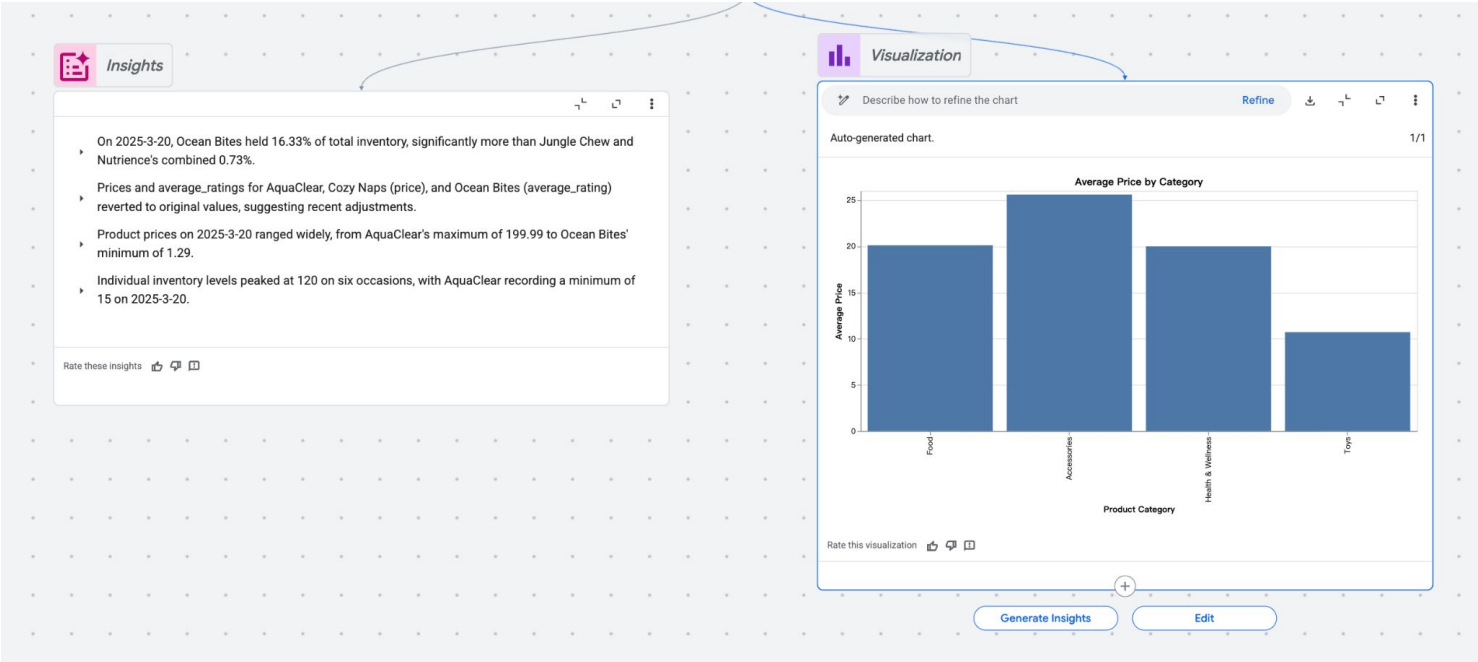
Dataform



Build Pipelines

Data Canvas

Dataform



Build Pipelines

Data Canvas

Dataform

Development Workspaces | Workflow Execution Logs | Releases & scheduling | Settings

+ Create development workspace | Delete selected workspaces

Development workspaces contain an editable copy of your team's repository. Using development workspaces, you can develop code without affecting other users, commit changes, and push commits to your remote Git repository. If your repository is connected to a remote Git provider, changes from your development workspace will be pushed to a remote branch named after your development workspace. Otherwise, your changes will be pushed to your default branch. Learn more

Name ↑
jackson
paggie

branch

Development Workspaces | Workflow Execution Logs | Releases & scheduling | Settings

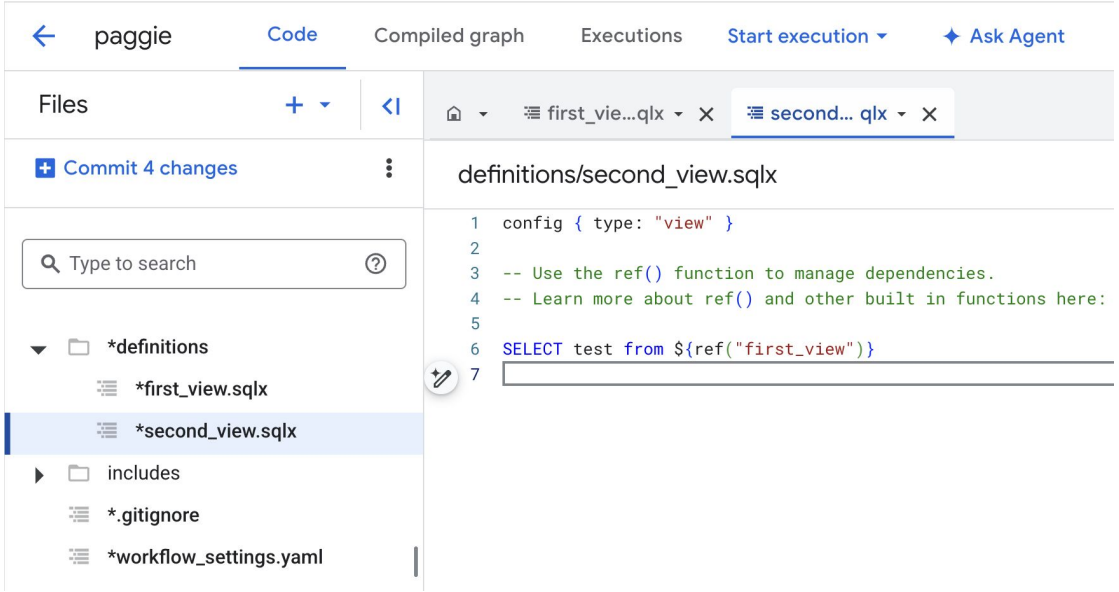
Connect with Git | Configure private npm packages

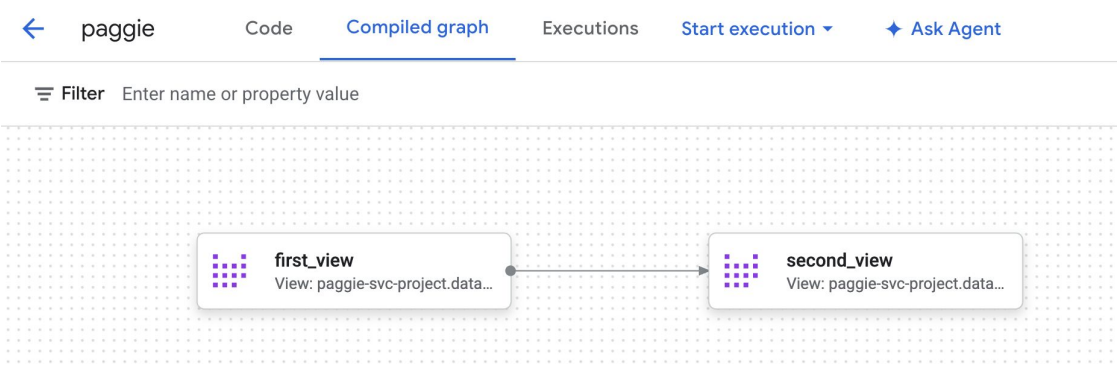
Name	demo
Location	me-central1
Service account	Default Dataform service account
actAs permission checks	
Enforcement	Don't enforce
Encryption Settings	
Type	Google-managed

Build Pipelines

Data Canvas

Dataform





Augment with BigQuery ML

Scenario 1: [Remote Model] Document AI with BQML

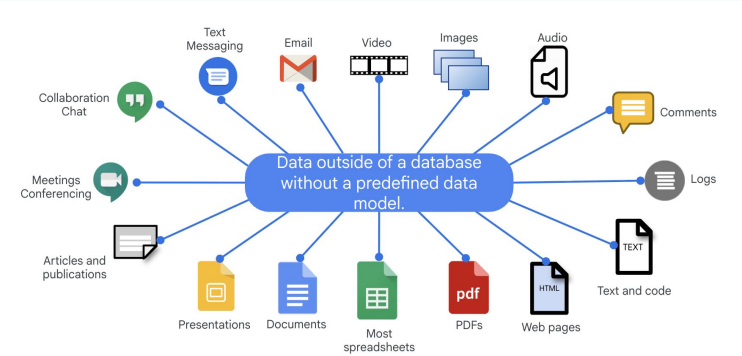
Scenario 2: [Remote Model] Image Data Summarization with BQML

Scenario 3: [Native Model] Custom BQML Model

The Unstructured Challenge

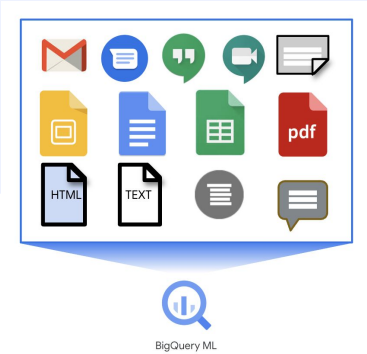
80% of enterprise data

is unstructured, trapped in images, PDFs, and audio files. Traditionally, this data was inaccessible to SQL-based analytics.



Multimodal BQML

We bridge the gap between data engineering and cognitive inference, enabling massive scalability without leaving the BigQuery ecosystem.



Introducing Object Tables

An **Object Table** is a specialized **external table** in BigQuery that manages **ObjectRef** metadata for unstructured data stored in Cloud Storage, let you analyze unstructured data in Cloud Storage.

Consistent Metadata

Automatic tracking of size, URI, and timestamps.

Unified Security

IAM permissions govern both SQL access and file reads.

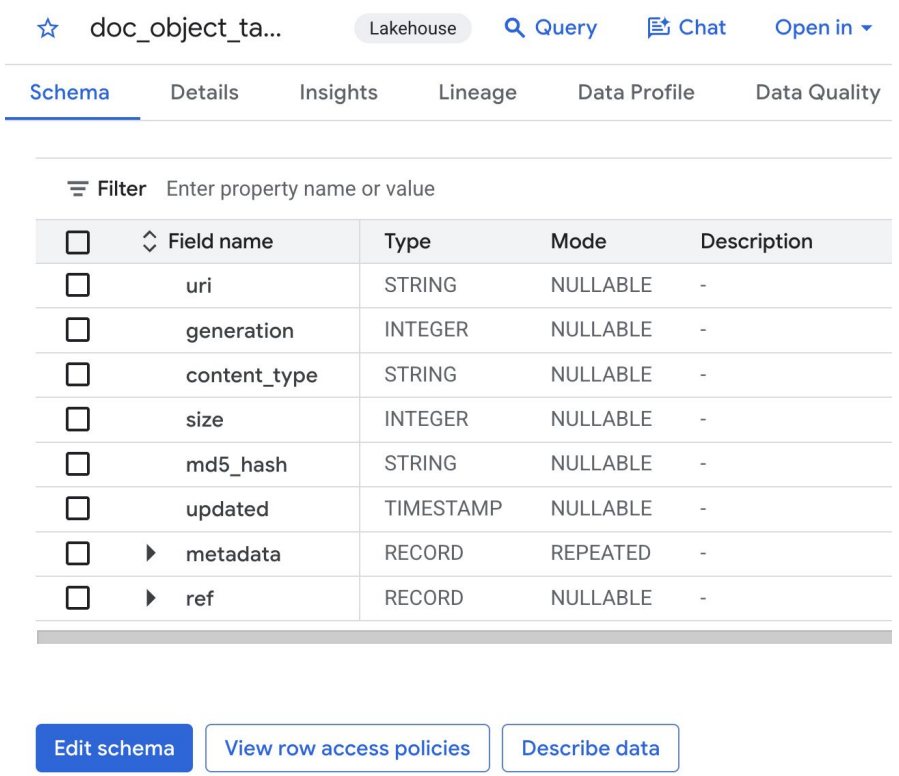
AI Native

Directly compatible with ML.GENERATE_TEXT.

Introduction to object tables | BigQuery | Google Cloud Documentation

Introducing Object Tables

An **Object Table** is a specialized **external table** in BigQuery that manages **ObjectRef** metadata for unstructured data stored in Cloud Storage, let you analyze unstructured data in Cloud Storage.



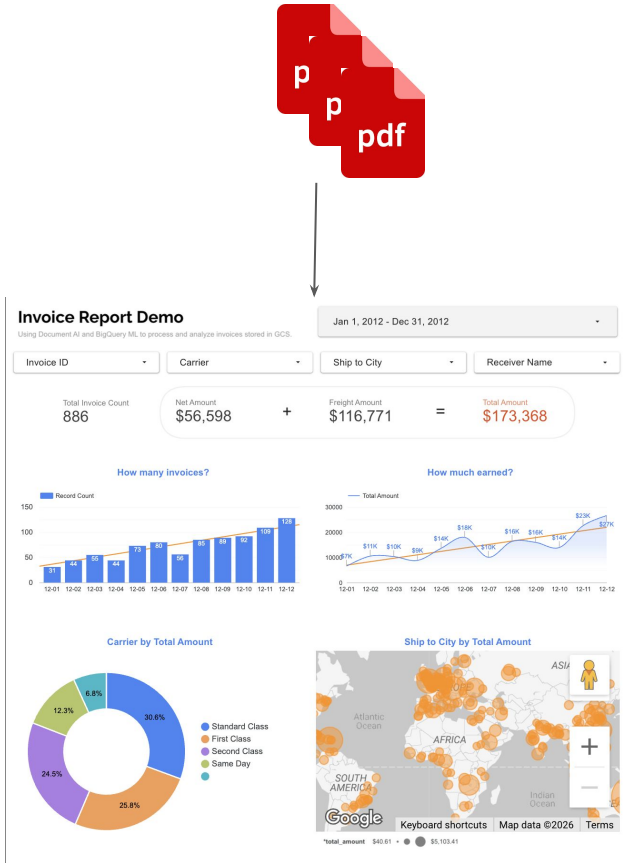
[Introduction to object tables | BigQuery | Google Cloud Documentation]

Augment with BigQuery ML

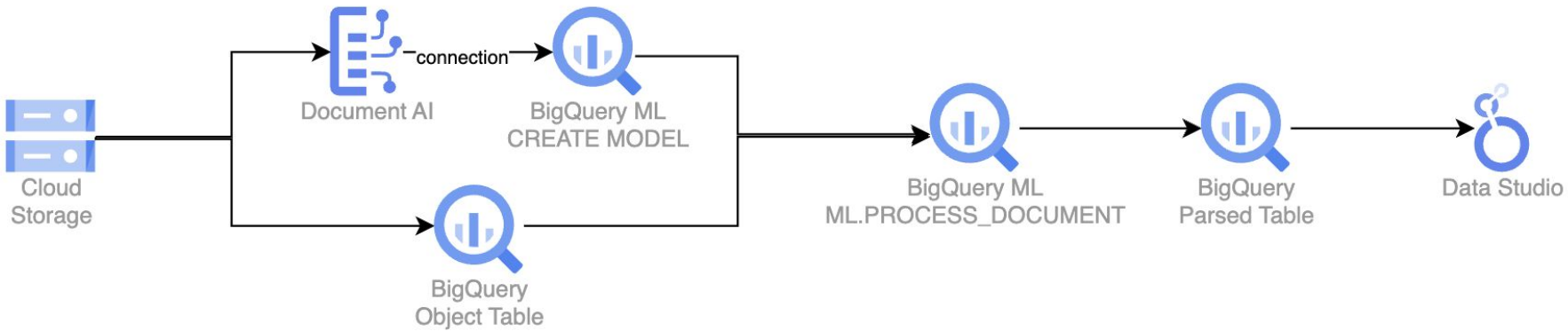
Scenario 1: [Prebuilt API] Document AI with BQML

Scenario 2: [Vertex AI] Image Data Summarization with BQML

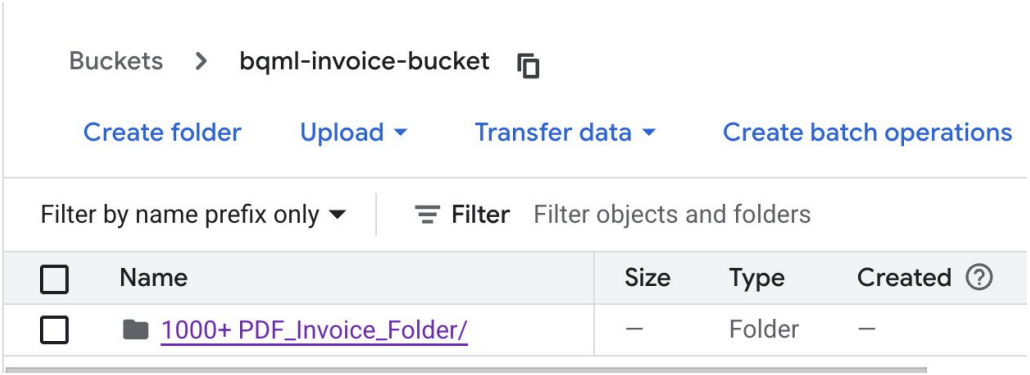
Scenario 3: [Native Model] Custom BQML Model



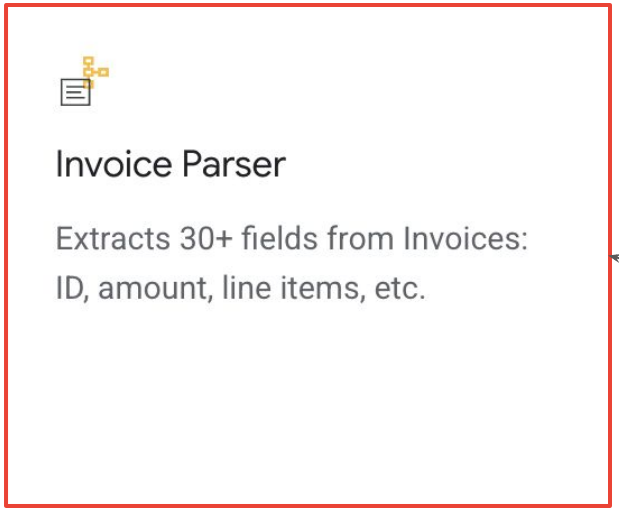
Architecture Overview

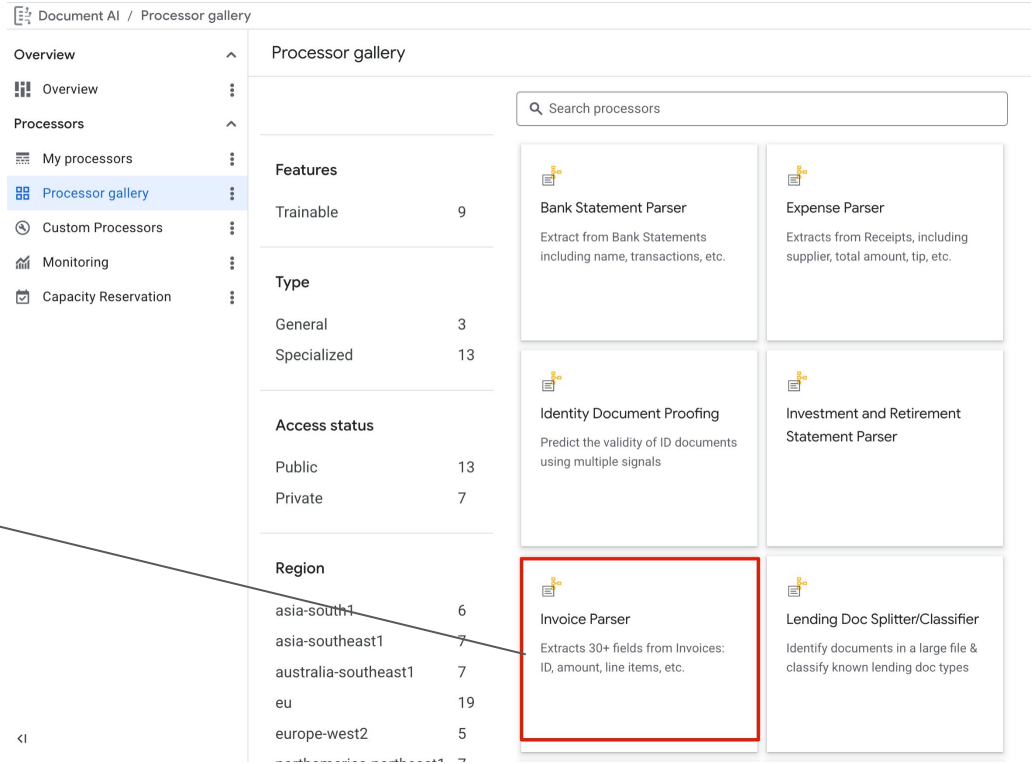


Step1. Upload invoice PDFs to GCS

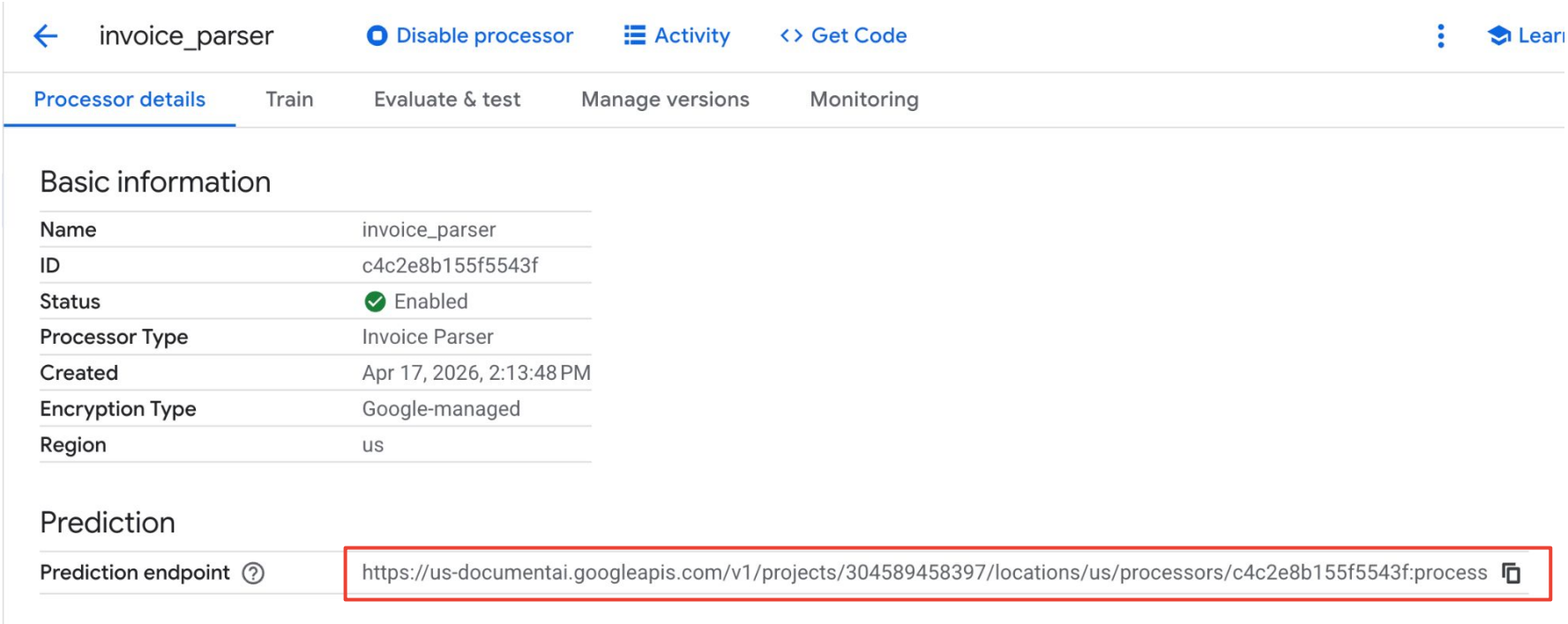


Step2. Create Document AI Processor: Invoice Parser





Step2. Create Document AI Processor: Invoice Parser



Step3. Create BigQuery External Connection & Assign IAM

docc-connection

Query Share Delete Refresh

Connection info

Edit Details

Connection ID	projects/paggie-svc-project/locations/us/connections/docc-connection
Friendly name	
Created	Apr 17, 2026, 2:16:14 PM UTC+3
Last modified	Apr 17, 2026, 2:16:14 PM UTC+3
Data location	us
Description	Document AI
Connection type	Vertex AI remote models, remote functions, Lakehouse and Spanner (Cloud Resource)
Service account id	bqcx-304589458397-ij5i@gcp-sa-bigquery-condel.iam.gserviceaccount.com

Principal ↑	Name	Role
bqcx-304589458397-ij5i@gcp-sa-bigquery-condel.iam.gserviceaccount.com		Document AI Viewer (Beta) Storage Object Viewer

Step 4. Create BigQuery Model

SQL Editor - BigQuery

```
CREATE OR REPLACE MODEL
`<Project_Name>.<Dataset_Name>.doc_model`
REMOTE WITH CONNECTION `<Project_Name>.us.doc_conn`
OPTIONS (
REMOTE_SERVICE_TYPE = 'CLOUD_AI_DOCUMENT_V1',
DOCUMENT_PROCESSOR = '<Processor_ID>'
);
```

Query ready to run

Specifies the service to use to create the model:

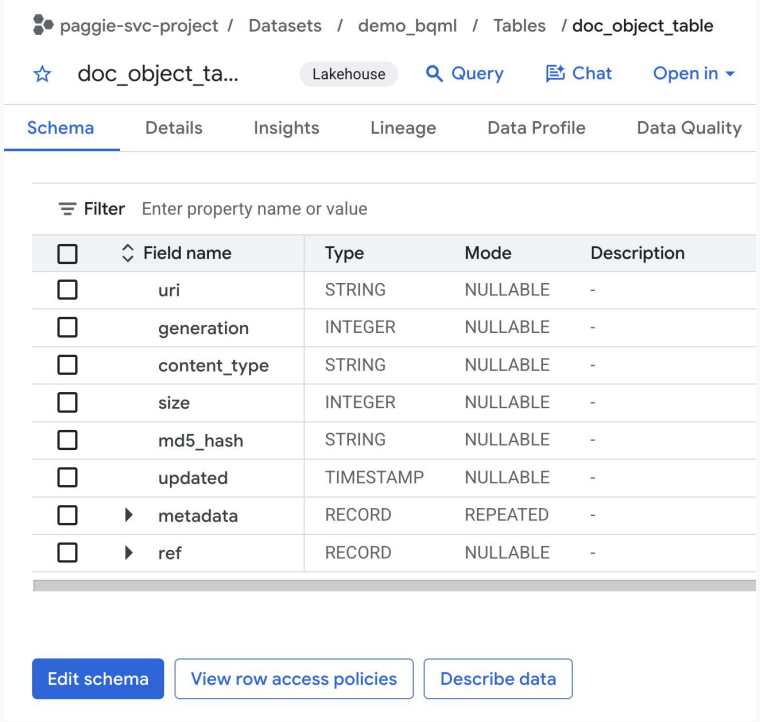
- Cloud Natural Language API
- Cloud Translation API
- Cloud Vision API
- Document AI API
- Speech-to-Text API

Step 5. Create Object Table

SQL Editor - BigQuery

```
CREATE EXTERNAL TABLE
`<Project_Name>.<Dataset_Name>.doc_object_table`
WITH CONNECTION `<Project_Name>.us.doc_conn`
OPTIONS(
 object_metadata = 'SIMPLE',
 uris = ['gs://<Bucket_Name>/*']
);
```

Query ready to run

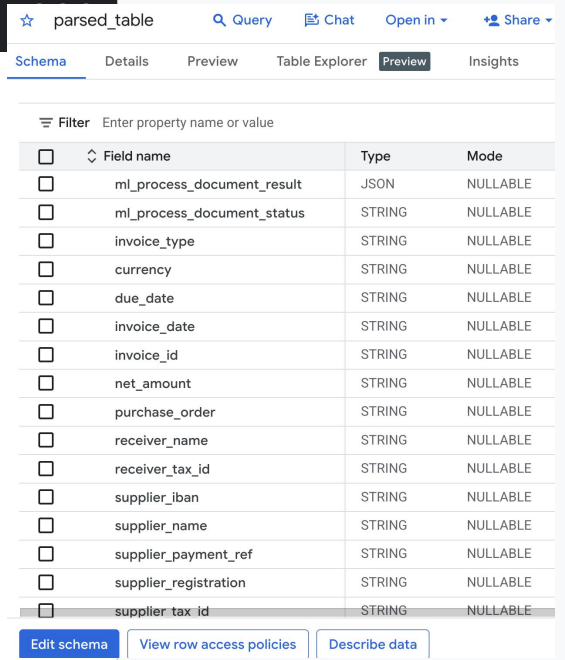


Step 6. ML.PROCESS_DOCUMENT

SQL Editor - BigQuery

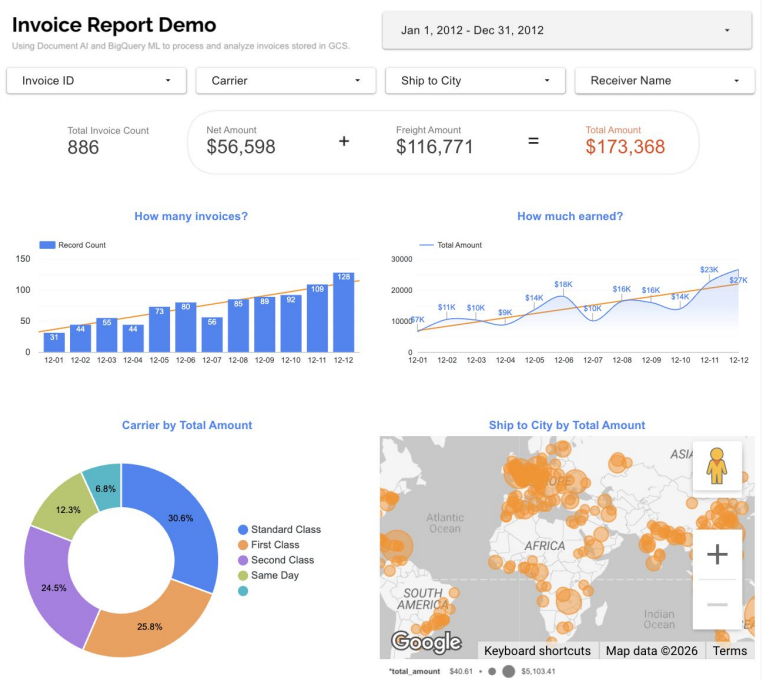
```
CREATE OR REPLACE TABLE `<Project_Name>.<Dataset_Name>.parsed_table`
AS
SELECT *
FROM ML.PROCESS_DOCUMENT(
 MODEL `<Project_Name>.<Dataset_Name>.doc_model`,
 TABLE `<Project_Name>.<Dataset_Name>.doc_object_table`
);
```

Query ready to run

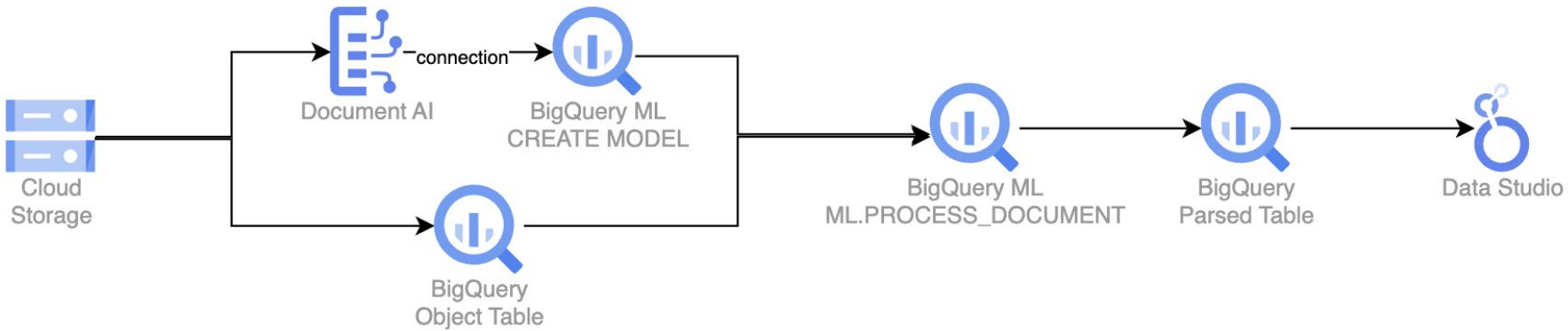


Scenario 1: Document AI with BQML

Step 7. Data Studio Dashboard



Architecture Overview



Augment with BigQuery ML

Scenario 1: [Prebuilt API] Document AI with BQML

Scenario 2: [Vertex AI] Image Data Summarization with BQML

Scenario 3: [Native Model] Custom BQML Model

https://storage.googleapis.com/cloud-samples-data/bigquery/tutorials/cymbal-pets/images/aquaclear-aquarium-co2-system.png

↓

Image Description

Image Public GCS

Step 1. Create Connection

SQL Editor - BigQuery

```
CREATE OR REPLACE CONNECTION
`.us.cymbal_connection` OPTIONS( connection_type
= 'CLOUD_RESOURCE' );
```

Query ready to run

This creates a **Resource Connection Service Account**. Grant this account the **Vertex AI User** role in IAM.

Step 2. Load Product Info

SQL Editor - BigQuery

```
LOAD DATA OVERWRITE demo_bqml.cymbal_products
FROM
 FILES(
   format = 'avro',
   uris = [

'gs://cloud-samples-data/bigquery/tutorials/cymbal-pets/tables/products/products_*.avro']);
```

Query ready to run

Field name	Type	Mode
product_id	INTEGER	NULLABLE
product_name	STRING	NULLABLE
brand	STRING	NULLABLE
category	STRING	NULLABLE
subcategory	STRING	NULLABLE
animal_type	STRING	NULLABLE
search_keywords	STRING	REPEATED
price	FLOAT	NULLABLE
description	STRING	NULLABLE
inventory_level	INTEGER	NULLABLE
supplier_id	INTEGER	NULLABLE
average_rating	FLOAT	NULLABLE
uri	STRING	NULLABLE

Step 3. Create Product Image Object Table

SQL Editor - BigQuery

```
CREATE EXTERNAL TABLE
demo_bqml.cymbal_product_images
  WITH CONNECTION `us.cymbal_connection`
  OPTIONS (
    object_metadata = 'SIMPLE',
    uris =
['gs://cloud-samples-data/bigquery/tutorials/cymbal
-pets/images/*.png'],
    max_staleness = INTERVAL 30 MINUTE,
    metadata_cache_mode = AUTOMATIC);
```

Query ready to run

Field name	Type	Mode
uri	STRING	NULLABLE
generation	INTEGER	NULLABLE
content_type	STRING	NULLABLE
size	INTEGER	NULLABLE
md5_hash	STRING	NULLABLE
updated	TIMESTAMP	NULLABLE
▶ metadata	RECORD	REPEATED
▶ ref	RECORD	NULLABLE

Step 4. Inner Join both

SQL Editor - BigQuery

```
CREATE OR REPLACE TABLE
demo_bqml.cymbal_products_mm
AS
SELECT cymbal_products.* EXCEPT (uri), ot.ref AS
image FROM demo_bqml.cymbal_products
INNER JOIN demo_bqml.cymbal_product_images ot
ON ot.uri = cymbal_products.uri;
```

Query ready to run

Step 5. Register Gemini Model

SQL Editor - BigQuery

```
CREATE OR REPLACE MODEL `demo_bqml.gemini`
  REMOTE WITH CONNECTION `us.cymbal_connection`
  OPTIONS (ENDPOINT = 'gemini-2.0-flash');
```

Query ready to run

Vertex AI Integration

This entry in the **Model Registry** allows BigQuery to send inference requests to Vertex AI endpoints via standard SQL syntax.

Step 6. Generate image_description by inferencing Gemini

```
CREATE OR REPLACE TABLE demo_bqml.cymbal_products_mm
AS
SELECT
  product_id,
  product_name,
  brand,
  category,
  subcategory,
  animal_type,
  search_keywords,
  price,
  description,
  inventory_level,
  supplier_id,
  average_rating,
  image,
  image_description
FROM
  AI.GENERATE_TABLE(
    MODEL `demo_bqml.gemini`,
    (
      SELECT
        ('Can you describe the following image? ', OBJ.GET_ACCESS_URL(image, 'r')) AS prompt,
        *
      FROM
        demo_bqml.cymbal_products_mm
    ),
    STRUCT('image_description STRING' AS output_schema));
```

The `AI.GENERATE_TABLE` function is a BigQuery SQL command that allows you to use generative AI model to transform unstructured data, such as raw text or images, directly into a structured table format.

Step 7. Review

https://storage.googleapis.com/cloud-samples-data/bigquery/tutorials/cymbal-pets/images/aquaclear-aquarium-co2-system.png

Gemini Generated Image Description:

The image displays a laboratory setup with a clear glass Aquaclear brand device featuring a spiral internal structure. A transparent tube connects this device to a regulator with a gauge and a black adjustment knob, all set against a gradient gray background.

Augment with BigQuery ML

Scenario 1: [Prebuilt API] Document AI with BQML

Scenario 2: [Vertex AI] Image Data Summarization with BQML

Scenario 3: [Native Model] Custom BQML Model

Demo BQML–Predict NYC taxi fares using BQML

In this demo, we'll build a model to predict taxi fares in NYC using one of BigQuery's public datasets *nyc-tlc.yellow.trips*.

Category	Key Columns	Description
Time	pickup_datetime, dropoff_datetime	The precise date and time when the meter was engaged and disengaged.
Location	pickup_longitude, pickup_latitude / PULocationID	The longitude/latitude (older data) or TLC Taxicab Boundary Zone (newer data) where the trip began/ended.
Fare	fare_amount, tip_amount, tolls_amount	The cost components: base fare, credit card tips, and bridge/tunnel tolls.
Trip Details	passenger_count, trip_distance	The number of passengers (driver-entered) and the elapsed trip distance reported by the taximeter.
Payment	payment_type	A numeric code signifying how the passenger paid (e.g., 1=Credit Card, 2=Cash).

Step 1. Create Dataset

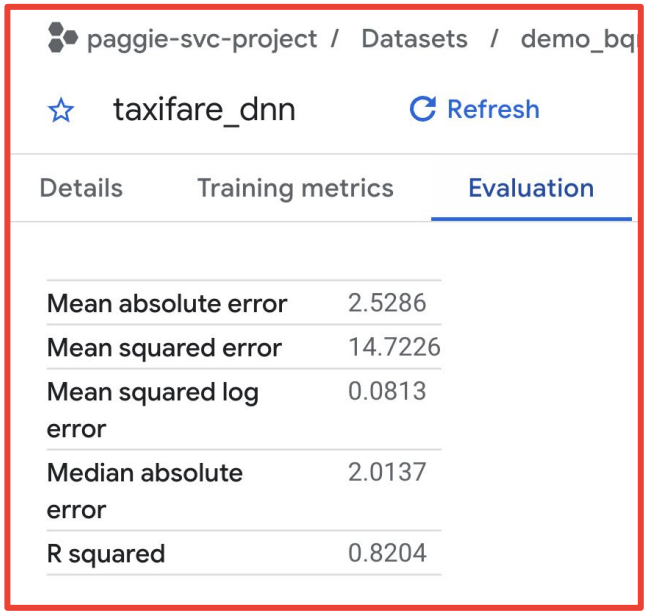
```
#standardSQL
create or replace table `taxi.taxi300k` as
WITH taxi_preproc AS (
SELECT
  ABS(MOD(FARM_FINGERPRINT(STRING(pickup_datetime)), 10000)) AS dataset,
  (tolls_amount + fare_amount) AS fare_amount,
  pickup_datetime,
  EXTRACT(DAYOFWEEK FROM pickup_datetime) AS dayofweek,
  EXTRACT(HOUR FROM pickup_datetime) AS hourofday,
  pickup_longitude AS pickuplon,
  pickup_latitude AS pickuplat,
  dropoff_longitude AS dropofflon,
  dropoff_latitude AS dropofflat,
  passenger_count
FROM
  `nyc-tlc.yellow.trips`
```

```
WHERE
  trip_distance > 0
  AND fare_amount >= 2.5
  AND fare_amount < 200
  AND pickup_longitude > -78
  AND pickup_longitude < -70
  AND dropoff_longitude > -78
  AND dropoff_longitude < -70
  AND pickup_latitude > 37
  AND pickup_latitude < 45
  AND dropoff_latitude > 37
  AND dropoff_latitude < 45
  AND passenger_count > 0
  AND ABS(MOD(FARM_FINGERPRINT(STRING(pickup_datetime)), 10000)) < 3
  )
  SELECT
  dataset,
  fare_amount,
  pickup_datetime,
  hourofday,
  dayofweek,
  CAST(dayofweek * 24 + hourofday AS STRING) AS dayhour,
  pickuplon,
  pickuplat,
  dropofflon,
  dropofflat,
  SQRT(POW((pickuplon - dropofflon),2) + POW(( pickuplat - dropofflat), 2)) AS dist,
  #Euclidean distance between pickup and drop off
  pickuplon - dropofflon AS londiff,
  pickuplat - dropofflat AS latdiff,
  passenger_count
  FROM taxi_preproc
  WHERE dataset < 3
```

Step 2. Create/Train Model

```
CREATE OR REPLACE MODEL
  taxi.taxifare_dnn OPTIONS (model_type='dnn_regressor',
    hidden_units=[144, 89, 55],
    labels=['fare_amount']) AS
SELECT
    fare_amount,
    hourofday,
    dayofweek,
    pickuplon,
    pickuplat,
    dropofflon,
    dropofflat,
    passenger_count
  FROM
    `taxi.taxi300k`
  WHERE
    dataset = 0;
```

Step 3. Evaluation



Metric	Definition & Meaning
Mean absolute error (MAE)	The average of the absolute differences between the predicted fare and the actual fare. On average, your predictions are off by **$2.53**.
Mean squared error (MSE)	The average of the squared differences. Because it squares the error, it penalizes large outliers more heavily than MAE.
Mean squared log error	Measures the relative error between the log of the predicted and actual values. It is useful when the target values have a wide range.
Median absolute error	The median of all error magnitudes. 50% of your predictions have an error of less than **$2.01**. It is more robust to outliers than MAE.
R squared	The "Coefficient of Determination". It indicates that **82.04%** of the variance in the taxi fare is explained by your model. Closer to 1.0 is better.

Step 4. Predict

```
SELECT
    fare_amount,
    predicted_fare_amount,
    hourofday,
    dayofweek,
    pickuplon,
    pickuplat,
    dropofflon,
    dropofflat,
    passenger_count
FROM
  ML.PREDICT(MODEL taxi.taxifare_dnn, (SELECT
    fare_amount,
    hourofday,
    dayofweek,
    pickuplon,
    pickuplat,
    dropofflon,
    dropofflat,
    passenger_count
  FROM
    `taxi.taxi300k`
  WHERE
    dataset = 2
    ))
    LIMIT 10
```

```
1  SELECT
2      fare_amount,
3      predicted_fare_amount,
```

This query will process 23.22 MB when run.

Using on-demand processing quota | Processing location: US

Query results | + Create conversation | Save results | Open in

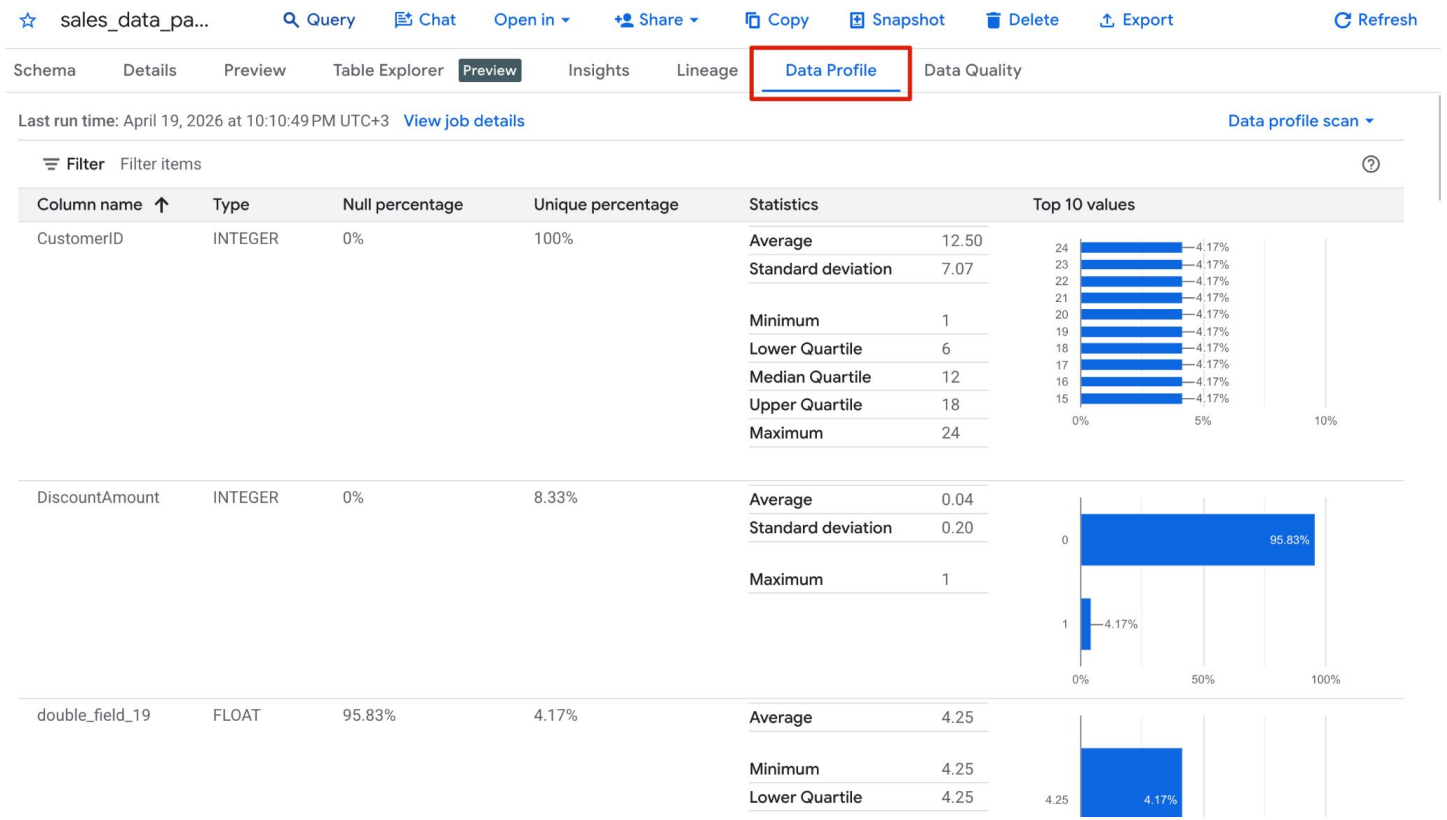
Job information | Results | Visualization | JSON | Execution details | Execution graph

Row	fare_amount	predicted_fare_a...	hourofday	dayofweek	pickuplon	pickuplat
1	4.6	7.020443916320...	4	4	-73.958026	40.77446
2	12.2	14.52977752685...	4	4	-74.006158	40.74366
3	9.8	12.53529262542...	4	4	-73.999302	40.72825
4	6.1	8.04678726196289	4	4	-74.005826	40.7402
5	4.1	7.401981353759...	4	4	-73.962267	40.76062
6	6.9	10.7291841506958	4	4	-74.006501	40.73972
7	5.7	10.26519107818...	4	4	-73.949382	40.80229
8	4.9	9.40158748626709	4	4	-73.978439	40.7413
9	4.1	9.242655754089...	4	4	-73.987827	40.74882
10	4.5	7.433818340301...	4	4	-73.960693	40.77812

Data Governance

Data Profiling

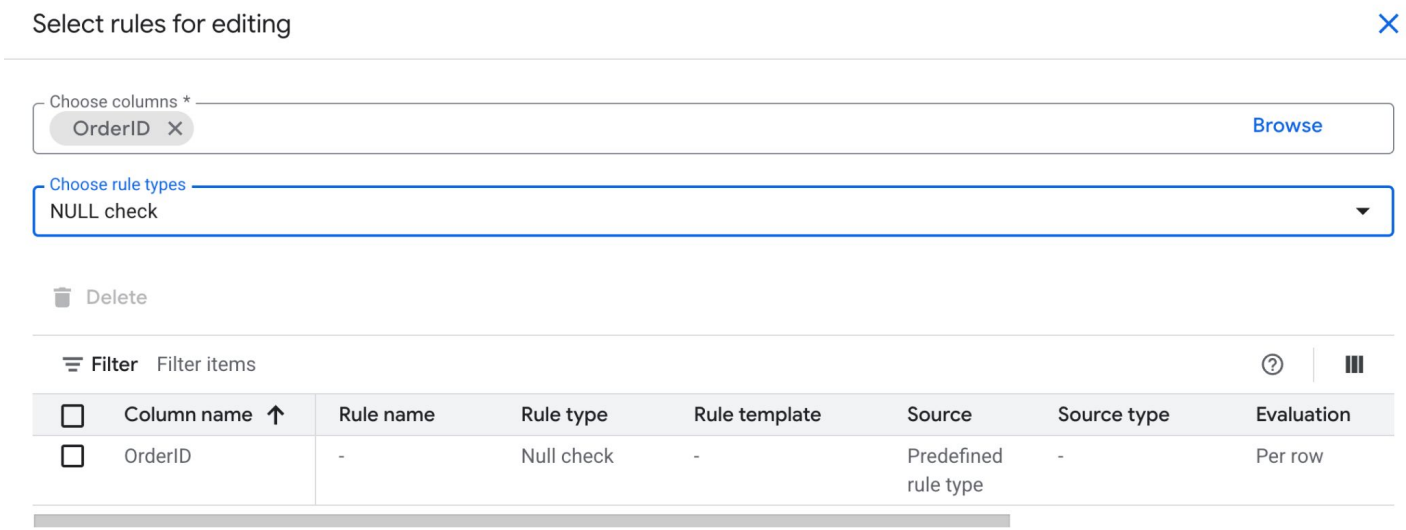
Data Quality Scan



Data Governance

Data Profiling

Data Quality Scan



Thank You!