



Data & AI Boot-Kon Event

Title: Data Canvas & Looker Studio

Goal of the lab

- To get data to insights with Natural Languages (NL)
- Without SQL Knowledge, only with NL to generate dashboards from the data

Author: Dinesh Sandra

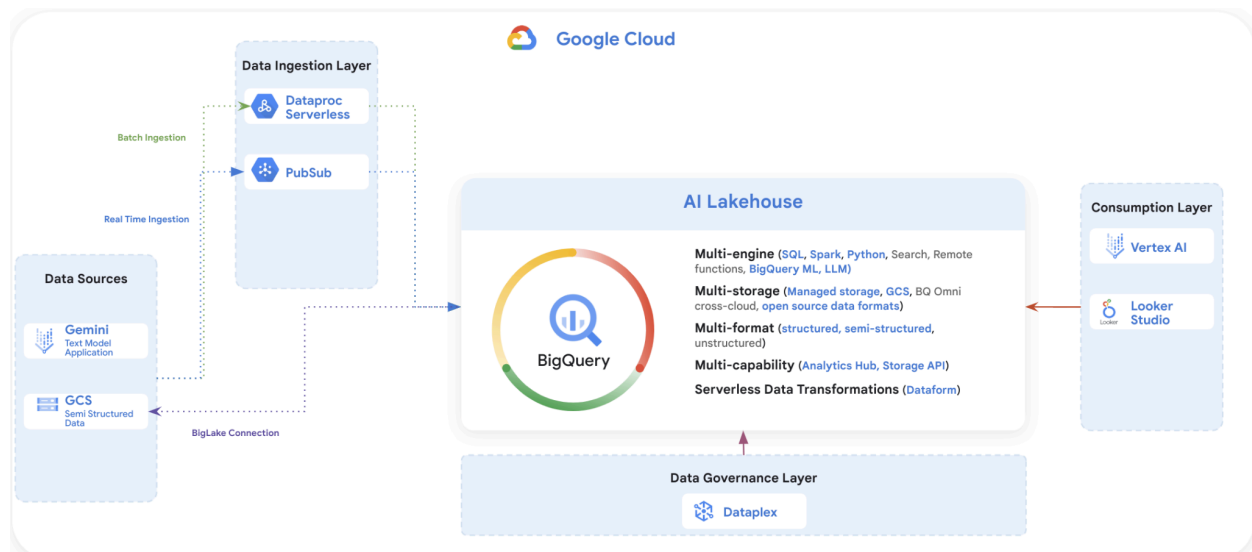
Date: 2024-04-01

Estimated Completion Time: 15 Minutes

CAUTION:

This lab is for educational purposes only and should be used with caution in production environments. Google Cloud Platform (GCP) products are changing frequently, and screenshots and instructions might become inaccurate over time. Always refer to the latest GCP documentation for the most up-to-date information.

Architecture Diagram:



What is Data Canvas?

Data Canvas is a reimagined, NL-driven experience for data professionals to work with their data. Advances in transformer based models for LLMs, retrieval, NL2SQL present an opportunity to radically simplify the data to insights work that data professionals do.

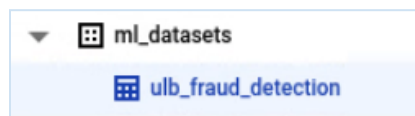
The data to insights journey for data professionals is slow and inefficient. One of the main reasons is that it's a tools-first approach. People with specialized skills with specific tools need to collaborate to be able to get insights. This orchestration between people and tools creates a lot of friction, leading to many underutilized or unutilized opportunities to get insights from an enterprise's growing data corpus.



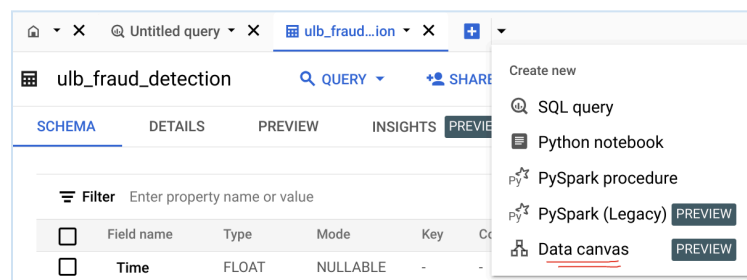
Data Canvas takes a *jobs-to-be-done* approach and offers a reimagined and context-rich user experience to help data professionals and enterprises get most out of their data. Users can use natural language to go through all the stages from data to insights. They can use it to discover data, build ingestion pipelines, prep or transform data, do advanced processing, compose analytics queries, do visualization and much more. And they can do all of this in a deeply collaborative manner.

[LAB] Data Canvas

1. Goto Google Cloud Console, under BigQuery, navigate to Datasets, Tables(to the specific table `ulb_fraud_detection`)



2. Click on dropdown button and select create data canvas as shown in the image below



3. Select the Region for example: us-central1

Select default region to store your code assets

i This setting defines the default region where your code assets will be stored. It can be updated at any time. If unsure, choose the closest region to where your data is stored. [Learn more](#)

Region *
us-central1 (Iowa)

[CANCEL](#) [SELECT](#)

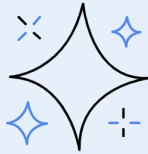
4. If you will be asked to enable the API, click on Enable API. Otherwise skip to step 5.



Introducing BigQuery data canvas

BigQuery data canvas lets you discover, transform, query, and visualize data with natural language powered by Gemini in BigQuery. You can work with data canvases in a directed acyclic graph (DAG) experience to help you visualize your non-linear workflows. To use data canvas in your Google Cloud projects, you must enable an API. [Learn more](#)

ENABLE DATA CANVAS



5. Enable Gemini APIs

Enable Gemini

✓ **Join the Gemini for Program**

2 **Required APIs**

i Gemini is a collection of AI models on Google Cloud. You can use Gemini with code completion and more.

[Cloud AI Companion API](#) ?

ENABLE

6. Type the **table name** that you want to explore data

🔍 **Search for tables using Data Catalog keyword search. E.g. 'trees California;**

- 🗄 fraudfix-bcootkon.ml_datasets.ulb_fraud_detection
- 🗄 fraudfix-bcootkon.ml_datasets.ulb_fraud_detection_blake
- 🗄 fraudfix-bcootkon.ml_datasets.ulb_fraud_detection_parquet
- 🗄 dineshsandra-62-20240220110632.tx.customers_20240220

7.

8. Click on Query and type the question, let's ask the NL question, example...

Which top 10 transaction had the highest amount and time and v1?



ulb_fraud_detection_parquet

SCHEMA DETAILS

Filter Enter property name or value

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
Time	FLOAT	NULLABLE	-	-	-	-	-
V1	FLOAT	NULLABLE	-	-	-	-	-
V2	FLOAT	NULLABLE	-	-	-	-	-
V3	FLOAT	NULLABLE	-	-	-	-	-
V4	FLOAT	NULLABLE	-	-	-	-	-
V5	FLOAT	NULLABLE	-	-	-	-	-
V6	FLOAT	NULLABLE	-	-	-	-	-
V7	FLOAT	NULLABLE	-	-	-	-	-
V8	FLOAT	NULLABLE	-	-	-	-	-
V9	FLOAT	NULLABLE	-	-	-	-	-

+

Branch another node

QUERY JOIN

SQL

Prompt to generate an SQL query. E.g. 'Find total items by species in descending order'

RUN SAVE SCHEDULE

Type a query to get started

1

Press Alt+F1 for Accessibility Options.

9. Click **Run**



SQL

Which top 10 transaction had the highest amount and time and v1?

[RUN](#) [SAVE](#) [SCHEDULE](#) Query completed.

```
1 # prompt: Which top 10 transaction had the highest amount and time and v1?
2
3 SELECT
4   t1.Time,
5   t1.Amount,
6   t1.V1
7 FROM
8   'genai-demo-2024.ml_datasets.ulb_fraud_detection_parquet' AS t1
9 ORDER BY
10  t1.Amount DESC
11 LIMIT
12  10
```

Rate this suggestion [👍](#) [👎](#) [📄](#) Press Alt+F1 for Accessibility Options.

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

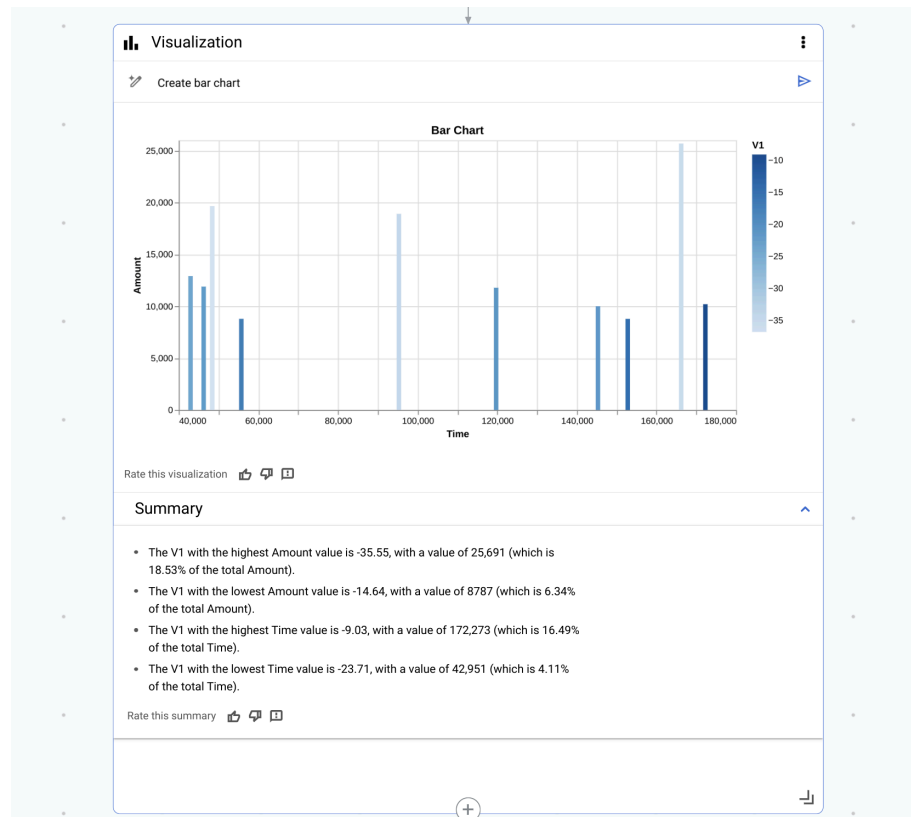
Row	Time	Amount	V1
1	166198.0	25691.16	-35.5485392...
2	48401.0	19656.53	-36.8023199...
3	95286.0	18910.0	-34.5492960...
4	42951.0	12910.93	-23.7128394...
5	46253.0	11898.09	-21.7806653...
6	119713.0	11789.84	-20.9248974...
7	172273.0	10199.44	-9.03053758...
8	145283.0	10000.0	-21.5324775...
9	55709.0	8790.26	-16.9500636...
10	152763.0	8787.0	-14.6417095...

[+](#)

Branch another node

[QUERY THESE RESULTS](#) [VISUALIZE](#) [JOIN](#)

10. Then Click **Visualize**, select whatever chart you would like to visualize. Notice the Summary which gives you a quick summary of observations,



Conclusion: You see there are many options and possibilities to explore the data from BigQuery into reports via Natural Language as input.



ub_fraud_detection_parquet

SCHEMA DETAILS

Filter Enter property name or value

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
Time	Float	Nullable	-	-	-	-	-
V1	Float	Nullable	-	-	-	-	-
V2	Float	Nullable	-	-	-	-	-
V3	Float	Nullable	-	-	-	-	-
V4	Float	Nullable	-	-	-	-	-
V5	Float	Nullable	-	-	-	-	-
V6	Float	Nullable	-	-	-	-	-
V7	Float	Nullable	-	-	-	-	-
V8	Float	Nullable	-	-	-	-	-
V9	Float	Nullable	-	-	-	-	-

SQL

Which top 10 transaction had the highest amount and time and v1?

Query completed

```
1 # prompt: Which top 10 transaction had the highest amount and time and v1?
2
3 SELECT
4   t1.Time,
5   t1.Amount,
6   t1.V1
7 FROM
8   `gcp://demo-2024-01-datasets.ub_fraud_detection_parquet` AS t1
9 ORDER BY
10  t1.Amount DESC
11 LIMIT
12  10
```

Query results

Row	Time	Amount	V1
1	166198.0	25691.16	-35.5485392...
2	484491.0	19638.53	-36.8023196...
3	92384.0	18910.0	-34.5457960...
4	42951.0	12910.93	-23.7128394...
5	46253.0	11898.09	-21.7886653...
6	119713.0	11789.84	-20.9248974...
7	172273.0	10199.44	-9.0303758...
8	143288.0	10090.0	-21.0147751...
9	52759.0	8799.26	-16.9596306...
10	152763.0	8787.0	-14.6417095...

Visualization

Create chart

Bar Chart

Summary

- The V1 with the highest Amount value is -35.55, with a value of 25,691 (which is 18.53% of the total Amount).
- The V1 with the lowest Amount value is -14.64, with a value of 8,787 (which is 6.34% of the total Amount).
- The V1 with the highest Time value is -9.03, with a value of 172,273 (which is 16.49% of the total Time).
- The V1 with the lowest Time value is -23.71, with a value of 42,951 (which is 4.11% of the total Time).

🎉🎉 Congratulations on completing Lab 7! 🎉🎉