Final Project

Project 1 – Project 2 – Project 3 – Project 4

DigitalSkola

fantastic 4

Kelompok 13

project team ↘

- Muhammad Ramli
- Ozza Dinata
- Donna Setiawan
- Ginta Khairunisa
- Satria Said

Project Overview ↘

Page

04 Project 1 – Business Inteligence (BI)

10 Project 2 – Data Build Tool (DBT)

18 Project 3 – Airflow

29 Project 4 – Kafka (Stream Processing)

Project 1 : Business Inteligence

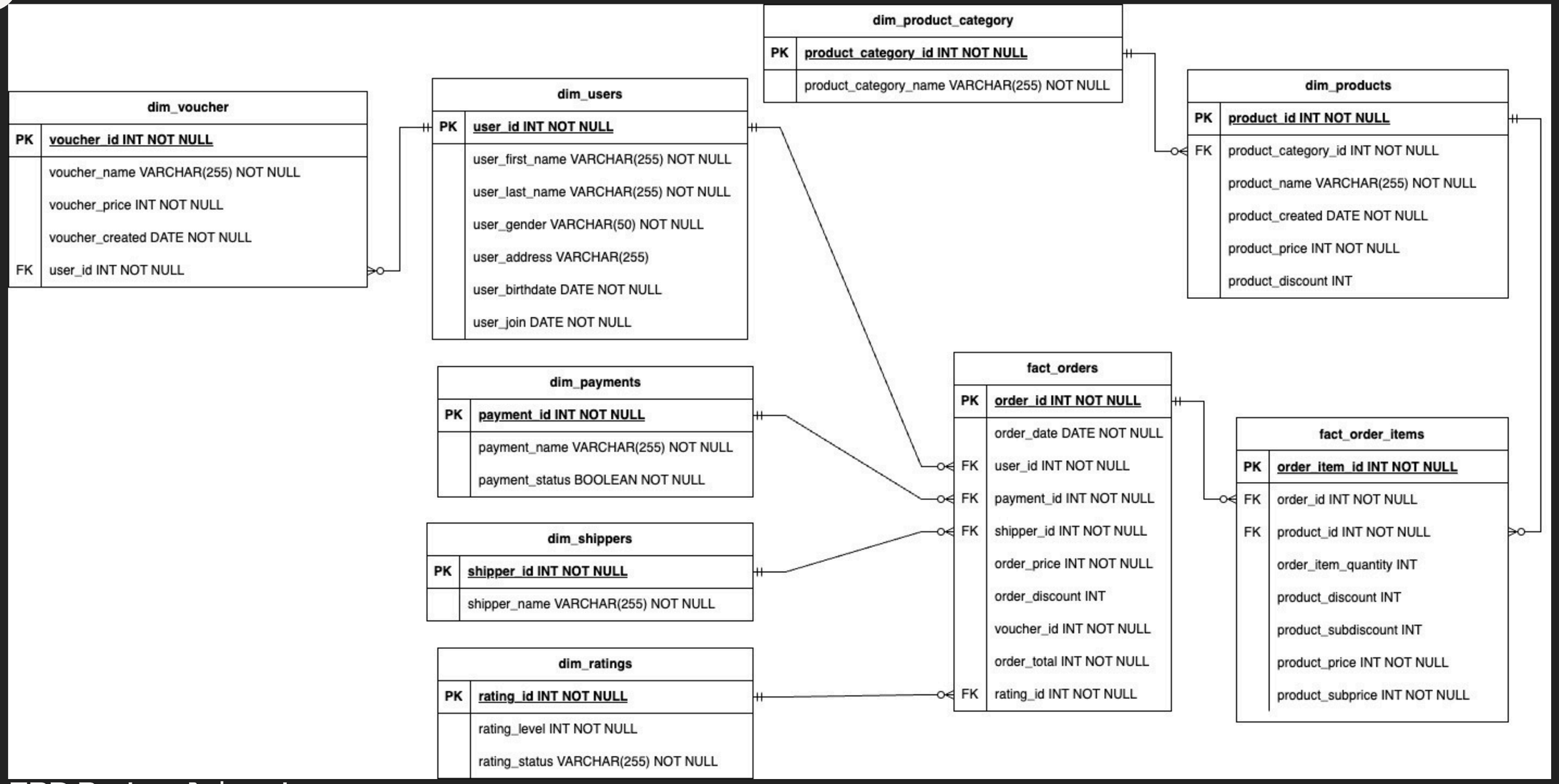
Objective Project

- Desain ERD DWH,
- Menjalankan ETL Datamart dengan python
- Desain Dashboard sesuai dengan Business Requirements Di Looker Studio

Tools yang digunakan dalam project

Project 1 : Result

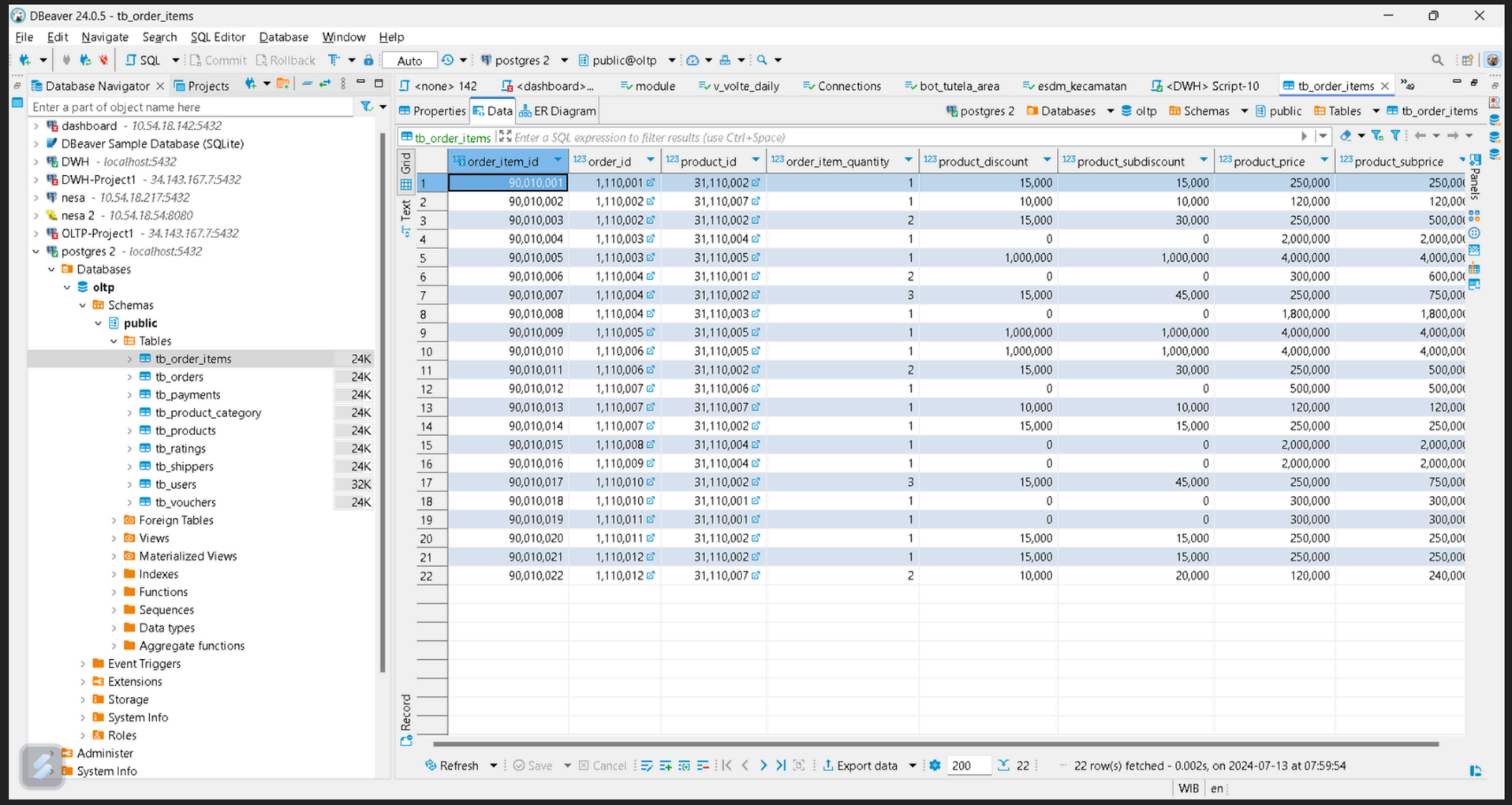
1 Membuat Desain ERD DWH untuk mengetahui alur data dan penentuan fact tables serta dim tables



ERD Project 1.drawio

Project 1 : Result

2 Lampiran screenshoot hasil database OLTP dan script ddl oltp yg dirunning berhasil dgn tools dbeaver



Project 1 : Result

3 Menjalankan etl data warehouse menggunakan jupyter notebook

etl.py

jupyter etl Last Checkpoint: 6 hours ago

File Edit View Run Kernel Settings Help

Code

```
Transform Data shippers Success
dim_shipper
Load Data dim_shipper Success
Extract Data tb_ratings Success
Transform Data ratings Success
dim_rating
Load Data dim_rating Success
Extract Data tb_vouchers Success
Transform Data vouchers Success
dim_voucher
Load Data dim_voucher Success
Extract Data tb_orders Success
Transform Data orders Success
fact_orders
Load Data fact_orders Success
Extract Data tb_product_category Success
Transform Data product_category Success
dim_product_category
Load Data dim_product_category Success
Extract Data tb_products Success
Transform Data products Success
dim_product
Load Data dim_product Success
Extract Data tb_order_items Success
Transform Data order_items Success
fact_order_items
Load Data fact_order_items Success
Data Mart Has Create Success
```

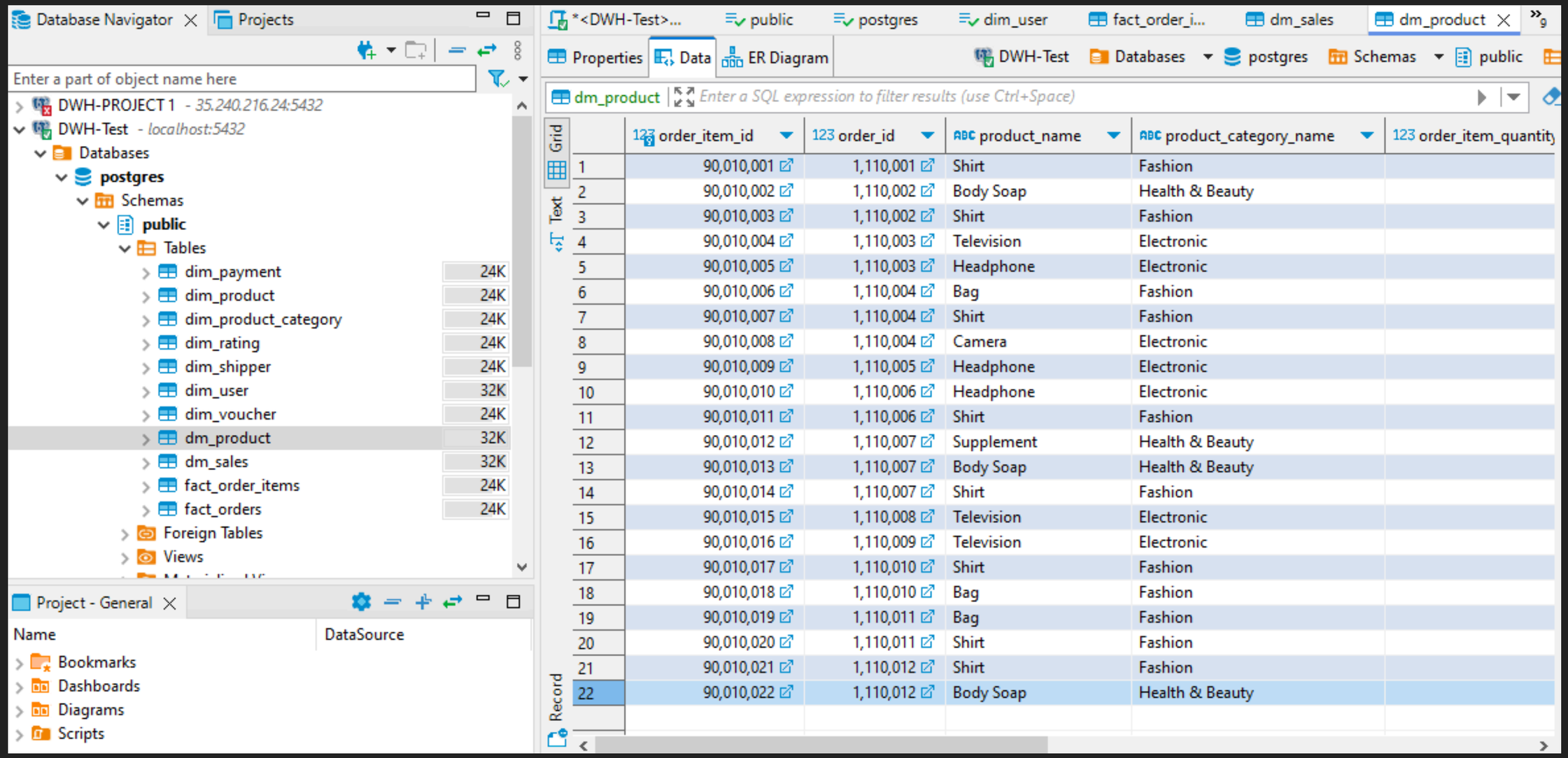
config.py

```
ddl_marts = {
    "dim_sales": """
        CREATE TABLE IF NOT EXISTS dm_sales (
            order_id INT NOT NULL PRIMARY KEY,
            order_date DATE NOT NULL,
            month_year DATE NOT NULL,
            user_id INT NOT NULL,
            user_name VARCHAR(255),
            user_gender VARCHAR(255),
            user_address VARCHAR(255),
            payment_type VARCHAR(255),
            shipper_name VARCHAR(255),
            order_price INT NOT NULL,
            order_discount INT,
            voucher_name VARCHAR(255),
            order_total INT NOT NULL,
            FOREIGN KEY (order_id) REFERENCES fact_orders(order_id),
            FOREIGN KEY (user_id) REFERENCES dim_user(user_id)
        );
    """,
    "dim_product_sales": """
        CREATE TABLE IF NOT EXISTS dm_product (
            order_item_id INT NOT NULL PRIMARY KEY,
            order_id INT NOT NULL,
            product_name VARCHAR(255) NOT NULL,
            product_category_name VARCHAR(255) NOT NULL,
            order_item_quantity INT,
            product_discount INT,
            product_subdiscount INT,
            product_price INT NOT NULL,
            product_subprice INT NOT NULL,
            margin_laba INT,
            FOREIGN KEY (order_item_id) REFERENCES fact_order_items(order_item_id),
            FOREIGN KEY (order_id) REFERENCES fact_orders(order_id)
        );
    """,
```

Lampiran screenshoot script etl yang dijalankan di jupyter notebook

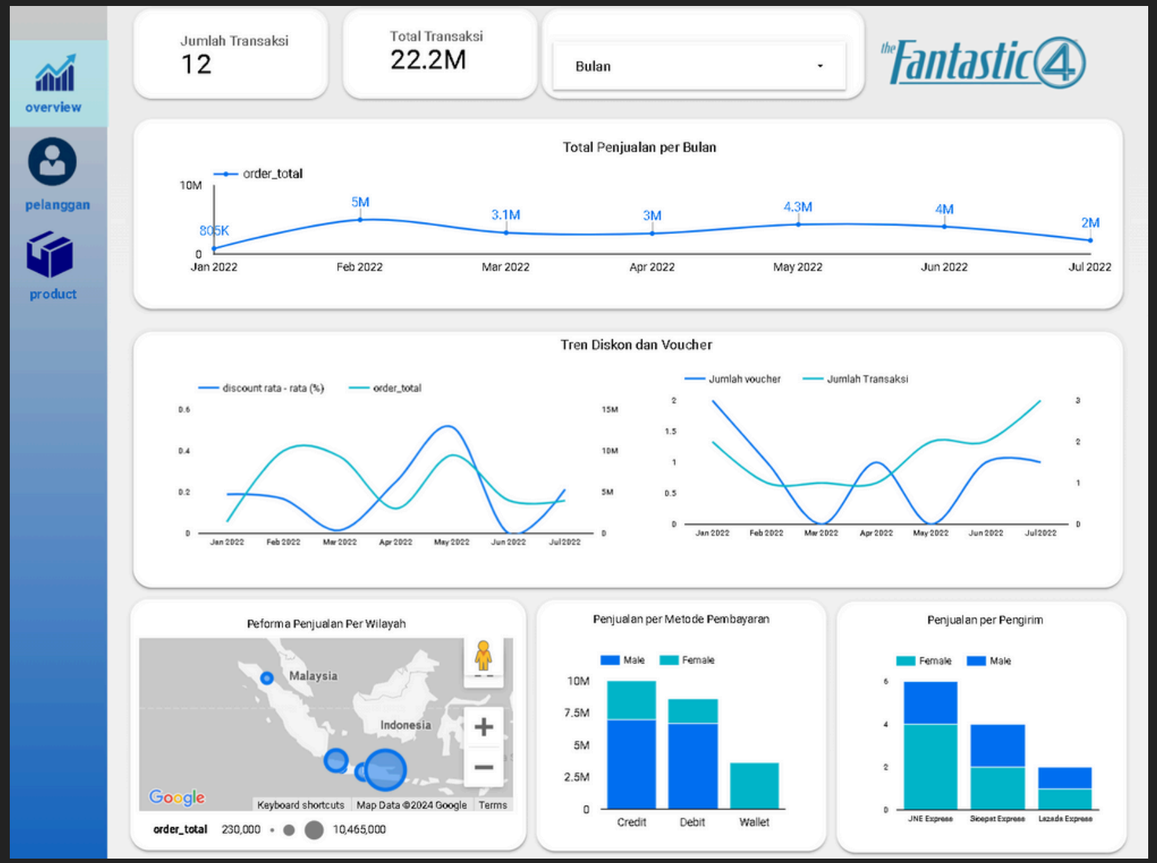
Project 1: Result

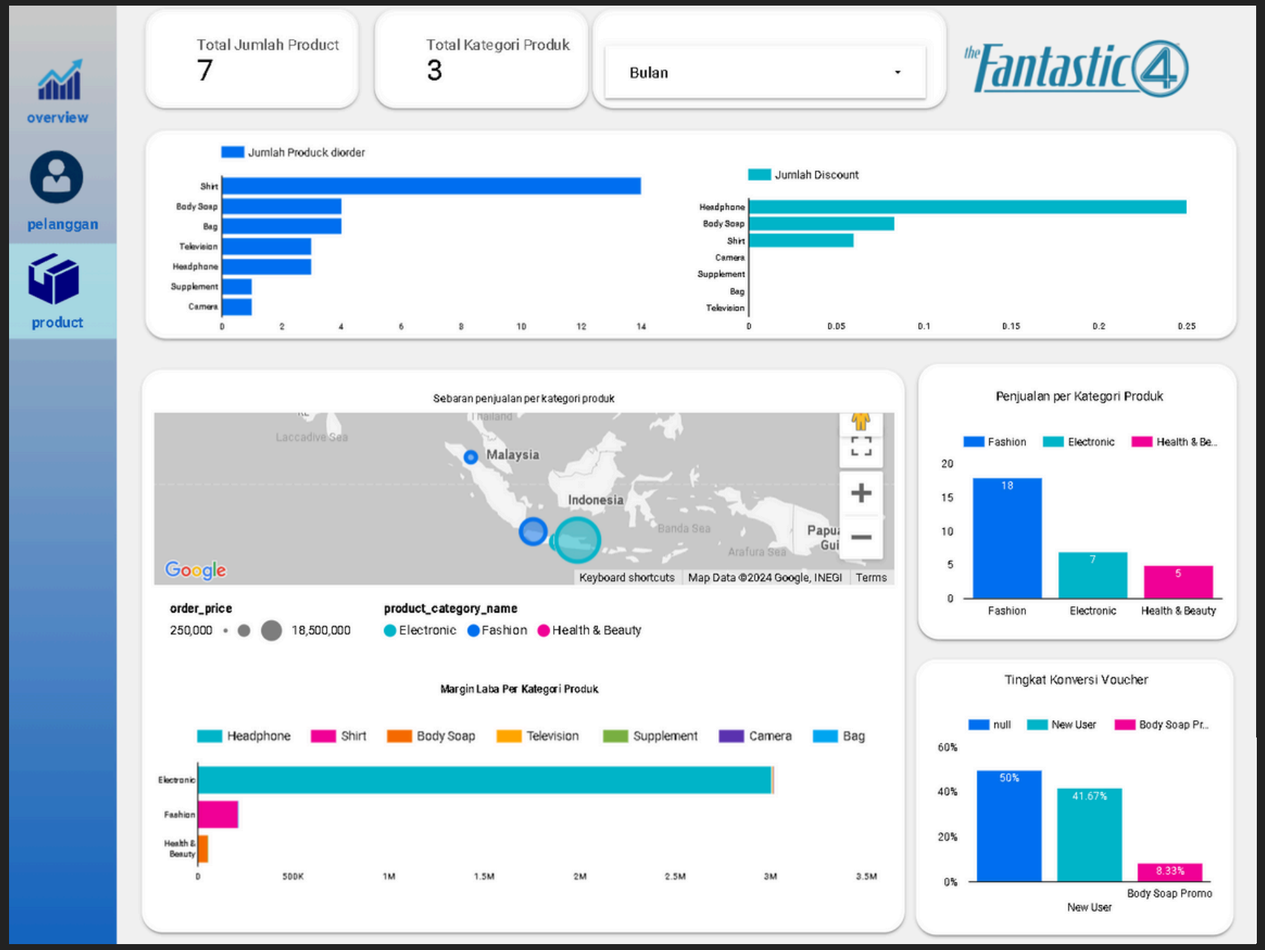
Output Data Mart



Project 1 : Result

4 Membuat report dashboard Data Penjualan E- Commerce di looker studio





Hasil Insight Report di Looker Studio

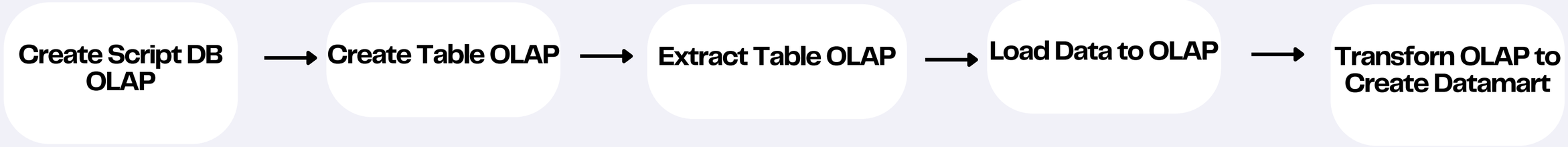
https://lookerstudio.google.com/s/jZHzA-xFNEE

Project 2 : DBT

Objective Project

Impelementasi Data Pipeline End to End dengan menggunakan DBT, Docker, postgre SQL, python

Data Pipeline



Tools yang digunakan dalam project

dbt

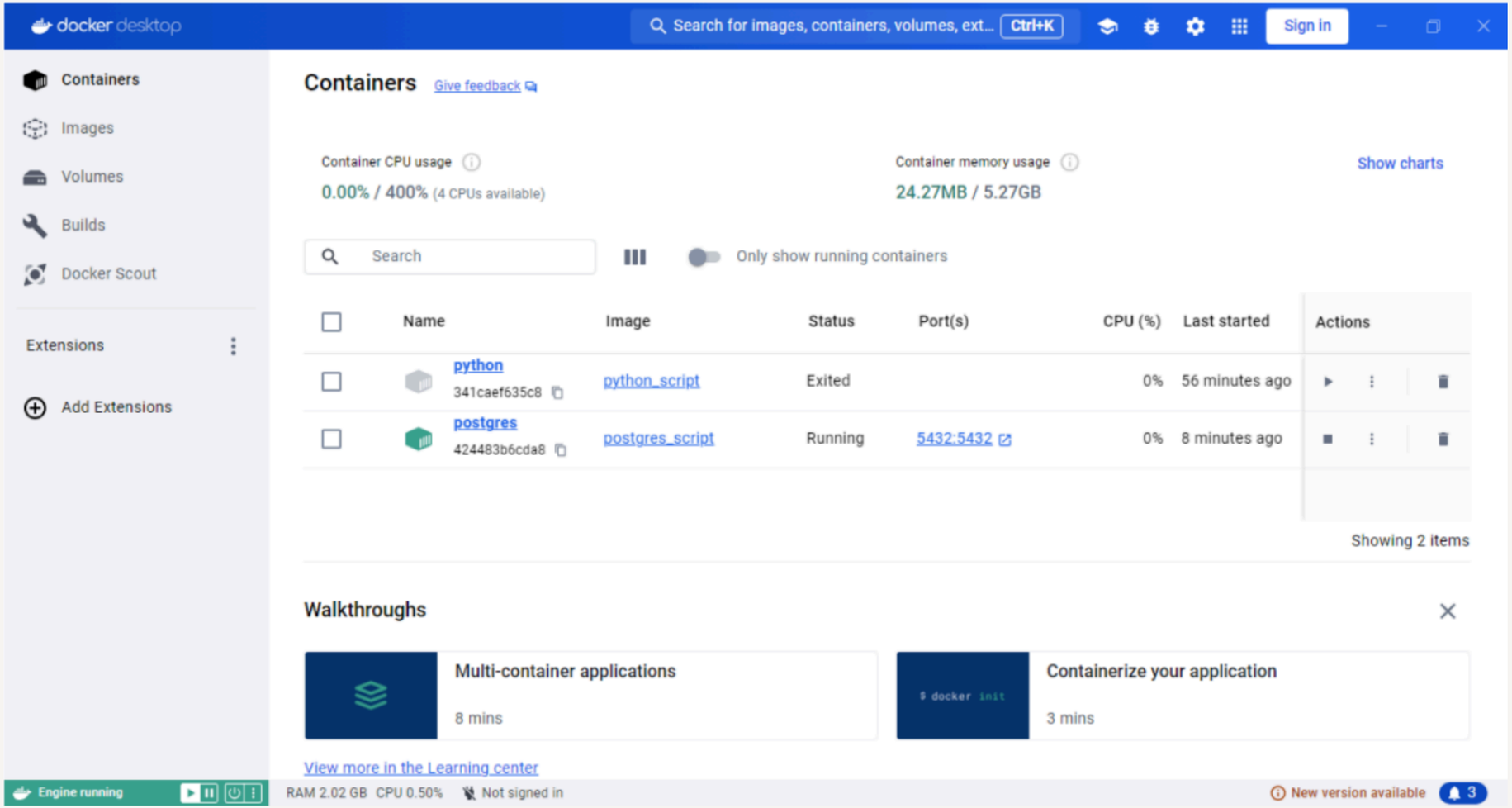
PostgreSQL

docker

python

Project 2 – DBT

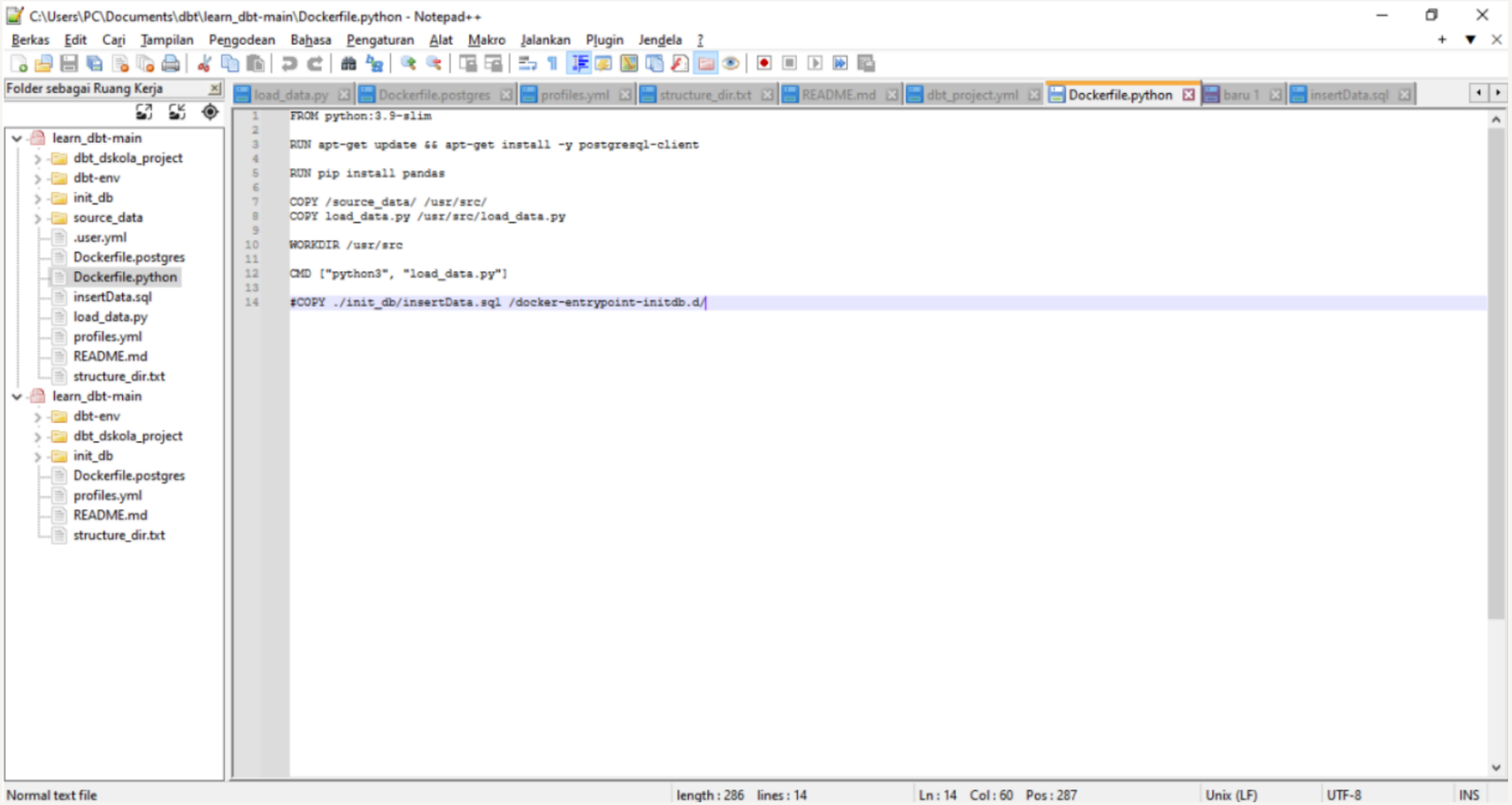
01 Menjalankan Docker Container



Lampiran Gambar Docker Container Run

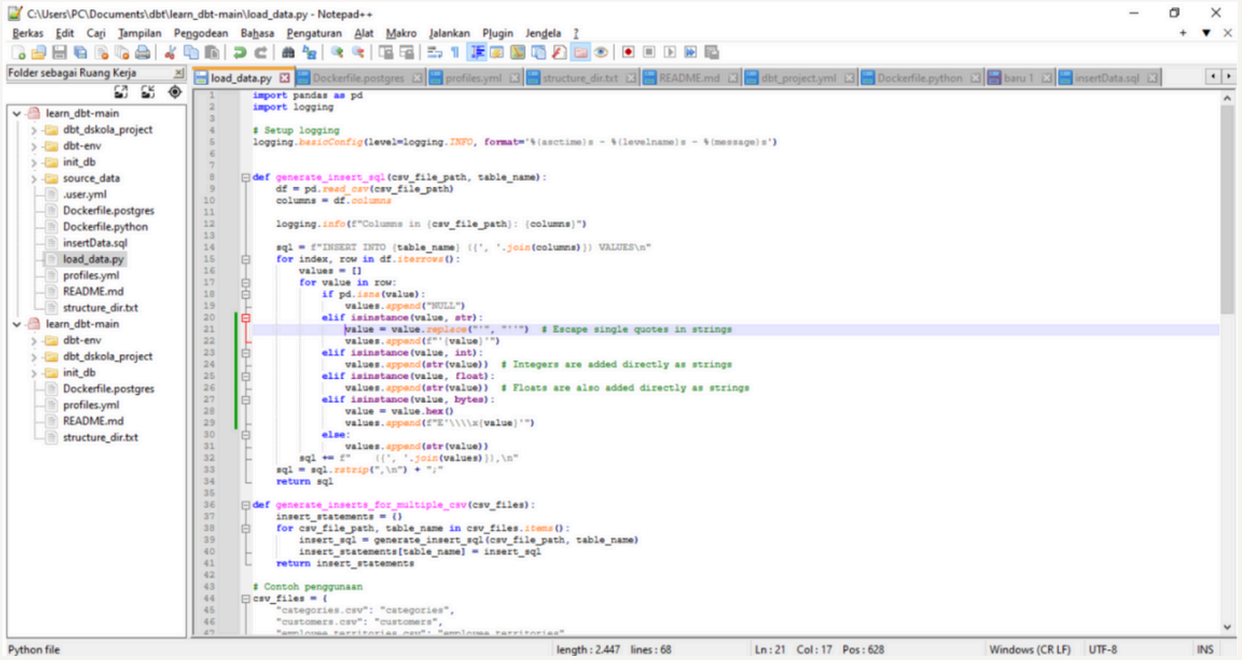
Project 2 – Result ↘

Lampiran Script Dockerfile Python yang akan dijalankan



Project 2 – Result ↘

Tampilan Script Python Load data dari file CSV

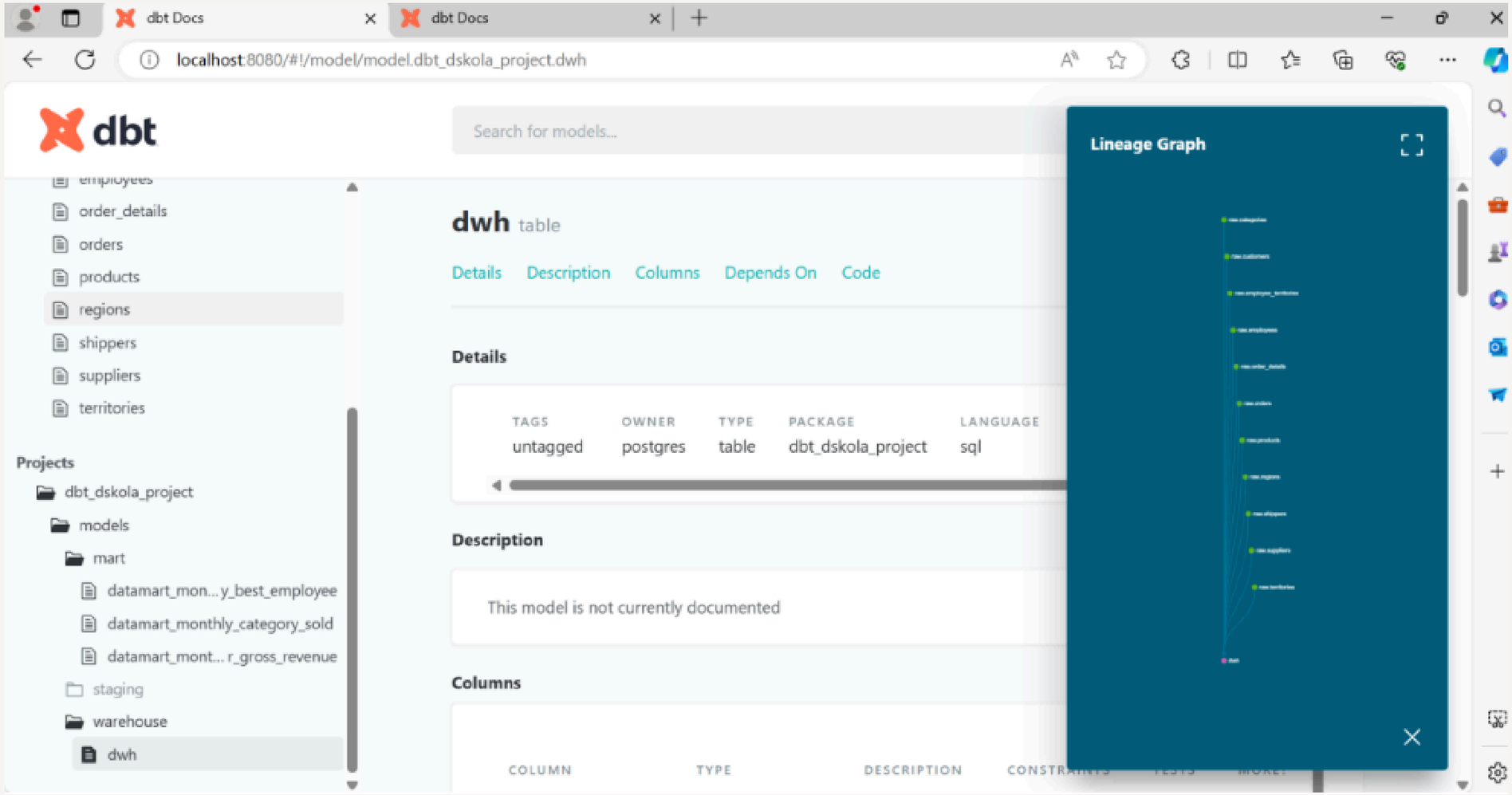



```
db_dskola=# \l
db_dskola=# \dt
                List of relations
 Schema |         Name          | Type  |  Owner
--------+-----------------------+-------+----------
 public | categories            | table | postgres
 public | customers             | table | postgres
 public | employee_territories  | table | postgres
 public | employees             | table | postgres
 public | order_details         | table | postgres
 public | orders                | table | postgres
 public | products              | table | postgres
 public | regions               | table | postgres
 public | shippers              | table | postgres
 public | suppliers             | table | postgres
 public | territories           | table | postgres
(11 rows)
```

Berikut tampilan Output Table Database db_skola

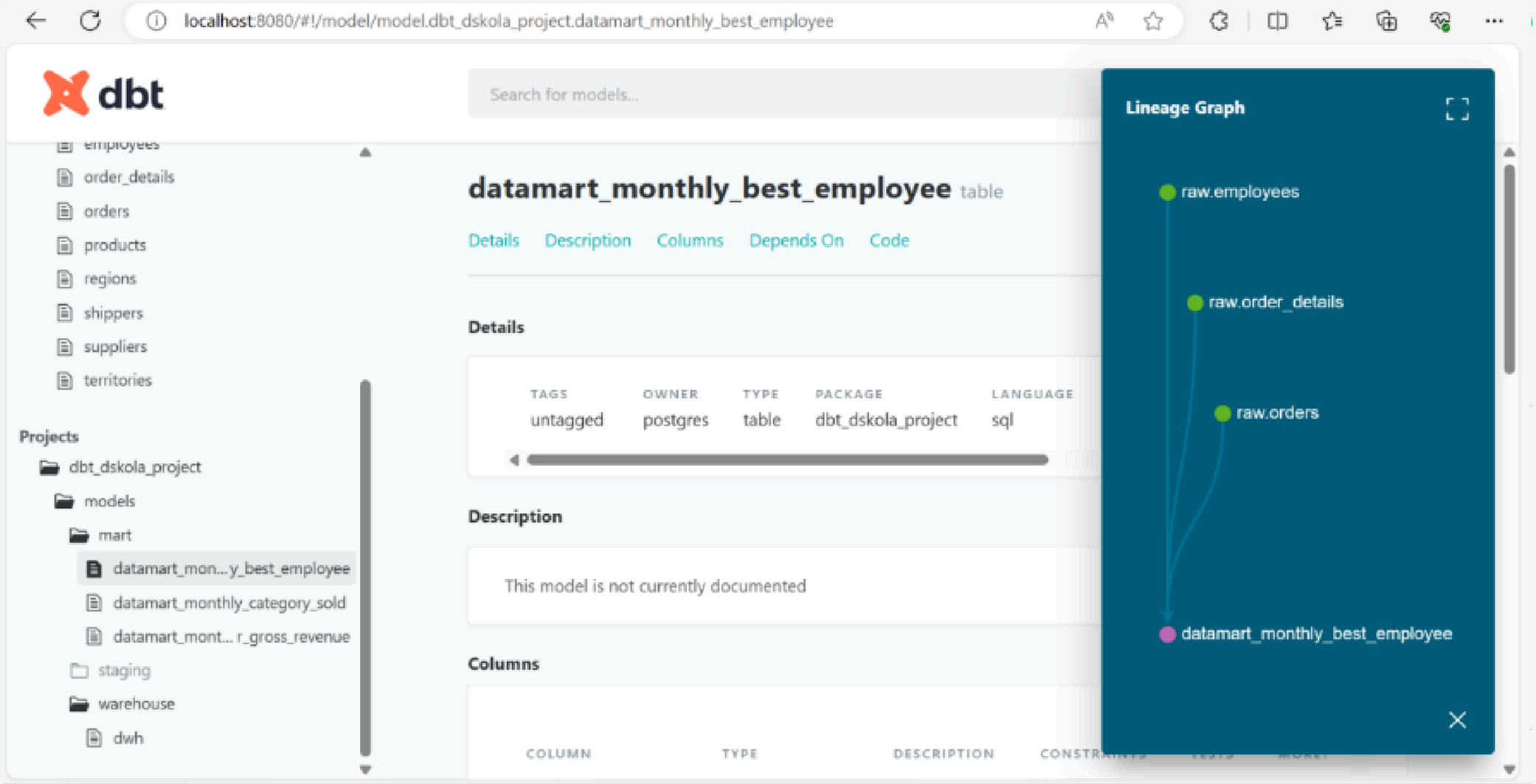
Project 2 – Result

02 Lampiran hasil data warehouse yang sudah dijalankan di DBT



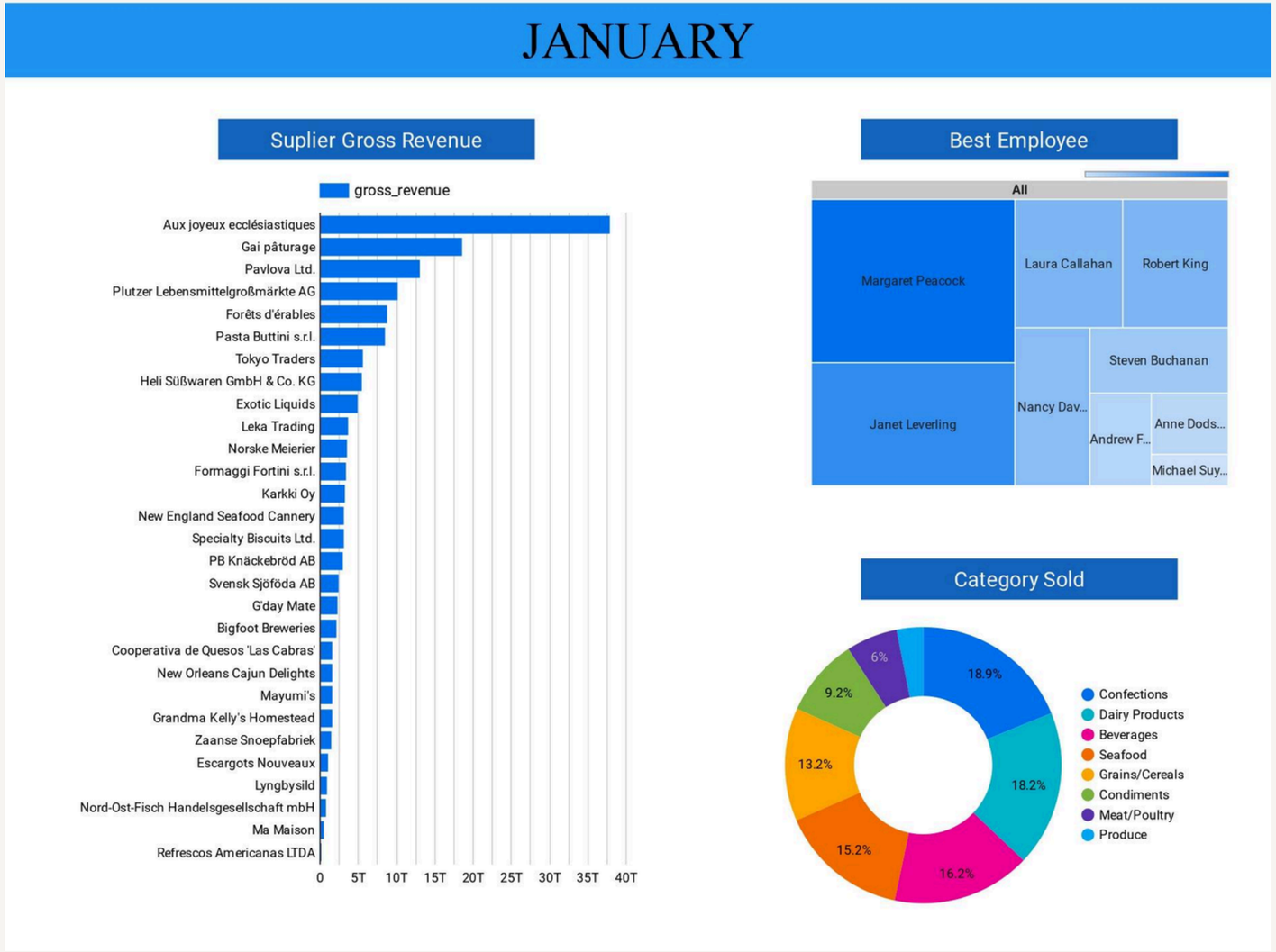
Project 2 – Result

03 Lampiran gambar data mart di DBT



Project 2 – Result

04 Dashboard hasil dari data end to end proses dari dbt menggunakan Looker Studio

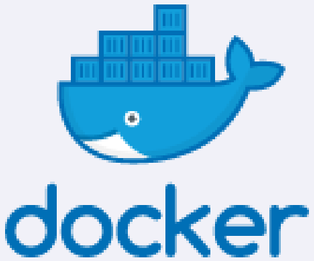


Project 3 : Airflow

Objective Project

Impelementasi Data Pipeline End to End dengan menggunakan Docker, Apache Airflow , Posgtre SQL dan Snowflake

Tools yang digunakan dalam project

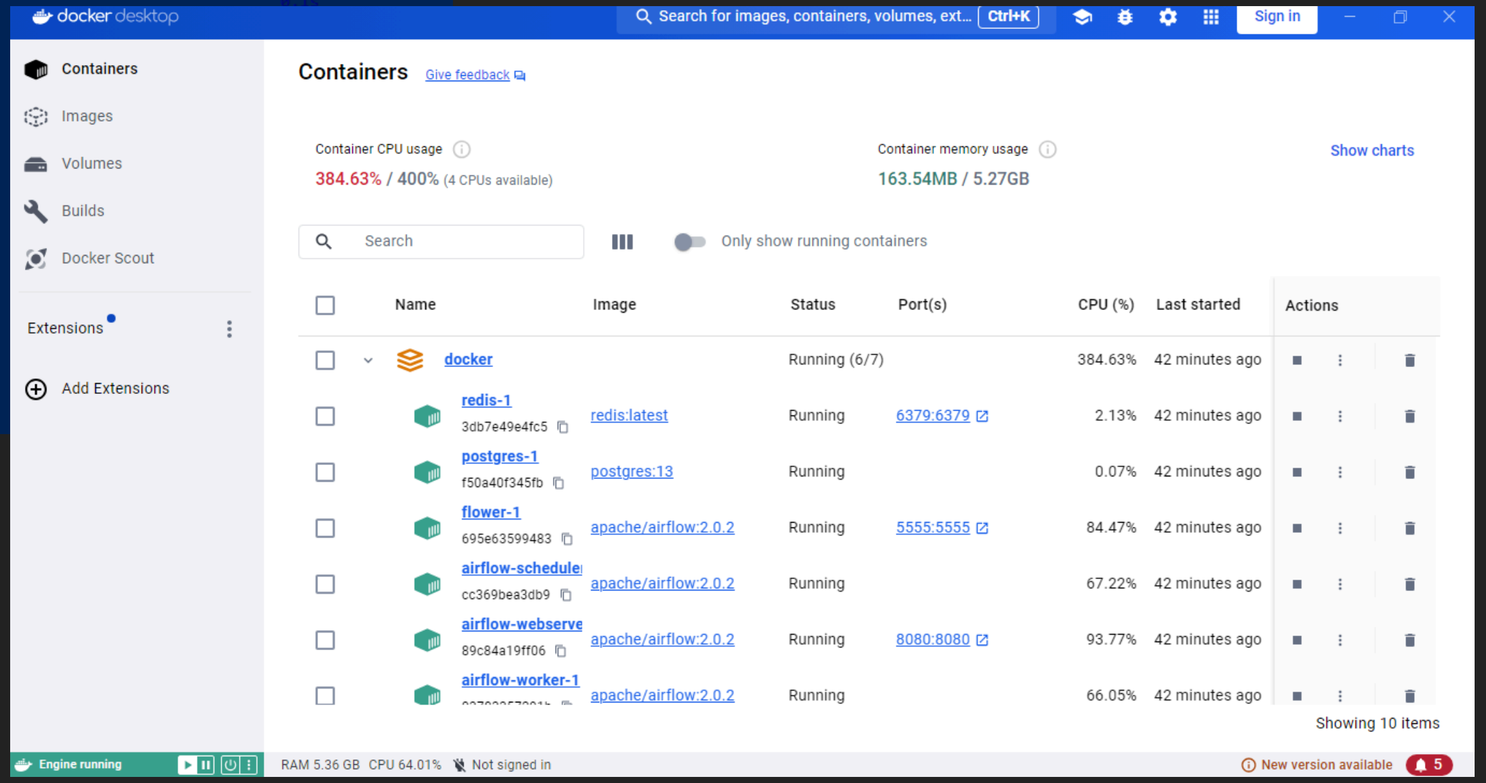



Project 3 : AIRFLOW

Windows PowerShell

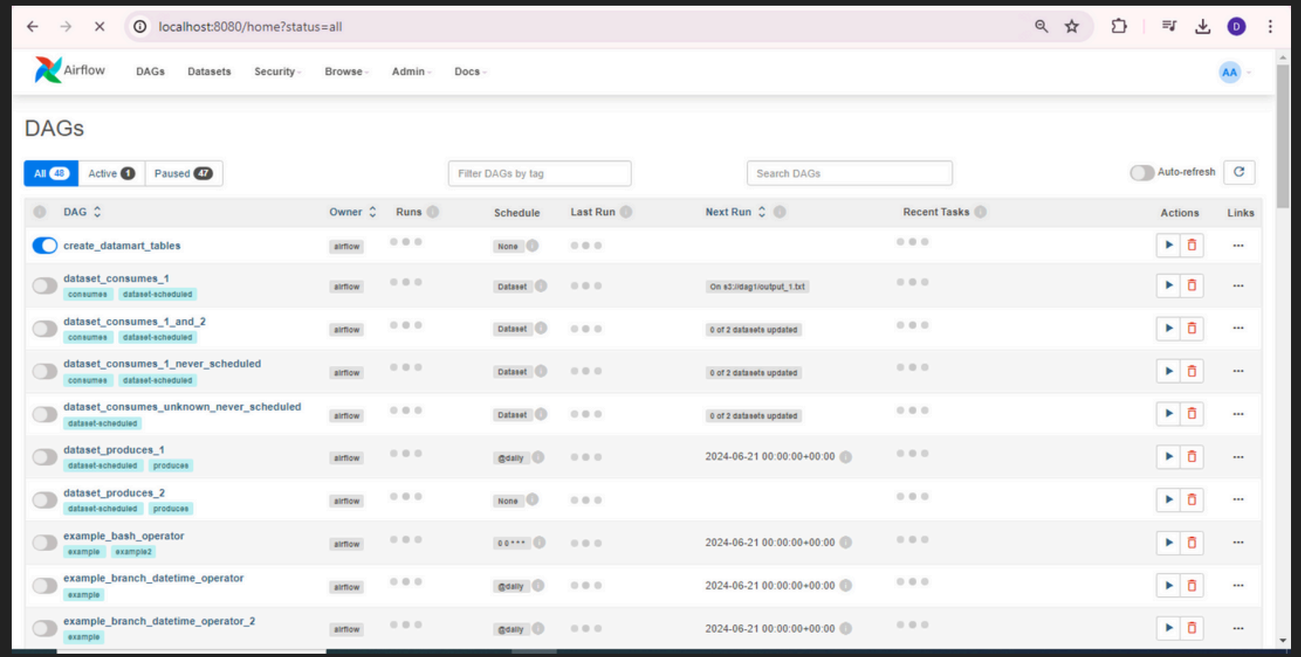
```
      --tls                Use TLS; implied by --tlsverify
      --tlscacert string   Trust certs signed only by this CA (default
                           "C:\\Users\\PC\\.docker\\ca.pem")
      --tlscert string     Path to TLS certificate file (default
                           "C:\\Users\\PC\\.docker\\cert.pem")
      --tlskey string      Path to TLS key file (default
                           "C:\\Users\\PC\\.docker\\key.pem")
      --tlsverify          Use TLS and verify the remote
  -v, --version            Print version information and quit

Run 'docker COMMAND --help' for more information on a command.

For more help on how to use Docker, head to https://docs.docker.com/go/guides/

[+] Running 8/8docker> docker-compose up -d
 ⠿ Network docker_default                 Created
 ⠿ Container docker-redis-1              Healthy
 ⠿ Container docker-postgres-1           Healthy
 ⠿ Container docker-airflow-worker-1     Started
 ⠿ Container docker-flower-1             Started
 ⠿ Container docker-airflow-webserver-1  Started
 ⠿ Container docker-airflow-init-1       Started
 ⠿ Container docker-airflow-scheduler-1  Started
PS C:\Users\PC\docker>
```

1 Tampilan running airflow di docker (Windows Shell)

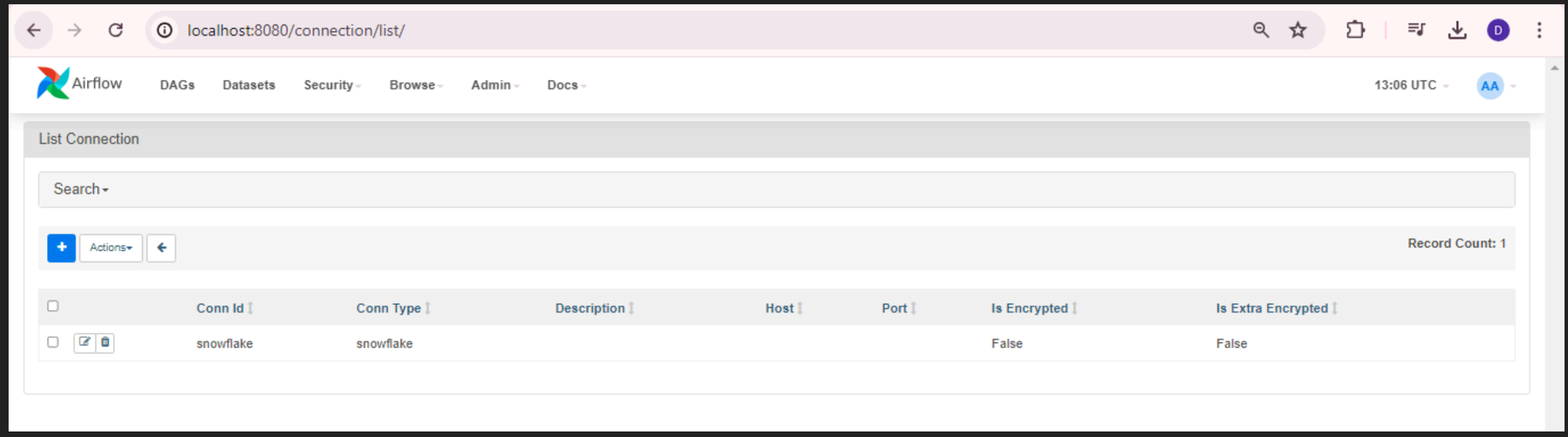


2 Tampilan Docker Running UI

Project 3 : AIRFLOW

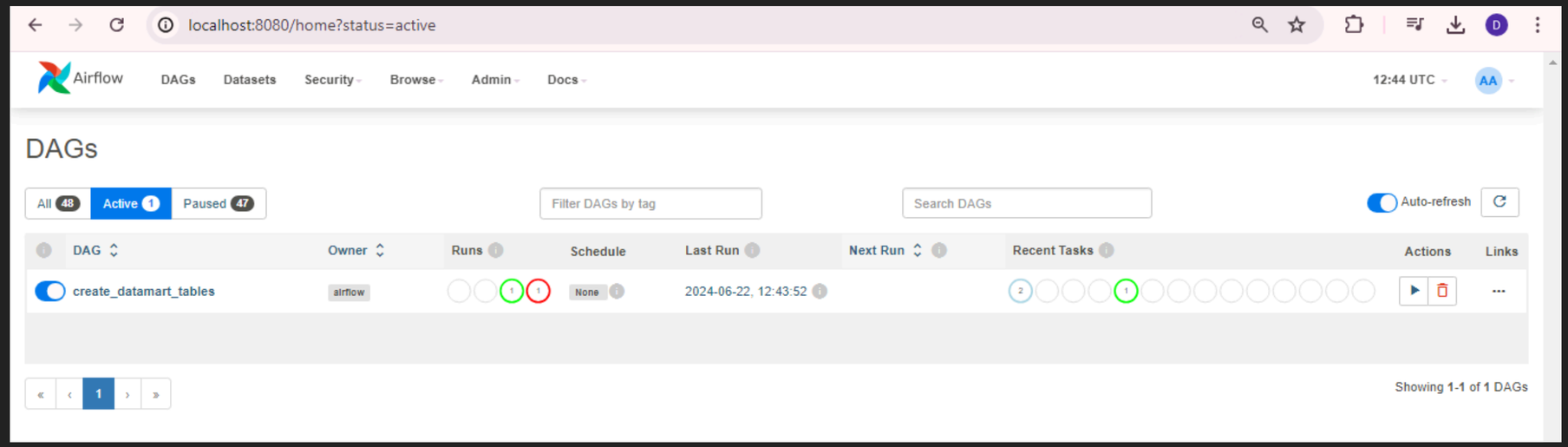


3 Tampilan halaman airflow

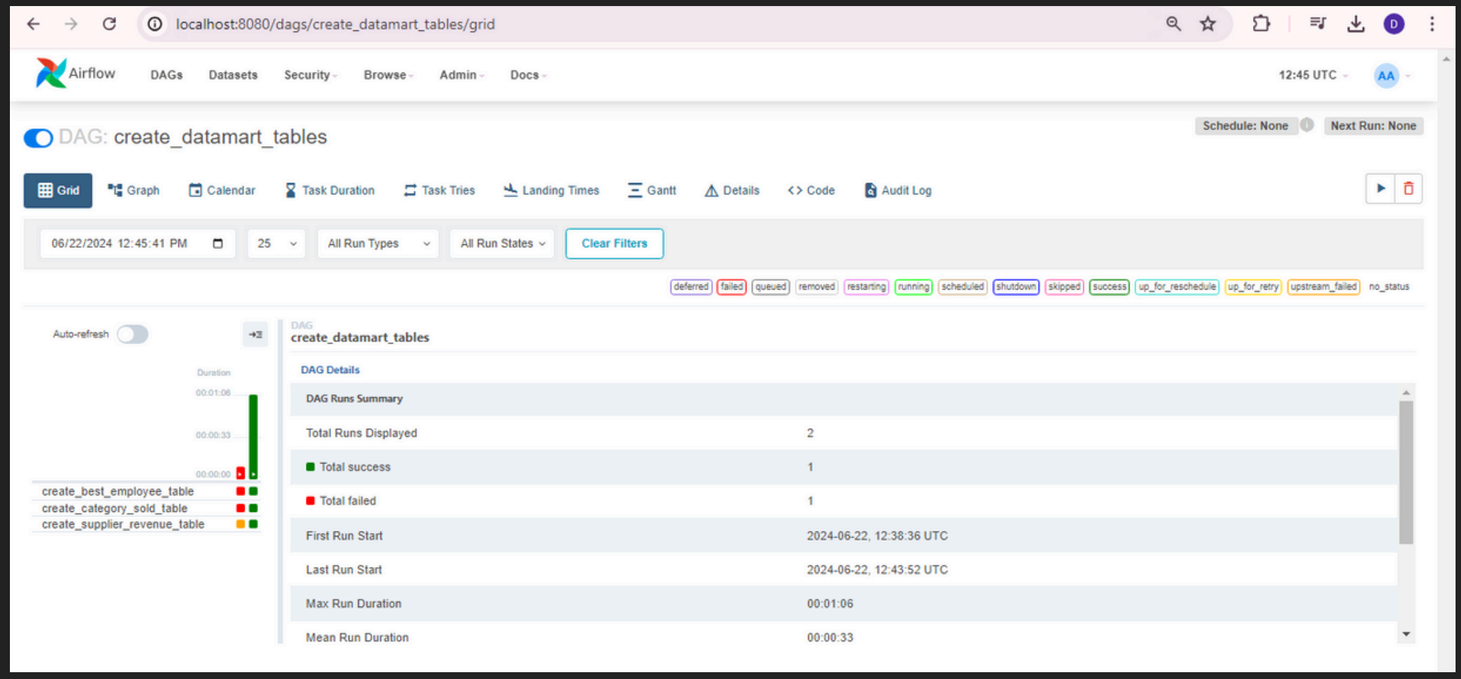


4 Tampilan Koneksi ke snowflake

Project 3 : AIRFLOW

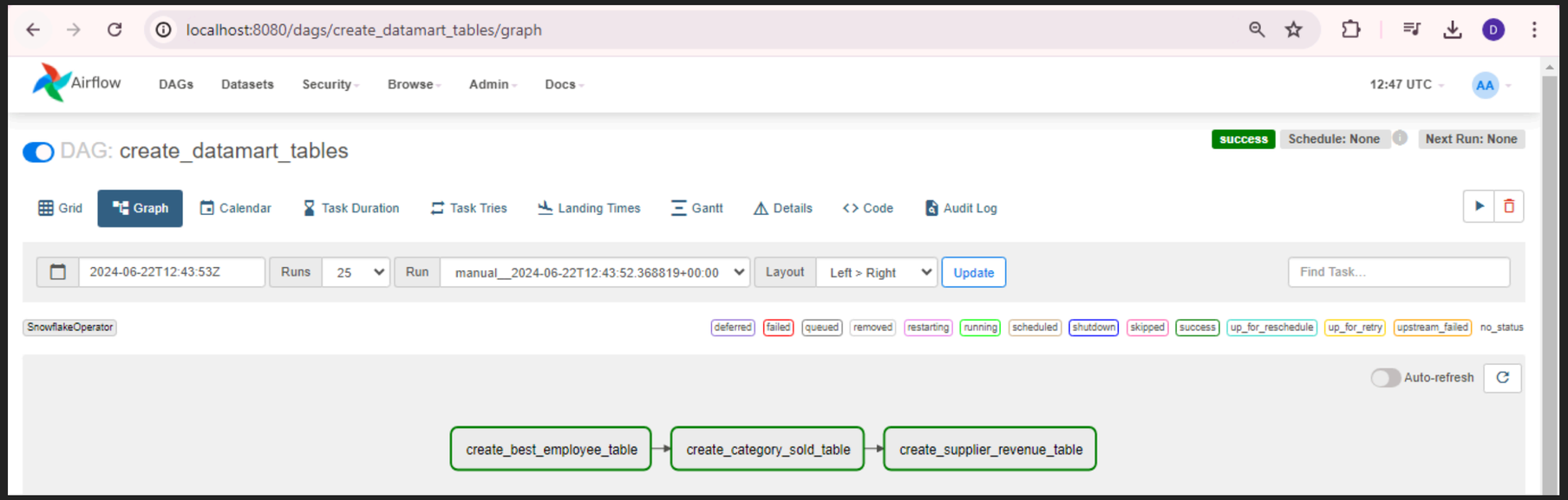


5 Tampilan saat runing DAG datamart

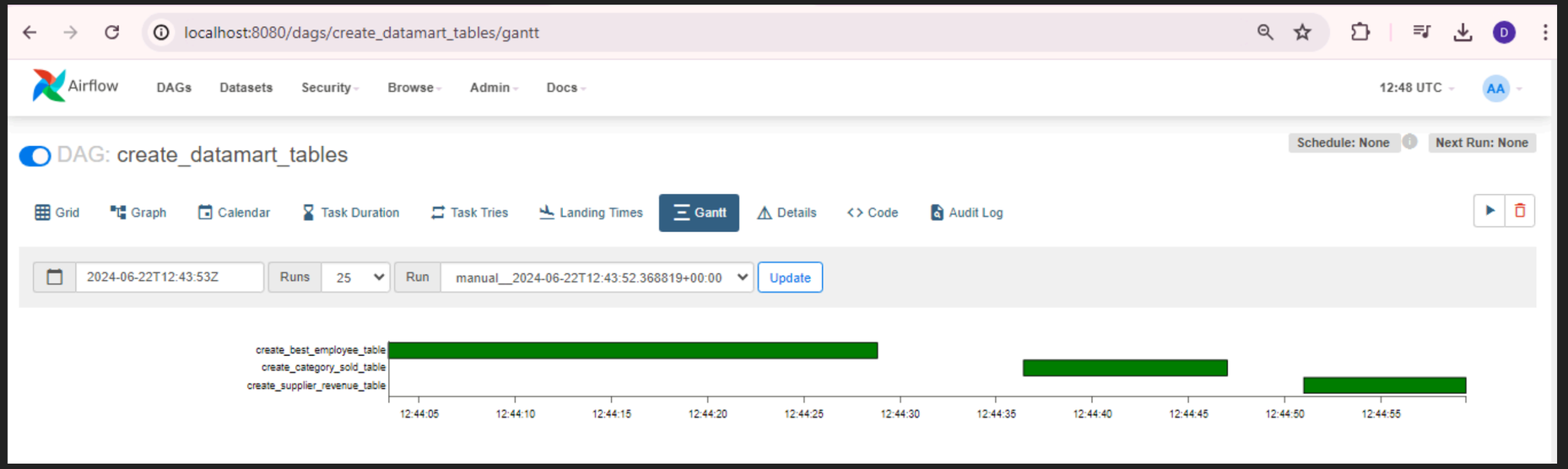


6 Tampilan di Grid Running Datamart

Project 3 : AIRFLOW

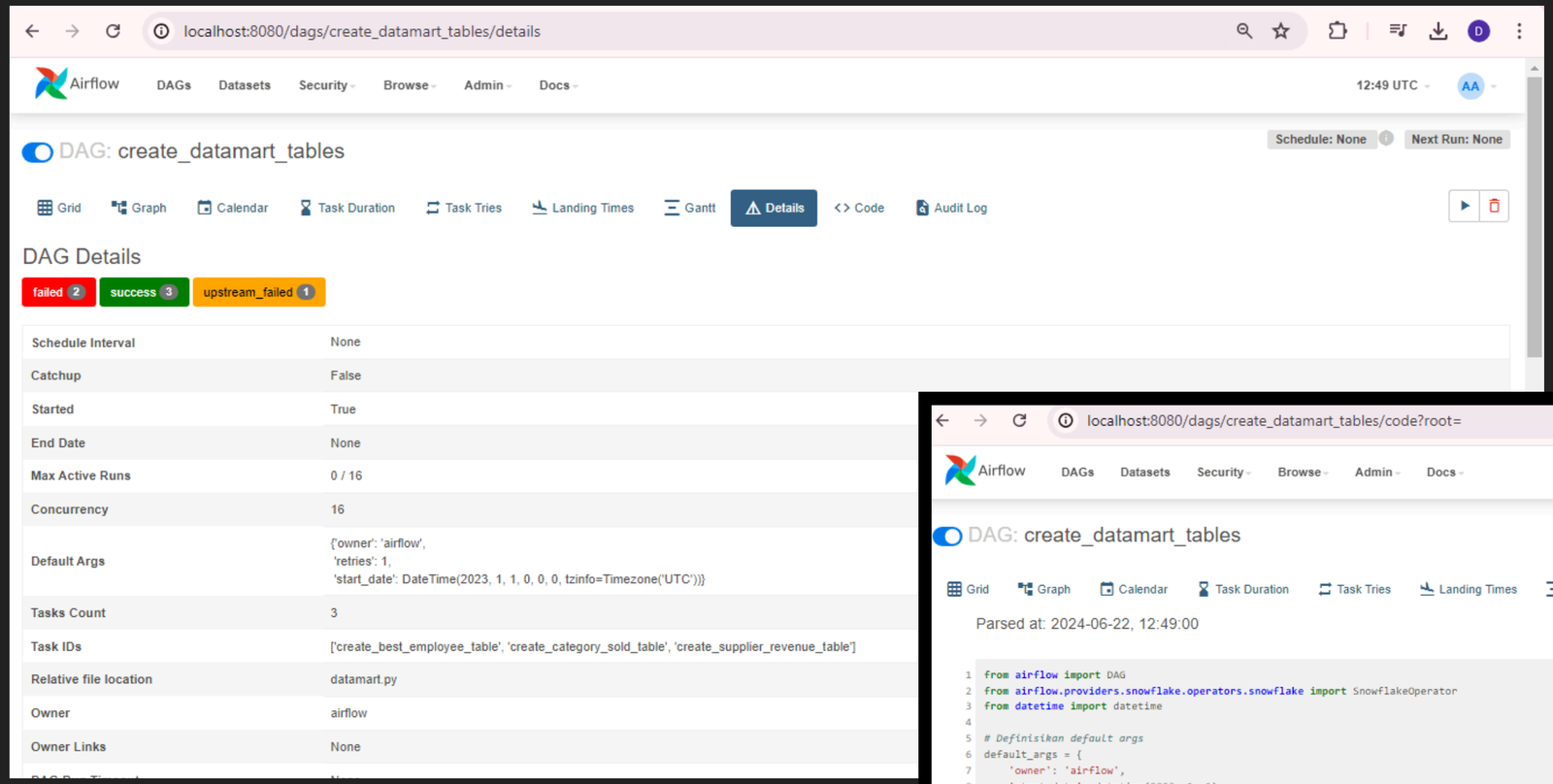


7 Tampilan di Graph Running Datamart

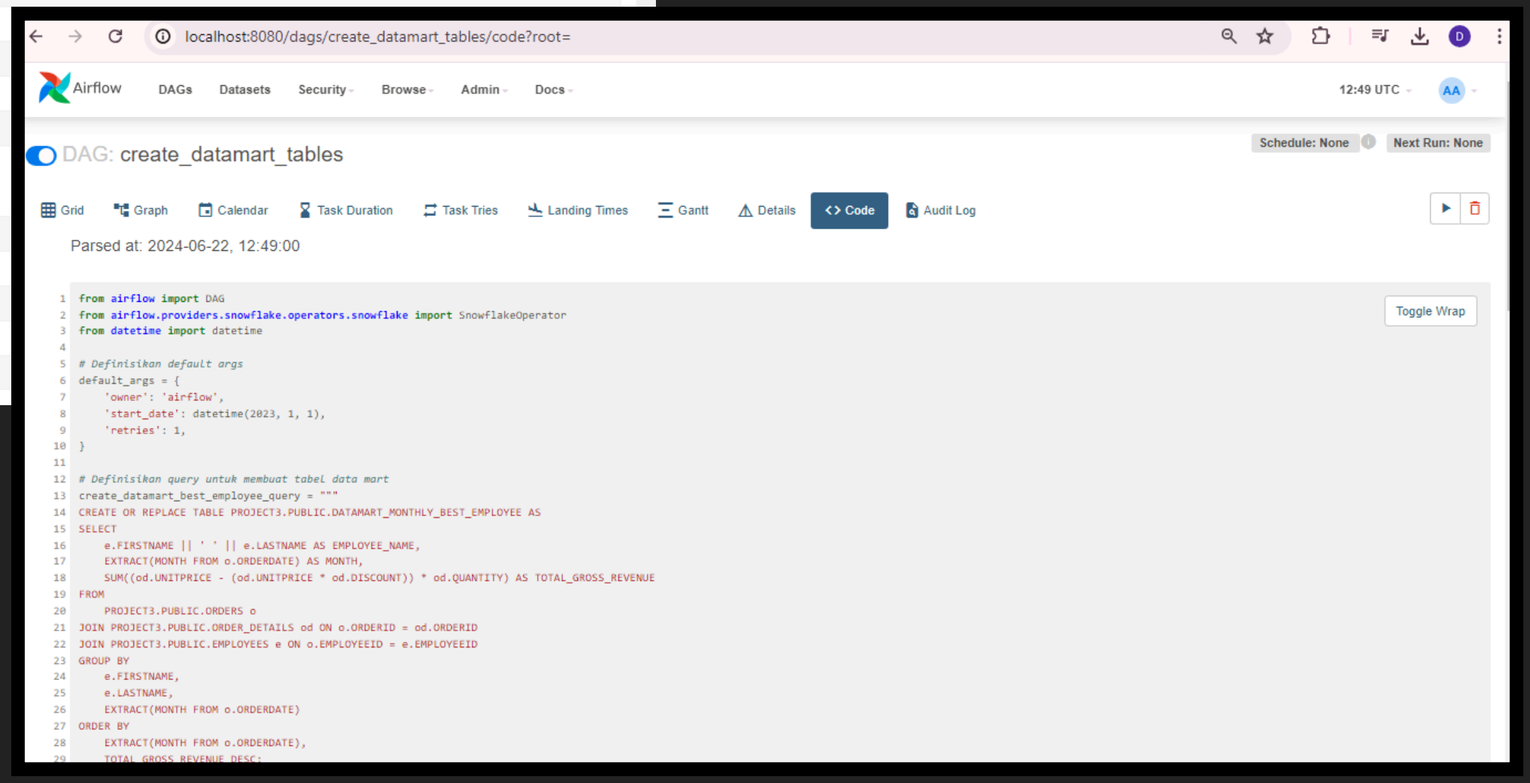


8 Tampilan di Grantt Running Datamart

Project 3 : AIRFLOW



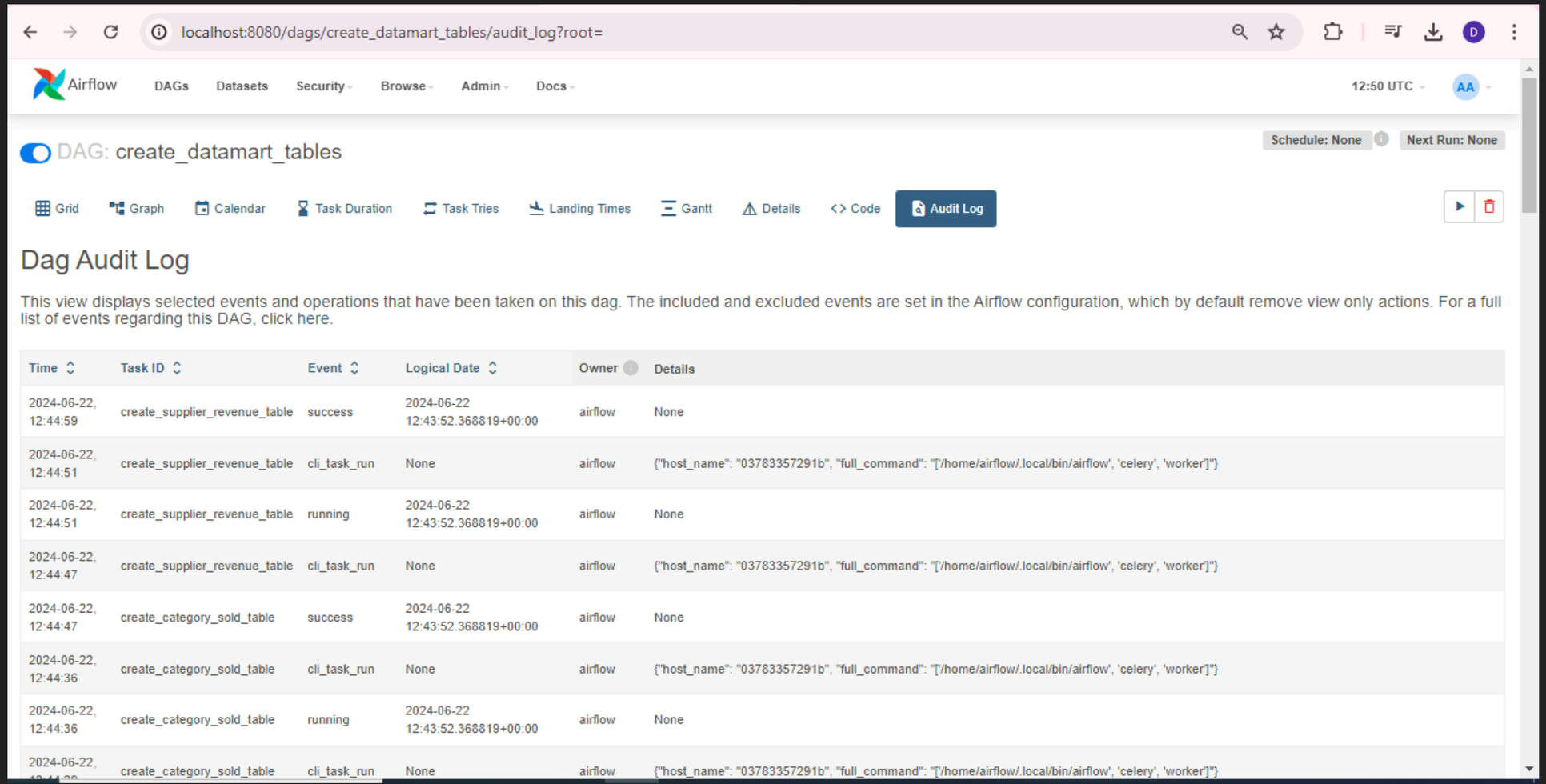
9 Tampilan di Detail Running Datamart



10 Tampilan di Code Running Datamart

Project 3 : AIRFLOW

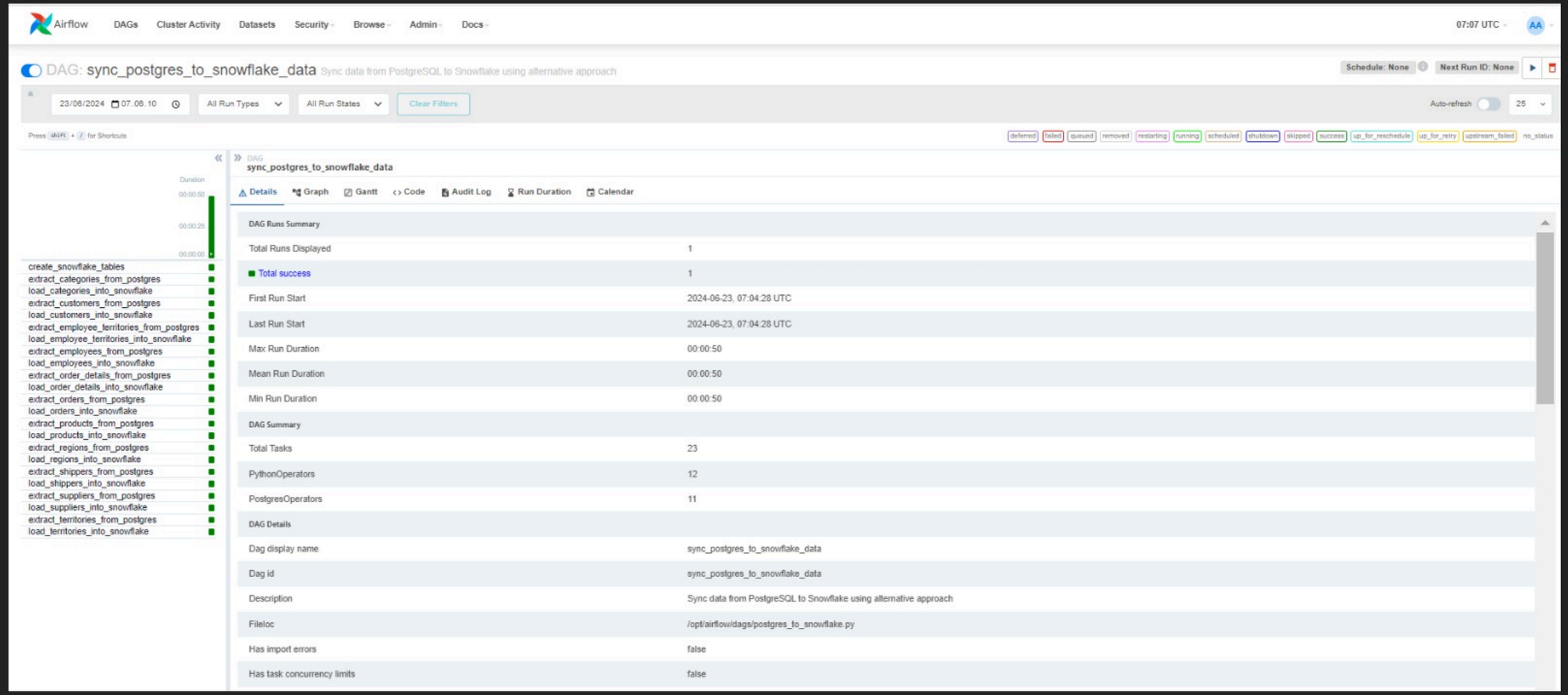
11 Tampilan Log Running Datamart



Time	Task ID	Event	Logical Date	Owner	Details
2024-06-22, 12:44:59	create_supplier_revenue_table	success	2024-06-22 12:43:52.368819+00:00	airflow	None
2024-06-22, 12:44:51	create_supplier_revenue_table	cli_task_run	None	airflow	{"host_name": "03783357291b", "full_command": "['/home/airflow/.local/bin/airflow', 'celery', 'worker']"}
2024-06-22, 12:44:51	create_supplier_revenue_table	running	2024-06-22 12:43:52.368819+00:00	airflow	None
2024-06-22, 12:44:47	create_supplier_revenue_table	cli_task_run	None	airflow	{"host_name": "03783357291b", "full_command": "['/home/airflow/.local/bin/airflow', 'celery', 'worker']"}
2024-06-22, 12:44:47	create_category_sold_table	success	2024-06-22 12:43:52.368819+00:00	airflow	None
2024-06-22, 12:44:36	create_category_sold_table	cli_task_run	None	airflow	{"host_name": "03783357291b", "full_command": "['/home/airflow/.local/bin/airflow', 'celery', 'worker']"}
2024-06-22, 12:44:36	create_category_sold_table	running	2024-06-22 12:43:52.368819+00:00	airflow	None
2024-06-22,	create_category_sold_table	cli_task_run	None	airflow	{"host_name": "03783357291b", "full_command": "['/home/airflow/.local/bin/airflow', 'celery', 'worker']"}

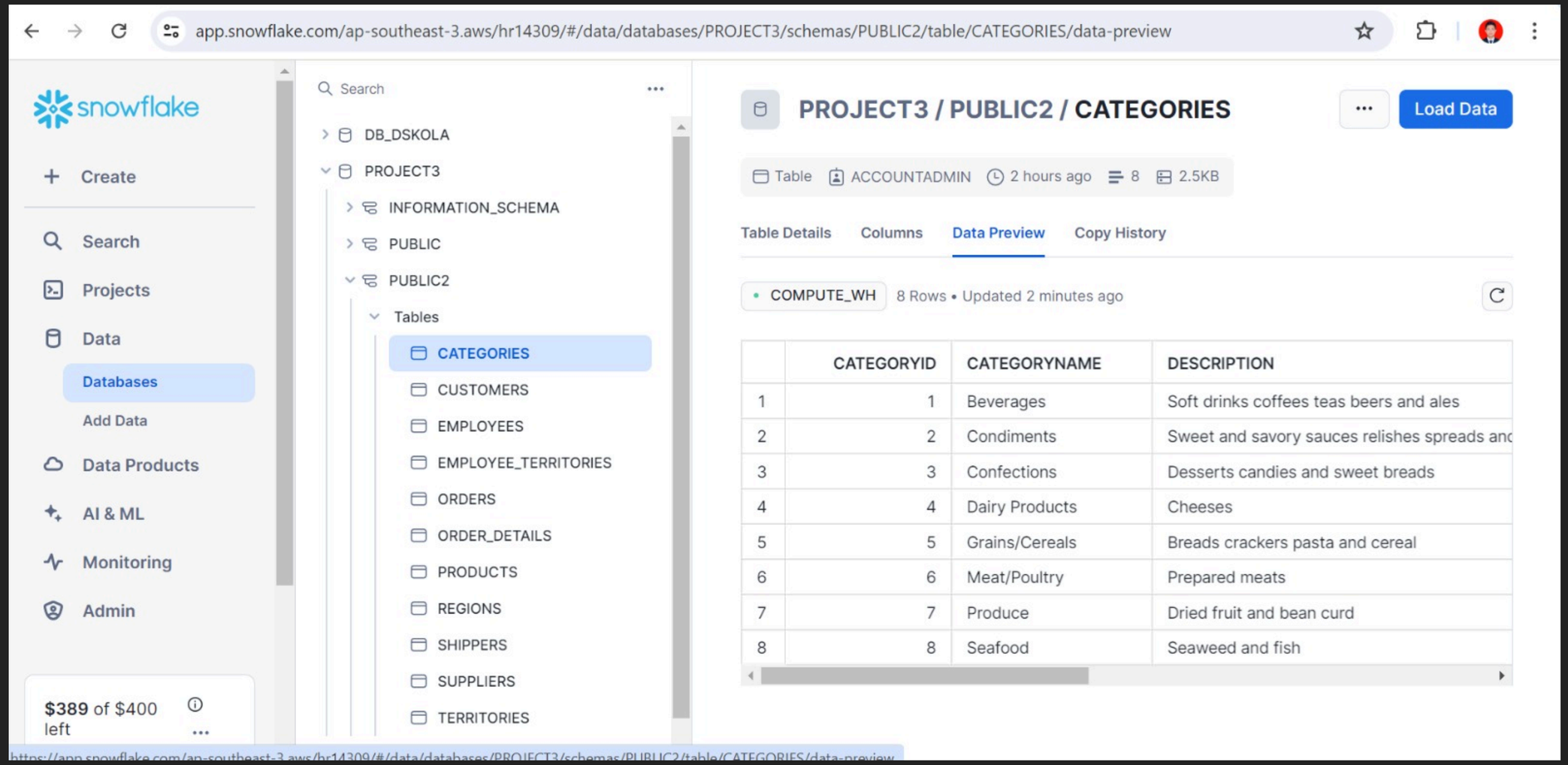
Project 3 : AIRFLOW

12 Tampilan singkronasi dari Postgress ke Snowflake



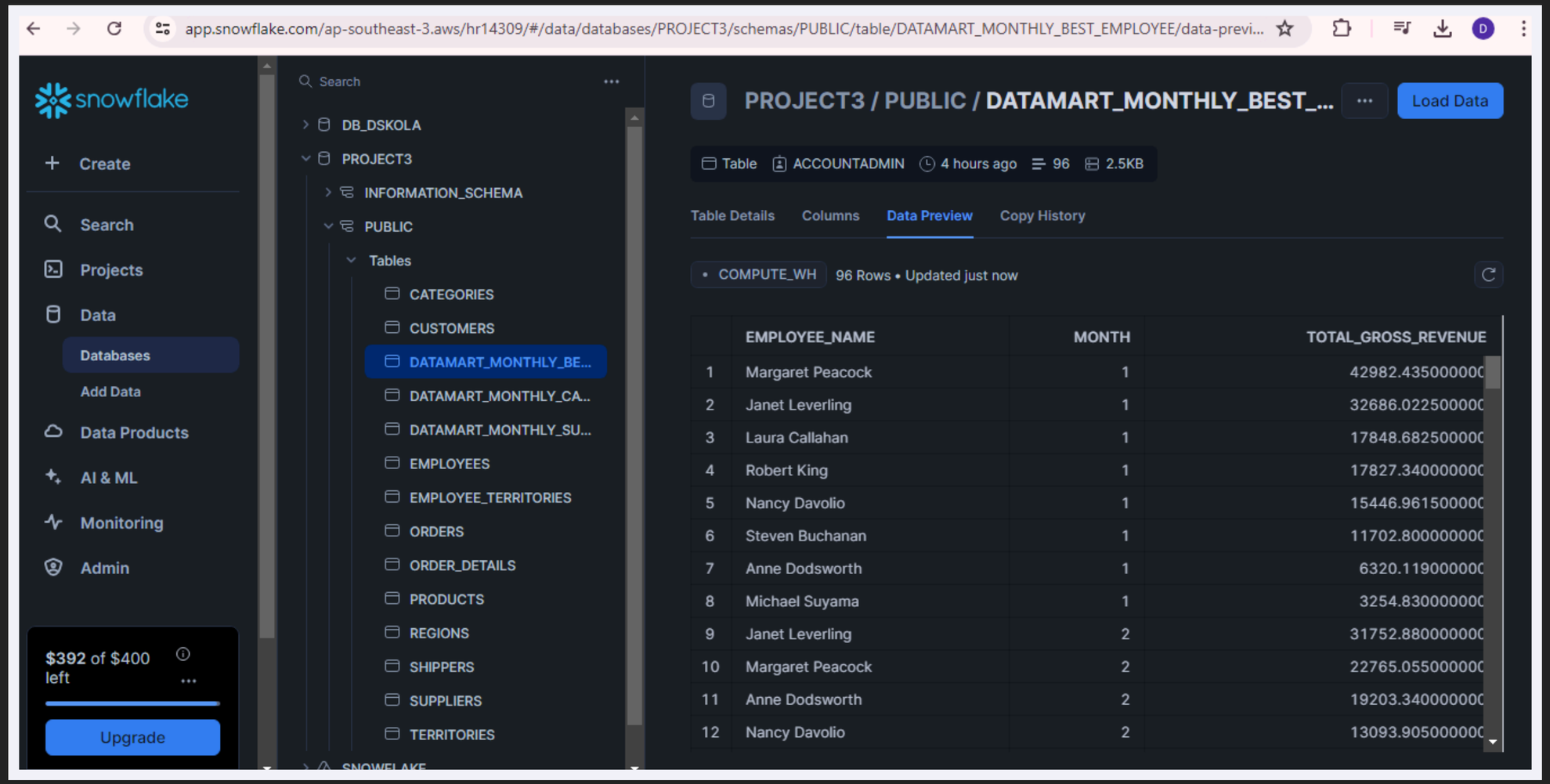
Project 3 : AIRFLOW

13 Tampilan Data Hasil categories di Snowflake



Project 3 : AIRFLOW

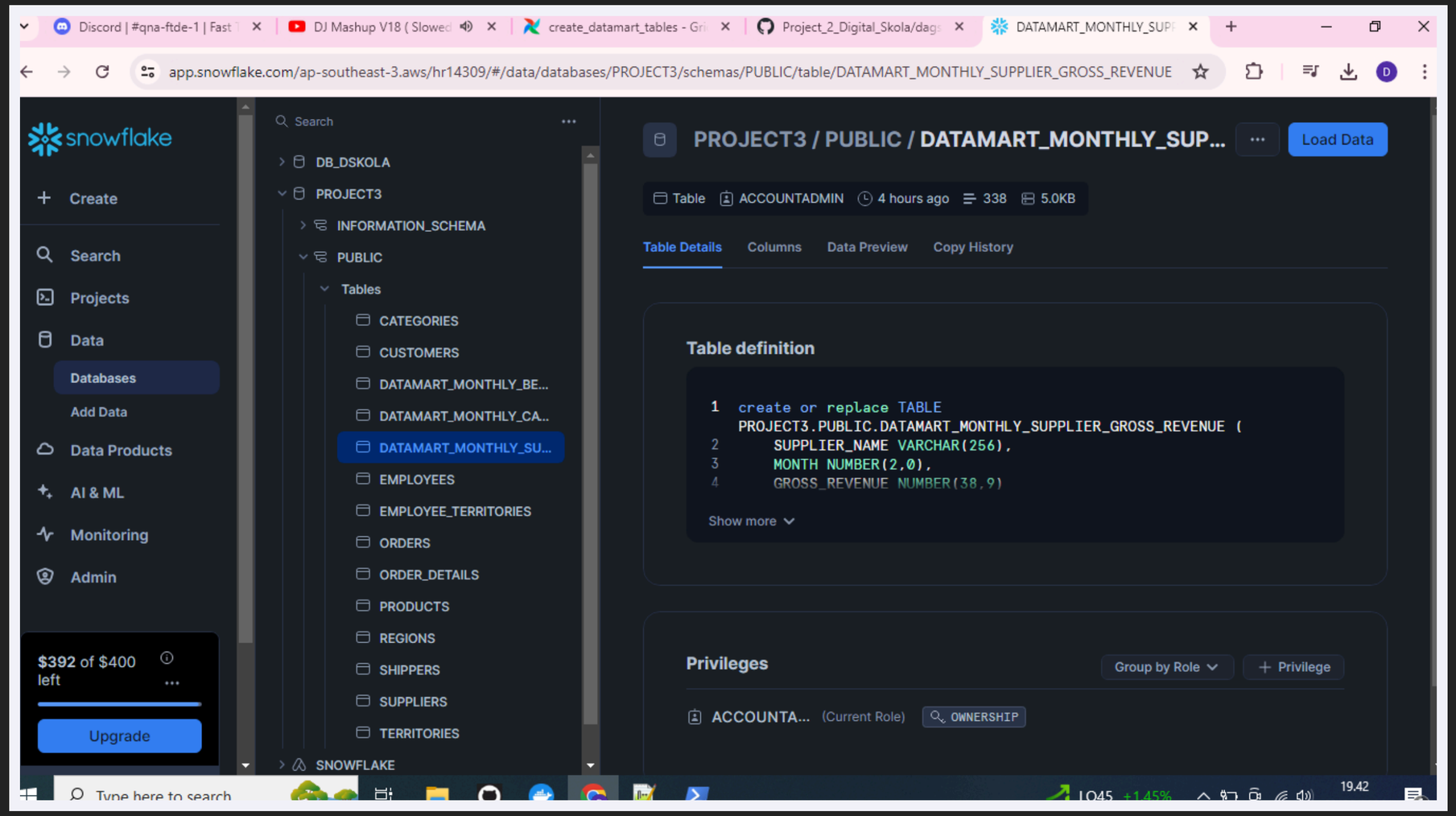
14 Tampilan data Hasil datamart di Snowflake



	EMPLOYEE_NAME	MONTH	TOTAL_GROSS_REVENUE
1	Margaret Peacock	1	42982.435000000
2	Janet Leverling	1	32686.022500000
3	Laura Callahan	1	17848.682500000
4	Robert King	1	17827.340000000
5	Nancy Davolio	1	15446.961500000
6	Steven Buchanan	1	11702.800000000
7	Anne Dodsworth	1	6320.119000000
8	Michael Suyama	1	3254.830000000
9	Janet Leverling	2	31752.880000000
10	Margaret Peacock	2	22765.055000000
11	Anne Dodsworth	2	19203.340000000
12	Nancy Davolio	2	13093.905000000

Project 3 : Result

15 Tampilan table Hasil datamart di Snowflake



Project 4 : Kafka

Objective Project

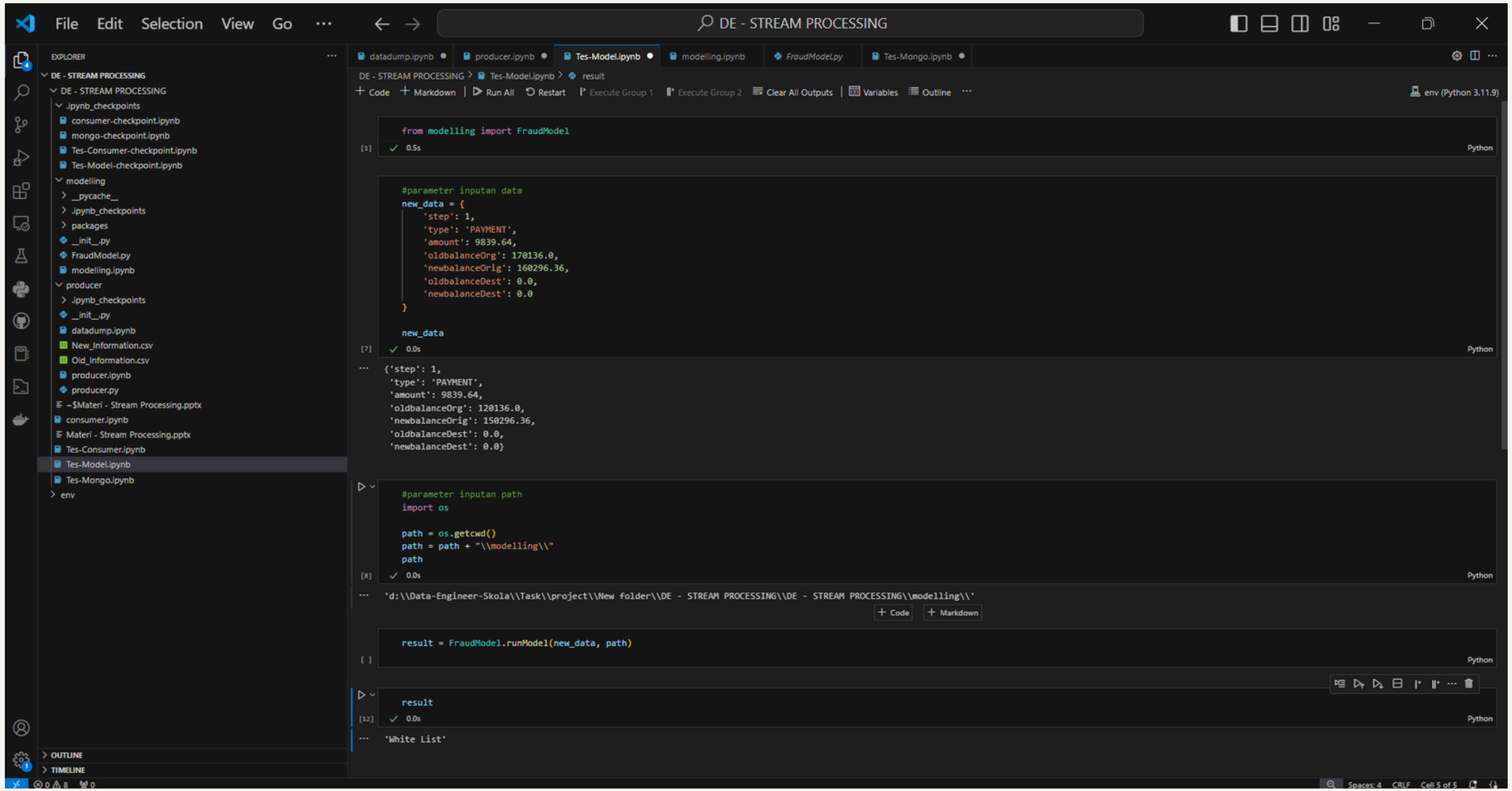
Streaming processing

- Check Module ML
- Masukan data ke dlm db
- Load data new ke kafka, time.sleep(10)
- Mengambil data dari kafka, procedure dan consumer dijalankan bersamaan
- Get data dalam database
- Join data kafka Dan DB postgres
- Panggil func Predict Detect
- Insert di mongodb
- Save data CSV

Tools yang digunakan dalam project

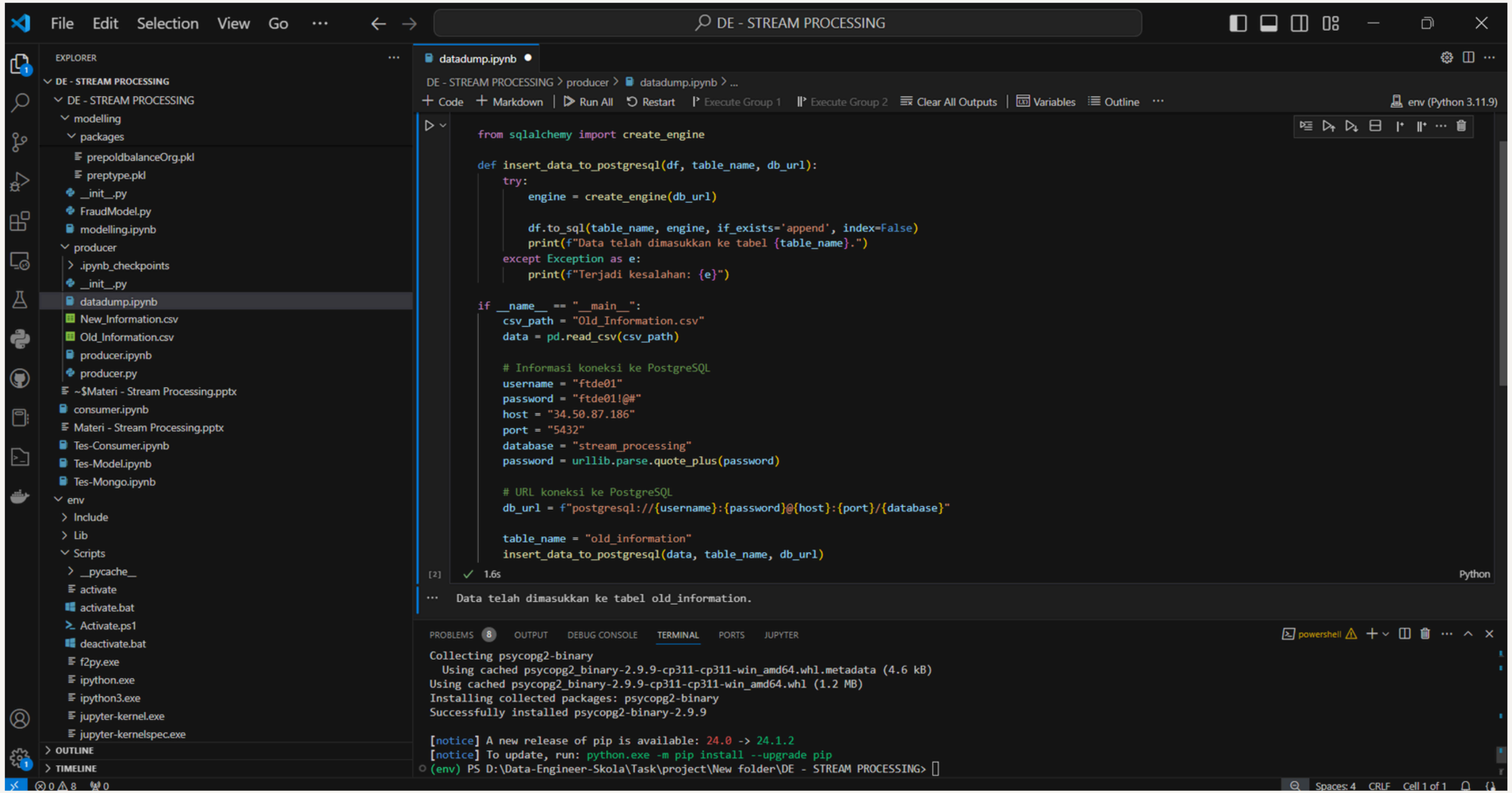
Project 4 – Kafka

01 Modeling/Fraudmodul.py --> Check Module ML



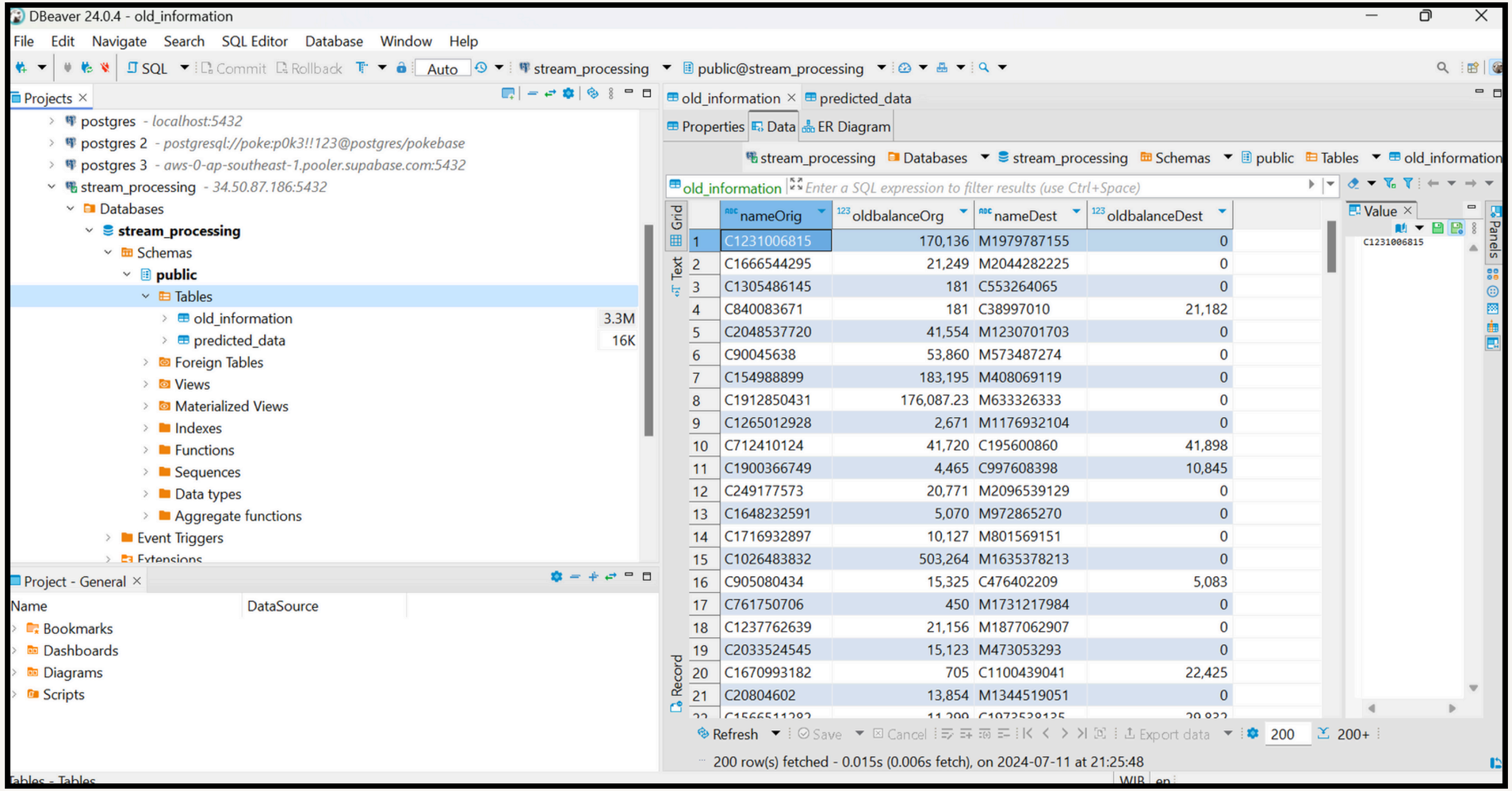
Project 4 – Kafka

02 Producer/datadump --> masukan data ke dlm db



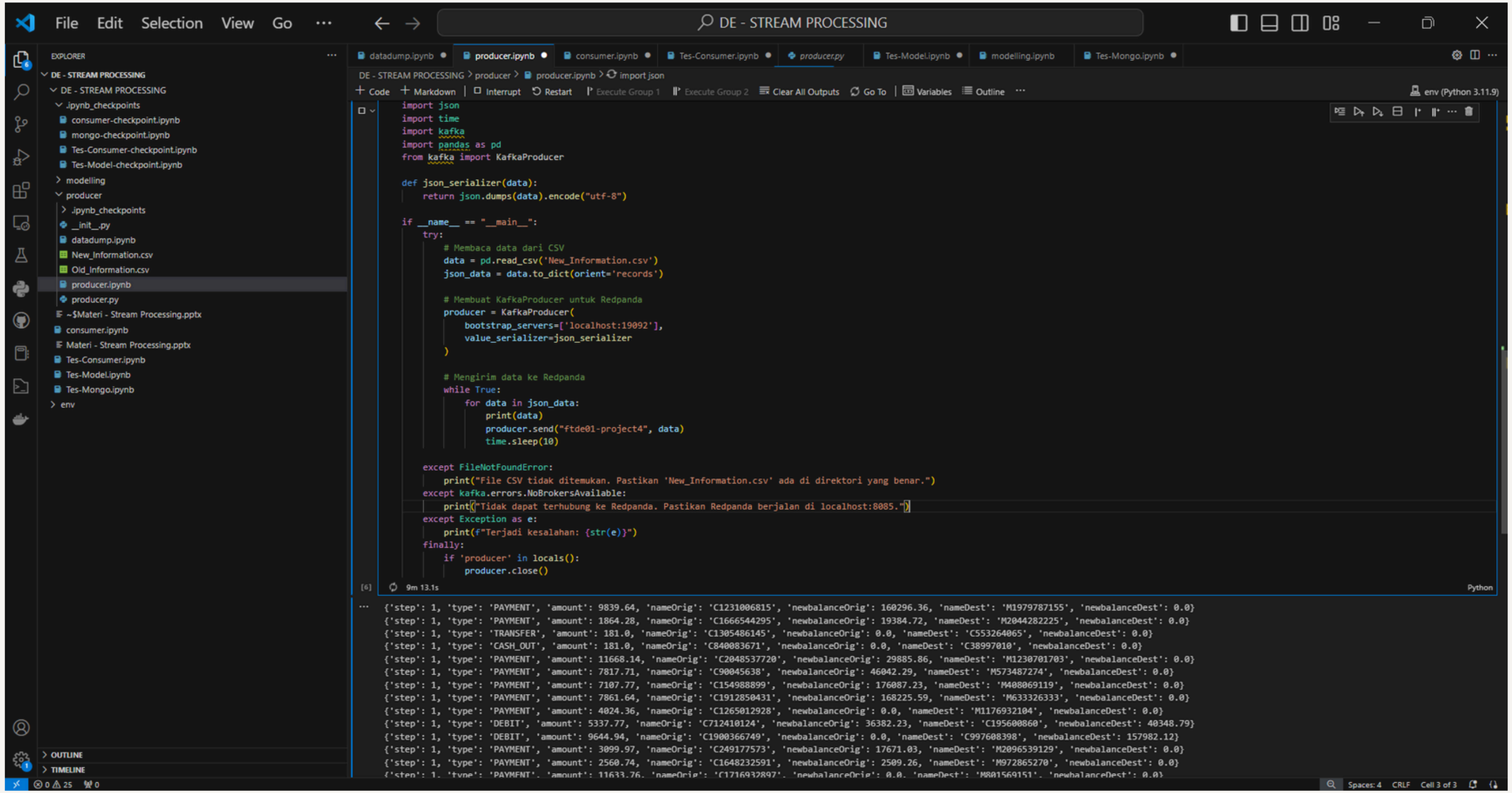
Project 4 – Kafka

03 Data yang sudah masuk di DB



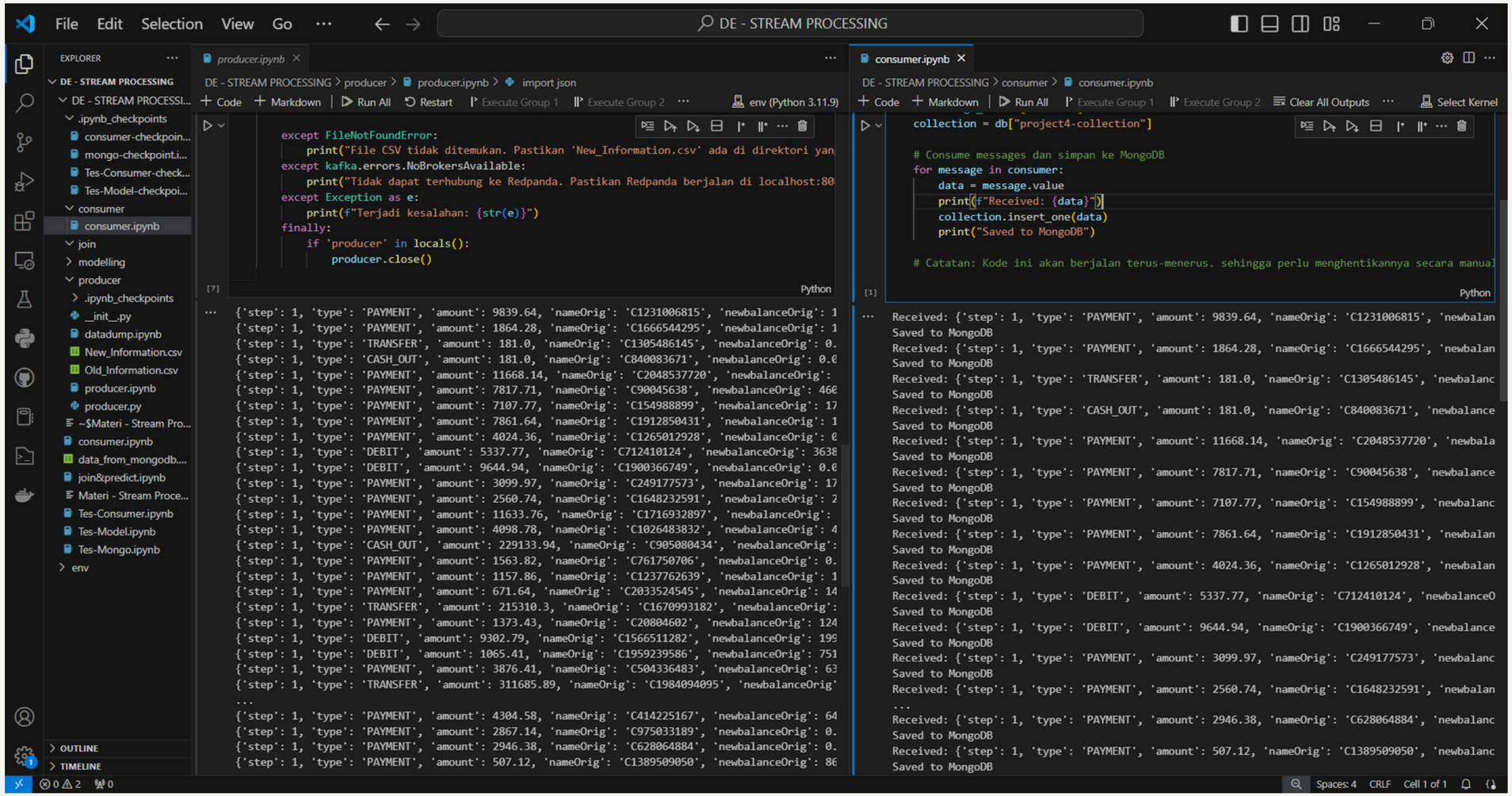
Project 4 – Kafka

04 Producer/producer.ipynb --> lempar data new ke kafka, time.sleep(10)



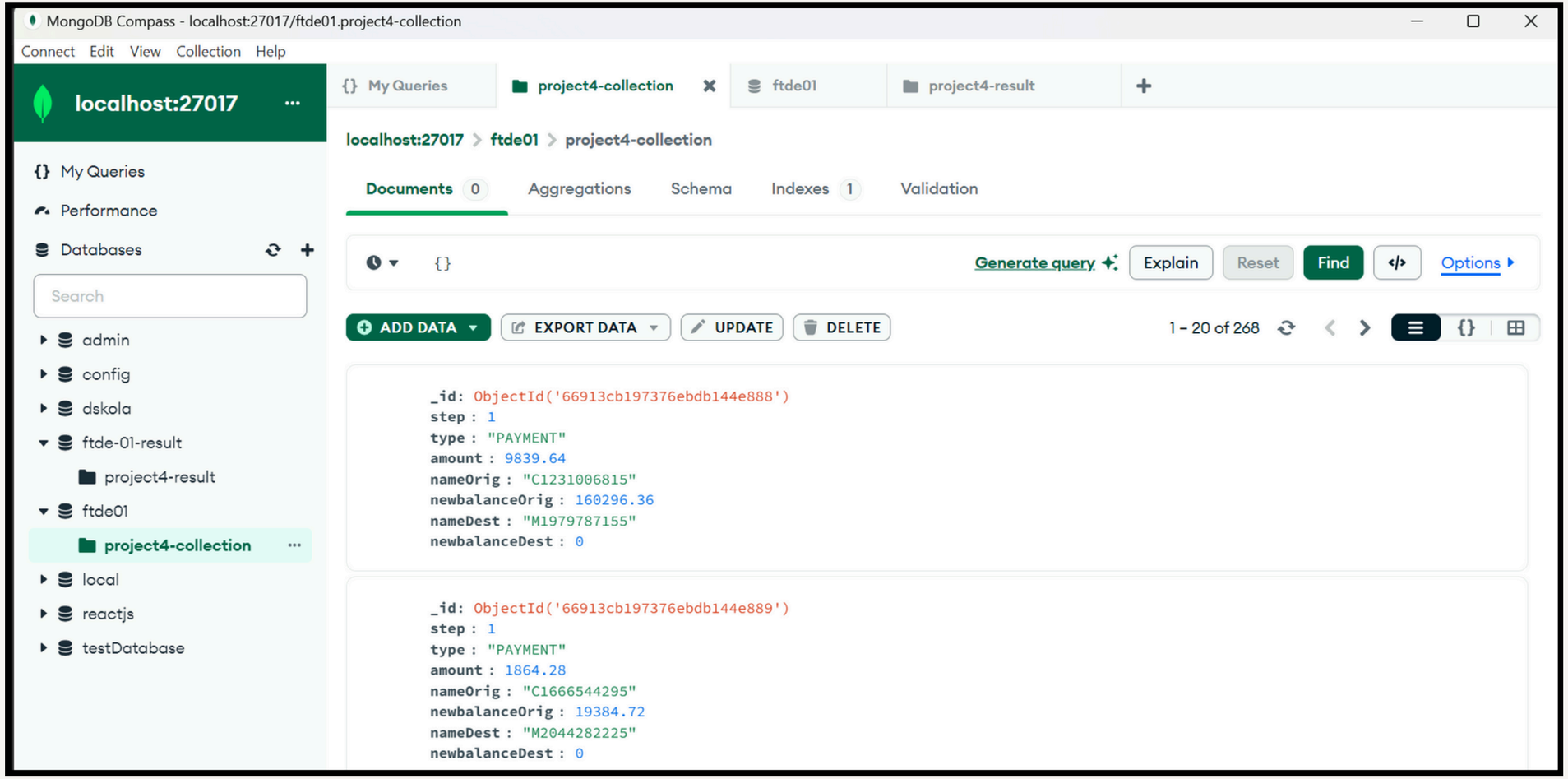
Project 4 – Kafka

05 producer/producer.ipynb & consumer/consumer.ipynb--> Mengambil data dari kafka, producer dan consumer dijalankan bersamaan serta menyimpan consume dari kafka ke mongodb



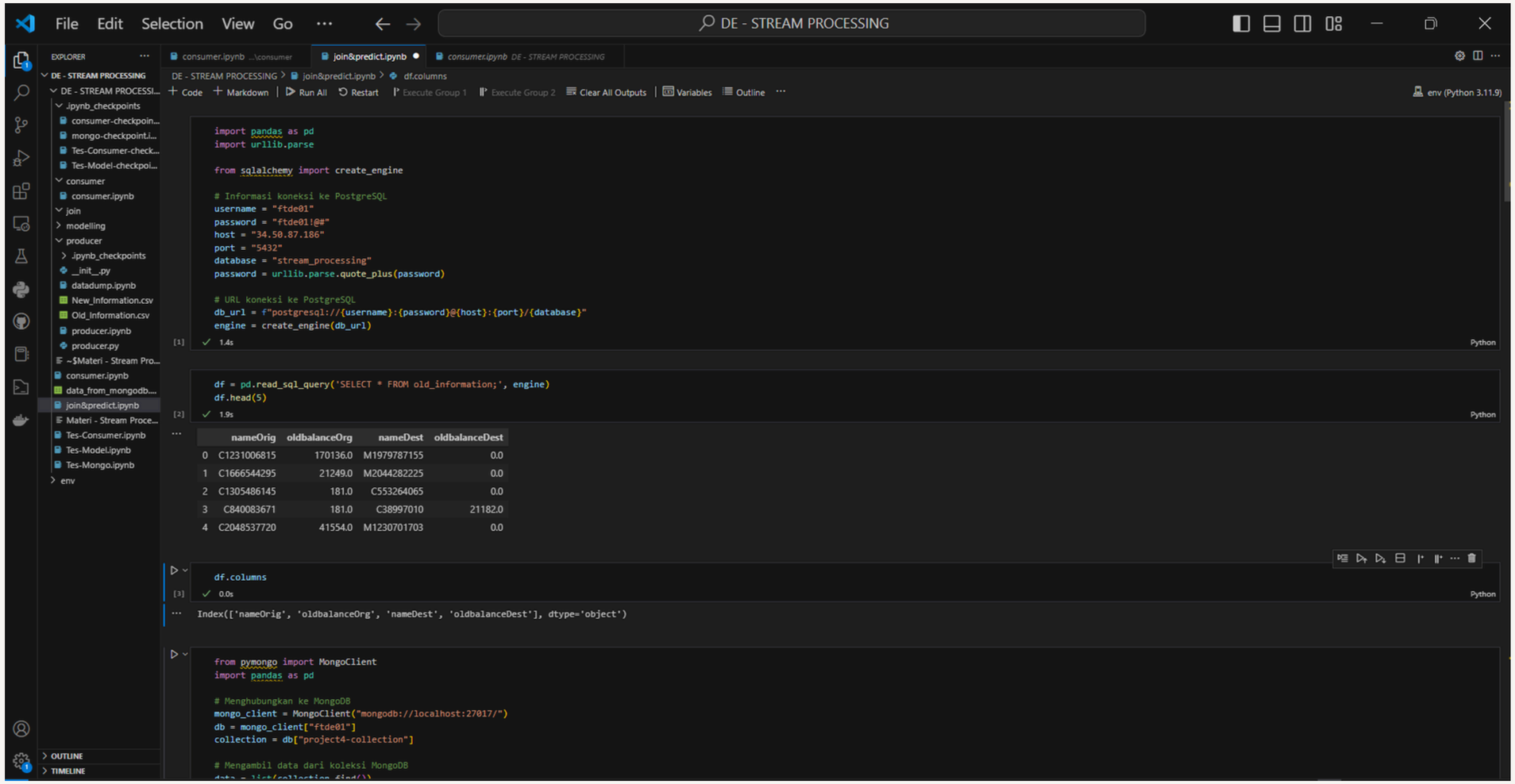
Project 4 – Kafka

06 Consume messages dan simpan ke MongoDB



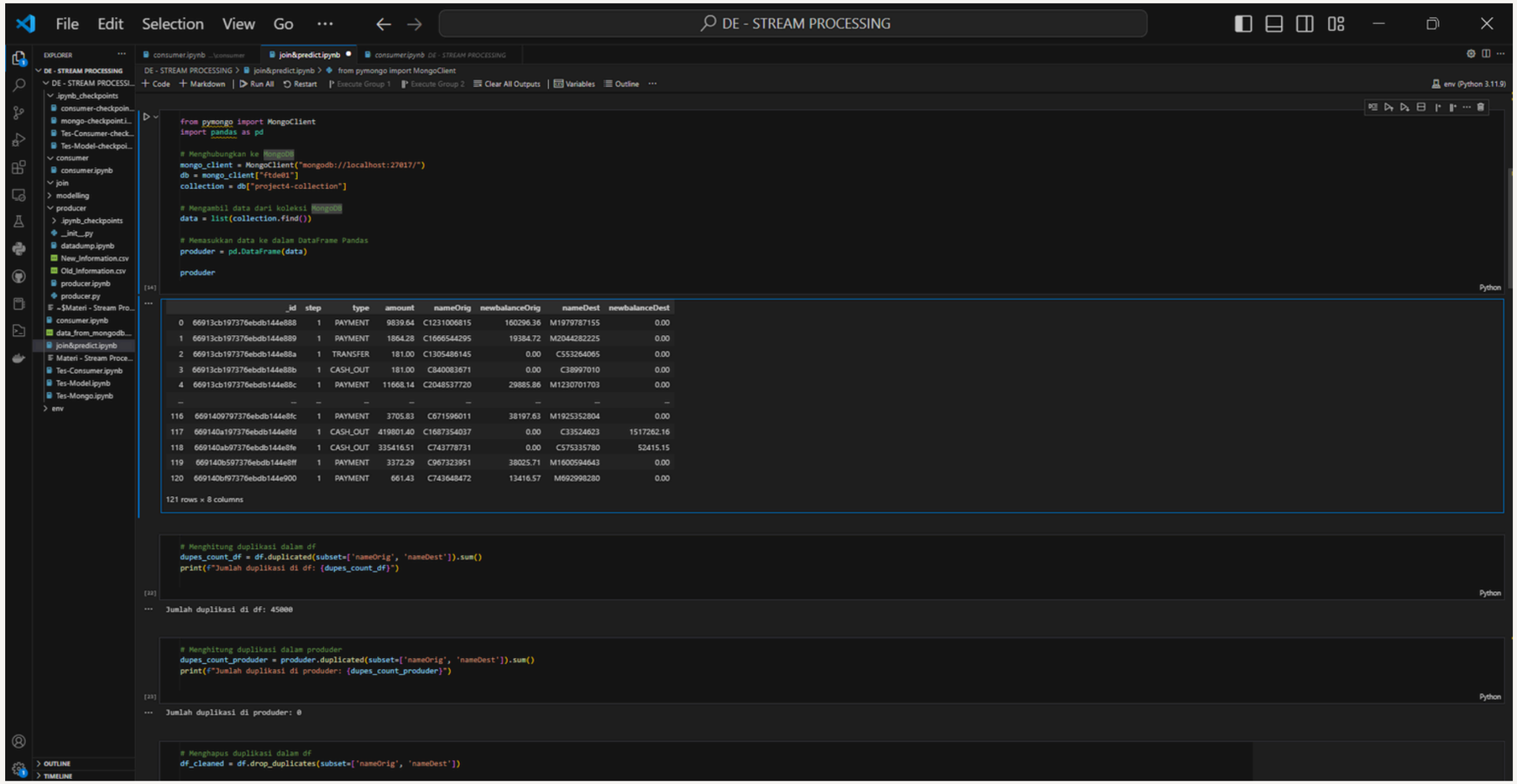
Project 4 – Kafka

07 Join&predict.ipynb --> ambil data dalam database postgre



Project 4 – Kafka

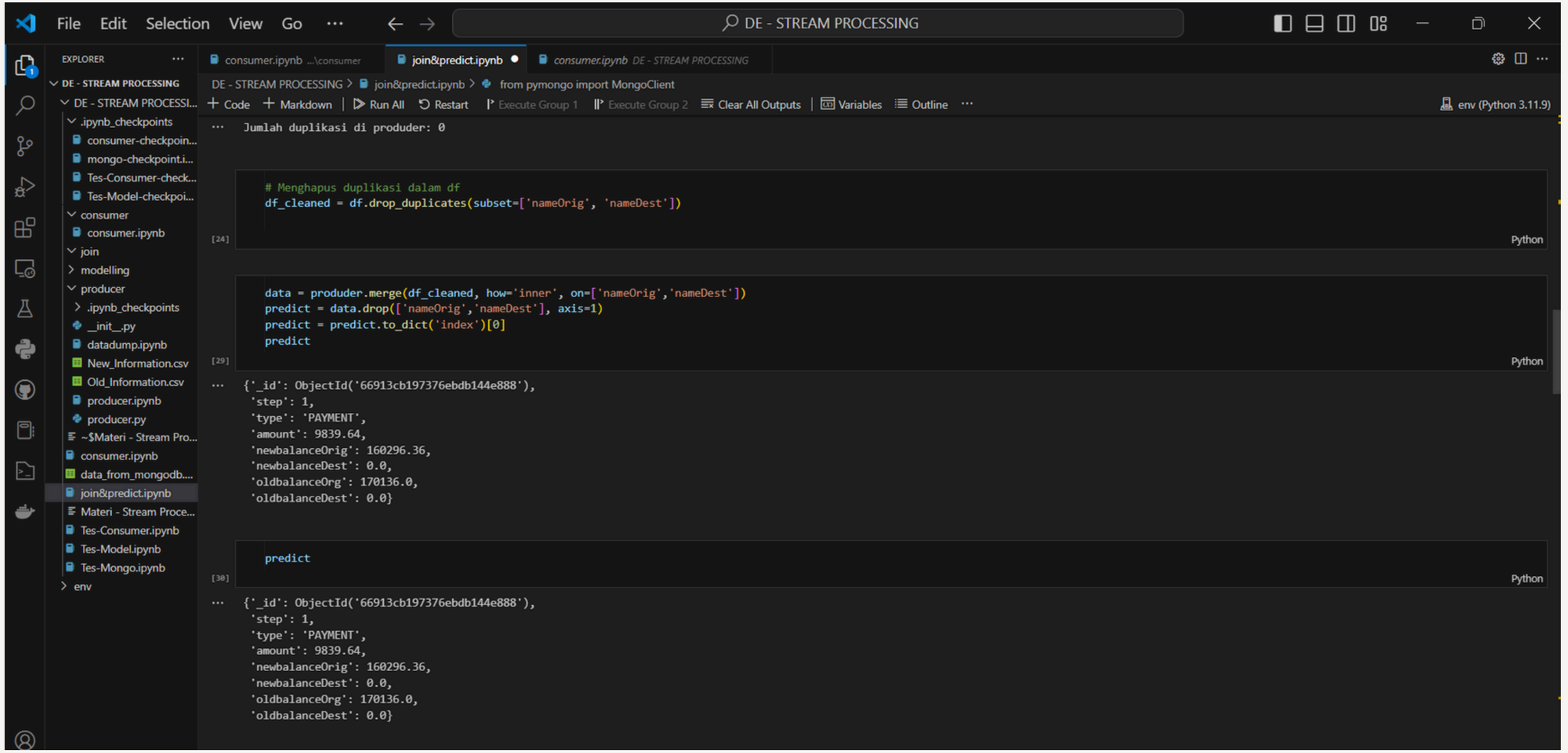
08 Join&predict.ipynb –-> ambil data dari consume kafka yang telah disimpan ke mongodb



*Cek duplicate data, ternyata database postgre terdapat data duplikat

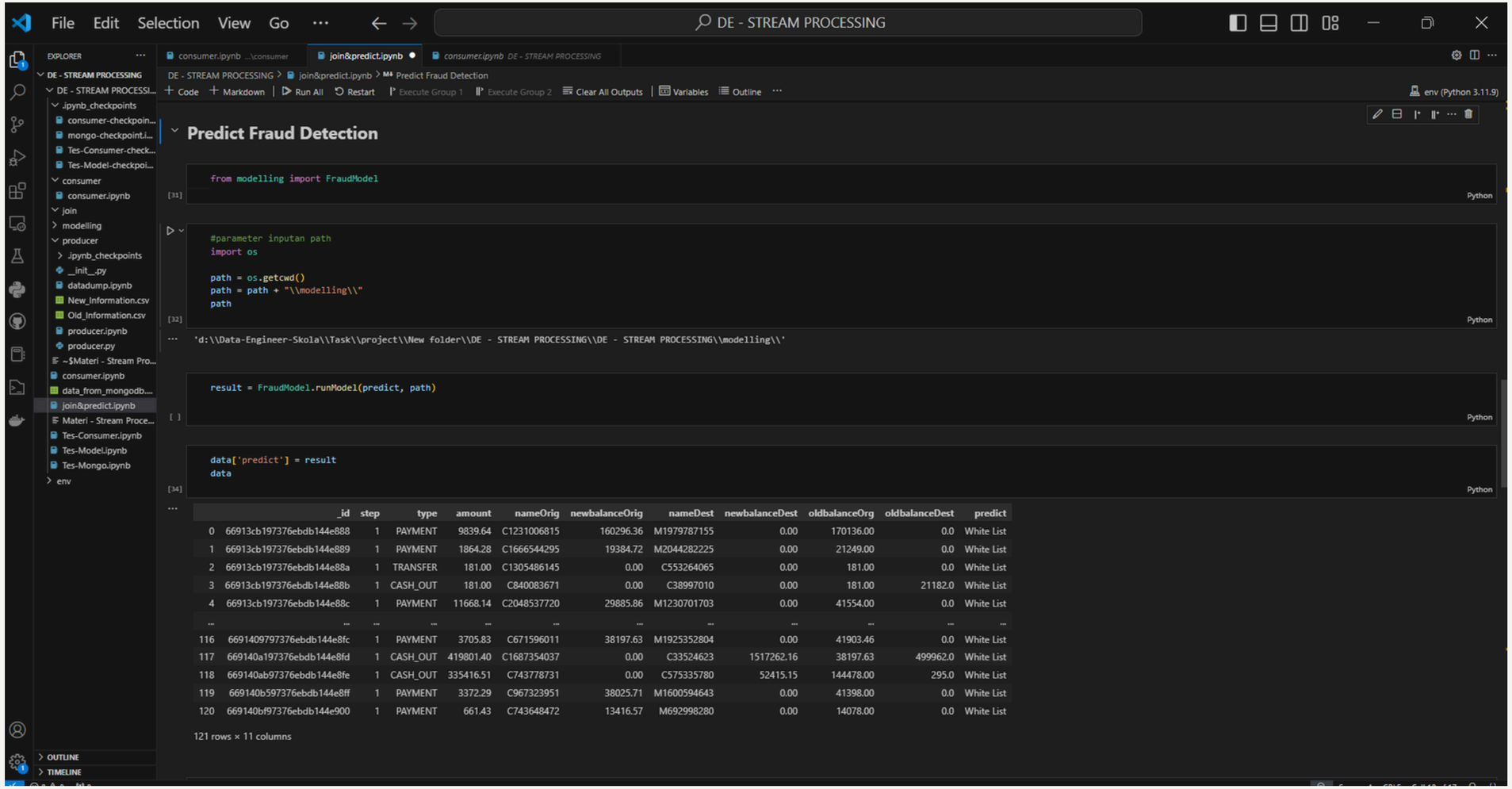
Project 4 – Kafka

09 Join&predict.ipynb --> Join data dari kakfa Dan DB postgres



Project 4 – Kafka

10 join&predict.ipynb -->panggil func Predict Detect

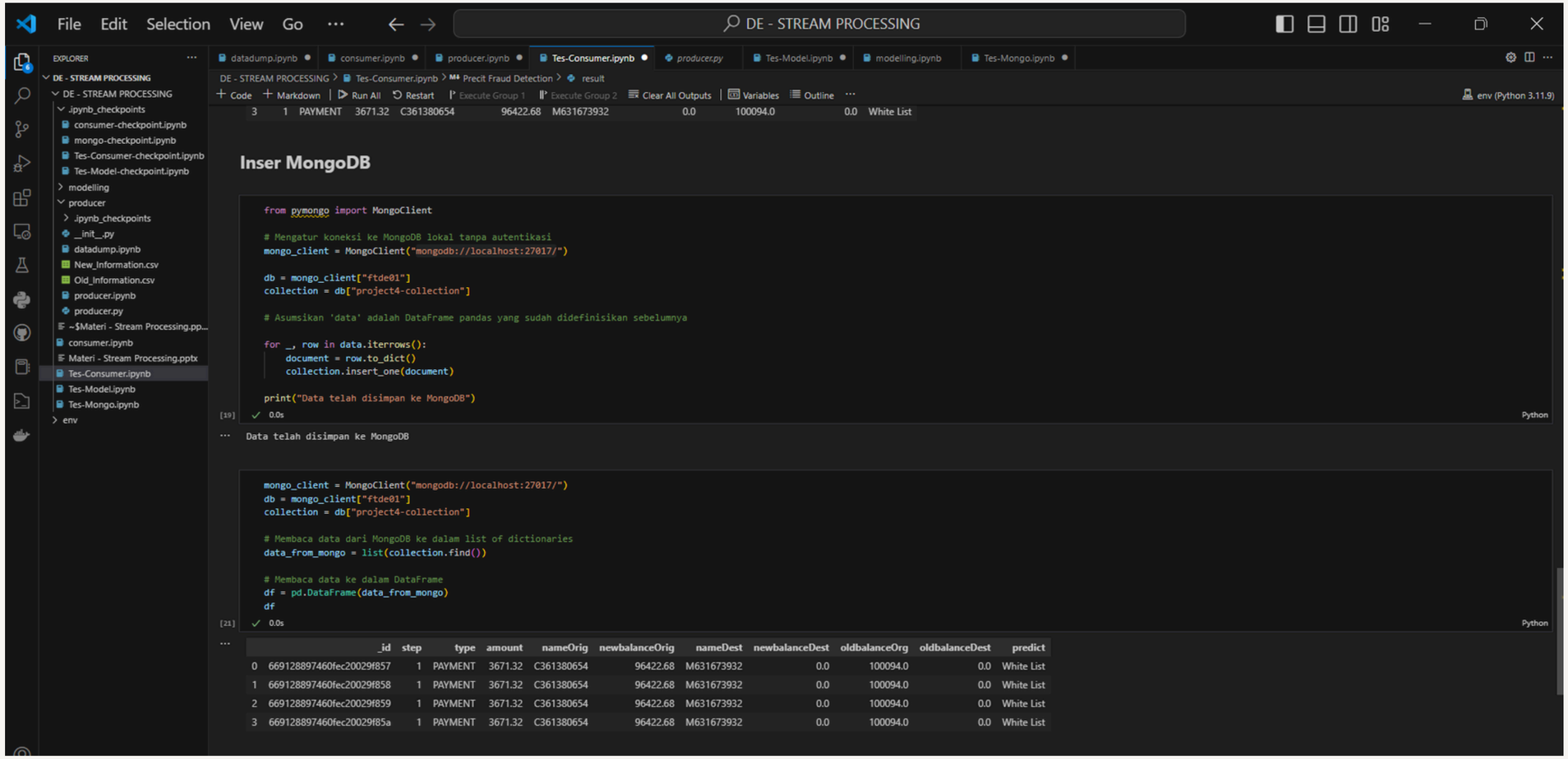


	_id	step	type	amount	nameOrig	newbalanceOrig	nameDest	newbalanceDest	oldbalanceOrg	oldbalanceDest	predict
0	66913cb197376ebdb144e888	1	PAYMENT	9839.64	C1231006815	160296.36	M1979787155	0.00	170136.00	0.0	White List
1	66913cb197376ebdb144e889	1	PAYMENT	1864.28	C1666544295	19384.72	M2044282225	0.00	21249.00	0.0	White List
2	66913cb197376ebdb144e88a	1	TRANSFER	181.00	C1305486145	0.00	C553264065	0.00	181.00	0.0	White List
3	66913cb197376ebdb144e88b	1	CASH_OUT	181.00	C840083671	0.00	C38997010	0.00	181.00	21182.0	White List
4	66913cb197376ebdb144e88c	1	PAYMENT	11668.14	C2048537720	29885.86	M1230701703	0.00	41554.00	0.0	White List
...	...	...	...	...	...	...	...	...	...	...	...
116	6691409797376ebdb144e8fc	1	PAYMENT	3705.83	C671596011	38197.63	M1925352804	0.00	41903.46	0.0	White List
117	669140a197376ebdb144e8fd	1	CASH_OUT	419801.40	C1687354037	0.00	C33524623	1517262.16	38197.63	499962.0	White List
118	669140ab97376ebdb144e8fe	1	CASH_OUT	335416.51	C743778731	0.00	C575335780	52415.15	144478.00	295.0	White List
119	669140b597376ebdb144e8ff	1	PAYMENT	3372.29	C967323951	38025.71	M1600594643	0.00	41398.00	0.0	White List
120	669140bf97376ebdb144e900	1	PAYMENT	661.43	C743648472	13416.57	M692998280	0.00	14078.00	0.0	White List

121 rows × 11 columns

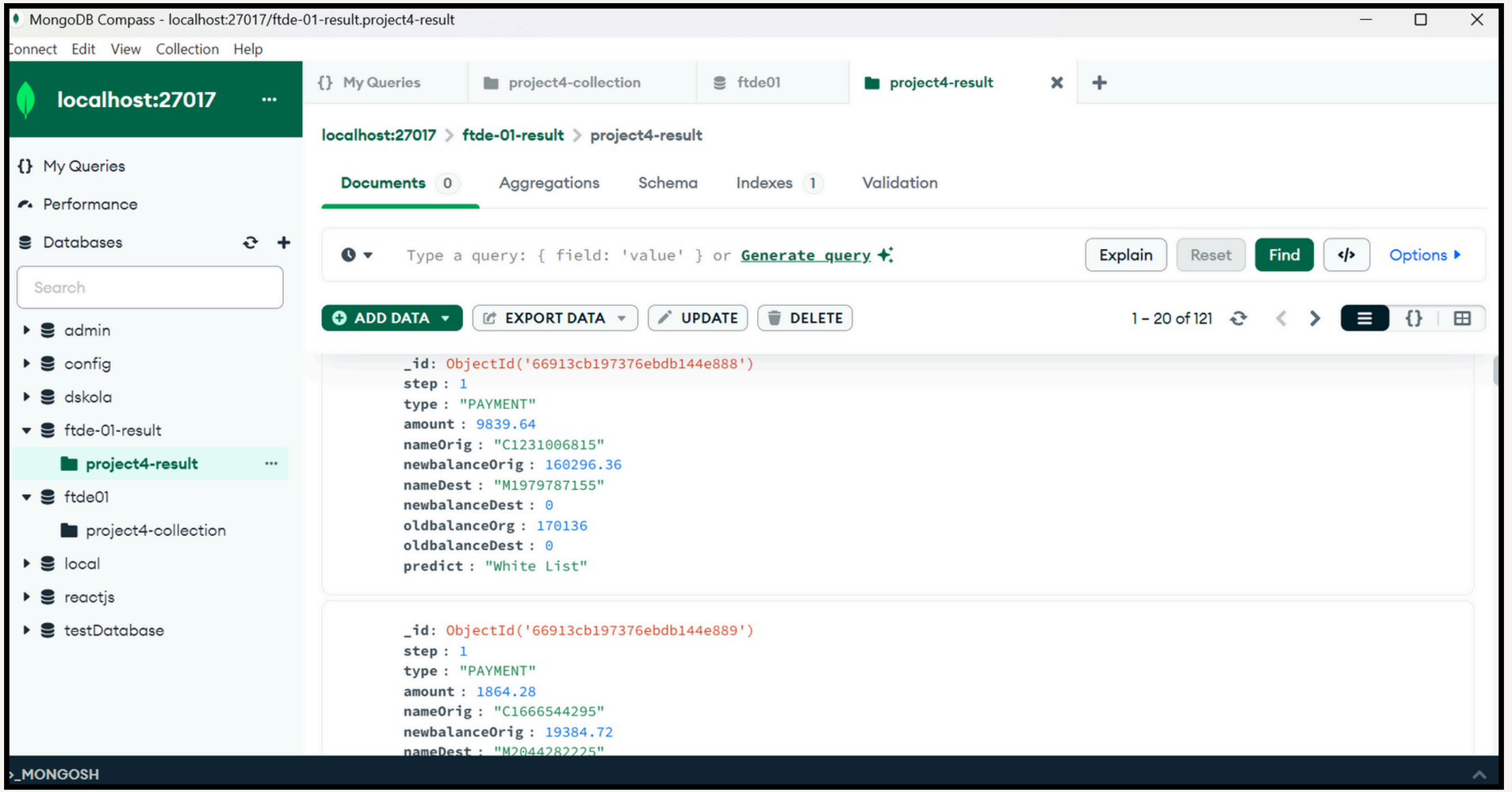
Project 4 – Kafka

11 Join&predict.ipynb --> insert hasil prediksi di mongodb



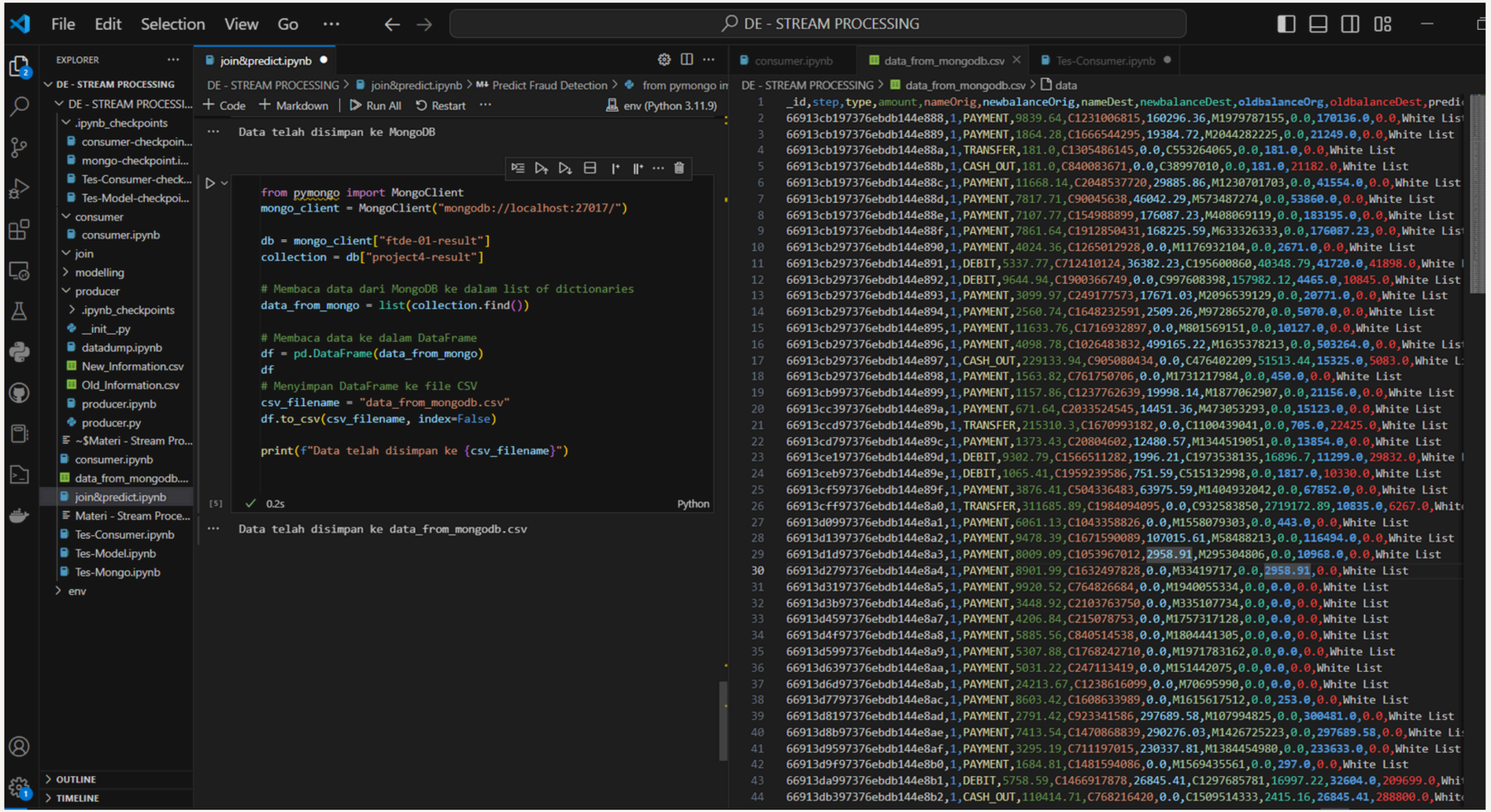
Project 4 – Kafka

12 Tampilan pada Mongo DB



Project 4 – Result

13 Join&predict.ipynb Save data CSV



Sekian & Terima Kasih

Kelompok 13

fantastic 4