

PROJECT 2 AIRFLOW DAN SNOWFLAKE

FAST TRACK DATA ENGINEER

DIGITAL SKOLA

Nama Team: Fantastic 4

Nama Anggota:

1. Donna Setiawan
2. Muhammad Ramli
3. Ozza Dinanta
4. Satria Said
5. Ginta Khoirunisa

Requirment/Task:

1. Instalasi Airflow di Docker (plugin **Snowflake**)
2. Membuat Database dan Insert Data dari Project 2 (Postgress), kemudian masukan ke Snowflake
3. Membuat Code DAG untuk tugas menjalankan query Data Mart dan dan menambahkan koneksi dari dari Snowflake
4. Merekam Hasil Running Airflow secara keseluruhan (Cara Kerja dan Hasil running)
5. Membuat Dokumentasi hasil yang berada Snowflake

Link Github :

https://github.com/dannasetia11/Project_2_Digital_Skola

Video Rekaman : **terdapat di Github**

Lampiran Setiap Task

Tampilan running airflow di docker (Windows Shell) :

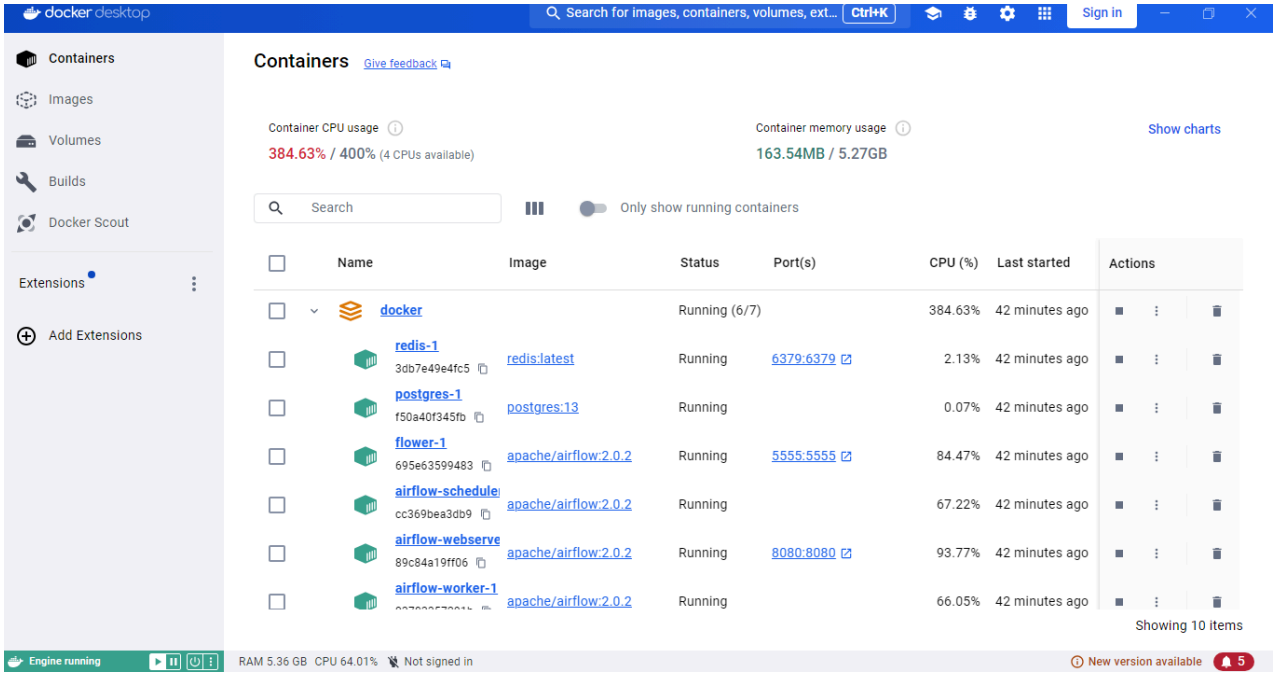
```
Windows PowerShell
--tls               Use TLS; implied by --tlsverify
--tlscacert string  Trust certs signed only by this CA (default
                    "C:\\Users\\PC\\.docker\\ca.pem")
--tlscert string    Path to TLS certificate file (default
                    "C:\\Users\\PC\\.docker\\cert.pem")
--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.

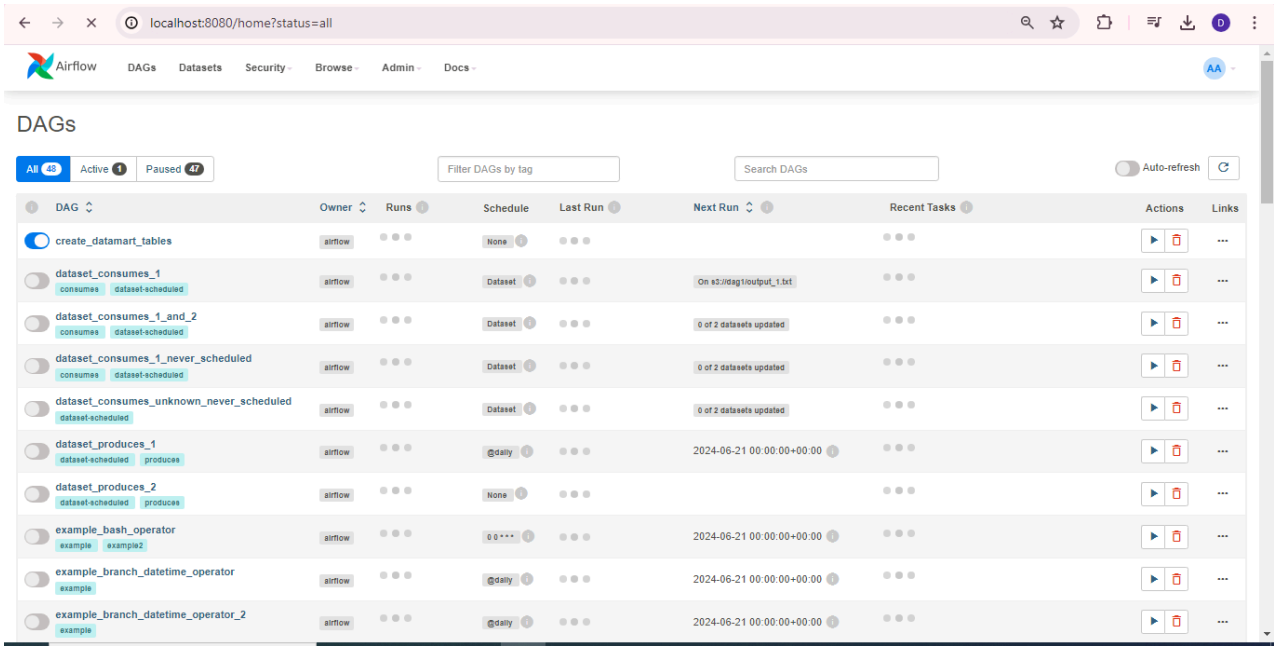
For more help on how to use Docker, head to https://docs.docker.com/go/guides/

docker> 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>
```

Tampilan Docker Running UI :



Tampilan halaman airflow :



Tampilan Koneksi ke snowflake :

localhost:8080/connection/list/

Airflow

DAGs

Datasets

Security

Browse

Admin

Docs

13:06 UTC

List Connection

Search

+

Actions

←

Record Count: 1

	Conn Id	Conn Type	Description	Host	Port	Is Encrypted	Is Extra Encrypted
<input type="checkbox"/>	snowflake	snowflake				False	False

Version: v2.5.0

Git Version: .release:2.5.0+fa2bec042995004f45b914dd1d66b466ccced410

Tampilan saat runing DAG datamart :

localhost:8080/home?status=active

Airflow

DAGs

Datasets

Security

Browse

Admin

Docs

12:44 UTC

DAGs

All46Active1Paused47

Filter DAGs by tag

Search DAGs

Auto-refresh

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
<input checked="" type="checkbox"/> create_datamart_tables	airflow	<div><div></div><div>1</div><div>1</div></div>	None	2024-06-22, 12:43:52		<div><div>2</div><div></div><div>1</div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div>	<div><div>▶</div><div>🔍</div><div>⋮</div></div>	

◀

<

1

>

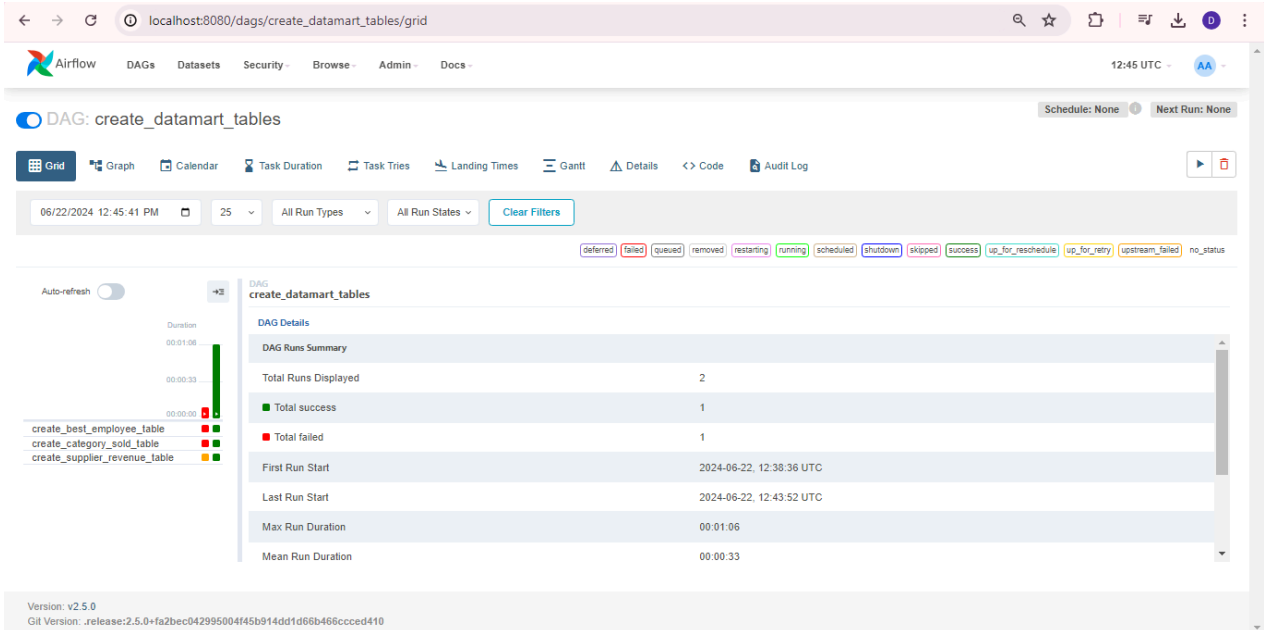
▶

Showing 1-1 of 1 DAGs

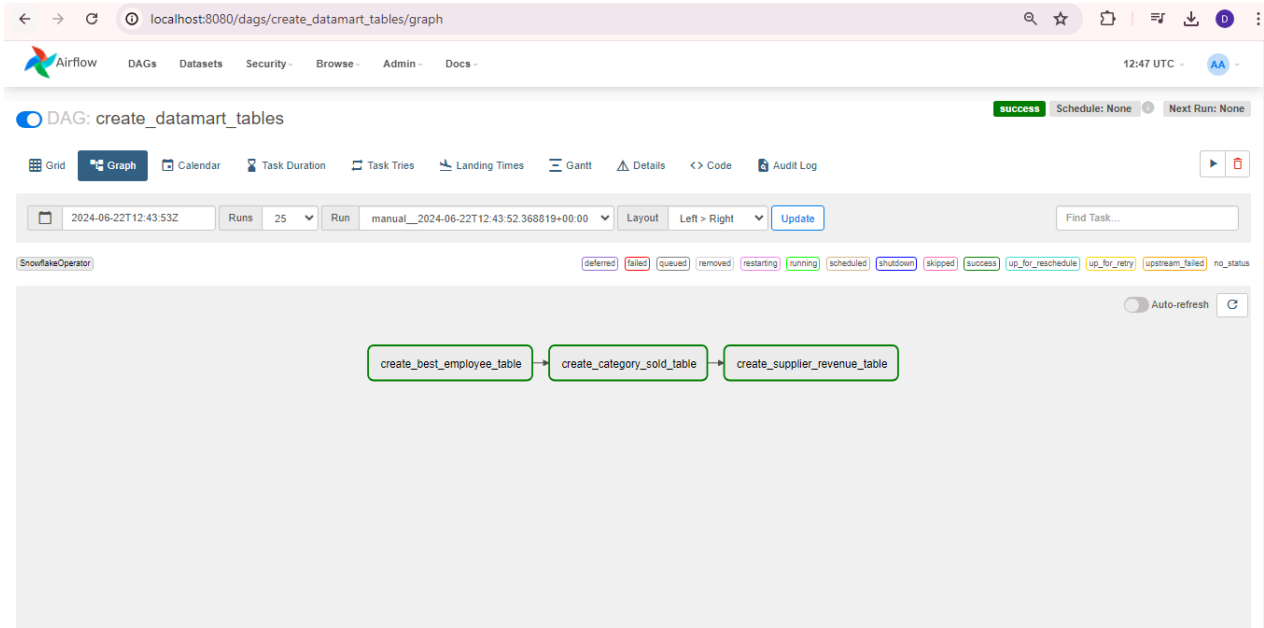
Version: v2.5.0

Git Version: .release:2.5.0+fa2bec042995004f45b914dd1d66b466ccced410

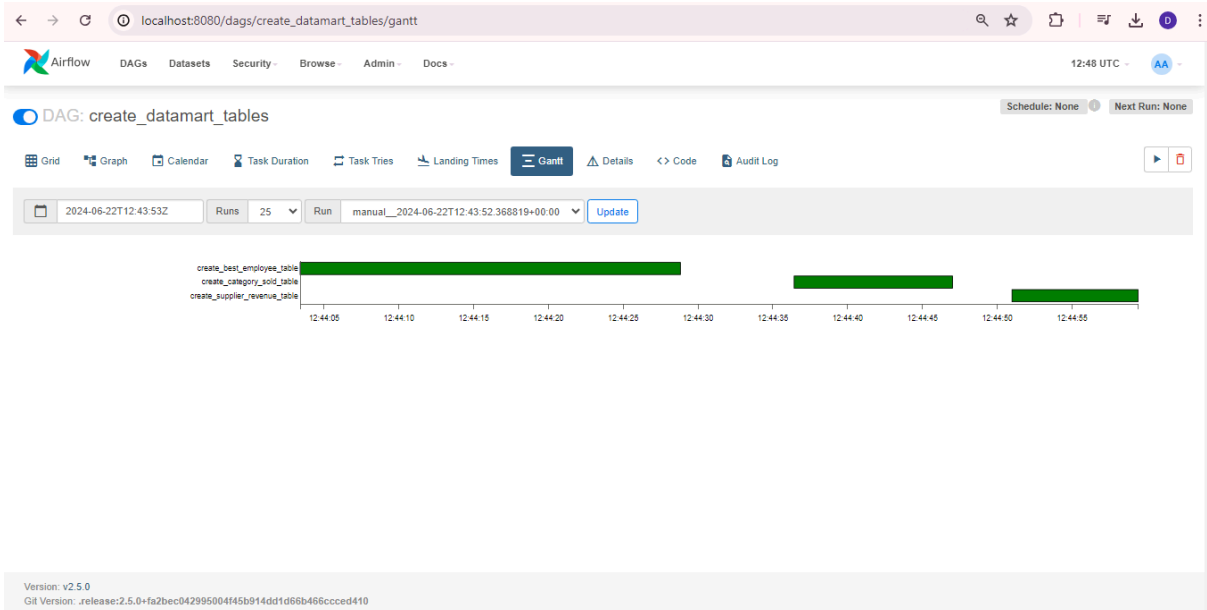
Tampilan di Grid Running Datamart :



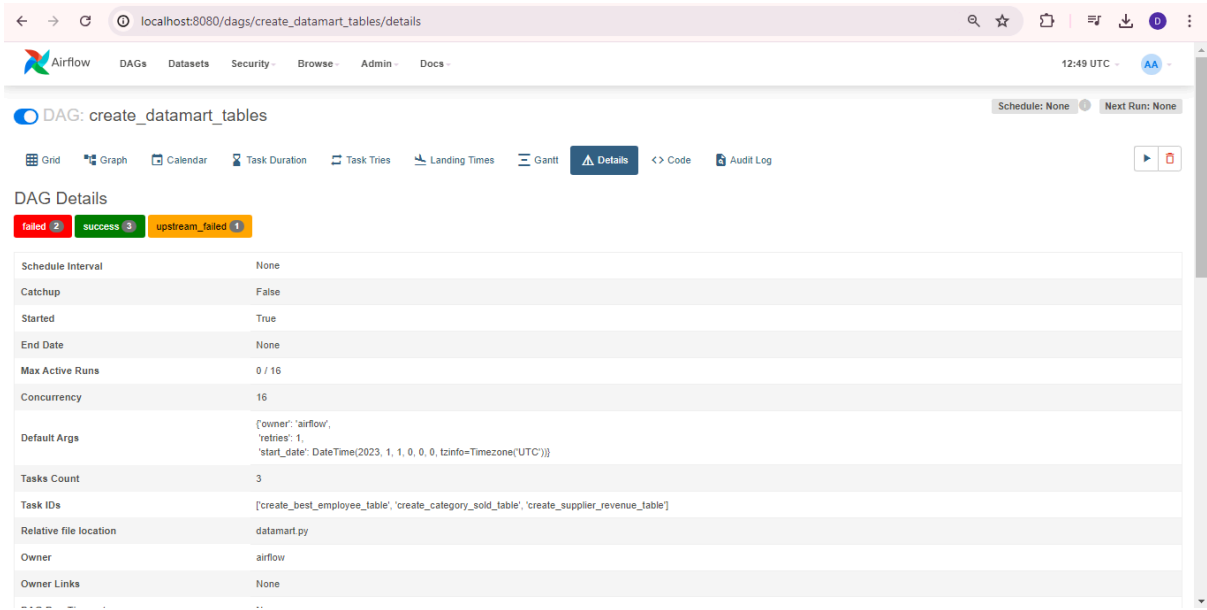
Tampilan di Graph Running Datamart :



Tampilan di Grantt Running Datamart :



Tampilan di Detail Running Datamart :




Tampilan di Code Running Datamart :

← → ↺

localhost:8080/dags/create_datamart_tables/code?root=

🔍 ☆ 📄 ⌵ ⓘ

 DAGs Datasets Security - Browse - Admin - Docs -

12:49 UTC - AA -

DAG: create_datamart_tables

Schedule: None | Next Run: None

Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details <> Code Audit Log

Parsed at: 2024-06-22, 12:49:00

```
1 from airflow import DAG
2 from airflow.providers.snowflake.operators.snowflake import SnowflakeOperator
3 from datetime import datetime
4
5 # Definiskan default args
6 default_args = {
7     'owner': 'airflow',
8     'start_date': datetime(2023, 1, 1),
9     'retries': 1,
10 }
11
12 # Definiskan query untuk membuat tabel data mart
13 create_datamart_best_employee_query = """
14 CREATE OR REPLACE TABLE PROJECT3.PUBLIC.DATAMART_MONTHLY_BEST_EMPLOYEE AS
15 SELECT
16     e.FIRSTNAME || ' ' || e.LASTNAME AS EMPLOYEE_NAME,
17     EXTRACT(MONTH FROM o.ORDERDATE) AS MONTH,
18     SUM((od.UNITPRICE - (od.UNITPRICE * od.DISCOUNT)) * od.QUANTITY) AS TOTAL_GROSS_REVENUE
19 FROM
20     PROJECT3.PUBLIC.ORDERS o
21 JOIN PROJECT3.PUBLIC.ORDER_DETAILS od ON o.ORDERID = od.ORDERID
22 JOIN PROJECT3.PUBLIC.EMPLOYEES e ON o.EMPLOYEEID = e.EMPLOYEEID
23 GROUP BY
24     e.FIRSTNAME,
25     e.LASTNAME,
26     EXTRACT(MONTH FROM o.ORDERDATE)
27 ORDER BY
28     EXTRACT(MONTH FROM o.ORDERDATE),
29     TOTAL_GROSS_REVENUE DESC;
```


Toggle Wrap

Tampilan Log Running Datamart :

← → ↺

localhost:8080/dags/create_datamart_tables/audit_log?root=

🔍 ☆ 📄 ⌵ ⓘ

 DAGs Datasets Security - Browse - Admin - Docs -

12:50 UTC - AA -

DAG: create_datamart_tables

Schedule: None | Next Run: None

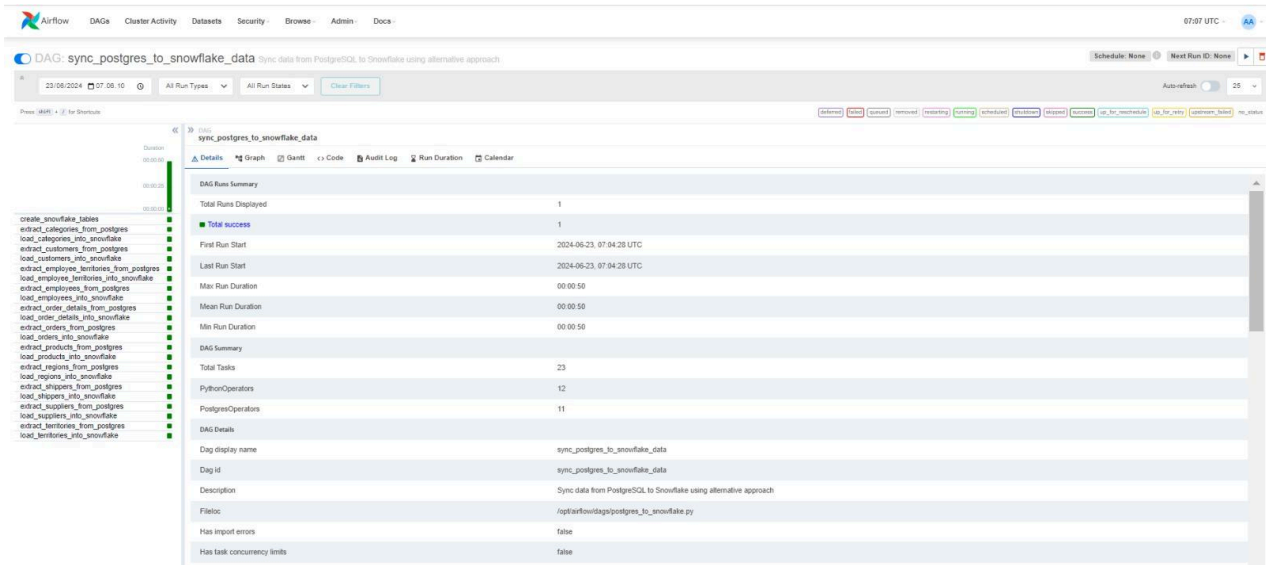
Grid Graph Calendar Task Duration Task Tries Landing Times Gantt Details <> Code Audit Log

Dag Audit Log

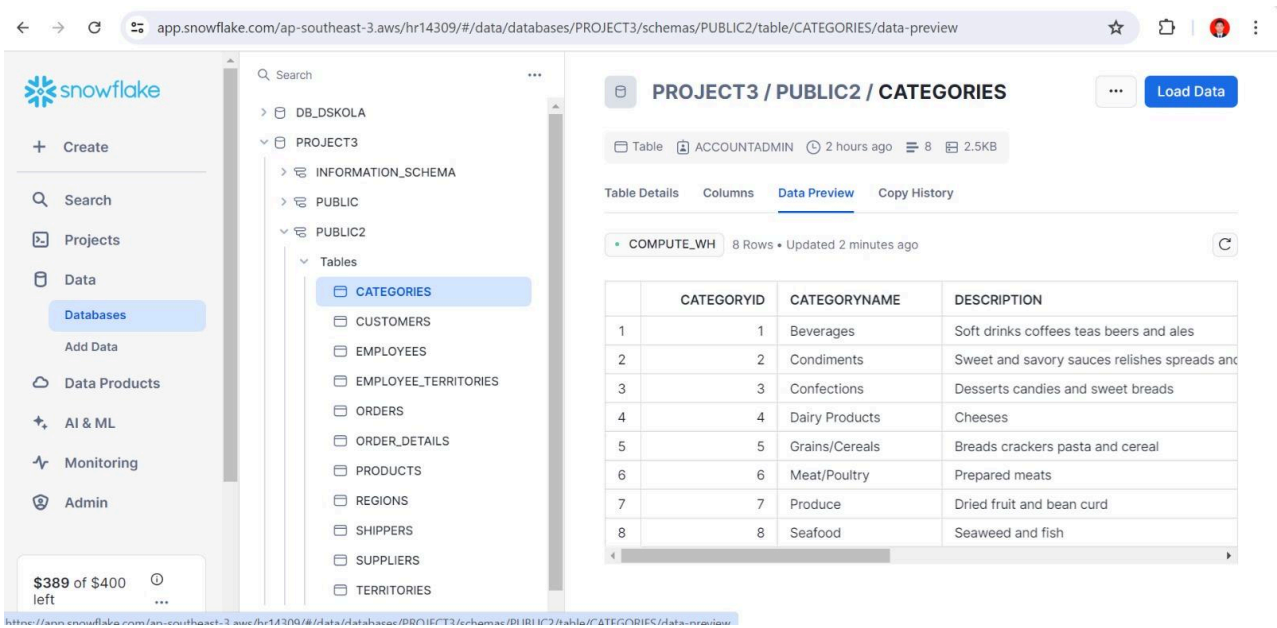
This view displays selected events and operations that have been taken on this dag. The included and excluded events are set in the Airflow configuration, which by default remove view only actions. For a full list of events regarding this DAG, click here.

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, 12:44:36	create_category_sold_table	cli_task_run	None	airflow	{"host_name": "03783357291b", "full_command": "[/home/airflow/.local/bin/airflow, 'celery', 'worker']"}

Tampilan sinkronisasi dari Postgress ke Snowflake



Tampilan Data Hasil categories di Snowflake



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

Tampilan table Hasil datamart di Snowflake :

```
1 create or replace TABLE
2 PROJECT3.PUBLIC.DATAMART_MONTHLY_SUPPLIER_GROSS_REVENUE (
3   SUPPLIER_NAME VARCHAR(256),
4   MONTH NUMBER(2,0),
5   GROSS_REVENUE NUMBER(38,9)
```