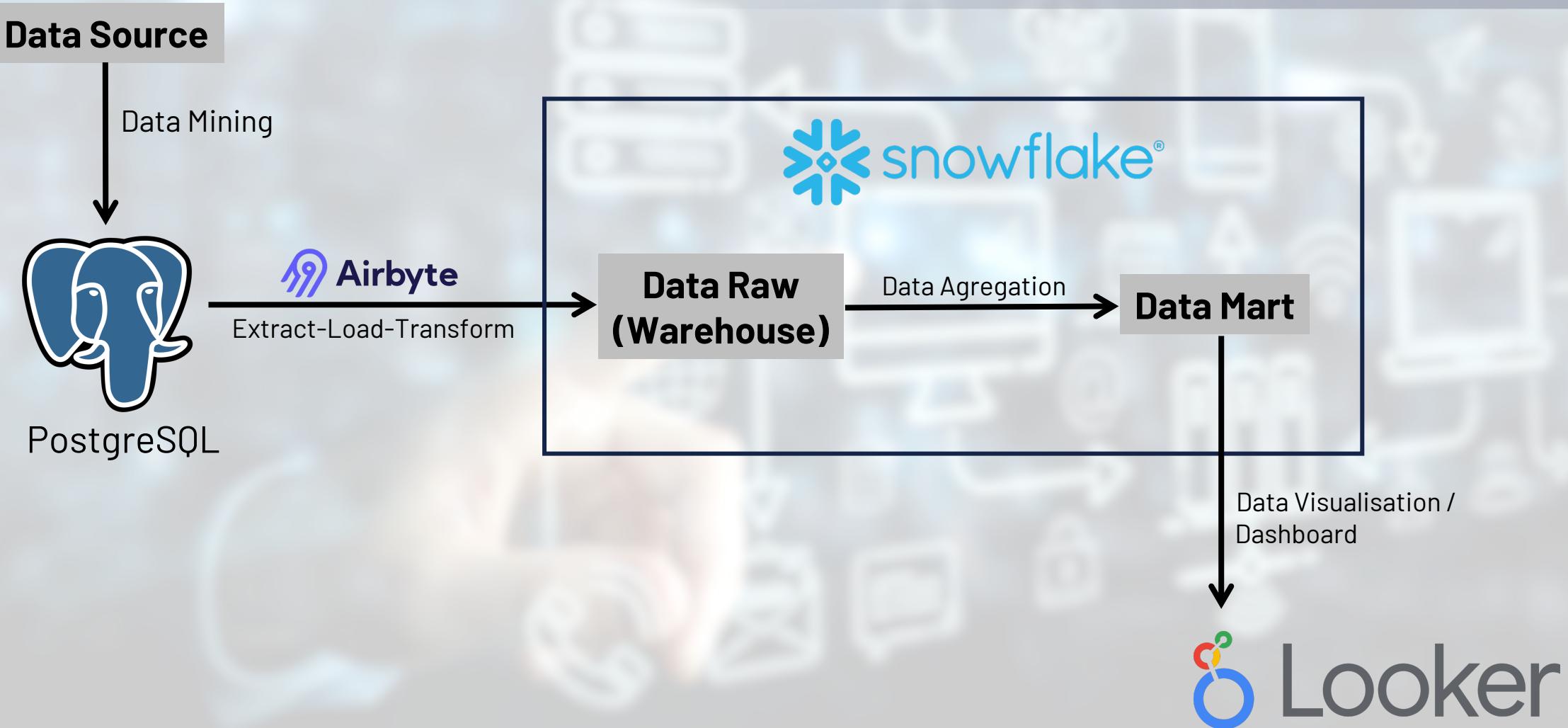


Final Project Documentation: Data Pipeline Architecture with PostgreSQL, Snowflake and Google Looker Studio

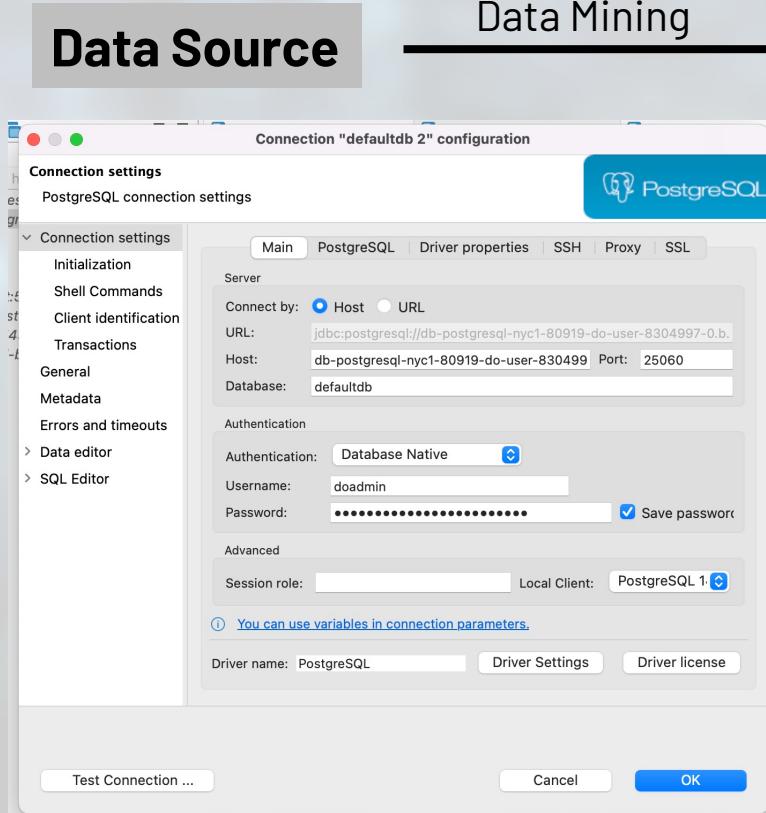
Rheza Paleva Uyanto
Digital Skola – Data Engineer Batch 12

-  uyantorheza@gmail.com
-  linkedin.com/in/rheza-uyanto/
-  github.com/rhezapal

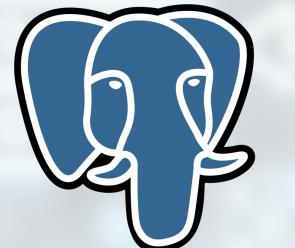
Data Pipeline Diagram



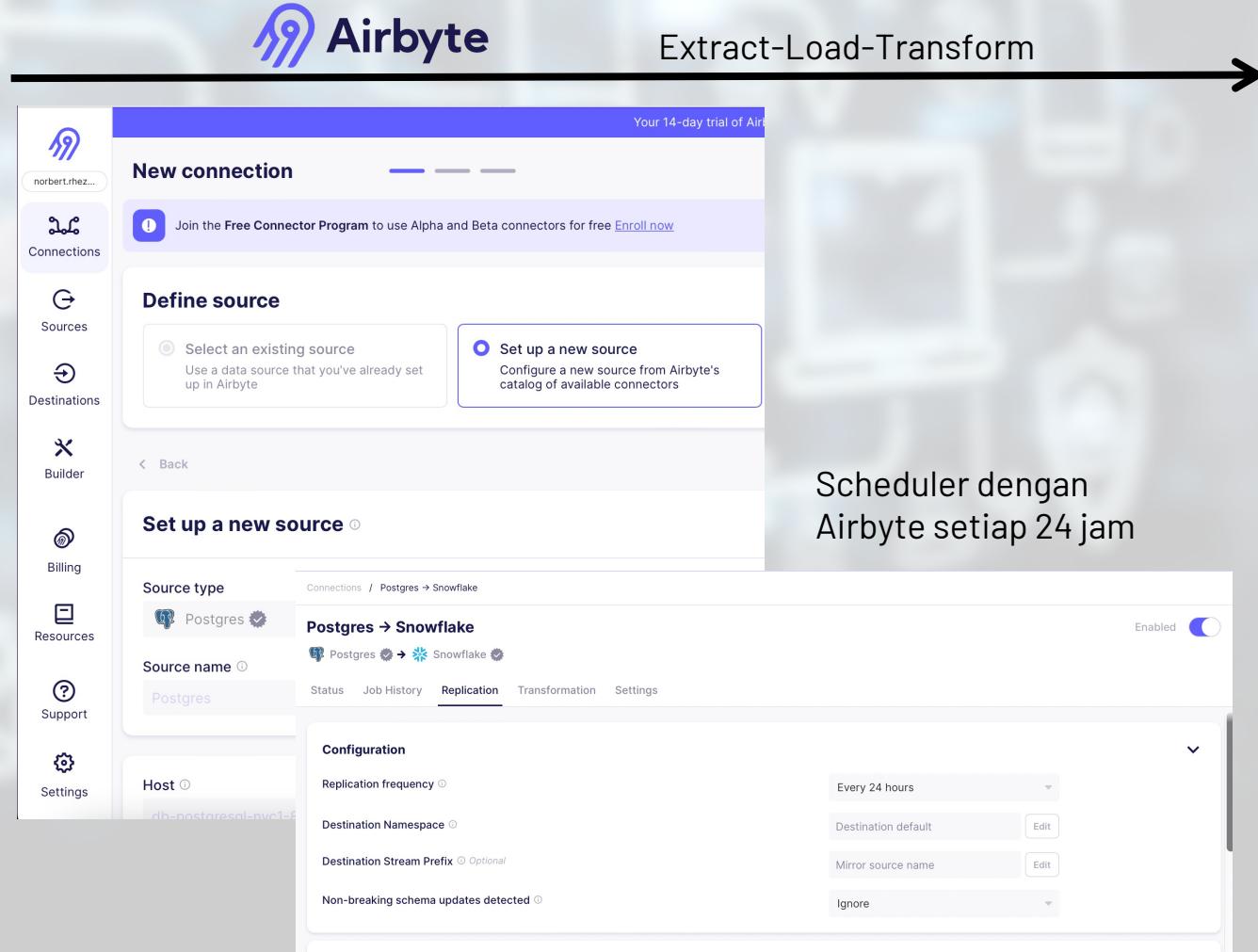
Explanation (1)



Data Mining



PostgreSQL



Airbyte

New connection

Join the Free Connector Program to use Alpha and Beta connectors for free [Enroll now](#)

Connections

Sources

Destinations

Builder

Billing

Resources

Support

Settings

Define source

Select an existing source

Use a data source that you've already set up in Airbyte

Set up a new source

Configure a new source from Airbyte's catalog of available connectors

Back

Set up a new source

Source type

Postgres → Snowflake

Postgres → Snowflake

Source name

Postgres

Status Job History Replication Transformation Settings

Configuration

Replication frequency

Every 24 hours

Destination Namespace

Destination Stream Prefix Optional

Non-breaking schema updates detected

Mirror source name

Ignore

Your 14-day trial of Air

Extract-Load-Transform

Scheduler dengan Airbyte setiap 24 jam

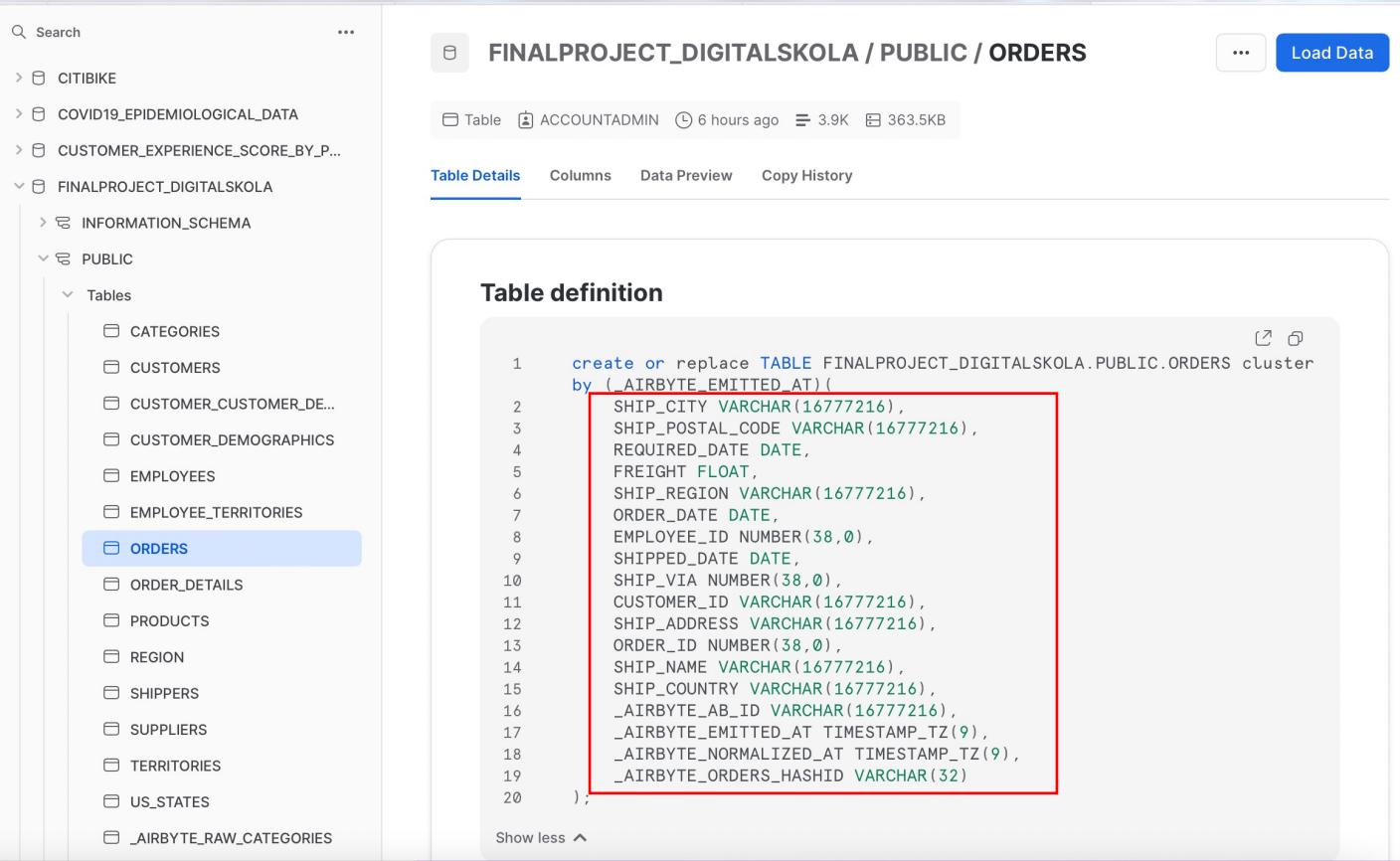
Explanation (2)



Data Raw
(Warehouse)

...Data Agregation

- Cek data type : sesuai, dan terdapat 14 tabel

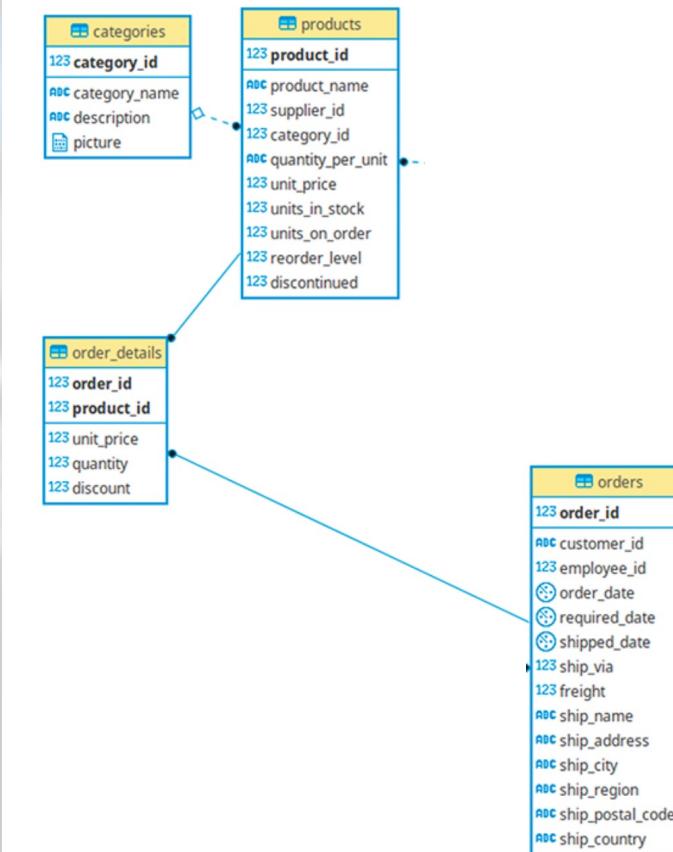


The screenshot shows the Snowflake interface with the following details:

- Left sidebar:** Shows the schema structure with nodes like CITIBIKE, COVID19_EPIDEMIOLOGICAL_DATA, CUSTOMER_EXPERIENCE_SCORE_BY_P..., FINALPROJECT_DIGITALSKOLA, INFORMATION_SCHEMA, PUBLIC, and various tables under PUBLIC.
- Table definition view:** For the ORDERS table in the FINALPROJECT_DIGITALSKOLA.PUBLIC schema. The code is as follows:

```
1  create or replace TABLE FINALPROJECT_DIGITALSKOLA.PUBLIC.ORDERS cluster
by (_AIRBYTE_EMITTED_AT)
2
3      SHIP_CITY VARCHAR(16777216),
4      SHIP_POSTAL_CODE VARCHAR(16777216),
5      REQUIRED_DATE DATE,
6      FREIGHT FLOAT,
7      SHIP_REGION VARCHAR(16777216),
8      ORDER_DATE DATE,
9      EMPLOYEE_ID NUMBER(38,0),
10     SHIPPED_DATE DATE,
11     SHIP_VIA NUMBER(38,0),
12     CUSTOMER_ID VARCHAR(16777216),
13     SHIP_ADDRESS VARCHAR(16777216),
14     ORDER_ID NUMBER(38,0),
15     SHIP_NAME VARCHAR(16777216),
16     SHIP_COUNTRY VARCHAR(16777216),
17     _AIRBYTE_AB_ID VARCHAR(16777216),
18     _AIRBYTE_EMITTED_AT TIMESTAMP_TZ(9),
19     _AIRBYTE_NORMALIZED_AT TIMESTAMP_TZ(9),
20     _AIRBYTE_ORDERS_HASHID VARCHAR(32)
```

A red box highlights the column definitions from SHIP_CITY to _AIRBYTE_ORDERS_HASHID.



Explanation (3)



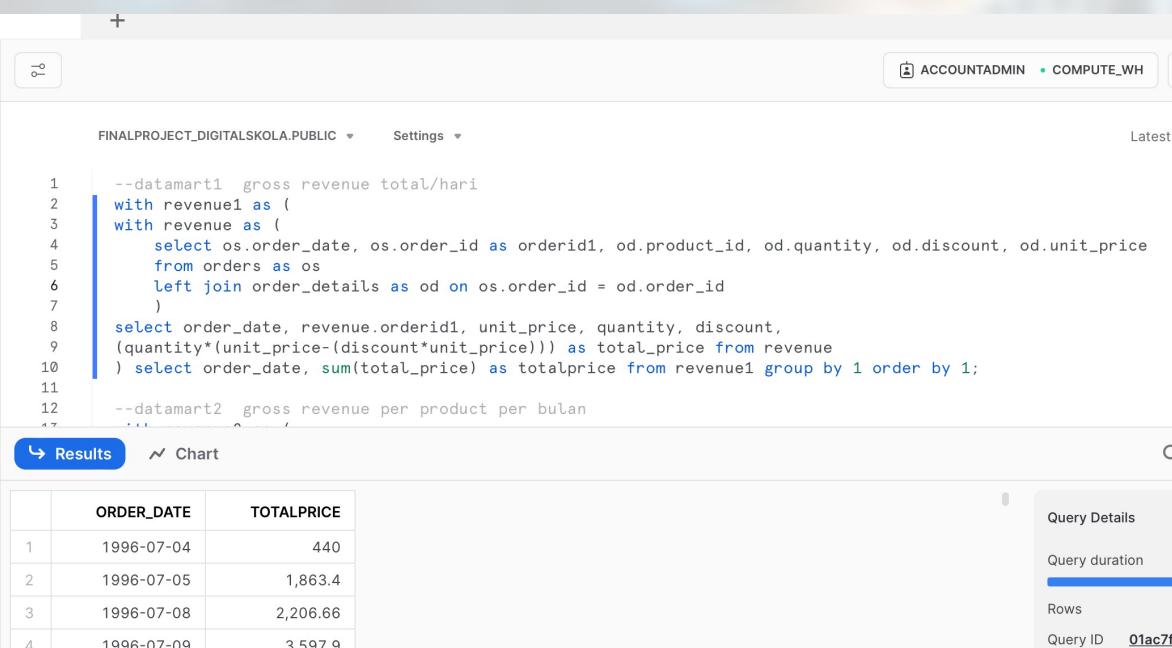
Data Aggregation

→ Data Mart

...Data Visualization

4 tabel yang digunakan, dan ada 5 report yang diminta :

- Gross revenue total per hari
- Gross Revenue per product per bulan
- Jumlah total pembelian per product per bulan
- Jumlah total pembelian per kategori product per bulan
- Jumlah total pembelian per negara per bulan



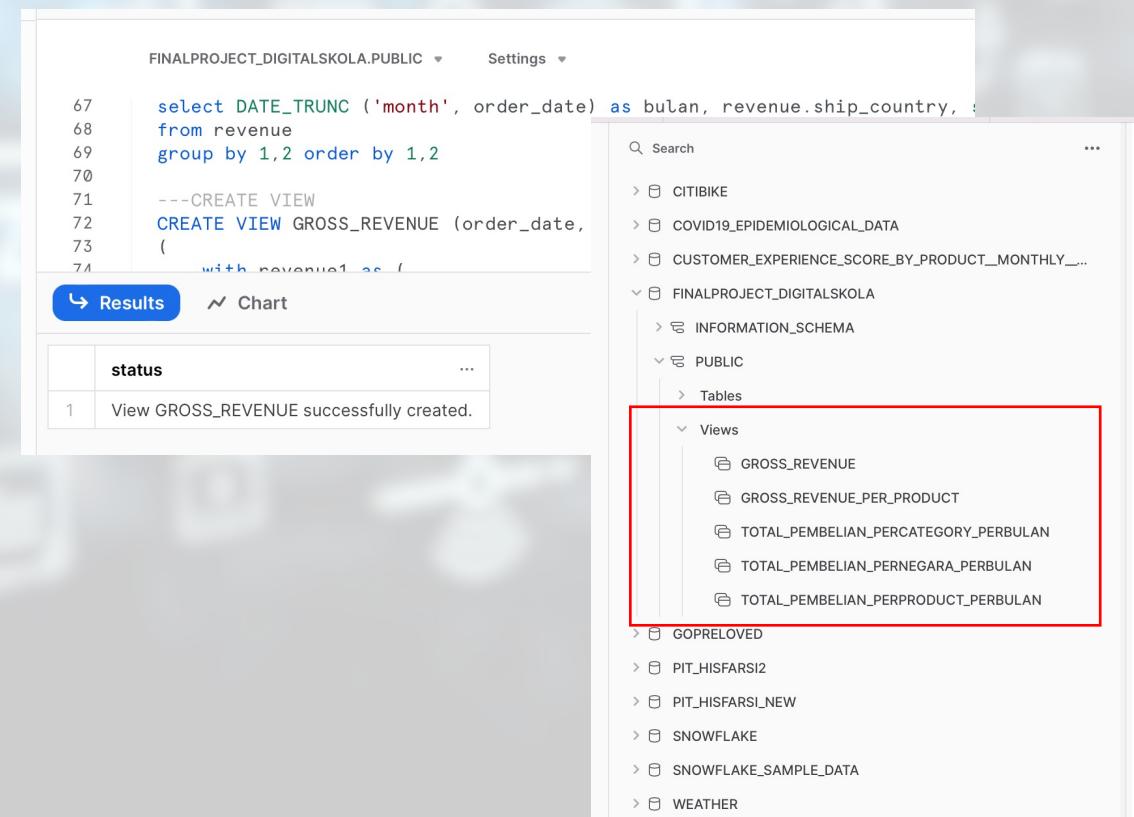
The screenshot shows the Snowflake interface. On the left, a code editor displays a SQL script for creating two data mart views. The right side shows the results of the first query, which is a table with columns ORDER_DATE and TOTALPRICE.

ORDER_DATE	TOTALPRICE
1996-07-04	440
1996-07-05	1,863.4
1996-07-08	2,206.66
1996-07-09	3,597.9

Query Details:

- Query duration: 01ac7f2
- Rows: 4

Membuat create view



The screenshot shows the Snowflake interface. A query editor on the left contains SQL code for creating a view named GROSS_REVENUE. The right side shows the database navigation pane with a red box highlighting the 'Views' section under the 'FINALPROJECT_DIGITALSKOLA' schema, which contains the newly created view.

```
select DATE_TRUNC ('month', order_date) as bulan, revenue.ship_country, ...  
from revenue  
group by 1,2 order by 1,2  
---CREATE VIEW  
CREATE VIEW GROSS_REVENUE (order_date,  
(  
with revenue1 as (
```

status

1 View GROSS_REVENUE successfully created.

Views

- GROSS_REVENUE
- GROSS_REVENUE_PER_PRODUCT
- TOTAL PEMBELIAN_PERCATEGORY_PERBULAN
- TOTAL PEMBELIAN_PERNEGARA_PERBULAN
- TOTAL PEMBELIAN_PERPRODUCT_PERBULAN

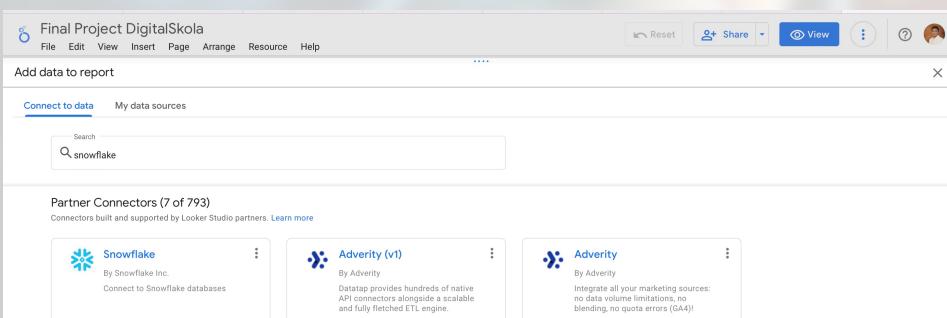
Explanation (4)

Data Visualization

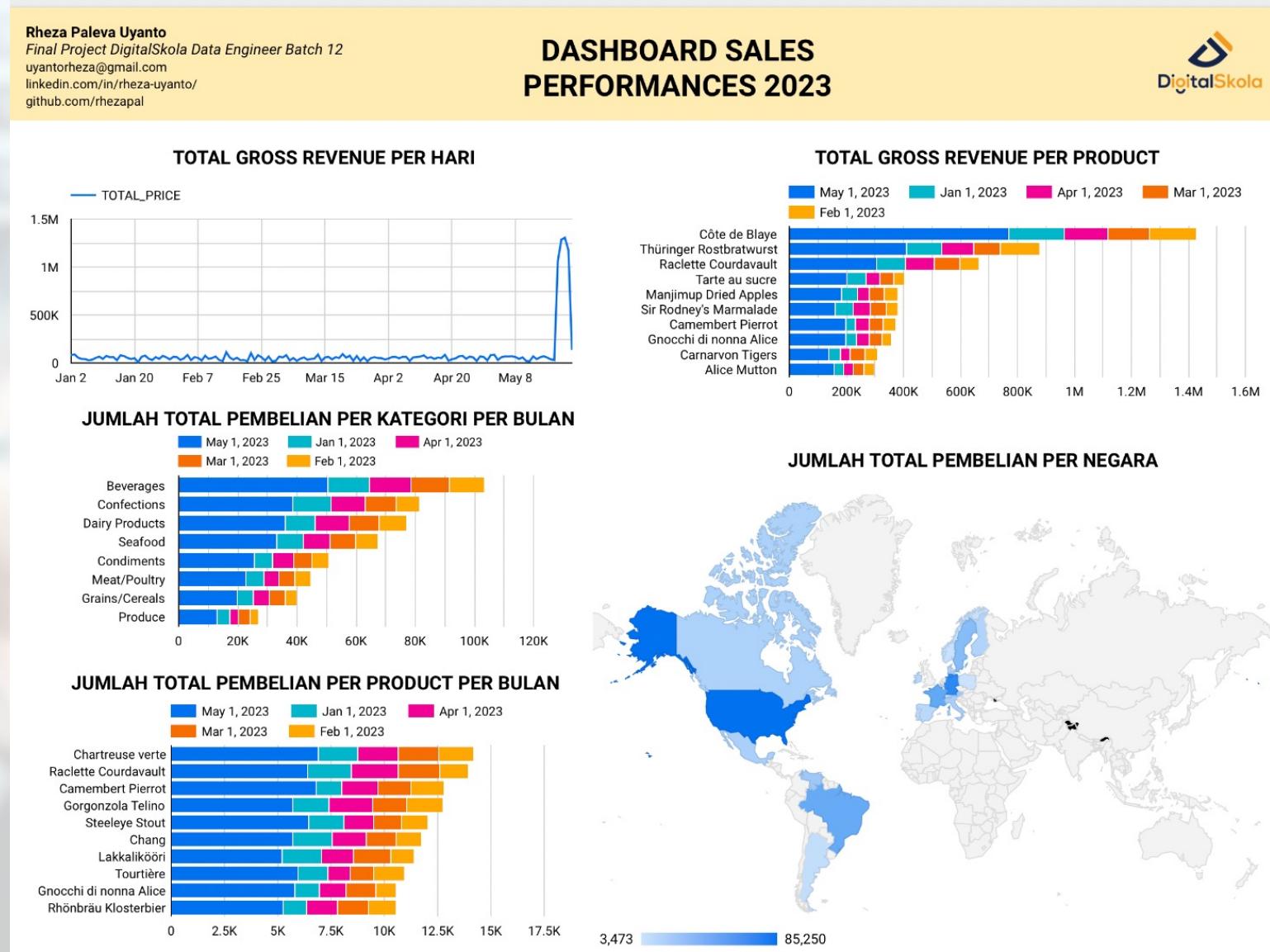


5 grafik yang diminta :

- **Gross revenue total per hari** (line chart)
- **Gross Revenue per product per bulan** (bar chart)
- **Jumlah total pembelian per product per bulan** (bar chart, top 10 highest)
- **Jumlah total pembelian per kategori product per bulan** (bar chart, top 10 highest)
- **Jumlah total pembelian per negara per bulan** (map chart)



Untuk akses dashboard, silakan klik [link](#) berikut ini



References

- <https://docs.snowflake.com/en/sql-reference/constructs/join>
- <https://docs.snowflake.com/en/sql-reference/sql/create-view>
- <https://docs.airbyte.com>
- <https://cloud.google.com/looker/docs>
- Materi – PostgreSQL