

## **Digitalskola Data Engineer batch 12 final project**

Project template dapat dilihat dan di fork di:

<https://github.com/saungkertas/digitalskolade12>

1. Buatlah arsitektur end to end dari source (postgresql), ke datawarehouse(Snowflake), hingga data visualization.
2. Buatlah scheduler untuk ingest data dari Postgresql ke datawarehouse(Snowflake), dapat menggunakan ETL seperti (airbyte, python script, other ETL tools)
3. Buatlah scheduler menggunakan airflow untuk membentuk table Datamart.
4. Buatlah datamart view berikut:
  - 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
4. Buatlah dashboard chart dari poin 3. (dapat menggunakan Snowsight atau Looker Data Studio)
  - 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)
5. Setelah berhasil, masukkan python code dan airflow dags di dalam directory dags di repository ini airflow/dags/
6. Presentasikan hasil pengerjaan, termasuk arsitektur pipeline, step pengerjaan, hingga kendala yang dialami selama pengerjaan.

Catatan:

- database source dapat diakses:

username = doadmin

password = AVNS\_Lx9PbZB62Y5BMnMI6UA

host = db-postgresql-nyc1-80919-do-user-8304997-0.b.db.ondigitalocean.com

port = 25060

database = defaultdb

- airflow dapat diakses pada:

<http://35.197.159.19:8080>

- dags airflow yang sudah jadi dapat diletakkan di google cloud storage (gcs) berikut:

<gs://digitalskola-de12-final/dags/>

- dataset yang digunakan adalah northwind dataset  
([https://raw.githubusercontent.com/pthom/northwind\\_psql/master/ER.png](https://raw.githubusercontent.com/pthom/northwind_psql/master/ER.png))
- fork project <https://github.com/saungkertas/digitalskolade12> dan submit PR jika sudah selesai
- script dapat dijalankan manual terlebih dahulu di server sebelum disubmit(example:python3 ingest\_orders.py 2022-08-04)
- untuk develop script atau coba2, bisa dilakukan di directory /root/coba2.
- folder dags akan di copy tiap 2 menit dari gs://digitalskola-de12-final/dags/
- silakan japri jika ada perubahan script/dags, akan diapprove sore hari tiap harinya.
- Fact table menggunakan daily run, sedangkan dimension table hanya perlu init run (sekali run saja untuk mengambil semua data)
- setiap table yang dicreate, beri prefix <nama>\_nama\_table

