

New Assignment - Airflow Connection & DAGs Running

1. (10 point) Jalankan docker apache airflow kemudian lampirkan bukti:

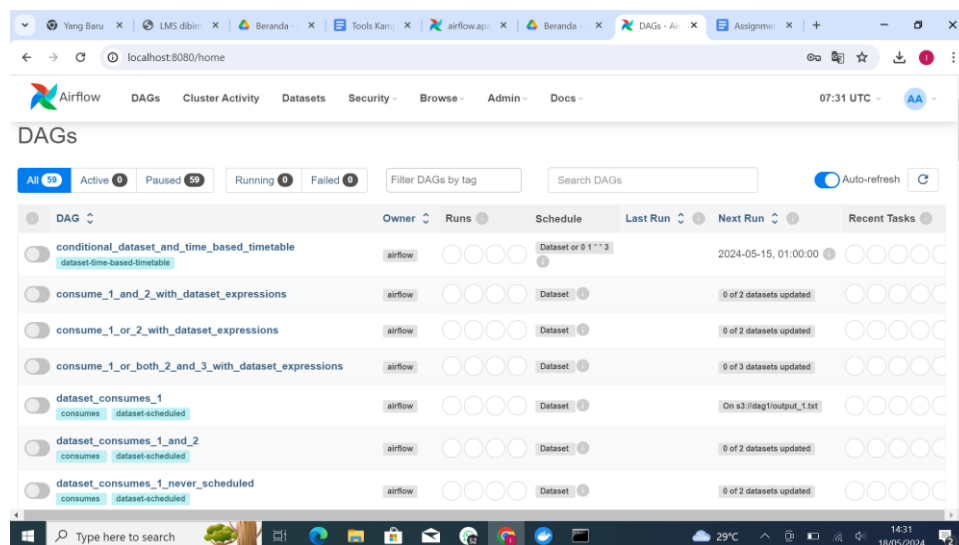
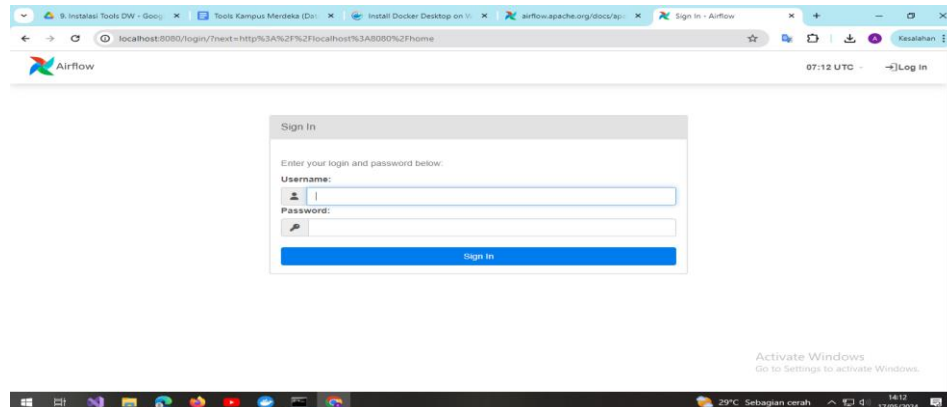
a. Screenshot terminal

```
C:\Windows\System32\cmd.exe
C:\Users\LABKON2\Downloads\Airflow>docker compose down
time="2024-05-18T10:09:59+07:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
time="2024-05-18T10:09:59+07:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
[+] Running 8/8
  Container airflow-airflow-worker-1   Removed    3.0s
  Container airflow-airflow-triggerer-1 Removed    1.0s
  Container airflow-airflow-webserver-1 Removed    4.2s
  Container airflow-airflow-scheduler-1 Removed    2.4s
  Container airflow-airflow-init-1     Removed    0.0s
  Container airflow-postgres-1         Removed    0.6s
  Container airflow-redis-1            Removed    0.7s
  Network airflow_default              Removed    0.2s

C:\Users\LABKON2\Downloads\Airflow>docker compose up -d --build
time="2024-05-18T10:10:39+07:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
time="2024-05-18T10:10:39+07:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
[+] Running 8/8
  Network airflow_default              Created    0.0s
  Container airflow-postgres-1         Healthy    0.2s
  Container airflow-redis-1            Healthy    0.2s
  Container airflow-airflow-init-1     Exited     0.1s
  Container airflow-airflow-scheduler-1 Started     0.2s
  Container airflow-airflow-webserver-1 Started     0.2s
  Container airflow-airflow-triggerer-1 Started     0.3s
  Container airflow-airflow-worker-1   Started     0.3s

C:\Users\LABKON2\Downloads\Airflow>
```

b. Screenshot localhost:8080

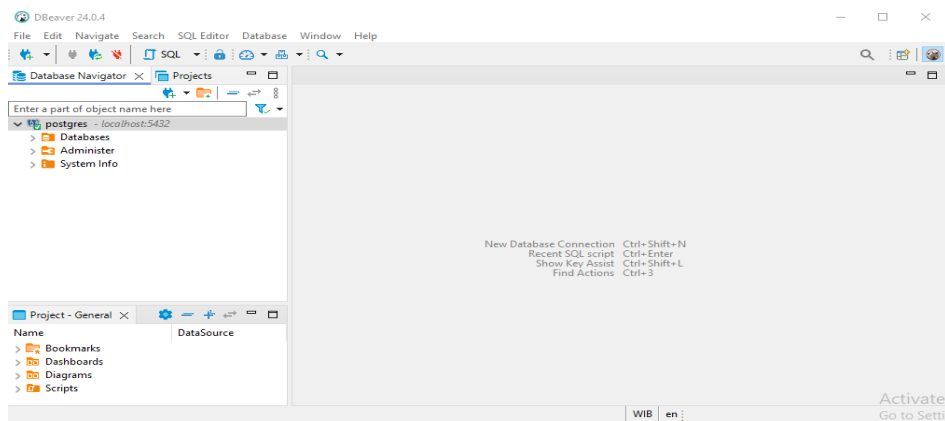


2. (10 point) Jalankan docker database mysql dan postgresql kemudian lampirkan bukti:

a. Screenshot terminal

```
postgres /usr/local/bin/docker-entrypoint.sh: ignoring /docker-entrypoint-initdb.d/*
postgres
postgres waiting for server to shut down... 2024-05-18 07:02:15.611 UTC [48] LOG: received fast shutdown request
postgres 2024-05-18 07:02:15.614 UTC [48] LOG: aborting any active transactions
postgres 2024-05-18 07:02:15.615 UTC [48] LOG: background worker "logical replication launcher" (PID 55) exited with exit cod
e 1
postgres 2024-05-18 07:02:15.615 UTC [50] LOG: shutting down
postgres 2024-05-18 07:02:15.646 UTC [48] LOG: database system is shut down
postgres done
postgres server stopped
postgres PostgreSQL init process complete; ready for start up.
postgres
postgres 2024-05-18 07:02:15.729 UTC [1] LOG: starting PostgreSQL 13.15 (Debian 13.15-1.pgdg120+1) on x86_64-linux-gnu, co
mpiled by gcc (Debian 12.2.0-14) 12.2.0, 64-bit
postgres 2024-05-18 07:02:15.730 UTC [1] LOG: listening on IPv4 address "0.0.0.0", port 5432
postgres 2024-05-18 07:02:15.730 UTC [1] LOG: listening on IPv6 address ":::", port 5432
postgres 2024-05-18 07:02:15.733 UTC [1] LOG: listening on Unix socket "/var/run/postgresql/.s.PGSQL.5432"
postgres 2024-05-18 07:02:15.740 UTC [63] LOG: database system was shut down at 2024-05-18 07:02:15 UTC
postgres 2024-05-18 07:02:15.744 UTC [1] LOG: database system is ready to accept connections
airflow-init .....
airflow-init ERROR! Maximum number of retries (20) reached.
airflow-init Last check result:
airflow-init $ airflow db check
airflow-init /home/airflow/.local/bin/airflow: line 3: import: command not found
airflow-init /home/airflow/.local/bin/airflow: line 4: import: command not found
airflow-init /home/airflow/.local/bin/airflow: line 5: from: command not found
airflow-init /home/airflow/.local/bin/airflow: line 7: syntax error near unexpected token `('
airflow-init /home/airflow/.local/bin/airflow: line 7: ` sys.argv[0] = re.sub(r'(-script|.pyw|\.exe)?$', '', sys.argv[0])'
airflow-init airflow-init exited with code 1
service "airflow-init" didn't complete successfully: exit 1
PS C:\Users\LABKOM2\Downloads\WSIB\WSIB-DW-Batch_6-Airflow\airflow> []
```

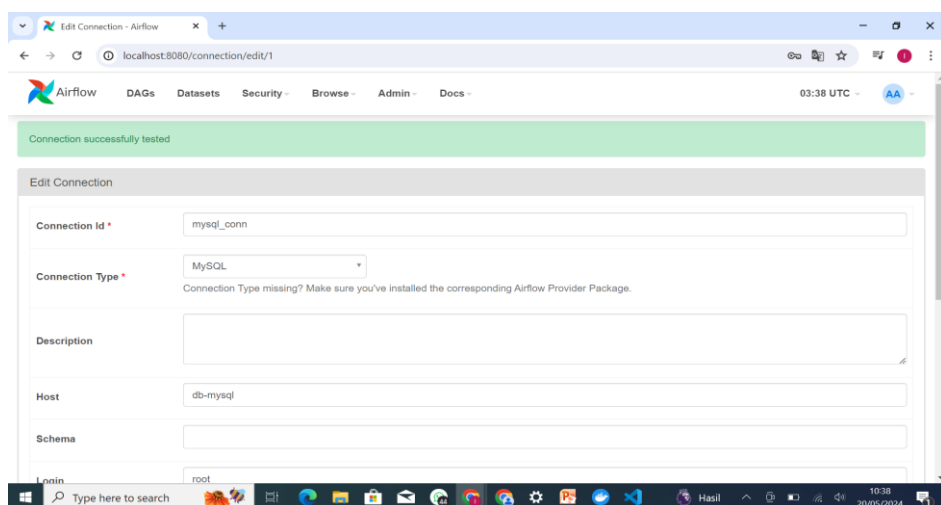
b. Screenshot Dbeaver / MySQL Connection Extension yang telah terhubung dengan postgresql dan mysql



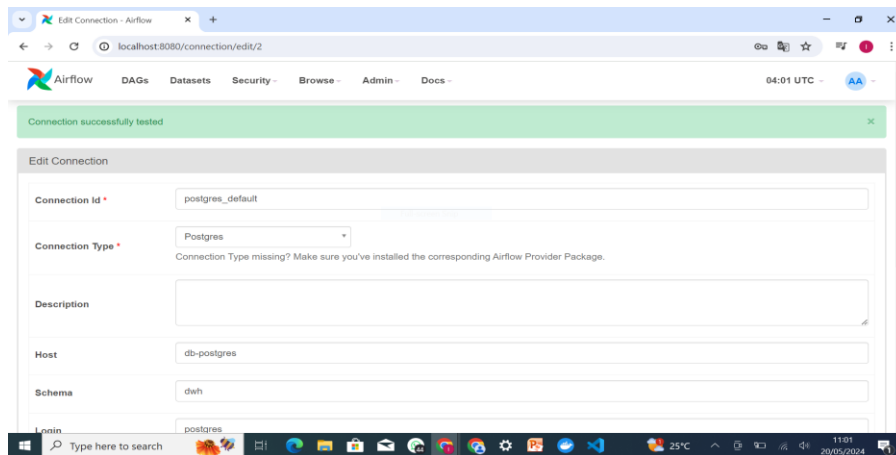
3. (10 point) Buat connection MySQL dan PostgreSQL pada Apache Airflow kemudian

a. Screenshot UI hasil testing yang menunjukkan connection successful

MySQL

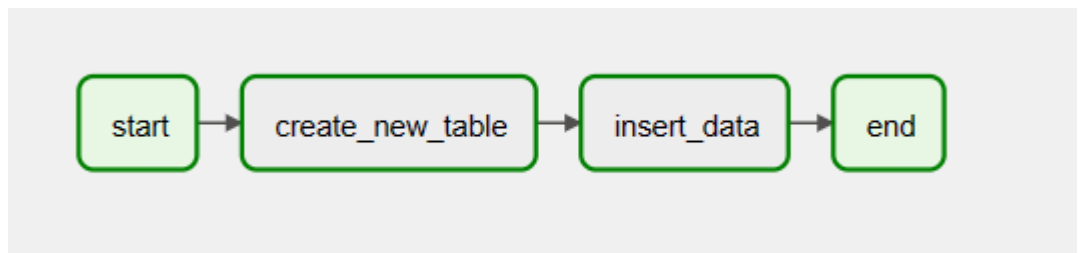


PostgreSQL

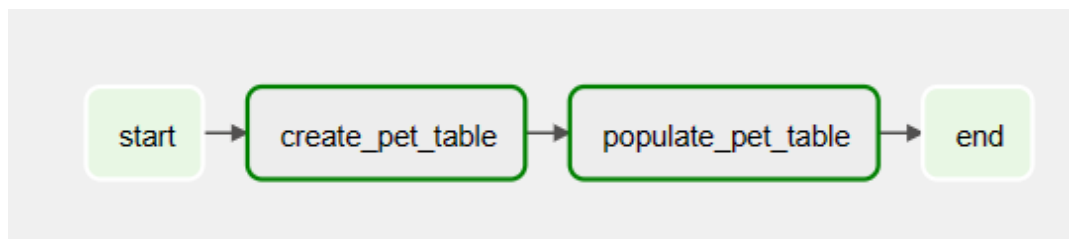


4. (10 point) Modifikasi DAGs `op_mysql` dan `op_postgresql` yang sudah ada dengan menambahkan EmptyOperator sehingga tampilan kedua DAG tersebut menjadi sebagai berikut:

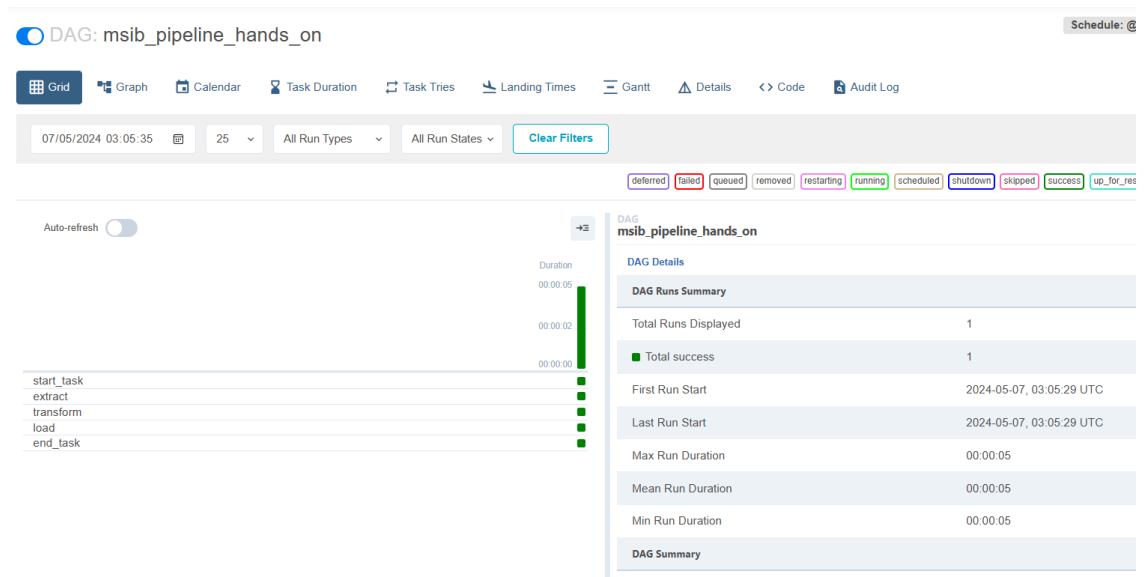
a. `op_mysql` :



b. `op_postgresql`



5. (10 point) Jalankan DAGs `op_mysql` dan `op_postgresql` hingga `success` kemudian lampirkan hasil screenshotnya. Usahakan sampai pipeline berjalan `success` semua



6. (10 point) Tampilkan hasil table yang dibuat oleh `op_mysql` dan `op_postgresql`
7. (30 point) Jelaskan apa yang dikerjakan oleh `op_mysql` dan `op_postgresql`? Apa perbedaan sebelum dan sesudah dijalankannya DAG? dan apa yang dapat di improve dari kedua DAG tersebut?

Op mysql dan PostgreSQL.

hal ini memungkinkan pengguna untuk menjalankan kueri SQL, mengelola koneksi database, dan mengelola data dalam saluran Airflow. `op_mysql` biasanya mengacu pada `MySQLOperator`, yang memungkinkan untuk kita menjalankan kueri SQL pada database MySQL.

Berikut ini ada beberapa tugas yang dapat dilakukan oleh `MySQLOperator`.

- Jalankan kueri `SELECT` untuk membaca data dari database.
 - Memodifikasi data dalam tabel dengan menjalankan kueri `INSERT`, `UPDATE`, atau `DELETE`.
 - Menjalankan prosedur atau fungsi tersimpan dalam database.
 - Jalankan skrip SQL dari file eksternal.
- `op_postgresql` biasanya mengacu pada `PostgresOperator`, yang memiliki fungsi serupa dengan `MySQLOperator`, tetapi untuk database PostgreSQL. dapat melakukan tugas-tugas berikut:
- Jalankan kueri `SELECT` untuk membaca data.
 - Memodifikasi data dengan menjalankan kueri `INSERT`, `UPDATE`, atau `DELETE`.
 - Menjalankan prosedur atau fungsi tersimpan.
 - Jalankan skrip SQL dari file eksternal.

8. (10 point) Buat Kesimpulan mengenai Apache Airflow

Apache Airflow adalah sumber platform yang terbuka atau sering disebut dengan open-source untuk mengembangkan, menjadwalkan dan memantau alur kerja berorientasi batch. Dan juga menggunakan kerangka kerja python yang bisa diperluas dan memungkinkan untuk membangun alur kerja yang terhubung dengan hampir semua teknologi. Airflow juga dapat diintegrasikan dengan layanan cloud seperti google cloud platform (GCP) untuk berinteraksi dengan layanan Bigquery, apache airflow juga menjadi alat yang sangat digunakan dalam mengelola alur kerja data.

9. Kumpulkan hasil pengerjaan berupa file .pdf dan upload code hasil pengerjaan ke dalam github pribadi masing-masing.