**Data Warehousing dibimbing.id**

**Assignment Day 40**

**ETL WITH AIRFLOW PART 3**



**Kelompok 5**

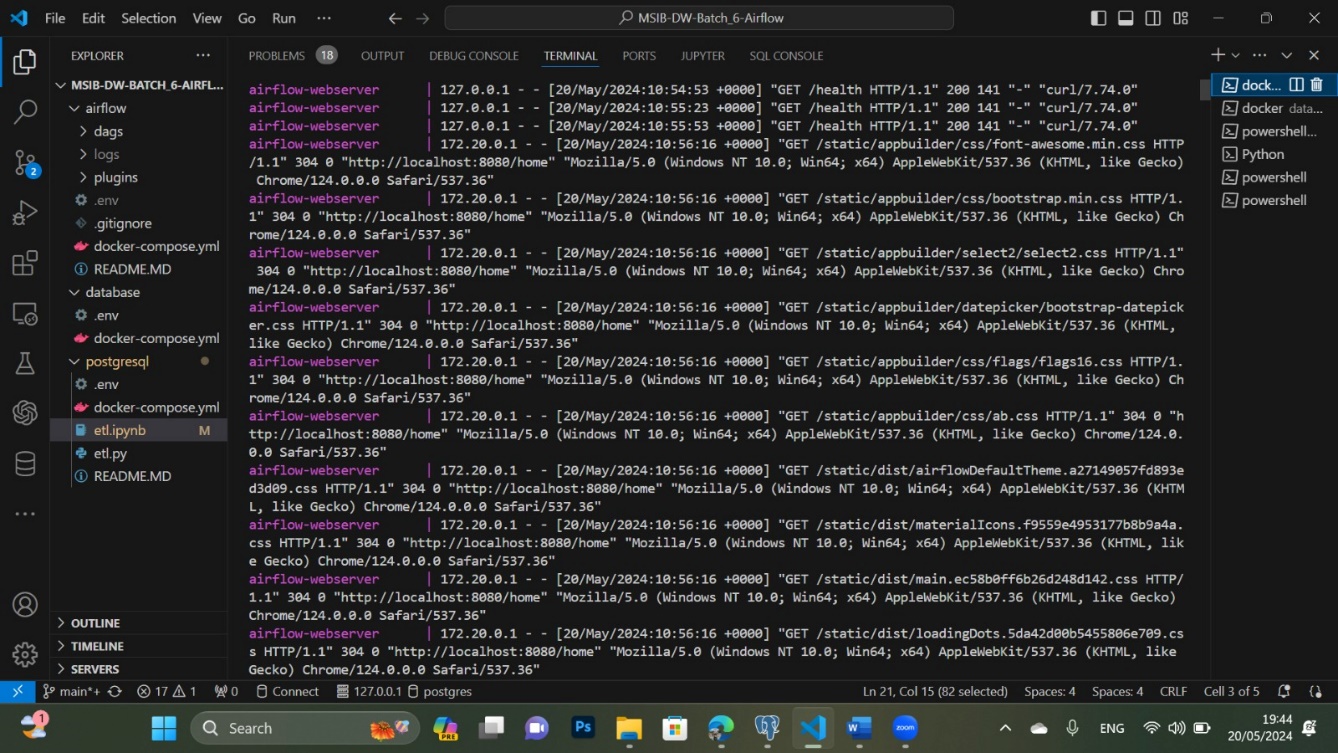
|  |
| --- |
| FATIMAH AZZAHRA |
| FRISKA MURNI KASIH LOALI |
| ISRA PUSPITA HARDANI |
| M RIZKI AZMI |
| TIARA WULANDARI |

**Program Studi Independen Bersertifikat 6 Data Warehousing dibimbing.id**

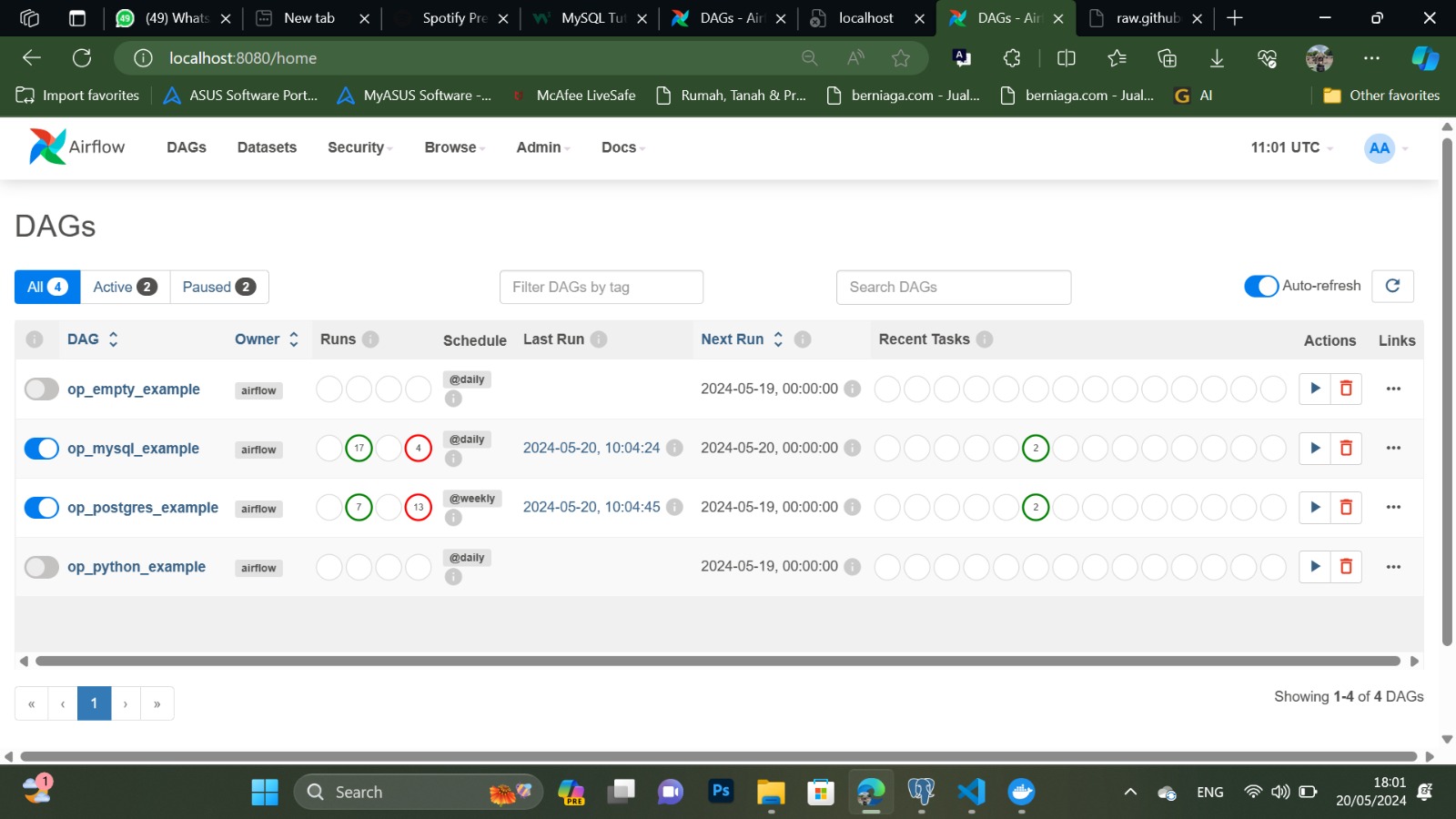
**2024**

**New Assignment - Airflow Connection & DAGs Running**

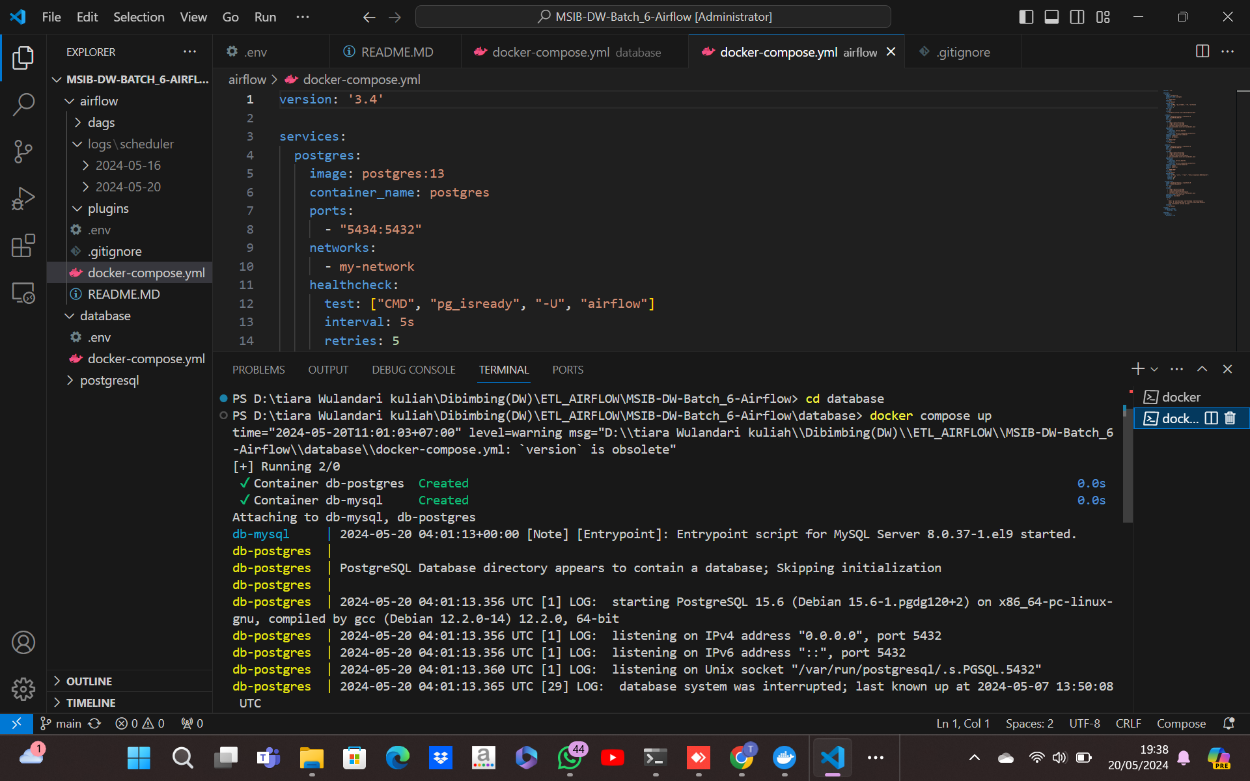
1. Jalankan docker apache airflow kemudian lampirkan bukti:
   * 1. Screenshot terminal



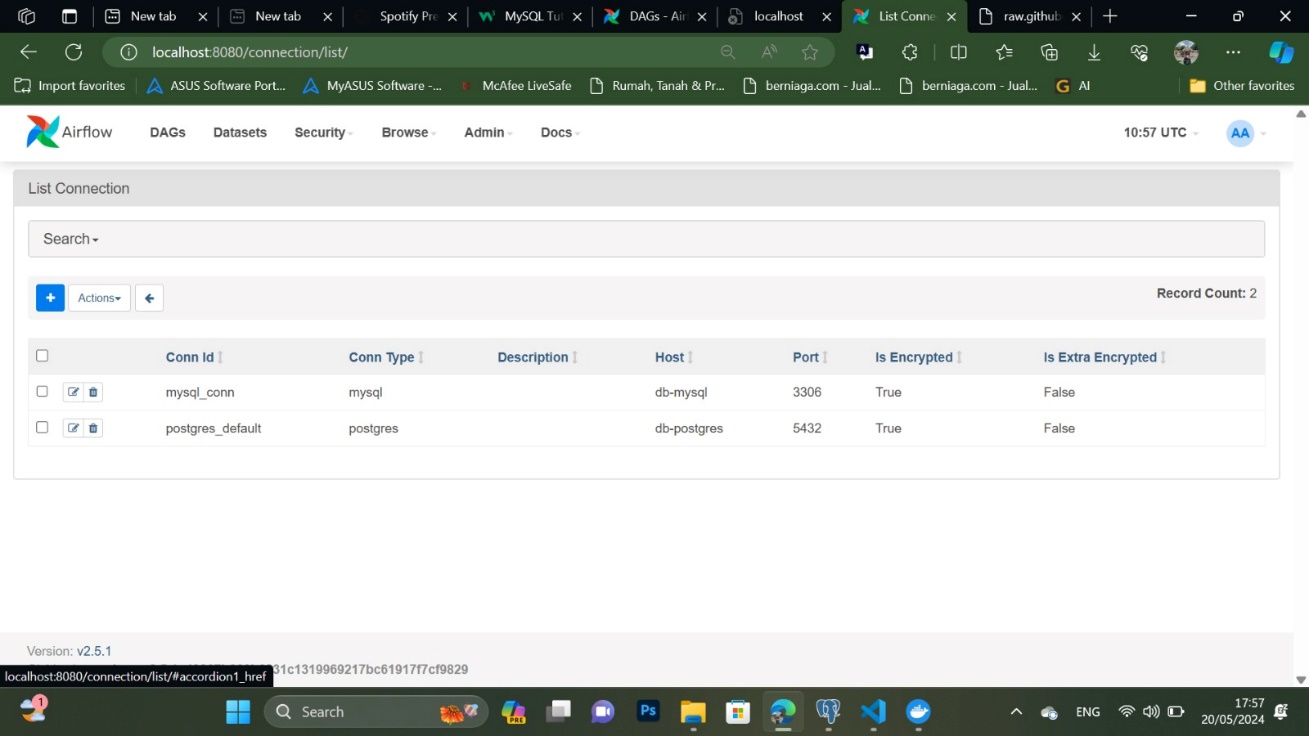
* + 1. Screenshot localhost:8080



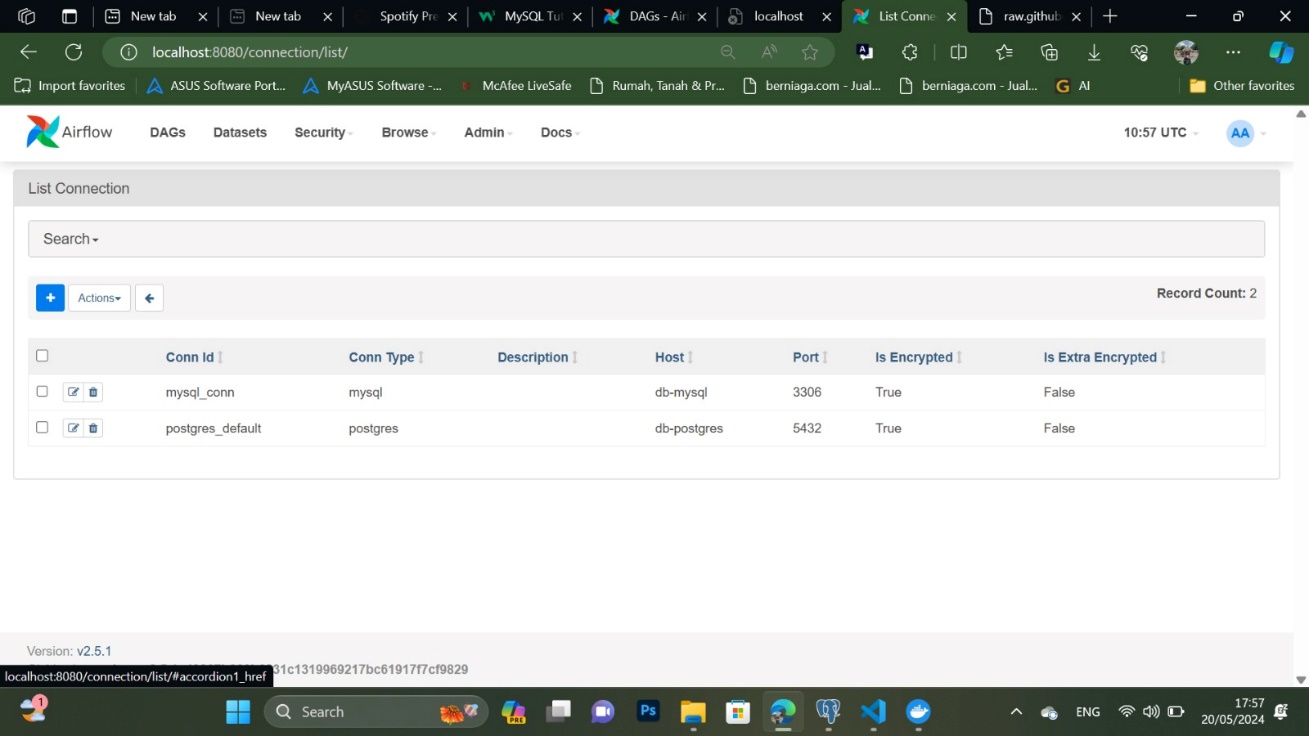
1. Jalankan docker database mysql dan postgresql kemudian lampirkan bukti:
   * 1. Screenshoot terminal



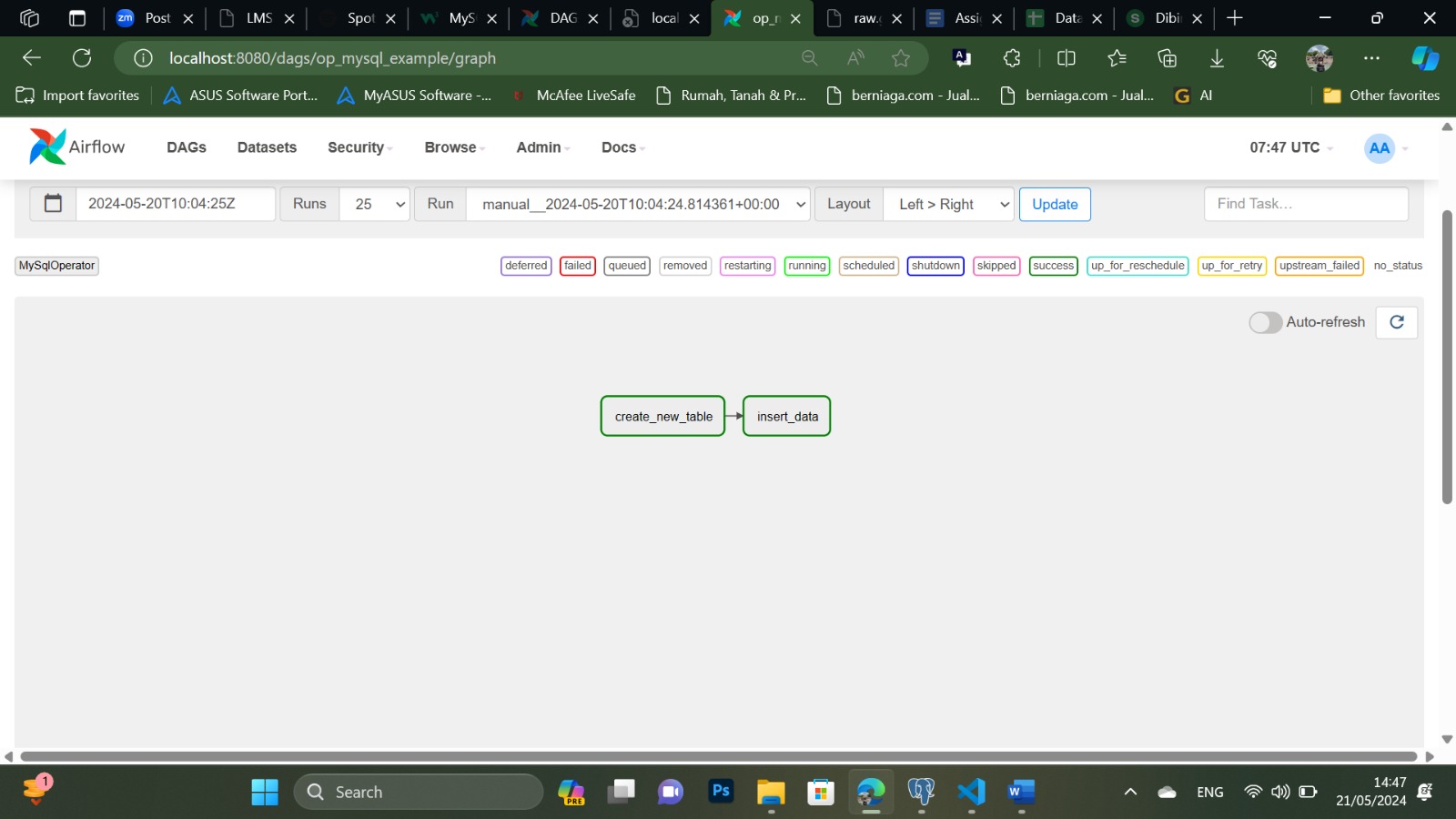
* + 1. Screenshoot Dbeaver / MySQL Connection Extension yang telah terhubung dengan postgresql dan mysql



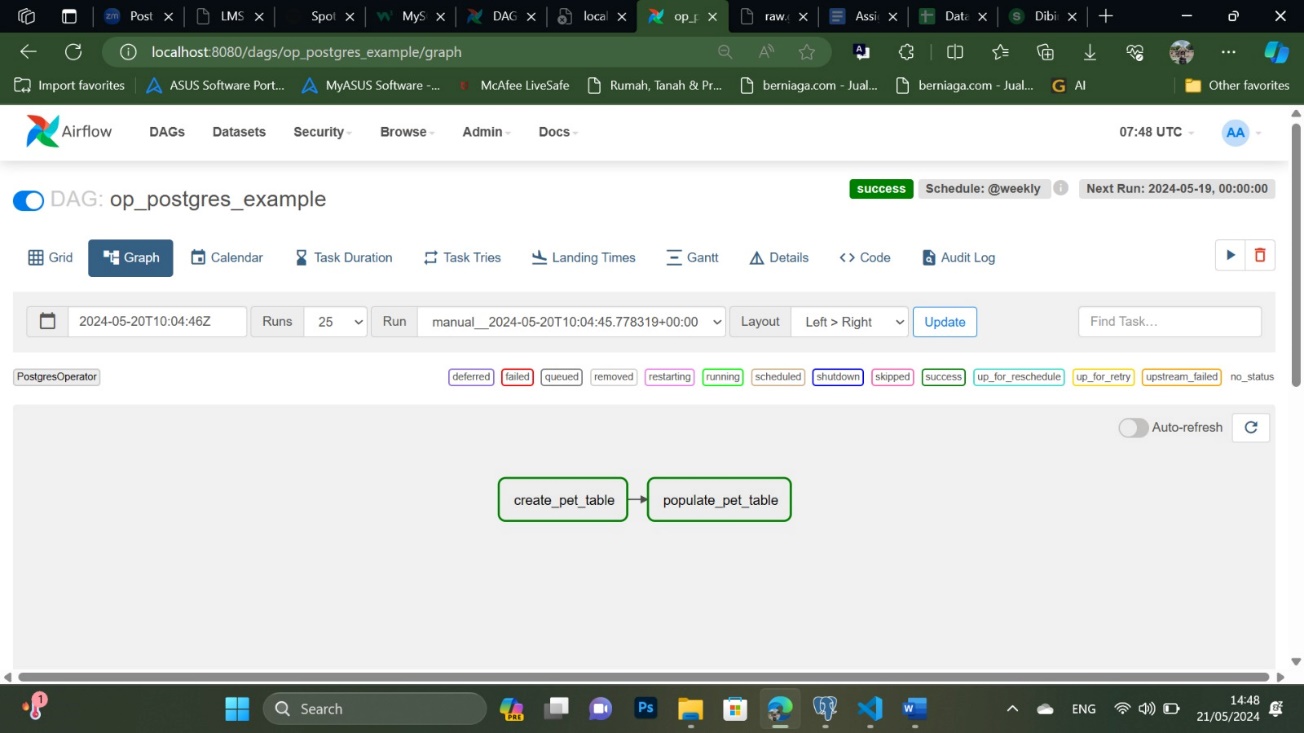
1. Buat connection MySQL dan PostgreSQL pada Apache Airflow kemudian
   * 1. Screenshoot UI hasil testing yang menunjukkan connection successful



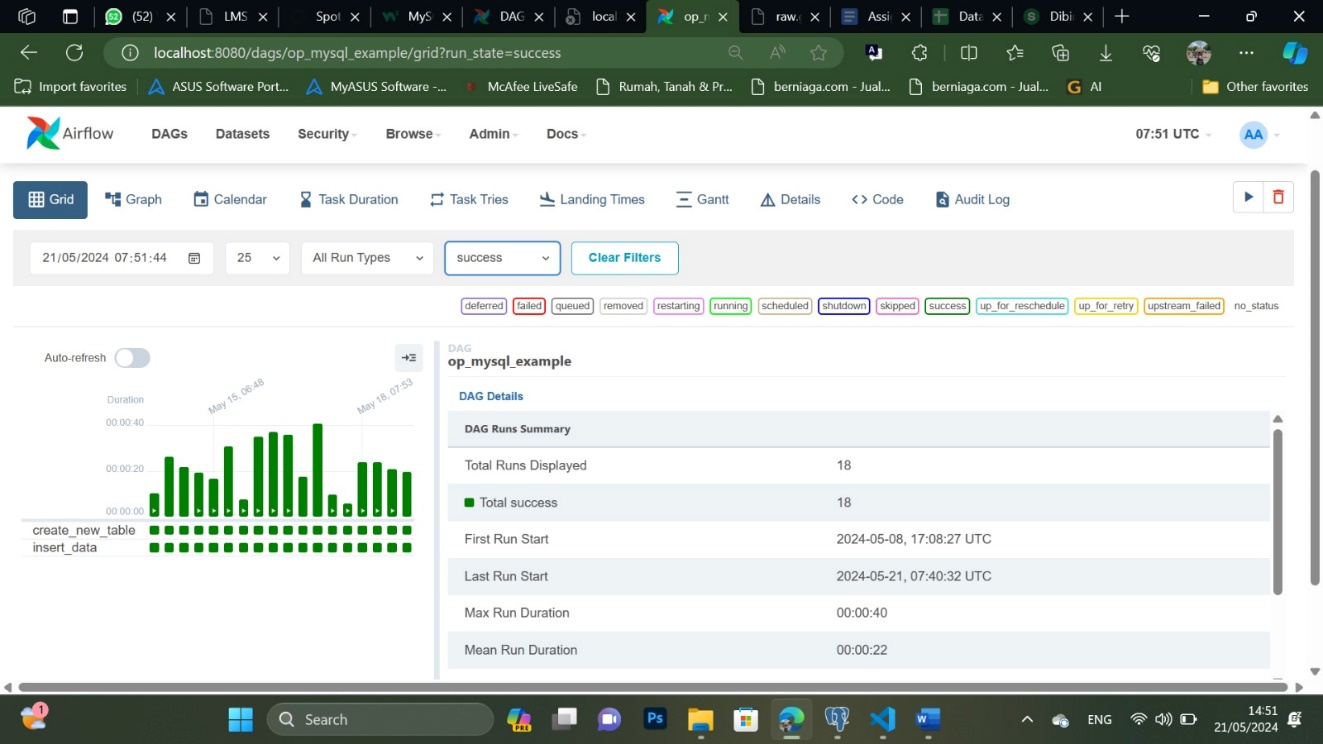
1. Modifikasi DAGs op\_mysql  dan op\_postgresql yang sudah ada dengan menambahkan EmptyOperator sehingga tampilan kedua DAG tersebut menjadi sebagai berikut:
   * 1. op\_mysql :

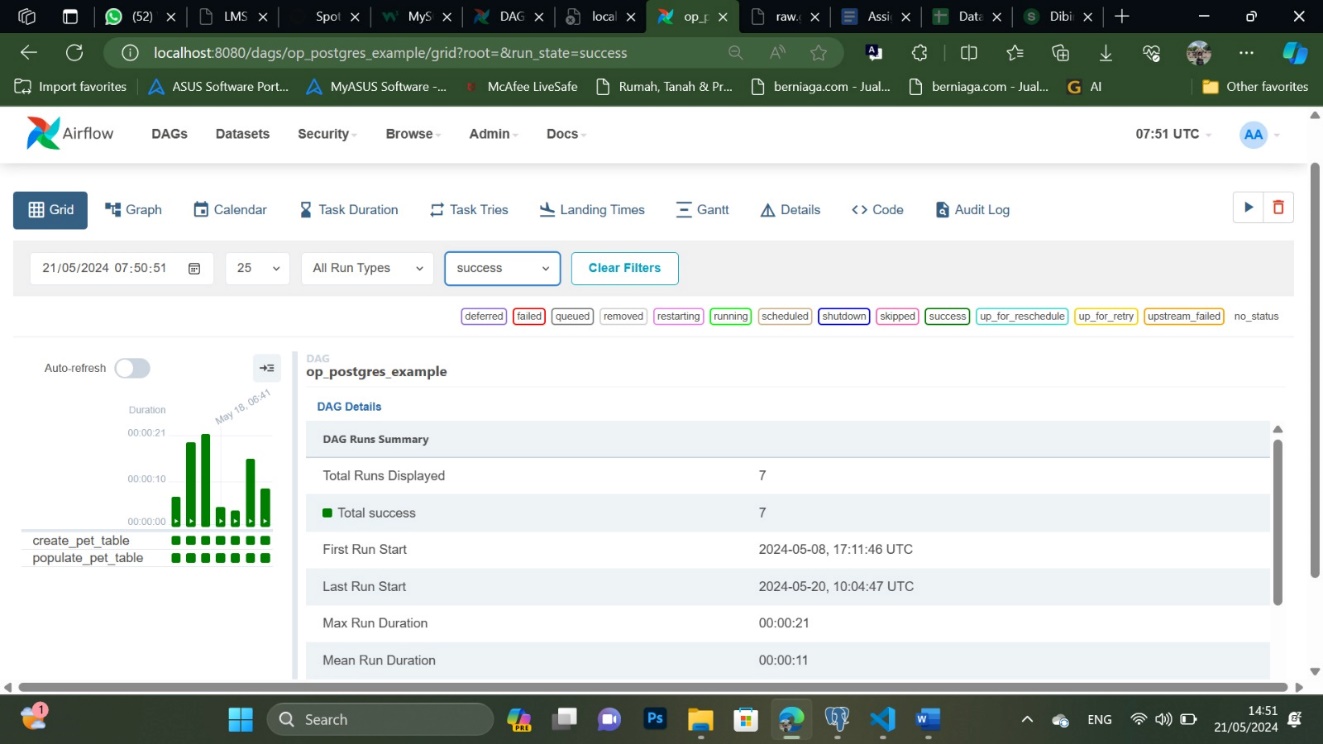


1. op\_postgresql



1. Jalankan DAGs op\_mysql dan op\_postgresql hingga success kemudian lampirkan hasil screenshotnya. Usahakan sampai pipeline berjalan success semua





1. Tampilkan hasil table yang dibuat oleh op\_mysql dan op\_postgresql
2. Jelaskan apa yang dikerjakan oleh op\_mysql dan op\_postgresql? Apa perbedaan sebelum dan sesudah dijalankannya DAG?

Perbedaan Sebelum dan Sesudah DAG Dijalankan:

Sebelum DAG dijalankan, perintah SQL hanya didefinisikan dalam kode.

Setelah DAG dijalankan, operator akan mengeksekusi perintah SQL di database MySQL sesuai dengan jadwal yang ditentukan.

Yang dikerjakan oleh op\_mysql dan op\_postgresql adalah

* Op\_mysql

Create\_new\_table, insert data,dan end

* Op\_postgresql

Create\_pet\_table, populate\_pet\_table, dan end

Cara Penggunaan:

Mirip dengan op\_mysql, Anda dapat memberikan perintah SQL langsung atau mengacu pada file eksternal.

Anda harus mengatur koneksi ke instance PostgreSQL menggunakan parameter op\_postgresql.

1. Buat Kesimpulan mengenai Apache Airflow

ringkasan tentang Apache Airflow:

Platform Open-Source: Apache Airflow adalah platform manajemen alur kerja sumber terbuka untuk pipeline rekayasa data.

DAGs & Tugas: Menggunakan DAG (Directed Acyclic Graph) untuk merepresentasikan alur kerja yang terdiri dari tugas-tugas yang diatur dengan mempertimbangkan ketergantungan dan aliran data.

Fleksibilitas & Ekstensibilitas: Memungkinkan pembuatan alur kerja yang dinamis dan dapat diperluas dengan kerangka kerja Python yang dapat terhubung dengan hampir semua teknologi.

Komponen Opsional: Termasuk komponen seperti worker opsional, triggerer, dan pengolah DAG untuk meningkatkan skalabilitas dan kinerja Airflow.

Airflow dirancang untuk memudahkan pengembangan, penjadwalan, dan pemantauan alur kerja yang berorientasi batch dengan antarmuka web yang membantu mengelola status alur kerja Anda