

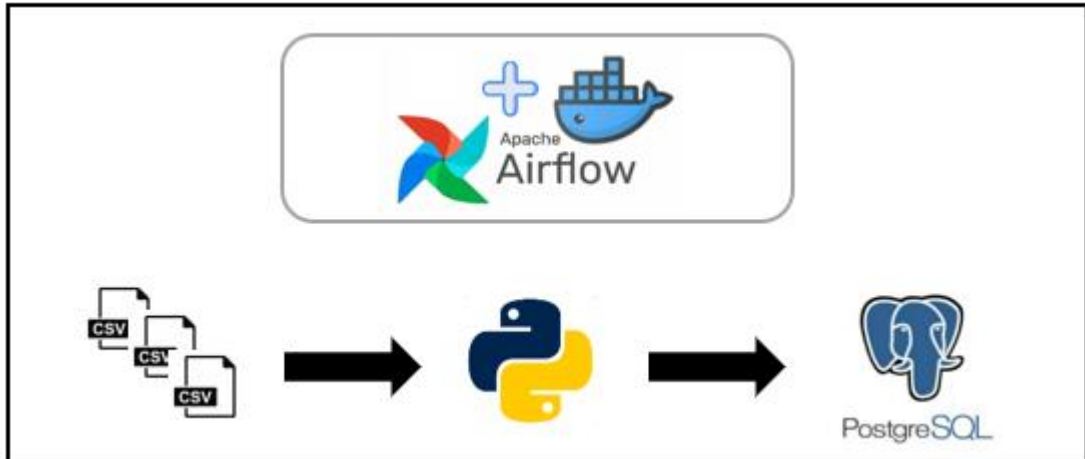
## Batch Processing :

### ETL Multiple File CSV to Database with Airflow

#### 1. Purpose

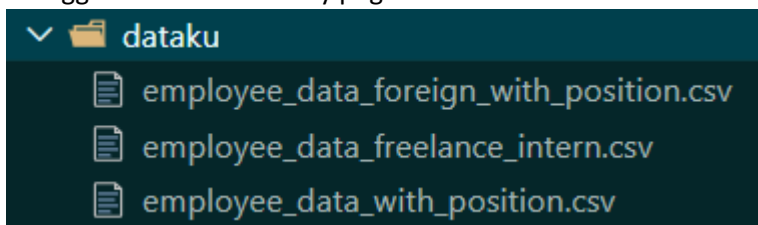
Mengextract semua file.csv dalam file dataku, menggabungkannya menjadi satu file.csv, mentransform data untuk menyelaraskan isi data yang kurang dari file sumber, kemudian meng-load data tersebut menjadi satu tabel dalam suatu database

#### 2. Architecture



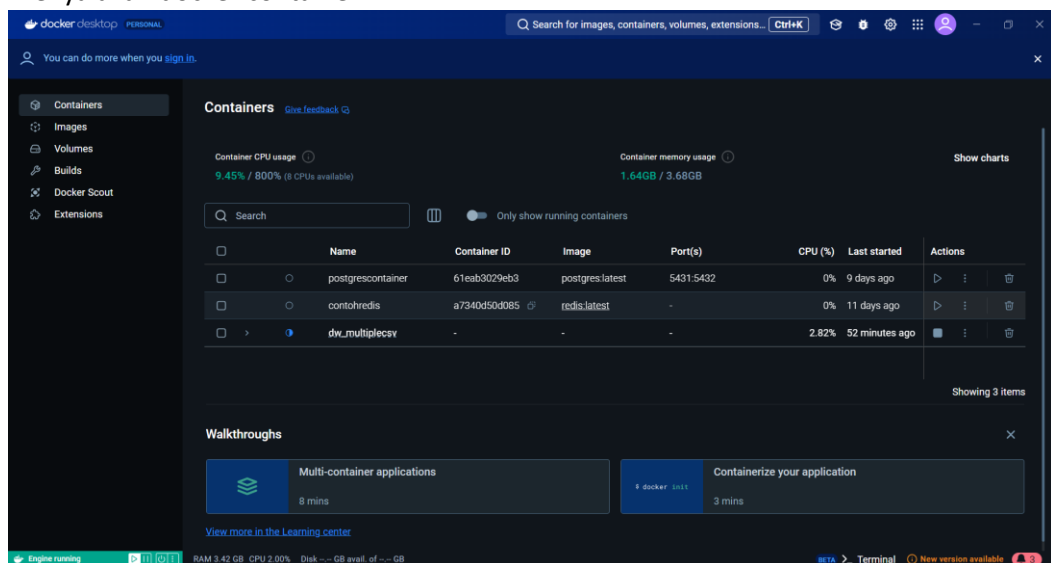
#### 3. Source data

Menggunakan data dummy pegawai PT.X

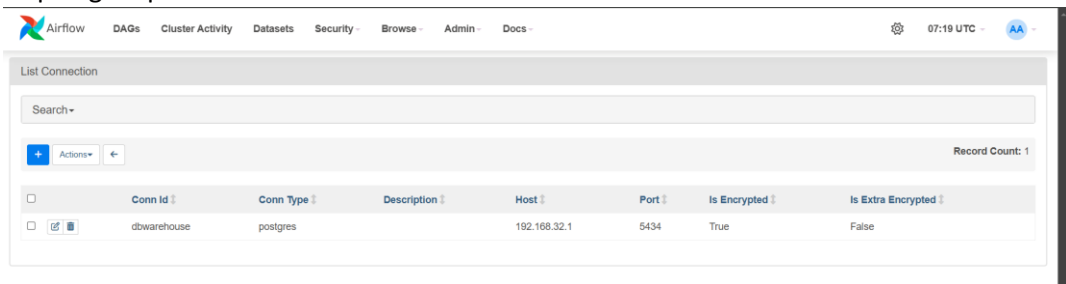


#### 4. Process

##### a. Menyalakan docker container

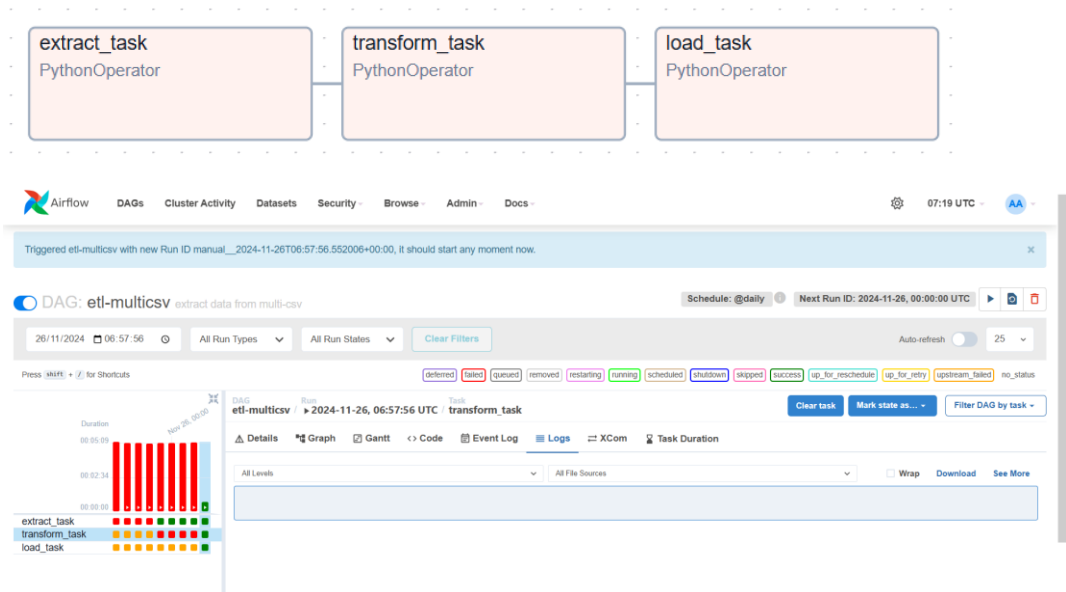


- b. Masuk ke Airflow, kemudian membuat connection airflow ke datawarehouse, dengan ini ke postgresql

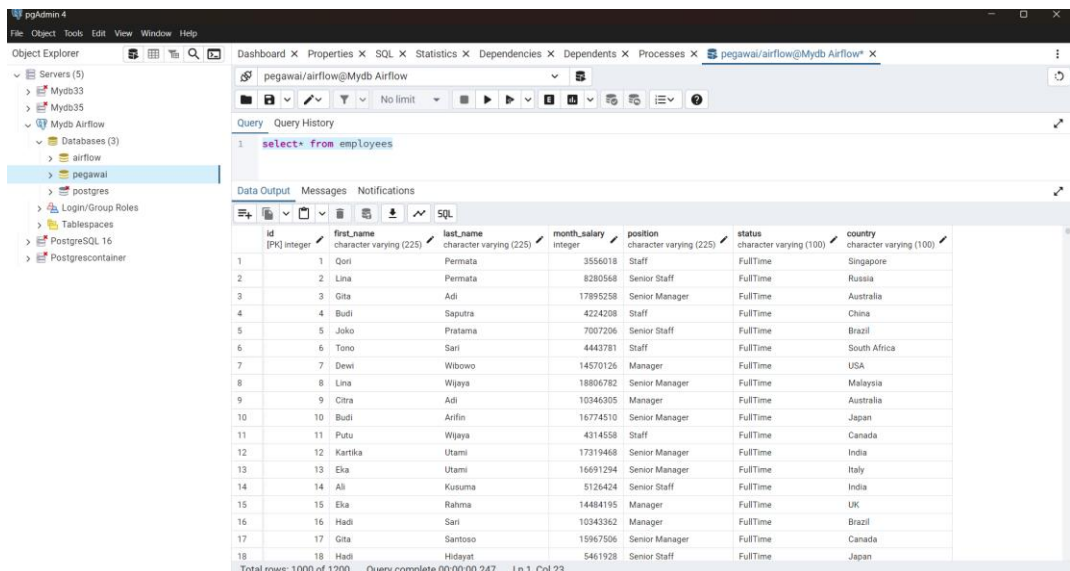


Conn Id	Conn Type	Description	Host	Port	Is Encrypted	Is Extra Encrypted
dbwarehouse	postgres		192.168.32.1	5434	True	False

- c. Menjalankan DAG



- d. Hasil



ID	first_name	last_name	month_salary	position	status	country
1	Qori	Permata	3556018	Staff	FullTime	Singapore
2	Lina	Permata	8280568	Senior Staff	FullTime	Russia
3	Gita	Adi	17895258	Senior Manager	FullTime	Australia
4	Budi	Saputra	4224208	Staff	FullTime	China
5	Joko	Pratama	7007206	Senior Staff	FullTime	Brazil
6	Tono	Sari	4443781	Staff	FullTime	South Africa
7	Dewi	Wibowo	14570126	Manager	FullTime	USA
8	Lina	Wijaya	18806782	Senior Manager	FullTime	Malaysia
9	Citra	Adi	10346305	Manager	FullTime	Australia
10	Budi	Arlin	16774510	Senior Manager	FullTime	Japan
11	Putu	Wijaya	4314558	Staff	FullTime	Canada
12	Kartika	Utami	17319468	Senior Manager	FullTime	Italy
13	Eka	Utami	16691294	Senior Manager	FullTime	India
14	Ali	Kusuma	5126424	Senior Staff	FullTime	India
15	Eka	Rahma	14484195	Manager	FullTime	UK
16	Hadi	Sari	10343362	Manager	FullTime	Brazil
17	Gita	Santoso	15967506	Senior Manager	FullTime	Canada
18	Hadi	Hidayat	5461928	Senior Staff	FullTime	Japan

- e. Analisis

Berdasarkan analisis pegawai PT.X, dapat diambil kesimpulan:

- Perbandingan jumlah pegawai local (Indonesia) dengan pegawai internasional adalah sama rata, yakni 50% : 50%
- Jumlah total pegawai PT.X adalah sebanyak 1.200 orang

- Pegawai dengan status “Full Time” mendominasi dengan 91.7% dari keseluruhan jumlah pegawai, yakni sebanyak 1.100 orang
- Pegawai dengan status Intern dan Freelance semuanya adalah pegawai local (pegawai Indonesia)
- Posisi Senior Manager mendominasi 30.4% dari keseluruhan jumlah pegawai, diikuti dengan Senior Staff 29%, Manajer 28,6% dan Staff 12%
- Gaji bulanan tertinggi adalah posisi Senior Manager dengan rata-rata 17 juta

Dashboard :

<https://lookerstudio.google.com/reporting/60a53a26-22b5-4f1b-86e8-b99853902297>

