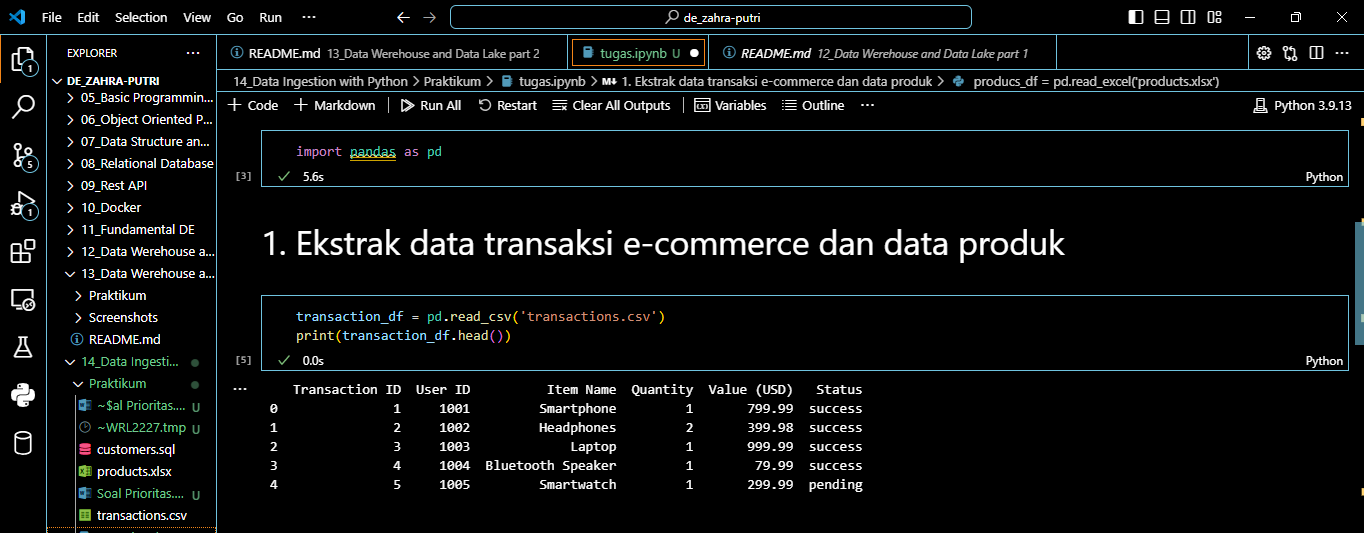
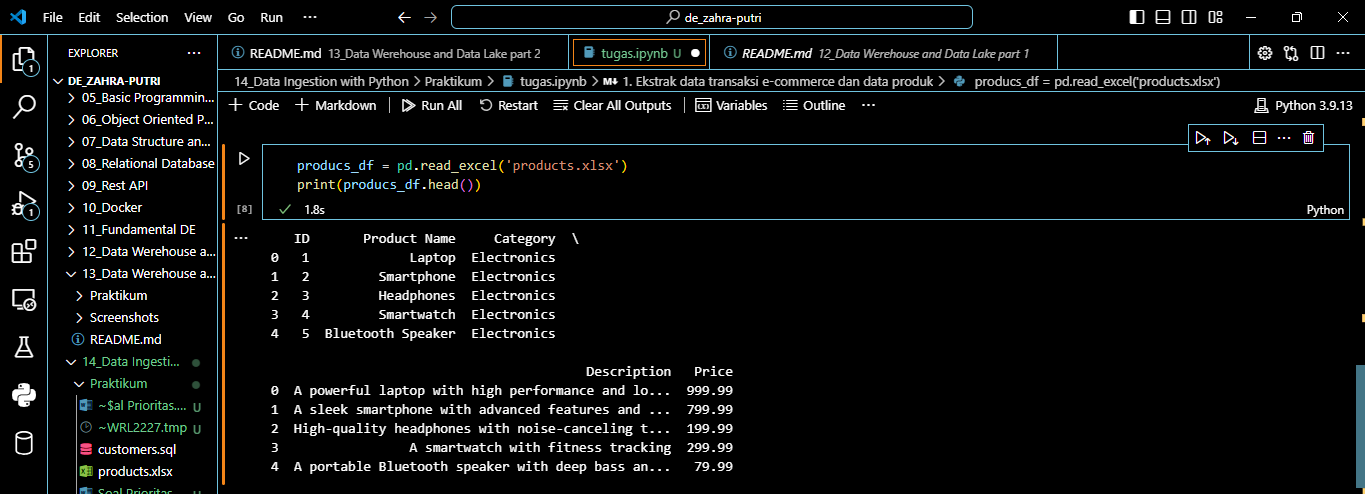
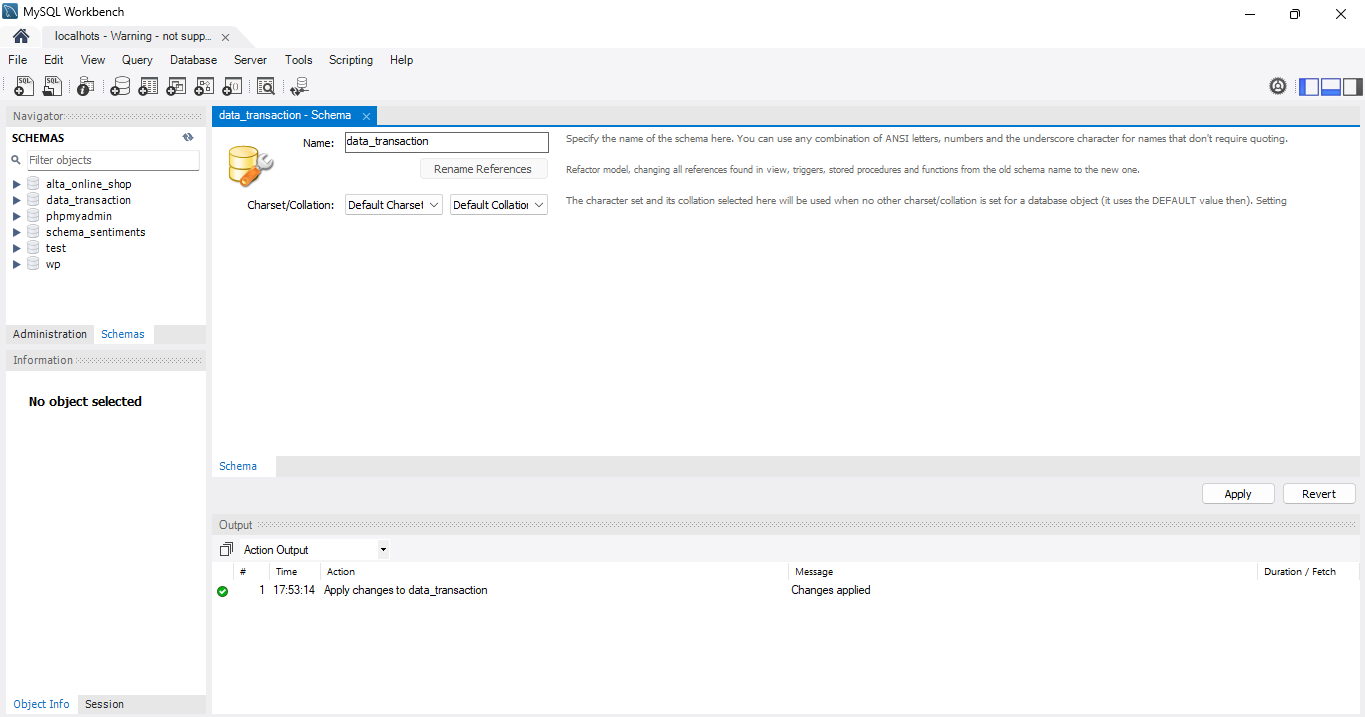
**Soal Prioritas 1**

1. Ekstrak data transaksi e-commerce dari file CSV dan data produk dari file Excel.





1. Gunakan pymysql untuk mengimpor data pelanggan dari database MySQL.



1. Ambil data pengiriman dari API berikut: https://fakestoreapi.com/products menggunakan requests.
2. Simpan semua data yang diingest ke dalam Pandas DataFrame dan ekspor ke file Parquet.

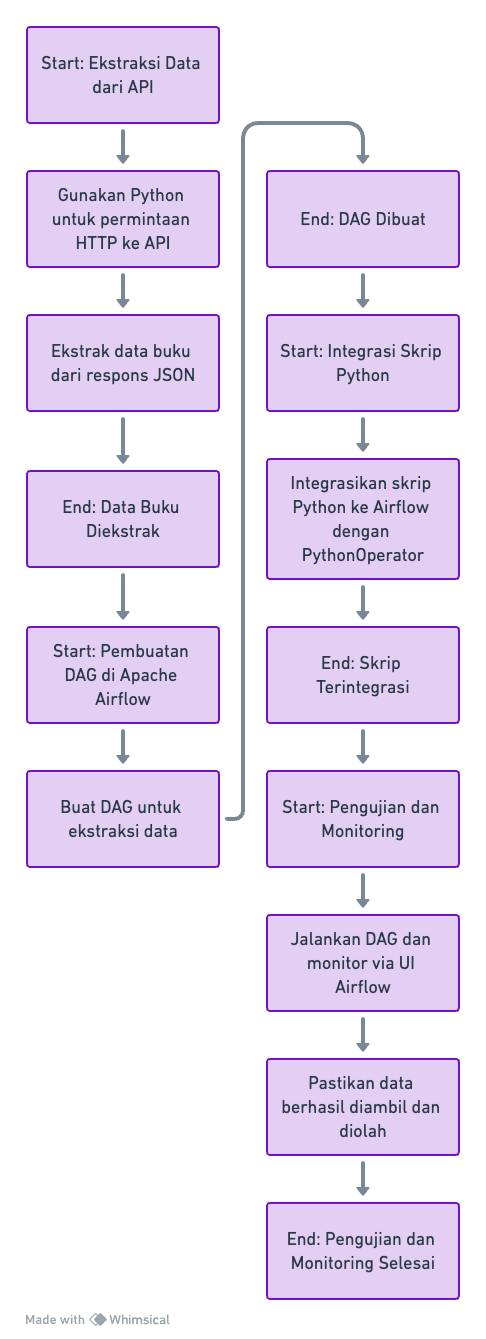
Catatan:

* Untuk soal nomor 1 gunakan dataset berikut dengan nama products dan transactions.csv: Dataset
* Untuk soal nomor 2, jalankan kode SQL yang telah disediakan dengan nama customers.sql untuk menambahkan data pelanggan ke dalam database MySQL.

**Soal Prioritas 2**

1. Persiapan Lingkungan Pengembangan:
2. Pastikan Python sudah terinstal di sistem Anda.
3. Install library yang diperlukan: requests untuk mengirim permintaan HTTP, dan beautifulsoup4 untuk parsing HTML.
4. Pemilihan Situs E-commerce dan Produk:
5. Pilih situs e-commerce yang akan di-scrape (misalnya, Amazon, Tokopedia, dll.).
6. Tentukan produk spesifik yang ulasannya ingin Anda kumpulkan.
7. Ekstraksi Data Ulasan:
8. Gunakan library requests untuk mengirim permintaan ke halaman produk.
9. Dengan BeautifulSoup, parse HTML yang diterima untuk mengekstrak informasi ulasan, seperti rating, teks ulasan, tanggal ulasan, dan nama pengguna.
10. Penanganan Pagination:
11. Jika ulasan produk terdapat di beberapa halaman, implementasikan logika untuk menelusuri setiap halaman.
12. Ekstrak data ulasan dari setiap halaman tersebut.
13. Pembersihan dan Penyimpanan Data:
14. Bersihkan data yang telah diekstrak dari tag HTML dan format yang tidak diperlukan.
15. Simpan data ulasan ke dalam format yang diinginkan, seperti CSV atau database.

**Soal Eksplorasi**



1. Ekstraksi Data dari API:
2. Gunakan Python untuk mengirim permintaan HTTP ke https://gist.githubusercontent.com/nadirbslmh/8922f71875948802321ef479a017f4c0/raw/1fbbc4b3d55f8ae717eb197d9aaf530ed1bc7ed2/sample.json dan terima respons dalam format JSON.
3. Ekstrak data buku yang relevan dari respons JSON, seperti judul, pengarang, tahun terbit, dan genre.
4. Pembuatan DAG di Apache Airflow:
5. Buat DAG di Apache Airflow untuk menjadwalkan dan mengotomatisasi proses ekstraksi data ini.
6. Tentukan jadwal eksekusi menggunakan cron expression, misalnya setiap minggu pada hari Senin jam 09:00.
7. Integrasi Skrip Python untuk Ekstraksi Data:
8. Integrasikan skrip Python yang telah Anda buat untuk ekstraksi data ke dalam task di Airflow menggunakan PythonOperator.
9. Pengujian dan Monitoring:
10. Jalankan DAG dan monitor prosesnya melalui UI Airflow.
11. Pastikan data berhasil diambil dan diolah sesuai jadwal yang ditentukan.

# Query to fetch customer data

query = "SELECT \* FROM customers"

# Execute the query and fetch data into a DataFrame

customers\_df = pd.read\_sql(query, connection)

# Display the first few rows of the customers DataFrame

print(customers\_df.head())

# Close the database connection

connection.close()

```

### 4. Ambil Data Pengiriman dari API

Gunakan library requests untuk mengambil data pengiriman dari API fakestoreapi.com.

```python

import requests

# API endpoint for shipping data

api\_url = 'https://fakestoreapi.com/products'

# Send GET request to the API

response = requests.get(api\_url)

# Convert JSON response to DataFrame

shipping\_df = pd.DataFrame(response.json())

# Display the first few rows of the shipping DataFrame

print(shipping\_df.head())

```

### 5. Simpan Data dalam Pandas DataFrame dan Ekspor ke File Parquet

Terakhir, simpan semua data yang telah diingest ke dalam Pandas DataFrame dan ekspor ke file Parquet.

```python

# Combine all DataFrames into a single DataFrame

combined\_df = pd.concat([transactions\_df, products\_df, customers\_df, shipping\_df], axis=1)

# Export combined DataFrame to Parquet file

combined\_df.to\_parquet('combined\_data.parquet')

```

Pastikan Anda telah mengganti placeholder seperti nama file dan informasi koneksi MySQL dengan nilai yang sesuai untuk lingkungan Anda. Semoga tutorial ini membantu! Jika ada pertanyaan lebih lanjut, jangan ragu untuk bertanya.