**LAPORAN UAS**

**DATA WAREHOUSE**

**Kelompok 4**

A logo of a building

Description automatically generated

Disusun Oleh:

**Arimbi Putri Hapsari 2341760016**

**Karina Ika Indasa 2341760042**

**Khuzaima Filla Januartha 2341760078**

**Paudra Akbar Buana 2341760063**

**Reza Angelina Febriyanti 2341760015**

**Jurusan Teknologi Informasi**

**[D4 Teknik Informatika/D4 Sistem Informasi Bisnis]**

**Politeknik Negeri Malang**

**2024**

1. **Cari Studi Kasus Pengelolaan Data**

* Link Dataset:

[**https://www.kaggle.com/datasets/andrexibiza/grocery-sales-dataset**](https://www.kaggle.com/datasets/andrexibiza/grocery-sales-dataset)

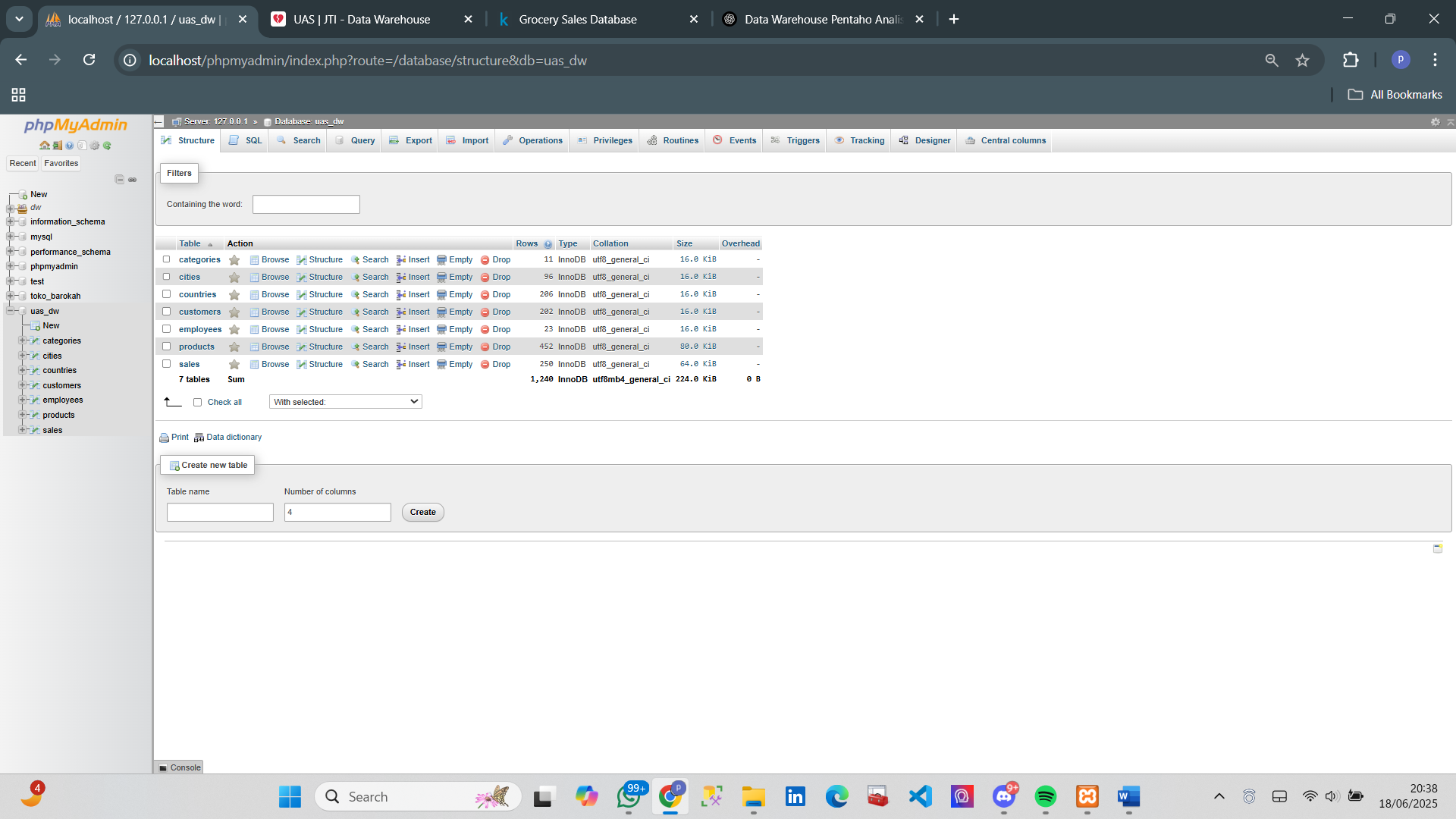
Dalam Dataset diatas telah memenuhi kriteria:  
 - Data Karyawan = employees  
 - Data Penjualan = sales  
 - Data Inventaris = products  
 - Data Pelanggan = customers

1. **Pilih Studi Kasus yang Memenuhi Kriteria**

* Studi Kasus 1:  
  Melakukan pengambilan data dari tiga tabel utama dalam data warehouse, yaitu *Facttotal, Dim\_Date,Dim\_Employees*, Dan *Dim\_Customers* untuk menyusun laporan penjualan karyawan. Laporan ini bertujuan menampilkan total penjualan selama tahun 2017 yang dilakukan oleh karyawan dengan EmployeeID = 22. Hasil ini digunakan sebagai dasar evaluasi kinerja individual dalam mendukung pengambilan keputusan manajerial.
* Studi Kasus 2:   
  Melakukan pengambilan data dengan memanfaatkan data dari tiga tabel utama dalam data warehouse, yaitu *FactQty, Dim\_Date,* dan *Dim\_Customers*, kita dapat menyusun laporan jumlah pembelian oleh customer dengan CustomerID = 87 selama tahun 2017. Laporan ini digunakan untuk mengevaluasi nilai dan frekuensi transaksi dari pelanggan tersebut sehingga dapat dijadikan dasar pengambilan keputusan strategis bagi tim penjualan dan manajemen.

1. **Ambil Data Dummy dari Studi Kasus Terpilih**

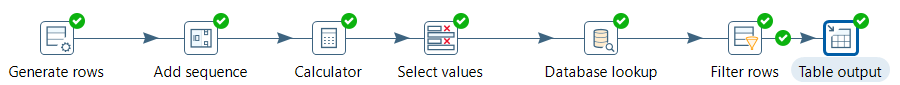
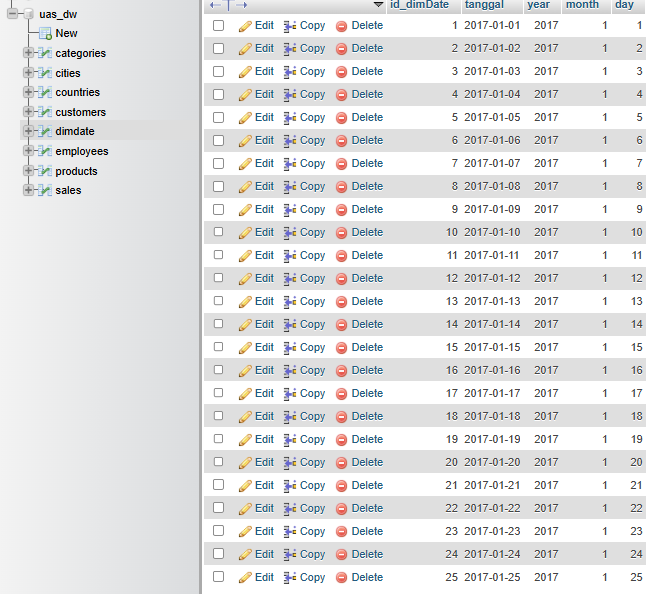
* Database:

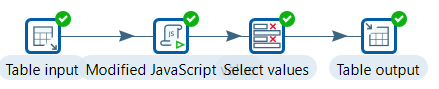


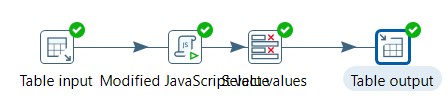
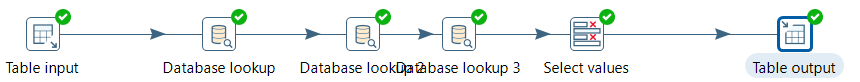
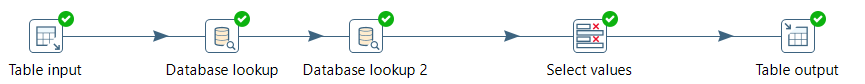
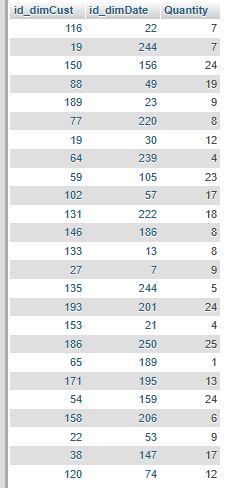
Impor data dummy yang bertipe (csv) dan buat database setelah itu masukkan tiap file csv menjadi 1 tabel. Dan pada dataset saya berisi 7 file/table yaitu:

* categories.csv
* cities.csv
* countries.csv
* customers.csv
* employees.csv
* products.csv
* sales.csv

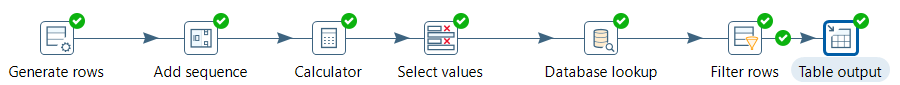
1. **Rancang Skema Bintang (Star Schema) untuk Data Warehouse**

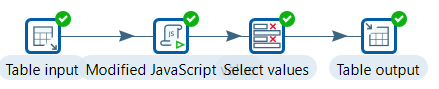
* Dimensi Waktu  
  ****
* ****
* Dimensi Pegawai

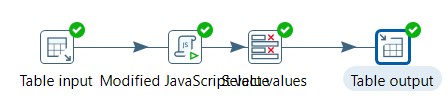
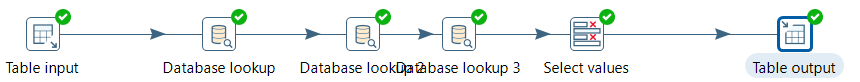
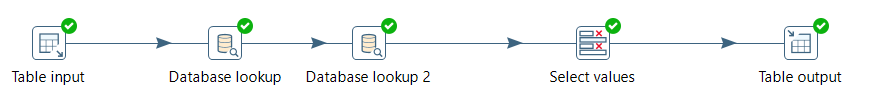
**  
**

* Dimensi Customer  
  ****  
  ****
* Fakta Total  
  ****  
  ****
* Fakta QTY  
  ****  
  ****

1. **Rancang Proyek ETL Berdasarkan Skema Bintang**

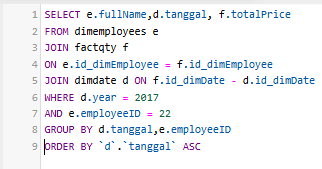
* **Extract**: Mengambil data dari sumber data dummy (file CSV, database, API, dsb).
* **Transform**: Membersihkan data, mengubah format data, konversi tipe data, agregasi, dan transformasi lainnya sesuai kebutuhan.
* **Load**: Memasukkan data yang sudah bersih dan siap ke dalam tabel fakta dan dimensi di data warehouse.
* Dimensi Waktu  
  ****
* Dimensi Pegawai

****

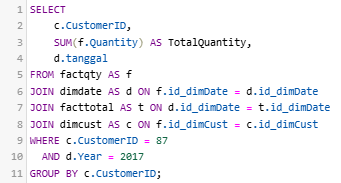
* Dimensi Customer  
  ****
* Fakta Total  
  ****
* Fakta QTY  
  

1. **Jelaskan Langkah-Langkah Proses ETL**

* **Extract**: Membaca data dummy dari sumber, misalnya membaca file CSV berisi data karyawan atau data lain.
  + Membuat table baru Bernama uas\_dw, setelah itu Impor data dummy yang telah di download pada Kaggle juga bertipe (csv) setelah itu masukkan tiap file csv menjadi 1 tabel
* **Transform**: Melakukan pembersihan data seperti memperbaiki format tanggal, normalisasi nama atau kategori, menghapus data duplikat, menyesuaikan struktur data sesuai skema star schema.
  + Memperbaiki tipe data yang berbeda pada tiap table sebelum membuat table dimensi, dan disamakan nantinya Ketika ingin membuat table dimensi
* **Load**: Memasukkan data ke tabel fakta dan dimensi dengan memperhatikan integritas data, kunci primer, dan relasi antar tabel.
  + Mengambil id pada tiap table nantinya pada table dimensi yang ingin dibuat, lalu beberapa value yang ingin diambil sesuai dengan studi kasus yang ada diatas.

1. **Studi Kasus**
2. Query:  
   ****

Output:  
****

1. Query:  
   ****

Output:  
****