

LAPORAN UAS
DATA WAREHOUSE

Kelompok 4



Disusun Oleh:

Arimbi Putri Hapsari	2341760016
Karina Ika Indasa	2341760042
Khuzaima Filla Januardha	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>

Dalam Dataset diatas telah memenuhi kriteria:

- Data Karyawan = **employees**
- Data Penjualan = **sales**
- Data Inventaris = **products**
- Data Pelanggan = **customers**

2. 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.

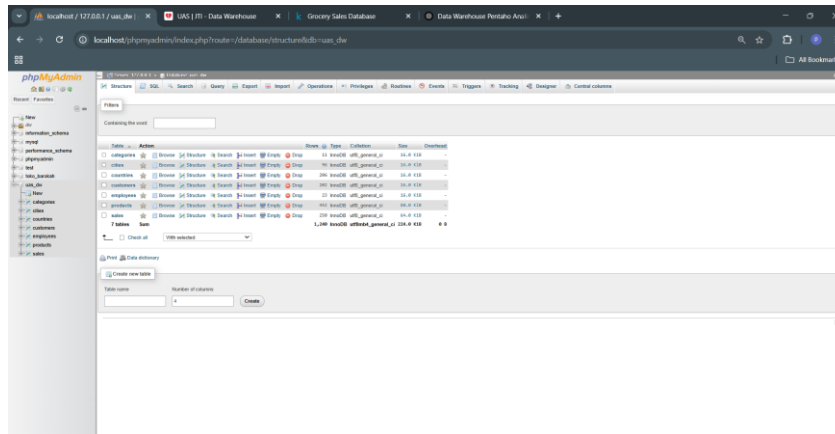
Hasil ini digunakan sebagai dasar evaluasi kinerja individual dalam mendukung pengambilan keputusan manajerial.

- 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.

3. Ambil Data Dummy dari Studi Kasus Terpilih

- Database:

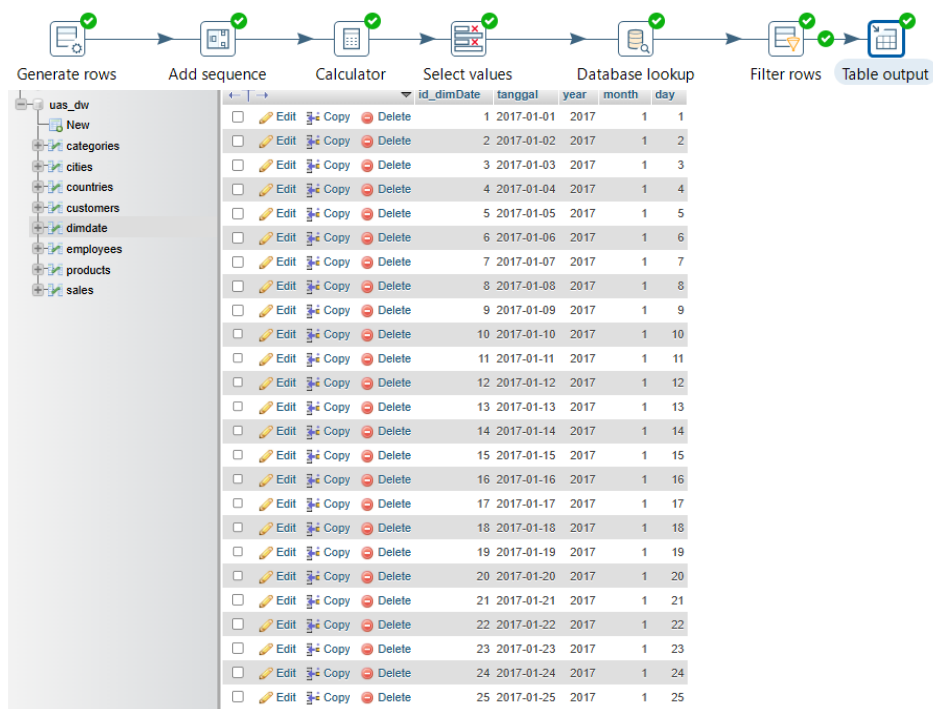


Impor data dummy yang bertipe (csv) dan buat database setelah itu masukkan tiap file csv menjadi 1 tabel. Dan pada dataset saya berisi 7 file/table yaitu:

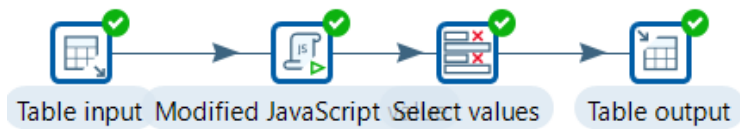
- categories.csv
- cities.csv
- countries.csv
- customers.csv
- employees.csv
- products.csv
- sales.csv

4. Rancang Skema Bintang (Star Schema) untuk Data Warehouse

- Dimensi Waktu

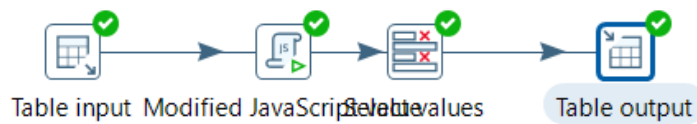


- Dimensi Pegawai



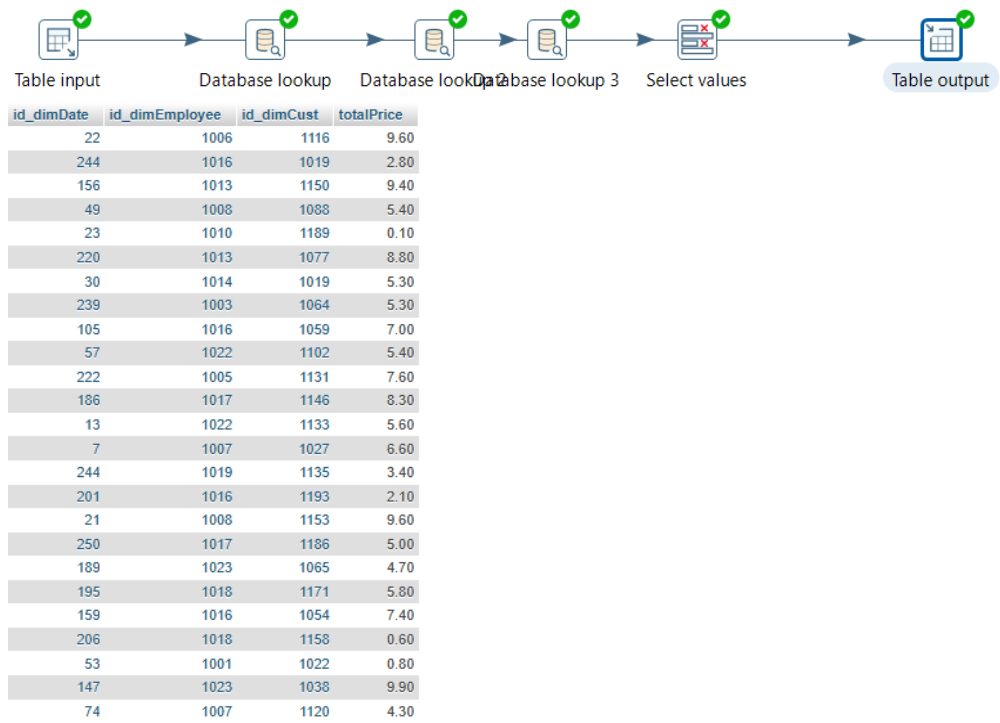
		id_dimEmployee	employeeID	fullName
<input type="checkbox"/>	Edit Copy Delete		1001	1 Nicole Fuller
<input type="checkbox"/>	Edit Copy Delete		1002	2 Christine Palmer
<input type="checkbox"/>	Edit Copy Delete		1003	3 Pablo Cline
<input type="checkbox"/>	Edit Copy Delete		1004	4 Darnell Nielsen
<input type="checkbox"/>	Edit Copy Delete		1005	5 Desiree Stuart
<input type="checkbox"/>	Edit Copy Delete		1006	6 Holly Collins
<input type="checkbox"/>	Edit Copy Delete		1007	7 Chadwick Cook
<input type="checkbox"/>	Edit Copy Delete		1008	8 Julie Dyer
<input type="checkbox"/>	Edit Copy Delete		1009	9 Daphne King
<input type="checkbox"/>	Edit Copy Delete		1010	10 Jean Vang
<input type="checkbox"/>	Edit Copy Delete		1011	11 Sonya Dickson
<input type="checkbox"/>	Edit Copy Delete		1012	12 Lindsay Chen
<input type="checkbox"/>	Edit Copy Delete		1013	13 Katina Marks
<input type="checkbox"/>	Edit Copy Delete		1014	14 Wendi Buckley
<input type="checkbox"/>	Edit Copy Delete		1015	15 Kari Finley
<input type="checkbox"/>	Edit Copy Delete		1016	16 Chadwick Walton
<input type="checkbox"/>	Edit Copy Delete		1017	17 Seth Franco
<input type="checkbox"/>	Edit Copy Delete		1018	18 Warren Bartlett
<input type="checkbox"/>	Edit Copy Delete		1019	19 Bernard Moody
<input type="checkbox"/>	Edit Copy Delete		1020	20 Shelby Riddle
<input type="checkbox"/>	Edit Copy Delete		1021	21 Devon Brewer
<input type="checkbox"/>	Edit Copy Delete		1022	22 Tonia Mc Millan
<input type="checkbox"/>	Edit Copy Delete		1023	23 Janet Flowers

- Dimensi Customer

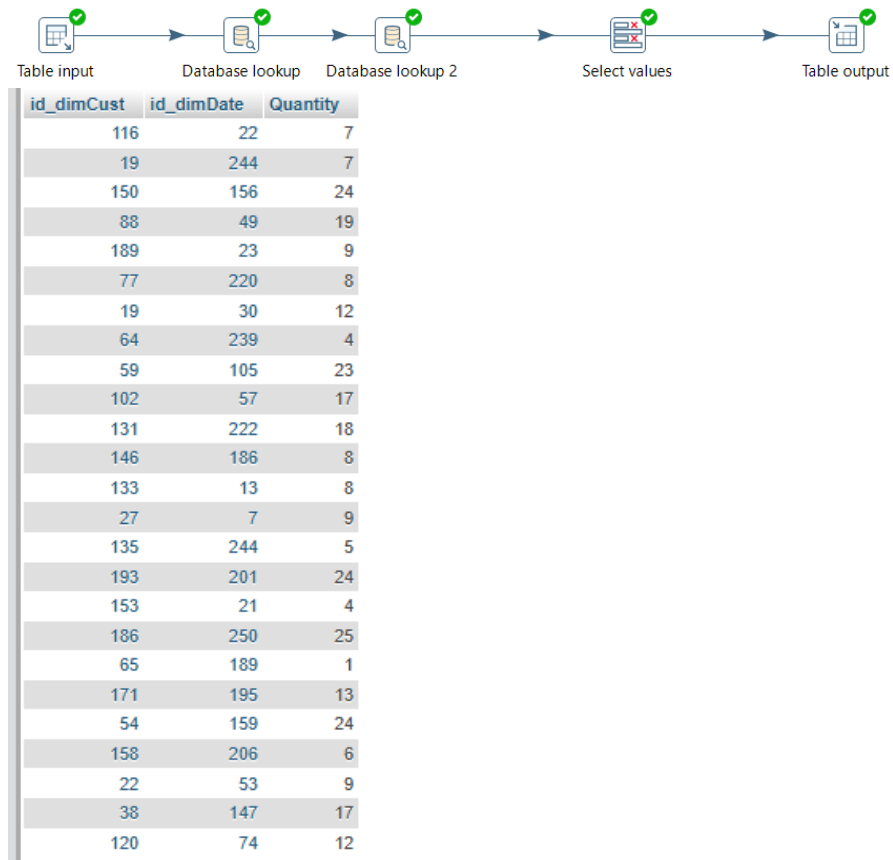


		id_dimCust	customerID	fullName
<input type="checkbox"/>	Edit Copy Delete		1001	0 FirstName LastName
<input type="checkbox"/>	Edit Copy Delete		1002	1 Stefanie Frye
<input type="checkbox"/>	Edit Copy Delete		1003	2 Sandy Kirby
<input type="checkbox"/>	Edit Copy Delete		1004	3 Lee Zhang
<input type="checkbox"/>	Edit Copy Delete		1005	4 Regina Avery
<input type="checkbox"/>	Edit Copy Delete		1006	5 Daniel Mccann
<input type="checkbox"/>	Edit Copy Delete		1007	6 Dennis Zuniga
<input type="checkbox"/>	Edit Copy Delete		1008	7 Myra Jefferson
<input type="checkbox"/>	Edit Copy Delete		1009	8 Teddy Parsons
<input type="checkbox"/>	Edit Copy Delete		1010	9 Annie Stein
<input type="checkbox"/>	Edit Copy Delete		1011	10 Herman Pope
<input type="checkbox"/>	Edit Copy Delete		1012	11 Jenifer Winters
<input type="checkbox"/>	Edit Copy Delete		1013	12 Stefanie Gordon
<input type="checkbox"/>	Edit Copy Delete		1014	13 Chastity Gillespie
<input type="checkbox"/>	Edit Copy Delete		1015	14 Evelyn Sweeney
<input type="checkbox"/>	Edit Copy Delete		1016	15 Jeannie Gardner
<input type="checkbox"/>	Edit Copy Delete		1017	16 Rickey Ho
<input type="checkbox"/>	Edit Copy Delete		1018	17 Bobbie Simon
<input type="checkbox"/>	Edit Copy Delete		1019	18 Lesley Orr
<input type="checkbox"/>	Edit Copy Delete		1020	19 Shawna Herring
<input type="checkbox"/>	Edit Copy Delete		1021	20 Theresa Berg
<input type="checkbox"/>	Edit Copy Delete		1022	21 Tasha Heath
<input type="checkbox"/>	Edit Copy Delete		1023	22 Karla Sutton
<input type="checkbox"/>	Edit Copy Delete		1024	23 Otis Miller
<input type="checkbox"/>	Edit Copy Delete		1025	24 Alisa Ruiz

- Fakta Total



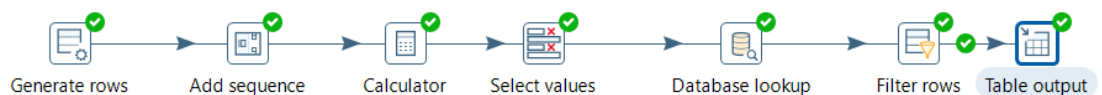
- Fakta QTY



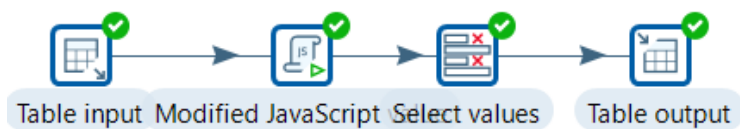
5. Rancang Proyek ETL Berdasarkan Skema Bintang

- **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.

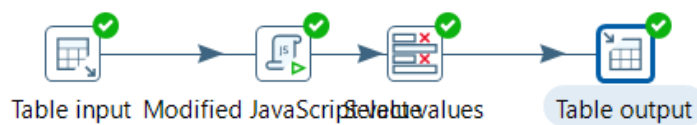
- Dimensi Waktu



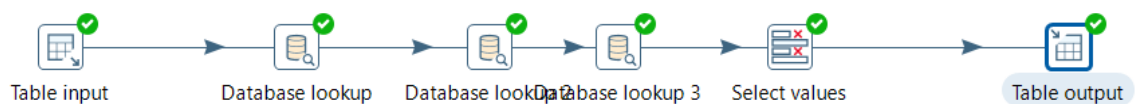
- Dimensi Pegawai



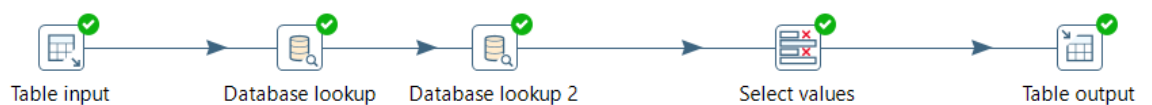
- Dimensi Customer



- Fakta Total



- Fakta QTY



6. 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.

7. Studi Kasus

1. Query:

```
1 SELECT e.fullName,d.tanggal, f.totalPrice
2 FROM dimemployees e
3 JOIN factqty f
4 ON e.id_dimEmployee = f.id_dimEmployee
5 JOIN dimdate d ON f.id_dimDate = d.id_dimDate
6 WHERE d.year = 2017
7 AND e.employeeID = 22
8 GROUP BY d.tanggal,e.employeeID
9 ORDER BY `d`.`tanggal` ASC
```

Output:

fullName	tanggal	totalPrice
Tonia Mc Millan	2017-01-01	5.40
Tonia Mc Millan	2017-01-02	5.40
Tonia Mc Millan	2017-01-03	5.40
Tonia Mc Millan	2017-01-04	5.40
Tonia Mc Millan	2017-01-05	5.40
Tonia Mc Millan	2017-01-06	5.40
Tonia Mc Millan	2017-01-07	5.40
Tonia Mc Millan	2017-01-08	5.40
Tonia Mc Millan	2017-01-09	5.40
Tonia Mc Millan	2017-01-10	5.40
Tonia Mc Millan	2017-01-11	5.40
Tonia Mc Millan	2017-01-12	5.40
Tonia Mc Millan	2017-01-13	5.40
Tonia Mc Millan	2017-01-14	5.40
Tonia Mc Millan	2017-01-15	5.40
Tonia Mc Millan	2017-01-16	5.40
Tonia Mc Millan	2017-01-17	5.40
Tonia Mc Millan	2017-01-18	5.40
Tonia Mc Millan	2017-01-19	5.40
Tonia Mc Millan	2017-01-20	5.40
Tonia Mc Millan	2017-01-21	5.40
Tonia Mc Millan	2017-01-22	5.40
Tonia Mc Millan	2017-01-23	5.40
Tonia Mc Millan	2017-01-24	5.40
Tonia Mc Millan	2017-01-25	5.40
Tonia Mc Millan	2017-01-26	5.40
Tonia Mc Millan	2017-01-27	5.40

2. Query:

```
1 SELECT
2     c.CustomerID,
3     SUM(f.Quantity) AS TotalQuantity,
4     d.tanggal
5 FROM factqty AS f
6 JOIN dimdate AS d ON f.id_dimDate = d.id_dimDate
7 JOIN facttotal AS t ON d.id_dimDate = t.id_dimDate
8 JOIN dimcust AS c ON f.id_dimCust = c.id_dimCust
9 WHERE c.CustomerID = 87
10 AND d.Year = 2017
11 GROUP BY c.CustomerID;
```

Output:

CustomerID	TotalQuantity	tanggal
87	19	2017-02-18