Laporan Praktikum Data Warehouse

Jobsheet 3 : Database Analytical

Dosen Pengampu : Vipkas Al Hadid Firdaus, ST., MT

Disusun Oleh:

Queenadhynar Azarine Dwipa A.

2341760109

SIB 2B

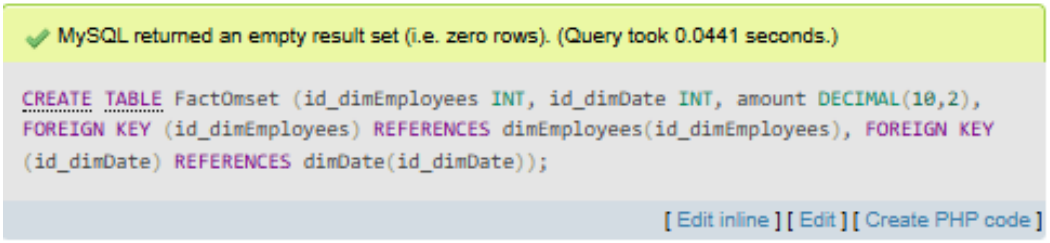
JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

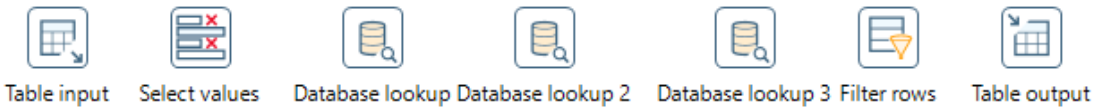
2023/2024

C. Fakta Pembayaran

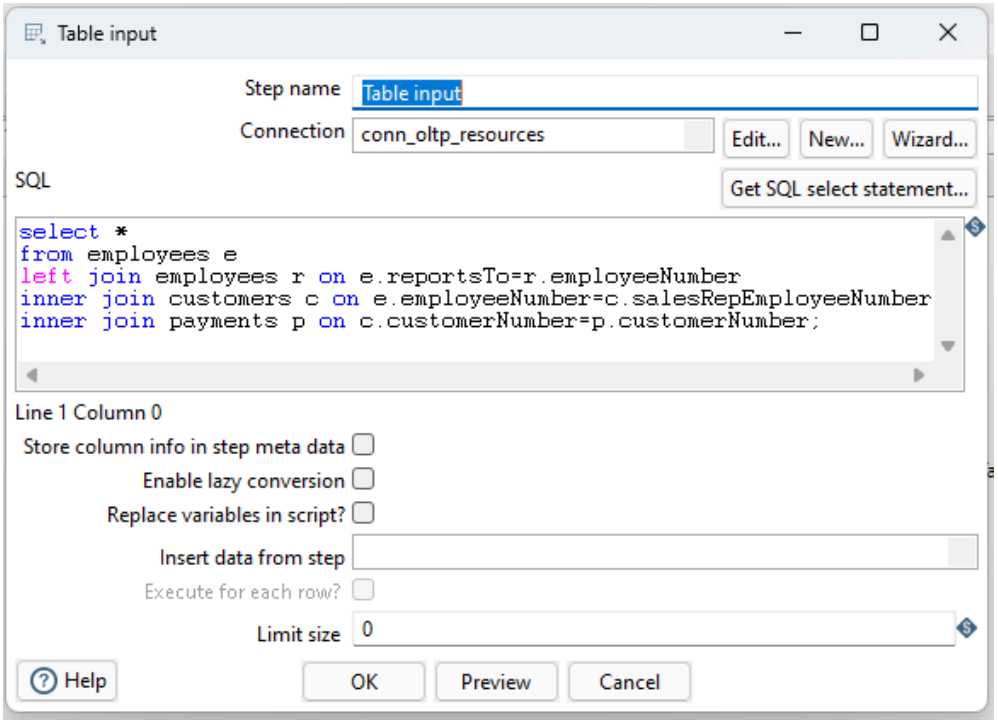
1. Buatlah sebuah tabel pada databse OLAP yang telah dibuat (database dw_legendVehicle).
   - nama tabel: FactOmset
   - Field: id_dimEmployees int FK tabel dimEmployee
   - Field: id_dimDate int FK tabel dimDate
   - Field: amount decimal(10.2)



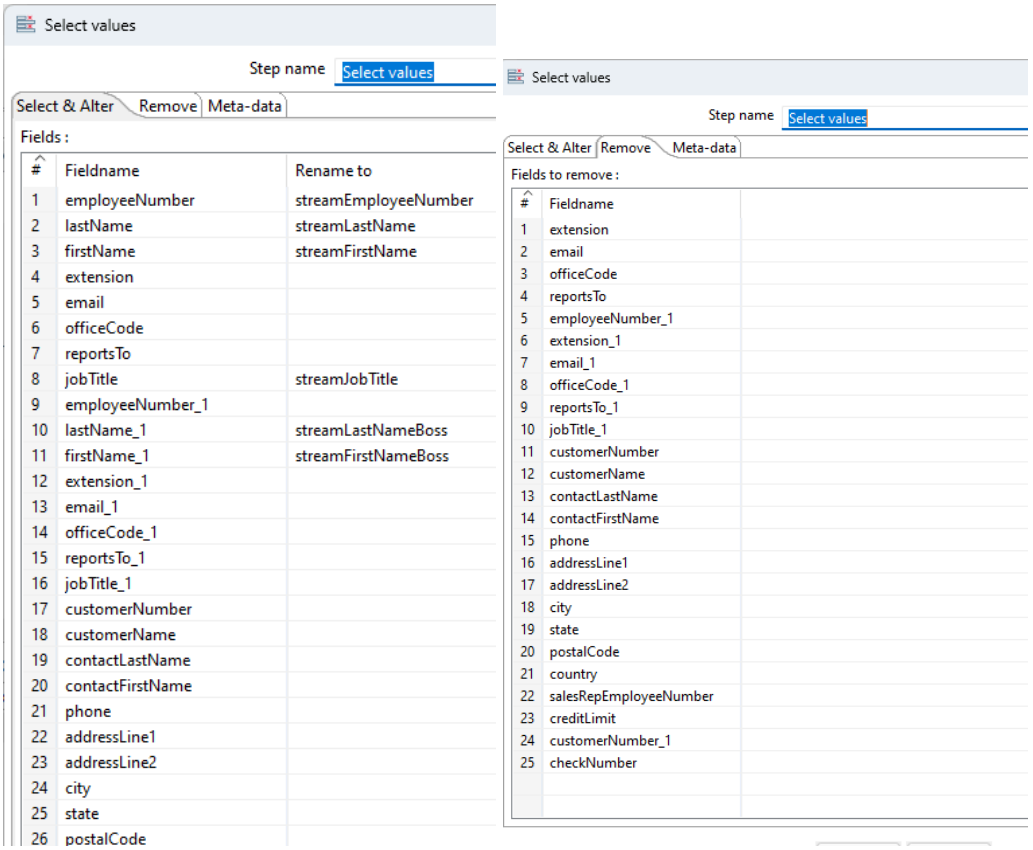
2. Buat Transformation baru pada PDI Spoon. Gunakan objevt-object yang dihubunugkan sebagai berikut:
   - Table Input: Dignakan untuk mengambil data transaksi dari OLTP
   - Select values: digunakan untuk memilih field yang akan digunakan untuk OLAP
   - Database lookup (1) : digunakan untuk mencocokkan data pada tabel dimEmployee untuk mengambil id_dimEmployee.
   - Database lookup (2) : digunakan untuk mencocokkan data pada tabel dimDate untuk mengambil id_dimDate.
   - Databse lookup (3): digunakan untuk mencocokkan data pada tabel factomset untuk melihat data yang sama atau tidak.
   - Filter rows: digunakan untuk memiih data yang sudah ada pada tabel factomset tidak dimasukkan lagi.
   - Table output: digunakan untuk memasukkan data pada tabel factOmset.



3. Konfigurasi pada Table input adalah untuk mendapatkan data dari OLTP sesuai dengan hasil dari query berikut. query tersebut akan menampilkan seluruh isi field dari employee hingga transaksi yang didapat dilihat dari payments yang didapat.

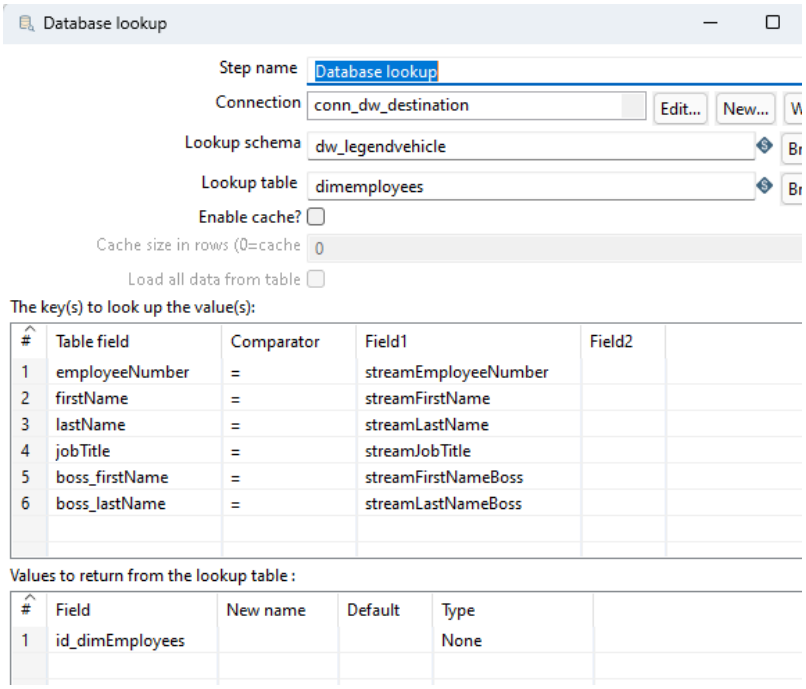


4. Konfigurasi pada Select Values adalah untuk menghapus semua field kecuali employeeNumber, lastname dari employee, firstname dari employee, jobTitle dari employee, lastname dari manager, firstname dari manager, payment date dan amount.

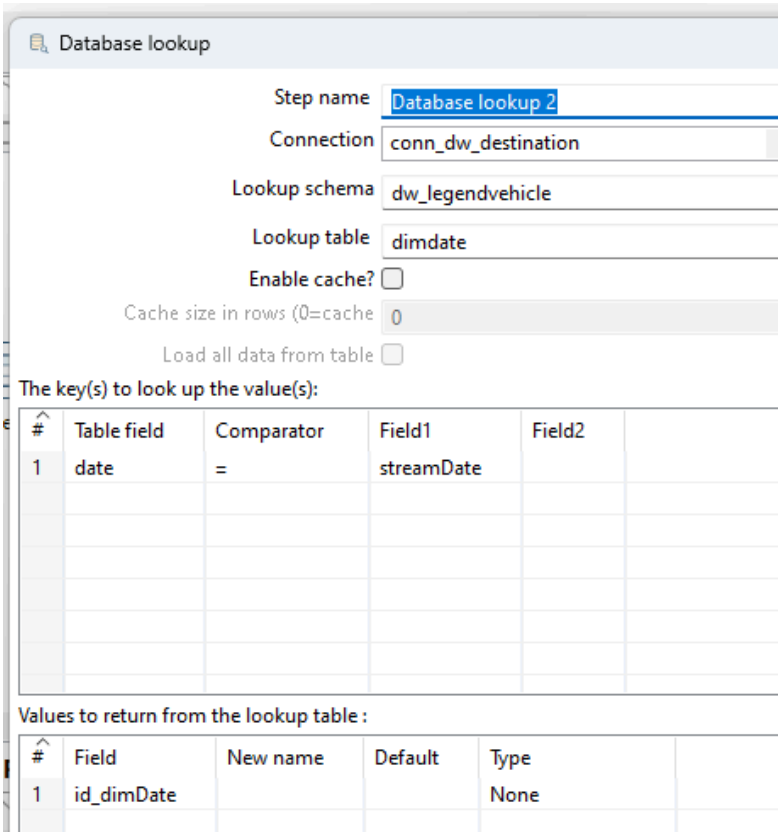


5. Konfigurasi pada tabel database lookup adalah dengan melakukan komparasi field stream (output dari proses sebelumnya) dengan field isi data pada tabel

dimEmployees. Jika data tersebut cocok maka akan diambil id_dimEmployees nya.

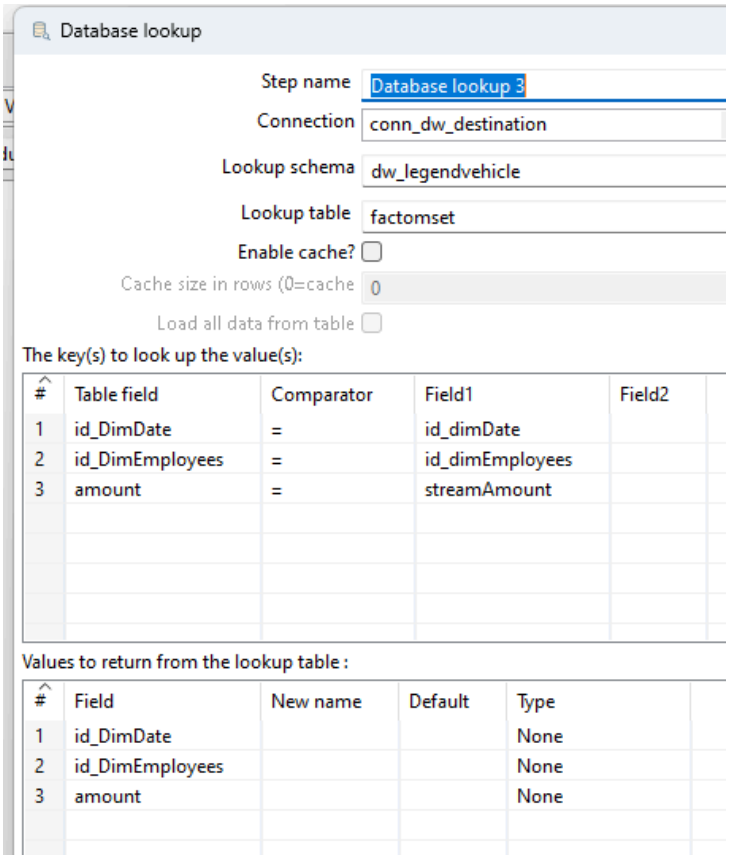


6. Konfigurasi pada tabel database lookup yang kedua adalah dengan melakukan komparasi field stream (output dari proses sebelumnya untuk field date) dengan field isi data pada tabel dimDate. Jika data tersebut cocok maka akan diambil id_dimDate nya.

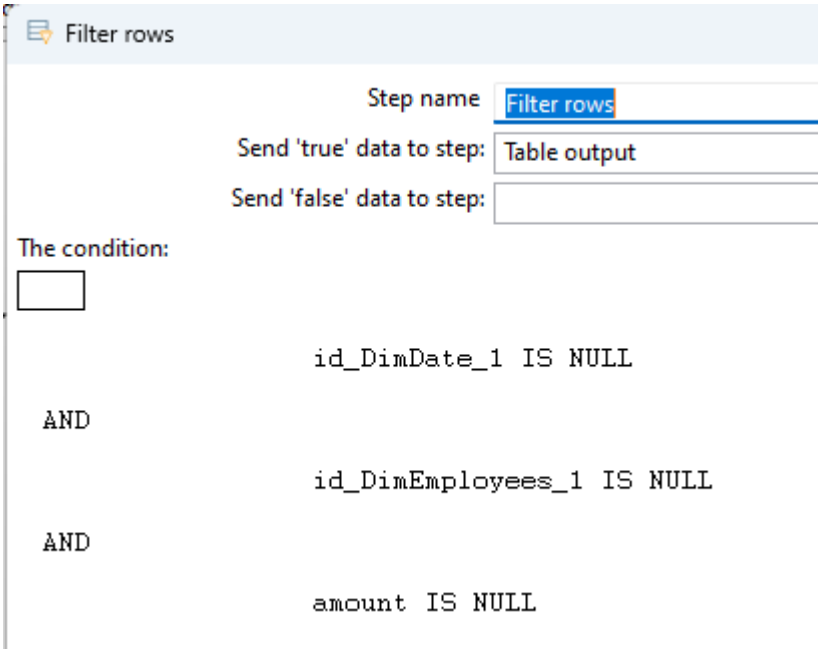


7. Konfigurasi pada tabel database lookup yang ketiga adalah dengan melakukan komparasi id_dimDate dan id_dimEmployees yang diambil dari proses lookup

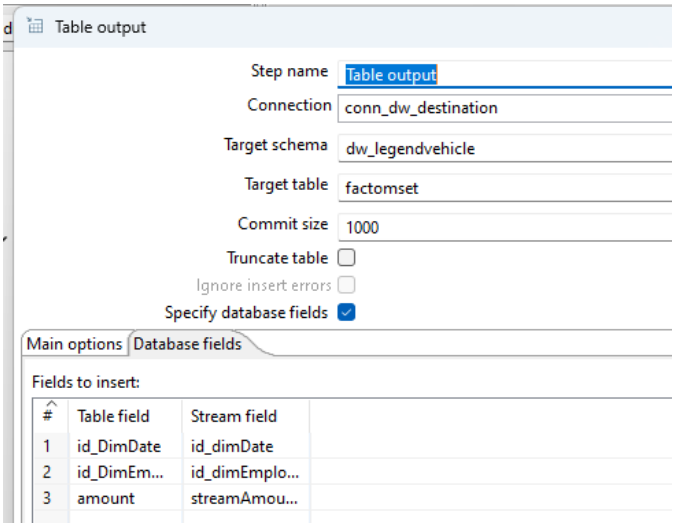
sebelumnya dengan field isi data pada tabel factOmset. Jika data tersebut cocok data tidak akan dimasukkan dalam tabel factOmset.



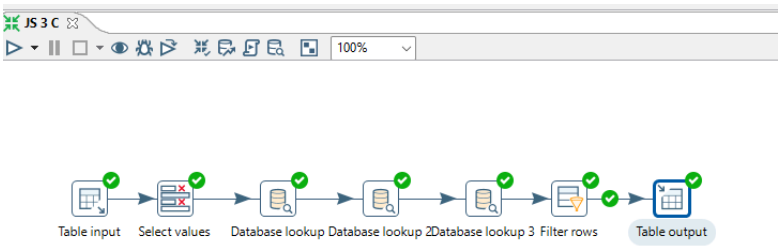
8. Pada bagian filter rows untuk melihat data belum ada pada tabel factOmset dengan melihat bahwa output dari lookup sebelumnya bernilai NULL.

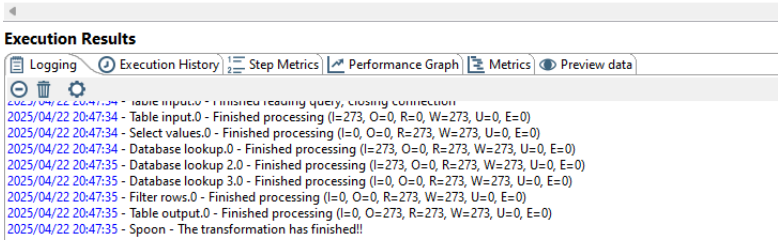


9. Konfigurasi terakhir pada table output adalah dengan melkukan mapping data output dari proses seblumnya kedalam field pada tabel factOmset.

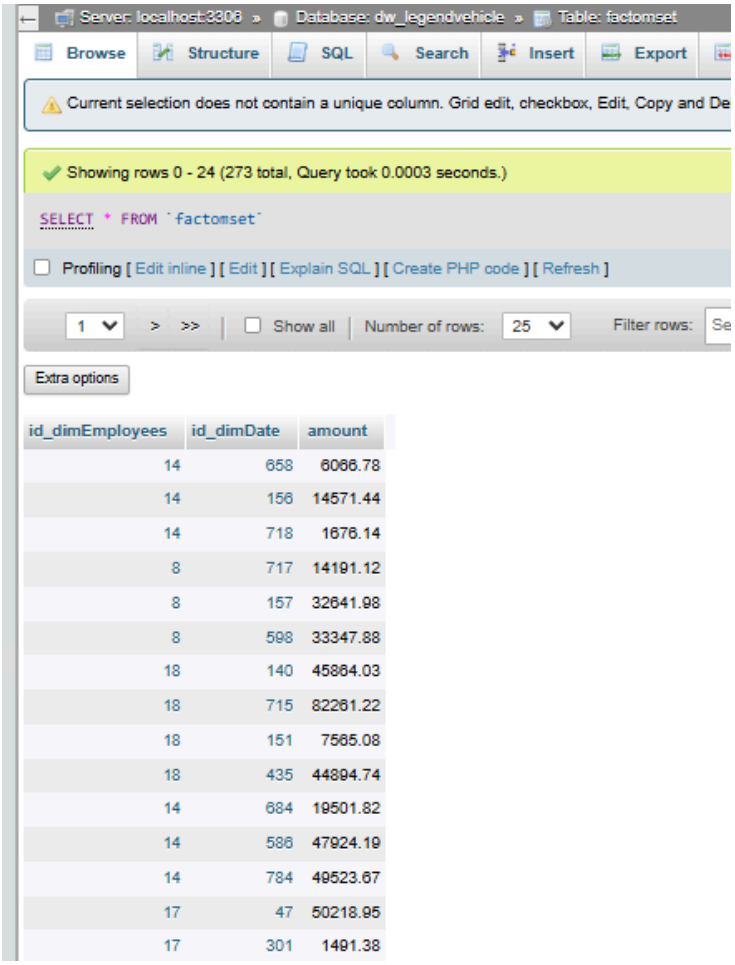


10. Jalankan proses transformation tersebut.





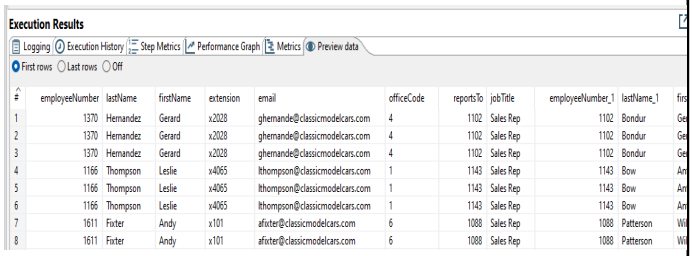
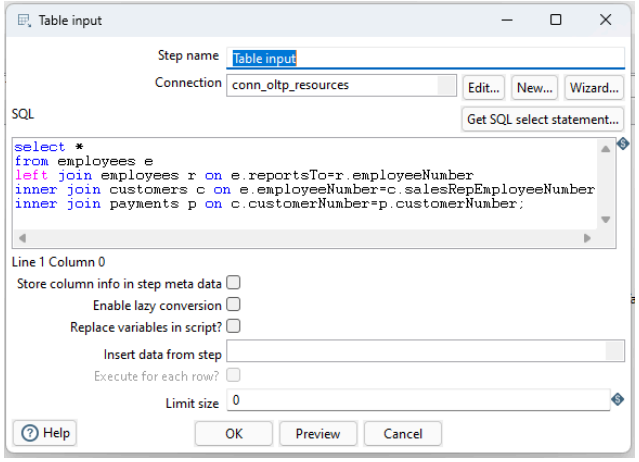
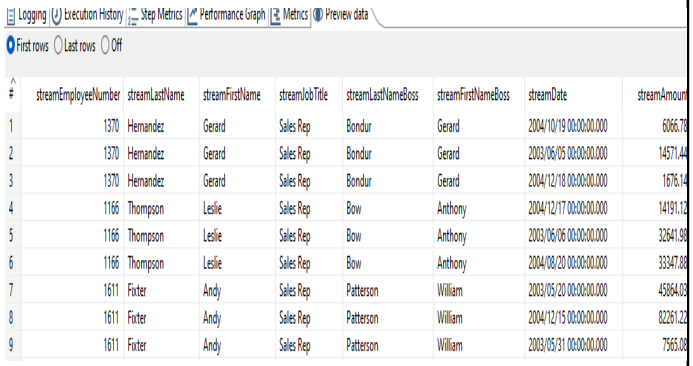
Hasil Transformation

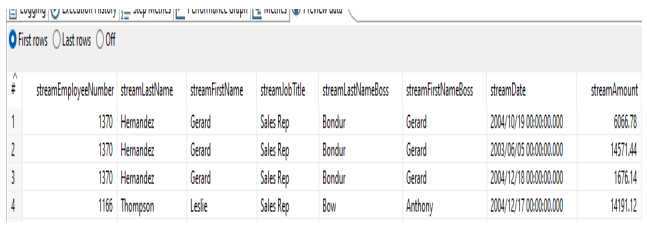
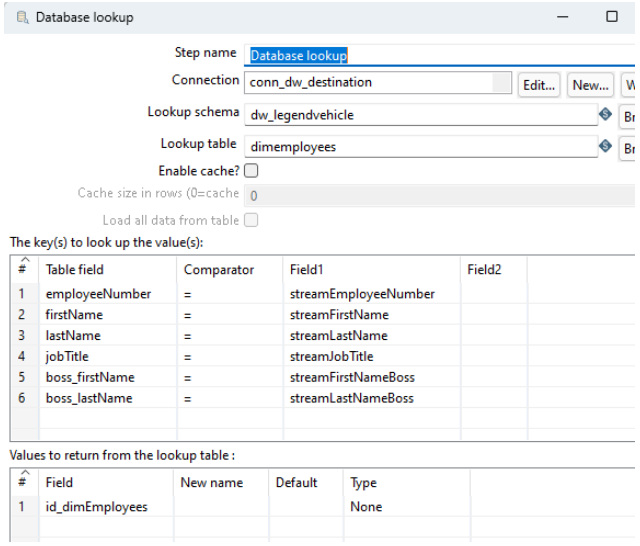
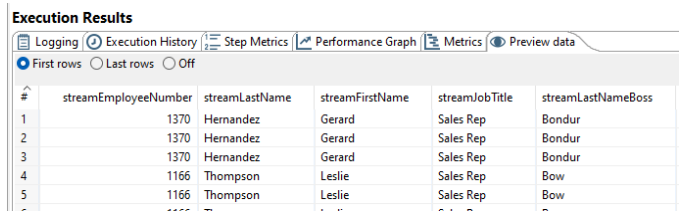


id_dimEmployees	id_dimDate	amount
14	658	6066.78
14	156	14571.44
14	718	1676.14
8	717	14191.12
8	157	32641.98
8	598	33347.88
18	140	45864.03
18	715	82261.22
18	151	7565.08
18	435	44894.74
14	684	19501.82
14	586	47924.19
14	784	49523.67
17	47	50218.95
17	301	1491.38

Tugas 3

1. Buka preview tab pada execution result area di setiap proses object. amati input dan output data yang ada. bandingkan di setiap prosesnya. jelaskan perbedaan disetiap prosesnya.

Proses Objek	SS data input	SS data output	Keterangan
Table Input	data transaksi dari OLTP yaitu (employee, manager, customer, dan payments	 data yang sama seperti query employees	mengambil data transaksi dari database OLTP
Select values	 semua field hasil query	 hanya field yang dipilih (employeeNumber, lastname, firstname, jobTitle, manager, paymentDate, amount)	menghapus field yang tidak dibutuhkan

Database lookup (1)	Select values Step name: Select values Select & Alter | Remove | Meta-data Fields : data dari select values	 data dengan tambahan id_dimEmployees dari dimEmployee	Mencari id_dimemployees berdasarkan employeeNumber
Database lookup (2)	 data dari database lookup 1	 data dengan tambahan id_dimdatedari dimDate	mencari id_dimDate berdasarkan payment

Select values

#	Fieldname	Rename to
1	employeeNumber	streamEmployeeNumber
2	lastName	streamLastName
3	firstName	streamFirstName
4	extension	
5	email	
6	officeCode	
7	reportsTo	
8	jobTitle	streamJobTitle

#	streamEmployeeNumber	streamLastName	streamFirstName	streamJobTitle	streamLastNameBoss	streamFirstNameBoss	streamDate	streamAmount
1	1370	Hernandez	Gerard	Sales Rep	Bondur	Gerard	2004/10/19 00:00:00.000	6066.78
2	1370	Hernandez	Gerard	Sales Rep	Bondur	Gerard	2003/06/05 00:00:00.000	14571.44
3	1370	Hernandez	Gerard	Sales Rep	Bondur	Gerard	2004/12/18 00:00:00.000	1676.14
4	1166	Thompson	Leslie	Sales Rep	Bow	Anthony	2004/12/17 00:00:00.000	14191.12

Database lookup

Step name: Database lookup
Connection: conn_dw_destination
Lookup schema: dw_legendvehicle
Lookup table: dimemployees
Enable cache?
Cache size in rows (0=cache): 0
Load all data from table

The key(s) to look up the value(s):

#	Table field	Comparator	Field1	Field2
1	employeeNumber	=	streamEmployeeNumber	
2	firstName	=	streamFirstName	
3	lastName	=	streamLastName	
4	jobTitle	=	streamJobTitle	
5	boss_firstName	=	streamFirstNameBoss	
6	boss_lastName	=	streamLastNameBoss	

Values to return from the lookup table :

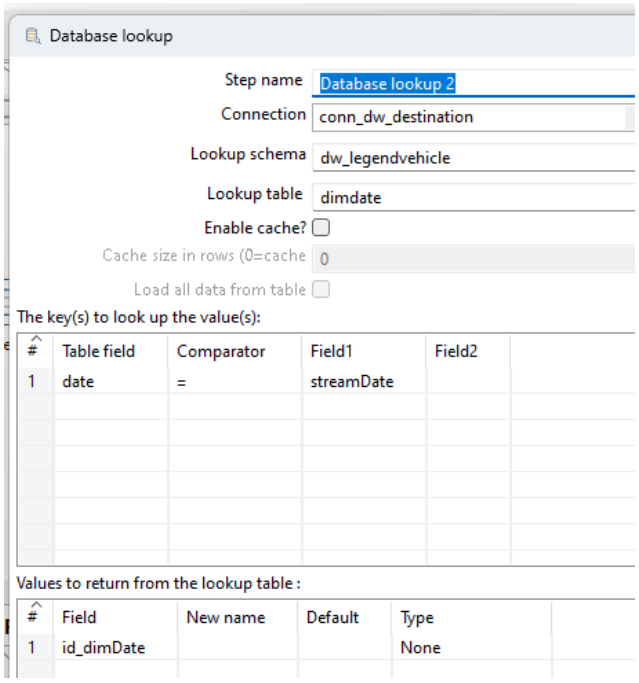
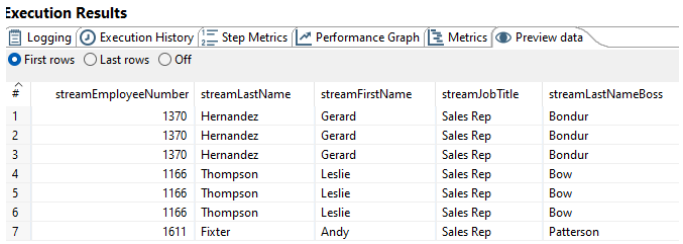
#	Field	New name	Default	Type
1	id_dimEmployees			None

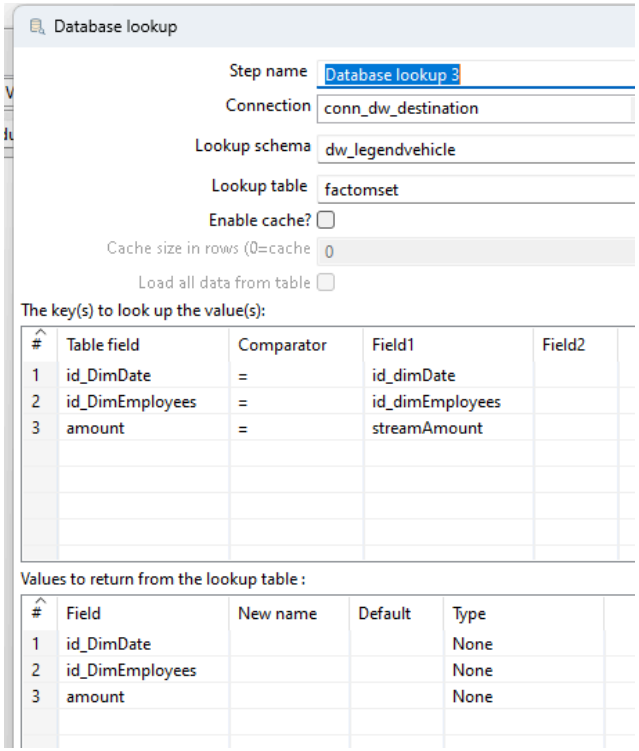
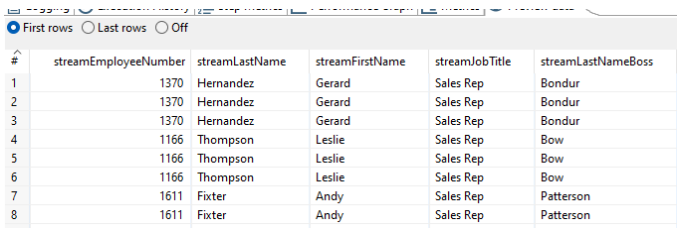
Execution Results

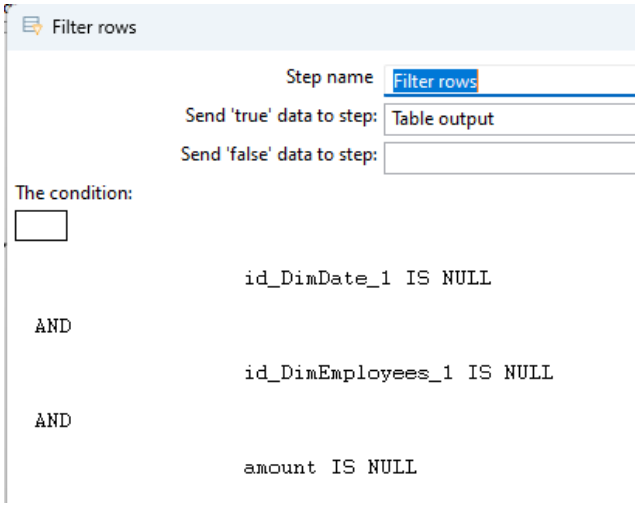
Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

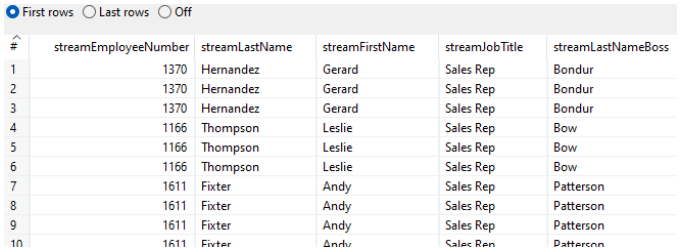
First rows | Last rows | Off

#	streamEmployeeNumber	streamLastName	streamFirstName	streamJobTitle	streamLastNameBoss
1	1370	Hernandez	Gerard	Sales Rep	Bondur
2	1370	Hernandez	Gerard	Sales Rep	Bondur
3	1370	Hernandez	Gerard	Sales Rep	Bondur
4	1166	Thompson	Leslie	Sales Rep	Bow
5	1166	Thompson	Leslie	Sales Rep	Bow

Database lookup (3)	 data dari database lookup 2	 data dengan tambahan id factomset jika sudah ada	mengecek apakah kombinasi id_dimemployees dari id_dimDate sudah ada di factomset

Filter rows	 Database lookup Step name: Database lookup 3 Connection: conn_dw_destination Lookup schema: dw_legendvehicle Lookup table: factomset Enable cache? Cache size in rows (0=cache: 0 Load all data from table The key(s) to look up the value(s): \# | Table field | Comparator | Field1 | Field2 1 | id_DimDate | = | id_dimDate 2 | id_DimEmployees | = | id_dimEmployees 3 | amount | = | streamAmount Values to return from the lookup table : \# | Field | New name | Default | Type 1 | id_DimDate | None 2 | id_DimEmployees | None 3 | amount | None data dari database lookup 3	 First rows / Last rows / Off \# | streamEmployeeNumber | streamLastName | streamFirstName | streamJobTitle | streamLastNameBoss 1 | 1370 | Hernandez | Gerard | Sales Rep | Bondur 2 | 1370 | Hernandez | Gerard | Sales Rep | Bondur 3 | 1370 | Hernandez | Gerard | Sales Rep | Bondur 4 | 1166 | Thompson | Leslie | Sales Rep | Bow 5 | 1166 | Thompson | Leslie | Sales Rep | Bow 6 | 1166 | Thompson | Leslie | Sales Rep | Bow 7 | 1611 | Fixter | Andy | Sales Rep | Patterson 8 | 1611 | Fixter | Andy | Sales Rep | Patterson data yang belum ada dalam factomset	menyaring data yang sudah ada agar tidak dimasukkan lagi

<table>
<tr>
<td>Table output</td>
<td> 

data hasil filter rows</td>
<td> 

#	streamEmployeeNumber	streamLastName	streamFirstName	streamJobTitle	streamLastNameBoss
1	1370	Hernandez	Gerard	Sales Rep	Bondur
2	1370	Hernandez	Gerard	Sales Rep	Bondur
3	1370	Hernandez	Gerard	Sales Rep	Bondur
4	1166	Thompson	Leslie	Sales Rep	Bow
5	1166	Thompson	Leslie	Sales Rep	Bow
6	1166	Thompson	Leslie	Sales Rep	Bow
7	1611	Fixter	Andy	Sales Rep	Patterson
8	1611	Fixter	Andy	Sales Rep	Patterson
9	1611	Fixter	Andy	Sales Rep	Patterson
10	1611	Fixter	Andy	Sales Rep	Patterson

data masuk ke tabel factomset</td>
<td>memasukkan data baru ke dalam tabel OLAP</td>
</tr>
</table>

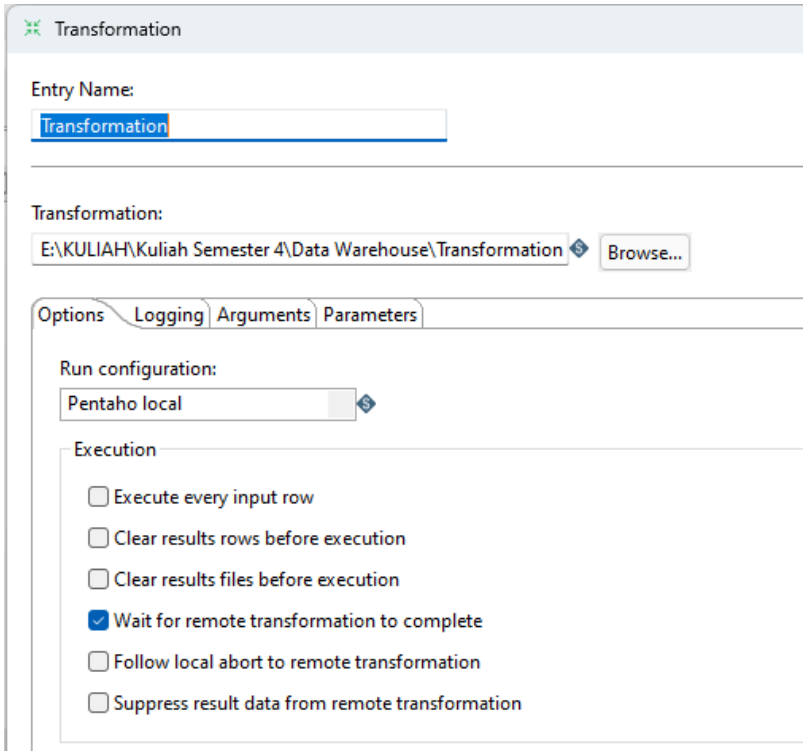
2. Jika proses itu di ulangi (di run kembali) apakah data akan redudant?
   - Tidak redundant
     karena database lookup 3 berfungsi untuk mengecek apakah kombinasi id_dimEmployees dan id_dimDate sudah ada di factomset filter rows memastikan hanya data baru yang akan masuk ke dalam tabel factOmset
     jika transformasi dijalankan ulang, data yang sudah ada tidak akan dimasukkan kembali, sehingga tidak ada duplikasi data (redundasi)

D. Jobs

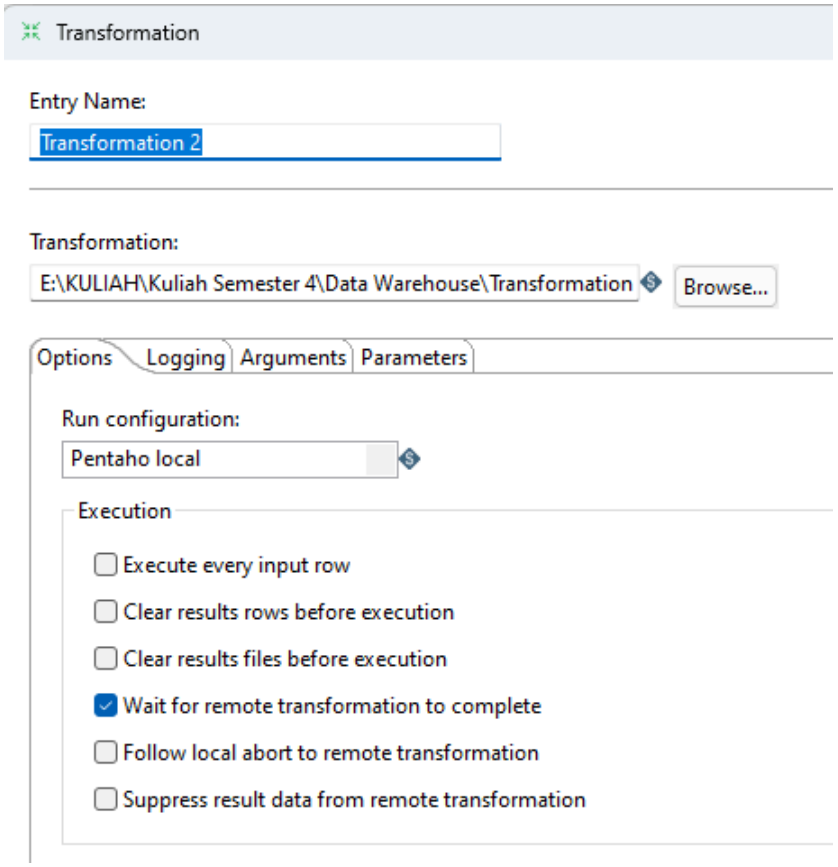
1. Buka Jobs pada **File - New - Jobs**
2. Gunakan 5 objects dan hububngkan sesuai urutan sebagai berikut:
   - **Start:** Objek untuk melakukan konfigurasi cron job dari proses ETL yag telah dibuat
   - **Transformation 1 :** digunakan untuk menjalankan transformation pembuatan dimDate.
   - **Transformation 2 :** digunakan untuk menjalankan transformation pembuatan dimEmployees.
   - **Transformation 3 :** digunakan untuk menjalankan transformation pembuatan factOmset.
   - **Success:** Objek untuk menandakan bahwa proses telah selesai.

3. Konfigurasi dilakukan pada setiap transformation untuk mengambil file tranformation yang telah dibuat pada bagian A, B dan C.

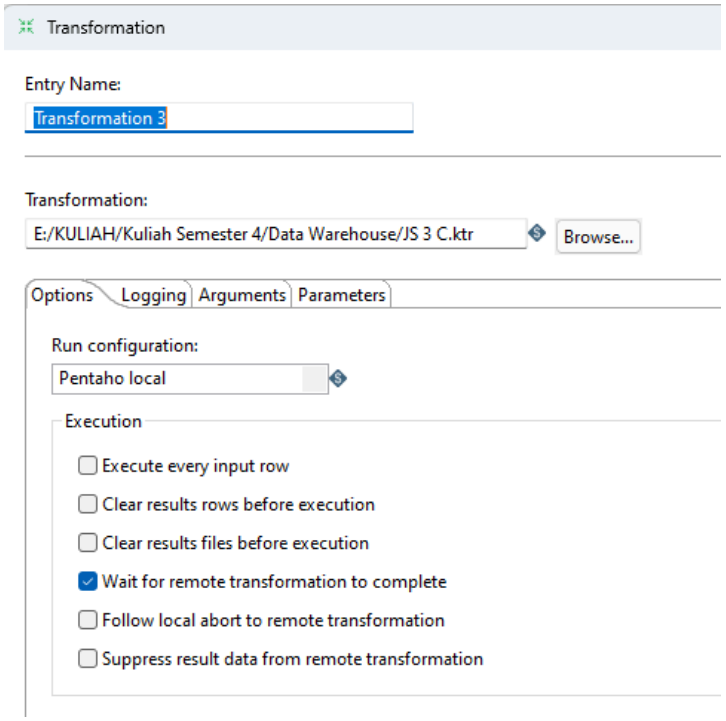
File Transformation A



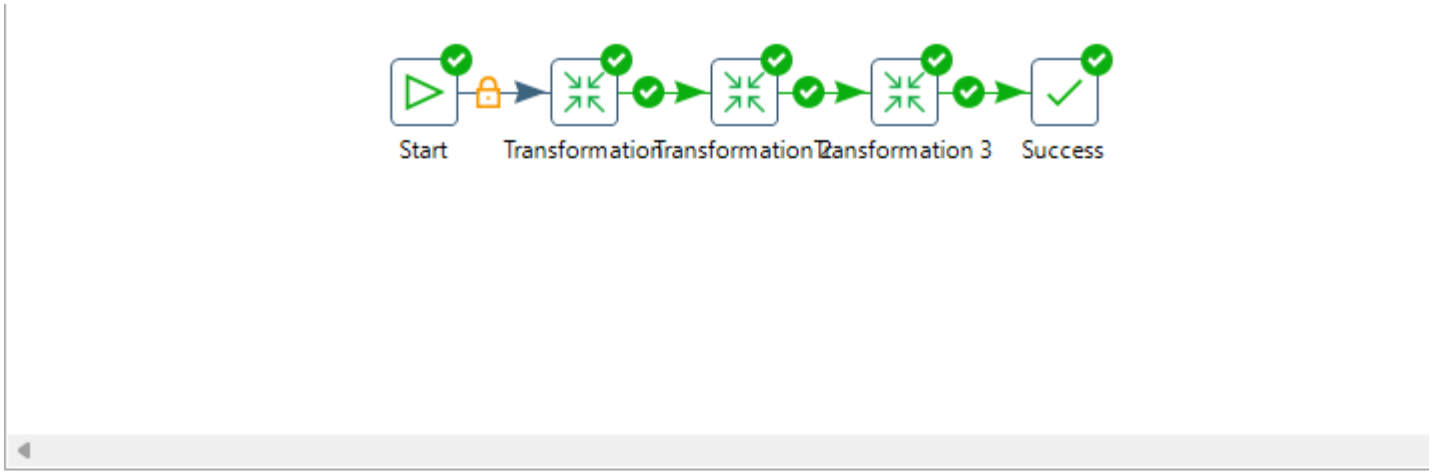
File Transformation B



File Transformation C



4. Jalankan jobs tersebut ketika semua transformation telah dihubungkan dengan file tranformation yang telah dibuat sebelumnya.



Execution Results

Logging | History | Job metrics | Metrics

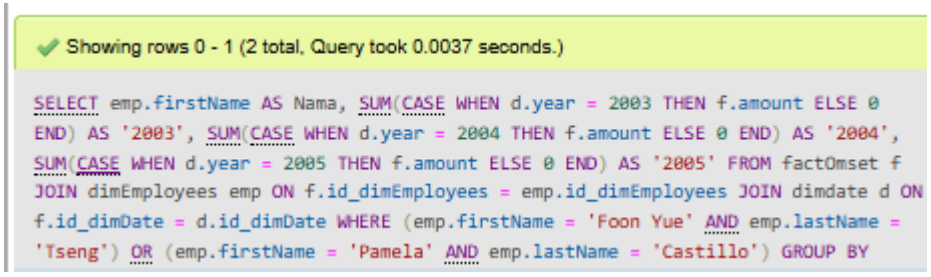
```
2025/04/22 22:21:03 - Spoon - Starting job...
2025/04/22 22:21:03 - JS 3 D - Start of job execution
2025/04/22 22:21:03 - JS 3 D - Starting entry [Transformation]
2025/04/22 22:21:04 - Transformation - Using run configuration [Pentaho local]
2025/04/22 22:21:04 - Transformation 1Js3 - Dispatching started for transformation [Transformation 1Js3]
2025/04/22 22:21:04 - Table output.0 - Connected to database [conn_dw_destination] (commit=1000)
2025/04/22 22:21:04 - Generate rows.0 - Finished processing (I=0, O=0, R=0, W=1825, U=0, E=0)
2025/04/22 22:21:04 - Add sequence.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
2025/04/22 22:21:04 - Calculator.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
```

Tugas 4

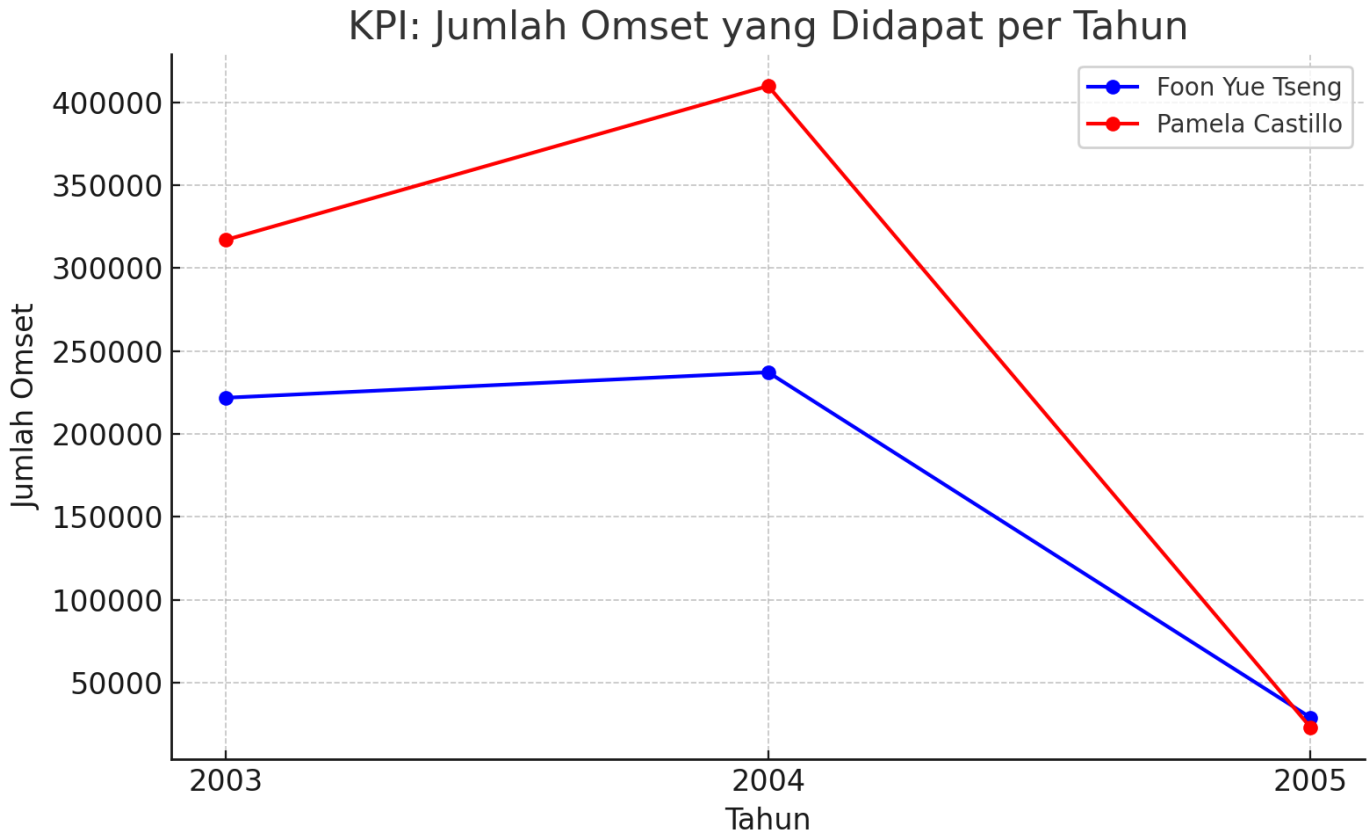
1. Buka desain database dari dw_legendvehicle pada DBMS, bandingkan design tersebut dengan desain db OLTP legendVehicle pada jobsheet 2. analisalah dan ceritakan perbedaannya.

No.	Aspek	OLTP (Operational Database)	OLAP (Data Warehouse)
1	Tujuan	Digunakan untuk transaksi harian seperti penjualan, update, data, dll	Digunakan untuk analisis data dalam jangka panjang
2	Struktur Data	Normalisasi tinggi banyak tabel dengan relasi untuk efisensi penyimpanan	Denormalisasi mengurangi jumlah join agar query lebih cepat
3	Contoh Tabel	customers, orders, vehicles	dimDate, dimEmployees, factOmset
4	Performansi	Dioptimalkan untuk transaksi cepat	Dioptimalkan untuk query analitik dan laporan

2. Buatlah report pertahun untuk KPI **"Jumlah omset yang didapat"** pada **Foon Yue Tseng** dan **Pamela Castillo**. Serta gambarkan grafiknya (grafik garis).



Nama	2003	2004	2005
Foon Yue	221887.03	237255.26	29070.38
Pamela	317104.78	409910.07	23187.02



3. Jelaskan perbedaan query saat mendapatkan data pada nomor 2 dengan query pada saat Jobsheet 2!
    - Query di JS 2 OLTP menggunakan tabel transaksi langsung misalnya orders. payments dll
    - Query di tugas ini OLAP menggunakan dimensional modeling yaitu factOmset, dimEmployees, dan dimDate sehingga lebih efisien dan fokus ke agregasi data
4. Simpulkan dengan bahasa sendiri, apa perbedaan OLTP dan OLAP?
    - OLTP (Online Transcation Processing) database untuk transaksi sehari - hari, cepat dalam operasi CRUD (create, read, update , delete)
    - OLAP (Online Analytical Processing) database untuk analisis dan pelaporan, lebih fokus pada agregasi dan performa query