

NIM : 2241720255

Nama Lengkap : Davis Maulana Hermanto

Daftar Isi

[Chapter 1(Data Warehouse Sederhana) 4](#_Toc26352)

[Tugas 1 5](#_Toc229)

[Sub Chapter 1 (Pengambilan Data (Extract)) 6](#_Toc23206)

[Tugas 2 6](#_Toc756)

[Sub Chapter 2 (Filter Data (Transform) dan Pengemasan data (Load)) 7](#_Toc9229)

[Tugas 3 7](#_Toc678)

[Sub Chapter 3 (Studi Kasus) 7](#_Toc12633)

[Chapter 2(Data Sumber) 11](#_Toc21818)

[Tugas 1 11](#_Toc27484)

[Sub Chapter 2 (Analisa Data) 13](#_Toc11924)

[Tugas 2 14](#_Toc29505)

[Tugas 3 17](#_Toc23722)

[Sub Chapter 3 (Studi Kasus) 22](#_Toc4819)

[Chapter 3(Database Analytical) 24](#_Toc3867)

[Sub Chapter 1 (Dimensi Waktu) 24](#_Toc7211)

[Tugas 1 31](#_Toc5292)

[Sub Chapter 2 (Dimensi Pegawai) 33](#_Toc7114)

[Tugas 2 39](#_Toc1408)

[Sub Chapter 3 (Fakta Pembayaran) 41](#_Toc6234)

[Tugas 3 48](#_Toc18505)

[Sub Chapter 4 (Jobs) 50](#_Toc13747)

[Tugas 4 51](#_Toc4548)

[Sub Chapter 5 (Studi Kasus) 53](#_Toc31959)

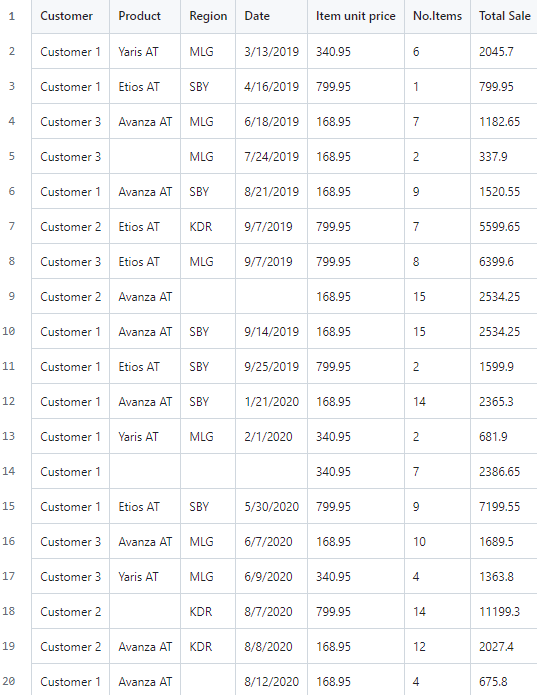
# Chapter 1(Data Warehouse Sederhana)

**Studi Kasus**

Pak Amir adalah Manager Sales Astro Mobil, yang bergerak di bidang distributor mobil wilayah Jawa Timur. Pak Amir meminta data penjualan dari beberapa cabang kepada supervisor. Data tersebut akan digunakan untuk membuat suatu Dashboard pengambilan keputusan. Dikarenakan proses permintaan ini dilakukan secara berulang setiap hari setelah jam kantor maka, Pak Amir akan membuat proses untuk menarik data yang ada pada file milik supervisor tersebut secara otomatis. Namun, kadang data tersebut tidak lengkap. Sehingga Pak Amir akan mengambil data yang lengkap saja dan mengembalikan data yang tidak lengkap.

Dari studi kasus tersebut maka, akan dilakukan:

1. Cek dan Analisa data penjualan
2. Import data dari file (Extract)
3. Identifikasi data yang tidak lengkap (missing data) dan meletakkan pada file yang berbeda (Transform)
4. Memindahkan data yang sudah lengkap ke file dashboard (Load)
5. Mengumpulkan data yang belum lengkap untuk dikembalikan



## Tugas 1

Analisa lah data tersebut

1. Berapa jumlah kolom pada data tersebut?

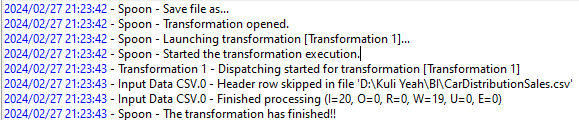
* Ada 7 kolom yang terdiri dari kolom customer, product, region, date, item unit price, no. Items, dan Total Sale.

1. Apa arti atau isi data dari setiap kolom yang ada?
2. Customer : Berisi nama lengkap pelanggan
3. Product : Berisi nama model mobil yang terjual
4. Region : Berisi wilayah penjualan mobil, seperti MLG (Malang), SBY (Surabaya), dan KDR (Kediri)
5. Date : Berisi tanggal penjualan mobil dalam format dd/mm/yyyy
6. Item Unit Price : Berisi harga satuan produk mobil
7. No. Items : Berisi jumlah produk mobil yang terjual pada tanggal tersebut
8. Total Sale : Berisi total penjualan produk mobil pada tanggal tersebut
9. Adakah data yang memiliki nilai null / data yang tidak lengkap?
   1. Pelanggan 3 memiliki kolom Product yang kosong pada tanggal 7/24/2019.
   2. Pelanggan 2 memiliki kolom Date dan Region yang kosong untuk produk Avanza AT.
   3. Pelanggan 1 memiliki kolom Product, Region, dan Date yang kosong pada satu kesempatan.
10. Adakah data yang memiliki tipe yang berbeda dengan data lainnya pada kolom yang sama?
    1. Tidak ada, Semua data dalam setiap kolom tampak memiliki tipe yang konsisten.

## Sub Chapter 1 (Pengambilan Data (Extract))

## Tugas 2

1. Apakah data hasil eksekusi sesuai dengan data aslinya?
2. Tidak
3. PDI Spoon melakukan proses extract, perhatikan pada **Tab Logging** di **Execution Results Area,** langkah-langkah apa sajakah yang dilakukan PDI Spoon untuk melakukan extract data?
4. DI Spoon membukaproyek transformasi yang akan dieksekusi.
5. PDI Spoon memulai eksekusi transformasi dengan nama “Transformation 1”.
6. Transformasi dimulai dan setiap langkah (step) dijalankan secara paralel.
7. PDI Spoon menginisiasi pengiriman data antara langkah-langkah dalam transformasi.
8. PDI Spoon membaca data dari file CSV yang terletak di jalur yang disebutkan.
9. Baris header pada file CSV dilewati (skipped).
10. Setelah memproses data, transformasi selesai.



1. Perhatikan gambar dibawah ini! Apa yang dimaksud dengan I,O,R,W,U,E?
   1. I : Input Data CSV.0, Menunjukkan bahwa data yang diproses berasal dari file CSV dengan nama "Data CSV.0"
   2. O : Finished processing, Menunjukkan bahwa proses pengolahan data telah selesai.
   3. R : 0, Menunjukkan jumlah baris data yang ditolak karena memiliki kesalahan
   4. W: 19, Menunjukkan jumlah baris data yang diterima dan disimpan
   5. U : 0, Menunjukkan jumlah baris data yang diperbarui
   6. E : 0, Menunjukkan jumlah baris data yang dihapus



1. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU ). Bandingkan dalam bentuk tabel.

* Saya hanya memerlukan waktu 1 detik saja untuk melakukan semua proses tersebut

## Sub Chapter 2 (Filter Data (Transform) dan Pengemasan data (Load))

## Tugas 3

1. Apa perbedaan isi data output dilihat dari isi file csv dan file excel?
   1. Pada file csv, data yang ditampilkan telah tersaring sehingga tidak ada data null pada suatu kolom setiap barisnya. Sedangkan pada file excel, data yang ditampilkan merupakan baris baris yang memiliki data null pada suatu kolomnya.
2. Jelaskan apa yang terjadi pada proses **Filter rows**!
   1. Pada proses “Filter rows”, baris data yang memenuhi kondisi tertentu akan disaring dan diproses lebih lanjut. Dalam kasus ini, baris yang memiliki nilai bukan NULL untuk kolom Customer, Product, Region, Date, Item unit price, No.Items, dan Total Sale akan dianggap sebagai ‘true’ dan diteruskan ke langkah selanjutnya. Dengan kata lain, proses ini bertujuan untuk menghilangkan baris data yang memiliki nilai kosong atau tidak lengkap. Data yang memenuhi syarat akan diteruskan ke langkah berikutnya, sedangkan data yang tidak memenuhi syarat akan diabaikan. Jadi, pada tahap ini, kita memfilter data berdasarkan kondisi-kondisi tertentu agar hanya data yang relevan dan lengkap yang digunakan dalam analisis lebih lanjut
3. Berapa waktu yang dibutuhkan untuk melakukan eksekusi Extract Transfer Load data ini? Bandingkan dengan 3 teman yang lain, identifikasi pula spesifikasi perangkat yang digunakan (Processor, RAM, Storage, VGA, CPU ). bandingkan dalam bentuk tabel

## Sub Chapter 3 (Studi Kasus)

Pak Pascanowo seorang ketua RW di Ibu Kota Nusabangsa (IKN). Terdapat 20 orang dari negara Konoha yang akan pindah dan berdomisili di IKN. Tata letak pemukiman pada IKN telah dikondisikan agar berkelompok disesuaikan dengan lokasi pekerjaannya agar akses terjangkau.

Sebagai contoh, mahasiswa akan dikumpulkan dan berdomisili di wilayah edukasi yang akses dekat dengan kampus-kampus perguruan tinggi. Sedangkan dokter akan berdomisili dekat dengan rumah sakit atau layanan kesehatan dan PNS akan berdomisili dekat dengan perkantoran dan layanan publik. Selain itu, orang yang bukan dari ketiga kategori tersebut bebas memilih domisili atau tempat tinggalnya di IKN.

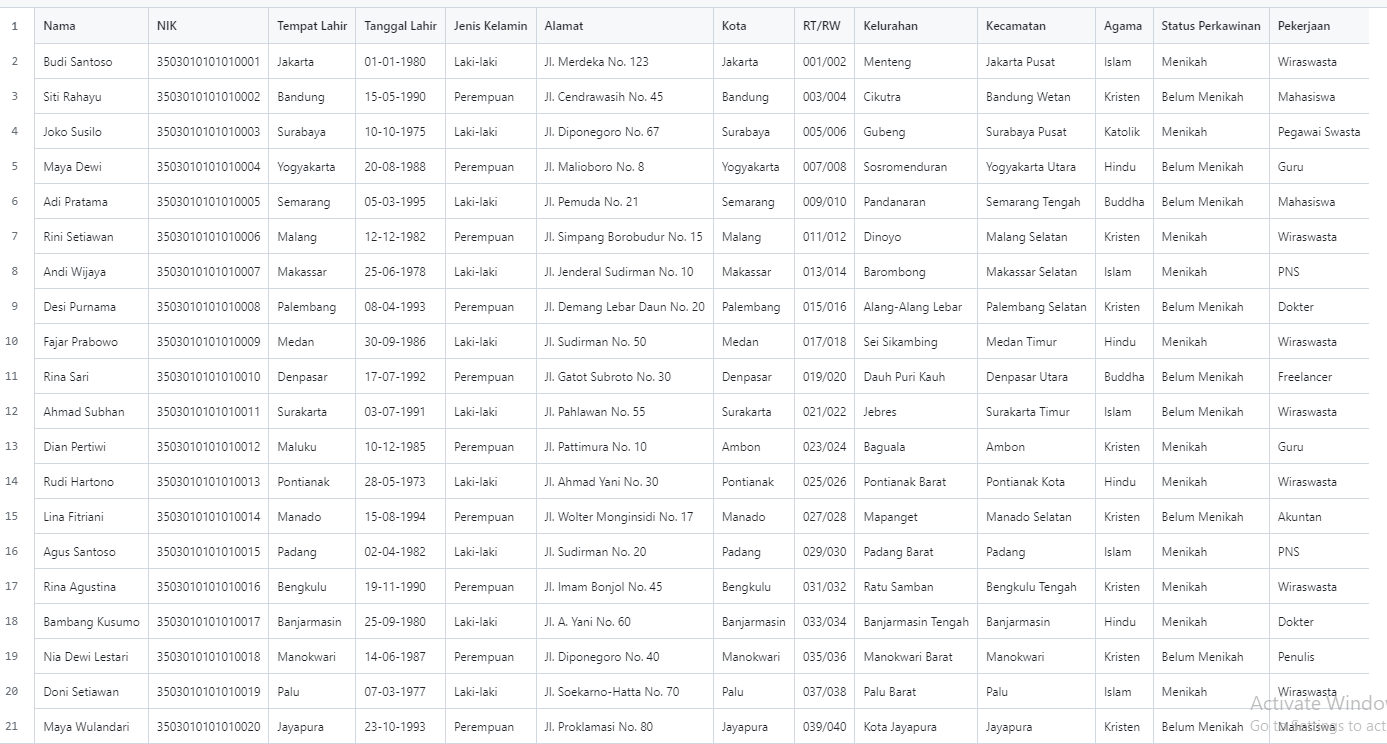
Pak Pascanowo akan mendapatkan data dari pemerintah pusat tentang Masyarakat Konoha yang akan pindah ke IKN. darai data tersebut Pak Pascanowo akan membagi menjadi 4 kelompok,

**Data masyarakat berprofesi mahasiswa** akan diberikan ke Pak Emir Makarena selaku ketua RT wilayah pendidikan.

**Data masyarakat berprofesi PNS** akan diberikan ke Pak Yasana Laili selaku ketua RT wilayah pemerintahan

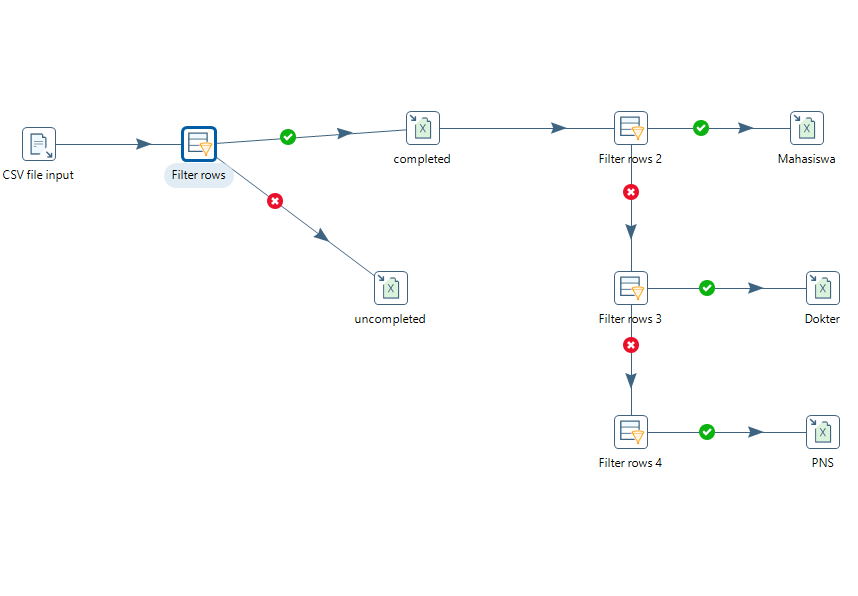
**Data masyarakat berprofesi Dokter** akan diberikan ke Pak Budi Pekerti selaku ketua RT wilayah kesehatan

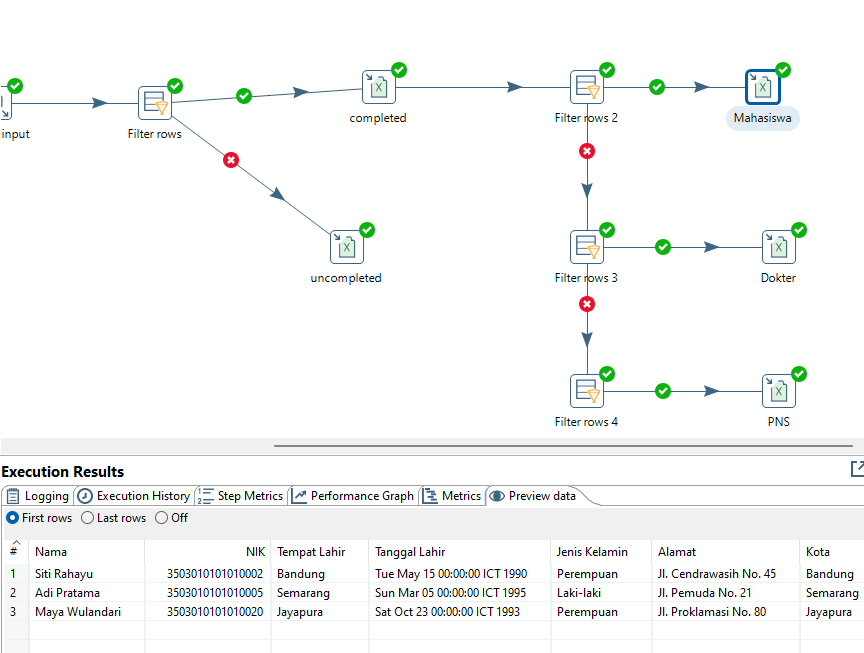
data-data tersbut akan digunakan untuk analisa di masing-masing wilayah dan juga Dashboard untuk emmbantu mengambil keputusan.

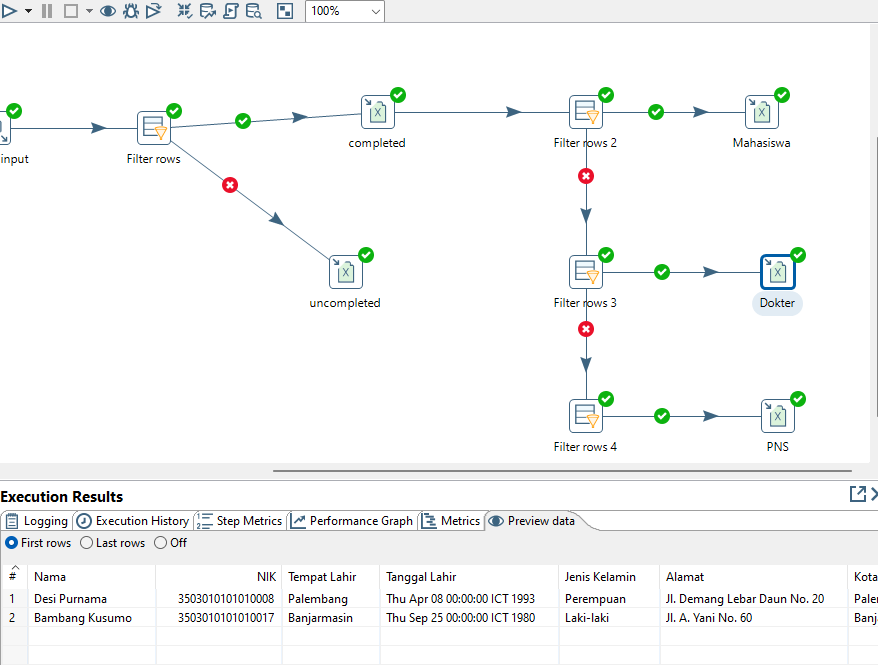


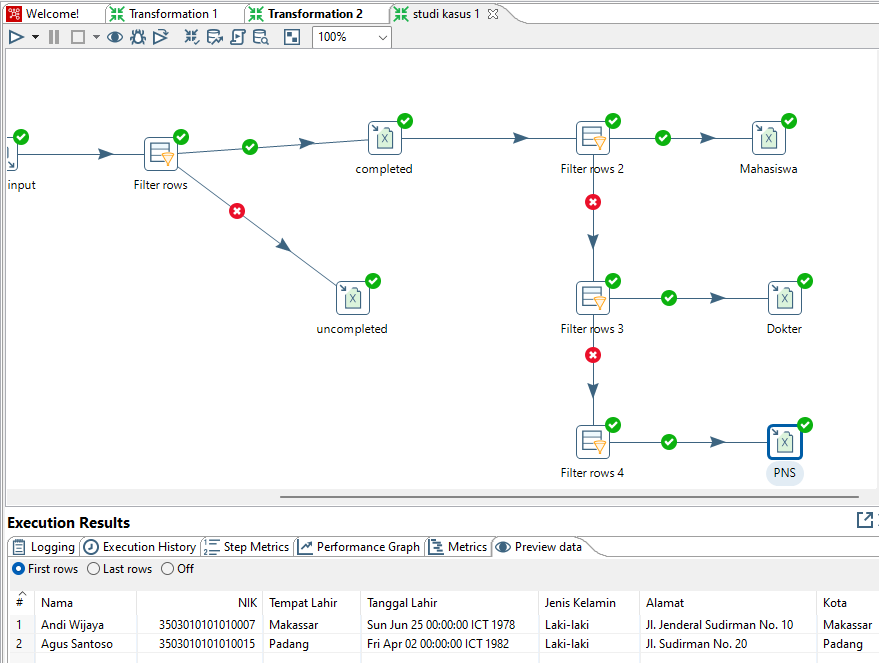
Bantulah Pak Pascanowo untuk memishkan data tersebut menjadi 4 file: MasyarakatMahasiswa, MasyarakatPNS, MasyarakatDokter, MasyarakatLainnya.

Dikarenakan proses tersebut akan repetisi setiap ada masyarakat Konoha yang berpindah ke IKN maka buatlah sistem tersebut menggunakan PDI Spoon.





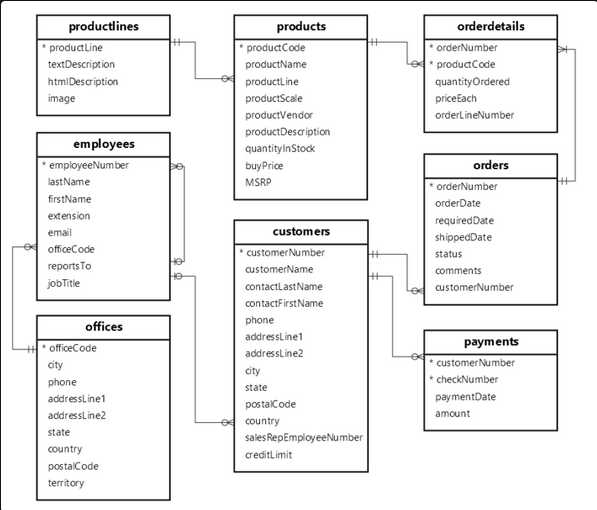




# Chapter 2(Data Sumber)

**Studi Kasus**

LegendVehicle merupakan perusahan jual-beli tukar-tambah kendaraan klasik. Perusahaan ini memiliki cabang di berbagai negara. LegendVehicle memiliki sistem informasi ERP sendiri. Salah satu modul dari sistem ERP tersebut adalah modul penjualan. Desain database dari modul tersebut adalah sebagai berikut:



Selain itu proses penjualan kendaraan pada perusahaan tersebut bukan hanya melalui showroom cabang, melainkan reseller-reseller bebas lainnya.

## Tugas 1

1. Import data perusahaan tersebut pada DBMS MySQL!



1. Analisa struktur data dari database perusahaan tersebut, dalam bentuk tabel, analisa hubungan setiap tabel nya!

|  |  |  |
| --- | --- | --- |
| **Tabel 1** | **Tabel 2** | **Jenis Relasi** |
| productlines | products | One to many |
| products | orderdetails | One to many |
| customers | orders | One to many |
| offices | employees | One to many |
| employees | customers | One to many |
| customers | payments | One to many |
| orders | orderdetails | One to many |

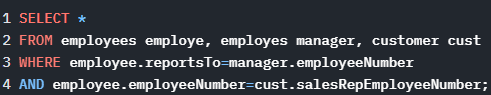
1. Analisa jumlah field pada setiap tabel!

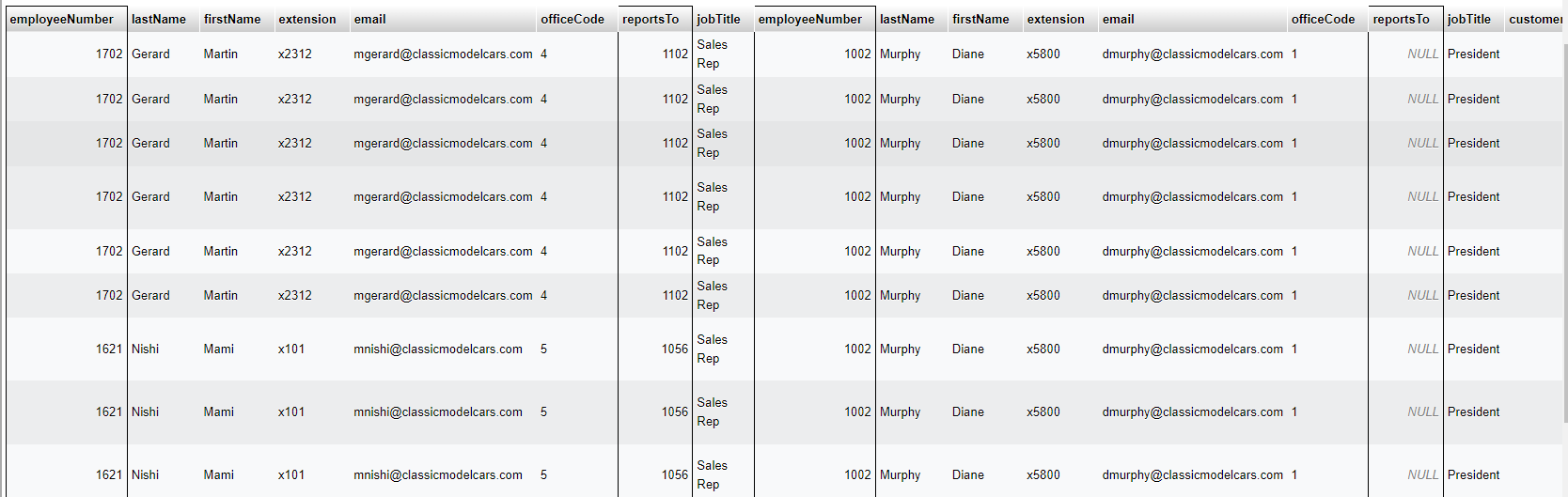
|  |  |
| --- | --- |
| **Nama Tabel** | **Jumlah Field** |
| productlines | 4 fields |
| products | 9 fields |
| orderdetails | 5 fields |
| employees | 8 fields |
| offices | 9 fields |
| customers | 13 fields |
| orders | 7 fields |
| payments | 4 fields |

## Sub Chapter 2 (Analisa Data)

**Praktikum 1**

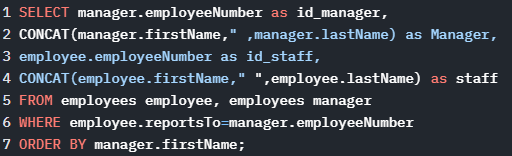
1. Jalankan **query** berikut pada **DBMS MySql** yang telah tersedia **data Perusahaan LegendVehicle.**

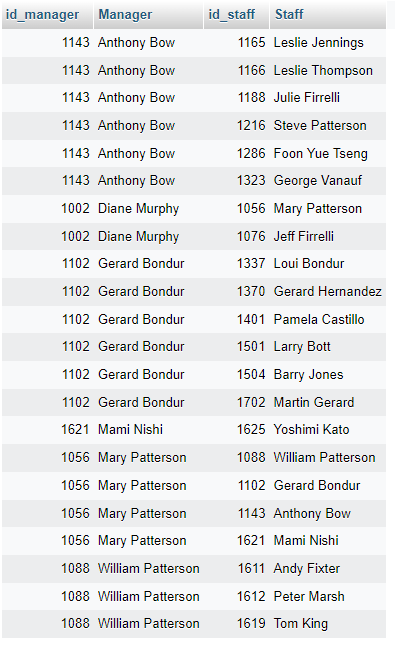




maka hasil dari query tersebut adalah data **Employee** beserta **Manajernya** dan **Customer** yang ia miliki. perhatikan hasil data dengan seksama.

1. Buka **tab baru** pada browser untuk melakukan eksekusi **query** berikut:





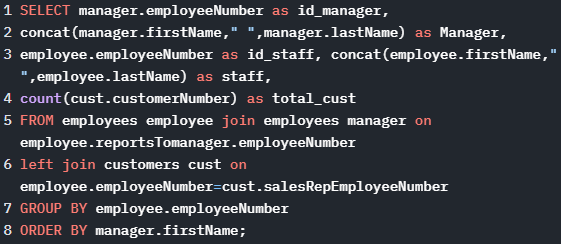
dari hasil **query** diatas maka akan ditemukan atasan dari setiap pegawai.

## Tugas 2

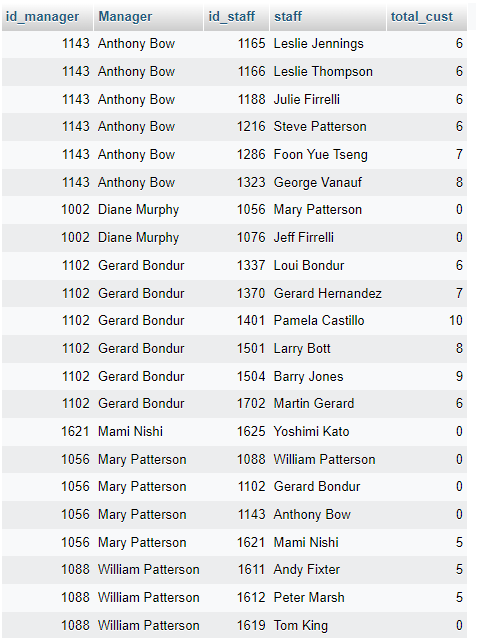
1. Gambarlah hirarki organisasi berdasarkan atasan dari setiap pegawai sesuai dengan hasil prkatikum diatas!

* Anthony Bow [1143]
  + Leslie Jennings [1165]
  + Leslie Thompson [1166]
  + Julie Firrelli [1188]
  + Steve Patterson [1216]
  + Foon Yue Tseng [1286]
  + George Vanauf [1323]
* Diane Murphy [1002]
  + Mary Patterson [1056]
  + Jeff Firrelli [1076]
* Gerard Bondur [1102]
  + Loui Bondur [1337]
  + Gerard Hernandez [1370]
  + Pamela Castillo [1401]
  + Larry Bott [1501]
  + Barry Jones [1504]
  + Martin Gerard [1702]
* Mami Nishi [1621]
  + Yoshimi Kato [1625]
* Mary Patterson [1056]
  + William Patterson [1088]
  + Gerard Bondur [1102]
  + Anthony Bow [1143]
  + Mami Nishi [1621]
* William Patterson [1088]
  + Andy Fixter [1611]
  + Peter Marsh [1612]
  + Tom King [1619]

1. Buka **tab baru** pada browser untuk melakukan eksekusi **query** berikut:



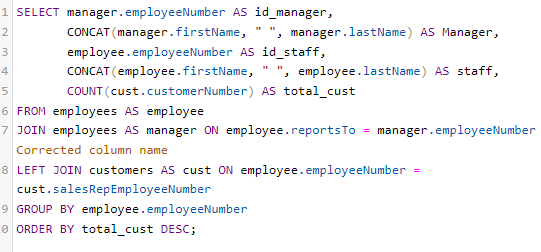
Dari query tersebut menghasilkan jumlah **customer** dari setiap **staff**.

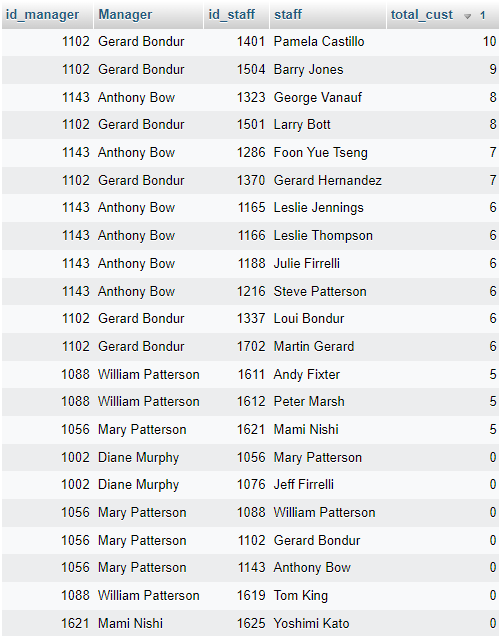


Jika perusahaan tersebut memiliki **KPI (Key Performances Indicator) "Jumlah customer yang bertransaksi"** maka jawablah pertanyaan-pertanyaan berikut!

## Tugas 3

1. Siapakah staff dengan hirarki paling bawah yang berprestasi dilihat dari jumlah customer terbanyak?





* Pamela Castillo

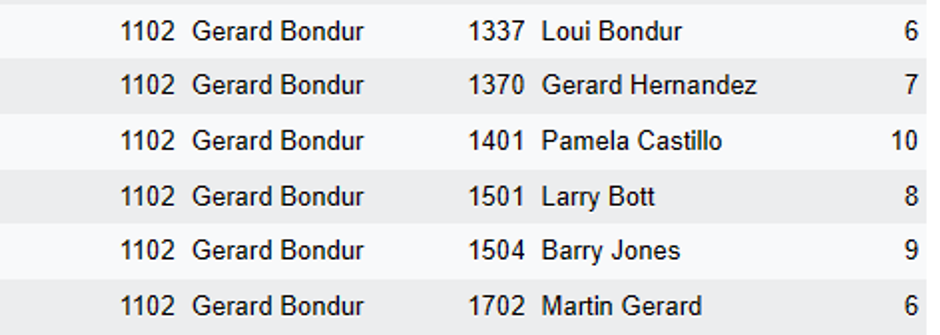
1. Jika KPI atasan dihitung dari customer yang dimilikinya dijumlah dengan customer dari staff dibawahnya, urutkan ranking prestasi keseluruhan pegawai beserta keterangan jumlah customer yang dimilikinya!



Total customer = 39



Total customer = 0



Total customer = 46



Total customer = 0



Total customer = 5

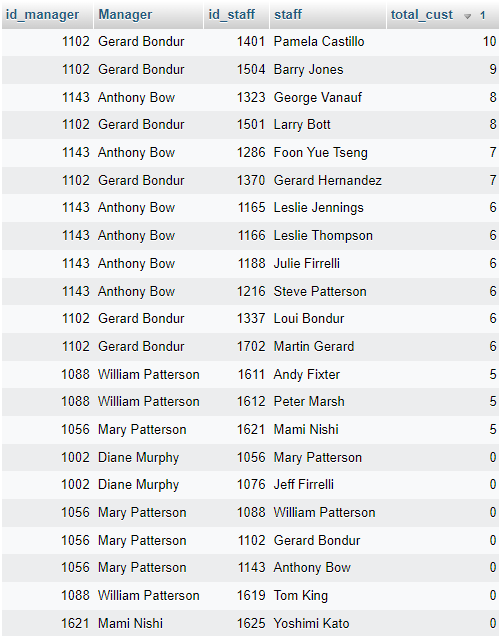


Total customer = 10

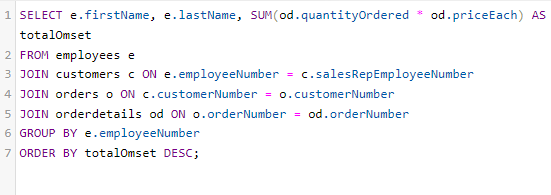
* Manager list :

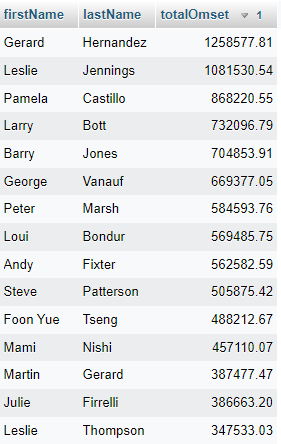
1. Gerald Bondur
2. Anthony Bow
3. William Patterson
4. Marry Peterson
5. Diane Murphy
6. Yoshimi Kato

* Staff list:



1. Analisa kembali data LegendVehicle untuk mendapatkan ranking pegawai berdasarkan KPI **"Jumlah omset yang didapat".** Urutkan ranking pegawai beserta keterangan dana yang didapat!

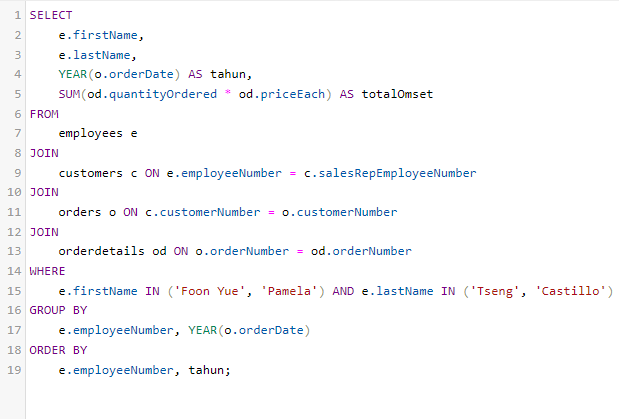


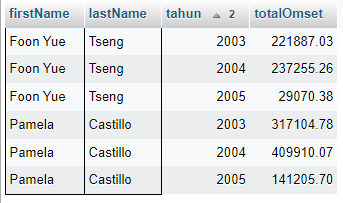


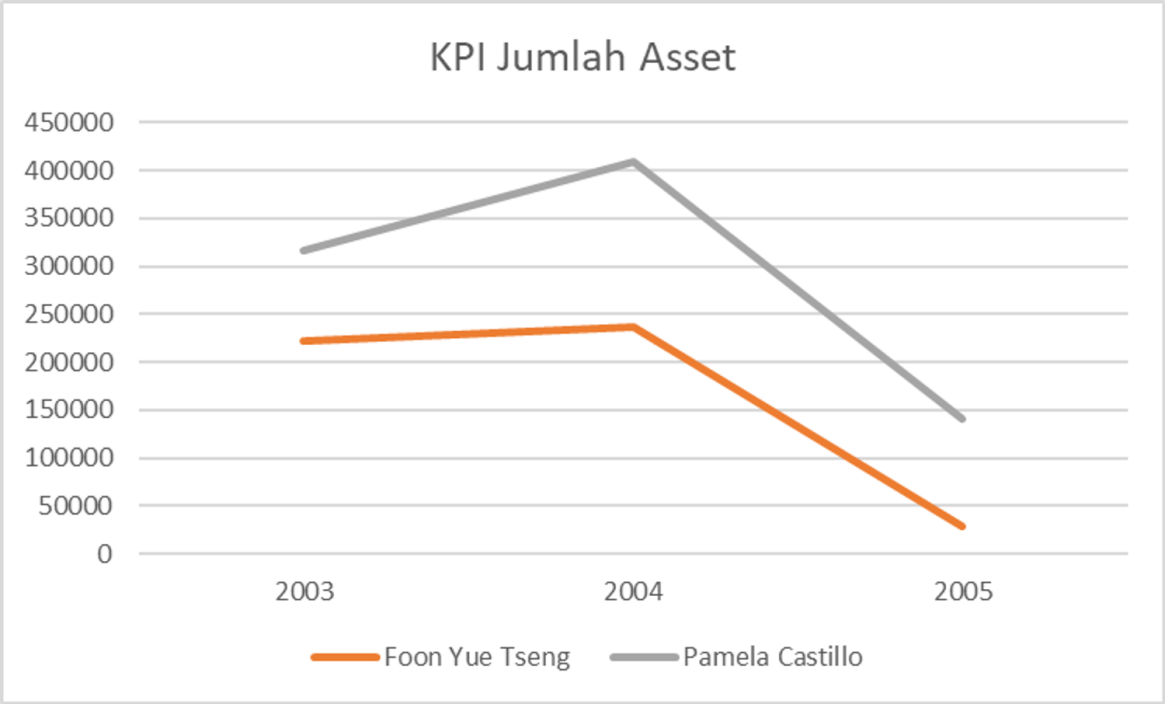
1. Jika KPI yang pertama merupakan "**Jumlah customer yang bertransaksi"** sedangkan KPI yang kedua **"Jumlah omset yang didapat".** Maka, berapakah jumlah field yang dibutuhkan untuk mendapatkan informasi tersebut?

|  |  |
| --- | --- |
| KPI | Total Field |
| Jumlah customer yang bertransaksi | 5 |
| Jumlah omset yang didapat | 3 |

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







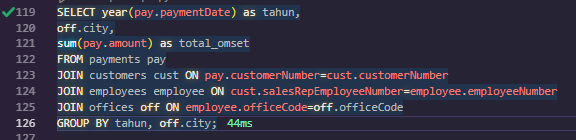
## Sub Chapter 3 (Studi Kasus)

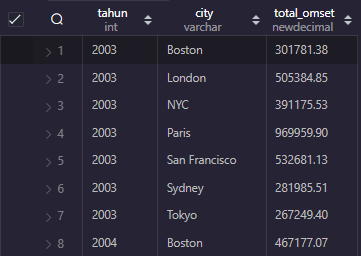
Pak Huhut merupakan pemegang saham LegendVehicle. dia membutuhkan dashboard untuk melihat perkembangan penjualan (omset) disetiap cabang di tiap tahunnya. Dikarenakan perusahaan tersebut belum merekrut Data Engineer maka, penarikan informasi hanya bisa dilakukan melaluai OLTP yang ada.

Hasil report yang diinginkan adalah grafik berdasarkan tabel berikut:

Analisalah terlbih dahulu:

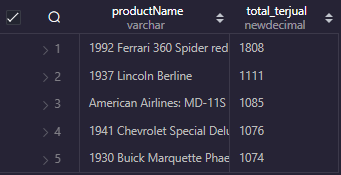
1. Field apa saja yang diperlukan untuk menampilkan penjualan di setiap cabang. field yang diperlukan adalah paymentDate, amount, city.
2. Bentuk query dengan memperhatikan relasi antar table.



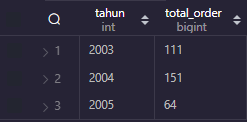


**SOAL BONUS**: buatlah report lain dengan sumber data OLTP yang sama, Analisa field yang digunakan, bentuk struktur query dan tuliskan dalam table serta grafiknya.

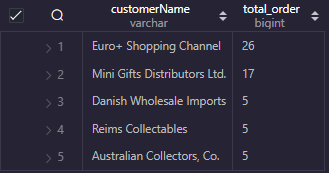
1. Produk paling banyak dijual



1. Total order berdasarkan tahun



1. Customer paling banyak bertransaksi



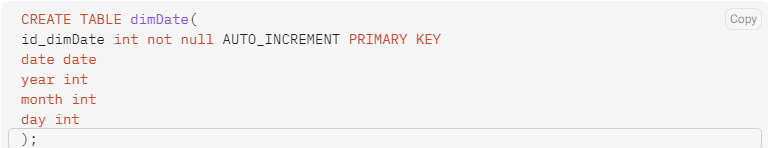
# Chapter 3(Database Analytical)

## Sub Chapter 1 (Dimensi Waktu)

1. Buatlah sebuah database yang digunakan sebagai ****OLAP**** dengan nama ****dw\_LegendVehicle****.



1. Buatlah table untuk menyimpan data master waktu atau yang disebut dengan ****tabel dimensi.**** Beri nama table tersebut dengan nama ****dimDate****.

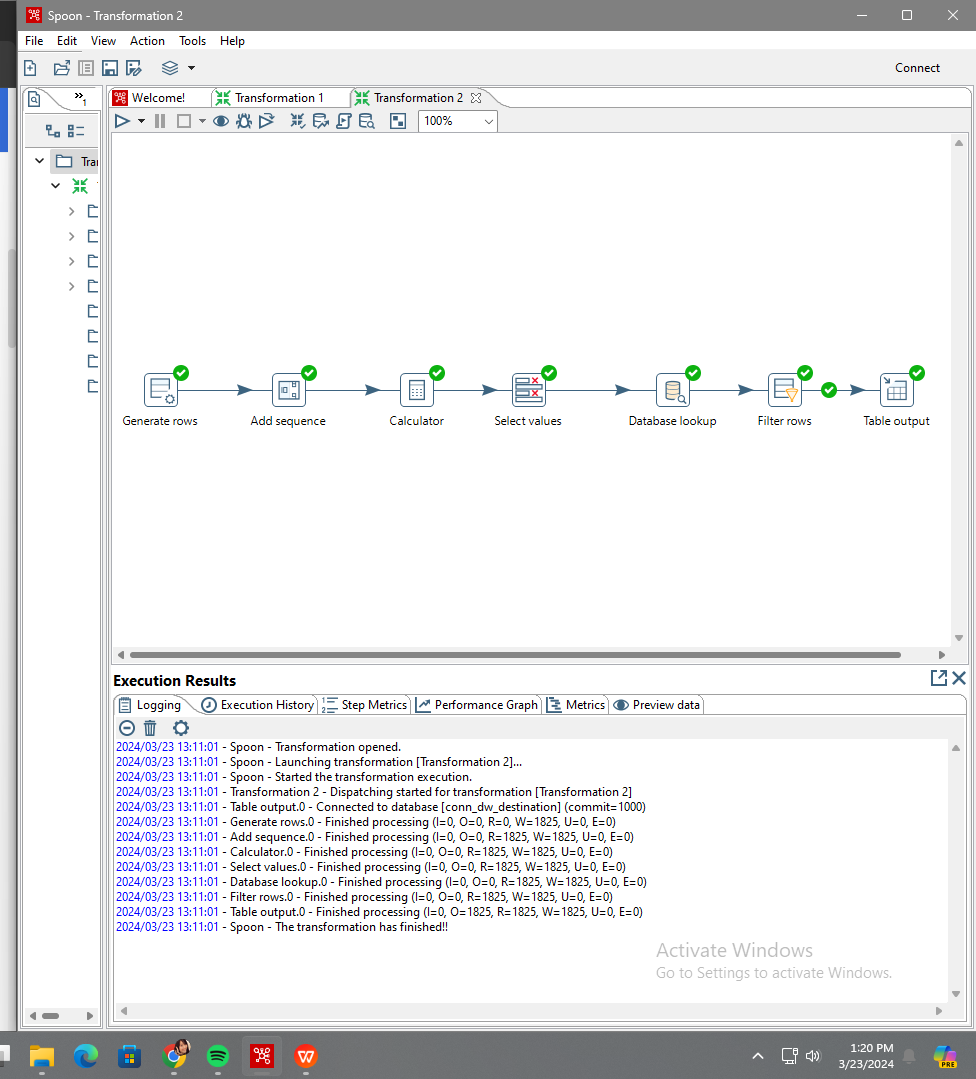


Pada tahapan selanjutnya, untuk membuat tabel dimensi dimDate , maka diperlukan generate data tanggal. Data tanggal yang disiapkan pada tabel dimDate menyesuaikan dengan proses bisnis yang berjalan.

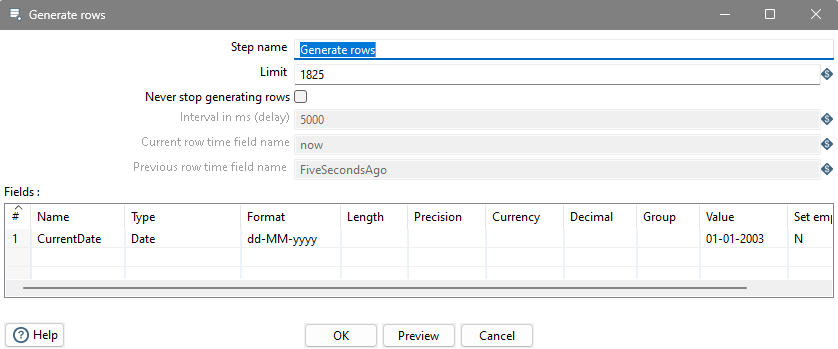
Proses bisnis pada LegendVehicle adalah 5 tahun. Sehingga data pada tabel dimdate yang harus tersedia adalah tanggal selama 5 tahun. Mulai dari 1 Januari 2023

1. Buka PDI Spoon. Buat Transformation baru -> ****File - New - Transformation.****
2. Drag and Drop beberapa objek yaitu:

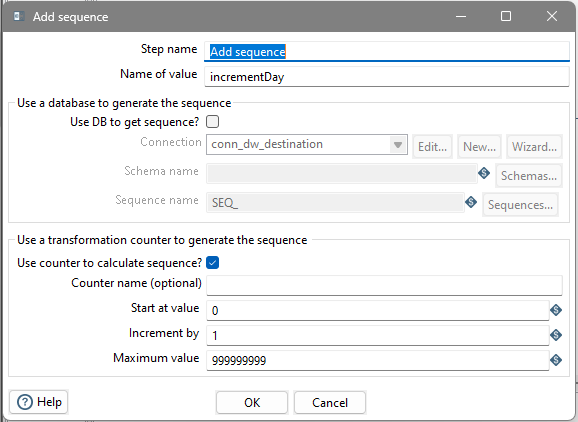
* ****Generate Rows:**** digunakan untuk membuat baris data baru.
* ****Add Sequence:**** digunakan untuk membuat sequence, dalam hal ini membuat data di setiap harinya.
* ****Calculator:**** digunakan untuk menjumlahkan hari dan mengambil data tahun, bulan dan hari.
* ****Select Values:**** digunakan untuk memilih field yang digunakan.
* ****Database Lookup:**** digunakan untuk melihat dan memastikan bahwa data yang akan dimasukkan kedalam tabel dimDate tidak kembar atau sama dengan data yang ada pada tabel dimDate itu sendiri.
* ****Filter Rows:**** digunakan untuk mengambil data yang belum ada pada table dimDate setelah dicek sebelumnya.
* ****Table Output:**** digunakan untuk menyimpan data pada tabel tujuan (dimDate).



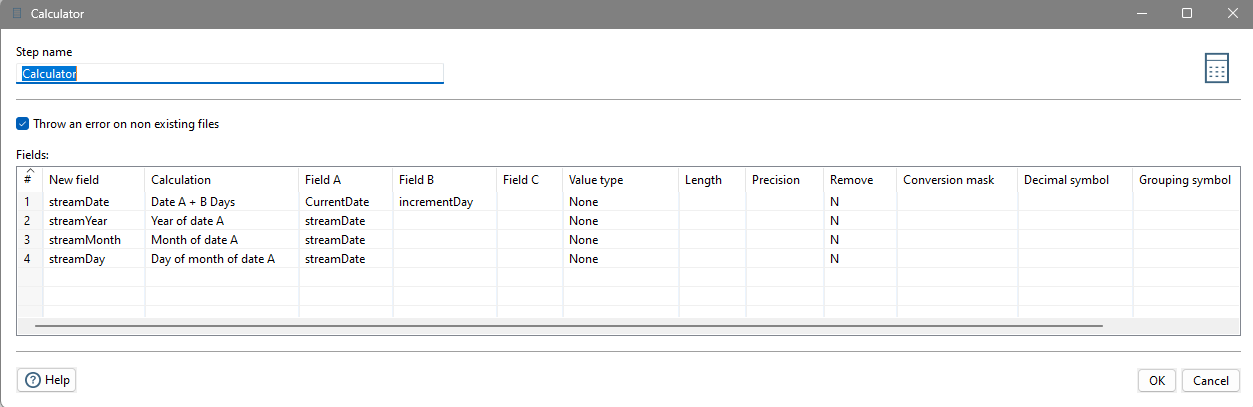
1. Konfigurasi pada ****Generate Rows**** adalah merubah ****limit**** menjadi ****1825**** dimana memiliki arti bahwa data yang akan dibuat sebanyak 1825 data. 1825 merupakan jumlah hari dalam 5 tahun ( 365 hari x 5 tahun ).
2. Membuat fields baru bernama ****CurrentDate**** dengan ****type**** data ****Date**** dan ****format dd-MM-yyyy**** serta ****value**** awal ****01-01-2003****.



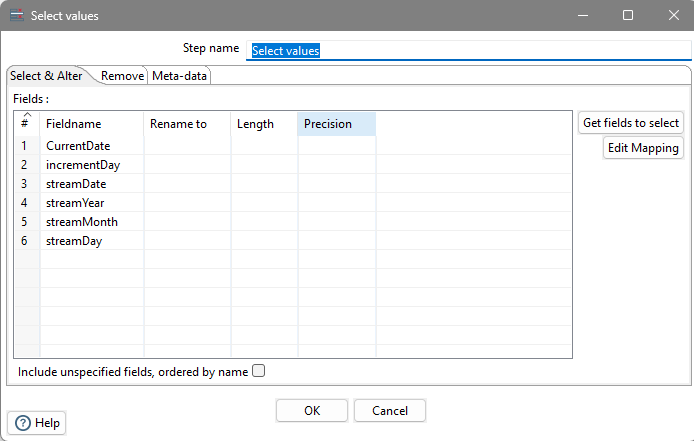
1. Hubungkan output dari ****Generate Rows**** menuju ****Add Sequence****.
2. Konfigurasi pada ****Add Sequences**** adalah merubah ****Name of value**** menjadi ****incrementDay**** dengan ****start value**** bernilai ****0**** dan ****increment by**** bernilai ****1****

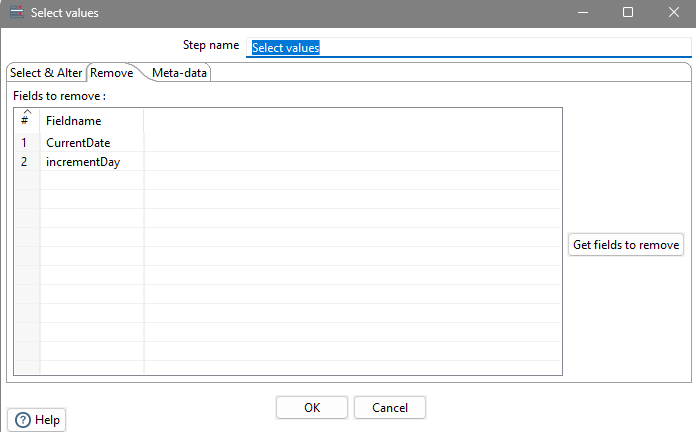


1. Hubungkan output dari ****add sequences**** menuju ****calculator.****
2. Konfigurasi pada calculator dengan membuat fields baru sebagai berikut:
   1. ****streamDate**** merupakan kalkulasi dari ****CurrentDate + incrementDay****
   2. ****streamYear**** merupakan ****Year**** dari ****streamDate****
   3. ****streamMonth**** merupakan ****Month**** dari ****streamDate****
   4. ****streamDay**** merupakan ****Day of month**** dari ****streamDate****

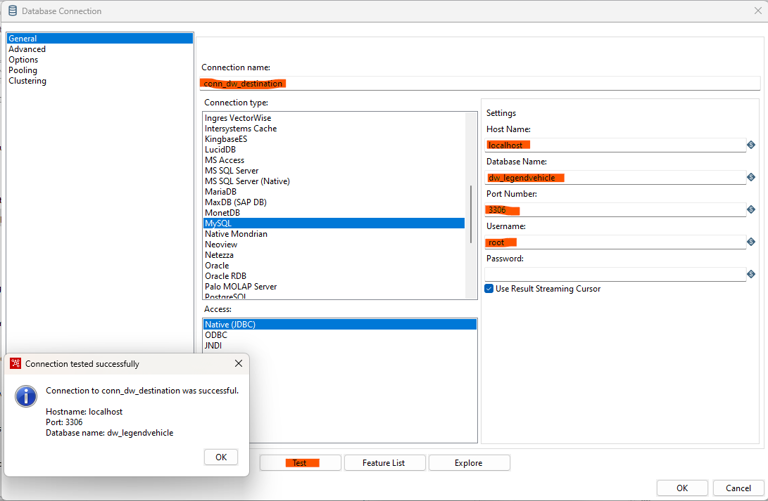


1. Hubungkan output dari ****calculator**** menuju ****Select values****
2. Konfigurasi pada ****select values**** adalah dengan menekan tombol ****Get fields to select**** pada tab ****Select & Alter****. Secara otomatis semua fields dari data input akan muncul.
3. Dikarenakan tidak semua fields digunakan, maka pada tab ****Remove**** diisikan fields ****CurrentDate**** dan ****incrementDay**** dikarenakan kedua fields tersebut tidak digunakan.

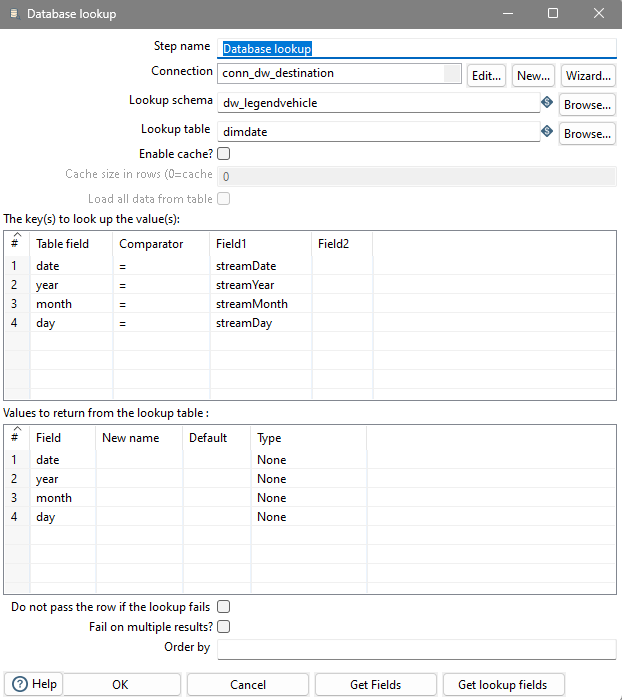




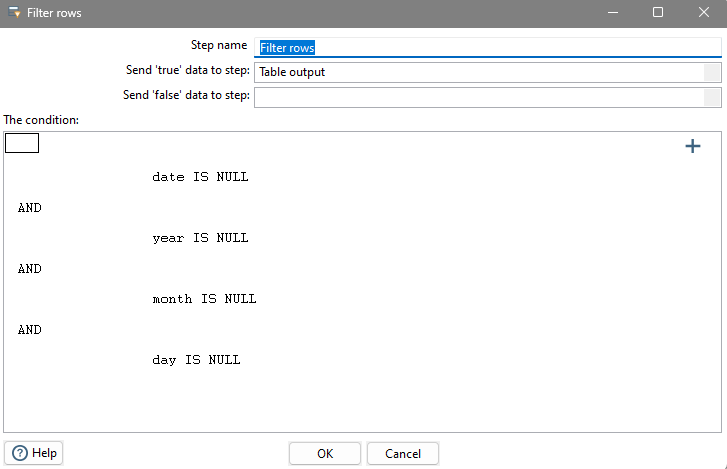
1. Hubungkan output select values menuju database lookup.
2. Sebelum melakukan konfigurasi pada ****database lookup****, buatlah koneksi terlebih dahulu pada database melalui ****File - New - Database Connection****. Gunakan ****Connection type MySQL**** dengan ****host name , database name, port number, username dan password**** sesuai konfigurasi MySQL pada device masing-masing. beri nama ****connection name**** tersebut dengan nama ****conn\_dw\_destination****.



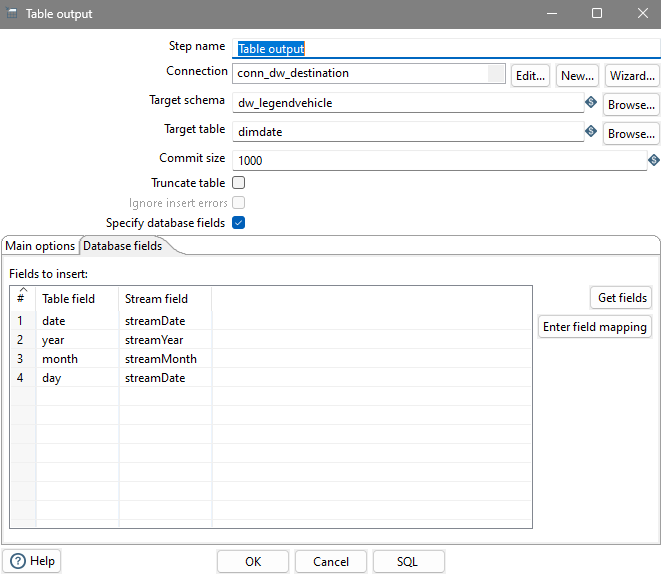
1. Konfigurasi pada ****database lookup**** adalah dengan memberikan ****connection**** dengan koneksi yang sudah dibuat pada step sebelumnya. dengan ****schema**** nama database yang digunakan dan ****tabel dimdate**** yang telah dibuat pada langkah pertama.
2. Field yang akan dicek untuk melihat kesamaan isi datanya agar tidak kembar adalah:
   1. field ****date**** pada table ****dimdate**** dengan field ****streamDate****
   2. field ****year**** pada table ****dimdate**** dengan field ****streamYear****
   3. field ****month**** pada table ****dimdate**** dengan field ****streamMonth****
   4. field ****day**** pada table ****dimdate**** dengan field ****streamDay****
3. Field yang akan di ****retrive**** adalah field yang ada pada table ****dimDate**** yaitu ****date, year, month,**** dan ****day.****



1. Hubungkan output dari ****database lookup**** dengan ****filter rows****
2. Konfigurasi pada ****filter rows**** adalah dengan melakukan konfigurasi ****output true data**** pada ****table output****. Pada bagian ini data yang tidak memiliki kesamaan pada tahapan sebelumnya akan dicek dimana jika ****fields Stream**** tidak memiliki kesamaan dengan ****field dimDate****, maka ****field dimDate**** tersebut akan bernilai ****null****. Pada pernyataan kondisi tuliskan ( ****date is null and year is null and month is null and day is nu****ll)

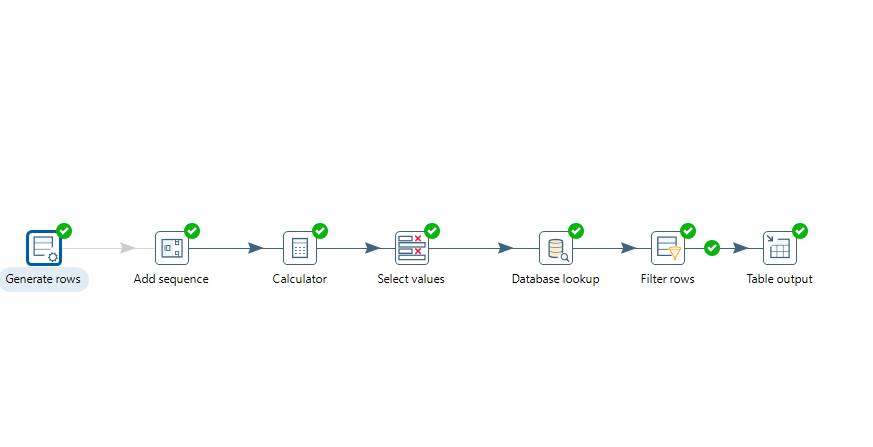


1. Hubungkan output dari ****filter rows**** menuju ****table output****.
2. Konfigurasi pada ****table output**** adalah memberikan koneksi pada ****conn\_dw\_destination**** dengan ****schema dw\_legendvehicle**** dan table ****dimdate****.
3. Aktifkan ****specify database fields****.
4. Pada tab ****Database fields****, mapping data input ****streamDate, streamYear, streamMonth**** dan ****streamDay**** dengan fields yang ada pada ****dimDate****. Pada tahapan ini akan dilakukan insert data menuju tabel ****dimDate****.



1. cek isi table dimdate pada database. Jika sukses maka pada table dimdate akan terisi 1825 data.





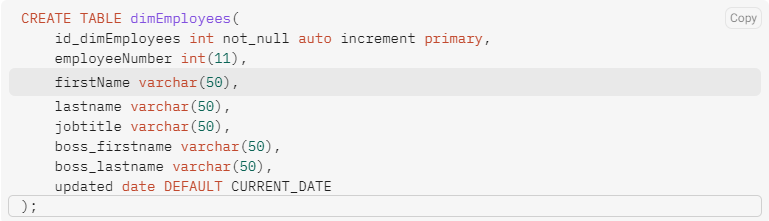
## Tugas 1

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.

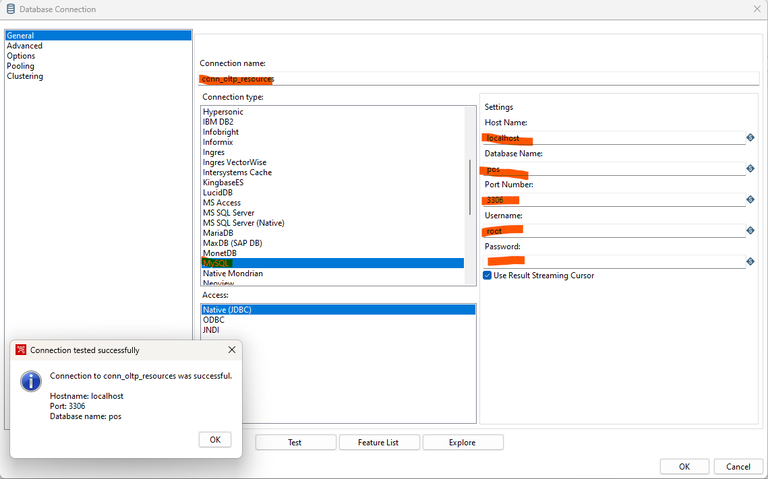
|  |  |  |  |
| --- | --- | --- | --- |
| **Proses Objek** | **SS data input** | **SS data output** | **Keterangan** |
| Generate rows |  |  | Object generate row akan mengenerate 1825 baris data dengan value bertipe date dan konstan 01-01-2003. |
| Add Sequences |  |  | Add sequences akan membuat sebuah field baru “incrementDay” kemudian akan diisi dengan value mulai dari 1 dengan increment(peningkatan) dengan nilai 1 sehingga akan menjadi 1, 2, 3, dst hingga baris terakhir. |
| Calculator |  |  | Pada calculator akan mengambil field currentDate lalu menjumlahkannya dengan incrementDay menjadi streamDate. Ini tidak akan membuat tanggal melebihi waktu aslinya karena calculation sudah diterapkan sebagai date. Kemudian |
| Select values |  |  | Pada bagian ini, select values akan memilih semua field pada input dan kemudian membuang beberapa field yang sudah tidak diperlukan yaitu currentDate dan incrementDay. |
| Database lookup |  |  | Database lookup akan membuat field baru yang akan disamakan dengan field yang ada di database. Field baru ini akan di assign ke field stream yang sudah dihasilkan sebelumnya. Kemudian database lookup akan Bersiap untuk menambahkan data ke database. |
| Filter rows |  |  | Filter rows akan mengambil input dari database lookup untuk di filter apakah field Stream tidak memiliki kesamaan dengan field dimDate, maka field dimDate tersebut akan bernilai null. |
| Table Output |  |  | Table output akan melakukan insert data pada database. |

## Sub Chapter 2 (Dimensi Pegawai)

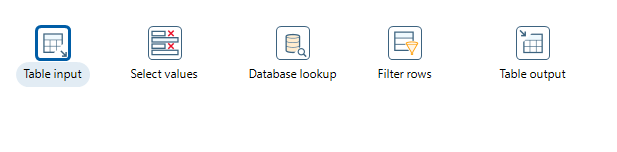
1. Buatlah tabel ****dimPegawai**** pada ****dw\_legendVehicle.****



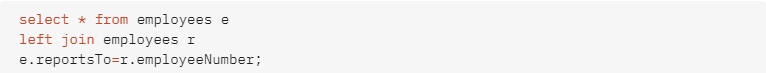
1. Pada PDI Spoon buatlah koneksi baru dengan nama conn\_oltp\_resources yang menghubungkan dengan database oltp. sesuaikan hostname, database name, port number, username dan password dengan keadaan pada device masing-masing.

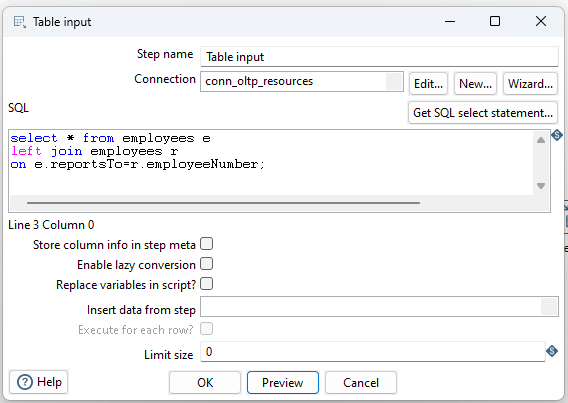


1. Drag and drop beberapa objek sebagai berikut:
   1. Table input: digunakan mengambil data dari database OLTP.
   2. Select values: memeilih field yang digunakan untuk proses Transform dan Load.
   3. Database lookup: digunakan untuk melihat data pada tabel dimEmployees untuk memastikan data tidak kembar
   4. Filter rows: digunakan untuk memilih data stream yang masih belum ada apada tabel dimEmployees.
   5. Table output: Memasukkan data ke dalam tavle dimEmployees

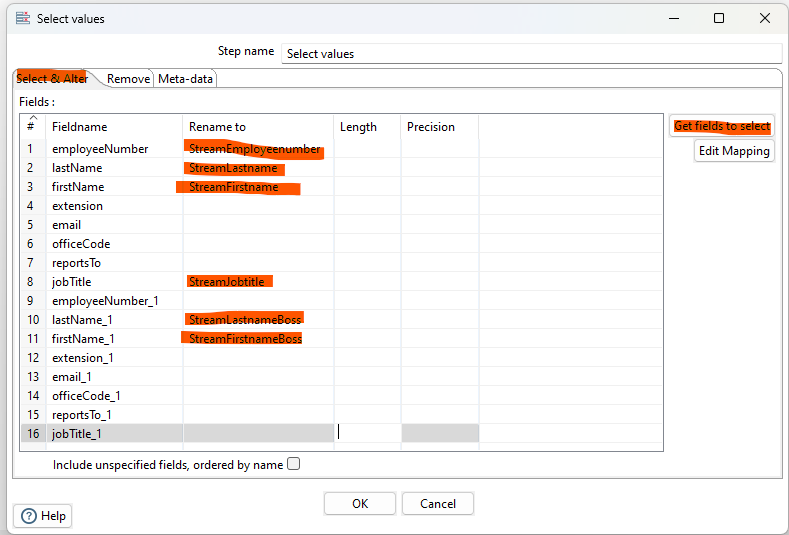


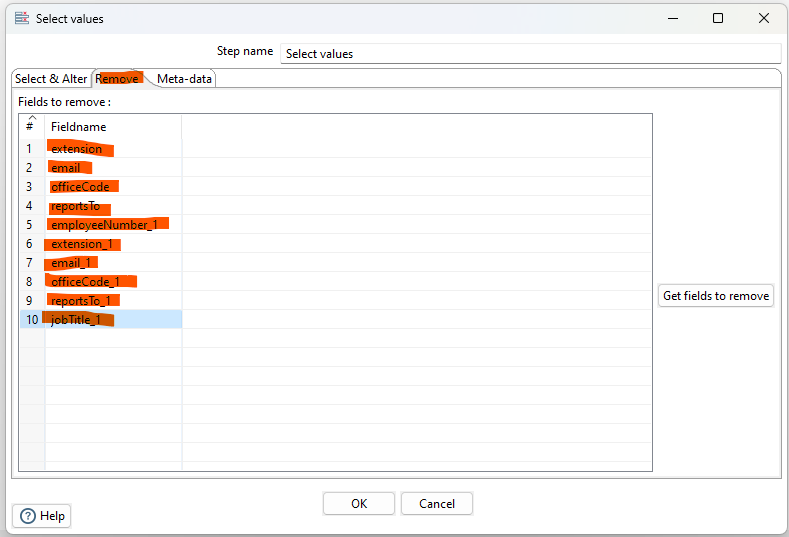
1. Konfigurasi pada table input dengan menghubungkan Connection pada konesi conn\_oltp\_resources. Untuk mengambil data sumber menggunakan query dibawah ini.



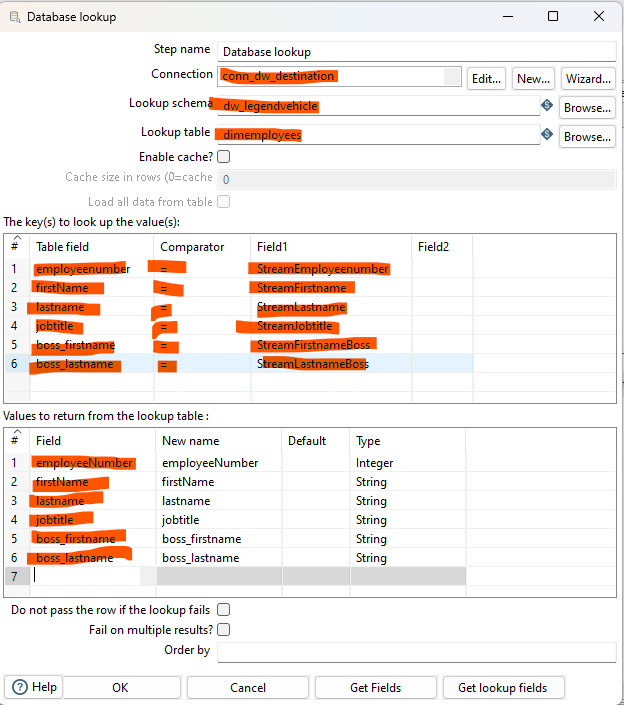


1. Hubungkan output table input pada select values.
2. Konfigurasi pada Select values yaitu mengambil data dari field employeenumber, lastname, firstname, jobtitle , lastname\_1 dan firstname\_1 sebagai data stream yang digunakan pada proses ETL pada tab select & alter.
3. Hilangkan field lain yang tidak digunakan pada tab remove.

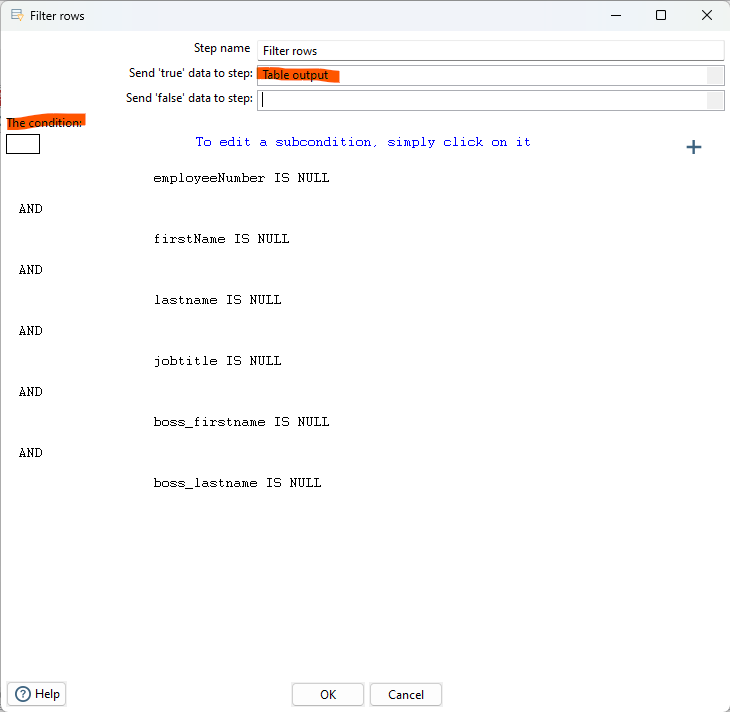




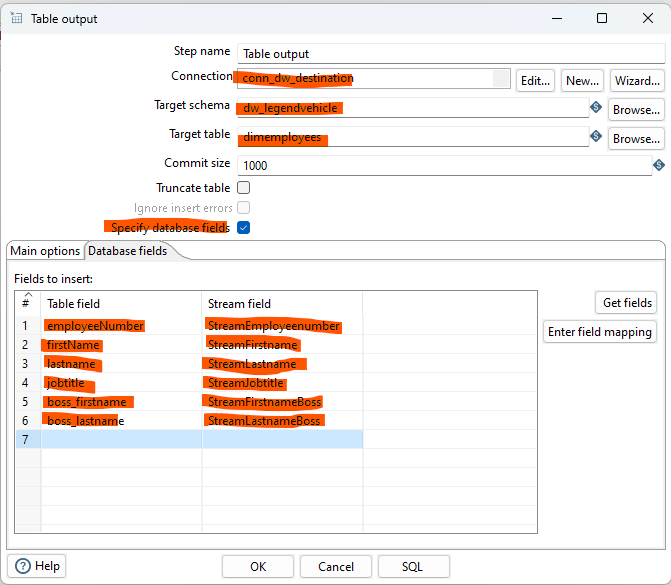
1. Hubungkan output select values pada database lookup
2. Konfigurasi pada database lookup adalah dengan menghubungkan koneksi pada conn\_dw\_destination dengan table lookup dimEmployees yang telah dibuat pada tahap pertama.
3. Field yang di lookup adalah field pada tabel dimEmployees dengan field stream input dari OLTP. sedangkan field yang di retrieve adalah field dari dimEmployees itu sendiri. Jika tidak ada data yang sama maka akan muncul null.



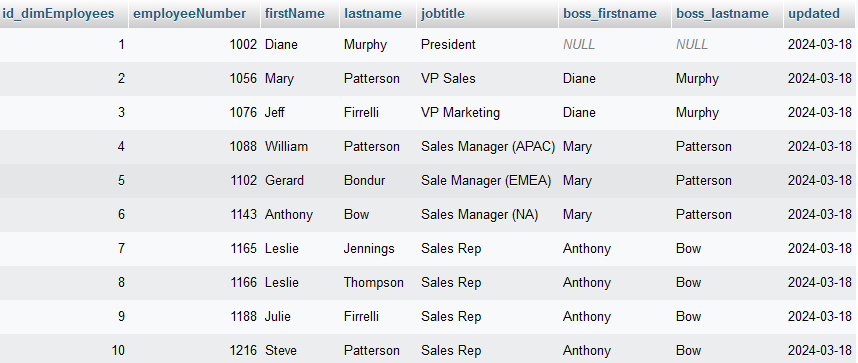
1. Hubungkan output databse lookup dengan filter rows.
2. Pada filter rows berikan kondisi field yang null pada field dimemployees untuk dimasukkan pada proses selanjutnya. Hal itu menandakan bahwa data stream belum memiliki kesamaan pada data di dimemployees.

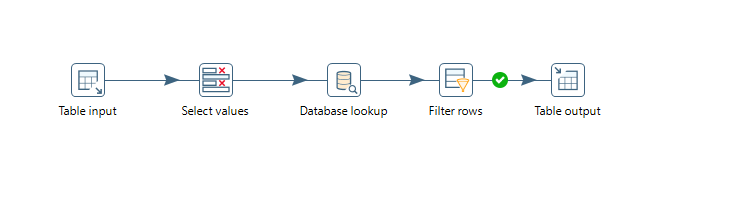


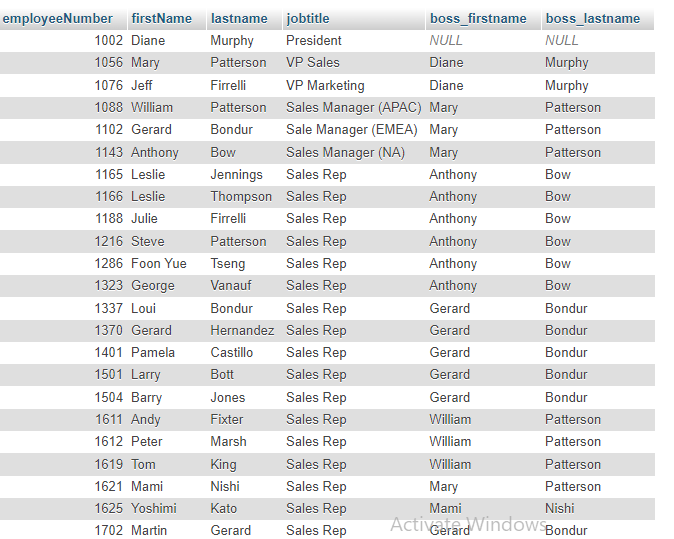
1. Hubungkan output dari filter rows dengan table output.
2. Pada table output, gunakan connection conn\_dw\_destination untuk memasukkan data pada tabel dimemployees.
3. Aktifkan specify databse fields, dan mapping data stream input dari oltp terhadap field yang ada pada dimemployees.



1. jika proses keseluruhan berhasil maka tabel dimemployees akan terisi data pegawai dari database OLTP.





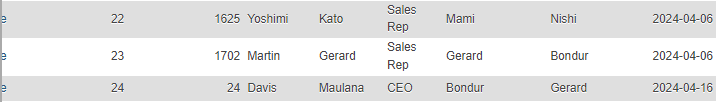


## Tugas 2

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.

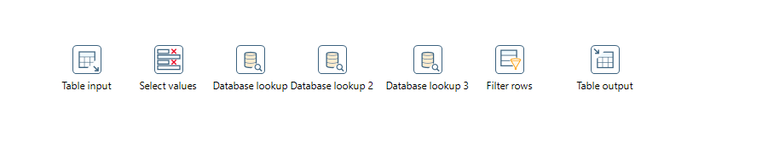
|  |  |  |  |
| --- | --- | --- | --- |
| Proses Objek | SS data input | SS data output | Keterangan |
| Table Input |  |  | Objek ini digunakan untuk mengambil data dari database classicmodels yang disini dijadikan sebagai data OLTP. |
| Select Values |  |  | Objek ini akan memilih field yang akan digunakan dan membuang field yang tidak diperlukan. |
| Database Lookup |  |  | Melakukan pengecekan pada table dimEmployee untuk memastikan tidak ada data yang sama. |
| Filter Rows |  |  | Melakukan filter terhadap data stream yang belum dimiliki oleh table dimEmployee. |
| Table Output |  |  | Memasukkan data dari table employee ke table dimEmployee disesuaikan dengan field yang ada juga. |

1. Jika proses itu di ulangi ( di run kembali ) apakah data akan redudant?
   1. Ya, akan terjadi duplikasi atau penumpukan data yang sama pada table dimEmployee
2. Tambahkan nama anda pada table employee di OLTP. jalankan kembali transformasi ini. Amati hasilnya, apa yang terjadi?

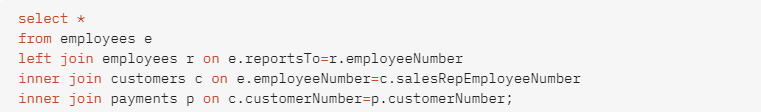


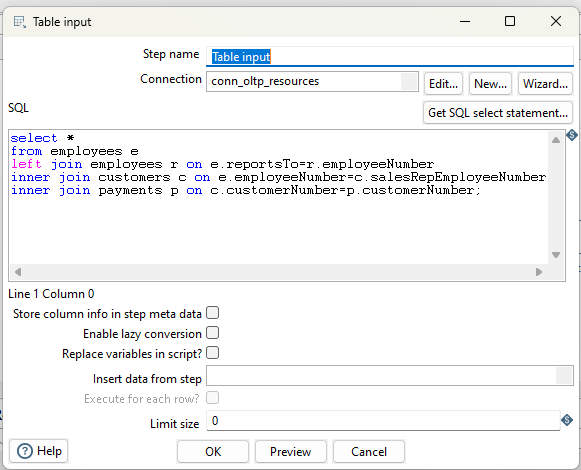
## Sub Chapter 3 (Fakta Pembayaran)

1. Buatlah sebuah tabel pada databse OLAP yang telah dibuat (database dw\_legendVehicle).
   1. nama tabel: FactOmset
   2. Field: id\_dimEmployees int FK tabel dimEmployee
   3. Field: id\_dimDate int FK tabel dimDate
   4. Field: amount decimal(10.2)
2. Buat Transformation baru pada PDI Spoon. Gunakan objevt-object yang dihubunugkan sebagai berikut:
   1. Table Input: Dignakan untuk mengambil data transaksi dari OLTP
   2. Select values: digunakan untuk memilih field yang akan digunakan untuk OLAP
   3. Database lookup (1) : digunakan untuk mencocokkan data pada tabel dimEmployee untuk mengambil id\_dimEmployee.
   4. Database lookup (2) : digunakan untuk mencocokkan data pada tabel dimDate untuk mengambil id\_dimDate.
   5. Databse lookup (3): digunakan untuk mencocokkan data pada tabel factomset untuk melihat data yang sama atau tidak.
   6. Filter rows: digunakan untuk memiih data yang sudah ada pada tabel factomset tidak dimasukkan lagi.
   7. Table output: digunakan untuk memasukkan data pada tabel factOmset.

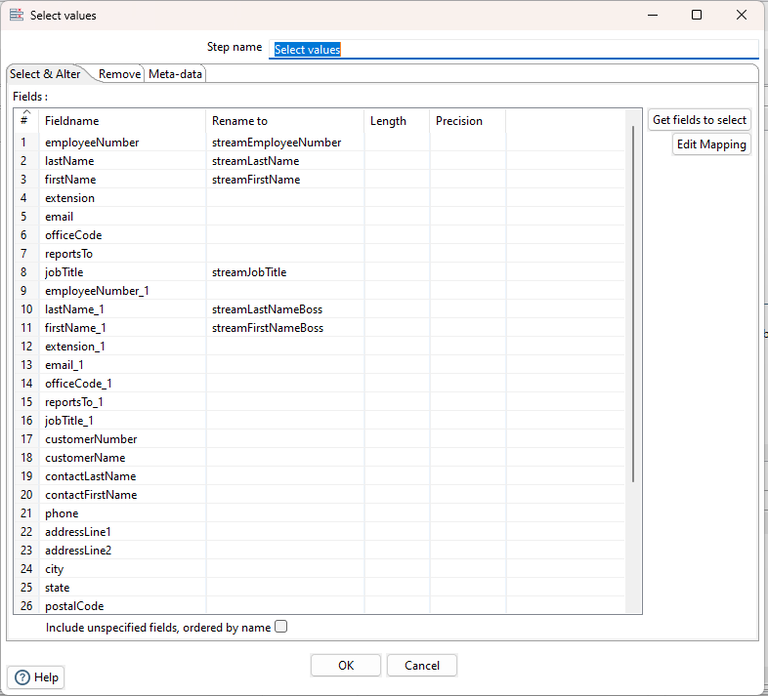


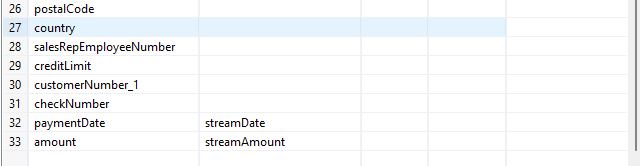
1. Konfigurasi pada Table input adalah untuk mendapatkan data dari OLTP sesuai dengan hasil dari query berikut. query tersebut akan menampilkan seluruh isi field dari employee hingga transaksi yang didapat dilihat dari payments yang didapat.

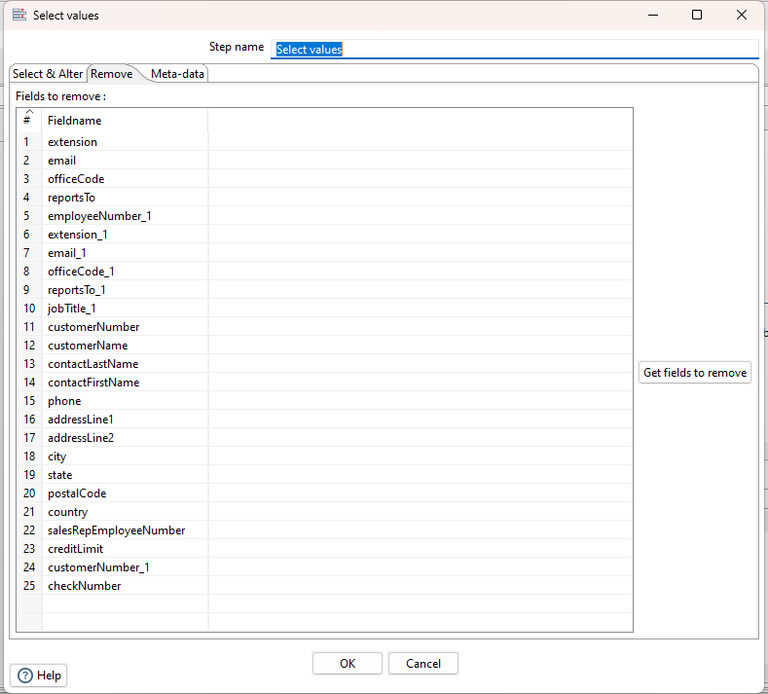




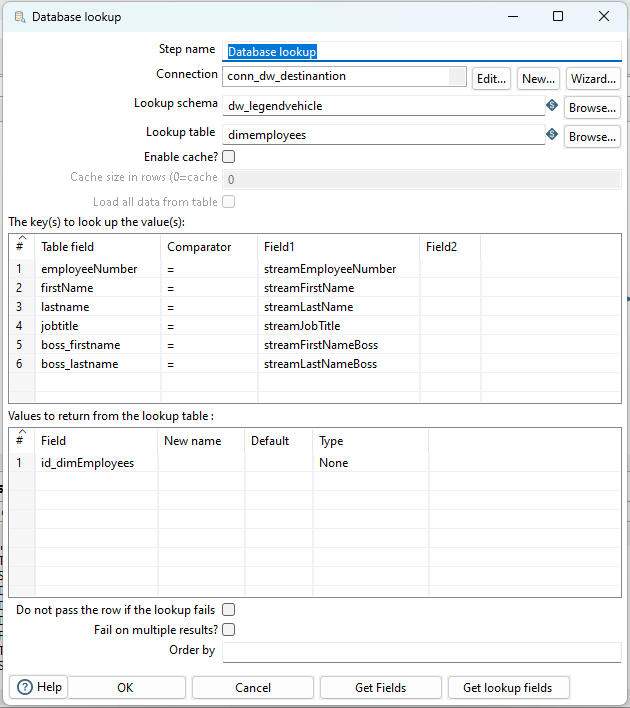
1. Konfigurasi pada Select Values adalah untuk menghapus semua field kecuali employeeNumber, lastname dari employee, firstname dari employee, jobTitle dari employee, lastname dari manager, firstname dari manager, payment date dan amount.



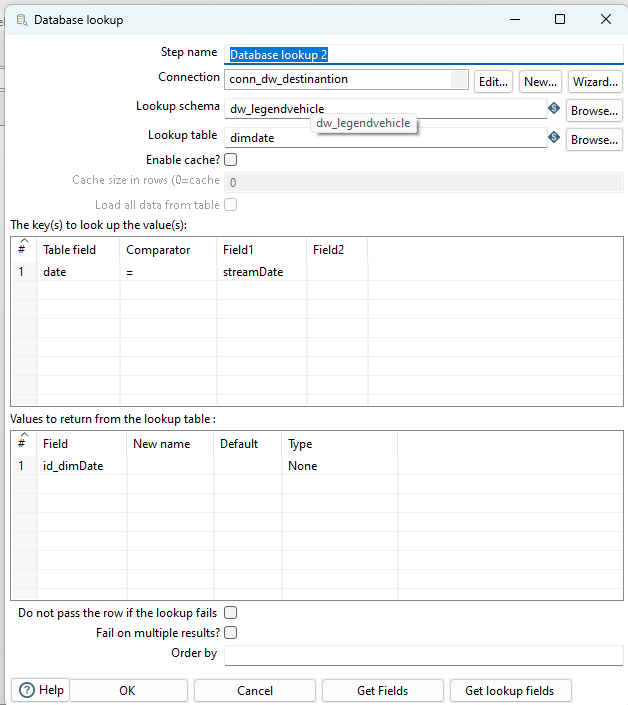




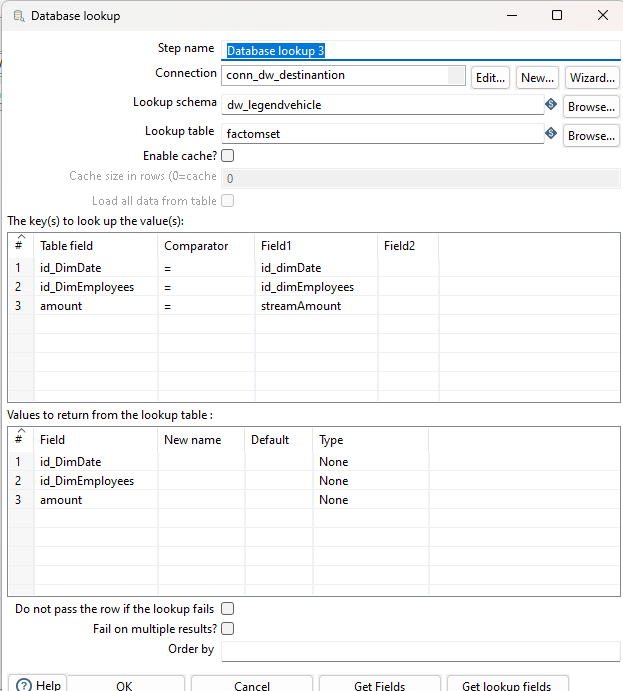
1. Konfigurasi pada tabel database lookup adalah dengan melakukan komparasi field stream (output dari proses sebelumnya) dengan field isi data pada tabel dimEmployees. Jika data tersebut cocok maka akan diambil id\_dimEmployees nya.



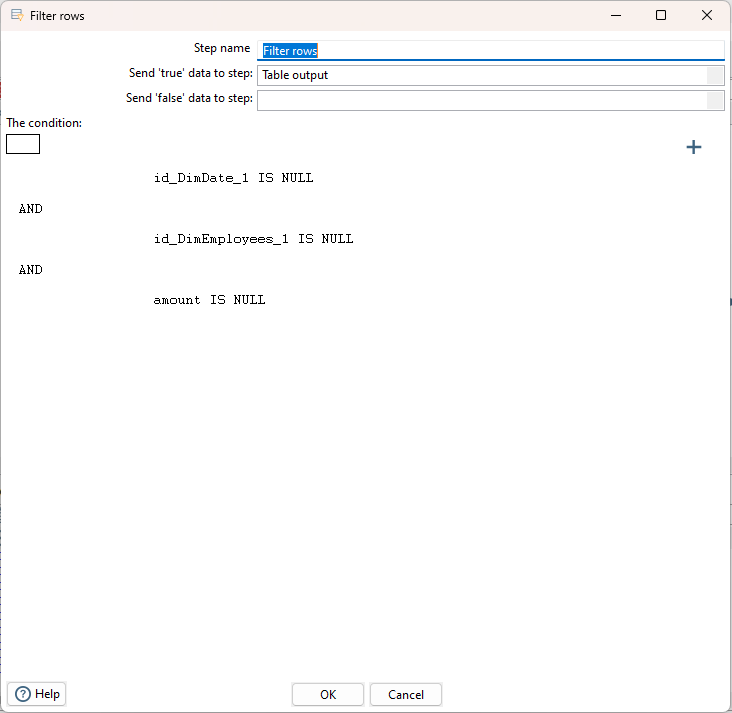
1. Konfigurasi pada tabel database lookup yang kedua adalah dengan melakukan komparasi field stream (output dari proses sebelumnya untuk field date) dengan field isi data pada tabel dimDate. Jika data tersebut cocok maka akan diambil id\_dimDate nya.



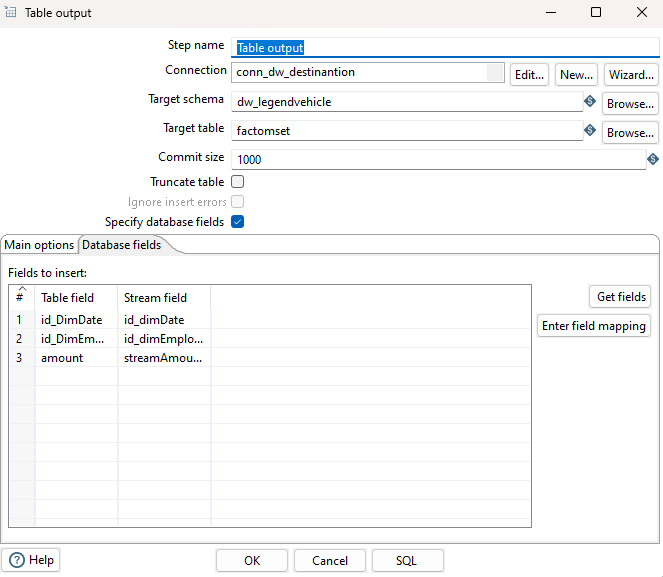
1. Konfigurasi pada tabel database lookup yang ketiga adalah dengan melakukan komparasi id\_dimDate dan id\_dimEmployees yang diambil dari proses lookup sebelumnya dengan field isi data pada tabel factOmset. Jika data tersebut cocok data tidak akan dimasukkan dalam tabel factOmset.



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



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



1. Jalankan proses transformation tersebut.

## Tugas 3

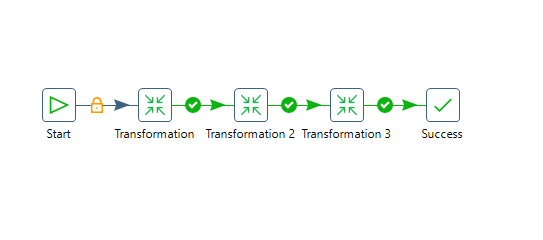
1. Buka preview tab pada execution result area di setiap proses object. amati input dan output data yang ada. bandingkan di setiap prosesnya. jelaskan perbedaan disetiap prosesnya.

|  |  |  |  |
| --- | --- | --- | --- |
| Proses Objek | SS data input | SS data output | Keterangan |
| Table Input |  |  | Table input digunakan untuk mendapatkan data dari OLTP sesuai dengan hasil dari query |
| Select Values |  |  | Select Values digunakan 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. |
| Database Lookup(1) |  |  | Proses ini mencocokan field stream dengan field isi data di table dimEmployees. Jika cocok diambil id\_dimEmployees nya |
| Database Lookup(2) |  |  | Proses ini mencocokan field stream dengan field isi data di table dimDate. Jika cocok diambil id\_dimDate nya |
| Database Lookup(3) |  |  | Proses ini membandingkan id\_dimDate dan id\_dimEmployees yang diambil dari proses lookup sebelumya dengan field isi data pada table factOmset. Jika cocok data tidak akan dimasukkan dalam table factOmset |
| Filter Rows |  |  | Proses ini melihat data belum ada pada table factOmset dengan melihat bahwa output dari lookup sebelumya bernilai NULL |
| Table Output |  |  | Proses ini melakukan mapping data output dari proses sebelumnya kedalam field pada table factOmset |

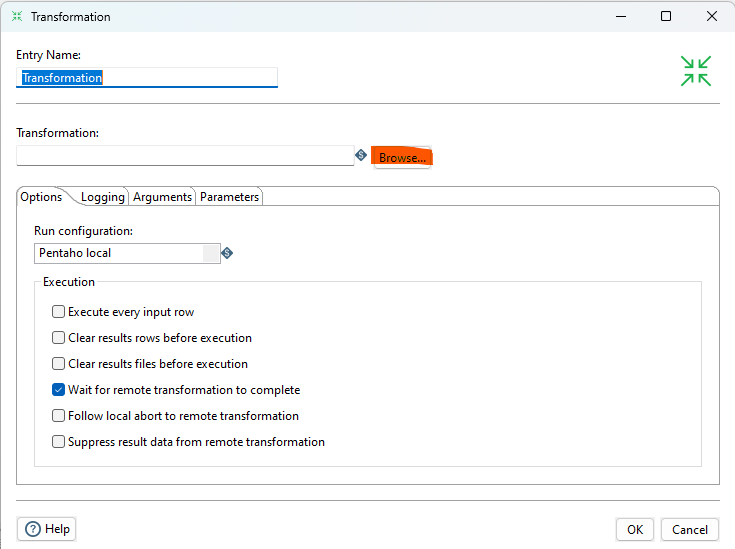
1. Jika proses itu di ulangi ( di run kembali ) apakah data akan redudant?
   1. Bisa jika proses tersebut dijalankan Kembali maka akan menginsert data Kembali dan terjadi duplicate data

## Sub Chapter 4 (Jobs)

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



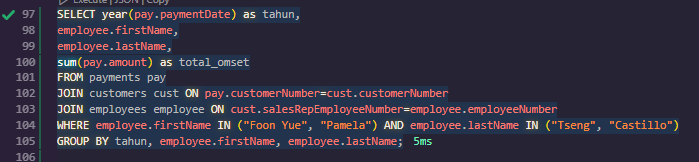
1. Konfigurasi dilakukan pada setiap transformation untuk mengambil file tranformation yang telah dibuat pada bagian A, B dan C.

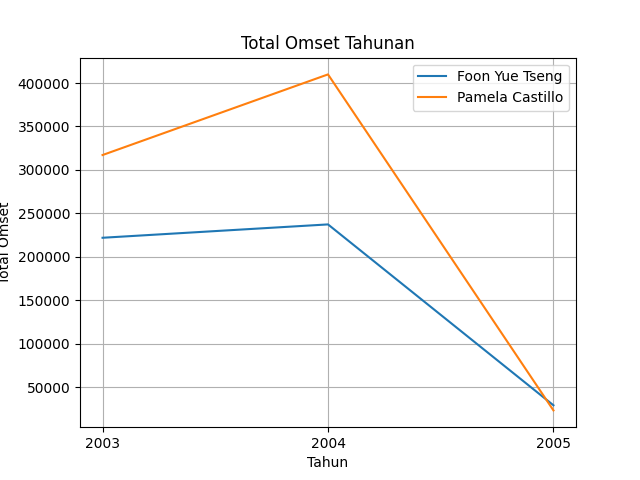


1. Jalankan jobs tersebut ketika semua transformation telah dihubungkan dengan file transformation yang telah dibuat sebelumnya.

## 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.
2. dw\_legendvehicle memiliki dimensi dimEmployees yang lebih ringkas dibandingkan dengan bi employees pada OLTP legendVehicle. Dimensi dimEmployees hanya menyertakan informasi yang diperlukan untuk analisis, sedangkan bi employees menyertakan semua informasi tentang karyawan.
3. dw\_legendvehicle memiliki dimensi dimDate yang menyertakan informasi granular tentang tanggal, seperti tahun, bulan, dan hari. OLTP legendVehicle tidak memiliki dimensi dedicated untuk tanggal.
4. dw\_legendvehicle hanya memiliki satu fakta, yaitu FactOmset, yang menyimpan informasi tentang omzet penjualan. OLTP legendVehicle memiliki beberapa fakta, seperti bi orderdetails, yang menyimpan informasi tentang detail pesanan
5. Buatlah report pertahun untuk KPI "**Jumlah omset yang didapat**" pada **Foon Yue Tseng** dan **Pamela Castillo**. Serta gambarkan grafiknya (grafik garis).





1. Jelaskan perbedaan query saat mendapatkan data pada nomor 2 dengan query pada saat Jobsheet 2!
2. Pada query nomor dua, sudah tidak lagi menggunakan dua query yang berbeda melainkan hanya satu query saja. Perbedaannya terletak di penggunaan syntax IN untuk mengambil dua data sekaligus.
3. Simpulkan dengan bahasa sendiri, apa perbedaan OLTP dan OLAP?

OLTP:

* Biasanya digunakan untuk transaksi bisnis sehari-hari, seperti penjualan atau pembaruan data.
* Data biasanya terfragmentasi dan dinormalisasi untuk menghindari redundansi dan memastikan integritas data.
* Operasi yang biasa dilakukan adalah INSERT, UPDATE, dan DELETE.

OLAP:

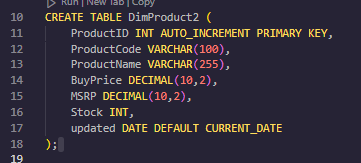
* Biasanya digunakan untuk analisis dan pelaporan data.
* Data biasanya teragregasi dan denormalisasi untuk mempercepat kinerja query.
* Operasi yang biasa dilakukan adalah SELECT (mengambil data).
* Tabel Dimensi (Dimension Tables): dimDate dan dimEmployees adalah tabel dimensi. Tabel ini biasanya memiliki atribut yang mendetail dan digunakan untuk menyaring, mengelompokkan, atau memberi label data dalam tabel fakta.
* Tabel Fakta (Fact Table): FactOmset adalah tabel fakta. Tabel ini biasanya berisi kunci asing dan ukuran. Dalam kasus ini, amount adalah ukuran dan id\_dimEmployees dan id\_dimDate adalah kunci asing.

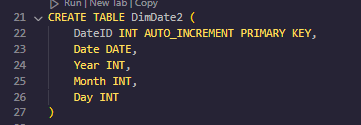
## Sub Chapter 5 (Studi Kasus)

Pak Marto merupakan komisaris dari LegendVehicle. Pak Marto butuh laporan untuk analisa dari hasil penjualan dilihat dari sisi barang / product.

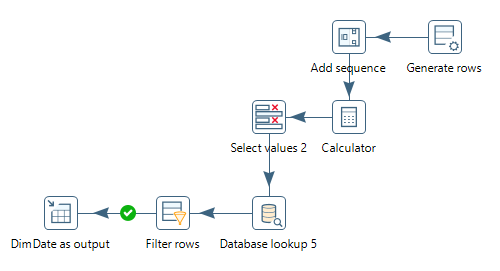
Buatlah databse OLAP tersebut untuk membantu pak marto menganalisa penjualan barang.

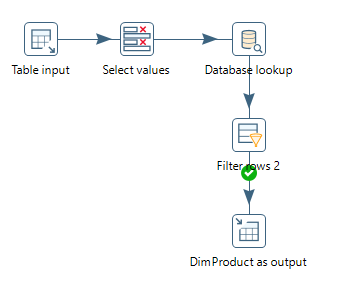
1. Pertama, yang disiapkan adalah table untuk database OLAP, berikut table yang dapat dibuat.

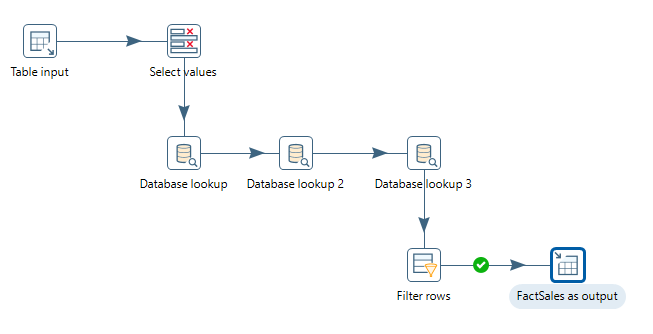




1. Kemudian membuat skema ETL menggunakan Pentaho Data Integration (PDI)







1. Hasil masing-masing table di database OLAP.

