­LAPORAN PRAKTIKUM DATA WAREHOUSE

JOBSHEET 3 : DATA BASE ANALYTICAL



DISUSUN OLEH :

AQILA NUR AZZA (2341760022)

KELAS 2A-SIB/04

**PROGRAM STUDI D-IV SISTEM INFORMASI BISNIS**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI MALANG**

Jl. Soekarno Hatta No. 9, Jattimulyo, Kec. Lowokwaru, Kota Malang, Jawa Timur

65141

**Studi Kasus**

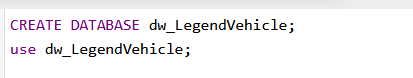
Berdasarkan studi kasus sebelumnya, Pimpinan dari LegendVehicle mulai mengenal teknologi dan berpandangan visioner. LegendVehicle akan menerapkan data warehouse pada proses bisnisnya untuk menganalisa proses transaksi yang dilihat dari pembayaran yang masuk. Proses ini digunakan untuk menganalisa KPI "jumlah omset yang dimiliki ".

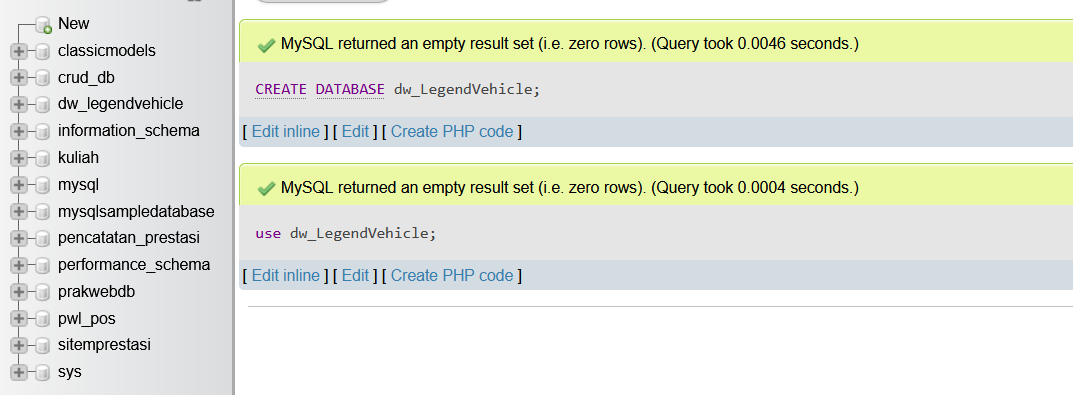
Untuk membentuk data warehouse tersebut, Data Engineer pada LegendVehicle perlu membuat sebuah database yang digunakan sebagai databasae OLAP.

Data dari OLTP yang ada akan di "ETL" kan menuju database OLAP.

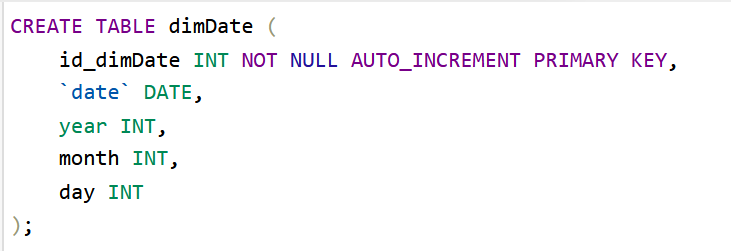
**A. Dimensi Waktu**

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**.

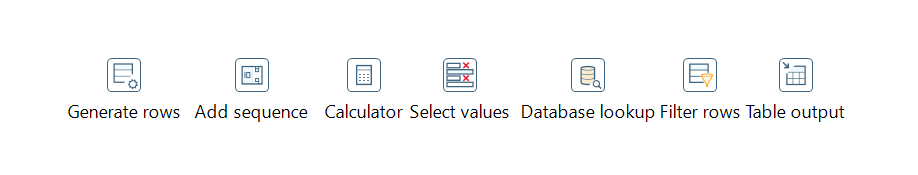


Pada tahapan selanjutnya, untuk membuat tabel dimensi dimDate , maka diperlukan generate data tanggal. Data tanggal yang disiapkan pada tabel dimDate menyesuaikan dengan proses bisnis yang berjalan.

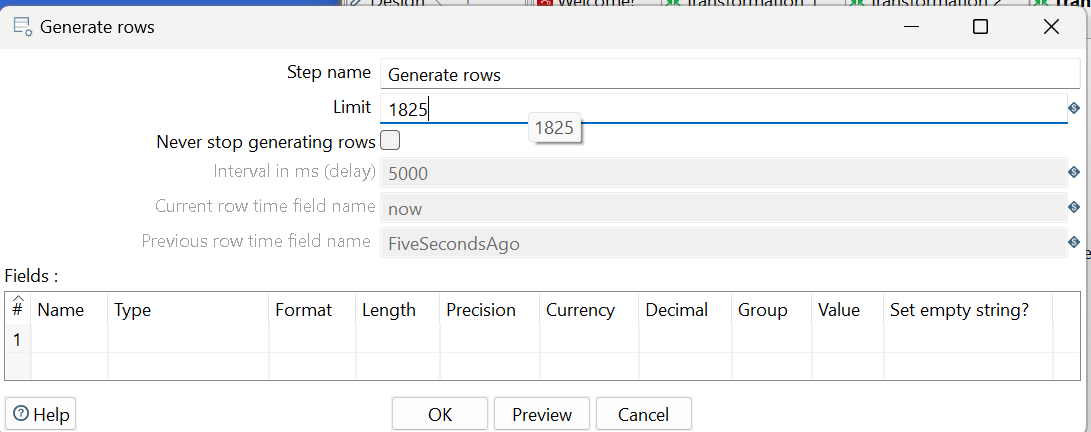
Proses bisnis pada LegendVehicle adalah 5 tahun. Sehingga data pada tabel dimdate yang harus tersedia adalah tanggal selama 5 tahun. Mulai dari 1 Januari 2023

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

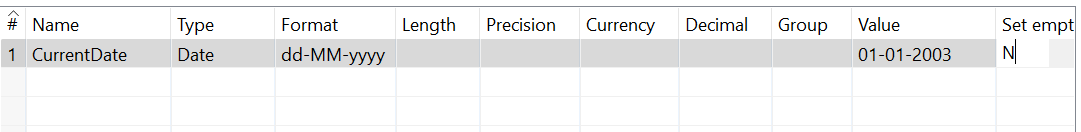
* **Generate Rows:** digunakan untuk membuat baris data baru.
* **Add Sequence:** digunakan untuk membuat sequence, dalam hal ini membuat data di setiap harinya.
* **Calculator:** digunakan untuk menjumlahkan hari dan mengambil data tahun, bulan dan hari.
* **Select Values:** digunakan untuk memilih 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).



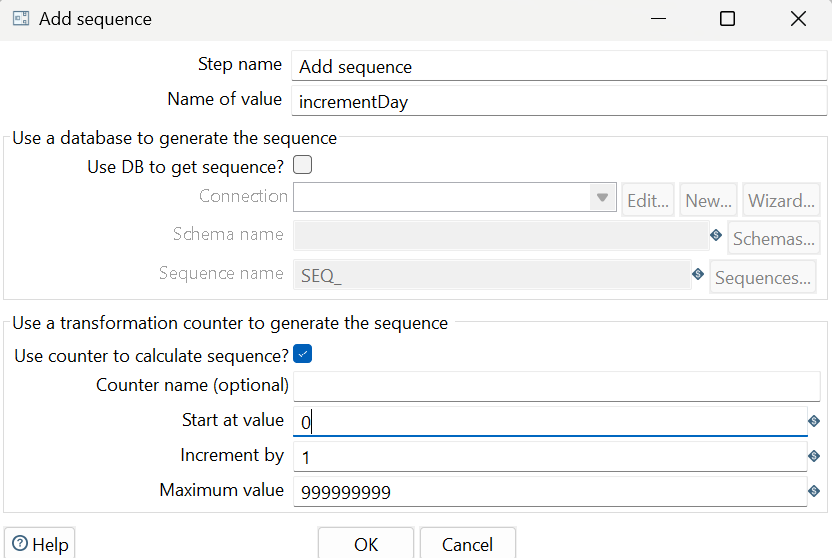
1. Konfigurasi pada **Generate Rows** adalah merubah **limit** menjadi **1825** dimana memiliki arti bahwa data yang akan dibuat sebanyak 1825 data. 1825 merupakan jumlah hari dalam 5 tahun ( 365 hari x 5 tahun ).



1. Membuat fields baru bernama **CurrentDate** dengan **type** data **Date** dan **format dd-MM-yyyy** serta **value** awal **01-01-2003**.

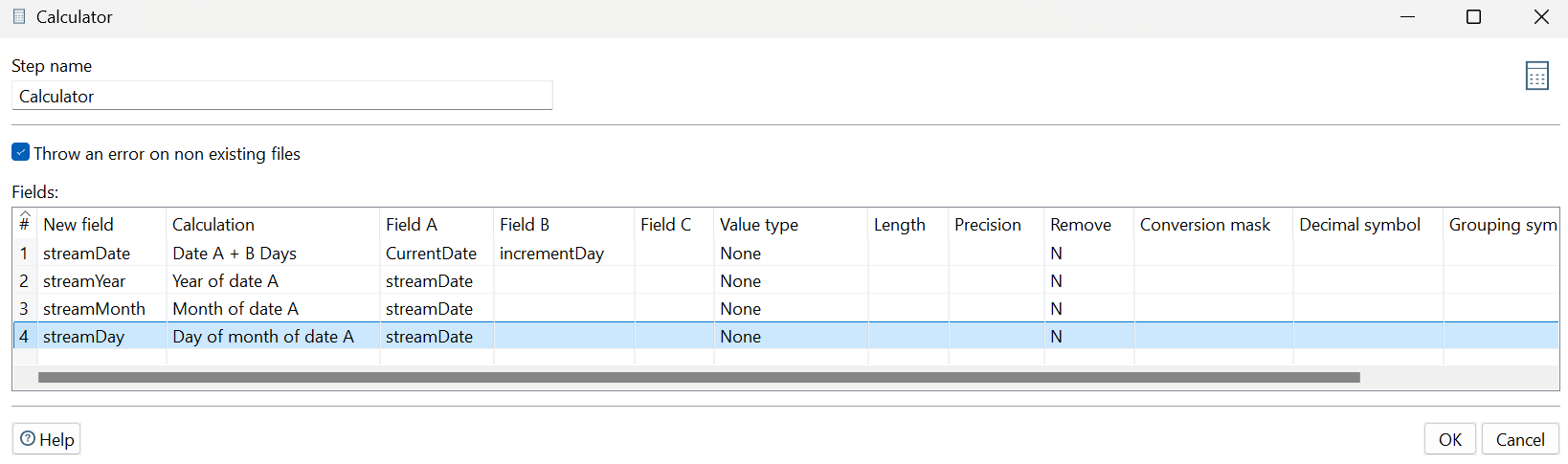


1. Hubungkan output dari **Generate Rows** menuju **Add Sequence**.
2. Konfigurasi pada **Add Sequences** adalah merubah **Name of value** menjadi **incrementDay** dengan **start value** bernilai **0** dan **increment by** bernilai **1**

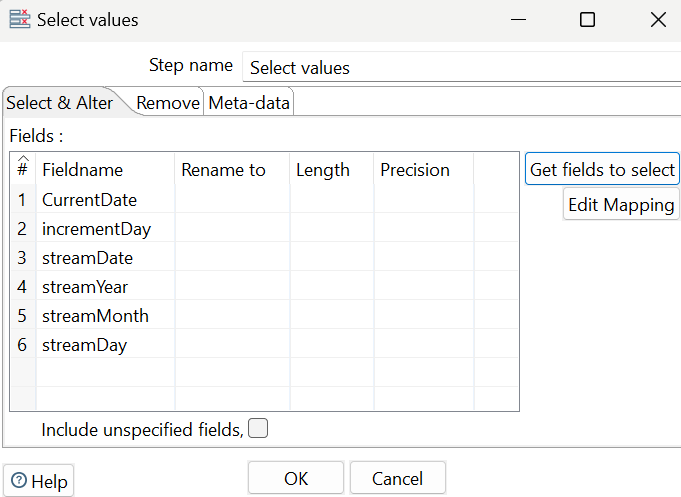
****

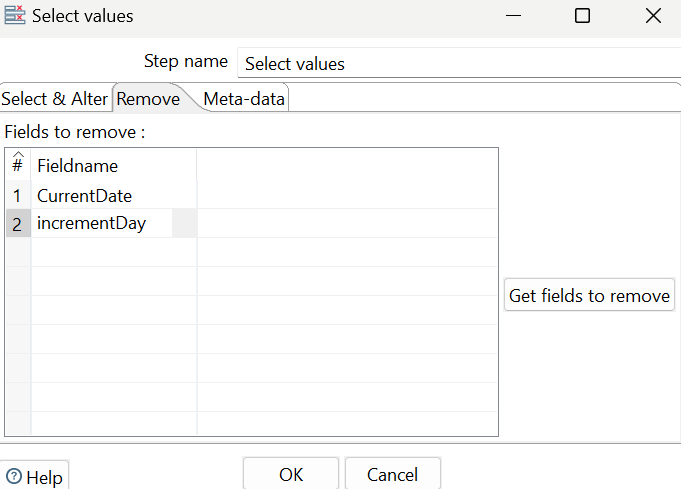
1. Hubungkan output dari **add sequences** menuju **calculator.**
2. Konfigurasi pada calculator dengan membuat fields baru sebagai berikut:

* **streamDate** merupakan kalkulasi dari **CurrentDate + incrementDay**
* **streamYear** merupakan **Year** dari **streamDate**
* **streamMonth** merupakan **Month** dari **streamDate**
* **streamDay** merupakan **Day of month** dari **streamDate**

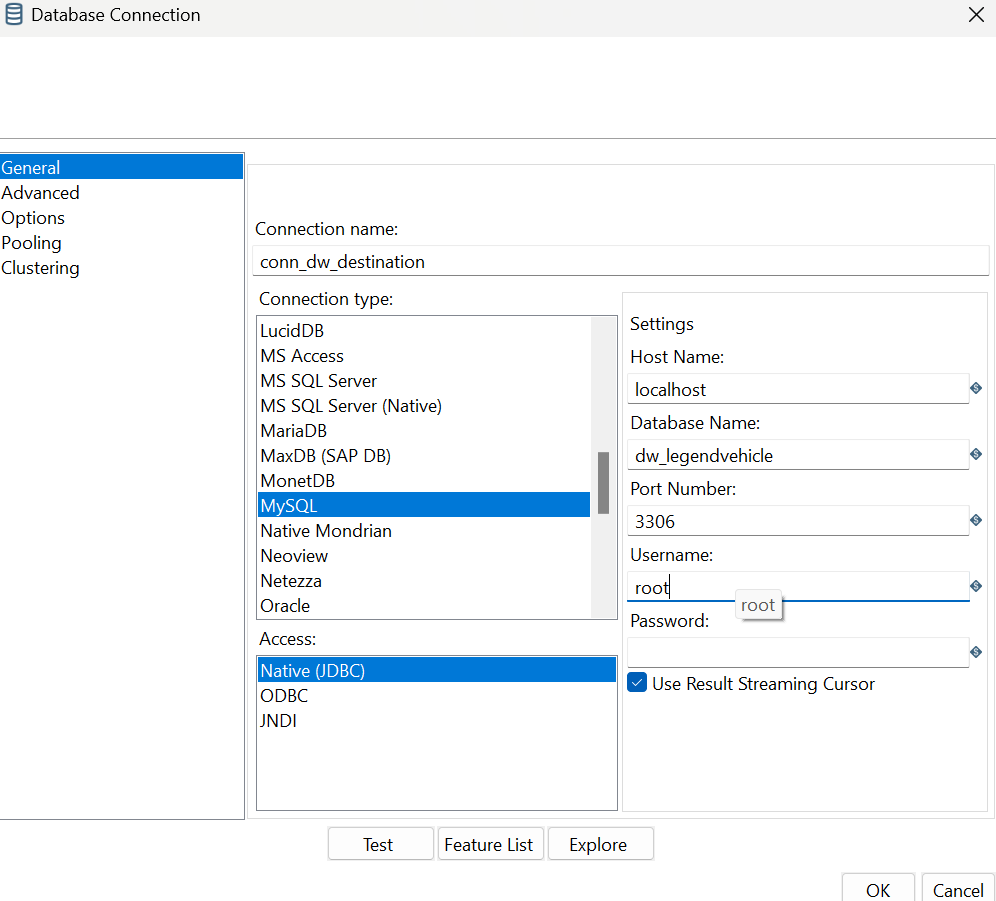
****

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.





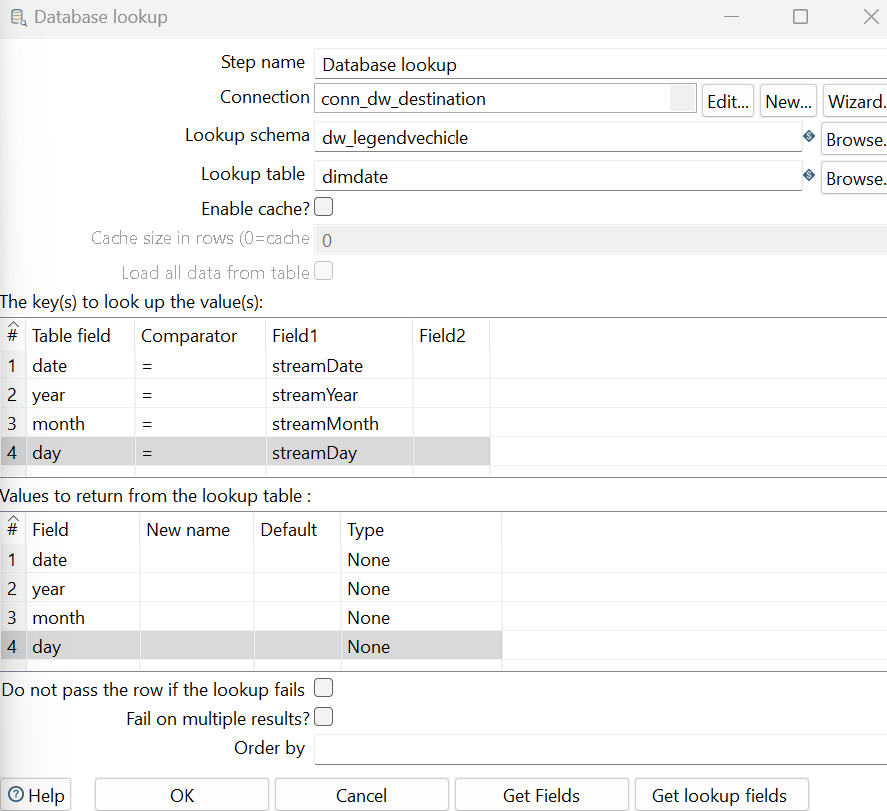
1. Hubungkan output select values menuju database lookup.
2. Sebelum melakukan konfigurasi pada **database lookup**, buatlah koneksi terlebih dahulu pada database melalui **File - New - Database Connection**. Gunakan **Connection type MySQL** dengan **host name , database name, port number, username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



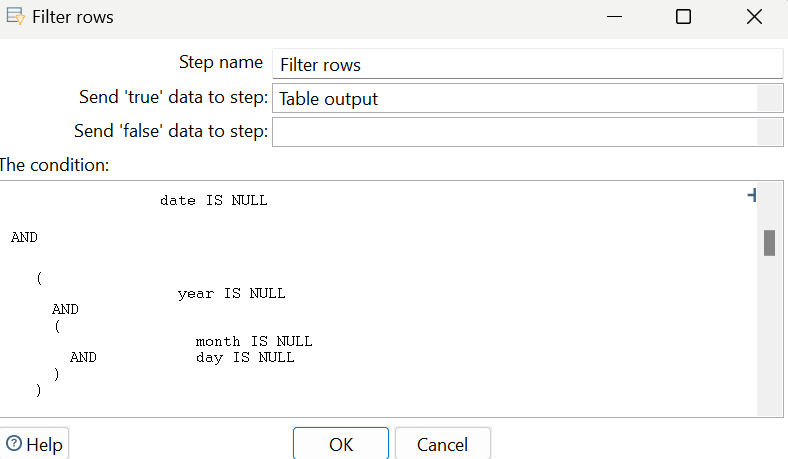
1. Konfigurasi pada **database lookup** adalah dengan memberikan **connection** dengan koneksi yang sudah dibuat pada step sebelumnya. dengan **schema** nama database yang digunakan dan **tabel dimdate** yang telah dibuat pada langkah pertama.
2. Field yang akan dicek untuk melihat kesamaan isi datanya agar tidak kembar adalah:

* field **date** pada table **dimdate** dengan field **streamDate**
* field **year** pada table **dimdate** dengan field **streamYear**
* field **month** pada table **dimdate** dengan field **streamMonth**
* field **day** pada table **dimdate** dengan field **streamDay**

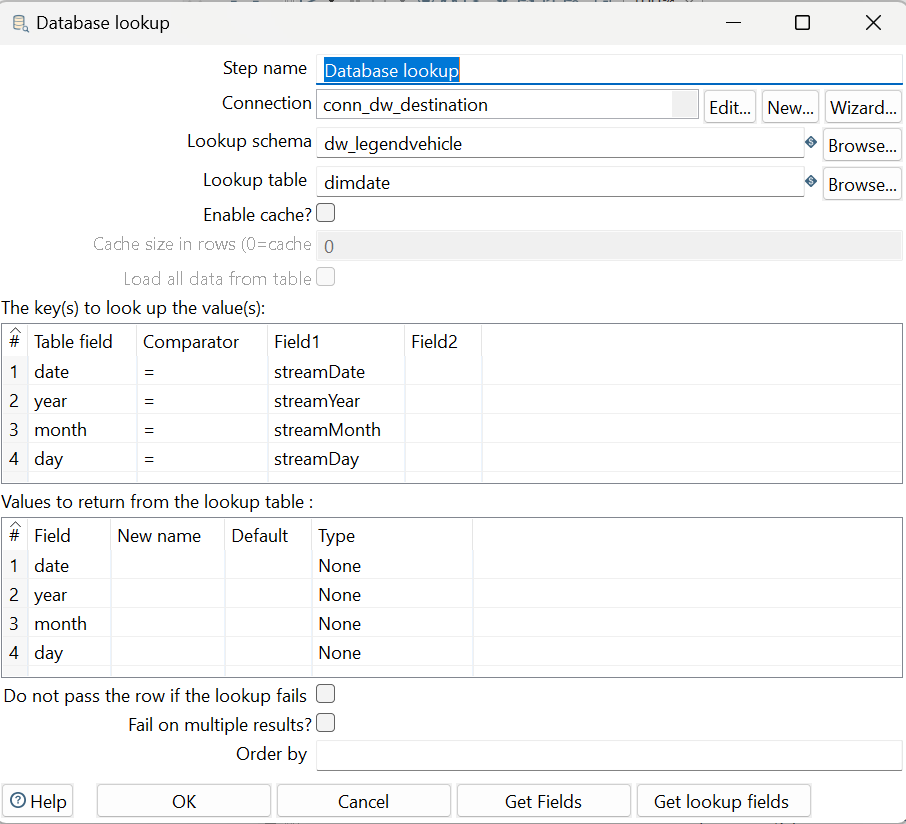
1. Field yang akan di **retrive** adalah field yang ada pada table **dimDate** yaitu **date, year, month,** dan **day**.



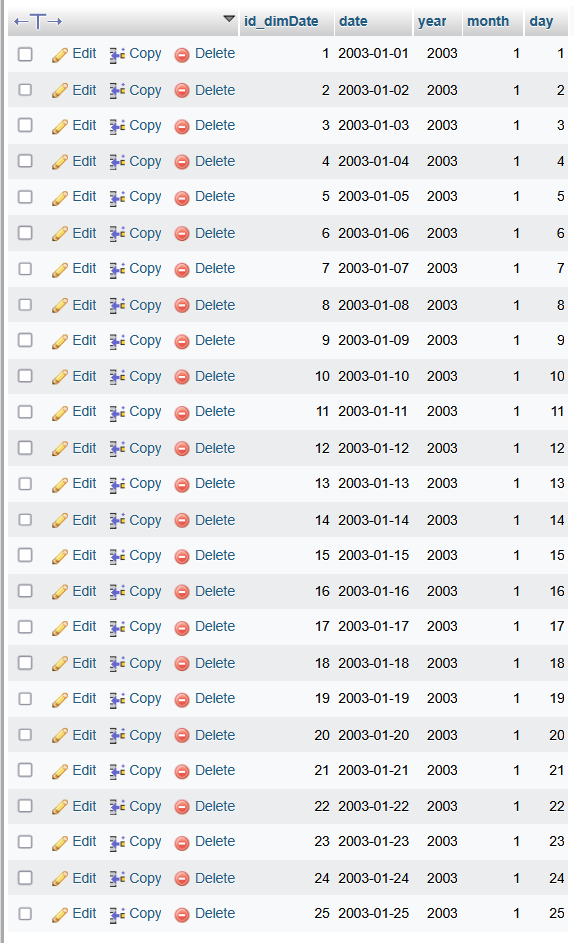
1. Hubungkan output dari **database lookup** dengan **filter rows**
2. Konfigurasi pada **filter rows** adalah dengan melakukan konfigurasi **output true data** pada **table output**. Pada bagian ini data yang tidak memiliki kesamaan pada tahapan sebelumnya akan dicek dimana jika **fields Stream** tidak 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.



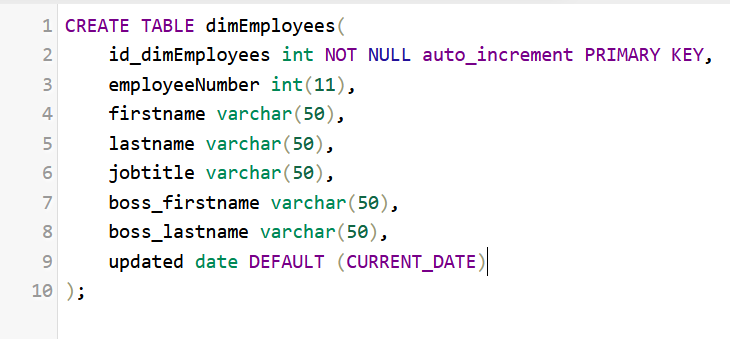
**TUGAS 1**

* 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.

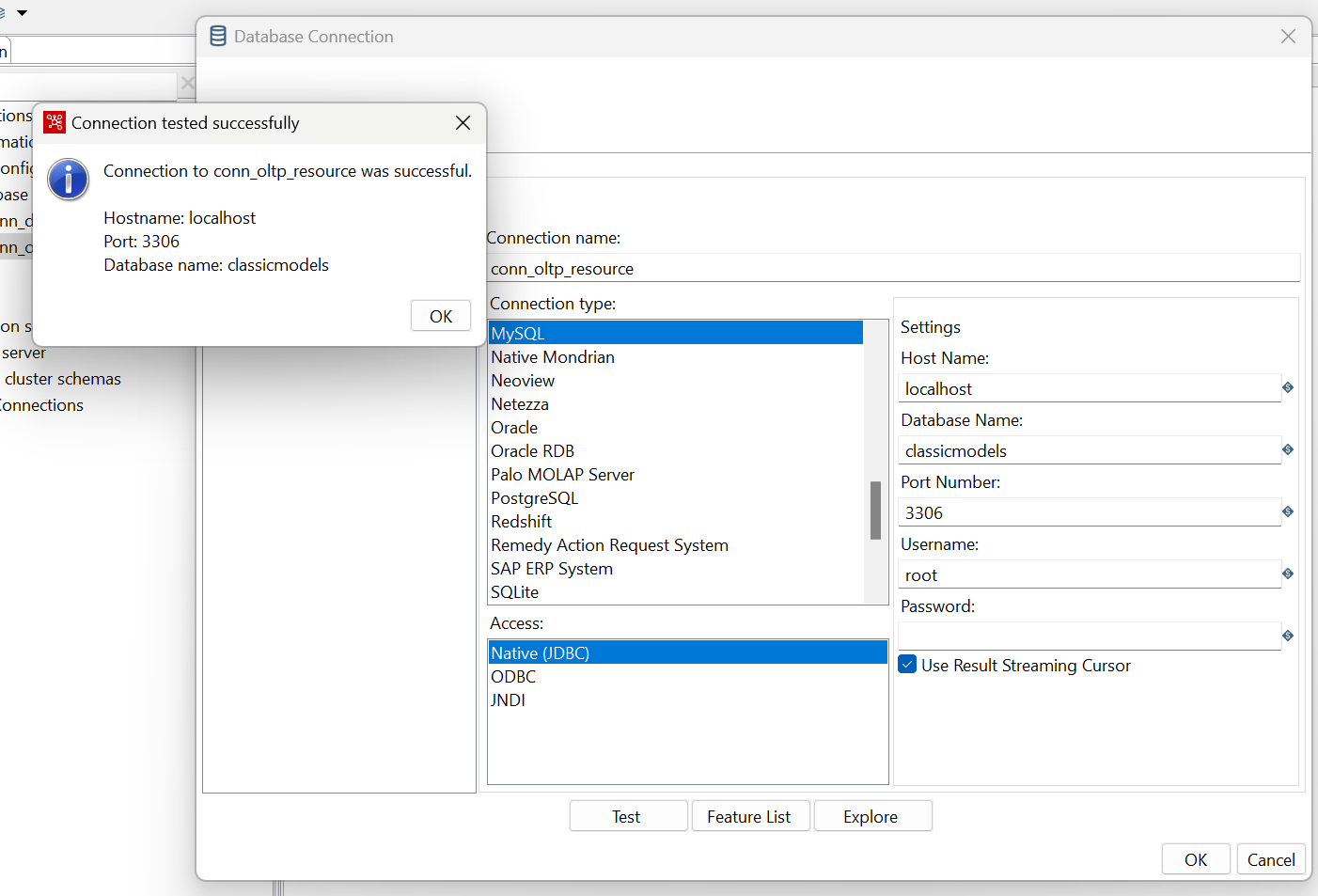
|  |  |  |  |
| --- | --- | --- | --- |
| Proses Objek | SS data input | SS data output | Keterangan |
| Generate rows | Tidak ada |  | Step ini digunakan untuk menghasilkan data tanggal awal secara otomatis dan tidak memiliki input karena merupakan sumber data pertama. |
| Add Sequences |  |  | Step ini akan menambahkan urutan angka yang bisa digunakan untuk perhitungan lebih lanjut dan berguna untuk membuat identifikasi unik atau perhitungan urutan. |
| Calculator |  |  | Step ini digunakan untuk melakukan perhitungan atau manipulasi data, seperti menambahkan hari atau memisahkan tanggal ke dalam komponen tahun, bulan, dan hari serta Berguna untuk analisis lebih lanjut atau pembuatan tabel dimensi waktu. |
| Select values |  |  | Step ini digunakan untuk memilih kolom yang diperlukan, sehingga hanya kolom yang relevan yang dipertahankan dan Bisa digunakan untuk mengganti nama kolom atau membuang kolom yang tidak diperlukan. |
| Database lookup |  |  | Step ini digunakan untuk mencari informasi tambahan dari database berdasarkan streamDate dan Jika nilai yang dicari tidak ditemukan dalam database, hasilnya <null>. |
| Filter rows |  |  | Step ini digunakan untuk menyaring (filter) baris berdasarkan kondisi jika tanggal, bulan, tahun serta hari null. |
| Table Output |  |  | Table Output menerima data yang memiliki nilai NULL pada kolom date, year, month, dan day serta data yang memiliki nilai valid (tidak NULL) akan diabaikan dan tidak masuk ke Table Output. |

**B. Dimensi Pegawai**

1. Buatlah tabel dimPegawai pada dw\_legendVehicle.

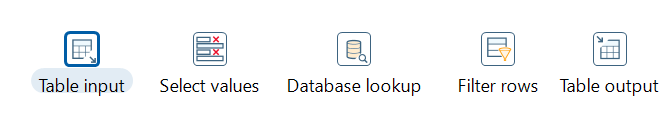


1. Pada PDI Spoon buatlah koneksi baru dengan nama **conn\_oltp\_resources** yang menghubungkan dengan database oltp. sesuaikan **hostname, database name, port number, username dan password dengan keadaan pada device masing-masing.**

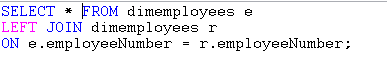


1. Drag and drop beberapa objek sebagai berikut:

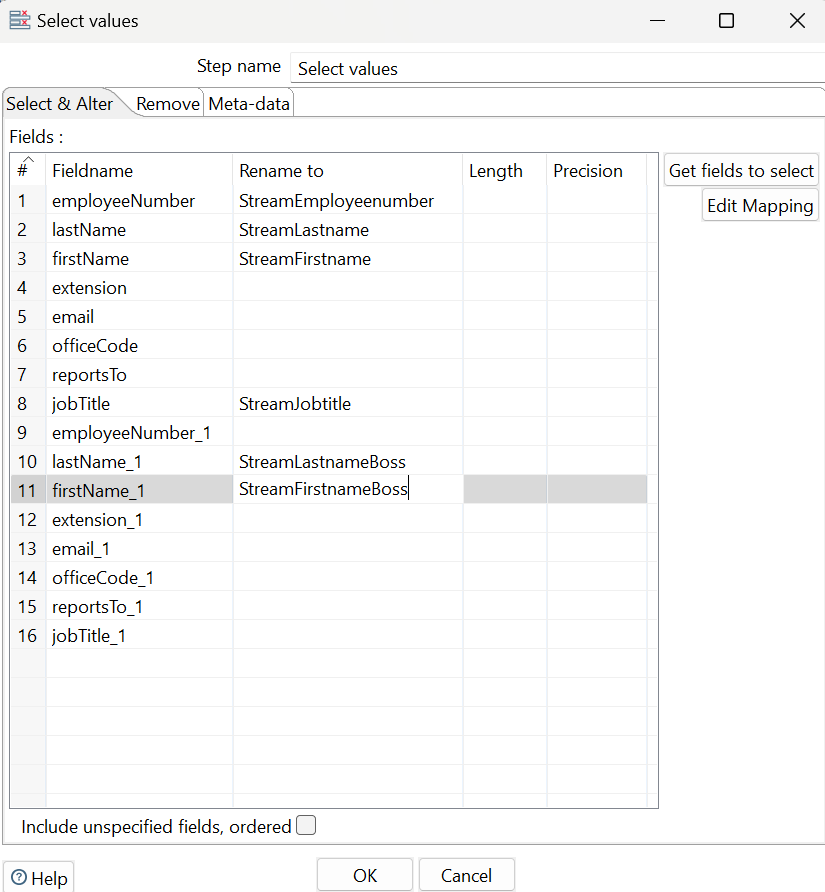
* **Table input:** digunakan mengambil data dari database OLTP.
* **Select values:** memeilih field yang digunakan untuk proses Transform dan Load.
* **Database lookup:** digunakan untuk melihat data pada tabel dimEmployees untuk memastikan data tidak kembar
* **Filter rows:** digunakan untuk memilih data stream yang masih belum ada apada tabel dimEmployees.
* **Table output:** Memasukkan data ke dalam tavle dimEmployees

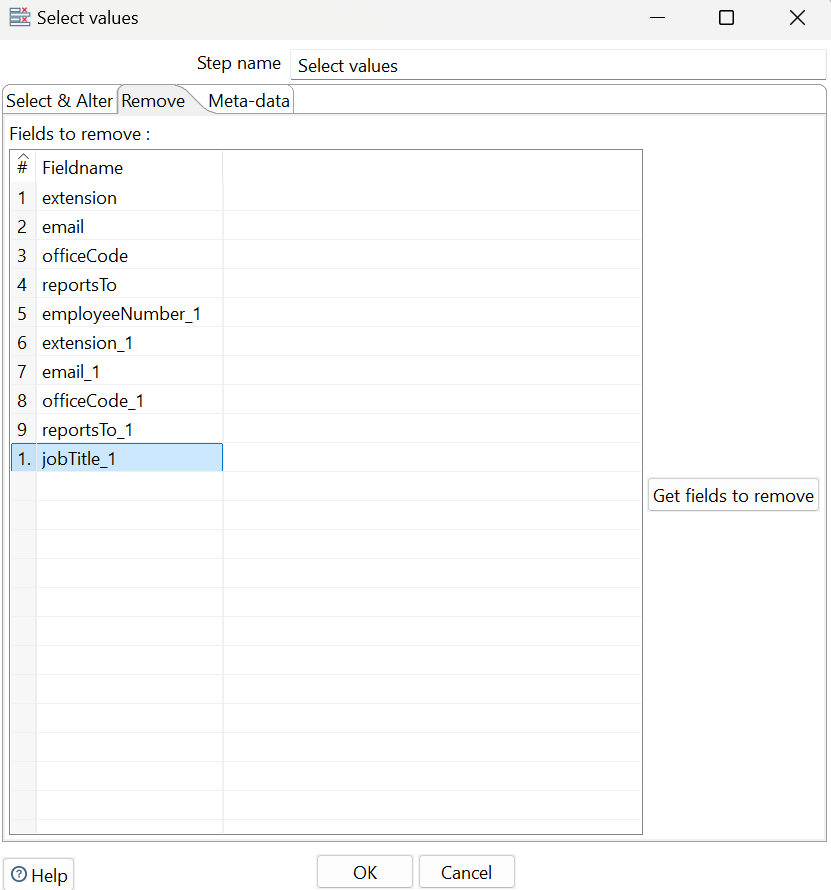


1. Konfigurasi pada table input dengan menghubungkan **Connection** pada konesi **conn\_oltp\_resources**. Untuk mengambil data sumber menggunakan query dibawah ini.

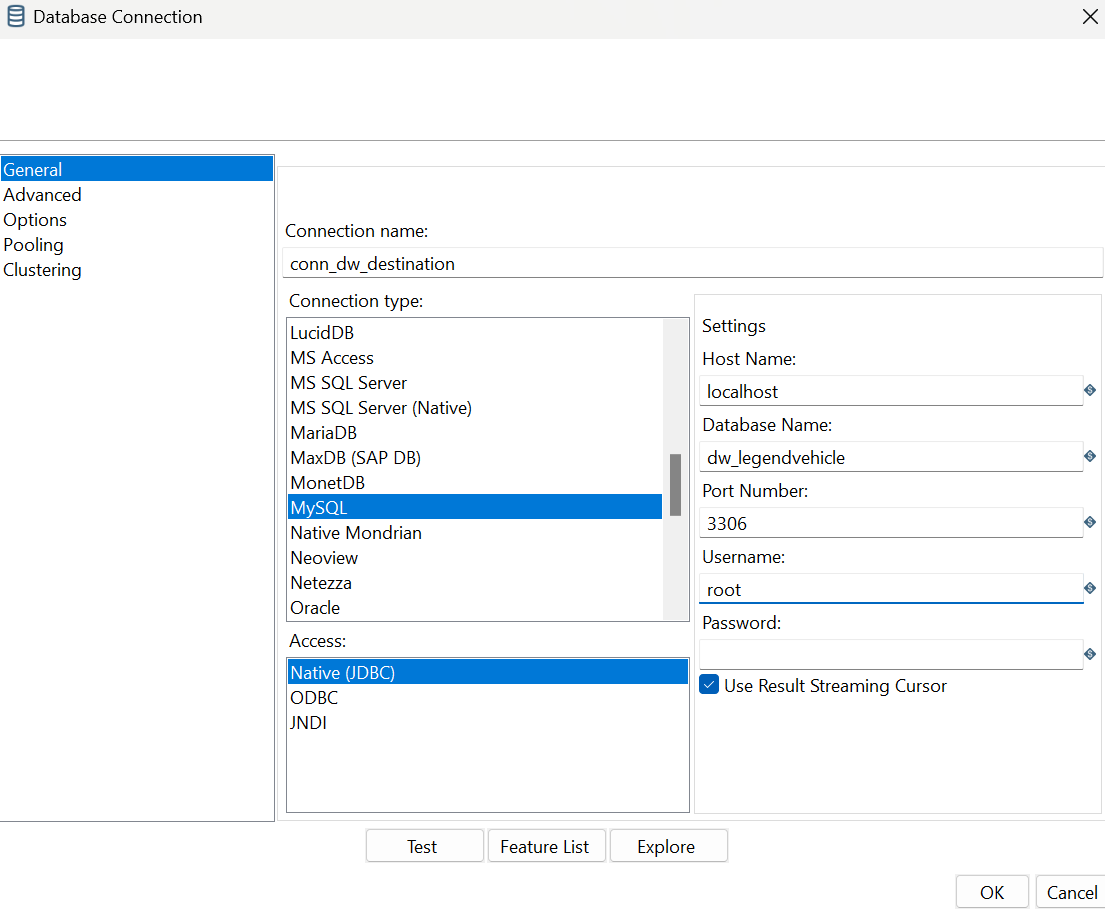


1. Hubungkan output **table input** pada **select values**.
2. Konfigurasi pada **Select values** yaitu mengambil data dari field **employeenumber, lastname, firstname, jobtitle , lastname\_1 dan firstname\_1** sebagai **data stream** yang digunakan pada proses ETL pada tab **select & alter**.
3. Hilangkan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

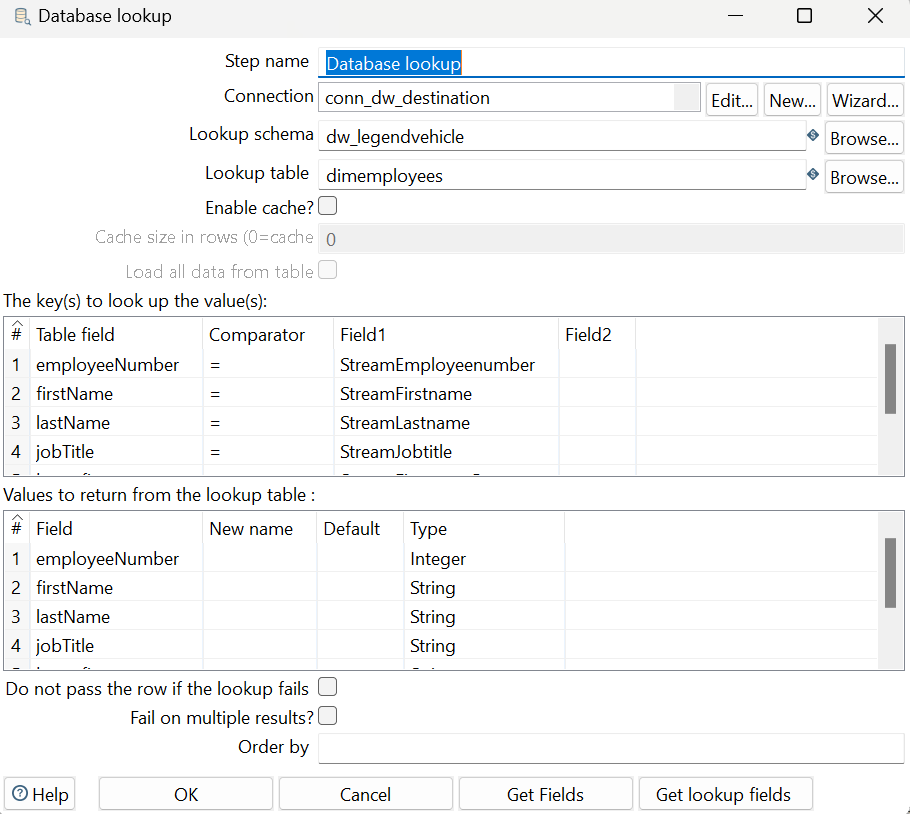




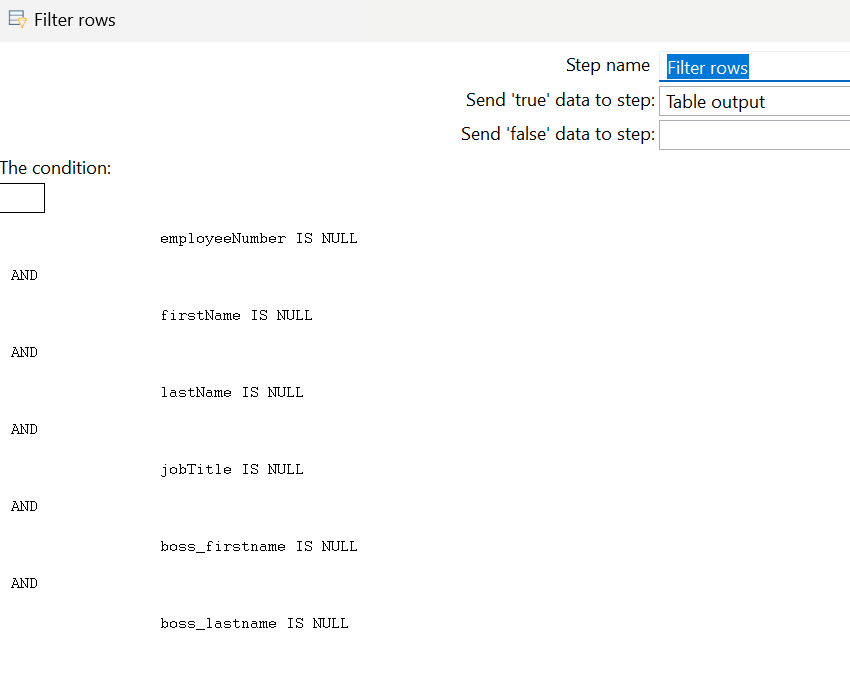
1. Hubungkan output **select values** pada **database lookup**
2. Konfigurasi pada database lookup adalah dengan menghubungkan koneksi pada **conn\_dw\_destination** dengan table lookup **dimEmployees** yang telah dibuat pada tahap pertama.



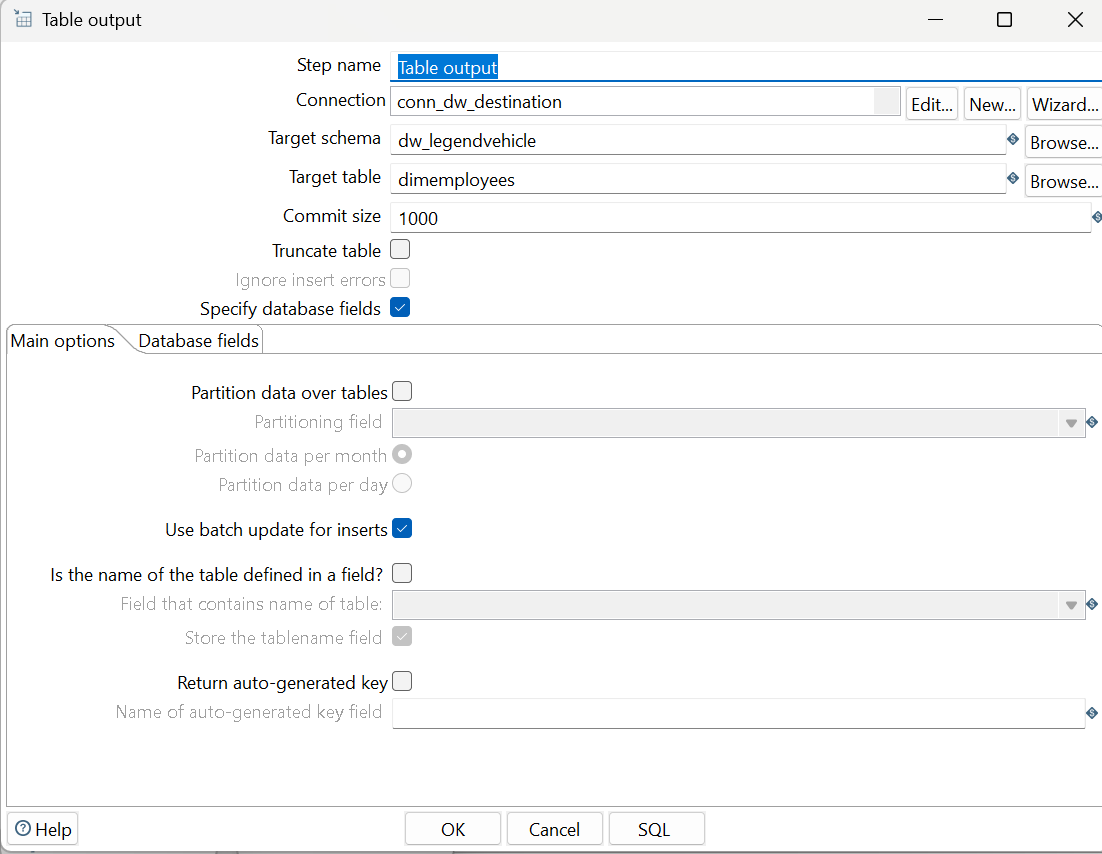
1. Field yang di lookup adalah field pada tabel **dimEmployees** dengan **field stream input** dari OLTP. sedangkan field yang di **retrieve** adalah field dari **dimEmployees** itu sendiri. Jika tidak ada data yang sama maka akan muncul null.



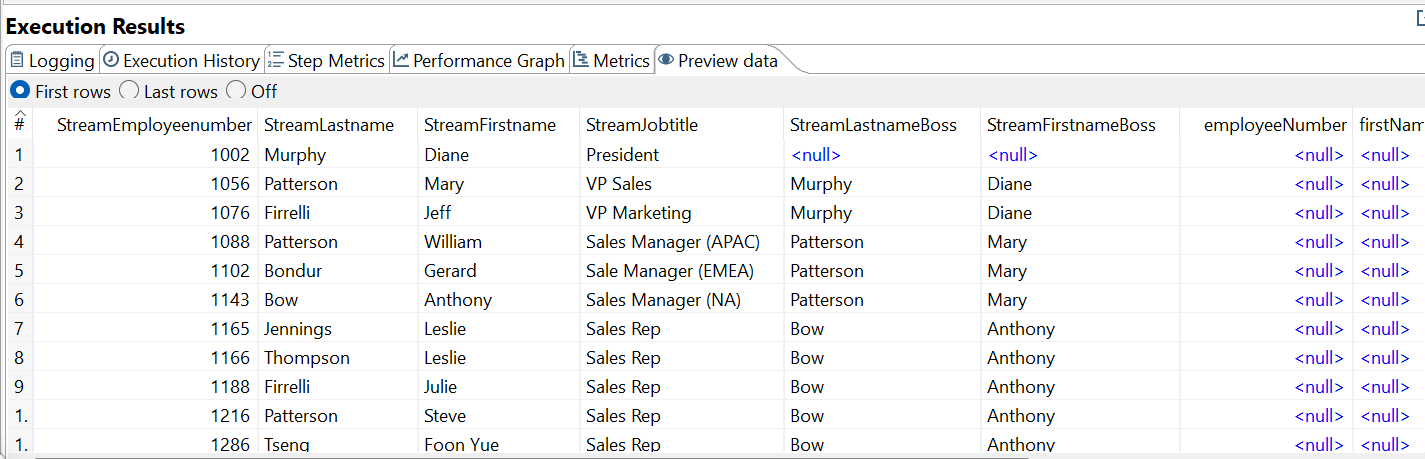
1. Hubungkan **output databse** lookup dengan **filter rows**.
2. Pada **filter rows** berikan kondisi field yang **null** pada **field dimemployees** untuk dimasukkan pada proses selanjutnya. Hal itu menandakan bahwa **data stream** belum memiliki kesamaan pada data di **dimemployees**.

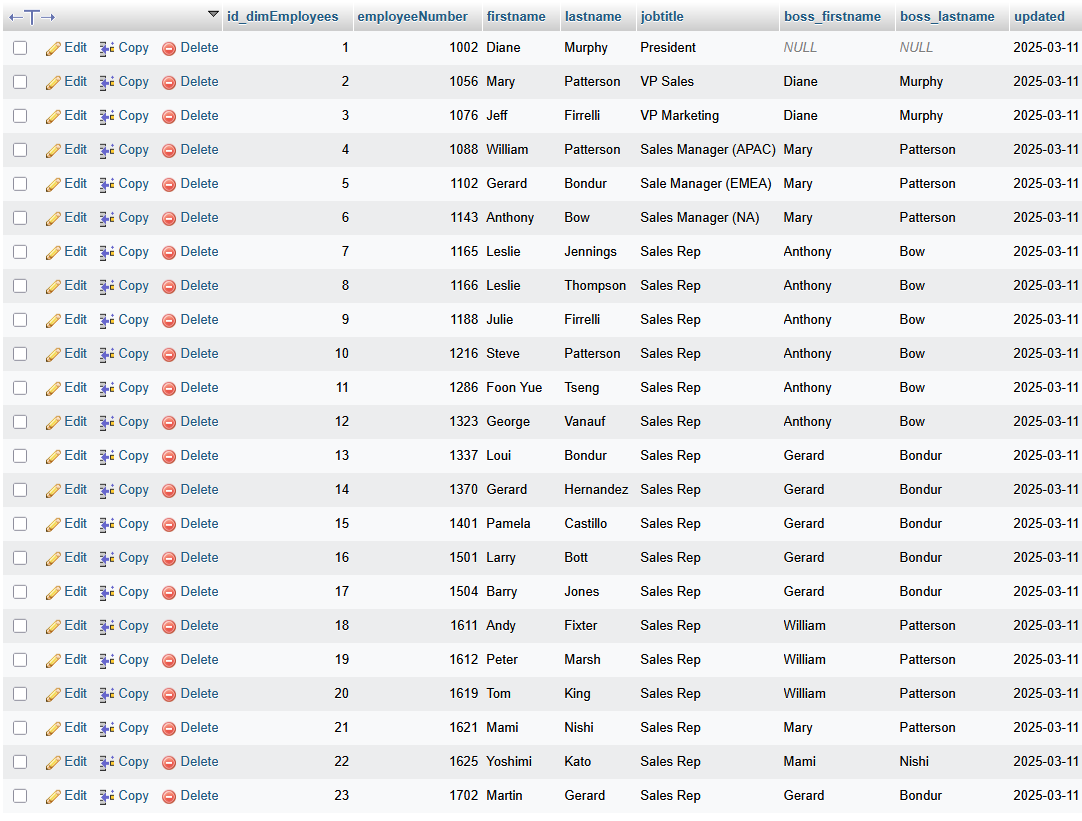


1. Hubungkan output dari **filter rows** dengan **table output.**
2. Pada **table output**, gunakan **connection conn\_dw\_destination** untuk memasukkan data pada tabel **dimemployees.**
3. Aktifkan **specify databse fields**, dan mapping **data stream input** dari oltp terhadap field yang ada pada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



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



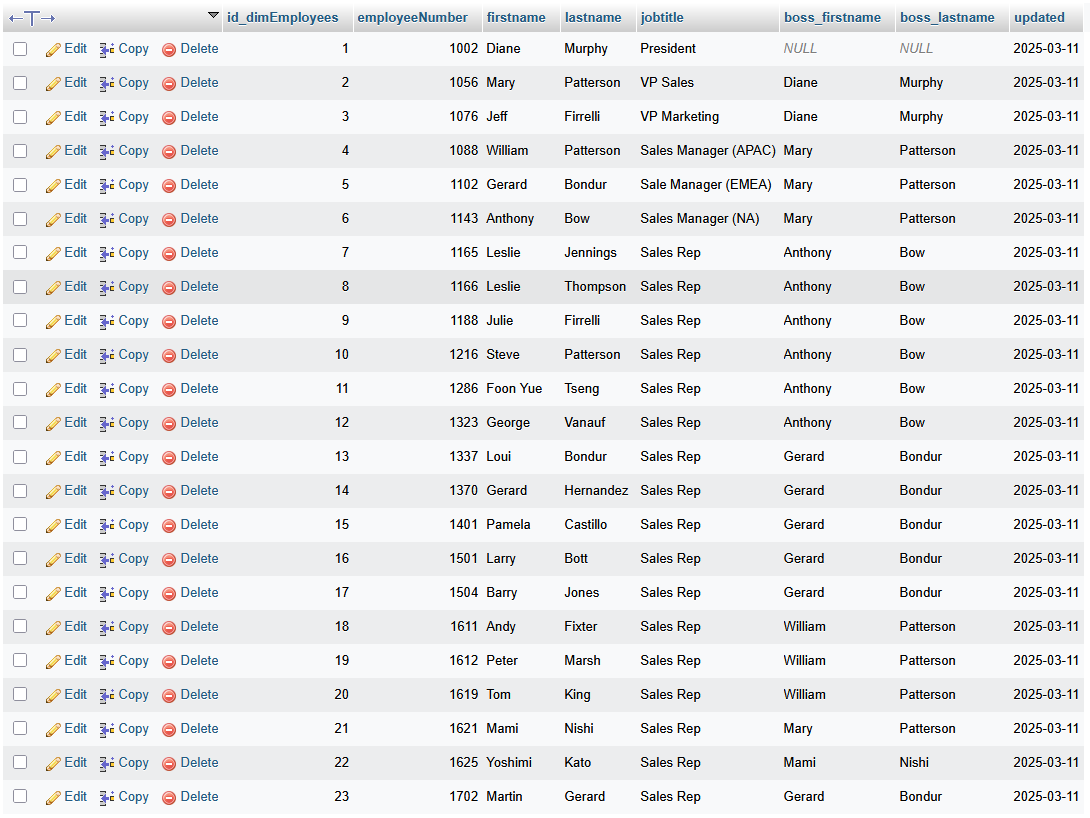


TUGAS 2

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.

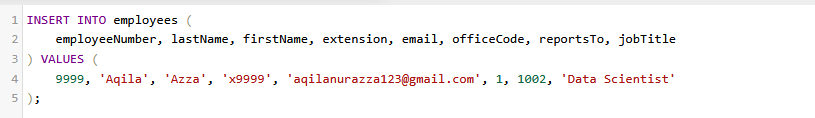
|  |  |  |
| --- | --- | --- |
| Proses Objek | SS data input | SS data output |
| Table Input | - |  |
| Select Values |  |  |
| Database Lookup |  |  |
| Filter Rows |  |  |
| Table Output |  |  |

1. Jika proses itu di ulangi ( di run kembali ) apakah data akan redudant?

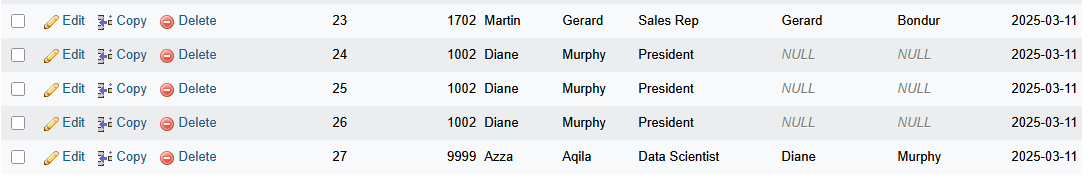
  
Penjelasan :

Dengan menggunakan data yang sama dan sudah di saring maka tidak akan terdapat redudansi

1. Tambahkan nama anda pada table employee di OLTP. jalankan kembali transformasi ini. Amati hasilnya, apa yang terjadi?



* + Dw\_employees



Penjelasan :

Terjadi redudansi setelah insert data.