Nama : Dini Elminingtyas

Kelas/No.absen : SIB 2B / 07

NIM : 2341760180

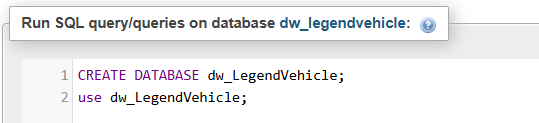
**DATA WAREHOUSE**

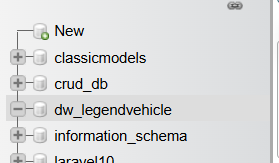
**JOBSHEET 3**

**DATABASE ANALYTICAL**

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

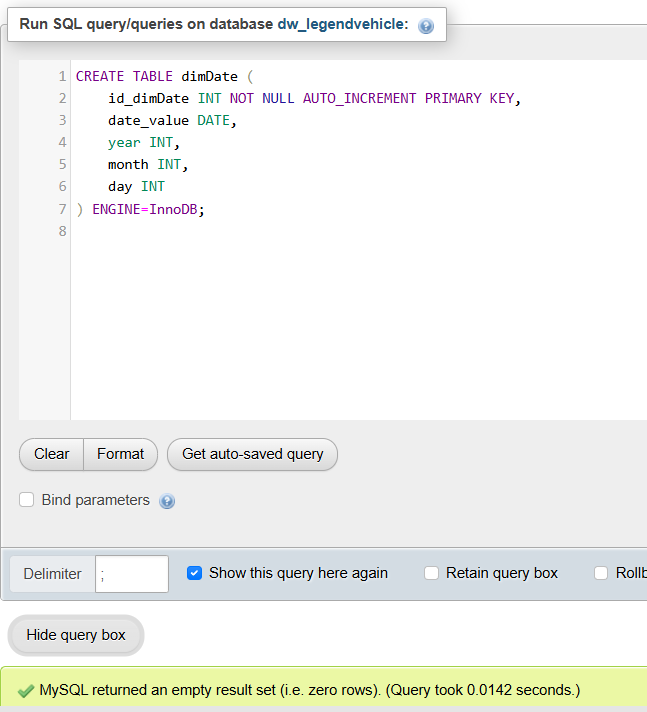
* Query





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

* Query





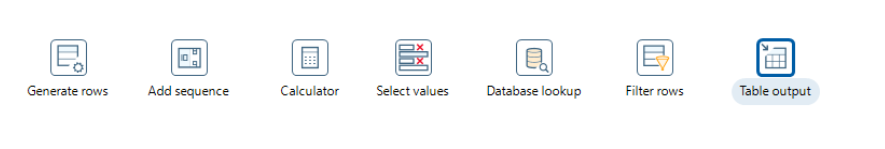
Pada tahapan selanjutnya, untuk membuat tabel dimensi dimDate , maka diperlukan generate data tanggal. Data tanggal yang disiapkan pada tabel dimDate menyesuaikan dengan proses bisnis yang berjalan. Proses bisnis pada LegendVehicle adalah 5 tahun. Sehingga data pada tabel dimdate yang harus tersedia adalah tanggal selama 5 tahun. Mulai dari 1 Januari 2023

1. Buka PDI Spoon. Buat Transformation baru -> **File - New - Transformation.**

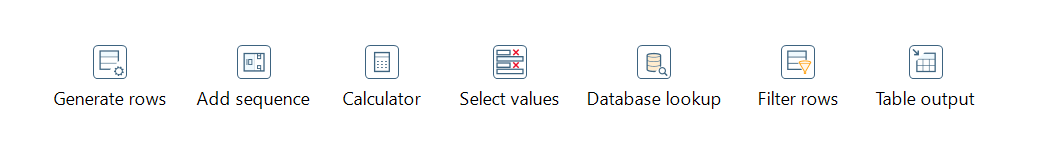


1. Drag and Drop beberapa objek yaitu:

* **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).

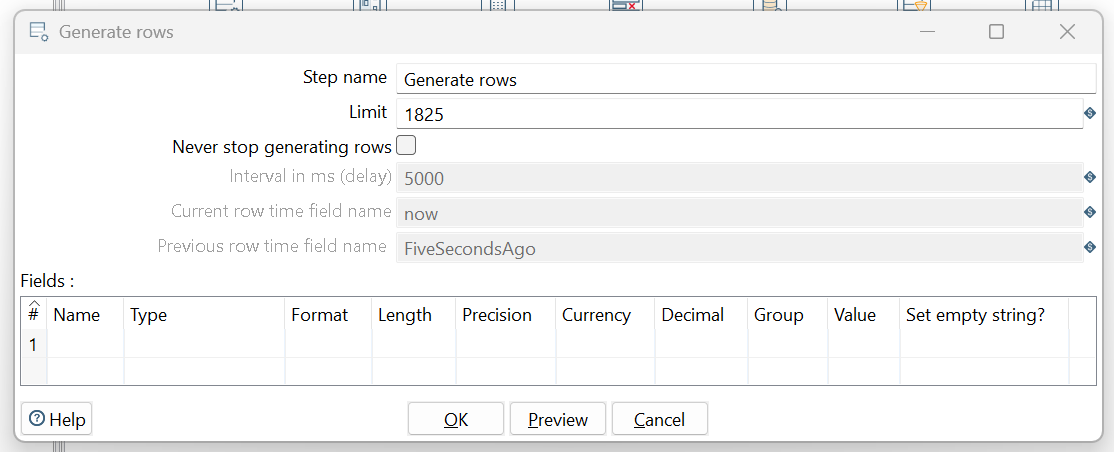


* Drag n Drop

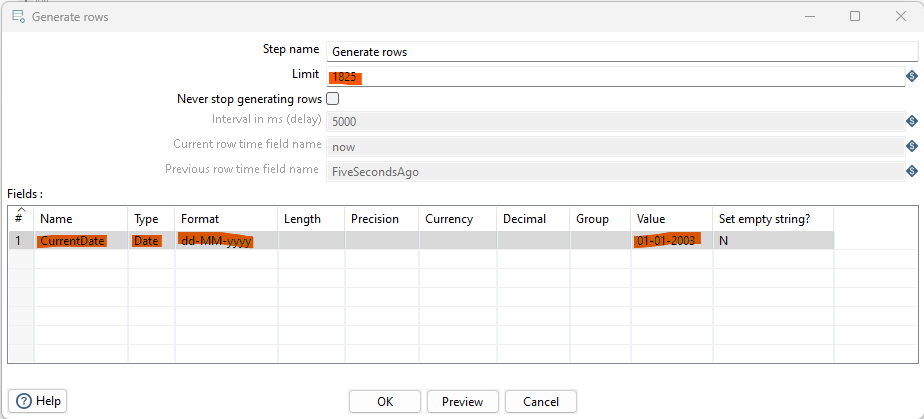


1. Konfigurasi pada Generate Rows adalah merubah limit menjadi 1825 dimana memiliki arti bahwa data yang akan dibuat sebanyak 1825 data. 1825 merupakan jumlah hari dalam 5 tahun ( 365 hari x 5 tahun ).

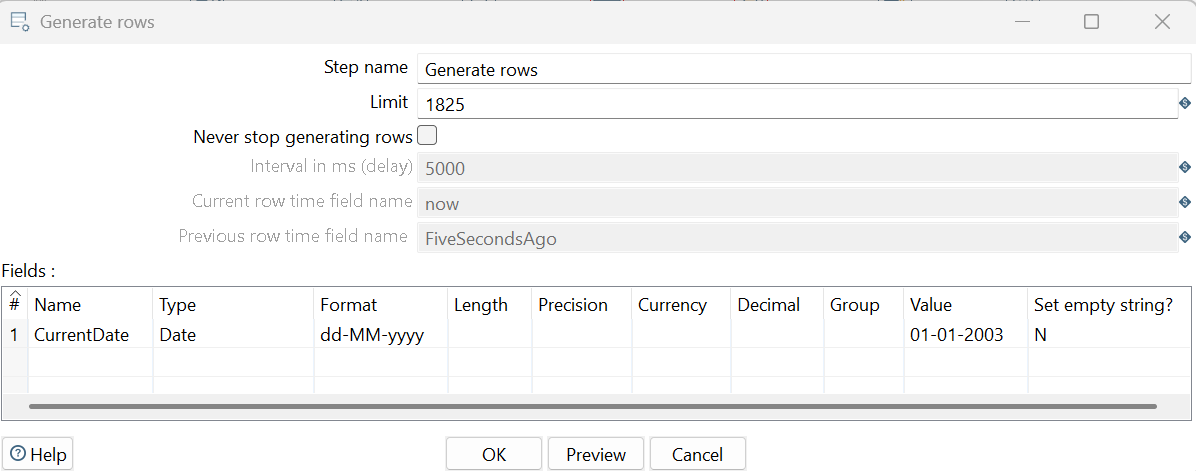
* Configure



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

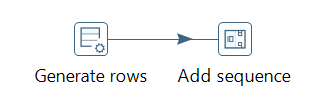


* Configuration

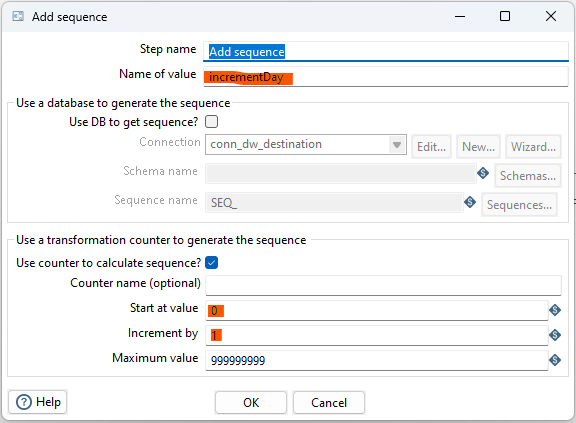


1. Hubungkan output dari Generate Rows menuju Add Sequence.

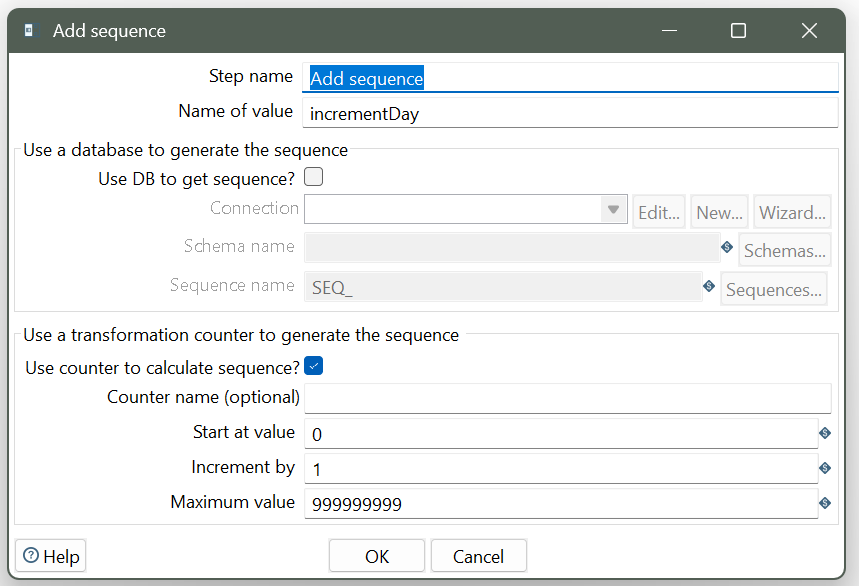
* Configure



1. Konfigurasi pada Add Sequences adalah merubah Name of value menjadi incrementDay dengan start value bernilai 0 dan increment by bernilai 1

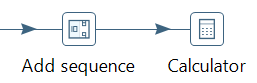


* Configure



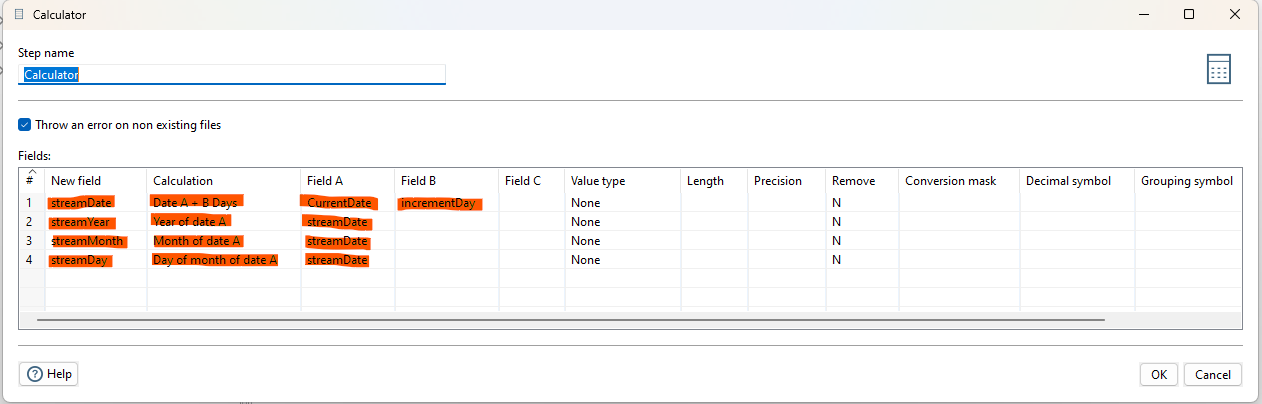
1. Hubungkan output dari add sequences menuju calculator.

* Comfigure

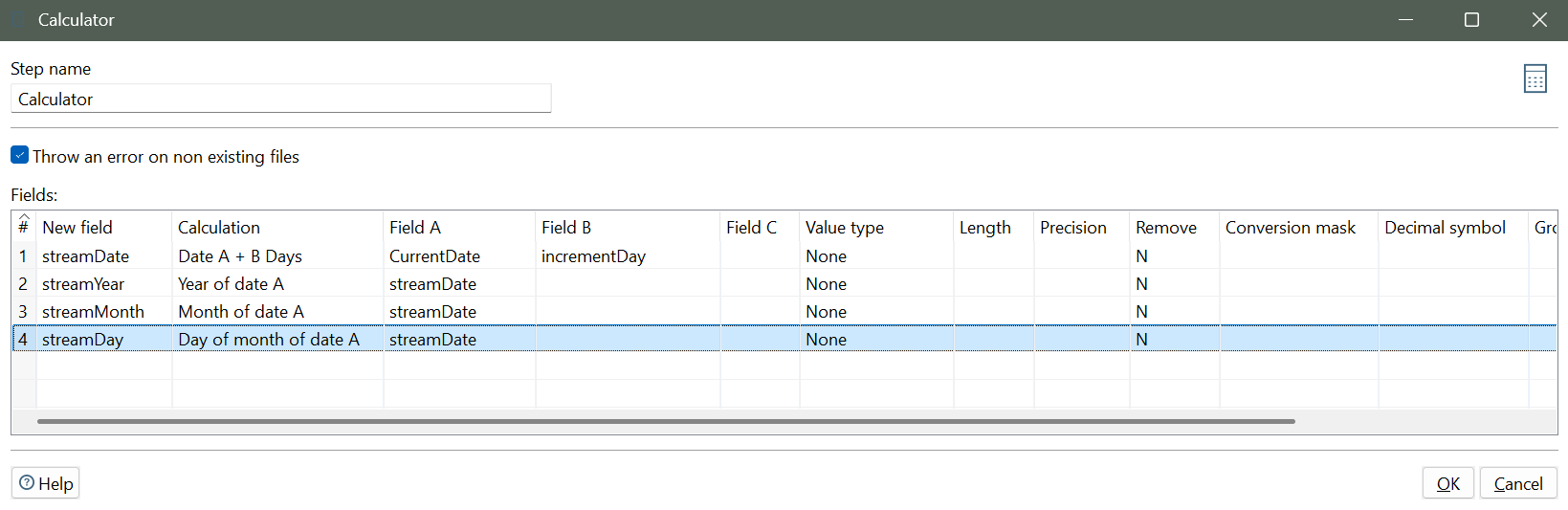


1. Konfigurasi pada calculator dengan membuat fields baru sebagai berikut:

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

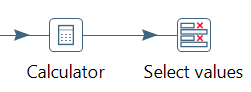


* Configure



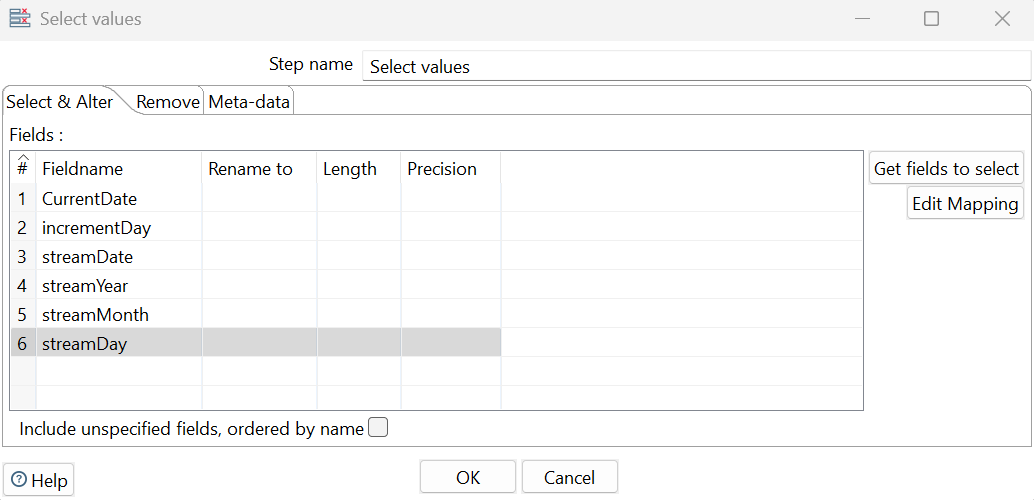
1. Hubungkan output dari calculator menuju Select values

* Configure

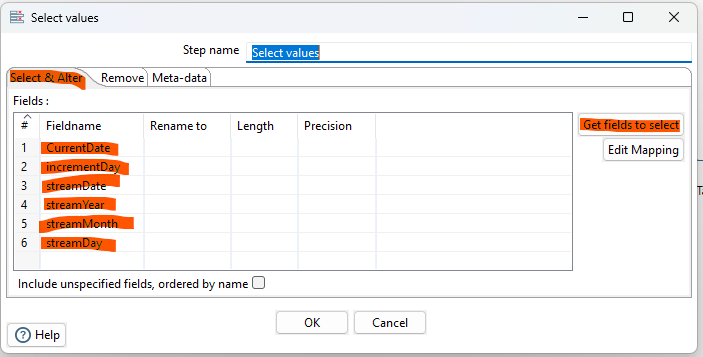


1. Konfigurasi pada select values adalah dengan menekan tombol Get fields to select pada tab Select & Alter. Secara otomatis semua fields dari data input akan muncul.

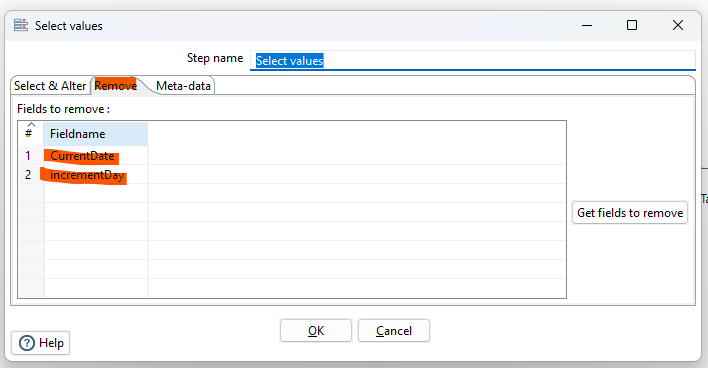
* Configure



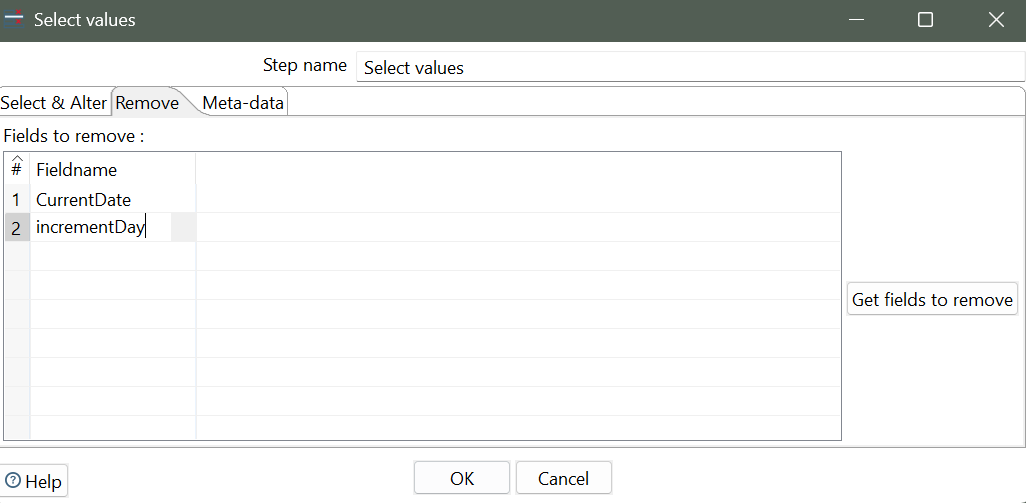
1. Dikarenakan tidak semua fields digunakan, maka pada tab Remove diisikan fields CurrentDate dan incrementDay dikarenakan kedua fields tersebut tidak digunakan.



Gambar konfigurasi tab select & alter pada select values

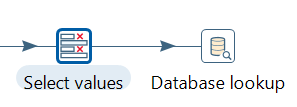


* Remove CurrentDate n incrmentDay

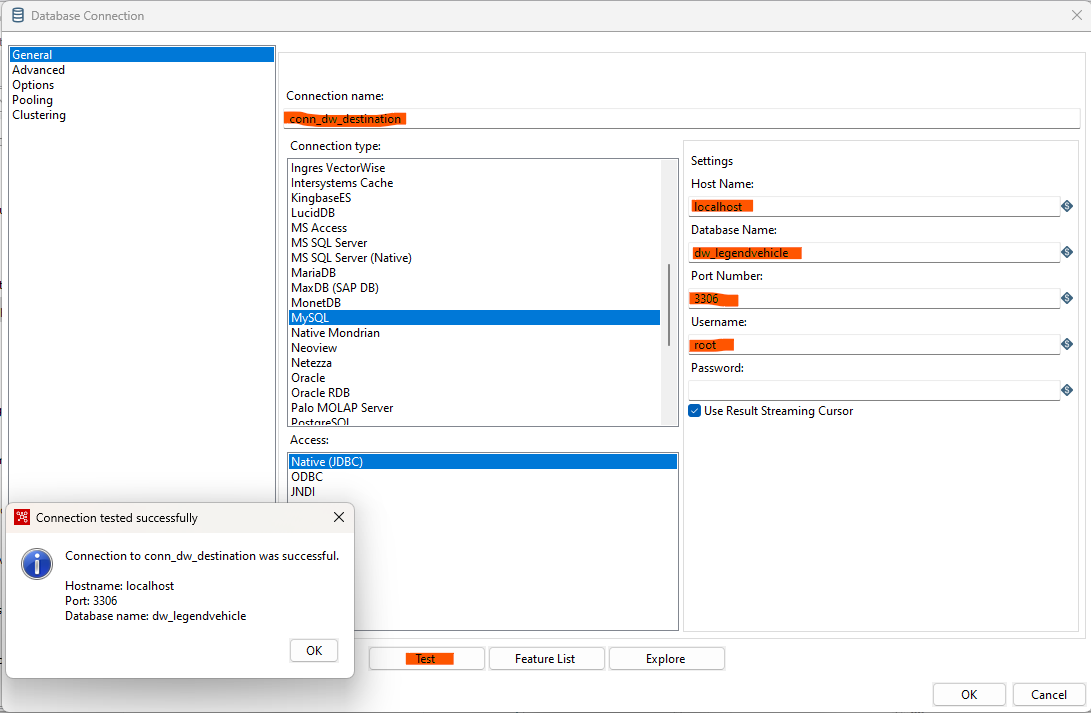


1. Hubungkan output select values menuju database lookup.

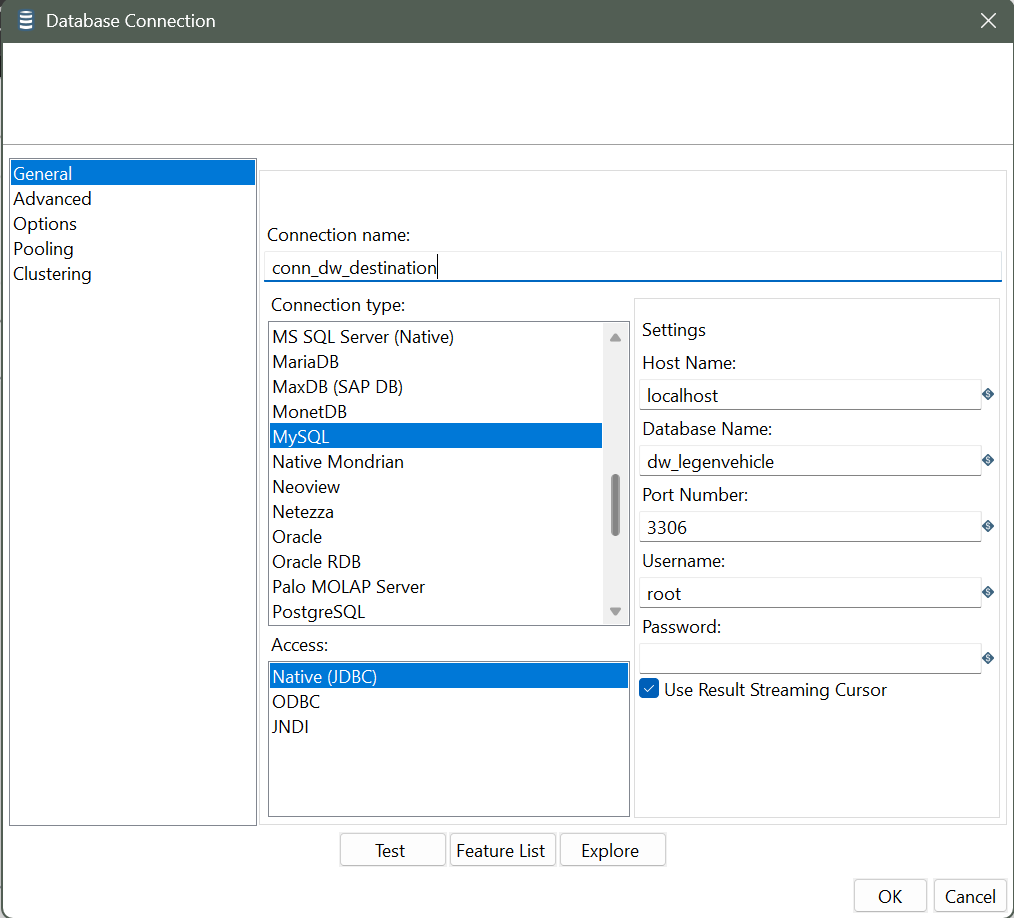
* Configure



1. 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.



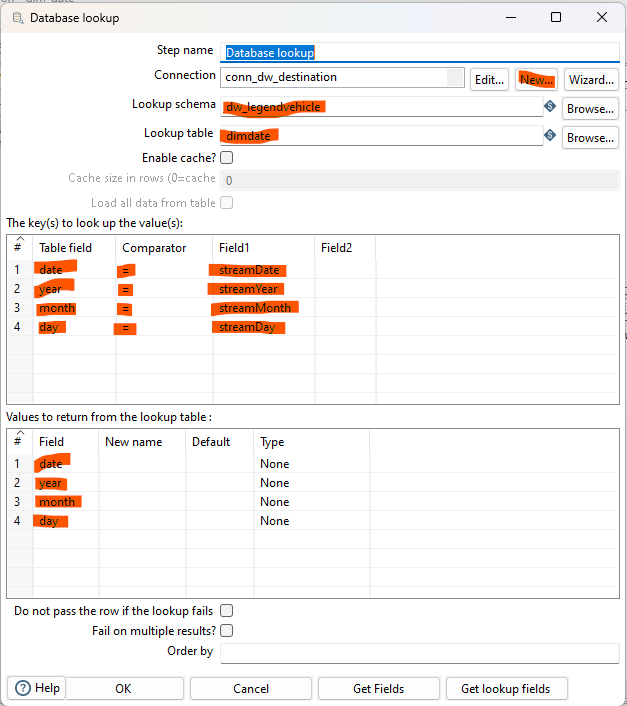
* Configure



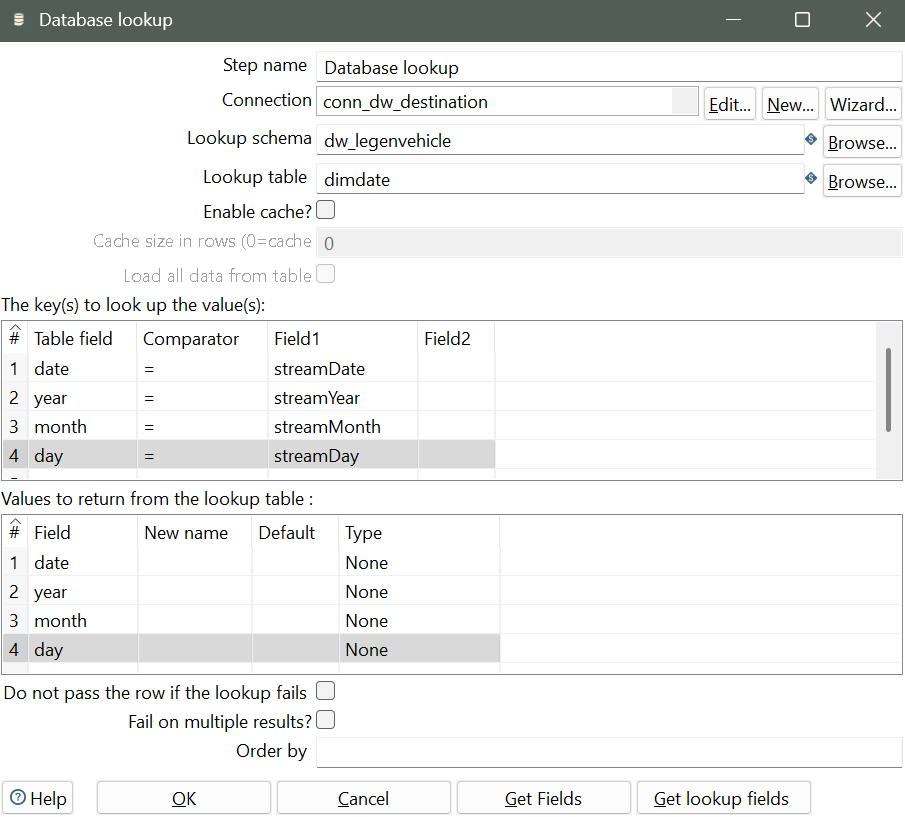
1. Konfigurasi pada database lookup adalah dengan memberikan connection dengan koneksi yang sudah dibuat pada step sebelumnya. dengan schema nama database yang digunakan dan tabel dimdate yang telah dibuat pada langkah pertama.
2. Field yang akan dicek untuk melihat kesamaan isi datanya agar tidak kembar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

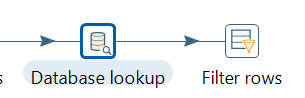


* Configure

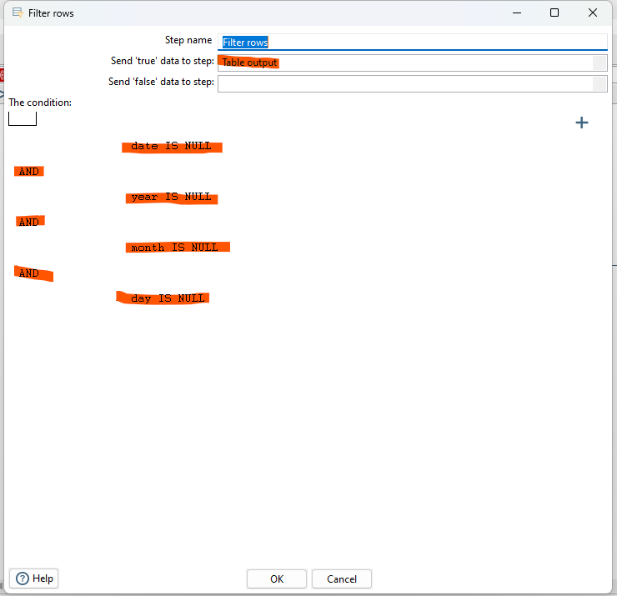


1. Hubungkan output dari database lookup dengan filter rows

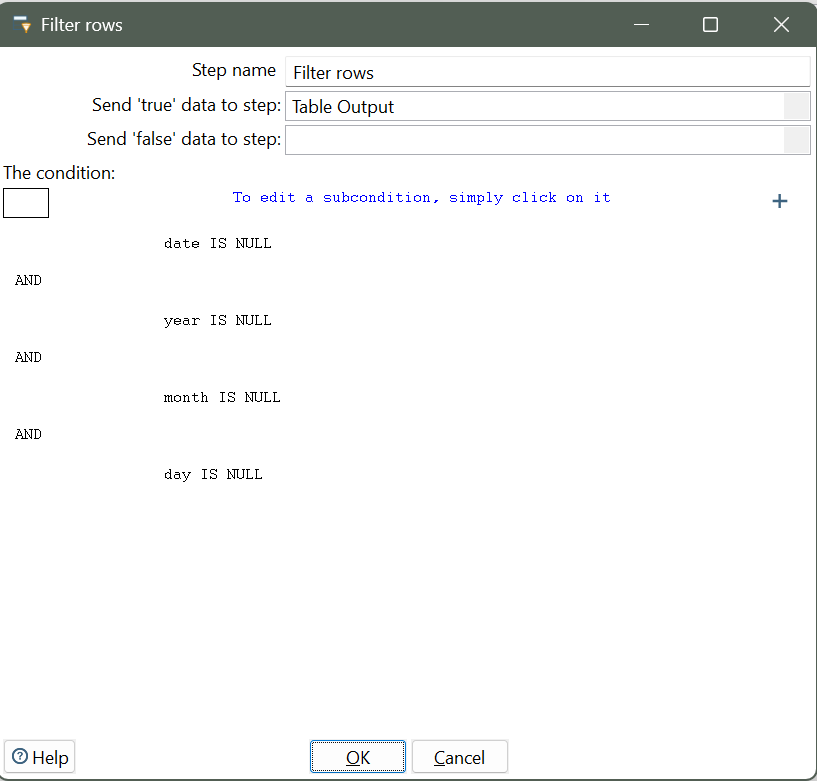
* Configure



1. 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 null)

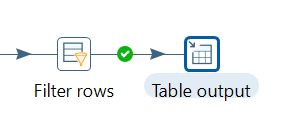


* Configure



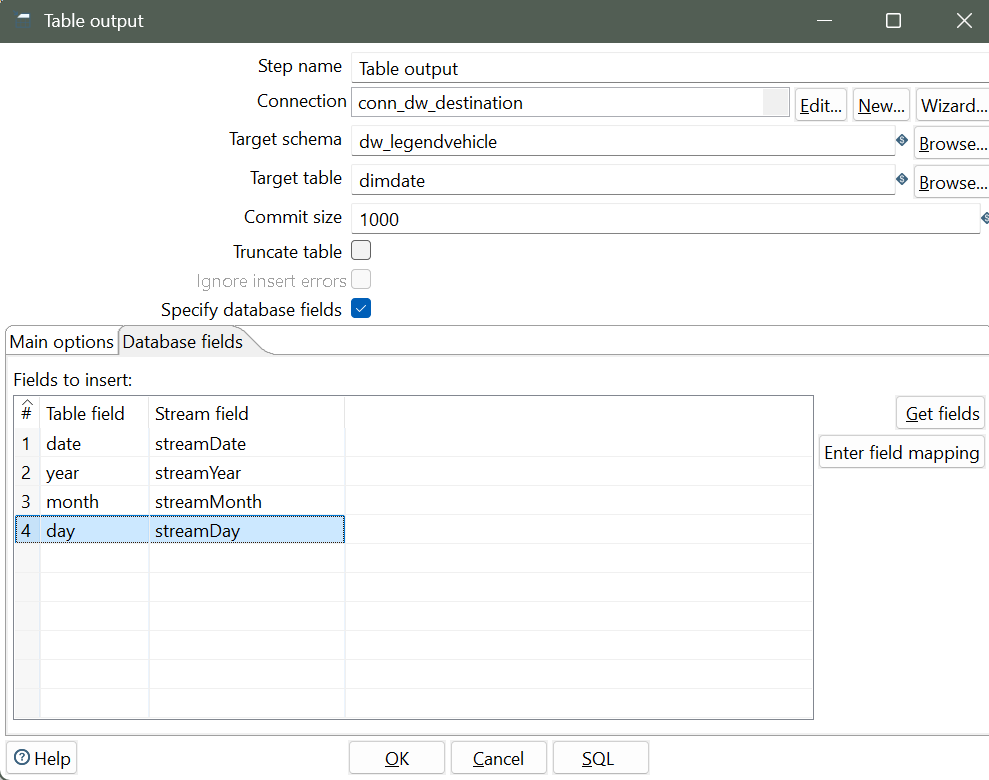
1. Hubungkan output dari filter rows menuju table output.

* Configure

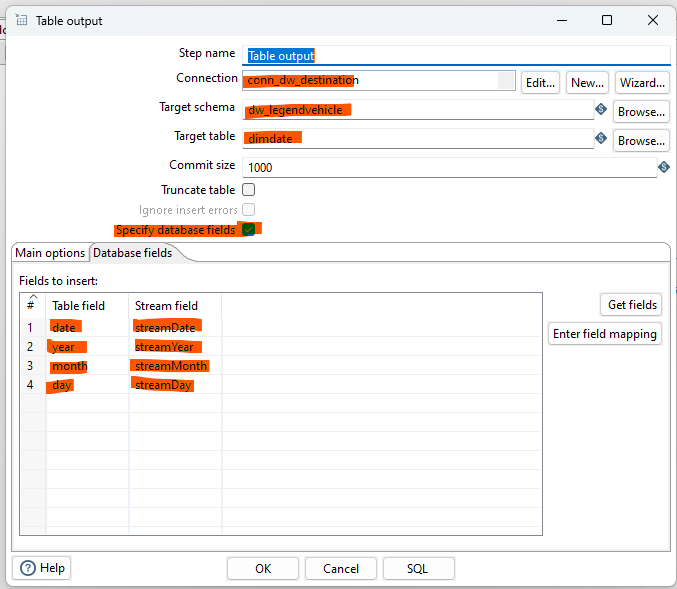


1. Konfigurasi pada table output adalah memberikan koneksi pada conn\_dw\_destination dengan schema dw\_legendvehicle dan table dimdate.

* Configure



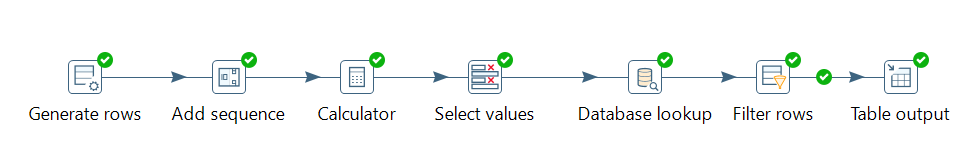
1. Aktifkan specify database fields.
2. Pada tab Database fields, mapping data input streamDate, streamYear, streamMonth dan streamDay dengan fields yang ada pada dimDate. Pada tahapan ini akan dilakukan insert data menuju tabel dimDate.



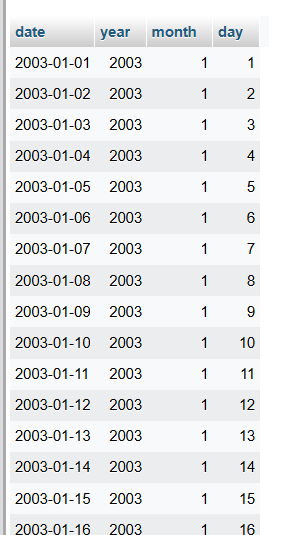
1. cek isi table dimdate pada database. Jika sukses maka pada table dimdate akan terisi 1825 data.



* Pentaho step



Database



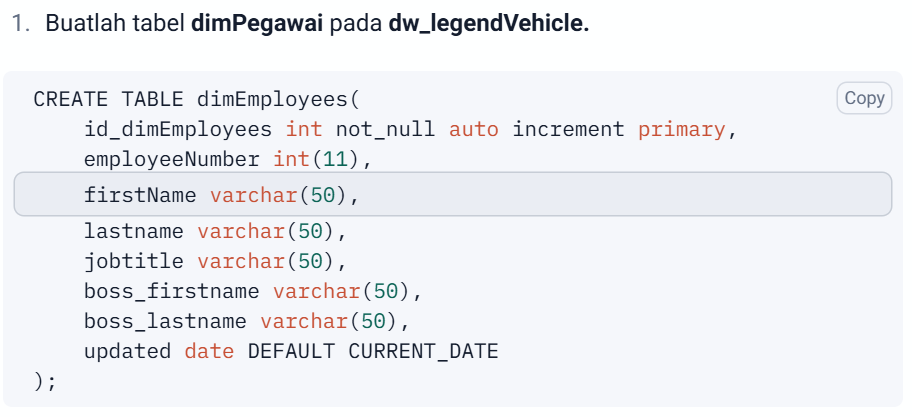
**TUGAS 1**

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.

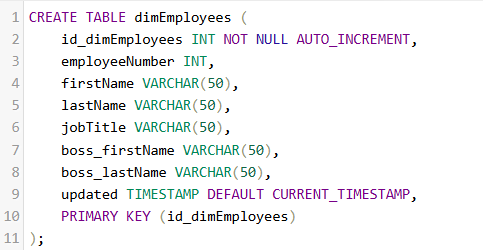
|  |  |  |  |
| --- | --- | --- | --- |
| Proses Objek | Data Input | Data Output | Keterangan |
| Generate Rows |  |  | Pada bagian input, pengguna mengatur parameter. Bagian output menunjukkan hasil eksekusi berupa daftar data yang telah dibuat. |
| Add Sequences |  |  | Pada bagian input, mengatur parameter "SEQ" dengan nilai awal 0 dan penambahan 1. Bagian output menunjukkan hasil eksekusi berupa daftar tanggal "01-01-2003" dengan nilai increment bertambah per baris. |
| Calculator |  |  | Pada bagian input mengonfigurasi kalkulator dengan beberapa kolom untuk memecah tanggal. Bagian output menunjukkan hasil eksekusi berupa daftar data tanggal yang dipecah menjadi nilai tahun, bulan, dan hari pada kolom terpisah. |

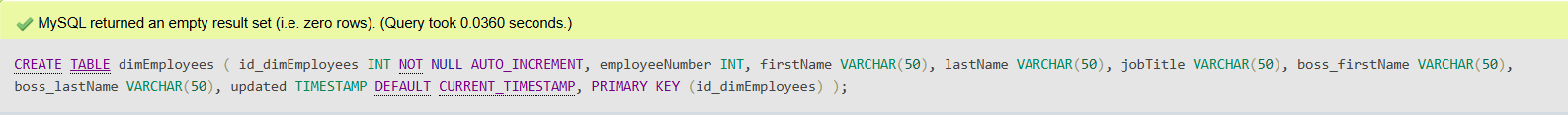
|  |  |  |  |
| --- | --- | --- | --- |
| Select Values |  |  |  |
| Database Lookup |  |  |  |
| Filter Rows |  |  |  |
| Table Output |  |  |  |

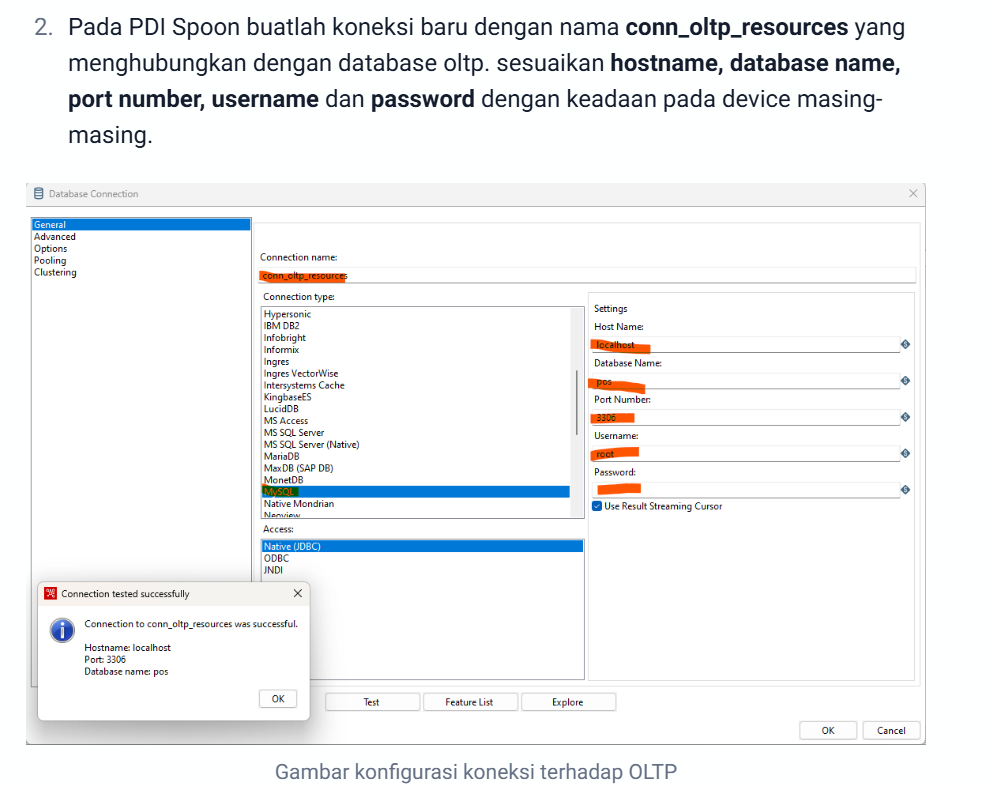
1. **DIMENSI PEGAWAI**

****

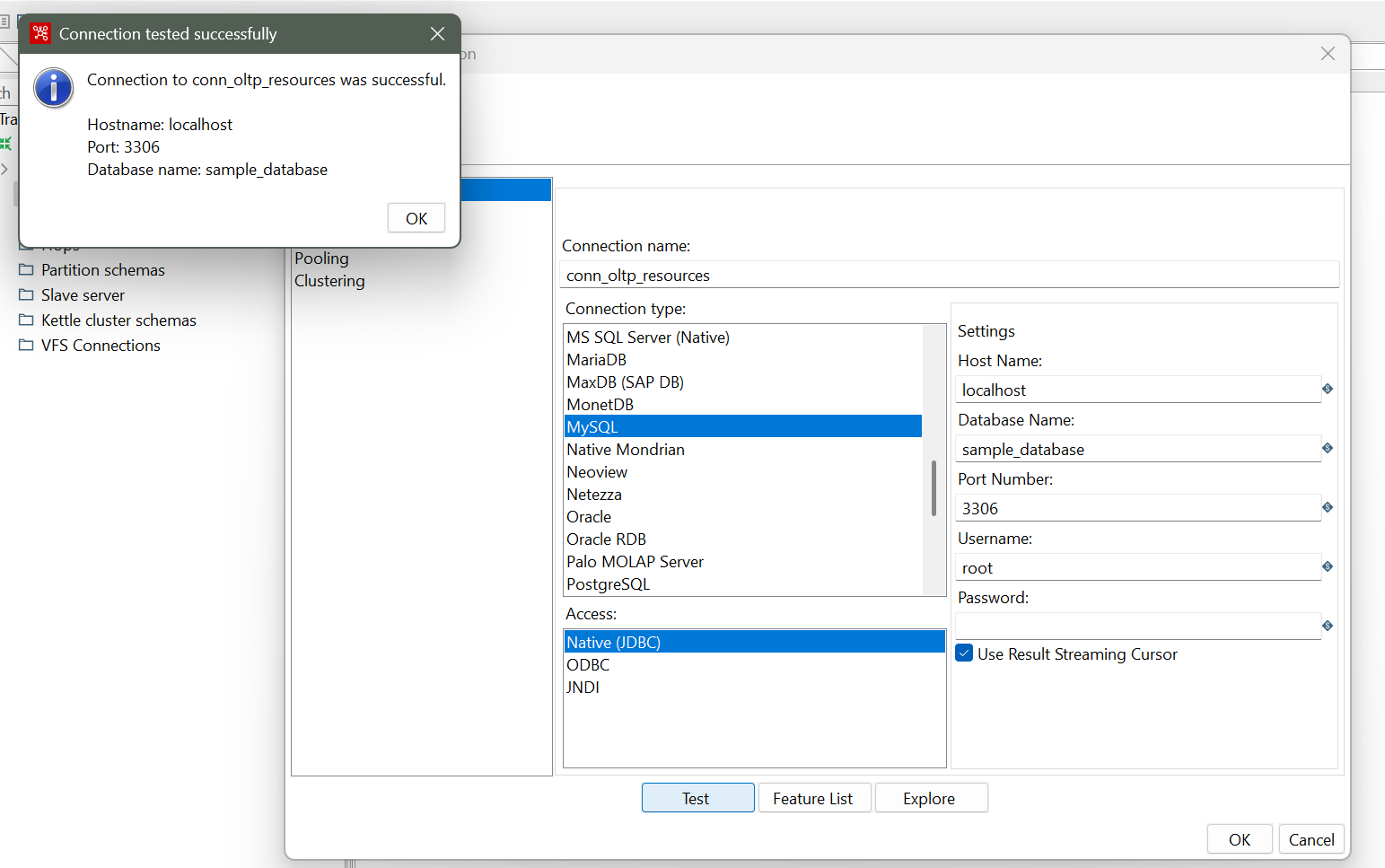
* fixed query

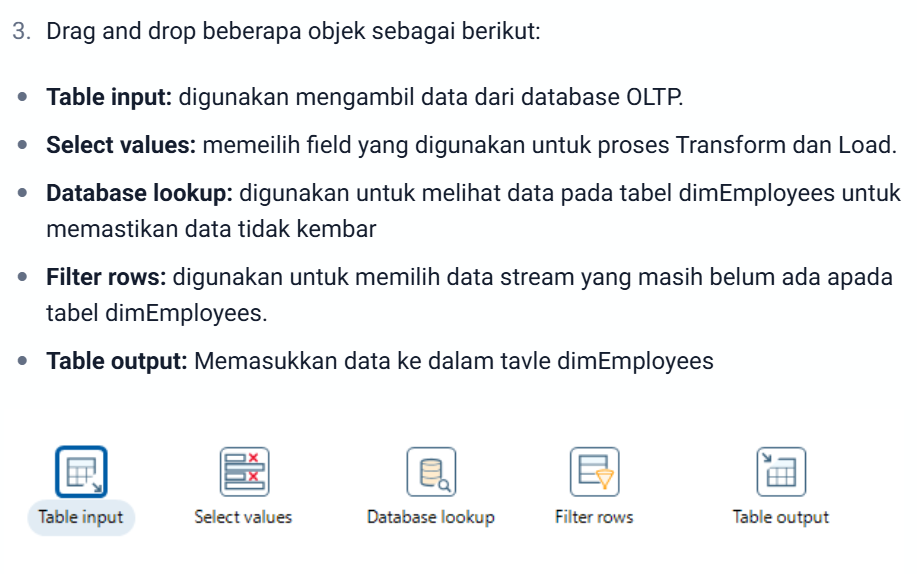
****

****

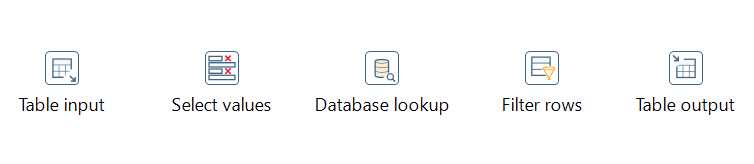


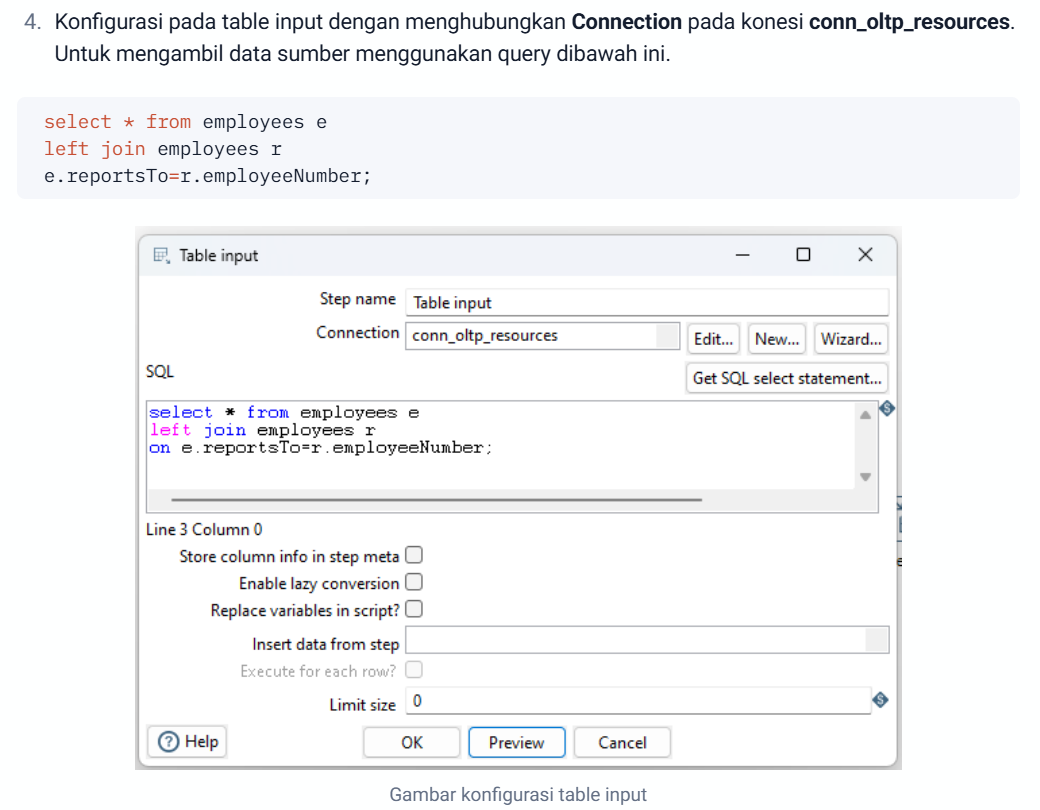
* Configure



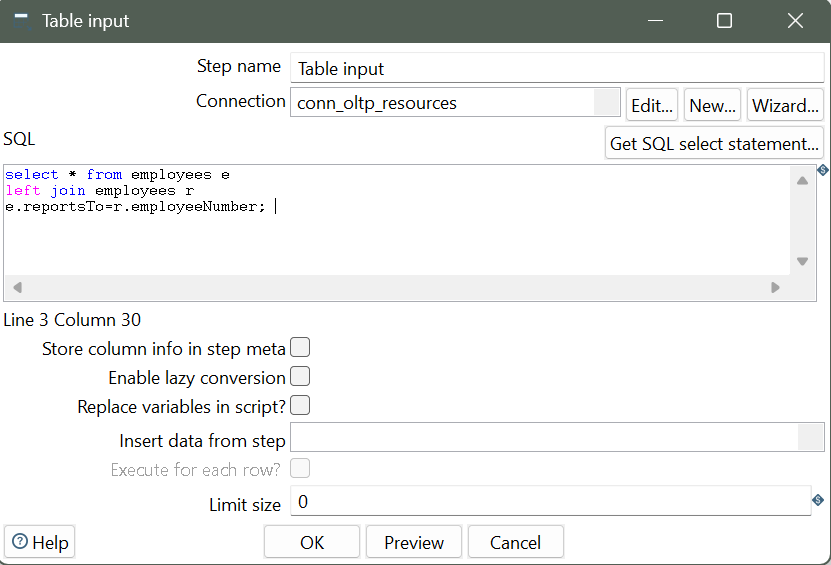


* Configure



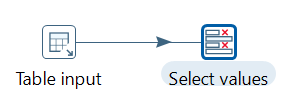


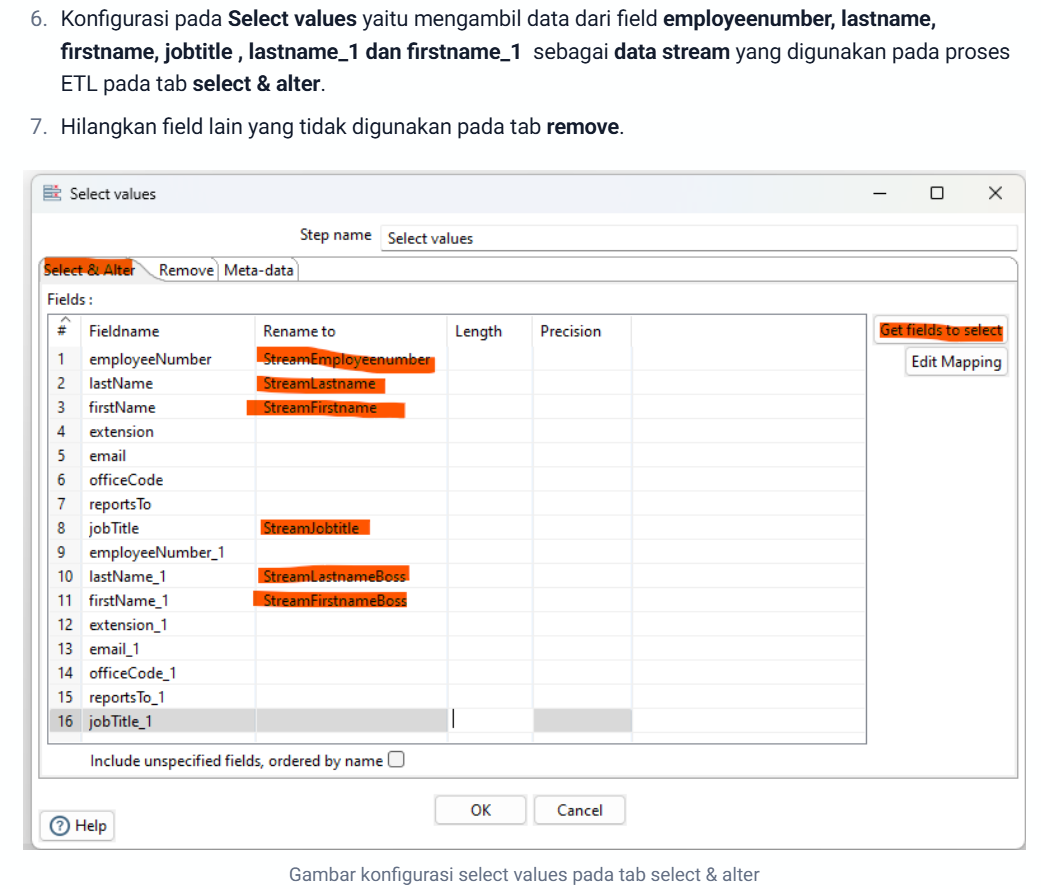
* Configure

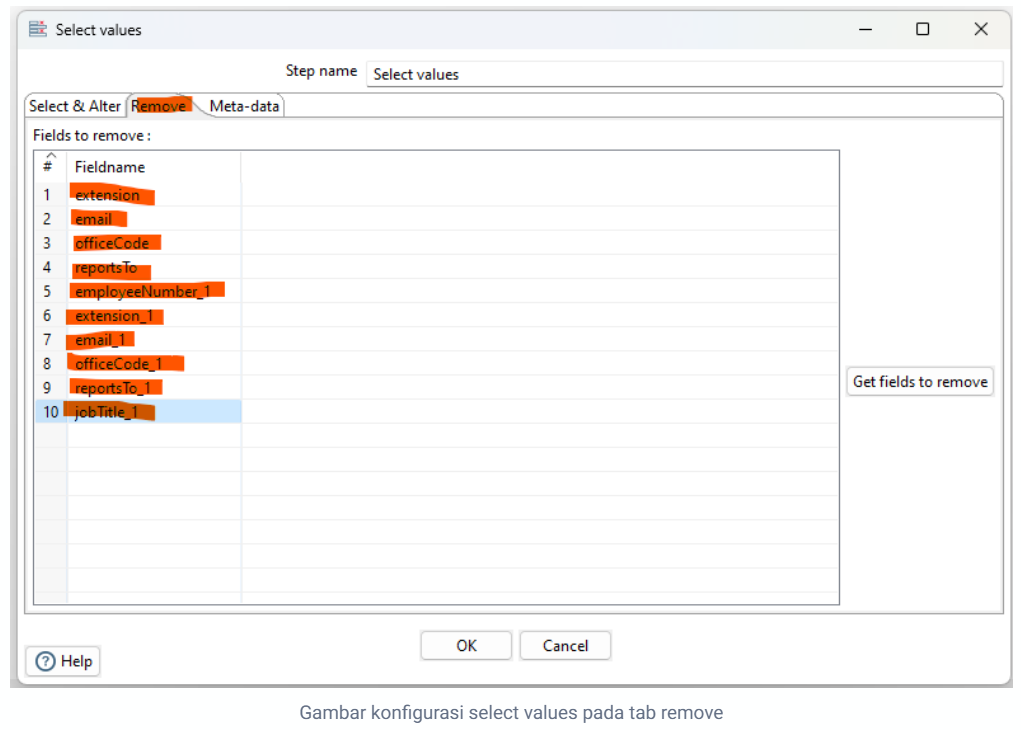




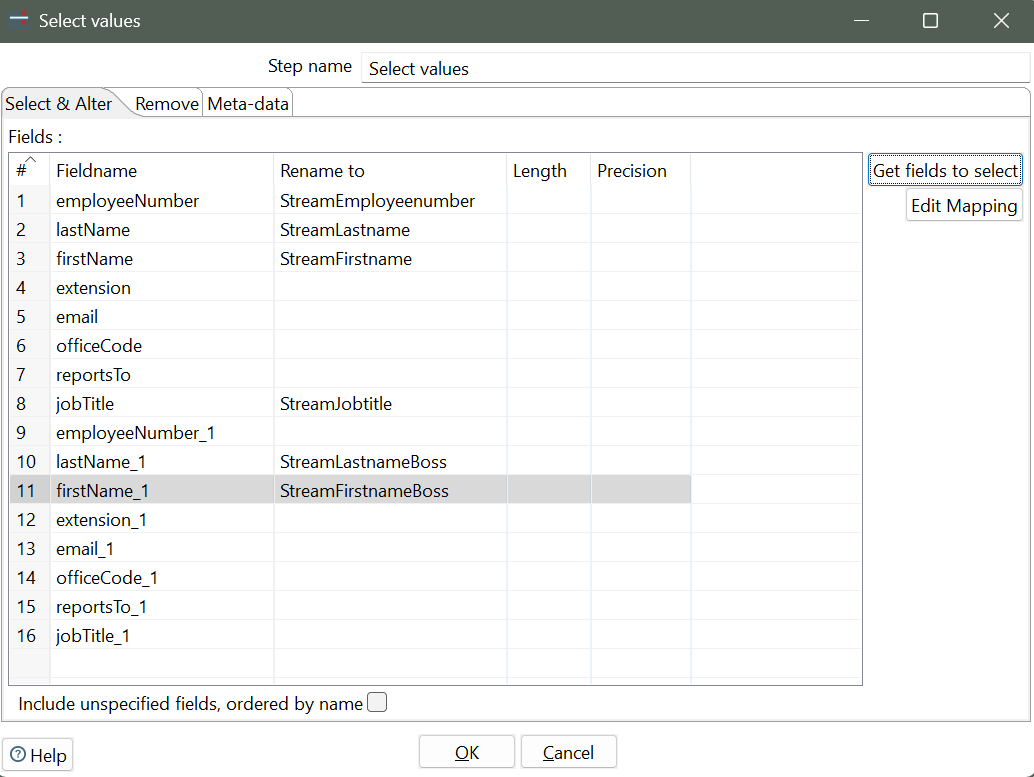
* Configure

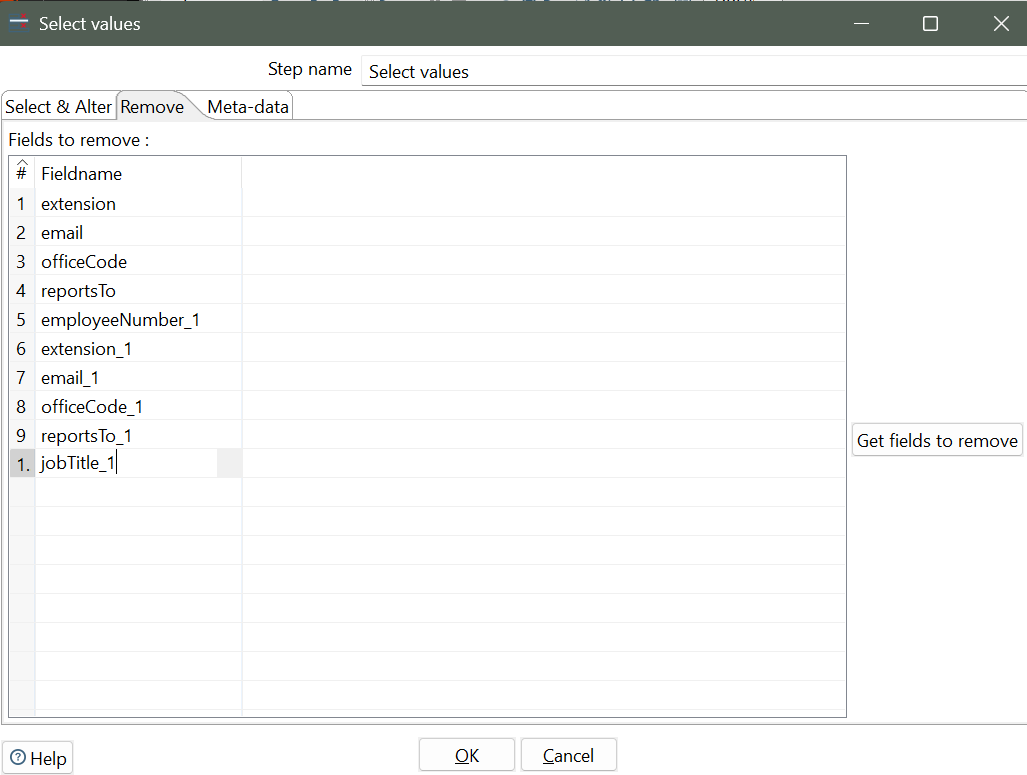






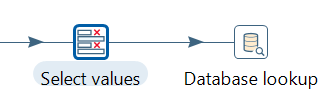
* Configure

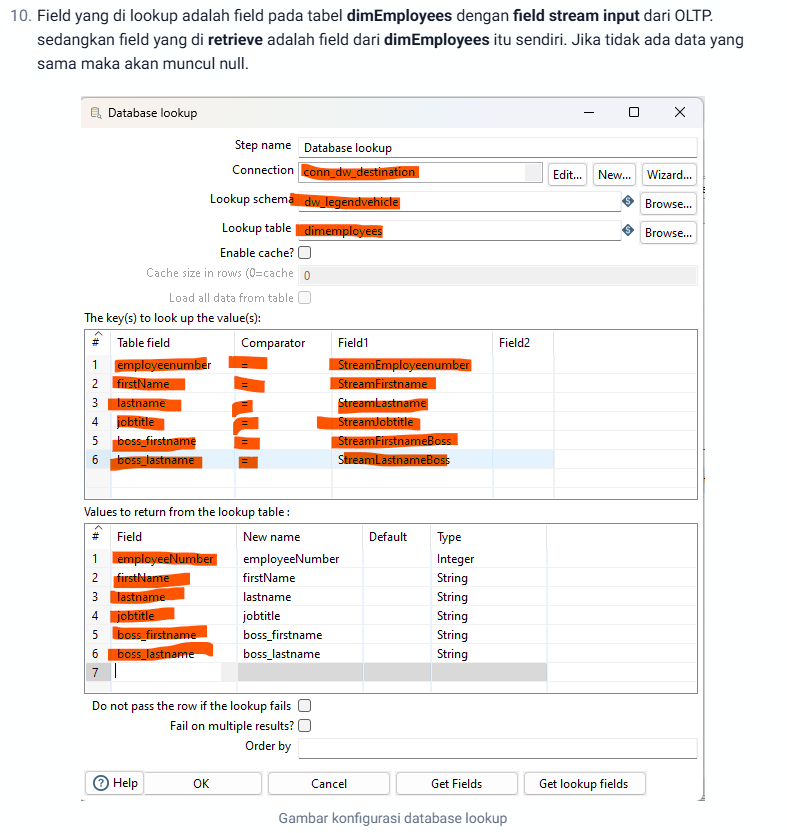


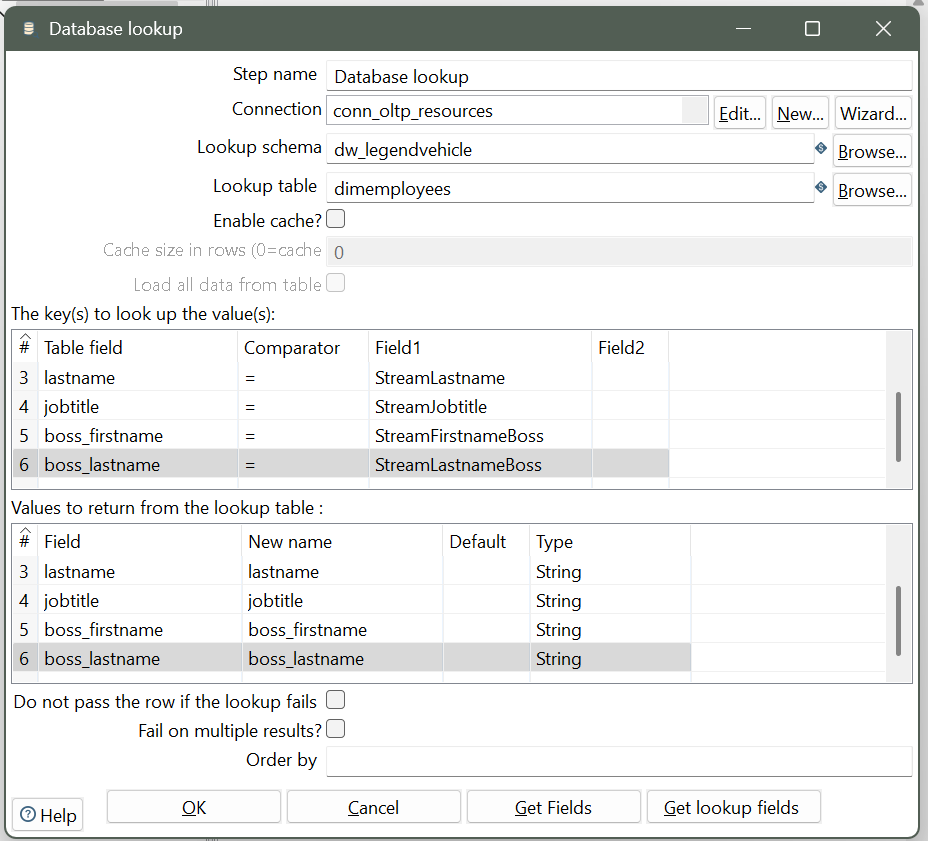




* Configure

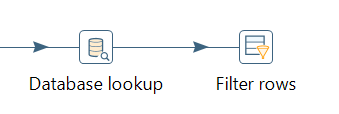


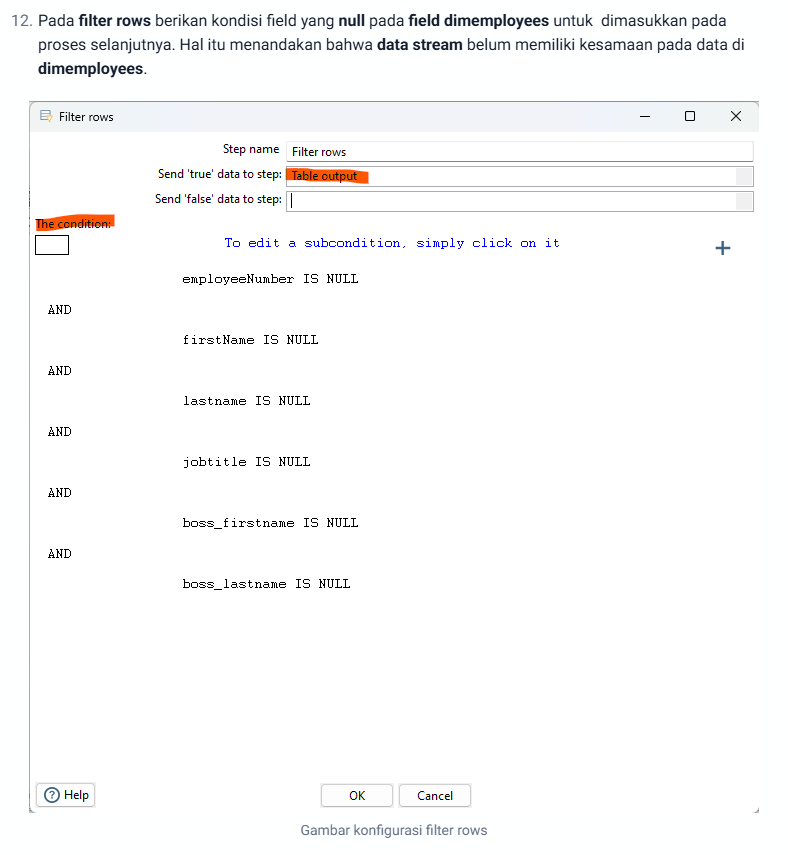


* Configure

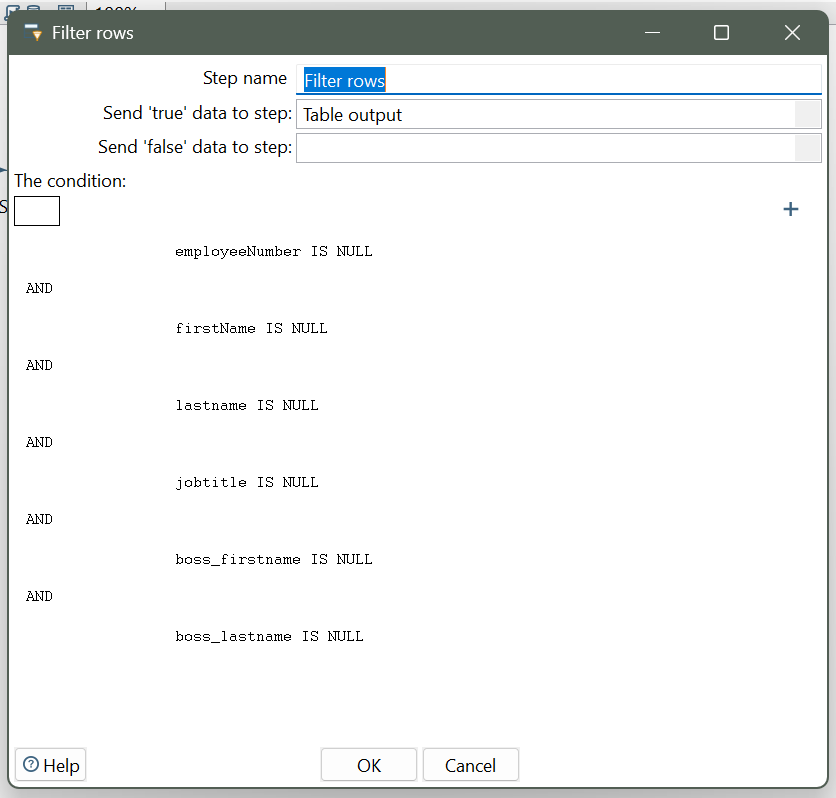


* Configure



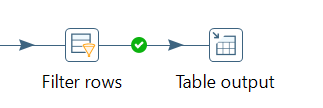


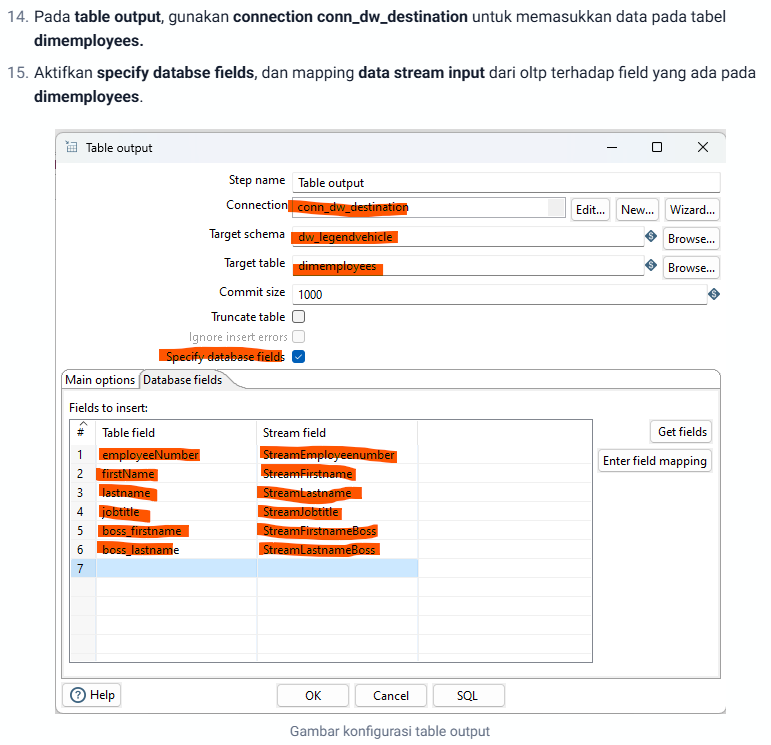
* Configure



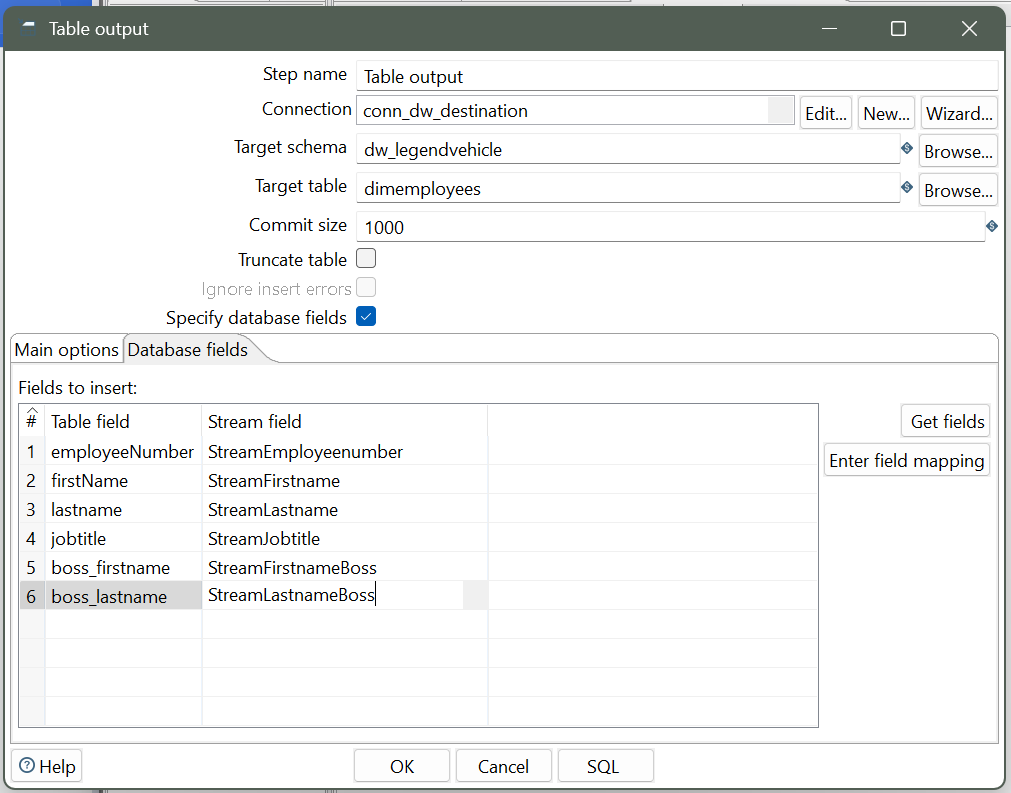


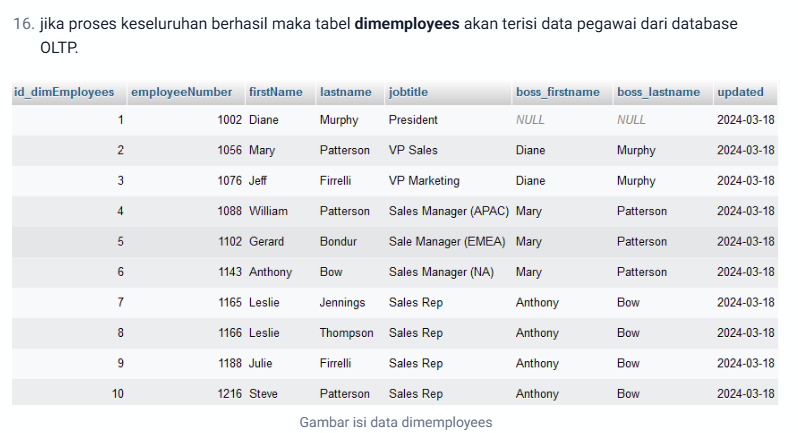
* Configure



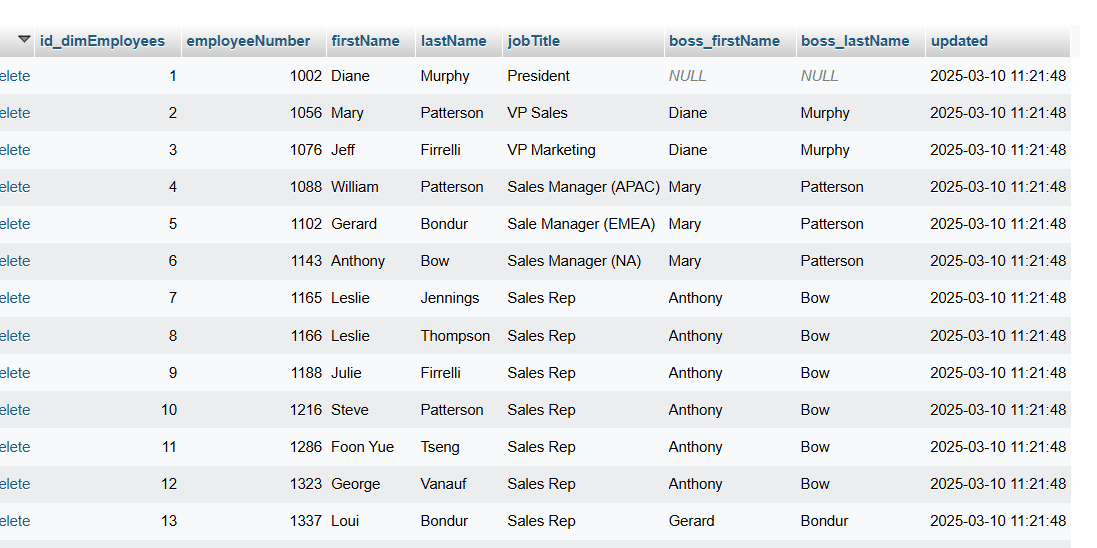


* Configure





* Database



**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.

* Table Execution Result

|  |  |  |  |
| --- | --- | --- | --- |
| Proses Object | SS data input | SS data output | Keterangan |
| Table input |  |  | Pertama konfigurasi Table Input dengan LEFT JOIN pada tabel employees.  Kemudian hasil querynya, yaitu daftar karyawan dan atasan mereka. |
| Select Values |  |  | pertama mapping kolom dengan mengganti nama beberapa kolom di langkah Select values. Kemudian hasil eksekusi querynya, menunjukkan data karyawan dan atasan mereka telah dipetakan sesuai nama baru. |
| Database Lookup |  |  | konfigurasi database Lookup untuk mencocokkan data karyawan dengan tabel referensi dimemployees. kemudian hasil eksekusi nenunjukkan data karyawan dengan informasi tambahan dari lookup table. |

|  |  |  |  |
| --- | --- | --- | --- |
| Filter Rows |  |  | konfigurasi Filter Rows untuk menyaring data dengan kondisi IS NULL pada beberapa kolom. kemudian hasil filter, hanya menunjukkan data yang memenuhi kondisi tetap ada dalam output. |
| Table Output |  |  | Pertama langkah Table Output yang digunakan untuk menyimpan data hasil transformasi ke dalam tabel dimemployees di skema dw\_legendvehicle. kemudian hasil eksekusi, menampilkan data yang akan disimpan, termasuk kolom StreamEmployeeNumber, StreamLastname, StreamFirstname, StreamJobtitle, dan informasi tambahan lainnya. |

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

* Jika proses dijalankan kembali, data kemungkinan akan menjadi redundan tergantung pada pengaturan di langkah Table Output. Jika opsi "Truncate table" tidak dicentang, maka data baru akan ditambahkan tanpa menghapus data lama, sehingga menyebabkan duplikasi. Namun, jika opsi tersebut dicentang, tabel akan dikosongkan terlebih dahulu sebelum memasukkan data baru, sehingga tidak terjadi duplikasi. Selain itu, jika terdapat primary key atau unique constraint, proses insert ulang bisa gagal atau hanya memasukkan data yang berbeda dari yang sudah ada.

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

* Ketika nama baru ditambahkan ke tabel employees di OLTP dan transformasi dijalankan kembali, hasilnya bergantung pada mekanisme penyimpanan data. Jika transformasi hanya menambahkan data baru tanpa menghapus yang lama, maka data baru akan muncul di hasil transformasi bersama dengan data sebelumnya. Jika ada mekanisme deduplikasi, hanya perubahan yang diperlukan yang akan dicatat. Namun, jika tabel tujuan dikosongkan terlebih dahulu sebelum memasukkan data baru, maka hanya data dari eksekusi terbaru yang tersimpan, termasuk data yang baru ditambahkan.