LAPORAN PRAKTIKUM JOBSHEET 3 (A & B)

DATA WAREHOUSE

Disusun Oleh:

Niriza Lailaumi Hidayat (2341760072)

SIB 2B

SISTEM INFORMASI BISNIS

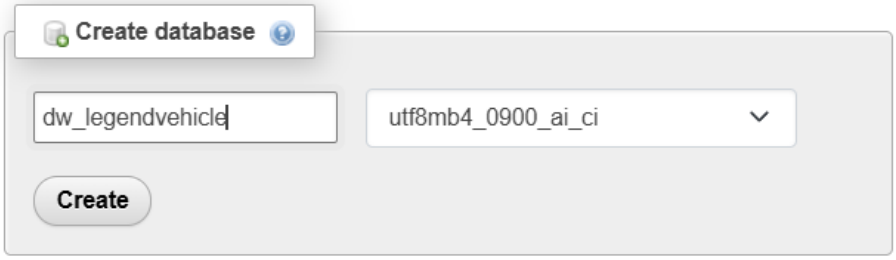
JURUSAN TEKNOLOGI INFORMASI

POLITEKNIK NEGERI MALANG

2025

A. Dimensi Waktu

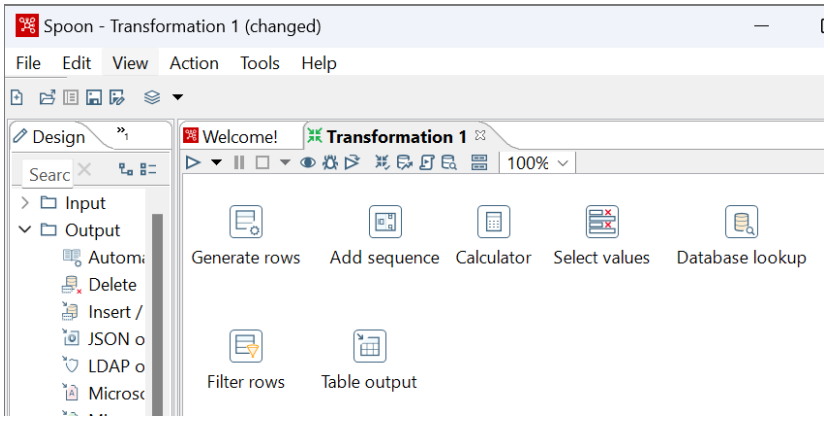
1. Membuat database yang digunakan sebagai OLAP dengan nama dw_legendvehicle.



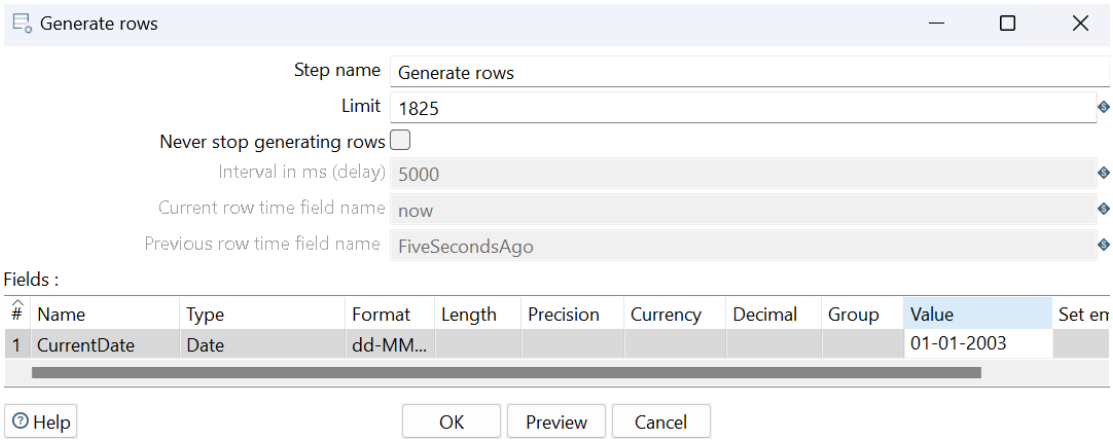
2. Membuat tabel untuk menyimpan data master waktu atau yang disbeut dengan tabel dimensi dengan nama dimDate.

```sql
CREATE TABLE dimDate (
    id_dimDate INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
    `date` DATE NOT NULL,
    year INT NOT NULL,
    month INT NOT NULL,
    day INT NOT NULL
);
```

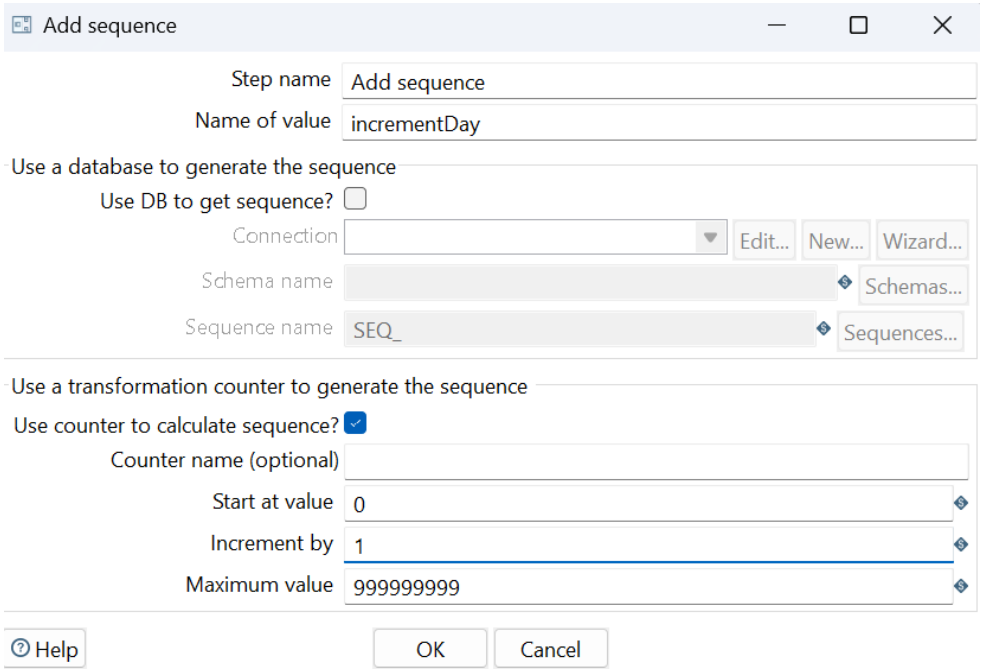
3. Drag and drop beberapa objek seperti gambar di bawah.



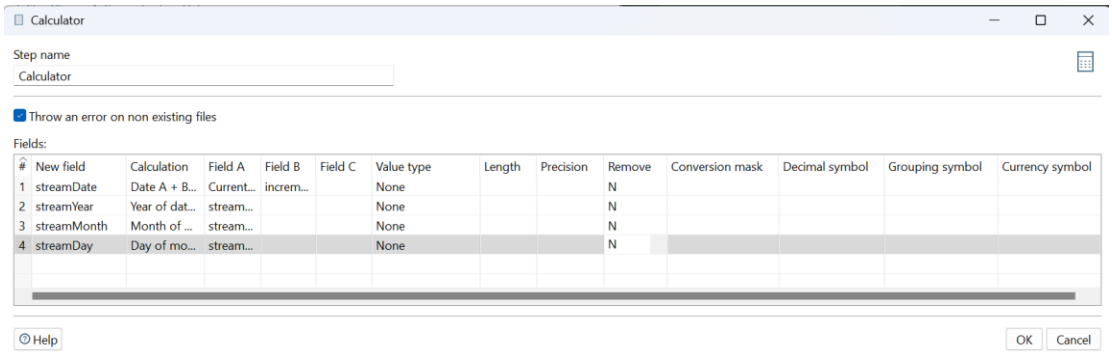
4. Konfiguransi Generate Rows dengan merubah limit menjadi 1825 dimana memiliki arti bahwa data yang akan dibuat sebanyak 1825 data. 1825 merupakan jumlah hari dalam 5 tahun (365 hari x 5 tahun).
5. Membuat fields baru bernama CurrentDate dengan type data Date dan format dd-MM-yyyy serta value awal 01-01-2003.



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

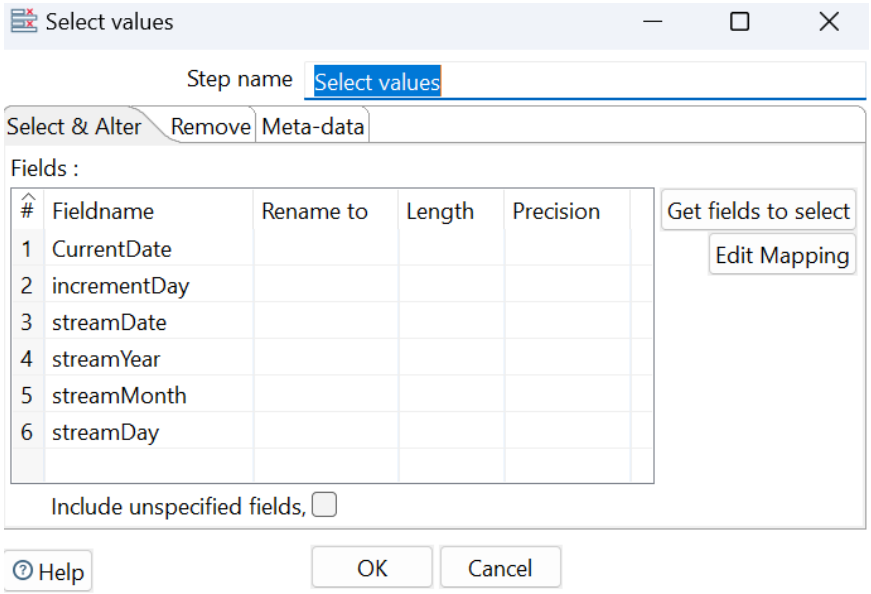


8. Konfigurasi Calculator

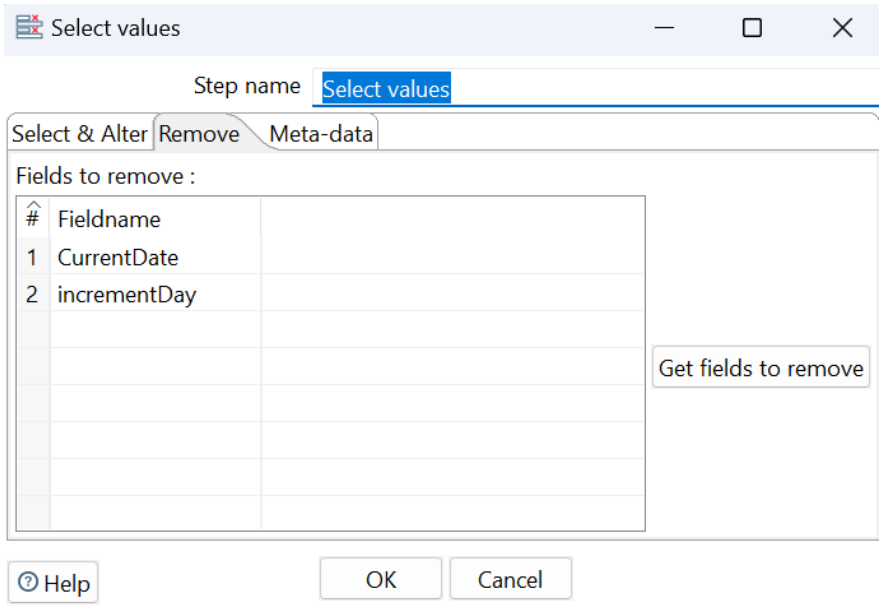


9. Hubungkan output dari calculator menuju Select values

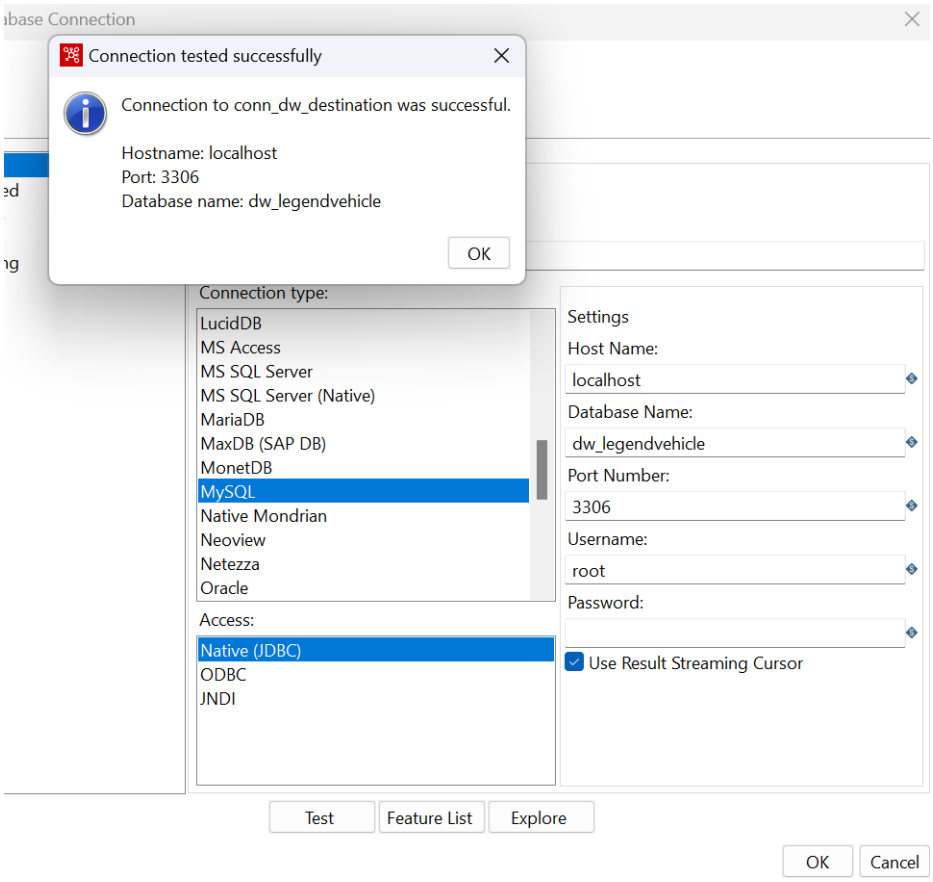
10. Konfigurasi pada select values adalah dengan menekan tombol Get fields to select pada tab Select & Alter. Secara otomatis semua fields dari data input akan muncul.



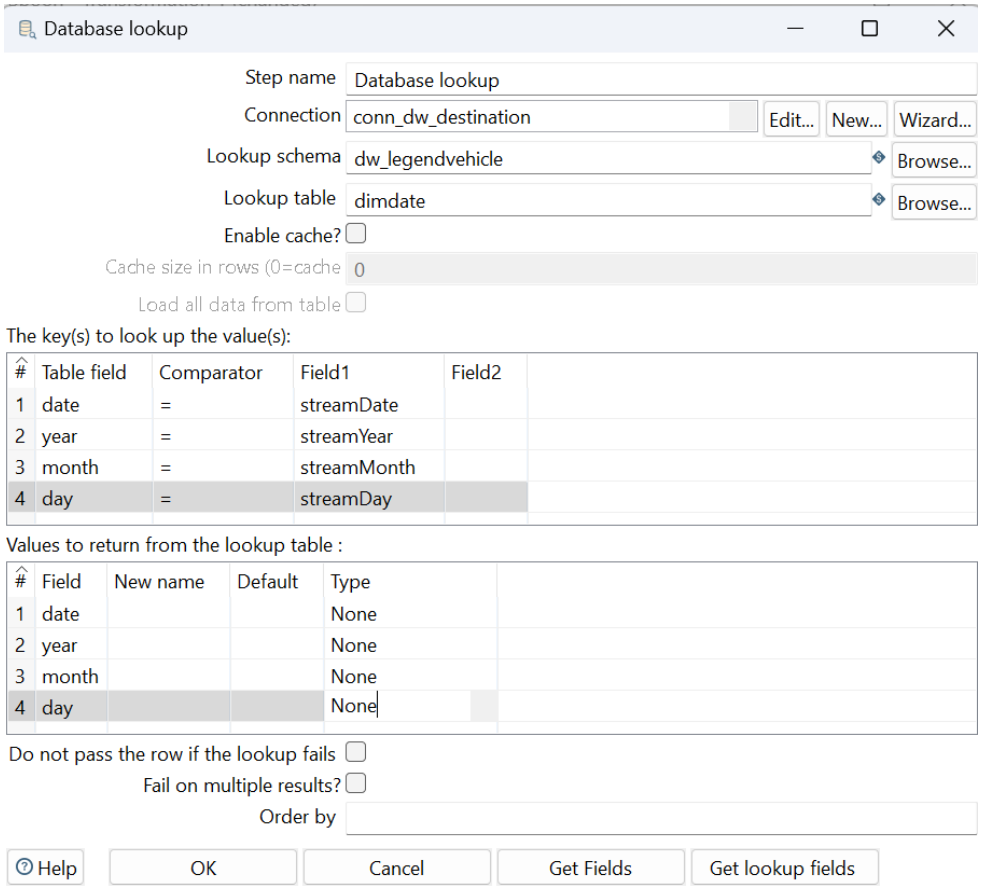
11. Dikarenakan tidak semua fields digunakan, maka pada tab Remove diisikan fields CurrentDate dan incrementDay dikarenakan kedua fields tersebut tidak digunakan.



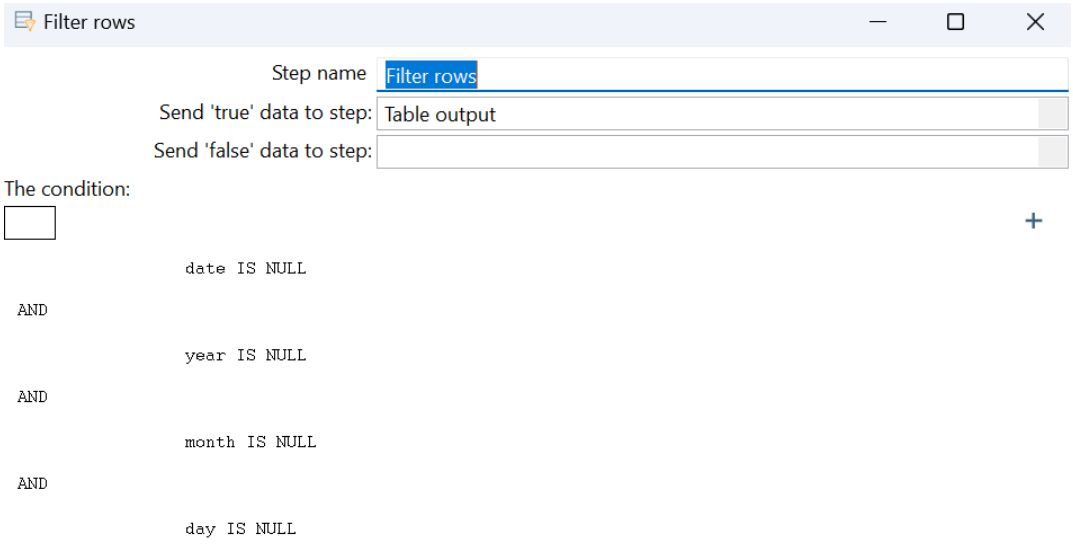
12. Hubungkan output select values menuju database lookup.
13. Sebelum melakukan konfigurasi pada database lookup, buatlah koneksi terlebih dahulu pada database melalui File - New - Database Connection. Gunakan Connection type MySQL dengan host name, database name, port number, username dan password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



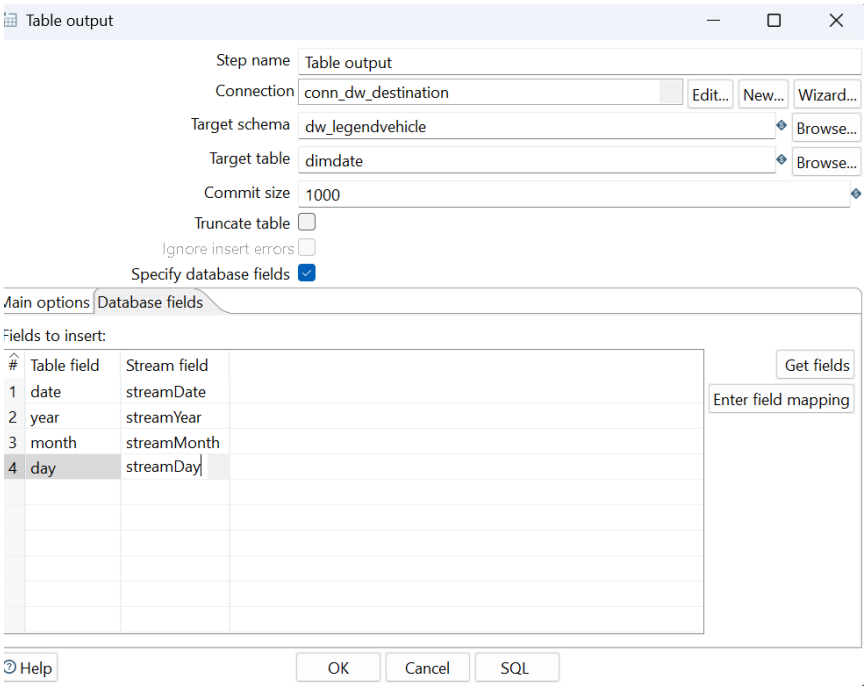
14. Konfigurasi pada database lookup adalah dengan memberikan connection dengan koneksi yang sudah dibuat pada step sebelumnya. dengan schema nama database yang digunakan dan tabel dimdate yang telah dibuat pada langkah pertama.
15. Field yang akan dicek untuk melihat kesamaan isi datanya agar tidak 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
16. Field yang akan di retrive adalah field yang ada pada table dimDate yaitu date, year, month, dan day



17. Hubungkan output dari database lookup dengan filter rows.
18. 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).



19. Hubungkan output dari filter rows menuju table output.
20. Konfigurasi pada table output adalah memberikan koneksi pada conn_dw_destination dengan schema dw_legendvehicle dan table dimdate.
21. Aktifkan specify database fields.
22. 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.

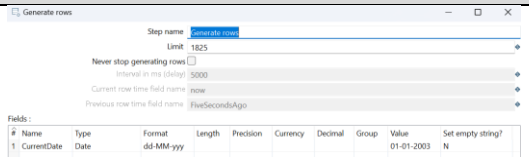
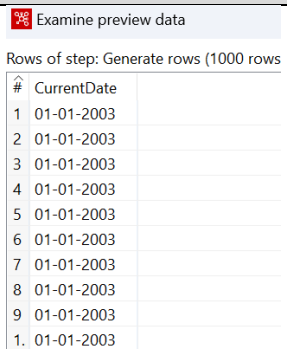
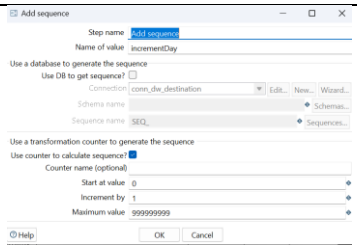
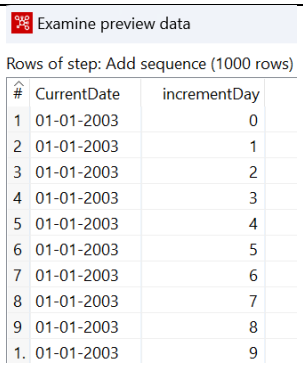
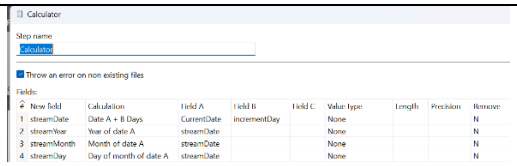
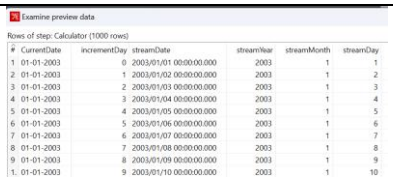


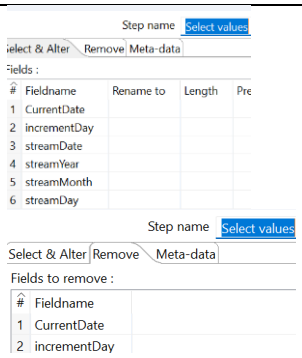
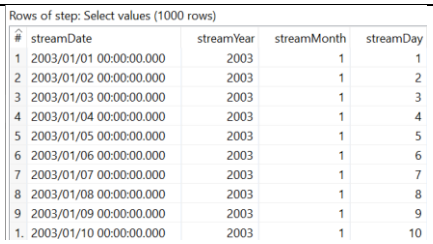
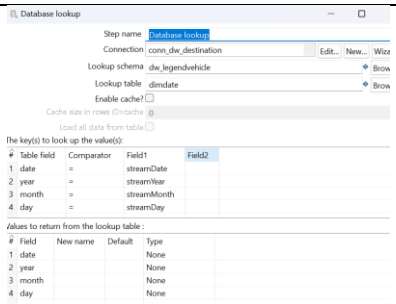
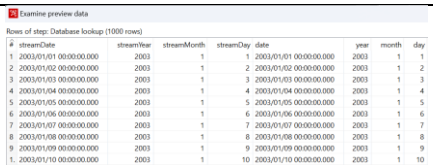
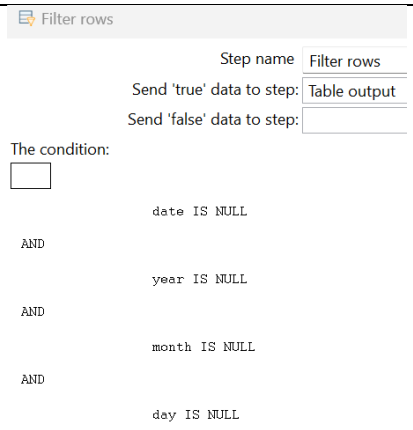
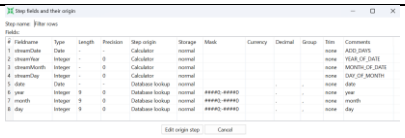
23. cek isi table dimdate pada database. Jika sukses maka pada table dimdate akan terisi 1825 data.

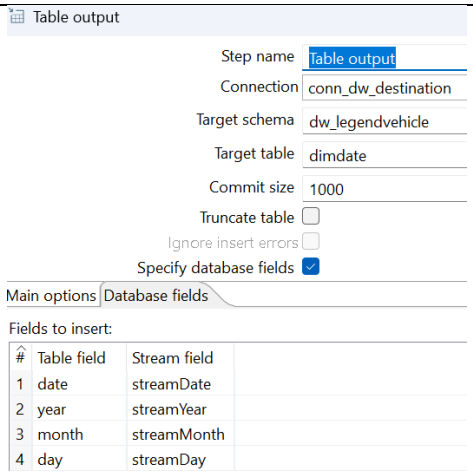
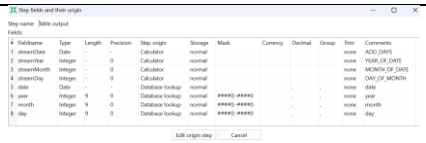
	id_dimDate	date	year	month	day
Edit Copy Delete	1	2003-01-01	2003	1	1
Edit Copy Delete	2	2003-01-02	2003	1	2
Edit Copy Delete	3	2003-01-03	2003	1	3
Edit Copy Delete	4	2003-01-04	2003	1	4
Edit Copy Delete	5	2003-01-05	2003	1	5
Edit Copy Delete	6	2003-01-06	2003	1	6
Edit Copy Delete	7	2003-01-07	2003	1	7
Edit Copy Delete	8	2003-01-08	2003	1	8
Edit Copy Delete	9	2003-01-09	2003	1	9
Edit Copy Delete	10	2003-01-10	2003	1	10
Edit Copy Delete	11	2003-01-11	2003	1	11

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			Input hanya menetapkan 01-01-2003 tanpa perubahan, sementara output menghasilkan 1000 baris dengan nilai yang sama. Interval 5000 ms tidak mempengaruhi karena tidak ada logika penambahan tanggal. Untuk perubahan nilai, diperlukan transformasi tambahan.
Add Sequence			Proses Add Sequence menambahkan kolom incrementDay yang dimulai dari 0 dan bertambah 1 setiap baris, sementara nilai CurrentDate tetap 01-01-2003. Meskipun ada urutan angka, tanggal belum berubah karena belum dikombinasikan dengan transformasi lain seperti Formula untuk menambahkan hari ke tanggal.
Calculator			Pada proses Calculator, kolom streamDate dibuat dengan menambahkan nilai incrementDay ke CurrentDate, sehingga menghasilkan tanggal yang bertambah setiap baris. Selain itu, streamYear, streamMonth, dan streamDay diekstrak dari streamDate untuk mendapatkan informasi tahun, bulan, dan hari secara terpisah. Hasilnya, tanggal kini berubah secara dinamis

			sesuai dengan urutan yang diberikan.
Select values			Pada proses Select Values, kolom CurrentDate dan incrementDay dihapus, sehingga hasil akhirnya hanya menyisakan streamDate, streamYear, streamMonth, dan streamDay. Data kini lebih bersih dengan hanya menampilkan informasi tanggal yang telah dihitung sebelumnya.
Database lookup			Proses Database Lookup mencocokkan data dari tabel referensi dimdate berdasarkan kolom year, month, dan day dengan nilai streamYear, streamMonth, dan streamDay. Pada output, kolom date berhasil ditambahkan berdasarkan kecocokan nilai dari lookup table, sementara data input tetap dipertahankan. Hal ini menunjukkan bahwa lookup berhasil menghubungkan data tanggal dengan informasi yang sesuai.
Filter rows			Proses Filter Rows menyaring data berdasarkan kondisi date IS NULL, year IS NULL, month IS NULL, dan day IS NULL. Jika semua nilai tersebut NULL, data akan dikirim ke langkah Table Output, sedangkan data yang memiliki nilai pada salah satu kolom akan diteruskan ke langkah lain yang belum ditentukan. Hal ini memastikan hanya data yang gagal ditemukan dalam proses Database Lookup

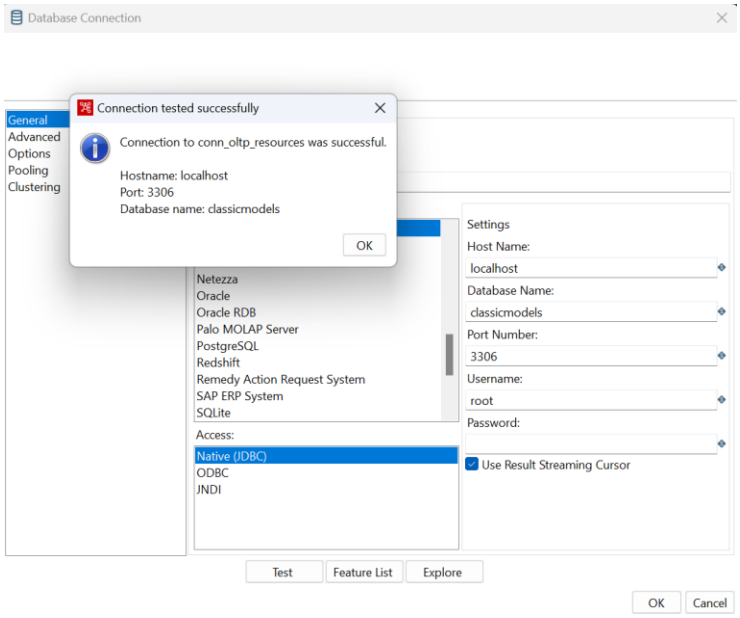
			yang akan difilter dan diproses lebih lanjut.
Table Output			Proses Table Output menyimpan data hasil transformasi ke dalam tabel dimdate pada skema dw_legendvehicle. Data yang dimasukkan terdiri dari streamDate, streamYear, streamMonth, dan streamDay, yang masing-masing dipetakan ke kolom date, year, month, dan day dalam tabel tujuan. Hal ini memastikan bahwa data yang telah melalui proses lookup dan filter dapat disimpan dengan struktur yang sesuai dalam data warehouse.

B. Dimensi Pegawai

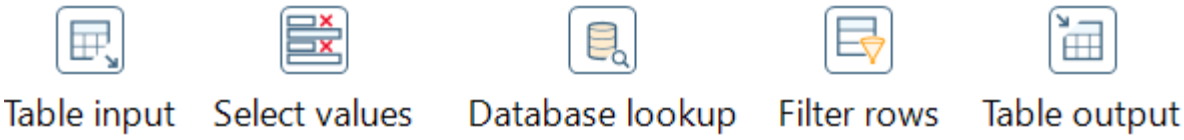
1. Membuat tabel dimPegawai pada dw_legendvehicle

```sql
CREATE TABLE dimEmployees (
    id_dimEmployees INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
    employeeNumber INT(11),
    firstName VARCHAR(50),
    lastName VARCHAR(50),
    jobTitle VARCHAR(50),
    boss_firstName VARCHAR(50),
    boss_lastName VARCHAR(50),
    updated DATE DEFAULT (CURRENT_TIMESTAMP)
);
```

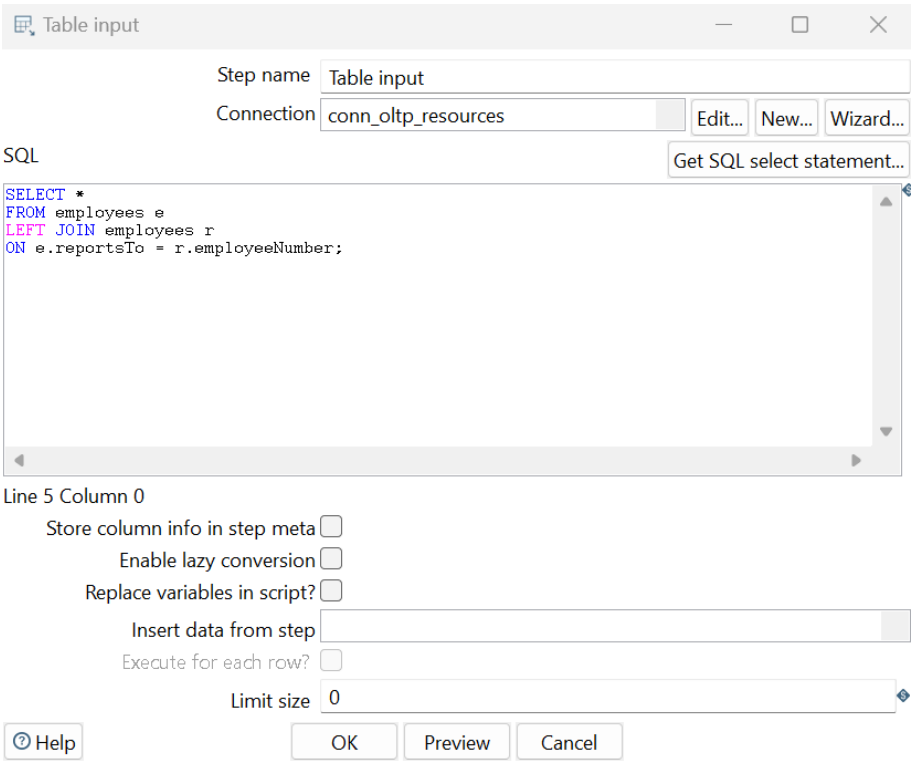
2. Pada PDI Spoon buatlah koneksi baru dengan nama conn_oltp_resources yang menghubungkan dengan database oltp. sesuaikan hostname, database name, port number, username dan password dengan keadaan pada device masing-masing.



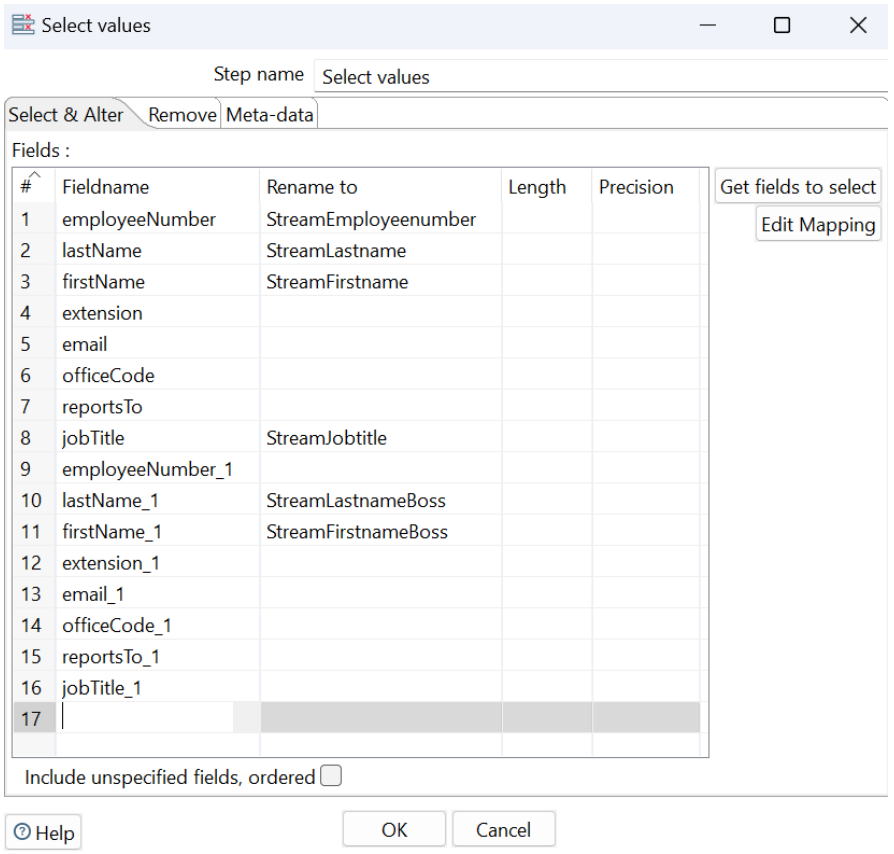
3. Drag and drop beberapa objek sebagai berikut:
    - Table input: digunakan mengambil data dari database OLTP.
    - Select values: memeilih field yang digunakan untuk proses Transform dan Load.
    - Database lookup: digunakan untuk melihat data pada tabel dimEmployees untuk memastikan data tidak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

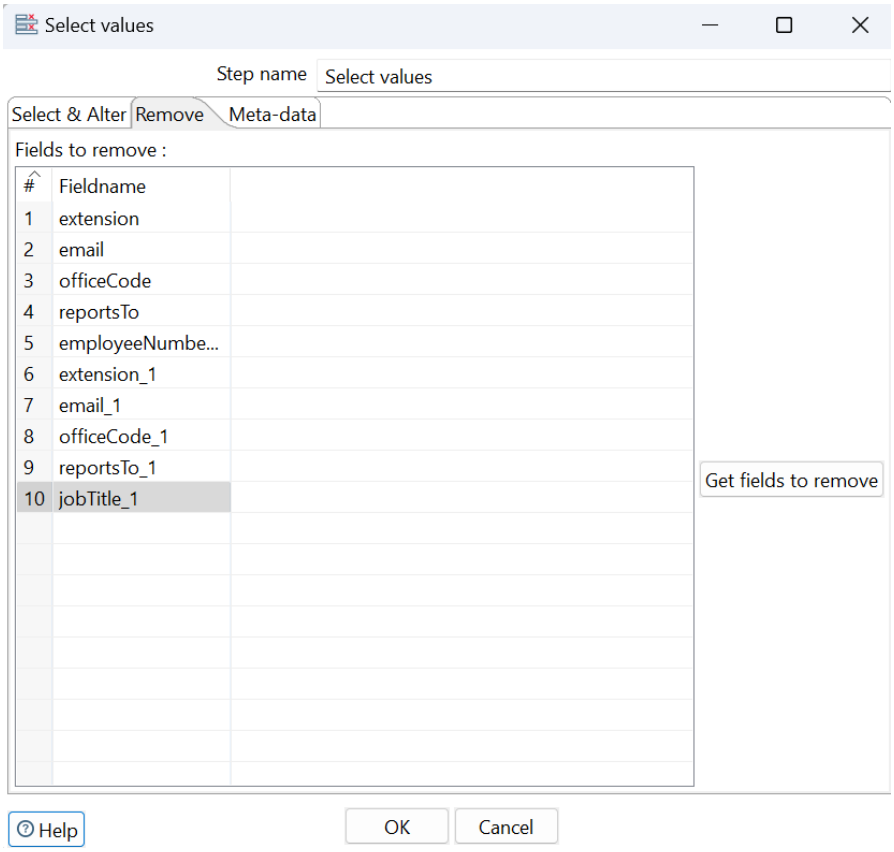


4. Konfigurasi tabel input dengan menghubungkan Connection pada konesi conn_oltp_resources. Untuk mengambil data sumber menggunakan query dibawah ini.

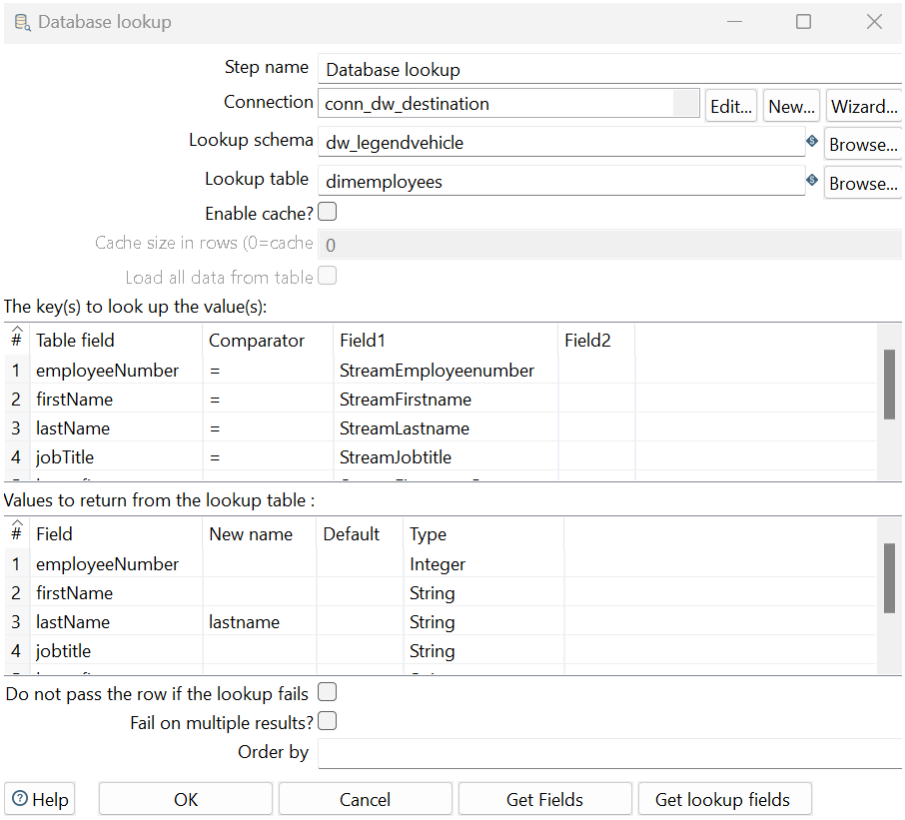


5. Hubungkan output table input pada select values.
6. Konfigurasi pada Select values yaitu mengambil data dari field employeenumber, lastname, firstname, jobtitle, lastname_1 dan firstname_1 sebagai data stream yang digunakan pada proses ETL pada tab select & alter.
7. Hilangkan field lain yang tidak digunakan pada tab remove.

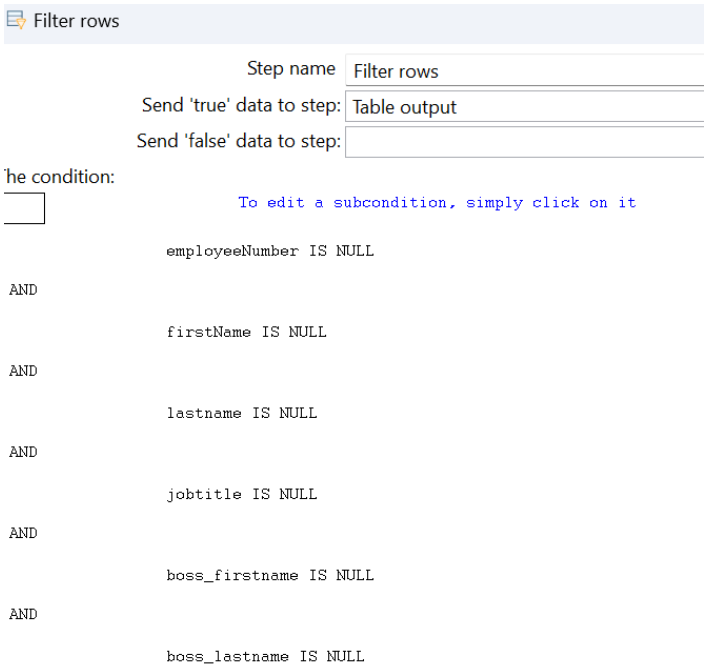

Select values
Step name
Select values
Select & Alter
Remove
Meta-data
Fields :
Fieldname Rename to Length Precision
1 employeeNumber StreamEmployeenumber
2 lastName StreamLastname
3 firstName StreamFirstname
4 extension
5 email
6 officeCode
7 reportsTo
8 jobTitle StreamJobtitle
9 employeeNumber_1
10 lastName_1 StreamLastnameBoss
11 firstName_1 StreamFirstnameBoss
12 extension_1
13 email_1
14 officeCode_1
15 reportsTo_1
16 jobTitle_1
17
Get fields to select
Edit Mapping
Include unspecified fields, ordered
Help
OK
Cancel


Select values
Step name
Select values
Select & Alter
Remove
Meta-data
Fields to remove :
Fieldname
1 extension
2 email
3 officeCode
4 reportsTo
5 employeeNumbe...
6 extension_1
7 email_1
8 officeCode_1
9 reportsTo_1
10 jobTitle_1
Get fields to remove
Help
OK
Cancel

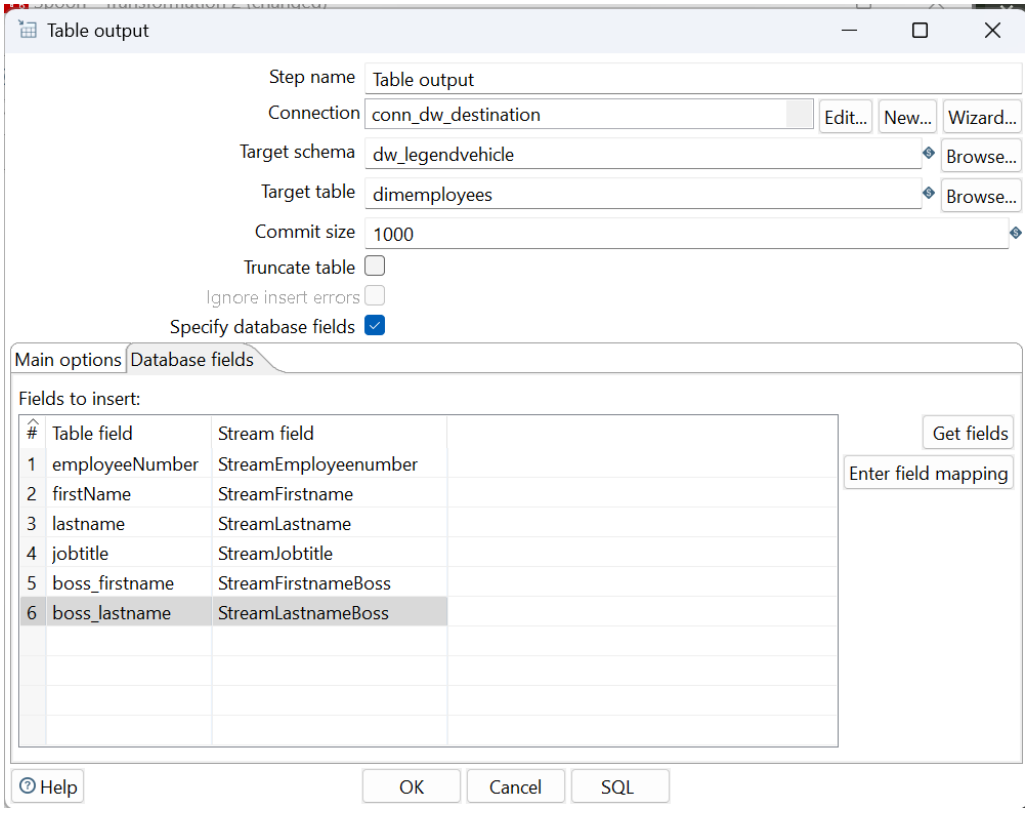
8. Hubungkan output select values pada database lookup
9. Konfigurasi pada database lookup adalah dengan menghubungkan koneksi pada conn_dw_destination dengan table lookup dimEmployees yang telah dibuat pada tahap pertama.
10. Field yang di lookup adalah field pada tabel dimEmployees dengan field stream input dari OLTP. sedangkan field yang di retrieve adalah field dari dimEmployees itu sendiri. Jika tidak ada data yang sama maka akan muncul null.



11. Hubungkan output databse lookup dengan filter rows.
12. 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.



13. Hubungkan output dari filter rows dengan table output.
14. Pada table output, gunakan connection conn_dw_destination untuk memasukkan data pada tabel dimemployees.
15. Aktifkan specify databse fields, dan mapping data stream input dari oltp terhadap field yang ada pada dimemployees



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

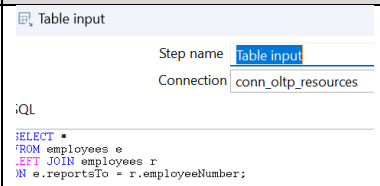
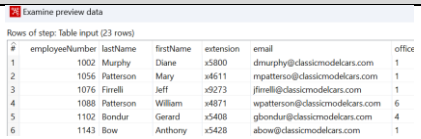
Database: dw_legendvehicle » Table: dimemployees

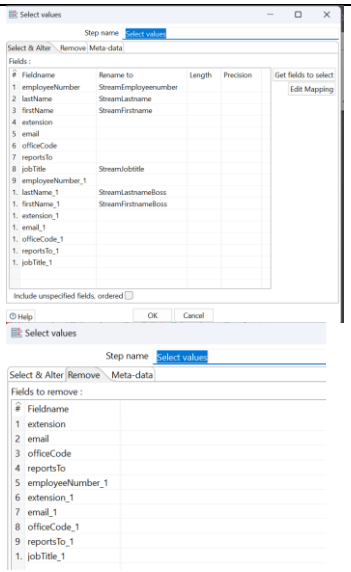
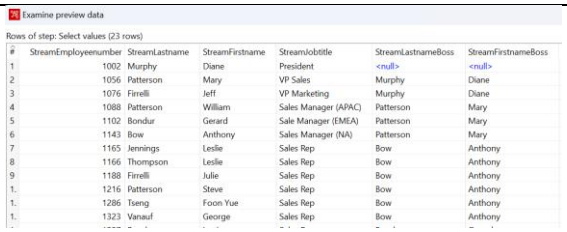
SQL | Search | Insert | Export | Import | Privileges | Operations | Triggers

	id_dimEmployees	employeeNumber	firstName	lastName	jobTitle	boss_firstName	boss_lastName	updated
e	1	1002	Diane	Murphy	President	*NULL*	*NULL*	2025-03-08
e	2	1056	Mary	Patterson	VP Sales	Diane	Murphy	2025-03-08
e	3	1076	Jeff	Firrelli	VP Marketing	Diane	Murphy	2025-03-08
e	4	1088	William	Patterson	Sales Manager (APAC)	Mary	Patterson	2025-03-08
e	5	1102	Gerard	Bondur	Sale Manager (EMEA)	Mary	Patterson	2025-03-08
e	6	1143	Anthony	Bow	Sales Manager (NA)	Mary	Patterson	2025-03-08
e	7	1165	Leslie	Jennings	Sales Rep	Anthony	Bow	2025-03-08
e	8	1166	Leslie	Thompson	Sales Rep	Anthony	Bow	2025-03-08
e	9	1188	Julie	Firrelli	Sales Rep	Anthony	Bow	2025-03-08
e	10	1216	Steve	Patterson	Sales Rep	Anthony	Bow	2025-03-08
e	11	1286	Foon Yue	Tseng	Sales Rep	Anthony	Bow	2025-03-08
e	12	1323	George	Vanauf	Sales Rep	Anthony	Bow	2025-03-08
e	13	1337	Loui	Bondur	Sales Rep	Gerard	Bondur	2025-03-08
e	14	1370	Gerard	Hernandez	Sales Rep	Gerard	Bondur	2025-03-08
e	15	1401	Pamela	Castillo	Sales Rep	Gerard	Bondur	2025-03-08
e	16	1501	Larry	Bott	Sales Rep	Gerard	Bondur	2025-03-08
e	17	1504	Barry	Jones	Sales Rep	Gerard	Bondur	2025-03-08
e	18	1611	Andy	Fixter	Sales Rep	William	Patterson	2025-03-08
e	19	1612	Peter	Marsh	Sales Rep	William	Patterson	2025-03-08

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			Proses Table Input mengambil data dari tabel employees dalam koneksi conn_oltp_resources. Query yang digunakan melakukan LEFT JOIN pada tabel yang sama untuk menghubungkaProses Table Input mengambil data dari tabel employees dalam

			koneksi conn_oltp_resources. Query yang digunakan melakukan LEFT JOIN pada tabel yang sama untuk menghubungkan setiap karyawan dengan atasannya berdasarkan kolom reportsTo. Hasil transformasi ini menghasilkan dataset yang mencakup informasi karyawan, termasuk employeeNumber, lastName, firstName, extension, email, dan lainnya, dengan total 23 baris.n setiap karyawan dengan atasannya berdasarkan kolom reportsTo. Hasil transformasi ini menghasilkan dataset yang mencakup informasi karyawan, termasuk employeeNumber, lastName, firstName, extension, email, dan lainnya, dengan total 23 baris.
Select values			Input memiliki lebih banyak kolom, termasuk extension, email, officeCode, dan beberapa versi duplikat dengan akhiran _1. Dalam proses Select Values, beberapa kolom diubah namanya, seperti employeeNumber menjadi StreamEmployeenumber dan lastName_1 menjadi StreamLastnameBoss, sementara kolom yang tidak diperlukan dihapus. Output akhirnya lebih

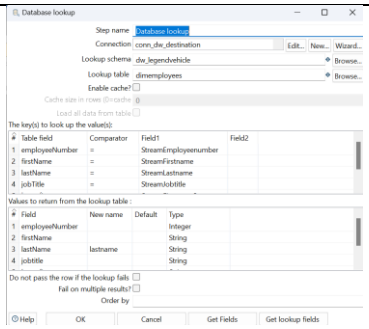
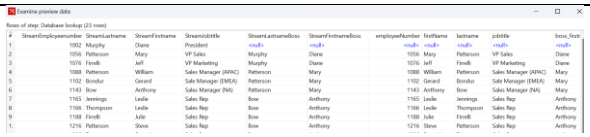
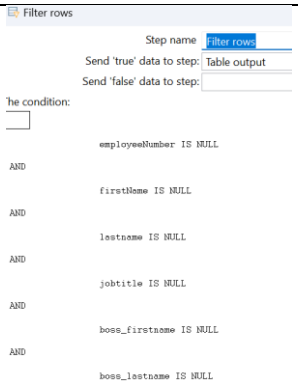
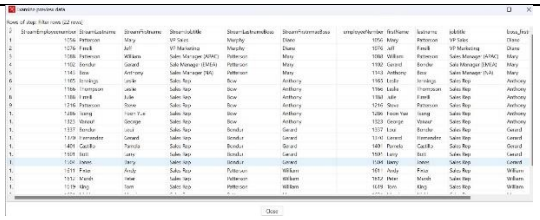
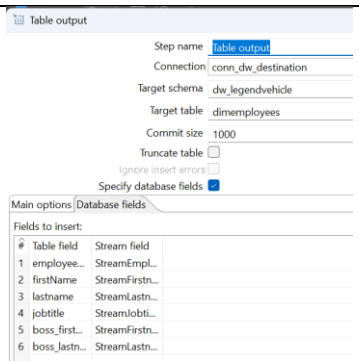
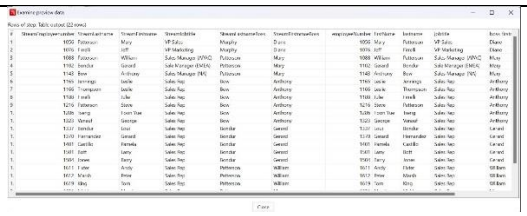
			ringkas, hanya menyisakan informasi penting tentang karyawan dan atasan mereka.
Database lookup			Proses Database Lookup mencocokkan data input berdasarkan employeeNumber, firstname, dan lastname, lalu menambahkan kolom dari tabel referensi, seperti jobtitle. Hasilnya, output berisi data awal yang diperkaya dengan informasi tambahan dari lookup table.
Filter rows			Proses Filter Rows menyaring data berdasarkan kondisi IS NULL untuk beberapa kolom, seperti employeeNumber, firstName, lastName, jobTitle, boss_firstname, dan boss_lastname. Data yang memenuhi semua kondisi ini akan difilter dan tidak muncul dalam output. Hasilnya, output hanya berisi baris dengan data lengkap, menghindari entri yang memiliki nilai kosong di kolom-kolom penting.

Table Output			Proses Table Output memasukkan data hasil filter ke dalam tabel dimemployees di database dw_legendvehicle. Data yang telah diperkaya melalui Database Lookup dan difilter melalui Filter Rows akan disimpan sesuai dengan pemetaan antara kolom sumber (Stream field) dan kolom tujuan (Table field). Hasilnya, hanya data yang valid dan lengkap yang masuk ke tabel target.

2. Jika proses itu di ulangi (di run kembali) apakah data akan redudant?
   - Jika proses di ulangi, maka tidak akan ada redudansi data karena terdapat mekanisme yang mencegah terjadinya duplikasi yaitu Database Lookup yang mengecek apakah data sudah ada sebelum dimasukkan dan Filter Rows yang menyaring data sehingga tidak terjadi duplikasi.
3. Tambahkan nama anda pada table employee di OLTP. jalankan kembali transformasi ini. Amati hasilnya, apa yang terjadi?
   - Menambahkan nama pada tabel employee di OLTP

Run SQL query/queries on table classicmodels.employees:

```
INSERT INTO `employees`(`employeeNumber`, `lastName`, `firstName`, `extension`,
`email`, `officeCode`, `reportsTo`, `jobTitle`) VALUES
(2508,'Lailaumi','Niriza','x8004','niriza@gmail.com', 1, NULL, 'Designer');
```

Data berhasil ditambahkan pada tabel employee di OLTP

1 row inserted. (Query took 0.0109 seconds.)

```
INSERT INTO `employees`(`employeeNumber`, `lastName`, `firstName`, `extension`, `email`,
`officeCode`, `reportsTo`, `jobTitle`) VALUES
(2508,'Lailaumi','Niriza','x8004','niriza@gmail.com', 1, NULL, 'Designer');
```

[Edit inline] [Edit] [Create PHP code]

Pada employeeNumber diisi dengan data yang unik agar tidak terjadi duplikasi data.