

DATA WEREHOUSE

JOBSHEET 3



Oleh:

BAGAS SATRIA YUDHO NUGRAHA

SIB 2B 04 2341760108

**D-IV SISTEM INFORMASI BISNIS
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG**

1.

Buatlah sebuah database yang digunakan sebagai OLAP dengan nama dw_LegendVehicle.

Copy

```
CREATE DATABASES dw_LegendVehicle;
```

```
use dw_LegendVehicle;
```

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

Copy

```
CREATE TABLE dimDate(
```

```
id_dimDate int not null AUTO_INCREMENT PRIMARY KEY
```

```
date date
```

```
year int
```

```
month int
```

```
day int
```

```
);
```

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.

Step name: Generate rows

Limit: 1825

Never stop generating rows: ☐

Interval in ms (delay): 5000

Current row time field name: now

Previous row time field name: FiveSecondsAgo

Fields:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	CurrentDate	Date	dd-MM-yyyy						01-01-2003	N

Buttons: ? Help, OK, Preview, Cancel

Gambar konfigurasi generate rows

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

Add sequence

Step name: Add sequence

Name of value: incrementDay

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection: conn_dw_destination Edit... New... Wizard...

Schema name: Schemas...

Sequence name: SEQ_ Sequences...

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value: 0

Increment by: 1

Maximum value: 999999999

Help OK Cancel

Gambar konfigurasi add sequences

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

Calculator

Step name: Calculator

☒ Throw an error on non existing files

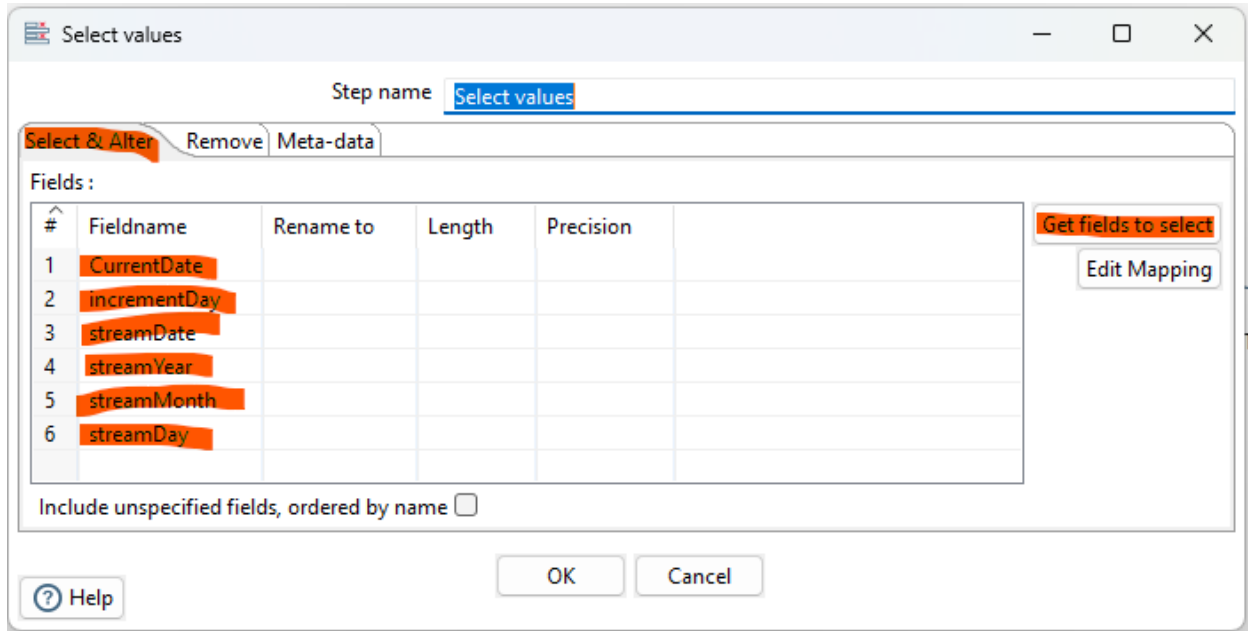
Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping symbol
1	<u>streamDate</u>	<u>Date A + B Days</u>	<u>CurrentDate</u>	<u>incrementDay</u>		None			N			
2	<u>streamYear</u>	<u>Year of date A</u>	<u>streamDate</u>			None			N			
3	<u>streamMonth</u>	<u>Month of date A</u>	<u>streamDate</u>			None			N			
4	<u>streamDay</u>	<u>Day of month of date A</u>	<u>streamDate</u>			None			N			

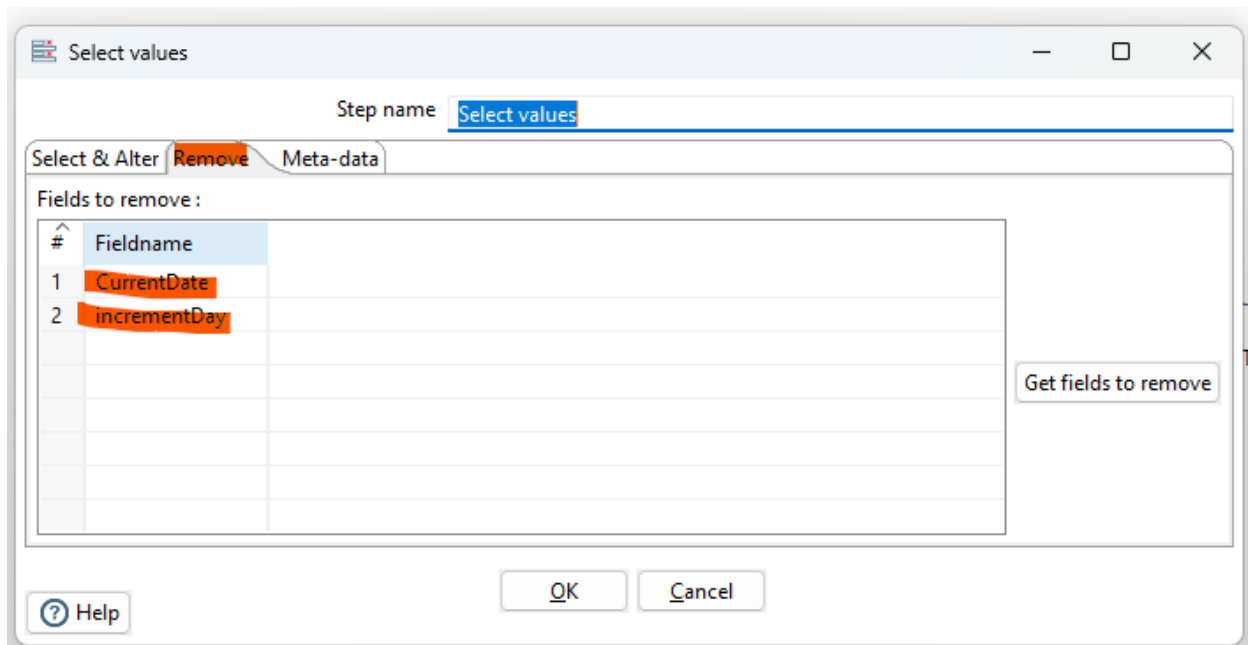
Help OK Cancel

Gambar konfigurasi kalkulator

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.

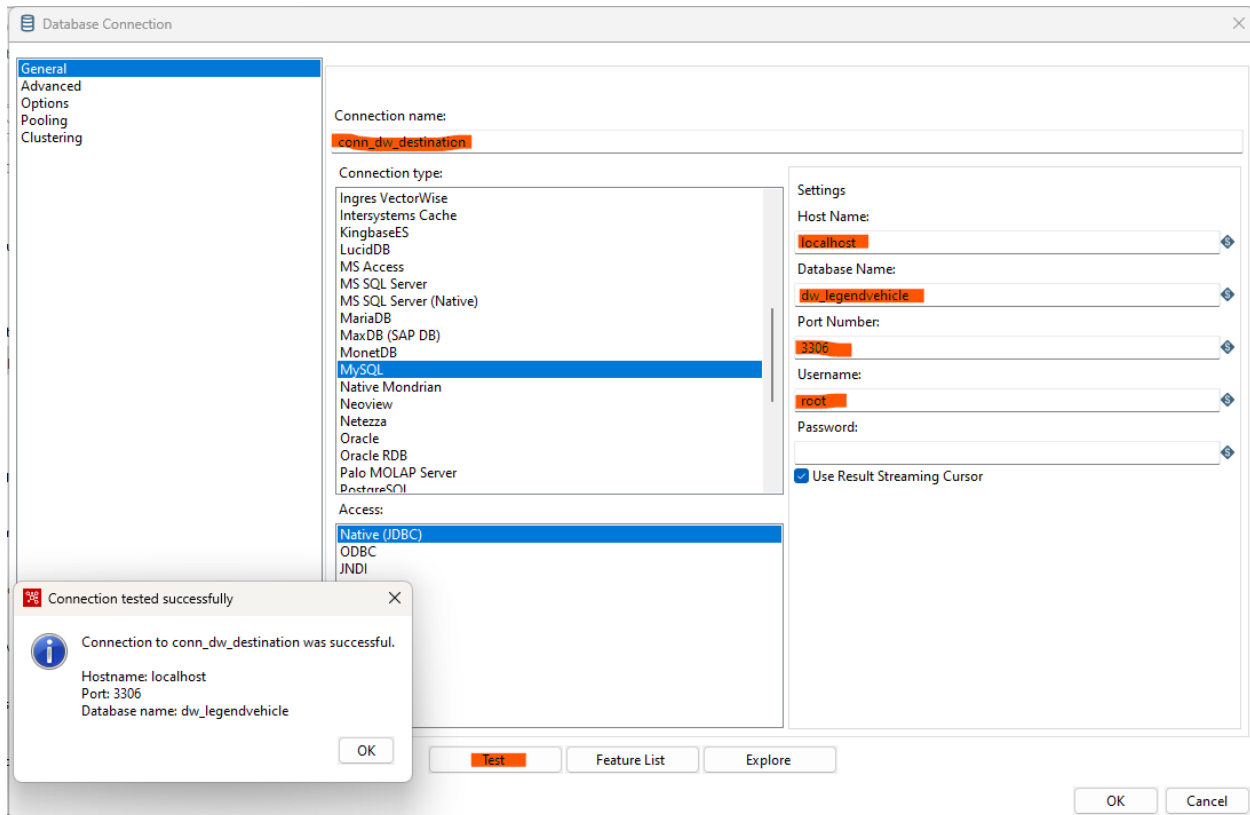


Gambar konfigurasi tab select & alter pada select values



Gambar konfigurasi tab remove pada select values

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.



Gambar konfigurasi database connection

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 retrieve adalah field yang ada pada table dimDate yaitu date, year, month, dan day.

Database lookup

Step name: Database lookup

Connection: conn_dw_destination [Edit...] [New...] [Wizard...]

Lookup schema: dw_legendvehicle [Browse...]

Lookup table: dimdate [Browse...]

Enable cache? ☐

Cache size in rows (0=cache): 0

Load all data from table ☐

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

#	Table field	Comparator	Field1	Field2
1	date	=	streamDate	
2	year	=	streamYear	
3	month	=	streamMonth	
4	day	=	streamDay	

Values to return from the lookup table :

#	Field	New name	Default	Type
1	date			None
2	year			None
3	month			None
4	day			None

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

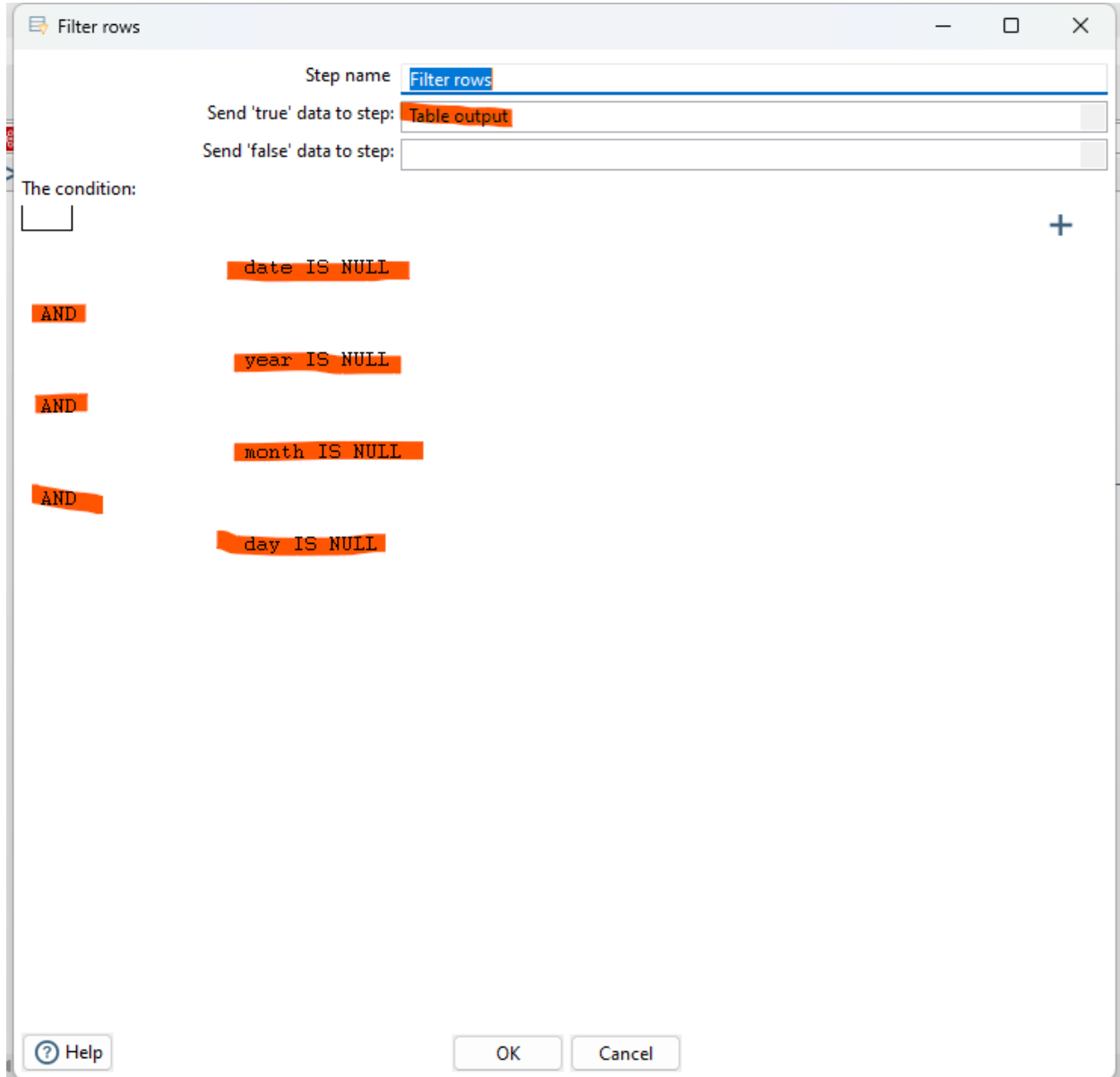
Order by:

[?] Help [OK] [Cancel] [Get Fields] [Get lookup fields]

Gambar konfigurasi database lookup

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 null)



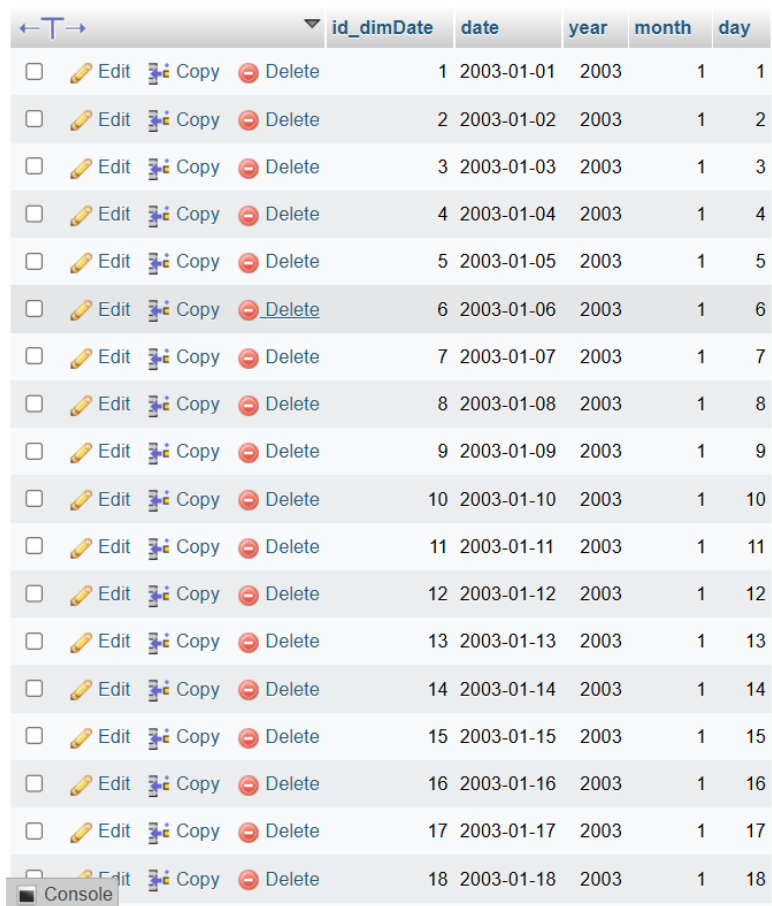
Gambar konfigurasi filter rows

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.

- [illegible]

Gambar Konfigurasi table output

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



	id_dimDate	date	year	month	day
<input type="checkbox"/> Edit Copy Delete	1	2003-01-01	2003	1	1
<input type="checkbox"/> Edit Copy Delete	2	2003-01-02	2003	1	2
<input type="checkbox"/> Edit Copy Delete	3	2003-01-03	2003	1	3
<input type="checkbox"/> Edit Copy Delete	4	2003-01-04	2003	1	4
<input type="checkbox"/> Edit Copy Delete	5	2003-01-05	2003	1	5
<input type="checkbox"/> Edit Copy Delete	6	2003-01-06	2003	1	6
<input type="checkbox"/> Edit Copy Delete	7	2003-01-07	2003	1	7
<input type="checkbox"/> Edit Copy Delete	8	2003-01-08	2003	1	8
<input type="checkbox"/> Edit Copy Delete	9	2003-01-09	2003	1	9
<input type="checkbox"/> Edit Copy Delete	10	2003-01-10	2003	1	10
<input type="checkbox"/> Edit Copy Delete	11	2003-01-11	2003	1	11
<input type="checkbox"/> Edit Copy Delete	12	2003-01-12	2003	1	12
<input type="checkbox"/> Edit Copy Delete	13	2003-01-13	2003	1	13
<input type="checkbox"/> Edit Copy Delete	14	2003-01-14	2003	1	14
<input type="checkbox"/> Edit Copy Delete	15	2003-01-15	2003	1	15
<input type="checkbox"/> Edit Copy Delete	16	2003-01-16	2003	1	16
<input type="checkbox"/> Edit Copy Delete	17	2003-01-17	2003	1	17
<input type="checkbox"/> Edit Copy Delete	18	2003-01-18	2003	1	18

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 di setiap prosesnya.

Proses Objek	SS data input	SS data output	Keterangan
Generate Rows	(tidak ada input)	1825	Membuat 1825 baris data kosong atau default
Add Squances	1825	1825	Menambahkan nomor urut atau sequence ke setiap baris
Calculator	1825	1825	Melakukan perhitungan

			berdasarkan nilai tertentu.
Select Values	1825	1825	Memilih kolom tertentu dari dataset.
Database Lookup	1825	1825	Mencari data dari database lain berdasarkan nilai kunci tertentu.
Filter Rows	1825	1825	Menyaring data berdasarkan kondisi tertentu.
Table Outputz	1825	1825 (ke database)	Menyimpan data hasil transformasi ke dalam tabel database.

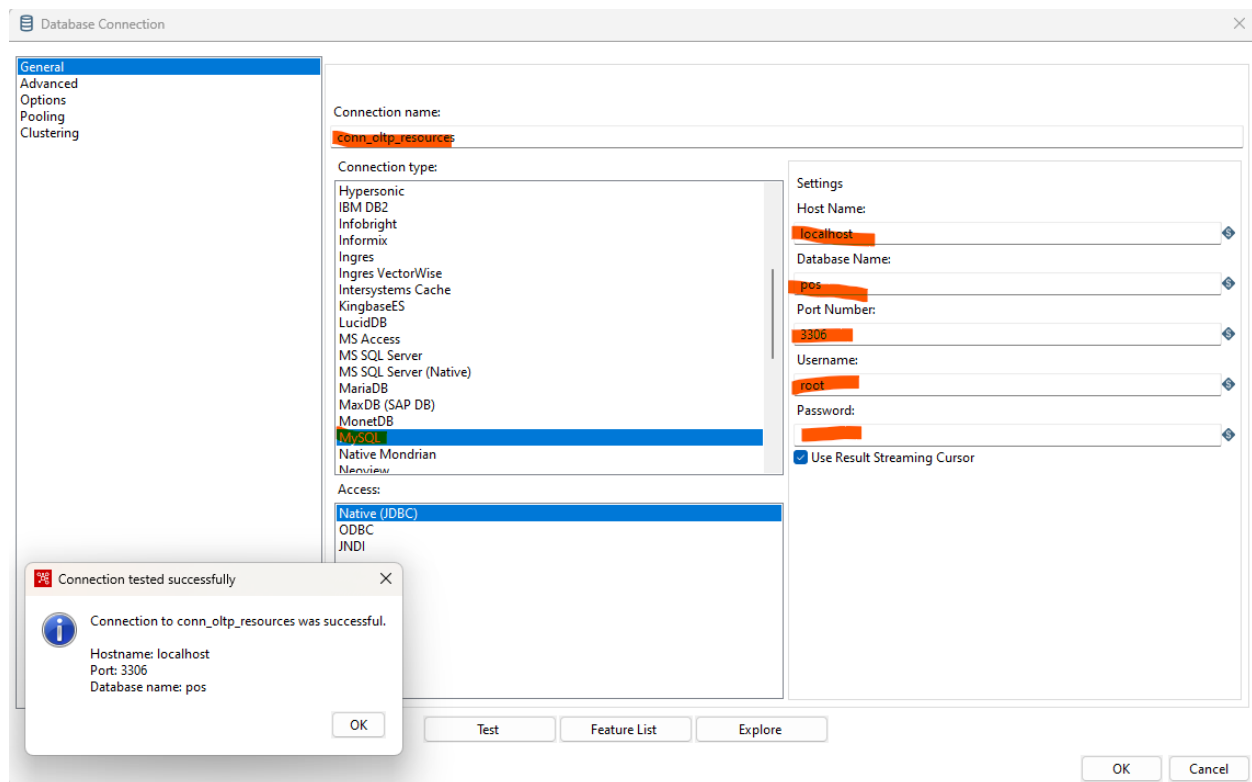
1.

Buatlah tabel **dimPegawai** pada **dw_legendVehicle**.

Copy

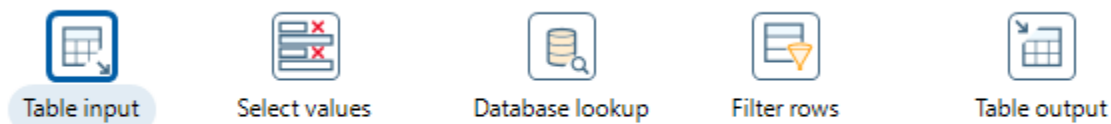
```
CREATE TABLE dimEmployees(
    id_dimEmployees int not_null auto increment primary,
    employeeNumber int(11),
    firstName varchar(50),
    lastname varchar(50),
    jobtitle varchar(50),
    boss_firstname varchar(50),
    boss_lastname varchar(50),
    updated date DEFAULT CURRENT_DATE
);
```

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.



Gambar konfigurasi koneksi terhadap OLTP

1. Drag and drop beberapa objek sebagai berikut:
 - **Table input:** digunakan mengambil data dari database OLTP.
 - **Select values:** memilih 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 table dimEmployees



1. Konfigurasi pada table input dengan menghubungkan **Connection** pada koneksi **conn_oltp_resources**. Untuk mengambil data sumber menggunakan query dibawah ini.

Copy

```
select * from employees e
```

```
left join employees r
```

```
on e.reportsTo=r.employeeNumber;
```

Table input

Step name: Table input

Connection: conn_oltp_resources

Buttons: Edit..., New..., Wizard..., Get SQL select statement...

SQL:

```
select * from employees e
left join employees r
on e.reportsTo=r.employeeNumber;
```

Line 3 Column 0

Store column info in step meta ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step:

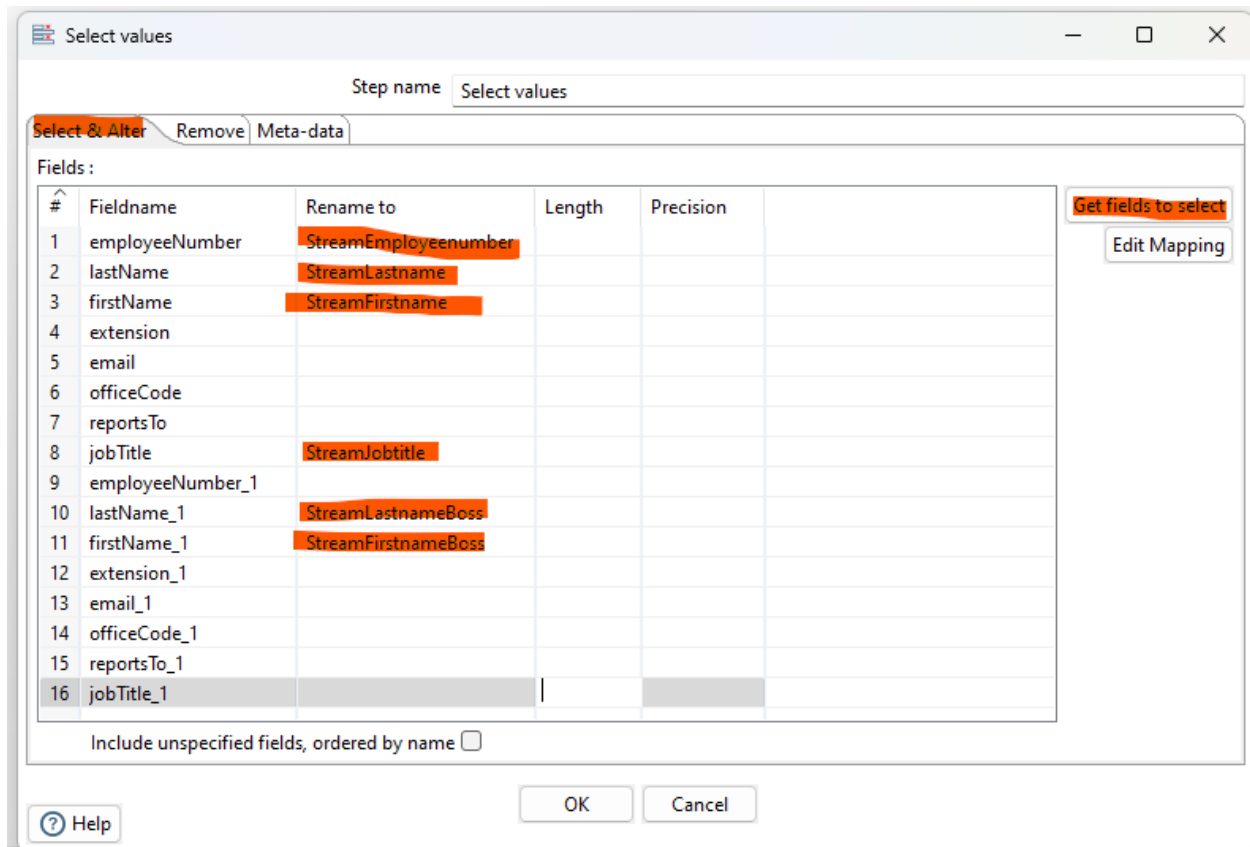
Execute for each row? ☐

Limit size: 0

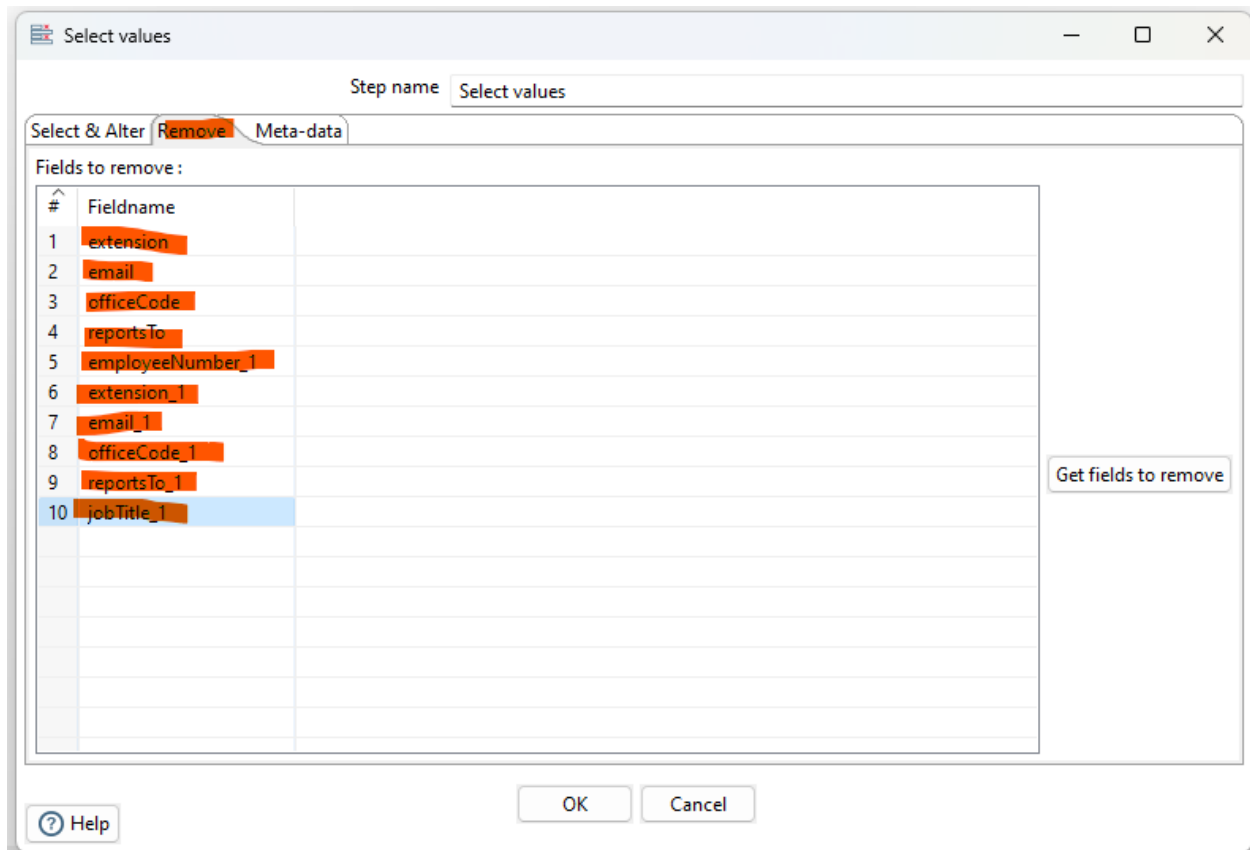
Buttons: Help, OK, Preview, Cancel

Gambar konfigurasi table input

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**.



Gambar konfigurasi select values pada tab select & alter



Gambar konfigurasi select values 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.

Database lookup

Step name: Database lookup

Connection: conn_dw_destination [Edit... New... Wizard...]

Lookup schema: dw_legendvehicle [Browse...]

Lookup table: dimemployees [Browse...]

Enable cache? ☐

Cache size in rows (0=cache): 0

Load all data from table ☐

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

#	Table field	Comparator	Field1	Field2
1	employeenumber	=	StreamEmployeeNumber	
2	firstName	=	StreamFirstname	
3	lastName	=	StreamLastname	
4	jobtitle	=	StreamJobtitle	
5	boss_firstname	=	StreamFirstnameBoss	
6	boss_lastname	=	StreamLastnameBoss	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	employeeNumber	employeeNumber		Integer
2	firstName	firstName		String
3	lastName	lastName		String
4	jobtitle	jobtitle		String
5	boss_firstname	boss_firstname		String
6	boss_lastname	boss_lastname		String
7				

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

Order by:

[?] Help OK Cancel Get Fields Get lookup fields

Gambar konfigurasi database lookup

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**.

Filter rows

Step name: Filter rows

Send 'true' data to step: Table output

Send 'false' data to step:

The condition:

☐ To edit a subcondition, simply click on it +

employeeNumber IS NULL

AND

firstName IS NULL

AND

lastname IS NULL

AND

jobtitle IS NULL

AND

boss_firstname IS NULL

AND

boss_lastname IS NULL

? Help OK Cancel

Gambar konfigurasi filter rows

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 database fields**, dan mapping **data stream input** dari oltp terhadap field yang ada pada **dimemployees**.

Table output

Step name: Table output

Connection: conn_dw_destination [Edit...] [New...] [Wizard...]

Target schema: dw_legendvehicle [Browse...]

Target table: dimemployees [Browse...]

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☒

Main options | **Database fields**

Fields to insert:

#	Table field	Stream field
1	employeeNumber	StreamEmployeeNumber
2	firstName	StreamFirstname
3	lastName	StreamLastname
4	jobtitle	StreamJobtitle
5	boss_firstname	StreamFirstnameBoss
6	boss_lastname	StreamLastnameBoss
7		

[Get fields]

[Enter field mapping]

[?] Help [OK] [Cancel] [SQL]

Gambar konfigurasi table output

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

id_dimEmployees	employeeNumber	firstName	lastname	jobtitle	boss_firstname	boss_lastname	updated
1	1002	Diane	Murphy	President	NULL	NULL	2024-03-18
2	1056	Mary	Patterson	VP Sales	Diane	Murphy	2024-03-18
3	1076	Jeff	Firrelli	VP Marketing	Diane	Murphy	2024-03-18
4	1088	William	Patterson	Sales Manager (APAC)	Mary	Patterson	2024-03-18
5	1102	Gerard	Bondur	Sale Manager (EMEA)	Mary	Patterson	2024-03-18
6	1143	Anthony	Bow	Sales Manager (NA)	Mary	Patterson	2024-03-18
7	1165	Leslie	Jennings	Sales Rep	Anthony	Bow	2024-03-18
8	1166	Leslie	Thompson	Sales Rep	Anthony	Bow	2024-03-18
9	1188	Julie	Firrelli	Sales Rep	Anthony	Bow	2024-03-18
10	1216	Steve	Patterson	Sales Rep	Anthony	Bow	2024-03-18

Gambar isi data dimemployees

TUGAS 2

Proses Objek	SS Data Input	SS Data Output	Keterangan
Table Input	Data awal dari sumber (OLTP)	Data mentah masuk ke proses ETL	Mengambil data dari datab operasional
Select Values	Data awal	Data dengan kolom yang dipilih	Memilih hanya kolom yang dibutuhkan
Database Lookup	Data dari tahap sebelumnya	Data dengan informasi tambahan dari lookup table	Menambahkan data dari ta berdasarkan kunci tertentu
Filter Rows	Data dengan lookup	Data yang sudah difilter sesuai kriteria	Menghapus data yang tidak memenuhi syarat
Table Output	Data hasil transformasi	Data akhir yang siap disimpan ke OLAP atau data warehouse	Menyimpan hasil transform tujuan akhir