

Nama : Adinda Ivanka Maysanda Putri

Kelas : SIB 2B

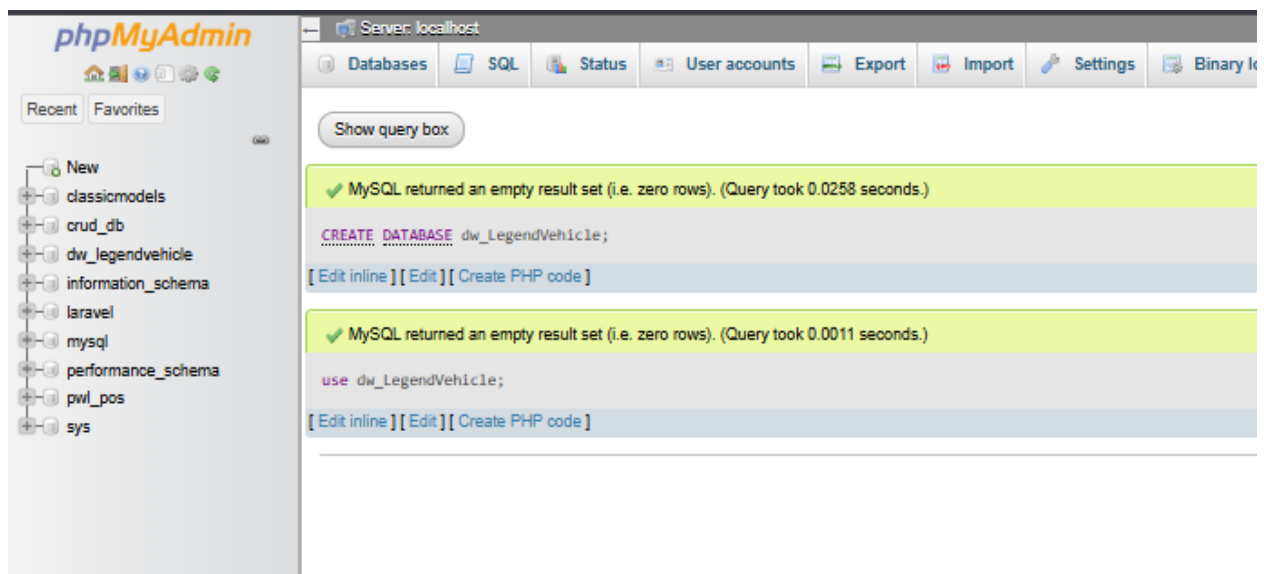
NIM : 2341760058

JOBSHEET 3

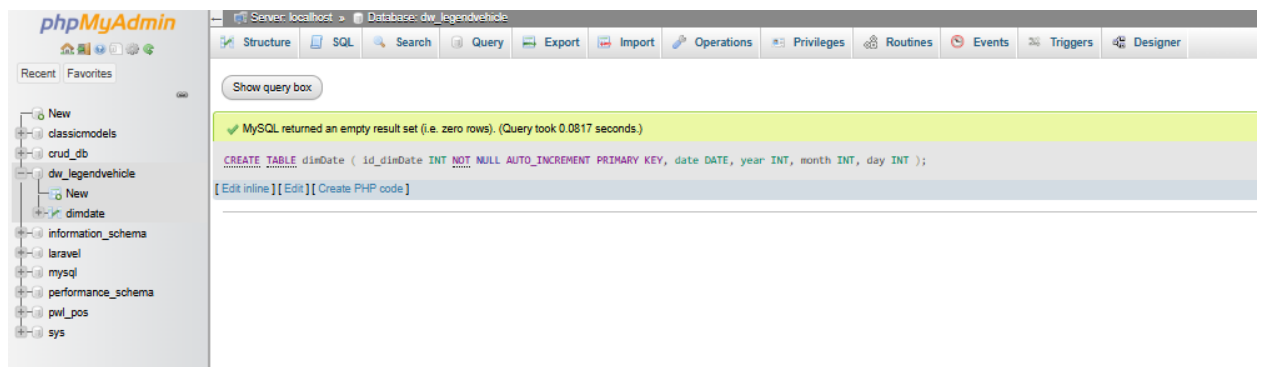
DATABASE ANALYTICAL

A. DIMENSI WAKTU

1. Buatlah sebuah database yang digunakan sebagai OLAP dengan nama dw_LegendVehicle.



2. 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.

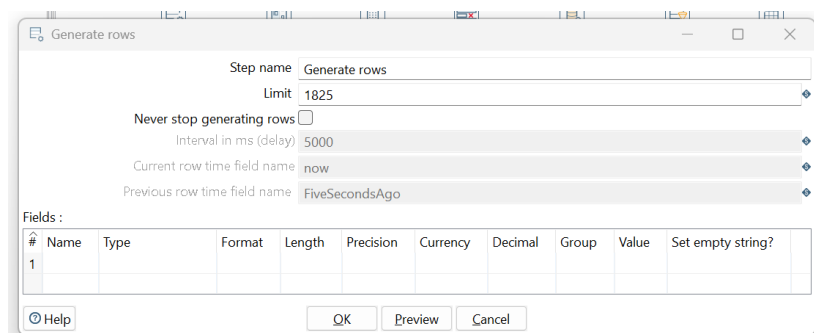
2. Drag and Drop beberapa objek yaitu:

- 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).



3. 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).

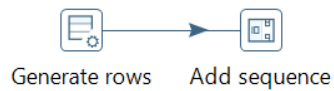
⇒ Configure



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

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

5. Hubungkan output dari Generate Rows menuju Add Sequence.



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

7. Hubungkan output dari add sequences menuju calculator.



8. 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

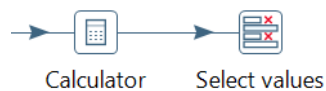
☒ Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping
1	streamDate	Date A + 8 Days	CurrentDate	incrementDay		None			N			
2	streamYear	Year of date A	streamDate			None			N			
3	streamMonth	Month of date A	streamDate			None			N			
4	streamDay	Day of month of date A	streamDate			None			N			

Help OK Cancel

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.

Select values

Step name Select values

Select & Alter Remove Meta-data

Fields :

#	Fieldname	Rename to	Length	Precision
1	CurrentDate			
2	incrementDay			
3	streamDate			
4	streamYear			
5	streamMonth			
6	streamDay			

Get fields to select
Edit Mapping

Include unspecified fields, ordered by name ☐

Help OK Cancel

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

Select values

Step name Select values

Select & Alter Remove Meta-data

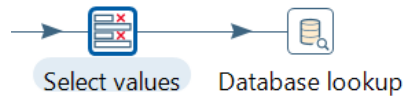
Fields to remove :

#	Fieldname
1	CurrentDate
2	incrementDay

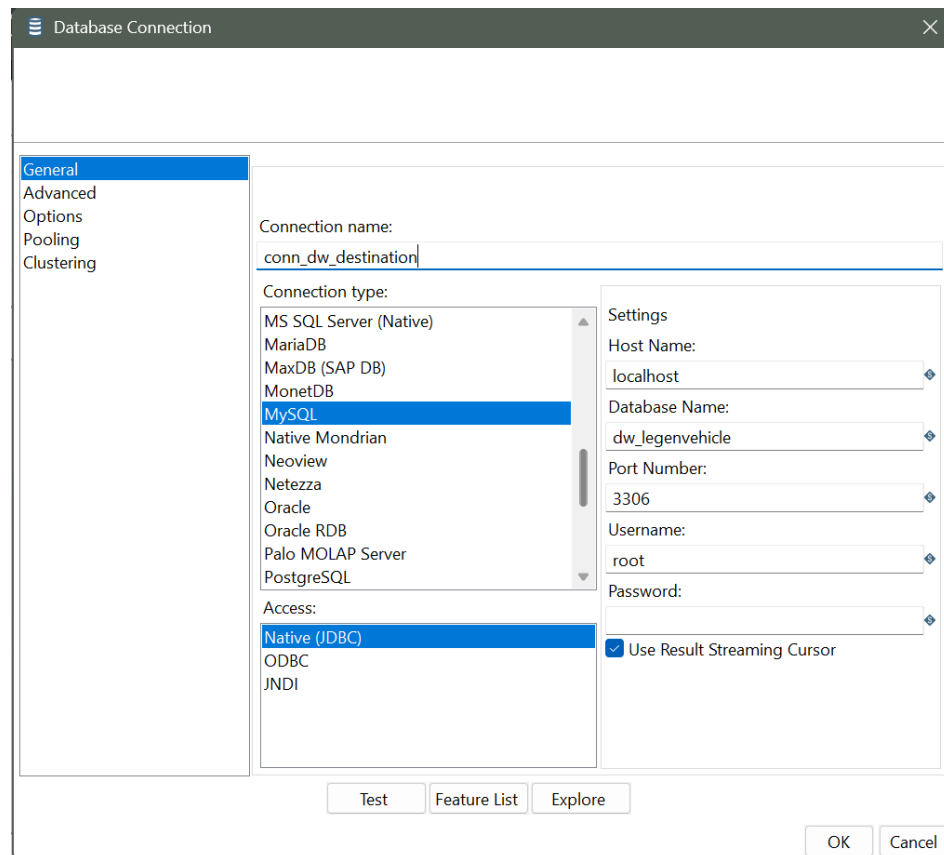
Get fields to remove

Help OK Cancel

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 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_legenvehicle [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 ☐

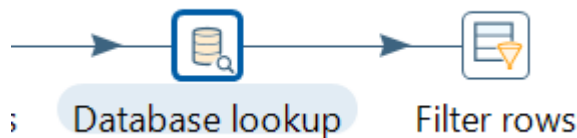
Fail on multiple results? ☐

Order by: [Dropdown]

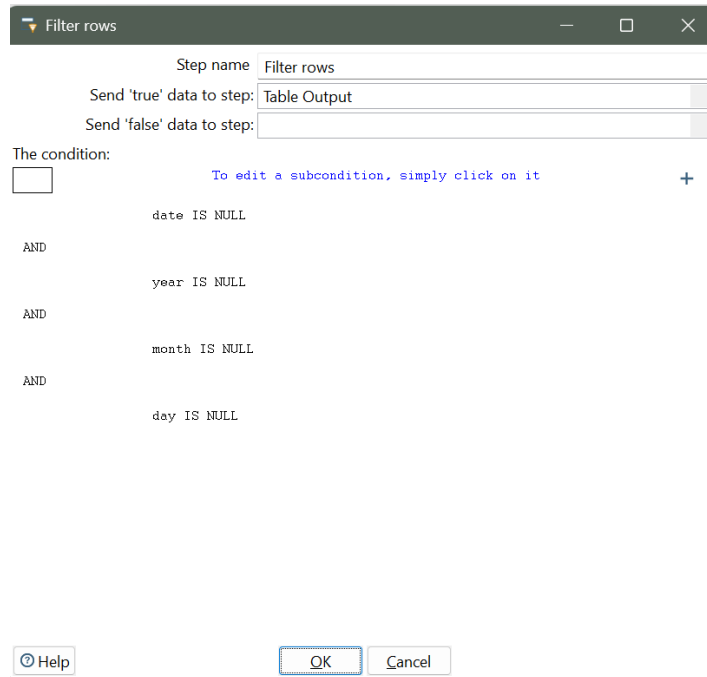
[Help] [OK] [Cancel] [Get Fields] [Get lookup fields]

17. Hubungkan output dari database lookup dengan filter rows

⇒ Configure

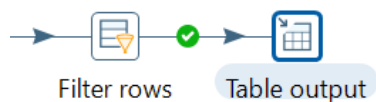


18. Konfigurasi pada filter rows adalah dengan melakukan konfigurasi output true data pada table output. Pada bagian ini data yang tidak memiliki kesamaan pada tahapan sebelumnya akan dicek dimana jika fields Stream tidak memiliki kesamaan dengan field dimDate, maka field dimDate tersebut akan bernilai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



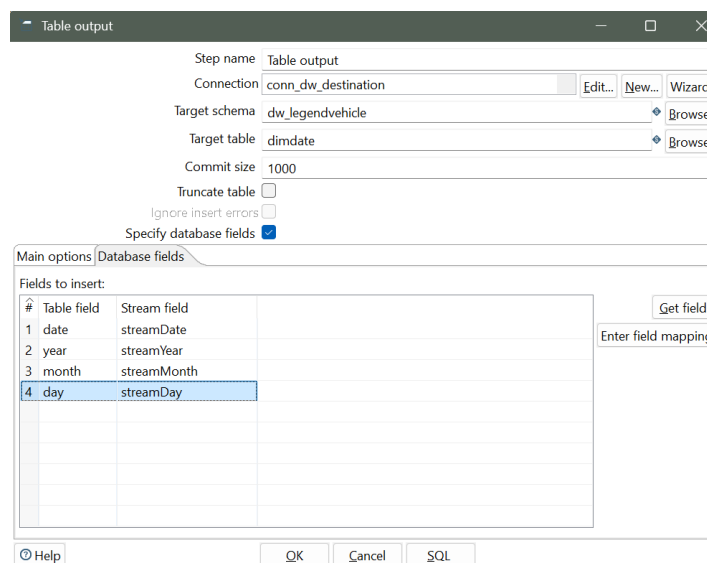
19. Hubungkan output dari filter rows menuju table output.

⇒ Configure



20. Konfigurasi pada table output adalah memberikan koneksi pada conn_dw_destination dengan schema dw_legendvehicle dan table dimdate.

⇒ Configure

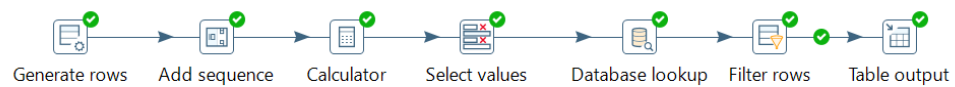


21. Aktifkan specify database fields.

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
1	2003-01-01	2003	1	1
2	2003-01-02	2003	1	2
3	2003-01-03	2003	1	3
4	2003-01-04	2003	1	4
5	2003-01-05	2003	1	5
6	2003-01-06	2003	1	6
7	2003-01-07	2003	1	7
8	2003-01-08	2003	1	8

⇒ Pentaho step

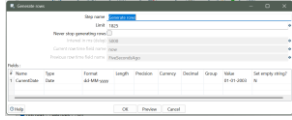
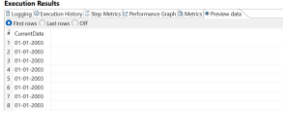
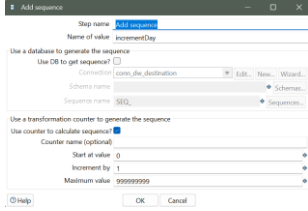
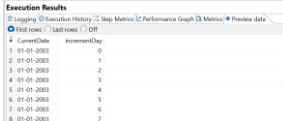
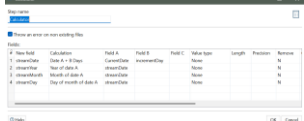
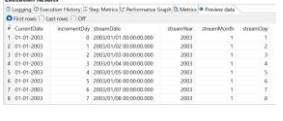


Database

date	year	month	day
2003-01-01	2003	1	1
2003-01-02	2003	1	2
2003-01-03	2003	1	3
2003-01-04	2003	1	4
2003-01-05	2003	1	5
2003-01-06	2003	1	6
2003-01-07	2003	1	7
2003-01-08	2003	1	8
2003-01-09	2003	1	9
2003-01-10	2003	1	10
2003-01-11	2003	1	11
2003-01-12	2003	1	12
2003-01-13	2003	1	13
2003-01-14	2003	1	14
2003-01-15	2003	1	15
2003-01-16	2003	1	16

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	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			

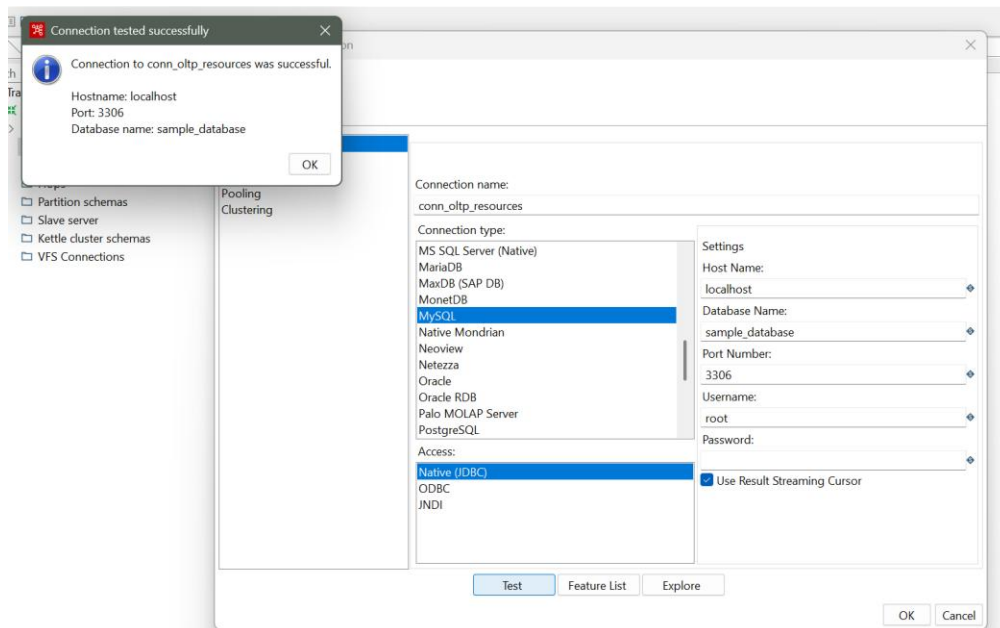
B. DIMENSI PEGAWAI

1. Buatlah tabel dimPegawai pada dw_legendVehicle.

✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0360 seconds.)

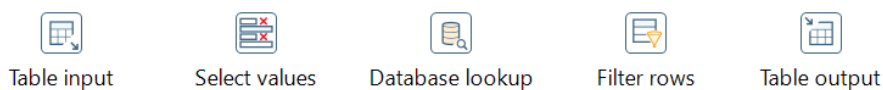
```
CREATE TABLE dimEmployees ( id_dimEmployees INT NOT NULL AUTO_INCREMENT, employeeNumber INT, firstName VARCHAR(50), lastName VARCHAR(50), jobTitle VARCHAR(50), boss_firstName VARCHAR(50), boss_lastName VARCHAR(50), updated TIMESTAMPT DEFAULT CURRENT_TIMESTAMP, PRIMARY KEY (id_dimEmployees) );
```

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: 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 tavle dimEmployees



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

Table input

Step name: Table input

Connection: conn_oltp_resources

SQL

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

Line 3 Column 30

Store column info in step meta ☐

Enable lazy conversion ☐

Replace variables in script? ☐

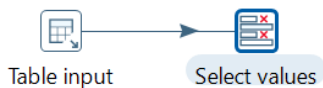
Insert data from step ☐

Execute for each row? ☐

Limit size: 0

Help OK Preview Cancel

5. Hubungkan output table input pada select values.



6. Konfigurasi pada Select values yaitu mengambil data dari field employeeenumber, 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 :

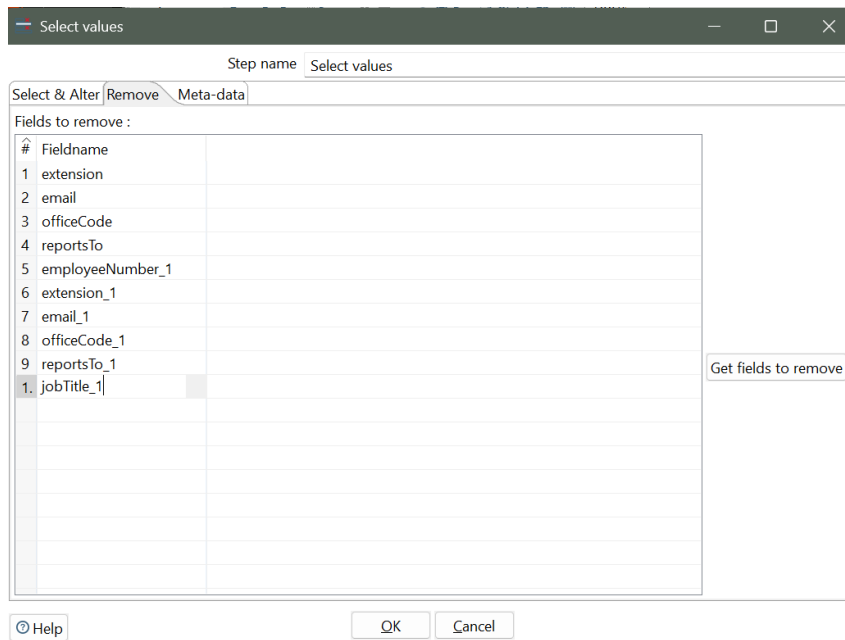
#	Fieldname	Rename to	Length	Precision
1	employeeNumber	StreamEmployeeenumber		
2	lastName	StreamLastname		
3	firstName	StreamFirstname		
4	extension			
5	email			
6	officeCode			
7	reportsTo			
8	jobTitle	StreamJobtitle		
9	employeeNumber_1			
10	lastName_1	StreamLastnameBoss		
11	firstName_1	StreamFirstnameBoss		
12	extension_1			
13	email_1			
14	officeCode_1			
15	reportsTo_1			
16	jobTitle_1			

Get fields to select

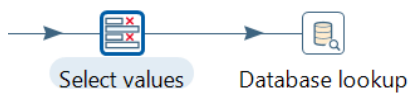
Edit Mapping

Include unspecified fields, ordered by name ☐

Help OK Cancel

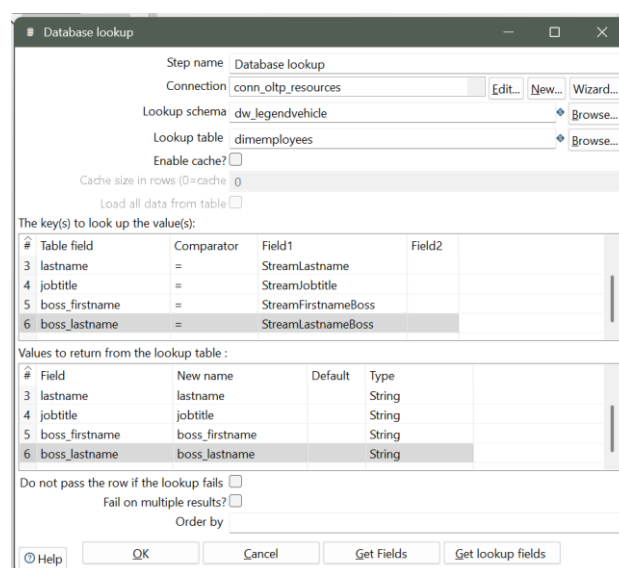


8. Hubungkan output select values pada database lookup

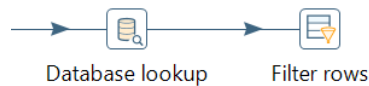


9. Konfigurasi pada database lookup adalah dengan menghubungkan koneksi pada conn_dw_destination dengan table lookup dimEmployees yang telah dibuat pada tahap pertama.

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 database 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.

The screenshot shows a window titled 'Filter rows'. It has a 'Step name' field with 'Filter rows' entered. Below it, 'Send 'true' data to step:' is set to 'Table output', and 'Send 'false' data to step:' is empty. Under 'The condition:', there is a list of conditions separated by 'AND' operators: 'employeeNumber IS NULL', 'firstName IS NULL', 'lastName IS NULL', 'jobtitle IS NULL', 'boss_firstname IS NULL', and 'boss_lastname IS NULL'. At the bottom are 'Help', 'OK', and 'Cancel' buttons.

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 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

[Get fields] [Enter field mapping]

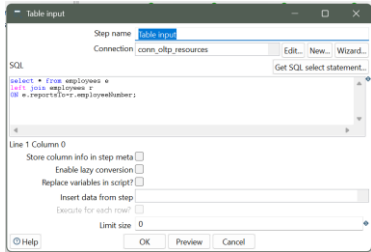

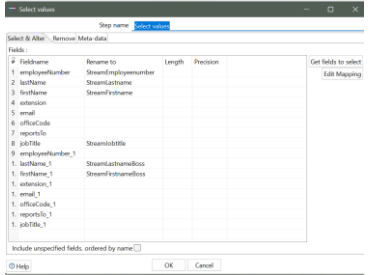

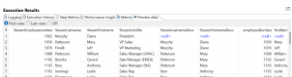
[Help] [OK] [Cancel] [SQL]

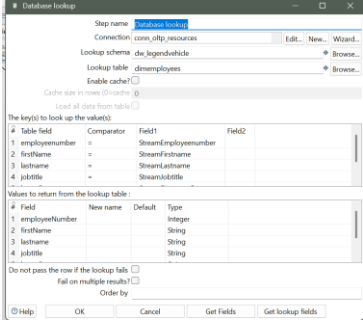
16. jika proses keseluruhan berhasil maka tabel dimemployees akan terisi data pegawai dari database OLTP.

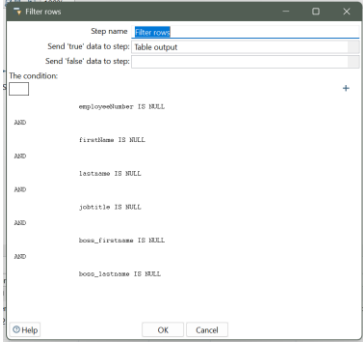

▼ id_dimEmployees	employeeNumber	firstName	lastName	jobTitle	boss_firstname	boss_lastname	updated
delete	1	1002 Diane	Murphy	President	NULL	NULL	2025-03-10 11:21:48
delete	2	1056 Mary	Patterson	VP Sales	Diane	Murphy	2025-03-10 11:21:48
delete	3	1076 Jeff	Firrelli	VP Marketing	Diane	Murphy	2025-03-10 11:21:48
delete	4	1088 William	Patterson	Sales Manager (APAC)	Mary	Patterson	2025-03-10 11:21:48
delete	5	1102 Gerard	Bondur	Sale Manager (EMEA)	Mary	Patterson	2025-03-10 11:21:48
delete	6	1143 Anthony	Bow	Sales Manager (NA)	Mary	Patterson	2025-03-10 11:21:48
delete	7	1165 Leslie	Jennings	Sales Rep	Anthony	Bow	2025-03-10 11:21:48
delete	8	1166 Leslie	Thompson	Sales Rep	Anthony	Bow	2025-03-10 11:21:48
delete	9	1188 Julie	Firrelli	Sales Rep	Anthony	Bow	2025-03-10 11:21:48
delete	10	1216 Steve	Patterson	Sales Rep	Anthony	Bow	2025-03-10 11:21:48
delete	11	1286 Foon Yue	Tseng	Sales Rep	Anthony	Bow	2025-03-10 11:21:48
delete	12	1323 George	Vanauf	Sales Rep	Anthony	Bow	2025-03-10 11:21:48
delete	13	1337 Loui	Bondur	Sales Rep	Gerard	Bondur	2025-03-10 11:21:48

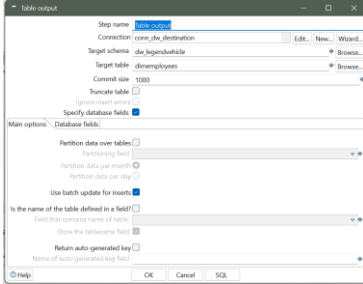

TUGAS 2

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 di setiap prosesnya.

Proses Object	SS data input	SS data output	Keterangan
Table input			<p>Pertama konfigurasi Table Input dengan LEFT JOIN pada tabel employees. Kemudian hasil querynya, yaitu daftar karyawan dan atasan mereka.</p>
Select Values			<p>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.</p>
Database Lookup			<p>konfigurasi database Lookup untuk mencocokkan data karyawan dengan</p>

		<p>tabel referensi dimemployees.</p> <p>kemudian hasil eksekusi menunjukkan data karyawan dengan informasi tambahan dari lookup table.</p>
--	---	--

<p>Filter Rows</p>		 <p>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.</p>
--------------------	--	--

<p>Table Output</p>		 <p>Pertama langkah Table Output yang digunakan untuk menyimpan data hasil transformasi ke dalam tabel dimemployees di skema dw_legendvehicle. kemudian hasil</p>
---------------------	---	---

			eksekusi, menampilkan data yang akan disimpan, termasuk kolom StreamEmployeeN umber, StreamLastname, StreamFirstname, StreamJobtitle, dan informasi tambahan lainnya.
--	--	--	--

2. 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.

3. 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.