

**DATA WAREHOUSE**

**JOBSHEET 3: DATABASE ANALYTICAL**

**(DIMENSI WAKTU & DIMENSI PEGAWAI)**



Disusun Oleh:

KARINA IKA INDASA (2341760042)

KELAS 2A-SIB/16

**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 database OLAP.

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

### A. Dimensi Waktu

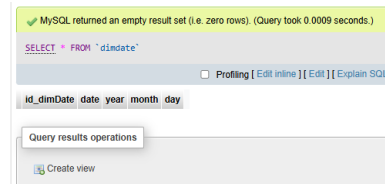
1. Buatlah sebuah database yang digunakan sebagai **OLAP** dengan nama **dw\_LegendVehicle**.

```
1 CREATE DATABASE dw_LegendVehicle;  
2 use dw_LegendVehicle;
```

☐ dw\_legendvehicle ut-f8mb4\_0900\_ai\_ci ☒ Replicated [Check privileges](#)

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

```
1 CREATE TABLE dimDate (  
2   id_dimDate INT NOT NULL AUTO_INCREMENT PRIMARY KEY,  
3   date DATE,  
4   year INT,  
5   month INT,  
6   day INT  
7 );
```



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).



- 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 ).
- 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

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

Step name: Add sequence

Name of value: incrementDay

Use a database to generate the sequence? ☐

Connection: (empty) Edit... New... Wizard...

Schema name: (empty) Schemas...

Sequence name: SEQ\_ Sequences...

Use a transformation counter to generate the sequence? ☒

Counter name (optional): (empty)

Start at value: 0

Increment by: 1

Maximum value: 999999999

Buttons: Help, OK, Cancel

- Hubungkan output dari **add sequences** menuju **calculator**.
- 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**

Step name: Calculator

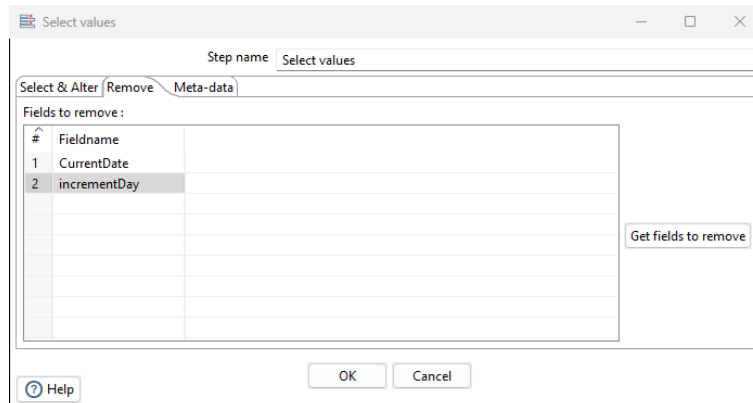
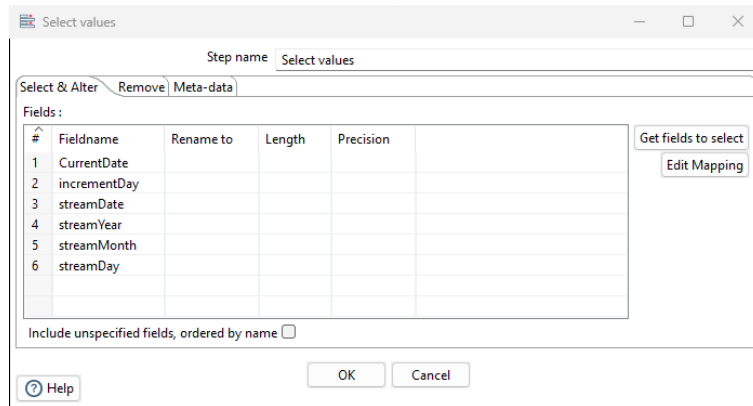
Throw an error on non existing files: ☒

Fields:

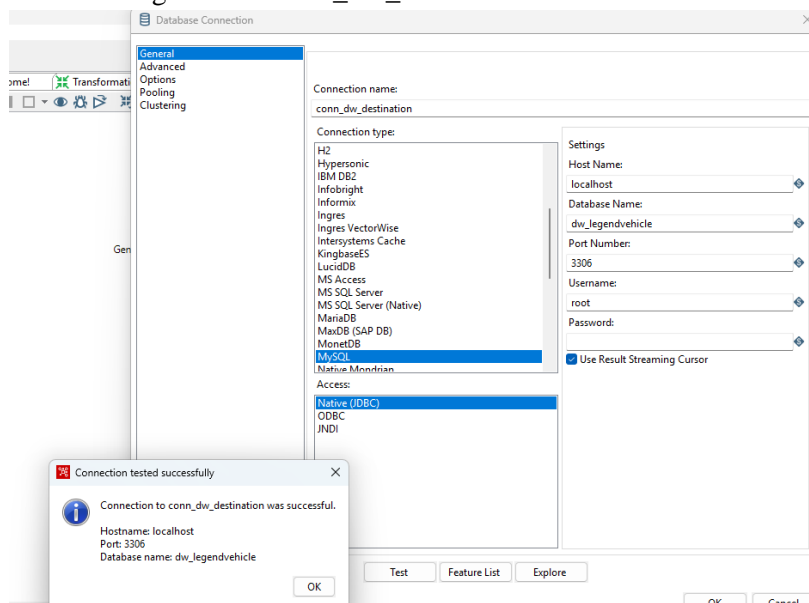
#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion
1	streamDate	Date A + B D...	Current...	increm...		None			N	
2	streamYear	Year of date A	stream...			None			N	
3	streamMonth	Month of da...	stream...			None			N	
4	streamDay	Day of mont...	stream...			None			N	

Buttons: 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.
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**.

Step name: Database lookup

Connection: conn\_dw\_destination

Lookup schema: dw\_legendvehicle

Lookup table: dimDate

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

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**)

Step name: Filter rows

Send 'true' data to step: Table output

Send 'false' data to step:

The condition:

date IS NULL

AND

year IS NULL

AND

month IS NULL

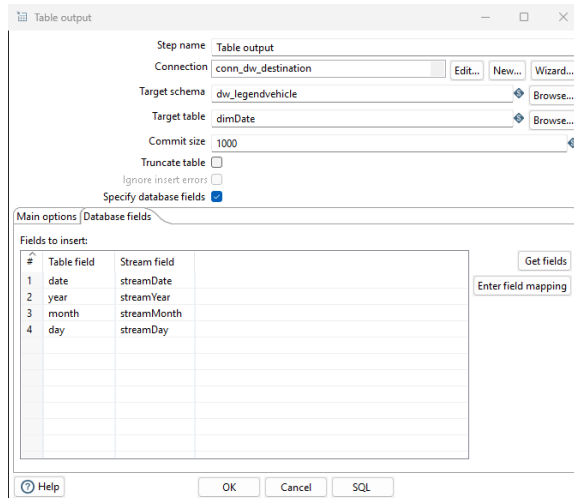
AND

day IS NULL

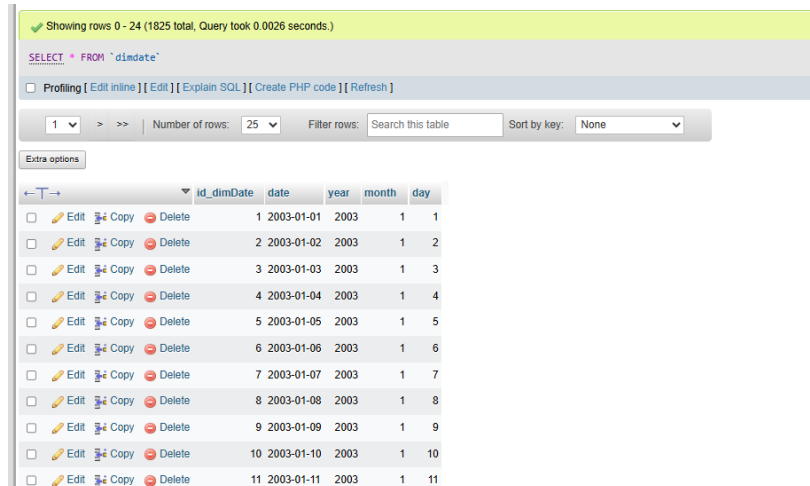
Help OK Cancel

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.

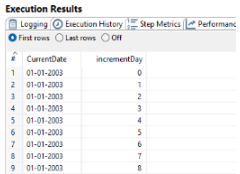
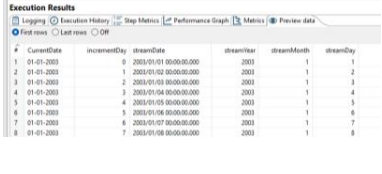
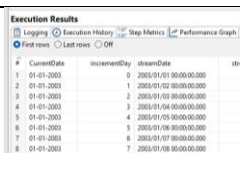
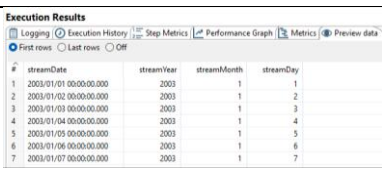
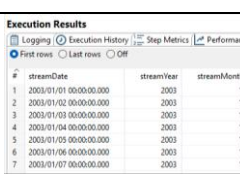
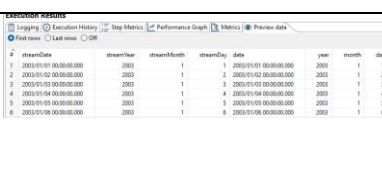
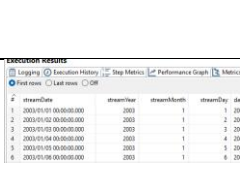
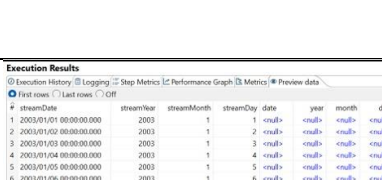
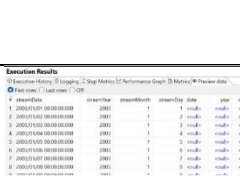
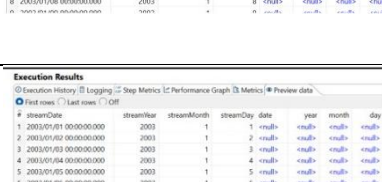


## ➤ Tugas 1

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.

**Jawab:**

Proses Objek	SS Data Input	SS Data Output	Keterangan
Generate Rows	- Tidak ada input		- Membuat data dummy dengan jumlah baris tertentu sebagai input untuk proses berikutnya.
Add Sequences			- Menambahkan nomor urut (sequence) pada setiap baris data.

Calculator			- Melakukan operasi matematika seperti penjumlahan, pengurangan, perkalian, dll.
Select Values			- Memilih hanya kolom yang diperlukan untuk proses selanjutnya, bisa juga mengganti nama kolom.
Database Lookup			- Mencocokkan data dengan tabel di database dan menambahkan informasi tambahan (seperti ID, nama, dll.).
Filter Rows			- Menyaring data berdasarkan kondisi tertentu, misalnya hanya mengambil data dengan tahun tertentu.
Table Output			- Menyimpan data hasil transformasi ke dalam database.

## B. Dimensi Pegawai

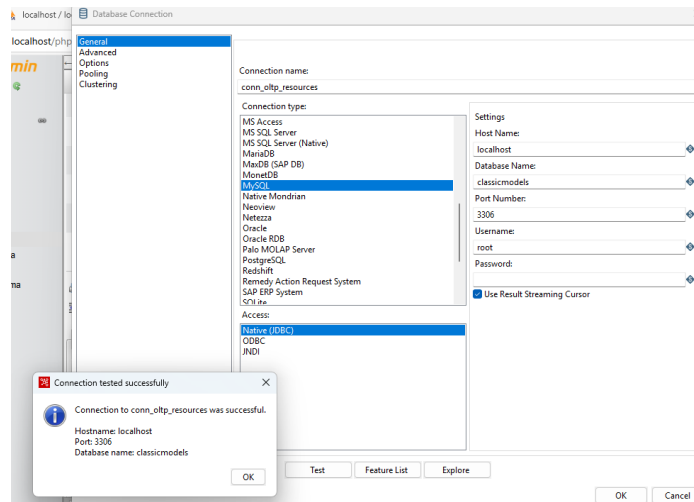
1. Buatlah tabel **dimPegawai** pada **dw\_legendVehicle**.

```

1 CREATE TABLE dimEmployees (
2     id_dimEmployees INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
3     employeeNumber INT,
4     firstName VARCHAR(50),
5     lastName VARCHAR(50),
6     jobTitle VARCHAR(50),
7     boss_firstName VARCHAR(50),
8     boss_lastName VARCHAR(50),
9     updated DATE DEFAULT (CURRENT_DATE())
10 );

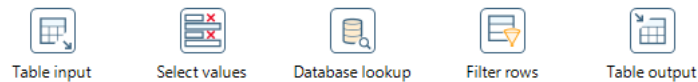
```

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.

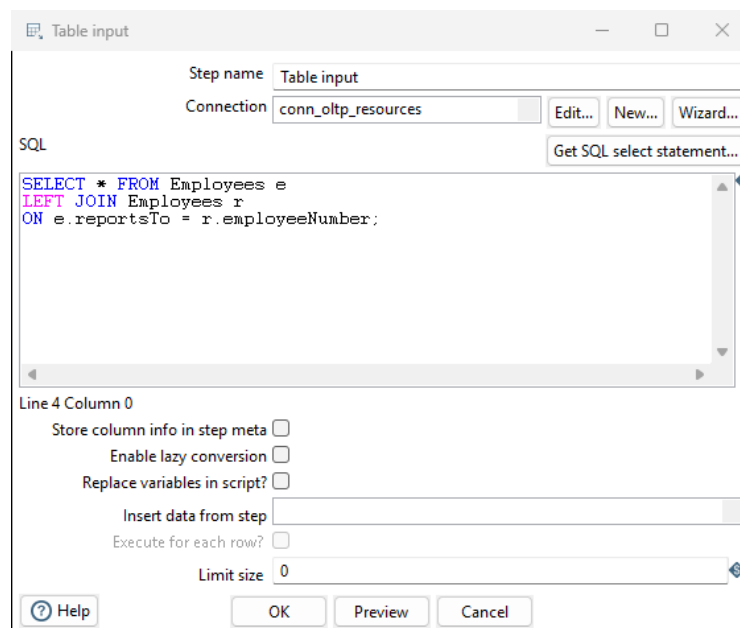


3. Drag and drop beberapa objek sebagai berikut:

- **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.



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**.

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			

Get fields to select  
Edit Mapping

Include unspecified fields, ordered by name ☐

Help OK Cancel

Step name: Select values

Select & Alter Remove Meta-data

Fields to remove :

#	Fieldname
1	extension
2	email
3	officeCode
4	reportsTo
5	employeeNumber_1
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.

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			Integer
2	firstName			String
3	lastName			String
4	jobTitle			String
5	boss_firstname			String
6	boss_lastname			String

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

Order by:

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

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**.

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

[Help] [OK] [Cancel]

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

Target schema: dw\_legendvehicle

Target table: dimEmployees

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.

Showing rows 0 - 22 (23 total, Query took 0.0037 seconds.)

SELECT \* FROM `dimemployees`

Profiling [ Edit inline ] [ Edit ] [ Explain SQL ] [ Create PHP code ] [ Refresh ]

Show all | Number of rows: 25 | Filter rows: Search this table | Sort by key: None

Extra options

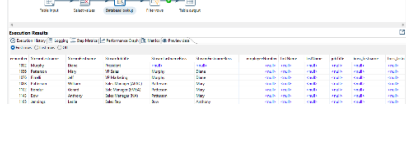
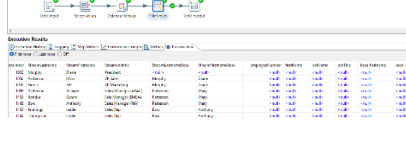
	id_dimEmployees	employeeNumber	firstName	lastName	jobTitle	boss_firstname	boss_lastname	updated
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	1	1002	Diane	Murphy	President	NULL	NULL	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	2	1056	Mary	Patterson	VP Sales	Diane	Murphy	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	3	1076	Jeff	Firrelli	VP Marketing	Diane	Murphy	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	4	1088	William	Patterson	Sales Manager (APAC)	Mary	Patterson	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	5	1102	Gerard	Bondur	Sale Manager (EMEA)	Mary	Patterson	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	6	1143	Anthony	Bow	Sales Manager (NA)	Mary	Patterson	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	7	1165	Leslie	Jennings	Sales Rep	Anthony	Bow	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	8	1166	Leslie	Thompson	Sales Rep	Anthony	Bow	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	9	1188	Julie	Firrelli	Sales Rep	Anthony	Bow	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	10	1216	Steve	Patterson	Sales Rep	Anthony	Bow	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	11	1286	Foon Yue	Tseng	Sales Rep	Anthony	Bow	2025-03-11
<input type="checkbox"/> Edit <input type="checkbox"/> Copy <input type="checkbox"/> Delete	12	1323	George	Vanauf	Sales Rep	Anthony	Bow	2025-03-11

## ➤ 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.




**Jawab:**

Proses Objek	SS Data Input	SS Data Output	Keterangan
Table Input	<p>Table Input</p> <p>Step name: Table input</p> <p>Connection: conn_oltp_resources</p> <p>SQL</p> <pre>SELECT * FROM Employees e LEFT JOIN Employees r ON e.reportsTo = r.employeeNumber;</pre>		Mengambil data mentah dari database sebagai sumber utama.

Select Values			Memilih, mengganti nama, atau mengubah tipe data kolom.
Database Lookup			Menambahkan informasi dari tabel lain menggunakan key tertentu.
Filter Rows			Menyaring data berdasarkan syarat tertentu (misalnya hanya data dengan officeCode = 1).
Table Output			Memasukkan data hasil transformasi ke dalam tabel database yang telah ditentukan.

2. Jika proses itu di ulangi ( di run kembali ) apakah data akan redundant?

**Jawab:** ada, Jika proses dijalankan ulang, data menjadi redundant karena tidak ada mekanisme untuk mencegah duplikasi, sehingga data yang sama tersimpan lebih dari satu kali di tabel

			24	1002 Diane	Murphy	President	NULL	NULL	2025-03-11
---	---	---	----	------------	--------	-----------	------	------	------------






























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

**Jawab:**

```

1 INSERT INTO employees (
2     employeeNumber, firstName, lastName, extension, email, officeCode, reportsTo, jobTitle
3 ) VALUES (
4     9999, 'Karina', 'Ika', 'x9999', 'karina11@gmail.com', 1, 1002, 'Data Scientist'
5 );

```

			16	1501 Larry	Bott	Sales Rep	Gerard	Bondur	2025-03-11
			17	1504 Barry	Jones	Sales Rep	Gerard	Bondur	2025-03-11
			18	1611 Andy	Fixter	Sales Rep	William	Patterson	2025-03-11
			19	1612 Peter	Marsh	Sales Rep	William	Patterson	2025-03-11
			20	1619 Tom	King	Sales Rep	William	Patterson	2025-03-11
			21	1621 Mami	Nishi	Sales Rep	Mary	Patterson	2025-03-11
			22	1625 Yoshimi	Kato	Sales Rep	Mami	Nishi	2025-03-11
			23	1702 Martin	Gerard	Sales Rep	Gerard	Bondur	2025-03-11
			24	1002 Diane	Murphy	President	NULL	NULL	2025-03-11
			25	1002 Diane	Murphy	President	NULL	NULL	2025-03-11
			26	9999 Karina	Ika	Data Scientist	Diane	Murphy	2025-03-11