

TUGAS PRAKTIKUM 3 DATA WAREHOUSE
MATERI DATA WAREHOUSE



Disusun Oleh:
Maisya Dhia (2341760114)
SIB 2B

SISTEM INFORMASI BISNIS
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
2024



Jobsheet 3: Database Analytical

Tujuan Praktikum

Setelah melakukan praktikum ini, mahasiswa diharapkan dapat lebih mengenal table dimensi, table fakta dan apa itu OLAP

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**.

```
CREATE DATABASES dw_LegendVehicle;
use dw_LegendVehicle;
```



dw_LegendVehicle

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

```
CREATE TABLE dimDate(
id_dimDate int not null AUTO_INCREMENT PRIMARY KEY
date date
year int
month int
day int
);
```

Table	Action	Rows	Type	Collation	Size	Overhead
<input type="checkbox"/> dimDate	★ Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
1 table	Sum	0	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	0 B

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).

Generate rows

Step name: Generate rows

Limit: 1825

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

Fields :

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string?
1	currentDate	Date	dd-mm-yyyy						01-01-2023 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**

Add sequence

Step name: Add sequence

Name of value: incrementDay

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection: 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: 99999999

Help OK Cancel

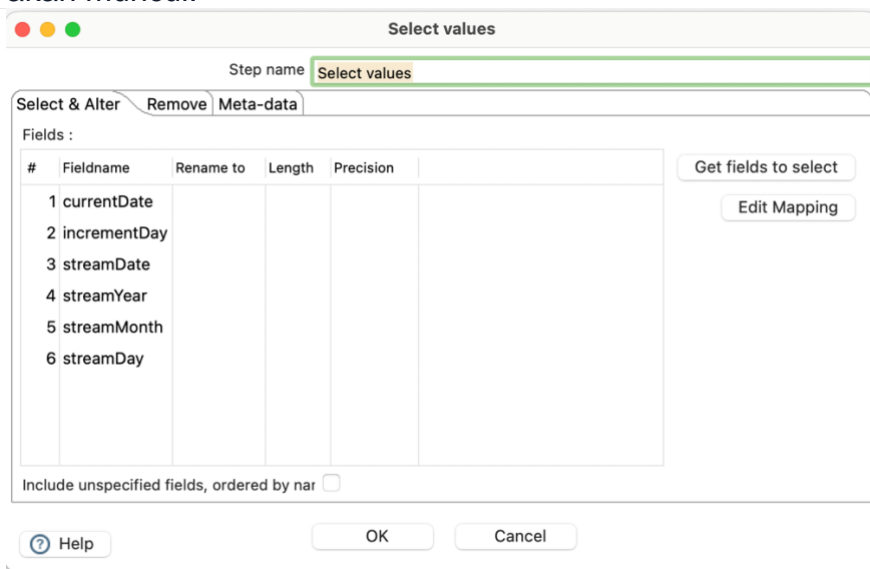
7. Hubungkan output dari **add sequences** menuju **calculator**.

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

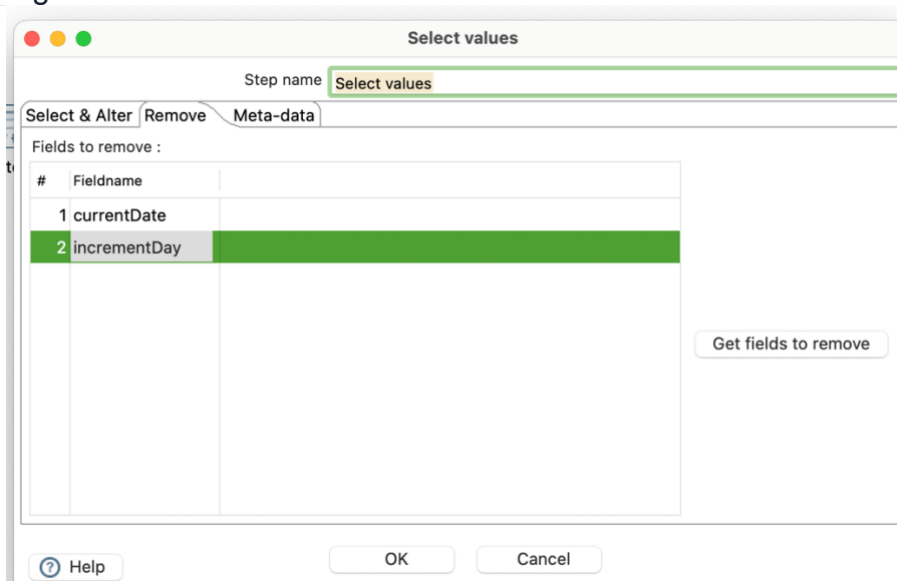
- **streamDate** merupakan kalkulasi dari **CurrentDate + incrementDay**
- **streamYear** merupakan **Year** dari **streamDate**
- **streamMonth** merupakan **Month** dari **streamDate**
- **streamDay** merupakan **Day of month** dari **streamDate**

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.

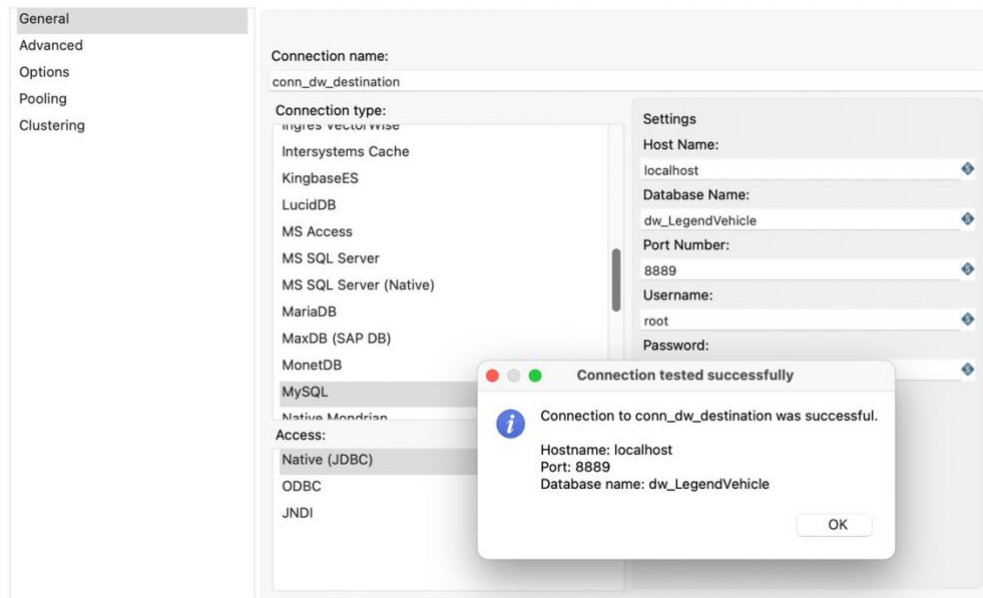


11. Dikarenakan tidak semua fields digunakan, maka pada tab **Remove** diisikan fields **CurrentDate** dan **incrementDay** dikarenakan kedua fields tersebut tidak digunakan.



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**.

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 everything) 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 ☐

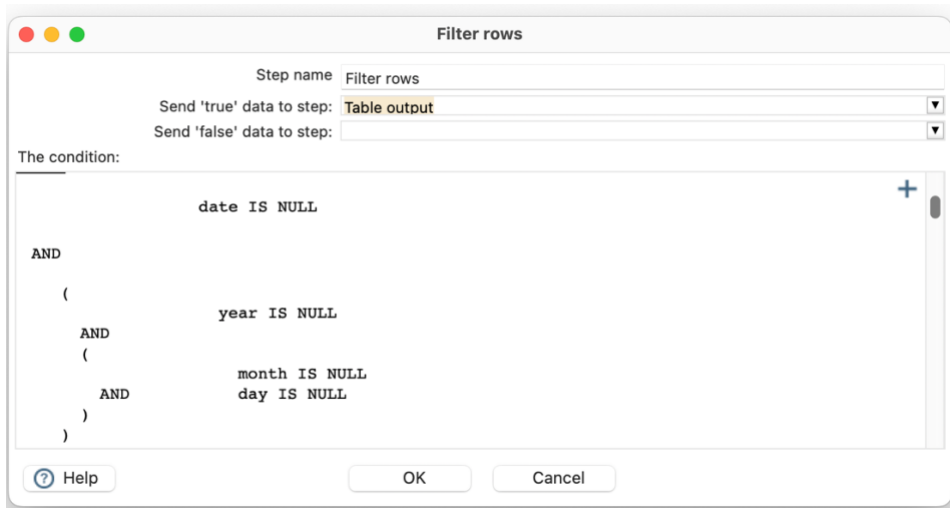
Fail on multiple results? ☐

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**)

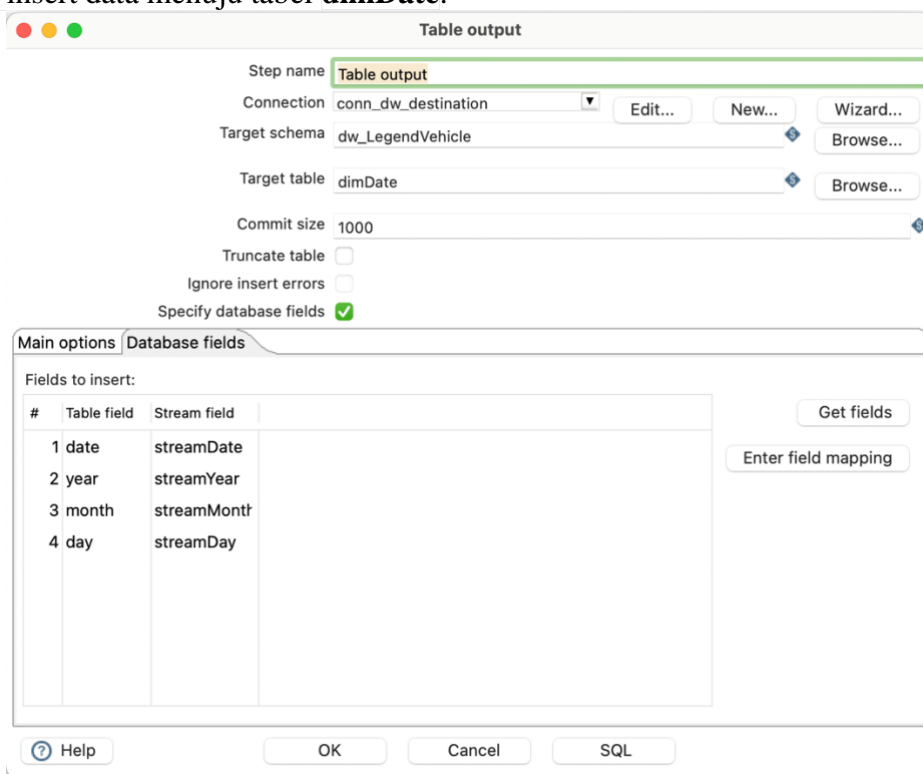


19. Hubungkan output dari **filter rows** menuju **table output**.

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**.





Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025-03-10 10:51:12.520 - Add sequence.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
2025-03-10 10:51:12.641 - Calculator.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
2025-03-10 10:51:12.644 - Select values.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
2025-03-10 10:51:13.378 - Filter rows.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
2025-03-10 10:51:13.382 - Database lookup.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
2025-03-10 10:51:13.457 - Table output.0 - Finished processing (I=0, O=1825, R=1825, W=1825, U=0, E=0)
2025-03-10 10:51:13.464 - Spoon - The transformation has finished!!

Showing rows 0 - 24 (1825 total, query took 0.0025 seconds.)

`SELECT * FROM `dimDate``

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

1 > >> | Number of rows: 25 Filter rows: Search this table

Extra options

	id_dimDate	date	year	month	day
<input type="checkbox"/> Edit Copy Delete	1	2023-01-01	2023	1	1
<input type="checkbox"/> Edit Copy Delete	2	2023-01-02	2023	1	2
<input type="checkbox"/> Edit Copy Delete	3	2023-01-03	2023	1	3
<input type="checkbox"/> Edit Copy Delete	4	2023-01-04	2023	1	4
<input type="checkbox"/> Edit Copy Delete	5	2023-01-05	2023	1	5
<input type="checkbox"/> Edit Copy Delete	6	2023-01-06	2023	1	6
<input type="checkbox"/> Edit Copy Delete	7	2023-01-07	2023	1	7
<input type="checkbox"/> Edit Copy Delete	8	2023-01-08	2023	1	8
<input type="checkbox"/> Edit Copy Delete	9	2023-01-09	2023	1	9
<input type="checkbox"/> Edit Copy Delete	10	2023-01-10	2023	1	10
<input type="checkbox"/> Edit Copy Delete	11	2023-01-11	2023	1	11
<input type="checkbox"/> Edit Copy Delete	12	2023-01-12	2023	1	12
<input type="checkbox"/> Edit Copy Delete	13	2023-01-13	2023	1	13
<input type="checkbox"/> Edit Copy Delete	14	2023-01-14	2023	1	14
<input type="checkbox"/> Edit Copy Delete	15	2023-01-15	2023	1	15
<input type="checkbox"/> Edit Copy Delete	16	2023-01-16	2023	1	16
<input type="checkbox"/> Edit Copy Delete	17	2023-01-17	2023	1	17
<input type="checkbox"/> Edit Copy Delete	18	2023-01-18	2023	1	18
<input type="checkbox"/> Edit Copy Delete	19	2023-01-19	2023	1	19
<input type="checkbox"/> Edit Copy Delete	20	2023-01-20	2023	1	20
<input type="checkbox"/> Edit Copy Delete	21	2023-01-21	2023	1	21
<input type="checkbox"/> Edit Copy Delete	22	2023-01-22	2023	1	22
<input type="checkbox"/> Edit Copy Delete	23	2023-01-23	2023	1	23
<input type="checkbox"/> Edit Copy Delete	24	2023-01-24	2023	1	24
<input type="checkbox"/> Edit Copy Delete	25	2023-01-25	2023	1	25

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	dibawah	dibawah	dibawah
Add Sequences	dibawah	dibawah	dibawah
Calculator	dibawah	dibawah	dibawah
Select values	dibawah	dibawah	dibawah
Database lookup	dibawah	dibawah	dibawah
Filter rows	dibawah	dibawah	dibawah
Table Output	dibawah	dibawah	dibawah

PROSES OBJEK :

GENERATE ROWS

-SS DATA INPUT

Kosong karena belum dibuat

-SS DATA OUTPUT

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty string
1	CurrentDate	Date	dd-mm-yyyy						01-01-2023	N

-KETERANGAN

Generate Rows digunakan untuk membuat baris data baru sesuai jumlah yang telah ditentukan. Pada tahap ini, data input masih kosong karena baru dibuat dalam proses ini. Outputnya berupa sekumpulan baris data awal.

ADD SEQUENCES

-SS DATA INPUT

#	CurrentDate
1	01-01-2023
2	01-01-2023
3	01-01-2023
4	01-01-2023
5	01-01-2023
6	01-01-2023
7	01-01-2023
8	01-01-2023
9	01-01-2023
10	01-01-2023
11	01-01-2023
12	01-01-2023

-SS DATA OUTPUT

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

-KETERANGAN

Add Sequence digunakan untuk menambahkan angka urut pada setiap baris yang dihasilkan dari Generate Rows. Inputnya adalah hasil dari Generate Rows, sedangkan outputnya memiliki tambahan kolom sequence yang berisi angka berurut.

CALCULATOR

-SS DATA INPUT

#	CurrentDate	incrementDay
1	01-01-2023	0
2	01-01-2023	1
3	01-01-2023	2
4	01-01-2023	3
5	01-01-2023	4
6	01-01-2023	5
7	01-01-2023	6

-SS DATA OUTPUT

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 syn
1	streamDate	Date A + B 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 / streamDate	streamDate			None			N			

-KETERANGAN

Calculator digunakan untuk menjumlahkan hari dan mengambil data tahun, bulan, dan hari dari data sebelumnya. Inputnya adalah hasil dari Add Sequence, dan outputnya adalah data dengan tambahan informasi tanggal yang lebih lengkap.

SELECT VALUES

-SS DATA INPUT

☒ First rows
 ☐ Last rows
 ☐ Off

#	CurrentDate	incrementDay	streamDate	streamYear	streamMonth	streamDay
1	01-01-2023	0	2023/01/01 00:01:00.00	2023	1	1
2	01-01-2023	1	2023/01/02 00:01:00.00	2023	1	2
3	01-01-2023	2	2023/01/03 00:01:00.00	2023	1	3
4	01-01-2023	3	2023/01/04 00:01:00.00	2023	1	4
5	01-01-2023	4	2023/01/05 00:01:00.00	2023	1	5
6	01-01-2023	5	2023/01/06 00:01:00.00	2023	1	6

-SS DATA OUTPUT

Select values

Step name

Fields to remove :

#	Fieldname
1	CurrentDate
2	incrementDay

-KETERANGAN

Select Values digunakan untuk memilih hanya field yang relevan untuk proses berikutnya. Inputnya adalah hasil dari Calculator, dan outputnya berisi field yang telah dipilih saja.

DATABASE LOOKUP

-SS DATA INPUT

#	streamDate	streamYear	streamMonth	streamDay
1	2023/01/01 00:01:00.00	2023	1	1
2	2023/01/02 00:01:00.00	2023	1	2
3	2023/01/03 00:01:00.00	2023	1	3
4	2023/01/04 00:01:00.00	2023	1	4
5	2023/01/05 00:01:00.00	2023	1	5
6	2023/01/06 00:01:00.00	2023	1	6
7	2023/01/07 00:01:00.00	2023	1	7

-SS DATA OUTPUT

Step name **Database lookup**

Connection **conn_dw_destination**

Lookup schema **dw_LegendVehicle**

Lookup table **dimDate**

Enable cache? ☐

Cache size in rows (0=cache everything) **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

-KETERANGAN

Database Lookup digunakan untuk memastikan bahwa data yang akan dimasukkan ke tabel dimDate tidak duplikat. Inputnya adalah data dari Select Values, dan outputnya bisa berupa data yang telah diperiksa terhadap tabel dimDate.

FILTER ROWS

-SS DATA INPUT

#	streamDate	streamYear	streamMonth	streamDay	date	year	month	day
1	2023/01/01 00:01:00.00	2023	1	1	<null>	<null>	<null>	<null>
2	2023/01/02 00:01:00.00	2023	1	2	<null>	<null>	<null>	<null>
3	2023/01/03 00:01:00.00	2023	1	3	<null>	<null>	<null>	<null>
4	2023/01/04 00:01:00.00	2023	1	4	<null>	<null>	<null>	<null>
5	2023/01/05 00:01:00.00	2023	1	5	<null>	<null>	<null>	<null>
6	2023/01/06 00:01:00.00	2023	1	6	<null>	<null>	<null>	<null>
7	2023/01/07 00:01:00.00	2023	1	7	<null>	<null>	<null>	<null>
8	2023/01/08 00:01:00.00	2023	1	8	<null>	<null>	<null>	<null>

-SS DATA OUTPUT

Filter r

Step name

Filter rows

Send 'true' data to step:

Table output

Send 'false' data to step:

he condition:

date IS NULL

AND

(

year IS NULL

AND

(

month IS NULL

AND

day IS NULL

)

)

-KETERANGAN

Filter Rows digunakan untuk menyaring data yang belum ada dalam tabel dimDate.

Inputnya adalah hasil dari Database Lookup, dan outputnya hanya menyisakan data yang belum ada di tabel tujuan.

TABLE OUTPUT

-SS DATA INPUT

#	streamDate	streamYear	streamMonth	streamDay	date	year	month	day
1	2023/01/01 00:01:00.00	2023	1	1	<null>	<null>	<null>	<null>
2	2023/01/02 00:01:00.00	2023	1	2	<null>	<null>	<null>	<null>
3	2023/01/03 00:01:00.00	2023	1	3	<null>	<null>	<null>	<null>
4	2023/01/04 00:01:00.00	2023	1	4	<null>	<null>	<null>	<null>
5	2023/01/05 00:01:00.00	2023	1	5	<null>	<null>	<null>	<null>
6	2023/01/06 00:01:00.00	2023	1	6	<null>	<null>	<null>	<null>
7	2023/01/07 00:01:00.00	2023	1	7	<null>	<null>	<null>	<null>
8	2023/01/08 00:01:00.00	2023	1	8	<null>	<null>	<null>	<null>

-SS DATA OUTPUT

Table output

Step name: **Table output**

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

Target schema: **dw_LegendVehicle** [Browse...]

Target table: **dimDate** [Browse...]

Commit size: **1000**

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☒

Main options | **Database fields**

Partition data over tables: ☐

Partitioning field: [Select]

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field? ☐

Field that contains name of table: [Select]

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field: [Select]

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

	id_dimDate	date	year	month	day
1	1	2023-01-01	2023	1	1
2	2	2023-01-02	2023	1	2
3	3	2023-01-03	2023	1	3
4	4	2023-01-04	2023	1	4
5	5	2023-01-05	2023	1	5
6	6	2023-01-06	2023	1	6
7	7	2023-01-07	2023	1	7
8	8	2023-01-08	2023	1	8
9	9	2023-01-09	2023	1	9
10	10	2023-01-10	2023	1	10
11	11	2023-01-11	2023	1	11
12	12	2023-01-12	2023	1	12
13	13	2023-01-13	2023	1	13
14	14	2023-01-14	2023	1	14
15	15	2023-01-15	2023	1	15
16	16	2023-01-16	2023	1	16
17	17	2023-01-17	2023	1	17
18	18	2023-01-18	2023	1	18
19	19	2023-01-19	2023	1	19
20	20	2023-01-20	2023	1	20
21	21	2023-01-21	2023	1	21
22	22	2023-01-22	2023	1	22
23	23	2023-01-23	2023	1	23
24	24	2023-01-24	2023	1	24
25	25	2023-01-25	2023	1	25

-KETERANGAN

Table Output digunakan untuk menyimpan hasil akhir ke dalam tabel dimDate. Inputnya adalah hasil Filter Rows, dan outputnya adalah data yang telah disimpan ke dalam database.



B. Dimensi Pegawai

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

```
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
);
```

```
CREATE TABLE dimPegawai (
  id_dimPegawai INT NOT NULL AUTO_INCREMENT PRIMARY KEY,
  employeeNumber INT(11) UNIQUE,
  firstName VARCHAR(50),
  lastName VARCHAR(50),
  jobTitle VARCHAR(50),
  boss_firstName VARCHAR(50),
  boss_lastName VARCHAR(50),
  updated TIMESTAMP DEFAULT CURRENT_TIMESTAMP ON UPDATE 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.

The screenshot shows the SAP Data Services Designer (SDD) interface. On the left, there is a sidebar with tabs: General, Advanced, Options, Pooling, and Clustering. The 'General' tab is selected. The main area is divided into two sections: 'Connection type' and 'Settings'.

Connection type:

- MS SQL Server
- MS SQL Server (Native)
- MariaDB
- MaxDB (SAP DB)
- MonetDB
- MySQL
- Native Mondrian
- Neoview
- Netezza
- Oracle
- Oracle RDB

Access:

- Native (JDBC)
- ODBC
- JNDI

Settings:

- Host Name: localhost
- Database Name: dw_LegendVehicle
- Port Number: 8889
- Username: root
- Password: [masked]
- ☒ Use Result Streaming Cursor

At the bottom of the window, there are three buttons: Test, Feature List, and Explore.

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.

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

The screenshot shows the 'Table input' configuration window. It has a title bar with three colored buttons (red, yellow, green). The window contains the following elements:

- Step name:** A text field containing 'Table input'.
- Connection:** A dropdown menu showing 'conn_oltp_resources'.
- Buttons:** 'Edit...', 'New...', 'Wizard...', and 'Get SQL select statement...'.
- SQL:** A text area containing the following SQL query:


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


General
Advanced
Options
Pooling
Clustering

Connection name:
conn_oltp_resources

Connection type:
MySQL
Native Mondrian
Neoview
Nettezza
Oracle
Oracle RDB
PostgreSQL
Redshift
Remedy Action Request System
SAP ERP System
SQLite

Access:
Native (JDBC)
ODBC
JNDI

Settings
Host Name:
localhost
Database Name:
classmodels
Port Number:
8889
Username:
root
Password:

☒ Use Result Streaming Cursor

5. Hubungkan output **table input** pada **select values**.

Konfigurasi pada **Select values** yaitu mengambil data dari field

6. **employeenumber, lastname, firstname, jobtitle , lastname_1 dan firstname_1** sebagai **data stream** yang digunakan pada proses ETL pada tab **select & alter**.

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			

Get fields to select
Edit Mapping

Include unspecified fields, ordered by name ☐

7. Hilangkan field lain yang tidak digunakan pada tab **remove**.

The screenshot shows a window titled 'Select values' with a 'Step name' field containing 'Select values'. Below this are three tabs: 'Select & Alter', 'Remove', and 'Meta-data'. The 'Remove' tab is selected, and it displays a section 'Fields to remove :'. This section contains a table with two columns: '#' and 'Fieldname'. The table lists 10 fields, with the 10th field, 'jobTitle_1', highlighted in green. To the right of the table is a button labeled 'Get fields to remove'.

#	Fieldname
1	extension
2	email
3	officeCode
4	reportsTo
5	employeeNumber
6	extension_1
7	email_1
8	officeCode_1
9	reportsTo_1
10	jobTitle_1

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_oltp_resources

Edit...

New...

Wizard...

Lookup schema

dw_LegendVehicle

Browse...

Lookup table

dimPegawai

Browse...

Enable cache?

☐

Cache size in rows (0=cache everything)

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	=	StreamLastname	

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

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**.

Step name **Filter rows**

Send 'true' data to step  **table output**

Send 'false' data to step:

The condition:

```
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
            )
        )
    )
)
```

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**.

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



Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025-03-10 12:10:31.145 - Spoon - Transformation opened.
2025-03-10 12:10:31.146 - Spoon - Launching transformation [dw_js3_B]...
2025-03-10 12:10:31.146 - Spoon - Started the transformation execution.
2025-03-10 12:10:31.422 - dw_js3_B - Dispatching started for transformation [dw_js3_B]
2025-03-10 12:10:31.449 - Table output.0 - Connected to database [conn_dw_destination] (commit=1000)
2025-03-10 12:10:31.750 - Table input.0 - Finished reading query, closing connection
2025-03-10 12:10:31.769 - Table input.0 - Finished processing (I=23, O=0, R=0, W=23, U=0, E=0)
2025-03-10 12:10:31.774 - Select values.0 - Finished processing (I=0, O=0, R=23, W=23, U=0, E=0)
2025-03-10 12:10:31.839 - Filter rows.0 - Finished processing (I=0, O=0, R=23, W=23, U=0, E=0)
2025-03-10 12:10:31.843 - Database lookup.0 - Finished processing (I=0, O=0, R=23, W=23, U=0, E=0)
2025-03-10 12:10:31.871 - Table output.0 - Finished processing (I=0, O=23, R=23, W=23, U=0, E=0)
2025-03-10 12:10:31.874 - Spoon - The transformation has finished!!

Extra options												
+ -		id_dimPegawai	employeeNumber	firstName	lastName	jobTitle	boss_firstName	boss_lastName	updated			
<input type="checkbox"/>	 Edit	 Copy	 Delete	1	1002	Diane	Murphy	President	NULL	NULL	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	2	1056	Mary	Patterson	VP Sales	Diane	Murphy	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	3	1076	Jeff	Firrelli	VP Marketing	Diane	Murphy	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	4	1088	William	Patterson	Sales Manager (APAC)	Mary	Patterson	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	5	1102	Gerard	Bondur	Sale Manager (EMEA)	Mary	Patterson	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	6	1143	Anthony	Bow	Sales Manager (NA)	Mary	Patterson	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	7	1165	Leslie	Jennings	Sales Rep	Anthony	Bow	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	8	1166	Leslie	Thompson	Sales Rep	Anthony	Bow	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	9	1188	Julie	Firrelli	Sales Rep	Anthony	Bow	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	10	1216	Steve	Patterson	Sales Rep	Anthony	Bow	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	11	1286	Foon Yue	Tseng	Sales Rep	Anthony	Bow	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	12	1323	George	Vanauf	Sales Rep	Anthony	Bow	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	13	1337	Loui	Bondur	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	14	1370	Gerard	Hernandez	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	15	1401	Pamela	Castillo	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	16	1501	Larry	Bott	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	17	1504	Barry	Jones	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	18	1611	Andy	Fixter	Sales Rep	William	Patterson	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	19	1612	Peter	Marsh	Sales Rep	William	Patterson	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	20	1619	Tom	King	Sales Rep	William	Patterson	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	21	1621	Mami	Nishi	Sales Rep	Mary	Patterson	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	22	1625	Yoshimi	Kato	Sales Rep	Mami	Nishi	2025-03-10 12:10:31	
<input type="checkbox"/>	 Edit	 Copy	 Delete	23	1702	Martin	Gerard	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31	

TUGAS 2

- 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			
Select values			
Database lookup			
Filter rows			
Table Output			

PROSES OBJEK :

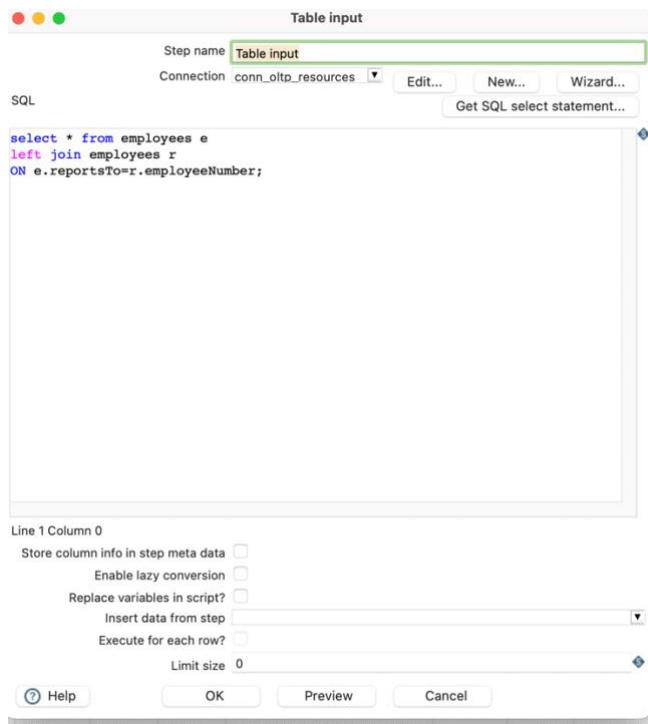
TABLE INPUT

-SS DATA INPUT

Berasal dari table classicmodel

Table	Action	Rows	Type	Collation	Size	Overhead
customers		122	InnoDB	utf8mb4_0900_ai_ci	64.0 KiB	-
dimDate		0	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
employees		23	InnoDB	utf8mb4_0900_ai_ci	48.0 KiB	-
offices		7	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
orderdetails		2,996	InnoDB	utf8mb4_0900_ai_ci	240.0 KiB	-
orders		326	InnoDB	utf8mb4_0900_ai_ci	64.0 KiB	-
payments		273	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
productlines		7	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
products		110	InnoDB	utf8mb4_0900_ai_ci	80.0 KiB	-

-SS DATA OUTPUT



-KETERANGAN

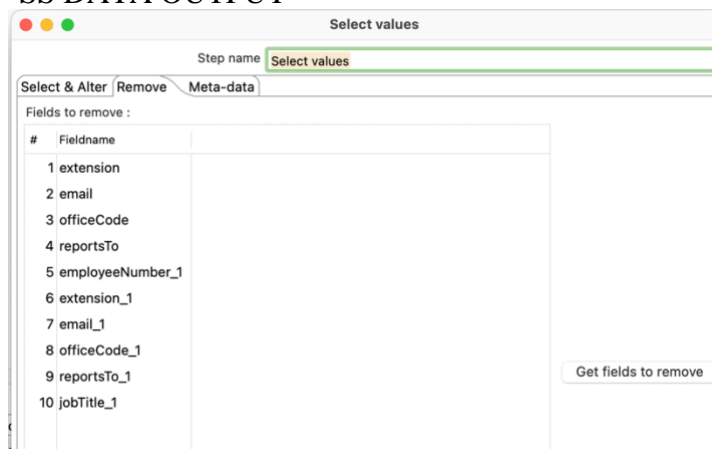
Table Input digunakan untuk mengambil data dari database OLTP. Data inputnya berasal dari tabel operasional, dan outputnya adalah hasil ekstraksi data ke dalam stream transformasi.

SELECT VALUES

-SS DATA INPUT

employeeNumber	lastName	firstName	extension	email	officeCode	reportsTo	jobTitle	employeeNumber_1	lastName_1	firstName_1	extension_1	email_1
1002	Murphy	Diane	x5800	dmurphy@classicmodelcars.com	1	<null>	President	<null>	<null>	<null>	<null>	<null>
1056	Patterson	Mary	x4611	mpatterso@classicmodelcars.com	1	1002	VP Sales	1002	Murphy	Diane	x5800	dmurphy@classicmodelcars.com
email_1				officeCode_1		reportsTo_1		jobTitle_1				
<null>				<null>		<null>		<null>				
dmurphy@classicmodelcars.com				1		<null>		President				

-SS DATA OUTPUT



-KETERANGAN

Select Values digunakan untuk memilih hanya field yang relevan untuk proses Transform dan Load. Inputnya adalah hasil dari Table Input, dan outputnya adalah data dengan kolom yang sudah difilter.

DATABASE LOOKUP

-SS DATA INPUT

	StreamEmployeeNumber	StreamLastName	StreamFirstname	StreamJobtitle	StreamLastnameBoss	StreamFirstnameBoss
1	1002	Murphy	Diane	President	<null>	<null>
2	1056	Patterson	Mary	VP Sales	Murphy	Diane
3	1076	Firrelli	Jeff	VP Marketing	Murphy	Diane

-SS DATA OUTPUT

-KETERANGAN

Database Lookup digunakan untuk memastikan data yang akan dimasukkan ke tabel dimEmployees tidak kembar dengan data yang sudah ada. Inputnya adalah hasil dari Select Values, dan outputnya menunjukkan apakah data sudah ada dalam tabel tujuan atau belum.

FILTER ROWS

-SS DATA INPUT

#	StreamEmployeeNumber	StreamLastName	StreamFirstname	StreamJobtitle	StreamLastnameBoss	StreamFirstnameBoss	employeeNumber	firstName	lastName	jobTitle
1	1002	Murphy	Diane	President	<null>	<null>	<null>	<null>	<null>	<null>
2	1056	Patterson	Mary	VP Sales	Murphy	Diane	<null>	<null>	<null>	<null>
3	1076	Firrelli	Jeff	VP Marketing	Murphy	Diane	<null>	<null>	<null>	<null>
4	1088	Patterson	William	Sales Manager (APAC)	Patterson	Mary	1088	William	Patterson	Sales Manager

-SS DATA OUTPUT

Filter rows

Step name

Filter rows

Send 'true' data to step:

Table output

Send 'false' data to step:

The condition:

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

-KETERANGAN

Filter Rows digunakan untuk menyaring hanya data yang belum ada dalam tabel dimEmployees. Inputnya adalah hasil dari Database Lookup, dan outputnya adalah data yang benar-benar baru untuk dimasukkan.

TABLE OUTPUT

-SS DATA INPUT

StreamErr	StreamLastname	StreamFirstname	StreamJobtitle	StreamLastnameBoss	StreamFirstnameBoss	employeeNumber	firstName	lastName	jobTitle	boss_firstname	boss_lastName
1002	Murphy	Diane	President	<null>	<null>		<null>	<null>	<null>	<null>	<null>
1056	Patterson	Mary	VP Sales	Murphy	Diane		<null>	<null>	<null>	<null>	<null>
1076	Firrelli	Jeff	VP Marketing	Murphy	Diane		<null>	<null>	<null>	<null>	<null>

-SS DATA OUTPUT

		id_dimPegawai	employeeNumber	firstName	lastName	jobTitle	boss_firstName	boss_lastName	updated	
			1	1002	Diane	Murphy	President	NULL	NULL	2025-03-10 12:10:31
			2	1056	Mary	Patterson	VP Sales	Diane	Murphy	2025-03-10 12:10:31
			3	1076	Jeff	Firrelli	VP Marketing	Diane	Murphy	2025-03-10 12:10:31
			4	1088	William	Patterson	Sales Manager (APAC)	Mary	Patterson	2025-03-10 12:10:31
			5	1102	Gerard	Bondur	Sale Manager (EMEA)	Mary	Patterson	2025-03-10 12:10:31
			6	1143	Anthony	Bow	Sales Manager (NA)	Mary	Patterson	2025-03-10 12:10:31
			7	1165	Leslie	Jennings	Sales Rep	Anthony	Bow	2025-03-10 12:10:31
			8	1166	Leslie	Thompson	Sales Rep	Anthony	Bow	2025-03-10 12:10:31
			9	1188	Julie	Firrelli	Sales Rep	Anthony	Bow	2025-03-10 12:10:31
			10	1216	Steve	Patterson	Sales Rep	Anthony	Bow	2025-03-10 12:10:31
			11	1286	Foon Yue	Tseng	Sales Rep	Anthony	Bow	2025-03-10 12:10:31
			12	1323	George	Vanau	Sales Rep	Anthony	Bow	2025-03-10 12:10:31
			13	1337	Loui	Bondur	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31
			14	1370	Gerard	Hernandez	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31
			15	1401	Pamela	Castillo	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31
			16	1501	Larry	Bott	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31
			17	1504	Barry	Jones	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31
			18	1611	Andy	Fixter	Sales Rep	William	Patterson	2025-03-10 12:10:31
			19	1612	Peter	Marsh	Sales Rep	William	Patterson	2025-03-10 12:10:31
			20	1619	Tom	King	Sales Rep	William	Patterson	2025-03-10 12:10:31
			21	1621	Mami	Nishi	Sales Rep	Mary	Patterson	2025-03-10 12:10:31
			22	1625	Yoshimi	Kato	Sales Rep	Mami	Nishi	2025-03-10 12:10:31
			23	1702	Martin	Gerard	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31

-KETERANGAN

Table Output digunakan untuk menyimpan hasil akhir ke dalam tabel dimEmployees. Inputnya adalah hasil Filter Rows, dan outputnya adalah data yang sudah tersimpan dalam database.

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

- Tidak, karena proses Database Lookup dan Filter Rows memastikan bahwa hanya data yang belum ada di dimEmployees yang akan dimasukkan. Data yang sudah ada akan difilter sehingga tidak terjadi duplikasi.

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

```
1 INSERT INTO dimPegawai (employeeNumber, firstName, lastName, jobTitle, boss_firstName, boss_lastName)
2 VALUES (1411, 'Maisya', 'Azzahra', 'Sales Rep', 'Gerard', 'Jones');
3
```

 Edit  Copy  Delete	23	1702 Martin	Gerard	Sales Rep	Gerard	Bondur	2025-03-10 12:10:31
 Edit  Copy  Delete	24	1411 Maisya	Azzahra	Sales Rep	Gerard	Jones	2025-03-12 22:40:37

- Jika nama baru ditambahkan ke tabel Pegawai di OLTP, maka saat transformasi dijalankan kembali, nama tersebut akan ikut masuk ke dalam proses ETL. Data ini akan lolos dari tahap Database Lookup dan Filter Rows karena belum ada di dimPegawai, sehingga akan tersimpan sebagai data baru di tabel dimPegawai setelah proses Table Output selesai.