

LAPORAN PRAKTIKUM DATA WAREHOUSE

JOBSHEET 3 : DATA BASE ANALYTICAL



DISUSUN OLEH :

AQILA NUR AZZA (2341760022)

KELAS 2A-SIB/04

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

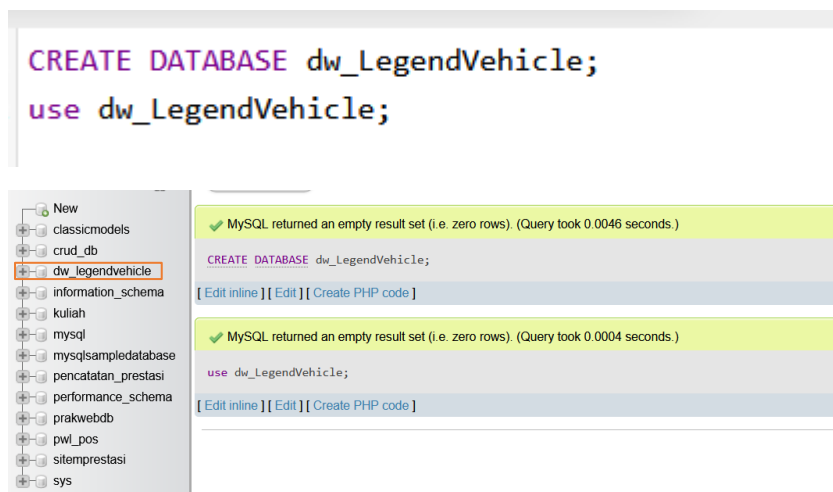
Berdasarkan studi kasus sebelumnya, Pimpinan dari LegendVehicle mulai mengenal teknologi dan berpandangan visioner. LegendVehicle akan menerapkan data warehouse pada proses bisnisnya untuk menganalisa proses transaksi yang dilihat dari pembayaran yang masuk. Proses ini digunakan untuk menganalisa KPI "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 database OLAP.

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**.

```
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).



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).

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										

Buttons: Help, OK, Preview, Cancel

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
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**

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: 999999999

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**

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 sym
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 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.

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 :

#	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, ☐

Help OK Cancel

Select values

Step name: Select values

Select & Alter Remove Meta-data

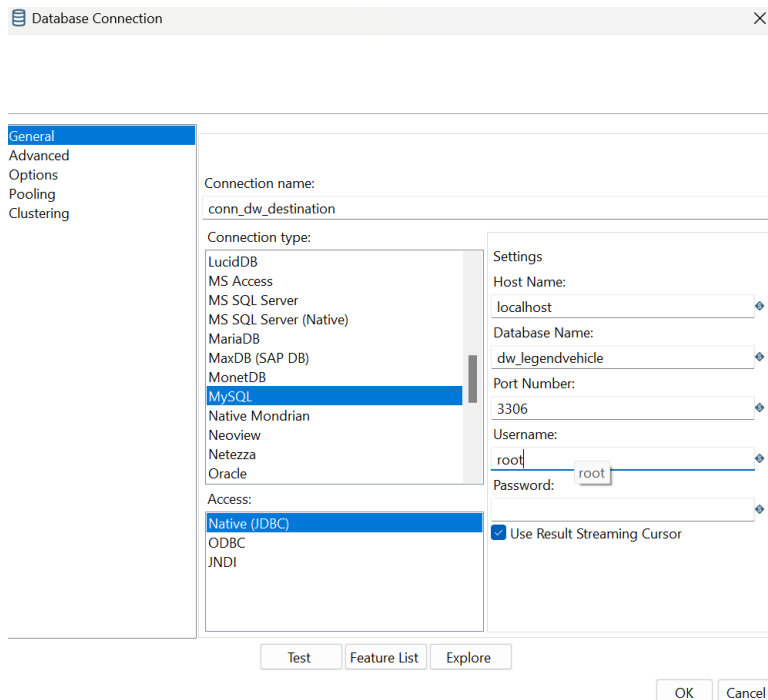
Fields to remove :

#	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 **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): 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**)

Filter rows

Step name: Filter rows

Send 'true' data to step: Table output

Send 'false' data to step:

The condition:

```

date IS NULL
AND
(
  AND
  (
    AND
    (
      month IS NULL
      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**.

Proses Objek	SS data input	SS data output	Keterangan																																																																																																								
Generate rows	Tidak ada	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># CurrentDate</div> <table><tr><td>1</td><td>01-01-2003</td></tr><tr><td>2</td><td>01-01-2003</td></tr><tr><td>3</td><td>01-01-2003</td></tr><tr><td>4</td><td>01-01-2003</td></tr><tr><td>5</td><td>01-01-2003</td></tr><tr><td>6</td><td>01-01-2003</td></tr><tr><td>7</td><td>01-01-2003</td></tr><tr><td>8</td><td>01-01-2003</td></tr><tr><td>n</td><td>01-01-2003</td></tr></table>	1	01-01-2003	2	01-01-2003	3	01-01-2003	4	01-01-2003	5	01-01-2003	6	01-01-2003	7	01-01-2003	8	01-01-2003	n	01-01-2003	Step ini digunakan untuk menghasilkan data tanggal awal secara otomatis dan tidak memiliki input karena merupakan sumber data pertama.																																																																																						
1	01-01-2003																																																																																																										
2	01-01-2003																																																																																																										
3	01-01-2003																																																																																																										
4	01-01-2003																																																																																																										
5	01-01-2003																																																																																																										
6	01-01-2003																																																																																																										
7	01-01-2003																																																																																																										
8	01-01-2003																																																																																																										
n	01-01-2003																																																																																																										
Add Sequences	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># CurrentDate</div> <table><tr><td>1</td><td>01-01-2003</td></tr><tr><td>2</td><td>01-01-2003</td></tr><tr><td>3</td><td>01-01-2003</td></tr><tr><td>4</td><td>01-01-2003</td></tr><tr><td>5</td><td>01-01-2003</td></tr><tr><td>6</td><td>01-01-2003</td></tr><tr><td>7</td><td>01-01-2003</td></tr><tr><td>8</td><td>01-01-2003</td></tr><tr><td>n</td><td>01-01-2003</td></tr></table>	1	01-01-2003	2	01-01-2003	3	01-01-2003	4	01-01-2003	5	01-01-2003	6	01-01-2003	7	01-01-2003	8	01-01-2003	n	01-01-2003	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># CurrentDate incrementDay</div> <table><tr><td>1</td><td>01-01-2003</td><td>0</td></tr><tr><td>2</td><td>01-01-2003</td><td>1</td></tr><tr><td>3</td><td>01-01-2003</td><td>2</td></tr><tr><td>4</td><td>01-01-2003</td><td>3</td></tr><tr><td>5</td><td>01-01-2003</td><td>4</td></tr><tr><td>6</td><td>01-01-2003</td><td>5</td></tr><tr><td>7</td><td>01-01-2003</td><td>6</td></tr><tr><td>8</td><td>01-01-2003</td><td>7</td></tr><tr><td>n</td><td>01-01-2003</td><td>n</td></tr></table>	1	01-01-2003	0	2	01-01-2003	1	3	01-01-2003	2	4	01-01-2003	3	5	01-01-2003	4	6	01-01-2003	5	7	01-01-2003	6	8	01-01-2003	7	n	01-01-2003	n	Step ini akan menambahkan urutan angka yang bisa digunakan untuk perhitungan lebih lanjut dan berguna untuk membuat identifikasi unik atau perhitungan urutan.																																																											
1	01-01-2003																																																																																																										
2	01-01-2003																																																																																																										
3	01-01-2003																																																																																																										
4	01-01-2003																																																																																																										
5	01-01-2003																																																																																																										
6	01-01-2003																																																																																																										
7	01-01-2003																																																																																																										
8	01-01-2003																																																																																																										
n	01-01-2003																																																																																																										
1	01-01-2003	0																																																																																																									
2	01-01-2003	1																																																																																																									
3	01-01-2003	2																																																																																																									
4	01-01-2003	3																																																																																																									
5	01-01-2003	4																																																																																																									
6	01-01-2003	5																																																																																																									
7	01-01-2003	6																																																																																																									
8	01-01-2003	7																																																																																																									
n	01-01-2003	n																																																																																																									
Calculator	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># CurrentDate incrementDay</div> <table><tr><td>1</td><td>01-01-2003</td><td>0</td></tr><tr><td>2</td><td>01-01-2003</td><td>1</td></tr><tr><td>3</td><td>01-01-2003</td><td>2</td></tr><tr><td>4</td><td>01-01-2003</td><td>3</td></tr><tr><td>5</td><td>01-01-2003</td><td>4</td></tr><tr><td>6</td><td>01-01-2003</td><td>5</td></tr><tr><td>7</td><td>01-01-2003</td><td>6</td></tr><tr><td>8</td><td>01-01-2003</td><td>7</td></tr><tr><td>n</td><td>01-01-2003</td><td>n</td></tr></table>	1	01-01-2003	0	2	01-01-2003	1	3	01-01-2003	2	4	01-01-2003	3	5	01-01-2003	4	6	01-01-2003	5	7	01-01-2003	6	8	01-01-2003	7	n	01-01-2003	n	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># CurrentDate incrementDay streamDate streamYear streamMonth streamDay</div> <table><tr><td>1</td><td>01-01-2003</td><td>0</td><td>2003/01/01 00:00:00.000</td><td>2003</td><td>1</td><td>1</td></tr><tr><td>2</td><td>01-01-2003</td><td>1</td><td>2003/01/02 00:00:00.000</td><td>2003</td><td>1</td><td>2</td></tr><tr><td>3</td><td>01-01-2003</td><td>2</td><td>2003/01/03 00:00:00.000</td><td>2003</td><td>1</td><td>3</td></tr><tr><td>4</td><td>01-01-2003</td><td>3</td><td>2003/01/04 00:00:00.000</td><td>2003</td><td>1</td><td>4</td></tr><tr><td>5</td><td>01-01-2003</td><td>4</td><td>2003/01/05 00:00:00.000</td><td>2003</td><td>1</td><td>5</td></tr><tr><td>6</td><td>01-01-2003</td><td>5</td><td>2003/01/06 00:00:00.000</td><td>2003</td><td>1</td><td>6</td></tr><tr><td>7</td><td>01-01-2003</td><td>6</td><td>2003/01/07 00:00:00.000</td><td>2003</td><td>1</td><td>7</td></tr><tr><td>8</td><td>01-01-2003</td><td>7</td><td>2003/01/08 00:00:00.000</td><td>2003</td><td>1</td><td>8</td></tr><tr><td>n</td><td>01-01-2003</td><td>n</td><td>2003/01/01 00:00:00.000</td><td>2003</td><td>1</td><td>n</td></tr></table>	1	01-01-2003	0	2003/01/01 00:00:00.000	2003	1	1	2	01-01-2003	1	2003/01/02 00:00:00.000	2003	1	2	3	01-01-2003	2	2003/01/03 00:00:00.000	2003	1	3	4	01-01-2003	3	2003/01/04 00:00:00.000	2003	1	4	5	01-01-2003	4	2003/01/05 00:00:00.000	2003	1	5	6	01-01-2003	5	2003/01/06 00:00:00.000	2003	1	6	7	01-01-2003	6	2003/01/07 00:00:00.000	2003	1	7	8	01-01-2003	7	2003/01/08 00:00:00.000	2003	1	8	n	01-01-2003	n	2003/01/01 00:00:00.000	2003	1	n	Step ini digunakan untuk melakukan perhitungan atau manipulasi data, seperti menambahkan hari atau memisahkan tanggal ke dalam komponen tahun, bulan, dan hari serta Berguna untuk analisis lebih lanjut atau pembuatan tabel dimensi waktu.														
1	01-01-2003	0																																																																																																									
2	01-01-2003	1																																																																																																									
3	01-01-2003	2																																																																																																									
4	01-01-2003	3																																																																																																									
5	01-01-2003	4																																																																																																									
6	01-01-2003	5																																																																																																									
7	01-01-2003	6																																																																																																									
8	01-01-2003	7																																																																																																									
n	01-01-2003	n																																																																																																									
1	01-01-2003	0	2003/01/01 00:00:00.000	2003	1	1																																																																																																					
2	01-01-2003	1	2003/01/02 00:00:00.000	2003	1	2																																																																																																					
3	01-01-2003	2	2003/01/03 00:00:00.000	2003	1	3																																																																																																					
4	01-01-2003	3	2003/01/04 00:00:00.000	2003	1	4																																																																																																					
5	01-01-2003	4	2003/01/05 00:00:00.000	2003	1	5																																																																																																					
6	01-01-2003	5	2003/01/06 00:00:00.000	2003	1	6																																																																																																					
7	01-01-2003	6	2003/01/07 00:00:00.000	2003	1	7																																																																																																					
8	01-01-2003	7	2003/01/08 00:00:00.000	2003	1	8																																																																																																					
n	01-01-2003	n	2003/01/01 00:00:00.000	2003	1	n																																																																																																					
Select values	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># CurrentDate incrementDay streamDate streamYear streamMonth streamDay</div> <table><tr><td>1</td><td>01-01-2003</td><td>0</td><td>2003/01/01 00:00:00.000</td><td>2003</td><td>1</td><td>1</td></tr><tr><td>2</td><td>01-01-2003</td><td>1</td><td>2003/01/02 00:00:00.000</td><td>2003</td><td>1</td><td>2</td></tr><tr><td>3</td><td>01-01-2003</td><td>2</td><td>2003/01/03 00:00:00.000</td><td>2003</td><td>1</td><td>3</td></tr><tr><td>4</td><td>01-01-2003</td><td>3</td><td>2003/01/04 00:00:00.000</td><td>2003</td><td>1</td><td>4</td></tr><tr><td>5</td><td>01-01-2003</td><td>4</td><td>2003/01/05 00:00:00.000</td><td>2003</td><td>1</td><td>5</td></tr><tr><td>6</td><td>01-01-2003</td><td>5</td><td>2003/01/06 00:00:00.000</td><td>2003</td><td>1</td><td>6</td></tr><tr><td>7</td><td>01-01-2003</td><td>6</td><td>2003/01/07 00:00:00.000</td><td>2003</td><td>1</td><td>7</td></tr><tr><td>8</td><td>01-01-2003</td><td>7</td><td>2003/01/08 00:00:00.000</td><td>2003</td><td>1</td><td>8</td></tr><tr><td>n</td><td>01-01-2003</td><td>n</td><td>2003/01/01 00:00:00.000</td><td>2003</td><td>1</td><td>n</td></tr></table>	1	01-01-2003	0	2003/01/01 00:00:00.000	2003	1	1	2	01-01-2003	1	2003/01/02 00:00:00.000	2003	1	2	3	01-01-2003	2	2003/01/03 00:00:00.000	2003	1	3	4	01-01-2003	3	2003/01/04 00:00:00.000	2003	1	4	5	01-01-2003	4	2003/01/05 00:00:00.000	2003	1	5	6	01-01-2003	5	2003/01/06 00:00:00.000	2003	1	6	7	01-01-2003	6	2003/01/07 00:00:00.000	2003	1	7	8	01-01-2003	7	2003/01/08 00:00:00.000	2003	1	8	n	01-01-2003	n	2003/01/01 00:00:00.000	2003	1	n	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># streamDate streamYear streamMonth streamDay</div> <table><tr><td>1</td><td>2003/01/01 00:00:00.000</td><td>2003</td><td>1</td><td>1</td></tr><tr><td>2</td><td>2003/01/02 00:00:00.000</td><td>2003</td><td>1</td><td>2</td></tr><tr><td>3</td><td>2003/01/03 00:00:00.000</td><td>2003</td><td>1</td><td>3</td></tr><tr><td>4</td><td>2003/01/04 00:00:00.000</td><td>2003</td><td>1</td><td>4</td></tr><tr><td>5</td><td>2003/01/05 00:00:00.000</td><td>2003</td><td>1</td><td>5</td></tr><tr><td>6</td><td>2003/01/06 00:00:00.000</td><td>2003</td><td>1</td><td>6</td></tr><tr><td>7</td><td>2003/01/07 00:00:00.000</td><td>2003</td><td>1</td><td>7</td></tr><tr><td>8</td><td>2003/01/08 00:00:00.000</td><td>2003</td><td>1</td><td>8</td></tr></table>	1	2003/01/01 00:00:00.000	2003	1	1	2	2003/01/02 00:00:00.000	2003	1	2	3	2003/01/03 00:00:00.000	2003	1	3	4	2003/01/04 00:00:00.000	2003	1	4	5	2003/01/05 00:00:00.000	2003	1	5	6	2003/01/06 00:00:00.000	2003	1	6	7	2003/01/07 00:00:00.000	2003	1	7	8	2003/01/08 00:00:00.000	2003	1	8	Step ini digunakan untuk memilih kolom yang diperlukan, sehingga hanya kolom yang relevan yang dipertahankan dan Bisa digunakan untuk mengganti nama kolom atau membuang kolom yang tidak diperlukan.	
1	01-01-2003	0	2003/01/01 00:00:00.000	2003	1	1																																																																																																					
2	01-01-2003	1	2003/01/02 00:00:00.000	2003	1	2																																																																																																					
3	01-01-2003	2	2003/01/03 00:00:00.000	2003	1	3																																																																																																					
4	01-01-2003	3	2003/01/04 00:00:00.000	2003	1	4																																																																																																					
5	01-01-2003	4	2003/01/05 00:00:00.000	2003	1	5																																																																																																					
6	01-01-2003	5	2003/01/06 00:00:00.000	2003	1	6																																																																																																					
7	01-01-2003	6	2003/01/07 00:00:00.000	2003	1	7																																																																																																					
8	01-01-2003	7	2003/01/08 00:00:00.000	2003	1	8																																																																																																					
n	01-01-2003	n	2003/01/01 00:00:00.000	2003	1	n																																																																																																					
1	2003/01/01 00:00:00.000	2003	1	1																																																																																																							
2	2003/01/02 00:00:00.000	2003	1	2																																																																																																							
3	2003/01/03 00:00:00.000	2003	1	3																																																																																																							
4	2003/01/04 00:00:00.000	2003	1	4																																																																																																							
5	2003/01/05 00:00:00.000	2003	1	5																																																																																																							
6	2003/01/06 00:00:00.000	2003	1	6																																																																																																							
7	2003/01/07 00:00:00.000	2003	1	7																																																																																																							
8	2003/01/08 00:00:00.000	2003	1	8																																																																																																							
Database lookup	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># streamDate streamYear streamMonth streamDay</div> <table><tr><td>1</td><td>2003/01/01 00:00:00.000</td><td>2003</td><td>1</td><td>1</td></tr><tr><td>2</td><td>2003/01/02 00:00:00.000</td><td>2003</td><td>1</td><td>2</td></tr><tr><td>3</td><td>2003/01/03 00:00:00.000</td><td>2003</td><td>1</td><td>3</td></tr><tr><td>4</td><td>2003/01/04 00:00:00.000</td><td>2003</td><td>1</td><td>4</td></tr><tr><td>5</td><td>2003/01/05 00:00:00.000</td><td>2003</td><td>1</td><td>5</td></tr><tr><td>6</td><td>2003/01/06 00:00:00.000</td><td>2003</td><td>1</td><td>6</td></tr><tr><td>7</td><td>2003/01/07 00:00:00.000</td><td>2003</td><td>1</td><td>7</td></tr><tr><td>8</td><td>2003/01/08 00:00:00.000</td><td>2003</td><td>1</td><td>8</td></tr></table>	1	2003/01/01 00:00:00.000	2003	1	1	2	2003/01/02 00:00:00.000	2003	1	2	3	2003/01/03 00:00:00.000	2003	1	3	4	2003/01/04 00:00:00.000	2003	1	4	5	2003/01/05 00:00:00.000	2003	1	5	6	2003/01/06 00:00:00.000	2003	1	6	7	2003/01/07 00:00:00.000	2003	1	7	8	2003/01/08 00:00:00.000	2003	1	8	<div>Execution Results</div> <div>Execution History Logging Step Metrics Performance Graph Metrics Preview data</div> <div>First rows Last rows Off</div> <div># streamDate streamYear streamMonth streamDay date year month day</div> <table><tr><td>1</td><td>2003/01/01 00:00:00.000</td><td>2003</td><td>1</td><td>1</td><td>valid</td><td>valid</td><td>valid</td></tr><tr><td>2</td><td>2003/01/02 00:00:00.000</td><td>2003</td><td>1</td><td>2</td><td>valid</td><td>valid</td><td>valid</td></tr><tr><td>3</td><td>2003/01/03 00:00:00.000</td><td>2003</td><td>1</td><td>3</td><td>valid</td><td>valid</td><td>valid</td></tr><tr><td>4</td><td>2003/01/04 00:00:00.000</td><td>2003</td><td>1</td><td>4</td><td>valid</td><td>valid</td><td>valid</td></tr><tr><td>5</td><td>2003/01/05 00:00:00.000</td><td>2003</td><td>1</td><td>5</td><td>valid</td><td>valid</td><td>valid</td></tr><tr><td>6</td><td>2003/01/06 00:00:00.000</td><td>2003</td><td>1</td><td>6</td><td>valid</td><td>valid</td><td>valid</td></tr><tr><td>7</td><td>2003/01/07 00:00:00.000</td><td>2003</td><td>1</td><td>7</td><td>valid</td><td>valid</td><td>valid</td></tr><tr><td>8</td><td>2003/01/08 00:00:00.000</td><td>2003</td><td>1</td><td>8</td><td>valid</td><td>valid</td><td>valid</td></tr></table>	1	2003/01/01 00:00:00.000	2003	1	1	valid	valid	valid	2	2003/01/02 00:00:00.000	2003	1	2	valid	valid	valid	3	2003/01/03 00:00:00.000	2003	1	3	valid	valid	valid	4	2003/01/04 00:00:00.000	2003	1	4	valid	valid	valid	5	2003/01/05 00:00:00.000	2003	1	5	valid	valid	valid	6	2003/01/06 00:00:00.000	2003	1	6	valid	valid	valid	7	2003/01/07 00:00:00.000	2003	1	7	valid	valid	valid	8	2003/01/08 00:00:00.000	2003	1	8	valid	valid	valid	Step ini digunakan untuk mencari informasi tambahan dari database berdasarkan streamDate dan Jika nilai yang dicari tidak ditemukan dalam
1	2003/01/01 00:00:00.000	2003	1	1																																																																																																							
2	2003/01/02 00:00:00.000	2003	1	2																																																																																																							
3	2003/01/03 00:00:00.000	2003	1	3																																																																																																							
4	2003/01/04 00:00:00.000	2003	1	4																																																																																																							
5	2003/01/05 00:00:00.000	2003	1	5																																																																																																							
6	2003/01/06 00:00:00.000	2003	1	6																																																																																																							
7	2003/01/07 00:00:00.000	2003	1	7																																																																																																							
8	2003/01/08 00:00:00.000	2003	1	8																																																																																																							
1	2003/01/01 00:00:00.000	2003	1	1	valid	valid	valid																																																																																																				
2	2003/01/02 00:00:00.000	2003	1	2	valid	valid	valid																																																																																																				
3	2003/01/03 00:00:00.000	2003	1	3	valid	valid	valid																																																																																																				
4	2003/01/04 00:00:00.000	2003	1	4	valid	valid	valid																																																																																																				
5	2003/01/05 00:00:00.000	2003	1	5	valid	valid	valid																																																																																																				
6	2003/01/06 00:00:00.000	2003	1	6	valid	valid	valid																																																																																																				
7	2003/01/07 00:00:00.000	2003	1	7	valid	valid	valid																																																																																																				
8	2003/01/08 00:00:00.000	2003	1	8	valid	valid	valid																																																																																																				

			database, hasilnya <null>.
Filter rows			Step ini digunakan untuk menyaring (filter) baris berdasarkan kondisi jika tanggal, bulan, tahun serta hari null.
Table Output			Table Output menerima data yang memiliki nilai NULL pada kolom date, year, month, dan day serta data yang memiliki nilai valid (tidak NULL) akan diabaikan dan tidak masuk ke Table Output.

B. Dimensi Pegawai

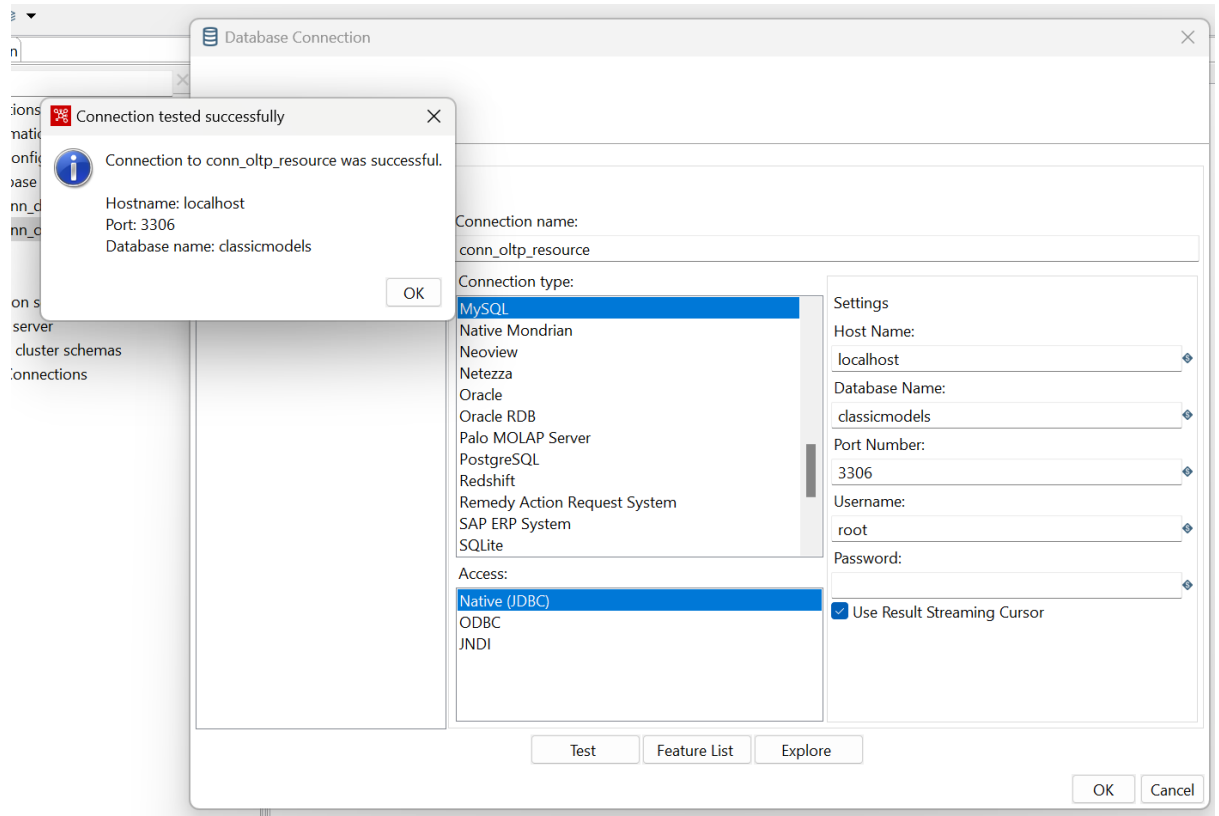
1. Buatlah tabel dimPegawai pada dw_legendVehicle.

```

1 CREATE TABLE dimEmployees(
2     id_dimEmployees int NOT NULL auto_increment PRIMARY KEY,
3     employeeNumber int(11),
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 table dimEmployees



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

```
SELECT * FROM dimemployees e
LEFT JOIN dimemployees r
ON e.employeeNumber = r.employeeNumber;
```

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 ☐

Help OK Cancel

Step name: Select values

Select & Alter Remove Meta-data

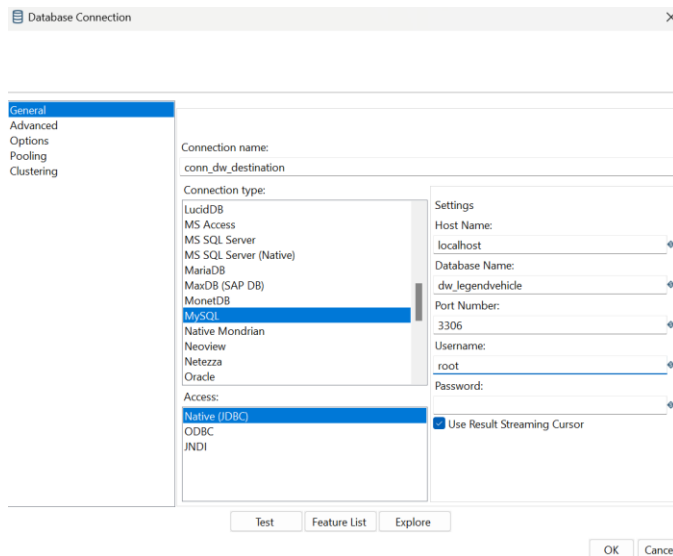
Fields to remove :

#	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

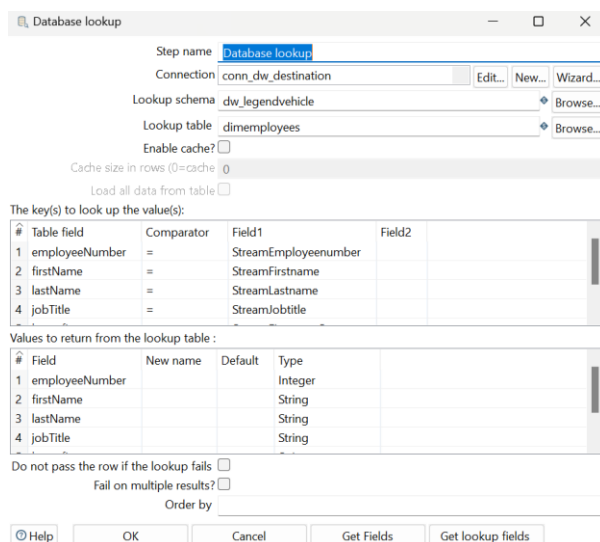
Get fields to remove

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.



11. Hubungkan **output databse** lookup dengan **filter rows**.
12. Pada **filter rows** berikan kondisi field yang **null** pada **field dimemployees** untuk dimasukkan pada proses selanjutnya. Hal itu menandakan bahwa **data stream** belum memiliki kesamaan pada data di **dimemployees**.

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

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

Partition data over tables: ☐

Partitioning field:

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:

Store the tablename field: ☐

Return auto-generated key: ☐

Name of auto-generated key field:

OK Cancel SQL

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

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

First rows Last rows Off

#	StreamEmployeeNumber	StreamLastname	StreamFirstname	StreamJobtitle	StreamLastnameBoss	StreamFirstnameBoss	employeeNumber	firstNarn
1	1002	Murphy	Diane	President	<null>	<null>	<null>	<null>
2	1056	Patterson	Mary	VP Sales	Murphy	Diane	<null>	<null>
3	1076	Firrelli	Jeff	VP Marketing	Murphy	Diane	<null>	<null>
4	1088	Patterson	William	Sales Manager (APAC)	Patterson	Mary	<null>	<null>
5	1102	Bondur	Gerard	Sale Manager (EMEA)	Patterson	Mary	<null>	<null>
6	1143	Bow	Anthony	Sales Manager (NA)	Patterson	Mary	<null>	<null>
7	1165	Jennings	Leslie	Sales Rep	Bow	Anthony	<null>	<null>
8	1166	Thompson	Leslie	Sales Rep	Bow	Anthony	<null>	<null>
9	1188	Firrelli	Julie	Sales Rep	Bow	Anthony	<null>	<null>
1.	1216	Patterson	Steve	Sales Rep	Bow	Anthony	<null>	<null>
1.	1286	Tsenq	Foon Yue	Sales Rep	Bow	Anthony	<null>	<null>

Table Output									
Execution Results									
idempotent Execution History Step Metrics Performance Graph Metrics Previous data									
idempotent Last rows DDL									
StreamEmployeeNumber	StreamLastname	StreamFirstname	StreamJobtitle	StreamLastnameBoss	StreamFirstnameBoss	employeeNumber	StreamEmployeeNumber	StreamLastname	StreamFirstname
1	1002	Murphy	Diane	President	murphy	1002	1002	Murphy	Diane
2	1056	Patterson	Mary	VP Sales	Murphy	1056	1056	Patterson	Mary
3	1076	Firrelli	Jeff	VP Marketing	Murphy	1076	1076	Firrelli	Jeff
4	1088	Patterson	William	Sales Manager (APAC)	Patterson	1088	1088	Patterson	William
5	1102	Bondur	Gerard	Sale Manager (EMEA)	Patterson	1102	1102	Bondur	Gerard
6	1143	Bow	Anthony	Sales Manager (NA)	Patterson	1143	1143	Bow	Anthony
7	1165	Jennings	Leslie	Sales Rep	Bow	1165	1165	Jennings	Leslie
8	1166	Thompson	Leslie	Sales Rep	Bow	1166	1166	Thompson	Leslie
9	1188	Firrelli	Julie	Sales Rep	Bow	1188	1188	Firrelli	Julie
10	1216	Patterson	Steve	Sales Rep	Bow	1216	1216	Patterson	Steve
11	1286	Tseng	Foon Yue	Sales Rep	Bow	1286	1286	Tseng	Foon Yue

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

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














Penjelasan :

Dengan menggunakan data yang sama dan sudah di saring maka tidak akan terdapat redundansi

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

```
1 INSERT INTO employees (
2     employeeNumber, lastName, firstName, extension, email, officeCode, reportsTo, jobTitle
3 ) VALUES (
4     9999, 'Aqila', 'Azza', 'x9999', 'aqilanurazza123@gmail.com', 1, 1002, 'Data Scientist'
5 );
```

➤ Dw_employees

<input type="checkbox"/>	 Edit	 Copy	 Delete	23	1702	Martin	Gerard	Sales Rep	Gerard	Bondur	2025-03-11
<input type="checkbox"/>	 Edit	 Copy	 Delete	24	1002	Diane	Murphy	President	NULL	NULL	2025-03-11
<input type="checkbox"/>	 Edit	 Copy	 Delete	25	1002	Diane	Murphy	President	NULL	NULL	2025-03-11
<input type="checkbox"/>	 Edit	 Copy	 Delete	26	1002	Diane	Murphy	President	NULL	NULL	2025-03-11
<input type="checkbox"/>	 Edit	 Copy	 Delete	27	9999	Azza	Aqila	Data Scientist	Diane	Murphy	2025-03-11

Penjelasan :

Terjadi redudansi setelah insert data.