

DATA WAREHOUSE

JOBSHEET 3

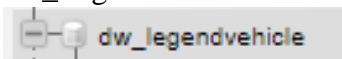
Nama : My Babby Findia R.S

No. Absen : 16

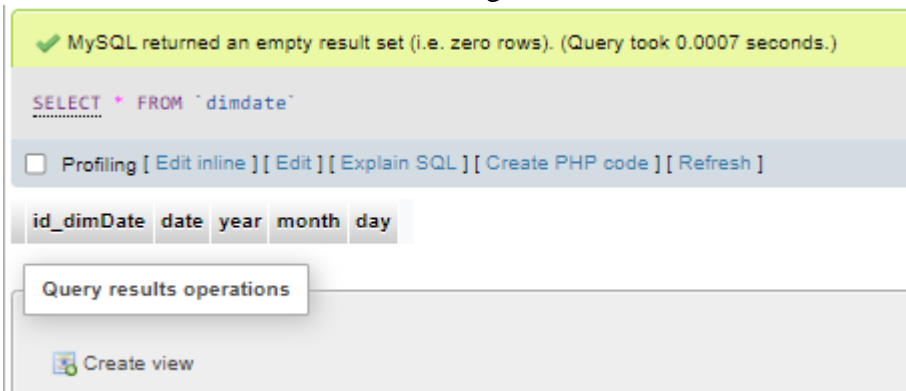
Kelas : SIB-2B

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.



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).



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

Generate rows

Step name:

Limit:

Never stop generating rows: ☐

Interval in ms (delay):

Current row time field name:

Previous row time field name:

Fields:

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

Help OK Preview Cancel

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:

Name of value:

Use a database to generate the sequence

Use DB to get sequence? ☐

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

Schema name: Schemas...

Sequence name: Sequences...

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value:

Increment by:

Maximum value:

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:

☒ Throw an error on non-existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping symbol
1	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			

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 & Alter Remove Meta-data

Fields:

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

Include unspecified fields, ordered by name ☐

Select values

Step name:

Select & Alter Remove Meta-data

Fields to remove:

#	Fieldname
1	CurrentDate
2	incrementDay

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.

Database Connection

General
Advanced
Options
Pooling
Clustering

Connection name:
conn_dw_destination

Connection type:
Infobright
Informix
Ingres
Ingres VectorWise
Intersystems Cache
KingbaseES
LucidDB
MS Access
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_jelegendvehicle
Port Number: 3306
Username: root
Password:
☒ Use Result Streaming Cursor

Test Feature List Explore

OK Cancel


Connection tested successfully

Connection to conn_dw_destination was successful.

Hostname: localhost
Port: 3306
Database name: dw_jelegendvehicle

OK

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 retriive 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	date	=	streamYear	
3	month	=	streamMonth	
4	day	=	streamDay	
5				

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

- 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

(

AND

day IS NULL

null = []

)

AND

)

)

)

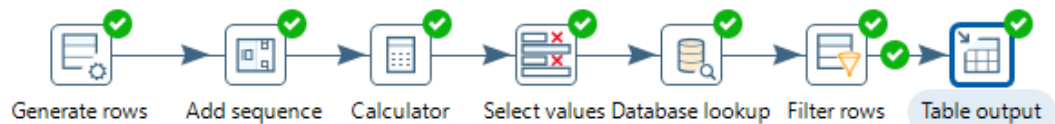
)

Help

OK

Cancel

- [illegible]



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data



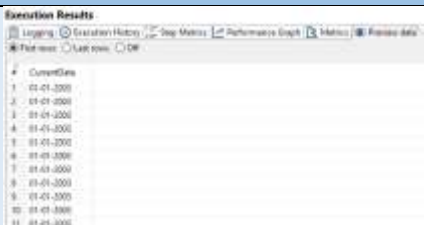

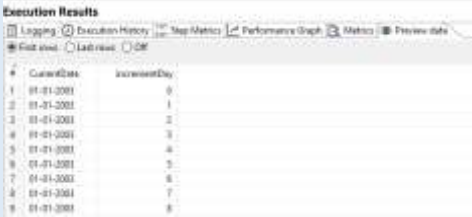

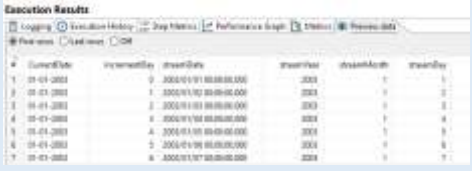
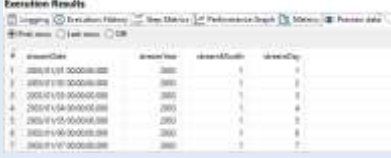
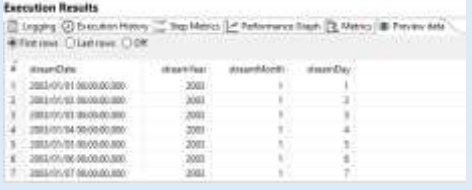


2025/03/10 19:07:44 - Spoon - Transformation opened.
 2025/03/10 19:07:44 - Spoon - Launching transformation [Transformation 1]...
 2025/03/10 19:07:44 - Spoon - Started the transformation execution.
 2025/03/10 19:07:45 - Transformation 1 - Dispatching started for transformation [Transformation 1]
 2025/03/10 19:07:46 - Table output.0 - Connected to database [conn_dw_destination] (commit=1000)
 2025/03/10 19:07:46 - Generate rows.0 - Finished processing (I=0, O=0, R=0, W=1825, U=0, E=0)
 2025/03/10 19:07:46 - Add sequence.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
 2025/03/10 19:07:46 - Calculator.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
 2025/03/10 19:07:46 - Select values.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)
 2025/03/10 19:07:49 - Database lookup.0 - Finished processing (I=0, O=0, R=1825, W=1825, U=0, E=0)





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

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

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			Pada output data menghasilkan sejumlah baris sesuai dengan limit yang ditentukan
Add Sequences			Pada input data kosong dari generate rows maka outputnya menambahkan kolom sequence yang berisi angka bertambah mulai dari 1 hingga jumlah baris yang dihasilkan
Calculator			Pada input data dari add sequence hanya berisi ID dan pada output data menambahkan kolom tanggal yang dihitung dari tanggal awal dengan penambahan squence sebagai offset hari.
Select values			Pada input data semua kolom yang dihasilkan oleh calculator dan pada output hanya kolom yang diilih.
Database lookup			Pada inputan data dari select values maka output data yang telah dicek apakah sudah ada di database

Filter rows			Pada inputan data dari database lookup maka output data hanya menyimpan data yang belum ada dalam tabel untuk menghindari duplikasi.
Table Output			Pada inputan data yang lolos filter data ditulis ke tabel dalam mysql.

B. DIMENSI PEGAWAI

1. Buatlah tabel dimPegawai pada dw_legendVehicle.

✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0015 seconds.)

```
SELECT * FROM `dimemployees`
```

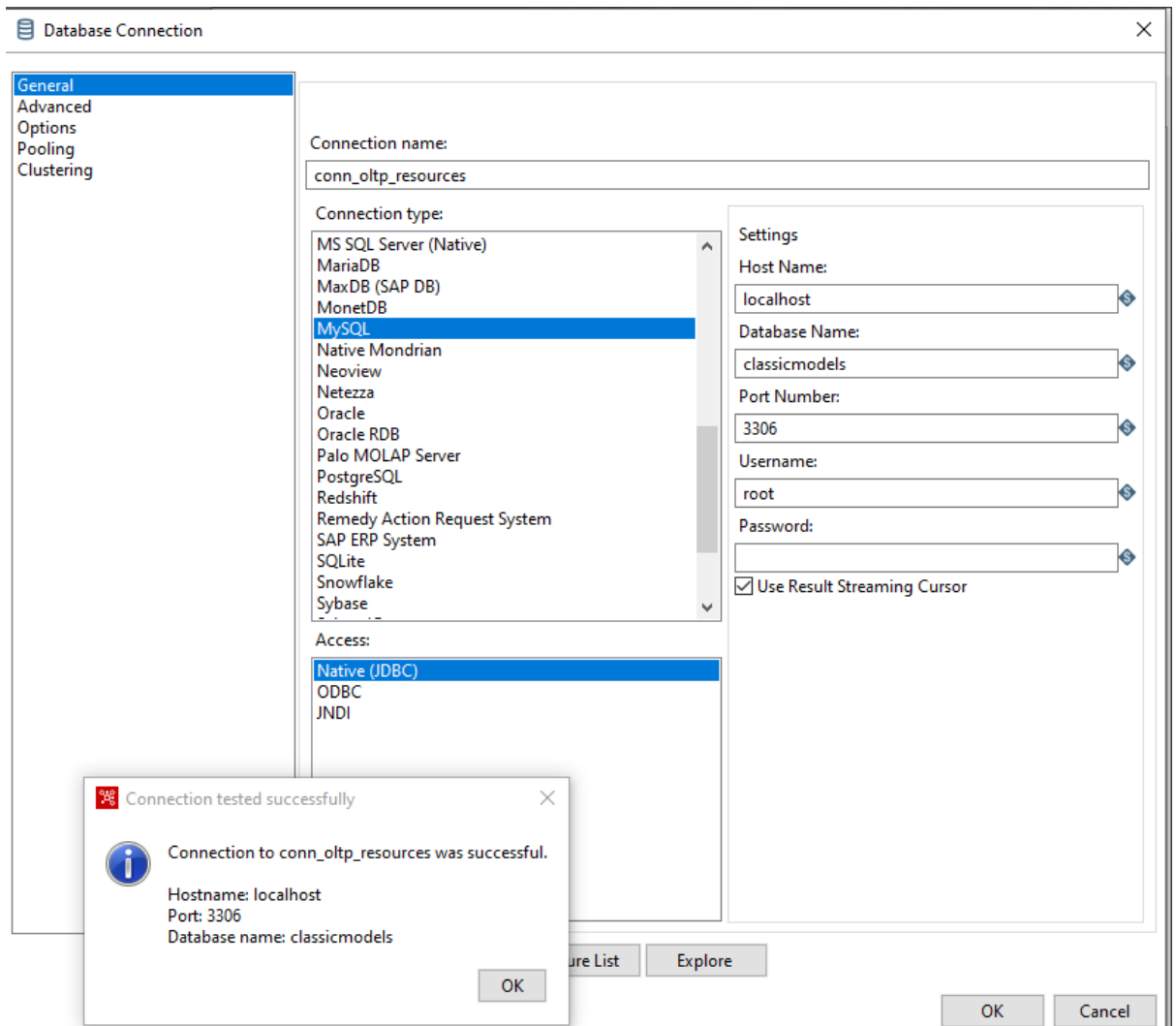
☐ Profiling [[Edit inline](#)] [[Edit](#)] [[Explain SQL](#)] [[Create PHP code](#)] [[Refresh](#)]

id_dimEmployees	employeeNumber	firstName	lastname	jobtitle	boss_firstname	boss_lastname	updated
-----------------	----------------	-----------	----------	----------	----------------	---------------	---------

Query results operations

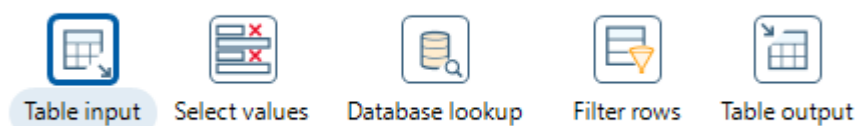
[Create view](#)

2. Pada PDI Spoon buatlah koneksi baru dengan nama conn_oltp_resources yang menghubungkan dengan database oltp. sesuaikan hostname, database name, port number, username dan password dengan keadaan pada device masing-masing.



3. Drag and drop beberapa objek sebagai berikut:

- Table input: digunakan mengambil data dari database OLTP.
- Select values: memilih field yang digunakan untuk proses Transform dan Load.
- Database lookup: digunakan untuk melihat data pada tabel dimEmployees untuk memastikan data tidak kembar
- Filter rows: digunakan untuk memilih data stream yang masih belum ada apada tabel dimEmployees.
- Table output: Memasukkan data ke dalam tavle dimEmployees



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

Table input

Step name: Table input

Connection: conn_oltp_resources Edit... New... Wizard...

SQL Get SQL select statement...

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

Line 1 Column 0

Store column info in step meta ☐

Enable lazy conversion ☐

Replace variables in script? ☐

Insert data from step

Execute for each row? ☐

Limit size: 0

Help OK Preview Cancel

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.

Select values

Step nameSelect values

Select & AlterRemoveMeta-data

Fields :

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

Get fields to selectEdit Mapping

☐ Include unspecified fields, ordered by name

Help

OKCancel

Select values

Step nameSelect values

Select & AlterRemoveMeta-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

OKCancel

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: dimemployees [Browse...]

Enable cache? ☐

Cache size in rows (0=cache): 0

Load all data from table ☐

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

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

Values to return from the lookup table:

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

Send 'true' data to step:

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:

Connection:

Target schema:

Target table:

Commit size:

Truncate table: ☐

Ignore insert errors: ☐

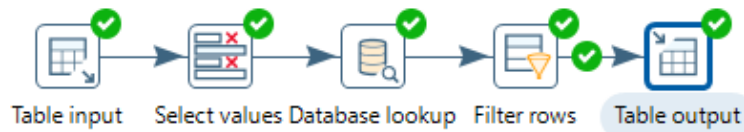
Specify database fields: ☒

Main options Database fields

Fields to insert:

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

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



2025/03/11 16:28:04 - Transformation 2 - Dispatching started for transformation [Transformation 2]
2025/03/11 16:28:05 - Table output.0 - Connected to database [conn_dw_destination] (commit=1000)
2025/03/11 16:28:05 - Table input.0 - Finished reading query, closing connection
2025/03/11 16:28:05 - Table input.0 - Finished processing (I=23, O=0, R=0, W=23, U=0, E=0)
2025/03/11 16:28:05 - Select values.0 - Finished processing (I=0, O=0, R=23, W=23, U=0, E=0)
2025/03/11 16:28:05 - Database lookup.0 - Finished processing (I=0, O=0, R=23, W=23, U=0, E=0)
2025/03/11 16:28:05 - Filter rows.0 - Finished processing (I=0, O=0, R=23, W=23, U=0, E=0)
2025/03/11 16:28:05 - Table output.0 - Finished processing (I=0, O=23, R=23, W=23, U=0, E=0)
2025/03/11 16:28:05 - Spoon - The transformation has finished!!

Showing rows 0 - 22 (23 total, Query took 0.0012 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

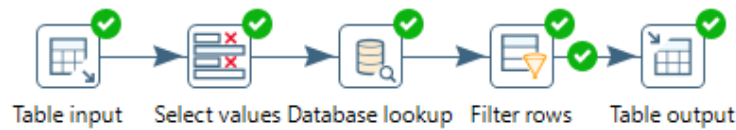
			id_dimEmployees	employeeNumber	firstName	lastName	jobtitle	boss_firstname	boss_lastname	updated
<input type="checkbox"/>	Edit	Copy	Delete	1	1002	Diane	Murphy	President	NULL	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	2	1056	Mary	Patterson	VP Sales	Diane	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	3	1076	Jeff	Firrelli	VP Marketing	Diane	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	4	1088	William	Patterson	Sales Manager (APAC)	Mary	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	5	1102	Gerard	Bondur	Sale Manager (EMEA)	Mary	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	6	1143	Anthony	Bow	Sales Manager (NA)	Mary	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	7	1165	Leslie	Jennings	Sales Rep	Anthony	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	8	1166	Leslie	Thompson	Sales Rep	Anthony	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	9	1188	Julie	Firrelli	Sales Rep	Anthony	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	10	1216	Steve	Patterson	Sales Rep	Anthony	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	11	1286	Foon Yue	Tseng	Sales Rep	Anthony	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	12	1323	George	Vanauf	Sales Rep	Anthony	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	13	1337	Loui	Bondur	Sales Rep	Gerard	2025-03-11 16:28:05
<input type="checkbox"/>	Edit	Copy	Delete	14	1370	Gerard	Hernandez	Sales Rep	Gerard	2025-03-11 16:28:05

TUGAS 2

1. Buka preview tab pada execution result area di setiap proses object. amati input dan output data yang ada. bandingkan di setiap prosesnya. jelaskan perbedaan di setiap prosesnya.

Proses Objek	SS data input	SS data output	Keterangan
Table Input			Mengambil data langsung dari database sumber
Select values			Memilih kolom yang dibutuhkan dan mengganti nama field
Database lookup			Melakukan lookup data dari tabel referensi misalnya mencari nama atasan dari reportsTo
Filter rows			Memfilter data berdasarkan kondisi tertentu
Table Output			Menyimpan hasil transformasi ke database legendvehicle

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



Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025/03/12 16:02:49 - Transformation 2 - Dispatching started for transformation [Transformation 2]
2025/03/12 16:02:49 - Table output.0 - Connected to database [conn_dw_destination] (commit=1000)
2025/03/12 16:02:49 - Table input.0 - Finished reading query, closing connection
2025/03/12 16:02:49 - Table input.0 - Finished processing (I=23, O=0, R=0, W=23, U=0, E=0)
2025/03/12 16:02:49 - Select values.0 - Finished processing (I=0, O=0, R=23, W=23, U=0, E=0)
2025/03/12 16:02:49 - Database lookup.0 - Finished processing (I=22, O=0, R=23, W=23, U=0, E=0)
2025/03/12 16:02:49 - Filter rows.0 - Finished processing (I=0, O=0, R=23, W=1, U=0, E=0)
2025/03/12 16:02:49 - Table output.0 - Finished processing (I=0, O=1, R=1, W=1, U=0, E=0)
2025/03/12 16:02:49 - Spoon - The transformation has finished!!

- Iya, bisa redundant tidak memiliki mekanisme untuk menghindari duplikasi data. Jika table output langsung menambahkan data tanpa validasi.
3. Tambahkan nama anda pada table employee di OLTP. jalankan kembali transformasi ini. Amati hasilnya, apa yang terjadi?

	id_dimEmployees	employeeNumber	firstName	lastname	jobtitle	boss_firstname	boss_lastname	updated
		27	0000	Babby	Findia	Software Engineer	John Doe	2025-03-12 16:11:55