


<p>Nama: Adam Zannuba</p> <p>NIM: 065001900022</p>	 <p>Praktikum Data Warehouse</p>	<p>MODUL 6</p> <p>Nama Dosen: Ir. Teddy Siswanto, MMSi</p>
<p>Hari/Tanggal: 19 May 2022</p>		<p>Nama Asisten Labratorium: 1. Azhar Rizki Zulma 065001900001 2. Nadiya Amanda Rizkania 064001900003</p>

Merge Data pada Spoon

1. Teori Singkat

Data warehouse adalah jenis sistem manajemen data yang dirancang untuk memungkinkan dan mendukung kegiatan business intelligence (BI), terutama analitik. Gudang data semata-mata dimaksudkan untuk melakukan kueri dan analisis dan sering berisi sejumlah besar data historis. Data dalam gudang data biasanya berasal dari berbagai sumber seperti file log aplikasi dan aplikasi transaksi. Gudang data memusatkan dan mengkonsolidasikan sejumlah besar data dari berbagai sumber. Kemampuan analitisnya memungkinkan organisasi untuk memperoleh wawasan bisnis yang berharga dari data mereka untuk meningkatkan pengambilan keputusan. Seiring waktu, ia membangun catatan sejarah yang dapat sangat berharga bagi para ilmuwan data dan analis bisnis. Karena kemampuan ini, gudang data dapat dianggap sebagai "sumber kebenaran tunggal" organisasi.

2. Alat dan Bahan

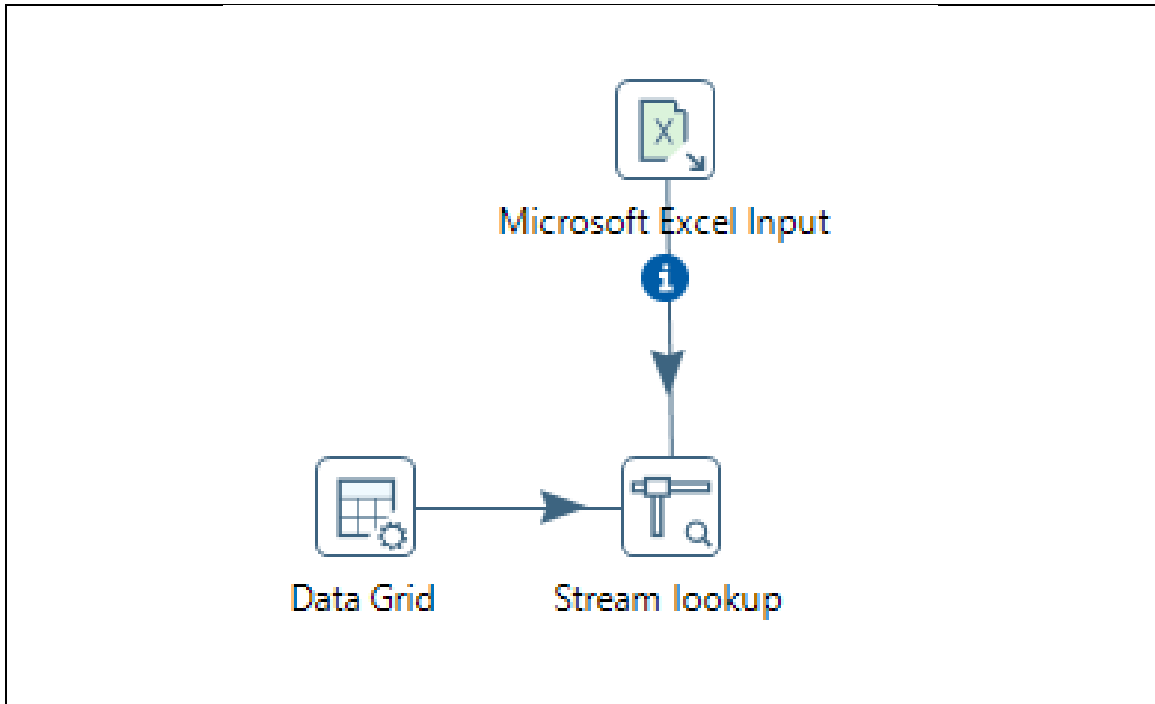
Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

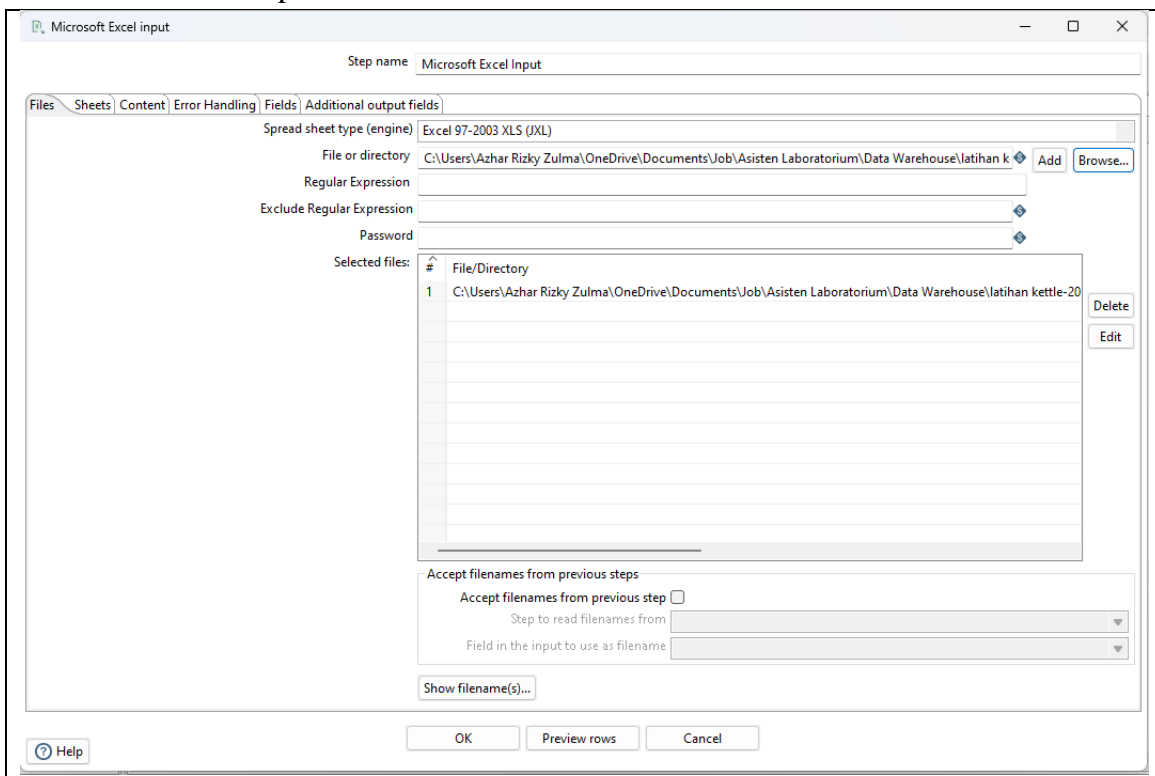


3. Elemen Kompetensi

- a. Latihan pertama – Melakukan penggabungan data menggunakan Stream Lookup
 1. Buat transformation baru dengan struktur seperti pada gambar dibawah ini.



2. Microsoft Excel Input.



Microsoft Excel input

Step name Microsoft Excel Input

Files Sheets Content Error Handling Fields Additional output fields

List of sheets to read

#	Sheet name	Start row	Start column
1	Sheet1	0	0

Get sheetname(s)...

Help OK Preview rows Cancel

Microsoft Excel input

Step name Microsoft Excel Input

Files Sheets Content Error Handling Fields Additional output fields

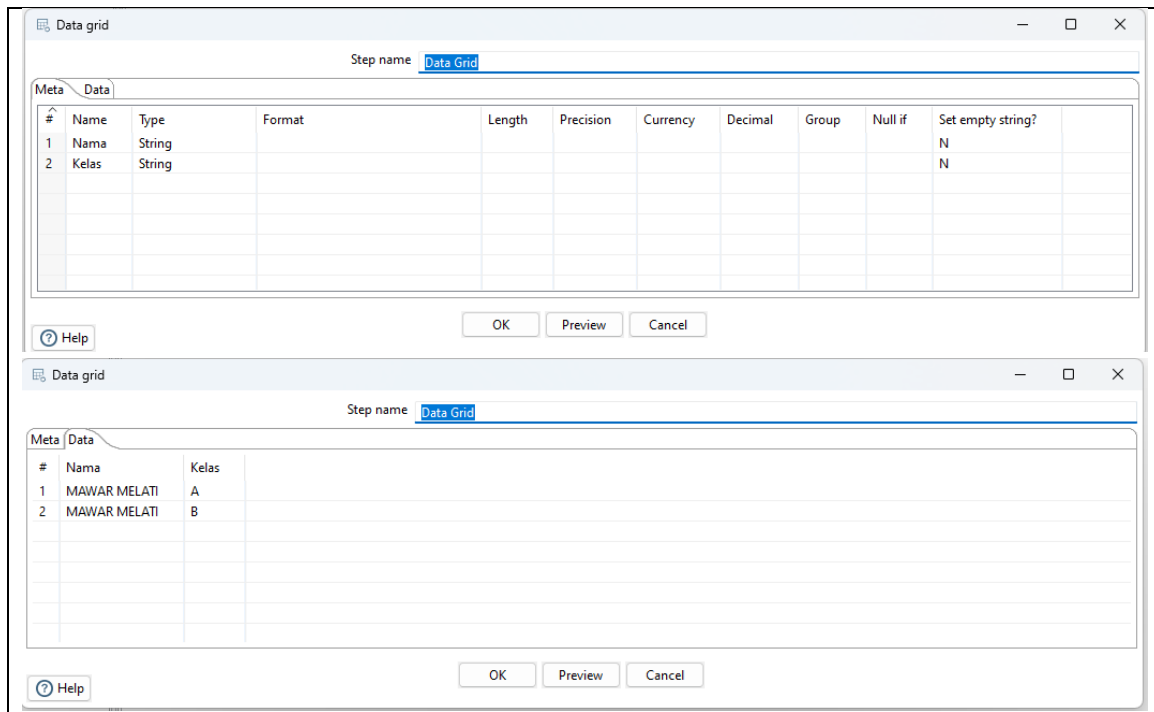
#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	NIS	String	-1	-1	none	N				
2	Nama	String	-1	-1	none	N				
3	Jenis Kelamin	String	-1	-1	none	N				
4	Kelas	String	-1	-1	none	N				

Get fields from header row...

Help OK Preview rows Cancel



3. Data Grid.



Step name: Data Grid

Meta Data

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Null if	Set empty string?
1	Nama	String								N
2	Kelas	String								N

Help OK Preview Cancel

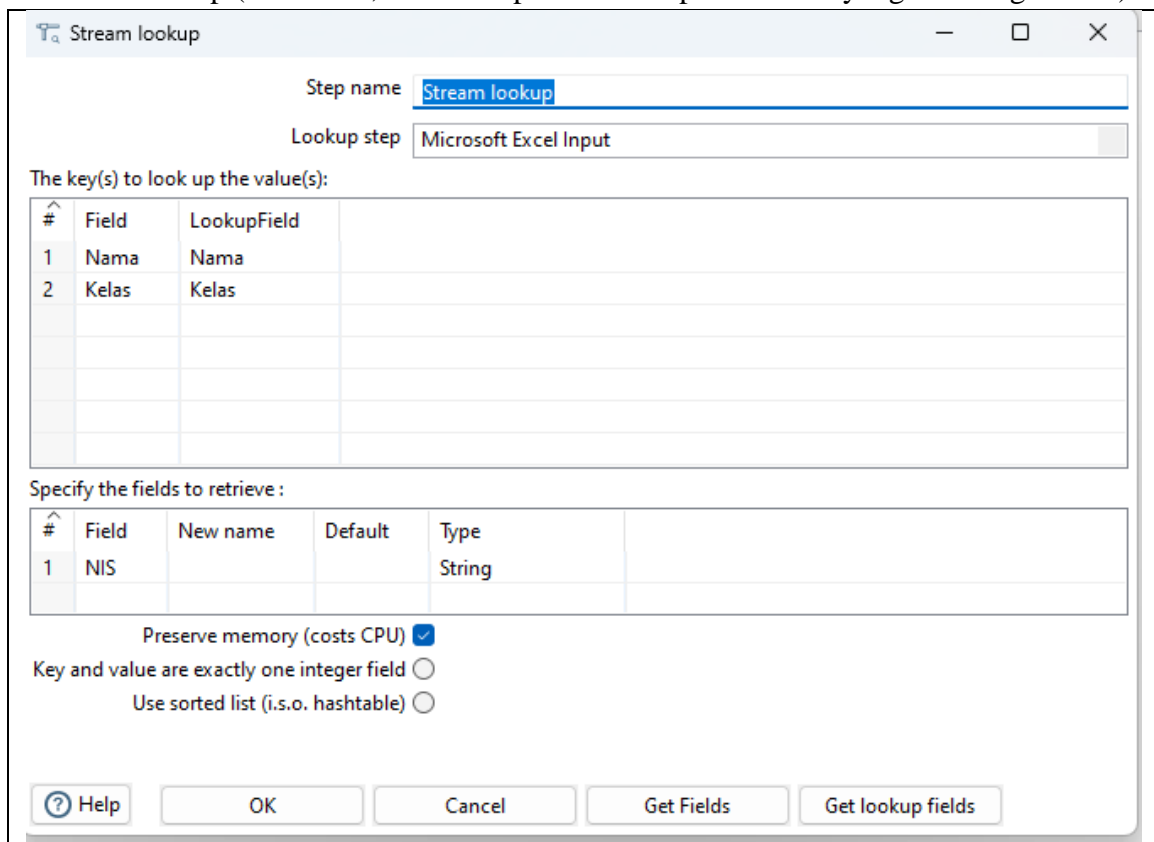
Step name: Data Grid

Meta Data

#	Nama	Kelas
1	MAWAR MELATI	A
2	MAWAR MELATI	B

Help OK Preview Cancel

4. Stream lookup (Get fields, Get lookup fields & hapus variabel yang tidak digunakan).



Stream lookup

Step name: Stream lookup

Lookup step: Microsoft Excel Input

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

#	Field	LookupField
1	Nama	Nama
2	Kelas	Kelas

Specify the fields to retrieve :

#	Field	New name	Default	Type
1	NIS			String

Preserve memory (costs CPU) ☒

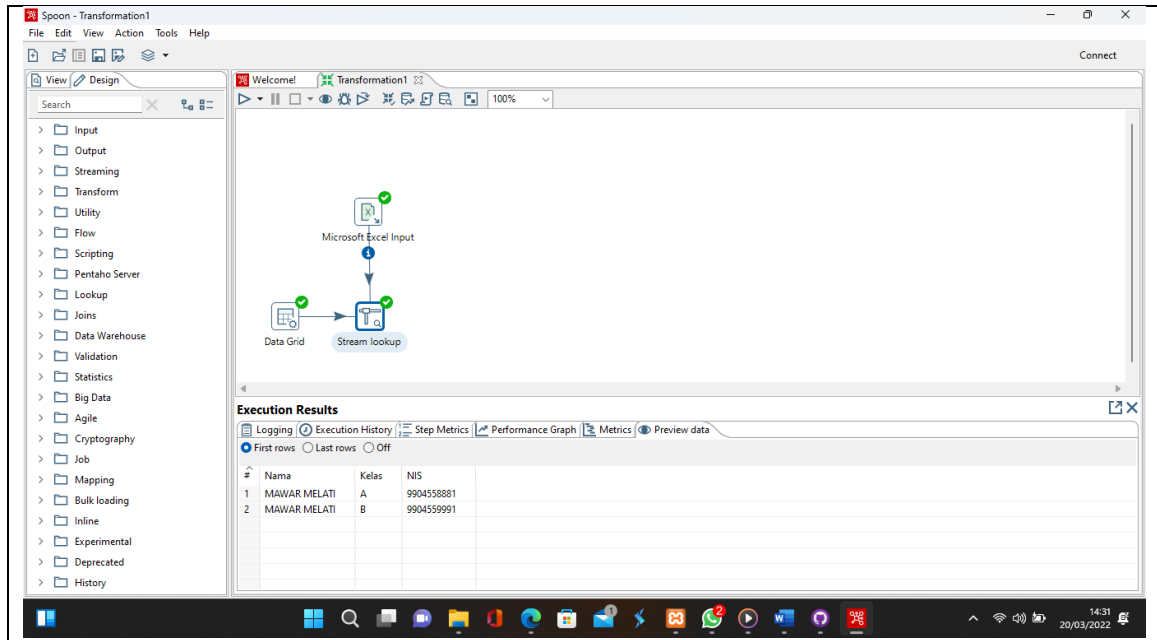
Key and value are exactly one integer field ☐

Use sorted list (i.s.o. hashtable) ☐

Help OK Cancel Get Fields Get lookup fields

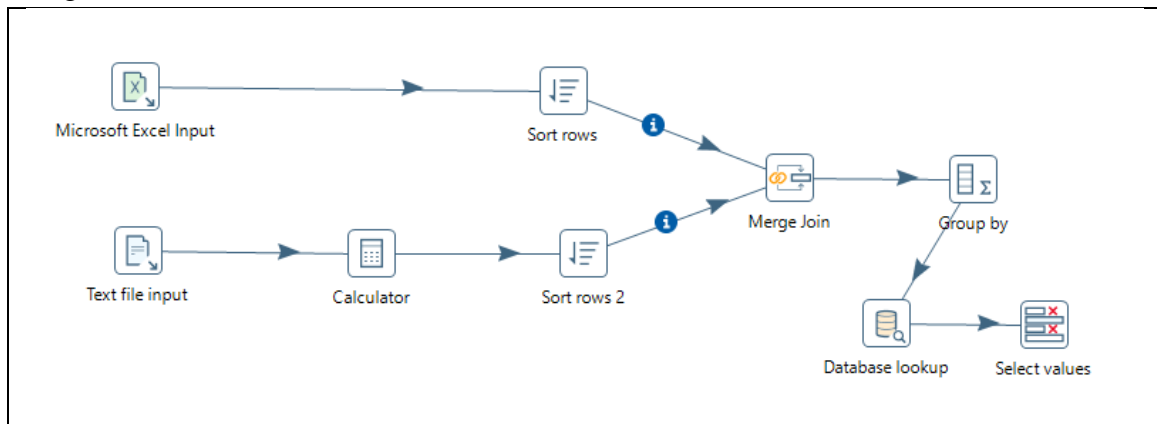


5. Ketika dijalankan maka outputnya akan seperti gambar berikut ini.



b. Latihan Kedua – Penggabungan dengan Merge dan Database Keylookup

1. Nyalakan Apache dan MySQL pada XAMPP dan buat transformasi baru dengan struktur sebagai berikut.



2. Microsoft Excel Input (Get sheetsname dan Get fields).

Microsoft Excel input

Step name: Microsoft Excel Input

Files | Sheets | Content | Error Handling | Fields | Additional output fields

Spread sheet type (engine): Excel 97-2003 XLS (JXL)

File or directory: C:\Users\Azhar Rizky Zulma\OneDrive\Documents\Job\Asisten Laboratorium\Data Warehouse\latihan k

Regular Expression:

Exclude Regular Expression:

Password:

Selected files:

#	File/Directory
1	C:\Users\Azhar Rizky Zulma\OneDrive\Documents\Job\Asisten Laboratorium\Data Warehouse\latihan kettle-20

Accept filenames from previous steps

Accept filenames from previous step ☐

Step to read filenames from:

Field in the input to use as filename:

Show filename(s)...

OK Preview rows Cancel

Microsoft Excel input

Step name: Microsoft Excel Input

Files | Sheets | Content | Error Handling | Fields | Additional output fields

List of sheets to read

#	Sheet name	Start row	Start column
1	Sheet1	0	0

Get sheetname(s)...

OK Preview rows Cancel



Microsoft Excel input

Step name: Microsoft Excel Input

Files | Sheets | Content | Error Handling | Fields | Additional output fields

#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	orderNumber	Integer	-1	-1	none	N				
2	orderDate	Date	-1	-1	none	N				
3	requiredDate	Date	-1	-1	none	N				
4	shippedDate	Date	-1	-1	none	N				
5	status	String	-1	-1	none	N				
6	customerNumber	Integer	-1	-1	none	N				
7	comments	String	-1	-1	none	N				

Get fields from header row...

Help OK Preview rows Cancel

3. Text file input (Get fields pada tab fields).

Text file input

Step name: Text file input

File | Content | Error Handling | Filters | Fields | Additional output fields

File or directory: \${Internal.Entry.Current.Directory}/orderdetails.txt

Regular Expression:

Exclude Regular Expression:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Require
1	\${Internal.Entry.Current.Directory}/orderdetails.txt			N

Accept filenames from previous steps

Accept filenames from previous step ☐

Pass through fields from previous step ☐

Step to read filenames from:

Field in the input to use as filename:

Show filename(s)... Show file content Show content from first data line

Help OK Preview rows Cancel



Text file input

Step name Text file input

File Content Error Handling Filters Fields Additional output fields

#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type	Repeat
1	orderNumber	Integer	#		15	0	\$.	,	-		none	N
2	productCode	String			9		\$.	,	-		none	N
3	quantityOrdered	Number	#		15	0	\$.	,	-		none	N
4	priceEach	Number	##		15	2	\$.	,	-		none	N
5	orderLineNumber	Integer	#		15	0	\$.	,	-		none	N

Get Fields Minimal width

Help OK Preview rows Cancel

4. Calculator.

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 symbol	Currency sy
1	Total	A * B	quantityOrdered	priceEach		Number			N	##			

Help OK Cancel



6. Sort rows 2 (Get fields 2).

Step name: Sort rows 2

Sort directory: %%java.io.tmpdir%% Browse...

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

Only pass unique rows? (verifies keys only) ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	orderNumber	Y	N	N	2	N

Help OK Cancel Get Fields

7. Merge join.

Step name: Merge Join

First Step: Sort rows

Second Step: Sort rows 2

Join Type: INNER

Keys for 1st step:

#	Key field
1	orderNu...

Keys for 2nd step:

#	Key field
1	orderNu...

Get key fields Get key fields

Help OK Cancel



8. Group by (Get fields & Get lookup fields).

Group by

Step name

Group by

Include all rows?

☐

Temporary files directory

%%java.io.tmpdir%%

Browse...

TMP-file prefix

grp

Add line number, restart in each group

☐

Line number field name

Always give back a result row

☐

The fields that make up the group:

#	Group field
1	orderNumber

Get Fields

Aggregates :

#	Name	Subject	Type
1	customerNumber	customerNumber	First value
2	Total	Total	Sum

Get lookup fields

Help

OK

Cancel

9. Database lookup (Database connection, Database explorer)

The image displays a sequence of three database-related windows:

- Database Connection:** The 'General' tab is active. The 'Connection name' is 'MySQL-ClassicalModels 3'. The 'Connection type' is 'MySQL'. The 'Settings' section shows 'Host Name' as 'localhost', 'Database Name' as 'classicalmodels', 'Port Number' as '3306', 'Username' as 'root', and 'Password' as '3306'. The 'Use Result Streaming Cursor' checkbox is checked. A 'Test' button is at the bottom.
- Connection tested successfully:** A small dialog box indicating the connection was successful. It lists: Hostname: localhost, Port: 3306, Database name: classicalmodels.
- Database lookup:** The 'Step name' is 'Database lookup'. The 'Connection' is 'MySQL-ClassicalModels 3'. The 'Lookup schema' is empty, and the 'Lookup table' is 'customers'. The 'Enable cache?' checkbox is checked. The 'Cache size in rows (0=cache)' is 0.
- Database Explorer:** A tree view showing the database structure. Under 'MySQL-ClassicalModels 3', the 'Tables' folder is expanded, and 'customers' is selected. Other tables listed include employees, offices, orderdetails, orders, payments, productlines, and products. Views and Synonyms are also listed.



Database lookup

Step name: Database lookup

Connection: MySQL-ClassicModels 3 [Edit...] [New...] [Wizard...]

Lookup schema: classicmodels [Browse...]

Lookup table: customers [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	customerNumber	=	customerNumber	

Values to return from the lookup table:

#	Field	New name	Default	Type
1	customerName			String

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

Order by:

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

8. Select values (Get fields to select, Get fields to change).

Select values

Step name: Select values

Select & Alter [Remove] [Meta-data]

Fields:

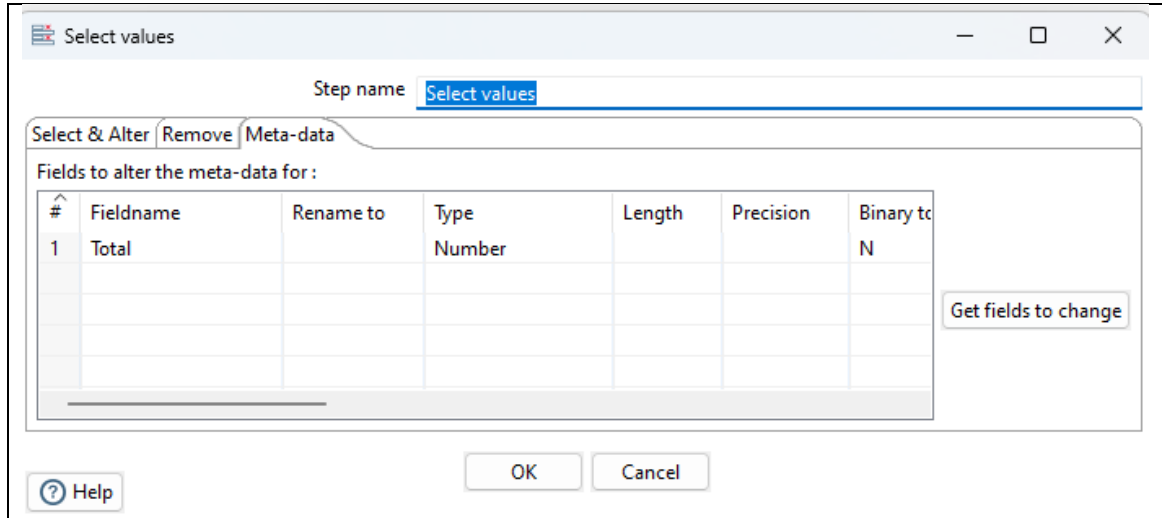
#	Fieldname	Rename to	Length	Precision
1	orderNumber			
2	customerNumber			
3	customerName			
4	Total			

[Get fields to select] [Edit Mapping]

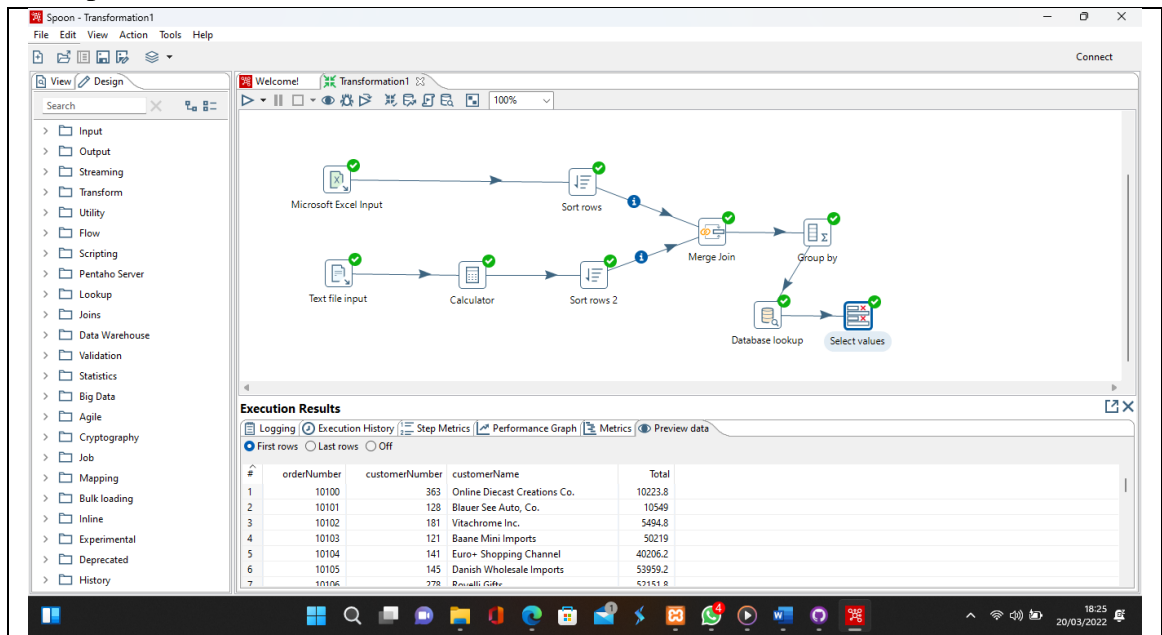
Include unspecified fields, ordered by name ☐

[?] Help [OK] [Cancel]





9. Output



4. File Praktikum

Github Repository:

5. Soal Latihan

Soal:

1. Apa yang dimaksud dengan Merge Data?
2. Apa tujuan dan fungsi dari menggabungkan data?



Jawaban:

- 1.
- 2.

6. Kesimpulan

- a. Dalam pengerjaan praktikum Data Warehouse, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
- b. Kita dapat mengetahui...

7. Cek List (✓)

8.

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Latihan Pertama	...	
2.	Latihan Kedua	...	

9. Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Latihan Pertama	... Menit	...
2.	Latihan Kedua	... Menit	...

Keterangan:

1. Menarik
2. Baik
3. Cukup
4. Kurang

