



# **PRAKTIKUM DATA WAREHOUSE**

## **LAPORAN TUGAS 2 : TUTORIAL ETL DATASET**

By

**MOCH. TORIQUIL MUCHLISSIN**

**2 - D4 SD A | 3321600001**

**PROGRAM STUDI D4 SAINS DATA TERAPAN  
POLITEKNIK ELEKTRONIKA NEGERI SURABAYA**

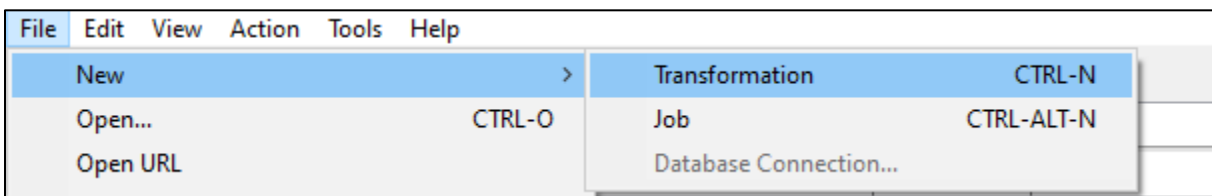
**2023**

## Task Pertama: Ekspor CSV ke XML

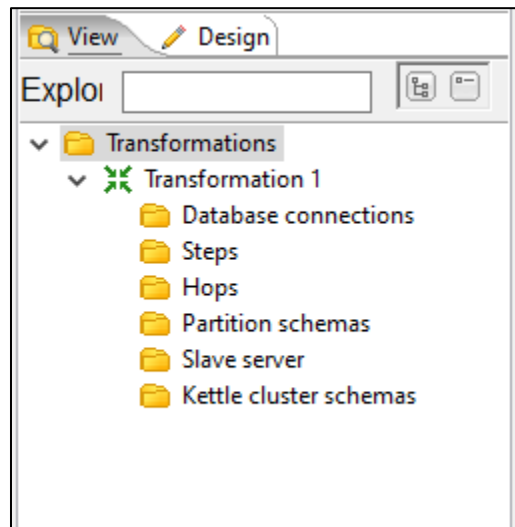
1. Siapkan file csv yang akan digunakan untuk ekspor file ke xml. File yang saya gunakan adalah data **e-commerce** yang didapatkan dari kaggle untuk melakukan percobaan ini

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ID	Warehouse	Mode_of	Customer	Customer	Cost_of_t	Prior_pur	Product_i	Gender	Discount	Weight_ir	Reached.on	Time_Y.N		
2	1	D	Flight	4	2	177	3	low	F	44	1233	1			
3	2	F	Flight	4	5	216	2	low	M	59	3088	1			
4	3	A	Flight	2	2	183	4	low	M	48	3374	1			
5	4	B	Flight	3	3	176	4	medium	M	10	1177	1			
6	5	C	Flight	2	2	184	3	medium	F	46	2484	1			
7	6	F	Flight	3	1	162	3	medium	F	12	1417	1			
8	7	D	Flight	3	4	250	3	low	F	3	2371	1			
9	8	F	Flight	4	1	233	2	low	F	48	2804	1			
10	9	A	Flight	3	4	150	3	low	F	11	1861	1			
11	10	B	Flight	3	2	164	3	medium	F	29	1187	1			
12	11	C	Flight	3	4	189	2	medium	M	12	2888	1			
13	12	F	Flight	4	5	232	3	medium	F	32	3253	1			
14	13	D	Flight	3	5	198	3	medium	F	1	3667	1			
15	14	F	Flight	4	4	275	3	high	M	29	2602	1			
16	15	A	Flight	4	3	152	3	low	M	43	1009	1			
17	16	B	Flight	4	3	227	3	low	F	45	2707	1			
18	17	C	Flight	3	4	143	2	medium	F	6	1194	1			
19	18	F	Ship	5	5	227	3	medium	M	36	3952	1			
20	19	D	Ship	5	5	239	3	high	M	18	2495	1			
21	20	F	Ship	4	5	145	3	medium	M	45	1059	1			
22	21	A	Ship	3	3	161	2	medium	F	38	1521	1			
23	22	B	Ship	3	1	232	4	medium	F	51	2899	1			

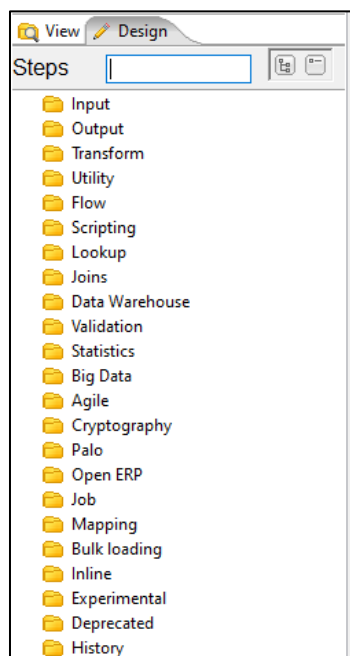
2. Selanjutnya jalankan file **spoon.bat**, kemudian buat **Transformation** melalui menu **File → New → Transformation** atau bisa menggunakan shortcut (**CTRL-N**).



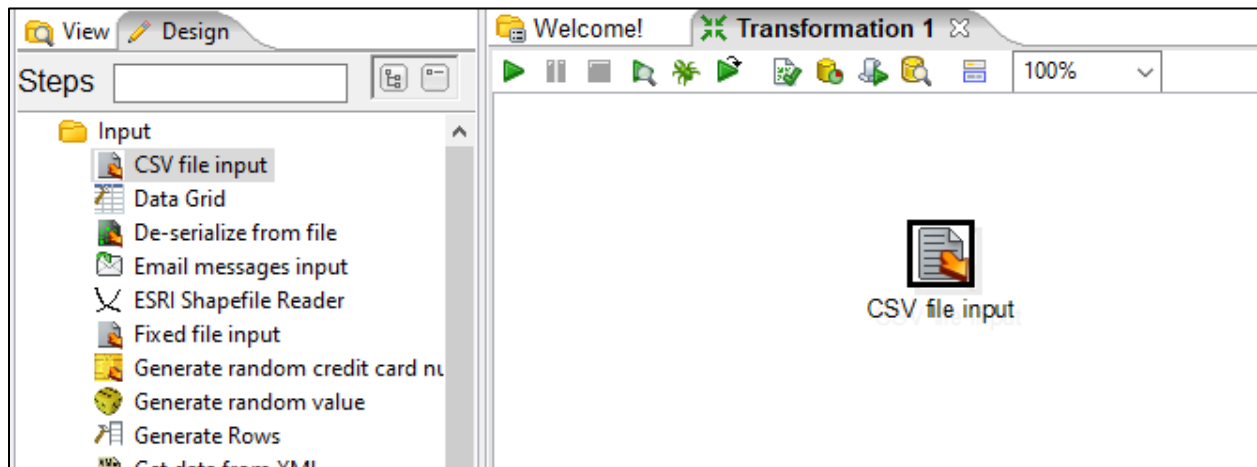
3. Pada panel disebelah kiri (section **View**) adalah tampilan dari folder transformation yang telah dibuat.



4. Kemudian pada panel disebelah kiri (section **Design**) adalah tampilan pilihan step yang bisa digunakan. Step ini merupakan elemen penyusun transformasi, yang dapat berupa input, proses atau output.



5. Buka folder Input kemudian pilih **CSV file input**. Setelah itu *drag* ikon ke bagian tengah.



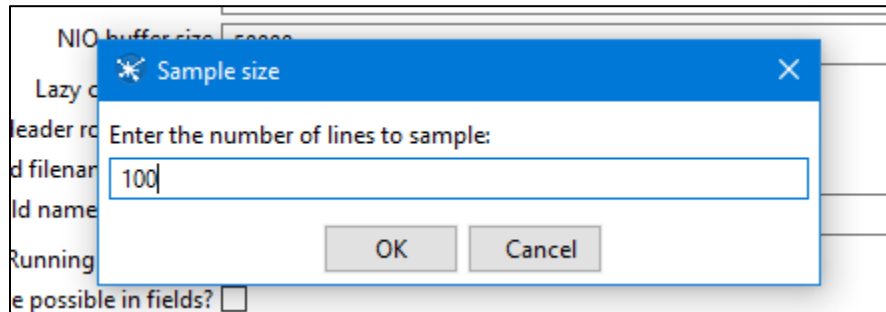
5. Double click CSV file yang sudah *didrag*. Setelah itu, akan muncul kotak seperti gambar dibawah. Isilah **Filename** dengan direktori data CSV yang digunakan dan atur **delimiter** sesuai format data. Apabila CSV memiliki header centanglah kotak **Header row present?**.

Step name	CSV file input
Filename	D:\Kuliah\Semester 4\Data Warehouse\Dataset\E-Commerce Shipping Data.csv
Delimiter	;
Enclosure	"
NIO buffer size	50000
Lazy conversion?	<input checked="" type="checkbox"/>
Header row present?	<input checked="" type="checkbox"/>
Add filename to result	<input type="checkbox"/>
The row number field name (optional)	
Running in parallel?	<input type="checkbox"/>
New line possible in fields?	<input type="checkbox"/>
File encoding	

6. Pilih **Get Fields**

OK	Get Fields	Preview	Cancel
----	------------	---------	--------

7. Selanjutnya, akan muncul kotak **Sample Size** yang akan menanyakan berapa inputan data yang akan user masukkan untuk mendapatkan field



8. Setelah field dibangkitkan, edit sesuai dengan nama field yang cocok dengan lengthnya gambar bawah.

#	^	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1		ID	Integer	#	15		IDR	.	,	none
2		Warehouse_block	String		1		IDR	.	,	none
3		Mode_of_Shipment	String		6		IDR	.	,	none
4		Customer_care_calls	Integer	#	15		IDR	.	,	none
5		Customer_rating	Integer	#	15		IDR	.	,	none
6		Cost_of_the_Product	Integer	#	15		IDR	.	,	none
7		Prior_purchases	Integer	#	15		IDR	.	,	none
8		Product_importance	String		6		IDR	.	,	none
9		Gender	String		1		IDR	.	,	none
10		Discount_offered	Integer	#	15		IDR	.	,	none
11		Weight_in_gms	Integer	#	15		IDR	.	,	none
12		Reached.on.Time_Y.N	Integer	#	15		IDR	.	,	none

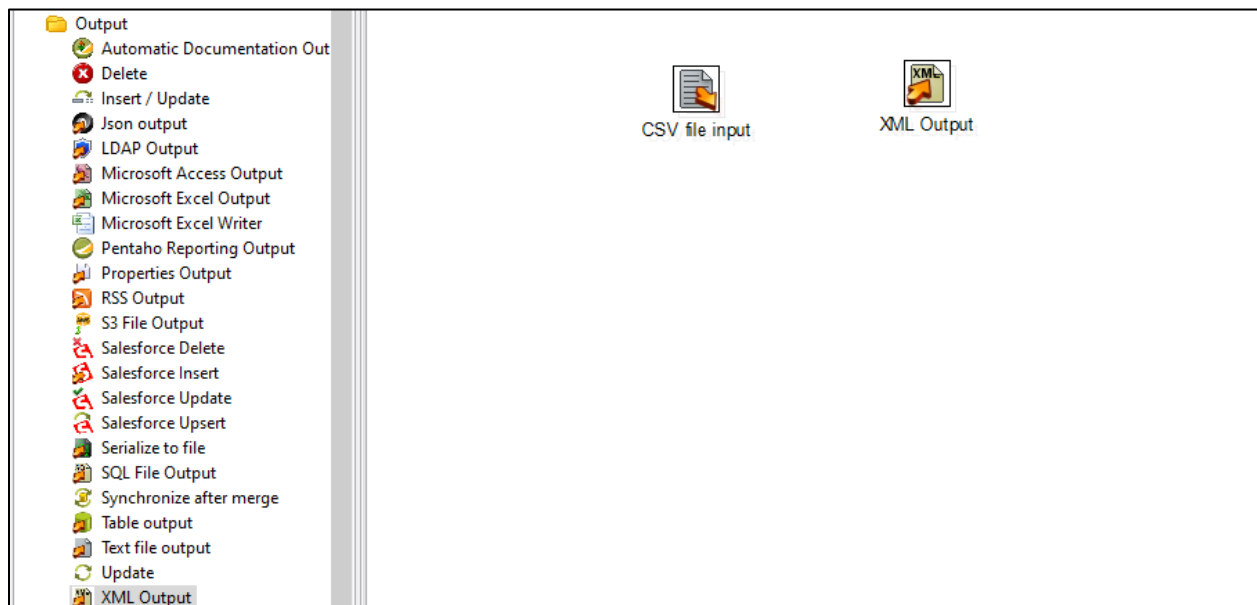
9. Click **Preview** untuk melihat keluaran proses loading. Setelah itu tekan OK

Examine preview data

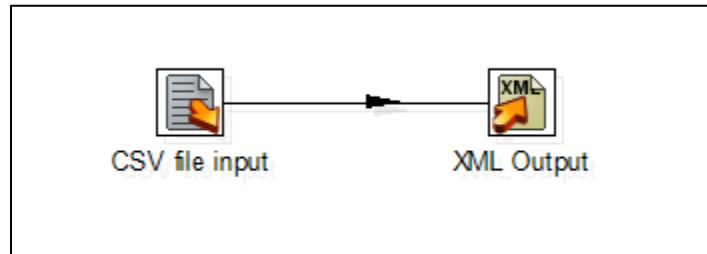
Rows of step: CSV file input (100 rows)

#	ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Product_importance	Gender	Discount_offered	V
1	1	D	Flight	4	2	177	3	low	F	44	
2	2	F	Flight	4	5	216	2	low	M	59	
3	3	A	Flight	2	2	183	4	low	M	48	
4	4	B	Flight	3	3	176	4	medium	M	10	
5	5	C	Flight	2	2	184	3	medium	F	46	
6	6	F	Flight	3	1	162	3	medium	F	12	
7	7	D	Flight	3	4	250	3	low	F	3	
8	8	F	Flight	4	1	233	2	low	F	48	
9	9	A	Flight	3	4	150	3	low	F	11	
10	10	B	Flight	3	2	164	3	medium	F	29	
11	11	C	Flight	3	4	189	2	medium	M	12	
12	12	F	Flight	4	5	232	3	medium	F	32	
13	13	D	Flight	3	5	198	3	medium	F	1	
14	14	F	Flight	4	4	275	3	high	M	29	
15	15	A	Flight	4	3	152	3	low	M	43	
16	16	B	Flight	4	3	227	3	low	F	45	
17	17	C	Flight	3	4	143	2	medium	F	6	
18	18	F	Ship	5	5	227	3	medium	M	36	
19	19	D	Ship	5	5	239	3	high	M	18	
20	20	F	Ship	4	5	145	3	medium	M	45	
21	21	A	Ship	3	3	161	2	medium	F	38	
22	22	B	Ship	3	1	232	4	medium	F	51	
23	23	C	Ship	2	5	156	2	low	M	2	
24	24	F	Ship	4	3	211	3	high	M	12	
25	25	D	Ship	4	5	251	2	medium	F	28	
26	26	F	Ship	3	1	225	4	low	M	29	
27	27	A	Ship	4	1	172	3	high	F	24	
28	28	B	Ship	5	1	162	3	medium	M	31	
29	29	C	Ship	2	3	234	4	low	M	44	
30	30	F	Ship	5	4	183	2	low	F	36	

10. Kemudian pilih folder **Output** → **XML Output**. Setelah itu, *drag* ikon ke bagian tengah.



11. Selanjutnya, kita membuat garis **Hop** yang merupakan garis penghubung antara CSV file input dan XML Output. Saat melakukan step ini **klik file + SHIFT** dan arahkan panah penghubung ke arah file XML



12. Double click file XML kemudian isi **Filename** dengan direktori yang akan digunakan untuk menyimpan hasil konversi file ke format XML

Step name: XML Output

Filename: D:\Kuliah\Semester 4\Data Warehouse\Dataset\E-Commerce Shipping Data.xml Browse...

Do not create file at start: ☐

Pass output to servlet: ☐

Extension: xml

Include stepnr in filename?: ☐

Include date in filename?: ☐

Include time in filename?: ☐

Specify Date time format: ☐

Date time format:

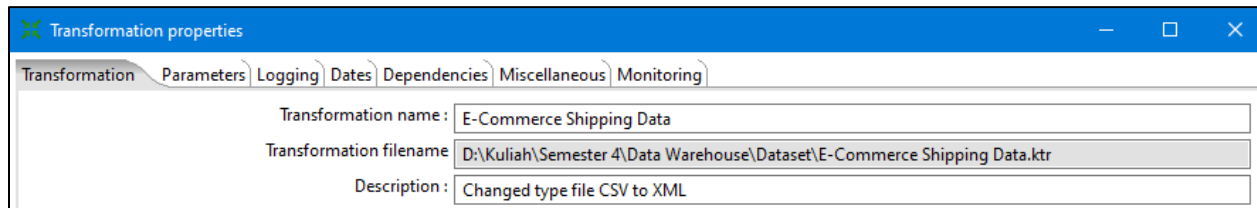
Show filename(s)... ...

Add filenames to result: ☐

13. Pilih tab **Fields** (seperti gambar di bawah) lalu pilih **Get Fields**. Fields akan terisi. isi jenis content type, lalu pilih OK.

#	Fieldname	Element name	Content type	Type	Format	Length	Precision	Currency	Decimal	Group	Null
1	ID		Element	Integer		15	0				
2	Warehouse_block		Element	String		1					
3	Mode_of_Shipment		Element	String		6					
4	Customer_care_calls		Element	Integer		15	0				
5	Customer_rating		Element	Integer		15	0				
6	Cost_of_the_Product		Element	Integer		15	0				
7	Prior_purchases		Element	Integer		15	0				
8	Product_importance		Element	String		6					
9	Gender		Element	String		1					
10	Discount_offered		Element	Integer		15	0				
11	Weight_in_gms		Element	Integer		15	0				
12	Reached.on.Time_Y.N		Element	Integer		15	0				

14. Sebelum menjalankan ekstraksi file. Simpan file Transformasi terlebih dahulu (**CRTL + S** atau menggunakan menu **File**). Isilah nama file transformasi dan direktori yang diinginkan



Transformation properties

Transformation name: E-Commerce Shipping Data

Transformation filename: D:\Kuliah\Semester 4\Data Warehouse\Dataset\E-Commerce Shipping Data.ktr

Description: Changed type file CSV to XML

15. Dalam melakukan transformasi pilih tombol **play** (seperti gambar di bawah). Lalu ekstraksi file akan berjalan.



16. Hasil file XML yang telah diekstraksi

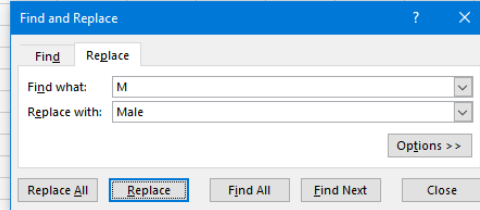
```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <Rows>
3 <Row><ID> 000000000000001</ID> <Warehouse block>D</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000004</Customer care calls>
4 <Row><ID> 000000000000002</ID> <Warehouse block>F</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000004</Customer care calls>
5 <Row><ID> 000000000000003</ID> <Warehouse block>A</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000002</Customer care calls>
6 <Row><ID> 000000000000004</ID> <Warehouse block>B</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000003</Customer care calls>
7 <Row><ID> 000000000000005</ID> <Warehouse block>C</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000002</Customer care calls>
8 <Row><ID> 000000000000006</ID> <Warehouse block>F</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000003</Customer care calls>
9 <Row><ID> 000000000000007</ID> <Warehouse block>D</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000003</Customer care calls>
10 <Row><ID> 000000000000008</ID> <Warehouse block>F</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000004</Customer care calls>
11 <Row><ID> 000000000000009</ID> <Warehouse block>A</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000003</Customer care calls>
12 <Row><ID> 000000000000010</ID> <Warehouse block>B</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000003</Customer care calls>
13 <Row><ID> 000000000000011</ID> <Warehouse block>C</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000003</Customer care calls>
14 <Row><ID> 000000000000012</ID> <Warehouse block>F</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000004</Customer care calls>
15 <Row><ID> 000000000000013</ID> <Warehouse block>D</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000003</Customer care calls>
16 <Row><ID> 000000000000014</ID> <Warehouse block>F</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000004</Customer care calls>
17 <Row><ID> 000000000000015</ID> <Warehouse block>A</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000004</Customer care calls>
18 <Row><ID> 000000000000016</ID> <Warehouse block>B</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000004</Customer care calls>
19 <Row><ID> 000000000000017</ID> <Warehouse block>C</Warehouse block> <Mode of Shipment>Flight</Mode of Shipment> <Customer care calls> 000000000000003</Customer care calls>
20 <Row><ID> 000000000000018</ID> <Warehouse block>F</Warehouse block> <Mode of Shipment>Ship</Mode of Shipment> <Customer care calls> 000000000000005</Customer care calls>
21 <Row><ID> 000000000000019</ID> <Warehouse block>D</Warehouse block> <Mode of Shipment>Ship</Mode of Shipment> <Customer care calls> 000000000000005</Customer care calls>
```



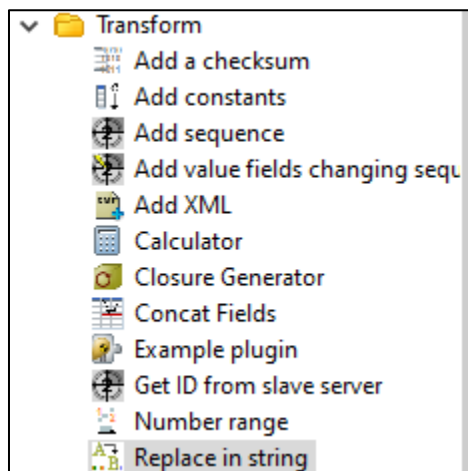
## Task Kedua: Filter (Ganti Kata)

1. Sebelum melakukan filter (ganti kata), disini terlebih dahulu dilakukan perubahan dalam data input untuk atribut gender yaitu merubah data **M** menjadi **Male**. Tujuannya disini data **Male** akan dikembalikan seperti semula menjadi data **M** dalam proses filter nantinya.

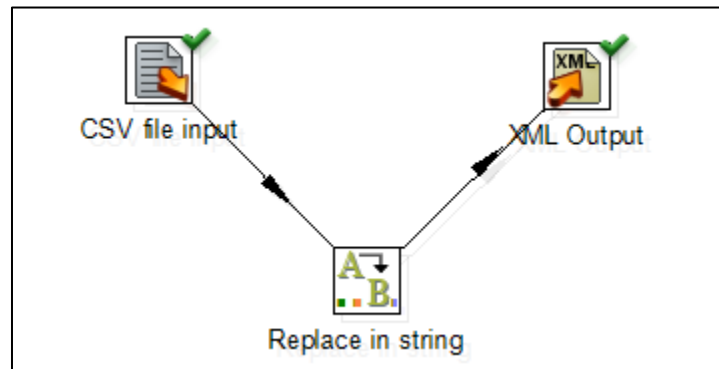
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	ID	Warehouse	Mode_of_Customer	Customer	Cost_of_t	Prior_purt	Product	i	Gender	Discount	Weight_ir	Reached.on	Time_Y,N					
2	1	D	Flight	4	2	177	3	low	F	44	1233	1						
3	2	F	Flight	4	5	216	2	low	Male	59	3088	1						
4	3	A	Flight	2	2	183	4	low	M									
5	4	B	Flight	3	3	176	4	medium	Male									
6	5	C	Flight	2	2	184	3	medium	F									
7	6	F	Flight	3	1	162	3	medium	F									
8	7	D	Flight	3	4	250	3	low	F									
9	8	F	Flight	4	1	233	2	low	F									
10	9	A	Flight	3	4	150	3	low	F									
11	10	B	Flight	3	2	164	3	medium	F									
12	11	C	Flight	3	4	189	2	medium	M									
13	12	F	Flight	4	5	232	3	medium	F									
14	13	D	Flight	3	5	198	3	medium	F	1	3667	1						
15	14	F	Flight	4	4	275	3	high	Male	29	2602	1						
16	15	A	Flight	4	3	152	3	low	M	43	1009	1						
17	16	B	Flight	4	3	227	3	low	F	45	2707	1						
18	17	C	Flight	3	4	143	2	medium	F	6	1194	1						
19	18	F	Ship	5	5	227	3	medium	Male	36	3952	1						
20	19	D	Ship	5	5	239	3	high	M	18	2495	1						
21	20	F	Ship	4	5	145	3	medium	Male	45	1059	1						
22	21	A	Ship	3	3	161	2	medium	F	38	1521	1						
23	22	B	Ship	3	1	232	4	medium	F	51	2899	1						



2. Melanjutkan dari Task 1. Kita hapus terlebih dahulu garis **Hop** yang telah dibuat, kemudian pilih folder **Transform** dan **drag Replace in String**.



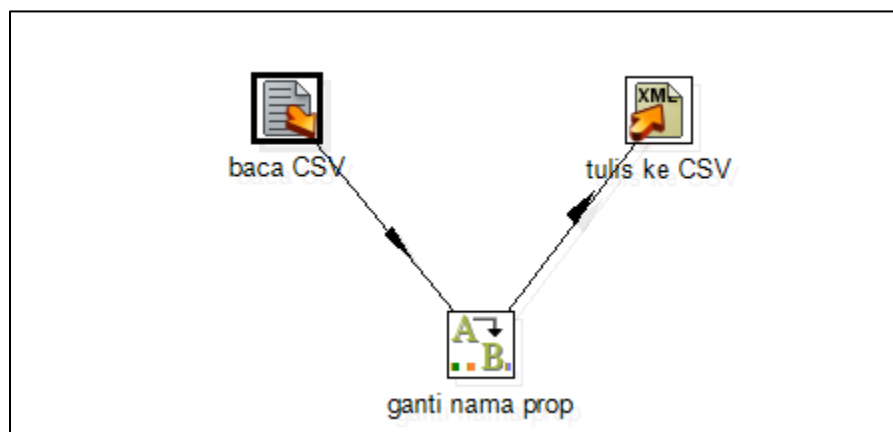
3. Tambahkan kembali garis **Hop** yang menghubungkan **CSV file input** – **Replace in String** – **XML Output**



4. Double click **Replace in String** step dan isi In stream field, **Search** dan **Replace with** (seperti gambar bawah), dan pilih **OK**. (Jalankan transformasi dan dapat dilihat hasilnya kemunculan “Male” maenjadi “M”).

Step name   Replace in string									
Fields string									
#	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field	Whole Word	Case sensitive
1	Gender		N	Male	M			N	N

5. Untuk memperjelas proses jalannya transformasi, anda dapat merubah nama step file menjadi seperti gambar dibawah.



### Task Ketiga: Output ke Database

1. Download terlebih dahulu **JDBC connector** dari MYSQL melalui link berikut :  
<http://download.nust.na/pub6/mysql/downloads/connector/j/index.html>.

2. Select Platform : **Platform Independent**, kemudian download file **ZIP Archive**

**Connector/J 5.1.12**

Select Platform:

Platform Independent

[Looking for previous GA versions?](#)

<b>Platform Independent (Architecture Independent), Compressed TAR Archive</b> (mysql-connector-java-5.1.12.tar.gz)	5.1.12	3.6M	<a href="#">Download</a>
<b>Platform Independent (Architecture Independent), ZIP Archive</b> (mysql-connector-java-5.1.12.zip)	5.1.12	3.8M	<a href="#">Download</a>

MD5: d2f836c761614a3fdf39f7a6c7c1acb5

MD5: 66e2058a1f99a50b8e98ced72b39b372

3. Ekstraksi hasil zip, pindahkan file mysql-connector-xxxx.bin.jar ke dalam folder file pdi/data-integration/lib/(mysql\_connector). Restart spoon.bat agar JDBC dapat digunakan.

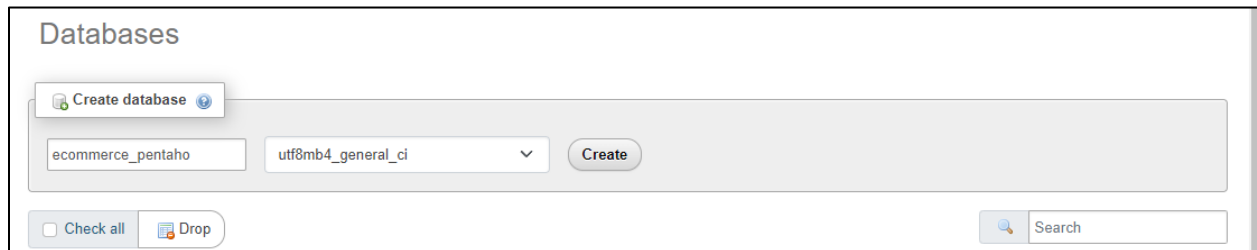
PC > Downloads > mysql-connector-j-8.0.32 > mysql-connector-j-8.0.32

Name	Date modified	Type	Size
src	07/12/2022 13:05	File folder	
build	07/12/2022 13:05	Microsoft Edge H...	95 KB
CHANGES	07/12/2022 13:05	File	273 KB
INFO_BIN	07/12/2022 13:05	File	1 KB
INFO_SRC	07/12/2022 13:05	File	1 KB
LICENSE	07/12/2022 13:05	File	70 KB
mysql-connector-j-8.0.32	07/12/2022 13:05	Executable Jar File	2.423 KB
README	07/12/2022 13:05	File	2 KB

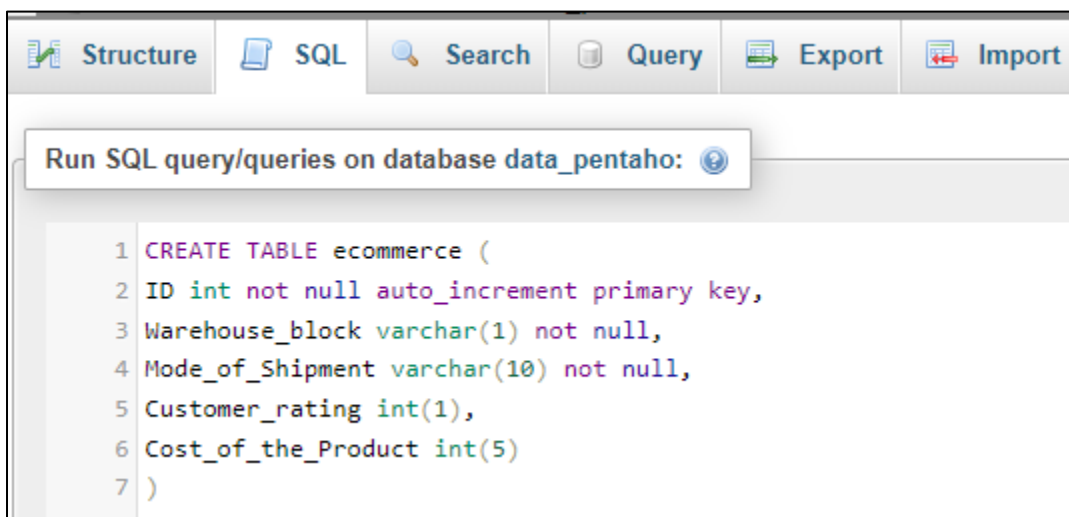
PC > Local Disk (E:) > Program Files > pdi-ce-5.0.1-stable > data-integration > lib

Name	Date modified	Type	Size
mof-200507110943	15/11/2013 16:18	Executable Jar File	171 KB
mondrian-3.6.1	15/11/2013 16:18	Executable Jar File	3.271 KB
monetdb-jdbc-2.8	15/11/2013 16:18	Executable Jar File	115 KB
mstor-0.9.13	15/11/2013 16:18	Executable Jar File	192 KB
mvel2-2.0.10	15/11/2013 16:18	Executable Jar File	594 KB
mysql-connector-j-8.0.32	07/12/2022 13:05	Executable Jar File	2.423 KB

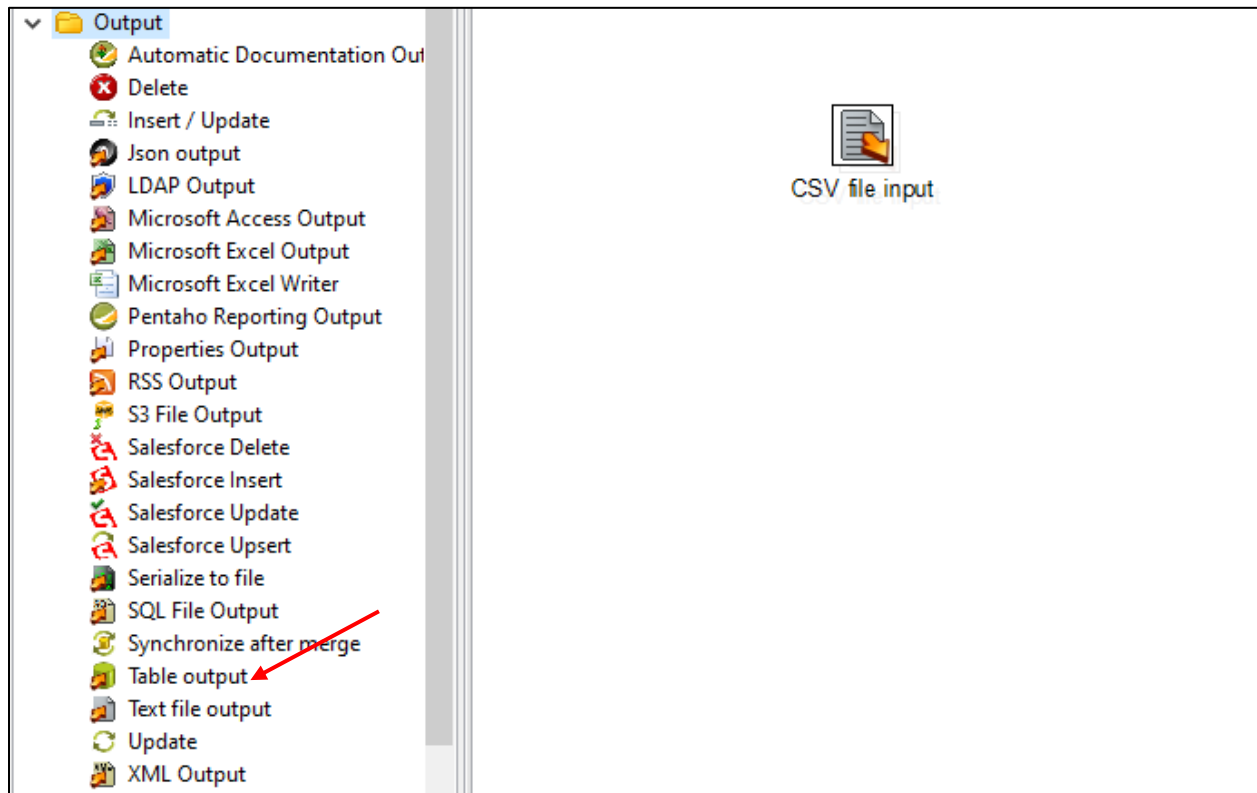
4. Pastikan server MySQL anda telah berjalan. Menggunakan phpMyAdmin atau tools yang lain, kemudian *create* database yang akan digunakan.



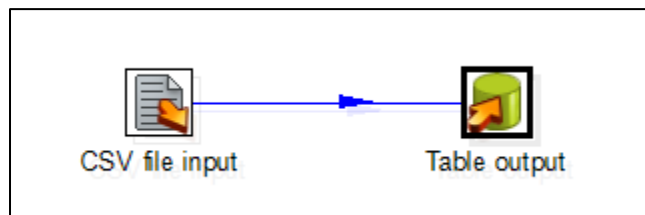
5. *Create table* dalam database ecommerce\_pentaho dengan sintaks seperti gambar dibawah :



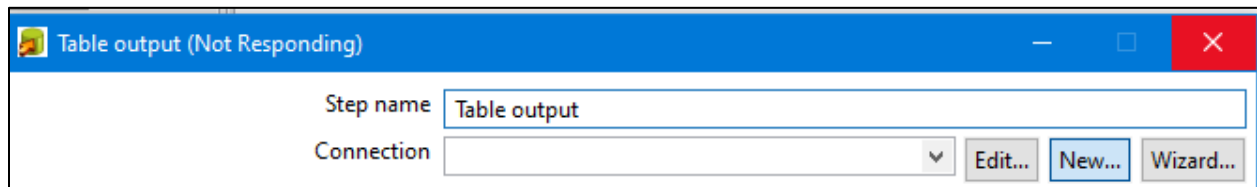
6. Kembali ke dalam spoon , buatlah transformasi baru, lalu dengan cara yang sama seperti task 1 dan task 2, tambahkan step **CSV file input**. Sedangkan untuk output, pilih **Table output**.



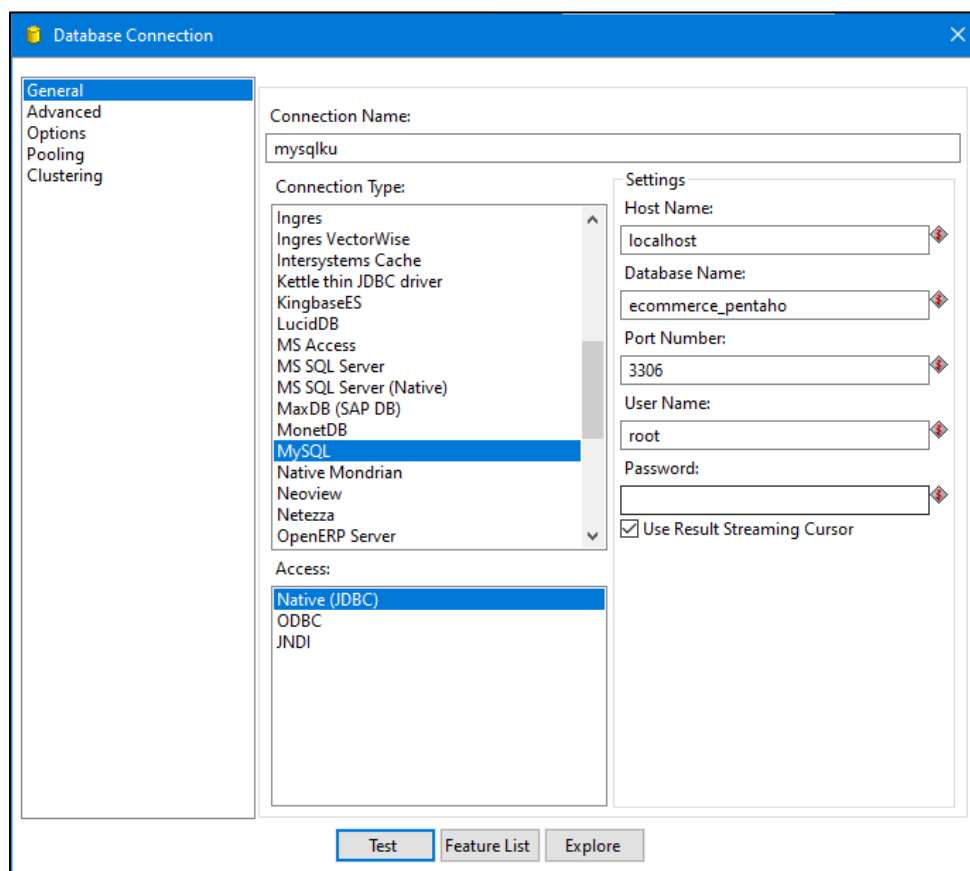
7. Buatlah garis **Hop** antara **CSV file input** dan **Table output**



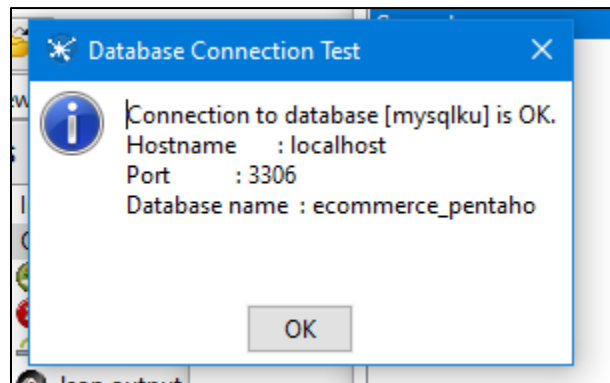
8. Double click **Table output** untuk membuat koneksi ke dalam database. Selanjutnya pilih **New**



9. Selanjutnya akan muncul kotak seperti gambar dibawah. Isilah bagian **Connection name**, **Connection Type** (MySQL), **Access** (Native JDBC), dan **Settings** sesuai phpMyAdmin user accounts. Lalu pilih tombol **Test**



10. Apabila sudah terkoneksi dengan database akan muncul tampilan seperti berikut :



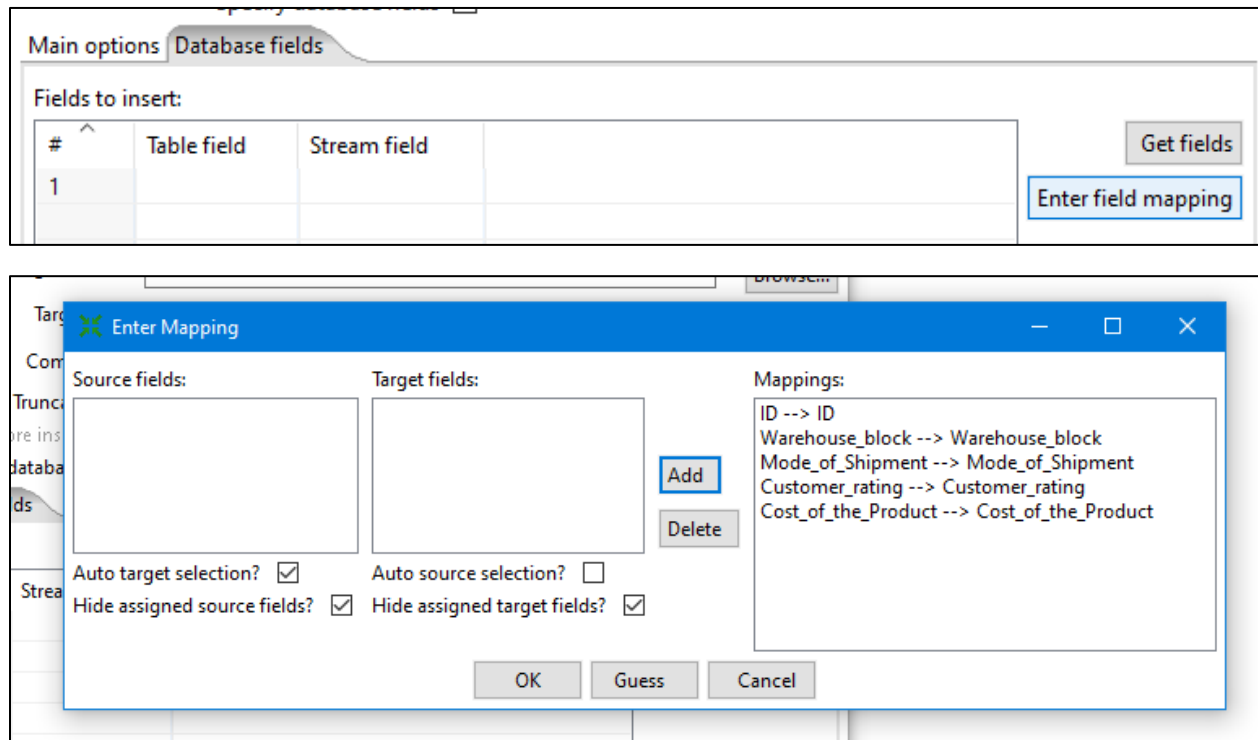
11. Isilah bagian *Target table* sesuai dengan nama tabel yang telah dibuat dalam database.

Connection	mysqlku	▼	Edit...	New...	Wizard...
Target schema		⬇		Browse...	
Target table	ecommerce	⬇		Browse...	
Commit size	1000	⬇			
Truncate table	<input type="checkbox"/>				

12. Karena nama field input tidak sama dengan nama field pada tabel, jangan lupa check “specify database field”

truncate table	<input type="checkbox"/>
Ignore insert errors	<input type="checkbox"/>
Specify database fields	<input checked="" type="checkbox"/>
Main options	Database fields

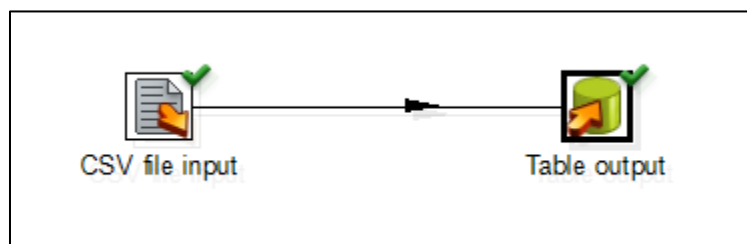
13. Pindahkan tab ke Database fields kemudian lakukan mapping atribut



14. Jalankan transformasi



15. Jika semua berjalan lancar, maka akan keluar tanda check hijau pada tabel output. Periksa tabel untuk hasil.





16. Hasil transformasi ke database.

←T→	ID	Warehouse_block	Mode_of_Shipment	Customer_rating	Cost_of_the_Product
<input type="checkbox"/> Edit Copy Delete	1	D	Flight	2	177
<input type="checkbox"/> Edit Copy Delete	2	F	Flight	5	216
<input type="checkbox"/> Edit Copy Delete	3	A	Flight	2	183
<input type="checkbox"/> Edit Copy Delete	4	B	Flight	3	176
<input type="checkbox"/> Edit Copy Delete	5	C	Flight	2	184
<input type="checkbox"/> Edit Copy Delete	6	F	Flight	1	162
<input type="checkbox"/> Edit Copy Delete	7	D	Flight	4	250
<input type="checkbox"/> Edit Copy Delete	8	F	Flight	1	233
<input type="checkbox"/> Edit Copy Delete	9	A	Flight	4	150
<input type="checkbox"/> Edit Copy Delete	10	B	Flight	2	164
<input type="checkbox"/> Edit Copy Delete	11	C	Flight	4	189
<input type="checkbox"/> Edit Copy Delete	12	F	Flight	5	232
<input type="checkbox"/> Edit Copy Delete	13	D	Flight	5	198
<input type="checkbox"/> Edit Copy Delete	14	F	Flight	4	275
<input type="checkbox"/> Edit Copy Delete	15	A	Flight	3	152
<input type="checkbox"/> Edit Copy Delete	16	B	Flight	3	227
<input type="checkbox"/> Edit Copy Delete	17	C	Flight	4	143

## Task Keempat: Lookup tabel

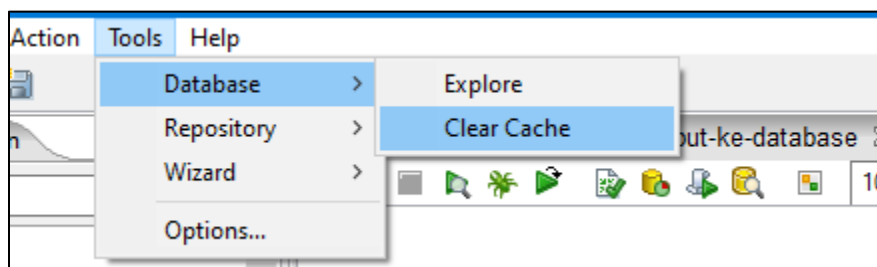
1. *Create table* Lookup Mode of Shipment. Tujuan dari langkah ini adalah mengubah “Road” menjadi kode 1, “Ship” menjadi kode 2 dan “Flight” menjadi kode 3. Eksekusi query berikut untuk menambahkan tabel lookup dan datanya:

```
1 CREATE TABLE `Lookup_Mode_of_Shipment` (  
2   `kode_Mode_of_Shipment` int(10) NOT NULL,  
3   `nama_Mode_of_Shipment` varchar(10) NOT NULL  
4 ) ENGINE=InnoDB DEFAULT CHARSET=latin1;  
5 INSERT INTO `Lookup_Mode_of_Shipment` (`kode_Mode_of_Shipment`, `nama_Mode_of_Shipment`) VALUES  
6 (1, 'Road'),  
7 (2, 'Ship'),  
8 (3, 'Flight');  
9
```

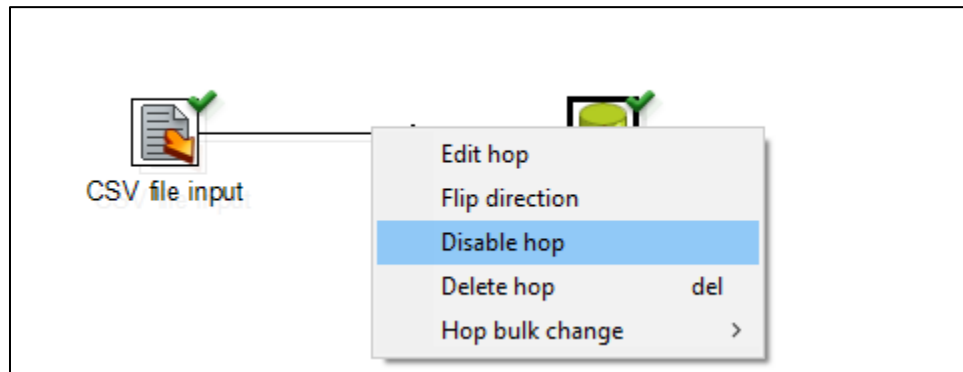
2. Tabel ecommerce juga perlu ditambahkan kode Mode of Shipment. Lakukan query berikut untuk menambahkan field kode\_Mode\_of\_Shipment

```
1 alter TABLE ecommerce  
2   add kode_Mode_of_Shipment int  
3
```

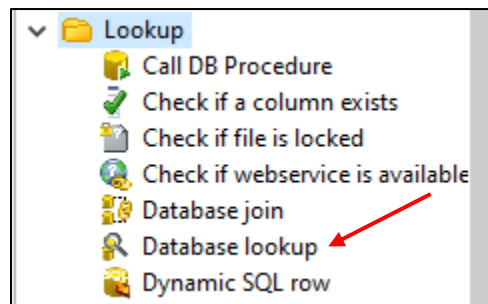
3. Kembali ke Spoon, agar perubahan database ini muncul di Spoon, bersihkan dulu cache dengan cara Tools → Database → Clear Cache.



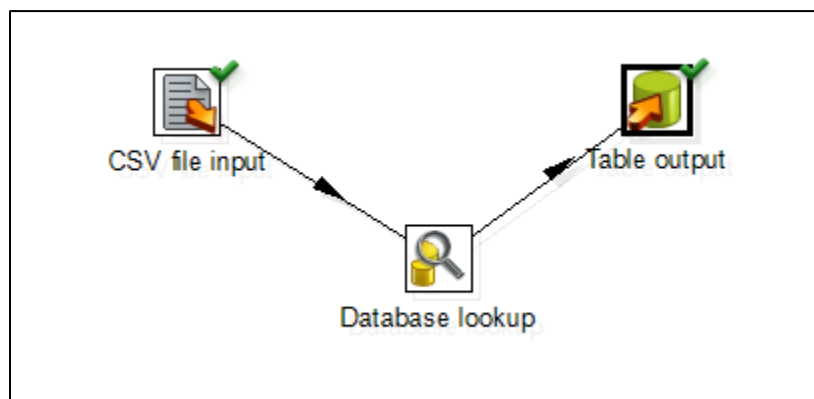
4. Lakukan **save as** terlebih dahulu untuk menyimpan transformasi. Selanjutnya disable Hop yang menghubungkan CSV file input dan Table output



5. Buka folder Lookup kemudian Database lookup



6. Buatlah garis Hop yang menghubungkan CSV file input – Database lookup – Table output



7. Double klik Database lookup untuk mengedit property. Isi Lookup table, key dan lookup dan jangan lupa tipe-nya diisi (gambar bawah)

Database Value Lookup

Step name: Database lookup

Connection: mysqlku [Edit...] [New...] [Wizard...]

Lookup schema: [Browse...]

Lookup table: lookup\_Mode\_of-Shipment [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	nama_Mod...	=	Mode_...	Field2

Values to return from the lookup table:

#	Field	New name	Default	Type
1	kode_...			Integer

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

Order by: [ ]

8. Double klik Table output untuk menambahkan field kode\_propinsi, jika belum muncul di dropbox, ketikkan secara manual.

Main options Database fields		
Fields to insert:		
#	Table field	Stream field
1	ID	ID
2	Warehouse_...	Warehouse_bl...
3	Mode_of_S...	Mode_of_Ship...
4	Customer_r...	Customer_rati...
5	Cost_of_the...	Cost_of_the_P...
6	kode_mode...	kode_mode_o...
7		
8		

## Task Kelima: Lookup tabel

1. Sebagai kita memiliki file MS excel sebagai berikut :

ID	Warehouse	Mode_of	Customer	Customer	Cost_of_t	Prior_pur	Product_i	Gender	Discount	Weight_ir	Reached.on	Time_Y.N
1	D	Flight	4	2	177	3	low	F	44	1233	1	
2	F	Flight	4	5	216	2	low	M	59	3088	1	
3	A	Flight	2	2	183	4	low	M	48	3374	1	
4	B	Flight	3	3	176	4	medium	M	10	1177	1	
5	C	Flight	2	2	184	3	medium	F	46	2484	1	
6	F	Flight	3	1	162	3	medium	F	12	1417	1	
7	D	Flight	3	4	250	3	low	F	3	2371	1	
8	F	Flight	4	1	233	2	low	F	48	2804	1	
9	A	Flight	3	4	150	3	low	F	11	1861	1	
10	B	Flight	3	2	164	3	medium	F	29	1187	1	
11	C	Flight	3	4	189	2	medium	M	12	2888	1	
12	F	Flight	4	5	232	3	medium	F	32	3253	1	
13	D	Flight	3	5	198	3	medium	F	1	3667	1	
14	F	Flight	4	4	275	3	high	M	29	2602	1	
15	A	Flight	4	3	152	3	low	M	43	1009	1	
16	B	Flight	4	3	227	3	low	F	45	2707	1	
17	C	Flight	3	4	143	2	medium	F	6	1194	1	
18	F	Ship	5	5	227	3	medium	M	36	3952	1	
19	D	Ship	5	5	239	3	high	M	18	2495	1	

2. Pilih folder Input → Microsoft Excel Input, lalu edit propertynya. Pilih Browse untuk memilih direktori data kemudian pilih file excel.

Spread sheet type (engine)

File or directory

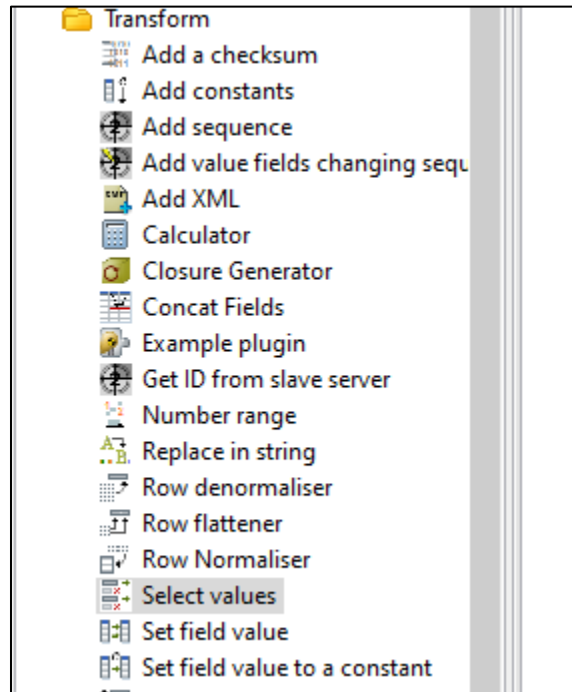
Regular Expression

Exclude Regular Expression

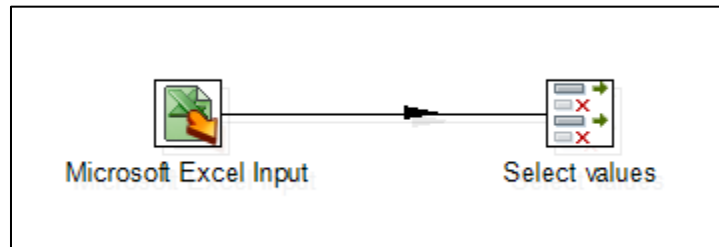
3. Pilih tab Fields lalu pilih Get fields from header row..

#	Name	Type	Length	Precision	Trim type	Repeat	Format	Current
1	ID	Number			none	N		
2	Warehouse_block	String			none	N		
3	Mode_of_Shipment	String			none	N		
4	Customer_care_calls	Number			none	N		
5	Customer_rating	Number			none	N		
6	Cost_of_the_Product	Number			none	N		
7	Prior_purchases	Number			none	N		
8	Product_importance	String			none	N		
9	Gender	String			none	N		
10	Discount_offered	Number			none	N		
11	Weight_in_gms	Number			none	N		
12	Reached.on.Time_Y.N	Number			none	N		

3. Buka folder Transform kemudian pilih dan *drag* Select Values



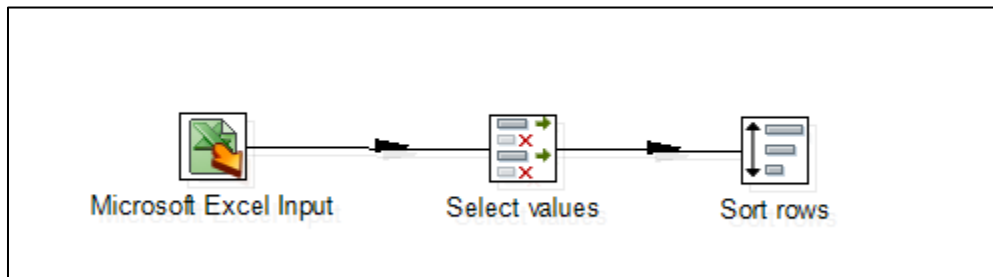
4. Buatlah garis **Hop** antara **Microsoft Excel Input** dengan **Select values**



5. Double click **Select values** kemudian pilih tab **Remove** tambahkan semua fields kecuali field **Mode of Shipment** dan **Product importance**

Select & Alter Remove Meta-data		
Fields to remove :		
#	Fieldname	
1	Cost_of_the_Product	
2	Customer_care_calls	
3	Customer_rating	
4	Discount_offered	
5	Gender	
6	ID	
7	Prior_purchases	
8	Reached.on.Time_Y.N	

6. Tambahkan **Sort rows** kemudian buat garis **Hop** lagi sehingga **Select values** terhubung ke **Sort rows**



7. Double click untuk mengedit property **Sort rows**, tambahkan bahwa pengurutan berdasarkan atribut **Product importance**

Sort rows

Step name: Sort rows

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

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

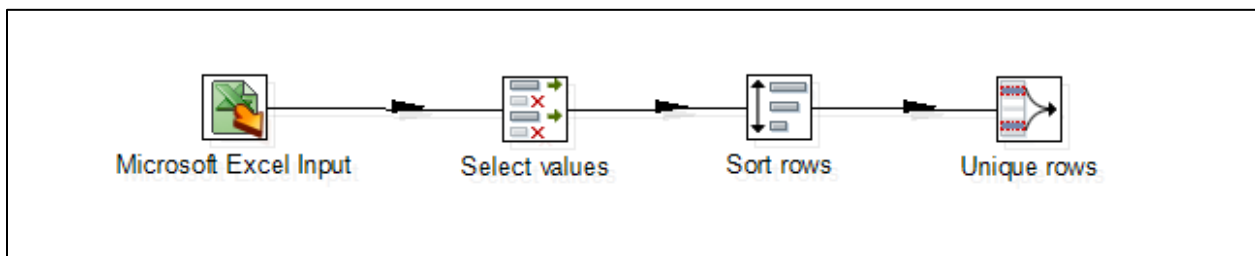
Compress TMP Files? ☐

Only pass unique rows? (verifies keys only) ☐

Fields :

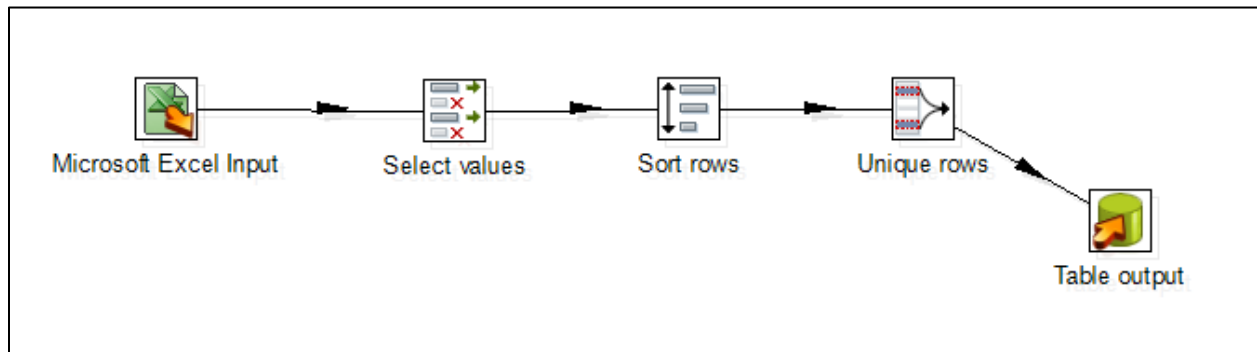
#	Fieldname	Ascending	Case sensitive compare?	Presorted?
1	Product_importance			
2				
3				

8. Tambahkan **Transfrom** → **Unique rows** dan hubungkan dengan garis **Hop** dari **Sort rows**. Tidak mengedit property pada langkah ini.





9. Selanjutnya tambahkan Output → Table Output seperti pada task 2. Isi koneksi dan nama tabelnya.



10. Struktur tabel adalah sebagai berikut:

```
1 create table product_importance (Product_importance varchar(10),
2                                Mode_of_Shipment varchar(20));
```

11. Jalankan transformasi dan hasilnya akan menjadi seperti berikut.

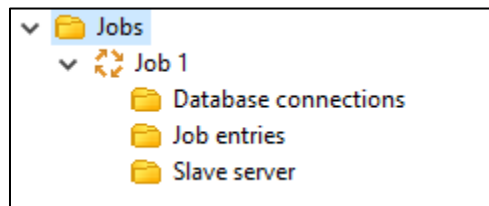
#	Mode_of_Shipment	Product_importance
1	Flight	high
2	Ship	high
3	Flight	low
4	Flight	low
5	Flight	low
6	Flight	low
7	Flight	low
8	Flight	low
9	Flight	low
10	Flight	low
11	Flight	medium
12	Flight	medium
13	Flight	medium
14	Flight	medium
15	Flight	medium
16	Flight	medium
17	Flight	medium
18	Flight	medium
19	Ship	medium

## Job

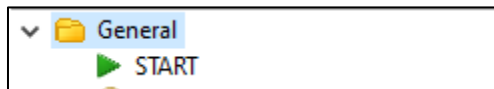
Dalam PDI, Job digunakan untuk mengkoordinasikan proses ETL. Fungsi Job adalah:

- Mengatur urutan transformasi.
- Penjadwalan transformasi.
- Pengecekan kondisi sebelum dilakukan transformasi. Misalnya apakah file atau tabel input sudah tersedia.
- Pengelolaan file (FTP, copy, delete file)
- Mengirimkan notifikasi melalui email.

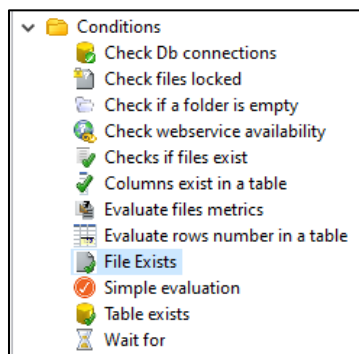
1. Pilih **New** → **Job** atau **CTRL + ALT + N**.



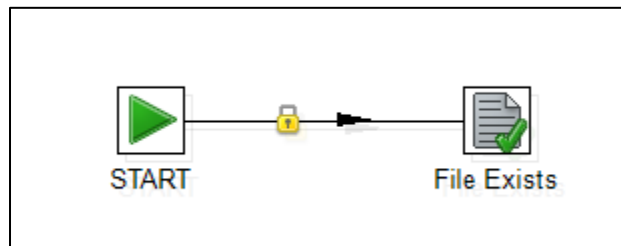
2. Pilih (section **Design**) lalu folder **General** kemudian pilih **START**



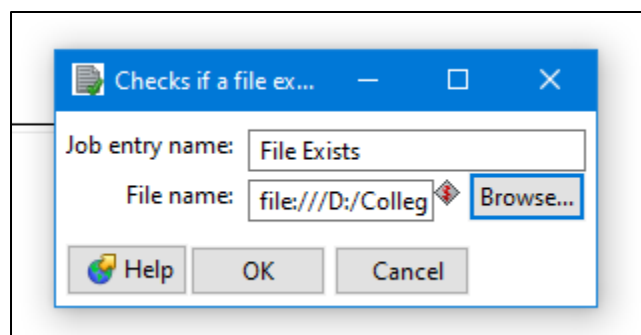
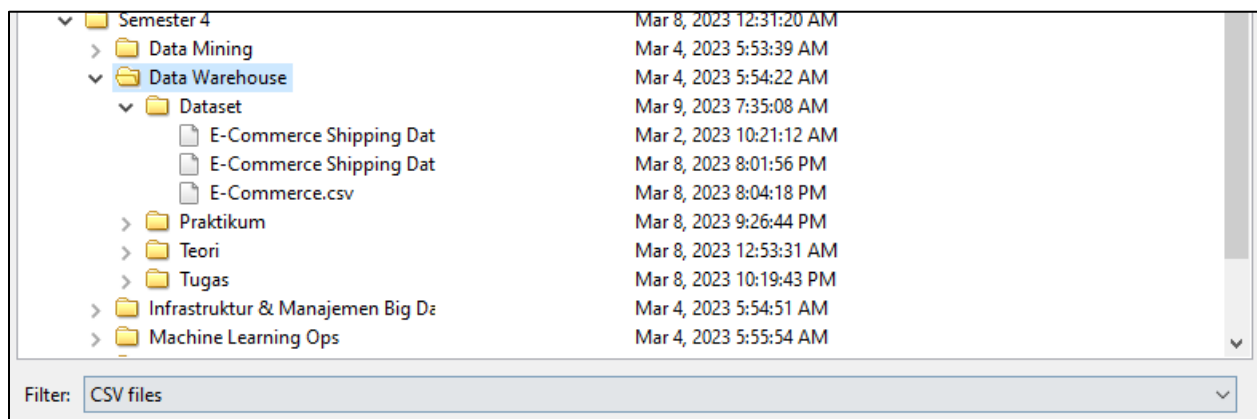
3. Pilih folder **Conditions** → **File Exits**. Step ini untuk mengecek apakah file input sudah ada.



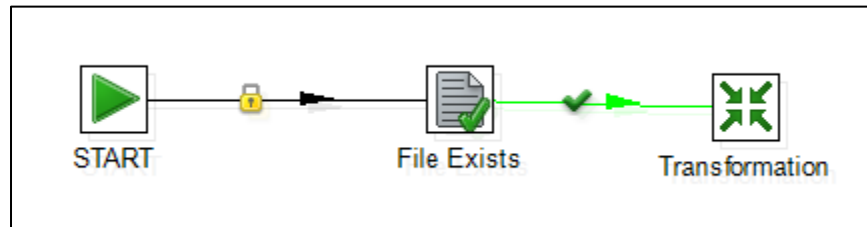
4. Buat garis **Hop** antara **START** menuju **File exists**



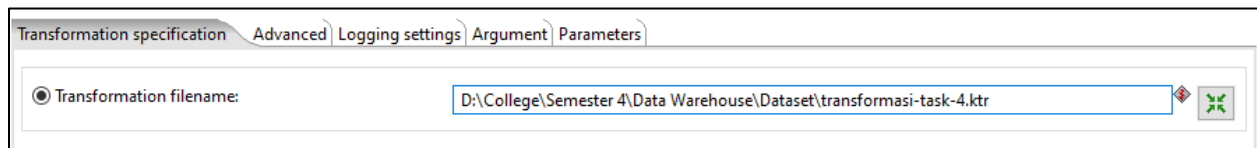
5. Pilih filter **CSV Files** dan pilih file E-Commerce.csv



6. Pilih folder **General** kemudian pilih **Transformation**.



7. Double click **Transformation** kemudian pilih file transformasi dari task 4



8. Silahkan jalankan Job yang telah dibuat