



# **DATA WAREHOUSE**

## **LAPORAN TUGAS 1 : ETL DATASET**

By

**MOCH. TORIQUIL MUCHLISSIN**

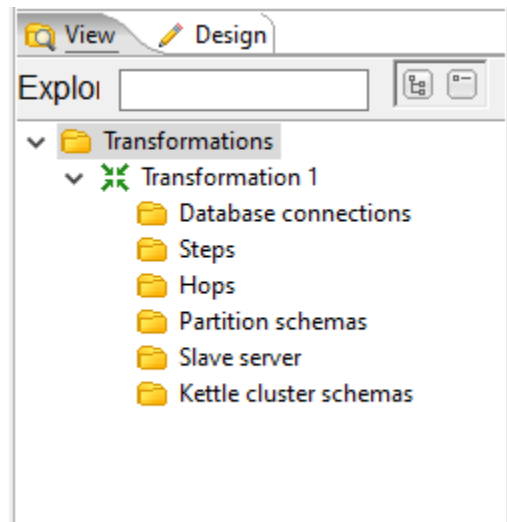
**2 - D4 SD A | 3321600001**

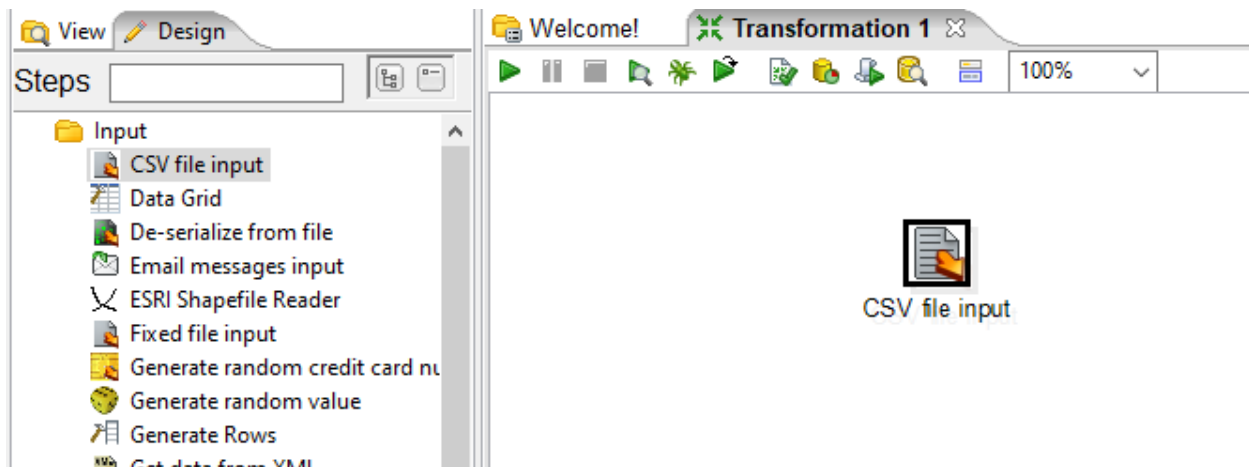
**PROGRAM STUDI D4 SAINS DATA TERAPAN  
POLITEKNIK ELEKTRONIKA NEGERI SURABAYA**

**2023**

## Task Pertama: Ekspor CSV ke XML

	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			





Step name: CSV file input

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

Delimiter: ; Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result? ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

File encoding:

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result? ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

**Sample size**

Enter the number of lines to sample:

100

OK Cancel

#	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

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	
..	..	..	..	..	..	..	..	..	..	..	..

- Output
- Automatic Documentation Out
- Delete
- Insert / Update
- Json output
- LDAP Output
- Microsoft Access Output
- Microsoft Excel Output
- Microsoft Excel Writer
- Pentaho Reporting Output
- Properties Output
- RSS Output
- S3 File Output
- Salesforce Delete
- Salesforce Insert
- Salesforce Update
- Salesforce Upsert
- Serialize to file
- SQL File Output
- Synchronize after merge
- Table output
- Text file output
- Update
- XML Output



CSV file input



XML Output



CSV file input



XML Output

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 ☐

File Content Fields

#	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				

Transformation properties

Transformation Parameters Logging Dates Dependencies Miscellaneous Monitoring

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



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <Rows>
3 <Row><ID> 0000000000000001</ID> <Warehouse_block>D</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000004</Customer_care_calls>
4 <Row><ID> 0000000000000002</ID> <Warehouse_block>F</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000004</Customer_care_calls>
5 <Row><ID> 0000000000000003</ID> <Warehouse_block>A</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000002</Customer_care_calls>
6 <Row><ID> 0000000000000004</ID> <Warehouse_block>B</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000002</Customer_care_calls>
7 <Row><ID> 0000000000000005</ID> <Warehouse_block>C</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000002</Customer_care_calls>
8 <Row><ID> 0000000000000006</ID> <Warehouse_block>F</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000003</Customer_care_calls>
9 <Row><ID> 0000000000000007</ID> <Warehouse_block>D</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000003</Customer_care_calls>
10 <Row><ID> 0000000000000008</ID> <Warehouse_block>F</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000004</Customer_care_calls>
11 <Row><ID> 0000000000000009</ID> <Warehouse_block>A</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000003</Customer_care_calls>
12 <Row><ID> 0000000000000010</ID> <Warehouse_block>B</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000003</Customer_care_calls>
13 <Row><ID> 0000000000000011</ID> <Warehouse_block>C</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000003</Customer_care_calls>
14 <Row><ID> 0000000000000012</ID> <Warehouse_block>F</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000004</Customer_care_calls>
15 <Row><ID> 0000000000000013</ID> <Warehouse_block>D</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000003</Customer_care_calls>
16 <Row><ID> 0000000000000014</ID> <Warehouse_block>F</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000004</Customer_care_calls>
17 <Row><ID> 0000000000000015</ID> <Warehouse_block>A</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000004</Customer_care_calls>
18 <Row><ID> 0000000000000016</ID> <Warehouse_block>B</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000004</Customer_care_calls>
19 <Row><ID> 0000000000000017</ID> <Warehouse_block>C</Warehouse_block> <Mode_of_Shipment>Flight</Mode_of_Shipment> <Customer_care_calls> 0000000000000003</Customer_care_calls>
20 <Row><ID> 0000000000000018</ID> <Warehouse_block>F</Warehouse_block> <Mode_of_Shipment>Ship</Mode_of_Shipment> <Customer_care_calls> 0000000000000005</Customer_care_calls>
21 <Row><ID> 0000000000000019</ID> <Warehouse_block>D</Warehouse_block> <Mode_of_Shipment>Ship</Mode_of_Shipment> <Customer_care_calls> 0000000000000005</Customer_care_calls>
```

## Task Kedua: Filter (Ganti Kata)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	ID	Warehous	Mode_of	Customer	Customer	Cost_of_t	Prior_purt	Product	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						

Find and Replace

Find Replace

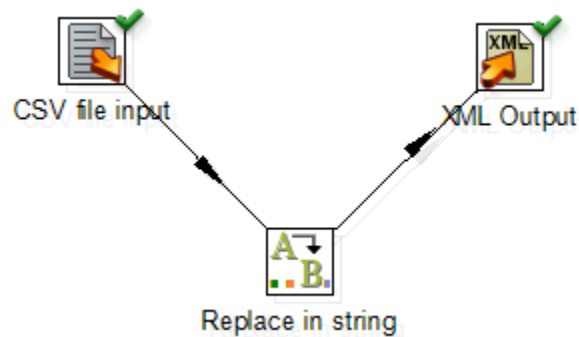
Find what: M

Replace with: Male

Options >>

Replace All Replace Find All Find Next Close

- Transform
- Add a checksum
  - Add constants
  - Add sequence
  - Add value fields changing sequ
  - Add XML
  - Calculator
  - Closure Generator
  - Concat Fields
  - Example plugin
  - Get ID from slave server
  - Number range
  - Replace in string



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

