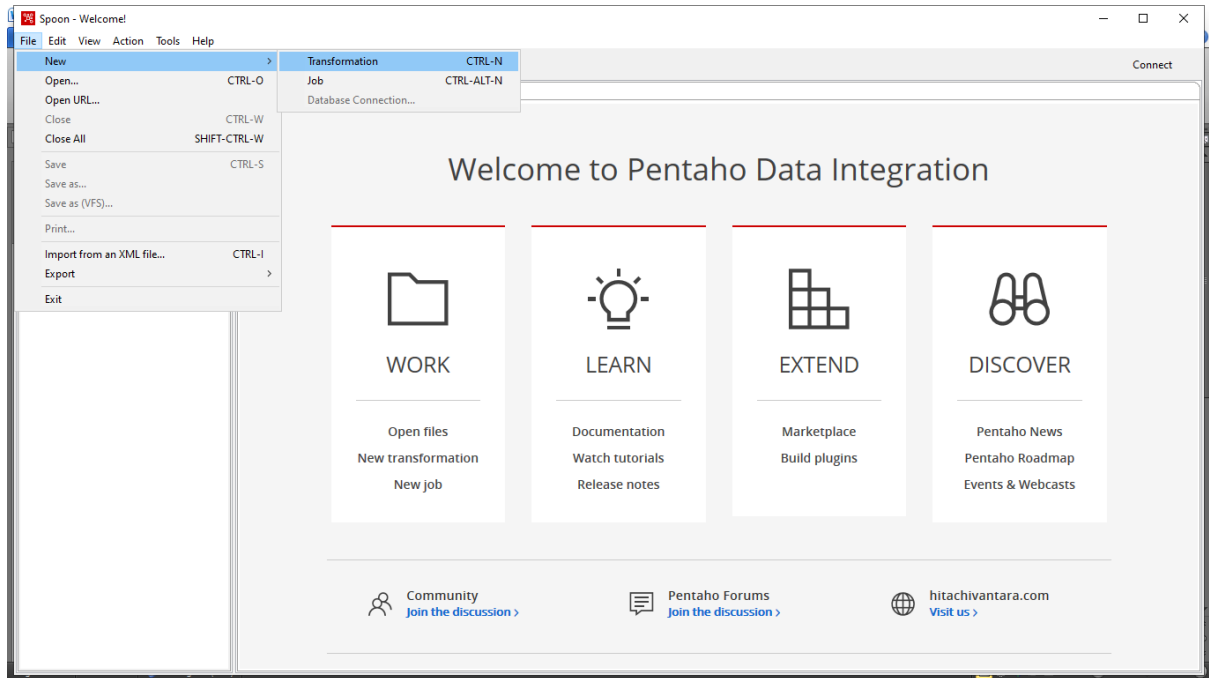
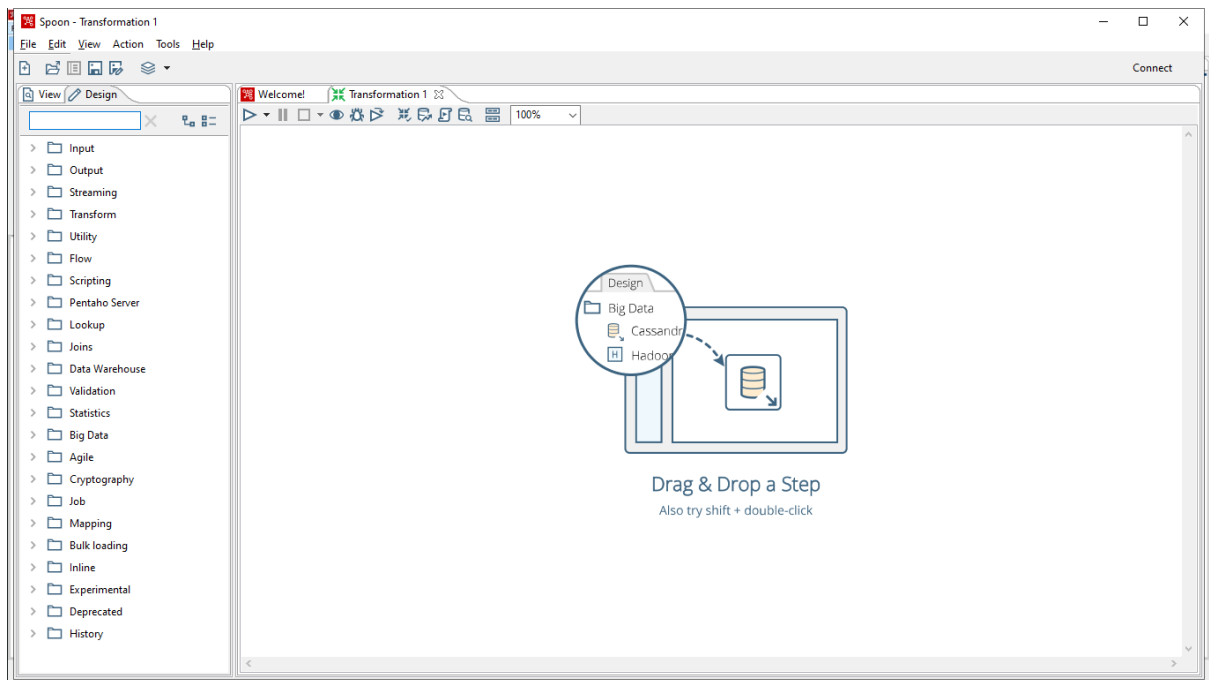


Pentaho - Extract and Load

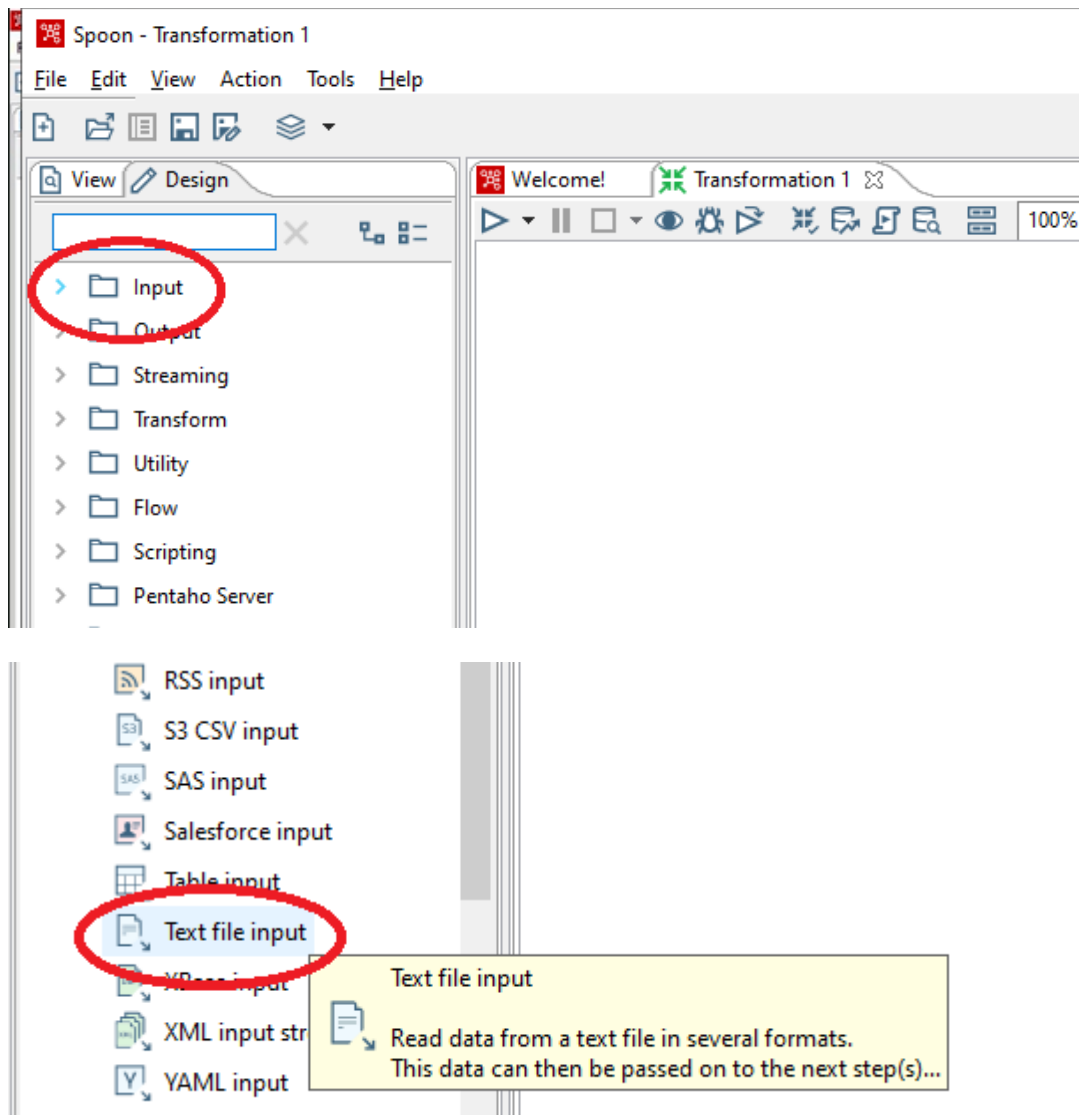
1. Untuk memulai penggunaan pentaho, pilih menu file > new > transformation



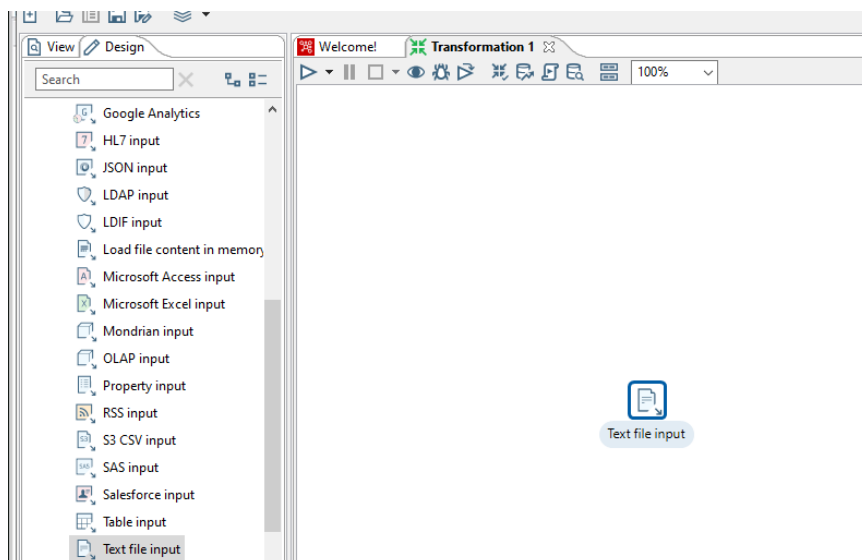
2. Akan muncul tampilan seperti berikut :



3. Pada menu input, drag pilihan 'Text file input' ke blank space.



Tampilan hasil :



4. Klik 2 kali icon 'Text file input' yang dihasilkan tadi, maka akan muncul pop up berikut

Text file input

Step name: Read sales data

File or directory:

Regular Expression:

Exclude Regular Expression:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include subfolders
1					

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

OK Preview rows Cancel

Untuk kasus yang saya lakukan, isi nama sesuai gambar. Lalu browse data csv untuk sales data sesuai dengan path yang ada di gambar:

Text file input

Step name: Read sales data

File or directory: C:\data-integration\samples\transformations\files\sales_data.csv

Regular Expression:

Exclude Regular Expression:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include subfolders
---	----------------	-------------------	------------------	----------	--------------------

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

OK Preview rows Cancel

Tekan add sehingga akan muncul file baru dalam tabel 'Selected Files' :

Text file input

Step name: Read sales data

File or directory: C:\data-integration\samples\transformations\files\sales_data.csv

Regular Expression:

Exclude Regular Expression:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include subfolders
1	C:\data-integration\samples\transformations\files\sales_data.csv				
2					

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

OK Preview rows Cancel

5. Pada content, ubah format ke Unix :

The screenshot shows the 'Content' tab of a data tool interface. The 'Filetype' is set to 'CSV'. The 'Separator' is a semicolon (;). The 'Enclosure' is a double quote ("). The 'Escape' is empty. The 'Header' checkbox is checked, and 'Number of header lines' is set to 1. The 'Footer' checkbox is unchecked, and 'Number of footer lines' is set to 1. The 'Wrapped lines?' checkbox is unchecked, and 'Number of times wrapped' is set to 1. The 'Paged layout (printout)?' checkbox is unchecked, and 'Number of lines per page' is set to 80. The 'Document header lines' is set to 0. The 'Compression' is set to 'None'. The 'No empty rows' checkbox is checked. The 'Include filename in output?' checkbox is unchecked. The 'Rownum in output?' checkbox is unchecked. The 'Format' dropdown is set to 'Unix', with 'DOS' and 'mixed' also visible in the list. The 'Help' button is at the bottom left.

Kembali ke file dan lakukan langkah berikut untuk menentukan jumlah baris atas konten yang diambil:

The screenshot shows the 'Content' tab of a data tool interface. The 'File or directory' field is empty. The 'Regular Expression' and 'Exclude Regular Expression' fields are empty. The 'Selected files' list shows 'C:\data-1*.csv' with 'Nr of lines to view...' set to 100. The 'Show filename(s)...' button is circled in red. The 'Show file content' button is also visible. The 'Preview rows' button is at the bottom right.

Hasil :

The screenshot shows the 'Content of first file' window. The table has columns: ORDERNUMBER, QUANTITYORDERED, PRICEEACH, ORDERLINENUMBER, SALES, ORDERDATE, STATUS, QTR_ID, MONTH_ID, YEAR_ID, PRODUCTLINE, MSRP, PRODUCTCODE, CUSTOMERNAME, PHONE, ADDRESS, CITY, STATE, COUNTRY. The data is a large list of orders, including details like order numbers, quantities, prices, dates, and customer information.

6. Keluar dari tampilan konten lalu pilih menu konten. Menu ini menjadi penentu atas apa saja yang ditampilkan di konten tadi.

7. Pada menu fields, pilih Get Fields dan isi 0, hasilnya :

#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type	Repeat
1	ORDERNUMBER	Integer	#		15	0	IDR	.	,	-		none	N
2	QUANTITYORDERED	Integer	#		15	0	IDR	.	,	-		none	N
3	PRICEEACH	Number	#,.		5	2	IDR	.	,	-		none	N
4	ORDERLINENUMBER	Integer	#		15	0	IDR	.	,	-		none	N
5	SALES	Number	#,.		7	2	IDR	.	,	-		none	N
6	ORDERDATE	String			15		IDR	.	,	-		none	N
7	STATUS	String			10		IDR	.	,	-		none	N
8	QTR_ID	Integer	#		15	0	IDR	.	,	-		none	N
9	MONTH_ID	Integer	#		15	0	IDR	.	,	-		none	N
10	YEAR_ID	Integer	#		15	0	IDR	.	,	-		none	N
11	PRODUCTLINE	String			16		IDR	.	,	-		none	N
12	MSRP	Integer	#		15	0	IDR	.	,	-		none	N
13	PRODUCTCODE	String			9		IDR	.	,	-		none	N
14	CUSTOMERNAME	String			34		IDR	.	,	-		none	N
15	PHONE	String			17		IDR	.	,	-		none	N

Hasil preview rows :

#	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCTLINE	MSRP	PRODUCTCODE	CUSTID
1	10107	30	95.7	2	2871	2/24/2003 0:00	Shipped	1	2	2003	Motorcycles	95	S10_1678	Land
2	10121	34	81.3	5	2765.9	5/7/2003 0:00	Shipped	2	5	2003	Motorcycles	95	S10_1678	Reim:
3	10134	41	94.7	2	3884.3	7/1/2003 0:00	Shipped	3	7	2003	Motorcycles	95	S10_1678	Lyon
4	10145	45	83.3	6	3746.7	8/25/2003 0:00	Shipped	3	8	2003	Motorcycles	95	S10_1678	Toys4
5	10159	49	100	14	5205.3	10/10/2003 0:00	Shipped	4	10	2003	Motorcycles	95	S10_1678	Corpr
6	10168	36	96.7	1	3479.8	10/28/2003 0:00	Shipped	4	10	2003	Motorcycles	95	S10_1678	Techn
7	10180	29	86.1	9	2497.8	11/11/2003 0:00	Shipped	4	11	2003	Motorcycles	95	S10_1678	Daed.
8	10188	48	100	1	5512.3	11/18/2003 0:00	Shipped	4	11	2003	Motorcycles	95	S10_1678	Herkk
9	10201	22	98.6	2	2168.5	12/1/2003 0:00	Shipped	4	12	2003	Motorcycles	95	S10_1678	Mini
10	10211	41	100	14	4708.4	1/15/2004 0:00	Shipped	1	1	2004	Motorcycles	95	S10_1678	Auto
11	10223	37	100	1	3965.7	2/20/2004 0:00	Shipped	1	2	2004	Motorcycles	95	S10_1678	Austr
12	10237	23	100	7	2333.1	4/5/2004 0:00	Shipped	2	4	2004	Motorcycles	95	S10_1678	Vitac
13	10251	28	100	2	3188.6	5/18/2004 0:00	Shipped	2	5	2004	Motorcycles	95	S10_1678	Tekni
14	10263	34	100	2	3676.8	6/28/2004 0:00	Shipped	2	6	2004	Motorcycles	95	S10_1678	Gift D
15	10275	45	92.8	1	4177.4	7/23/2004 0:00	Shipped	3	7	2004	Motorcycles	95	S10_1678	La Ro
16	10285	36	100	6	4099.7	8/27/2004 0:00	Shipped	3	8	2004	Motorcycles	95	S10_1678	Marte
17	10299	23	100	9	2597.4	9/30/2004 0:00	Shipped	3	9	2004	Motorcycles	95	S10_1678	Toys
18	10309	41	100	5	4394.4	10/15/2004 0:00	Shipped	4	10	2004	Motorcycles	95	S10_1678	Baani
19	10318	46	94.7	1	4398	11/2/2004 0:00	Shipped	4	11	2004	Motorcycles	95	S10_1678	Dieca
20	10329	42	100	1	4396.1	11/15/2004 0:00	Shipped	4	11	2004	Motorcycles	95	S10_1678	Land
21	10341	41	100	9	7737.9	11/24/2004 0:00	Shipped	4	11	2004	Motorcycles	95	S10_1678	Salzbr
22	10361	20	72.5	13	1451	12/17/2004 0:00	Shipped	4	12	2004	Motorcycles	95	S10_1678	Souvi
23	10375	21	34.9	12	733.1	2/3/2005 0:00	Shipped	1	2	2005	Motorcycles	95	S10_1678	La Ro
24	10388	42	76.4	4	3207.1	3/3/2005 0:00	Shipped	1	3	2005	Motorcycles	95	S10_1678	FunG
25	10403	24	100	7	2434.6	4/8/2005 0:00	Shipped	2	4	2005	Motorcycles	95	S10_1678	UK Cr
26	10417	66	100	2	7516.1	5/13/2005 0:00	Disputed	2	5	2005	Motorcycles	95	S10_1678	Euro+
27	10103	26	100	11	5404.6	1/29/2003 0:00	Shipped	1	1	2003	Classic Cars	214	S10_1949	Baani
28	10112	29	100	1	7209.1	3/24/2003 0:00	Shipped	1	3	2003	Classic Cars	214	S10_1949	Volvo
29	10126	38	100	11	7329.1	5/28/2003 0:00	Shipped	2	5	2003	Classic Cars	214	S10_1949	Corric
30	10140	37	100	11	7374.1	7/24/2003 0:00	Shipped	3	7	2003	Classic Cars	214	S10_1949	Techn
31	10150	45	100	8	10993.5	9/19/2003 0:00	Shipped	3	9	2003	Classic Cars	214	S10_1949	Drago
32	10163	21	100	1	4860.2	10/20/2003 0:00	Shipped	4	10	2003	Classic Cars	214	S10_1949	Classi

Masih ada data yang tidak sesuai :

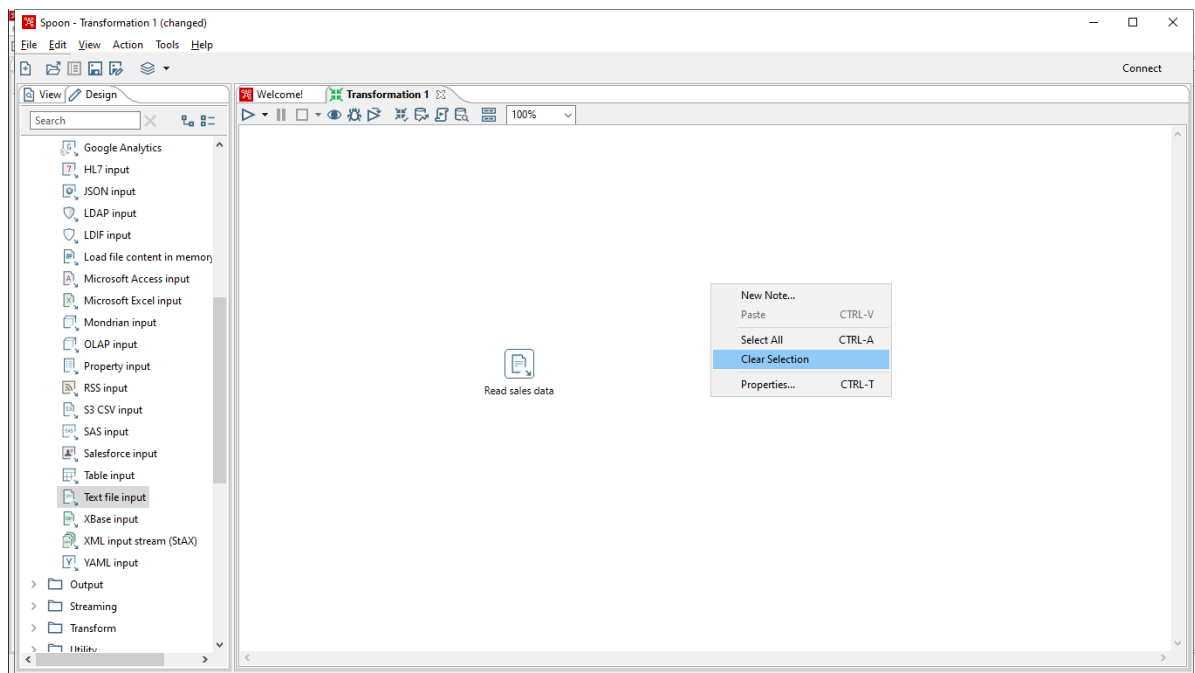
Examine preview data

Rows of steps: Read sales data (1000 rows)

PHONE	ADDRESSLINE1	ADDRESSLINE2	CITY	STATE	POSTALCODE	COUNTRY	TERRITORY	CONTACTLASTNAME	CONTACTFIRSTNAME
2125557818	897 Long Airport Avenue	<null>	NYC	NY	10022	United States	NA	Yu	Kwai
26.47.1555	59 rue de l'Abbaye	<null>	Reims	<null>	51100	France	EMEA	Henriot	Paul
+33 1 46 62 7555	27 rue du Colonel Pierre Avia	<null>	Paris	<null>	75508	France	EMEA	Da Cunha	Daniel
6265557265	78934 Hillside Dr.	<null>	Pasadena	CA	90003	USA	NA	Young	Julie
6505551386	7734 Strong St.	<null>	San Francisco	CA	<null>	United States	NA	Brown	Julie
6505556809	9408 Furth Circle	<null>	Burlingame	CA	94217	USA	NA	Hirano	Juri
20.16.1555	184, chaussée de Tournai	<null>	Lille	<null>	59000	France	EMEA	Ranc.	Martine
+47 2267 3215	Drammen 121, PR 744 Sentrum	<null>	Bergen	<null>	N 5804	Norway	EMEA	Oestan	Veysel
6505555787	5557 North Pendale Street	<null>	San Francisco	CA	<null>	United States	NA	Murphy	Julie
(1) 47.55.6555	25, rue Lauriston	<null>	Paris	<null>	75016	France	EMEA	Perrier	Dominique
03 9520 4555	636 St Kilda Road	<null>	Melbourne	Victoria	3004	Australia	APAC	Ferguson	Peter
2125551500	2678 Kingston Rd.	<null>	NYC	NY	10022	USA	NA	Frick	Michael
2015559350	7476 Moss Rd.	<null>	Newark	NJ	94019	USA	NA	Brown	William
203555270	25593 Spring Bay Ln.	<null>	Bridgewater	CT	97562	United States	NA	King	Julie
40.67.8555	67, rue des Cinquante Otages	<null>	Nantes	<null>	44000	France	EMEA	Labrun	Janine
6175558555	39323 Spinnaker Dr.	<null>	Cambridge	MA	51247	USA	NA	Hernandez	Marta
90-224 8555	Keskuskatu 45	<null>	Helsinki	<null>	21240	Finland	EMEA	Karttunen	Matti
07-98 9555	Erling Skakkes gate 78	<null>	Stavern	<null>	4110	Norway	EMEA	Bergulfen	Jonas
2155551555	7586 Pompton St.	<null>	Allentown	PA	70267	USA	NA	Yu	Kyung
2125557818	897 Long Airport Avenue	<null>	NYC	NY	10022	United States	NA	Yu	Kwai
6562-9555	Geislweg 14	<null>	Salzburg	<null>	5020	Austria	EMEA	Pipps	Georg
+61 2 9495 8555	Monitor Money Building, 815 Pacific Hwy	<null>	Chatswood	NSW	2067	Australia	APAC	Huxley	Adrian
40.67.8555	67, rue des Cinquante Otages	<null>	Nantes	<null>	44000	France	EMEA	Labrun	Janine
5085552555	1785 First Street	<null>	New Bedford	MA	50553	USA	NA	Benitez	Violeta
(171) 555-2282	Berkeley Gardens 12 Brewery	<null>	Liverpool	<null>	WX1 6LT	UK	EMEA	Devon	Elizabeth
(91) 555 94 44	C/ Moralarzal, 86	<null>	Madrid	<null>	28034	Spain	EMEA	Freyre	Diego
07-98 9555	Erling Skakkes gate 78	<null>	Stavern	<null>	4110	Norway	EMEA	Bergulfen	Jonas
0921-12 3555	Berguvsv. gen 8	<null>	Luleå	<null>	S-958 22	Sweden	EMEA	Berglund	Christina
(91) 555 22 82	C/ Araquil, 67	<null>	Madrid	<null>	28023	Spain	EMEA	Sommer	Martyn
6505556809	9408 Furth Circle	<null>	Burlingame	CA	94217	USA	NA	Hirano	Juri
+65 221 7555	Bronz Sok, Bronz Apt. 3/6 Tesvikiye	<null>	Singapore	<null>	079903	Singapore	Japan	Natividad	Eric
2125558022	4005 Dromedon St	<null>	NYC	NY	10022	United States	NA	Hernandez	Marta

Close Show Log

8. Kembali ke tampilan awal, buka transformation dengan klik kanan > transformation, atau double klik kiri, atau bisa juga dengan ctrl+t.



Transformation properties

Transformation Parameters Logging Dates Dependencies Miscellaneous Monitoring

Transformation name: Getting started Transformation

Transformation filename

Description:

Extended description:

Status:

Version:

Directory: /

Created by: -

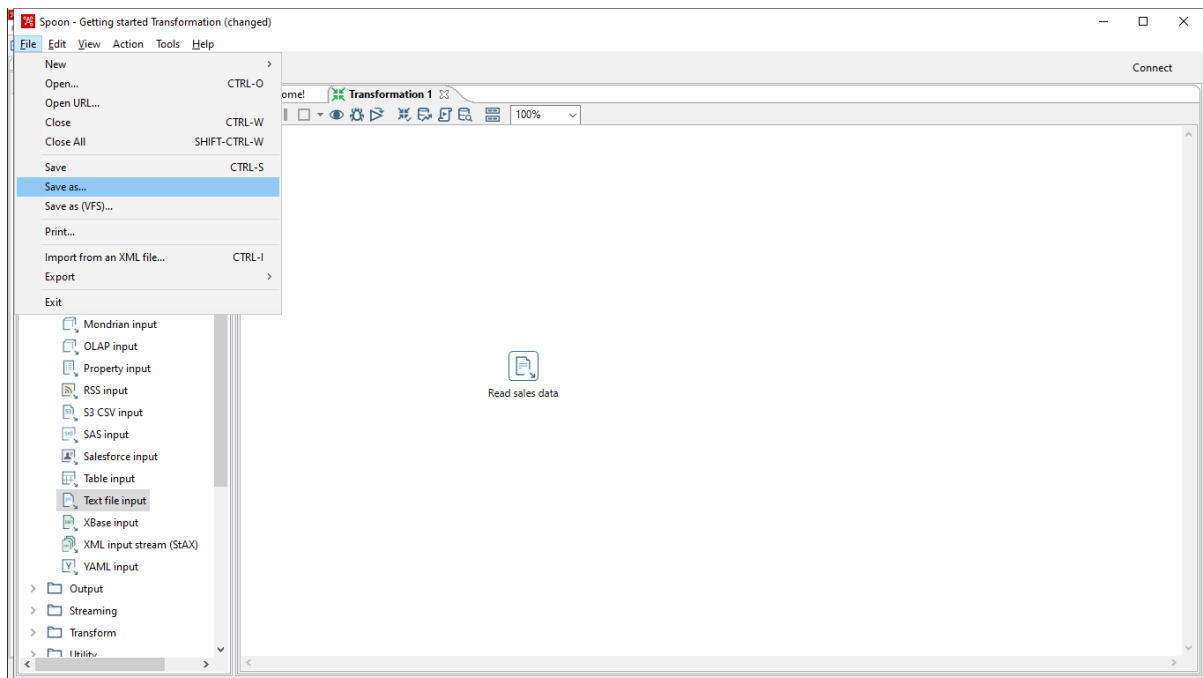
Created at: Tue Oct 11 23:25:40 ICT 2022

Last modified by: -

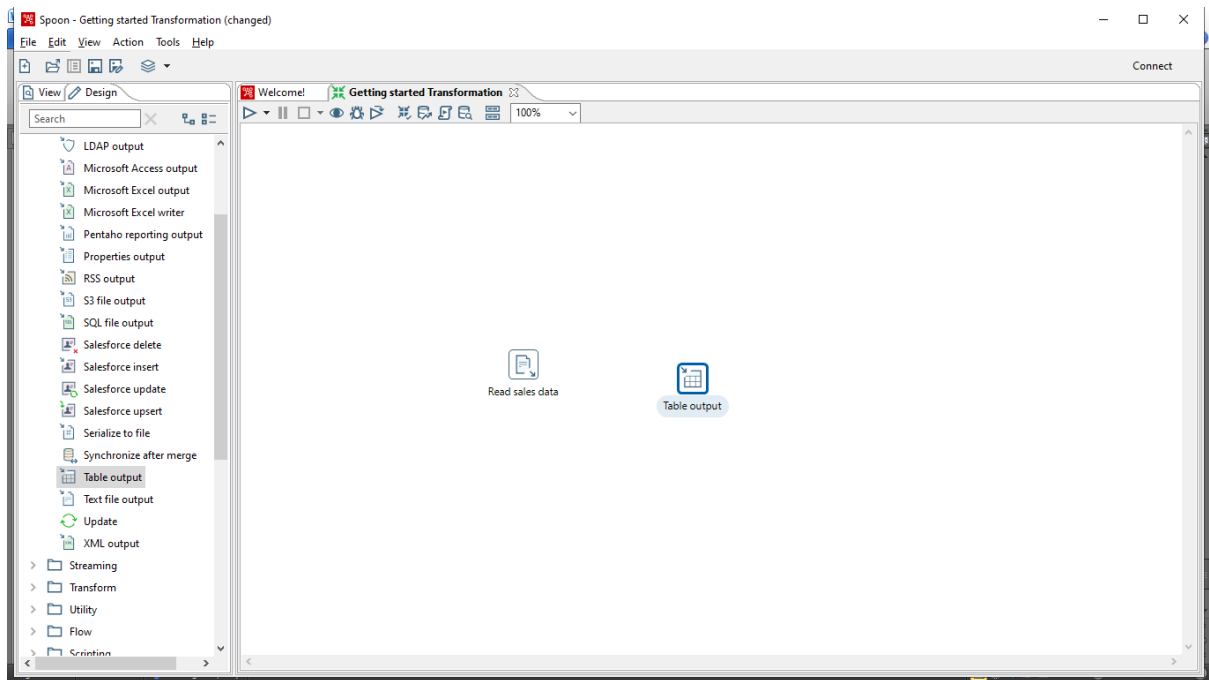
Last modified at: Tue Oct 11 23:25:40 ICT 2022

OK SQL Cancel

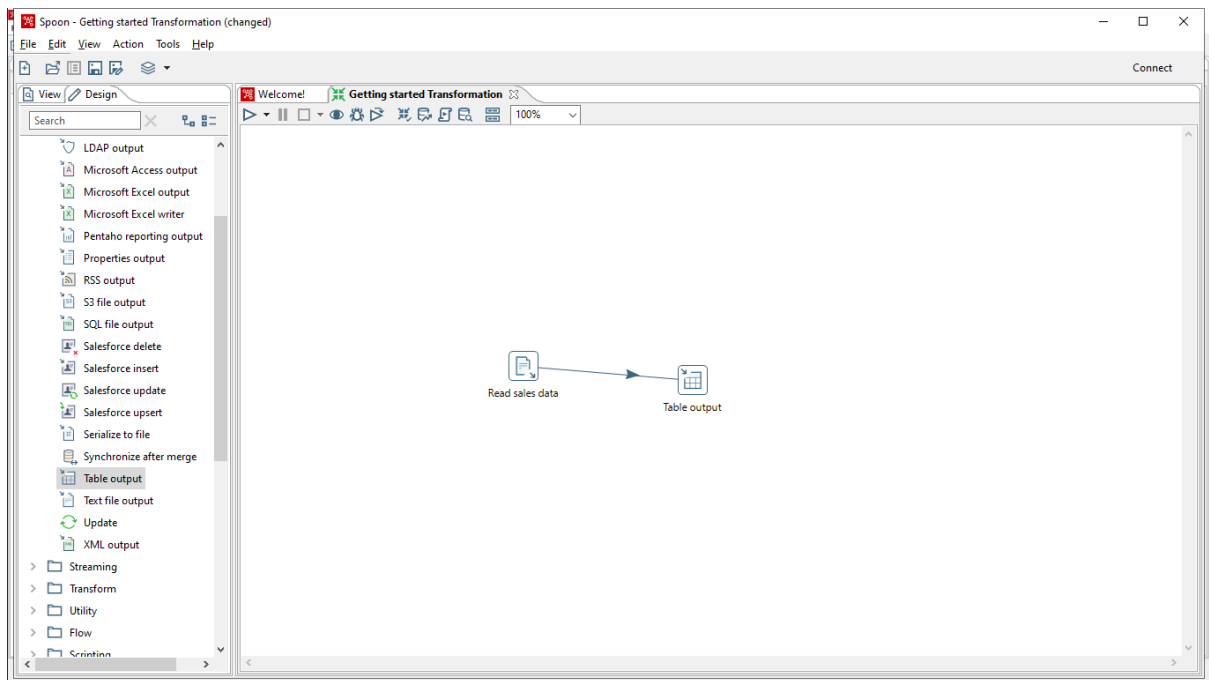
Klik OK lalu save



9. Setelah di-save, pada output, drag dan drop table output :



Gunakan tombol shift+K lalu tekan read sales data dan hubungkan dengan table output :



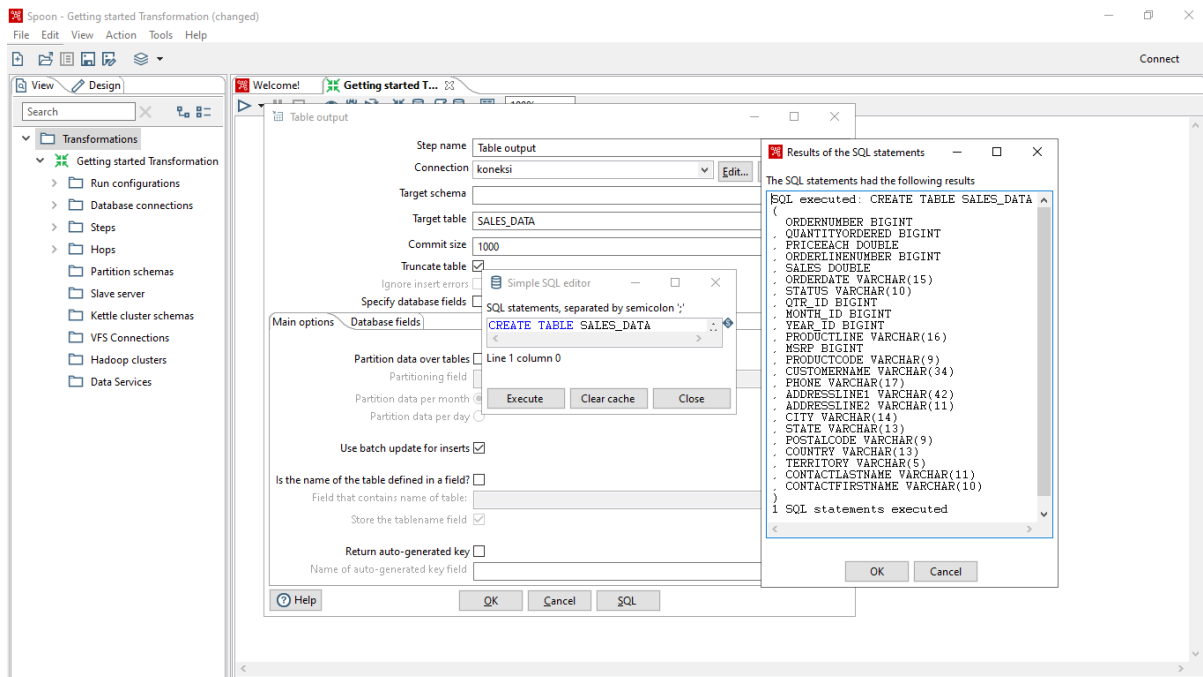
10. Double click table output tadi lalu ubah namanya. Setelah itu klik new pada bagian connection:

The screenshot shows the 'Table output' dialog box. The 'Step name' is 'Write to Database'. The 'Connection' dropdown is set to 'New...'. The 'Target schema' and 'Target table' fields are empty, with 'Browse...' buttons next to them. The 'Commit size' is set to '1000'. There are checkboxes for 'Truncate table', 'Ignore insert errors', and 'Specify database fields', all of which are currently unchecked. Below these is a tabbed interface with 'Main options' and 'Database fields' tabs. The 'Database fields' tab is active, showing options for 'Partition data over tables', 'Partitioning field', 'Partition data per month', 'Partition data per day', 'Use batch update for inserts' (checked), 'Is the name of the table defined in a field?' (unchecked), 'Field that contains name of table', 'Store the tablename field' (checked), 'Return auto-generated key' (unchecked), and 'Name of auto-generated key field'. At the bottom are 'Help', 'OK', 'Cancel', and 'SQL' buttons.

Isi nama, jenis koneksi, dan lainnya, saya menggunakan mysql dan database yang sudah ada di mysql untuk testing :

The screenshot shows the 'Database Connection' dialog box. The 'Connection name' is 'koneksi'. The 'Connection type' dropdown is set to 'MySQL'. The 'Settings' tab is active, showing fields for 'Host Name' (localhost), 'Database Name' (mahasiswa), 'Port Number' (3306), 'Username' (root), and 'Password'. There is a checkbox for 'Use Result Streaming Cursor' which is checked. At the bottom are 'Test', 'Feature List', 'Explore', 'OK', and 'Cancel' buttons. In the background, a 'Spoon - Getting started Transformation' window is visible, showing a 'Connection tested successfully' message with details: 'Hostname: localhost', 'Port: 3306', and 'Database name: mahasiswa'.

11. Pilih tombol SQL dan lakukan execute untuk memperlihatkan hasil SQL statements :



12. Keluar dari segala tampilan dan save progress untuk 'Extract and Load'

