

**Aim :-** ETL transformations with pentaho.

### **Theory :-**

What is Pentaho?

Pentaho is a Business Intelligence tool which provides a wide range of business intelligence solutions to the customers. It is capable of reporting, data analysis, data integration, data mining, etc. Pentaho also offers a comprehensive set of BI features which allows you to improve business performance and efficiency.

#### **Features of Pentaho**

Following, are important features of Pentaho:

- ETL capabilities for business intelligence needs
- Understanding Pentaho Report Designer
- Product Expertise
- Offers Side-by-side subreports
- Unlocking new capabilities
- Professional Support
- Query and Reporting
- Offers Enhanced Functionality
- Full runtime metadata support from data sources

#### **Introduction to Pentaho ETL tool**

The Data Integration perspective of Spoon allows you to create two basic file types: transformations and jobs.

1. **Transformations** are used to describe the data flows for ETL such as reading from a source, transforming data and loading it into a target location.
2. **Jobs** are used to coordinate ETL activities such as defining the flow and dependencies for what order transformations should be run, or prepare for execution by checking conditions such as, "Is my source file available?" or "Does a table exist in my database?"

## 5 Different Transformations using CSV and Microsoft Excel

Csv to excel:

**Execution Results**

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	36275	36276	0	0	0	0	Finished	0.3s	127,284	
2	Microsoft Excel Output	0	36275	36275	0	36275	0	0	0	Finished	0.6s	56,857	

Booking_ID	no_of_adul	no_of_chi	no_of_wed	type_of_m	required_c	room_type	lead_time	arrival_yea	arrival_mor	arrival_dat	market_se	repeated	no_of_prev	no_of_prev	avg_price	no_of_spe	booking_status
INN00001	2.00	.00	1.00	2.00 Meal Plan	.00	Room_Typ	224.00	2.017.00	10.00	2.00 Offline	.00	.00	.00	.00	65.00	.00	Not_Canceled
INN00002	2.00	.00	2.00	3.00 Not Select	.00	Room_Typ	5.00	2.018.00	11.00	6.00 Online	.00	.00	.00	.00	106.68	1.00	Not_Canceled
INN00003	1.00	.00	2.00	1.00 Meal Plan	.00	Room_Typ	1.00	2.018.00	2.00	28.00 Online	.00	.00	.00	.00	60.00	.00	Canceled
INN00004	2.00	.00	.00	2.00 Meal Plan	.00	Room_Typ	211.00	2.018.00	5.00	20.00 Online	.00	.00	.00	.00	100.00	.00	Canceled
INN00005	2.00	.00	1.00	1.00 Not Select	.00	Room_Typ	48.00	2.018.00	4.00	11.00 Online	.00	.00	.00	.00	94.50	.00	Canceled
INN00006	2.00	.00	.00	2.00 Meal Plan	.00	Room_Typ	346.00	2.018.00	9.00	13.00 Online	.00	.00	.00	.00	115.00	1.00	Canceled
INN00007	2.00	.00	1.00	3.00 Meal Plan	.00	Room_Typ	34.00	2.017.00	10.00	15.00 Online	.00	.00	.00	.00	107.55	1.00	Not_Canceled
INN00008	2.00	.00	1.00	3.00 Meal Plan	.00	Room_Typ	83.00	2.018.00	12.00	26.00 Online	.00	.00	.00	.00	105.61	1.00	Not_Canceled
INN00009	3.00	.00	.00	4.00 Meal Plan	.00	Room_Typ	121.00	2.018.00	7.00	6.00 Offline	.00	.00	.00	.00	96.90	1.00	Not_Canceled
INN00010	2.00	.00	.00	5.00 Meal Plan	.00	Room_Typ	44.00	2.018.00	10.00	18.00 Online	.00	.00	.00	.00	133.44	3.00	Not_Canceled
INN00011	1.00	.00	1.00	.00 Not Select	.00	Room_Typ	.00	2.018.00	9.00	11.00 Online	.00	.00	.00	.00	85.03	.00	Not_Canceled
INN00012	1.00	.00	2.00	1.00 Meal Plan	.00	Room_Typ	35.00	2.018.00	4.00	30.00 Online	.00	.00	.00	.00	140.40	1.00	Not_Canceled
INN00013	2.00	.00	2.00	1.00 Not Select	.00	Room_Typ	30.00	2.018.00	11.00	26.00 Online	.00	.00	.00	.00	88.00	.00	Canceled
INN00014	1.00	.00	2.00	.00 Meal Plan	.00	Room_Typ	95.00	2.018.00	11.00	20.00 Online	.00	.00	.00	.00	90.00	2.00	Canceled
INN00015	2.00	.00	.00	2.00 Meal Plan	.00	Room_Typ	47.00	2.017.00	10.00	20.00 Online	.00	.00	.00	.00	94.50	2.00	Not_Canceled
INN00016	2.00	.00	.00	2.00 Meal Plan	.00	Room_Typ	256.00	2.018.00	6.00	15.00 Online	.00	.00	.00	.00	115.00	1.00	Canceled
INN00017	1.00	.00	1.00	.00 Meal Plan	.00	Room_Typ	.00	2.017.00	10.00	5.00 Offline	.00	.00	.00	.00	96.00	.00	Not_Canceled
INN00018	2.00	.00	1.00	3.00 Not Select	.00	Room_Typ	1.00	2.017.00	8.00	10.00 Online	.00	.00	.00	.00	96.00	1.00	Not_Canceled
INN00019	2.00	.00	2.00	2.00 Meal Plan	.00	Room_Typ	99.00	2.017.00	10.00	30.00 Online	.00	.00	.00	.00	65.00	.00	Canceled
INN00020	2.00	.00	1.00	.00 Meal Plan	.00	Room_Typ	12.00	2.017.00	10.00	4.00 Offline	.00	.00	.00	.00	72.00	.00	Not_Canceled
INN00021	2.00	.00	2.00	2.00 Meal Plan	.00	Room_Typ	99.00	2.017.00	10.00	30.00 Online	.00	.00	.00	.00	65.00	.00	Canceled
INN00022	1.00	.00	.00	1.00 Meal Plan	.00	Room_Typ	122.00	2.018.00	11.00	25.00 Corporate	.00	.00	.00	.00	67.00	.00	Not_Canceled
INN00023	2.00	.00	2.00	4.00 Meal Plan	.00	Room_Typ	2.00	2.018.00	3.00	20.00 Offline	.00	.00	.00	.00	85.00	.00	Not_Canceled
INN00024	2.00	.00	.00	3.00 Meal Plan	.00	Room_Typ	37.00	2.018.00	10.00	13.00 Offline	.00	.00	.00	.00	105.00	.00	Not_Canceled
INN00025	2.00	.00	2.00	1.00 Not Select	.00	Room_Typ	130.00	2.018.00	5.00	22.00 Online	.00	.00	.00	.00	94.50	1.00	Not_Canceled
INN00026	2.00	.00	.00	2.00 Meal Plan	.00	Room_Typ	99.00	2.018.00	4.00	28.00 Online	.00	.00	.00	.00	114.30	1.00	Not_Canceled
INN00027	2.00	.00	1.00	1.00 Meal Plan	.00	Room_Typ	60.00	2.017.00	9.00	21.00 Offline	.00	.00	.00	.00	65.00	.00	Not_Canceled

**1)Sort Rows:**

It sorts rows based on the fields you specify and on whether they should be sorted in ascending or descending order.

For Sorting we will take csv input, Excel output and drag and drop Sort transformation.

Step name: Sort rows

Sort directory: %%java.io.tmpdir%% [Browse...](#)

TMP-file prefix: out

Sort size (rows in memory): 1000000

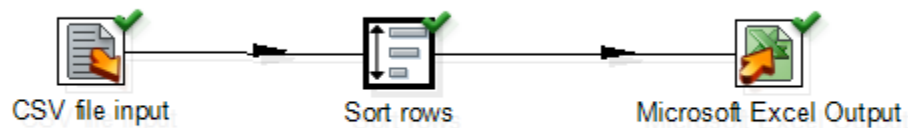
Free memory threshold (in %):

Compress TMP Files? ☐

Only pass unique rows? (verifies keys only) ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Presorted?
1	arrival_year	Y		
2				



Execution Results													
Execution History / Logging / Step Metrics / Performance Graph / Metrics / Preview data													
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	36275	36276	0	0	0	0	Finished	1.5s	24,314	
2	Sort rows	0	36275	36275	0	0	0	0	0	Finished	2.2s	16,474	
3	Microsoft Excel Output	0	36275	36275	0	36275	0	0	0	Finished	2.6s	13,888	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	Booking ID	no_of_adul	no_of_chi	no_of_wee	no_of_wee	type_of_m	required_c	room_type	lead_time	arrival_yea	arrival_mor	arrival_dat	market_seg	repeated_c	no_of_prev	no_of_prev	avg_price	no_of_s
2	INN00001	2	0	1	2	Meal Plan 1	0	Room_Type	224	2017	10	2	Offline	0	0	0	65	
3	INN00002	2	0	2	3	Not Selecte	0	Room_Type	5	2018	11	6	Online	0	0	0	106.68	
4	INN00003	1	0	2	1	Meal Plan 1	0	Room_Type	1	2018	2	28	Online	0	0	0	60	
5	INN00004	2	0	0	2	Meal Plan 1	0	Room_Type	211	2018	5	20	Online	0	0	0	100	
6	INN00005	2	0	1	1	Not Selecte	0	Room_Type	48	2018	4	11	Online	0	0	0	94.5	
7	INN00006	2	0	0	2	Meal Plan 2	0	Room_Type	346	2018	9	13	Online	0	0	0	115	
8	INN00007	2	0	1	3	Meal Plan 1	0	Room_Type	34	2017	10	15	Online	0	0	0	107.55	
9	INN00008	2	0	1	3	Meal Plan 1	0	Room_Type	83	2018	12	26	Online	0	0	0	105.61	
10	INN00009	3	0	0	4	Meal Plan 1	0	Room_Type	121	2018	7	6	Offline	0	0	0	96.9	
11	INN00010	2	0	0	5	Meal Plan 1	0	Room_Type	44	2018	10	18	Online	0	0	0	133.44	
12	INN00011	1	0	1	0	Not Selecte	0	Room_Type	0	2018	9	11	Online	0	0	0	85.03	
13	INN00012	1	0	2	1	Meal Plan 1	0	Room_Type	35	2018	4	30	Online	0	0	0	140.4	
14	INN00013	2	0	2	1	Not Selecte	0	Room_Type	30	2018	11	26	Online	0	0	0	88	
15	INN00014	1	0	2	0	Meal Plan 1	0	Room_Type	95	2018	11	20	Online	0	0	0	90	
16	INN00015	2	0	0	2	Meal Plan 1	0	Room_Type	47	2017	10	20	Online	0	0	0	94.5	
17	INN00016	2	0	0	2	Meal Plan 2	0	Room_Type	256	2018	6	15	Online	0	0	0	115	
18	INN00017	1	0	1	0	Meal Plan 1	0	Room_Type	0	2017	10	5	Offline	0	0	0	96	
19	INN00018	2	0	1	3	Not Selecte	0	Room_Type	1	2017	8	10	Online	0	0	0	96	
20	INN00019	2	0	2	2	Meal Plan 1	0	Room_Type	99	2017	10	30	Online	0	0	0	65	
21	INN00020	2	0	1	0	Meal Plan 1	0	Room_Type	12	2017	10	4	Offline	0	0	0	72	
22	INN00021	2	0	2	2	Meal Plan 1	0	Room_Type	99	2017	10	30	Online	0	0	0	65	
23	INN00022	1	0	0	1	Meal Plan 1	0	Room_Type	122	2018	11	25	Corporate	0	0	0	67	
24	INN00023	2	0	2	4	Meal Plan 1	0	Room_Type	2	2018	3	20	Offline	0	0	0	85	

Hotel Reservations

A1	fx	Booking_ID																	
1	INN00001	2.00	0.00	1.00	2.00	Meal Plan	0.00	Room_Type	224.00	2,017.00	10.00	2.00	Offline	0.00	0.00	0.00	65.00	0.00	Not_Canceled
3	INN00007	2.00	0.00	1.00	3.00	Meal Plan	0.00	Room_Type	34.00	2,017.00	10.00	15.00	Online	0.00	0.00	0.00	107.55	1.00	Not_Canceled
4	INN00015	2.00	0.00	0.00	2.00	Meal Plan	0.00	Room_Type	47.00	2,017.00	10.00	20.00	Online	0.00	0.00	0.00	94.50	2.00	Not_Canceled
5	INN00017	1.00	0.00	1.00	0.00	Meal Plan	0.00	Room_Type	0.00	2,017.00	10.00	5.00	Offline	0.00	0.00	0.00	96.00	0.00	Not_Canceled
6	INN00018	2.00	0.00	1.00	3.00	Not Select	0.00	Room_Type	1.00	2,017.00	8.00	10.00	Online	0.00	0.00	0.00	96.00	1.00	Not_Canceled
7	INN00019	2.00	0.00	2.00	2.00	Meal Plan	0.00	Room_Type	99.00	2,017.00	10.00	30.00	Online	0.00	0.00	0.00	65.00	0.00	Canceled
8	INN00020	2.00	0.00	1.00	0.00	Meal Plan	0.00	Room_Type	12.00	2,017.00	10.00	4.00	Offline	0.00	0.00	0.00	72.00	0.00	Not_Canceled
9	INN00021	2.00	0.00	2.00	2.00	Meal Plan	0.00	Room_Type	99.00	2,017.00	10.00	30.00	Online	0.00	0.00	0.00	65.00	0.00	Canceled
10	INN00027	2.00	0.00	1.00	1.00	Meal Plan	0.00	Room_Type	60.00	2,017.00	9.00	21.00	Offline	0.00	0.00	0.00	65.00	0.00	Not_Canceled
11	INN00029	1.00	0.00	1.00	2.00	Meal Plan	0.00	Room_Type	37.00	2,017.00	11.00	6.00	Online	0.00	0.00	0.00	37.33	0.00	Canceled
12	INN00030	2.00	0.00	0.00	2.00	Meal Plan	0.00	Room_Type	56.00	2,017.00	9.00	17.00	Offline	0.00	0.00	0.00	82.00	0.00	Not_Canceled
13	INN00031	2.00	0.00	1.00	1.00	Meal Plan	0.00	Room_Type	3.00	2,017.00	9.00	19.00	Online	0.00	0.00	0.00	177.50	1.00	Not_Canceled
14	INN00038	2.00	0.00	0.00	1.00	Meal Plan	0.00	Room_Type	289.00	2,017.00	10.00	17.00	Offline	0.00	0.00	0.00	67.00	0.00	Not_Canceled
15	INN00045	2.00	0.00	0.00	4.00	Meal Plan	0.00	Room_Type	55.00	2,017.00	11.00	18.00	Offline	0.00	0.00	0.00	60.00	1.00	Not_Canceled
16	INN00046	2.00	0.00	0.00	3.00	Meal Plan	0.00	Room_Type	146.00	2,017.00	12.00	29.00	Offline	0.00	0.00	0.00	75.00	0.00	Not_Canceled
17	INN00047	1.00	0.00	0.00	2.00	Meal Plan	0.00	Room_Type	32.00	2,017.00	11.00	20.00	Offline	0.00	0.00	0.00	73.00	0.00	Not_Canceled
18	INN00054	2.00	0.00	0.00	4.00	Meal Plan	0.00	Room_Type	51.00	2,017.00	11.00	11.00	Offline	0.00	0.00	0.00	60.00	1.00	Not_Canceled
19	INN00064	1.00	0.00	0.00	1.00	Meal Plan	0.00	Room_Type	2.00	2,017.00	9.00	10.00	Compleme	0.00	0.00	0.00	0.00	1.00	Not_Canceled
20	INN00075	2.00	0.00	2.00	3.00	Meal Plan	0.00	Room_Type	34.00	2,017.00	10.00	25.00	Offline	0.00	0.00	0.00	75.00	0.00	Not_Canceled
21	INN00104	2.00	0.00	0.00	1.00	Meal Plan	0.00	Room_Type	16.00	2,017.00	10.00	31.00	Online	0.00	0.00	0.00	95.00	1.00	Not_Canceled
22	INN00105	2.00	0.00	0.00	3.00	Meal Plan	0.00	Room_Type	24.00	2,017.00	10.00	8.00	Online	0.00	0.00	0.00	107.00	1.00	Not_Canceled
23	INN00107	2.00	0.00	1.00	4.00	Meal Plan	1.00	Room_Type	10.00	2,017.00	11.00	2.00	Online	0.00	0.00	0.00	82.95	0.00	Not_Canceled
24	INN00108	2.00	0.00	1.00	3.00	Meal Plan	0.00	Room_Type	0.00	2,017.00	9.00	21.00	Online	0.00	0.00	0.00	155.00	1.00	Not_Canceled
25	INN00110	2.00	0.00	2.00	1.00	Meal Plan	0.00	Room_Type	32.00	2,017.00	9.00	5.00	Online	0.00	0.00	0.00	94.50	3.00	Not_Canceled
26	INN00120	2.00	0.00	2.00	3.00	Meal Plan	0.00	Room_Type	2.00	2,017.00	11.00	14.00	Online	0.00	0.00	0.00	103.00	1.00	Not_Canceled
27	INN00134	2.00	0.00	1.00	4.00	Meal Plan	0.00	Room_Type	28.00	2,017.00	10.00	19.00	Offline	0.00	0.00	0.00	75.00	0.00	Not_Canceled
28	INN00136	2.00	0.00	0.00	1.00	Meal Plan	0.00	Room_Type	11.00	2,017.00	10.00	9.00	Online	0.00	0.00	0.00	155.00	2.00	Not_Canceled

## 2) Calculator:

The Calculator step provides you with predefined functions that you can execute on input field values. To use, specify the input fields and type of function to perform and return results.

For calculating we will take csv from input, Excel from output and drag and drop calculator transformation.

Calculator

Step name

Calculator

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove	Conversion mask	Decimal symbol	Grouping symbol	Currency symbol
1	Sum	A + B	no_of_a...	avg_pric...									

Help

OK

Cancel

CSV file input

Calculator

Microsoft Excel Output

Execution Results

Execution History

Logging

Step Metrics

Performance Graph

Metrics

Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	36275	36276	0	0	0	0	Finished	0.1s	453,450	
2	Calculator	0	36275	36275	0	0	0	0	0	Finished	0.2s	153,059	
3	Microsoft Excel Output	0	36275	36275	0	36275	0	0	0	Finished	0.5s	75,259	

A1

fx

Booking\_ID

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		no_of_adul	no_of_chi	no_of_w	no_of_w	type_of_m	required_c	room_type	lead time	arrival_yea	arrival_mor	arrival_dati	market_se	repeated	no_of_prev	no_of_prev	avg_price	no_of_spe	booking_sl	Sum	
2	2.00	0.00	1.00	2.00	Meal Plan	0.00	Room_Type	224.00	2.017.00	10.00	2.00	Offline	0.00	0.00	0.00	65.00	0.00	Not_Cance	67.00		
3	2.00	0.00	2.00	3.00	Not Select	0.00	Room_Type	5.00	2.018.00	11.00	6.00	Online	0.00	0.00	0.00	106.68	1.00	Not_Cance	109.00		
4	1.00	0.00	2.00	1.00	Meal Plan	0.00	Room_Type	1.00	2.018.00	2.00	28.00	Online	0.00	0.00	0.00	60.00	0.00	Canceled	61.00		
5	2.00	0.00	0.00	2.00	Meal Plan	0.00	Room_Type	211.00	2.018.00	5.00	20.00	Online	0.00	0.00	0.00	100.00	0.00	Canceled	102.00		
6	2.00	0.00	1.00	1.00	Not Select	0.00	Room_Type	48.00	2.018.00	4.00	11.00	Online	0.00	0.00	0.00	94.50	0.00	Canceled	97.00		
7	2.00	0.00	0.00	2.00	Meal Plan	0.00	Room_Type	346.00	2.018.00	9.00	13.00	Online	0.00	0.00	0.00	115.00	1.00	Canceled	117.00		
8	2.00	0.00	1.00	3.00	Meal Plan	0.00	Room_Type	34.00	2.017.00	10.00	15.00	Online	0.00	0.00	0.00	107.55	1.00	Not_Cance	110.00		
9	2.00	0.00	1.00	3.00	Meal Plan	0.00	Room_Type	83.00	2.018.00	12.00	26.00	Online	0.00	0.00	0.00	105.61	1.00	Not_Cance	108.00		
10	3.00	0.00	0.00	4.00	Meal Plan	0.00	Room_Type	121.00	2.018.00	7.00	6.00	Offline	0.00	0.00	0.00	96.90	1.00	Not_Cance	100.00		
11	2.00	0.00	0.00	5.00	Meal Plan	0.00	Room_Type	44.00	2.018.00	10.00	18.00	Online	0.00	0.00	0.00	133.44	3.00	Not_Cance	135.00		
12	1.00	0.00	1.00	0.00	Not Select	0.00	Room_Type	0.00	2.018.00	9.00	11.00	Online	0.00	0.00	0.00	85.03	0.00	Not_Cance	86.00		
13	1.00	0.00	2.00	1.00	Meal Plan	0.00	Room_Type	35.00	2.018.00	4.00	30.00	Online	0.00	0.00	0.00	140.40	1.00	Not_Cance	141.00		
14	2.00	0.00	2.00	1.00	Not Select	0.00	Room_Type	30.00	2.018.00	11.00	26.00	Online	0.00	0.00	0.00	88.00	0.00	Canceled	90.00		
15	1.00	0.00	2.00	0.00	Meal Plan	0.00	Room_Type	95.00	2.018.00	11.00	20.00	Online	0.00	0.00	0.00	90.00	2.00	Canceled	91.00		
16	2.00	0.00	0.00	2.00	Meal Plan	0.00	Room_Type	47.00	2.017.00	10.00	20.00	Online	0.00	0.00	0.00	94.50	2.00	Not_Cance	97.00		
17	2.00	0.00	0.00	2.00	Meal																

**3)Concatenate:**

This transformation is used to concatenate two different fields

**Concat Fields**

Step name: Concat Fields

Target Field Name: Market Booking

Length of Target Field: 0

Separator: ;

Enclosure: "

Insert TAB

Fields Advanced

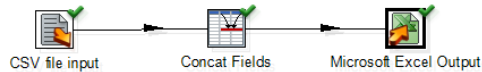
#	Name	Type	Format	Length	Precision	Currency	Decimal	Gro
1	market_segment_type							
2	booking_status							

Get Fields Minimal width

Help

OK

Cancel

**Execution Results**

Execution History Logging Step Metrics Performance Graph Metrics Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	36275	36276	0	0	0	0	Finished	0.1s	518,228	
2	Concat Fields	0	36275	36275	0	0	0	0	0	Finished	0.2s	174,399	
3	Microsoft Excel Output	0	36275	36275	0	36275	0	0	0	Finished	0.4s	83,583	

	A1		Q fx		Booking_ID																			
#	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U				
1	no_of_adul	no_of_chi	no_of_w	no_of_w	type_of_m	required_c	room_type	lead_time	arrival_yea	arrival_mor	arrival_data	market_se	repeated	no_of_prev	no_of_prev	avg_price	no_of_spe	booking_sl	Market Booking					
2	2.00	.00	1.00	2.00	Meal Plan	.00	Room_Typ	224.00	2.017.00	10.00	2.00	Offline	.00	.00	.00	65.00	.00	Not_Cancel	Offline;Not_Canceled					
3	2.00	.00	2.00	3.00	Not Select	.00	Room_Typ	5.00	2.018.00	11.00	6.00	Online	.00	.00	.00	106.68	1.00	Not_Cancel	Online;Not_Canceled					
4	1.00	.00	2.00	1.00	Meal Plan	.00	Room_Typ	1.00	2.018.00	2.00	28.00	Online	.00	.00	.00	60.00	.00	Canceled	Online;Canceled					
5	2.00	.00	.00	2.00	Meal Plan	.00	Room_Typ	211.00	2.018.00	5.00	20.00	Online	.00	.00	.00	100.00	.00	Canceled	Online;Canceled					
6	2.00	.00	1.00	1.00	Not Select	.00	Room_Typ	48.00	2.018.00	4.00	11.00	Online	.00	.00	.00	94.50	.00	Canceled	Online;Canceled					
7	2.00	.00	.00	2.00	Meal Plan	.00	Room_Typ	346.00	2.018.00	9.00	13.00	Online	.00	.00	.00	115.00	1.00	Canceled	Online;Canceled					
8	2.00	.00	1.00	3.00	Meal Plan	.00	Room_Typ	34.00	2.017.00	10.00	15.00	Online	.00	.00	.00	107.55	1.00	Not_Cancel	Online;Not_Canceled					
9	2.00	.00	1.00	3.00	Meal Plan	.00	Room_Typ	83.00	2.018.00	12.00	26.00	Online	.00	.00	.00	105.61	1.00	Not_Cancel	Online;Not_Canceled					
10	3.00	.00	.00	4.00	Meal Plan	.00	Room_Typ	121.00	2.018.00	7.00	6.00	Offline	.00	.00	.00	96.90	1.00	Not_Cancel	Offline;Not_Canceled					
11	2.00	.00	.00	5.00	Meal Plan	.00	Room_Typ	44.00	2.018.00	10.00	18.00	Online	.00	.00	.00	133.44	3.00	Not_Cancel	Online;Not_Canceled					
12	1.00	.00	1.00	.00	Not Select	.00	Room_Typ	.00	2.018.00	9.00	11.00	Online	.00	.00	.00	85.03	.00	Not_Cancel	Online;Not_Canceled					
13	1.00	.00	2.00	1.00	Meal Plan	.00	Room_Typ	35.00	2.018.00	4.00	30.00	Online	.00	.00	.00	140.40	1.00	Not_Cancel	Online;Not_Canceled					
14	2.00	.00	2.00	1.00	Not Select	.00	Room_Typ	30.00	2.018.00	11.00	26.00	Online	.00	.00	.00	88.00	.00	Canceled	Online;Canceled					
15	1.00	.00	2.00	.00	Meal Plan	.00	Room_Typ	95.00	2.018.00	11.00	20.00	Online	.00	.00	.00	90.00	2.00	Canceled	Online;Canceled					
16	2.00	.00	.00	2.00	Meal Plan	.00	Room_Typ	47.00	2.017.00	10.00	20.00	Online	.00	.00	.00	94.50	2.00	Not_Cancel	Online;Not_Canceled					
17	2.00	.00	.00	2.00	Meal Plan	.00	Room_Typ	256.00	2.018.00	6.00	15.00	Online	.00	.00	.00	115.00	1.00	Canceled	Online;Canceled					
18	1.00	.00	1.00	.00	Meal Plan	.00	Room_Typ	.00	2.017.00	10.00	5.00	Offline	.00	.00	.00	96.00	.00	Not_Cancel	Offline;Not_Canceled					
19	2.00	.00	1.00	3.00	Not Select	.00	Room_Typ	1.00	2.017.00	8.00	10.00	Online	.00	.00	.00	96.00	1.00	Not_Cancel	Online;Not_Canceled					
20	2.00	.00	2.00	2.00	Meal Plan	.00	Room_Typ	99.00	2.017.00	10.00	30.00	Online	.00	.00	.00	65.00	.00	Canceled	Online;Canceled					
21	2.00	.00	1.00	.00	Meal Plan	.00	Room_Typ	12.00	2.017.00	10.00	4.00	Offline	.00	.00	.00	72.00	.00	Not_Cancel	Offline;Not_Canceled					
22	2.00	.00	2.00	2.00	Meal Plan	.00	Room_Typ	99.00	2.017.00	10.00	30.00	Online	.00	.00	.00	65.00	.00	Canceled	Online;Canceled					
23	1.00	.00	.00	1.00	Meal Plan	.00	Room_Typ	122.00	2.018.00	11.00	25.00	Corporate	.00	.00	.00	67.00	.00	Not_Cancel	Corporate;Not_Canceled					
24	2.00	.00	2.00	4.00	Meal Plan	.00	Room_Typ	2.00	2.018.00	3.00	20.00	Offline	.00	.00	.00	85.00	.00	Not_Cancel	Offline;Not_Canceled					
25	2.00	.00	.00	3.00	Meal Plan	.00	Room_Typ	37.00	2.018.00	10.00	13.00	Offline	.00	.00	.00	105.00	.00	Not_Cancel	Offline;Not_Canceled					
26	2.00	.00	2.00	1.00	Not Select	.00	Room_Typ	130.00	2.018.00	5.00	22.00	Online	.00	.00	.00	94.50	1.00	Not_Cancel	Online;Not_Canceled					
27	2.00	.00	.00	2.00	Meal Plan	.00	Room_Typ	99.00	2.018.00	4.00	28.00	Online	.00	.00	.00	114.30	1.00	Not_Cancel	Online;Not_Canceled					
28	2.00	.00	1.00	1.00	Meal Plan	.00	Room_Typ	60.00	2.017.00	9.00	21.00	Offline	.00	.00	.00	65.00	.00	Not_Cancel	Offline;Not_Canceled					

#### 4)Split field:

Split Fields component split the value of a variable into multiple variables based on delimiter. For example, consider a variable var1 having value a;b;c. Now i need to split into three different values a,b and c. In order to achieve this, use split Fields component in Pentaho with delimiter as “;”

The screenshot displays the Pentaho Data Integration (Spoon) interface. The top window, titled 'Field splitter', shows the configuration for the 'Split Fields' step. The 'Field to split' is set to 'type\_of\_meal\_plan', the 'Delimiter' is a semicolon (;), and the 'Enclosure' is empty. Below this is a table for defining the output fields.

#	New field	ID	Remove ID?	Type	Length	Precision	Format	Group	Decimal	Currency	Nullif	Default	Trim type
1	newfield		N	String									none

The bottom window shows the 'Spoon - transform3' interface with a design view. The workflow consists of three steps: 'CSV file input' → 'Split Fields' → 'Microsoft Excel Output'. The 'Execution Results' tab is active, showing the following data:

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	36275	36276	0	0	0	0	Finished	0.1s	525,739	
2	Split Fields	0	36275	36275	0	0	0	0	0	Finished	0.1s	259,107	
3	Microsoft Excel Output	0	36275	36275	0	36275	0	0	0	Finished	0.4s	103,643	

WPS Office

S Hotel Reservations.csv

file4.xlsx

file5.xlsx

**5)Row flattener:**

The Row flattener step allows you to flatten sequentially provided data. Use this step when you have a file organized in consecutive rows where each row is an attribute, and you want to structure those rows of data into columns.

**Flattener**

Step name:

The field to:

The target fields after flattening:

#	Target field
1	newfield

OK Cancel



The screenshot displays the SAP Data Integration Studio interface. The top menu bar includes File, Edit, View, Action, Tools, and Help. The main workspace shows a data flow diagram with three components: CSV file input, Row flattener, and Microsoft Excel Output. The Execution Results table at the bottom provides a detailed view of the data load process.

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	CSV file input	0	0	36275	36276	0	0	0	0	Finished	0.1s	510,930	
2	Row flattener	0	36275	36275	0	0	0	0	0	Finished	0.1s	255,458	
3	Microsoft Excel Output	0	36275	36275	0	36275	0	0	0	Finished	0.3s	123,384	

The bottom of the image shows a preview of the data loaded into the Excel output, with columns A through T. The data includes booking IDs, arrival dates, room types, and various status indicators.

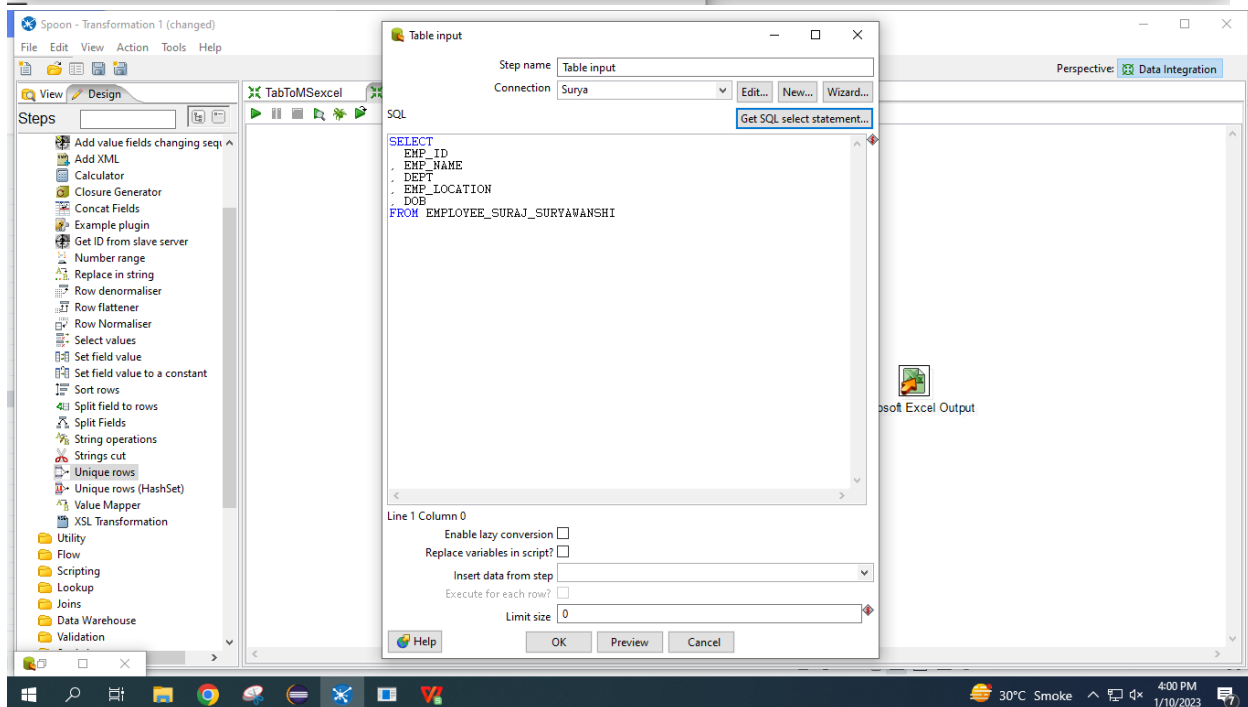
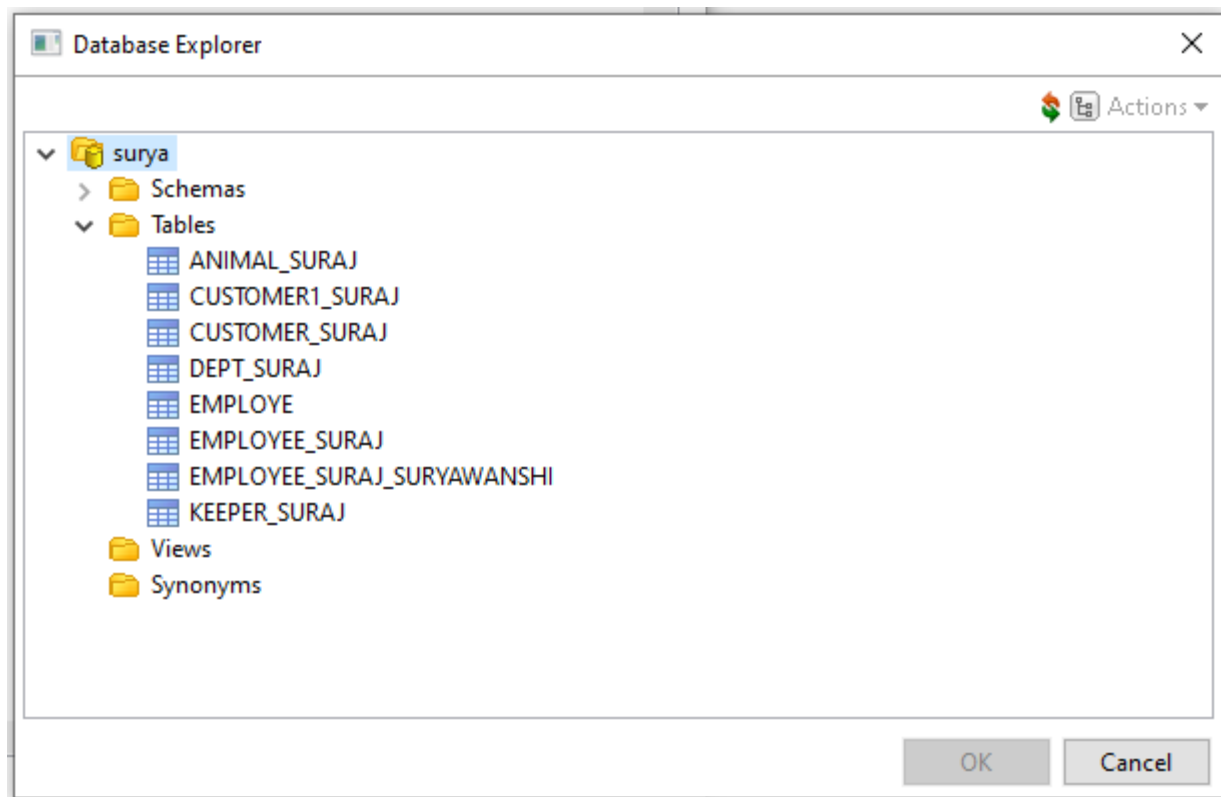
For Database table input to Microsoft Excel output with different type of transformation:

Create database connection:

The screenshot shows the 'Database Connection' dialog box with the following configuration:

- General** tab is selected in the left sidebar.
- Connection Name:** surya
- Connection Type:** Oracle (selected from the list: Netezza, OpenERP Server, Oracle, Oracle RDB, Palo MOLAP Server, PostgreSQL, Remedy Action Request System, SAP ERP System, SQLite, Sybase, SybaseIQ, Teradata, UniVerse database, Vertica, Vertica 5+, dBase III, IV or 5)
- Access:** Native (JDBC) (selected from the list: Native (JDBC), ODBC, OCI, JNDI)
- Settings:**
  - Host Name:** mcaserver
  - Database Name:** sem1
  - Tablespace for Data:** (empty)
  - Tablespace for Indices:** (empty)
  - Port Number:** 1521
  - User Name:** user60b
  - Password:** (masked with asterisks)

Buttons at the bottom: Test, Feature List, Explore, OK, Cancel.





The screenshot shows the Spoon - TabToMSExcel application window. The main design canvas displays a workflow with three steps: 'Table input', 'Sort rows', and 'Microsoft Excel Output', connected sequentially. The 'Execution Results' tab is active, showing a table with the following data:

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	Table input	0	0	6	6	0	0	0	0	Finished	0.0s	1,000	
2	Sort rows	0	6	6	0	0	0	0	0	Finished	0.0s	162	
3	Microsoft Excel Output	0	6	6	0	6	0	0	0	Finished	0.3s	17	

## Table Output-

Examine preview data

Rows of step: Microsoft Excel Output (6 rows)

#	EMP_ID	EMP_NAME	DEPT	EMP_LOCATION	DOB	SALARY	COMM
1	1	Suraj	A	Mumbai	2001-02-19 00:00:00.000000000	52000	900
2	2	Shiva	A	Pune	2001-01-21 00:00:00.000000000	32000	700
3	3	Shreeya	B	Pune	2001-03-21 00:00:00.000000000	42000	600
4	4	Dwyane	A	Chennai	2001-06-23 00:00:00.000000000	52000	200
5	5	John	B	Delhi	2002-06-23 00:00:00.000000000	33000	300
6	6	Kelvin	C	Banglore	2002-07-25 00:00:00.000000000	33000	300

**2) Replace in String:****Input-**

Examine preview data

Rows of step: Add sequence (7 rows)

#	EMP_ID	EMP_NAME	DEPT	EMP_LOCATION	DOB	SALARY	COMM	Sequence
1	5	John	B	Delhi	2002-06-23 00:00:00.000000000	33000	300	1
2	1	Suraj	A	Mumbai	2001-02-19 00:00:00.000000000	52000	900	2
3	2	Shiva	A	Pune	2001-01-21 00:00:00.000000000	32000	700	3
4	3	Shreeya	B	Pune	2001-03-21 00:00:00.000000000	42000	600	4
5	4	Dwyane	A	Chennai	2001-06-23 00:00:00.000000000	52000	200	5
6	5	John	B	Delhi	2001-06-23 00:00:00.000000000	22000	300	6
7	6	Kelvin	C	Banglore	2002-07-25 00:00:00.000000000	33000	300	7

**Transformation-**

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	EMP_NAME		N	John	Atharva	N		N	N
2	DEPT		N			N		N	N
3	EMP_LOCATION		N			N		N	N

Help OK Get fields Cancel

**Output-****Execution Results**

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

☒ First rows ☐ Last rows ☐ Off

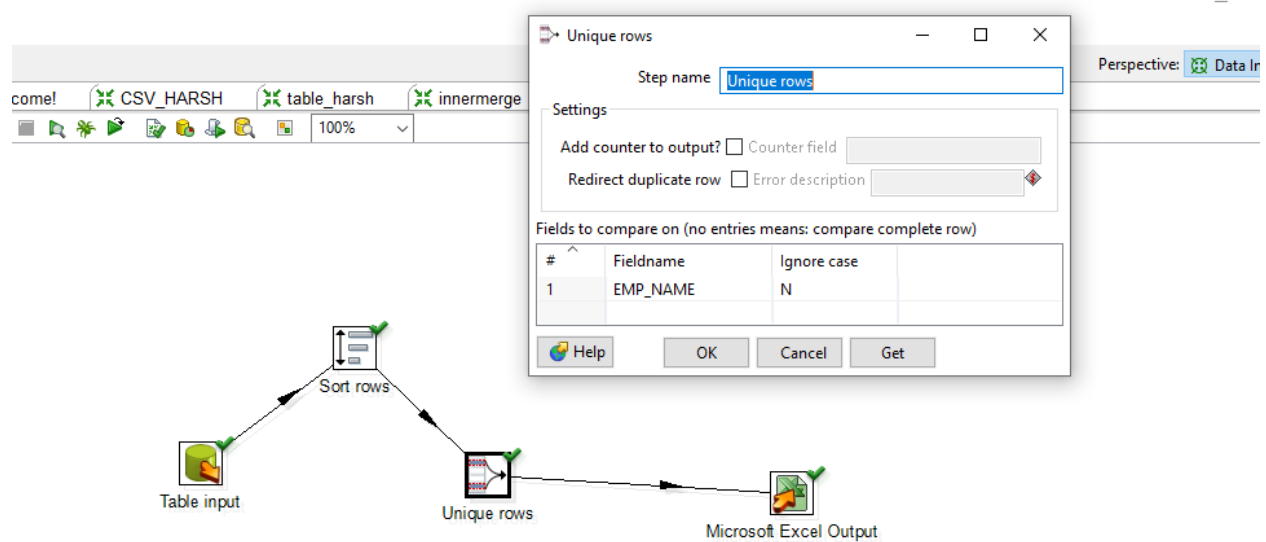
#	EMP_ID	EMP_NAME	DEPT	EMP_LOCATION	DOB	SALARY	COMM
1	5	Atharva	B	Delhi	2002-06-23 00:00:00.000000000	33000	300
2	1	Suraj	A	Mumbai	2001-02-19 00:00:00.000000000	52000	900
3	2	Shiva	A	Pune	2001-01-21 00:00:00.000000000	32000	700
4	3	Shreeya	B	Pune	2001-03-21 00:00:00.000000000	42000	600
5	4	Dwyane	A	Chennai	2001-06-23 00:00:00.000000000	52000	200
6	5	Atharva	B	Delhi	2001-06-23 00:00:00.000000000	22000	300
7	6	Kelvin	C	Banglore	2002-07-25 00:00:00.000000000	33000	300

**3)Unique Rows:****Input-**

Examine preview data

Rows of step: Table input (7 rows)


#	EMP_ID	EMP_NAME	DEPT	EMP_LOCATION	DOB	SALARY	COMM
1	5	John	B	Delhi	2002-06-23 00:00:00.000000000	33000	300
2	1	Suraj	A	Mumbai	2001-02-19 00:00:00.000000000	52000	900
3	2	Shiva	A	Pune	2001-01-21 00:00:00.000000000	32000	700
4	3	Shreeya	B	Pune	2001-03-21 00:00:00.000000000	42000	600
5	4	Dwyane	A	Chennai	2001-06-23 00:00:00.000000000	52000	200
6	5	John	B	Delhi	2001-06-23 00:00:00.000000000	22000	300
7	6	Kelvin	C	Banglore	2002-07-25 00:00:00.000000000	33000	300

**Transformation****Output-**

Examine preview data

Rows of step: Microsoft Excel Output (6 rows)

#	EMP_ID	EMP_NAME	DEPT	EMP_LOCATION	DOB	SALARY	COMM
1	1	Suraj	A	Mumbai	2001-02-19 00:00:00.000000000	52000	900
2	2	Shiva	A	Pune	2001-01-21 00:00:00.000000000	32000	700
3	3	Shreeya	B	Pune	2001-03-21 00:00:00.000000000	42000	600
4	4	Dwyane	A	Chennai	2001-06-23 00:00:00.000000000	52000	200
5	5	John	B	Delhi	2002-06-23 00:00:00.000000000	33000	300
6	6	Kelvin	C	Banglore	2002-07-25 00:00:00.000000000	33000	300

**4) Sequence:****Input-** Examine preview data

Rows of step: Table input (7 rows)

#	EMP_ID	EMP_NAME	DEPT	EMP_LOCATION	DOB	SALARY	COMM
1	5	John	B	Delhi	2002-06-23 00:00:00.000000000	33000	300
2	1	Suraj	A	Mumbai	2001-02-19 00:00:00.000000000	52000	900
3	2	Shiva	A	Pune	2001-01-21 00:00:00.000000000	32000	700
4	3	Shreeya	B	Pune	2001-03-21 00:00:00.000000000	42000	600
5	4	Dwyane	A	Chennai	2001-06-23 00:00:00.000000000	52000	200
6	5	John	B	Delhi	2001-06-23 00:00:00.000000000	22000	300
7	6	Kelvin	C	Banglore	2002-07-25 00:00:00.000000000	33000	300

**Transformation-**

Concat Fields

Step name

Concat Fields

Target Field Name

Full Name

Length of Target Field

0

Separator

,

Enclosure

"

Insert TAB

Fields

Advanced

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim Type	Null
1	EMPNAME	String		15						
2	EMPLNAME	String		15						

Get Fields


Minimal width

Help

OK

Cancel

**Output-**

 Examine preview data

Rows of step: Add sequence (7 rows)

#	EMP_ID	EMP_NAME	DEPT	EMP_LOCATION	DOB	SALARY	COMM	Sequence
1	5	John	B	Delhi	2002-06-23 00:00:00.000000000	33000	300	1
2	1	Suraj	A	Mumbai	2001-02-19 00:00:00.000000000	52000	900	2
3	2	Shiva	A	Pune	2001-01-21 00:00:00.000000000	32000	700	3
4	3	Shreeya	B	Pune	2001-03-21 00:00:00.000000000	42000	600	4
5	4	Dwyane	A	Chennai	2001-06-23 00:00:00.000000000	52000	200	5
6	5	John	B	Delhi	2001-06-23 00:00:00.000000000	22000	300	6
7	6	Kelvin	C	Banglore	2002-07-25 00:00:00.000000000	33000	300	7



## 5) Concat Fields

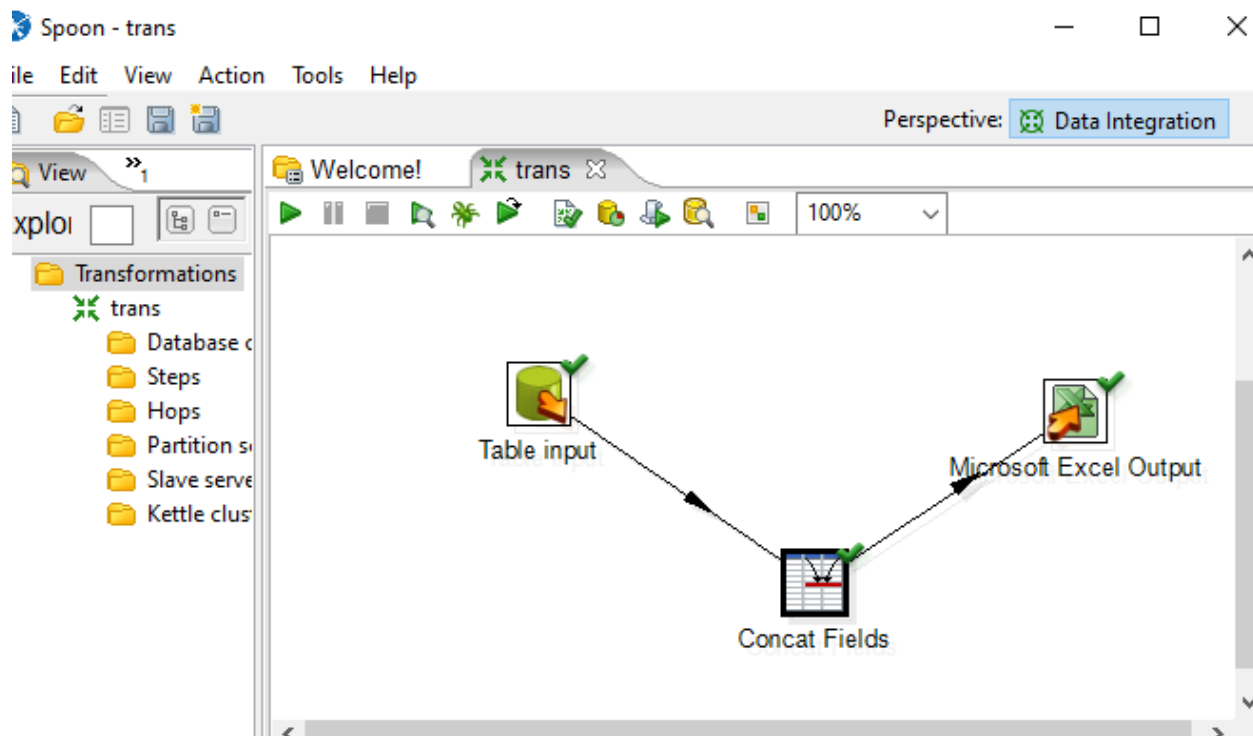
**Input-**

 [Examine preview data](#)

Rows of step: Table input (7 rows)

[illegible]

**Transformation:**



Concat Fields

Step name: Concat Fields

Target Field Name: ID-DEPT

Length of Target Field: 0

Separator: , Insert TAB

Enclosure:

Fields | Advanced

#	Name	Type	Format	Length
1	EMP_ID	Integer		
2	DEPT	Integer		
3		None		

< Get Fields Minimal width >

OK Cancel

Help

## Output-

Execution Results

Execution History

Logging

Step Metrics

Performance Graph

Metrics

Preview data

☒ First rows
 ☐ Last rows
 ☐ Off

#	EMP_ID	EMP_NAME	DEPT	EMP_LOCATION	DOB	SALARY	COMM	ID-DEPT
1	5	John	B	Delhi	2002-06-23 00:00:00.000000000	33000	300	5,B
2	1	Suraj	A	Mumbai	2001-02-19 00:00:00.000000000	52000	900	1,A
3	2	Shiva	A	Pune	2001-01-21 00:00:00.000000000	32000	700	2,A
4	3	Shreeya	B	Pune	2001-03-21 00:00:00.000000000	42000	600	3,B
5	4	Dwyane	A	Chennai	2001-06-23 00:00:00.000000000	52000	200	4,A
6	5	John	B	Delhi	2001-06-23 00:00:00.000000000	22000	300	5,B
7	6	Kelvin	C	Banglore	2002-07-25 00:00:00.000000000	33000	300	6,C