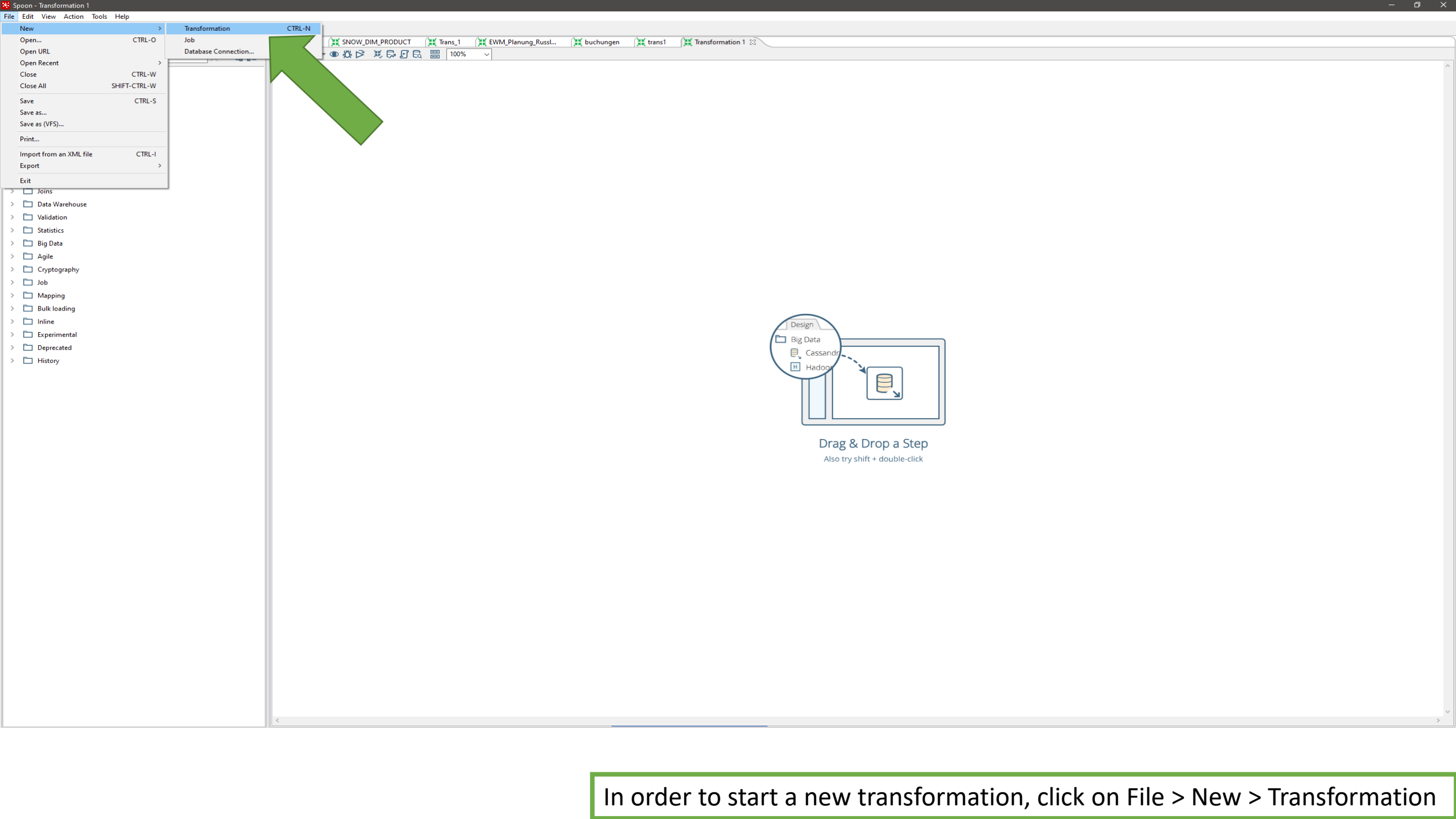
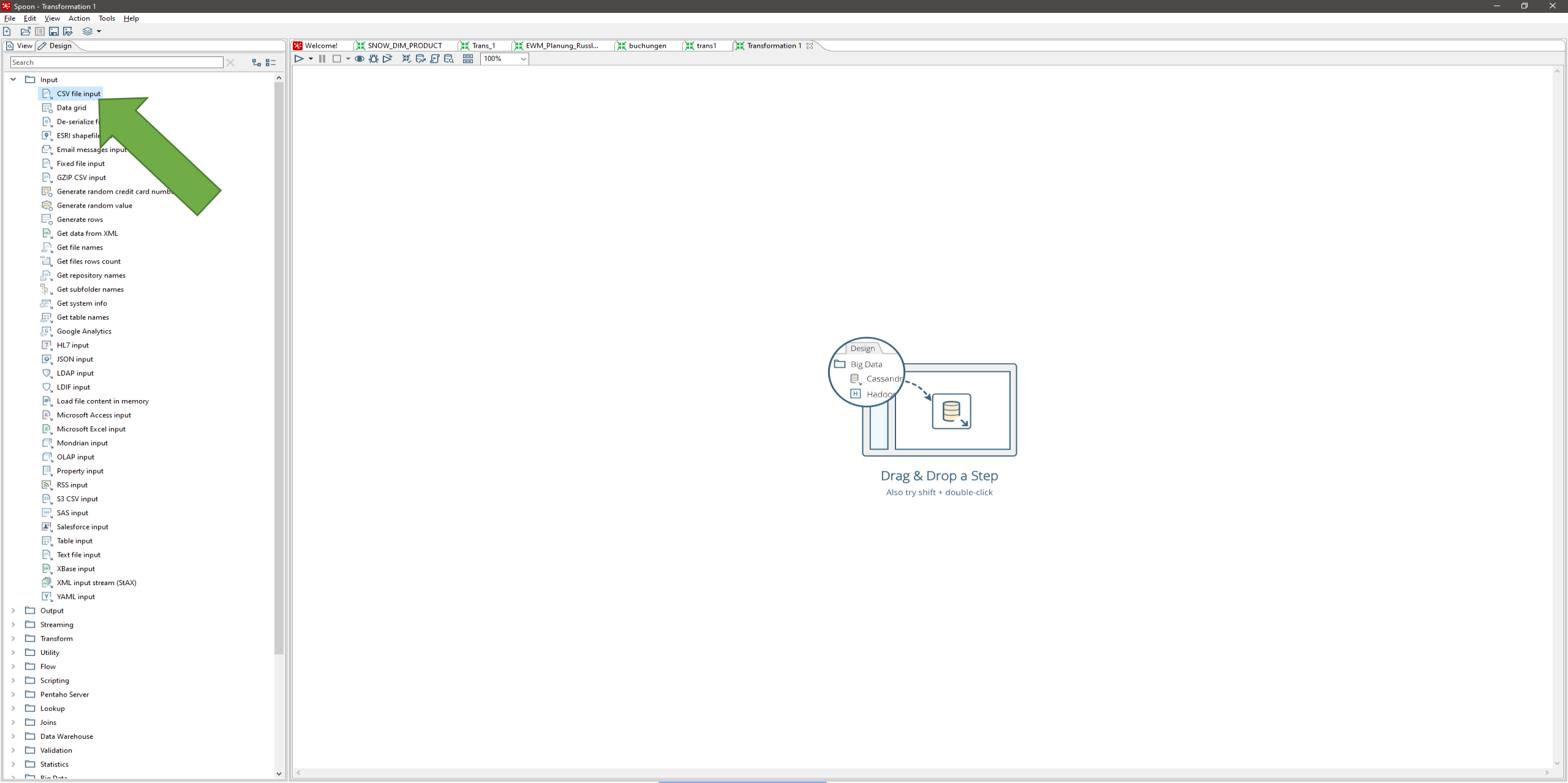
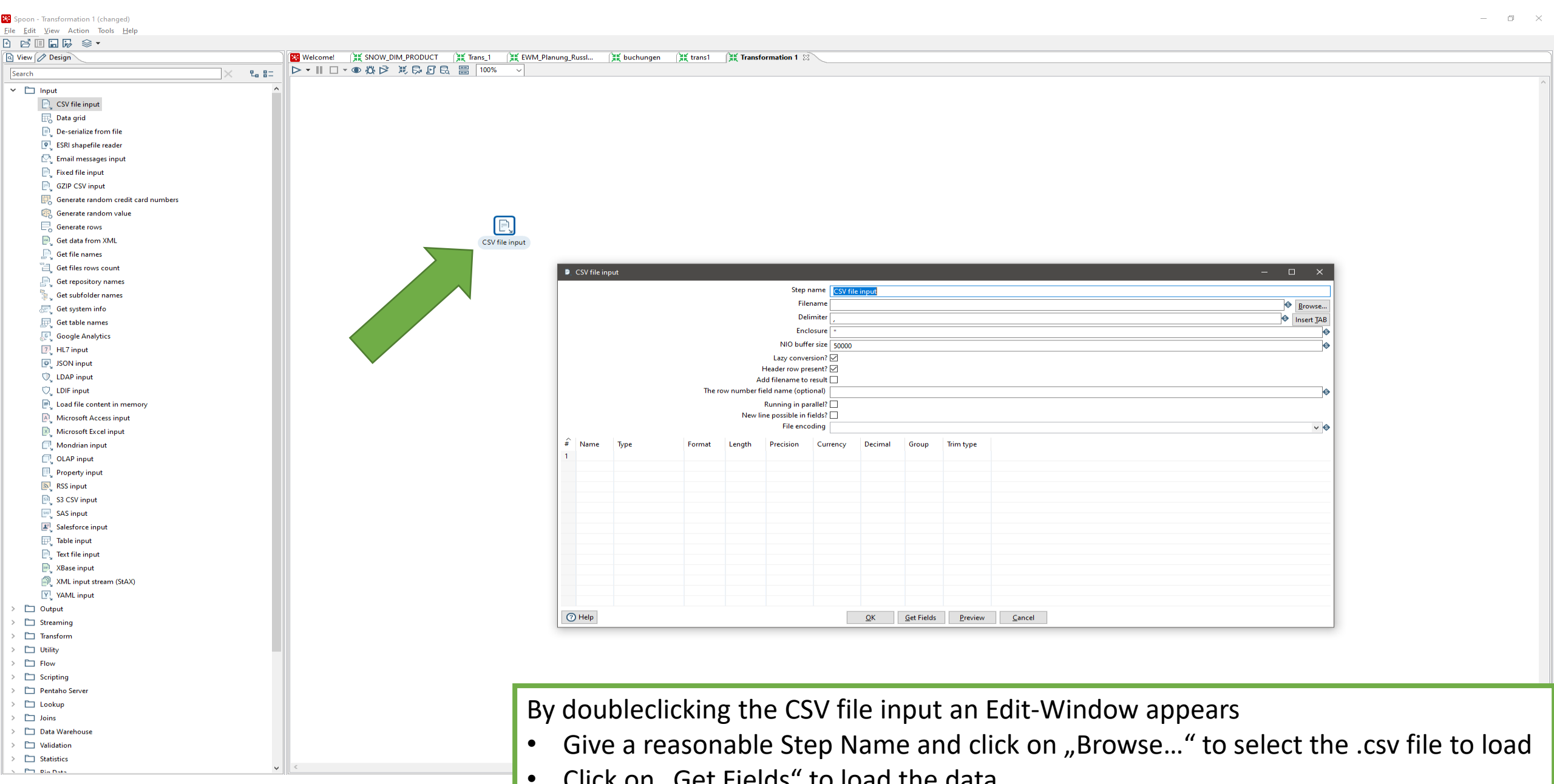
Tutorial ETL - Processing



In order to start a new transformation, click on File > New > Transformation

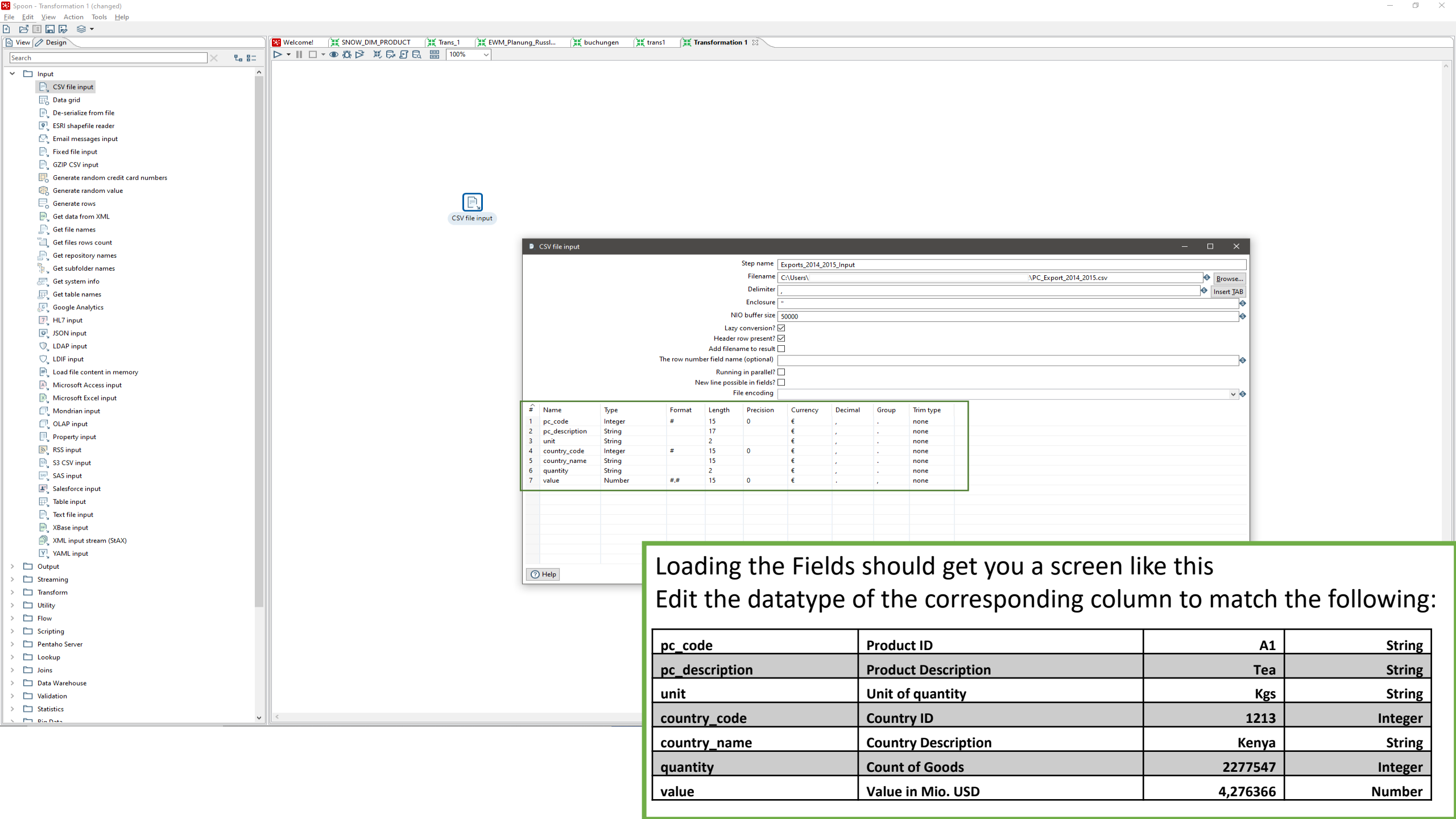


- To import the operational data, click on Input and drag the CSV file input into the main window



By doubleclicking the CSV file input an Edit-Window appears

- Give a reasonable Step Name and click on „Browse…“ to select the .csv file to load
- Click on „Get Fields“ to load the data
- Press OK



Loading the Fields should get you a screen like this

Edit the datatype of the corresponding column to match the following:

pc_code	Product ID	A1	String
pc_description	Product Description	Tea	String
unit	Unit of quantity	Kgs	String
country_code	Country ID	1213	Integer
country_name	Country Description	Kenya	String
quantity	Count of Goods	2277547	Integer
value	Value in Mio. USD	4,276366	Number

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

View Design

Search

- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
  - Execute SQL script
  - Execute row SQL script
  - Formula
  - Modified JavaScript value
  - Regex evaluati
  - Rules accumul
  - Rules executor
  - User defined Java class
  - User defined Java expression
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

Modified JavaScript value

This is a modified plugin for the Scripting Values with improved interface and performance.
Written _donated to open source by Martin Lange, Proconis : http://www.proconis.de

Welcome! | SNOW_DIM_PRODUCT | Trans_1 | EWM_Planung_Russl... | buchungen | trans1 | Transformation 1

100%

Exports_2014_2015_Input

The first step is to add a random date to each transaction
To achieve this, we integrate a step named „Modified JavaScript value“
This step can be found under Scripting > Modified JavaScrip value
Drag it into the main screen

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

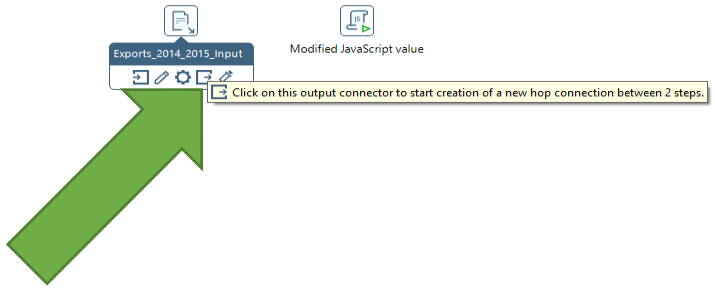
View Design

Search

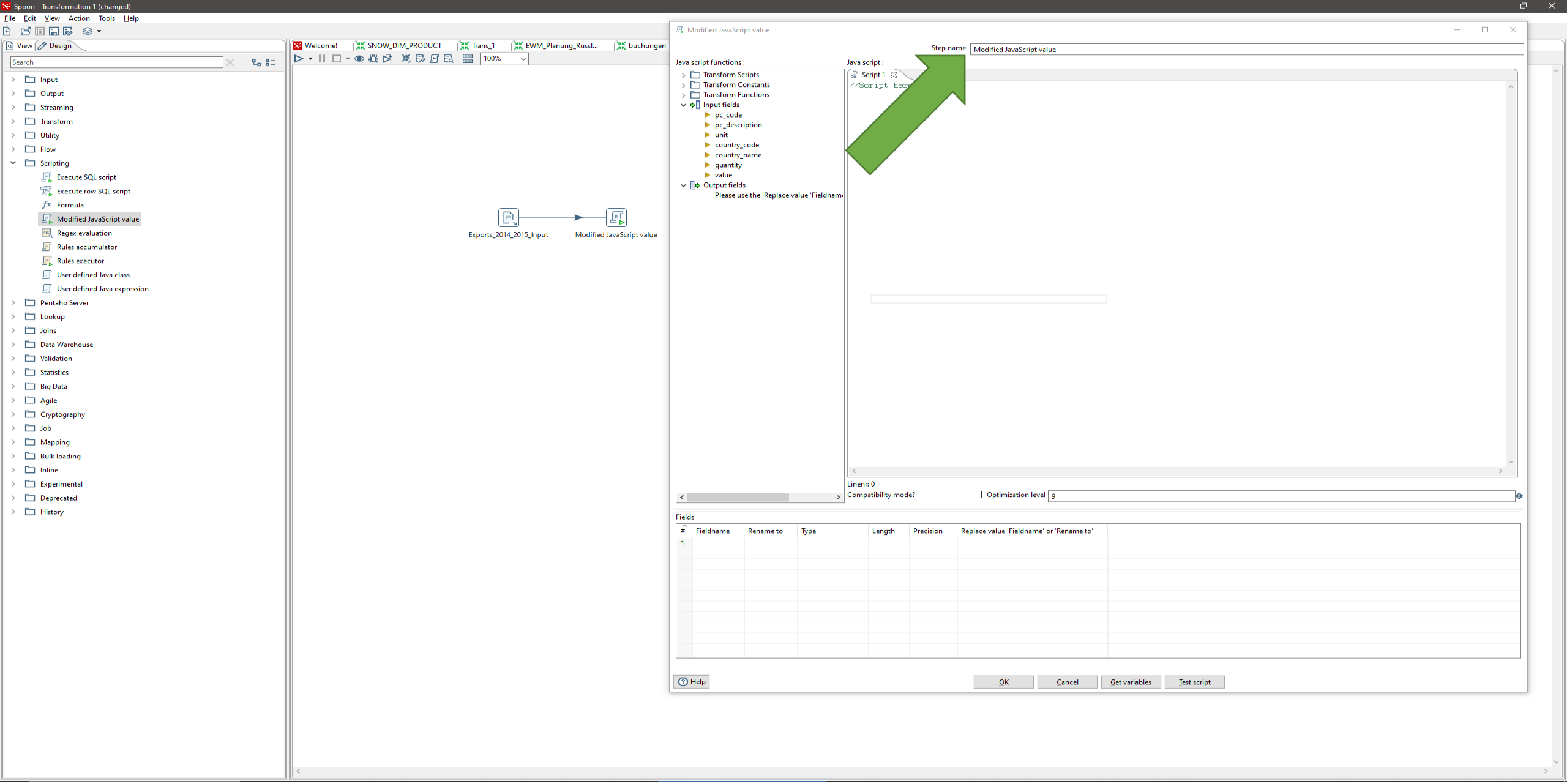
- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
  - Execute SQL script
  - Execute row SQL script
  - Formula
  - Modified JavaScript value
  - Regex evaluation
  - Rules accumulator
  - Rules executor
  - User defined Java class
  - User defined Java expression
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

Welcome! | SNOW_DIM_PRODUCT | Trans_1 | EWM_Planung_Russl... | buchungen | trans1 | Transformation 1

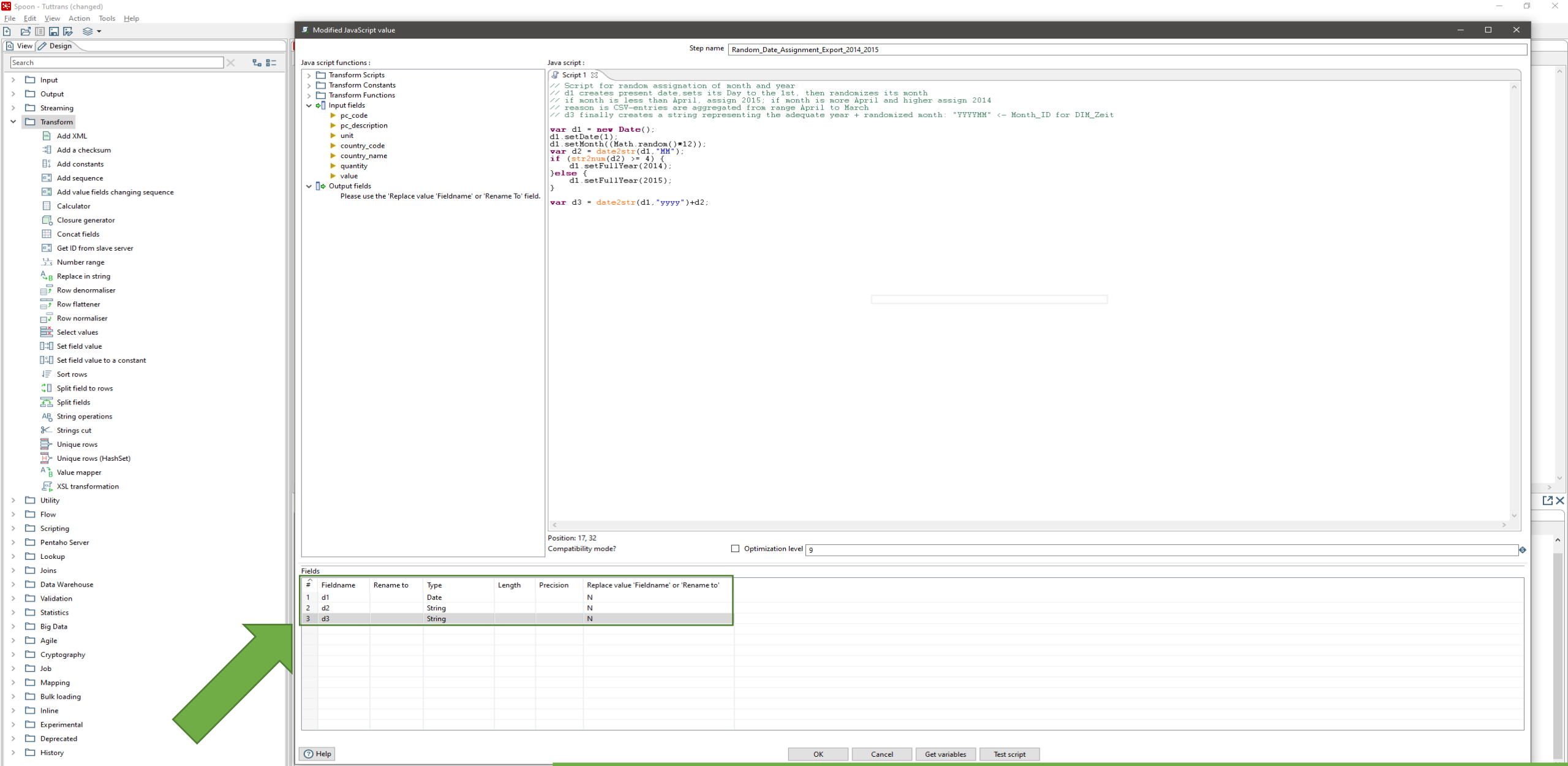
100%



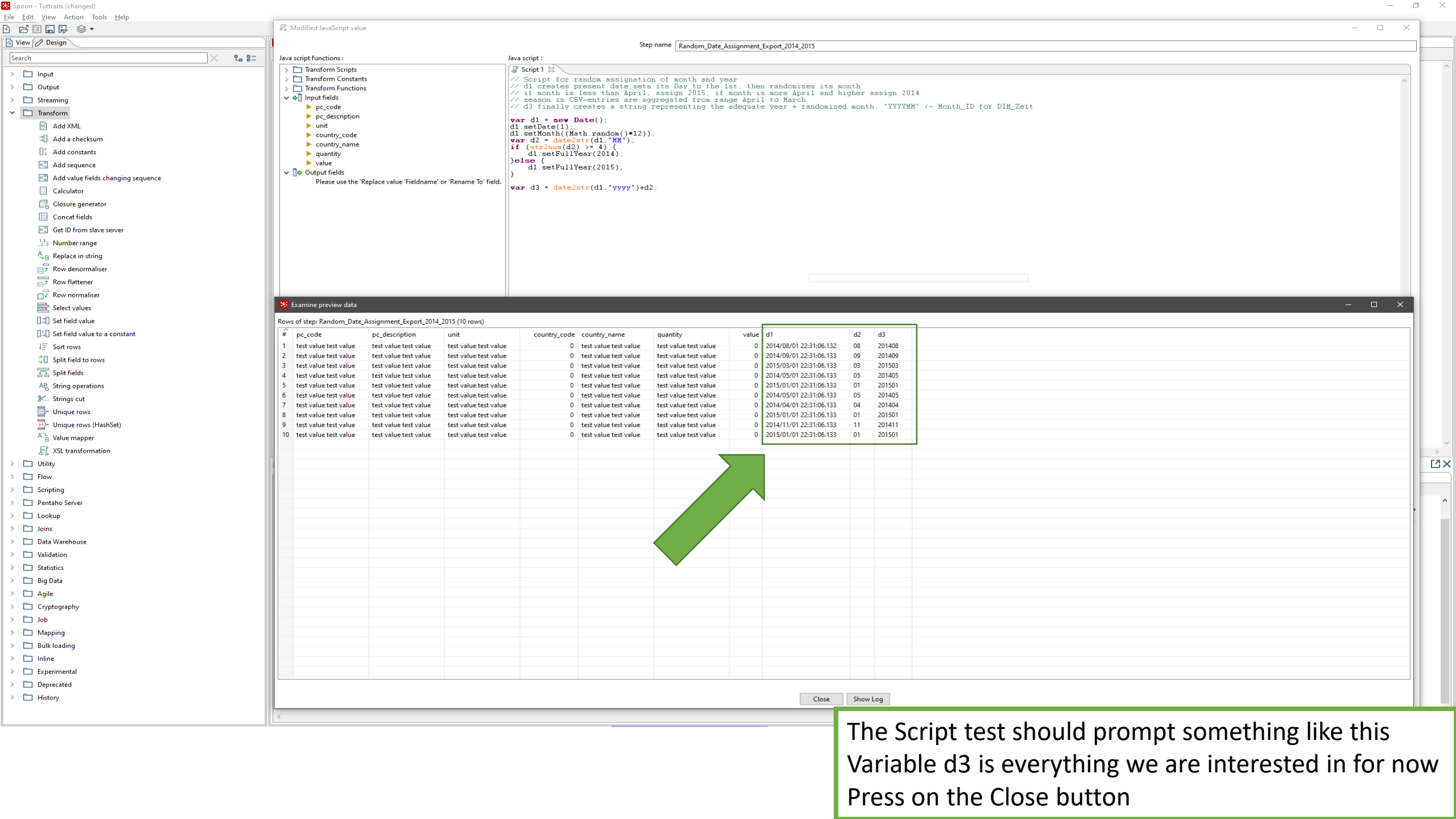
To connect both steps, simply hover over the already created input step
Then click on the output connector and click on the Modified JavaScript value – Step
Select „Main Output of Step“
This should create a link between both steps
Finally click on Modified JavaScript value

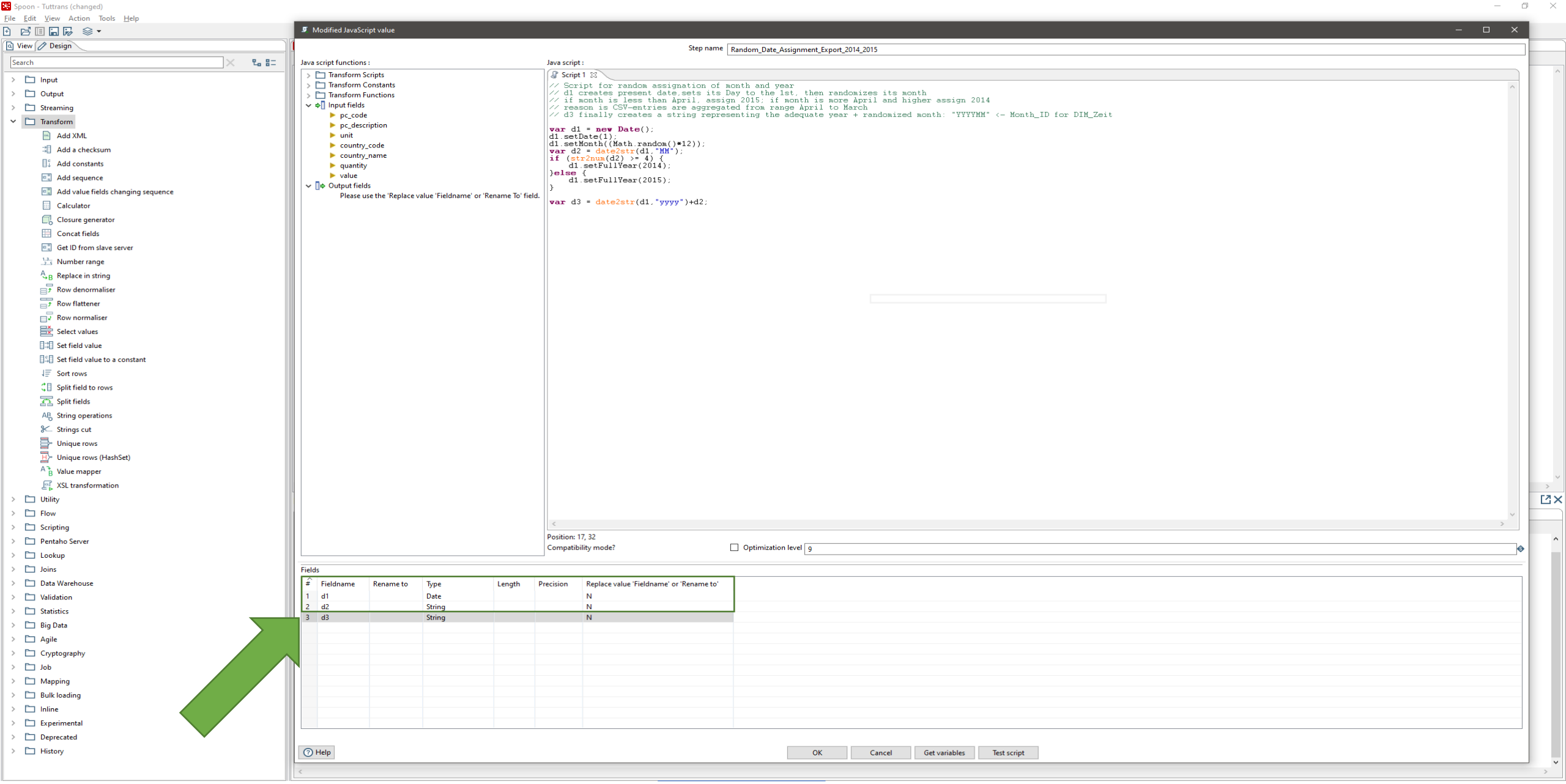


Again give the step a reasonable name
We will now proceed with scripting the random date assignment

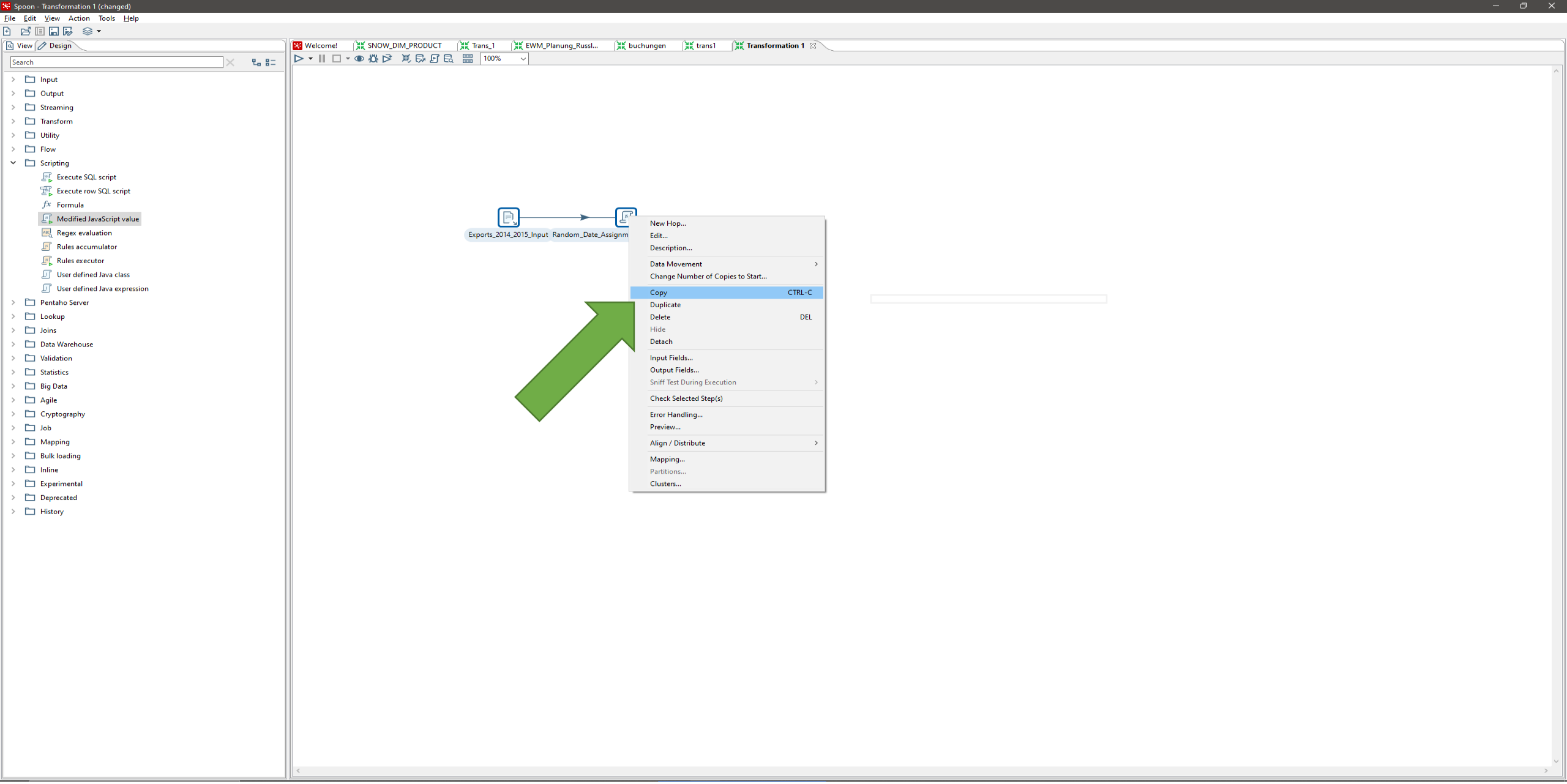


The procedure of the script is described in the comment section above the code
After typing the script a click on „Get variables“ should give you an output like this
Change the datatype of d3 to Integer and click on „Test script“ and press OK

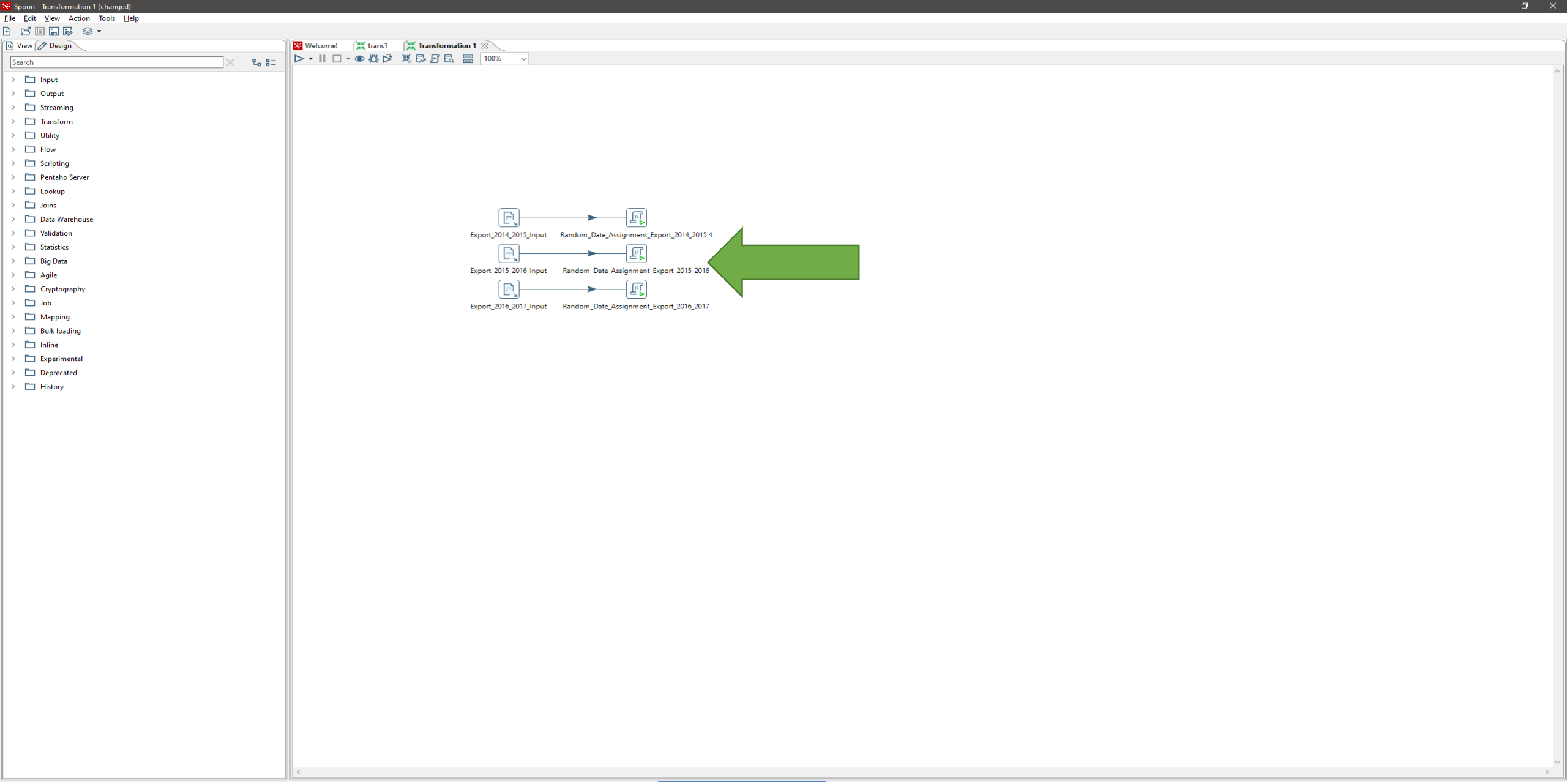

Spoon - Tuttrans (changed)
File Edit View Action Tools Help
View Design
Search
Input
Output
Streaming
Transform
Add XML
Add a checksum
Add constants
Add sequence
Add value fields changing sequence
Calculator
Closure generator
Concat fields
Get ID from slave server
Number range
Replace in string
Row denormaliser
Row flattener
Row normaliser
Select values
Set field value
Set field value to a constant
Sort rows
Split field to rows
Split fields
String operations
Strings cut
Unique rows
Unique rows (HashSet)
Value mapper
XSL transformation
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Big Data
Agile
Cryptography
Job
Mapping
Bulk loading
Inline
Experimental
Deprecated
History
Modified JavaScript value
Step name
Random_Date_Assignment_Export_2014_2015
Java script functions :
Transform Scripts
Transform Constants
Transform Functions
Input fields
pc_code
pc_description
unit
country_code
country_name
quantity
value
Output fields
Please use the 'Replace value 'Fieldname' or 'Rename To' field.
Java script :
Script 1
// Script for random assignation of month and year
// d1 creates present date,sets its Day to the 1st, then randomizes its month
// if month is less than April, assign 2015; if month is more April and higher assign 2014
// reason is CSV-entries are aggregated from range April to March
// d3 finally creates a string representing the adequate year + randomized month: "YYYYMM" <- Month_ID for DIM_Zeit
var d1 = new Date();
d1.setDate(1);
d1.setMonth((Math.random()*12));
var d2 = date2str(d1,"MM");
if (str2num(d2) >= 4) {
d1.setFullYear(2014);
}else {
d1.setFullYear(2015);
}
var d3 = date2str(d1,"yyyy")+d2;
Examine preview data
Rows of step: Random_Date_Assignment_Export_2014_2015 (10 rows)
pc_code pc_description unit country_code country_name quantity value d1 d2 d3
1 test value test value test value test value test value test value test value test value 0 2014/08/01 22:31:06.132 08 201408
2 test value test value test value test value test value test value test value test value 0 2014/09/01 22:31:06.133 09 201409
3 test value test value test value test value test value test value test value test value 0 2015/03/01 22:31:06.133 03 201503
4 test value test value test value test value test value test value test value test value 0 2014/05/01 22:31:06.133 05 201405
5 test value test value test value test value test value test value test value test value 0 2015/01/01 22:31:06.133 01 201501
6 test value test value test value test value test value test value test value test value 0 2014/05/01 22:31:06.133 05 201405
7 test value test value test value test value test value test value test value test value 0 2014/04/01 22:31:06.133 04 201404
8 test value test value test value test value test value test value test value test value 0 2015/01/01 22:31:06.133 01 201501
9 test value test value test value test value test value test value test value test value 0 2014/11/01 22:31:06.133 11 201411
10 test value test value test value test value test value test value test value test value 0 2015/01/01 22:31:06.133 01 201501
Close
Show Log
The Script test should prompt something like this
Variable d3 is everything we are interested in for now
Press on the Close button



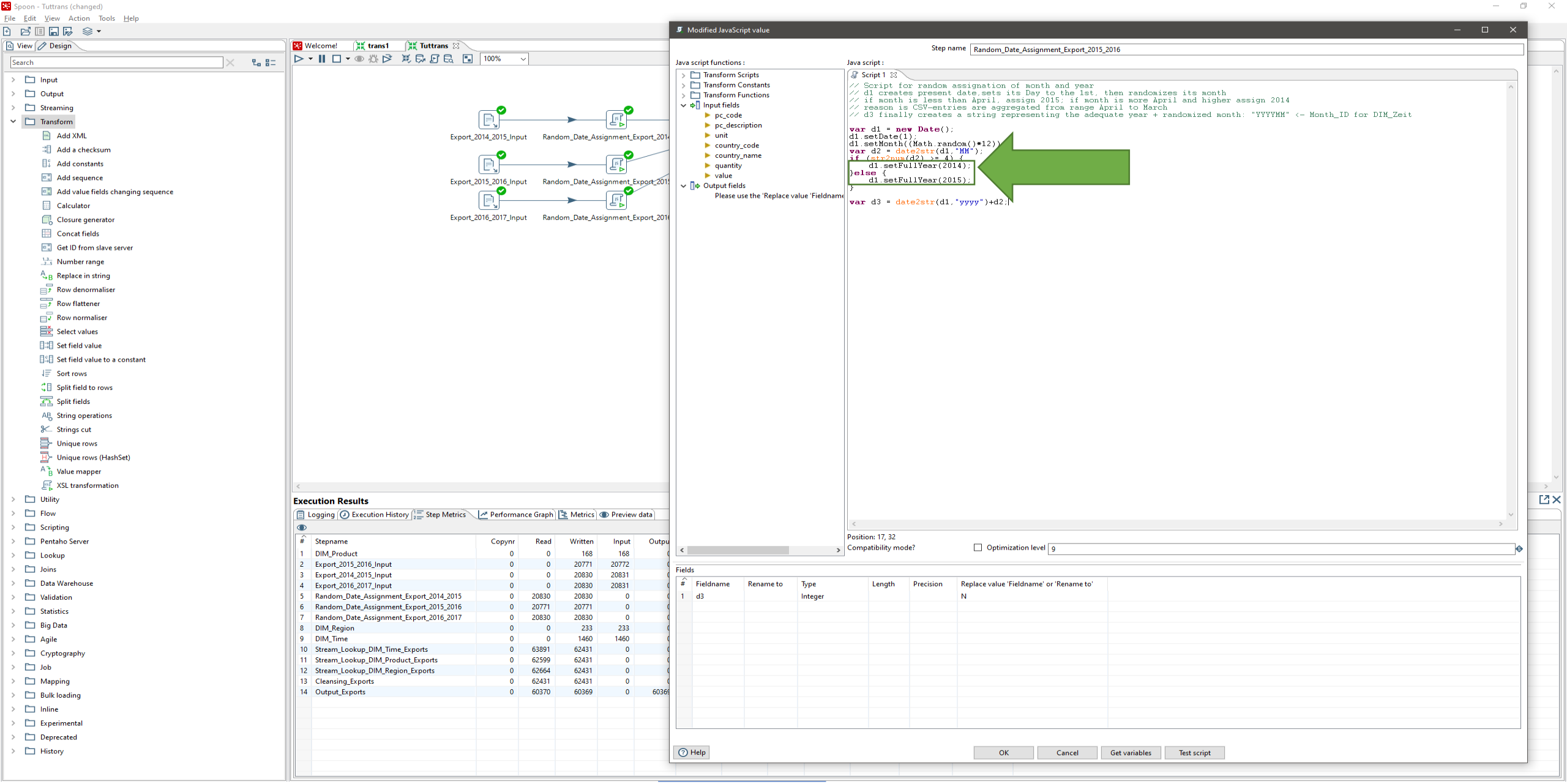
Delete both variables d1 and d2 since we are not interested in it by now
Press OK



Now copy both steps and paste them to create inputs for the other two export files



Rearrange the steps and rename them according to the picture above
Then double-click on „Random_Date_Assignment_Export_2015_2016“



Adjust d1.setFullYear as described above and press OK

Do the same for „Random_Date_Assignment_Export_2016_2017“

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

View Design

Search

- Input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
  - Call DB procedure
  - Check if file is locked
  - Check if webservice is available
  - Column exists
  - Database join
  - Database lookup
  - Dynamic SQL row
  - File exists
  - Fuzzy match
  - HTTP client
  - HTTP post
  - REST client
  - Stream lookup
  - Table exists
  - Web services l
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

n lookup

alues coming from another stream in the transformation.

Welcome! trans1 Transformation 1

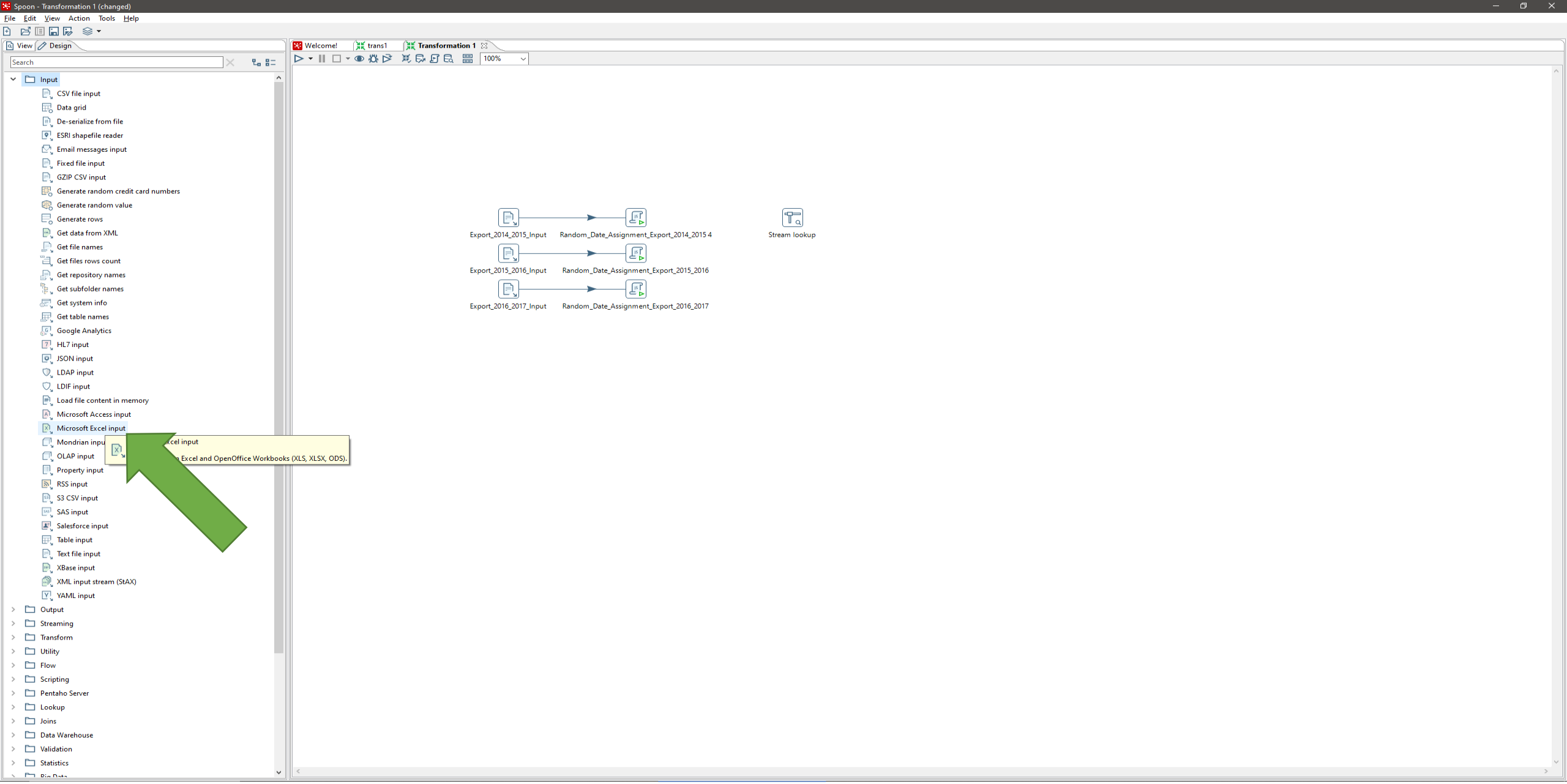
100%

Export_2014_2015_Input → Random_Date_Assignment_Export_2014_2015 4

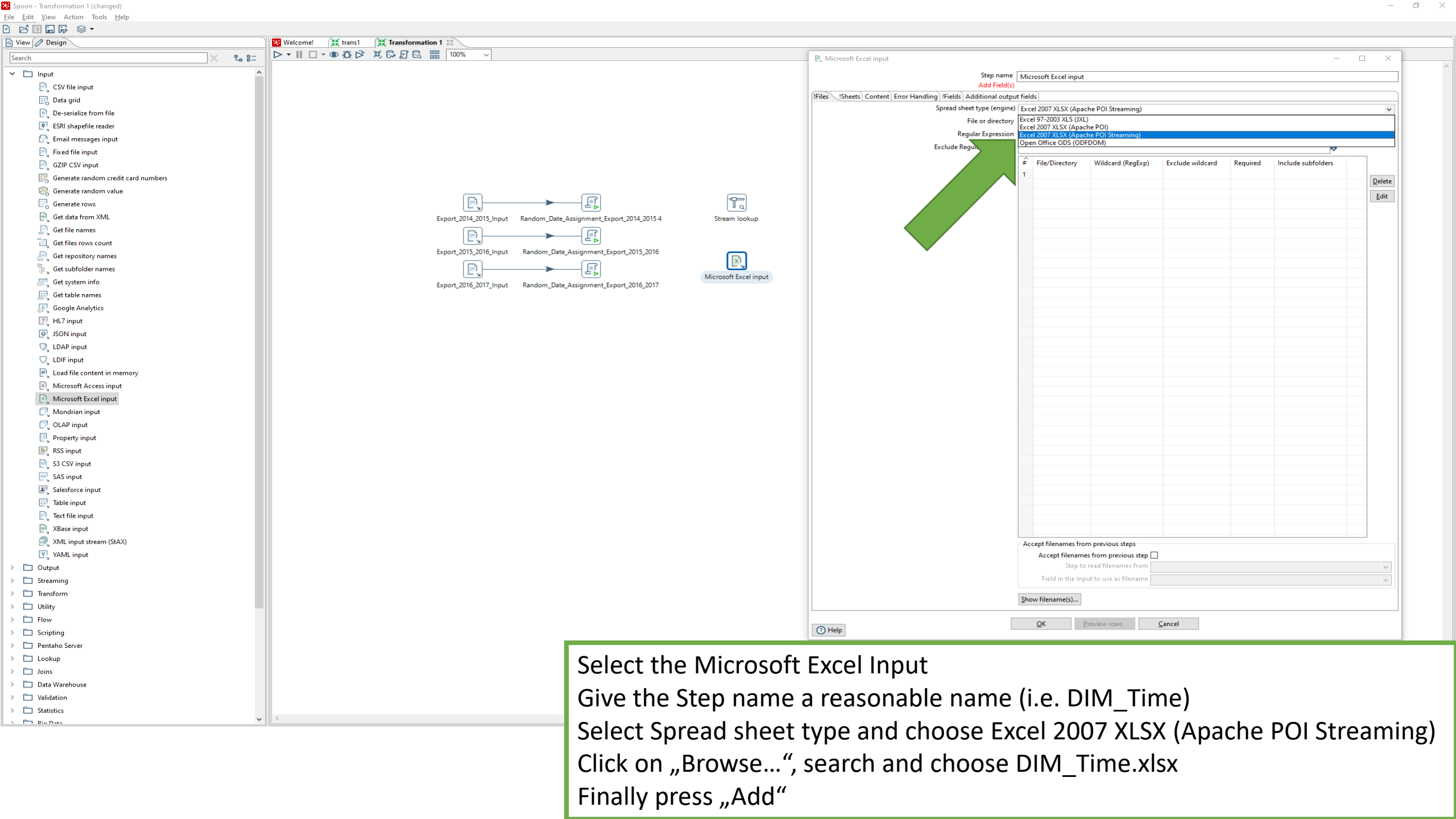
Export_2015_2016_Input → Random_Date_Assignment_Export_2015_2016

Export_2016_2017_Input → Random_Date_Assignment_Export_2016_2017

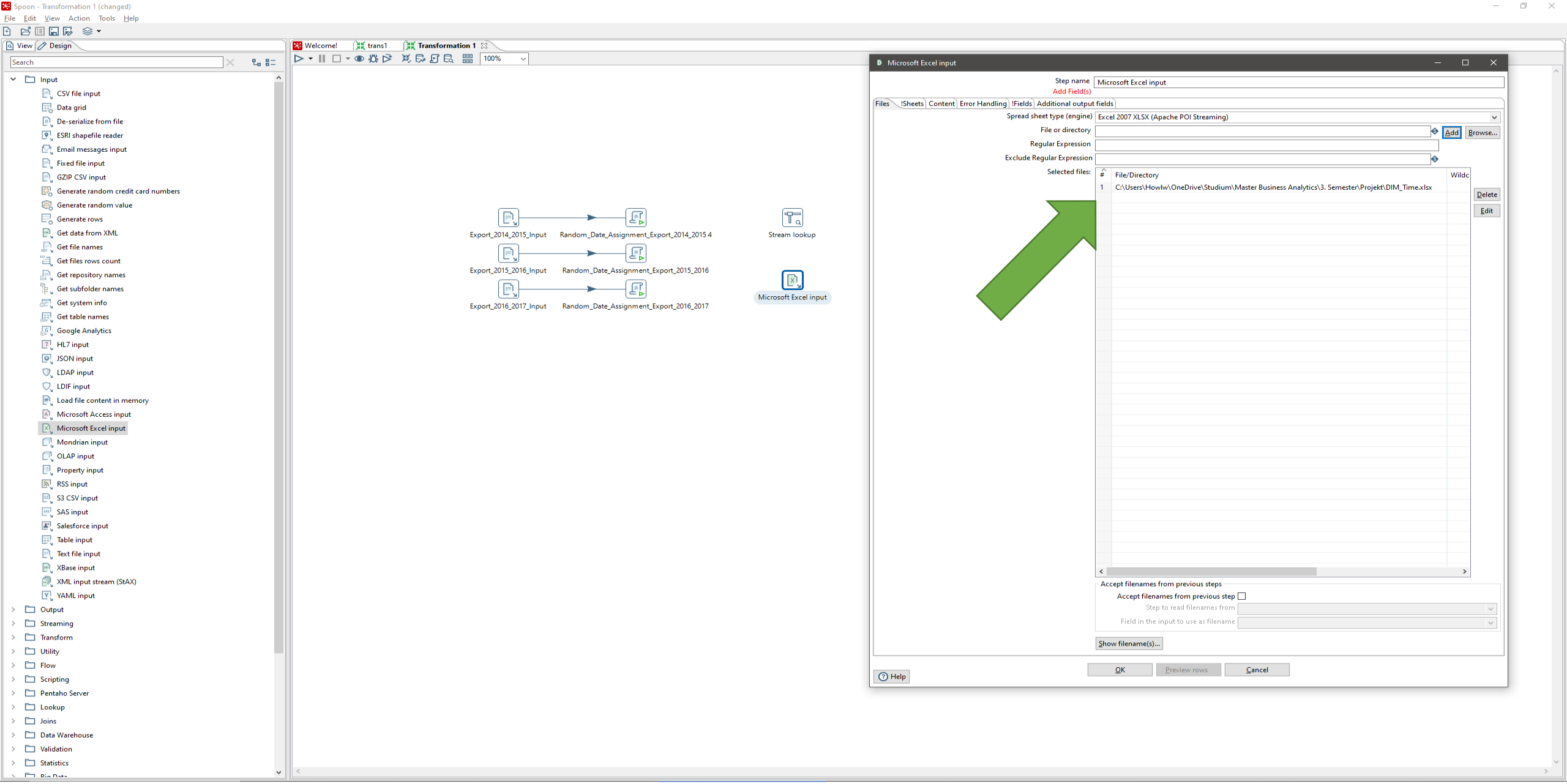
We will now create the first Lookup for our inputs to enrich dimensions
Find Stream Lookup under Lookup and drag it onto the main screen



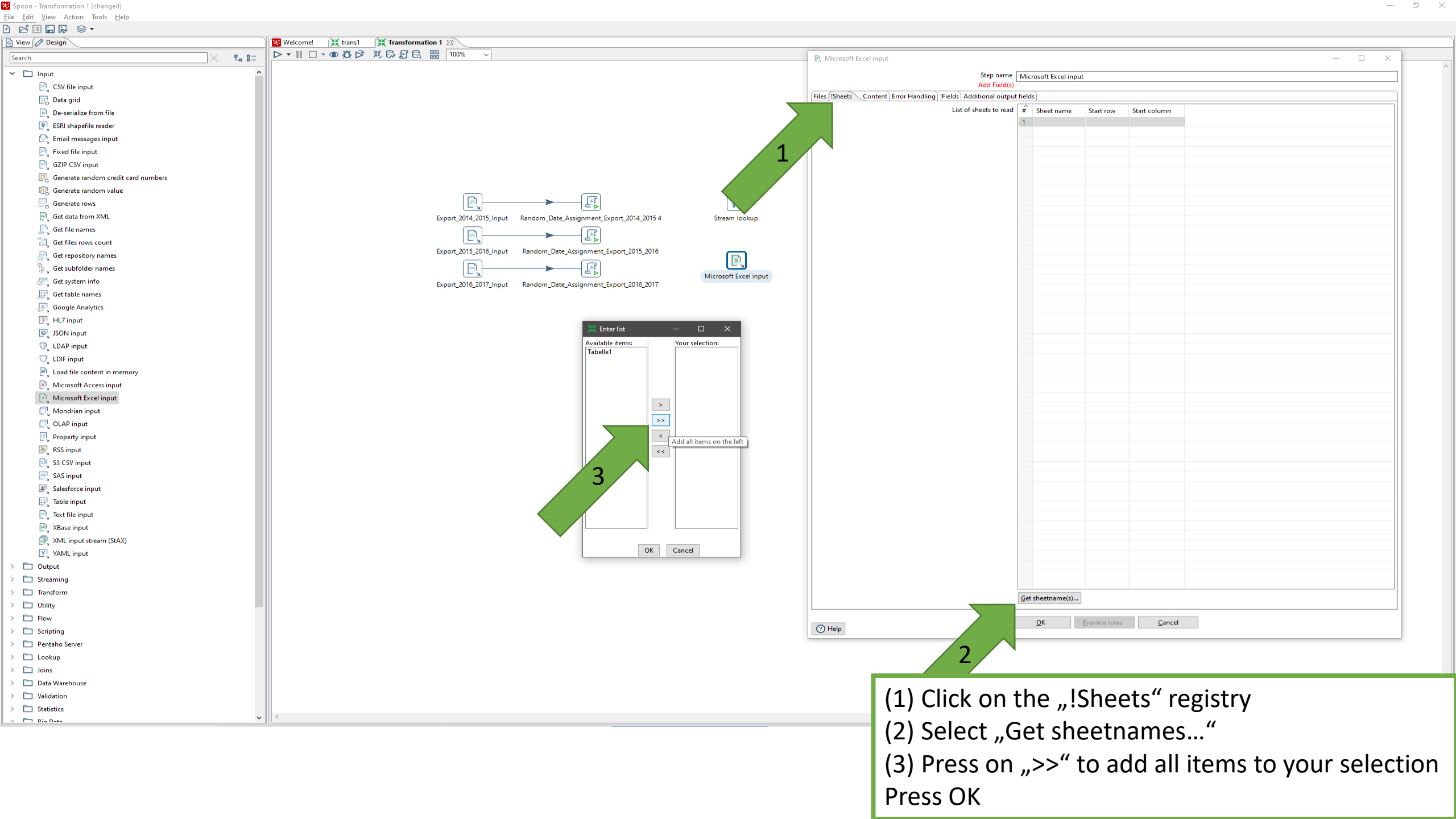
Since our Time dimension DIM_Time is an Excel file, we additionally need a MS Excel Input Step
Drag the MS Input Step onto the main screen

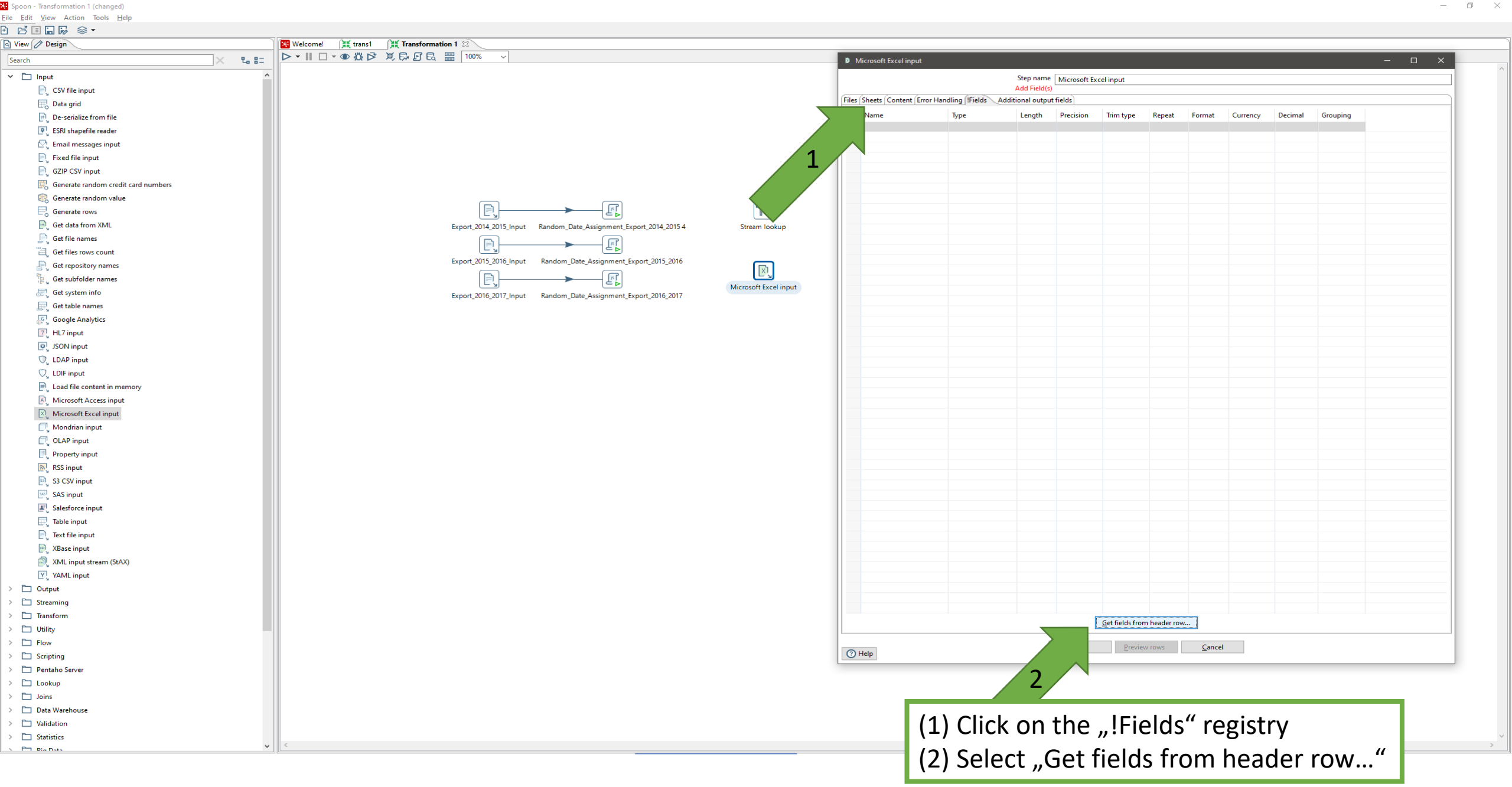


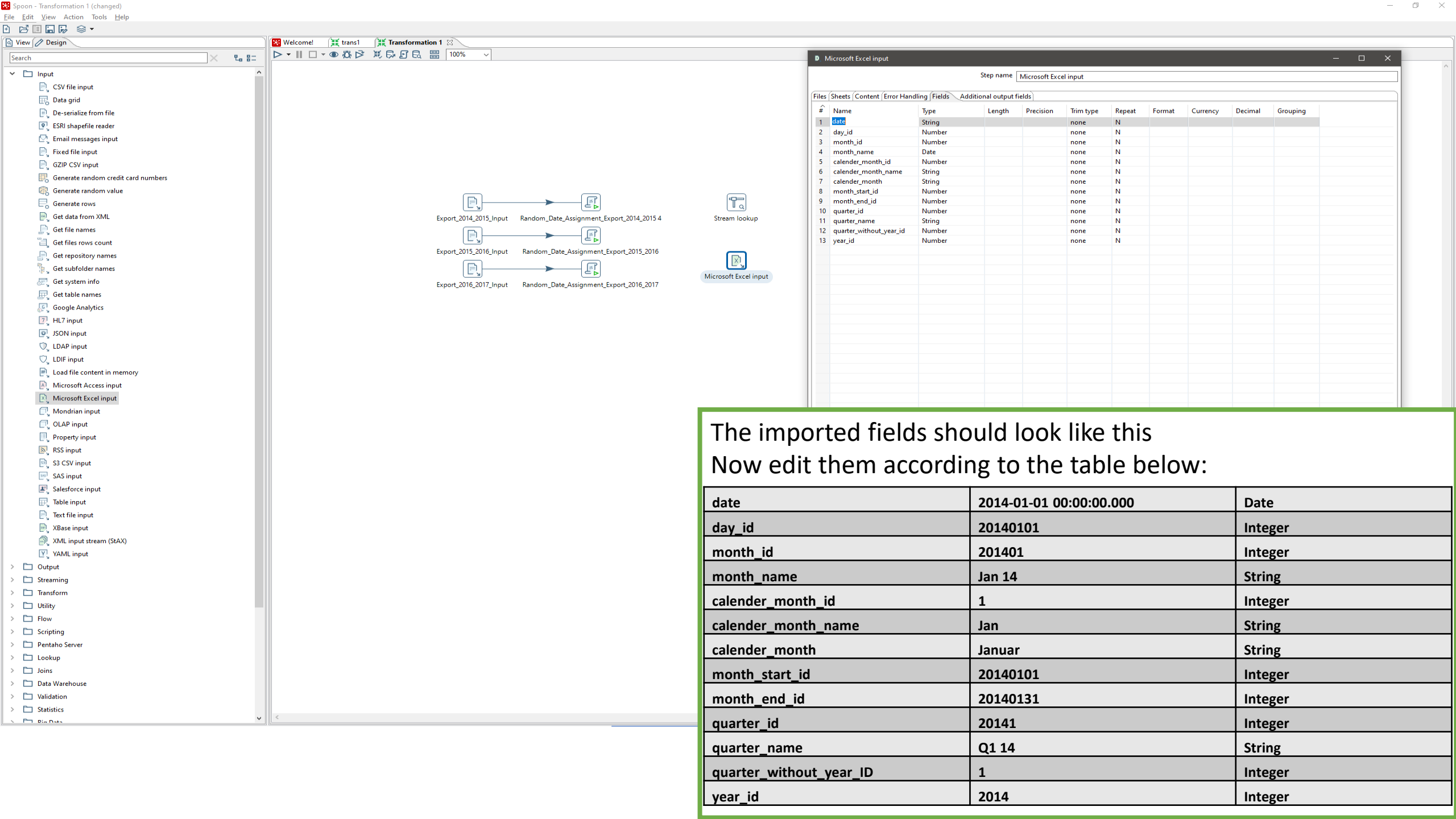
Select the Microsoft Excel Input
Give the Step name a reasonable name (i.e. DIM_Time)
Select Spread sheet type and choose Excel 2007 XLSX (Apache POI Streaming)
Click on „Browse…“, search and choose DIM_Time.xlsx
Finally press „Add“



DIM_Time.xlsx should be added to the selected files
Screen should look like stated above

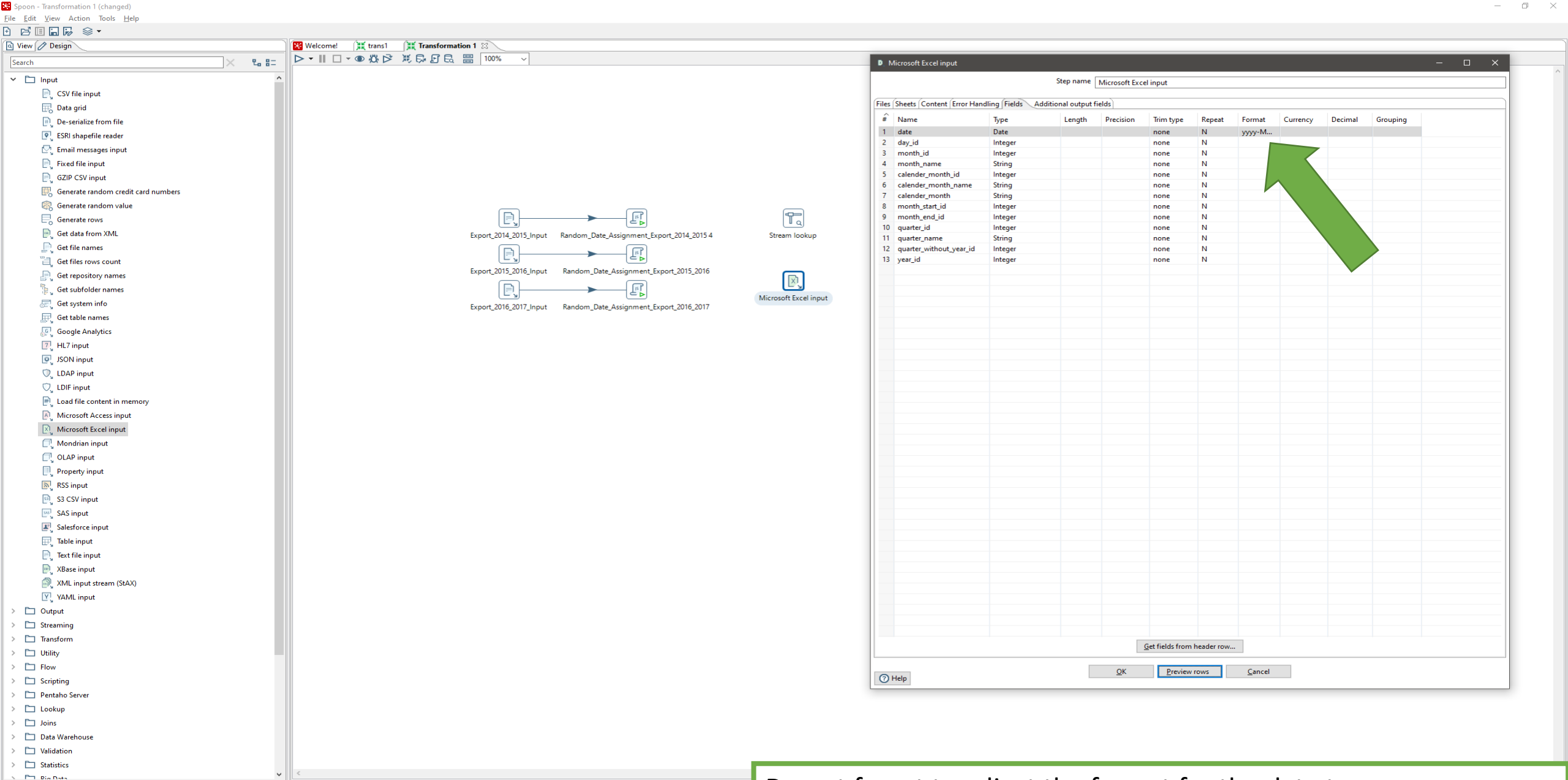

Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
View Design
Search
Input
CSV file input
Data grid
De-serialize from file
ESRI shapefile reader
Email messages input
Fixed file input
GZIP CSV input
Generate random credit card numbers
Generate random value
Generate rows
Get data from XML
Get file names
Get files rows count
Get repository names
Get subfolder names
Get system info
Get table names
Google Analytics
HL7 input
JSON input
LDAP input
LDIF input
Load file content in memory
Microsoft Access input
Microsoft Excel input
Mondrian input
OLAP input
Property input
RSS input
S3 CSV input
SAS input
Salesforce input
Table input
Text file input
XBase input
XML input stream (StAX)
YAML input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Welcome!
trans1
Transformation 1
100%
Export_2014_2015_Input
Random_Date_Assignment_Export_2014_2015 4
Export_2015_2016_Input
Random_Date_Assignment_Export_2015_2016
Export_2016_2017_Input
Random_Date_Assignment_Export_2016_2017
Stream lookup
Microsoft Excel input
Enter list
Available items:
Tabelle1
Your selection:
>
>>
<
<<
Add all items on the left
OK
Cancel
1
2
3
Microsoft Excel input
Step name
Microsoft Excel input
Add Field(s)
Files !Sheets Content Error Handling !Fields Additional output fields
List of sheets to read
Sheet name Start row Start column
1
Get sheetname(s)...
Help
OK
Preview rows
Cancel
(1) Click on the „!Sheets“ registry
(2) Select „Get sheetnames…“
(3) Press on „>>“ to add all items to your selection
Press OK


Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
View Design
Search
Input
CSV file input
Data grid
De-serialize from file
ESRI shapefile reader
Email messages input
Fixed file input
GZIP CSV input
Generate random credit card numbers
Generate random value
Generate rows
Get data from XML
Get file names
Get files rows count
Get repository names
Get subfolder names
Get system info
Get table names
Google Analytics
HL7 input
JSON input
LDAP input
LDIF input
Load file content in memory
Microsoft Access input
Microsoft Excel input
Mondrian input
OLAP input
Property input
RSS input
S3 CSV input
SAS input
Salesforce input
Table input
Text file input
XBase input
XML input stream (StAX)
YAML input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Welcome!
trans1
Transformation 1
100%
Export_2014_2015_Input
Random_Date_Assignment_Export_2014_2015 4
Export_2015_2016_Input
Random_Date_Assignment_Export_2015_2016
Export_2016_2017_Input
Random_Date_Assignment_Export_2016_2017
Stream lookup
Microsoft Excel input
Microsoft Excel input
Step name
Microsoft Excel input
Add Field(s)
Files
Sheets
Content
Error Handling
!Fields
Additional output fields
Name
Type
Length
Precision
Trim type
Repeat
Format
Currency
Decimal
Grouping
Get fields from header row...
Help
Preview rows
Cancel
1
2
(1) Click on the „!Fields“ registry
(2) Select „Get fields from header row…“



The imported fields should look like this
Now edit them according to the table below:

date	**2014-01-01 00:00:00.000**	**Date**
day_id	**20140101**	**Integer**
month_id	**201401**	**Integer**
month_name	**Jan 14**	**String**
calender_month_id	**1**	**Integer**
calender_month_name	**Jan**	**String**
calender_month	**Januar**	**String**
month_start_id	**20140101**	**Integer**
month_end_id	**20140131**	**Integer**
quarter_id	**20141**	**Integer**
quarter_name	**Q1 14**	**String**
quarter_without_year_ID	**1**	**Integer**
year_id	**2014**	**Integer**



Do not forget to adjust the format for the date to:
„yyyy-MM-dd HH:mm:ss.SSS“
Press OK

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

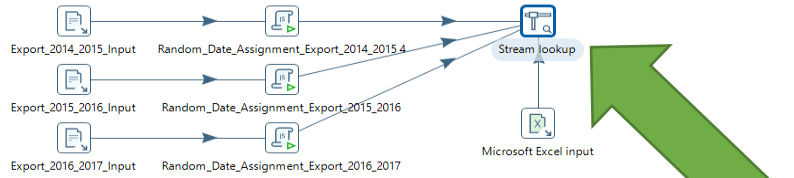
View Design

Search

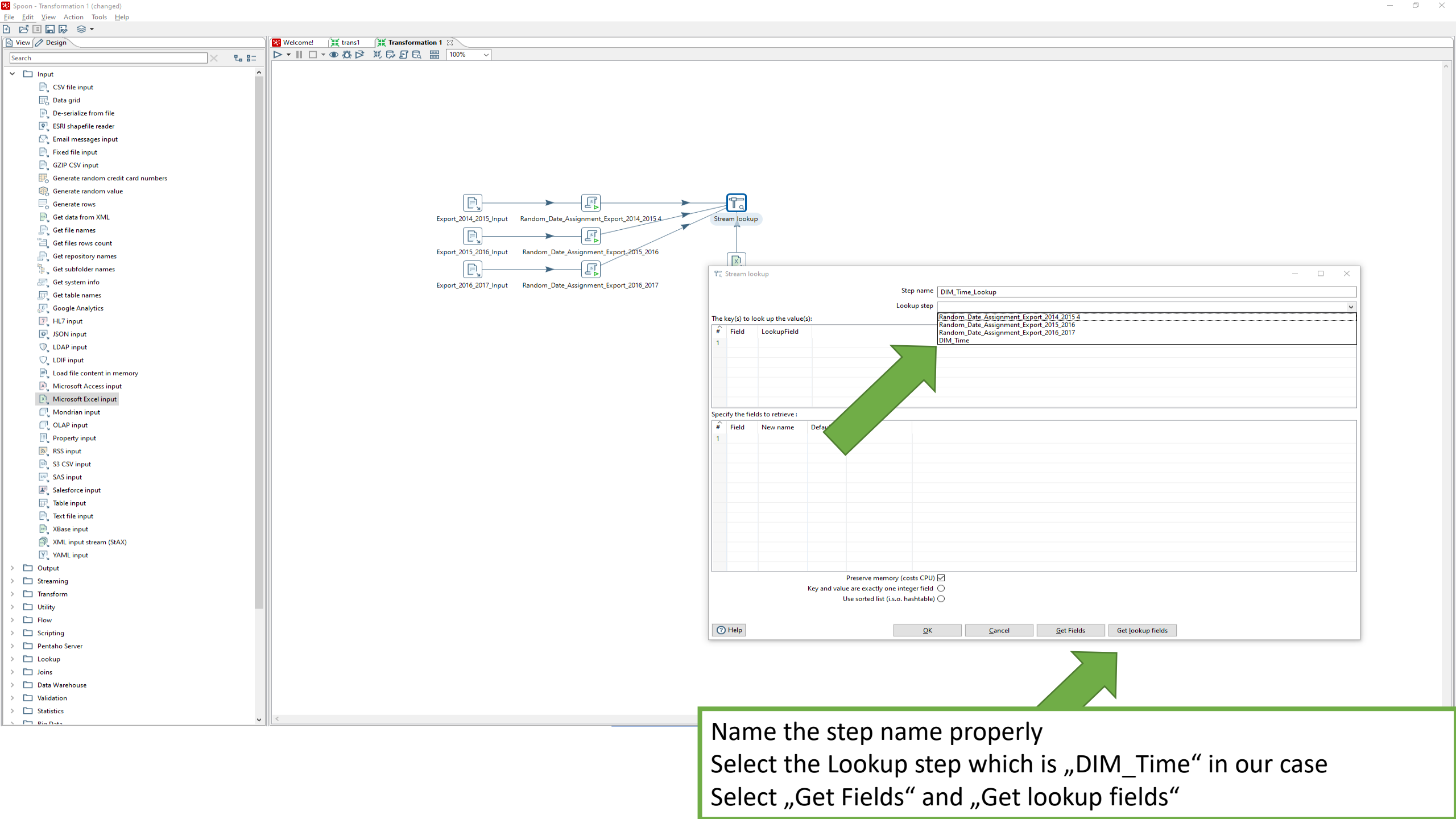
- Input
  - CSV file input
  - Data grid
  - De-serialize from file
  - ESRI shapefile reader
  - Email messages input
  - Fixed file input
  - GZIP CSV input
  - Generate random credit card numbers
  - Generate random value
  - Generate rows
  - Get data from XML
  - Get file names
  - Get files rows count
  - Get repository names
  - Get subfolder names
  - Get system info
  - Get table names
  - Google Analytics
  - HL7 input
  - JSON input
  - LDAP input
  - LDIF input
  - Load file content in memory
  - Microsoft Access input
  - Microsoft Excel input
  - Mondrian input
  - OLAP input
  - Property input
  - RSS input
  - S3 CSV input
  - SAS input
  - Salesforce input
  - Table input
  - Text file input
  - XBase input
  - XML input stream (StAX)
  - YAML input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics

Welcome! trans1 Transformation 1

100%

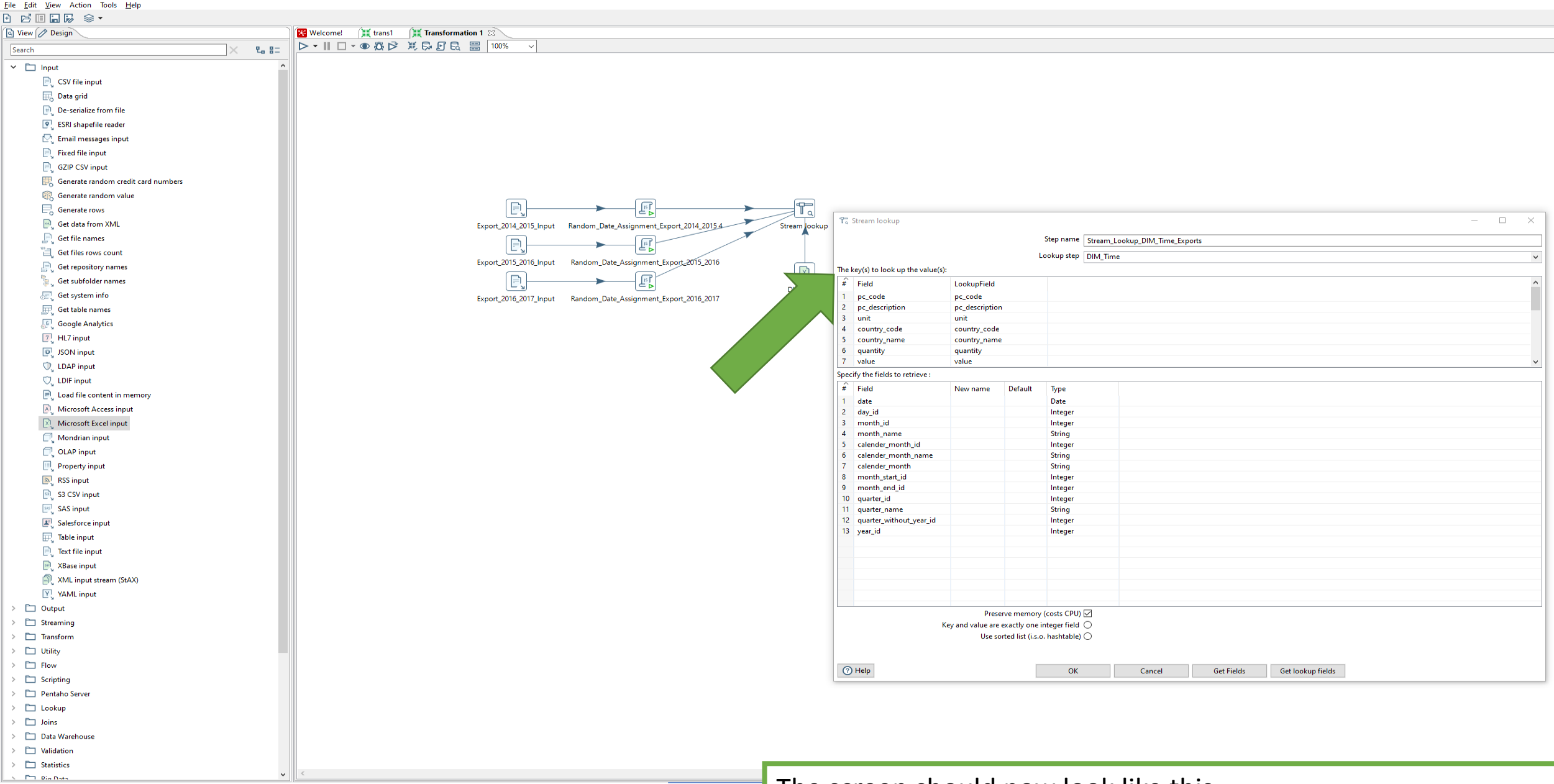


Now connect the steps as shown in the picture above
Then double-click „Stream lookup“ to edit it

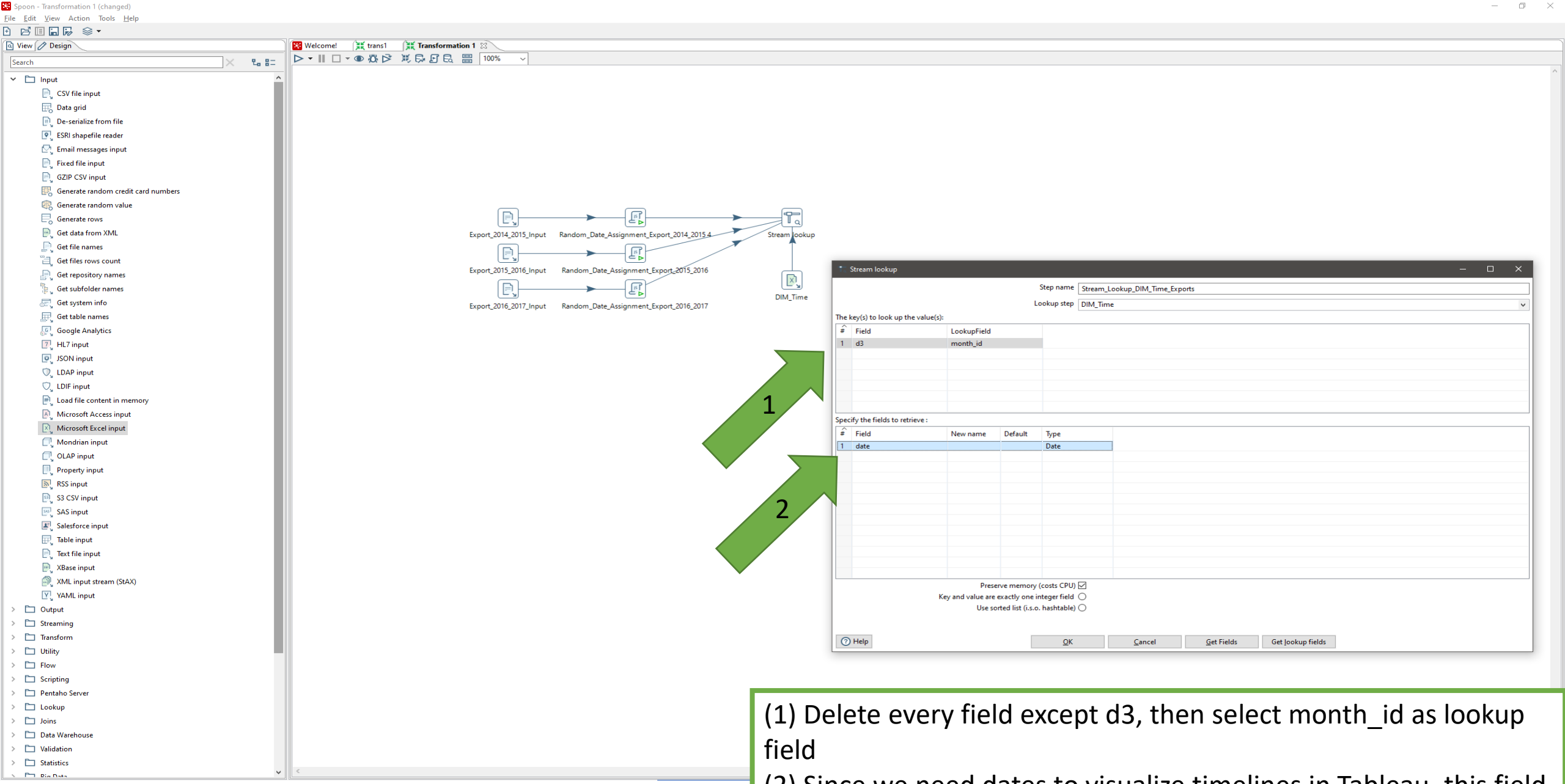

Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
View Design
Search
Input
CSV file input
Data grid
De-serialize from file
ESRI shapefile reader
Email messages input
Fixed file input
GZIP CSV input
Generate random credit card numbers
Generate random value
Generate rows
Get data from XML
Get file names
Get files rows count
Get repository names
Get subfolder names
Get system info
Get table names
Google Analytics
HL7 input
JSON input
LDAP input
LDIF input
Load file content in memory
Microsoft Access input
Microsoft Excel input
Mondrian input
OLAP input
Property input
RSS input
S3 CSV input
SAS input
Salesforce input
Table input
Text file input
XBase input
XML input stream (StAX)
YAML input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Welcome!
trans1
Transformation 1
100%
Export_2014_2015_Input
Random_Date_Assignment_Export_2014_2015 4
Export_2015_2016_Input
Random_Date_Assignment_Export_2015_2016
Export_2016_2017_Input
Random_Date_Assignment_Export_2016_2017
Stream lookup
Stream lookup
Step name
DIM_Time_Lookup
Lookup step
The key(s) to look up the value(s):
Field LookupField
1
Random_Date_Assignment_Export_2014_2015 4
Random_Date_Assignment_Export_2015_2016
Random_Date_Assignment_Export_2016_2017
DIM_Time
Specify the fields to retrieve :
Field New name
1
Preserve memory (costs CPU)
Key and value are exactly one integer field
Use sorted list (i.s.o. hashtable)
Help
OK
Cancel
Get Fields
Get lookup fields
Name the step name properly
Select the Lookup step which is „DIM_Time“ in our case
Select „Get Fields“ and „Get lookup fields“

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help



The screen should now look like this
Since we look up the columns of DIM_Time, we already know that d3 is the primary key to connect with DIM_Time



(1) Delete every field except d3, then select month_id as lookup field

(2) Since we need dates to visualize timelines in Tableau, this field is the only one we are interested in: Delete every field except date in the fields to retrieve

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

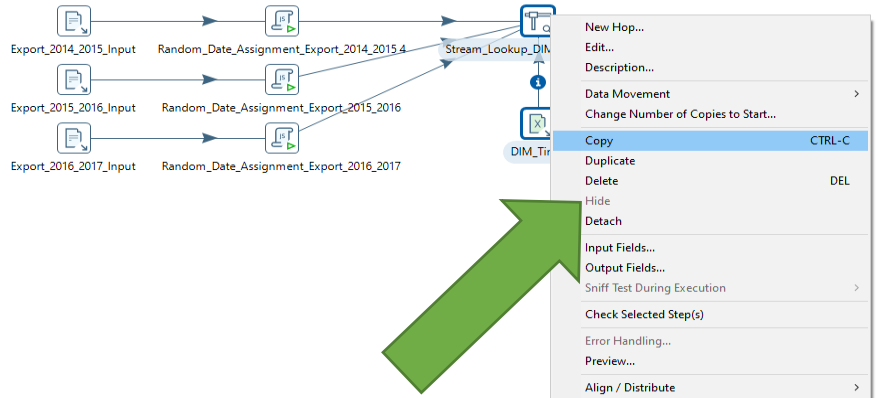
View Design

Search

- Input
  - CSV file input
  - Data grid
  - De-serialize from file
  - ESRI shapefile reader
  - Email messages input
  - Fixed file input
  - GZIP CSV input
  - Generate random credit card numbers
  - Generate random value
  - Generate rows
  - Get data from XML
  - Get file names
  - Get files rows count
  - Get repository names
  - Get subfolder names
  - Get system info
  - Get table names
  - Google Analytics
  - HL7 input
  - JSON input
  - LDAP input
  - LDIF input
  - Load file content in memory
  - Microsoft Access input
  - Microsoft Excel input
  - Mondrian input
  - OLAP input
  - Property input
  - RSS input
  - S3 CSV input
  - SAS input
  - Salesforce input
  - Table input
  - Text file input
  - XBase input
  - XML input stream (StAX)
  - YAML input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data

Welcome! trans1 Transformation 1

100%



Execution Results

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

```
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
```

Copy and Paste both the lookup step and the Excel Input Step right next to the first ones

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

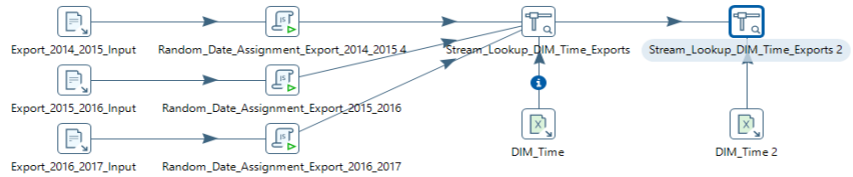
View Design

Search

- Input
  - CSV file input
  - Data grid
  - De-serialize from file
  - ESRI shapefile reader
  - Email messages input
  - Fixed file input
  - GZIP CSV input
  - Generate random credit card numbers
  - Generate random value
  - Generate rows
  - Get data from XML
  - Get file names
  - Get files rows count
  - Get repository names
  - Get subfolder names
  - Get system info
  - Get table names
  - Google Analytics
  - HL7 input
  - JSON input
  - LDAP input
  - LDIF input
  - Load file content in memory
  - Microsoft Access input
  - Microsoft Excel input
  - Mondrian input
  - OLAP input
  - Property input
  - RSS input
  - S3 CSV input
  - SAS input
  - Salesforce input
  - Table input
  - Text file input
  - XBase input
  - XML input stream (StAX)
  - YAML input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data

Welcome! trans1 Transformation 1

100%

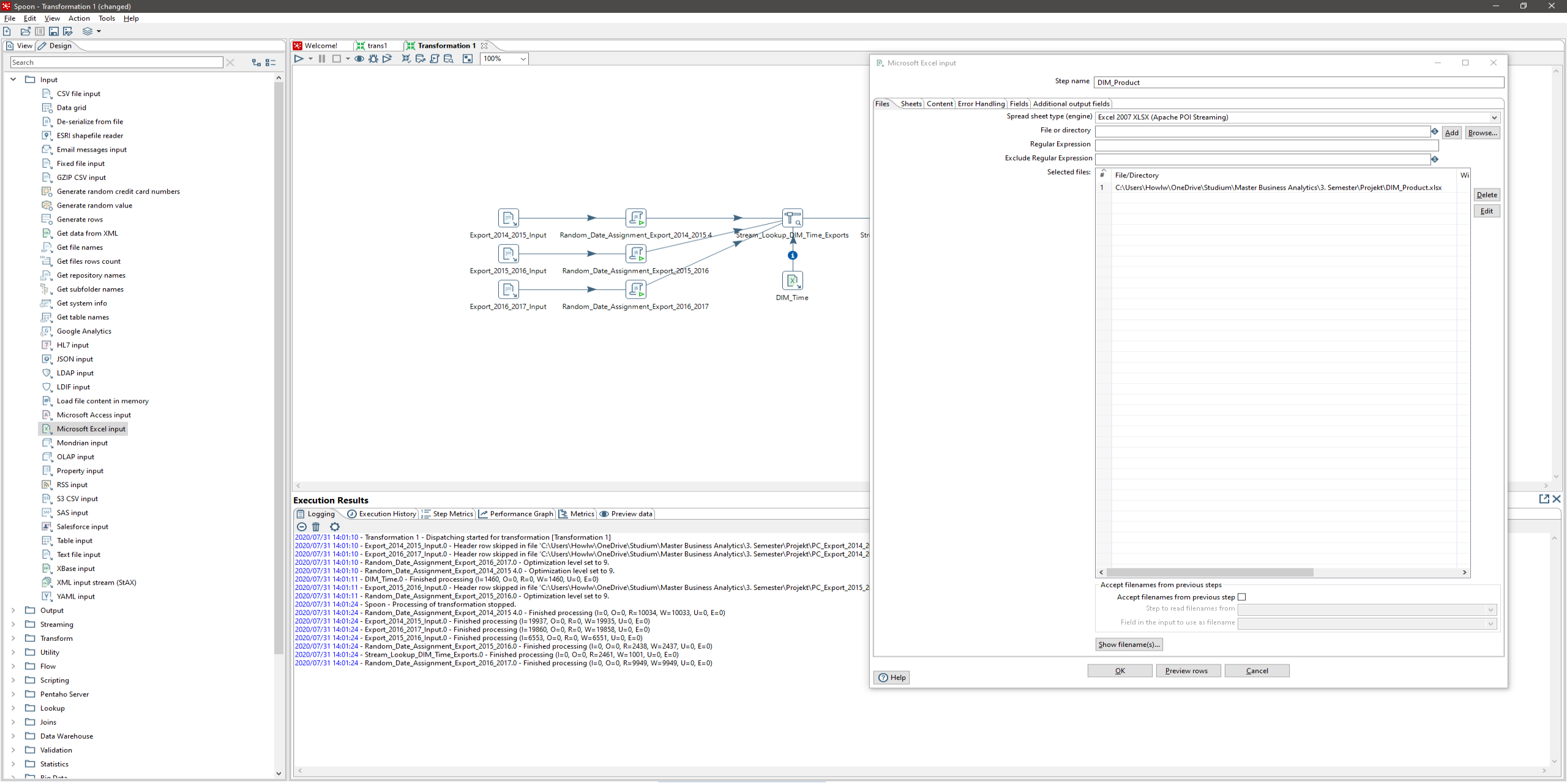


Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

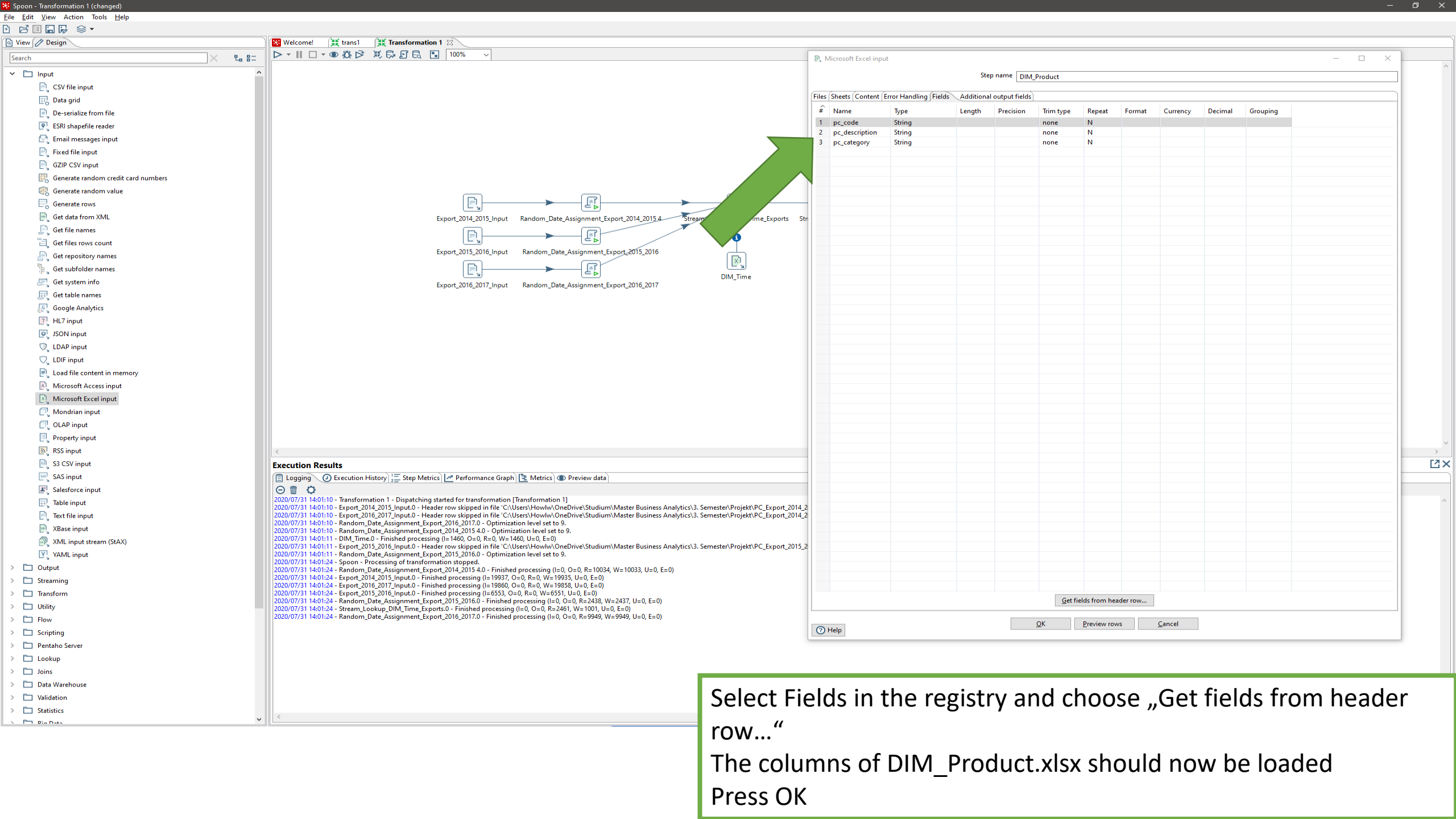
```
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
```

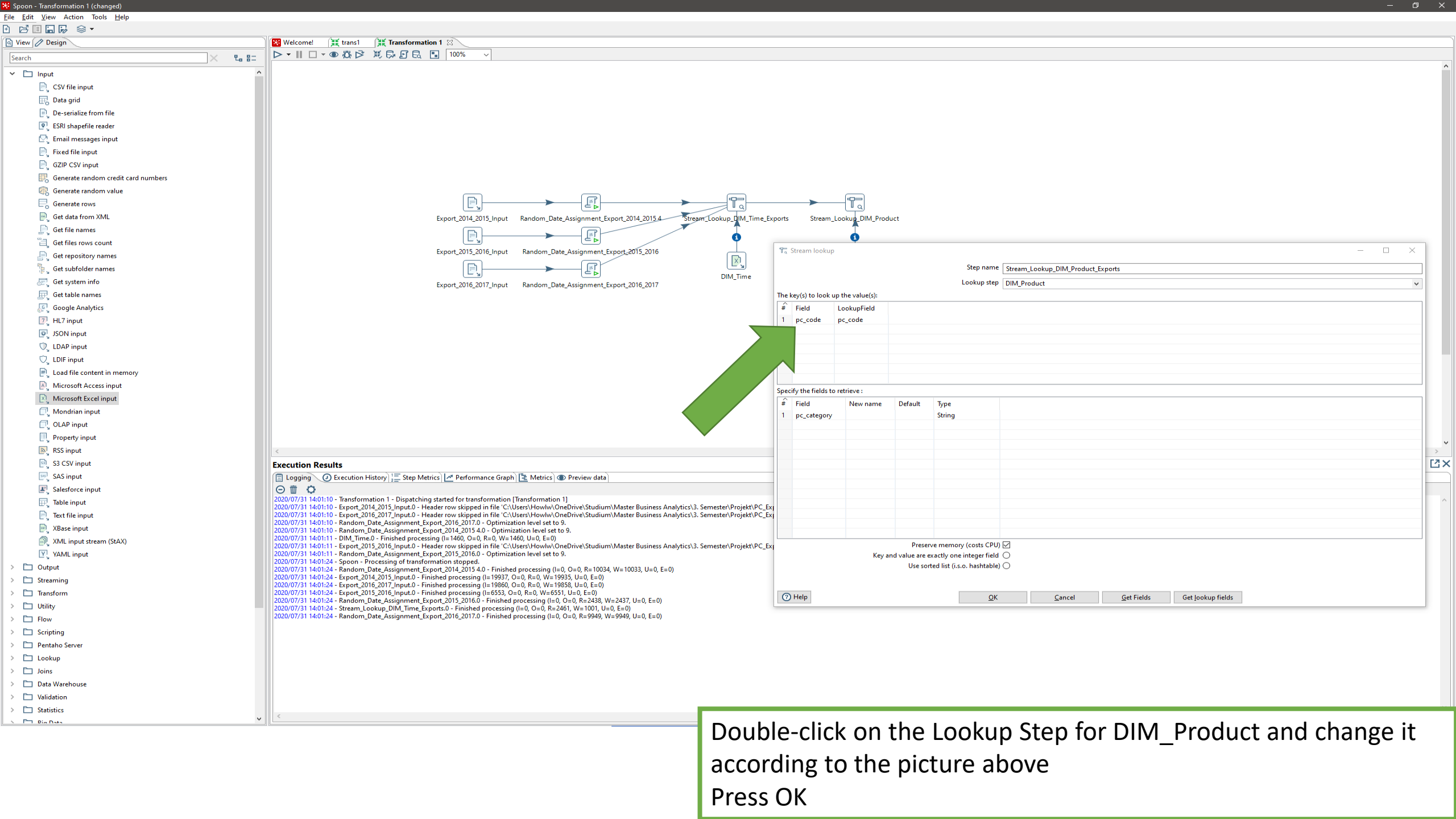
Connect the lookup steps and double-click the newly created Excel Input Step to set up DIM_Product



Rename it properly

Select „Browse…“ to select DIM_Product.xlsx and add it


Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
Welcome! trans1 Transformation 1
Export_2014_2015_Input
Random_Date_Assignment_Export_2014_2015 4
Export_2015_2016_Input
Random_Date_Assignment_Export_2015_2016
Export_2016_2017_Input
Random_Date_Assignment_Export_2016_2017
DIM_Time
Microsoft Excel input
Step name DIM_Product
Files Sheets Content Error Handling Fields Additional output fields
Name Type Length Precision Trim type Repeat Format Currency Decimal Grouping
1 pc_code String none N
2 pc_description String none N
3 pc_category String none N
Get fields from header row...
Help
OK Preview rows Cancel
Execution Results
Logging Execution History Step Metrics Performance Graph Metrics Preview data
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
Select Fields in the registry and choose „Get fields from header row...“
The columns of DIM_Product.xlsx should now be loaded
Press OK


Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
Welcome! trans1 Transformation 1
100%
Export_2014_2015_Input
Random_Date_Assignment_Export_2014_2015 4
Export_2015_2016_Input
Random_Date_Assignment_Export_2015_2016
Export_2016_2017_Input
Random_Date_Assignment_Export_2016_2017
Stream_Lookup_DIM_Time_Exports
DIM_Time
Stream_Lookup_DIM_Product
Stream lookup
Step name
Stream_Lookup_DIM_Product_Exports
Lookup step
DIM_Product
The key(s) to look up the value(s):
Field LookupField
1 pc_code pc_code
Specify the fields to retrieve :
Field New name Default Type
1 pc_category String
Preserve memory (costs CPU)
Key and value are exactly one integer field
Use sorted list (i.s.o. hashtable)
Help
OK
Cancel
Get Fields
Get lookup fields
Execution Results
Logging Execution History Step Metrics Performance Graph Metrics Preview data
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Ex
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Ex
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Ex
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
View Design
Search
Input
CSV file input
Data grid
De-serialize from file
ESRI shapefile reader
Email messages input
Fixed file input
GZIP CSV input
Generate random credit card numbers
Generate random value
Generate rows
Get data from XML
Get file names
Get files rows count
Get repository names
Get subfolder names
Get system info
Get table names
Google Analytics
HL7 input
JSON input
LDAP input
LDIF input
Load file content in memory
Microsoft Access input
Microsoft Excel input
Mondrian input
OLAP input
Property input
RSS input
S3 CSV input
SAS input
Salesforce input
Table input
Text file input
XBase input
XML input stream (StAX)
YAML input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Double-click on the Lookup Step for DIM_Product and change it according to the picture above
Press OK

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

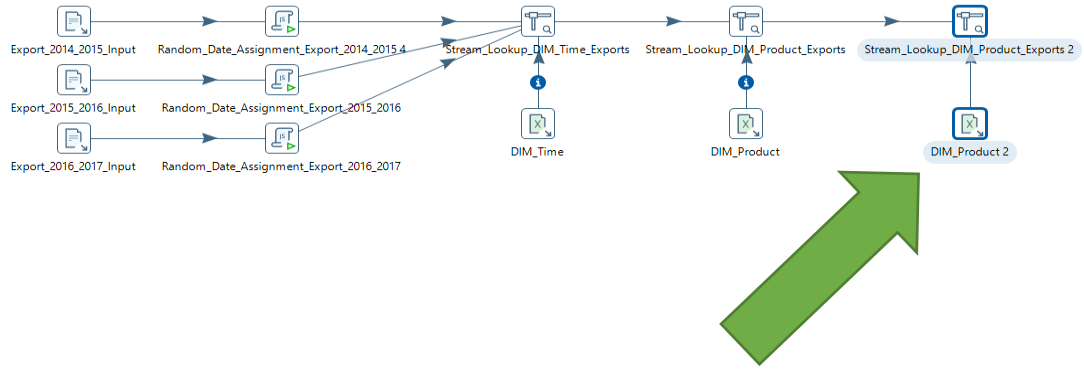
View Design

Search

- Input
  - CSV file input
  - Data grid
  - De-serialize from file
  - ESRI shapefile reader
  - Email messages input
  - Fixed file input
  - GZIP CSV input
  - Generate random credit card numbers
  - Generate random value
  - Generate rows
  - Get data from XML
  - Get file names
  - Get files rows count
  - Get repository names
  - Get subfolder names
  - Get system info
  - Get table names
  - Google Analytics
  - HL7 input
  - JSON input
  - LDAP input
  - LDIF input
  - Load file content in memory
  - Microsoft Access input
  - Microsoft Excel input
  - Mondrian input
  - OLAP input
  - Property input
  - RSS input
  - S3 CSV input
  - SAS input
  - Salesforce input
  - Table input
  - Text file input
  - XBase input
  - XML input stream (StAX)
  - YAML input
- Output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics

Welcome! trans1 Transformation 1

100%



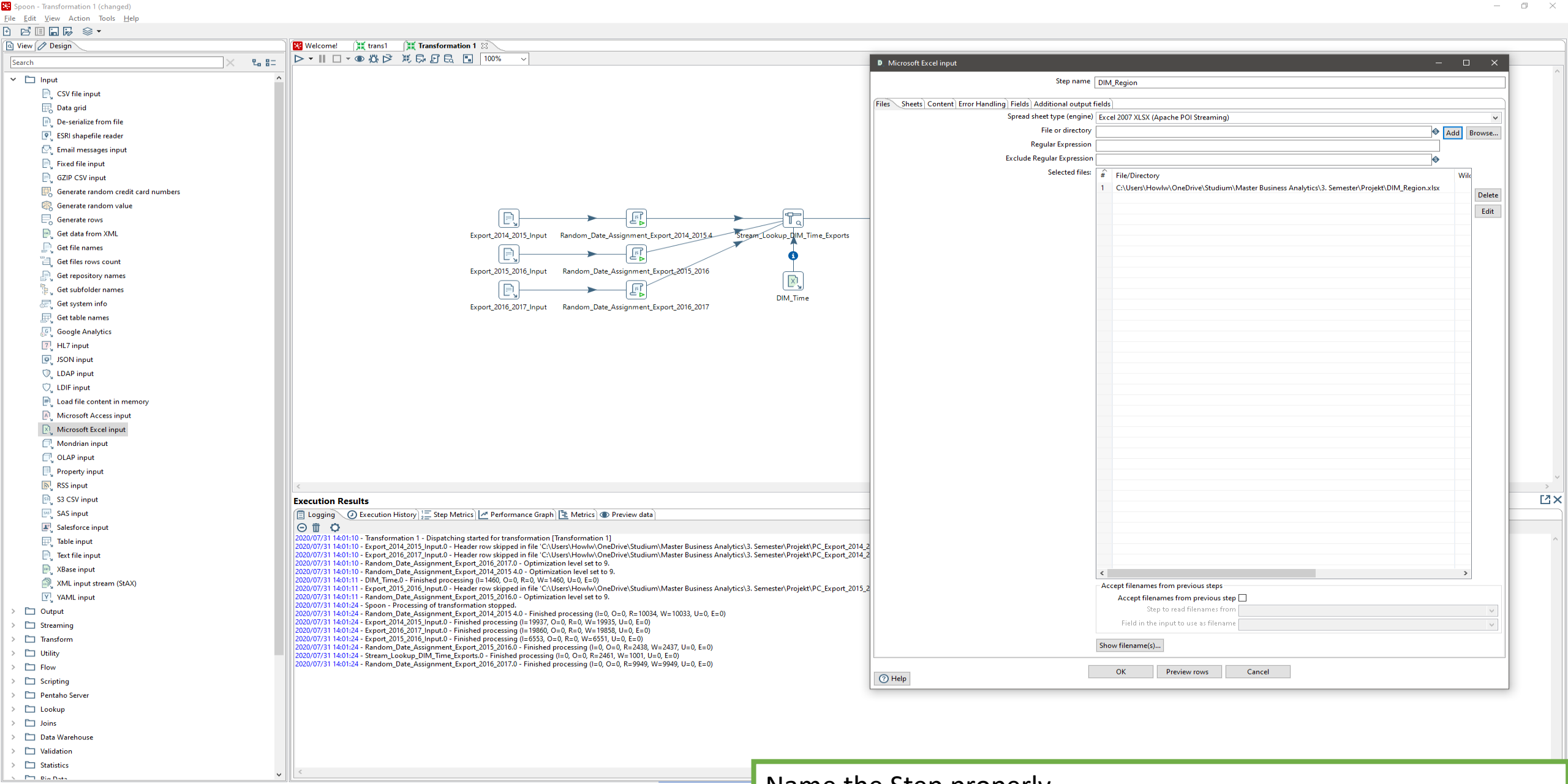
Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

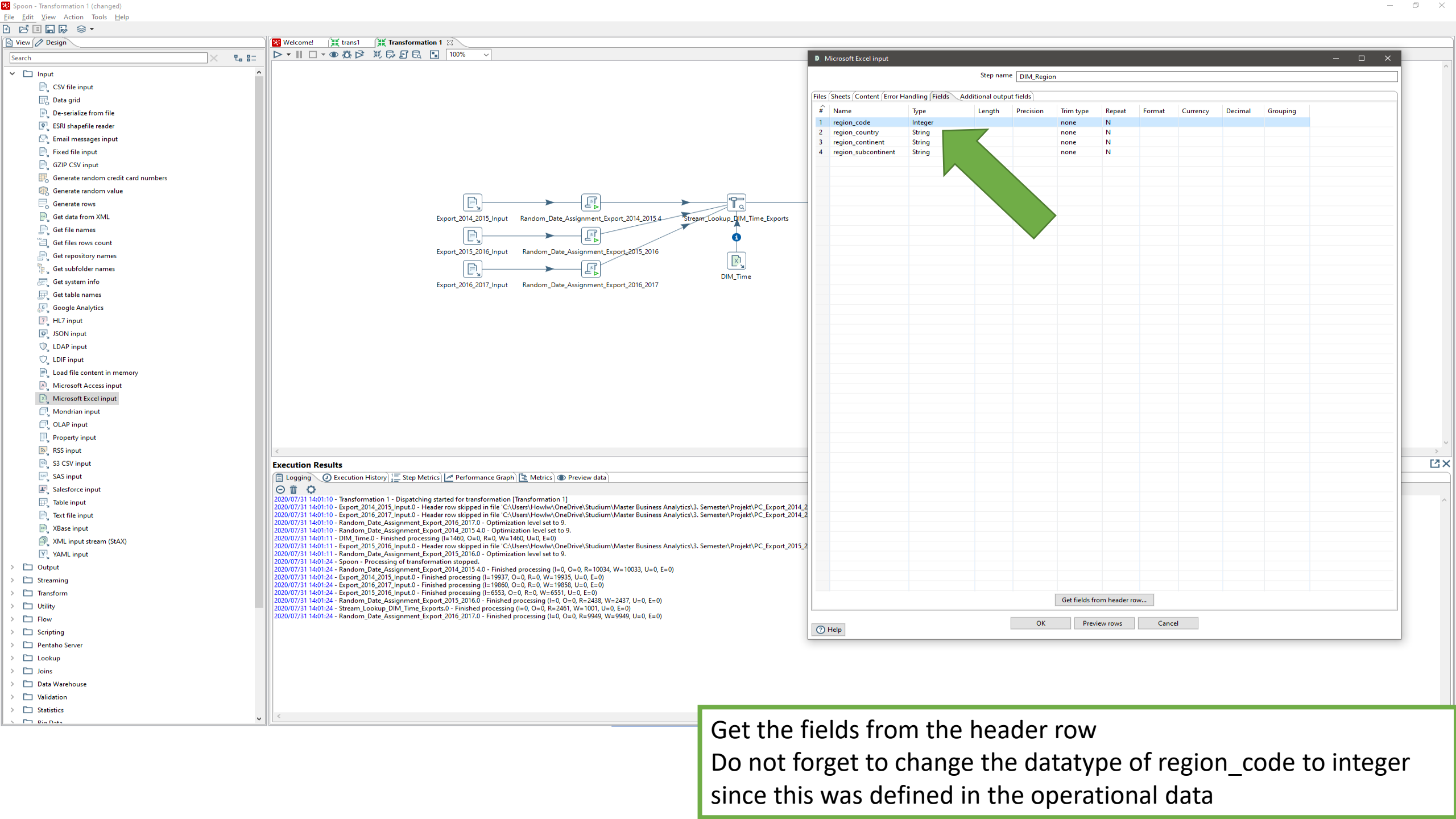
```
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
```

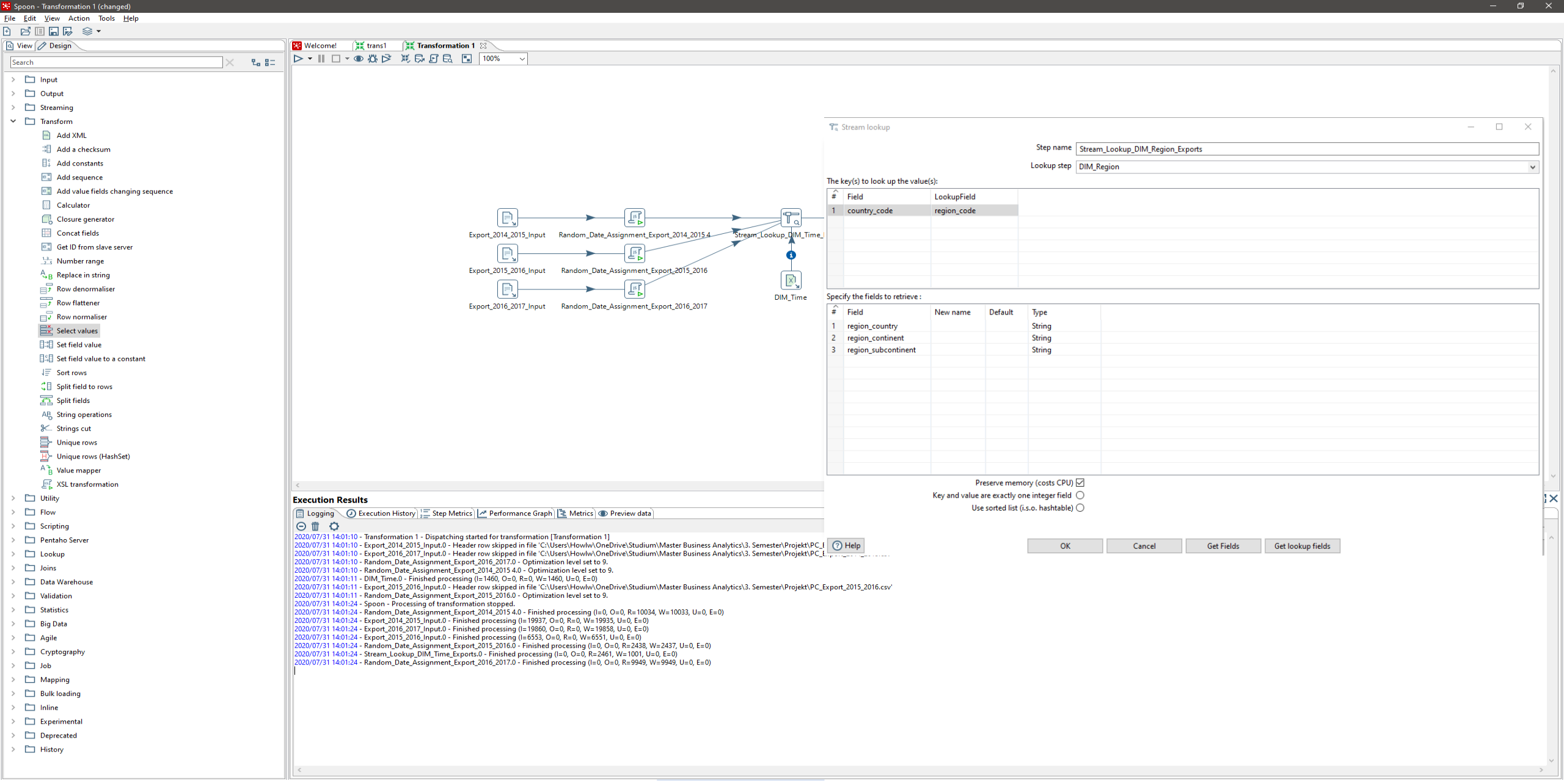
Same procedure for the third and final Dimensiontable integration

Again copy and paste both Stream Lookup and Microsoft Excel Input, connect the steps and then edit them



Name the Step properly
Select the DIM_Region.xlsx file via „Browse…“ and add it
Click on Fields


Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
View Design
Search
Input
CSV file input
Data grid
De-serialize from file
ESRI shapefile reader
Email messages input
Fixed file input
GZIP CSV input
Generate random credit card numbers
Generate random value
Generate rows
Get data from XML
Get file names
Get files rows count
Get repository names
Get subfolder names
Get system info
Get table names
Google Analytics
HL7 input
JSON input
LDAP input
LDIF input
Load file content in memory
Microsoft Access input
Microsoft Excel input
Mondrian input
OLAP input
Property input
RSS input
S3 CSV input
SAS input
Salesforce input
Table input
Text file input
XBase input
XML input stream (StAX)
YAML input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Welcome!
trans1
Transformation 1
100%
Export_2014_2015_Input
Random_Date_Assignment_Export_2014_2015 4
Export_2015_2016_Input
Random_Date_Assignment_Export_2015_2016
Export_2016_2017_Input
Random_Date_Assignment_Export_2016_2017
Stream_Lookup_DIM_Time_Exports
DIM_Time
Execution Results
Logging
Execution History
Step Metrics
Performance Graph
Metrics
Preview data
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
Microsoft Excel input
Step name
DIM_Region
Files
Sheets
Content
Error Handling
Fields
Additional output fields
Name Type Length Precision Trim type Repeat Format Currency Decimal Grouping
1 region_code Integer none N
2 region_country String none N
3 region_continent String none N
4 region_subcontinent String none N
Get fields from header row...
Help
OK
Preview rows
Cancel
Get the fields from the header row
Do not forget to change the datatype of region_code to integer
since this was defined in the operational data



Edit the Lookup Step as stated on the picture above
Press OK

Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

View Design

Search

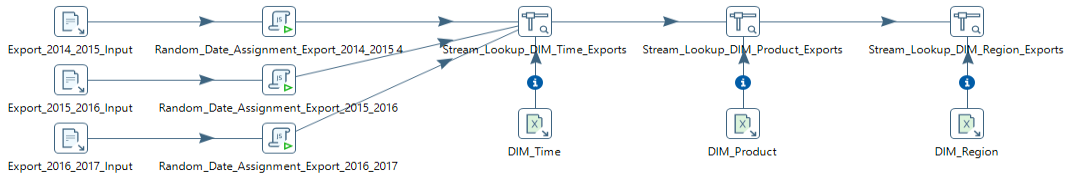
- Input
- Output
- Streaming
- Transform
  - Add XML
  - Add a checksum
  - Add constants
  - Add sequence
  - Add value fields changing sequence
  - Calculator
  - Closure generator
  - Concat fields
  - Get ID from slave server
  - Number range
  - Replace in string
  - Row denormaliser
  - Row flattener
  - Row normaliser
  - Select values
  - Set field value
  - Set field value
  - Sort rows
  - Split field to rows
  - Split fields
  - String operations
  - Strings cut
  - Unique rows
  - Unique rows (HashSet)
  - Value mapper
  - XSL transformation
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

values

move fields in a row.
t the field meta-data: type, length and precision.

Welcome! trans1 Transformation 1

100%

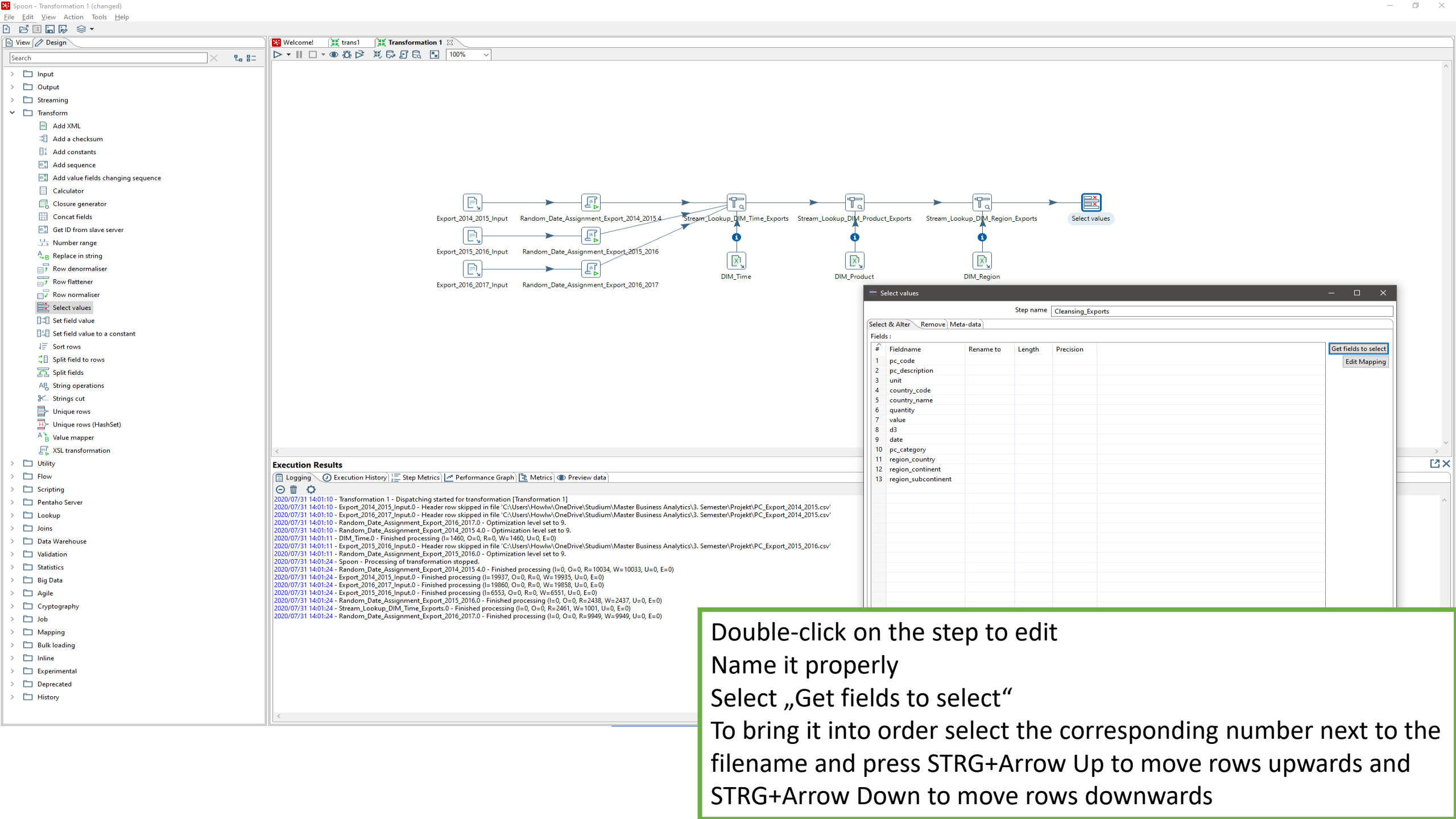


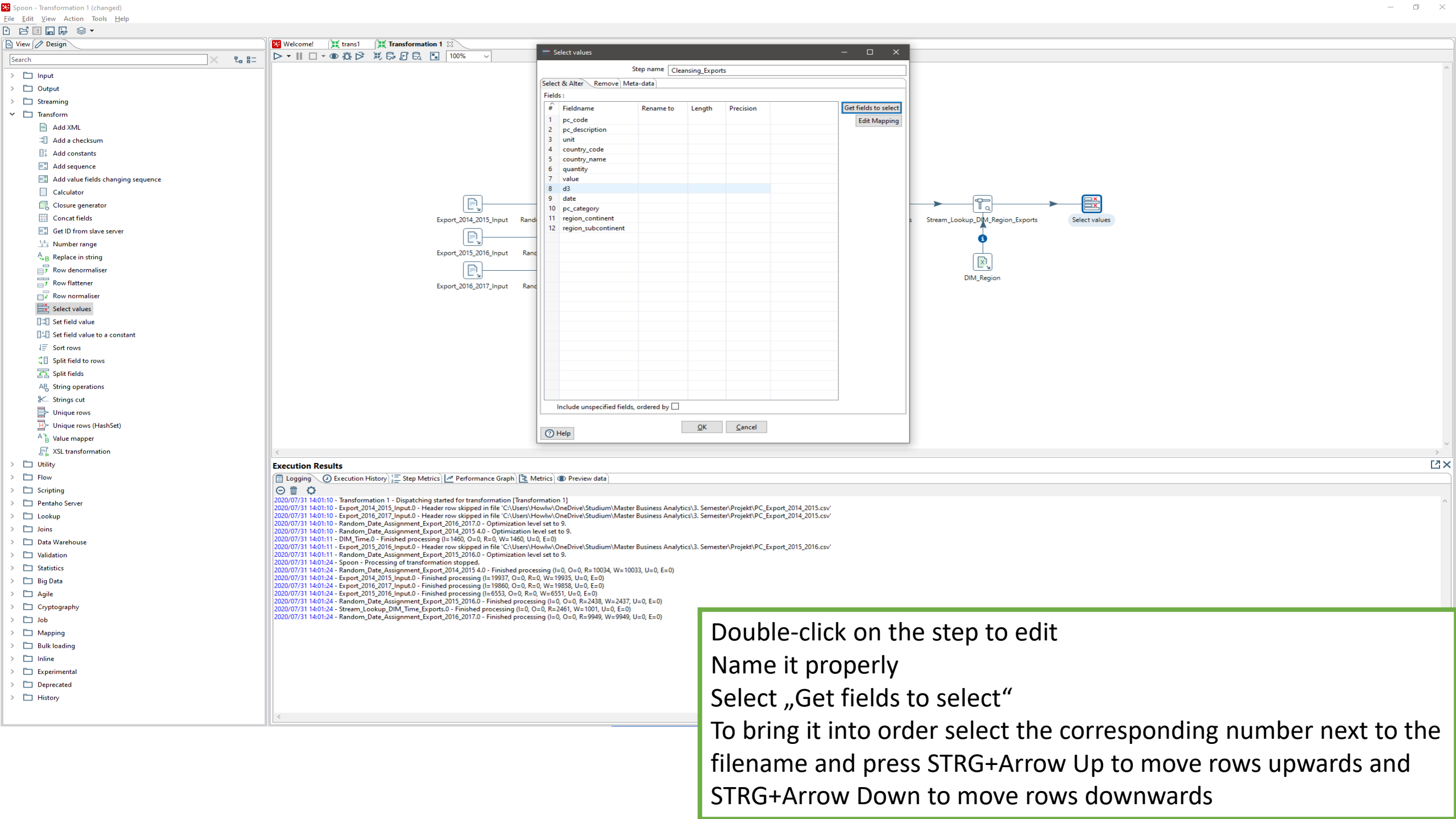
Execution Results

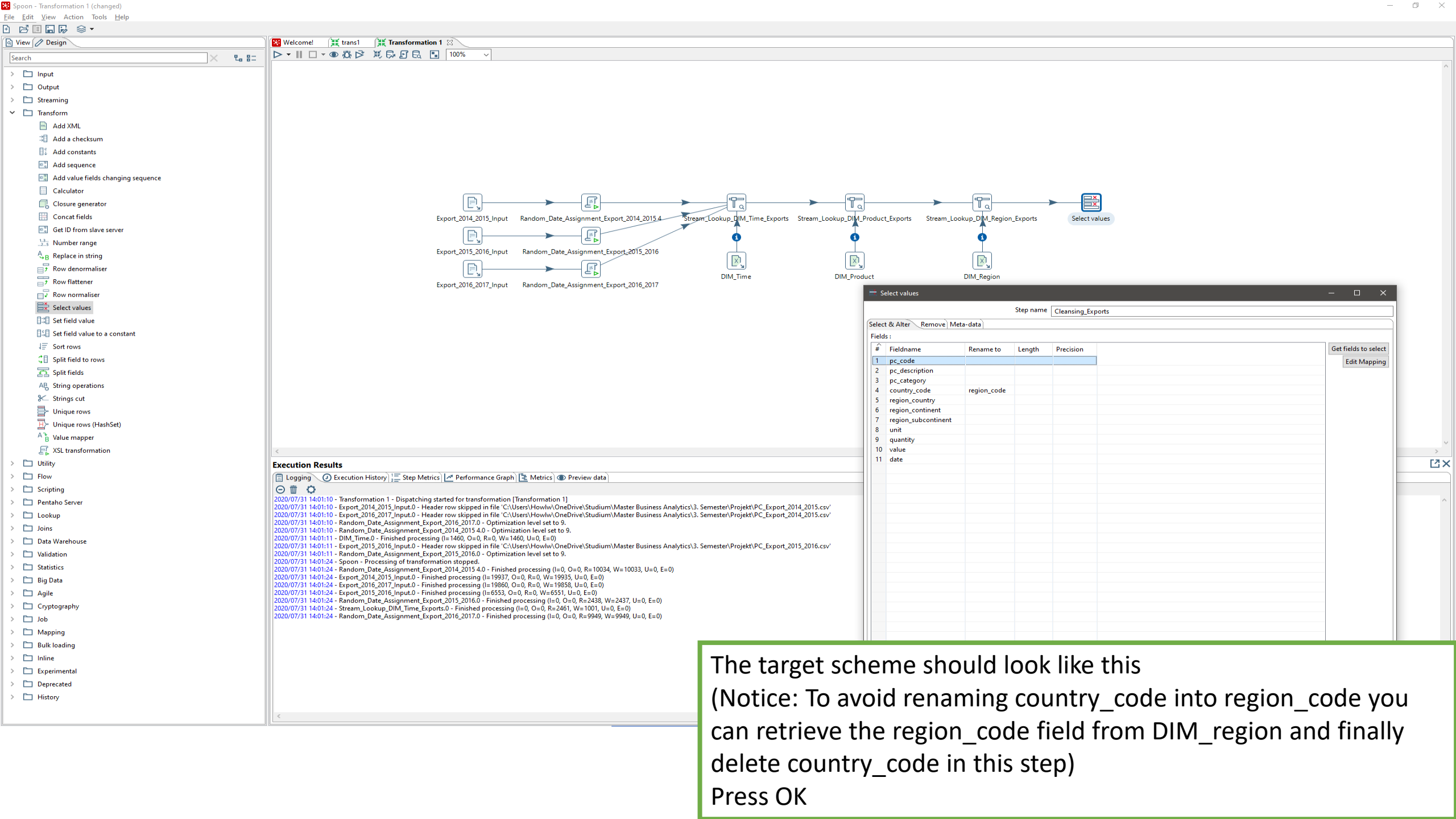
Logging Execution History Step Metrics Performance Graph Metrics Preview data

```
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
```

With completed integration of all three dimension tables the only thing left is cleansing the data and finally outputting it
Drag the „Select values“ step onto the main screen


Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
View Design
Search
Input
Output
Streaming
Transform
Add XML
Add a checksum
Add constants
Add sequence
Add value fields changing sequence
Calculator
Closure generator
Concat fields
Get ID from slave server
Number range
Replace in string
Row denormaliser
Row flattener
Row normaliser
Select values
Set field value
Set field value to a constant
Sort rows
Split field to rows
Split fields
String operations
Strings cut
Unique rows
Unique rows (HashSet)
Value mapper
XSL transformation
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Big Data
Agile
Cryptography
Job
Mapping
Bulk loading
Inline
Experimental
Deprecated
History
Welcome!
trans1
Transformation 1
100%
Export_2014_2015_Input
Export_2015_2016_Input
Export_2016_2017_Input
Random_Date_Assignment_Export_2014_2015 4
Random_Date_Assignment_Export_2015_2016
Random_Date_Assignment_Export_2016_2017
Stream_Lookup_DIM_Time_Exports
Stream_Lookup_DIM_Product_Exports
Stream_Lookup_DIM_Region_Exports
Select values
DIM_Time
DIM_Product
DIM_Region
Execution Results
Logging
Execution History
Step Metrics
Performance Graph
Metrics
Preview data
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
Select values
Step name
Cleansing_Exports
Select & Alter
Remove
Meta-data
Fields :
Fieldname Rename to Length Precision
1 pc_code
2 pc_description
3 unit
4 country_code
5 country_name
6 quantity
7 value
8 d3
9 date
10 pc_category
11 region_country
12 region_continent
13 region_subcontinent
Get fields to select
Edit Mapping
Double-click on the step to edit
Name it properly
Select „Get fields to select“
To bring it into order select the corresponding number next to the filename and press STRG+Arrow Up to move rows upwards and STRG+Arrow Down to move rows downwards


Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
Welcome! trans1 Transformation 1
Select values
Step name Cleansing_Exports
Select & Alter Remove Meta-data
Fields :
Fieldname Rename to Length Precision
1 pc_code
2 pc_description
3 unit
4 country_code
5 country_name
6 quantity
7 value
8 d3
9 date
10 pc_category
11 region_continent
12 region_subcontinent
Get fields to select
Edit Mapping
Include unspecified fields, ordered by
Help
OK
Cancel
Export_2014_2015_Input
Export_2015_2016_Input
Export_2016_2017_Input
Stream_Lookup_DIM_Region_Exports
Select values
DIM_Region
Execution Results
Logging Execution History Step Metrics Performance Graph Metrics Preview data
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
Double-click on the step to edit
Name it properly
Select „Get fields to select“
To bring it into order select the corresponding number next to the filename and press STRG+Arrow Up to move rows upwards and STRG+Arrow Down to move rows downwards


Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
Welcome! trans1 Transformation 1
100%
Export_2014_2015_Input
Export_2015_2016_Input
Export_2016_2017_Input
Random_Date_Assignment_Export_2014_2015 4
Random_Date_Assignment_Export_2015_2016
Random_Date_Assignment_Export_2016_2017
Stream_Lookup_DIM_Time_Exports
Stream_Lookup_DIM_Product_Exports
Stream_Lookup_DIM_Region_Exports
Select values
DIM_Time
DIM_Product
DIM_Region
Execution Results
Logging Execution History Step Metrics Performance Graph Metrics Preview data
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
Select values
Step name Cleansing_Exports
Select & Alter Remove Meta-data
Fields :
Fieldname Rename to Length Precision
1 pc_code
2 pc_description
3 pc_category
4 country_code region_code
5 region_country
6 region_continent
7 region_subcontinent
8 unit
9 quantity
10 value
11 date
Get fields to select
Edit Mapping
The target scheme should look like this
(Notice: To avoid renaming country_code into region_code you can retrieve the region_code field from DIM_region and finally delete country_code in this step)
Press OK

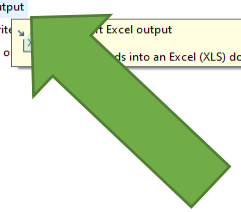
Spoon - Transformation 1 (changed)

File Edit View Action Tools Help

View Design

Search

- Input
- Output
  - Automatic documentation output
  - 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
  - SQL file output
  - Salesforce delete
  - Salesforce insert
  - Salesforce update
  - Salesforce upsert
  - Serialize to file
  - Synchronize after merge
  - Table output
  - Text file output
  - Update
  - XML output
- Streaming
- Transform
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

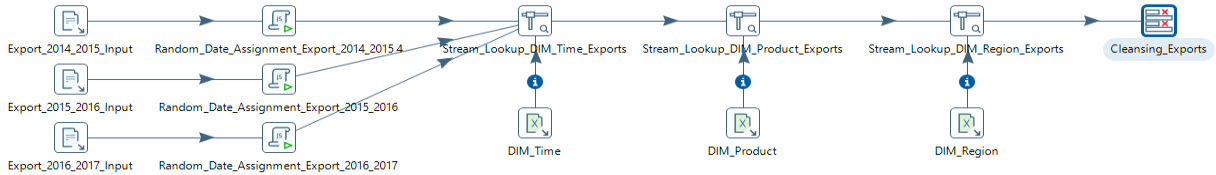


Microsoft Excel output

ds into an Excel (XLS) document with formatting information.

Welcome! trans1 Transformation 1

100%

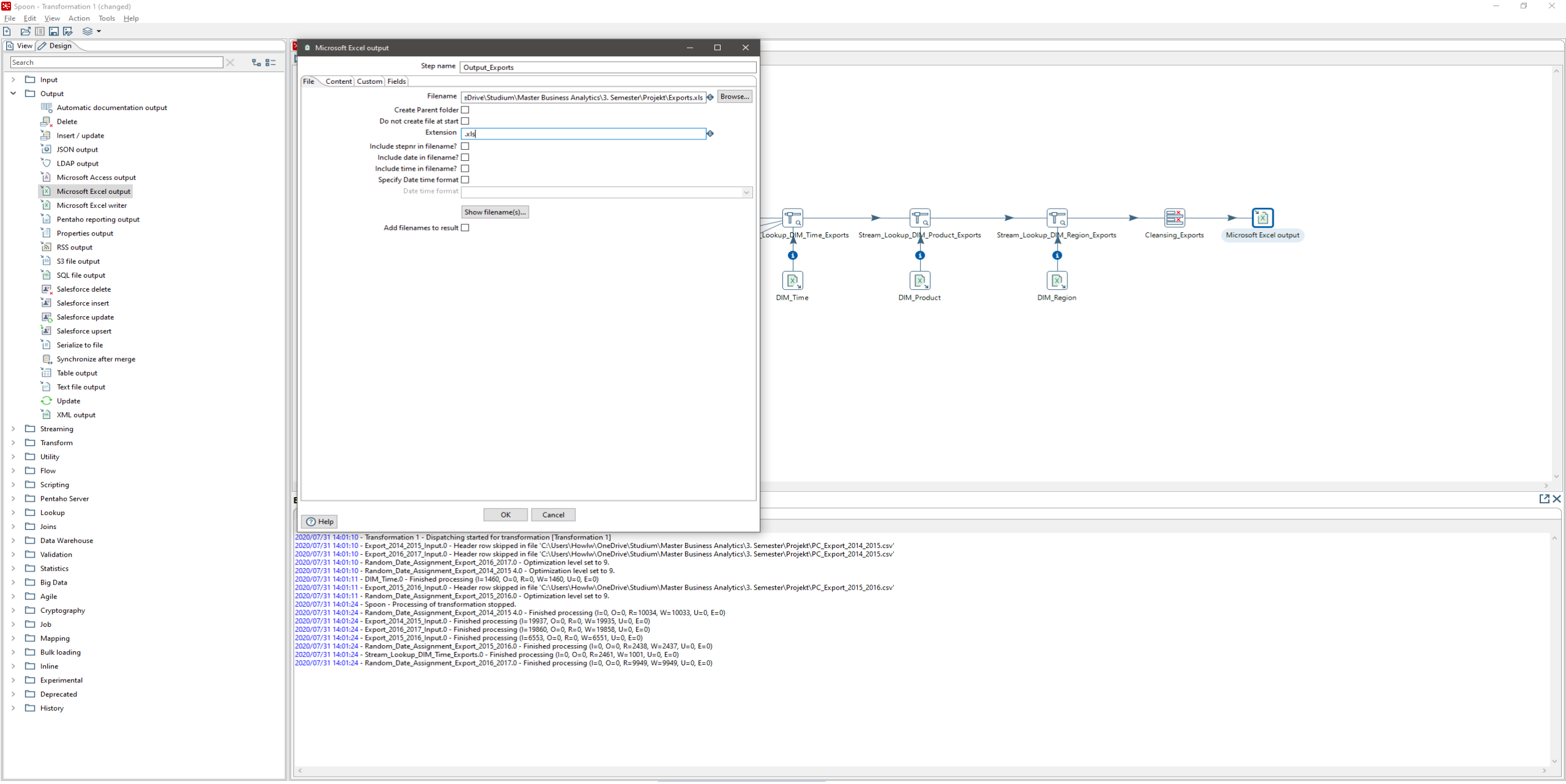


Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

```
2020/07/31 14:01:10 - Transformation 1 - Dispatching started for transformation [Transformation 1]
2020/07/31 14:01:10 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 14:01:10 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 14:01:11 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 14:01:11 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 14:01:11 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 14:01:24 - Spoon - Processing of transformation stopped.
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=10034, W=10033, U=0, E=0)
2020/07/31 14:01:24 - Export_2014_2015_Input.0 - Finished processing (I=19937, O=0, R=0, W=19935, U=0, E=0)
2020/07/31 14:01:24 - Export_2016_2017_Input.0 - Finished processing (I=19860, O=0, R=0, W=19858, U=0, E=0)
2020/07/31 14:01:24 - Export_2015_2016_Input.0 - Finished processing (I=6553, O=0, R=0, W=6551, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=2438, W=2437, U=0, E=0)
2020/07/31 14:01:24 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=2461, W=1001, U=0, E=0)
2020/07/31 14:01:24 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=9949, W=9949, U=0, E=0)
```

Finally for the Output Step drag Microsoft Excel output onto the main screen and double-click it



Name it properly

Click „Browse…“ to choose the destination for saving the file

Spoon - Tuttrans

File Edit View Action Tools Help

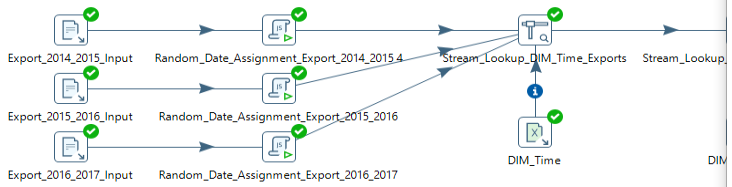
View Design

Search

- Input
- Output
- Streaming
- Transform
  - Add XML
  - Add a checksum
  - Add constants
  - Add sequence
  - Add value fields changing sequence
  - Calculator
  - Closure generator
  - Concat fields
  - Get ID from slave server
  - Number range
  - Replace in string
  - Row denormaliser
  - Row flattener
  - Row normaliser
  - Select values
  - Set field value
  - Set field value to a constant
  - Sort rows
  - Split field to rows
  - Split fields
  - String operations
  - Strings cut
  - Unique rows
  - Unique rows (HashSet)
  - Value mapper
  - XSL transformation
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

Welcome! | trans1 | Tuttrans

100%



Execution Results

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

```
2020/07/31 21:36:16 - Spoon - Using legacy execution engine
2020/07/31 21:36:16 - Spoon - Transformation opened.
2020/07/31 21:36:16 - Spoon - Launching transformation [Tuttrans]...
2020/07/31 21:36:16 - Spoon - Started the transformation execution.
2020/07/31 21:36:16 - Tuttrans - Dispatching started for transformation [Tuttrans]
2020/07/31 21:36:16 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 21:36:16 - DIM_Product.0 - Finished processing (I=168, O=0, R=0, W=168, U=0, E=0)
2020/07/31 21:36:16 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 21:36:16 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 21:36:16 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 21:36:16 - DIM_Region.0 - Finished processing (I=233, O=0, R=0, W=233, U=0, E=0)
2020/07/31 21:36:16 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 21:36:16 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 21:36:16 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 21:36:17 - Export_2014_2015_Input.0 - Finished processing (I=20831, O=0, R=0, W=20830, U=0, E=0)
2020/07/31 21:36:17 - Export_2016_2017_Input.0 - Finished processing (I=20831, O=0, R=0, W=20830, U=0, E=0)
2020/07/31 21:36:17 - Export_2015_2016_Input.0 - Finished processing (I=20772, O=0, R=0, W=20771, U=0, E=0)
2020/07/31 21:36:17 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=20830, W=20830, U=0, E=0)
2020/07/31 21:36:17 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=20830, W=20830, U=0, E=0)
2020/07/31 21:36:17 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=20771, W=20771, U=0, E=0)
2020/07/31 21:36:32 - Stream_Lookup_DIM_Time_Exports.0 - linenr 50000
2020/07/31 21:37:46 - Stream_Lookup_DIM_Product_Exports.0 - linenr 50000
2020/07/31 21:38:16 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=63891, W=62431, U=0, E=0)
```

Microsoft Excel output

Step name: Output_Exports

File | Content | Custom | Fields

#	Name	Type	Format
1	pc_code	String	
2	pc_description	String	
3	pc_category	String	
4	region_code	Integer	
5	region_country	String	
6	region_continent	String	
7	region_subcontinent	String	
8	unit	String	
9	quantity	Integer	
10	value	Number	0000000000000.00
11	date	Date	

Get Fields | Minimal width

Help | OK | Cancel

Switch to Fields and edit the Fields accordingly
Press OK

Spoon - Tuttrans

File Edit View Action Tools Help

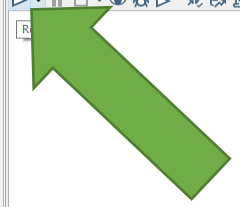
View Design

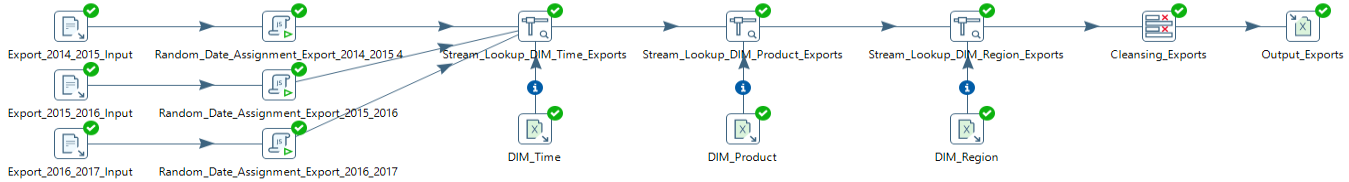
Search

- Input
- Output
- Streaming
- Transform
  - Add XML
  - Add a checksum
  - Add constants
  - Add sequence
  - Add value fields changing sequence
  - Calculator
  - Closure generator
  - Concat fields
  - Get ID from slave server
  - Number range
  - Replace in string
  - Row denormaliser
  - Row flattener
  - Row normaliser
  - Select values
  - Set field value
  - Set field value to a constant
  - Sort rows
  - Split field to rows
  - Split fields
  - String operations
  - Strings cut
  - Unique rows
  - Unique rows (HashSet)
  - Value mapper
  - XSL transformation
- Utility
- Flow
- Scripting
- Pentaho Server
- Lookup
- Joins
- Data Warehouse
- Validation
- Statistics
- Big Data
- Agile
- Cryptography
- Job
- Mapping
- Bulk loading
- Inline
- Experimental
- Deprecated
- History

Welcome! trans1 Tuttrans

100%





Execution Results

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

```
2020/07/31 21:36:16 - Tuttrans - Dispatching started for transformation [Tuttrans]
2020/07/31 21:36:16 - Export_2015_2016_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2015_2016.csv'
2020/07/31 21:36:16 - DIM_Product.0 - Finished processing (I=168, O=0, R=0, W=168, U=0, E=0)
2020/07/31 21:36:16 - Random_Date_Assignment_Export_2015_2016.0 - Optimization level set to 9.
2020/07/31 21:36:16 - Export_2016_2017_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 21:36:16 - Random_Date_Assignment_Export_2016_2017.0 - Optimization level set to 9.
2020/07/31 21:36:16 - DIM_Region.0 - Finished processing (I=233, O=0, R=0, W=233, U=0, E=0)
2020/07/31 21:36:16 - Export_2014_2015_Input.0 - Header row skipped in file 'C:\Users\Howlw\OneDrive\Studium\Master Business Analytics\3. Semester\Projekt\PC_Export_2014_2015.csv'
2020/07/31 21:36:16 - Random_Date_Assignment_Export_2014_2015 4.0 - Optimization level set to 9.
2020/07/31 21:36:16 - DIM_Time.0 - Finished processing (I=1460, O=0, R=0, W=1460, U=0, E=0)
2020/07/31 21:36:17 - Export_2014_2015_Input.0 - Finished processing (I=20831, O=0, R=0, W=20830, U=0, E=0)
2020/07/31 21:36:17 - Export_2016_2017_Input.0 - Finished processing (I=20831, O=0, R=0, W=20830, U=0, E=0)
2020/07/31 21:36:17 - Export_2015_2016_Input.0 - Finished processing (I=20772, O=0, R=0, W=20771, U=0, E=0)
2020/07/31 21:36:17 - Random_Date_Assignment_Export_2014_2015 4.0 - Finished processing (I=0, O=0, R=20830, W=20830, U=0, E=0)
2020/07/31 21:36:17 - Random_Date_Assignment_Export_2016_2017.0 - Finished processing (I=0, O=0, R=20830, W=20830, U=0, E=0)
2020/07/31 21:36:17 - Random_Date_Assignment_Export_2015_2016.0 - Finished processing (I=0, O=0, R=20771, W=20771, U=0, E=0)
2020/07/31 21:36:32 - Stream_Lookup_DIM_Time_Exports.0 - linenr 50000
2020/07/31 21:37:46 - Stream_Lookup_DIM_Product_Exports.0 - linenr 50000
2020/07/31 21:38:16 - Stream_Lookup_DIM_Time_Exports.0 - Finished processing (I=0, O=0, R=63891, W=62431, U=0, E=0)
2020/07/31 21:39:57 - Stream_Lookup_DIM_Region_Exports.0 - linenr 50000
2020/07/31 21:40:40 - Stream_Lookup_DIM_Product_Exports.0 - Finished processing (I=0, O=0, R=62599, W=62431, U=0, E=0)
2020/07/31 21:42:55 - Cleansing_Exports.0 - linenr 50000
2020/07/31 21:43:45 - Stream_Lookup_DIM_Region_Exports.0 - Finished processing (I=0, O=0, R=62664, W=62431, U=0, E=0)
2020/07/31 21:46:40 - Output_Exports.0 - linenr 50000
2020/07/31 21:47:43 - Cleansing_Exports.0 - Finished processing (I=0, O=0, R=62431, W=62431, U=0, E=0)
2020/07/31 21:52:31 - Output_Exports.0 - Finished processing (I=0, O=62431, R=62431, W=62431, U=0, E=0)
2020/07/31 21:52:31 - Spoon - The transformation has finished!!
```

Finally press „Run“ and then again „Run“ without changing anything
The transformation should start and finish successfully