Департамент образования и науки города Москвы
Государственное автономное образовательное учреждение
высшего образования города Москвы
«Московский городской педагогический университет»
Институт цифрового образования
Департамент информатики, управления и технологий

ДИСЦИПЛИНА:

Инструменты для хранения и обработки больших данных

Домашнее задание 1

Тема:

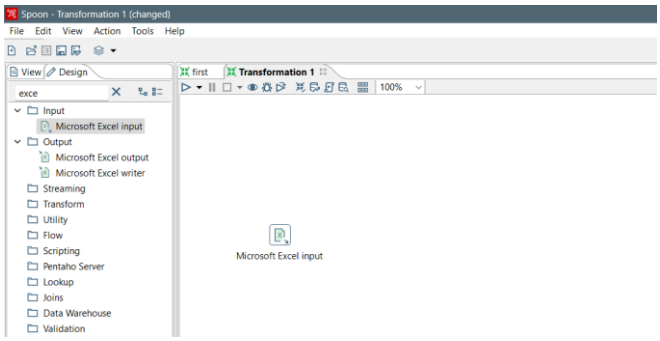
Pentaho.

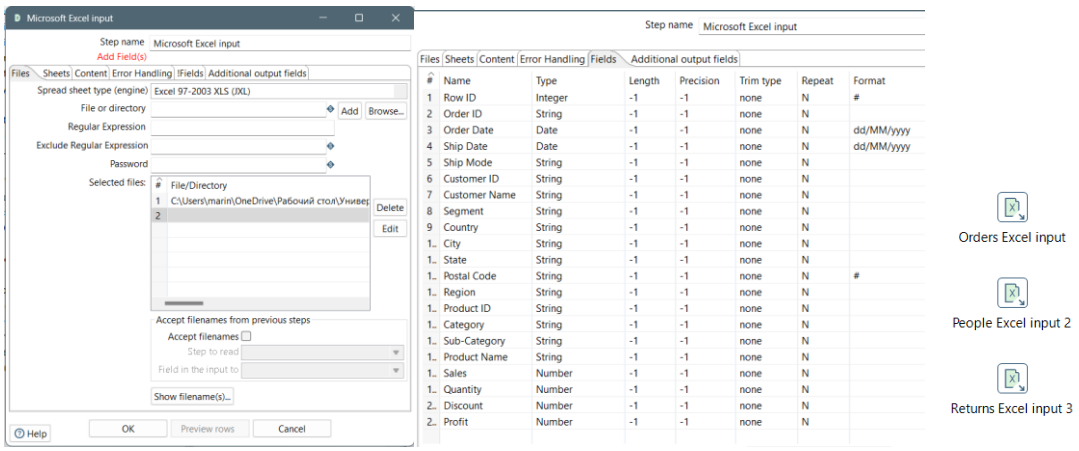
Выполнила: Соколова М. С., группа: АДЭУ-201

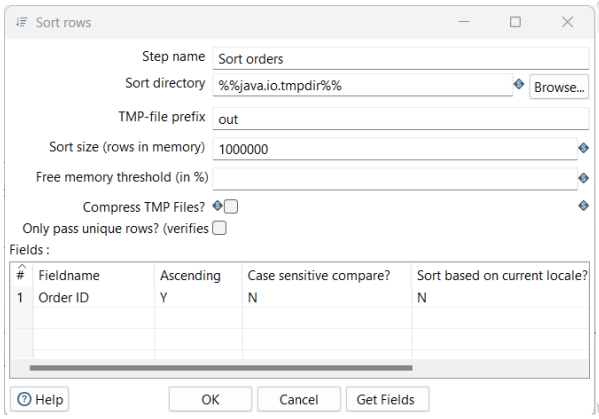
Преподаватель: Босенко Т. М.

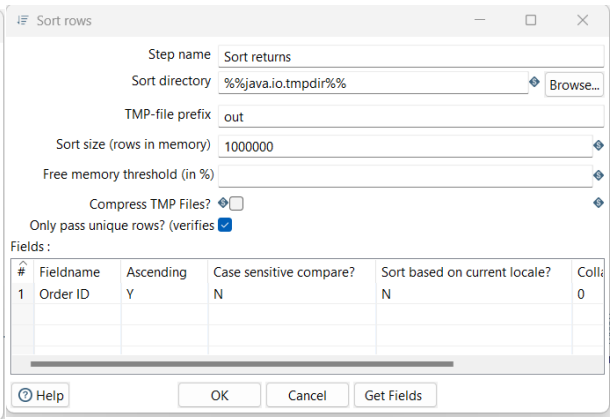
Москва

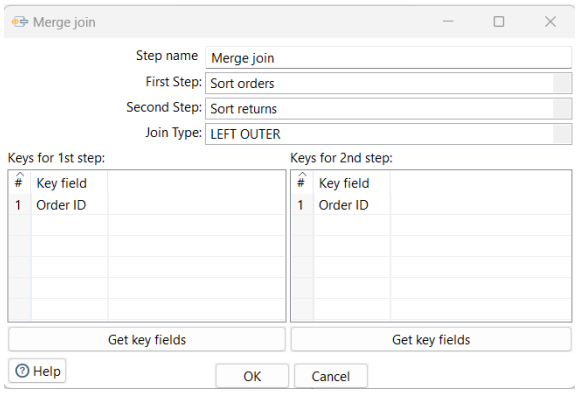
2023

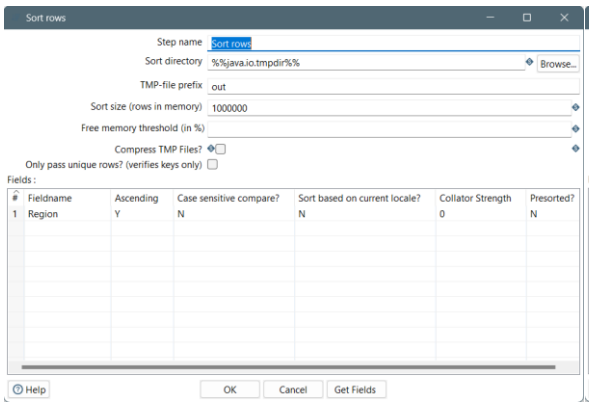

Spoon - Transformation 1 (changed)
File Edit View Action Tools Help
View Design
Input
Microsoft Excel input
Output
Microsoft Excel output
Microsoft Excel writer
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
first
Transformation 1
100%
Microsoft Excel input

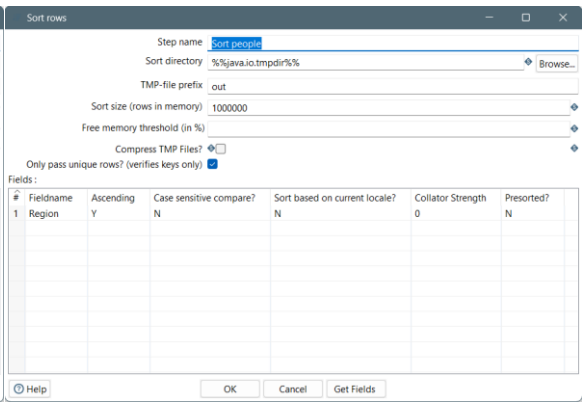

Microsoft Excel input
Step name Microsoft Excel input
Add Field(s)
Files Sheets Content Error Handling Fields Additional output fields
Spread sheet type (engine) Excel 97-2003 XLS (JXL)
File or directory
Add Browse...
Regular Expression
Exclude Regular Expression
Password
Selected files:
File/Directory
C:\Users\marin\OneDrive\Рабочий стол\Универ
Delete
Edit
Accept filenames from previous steps
Accept filenames
Step to read
Field in the input to
Show filename(s)...
Help
OK
Preview rows
Cancel
Step name Microsoft Excel input
Files Sheets Content Error Handling Fields Additional output fields
Name Type Length Precision Trim type Repeat Format
1 Row ID Integer -1 -1 none N #
2 Order ID String -1 -1 none N
3 Order Date Date -1 -1 none N dd/MM/yyyy
4 Ship Date Date -1 -1 none N dd/MM/yyyy
5 Ship Mode String -1 -1 none N
6 Customer ID String -1 -1 none N
7 Customer Name String -1 -1 none N
8 Segment String -1 -1 none N
9 Country String -1 -1 none N
1.. City String -1 -1 none N
1.. State String -1 -1 none N
1.. Postal Code String -1 -1 none N #
1.. Region String -1 -1 none N
1.. Product ID String -1 -1 none N
1.. Category String -1 -1 none N
1.. Sub-Category String -1 -1 none N
1.. Product Name String -1 -1 none N
1.. Sales Number -1 -1 none N
1.. Quantity Number -1 -1 none N
2.. Discount Number -1 -1 none N
2.. Profit Number -1 -1 none N
Orders Excel input
People Excel input 2
Returns Excel input 3

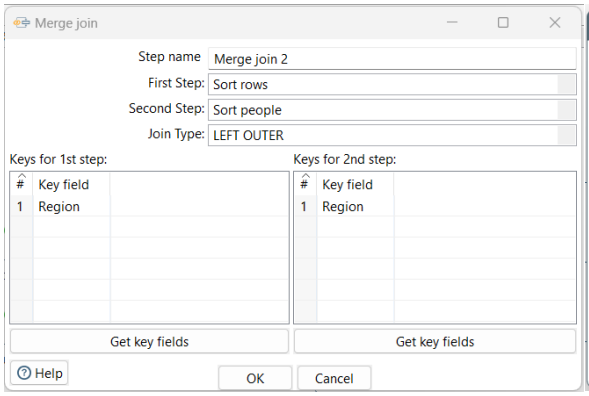

Sort rows
Step name Sort orders
Sort directory %%java.io.tmpdir%%
Browse...
TMP-file prefix out
Sort size (rows in memory) 1000000
Free memory threshold (in %)
Compress TMP Files?
Only pass unique rows? (verifies
Fields :
Fieldname Ascending Case sensitive compare? Sort based on current locale?
1 Order ID Y N N
Help
OK
Cancel
Get Fields

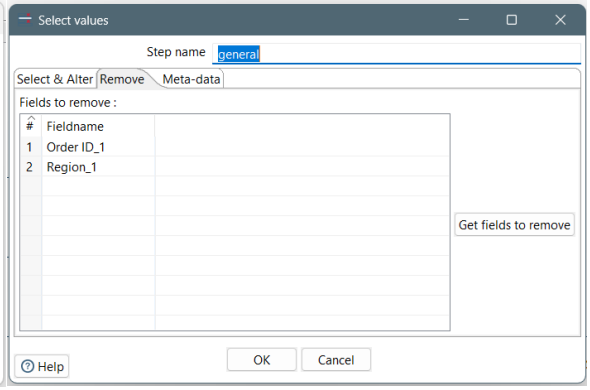

Sort rows
Step name Sort returns
Sort directory %%java.io.tmpdir%%
Browse...
TMP-file prefix out
Sort size (rows in memory) 1000000
Free memory threshold (in %)
Compress TMP Files?
Only pass unique rows? (verifies
Fields :
Fieldname Ascending Case sensitive compare? Sort based on current locale? Colla
1 Order ID Y N N 0
Help
OK
Cancel
Get Fields

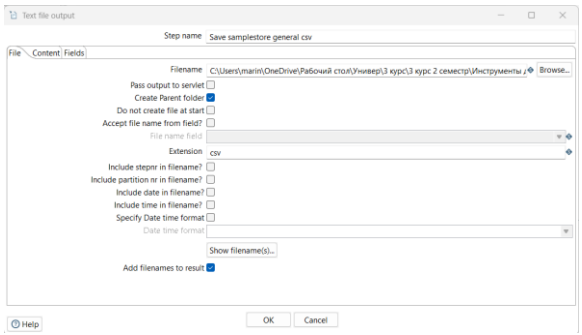

Merge join
Step name Merge join
First Step: Sort orders
Second Step: Sort returns
Join Type: LEFT OUTER
Keys for 1st step:
Key field
1 Order ID
Keys for 2nd step:
Key field
1 Order ID
Get key fields
Get key fields
Help
OK
Cancel

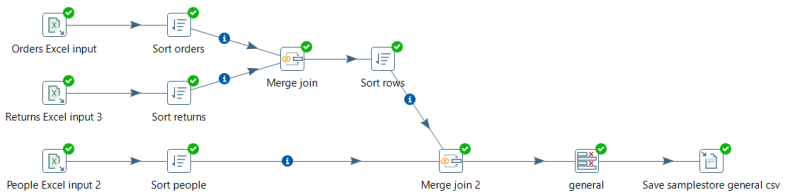

Sort rows
Step name Sort rows
Sort directory %%java.io.tmpdir%%
Browse...
TMP-file prefix out
Sort size (rows in memory) 1000000
Free memory threshold (in %)
Compress TMP Files?
Only pass unique rows? (verifies keys only)
Fields :
Fieldname Ascending Case sensitive compare? Sort based on current locale? Collator Strength Presorted?
1 Region Y N N 0 N
Help
OK
Cancel
Get Fields

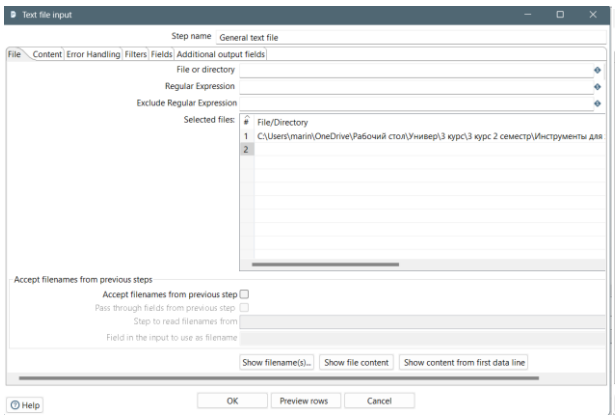

Sort rows
Step name Sort people
Sort directory %%java.io.tmpdir%%
Browse...
TMP-file prefix out
Sort size (rows in memory) 1000000
Free memory threshold (in %)
Compress TMP Files?
Only pass unique rows? (verifies keys only)
Fields :
Fieldname Ascending Case sensitive compare? Sort based on current locale? Collator Strength Presorted?
1 Region Y N N 0 N
Help
OK
Cancel
Get Fields

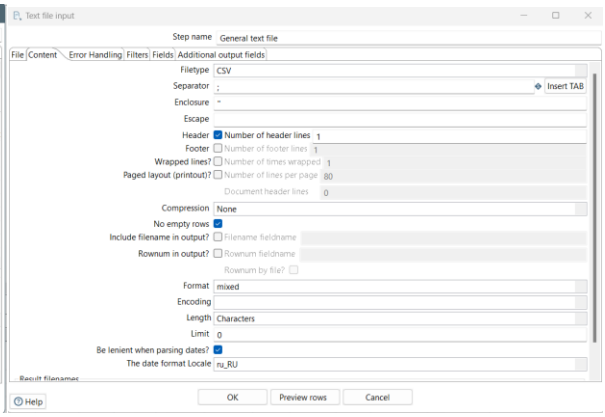

Merge join
Step name Merge join 2
First Step: Sort rows
Second Step: Sort people
Join Type: LEFT OUTER
Keys for 1st step:
Key field
1 Region
Keys for 2nd step:
Key field
1 Region
Get key fields
Get key fields
Help
OK
Cancel

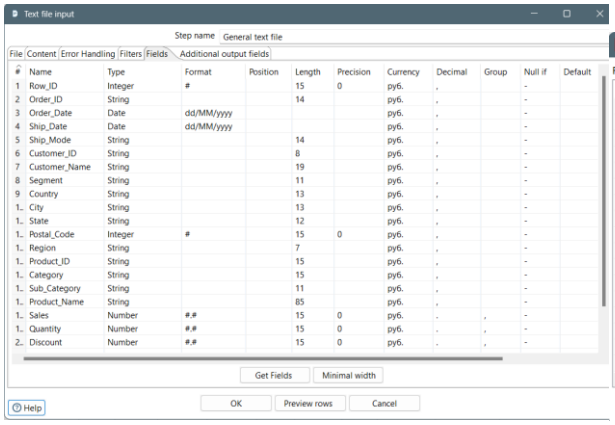

Select values
Step name general
Select & Alter
Remove
Meta-data
Fields to remove :
Fieldname
1 Order ID_1
2 Region_1
Get fields to remove
Help
OK
Cancel

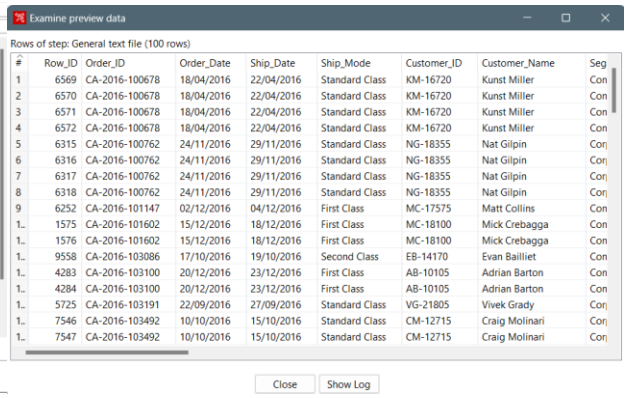

Text file output
Step name Save samplestore general csv
File
Content
Fields
Filename
Browse...
Pass output to servlet
Create Parent folder
Do not create file at start
Accept file name from field?
File name field
Extension csv
Include stepnr in filename?
Include partition nr in filename?
Include date in filename?
Include time in filename?
Specify Date time format
Date time format
Show filename(s)...
Add filenames to result
Help
OK
Cancel

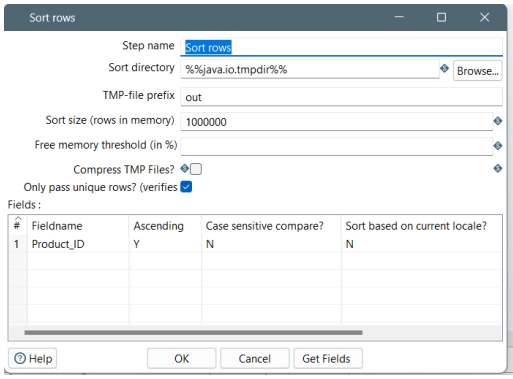

Orders Excel input
Sort orders
Merge join
Sort rows
Returns Excel input 3
Sort returns
People Excel input 2
Sort people
Merge join 2
general
Save samplestore general csv

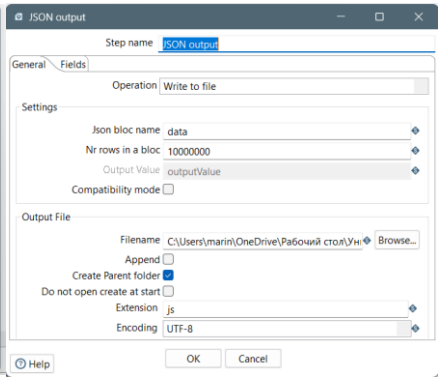


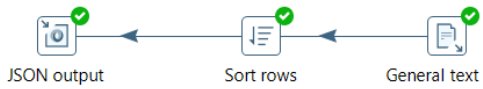








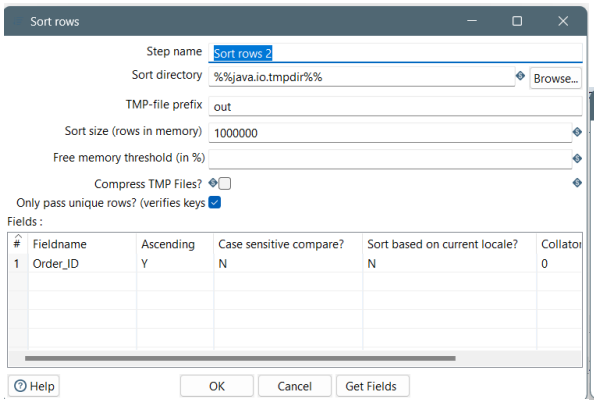


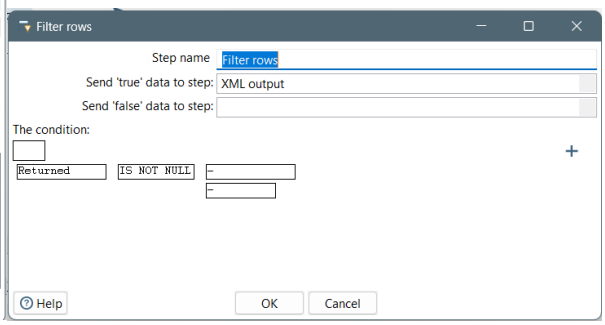


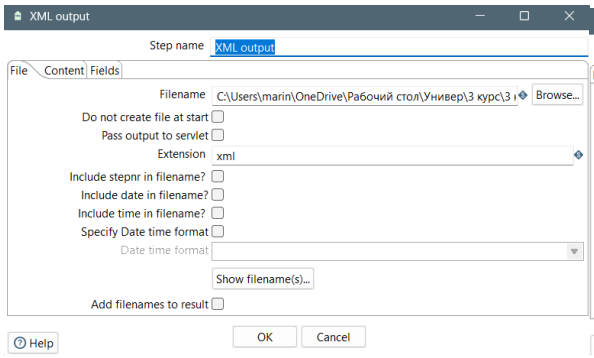
products_0.js

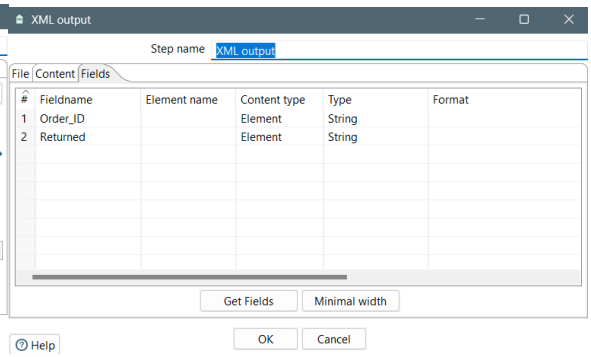
Файл Изменить Просмотр

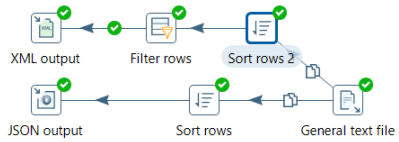

```
{"data":[{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10000112","Product_Name":"Bush Birmingham Collection Bookcase, Dark
Cherry"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10000330","Product_Name":"Sauder Camden County Barrister Bookcase, Planked Cherry
Finish"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10000362","Product_Name":"Sauder Inglewood Library
Bookcases"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10000468","Product_Name":"O'Sullivan 2-Shelf Heavy-Duty
Bookcases"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10000711","Product_Name":"Hon Metal Bookcases,
Gray"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10000780","Product_Name":"O'Sullivan Plantations 2-Door Library in Landvery
Oak"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001337","Product_Name":"O'Sullivan Living Dimensions 2-Shelf
Bookcases"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001519","Product_Name":"O'Sullivan 3-Shelf Heavy-Duty
Bookcases"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001567","Product_Name":"Bush Westfield Collection Bookcases, Dark Cherry Finish, Fully
Assembled"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001601","Product_Name":"Sauder Mission Library with Doors, Fruitwood
Finish"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001608","Product_Name":"Hon Metal Bookcases,
Black"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001619","Product_Name":"O'Sullivan Cherrywood Estates Traditional
Bookcase"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001798","Product_Name":"Bush Somerset Collection
Bookcase"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001811","Product_Name":"Atlantic Metals Mobile 5-Shelf Bookcases, Custom
Colors"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001918","Product_Name":"Sauder Forest Hills Library with Doors, Woodland Oak
Finish"},{"Sub_Category":"Bookcases","Category":"Furniture","Product_ID":"FUR-BO-10001972","Product_Name":"O'Sullivan 4-Shelf Bookcase in Odessa
```



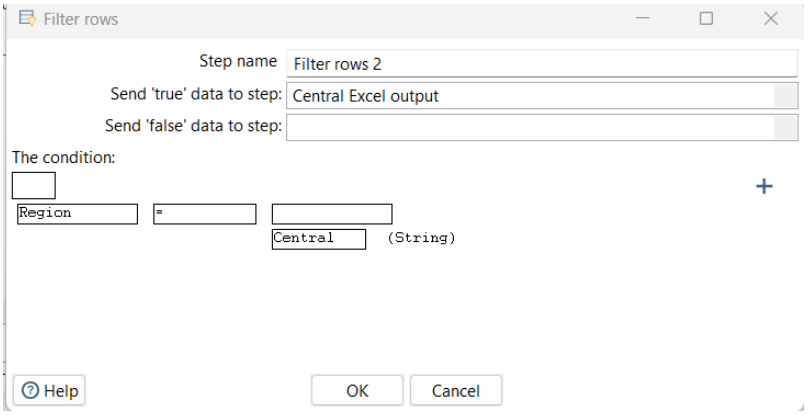


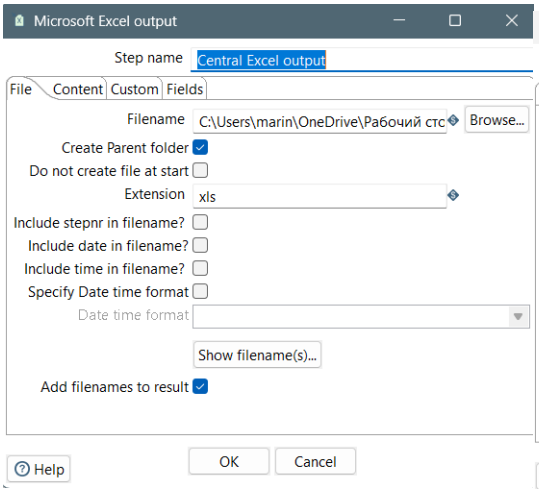


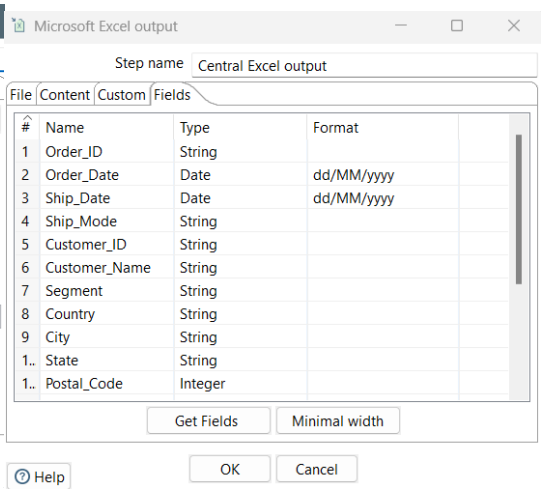


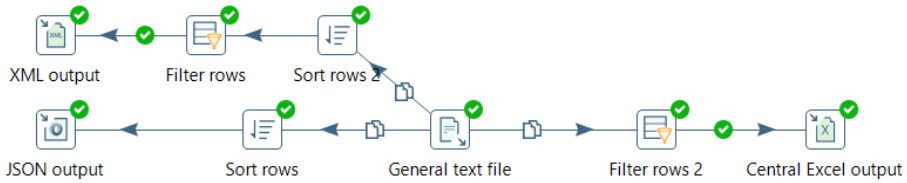


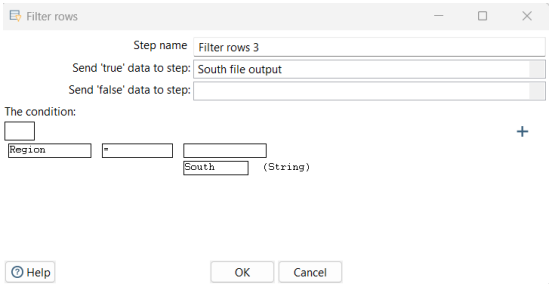

```
<?xml version='1.0' encoding='UTF-8'?>
<Rows>
<Row><Order_ID>CA-2016-100762</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-100867</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-102652</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-103373</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-103744</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-103940</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-104829</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-105270</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-108609</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-108861</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-109918</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-110786</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-111871</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-116785</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-123225</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-123253</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-123498</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-124688</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-126361</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-126403</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-126522</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-127012</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-127131</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-133690</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-134726</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-135657</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-135699</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-140816</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-141726</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-142769</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-143336</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-143840</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-148614</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-148950</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-151162</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-152345</Order_ID> <Returned>Yes</Returned></Row>
<Row><Order_ID>CA-2016-153150</Order_ID> <Returned>Yes</Returned></Row>
```

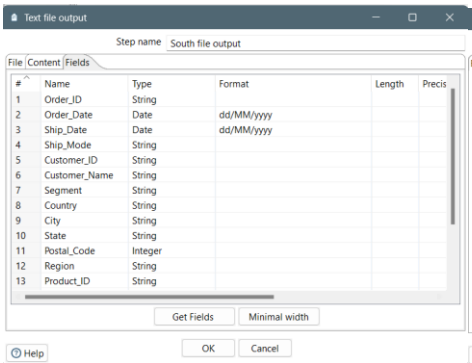

Filter rows
Step name
Filter rows 2
Send 'true' data to step:
Central Excel output
Send 'false' data to step:
The condition:
Region
=
Central
(String)
Help
OK
Cancel

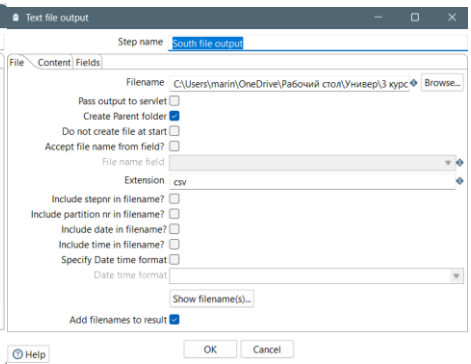

Microsoft Excel output
Step name
Central Excel output
File
Content
Custom
Fields
Filename
C:\Users\marin\OneDrive\Рабочий стс
Browse...
Create Parent folder
Do not create file at start
Extension
xls
Include stepnr in filename?
Include date in filename?
Include time in filename?
Specify Date time format
Date time format
Show filename(s)...
Add filenames to result
Help
OK
Cancel

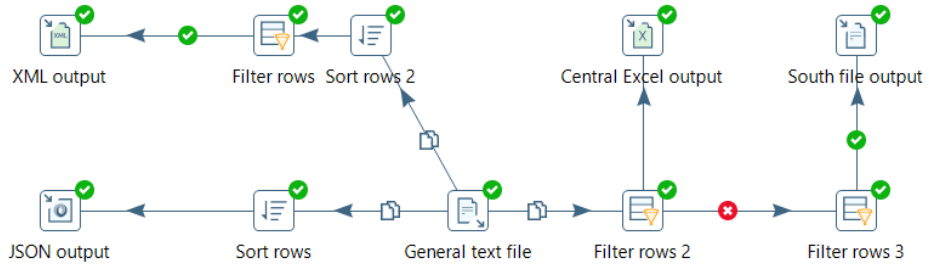

Microsoft Excel output
Step name
Central Excel output
File
Content
Custom
Fields
Name Type Format
1 Order_ID String
2 Order_Date Date dd/MM/yyyy
3 Ship_Date Date dd/MM/yyyy
4 Ship_Mode String
5 Customer_ID String
6 Customer_Name String
7 Segment String
8 Country String
9 City String
1.. State String
1.. Postal_Code Integer
Get Fields
Minimal width
Help
OK
Cancel

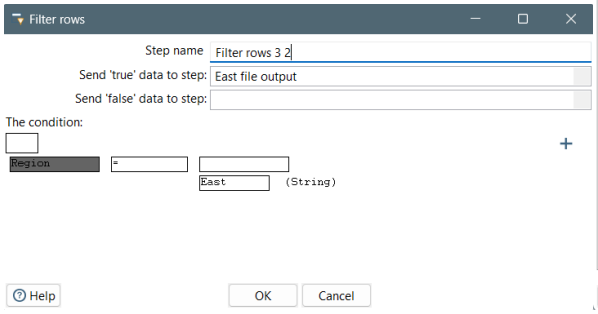

XML output
Filter rows
Sort rows 2
JSON output
Sort rows
General text file
Filter rows 2
Central Excel output

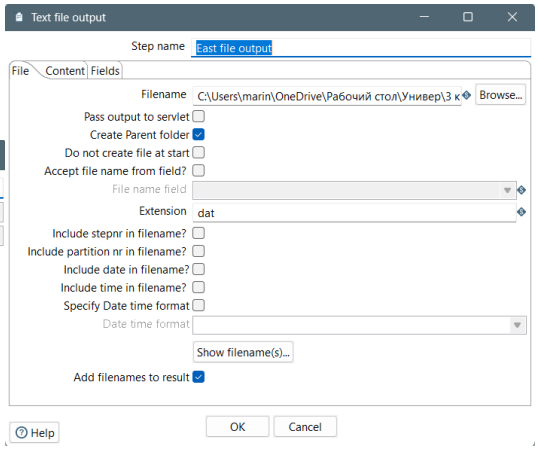

Filter rows
Step name
Filter rows 3
Send 'true' data to step:
South file output
Send 'false' data to step:
The condition:
Region
=
South
(String)
Help
OK
Cancel

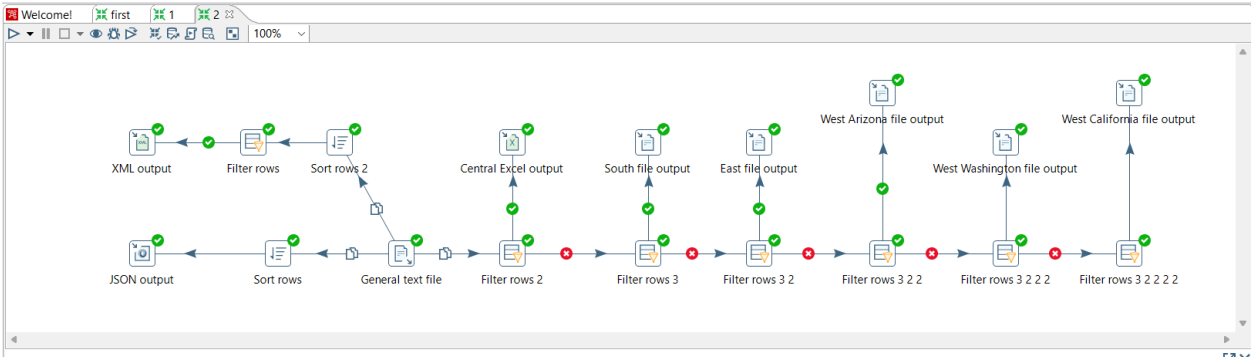

Text file output
Step name
South file output
File
Content
Fields
Name Type Format Length Precis
1 Order_ID String
2 Order_Date Date dd/MM/yyyy
3 Ship_Date Date dd/MM/yyyy
4 Ship_Mode String
5 Customer_ID String
6 Customer_Name String
7 Segment String
8 Country String
9 City String
10 State String
11 Postal_Code Integer
12 Region String
13 Product_ID String
Get Fields
Minimal width
Help
OK
Cancel


Text file output
Step name
South file output
File
Content
Fields
Filename
C:\Users\marin\OneDrive\Рабочий стол\Универ\3 курс
Browse...
Pass output to servlet
Create Parent folder
Do not create file at start
Accept file name from field?
File name field
Extension
csv
Include stepnr in filename?
Include partition nr in filename?
Include date in filename?
Include time in filename?
Specify Date time format
Date time format
Show filename(s)...
Add filenames to result
Help
OK
Cancel









Execution Results

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

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	General text file	0	0	29982	9995	0	1	0	0	Finished	0.9s	32 273	-
2	Filter rows 2	0	9994	9994	0	0	0	0	0	Finished	0.9s	10 677	-
3	Filter rows 3	0	7671	7671	0	0	0	0	0	Finished	0.9s	8 152	-
4	Sort rows 2	0	9994	5009	0	0	0	0	0	Finished	1.0s	9 974	-
5	Filter rows 3 2	0	6051	6051	0	0	0	0	0	Finished	0.9s	6 376	-
6	Filter rows	0	5009	296	0	0	0	0	0	Finished	1.0s	4 863	-
7	Filter rows 3 2 2	0	3203	3203	0	0	0	0	0	Finished	1.0s	3 357	-
8	Filter rows 3 2 2 2	0	2979	2473	0	0	0	0	0	Finished	1.0s	3 077	-
9	South file output	0	1620	1620	0	1621	0	0	0	Finished	0.9s	1 712	-
10	Sort rows	0	9994	1862	0	0	0	0	0	Finished	1.0s	9 703	-
11	XML output	0	296	296	0	296	0	0	0	Finished	1.1s	279	-
12	West Washington file output	0	0	0	0	0	0	0	0	Finished	1.0s	0	-
13	Central Excel output	0	2323	2323	0	2323	0	0	0	Finished	2.4s	974	-
14	West Arizona file output	0	224	224	0	225	0	0	0	Finished	1.0s	232	-