

PROJECT 9 TALEND JOIN

First we create a job called Project 9 in the Talend Appliaction.

The next step is to create 4 delimited files which are fact_sales, product, time, geography.

The below images shows how the 4 delimited files were configured,

Fact_sales

fact_sales(tFileInputDelimited_1)

Basic settings	Property Type	Built-In	
Advanced settings	Schema	Repository	DELIM:fact_sales_1 - metadata
Dynamic settings	"When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0."		
View	File name/Stream	"C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/FACT_SALES.csv"	
Documentation	Row Separator	"\n"	Field Separator " , "
<input type="checkbox"/> CSV options			
	Header	1	Footer 0 Limit <input type="text"/>
<input checked="" type="checkbox"/> Skip empty rows <input type="checkbox"/> Uncompress as zip file <input type="checkbox"/> Die on error			

Product

product(tFileInputDelimited_2)

Basic settings	Property Type	Built-In	
Advanced settings	Schema	Repository	DELIM:dim_product_ - metadata
Dynamic settings	"When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0."		
View	File name/Stream	"C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/DIM_PRODUCTS.csv"	
Documentation	Row Separator	"\n"	Field Separator " , "
<input type="checkbox"/> CSV options			
	Header	1	Footer 0 Limit <input type="text"/>
<input checked="" type="checkbox"/> Skip empty rows <input type="checkbox"/> Uncompress as zip file <input type="checkbox"/> Die on error			

Time

time(tFileInputDelimited_3)

Basic settings	Property Type	Built-In	
Advanced settings	Schema	Repository	DELIM:dim_time_ - metadata
Dynamic settings	"When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0."		
View	File name/Stream	"C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/DIM_TIME.csv"	
Documentation	Row Separator	"\n"	Field Separator " , "
<input type="checkbox"/> CSV options			
	Header	1	Footer 0 Limit <input type="text"/>
<input checked="" type="checkbox"/> Skip empty rows <input type="checkbox"/> Uncompress as zip file <input type="checkbox"/> Die on error			

Geography

geography(tFileInputDelimited_4)

Basic settings	Property Type	Built-In	
Advanced settings	Schema	Repository	DELIM:dim_geo_ - metadata
Dynamic settings	"When the input source is a stream or a zip file, footer and random shouldn't be bigger than 0."		
View	File name/Stream	"C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/DIM_GEOGRAPHY.csv"	
Documentation	Row Separator	"\n"	Field Separator " , "
<input type="checkbox"/> CSV options			
	Header	1	Footer 0 Limit <input type="text"/>
<input checked="" type="checkbox"/> Skip empty rows <input type="checkbox"/> Uncompress as zip file <input type="checkbox"/> Die on error			

PROJECT 9 TALEND JOIN

After creating the 4 delimited files we create a folder named project 9 in File delimited and then we create 4 delimited files for the fact sales, dim_geography, dim_time, dim_product a sample how to do that is shown in the below images,

Edit an existing Metadata File on repository
Update the properties

Name	fact_sales_1		
Purpose			
Description			
Author	user@talend.com		
Locker	user@talend.com		
Version	0.1	M	m
Status			
Path	project9		

File - Step 2 of 3

Edit an existing Metadata File on repository
Update the path of the file and the format settings

File Settings

Server	Localhost 127.0.0.1
File	C:/Users/mohit/Downloads/fact_sales-1.csv
Format	WINDOWS

File Viewer

```
product_id,geography_id,time_id,SalesAmt
1,1,1,673
1,1,2,149
1,1,3,824
1,1,4,96
1,2,1,677
1,2,2,343
1,2,3,451
1,2,4,994
1,3,1,28
1,3,2,418
1,3,3,793
```

PROJECT 9 TALEND JOIN

File - Step 3 of 3

Update an existing Metadata File on repository
Define the setting of the parse job



File Settings		Rows To Skip	
Encoding	US-ASCII	If any rows must be ignored, specify the following parameters	
Field Separator	Comr	Corresponding Character	","
Row Separator	Stand	Corresponding Character	"\n"
Escape Char Settings		Header	<input checked="" type="checkbox"/> 1
<input type="radio"/> CSV	<input checked="" type="radio"/> Delimited	Footer	<input type="checkbox"/>
Escape Char	Empty	<input type="checkbox"/> Skip empty row	
Text Enclosure	Empty	Limit Of Rows	
<input type="checkbox"/> Split row before field		If the number of lines must be limited, specify this number	
		Limit	<input type="checkbox"/>

Preview	Output																				
<input checked="" type="checkbox"/> Set heading row as column names Refresh Preview																					
<table border="1"><thead><tr><th>product_id</th><th>geography_id</th><th>time_id</th><th>SalesAmt</th></tr></thead><tbody><tr><td>1</td><td>1</td><td>1</td><td>673</td></tr><tr><td>1</td><td>1</td><td>2</td><td>149</td></tr><tr><td>1</td><td>1</td><td>3</td><td>824</td></tr><tr><td>1</td><td>1</td><td>4</td><td>96</td></tr></tbody></table>		product_id	geography_id	time_id	SalesAmt	1	1	1	673	1	1	2	149	1	1	3	824	1	1	4	96
product_id	geography_id	time_id	SalesAmt																		
1	1	1	673																		
1	1	2	149																		
1	1	3	824																		
1	1	4	96																		
Export as context Revert Context																					

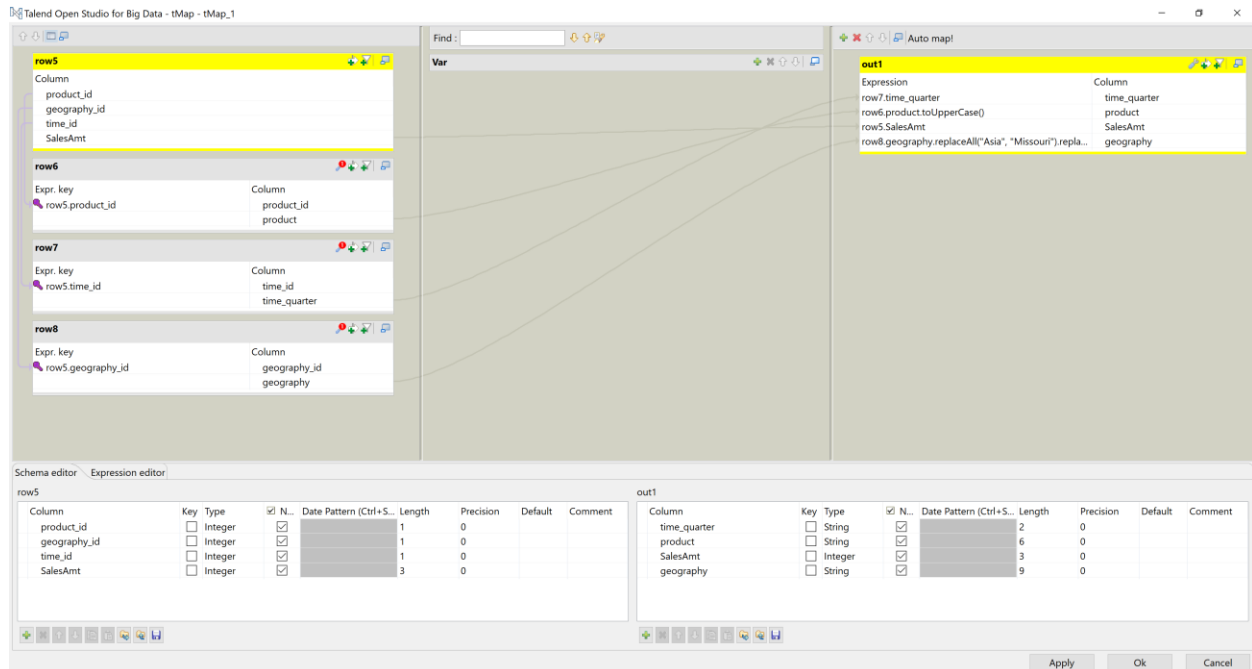
Project 9 Delimited folder after creating the delimited files for all the data we have it can seen in the image below,

- ▼ **File delimited**
 - ▼ project9
 - ▼ dim_geo_0.1
 - > metadata
 - ▼ dim_product_0.1
 - > metadata
 - ▼ dim_time_0.1
 - > metadata
 - > fact_sales_1 0.1

PROJECT 9 TALEND JOIN

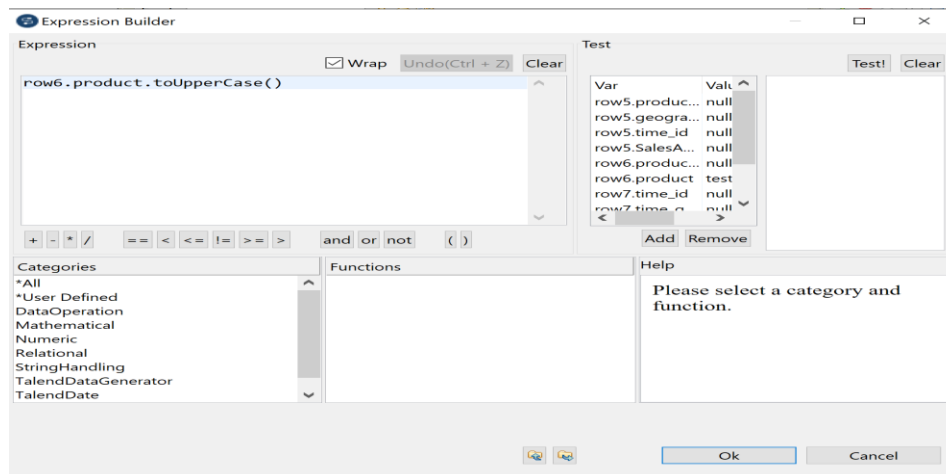
Configure the components tFileInputDelimited for the files "fact_sales," "dim_products," "dim_geography," and "dim_time." Connect them to a tMap component, followed by a tLogRow and finally to a tOutputBuffer.

The below image indicates the tmap components configuration,



CAPITALIZING PRODUCT NAMES:

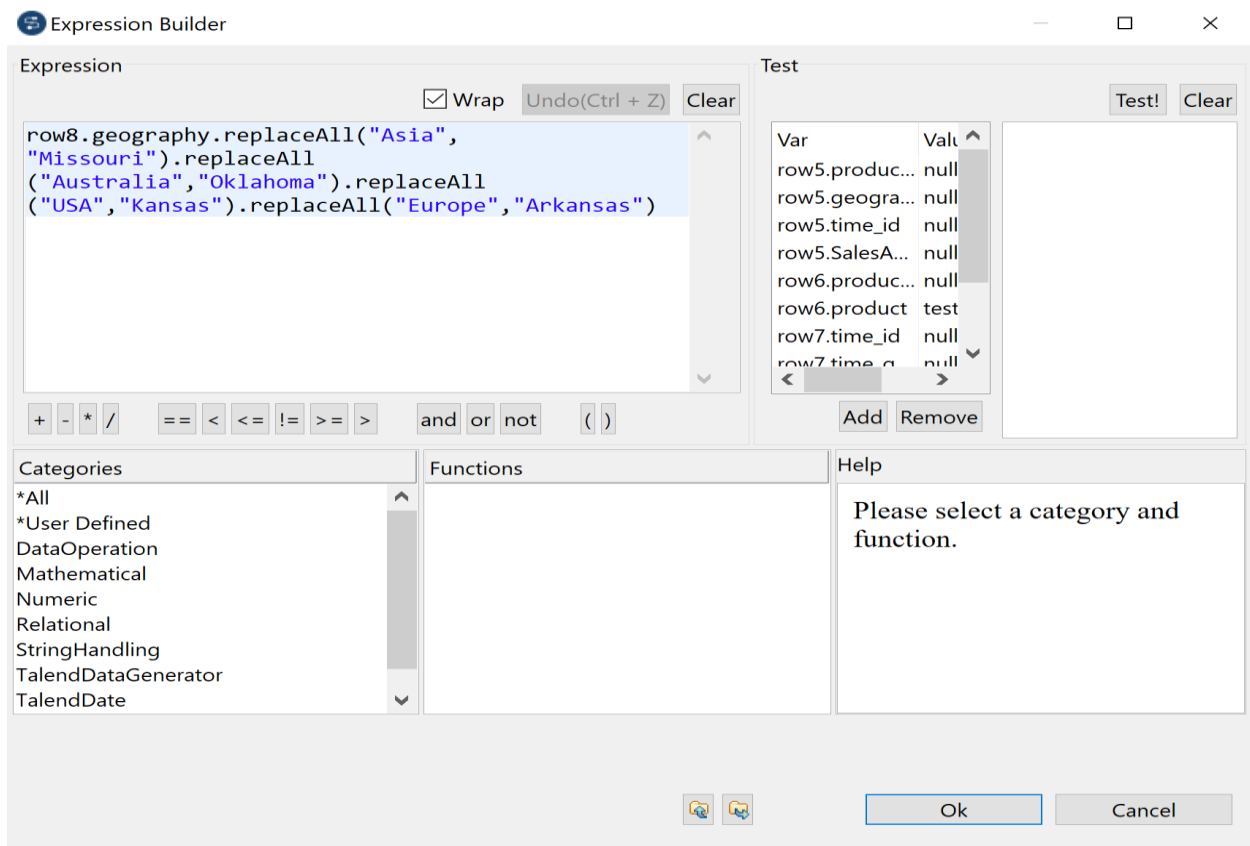
In order to capitalize product names in the out1 table the following expression is provided which converts any case letter to UPPERCASE as shown in the image below,



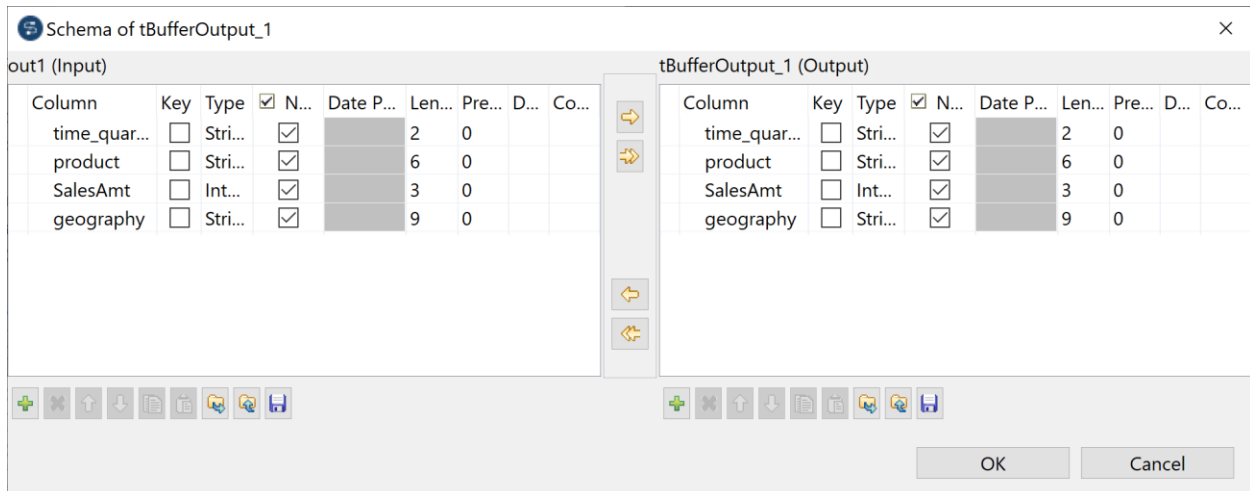
PROJECT 9 TALEND JOIN

Geographic locations replacement:

We replace the current geographic locations with the location given as part of task the expression can be seen in the image below,

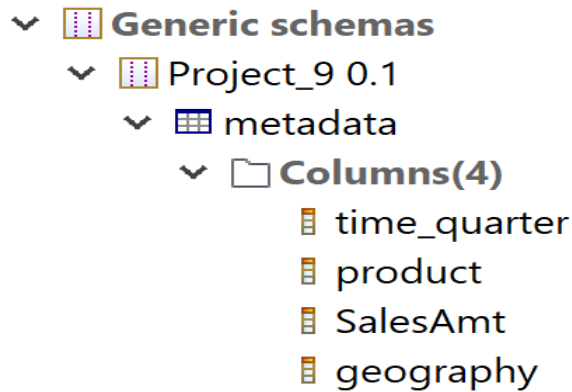


Begin by syncing the schema from the tMap component in the tBufferOutput component by selecting "Sync columns." Then, proceed to edit the schema and save it in the "tBufferOutput" schema. After saving, the schema will be stored in the "Generic Schemas" destination.



PROJECT 9 TALEND JOIN

Next, create tBufferInput components to read the data from the tBufferOutputComponent. Connect a tFilterRow component, a tLogRow component, and a tFinalOutputExcel component. The image of generic schema we have created can be seen below,



After configuring the schema for all four tBufferInput components, proceed to configure the tFileRow components. Once all four tFilterRow components are configured, set up the corresponding LogRow components connected to tFilterRows to read the data. The images for the configuration of tFilterRows are as follows.

The screenshot shows the configuration window for the tFilterRow_1 component. The 'Basic settings' tab is selected. The 'Schema' is set to 'Built-In'. The 'Logical operator used to combine conditions' is set to 'And'. The 'Conditions' table is configured as follows:

InputColumn	Function	Operator	Value
product	Empty	Equals	"LAPTOP"

There are icons for adding, removing, and reordering conditions. The 'Use advanced mode' checkbox is unchecked.

The screenshot shows the configuration window for the tFilterRow_2 component. The 'Basic settings' tab is selected. The 'Schema' is set to 'Built-In'. The 'Logical operator used to combine conditions' is set to 'And'. The 'Conditions' table is configured as follows:

InputColumn	Function	Operator	Value
product	Empty	Equals	"MOBILE"

There are icons for adding, removing, and reordering conditions. The 'Use advanced mode' checkbox is unchecked.

PROJECT 9 TALEND JOIN

tFilterRow_3

Basic settings

Schema: Built-In Edit schema Sync columns

Logical operator used to combine conditions: And *

Conditions	InputColumn	Function	Operator	Value
	product	Empty	Equals	"TV"

☐ Use advanced mode

tFilterRow_4

Basic settings

Schema: Built-In Edit schema Sync columns

Logical operator used to combine conditions: And *

Conditions	InputColumn	Function	Operator	Value
	product	Empty	Equals	"TABLET"

☐ Use advanced mode

Once the logRow files are configured we then configure tOutputFileExcel component. All data is directed to a single Excel file named "sales_data.xls" with independent sheets named "LAPTOP," "TABLET," "TV," and "MOBILE" configured as shown below,

tFileOutputExcel_2

Basic settings

Property Type: Built-In

☒ Write excel2007 file format (xlsx)

☐ Use Output Stream

File Name: "C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/sales_data.xls"

Sheet name: "Laptop"

☒ Include header

☒ Append existing file ☒ Append existing sheet

☐ Is absolute Y pos.

Font: Default

☐ Define all columns auto size

Define column auto size	Column	Auto size
	time_quarter	<input type="checkbox"/>
	product	<input type="checkbox"/>
	SalesAmt	<input type="checkbox"/>
	geography	<input type="checkbox"/>

tFileOutputExcel_3

Basic settings

Property Type: Built-In

☒ Write excel2007 file format (xlsx)

☐ Use Output Stream

File Name: "C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/sales_data.xls"

Sheet name: "Mobile"

☒ Include header

☒ Append existing file ☒ Append existing sheet

☐ Is absolute Y pos.

Font: Default

☐ Define all columns auto size

Define column auto size	Column	Auto size
	time_quarter	<input type="checkbox"/>
	product	<input type="checkbox"/>
	SalesAmt	<input type="checkbox"/>
	geography	<input type="checkbox"/>

PROJECT 9 TALEND JOIN

The image shows two screenshots of the Talend Studio configuration interface for the **tFileOutputExcel** component. The top screenshot is for **tFileOutputExcel_4** and the bottom is for **tFileOutputExcel_5**. Both are configured with the following settings:

- Property Type:** Built-In
- Basic settings:**
 - ☒ Write excel2007 file format (xlsx)
 - ☐ Use Output Stream
 - File Name:** "C:/Users/mohit/OneDrive/Desktop/TOS_BD-20211109_1610-V8.0.1/workspace/sales_data.xls"
 - Sheet name:** "tv"
 - ☒ Include header
 - ☒ Append existing file ☒ Append existing sheet
 - ☐ Is absolute Y pos.
 - Font:** Default
 - ☐ Define all columns auto size
 - Define column auto size:** A table with columns for data types and checkboxes for auto-sizing.

Column	Auto size
time_quarter	<input type="checkbox"/>
product	<input type="checkbox"/>
SalesAmt	<input type="checkbox"/>
geography	<input type="checkbox"/>

The bottom screenshot for **tFileOutputExcel_5** includes an additional setting:

- ☐ Protect file

Finally we run the job Once the job runs we can see the following consoles in the logRow components,

Job project_9

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution



Run



Kill



Clear

Starting job project_9 at 20:07 04/11/2023.

[statistics] connecting to socket on port 3705

[statistics] connected

.-----+-----.

| tLogRow_3 |

|=-----+-----=|

|product_id|product|

|=-----+-----=|

|1 |Laptop|

|2 |Mobile|

|3 |TV |

|4 |Tablet|

|'-----+-----'|

.-----+-----.

| tLogRow_4 |

|=-----+-----=|

|time_id|time_quarter|

|=-----+-----=|

|1 |Q1 |

|2 |Q2 |

|3 |Q3 |

|4 |Q4 |

|'-----+-----'|

.-----+-----.

| tLogRow_5 |

|=-----+-----=|

|geography_id|geography|

<

PROJECT 9 TALEND JOIN

The snip of tLogRow_1 was not taken completely because it has lot of data,

tLogRow_5	
geography_id	geography
1	USA
2	Asia
3	Australia
4	Europe

tLogRow_1			
product_id	geography_id	time_id	SalesAmt
1	1	1	797
1	1	2	54
1	1	3	886
1	1	4	693
1	2	1	673
1	2	2	74
1	2	3	884
1	2	4	268
1	3	1	28
1	3	2	797
1	3	3	139
1	3	4	733
1	4	1	258

<

PROJECT 9 TALEND JOIN

tLogRow_2			
time_quarter	product	SalesAmt	geography
Q1	LAPTOP	797	Kansas
Q2	LAPTOP	54	Kansas
Q3	LAPTOP	886	Kansas
Q4	LAPTOP	693	Kansas
Q1	LAPTOP	673	Missouri
Q2	LAPTOP	74	Missouri
Q3	LAPTOP	884	Missouri
Q4	LAPTOP	268	Missouri
Q1	LAPTOP	28	Oklahoma
Q2	LAPTOP	797	Oklahoma
Q3	LAPTOP	139	Oklahoma
Q4	LAPTOP	733	Oklahoma
Q1	LAPTOP	258	Arkansas
Q2	LAPTOP	561	Arkansas
Q3	LAPTOP	117	Arkansas
Q4	LAPTOP	564	Arkansas

tLogRow_6			
time_quarter	product	SalesAmt	geography
Q1	MOBILE	231	Kansas
Q2	MOBILE	760	Kansas
Q3	MOBILE	677	Kansas
Q4	MOBILE	707	Kansas
Q1	MOBILE	329	Missouri
Q2	MOBILE	366	Missouri
Q3	MOBILE	26	Missouri
Q4	MOBILE	759	Missouri
Q1	MOBILE	443	Oklahoma
Q2	MOBILE	117	Oklahoma
Q3	MOBILE	257	Oklahoma
Q4	MOBILE	960	Oklahoma
Q1	MOBILE	289	Arkansas
Q2	MOBILE	811	Arkansas
Q3	MOBILE	176	Arkansas
Q4	MOBILE	728	Arkansas



PROJECT 9 TALEND JOIN

tLogRow_7			
time_quarter	product	SalesAmt	geography
Q1	TV	867	Kansas
Q2	TV	727	Kansas
Q3	TV	958	Kansas
Q4	TV	72	Kansas
Q1	TV	576	Missouri
Q2	TV	830	Missouri
Q3	TV	316	Missouri
Q4	TV	812	Missouri
Q1	TV	124	Oklahoma
Q2	TV	147	Oklahoma
Q3	TV	840	Oklahoma
Q4	TV	91	Oklahoma
Q1	TV	502	Arkansas
Q2	TV	675	Arkansas
Q3	TV	942	Arkansas
Q4	TV	822	Arkansas

tLogRow_8			
time_quarter	product	SalesAmt	geography
Q1	TABLET	394	Kansas
Q2	TABLET	796	Kansas
Q3	TABLET	474	Kansas
Q4	TABLET	190	Kansas
Q1	TABLET	629	Missouri
Q2	TABLET	419	Missouri
Q3	TABLET	437	Missouri
Q4	TABLET	322	Missouri
Q1	TABLET	636	Oklahoma
Q2	TABLET	650	Oklahoma
Q3	TABLET	717	Oklahoma
Q4	TABLET	280	Oklahoma
Q1	TABLET	995	Arkansas
Q2	TABLET	796	Arkansas
Q3	TABLET	459	Arkansas
Q4	TABLET	758	Arkansas

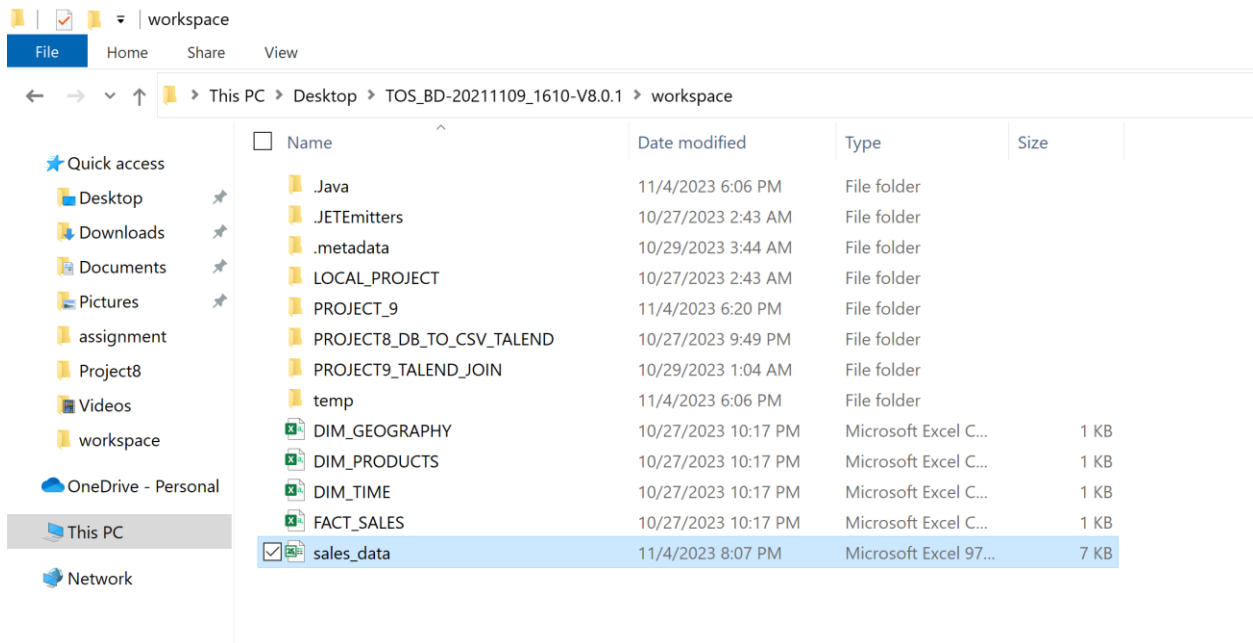
[statistics] disconnected

Job project_9 ended at 20:07 04/11/2023. [E

PROJECT 9 TALEND JOIN

The above picture indicates that our job was done successfully.

As we can see that the generated output file sales_data.xls will be created if the job is run,



The below images show the data generated in 4 sheets when the job is run,

The screenshot shows the 'time_quarter' sheet in Microsoft Excel. The data is organized into columns: 'time_quarter', 'product', 'SalesAmt', and 'geography'. The data is listed in rows 1 through 17.

time_quarter	product	SalesAmt	geography
Q1	LAPTOP	797	Kansas
Q2	LAPTOP	54	Kansas
Q3	LAPTOP	886	Kansas
Q4	LAPTOP	693	Kansas
Q1	LAPTOP	673	Missouri
Q2	LAPTOP	74	Missouri
Q3	LAPTOP	884	Missouri
Q4	LAPTOP	268	Missouri
Q1	LAPTOP	28	Oklahoma
Q2	LAPTOP	797	Oklahoma
Q3	LAPTOP	139	Oklahoma
Q4	LAPTOP	733	Oklahoma
Q1	LAPTOP	258	Arkansas
Q2	LAPTOP	561	Arkansas
Q3	LAPTOP	117	Arkansas
Q4	LAPTOP	564	Arkansas

PROJECT 9 TALEND JOIN

sales_data - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Clipboard Font Alignment Number Styles

time_quarter

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	time_quarter	product	SalesAmt	geography											
2	Q1	MOBILE	231	Kansas											
3	Q2	MOBILE	760	Kansas											
4	Q3	MOBILE	677	Kansas											
5	Q4	MOBILE	707	Kansas											
6	Q1	MOBILE	329	Missouri											
7	Q2	MOBILE	366	Missouri											
8	Q3	MOBILE	26	Missouri											
9	Q4	MOBILE	759	Missouri											
10	Q1	MOBILE	443	Oklahoma											
11	Q2	MOBILE	117	Oklahoma											
12	Q3	MOBILE	257	Oklahoma											
13	Q4	MOBILE	960	Oklahoma											
14	Q1	MOBILE	289	Arkansas											
15	Q2	MOBILE	811	Arkansas											
16	Q3	MOBILE	176	Arkansas											
17	Q4	MOBILE	728	Arkansas											
18															
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															

Laptop Mobile tv tablet

PROJECT 9 TALEND JOIN

sales_data - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Cut Copy Format Painter Clipboard Font Alignment Number Conditional Formatting

Calibri 11 A A B I U Merge & Center General \$ % .00 .00

A1 time_quarter

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	time_quarter	product	SalesAmt	geography									
2	Q1	TV	867	Kansas									
3	Q2	TV	727	Kansas									
4	Q3	TV	958	Kansas									
5	Q4	TV	72	Kansas									
6	Q1	TV	576	Missouri									
7	Q2	TV	830	Missouri									
8	Q3	TV	316	Missouri									
9	Q4	TV	812	Missouri									
10	Q1	TV	124	Oklahoma									
11	Q2	TV	147	Oklahoma									
12	Q3	TV	840	Oklahoma									
13	Q4	TV	91	Oklahoma									
14	Q1	TV	502	Arkansas									
15	Q2	TV	675	Arkansas									
16	Q3	TV	942	Arkansas									
17	Q4	TV	822	Arkansas									
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													

Laptop Mobile tv tablet

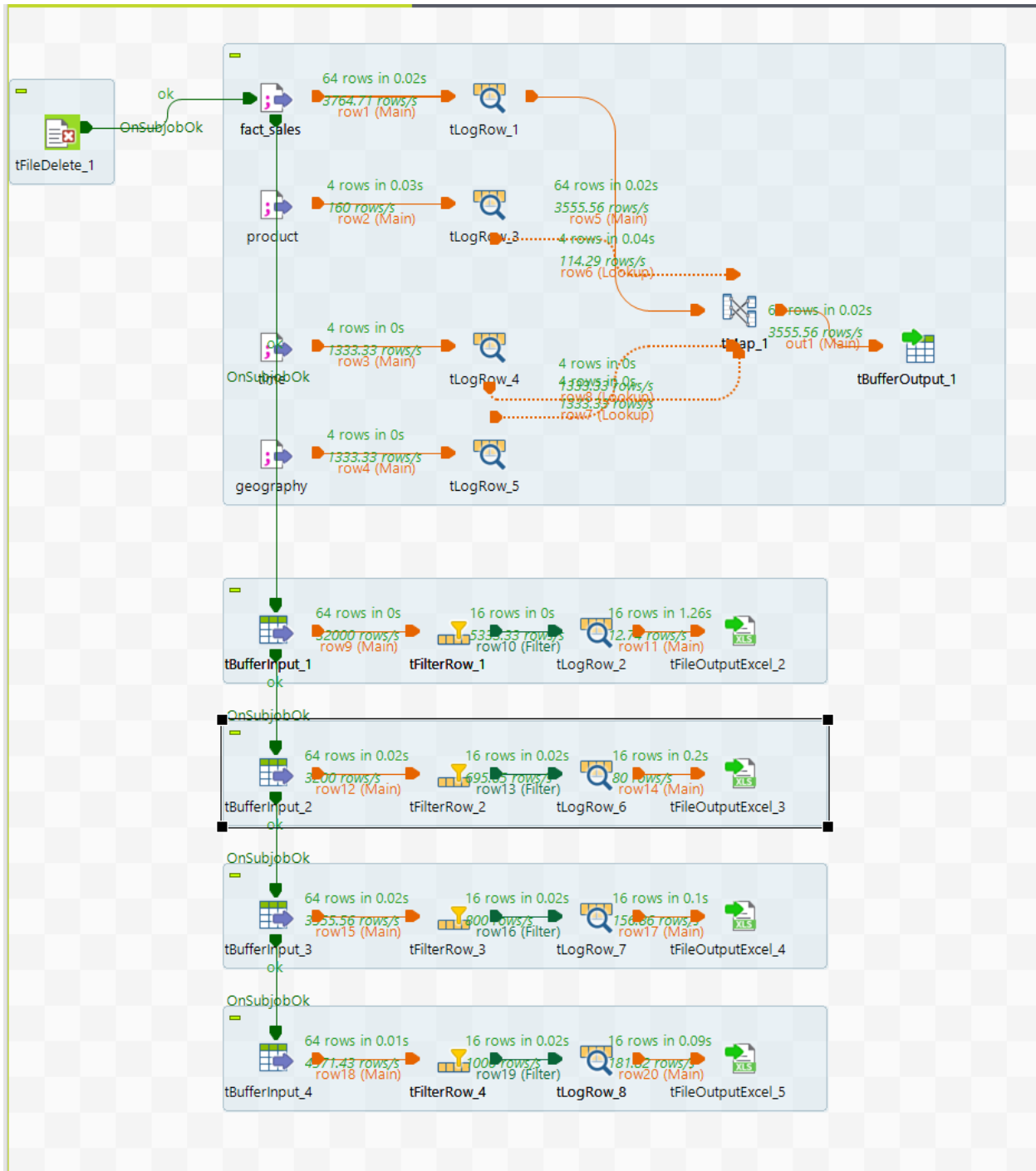
PROJECT 9 TALEND JOIN

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	time_quarter	product	SalesAmt	geography										
2	Q1	TABLET	394	Kansas										
3	Q2	TABLET	796	Kansas										
4	Q3	TABLET	474	Kansas										
5	Q4	TABLET	190	Kansas										
6	Q1	TABLET	629	Missouri										
7	Q2	TABLET	419	Missouri										
8	Q3	TABLET	437	Missouri										
9	Q4	TABLET	322	Missouri										
10	Q1	TABLET	636	Oklahoma										
11	Q2	TABLET	650	Oklahoma										
12	Q3	TABLET	717	Oklahoma										
13	Q4	TABLET	280	Oklahoma										
14	Q1	TABLET	995	Arkansas										
15	Q2	TABLET	796	Arkansas										
16	Q3	TABLET	459	Arkansas										
17	Q4	TABLET	758	Arkansas										
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														
29														
30														

The below image shows the picture of entire job that was created,

PROJECT 9 TALEND JOIN



Note

We additionally create `tFileDelete` component to clear the buffer data before running the job.

So when the task is run for the first time this `tFileDelete` component has to be deleted or the task will throw an error if this component is not removed.