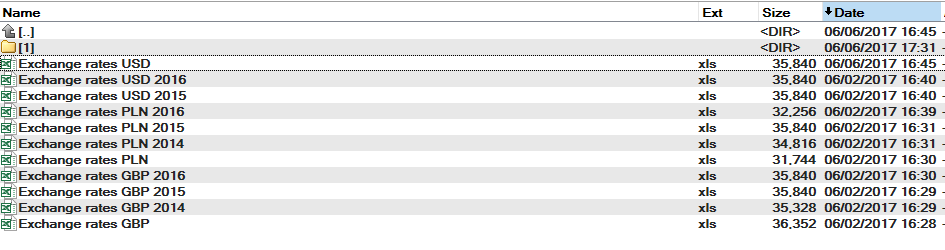
**Module2\_3** [create\_user/schema\_script](create_user.sql) [creat\_table\_script](create_table_and_constraint.sql)

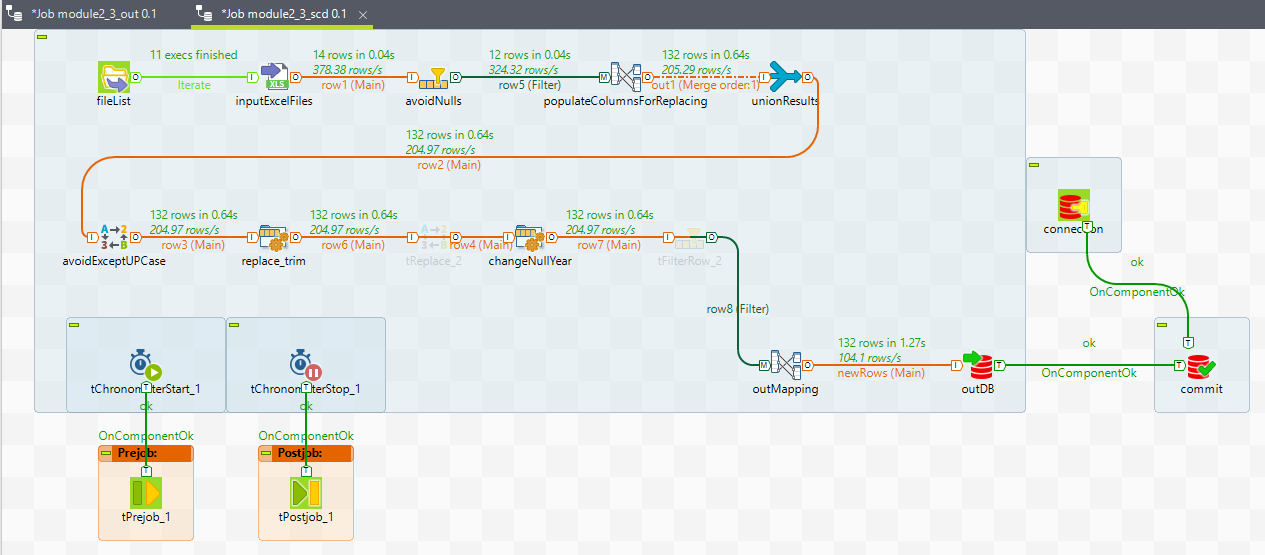
This input data set has some features.



On the picture above you can see that those files which don’t have Year in a title are current (2017) year. There are two ways how to get Year and Currency – from the title of file and from the file contain.

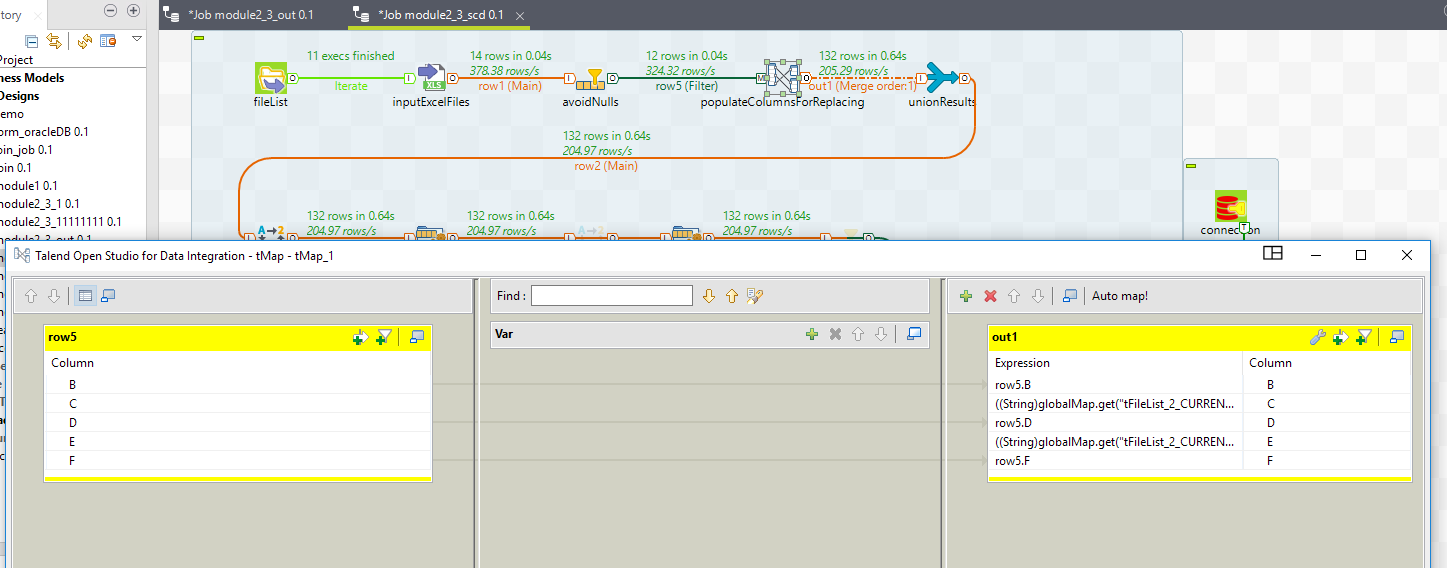
let’s explorer both of them.

[**Method 1**](dzianis_dziadkou_Module2_method1.zip)

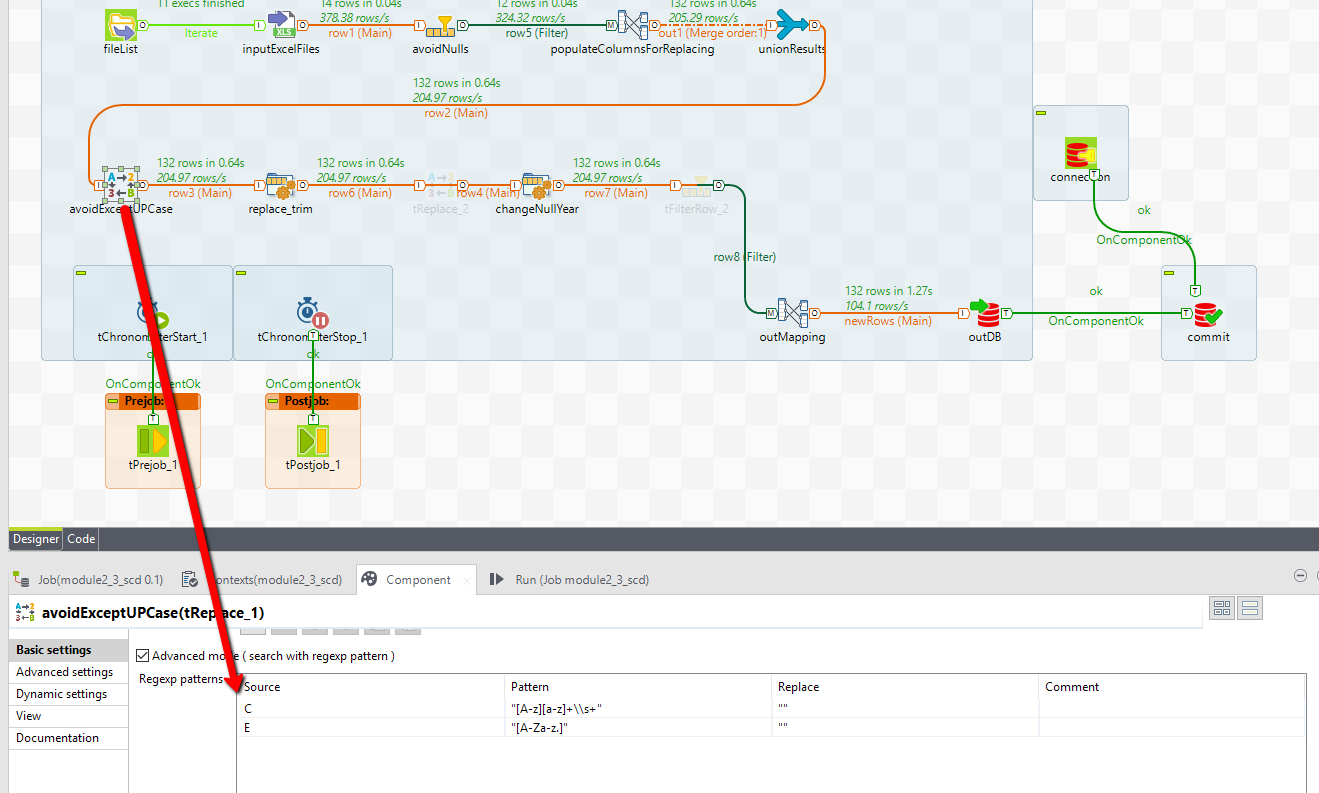


This way Year and Currency values will be come from the title.

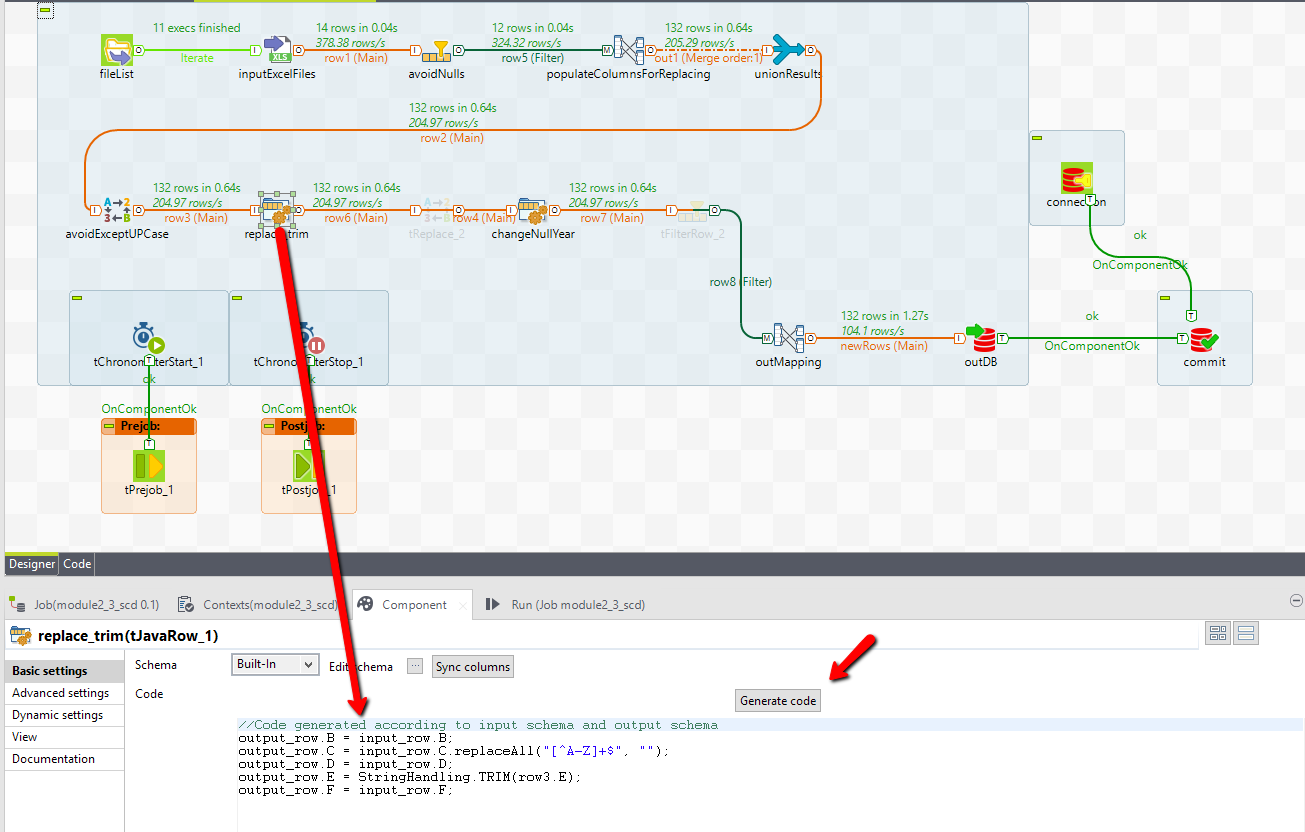
Use FileName to get both columns thus copying them (out1.C, out1.E).



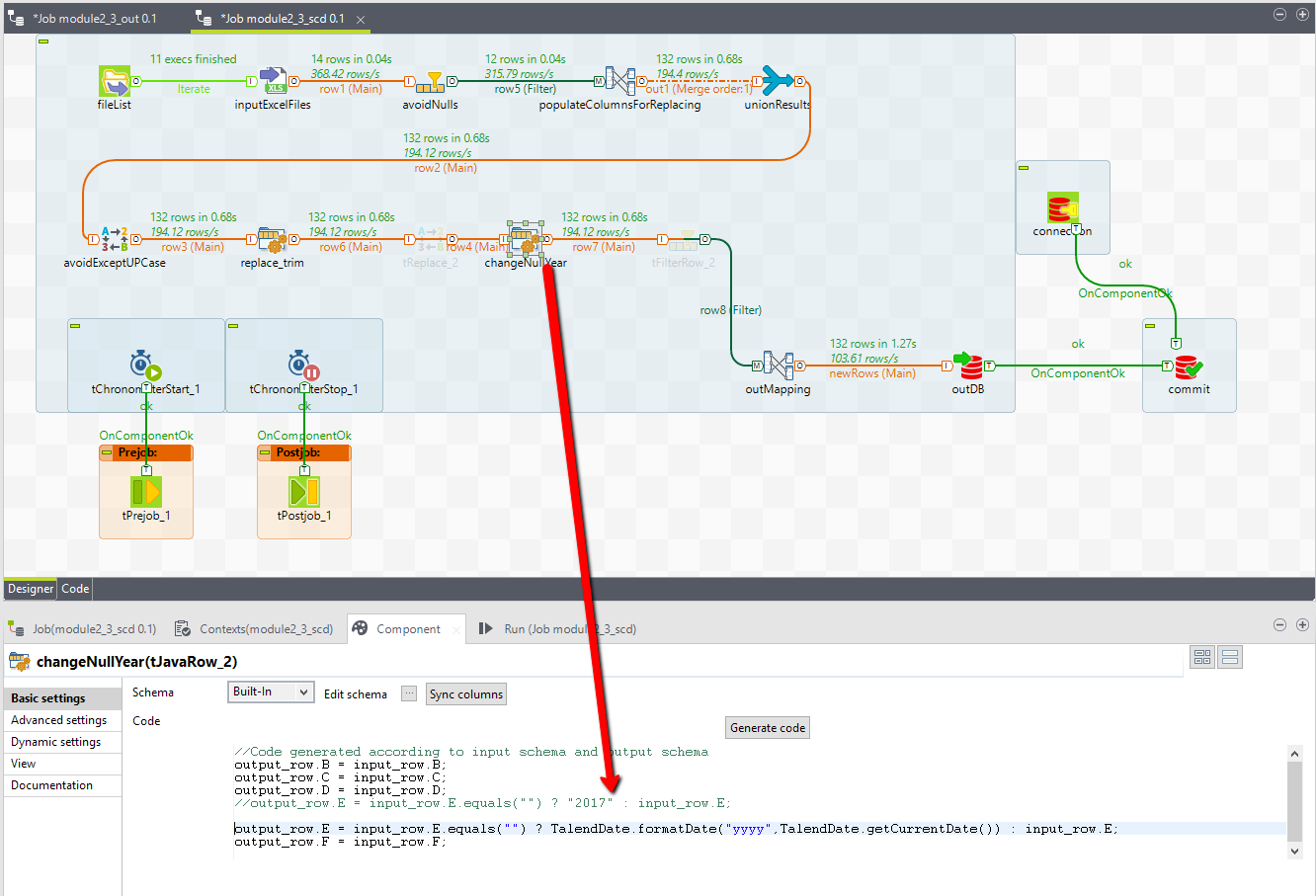
Next do replace statement. In the out1.C column would store Currency, in the out1.E – Year.



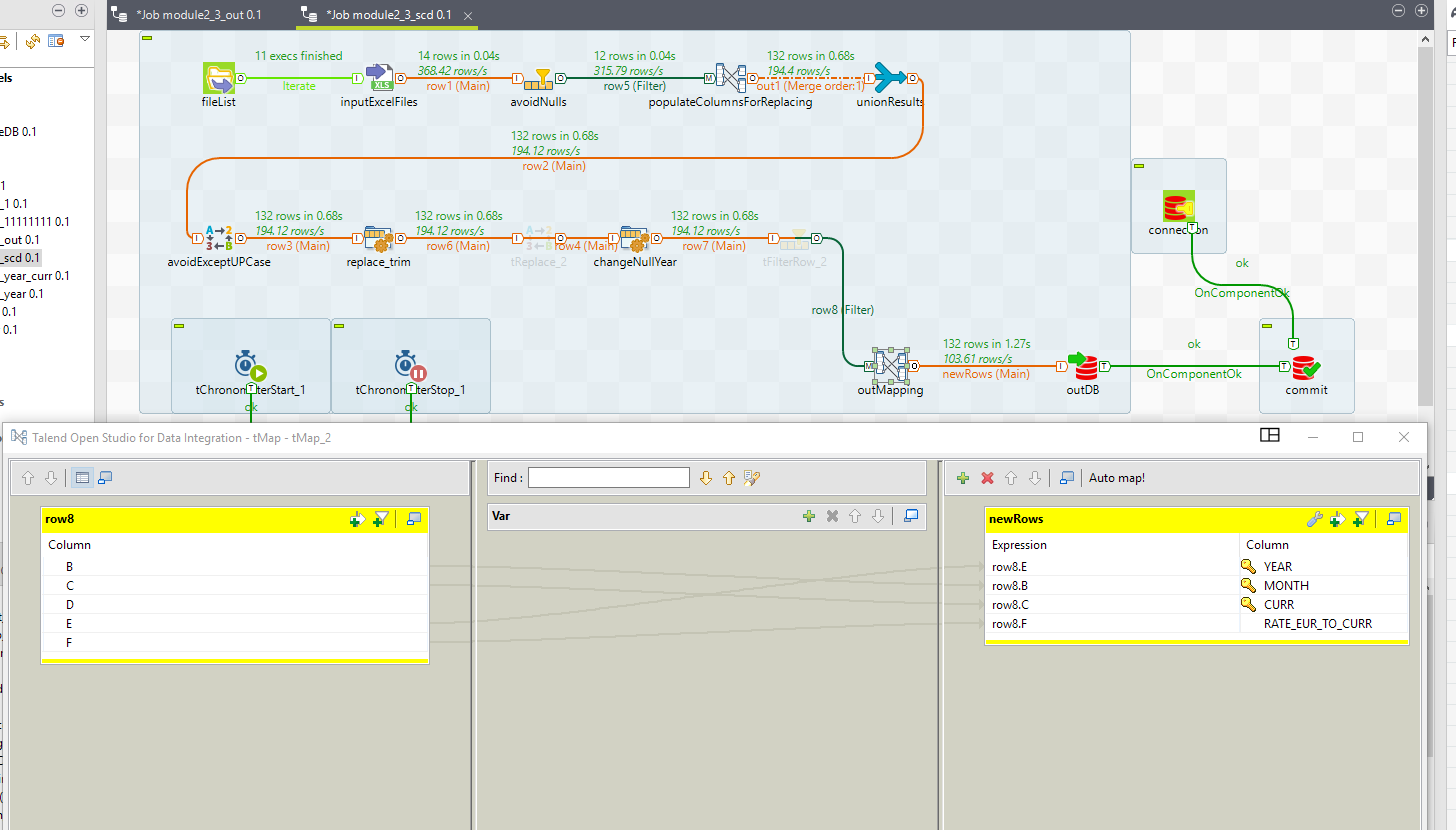
Then use tJavaRow statement. On the picture below you can see that statement for the out.C column which is the same as if the tReplace statement was used for. On the other hand for the out.E column use trim function to delete gaps.



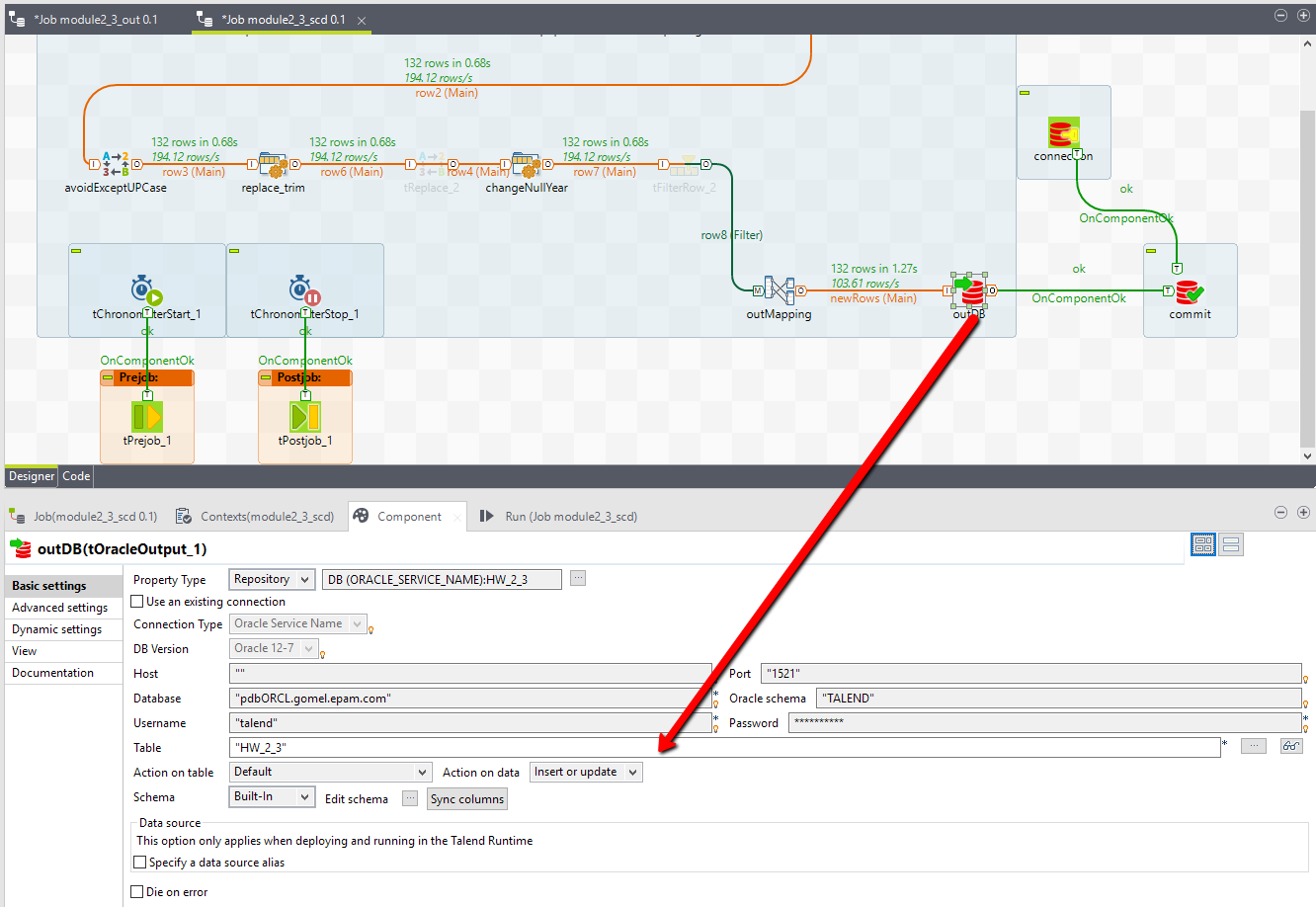
Such as there is no current year value in the title file (2017) it’s possible to change it by hands. Just extract the Year value from the current date. Worth mentioning it’s possible that not only files related to the current year would not have Year value in the title. If so, this methods is not right.



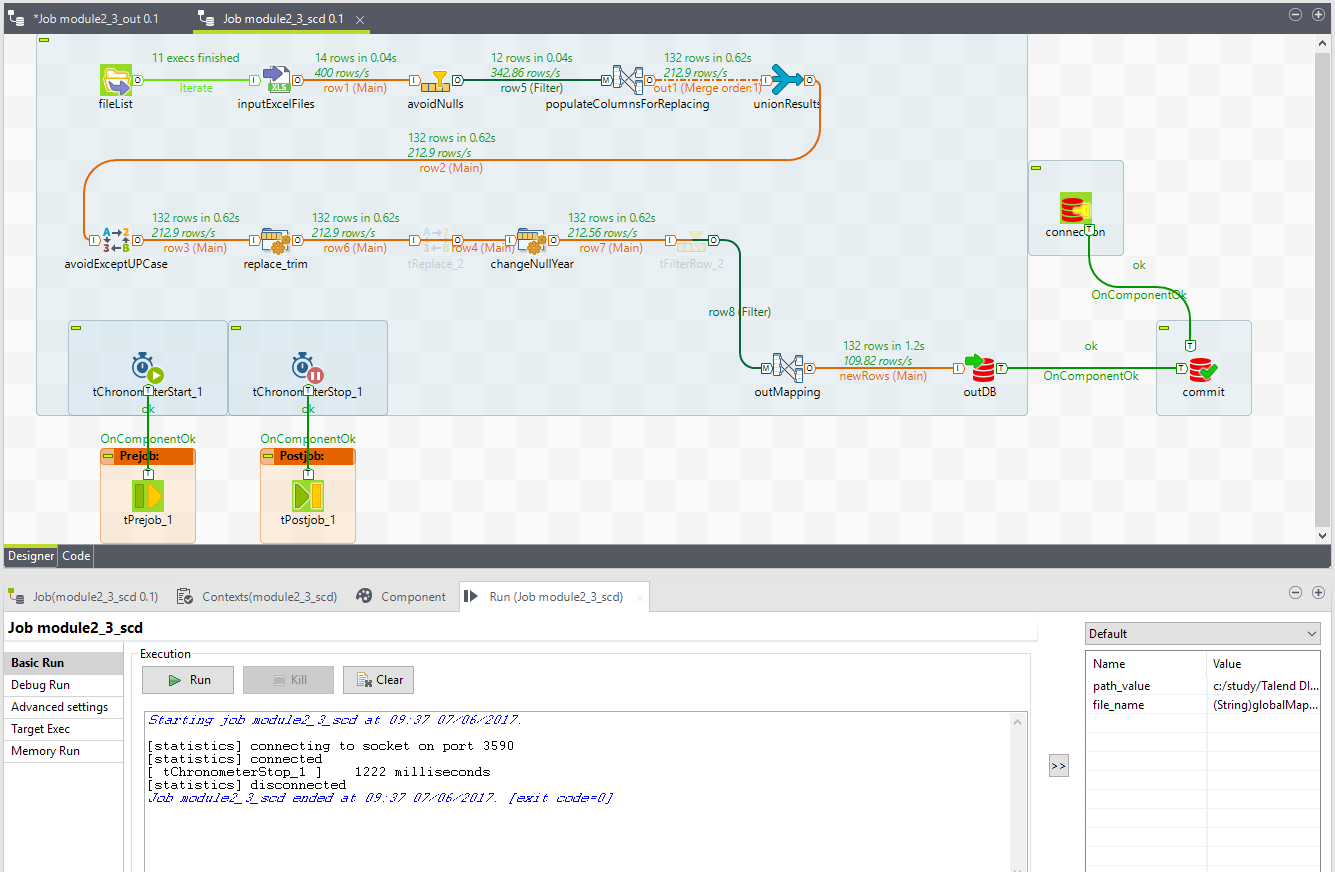
Regarding the requirement for this task new data should be **updated**. To do it’s necessary to specify Key columns. Also to order columns regarding the task requirement and rename them.



Next step is to write data set to the DB. Here it’s Oracle DB. Such as new data coming to the table should be inserted use **Insert and Update** action. For strings which have null values in the Balance column (for Jun-Dec 2017, because there are not any data till these months will come). let’s imagine this data set is using for building the report with charts monthly. In case when strings with null values are absent in DB it will require to make additional steps to retrieve charts from Jan to Dec. Finally, these null values should be updated when the month data would be received.



When data are come to the DB do commit and close the connection. This job has average running time approximately 1.2 sec.



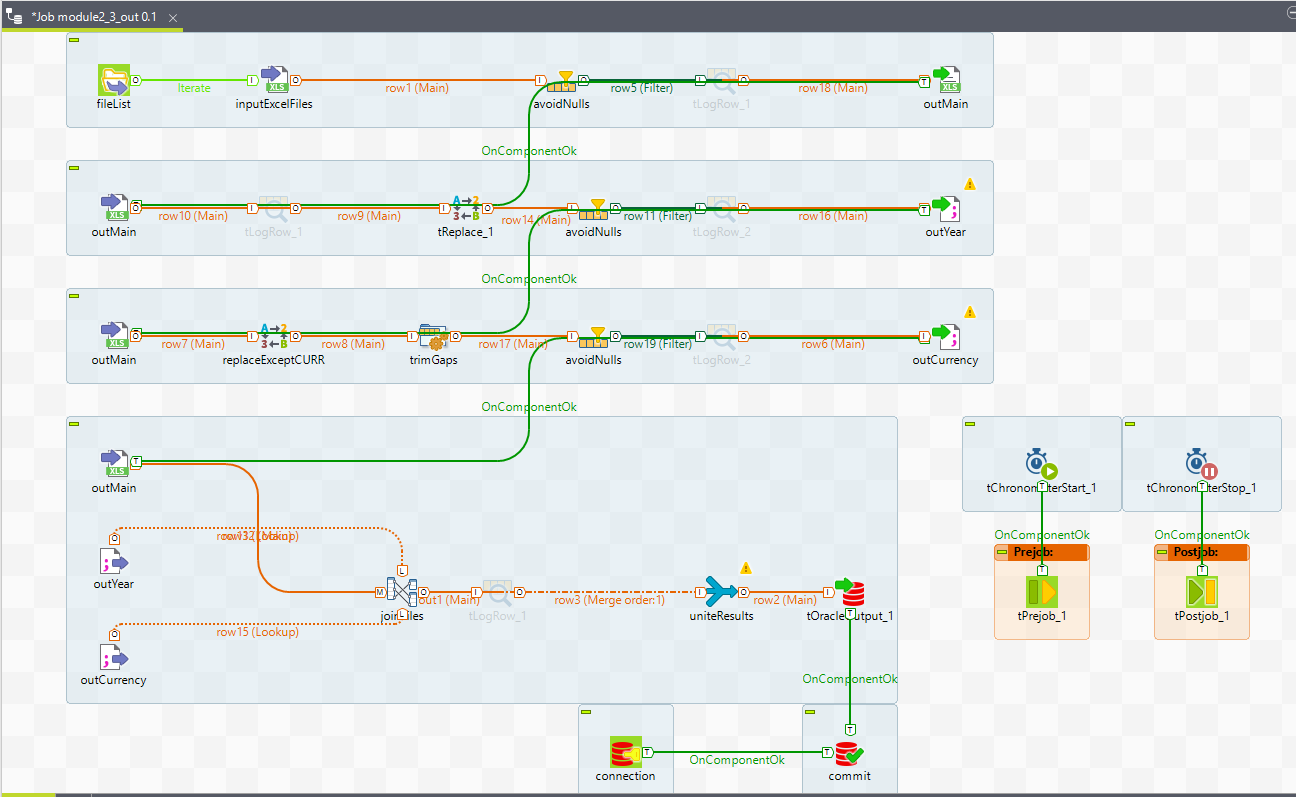
**Method 2** is below.

[**Method 2**](dzianis_dziadkou_Module2_method2.zip)

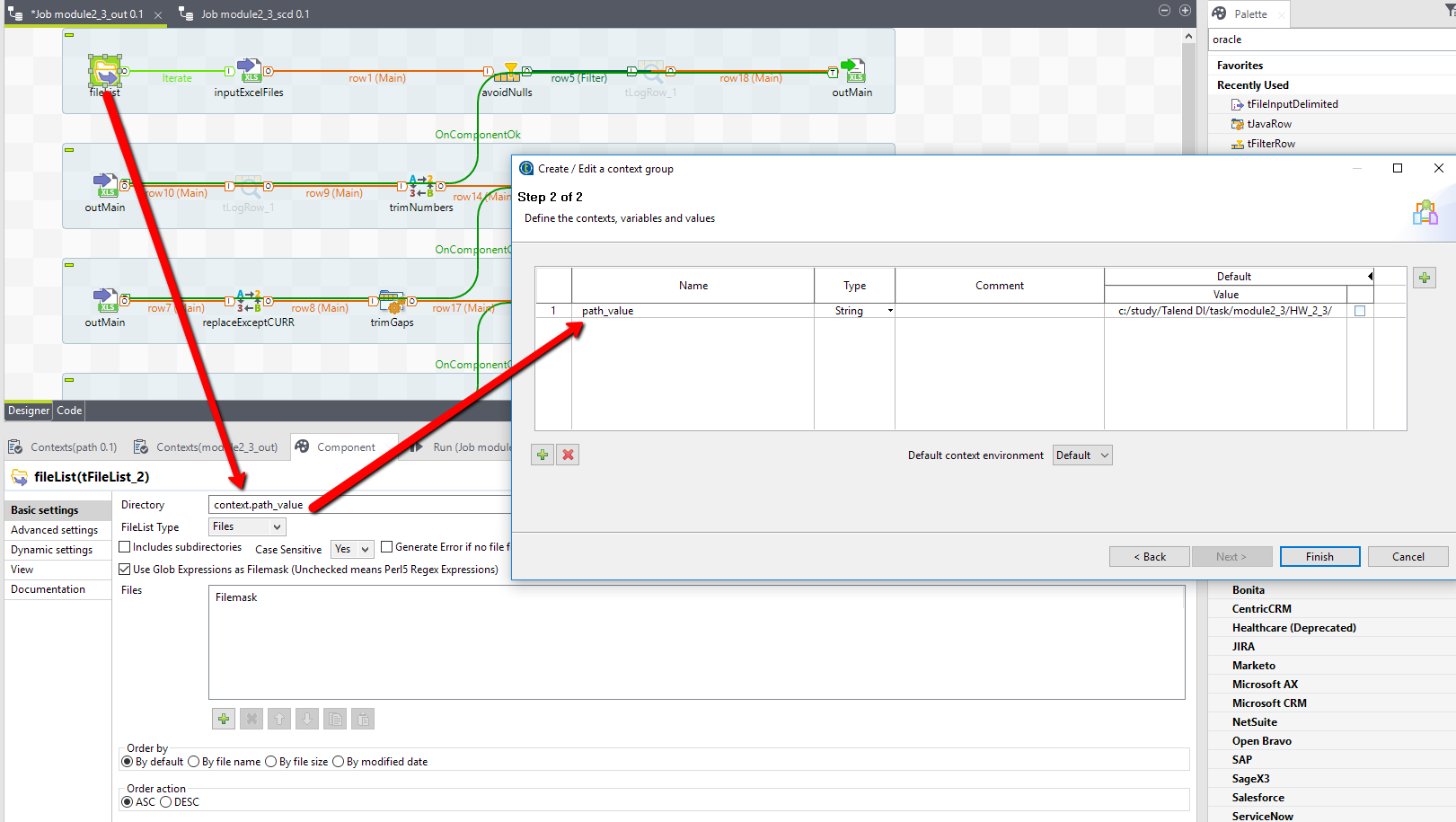
This is another way on how to solve this task. Even so, the Method 1 has right result we can’t be make sure that files with current year always would be come without Year value in the title.

This method is based on the fact that file name could be changed as you wish in spite of the file content

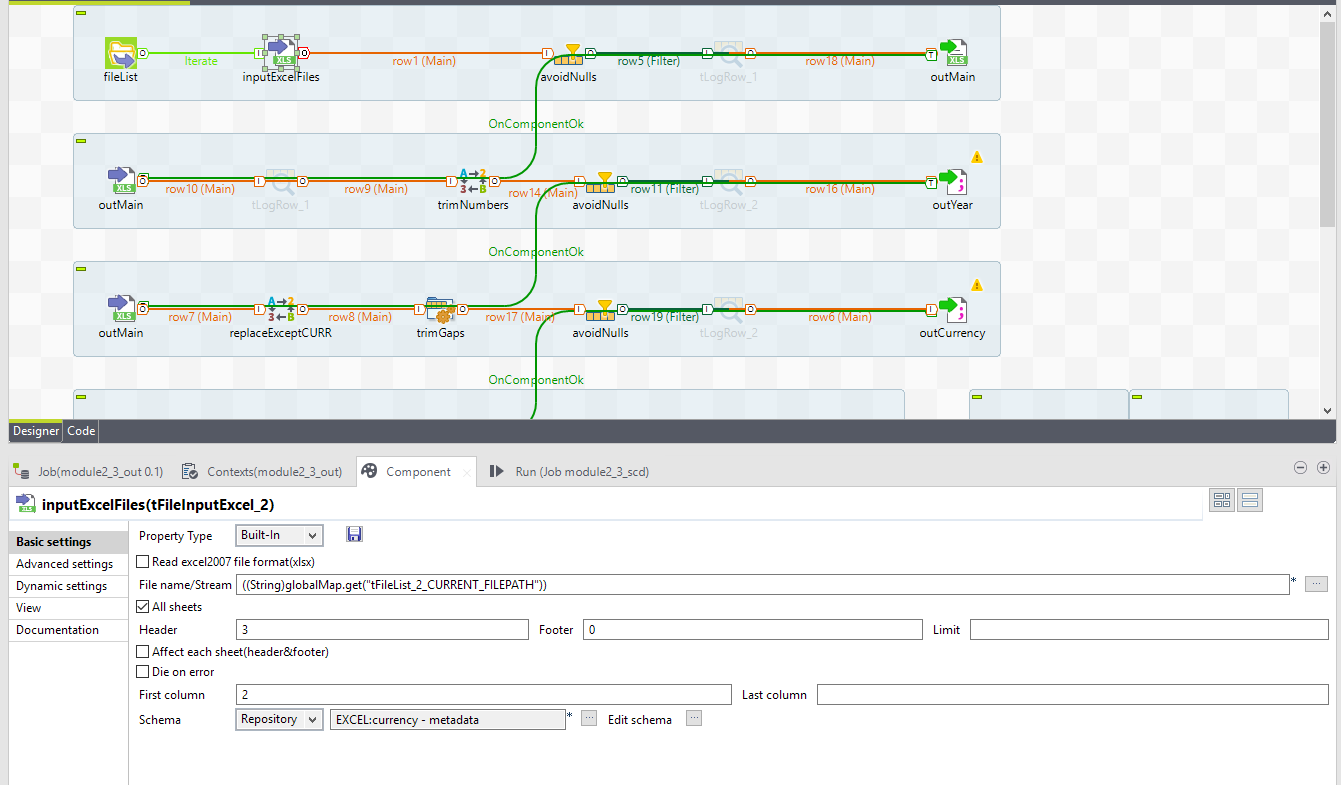
is always relevant. Current values of Year and Currency are stored in the files content.



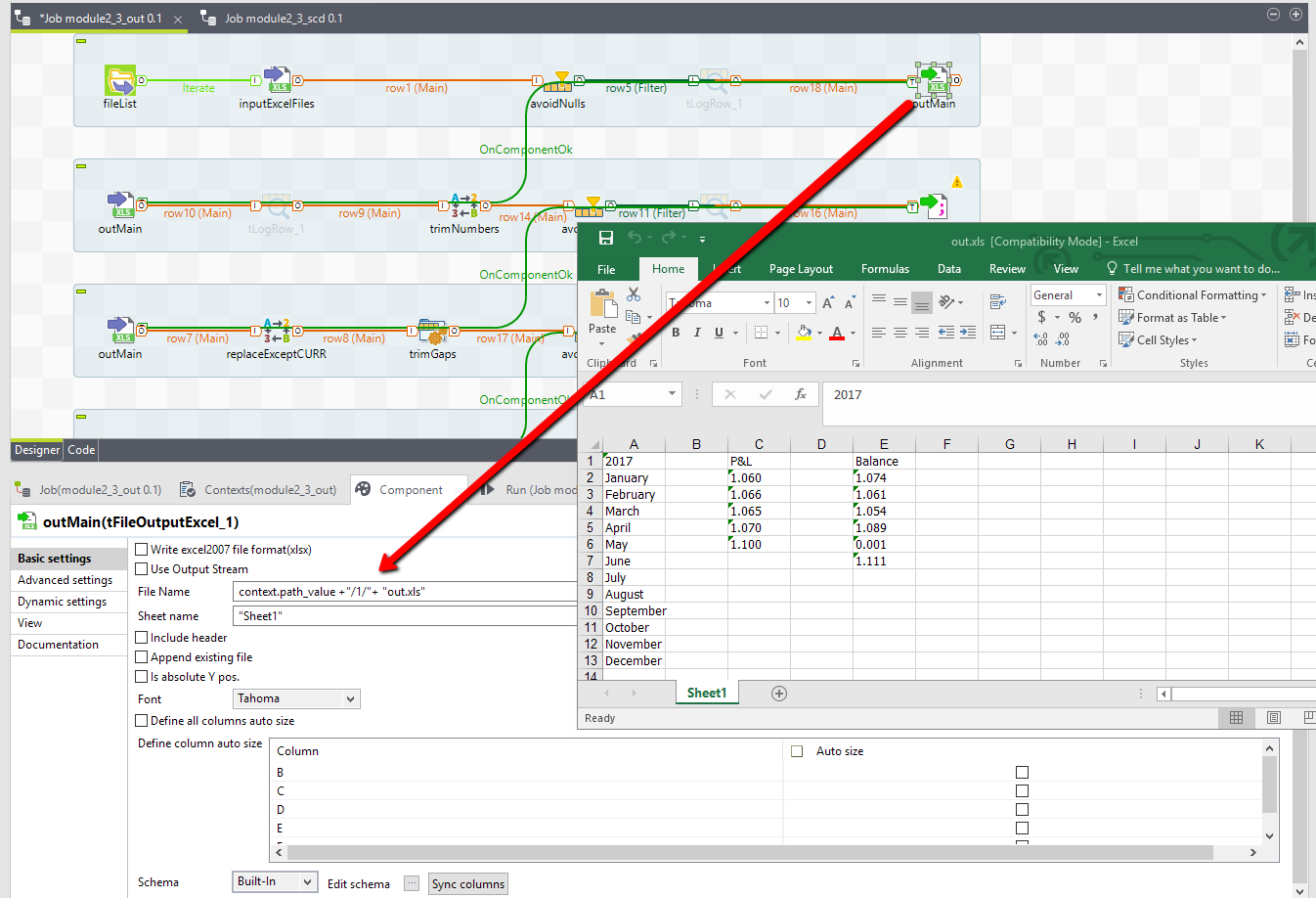
We have more than one file as an input file, thus use tFileList statement to read all of them. To avoid using path every time use Context.



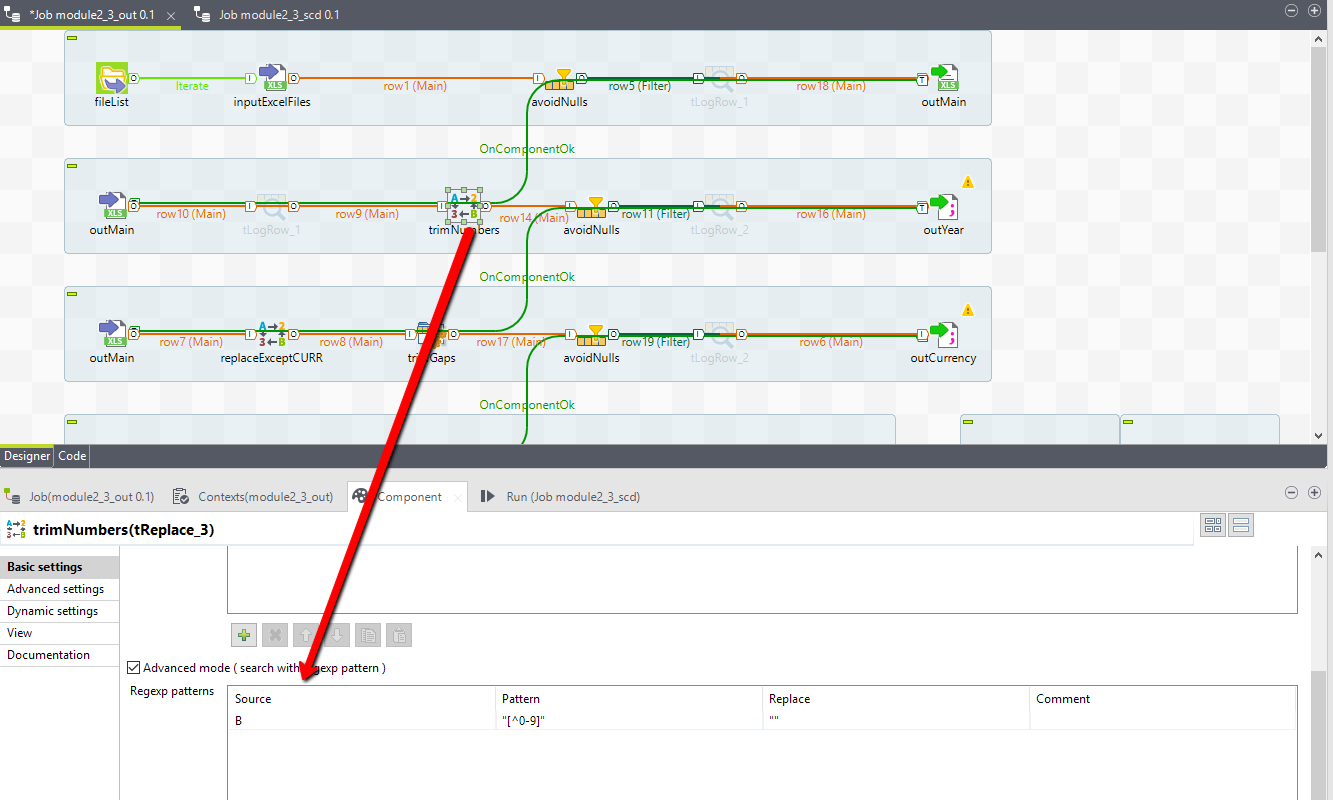
To get file content without unnecessary columns and rows excluding null values do next:



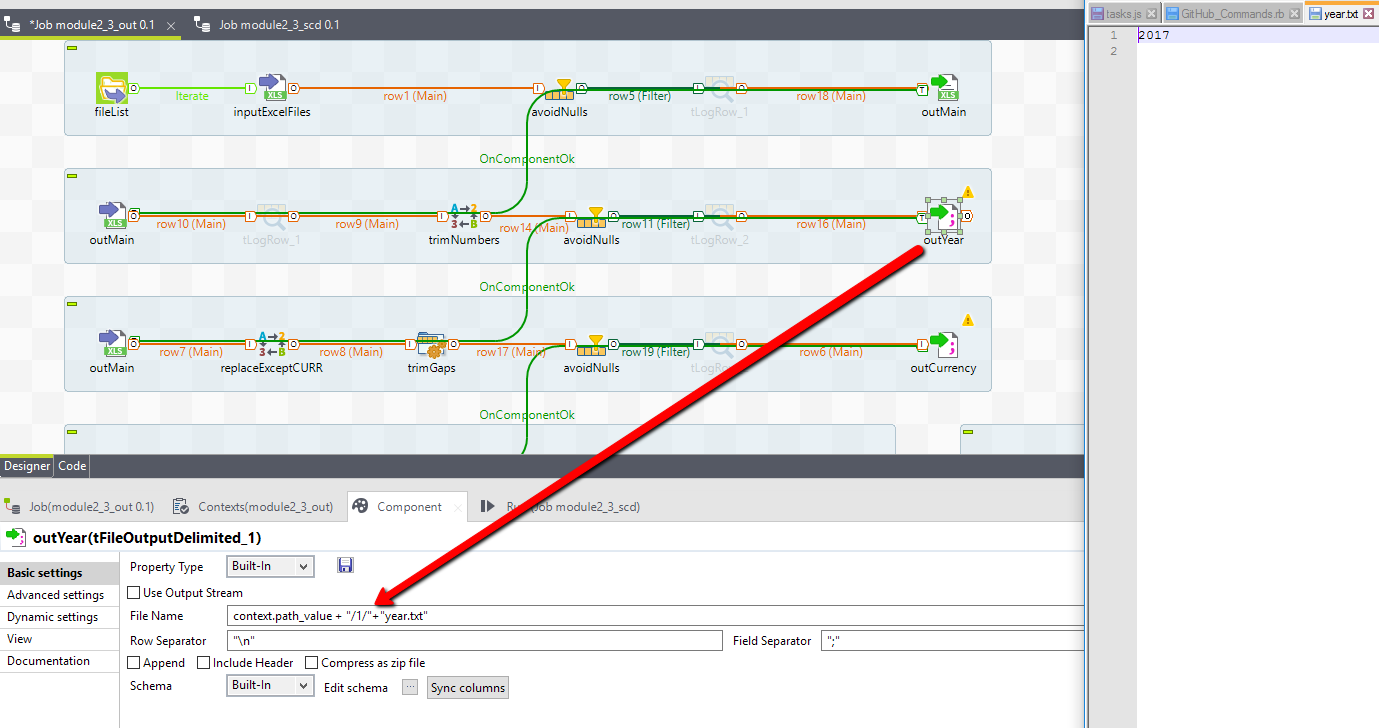
All intermediate file would be stored in current\_path + /1/ folder.



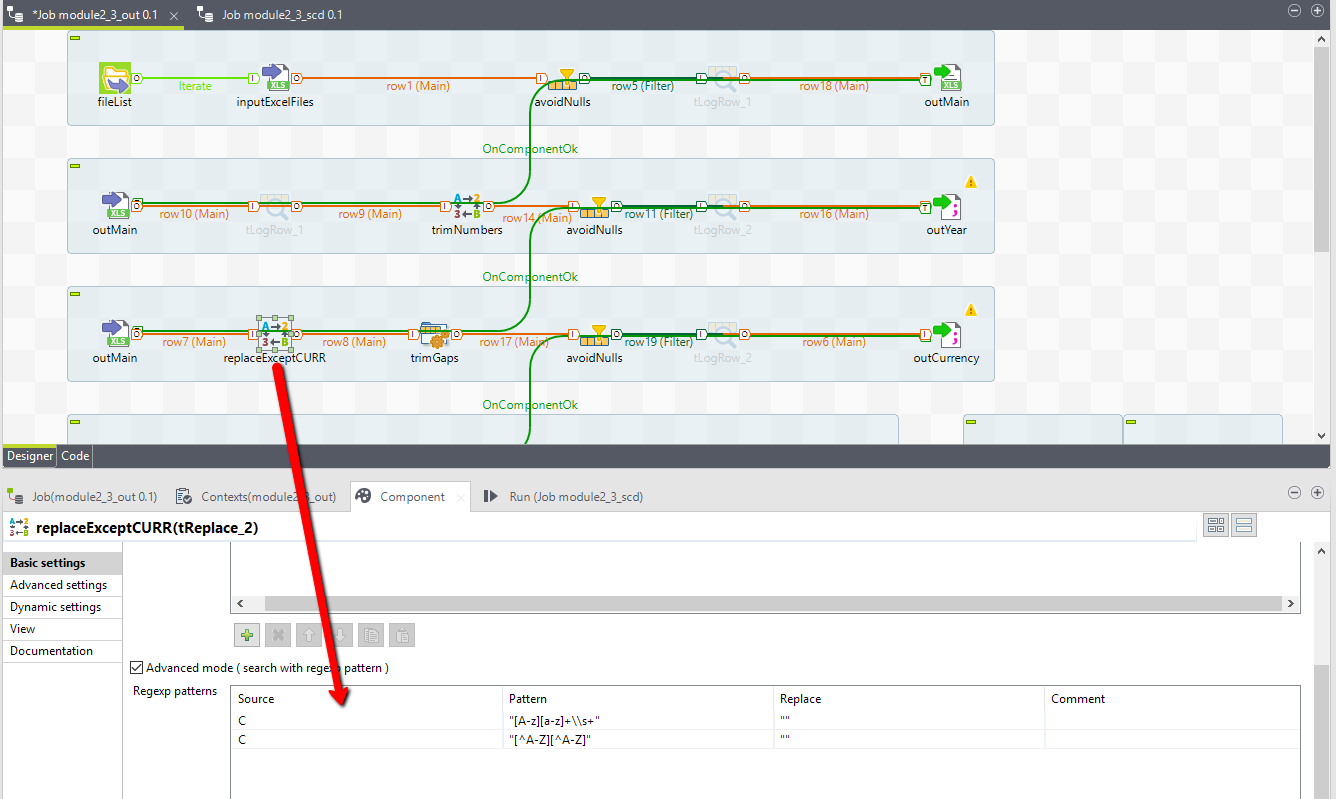
Next, pick the Year from the Year and Month column (out.xls). Using tReplace statement trim all not number symbols. In the avoidNull filter stage delete null values.



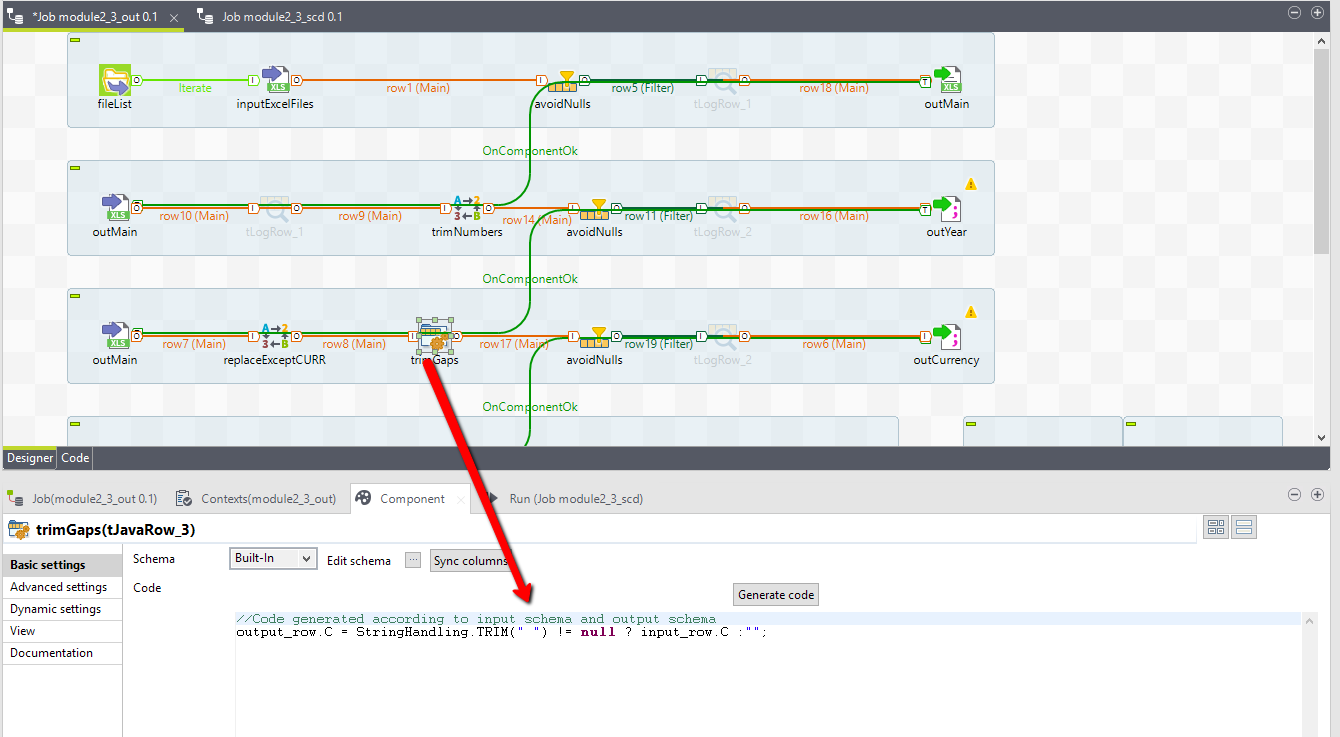
Save result to the .txt file just like single year value.



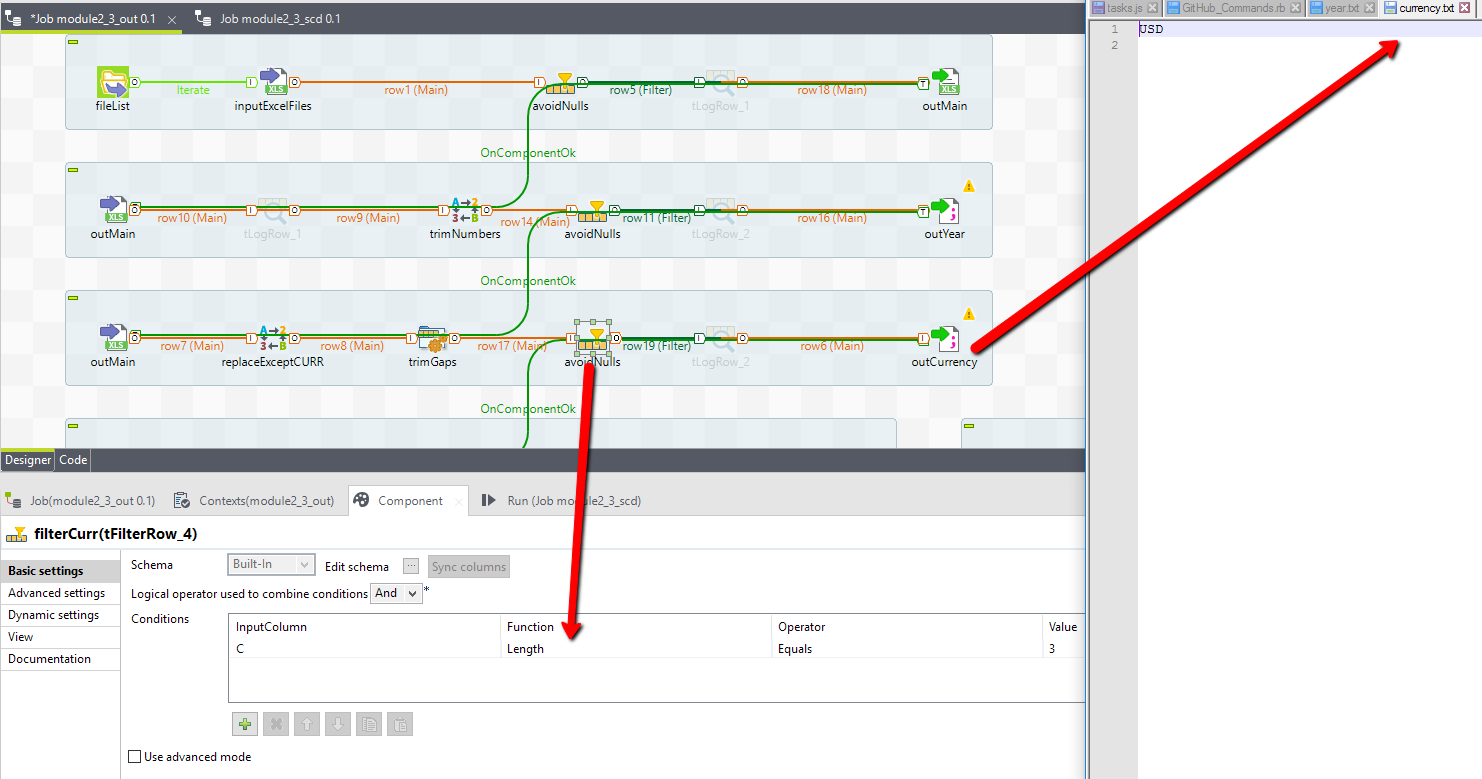
Third step – pick Currency abbreviation. Do two replace statements one after another.



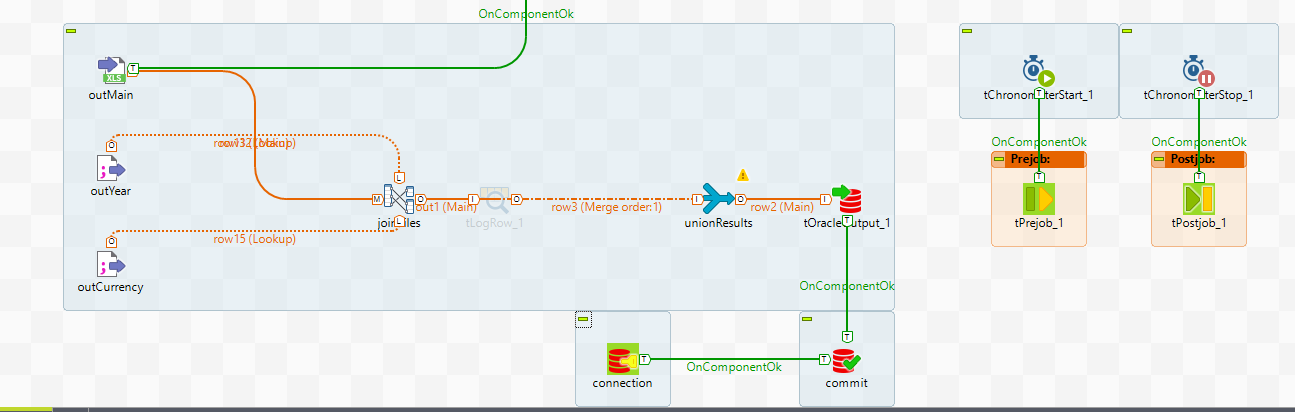
Delete gaps and process the null values case.



Next pick only those strings which have length = 3 symbols. Every currency code should have three symbols (regarding [this](https://en.wikipedia.org/wiki/ISO_4217) information).

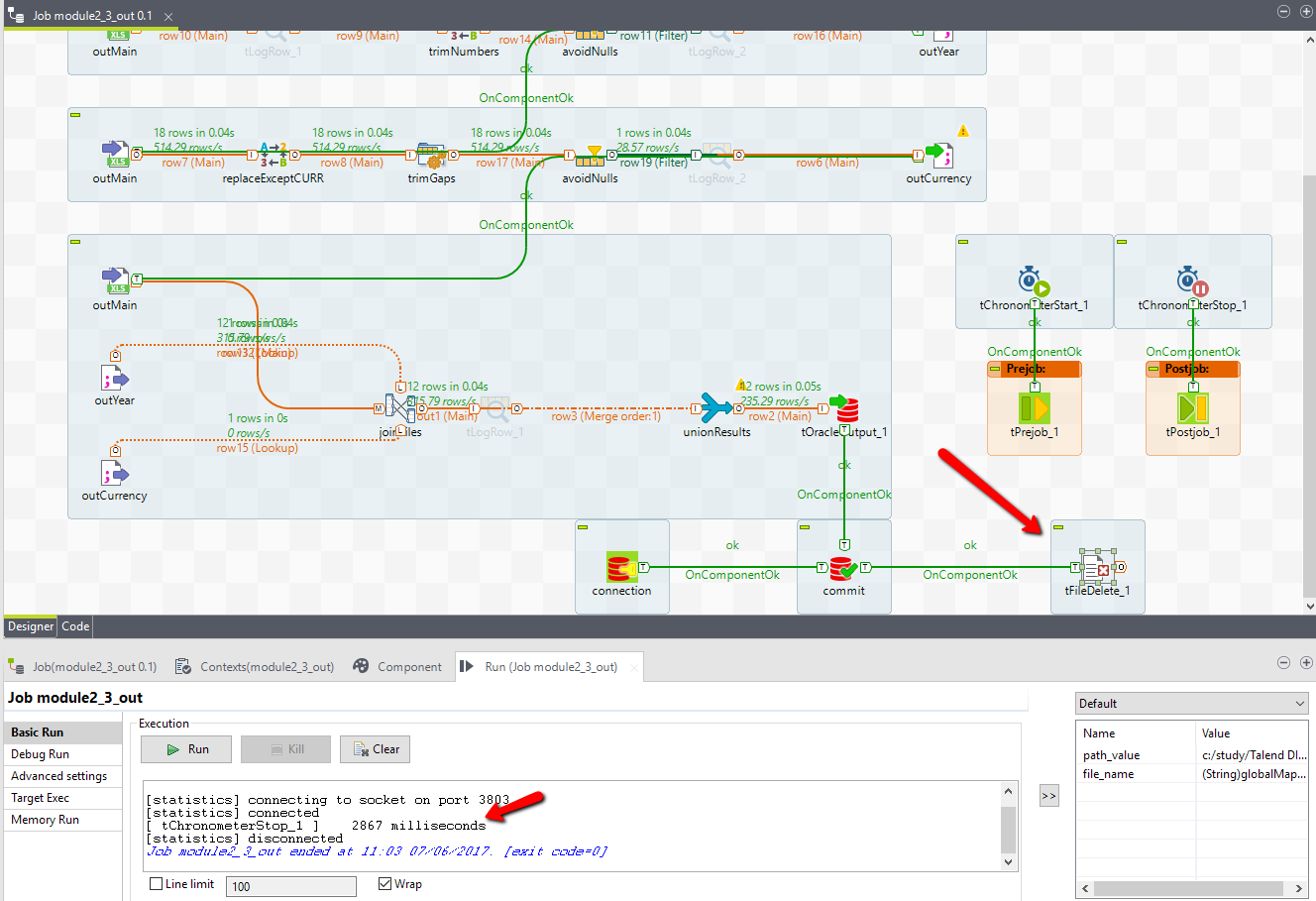


Finally, join three files between each other, make unite for every file in the tFileList list and write it to the DB (this step has the same logic as on the Method 1).



This job has average running time approximately 1.2 sec. (Previous tests showed running time for Method 2 is about 3.1 sec vs 1.2 sec for Method 1).

Intermediate files were created to make the process better understanding. On the other hand these files unnecessary. So, it’s possible to delete them using tFileDelete like on the picture below way.



**Conclusion:**

In spite of the fact that both methods give the same result and comparable running time (last comparing) the Method 2 is more appropriate and more safety for data content.