

Talend Documentation

Title: Extracting Data of CSV files Using tFilelist, tFlowtoiterrate, tIterratetoflow

Submitted By

Pratik Meshram (BDSA255)

Abhinav Kumar (BDSA243)

Waibhav Kumar (BDSA254)

DEPARTMENT OF COMPUTER ENGINEERING

MIT ACADEMY OF ENGINEERING

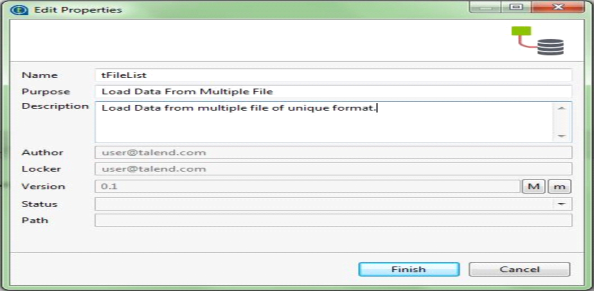
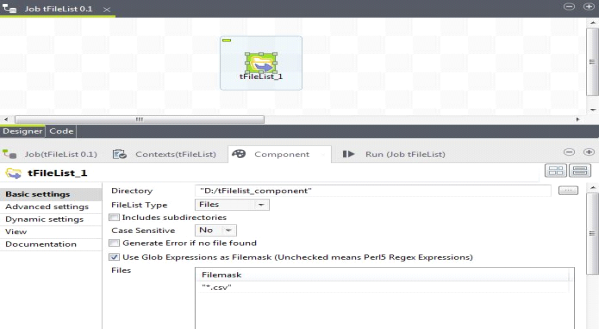
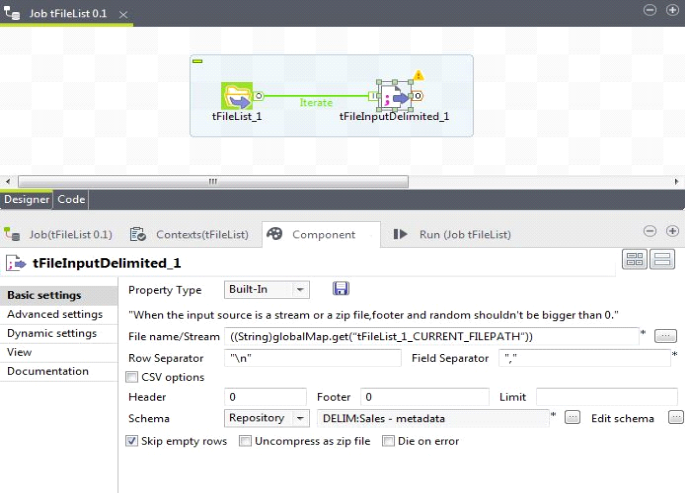
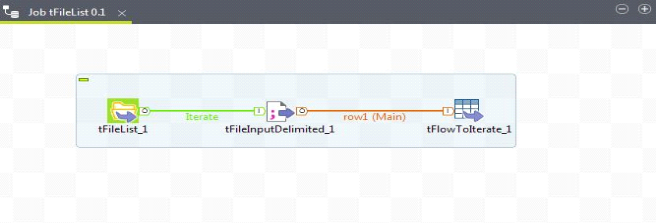
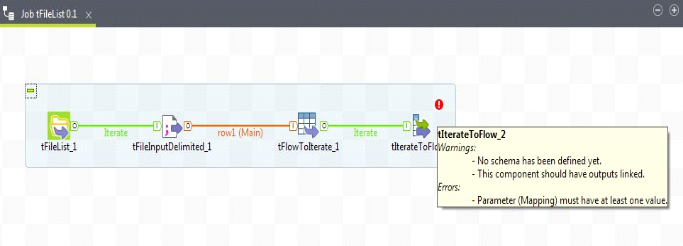
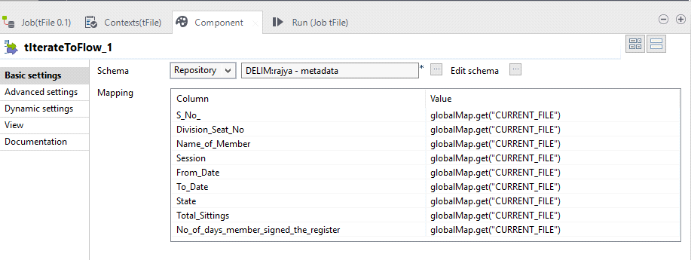
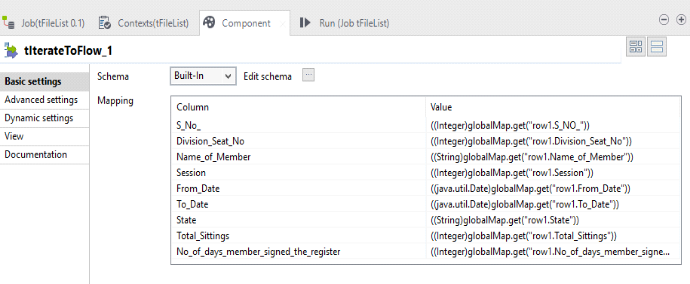
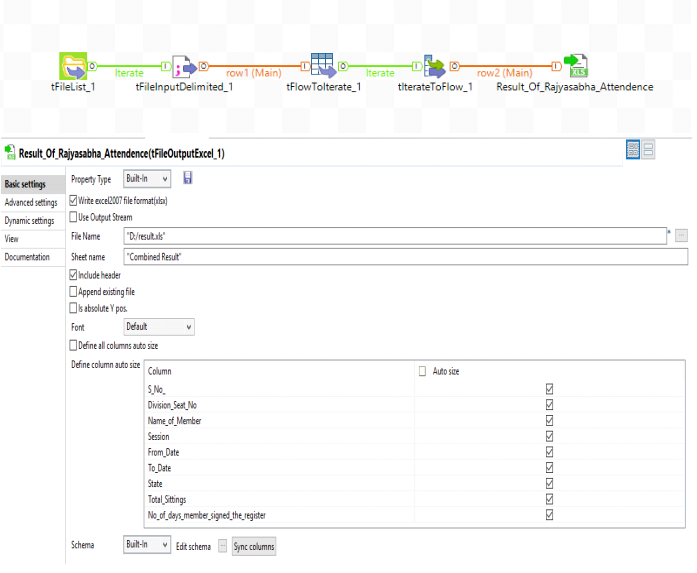
ALANDI (D), PUNE

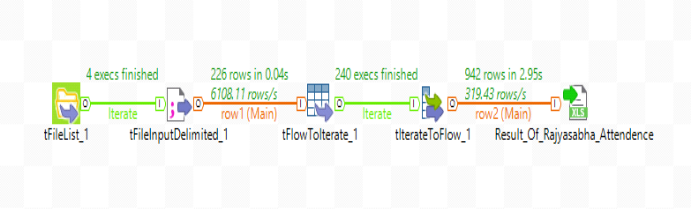
**Introduction**

tFileList can be used to merge data of multiple files having same schema and same file mask into a single file or into a database. It supports four file extensions viz .xls, .xlsx, .csv, .txt. This component iterates over the files present in given folder. Remember – tFileList will not iterate the files recursively.

However, if we want to merge the data into a file instead of database, the output file will have data from the last read file only instead of the merged data.

To solve this problem we must use two new components – tFlowToIterate and tIterateToFlowto iterate all the files and store all the data  in a single output file.

* tFlowToIterate will process the files one by one.
* tIterateToFlow will store the data till all the files are processed and data from all those files has been fetched.
* Now, merge data from all above csv files, stored in one directory into one single excel file we need to follow this steps:
* Create new job and name it (Say “**tFileList”**)[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture6.png)
* Add **tFileList** component from the Palette.
* Configure the Basic settings of **tFileList** component.
  + Browse the directory where the desired files, to be merged are stored.
  + In filemask mention the type of file that is their extension, in our case it is .csv
  + Thus, all .csv files present in given location will be read one after other.[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture7.png)
* Select “**tFileInputDelimited”** from the Palette. Connect tFileList with **tFileInputDelimited** using iterator link.
  + Configure the properties of **tFileInputDelimited**.
  + In file name, specify/set global variable which will fetch the path of the current file that has iterated which can be something like this – **((String)globalMap.get(“tFileList\_1\_CURRENT\_FILEPATH”))**
  + Change Field Separator: “,”
  + Row Separator: “\n”
  + Schema can be taken from repository.[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture8.png)
* Add **tFlowToIterate**component and connect **tFileInputDelimited** with **tFlowToIterate.**[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture9.png)
* Add **tIterateToFlow** from Palette and connect **tFlowToIterate** with it using Itrate link.[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture10.png)
* For tIterateToFlow component, select same schema from repository which we have taken for **tFileInputDelimited**component as shown below.[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture11.png)
  + Here, you can see that value of columns are **globalMap.get(“CURRENT\_FILE”).**
  + This will return object but our values are **not** in form of objects. So, we will convert (cast) them to required data types.[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture12.png)
* Add**tFileOuputExcel** from Palette and connect it to**tIterateToFlow**and set all properties as shown below. This is where we will have merged data.[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture13.png)
* Run the Job.

**[](http://www.cannyinformatics.com/wp-content/uploads/2017/07/Picture15.png)**

After successful execution of job the data of all the sheets will be combined and stored at  given location of **tFileOutputExcel**.