# Preparations

Operating system: **Windows**

Softwares:

1. Talend: Talend Open Studio for Big Data (7.3)
2. MongoDB: MongoDB server version 4.4.3

Home directory: **D:\Talend\TOS\_BD-20200219\_1130-V7.3.1\workspace\JTC**

Files to be placed in Home directory:

1. Raw data file: **2009.zip**
2. Input file for DimProducts: **DimProducts.txt**
3. Input file for DimStores: **DimStores.txt**



# Process description

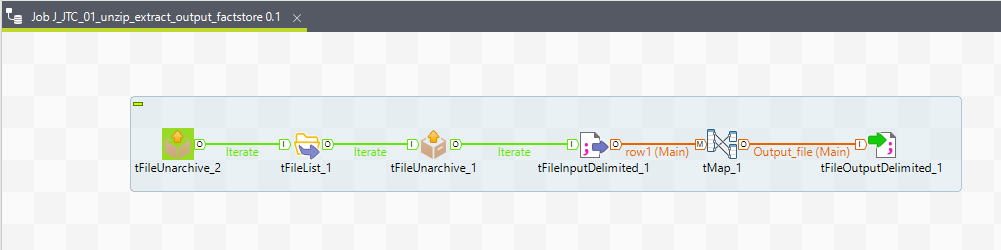
## Steps:

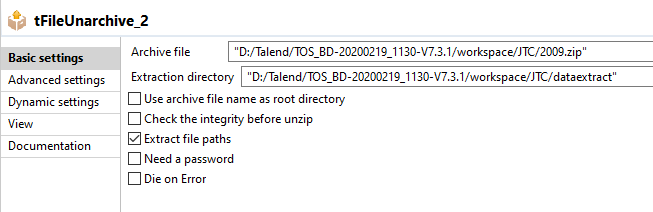
1. Unzip 2009.zip
2. Extract by converting the “.gz” files into “.txt” files
3. Read the input txt files into pipe delimited files and store in factstore folders
4. Combile all the factstore files in single factstore file
5. Upload the factstore file into MongoDB collection, “factstore” in database “JTC”
6. Upload the DimProducts.txt into MongoDB collection, “DimProducts” in database “JTC”
7. Upload the DimStores.txt into MongoDB collection, “DimStores” in database “JTC”
8. Execute to query average number of stocks on hand by category using collections “factstore” and “DimProducts”
9. Execute to query average number of stocks on hand by store by using collections “factstore” and “DimStores”

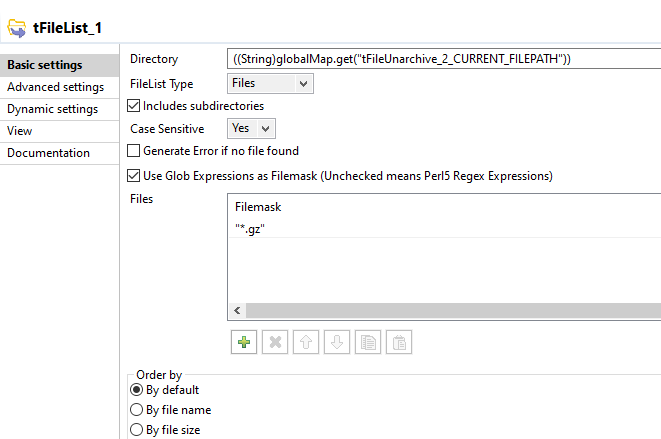
## All these are accomplished via the below jobs:

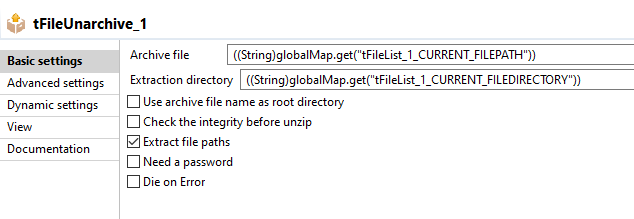
* J\_JTC\_01\_unzip\_extract\_output\_factstore
* J\_JTC\_02\_combine\_factstore\_into\_single\_file
* J\_JTC\_03\_create\_collection\_factstore
* J\_JTC\_04\_create\_collection\_DimProducts
* J\_JTC\_05\_create\_collection\_DimStores
* J\_JTC\_06\_Query\_avg\_stksOnHand\_by\_Category
* J\_JTC\_07\_Query\_avg\_stksOnHand\_by\_Store

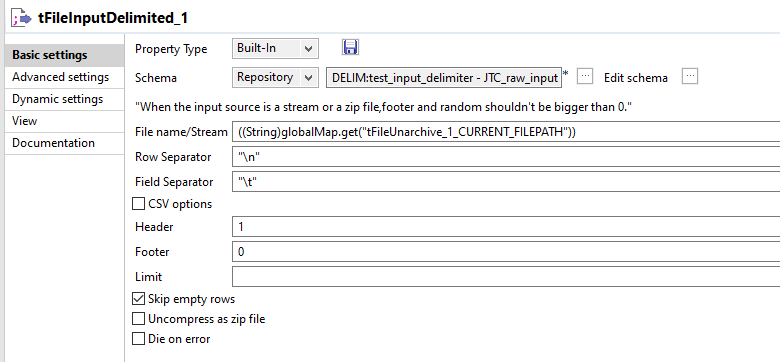
# Job: J\_JTC\_01\_unzip\_extract\_output\_factstore

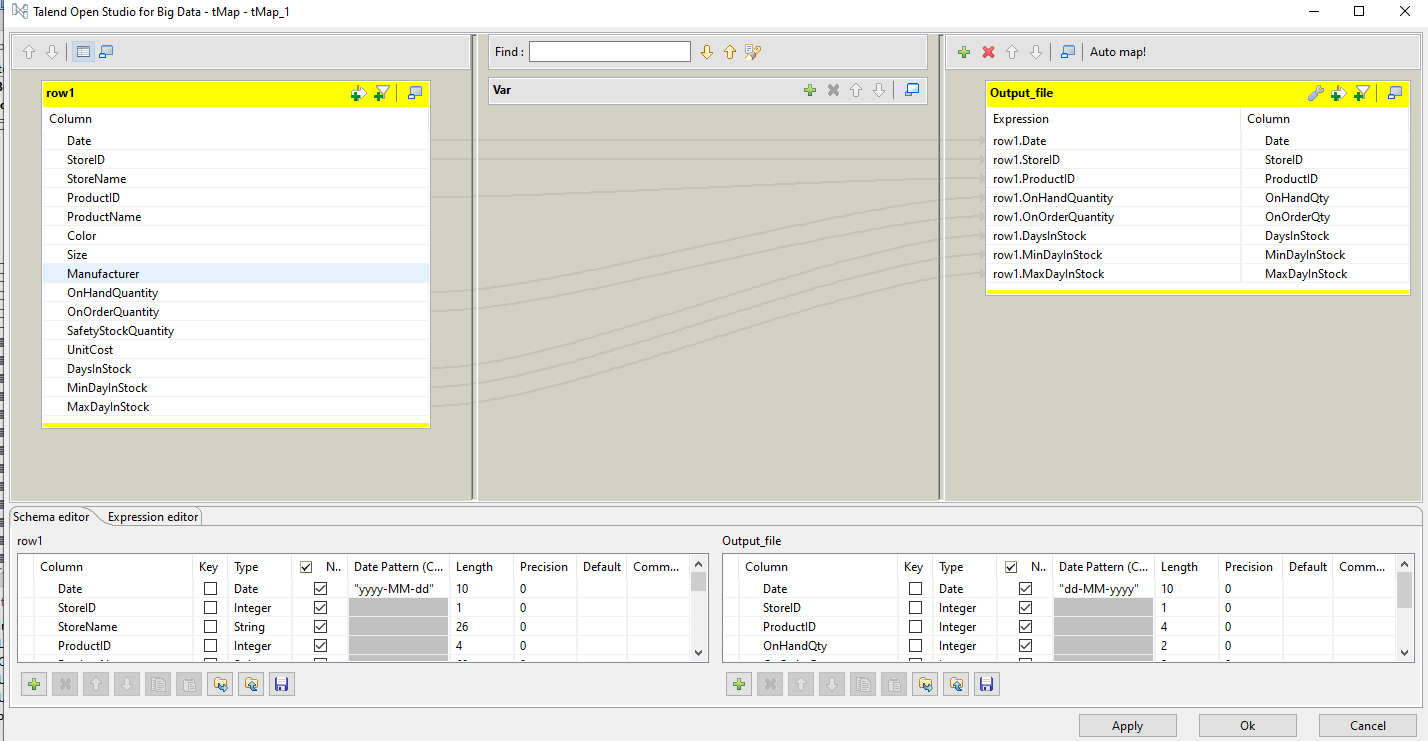


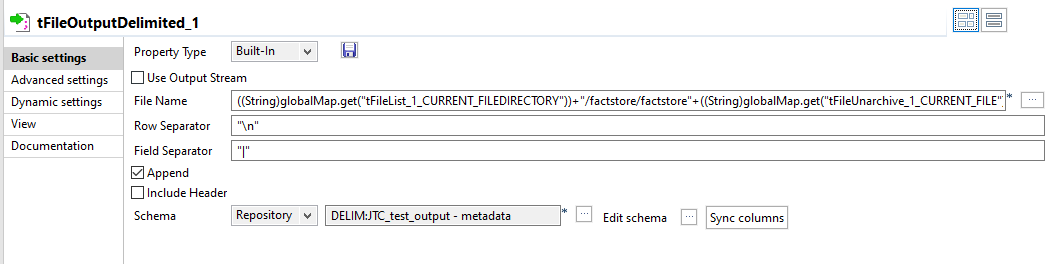




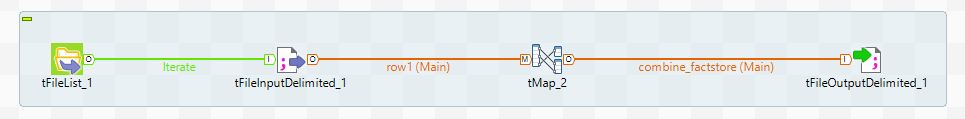


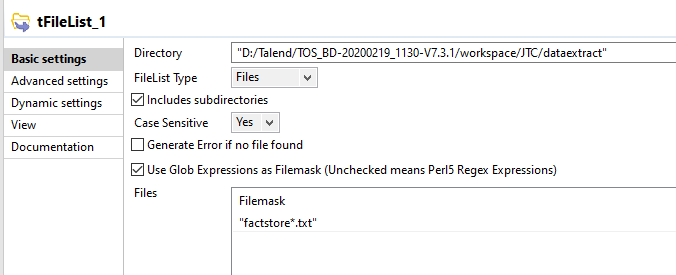


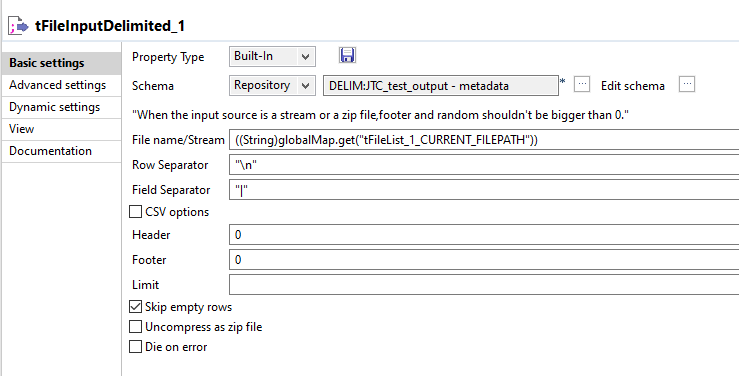


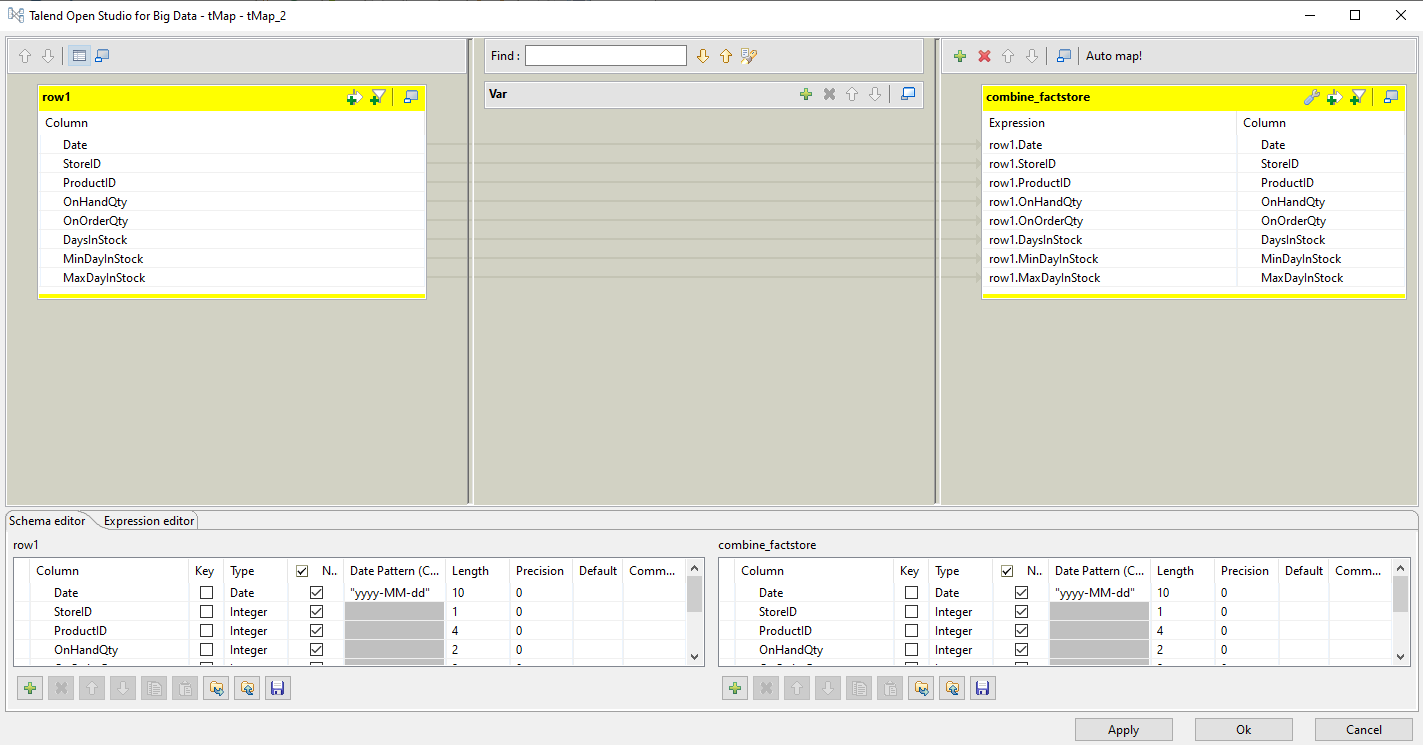


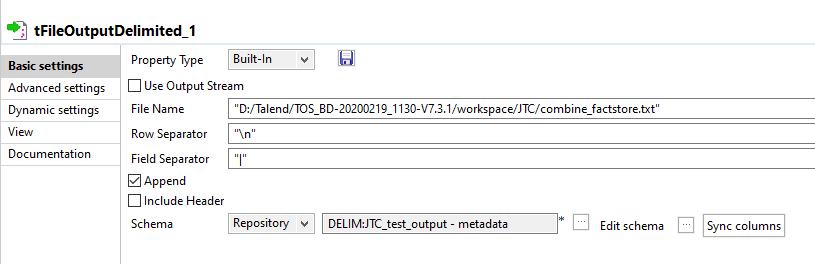
# Job: J\_JTC\_02\_combine\_factstore\_into\_single\_file



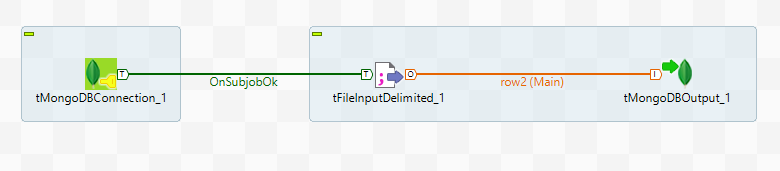


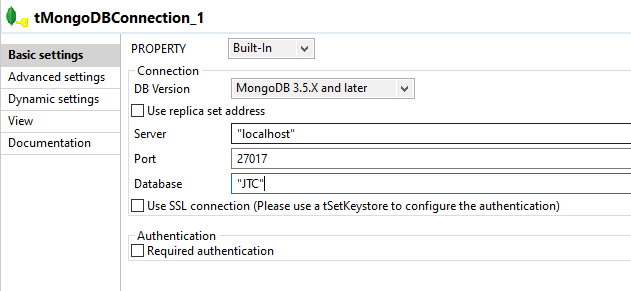


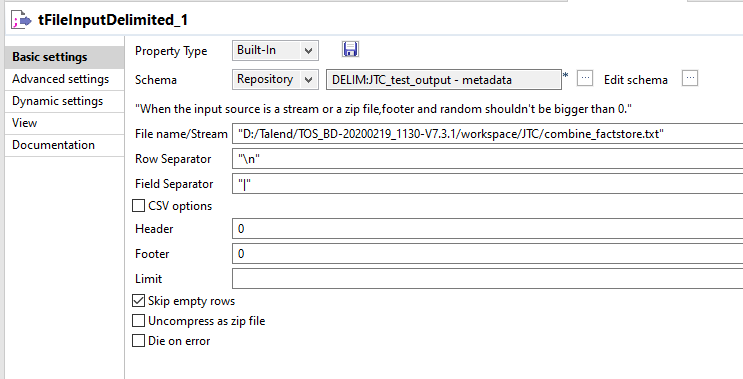


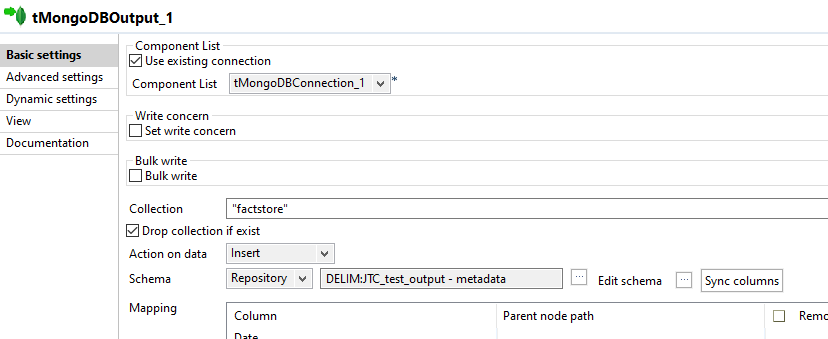


# Job: J\_JTC\_03\_create\_collection\_factstore

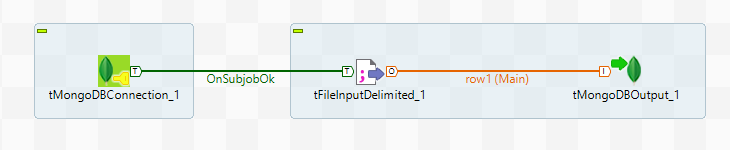


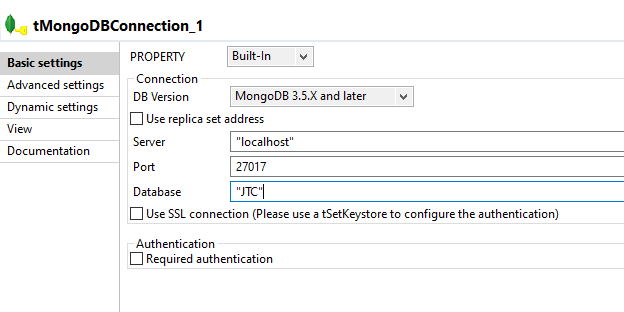


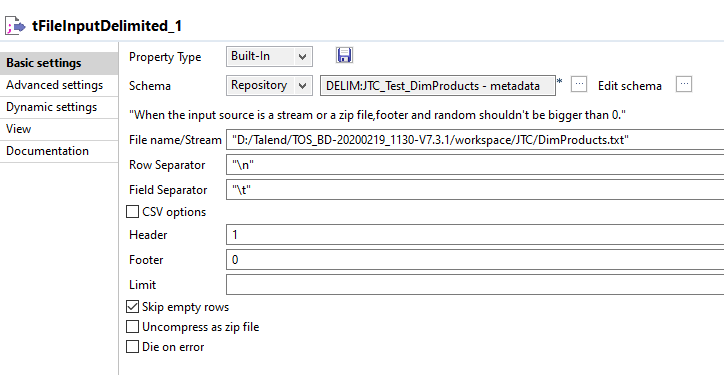


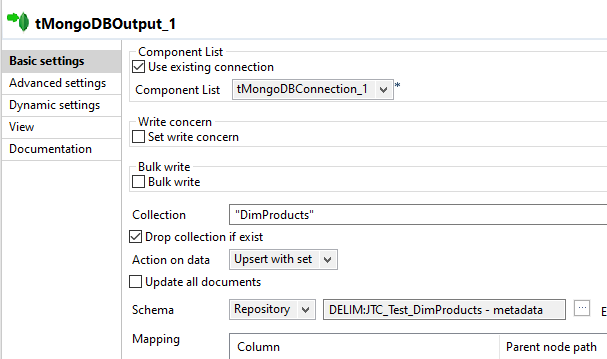


# Job: J\_JTC\_04\_create\_collection\_DimProducts

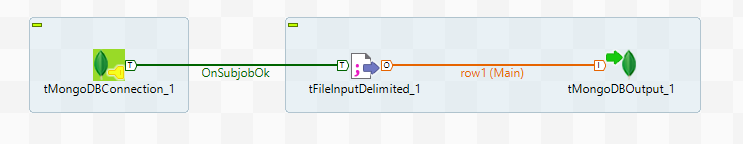


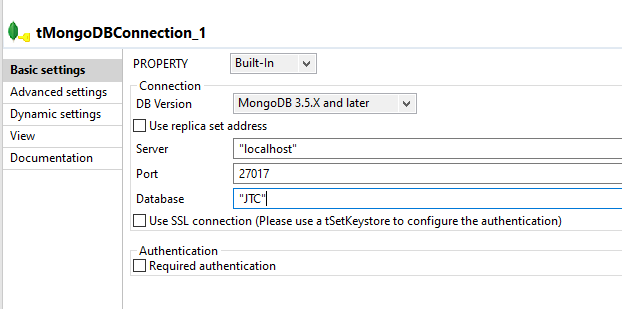


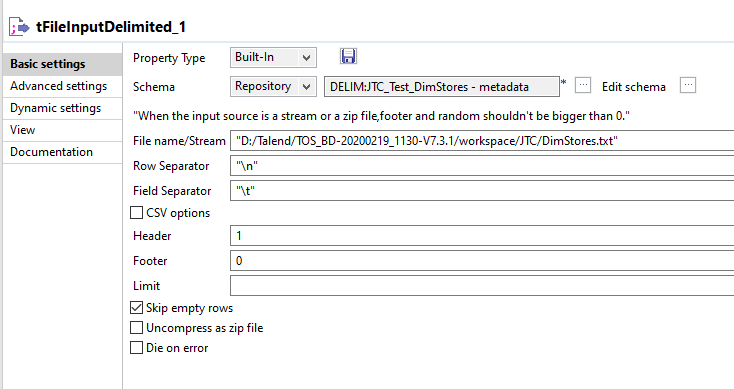


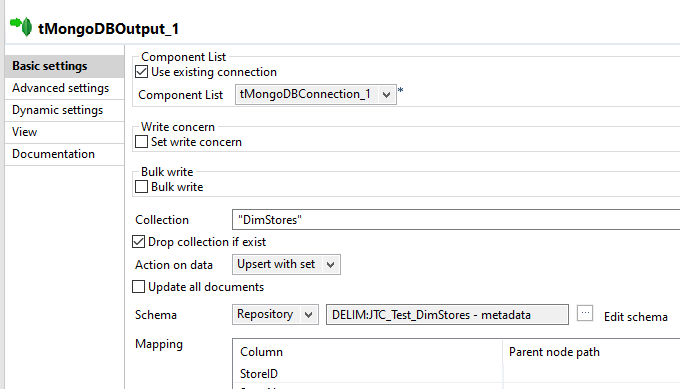


# Job: J\_JTC\_05\_create\_collection\_DimStores



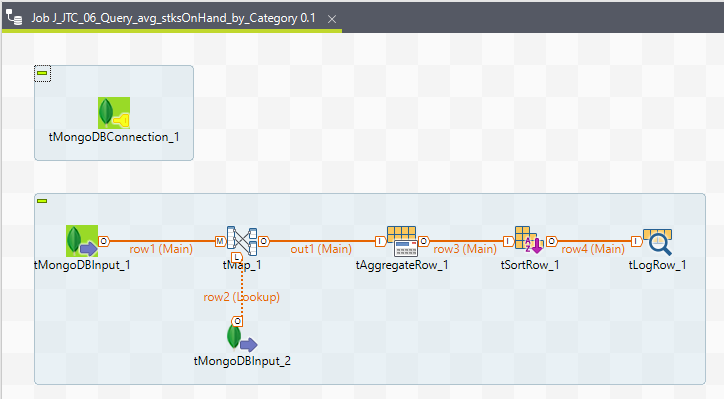


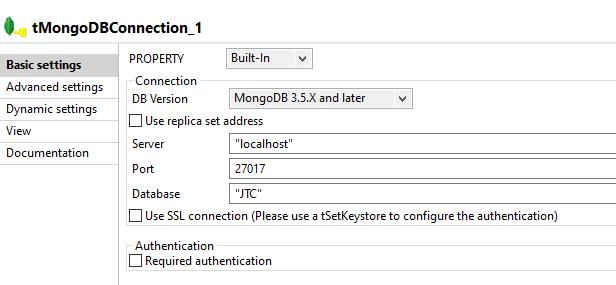


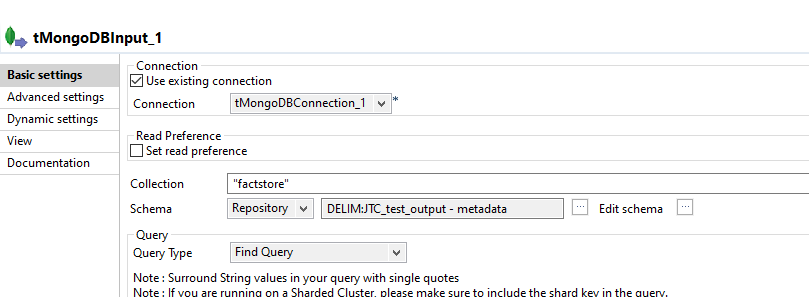


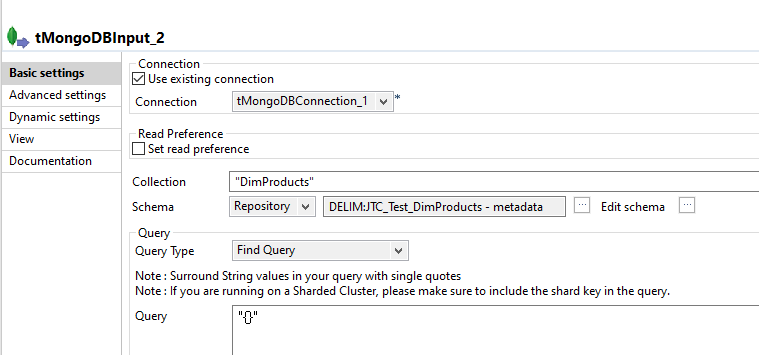
# Job: J\_JTC\_06\_Query\_avg\_stksOnHand\_by\_Category

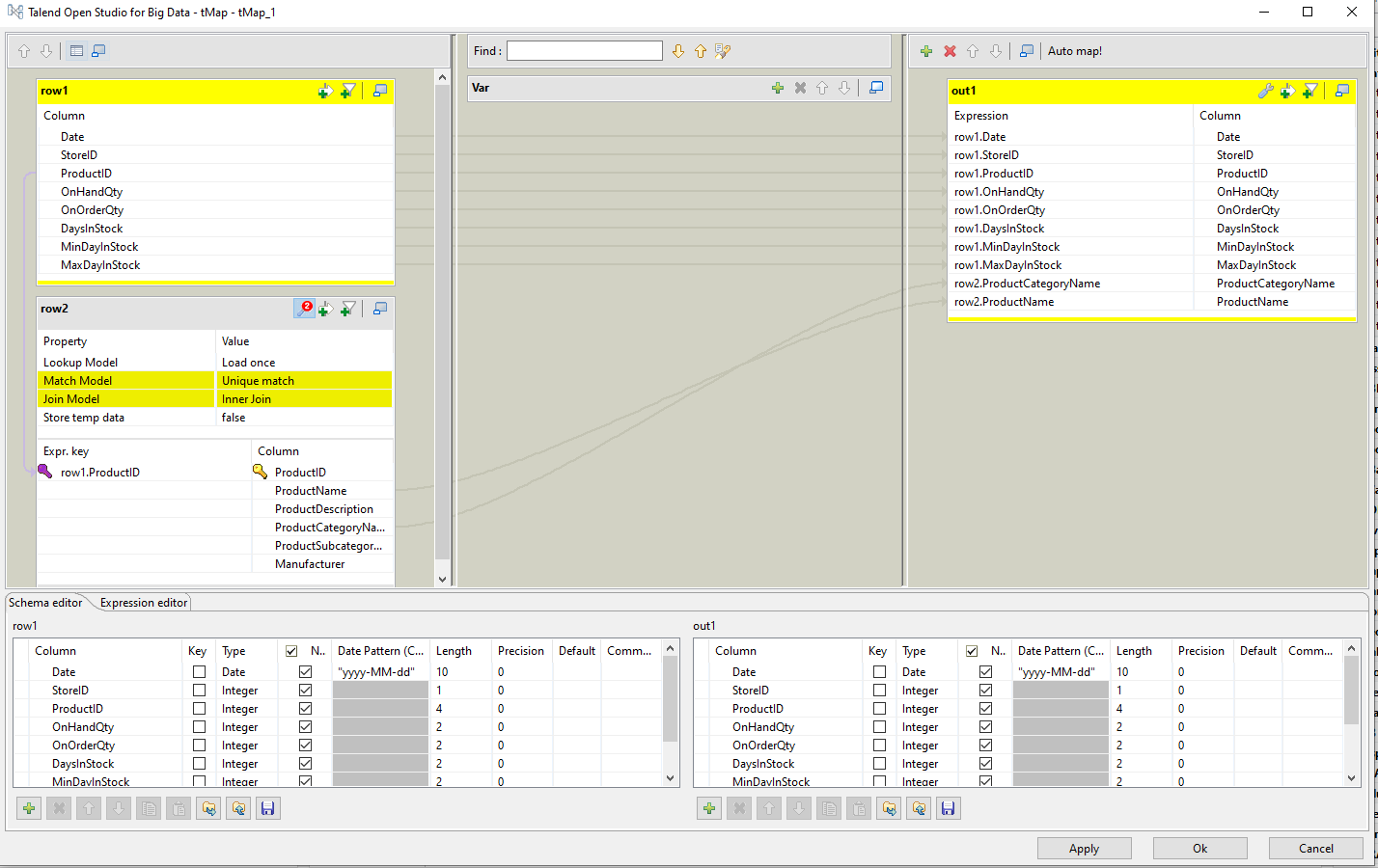
Query average number of stocks on hand by category

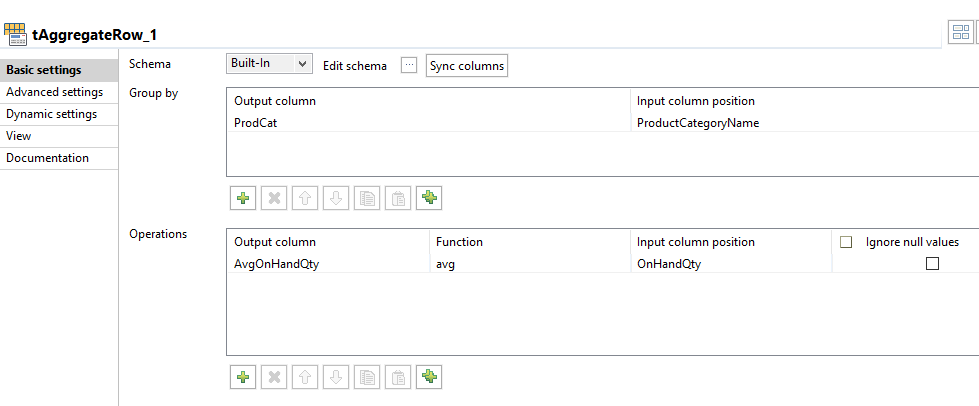


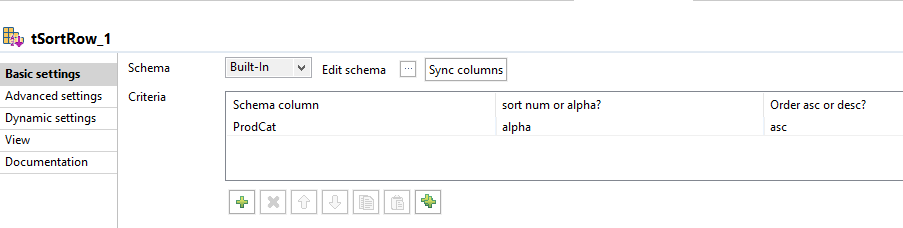


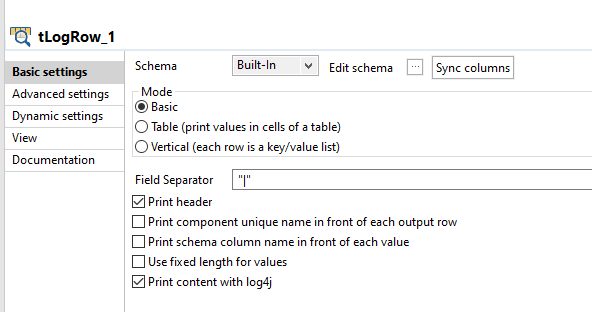




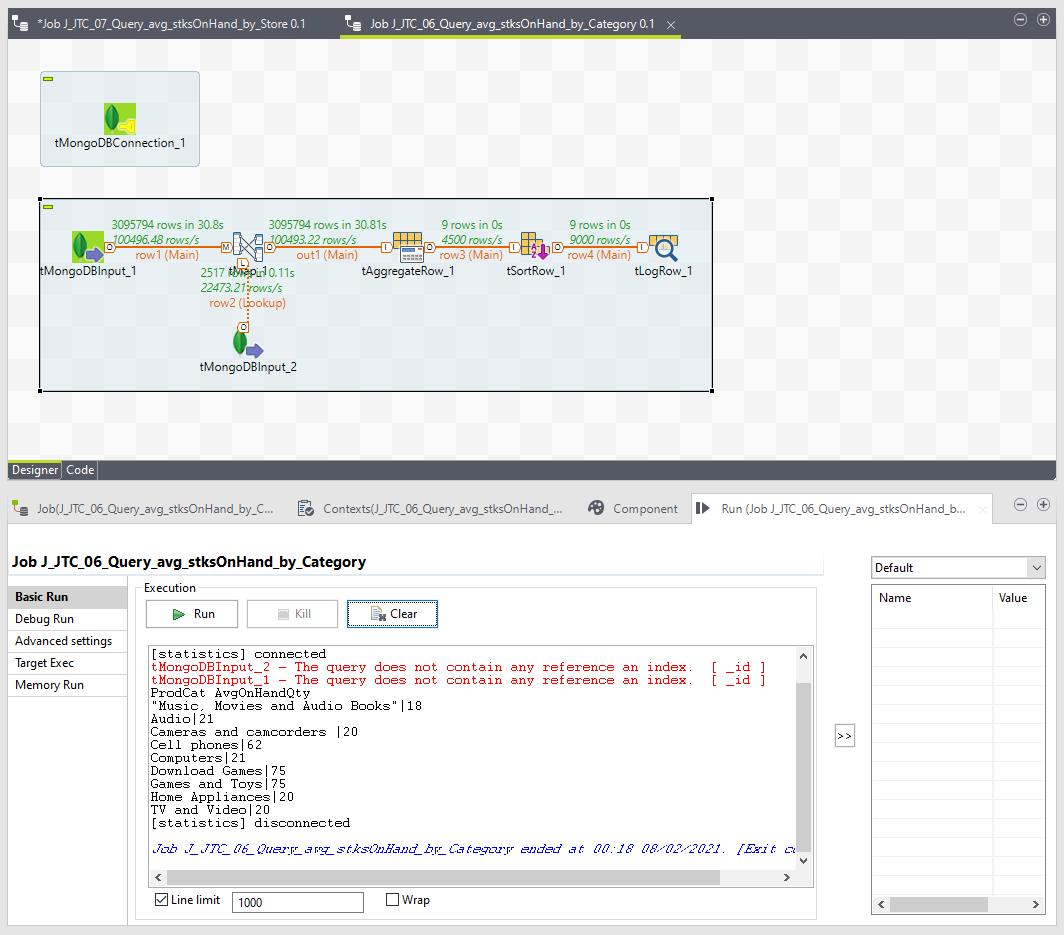






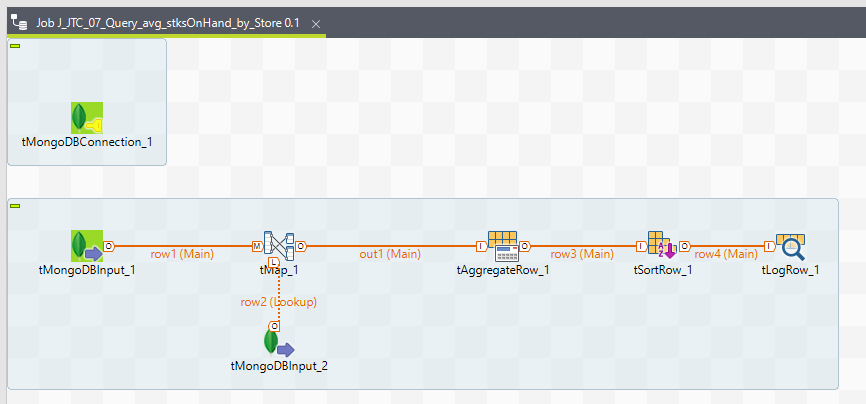


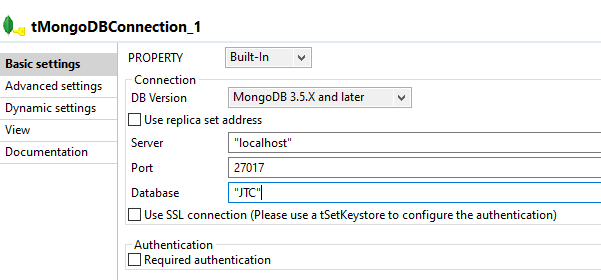
## Expected Result #1: Query average number of stocks on hand by category

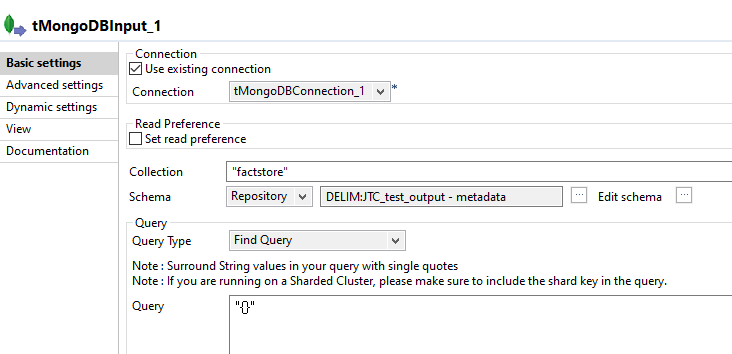


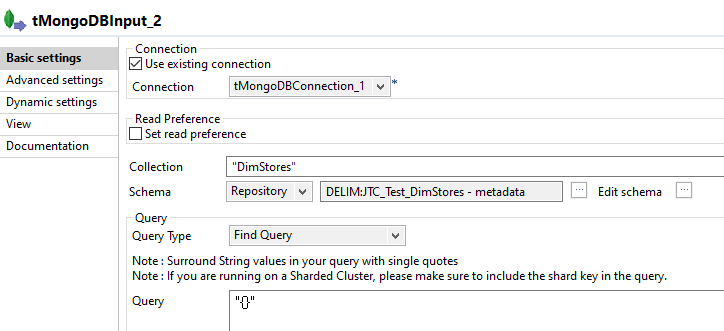
# Job: J\_JTC\_07\_Query\_avg\_stksOnHand\_by\_Store

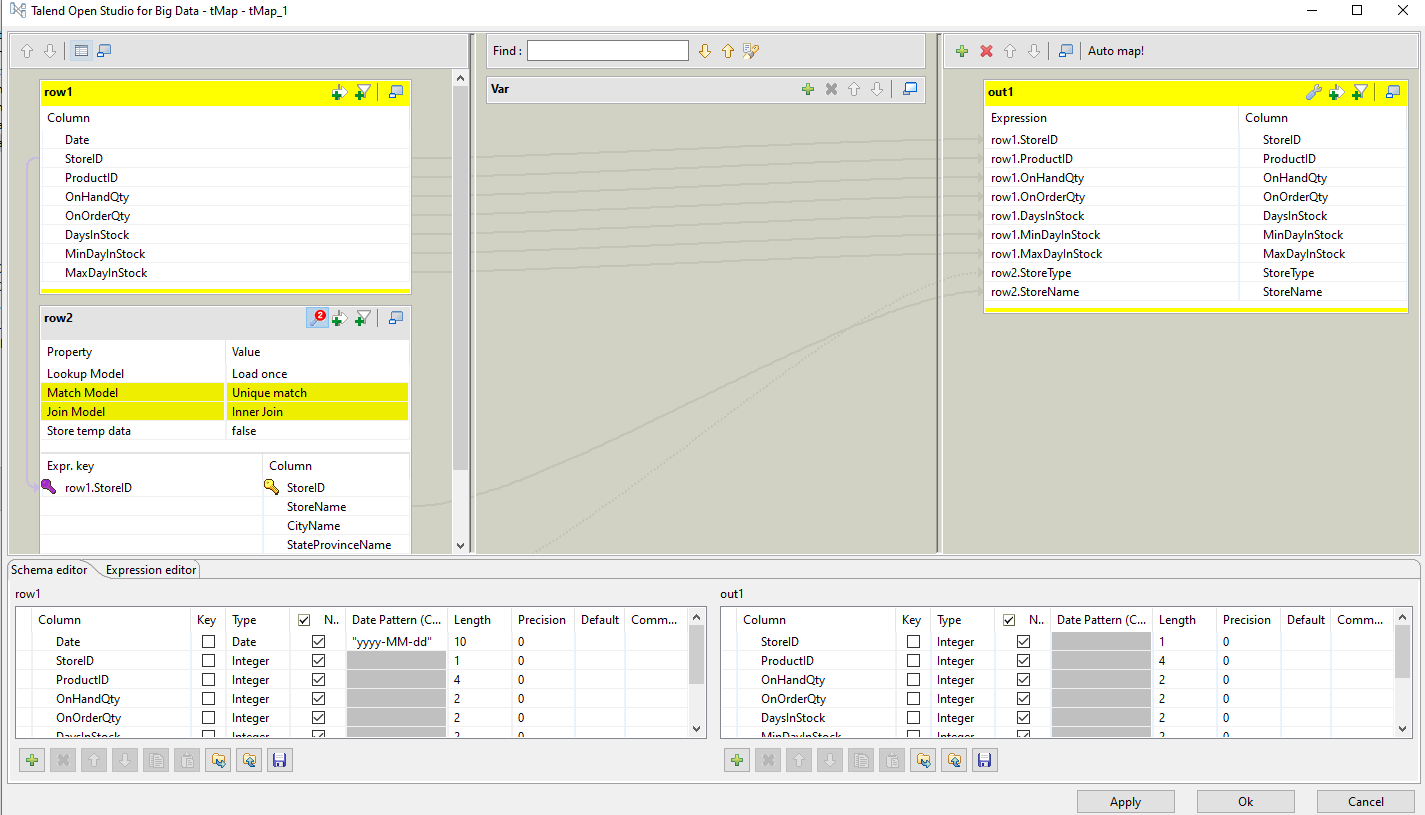
Query average number of stocks on hand by store

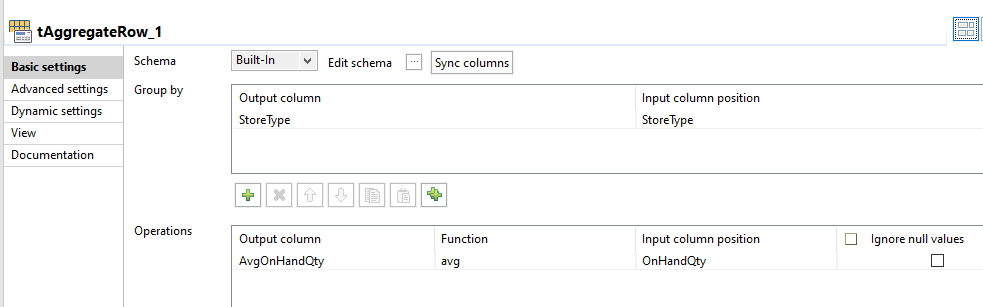


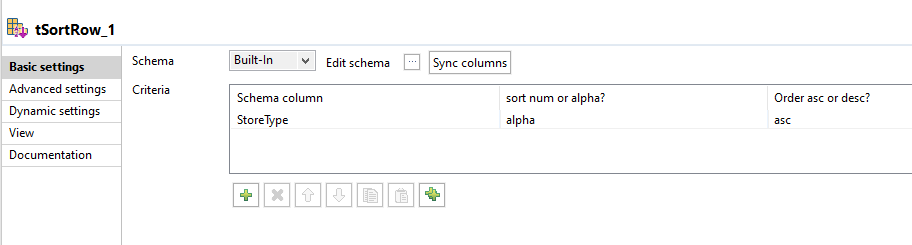


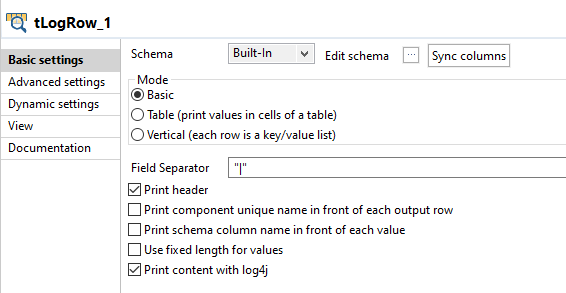












## Expected Result #2: Query average number of stocks on hand by store

