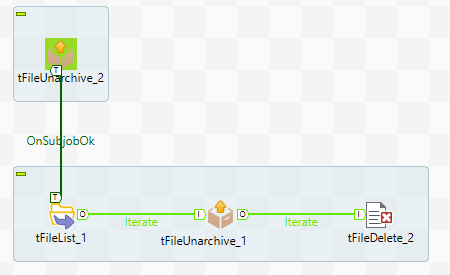
### Install MongoDB version 5 and Talend Open Studio for Big Data v7.3

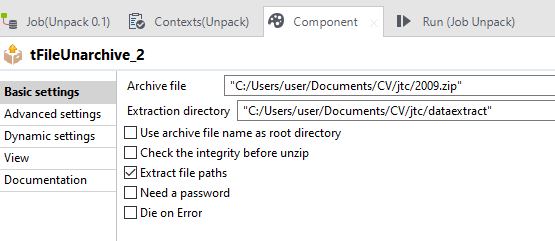
### Create database factstore in MongoDB

### In Talend, Create DB connection to MongoDB

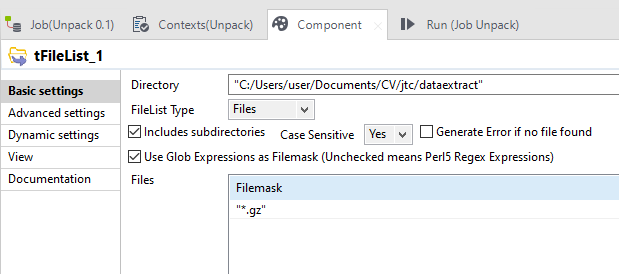
### In Talend, create a job to unpack the data files into C:\Users\user\Documents\CV\jtc\dataextract



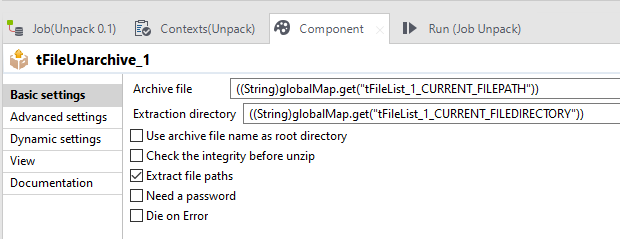
Unzip 2009.zip to C:\Users\user\Documents\CV\jtc\dataextract



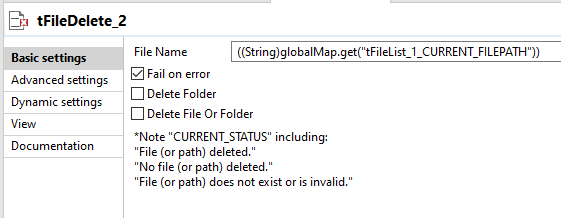
List the files in C:\Users\user\Documents\CV\jtc\dataextract



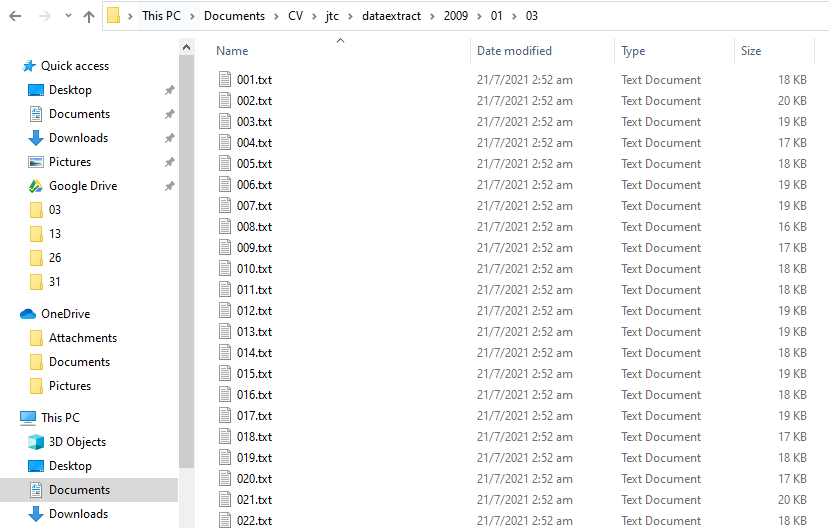
Based on the list in the previous step, iterate through the gz files and upack them to text files, retaining directory structure



Delete the gz files



Sample result of text files extracted

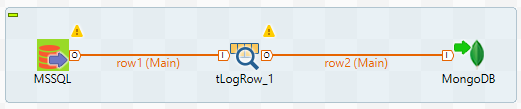


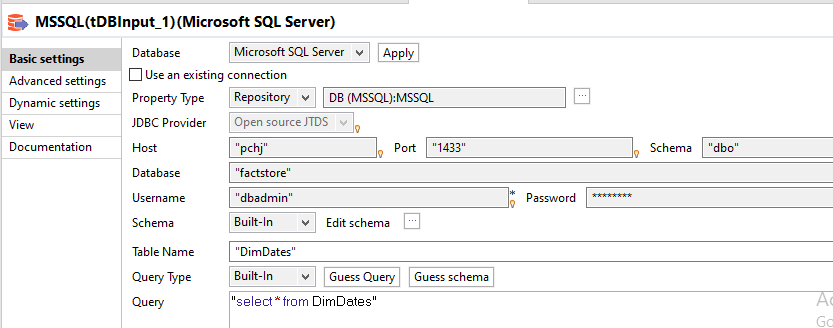
### MSSQL

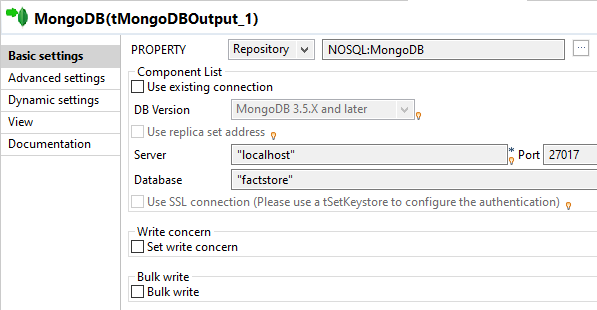
### I made use of MSSQL for testing, verification and for setting up MongoDB database with the required data types and sizes.

### In MSSQL, I run DDL-DML.sql to create the FactStore table

In Talend, I create jobs to setup/copy MSSQL tables (DimDates, DimStores, DimProducts) to MongoDB collections

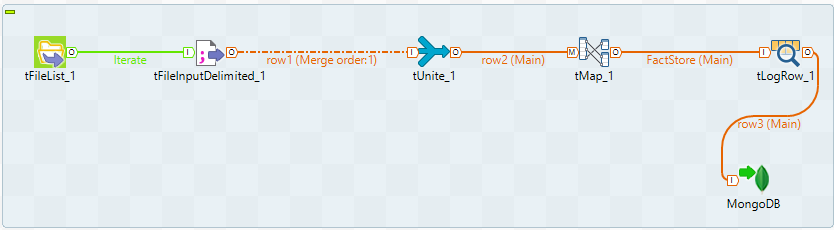


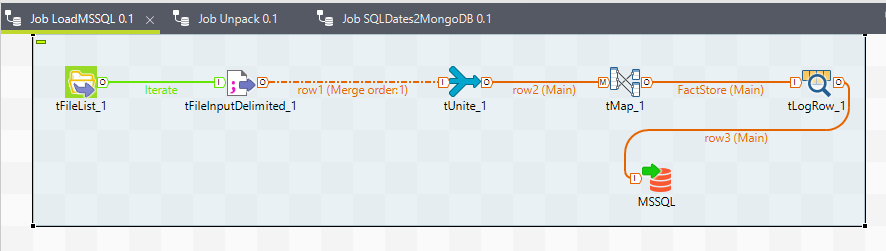




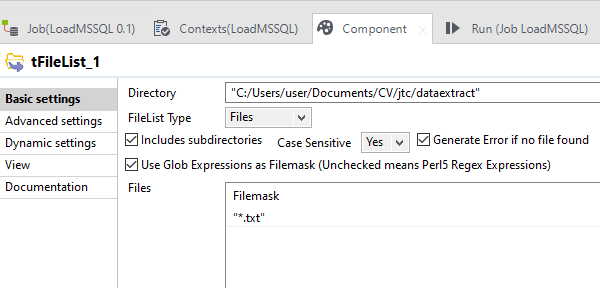
1. Load data from delimited files to MongoDB and MSSQL

Jobs for MongoDB and MSSQL are the same, except for the last component

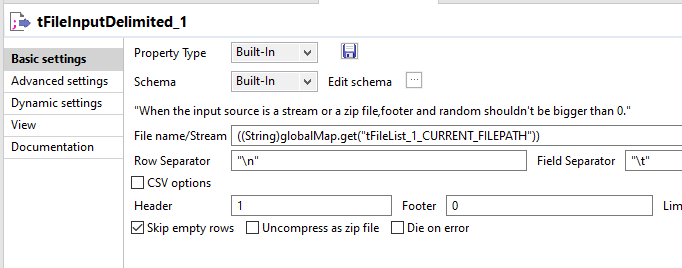




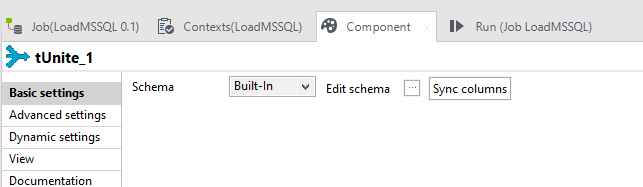
List text files in C:\Users\user\Documents\CV\jtc\dataextract



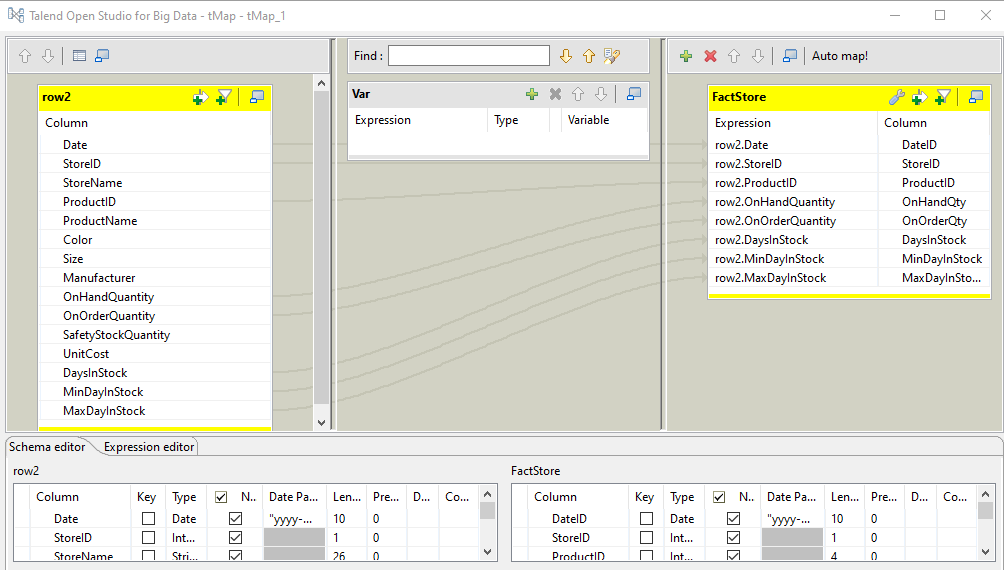
From the list from previous step, parse the delimited files one at a time



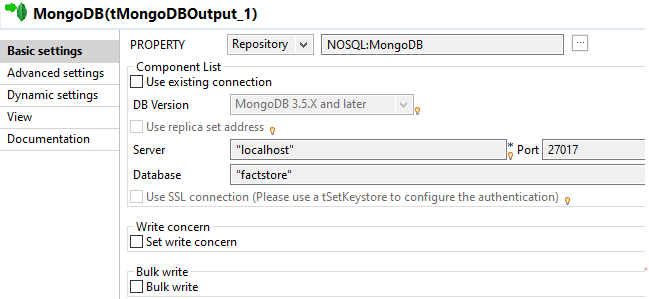
Merge the parsed files into one big file



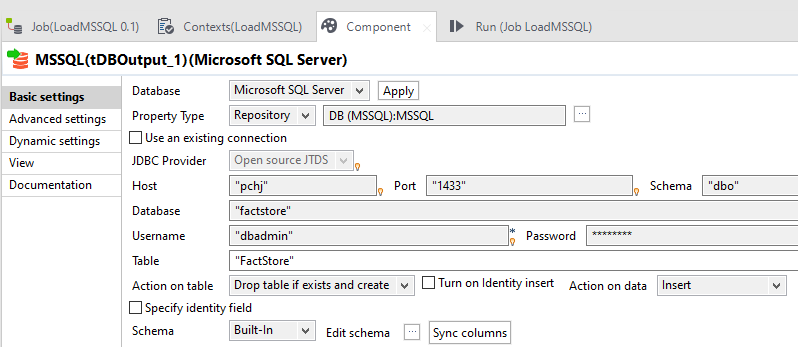
Map selected fields for FactStore - DateID, StoreID, ProductID, OnHandQty, OnOrderQty, DaysInStock, MinDayInStock, MaxDayInStock



Export Data to MongoDB



Export Data to MSSQL



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

(Unfortunately I cannot find a solution to change Average OnHandQty to integer)

**Query:**

db.FactStore.aggregate([

{"$lookup":{"from":"DimProducts","localField":"ProductID","foreignField":"ProductID","as":"StoreProducts"}},

{"$unwind":"$StoreProducts"},

{"$group":{"\_id":{"DimProducts":"$StoreProducts.ProductCategoryName"},"AvgQty":{"$avg":"$OnHandQty"}}}])

**Results:**

{ \_id: { DimProducts: 'Home Appliances' },

AvgQty: 20.336342483291062 }

{ \_id: { DimProducts: 'Audio' }, AvgQty: 21.75158076962575 }

{ \_id: { DimProducts: 'Cameras and camcorders ' },

AvgQty: 20.67926907486834 }

{ \_id: { DimProducts: 'Music, Movies and Audio Books' },

AvgQty: 18.86934555493405 }

{ \_id: { DimProducts: 'TV and Video' },

AvgQty: 20.413036681950814 }

{ \_id: { DimProducts: 'Cell phones' },

AvgQty: 62.32336163476759 }

{ \_id: { DimProducts: 'Games and Toys' },

AvgQty: 75.91057615846165 }

{ \_id: { DimProducts: 'Computers' }, AvgQty: 21.73863693874828 }

### **Expected Result #2: Query average number of stocks on hand by store (Type).**

**Query**

db.FactStore.aggregate([

{"$lookup":{"from":"DimStores","localField":"StoreID","foreignField":"StoreID","as":"Stores"}},

{"$unwind":"$Stores"},

{"$group":{"\_id":{"DimStores":"$Stores.StoreType"},"AvgQty":{"$avg":"$OnHandQty"}}},

{ $sort : {"\_id" : 1 } }

])

**Results:**

{ \_id: { DimStores: 'Catalog' }, AvgQty: 59.49194060159919 }

{ \_id: { DimStores: 'Online' }, AvgQty: 47.97443884369705 }

{ \_id: { DimStores: 'Reseller' }, AvgQty: 37.56381996182501 }

{ \_id: { DimStores: 'Store' }, AvgQty: 24.26701030622914 }