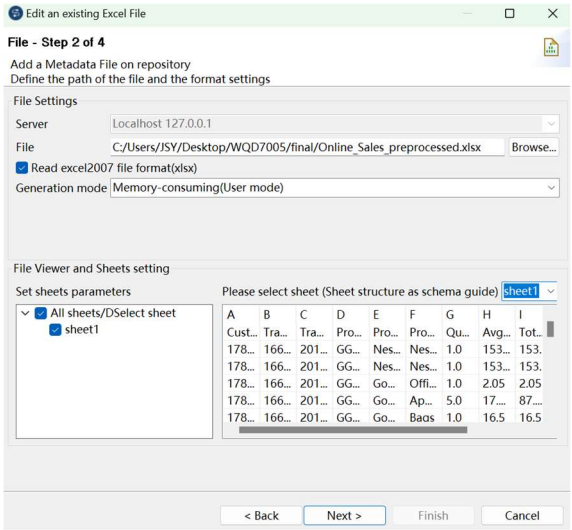
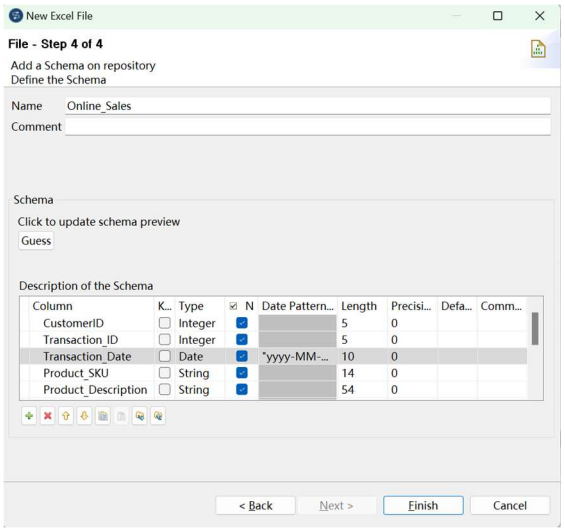
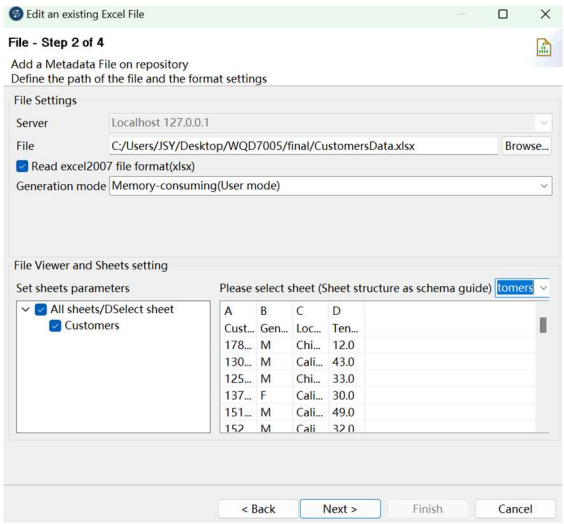
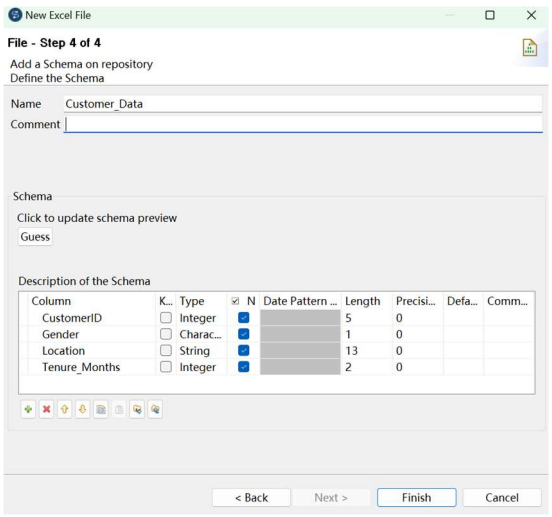
Data Integration Using Talend Open Studio

Import CustomerData.xlsx and Online_Sales_preprocessed.xlsx, and justify the data pattern:

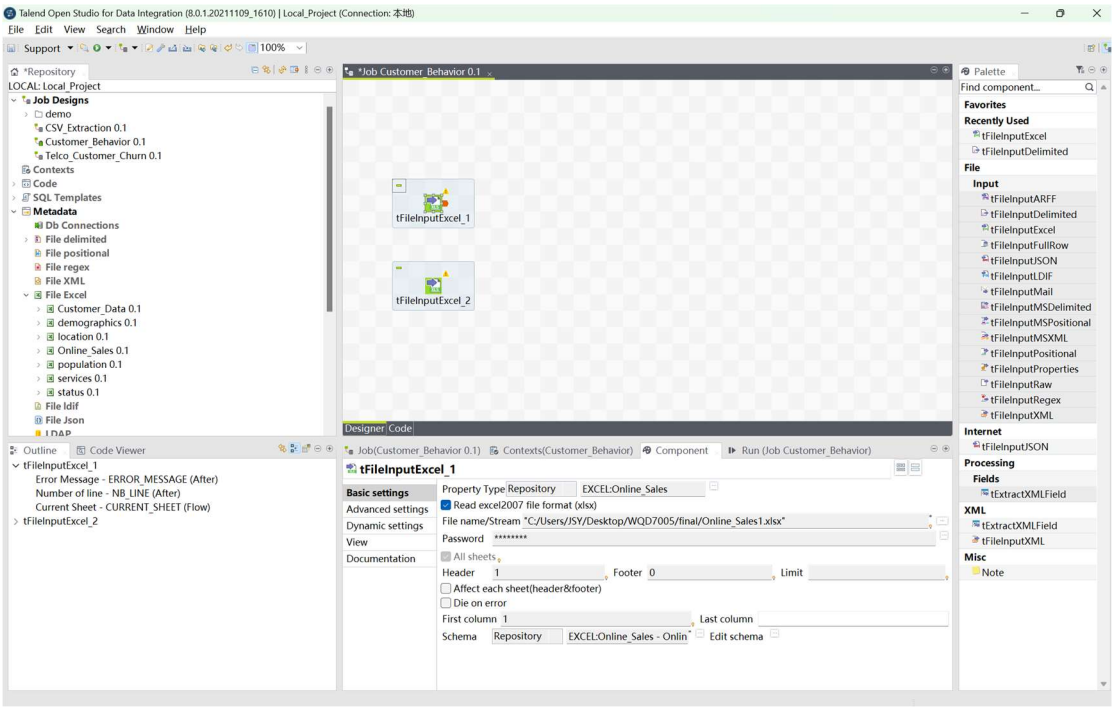




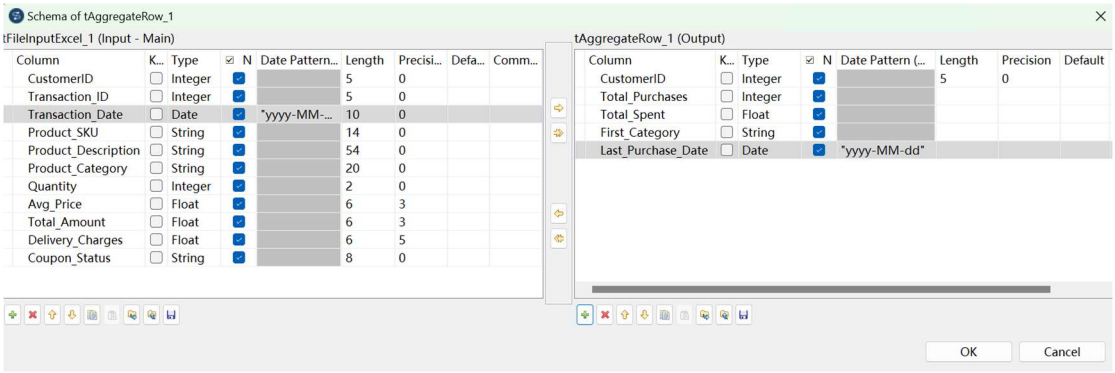


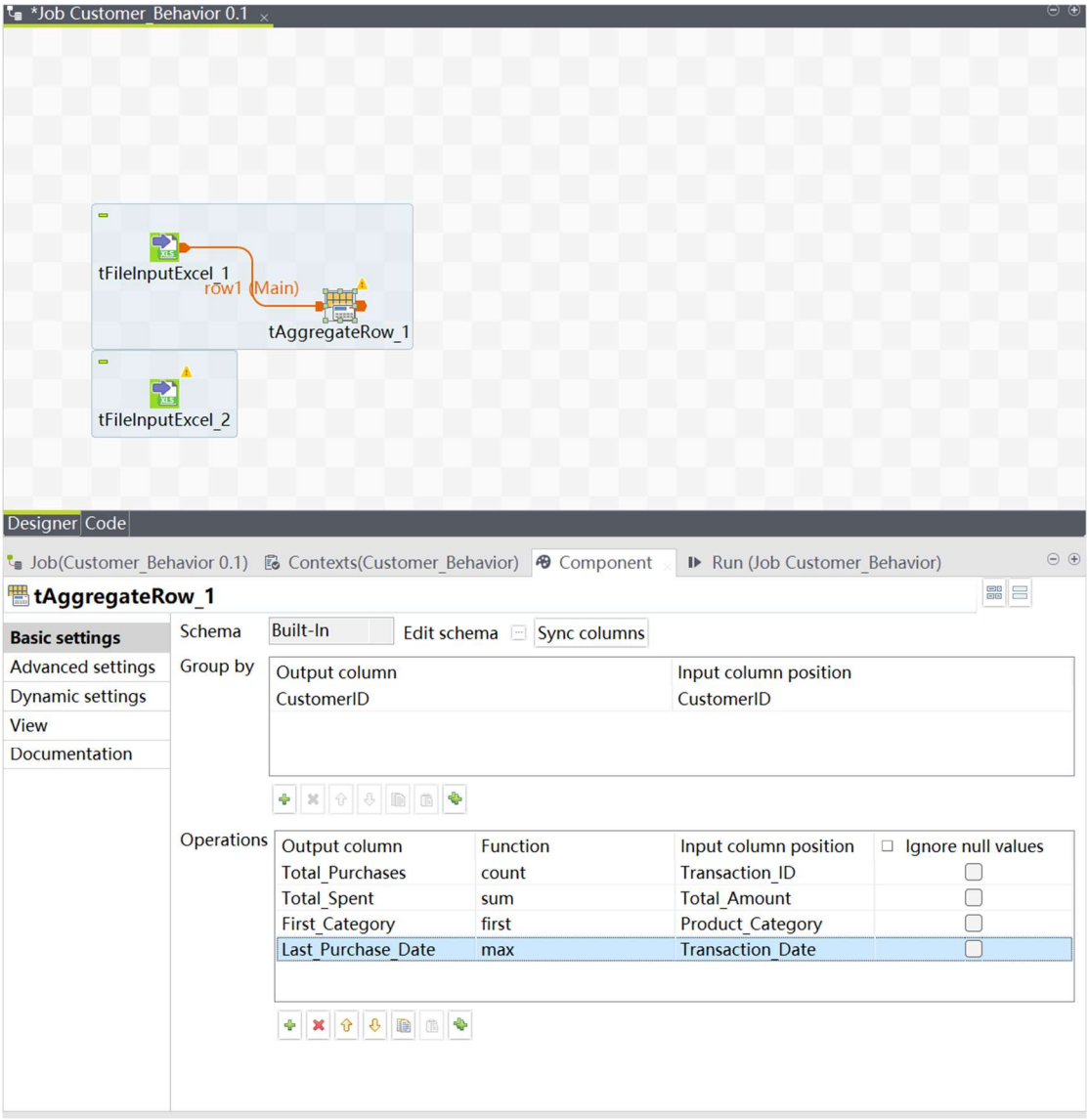


Create a new job named ‘Customer_Behavior’ and load Online_Sales and Customer_Data files in the tFileInputExcel components:

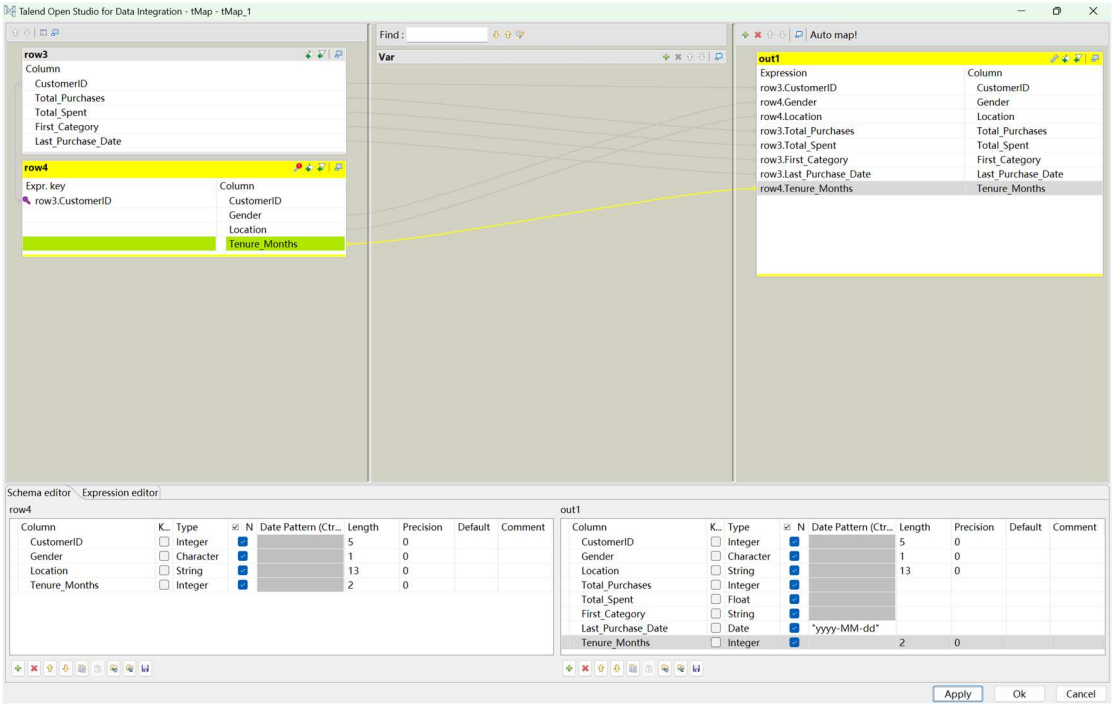


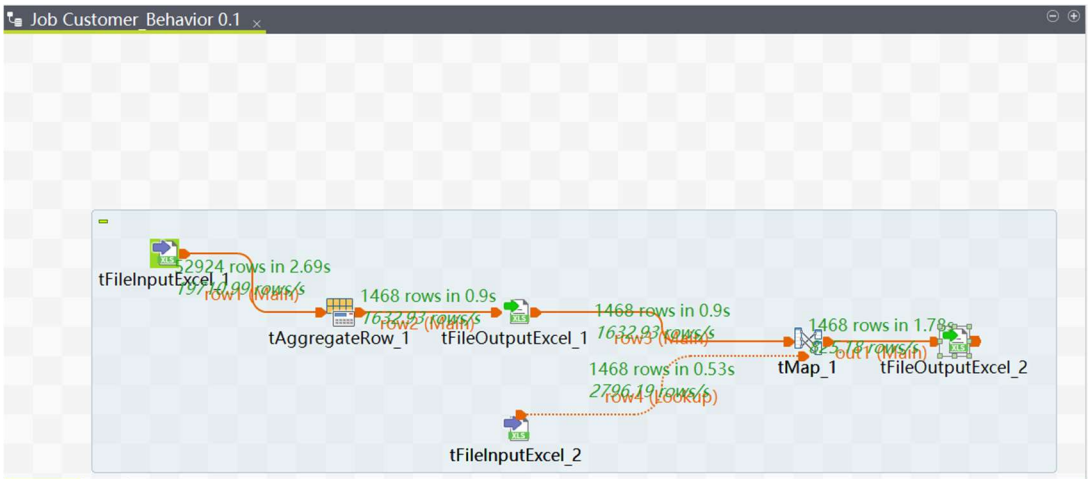
Add columns of Total_Purchases, Total_Spent, First_Category, Last_Purchased_Date, and get these data from count of TransactionID, sum of Total_Amount, first of Product_Category, and max of Transaction_Data, separately using Operations in tAggregateRow component:

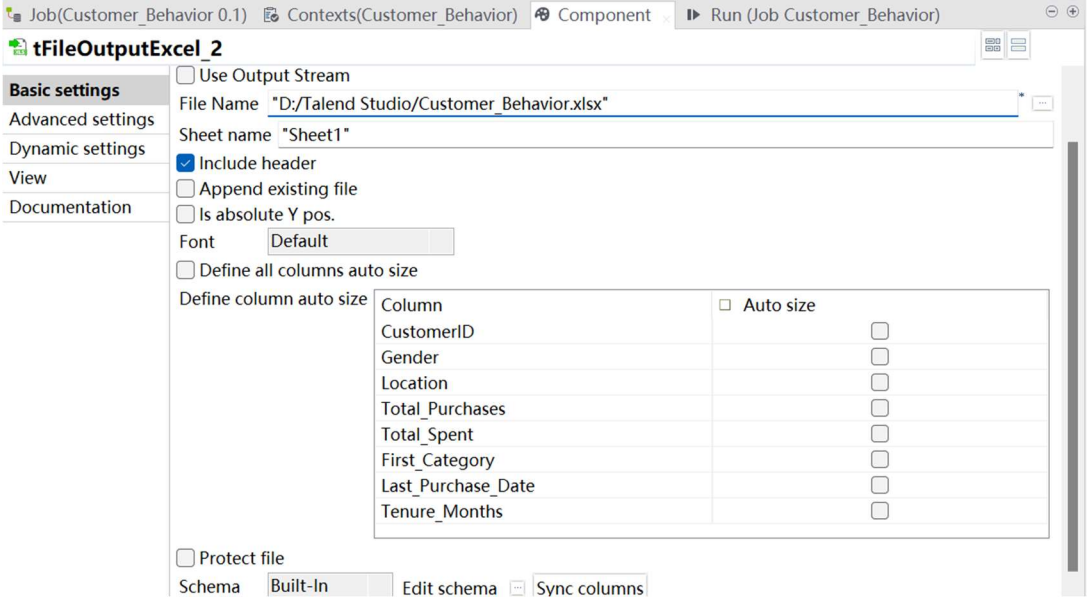




Join the output of Aggregation and the CustomerData file using tMap component, and run the whole working process, then we get one combined dataset named Customer_Behavior.xlsx, having 8 columns of CustomerID, Gender, Location, Total_Purchases, Total_Spent, First_Category, Last_Purchased_Date, and Tenure_Months, with 1468 customers' rows.


Talend Open Studio for Data Integration - tMap - tMap_1
row3
Column
CustomerID
Total_Purchases
Total_Spent
First_Category
Last_Purchase_Date
row4
Expr. key
row3.CustomerID
Column
CustomerID
Gender
Location
Tenure_Months
Find :
Var
Auto map!
out1
Expression
Column
row3.CustomerID
CustomerID
row4.Gender
Gender
row4.Location
Location
row3.Total_Purchases
Total_Purchases
row3.Total_Spent
Total_Spent
row3.First_Category
First_Category
row3.Last_Purchase_Date
Last_Purchase_Date
row4.Tenure_Months
Tenure_Months
Schema editor
Expression editor
row4
out1
Column
K...
Type
N
Date Pattern (Ctr...
Length
Precision
Default
Comment
CustomerID
Integer
5
0
Gender
Character
1
0
Location
String
13
0
Tenure_Months
Integer
2
0
Total_Purchases
Integer
Total_Spent
Float
First_Category
String
Last_Purchase_Date
Date
"yyyy-MM-dd"
Apply
Ok
Cancel


Job Customer_Behavior 0.1
tFileInputExcel_1
52924 rows in 2.69s
row1 (Main)
tAggregateRow_1
1468 rows in 0.9s
row2 (Main)
tFileOutputExcel_1
1468 rows in 0.9s
row3 (Main)
tMap_1
1468 rows in 1.78s
out1 (Main)
tFileOutputExcel_2
1468 rows in 0.53s
row4 (Lookup)
tFileInputExcel_2
Designer
Code


Job(Customer_Behavior 0.1)
Contexts(Customer_Behavior)
Component
Run (Job Customer_Behavior)
tFileOutputExcel_2
Basic settings
Advanced settings
Dynamic settings
View
Documentation
Use Output Stream
File Name "D:/Talend Studio/Customer_Behavior.xlsx"
Sheet name "Sheet1"
Include header
Append existing file
Is absolute Y pos.
Font Default
Define all columns auto size
Define column auto size
Column
Auto size
CustomerID
Gender
Location
Total_Purchases
Total_Spent
First_Category
Last_Purchase_Date
Tenure_Months
Protect file
Schema Built-In
Edit schema
Sync columns