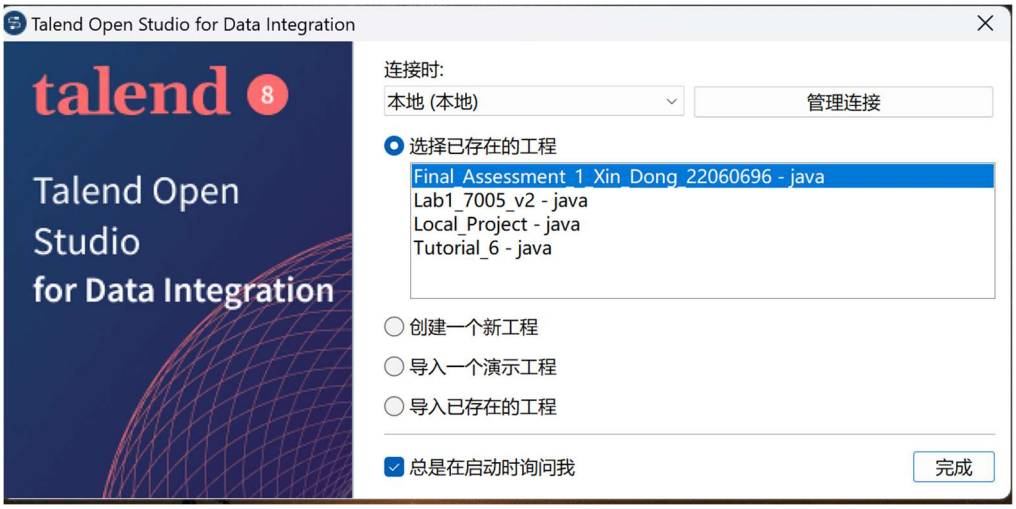
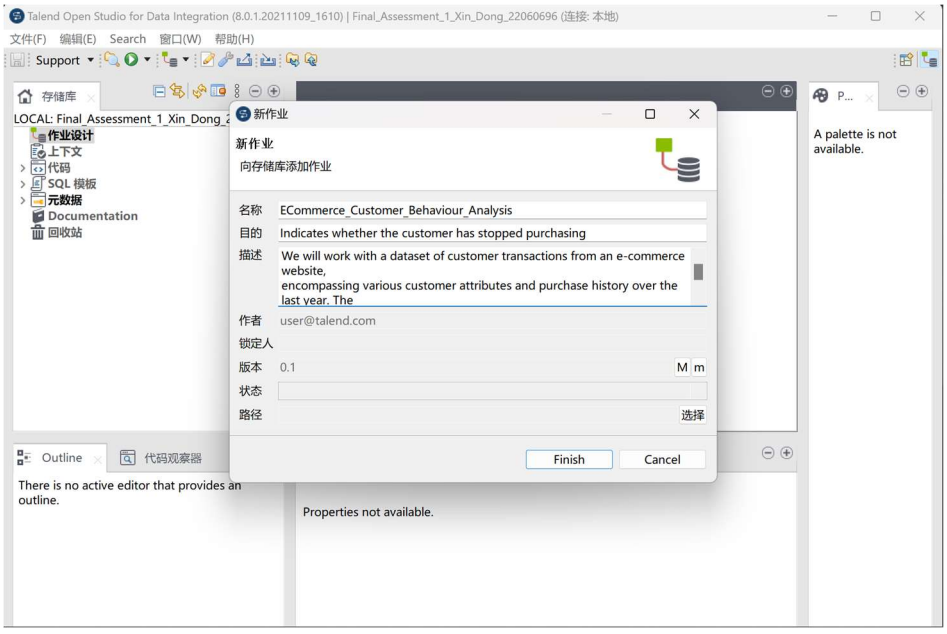
Talend Data Integration: Data Pre-Processing

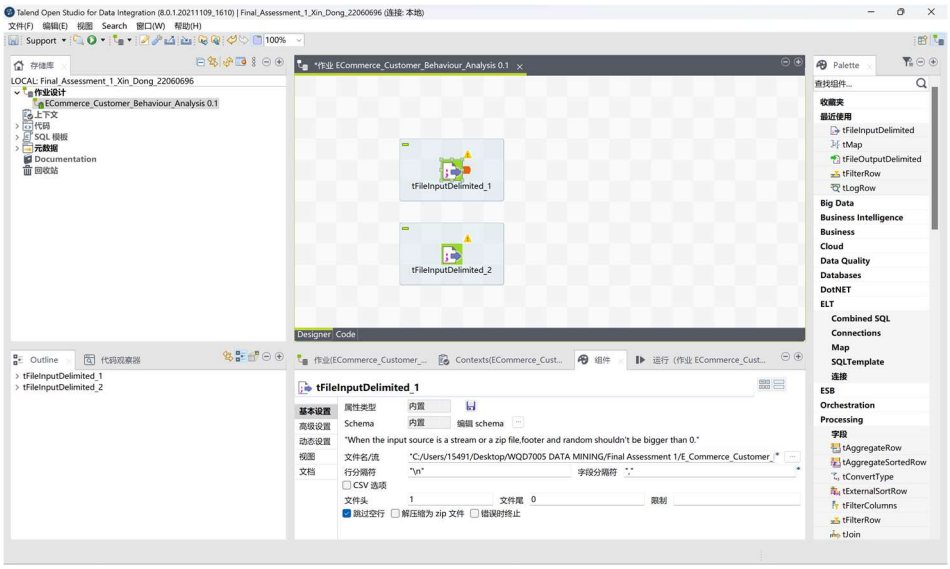
Talend Data Preparation:

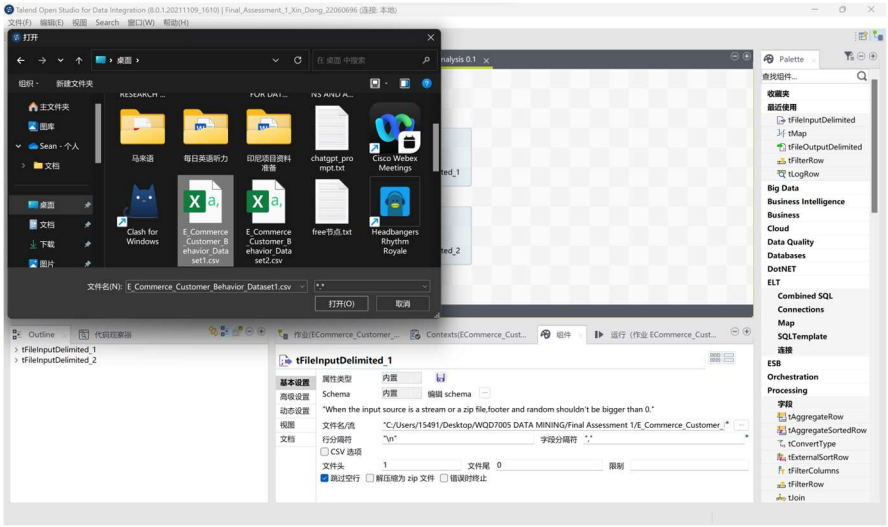
Create a project

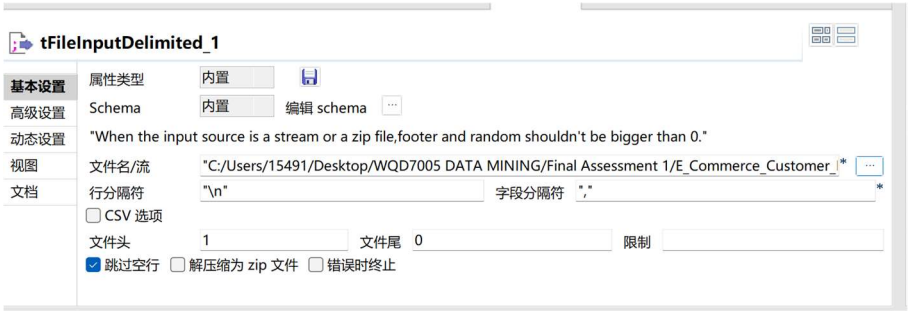


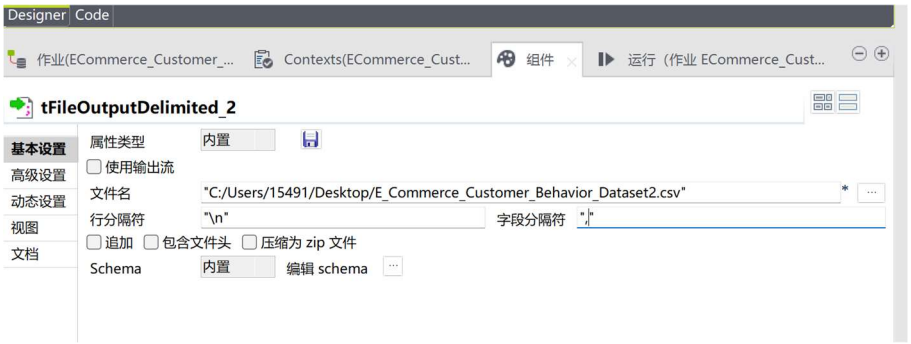


Import the 2 files to be integrated and set configuration:

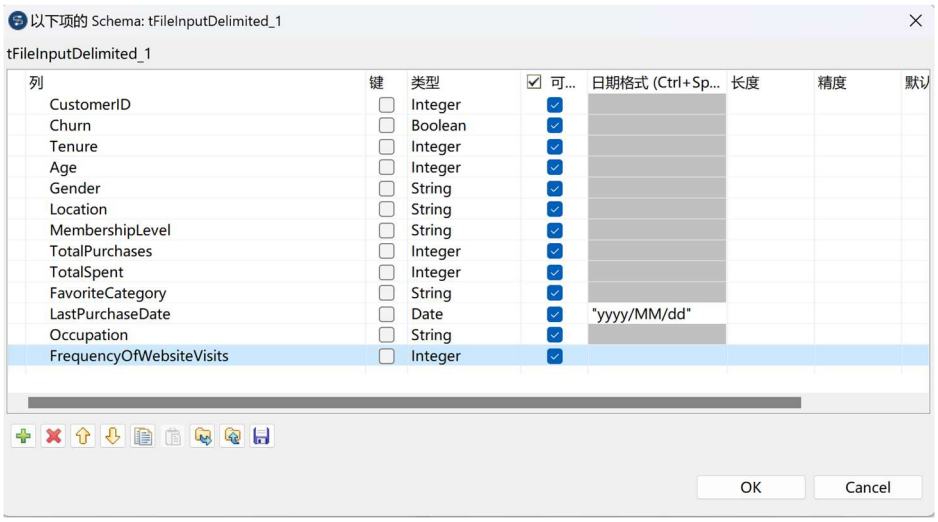

tFileInputDelimited_1
tFileInputDelimited_2
"When the input source is a stream or a zip file,footer and random shouldn't be bigger than 0."
"C:/Users/15491/Desktop/WQD7005 DATA MINING/Final Assessment 1/E_Commerce_Customer_"
"\n"
","

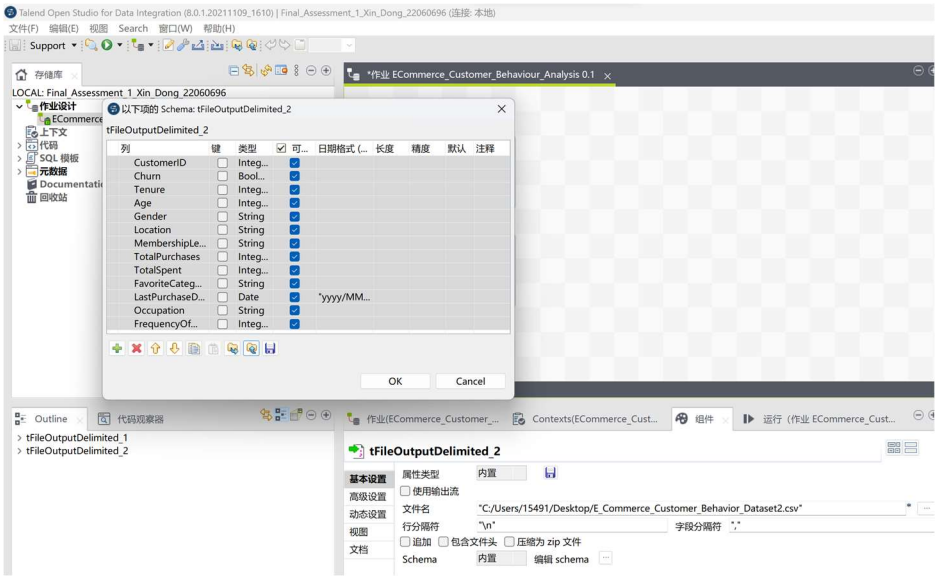

E_Commerce_Customer_Behavior_Dataset1.csv
tFileInputDelimited_1
tFileInputDelimited_2
"When the input source is a stream or a zip file,footer and random shouldn't be bigger than 0."


tFileInputDelimited_1
基本设置
高级设置
动态设置
视图
文档
属性类型
内置
Schema
内置
编辑 schema
"When the input source is a stream or a zip file,footer and random shouldn't be bigger than 0."
文件名/流
"C:/Users/15491/Desktop/WQD7005 DATA MINING/Final Assessment 1/E_Commerce_Customer_"
行分隔符
"\n"
字段分隔符
","
CSV 选项
文件头
1
文件尾
0
限制
跳过空行
解压缩为 zip 文件
错误时终止

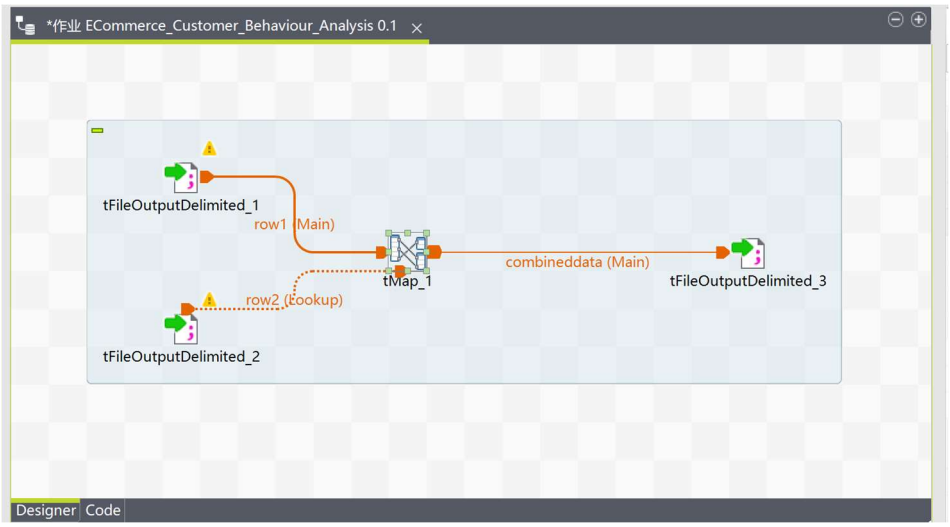


Set schema for the 2 files separately:

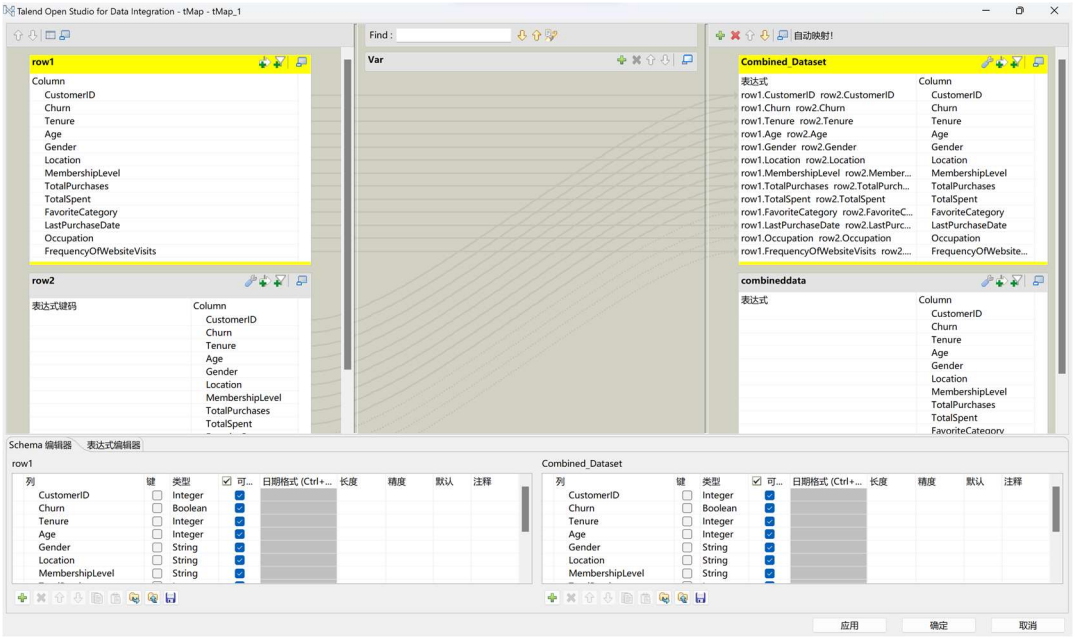




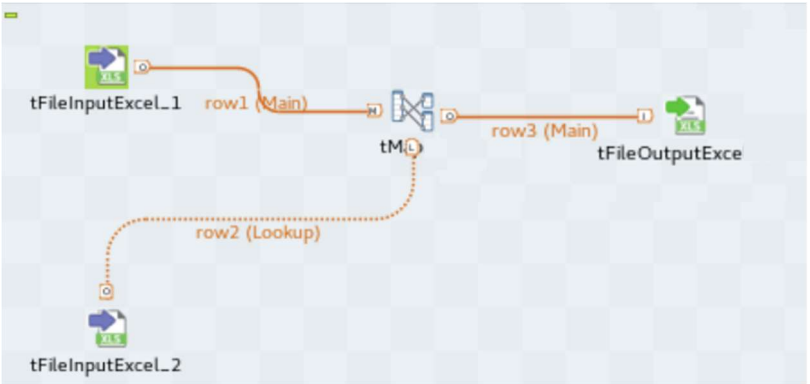
And then we put tMap and set tMap configuration, and finally set the tFileOutput.



tMap configuration:



We changed the name and get the final output pic:



So we successfully combined 2 files of dataset into one.