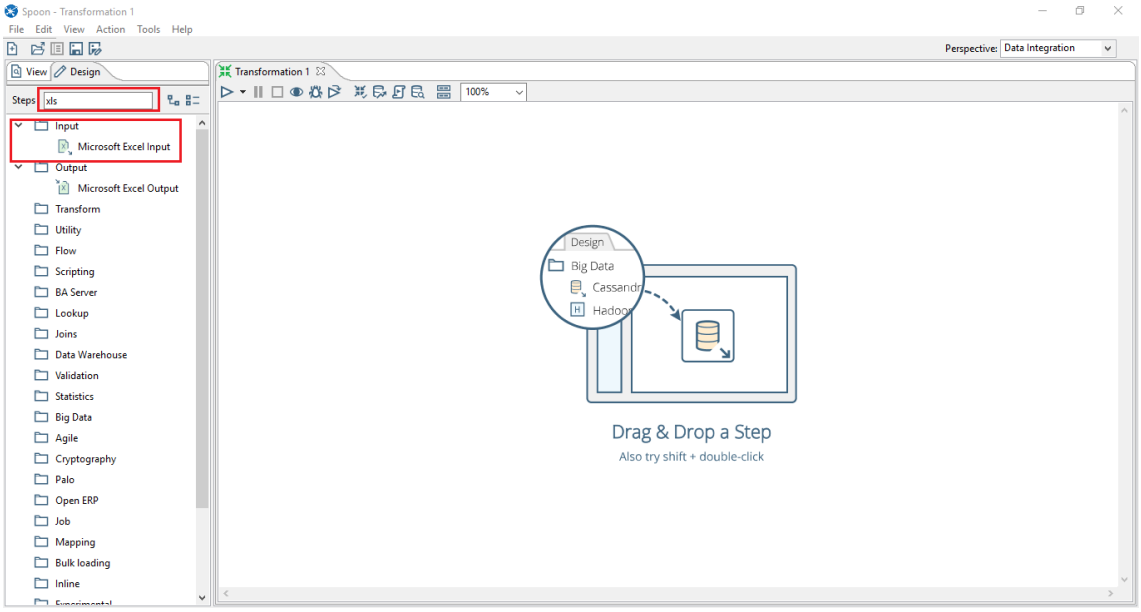
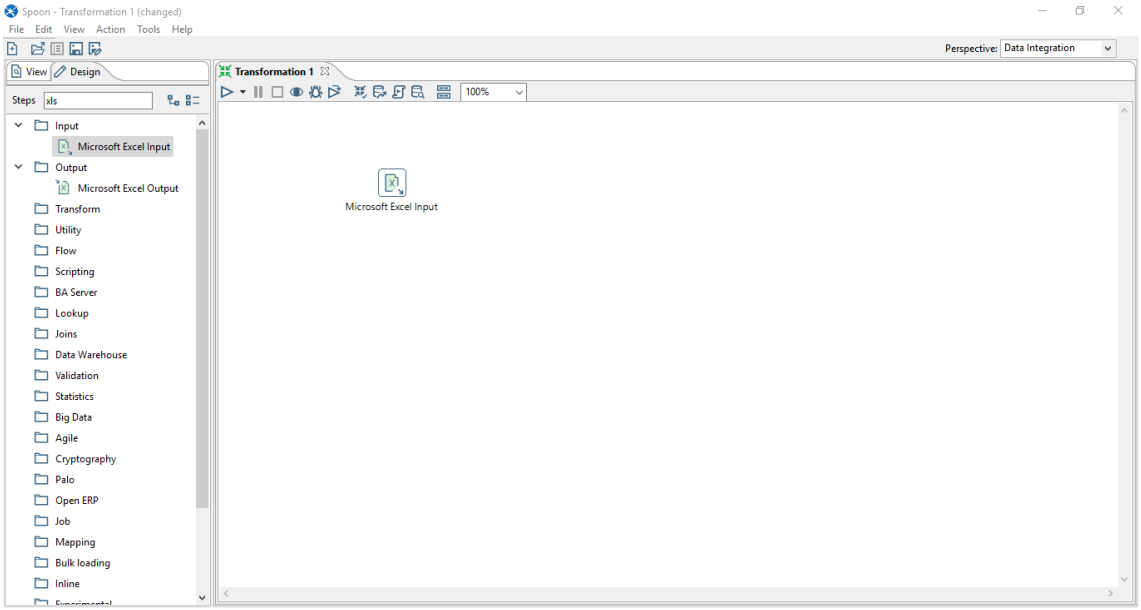
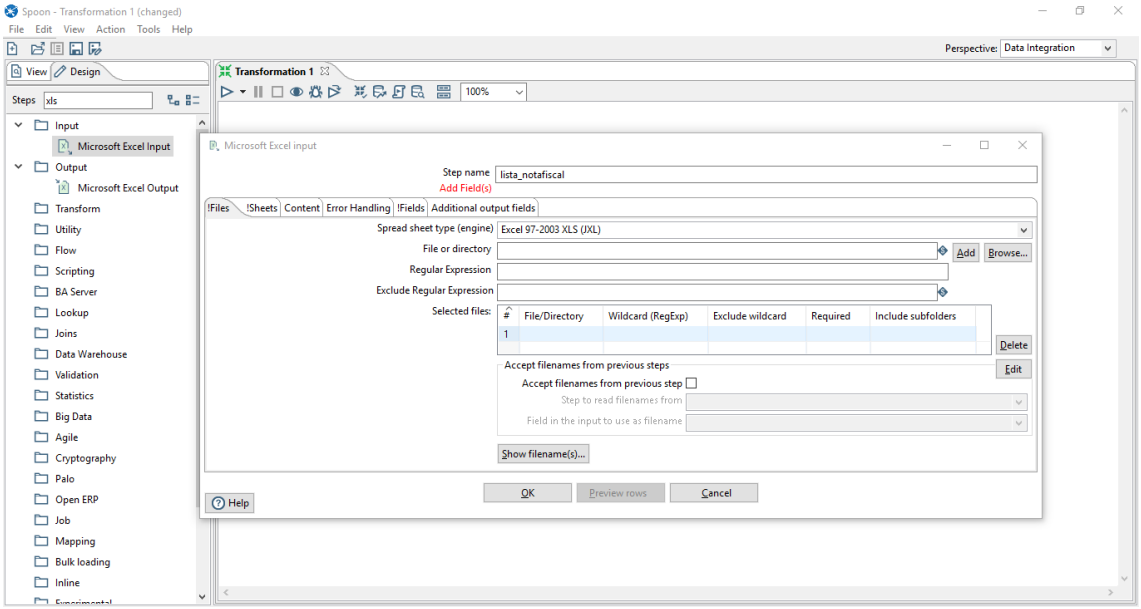
Tutorial - Realizando o ETL através do PDI Pentaho em um banco de dados relacional

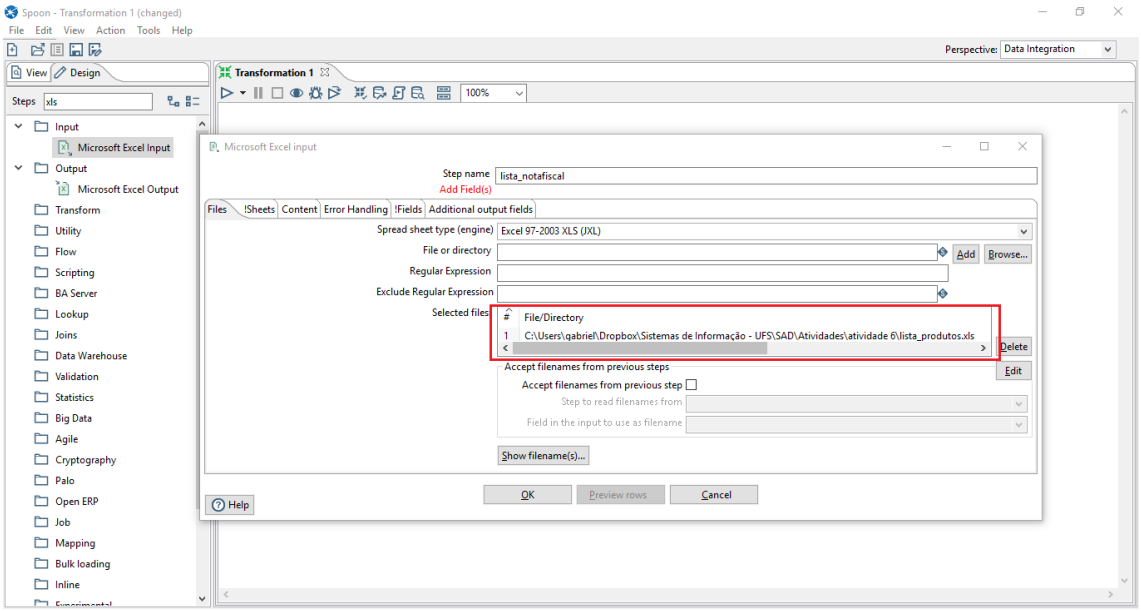
Passo 1: Selecionando o tipo de arquivo para a entrada de dados no formato xls (Excel):



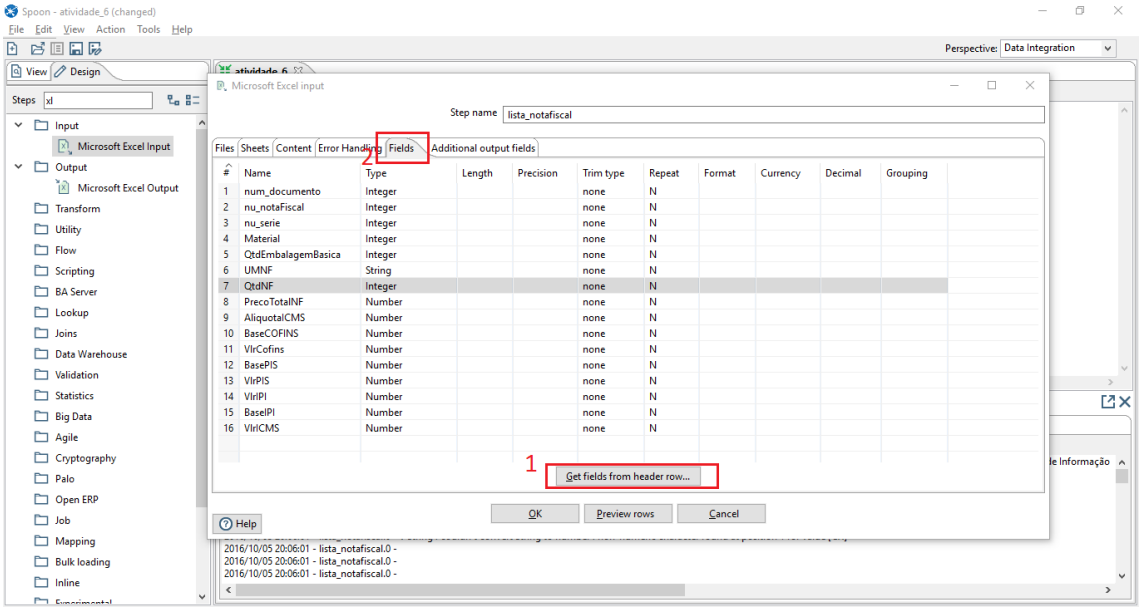


Passo 2: Configurar step de entrada para receber os dados da planilha

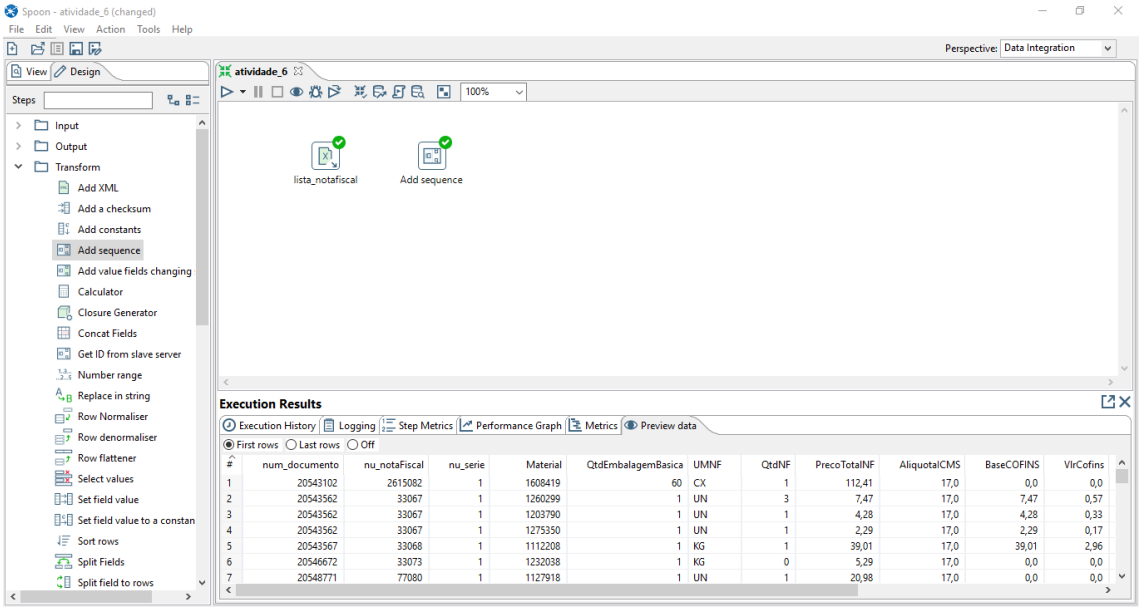




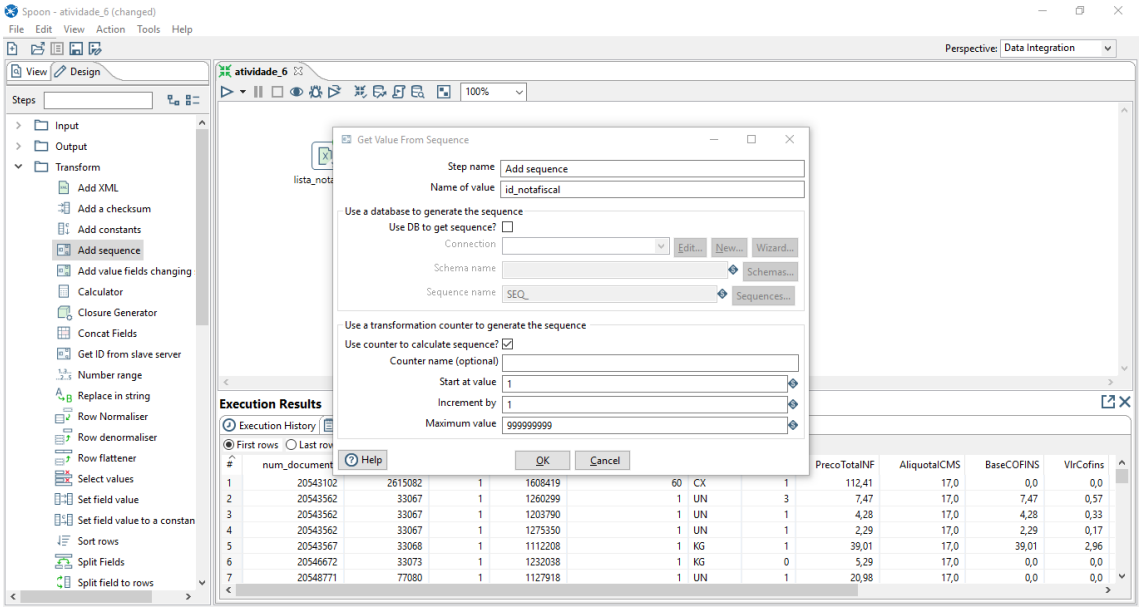
Ir na aba *Fields* para configurar o formato das colunas da planilha



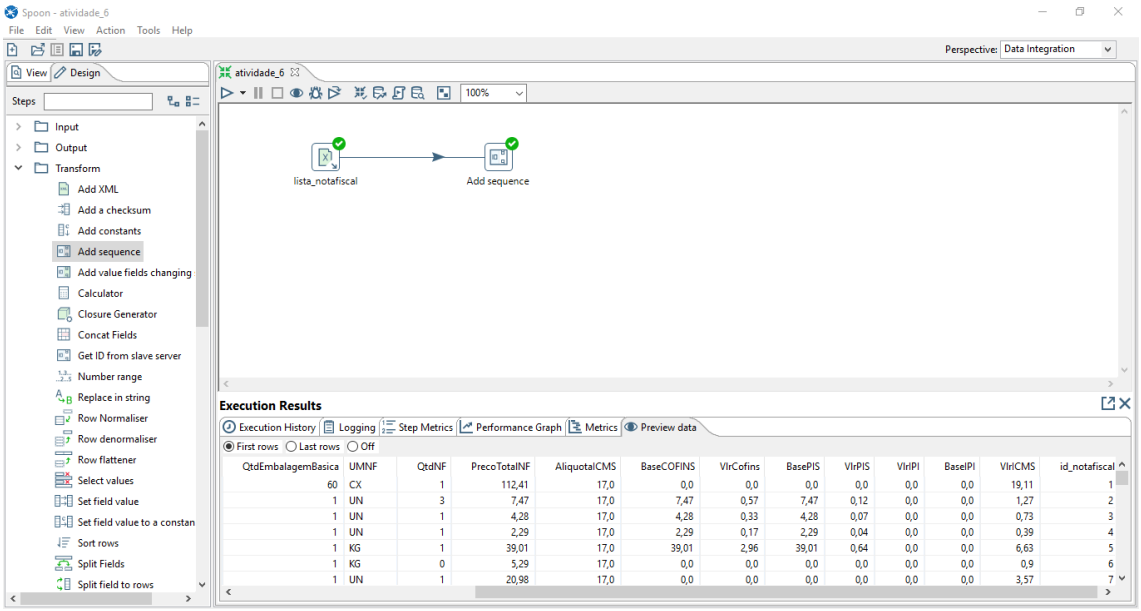
Passo 3: Adicionar um step de transformação *Add sequence*, para adicionar um colunar com números sequências



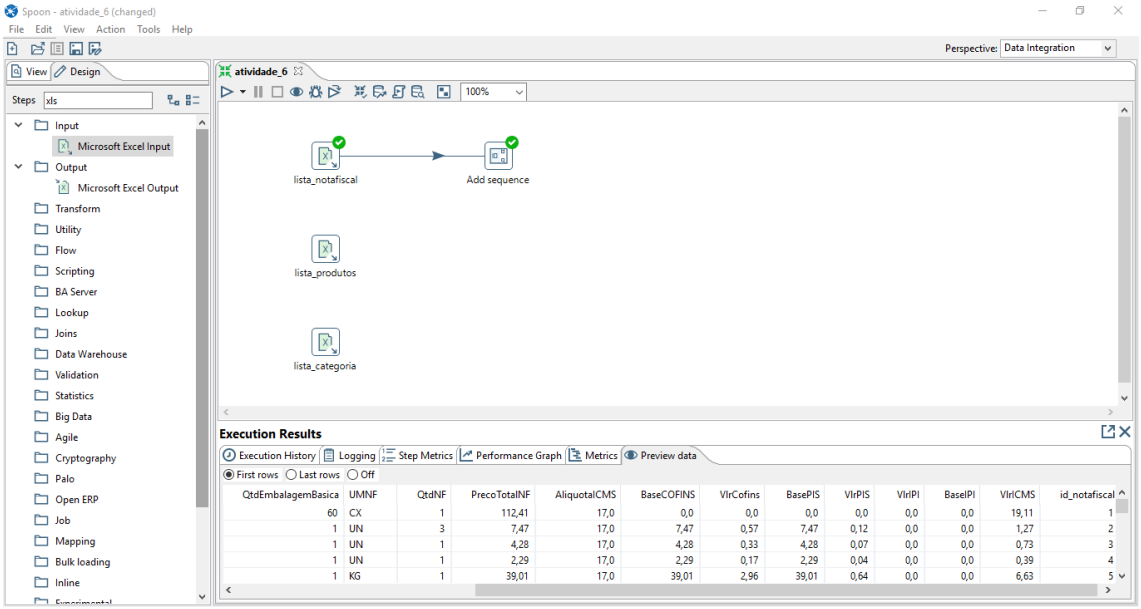
Definir um nome para a nova coluna



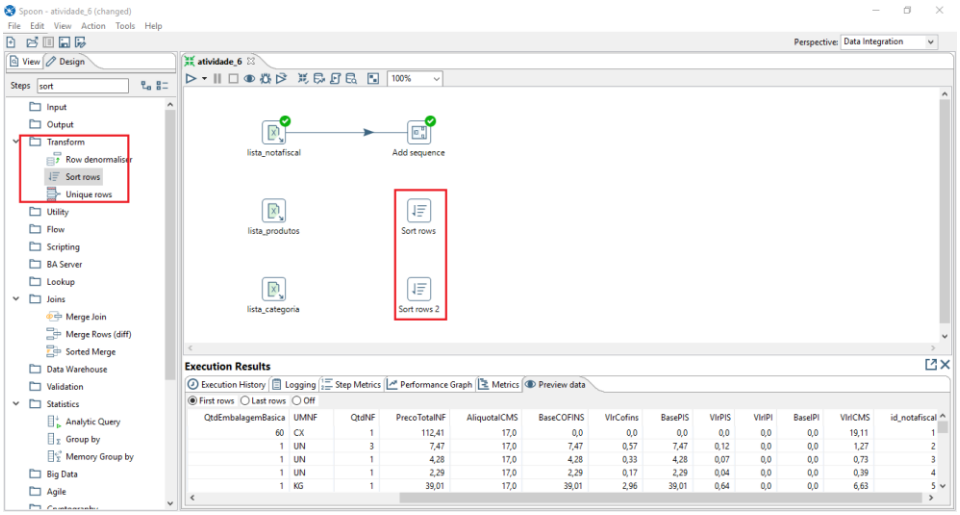
Estabelecer a conexão entres os steps

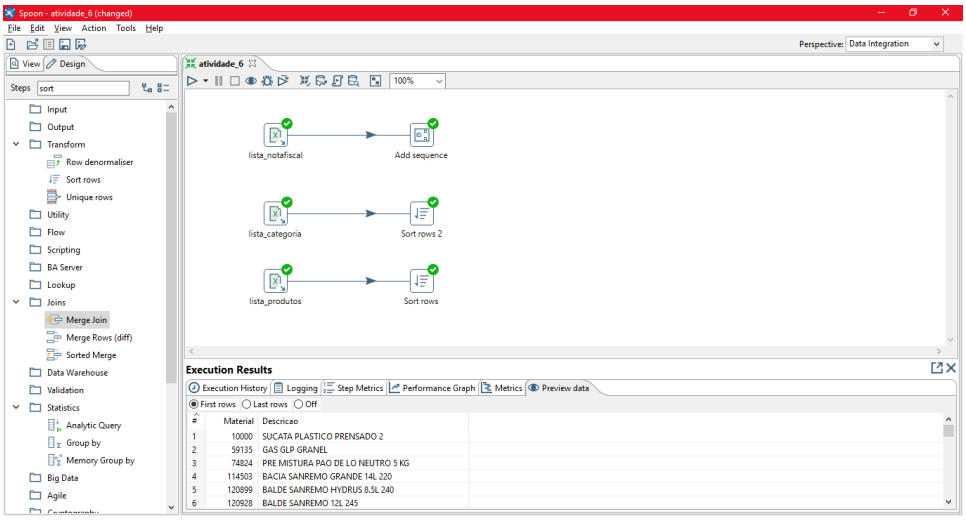


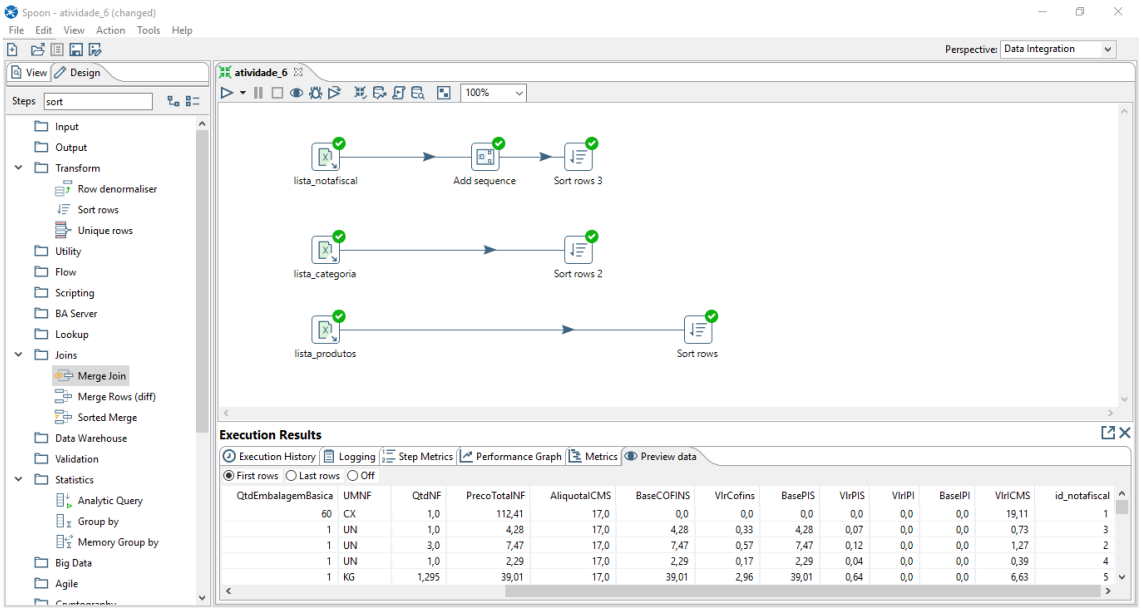
Adicionar os demais steps de entrada no formato de dados xls



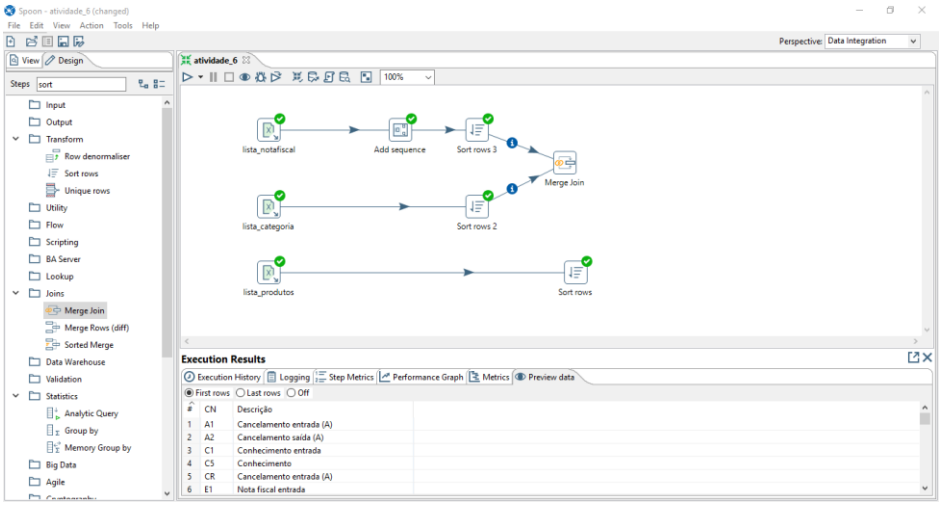
Passo 4: Adicionar os stpes de transformação para ordenar os dados da planilha

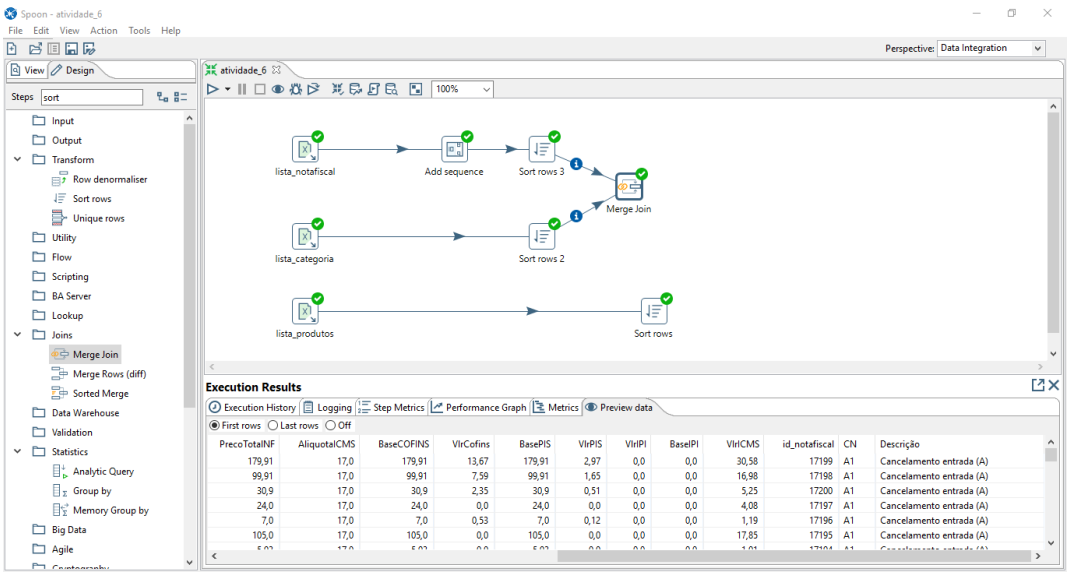




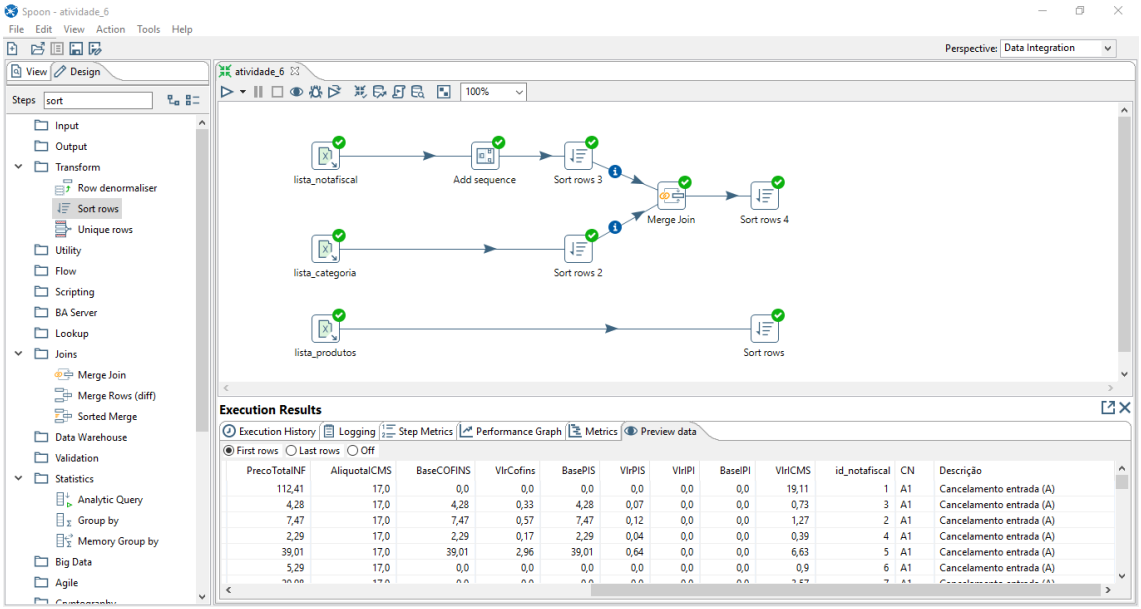


Passo 5: Realizar o step de *Merge Join* nas duas planilhas, lista_notafiscal e lista_categoria.

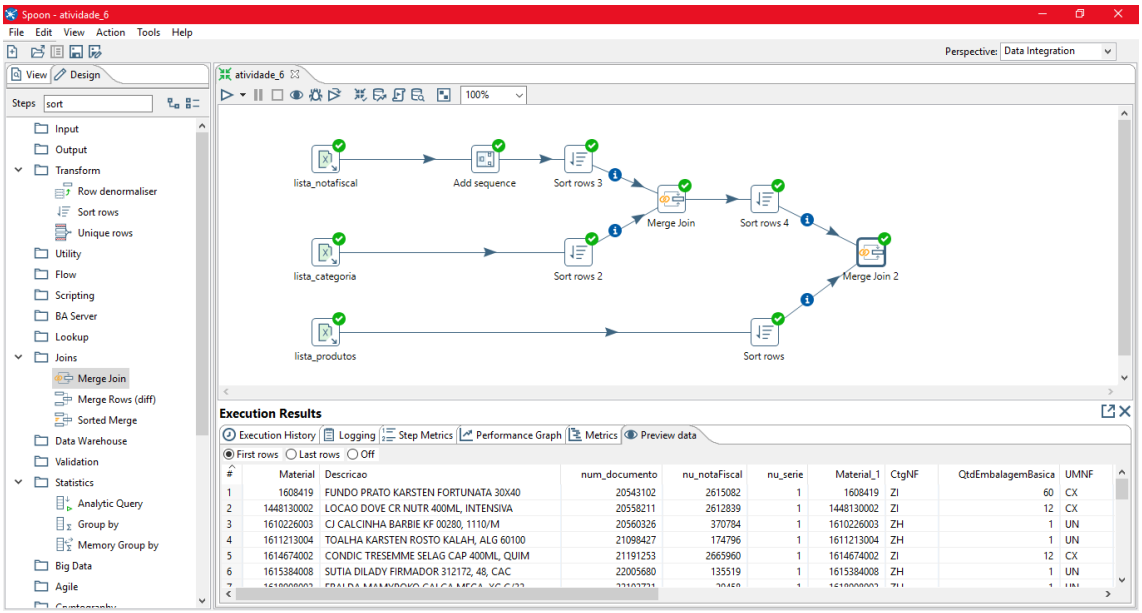




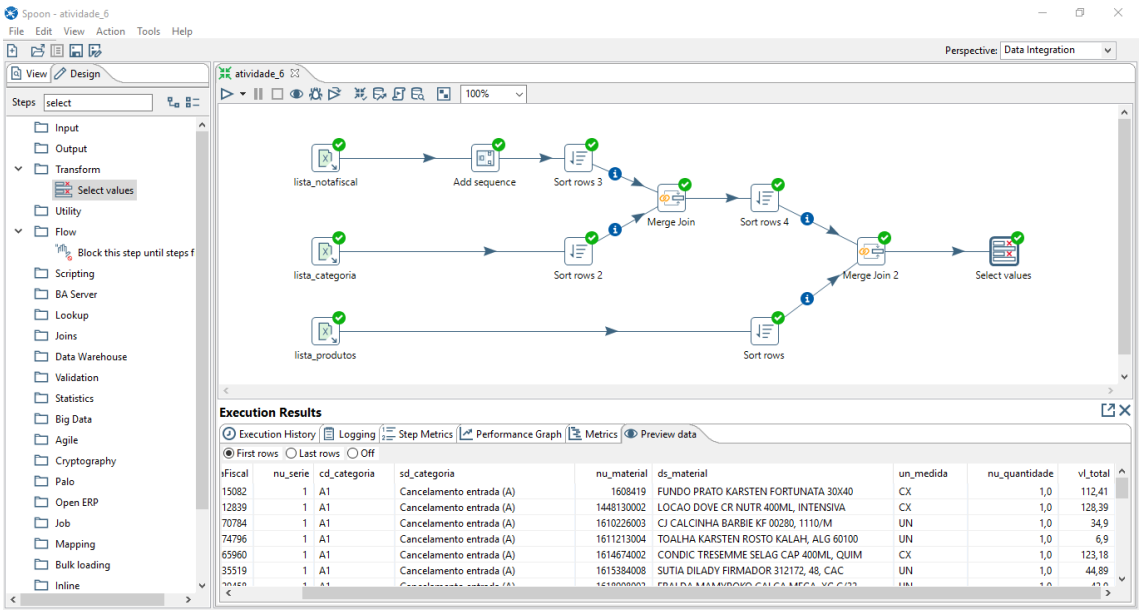
Passo 6: Realizar novamente a ordenação no resultado do step *Merge Join*



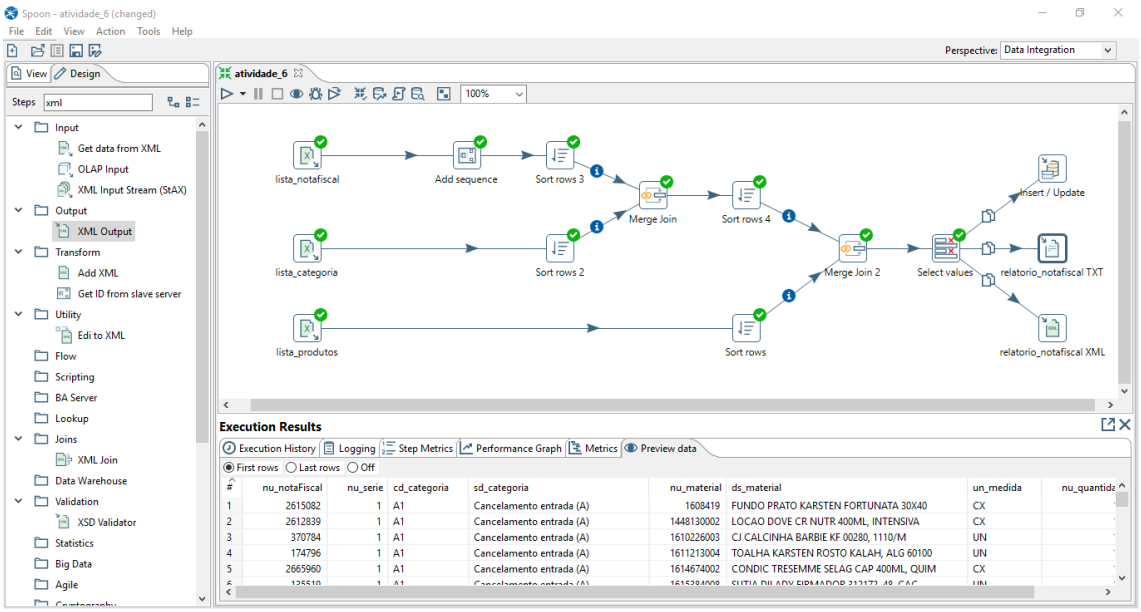
Passo 7: Realizar o *Merge Join* nas duas saídas ordenadas



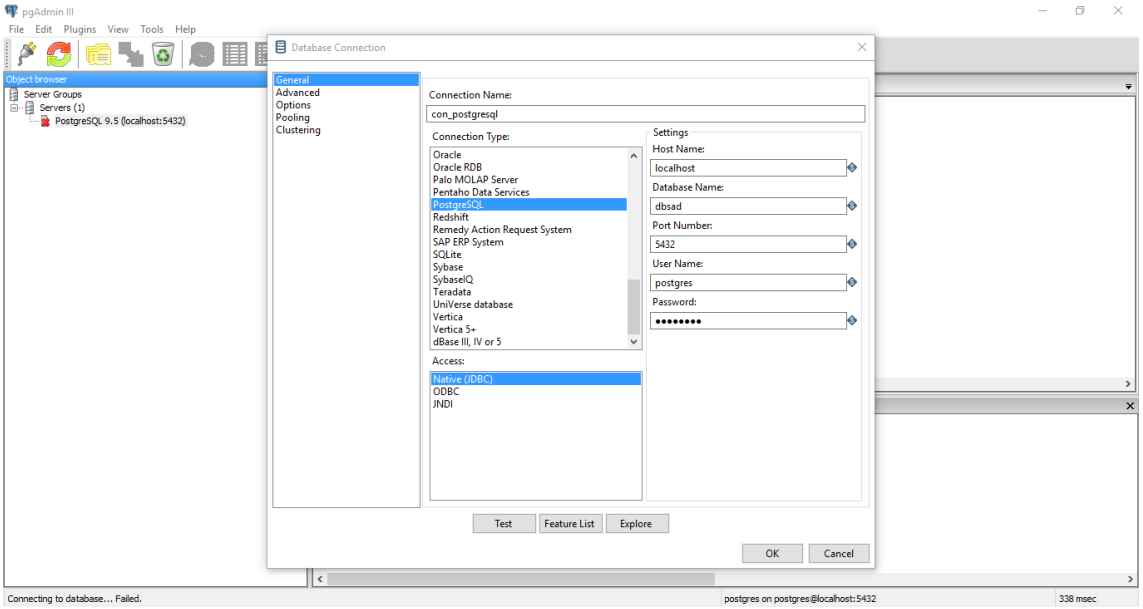
Passo 8: Realizar a transformação *Select values,* para selecionar as colunas que irá conter a nova saída

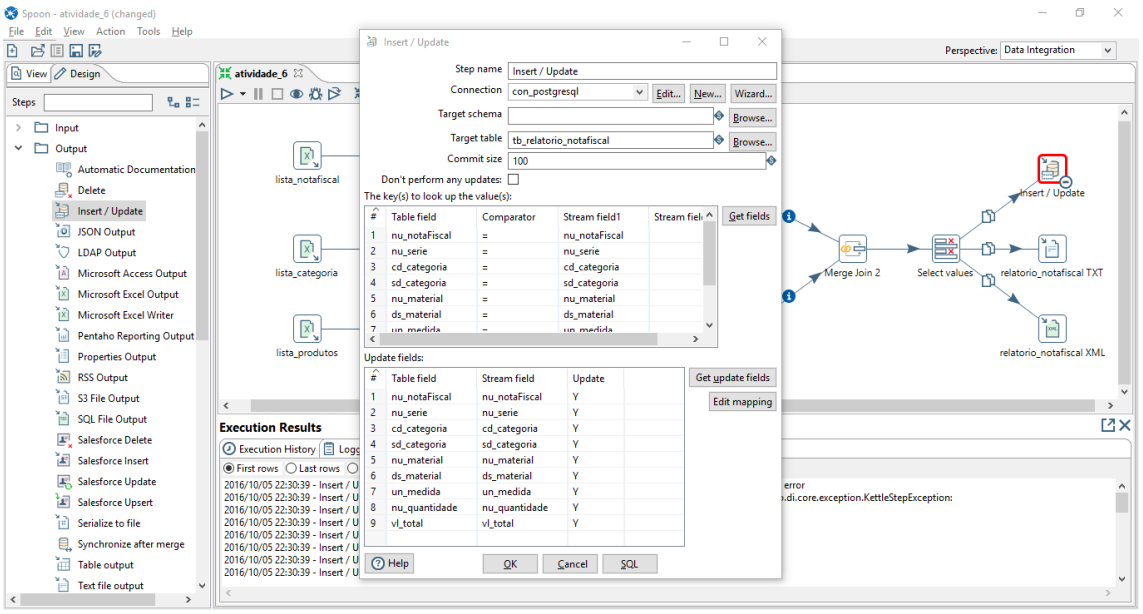


Passo 9: Adicionar os stpes de saída para os formatos XML, TXT e gerar uma carga no banco de dados PostgreSQL

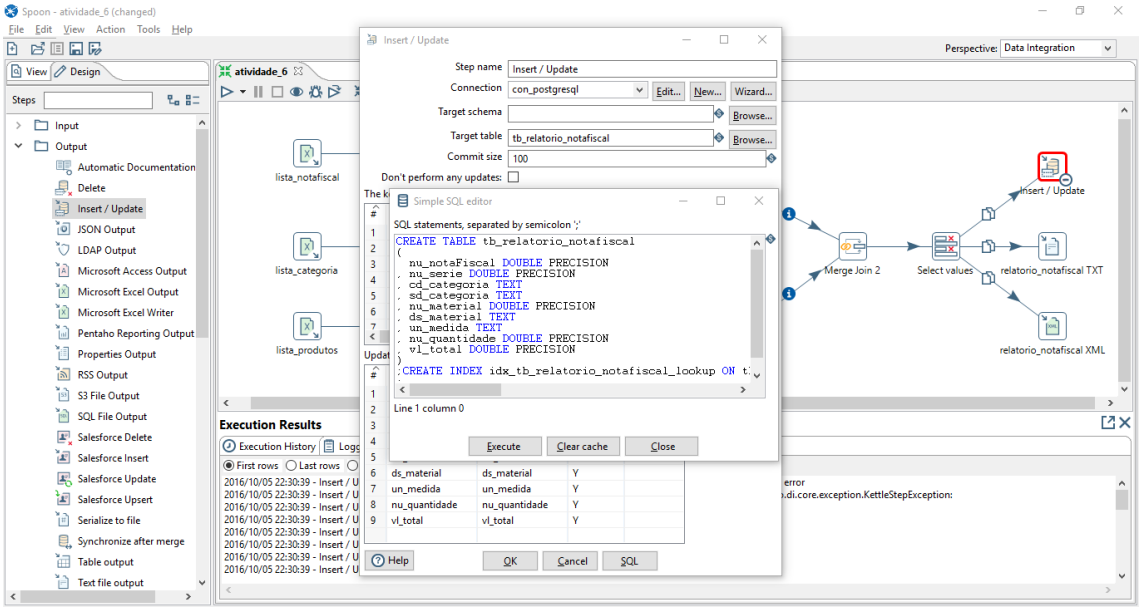


Passo 10: Configurar a conexão com o banco de dados





Criar a base e a tabela com o script gerado pela ferramenta



Executar os steps