**Sistemas de Apoio a Decisão - 2016.1**

**André Teixeira de Frades**

**Atividade 6 – 21/09/2016**

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

1-    Crie uma pequena documentação/tutorial de como realizar os seguintes tópicos (tire snapshots das telas e escreva de forma explicativa):

a)     Utilize, no mínimo três, arquivos no formato xls (Excel) para dar carga em um processo de ETL usando o PDI;

b)    Utilize os processos de transformação para tratar os dados de carga;

c)     Salve a saída em um banco de dados relacional de sua escolha (MySQL, PostgreSQL, etc..);

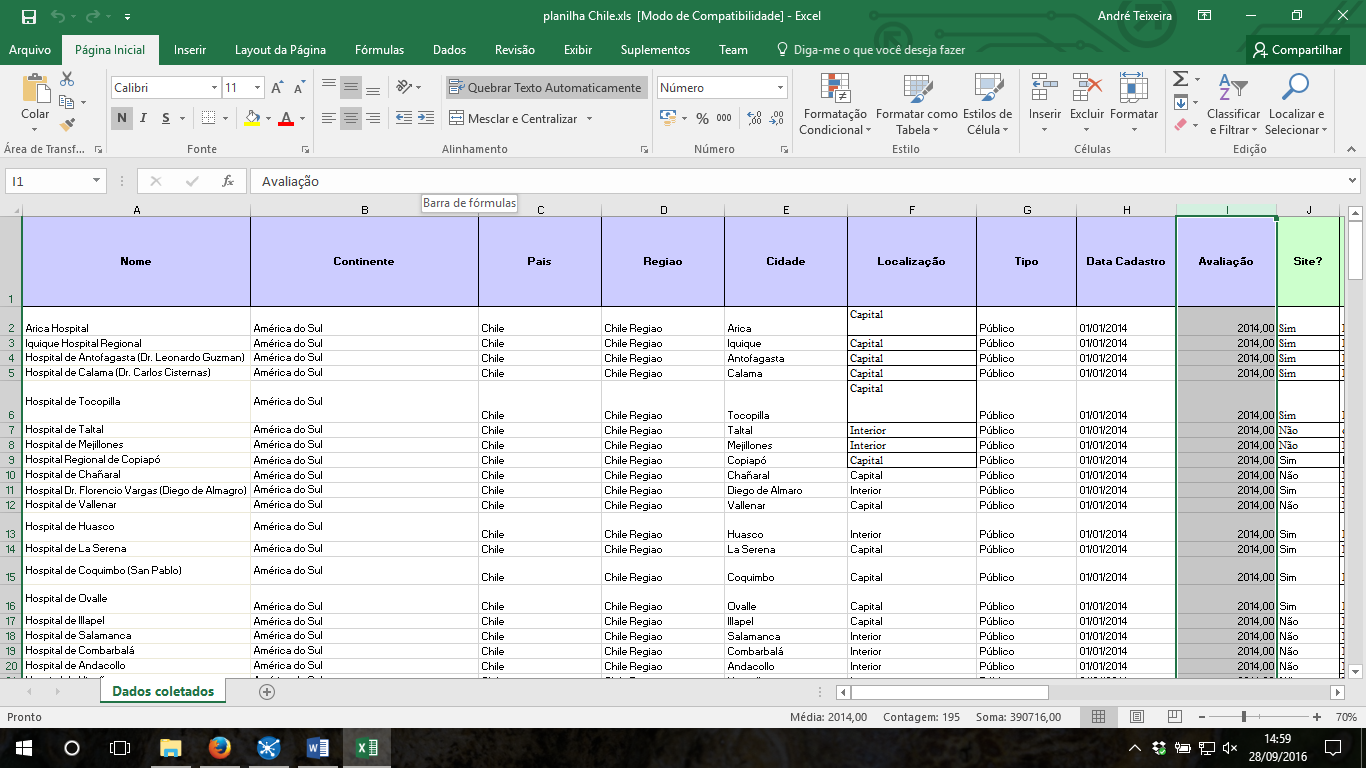
d)    Salve os arquivos em formato TXT;

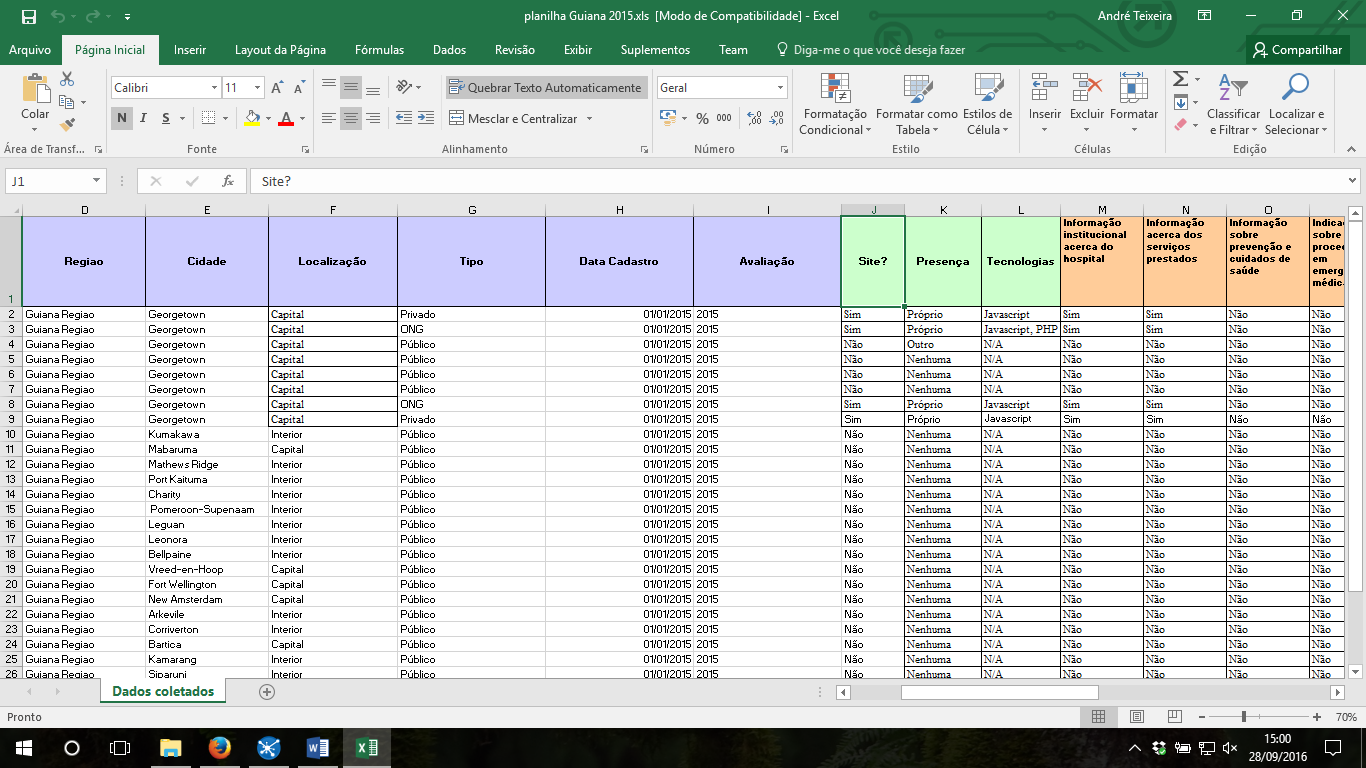
e)     Salve a saída em formato XML;

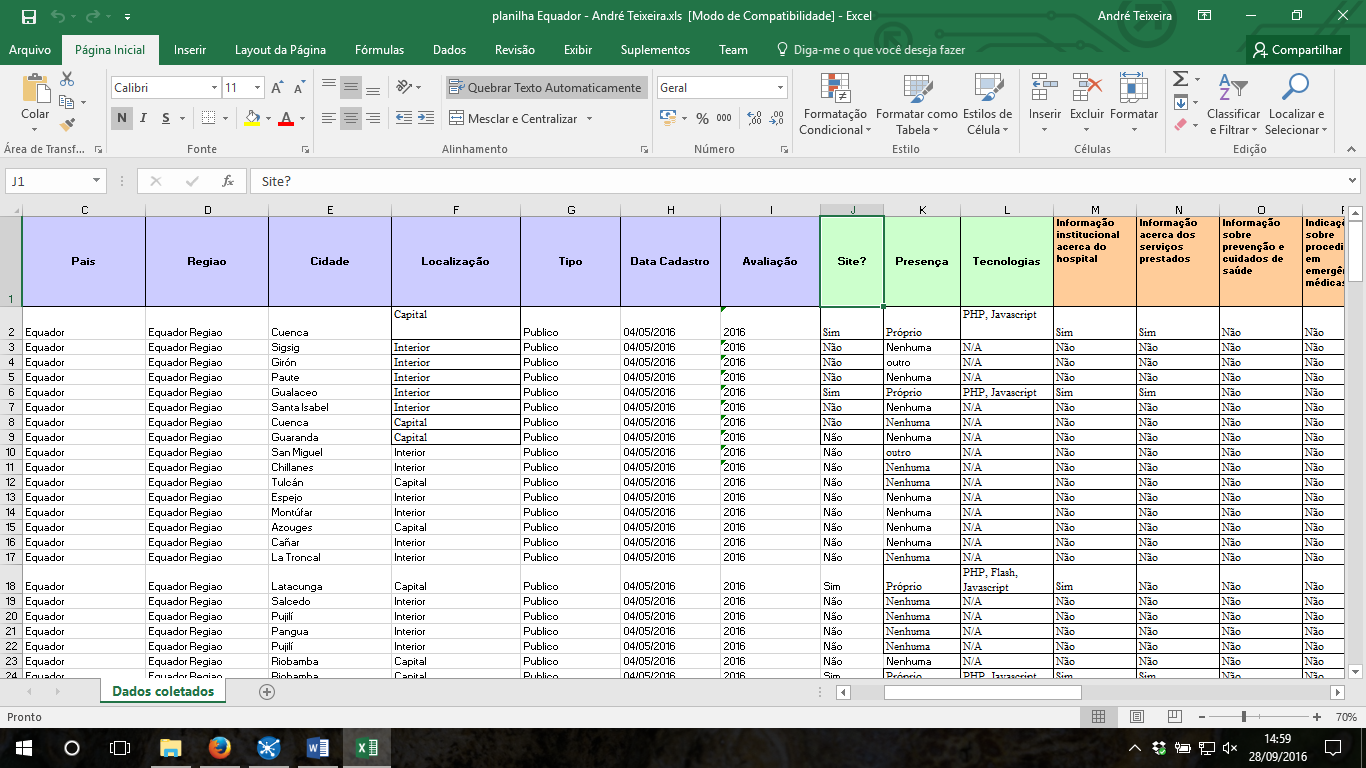
f)      Salve o arquivo do projeto no repositório do GIT;

g)     Envie o link para o SIGAA.

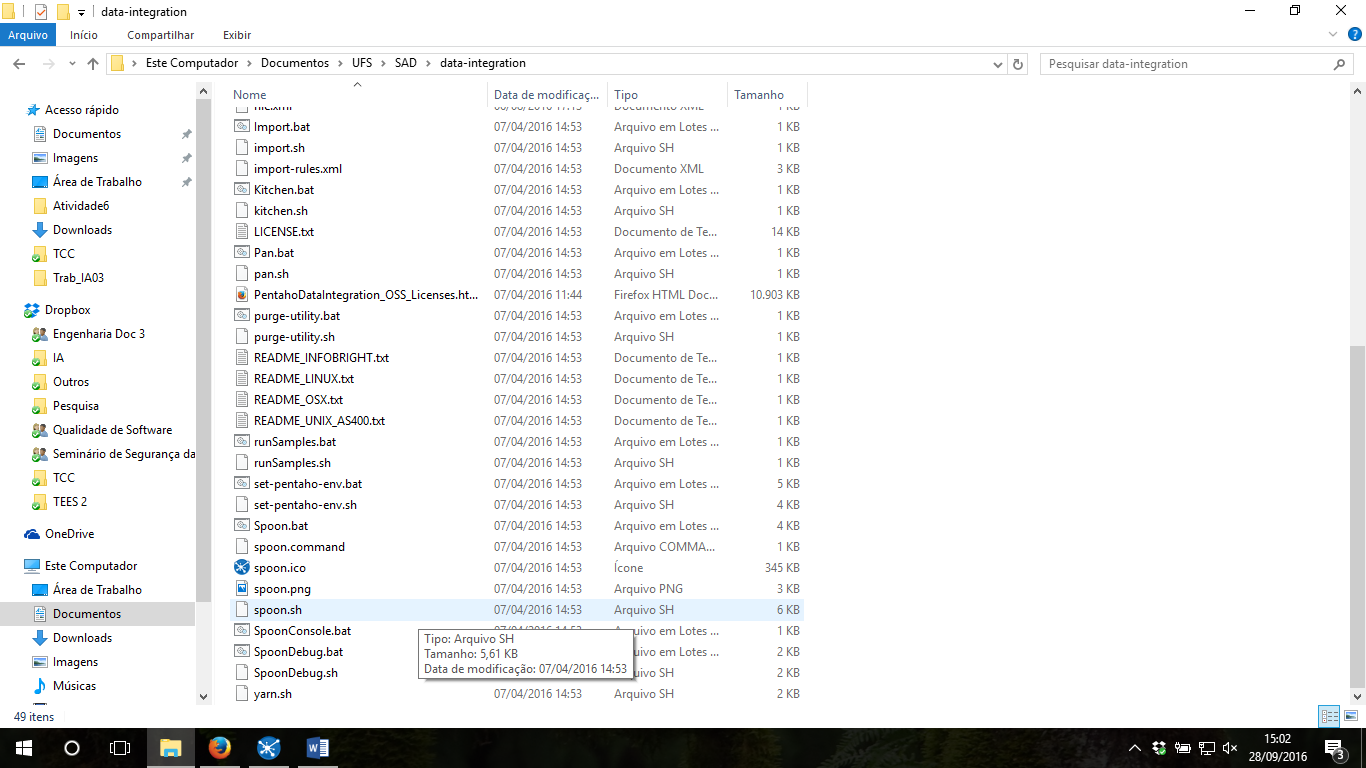
**a)** Arquivos de entrada escolhidos, planilhas Excel com dados de hospitais de diferentes países contendo os mesmos tipos de informações, deseja-se junta-los e fazer um tratamento para pequenas correções nos dados evitando inconsistências no banco:



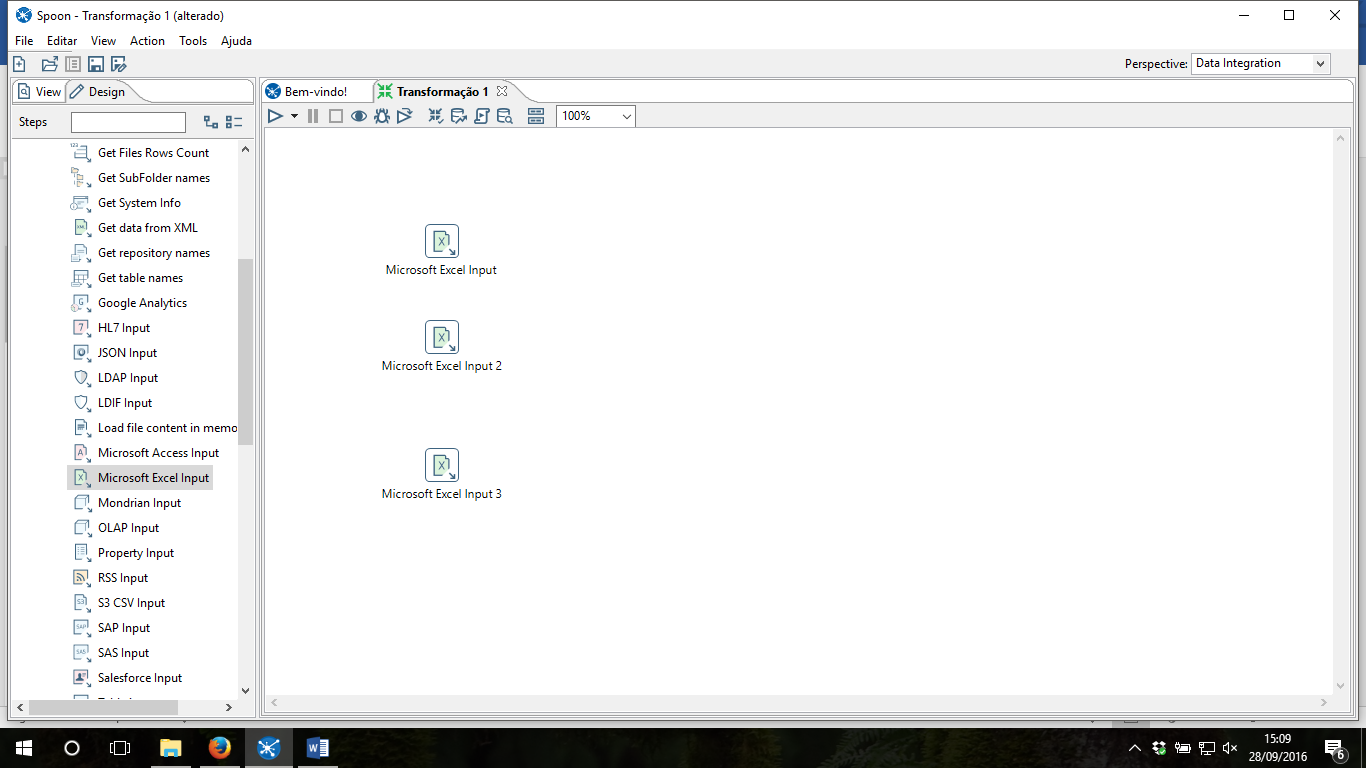




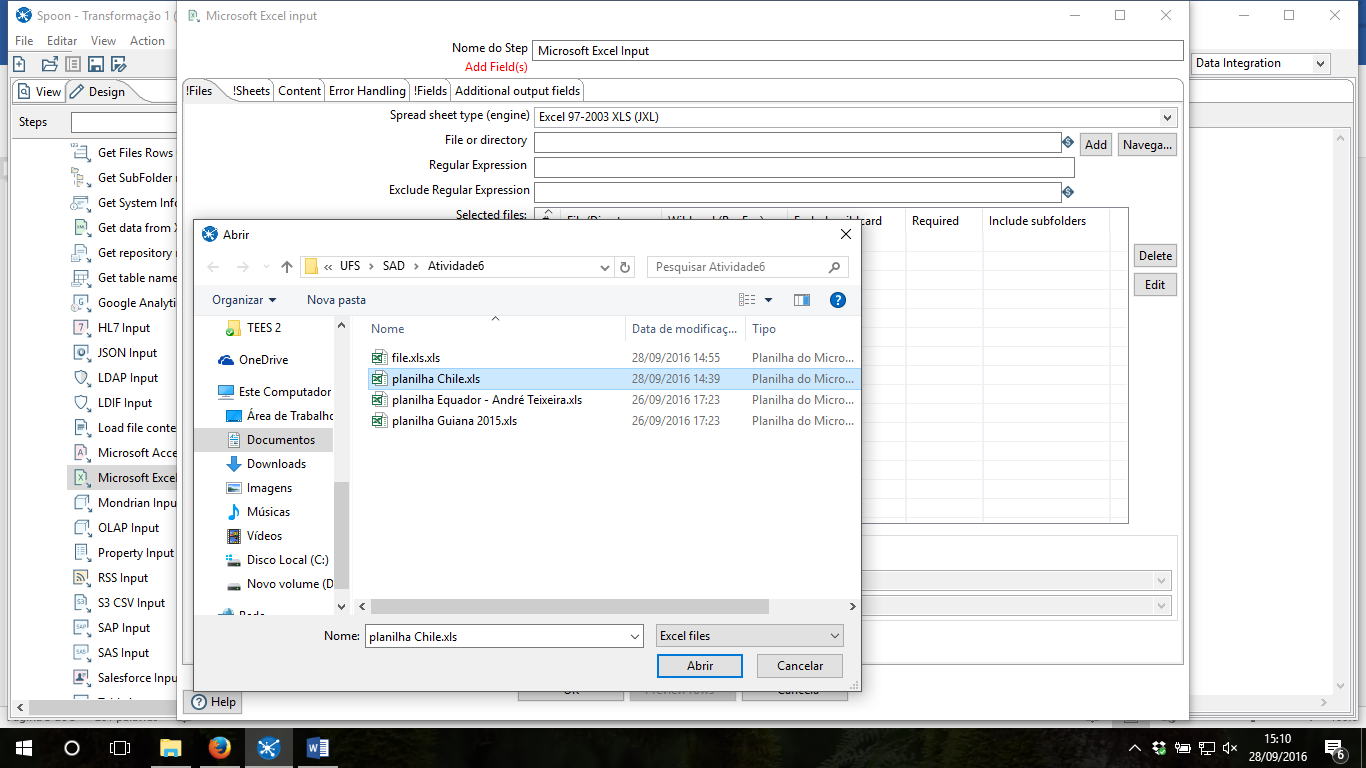
Em seguida inicie o Pentaho clicando em spoon.bat (Windows) ou via terminal de comando spoon.sh (Linux).



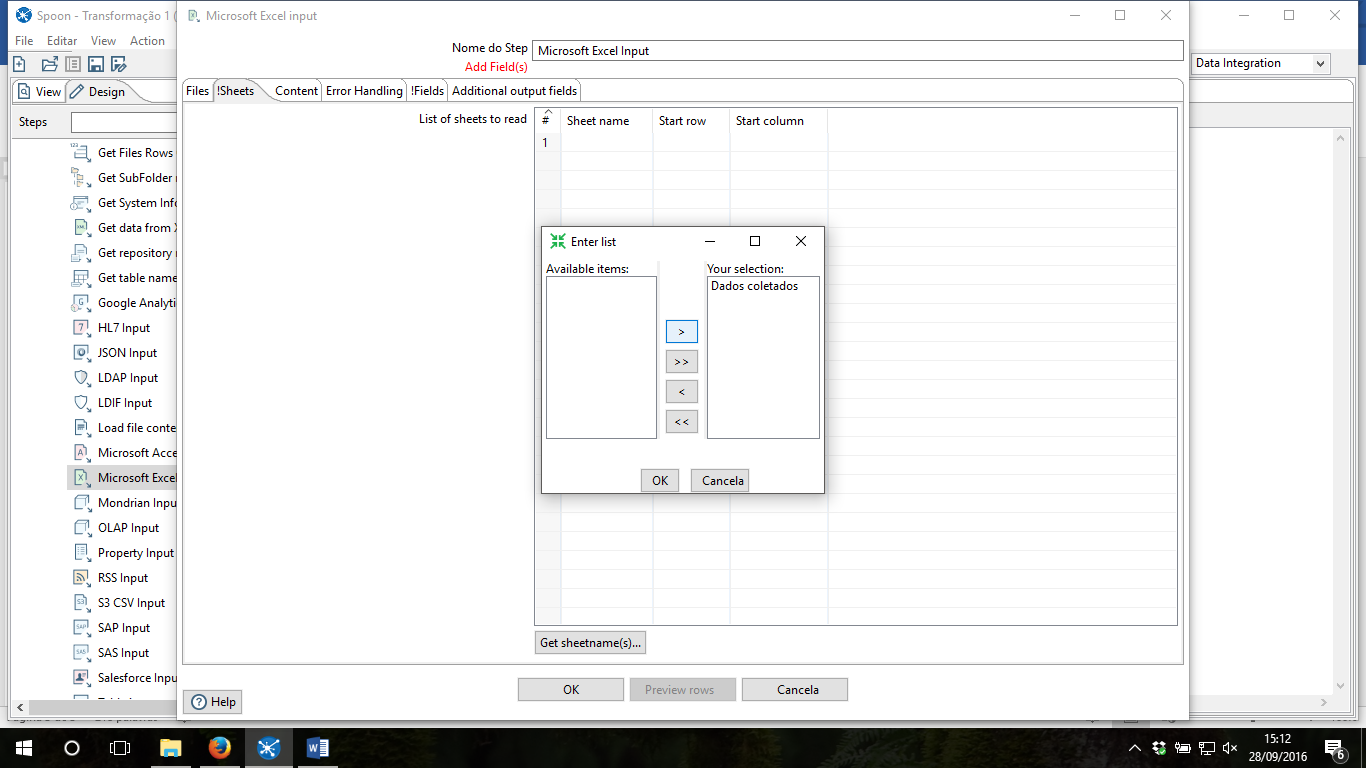
No ambiente do pentaho, clique no menu File-> Novo-> transformação. Na aba design, pasta input, selecione e arrasta três entradas Excel, uma para cada arquivo.



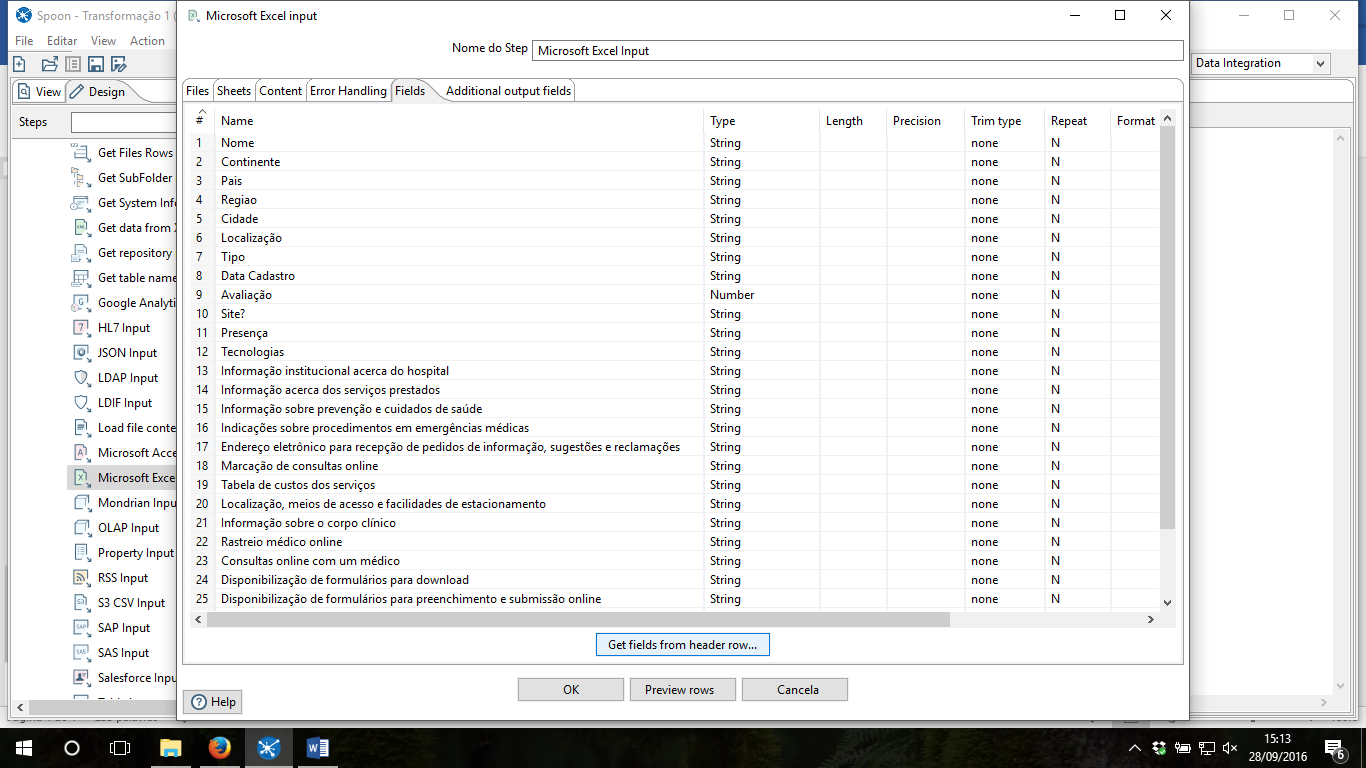
Em seguida, faça o seguinte processo para cada entrada, clicar com botão direito, selecionar editar step, na janela aberta aba files, clique em navegar e selecione arquivo de entrada, depois no botão Add.



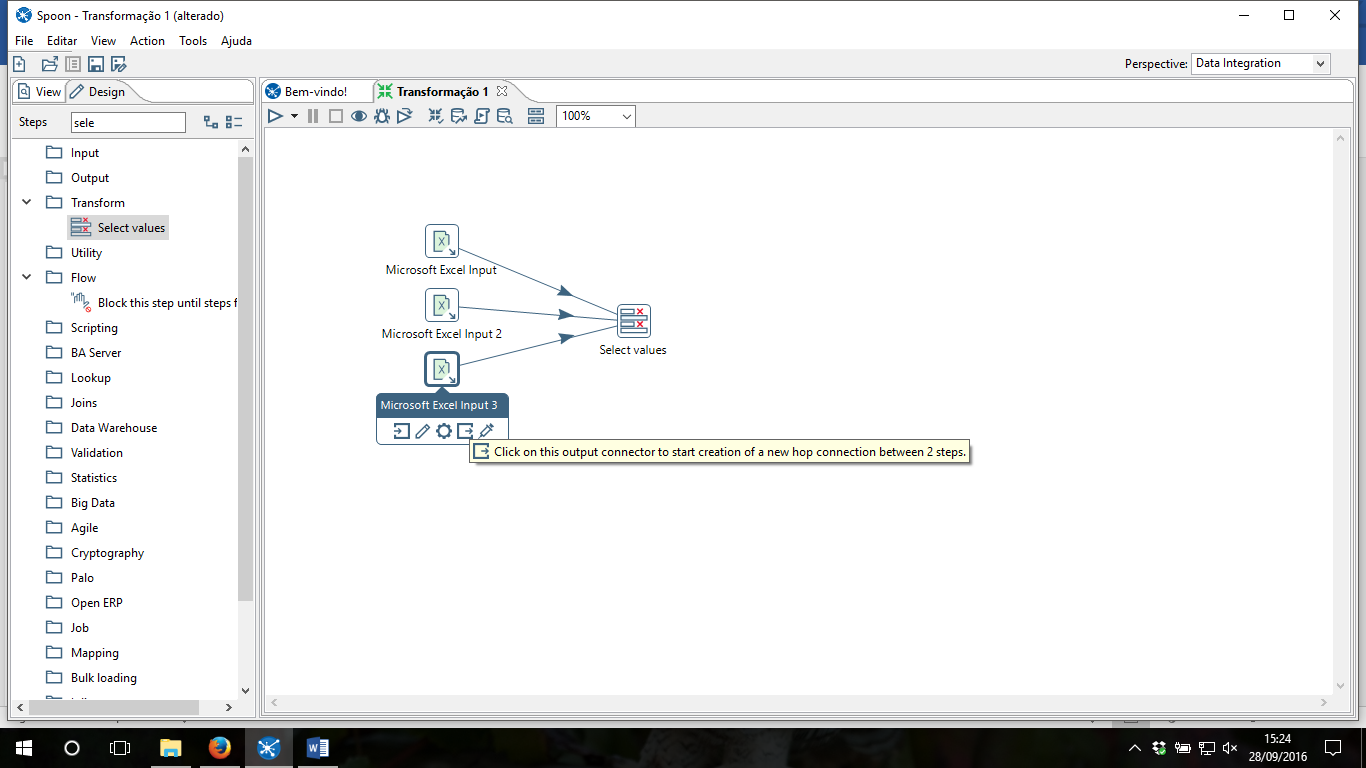
Na aba Sheets, clique em get sheetnames, selecione a pasta de trabalho excel:

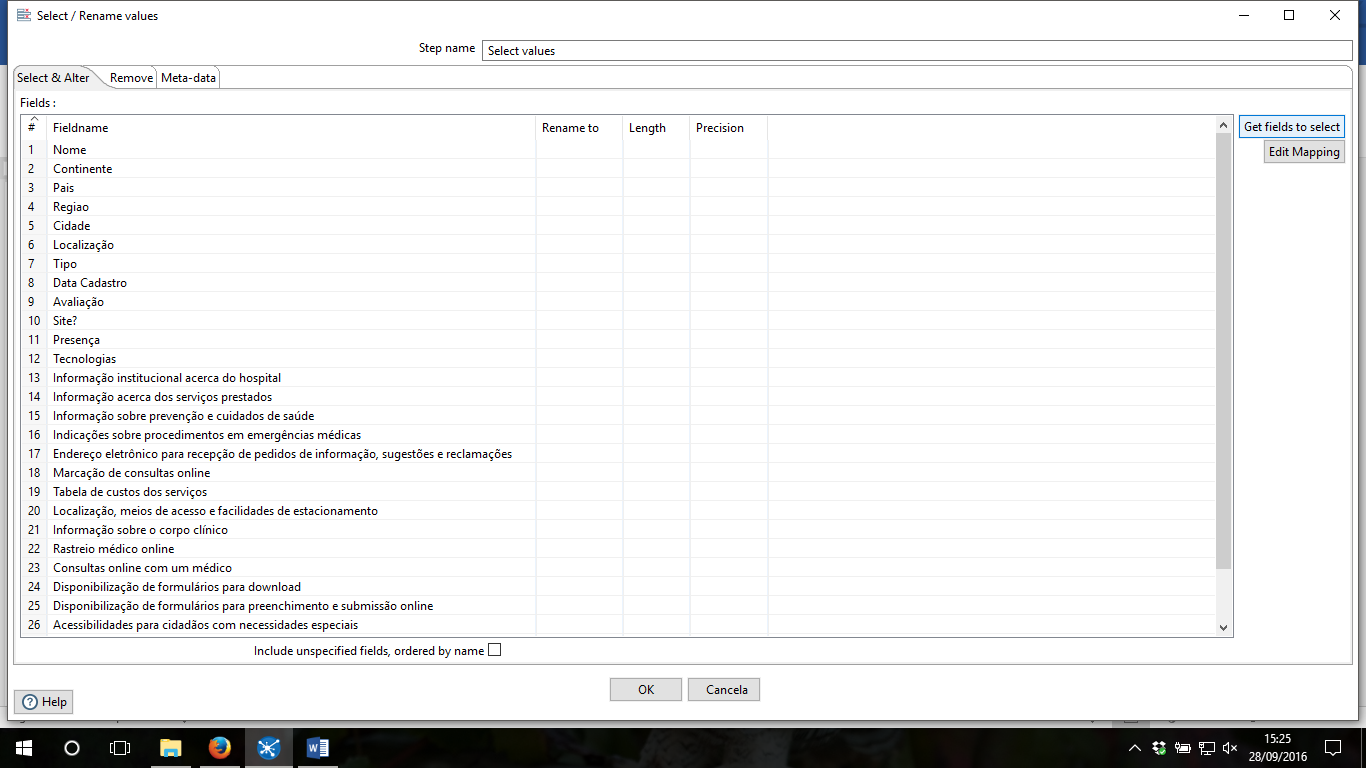


Na aba Fields, clique em “get fields...”, se desejar altera os tipos de dados na coluna type (também alterar para evitar erro nos tipos de dados, precisam estar em comum), terminado clique em ok:

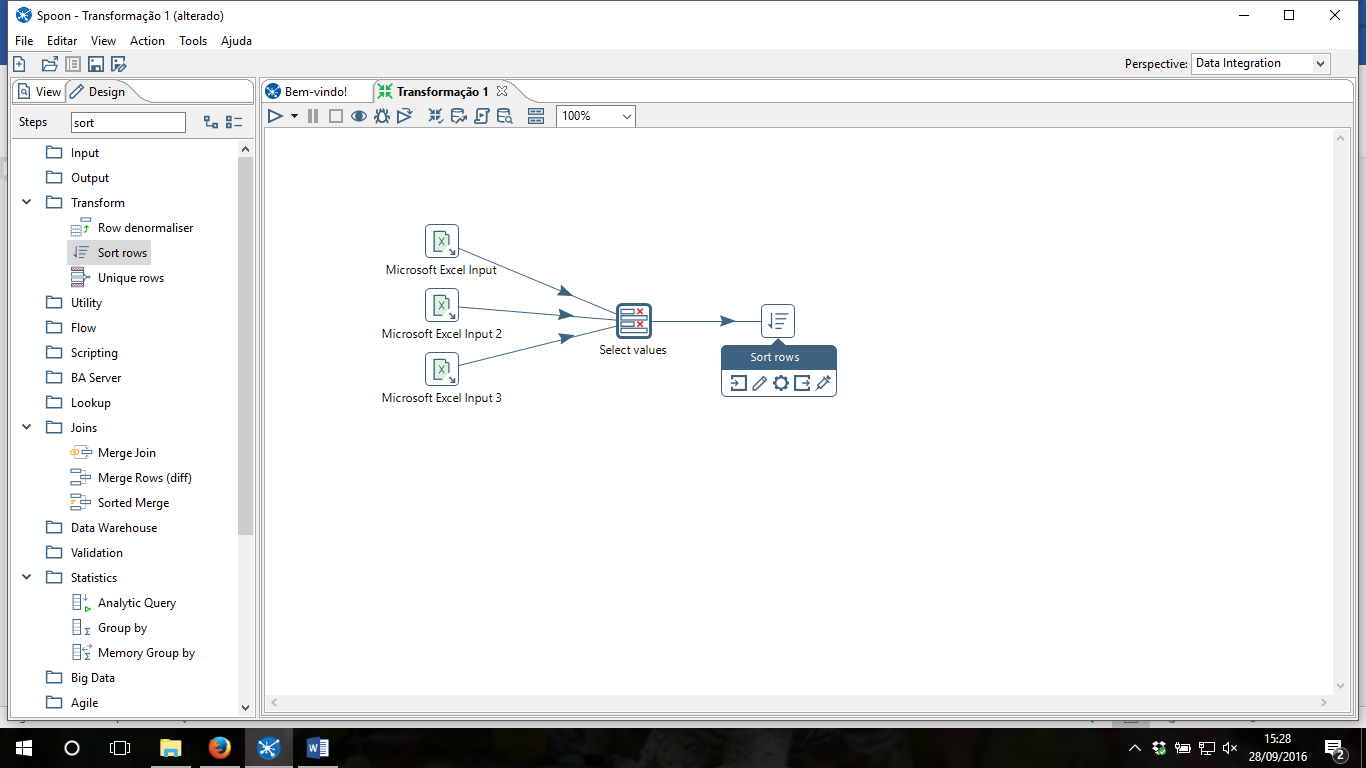


**b)** Para juntar os dados selecione a step select values, clique sobre as entradas Excel e crie um job de cada uma (seta) conforme imagem:

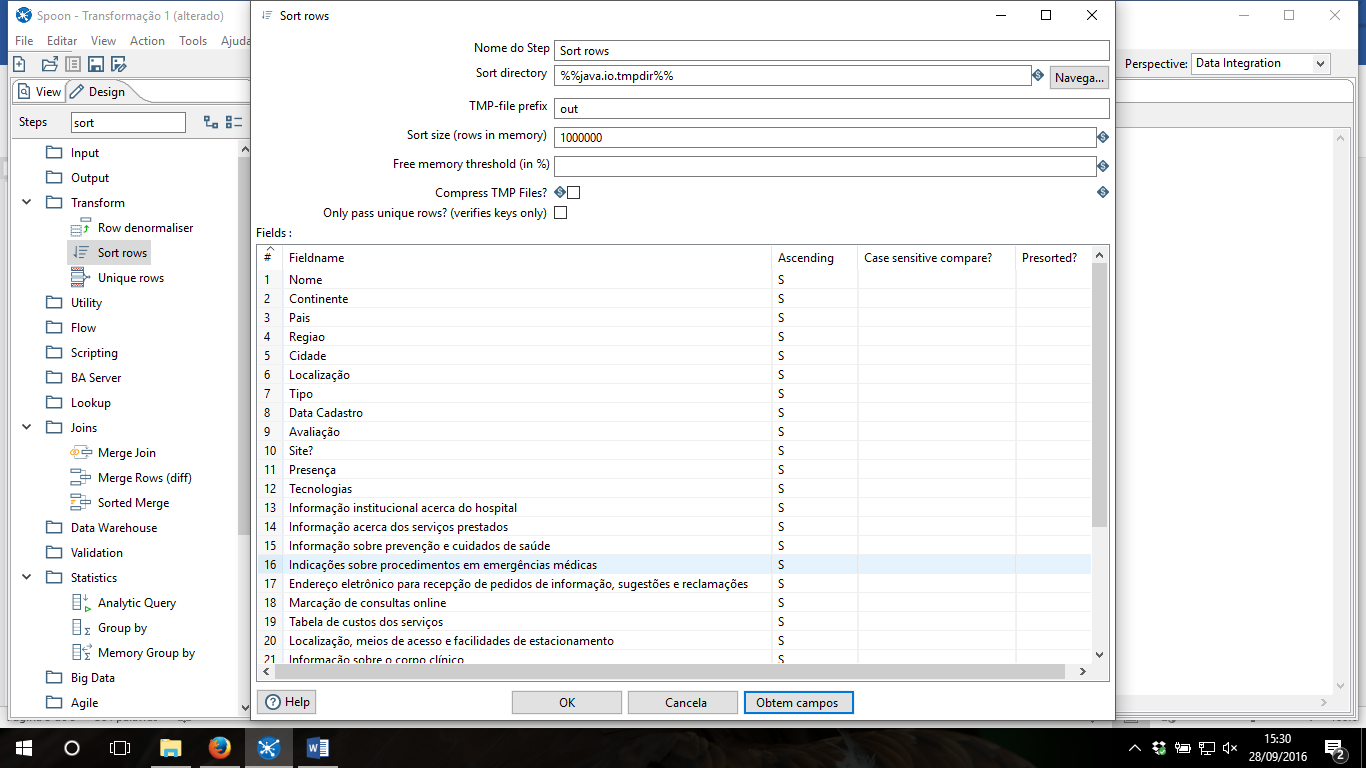


Clique com lado direito sobre a step select values, selecione editar step, clique em “get fields to select” (imagem abaixo) Se desejar remover alguma coluna carregue a na aba remove.

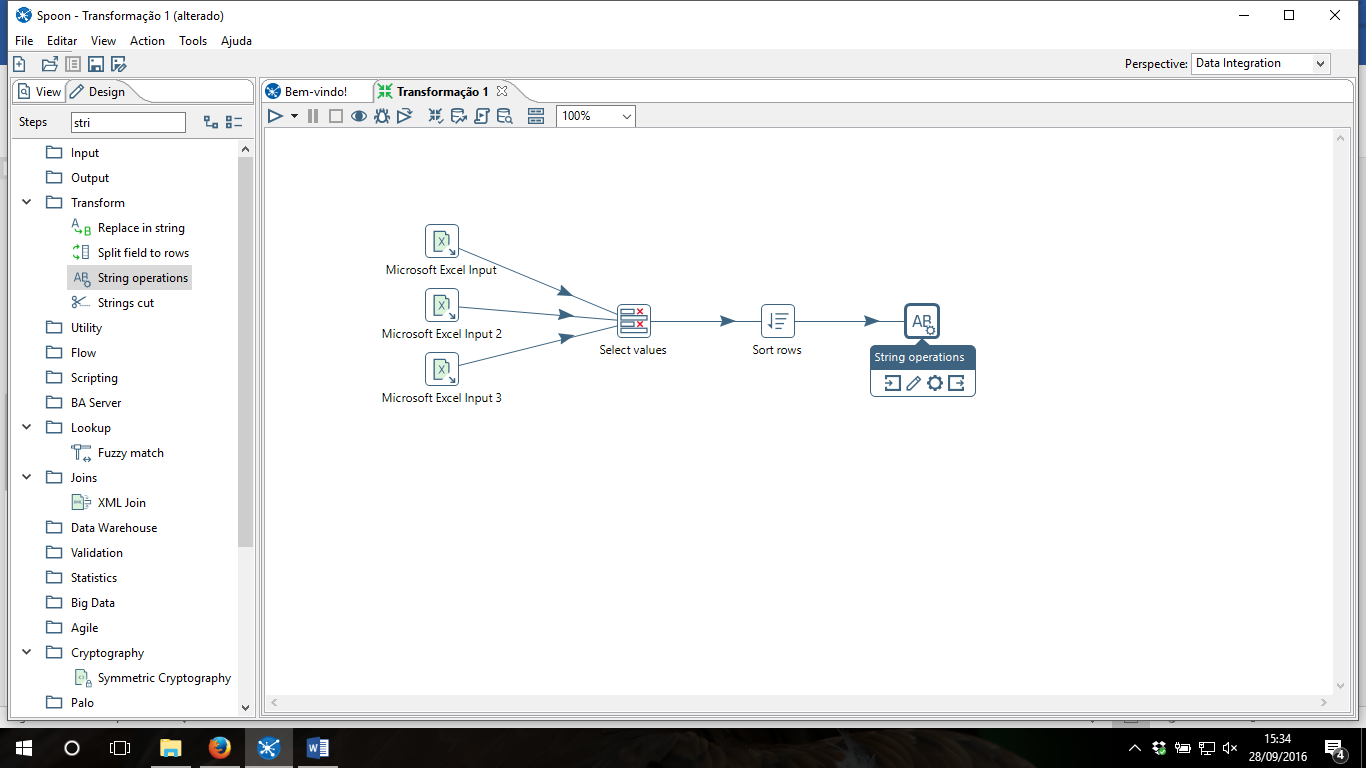
Para ordenar os dados selecione a step sort rows, clique sobre as select values e crie um job (seta) conforme imagem:



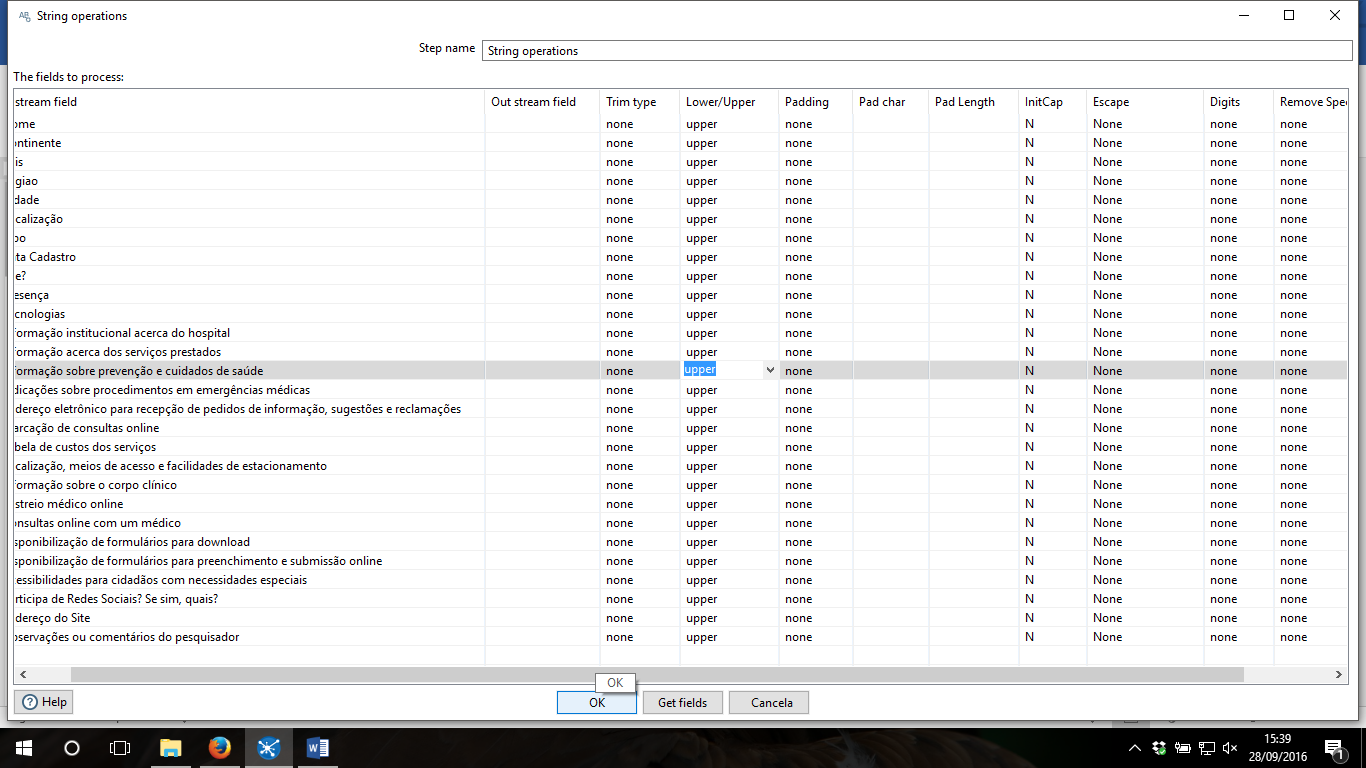
Clique com lado direito sobre a step sort rows, selecione editar step, clique em “obter campos” (imagem abaixo), configure ordenação nas colunas disponíveis e clique em ok quando terminado.



Para tratar os dados selecione a step string operations, clique sobre sort rows e crie um job (seta) conforme imagem:

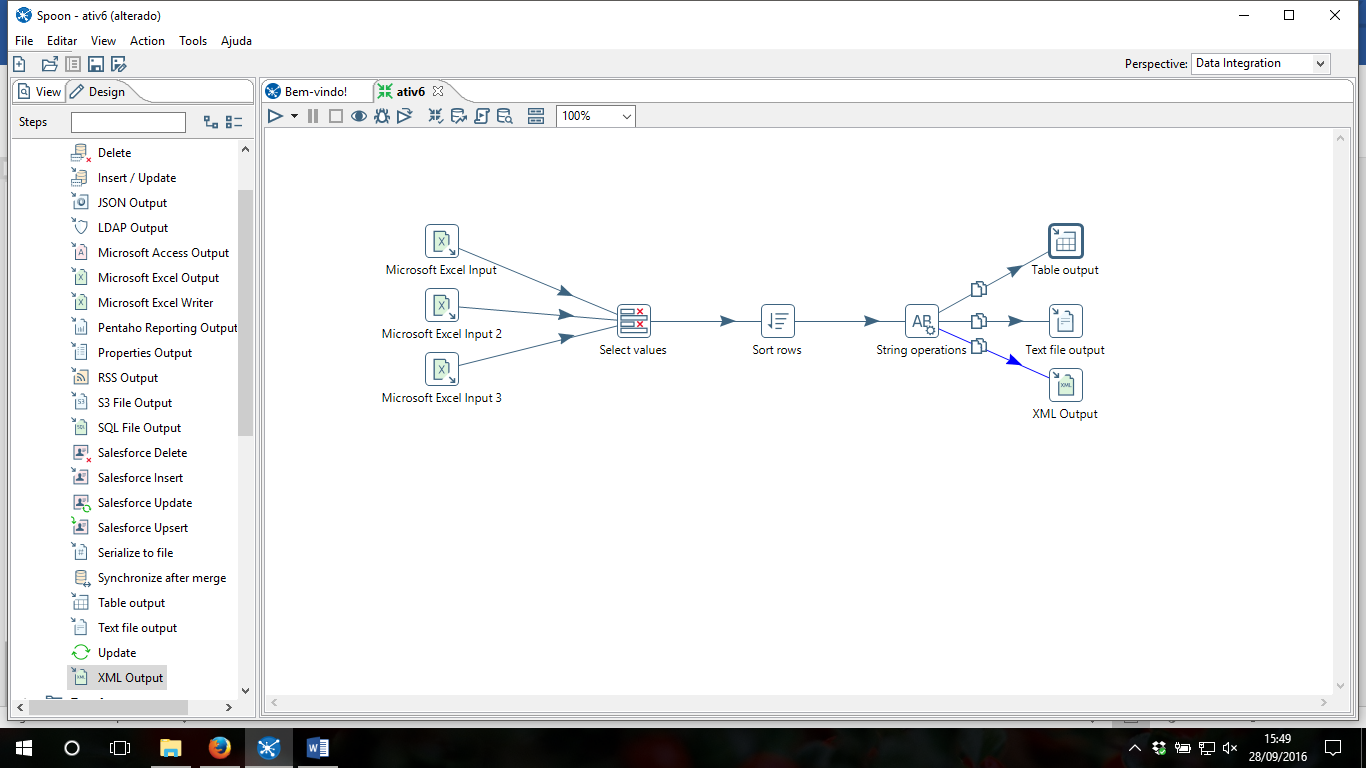


Clique com lado direito sobre a step string operations, selecione editar step, clique em “get fields” (imagem abaixo), configure (nesse caso tudo para upper case) nas colunas disponíveis e clique em ok quando terminado.

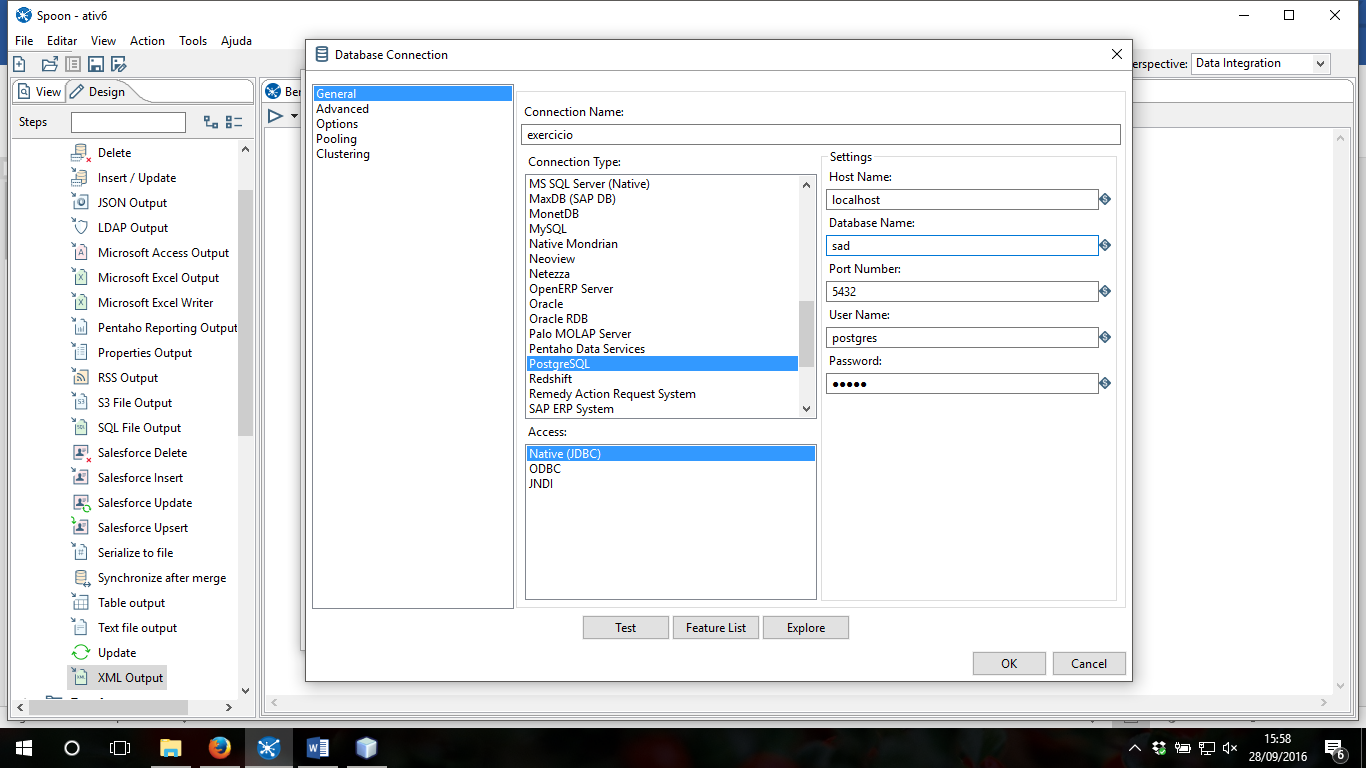


**c) d)** e **e)**

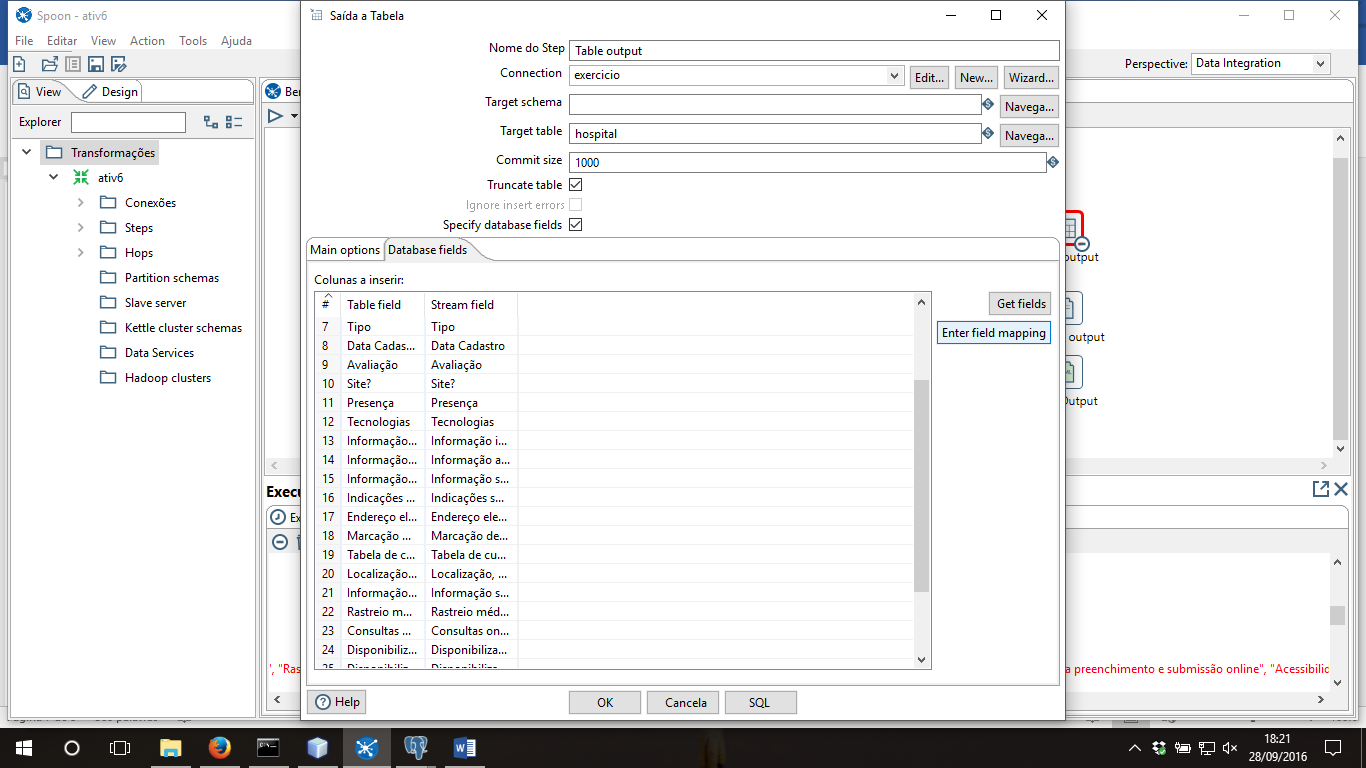
Para as três letras, selecione as respectivas saídas, ligue um job da step de string operations para cada step de saída (selecione copiar quando indagado):

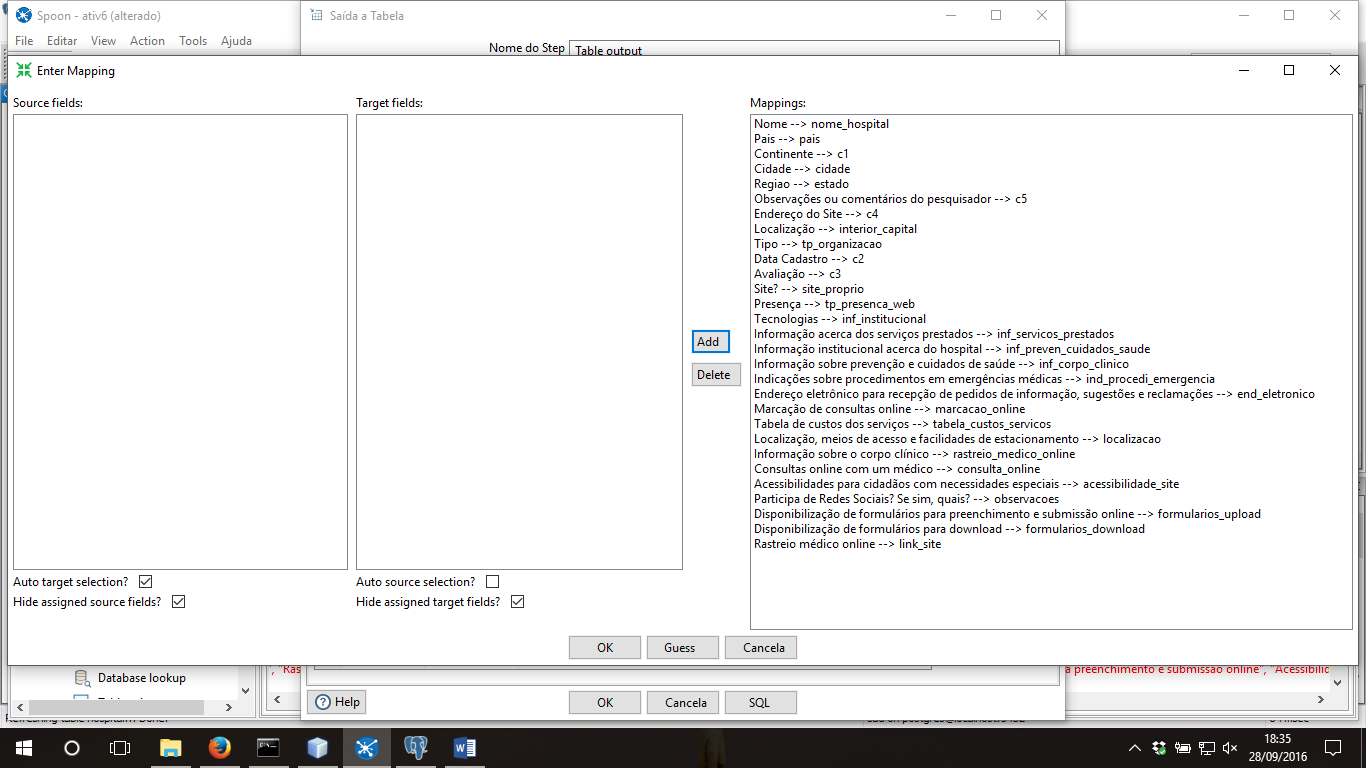


Para letra c, clique com lado direito na step table output, clique com lado direito, em seguida editar step. No campo connection clique em new, configure os dados de acordo com banco de dados escolhido, nesse caso o PostgreSQL, clique em ok ao terminar.

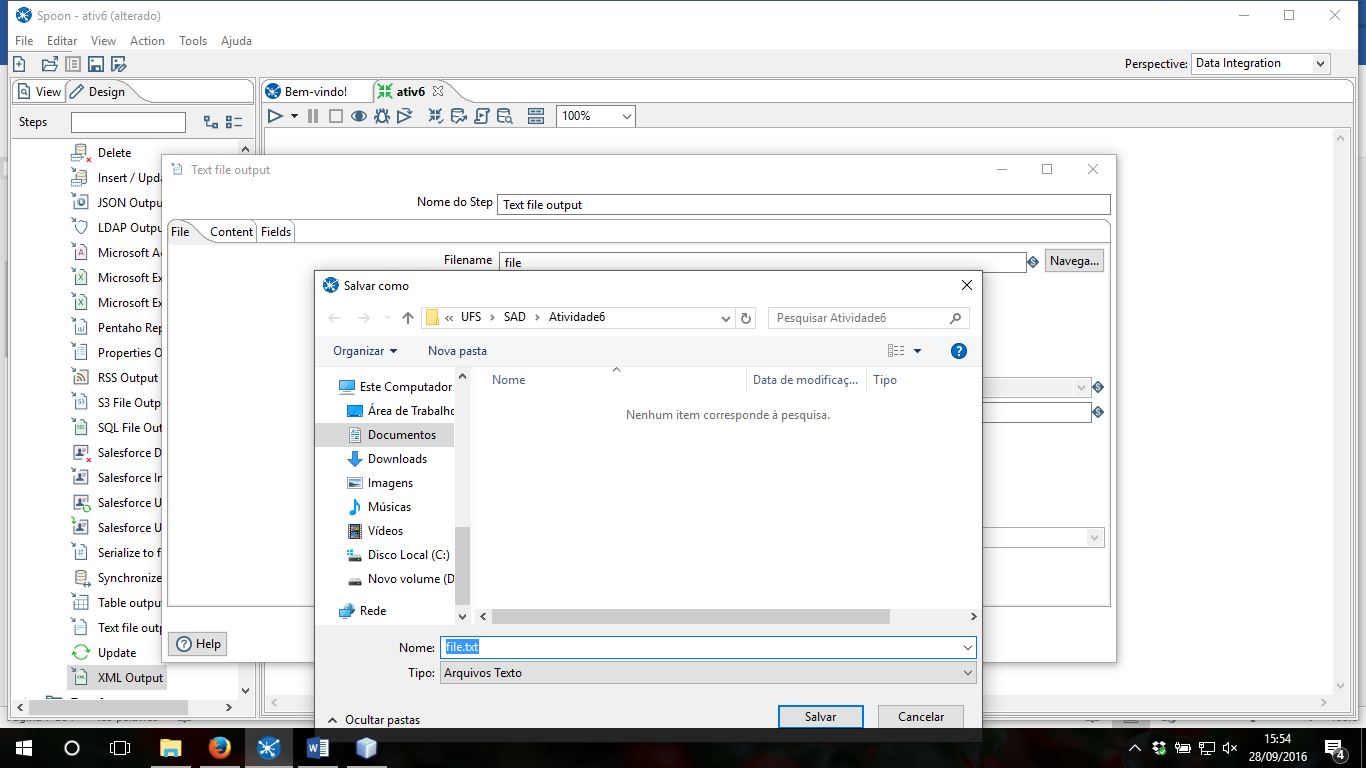


Importante já existir uma database no banco de dados com o nome selecionado, após disso será necessário selecionar a opção “specify database fields”, na aba database fields mapear (opção enter field mapping) os fields para o banco (ver figuras abaixo). Obs(a ordem do mapeamento foi aleatória, a finalidade é só demosntrar).

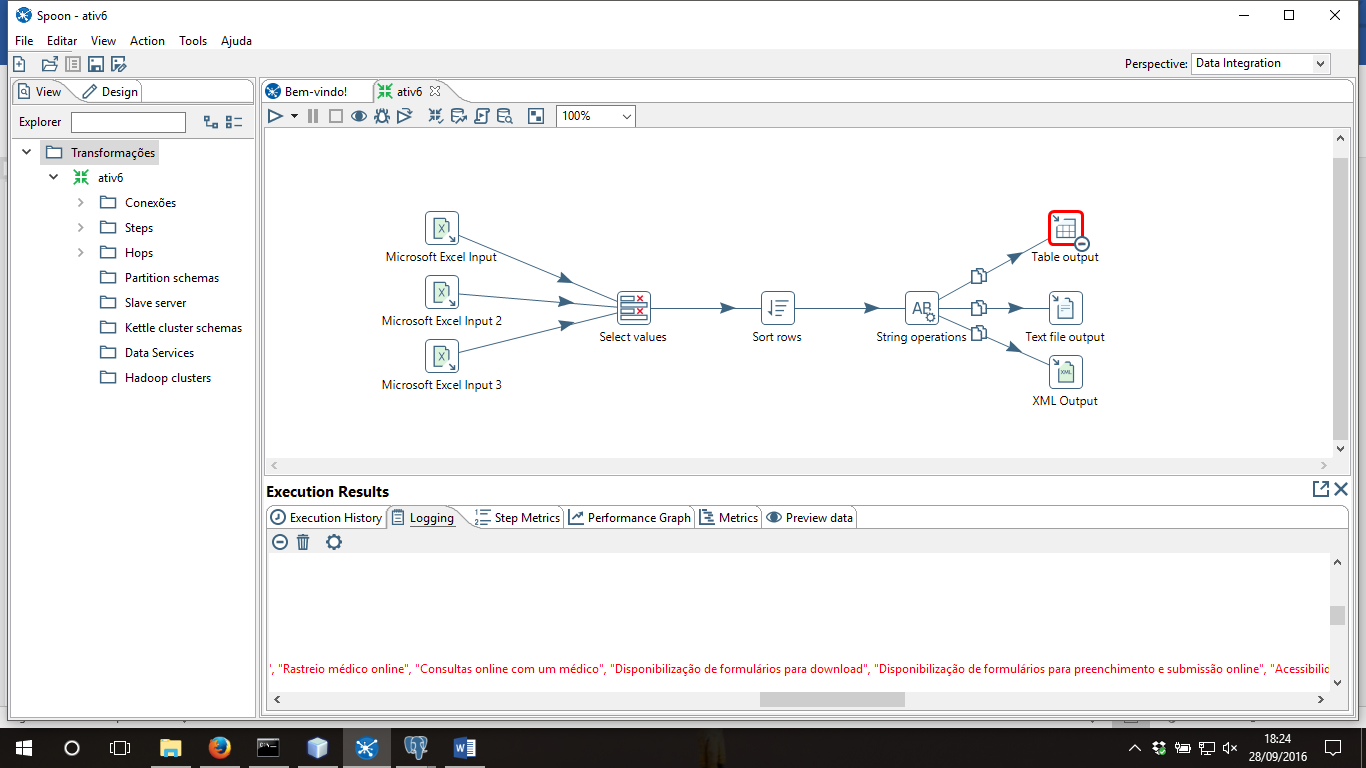




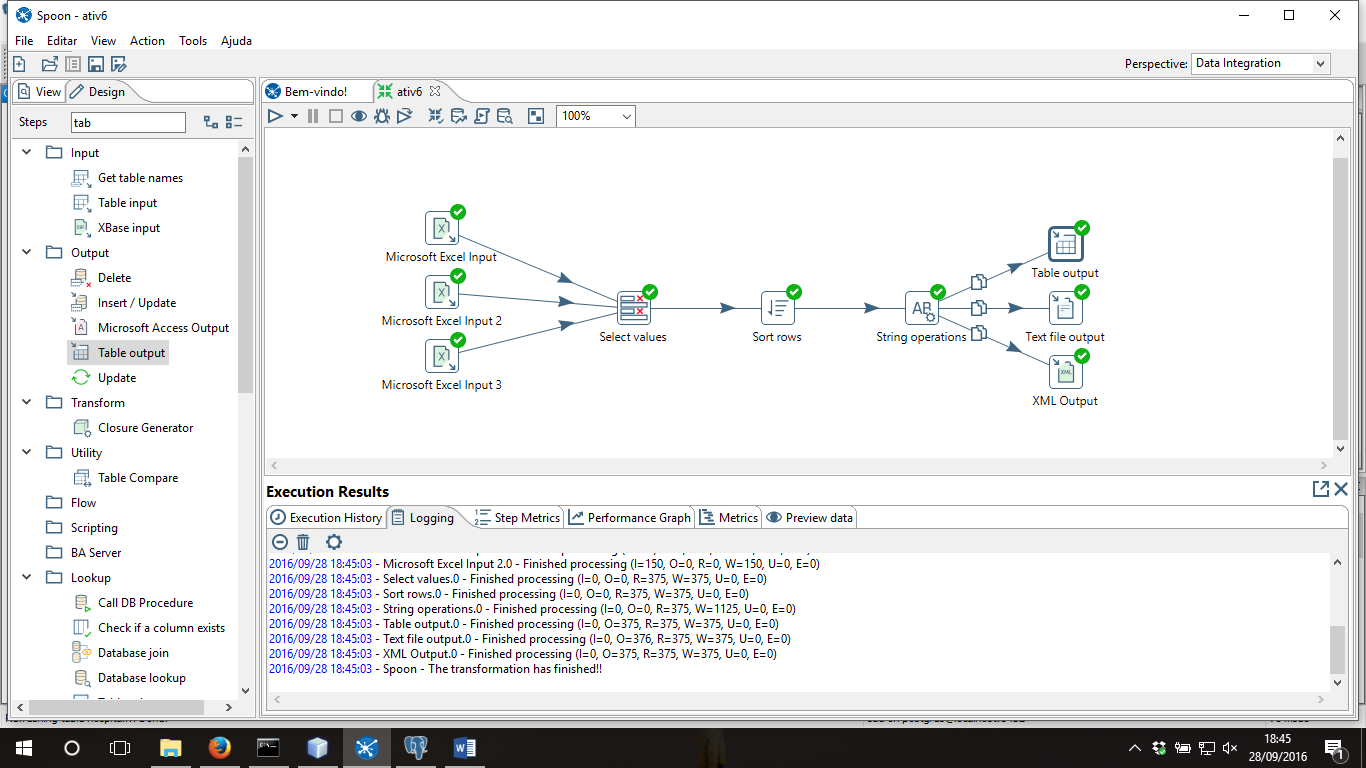
Para letra “d” e “e”, clique com lado direito em cada step de saída, em seguida editar step. No campo filename, clique me navegar, escolha pasta e nome arquivo saída.

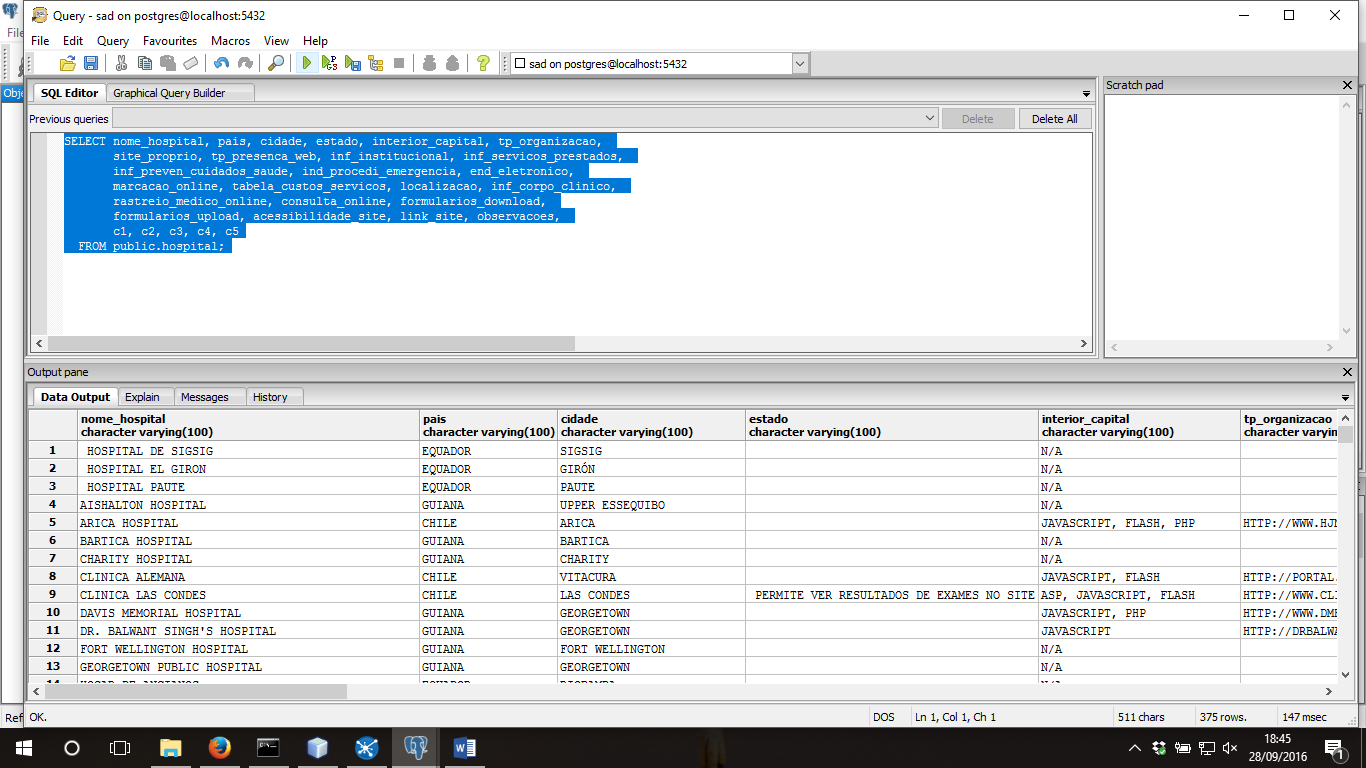


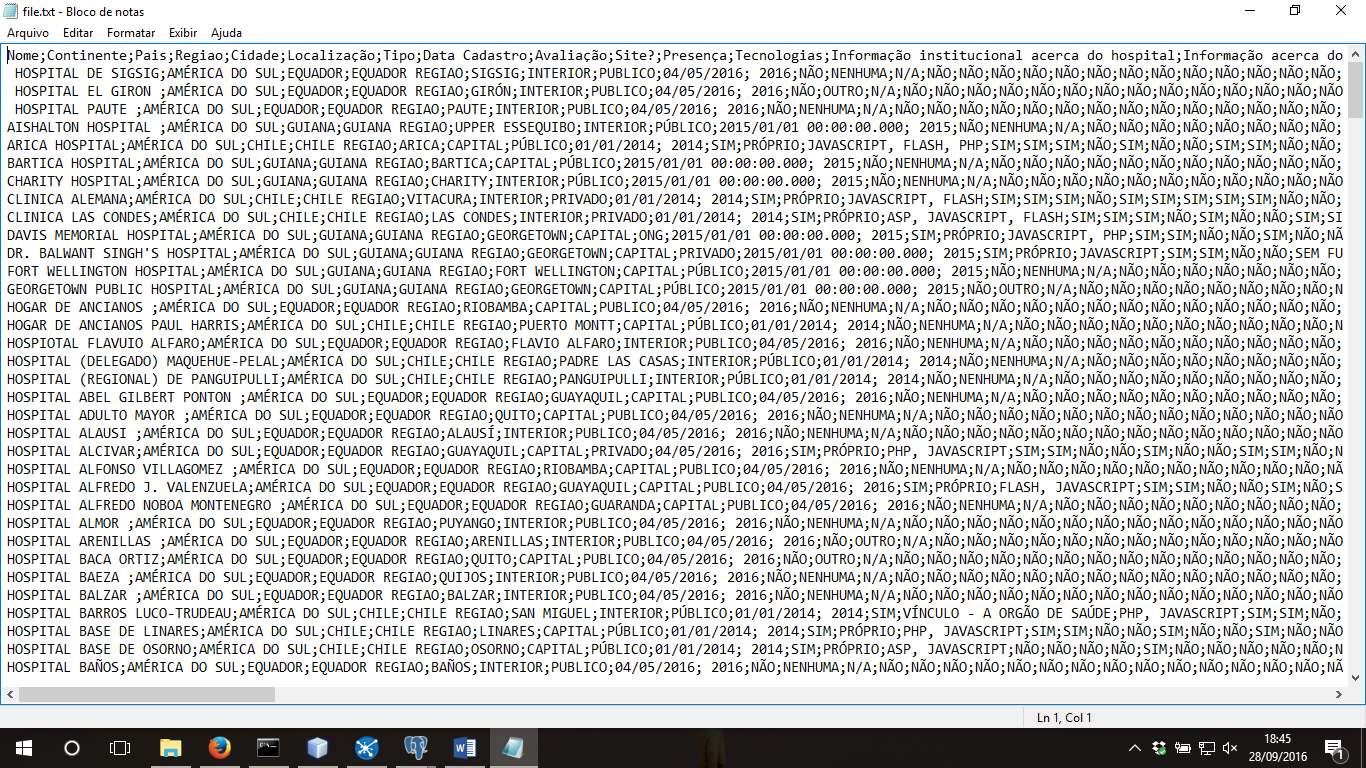
Após finalizar todas as etapas, clique em Run (botão play):

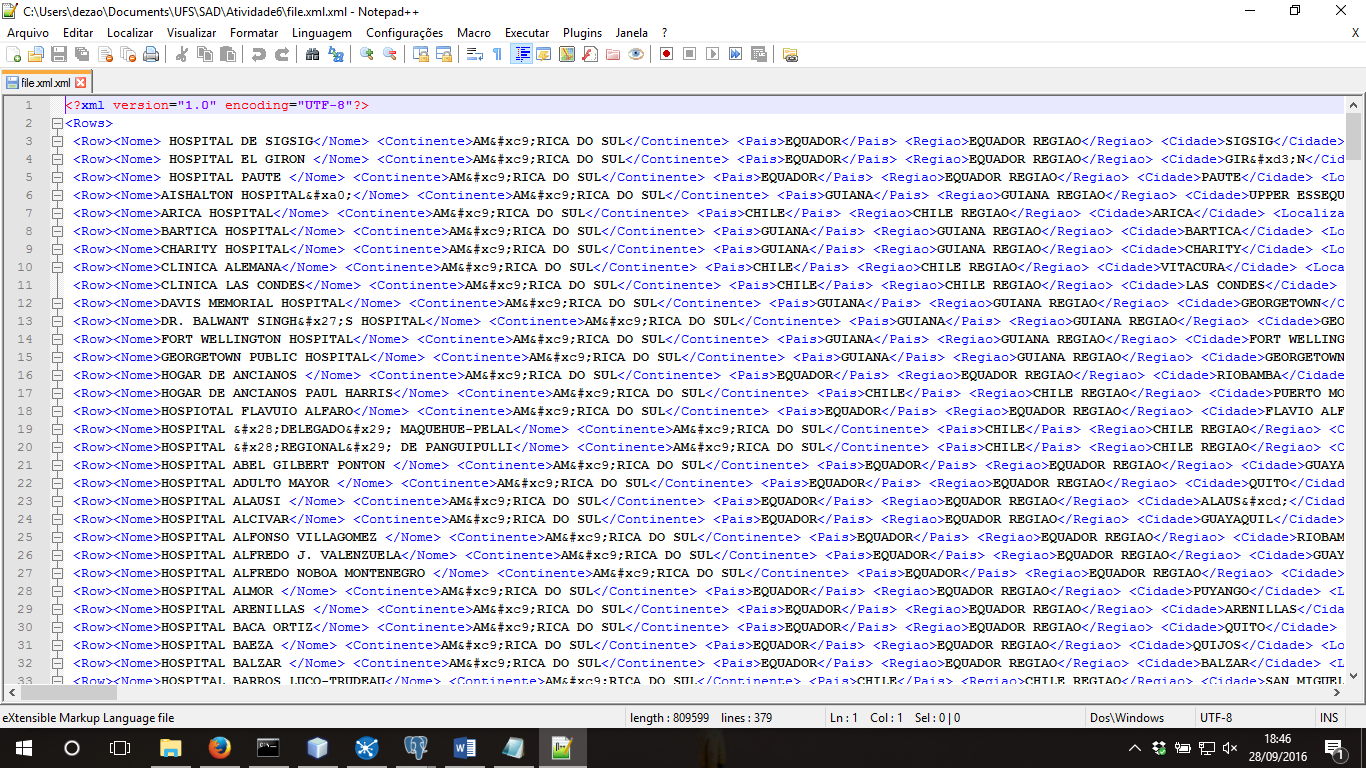


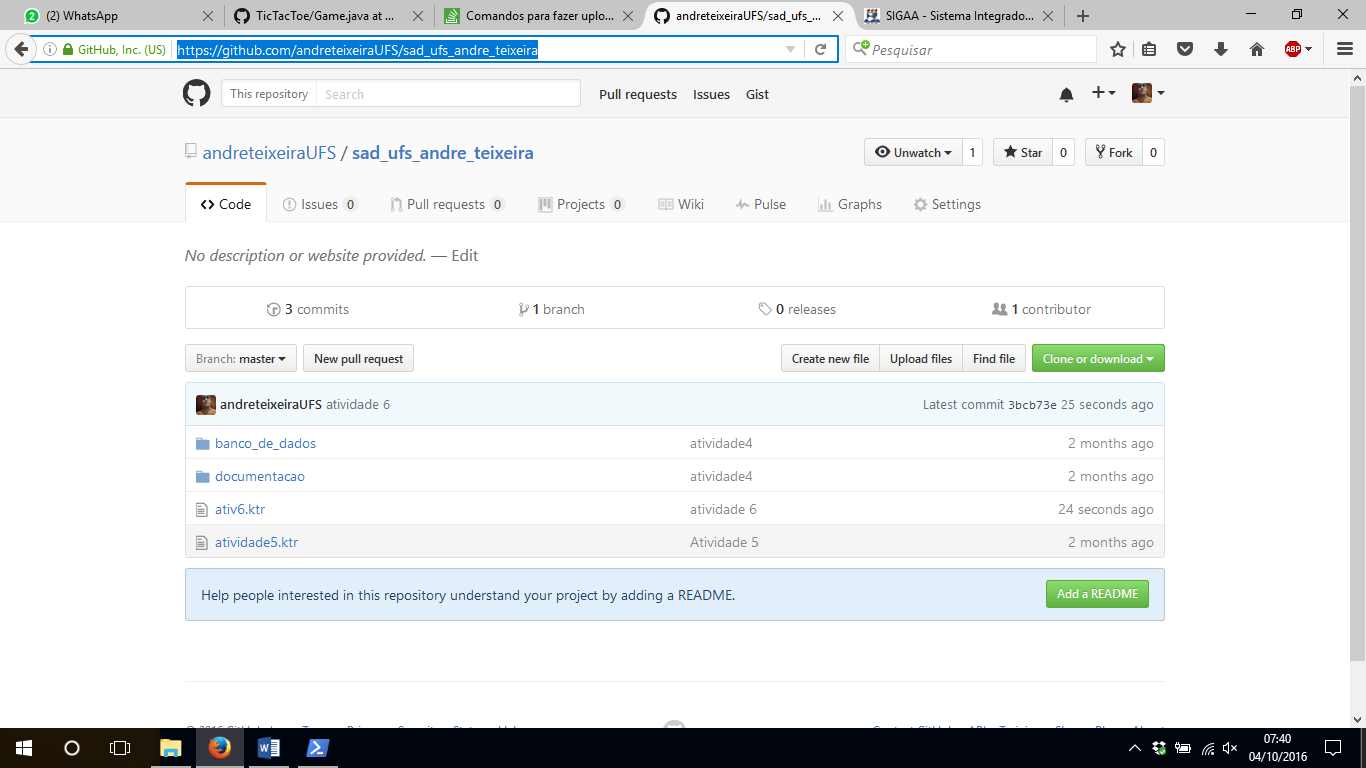
Verifique se tudo deu certo, se o banco foi carregado corretamente, bem como a geração dos arquivos.









**f)** 

**g)** [**https://github.com/andreteixeiraUFS/sad\_ufs\_andre\_teixeira**](https://github.com/andreteixeiraUFS/sad_ufs_andre_teixeira)