

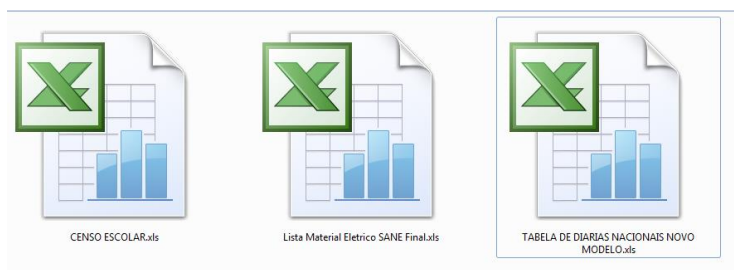
Atividade 6

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

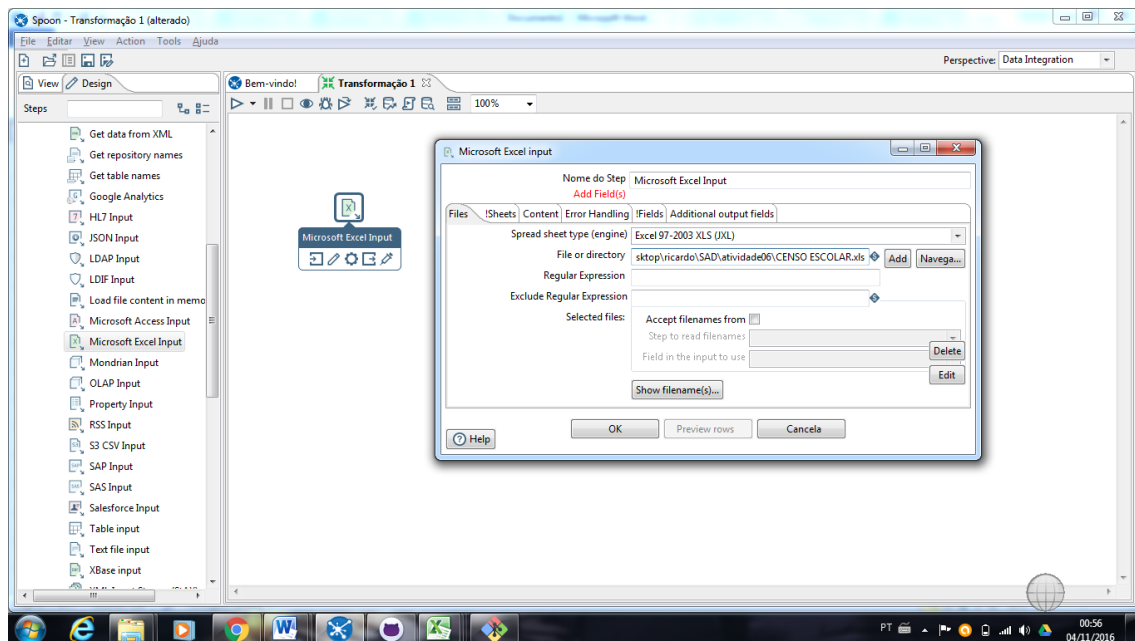
1- Crie uma pequena documentação/tutorial de como realizar os seguintes tópicos (tire snapshots das telas e escreva de forma explicativa):

- 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

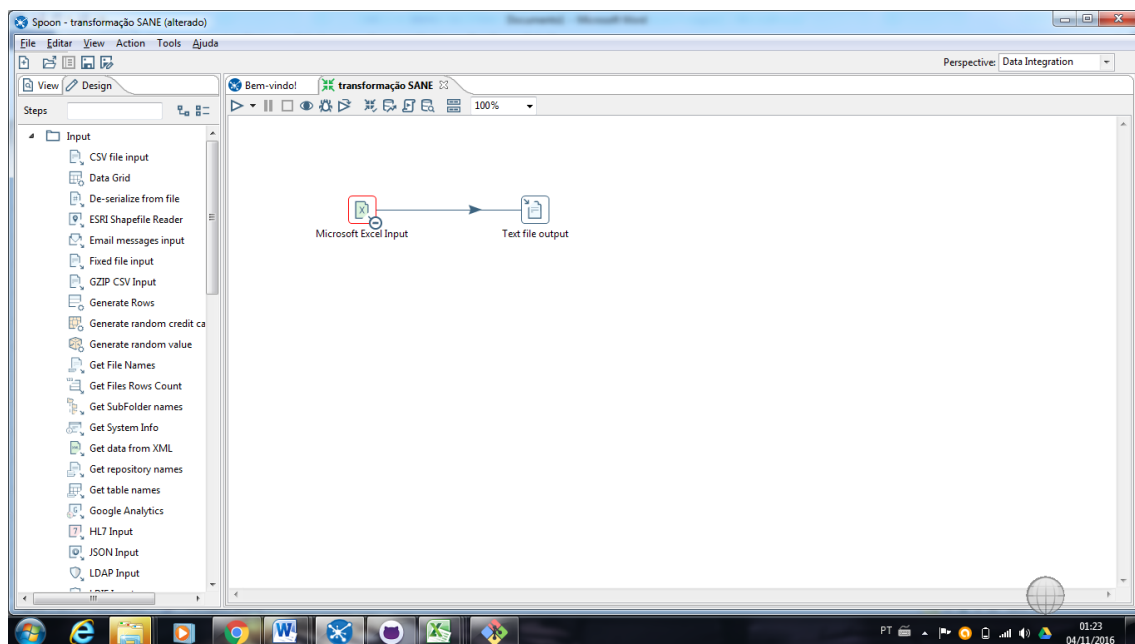
Utilizaremos os seguintes arquivos, disponíveis na internet: CENSO ESCOLAR.xls, Lista Material Eletrico SANE Final.xls e TABELA DE DIARIAS NACIONAIS NOVO MODELO.xls



Em seguida vamos carrega-los no PDI, os mesmos serão Input para transformações no processo de ETL, como visto a seguir. Entre as opções de “Input” escolhemos o formato do arquivo desejado, clicamos e arrastamos a opção até a tela ao lado, em seguida damos um duplo clique sobre o ícone gerado, será exibida a janela a seguir, onde navegamos até o arquivo desejado e clicamos em “OK” para carregar.

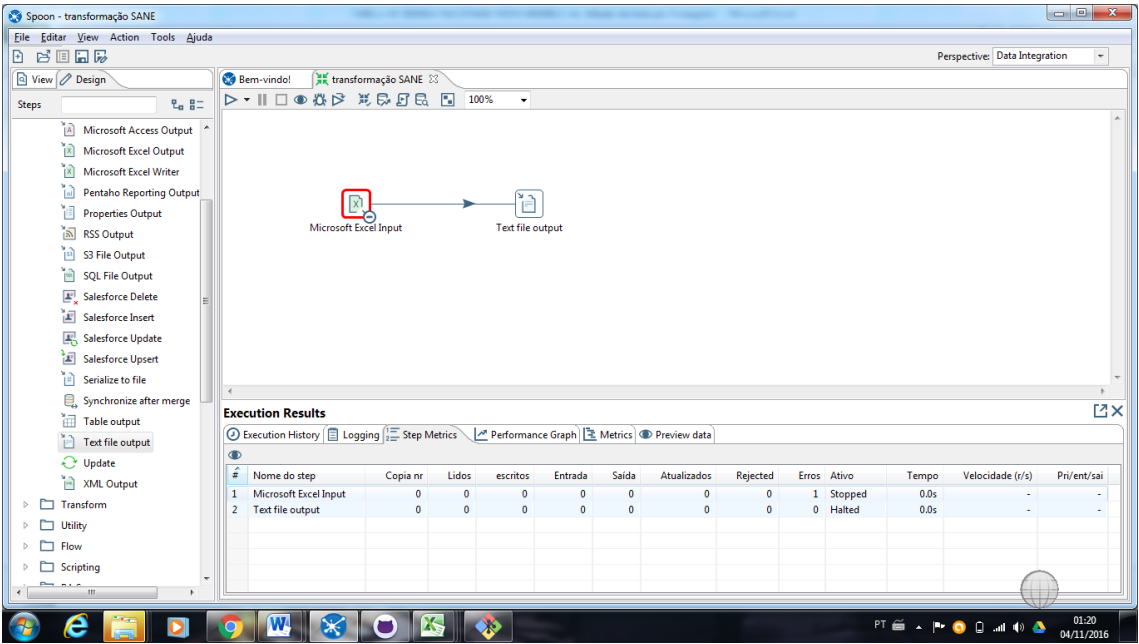


Em seguida escolhemos o formato de saída, neste caso, .txt.



Em seguida executamos a transformação, porém ocorreu um erro inesperado! Este erro se repetiu nas demais tentativas de transformação.

Lista Material Eletrico SANE Final.xls



CENSO ESCOLAR.xls

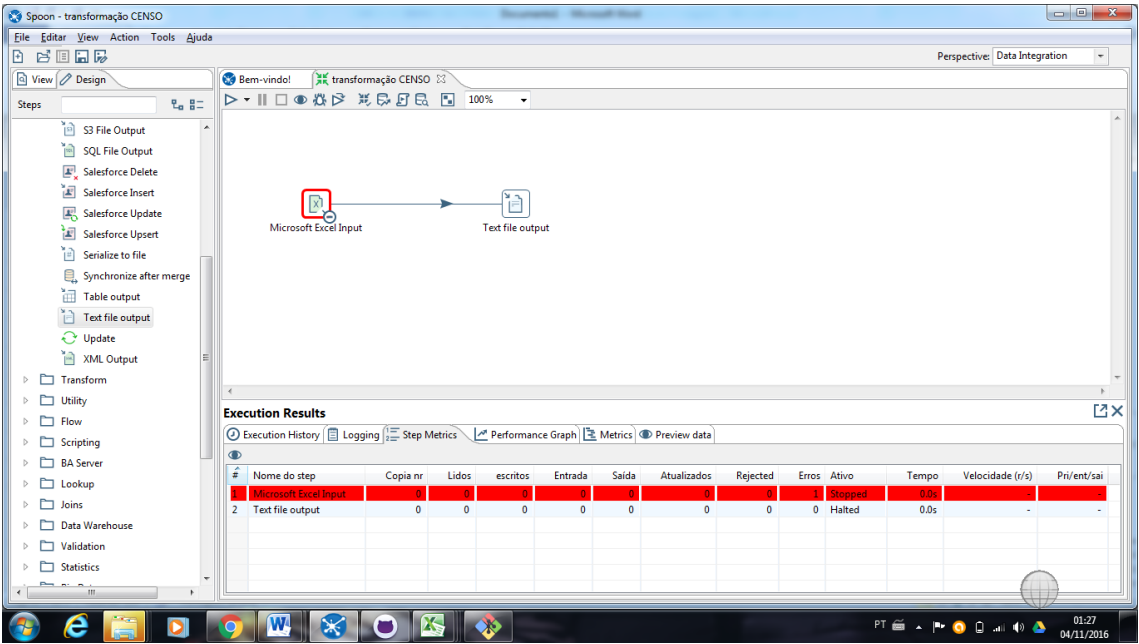


TABELA DE DIARIAS NACIONAIS NOVO MODELO.xls

The screenshot displays the Apache NiFi web console for a workflow named 'transformação TABELA'. The design canvas shows a flow from 'Microsoft Excel Input' to 'Text file output'. The 'Execution Results' table provides a detailed log of the workflow's execution.

| # | Nome do step | Copia nr | Lidos | escritos | Entrada | Saida | Atualizados | Rejected | Erros | Ativo | Tempo | Velocidade (r/s) | Pri/ent/sai |
|---|-----------------------|----------|-------|----------|---------|-------|-------------|----------|-------|---------|-------|------------------|-------------|
| 1 | Microsoft Excel Input | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | Stopped | 0.0s | - | - |
| 2 | Text file output | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Halted | 0.0s | - | - |

Mesmo com os erros, as saídas foram salvas no formato .xml.