

Documentação das Transformações das bases de Servidores

Yago Alves da Silva

Pré-processamento dos arquivos

1. O primeiro passo aqui é utilizar o script de para a extração dos dados que são protegidos pelo governo, **Replace.py** (<https://github.com/npi-ufc-qxd/dados-abertos/blob/master/Scripts/Replace.py>).

O segundo, consiste em usar mais um script, o **Dividir.py** (<https://github.com/npi-ufc-qxd/dados-abertos/blob/master/Scripts/Dividir.py>). Dessa vez com o objetivo de dividir o arquivo e diminuir a quantidade de linhas passadas, para evitar travamentos no Pentaho.

O comando para a execução segue a estrutura :

```
python Script.py
```

Deve-se passar o diretório onde se encontram os arquivos .csv dentro do método no caso do Replace.py como mostra a Figura 1 e no Dividir.py como mostra na Figura 2.

Figura 1

```
for fileName in glob.glob('Diretorio do arquivo.csv'):
```

Figura 2

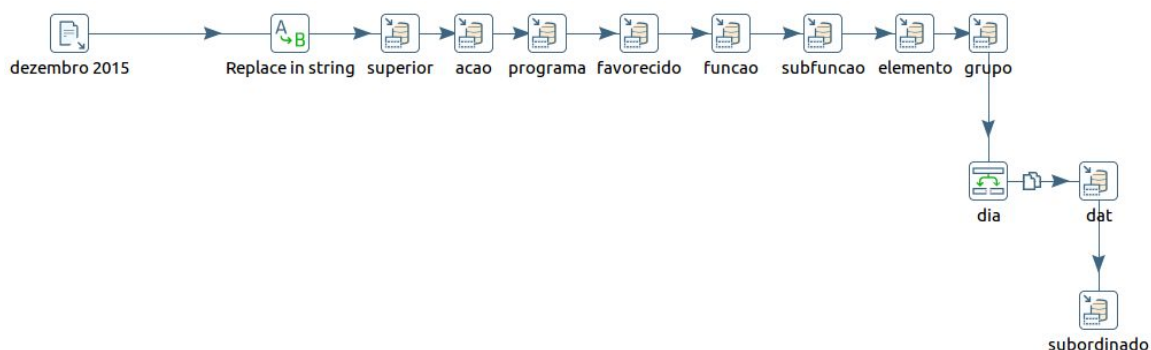
```
for fileName in glob.glob(options.folderName + 'Diretorio do arquivo.csv'):
```

Transformações do Pentaho

1. T1.ktr

A transformação T1 mostrada na Figura 3 é responsável por fazer a carga tabelas de Órgão superior, Ação, Programa, Favorecido, Função, Subfunção, Elemento, Grupo e Data.

Figura 3



dezembro 2015 : Essa *step* é deve ser usada para inserir as partes divididas do arquivo GastosDiretos.csv

Replace in string : Substitui os valores dos campos que possuem “ \00” por um espaço em branco no atributo `Lingagem_cidadã`.

superior : Insere os dados do arquivo csv na tabela `orgao_superior`.

acao : Insere os dados do arquivo csv na tabela `acao`.

programa : Insere os dados do arquivo csv na tabela `programa`

favorecido : Insere os dados do arquivo csv na tabela `favorecido`

funcao : Insere os dados do arquivo csv na tabela `funcao`

subfuncao : Insere os dados do arquivo csv na tabela `sub_funcao`

elemento : Insere os dados do arquivo csv na tabela `elemento`

grupo : Insere os dados do arquivo csv na tabela `grupo`

dia : Separa os valores de `Data_Pagamento` em campos distintos de “dd”, “mm” e “aaaa”.

dat : Insere os dados quebrados na step **dia** na tabela `data`.

subordinado : Insere os dados do arquivo csv na tabela `orgao_subordinado`.

2. T3.ktr

A transformação T3 mostrada na Figura 4, deve ser inserida antes da T2 e é responsável por fazer a carga tabelas de `unidade_gestora`.

Figura 4



dezembro 2015 : Essa *step* deve ser usada para inserir as partes divididas do arquivo `GastosDiretos.csv`

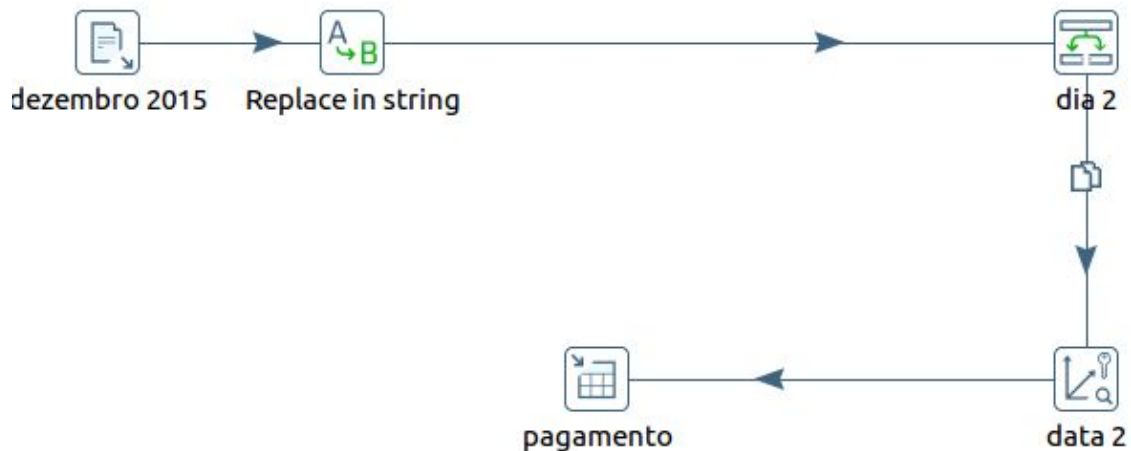
Replace in string : Substitui os valores dos campos que possuem “ \00” por um espaço em branco no atributo `Lingagem_cidadã`.

unidade gestora : Insere os dados do arquivo csv na tabela `unidade_gestora`

3. T2.ktr

A transformação T2 mostrada na Figura 5, deve ser inserida após o término de todas as outras e é responsável por fazer a carga da tabela de pagamento.

Figura 5



dezembro 2015 : Essa *step* é deve ser usada para inserir as partes divididas do arquivo GastosDiretos.csv.

Replace in string : Substitui os valores dos campos que possuem “ \00” por um espaço em branco no atributo Linguagem_cidadã.

dia 2: Separa os valores de Data_Pagamento em campos distintos de “dd”, “mm” e “aaaa”

data 2 : Faz o relacionamento das tabelas “data” e “pagamento”

pagamento : Insere os dados do arquivo csv na tabela pagamento.