

## 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

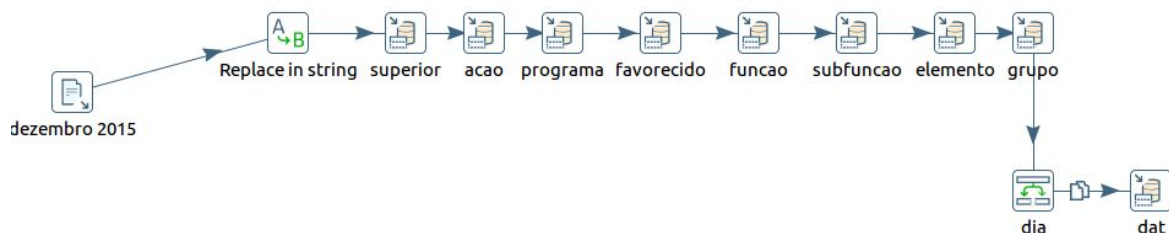
```
for fileName in glob.glob(options.folderName + 'Diretorio do arquivo.csv'):
```

### 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 Linguagem\_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

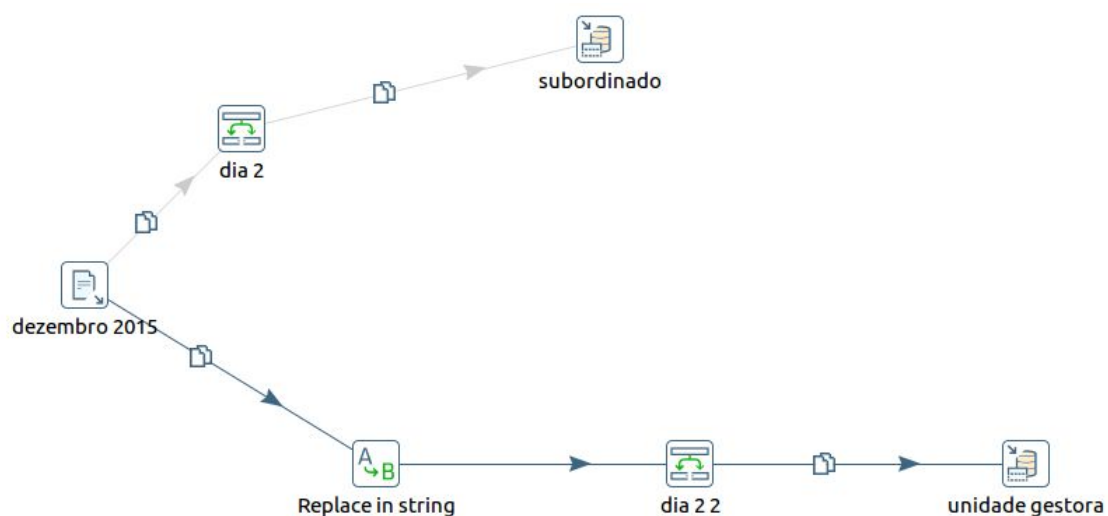
**dia** : Separa os valores de Data\_Pagamento em campos distintos de “dd”, “mm” e “aaaa”.

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

## 2. T3.ktr

A transformação T2 deve ser inserida após o término da T3 e T4, sendo assim, a Figura 4 mostra uma imagem da transformação completa.

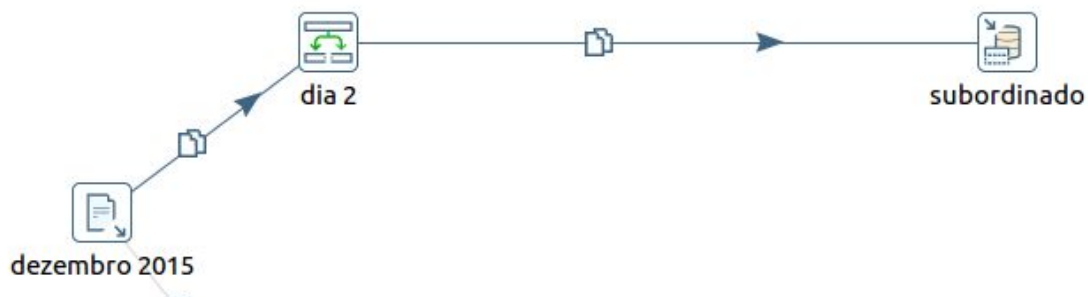
Figura 4



A primeira parte da transformação T3, representada na Figura 5 é responsável por fazer a carga da tabela de orgao\_subordinado.

**OBS :** Para que a transformação possa ser feita por partes, as setas que ligam as steps para os arquivos de saída devem estar azul.

Figura 5



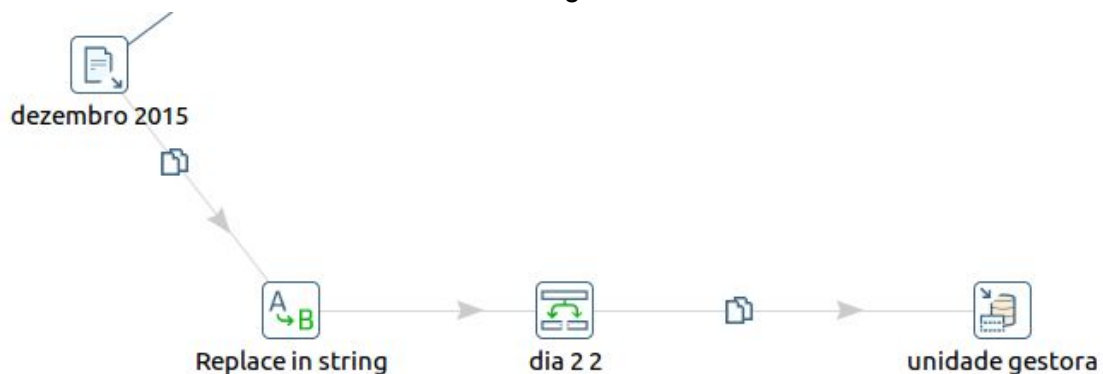
**dezembro 2015 :** Essa *step* é deve ser usada para inserir as partes divididas do arquivo GastosDiretos.csv

**dia 2 :** Separa os valores de Data\_Pagamento em campos distintos de “dd”, “mm” e “aaaa”

**subordinado :** Insere os dados do arquivo csv na tabela grupo

A Figura 6 mostra a segunda parte da transformação T3.ktr, responsável por preencher a tabela unidade\_gestora.

Figura 6



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

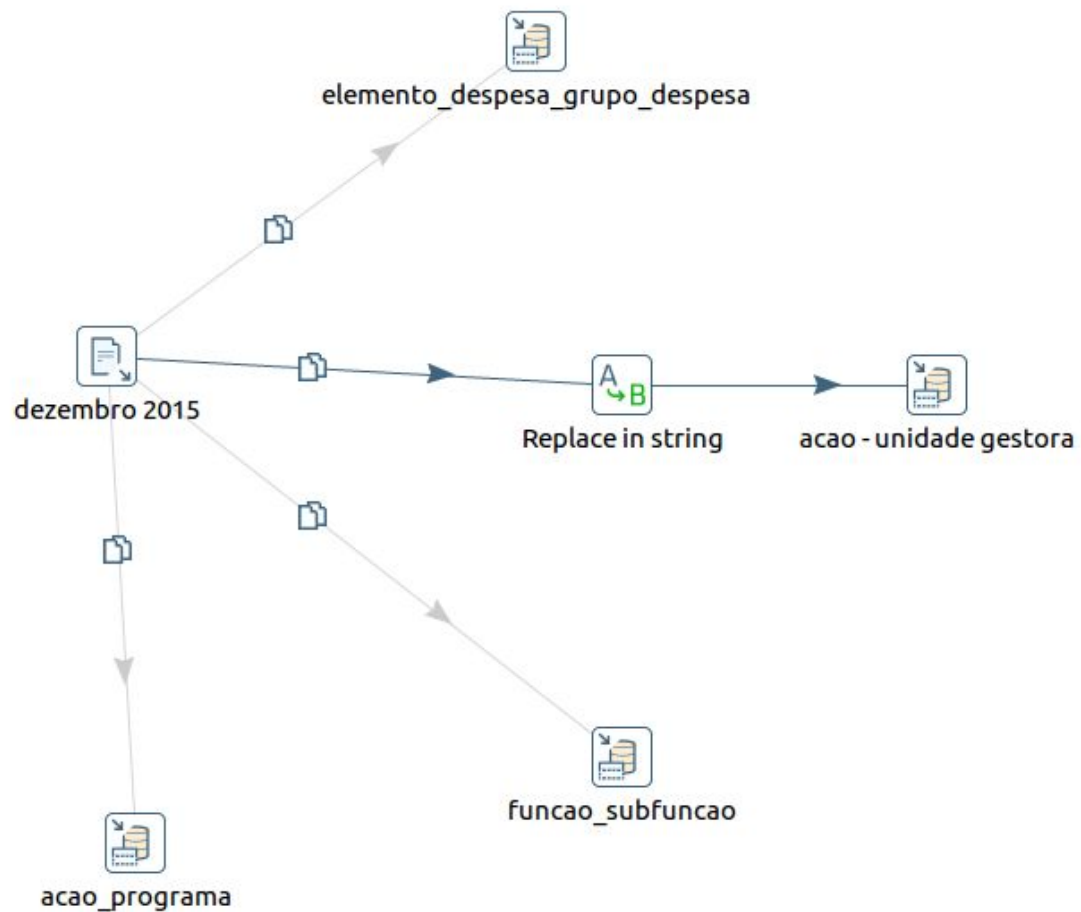
**Replace in string :** Substitui os valores dos campos que possuem “\00” por um espaço em branco no atributo Linguagem\_cidadã.

**dia 2 2** : Separa os valores de Data\_Pagamento em campos distintos de “dd”, “mm” e “aaaa”

### 3. T4.ktr

A Figura 7 mostra transformação T4 completa.

Figura 7



Assim como a transformação T3, as T4 precisa ser inserida por ordem. A primeira parte, como mostra a Figura 8, é responsável por inserir os dados na tabela de elemento\_despesa\_grupo\_despesa.

Figura 8



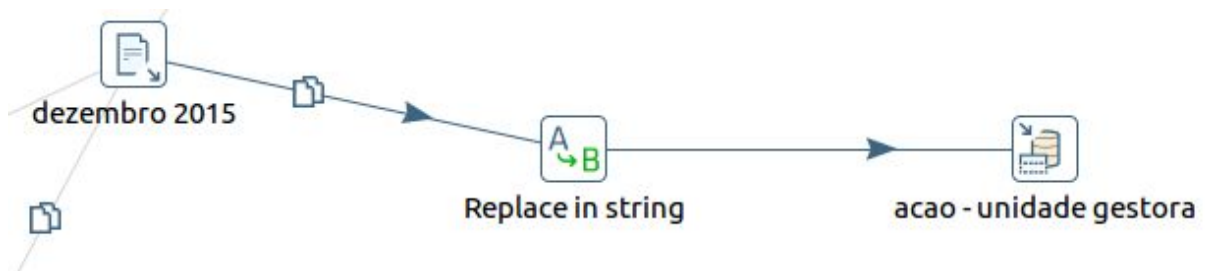
**dezembro 2015** : Essa *step* é deve ser usada para inserir as partes divididas do arquivo GastosDiretos.csv.

**elemento\_despesa\_grupo\_despesa** : Insere os dados do arquivo csv na tabela elemento\_despesa\_grupo\_despesa.

A segunda parte da transformação, representada na Figura 9, é responsável por fazer a carga dos dados na tabela acao\_unidade gestora.

0

Figura 9



**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.

**acao-unidade gestora** : Insere os dados do arquivo csv na tabela acao-unidade gestora.

A terceira parte da transformação, representada na Figura 10, é responsável por fazer a carga dos dados na tabela funcao\_subfuncao.

**funcao\_subfuncao** : Insere os dados do arquivo csv na tabela funcao\_subfuncao.

Figura 11



**acao\_programa** : Insere os dados do arquivo csv na tabela acao\_programa.

### 3. T2.ktr

A transformação T2 mostrada na Figura 12 é responsável por fazer a carga tabela de pagamento, e deve ser executada apenas depois que o resto das transformações já tenham sido feitas.

Figura 12



**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`.

**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”