

Professor:
Fabiano Tavares



EXTRAÇÃO - LAB II

ANO: 2019

LAB II – Extração de Dados

Extração de diversas fontes, incremental e total

Pré-requisitos

- ❖ JAVA;
- ❖ Instalar Pentaho Data Integration;
- ❖ Possuir o diretório de “Bases” de teste;
- ❖ Instalar MySQL;

Exercício 1

Extraindo dados de Banco de Dados

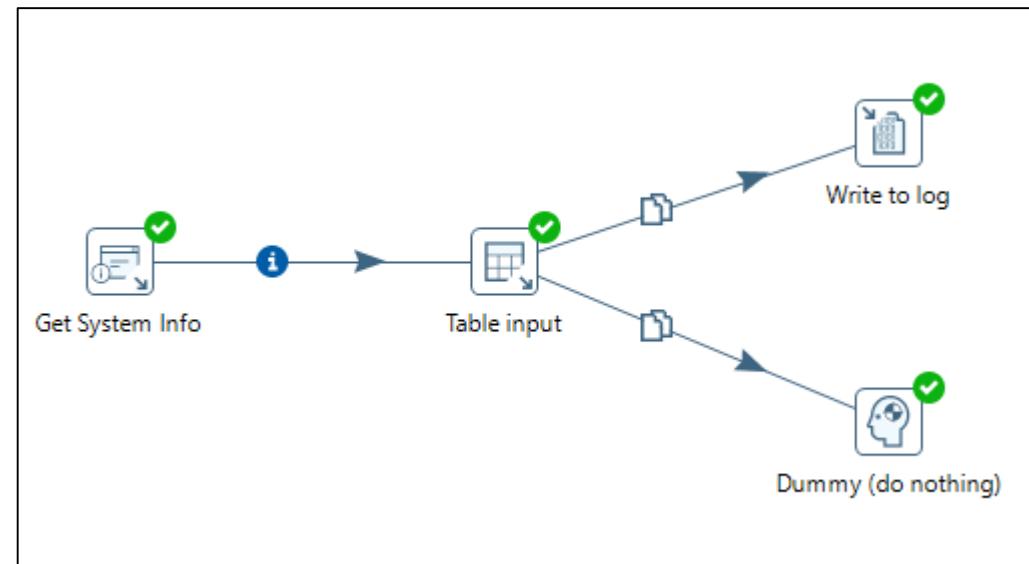
Exercício 1 – Extraindo dados de Banco de Dados

Exercício: Extraindo dados de Banco de Dados

Objetivo: Extrair dados banco de dados de forma incremental.

Depende: Para este exercício será necessário restaurar a base dados erp_oltp
(verificar arquivo de instalação do mysql);

Resultado:



Exercício 1 – Extraindo dados de Banco de Dados

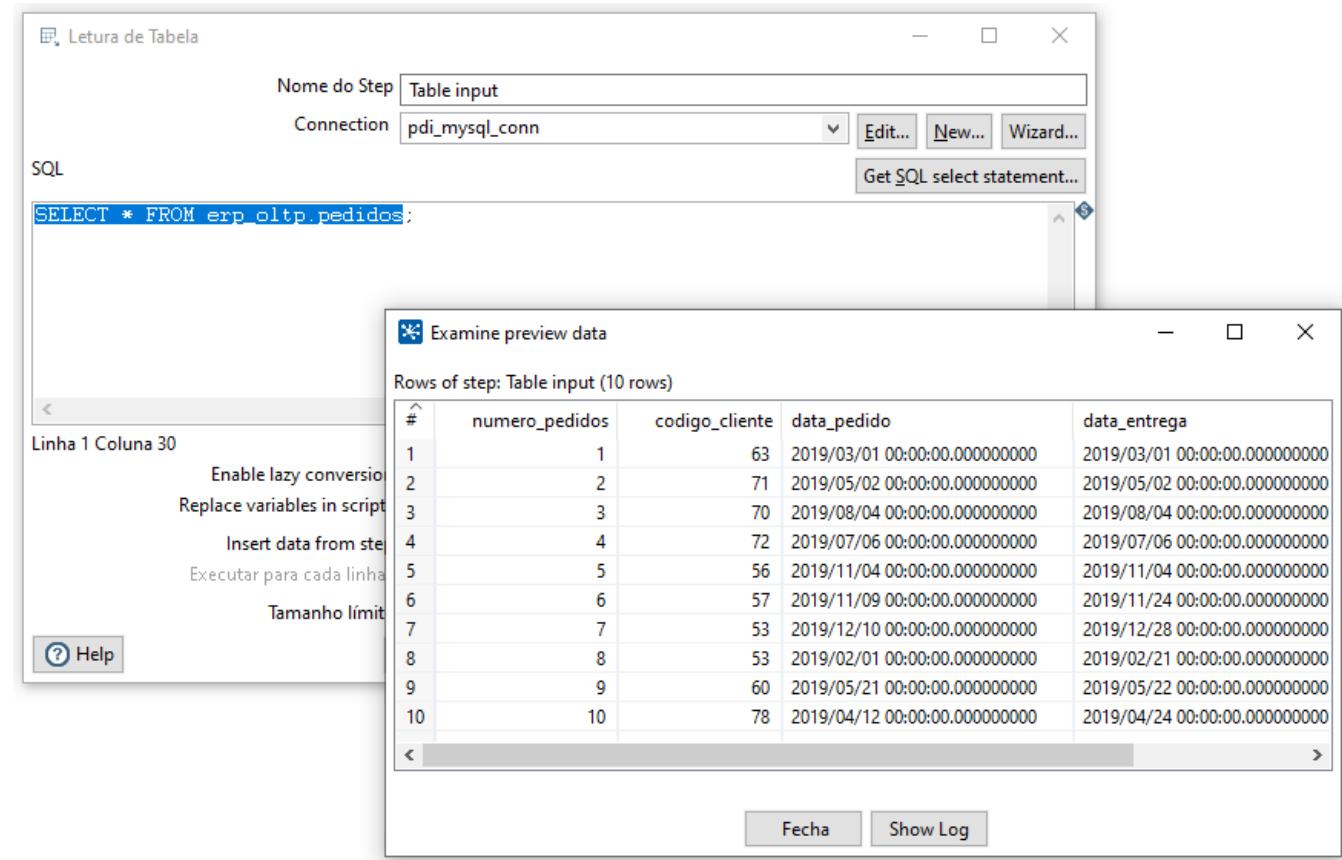
1 – Abra uma nova transformação e adicione um step *Table input*;

2 - Edite o step *Table input* e escolha a Connection *pdi_mysql_conn* (Criada no Lab I);

3 – Em SQL digite a consulta para recuperar todos os dados da tabela de pedidos;

“SELECT * FROM
erp.oltp_pedidos”

Clique em *preview data* para verificar uma prévia dos dados



The screenshot shows the PDI Data Transformer interface. The top window is titled "Letura de Tabela" (Table Reader) with the step name "Table input". It shows the connection "pdi_mysql_conn" selected. The SQL query entered is "SELECT * FROM erp.oltp_pedidos;". The bottom window is titled "Examine preview data" and displays a preview of 10 rows from the "Table input" step. The columns are labeled: #, numero_pedidos, codigo_cliente, data_pedido, and data_entrega. The data is as follows:

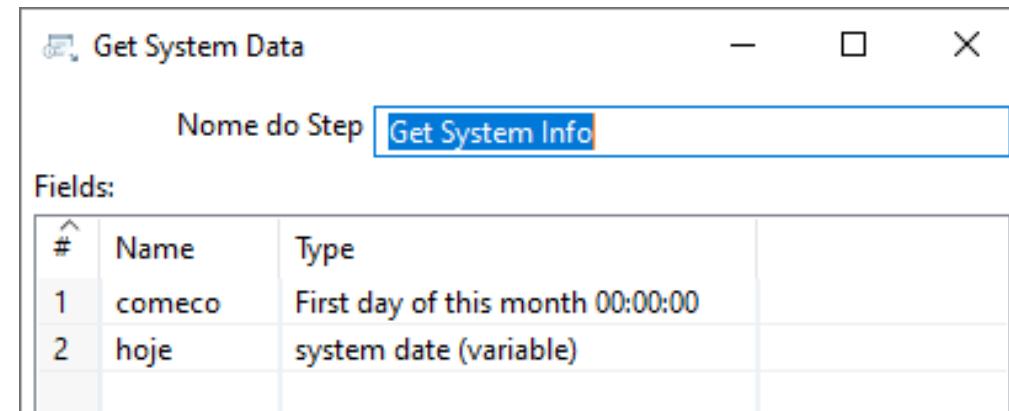
#	numero_pedidos	codigo_cliente	data_pedido	data_entrega
1	1	63	2019/03/01 00:00:00.000000000	2019/03/01 00:00:00.000000000
2	2	71	2019/05/02 00:00:00.000000000	2019/05/02 00:00:00.000000000
3	3	70	2019/08/04 00:00:00.000000000	2019/08/04 00:00:00.000000000
4	4	72	2019/07/06 00:00:00.000000000	2019/07/06 00:00:00.000000000
5	5	56	2019/11/04 00:00:00.000000000	2019/11/04 00:00:00.000000000
6	6	57	2019/11/09 00:00:00.000000000	2019/11/24 00:00:00.000000000
7	7	53	2019/12/10 00:00:00.000000000	2019/12/28 00:00:00.000000000
8	8	53	2019/02/01 00:00:00.000000000	2019/02/21 00:00:00.000000000
9	9	60	2019/05/21 00:00:00.000000000	2019/05/22 00:00:00.000000000
10	10	78	2019/04/12 00:00:00.000000000	2019/04/24 00:00:00.000000000

Exercício 1 – Extraindo dados de Banco de Dados

4 – Agora adicione os demais *Get System Info* antes da step *Table Input* e crie um hop ligando-os;

5 - Edite o step *Get System Info* e adicione dois atributos conforme figura ao lado:

- Comeco: *First day of this month*
- Hoje: *system date*



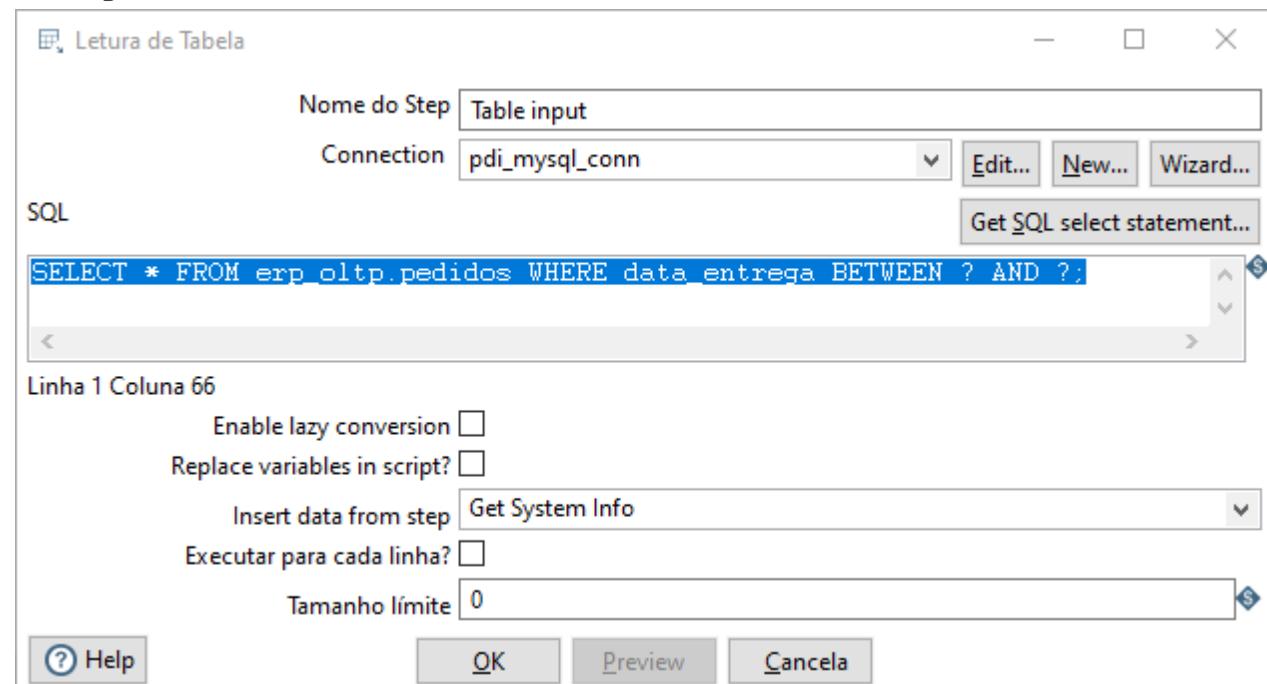
Exercício 1 – Extraindo dados de Banco de Dados

6 – Esses dois campos servirão de parâmetros para nossa extração incremental.

7 – Agora no step *Table Input* edite o SQL digite a consulta para recuperar todos os dados da tabela de pedidos de acordo com dois parâmetros:

```
"SELECT * FROM erp.oltp.pedidos WHERE data_entrega BETWEEN ? AND ?;"
```

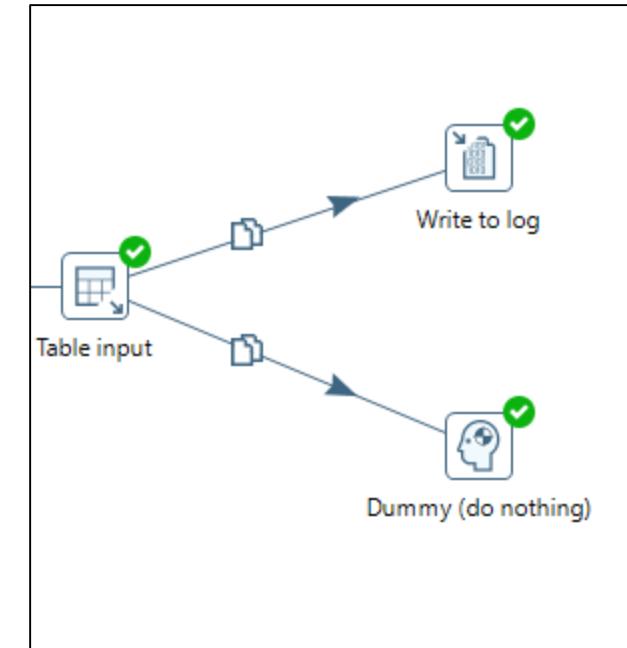
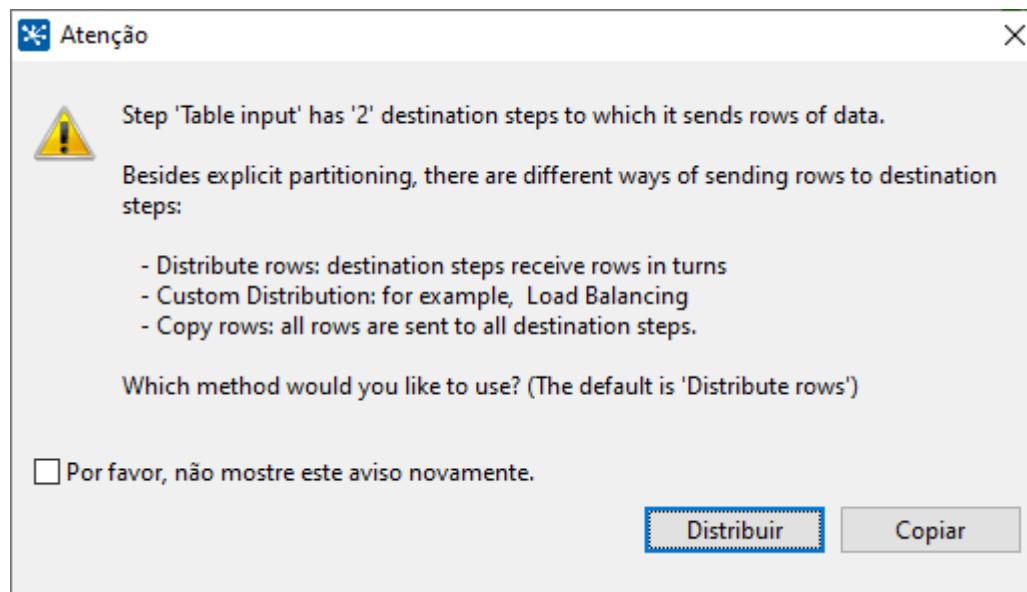
Obs: Escolha a opção “*Insert data from step*”



Exercício 1 – Extraindo dados de Banco de Dados

8 - Adicione mais dois step “Write to Log” e “Dummy”. Crie dois hops saindo do step “Table input”

9 - Ao aparecer a janela de diálogo, responda que você quer realizar uma cópia. Nesse momento você está realizando uma cópia do *stream* do *dataset*:



Exercício 1 – Extraindo dados de Banco de Dados

- 10 – No step *Write Log*, edite e clique em obter campos;
- 11 - Execute a transformação e observe a quantidade de registros lidos por cada step *Dummy* e *Write Log*;

Neste exercícios vimos uma **cópia de *stream*** e como **extrair dados de forma incremental** de uma base de dados;

Exercício 2

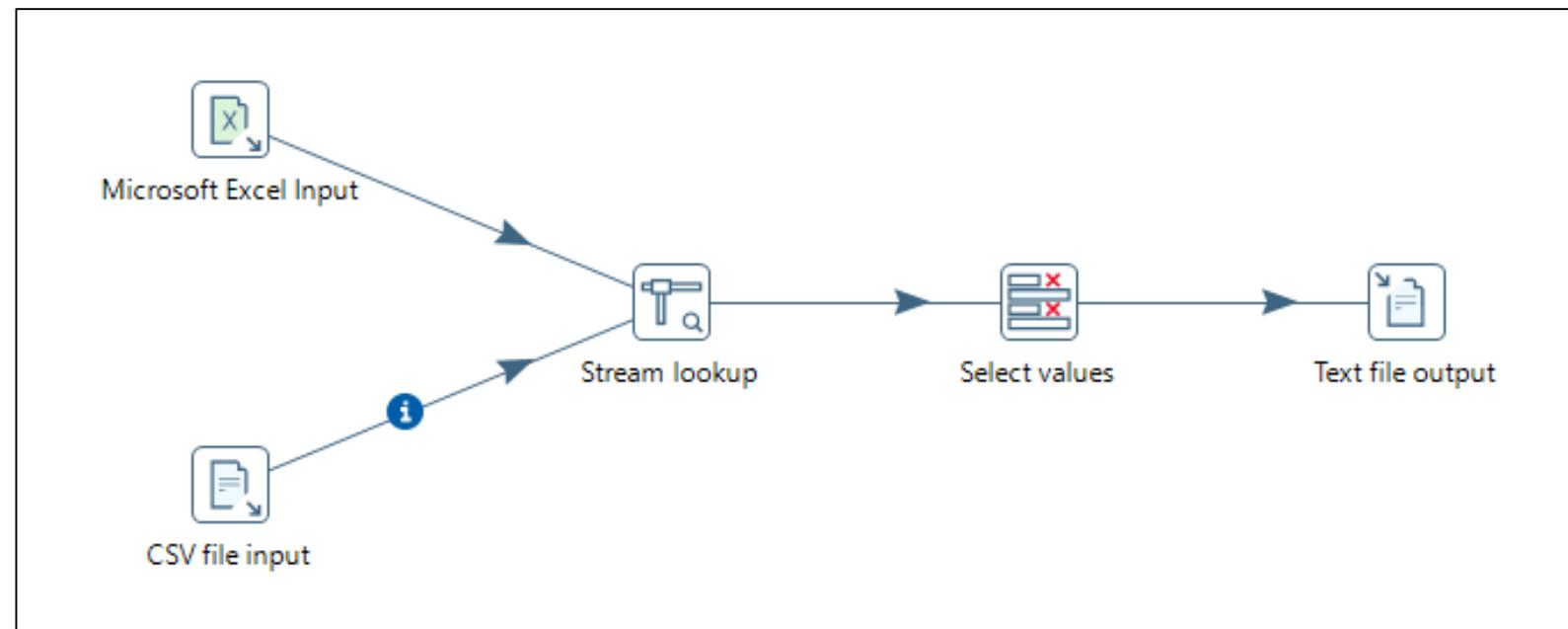
Extraindo de duas bases

Exercício 2 – Extrairindo de duas bases

Exercício: Extrairindo dados de um arquivo texto, realizando uma transformação e carregando o resultado em um arquivo texto.

Objetivo: **Extrair de bases diversas** e combiná-las.

Resultado:



Exercício 2 – Extraíndo de duas bases

1 – Para realizar esse exercício você precisará de dois arquivos armazenados na pasta *Bases*: *capes.xls* e *ies.csv*.

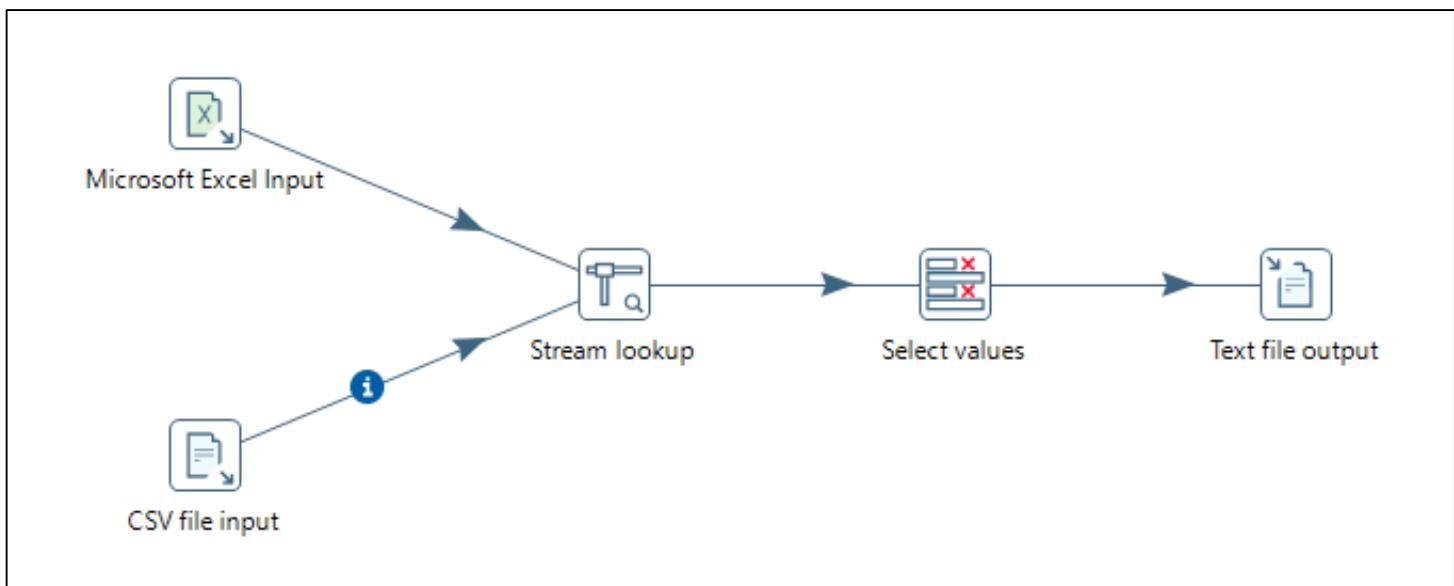
- O primeiro contém informações sobre a avaliação dos programas de Pós-Graduação das Instituições de Ensino Superior (IES) do país.
- O segundo arquivo é uma tabela de lookup contendo o código e a sigla das IES's. Vamos mostrar a extração de dados a partir de dois arquivos (Excel e CSV), uma pequena transformação e o carregamento dos dados em um arquivo texto.
- Abra a planilha e veja que o campo *ies* possui a sigla da instituição. Queremos gravar em um arquivo texto parte dos dados da planilha e o código da instituição no lugar de sua sigla.

Para iniciar, abra o Spoon e crie uma nova transformação.

Exercício 2 – Extraindo de duas bases

2 – Abra a categoria *Input* e adicione os steps *Excel Input* e *CSV file input*. Em seguida, expanda a categoria *Lookup* e adicione o step *Stream lookup*.

- Da categoria *Transform*, adicione o step *Select values*.
- Da categoria *Output*, adicione o step *Text file output*. Por fim, crie os hop's para conectar os steps, de acordo



Exercício 2 – Extraindo de duas bases

3 – Edite o step Excel input com os seguintes parâmetros:

❖ Aba Files:

- *File or directory*: localize o arquivo *capes.xls* com o botão Navegar. Em seguida, clique em *Add* para adicionar o arquivo ao grid. Para ter certeza que o arquivo foi localizado, clique no botão *Show filename(s)*.
- Clique no botão *Get sheetname* e escolha a planilha desejada. Se o nome da planilha não aparecer na lista, reveja os parâmetros da aba *Files*.
- Certifique-se que o campo Header esteja marcado (vamos precisar dele na próxima aba).
- Esse arquivo foi gravado no Linux! Mude o campo *Encoding* para UTF-8.

Exercício 2 – Extraíndo de duas bases

❖ Aba Fields

- Clique no botão *Get fields from header now* e veja todos os campos disponíveis no arquivo.
- Dê uma olhada nos dados que serão extraídos do arquivo, clicando no botão *Preview rows*. Clique Ok e salve a transformação.

4 – Edite o *step CSV file input* com os seguintes parâmetros:

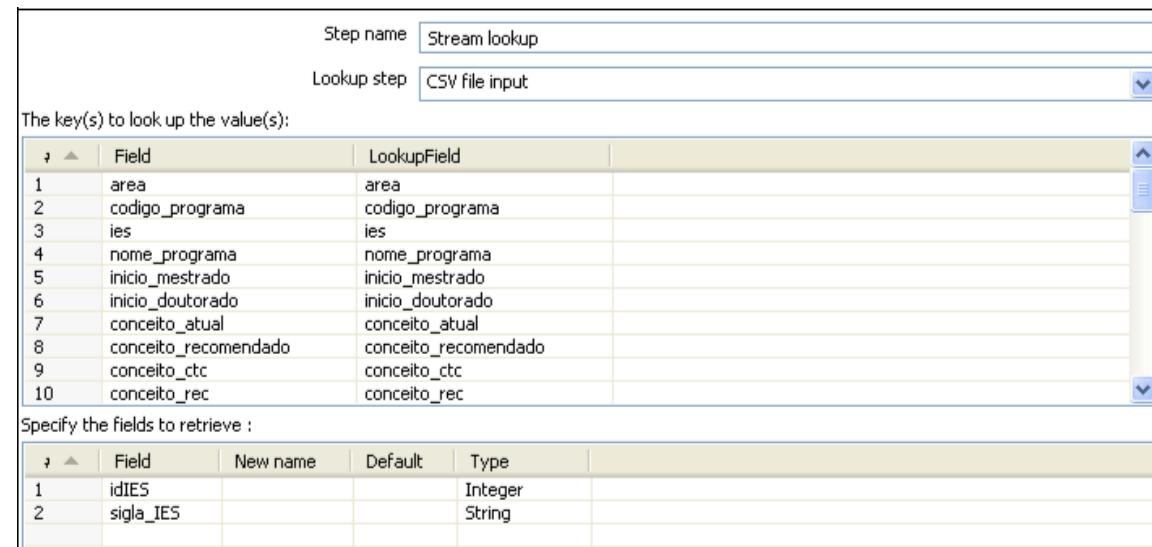
- Filename: localize o arquivo ies.csv com o botão Navegar.
- Delimiter: ; (ponto-e-vírgula)
- Desmarque a opção Lazy conversion.
- Clique no botão *Obtém campos* e veja os campos que serão lidos.
- Clique no botão *Preview* para visualizar uma amostra dos dados.
 - Na grade com os campos, diminua o tamanho do campo idIES para 1(propriedade Length).
 - Rode novamente o preview.
 - Retire o símbolo da moeda (R\$) da propriedade Currency.

Exercício 2 – Extraíndo de duas bases

5 – Antes de editar o step *Stream lookup*, dê uma olhada no fluxo de registros de entrada. Clique com o botão direito em cima do step e escolha a opção *Mostra campos de entrada*. Deverão ser exibidos 41 campos (39 da planilha e 2 do arquivo texto).

6 – Edite o step *Stream lookup* com os seguintes parâmetros:

- *Lookup step*: escolha o step *CSV file input*.
- Clique nos botões *Get fields* e *Get lookup fields*. As grades de campos deverão ter a configuração da figura.



The screenshot shows the configuration of a Stream lookup step. At the top, it displays the Step name as "Stream lookup" and the Lookup step as "CSV file input". Below this, there are two tables:

The key(s) to look up the value(s):

#	Field	LookupField
1	area	area
2	codigo_programa	codigo_programa
3	ies	ies
4	nome_programa	nome_programa
5	inicio_mestrado	inicio_mestrado
6	inicio_doutorado	inicio_doutorado
7	conceito_atual	conceito_atual
8	conceito_recomendado	conceito_recomendado
9	conceito_ctc	conceito_ctc
10	conceito_rec	conceito_rec

Specify the fields to retrieve :

#	Field	New name	Default	Type
1	idIES			Integer
2	siglaIES			String

Exercício 2 – Extraindo de duas bases

- ❖ Na grade de cima, remova todos os campos, deixando apenas o campo ies. Em Lookup Field escolha o campo *siglaIES*.
 - ❖ Na grade de baixo, remova o campo *siglaIES*. No campo *idIES*, digite o valor *id_instituicao* na propriedade New name. As grades deverão ter a configuração abaixo. Salve a transformação.

Exercício 2 – Extraindo de duas bases

- ❖ Na grade de cima, remova todos os campos, deixando apenas o campo ies. Em Lookup Field escolha o campo *siglaIES*.
- ❖ Na grade de baixo, remova o campo *siglaIES*. No campo *idIES*, digite o valor *id_instituicao* na propriedade New name. As grades deverão ter a configuração abaixo. Salve a transformação.

The key(s) to look up the value(s):				
#	Field	LookupField		
1	ies	siglaIES		
Specify the fields to retrieve :				
#	Field	New name	Default	Type
1	idIES	id_instituicao		Integer

- ❖ Na área de trabalho, clique com o botão direito em cima do step *Stream lookup* e escolha a opção *Mostra campos de saída*. Observe que temos agora 40 campos (o último é o campo que denota o *id* da instituição).

❖ 7 – Suponha que não precisamos de todos os campos vindos da planilha. Além disso, queremos modificar o nome dos campos que serão carregados ao final do processo de transformação. Edite o step *Select values* com os seguintes parâmetros:

- Aba Meta-data
 - Fieldname: escolha o campo area digite id_area na propriedade Rename to e escolha Integer na propriedade Type.
 - Repita a operação com o campo codigo_programa, com o nome id_programa. A grade deverá ter a configuração abaixo.

Fields to alter the meta-data for :			
#	Fieldname	Rename to	Type
1	area	id_area	Integer
2	codigo_programa	id_programa	-

❖ Aba Remove

- Clique no botão *Get fields to remove* e exclua os seguintes campos:
 - area, codigo_programa, ies, nome_programa, inicio_mestrado, inicio_doutorado, conceito_recomendado e id_instituicao.
 - A lista de campos que queremos excluir deverá ter a configuração abaixo.

Fields to remove :	
#	Fieldname
1	conceito_atual
2	conceito_ctc
3	conceito_rec
4	q2_q5
5	itens_q2_q5
6	itens_q3_q4
7	docentes_ano_2001_2003
8	docentes_ano_2004_2006
9	docentes_ano_2001_2006
10	Teses_2001_2003

- Salve a transformação e veja os campos de saída do step, clicando com o botão direito em cima dele.

- ❖ 8 – Vamos editar o step Text file output com os seguintes parâmetros:
 - Aba *File*
 - *Filename:* <path>\resultado
 - Aba *Fields*
 - Clique no botão *Obtém campos* e veja os campos que serão gravados. Alterações de formato, tamanho, etc. poderão ser feitas diretamente na grade.
 - Clique no botão *Minimal width* e veja que o step fornece um formato padrão para os campos.
- ❖ 9 – Execute a transformação e veja o arquivo gerado.

❖ O que fizemos nessa transformação?

- Mostramos como extrair dados de dois arquivos com formatos diferentes (.xls e .csv).
- Fizemos duas transformações nos dados extraídos: trocamos o nome da IES por seu código e retiramos os campos que não queríamos gravar. Além disso, alteramos os metadados de dois campos.
- Carregamos o resultado da transformação em um arquivo texto.

Exercício 3

Extraindo dados de um conjunto de arquivos

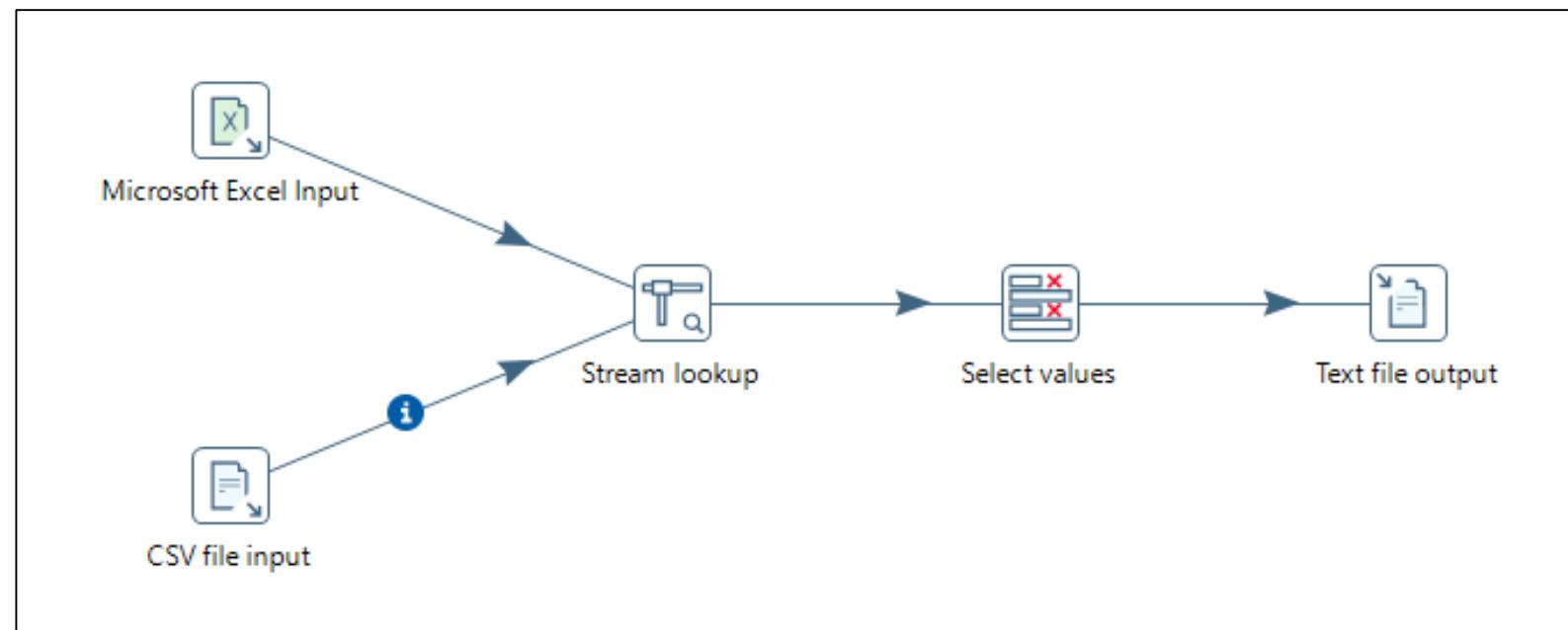
Exercício 3 – Extrair dados de um conjunto de arquivos

Exercício: Extrair dados de um conjunto de arquivos em diretório especificado.

Objetivo: Leitura de múltiplos arquivos.

Depende: Exercício 2

Resultado:



Exercício 3 – Extraindo dados de um conjunto de arquivos

- 1 – Para esse exercício vamos utilizar as planilhas armazenadas na pasta `/etl_labs/base/planilhas`. Crie uma nova transformação e copie/cole os steps e hops da transformação do exercício anterior.
- 2 – Edite o step *Excel input*. Na grade de arquivos selecionados da aba *Files*, marque o arquivo *capes.xls* e clique no botão *Delete*.
- 3 – Acesse a pasta *planilhas* e adicione todos os arquivos na grade usando o botão *Add* (Ok...não precisa adicionar todos. São 45 arquivos. Verifique se o step carregou os arquivos clicando no botão *Show filename(s)*).

Exercício 3 – Extraindo dados de um conjunto de arquivos

Selected files:	#	File/Directory	Wildcard (RegExp)	Required
	1	E:\pdi_labs\bases\planilhas\01_Mat_Estat.xls		
	2	E:\pdi_labs\bases\planilhas\02_Computacao.xls		
	3	E:\pdi_labs\bases\planilhas\03_Fisica_Astronomia.xls		
	4	E:\pdi_labs\bases\planilhas\04_Quimica.xls		
	5	E:\pdi_labs\bases\planilhas\05_Geociencias.xls		
	6	E:\pdi_labs\bases\planilhas\06_Bio1.xls		
	7	E:\pdi_labs\bases\planilhas\07_Ecologia.xls		
	8	E:\pdi_labs\bases\planilhas\08_Bio2.xls		
	9	E:\pdi_labs\bases\planilhas\09_Bio3.xls		
	10	E:\pdi_labs\bases\planilhas\10_Eng1.xls		
	11	E:\pdi_labs\bases\planilhas\11_Artes.xls		
	12	E:\pdi_labs\bases\planilhas\12_Eng2.xls		
	13	E:\pdi_labs\bases\planilhas\13_Eng3.xls		
	14	E:\pdi_labs\bases\planilhas\14_Eng4.xls		
	15	E:\pdi_labs\bases\planilhas\15_Med1.xls		
	16	E:\pdi_labs\bases\planilhas\16_Med2.xls		
	17	E:\pdi_labs\bases\planilhas\17_Med3.xls		
	18	E:\pdi_labs\bases\planilhas\18_Odontologia.xls		
	19	E:\pdi_labs\bases\planilhas\19_Farmacia.xls		

4 – Na aba *Sheets* exclua a planilha do exercício anterior e clique em *Get sheetname(s)* para pesquisar pelas planilhas de **todos** os arquivos selecionados no passo anterior. Escolha a planilha *plan1*.

Exercício 3 – Extraindo dados de um conjunto de arquivos

5 - Na aba *Content* certifique-se que o campo *Header* esteja marcado e que o *Encoding* do arquivo seja UTF-8 ou ISO-8859-1 (experimente colocar outra codificação e veja o que irá ocorrer).

6 – Na aba *Fields* clique no botão *Get fields from header now*, lembrando-se de limpar a lista de campos do exercício anterior (confirme na janela de diálogo).

7 – Salve a transformação e execute-a. Veja o resultado do arquivo gerado pela transformação.

Exercício 4

Extraindo dados usando expressões regulares

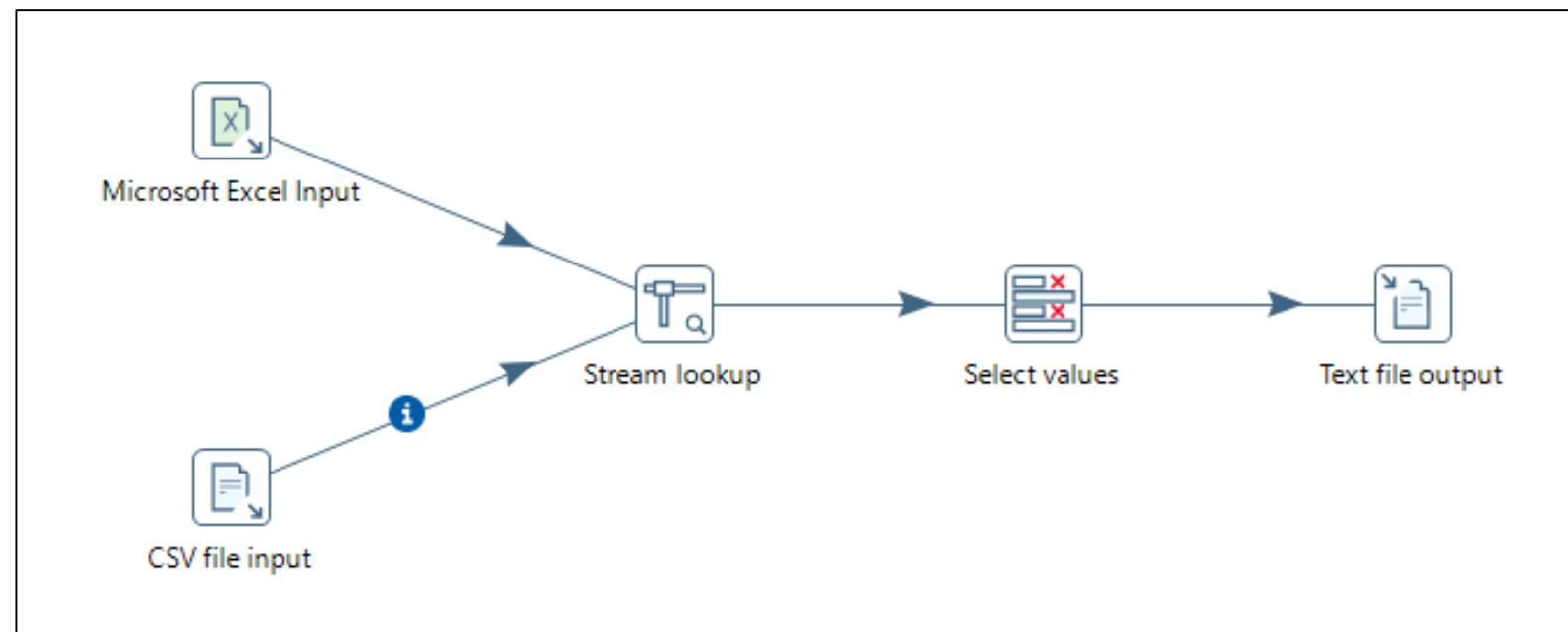
Exercício 4 – Extraindo dados usando expressões regulares

Exercício: Extraindo dados de um conjunto de arquivos em diretório especificado.

Objetivo: Leitura de múltiplos arquivos com expressão regular.

Depende: Exercício 3

Resultado:



Exercício 4 – Extraindo dados usando expressões regulares

- ❖ Em vários steps do PDI podemos usar expressões regulares:
- ❖ Exemplos:

Expressão regular	Resultado	Exemplos
.*\txt	Qualquer arquivo .txt	Arquivo.txt
test(19 20)\d\d-(0[19] 1[012])\txt	Qualquer arquivo começando com test, seguido por uma data usando o formato yyyy-mm	test2009-12.txt test2009-01.txt
(?i)test.+txt	Qualquer arquivo .txt começando com test escrito em maiúsculo ou minúsculo	TeSTcaseinsensitive.txt

- ❖ Para mais expressões regulares: <http://aprenda.vidageek.net/aprenda/regex>

Exercício 4 – Extraindo dados usando expressões regulares

- 1 – A solução adotada no exercício anterior pode ser refeita para evitar a inclusão de cada arquivo manualmente (imagine uma aplicação real com milhares de arquivos). Crie uma nova transformação e copie/cole os steps e hops da transformação do exercício anterior e edite o step *Excel input*.
- 2 – Na aba *Files*, exclua todos os arquivos da grade *Selected files*. No campo *File ou directory* digite o caminho para a pasta onde as planilhas estão armazenadas (ex.: E:\etl_labs\bases\planilhas\). Adicione o caminho à lista de arquivos clicando no botão *Add*.
- 3 – No campo *Selected files*, digite a seguinte expressão na propriedade *Wildcard*:

`.*\.xls`

Exercício 4 – Extraindo dados usando expressões regulares

Selected files:	#	File/Directory	Wildcard (RegExp)	Required
	1	E:\pdi_labs\bases\planilhas\	.*\.xls	

4 – Para garantir que os arquivos serão lidos usando a expressão regular, clique no botão *Show filename(s)* e veja a lista de arquivos que serão lidos. Salve a transformação e execute-a, observando o arquivo gerado.

5 – Salve a transformação e execute-a. Veja o resultado do arquivo gerado pela transformação.

Exercício 5

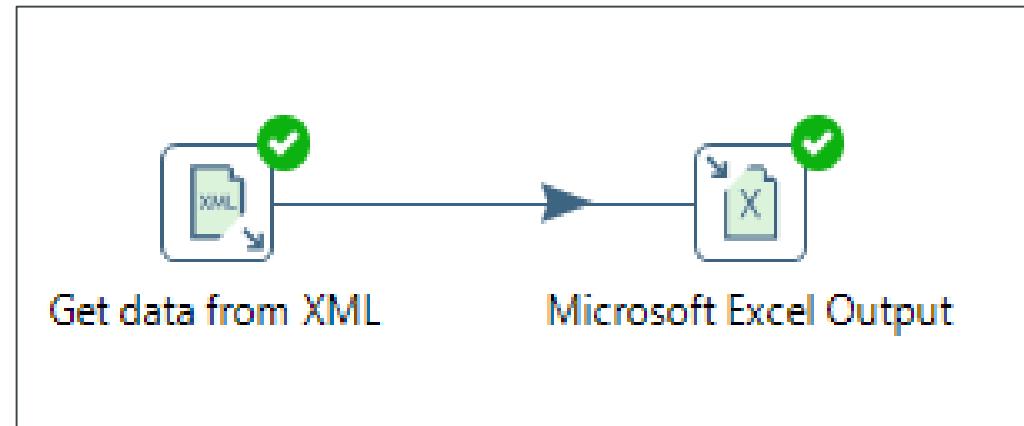
Extraindo de um arquivo XML

Exercício 5 - Extraíndo de um arquivo XML

Exercício: Extraíndo uma lista com dados de países de um arquivo XML

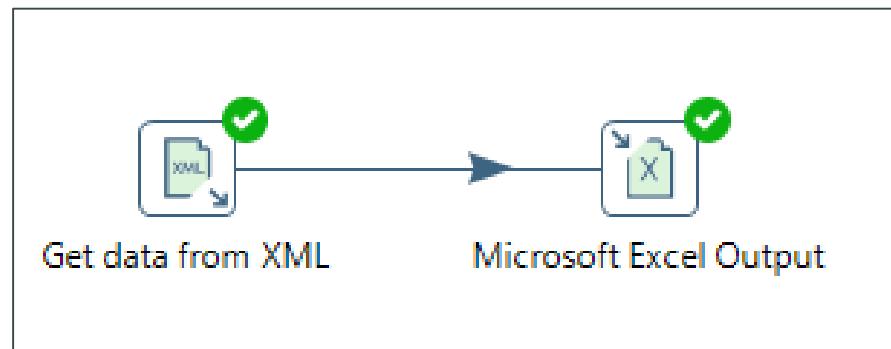
Objetivo: Leitura de um arquivos XML.

Resultado:



Exercício 5 - Extraindo de um arquivo XML

- 1- Para esse exercício vamos utilizar o arquivo *countries.xml* que se encontra na pasta */etl_labs/bases/xml*. Abra o arquivo e observe a sua estrutura.
- 2- Crie uma nova transformação e adicione os steps *Get data from XML* e *Excel output*. Crie um hop ligando os steps.



Exercício 5 - Extraíndo de um arquivo XML

3- Edite o step Get data from XML, com os seguintes parâmetros:

❖ Aba File

- *File or directory*: encontre o arquivo *countries.xml* e adicione à lista.

❖ Aba Content

- Clique no botão *Get Xpath nodes* e selecione */world/country*
- *Get Xpath nodes* é o caminho percorrido nos nós do XML

❖ Aba Fields

- Preencha a grade de acordo com a figura abaixo.

Exercício 5 - Extraiendo de um arquivo XML

Get data from XML

Nome do Step Get data from XML

File Content Fields Additional output fields

#	Name	XPath	Element	Result type	Type
1	country	name	Node	Value of	String
2	capital	capital	Node	Value of	String
3	language	language/name	Node	Value of	String
4	isofficial	language/isofficial	Attribute	Value of	String
5	percentage	language/percentage	Node	Value of	Number
6					

Get fields

Help OK Preview rows Cancela

Exercício 5 - Extraindo de um arquivo XML

4 - Clique em *Preview rows* para visualizar uma prévia dos dados extraídos:

Examine preview data

Rows of step: Get data from XML (100 rows)

#	country	capital	language	isofficial	percentag
1	Afghanistan	Kabul	Pashto	T	524,
2	Albania	Tirana	Albanian	T	979,
3	Algeria	Alger	Arabic	T	86,
4	American Samoa	Fagatogo	Samoan	T	906,
5	Andorra	Andorra la Vella	Spanish	F	446,
6	Angola	Luanda	Ovimbundu	F	372,
7	Anguilla	The Valley	English	T	0,
8	Antigua and Barbuda	Saint Johns	Creole English	F	957,
9	Argentina	Buenos Aires	Spanish	T	968,
10	Armenia	Yerevan	Armenian	T	934, ▾

< >

Fecha **Show Log**

5 - Salve a transformação e edite o *Step Excel output* para gravar o resultado em uma planilha.

6 - Rode a transformação e veja os dados dos países em formato de planilha.

Exercício 6

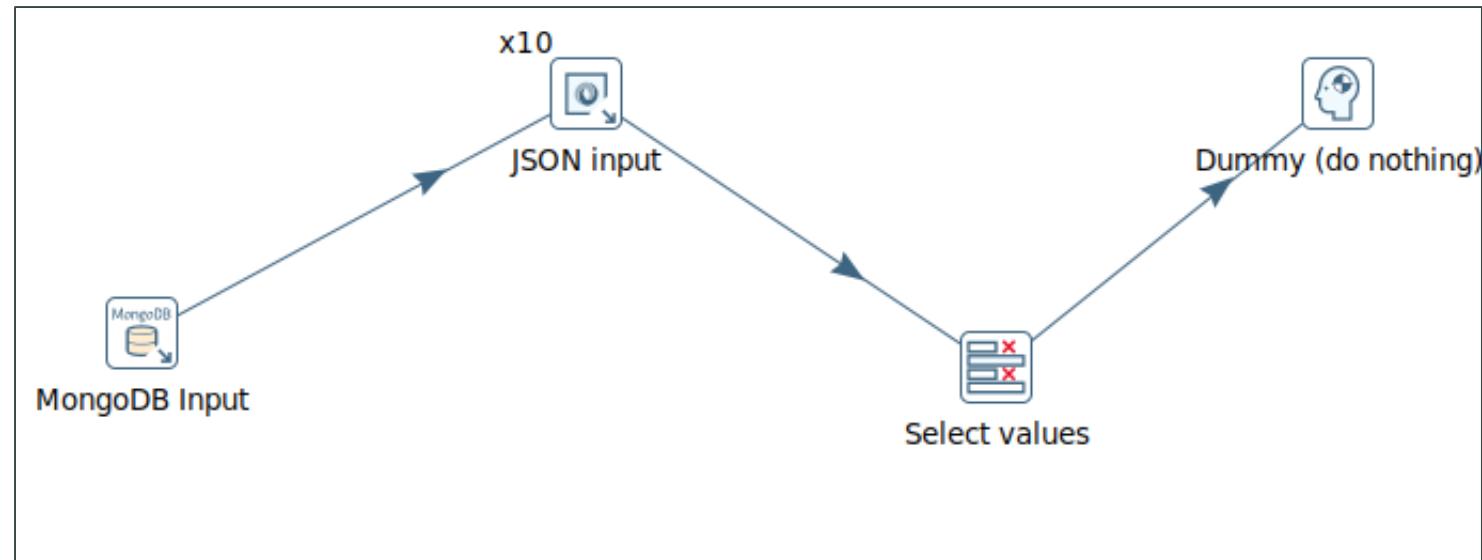
Extraindo dados de Banco Não Relacional

Exercício 6 - Extraindo dados de Banco Não Relacional

Exercício: Extraindo dados do banco MongoDB e lendo entrada Json

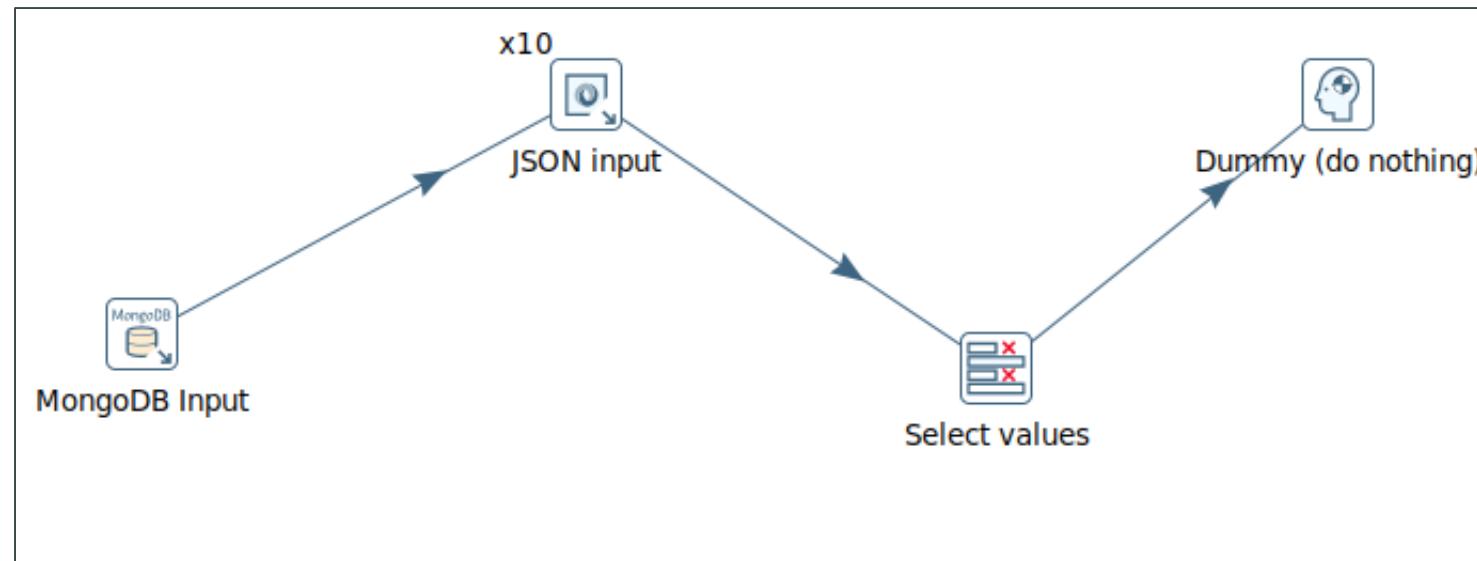
Objetivo: Leitura de um banco não relacional e paralelismo.

Resultado:



Exercício 6 - Extraindo dados de Banco Não Relacional

- ❖ Crie uma nova transformação. Insira os seguintes *steps* e *hops* conforme figura abaixo:

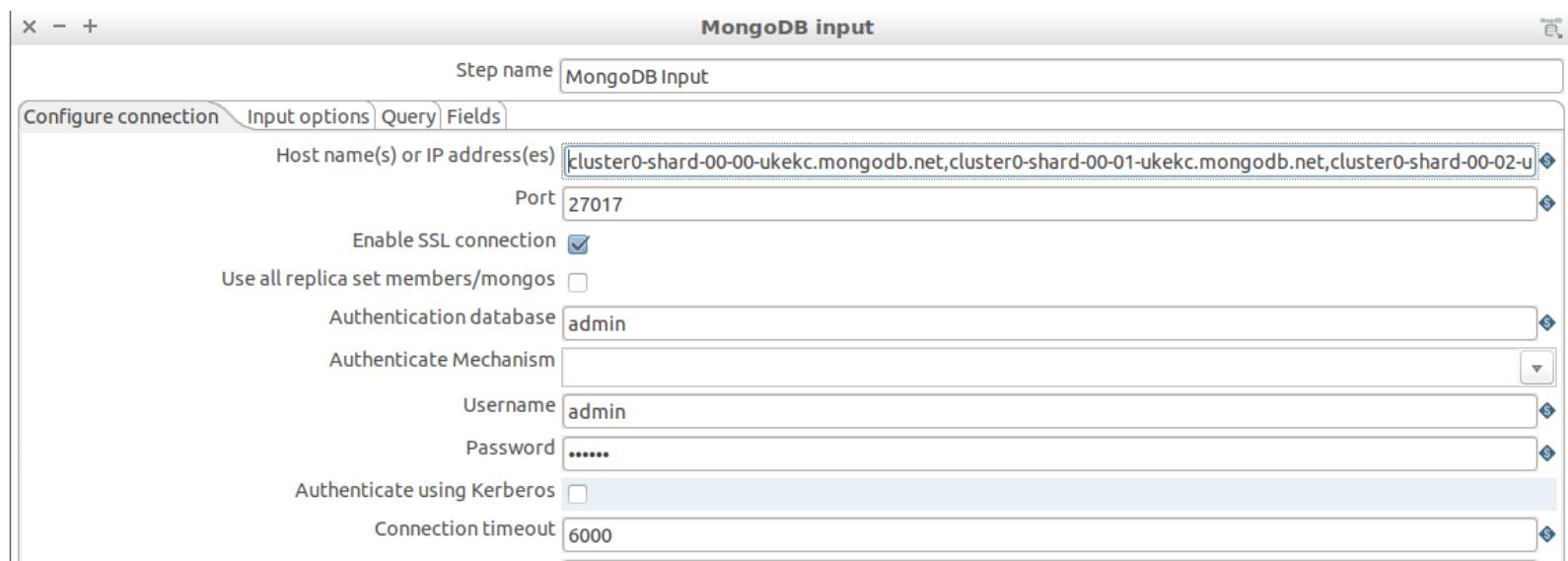


Exercício 6 - Extraindo dados de Banco Não Relacional

❖ *MongoDb Input:*

■ *Aba Configure connection:*

- *Hosts names:* cluster0-shard-00-00-ukekc.mongodb.net,cluster0-shard-00-01-ukekc.mongodb.net,cluster0-shard-00-02-ukekc.mongodb.net
- *Port:*27017
- *Enable SS Connection:* sim
- *Username:* admin
- *Password:* asd123

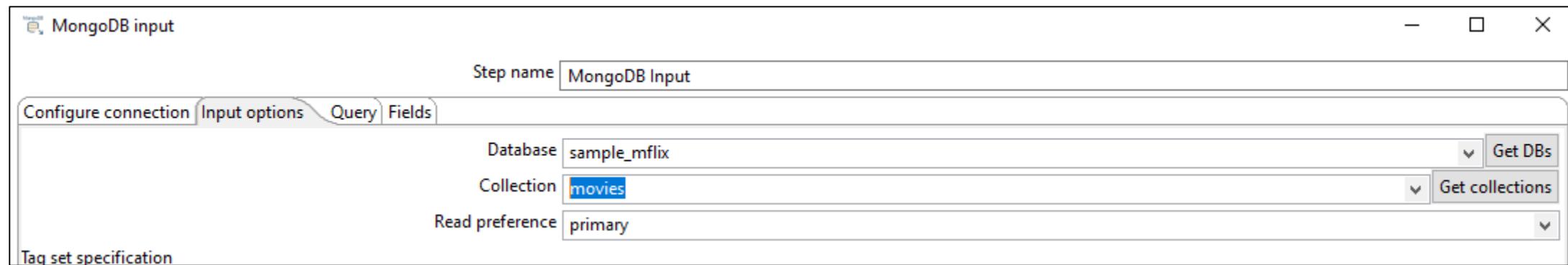


Exercício 6 - Extraindo dados de Banco Não Relacional

❖ *MongoDb Input:*

■ *Aba Input Options:*

- Clique no botão *GetDBs* e escolha o database: *sample_mflix*
- Clique no botão *GetCollection* e escolha a collection: *movies*



Exercício 6 - Extraindo dados de Banco Não Relacional

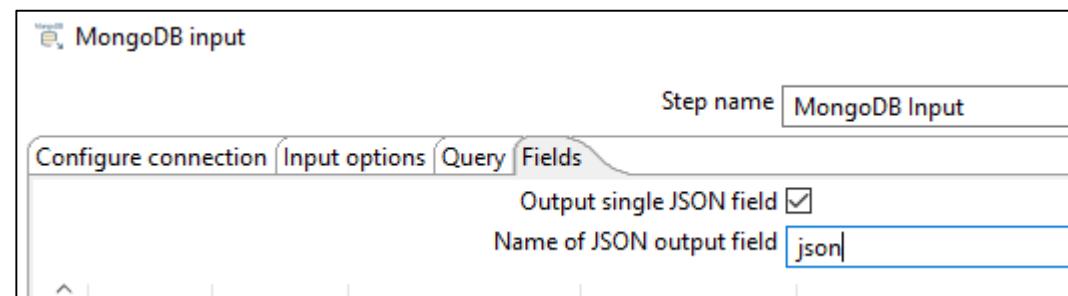
❖ *MongoDb Input:*

- *Aba Query:*

- *Query Expression: {}*

- *Aba Fields:*

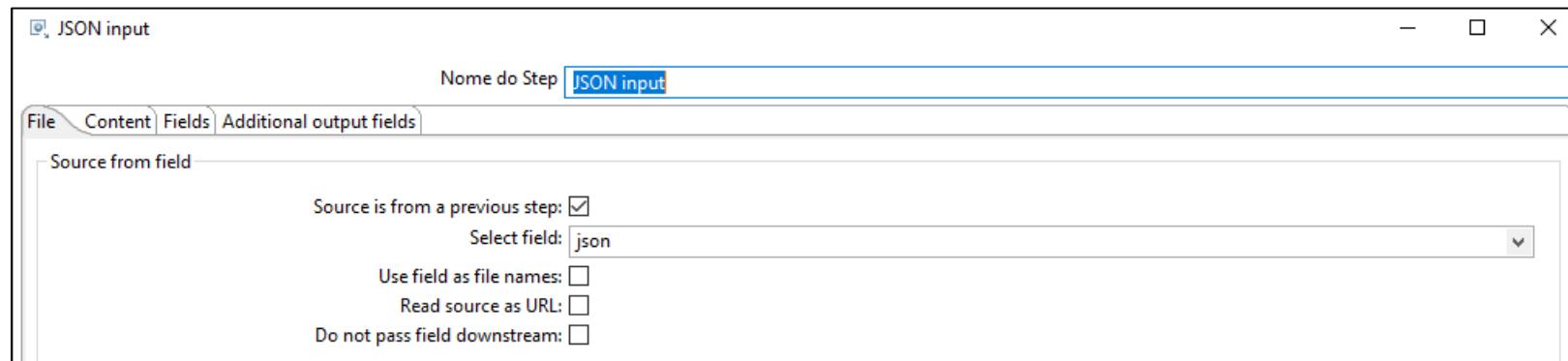
- *Output single JSON Field: selecionar a opção*
 - *Name of json output field: json*



Exercício 6 - Extraindo dados de Banco Não Relacional

❖ *JSON Input:*

■ Aba File:



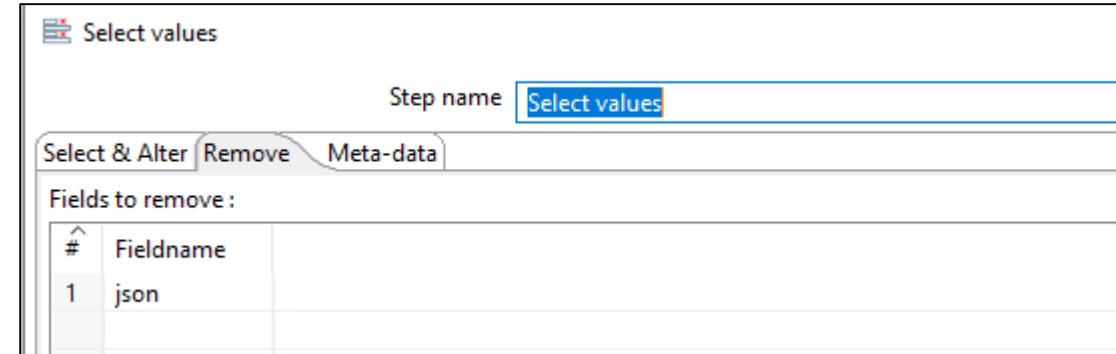
■ Aba *Fields*:

#	Name	Path	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type	Repeat
1	title	\$.title	String							nenhum	N
2	year	\$.year	String							nenhum	N
3	type	\$.type	String							nenhum	N

Exercício 6 - Extraiendo dados de Banco Não Relacional

❖ *Select Values:*

- *Aba Remove:*



❖ *Execute e veja o preview no step Dummy*

#	title	year	type
1	Blacksmith Scene	1893	movie
2	A Corner in Wheat	1909	movie
3	The Great Train Robbery	1903	movie
4	The Italian	1915	movie
5	Winsor McCay, the Famous Cartoonist of the N.Y. Herald and His Moving Comics	1911	movie
6	Traffic in Souls	1913	movie
7	Hell's Hinges	1916	movie
8	Gertie the Dinosaur	1914	movie
9	In the Land of the Head Hunters	1914	movie
10	Where Are My Children?	1916	movie
11	The Perils of Pauline	1914	movie
12	The Birth of a Nation	1915	movie
13	The Poor Little Rich Girl	1917	movie

Exercício 7

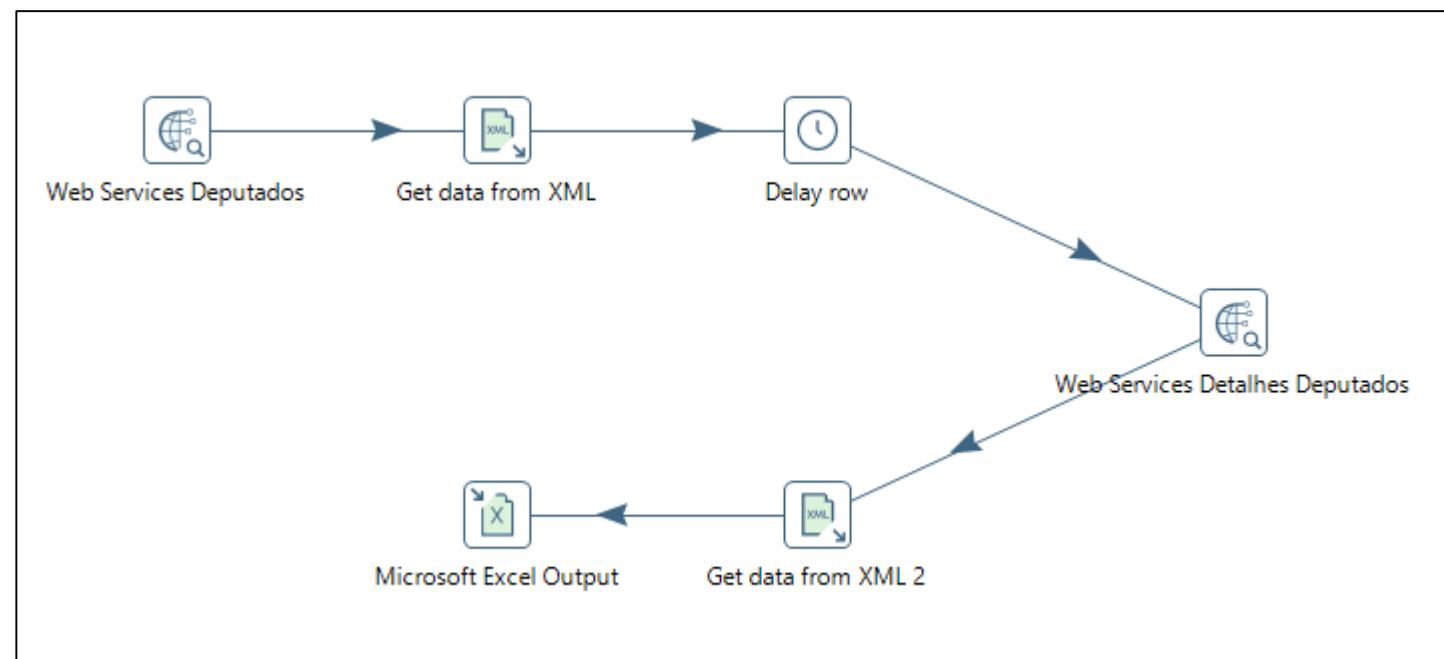
Extraindo dados de Web Service

Exercício 7 - Extraindo dados de Web Service

Exercício: Extraindo dados de um Web Service com extração de um XML

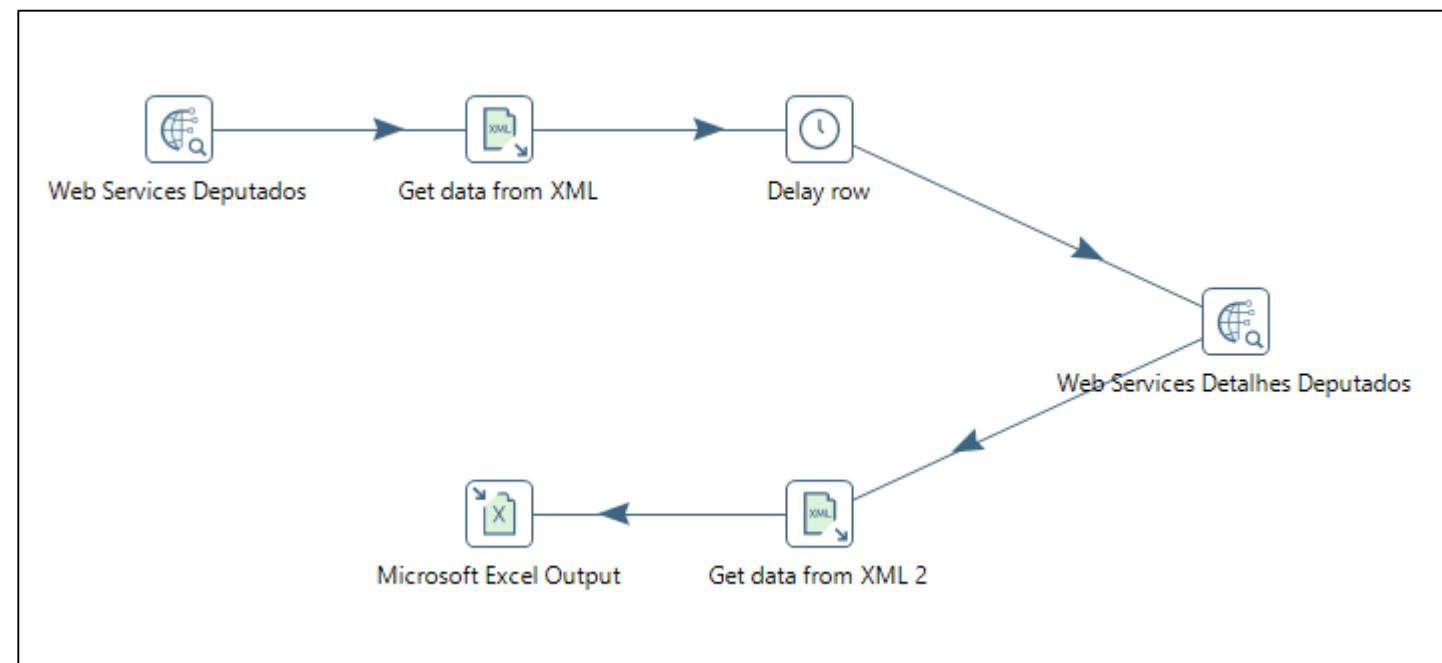
Objetivo: Interoperabilidade entre sistemas.

Resultado:



Exercício 7 - Extraindo dados de Web Service

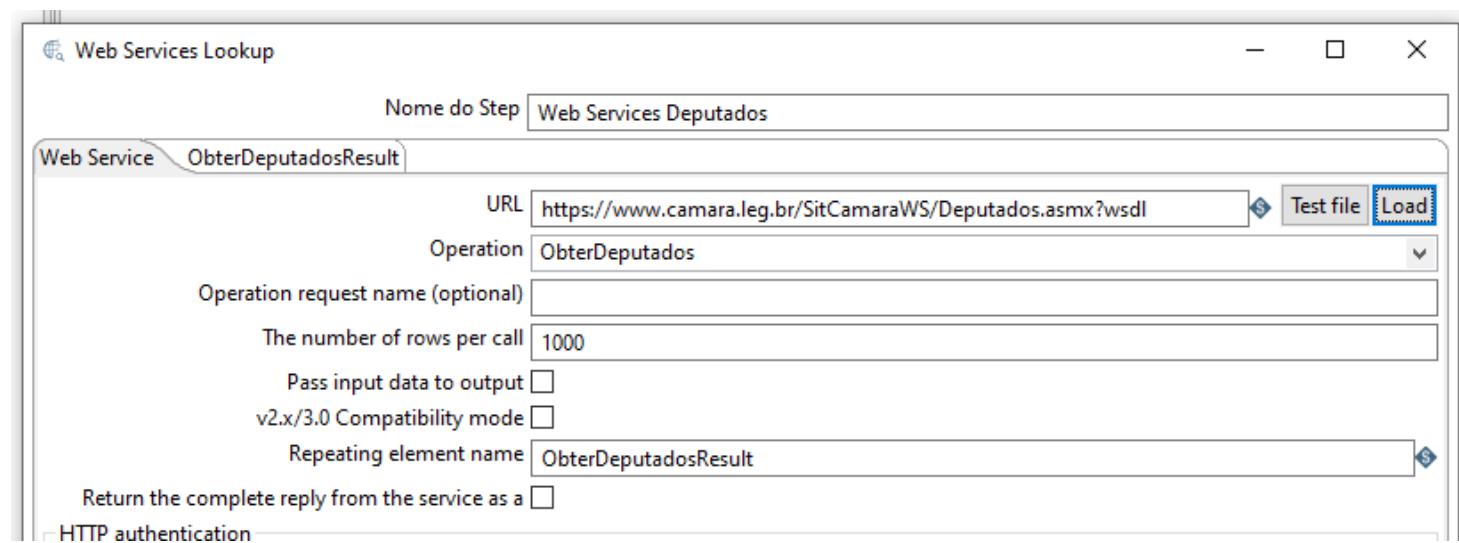
1 - Procure na aba de *Design* os seguintes *steps* e construa conforme a figura abaixo:



Exercício 7 - Extraiendo dados de Web Service

2 - No primeiro step Web Service Lookup configure a URL do Web Service

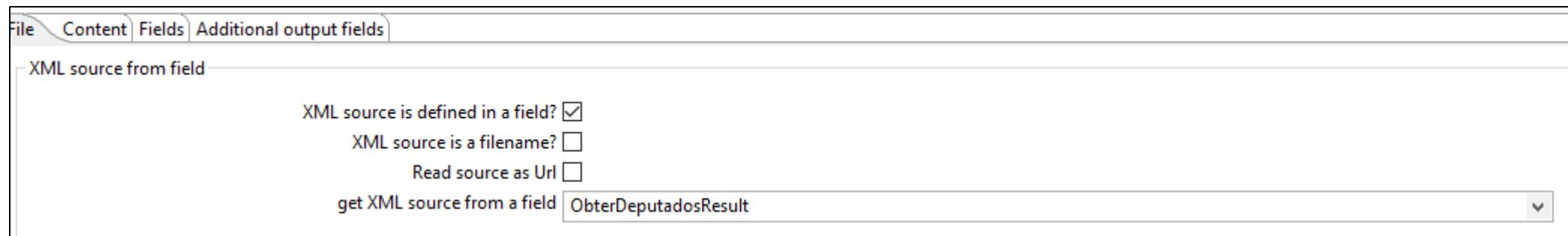
- Na aba *Web Service*:
 - *https://www.camara.leg.br/SitCamaraWS/Deputados.asmx?wsdl* e clique no botão *load*
 - Escolha a *Operation*: opção *ObterDeputados*
 - Em *Repeating elemento name*: *ObterDeputadosResult*
- Na aba próxima Aba clique em *obter campos*



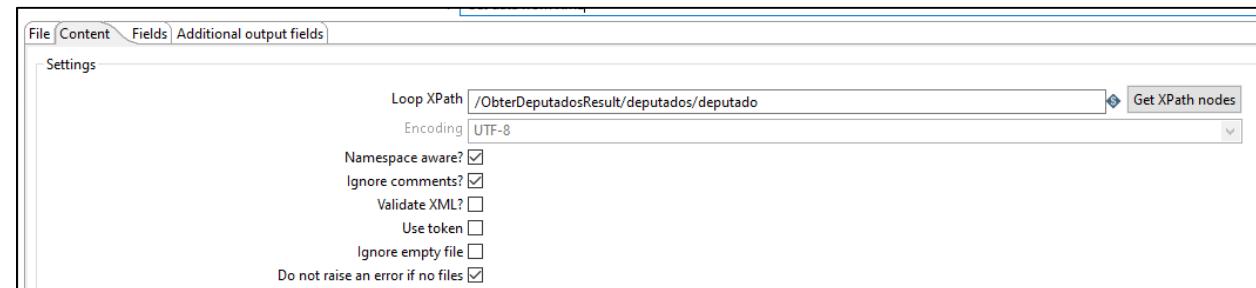
Exercício 7 - Extraindo dados de Web Service

3 – Como Web Service retorna um XML o próximo step é *Get data from XML*

- Escolha *XML source is defined in a field* na aba *File*;
- Escolha o *field* *ObterDeputadosResult*



- Na aba *Content* defina o Loop Xpath:
 - /ObterDeputadosResult/deputados/deputado
- Na aba *Fields* clique em *Get fields*
- Adicionar fragmento de XML do próximo slide.



```
<ObterDeputadosResult>
  <deputados>
    <deputado>
      <ideCadastro>73701</ideCadastro>
      <codOrcamento>1310</codOrcamento>
      <condicao>Titular</condicao>
      <matricula>291</matricula>
      <idParlamentar>1637158</idParlamentar>
      <nome>BENEDITA SOUZA DA SILVA SAMPAIO</nome>
      <nomeParlamentar>BENEDITA DA SILVA</nomeParlamentar>
      <urlFoto>http://www.camara.gov.br/internet/deputado/bandep/73701.jpg</urlFoto>
      <sexo>feminino</sexo>
      <uf>RJ</uf>
      <partido>PT</partido>
      <gabinete>330</gabinete>
      <anexo>4</anexo>
      <fone>3215-5330</fone>
      <email>dep.beneditadasilva@camara.leg.br</email>
      <comissoes>
        <titular/>
        <suplente/>
      </comissoes>
    </deputado>
  </deputados>
</ObterDeputadosResult>
```

Exercício 7 - Extraindo dados de Web Service

4 – Você pode clicar em *preview* (barra de ferramentas) e visualizar os dados do Web Service

Examine preview data

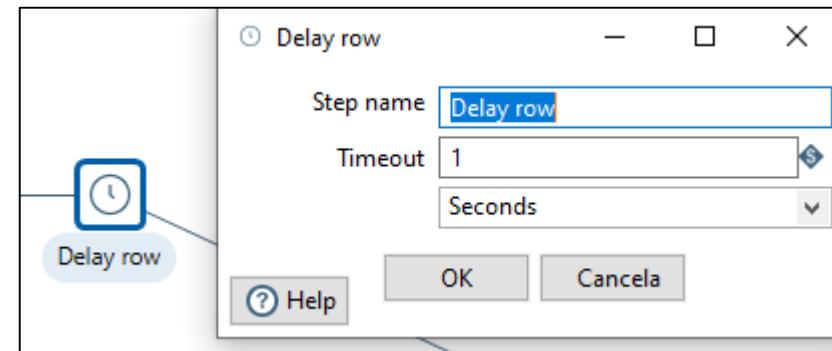
Rows of step: Get data from XML (10 rows)

#	ObterDeputadosResult	ideCadastro	codOrcamento	condicao	matricula	idParlamentar	nome	nomeParlamentar
1	<ObterDeputadosResult xmlns=...	73701	1310	Titular	291	1637158	BENEDITA SOUZA DA SILVA SAMPAIO	BENEDITA DA SILVA
2	<ObterDeputadosResult xmlns=...	73696	1880	Titular	471	1637159	ANGELA REGINA HEINZEN AMIN HELOU	ANGELA AMIN
3	<ObterDeputadosResult xmlns=...	73801	1690	Titular	154	1637160	RENILDO VASCONCELOS CALHEIROS	RENILDO CALHEIROS
4	<ObterDeputadosResult xmlns=...	74848	1775	Titular	305	1637161	JANDIRA FEGHALI	JANDIRA FEGHALI
5	<ObterDeputadosResult xmlns=...	74459	1935	Titular	109	1637162	ÁTILA FREITAS LIRA	ÁTILA LIRA
6	<ObterDeputadosResult xmlns=...	73466	1876	Titular	460	1637163	RUBENS BUENO	RUBENS BUENO
7	<ObterDeputadosResult xmlns=...	74419	1218	Titular	143	1637164	LUIZ GONZAGA PATRIOTA	GONZAGA PATRIOTA
8	<ObterDeputadosResult xmlns=...	141401	2547	Titular	400	1637165	CARLOS GOMES BEZERRA	CARLOS BEZERRA
9	<ObterDeputadosResult xmlns=...	74317	1246	Titular	114	1637166	JÚLIO CÉSAR DE CARVALHO LIMA	JÚLIO CESAR
10	<ObterDeputadosResult xmlns=...	74305	3340	Titular	107	1637167	ROBERTO SOARES PESSOA	ROBERTO PESSOA

Exercício 7 - Extraindo dados de Web Service

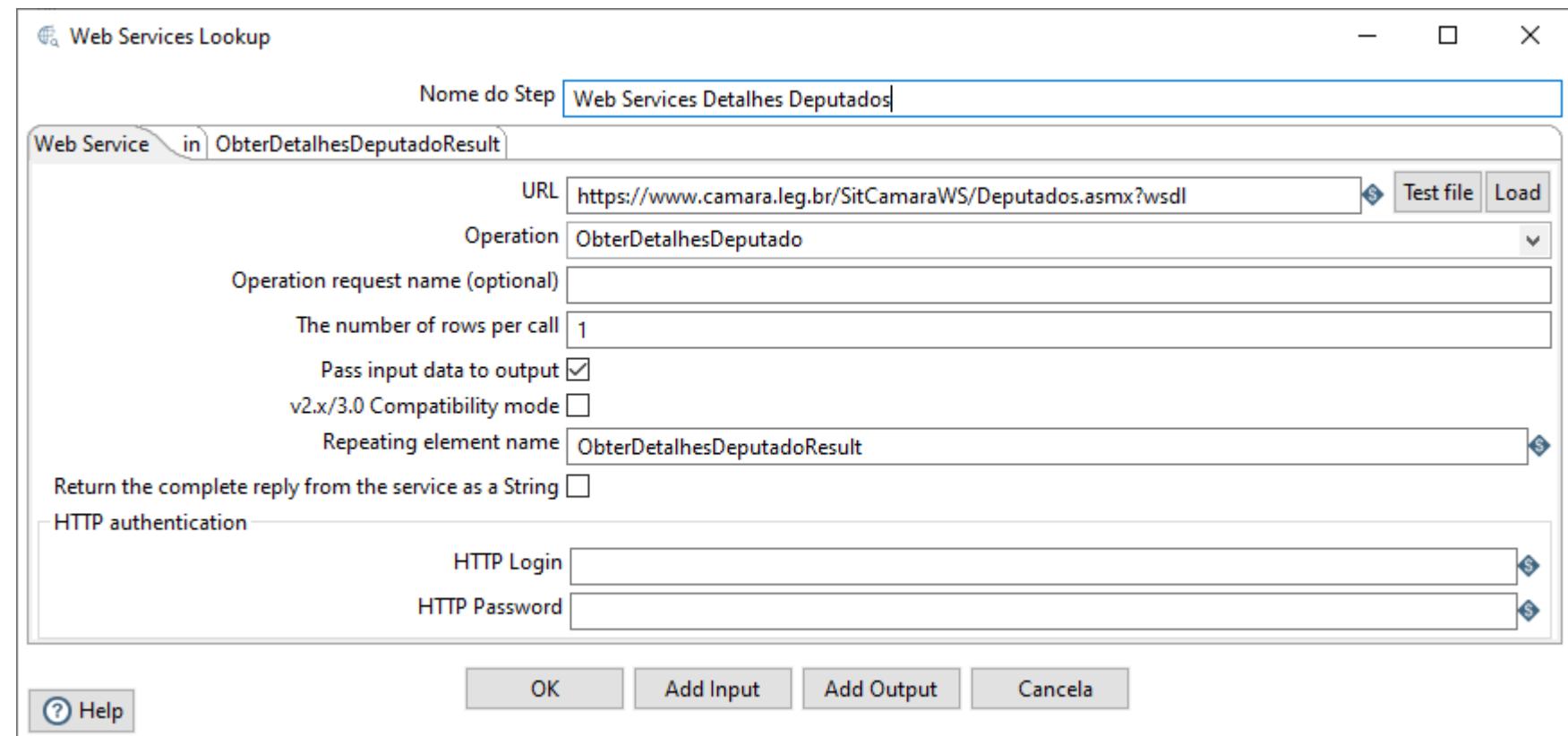
5 – Agora vamos adicionar os seguintes steps com objetivo integrar mais detalhes sobre os deputados.

- ❖ *Delay now:* configure um tempo de espera para cada linha da próxima chamada de Web Service



Exercício 7 - Extraindo dados de Web Service

- ❖ No segundo step de *Web service* configure conforme figura abaixo. O que modificará do primeiro a *Operation*: *ObterDetalhesDeputado*



Exercício 7 - Extraindo dados de Web Service

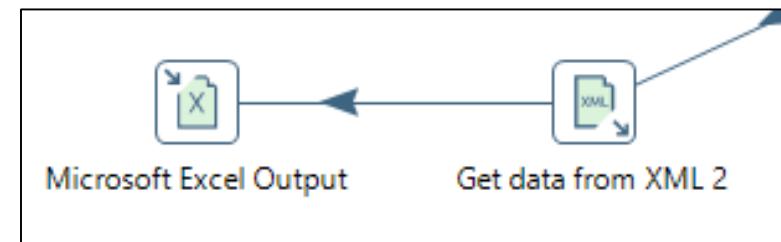
- ❖ Na aba *In* é esperado o dado de entrada do Web Service. Utilizaremos o campo “ideCadastro”.
 - Clique em *Obtem campos*
 - Remova a linha *numLegislatura*
 - Escolha valor de entrada com sendo *ideCadastro*

The screenshot shows the 'Web Services Lookup' interface. At the top, there are tabs for 'Nome do Step' (Nome do Step) and 'Web Services Detalhes Deputados' (Web Services Detalhes Deputados), with the latter being the active tab. Below the tabs, there is a sub-tab 'Web Service' (Web Service). The main area displays a table titled 'ObterDetalhesDeputadoResult'. The table has columns: '#', 'Name', 'WS Name', and 'WS Type'. There is one row with the value '1' in the '#' column, 'ideCadastro' in the 'Name' column, 'ideCadastro' in the 'WS Name' column, and 'string' in the 'WS Type' column.

#	Name	WS Name	WS Type
1	ideCadastro	ideCadastro	string

Exercício 7 - Extraiendo dados de Web Service

6 – Os próximos steps são de *Get Data from XML* e *Microsoft Excel Output*:



- ❖ No step *Get Data from XML*: configure semelhante ao passo 3 deste exercício com as opções:
 - Escolha em *Get XML source from a field*: ObterDetalhesDeputadoResult
 - Na aba *Content* : Loop Xpath: /ObterDetalhesDeputadoResult/Deputados/Deputado
 - Na aba *Fields* clique em *Get fields* (*utilizar fragmento de XML do próximo slides*)
- ❖ No step *Excel Output* configure onde salvar e aba *Fields* com a colunas de saída.
 - Não esqueça de excluir: ObterDeputadosResult, ObterDetalhesDeputadoResult e o campos duplicados que aparecem _1 na frente do nome.

7 - Execute a transformação e veja o resultado no arquivo excel.

Caso algum **ERROR** ocorra no STEP do segundo Web Service: aumente o tempo de em *Delay Rows*

```
<ObterDetalhesDeputadoResult>
<Deputados>
<Deputado>
<numLegislatura>56</numLegislatura>
<email>dep.beneditasilva@camara.leg.br</email>
<nomeProfissao>Assistente social, Auxiliar de enfermagem, Professor, Servidor público</nomeProfissao>
<dataNascimento>26/04/1942</dataNascimento>
<dataFalecimento/>
<ufRepresentacaoAtual>RJ</ufRepresentacaoAtual>
<situacaoNaLegislaturaAtual>Em Exercício</situacaoNaLegislaturaAtual>
<idCadastro>73701</idCadastro>
<idParlamentarDeprecated>76</idParlamentarDeprecated>
<nomeParlamentarAtual>BENEDITA DA SILVA</nomeParlamentarAtual>
<nomeCivil>BENEDITA SOUZA DA SILVA SAMPAIO</nomeCivil>
<sexo>F</sexo>
<partidoAtual>
<idPartido>PT</idPartido>
<sigla>PT</sigla>
<nome>Partido dos Trabalhadores</nome>
</partidoAtual>
<gabinete>
<numero>330</numero>
<anexo></anexo>
<telefone>3215-5330</telefone>
</gabinete>
<comissoes>
<comissao>
<idOrgaoLegislativoCD>537870</idOrgaoLegislativoCD>
<siglaComissao>CMULHER</siglaComissao>
<nomeComissao>Comissão de Defesa dos Direitos da Mulher</nomeComissao>
<condicaoMembro>Suplente</condicaoMembro>
<dataEntrada>13/03/2019</dataEntrada>
<dataSaida/>
</comissao>
</comissoes>
<cargosComissoes>
<cargoComissao>
<idOrgaoLegislativoCD>536996</idOrgaoLegislativoCD>
<siglaComissao>CCULT</siglaComissao>
<nomeComissao>Comissão de Cultura</nomeComissao>
<idCargo>1</idCargo>
<nomeCargo>Presidente</nomeCargo>
<dataEntrada>14/03/2019</dataEntrada>
<dataSaida/>
</cargoComissao>
</cargosComissoes>
<periodosExercicio>
<periodoExercicio>
<siglaUFRepresentacao>RJ</siglaUFRepresentacao>
<situacaoExercicio>Titular</situacaoExercicio>
<dataInicio>01/02/2019</dataInicio>
<dataFim>
<idCausaFimExercicio/>
<descricaoCausaFimExercicio/>
<idCadastroParlamentarAnterior/>
</periodoExercicio>
</periodosExercicio>
<historicoNomeParlamentar/>
<filiacoesPartidarias/>
<historicoLider>
<itemHistoricoLider>
<idHistoricoLider>73701</idHistoricoLider>
<idCargoLideranca>V</idCargoLideranca>
<descricaoCargoLideranca>Vice-Líder</descricaoCargoLideranca>
<numOrdemCargo>2021</numOrdemCargo>
<dataDesignacao>19/11/2019</dataDesignacao>
<dataTermino>19/11/2019</dataTermino>
<codigoUnidadeLideranca>P</codigoUnidadeLideranca>
<siglaUnidadeLideranca>PT</siglaUnidadeLideranca>
<idBlocoPartido>PT</idBlocoPartido>
</itemHistoricoLider>
<historicoLider>
</historicoLider>
</Deputado>
</Deputados>
</ObterDetalhesDeputadoResult>
```

Exercício 8

Crie uma extração.

Exercício 8

Escolha alguns *steps inputs* e crie uma extração de dados. Quanto mais *steps* melhor para o treinamento. Use a criatividade !



