



# ISI - TRABALHO PRÁTICO 1

Integração de Sistemas de Informação

José Moura – a13742

Paulo Meneses – a17611

## Índice

<b>Introdução .....</b>	<b>2</b>
<b>Objetivos no âmbito da disciplina .....</b>	<b>2</b>
<b>Resumo .....</b>	<b>3</b>
<b>Desenvolvimento.....</b>	<b>4</b>
Transformações.....	4
ExtractPrevisões.....	4
htmlAsAmbient .....	11
InsereBD.....	12
Jobs.....	14
EmailPrevisão.....	14
<b>Conclusão.....</b>	<b>19</b>

# Introdução

Este trabalho foi realizado no âmbito da unidade curricular, Integração de Sistemas de Informação, no curso de Licenciatura em Engenharia em sistemas informáticos pós-laboral.

O trabalho foi desenvolvido uma solução de ETL (Extract, Transform, Load) com a ferramenta Pentaho Kettle, que permite a criação de diversos processos de ETL, neste caso o envio de um email com a previsão meteorológica de braga.

## Objetivos no âmbito da disciplina

- Consolidar conceitos associados à Integração de Sistemas de Informação usando Dados;
- Analisar e especificar cenários de aplicação de processos de ETL;
- Explorar ferramentas de suporte a processos de ETL;
- Explorar novas Tecnologias, Frameworks ou Paradigmas;
- Potenciar a experiência no desenvolvimento de software;
- Facilitar a assimilação do conteúdo da Unidade Curricular.

# Resumo

Este trabalho foi desenvolvido na ferramenta Pentaho Kettle, na qual o resultado do processo ETL é o envio de um email com a previsão meteorológica de Braga para os próximos 5 dias.

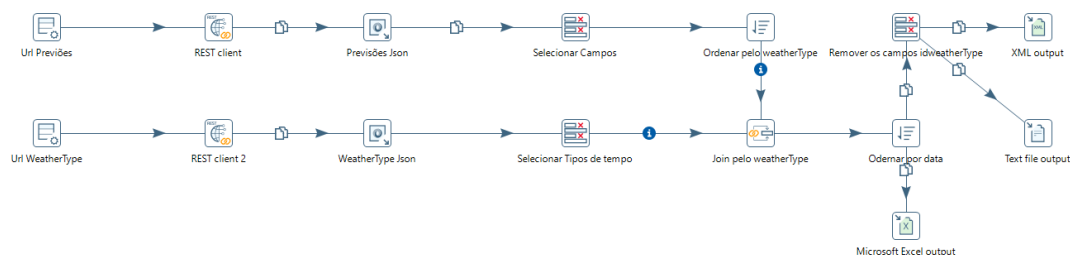
Os dados utilizados são retirados da API do IPMA (Instituto Português do Mar e da Atmosfera), através de um RestClient na ferramenta, sendo selecionados somente os valores que nos interessam, tais como, temperatura máxima, temperatura mínima, data da previsão e descrição da previsão, visto que a descrição da previsão está separada dos outros dados na API, será necessário mais um dado que é o id do tipo de tempo, para posteriormente ser executada a reunião das tabelas. A tabela resultante será depois exportada em três formatos de ficheiro como: txt, xml e xls. Esses ficheiros poderão depois ser tratados por outras transformações, neste caso o xml é transformado numa tabela html com as previsões através de uma XSLT, para seguidamente ser enviada por email, já no caso do ficheiro de texto é lido e inserido numa base de dados PostgreSQL

# Desenvolvimento

## 1. Transformações

### 1.1. ExtractPrevisões

Esta transformação faz a extração dos dados da API e gera 3 outputs com as previsões.



É utilizado um Generate Rows (Url Previsões) para gerar um linha com o link (<http://api.ipma.pt/open-data/forecast/meteorology/cities/daily/1030300.json>) para as previsões de braga.

Depois o Rest Client utiliza esse link para receber o json através do metodo GET e o output é inserido num Json Input (Previsões Json).



Generate rows

Nome do Step: Url Previsões

Limit: 1

Never stop generating rows: ☐

Interval in ms (delay): 5000

Current row time field name: now

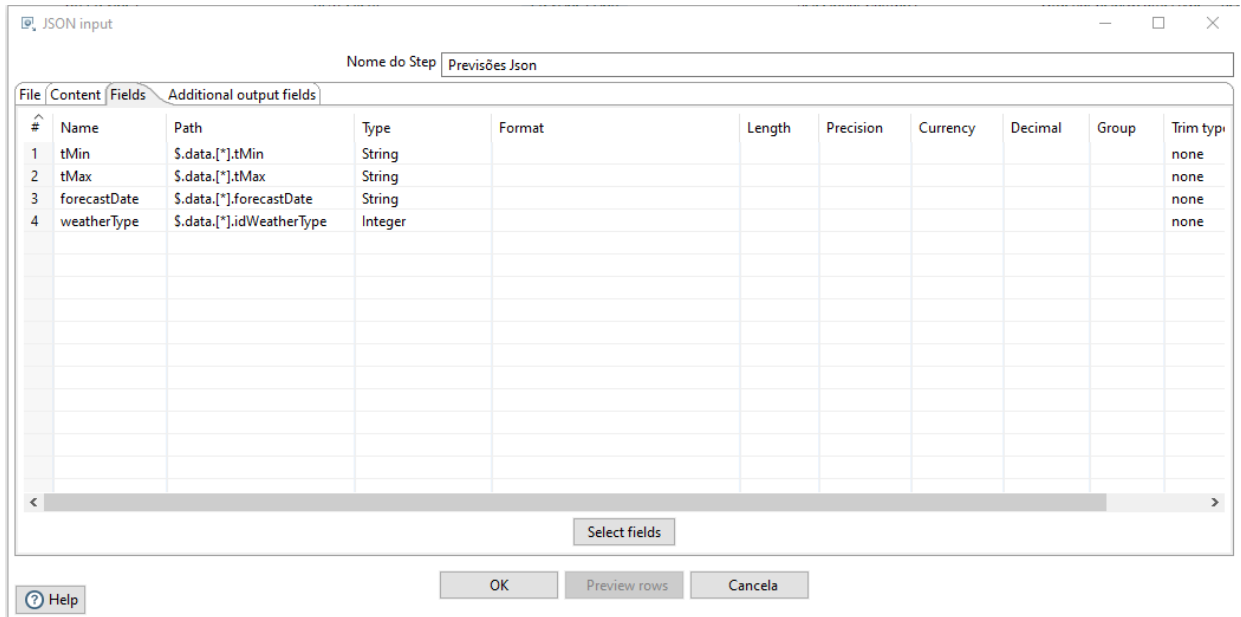
Previous row time field name: FiveSecondsAgo

Fields:

#	Nome	Tipo	Formato	Tamanho	Precisão	Moeda	Decimal	Grupo	Valor
1	URL	String							http://api.ipma.pt/open-data/forecast/
2									

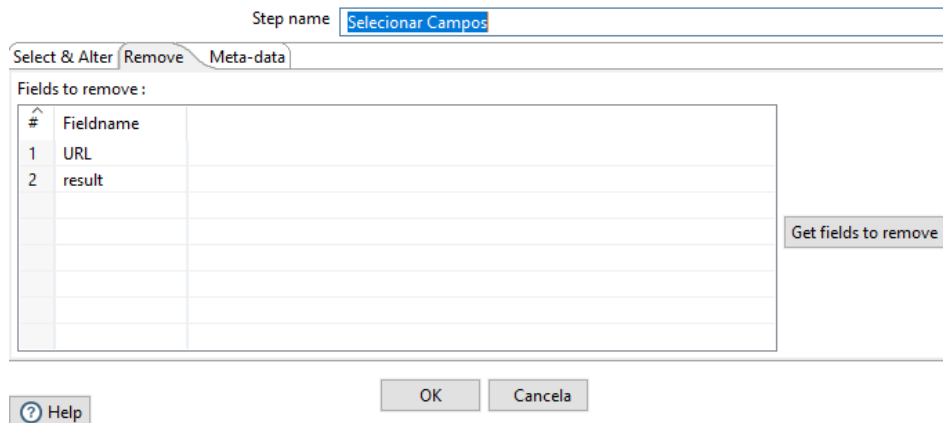
Help OK Preview Cancela

Os campos do Previsões Json que foram configurados foram: tMin, tMax, forecastDate e weatherType.



#	Name	Path	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	tMin	\$.data.[*].tMin	String							none
2	tMax	\$.data.[*].tMax	String							none
3	forecastDate	\$.data.[*].forecastDate	String							none
4	weatherType	\$.data.[*].idWeatherType	Integer							none

O output desse json é inserido num Select Values (Selecionar Campos) para ser selecionado somente os valores de tMin, tMax, forecastDate e weatherType.

#	Fieldname
1	URL
2	result

## ISI - Trabalho Prático 1

Select values

Step name: Selecionar Campos

Select & Alter Remove Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	URL			
2	result			
3	tMin			
4	tMax			
5	forecastDate			
6	weatherType			

Get fields to select  
Edit Mapping

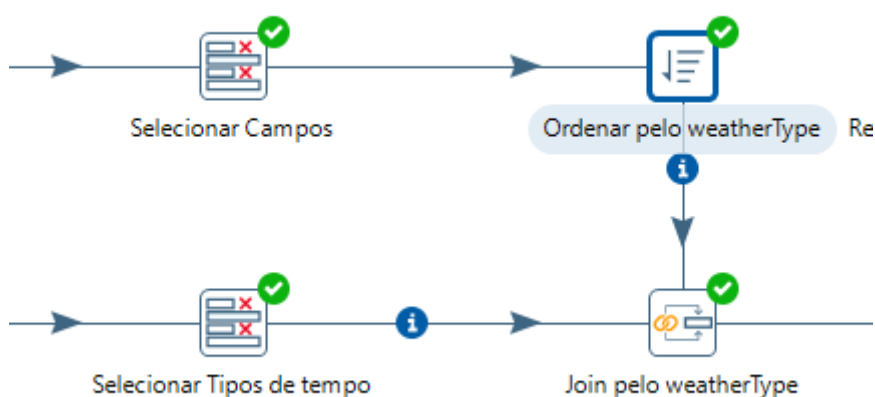
Include unspecified fields, ordered by name ☐

Help OK Cancela

O resultado deste Select Values é o seguinte:

#	tMin	tMax	forecastDate	weatherType
1	11.3	17.3	2020-11-07	10
2	13.1	17.4	2020-11-08	9
3	11.0	16.7	2020-11-09	9
4	8.1	18.2	2020-11-10	3
5	8.1	17.7	2020-11-11	5

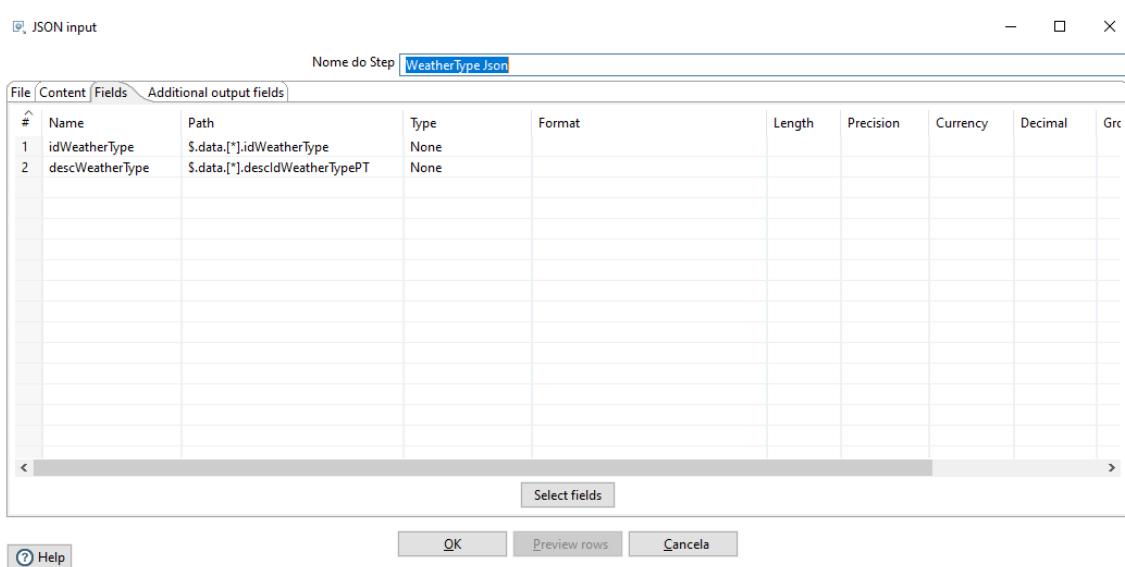
Os valores são ordenados pela ordem crescente pelo campo weatherType no Sort Rows (Ordenar pelo weatherType) para depois serem juntos a uma tabela com a descrição para cada tipo de tempo.



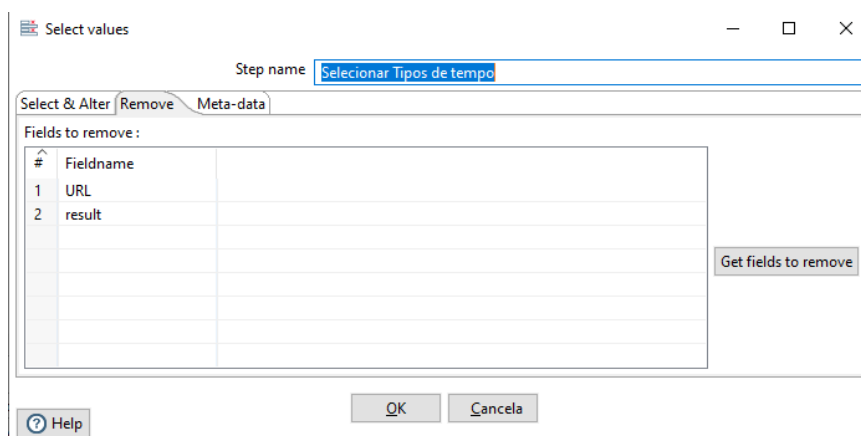
Depois da ordenação a tabela resultado é a seguinte:

#	tMin	tMax	forecastDate	weatherType
1	8.1	18.2	2020-11-10	3
2	8.1	17.7	2020-11-11	5
3	13.1	17.4	2020-11-08	9
4	11.0	16.7	2020-11-09	9
5	11.3	17.3	2020-11-07	10

O processo de obter a descrição dos tipos de tempo é semelhante ao anterior, é criado uma linha com um Generate Rows (Url weatherType), com o link (<http://api.ipma.pt/open-data/weather-type-classe.json>), depois o Rest Client 2 efetua o download do json, os campos inseridos no Json Input (WeatherType Json) são: idWeatherType e descWeatherType.



Depois é utilizado um Select Values (Selecionar Tipos Tempo) para excluir os campos URL e result.

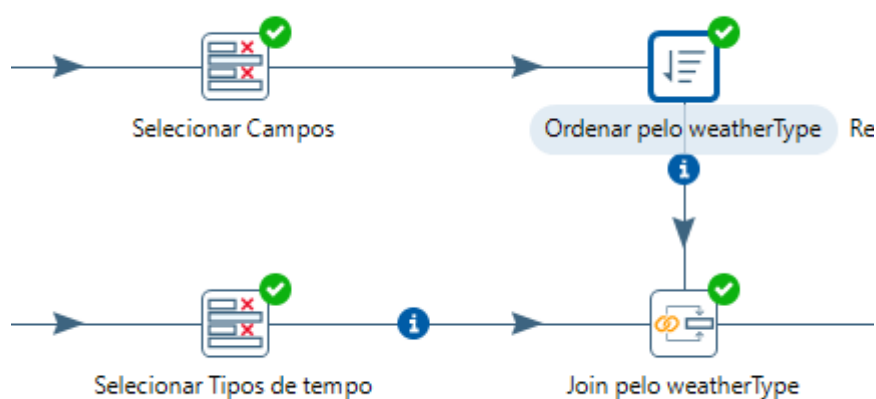




A tabela resultado dos tipo de tempo é:

#	idWeatherType	descWeatherType
1	-99	---
2	0	Sem informação
3	1	Céu limpo
4	2	Céu pouco nublado
5	3	Céu parcialmente nublado
6	4	Céu muito nublado ou encoberto
7	5	Céu nublado por nuvens altas
8	6	Aguaceiros
9	7	Aguaceiros fracos
10	8	Aguaceiros fortes
11	9	Chuva
12	10	Chuva fraca ou chuvisco
13	11	Chuva forte
14	12	Períodos de chuva
15	13	Períodos de chuva fraca
16	14	Períodos de chuva forte
17	15	Chuvisco
18	16	Nebulosa
19	17	Nevoeiro ou nuvens baixas
20	18	Neve
21	19	Trovoada
22	20	Aguaceiros e trovoada
23	21	Granizo
24	22	Gelada
25	23	Chuva e trovoada
26	24	Nebulosidade convectiva
27	25	Céu com períodos de muito nublado
28	26	Nevoeiro
29	27	Céu nublado

No fim o resultado é inserido no Merge Rows (Join pelo weatherType) visto que ele já está ordenado.



O Merge Rows permite fazer o cruzamento dos dados de uma tabela com a outra por determinado valor chave, neste caso o weatherType.

Merge join

Step name:

First Step:

Second Step:

Join Type:

Keys for 1st step:

#	Key field
1	weatherTy...

Get key fields

Keys for 2nd step:

#	Key field
1	idWeathe...

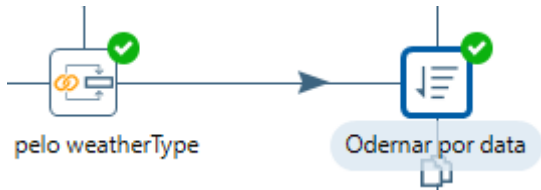
Get key fields

Help OK Cancela

O resultado do Merge Rows é:

#	tMin	tMax	forecastDate	weatherType	idWeatherType	descWeatherType
1	8.1	18.2	2020-11-10	3	3	Céu parcialmente nublado
2	8.1	17.7	2020-11-11	5	5	Céu nublado por nuvens altas
3	13.1	17.4	2020-11-08	9	9	Chuva
4	11.0	16.7	2020-11-09	9	9	Chuva
5	11.3	17.3	2020-11-07	10	10	Chuva fraca ou chuvisco

Esta tabela é de seguida ordenada por ordem crescente pelo campo forecastDate, no Sort Rows (Ordenar por data).



Nome do Step: Ordenar por data

Sort directory: %%java.io.tmpdir%%

TMP-file prefix: out

Sort size (rows in memory): 1000000

Free memory threshold (in %):

Compress TMP Files? ☐

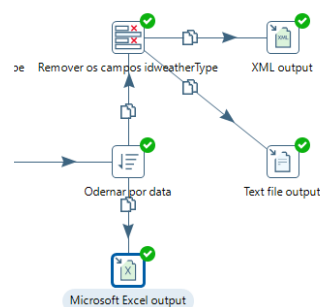
Only pass unique rows? (verifies keys only) ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	forecastDate	S	N	N	0	N

Buttons: Help, OK, Cancela, Obtem campos

Depois é efetuada a remoção dos campos weatherType e idWeatherType através do Select Values (Remover os campos idweatherType) e o output é enviado para três ficheiros nos formatos: xml, txt e xls.



## Remoção dos campos weatherType e IdWeatherType

Select values

Step name: Remover os campos idweatherType

Select & Alter Remove Meta-data

Fields to remove:

#	Fieldname
1	weatherType
2	idWeatherType

Get fields to remove

Help OK Cancela

A tabela que resulta é :

#	tMin	tMax	forecastDate	descWeatherType
1	11.3	17.3	2020-11-07	Chuva fraca ou chuvisco
2	13.1	17.4	2020-11-08	Chuva
3	11.0	16.7	2020-11-09	Chuva
4	8.1	18.2	2020-11-10	Céu parcialmente nublado
5	8.1	17.7	2020-11-11	Céu nublado por nuvens altas

No fim é gerado um xml com o nome de Previsões que contem a tabela.

XML output

Nome do Step: XML output

File Content Fields

Filename: \${Internal.Entry.Current.Directory}/Previsões Browse...

Do not create file at start ☐

Pass output to servlet ☐

Extension: xml

Include stepnr in filename? ☐

Include date in filename? ☐

Include time in filename? ☐

Specify Date time format ☐

Date time format:  

Show filename(s)...

Add filenames to result ☐

Help OK Cancela

O conteúdo do xml é o seguinte:

```
<?xml version="1.0" encoding="UTF-8"?>
<Rows>
  <Row>
    <tMin>7.1</tMin>
    <tMax>19.0</tMax>
    <forecastDate>2020-11-10</forecastDate>
    <descWeatherType>Céu nublado por nuvens altas</descWeatherType>
  </Row>
  <Row>
    <tMin>7.4</tMin>
    <tMax>19.0</tMax>
    <forecastDate>2020-11-11</forecastDate>
    <descWeatherType>Céu parcialmente nublado</descWeatherType>
  </Row>
  <Row>
    <tMin>8.6</tMin>
    <tMax>21.2</tMax>
    <forecastDate>2020-11-12</forecastDate>
    <descWeatherType>Céu nublado por nuvens altas</descWeatherType>
  </Row>
  <Row>
    <tMin>8.9</tMin>
    <tMax>19.0</tMax>
    <forecastDate>2020-11-13</forecastDate>
    <descWeatherType>Aguaceiros fracos</descWeatherType>
  </Row>
  <Row>
    <tMin>11.0</tMin>
    <tMax>18.2</tMax>
    <forecastDate>2020-11-14</forecastDate>
    <descWeatherType>Chuva</descWeatherType>
  </Row>
</Rows>
```

## 1.2 htmlAsAmbient

A transformação htmlAsAmbient consiste em pegar no conteúdo de um ficheiro html e defini-lo como variável de ambiente chama html\_body.



O ficheiro é lido como um ficheiro de texto, neste caso o ficheiro lido é Previsão.html, na configuração do Content é preciso mudar o fileType para Fixed e nos Fields é preciso a configuração do campo com Name html, Type string, Position 0, Length 100, Trim Type None e Repeat N.

Text file input

Nome do Step: **Text file input**

Filetype: Fixed

Separator: ;

Enclosure: "

Exclude Regular Expression:

Selected files:

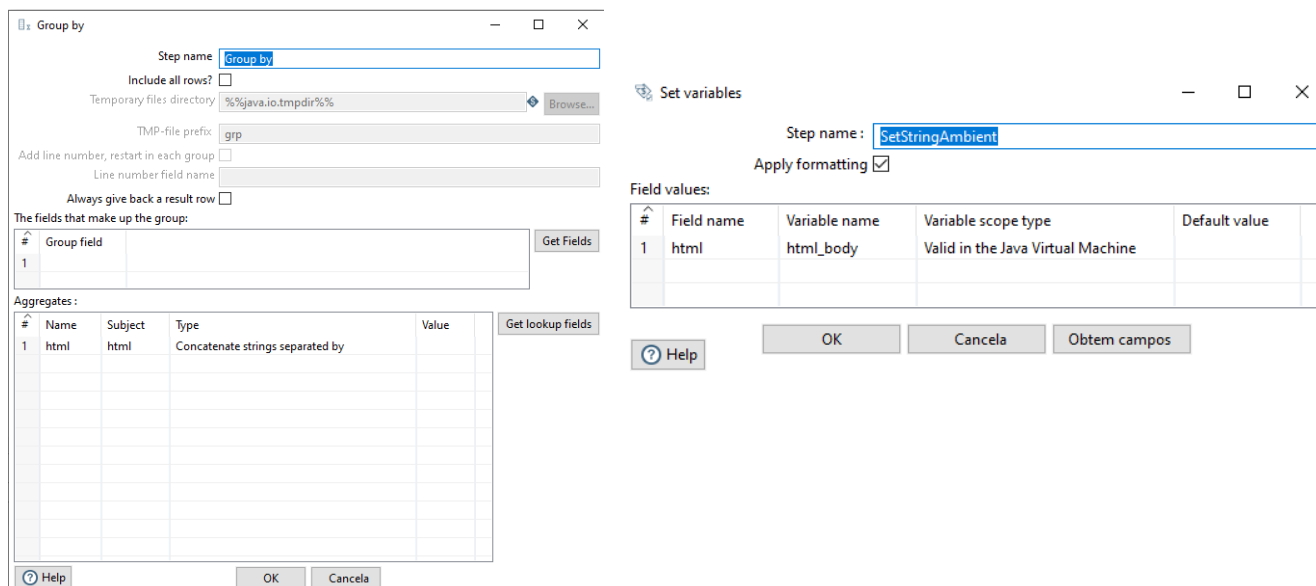
#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include subfolders
1	\$(Internal.Entry.Current.Directory)/Previsão.html			N	N

Text file input

Nome do Step: **Text file input**

#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type	Repeat
1	html	String		0	100							none	N

É utilizado um Group by para a concatenação das strings lidas numa só e depois essa string final é passada para um Set Variables (SetStringAmbient) para pôr a string como variável de ambiente para ser passada para o email.



**Group by**

Step name: Group by

Include all rows? ☐

Temporary files directory: %%java.io.tmpdir%% Browse...

TMP-file prefix: grp

Add line number, restart in each group ☐

Line number field name:

Always give back a result row ☐

The fields that make up the group:

#	Group field
1	

Get Fields

Aggregates:

#	Name	Subject	Type	Value
1	html	html	Concatenate strings separated by	

Get lookup fields

Help OK Cancela

**Set variables**

Step name: SetStringAmbient

Apply formatting ☒

Field values:

#	Field name	Variable name	Variable scope type	Default value
1	html	html_body	Valid in the Java Virtual Machine	

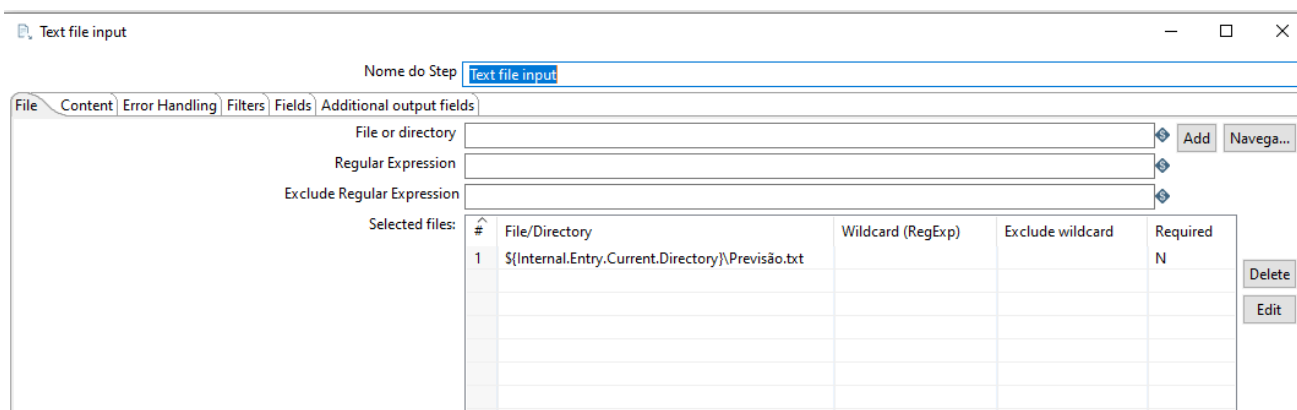
Help OK Cancela Obtem campos

### 1.3 InserirBD

Esta transformação, insere as previsões numa base de dados PostgreSQL, através da leitura do ficheiro de texto gerado anteriormente.



Configuração do Text file input:



Nome do Step: Text file input

File Content Error Handling Filters Fields Additional output fields

File or directory:  Add Navega...

Regular Expression:

Exclude Regular Expression:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required
1	\$(Internal.Entry.Current.Directory)\Previsão.txt			N

Delete Edit

Text file input

Nome do Step: **Text file input**

#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type	Repeat
1	tMin	Number	#, #		15	0	€	.	,	-		none	N
2	tMax	Number	#, #		15	0	€	.	,	-		none	N
3	forecastDate	Date	yyyy-MM-dd				€	.	,	-		none	N
4	descWeatherType	String			28		€	.	,	-		none	N

### Configuração do Insert / Update:

Insert / update

Step name: **Insert / update**

Connection: testePrevisao [Edit... New... Wizard...]

Target schema: public [Navega...]

Target table: previsoes [Browse...]

Commit size: 100

Don't perform any updates: ☐

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

#	Table field	Comparator	Stream field1	Stream field2
1	tmin	=	tMin	
2	tmax	=	tMax	
3	desWeather	=	descWeatherType	
4	forecastdate		forecastDate	

Update fields:

#	Table field	Stream field	Update
1	tmin	tMin	Y
2	tmax	tMax	Y
3	forecastdate	forecastDate	Y
4	desWeather	descWeatherType	Y

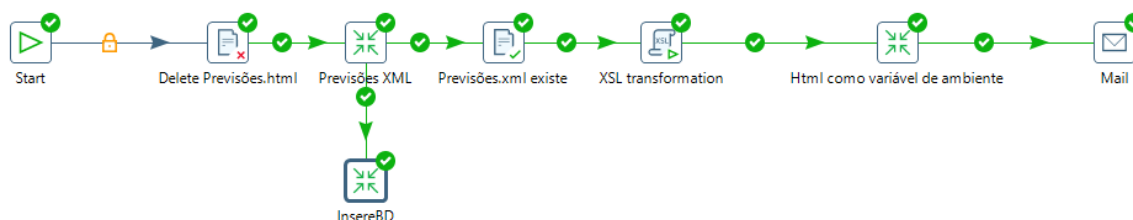
Buttons: [Help] [OK] [Cancela] [SQL]

### Tabela resultante na base de dados:

Data Output	Explain	Messages	Notifications
<b>tmin</b> bigint	<b>tmax</b> bigint	<b>desweather</b> character varying (28)	<b>forecastdate</b> timestamp without time zone
1	7	19	Céu nublado por nuvens altas
2	7	19	Céu parcialmente nublado
3	9	21	Céu nublado por nuvens altas
4	9	19	Aguaceiros fracos
5	11	18	Chuva

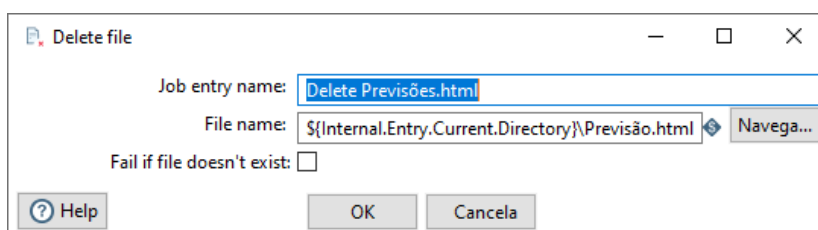
## 2. Jobs

### 2.1 EmailPrevisão

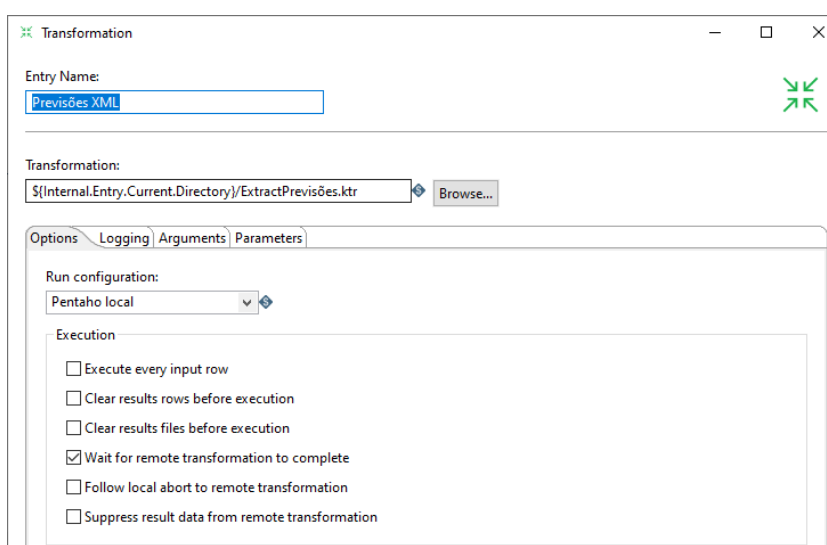


O job é iniciado pelo Start, que irá apagar o ficheiro Previsões.html e depois executar a transformação Previsões XML que é a transformação Email com previsão, de seguida é iniciado a transformação InsereBD e a verificação se o xml existe para ser aplicado uma XSL transformation no xml gerado pela transformação que irá gerar uma página html com as previsões, de seguida é executada a transformação Html como variável de ambiente que põe essa pagina como variável de ambiente, para, por fim ser usada no corpo do email para mostrar os dados no email.

Configuração do Delete Previsões.html:

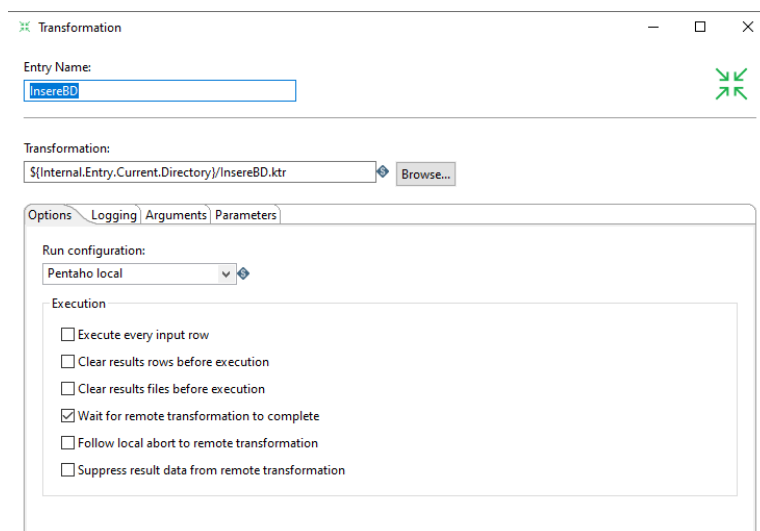


Configuração da transformação ExtractPrevisões:

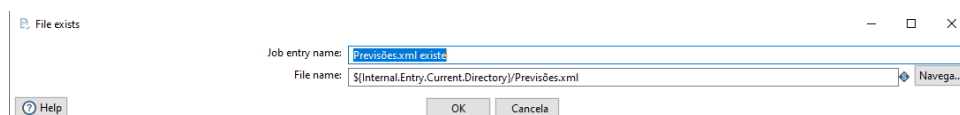


No fim da execução da transformação ExtractPrecisões é verificado se o xml existe e executada a transformação InsereBD.

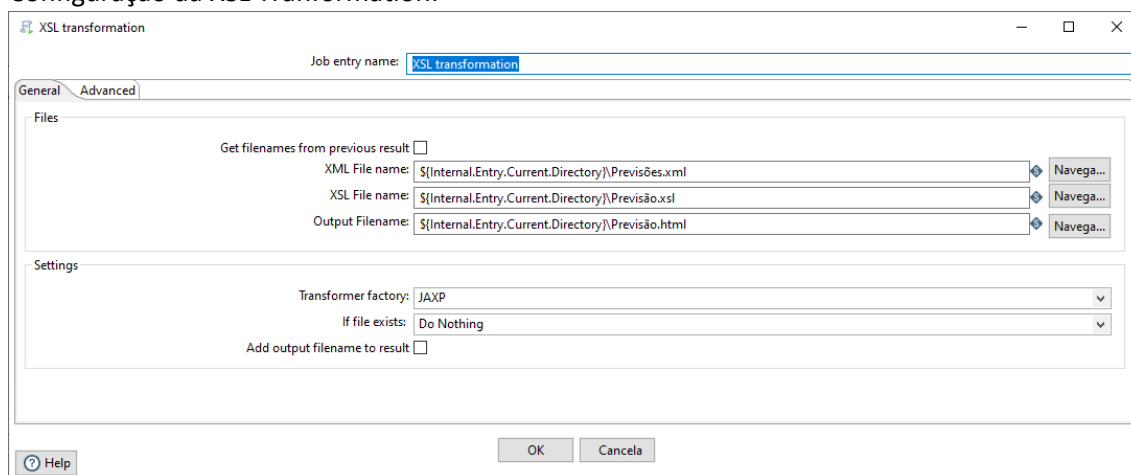
Configuração de InsereBD:



Configuração de Previsões.xml existe:



Configuração da XSL Transformation:



A transformação XSL é feita através do ficheiro Previsão.xsl, cria uma tabela com o dia da previsão, temperatura máxima, temperatura mínima e a descrição do tempo. O ficheiro Previsão.xsl contém o seguinte código:

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="/">
```



```

<html lang="pt-PT">
<meta charset="utf-8"/>
<head>
<style>
table {
    font-family: arial, sans-serif;
    border-collapse: collapse;
    width: 100%;
}

td, th {
    border: 1px solid #dddddd
    text-align: left
    padding: 8px
}

tr:nth-child(even) {
    background-color: #dddddd
}
</style>
</head>
<body>
    <h2>Previsao do tempo para Braga</h2>
    <table>
        <tr>
            <th style="text-align:center" >Dia</th>
            <th style="text-align:center" >Previsao</th>
            <th style="text-align:center" >Temperatura Minima</th>
            <th style="text-align:center" >Temperatura Maxima</th>
        </tr>
        <xsl:for-each select="Rows/Row">
            <tr>
                <td style="text-align:center" ><xsl:value-of select="forecastDate"/></td>
                <td style="text-align:center" ><xsl:value-of select="descWeatherType"/></td>
                <td style="text-align:center" ><xsl:value-of select="tMin"/></td>
                <td style="text-align:center" ><xsl:value-of select="tMax"/></td>
            </tr>
        </xsl:for-each>
    </table>
</body>
</html>
</xsl:template> </xsl:stylesheet>

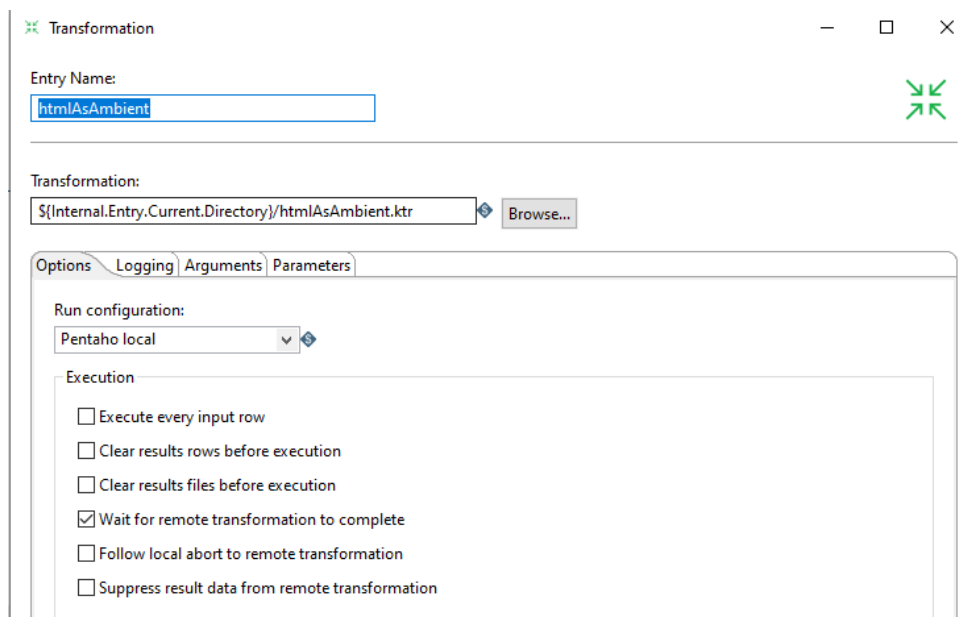
```

Depois da transformação obtemos uma tabela da seguinte forma:

#### Previsao do tempo para Braga

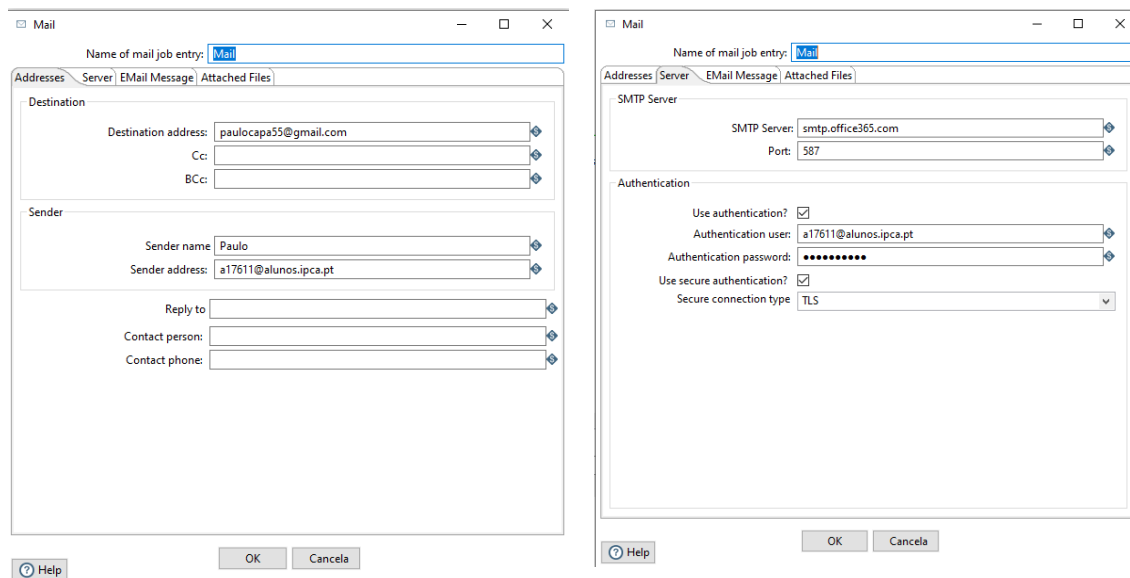
Dia	Previsao	Temperatura Minima	Temperatura Maxima
2020-11-10	Céu nublado por nuvens altas	7.1	19.0
2020-11-11	Céu parcialmente nublado	7.4	19.0
2020-11-12	Céu nublado por nuvens altas	8.6	21.2
2020-11-13	Aguaceiros tracos	8.9	19.0
2020-11-14	Chuva	11.0	18.2

De seguida a tabela html é definida como variável de ambiente através da transformação htmlAsAmbient com a seguinte configuração:



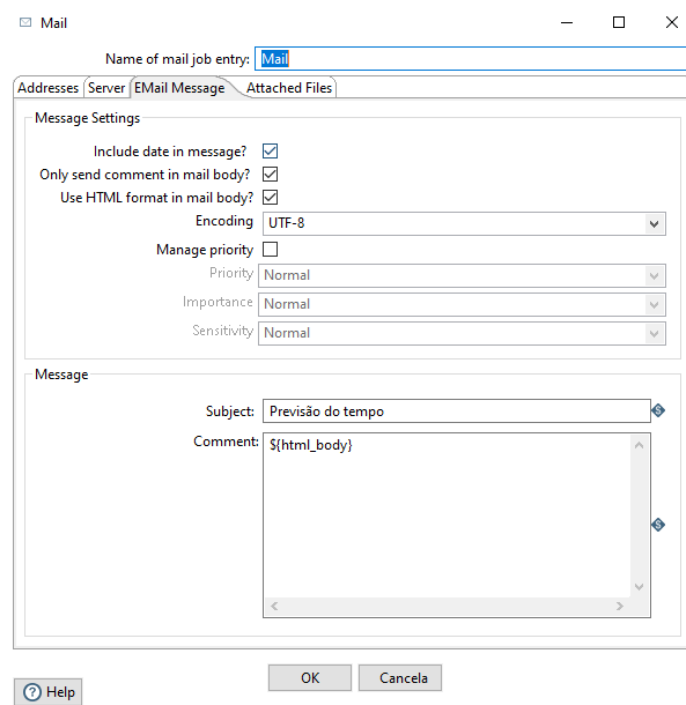
The screenshot shows the 'Transformation' configuration window. The 'Entry Name' is 'htmlAsAmbient'. The 'Transformation' field is set to '\$[Internal.Entry.Current.Directory]/htmlAsAmbient.ktr'. The 'Options' tab is selected, showing 'Run configuration' set to 'Pentaho local'. Under 'Execution', the following options are checked: 'Wait for remote transformation to complete'. Other options like 'Execute every input row', 'Clear results rows before execution', 'Clear results files before execution', 'Follow local abort to remote transformation', and 'Suppress result data from remote transformation' are unchecked.

Por fim o email é enviado, com a seguinte configuração:



The first screenshot shows the 'Mail' configuration window with the 'Name of mail job entry' set to 'Mail'. The 'Destination' tab is selected, showing 'Destination address' as 'paulocapa55@gmail.com', 'Sender name' as 'Paulo', and 'Sender address' as 'a17611@alunos.ipca.pt'. The second screenshot shows the 'SMTP Server' and 'Authentication' tabs. The 'SMTP Server' is 'smtp.office365.com' and the 'Port' is '587'. Under 'Authentication', 'Use authentication?' is checked, 'Authentication user' is 'a17611@alunos.ipca.pt', 'Authentication password' is masked with dots, 'Use secure authentication?' is checked, and 'Secure connection type' is 'TLS'.

Na mensagem do email é enviado o html com a tabela:



No final o email recebido é o seguinte:



**Paulo** <a17611@alunos.ipca.pt>

para mim ▾



## Previsao do tempo para Braga

Dia	Previsao	Temperatura Minima	Temperatura Maxima
2020-11-10	CÃ©u nublado por nuvens altas	7.1	19.0
2020-11-11	CÃ©u parcialmente nublado	7.4	19.0
2020-11-12	CÃ©u nublado por nuvens altas	8.6	21.2
2020-11-13	Aguaceiros fracos	8.9	19.0
2020-11-14	Chuva	11.0	18.2

Message date: 2020/11/10 10:56:31.772

# Conclusão

Com o trabalho desenvolvido foram exploradas diversas funcionalidades na ferramenta, bem como ajudou na consolidação e assimilação do conteúdo da disciplina, também aumentou a nossa experiência com ETL e permitiu-nos perceber a as diferentes utilidades das ferramentas ETL podem trazer, e a sua grande capacidade de criar uma grande variedade de processos de uma forma clara e com uma a utilização simples.