ISI - Trabalho Prático 1

Integração de Sistemas de Informação



José Moura – a13742 Paulo Meneses – a17611

Índice

**Introdução2**

**Objetivos no âmbito da disciplina**2

**Resumo**3

**Desenvolvimento4**

Transformações4

ExtractPrevisões4

htmlAsAmbient11

InsereBD12

Jobs14

EmailPrevisão14

**Conclusão19**

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.

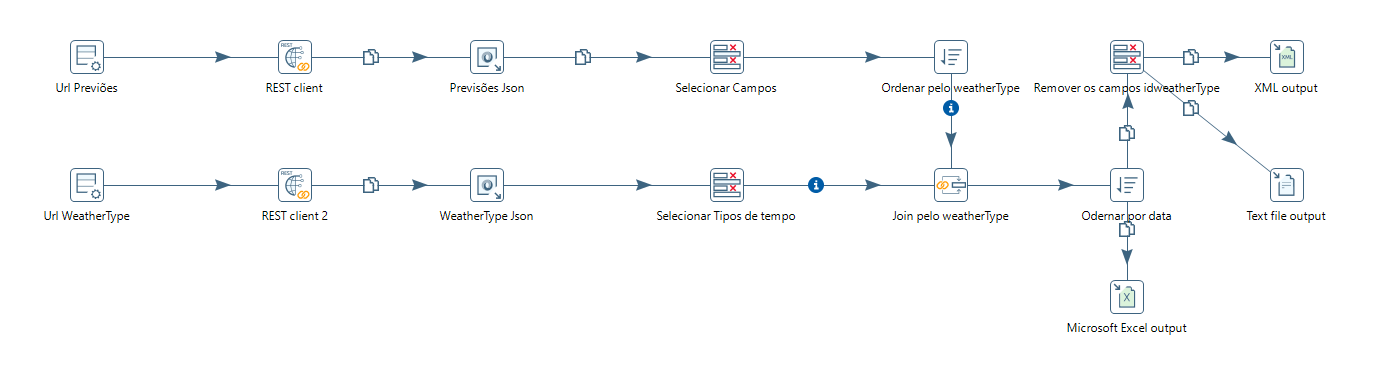
Resumo

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 do 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

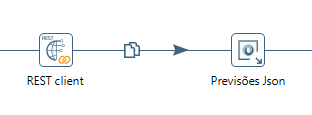
Desenvolvimento

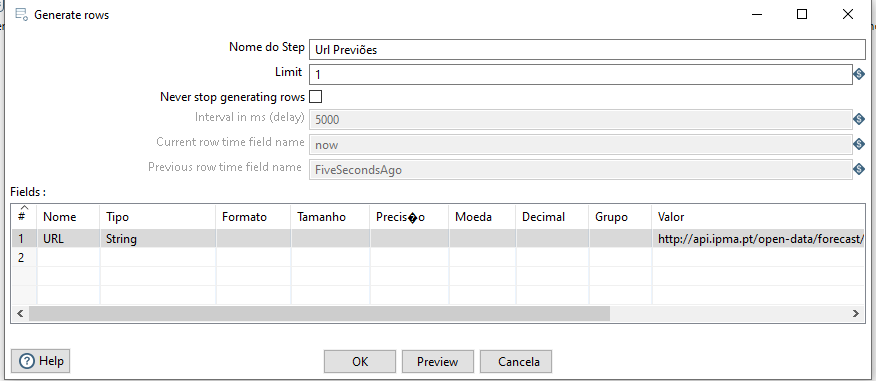
1. Transformações
   1. ExtractPrevisões

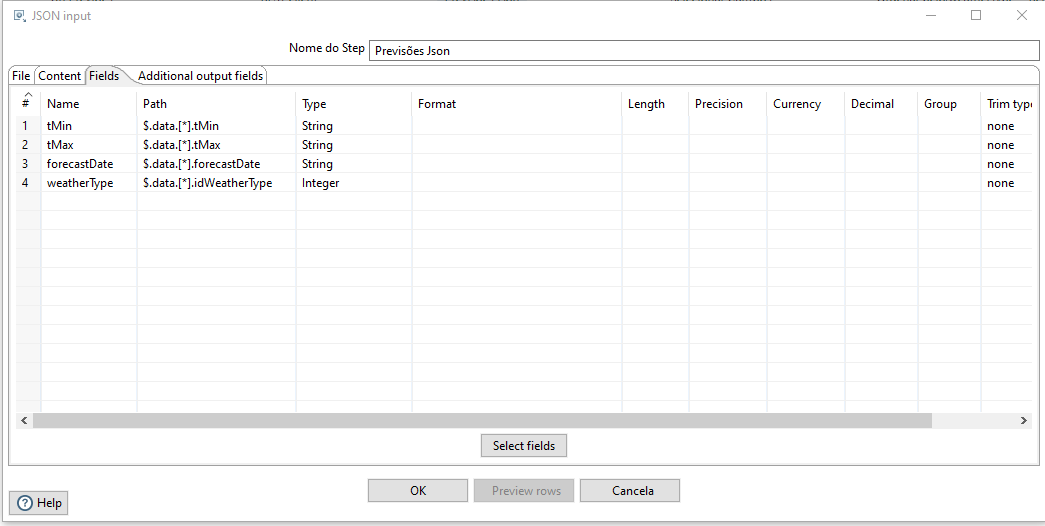
Esta transformação faz a extração dos dados da API e gera 3 outputs com as previsões.

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.

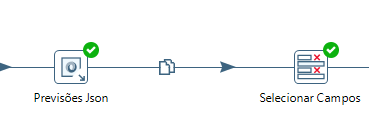
Depois o Rest Client utiliza esse link para receber o json através do metodo GET e o output é inserido num Json Input (Previsões Json).

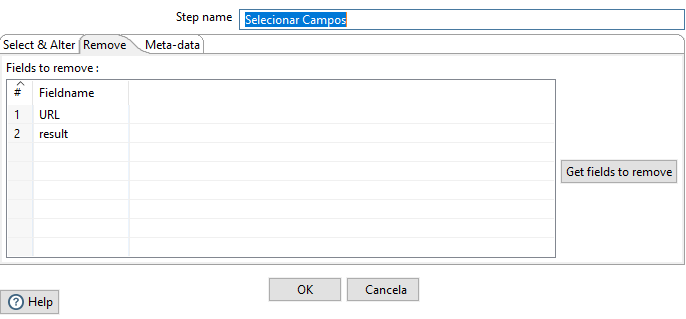


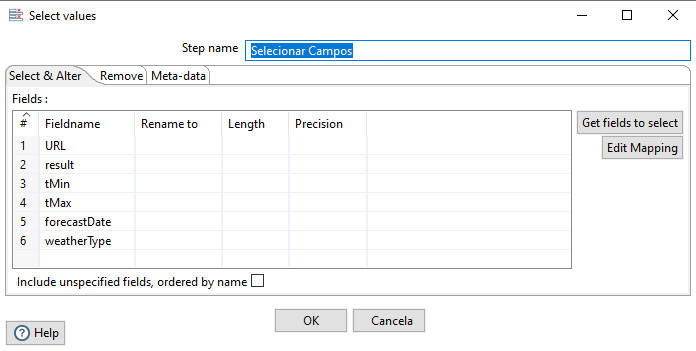


Os campos do Previsões Json que foram configurados foram: tMin, tMax, forecastDate e weatherType.

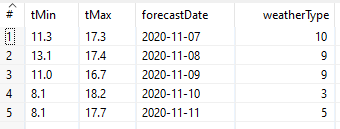
O output desse json é inserido num Select Values (Selecionar Campos) para ser selecionado somente os valores de tMin, tMax, forecastDate e weatherType.



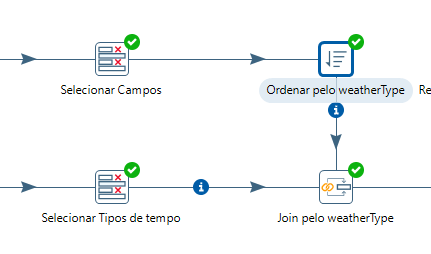




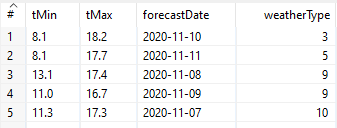
O resultado deste Select Values é o seguinte:

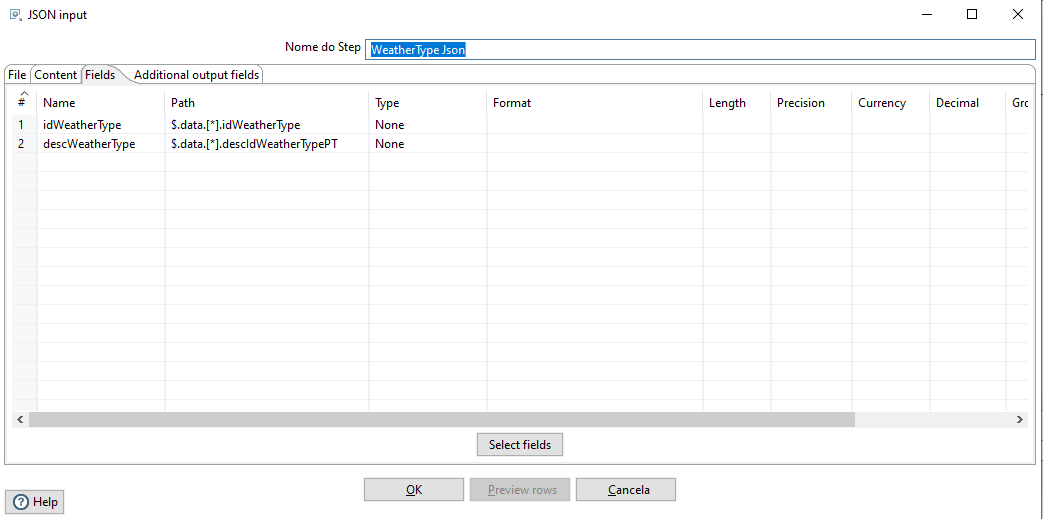


Os valores são ordenados pela ordem crescente pelo campo weatherType no Sort Rows (Ordenar pelo weatherType) para depois serem juntos a uma tabela com a descrição para cada tipo de tempo.

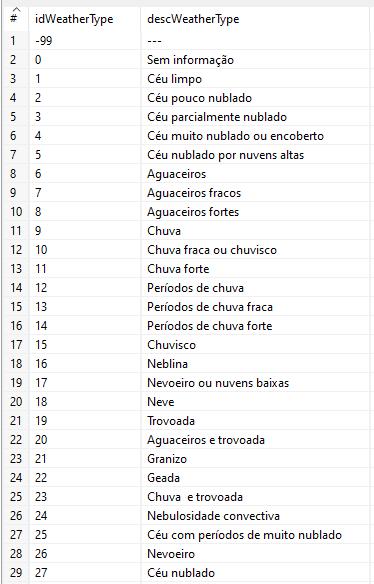
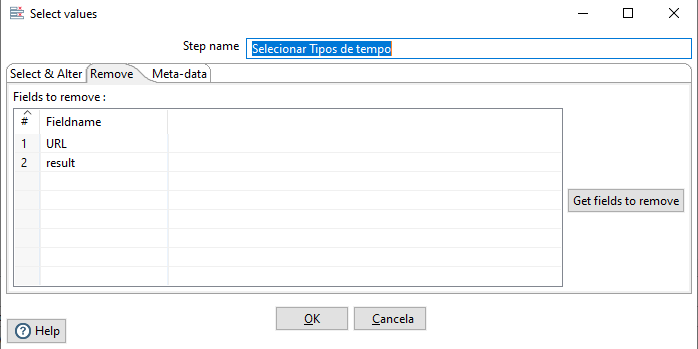


Depois da ordenação a tabela resultado é a seguinte:

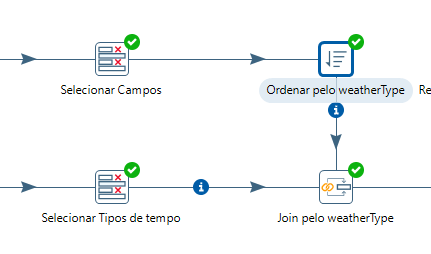


O processo de obter a descrição dos tipos de tempo é semelhante ao anterior, é criado uma linha com um Generate Rows (Url weatherType), com o link (<http://api.ipma.pt/open-data/weather-type-classe.json>), depois o Rest Client 2 efetua o download do json, os campos inseridos no Json Input (WeatherType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

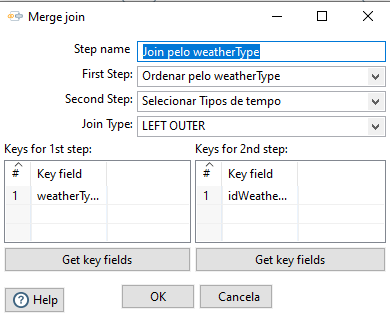
Depois é utilizado um Select Values (Selecionar Tipos Tempo) para excluir os campos URL e result.

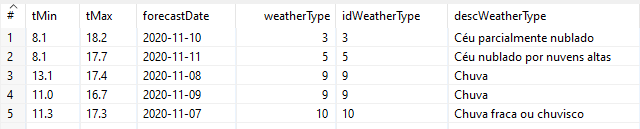
A tabela resultado dos tipo de tempo é:

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

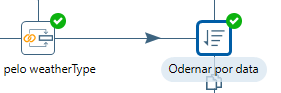


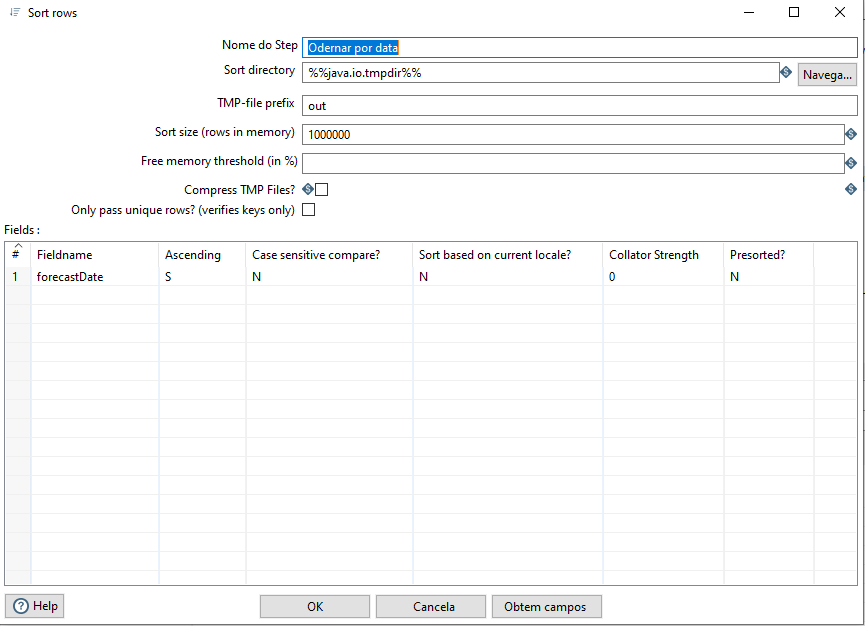
O Merge Rows permite fazer o cruzamento dos dados de uma tabela com a outra por determinado valor chave, neste caso o weatherType.

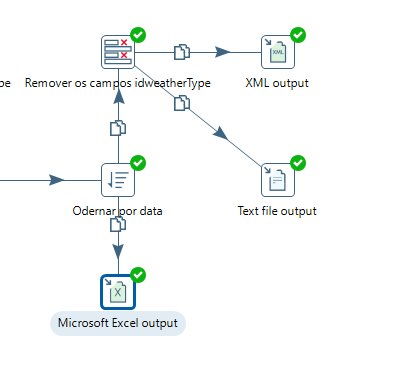


O resultado do Merge Rows é:

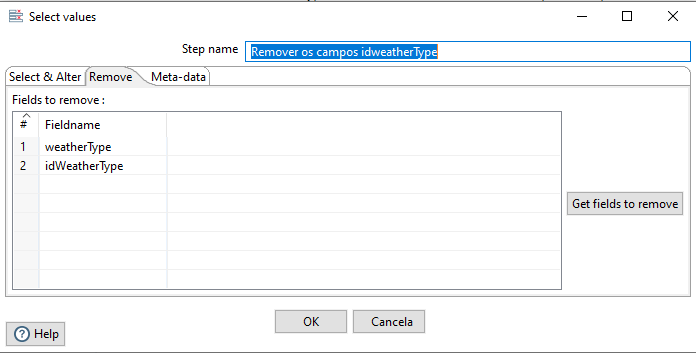
Esta tabela é de seguida ordenada por ordem crescente pelo campo forecastDate, no Sort Rows (Ordenar por data).

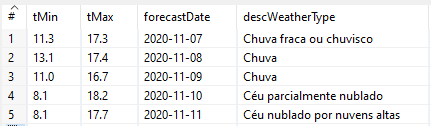




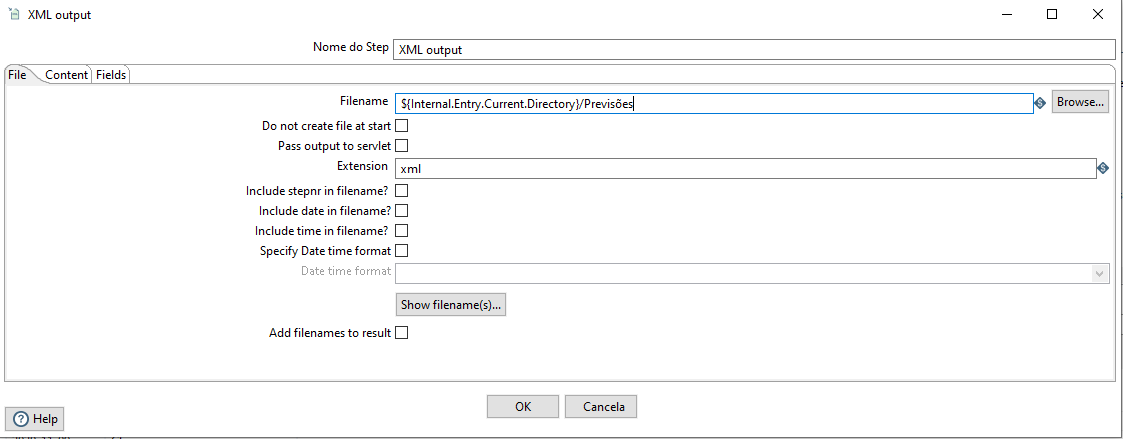


Depois é efetuada a remoção dos campos weatherType e idWeatherType através do Select Values (Remover os campos idweatherType) e o output é enviado para três ficherios nos formatos: xml, txt e xls.

Remoção dos campos weatherType e IdWeatherType

A tabela que resulta é :

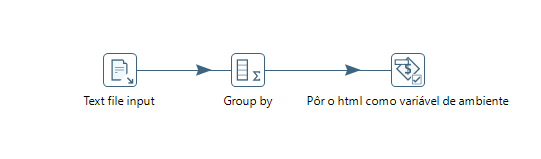
No fim é gerado um xml com o nome de Previsões que contem a tabela.

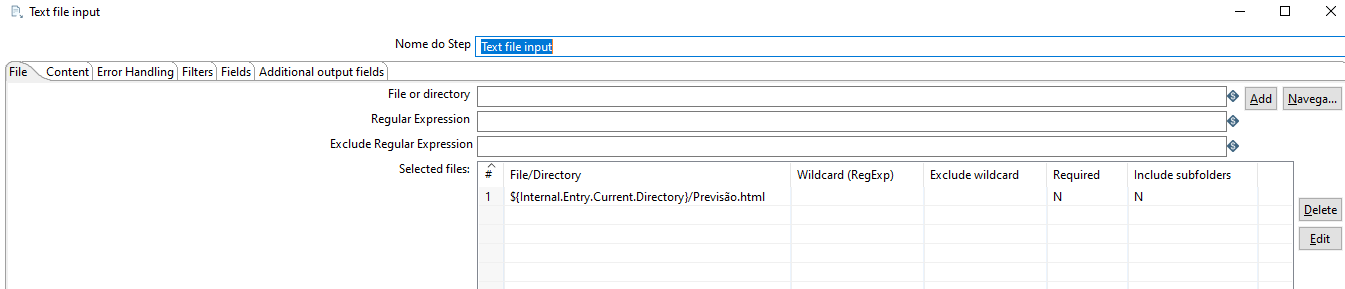
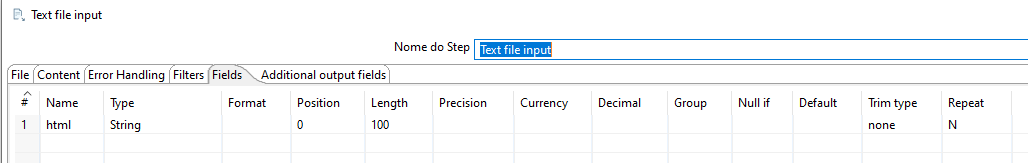


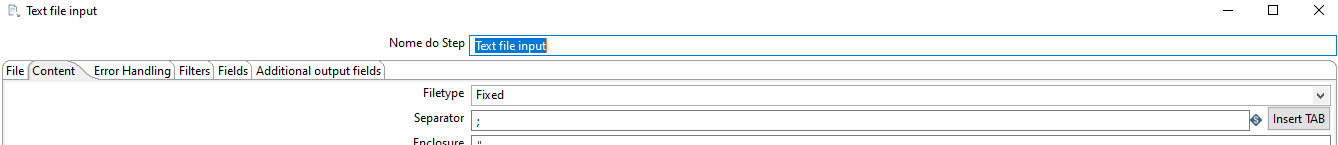
O conteúdo do xml é o seguinte:



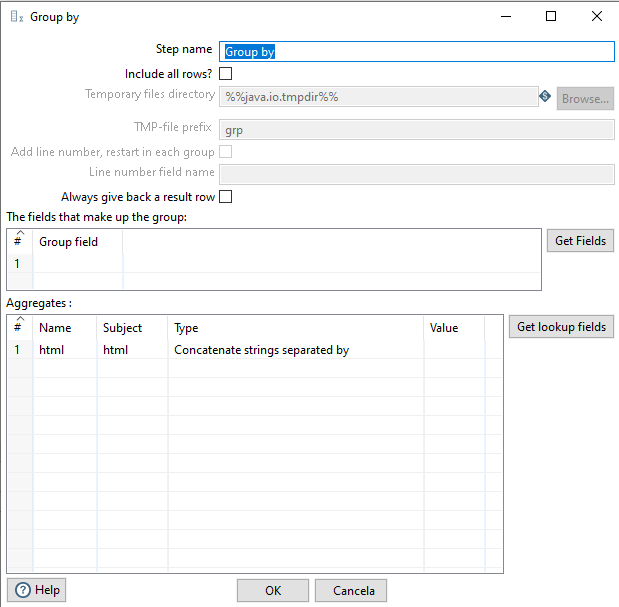
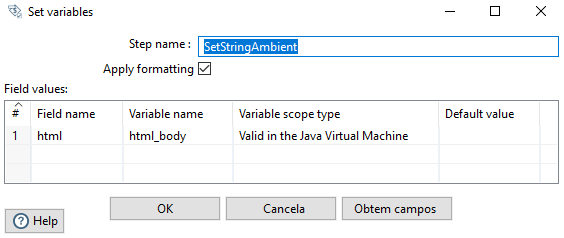
1.2 htmlAsAmbient

 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.

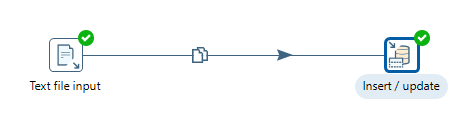


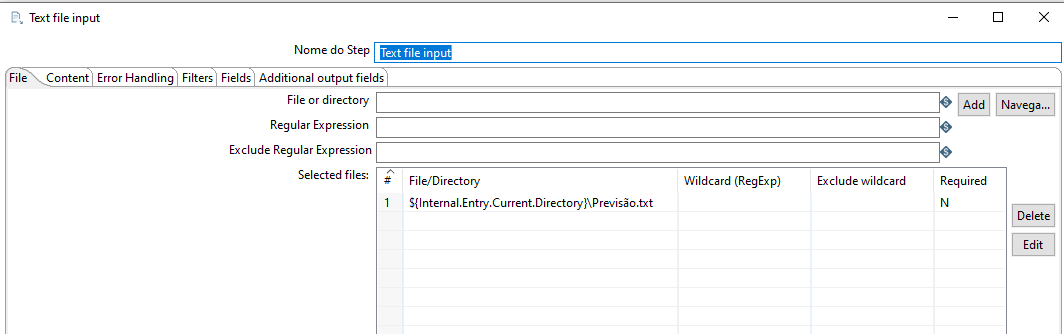
É utilizado um Group by para a concatenação das strings lidas numa só e depois essa string final é passada para um Set Variables (SetStringAmbient) para pôr a string como variável de ambiente para ser passada para o email.

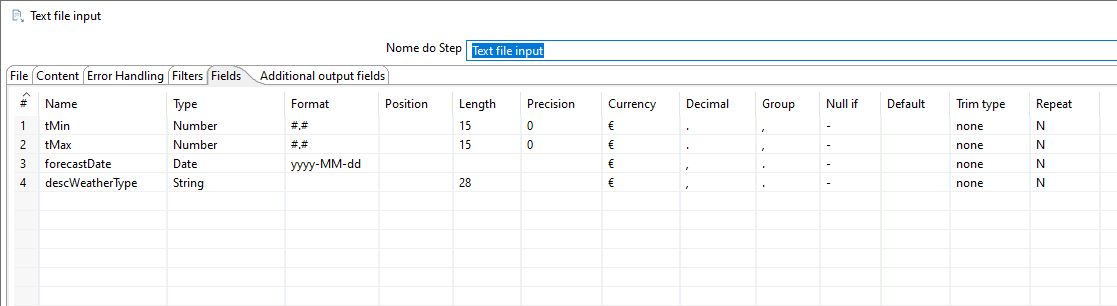


1.3 InsereBD

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



Configuração do Text file input:



Configuração do Insert / Update:

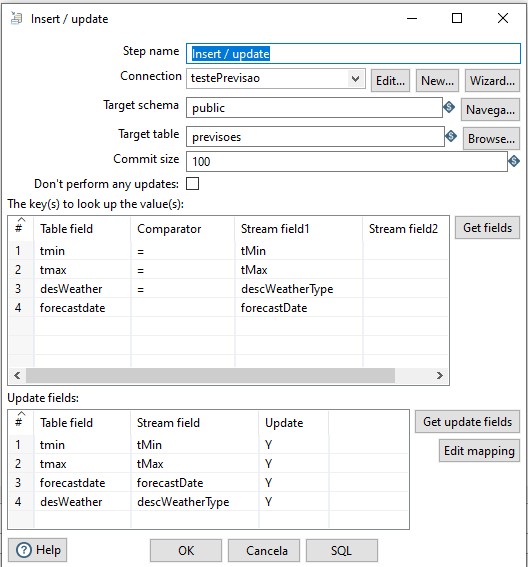
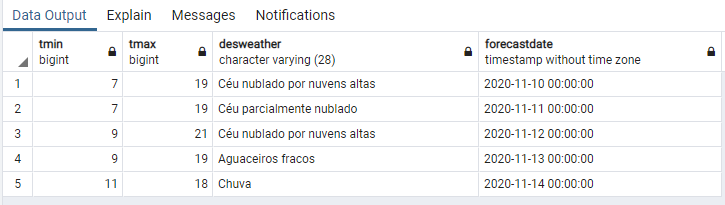
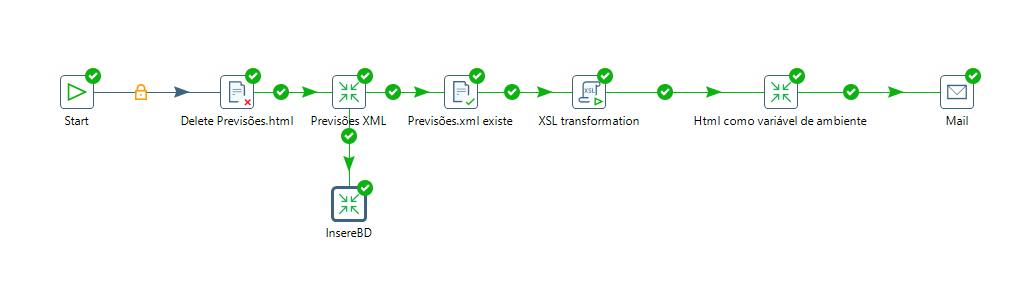


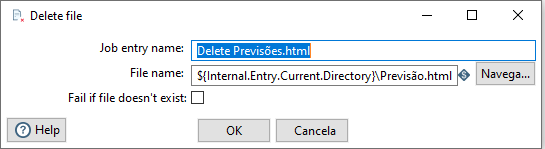
Tabela resultante na base de dados:

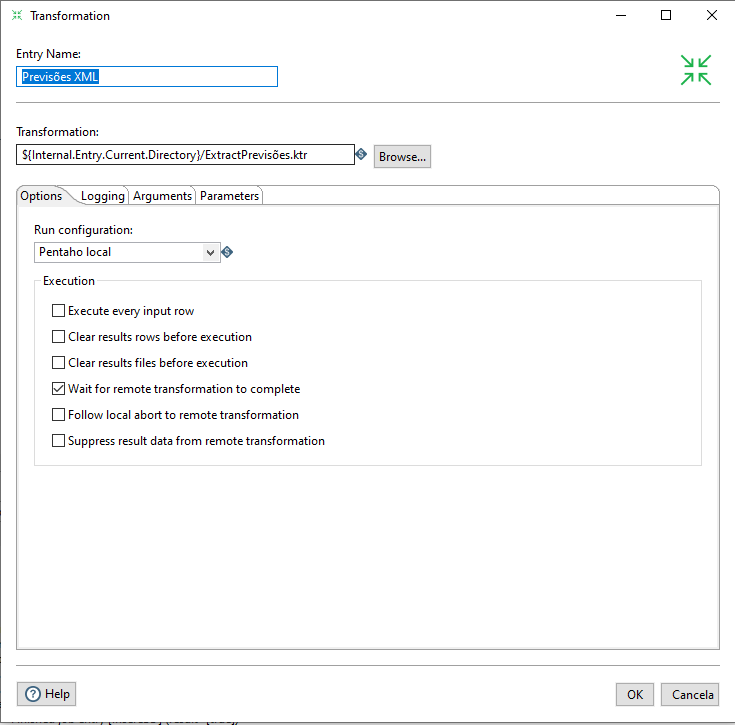


1. Jobs
   1. EmailPrevisão



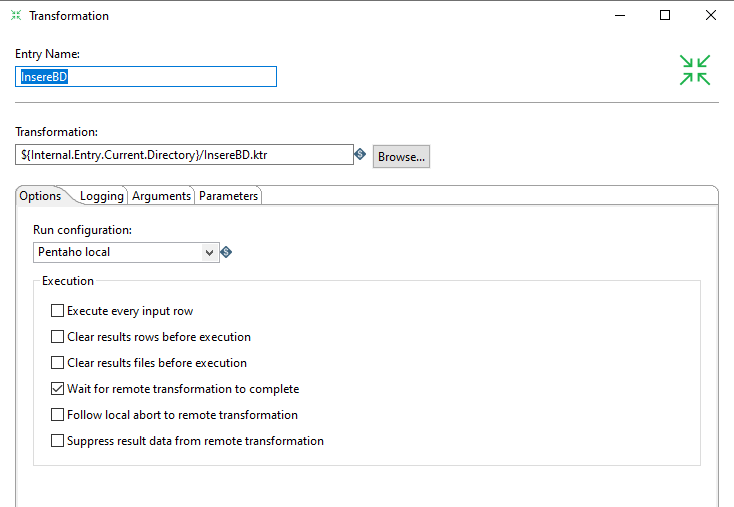
O job é iniciado pelo Start, que irá apagar o ficheiro Previsões.html e depois executar a transformação Previsões XML que é a transformação Email com previsão, de seguida é iniciado a transformação InsereBD e a verificação se o xml existe para ser aplicado uma XSL transformation no xml gerado pela transformação 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:

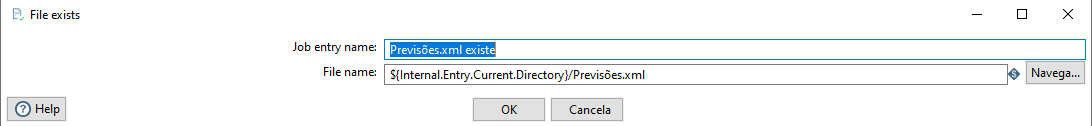
Configuração da transformação ExtractPrevisões:

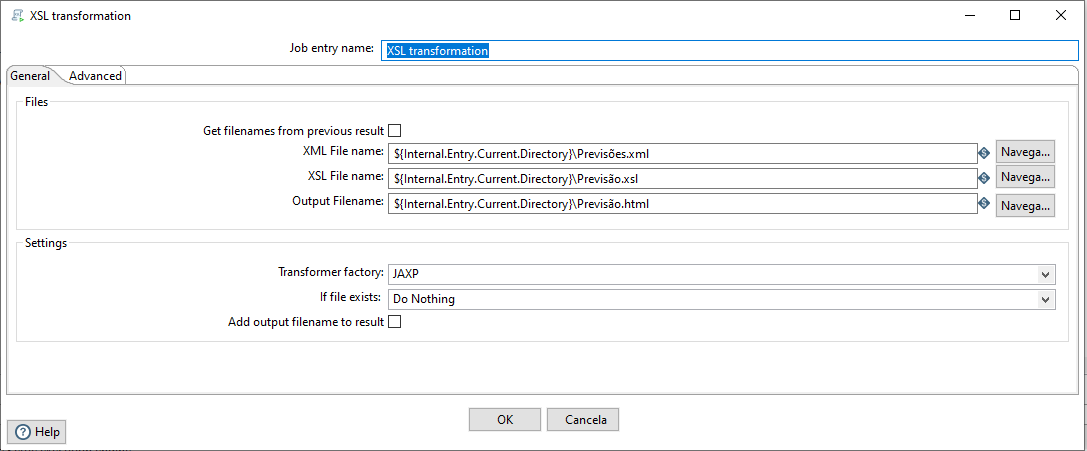
No fim da execução da transformação ExtractPrecisões é verificado se o xml existe e executada a transformação InsereBD.

Configuração de InsereBD:



Configuração de Previsões.xml existe:



Configuração da XSL Tranformation:

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>

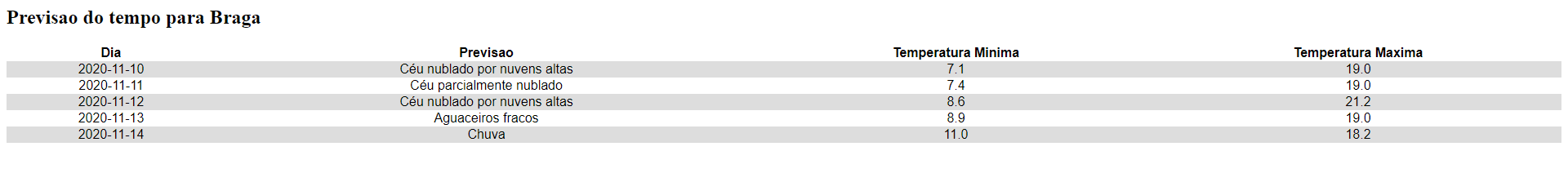
    </xsl:for-each>

  </table>

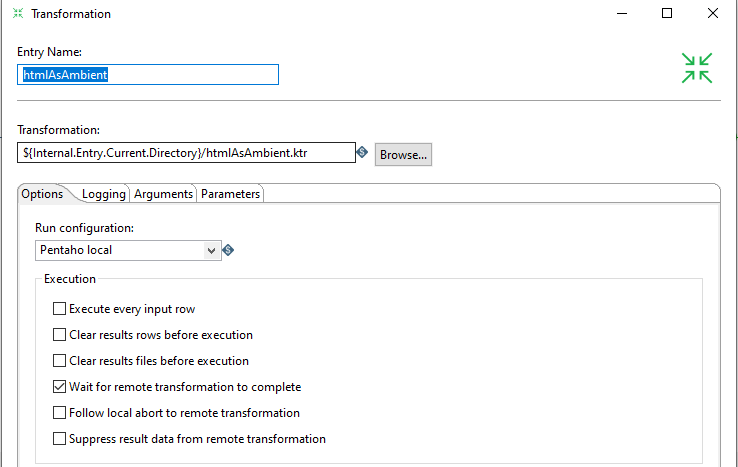
</body>

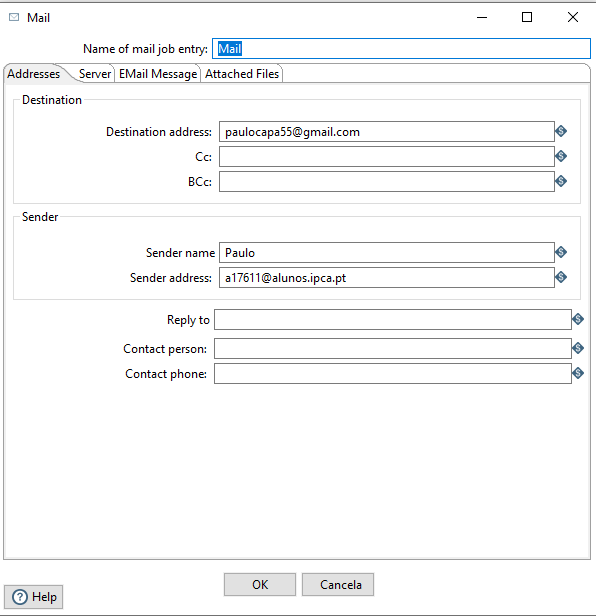
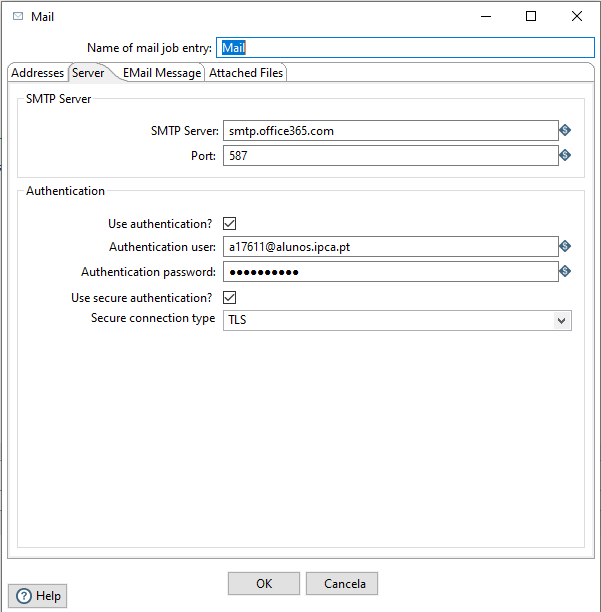
</html>

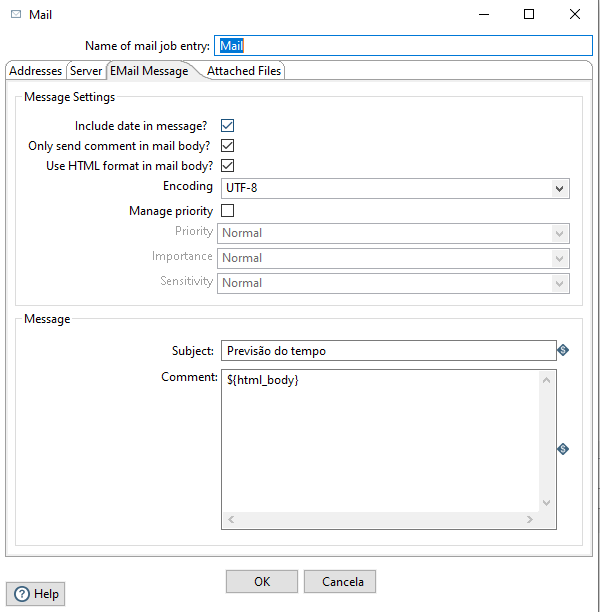
</xsl:template> </xsl:stylesheet>

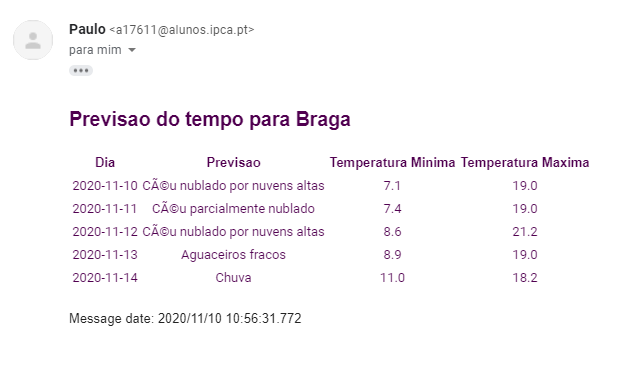
Depois da transformação obtemos uma tabela da seguinte forma:

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



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

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

No final o email recebido é o seguinte:

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.