



**Instituto Politécnico do Cávado e do Ave**

Licenciatura em Engenharia Sistemas Informáticos

Disciplina de Integração de Sistemas Informáticos

Docente: Luis Gonzaga Martins Ferreira

## **Relatório de Trabalho Prático**

Oleksandr Sierov nº16991

Barcelos, novembro 2020

# Índice

Índice.....	2
Índice de Figuras.....	3
Introdução.....	4
Tema .....	4
Estrutura do Trabalho (Job).....	5
Filter.....	6
Adicionar a Base de Dados .....	6
GetFiles.....	7
Conclusão.....	8
Bibliografia.....	9

## Índice de Figuras

Figura 1 - Do Job.....	5
Figura 2 - Transformação “Filter” .....	6
Figura 3 - Transformação “Adicionar a Base de Dados” .....	6
Figura 4 - Transformação GetFiles .....	7

## **Introdução**

No âmbito da disciplina de Integração de Sistemas de Informação, nos foi proposto a realização de um trabalho pretendendo-se focar a aplicação e experimentação de ferramentas em processos de ETL com um tema a nossa escolha. O objetivo de trabalho é, utilizando ferramentas como Pentaho Kettle desenvolver um trabalho para explorar processos de ETL.

### **Tema**

O tema escolhido por mim será uma lista de filmes da Netflix que vai ser processada por algumas transformações, submetida em uma base de dados, zipada e enviada para o email.

## Estrutura do Trabalho (Job)

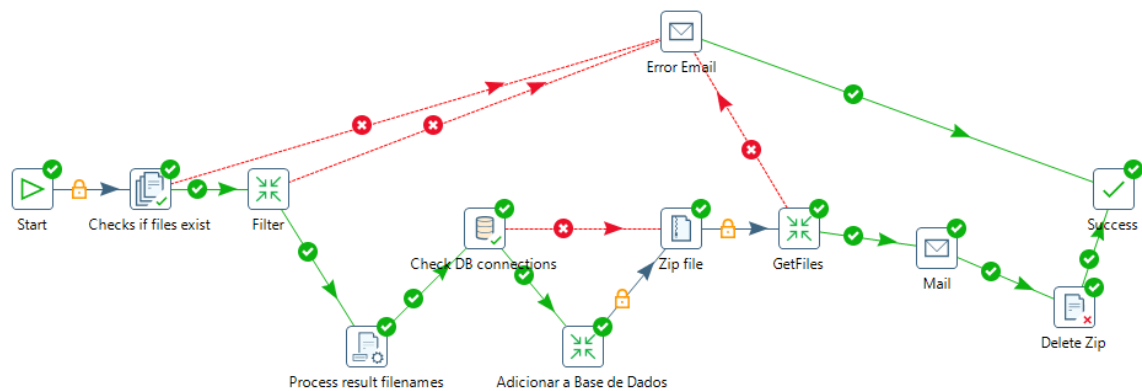


Figura 1 – Do Job

Neste passo inicial temos o Job, é dos arquivos mais importantes para o funcionamento correto do trabalho. Há um começo, após faz uma verificação a existência do arquivo que ira ser utilizado no “Filter”, caso o arquivo não exista será enviado um email com um error report. Após a passagem pelo Filter1 que (será explicada a seguir) temos um “Process result filenames”, este passo server para apagar os ficheiros que vão sair do “Filter”, mas só irão ser apagados da memoria (Eu efetuei este passo para depois ao enviar o email não serem enviados arquivos que não serão necessários).

No “Check DB connections” como o nome diz é verificada a conexão a base de dados, para onde posteriormente serão enviados ficheiros para tabelas. Caso a conexão esta estabelecida corretamente entra na transformação “Adicionar a Base de Dados”, caso não então salta esse passo.

Entrando agora na parte “Zip File”, é criado um Zip onde vai guardar dados executados na “Filter”, após isso, em “GetFiles” é efetuada a seleção dos ficheiros a serem enviados por email em “Mail”, após apaga o zip e termina a Job.

## Filter

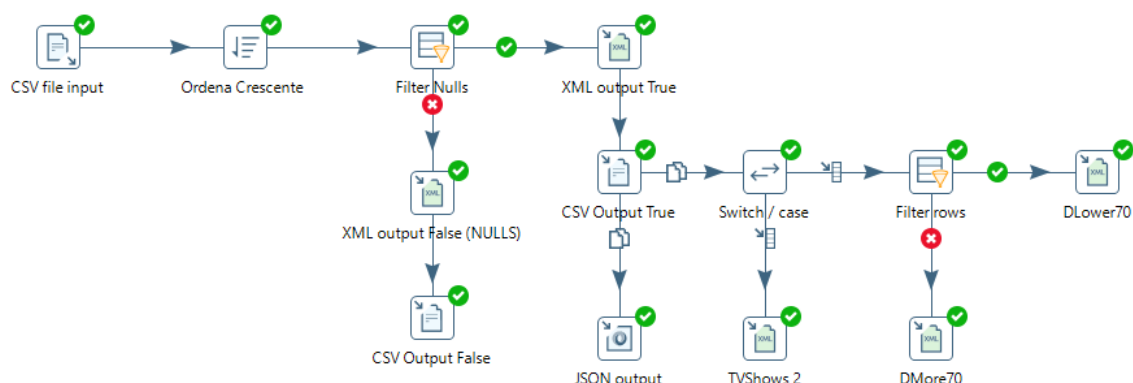


Figura 2 – Transformação “Filter”

Nesta transformação entra o arquivo CSV que foi verificado no Job, ordena-o de forma crescente, após isso faz uma filtragem e retira todas as linhas que tenham campos em Null, envia para um output de XML, CSV e de Json apenas a que foi limpa. Após isso passa por um switch case e divide em dois arquivos, um apenas com TV Shows e outro apenas com Movies. Os Movies não têm logo um output, usando uma expressão regular são divididos em filmes com menos de 70 min e filmes com mais de 70 min.

## Adicionar a Base de Dados

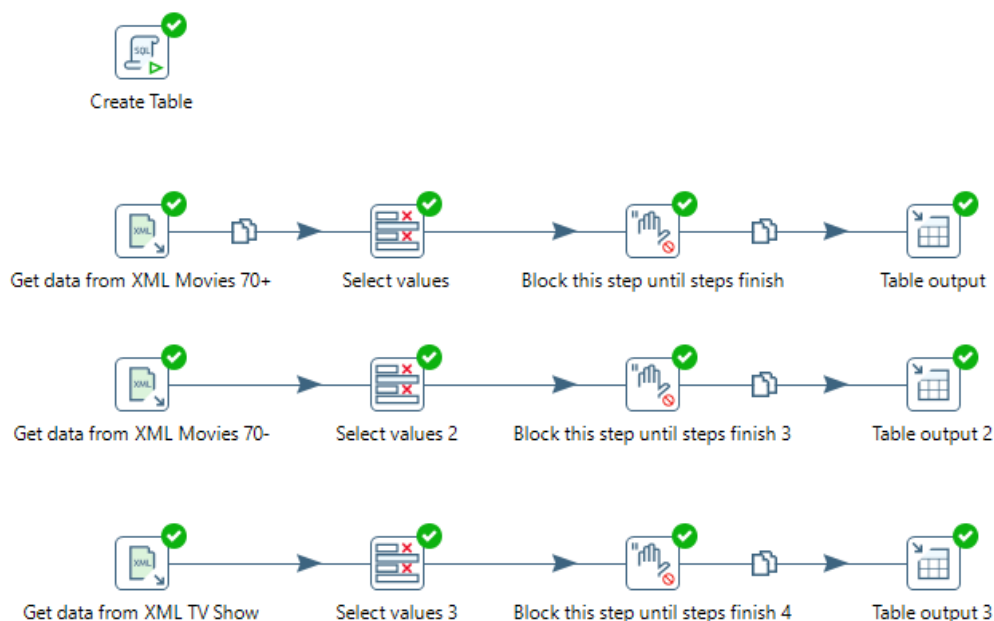


Figura 3 – Transformação “Adicionar a Base de Dados”

“Create Table”- Executa o SQL para Dropar e Criar tabelas onde serão inseridos os valores. De seguida, pega nos XML criados em “Filter”, escolhe as colunas a ser enviadas para a base de dados, espera que o “Create table” acabe de executar e insere nas tabelas da base de dados.

## GetFiles



*Figura 4 – Transformação GetFiles*

Esta transformação pega nos ficheiros seleccionados em Get File Names e coloca como um ficheiro de Result para posteriormente enviar no email.

## Conclusão

O trabalho foi desenvolvido com bastante sucesso. Houve algumas dificuldades como por exemplo uma tentativa de enviar ficheiros xml para o mail, mas estes ficavam bloqueados, a solução foi zipar os arquivos. Houve uma tentativa de usar uma expressão regular para dividir os TVShows dos Movies mas conclui que talvez usar um switch iria variar um pouco o trabalho. Apesar de algumas pequenas dificuldades o trabalho, a meu ver, contem uma boa parte das ferramentas disponibilizadas para a trabalhar com ETL. Seria possível adicionar mais qualquer coisa? Definitivamente sim. Concluindo assim, eu gostei de trabalhar com este software.



## **Bibliografia**