



PROJETO nrº 1 - ISI - LESI

(Engenharia de Sistemas Informáticos)

13.11.2020

Daniel Silva

Nr 17637

Índice

Índice	1
Índice Figuras	2
Introdução	3
Contextualização	4
Implementação de funcionalidades	5
Job	5
Transformation	6
Conclusão	8
Bibliografia	9

Índice Figuras

Figura 1 - Contextualização	3
Figura 2 - Job	4
Figura 3 - Envia Email	4
Figura 4 - JSON Input	5
Figura 5 - Dados que serão carregados	5
Figura 6 - Cálculo dos dados	6
Figura 7 - Contexto Geral	6

Introdução

Este documento tem como objetivo relatar de uma forma sucinta o primeiro trabalho prático realizado no âmbito da unidade curricular, integração de sistemas de informação, lecionada no curso de licenciatura em engenharia de sistemas informáticos, pós-laboral no Instituto Politécnico do Cávado e do Ave.

O trabalho consistiu na implementação de um “extract, transform, load” (ETL), com recurso à ferramenta Pentaho – Kettle, onde é realizado o tratamento de informação proveniente de terceiros e que gera um 2 ficheiros finais como output. Este ficheiro contém todas as transformações aplicadas de acordo com uma estrutura resultante previamente definida.

Contextualização

Trabalho Realizado no contexto da disciplina de ISI curso LESI (Licenciatura em Sistemas de Engenharia Informática), aprendemos até então a utilizar xml/json formatos para manipular informação, transportá-la, guardá-la e aprendemos também ferramentas que permitem essa mesma manipulação. Este Projecto consiste nisso mesmo, a utilização, prática destas ferramentas e ter uma melhor percepção de como funcionam que vantagens podem ferramentas como estas oferecer. Eu optei por utilizar o Pentaho Data Integration (kettle) o objectivo é gerar tabelas classificativas de Futebol para qualquer liga, apenas é preciso de input dados sobre as equipas dessa mesma liga (nome da equipa, vitórias, derrotas, empates, etc ..) para este projecto os dados provieram de uma api (<https://www.api-futebol.com.br/>) em formato json.

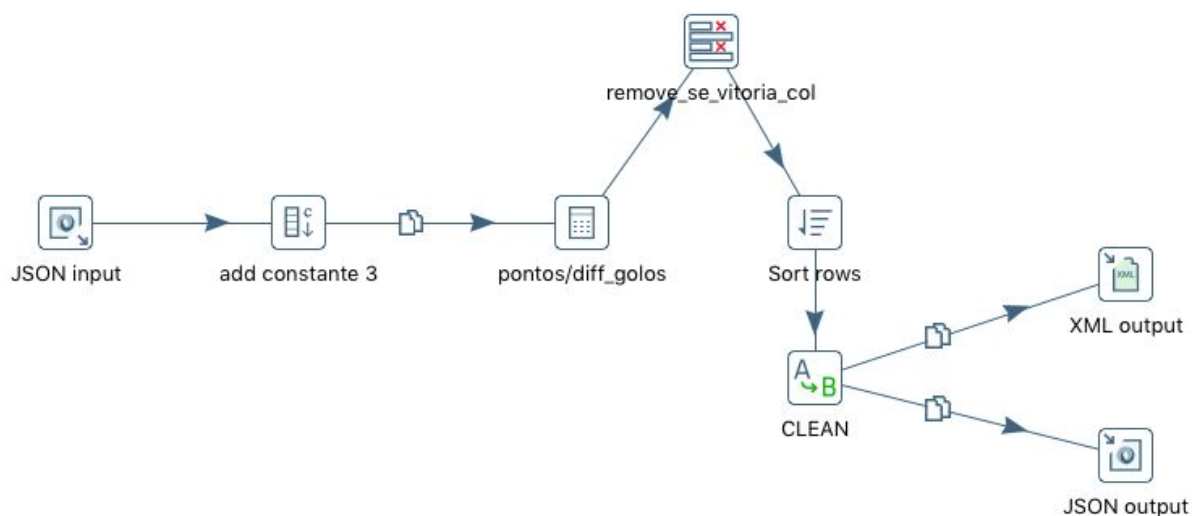


Figure 1 - Contextualização

Implementação de funcionalidades

- Job

O job serve para dar início ao programa e correr o processo de transformação onde toda a manipulação dos ficheiros e dados será efectuada.



Figure 2 - Job

Ainda na parte de job é onde será no fim de toda a execução se for efectuada com sucesso enviado um email notificando que a tabela classificativa foi gerada.



Figure 3 - Envia Email

● Transformation

Os dados serão carregados para o processo provenientes de um ficheiro json. Apenas alguns dados é que vão ser levados em consideração, ou seja, vai haver inicialmente uma filtragem.



Figura 4 - JSON Input

Step name JSON input				
File	Content	Fields	Additional output fields	
#	Name	Path	Type	Field
1	jogos	\$.taca-rio-fase-de-grupos.grupo-a[*].jogos	Integer	
2	vitorias	\$.taca-rio-fase-de-grupos.grupo-a[*].vitorias	Integer	
3	empates	\$.taca-rio-fase-de-grupos.grupo-a[*].empates	Integer	
4	derrotas	\$.taca-rio-fase-de-grupos.grupo-a[*].derrotas	Integer	
5	gols_pro	\$.taca-rio-fase-de-grupos.grupo-a[*].gols_pro	Integer	
6	gols_contra	\$.taca-rio-fase-de-grupos.grupo-a[*].gols_contra	Integer	
7	nome_popular	\$.taca-rio-fase-de-grupos.grupo-a[*].time.nome...	String	

Figura 5 - Dados que serão carregados

Posteriormente adicionamos uma constante pela qual iremos multiplicar os dados filtrados mais nomeadamente (vitórias) de maneira a sabermos a devida pontuação de cada equipa, somamos este resultado com o número de empates, fazemos também o mesmo para a diferença de golos em que subtraímos golos marcados pelos sofridos.



Figura 6 - Cálculo dos dados

Depois dessa operação, retiramos a coluna com a nossa constante e ordenamos os dados do 1º classificado para o último.

Fazemos também uma operação que é uniformizar o nome das equipas, aplicando uma expressão regular.

No fim guardamos os dados já tratados e filtrados em 1 ficheiro JSON e XML.

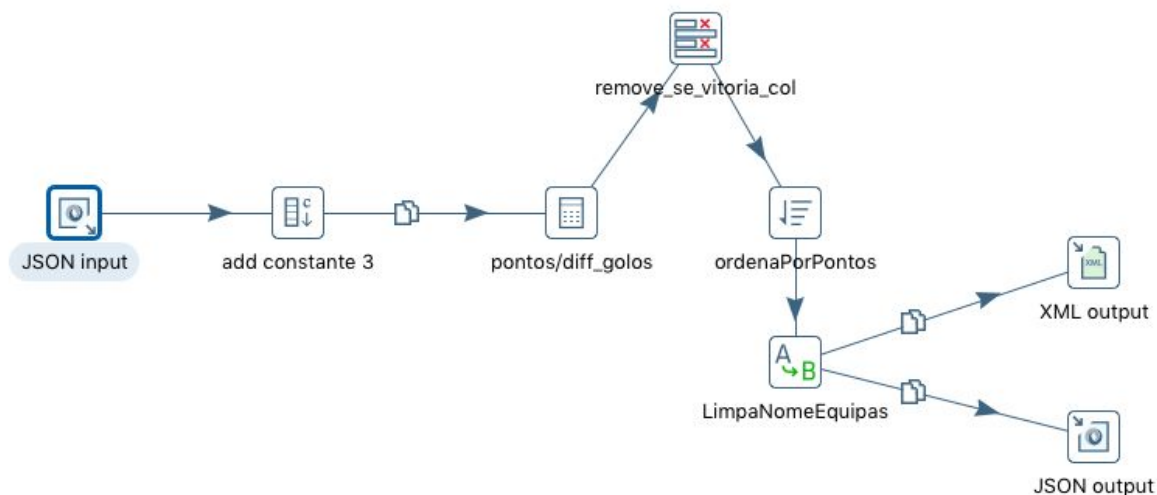


Figura 7 - Contexto Geral

Conclusão

A implementação deste ETL foi feita com facilidade, inicialmente pensei que viesse a dar mais trabalho por não entender bem o software (kettle) mas com algum trabalho de pesquisa foi fácil de entender o funcionamento e claramente que usar softwares deste tipo são uma grande vantagem.

Antes de trabalhar com o software tentei fazer o mesmo mas com uma linguagem de programação, usei PHP por nunca ter usado e pareceu-me a oportunidade perfeita, mas admito que perdi muito tempo só a tentar entender a linguagem e aplicar exatamente o pretendido com ela, contudo penso que estive muito perto de uma solução, e pretendo ainda encontrá-la.

Bibliografia

<https://www.api-futebol.com.br/>

<https://help.pentaho.com/Documentation/9.1/>

[https://wiki.pentaho.com/display/EAI/Mail+\(step\)](https://wiki.pentaho.com/display/EAI/Mail+(step))

<https://www.youtube.com/watch?v=oSap0SUueqg>

<https://ipca.pt/> (moodle)