

Relatório

Trabalho prático 1

Integração de Sistemas de Informação

Docente: Óscar Ribeiro

Realizado por:

Diogo Gonçalves (a26006)

Licenciatura em Engenharia Sistemas Informáticos

3º ano

Barcelos, 25 de outubro de 2024

Índice

Índice de figuras	3
Introdução	4
Contextualização	5
Transformações	6
getFromAPI	6
filterDataFromAPI	7
htmlSetAsAmbient.....	8
sendEmailFromDB	9
Utilização de expressões regulares(regex)	9
getLogs.....	10
Trabalhos	11
project.....	11
Demonstração	12
Conclusão.....	13
Bibliografia.....	14

Índice de figuras

Figura 1 - Transformação getFromAPI.....	6
Figura 2- Transformação filterDataFromAPI	7
Figura 3 - Transformação htmlSetAsAmbient	8
Figura 4 - Transformação sendEmailFromDB	9
Figura 5 - Expressões regulares: resultado da procura, com vários casos de uso	9
Figura 6 - Transformação getLogs	10
Figura 7 - Trabalho project	11
Figura 8 - QRCode com vídeo comprovativo do funcionamento do trabalho	12
Figura 9 - Comprovativo email de receção de produtos em desconto, de uma categoria específica, com valor inferior a 80€	12

Introdução

No contexto atual de crescente digitalização e transformação dos negócios, a integração eficaz de sistemas de informação tornou-se um fator estratégico para o sucesso organizacional. Processos de integração que garantam a correta extração, transformação e carregamento de dados (ETL) são fundamentais para permitir que empresas mantenham a coesão e o valor dos dados entre diferentes sistemas, ferramentas e plataformas.

Este trabalho tem como objetivo consolidar o conhecimento sobre integração de sistemas de informação a nível de dados, explorar novas ferramentas e tecnologias, e desenvolver as capacidades técnicas na construção de soluções ETL, contribuindo assim para a melhoria da assimilação de conceitos e metodologias apresentadas ao longo da unidade curricular.

Contextualização

O tema escolhido foi a obtenção de produtos, com certas características de uma loja online, utilizando a plataforma Pentaho Data Integration.

Neste processo, pode se verificar a obtenção de dados de uma API, a criação e respetiva configuração de um ficheiro XSL, a filtragem, seleção, ordenação e eliminação de dados, verificação de existência de ficheiros, transformação de variáveis em variáveis de ambiente e importação/exportação de dados entre ficheiros CSV, JSON, XML, XLS, HTML. Também é possível comprovar o acesso à base de dados, utilização de expressões regulares(regex) com o objetivo de filtragem, utilização de script em javascript para criação de variáveis, encriptação de palavra-passe e o envio de emails. Por fim, ainda é possível ver a geração de ficheiros de logs.

Transformações

getFromAPI

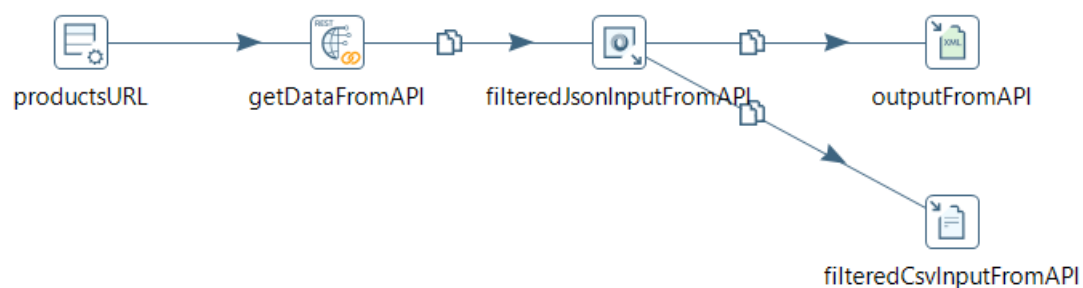


Figura 1 - Transformação `getFromAPI`

Nesta transformação foi realizada a obtenção de dados de uma API(<https://dummyjson.com/products>), que foram carregados para um ficheiro JSON.

A partir desse JSON, apenas os dados desejados, são carregados para dois tipos de ficheiros diferentes, CSV e XML.

filterDataFromAPI

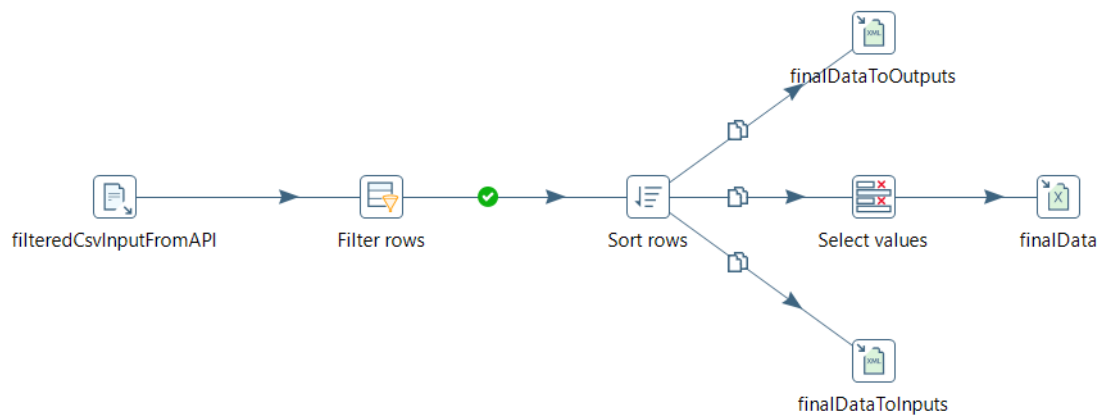


Figura 2- Transformação filterDataFromAPI

Nesta transformação, foi utilizado o ficheiro(filteredCsvInputFromAPI.csv) gerado na transformação anterior (getFromAPI), para a filtragem de dados. De seguida, os dados foram organizados, por ordem ascendente, pelo seu preço. Estes dados são guardados em 2 ficheiros XML diferentes, para serem futuramente consultados/utilizados na criação de um ficheiro html. Pelo seguimento da linha central, foram removidas colunas que já não eram relevantes e posteriormente, os dados relevantes foram guardados num ficheiro excel(XLS).

htmlSetAsAmbient

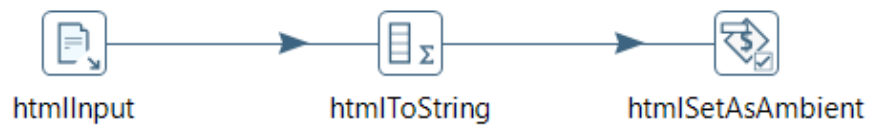


Figura 3 - Transformação htmlSetAsAmbient

Nesta transformação, foi utilizado o conteúdo do ficheiro html, gerado num passo anterior no trabalho, para passar para string, que posteriormente foi definida como uma variável ambiente para ser utilizada no envio de um email.

sendEmailFromDB



Figura 4 - Transformação sendEmailFromDB

Nesta transformação, é realizada a obtenção de dados de utilizadores de uma base de dados, de uma tabela relacionada a utilizadores. Primeiro, é verificada a validade dos emails, utilizando a expressão regular apresentada em baixo (Figura 5). De seguida, foi realizada uma filtragem e uma seleção/eliminação para obter apenas os campos desejados. Como era necessária a existência de mais dados, foi utilizada a função de script para definir umas variáveis constantes para serem usadas de seguida, no envio do email. Por razões de segurança, foi necessária a encriptação da palavra-passe, como se pode verificar no script, apenas é apresentada a versão encriptada.

Utilização de expressões regulares(regex)

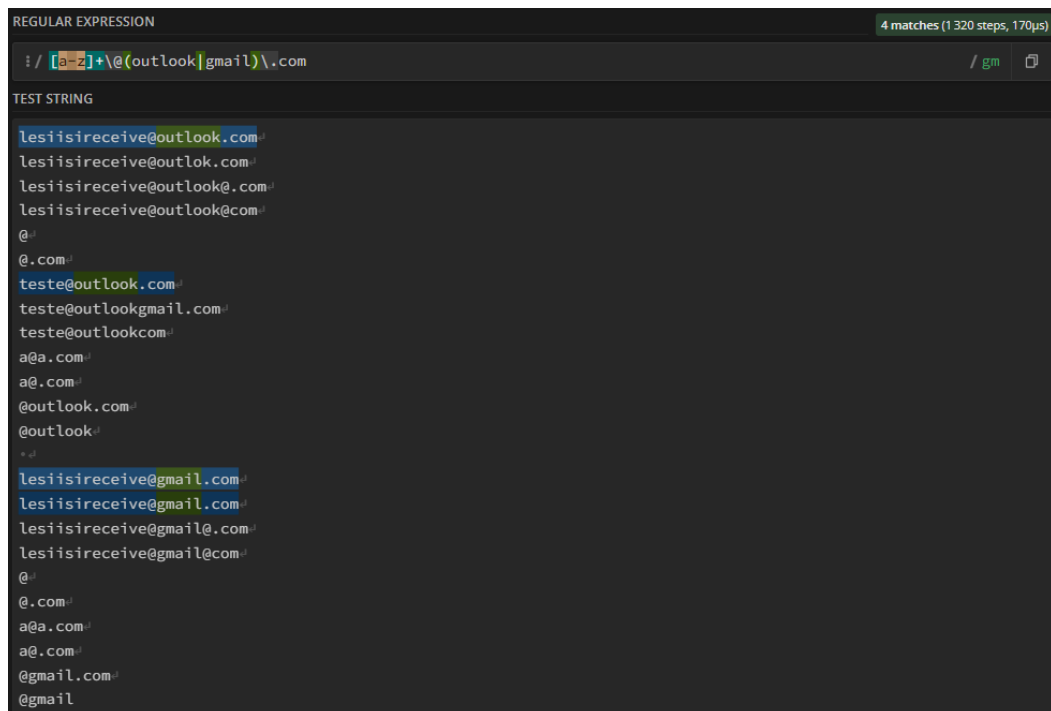


Figura 5 - Expressões regulares: resultado da procura, com vários casos de uso

getLogs

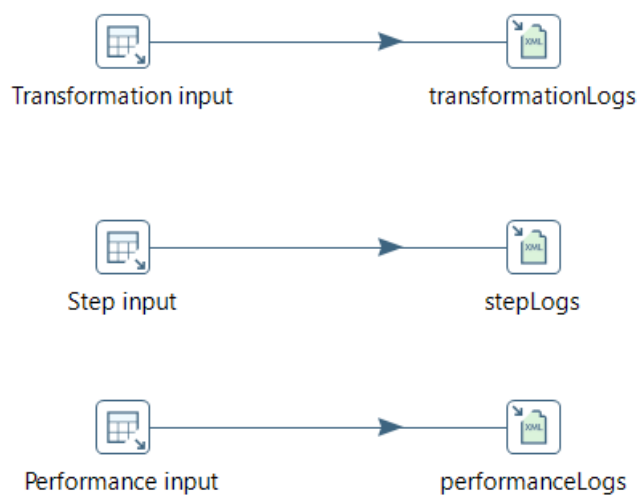


Figura 6 - Transformação getLogs

Nesta transformação é realizada a obtenção de dados de uma base de dados, de 3 tabelas diferentes, todas relacionadas a logs diferentes. Cada tabela vai dar origem a um ficheiro XML output diferente, com os respetivos dados da tabela.

Trabalhos

project

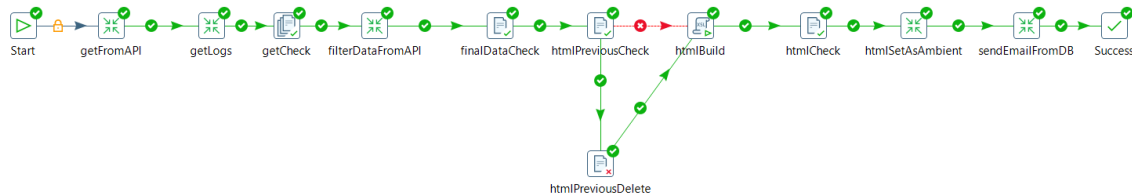


Figura 7 - Trabalho project

Neste trabalho, são utilizadas as transformações mencionadas anteriormente (figura 1, figura 2, figura 3, figura 4 e figura 6), de forma sequencial, junto com verificações de existência de ficheiros, antes de eles serem utilizados e a eliminação do ficheiro html, caso ele já exista, para caso exista alguma alteração no ficheiro xsl ou xml que o geram.

Neste caso é importante reparar que, os logs, apenas estão a ser obtidos da transformação getFromAPI, como forma de exemplo, porém era possível obter de todas as transformações.

Demonstração

De seguida, encontra-se um QR Code com um vídeo do trabalho a funcionar.

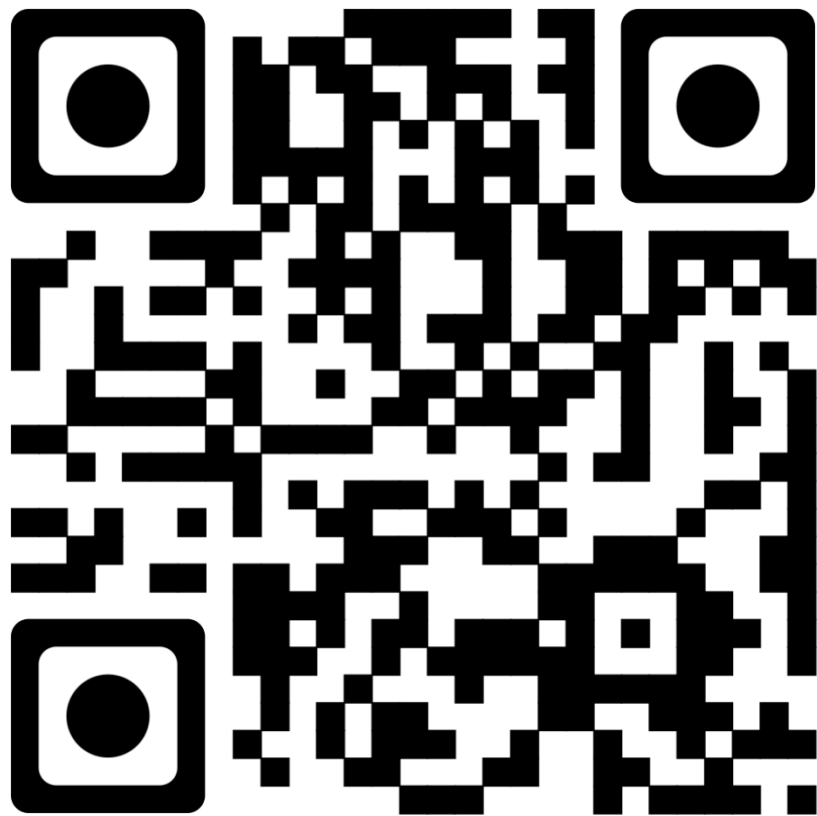


Figura 8 - QRCode com vídeo comprovativo do funcionamento do trabalho

Diogo Gonçalves

Para: Você

Responder

Responder a todos

Reencaminhar

...

seg, 25/10/2024 14:01

Olá Ana Santos os descontos são:

Fragrâncias com valores inferiores a 80€, em desconto

Imagem	Id	Título	Descrição	Categoria	Preço (€)	Desconto (%)	Stock
	6	Calvin Klein CK One	CK One by Calvin Klein is a classic unisex fragrance, known for its fresh and clean scent. It's a versatile fragrance suitable for everyday wear.	frangrances	49.99	0.32	17
	9	Dolce Shine Eau de	Dolce Shine by Dolce & Gabbana is a vibrant and fruity fragrance, featuring notes of mango, jasmine, and blonde woods. It's a joyful and youthful scent.	frangrances	69.99	11.47	3
	10	Gucci Bloom Eau de	Gucci Bloom by Gucci is a floral and captivating fragrance, with notes of tuberose, jasmine, and Rangoon creeper. It's a modern and romantic scent.	frangrances	79.99	8.90	93

Message date: 2024/10/25 15:01:41.431

Figura 9 - Comprovativo email de receção de produtos em desconto, de uma categoria específica, com valor inferior a 80€

Conclusão

Neste trabalho consegui consolidar e aprofundar os meus conhecimentos sobre os processos ETL, e sobre o que é integração de dados e as suas aplicações. Aprendi também diferentes tipos de ficheiros, as melhores maneiras para usufruir deles e a importar/exportar dados entre eles.

No futuro pretendo melhorar o meu conhecimento em outras plataformas

Bibliografia

<https://pentaho.com/products/pentaho-data-integration/>

<https://dummyjson.com>

<https://regex101.com/>

<https://wettransfer.com/>

<https://www.freeconvert.com/mov-to-mp4>

<https://www.freeconvert.com/pt/video-compressor>

<https://www.qrcode-tiger.com/>

<https://github.com/25bczz/pentaho-data-integration>