



# INTEGRAÇÃO DE SISTEMAS DE INFORMAÇÃO

Docente: Óscar Ribeiro

## Trabalho Prático I

Com este trabalho da Disciplina de Integração de Sistemas de Informação (ISI) pretende-se focar a aplicação e experimentação de ferramentas em processos de ETL (Extract, Transformation and Load), inerentes a processos de Integração de Sistemas de informação ao nível dos dados.

Marco António Ramoa Macedo

Número 26874

22 de outubro de 2024

## Conteúdo

1	Problema.....	3
2	Estratégia.....	4
3	Transformações.....	5
3.1	Obter dados da API.....	6
3.2	Junção de dados.....	7
3.3	Filtração de dados .....	8
3.4	Definir ambiente de string.....	9
4	Jobs .....	10
4.1	Enviar dados processado por email.....	10
5	Node-Red .....	12
6	Vídeo .....	14
7	Conclusão .....	15

## Índice de Figuras

Figura 1 Primeira Transformação - Extração de dados.....	6
Figura 2 Segunda Transformação - Junção de dados.....	7
Figura 3 Terceira transformação - Filtragem de dados.....	8
Figura 4 Quarta transformação - string de ambiente .....	9
Figura 5 Primeiro Job - Enviar Email.....	10
Figura 6 Definições Email.....	11
Figura 7 Node-Red .....	12
Figura 8 Node-Red - Dados utilizadores.....	12
Figura 9 Node-Red - Dados de moradas.....	13

# 1 Problema

Este projeto, pretende demonstrar o uso de ferramentas ETL para que seja possível automatizar a extração, transformação e carregamento de dados. No caso deste projeto, poderá ser possível aceder às informações de alguma pessoa. O objetivo é desenvolver um processo automatizado, que sendo executado, realize a consulta a dados externos, obtendo assim detalhes acerca de várias pessoas.

A integração de sistemas de informação é complexa, especialmente no contexto da gestão de dados pessoais. A diversidade de fontes de dados e a necessidade de garantir a qualidade e a consistência das informações são desafios significativos. Com o crescimento exponencial da quantidade de dados disponíveis, um processo de ETL bem implementado torna-se essencial para garantir que os dados sejam extraídos, organizados e normalizados de maneira eficaz, possibilitando a utilização confiável e informada dessas informações.

## 2 Estratégia

A estratégia adotada envolve a utilização do Pentaho como ferramenta principal para o desenvolvimento do processo de ETL.

Os dados extraídos da API, vêm em formato JSON que vão ser processados de maneira que seja possível no final a elaboração de um pequeno ficheiro HTML, com todas as informações modificadas. Alguns destes campos serão, número telemóvel, primeiro nome, email, último nome, etc. Estes dados também serão inseridos na base de dados, na qual foi utilizada o SQL SERVER.

A visualização dos dados processados, poderão ser revistos por email ou quando utilizada a ferramenta Node Red.

### 3 Transformações

Foram realizadas quatro transformações para a execução do trabalho prático. Resumidamente, uma das transformações trata de extrair dados da API, de seguida existe uma transformação para a junção de dados, ocorre também filtragem de dados numa outra transformação e finalmente temos uma transformação que trata de converter os dados processados para html.

### 3.1 Obter dados da API

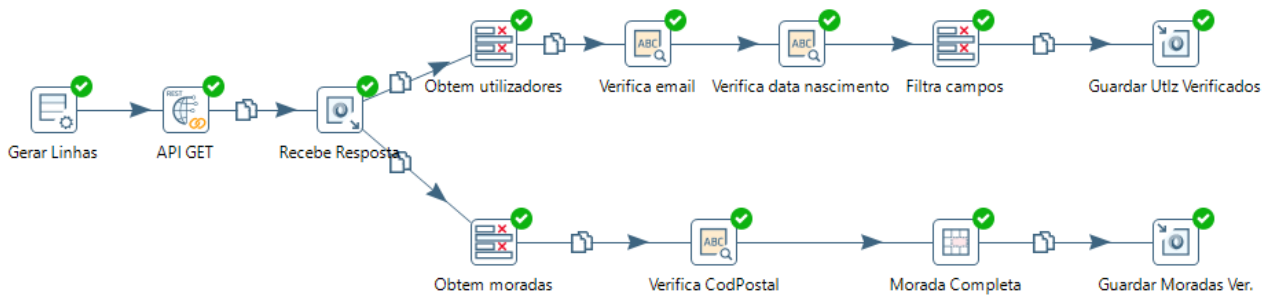


Figura 1 Primeira Transformação - Extração de dados

Na primeira transformação começamos por criar um step para gerar linhas que irá conter o URL para conseguir fazer o request à API. Obtemos a resposta com os campos que pretendemos. Depois de obtidos os dados, são seleccionados conforme o necessário, para o caminho de cima foram seleccionados todos os dados relevantes a utilizadores, para o caminho de baixo, todos os dados relacionados a moradas. Ambos os caminhos têm steps que utilizam expressões regex, para verificar dados que são necessários estarem corretos. No caso das moradas é também concatenado a morada, obtendo assim um novo campo que irá ter a morada completa. No final ambos os ficheiros são guardados.

### 3.2 Junção de dados

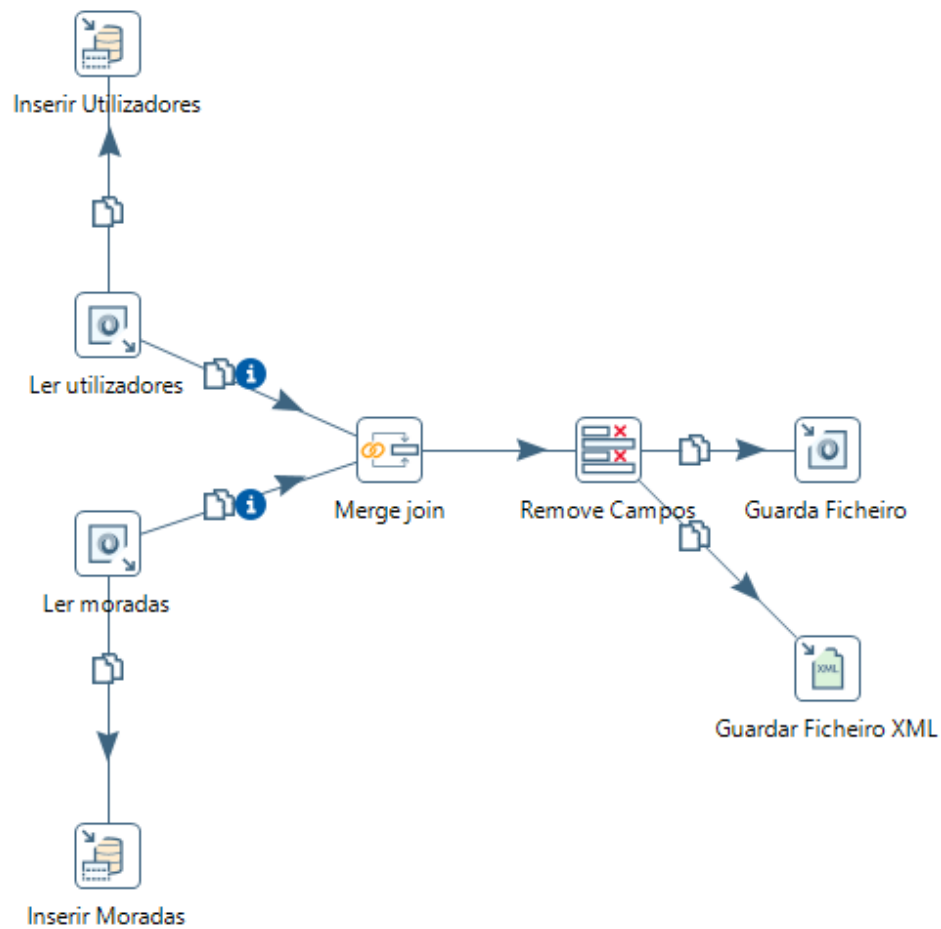
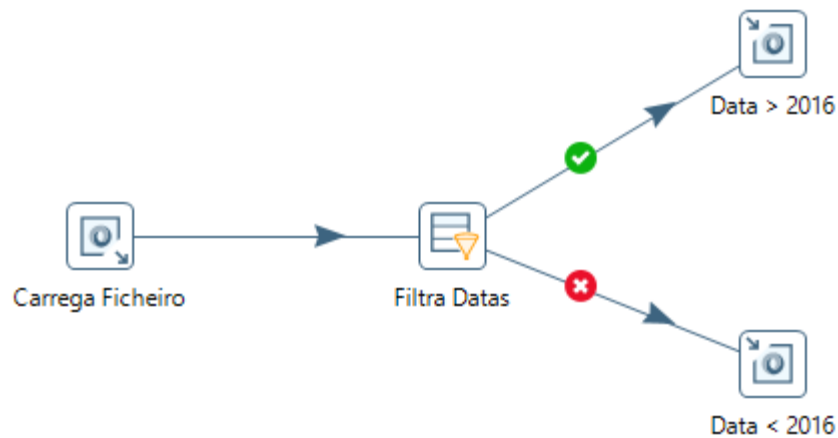


Figura 2 Segunda Transformação - Junção de dados

Na segunda transformação, são lidos os dois ficheiros anteriores, onde inicialmente são logo inseridos na base de dados. De seguida ambos os ficheiros são juntos, são removidos campos duplicados e não necessários e é criado um ficheiro json e um ficheiro XML.



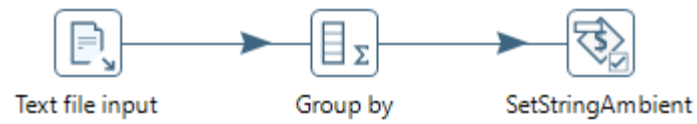
### 3.3 Filtração de dados



*Figura 3 Terceira transformação - Filtragem de dados*

Na terceira transformação, é carregado o ficheiro que foi gerado na segunda transformação, onde todas as informações foram juntas. Filtramos as datas de nascimento das pessoas, conforme maior ou menor do que 2016, obtendo assim dois ficheiros diferentes com resultados adequados à condição pedida.

### 3.4 Definir ambiente de string



*Figura 4 Quarta transformação - string de ambiente*

Na quarta e última transformação, que foi fornecido como exemplo pelo professor, é carregado um ficheiro que é gerado no Job (ver Figura 5) onde decorre uma transformação de ficheiro xml para html. Aqui definimos que queremos colocar todo o texto como html, assim sendo possível no Job a realização do step para enviar email.

## 4 Jobs

Foi realizada apenas um job, no qual consiste em percorrer todas as quatro transformações construídas e no final é enviado um email com os dados processados.

### 4.1 Enviar dados processado por email

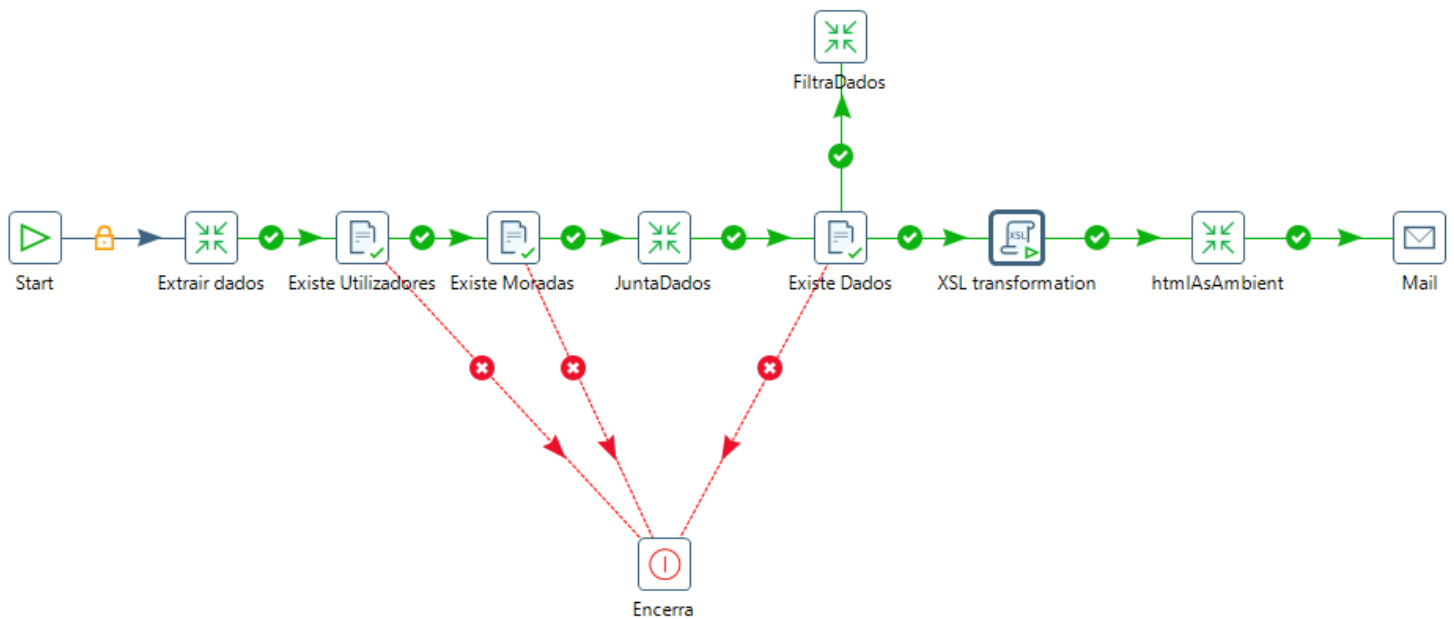


Figura 5 Primeiro Job - Enviar Email

Como foram citados anteriormente, este Job irá executar todas as transformações construídas e vai verificando se o ficheiro é gerado, para conseguir avançar para o próximo step, já na parte final do Job é feita uma transformação de um ficheiro xml para html, que foi referido na descrição da quarta transformação. Por final é enviado o email para a pessoa pretendida

Addresses		Server	Email Message	Attached Files
Destination		Destination address: a26874@alunos.ipca.pt		SMTP Server: smtp.office365.com
Cc:				Port: 587
Bcc:				Use authentication? <input checked="" type="checkbox"/>
Sender		Sender name: Marco		Authentication user: a26874@alunos.ipca.pt
Sender address: a26874@alunos.ipca.pt				Authentication password: *****
		Use secure authentication? <input checked="" type="checkbox"/>		Secure connection type: TLS

**Message Settings**

Include date in message? ☒  
Only send comment in mail body? ☒  
Use HTML format in mail body? ☒  
Encoding: UTF-8  
Manage priority: ☐  
Priority: Normal  
Importance: Normal  
Sensitivity: Normal

**Message**

Subject: Dados  
Comment: \${html\_body}

Figura 6 Definições Email

Na Figura 6 está representado todos os campos necessários para o envio de email a partir do Job desenvolvido. Foram definidos o servidor SMTP, a porta, o endereço de destino, nome, endereço de quem envia, password, tipo de conexão e autenticação segura. Na parte de email message foi incluído a data na mensagem e usado o formato html no corpo da mensagem. No subject é apenas o título da mensagem e no comment é passado o html\_body que foi recebido da transformação numero quatro.

# 5 Node-Red

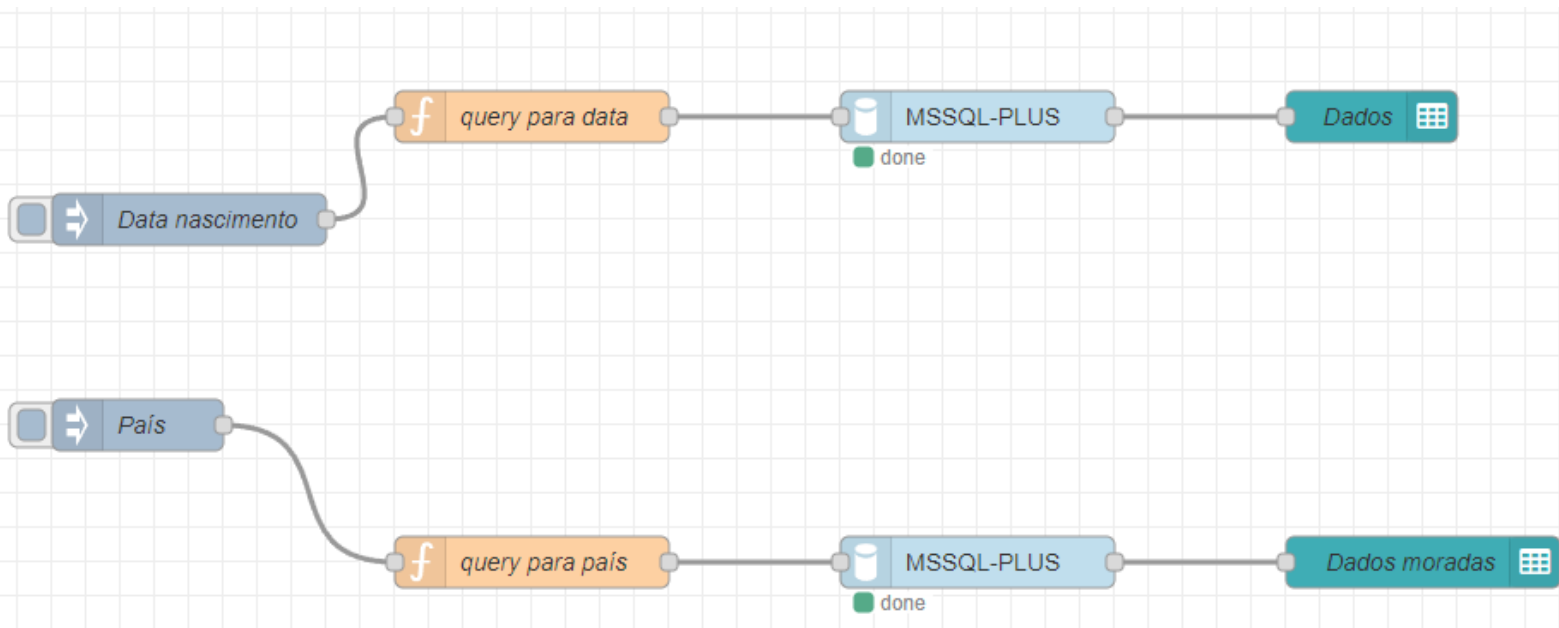


Figura 7 Node-Red

Na Figura 7 é representado o fluxo criado em Node-Red. Podemos Inserir alguma data de nascimento e verificar quais dos utilizadores nasceram antes ou depois dessa determinada data. No outro fluxo, podemos filtrar por países, visto que podemos ter vários utilizadores com o mesmo país. No final é apresentado para ambos os fluxos uma tabela com os dados encontrados.

## Dados utilizadores

id	nome	sobrenome	email	telemovel	data_nascimento
10	Ricardo	Cummings	sylvia.wiegand@yaho...	+17858279937	2007-10-30T00:00:00...
15	Raymundo	Bergnaum	oschamberger@farre...	+17165559143	2008-06-14T00:00:00...
16	Duncan	Koepp	morissette.jovany@w...	+14803317639	2006-05-11T00:00:00...
22	Mathias	Kertzmann	charlene90@gmail.co...	+16513871499	2005-01-05T00:00:00...
29	Zachary	Howe	sydney.reilly@mclau...	+17432562699	2007-08-15T00:00:00...
30	Vidal	McLaughlin	ibayer@gmail.com	+19514627566	2007-10-01T00:00:00...
34	Gunnar	Howe	zieme.harrison@seng...	+18145048907	2005-10-05T00:00:00...
38	Jake	Cartwright	robyn74@mclaughlin...	+12695563682	2005-09-15T00:00:00...

Figura 8 Node-Red - Dados utilizadores

Na Figura 8 temos a tabela gerada pelo Node-Red que representa todos os utilizadores que nasceram antes da data 2008-10-10.

### Dados moradas

id	↑	rua	↑	nome_rua	↑	numero...	↑	cidade	↑	cod_pos...	↑	pais	↑	cod_pais	↑	morada_...	↑
13		1404 Spence...		Abdullah Ports		66559		Simburgh		73911-9505		Austria		AT		1404 Spencer...	
20		1185 Altenw...		Rebekah Ra...		2880		West Maximi...		97092-0102		Austria		AT		1185 Altenwe...	

*Figura 9 Node-Red - Dados de moradas*

Na Figura 9 temos a tabela gerada pelo Node-Red que representa todos as moradas que tenham em comum o país, que no caso é Austria.

## 6 Vídeo



## 7 Conclusão

Em conclusão, este projeto evidenciou a relevância de uma solução ETL automatizada para a integração e gestão de dados pessoais, especialmente num cenário onde a diversidade de fontes e a qualidade dos dados são desafios recorrentes. Ao implementar um processo de ETL eficaz, torna-se possível extrair, transformar e carregar dados de forma automatizada, garantindo que as informações obtidas sejam consistentes, organizadas e normalizadas. Para concluir podemos dizer que um ETL bem implementado, é essencial para a eficiência e segurança na gestão de grandes volumes de dados.