

Trabalho 1

Integração de Sistemas de Informação

Aluno:
Francisco Moreira Rebêlo, a16443

Docente:
Óscar Ribeiro

Escola Superior de Tecnologia
Licenciatura em Engenharia de Sistemas Informáticos - PL

Barcelos, 26 de outubro de 2023

Integração de Sistemas de Informação
Trabalho 1

Índice

1. Contextualização	5
2. Criação de dados aleatórios	6
3. Criação das conversões no Kettle.....	7
3.1. Transformação 1 – CSV para XML	7
3.2. Transformação 2 – XML para Excel.....	9
3.3. Transformação 3 – Filtrar campos Excel.....	11
4. Criação do Job.....	13
5. Prova envio de email	15
6. Vídeo do funcionamento do programa	16
7. Conclusão.....	17

Índice de Imagens

Figura 1 - Mockaroo campos	6
Figura 2 - Transformação 1 – Geral	7
Figura 3 - Transformação 1 - Detalhe importação CSV	8
Figura 4 - Transformação 1 - Detalhe criação XML	8
Figura 5 - Transformação 2 – Geral	9
Figura 6 - Transformação 2 - Detalhe importação XML	9
Figura 7 - Transformação 2 - Detalhe filtro de linhas	10
Figura 8 - Transformação 2 - Detalhe criação Excel	10
Figura 9 - Transformação 3 – Geral	11
Figura 10 - Transformação 3 - Importação XLS	11
Figura 11 - Transformação 3 - Campos a remover	12
Figura 12 - Transformação 3 - Criação Excel editado	12
Figura 13 - Job - Vista Geral	14
Figura 14 - Sapo - Conta suspensa	15
Figura 15 - Formato email	15
Figura 16 - QR Code - Vídeo Funcionamento	16

1. Contextualização

O primeiro trabalho da Unidade Curricular de Integração de Sistemas de Informação prevê a utilização de ferramentas de transformação de dados onde é esperada que sejam utilizadas ferramentas utilizadas na aula ou outras de nosso conhecimento de modo a consolidar a matéria que foi abordada nas aulas. O programa que foi utilizado para fazer o trabalho foi o Pentaho Data Integration da Hitachi AKA Kettle AKA Spoon. Este programa dispõe de diversas ferramentas de manipulação de dados, tabelas, bases de dados e permite também a criação de pastas, comparação de ficheiros, envio de emails, pings entre outras ferramentas interessantes.

2. Criação de dados aleatórios

O trabalho baseia-se no trabalho que temos previsto para Projeto Aplicado onde no meu caso, trata-se de registos de temperaturas de quartos de hotel.

Para gerar dados aleatórios utilizei o website Mockaroo que permite gerar dados aleatórios tendo em conta o tipo de dados de cada campo. Ele também permite gerar os campos através de Inteligência Artificial, basta dar uma descrição do problema que temos, especificar a quantidade de campos e ele devolve os campos. Foi através desta abordagem que criei os campos.

Field Name	Type	Options
room_number	Row Number	blank: 0% Σ X
temperature	Number	min: 0 max: 100 decimals: 2 blank: 0% Σ X
humidity	Number	min: 0 max: 100 decimals: 2 blank: 0% Σ X
timestamp	Datetime	01/01/2022 to 12/31/2022 format: mm/dd/yyyy blank: 0% Σ X
occupancy_status	Custom List	occupied, vacant random blank: 0% Σ X
thermostat_setting	Number	min: 60 max: 80 decimals: 1 blank: 0% Σ X
fan_status	Boolean	blank: 0% Σ X
heating_status	Boolean	blank: 0% Σ X
cooling_status	Boolean	blank: 0% Σ X
window_status	Boolean	blank: 0% Σ X
door_status	Boolean	blank: 0% Σ X
light_status	Boolean	blank: 0% Σ X
noise_level	Number	min: 0 max: 10 decimals: 1 blank: 0% Σ X
power_consumption	Number	min: 0 max: 1000 decimals: 2 blank: 0% Σ X
energy_saving_mode	Boolean	blank: 0% Σ X
occupant_count	Number	min: 0 max: 10 decimals: 0 blank: 0% Σ X
occupant_age	Number	min: 0 max: 100 decimals: 0 blank: 0% Σ X
occupant_gender	Custom List	male, female, other random blank: 0% Σ X
occupant_nationality	Country	restrict countries... blank: 0% Σ X
occupant_language	Language	blank: 0% Σ X
occupant_mood	Custom List	happy, sad, neutral random blank: 0% Σ X

Figura 1 - Mockaroo campos

3. Criação das conversões no Kettle

No Kettle, temos dois tipos de ficheiros. Temos o ficheiro .ktr que pertence a uma Transformação e o ficheiro .kjb que corresponde a um Job. Podemos fazer uma analogia do Job e a Transformação ao Main e a uma Função de um programa respetivamente. Da mesma forma que no Main de um programa chamamos cada função, no Kettle no Job chamamos cada Transformação que precisemos de usar.

Em primeiro vamos falar sobre cada uma das três transformações que fiz.

3.1. Transformação 1 – CSV para XML

Esta primeira transformação converte o ficheiro CSV que foi criado no Mockaroo e converte-o em XML. Houve uma dificuldade neste passo relativamente aos campos cuja variável era um inteiro e a causa era que o Kettle colocava um limite máximo de 15 caracteres para o inteiro quando o máximo que um inteiro suporta é 10 e isso gerava problemas mais à frente. Para ultrapassar isto, removi o limite deixando assim o campo do limite vazio. Em cada local e isto serve de regra para todo o trabalho onde era necessário colocar caminhos para ficheiros, foram sempre utilizados caminhos relativos de modo a garantir o funcionamento do Job em diferentes computadores. Para cada importação ou criação de ficheiros é sempre necessário clicar onde diz *Get Fields* para obtermos os campos que vamos importar ou criar de ou num ficheiro.



Figura 2 - Transformação 1 – Geral

Integração de Sistemas de Informação

Trabalho 1

CSV file input

Step name: CSV file input

Filename: \$(Internal.Entry.Current.Directory)\MOCK_DATA.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding: UTF-8

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	room_number	Integer	#		0	€	.	.	none
2	temperature	Number	#,.	15	0	€	.	.	none
3	humidity	Number	#,.	15	0	€	.	.	none
4	timestamp	Date	MM/dd/yyyy			€	.	.	none
5	sensor_id	Integer	#		0	€	.	.	none
6	occupancy_status	String		11		€	.	.	none
7	power_status	Boolean				€	.	.	none
8	fan_speed	Integer	#		0	€	.	.	none
9	window_open	Boolean				€	.	.	none
10	heating_status	Boolean				€	.	.	none
11	cooling_status	Boolean				€	.	.	none
12	thermostat_mode	String		4		€	.	.	none
13	set_temperature	Number	#,.	15	0	€	.	.	none
14	energy_consumption	Number	#,.	15	0	€	.	.	none
15	occupant_name	String		18		€	.	.	none
16	occupant_age	Integer	#		0	€	.	.	none
17	occupant_gender	String		11		€	.	.	none
18	occupant_nationality	String		21		€	.	.	none
19	occupant_email	String		27		€	.	.	none
20	occupant_phone	String		12		€	.	.	none
21	occupant_address	String		28		€	.	.	none

Help OK Obtem campos Preview Cancela

Figura 3 - Transformação 1 - Detalhe importação CSV

XML output

Nome do Step: XML output

File Content Fields

Filename: \$(Internal.Entry.Current.Directory)\Dados Gerados\Generated_XML

Do not create file at start ☐

Pass output to servlet ☐

Extension: xml

Include stepnr in filename? ☐

Include date in filename? ☐

Include time in filename? ☐

Specify Date time format ☐

Date time format:

Show filename(s)...

Add filenames to result ☐

Help OK Cancela

Figura 4 - Transformação 1 - Detalhe criação XML

3.2. Transformação 2 – XML para Excel

Nesta segunda transformação, estamos a importar o ficheiro que foi criado na transformação anterior e filtramos as linhas onde o valor do campo *Energy Consumption* é superior a 500, ou seja, no ficheiro a ser criado de seguida restam apenas os valores superiores a 500. No final desta transformação é criado um ficheiro em formato .xls que é o tipo de ficheiro utilizado pelo Excel.

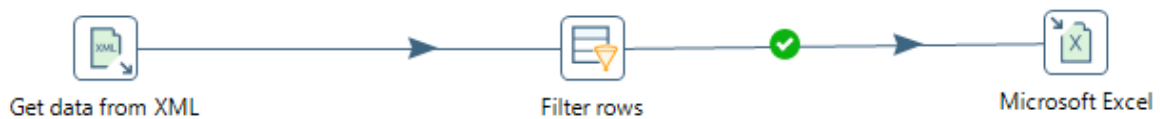


Figura 5 - Transformação 2 – Geral

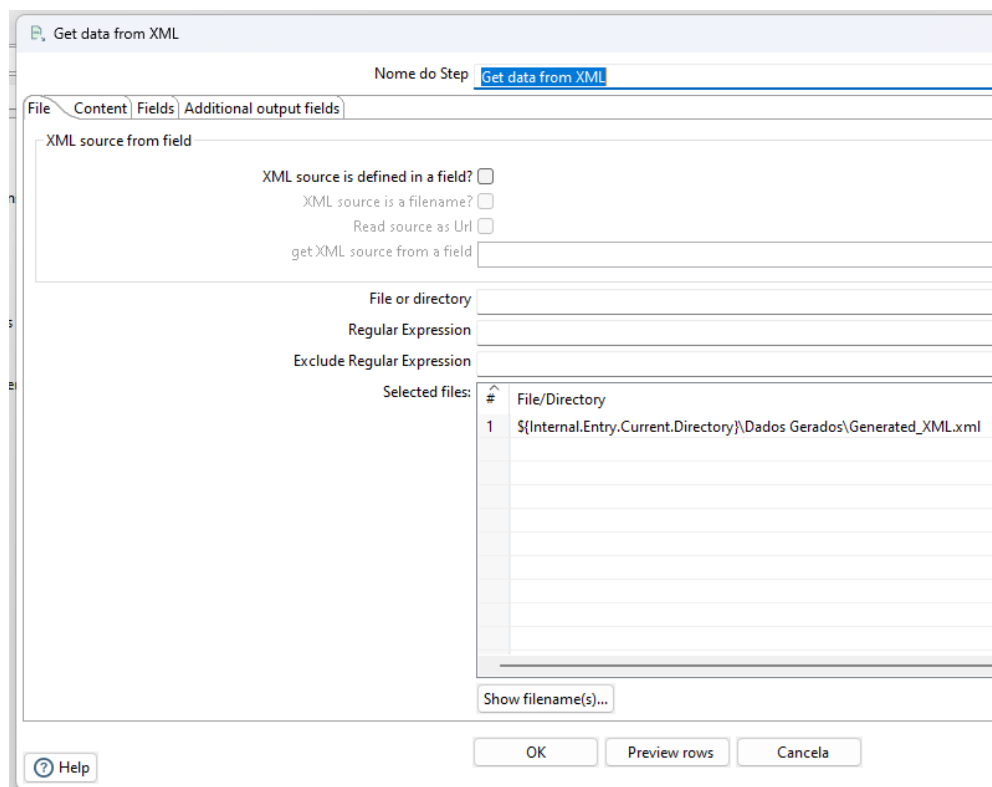


Figura 6 - Transformação 2 - Detalhe importação XML

Integração de Sistemas de Informação Trabalho 1

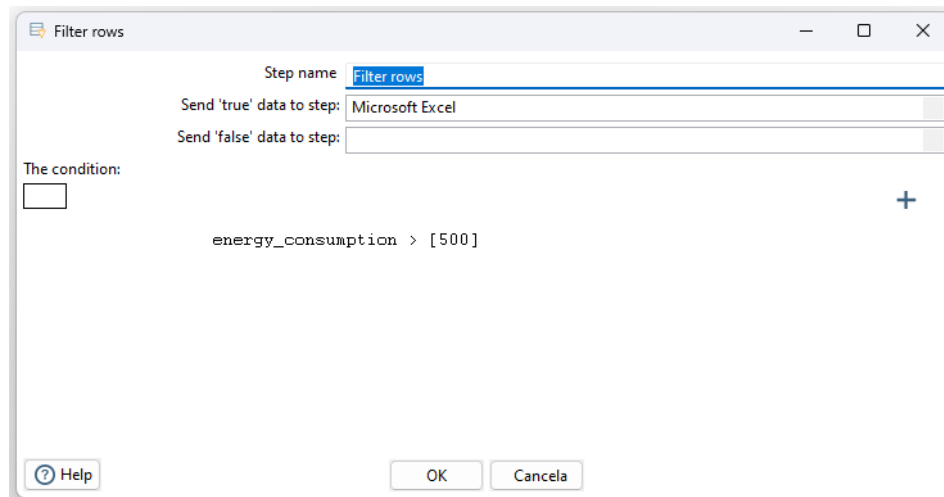


Figura 7 - Transformação 2 - Detalhe filtro de linhas

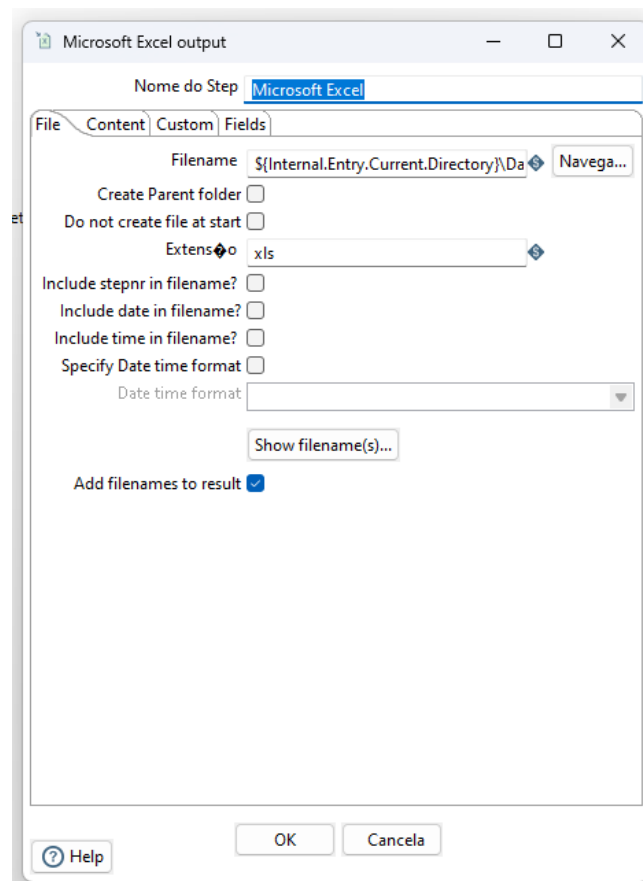


Figura 8 - Transformação 2 - Detalhe criação Excel

3.3. Transformação 3 – Filtrar campos Excel

Esta é a última transformação deste trabalho. Nesta transformação são apagadas os campos ou colunas que são irrelevantes ao utilizador final do ficheiro Excel.

Neste caso sobraram apenas os campos Room Number, Energy Consumption e Occupant Name.

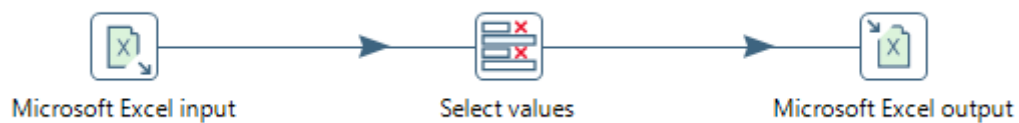


Figura 9 - Transformação 3 – Geral

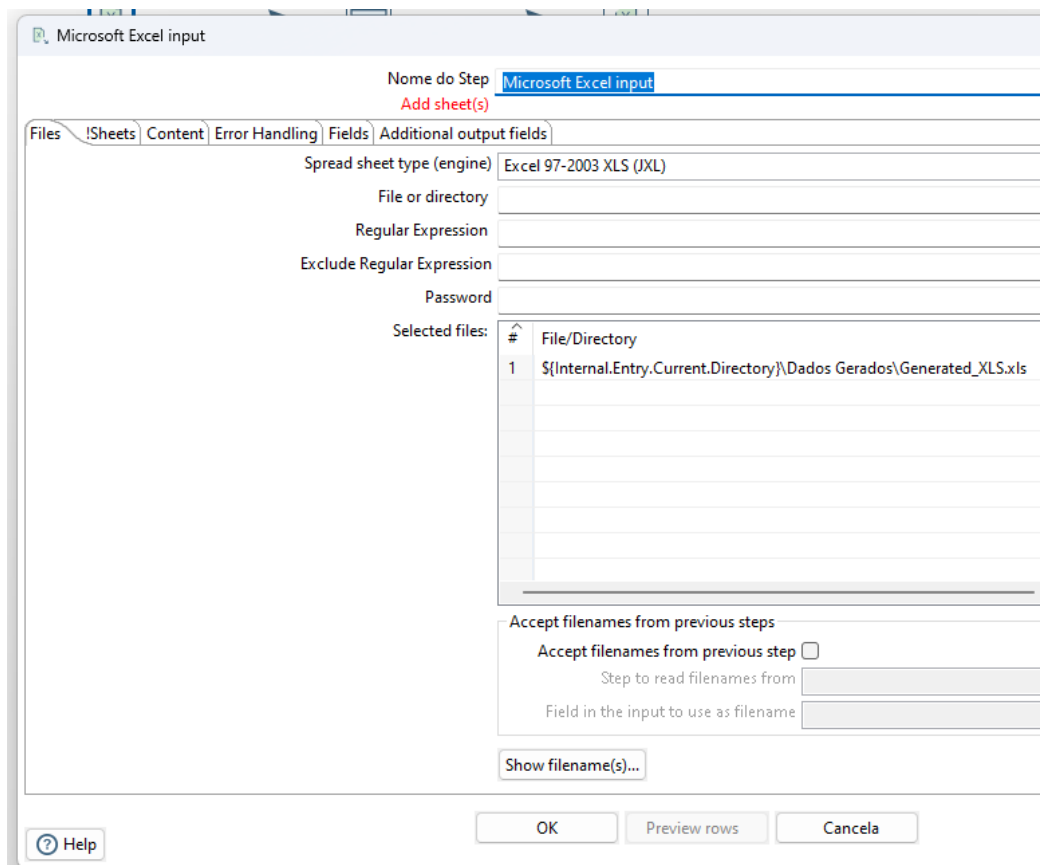


Figura 10 - Transformação 3 - Importação XLS

Integração de Sistemas de Informação Trabalho 1

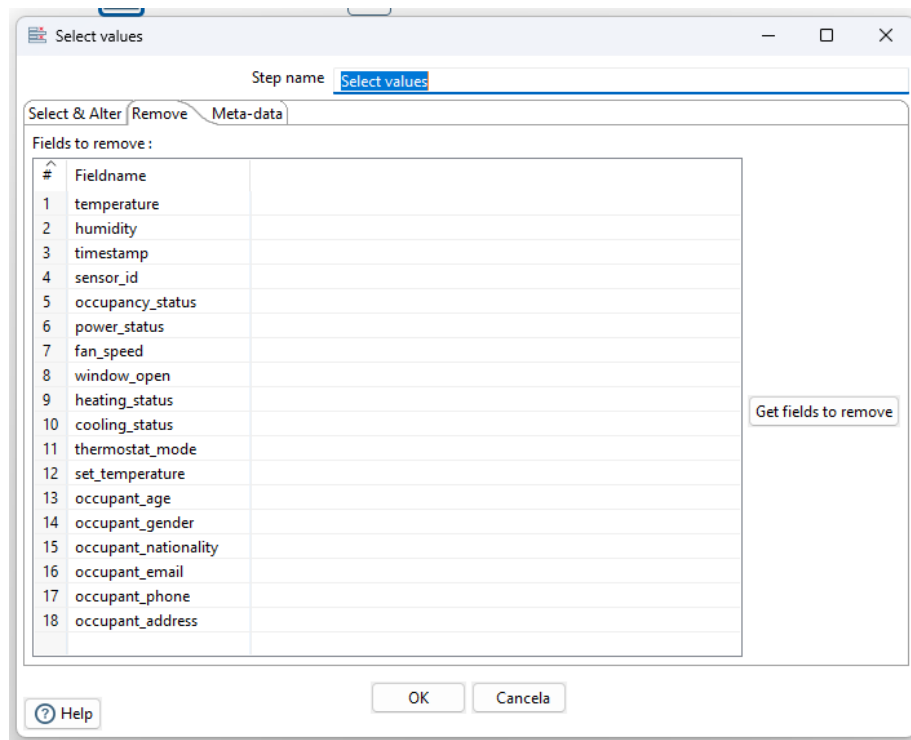


Figura 11 - Transformação 3 - Campos a remover

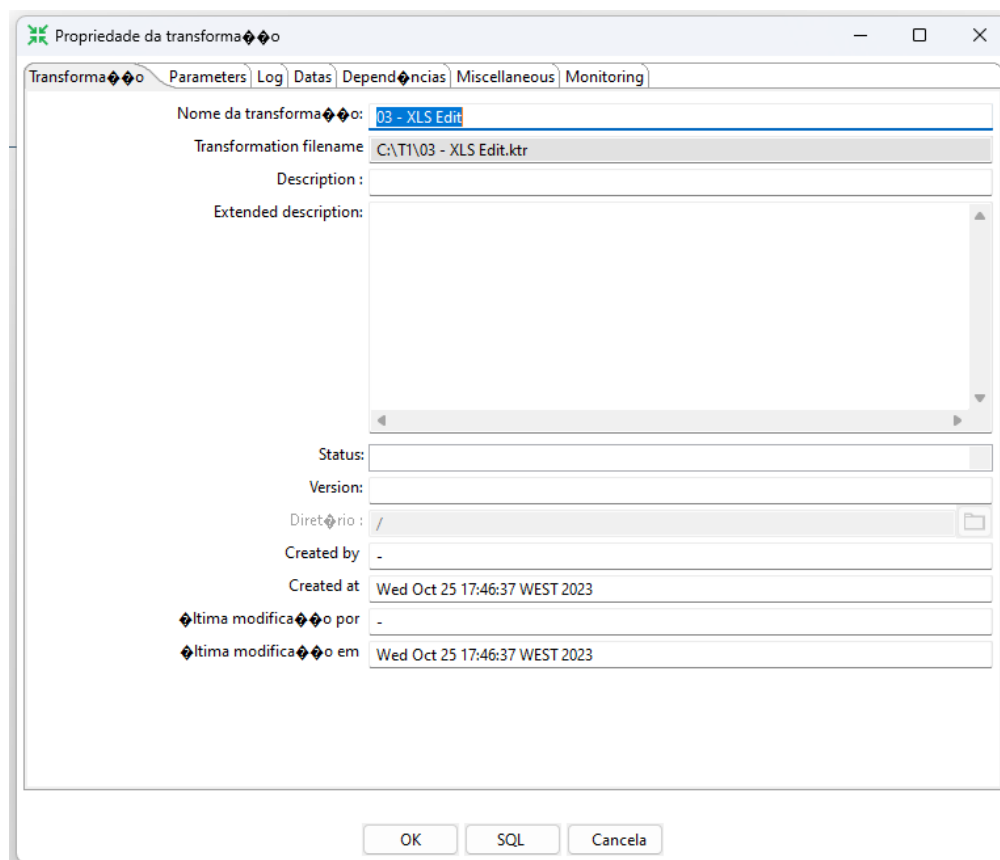


Figura 12 - Transformação 3 - Criação Excel editado

4. Criação do Job

No Job iremos chamar todas as transformações referidas anteriormente e utilizaremos também algumas ferramentas do Kettle.

Para iniciar o Job, basta apenas abrir com o Kettle o ficheiro START_HERE.kjb.

O Job é executado segundo a seguinte ordem:

1. Start
2. Cria pasta para armazenar ficheiros a serem criados
 - a. Caso não seja possível por alguma razão a criação da pasta, o programa mostra um erro e o Job é terminado sem sucesso.
3. Verifica se o CSV existe, se sim avança para a próxima etapa.
 - a. Caso o CSV não exista, é mostrada uma mensagem de erro e o Job é terminado.
4. É executada a transformação 1
5. Verifica se o XML existe, se sim avança para a próxima etapa.
 - a. Caso o XML não exista, é mostrada uma mensagem de erro e o Job é terminado.
6. É executada a transformação 2
7. Verifica se o XLS existe, se sim avança para a próxima etapa.
 - a. Caso o XLS não exista, é mostrada uma mensagem de erro e o Job é terminado.
8. É executada a transformação 3
9. Verifica se o XLS existe, se sim avança para a próxima etapa.
 - a. Caso o XLS não exista, é mostrada uma mensagem de erro e o Job é terminado.
10. É feito um ping ao Google para verificar se existe ligação à internet, se há sucesso, avança para a próxima etapa.
 - a. Caso não haja ligação à internet é mostrada uma mensagem de erro, o último ficheiro criado é copiado para a raiz do projeto para ficar guardado localmente, a pasta que foi criada ao início é apagada e o programa termina com sucesso
11. É enviado um email para a16443@alunos.ipca.pt através do endereço a16443@sapo.pt com o ficheiro criado anteriormente em anexo, se houver sucesso, avança para a etapa seguinte. Foi utilizado um email da Sapo pois a

Integração de Sistemas de Informação Trabalho 1

segurança da Sapo é menos apertada e surgem menos problemas ao configurar e enviar o email.

- a. Caso haja algum problema ao enviar o email é mostrada uma mensagem de erro, o último ficheiro criado é copiado para a raiz do projeto para ficar guardado localmente, a pasta que foi criada ao início é apagada e o programa termina com sucesso.

12. A pasta criada ao início é apagada

13. Job termina com sucesso!

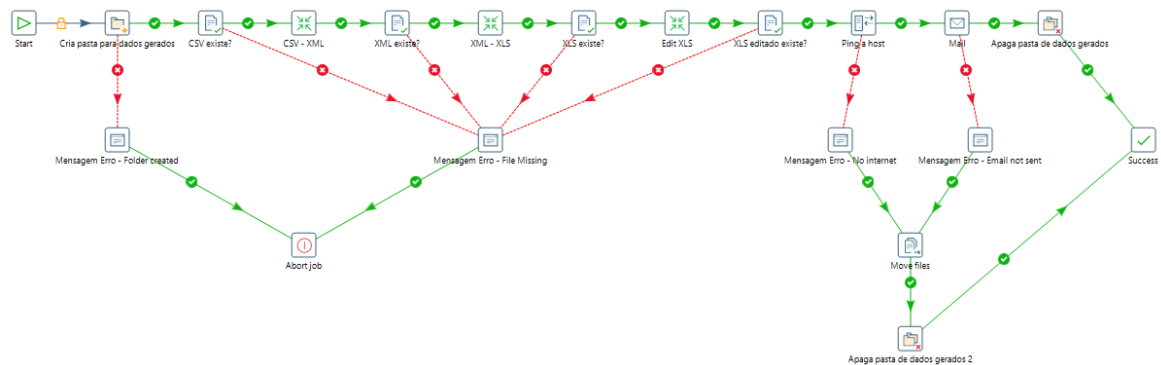


Figura 13 - Job - Vista Geral

5. Prova envio de email

Aquando da realização do vídeo, o email não era enviado. Fui averiguar a situação e apercebi-me de que a conta de email tinha sido suspensa.

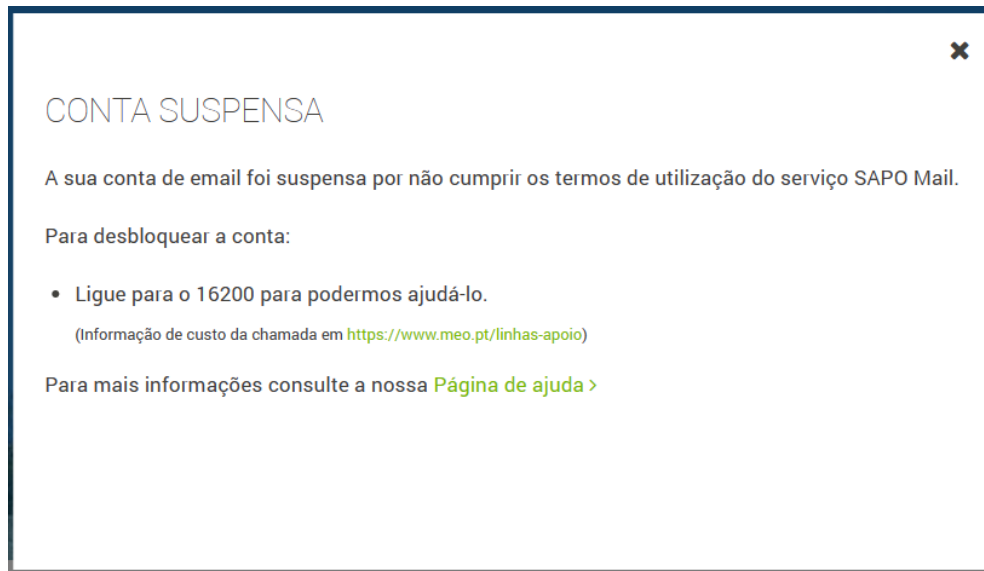


Figura 14 - Sapo - Conta suspensa

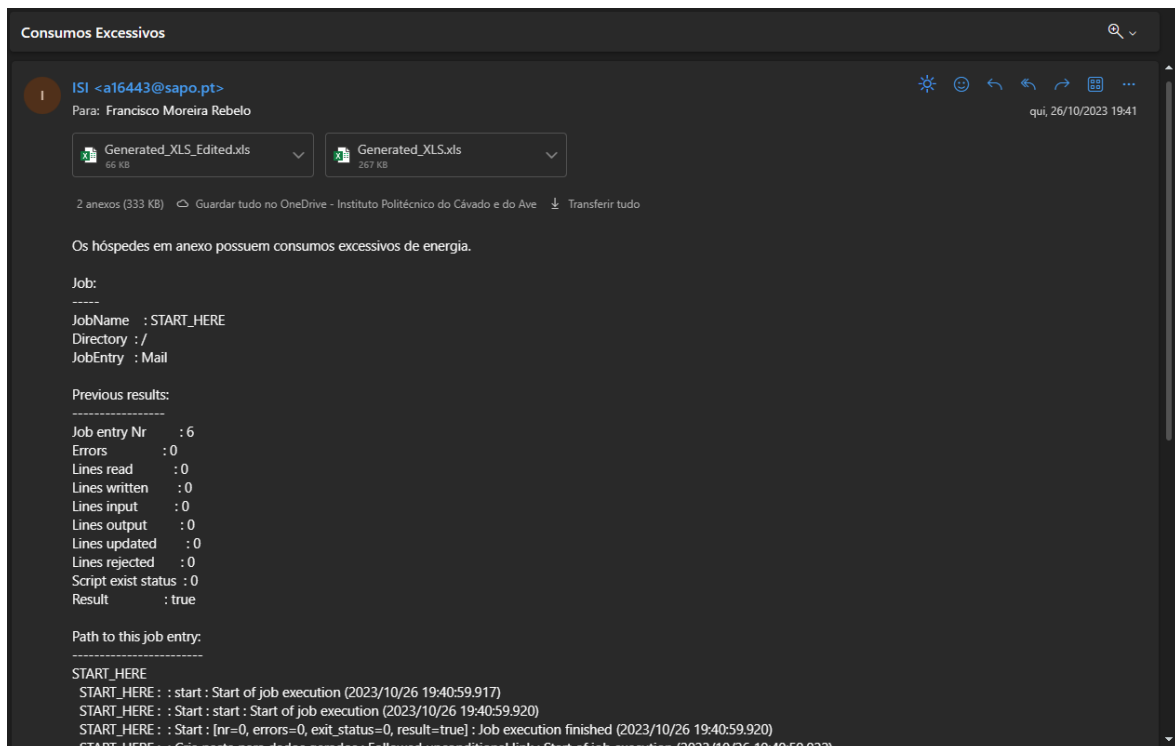


Figura 15 - Formato email

6. Vídeo do funcionamento do programa



Figura 16 - QR Code - Vídeo Funcionamento

<https://youtu.be/8vtfzru8vQs>

7. Conclusão

Com a realização deste trabalho, posso concluir que a disciplina de Integração de Sistemas de Informação oferece aos alunos a oportunidade de aplicar os conhecimentos adquiridos em aula através da utilização de ferramentas de transformação de dados. A escolha do programa Pentaho Data Integration, revela-se acertada devido à sua ampla gama de funcionalidades para manipulação de dados e integração de diversas fontes de informação. Esta experiência prática contribuiu significativamente para a consolidação da matéria dada em aula preparando-nos para enfrentar desafios reais no campo da integração de sistemas de informação.