



INSTITUTO POLITÉCNICO
DO CÁVADO E DO AVE
ESCOLA SUPERIOR
DE TECNOLOGIA

Trabalho Prático 01 - ISI

LESI - PL

Integração de Sistemas de
Informação

Diogo Gomes Silva - 23893

Barcelos, Braga 27/10/2023

Índice:

1. Introdução:	2
2. Motivação:	2
3. Objetivos:.....	3
4. Temas:	3
Resultados Formula 1 - 2012:.....	3
Dados “WorkersData”:.....	3
5. Arquivos Iniciais Utilizados:	4
Resultados Formula 1 - 2012:.....	4
Dados “WorkersData”:.....	4
6. Solução Desenvolvida:	5
Resultados Formula 1 - 2012:.....	5
Primeira Transformação “F1_Corridas_2012”:	5
Job “Executa_F1_2012”:	9
Transformação “Get_HTML”:	10
Resultados dos Emails obtidos:	12
Dados WorkersData:.....	13
Primeira Transformação “Obter_WorkersData”:.....	13
Job “WorkersData_Run”:	16
Resultados dos Emails obtidos:	17
7. Conclusão:	18
8. Bibliografia/Web-grafia:.....	18

Figura 1 - Transformação "F1_Corridas_2012"	5
Figura 2 - Job "Executa_F1_2012"	9
Figura 3 - Transformação "Get_HTML"	10
Figura 4 - Transformação "Obter_WorkersData"	13
Figura 5 - Job "WorkersData_Run"	16

1. Introdução:

No âmbito da Unidade Curricular de Integração de Sistemas de Informação (ISI) foi proposto 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. Foi então, efetuado o desenvolvimento de processos de ETL recorrendo a uma ferramenta chamada Kettle Pentaho.

2. Motivação:

Uma vez que os processos de negócio não param de se reconfigurar, as empresas enfrentam desafios constantes de análise e aquisição de novas soluções informáticas. A necessidade de rentabilizarem anteriores aquisições, tanto pelo investimento financeiro envolvido como pela dependência dos processos, dos dados, etc., que delas fazem parte, as empresas procuram formas capazes de analisarem mais-valias e menor impacto com as novas aquisições. Processos como: i) auditorias a dados, processos, segurança, outros; ii) migração e reorganização de dados; iii) análise e processamento de dados (datamining, etc.); iv) recomendações e previsões sobre estados com processamento recorrente a big data; são exemplo de cenários onde processos de ETL poderão desempenhar papel preponderante. No contexto emergente dos smart environments (Cities, Health Care, Social, etc.) a evidência deste tipo de soluções é clara e crescente. A integração de soluções ditas mais inteligentes, que reagem a eventos, em processos existentes ou mesmo legados, são desafios constantes. Pretende-se que o aluno defina o objetivo a explorar no trabalho que poderá ser real ou imaginário.

3. Objetivos:

Os objetivos do desenvolvimento deste trabalho são os seguintes:

- Consolidar conceitos associados à Integração de Sistemas de Informação usando Dados;
- Analisar e especificar cenários de aplicação de processos de ETL;
- Explorar ferramentas de suporte a processos de ETL;
- Explorar novas Tecnologias, Frameworks ou Paradigmas;
- Potenciar a experiência no desenvolvimento de software;
- Facilitar a assimilação do conteúdo da Unidade Curricular

4. Temas:

O presente trabalho foi dividido em dois temas:

RESULTADOS FORMULA 1 - 2012:






Esta solução desenvolvida foi gerada através de 5 tabelas em arquivo CSV com vários campos com Resultados de Corridas de Formula 1 de 1950 até 2023, onde apenas foram selecionados os resultados do ano 2012. Esta solução teve como objetivo a manipulação de dados e criação de Joins de tabelas, manipulação de arquivos XML e XSL para gerar um HTML com os dados da tabela final e enviar a mesma por E-mail.

DADOS “WORKERSDATA”:


Esta solução desenvolvida foi gerada através de 1 tabela em arquivo CSV com vários campos de informação fictícia de trabalhadores de várias empresas. Esta solução teve como objetivo a manipulação de dados da tabela através de Expressões Regulares (REGEX) e transformações de arquivos XLS, criar arquivos ZIP e enviar os mesmos por E-mail.

5. Arquivos Iniciais Utilizados:

RESULTADOS FORMULA 1 - 2012:

TP1ISI_23893 > F1_2012 > DadosInput				Procurar em
Nome	Data de modificação	Tipo	Tamanho	
 circuits.csv	25/10/2023 22:06	Ficheiro de Valore...	10 KB	
 constructors.csv	25/10/2023 22:06	Ficheiro de Valore...	17 KB	
 drivers.csv	25/10/2023 22:06	Ficheiro de Valore...	92 KB	
 races.csv	25/10/2023 22:06	Ficheiro de Valore...	156 KB	
 results.csv	25/10/2023 22:06	Ficheiro de Valore...	1 629 KB	

DADOS “WORKERSDATA”:

TP1ISI_23893 > WorkersData > InputData				Procurar e
Nome	Data de modificação	Tipo	Tamanho	
 WorkersData.csv	29/10/2023 13:30	Ficheiro de Valore...	22 KB	

A1	id,first_name,last_name,email,phone_number,job,company,department,city,postal_code,salary,car													
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id,first_name,last_name,email,phone_number,job,company,department,city,postal_code,salary,car													
2	1,Pip,Tyzack,ptyzack0@reverbnation.com,(451) 1970713,,Feedspan,Research and Development,Pole	687 37,"	~35645,48",Honda											
3	2,Carmon,Ellsworth,cellsworth1w3.org,(965) 1300204,Web Developer II,Devshare,Human Resources,Bicaj,,,"	~55313,82",Pontiac												
4	3,Fredrick,Noquet,fnoquet2@plala.o,(149) 1299616,Technical Writer,Oyonder,Accounting,Ljutomer,9240,"	~3368,18",Chevrolet												
5	4,Wilow,Rucklidge,,(335) 7693697,,Zava,,Okpoma,,,"	~40915,26",Porsche												
6	5,Jacinta,Hartnup,jhartnup4@phet,(780) 3087950,,Zoonder,,Makroch	~35426,32",Honda												
7	6,Henderson,Fidoe,hfidoe5tagramom,(430) 9743202,Help Desk Technician,Eidel,Marketing,La Puerta de San Jos	~21791,72",Ford												
8	7,Aluin,Kubik,akubik6@businessweek.com,,Brightbean,,Makale,,,"	~39917,24",Lotus												
9	8,Sheba,Roswarn,sroswarn7@sun.com,(133) 7739996,Developer II,Zoomzone,,Haozigang,,,"	~31237,18",Suzuki												
10	9,Meade,Guppie,(944) 1556689,,Skiba,,Cirateun,,,"	~56516,50",Chevrolet												
11	10,Tracy,Stidever,tstidever9dtoany.com,,Chief Design Engineer,Youspan,Services,	~74283,48",Audi												
12	11,Olia,Grouse,,(787) 2800419,Actuary,Eabox,,Lengkongbarang,,,"	~12072,32",Hyundai												
13	12,Itch,,ihulburdb@bloglines.com,(221) 4667087,,Photofeed,Research and Development,Longhua,,,"	~20474,47",Ford												
14	13,Romain,Braxton,,Pixoboo,Product Management,Samashki,366602,"	~2077,89",Nissan												
15	14,Inger,Arbuckel,iarbuckeld@privacy.gov.au,(678) 6973331,Systems Administrator IV,Rooxo,,Bluri,,,"	~35449,47",Ford												
16	15,Tally,,(429) 2139701,Desktop Support Technician,Edgeclub,Human Resources,Tanabe,610-0331,"	~78165,66",Chevrolet												
17	16,Buddy,Gotcliff,,(894) 5859133,Occupational Therapist,Browsezoom,Accounting,Ipor	~62120,10",Lincoln												
18	17,Darwin,Stilling,,Engineer III,Trudeo,Services,Santa Fe,6513,"	~77947,61",Jaguar												
19	18,Nobie,,njordih@microsoft.com,(142) 3970440,Research Nurse,Kazu,Sales,Budapest,1062,"	~83294,36",Chevrolet												
20	19,Roma,,rilyenkoi@techcrunch.com,(170) 2994197,Operator,DabZ,Business Development,Jaba	~63174,87",Ford												
21	20,Susannah,Titherington,stitheringtonj@intel.com,(220) 7882826,Photojam,,Sollefte	~10971,24",Ford												
22	21,Phillip,Bickerdyke,pbickerdykek@utexas.edu,(212) 6797362,Editor,Dabvine,Research and Development,Koungou,97600,"	~72478,75",Toyota												
23	22,Leisha,,Director of Sales,Kwimbee,Training,Ljusdal,827 24,"	~88444,53",Volkswagen												
24	23,Isis,,ilampreym@mtv.com,(130) 9029146,Project Manager,Agivu,,Kobleve,,,"	~93374,40",Hyundai												
25	24,Ninon,Fransoni,nfransonin@loc.gov,(966) 4410068,Analog Circuit Design manager,Devshare,,Jhumra,50355,"	~21765,23",Ford												
26	25,Honoria,Claesens,hclaesenso@cocolog-nifty.com,(721) 4463815,Assistant Media Planner,Tagpad,,R	~11490,74",Dodge												

6. Solução Desenvolvida:

RESULTADOS FORMULA 1 - 2012:

Primeira Transformação “F1_Corridas_2012”:

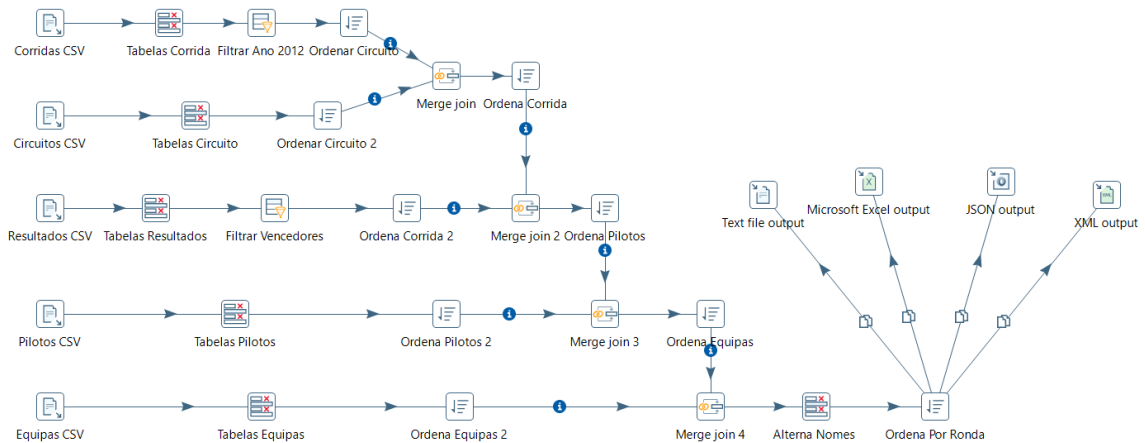
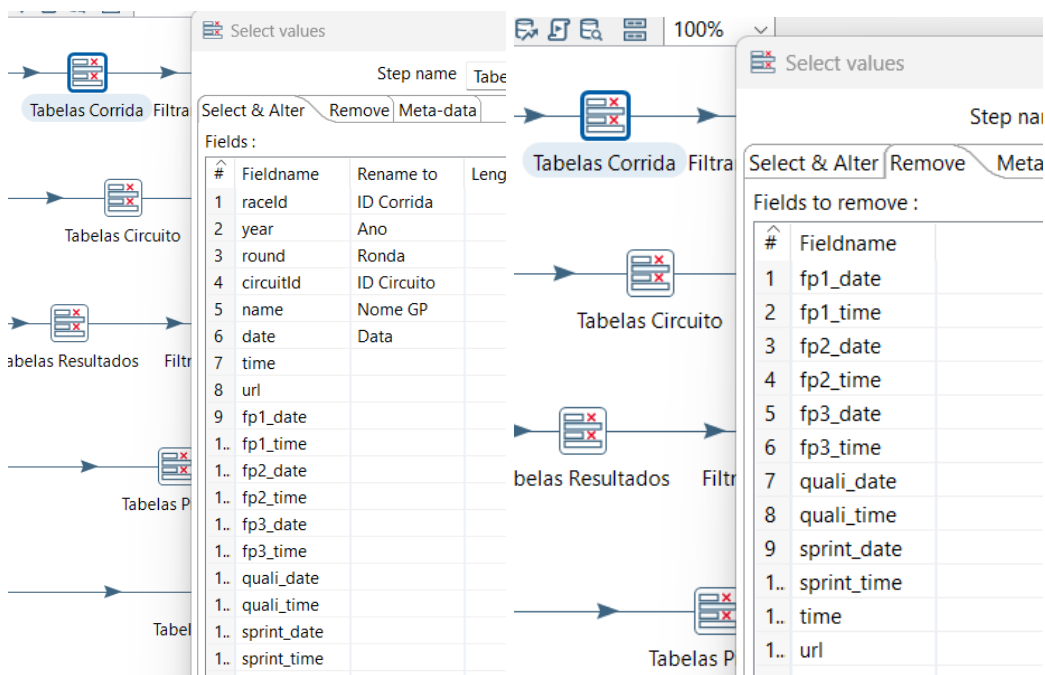
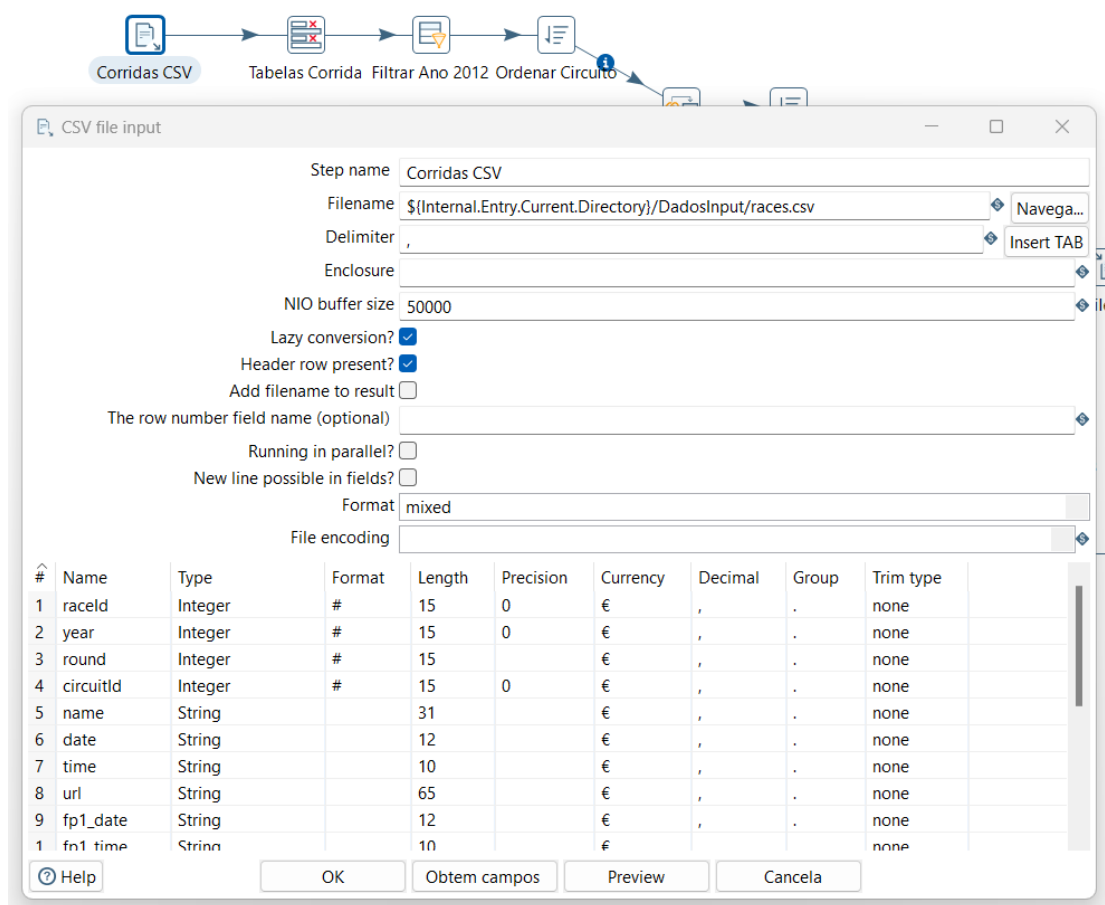
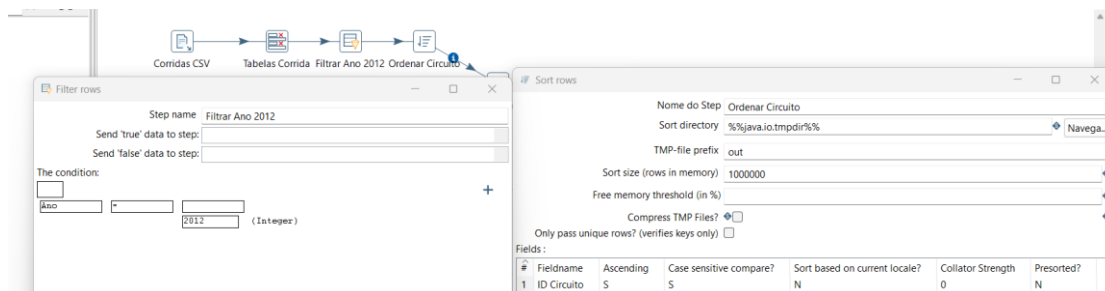


Figura 1 - Transformação "F1_Corridas_2012"

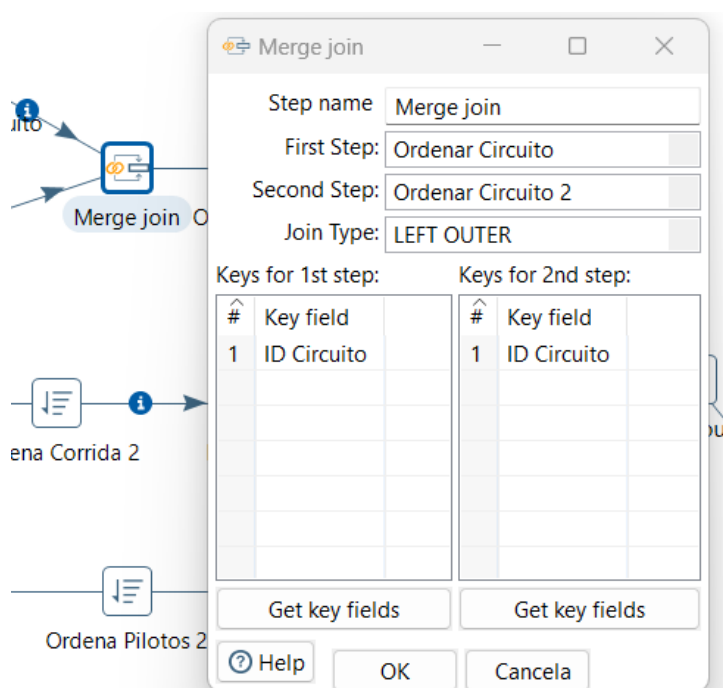
Nesta Primeira Transformação, começamos por obter os dados das tabelas dos arquivos CSV através da ferramenta “Csv File Input”, colocando o caminho para o ficheiro no campo “Filename” e obtendo os campos em baixo, como o seguinte exemplo:



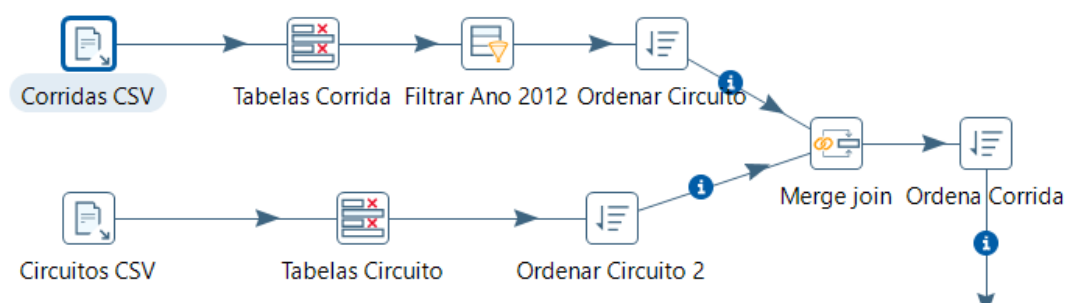
Em seguida, utilizamos a ferramenta “Select Values” para apagar os campos indesejados ou modificar os campos que vamos utilizar.




As ferramentas “Filter Rows” e “Sort Rows” foram utilizadas respetivamente para filtrar e ordenar os campos utilizados, no exemplo Filtramos o ano para termos apenas corridas do ano de 2012 e de seguida Ordenamos pelo ID Circuito para fazermos o Join de ambas as tabelas:



Com ambas as tabelas ordenadas pelo ID de Circuito podemos então fazer o Join das tabelas, utilizando o “LEFT OUTER JOIN”.



Assim obtemos o seguinte resultado:



Execution Results

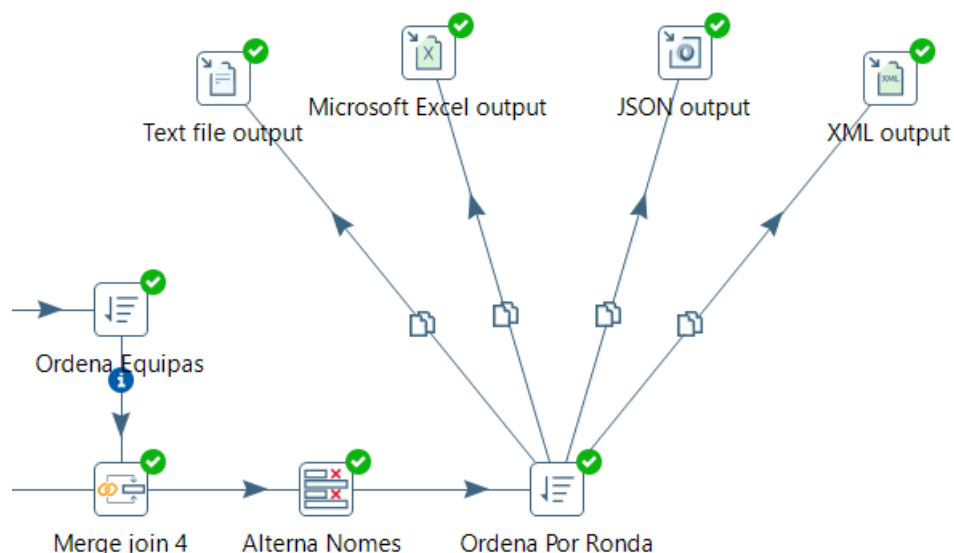
Logging Execution History Step Metrics Performance Graph Metrics Preview data

\$(TransPreview.FirstRows.Label) \$(TransPreview.LastRows.Label) \$(TransPreview.Off.Label)

#	ID Corrida	Ano	Ronda	ID Circuito	Nome GP	Data	ID Circuito_1	Nome	Localização	País
1	860	2012	1	1	"Australian Grand Prix"	"2012-03-18"	1	"Albert Park Grand Prix Circuit"	"Melbourne"	"Australia"
2	861	2012	2	2	"Malaysian Grand Prix"	"2012-03-25"	2	"Sepang International Circuit"	"Kuala Lumpur"	"Malaysia"
3	863	2012	4	3	"Bahrain Grand Prix"	"2012-04-22"	3	"Bahrain International Circuit"	"Sakhir"	"Bahrain"
4	864	2012	5	4	"Spanish Grand Prix"	"2012-05-13"	4	"Circuit de Barcelona-Catalunya"	"Montmeló"	"Spain"
5	865	2012	6	6	"Monaco Grand Prix"	"2012-05-27"	6	"Circuit de Monaco"	"Monte-Carlo"	"Monaco"
6	866	2012	7	7	"Canadian Grand Prix"	"2012-06-10"	7	"Circuit Gilles Villeneuve"	"Montreal"	"Canada"
7	868	2012	9	9	"British Grand Prix"	"2012-07-08"	9	"Silverstone Circuit"	"Silverstone"	"UK"
8	869	2012	10	10	"German Grand Prix"	"2012-07-22"	10	"Hockenheimring"	"Hockenheim"	"Germany"
9	870	2012	11	11	"Hungarian Grand Prix"	"2012-07-29"	11	"Hungaroring"	"Budapest"	"Hungary"
1..	867	2012	8	12	"European Grand Prix"	"2012-06-24"	12	"Valencia Street Circuit"	"Valencia"	"Spain"
1..	871	2012	12	13	"Belgian Grand Prix"	"2012-09-02"	13	"Circuit de Spa-Francorchamps"	"Spa"	"Belgium"
1..	872	2012	13	14	"Italian Grand Prix"	"2012-09-09"	14	"Autodromo Nazionale di Monza"	"Monza"	"Italy"
1..	873	2012	14	15	"Singapore Grand Prix"	"2012-09-23"	15	"Marina Bay Street Circuit"	"Marina Bay"	"Singapore"
1..	862	2012	3	17	"Chinese Grand Prix"	"2012-04-15"	17	"Shanghai International Circuit"	"Shanghai"	"China"
1..	879	2012	20	18	"Brazilian Grand Prix"	"2012-11-25"	18	"Autódromo José Carlos Pace"	"São Paulo"	"Brazil"
1..	874	2012	15	22	"Japanese Grand Prix"	"2012-10-07"	22	"Suzuka Circuit"	"Suzuka"	"Japan"
1..	877	2012	18	24	"Abu Dhabi Grand Prix"	"2012-11-04"	24	"Yas Marina Circuit"	"Abu Dhabi"	"UAE"
1..	875	2012	16	35	"Korean Grand Prix"	"2012-10-14"	35	"Korean International Circuit"	"Yeongam County"	"Korea"
1..	876	2012	17	68	"Indian Grand Prix"	"2012-10-28"	68	"Buddh International Circuit"	"Uttar Pradesh"	"India"
2..	878	2012	19	69	"United States Grand Prix"	"2012-11-18"	69	"Circuit of the Americas"	"Austin"	"USA"

Como podemos observar a tabela contém agora dois ID Circuito (ID_Circuito e ID_Circuito_1) e assim conseguimos juntar o nome, localização e país do circuito às Corridas.

Assim repetimos este processo para os outros arquivos CSV de forma a juntar todas as tabelas.



No fim após o último Join, fazemos as ultimas alterações de Nomes das tabelas e apagamos as tabelas com ID's desnecessários, ordenamos pela Ronda da Corrida e exportamos para ficheiros: txt, xls, json e xml.

Job “Executa_F1_2012”:

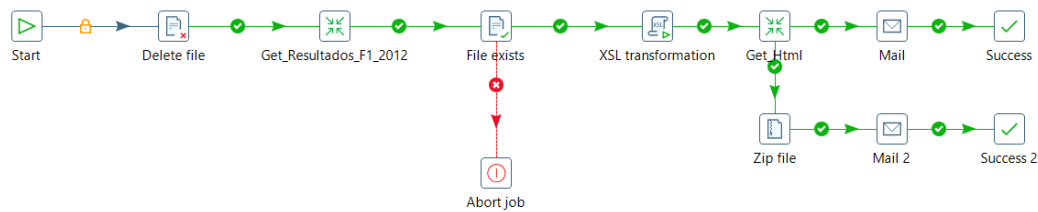
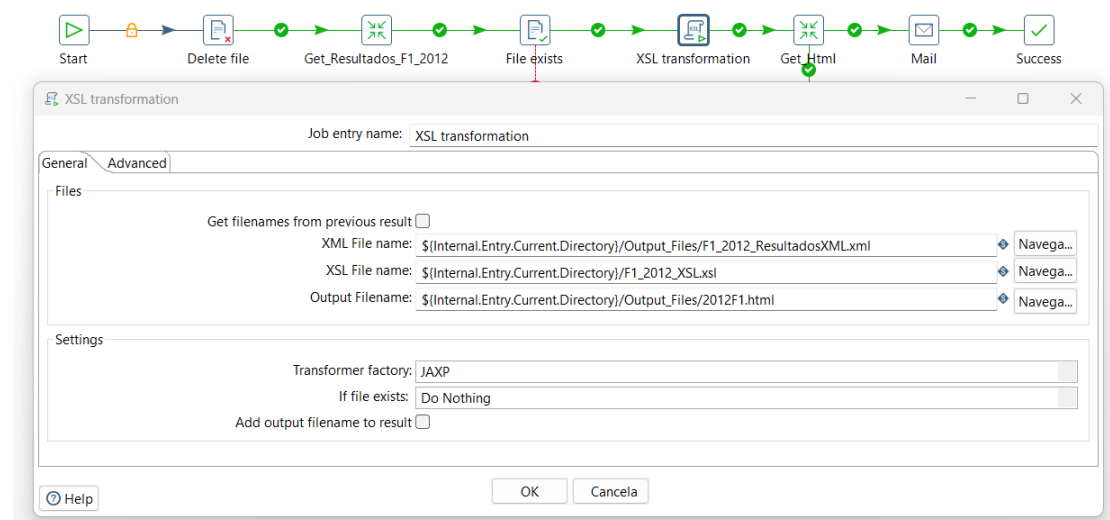


Figura 2 - Job "Executa_F1_2012"

Neste job, começamos por apagar, caso já exista, o ficheiro html que iremos gerar mais à frente, de seguida executamos a transformação “F1_Corridas_2012” para obter todos os ficheiros da tabela. Verificamos se o Ficheiro XML foi gerado, caso não tenha sido gerado abortamos o job, caso tenha sido gerado corretamente utilizamos o mesmo, bem como um arquivo XSL criado para obter um arquivo html através da ferramenta “XSL transformation”:



Arquivo XSL criado:

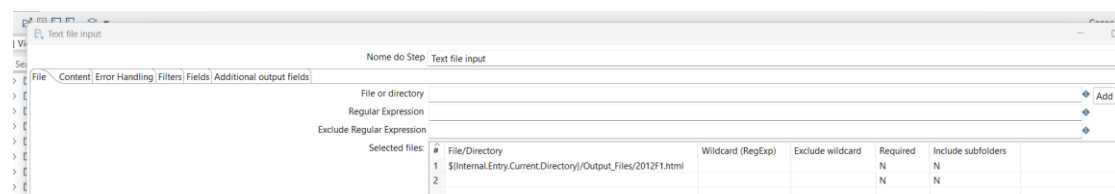
DadosInput	28/10/2023 16:54	Pasta de ficheiros	
Output_Files	29/10/2023 01:03	Pasta de ficheiros	
Executa_F1_2012.kjb	29/10/2023 01:09	Ficheiro KJB	19 KB
F1_2012_XSL.xsl	28/10/2023 15:24	XSLT Stylesheet	3 KB
F1_Corridas_2012.ktr	28/10/2023 23:25	Ficheiro KTR	71 KB
Get_Html.ktr	28/10/2023 23:10	Ficheiro KTR	17 KB

Depois da transformação para html feita, temos outra transformação para obter este html e armazenar numa variável.

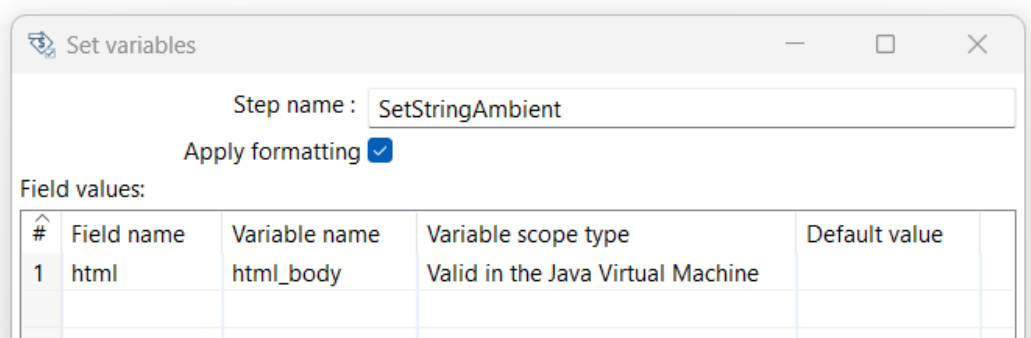
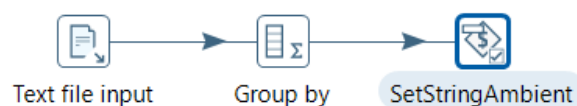
Transformação “Get_HTML”:



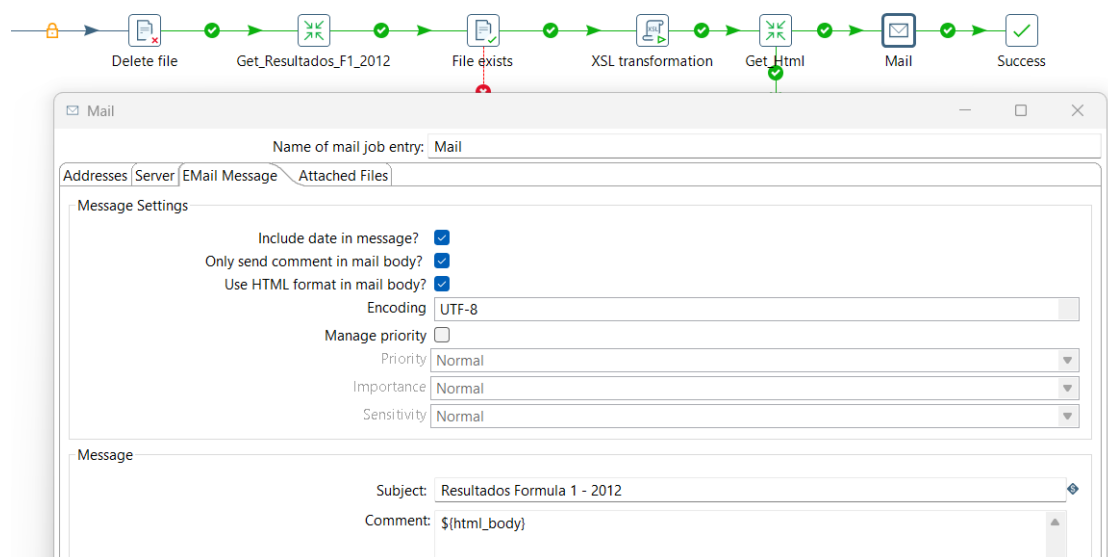
Figura 3 - Transformação "Get_HTML"



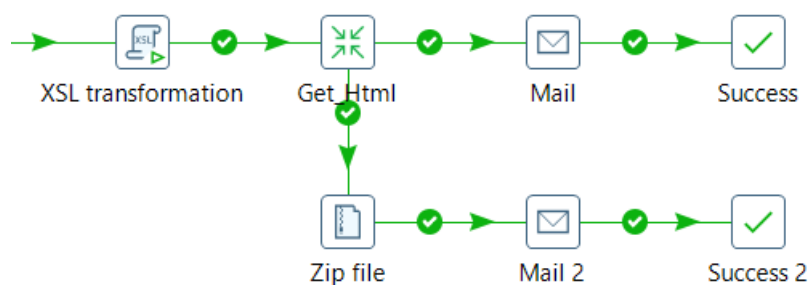
Obtemos assim o ficheiro html.



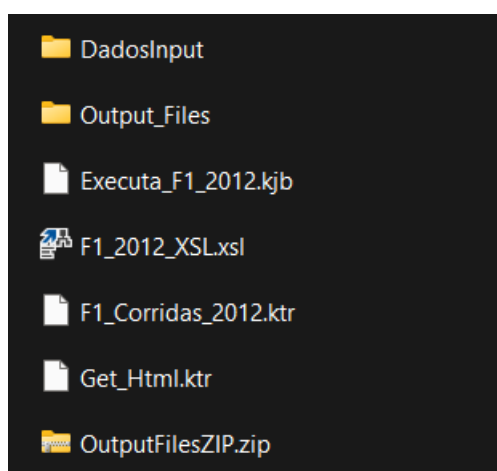
E depois criamos a variável “html_body” com o conteúdo do arquivo html.



Assim podemos enviar o conteúdo do html por email colocando a variável no espaço “Comment” do E-mail.



Enviamos também um segundo E-mail com um ficheiro Zip com o conteúdo gerado pelas transformações efetuadas, zipando a pasta Output_Files, e finalizamos o Job com Sucesso.



Resultados dos Emails obtidos:

Aluno

para mim

15:33 (há 0 minutos)

☆ ↻ ⓘ

🌐

inglês > português

Traduzir mensagem

Desativar para mensagens em inglês x

Resultados Formula 1 - 2012									
Ronda	Nome_GP	Data	Nome_Circuito	Localização	País	Numero_Piloto	Nome_Piloto	Apelido_Piloto	Nome_Equipa
1	"Australian Grand Prix"	"2012-03-18"	"Albert Park Grand Prix Circuit"	"Melbourne"	"Australia"	22	"Jenson"	"Button"	"McLaren"
2	"Malaysian Grand Prix"	"2012-03-25"	"Sepang International Circuit"	"Kuala Lumpur"	"Malaysia"	14	"Fernando"	"Alonso"	"Ferrari"
3	"Chinese Grand Prix"	"2012-04-15"	"Shanghai International Circuit"	"Shanghai"	"China"	6	"Nico"	"Rosberg"	"Mercedes"
4	"Bahrain Grand Prix"	"2012-04-22"	"Bahrain International Circuit"	"Sakhir"	"Bahrain"	5	"Sebastian"	"Vettel"	"Red Bull"
5	"Spanish Grand Prix"	"2012-05-13"	"Circuit de Barcelona-Catalunya"	"Montmeló"	"Spain"	13	"Pastor"	"Maldonado"	"Williams"
6	"Monaco Grand Prix"	"2012-05-27"	"Circuit de Monaco"	"Monte-Carlo"	"Monaco"	W	"Mark"	"Webber"	"Red Bull"
7	"Canadian Grand Prix"	"2012-06-10"	"Circuit Gilles Villeneuve"	"Montreal"	"Canada"	44	"Lewis"	"Hamilton"	"McLaren"
8	"European Grand Prix"	"2012-06-24"	"Valencia Street Circuit"	"Valencia"	"Spain"	14	"Fernando"	"Alonso"	"Ferrari"
9	"British Grand Prix"	"2012-07-08"	"Silverstone Circuit"	"Silverstone"	"UK"	W	"Mark"	"Webber"	"Red Bull"
10	"German Grand Prix"	"2012-07-22"	"Hockenheimring"	"Hockenheim"	"Germany"	14	"Fernando"	"Alonso"	"Ferrari"
11	"Hungarian Grand Prix"	"2012-07-29"	"Hungaroring"	"Budapest"	"Hungary"	44	"Lewis"	"Hamilton"	"McLaren"
12	"Belgian Grand Prix"	"2012-09-02"	"Circuit de Spa-Francorchamps"	"Spa"	"Belgium"	22	"Jenson"	"Button"	"McLaren"
13	"Italian Grand Prix"	"2012-09-09"	"Autodromo Nazionale di Monza"	"Monza"	"Italy"	44	"Lewis"	"Hamilton"	"McLaren"
14	"Singapore Grand Prix"	"2012-09-23"	"Marina Bay Street Circuit"	"Marina Bay"	"Singapore"	5	"Sebastian"	"Vettel"	"Red Bull"
15	"Japanese Grand Prix"	"2012-10-07"	"Suzuka Circuit"	"Suzuka"	"Japan"	5	"Sebastian"	"Vettel"	"Red Bull"
16	"Korean Grand Prix"	"2012-10-14"	"Korean International Circuit"	"Yeongam County"	"Korea"	5	"Sebastian"	"Vettel"	"Red Bull"
17	"Indian Grand Prix"	"2012-10-28"	"Buddh International Circuit"	"Uttar Pradesh"	"India"	5	"Sebastian"	"Vettel"	"Red Bull"
18	"Abu Dhabi Grand Prix"	"2012-11-04"	"Yas Marina Circuit"	"Abu Dhabi"	"UAE"	7	"Kimi"	"Räikkönen"	"Lotus F1"
19	"United States Grand Prix"	"2012-11-18"	"Circuit of the Americas"	"Austin"	"USA"	44	"Lewis"	"Hamilton"	"McLaren"
20	"Brazilian Grand Prix"	"2012-11-25"	"Autódromo José Carlos Pace"	"São Paulo"	"Brazil"	22	"Jenson"	"Button"	"McLaren"

Arquivo ZIP - Ficheiros Resultados F1 - 2012 ➤

Caixa de entrada x

Aluno




para mim

Arquivo ZIP com ficheiros gerados com os resultados de Formula 1 no ano de 2012.

Message date: 2023/10/29 15:32:59.358

Um anexo • Verificado pelo Gmail

OutputFilesZIP.zip

OutputFilesZIP.zip 3 itens		
Nome	Última modificação	Tamanho do ficheiro
 2012F1.html	29/10/2023	8 KB
 F1_Resultados_2012.xls	29/10/2023	19 KB
 Resultados_F1_2012.txt	29/10/2023	5 KB

DADOS WORKERSDATA:

Primeira Transformação “Obter_WorkersData”:

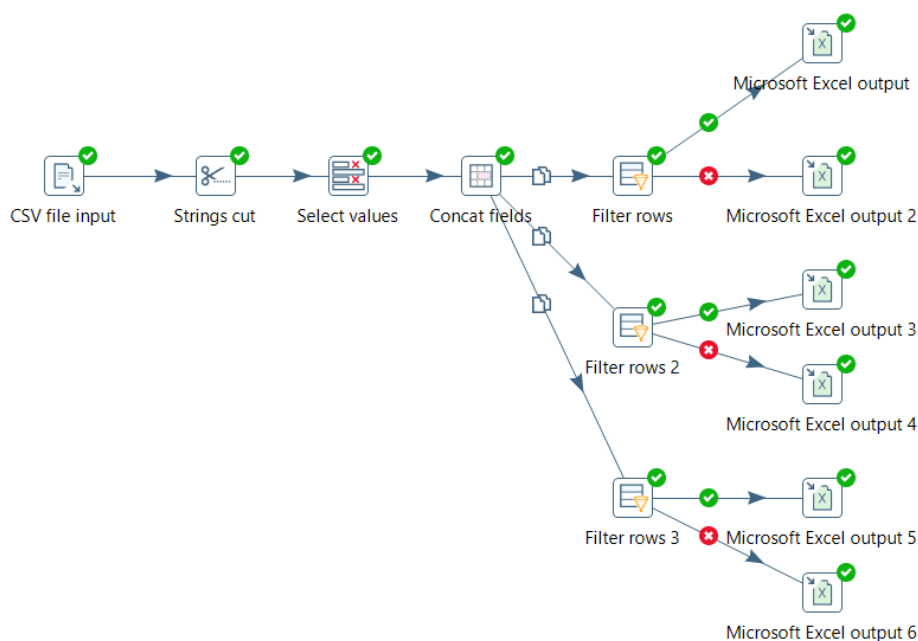
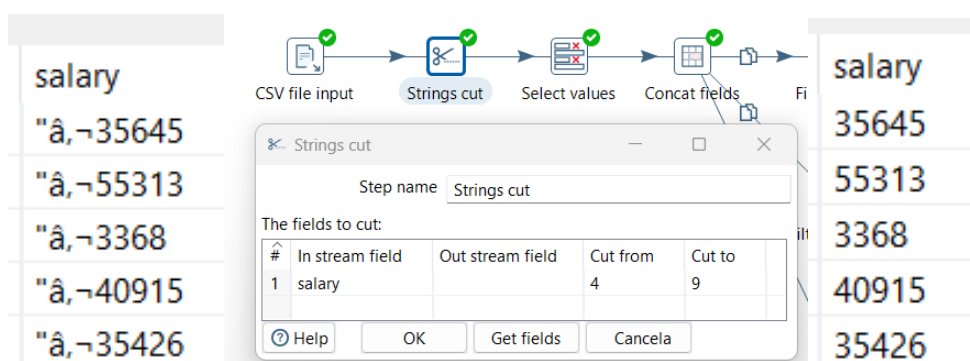
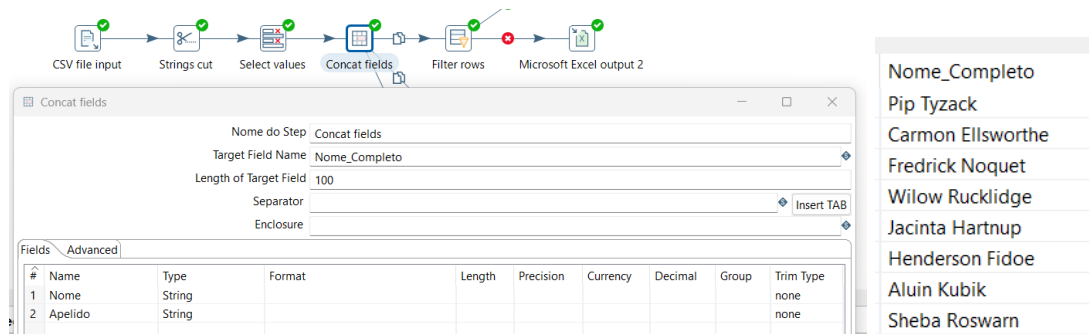


Figura 4 - Transformação "Obter_WorkersData"

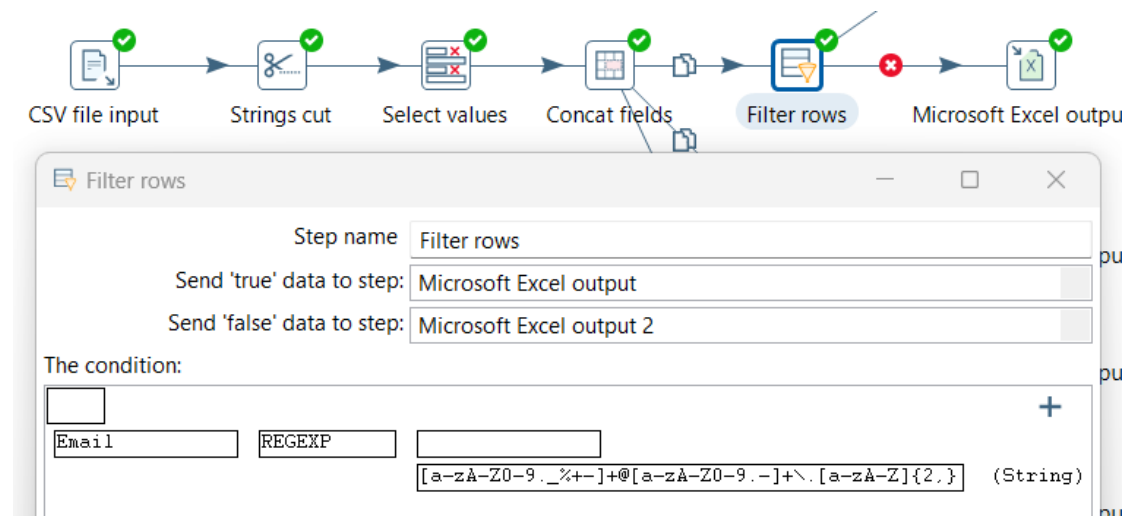
Nesta transformação começamos por importar o ficheiro CSV e utilizamos o campo “Strings cut” para cortar uma parte da tabela salário já que o símbolo do dólar ficou com uma formatação indesejada:



Depois, utilizamos o campo “Select Values” para mudar o nome dos campos e remover os campos indesejados, e utilizamos o campo Concat Fields para juntar o Nome e o Apelido das pessoas em apenas um campo:



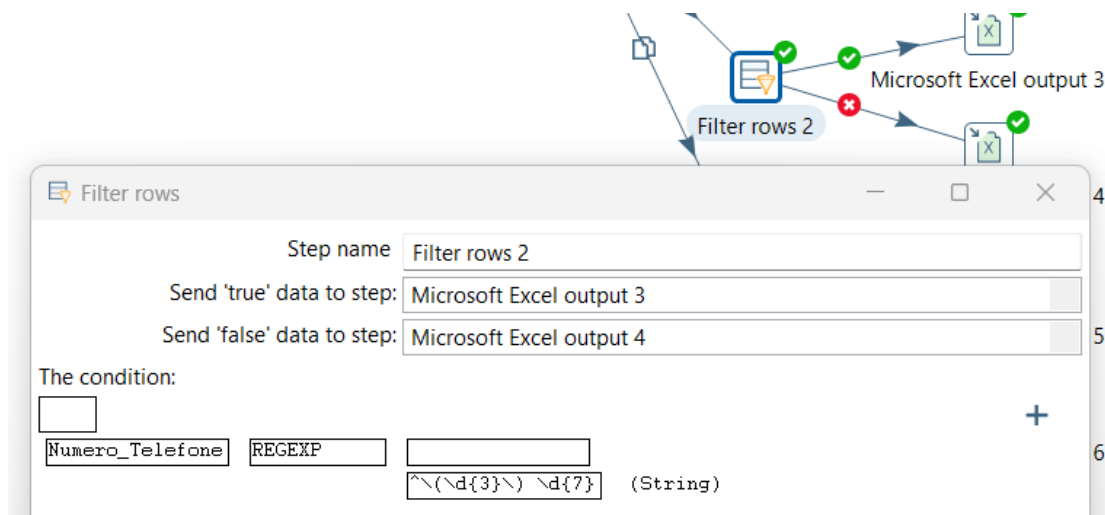
Para finalizar a transformação, filtramos os dados através da ferramenta “Filter Rows” utilizando as expressões regulares:



Aqui foi utilizada a expressão regular:

`[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}`

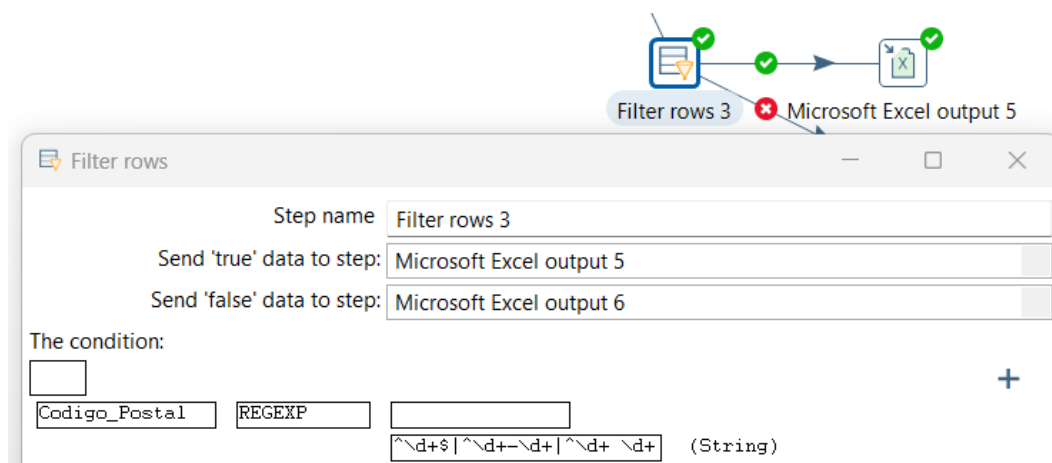
Para filtrar e-mails válidos e inválidos.



Aqui foi utilizada a expressão regular:

`^(\\d{3}\\) \\d{7}`

Para filtrar Números de telefone válidos e inválidos.



Por fim, aqui foi utilizada a expressão regular:

`^\\d+$|^\\d+-\\d+|^\\d+ \\d+`

Para filtrar Códigos Postais válidos e inválidos.

Nome	Data de modificação	Tipo	Tamanho
CodPostais_Invalidos.xls	29/10/2023 15:46	Folha Cálculo do ...	53 KB
CodPostais_Validos.xls	29/10/2023 15:46	Folha Cálculo do ...	48 KB
Emails_Invalidos.xls	29/10/2023 15:46	Folha Cálculo do ...	36 KB
Emails_Validos.xls	29/10/2023 15:46	Folha Cálculo do ...	66 KB
Numeros_Invalidos.xls	29/10/2023 15:46	Folha Cálculo do ...	25 KB
Numeros_Validos.xls	29/10/2023 15:46	Folha Cálculo do ...	76 KB

Obtemos assim os ficheiros XLS com os dados filtrados.

Job “WorkersData_Run”:

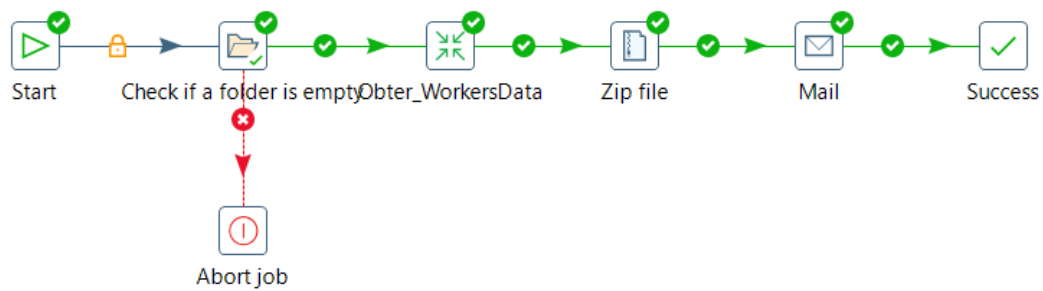
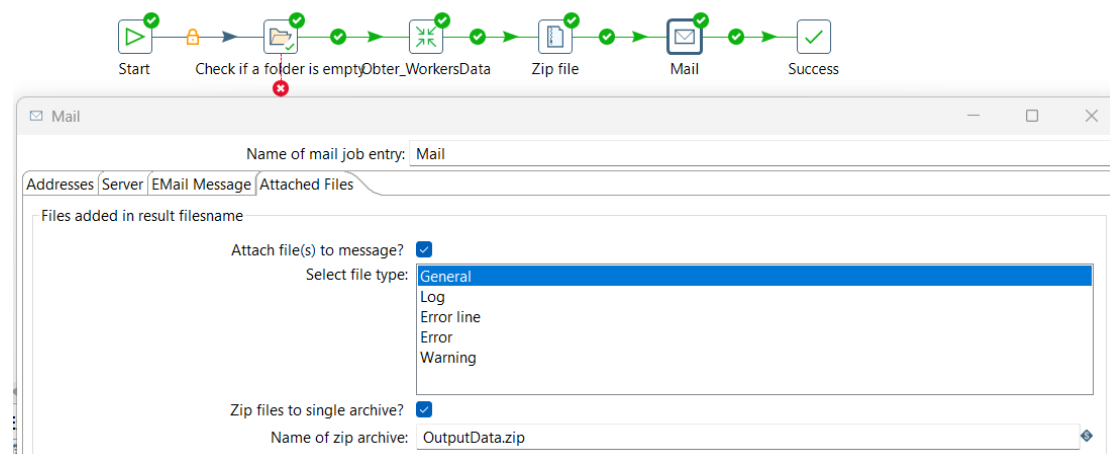


Figura 5 - Job "WorkersData_Run"

Com este Job é possível verificar se a pasta “OutputData” está vazia, caso esta não esteja o job é abortado, e caso a pasta esteja vazia é executada a transformação “Obter_WorkersData” para gerar os arquivos. De seguida a pasta “OutputData” é zipada e enviada por email.



Resultados dos Emails obtidos:

Arquivos ZIP WorkersData Gerados ➤ Caixa de entrada x



Aluno

para mim ▼



inglês ▼

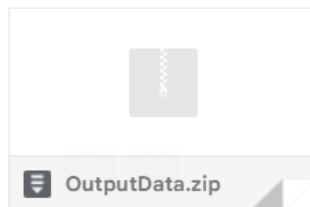


português ▼

[Traduzir mensagem](#)

Message date: 2023/10/29 16:02:06.235

Um anexo • Verificado pelo Gmail ⓘ



7. Conclusão:

O desenvolvimento deste projeto permitiu a aquisição de novas práticas no desenvolvimento de um projeto com base em ETL, permitindo adquirir conhecimentos e novas práticas na manipulação de dados e diferentes tipos de arquivos. Permitiu também ganhar novos conhecimentos em relação à manipulação de arquivos XML e XSL que permitem gerar arquivos HTML, e enviar os mesmos através de plataformas remotas como o E-mail. Foi também importante para desenvolver a utilização de Expressões Regulares na filtragem de dados.

Este Trabalho foi também bastante útil para desenvolver o conhecimento sobre a plataforma Kettle Pentaho, que permite efetuar vários tipos de transformações e manipulações de dados.

8. Bibliografia/Web-grafia:

<https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integration-analytics.html>

<https://stackoverflow.com>

<https://etl-tools.info/en/pentaho/kettle-etl.htm>