

Tabela de conteúdos

Metodologia.....	2
Modelo multidimensional - Steelwheels.....	2
Esquemas relacionais.....	2
Exportação das fontes de dados.....	6
Excel.....	6
MySQL remoto.....	7
CSV.....	7
Howtos.....	9
JAVA_HOME em sistemas UNIX.....	9
Considerações adicionais.....	9
Software Data Integration em sistemas UNIX.....	9
Considerações adicionais.....	10
Considerações MacOS:.....	10
Conexão à base de dados no data-integration.....	10
Registo de Trabalho.....	12

Metodologia

Nesta secção serão descritas todas as opções tomadas pelo grupo, seguidas de explicações, gráficos e imagens que os descrevam.

O software *BI server* e *Data Integration* foi instalado pelos membros do grupo em sistemas de base UNIX, nomeadamente Macintosh e Linux, como tal o processo de instalação poderá diferir relativamente aos *howtos* fornecidos pelo docente.

Modelo multidimensional - *Steelwheels*

Inicialmente foi desenvolvido um modelo multidimensional, baseado numa estrutura em estrela, resultado de uma simplificação da amostra *Steelwheels* explorada previamente na componente prática. Para tal, foram analisados os factos e dimensões da amostra, obtendo três dimensões distintas, unidas através de uma tabela de factos. Consoante o enunciado, foi seleccionada uma dimensão relativa à localização geográfica. O modelo desenvolvido está presente na figura 1 que se segue.

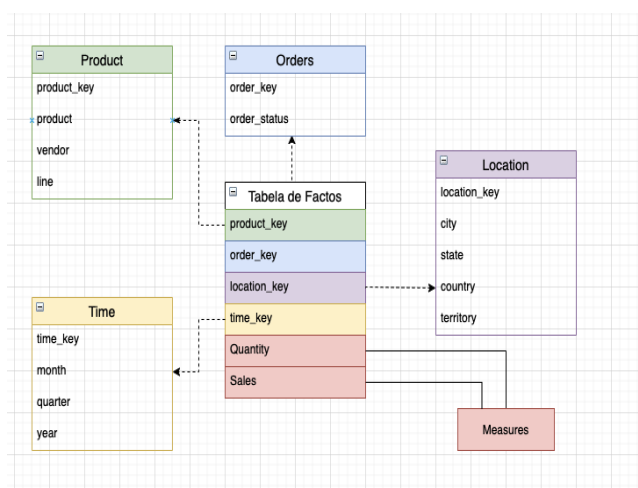


Illustration 1: Modelo Simplificado Steelwheels

Esquemas relacionais

Seguidamente, foram construídos 6 modelos de esquema relacional, resultantes de uma simplificação do modelo desenvolvido na secção anterior. Para tal foram exportadas quatro estruturas de dados do software Pentaho Data Integration, nomeadamente: (i) product; (ii) time; (iii) customer; (iv) orders. Adicionalmente foi introduzida uma tabela de factos, no centro do modelo, constituído pelas chaves estrangeiras das tabelas (i), (ii), (iii) e (iv), tal como por métricas de avaliação destas dimensões.

De acordo com o enunciado, foram retiradas algumas referências geográficas destas tabelas, tal como o estado e país, pois cada modelo representa os dados de um dado país. Adicionalmente foram realizadas alterações nos nomes de atributos (idiomas diferentes), valores monetários (EURO / DOLAR..), separação de atributos, entre outros. Os países utilizados para a criação destes modelos foram: (i) EUA; (ii) Espanha; (iii) França; (iv) Austrália; (v) Nova Zelândia; (vi) Reino Unido. Segue-se a figura 2 que representa a versão simplificada do ER da BD original e as variações criadas para cada um dos restantes países.

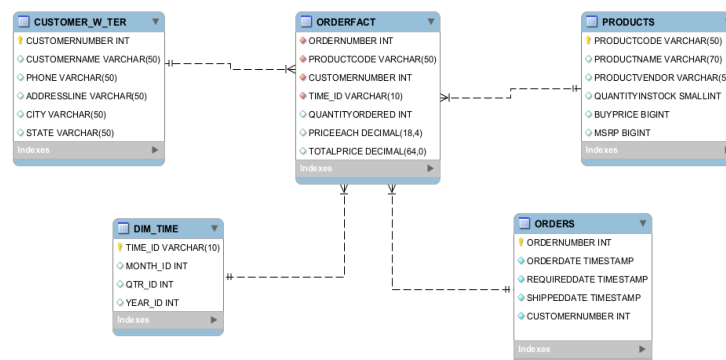


Illustration 2: ER - Tabela base

As seguintes estruturas demonstram os restantes países, seguido de uma breve explicação das alterações realizadas.

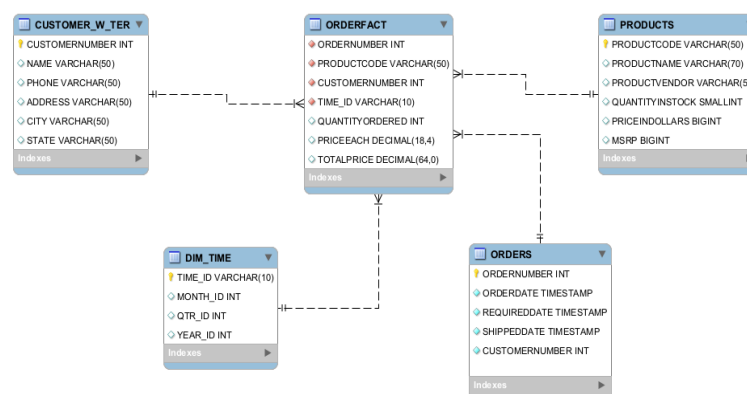


Illustration 3: ER – EUA

O esquema da figura 3 representa uma estrutura do país (i), onde os atributos ADDRESSLINE e CUSTOMERNAME foram modificados para ADDRESS e NAME, respetivamente. Adicionalmente a moeda utilizada foi definida para o DÓLAR.

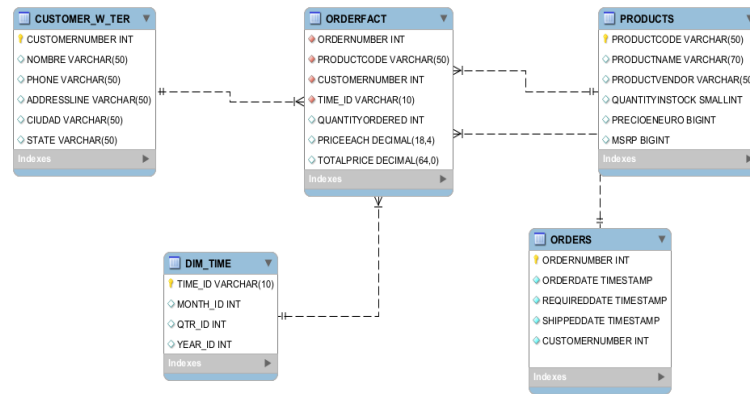


Illustration 4: ER - Espanha

O esquema da figura 4 representa uma estrutura do país (ii), onde os atributos CUSTOMERNAME, CITY e BUYPRICE foram traduzidos para espanhol. Adicionalmente a moeda utilizada foi definida para o EURO.

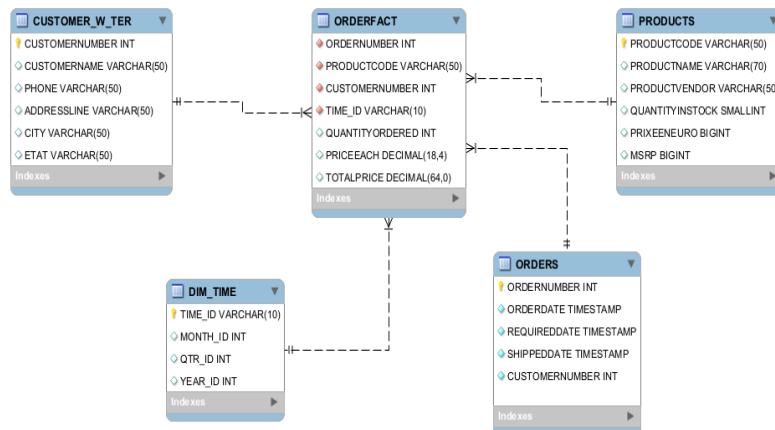


Illustration 5: ER - França

O esquema da figura 5 representa uma estrutura do país (iii), onde os atributos STATE e BUYPRICE foram traduzidos para francês. Adicionalmente a moeda utilizada foi definida para o EURO.

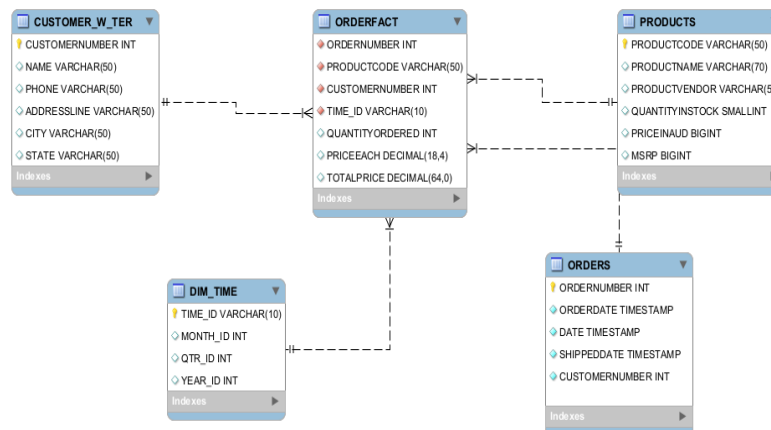


Illustration 6: ER - Austrália

O esquema da figura 6 representa uma estrutura do país (iv), onde os atributos CUSTOMERNAME e REQUIREDDATE foram modificados para NAME e DATE, respetivamente. Adicionalmente a moeda utilizada foi definida para o dólar australiano.

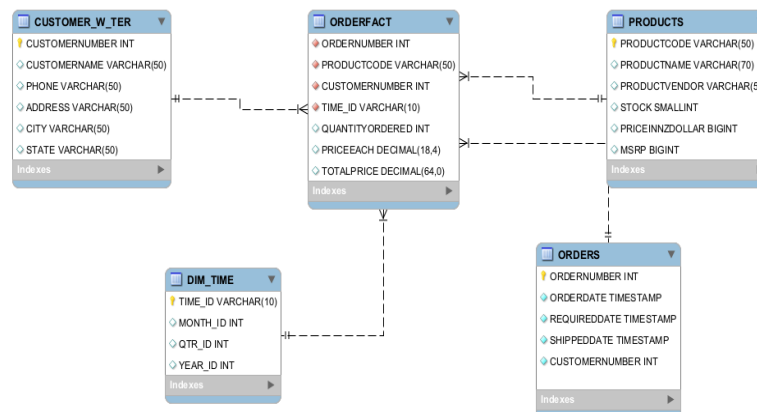


Illustration 7: ER - Nova Zelândia

O esquema da figura 7 representa uma estrutura do país (v), onde os atributos foram alterados de modo a respeitarem o contexto geográfico, ao alterar a moeda para dólares da Nova Zelândia. Adicionalmente o campo ADDRESSLINE da tabela CUSTOMER_W_TER foi alterado para ADDRESS e alterado o campo QUANTITYINSTOCK da tabela PRODUCTS para STOCK.

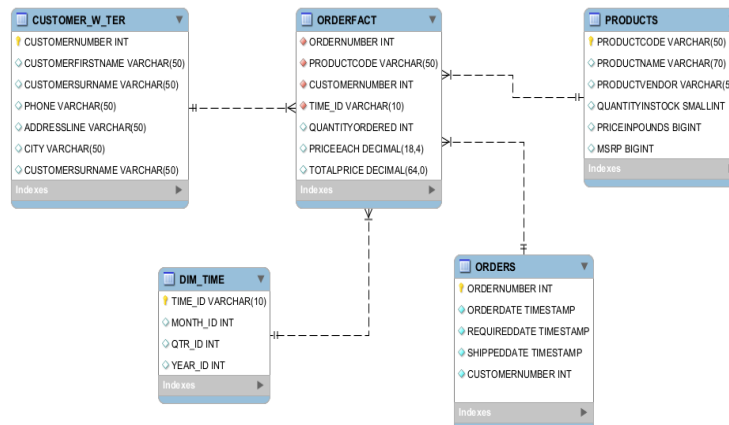


Illustration 8: ER - Reino Unido

O esquema da figura 8 representa uma estrutura do país (vi), onde os atributos foram alterados de modo a respeitarem o contexto geográfico, ao dividir o campo CUSTOMERNAME em dois, nomeadamente CUSTOMERFIRSTNAME e CUSTOMERSURNAME. Adicionalmente foi realizada uma alteração na moeda utilizada para *pound*.

Exportação das fontes de dados

Após uma breve discussão, o grupo decidiu criar 4 das 6 fontes para cada um dos países nos seguintes formatos de exportação:

- USA – exportação de dados no formato .csv;
- Espanha – exportação de dados no formato .mdb;
- França – exportação de dados no formato .xls;
- Austrália – exportação de dados em MySQL;

As exportações em falta já estão a ser desenvolvidas pelo grupo, mas aquando da entrega RT03 as mesmas não estavam finalizadas.

Excel

A figura 9 representa a criação da fonte de dados referente à França. Utilizando o Kettle, criou-se esta configuração de forma a exportar as tabelas da base de dados (simplificada) SteelWheels para um único ficheiro Excel. Cada tabela da base de dados irá corresponder a uma sheet deste ficheiro. Foi também configurada através de JavaScript a conversão de dólares americanos para euros, de forma a demonstrar a funcionalidade de Data Transformation.

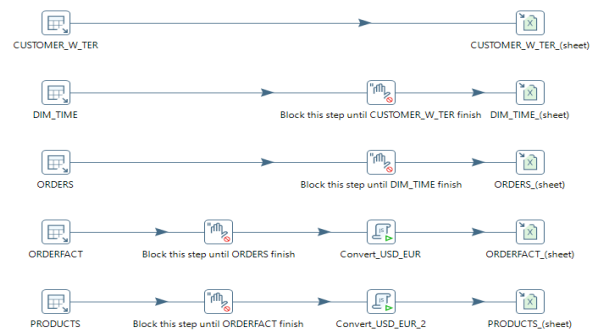


Illustration 9: Criação de fonte para exportação de dados em Excel.

MySQL remoto

A figura 10 representa a criação da fonte de dados referente à Austrália. Utilizando o Kettle, criou-se esta configuração de forma a exportar as tabelas da base de dados (simplificada) SteelWheels em SQL para formato de base de dados MySQL. Para tal foi necessário criar uma base de dados, neste caso “australia_db” através do XAMPP. Posteriormente, realizou-se uma conexão com a mesma no Kettle de forma a se poder exportar as tabelas. Foi também configurada através de JavaScript a conversão de dólares americanos para dólares australianos, de forma a demonstrar a funcionalidade de Data Transformation.

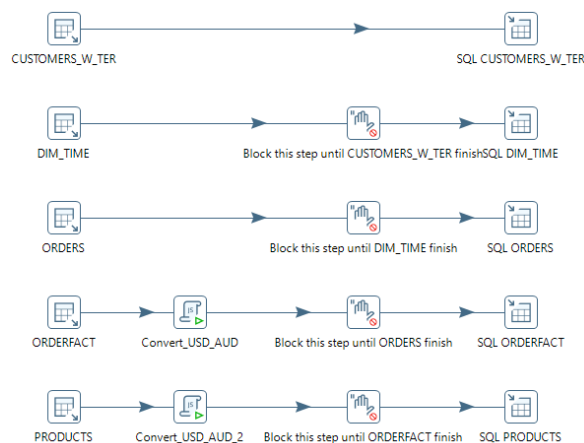


Illustration 10: Criação de fonte para exportação de dados em MySQL.

CSV

A exportação da fonte de dados csv foi realizada para o país EUA, como é possível verificar na figura 11. De forma semelhante aos casos anteriores, foram realizadas *queries* para obter os dados referentes de cada tabela deste país. Seguidamente foi feita uma verificação por tuplos nulos e renomeados os campos de modo a corresponderem ao ER da figura 3. Para finalizar os dados foram exportados em formato CSV para uma pasta EUA, onde cada ficheiro representa uma tabela distinta. Neste caso não foi realizada uma transformação monetária, pelo facto dos valores já estarem em dólar americano.

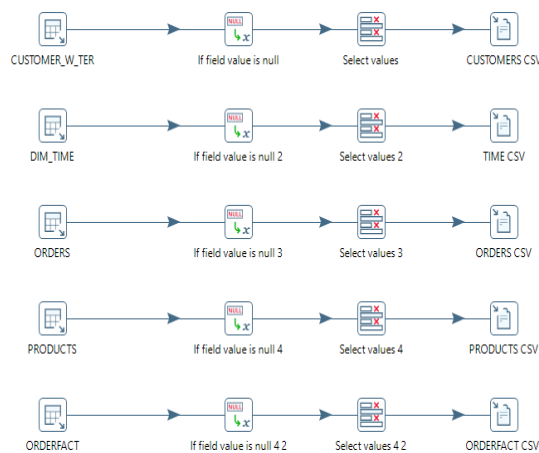


Illustration 11: Criação de fonte para exportação de dados em CSV.

Microsoft Access

Para a exportação de dados para MDB utilizou-se os dados referentes ao país Espanha, como realizado anteriormente realizou-se as devidas queries para cada tabela pretendida, sempre mantendo a ligação apenas com o país em questão, para que todos os resultados obtidos pertencem ao mesmo. Na tabela de ORDERFACT foi necessário realizar mais uma vez a conversão da moeda de USD para EURO, outra observação a ter em consideração nesta mesma tabela, foi necessário alterar o tipo das colunas PRICEEACH e TOTALPRICE de *number* para *string* pois não estava a permitir introduzir os valores no ficheiro MDB devido a um erro de precisão. No final deste processo obteve-se um único ficheiro MDB contendo este todas as tabelas referente ao país Espanha.

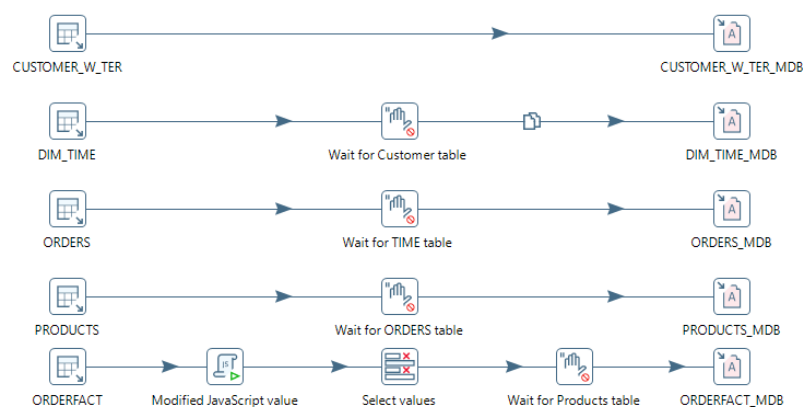


Illustration 12: Criação de fonte para exportação de dados em MDB

Howtos

De forma a sistematizar os problemas encontrados e facilitar a execução de trabalho futuro usando as ferramentas do Pentaho, esta secção contém uma listagem de itens que, no momento da realização do trabalho, não constavam nos howtos online disponibilizados, de modo a que outros grupos possam usufruir destas novas soluções.

JAVA_HOME em sistemas UNIX.

1. Instalar JAVA 8 - algumas funcionalidades não funcionam em outras versões após consulta de documentação
 1. ***sudo apt install openjdk-8-jdk***
2. Descobrir caminho para a pasta JAVA
 1. ***dirname \$(dirname \$(readlink -f \$(which javac)))***
3. Adicionar o caminho anterior ao ambiente de desenvolvimento
 1. Editar o ficheiro /etc/profile
 1. Adicionar export **JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64**
 2. Reiniciar sessão do utilizador
4. Verificar sucesso da operação
 1. ***echo \$JAVA_HOME*** - deverá retornar o caminho detetado anteriormente
 2. ***java -version*** - deverá retornar a versão atualmente instalada do java

Considerações adicionais

1. Verificar a versão do java necessária para o software Pentaho aquando da instalação, pois a versão 8 poderá não ser a utilizada atualmente.
2. Como opção poderá definir o caminho da variável num ficheiro .bashrc ou .zprofile, dependendo do interpretador de comandos UNIX que esteja a utilizar.
3. Poderá ser necessário exportar a variável PATH
 1. ***export PATH=\${PATH}:\${JAVA_HOME}/bin***

Software Data Integration em sistemas UNIX.

1. Requisitos
 1. Java
 2. Variáveis de ambiente configuradas
2. Download do ficheiro zip
 1. ***<https://sourceforge.net/projects/pentaho/>***
3. Extrair ficheiros para uma pasta à sua escolha
4. Executar o ficheiro spoon.sh que se encontra dentro da pasta data-integration
 1. ***./spoon.sh***

Considerações adicionais

1. Como opção poderá seguir o tutorial fornecido na documentação da plataforma Pentaho. O sistema pedirá que efetue o registo, mas na realidade pode fornecer dados aleatórios pois não é realizada uma verificação do email.

1. <https://www.hitachivantara.com/en-us/pdf/white-paper/pentaho-ce-installation-guide-on-linux-operating-system-whitepaper.pdf>

Considerações MacOS:

Ao tentar instalar o pentaho community edition, deparamo-nos com a informação que este ainda não estava disponível para MacOS, para tal teve-se de recorrer à versão enterprise.

Nota: A versão enterprise quando se desinstala e se volta a instalar efetua uma nova renovação de licença.

Pentaho from Hitachi Vantara Overview

End to end data integration and analytics platform

Pentaho tightly couples data integration with business analytics in a modern platform that brings together IT and business users to easily access, visualize and explore all data that impacts business results. Use it as a full suite or as individual components that are accessible on-premise in the cloud or on-the-go (mobile). Pentaho Kettle enables IT and developers to access and integrate data from any source, and deliver it to your business applications, all from within an intuitive and easy to use graphical tool.

Need help installing PDI? Access the installation guides for the following operations systems:

Windows: <https://www.hitachivantara.com/en-us/pdf/white-paper/pentaho-community-edition-installation-guide-for-windows-whitepaper.pdf>

Linux: <https://www.hitachivantara.com/en-us/pdf/white-paper/pentaho-ce-installation-guide-on-linux-operating-system-whitepaper.pdf>

Mac: Coming Soon

Figure 1: Mensagem suporte do Pentaho com a informação da ausência da versão community para MacOS.

Instalação plugin Saiku:

Nos sistemas MacOS tendo a versão enterprise instalada existe uma alteração na diretoria.

Sendo assim deve-se introduzir a pasta que está dentro do ficheiro zip (disponível [aqui](#)), na seguinte diretoria: `/Applications/Pentaho/server/pentaho-server/pentaho-solutions/system`, ficando assim `/Applications/Pentaho/server/pentaho-server/pentaho-solutions/system/saiku`. Os restantes passos podem ser seguidos no [Howto I](#).

Para iniciar o servidor local é executar o ficheiro **start.command**, disponível na diretoria `/Applications/Pentaho`, para desligar e salvar as alterações do servidor local é só executar o ficheiro **stop.command** disponível na mesma diretoria.

Conexão à base de dados no data-integration

1. Ao configurar uma nova ligação à base de dados que contém a amostra Steelwheels a password da conexão poderá ser não só “pentaho_user”, mas também “password”.

Sistemas de Apoio à Decisão

SAD20202021-RT-N.03-GRUPO-N.A05-

JOAOFRANCO.e.JOSEFREITAS.e.LEONARDOABREU



Registo de Trabalho

Entrega	Segmento	Global			Individual	Individual	Individual	Descrição
		Total ->	74,5			27,5	27,5	
		# Participantes				Leonardo Abreu	José Freitas	João Franco
RT03	89	3	X	X	X	Leitura e discussão do enunciado, divisão de tarefas		
RT03	98	1	O	X	O	Leitura e discussão do enunciado, divisão de tarefas		
RT03	99	1	O	X	O	Correção dos modelos ER desenvolvidos anteriormente		
RT03	100	1	O	X	O	Correção dos modelos ER desenvolvidos anteriormente		
RT03	101	1	O	X	O	Correção dos modelos ER desenvolvidos anteriormente		
RT03	102	1	O	X	O	Revisão do documento anterior		
RT03	103	1	O	X	O	Análise da exportação de dados em excel		
RT03	104	1	O	X	O	Análise da exportação de dados em excel		
RT03	105	1	O	X	O	Exportação das tabelas CSV para os EUA		
RT03	106	1	O	X	O	Exportação das tabelas CSV para os EUA		
RT03	107	1	O	X	O	Exportação das tabelas CSV para os EUA		
RT03	108	1	O	X	O	Formatação dos dados em CSV		
RT03	109	1	O	X	O	Escrita do documento RT03		
RT03	110	1	O	O	X	Escrita do documento RT03		
RT03	118	1	O	O	X	Criação de fonte de dados para exportação em xls: França (ligação)		
RT03	119	1	O	O	X	Criação de fonte de dados para exportação em xls (criação das tables: CUSTOMER_W_TER, DIM_TIM e ORDERS)		
RT03	120	1	O	O	X	Criação de fonte de dados para exportação em xls (teste exportação para xls Nota: Não estava a funcionar pois não tinha sido implementado do Block this step until finish)		
RT03	121	1	O	O	X	Criação de fonte de dados para exportação em xls (criação das tables: ORDERFACT e PRODUCTS e teste de exportação para xls.)		
RT03	122	1	O	O	X	Pesquisa de como realizar conversão de moedas usando JS no Data Integration (videos e Howtos disponibilizados)		
RT03	123	1	O	O	X	Criação de fonte de dados para exportação em xls (Script de conversão de USD para Euros)		
RT03	124	1	O	O	X	Criação de fonte de dados para exportação em xls: Australia (ligação)		
RT03	125	1	O	O	X	Criação de fonte de dados para exportação em xls: Australia (criação das tables)		
RT03	126	1	O	O	X	Criação de fonte de dados para exportação em xls: Australia (criação das tables)		
RT03	127	1	O	O	X	Criação de fonte de dados para exportação em xls: Australia (Script de conversão de USD para AUD e testes à exportação para a bd)		
RT03	128	1	O	O	X	Criação de fonte de dados para exportação em xls: Australia (Ligação para a bd com o auxilio do Leonardo, pois mysql-connector-java estava em falta)		
RT03	129	1	O	O	X	Instalação do Pentaho no windows para posteriormente testar exportação de dados com o MS Access.		
RT03	138	1	X	O	O	Tentativa de resolução de erro na versão enterprise do pentaho para MacOS (SEM SUCESSO)		
RT03	139	1	X	O	O	Instalação da versão community no MacOS do pentaho server		
RT03	140	1	X	O	O	Instalação da versão community no MacOS do data integration		
RT03	141	1	X	O	O	Tentativa de resulição de bug na versão no data integration, não		

						permitia realizar a conexão com o pentaho-server (S/sucesso)
RT03	142	1	X	O	O): Instalação da versão community no Windows do Pentaho-server :(
RT03	143	1	X	O	O): Instalação da versão community no Windows do Pentaho Data Integration :(
RT03	144	1	X	O	O	Instalação do Microsoft Access e Xampp para o windows.
RT03	145	1	X	O	O	Exploração das funcionalidade de script, calculadora, opções de output
RT03	146	1	X	O	O	Exploração das funcionalidade de script, calculadora, opções de output
RT03	147	1	X	O	O	No processo de aguardar pela as instalações de software estive em video chamada com os colegas a auxiliar nas tarefas de exportação dos mesmos
RT03	148	1	X	O	O	Criação da transformação para extração de dados para um ficheiro .MDS do país: Espanha
RT03	149	1	X	O	O	Criação da transformação para extração de dados para um ficheiro .MDS do país: Espanha