

## **Tabela de conteúdos**

Metodologia.....	2
Modelo multidimensional - Steelwheels.....	2
Esquemas relacionais.....	2
Howtos.....	6
JAVA_HOME em sistemas UNIX.....	6
Considerações adicionais.....	6
Software Data Integration em sistemas UNIX.....	6
Considerações adicionais.....	7
Considerações MacOS:.....	7
Registo de Trabalho.....	8
.....	8

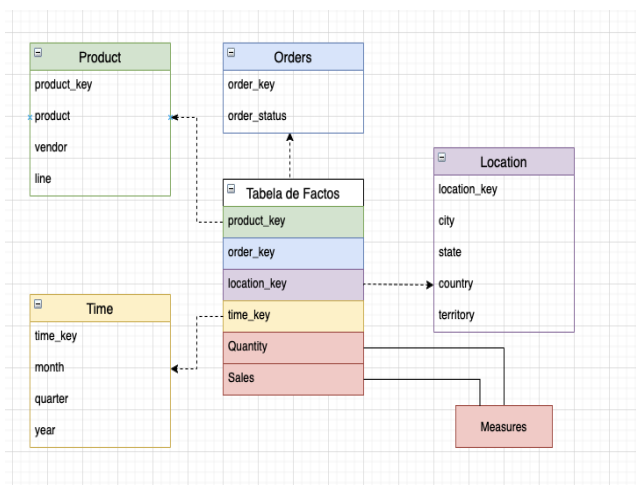
## Metodologia

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 todas as referências geográficas destas tabelas, pois cada modelo representa os dados de um dado país, e 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 que servirá como esquema do país (i) e as variações criadas para cada um dos restantes países.

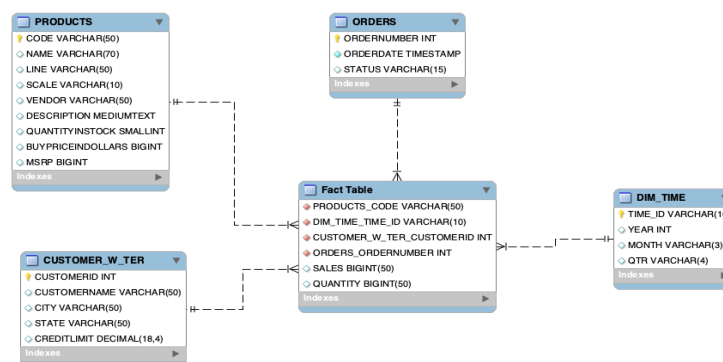


Illustration 2: ER - Tabela base e EUA

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

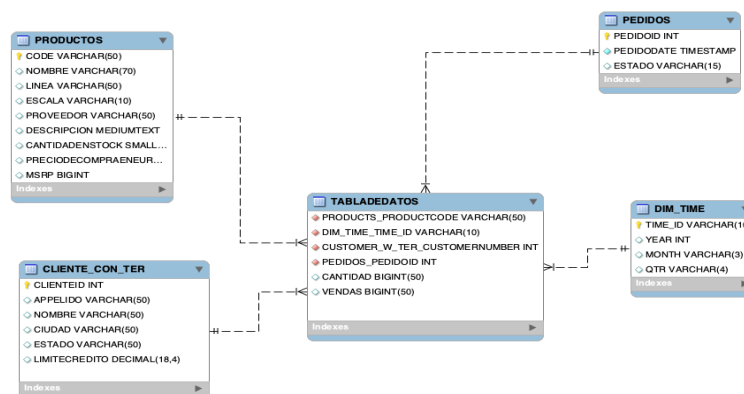


Illustration 3: ER - Espanha

O esquema da figura 3 representa uma estrutura do país (ii), onde os atributos foram traduzidos para espanhol, adicionalmente a moeda utilizada foi definida para o EURO. Adicionalmente o CUSTOMERNAME foi dividido em dois campos denominados por APELLIDO E NOMBRE

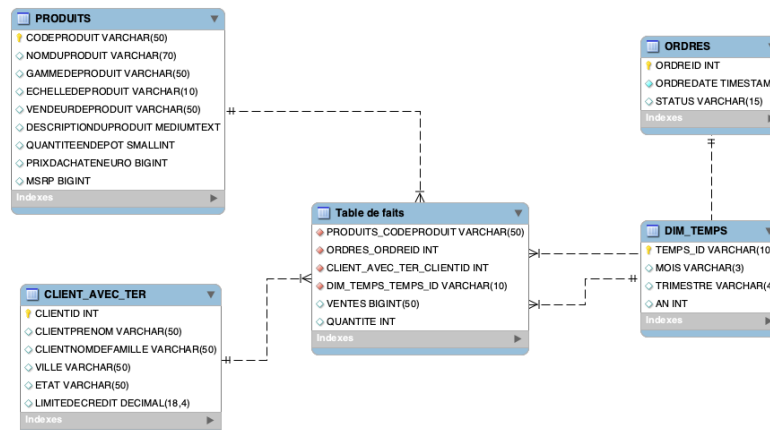


Illustration 4: ER - França

O esquema da figura 4 representa uma estrutura do país (iii), onde os atributos foram traduzidos para francês, adicionalmente a moeda utilizada foi definida para o EURO.

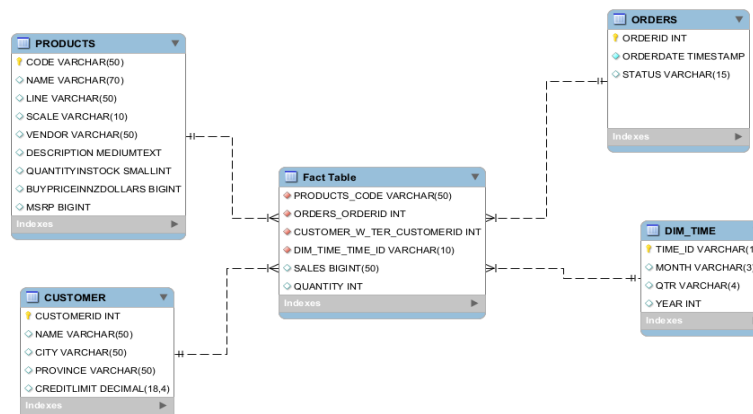


Illustration 5: ER - Nova Zelândia

O esquema da figura 5 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 foi retirado o campo REQUIREDDATE da tabela ORDERS e alterado o campo PROVINCE da tabela CUSTOMER.

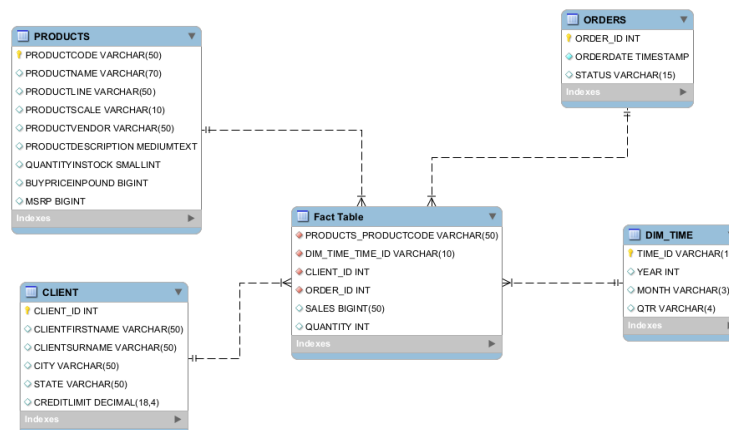


Illustration 6: ER - Reino Unido

O esquema da figura 6 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 CLIENTFIRSTNAME e CLIENTSURNAME. A tabela CUSTOMER foi alterada para CLIENT e foi ainda realizada uma alteração na moeda utilizada para *pound*.

## Criação das seis fontes

Foi iniciado o processo para a exportação de valores através do Kettle, de modo a transferir todas as tabelas necessárias para um ficheiro excel. Este processo foi testado de acordo com o digrama na figura 7, onde é possível visualizar o processo de obter as tabelas, normalizar os valores monetários e obter estes mesmos dados num ficheiros excel. Presente conceito será desenvolvido e explorado na próxima entrega no projeto.



Illustration 7: Diagrama de exportação

## 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 têm-se de 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.

# Registo de Trabalho

		Global			Individual	Individual	Individual	
	Total ->	44			17	17	17	
Entrega	Segmento	#	Participantes	Leonardo Abreu	José Freitas	João Franco	Descrição	
RT02	32	3	X	X	X		Leitura e discussão do enunciado, divisão de tarefas	
RT02	33	3	X	X	X		Exploração de interface e funcionalidades do data integration	
RT02	43	1	O	X	O		Instalação do software data integration	
RT02	44	1	O	X	O		Instalação do software data integration	
RT02	45	1	O	X	O		Definição de modelo E-R - UK	
RT02	46	1	O	X	O		Definição de modelo E-R - NZ	
RT02	47	1	O	X	O		Revisão dos 2 modelos E-R - UK e NZ	
RT02	48	1	O	X	O		Escrita do documento RT02	
RT02	49	1	O	X	O		Escrita do documento RT02	
RT02	50	1	O	X	O		Escrita do documento RT02	
RT02	51	1	O	X	O		Escrita do documento RT02	
RT02	52	1	O	X	O		Registo de horas	
RT02	61	1	O	O	X		Instalação do Pentaho Enterprise no MacOS	
RT02	62	1	O	O	X		Instalação do Pentaho Data Integration e MySQL Workbench	
RT02	63	1	O	O	X		Definição dos Modelos E-R para os países França e Austrália	
RT02	64	1	O	O	X		Definição dos Modelos E-R para os países França e Austrália	
RT02	65	1	O	O	X		Definição dos Modelos E-R para os países França e Austrália	
RT02	66	1	O	O	X		Escrita do documento RT02	
RT02	67	1	O	O	X		Escrita do documento RT02	
RT02	68	1	O	O	X		Visualização de tutoriais para a transformação de dados no Kettle	
RT02	69	1	O	O	X		Exploração da funcionalidade de transformação de dados no data integration (Calculator)	
RT02	70	1	O	O	X		Registo de horas	
RT02	79	1	X	O	O		Tentativa de instalação da versão community do pentaho para MacOS	
RT02	80	1	X	O	O		Reinstalação do pentaho enterprise para MacOS	
RT02	81	1	X	O	O		Exploração Interface Data integration	
RT02	82	1	X	O	O		Criação da Tabela de Factos	
RT02	83	1	X	O	O		Definição do ER - USA e SPAIN	



RT02	84	1	X	O	O	Data Integration -> Export em excel dos dados para o mesmo documento, apenas modificando a sheet.
RT02	85	1	X	O	O	Exploração da funcionalidade de transformação de dados no data integration (Calculator)
RT02	86	1	X	O	O	Criação de uma calculadora que faz a conversão do valor de Dolar para Euro
RT02	87	1	X	O	O	Criação da query com InnerJoins para obter a tabela de factos (Por terminar 23/03/2021 - 21:12)
RT02	88	1	X	O	O	Registo de horas