

Enunciado do Projeto de Sistemas de Suporte à Decisão 2024-25

ETL – Trabalho Prático 2/3

LEI/LGI

Introdução

A FoodMart é uma empresa multinacional de distribuição generalista com 24 lojas físicas instaladas nos Estados Unidos, México e Canadá às quais junta a distribuição porta-a-porta através de canais eletrónicos (mobile e web) disponibilizado aos clientes no final do ano passado. A FoodMart emprega cerca de 1150 empregados. Ao contrário da maioria das lojas físicas no mercado de distribuição, é solicitado aos clientes da Foodmart um registo / identificação para acesso às lojas ou canais eletrónicos, este processo facilita o processo de compra e entrega de mercadorias ao cliente assim como os pagamentos.

Os sistemas de informação de suporte à operação da FoodMart consistem num ERP (**Pegasus**) e numa plataforma de comércio eletrónico (**Morpheus**). A generalidade dos dados mestre são armazenados e geridos no sistema Pegasus que disponibiliza aos restantes sistemas da organização um conjunto de WebServices que permitem efetuar a interligação de dados mestre entre os vários sistemas. No caso de existirem dados mestre particulares de cada plataforma, os mesmos serão geridos pelos próprios sistemas individualmente.

Pré-requisitos

Para efetuar o setup das Bases de Dados necessárias ao Trabalho 2, deverá seguir os seguintes passos:

1. Download dos ficheiros:
 - a. [Script_BD_Pegasus.sql](#)
 - b. [Script_BD_FoodMartDW.sql](#)
 - c. [Script_BD_Pegasus_20130505.sql](#)
 - d. [Time.xls](#)
2. Abrir o **SQL Server Management Studio** ou **Azure Data Studio**.
3. Criar a BD **Pegasus** através do script: **Script_BD_Pegasus.sql**, ou de forma manual, crie a base de dados e depois os objetos presentes no mesmo script.
4. Criar a BD do Data Mart através do script: **Script_BD_FoodMartDW.sql**, ou de forma manual, crie a base de dados e depois os objetos presentes no mesmo script.
5. Confirmar no SQL Server Management Studio ou no **Azure Data Studio** que as Bases de Dados **Pegasus** e **FoodmartDW** foram criadas com sucesso. A Base de Dados **Pegasus** deverá conter dados.

Notas sobre a execução do trabalho

Na sequência do projeto de implementação da solução de Data Warehouse na Foodmart, foram tomadas as seguintes decisões em termos de modelos de informação:

- Não vai ser integrada na 1^a fase informação do sistema **Morpheus**, uma vez que o sistema está a sofrer um processo de atualização significativa ao nível de estruturas de dados.
- A implementação de processos de **ETL** já se iniciou, mas não foi concluída, pelo que é necessário implementar os processos de **ETL** remanescentes.
- Deverão ser considerados para efeitos de captura de alterações ao nível dos dados e referenciamento no tempo, os campos **ChangeDatetime** criados nas tabelas relevantes do sistema **Pegasus**.

A execução dos processos de ETL, deve ser efetuada em dois passos (carregamento **inicial** e carregamento **incremental**). Para tal, devem ser seguidos os seguintes procedimentos:

1. Execução do carregamento inicial do DataMart com todo o histórico de informação até 2013-05-04 (Dados no **Script_BD_Pegasus.sql**). No caso de usar-se datas de vigência nos registos, a correspondente data de início deverá ser “01/01/2011”.
2. Validação da informação carregada. Verificar no SQL Server se, de facto, os dados foram transferidos.
3. Execução do script file **Script_BD_Pegasus_20130505.sql** para atualização da base de dados no dia 2013-05-05.
4. Avaliação da informação sincronizada. Devem ser verificadas após a execução dos processos de ETL (carregamento **inicial** e carregamento **incremental**) a existência de novos registos e/ou alteração dos mesmos de acordo com o comportamento implementado ao nível do SCD.

Ex 1. Load Customer Dimension

Com recurso aos conhecimentos adquiridos na utilização da ferramenta de ETL **Pentaho Data Integration**, implemente os processos de ETL necessários para o carregamento da Dimensão **D_CUSTOMER** da base de dados **FoodmartDW**. Considere os seguintes pontos para implementação:

- **Nome da transformação:** Load_D_Customer.ktr
- **Sistema Fonte:** Pegasus
- **Tabela(s) relevantes do sistema Fonte:** Customer
- **Sistema Destino:** FoodMartDW
- **Tabela(s) destino:** D_Customers
- **Mapeamento dos campos:**

Pegasus.Customer	FoodMartDW.D_Customer
PK	PK
Customer_ID	CustomerKey
Company Name	Customer_ID
Contact Name	Company Name
Contact Title	Contact Name
Address	Contact Title
City	Address
Region	City
PostalCode	Region
Country	Country
Phone	
Fax	
ChangeDateTime	
	version
	date_from
	date_to

- Recorra ao campo **ChangeDatetime** para detetar alterações e reforçar a validação com recurso à comparação de chaves.

- Não é relevante manter o histórico da informação ao nível das características da dimensão.
- 1.1. Implemente o processo de **ETL** utilizando o Pentaho DI e guarde no ficheiro **Load_D_Customer.ktr**
 - 1.2. Descreva a transformação feita, os componentes (**steps**) utilizados e o resultado da mesma.
 - 1.3. De acordo com a especificação de dados, indique a técnica de **Slowly Changing Dimensions** que deve aplicar a este processo de **ETL**? Justifique a resposta.

Deverá responder às questões 1.2 e 1.3 utilizando o ficheiro de documentação [FoodmartDW_Doc.docx](#).

Ex 2. Load Product Dimension

Com recurso aos conhecimentos adquiridos na utilização da ferramenta de **ETL** Pentaho DI, implemente os processos de **ETL** necessários para o carregamento da Dimensão **D_PRODUCT** da base de dados **FoodmartDW**. Considere os seguintes pontos para implementação:

- **Nome da transformação:** Load_D_Product.ktr
- **Sistema Fonte:** Pegasus
- **Tabela(s) relevantes do sistema Fonte:** Products, ProductCategories
- **Sistema Destino:** FoodMartDW
- **Tabela(s) destino:** D_Product
- **Mapeamento dos campos:**

Pegasus.Products	FoodMartDW.D_Product
PK ProductID	PK ProductKey
ProductName	ProductID
FK SupplierID	ProductName
FK CategoryID	SupplierID
QuantityPerUnit	CategoryID
UnitCostPrice	UnitCostPrice
UnitPrice	UnitPrice
UnitsInStock	
UnitsOnOrder	
ReorderLevel	
Discontinued	
ChangeDateTime	
Pegasus.ProductCategories	
PK CategoryID	CategoryName
CategoryName	
Description	
Picture	
	version
	date_from
	date_to

- Utilize o campo **ChangeDatetime** para detetar alterações e reforçar a validação com recurso à comparação de chaves.

- No caso de alterações no preço dos produtos (UnitCostPrice e UnitPrice), deve ser preservado o histórico.

2.1 Implemente o processo de ETL utilizando o Pentaho DI e guarde no ficheiro **Load_D_Product.ktr**

2.2 Descreva a transformação feita, os componentes (steps) utilizados e o resultado da mesma.

2.3 De acordo com a especificação de dados, qual a técnica de **Slowly Changing Dimensions** que deve aplicar a este processo de ETL? Justifique a resposta.

Deve responder às questões 2.2 e 2.3 utilizando o ficheiro de documentação [FoodmartDW_Doc.docx](#).

Ex 3. Load Time Dimension

Com recurso aos conhecimentos adquiridos na utilização da ferramenta de ETL Pentaho DI, implemente os processos de ETL necessários para a criação e carregamento da Dimensão **D_TIME** da base de dados **FoodmartDW**. Considere os seguintes pontos para implementação:

- **Nome da transformação:** Load_D_Time.ktr
- **Sistema Fonte:** Time.xls
- **Sistema Destino:** FoodMartDW
- **Tabela(s) destino:** D_Time
- **Mapeamento dos campos:**

Time.xlsx	FoodMartDW.D_Time
Date	MonthKey
MonthNumberOfYear	MonthNumberOfYear
EnglishMonthName	EnglishMonthName
CalendarQuarter	CalendarQuarter
CalendarSemester	CalendarSemester
CalendarYear	CalendarYear

- O campo MonthKey resulta da conversão do campo **Date** respeitando o formato **YYYYMM**.

3.1 Implemente o processo de ETL utilizando o Pentaho DI e guarde no ficheiro **Load_D_Time.ktr**

3.2 Descreva a transformação feita, os componentes (steps) utilizados e o resultado da mesma.

3.3 De acordo com a especificação de dados, qual a técnica de **Slowly Changing Dimensions** que deve aplicar a este processo de ETL? Justifique a resposta.

Deve responder às questões 3.2 e 3.3 utilizando o ficheiro de documentação [FoodmartDW_Doc.docx](#).

Ex 4. Load Fact Sales_Month

De forma a tornar os relatórios sobre **Vendas Mensais** mais eficientes, a equipa de implementação do Data Warehouse tomou a decisão de implementar uma tabela Agregada Mensal de Vendas, esta tabela deve summarizar os valores de **Vendas Líquidas** registados ao longo de cada mês.

- **Nome da transformação:** Load_F_Sales_Month.ktr
- **Sistema Fonte:** Pegasus

- **Tabela(s) relevantes do sistema Fonte:** Orders, OrdersDetails
- **Sistema Destino:** FoodMartDW
- **Tabela(s) destino:** F_Sales_Month
- **Mapeamento dos campos:**

Pegasus.Orders	FoodMartDW.F_Sales_Month
	SK MonthKey
	SK CustomerKey
	SK ProductKey
PK OrderID	
FK CustomerID	
FK EmployeeID	
OrderDate	
RequiredDate	
ShippedDate	
ShipVia	
Freight	
ShipName	
ShipAddress	
ShipCity	
ShipRegion	
ShipPostalCode	
ShipCountry	
Pegasus.OrderDetails	
PK OrderID	
FK ProductID	
UnitPrice	
Quantity	
Discount	
	SalesQty
	SalesAmount
	NetSalesAmount
	VatAmount
	ShipmentDuration
	OnTime

- Considere a coluna **MonthKey** como chave lógica para a Dimensão **D_Time** que deve respeitar o formato “YYYYMM” e ser obtida a partir do **OrderDate**.
- Deve utilizar o campo **OrderDate** para detetar novos registo.
- O campo **SalesAmount** deve indicar o valor da venda original, enquanto o **NetSalesAmount** deve ter em conta o desconto aplicado.
- O campo **VatAmount** deve apresentar o valor da venda líquida acrescido de 23%.
- O campo **ShipmentDuration** deve ser obtido tendo em conta a diferença entre o ShippedDate e OrderDate.
- O campo **OnTime** deve indicar se a encomenda foi enviada na data prevista ou não, e deve ser obtida a partir dos campos ShippedDate e RequiredDate.

- 4.1 Implemente o processo de ETL e guarde no ficheiro **Load_F_Sales_Month.ktr**
- 4.2 Descreva a transformação feita, os componentes (**steps**) utilizados e o resultado da mesma.

Deve responder à questão 4.2 utilizando o ficheiro de documentação **FoodmartDW_Doc.docx**.

Ex 5. SQL queries ao FoodMartDW

Após a criação das tabelas de dimensão **D_Customer**, **D_Product** e **D_Time** e da tabela de factos **F_Sales_Month**, está em condições de dar algumas respostas à administração da **FoodMart** acerca dos resultados da empresa. Assim, recorrendo a **queries** SQL, responda às seguintes questões:

- 5.1 Qual o **produto** com o maior volume de vendas (em valor monetário)? A query deve devolver o nome do produto e o valor respetivo.
- 5.2 Quais os três **países** que registaram o maior e menor volume de vendas (em valor monetário)? As queries devem devolver o nome dos países e os respetivos valores.
- 5.3 Qual o **volume de vendas líquido** (NetSalesAmount) e a % desse volume para cada **categoria** e fornecedor (**supplier**) de produtos?
- 5.4 Qual o **prazo mínimo, médio e máximo** de entrega das encomendas por **país**?
- 5.5 Qual o **volume de vendas** e as **quantidades vendidas** para o **2º e 3º trimestres** de cada **ano**?

Coloque o código das **queries** e o resultado das mesmas, no ficheiro de documentação **FoodmartDW_Doc.docx**.

Indicações relativas à Entrega do Trabalho Prático

1. O projeto deve ser colocado numa pasta em formato ZIP com o nome, **GrupoXX_TP2_ETL.zip** onde XX representa o número do grupo (e.g. Grupo01_TP2_ETL.zip).
2. Deve entregar os seguintes ficheiros:
 - a. Ficheiro de transformação **Load_D_Customer.ktr**;
 - b. Ficheiro de transformação **Load_D_Product.ktr**;
 - c. Ficheiro de transformação **Load_D_Time.ktr**
 - d. Ficheiro de transformação **Load_F_Sales_Month.ktr**;
 - e. Documento **FoodmartDW_Doc.docx** com as respostas solicitadas ao longo do enunciado.
3. Após a entrega do trabalho poderá haver lugar a uma apresentação/discussão do trabalho com cada um dos elementos dos grupos.
4. O trabalho deve ser entregue até ao final do dia **12 de Maio de 2025(23h59)** através da tarefa disponível no **Moodle**. O não cumprimento do prazo estabelecido, remete os alunos para o exame de recurso.
5. **O incumprimento do prazo estipulado remete os alunos para o exame de recurso.**