

Enunciado do Projeto de Sistemas de Suporte à Decisão 2021-22

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 um total de 1155 empregados. Ao contrário da maioria das lojas físicas no mercado de distribuição, é requerido aos clientes do 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 à implementação do 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**
2. Abrir o SQL Server Management Studio.
3. Criar BD Pegasus através do script: **Script_BD_Pegasus.sql**
4. Criar BD do Data Mart através do script: **Script_BD_FoodMartDW.sql**
5. Confirmar no SQL Server Management 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 projecto 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 1ª 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 referênciação no tempo os campos **ChangeDatetime** criados nas tabelas relevantes do sistema Pegasus.

A execução dos processos de ETL, deverá ser efetuada em dois passos (carregamento inicial e carregamento incremental). Para tal deverão 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.
3. Execução do script file **Script_BD_Pegasus_20130505.sql** para inclusão de imagem do sistema 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 pelo processo de ETL.

Ex 1. Load Customer Dimension

Com recurso aos conhecimentos adquiridos na utilização da ferramenta de ETL Pentaho Data Integration, implementar os processos de ETL necessários para o preenchimento da Dimensão D_CUSTOMER da base de dados FoodmartDW. Considerar 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_Customers
	PK CustomerKey
PK Customer_ID	Customer_ID
Company Name	Company Name
Contact Name	Contact Name
Contact Title	Contact Title
Address	
City	City
Region	Region
PostalCode	
Country	Country
Phone	

Fax ChangeDateTime	version date_from date_to
-----------------------	---------------------------------

- Deverá ser utilizado o campo ChangeDatetime para detectar alterações e reforçada a validação com recurso à comparação de chaves.
- Não é relevante manter o histórico da informação ao nível das características da dimensão.

- 1.1. Implementar o processo de ETL utilizando o Pentaho DI e guardar no ficheiro **Load_D_Customer.ktr**
- 1.2. Descrever a transformação feita, os componentes (steps) utilizados e o resultado final.
- 1.3. De acordo com a especificação de dados. Qual a técnica de Slowly Changing Dimensions que deverá ser aplicada 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 Data Integration, implementar os processos de ETL necessários para o preenchimento da Dimensão D_PRODUCT da base de dados foodmartDW. Considerar 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 ProductKey
PK ProductID	ProductID
ProductName	ProductName
SupplierID	SupplierID
CategoryID	CategoryID
QuantityPerUnit	
UnitCostPrice	UnitCostPrice
UnitPrice	UnitPrice
UnitsInStock	

UnitsOnOrder	
ReorderLevel	
Discontinued	
ChangeDateTime	
Pegasus.ProductCategories	
PK CategoryID	
CategoryName	CategoryName
Description	
Picture	
	version
	date_from
	date_to

- Deverá ser utilizado o campo ChangeDatetime para detetar alterações e reforçada a validação com recurso à comparação de chaves.
- No caso de alterações no preço dos produtos (UnitCostPrice e UnitPrice) deverá ser guardado o histórico.

2.1 Implementar o processo de ETL utilizando o Pentaho DI e guardar no ficheiro **Load_D_Product.ktr**

2.2 Descrever a transformação feita, os componentes (steps) utilizados e o resultado final.

2.3 De acordo com a especificação de dados. Qual a técnica de Slowly Changing Dimensions que deverá ser aplicada a este processo de ETL? Justifique a resposta.

Deverá responder às questões 2.2 e 2.3 utilizando o ficheiro de documentação foodmartDW_Doc.docx.

Ex 3. 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 deverá sumarizar 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	
CustomerID	
EmployeeID	
OrderDate	
RequiredDate	
ShippedDate	
ShipVia	
Freight	
ShipName	
ShipAddress	
ShipCity	
ShipRegion	
ShipPostalCode	
ShipCountry	
Pegasus.OrderDetails	
PK OrderID	
ProductID	
UnitPrice	
Quantity	
Discount	
	SalesQty
	SalesAmount
	NetSalesAmount
	ShipmentDuration
	OnTime

- Deverá considerar a coluna **MonthKey** como chave lógica para a Dimensão D_Month (ainda que não visível) que deverá seguir a nomenclatura “YYYYMM” e ser apurada a partir da **OrderDate**.
- Deverá ser utilizado o campo **OrderDate** para detetar novos registos.
- O campo **SalesAmount** deverá indicar o valor da venda original enquanto o **NetSalesAmount** deverá ter em consideração o desconto aplicado.

- O campo **ShipmentDuration** deve ser obtido tendo em conta o ShippedDate e OrderDate.
- O campo **OnTime** deverá indicar se a encomenda foi enviada na data prevista ou não e deve ser obtida tendo em conta os campos ShippedDate e RequiredDate.

3.1 Implemente o processo de ETL **Load_F_Sales_Month.ktr**

3.2 Descrever a transformação feita, os componentes (steps) utilizados e o resultado final.

Deverá responder à questão 3.2 utilizando o ficheiro de documentação foodmartDW_Doc.docx.

Ex 4. SQL queries ao DW

Após a criação das dimensões D_Customers e D_Product e da tabela de factos F_Sales_Month, está em condições de dar algumas respostas à administração do FoodMart sobre a performance do negócio. Assim, recorrendo a queries SQL, responda às seguintes questões:

- 4.1 Qual o produto com mais vendas?
- 4.2 Em que países foram registadas vendas entre Junho e Julho de 2011?
- 4.3 Qual o nome dos produtos vendidos cujo preço unitário é superior a 100\$?
- 4.4 Qual o nome dos 3 produtos mais rentáveis?

Deverá colocar as queries e o resultado das mesmas, no ficheiro de documentação foodmartDW_Doc.docx.

Notas relativas à Entrega deste Trabalho Prático

1. O projeto deverá ser colocado numa pasta compactada em formato ZIP com o nome, GrupoXX TP2 ETL.zip onde XX indica o número do grupo (e.g. Grupo01_ TP2_ETL.zip).
2. Deverá 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_F_Sales_Month.ktr**;
 - d. Documento **foodmartDW_Doc.docx** com as respostas aos exercícios.
3. O trabalho deverá ser entregue até ao final do dia 08 de Maio de 2022 (23h59) para o email do professor correspondente às aulas práticas em que está inscrito: p1508@ulusofona.pt ou andrea.costigliola@ulusofona.pt. Os trabalhos entregues depois da última data referida não serão aceites, pelo que o incumprimento dos prazos estipulados obriga os alunos a realizarem a cadeira na 2ª época.