Data Mart Implementation (P01)

DECISION SUPPORT SYSTEMS, 2021-22

**André Cardoso (18848), Leonel Fernandes (18850)**

# Introduction

Este projeto tem como objetivo a implementação de um Data Mart baseado numa base de dados operacional. A estrutura dos dados deve ser analisada, assim como devem ser definidas as regras e requisitos de negócio de negócio. Deve ser ainda construído um modelo de dimensões desse Data Mart e implementado um processo de ETL.

A Base de dados a ser analisada neste projeto é de um negócio de fabrico e venda de bicicletas e de acessórios para as mesmas. Esta base de dados suporta, a compra de bicicletas em loja física, onde vendem para centenas de clientes, assim como através da internet onde o número de clientes é superior, e ainda o fabrico dos produtos. É possível, ainda através da base de dados, o registo de clientes, produtos e fornecedores, assim como a compra de produtos a fornecedores.

# Data sources

Neste capítulo serão estudados os dados provenientes da base de dados. Será criado e implementado um Data Mart de acordo com os requisitos de negócio definidos. Os dados provenientes das tabelas pertinentes serão analisados quanto à qualidade dos dados.

Table 1: Summary of the Database “Adventure Works”

|  |  |  |
| --- | --- | --- |
| **Event / object** | **Table** | **Nr. Records** |
| Produtos | *Product* | 504 |
| Funcionários | *Employees* | 290 |
| Clientes | *Store*  *Person*  *Customer* | 701  19972  19820 |
| Fornecedores | *Vendor* | 104 |
| Vendas no Território | *SalesTerritory* | 10 |
| Forma de transporte de produtos | *ShipMethod* | 5 |
| Compra de produtos | *PurchaseOrderHeader*  *PurchaseOrderDetail* | 4012  8845 |
| Venda de produtos | *SalesOrderHeader*  *SalesOrderDetail* | 31465  121317 |

# Dimensional modelling

Table 2: Data Warehouse Matrix

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DIMENSIONS**  **BUSINESS PROCESSES** | Employee | Customer | Vendor | Product | ShipMethod | SalesReason | SalesTerritory | Data |
| Registar Funcionários | X |  |  |  |  |  |  | X |
| Registar Clientes |  | X |  |  |  |  |  |  |
| Registar Fornecedores |  |  | X |  |  |  |  |  |
| Registar Produtos |  |  |  | X |  |  |  | X |
| Venda de Produtos | X | X |  | X | X | X | X | X |
| Compra de Produtos | X |  | X | X | X |  |  | X |

# Business Processes

Os principais processos de negócio presentes na base de dados:

* Registar Funcionários – Registo de funcionários funcionário na BD.
* Registar Produtos – Registo de novos produtos na BD.
* Registar Clientes – Registo de novos clientes na BD.
* Registar Fornecedores – Registo de novos fornecedores na BD.
* Venda de Produtos – A venda de produtos para clientes ou para lojas.
* Compra de Produtos – A compra de produtos a fornecedores.

# Business Requirements

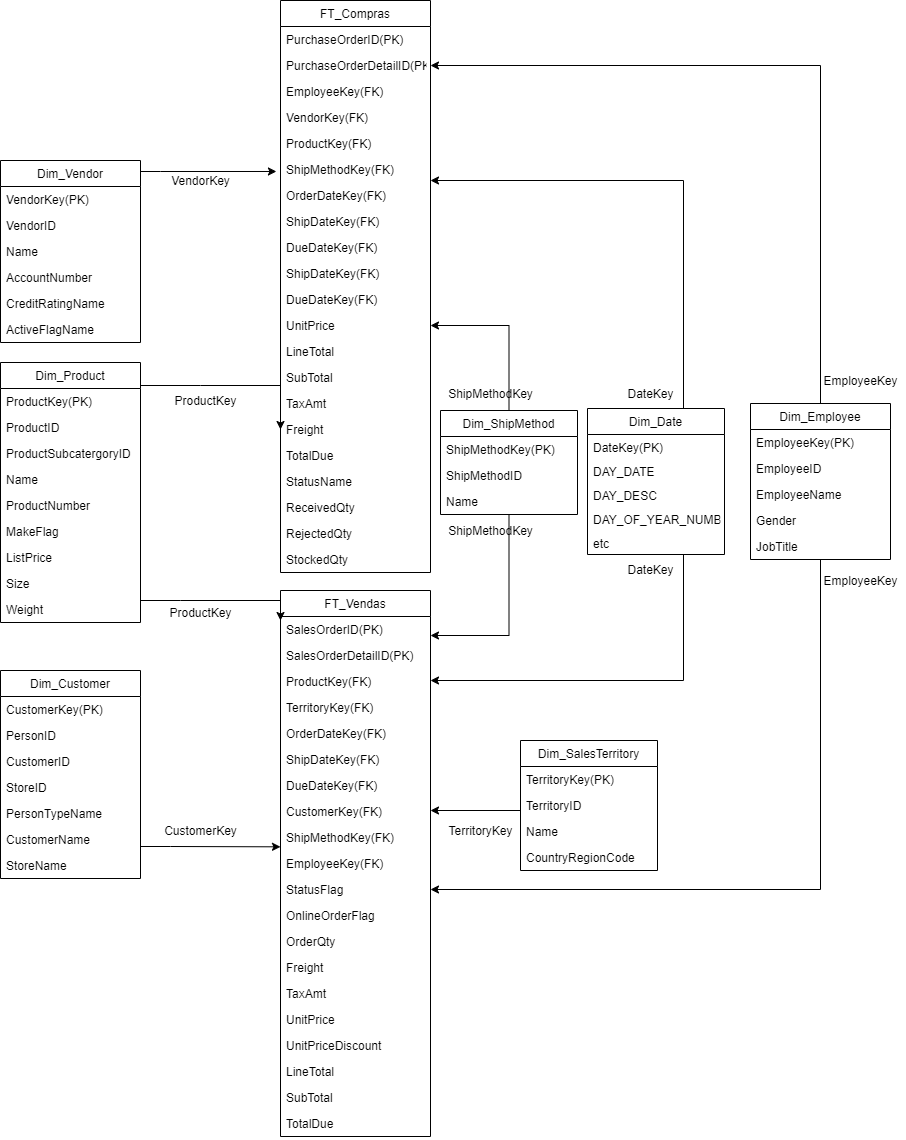
Com o Data Mart desenvolvido, o sistema deve ser capaz de responder a algumas questões sobre os dados lá presentes:

* Qual a quantidade total de produtos vendidos?
* Qual o valor total monetário gerado na venda de produtos?
* Qual o produto mais vendido?
* Quais os produtos mais vendidos por território?
* Quais os territórios onde mais produtos são vendidos?
* Quais os produtos mais rentáveis?
* Qual a percentagem de produtos vendidos online e em loja?
* Qual a quantidade de vendas a lojas e a clientes individuais?
* Qual a quantidade de produtos por cada venda?
* Qual o valor total em taxas de venda de produtos?
* Qual a quantidade de encomendas para cada estado (Também em percentagem)?
* Qual a quantidade de produtos cuja produção foi descontinuada?
* Quais os funcionários que mais produtos encomendam para stock?
* Qual o valor total em taxas de compra de produtos para stock?
* Qual a quantidade de compras para stock em cada estado (Também em percentagem)?
* Qual a percentagem de produtos rejeitados numa compra a fornecedor?
* Quais os fornecedores aos quais foram comprados mais produtos (Também em percentagem)?
* Quais os fornecedores onde foi gasto mais dinheiro na compra de produtos (Também em percentagem)?
* Quais as formas de transporte mais utilizadas?
* Qual o total gasto em cada empresa de transporte?

# Fact Tables

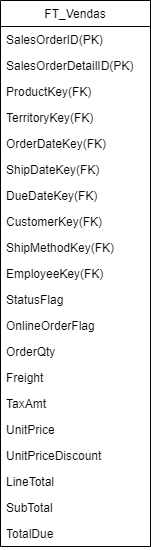
As Tabelas de factos utilizadas são as tabelas de “Vendas” e de “Compras”, duas tabelas de factos transacionais, sendo vendas a junção de “SalesOrderHeader” e “SalesOrderDetail” e compras a junção de “PurchaseOrderHeader” e “PurchaseOrderDetail”, onde a quantidade de registos de cada uma trata-se de quantidade de registos das tabelas “Detail”, sendo a sua granularidade cada produto num pedido de venda e compra, respetivamente.

# Design of the dimensional data model



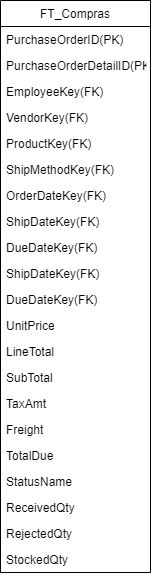
# Tabela de Factos “Vendas”

Esta tabela de factos trata-se da tabela referente à venda de produtos a clientes, cada linha desta tabela refere-se a um produto de uma venda.

* SalesOrderID(PK) – Primary Key da tabela.
* SalesOrderDetailID(PK) – Primary Key da tabela.
* ProductKey(FK) – Trata-se da chave estrangeira proveniente da tabela de produtos.
* TerrritoryKey(FK) – Trata-se da chave estrangeira proveniente da tabela de Território.
* OrderDateKeyFK) – Trata-se da chave estrangeira proveniente da tabela de Date que representa a data de encomenda.
* ShipDateKey(FK) – Trata-se da chave estrangeira proveniente da tabela de Date que representa a data de envio
* DueDateKey(FK) – Trata-se da chave estrangeira proveniente da tabela de Date que representa a data marcada para entrega.
* CustomerKey(FK) – Trata-se da chave estrangeira proveniente da tabela de Customer.
* ShipMethodKey(FK) – Trata-se da chave estrangeira proveniente da tabela de ShipMethod.
* EmployeeKey(FK) – Trata-se da chave estrangeira proveniente da tabela de Employee.
* StatusFlag – Campo que identifica o estado da encomenda.
* OnlineOrderFlag – Campo que identifica se a encomenda foi feita online ou em loja.
* OrderQty – Campo que identifica a quantidade de um produto.
* Freight – Campo que identifica o custo de transporte de uma encomenda.
* TaxAmt – Campo que identifica a quantidade total de taxas numa encomenda.
* UnitPrice – Campo que identifica o preço de cada unidade do produto.
* UnitPriceDiscount – Campo que identifica o desconto por cada produto.
* LineTotal – Campo que identifica o preço da quantidade toda de um produto na encomenda, calculado como UnitPrice \* (1 - UnitPriceDiscount) \* OrderQty.
* SubTotal – Campo que identifica o preço total de uma encomenda, calculado como SUM(LineTotal para cada produto na encomenda com o adequado SalesOrderID).
* TotalDue – Campo que identifica o preço total com taxas e preço de transporte, calculado como(Subtotal + TaxAmt + Freight).

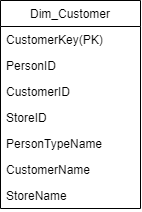
# Tabela de Factos “Compras”

Esta tabela de factos trata-se da tabela referente à compra de produtos a fornecedores, cada linha desta tabela refere-se a um produto de uma compra.

* PurchaseOrderID(PK) – Primary Key da Tabela.
* PurchaseOrderDetailID(PK) – Primary Key da Tabela.
* EmployeeKey(FK) – Trata-se da chave estrangeira proveniente da tabela Employee.
* VendorKey(FK) – Trata-se da chave estrangeira proveniente da tabela Vendor.
* ProductKey(FK) – Trata-se da chave estrangeira proveniente da tabela Product.
* ShipMethodKey(FK) – Trata-se da chave estrangeira proveniente da tabela ShipMethod.
* OrderDateKeyFK) – Trata-se da chave estrangeira proveniente da tabela de Date que representa a data de encomenda.
* ShipDateKey(FK) – Trata-se da chave estrangeira proveniente da tabela de Date que representa a data de envio
* DueDateKey(FK) – Trata-se da chave estrangeira proveniente da tabela de Date que representa a data marcada para entrega.
* UnitPrice – Campo que identifica o preço de cada unidade do produto.
* LineTotal – Campo que identifica o preço da quantidade toda de um produto na encomenda, calculado como UnitPrice \* (1 - UnitPriceDiscount) \* OrderQty.
* SubTotal - Campo que identifica o preço total de uma encomenda, calculado como SUM(LineTotal para cada produto na encomenda com o adequado SalesOrderID).
* TaxAmt – Campo que identifica a quantidade total de taxas numa encomenda.
* Freight – Campo que identifica o custo de transporte de uma encomenda.
* TotalDue – Campo que identifica o preço total com taxas e preço de transporte, calculado como (Subtotal + TaxAmt + Freight).
* StatusName - Campo que identifica o estado da encomenda.
* ReceivedQty – Campo que identifica a quantidade de produtos recebidos.
* RejectedQty – Campo que identifica a quantidade de produtos rejeitados.
* StockedQty – Campo que identifica a quantidade de produtos que foram postos em stock (ReceivedQty – RejectedQty).

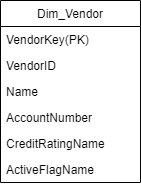
# Dimensão “Customer”

Esta tabela refere-se aos clientes aos quais vão ser feitas vendas de produtos, estes podem ser Lojas ou Individuais. Esta tabela tem uma Surrogate Key o campo “CustomerKey”.

* CustomerKey(PK) – Surrogate Key da tabela.
* PersonID – Identificador o tipo de Pessoa.
* CustomerID – Identificador da tabela Cliente, nulo se for uma loja.
* StoreID – Identificador da tabela loja, nulo se for um cliente individual.
* PersonTypeName – Campo que identifica o tipo de pessoa (cliente ou loja).
* CustomerName – Nome do cliente, nulo se for uma loja.
* StoreName – Nome da loja, nulo se for um cliente.

# Dimensão “Vendor”

Esta tabela refere-se aos fornecedores aos quais vão ser feitas compras de produtos. Esta tabela tem uma Surrogate Key o campo “VendorKey”.

* VendorKey(PK) – Surrogate Key da tabela.
* VendorID – Identificador da tabela Vendor.
* Name – Campo que identifica o nome do fornecedor.
* AccountNumber – Campo que identifica o número de conta de um fornecedor.
* CreditRatingName – Campo que identifica uma avaliação dada ao fornecedor.
* ActiveFlagName – Campo que identifica se o fornecedor se encontra ativo ou não.

# Dimensão “Employee”

Esta tabela refere-se aos empregados que vão fazer encomendas a fornecedores ou fazer vendas a clientes. Esta tabela tem uma Surrogate Key o campo “EmployeeKey”.

* Uma imagem com mesa

  Descrição gerada automaticamenteEmployeeKey(PK) – Surrogate Key da tabela.
* EmployeeID – Identificador da tabela de Employee.
* EmployeeName – Campo que identifica o nome do empregado.
* Gender – Campo que identifica o género do empregado.
* JobTitle – Campo que identifica o tipo de função que o empregado desempenha na empresa.

# Dimensão “ShipMethod”

Esta tabela refere-se aos tipos de transporte para encomendas. Esta tabela tem uma Surrogate Key “ShipMethodKey”.

* Uma imagem com mesa

  Descrição gerada automaticamenteShipMethodKey(PK) – Surrogate Key da tabela.
* ShipMethodID – Identificador da tabela ShipMethod.
* Name – Campo que identifica o nome da empresa responsável pelo transporte

# Dimensão “SalesTerritory”

Esta tabela refere-se aos territórios para encomendas. Esta tabela tem uma Surrogate Key “TerritoryKey”

* Uma imagem com mesa

  Descrição gerada automaticamenteTerritoryKey(PK) – Surrogate Key da tabela
* TerritoryID – Identificador da tabela SalesTerritory.
* Name – Campo que identifica o nome da região
* CountryRegionCode – Campo que identifica o código da região

# Dimensão “Product”

Esta tabela refere-se aos produtos que podem estar presentes numa encomenda a um fornecedor e ainda numa venda a um cliente ou loja. Esta tabela tem uma Surrogate Key “ProductKey”

* Uma imagem com mesa

  Descrição gerada automaticamente ProductKey(PK) – Surrogate Key da Tabela.
* ProductID – Identificador da tabela Product.
* ProductSubcategoryID – Identificador da tabela Subcategory.
* Name – Campo que identifica o nome do produto.
* ProductNumber – Campo que identifica o número do produto.
* MakeFlag – Campo que identifica se o produto foi fabricado na empresa ou não.
* ListPrice – Campo que identifica o preço de venda do produto.
* Size – Campo que identifica o tamanho do produto.
* Weight – Campo que identifica o peso do produto.

# Dimensão “Date”

Uma imagem com texto

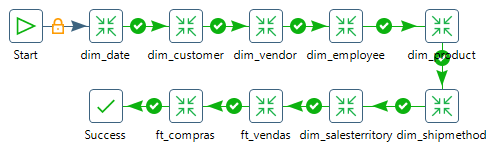
Descrição gerada automaticamente

# Data mart implementation

A solução desenvolvida usando a ferramenta Spoon do programa PDI envolve a criação de um *Job* – definição da sequência lógica do processo – e várias *Transformations* – definição do processo ETL individual a cada tabela (dimensão ou tabela de factos).

# Job

Definição da ordem de execução das Transformations para Dimensões e por fim Tabelas de Factos.



# Transformations

# Dimensão “Employee”

Para a dimensão *Employee* são necessários alguns campos da Tabela “HumanResources.Employee”.

De seguida utilizando o campo **ProductID,** obtido anteriormente, extraímos campos que consideramos relevantes da Tabela “Person.Person”.

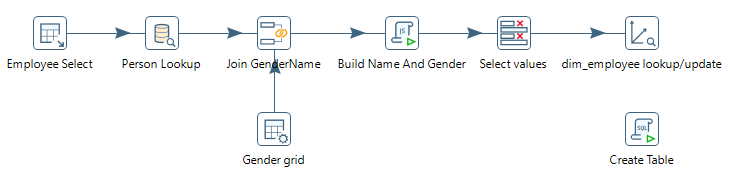
|  |  |
| --- | --- |
| **HumanResources.Employee** | **Person.Person** |
| * BusinessEntityID * Gender * JobTitle | * Title * FirstName * MiddleName * LastName * Suffix |

Tendo extraído todos os dados necessários, transformamo-los para que se tornem mais legíveis sem perda de informação.

Assim, são originados 2 campos: **GID** e **GName** (GID [M, F]; GName[Male, Female]).

Estes dados são cruzados com os já extraídos e posicionados corretamente.

De seguida os campos de nome extraídos da tabela “Person.Person” e os campos criados para o campo **Gender** são trabalhados de modo que seja criado o campo **EmployeeName** com todos os campos da tabela “Person.Person” na ordem correta e o campo **Gender** siga o seguinte formato: ‘Y/N – Yes/No’ .



# Dimensão “Customer”

Para dar origem a esta dimensão, selecionam-se alguns campos da Tabela “Sales.Customer”.

De seguida utilizando os campos **StoreID** e **PersonID,** obtidos anteriormente, extraímos campos que consideramos relevantes das Tabelas “Sales.Store” e “Person.Person”, respetivamente. Para isto é necessário filtrar a Tabela existente dando origem a duas novas tabelas, uma para procurar os dados em “Sales.Store” e outra para os dados em “Person.Person”.

|  |  |  |
| --- | --- | --- |
| **Sales.CustomerID** | **Sales.Store** | **Person.Person** |
| * CustomerID * PersonID * StoreID | * Name | * PersonType * Title * FirstName * MiddleName * LastName * Suffix |

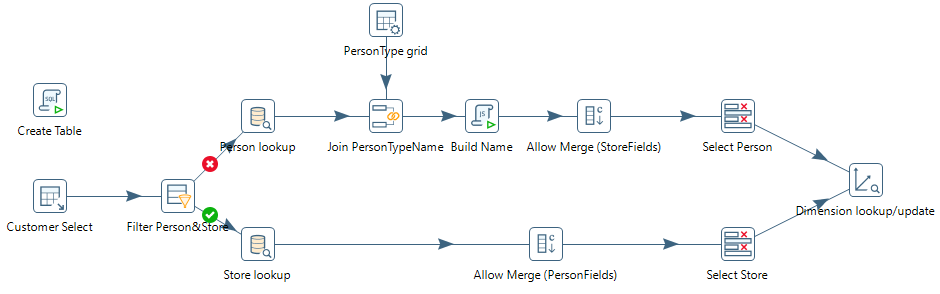
Tendo extraído todos os dados necessários, transformamo-los para que se tornem mais legíveis sem perda de informação.

Assim, são originados 2 campos: **PTID** e **PTName**.

Estes dados são cruzados com os já extraídos (**PersonType**) e posicionados corretamente.

De seguida os campos de nome extraídos da tabela “Person.Person” são trabalhados de modo que seja criado o campo **CustomerName** com todos os campos da tabela “Person.Person” na ordem correta **exceto** o campo **PersonType** que é agora composto por ‘PTID - PTName’.

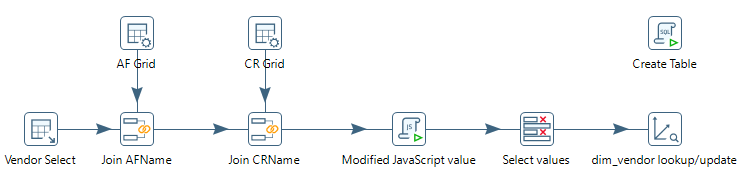
Finalmente, são adicionadas as colunas em falta e selecionadas as colunas necessárias em cada uma das tabelas para que estas sejam estruturalmente iguais para possibilitar a união dos dados numa só tabela.



# Dimensão “Vendor”

A dimensão “Vendor” é de um processo mais simples. Selecionam-se alguns campos da Tabela “Sales.Vendor” e criam-se campos novos (**AFID, AFName** e **CRID, CRName**) para transformação dos campos existentes **ActiveFlag** e **CreditRating.**

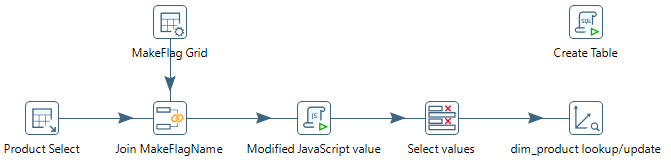
Por fim é feita a transformação (‘AFID/CRID – AFName/CRName’) e selecionam-se os campos importantes para criação da dimensão.



# Dimensão “Product”

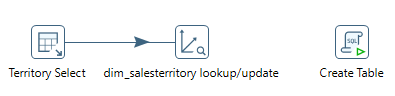
Em semelhança à dimensão anterior, selecionam-se alguns campos da Tabela “Production.Product” e criam-se campos novos (**MFID, MFName**) para transformação do campo existente **MakeFlag.**

Por fim é feita a transformação (‘MFID - MFName’) e selecionam-se os campos importantes para criação da dimensão.



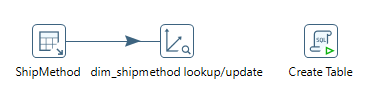
# Dimensão “Sales Territory”

As duas dimensões seguintes são fundamentalmente mais simples, sendo necessário apenas os campos disponíveis nas tabelas da base de dados de origem, neste caso “Sales.SalesTerritory”.



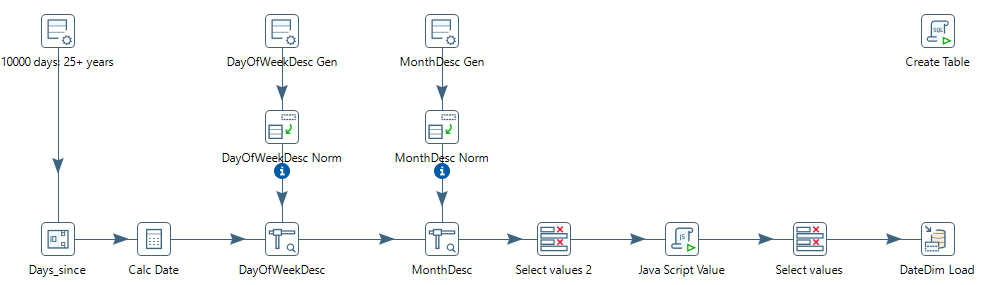
# Dimensão “Ship Method”

Tabela da base de dados de origem “Purchasing.ShipMethod”.



# Dimensão “Date”

Esta dimensão foi fornecida pelo professor, Joaquim Silva, e fornece uma tabela completa com informações diversas acerca cada dia num vasto espaço de tempo mais de 25 anos, como se pode ler na figura.



# Tabela de Factos “Vendas”

A seguinte Tabela de Factos seleciona campos das tabelas “Sales.SalesOrderDetail” e “Sales.SalesOrderHeader” juntando-as tendo por base a chave primária **SalesOrderID**.

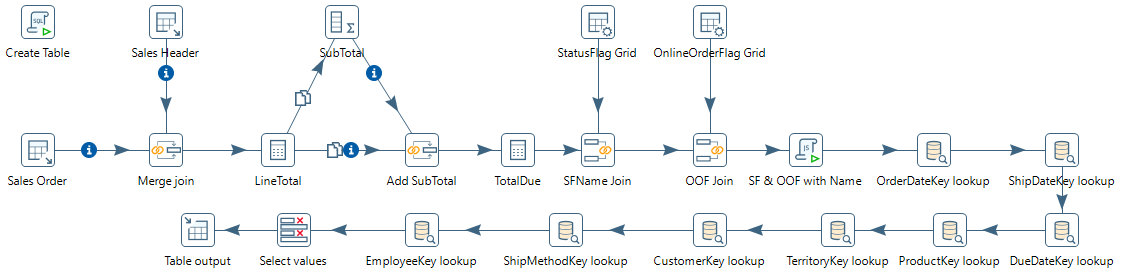
De seguida são calculados alguns campos **LineTotal**, **SubTotal** e **TotalDue**:

* **LineTotal** = ‘OrderQty’ \* ‘UnitPrice’
* **SubTotal** é a soma dos **LineTotal** pertencentes ao mesmo **SalesOrderID**, como total é feito um *Group By*  em função dessa chave.
* **TotalDue** = **SubTotal** + ‘Freight’ + ‘TaxAmt’

Novamente, são gerados campos novos (**SFID, SFName**  e **OOFID, OOFName**) para transformação dos campos existentes **StatusFlag** e **OnlineOrderFlag.**

Por fim é feita a transformação (‘SFID/OOFID – SFName/OOFName’).

De seguida são procuradas e angariadas as *Surrogate Keys* (OrderDateKey, ProductKey, etc.) necessárias e por fim são selecionados os dados importantes deixando de fora os auxiliares para a transformação e cálculos.

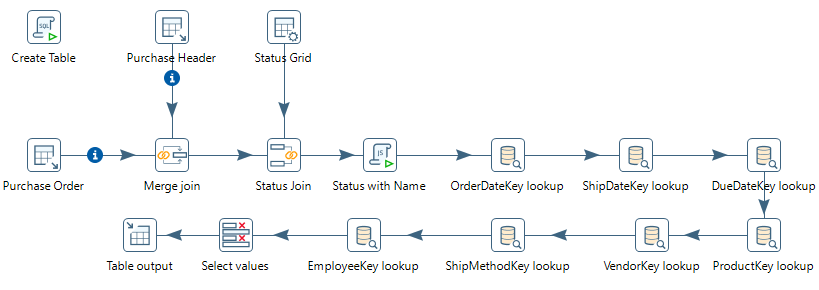


# Tabela de Factos “Compras”

Por último a seguinte Tabela de Factos seleciona campos das tabelas “Purchasing.PurchaseOrderDetail” e “Purchasing.PurchaseOrderHeader” juntando-as tendo por base a chave primária **PurchaseOrderID**.

Não sendo necessário cálculo de dados, são gerados campos novos (**SID, SName**) para transformação do campo existente **Status.**

De seguida são procuradas e angariadas as *Surrogate Keys* (OrderDateKey, ProductKey, etc.) necessárias e por fim são selecionados os dados importantes deixando de fora os auxiliares para a transformação e cálculos.



# Conclusion

Como nota final seria importante referir a importância dos conteúdos abordados neste trabalho prático. Agregação de dados de forma concisa e frequente permite a uma empresa, de qualquer área, analisar o seu desempenho, podendo assim adaptar-se de forma a melhor nas mais diversas áreas no seu negócio.

Por vezes o estudo dos dados é dificultado, ou até mesmo impossibilitado, pela sua fraca qualidade ou por vezes até inexistência. A construção e implementação de um DataMart depende acima de tudo da qualidade dos dados.

A realização deste trabalho prático faz com que sejam aplicados todos os conhecimentos aprendidos em aula e penso que todos os objetivos definidos no enunciado foram cumpridos.

# Appendix A – Data description maps

Table 3: Data description map of Fact Table “Vendas”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Vendas | Tabela Factos | 121317 | | Compras de produtos ao fornecedor por parte da empresa | | | | |
|  |  |  |  |  |  |  |  |  |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| SalesOrderID | Id da venda | Int | -- | SalesOrderHeader | SalesOrderID | Int |  | 1,2,3,4,5 |
| SalesOrderDetailID | Id de uma linha na venda | Int | -- | SalesOrderDetail | SalesOrderDetailID | Int |  | 1,2,3,4,5 |
| ProductKey | Chave de produto | Int | -- | SalesOrderDetail | ProductID | Int | Surrogate Key | 1,2,3,4,5 |
| TerritoryKey | Chave do território | Int | -- | SalesOrderHeader | TerritoryID | Int | Surrogate Key | 1,2,3,4,5 |
| OrderDateKey | Chave da data de encomenda | Int | -- | SalesOrderHeader | OrderDate | Datetime | Surrogate Key | 2018-05-31 |
| ShipDateKey | Chave da data de envio | Int | -- | SalesOrderHeader | ShipDate | Datetime | Surrogate Key | 2018-05-31 |
| DueDateKey | Chave da data prevista de entrega | Int | -- | SalesOrderHeader | DueDate | Datetime | Surrogate Key | 2018-05-31 |
| CustomerKey | Chave do Cliente | Int | -- | SalesOrderHeader | CustomerID | Int | Surrogate Key | 1,2,3,4,5 |
| ShipMethodKey | Chave da empresa de envio | Int | -- | SalesOrderHeader | ShipMethodID | Int | Surrogate Key | 1,2,3,4,5 |
| EmployeeKey | Chave do Funcionário | Int | -- | SalesOrderHeader | SalesPersonID | Int | Surrogate Key | 1,2,3,4,5 |
| Status | Estado da encomenda | Varchar(100) | 2 | SalesOrderHeader | Status | tinyInt | 1 = In process  2 = Approved  3 = Backordered  4 = Rejected  5 = Shipped  6 = Cancelled | 1,2,3,4,5,6 |
| OnlineOrderFlag | Caso seja uma encomenda online | Varchar(100) | 2 | SalesOrderHeader | OnlineOrderFlag | Flag(bit) | 1 – Online  0 – Directly | 1, 0 |
| OrderQty | Quantidade de um produto | int | 2 | SalesOrderDetail | OrderQty | Int |  | 1,2,3,4,5 |
| Freight | Custo de transporte total | Decimal | 2 | SalesOrderHeader | Freight | Money |  | 38.8276 |
| TaxAmt | Imposto totais | Decimal | 2 | SalesOrderHeader | TaxAmt | Money |  | 40.2681 |
| UnitPrice | Preço por unidade da linha | Decimal | 2 | SalesOrderDetail | UnitPrice | Money |  | 2024.994 |
| UnitPriceDiscount | Desconto por unidade da linha | Decimal | 2 | SalesOrderDetail | UnitPriceDiscount | Money |  | 0.10 |
| LineTotal | Total da linha | Decimal | 2 | SalesOrderDetail | LineTotal | Money |  | 2039.994000 |
| SubTotal | Total da encomenda | Decimal | 2 | SalesOrderHeader | SubTotal | Money |  | 20565.6206 |
| TotalDue | Total + taxas + transporte | Decimal | 2 | SalesOrderHeader | TotalDue | Money |  | 23153.2339 |

Table 4: Data description map of Fact Table “Compras”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Compras | Tabela Factos | 8845 | | Venda de produtos ao Cliente por parte da empresa | | | | |
|  |  |  |  |  |  |  |  |  |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| PurchaseOrderID | Id da compra | Int | -- | PurchaseOrderHeader | PurchaseOrderID | Int |  | 1,2,3,4,5 |
| PurchaseOrderDetailID | Id de uma linha na compra | Int | -- | PurchaseOrderDetail | PurchaseOrderDetailID | Int |  | 1,2,3,4,5 |
| EmployeeKey | Chave do Funcionário | Int | -- | PurchaseOrderHeader | EmployeeID | Int | Surrogate Key | 1,2,3,4,5 |
| VendorKey | Chave do Fornecedor | Int | -- | PurchaseOrderHeader | VendorID | Int | Surrogate Key | 1,2,3,4,5 |
| ProductKey | Chave de produto | Int | -- | PurchaseOrderDetail | ProductID | Int | Surrogate Key | 1,2,3,4,5 |
| ShipMethodKey | Chave da empresa de envio | Int | -- | PurchaseOrderHeader | ShipMethodID | Int | Surrogate Key | 1,2,3,4,5 |
| OrderDateKey | Chave da data de encomenda | Int | -- | PurchaseOrderHeader | OrderDate | Datetime | Surrogate Key | 2018-05-31 |
| ShipDateKey | Chave da data de envio | Int | -- | PurchaseOrderHeader | ShipDate | Datetime | Surrogate Key | 2018-05-31 |
| DueDateKey | Chave da data prevista de entrega | Int | -- | PurchaseOrderDetail | DueDate | Datetime | Surrogate Key | 2018-05-31 |
| UnitPrice | Preço por unidade da linha | Decimal | 2 | PurchaseOrderDetail | UnitPrice | Money |  | 2024.994 |
| LineTotal | Total da linha | Decimal | 2 | PurchaseOrderDetail | LineTotal | Money |  | 2039.994000 |
| SubTotal | Total da encomenda | Decimal | 2 | PurchaseOrderHeader | SubTotal | Money |  | 20565.6206 |
| TaxAmt | Imposto totais | Decimal | 2 | PurchaseOrderHeader | TaxAmt | Money |  | 40.2681 |
| Freight | Custo de transporte total | Decimal | 2 | PurchaseOrderHeader | Freight | Money |  | 38.8276 |
| TotalDue | Total + taxas + transporte | Decimal | 2 | PurchaseOrderHeader | TotalDue | Money |  | 23153.2339 |
| StatusName | Estado da encomenda | Varchar(100) | 2 | PurchaseOrderHeader | Status | tinyInt | 1 = In process  2 = Approved  3 = Backordered  4 = Rejected  5 = Shipped  6 = Cancelled | 1,2,3,4,5,6 |
| ReceivedQty | Quantidade do produto recebida | Int | 2 | PurchaseOrderDetail | ReceivedQty | Decimal |  | 550.00 |
| RejectedQty | Quantidade do produto rejeitado | Int | 2 | PurchaseOrderDetail | RejectedQty | Decimal |  | 550.00 |
| StockedQty | Quantidade do produto em stock | Int | 2 | PurchaseOrderDetail | StockedQty | Decimal |  | 550.00 |

Table 5: Data description map of the dimension “Customer”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Customer | Dimension | 19820 | | Clientes que fizeram compras na empresa, podem ser individuais ou lojas. | | | | |
|  |  |  |  |  |  |  |  |  |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| CustomerKey | Chave do Cliente | Int | -- |  |  |  | Surrogate Key | 1,2,3,4,5 |
| CsutomerID | Id de cliente, não diferencia se pessoa ou store | Int | -- | Customer | CustomerID | Int |  | 1,2,3,4,5 |
| PersonID | Id de pessoa, nulo se for loja | Int | 1 | Customer | PersonID | Int |  | 1,2,3,4,5 |
| StoreID | Id de loja, nulo se for pessoa | Int | 1 | Customer | StoreID | Int |  | 1,2,3,4,5 |
| PersonType | Identificador do tipo de pessoa (loja ou individual) | Varchar(100) | 1 | Person | PersonType | nchar | SC – *Store Contact*  IN – *Individual (retail) customer*  SP – *Sales Person*  EM – *Employee (non-sales)*  VC – *Vendor contact*  GC – *General contact* | SC, IN, SP, EM, VC, GC |
| CustomerName | Nome do cliente, nulo se for loja | Varchar(100) | 2 | Person | Title, FirstName, MiddleName, LastName, Suffix | Name | Title + FirstName + LastName + Suffix | Ms. Jo J. Zimmerman |
| StoreName | Nome da loja, nulo se for cliente | Varchar(100) | 1 | Store | Name | Name |  | Next-Door Bike Store |
| Version | Versão dos dados | Int | -- |  |  |  |  | 1 |
| Date\_from | Data a partir do qual os dados são válidos | datetime | -- |  |  |  |  | 1800-01-01 00:00:00:0000 |
| Date\_to | Data até quando os dados são validos | datetime | -- |  |  |  |  | 2500-01-01 00:00:00:0000 |

Table 5: Data description map of the dimension “Vendor”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Vendor | Dimension | 104 | | Fornecedores aos quais a empresa fez encomendas | | | | |
|  |  |  |  |  |  |  |  |  |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| VendorKey | Chave do Fornecedor | Int | -- |  |  |  | Surrogate Key | 1,2,3,4,5 |
| VendorID | Identificador do Fornecedor | Int | -- | Vendor | BusinessEntityID | Int |  | 1,2,3,4,5 |
| Name | Nome do fornecedor | Varchar(100) | 2 | Vendor | Name | Name |  | *Australia Bike Retailer* |
| AccountNumber | Número da conta do fornecedor | Varchar(100) | 2 | Vendor | AccountNumber | AccountNumber |  | *ADATUM0001* |
| CreditRating | Avaliação do fornecedor | Varchar(100) | 1 | Vendor | CreditRating | tinyint | 1 – *Superior*  2 – *Excellent*  3 – *Above average*  4 – *Average*  5 – *Below average* | *1,2,3,4,5* |
| ActiveFlag | Identificador de estado do fornecedor | Varchar(100) | 1 | Vendor | ActiveFlag | Flag(bit) | 0 – not active  1 - active | *0, 1* |
| Version | Versão dos dados | Int | -- |  |  |  |  | 1 |
| Date\_from | Data a partir do qual os dados são válidos | datetime | -- |  |  |  |  | 1800-01-01 00:00:00:0000 |
| Date\_to | Data até quando os dados são validos | datetime | -- |  |  |  |  | 2500-01-01 00:00:00:0000 |

Table 6: Data description map of the dimension “Employee”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Employee | Dimension | 290 | | Funcionários que trabalham para a empresa | | | | |
|  |  |  |  |  |  |  |  |  |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| EmployeeKey | Chave do Funcionário | Int | -- |  |  |  | Surrogate Key | 1,2,3,4,5 |
| EmployeeID | Identificador do Funcionário | Int | -- | Employee | BusinessEntityID | Int |  | 1,2,3,4,5 |
| EmployeeName | Nome do Funcionário | Varchar(100) | 1 | Person | Title, FirstName, MiddleName, LastName, Suffix | Name | Title + FirstName + LastName + Suffix | Ms. Jo J. Zimmerman |
| Gender | Género do Funcionário | Varchar(100) | 2 | Employee | Gender | nchar | M – Male  F - Female | M, F |
| JobTitle | Nome da função do funcionário | Varchar(100) | 2 | Employee | JobTitle | nvarchar |  | Chief Executive Officer |
| Version | Versão dos dados | Int | -- |  |  |  |  | 1 |
| Date\_from | Data a partir do qual os dados são válidos | datetime | -- |  |  |  |  | 1800-01-01 00:00:00:0000 |
| Date\_to | Data até quando os dados são validos | datetime | -- |  |  |  |  | 2500-01-01 00:00:00:0000 |

Table 6: Data description map of the dimension “ShipMethod”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| ShipMethod | Dimension | 5 | | Tabela dos métodos de transporte | | | | |
|  |  |  |  |  |  |  |  |  |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| ShipMethodKey | Chave do Método de Transporte | Int | -- |  |  |  | Surrogate Key | 1,2,3,4,5 |
| ShipMethodID | Identificador do Fornecedor | Int | -- | ShipMethod | ShipMethodID | Int |  | 1,2,3,4,5 |
| Name | Nome da transportadora | Varchar(100) | 2 | ShipMethod | Name | Name |  | CARGO TRANSPORT 5 |
| Version | Versão dos dados | Int | -- |  |  |  |  | 1 |
| Date\_from | Data a partir do qual os dados são válidos | datetime | -- |  |  |  |  | 1800-01-01 00:00:00:0000 |
| Date\_to | Data até quando os dados são validos | datetime | -- |  |  |  |  | 2500-01-01 00:00:00:0000 |

Table 7: Data description map of the dimension “Product”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Product | Dimension | 504 | | Tabela de produtos utilizados em encomendas e vendas | | | | |
|  |  |  |  |  |  |  |  |  |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| ProductKey | Chave do Produto | Int | -- |  |  |  | Surrogate Key | 1,2,3,4,5 |
| ProductID | Identificador do Produto | Int | -- | Product | ProductID | Int |  | 1,2,3,4,5 |
| ProductSubcategoryID | Identificador da Subcategoria | Int | 2 | Product | ProductSubcategoryID | Int |  | 1,2,3,4,5 |
| Name | Nome do Produto | Varchar(100) | 2 | Product | Name | Name |  | Chain |
| ProductNumber | Número do Produto | Varchar(100) | 2 | Product | ProductNumber | nvarchar |  | BE-2908 |
| MakeFlag | Indica se foi feito na empresa | Varchar(100) | 2 | Product | MakeFlag | Flag(bit) | *0 – Purchased*  *1 – Manufactured* | 0,1 |
| Size | Tamanho do produto | Varchar(100) | 2 | Product | Size | nvarchar |  | 1,2,3 |
| Weight | Peso do produto | decimal | 2 | Product | Weight | decimal |  | 1,2,3 |
| Version | Versão dos dados | Int | -- |  |  |  |  | 1 |
| Date\_from | Data a partir do qual os dados são válidos | datetime | -- |  |  |  |  | 1800-01-01 00:00:00:0000 |
| Date\_to | Data até quando os dados são validos | datetime | -- |  |  |  |  | 2500-01-01 00:00:00:0000 |

Table 8: Data description map of the dimension “SalesTerritory”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| SalesTerritory | Dimension | 10 | | Tabela de territórios onde foram efetuadas vendas | | | | |
|  |  |  |  |  |  |  |  |  |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| TerritoryKey | Chave do Território | Int | -- |  |  |  | Surrogate Key | 1,2,3,4,5 |
| TerritoryID | Identificador do Território | Int | -- | SalesTerritory | TerritoryID | Int |  | 1,2,3,4,5 |
| Name | Nome do território | Varchar(100) | 2 | SalesTerritory | Name | Name |  | Southwest |
| CountryRegionCode | Código do país | Varchar(3) | 2 | SalesTerritory | CountryRegionCode | nvarchar |  | US |