Data Mart Implementation (P01)

DECISION SUPPORT SYSTEMS, 2022-23

**João Apresentação (21152), Gonçalo Cunha (21145), Pedro Simões (21140)**

# Introdução

Este projeto de uma forma resumida terá a implementação de um Data Mart com recurso a uma base de dados que armazena atividades de uma empresa de importação e distribuição de produtos, no caso em estudo, Wide World Importers (WWI).

A empresa WWI é uma empresa de importação e distribuição de produtos, que atende principalmente a empresas que revendem para consumidores individuais (público geral). Os clientes da WWI incluem lojas de especialidades, supermercados, lojas de informática, lojas de atrações turísticas e alguns indivíduos.

O projeto passará pela análise de requisitos de negócios, montagem de dimensões e tabelas de factos e implementação de processos ETL.

Nesta introdução, será falada a visão geral do projeto, os objetivos e uma descrição dos processos de negócios envolvidos.

## Visão

O objetivo principal deste projeto é criar um Data Mart eficiente e preciso que possa ser usado para análises de negócios na empresa WWI.

O Data Mart será projetado para fornecer as respostas necessárias e relevantes para os utilizadores, permitindo que estes tomem decisões com base nessas informações.

## Objetivos

Os objetivos deste projeto passam por:

* Realizar a análise de requisitos de negócios para entender as necessidades da empresa WWI;
* Realizar a análise de perfil de dados para entender a qualidade e a integridade dos dados disponíveis;
* Projetar e implementar processos de extração, transformação e carregamento dos dados para a Data Mart;
* Projetar e implementar um modelo dimensional que atenda aos requisitos de negócios da WWI;
* Executar os processos ETL para garantir a integridade dos dados no Data Mart.

## Descrição de Processos de Negócio

Os processos de negócios da WWI inicia-se pelo pedido de produtos, da parte do cliente. Se a WWI não tiver stock suficiente, eles encomendam o stock adicional dos fornecedores. Se os clientes não quiserem esperar pelos produtos que não estão disponíveis, os produtos são enviados posteriormente em um envio separado.

A WWI rentabiliza através dos produtos em stock, convertendo o pedido em fatura. Quando os clientes fazem pedidos de produtos que não estão em stock, esses produtos são colocados em backorder. A WWI entrega os produtos em stock aos clientes, seja por meio de seus próprios veículos de entrega ou por meio de outros correios ou métodos de frete.

Os clientes pagam as faturas à WWI. As faturas e pagamentos são registados na tabela de transações de clientes.

# Recurso de dadosUma imagem com diagrama Descrição gerada automaticamente

Table 1: Summary of WWI database contents

|  |  |  |
| --- | --- | --- |
| **Event / object** | **Table** | **Nr. Records** |
| Clientes | *Customers* | 663 |
| Encomendas feitas | *Orders* | 73595 |
| Registo de Faturas | *Invoices* | 70510 |
| Transações de dinheiro entre WWI e Costumers | *Customertransactions* | 97147 |
| Inventário de produtos | *Stockitems* | 227 |
| Lista de descontos | *Specialdeals* | 2 |

# Modelação Dimensional

<< e. g.: Stores with the highest sales growth? Which products topped sales in the last month?

Then, for each process, identify the metrics and the corresponding measures to include in the fact table.>>

* Preencher matriz
* Identificar metricas e medidas para incluir na fact table

Table 2: Data Warehouse Matrix

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DIMENSIONS**  **BUSINESS PROCESSES** | Dim\_costumers | Dim\_stockitems | Dim\_costumertransactions | Ft\_orderlines | Ft\_invoicelines |  |  |  |  |  |  |
| Process 1 | X |  | X |  |  |  |  |  |  |  |  |
| Process 2 |  | X | X |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

## Design do modelo de dados dimensional

Abaixo está representado, em tabelas, informações sobre os atributos mais importantes de cada tabela

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| dim\_costumers | Dimension | 664 | | Dimensão que apresenta os dados pessoais de um cliente | | | | |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| Costumerkey |  | Int |  |  |  |  |  |  |
| Costumerid |  | Int |  |  |  |  |  |  |
| Costumername |  | String |  |  |  |  |  |  |
| Cityname |  | String |  |  |  |  |  |  |
| Formalname |  | String |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| dim\_stockitems | Dimension | 228 | | Dimensão que apresenta os dados sobre os items em stock e suas características | | | | |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| Stockitemkey |  | Int |  |  |  |  |  |  |
| Stockitemid |  | Int |  |  |  |  |  |  |
| Stockitemname |  | String |  |  |  |  |  |  |
| Colorname |  | String |  |  |  |  |  |  |
| Unitprice |  | Double |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| dim\_customertransaction | Dimension | 26638 | | Dimensão que apresenta todas as transações que ocorreram no WWI | | | | |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| Customertransactionkey |  | Int |  |  |  |  |  |  |
| Customertransactionid |  | Int |  |  |  |  |  |  |
| Costumerid |  | Int |  |  |  |  |  |  |
| Transactiontypename |  | String |  |  |  |  |  |  |
| Invoiceid |  | Int |  |  |  |  |  |  |
| Transactiondate |  | Datetime |  |  |  |  |  |  |
| Transactionamount |  | Double |  |  |  |  |  |  |
| Isfinalized |  | Bool |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Ft\_orderlines | Fact Table | 694236 | | Tabela de factos representante de cada linha de cada encomenda | | | | |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| Orderlineid |  | Int |  |  |  |  |  |  |
| Orderid |  | Int |  |  |  |  |  |  |
| Costumerkey |  | Int |  |  |  |  |  |  |
| Stockitemkey |  | Int |  |  |  |  |  |  |
| Orderdate |  | Datetime |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Ft\_invoicelines | Fact Table | 1141287 | | Tabela de factos representante de cada linha de cada fatura | | | | |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
| Invoicelineid |  | Int |  |  |  |  |  |  |
| Invoiceid |  | Int |  |  |  |  |  |  |
| Costumerkey |  | Int |  |  |  |  |  |  |
| Extendedprice |  | Double |  |  |  |  |  |  |
| Lineprofit |  | Double |  |  |  |  |  |  |
| Invoicedate |  | Datetime |  |  |  |  |  |  |

# Implementação da DataMart

* Explicar Kettle

# Conclusão

* Revisão crítica do trabalho feito
* Pontos fortes e fracos
* Se possível, o que se mudaria/adicionaria