Data Mart Implementation (P01)

DECISION SUPPORT SYSTEMS, 2022-23

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

<< The goal of this project is to implement a data mart based on a transactional data source. The team may use a different data source >>

# Introdução

Este projeto de uma forma resumida terá uma 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).

O projeto passará pela análise de requisitos de negócios, montagem de dimensões e tabelas de factos, implementação de processos ETL e documentação das transformações. Esta implementação foi realizada com o uso de servidor PostgreSQL.

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 nesses dados 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;
* Documentar as transformações ETL;
* Executar os trabalhos ETL para garantir a integridade dos dados no Data Mart.

## Descrição de Processos de Negócio

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.

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 Dados

* Diagrama da database do stor
* Preencher tabela tendo em conta a database
* Modelo relacional (?)

Table 1: Summary of WWI database contents

|  |  |  |
| --- | --- | --- |
| **Event / object** | **Table** | **Nr. Records** |
| Rentals | *Rental* | 99 999 |
| Bikes | *…* | … |
| … | *…* | … |

# Dimensional modelling

<< 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** | Dimension 1 | Dimension 2 | Dimension 3 |  |  |  |  |  |  |  |  |
| Process 1 | X |  | X |  |  |  |  |  |  |  |  |
| Process 2 |  | X | X |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

# Design of the dimensional data model

For each table (TF or Dim.), you should complete a data description map (see **Appendix A**)

# Appendix A – Data description maps

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Type of table** | **Nr. Records** | | **Description** | | | | |
| Dim\_A | Dimension | ?? | | ?? | | | | |
| **Target (Data mart)** | | | | **Source (OLTP)** | | | | |
| **Column** | **Description** | **Data type** | **SCD** | **Table** | **Column** | **Data type** | **ETL rules** | **Example of values** |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

Table 3: Data description map of Dim\_A

* Não entendi

# Data mart implementation

* Explicar Kettle

# Conclusion

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