Data Mart Implementation (P01)

DECISION SUPPORT SYSTEMS, 2021-22

**João Fernandes (18825) Carlos Martins (18836)**

<< 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, but this must be validated by the teacher.

This document has an suitable structure for the project, however chapters can be deleted, changed, or added. Keep formatting consistent throughout the document. **Comments**, such as this one, placed between "<<" and ">>", **should be removed**. >>

# Introduction

<<Describe the scope and general objectives of the project and summarize in 2 or 3 paragraphs the business processes supported by the database >>

The scope of this project is to implement a data mart based on an operation database. In this case the database in use is **Adventure Works 2021 database**. This database is from a fictional USA company, that is both a manufacturer and seller of bicycles/accessories.

The company has two main channels: **Internet Sales** and **Retail Sales**.Since we were given the option of choosing one of these two, we chose **Retail Sales**.

# Data sources

<<This section describes the data sources’ structure and content. Insert the relational model of the database and a content overview of the operational database. You can use a data profiling tool to get a quick overview of the data quality. You can use a table (see next) to list the main objects/events.>>

Table 1: Summary of AAA database contents

|  |  |  |
| --- | --- | --- |
| **Event / object** | **Table** | **Nr. Records** |
| E.g., Rentals | *Rental* | 99 999 |
| Films | *…* | … |
| … | *…* | … |

# Dimensional modelling

This database stores multiple information about various business processes like **purchases** and **sales**. Given these processes we defined some questions that the system should be able to answer:

* **Q1** - What was the most profitable month?
* **Q2** – Which store stole the most in a certain location?
* **Q3** - Who is the best salesperson for each product?
* **Q4** - Which store had the biggest sales growth?
* **Q5** - Which products were the most sold per month?
* **Q6** – Who was the client who spend the most money per month?

Table 2: Data Warehouse Matrix

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DIMENSIONS**  **BUSINESS PROCESSES** | Customer | Product | Date | Location | Store | Person | Salesperson |  |  |  |  |
| Most profitable month in a year |  |  | X |  | X |  | X |  |  |  |  |
| Store with the most sales in a certain location |  |  |  | X | X |  |  |  |  |  |  |
| Best salesperson for each product |  | X |  |  |  | X | X |  |  |  |  |
| Store with the biggest sales growth |  |  |  | X | X |  |  |  |  |  |  |
| Most sold products per month |  | X | X | X | X |  |  |  |  |  |  |
| Client who spent the most money | X |  | X |  | X | X |  |  |  |  |  |

1. **Fact table(s) and its granularity**

This database is about a fictional USA company, that is both a manufacturer and seller of bicycles and accessories so this is an indicator that the fact tables should be of transactional type. There are two fact tables. The first one is the junction of **SalesOrderHeader** with **SalesOrderDetail** and this FT contains all the information of a sales process. The other one is the junction of **PurchaseOrderHeader** and **PurchaseOrderDetail** and this contains all the information of a purchase process. Both fact tables are of transactional type and each line represents a sales or purchase order respectively.

**2. Dimensions and attributes relevant to the analysis.**

# Design of the dimensional data model

<<Definition of the fact tables (FT), describing the granularity and justifying the measures to be included. In the case of derived measures, it is necessary to present here how they are obtained.

Develop the relational model using a database modelling tool and include the ER diagram in this report. Rationalize the options that have been taken (dimensions, outriggers, DD, etc.). For each table (TF or Dim.), you should complete a data description map (see Appendix A).>>

# Data mart implementation

<<Describe the ETL process and highlight the most relevant aspects. Include the graphical representation of the integration transformations and jobs. At the end of this section, write the summary of the data mart content, e.g., number of records loaded into each table.>>

# Conclusion

<<In the closing chapter, elaborate a critical review of the work done, pointing out its strengths and weaknesses. In addition, if applicable, please list possible future tasks or new options to deepen the work done.>>

# Bibliography

<< In this section, you must present, in APA format, the list of bibliographic sources consulted during the execution of the work and that were relevant for its execution.>>

# Appendix A – Data description maps

<< Display a map/table for each table included in the data mart. See document DSS22\_ETL\_Process >>

Table 3: Data description map of Dim\_A

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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** |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |