|  |
| --- |
| EPAM Systems, RD Dep. |
| Publishing House "MIF" |



Contents

[1. Business Description 3](#_Toc412572569)

[1.1. Business background 3](#_Toc412572570)

[1.2. Problems because of poor data management 3](#_Toc412572571)

[1.3. Benefits from implementing a Data Warehouse 3](#_Toc412572572)

[2. Dimensions of a Business 3](#_Toc412572573)

[3. Logical Scheme 3](#_Toc412572574)

[4. Data Flow 3](#_Toc412572575)

[5. Fact Table Partitioning Strategy 3](#_Toc412572576)

[6. Strategy of Parallel Load 3](#_Toc412572577)

[7. Report Layouts 3](#_Toc412572578)

# Business Description

## Business background

MIF is one of the most progressive publishing houses in the domestic market right now. They have been working for 12 years already and published over 1928 books. The abbreviation MIF is stands for names of three founders: Igor Mann, Michael Ivanov and Michael Ferebr. The publishing house is specializing on motivational and personal development books for individual readers and companies, but also publishing children’s books, e-books, audiobooks, comics. They pursue active innovative marketing policy and actively invite new authors.

## Problems because of poor data management

Poor data management can lead to a lot of problems, waste corporate money and time and, even more, inflict irreparable damage.

Here a list of problems that may arise:

* Loss of customers because of out-of-date or incorrect data.
* Unsuccessful marketing and advertising campaigns.
* Loss of revenue because of errors, wrong decisions and lost opportunities.
* If company has multiple systems that don’t properly integrate it can lead to multiple versions of the truth.

## Benefits from implementing a Data Warehouse

* Time saving: users can quickly access critical data from a number of sources—all in one place, so they can rapidly make informed decisions on key initiatives.
* Historic data: A large amounts of historical data can be stored so users can analyze different time periods and trends in order to make future predictions.
* Consistent data: The conversion of data from numerous source systems and data files and transformation into a common format, which allows to make versatile analysis.
* Good performance: Data Warheouse is specially designed with a focus on speed of data retrieval and analysis.
* Forecasts: Provides an opportunity to forecast future growth, needs and problems.
* Customer relationships: Manage and enhance customer relationships.
* Marketing campaign: Monitor and modify a marketing campaign.
* Improve efficiency: Increase the efficiency and effectiveness of product management and development.

# Dimensions of a Business

/\* пока тут очень сырой вариан, так как все зависит от данных, которые будут получены для заполнения DWH\*/

## Бизнес-процесс

Бизнес-процесс для издательского дома «МИФ»: учет продаж изданных книг.

## The grain

Что представляет собой строка из таблицы фактов?

Одна строка из таблицы фактов представляет собой одну книгу, которую работник продал потребителю в конкретном магазине в какой-либо лень.

## Dimensions

Хранилище данных будет включать следующие измерения:

* Магазины
* Работники
* Клиенты
* Даты
* Печатные книги
* Аудиокниги
* Электронные книги
* Наши авторы
* Валюта (тк не все цены в бел рублях)

DimStores:

StoreID, StoreAlternateKey Manager, City, Subway\_station, Address

DimEmployess:

Employee\_id, First\_name, Last\_name, Gender,Department, Phone, Email

DimCustomers:

CustomerID, First\_name, Last\_name, Gender, Address, Phone, Email

DimDate:

Date\_key, Full\_date, Year, Quarter, Semester, Month, Month\_name, Day\_of\_month, Day\_of\_year, Day\_name

DimPrint

Print\_id, Name, Author, Genre, Publish\_year, Language, Description

DimAudio

Audio\_id, Name, Author, Genre, Narrator, Language, Description

DimEbook

Ebook\_id, Name, Author, Genre, Format(?), Size,Language, Description

DimOurAuth

Auth\_id, First\_name, Last\_name, Age, Gender, Phone, Email

DimCurrency

Cur\_key, Alternate\_key, Value

## Fact table

Фактовая таблица:

* Продажи

FactSales:

Sale\_id, Store\_id, Employee\_id, Customer\_id, Date\_key, Type\_of\_book, Book\_id, Auth\_id, Price, Amount

# Logical Scheme

# Data Flow

# Fact Table Partitioning Strategy

# Strategy of Parallel Load

# Report Layouts