|  |
| --- |
| EPAM Systems, RD Dep. |
| DWH |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| История изменений | | | | | |
| Версия | Изменения | Автор | Дата | Рецензировано | |
| Рецензент | Дата |
| 0.1 | Initial status | Gleb\_Kulvanovski | 15-Nov-2017 |  |  |
|  |  |  |  |  |  |

Содержание

[1. Business Description (Описание предметной области) 3](#_Toc498529088)

[1.1. Business background (Миссия проекта / Бизнесс основа) 3](#_Toc498529089)

[1.2. Problems because of poor data management / Недостатки существующих решений 3](#_Toc498529090)

[1.3. Benefits from implementing a Data Warehouse / Преимущества от внедрения разрабоываемого хранилища данных 3](#_Toc498529091)

[2. Dimensions of a Business 3](#_Toc498529092)

[2.1. Выбор зерна 3](#_Toc498529093)

[2.2. Выделение контекста 3](#_Toc498529094)

[2.3. Описание факта 3](#_Toc498529095)

[3. Logical Scheme 3](#_Toc498529096)

[4. Data Flow 3](#_Toc498529097)

[5. Fact Table Partitioning Strategy 3](#_Toc498529098)

[6. Strategy of Parallel Load 3](#_Toc498529099)

[7. Report Layouts 3](#_Toc498529100)

# Business Description (Описание предметной области)

## Business background (Миссия проекта / Бизнесс основа)

Сеть магазинов по продаже одежды и обуви.

## Problems because of poor data management / Недостатки существующих решений

Отсуттсвие актуальных данных о продажах и рентабельности компании.

## Benefits from implementing a Data Warehouse / Преимущества от внедрения разрабоываемого хранилища данных

Составление отчетов по продажаам, региона, улучшение логистики компании. Отчёты будут строиться для отслеживания продаж, оборота и прибыли магазина в зависимости от даты(по месяцам, возможно по дням недели и времени(утро, день, вечер)).

# Dimensions of a Business

## Выбор зерна

Зерном данного DWH явзлется прожажи вещей для каждого магазина, города, страны за определённый период.

## Выделение контекста

Измерение **Date**:

– значение даты (естественный первичный ключ),

– год, квартал (номер в году, номер вмете с годом), месяц (номер и название на разных языках), номер недели в году, день в месяце,

– день недели (номер и название на разных языках).

Измерение **Customers**:

-- id\_c первичный ключ,

-- id\_cust уникальный id клиена,

-- имя, фамилия,

-- адресс (улица, домашний телефон, почтовый индекс и т.п.),

-- телефон(мобильный телефон),

-- e-mail,

-- location(город, страна, континент),

Измерение **Stores**:

--id\_stores первичный ключ,

--address

-- Phone

-- Postcode

-- City

-- Country

-- Region

Измерение **Employees**:

-- id\_e первичный ключ

-- id\_emp уникальный id сотрудника

-- имя, фамилия

-- Phone мобильный телефон

-- Address (улица, дом.тел)

-- Location (город, страна)

-- Jobs (должность)

Измерение **Product**:

-- id\_p первичный ключ,

-- id\_prod уникальный id товара,

-- prod\_name имя товара,

-- prod\_desc описание товара,

-- prod\_sale цена товара,

-- subcategory\_id ключ подкатегории,

-- subcategory\_name имя подкатегории,

-- subcategory\_desc описание подкатегории,

-- categore\_id ключ категории,

-- categore\_name имя категории,

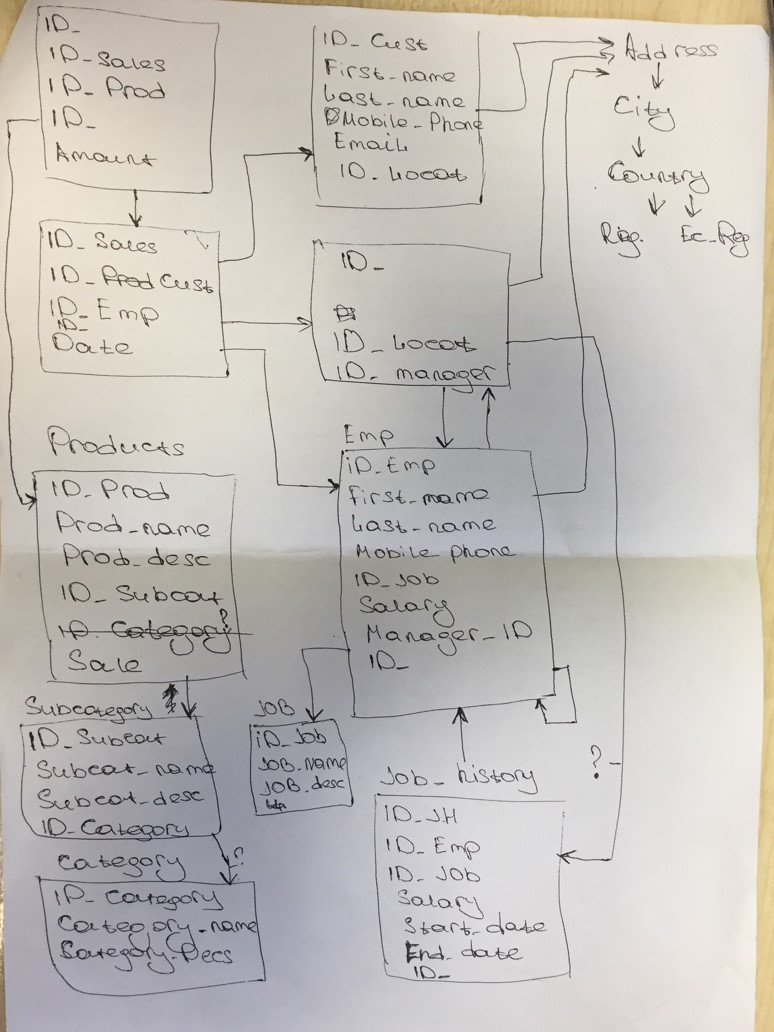
-- categore\_desc описание категории

Описание факта

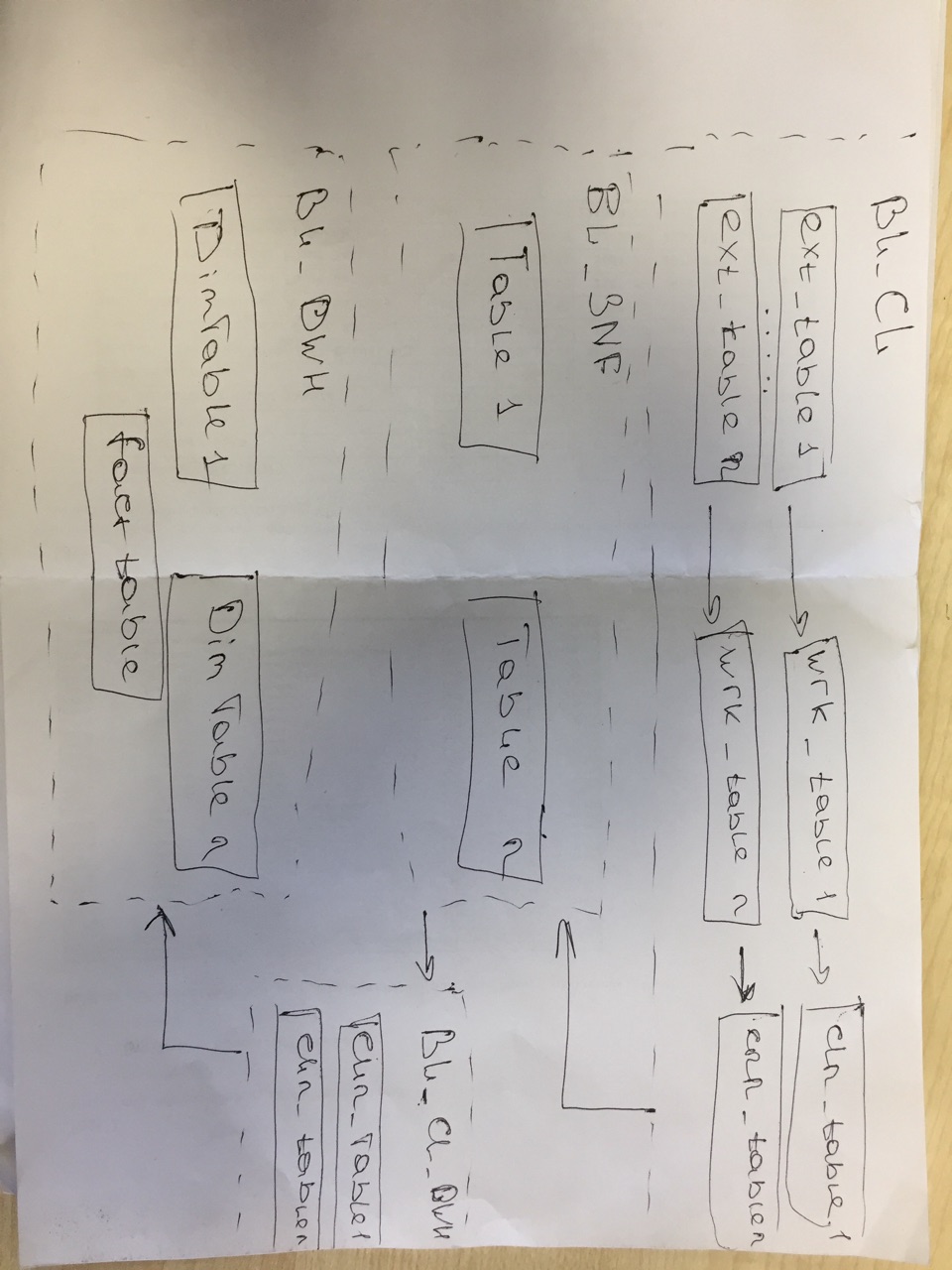
Фактовая таблица содержит id таблиц измерений, а так же сумму общую сумму платежа(Sales) и время платежа(Date)

# Logical Scheme

## Схема слоя BL\_3NF



# Data Flow



# Fact Table Partitioning Strategy

Разбиение таблицы фактов на партиции будет произведенно по датам. Это необходимо для того, что бы более эффективно строить отчёты и для более быстрого заполнения таблцы данными т.к. при партиционировании индексы могут быть не на всю таблицу, а на партиции и из-за этого обновление индекса происходит быстрее чем при обновлении индекса на всю таблицу.

# Strategy of Parallel Load

# Report Layouts