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

Image goes here.

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

В настоящий время многие операции совершаются онлайн, то есть в режиме реального времени. Это справедливо и для бронирования номеров в отелях. Из любой точки мира можно забронировать номер в любой стране. На основе наиболее популярных сайтов будет построено хранилище данных, в котором будет информация по каждому заказу, включающая в себя данные заказчика, ресурс и детали заказа.

## Problems because of poor data management

* Отсутствие единого общего хранилища для разных сайтов.
* Затруднения анализа данных, а как следствие низкая чувствительность к изменениям на рынке.
* Потеря клиентов из-за отсутствия стимулирования правильного сегмента потребителей.

## Benefits from implementing a Data Warehouse

* Вовремя отследить изменения и тенденции спроса на рынке поможет оставаться в тендре долгое время.
* Можно отслеживть сегменты, приносящие минимальную прибыль, и проводить политику по укреплению их на рынке.
* Возможность отслеживать активность каждого пользователя на определённом временном интервале.
* Используя соответсвующие витрины, репортинг станет проще, доступнее и появятся много возможностей.

# Dimensions of a Business

Let’s identify 4 steps of creating data warehouse:

1. Select the business process

Selected business process is online booking of hotels.

2. Declare the Grain

The grain is one order on site.

3. Identify the Dimensions

1. Hotels;
2. Information sources;
3. Date&Time;
4. Customers;

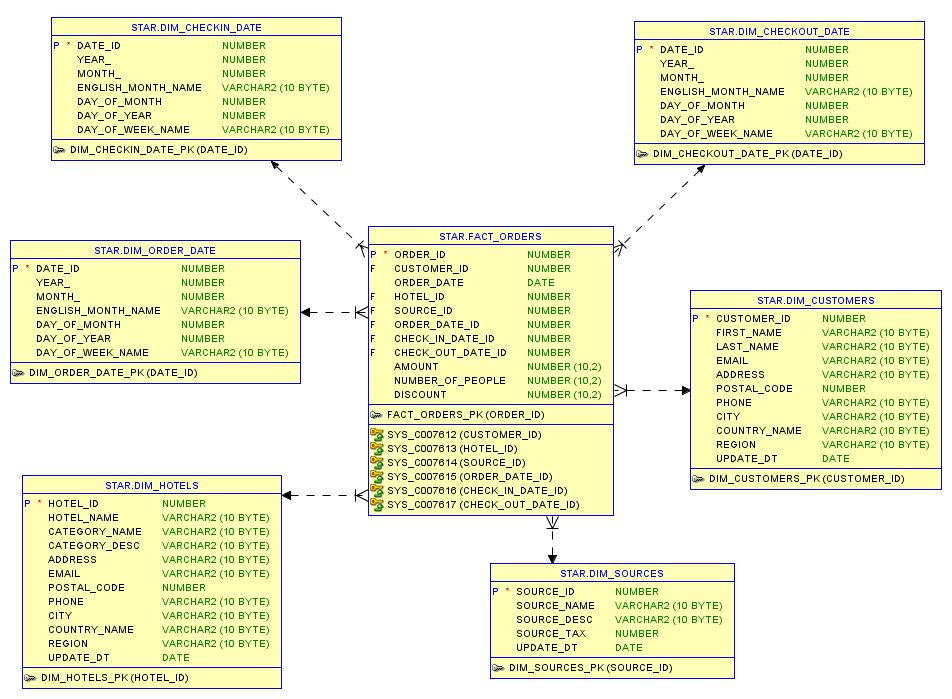
4. Identify the Facts

* The order and its description is a fact.

**Sale**

* Sale of some product to the customer

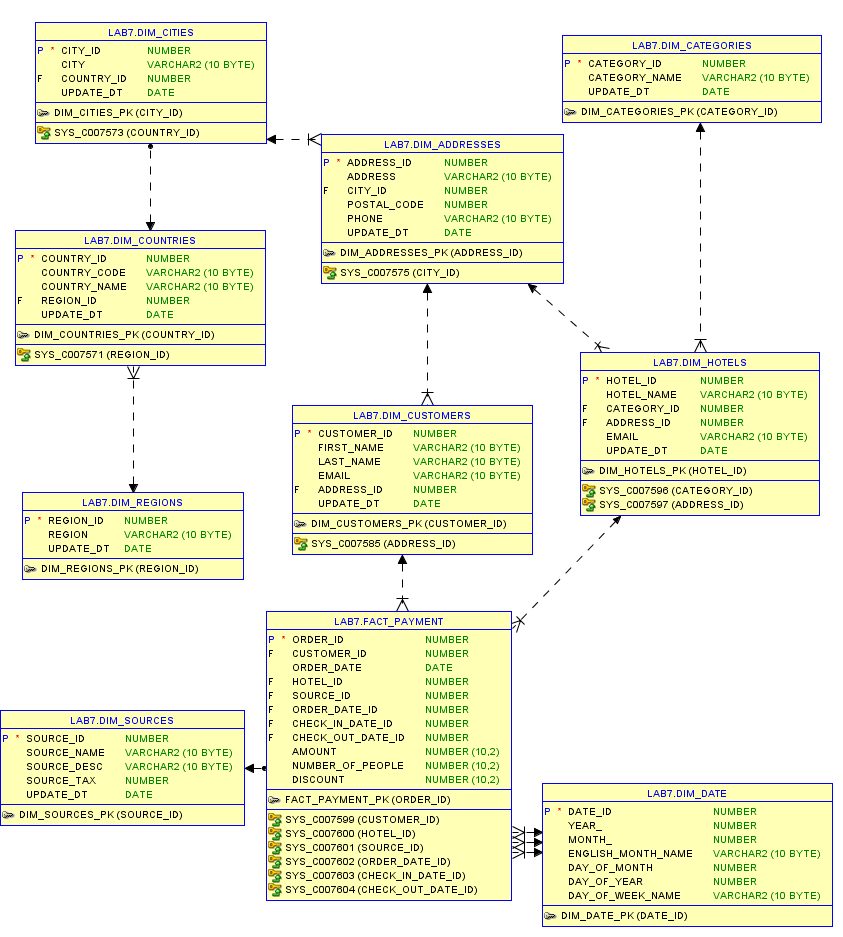
Star schema



|  |  |  |  |
| --- | --- | --- | --- |
| **Dimension** | **Field** | **Description** | **Example** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Fact** | **Field** | **Description** | **Example** |
| **Fact\_orders** | Order\_id |  |  |
| Customer\_id |  |  |
| Order\_date |  |  |
| Hotel\_id |  |  |
| Source\_id |  |  |
| Order\_date\_id |  |  |
| Check\_in\_date\_id |  |  |
|  | Check\_out\_date\_id |  |  |
|  | amount |  |  |
|  | Number\_of\_people |  |  |
|  | discount |  |  |

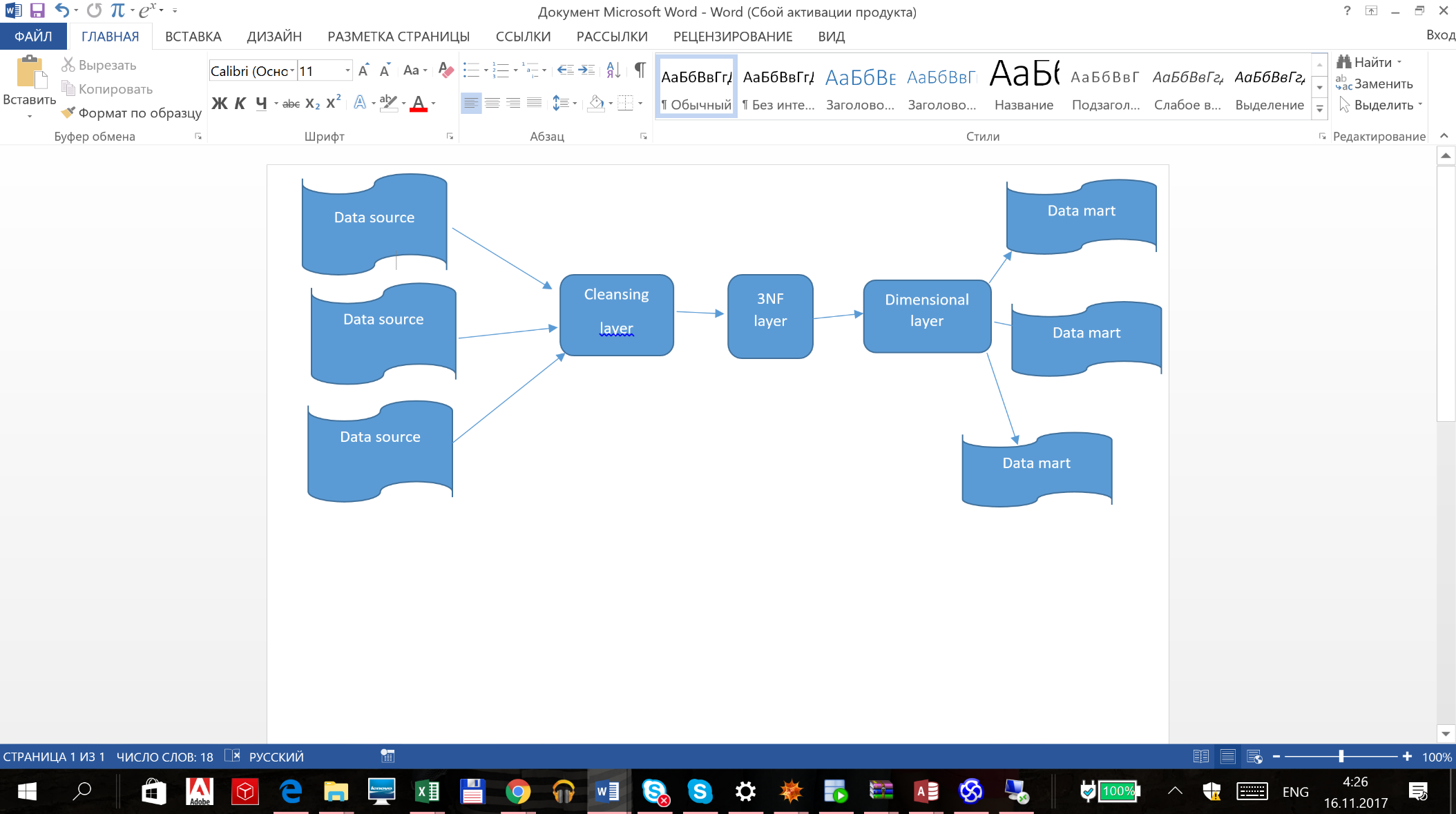
Snowflake schema:



Description of the methods, rules and the process of generation of your source data.

# Logical Scheme

# Data Flow



1. **Source layer.** A data warehouse system uses heterogeneous sources of data. That data is originally stored to corporate relational databases or legacy databases, or it may come from information systems outside the corporate walls.
2. **Cleansing Layer .**The Cleansing Layer is used for data cleansing, filtering wrong data, replace missing values with singletons and performing transformations like code lookups or currency conversions. As the Staging Area, the Cleansing Area contains only data of the last delivery, and data from different sources is not integrated
3. **Third Normal Form layer**

From staging, the data will transition into the foundation or integration layer via another set of ETL processes. Data begins to take shape and it is not uncommon to have some end-user application access data from this layer especially if they are time sensitive, as data will become available here before it is transformed into the dimension / performance layer. Traditionally this layer is implemented in the Third Normal Form (3NF).

1. **Dimensional layer.** The Dimensional layer consists of dimension and fact tables.It is used as a source for creating data marts.
2. **Data Marts.** You may want to customize your warehouse's architecture for different groups within your organization. You can do this by adding data marts, which are systems designed for a particular line of business.

# Fact Table Partitioning Strategy

# Strategy of Parallel Load

# Report Layouts