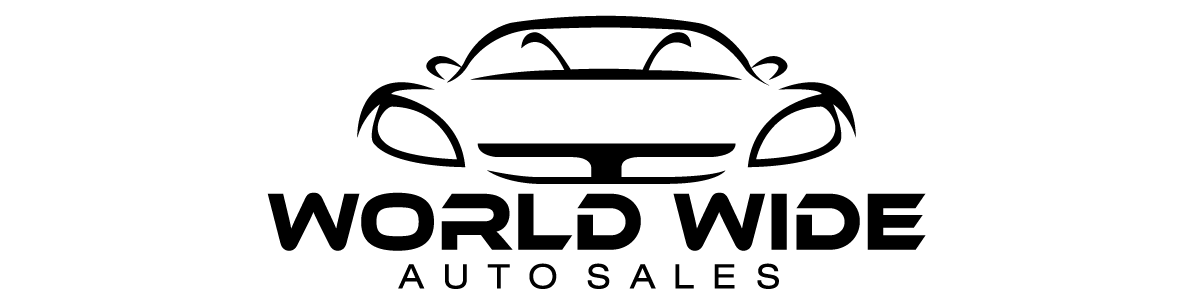
|  |
| --- |
| EPAM Systems, RD Dep. |
| DWH for Car Sales |



Contents

[1. Business Description 3](#_Toc498455020)

[1.1. Business background 3](#_Toc498455021)

[1.2. Problems because of poor data management 3](#_Toc498455022)

[1.3. Benefits from implementing a Data Warehouse 3](#_Toc498455023)

[2. Dimensions of a Business 3](#_Toc498455024)

[3. Dimensional model. Star Schema 5](#_Toc498455025)

[4. 3NF Scheme 5](#_Toc498455026)

[5. Logical Scheme 6](#_Toc498455027)

[6. Data Flow 6](#_Toc498455028)

[7. Fact Table Partitioning Strategy 6](#_Toc498455029)

[8. Strategy of Parallel Load 6](#_Toc498455030)

[9. Report Layouts 6](#_Toc498455031)

# Business Description

## Business background

Nowadays there are many different ways of buying cars. Thousands of cars are sold and bought daily. Moreover, sometimes it is complicated to keep track of all changes in car prices and the popularity of models especially because of the huge amount of different characteristic such as vehicle type, fuel type, and gearbox.

## Problems because of poor data management

The problems that business is facing because of poor data management are that it is so complicated to find all the information about the cars such as price, year of registration, model and kilometer in one place. Usually this information divided into small pieces in different places. In addition, it is hardly to find the information about different types of model in one place.

## Benefits from implementing a Data Warehouse

With the help of data warehouse, it would be easier to find the information about cars, building different types of reports. In addition, it could be possible to see the difference in sales per years or per model, or per price. Completed information about car sold will be in one place.

This DWH would be suitable either for people, who wanted to buy a car, or for that people, who are interested in evaluation of the situation in the car market.

# Dimensions of a Business

The business of the project is a car sale. Thousands of cars are sold and bought daily, so it becomes a problem to keep it in one place. Based on this, the grain of the model is a car for sale.

Fact table will include Seller\_id, Store\_id, Car\_id, Order\_Date\_id, Customer\_id, Order\_code, Cost, Min\_Price, Avg\_Price, SD\_Price.

Dimensions:

1. Seller. Here will be information about seller such as:

|  |  |  |
| --- | --- | --- |
| Seller\_id | Number(8) | PK |
| Seller\_name | Varchar2(200) |  |
| Seller\_surname | Varchar2(200) |  |
| Seller\_rating | Number(8) |  |
| Phone | Varchar2(200) |  |
| Email | Varchar2(200) |  |

1. Store. In Store there is information about the location:

|  |  |  |
| --- | --- | --- |
| Store\_id | Number(8) | PK |
| Store\_name | Varchar2(200) |  |
| Phone | Varchar2(200) |  |
| Email | Varchar2(200) |  |
| Street\_name | varchar2(200) |  |
| House\_number | Varchar2(200) |  |
| City\_name | Varchar2(200) |  |
| Country\_name | Varchar2(200) |  |

1. Cars. Here will be:

|  |  |  |
| --- | --- | --- |
| Car\_id | Number(8) | PK |
| Car\_number | Number(8) |  |
| Car\_name | Varchar2(200) |  |
| Car\_Registration\_Date | Date |  |
| Vehicle\_type\_name | Varchar2(200) |  |
| Engine\_Type\_name | Varchar2(200) |  |
| Gearbox\_type\_name | Varchar2(200) |  |
| Brand\_name | Varchar2(200) |  |
| Model\_name | Varchar2(200) |  |
| Repair\_status\_name | Varchar2(200) |  |
| Power\_PS | Number(8) |  |
| Kilometers | Number(8) |  |

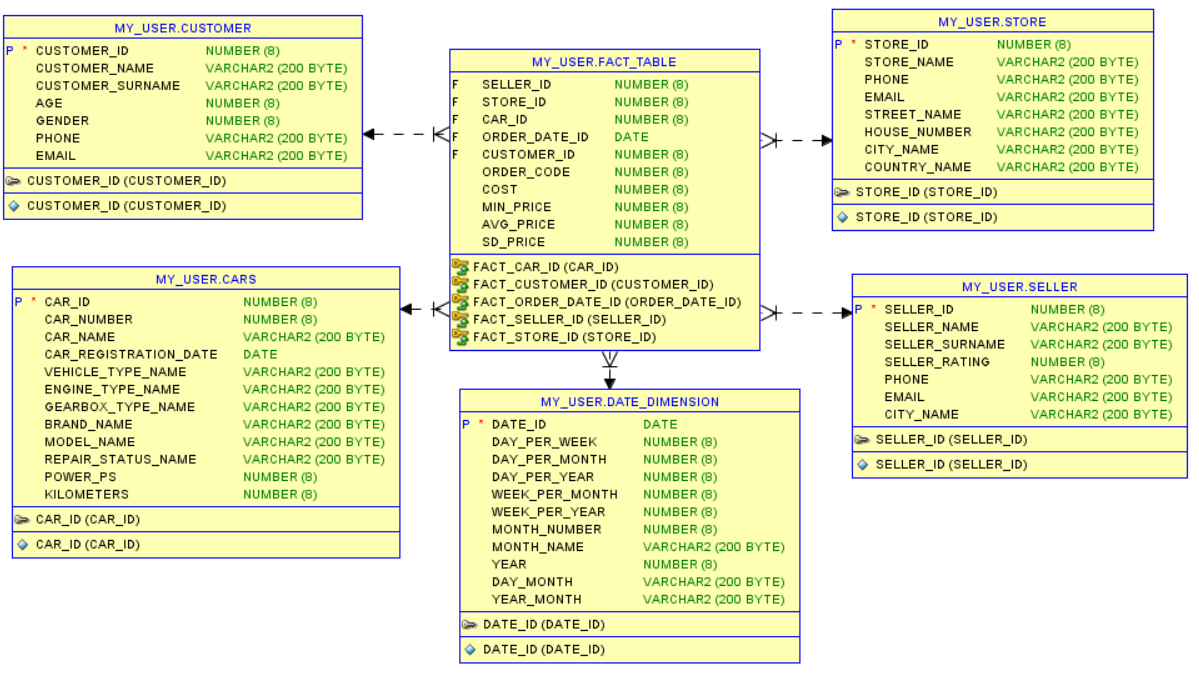
1. Date dimension. Here will be:

|  |  |  |
| --- | --- | --- |
| Date\_id | Date | PK |
| Day\_per\_week | Number(8) |  |
| Day\_per\_month | Number(8) |  |
| Day\_per\_year | Number(8) |  |
| Week\_per\_month | Number(8) |  |
| Week\_per\_year | Number(8) |  |
| Month\_number | Number(8) |  |
| Month\_name | Varchar2(200) |  |
| Year | Number(8) |  |
| Day-Month | Varchar2(200) |  |
| Year-Month | Varchar2(200) |  |

1. Customer. Here will be:

|  |  |  |
| --- | --- | --- |
| Customer\_id | Number(8) | PK |
| Customer\_name | Varchar2(200) |  |
| Customer\_surname | Varchar2(200) |  |
| Age | Number(8) |  |
| Gender | Number(8) |  |

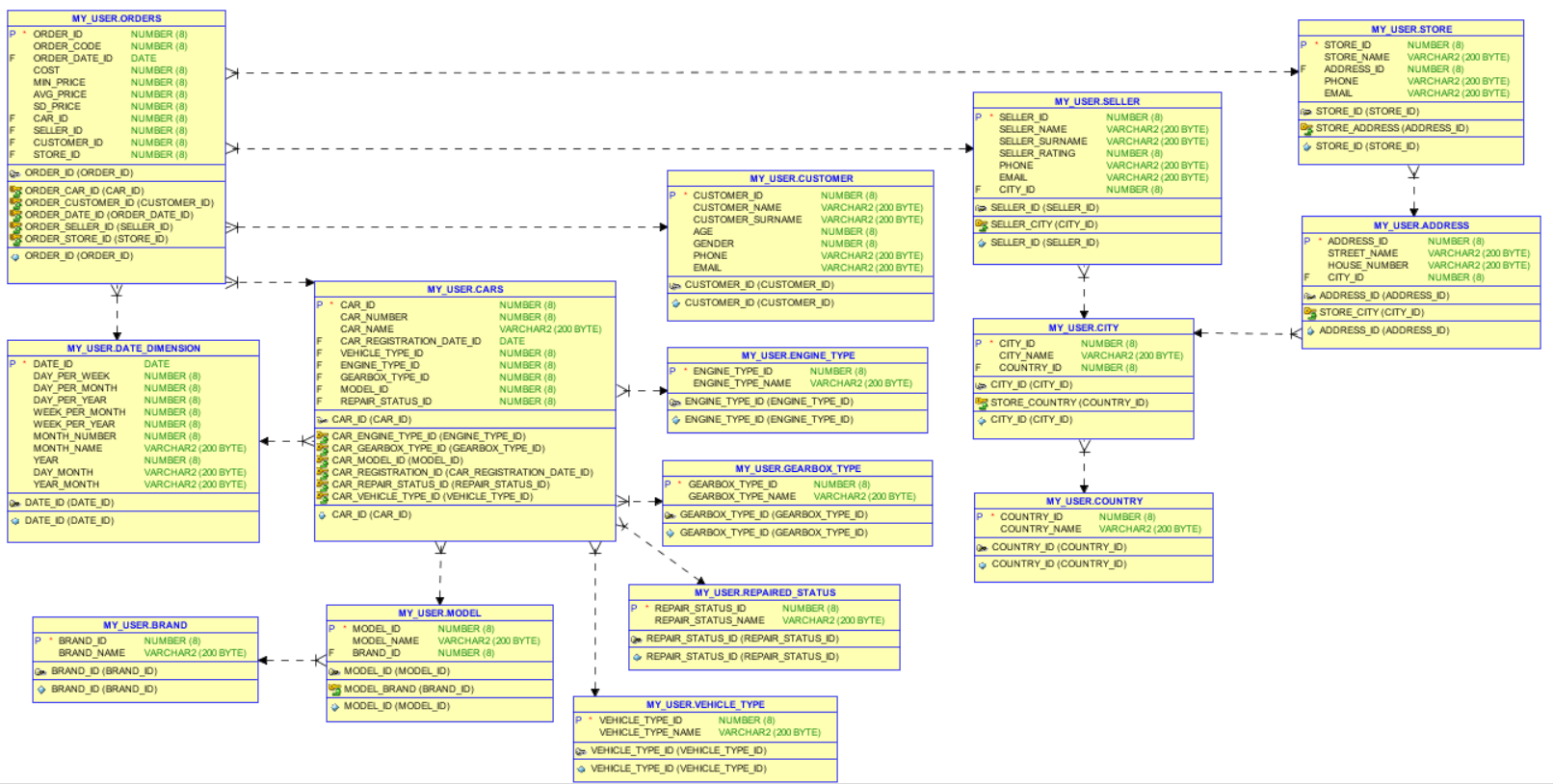
# Dimensional model. Star Schema



Advantages of star schema:

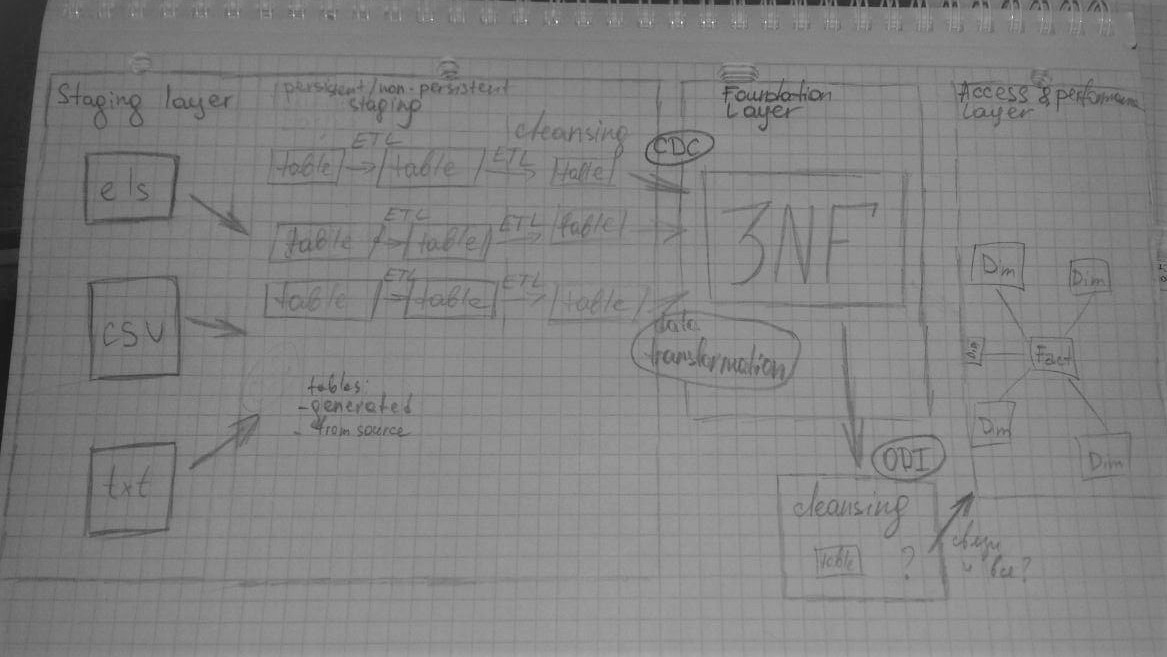
1. Difficult in understanding data structure
2. Higher productivity
3. While changes in one dimension, other will stay in the same conditional.

# 3NF Scheme



# Logical Scheme

# Data Flow



# Fact Table Partitioning Strategy

# Strategy of Parallel Load

# Report Layouts