

|  |
| --- |
| Sales analysis |
|  |

| RELATED ARTIFACTS | |
| --- | --- |
| Ref. | Artifact Name |
|  |  |
|  |  |

|  |  |
| --- | --- |
| Abbreviations and Acronyms | |
|  |  |
|  |  |

Contents

[Milavitsa's sales analysis 1](#_Toc498564835)

[1 Business Description 5](#_Toc498564836)

[1.1 Business background 5](#_Toc498564837)

[1.2 Problems because of poor data management 5](#_Toc498564838)

[1.3 Benefits from implementing a Data Warehouse 6](#_Toc498564839)

[2 Data Description 7](#_Toc498564840)

[2.1 Description of selected schema 7](#_Toc498564841)

[2.2 Description of data sources 7](#_Toc498564842)

[2.3 Description of business rules 7](#_Toc498564843)

[3 Dimensions of a Business 8](#_Toc498564844)

[3.1 Definition of selected Business Process 8](#_Toc498564845)

[3.2 Declaring the Grain 8](#_Toc498564846)

[3.3 Identifying the Dimensions 8](#_Toc498564847)

[3.4 Identifying the Facts 8](#_Toc498564848)

[3.5 Dimensional model. Star Schema 9](#_Toc498564849)

[3.5.1 Fact retail sales 10](#_Toc498564850)

[3.5.2 Fact product balances 11](#_Toc498564851)

[4 NF-layer of Data Warehouse 12](#_Toc498564852)

[4.1 3NF-model 16](#_Toc498564853)

[5 Object Partitioning 20](#_Toc498564854)

[6 Business processes 22](#_Toc498564855)

[6.1 Sales analysis 23](#_Toc498564856)

[6.2 Analysis of receipts 24](#_Toc498564857)

[6.3 Load analysis 25](#_Toc498564858)

[7 Data Modelling 26](#_Toc498564859)

[7.1 Detailing diagrams for 3NF and Star/snowflake layers 26](#_Toc498564860)

[7.2 Textual description of layers of data warehouse 26](#_Toc498564861)

[7.2.1 Source layer 26](#_Toc498564862)

[7.2.2 Data staging layer 27](#_Toc498564863)

[7.2.3 Data warehouse layer 27](#_Toc498564864)

[7.2.4 Analysis layer 27](#_Toc498564865)

[7.3 Visual description of layers of data warehouse 29](#_Toc498564866)

# Business Description

## Business background

For many years Milavitsa has been producing lady’s lingerie, being one of the biggest lingerie producer in Eastern Europe. The company's products are successfully sold in more than 25 countries around the world and are characterized by high quality, original and fashionable designs.

Company has four main brands:

1. Milavitsa.

The Milavitsa product portfolio includes bras, knickers, shapewear, knitwear and swimwear. The Milavitsa collection is divided into three categories: classic, fashion and swimwear. The basis of the classic collection is formed by a large variety of everyday styles, combining the basics of classical design, comfortable construction and functional materials. Milavitsa is an expert in creating the styles of large sizes. Fashion and swimwear collections are created for every season, following the fashion trends in design, materials and accessories.

1. Avelive Collection.

Bras, knickers, shapewear are also manufactured under the Aveline trademark. This is comfortable everyday lingerie at affordable prices. The Aveline products are developed and manufactured by Milavitsa, which is a guarantee of high quality and compliance with the company’s standards.

1. Alisee.

Alisee is a French lingerie brand acquired by Milavitsa. Alisee collection is designed and styled by European professionals. Tailored by Milavitsa to fit local market specifics.

1. Hidalgo.

The Hidalgo is men's underwear combines classic shapes and comfortable natural materials.

## Problems because of poor data management

Company faced next problems:

1. absence of business intelligence from several sources;
2. absence of sales and inventory information consolidation for the calculation of the optimal order and delivery;
3. decreasing query and system performance;
4. absence of timely access to data;
5. absence of historical intelligence.

## Benefits from implementing a Data Warehouse

1. Better decision-making.

Corporate decision makers will no longer have to make important business decisions based on limited data and hunches. Data warehouse will store credible facts and statistics, and decision makers will be able to retrieve that information from the data warehouse based on their personal needs.

1. Quick and easy access to data.

Speed is an important factor that sets company above its competitors. Business users can quickly access data from multiple sources from a data warehouse, meaning that precious time will not be wasted on retrieving data from multiple sources. This allows company to make quick and accurate decisions, with little or no support from its IT department.

1. Data quality and consistency.

Since data warehouses gather information from different sources and convert it into a single and widely used format, departments will produce results that are in line and consistent with each other. When data is standardized, company can have confidence in its accuracy, and accurate data is what makes for strong business decisions.

# Data Description

## Description of selected schema

The Star schema was chosen for business processes description.

The main reasons:

1. simple structure;
2. absence of a big number of tables to join;
3. denormalized tables are not too large in a specific case of this task;
4. widely support by a large number of business intelligence tools.

## Description of data sources

Customer and manager information was generated on the site www.mockaroo.com. It allows creating file with 1000 rows.

Additionally, manager information was modified in the Excel. There was added column “Position name”. Information in that column was created by the specific random formula.

Information for Collection, Line, Product type, Store dimensions is on the official company’s site www.milavitsa.com. Information about Sizes can be find via tne next links www.milavitsa.com/collections/converter and www.globebrand.com/sizing\_charts

## Description of business rules

Business rules:

1. Every Retail Sale must be associated with a valid Customer.
2. A Retail Sale is always associated with a Payment Method.
3. A Retail Sale can have one or many Products.
4. A Retail Sale is always associated with a specific Employee.

# Dimensions of a Business

## Definition of selected Business Process

Business Process for analysis is Milavitsa’s sales per different metrics.

## Declaring the Grain

The first fact granularity is Daily Sales Amount per certain Customer and Employee in a specific a Store. The second fact granularity is Monthly Product Stock Value per certain Product and Manager in a specific a Store.

## Identifying the Dimensions

Schema should contains next dimensions:

1. Dim\_customers\_scd:
2. Dim\_products\_scd:
3. Dim\_stores:
4. Dim\_payment\_methods\_scd;
5. Dim\_employees\_scd;
6. Dim\_promotions\_scd;
7. Dim\_time\_day.

## Identifying the Facts

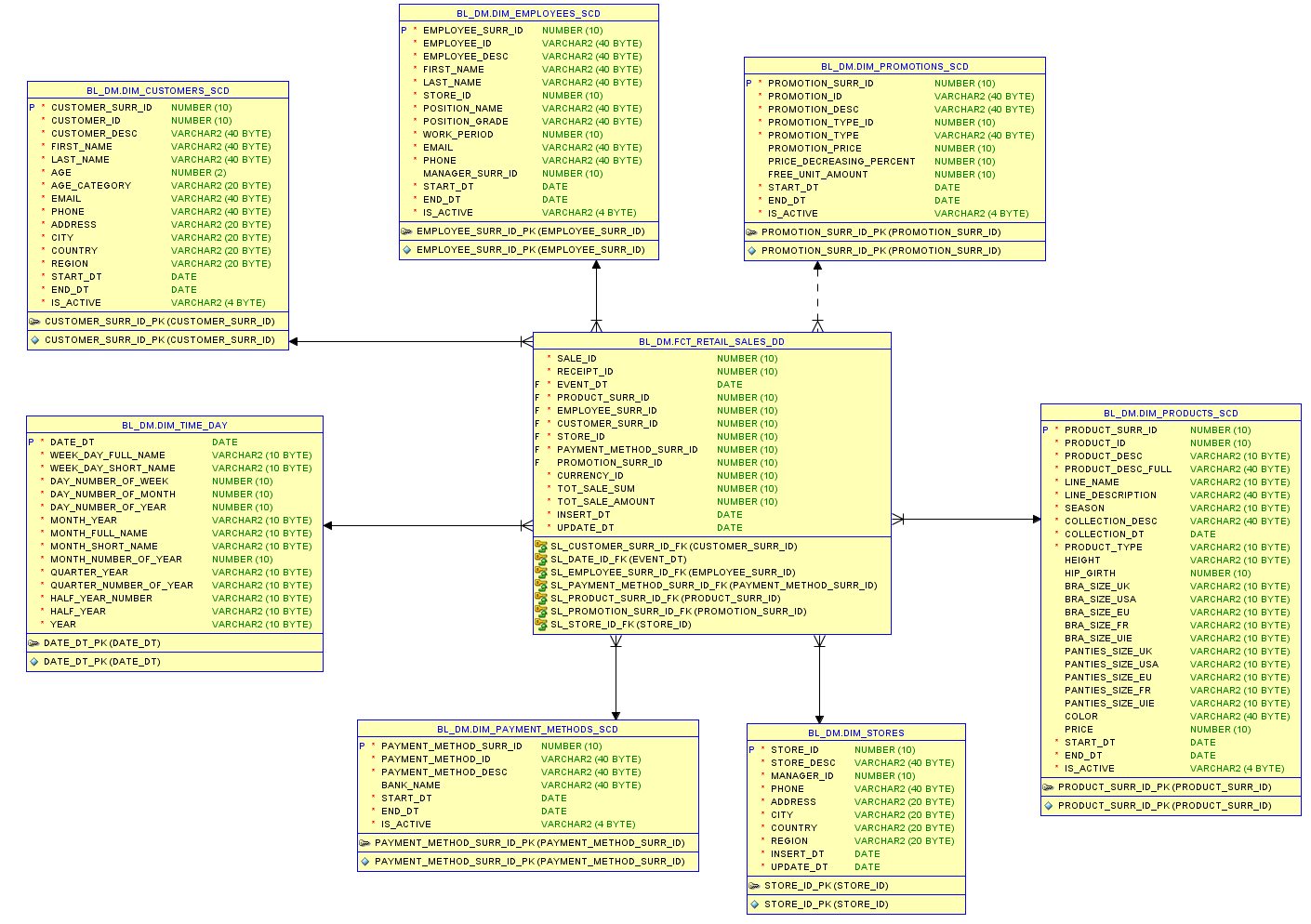
Schema should contains the next fact tables:

1. Fct\_retail\_sales;
2. Fct\_product\_balances;

## C:\Users\Valeryia\Desktop\dm_scheme.jpgDimensional model. Star Schema

Picture 1 Dimensional model

### Fact retail sales



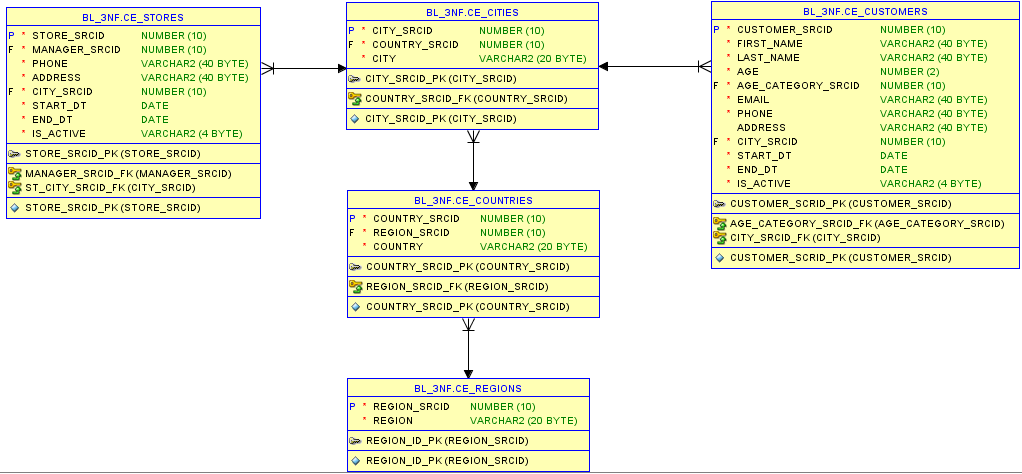
Picture 2 Retail sales fact table

### C:\Users\Valeryia\Desktop\bl_dm(2).pngFact product balances

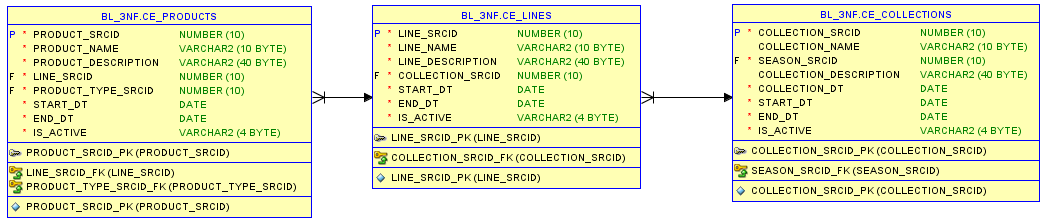
Picture 3 Balance product fact table

# NF-layer of Data Warehouse

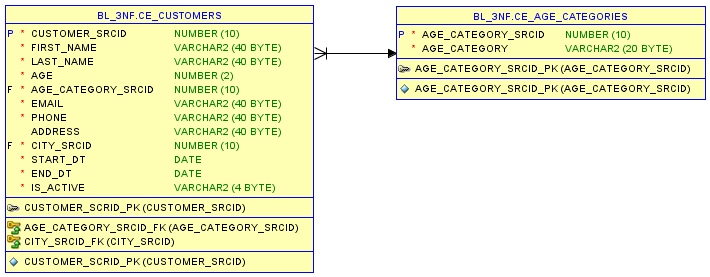
In this model the normalization was made with next steps:

1. Geographical information from the “Stores” and “Customers” was separated on different tables. Tables were created for each geographical object: region -> country -> city. These objects were connected in series: from the lowest in granularity to the highest.

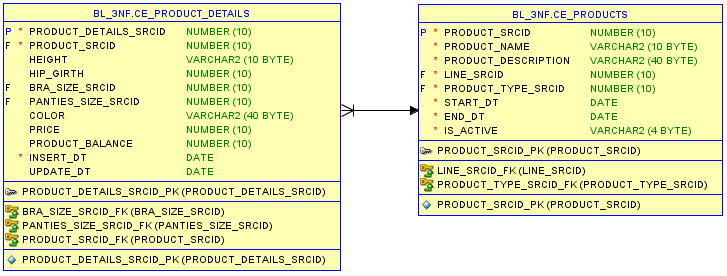
Picture 4 Geography part

1. Information about products was organized in the same way. There were separate tables for each object: collections -> lines -> products. These objects were connected in series: from the lowest in granularity to the highest.

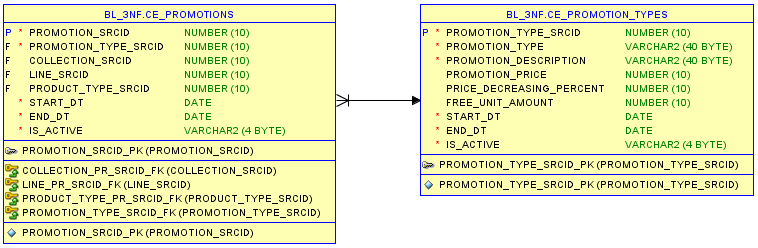
Picture 5 Product hierarchy part

1. A separate “Age\_categories” table was created for Customers. This table describes all possible age ranges that are important for business. Subsequently, each Customer can be assigned to a specific age category.

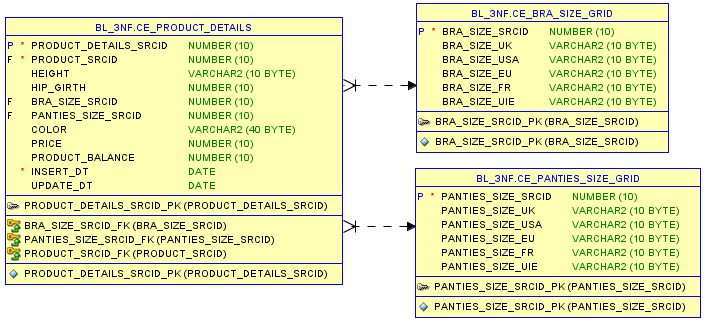
Picture 6 Customer part

1. Information about product characteristics was separated in a specific table “Product\_details”. It contains a complete description of each commodity unit of a certain size.

Picture 7 Product part

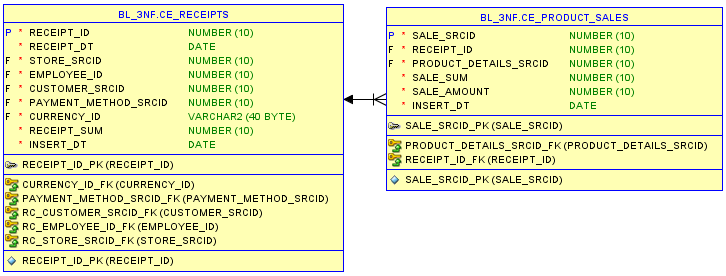
1. Also, the object “Promotions” was detailed as follows. Separately, a table was created on all possible types of promotions that were conducted on the network and was linked to a list of standard promotions.

Picture 8 Promotion part

1. The tables of the size grids “Bra\_size\_grid” and “Panities\_size\_grid”, which contain information on the international dimension matching for bra and penities, were separately created.

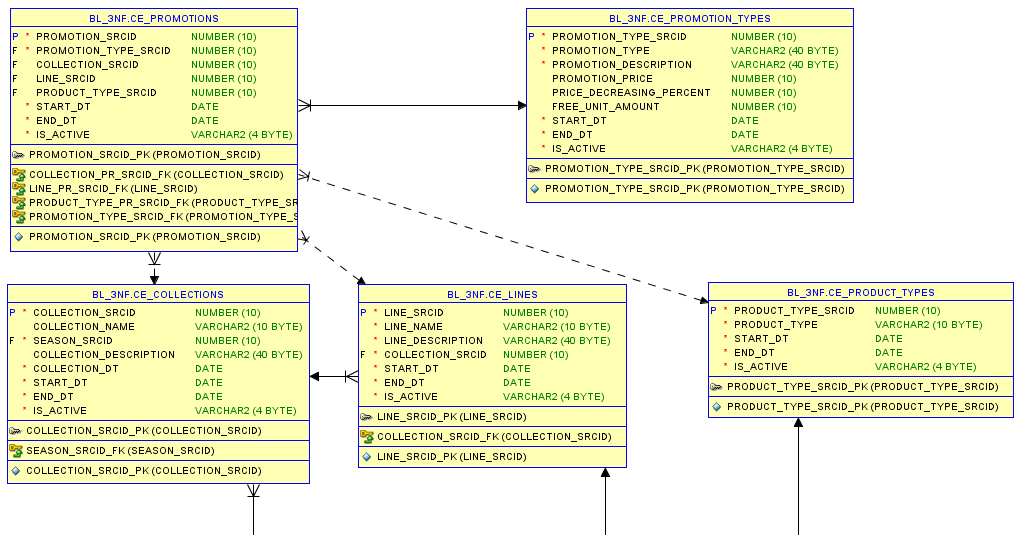
Picture 9 Size grid part

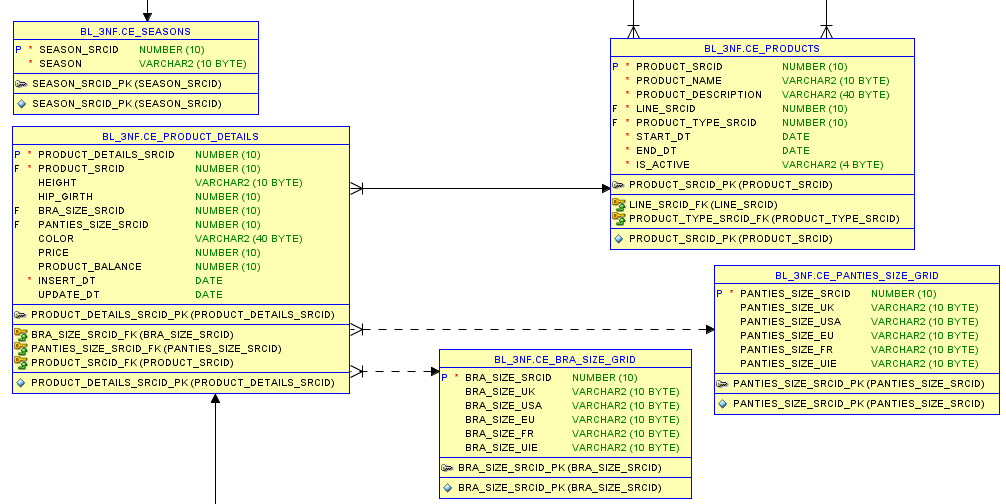
1. Additionally receipts information was add. It includes the calculation of their total cost.
2. The “Product\_sales” table details the check information, namely it describes a concrete product and amount of units what were purchased.

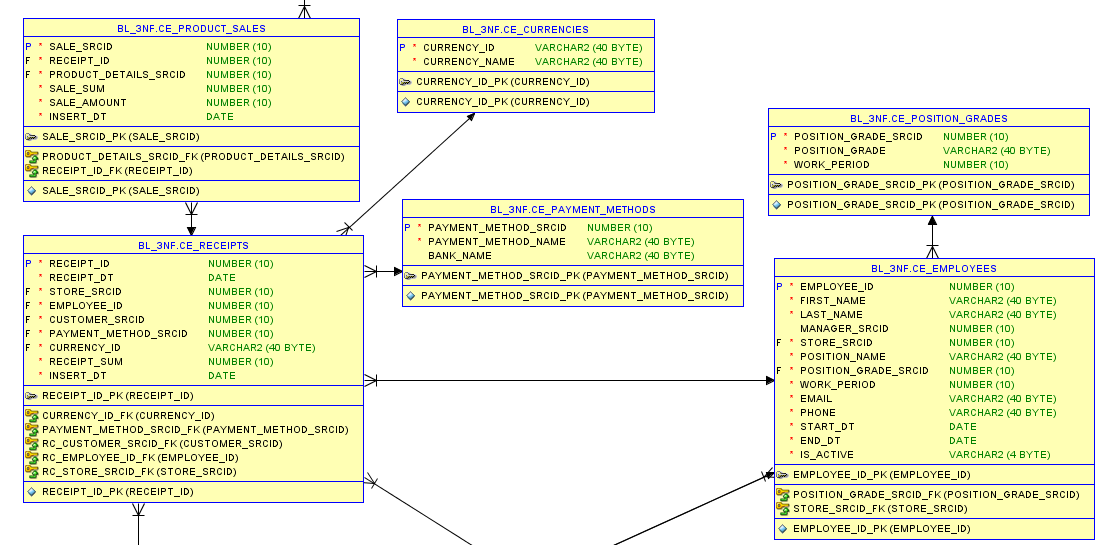


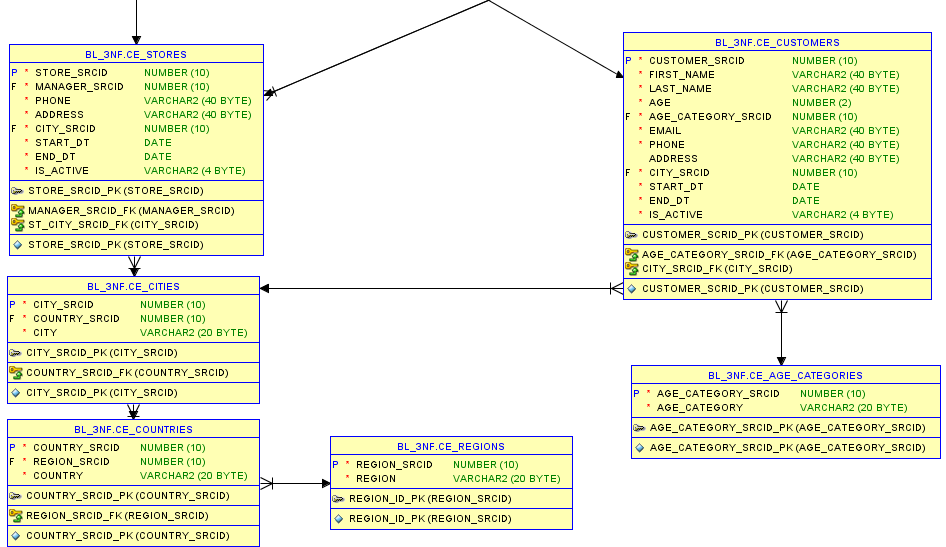
Picture 10 Sales part

## 3NF-model









Picture 11 3NF model

# Object Partitioning

A combination of two methods of data distribution is used to define a table with composite partitioning. Firstly, the table is partitioned by one method of data distribution. Then each section is subdivided into subsections using another method of data distribution. For example, if a range-list is specified for a table, then at the beginning the table is partitioned with keys ranges, and then each section is divided into subsections with the keys lists specified in the second partition. Sections of the table with composite partitioning exist only as metadata and do not provide actual data storage: the subsections of a particular section of a table or an index with composite partitioning are the physical segments of the database where the section data is stored.

Among the available partitioning methods, the above-described RANGE-LIST will be applied for the fact table FACT RETAIL SALES.

This partitioning method is most convenient for this fact table. It will be divided into decade’s partitions, and every decade, in turn, will contain partitions with stores per regions where sales are conducted. As the sales network is quite extensive, the granularity in the partitions to the level of the cities is not expedient. For business, it will be much more convenient to track sales within a specific region.

For the fact table FACT PRODUCT BALANCES, a similar partitioning strategy will be applied, as in this table commodity stocks will be analyzed.

CREATE TABLE fact\_retail\_sales (

sale\_id NUMBER(10) NOT NULL,

receipt\_id NUMBER(10) NOT NULL,

date\_id NUMBER(10) NOT NULL,

product\_details\_id NUMBER(10) NOT NULL,

employee\_id NUMBER(10) NOT NULL,

customer\_id NUMBER(10),

store\_id NUMBER(10) NOT NULL,

payment\_method\_id NUMBER(10) NOT NULL,

promotion\_id NUMBER(10),

currency\_id NUMBER(10) NOT NULL,

sale\_sum NUMBER(10) NOT NULL,

sale\_amount NUMBER(10) NOT NULL

)

PARTITION BY RANGE (date\_id) INTERVAL (NUMTOYMINTERVAL(3,'MONTH'))

SUBPARTITION BY LIST (store\_id) SUBPARTITION TEMPLATE

( SUBPARTITION local VALUES('...'),

SUBPARTITION Russia VALUES ('...'),

SUBPARTITION Europe VALUES ('...'),

SUBPARTITION Asia VALUES ('I')

)

(PARTITION half\_year\_2017 VALUES LESS THAN (TO\_DATE('01-JAN-2017','dd-MON-yyyy'))

)

PARALLEL;

# Business processes

In the analysis two business processes will be explored. The first of them is monthly sales of stores within regions, countries and cities, as well as sales by customers, the second - monthly commodity stocks of stores in different geographical sections and per network managers to identify the most valuable personnel.

Three reports are proposed as reports:

1. Sales analysis.

This table will be designed to provide general information about the functioning stores, sales, the number of products in stock, and the number of receipts. The lowest level of granularity will be represented by regions, and then DRILL-DOWN technology will make it possible to deepen the detailed analysis of analytics to countries and cities. The highest level of hierarchy will be presented by shops with a specific address in the city.

The following parameters will be presented in the table: the turnover for the previous period - the number of months of the previous period can be determined by specifying the corresponding value in the "Comparison period" field, the current period can be determined in the "Analysis period" field. The completion percentage will show what percentage of the revenue of the previous period is reached for the current period. The indicator quantity of goods determines the commodity remains, and the number of receipts shows how many sales were made for the analyzed period.

1. Analysis of receipts.

This report will allow user to view value of the main indicators per the metrics selected in the ListBox. In each ListBox, user can select multiple values at the same time. The analysis period can be determined by analogy with the previous report. The choice of month and year is provided using the ListBox.

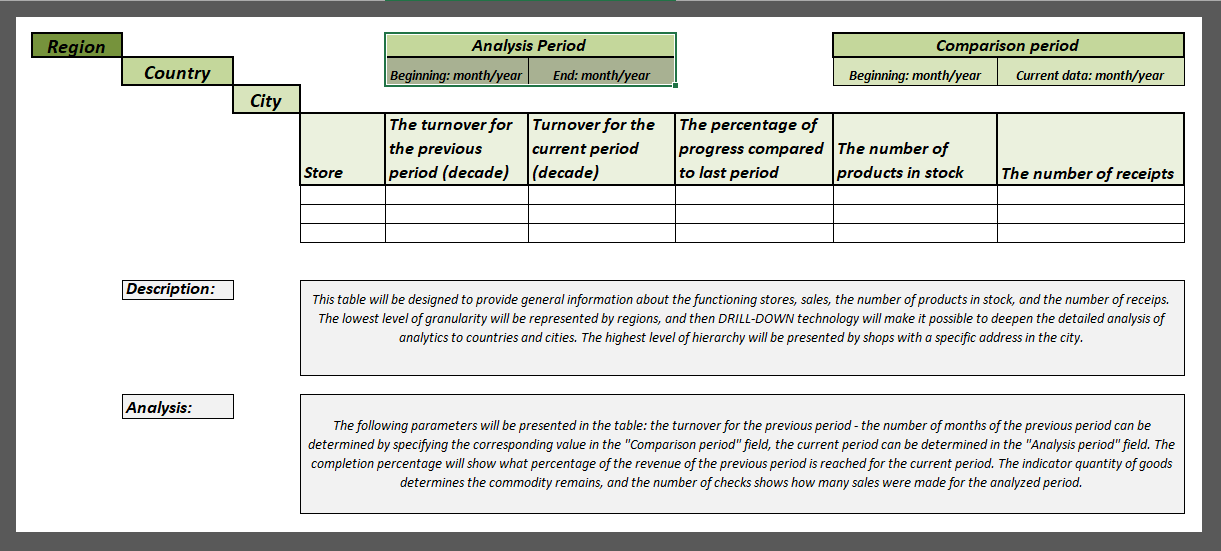
1. Load analysis.

This analysis is used to investigate customers’ influx during working days and working hours of the day what will be chosen for analysis. This analysis allows user to see the most favorable working periods for the promotions, as well as the time when it is necessary to have more staff in stores.

1. Data generation.

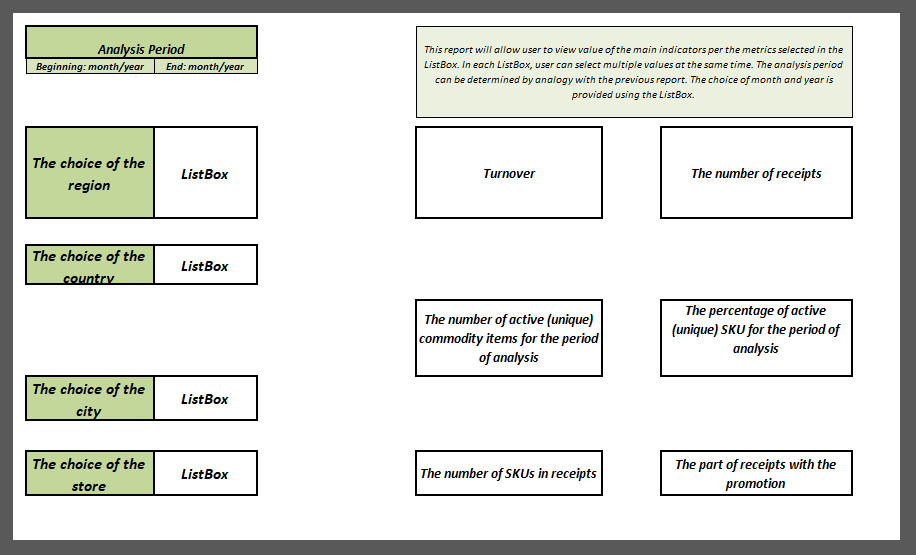
This tab contains tables with generated data to plot graphs in the "Load Analysis" report.

## Sales analysis



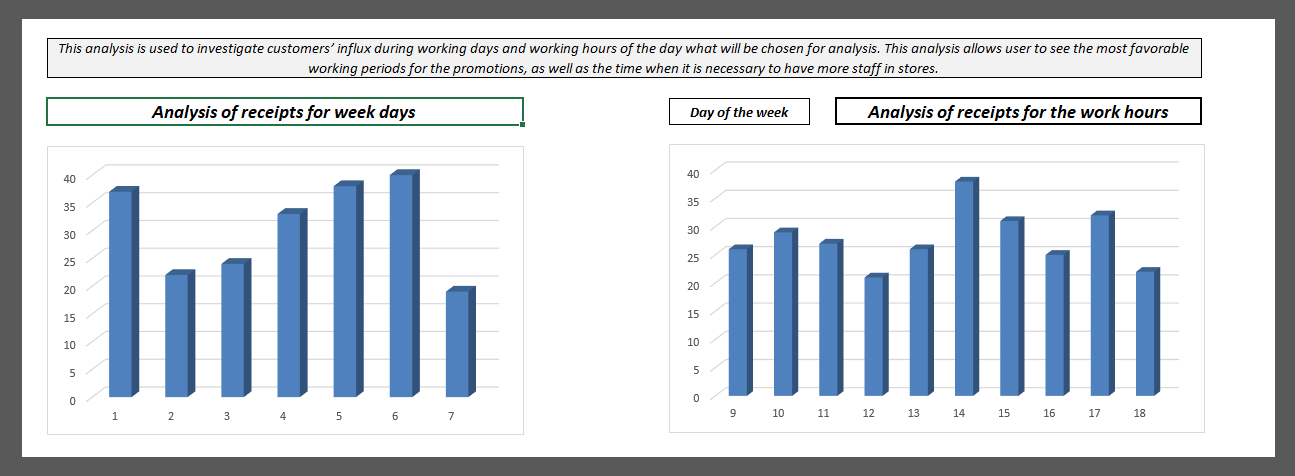
Picture 12 Sales analysis

## Analysis of receipts



Picture 13 Analysis of receipts

## Load analysis



Picture 14 Load analysis

# Data Modelling

## Detailing diagrams for 3NF and Star/snowflake layers

There were created seven dimensions. They are described in “Business Description #4” document. Several dimensions have a SCD2 type. It was chosen due to It is the next dimensions:

1. dim\_customers\_scd;
2. dim\_employees\_scd;
3. dim\_payment\_methods\_scd;
4. dim\_products\_scd;
5. dim\_promotions\_scd;

Dim\_time\_day has calendar type. Dim\_stores has a SCD1 type, because it is not necessary to have history about closed stores.

## Textual description of layers of data warehouse

In this work was used a two-layer architecture.

### Source layer

There were used flat files and html-code as data sources.



Picture 15 Source layer

### Data staging layer

The data stored to sources should be extracted, cleansed to remove inconsistencies and fill gaps, and integrated to merge heterogeneous sources into one common schema.

In this work will be used the following data staging layers step by step.

#### Cleansing layer

This layer was 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.

#### 3NF layer

Here reducing data redundancy and improving data integrity were made. Also, the process of simplifying the design of a database, so that it achieves the optimal structure composed of atomic elements, was achieved on this layer.

#### Dimensional model layer

The Star schema was chosen as a dimensional model for business processes description.

The main reasons:

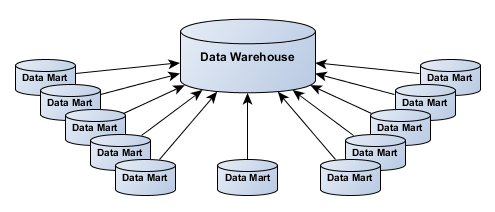
1. Simple structure;
2. Absence of a big number of tables to join;
3. Denormalized tables are not too large in a specific case of this task;
4. Widely support by a large number of business intelligence tools.

### Data warehouse layer

Information is stored to one logically centralized single repository: a data warehouse. The data warehouse can be directly accessed, but it can also be used as a source for creating data marts, which partially replicate data warehouse contents and are designed for specific departments. This layer is added after finishing previous steps.

### Analysis layer

In this layer, integrated data is efficiently and flexibly accessed to issue reports, dynamically analyze information, and simulate hypothetical business scenarios. Template of this layer is represented in “Retail Analysis” document.



Picture 16 DWH layer

## Visual description of layers of data warehouse



Picture 17 Layers model