|  |
| --- |
|  |
| Analysis of Milavitsa’s Sales |



Contents

[1. Business Description 3](#_Toc498389777)

[1.1. Business background 3](#_Toc498389778)

[1.2. Problems because of poor data management 3](#_Toc498389779)

[1.3. Benefits from implementing a Data Warehouse 3](#_Toc498389780)

[2. Data Description 5](#_Toc498389781)

[2.1. Description of selected schema 5](#_Toc498389782)

[2.2. Description of data sources 5](#_Toc498389783)

[2.3. Description of business rules 5](#_Toc498389784)

[3. Dimensions of a Business 6](#_Toc498389785)

[3.1. Definition of selected Business Process 6](#_Toc498389786)

[3.2. Declaring the Grain 6](#_Toc498389787)

[3.3. Identifying the Dimensions 6](#_Toc498389788)

[3.4. Identifying the Facts 8](#_Toc498389789)

[3.5. Dimensional model. Star Schema 10](#_Toc498389790)

[3.5.1. Fact\_retail\_sales 11](#_Toc498389791)

[3.5.2. Facs\_product\_balances 12](#_Toc498389792)

[4. 3NF-layer of Data Warehouse 13](#_Toc498389793)

[4.1. 3NF-model 16](#_Toc498389794)

[5. Object Partitioning 17](#_Toc498389795)

[6. Business processes 18](#_Toc498389796)

[6.1. Sales analysis 19](#_Toc498389797)

[6.2. Analysis of receipts 20](#_Toc498389798)

[6.3. Load analysis 21](#_Toc498389799)

# 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.
2. **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.
3. **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:

* Simple structure;
* Absence of a big number of tables to join;
* Denormalized tables are not too large in a specific case of this task;
* 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](http://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](http://www.milavitsa.com/). Information about Sizes can be find via tne next links [www.milavitsa.com/collections/converter](http://www.milavitsa.com/collections/converter/) and [www.globebrand.com/sizing\_charts](https://www.globebrand.com/sizing_charts)

## Description of business rules

Business rules:

* Every **Retail Sale** must be associated with a valid **Customer**.
* A **Retail Sale** is always associated with a **Payment Method**.
* A **Retail Sale** can have one or many **Products**.
* 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

Fact granularity is Daily Sales Amount per certain Customer and Employee in a specific a Store.

## Identifying the Dimensions

Schema should contains next dimensions:

1. Dim\_customers:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| CUSTOMER\_ID | NUMBER(10) | Primary Key |
| FIRST\_NAME | VARCHAR2(40 BYTE) |  |
| LAST\_NAME | VARCHAR2(40 BYTE) |  |
| AGE | NUMBER(2) |  |
| AGE\_CATEGORY | VARCHAR2(20 BYTE) |  |
| EMAIL | VARCHAR2(40 BYTE) |  |
| PHONE | VARCHAR2(40 BYTE) |  |
| ADDRESS | VARCHAR2(20 BYTE |  |
| CITY | VARCHAR2(20 BYTE) |  |
| COUNTRY | VARCHAR2(20 BYTE) |  |
| REGION | VARCHAR2(20 BYTE) |  |

1. Dim\_products:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| PRODUCT\_DETAILS\_ID | NUMBER(10) | Primary Key |
| PRODUCT\_ID | NUMBER(10) |  |
| PRODUCT\_NAME | VARCHAR2(10 BYTE) |  |
| PRODUCT\_DESCRIPTION | VARCHAR2(40 BYTE) |  |
| LINE\_ID | NUMBER(10) |  |
| LINE\_NAME | VARCHAR2(10 BYTE) |  |
| LINE\_DESCRIPTION | VARCHAR2(40 BYTE) |  |
| COLLECTION\_ID | NUMBER(10) |  |
| SEASON\_ID | NUMBER(10) |  |
| SEASON | VARCHAR2(10 BYTE) |  |
| COLLECTION\_DESCRIPTION | VARCHAR2(40 BYTE) |  |
| COLLECTION\_DATE | DATE |  |
| PRODUCT\_TYPE\_ID | NUMBER(10) |  |
| PRODUCT\_TYPE | VARCHAR2(10 BYTE) |  |
| HEIGHT | VARCHAR2(10 BYTE) |  |
| HIP\_GIRTH | NUMBER(10) |  |
| BRA\_SIZE\_UK | VARCHAR2(10 BYTE) |  |
| BRA\_SIZE\_USA | VARCHAR2(10 BYTE) |  |
| BRA\_SIZE\_EU | VARCHAR2(10 BYTE) |  |
| BRA\_SIZE\_FR | VARCHAR2(10 BYTE) |  |
| BRA\_SIZE\_UIE | VARCHAR2(10 BYTE) |  |
| PANTIES\_SIZE\_UK | VARCHAR2(10 BYTE) |  |
| PANTIES\_SIZE\_USA | VARCHAR2(10 BYTE) |  |
| PANTIES\_SIZE\_EU | VARCHAR2(10 BYTE) |  |
| PANTIES\_SIZE\_FR | VARCHAR2(10 BYTE) |  |
| PANTIES\_SIZE\_UIE | VARCHAR2(10 BYTE) |  |
| COLOR | VARCHAR2(40 BYTE) |  |
| PRICE | NUMBER(10) |  |

1. Dim\_stores:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| STORE\_ID | NUMBER(10) | Primary Key |
| MANAGER\_ID | NUMBER(10) |  |
| PHONE | VARCHAR2(40 BYTE) |  |
| ADDRESS | VARCHAR2(20 BYTE) |  |
| CITY | VARCHAR2(20 BYTE) |  |
| COUNTRY | VARCHAR2(20 BYTE) |  |
| REGION | VARCHAR2(20 BYTE) |  |

1. Dim\_payment\_methods;

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| PAYMENT\_METHOD\_ID | NUMBER(10,0) | Primary Key |
| PAYMENT\_METHOD\_NAME | VARCHAR2(40 BYTE) |  |
| BANK\_NAME | VARCHAR2(40 BYTE) |  |

1. Dim\_employees;

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| EMPLOYEE\_ID | NUMBER(10) | Primary Key |
| FIRST\_NAME | VARCHAR2(40 BYTE) |  |
| LAST\_NAME | VARCHAR2(40 BYTE) |  |
| STORE\_ID | NUMBER(10) |  |
| POSITION\_NAME | VARCHAR2(40 BYTE) |  |
| POSITION\_GRADE | VARCHAR2(40 BYTE) |  |
| HIRE\_DATE | DATE |  |
| FIRE\_DATE | DATE |  |
| WORK\_PERIOD | NUMBER(10) |  |
| EMAIL | VARCHAR2(40 BYTE) |  |
| PHONE | VARCHAR2(40 BYTE) |  |
| MANAGER\_ID | NUMBER(10) |  |
| M\_FIRST\_NAME | VARCHAR2(40 BYTE) |  |
| M\_LAST\_NAME | VARCHAR2(40 BYTE) |  |
| M\_POSITION\_NAME | VARCHAR2(40 BYTE) |  |

1. Dim\_promotions;

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| PROMOTION\_ID | NUMBER(10) | Primary Key |
| PROMOTION\_TYPE\_ID | NUMBER(10) |  |
| PROMOTION\_TYPE | VARCHAR2(40 BYTE) |  |
| PROMOTION\_DESCRIPTION | VARCHAR2(40 BYTE) |  |
| PROMOTION\_PRICE | NUMBER(10) |  |
| PRICE\_DECREASING\_PERCENT | NUMBER(10) |  |
| FREE\_UNIT\_AMOUNT | NUMBER(10) |  |
| START\_DATE | DATE |  |
| END\_DATE | DATE |  |

1. Dim\_date.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| DATE\_ID | NUMBER(10) | Primary Key |
| FULL\_DATE | VARCHAR2(10) |  |
| WEEK\_DAY\_FULL\_NAME | VARCHAR2(36 BYTE) |  |
| WEEK\_DAY\_SHORT\_NAME | VARCHAR2(12 BYTE) |  |
| DAY\_NUMBER\_OF\_WEEK | NUMBER |  |
| DAY\_NUMBER\_OF\_MONTH | NUMBER |  |
| DAY\_NUMBER\_OF\_YEAR | NUMBER |  |
| MONTH\_END\_DATE | NUMBER |  |
| MONTH\_FULL\_NAME | VARCHAR2(36 BYTE) |  |
| MONTH\_SHORT\_NAME | VARCHAR2(12 BYTE) |  |
| MONTH\_NUMBER\_OF\_YEAR | NUMBER |  |
| MONTH\_YEAR | VARCHAR2(17 BYTE) |  |
| QUARTER\_YEAR | VARCHAR2(7 BYTE) |  |
| QUARTER\_NUMBER\_OF\_YEAR | NUMBER |  |
| HALF\_YEAR\_NUMBER | NUMBER |  |
| HALF\_YEAR | VARCHAR2(7 BYTE) |  |
| YEAR | VARCHAR2(4 BYTE) |  |
| YEAR\_END\_DATE | NUMBER |  |

## Identifying the Facts

Schema should contains the next fact tables:

1. Fact\_retail\_sales;

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| SALE\_ID | NUMBER(10,0) | Surrogate identifier of one purchase |
| RECEIPT\_ID | NUMBER(10,0) | Unique number of receipt – natural identifier of one purchase |
| DATE\_ID | NUMBER(10,0) | Foreign Key |
| PRODUCT\_DETAILS\_ID | NUMBER(10,0) | Foreign Key |
| EMPLOYEE\_ID | NUMBER(10,0) | Foreign Key |
| CUSTOMER\_ID | NUMBER(10,0) | Foreign Key |
| STORE\_ID | NUMBER(10,0) | Foreign Key |
| PAYMENT\_METHOD\_ID | NUMBER(10,0) | Foreign Key |
| PROMOTION\_ID | NUMBER(10,0) | Foreign Key |
| CURRENCY\_ID | NUMBER(10,0) | Foreign Key |
| SALE\_SUM | NUMBER(10,0) | Total sum per a certain sale |
| SALE\_AMOUNT | NUMBER(10,0) | Total amount of units per a certain sale |

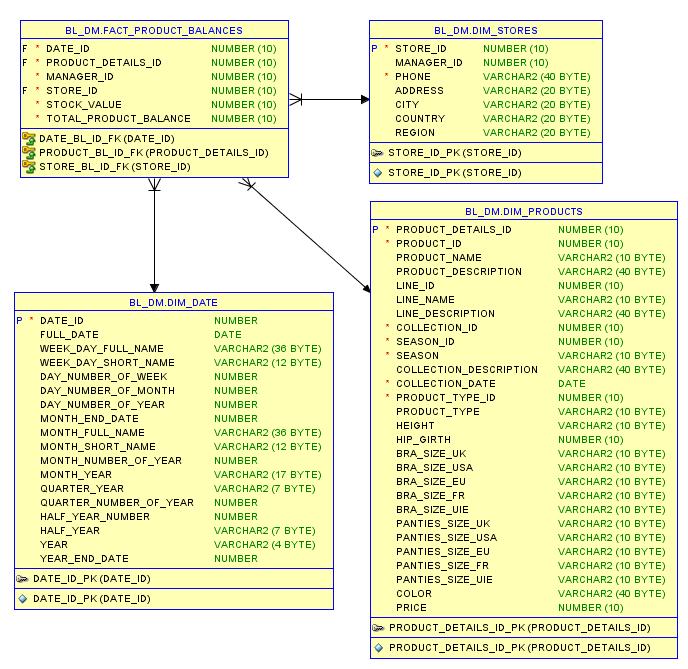
1. Fact\_product\_balances;

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Descroption** |
| DATE\_ID | NUMBER(10,0) | Foreign Key |
| PRODUCT\_DETAILS\_ID | NUMBER(10,0) | Foreign Key |
| MANAGER\_ID | NUMBER(10,0) | Foreign Key |
| STORE\_ID | NUMBER(10,0) | Foreign Key |
| STOCK\_VALUE | NUMBER(10,0) | Total stock sum per a product |
| TOTAL\_PRODUCT\_BALANCE | NUMBER(10,0) | Total stock amount of units per a product |

## C:\Users\Valeryia\Desktop\Task 07\bl_dm_model.jpgDimensional model. Star Schema

### C:\Users\Valeryia\Desktop\Task 07\fact_retail_sales.jpgFact\_retail\_sales

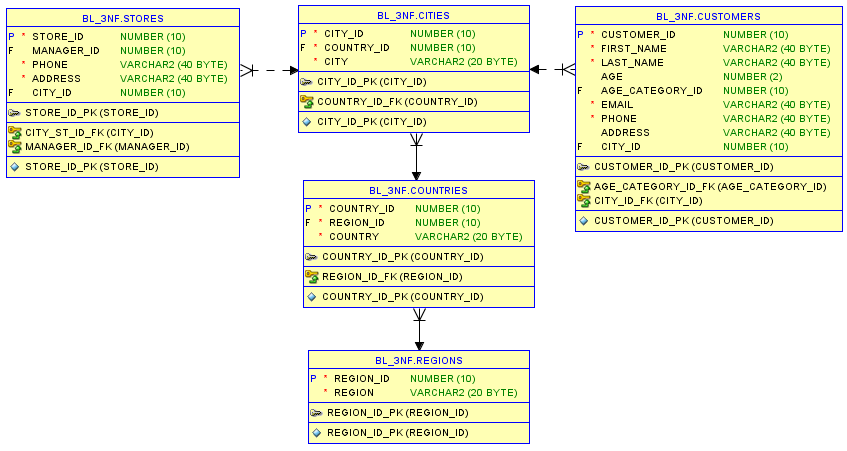
### Facs\_product\_balances



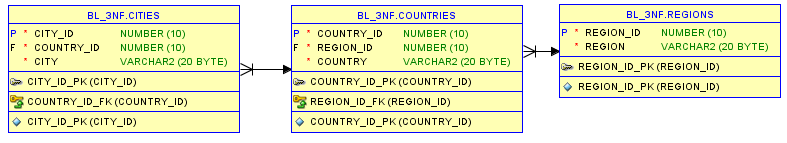
# 3NF-layer of Data Warehouse

In this model the normalization was made with next steps:

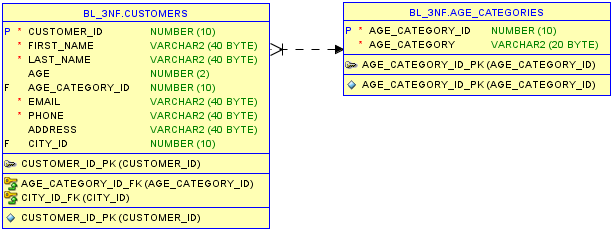
1. Geographical information from the “Stores” and “Customers” was separated on different tables. Tables were created for each geographical object: region -> country -> city. These objects were connected in series: from the lowest in granularity to the highest.



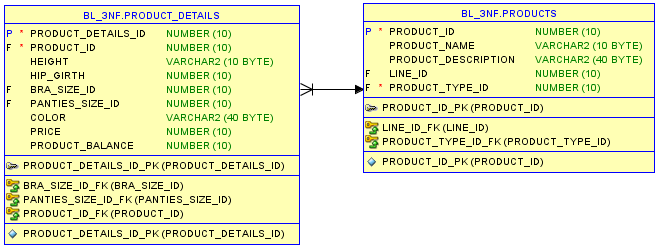
1. Information about products was organized in the same way. There were separate tables for each object: collections -> lines -> products. These objects were connected in series: from the lowest in granularity to the highest.



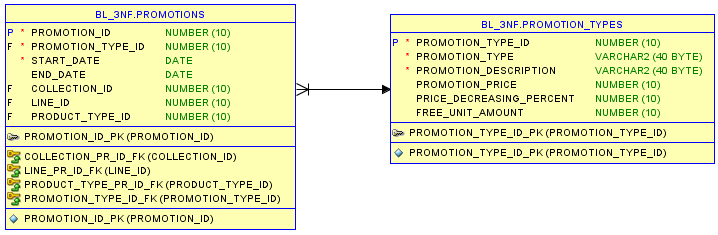
1. A separate “Age\_categories” table was created for Customers. This table describes all possible age ranges that are important for business. Subsequently, each Customer can be assigned to a specific age category.



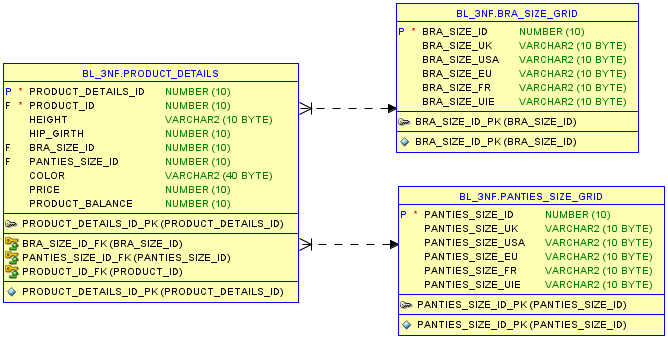
1. Information about product characteristics was separated in a specific table “Product\_details”. It contains a complete description of each commodity unit of a certain size.



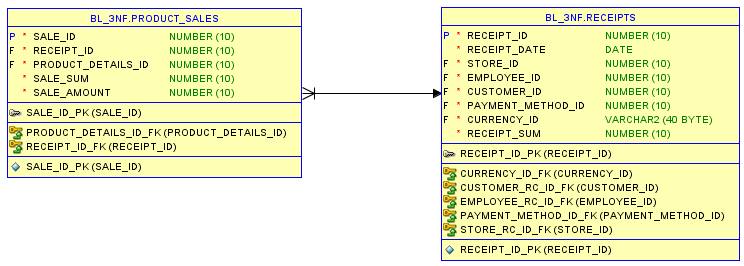
1. Also, the object “Promotions” was detailed as follows. Separately, a table was created on all possible types of promotions that were conducted on the network and was linked to a list of standard promotions.



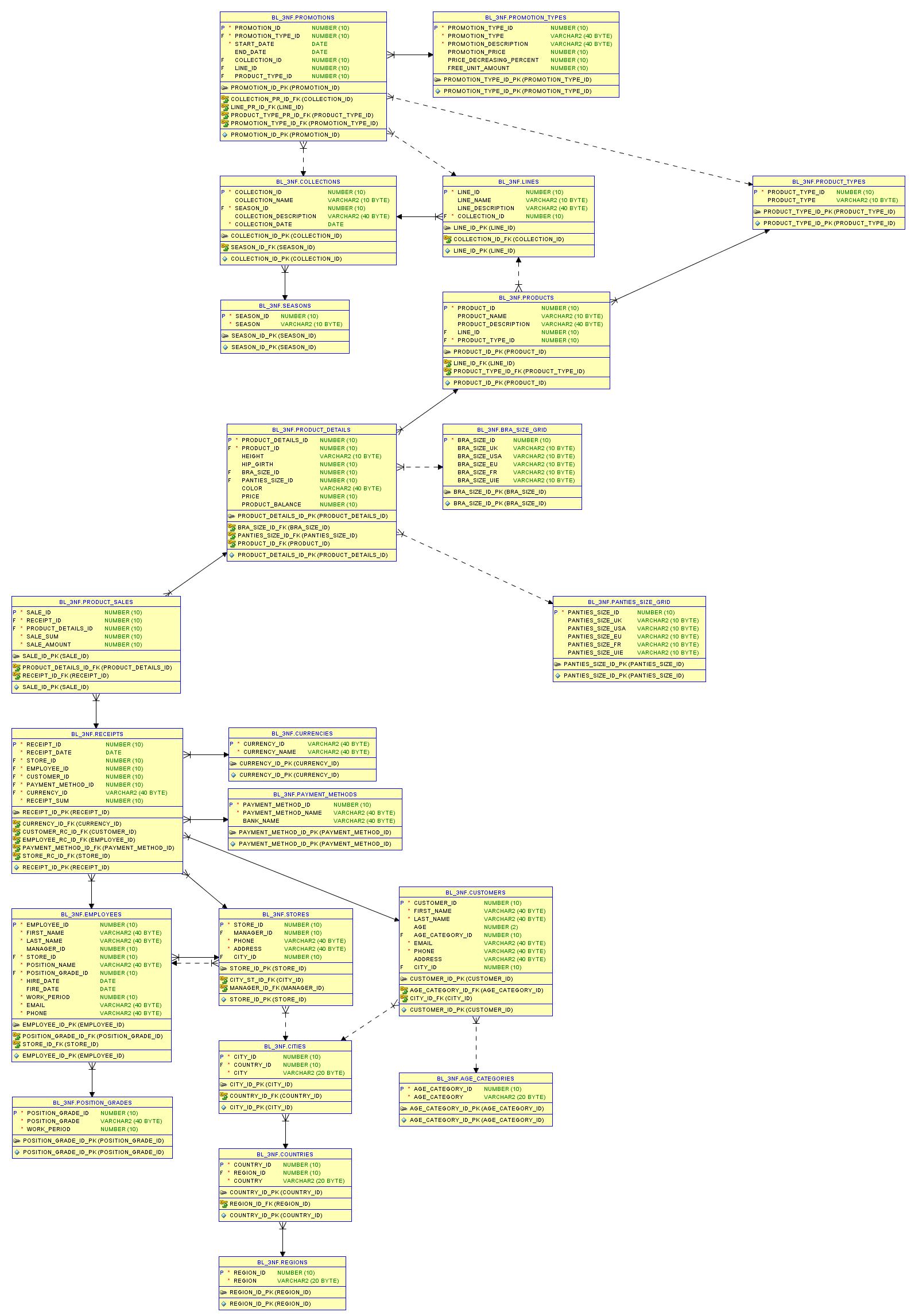
1. The tables of the size grids “Bra\_size\_grid” and “Panities\_size\_grid”, which contain information on the international dimension matching for bra and penities, were separately created.



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.



## 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.

# 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.

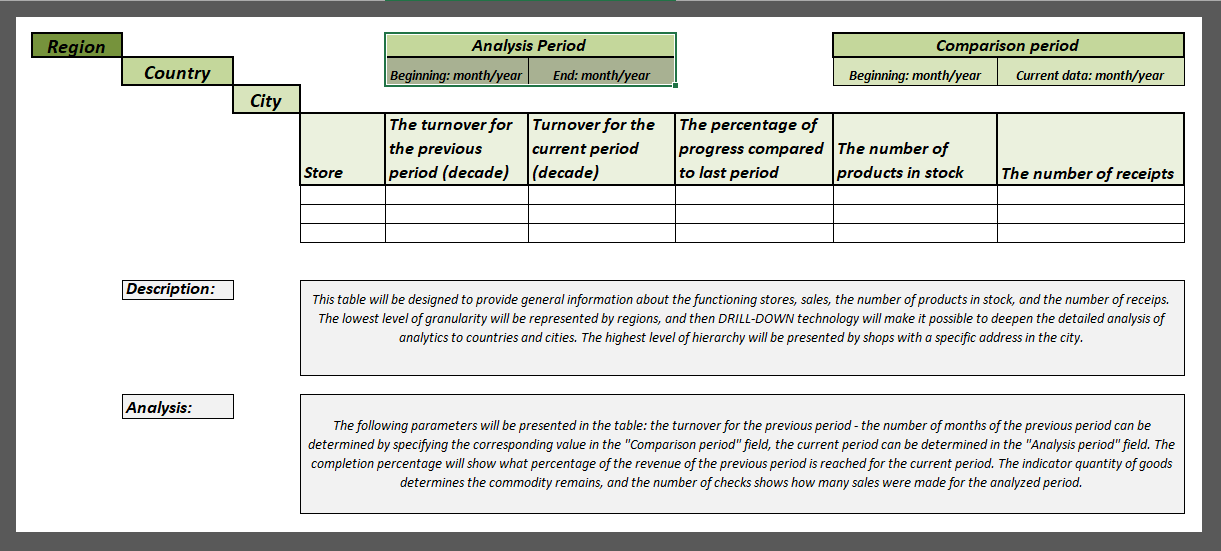
1. **Load analysis.**

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.

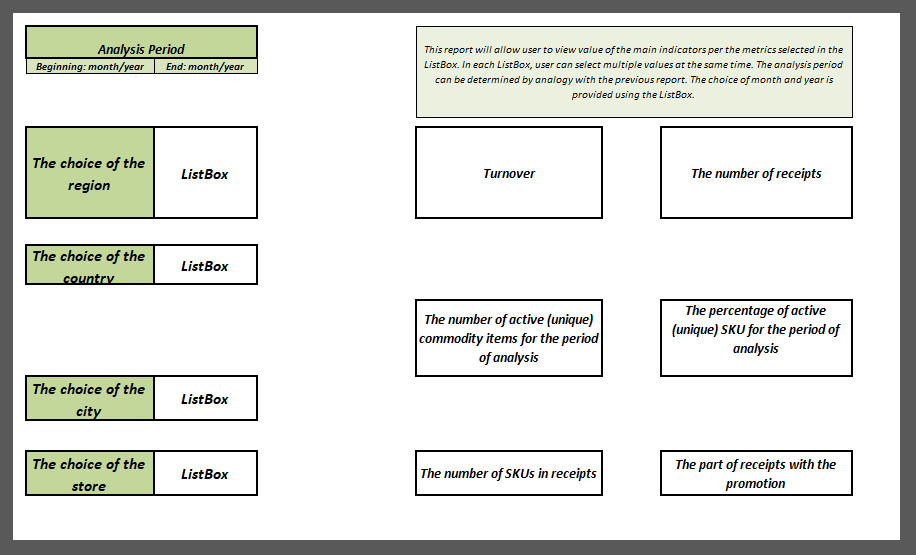
1. **Data generation.**

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

## Sales analysis



## Analysis of receipts



## Load analysis

