


| | | | | |
|--|----------------|---|----------|---------------------|
|  | Course | Databases and Information Systems 2016 | | |
| | Exercise Sheet | 7 | | |
| | Points | – | | |
| | Release Date | June 21 2016 | Due Date | July 05 2016 |

1 Data Warehousing

In this exercise, you create a data warehouse to analyse sales of various items of the department store chain *Superstore*. It is also your task to implement a tool that puts the manager of the department store chain in the position to analyse the sales made in the different department stores. The sales development of products or product families per shop, region and country over different periods such as the day, month, and quarter are of special interest.

Note

- The cross table in Exercise 1.2 can simply be printed to the console. You can choose a different table layout for your convenience as long as your table contains the required information.
- To avoid performance problems, execute inserts in batches (see `java.sql.PreparedStatement.addBatch()`).

1.1 ETL Process

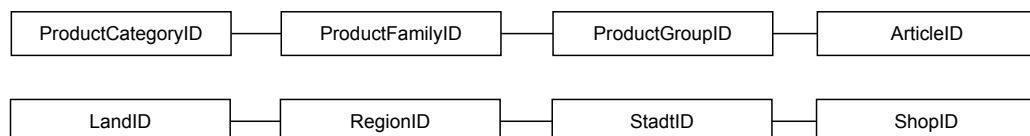
The first step is to implement an ETL process to populate the data warehouse prior to data analysis. The data required for analysis are available in different systems:

- The data for the individual stores and the product ranges are located in a corporate database. You can connect to the database via JDBC:

```
jdbc:db2://vsisls4.informatik.uni-hamburg.de:50001/VSISP
```

All relevant data reside in the schema DB2INST1.

Unfortunately, there is no proper documentation for the tables in the corporate database. However, the following hierarchies can be reconstructed easily from the foreign key relationships:




- The number of sold products and their turnover per department store, day and product are given in form of a CSV file. You can find the sales data of the last 5 months in:

<http://vsis-www.informatik.uni-hamburg.de/teaching/ss-14/dis/aufgaben/Sales.zip>

In the first line of the CSV file, there is a short description of all columns. The turnover is given in Euros.

Implement the ETL process as a Java application that extracts data from both data sources and loads them into a destination schema that you also have to define; the destination schema in the data warehouse is to be realised as a star schema.

During the transformation, be careful when converting data (number and date formats) and schemas (hierarchies in flat dimension tables) for the destination schema of the data warehouse. Please provide your program code with console printouts, so that your ETL process is easily understandable. Finally, populate the data warehouse with your ETL tool.

| | | | | |
|--|----------------|---|----------|---------------------|
|  | Course | Databases and Information Systems 2016 | | |
| | Exercise Sheet | 7 | | |
| | Points | – | | |
| | Release Date | June 21 2016 | Due Date | July 05 2016 |

1.2 Data Analysis

For reasonable data analysis, the manager requires another application that enables him to navigate in the data cube. Implement a Java application that outputs the data as in the following cross table:

| | sales | product 1 | product 2 | ... | total |
|---------|-----------------|-----------|-----------|-----|-------|
| Hamburg | quarter 1, 2011 | 12 | 48 | ... | ... |
| | quarter 2, 2011 | 31 | 12 | ... | ... |
| | quarter 3, 2011 | 50 | 1 | ... | ... |
| | quarter 4, 2011 | 2 | 0 | ... | ... |
| | total | 95 | 61 | ... | ... |
| Bayern | quarter 1, 2011 | 11 | 88 | ... | ... |
| | quarter 2, 2011 | 12 | 99 | ... | ... |
| | quarter 3, 2011 | 15 | 75 | ... | ... |
| | quarter 4, 2011 | 9 | 12 | ... | ... |
| | total | 47 | 247 | ... | ... |
| ... | ... | ... | ... | ... | ... |
| total | | 142 | 335 | ... | ... |

Furthermore, the Java application should be able to navigate along the dimensional hierarchies (drill down, roll up). You may use the star query and DB2 SQL extensions like **GROUPING SETS**, **ROLLUP** or **CUBE**.