Data Warehousing Assignment—Part I

Toon Calders Hatem Haddad

Deadline: 14 November 2017

1 Practical information

Deadline:	14 November 2017
Group size:	Up to four students
How to submit:	Upload solution at `uv.ulb.ac.be`

2 Objectives

The goals of this assignment are:

1. Create a dimensional fact model;
2. Translate it to the relational model;
3. Change your model considering the changes that may take place.

3 Database Description

In order to run the salary payment process for its employees, the human resources department of a company maintains a database with the current information about its employees. For every employee personal information is stored (first name, last name, gender, date of birth, marital status, number of children), as well as the date he or she was hired, his or her current salary, title, and department. There is a current work location for every employee and also his or her base location (city he or she was hired from). The employees can work for one or more clients. For client, the client name and the category (large,medium,small) is stored. For every city the state and for every state the country is stored. Figure 1 shows the database tables used by human resources. A copy of this database is available on the course web-site under the name **Assignment1**. This is the database snapshot of January 1st 1991 (`"1991-01-01 00:00:00.000"`). Later on further snapshots will be released.

4 Problem Description

The company would like to keep track of its work force and how it evolves over time. Therefore it is decided to extract the salary information at regular intervals (monthly) from the human resources database and to store it into a small data warehouse to allow for historical analysis of the salary evolution. The data warehouse should allow for storing the historical data over a long time and should anticipate potential changes in the database. Based on the data stored in the data warehouse it should be possible to answer analysis queries such as the following:

1. Give the average salary of managers per year for the sales and marketing departments;
2. Give the monthly average salary per gender;
3. Give geography location (city, state, country)-wise employee distribution.
4. Give, for every month in the period from January 1990 till July 1990, the average salary for all employees working with a specific client.
5. Give the total salary spent, per job title, since 1989.

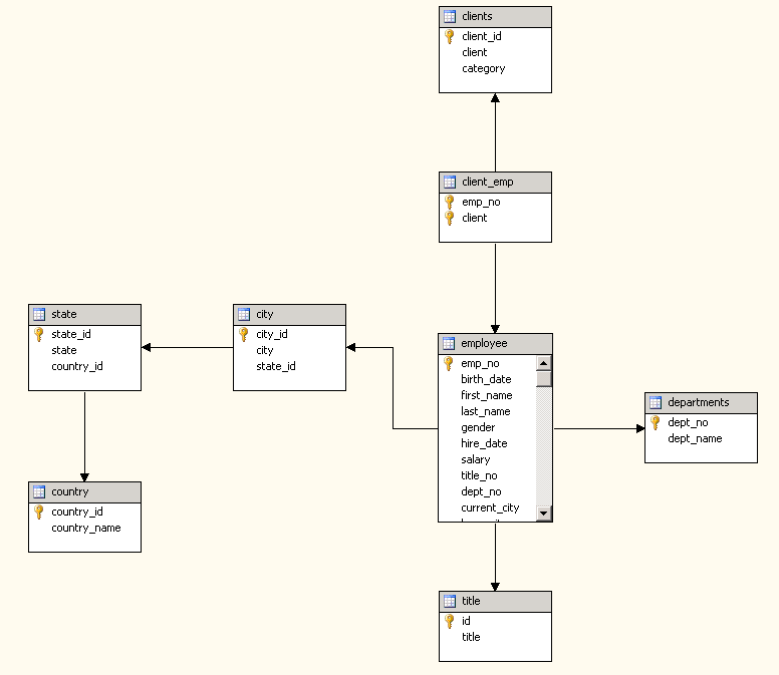

Figure 1: Database tables of the Assignment1 database.

5 Deliverables

You should deliver a report (as a.pdf), containing the following elements:

1. A **cover page** with the list of group members, including student ID,
2. Your **dimensional fact model** with a short explanation (total: 8 points). Make a dimensional model for the data warehouse. Indicate which cube(s) are needed, what are the dimensions, measures, hierarchies, etc.
3. Your explanation how **to anticipate potential changes** in the database (total: 4 points),
4. Your **relational model** with a short explanation where necessary (total: 8 points). Describe the tables for storing the data in a relational database; that is, in a ROLAP solution; make sure that your tables can accommodate changes in the data as much as possible.

Submit both files as a single .zip-file on the université virtuel course website.