Data Warehousing Assignment—Part I

Toon Calders Rohit Kumar

Deadline: 6 November 2015

1 Practical information

Deadline:	6 November 2014
Group size:	Three 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 Database Description

The iNG bank is maintaining a simple MSSQL database of its customers and their daily transactions. For every customer personal information is stored (First name, last name, gender, date of birth, marital status, number of children), as well as the accounts he or she has. For every account the date it was opened, the current balance, account type, and branch where it was opened is stored. There is a current city for every customer (where the customer is currently residing), and for every branch the city where it is located. The customer can have more than one account in more than one branch. Later on there will be some accounts which are external accounts and do not belong to any customer of the bank. For those the owner id of the account will be *null*. Every transaction (deposit or withdraw) done by every customer is stored in a transaction table. This table contains the account for which the transaction was performed, the date and the amount of the transaction. Further, the transactions have a 'from account' which maintains the account id from where the money came (if it was a deposit) or went (if it was a withdraw). For every city the state and for every state the country is stored. Figure 1 shows the database tables used by the bank. A copy of this database is available on the SQL Server CS-MSSQL under the name dw2015. This is the database snapshot of January 1st 2011. Notice that the transaction table is empty because the bank will start recording transactions in the data warehouse only from January 1st 2011 on. Later on further snapshots will be released. In these future snapshots the transaction table will gradually be filled.

4 Problem Description

The bank would like to keep track of its different accounts and how they evolve over time. Therefore it is decided to extract the transaction information on a daily basis from the database and to store it into a small data warehouse for historical analysis of the account evolution. The data warehouse should allow for storing the historical data over a long period of time and should anticipate potential changes in the database such as in current city, number of children, marital status, ... of its customers, and branch, type and balance of an account. 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 transaction amount over all customer per month for a given branch;
2. Give the average total transaction amount per gender;
3. Give the average and total transaction amount per region and month;

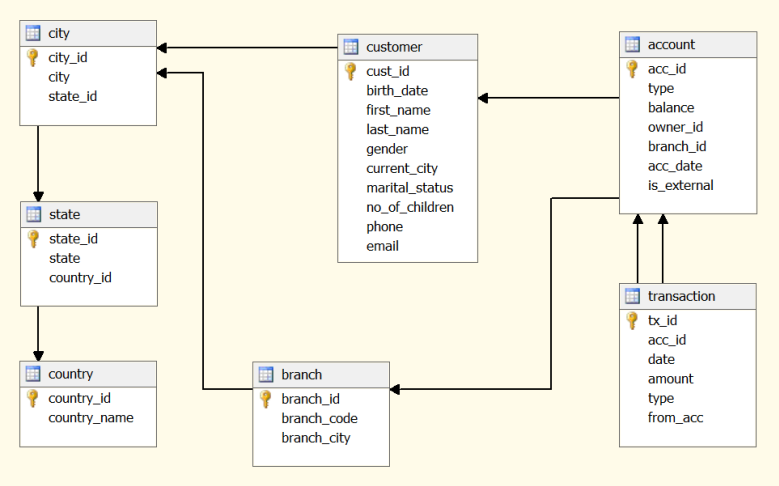

Figure 1: Database schema of the database for assignment part I. This database is available on the CS-MSSQL server as dw2015.

4. Give the total money deposited per branch during last month;
5. Give for every day the average amount deposited or withdrawn per branch for a given customer country.

5 Deliverables

You should deliver the following elements.

1. A report (as a.pdf), containing (length indication is purely indicative):
   (a) A **cover page** with the list of group members, including student ID,
   (b) Your **Dimensional fact model** with a short explanation where necessary (total: 1p to 1.5p),
   (c) Your **relational model** with a short explanation where necessary (total: 1p),
2. A .sql file containing the create table statements of your relational model.

Submit both files in a single .zip-file on the Université Virtuelle course website `uv.ulb.ac.be`.