TDT4300 Datavarehus og datagruvedrift - Spring 2013
Assignment 1: Data Warehousing

Make a complete report of the whole assignment. The assignment can be solved in groups of two students (or you can solve it alone).

Choose between two alternatives for exercise 2: one where you have to elaborate on a case and create fictitious data, and an alternative where we ask you to look for available relevant online dataset.

Exercise 1

a) Explain in your own words the concepts of OLTP (Online Transaction Processing) and OLAP (Online Analytical Processing). Emphasize on the differences between the two concepts in terms of properties and usage.

b) Explain the concept of data cube and the meaning of the term "cuboids".

c) Explain the data cube operations slice, dice, rollup and drill-down.

Exercise 2 (alternative a)

You are asked to create a data warehouse of traffic accidents in Norway to investigate the arterial routes that are most essential for the society to improve or set lower speed limits etc. We will be looking at direct costs of accidents and we will not take into account injuries etc. The data come from various insurance companies and they contain:

- When (date) and where the accident occurred (street and city, or such section of road and county).
- Driver related data (we are mostly interested in the age of the driver and whether he was drunk or not).
- Type of insurance of the car and insurance fees.

The data are imprecisely formulated and it is part of the task to select which information is necessary to include, or find a way to express the facts of the accidents. The main goal of the exercise is to practice modeling principles for data warehousing. You should mention explicitly any assumptions you may make.

a) Make a star or snowflake schema for this case description.

b) Define two different concept hierarchies (freely chosen dimensions).

Exercise 2 (alternative b)

Look for datasets that are available online which can be of interest to store and process using a data cube.

For possible data you look at: data.norge.no or similar listings in other countries. You can also use world population prospects data provided with this exercise.

a) Make a case description for the chosen dataset where you explain what facts and dimensions you find interesting. It is of course allowed to add fictitious dimensions to the data as long as it is realistic.

b) Make a star or snowflake schema for this case description.

c) Define two different concept hierarchies (freely chosen dimensions).

Exercise 3

Install the community edition of icCube tool (http://www.iccube.com) and implement a solution for your data. It is very convenient to use an Excel file as data source.

If you have chosen alternative a) of assignment 2, you will need to create some fictitious data. Use the dataset you found online if you have chose alternative b).

a) Use the builder and create a scheme for your solution that should include an Excel-file as data source, at least two dimensions with one hierarchy in each dimension, and one cube containing the facts (measures).

b) Write at lest two MDX queries that will illustrate your hierarchies and the aggregation in your cube.

c) Document your work e.g. by taking screenshots from the builder and med editor.

Final notes

If you have trouble installing and using icCube, you can ask questions in the discussion forum on It´s Learning. Please also look in the discussion forum for solutions to some intitial problems we already are aware of for Mac users.