

TDT4300 Datavarehus og datagruvedrift - Spring 2013

Assignment 1: Data Warehousing

Exercise 1

- a) Explain 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

You are asked to create a data warehouse of traffic accidents in Norway to investigate the arterial routes that are most essential for the to 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) Create two different concept hierarchies (freely chosen dimensions).

Exercise 3

Suppose that a data warehouse consists of the four dimensions date, spectator, location, and game, and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors, with each category having its own charge rate.

- a) Draw a star schema diagram for the data warehouse.
- b) Starting with the base cuboid [date, spectator, location, game], what specific OLAP operations should you perform in order to list the total charge paid by student spectators at GM Place in 2010?
- c) Bitmap indexing is useful in data warehousing. Taking this cube as an example, briefly discuss advantages and problems of using a bitmap index structure.

Notes

- Provide a **pdf** report. Include **two** diagrams (star and snowflake) for Exercise 2 and **one** diagram (star diagram) for Exercise 3.