

Introduction to Databases, ITU, Fall 2003

Rasmus Pagh

October 30, 2003

Exercises on November 4

You are unlikely to be able to do all the exercises in two hours. Therefore you are encouraged to prepare at home, and use the exercises for those parts you find difficult.

The exercises this week concerns building a data warehouse for **real** warehouses, i.e., a warehouse storing physical goods. Analyzing such a business involves considering things such as inventory, vendors, profits, storage time, and perhaps irregularities (e.g., returned, damaged, or lost items).

1. What business process(es) should you model (each process becomes one fact table)?
2. What grain should be the business process(es) be modeled at, supposing we want the data warehouse to run on a standard workstation with, e.g., 200 GB of disk space?
3. Choose the relevant dimensions for the model(s) of the business process(es) and put content (levels, attributes) into them.
4. Choose the measures for the fact table(s).
5. Document your design as a multidimensional diagram and a relational star schema design.
6. Write SQL queries (working on the star schema) that compute information such as profit during the last month, average number of shipments per day, or whatever you can think of.

In cases of doubt about what is to be done, make reasonable assumptions.

To be handed in no later than November 13, 11.59 AM:

Last hand-in!

Consider a university database with the following schema:

```
Student(id, name, birthdate)
Course(id, name, semester, weekday, room, teacher-id)
Enrolled(student-id, course-id)
Teacher(id, name, department)
```

The meaning of the relations is the obvious one, for example, `Enrolled` contains one tuple for each student taking a particular class. The database is used in a multi-user environment. Explain what SQL isolation level would you use, and why, for each of the following transactions:

- Enroll a student (identified by id) to “Introduction to databases” (having id IDBIF03).
- Change the enrollment of a student (identified by id) from “Networks and protocols” (having id NPF03) to “Introduction to databases” (having id IDBIF03).
- Assign a new teacher (identified by id) to the class with the highest number of students.
- For each class, show the number of students enrolled in the class.

As always, your hand-in must be completed individually.