

TENTAMEN

TDDD46 Databasteknik

January 17, 2014, 8.00-12.00

Help

Dictionary.

Grades

You can get max 34 points. To pass the exam, grade 3, you need 7.5 and 9.5 points in the practical and theoretical part of the exam, respectively.

Questions

Fang Wei-Kleiner will visit the room at 10.00.

Instructions

Write clearly. Give relevant and motivated answers only to the questions asked. State the assumptions you make besides those in the questions. None of these additional assumptions should change the spirit of the exercises. Please, answer in English.

Good luck!

Practical part (15 points)

Question 1. Data modeling with EER diagram (5 p):

We want to create a database to store information about the friendship relations among a group of people. In particular, for each male, we want to store his Swedish friends, and also his non-Swedish friends. We want to store the same information for each female. Bear in mind that every male and female is characterized by a single attribute, personal number. If you use any other attribute in your EER diagram, we will withdraw one point.

Draw an EER diagram for the description above. Clearly write down your choices and assumptions in case you find that something in the information above is not clear.

Question 2. SQL (1 + 2 + 2 = 5 p):

Consider the following database schema

Country(Name, Code, Capital, Area, Population)

Organization(Name, Abbreviation, Established)

IsMember(Organization, Country, Joined)

The attribute *Organization* in the table IsMember is a foreign key reference to *Abbreviation* in the table Organization.

The attribute *Country* in table IsMember is a foreign key reference to *Code* in the table Country.

Examples of the tuples from the above relational schema are as follows:

Country(Sweden, SWE, Stockholm, 449964, 9514000)

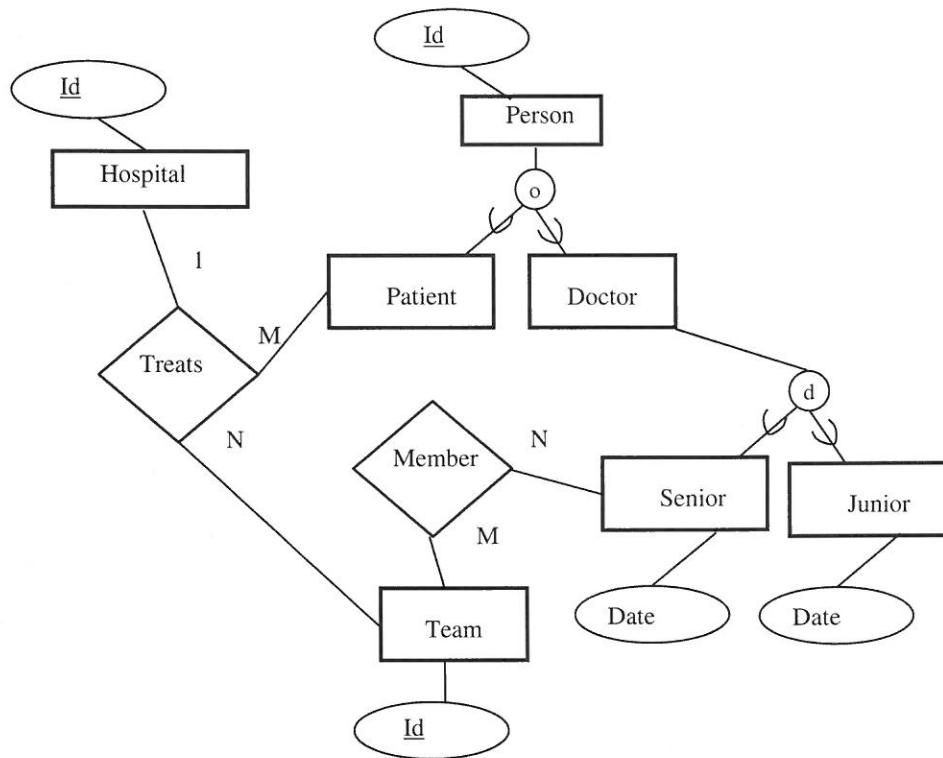
Organization(European Union, EU, 1952)

IsMember(EU, SWE, 1995-01-01)

1. List the number of organizations each country belongs to.
2. For each country and organization the country belongs to, list all the other countries that belong to the same organization.
3. For each organization, list the names of the countries that joined the organization first. Note that several such countries may exist. You can assume that dates can be compared with the operators <, >, and =.

Question 3. EER diagram and relational schema (5 p):

Translate the EER diagram below to a relational schema (use the algorithm seen in the course).



Theoretical part (19 points)

Question 4. Normalization (4 p):

Given the relation $R(A, B, C, D, E)$ with functional dependencies
 $\{AB \rightarrow C, C \rightarrow BD, D \rightarrow E, E \rightarrow D\}$.

1. Find all the candidate keys. Bear in mind that a relation can have several candidate keys.
2. Normalize (1NF \rightarrow 2NF \rightarrow 3NF \rightarrow BCNF). Explain your solution step by step.

Question 5. Data structures (2 + 3 = 5 p):

We have a file with 1000000 records. Each record is 40 bytes long. The records have two key attributes X and Y. The file is ordered on X. The database uses a block size of $B=40000$ bytes and unspanning allocation. Each index record is 4 bytes long.

1. Calculate the number of block access needed to find a record with a given value for X when using (a) the primary access method and (b) a single level index.
2. (a) Calculate the average number of block access needed to find a record with a given value for Y when using the primary access method.
(b) Calculate the number of block access needed to find a record with a given value for Y when using a single level index.
(c) Calculate the number of block access needed to find a record with a given value for Y when using static multi-level index.

Recall that $\log_2 2^x = x$. That is, $\log_2 1 = 0$, $\log_2 2 = 1$, $\log_2 4 = 2$, $\log_2 8 = 3$, $\log_2 16 = 4$, $\log_2 32 = 5$, $\log_2 64 = 6$, $\log_2 128 = 7$, $\log_2 256 = 8$, $\log_2 512 = 9$, $\log_2 1024 = 10$, $\log_2 2048 = 11$, $\log_2 4096 = 12$, $\log_2 8192 = 13$, $\log_2 16384 = 14$ etc.

Question 6. Transactions and concurrency control (1 + 1 + 1 = 3 p):

1. Give two schedules that are not serial but such that one is serializable and the other not. Justify your answer.
2. Apply the two-phase locking protocol to all the transactions in the two schedules you gave in exercise 1.
3. Show how the transactions in each of the two schedules in the exercise 2 (i.e. after you applied the two-phase locking protocol) are executed in a single-processor computer (i.e. show the ordering in which they are executed).

Question 7. Database recovery (3 + 1 = 4 p):

1. Describe the deferred update recovery method and the immediate update recovery method (versions 1 and 2) seen in the course.
2. Why do we not store read operations in the log file?

Question 8. Optimization (3 p):

Let *Student*, *Professor* and *Supervise* be three tables with 10 attributes respectively. The attribute *id* is the primary key of tables *Student* and *Professor* respectively. (*student_id*, *professor_id*) is the primary key of the table *Supervise*, where *student_id* is the foreign key to the *id* attribute in *Student* and *professor_id* is the foreign key to the *id* attribute in *Professor*. Show the canonical and the optimized query tree for the following MySQL query:

```
SELECT Student.name, Professor.name
FROM Student, Professor, Supervise
WHERE Student.salary>10000 Professor.salary>20000 and Professor.age -
Student.age < 10 and Student.id = Supervise.student_id and Professor.id =
Supervise.professor_id;
```