5

Institutionen för datavetenskap
Linköpings universitet

EXAM
TDDD37 Database Technology
TDDD46 Database Technology

January 17, 2015, 8.00-12.00

Help

Dictionary.

Grades

You can get max 30 points. To pass the exam, grade 3, you need 7.5 points in both the practical and theoretical part of the exam. For grade 4 and 5, you need 21 and 27 points, respectively.

Questions

Jose M. Peña will visit the room at 10.00.

Instructions

You can answer in Swedish or English. 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.

Good luck!

Practical part (15 points)

Question 1. Data modeling with EER diagram (4 + 1 = 5 p):

1. We want to create a database to store information about a company that has many departments. Each department is of exactly one out of three possible types. Every department has another department as supervisor. However, a supervisor department is always of the first or second type, and never of the third type. Note that not all the departments of first and second types necessarily supervise some other department. Since the supervisor department of a department may change from time to time, we want to store the date when the current supervisor department was appointed as such.

   Draw an EER diagram for the description above. Feel free to add the attributes that you consider necessary. Clearly write down your choices and assumptions in case you find that something in the information above is not clear.

2. Is the following statement true or false: All the entity types in an EER diagram must have a primary key ?

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

Consider the following relational 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:

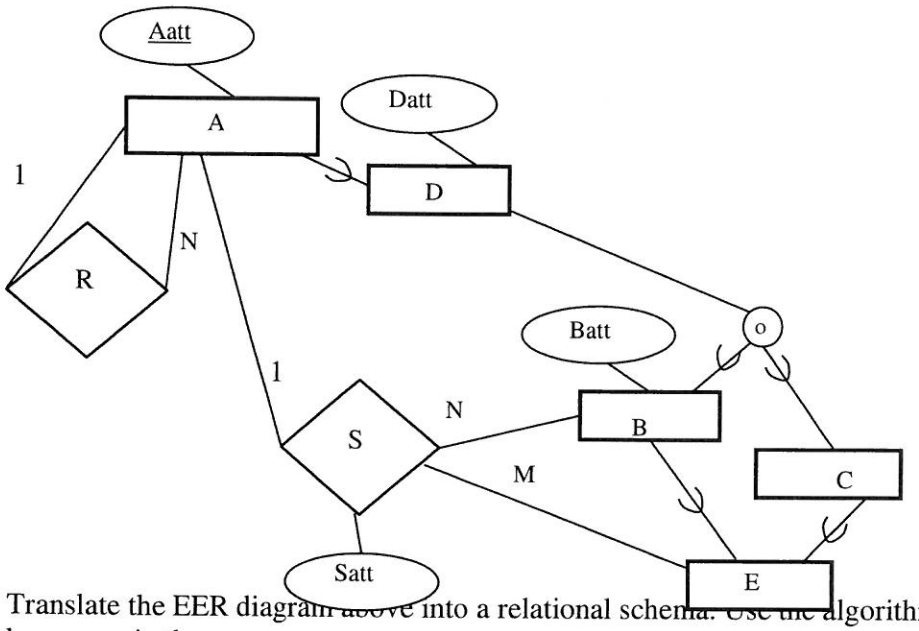
Country(Sweden, SWE, Stockholm, 449964, 9514000)
Organization(European Union, EU, 1952)
IsMember(EU, SWE, 1995-01-01)

1. List the names of all the organizations in ascending order of establishing date.

2. List the names of the organizations with a population of more than 100 millions. The population of an organization is the sum of the populations of its member countries.

3. For each organization, list the name of the country with smallest population in the organization.

Question 3. EER diagram and relational schema (4 + 1 = 5 p):



1. Translate the EER diagram above into a relational schema. Use the algorithm you have seen in the course.
2. In the course, you have seen two different ways of translating a 1-N relationship type. Describe them and discuss their relative advantages and disadvantages.

Theoretical part (15 points)

Question 4. Normalization (2 + 1 = 3 p):

1. Normalize up to Boyce-Codd normal form (BCNF) the relation R(A, B, C, D, E, F, G, H, I) with functional dependencies {ABC→DEFGHI, BC→DE, DE→FG, HI→BC}. Explain your solution step by step. Bear in mind that a relation can have several candidate keys.
2. Why do we normalize ?

Question 5. Data structures (2 + 2 = 4 p):

We have a file with 2000000 records. Each record is 20 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=4000 bytes and unspanning allocation. Each index record is 4 bytes long.

1. Calculate the average (or the maximum, if you prefer) 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. Calculate the average (or the maximum, if you prefer) number of block access needed to find a record with a given value for Y when using (a) the primary access method and (b) a single 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 + 2 = 3 p):

1. Give a schedule that is serializable but not serial. The schedule must have at least three transactions, and each transaction must have at least one write operation.

2. Describe the two-phase locking protocol (2PL) and what it is used for.

Question 7. Database recovery (3 p):

Consider the system log below:

Start-transaction T2
Write-item T2, B, 3, 4
Start-transaction T3
Write-item T3, A, 7, 8
Write-item T3, A, 8, 1
Write-item T3, A, 1, 5
Start-transaction T4
Write-item T4, B, 4, 5
Write-item T4, B, 5, 10
Start-transaction T1
Write-item T1, C, 8, 9
Write-item T1, C, 9, 10
Checkpoint

→system crash

Now, add somewhere in the system log above the instructions Commit T2, Commit T3, and Commit T4 so that recovery with deferred update performs as few operations as possible. Repeat the exercise for the other two recovery methods you have seen in the course.

Question 8. Optimization (2 p):

Consider the relational schema in Question 2. Show the canonical and the optimized query tree for the following SQL query:

```
SELECT Organization.Name
FROM Organization, IsMember
WHERE IsMmeber.Country = 'SWE' and Organization.Abbreviation =
IsMember.Organization;
```