

EXAM  
TDDD37 Database Technology  
TDDD46 Database Technology

April 10, 2015, 14.00-18.00

**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 16.00.

**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 (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.

**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.

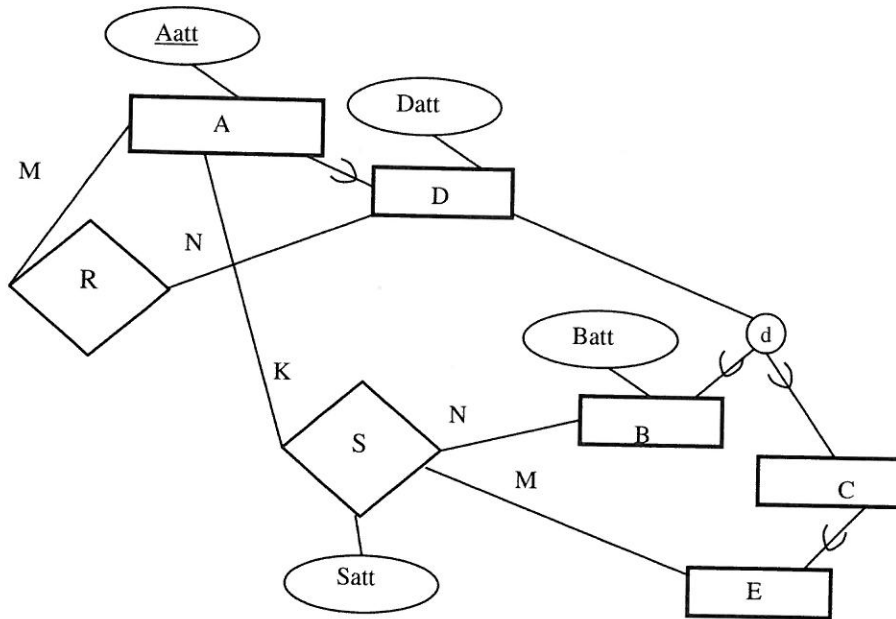
Examples of the tuples from the above relational schema are as follows:

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

1. List the names of all the countries in ascending order of population density. The population density is calculated as the population of the country divided by the area of the country.
2. List the names of the countries that belong to more than three organizations.
3. For each country, list the name of the oldest (i.e. first established) organization to which the country belongs.

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

Translate the EER diagram below into a relational schema. Use the algorithm you have seen in the course.



**Theoretical part (15 points)**

**Question 4. Normalization (3 p):**

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

**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 multilevel 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 (2 + 1 = 3 p):**

Consider the following schedule:

T1	T2	T3
read(x)		
x=x+1		
write(x)		
	read(x)	
	x=x+1	
	write(x)	
	read(z)	
	z=z+1	
	write(z)	
		read(z)
		z=z+1
		write(z)
		read(y)
		y=y+1
		write(y)
read(y)		
y=y+1		
write(y)		

1. Is the schedule above serializable? Justify your claim.
2. Apply the two-phase locking protocol to the schedule above.

**Question 7. Database recovery (3 p):**

Describe the three recovery methods you have seen in the course. Do not give examples but describe the methods.

**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;
```