Försättsblad till skriftlig tentamen vid Linköpings universitet

(fylls i av ansvarig)

Datum för tentamen	27/05/2013
Sal	T1, u1, Kåra, T2
Tid	14.00-18.00
Kurskod	TDDD12
Provkod	TEN1
Kursnamn/benämning	Databasteknik
Institution	*IDA*
Antal uppgifter som ingår i tentamen	7
Antal sidor på tentamen (inkl. försättsbladet)	5
Jour/Kursansvarig	Jose M. Peña, jose.m.pena@liu.se
Telefon under skrivtid	013 281651
Besöker salen ca kl.	15.00 och 16.30
Kursadministratör (namn + tfnnr + mailadress)	Madeleine Häger Dahlqvist, madeleine.hager.dahlqvist@liu.se, 013 282360
Tillåtna hjälpmedel	Ordbok
Övrigt (exempel när resultat kan ses på webben, betygsgränser, visning, övriga salar tentan går i m.m.)	
Vilken typ av papper ska användas, rutigt eller linjerat	
Antal exemplar i påsen	

Institutionen för datavetenskap
Linköpings universitet

EXAM
TDDD12 Databasteknik
TDDD46 Databasteknik

May 27, 2013, 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

Fang Wei-Kleiner and Jose M. Peña will visit the room at 15.00 and 16.30.

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 some companies, their employees and their jobs. Each company, employee and job can be classified as of type A or B. Companies of type A only have employees of type A, who only work in jobs of type A. Likewise, companies of type B only have employees of types B, who only work in jobs of type B. We want to create a database to store who works for which company in which job. Draw an EER diagram for such a database. Clearly write down your choices and assumptions in case you find that something in the information above is not clear.

2. What is the difference between an entity and an entity type ? And between a relationship and a relationship type ?

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)

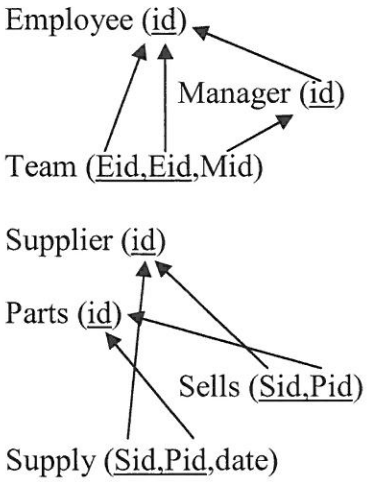
1. List the country names of all the European Union (Abbreviation: 'EU') members.

2. Compute the sum of the populations from all the 'EU' countries.

3. List the name of every country that is only the member of United Nations (Abbrevation: 'UN'). That is, the country is the member of 'UN', and is not the member of any other organisation.

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

Draw an EER diagram that, when translated using the algorithm seen in the course, may result in the following relational model.



Theoretical part (15 points)

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

1. Normalize (1NF→2NF→3NF→BCNF) the relation R(A, B, C, D) with functional dependencies {AB→CD, C→B, D→C}. Explain your solution step by step. Bear in mind that a relation can have several candidate keys.
2. Give an example of a relation that is in BCNF but not in 3NF. If this is not possible, explain why.

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

We have a file with 1000000 records. Each record is 50 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=10000 bytes and unspanning allocation. Each index record is 5 bytes long.

1. Calculate the average 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 number of block access needed to find a record with a given value for Y when using (a) the primary access method, (b) a single level index and (c) 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 (1.5 + 1.5 = 3 p):

Consider the following schedule:

T1	T2	T3
read(x)		
x=x+1		
write(x)		
		read(x)
		x=x+1
		write(x)
	read(x)	
	x=x+1	
	write(x)	
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 serializable? Justify your claim.

2. Does this schedule permits the two-phase locking protocol, i.e. can you apply the protocol so that the transactions interleave as in the schedule above ? Justify your answer.

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

1. Apply the recovery method for the three update methods seen in the course to the system log below. Show all operations (in the correct order) that are performed during the recovery.

   Part of system log:
   Start-transaction T1
   Start-transaction T2
   Start-transaction T3
   Start-transaction T4
   Start-transaction T5
   Write-item T1, B, 3, 4
   Commit T1
   Checkpoint
   Write-item T2, B, 3, 4
   Commit T2
   Checkpoint
   Write-item T3, B, 3, 4
   Commit T3
   Write-item T4, B, 3, 4
   →system crash

2. Assume that system crashes are rare and, thus, their influence on performance can be ignored. Assume that all the transactions always write on the same data item. Then, which update method is to be preferred and why ?