

EXAM
TDDD12 Databasteknik
TDDD81 Databasteknik för Kandidatprojekt

October 20, 2014, 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 (3 + 1 + 1 = 5 p):

1. We want to create a database to store information about people, companies and their jobs in these companies. Specifically, a person can work in several companies and have different jobs in the same or in different companies. A person can have another person as boss for each job he/she performs in a company.

Draw an EER diagram to represent as much information described above as possible. You are allowed to use only three regular entity types: Person, Company, and Job. Please, add the attributes that you consider necessary to the given entity types. Clearly write down your choices and assumptions in case you find that something in the information above is not clear.

2. What is a weak entity type ?
3. What is a subclass ?

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 name of the organization that was first established.
2. List the name of all the organizations of which Sweden is a member.
3. List the total area of each organization. The total area of an organization is the sum of the areas of its member countries.

Question 3. EER diagram and relational schema (3 + 2 = 5 p):

1. Draw an EER diagram whose translation results in the following relational schema:

A(Aatt, isD, Datt, RAatt, SAatt, Satt) where the foreign keys are to A itself.

B(Aatt, Batt) where Aatt is also a foreign key is to A.

C(Aatt) where Aatt is also a foreign key is to A.

2. Draw an EER diagram whose translation results in the following relational schema:

A(Aatt, isD, Datt, isB, Batt, isC, RAatt, SAatt, Satt) where the foreign keys are to A itself.

Theoretical part (15 points)

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

Given the relation R(A, B, C, D, E, F, G, H) with functional dependencies { $AB \rightarrow CDEFGH$, $CD \rightarrow B$, $D \rightarrow EFGH$, $E \rightarrow FGH$, $FG \rightarrow E$, $G \rightarrow H$ }.

1. Find all the candidate keys of R. Use the inference rules in the course to reach your conclusion. Do not use more than one rule in each derivation step.
2. Normalize R to 3NF. Explain the process step by step.

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 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 multilevel 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 and (b) 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) | | |
| read(z) | | |
| $z=z+x$ | | |
| write(z) | | |
| | read(x) | |
| | read(y) | |
| | | read(y) |
| | | $y=y+1$ |
| | | write(y) |
| | $y=y+x$ | |
| | write(y) | |
| read(y) | | |
| $y=y+1$ | | |
| write(y) | | |

1. Is the schedule above serializable ? Explain why or why not.
2. If the transactions in a schedule interleave the same before and after applying the two-phase locking protocol, then the schedule is serializable. True or false ?

Question 7. Database recovery (2 + 1 + 1 = 4 p):

1. Apply the three recovery 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 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

Commit T1

Commit T2

Checkpoint

Commit T4

→system crash

2. What does a commit instruction mean ?
3. What does a checkpoint instruction mean ?