Försättsblad till skriftlig tentamen vid Linköpings universitet

(fylls i av ansvarig)

Datum för tentamen	Oktober 17, 2011
Sal	G32-36
Tid	14.00-18.00
Kurskod	TDDD37
Provkod	TEN 1
Kursnamn/benämning	Databas teknik
Institution	*IDA*
Antal uppgifter som ingår i tentamen	8
Antal sidor på tentamen (inkl. försättsbladet)	6
Jour/Kursansvarig	Jose M. Peña
Telefon under skrivtid	013 281651
Besöker salen ca kl.	15.00 och 17.00
Kursadministratör (namn + tfnnr + mailadress)	Madeleine Hägg Dahlqvist 013 282360 madeleine.hagg.dahlqvist@liu.se
Tillåtna hjälpmedel	Ordbok
Övrigt (exempel när resultat kan ses på webben, betygsgränser, visning, övriga salar tentan går i m.m.)	G(3): 7.5+9 poäng VG(4): 23 poäng MVG(5): 29 poäng
Vilken typ av papper ska användas, rutigt eller linjerat	—
Antal exemplar i påsen	—

LiU, Linköping University
IDA, Department of Computer and Information Systems
Jose M. Peña
2011-10-06

EXAM

TDDD37 Database technology

17 October 2011, 14.00-18.00

Help

Dictionary.

Grades

You can get max 33 points. To pass the exam, grade 3, you need 7,5 and 9 points in the practical and theoretical part of the exam, respectively. For grade 4 and 5, you need 23 and 29 points, respectively.

Questions

During the exam, there is possibility of asking questions and clarifications from Jose M. Peña, tel. 013 281651, Patrick Lambrix, tel. 013 282605, He Tan 013 285841, and Fang Wei-Kleiner 013 284604. They will also visit the room at 15.00 och 17.00.

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 p)

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

Read the whole exercise before starting.

An online photo forum has much information which it must keep track of. Each of its members has a unique id, user name, registration date and the photography equipments he/she is using. Photography equipments include different cameras, lens and filter. Members can upload their own photos for sharing. Each photo has its unique id, title, exposure date, post date, category, and whether it is manipulated. The forum keeps track of the member who uploads a photo, and the photography equipments that were used for shooting the photo. Every member can comment on photos. Each comment has its post date. The forum invites professional photographers, such as journalists, to become their expert members. Each expert member has a biography. Every week one photo is selected as "*the photo of week*". A photo can be "*the photo of week*" only once. The forum invites one expert member to write a short review on "*the photo of week*". For "*the photo of week*" the forum stores the week number, the review and the reviewer.

Draw an EER diagram for the photo forum for the data described above.

Question 2. SQL (1 + 1.5 + 1.5 + 2 = 6 p):

Study the following relations describing suppliers of chips, customers and orders:

Supplier:

Id	**Name**
1	Estrella
2	OLW
3	Eldorado
…	…

Customer:

Id	**Name**
1	ICA
2	Hemköp
3	Willys
…	…

Order:

Supplier	**Customer**	**Date**	**Amount (kkr)**
1	1	16/10/2011	100
1	2	08/10/2011	95
2	1	08/10/2011	105
3	2	16/10/2011	100
3	3	08/10/2011	80
…	…	…	…

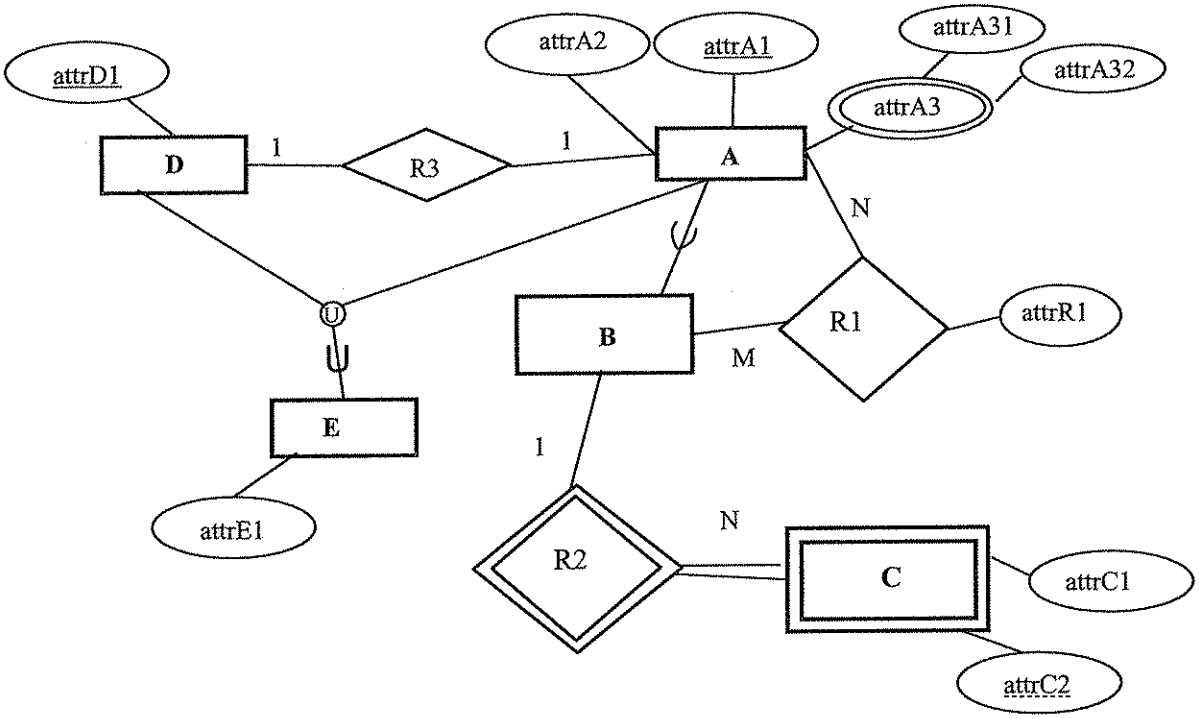
Supplier is a foreign key referring to Supplier (id)
Customer is a foreign key referring to Custom (id)

Write SQL queries for the following:

1. List the names of Suppliers that have at least one order with ICA.
2. List the names of Suppliers that do not have any order with ICA.
3. List the names of Suppliers that have more than 10 orders with ICA.
4. List the pairs of Suppliers that do not have orders with the same customer. For the data above we expect the following answer:

Supplier1	Supplier2
Eldorado	OLW
OLW	Eldorado

Question 3. Translation of EER to relational schema (4 p):



Translate the EER diagram above into a relational model (you have to follow the algorithm seen in the course). Mark the primary keys with solid underlining and the foreign keys with dotted underlining and an arrow from the foreign key to the attribute(s) pointed by the foreign key.

Theoretical part (18 points)

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

Given the relation R(A, B, C, D, E, F, G, H) with functional dependencies {AB→CDEFGH, CD→B, D→EFGH, E→FGH, FG→E, G→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 2NF. Explain the process step by step.
3. Why do we normalize relations ?

Question 5. Data structures (2 + 2 + 1 = 5 p):

We have a file with 30000 records. Each record is 5 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=100 bytes and unspanning allocation. Each index record is 4 bytes long.

1. Our goal is to perform at most 11 block accesses to find a record with a given value for X. Do we need to create a primary index or a static multilevel index to reach our goal ?
2. Our goal is to perform at most 11 block accesses to find a record with a given value for Y. Do we need to create a secondary index or a static multilevel index to reach our goal ?
3. What are B-trees and B+-trees ?

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

Question 6. Transactions and concurrency control (2 + 1 = 3 p):

1. Is the following transaction schedule serializable? Motivate your answer.

T1	T2	T3
	read(x)	
	x:=x+1	
	write(x)	
read(x)		
x:=x+1		
write(x)		

```
read(y)
y:=y+1
write(y)
```

```
                read(y)
                y:=y+1
                write(y)
```

```
read(y)
y:=y+1
write(y)
```

2. Does this schedule permit 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 p):

Apply the three recovery methods seen in the course to the system log below. Show all operations that are performed during the recovery. In the correct order!

Part of system log:

Start-transaction T1
Write-item T1, A, 5, 6
Start-transaction T2
Write-item T2, B, 2, 4
Write-item T2, B, 4, 7
Commit T1
Start-transaction T3
Write-item T3, A, 6, 8
Write-item T3, A, 8, 10
Write-item T2, B, 7, 2
Checkpoint
Start-transaction T4
Commit T2
Write-item T4, C, 1, 2

→system crash

Question 8. Optimization (1 + 1 + 1 = 3 p):

1. Let R(A,X), S(B, Y), and T(A, B, C) be three tables with the underlined attributes as keys. Optimize the following MySQL query:

```
SELECT *
FROM R, S, T
WHERE R.A = T.A AND S.B = T.B AND T.C > 50;
```

2. Assume that the tables do not contain any NULL value. Assume also that each table contains 1000 tuples and that each attribute is of size 4 byte. Show that the optimized query tree is more efficient than the canonical query tree.
3. Why does query optimization replace a selection followed by a Cartesian product with a join operation ?

LiU, Linköping University
IDA, Department of Computer and Information Systems
Jose M. Peña
2011-10-06

EXAM

TDDD37 Database technology

17 October 2011, 14.00-18.00

Help

Dictionary.

Grades

You can get max 33 points. To pass the exam, grade 3, you need 7,5 and 9 points in the practical and theoretical part of the exam, respectively. For grade 4 and 5, you need 23 and 29 points, respectively.

Questions

During the exam, there is possibility of asking questions and clarifications from Jose M. Peña, tel. 013 281651, Patrick Lambrix, tel. 013 282605, He Tan 013 285841, and Fang Wei-Kleiner 013 284604. They will also visit the room at 15.00 och 17.00.

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 p)

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

Read the whole exercise before starting.

An online photo forum has much information which it must keep track of. Each of its members has a unique id, user name, registration date and the photography equipments he/she is using. Photography equipments include different cameras, lens and filter. Members can upload their own photos for sharing. Each photo has its unique id, title, exposure date, post date, category, and whether it is manipulated. The forum keeps track of the member who uploads a photo, and the photography equipments that were used for shooting the photo. Every member can comment on photos. Each comment has its post date. The forum invites professional photographers, such as journalists, to become their expert members. Each expert member has a biography. Every week one photo is selected as "*the photo of week*". A photo can be "*the photo of week*" only once. The forum invites one expert member to write a short review on "*the photo of week*". For "*the photo of week*" the forum stores the week number, the review and the reviewer.

Draw an EER diagram for the photo forum for the data described above.

Question 2. SQL (1 + 1.5 + 1.5 + 2 = 6 p):

Study the following relations describing suppliers of chips, customers and orders:

Supplier:

Id	**Name**
1	Estrella
2	OLW
3	Eldorado
…	…

Customer:

Id	**Name**
1	ICA
2	Hemköp
3	Willys
…	…

Order:

Supplier	**Customer**	**Date**	**Amount (kkr)**
1	1	16/10/2011	100
1	2	08/10/2011	95
2	1	08/10/2011	105
3	2	16/10/2011	100
3	3	08/10/2011	80
…	…	…	…

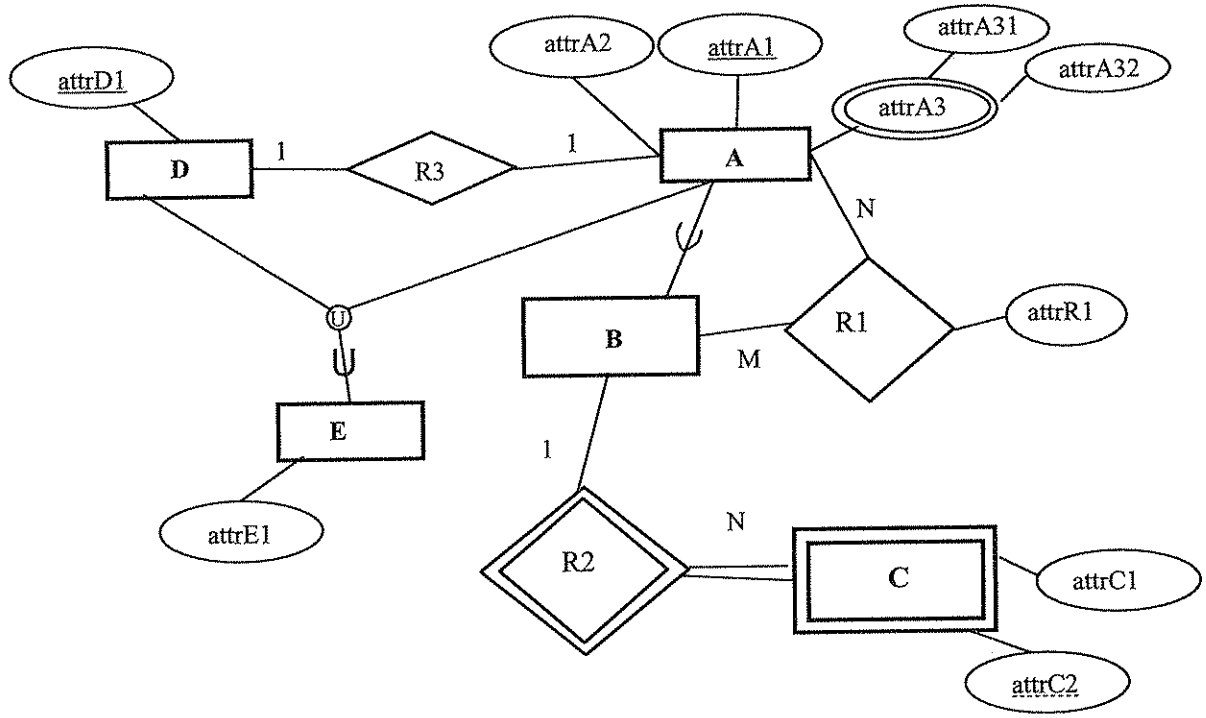
Supplier is a foreign key referring to Supplier (id)
Customer is a foreign key referring to Custom (id)

Write SQL queries for the following:

1. List the names of Suppliers that have at least one order with ICA.
2. List the names of Suppliers that do not have any order with ICA.
3. List the names of Suppliers that have more than 10 orders with ICA.
4. List the pairs of Suppliers that do not have orders with the same customer. For the data above we expect the following answer:

Supplier1	Supplier2
Eldorado	OLW
OLW	Eldorado

Question 3. Translation of EER to relational schema (4 p):



Translate the EER diagram above into a relational model (you have to follow the algorithm seen in the course). Mark the primary keys with solid underlining and the foreign keys with dotted underlining and an arrow from the foreign key to the attribute(s) pointed by the foreign key.

Theoretical part (18 points)

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

Given the relation R(A, B, C, D, E, F, G, H) with functional dependencies {AB→CDEFGH, CD→B, D→EFGH, E→FGH, FG→E, G→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 2NF. Explain the process step by step.
3. Why do we normalize relations ?

Question 5. Data structures (2 + 2 + 1 = 5 p):

We have a file with 30000 records. Each record is 5 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=100 bytes and unspanning allocation. Each index record is 4 bytes long.

1. Our goal is to perform at most 11 block accesses to find a record with a given value for X. Do we need to create a primary index or a static multilevel index to reach our goal ?
2. Our goal is to perform at most 11 block accesses to find a record with a given value for Y. Do we need to create a secondary index or a static multilevel index to reach our goal ?
3. What are B-trees and B+-trees ?

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

Question 6. Transactions and concurrency control (2 + 1 = 3 p):

1. Is the following transaction schedule serializable? Motivate your answer.

T1	T2	T3
	read(x)	
	x:=x+1	
	write(x)	
read(x)		
x:=x+1		
write(x)		

```
read(y)
y:=y+1
write(y)

                    read(y)
                    y:=y+1
                    write(y)

read(y)
y:=y+1
write(y)
```

2. Does this schedule permit 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 p):

Apply the three recovery methods seen in the course to the system log below. Show all operations that are performed during the recovery. In the correct order!

Part of system log:

Start-transaction T1
Write-item T1, A, 5, 6
Start-transaction T2
Write-item T2, B, 2, 4
Write-item T2, B, 4, 7
Commit T1
Start-transaction T3
Write-item T3, A, 6, 8
Write-item T3, A, 8, 10
Write-item T2, B, 7, 2
Checkpoint
Start-transaction T4
Commit T2
Write-item T4, C, 1, 2

→system crash

Question 8. Optimization (1 + 1 + 1 = 3 p):

1. Let R(A,X), S(B, Y), and T(A, B, C) be three tables with the underlined attributes as keys. Optimize the following MySQL query:

```
SELECT *
FROM R, S, T
WHERE R.A = T.A AND S.B = T.B AND T.C > 50;
```

2. Assume that the tables do not contain any NULL value. Assume also that each table contains 1000 tuples and that each attribute is of size 4 byte. Show that the optimized query tree is more efficient than the canonical query tree.
3. Why does query optimization replace a selection followed by a Cartesian product with a join operation ?