Örebro universitet
Institutionen för teknik
Thomas Padron-McCarthy (Thomas.Padron-McCarthy@tech.oru.se)

Tentamen i Databaser, fortsättningskurs för D3 m fl

onsdag 29 augusti 2007 kl 08:00 - 12:00 i L001

Gäller som tentamen för:
TDD119 Databaser, fortsättningskurs, provkod 0100

Hjälpmedel:	Inga hjälpmedel.
Poängkrav:	Maximal poäng är 44. För godkänt betyg (3 respektive G) krävs 22 poäng.
Resultat och lösningar:	Meddelas på kursens hemsida senast måndag 10 september 2007.
Visning och frågestund:	Tisdag 11 september 2007 kl 12:00-12:30 i mitt rum (T2220). Efter visningen kan tentorna hämtas på expeditionen. Man kan också få sin rättade tenta hemskickad.
Examinator och jourhavande:	Thomas Padron-McCarthy, telefon 070-73 47 013.

- Skriv tydligt och klart. Lösningar som inte går att läsa kan naturligtvis inte ge några poäng. Oklara och tvetydiga formuleringar kommer att misstolkas.
- Skriv den personliga tentamenskoden på varje inlämnat blad. Skriv *inte* namn eller personnummer på bladen.
- Skriv bara på en sida av papperet. Använd inte röd skrift.
- Antaganden utöver de som står i uppgifterna måste anges.
- Skriv gärna förklaringar om hur du tänkt. Även ett svar som är fel kan ge poäng, om det finns med en förklaring som visar att huvudtankarna var rätt.

LYCKA TILL!

Uppgift 1 (5 p)

Du ska skriva ett datorprogram som ska köras av vanliga användare på vanliga skrivbordsdatorer. Programmet ska arbeta med data som finns i en databas. Vi vill kunna byta databashanterare. Till att börja med körs programmet på Linux-datorer, men så småningom vill vi helst kunna flytta programmet till andra typer av datorer.

Det finns olika sätt för ett program att kommunicera med en databas. Vilket kan vara lämpligt i det här fallet? Diskutera några olika alternativ, och förklara varför de är, eller inte är, lämpliga.

Uppgift 2 (10 p)

En databas innehåller data om **Bilar**, och de **Personer** som äger bilarna. Varje bil ägs av en enda person.

Bilar

Bnr	**Märke**	**Ägare**
RFN540	Renault	2
ABC802	Luxor	2
BOK118	Saab	1

Personer

Pnr	**Namn**	**Stad**
1	Gustav	Örebro
2	Thomas	Örebro
3	Olle	Åmål
4	Margareta	Örebro

Databasen ovan beskriver världen som den ser ut vid ett visst tillfälle, men saker ändras ju med tiden. Till exempel kan bilar tillkomma, skrotas och säljas, och personer kan få nya telefonnummer.

a) (3p)

Vi vill kunna lagra data i databasen, inte bara om hur världen ser ut just nu, utan också om hur den sett ut tidigare. (Vem var det till exempel som ägde bilen RFN540 den 1 januari 2001?) Visa hur man kan komplettera databasen för att det ska fungera!

b) (1p)

En databas med en tidsdimension, som i uppgift a, brukar kallas för en **temporal databas**. Vad brukar man kalla en "vanlig" databas, som beskriver världen som den ser ut vid ett visst tillfälle?

c) (2p)

I originaldatabasen är kolumnen **Bilar.Ägare** ett referensattribut, som refererar till kolumnen **Personer.Pnr**. För att få referensintegritet deklarerar vi den förstås som en **foreign key** (främmande nyckel) med hjälp av ett SQL-kommando. Hur fungerar referensintegriteten i den nya databasen från deluppgift a? Blir det några särskilda problem?

d) (3p)

Vi vill kunna behålla gamla uppgifter i databasen, när uppgifterna ändras. Visa hur man kan komplettera databasen för att det ska fungera!

e) (1p)

Delfråga a och d handlar om två olika tidsdimensioner i databasen. Vad kallas de två tidsdimensionerna?

Uppgift 3 (4 p)

En annan form av databas där man bland annat behåller gamla uppgifter är ett **datalager** (på engelska **data warehouse**). Vilka likheter och skillnader finns det mellan ett datalager och en temporal databas?

Uppgift 4 (10 p)

Utgå från den ursprungliga databasen med bilar och personer, i uppgift 2. Följande fråga ställs mot den databasen:

```
select Namn
from Bilar, Personer
where Bilar.Ägare = Personer.Pnr
and Bilar.Märke = 'Renault'
and Personer.Stad = 'Örebro'
```

Vi vill alltså veta namnen på de personer i Örebro som äger (minst) en Renault.

a) (2p)

Visa den systematiska översättningen till relationsalgebra, även kallad "kanonisk form", av SQL-frågan.

b) (5p)

Visa hur en heuristisk frågeoptimerare kan optimera den kanoniska formen från deluppgift a. Visa både vilka olika optimeringar som görs, och vad slutresultatet blir.

c) (3p)

En kostnadsbaserad frågeoptimerare kan ta hänsyn till en del saker som en heuristisk frågeoptimerare normalt inte bryr sig om. Vilka? Förklara, med exempel från databasen och frågan i exemplet!

Uppgift 5 (8 p)

I en databashanterare körs två transaktioner (**T1** och **T2**) enligt följande tidsschema. Transaktionerna arbetar med tre olika dataobjekt (**X**, **Y** och **Z**).

Tid	T1	T2
1	Start	
2	Läs(X)	
3		Start
4		Läs(X)
5	Skriv(X)	
6		Läs(Y)
7		Läs(Z)
8		Skriv(Y)
9		Skriv(Z)
10		Commit
11	Commit	

a) (2p)

Här användes ingen metod för att hålla transaktionerna isolerade från varandra. Kan det ha uppstått fel i databasen, efter just detta tidschema. I så fall vilka?

b) (6p)

Välj *två* av följande metoder för isolering av transaktioner, och visa hur tidsschemat ovan skulle förändras om respektive metod användes. (Observera: Välj alltså *två* av dessa metoder!)

- Tvåfaslåsning med enkla binära lås
- Tvåfaslåsning med läs- och skrivlås
- Tidsstämpelmetoden, utan Thomas skrivregel (Thomas's Write Rule)
- Tidsstämpelmetoden, med Thomas skrivregel (Thomas's Write Rule)
- En optimistisk metod som i Mimer

Uppgift 6 (7 p)

Vi skapar en tabell som ska innehålla data om stjärnor:

```
create table star
(id integer,
x float, y float, z float,
namn char(20));
```

Vintergatan innehåller ungefär 200 miljarder stjärnor, dvs 200 * 10^9. Vi lagrar dem i tabellen med hjälp av 200 miljarder **insert**-kommandon i SQL. Det finns till att börja med inga index.

a) (4p)

Ungefär hur lång tid tar det att köra följande fråga? Gör rimliga antaganden om olika saker, och ange dessa antaganden. Förklara också hur du räknat.

```
select x, y, z
from star
where namn = 'Betelgeuse';
```

b) (3p)

Vi skapar ett index:

```
create index starnames on star(namn);
```

Det index som skapas är ett sekundärindex i form av ett B^+-träd. Ungefär hur lång tid tar det nu att köra frågan från a-uppgiften? Gör även här rimliga antaganden om olika saker, och ange dessa antaganden. Förklara också hur du räknat.