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

Tentamen i

Databaser, fortsättningskurs

för D3 m fl

lördag 24 mars 2007 kl 08:00 - 12:00 i L001

Hjälpmedel:	Inga hjälpmedel.
Poängkrav:	Maximal poäng är 43. För betyget 3 krävs 21 poäng.
Resultat och lösningar:	Meddelas på kursens hemsida senast tisdag 3 april 2007.
Visning:	Tisdag 3 april 2007 kl 12:00-12:30 i mitt rum (T2220). Efter visningen kan tentorna hämtas på expeditionen.
Examinator och jourhavande:	Thomas Padron-McCarthy, telefon 070-7347013.

- Skriv tydligt och klart. Lösningar som inte går att läsa kan naturligtvis inte ge några poäng. Oklara formuleringar kommer att misstolkas.
- Skriv den personliga tentamenskoden på varje inlämnat blad. Skriv *inte* namn eller personnummer på bladen.
- Lös bara en uppgift per blad. Skriv bara på en sida av papperet. Använd inte röd skrift.
- Antaganden utöver de som står i uppgifterna måste anges. Gjorda antaganden får inte förändra den givna uppgiften.
- 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 (6 p)

Förklara kort vad följande är:

a) (1p) ODBC
b) (1p) ESQL
c) (1p) ADO.NET

d) (2p) Vilka är de största skillnaderna mellan ODBC och ADO.NET?

Uppgift 2 (4 p)

En **temporal databas** är en databas som innehåller en tidsdimension. Man brukar tala om två olika tidsdimensioner, som man kan använda i databaser. Vad kallas de, och vad betyder de? (Förklara med exempel.)

Uppgift 3 (3 p)

Vilka är de största skillnaderna mellan objektorienterade databaser (ibland kallade "första generationens objektorienterade databaser") och objektrelationella databaser (ibland kallade "andra generationens objektorienterade databaser")?

Uppgift 4 (5 p)

Förklara kort följande termer:

a) datautvinning ("data mining")
b) datalager ("data warehouse")
c) dimensionsdatabas ("dimensional database")
d) faktatabell
e) dimensionstabell

Uppgift 5 (4 p)

Visa med ett exempel hur tvåfaslåsning med läs/skriv-läs fungerar.

Uppgift 6 (6 p)

Vi skapar en tabell som ska innehålla data om sandkorn:

```
create table sandkorn
(id integer,
x float, y float, z float,
vikt float);
```

En typisk sandstrand innehåller 10 biljoner sandkorn, dvs 10^{13} stycken, som vi lagrar i tabellen med hjälp av 10 biljoner **insert**-kommandon i SQL. (Databasen lagras på en riktigt, riktigt stor disk.) Det finns till att börja med inga index.

a) 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 * from sandkorn
where id = 1000000;
```

b) Vi skapar ett index:

```
create index sid on sandkorn(id);
```

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.

Uppgift 7: Relationsalgebra, frågeexekvering, frågeoptimering (15 p)

En databas innehåller tabellerna **A** (som i **Arbetare**), **P** (som i **Projekt**) och **AP** (som i **Arbetar På**):

A

Anr	Anamn	Lön	Telefon
1	Angus	29000	1710
2	Bruce	25000	1711
3	Cyril	32400	1029
...	...	...	...

AP

A	P
1	1
1	2
2	2
3	5
...	...

P

Pnr	Pnamn	Adress
1	Apollo	Houston
2	Manhattan	Los Alamos
...	...	...

Angus arbetar på Apollo-projektet och Manhattan-projektet, Bruce arbetar på Manhattan-projektet, och så vidare. Det finns tusentals projekt och många tusen personer.

Vi vill vilka av de personer som (ibland) arbetar i Gnesta som tjänar mer än 30000 kronor, och ställer följande SQL-fråga:

```
select Anr, Anamn
from A, AP, P
where A.Anr = AP.A
and AP.P = P.Pnr
and A.Lön > 30000
and P.Adress = 'Gnesta';
```

a) (2p) Visa den systematiska översättningen till relationsalgebra, även kallad "kanonisk form", av SQL-frågan.

b) (1p) Vi antar att databashanteraren arbetar internt med relationsalgebra. Förklara skillnaden mellan att köra relationsalgebran med materialiserade mellanresultat och med strömmade mellanresultat ("pipe-lining").

c) (2p) Skulle den kanoniska formen vara ett bra sätt att köra frågan? Motivera svaret, såväl för en körning med materialiserade mellanresultat som med strömmade mellanresultat.

d) (1p) Vad innebär frågeoptimering i databaser?

e) (1p) Vad innebär heuristisk optimering?

f) (5p) Visa hur en heuristisk optimerare kan optimera den kanoniska formen från deluppgift a. Visa både vilka olika optimeringar som görs, och vad slutresultatet blir.

g) (1p) Vad innebär kostnadsbaserad optimering, och hur skiljer sig den från heuristisk optimering?

h) (2p) Vilka fördelar och nackdelar finns med kostnadsbaserad optimering, jämfört med heuristisk optimering?