

Örebro universitet
Akademin för naturvetenskap och teknik
[Thomas Padron-McCarthy \(Thomas.Padron-McCarthy@oru.se\)](mailto:Thomas.Padron-McCarthy@oru.se)

Tentamen i Databasteknik II för D3 m fl

lördag 4 april 2009 kl 08:15 - 12:15

Gäller som tentamen för:
DT3001 Datateknik C, Databasteknik II, provkod 0100
TDD119 Databaser, fortsättningskurs, provkod 0100

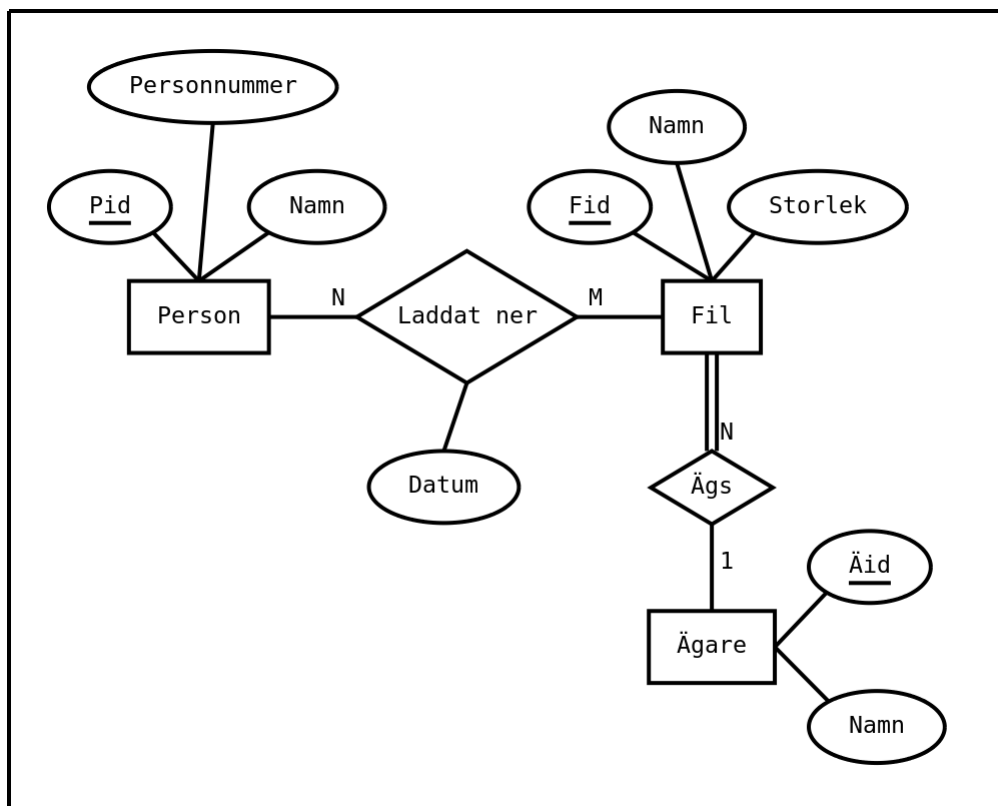
Hjälpmedel:	Miniräknare.
Poängkrav:	Maximal poäng är 40. För betyget 3 krävs 20 poäng.
Resultat och lösningar:	Meddelas via e-post eller på kursens hemsida, http://basen.oru.se/kurser/db2/2008-2009-p3/ , senast lördag 25 april 2009.
Återlämning av tentor:	Efter att resultatet meddelats kan tentorna hämtas på institutionen. 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. 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!

Scenario till uppgifterna

Antipiratbyrån, vars främsta uppgift är att värna upphovsrätten till filmer och annan media, gläder sig åt de utökade möjligheter de fått att registrera personer som laddar ner filer. De har nu byggt en databas för att hålla reda på filer, nerladdningar och misstänkta personer. Här är ett ER-diagram för Antipiratbyråns databas:



Här är en implementation av Antipiratbyråns databas i form av en relationsdatabas:

Tabellen **Personer**
10 miljoner rader

Pid	Personnummer	Namn
1	621010-2219	Olle Olsson
2	170114-1440	Lotta Lottén
3	081213-1658	Kasper Svensson
...

Tabellen **LaddatNer**
100 miljoner rader

Person	Fil	Datum
3	7181	2009-04-03
1	1	2009-04-01
3	1	2009-04-03
...

Tabellen **Filer**
10 miljoner rader

Fid	Namn	Storlek	Ägare
1	OfficeXP.zip	3554287616	13
2	CoqRouge.doc	67190208	3
3	SeekAmy.mp3	8709120	1
...

Tabellen **Ägare**
100 rader

Äid	Namn
1	Sony Music
2	Warner Brothers
3	Jan Guillou
...	...

Uppgift 1 (10 p)

Antipiratbyrån vill veta vilka som idag (**2009-04-04**) laddat ner någon fil som tillhör **Sony Music**, och ställer därför följande SQL-fråga:

```
select Personer.Personnummer, Personer.Namn
from Personer, LaddatNer, Filer, "Ägare"
where Personer.Pid = LaddatNer.Person
and LaddatNer.Fil = Filer.Fid
and Filer."Ägare" = "Ägare"."Äid"
and "Ägare".Namn = 'Sony Music'
and LaddatNer.Datum = '2009-04-04';
```

- (2p) Ange den systematiska översättningen av SQL-frågan till relationsalgebra, även kallad kanonisk form. (Det är alltså den o-optimerade formen.) Skriv upp den som ett relationsalgebrauttryck.
- (1p) Rita upp samma relationsalgebrauttryck som ett frågeträd.
- (4p) Visa hur en heuristisk frågeoptimerare optimerar den kanoniska formen från frågan ovan. Visa både vilka olika optimeringar som görs, och vad slutresultatet blir.
- (1p) Hur fungerar en kostnadsbaserad frågeoptimerare?
- (2p) 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 2 (6 p)

Antipiratbyrån vill söka efter en misstänkt nerladdare med ett visst personnummer, exempelvis så här:

```
select Namn from Personer where Personnummer = '081213-1658';
```

Databashanteraren arbetar med en blockstorlek på 2048 byte, och det tar i genomsnitt 10 millisekunder att hitta och läsa ett godtyckligt block från disken. Gör rimliga antaganden i övrigt (och ange dem!) och räkna sen ut hur lång tid det i genomsnitt tar att köra denna fråga, om tabellen är lagrad:

- (2p) enbart med ett primärindex i form av ett B⁺-träd på **Id**
- (2p) med ett primärindex enligt a-uppgiften, och dessutom med ett sekundärindex i form av ett B⁺-träd på **Personnummer**.
- (2p) Antipiratbyrån kommer att arbeta mycket flitigt, och göra flera hundra sådana sökningar per sekund. Är det troligt att deras databas klarar detta? Motivera svaret!

Visa (i alla deluppgifterna) hur du räknat!

Uppgift 3 (5 p)

I en databashanterare körs tre olika transaktioner (**T1**, **T2** och **T3**) enligt följande tidsschema. Transaktionerna arbetar med tre olika dataobjekt (**X**, **Y** och **Z**).

Tid	T1	T2	T3
1	Start		
2		Start	
3		Läs(Y)	
4		Läs(Z)	
5	Läs(X)		
6			Start
7			Läs(Z)
8			Läs(Y)
9	Skriv(X)		
10	Commit		
11		Skriv(Y)	
12		Skriv(Z)	
13		Commit	
14			Skriv(Z)
15			Commit

Precis före **commit**-operationen på rad 13 går plötsligt strömmen.

a) (2p) Hur ser loggfilen ut då?

b) (3p) Vilka operationer kommer att utföras under recovery-processen, när databasen återstartas?

Uppgift 4 (7 p)

a) (1p) Är tidsschemat i uppgiften ovan serialiserbart? Motivera svaret!

b) (1p) Och vad innebär egentligen serialiserbarhet?

c) (5p) 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 5 (6 p)

- a) (2p) Vad är ett datalager (på engelska data warehouse), och hur skiljer det sig från en vanlig databas?
- b) (2p) Vad är ett stjärnschema, och varför använder man det?
- c) (2p) Vad är en dimensionstabell, och varför uppfyller den ofta bara andra normalformen?

Uppgift 6 (6 p)

Här är sex olika tekniker för att kommunicera med en relationsdatabas från ett vanligt program:

- ODBC
- ESQL
- JDBC
- ADO.NET
- databasegna ("native") gränssnitt
- objekt-relationella översättningsramverk som till exempel Hibernate

(Försök) ange en stor fördel med var och en av dessa, jämfört med de andra.
