

Linköpings Universitet  
Institutionen för datavetenskap  
Eva Ragnemalm

# TENTAMEN

## 732G16 Databaser Design och programmering

**Datum: 2019-06-07 kl 8-12**

**Jourhavande lärare: Eva Ragnemalm**

**Tel: 070-1907391**

**Besöker tentamenslokalen: c:a kl 10**

**Hjälpmedel: ordbok för översättning från svenska till modersmål, vid behov.**

**Följ instruktionerna på tentaomslaget. Svara på EN uppgift per papper.**

Uppgift	Poäng
1	8
2	8
3	2
4	8
5	1
6	6
7	20
8	8
9	12

Betyg: G: 38p VG: 56p Max: 73p

1. **Begrepp:** För varje nedanstående påstående, är det sant eller falskt? Rätt svar ger 1p. Fel svar ger -1p men du kan inte få mindre än 0p på uppgiften som helhet. (8p)

- Ett ER-diagram är en *konceptuell datamodell*.
- Ett *referensattribut* är ett villkor som gäller för ett attribut, t.ex. att ett lagersaldo inte får ha ett negativt värde.
- Ett *fullt funktionellt beroende* är ett *funktionellt beroende* där man inte kan ta bort något attribut ur *determinanten* och fortfarande ha ett funktionellt beroende.
- Att en transaktion måste avbrytas och återställa tidigare värden kallas *rollback*.
- En *checkpoint* i *loggfilen* betyder att datat som pekats ut kontrollerats och är korrekt beräknad.
- Ett *primärindex* har alltid lika många indexposter som datafilen har block på disken.
- I en fil som är sorterad på det attribut man söker på, måste man alltid söka genom halva filen för att hitta den sökta posten.
- En anledning till att vi inte lagrar databaser bara i datorns primärminne är att då uppfylls inte D i ACID.

2. **Relationsmodellen, SQL:** antag att du har nedanstående tabeller:

Vara

ArtNr	Namn	Antal	Leverantor
1	Jackor med fler fickor än du kan hitta	20	1
2	Trebenta byxor	150	3
3	Bapelsiner	35	4
4	Prins Korv	50	4
5	Drottning Sylt	10	4
6	Molljoxer	250	3
7	Mojängar	55	3

Leverantor

IDnr	Namn	Adress	Land
1	Inteenfemma	Majvägen 5	Sverige
2	Tidig	Köbenhavnsvej 2	Danmark
3	WaitTulip	Allroad	USA
4	Kungsverket	Galavägen	Sverige

Tabellerna är skapade med följande kommandon:

```
create table Levarantor (IDnr integer, Namn varchar (25), Adress varchar (30), Land varchar (15), Primary key (IDnr)) Engine=InnoDB;
```

```
create table Vara (ArtNr integer, Namn varchar (50), Antal integer, Leverantor integer, Primary key (ArtNr), constraint varuLeverantor
```

```
Foreign key (Leverantor) references Leverantor (IDnr) Engine=In-
noDB;
```

Utgå från ovanstående tabeller.

(8p)

- a. Skriv ett SQL-kommando som listar namnet på alla varor och för varje vara även namnet på leverantören, för alla leverantörer från Sverige. (4p)
- b. Skriv ett SQL-kommando som listar leverantörernas IDnr och för varje leverantör hur många olika sorters varor de levererar (t.ex. nr 4 levererar 3 olika varor: Bapelsiner, Prins Korv och Drottning Sylt). (4p)

**Kombinationsuppgift (Design, Begrepp, Fysiska databasen, Normalisering): följande innehåller information till uppgift 3-7.**

En forskningsintresserad bonde vill undersöka vilka gödselblandningar som passar vilka växter och växtplatser som hon odlar. På nordöstra Lambohovs-fältet odlar hon spelt-vete, Trinitum spelta, och gödslar med blandning 38, men på det södra Labohovsfältet är det mindre sol, så där har hon på prov gödslat med mer fosfor, blandning 39 fastän det är spelt hon odlar där också. Det moderna brödvetet (Trinitum aestivum) som hon odlar i Slaka norra har hon provat att gödsla med mindre kväve, blandning 21. Hon bokför noggrant hur mycket utsäde hon använde på vilket fält och hur stor skörden sedan blir. Hon vill inte använda samma kombination av gröda och gödsel på samma fält men skulle hon råka göra det vill hon att bara det bästa resultatet sparas. Nu vill även några bond-kollegor på andra orter vara med och undersöka detta så de behöver dokumentera detta i några år och försöka hitta sambanden mellan gödsling, gröda och odlingsplats.

För de olika gödselblandningarna måste man hålla reda på hur stor andel fosfor, kväve och kalk de innehåller, samt ytterligare spårämnen som finns i signifikant mängd. Antalet signifikanta spårämnen kan variera mellan olika blandningar och man anger ingen mängd utan bara att de finns. Varje blandning har också en beskrivande text för att bonden ska komma ihåg vad syftet med blandningen är.

Gröda har ett botaniskt namn (det latinska namnet för växten, t.ex. Trinitum spelta). Varje gröda har också ett vardags-namn, det namn man i dagligt tal kallar växten. Grödan har också en beskrivande text. Vardags-namnen är inte säkert unika, både spelt och dinkel kallas gammelvete.

Odlingsplatserna är bondens olika fält och har namn som Södra Lambohovsfältet eller Slaka norra. Varje fält har en mark-karaktär (lerjord, sandjord osv), belyningskaraktär (helsoligt, halvsoligt osv) och en gps-koordinat t.ex. N58.23.432 E15.32.321 (gps-koordinater består alltid av två delar: nord/syd resp öst/väst som i sin tur delas upp i grader, minuter och sekunder) som anger var jordbruksredskap kan köra in på fältet (oftast ett hörn vid en väg). GPS-koordinaten fungerar inte som identifierare eftersom noggrannheten inte räcker för att skilja två fält som man når från samma infart.

Den första programmeraren hon konsulterade för att göra en förstudie, en student från Linköpings universitet, föreslog att allt skulle lagras i en enda dataframe för att det skulle bli enkelt att utföra olika beräkningar. Den data frame han föreslog såg ut så här:

```
Odling1(botanisktNamn, växtbeskrivning, vanligtNamn, fält, mark-
typ, ljustyp, GPS, blandningsID, blandningsbeskrivning, fosfor-
mängd, kvävemängd, kalkmängd, spårämnen, såddMängd, skördadMängd)
```

3. Vilken normalform uppfyller ovannämnda data frame, Odling1? Varför? (2p)

Efter en tid blev data ramen så stor att den tog för lång tid att ladda in och bonden frågade om man inte kunde lyfta ut informationen om gödselblandningarna i en egen data frame, skulle det inte bli mindre då? Programmeraren delade då upp informationen i följande två data frames:

```
Gödsel(blandningsID, blandningsbeskrivning, fosformängd,
kvävemängd, kalkmängd, spårämnen)
```

```
Odling2(botanisktNamn, växtbeskrivning, vanligtNamn, fält,
marktyp, ljustyp, GPS, blandningsID, såddMängd, skördadMängd)
```

4. Posterna (=raderna=observationerna) i den gamla data ramen Odling1 tog 500 byte per styck och i den nya Gödsel tar de 300 byte och i den nya Odling2 300 byte. Programmeraren blev väldigt paff när han såg att de två nya filerna tillsammans ändå tog mindre plats totalt. Den ursprungliga filen Odling1 hade 12 000 rader, den nya Odling2 har 12 000 rader och Gödsel 60 rader. Bonden har en hårddisk med accesstid på 10 ms och 2048 byte stora block. Räkna som om dessa data-frames lagras som tabeller i en databas. (8p)

- Hur många block tog den ursprungliga filen upp?
- Hur många block tar de två nya upp tillsammans?

5. Bonden insåg när hon tittade i data-ramen Gödsel att det fanns ett par blandningar som hade samma blandningsID men olika fosformängd. Uppenbarligen hade hon skrivit fel när hon skulle mata in informationen. Vad kallas det när samma sak händer i en databas? (1p)

6. Betrakta Odling2 enligt ovan som en relation, med den enda kandidatnyckeln {botanisktNamn, fält, blandningsID}. Identifiera fulla funktionella beroenden och rita ut som vi gjort i kursen (med pilar). Normalisera Odling2 så att den uppfyller 3NF. Motivera varje steg. (6p)

När man sedan skulle utvidga denna datalagring för att inkludera andra bönders gårdar upptäckte programmeraren att bonden Pettson också har ett fält som heter Slaka norra. Fält kan alltså heta likadant men tillhöra olika gårdar. Gårdarna i sig har namn, adress och en ägare, vars namn och telefonnummer man vill lagra. Inom varje bondgård är fältens namn unika.

7. Nu vill bonden ha en riktig databas för detta problem. Gör en databasdesign som minimerar utrymmesbehovet samtidigt som all information som beskrivs ovan bevaras (dvs all information som representeras i Odling1 ska finnas med, samt information om gårdarna enligt ovan). Rita upp ett ER-diagram, markera nycklar, kardinalitet och deltagande som vi gjort i kursen. (20p)

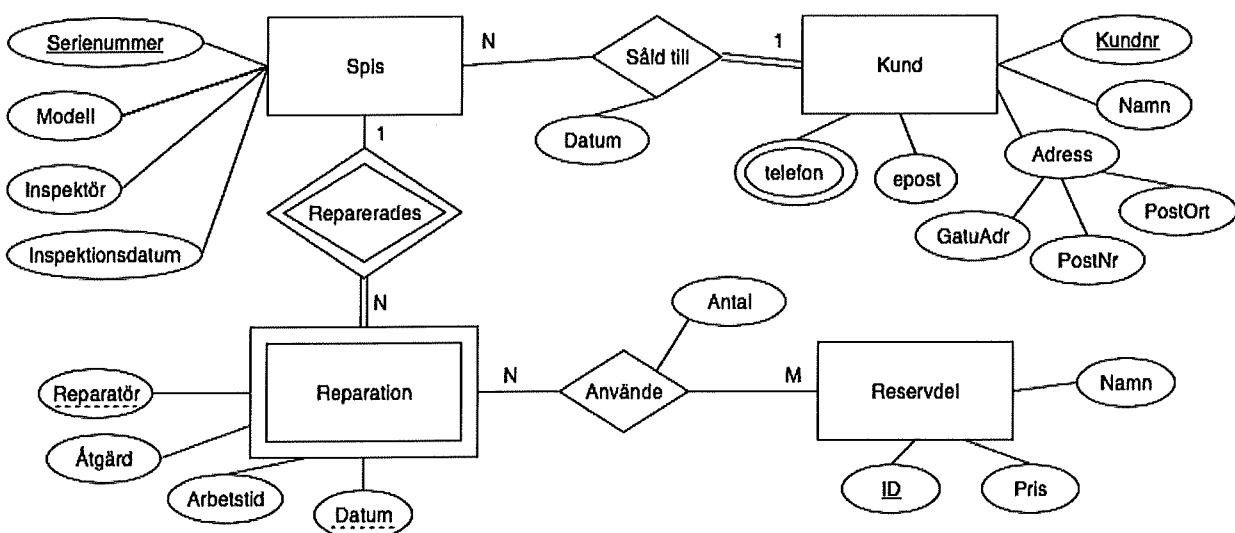
8. **Transaktionshantering:** Givet transaktionerna T1 och T2 som ser ut enligt nedan pseudokod (där X och Y refererar till två celler i en tabell och read och write är interaktioner med

databasen). Transaktion T1 kräver alltså att ökningen av X och minskningen av Y är atomär (odelbar): (8p)

Transaktion T1	Transaktion T2
Read (X)	Read(X)
X=X+1	X=X+5
Write(X)	Write(X)
Read(Y)	
Y=Y-1	
Write(Y)	

- Ge ett exempel på problem som kan uppstå när dessa körs parallellt. Visa på problemet genom att skriva operationerna i T1 resp T2 i den ordning de behöver hända för att problemet ska uppstå (skilj operationerna i de två transaktionerna åt genom att skriva transaktionsnamnet före operationen, t.ex. "T1.Read(X)"). Ange namnet på det problem du visar på.
- Visa, genom att lägga in operationerna Lock( ) och Unlock( ) på lämpliga platser i pseudokoden hur man kan få transaktionerna isolerade från varandra så att problemet inte kan uppstå samtidigt som man låser cellerna så kort tid som möjligt. Ange namnet på det protokoll du tänker att du följer.

9. **Relationsmodell:** Givet nedanstående ER-diagram, konvertera det till relationsschema. Ge relationer och attribut namn så att det framgår var de kommer ifrån och markera primärnycklar och främmande nycklar som vi gjort i kursen. (12p)



Lycka till!