



Försättsblad till skriftlig tentamen
vid Linköpings universitet



Datum för tentamen	2015-08-14
Sal (1)	<u>TER2</u>
Tid	14-18
Kurskod	732G16
Provkod	TEN1
Kursnamn/benämning Provnamn/benämning	Databaser: design och programmering Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	10
Jour/Kursansvarig Ange vem som besöker salen	Erik Prytz
Telefon under skrivtiden	070-2762748
Besöker salen ca klockan	14:45-15:00
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Annelie Almquist, 013-282934, annelie.almquist@liu.se
Tillåtna hjälpmedel	Inga hjälpmedel tillåtna.
övrigt	Instruktioner: Skriv läsligt. Visa ditt arbete när så krävs av uppgiften. Läs igenom tentan innan du börjar, så att du kan fråga jourhavande lärare om det finns oklarheter. Använd ett papper per fråga. Totalt antal poäng: 27 Poäng för G: 13 Poäng för VG: 20 Lycka till!
Antal exemplar i påsen	

1. (1p) **Begrepp:**
 - a. **Flervalsfråga, välj endast ett alternativ:** I databassammanhang skiljer man på *instans* och *schema*. Vad är en instans av en databas?
 - A: En beskrivning av de tabeller som kan finnas i databasen
 - B: En tabell bestående av en mängd tupler i databasen
 - C: Den ordnade listan av attributvärden i databasen
 - D: Den samling data som just nu finns i databasen

2. (2p) **Begrepp:** Ett för många förvirrande begrepp i relationsmodellen är just begreppet "relation". Vad är en "relation" enligt relationsmodellen? Hur förhåller sig detta till begreppet "relationships" (sv. "samband" eller "sambandstyper") i ER-modellen? (Svara med 1-2 meningar per fråga)

3. (1p) **Relationsalgebra:** Översätt följande relationalalgebraiska uttryck till en SQL-fråga:
$$\pi_{\text{Kapitel, CustomerRating}}(\sigma_{\text{AntalOrd} > 22000 \wedge \text{AntalBilder} < 5}(\text{MANUAL}))$$

4. (3p) **Fysisk design:** Antag att en databasfil ska lagras på en hårddisk som har en blockstorlek på 1024 byte och där adressen till varje block är 10 byte stor. Posterna i filen är 40 byte styck och primärnyckeln, som filen är sorterad efter, tar upp 6 av dessa 40 byte. Det finns just nu 200 000 poster i filen. För denna databasfil vill man nu skapa en indexfil med ett glest index.
 - a. Vad är blockningsfaktorn, och hur många block krävs för att lagra denna fil? Visa ditt arbete.
 - b. Hur stora blir indexfilens poster? Visa ditt arbete.
 - c. I detta (glesa) index krävs en post i indexfilen per block i databasfilen. Hur många block krävs således för att lagra denna indexfil? Visa ditt arbete.

5. (2p) **Transaktioner:** För att förhindra transaktionsproblem med parallella körningar av databasen kan man använda sig av olika *låsfunktioner*. Vad innebär binära lås, och hur skiljer det sig från läs- och skrivlås?

6. (4p) **ER-Design:** Ett nytt rapporteringssystem för folkhälsa planeras i Sverige, och du har i uppdrag att ta fram ett förslag på en lämplig databasstruktur i form av ett ER-diagram för databasen som ska spara information. Det finns en initial kravspecifikation nedan som du ska använda dig av för att designa ER-diagrammet.

Grunden i systemet är rapporter, som ska skickas in av sjukhus. Varje rapport innehåller ett anonymt ID-nummer för patienten som behandlats, vilket kön patienten har, åldern på patienten (endast år), samt det datum som patienten behandlats, och datumet som rapporten skickats in.

För att underlätta rapporteringen ska det finnas separat information om specifika sjukdomar också, och det ska vara sjukdomens "vanliga" namn, sjukdomens medicinska namn, en ID-kod, samt en beskrivande text om sjukdomen. På så sätt kan sjukhusen använda sig av ID-koden när de skickar in sina rapporter. Alla rapporter måste ange minst en sjukdom (med hjälp av ID-koder).

Information måste också sparas i databasen om sjukhusen som rapporterar sjukdomarna. Sjukhusen är centrala i databasen, och den information man initialt vill fånga är sjukhusets namn (t.ex. "Karolinska Institutet" eller "Linköpings Universitetssjukhus"), adressen till sjukhuset, samt Socialstyrelsens identifikationsnummer (som är unikt för varje sjukhus/vårdinrättning i Sverige). Man vill även kunna ta fram det totala antalet vårdplatser, för det är viktigt, men man vill göra detta genom att räkna på antalet vårdplatser per avdelning per sjukhus. Därför måste man även spara en hel del information om sjukhusens vårdavdelningar. Alla sjukhus har minst en, men oftast många fler, avdelningar. Man vill spara namnen på avdelningarna, och de är ofta samma för olika sjukhus (många sjukhus har avdelningar som heter till exempel "Ortopedi" eller "Öron, Näsa, Hals"). Däremot finns det inget sjukhus som har flera avdelningar som heter samma sak inom sjukhuset. För varje avdelning vill man, förutom namnet, spara antalet vårdplatser, antalet anställda, samt hur stor deras årliga budget är.

Slutligen, för att underlätta statistiska sammanställningar, vill man även spara information om regionerna sjukhusen finns i och drivs av (t.ex. Östergötland, Gotland, etc). Alla regioner i Sverige har inte sjukhus, men alla sjukhus finns förstås i en (och endast en!) region. Man vill ha namn och invånarantal för varje region i detta första skede.

Rita ett ER-diagram baserat på denna kravspecifikation. Markera nycklar, deltagande, samt kardinalitet där det behövs. Du behöver **inte** översätta ER-diagrammet till relationstabeller. Om information saknas eller kravspecifikationen är för vag i något avseende (ex. rörande unika värden för nycklar, deltagande, eller kardinalitet) när du bestämmer din design måste du specificera de antaganden du gör.

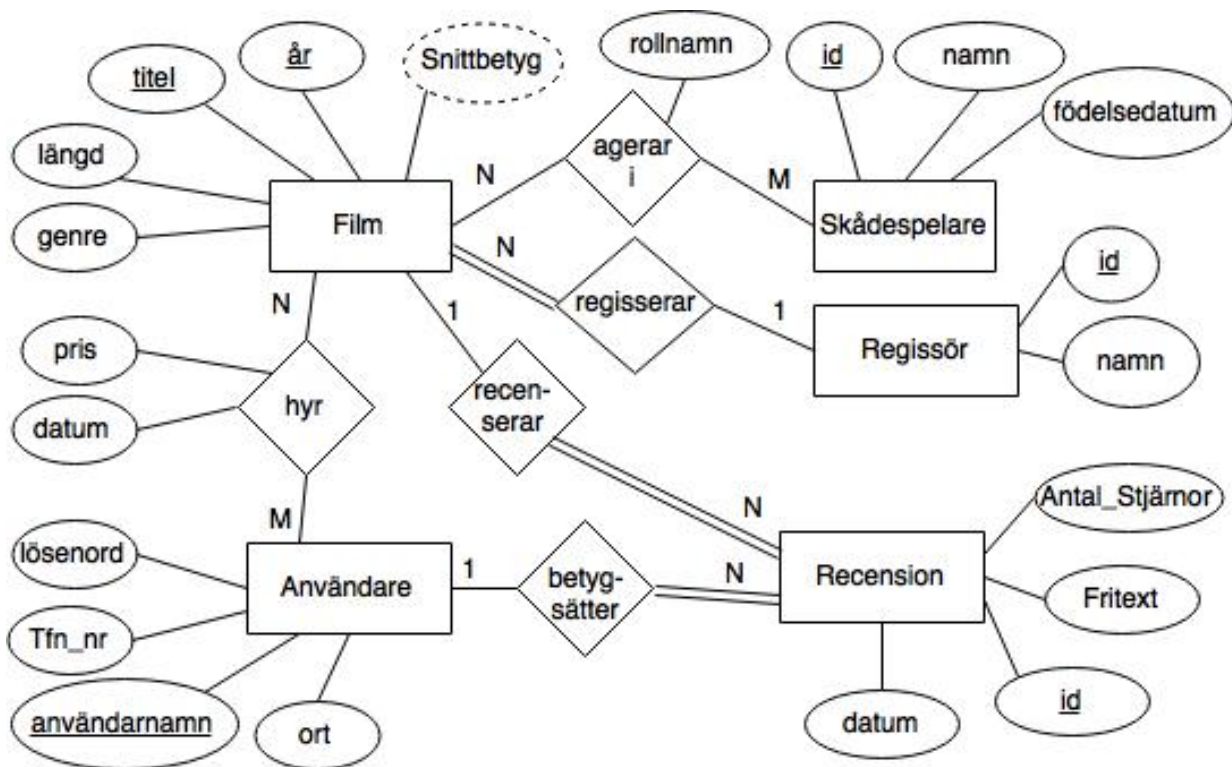
7. (4p) **ER till Relation:** Det internet-baserade företaget VHS4Ever, som hyr ut film över nätet, behöver en ny databas. De har redan designat ett ER-diagram över hur de vill att databasen ska vara strukturerad, men har inte lyckats översätta ER-diagrammet till en relationsdatabas. Hjälp VHS4Ever genom att översätta diagrammet till en lämplig relationsdatabas i enlighet med översättningsstegen beskrivna i kursboken och/eller föreläsningarna. För denna fråga behöver du **inte** formellt normalisera relationerna, men bör ändå sträva mot en enkel och icke-redundant lösning. Skriv ut relationerna enligt exempelformatet nedan och stryk under alla primärnycklar. Skriv även separat alla främmande nycklar med pilar till de primärnycklar dessa refererar till. Skriv ut de antaganden du gör, om några.

Exempelformat för relationerna:

Relationsnamn(Attribut 1, Attribut 2, Attribut 3)

Exempelformat för främmande nycklar:

RelationEtt(FrämmandeNyckel) → RelationTvå(Attribut).



8. (4p) **Normalisering:** En klantig databasdesigner har implementerat en dåligt designad relationsdatabas åt affärskedjan Freudian Slippers, som säljer tofflor med roliga motivtryck. Nu är det upp till dig att få lite fason och ordning på denna databas. Den dåliga databasen består av en enda stor tabell:

BUTIK(AnställningsID, TransaktionsID, VaruID, Förnamn, Färg, Efternamn, Timmar, Adress, Pris, Märke, Datum, Storlek, Butiksnamn, Lagerantal, Timlön, Antal, Tryck)

I den ligger alltså all information, från information om de anställda (som deras namn och lön), till butikerna med sina namn och adresser, lagerstatus, till alla transaktioner (försäljningar) och så vidare. Du vet att följande fullt funktionella beroenden (FFB) finns:

{AnställningsID} → {Förnamn, Efternamn, Timlön}
{AnställningsID, Butiksnamn} → {Timmar}
{Butiksnamn} → {Adress}
{Butiksnamn, VaruID} → {Pris, Lagerantal}
{VaruID, TransaktionsID} → {Antal}
{TransaktionsID} → {Datum, Butiksnamn}
{VaruID} → {Märke, Färg, Tryck, Storlek}

Normalisera tabellen BUTIK så att den uppfyller Boyce-Codd Normalform (BCNF). Skriv ut relationer inklusive attribut. Lista alla främmande nycklar (se fråga 7 för exempelformat). Motivera varför din lösning uppfyller varje normalform. Ange eventuella antaganden du gör.

9. (2p) **Integritet:** Man brukar tala om två typer av villkor som man måste ta hänsyn till för databasens integritet. Vad kallas dessa två villkorstyper, och vad innebär de?
10. (4p) **SQL:** En databas över ett bibliotek ser ut som följer:

BOK(ISBN, Titel, Utgivningsår, AntalExemplar, Förlag)
 FÖRLAG(Namn, Ort, Land)
 FÖRFATTARE(FörfattarID, Förnamn, Efternamn, Födelseår, Dödsår)
 LÅNTAGARE(Lånekortsnummer, Förnamn, Efternamn, Födelseår, Telefon)
 SKRIVENAV(Bok, Författare)
 LÅN(Nr, Låntagare, Bok, Antal, Datum, Returdatum)

Primärnycklarna är understrukna, och det finns även följande främmande nycklar:

SKRIVENAV(Författare) → FÖRFATTARE(FörfattarID)
 SKRIVENAV(Bok) → BOK(ISBN)
 LÅN(Låntagare) → Låntagare(Lånekortsnummer)
 LÅN(Bok) → BOK(ISBN)
 BOK(Förlag) → FÖRLAG(Namn)

Skriv ned de SQL-frågor som behövs för att besvara de tre nedanstående frågorna. Om du tycker det saknas information så att du behöver göra antaganden ange då samtliga sådana antaganden du gör.

- Vad är titlarna på de böcker som biblioteket har i sin databas som är utgivna av förlaget "Studentlitteratur" och som de har mer än 5 exemplar av?
- Hur många böcker, totalt, är utlånade till låntagare som har ett efternamn som *slutar* på "son" (t.ex. Svensson, Johansson, Simpson, Son, etc, men *inte* Sonby eller liknande.).
- Vem är den äldsta författaren som fortfarande är i livet, och hur många böcker har hen skrivit? (Tips: Levande författare har ett nullvärde på Dödsår, avlidna författare har ett numeriskt värde ex. 1923 eller 2015. Observera också att det är år 2015 just nu.)