

# Tentamen 2004-05-28

## DATABASTEKNIK - 1DL116, 1MB025

Datum .....Fredagen den 28 Maj, 2004  
Tid .....13:00-18:00  
Jourhavande lärare ...Kjell Orsborn, tel. 471 11 54 eller 070 425 06 91  
Hjälpmedel .....miniräknare

### Anvisningar:

- Läs igenom hela skrivningen och notera eventuella oklarheter innan du börjar lösa uppgifterna. Förutom anvisningarna på skrivningsomslaget så gäller följande:
  - Skriv tydligt och klart. Lösningar som inte går att läsa kan naturligtvis inte ge några poäng och oklara formuleringar kan dessutom misstolkas.
  - Antaganden utöver de som står i uppgiften måste anges. Gjorda antaganden får förstås inte förändra den givna uppgiften.
  - Skriv endast på en sida av papperet och använd ett nytt papper för varje uppgift för att underlätta rättning och minska risken för missförstånd.
- För godkänt krävs det cirka 50% av maxpoäng.

### 1. Database terminology:

2p

Concisely explain the following concepts (in a database context):

- (a) primärindex (eng. primary index)

Svar: Primärindex består av en ordnad fil av dataposter med 2 fält. Första fältet är av samma typ som ordningsfältet (indexeringsfältet) för datafilen och det andra fältet är en pekare till ett datablock (blockpekare). Primärindex är ett gles index då det har en indexpost för varje block i data-filen. Primärindex kräver mycket mindre plats än motsvarande datafil och kan utnyttjas för att snabba upp sökningen av dataposter i datafilen med avseende på indexeringsfältet.

- (b) naturlig join (eng. natural join)

Svar: En naturlig join är en sammansättningsoperation mellan två relationer (tabeller) där villkoret för att kombinationen av två tupler (en från varje relation) skall ingå i den resulterande och sammansatta relationen är ett likhetsvillkor mellan ett/ flera attribut. De attribut från den högra relationen som deltar i likhetsvillkoret ingår ej i den resulterande tabellen, dvs redundanta attribut elimineras.

### 2. ER/EER modellering:

4p

- (a) Hur representeras ett flervärt attribut från ER-modellen i relationsmodellen? (1p)

Svar: Ett flervärt attribut representeras som en egen relation (tabell) med en nyckel som består av attributet själv tillsammans med primärnyckeln från relationen (representerande en entitetstyp eller relationstyp) där det flervärda attributet ingår.

- (b) Förklara begreppen specialisering och generalisering (eng. specialization och generalization) samt förklara skillnaden mellan dessa. (3p)

Svar:

Specialisering är en process för att specificera en mängd av subklasser utifrån en entitetstyp som kallas superklass. Subtyperna erhålls härvid genom att någon/några egenskap/er hos superklassen som skiljer sig åt hos subklasserna.

Generalisering är den motsatta processen att specificera en superklass utifrån ett antal subklasser där man finner ett antal gemensamma egenskaper som kan extraheras och defineras att utgöra egenskaperna i en gemensam superklass.

### 3. Relationsalgebra:

2p

Antag att vi har två relationer (tabeller) med följande relationsscheman:

PUNKT(PNAMN, X, Y)  
CIRKEL(CNAMN, PKTNAMN, RADIE)

Uttryck följande frågor i relationsalgebra:

- (a) vilka namn och radier har cirklar med radier större än 35.0 (cm)?

Svar:

$\pi_{\langle CNAMN \rangle}(\sigma_{RADIE \geq 35.0}(CIRKEL))$

- (b) `SELECT CNAMN, PNAMN, X  
FROM CIRKEL, PUNKT  
WHERE PKTNAMN = PNAMN AND X > 0.0;`

Svar:

$\pi_{\langle CNAMN, PNAMN, X \rangle}(\sigma_{X > 0.0}(CIRKEL \bowtie_{\langle PNAMN = PKTNAMN \rangle} PUNKT))$

#### 4. Normalisering:

4p

Förklara följande begrepp för relationsmodellen:

- (a) första normalform - 1NF (1p)

Svar: Första normalform (1NF) är den grundläggande definitionen för relationsmodellen som säger att attributens domäner endast får bestå av atomiska (enkla odelbara) värden och att värdet för varje attribut i en tupel endast får bestå av enkla (endast ett) värde från attributets domän.

- (b) funktionellt beroende och fullt funktionellt beroende (3p)

Svar:

Ett funktionellt beroende så att X bestämmer Y,  $X \rightarrow Y$ , existerar då om för varje par av tupler  $t_1, t_2 \in r(R)$  och för alla  $r(R)$  följande gäller: om  $t_1[X] = t_2[X]$  så gäller att  $t_1[Y] = t_2[Y]$

Fullt funktionellt beroende anger att för ett funktionellt beroende gäller att det inte finns någon delmängd attribut  $A \subseteq X$  så att  $(X - \{A\}) \rightarrow Y$ .

Här gäller att  $R$  är ett relationsschema och  $r(R)$  är en instans av schemat  $R$  med attributen  $A_1, \dots, A_n$  och  $X, Y \subseteq \{A_1, \dots, A_n\}$ .

Alltså ett fullt funktionellt beroende är ett funktionellt beroende som inte innehåller något onödigt attribut i determinanten (vänsterledet i beroendet).

#### 5. Samtidighetskontroll (eng. concurrency control):

4p

Beskriv principerna för låshanteringen i samband med transaktioner för ett två-fas låsningsprotokoll (eng. two-phase locking protocol).

Svar: En transaktion sägs följa ett två-fas låsningsprotokoll om alla låsningsoperationer föregår den första upplåsningsoperationen (unlock) i transaktionen. Alltså en sådan transaktion genomgår en expanderande fas där nya lås kan utfärdas men inga lås kan släppas; och en krympande fas där existerande lås kan läsas upp men inga nya lås kan erhållas.

Två-fas låsningsprotokoll garanterar serialiserbara transaktionsscheman.

#### 6. Aktiva Databaser:

5p

Vi har två tabeller:

`EMPLOYEE(PNR, DNO, SALARY), nyckel PNR  
DEPT(DNO, NAME), nyckel DNO`

PNR: Personnummer  
DNO: Avdelningsnummer  
SALARY: Lön

- (a) Man vill göra en s.k. mastertabell som alltid är aktuell och som innehåller summan SUM av resp. avdelnings löner:

SALSUM(DNO,SUM), nyckel DNO

Visa hur triggers specificeras för att upprätthålla SUM. (2 p)

Svar:

```
create trigger propagatesum
  after update on employee
  referencing new row as nrow, old row as orow
  for each row
  begin
    update salsum s
      set sum = sum - orow.salary
      where s.dno = orow.dno
    update salsum s
      set sum = sum + nrow.salary
      where s.dno = nrow.dno
  end
```

- (b) Antag att vi också har en annan tabell med maximal tillåten summa av löner TOTSAL för varje avdelning:

BUDGET(DNO,TOTSAL), nyckel DNO

Visa hur triggers specificeras för att garantera att ingen avdelning någonsin har högre totala löner än TOTSAL. (2 p)

Svar:

```
create trigger checksalbudget
  after update on salsum
  referencing new row as nrow, old row as orow
  for each row
  when nrow.salary > (select totalsal from budget b where b.dno = nrow.dno)
  begin
    rollback
  end
```

- (c) Vad står ECA för i samband med aktiva databaser? (1 p)

Svar:

E = Event  
C = Condition  
A = Action

## 7. Frågeoptimering:

5p

- (a) Hur används *distinkta värden* för frågeoptimering? (1 p)

Svar:

Distinkta värden är hur många olika värden som finns i en given kolumn av en tabell. Distinkta värden används för att uppskatta hur många rader matchar jämförelser med kolumnen. Detta behövs för att uppskatta tiden att utföra frågan Om t.ex. det finns C rader och D distinkta värden för kolumn K i tabell T så får man D/C svar från frågan

```
SELECT * FROM T WHERE T.K=...
```

- (b) Vad är 'kostnadsbaserad optimering' (1 p)

Svar:

7b: Kostnadsbaserad optimering går ut på att hitta den billigaste exekveringsplanen för en given fråga, baserat på modell av konstnaden att utföra de olika operatorerna i varje exekveringsplan. För att uppskatta kostnaden håller databashanteraren statistik om fördelningen av datavärlden mm. för varje kolumn.

(c) Vilka är de tre vanligaste join-metoderna? (3 p)

Svar:

Nested-loop join

Sort-merge join

Hash join

## 8. XML-databaser:

4p

(a) Vad skiljer DTD från XMLSchema? (2 p)

Svar:

DTD beskriver struktur (syntax) hos XML dokument. Enda datatyp är sträng. XMLSchema beskriver datatyper och poststrukturer hos XML dokument. Många olika datatyper tillgängliga, ett rikt typsystem. Begränsningar i datavärden kan också beskrivas.

(b) Vilka är de två viktigaste frågespråken för XML? (2 p)

Svar:

XPath och XQuery

## 9. Multimedia-databaser:

2p

Vad är 'feature extraction' och hur används det för multi-mediadatabaser?

Svar:

Feature extraction går ut på att extrahera egenskaper från t.ex. bilder. När man sedan lagrar bilder i databasen lagras man också dessa egenskaper i vanliga tabeller. Man kan sedan söka efter bilder med visst innehåll specificerat genom dessa egenskaper genom att först extrahera egenskaperna för eftersökt bild och sedan söka i databasen efter bilder med matchande egenskaper.

Lycka till!

/ Kjell och Tore