

Tentamen

TI2506 Informatie- en Datamodellering

Donderdag, 11 december 2014
13u45 - 15u45

Dit tentamen bestaat uit 6 delen, elk met een aantal open vragen.
Van alle vragen dienen de antwoorden op het gewone antwoordblad(en) ingevuld te
worden. Kladpapier wordt niet nagekeken

Deel	Punten
1	10
2	10
3	20
4	10
5	15
6	35
Totaal	100

Het gebruik van de boeken “*Database Systems: Global Edition, 6/E*” en “*Fundamentals of Database Systems*” door Ramez Elmasri en Shamkant Navathe is toegestaan, evenals slides, oude tentamens met uitwerkingen en eigen aantekeningen.

Vul je naam, studienummer en studierichting in op ieder antwoordblad.

Veel succes

Deel 1 – Relationale Algebra (10 punten)

Vraag 1. Beschouw het volgende relationele schema zoals ook gebruikt in het boek:

EMPLOYEE(*Fname*, *Minit*, *Lname*, *Ssn*, *Bdate*, *Address*, *Sex*, *Salary*, *Super_ssn*, *Dno*)
DEPARTMENT(*Dname*, *Dnumber*, *Mgr_ssn*, *Mgr_start_date*)
DEPT_LOCATIONS(*Dnumber*, *Dlocation*)
PROJECT(*Pname*, *Pnumber*, *Plocation*, *Dnum*)
WORKS_ON(*Essn*, *Pno*, *Hours*)
DEPENDENT(*Essn*, *Dependent_name*, *Sex*, *Bdate*, *Relationship*)

Geef voor de volgende queries de algebra-expressie die deze query uitdrukt met de operatoren zoals besproken in Hfdst. 6 van het boek. Dit mag net zoals in het boek in de vorm van een enkele expressie of voor ingewikkelde queries in de vorm

<tabel> $\leftarrow$ *<algebra expressie>*;

<tabel> $\leftarrow$ *<algebra expressie>*;

... ;

RESULT $\leftarrow$ *<algebra expressie>*.

- a) Geef de namen van de departementen waarvan alle projecten dezelfde locatie hebben.
- b) Geef de namen van de projecten waarin de meerderheid van de werknemers niet getrouwd is, dwz. geen 'dependent' met *relationship* 'husband' of 'wife' heeft.

Deel 2 – Normaalvormen (10 punten)

Vraag 2. Beschouw een relatie met schema *R*(*A*, *B*, *C*, *D*, *E*) met sleutel (candidate key) *ABC*, en functionele dependencies $\{AB \rightarrow D, BC \rightarrow E, D \rightarrow E\}$. Wat is de hoogste normaalvorm waarin dit schema zit? Beargumenteer je antwoord.

Vraag 3. Stel we hebben een relatie *R*(*A*, *B*, *C*, *D*, *E*, *F*) met *join dependency* *JD*(*AB*, *BC*, *ACDE*, *ACF*) en sleutels (*candidate keys*) *B* en *AC*. Is deze relatie in 5NF? Laat zien waarom wel of niet.

Deel 3 – Normalizatie (20 punten)

Vraag 4. Beschouw een relatie met schema *R*(*A*, *B*, *C*, *D*, *E*, *F*) en de set van functionele dependencies $S = \{A \rightarrow DE, E \rightarrow B, BD \rightarrow F\}$. Bewijs met de regels van Armstrong (dus alleen IR1, IR2 en IR3) dat de functionele dependency $AC \rightarrow F$ in S^+ zit.

Vraag 5. Beschouw een relatie met schema *R*(*A*, *B*, *C*, *D*, *E*) met de set van functionele dependencies $S = \{B \rightarrow CD, D \rightarrow A\}$. Bepaal of de decompositie $\{AB, BCDE\}$ de *lossless join property* heeft. Laat zien hoe je dit bepaald hebt.

Vraag 6. Beschouw een relatie met schema $R(A, B, C, D, E, F, G)$ en de set van functionele dependencies $S = \{ A \rightarrow BCDG, BC \rightarrow D, AB \rightarrow E, AD \rightarrow EG \}$. Bepaal een decompositie naar een 3NF schema met de *nonadditive join property* volgens algoritme 6 (hfdst. 16) in het boek en slides. Geef per component ook de sleutels.

Deel 4 – File organisatie (10 punten)

Vraag 7. Een PARTS bestand met Part# als de hash sleutel bevat records met de volgende Part# waarden (binair gepresenteerd): 010101, 011111, 101111, 001100 en 110000. Laad deze records in een lege *extendible hash* file gebaseerd op *extendible hashing*. Neem aan dat buckets maximal twee records kunnen bevatten. Laat zowel de globale als lokale diepte zien. Begin met globale diepte $d=0$. Gebruik de hash-functie $h(K) = K \bmod 16$.

Deel 5 – Indexering (15 punten)

Vraag 8. Een PARTS bestand met Part# als sleutelveld bevat records met de volgende Part# waarden: 13, 12, 15, 9, 1, 8, 6, 3, 11, 7, 10, 14. Stel dat deze zoekwaarden in deze volgorde ingevoegd worden in een lege B^+ -boom van orde $p=4$ en $p_{\text{leaf}}=3$. Laat zien hoe de uiteindelijke B^+ -boom er uit ziet.

Deel 6 – Queryoptimalisatie (35 punten)

Vraag 9. Aan welke eis moeten twee relaties R en S voldoen wil $R \cup S$ een geldige relationele algebra expressie zijn? Geef een voorbeeld van relaties R en S waar dit **niet** geldt en geef aan wat er dan fout gaat.

Vraag 10. Een bestand van 15.000 blocks moet gesorteerd worden. Met hoeveel blocks aan bufferruimte zou dit gesorteerd kunnen worden door een 'external sort-merge' algoritme met ten hoogste 4 'passes' in de merge fase?

Vraag 11. Beschouw de volgende SQL-query voor het schema van vraag 1:

```
SELECT Fname, Lname, Dlocation
FROM EMPLOYEE, WORKS_ON, PROJECT
WHERE Sex='F' AND Ssn=Essn AND Pno=Pnumber AND Hours>50;
```

Geef twee verschillende *query trees* voor deze query.

Vraag 12. Stel we hebben relaties $R(a, b)$, $S(c, d)$ en $T(e, f)$. Geef precies alle algebraïsche identiteiten voor de relationele algebra die nodig zijn om de expressie

$$\pi_{a,f}(\sigma_{a=5 \text{ AND } c=3 \text{ AND } b=d \text{ AND } a=e}((R \times S) \times T))$$

om te schrijven in

$$\pi_{a,f}(\pi_a(\sigma_{a=5}(R) \bowtie_{b=d} \pi_d(\sigma_{c=3}(S))) \bowtie_{a=e} T).$$

Je hoeft alleen de nummers van de regels in de slides te geven, niet de stappen waarin ze toegepast worden.

Einde van het tentamen