

Eksamensoppgave i**TDT4225 Lagring og behandling av store datamengder****Fredag 2. desember 2011, kl. 0900-1300***Oppgaven er utarbeidet av faglærer Kjell Bratbergsengen og kvalitetssikret av Svein-Olaf Hvasshovd**Kontaktperson under eksamen: Kjell Bratbergsengen, telefon 7359 3439 og 906 17 185**Språkform: Bokmål**Tillatte hjelpemidler: D**Ingen trykte eller håndskrevne hjelpemiddel er tillatt.**Bestemt, enkel kalkulator tillatt.**Sensur: Fredag 23. desember 2011***Oppgave 1, Videoserver (10 %)**

En videoserver skal levere film til enkeltkunder hvor hver videostrøm i gjennomsnitt krever 3 MB/sekund. Data hentes fra disk som gjør 10000 rotasjoner per minutt og har en gjennomsnittlig sporlengde på 800 KB. Disken er koplet til en fiberoptisk buss med maksimal kapasitet 200 MB/s. Serveren har en intern buss med maksimal kapasitet 400 MB/s.

Ingen enheter skal belastes mer enn 70 % av full kapasitet.

Serverens arbeidslager er 8 GB.

Hvor mange kunder kan serveren maksimalt betjene samtidig?

Vis hvordan du kommer fram til resultatet og presiser forutsetningene.

Oppgave 2, R-tre for lagring av to-dimensjonale objekter (20 %)

a) Forklar oppbyggingen av en R-tre-fil.

b) Tabellen viser en indeksblokk i et R-tre.

| Koordinater | | | | |
|-------------|----|----|----|-------|
| x0 | y0 | x1 | y1 | Blokk |
| 7 | 1 | 11 | 3 | A |
| 10 | 7 | 15 | 20 | B |
| 5 | 4 | 6 | 17 | C |
| 15 | 5 | 20 | 6 | D |
| 4 | 4 | 16 | 9 | E |

Du skal søke etter et objekt med koordinat $\langle 8,13 \rangle$. Hvilke subtrær må søkes gjennom for å finne objektet.

c) Indeksblokken må splittes. Bruk kvadsplitt. Hvilke omhyllingsrektangler (bounding boxes) blir kimer i de to nye blokkene?

d) Ingen blokk skal ha færre enn to indekser. Fullfør splittingen av indeksblokken.

Oppgave 3, Sortering (15 %)

a) Du skal gjøre initiell sortering av følgende poster med reservoarmetoden, reservoaret har plass for 4 poster.

Innfilens nøkler ser slik ut: 32,44,10,3, 5,79,64,43,98,33,8,3,5,2.

Vis resultatet ved å skrive ut alle delfilene.

b) Du har fått 370 delfiler og kan flette maksimalt 90 delfiler. Hvor mange "dummy" delfiler må du legge til i første fletting for å få et optimalt flettetre?

c) Alle delfilene har volum 1. Hva blir samlet I/O-volum for hele fletteprosessen?

Oppgave 4, Lagring av matriser. (20 %)

Et system for lagring av store matriser lagrer matrisen som kvadratiske submatriser. Hver submatrise lagres som en blokk på disk, 16 KB stor. Hvert element i matrisen krever 4 byte.

a) Hvor mange rader og kolonner blir det i hver submatrise?

b) Systemet brukes til å lagre to matriser, $R(2000,3000)$ og $S(3000,4000)$. Hvor mye plass på disken opptar de to matrisene? Tallene i parentes angir antall linjer henholdsvis kolonner i matrisen.

c) Matrisene skal multipliseres: $T=R*S$. Skisser algoritmen du vil bruke for multiplikasjonen. Hvor stor blir T?

d) Anta at primærlageret bare har plass til tre submatriser. Hvor stort blir I/O-volumet for matrisemultiplikasjonen?

e) Hvis du har plass for 100 submatriser i arbeidslageret, hva blir nå I/O-volumet? Begrunn svaret.

Oppgave 5, Relasjonsalgebra (25 %)

a) Forklar relasjonsalgebraoperasjonen snitt: $R = A \cap B$. Hvilke krav stilles til operandene A og B?

b) Forklar hvordan snitt blir utført med gjentatte gjennomløp. Finn også et generelt uttrykk for operasjonens totale transportvolum når en bruker gjentatte gjennomløp.

c) Tabellen $A(a,b,c,d)$ har 3 millioner poster á 500 byte. a er nøkkel med en gjennomsnittlig lengde på 15 byte.

Tabellen $B(r,s,t,u)$ har 1 million poster á 500 byte. r er nøkkel med gjennomsnittlig nøkkellengde på 15 byte. Hva blir operasjonens I/O-volum? Tilgjengelig reservoar er 10 MB.

d) For begge tabeller eksisterer det en lagret signaturtabell på primærnøkkelen. Signaturtabellene er enklest mulig og er uten postreferanse. En signatur tar 4 byte. Forklar hvordan du vil benytte signaturtabellene til å effektivisere snittoperasjonen.

e) Hva blir I/O-volumet for den modifiserte snitt-algoritmen?

Oppgave 6, Optimalisering (10 %)

a) Hvor mange forskjellig algebra –trær er det mulig å få til når en forener tre tabeller?

b) Gitt:

```
SELECT A.x,B.s FROM A,B,C WHERE
```

```
    A.x>21 AND A.x=C.y AND C.y<B.s AND B.s <100;
```

Tegn opp algebra-treet for utførelsen av uttrykket slik det står.

c) Kan uttrykket omskrives slik at utførelsen blir mer effektiv?

Tegn opp det modifiserte algebra-treet.

Kjb/17.11.2011