

**Eksamensoppgave i****TDT4225 Lagring og behandling av store datamengder**

**Onsdag 9. desember 2009, kl. 0900-1200, OBS bare tre timer**

*Oppgaven er utarbeidet av faglærer Kjell Bratbergsengen og kvalitetssikrer 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: 11. januar 2010*

**Oppgave 1, Diskkontroller (10%)**

Beskriv de oppgavene som ligger i en diskkontroller.

**Oppgave 2, Lagring av koordinater (30%)**

a) Forklar oppbyggingen av et k-d-tre.

b) Forklar oppbyggingen av en gridfil.

Vi har følgende sett av koordinater som skal settes inn (x,y):

(36,91) (3,87) (45,23) (45,21) (98,1) (72,12) (33,76) (45,25) (26,25) (22,50)

c) Sett verdiene over inn i et k-d-tre

d) Sett verdiene over inn i gridfilen hvor blokkstørrelsen er 4. Vis hvordan treet bygges opp og blokker splittes.

e) Til hvilke oppgaver er gridfil egnet?

### Oppgave 3, Relasjonsalgebra (30%)

- a) Forklar relasjonsalgebraoperasjonen differanse eller minus:  $R = A - B$ .
- b) Forklar hvordan differanse blir utført med gjentatte gjennomløp. Finn også et uttrykk for totalt transportvolum.
- c) Forklar hva vi mener med en signatur og hvordan signaturer kan brukes i utførelsen av relasjonsalgebra.
- d) Forklar algoritmen i detalj når du utnytter signaturer under utførelse av differanse.

### Oppgave 4, System (30%)

Vi er i et land hvor det er en rekke bomveger hvor betalingen skjer ved et køfrisisystem. Ved passering av en bomstasjon lagres blant annet brikkenummer, tidspunkt, bomstasjonens identitet, mm. Posten for hver passering tar 100 byte. Fordi kundene kan be om spesifiserte utskrifter, skal data for alle passeringer lagres i minst ett år. Systemet har 80 millioner kunder og 100 millioner biler (brikker). For hver kunde har vi kundens identifikator (64 bit heltall), betalingsmetode, saldo, og betalingshistorikk, til sammen 500 byte per kunde. Disse data er samlet i kunderegistret.

For hver brikke har vi brikkens identifikatorer (64 bit heltall), kundenummer til eier av brikken og en status, totalt ca. 60 byte per brikke. Status bestemmes av betalingsform og om brikken er meldt stjålet. For eksempel vil kunder som betaler på forskudd, få gult lys når saldo er under en viss grense, og rødt lys når alt er brukt opp. Disse dataene er samlet i brikkeregistret.

Det genereres i gjennomsnitt 100 millioner passeringer hver dag. Prisen per passering avhenger av hvilken bomstasjon som passerer, tid på dagen og kjøretøytype. Kjøretøytypen står oppgitt i brikkeregistret.

Alle data beskrevet over, er lagret i en felles sentralmaskin for hele systemet.

- a) Hvor stort er systemets lagerbehov (ett års lagring).
- b) Når et kjøretøy passerer har vi bare 0,1 sekund fra brikken blir registrert til passeringslyset (grønt, gult, rødt) skal tennes (72 km/t tilsvarende 20 m/s). Dette er for lite tid til at sentralmaskinen kan involveres og det er en lokal maskin som gjør jobben. Den lokale maskinen kan også samle opp passeringer og sende disse periodisk til sentralmaskinen. Hvilke datastrukturer vil du ha i den lokale maskinen?
- c) Sentralmaskinen holder totaloversikten og sender for eksempel ut statusendring til aktuelle brikkenummer. Kravene til forsinkelser er at kundens konto skal oppdateres minst en gang per døgn. Alle passeringer som er eldre enn ett døgn skal altså være reflektert i kundens saldo. Hvordan vil du organisere registrene i sentralmaskinen: kunderegister, brikkeregister og passeringsfil?
- d) Hvordan vil du utføre oppdateringene – de som kommer fra de lokale maskinene nevnt i b) – til de sentrale registrene?