

**Norges teknisk-naturvitenskapelige universitet
Institutt for datateknikk og informasjonsvitenskap**



EKSAMENSOPPGAVE I FAG TDT4150 – AVANSERTE DATABASESYSTEMER

Faglig kontakt under eksamen: Kjetil Nørvåg

Tlf.: 93440

Eksamensdato: 2. juni 2008

Eksamenstid: 09.00-13.00

Tillatte hjelpemiddel: D: Ingen trykte eller håndskrivne hjelpemiddel tillatt. Bestemt, enkel kalkulator tillatt.

Språkform: Bokmål

Sensurdato: 23. juni 2008

Oppgave 1 – Utføring av spørringer– 15 %

- Hva bruker man histogram til i forbindelse med spørre-optimalisering?
- Hva er forskjellen mellom *equiwidth* og *equidepth*-histogram?
- Hva er forutseningen(e) for at relasjonsalgebra-uttrykkene Q1 og Q2 skal være ekvivalente?
 Q1: $\pi_a (R \bowtie_c S)$
 Q2: $(\pi_{a1} (R)) \bowtie_c (\pi_{a2} (S))$

Oppgave 2 – Parallele og distribuerte databaser – 25 %

- Beskriv 3 forskjellige partisjonerings-teknikker og nevnt stikkordmessig hva som er gode og dårlige egenskaper for hver av dem.
- Beskriv *synkron* og *asynkron* oppdatering i forbindelse med replisering, og diskuter hva som er gode og dårlige egenskaper for hver av dem.
- Tegn en skisse med tilstander og meldinger ved utførelse av tofase-kommit (2PC).

Oppgave 3 – Datavarehus og datagravedrift– 30 % (5% på a, 10% på b, 15% på e)

- Nevn 3 vanlige datavarehus-operasjoner/operatorer som kan brukes på multidimensjonale data og forklar hva disse gjør.
- Hva ønsker man å oppnå ved klynging? Vis algoritmen til en vanlig klyngings-metode ved hjelp av pseudo-kode.
- Anta handlekorg-data til høyre. Bruk apriori-algoritmen for å finne hvilke assosiasjonsregler som gjelder, gitt at minimum støtte er 50% og konfidens er 80%.

TransaksjonsID	Kjøpte produkt
1000	A,B,C
2000	A,C
4000	A,D
5000	B,E,F

Oppgave 4 – Diverse– 30 % (15% på a og 5% på resten)

- Anta en relasjon med todimensjonale data. Forklar hva som er vilkåret for at et punkt i denne relasjonen skal være med i resultatet av *skyline*-operatoren anvendt på denne relasjonen. Bruk deretter *skyline* på relasjonen til venstre, og vis hvilke tupler som vil være med i resultatet.

X	Y
1	9
2	11
3	2
4	4
5	9
6	6
7	11
7	8
8	5
9	1

- Hva er prinsippene bak kolonnebaserte databasesystem, og for hvilke aksessmønstre er slike databasesystem mest egnet?
- Hva er forskjellen mellom *valid-time* og *transaction-time* i forbindelse med temporale databaser?
- Tegn en skisse som viser datastrukturene på en POSTGRES-side, og beskriv kort hva som er innholdet/hensikten med hver av dem.