

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



EKSAMENSOPPGAVE I FAG TDT4150 – AVANSERTE DATABASESYSTEMER

Faglig kontakt under eksamen: Kjetil Nørvåg

Tlf.: 93440

Eksamensdato: 8. juni 2012

Eksamenstid: 09.00-13.00

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

Språkform: Bokmål

Sensurdato: 29. juni 2012

Oppgave 1 – Datamodeller og arkitektur – 25 %

- Gi en oversikt over hovedkomponentene i et databasesystem og forklar kort funksjonen til hver av dem under utføring av en spørring.
- Forklar alternative prosessmodeller for et databasesystem og deres fordeler og ulemper.

Oppgave 2 – Spørreoptimalisering – 10 %

- Anta at en relasjon R har en svært ikke-uniform distribusjon av verdier for en gitt attributt. I utgangspunktet lagret man kun statistikk med tre verdier (minste verdi, største verdi, antall tupler). Så ble databasen endret til å bruke histogram i stedet. Hvorfor vil spørringer på R nå kunne bli utført mer effektivt? Gi et eksempel på en slik spørring.
- Hvordan kan valget av joinalgoritme bli påvirket av om resultatet av spørringen skal være sortert eller ikke?

Oppgave 3 – Parallele og distribuerte databaser – 10 %

- Diskuter i hvilken grad join er velegnet for samlebåndsparallelitet og partisjonert parallelitet.
- Hva er fordeler og ulemper med asynkron ("lazy") gruppe-oppdatering ("update anywhere") av replikater?

Oppgave 4 – Skyline – 30 %

- Anta en relasjon med todimensjonale data. Forklar hva som er vilkåret for at et tuppel i denne relasjonen skal være med i resultatet av *skyline*-operatoren anvendt på denne relasjonen. Bruk deretter *skyline* på relasjonen til høyre, og vis hvilke tupler som vil være med i resultatet.
- Forklar splitt-og-hersk-algoritmen ("divide and conquer") for prosessering av skyline-spørringer.

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

Oppgave 5 – Diverse – 25 %

- Forklar prinsippene bak søk i relasjonsdatabaser ("keyword search in relational databases") og hva man får som resultat av et slikt søk.
- Forklar top-*k* spørringer. Gi et eksempel på en top-2-spørring og resultat basert på relasjonen til høyre. Gjør de antagelser som er nødvendig.