

Institutt for datateknikk og informasjonsvitenskap

Eksamensoppgave i TDT4150 Avanserte Databasesystemer

Faglig kontakt under eksamen: Kjetil Nørvåg

Tlf.: 41440433

Eksamensdato: 13. desember 2014

Eksamenstid (fra-til): 09.00-13.00

**Hjelpemiddelkode/Tillatte hjelpemidler: D: Ingen trykte eller håndskrivne hjelpemiddel
tillatt. Bestemt, enkel kalkulator tillatt.**

Annen informasjon:

Målform/språk: Bokmål

Antall sider (uten forside): 1

Antall sider vedlegg: 0

Kontrollert av:

Dato

Sign

Oppgave 1 – Spørreoptimalisering – 25 % (5 % på a/c/d, 10 % på b)

- a) Anta at en spørring inneholder bare en seleksjon og en projeksjon. Når kan det være lurt å utføre seleksjon før projeksjon og omvendt?
- b) Gitt spørringen: `SELECT * FROM r1, r2 WHERE r1.c1 = r2.c2 AND r2.c3 > 5;`
Hvilken statistikk bør databasesystemet ha for å kunne estimere størrelsen på resultatet av denne spørringen? Lag et symbolsk uttrykk for estimatet. Beskriv de antagelsene du gjør.
- c) Hva er fordeler og ulemper med *equiwidth* histogram i forhold til *equidepth*?
- d) Anta at tre relasjoner skal joins ved hjelp av *nested loop*. Hvordan bør utføringsplanen være for at man skal oppnå best mulig samlebåndsparallelitet? Begrunn svaret.

Oppgave 2 – Parallele og distribuerte databaser – 15 % (10 % på a og 5 % på b)

- a) Beskriv de fire meldingene som blir brukt i tofase-commit (2PC).
- b) Ved bruk av masterkopi (primary copy) replikering har alle objekter (f.eks. en relasjon) en masternode som mottar alle oppdateringer. Bør man bruke samme masternode for alle objekter?

Oppgave 3 – Skyline – 15 %

- a) 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 R til høyre, og vis hva som er resultatet. Begrunn svaret.
- b) Lag et SQL-uttrykk som gjør en skyline-spørring på en relasjon med to attributter som den til høyre (med SQL uten støtte for skyline-operatoren).

X	Y
3	10
5	3
1	6
1	1
1	10
6	1
10	1
1	4
4	3
3	8

Oppgave 4 – Kolonnebaserte databasesystemer – 15 % (5 % på a og 10 % på b)

- a) Hva er en *projeksjon* i C-Store?
- b) Forklar de tre forskjellige metodene for komprimering av kolonner i C-Store og når disse bør brukes.

Oppgave 5 – Diverse – 30 % (12 % på a/b, 6 % på c)

- a) Forklar de 4 hovedmodulene i en *lagringshandler* (storage manager) i et databasesystem.
- b) Forklar arkitekturen til SkimpyStash (datastrukturer og hvordan de blir brukt).
- c) Forklar teknikken ”swizzling”.