

Institutt for datateknikk og informasjonsvitenskap

Eksamensoppgave i TDT4150 Avanserte databasesystemer

Faglig kontakt under eksamen: Jon Olav Hauglid

Tlf.: 93 80 58 51

Eksamensdato: Onsdag 1. juni 2016

Eksamenstid (fra-til): 9.00 – 13.00

Hjelpemiddelkode/Tillatte hjelpemidler: D. Ingen trykte eller håndskrevne hjelpemidler tillatt. Bestemt, enkel kalkulator tillatt.

Annen informasjon: Oppgavesettet inneholder 6 oppgaver. Det er angitt i prosent hvor mye hver (del-)oppgave teller ved sensur. Gjør rimelige antagelser der du mener oppgaveteksten er ufullstendig og skriv kort hva du antar. Lykke til!

Målform/språk: Bokmål

Antall sider (uten forside): 2

Antall sider vedlegg: 0

Kontrollert av:

Dato

Sign

Oppgave 1 – Datamodeller og arkitektur (15 %)

- a) Hva er fysisk og logisk data-uavhengighet? Hvorfor er det ønskelige egenskaper ved datamodeller?
- b) Hvorfor er det viktig at et databasesystem (DBMS) har kontroll over når data blir skrevet til permanent lager (dvs. at operativsystemet ikke påvirker dette)?

Oppgave 2 – Lese- og skriveoptimaliserte databasesystemer (10 %)

- a) Hvilke egenskaper ved data i en database og ved spørringer mot databasen gjør at kolonnebaserte databaser, som f.eks. C-store, er en bra løsning?
- b) Forklar hvorfor LSM-trær regnes som bedre på innsetting («writes») enn tradisjonelle B-trær.

Oppgave 3 – Spørreoptimalisering (25 %)

- a) Lag et eksempel som illustrerer at det kan være voldsom stor forskjell mellom en god og en dårlig utføringsplan for en gitt spørring.
- b) Anta spørringen «SELECT * FROM varer WHERE pris > 4000». Hvordan kan et databasesystem (DBMS) estimere antall tupler i svaret? Hva påvirker hvor nøyaktig dette estimatet blir?
- c) Anta en spørring som kun består av joins mellom tre tabeller der kun «left-deep» jointre blir vurdert. Vil valget mellom nested loop eller merge join kunne påvirke om samlebåndutføring av spørringen kan brukes? Begrunn svaret.
- d) Hvorfor er det ønskelig å transformere en nøstet spørring til en spørring uten nøsting? Hva er en korrelert subspørring?

Oppgave 4 – Distribuerte databasesystemer (15 %)

- a) Hva er forskjellen mellom partisjonering av en tabell og replikering av en tabell?
- b) Lag et eksempel som illustrerer at asynkron («lazy») oppdatering av replikater kan medføre at ACID ikke blir overholdt.

Oppgave 5 – Rangering og skyline (20 %)

- a) Forklar hva Skyline-operatoren gjør.
- b) Hvorfor blir ytelsen til Block Nested Loop (BNL)-algoritmen for skyline dårlig hvis vindusstørrelsen er for liten?
- c) Hva er en top-k spørring? Hva gjør Rank Join bedre enn vanlig join (f.eks. nested loop) ved utføring av top-k spørringer?

Oppgave 6 – Tema fra seminarartikler (15 %)

- a) Hvorfor bruker Earlybird to forskjellige indeksformat?
- b) Gitt en prosessor med mange kjerner. Hvis man ønsker god skalerbarhet, er det en fordel å bruke 2PL med vranglåsdetektor fremfor å abortere transaksjoner med en gang de må vente på en lås? Diskuter.
- c) Både et tradisjonelt databasesystem (DBMS) og det foreslåtte Anti-Caching databasesystemet har tupler både i minne og på disk. Hva er da forskjellen mellom dem?