

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



EKSAMENSOPPGAVE I FAG TDT4150 – AVANSERTE DATABASESYSTEMER

Faglig kontakt under eksamen: Kjetil Nørvåg

Tlf.: 41440433

Eksamensdato: 15. desember 2012

Eksamenstid: 09.00-13.00

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

Språkform: Bokmål

Sensurdato: 12. januar 2013

Oppgave 1 – Datamodeller – 10 %

Forklar logisk og fysisk data-uavhengighet ("logical and physical data independence") og hvorfor dette kan være viktige egenskaper for en datamodell.

Oppgave 2 – Spørreoptimalisering – 20 %

- a) Forklar hvorfor spørreoptimalisering ("query optimization") er en viktig komponent av et databasehåndteringsystem (DBMS).
- b) Anta at vi har to spørringer:
 Q1: SELECT * FROM r1 NATURAL JOIN r2;
 Q2: SELECT * FROM r1 NATURAL JOIN r2 ORDER BY r2.c1;
 Nevn noen måter utføringsplanen for Q2 kan skille seg fra utføringsplanen for Q1. Beskriv de antagelsene du gjør.

Oppgave 3 – Parallele og distribuerte databaser – 20 %

- a) Er horisontal eller vertikal partisjonering å foretrekke?
- b) Anta at to transaksjoner utføres samtidig fra to forskjellige noder og at de samtidig ønsker å skrive til samme objekt. Dette objektet finnes replisert på begge nodene. Hva skjer:
 1) Ved bruk av synkron masterkopi ("eager primary copy") replisering?
 2) Ved bruk av synkron gruppekopi ("eager update anywhere") replisering?
 3) Ved bruk av asynkron gruppekopi ("lazy update anywhere") replisering?

Oppgave 4 – Rank-join – 20 %

En turist ønsker å besøke Trondheim og ønsker å leie en bil. Han ønsker at bilutleiefirmaet skal være i samme området som hotellet, og har alternativene som vises i tabellene nedenfor. Kjør rank-join algoritmen for $k = 3$ på de to relasjonene nedenfor og forklar på hvilket steg algoritmen vil avslutte og terskelverdien ("threshold value") på dette trinnet (join på attributtet Area). Scorings-funksjonen er $f = 0.5 * PriceA + 0.5 * PriceB$.

ID	Rent a Car	Area	PriceA
a1	Wreck	Sentrum	100
a2	Lux	Flatåsen	110
a3	Bil	Sentrum	120
a4	RentBil	Flatåsen	130
a5	SuperBil	Flatåsen	140

ID	Hotel	Area	PriceB
b1	Cozy	Flatåsen	90
b2	Cheap	Sentrum	100
b3	Good	Sentrum	120
b4	Better	Flatåsen	120
b5	Best	Sentrum	150

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

- a) Forklar forskjellene mellom et tradisjonelt databasesystem og et strømdatahåndterings-system ("data stream management system").
- b) Forklar hovedprinsippene bak MapReduce (ta gjerne utgangspunkt i figuren i pensum-artikkelen).
- c) Forklar hovedprinsippene bak Google File System (ta gjerne utgangspunkt i arkitektur-figuren i pensum-artikkelen).
- d) Forklar motivasjonen bak Hive, og gi en kort oversikt over funksjonalitet som Hive tilbyr.