

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



**EKSAMENSOPPGAVE I FAG TDT4150 – AVANSERTE DATABASESYSTEMER**

**Faglig kontakt under eksamen: Kjetil Nørvåg**

**Tlf.: 93440**

**Eksamensdato: 30. mai 2011**

**Eksamenstid: 09.00-13.00**

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

**Språkform: Bokmål**

**Sensurdato: 20. juni 2011**

### Oppgave 1 – Spørreoptimalisering – 15 %

- Hvorfor ser System R sin optimalisator kun på "left-deep" join-trær?
- Kan man alltid endre rekkefølgen på seleksjon og projeksjon? Hva kan være årsaken til å bytte rekkefølge på slike operatører?

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

- Hva er partisjonert parallellitet i forbindelse med utføring av spørringer i distribuerte databaser? Vurder hvor enkelt det er å utnytte partisjonert parallellitet for seleksjon, join og aggregering.
- Beskriv kort i hvilken grad asynkron gruppe-replisering ("lazy update anywhere") oppfyller ACID-egenskapene.

### Oppgave 3 – Datagruvedrift – 15 %

Anta handlekurv-data til høyre. Bruk apriori-algoritmen for å finne hvilke assosiasjonsregler som gjelder, gitt at minimum støtte er 50 % (dvs. *minimum support count* er 2) og konfidens er 90 %.

TransaksjonsID	Element
T100	A,C,D
T200	B,C,E
T300	A,B,C,E
T400	B,E

### Oppgave 4 – Datastrømmer – 15 %

- Forklar konseptet kontinuerlige spørringer ("continuous queries") og hvordan de er forskjellige fra tradisjonelle spørringer. Gi tre eksempler på kontinuerlige spørringer.
- Forklar konseptet glidende vindu ("sliding windows") og deres rolle i et datastrøm-system ("Data Stream Management System").

### Oppgave 5 – Skyline og ranking – 25 % (5% på a og b og 15% på c)

- Er det mulig å skrive en SQL-spørring som gir samme resultat som en "skyline"-spørring (uten å bruke en SKYLINE-operator)? Om nei, beskriv hvorfor. Om ja, beskriv ulemper med denne metoden.
- Diskuter fordeler og ulemper med "Block Nested Loop (BNL)"-algoritmen for prosessering av skyline-spørringer. Forklar spesielt i hvilke scenarioer den fungerer bra og i hvilke scenarioer den ikke bør brukes.
- Forklar rank-join-algoritmen. Hva er dens fordeler i forhold til en naiv algoritme?

### Oppgave 6 – Diverse – 15 % (10% på a og 5% på b)

- Diskuter likheter mellom nettverksmodellen og XML, og hva som er potensielle problemer med nettverksmodellen.
- Beskriv problemene relatert til "under-provisioning" og "over-provisioning" og forklar hvordan "cloud computing" kan brukes for å minimere dem.