



Høgskolen i Østfold

EKSAMEN

Emnekode: ITF30307	Emne: Databaseadministrasjon og databasesystemer
Dato: 10.12.12	Eksamenstid: 09.00 - 12.00.
Hjelpemidler: ingen	Faglærer: Edgar Bostrøm / Per O. Bisseberg
Oppgavesettet består av 2 sider. Vedlegget består av 1 side. På mange av oppgavene kan det lønne seg å svare punktvis. I noen tilfeller holder det med en setning eller tre, i andre tilfeller bør det gjøres en beskrivelse/kommentar/drøfting på hvert av disse punktene. Tidsangivelsen pr. oppgave gir indikasjon på hvor mye man bør svare. Hver deloppgave teller likt.	
Sensurdato: 7. januar 2013 Karakterene er tilgjengelige for studenter på studentweb senest dagen etter oppgitt sensurfrist.	

Oppgave 1. Tid: 45 minutter.

Denne oppgaven tar for seg oppgaver knyttet til drift av ett eller flere DBMS, oppgave a og b baserer seg på kunnskap dere har tilegnet dere i administrasjon av ulike DBMS.

a) Backup

- Hvilke ulike backup-metoder har en DBA til rådighet?
 - i. Full backup/dump
 - ii. Inkrementell
 - iii. Differensiell
 - iv. Tabellkopiering
 - v. Replikasjon
 - vi. ++
- Drøft fordeler og ulemper ved disse backup-metodene.
 - i. Beskrive de ulike metodene.... Samt drøfting av fordeler og ulemper (om noen finnes)

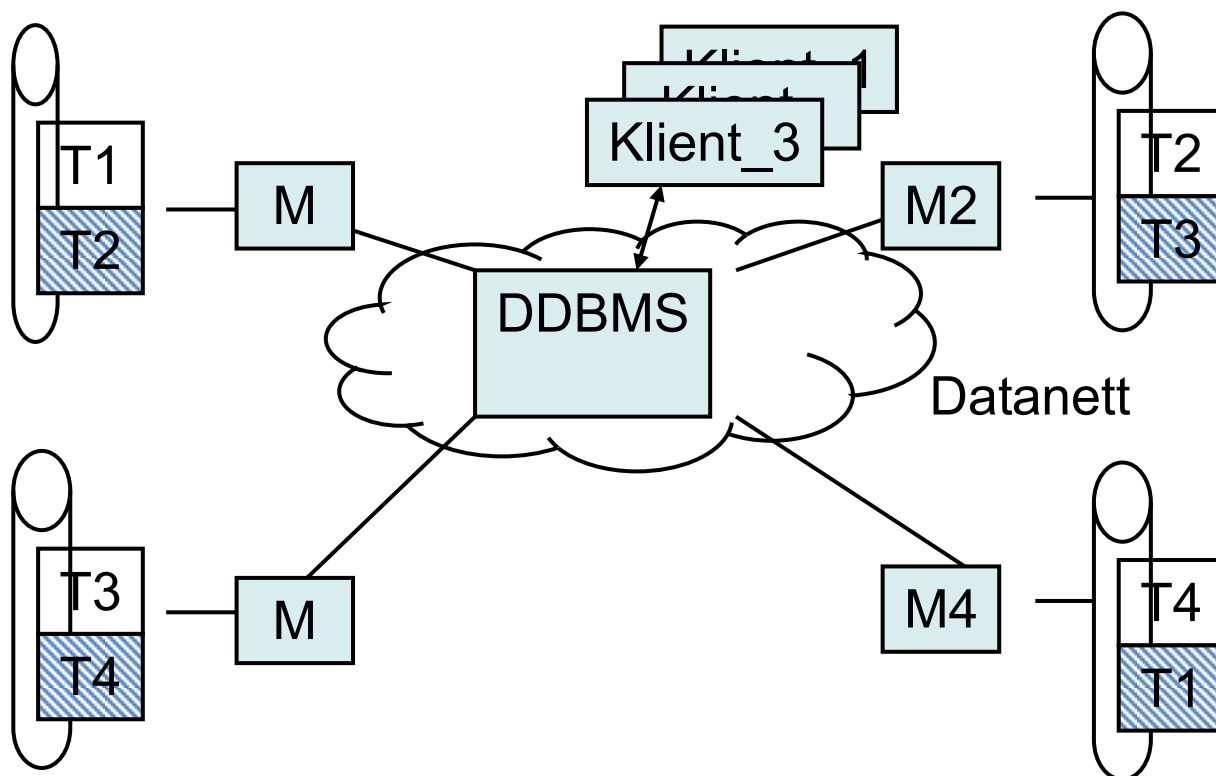
b) Hvilke metoder for spesifisering av brukertilgang/rettigheter er tilgjengelig for en DBA?

- Sette privilegier(tabelltilgang og host) på:
 - i. Brukernivå
 - ii. Roller
- Benytte Views
- etc

c) DDBMS:

- Hva er et DDBMS og hva karakteriserer et slikt system?
 - i. En distribuert database er en samling av flere logisk relaterte databaser som er forbundet med et datanettverk, hvor ddbms'et håndterer semantikk, transparens etc
- Lag en illustrasjon som viser en tenkt DDBMS arkitektur.

Noe slikt:



Oppgave 2. Tid: 45 minutter.

Velg 3 av disse 4:

- a) Bill Inmon definerte begrepet datavarehus som: "A subject-oriented, integrated, time-variant, and non-volatile¹ collection of data in support of management's decision-making process". Forklar hva som ligger i dette.

Se boka m.m.

- b) Hva ligger i begrepet Data Staging Area (DSA)?

Normalisert, renset, metadatakonsolidert mellomlagringsområde fra (mange) OLTP-systemer som deretter kan brukes for å lage et DW (evt. flere DM).

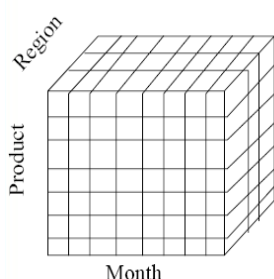
NB! Ønsker ikke et slikt svar, men dette er svært komprimert beskrivelse.

- c) Hva ligger i begrepet MOLAP (Multidimensional OLAP)?

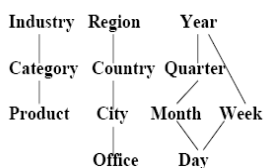
Litt om hva et OLAP er, deretter MOLAP. Har bl.a. figuren

Multidimensjonale data

- ♦ Sals-volum som funksjon av produkt, måned, og region (dimensjonar)



Dimensions: Product, Location, Time
Hierarchical summarization paths



i forelesningsnotatene. Gjerne litt om at det kan være ≥ 4 dimensjoner og bruk bl.a. til trenddata.

¹ Ikke-flyktig.

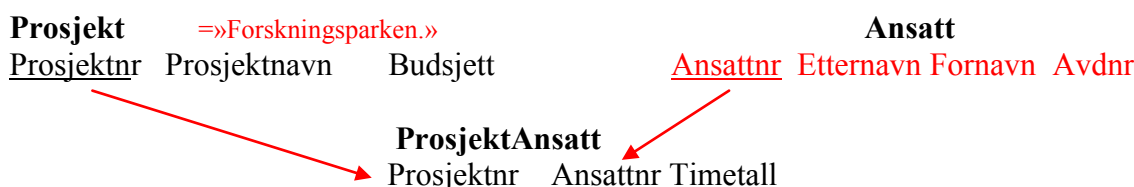
- d) Inmon står i utgangspunktet for en tanke om et felles datavarehus for en hel virksomhet. Finnes det alternative arkitekturer til dette? (Det er en fordel om du har med begrep, personnavn o.l. i forklaringen).

Her bør temaer som DM og mulighet for å analysere data uten fysisk å flytte de over i et eget system nevnes – kommer jo for en stor del an på ting som graden av ønske om samtidighet, prosesseringshastighet, behovet for datarensning m.m. – altså «virtuelle datavarehus/datamarts». Inmon vs. Kimball, bl.a. top-down & felles varehus vs. bottom-up & data marts som evt. kan integreres til felles DW.

Kommentar: Ikke forelest, men sto litt i boka og mye på nettet. Blir kanskje tynt på denne.

Oppgave 3. Tid: 45 minutter.

I deloppgave a) og b) tar vi for oss samme relasjoner som i undervisningen. Syntaks for den delen av relasjonsalgebra som vi har brukt finnes i vedlegget.



- a) Lag utsagn i relasjonsalgebra for:
- Ansattnr, etternavn, fornavn og avdnr på ansatte som har jobbet i prosjektet med prosjektnavn 'Forskningsparken'. Ikke bruk semijoin.

$$\pi_{Ansattnr, Etternavn, Fornavn, Avdnr} (Ansatt \bowtie_{Projektnavn = 'Forskningsparken'} ProjektAnsatt)$$
evt., mer elegant i dette tilfelle:

$$Ansatt \bowtie_{\pi_{Ansattnr} (ProjektAnsatt \bowtie_{Projektnavn = 'Forskningsparken'} Projekt))}$$
 - Som over, men med bruk av semijoin.

$$Ansatt \bowtie_{(ProjektAnsatt \bowtie_{Projektnavn = 'Forskningsparken'} Projekt)}$$
Også her kunne man ha projisert mer.
- b) Lag utsagn i relasjonsalgebra for:
- Prosjekter som ikke er kommet ordentlig i gang, i betydningen at de ikke har noen tilsvarende tupler i ProjektAnsatt. Projektnr og prosjektnavn skal være med.

$$Projekt \bowtie_{(\pi_{Projektnr} (Projekt) \setminus \pi_{Projektnr} (ProjektAnsatt))}$$
 - Eventuelle prosjekter som har involvert alle ansatte. Alt fra relasjonen Projekt skal være med.

$$Projekt \bowtie_{(\pi_{Projektnr, Ansattnr} (ProjektAnsatt) / \pi_{Ansattnr} (Ansatt))}$$
- c) Forklar sammenhengen mellom relasjonsalgebra og optimalisering.
I optimalisering trenger man å splitte opp f.eks. et SQL-utsagn i mindre deler, for deretter å omrokere det og finne en ekvivalent, men mest mulig optimal algoritme for gjennomføring. Relasjonsalgebra er velegnet for en slik oppsplitting.

Tips for del a) og b): Se syntaks i vedlegget.

Oppgave 4. Tid: 45 minutter.

- a) Du har følgende tabell(ANSATT) i en relasjonsdatabase. Skriv tilsvarende datasett som et XML-dokument.

Ansatt_id	Fornavn	Etternavn	Tlf_nr	Avd_id
ANS001	Per	Biffeberg	62916299	AVD007
ANS007	Edvald	Boastrøm Hagen	62616123	AVD001
ANS123	Hugo	Boss	62312234	AVD001

NB! Krever ikke at de har med header

```
<ansatte>
  <ansatt Ansatt_id="ANS001" Avd_id="AVD007">
    <fornavn>Per</fornavn>
    <etternavn>Biffeberg</etternavn>
    <tlf_nr>62916299</tlf_nr>
  </ansatt>
  <ansatt Ansatt_id="ANS007" Avd_id="AVD001">
    <fornavn>Edvald</fornavn>
    <etternavn>Boastrøm Hagen</etternavn>
    <tlf_nr>62616123</tlf_nr>
  </ansatt>
  <ansatt Ansatt_id="ANS123" Avd_id="AVD001">
    <fornavn>Hugo</fornavn>
    <etternavn>Boss</etternavn>
    <tlf_nr>62312234</tlf_nr>
  </ansatt>
</ansatte>
```

alternativt men kanskje lavere score??:

```
<ansatte>
  <ansatt>
    <ansatt_id>ANS001</ansatt_id>
    <fornavn>Per</fornavn>
    <etternavn>Biffeberg</etternavn>
    <tlf_nr>62916299</tlf_nr>
    <avd_id>AVD007</avd_id>
  </ansatt>
  :
</ansatte>
```

- b) Ved å benytte XQuery skal du hente ut all informasjon om ansatte som jobber i AVD001 fra XML-dokumentet du laget i oppgave a.
- Med Xpath (baserer her på mitt første eksempel)
 - ansatte/ansatt[@Avd_id="AVD001"] eller
 - //ansatt[@Avd_id="AVD001"]
 - Med FLWOR
for \$e in /ansatte/ansatt
where \$e/@Avd_id="AVD001"
return \$e
- c) XQuery benytter både XPath og FLWOR.
- Hva er XPath og FLWOR
 - Hva skiller disse?
 - Se boka s 1103 - 1108 i 4de ed
- d) Hvilke fordeler kan datalagring i XML tilby?
- Menneskeleselig
 - Godt egnet for å lagre semistrukturerte data, dvs. mindre strikt enn vanlige RDBMS-er.
 - Velegnet for konfigurering (f.eks. i stedet for .INI-filer)
 - Dokumentformat
 - Anvendelig i Web applikasjoner
 - Transformasjonsvennlig
 - Open Source
 - etc