

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i                    INF3100/INF4100 — Databasesystemer

Eksamensdag:            8. juni 2007

Tid for eksamen:        14.30 – 17.30

Oppgavesettet er på 4 sider.

Vedlegg:                 Ingen

Tillatte hjelpemidler: Kalkulator og ordbok

Kontroller at oppgavesettet er komplett før  
du begynner å besvare spørsmålene.

*Les oppgavene nøye, og lykke til!*

### Oppgave 1 Relasjonsdatabasespråk (45%)

En akvariebutikk har en relasjonsdatabase med informasjon om hvilke fiskearter de enkelte kundene har kjøpt. Dessuten er det informasjon om de problemene kundene melder inn om sykdom på fisken og liknende. I den forbindelsen skal vi se på følgende tre relasjoner:

`Salg(salgID, kundeID, artsnavn, dato, antall)`

`Hendelse(hendID, salgID)`

`Epikrise(hendID, dato, status, info)`

`Salg` inneholder opplysninger om salg av fisk. For hver fisketype solgt til en kunde på en gitt dato, tilordnes en `salgID`. Denne utgjør primærnøkkelen i `Salg`. Dessuten angis kundens `kundeID`, hvilken fisketype (`artsnavn`) salget gjaldt, `dato` og `antall` solgt av denne arten. (`kundeID`, `artsnavn`, `dato`) er kandidatnøkkel i `Salg`.

For å holde rede på de problemene kundene påtreffer, tilordnes hvert problem (`hendelse`) en hendelses-id, `hendID`. I relasjonen `Hendelse` angis hvilket salg hendelsen gjaldt. I relasjonen `Epikrise` noteres fortløpende informasjon knyttet til hver hendelse, i form av et stikkord `status` og en mer utførlig beskrivelse `info` samt `datoen` for informasjonen. Når en ny hendelse oppstår, settes det inn ett tuppel i `Hendelse` og ett i `Epikrise`, det sistnevnte med status `initieell`, en beskrivelse av hendelsen og dato for hendelsen. På senere tidspunkter legges så for samme hendelses-id inn ett eller flere nye tupler i `Epikrise` med andre statusverdier, som `diagnose/frisk/død`, en beskrivelse og dato for statusen.

I `Hendelse` er `hendID` primærnøkkel; attributtet `salgID` er fremmednøkkel til `Salg`. I `Epikrise` er (`hendID`, `dato`) primærnøkkel.

*(Fortsettes på side 2.)*

**1a (5%)**

Bruk SQL til å definere relasjonene `Salg` og `Hendelse` med integritetsregler.

**1b (5%)**

Bruk SQL til å finne antall forskjellige kunder som kjøpte fiskearten 'Black Molly' i 2006.

**1c (10%)**

Bruk SQL til å finne salg foretatt i november 2006 der fiskene i salget har vært gjenstand for mer enn to hendelser. I svaret skal du ta med `salgsID`-en, artsnavnet salget gjaldt og antall hendelser.

**1d (15%)**

Bruk SQL til å finne alle salg (identifisert ved deres `salgsID`-er) foretatt i januar 2007 der fiskene i salget senere har vært gjenstand for minst én hendelse og samtlige hendelser har `diagnose` som en av statusene. Sorter resultatet etter salgsdato.

**1e (10%)**

Bruk relasjonsalgebra til å besvare spørsmålet i 1c.

## Oppgave 2 FDer og normalformer(20%)

Gitt relasjonen

Adresse(by, postnr, gate)

der by, gate  $\rightarrow$  postnr og postnr  $\rightarrow$  by.

### 2a (5%)

Bestem kandidatnøkklene til Adresse.

### 2b (5%)

Angi brudd på 3NF og BCNF i Adresse.

### 2c (5%)

Dekomponer Adresse til BCNF. Er dekomposisjonen FD-bevarende? Begrunn svaret ditt.

### 2d (5%)

Hvilke grunner kan det være til å beholde Adresse uendret snarere enn å dekomponere den til BCNF?

## Oppgave 3 Treprotokollen (20%)

### 3a (5%)

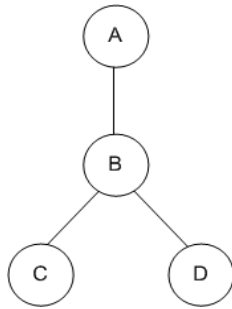
I hvilke situasjoner benytter man treprotokollen for å serialisere?

### 3b (5%)

Beskriv prinsippet for treprotokollen. Du kan anta at vi har én låstype (eksklusive låser).

(Fortsettes på side 4.)

Anta at vi har datastrukturen



Gitt følgende eksekveringsplan for to transaksjoner  $T_1$  og  $T_2$ :

$r_2(A); r_2(B); r_1(A); r_2(D); r_1(B); r_1(C); w_2(D); w_1(C); w_2(B)$

### 3c (10%)

Sett inn forespørsler om eksklusive låser og frigivelser av låser i eksekveringsplanen i henhold til treprotokollen. Låsene skal frigis så snart som mulig. Beskriv hvordan eksekveringen av planen vil forløpe.

## Oppgave 4 TPMMS (15%)

### 4a (10%)

Beskriv algoritmen two-phase, multiway merge-sort (TPMMS). Hva er kostnaden til algoritmen (i antall blokk-I/O-er)?

### 4b (5%)

Hva er det som avgjør når man skal bruke TPMMS til å sortere en fil/relasjon, og når man bør/må benytte andre algoritmer? Hvilke alternative algoritmer er aktuelle, og når?