

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i INF3100 — Databasesystemer

Eksamensdag: 8. juni 2015

Tid for eksamen: 14.30 – 18.30

Oppgavesettet er på 4 sider.

Vedlegg: ingen

Tillatte hjelpemidler: Kalkulator og ordbok

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

Vi skal se på en liten sykehusdatabase. Relasjonene i databasen er som følger:

Pasient(*pnr*, *fnr*, *navn*, *adr*)

Sykehusopphold(*pnr*, *onr*, *inndato*, *utdato*, *avd*, *adato*)

Labprøve(*prvnr*, *pnr*, *prdato*, *prtype*, *atype*, *resultat*)

Hver pasient har et entydig pasientnummer (*pnr*). I relasjonen **Pasient** er attributtene *fnr*, *navn* og *adr* henholdsvis pasientens fødselsnummer, navn og adresse; hvis noen av opplysningene er ukjente, kan ett eller flere av disse attributtene være tomme (inneholde **null**).

En pasient kan ha flere opphold ved sykehuset, der et opphold er et dagsbesøk eller en innleggelse. Hvert opphold identifiseres ved pasientnummer kombinert med et oppholdsnummer (*onr*). Informasjon om oppholdene er samlet i relasjonen **Sykehusopphold**. Her er *inndato* og *utdato* henholdsvis innleggings- og utskrivelsesdato. Hvis pasienten bare har vært på dagsbesøk, er disse to datoene like. Attributtet *avd* angir hvilken avdeling pasienten er på. Ved langvarige sykehusopphold kan en pasient bli flyttet mellom avdelinger, derfor er det også tatt med et attributt *adato* som angir når pasienten ble lagt inn på (eller flyttet til) en avdeling.

Relasjonen **Labprøve** inneholder informasjon om alle laboratorieprøver. Hver prøve tilordnes et entydig nummer (*prvnr*). Attributtet *pnr* angir hvilken pasient prøven gjelder, *prdato* er datoen da prøven ble tatt, *prtype* hva slags prøve det er snakk om, *atype* hva slags analyse som skal gjøres, og *resultat* svaret på analysen. Et eksempel er at *prtype* = 'blodprøve' og *atype* = 'LPK', der LPK er en forkortelse for leukocytt partikkelkonsentrasjon (dvs. antall hvite blodlegemer pr. liter blod).

(Fortsettes på side 2.)

Oppgave 1 FDer (20%)

I relasjonen **Sykehusopphold** gjelder følgende funksjonelle avhengigheter (FDer):

$pnr, onr \rightarrow inndato, utdato$

$pnr, inndato \rightarrow onr$

$pnr, adato \rightarrow onr, avd$

- (i) (10%) Finn høyeste normalform som **Sykehusopphold** oppfyller. Begrunn svaret.
- (ii) (5%) Dekomponer **Sykehusopphold** tapsfritt til BCNF. Er dekomposisjonen FD-bevarende? Begrunn svaret.

Ofte utføres det flere analyser fra hver laboratorieprøve, f.eks. kan én blodprøve brukes til å måle både blodprosent og senkning. Imidlertid er det aldri aktuelt å foreta to identiske analyser av en og samme laboratorieprøve, så det er aldri mer enn ett resultat pr. analysetype.

- (iii) (5%) Formuler denne restriksjonen i form av en FD.

Oppgave 2 SQL (25%)

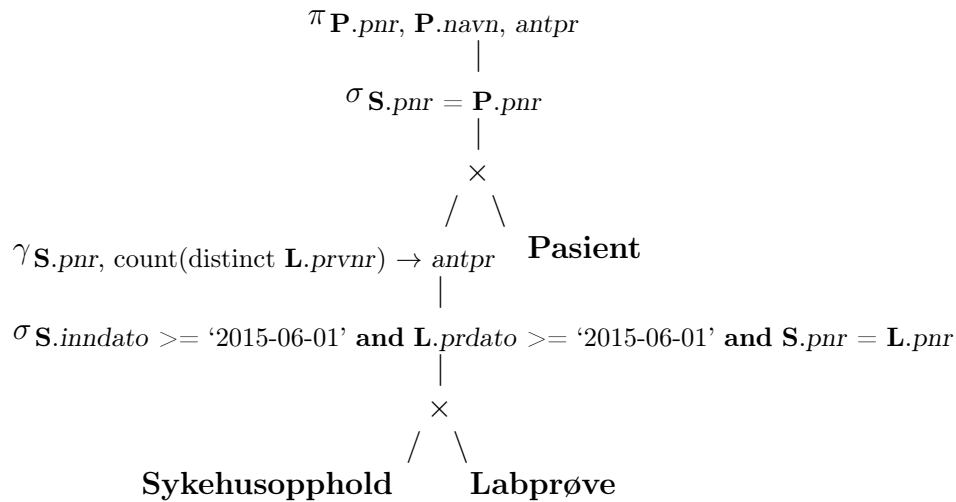
Besvar følgende spørsmål ved hjelp av SQL:

- (i) (10%) Finn ut om det er noen pasienter som har hatt mer enn ett opphold ved sykehuset, men der sykehuset ikke kjenner navnet på pasienten og heller ikke har tatt noen laboratorieprøver av pasienten. Skriv ut pasientnummeret og antall opphold for hver slik pasient.
- (ii) (15%) Blant de pasientene som på hvert av sine dagsbesøk har tatt nøyaktig én prøve, finn de pasientene som tar samme prøvetype hver gang. Ta bare med pasienter som har hatt 5 eller flere dagsbesøk. Skriv for hver slik pasient ut pasientnummeret, prøvetype og analysetyper. (Det kan være at det varierer fra gang til gang hvilken analyse som er foretatt av prøven selv om prøvetypen alltid er den samme. Det kan også være at det er foretatt flere analyser av en prøve.)

(Fortsettes på side 3.)

Oppgave 3 Relasjonsalgebra (20%)

Under finner du et relasjonsalgebrauttrykk. (For å spare plass har vi tildels skrevet **P** for **Pasient**, **S** for **Sykehusopphold** og **L** for **Labprøve**.)



- (i) (5%) Forklar med ord hvilken spørring uttrykket beskriver.
- (ii) (5%) Uttrykket utgjør en logisk spørreplan. Optimaliser den logiske spørreplanen.
- (iii) (10%) Uttrykk i relasjonsalgebra at en pasient ikke kan ha to forskjellige opphold med samme utskrivelsesdato.

Oppgave 4 Transaksjonsprotokoller (25%)

Vi skal utføre følgende to transaksjoner, der $r_i(y)$ og $w_i(y)$ som vanlig betyr at transaksjonen T_i henholdsvis leser og skriver et dataelement y , og c_i betyr at T_i committer:

$$\begin{aligned}
 T_1 &= r_1(a); r_1(c); w_1(a); r_1(b); w_1(b); c_1 \\
 T_2 &= r_2(a); r_2(c); w_2(c); c_2
 \end{aligned}$$

(Fortsettes på side 4.)

Gitt eksekveringsplanen

$$S = r_1(a); r_1(c); w_1(a); r_2(a); r_2(c); w_2(c); r_1(b); w_1(b); c_1; c_2$$

- (i) (5%) Avgjør om S er konfliktserialiserbar.

Anta at systemet tilbyr eksklusive låser. La $l_i(y)$ betegne at en transaksjon T_i tar en (eksklusiv) lås på dataelementet y , og la $u_i(y)$ betegne at T_i frigir låsen på y .

Vi minner om at strikt tofaselåsing (strikt 2PL) skiller seg fra vanlig 2PL ved at en transaksjon i strikt 2PL ikke frigir eksklusive låser før den har gjort commit (eller abort).

- (ii) (5%) Sett inn låseaksjoner i hver av T_1 og T_2 i henhold til strikt 2PL.
- (iii) (5%) Begrunn hvorfor S ikke kan genereres av en strikt 2PL-scheduler.
- (iv) (10%) Finnes det en ikke-seriell eksekveringsplan for T_1 og T_2 som kan genereres av en strikt 2PL-scheduler? Begrunn svaret.

Oppgave 5 RAID-teknologier (10%)

RAID 51 er en RAID-teknologi der RAID 5 og RAID 1 kombineres, dvs. en gruppe av n disker er organisert som RAID 5 (paritetsblokker fordelt på alle diskene i gruppen), og så speiler vi hele gruppen. (Totalt trengs altså $2n$ disker. Vanligvis er $n = 4$ eller $n = 5$ slik at det i alt er 8 eller 10 disker.)

- (i) (5%) Hva er det minste antall samtidige diskkræsje som kan gi varig tap av data? Begrunn svaret.
- (ii) (5%) Hva er det største antall disker som kan kræsje samtidig uten at vi taper data? Begrunn svaret.