

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i INF3100 — Databasesystemer

Eksamensdag: 10. juni 2014

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.

*Les oppgavene nøye, og lykke til!*

Alle boliger (eneboliger, leiligheter osv.) i Norge har et entydig *matrikkelnummer* som består av fem heltall (*knr*, *gnr*, *bnr*, *fnr*, *snr*). I tillegg kan alle boliger identifiseres entydig ved en adresse på formen (*adr*, *bolignr*) der *adr* er gateadresse og *bolignr* er bolignummer (*bolignr* brukes bare for å skille mellom leiligheter med lik gateadresse; hvis det ikke trengs, har det verdien 0).

Noen eiendomsmeglere har følgende felles (ikke perfekte) database over salg av boliger:

Matrikkel(*mnr*, *knr*, *gnr*, *bnr*, *fnr*, *snr*)

Boligsalg(*salgsnr*, *mnr*, *adr*, *bolignr*, *salgsdato*, *boligtype*, *areal*, *pris*)

Salgspart(*salgsnr*, *personnr*, *salgsrolle*)

Person(*personnr*, *navn*, *tlf*)

Relasjonen Matrikkel inneholder matrikkelnummere for alle boliger i landet. Hvert matrikkelnummer (*knr*, *gnr*, *bnr*, *fnr*, *snr*) er tilordnet en entydig verdi *mnr* som utgjør primærnøkkelen i relasjonen. I Matrikkel er derfor både (*mnr*) og (*knr*, *gnr*, *bnr*, *fnr*, *snr*) kandidatnøkler.

For hvert boligsalg tilordnes et entydig salgsnummer (*salgsnr*). Relasjonen Boligsalg inneholder i tillegg til salgsnummer og attributter som identifiserer boligen, informasjon om når salgskontrakten ble undertegnet (*salgsdato*), *boligtype* ('enebolig', 'tomannsbolig', 'rekkehus' eller 'leilighet'), boligens *areal* og salgsprisen (*pris*). Vi antar at relasjonen Boligsalg har følgende egenskaper: (i) En bolig kan selges flere ganger, men ikke på samme dag. (ii) En bolig endrer aldri matrikkelnummer eller adresse, så man kan bruke enten *mnr* eller *adr* og *bolignr* til å finne alle salg av en gitt bolig.

(Fortsettes på side 2.)

Relasjonen Salgspart inneholder informasjon om hvem som var involvert i salget av en bolig. Attributtet *salgsrolle* har en av verdiene ‘selger’, ‘kjøper’ og ‘megler’. Selgere, kjøpere og meglere identifiseres ved personnummer (*personnr*). Ved et gitt boligsalg kan maksimalt én person ha salgsrollen ‘megler’.

Relasjonen Person inneholder tilleggsopplysninger om personer (kjøpere, selgere og meglere).

Primærnøkler er understreket. Fremmednøkler er angitt ved likelydende attributtnavn; f.eks. er *salgsnr* i Salgspart fremmednøkkel til *salgsnr* i Boligsalg.

## Oppgave 1 FDer (20%)

- (i) (5%) Angi hvilke funksjonelle avhengigheter (FDer) som gjelder i relasjonen Boligsalg.
- (ii) (10%) Finn høyeste normalform som Boligsalg oppfyller. Begrunn svaret.
- (iii) (5%) Dekomponer Boligsalg tapsfritt til BCNF.

## Oppgave 2 SQL (25%)

Ta utgangspunkt i de opprinnelige relasjonene Matrikkel, Boligsalg, Salgspart og Person. Besvar følgende spørsmål ved hjelp av SQL (du kan gjerne bruke views som del av besvarelsen):

- (i) (5%) Finn ut om det er noen som har alle tre salgsroller (dvs. både selger, kjøper og megler) for en og samme bolig. (Det kan være flere slike personer.) Skriv ut navn, personnummer og hvilken/hvilke boliger det gjelder i form av *adr* og *bolignr*.
- (ii) (10%) Finn den/de boligene som har endret boligtype flest ganger. Skriv for hver ut *mnr* og antall endringer.
- (iii) (10%) Finn de boligene som har vært solgt minst én gang uten at det er oppgitt noen megler i forbindelse med salget. Skriv for hver ut matrikkelnummeret (dvs. *knr*, *gnr*, *bnr*, *fnr* og *snr*).

(Fortsettes på side 3.)

### Oppgave 3 Relasjonsalgebra (20%)

Uttrykket under finner antall salg pr. megler i 2013. (For å spare plass har vi skrevet B for Boligsalg, S for Salgspart og P for Person.)

$$\begin{array}{c}
 \pi_{P.navn, antsalg} \\
 | \\
 \gamma_{P.personnr, P.navn, \text{count}(B.salgsnr) \rightarrow antsalg} \\
 | \\
 \sigma_{S.salgsrolle = \text{'megler'} \text{ and } B.salgsdato \text{ like } \text{'2013\%'}} \\
 | \\
 \sigma_{B.salgsnr = S.salgsnr \text{ and } S.personnr = P.personnr} \\
 | \\
 \times \\
 / \backslash \\
 B \quad S \quad P
 \end{array}$$

- (i) (10%) Uttrykket utgjør en logisk spørreplan. Optimaliser den logiske spørreplanen.
- (ii) (10%) Uttrykk i relasjonsalgebra at det for et gitt boligsalg maksimalt kan være én megler.

### Oppgave 4 Transaksjonsprotokoller (20%)

Gitt tre transaksjoner

$$T_1 = r_1(a); r_1(b); w_1(a); w_1(b)$$

$$T_2 = r_2(c); r_2(a); w_2(c); w_2(a)$$

$$T_3 = r_3(a); r_3(b); r_3(c); w_3(c)$$

og eksekveringsplanen

$$S = r_1(a); r_1(b); r_2(c); r_2(a); r_3(a); r_3(b); r_3(c); w_1(a); w_1(b); w_2(c); w_2(a); w_3(c)$$

Vi skal se på bruk av snapshot isolation (SI) til samtidighetskontroll av transaksjonene. La  $l_i(x)$  bety at en transaksjon  $T_i$  ber om skrive-lås på elementet  $x$ . La  $u_i(x)$  bety at  $T_i$  frigir sin lås på  $x$ .

- (i) (10%) Sett inn aksjoner på formen  $l_i(x)$  og  $u_i(x)$  i  $T_1$ ,  $T_2$  og  $T_3$  i henhold til SI-protokollen FUW (første oppdaterer vinner).

(Fortsettes på side 4.)

- (ii) (10%) Beskriv hvordan eksekveringsforløpet blir i henhold til FUW-protokollen hvis eksekveringen i størst mulig grad skal følge eksekveringsplanen  $S$ . Ta for hver  $T_i$  med en aksjon av formen  $c_i$  ( $T_i$  committer) eller  $a_i$  ( $T_i$  aborteres/må rulles tilbake) på passende sted i eksekveringsforløpet.

## Oppgave 5 RAID-teknologier (15%)

RAID 1E er en diskteknologi der man kombinerer striping og speiling over et odde antall disk, dvs. antall disk er  $2m + 1$  for en  $m \geq 1$ . Hver datablokk er speilet på neste disk i RAIDet (unntaket er når datablokken er på den siste disken; da er speilingen på første disk i RAIDet). Hvis det f.eks. er 5 disk  $d_0, d_1, d_2, d_3, d_4$  og 10 datablokker A, B, C, D, E, F, G, H, I og J, så fordeler blokkene seg slik:

A	A	B	B	C
C	D	D	E	E
F	F	G	G	H
H	I	I	J	J
$d_0$	$d_1$	$d_2$	$d_3$	$d_4$

(Formelt kan RAID 1E beskrives slik: Hvis diskene er  $d_0, d_1, \dots, d_{2m}$  og blokkene er  $b_0, b_1, b_2, \dots$ , så er blokk  $b_i$  og dens speiling plassert på henholdsvis disk  $d_{2i}$  og disk  $d_{2i+1}$ , der  $2i$  og  $2i + 1$  beregnes modulo  $2m + 1$ . Så hvis det er 5 disk, er blokk  $b_0$  – A i figuren – plassert på disk  $d_{0 \bmod 5} = d_0$  og disk  $d_{1 \bmod 5} = d_1$ , mens f.eks. blokk  $b_7$  – H i figuren – er plassert på disk  $d_{14 \bmod 5} = d_4$  og disk  $d_{15 \bmod 5} = d_0$ .)

Gitt et RAID 1E med  $2m + 1$  disk.

- (i) (5%) Hva er minste antall diskkræsje som kan gi varig tap av data? Begrunn svaret.
- (ii) (10%) Hva er det største antall disk som kan kræsje samtidig uten at vi taper data? Begrunn svaret.