

Institutt for datateknikk og informatikk

Eksamensoppgave i TDT4145 Datamodellering og databasesystemer

Faglig kontakt under eksamen:

Roger Midtstraum: 995 72 420

Svein Erik Bratsberg: 995 39 963

Eksamensdato: 7. august 2017

Eksamenstid (fra-til): 09:00-13:00

Hjelpemiddelkode/Tillatte hjelpemidler:

D – Ingen trykte eller håndskrevne hjelpemidler tillatt. Bestemt, enkel kalkulator tillatt.

Annen informasjon:

Målform/språk: Norsk bokmål

Antall sider (uten forside): 5

Antall sider vedlegg: 0

Informasjon om trykking av eksamensoppgave

Originalen er:

1-sidig ☐ **2-sidig X**

sort/hvit X **farger** ☐

skal ha flervalgskjema ☐

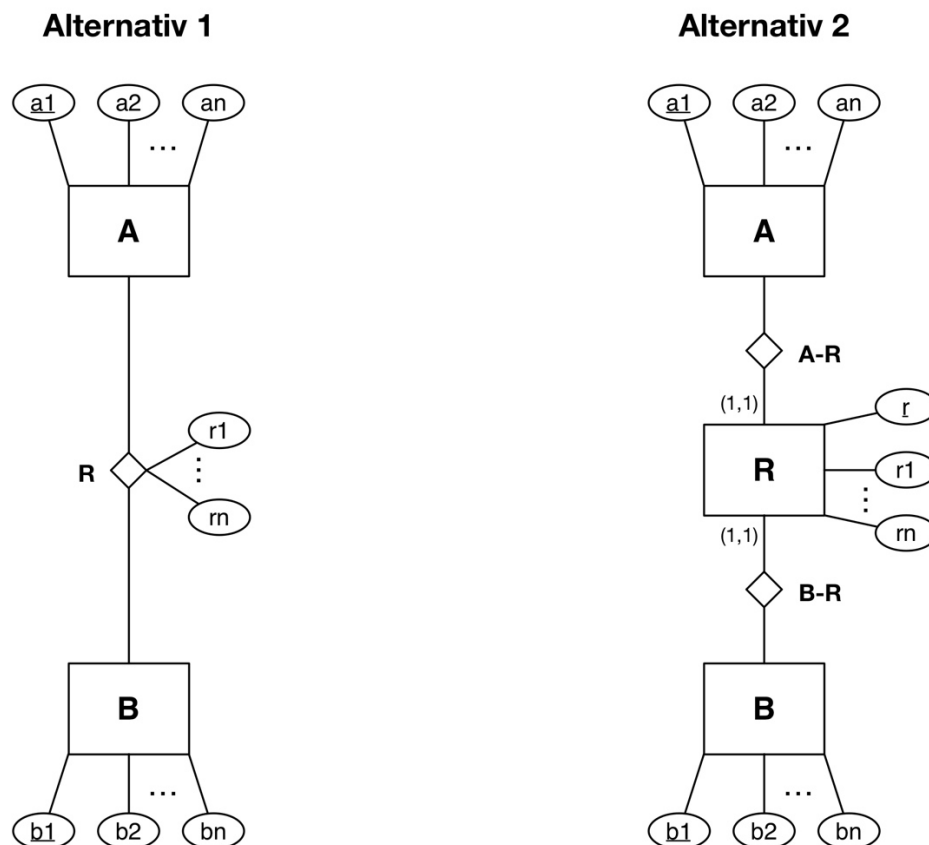
Kontrollert av:

Dato

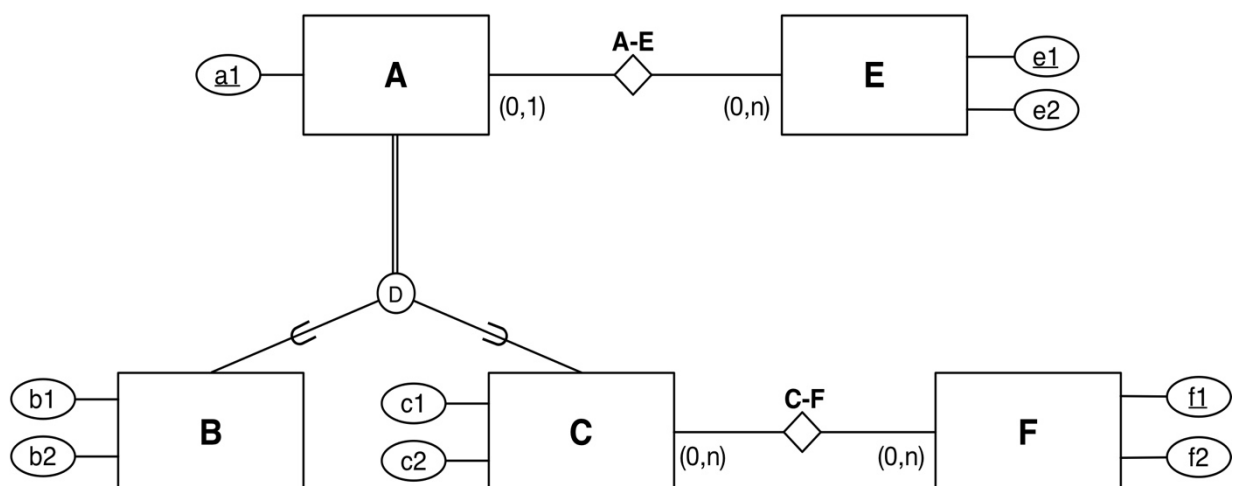
Sign

Oppgave 1 – Datamodeller og relasjonsdatabaser (15 %)

a) I figuren under har vi vist to ulike måter å representere en binær relasjon mellom entiteter i entitetsklassene A og B. Diskuter fordeler og ulemper ved de to alternativene.



b) I figuren under har vi vist en ER-modell. Oversett denne til en hensiktsmessig relasjonsdatabasemodell. Gjør kort rede for eventuelle forutsetninger som du finner det nødvendig å gjøre.



Oppgave 2 – Teori (10 %)

- a) Når vi designer relasjonsdatabaser kan vi velge om vi vil tillate *redundans* i datalagringen. Diskuter fordeler og ulemper ved å tillate redundans.
- b) I relasjonsdatabaser har vi en spesiell NULL-verdi. Hva er hensikten med å ha en slik spesiell verdi? Kunne vi klart oss uten?

Oppgave 3 – Relasjonsdatabaser, ER-modeller, relasjonsalgebra og SQL (20 %)

Gitt følgende relasjonsdatabase-skjema for en database over verdensdeler, land og byer. Primærnøkler er understreket.

Verdensdel (VID, VNavn)

Land (LID, LNavn, Areal, Folketall, VDelID)

VDelID er fremmednøkkel mot Verdensdel-tabellen. Kan ikke ha NULL-verdi.

By (BID, BNavn, Folketall, LandID, HovedstadFraAAR)

LandID er fremmednøkkel mot Land-tabellen. Kan ikke ha NULL-verdi. Dersom byen er hovedstad i landet, har den et årstall som verdi for HovedstadFraAAR, som ellers har NULL-verdi.

Relasjonsalgebra kan formuleres som tekst eller grafer. Hvis du behersker begge notasjonene foretrekker vi at du svarer med grafer, men du blir ikke trukket for å svare med tekst.

- a) Lag et ER-diagram (du kan bruke alle virkemidler som er med i pensum) som i størst mulig grad samsvarer med relasjonsskjemaet. Gjør rede for eventuelle antagelser du finner det nødvendig å gjøre.
- b) Lag en spørring i *relasjonsalgebra* som finner navn på by, navn på land og byens folketall for alle hovedsteder i Afrika.
- c) Lag en spørring i *SQL* som finner verdensdel-ID, verdensdel-navn, antall land i verdensdelen, verdensdelens areal og verdensdelens folketall.
- d) Lag en spørring i *SQL* som finner by-navn, byens folketall, land-navn og verdensdel-navn for alle byer som har flere enn 5 millioner innbyggere. Resultatet skal være sortert etter folketall i synkende ordning, deretter på byens navn i stigende rekkefølge.
- e) Lag et *SQL-skript* som oppretter Land-tabellen. Velg hensiktsmessige datatyper for attributtene (vi vektlegger ikke detaljer). Gjør rede for eventuelle antagelser du finner det nødvendig å gjøre.

Oppgave 4 - Normaliseringsteori (15 %)

- a) Anta $R = \{A, B, C, D\}$ og $F = \{A \rightarrow B; BC \rightarrow D\}$. Under er vist en tabellforekomst for R. Bestem hvilke verdier S, T, U, V og W kan ha for at tabellforekomsten skal være gyldig.

A	B	C	D
3	1	2	2
2	1	2	S
T	1	U	3
4	6	1	0
5	V	W	4

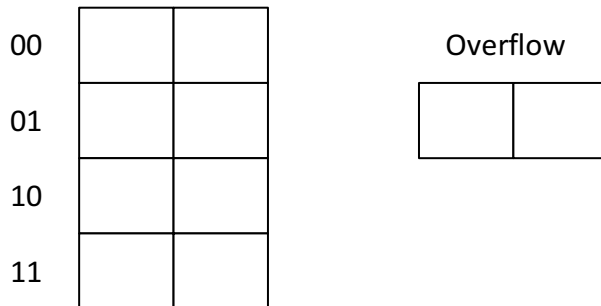
- b) Gitt $R = \{A, B, C, D, E\}$ og $F = \{AC \rightarrow BD; EC \rightarrow BD; A \rightarrow E; E \rightarrow A\}$. Forutsett at R er på første normalform. Hva er den høyeste normalformen som er oppfylt av R? Svaret må begrunnes.
- c) Gitt tabellen $R = \{A, B, C\}$. Attributtene har samme verdimengde:
 $\text{dom}(A) = \text{dom}(B) = \text{dom}(C) = \{1, 2, 3\}$.
- i) Anta $F = \Phi$. Hva er det maksimale antall rader som kan finnes i en tabellforekomst ($r(R)$)?
- ii) Anta $F = \{A \rightarrow BC\}$. Hva er det maksimale antall rader som kan finnes i en tabellforekomst?
- d) En tabell, R, splittes i to (komponent-)tabeller, R_1 og R_2 . Forklar hva det vil si at dekomponeringen har *tapsløst-join egenskapen* (eng: lossless join). Vurder om det er nødvendig å ha denne egenskapen eller ikke. Svaret må begrunnes.
- e) Gitt $R = \{A, B, C\}$ og $F = \{A \rightarrow B\}$. En vanlig feil er å tro at $A \rightarrow B$ medfører at "A fører til eller bestemmer B". Forklar hvorfor dette ikke er en god forklaring på den funksjonelle avhengigheten.

Oppgave 5 – Statisk hashing (5 %)

Vi skal sette inn følgende nøkler i en statisk hashing-struktur: 27, 18, 9, 7, 16, 13, 11. Vi har 4 blokker og en overflytsblokk. Hver blokk har plass til to nøkler. Se figuren under. Du kan bruke hashfunksjonen:

$$h(K) = K \text{ MOD } 4$$

Vis hvordan strukturen ser ut til slutt når du har satt inn alle nøklene.



Oppgave 6 – Lagring, indeksering og queries (10 %)

Vi har en database som lagrer webclicks i en tabell:

Click (clickId, user, url, time)

Hver post (record) i tabellen er 100 byte lang og hver blokk er 8 KB (8192 bytes). Vi har registrert 300 000 klikk i databasen.

- Vi ønsker å lagre tabellen i et clustered B+-tree med clickId (8 byte) som søkenøkkel.
Hvor mange blokker er det på hvert nivå i B+-treet? Beskriv eventuelle antagelser du tar.
- Vi ønsker å utføre følgende query:

```
SELECT url, count(*) AS clickcount  
FROM Click  
GROUP BY url  
ORDER BY clickcount DESC;
```

Hvordan vil du indeksere tabellen for å kunne svare på queriet?
Begrunn svaret ditt.

Oppgave 7 – Transaksjoner (15 %)

- a) Forklar de fire begrepene som forkortes ACID.
- b) Vi har en historie:
H₁: r₁(A); w₂(A); w₂(B); w₃(B); w₁(B); C₁; C₂; C₃;
Vi innfører tofaselåsing (rigorous 2PL). Skriv om historien slik at den bruker 2PL. Innfør operasjonene wl(X) – write_lock(X) og rl(X) – read_lock(X).
- c) For hver klasse av samtidighetsproblemer gitt under, hvilke ser du i historien H₁ i b) og hvorfor / hvorfor ikke?
 - Dirty read
 - Lost update
 - Unrepeatable read
 - Incorrect summary

Oppgave 8 – Transaksjoner: Recovery – ARIES (10 %)

- a) Beskriv og forklar hva som skjer i de tre fasene av recovery etter krasj i ARIES.
- b) Hva er PageLSN og hvorfor er dette en viktig oppfinnelse?