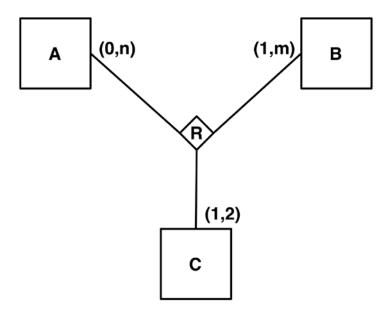


Institutt for datateknikk og informasjonsvitenskap

Eksamensoppgave i TDT4145 Datamodellering og databa	asesvste	emer
	<b>,</b>	
Faglig kontakt under eksamen:		
Svein Erik Bratsberg: 995 39 963		
Roger Midtstraum: 995 72 420		
Eksamensdato: 26. mai 2014		
Eksamenstid (fra-til): 09:00 - 13:00		
Hjelpemiddelkode/Tillatte hjelpemidler:		
D – Ingen trykte eller håndskrevne hjelpemidler tillatt. Bes	stemt, enkel k	alkulator tillatt.
Annen informasjon:		
Målform/språk: Norsk bokmål		
Antall sider: 6		
Antall sider vedlegg: 0		
		Kontrollert av:
	Svein-0	Olaf Hvasshovd (sign.)
	Dato	Sign.

## Oppgave 1 - Datamodeller (20 %)

a) (5 %) Restriksjoner på tertiære relasjonsklasser kan for eksempel være som i ERdiagrammet under. Forklar hvilke praktiske restriksjoner dette gir for entitetene i entitetsklassene som deltar i relasjonsklassen R.



b) (15 %) Lag en ER-modell (du kan bruke alle virkemidler som er med i pensum) for følgende situasjon:

Det skal lages en database for å registrere fugleobservasjoner. Fugleartene er delt inn i et antall fuglegrupper, som trostefugler eller skarver, og alle fuglearter er med i en bestemt fuglegruppe. Fuglegrupper har et entydig gruppenummer, et norsk navn og et latinsk navn. Fuglearter, som Rødstrupe eller Storskarv, har et norsk navn, et latinsk navn, en tekstlig beskrivelse, en forekomstkode (standfugl, trekkfugl osv.) og en antallsangivelse (1-7) som forteller hvor vanlig fuglen er i Norge. Observasjoner av en fugleart blir registrert med dato og tid, en kode for kjønn som er observert (hunn, hann eller begge), et anslag på antall individer og observatørens sikkerhet i forhold til observasjonen (usikker, sannsynlig, sikker). Det registreres alltid et bestemt observasjonssted som ligger i en kommune, det kan for eksempel være en bydel eller en kjent fuglelokalitet. Databaseeier vedlikeholder en oversikt over observasjonssteder, og observatører må velge blant de forhåndsregistrerte stedene. Alle observasjoner knyttes til en person som må være registrert med navn, adresse, telefonnummer og e-postadresse. Personer som er registrert i databasen, kan være medlem i en eller flere fugletitter-foreninger. Fugletitterforeningene klassifiserer sine medlemmer i kategoriene nybegynner, erfaren og ekspert. En person kan ha ulik status i forskjellige foreninger. Enkelte foreninger har også en ordning der medlemmer får status som artsekspert for en eller flere fuglearter. Klassifiseringen av observatører er viktig for å kunne vurdere kvaliteten og troverdigheten til de observasjonene som er registrert i databasen.

Gjør kort rede for eventuelle forutsetninger som du finner det nødvendig å gjøre.

### Oppgave 2 – Relasjonsalgebra og SQL (20 %)

Ta utgangspunkt i følgende relasjonsdatabase (primærnøkler er understreket) for en enkel mobiltelefon-database:

**Subscriber**(<u>ID</u>, Name, Address, PostalCode, City)

**PhoneNumber**(PNumber, SubscriberID)

SubscriberID er fremmednøkkel mot Subscriber-tabellen. SubscriberID kan ikke ha NULLverdi.

PhoneCall(CallID, FromNo, ToNo, CYear, CMonth, CDate, StartTime, Duration)

FromNo og ToNo er fremmednøkkel mot PhoneNumber-tabellen. FromNo og ToNo kan ikke ha 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) (4 %) 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) (3 %) Lag en spørring i *relasjonsalgebra* som finner ID, Name og Address for abonnenter som har ringt nummeret '11223344' i 2014 (samtalen startet i 2014).
- c) (3 %) Lag en spørring i *SQL* som finner alle telefonnumre (PNumber) som er registrert på abonnenter (eng. subscribers) som bor i gaten Buckhaugen i Trondheim ('Buckhaugen 1', 'Buckhaugen 2', osv.).
- d) (3%) Lag en spørring i *SQL* som finner antall abonnenter i hver by som er registrert med en eller flere abonnenter. Resultatet (city, antall abonnenter) skal sorteres etter synkende antall abonnenter.
- e) (3 %) Abonnenten med ID 100 flytter til Buckhaugen 99 som har postnummer (eng: PostalCode) 7046 og ligger i Trondheim. Skriv en *SQL-setning* som registrerer denne endringen i databasen.
- f) (4 %) Lag en *SQL-spørring* som finner PNumber og abonnentens ID, Name og Address for telefonnummer som ikke har hatt noen utgående samtaler i 2014 (samtalen startet i 2014). Resultatet skal sorteres på abonnentens ID i stigende rekkefølge. Dersom samme abonnent har flere ubrukte telefonnummer, skal disse sorteres i stigende rekkefølge.

# Oppgave 3 - Teori (20 %)

- a) (3 %) Hvilke krav må være oppfylt for at en mengde attributter skal være en *kandidatnøkkel* (eng. candidate key) for en tabell?
- b) (4 %) Gitt R =  $\{A, B, C, D\}$  og F =  $\{A \rightarrow B, C \rightarrow D, D \rightarrow C\}$ . Finn alle kandidatnøkler i R.
- c) (3 %) Gitt R = {A, B, C, D} og F = {A -> B, C -> D, D-> C}. Gå ut fra at R oppfyller 1. normalform. Bestem den høyeste normalformen som oppfylles av R. Svaret må begrunnes.
- d) (4 %) Gitt R = {A, B, C, D} og F = {A -> B, C -> D, D-> C}. R skal om nødvendig dekomponeres slik at alle komponenter er på Boyce-Codd normalform (BCNF). Vurder om dekomponeringen du kommer fram til er en god løsning.

А	В	С	
a <sub>1</sub>	b <sub>1</sub>	c <sub>1</sub>	
a <sub>1</sub>	<sup>b</sup> 2	c <sub>2</sub>	
<sup>a</sup> 2	b <sub>1</sub>	c <sub>2</sub>	
a <sub>1</sub>	b <sub>1</sub>	c <sub>2</sub>	
×	Y	z	

- e) (3 %) Over er det vist en tabellforekomst for en tabell med attributtene A, B og C. Gå ut fra at MVD-en A ->> B gjelder, hva må da være verdiene til X, Y og Z?
- f) (3 %) Over er det vist en tabellforekomst for en tabell med attributtene A, B og C. Gi et eksempel på verdier for X, Y og Z som gjør at MVD-en A ->> C *ikke* kan gjelde for tabellen.

#### **Oppgave 4 - B+-trær (10 %)**

Vi skal sette inn følgende nøkler i et B+-tre i den gitte rekkefølgen: 1, 2, 3, 4, 5, 6, 7, 8, 9 og 10. Det er plass til maksimalt tre nøkler i hver blokk. Vis tilstanden til B+-treet hver gang du skal til å splitte en blokk. Vis også hvordan B+-treet ser ut til slutt.

# Oppgave 5 - Lagring og queryutføring (15 %)

Anta følgende tabell:

CREATE TABLE Student (studno INT, lastname CHAR(30), firstname CHAR(30), email CHAR(30), startyear INT);

Anta tabellen er lagret i en heapfil med 1000 blokker hvor hver blokk får plass til 20 studentposter. Videre er det laget en unclustered B+-tre-indeks på 'lastname'. Vi antar B+-treet har 500 blokker på løvnivå og har høyde 3.

- a) (10 %) Gjør et estimat på hvor mange blokker som aksesseres (leses) ved de følgende SQL-setningene:
  - i) INSERT INTO Student VALUES (12123, 'Hansen', 'Hans', 'hans@email.org', 2013);
  - ii) SELECT lastname, firstname, email, startyear FROM Student WHERE lastname='Hansen';
  - iii) SELECT \* FROM Student:
  - iv) SELECT DISTINCT lastname FROM Student ORDER BY lastname;

Gi en begrunnelse/forklaring for hvert svar.

b) (5 %) Hvis den samme tabellen hadde fått kravet om at 'studno' skal være PRIMARY KEY, hvordan ville du ha lagret tabellen? Begrunn svaret.

# Oppgave 6 - Transaksjoner - vranglåser (5 %)

Når vi bruker låser av dataelementer kan transaksjoner oppleve vranglåser. Hvordan løses vranglåser mellom transaksjoner?

NB! Oppgavesettet fortsetter på neste side.

# Oppgave 7 - Transaksjoner - recovery (10 %)

Anta at vi bruker ARIES-recovery og la A,B,C og D være dataelementer med angitte startverdier (30, 15, 40, 20). Etter en krasj ble følgende logg funnet med loggposter på formatet:

[LSN,Operation,Transaction,DataItem,BeforeImage,AfterImage]

	A	В	С	D
	30	15	40	20
[101,start_trans,T3]				
[102,write,T3,A,30,33]	33			
[103,write,T3,B,15,12]		12		
[104,start_trans,T2]				
[105,commit,T3]				
[106,write,T2,B,12,18]		18		
[107,start_trans,T1]				
[108,start_checkpoint]				
[109,end_checkpoint]				
[110,write,T1,C,40,45]			45	
[111,write,T2,D,20,26]				26
[112,commit,T2]				

- a) Anta T1, T2 og T3 er alle transaksjonene som finnes. Hvordan ser transaksjonstabellen ut etter analysefasen av recovery?
- b) Hva er verdiene til A, B, C og D etter at recovery (alle faser) er ferdig?