

**Norges teknisk-naturvitenskapelige universitet
Institutt for datateknikk og informasjonsvitenskap**



**EKSAMENSOPPGAVE I FAG TDT4145 – DATAMODELLERING OG
DATABASESYSTEMER**

Faglig kontakt under eksamen: Svein Erik Bratsberg og Roger Midtstraum

Tlf.: 99539963 (Bratsberg) og 99572420 (Midtstraum)

Eksamensdato: 1. juni 2011

Eksamenstid: 09.00-13.00

Tillatte hjelpemiddel: D: Ingen trykte eller håndskrevne hjelpemiddel tillatt. Bestemt, enkel kalkulator tillatt.

Språkform: Bokmål

Sensurdato: 24. juni 2011

Oppgave 1 – Datamodellering – 25 %

Kinofabrikken driver et antall kinosentre som består av et antall saler der det vises filmer. Kundene kjøper billetter for å se en bestemt film, i en bestemt kinosal, til et bestemt tidspunkt. En billett reserverer et bestemt sete i kinosalen for den aktuelle forestillingen. Kinofabrikken har et kunderegister med kontaktinformasjon om kundene (navn, adresse, telefonnummer og e-postadresse) og en oversikt over hvilke filmsjangre (drama, komedie, western, barnefilm, etc.) hver kunde er mest interessert i. Kinofabrikken har informasjon om de filmene som blir vist – tittel, nasjonalitet, spilletid, produksjonsår, distribusjonsselskap, filmsensur og filmsjanger. En film er av en bestemt filmsjanger, og Kinofabrikken vedlikeholder en oversikt over filmsjangre med sjangernavn og sjangerbeskrivelser. Filmsensur er en aldersgrense for filmen, fastsatt av Medietilsynet. De aldersgrensene som gjelder er "7 år", "11 år", "15 år" og "18 år". Distribusjonsselskap er selskap som eier rettigheter for filmer på det norske markedet. Kinosentrene har navn, adresse og telefonnummer. Kinosalene har navn, et antall seter og er klassifisert i forhold til tilgjengelighet for rullestolbrukere. Kinofabrikken holder oversikt over alle kundeordrer med informasjon om kjøpstidspunkt, antall billetter av hver billettype ("barn", "student", "ordinær") og totalpris. En kundeordre gjelder for en bestemt forestilling. Prisene for de ulike billetttypene varierer og fastsettes for hver forestilling.

Du skal lage en ER-modell (du kan bruke alle elementer som er undervist) for en database som holder oversikt over Kinofabrikkens virksomhet, som beskrevet over. Databasen skal blant annet kunne brukes for å finne svar på spørsmål som:

- Hvor mange billetter er solgt til forestillinger med filmen *Sykkeltyven*?
- Hvilke seter er ledig for en bestemt forestilling?
- Hvilke distribusjonsselskaper har western-filmer?
- Hva var gjennomsnittlig omsetning pr. kunde i januar 2011?
- Hva er total setekapasitet pr. kinosenter?
- Hva var de 10 mest populære filmene i 1. kvartal 2011 målt i antall solgte billetter?
- Hvor mange billetter har kunden med kundenummer 12345 kjøpt i 2011?
- Hvordan er fordelingen av billettsalg (antall solgte billetter) for setene i salen *Woody Allen* i Spektrum kinosenter i 2011?

Når det gjelder egenskaper som naturlig modelleres som attributter, skal du ta med de attributtene som fremgår av beskrivelsen over og de attributtene som du anser som viktigst å ha med. Det er ikke nødvendig å ta med alle attributter som ville fremkommet gjennom en virkelig modelleringsprosess.

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

Oppgave 2 – Relasjonsalgebra og SQL – 20 %

Ta utgangspunkt i følgende relasjonsdatabase (primærnøkler er understreket):

Kunde(Kundenr, Navn, Adresse, Postnr, Poststed, Kredittgrense)

Produkt(Varenr, Betegnelse, Pris)

Ordre(Ordrenr, Dato, Status, Totalsum, Kundenr)

Ordrelinje(Ordrenr, Linjenr, Varenr, Antall, Delsum)

- Lag et *ER-skjema* som i størst mulig grad samsvarer med relasjonsskjemaet. Gjør rede for eventuelle antagelser som du finner det nødvendig å gjøre.
- Lag en spørring i *relasjonsalgebra* som finner alle kunder (Kundenr, Navn og Poststed) som har bestilt et produkt med betegnelsen "iPad 2".
- Lag en spørring i *relasjonsalgebra* som finner kunder (Kundenr og Navn) som ikke har noen ordrer.
- Lag en *SQL-spørring* som finner kundenummer, navn og adresse for kunder som ikke har noen ordrer på 1000 kroner eller mer.
- Lag en *SQL-spørring* som finner samlet omsetning (i kroner) for hvert av produktene 'grill', 'grillkull' og 'tennvæske'. Resultatet skal være sortert alfabetisk etter betegnelse.

Oppgave 3 – Normalisering – 15 %

- Forklar hva det betyr at en mengde attributter (Y) er funksjonelt avhengig av en annen mengde attributter (X), $X \rightarrow Y$.

Ta utgangspunkt i tabellen Bil(regNr, merke, modell, forsteAarReg, vekt). En mulig forekomst av tabellen er vist under:

Bil	regNr	merke	modell	forsteAarReg	vekt
	VH13589	Honda	CRV	2009	1707
	VT16092	Toyota	Yaris	2007	1055

- Anta som en hypotese at $vekt \rightarrow merke$. Hvilke restriksjoner ville dette forutsette i miniverdenen?

Ta utgangspunkt i at følgende funksjonelle avhengigheter gjelder i tabellen Bil:

$F = \{ regNr \rightarrow modell, forsteAarReg; modell \rightarrow merke, vekt \}$.

- Tabellen Bil er minimum på 1. normalform, hva er den høyeste normalformen som Bil tilfredsstillter? Svaret må begrunnes.
- Tabellen $R(a\ b\ c)$ der $F = \{ a \rightarrow b \}$ har tre mulige dekomponeringer:
 - $R_1(a\ b)$ og $R_2(a\ c)$
 - $R_1(a\ b)$ og $R_2(b\ c)$
 - $R_1(a\ c)$ og $R_2(b\ c)$

Hvilke av disse dekomponeringene har tapsløst-join-egenskapen og hvilke har bevaring av funksjonelle avhengigheter? Svaret må begrunnes.

Oppgave 4 – Lagring, indekser og queryutføring – 20 %

- a) Tegn en (eller flere) figur(er) som forklarer sammenhengen mellom de følgende begrepene:
- SQL-dictionary
 - B-trær
 - Datadevicer
 - Blokker
 - Poster
- b) Gi et eksempel på et SQL-query som har stor hjelp av en indeks. Forklar hvorfor denne indeksen hjelper i utførelsen av queriet.
- c) Anta vi har extendible hashing. I utgangspunktet er det fire blokker i hashindeksen: blokkene 0-3. Hver hashblokk kan inneholde inntil 4 poster. Vis den resulterende hashstrukturen, dvs. hvilke blokker de forskjellige verdiene havner i når vi setter inn de følgende *hashede* verdiene: 4, 6, 0, 8, 3, 7, 11, 12, 16, 13.
- d) Gi en kort oversikt over forskjellige måter å utføre join på.

Oppgave 5 – Transaksjoner – 20 %

- a) Gitt de følgende historiene
- (1) $r_1(X), r_2(Y), w_3(X), r_2(X), r_1(Y), c_1, c_3, c_2$
 - (2) $r_1(X), r_1(Y), w_1(X), r_2(Y), w_3(Y), w_1(X), r_2(Y), c_1, c_2, c_3$
 - (3) $r_2(X), w_3(X), c_3, w_1(Y), c_1, r_2(Y), w_2(Z), c_2$
 - (4) $r_1(X), w_2(X), c_2, w_1(X), c_1, r_3(X), c_3$

Tegn presedensgrafene for historiene og avgjør hvilke som er konfliktserialiserbare.

- b) Avgjør for hver av historiene over hvilken klasse de tilhører: ikke gjenopprettbar, gjenopprettbar, ACA og strict.
- c) Forklar fordelene med no-force/steal-strategien for logging/recovery.
- d) Forklar formålet med en CLR (Compensating Log Record) og hvilken annen nytte vi har av den?