

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



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

Faglig kontakt under eksamen: Svein Erik Bratsberg

Tlf.: 99539963 (Bratsberg)

Eksamensdato: 6. august 2012

Eksamenstid: 09.00-13.00

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

Språkform: Bokmål

Sensurdato: 27. august 2012

Oppgave 1 – Datamodellering – 20 %

PizzaSkuret lager skreddersydde pizzaer. Kundene kan velge pizzabunn blant et antall ulike deiger som leveres i et antall størrelser, uttrykt i diameteren på den ferdige pizzaen. Når deigtype og størrelse er valgt, kan kunden velge en pizzasaus (antall ml) og et antall innholdselementer (for eksempel skinke). For hvert innholdselement kan kunden bestemme mengden (gram). For å toppe pizzaen kan kunden velge et antall oster og mengde (gram) av hver ost. Når kunden har bestemt steketemperatur (grader) og steketid (minutter), er pizzaen ferdig konfigurert og klar til produksjon. De ulike ingrediensene i en pizza er prissatt slik at det beregnes en pris for hver enkelt pizza, med utgangspunkt i prisene for ingrediensene.

Pizzaskuret tar bare nettbestillinger og holder oversikt over alle kunder med kundenummer, navn og mobiltelefonnummer. Når en kunde bestiller en eller flere pizzaer blir det laget en ordre. Pizzaskuret holder rede på tiden en ordre opprettes og når kunden får levert de pizzaene som inngår i ordren. Ved utkjøring av pizzabestillinger registrerer man leveringsadresse og status for leveringen ("levert", "fant ikke fram", "kunne ikke betale" eller "nektet mottatt"), samt utkjøringsbil og sjåfør.

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

- Hva er leveringsstatus for alle bestillinger fra et bestemt kundenummer?
- Hva er total omsetning per kunde?
- Finn gjennomsnittstiden fra bestilling til leveranse.
- Finn andelen utkjøringer med status "fant ikke fram" per sjåfør.
- Hvilken pizzasaus er mest populær?
- Hva var gjennomsnittsprisen for pizzaer i hver måned 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 eventuelle forutsetninger som du finner det nødvendig å gjøre.

Oppgave 2 – ER, relasjonsalgebra og SQL – 20 %

Ta utgangspunkt i følgende relasjonsdatabase (primærnøkler er understreket) for skøyteLøp:

SkøyteLøp(LID, Klasse, Distanse, Dato, Tid)

SkøyteLøper(SID, Navn, FødselsÅr, Nasjonalitet, Klasse)

Passeringer(LID, SID, Lengde, PasseringsTid, RundeTid) – LID er fremmednøkkel mot SkøyteLøp, SID er fremmednøkkel mot SkøyteLøper.

Resultat(LID, SID, SluttTid, Plassering) – LID er fremmednøkkel mot SkøyteLøp, SID er fremmednøkkel mot SkøyteLøper.

Attributtet *Distanse* forteller hvor langt et skøyteLøp er, for eksempel 500 m, 1500 m eller 5000 m. I passeringer forteller *Lengde* hvor mange meter som er tilbakelagt. Skøytebaner er rundbaner på 400 m; på en 1500 m vil man derfor ha passeringspunkt etter 300 m, 700 m, 1100 m og 1500 m. Rundetid er

tid siden start for den første passeringen og tid siden forrige passering for de etterfølgende passeringene. En skøyteløper må fullføre et skøyteløp for å bli registrert i Resultat-tabellen for dette skøyteløpet.

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.

- Lag et *ER-diagram* som i størst mulig grad samsvarer med relasjonsskjemaet. Gjør rede for eventuelle antagelser du finner det nødvendig å gjøre.
- Lag en spørring i *relasjonsalgebra* som lager en resultatliste (Plassering, SID, Navn, Nasjonalitet, SluttTid) for løpet med LID lik 100.
- Lag en spørring i *relasjonsalgebra* som finner alle skøyteløpere (SID, Navn, Nasjonalitet) som er født i 1992 eller senere og som har fullført minst en 5000 m på en tid på 6,40 eller bedre.
- Lag en spørring i *SQL* som lager en resultatliste (Plassering, SID, Navn, Nasjonalitet, SluttTid) for løpet med LID lik 100. Resultatet skal være ordnet etter plassering, i stigende rekkefølge. Dersom flere løpere har samme plassering, skal disse innbyrdes sorteres alfabetisk etter navn.
- Lag en *SQL-spørring* for løpet med LID lik 100, som finner antall deltakere per nasjon som hadde deltakere som fullførte løpet. Resultatet skal sorteres etter antall deltakere i synkende rekkefølge.

Oppgave 3 – Teori – 20 %

- Ta utgangspunkt i tabellen $R(A\ B\ C\ D\ E)$ der vi har følgende tabellforekomst $\{(a_1, b_1, c_1, d_1, e_1)\}$. Legg til ett tuppel (rad) som gjør at den funksjonelle avhengigheten (engelsk: functional dependency) $BC \rightarrow D$ ikke kan gjelde for tabellen.
- Gitt to mengder funksjonelle avhengigheter $F = \{a \rightarrow b, b \rightarrow c, c \rightarrow a\}$ og $G = \{a \rightarrow bc, c \rightarrow a\}$. Hva betyr det at to mengder funksjonelle avhengigheter er ekvivalente? Er F og G ekvivalente mengder funksjonelle avhengigheter?
- Gitt $R(A\ B\ C\ D)$ og $F = \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$. Forklar begrepet kandidatnøkkel. Finn alle kandidatnøkler i R .
- Anta at R fra deloppgave c oppfyller første normalform (1NF). Hva er den høyeste normalformen som oppfylles av R ? Svaret må begrunnes.

Oppgave 4 – Hashing – 15 %

- a) Hva er fordelene og ulempene med extendible hashing sammenlignet med statisk hashing?
- b) Hva er fordelene og ulempene med lineær hashing sammenlignet med extendible hashing?
- c) Anta en extendible hashfil, hvor vi starter med 4 blokker i fila. Hver blokk har plass til tre nøkler. Sett inn følgende nøkler i den gitte rekkefølgen: 121, 204, 731, 547, 800, 221, 932, 145, 112, 721, 412. Bruk hashfunksjonen $H(K) = K \text{ MOD } 8$, men bruk kun det antall bits som er nødvendig til enhver tid. Vis tilstanden til hashfila etter at alle nøklene er satt inn.

Oppgave 5 – Transaksjonsteori – 10 %

- a) Lag en historie (schedule) som er strict og involverer to transaksjoner.
- b) Lag en historie (schedule) som er ACA (avoid cascading aborts), men ikke strict. Historien skal involvere to transaksjoner.
- c) Lag en historie (schedule) som er gjenopprettbar, men ikke ACA. Historien skal involvere to transaksjoner.
- d) Lag en historie (schedule) som ikke er gjenopprettbar og som involverer to transaksjoner.

Oppgave 6 – Låsing – 10 %

- a) Hvilke varianter av tofaselåsing finnes? Forklart kort fordeler og ulemper med de forskjellige variantene.
- b) Forklar hvordan vranglås kan oppstå med tofaselåsing.

Oppgave 7 – Logging og recovery – 5 %

Forklar hvorfor ARIES har sjekkpunkting og hva innholdet i sjekkpunktloggpostene er.