

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i : IN 212 - Databaseteori
Eksamensdag : Onsdag 8. juni 1994
Tid for eksamen : 09.00 - 15.00
Oppgavesettet er på : 5 sider
Vedlegg : Ingen
Tillatte hjelpemidler : Alle trykte og skrevne

Kontroller at oppgavesettet er komplett før du begynner å besvare det.

Oppgave 1 (35 %)

Nedenfor følger et begrepsmessig skjema for en liten relasjonsdatabase som brukes til å registrere utlån fra biblioteket i en bedrift:

AVDELING (AvdKode, AvdNavn, Adresse)
LÅNER (Låner-nr, Navn, AvdKode, Telefon)
BOK (ISBN, Eksemplar, Forfatter, Tittel, Utgivelsesår, Kjøpsår)
UTLÅN (ISBN, Eksemplar, Lånedato, Låner-nr, Returdato)

AvdKode er primærnøkkel i AVDELING
Låner-nr er primærnøkkel i LÅNER
Telefon er kandidatnøkkel i LÅNER
(ISBN, Eksemplar) er primærnøkkel i BOK
(ISBN, Eksemplar, Lånedato) er primærnøkkel i UTLÅN
Nullverdier er tillatt for Returdato i UTLÅN
AvdKode i LÅNER er fremmednøkkel til AVDELING
(ISBN, Eksemplar) i UTLÅN er fremmednøkkel til BOK
Låner-nr i UTLÅN er fremmednøkkel til LÅNER

I tillegg har vi to FD-er (funksjonelle avhengigheter):

ISBN → (Forfatter, Tittel, Utgivelsesår) i BOK
(ISBN, Låner-nr, Lånedato) → Eksemplar i UTLÅN

Databasen inneholder ingen JD-er (join-avhengigheter) ut over de som følger av det ovenstående.

Fortsettes neste side

Databasesystemet inneholder en Boolsk systemrutine ER-NULL som tester om et gitt datafelt (attributt) har null-verdien.

Databasesystemet håndterer datoer slik at differansen mellom to datoer gir antall dager mellom de to datoene.

Eksempler: 18.6.1994 - 8.6.1994 = 10 og 1.1.1994 - 31.12.1993 = 1

Dessuten finnes det en systemvariabel IDAG som gir dagens dato.

Oppgave 1 A Hvilken normalform har hver av tabellene i skjemaet? Begrunn svaret og beskriv hvilke endringer som eventuelt må gjøres for å få hele skjemaet på 5NF.

Oppgave 1 B Lag en ER-modell av den datastrukturen du kom frem til i oppgave 1 A. Gjør kort rede for eventuelle svake entitetstyper og hvilke nøkler de har. Oversett ER-modellen til et nettverksdatabaseskjema. (Du trenger ikke skrive opp posttypene, det er én for hver entitetstype. Det holder med beskrivelse av de ikke-singulære settypene.)

I resten av oppgaven skal du ta utgangspunkt i den opprinnelige datastrukturen og ikke ta hensyn til hva du måtte ha kommet frem til i oppgave 1 A.

Oppgave 1 C Bruk SQL til å lage en liste over alle utlån som er eldre enn fire uker uten at boken er returnert. Listen skal inneholde navn og adresse til lånere, og forfatter og tittel på bøker. Listen skal være sortert på avdelingsadresse og så på låners navn innenfor hver avdeling.

Oppgave 1 D Forklar hvordan optimalisereren vil behandle spørsmålet ovenfor, både hva slags relasjonsalgebrauttrykk den vil omforme SQL-uttrykket til, og hvordan den kommer frem til det. (Se bort fra sorteringen.)

Oppgave 1 E Skriv et relasjonsalgebrauttrykk, et tuppelkalkyleuttrykk og en SELECT-setning som hver for seg finner forfatter og tittel på de bøker som en eller annen gang har vært utlånt til minst en ansatt (låner) fra hver avdeling.

Oppgave 1 F Bruk SQL til å lage en liste over hyppig utlånte bøker i 1993. Listen skal inneholde forfatter, tittel og antall eksemplarer av hver bok og antall ganger boken har vært utlånt. Ta bare med bøker som er utlånt minst tre ganger. Sorter listen etter fallende popularitet.

Oppgave 2 (20 %)

Du skal nå bevise en slutningsregel for MVD-er:

La R være et relasjonsskjema og la X , Y og Z være attributter i R .
(Merk at vi ikke har forutsatt at X , Y og Z er disjunkte, og at de heller ikke trenger å være noen dekomposisjon av R .)

Bevis følgende påstand:

$$(X \twoheadrightarrow Y \wedge X \twoheadrightarrow Z) \Rightarrow (X \twoheadrightarrow Y \cap Z \wedge X \twoheadrightarrow Y \setminus Z \wedge X \twoheadrightarrow Z \setminus Y)$$

Oppgave 3 (15 %)

Vi skal i denne oppgaven se litt på en serialiseringsalgoritme for klassiske hierarkiske databaser (klassisk betyr at vi ikke tillater VPCR-er).

Vi kan betrakte en hierarkisk databasetilstand som ett enkelt tre:

Tilstanden for hvert hierarki omdannes til et tre ved at barna til en posttype i hierarkiet er alle forekomster av posttypen og alle barna (som er null eller flere posttyper) til posttypen, og ved at barna til en forekomst er barna i den PCR-forekomsten som denne forekomsten er forelder i. Alle hierarkiene fornes til ett tre ved å gjøre hele databasen til felles rot. (Bladnodene i dette tilstandstreet blir forekomstene av bladnodene i hierarkiene.)

Den algoritmen vi skal se på, kalles trelåseprotokollen, og den baserer seg på binære låser på nodene i tilstandstreet beskrevet ovenfor. Protokollen har to regler:

- Med unntak av den første låsen en transaksjon setter, kan ikke en transaksjon låse en node med mindre den har låst foreldrenoden
- Ingen noder kan låses to ganger av samme transaksjon

Merk at en transaksjon kan låse hvilken som helst node første gang den setter en lås, og at det å låse en node ikke låser noder i subtreet (de kan til og med være låst av andre transaksjoner). Hensikten med protokollen er at transaksjonene kan frigi låser etter hvert som de ikke trenger dem mer.

Det er bevist at trelåseprotokollen sikrer serialiserbarhet og at den forhindrer vranglås (deadlock).

Oppgaven fortsetter på neste side

Oppgave 3 A Lag et eksempel på to transaksjoner som kan gi vranglås ved to-faselåsing, og vis at disse to transaksjonene ikke kan gi vranglås ved trelåseprotokollen.

Oppgave 3 B Gi en kort forklaring av begrepet „vrangløp” (live-lock) og en kort kommentar om vrangløp og trelåseprotokollen.

Oppgave 4 (30 %)

Merk: Flere av observasjonene i oppgave 4 A, B og C er nokså trivielle. Kast ikke bort tiden på å lete etter vanskeligheter som ikke er der!

Vi skal i denne oppgaven utforske hvor kompliserte små relasjonsskjemaer kan bli. Vi vil i den forbindelse ha nytte av følgende definisjoner hvor de fem første burde være gamle kjenninger, mens den siste er ny:

Def 1 Et atomært attributt i et relasjonsskjema er et attributt som består av nøyaktig ett attributtatom.

Def 2 Et relasjonsskjema som består av ett atomært attributt, kalles unært.

Def 3 Et relasjonsskjema med eksakt to attributtatomer kalles binært. Vi skal bruke notasjonen $R = R(A,B)$ for binære relasjonsskjemaer der A og B betegner de to (disjunkte) atomære attributtene i R.

Def 4 Et relasjonsskjema med tre attributtatomer kalles ternært. Vi skal bruke notasjonen $R = R(A,B,C)$ for ternære relasjonsskjemaer der A, B og C betegner de tre (disjunkte) atomære attributtene i R.

Def 5 En MVD, $X \twoheadrightarrow Y$, i R kalles triviell hvis $Y \subseteq X$, eller hvis MVD-en ikke er en FD og $(R \setminus X) \setminus Y = \emptyset$ (det tomme attributtet).
(Merk: Denne definisjonen medfører at $X \rightarrow \emptyset$ er triviell.)

Def 6 En MVD, $X \twoheadrightarrow Y$, kalles patologisk hvis $X = \emptyset$.
(Begrepet „patologisk MVD” er ikke noe vanlig faguttrykk. Det er laget spesielt for denne eksamensoppgaven.)

La oss som eksempel se på hva det vil si å ha en patologisk FD, $\emptyset \rightarrow Z$:

Siden alle forekomster er like på det tomme attributtet, må alle forekomster ha samme verdi i Z. Z blir altså et konstant attributt.

Oppgaven fortsetter på neste side

Vi skal først betrakte det binære relasjonsskjemaet $R = R(A,B)$.

Oppgave 4 A Anta at vi har en patologisk MVD, $\emptyset \twoheadrightarrow A$, i R . Begrunn at R ikke er på 4NF og finn en tapsfri dekomposisjon av R til 4NF. Forklar kort hvorfor denne dekomposisjonen også er på 5NF. Lag et eksempel med tre A -verdier og fire B -verdier og vis hva som skjer.

Oppgave 4 B Vis at alle ikke-trivielle patologiske MVD-er i R enten er identisk med MVD-en i oppgave 4 A eller et spesialtilfelle av denne. For hvert av spesialtilfellene, som nødvendigvis må være FD-er (begrunn det kort), skal du angi konsekvensen for 5NF-dekomposisjonen i oppgave 4 A. Bruk gjerne samme eksempel. (Her trenger du ikke begrunne svaret.)

Resultatene og ditt eksempel ovenfor forklarer forhåpentlig hvorfor databaselitteraturen totalt ignorerer patologiske MVD-er. Det skal vi også gjøre heretter.

I resten av denne oppgaven skal du altså se bort fra patologiske MVD-er.

Oppgave 4 C Vis at binære relasjonsskjemaer er på 5NF (igjen forutsatt at de ikke har ikke-trivielle patologiske MVD-er).

Som kjent, kan vi alltid dekomponere et relasjonsskjema tapsfritt og FD-bevarende til 3NF, men ikke alltid til BCNF. (Det finnes, som du sikkert vet, et ternært moteksempel.)

Vi skal nå se på FD-bevaring ved dekomposisjon til høyere normalformer.

Oppgave 4 D Anta at R er et ternært relasjonsskjema som er på BCNF, men ikke på 4NF. Vis at da kan R dekomponeres tapsfritt og FD-bevarende til 5NF.

Oppgave 4 E Finn et eksempel på et relasjonsskjema som er på BCNF, men som ikke kan dekomponeres tapsfritt og FD-bevarende til 4NF.

Slutt på oppgavesettet

Ragnar Normann