

Relasjonsmodellen - Codd 1970

- Målet bak modellen var å redusere dataavhengighet.

Datauavhengighet: Slipper å endre programmer som bruker f.eks en database, når denne databasen blir endret, så skal ikke programmene som bruker databasen behøves å endres. Programmet skal ikke ta hensyn til hvordan databaser ser ut.

Oracle og Microsoft SQL er store relasjonsdatabaser. Vi skal bruke sqlite.

Begreper

Domener

Relasjonen $R = \{x_1...x_n | (x_1 \in S_1...x_n \in S_n)\}$

Hvor $x_1...x_n$ er n-tuppler og $S_1...S_n$ er domenene til tupplene.

I database relasjon så er ikke rekkefølgen på tupplene viktig, men i matematikken er dette nødvendig.

R er ett sett med mange n-tuppler.

$R = \{T_1...T_n\}$ hvor $T_i = \{A_n : d_n\}$

Der A_n er antall attributter og d_n er antall domener. Hvert attributt tilhører en domene, og hver tuppel inneholder attributtverdier og domener. Relasjonen er altså alle tupplene, altså hele tabellen.

Domene

- Mengden av lovlige verdier til attributtene

Attributter

- Navngir tupplenes elementer

Relasjonen R består av T_i tuppler = $\{T_1..T_n\}$

Tupplene inneholder $\{n - domener \text{ og } n - attributter\}$

En database inneholder alle relasjonene / tabellene.

$D = \{R_1, R_2...R_n\}$

En relasjon inneholder $n - tuppler$ og er en tabell av tuppler.

$R = \{T_1...T_n\}$

En tupplen inneholder n-domener og n-attributter

$T_i = \{A_n : d_n\}$

6 egenskaper til relasjonsmodellen

Navngivning

1. Unike relasjonsnavn i hver database
2. Unike attributtnavn i hver relasjon

Verdier i attributter / verdi

3. Alle verdier hentes ifra domenet, hvis de ikke eksisterer i domenet, kan det ikke lages ett attributt med dette navnet.
4. Atomiske og en enkel verdi (kan ikke består av flere deler). For eksempel fornavn og etternavn burde være i hvert sitt attributt, og ikke i samme attributt.

Rekkefølge

5. Rekkefølge i ar ingen betydning for tuppler og attributter

6. Ingen duplikate tuppler er lov. Ingen duplikate mengder, altså $\{1, 2, 2\}$ er ikke godkjent.

Tuppler er rader, attributter er kolonner og hele tabellen er relasjonen.

Grad

Relasjonen har en bestemt grad, graden er definert for antall attributter. n -grad = antall attributter i en relasjon

Kardinalitet

- Antall tuppler i en relasjon
- Endrer seg ettersom tuppler blir lagt til / slettet.

Identifisere de forskjellige begrepene

<i>Holt</i>	<i>Anne</i>	<i>norsk</i>
<i>Nordmann</i>	<i>Jan</i>	<i>norsk</i>
<i>Kristoffersen</i>	<i>Jan</i>	<i>svensk</i>
<i>Halvorsen</i>	<i>Stieg</i>	<i>svensk</i>

En tuppel er da $T_1 = \{\text{Holt, Anne, norsk}\}$

Domene er $\{\text{norsk, svensk}\}$. Et domene er alle mulige attributter til en tuppel som kan velges. Altså hvis dansk ikke befinner seg i domenet, så kan man ikke legge inn dansk nasjonalitet på en tuppel.
 $\{\text{Anne, Jan, Stieg}\}$ = Fornavns domene.

Attributtet her kunne vært nasjonalitet, fornavn og etternavn.

Hele tabellen er relasjonen.

Relasjonen er av grad 3, pga. 3 attributter, og har kardinaliteten 4 pga. 4 tuppler.

Relasjon-intensjon: Strukturen og domenespesifikasjoner (andre begrensninger)

Relasjon-ekstensjon: Selve dataen i relasjonen, altså tupplene.

Relasjonsdatabase

- Samling av navngitte og normaliserte relasjoner

Normaliserte relasjon er at relasjonen er strukturert på en bestemt måte.

Schema - Relasjonsskjema

Sett med par

Skjema består av attributter og domener = $\{A_n : d_n\}$

Hver attributt er knyttet til tilhørende domene.

Attributt $Farge = A_1$ og $d_1 =$ domenet. Domenet inneholder alle mulige farger $d_1 = \{\text{rød, blå, rosa...}\}$

Relasjons database skjema

$R = (R_1, R_2)$ hvor R_1 feks er forfatter og R_2 er bok. = Alle relasjonene/tabellen i databasen = Relasjonsdatabaseskjema

Nøkler - Relasjonsnøkler

Med nøkkel menes en verdi som gjør det mulig å finne et bestemt tuppel i en relasjon.

Supernøkkel

- Alle attributtene vil til sammen danne en supernøkkel
- Kombinasjon av attributter som unkt identifiserer en tuppel danner supernøkkel.

La oss se på ett eksempel:

$\{002, \text{Mehlum, Jan, norsk}\}$ er supernøkkel i relasjonen: Forfatter(forfatterID, etternavn, fornavn, nasjonalitet)

Supernøkkel til tuppleen gjør det mulig å finne tilbake til denne tuppleen på en enkel måte.

Kandidatnøkkel

- Minimal supernøkkel
- Unikt identifiserer et tuppel
- Ikke reduserbar

Primærnøkkel

- I værste fall vil alle attributtene utgjøre primærnøkkel til tupplet

Alternative nøkler

Resten av kandidatnøkler - de som ikke ble brukt, blir valgt til å være primærnøkler.

Fremmednøkkel

- En attributt-kombinasjon som tilsvarer en kandidatnøkkel i en annen eller i samme relasjon.

Null

-Verdien for ett attributt med ukjent verdi har verdien NULL. Altså det eksisterer ikke.

Ingen verdi er satt enda

Relasjonsmodellen er bygd på 2 verdier. TRUE/FALSE, men noen ganger kommer NULL i tillegg.

Relasjonsmodellen var hovedsaklig bygd på bools logikk.

Referanseintegritet

- Fyller ut verdier til en primærnøkkel.

Må referere til et tuppel som eksisterer

Andre begrensninger som systemet må godkjenne.

Views

ANSII SPARC modellen

External - Hvordan brukere ser fila

Internal - Hvordan ting faktisk er lagret. (Teknikere)

Forskjellige views for forskjellige brukere, teknikere, forbrukere osv.

Sikkerhetstiltak; brukes views for å hindre at ikke alle kan se all informasjonen i relasjonsdatabasen.

Kapittel 5.1

Relasjonsalgebra

Relasjon kan sees på som en tabell

- Består av en mengde med n-tuppler

$$R = \{T_1, T_2 \dots T_n\}$$

Definisjon på en tuppel

Tuppel er en ordnet liste i matematikken

Tuppel er en uordnet liste i databaser

Gikk fra ordnet til uordnete tuppler i relasjonsdatabaser for å oppnå data uavhengighet. For at det skal være lettere å fjerne og legge til tuppler. Når dataen er uavhengig, så har ikke posisjonen til tupplene lengre noe å si).

Nøkler

- Benyttes til å finne fram til forskjellige tupler

Supernøkkel

- En supernøkkel er unik slik at den alltid bare finner fram til en unik tuppel.

Referer til et tuppel i relasjonen / eller ingen. En supernøkkel kan inneholde opp til alle attributter til tupplene.

Eksempel på en supernøkkel kan være alle attributtene til en tuppel: (*Klemetsen, Gard*, 23). Disse 3 attributtene kan være en supernøkkel.

Kandidatnøkkel

- Er en minimal supernøkkel for en relasjon slik at relasjonen ikke har flere like tupler med de samme verdiene for disse attributene. Settet av attributter må være en supernøkkel.
- Alle kandidatenøkler er supernøkler.

Primærnøkkel

Alle relasjoner i en database skal ha en og bare en primærnøkkel. Primærnøkkel er den foretrukne metoden for å identifisere en rad i en tabell i en relasjonsdatabase. For eksempel kan man ha en oversikt/tabell over bøker med en kolonne for ISBN. Alle bøker har ett ISBN, og ingen bøker har samme ISBN. En primærnøkkel er den beste kandidatnøkkel.

Fremmednøkkel

En fremmednøkkel er en henvisning til andre data i en database. I en relasjonsdatabase peker fremmednøkkel til en kandidatnøkkel. Vanligvis vil dette være en primærnøkkel i en annen tabell, men det er også tillatt at fremmednøkler peker til oppføringer i samme tabell som de forekommer i.

Relasjonsalgebra

Relasjonsalgebra er et matematisk system som er bestående av:

- Er et sett med operasjoner som utføres på relasjoner (tabeller). Resultatet av disse operasjonene er alltid en ny relasjon.

Primitive typer

$$R_1 = \{1, 2, 3\}, R_2 = \{3, 4, 5\}$$

$$\text{UNION} - R_3 = R_1 \cup R_2 = \{1, 2, 3, 4, 5\}$$

$$\text{SNITT} - R_3 = R_1 \cap R_2 = \{3\}$$

$$\text{Differanse} - R_3 = R_2 - R_1 = \{4, 5\}$$

Kryssprodukt

- Kan finne antall tuppler i den nye relasjonen ved å multiplisere antall tuppler i den ene relasjonen med den andre.
- $R_3 = R_1 \times R_2 = \{(1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 3), (3, 4), (3, 5)\}$
- Kryssproduktet kan ikke utføres på relasjoner som inneholder like attributter.

Seleksjon

- Seleksjonsoperatoren brukt på en relasjon R gir en ny relasjon som har en undermengde av tuplene i R. Tuplene som blir med er de som tilfredsstiller en betingelse C som går på attributter i R. Seleksjon skrives $\sigma_C(R)$, der C er betingelsen og R en relasjon.
- Seleksjon brukes for å plukke ut tuppler som har felles attributtverdier.

$\sigma_C(R)$

For eksempel: $\sigma_{A=1}(R)$, betyr at alle tupler som ikke har verdien 1 for attributtet A forsvinne. For eksem er tuppelen (A, B, C) hvor A B og C er de forskjellige attributtverdiene. Dersom første posisjon, altså attributtnavn A ikke er 1, så vil ikke denne tuppelen bli valgt ut til den nye mengden R_2 . A sier bare noe om hvilken posisjon attributtet vi skal sjekke er i.

Projeksjon

Projeksjonsoperatoren brukt på en relasjon R vil frembringe en ny relasjon som bare har enkelte av attributtene til R. En projeksjon skrives $\pi_{a_1, \dots, a_n}(R)$, der a_1, \dots, a_n er en mengde attributtnavn og R er en relasjon. Brukes for å liste opp alle attributtverdier, altså på kolonner.

Gjør man projeksjonen $\pi_{A,B}(R)$ vil bare kolonnene A og B fra R komme med i den nye relasjonen. Med projeksjonen $\pi_A(R)$ vil bare kolonne A fra R beholdes. Altså den nye relasjonen vil bare innehold kolonne A i relasjonen.

Omnavning

Man kan ofte ønske å gi relasjoner eller attributter nye navn. Vil man gi relasjonen R det nye navnet S skrives dette $\rho_{S(A_1, \dots, A_n)}(R)$, der A_1, \dots, A_n er attributtnavnene i den nye relasjonen S.

Her har relasjonen fått det nye navnet S, samtidig som attributtene har fått nye navn. Verdiene i tuplene har ikke blitt endret.

Join / Forening

- Operasjonene projeksjon og seleksjon gjort på kryssprodukt
- En join er en spesiell form for produkt der relasjoner pares på bestemte måter.

I en naturlig join mellom relasjonene R_1 og R_2 pares tuplene i de to relasjonene på de attributtene de har felles. Dette skrives $R_1 \bowtie R_2$. Tupler som ikke matcher tupler i den andre relasjonen på ett eller flere felles attributter blir ikke med i den nye relasjonen.

En Theta-join mellom to relasjoner R_1 og R_2 fungerer som en naturlig join, med det unntak at tupler pares på en bestemt betingelse, kalt θ . Dette skrives $R \bowtie_{\theta} S$.

En Outer join er en join der tupler som ikke overholder kravet i joinen likevel blir med i produktet. En outer join på relasjonene R og S gjøres ved å gjøre en join mellom de to relasjonene, deretter legges de mistede tuplene inn igjen med en nullverdi i attributtene de mangler.

Grupering

AL er en liste med aggregeringsoperatorer – hver med et attributt som operand

GA er en liste med grupperingsattributter.
Eksempel: Antall boktitler pr. nasjonalitet.

Aggregering (oppsamling)

AL er en liste med aggregeringsoperatorer – hver med et attributt som operand

Typiske aggregeringsoperatorer:

Sum

Count

Average

Maximum

Minimum

Eksempel: Antall boktitler.