

IN2090 – Databaser og datamodellering

08 – Tapsfri dekomposisjon

Leif Harald Karlsen
leifhka@ifi.uio.no



Universitetet i Oslo

Hvordan oppnå BCNF?

- ◆ Hvordan lage et nytt skjema på BCNF som inneholder den samme informasjonen?
- ◆ Problemet er, grovt sett, at data som ikke hører sammen er i samme tabell.
- ◆ Kan løses ved å **dekomponere** tabellen til mindre tabeller.
- ◆ Kan ikke dekomponere som vi vil – dekomposisjonen må være **tapsfri**.

Tapsfri dekomponering

La $R(X)$ være en relasjon. En dekomponering av R er en mengde nye relasjoner $\{S_1(Y_1), \dots, S_n(Y_n)\}$ slik at

1. $Y_i \subseteq X$

2. $\bigcup_{i=1}^n Y_i = X$

En dekomponering er tapsfri hvis vi alltid kan ta en instans IR av R , projisere ned til instanser IS_i av S_i , og så rekonstruere IR via naturlig join, altså:

$$\pi_{Y_1}(IR) \bowtie \pi_{Y_2}(IR) \bowtie \dots \bowtie \pi_{Y_n}(IR) = IR$$

Ikke tapsfri dekomponering

Opprinnelig tabell: `Ansatte(AvdID, AvdNavn, AnsattID, Navn, Etternavn)`

Dekomponert til:

- ◆ `Avdeling(AvdID, AvdNavn)`
- ◆ `Ansatt(AnsattId, Navn, Etternavn)`

Alle attributter er med, men vi har mistet noe viktig, nemlig **forholdet mellom** ansatt og avdeling.

Tapsfri dekomponering, eksempel

Brnavn	Navn	Etternavn	Adresse	Kurskode	Tittel	Beskrivelse	AntSP	Kara
evgenit	Evgenij	Thorstensen	Addr1	IN2090	Databaser	EnBeskr...	10	B
peternl	Petter	Nilsen	Addr2	IN2090	Databaser	EnBeskr...	10	A
evgenit	Evgenij	Thorstensen	Addr1	IN2080	Beregn...	Descr...	10	A
leifhka	Leif H.	Karlsen	Addr3	IN2090	Databaser	EnBeskr...	10	B
leifhka	Leif H.	Karlsen	Addr3	IN3110	Program...	EnBeskr2...	5	C

Student

Brnavn	Navn	Etternavn	Adresse
evgenit	Evgenij	Thorstensen	Addr1
peternl	Petter	Nilsen	Addr2
leifhka	Leif H.	Karlsen	Addr3

Kurs

Kurskode	Tittel	Beskrivelse	AntSP
IN2090	Databaser	EnBeskr...	10
IN2080	Beregn...	Descr...	10
IN3110	Program...	EnBeskr2...	5

Karakter

Brnavn	Kurskode	Kara
evgenit	IN2090	B
peternl	IN2090	A
evgenit	IN2080	B
leifhka	IN2090	B
leifhka	IN3110	C

Alle attributter er med, og naturlig join gir opprinnelig tabell.

Hvordan garantere tapsfri dekomponering?

- ◆ Fagins teorem: En dekomponering av $R(X, Y, Z)$ til $S_1(X, Y), S_2(X, Z)$ er tapsfri hvis og bare hvis $X \rightarrow Y$
- ◆ Med andre ord, vi kan skille ut noen attributter og det de alle er avhengige av
- ◆ Dette gir opphav til dekomponeringsalgoritmen for BCNF

Tapsfri dekomponering til BCNF

Tapsfri dekomponering av $R(X)$ med FDer F :

1. Beregn nøklene til R
2. For hver FD $Y \rightarrow A \in F$, hvis FDen er et brudd på BCNF:
 - 2.1 beregn Y^+ ,
 - 2.2 og dekomponer R til $S_1(Y^+)$ og $S_2(Y, X/Y^+)$.
3. Fortsett rekursivt (over S_1 og S_2) til ingen brudd på BCNF

(Hvis en FD inneholder attributter fra ulike tabeller kan den ignoreres)

Eksempel 1

Tapsfri dekomponering av $R(X)$ med FDer F :

1. For hver FD $Y \rightarrow A \in F$, hvis FDen er et brudd på BCNF:
 - 1.1 beregn Y^+ ,
 - 1.2 og dekomponer R til $S_1(Y^+)$ og $S_2(Y, X/Y^+)$.
2. Fortsett rekursivt (over S_1 og S_2) til ingen brudd på BCNF

La $R(A, B, C)$ ha FDer $F = \{AB \rightarrow C, C \rightarrow A\}$.

- ◆ Kandidatnøkkel: B forekommer ikke på høyresider, så B er med i alle nøkler. $\{A, B\}$ og $\{B, C\}$ er nøklene
- ◆ $AB \rightarrow C$ er ikke brudd på BCNF, siden AB er en supernøkkel
- ◆ $C \rightarrow A$ er brudd på BCNF, men ikke på 3NF, siden A er et nøkkelattributt
- ◆ Beregner $C^+ = CA$
- ◆ Dekomponerer R til $S_1(C, A)$ og $S_2(C, B)$
- ◆ Kun én FD som holder for S_1 ($C \rightarrow A$) og bryter ikke med BCNF og ingen FDer for S_2 , altså begge på BCNF
- ◆ $R(A, B, C)$ dekomponeres dermed til $S_1(C, A)$ og $S_2(C, B)$

Eksempel 2

Tapsfri dekomponering av $R(X)$ med FDer F :

1. For hver FD $Y \rightarrow A \in F$, hvis FDen er et brudd på BCNF:
 - 1.1 beregn Y^+ ,
 - 1.2 og dekomponer R til $S_1(Y^+)$ og $S_2(Y, X/Y^+)$.
2. Fortsett rekursivt (over S_1 og S_2) til ingen brudd på BCNF

$S(\text{Brnavn}, \text{Navn}, \text{Etternavn}, \text{Kurskode}, \text{Kurstittel}, \text{Karakter})$

FDer:

1. $\text{Brnavn}, \text{Kurskode} \rightarrow \text{Karakter}$
 2. $\text{Brnavn} \rightarrow \text{Navn}$
 3. $\text{Brnavn} \rightarrow \text{Etternavn}$
 4. $\text{Kurskode} \rightarrow \text{Kurstittel}$
- ◆ Kandidatnøkkel: $\{\text{Brnavn}, \text{Kurskode}\}$
 - ◆ $\text{Brnavn} \rightarrow \text{Navn}$ bryter med BCNF
 - ◆ Beregner $\text{Brnavn}^+ = \{\text{Brnavn}, \text{Navn}, \text{Etternavn}\}$
 - ◆ Får da $S_1(\text{Brnavn}, \text{Navn}, \text{Etternavn})$ og $S_2(\text{Brnavn}, \text{Kurskode}, \text{Kurstittel}, \text{Karakter})$
 - ◆ S_1 har FDene 2. og 3., men ingen av disse bryter BCNF
 - ◆ S_2 har FDene 1. og 4.

Eksempel 2 (forts.)

Tapsfri dekomponering av $R(X)$ med FDer F :

1. For hver FD $Y \rightarrow A \in F$, hvis FDen er et brudd på BCNF:
 - 1.1 beregn Y^+ ,
 - 1.2 og dekomponer R til $S_1(Y^+)$ og $S_2(Y, X/Y^+)$.
2. Fortsett rekursivt (over S_1 og S_2) til ingen brudd på BCNF

$S_2(\text{Brnavn}, \text{Kurskode}, \text{Kurstittel}, \text{Karakter})$

FDer:

1. Kurskode \rightarrow Kurstittel
2. Brnavn, Kurskode \rightarrow Karakter
 - ◆ Kandidatnøkkel: {Brnavn, Kurskode}
 - ◆ Kurskode \rightarrow Kurstittel bryter med BCNF
 - ◆ $\text{Kurskode}^+ = \text{Kurskode}, \text{Kurstittel}$
 - ◆ Får $S_{21}(\text{Kurskode}, \text{Kurstittel})$ og $S_{22}(\text{Kurskode}, \text{Brnavn}, \text{Karakter})$
 - ◆ S_{21} har kun første FD som ikke bryter med BCNF
 - ◆ S_{22} har kun andre FD som ikke bryter med BCNF

Eksempel 2 (forts.)

$S(\text{Brnavn}, \text{Navn}, \text{Etternavn}, \text{Kurskode}, \text{KursTittel}, \text{Karakter})$

FDer:

1. $\text{Brnavn}, \text{Kurskode} \rightarrow \text{Karakter}$
2. $\text{Brnavn} \rightarrow \text{Navn}$
3. $\text{Brnavn} \rightarrow \text{Etternavn}$
4. $\text{Kurskode} \rightarrow \text{KursTittel}$

Dekomponeres altså til:

- ◆ $S_1(\text{Brnavn}, \text{Navn}, \text{Etternavn})$
- ◆ $S_{21}(\text{Kurskode}, \text{Kurstittel})$
- ◆ $S_{22}(\text{Kurskode}, \text{Brnavn}, \text{Karakterer})$

Dekomponering i praksis

- ◆ Algoritmen gir oss den riktige strukturen på tabellene
- ◆ Må etterpå gi tabellene meningsfulle navn og sette skranker på kolonner
- ◆ Dersom man startet med en skjema som inneholdt data må disse flyttes over
- ◆ Gitt en tabell $R(X)$ som dekomponeres til $S_1(Y_1), \dots, S_n(Y_n)$ kan dette gjøres ved å kjøre følgende for hver S_i :

```
INSERT INTO S_i  
SELECT DISTINCT Y_i  
FROM R;
```

Takk for nå!

Neste video vil handle om design i praksis.