



Oppskriftsbok

- **FDer og MVDer** - oversikt: se s. 2
- **Relasjonsalgebra** - oversikt: se s. 45
- Oppskriftsboken er ment brukt som oppslagsverk og eksempelsamling. Ikke alt er like viktig. I pensumlisten står detaljer om hva som er pensum.

FDer og MVDer

- Tillukningsalgoritmen (s. 3)
- Hvordan finne alle kandidatnøkler (s. 5)
- Hvordan finne normalformen til en relasjon (s. 7)
- Hvordan avgjøre om en dekomposisjon er tapsfri (s. 10)
- Hvordan avgjøre om en dekomposisjon er FD-bevarende (s. 13)
- Hvordan dekomponere tapsfritt til BCNF (s. 16)
- Hvordan dekomponere tapsfritt til EKNF (s. 18)
- Hvordan avgjøre om en FD følger fra en gitt mengde integritetsregler (s. 23)
- Hvordan avgjøre om en MVD følger fra en gitt mengde integritetsregler (s. 30)
- Hvordan finne alle kandidatnøkler i nærvær av MVDer (s. 33)
- Hvordan avgjøre om en dekomposisjon er tapsfri når noen av reglene er MVDer (s. 35)
- Hvordan dekomponere tapsfritt til 4NF (s. 37)
- Hvordan finne ut om en dekomposisjon kan ha støyinstanser (s. 40)

Tillukningsalgoritmen

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} . La X være en mengde av attributter i \mathcal{R} . X^+ , tillukningen av X , er de attributtene som avhenger funksjonelt av X (dvs. de attributtene som, gitt en verdi til hvert av attributtene i X , maksimalt kan ha én verdi).

Algoritme som beregner X^+ mhp. \mathcal{F} .

1. $T := X$
2. Så lenge T forandres:
Hvis det finnes en FD $V \rightarrow W$ i \mathcal{F} hvor V er en delmengde av T , sett $T := T \cup W$
3. $X^+ := T$

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn tillukningen av AB.

1. $T := AB$

2. Siden $AB \subseteq T$ og $AB \rightarrow DE$: Sett $T := ABDE$

Siden $AE \subseteq T$ og $AE \rightarrow BF$: Sett $T := ABDEF$

3. $AB^+ = ABDEF$

Hvordan finne alle kandidatnøkler

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} .

1. Sett X lik mengden av attributter som ikke forekommer i noen høyreside i \mathcal{F} .
2. Utvid systematisk X på alle mulige måter med attributter som forekommer i minst én venstreside i \mathcal{F} . For hver slik X , beregn tillukningen X^+ . Stans utvidelsen av en X hvis X^+ er samtlige attributter.

Hvis X^+ er samtlige attributter i \mathcal{R} , sjekk at uansett valg av en A i X , er $(X-A)^+$ **ikke** alle attributtene i \mathcal{R} . Hvis dette er tilfellet, er X en kandidatnøkkel.

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn kandidatnøklene til \mathcal{R} .

Attributter som ikke forekommer i noen høyreside: C

Attributter som bare forekommer i høyresider: F

Attributter som forekommer i minst én venstreside: A, B, C, D, E

1. $X := C$. $C^+ = CA$. C er ikke en kandidatnøkkel.
2. Prøv å utvide X med B, D, E. (Siden A er i C^+ , er det ikke noe poeng å utvide C med A.)
 1. $X := BC$. $BC^+ = BCDEF$. BC er en kandidatnøkkel.
 2. $X := CD$. $CD^+ = CDA$. CD er ikke en kandidatnøkkel.
 3. $X := CE$. $CE^+ = BCDEF$. CE er en kandidatnøkkel.

Fortsett med $X = CD$. Prøv å utvide med B, E.

4. $X := BCD$. $BCD^+ = BCDEF$. Men BC er en kandidatnøkkel, så BCD er ikke en kandidatnøkkel.
5. $X := CDE$. $CDE^+ = BCDEF$. Men CE er en kandidatnøkkel, så CDE er ikke en kandidatnøkkel.

Kandidatnøklene er BC og CE.

Hvordan finne normalformen til en relasjon

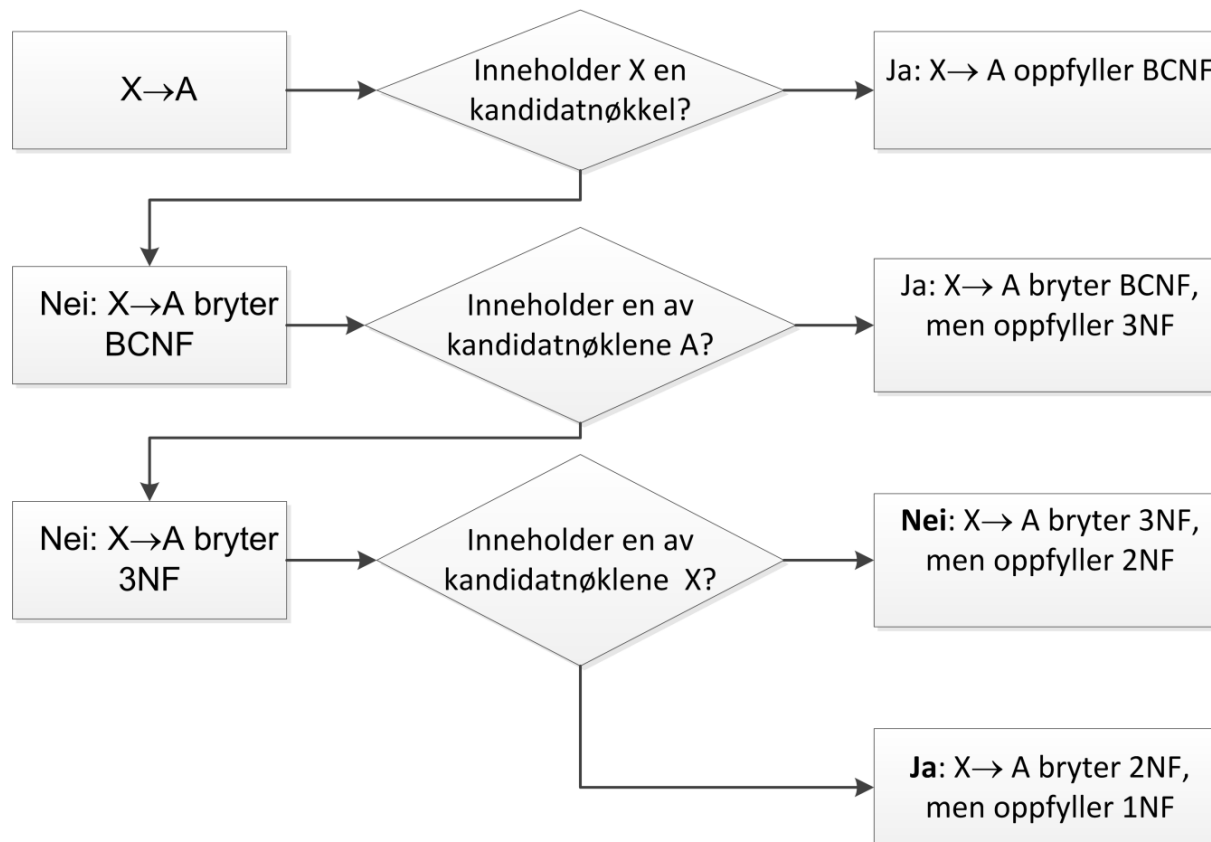
Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} .

1. Finn alle kandidatnøkklene i \mathcal{R} .
2. Lag atomære høyresider ved hjelp av splitting:
Hvis $X \rightarrow B_1 B_2 \dots B_n$ er en FD i \mathcal{F} , erstatt den med FDene $X \rightarrow B_1, X \rightarrow B_2, \dots, X \rightarrow B_n$.
3. For hver FD $X \rightarrow A$:
 1. Undersøk om X er en supernøkkel, dvs. om X inneholder en kandidatnøkkel. Hvis den gjør det, så oppfyller $X \rightarrow A$ BCNF.
 2. Hvis ikke, bryter $X \rightarrow A$ BCNF. Undersøk i såfall om A er et nøkkelattributt, dvs. om noen av kandidatnøkklene inneholder A . Hvis dette er tilfellet, så oppfyller $X \rightarrow A$ 3NF, men bryter BCNF.
 3. Hvis ikke, bryter $X \rightarrow A$ 3NF. Undersøk i såfall om X er inneholdt i noen kandidatnøkkel. Hvis dette er tilfellet, så **bryter** $X \rightarrow A$ 2NF (men oppfyller 1NF). Hvis det **ikke** er tilfellet, oppfyller $X \rightarrow A$ 2NF, men bryter 3NF.

Hvis minst én FD bryter en normalform, bryter også \mathcal{R} denne normalformen.

Hvis alle FDene oppfyller en normalform, oppfyller også \mathcal{R} denne normalformen.

Hvordan finne normalformen – flytdiagrambeskrivelse



Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn normalformen til \mathcal{R} . (Dvs. finn **høyeste** normalform som \mathcal{R} tilfredsstillers.)

1. Kandidatnøkklene er BC og CE (se side 6).
2. $\mathcal{F} = \{AB \rightarrow D, AB \rightarrow E, C \rightarrow A, BD \rightarrow E, AE \rightarrow B, AE \rightarrow F\}$
3. $AB \rightarrow D$: AB inneholder ikke BC eller CE. Bryter BCNF.
D er ikke med i BC eller CE. Bryter 3NF.
BC og CE inneholder ikke AB. Er på 2NF.
(Så \mathcal{R} er på 2NF eller lavere.)
 $AB \rightarrow E$: AB inneholder ikke BC eller CE. Bryter BCNF.
E er med i CE. Er på 3NF.
 $C \rightarrow A$: C inneholder ikke BC eller CE. Bryter BCNF.
A er ikke med i BC eller CE. Bryter 3NF.
BC (og CE) inneholder C. Bryter 2NF, men er på 1NF.
(Så \mathcal{R} er på 1NF, men ikke 2NF.)

Behøver i dette tilfellet ikke å undersøke resten av FDene, for uansett hvilken normalform de øvrige FDene har, vil \mathcal{R} uansett bryte 2NF og bare være på 1NF.

\mathcal{R} er på 1NF, men bryter/er ikke på 2NF.

Hvordan avgjøre om en dekomposisjon er tapsfri

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} . Gitt en dekomposisjon av \mathcal{R} til relasjonene S_1, \dots, S_k .

Chasealgoritme for å avgjøre om dekomposisjonen er tapsfri:

1. Lag en tabell med én kolonne for hvert attributt i \mathcal{R} og én rad for hver S_i .
2. I kolonnen for attributt A, for hver rad i:
 - Skriv a hvis A er et attributt i S_i
 - Skriv a_i hvis A ikke er et attributt i S_i
3. Så lenge det skjer forandringer i tabellen:
 - For hver FD $X \rightarrow Y$, for alle rader i tabellen med lik X-verdi, gjør Y-verdiene like. Hvis en av Y-ene er en verdi uten subskript, skal denne velges.
5. Hvis en rad er uten subskriptverdier, er dekomposisjonen tapsfri, ellers ikke.

Eksempel 1

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn ut om $\{ABDE, AC, BDE, ABEF\}$ er en tapsfri dekomposisjon.

	A	B	C	D	E	F
ABDE	a	b	c ₁	d	e	f ₁ f
AC	a	b ₂	c	d ₂	e ₂	f ₂
BDE	a ₃	b	c ₃	d	e	f ₃
ABEF	a	b	c ₄	d ₄ d	e	f

Ingen rader er fullstendig uten subskriptverdier.

Dekomposisjonen er ikke tapsfri.

Eksempel 2

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn ut om $\{ACE, BDE, ABEF, ABD\}$ er en tapsfri dekomposisjon.

	A	B	C	D	E	F
ACE	a	b ₁ b	c	d ₁ d	e	f ₁ f
BDE	a ₂	b	c ₂	d	e	f ₂
ABEF	a	b	c ₃	d ₃ d ₁ d	e	f
ABD	a	b	c ₄	d	e ₄	f ₄

Første rad er uten subskriptverdier.

Dekomposisjonen er tapsfri.

Hvordan avgjøre om en dekomposisjon er FD-bevarende

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} .

Gitt en dekomposisjon av \mathcal{R} til relasjonene S_1, \dots, S_k der S_i har attributtene W_i ($i=1, \dots, k$).

Den komplette algoritmen for å avgjøre om dekomposisjonen er FD-bevarende, er ikke pensum. Følgende forenklete algoritme kan benyttes på de eksemplene dere kommer borti i dette kurset[†]:

1. Lag atomære høyresider ved hjelp av splitting:
Hvis $X \rightarrow B_1 B_2 \dots B_n$ er en FD i \mathcal{F} , erstatt den med FDene
 $X \rightarrow B_1, X \rightarrow B_2, \dots, X \rightarrow B_n$.
2. For hver FD $X \rightarrow A$: Sjekk om det finnes minst én S_i der $XA \subseteq W_i$. Hvis dette er tilfellet for alle FDene, er dekomposisjonen FD-bevarende. Hvis det er minst én FD der $XA \not\subseteq W_i$ for alle i , er dekomposisjonen ikke FD-bevarende.

[†]Hvis den forenklete algoritmen konkluderer med at dekomposisjonen er FD-bevarende, så er dette alltid korrekt. Men den forenklete algoritmen kan noen ganger konkludere med at dekomposisjonen **ikke** er FD-bevarende selv om det korrekte er at den **er** FD-bevarende.

Eksempel 1

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn ut om dekomposisjonen $\{BDE, ABDF, AC, BC\}$ er FD-bevarende.

- $\mathcal{F} = \{AB \rightarrow D, AB \rightarrow E, C \rightarrow A, BD \rightarrow E, AE \rightarrow B, AE \rightarrow F\}$
- $AB \rightarrow D: ABD \subseteq ABDF$
 $AB \rightarrow E: ABE \not\subseteq BDE, ABE \not\subseteq ABDF, ABE \not\subseteq AC, ABE \not\subseteq BC$
 $AB \rightarrow E$ kan ikke sjekkes lokalt i noen av komponentene.[†]

$C \rightarrow A: AC \subseteq AC$
 $BD \rightarrow E: BDE \subseteq BDE$
 $AE \rightarrow B: ABE \not\subseteq BDE, ABE \not\subseteq ABDF, ABE \not\subseteq AC, ABE \not\subseteq BC$
 $AE \rightarrow B$ kan ikke sjekkes lokalt i noen av komponentene.
 $AE \rightarrow F: AEF \not\subseteq BDE, AEF \not\subseteq ABDF, AEF \not\subseteq AC, AEF \not\subseteq BC$
 $AE \rightarrow F$ kan ikke sjekkes lokalt i noen av komponentene.

Dekomposisjonen er ikke FD-bevarende.

[†] $AB \rightarrow E$ følger fra FDene $AB \rightarrow D$ og $BD \rightarrow E$ som begge kan sjekkes lokalt, så hvis det bare hadde vært $AB \rightarrow E$ som gikk "på tvers" av komponentene, ville dekomposisjonen likevel ha vært FD-bevarende. Men det tilsvarende er ikke tilfellet for $AE \rightarrow B$ og $AE \rightarrow F$.

Eksempel 2

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn ut om dekomposisjonen $\{ACE, BDE, ABEF, ABD\}$ er FD-bevarende.

1. $\mathcal{F} = \{AB \rightarrow D, AB \rightarrow E, C \rightarrow A, BD \rightarrow E, AE \rightarrow B, AE \rightarrow F\}$

2. $AB \rightarrow D: ABD \subseteq ABD$

$AB \rightarrow E: ABE \subseteq ABEF$

$C \rightarrow A: AC \subseteq ACE$

$BD \rightarrow E: BDE \subseteq BDE$

$AE \rightarrow B: ABE \subseteq ABEF$

$AE \rightarrow F: AEF \subseteq ABEF$

Alle FDene kan sjekkes lokalt i en komponent.

Dekomposisjonen er FD-bevarende.

Hvordan dekomponere tapsfritt til BCNF

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} .

1. Hvis $X \rightarrow A$ er et brudd på BCNF:
 - A. Beregn X^+
 - B. Dekomponer \mathcal{R} i S og \mathcal{T} der S er lik X^+ og \mathcal{T} er lik X samt de attributtene i \mathcal{R} som ikke er i X^+ .
2. Fortsett på samme måte med S og \mathcal{T} inntil alle relasjonene i dekomposisjonen tilfredsstill BCNF.
(Brudd på BCNF må beregnes på nytt[†] for hver av S og \mathcal{T} .)

[†]For å beregne eventuelle brudd på BCNF i hver av komponentene, må vi vite hvilke FDer som gjelder lokalt i hver komponent. Den generelle algoritmen for å finne de lokale FDene, er ikke pensum. De eksemplene dere kommer borti i dette kurset, vil ikke være mer kompliserte enn at det stort sett gir seg selv hvilke FDer som gjelder lokalt.

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

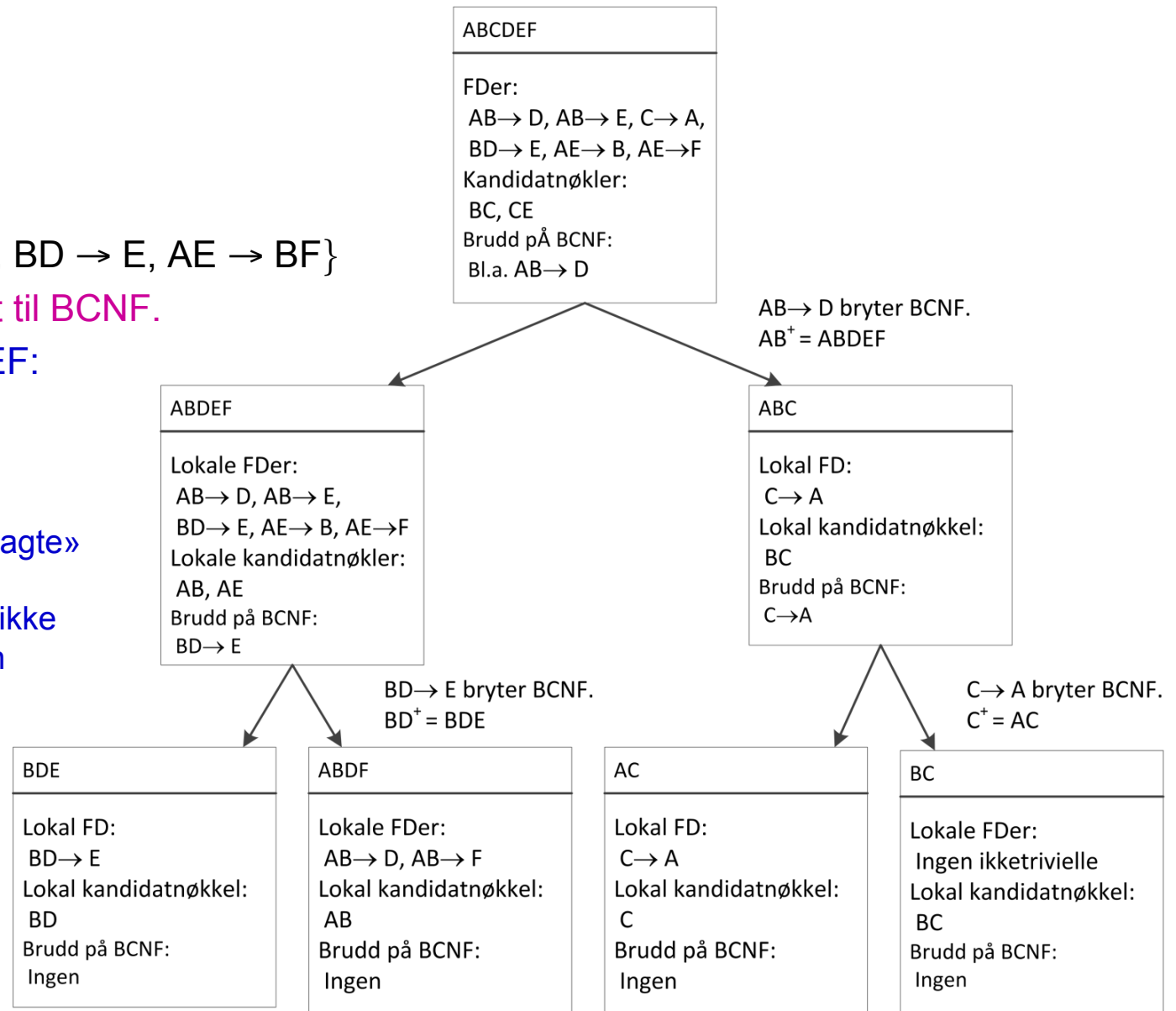
$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Dekomponer \mathcal{R} tapsfritt til BCNF.

Kandidatnøkler ABCDEF:
BC og CE, se side 6.

Merk at vi i komponenten ABDF i tillegg til den «opplagte» FDen $AB \rightarrow D$ har oppgitt FDen $AB \rightarrow F$. $AB \rightarrow F$ er ikke i \mathcal{F} , men følger fra \mathcal{F} siden $AB^+ = ABDEF$. Så lokalt i komponenten gjelder $AB \rightarrow F$.

$\{BDE, ABDF, AC, BC\}$ er en BCNF-dekomposisjon.



Hvordan dekomponere tapsfritt til EKNF

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} .

1. Finn en minimal overdekning \mathcal{G} for \mathcal{F} (se side 20).
2. For hver X som opptrer som venstreside i \mathcal{G} , finn alle FDer på formen $X \rightarrow A_i$ i \mathcal{G} og lag en relasjon \mathcal{R}_X som inneholder X samt alle høyresidene A_i .
3. Hvis ingen av \mathcal{R}_X -ene inneholder en kandidatnøkkel for \mathcal{R} , så utvid med en relasjon \mathcal{R}_0 der attributtene er en kandidatnøkkel.

Relasjonene i dekomposisjonen er minst EKNF.

(I beste fall oppnås BCNF i alle relasjonene.)

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Dekomponer \mathcal{R} tapsfritt til EKNF.

1. En minimal overdekning av \mathcal{F} er

$\mathcal{G} = \{AB \rightarrow D, C \rightarrow A, BD \rightarrow E, AE \rightarrow B, AE \rightarrow F\}$ (se side 21).

2. $AB \rightarrow D$ gir $\mathcal{R}_{AB}(ABD)$
 $C \rightarrow A$ gir $\mathcal{R}_C(AC)$
 $BD \rightarrow E$ gir $\mathcal{R}_{BD}(BDE)$
 $AE \rightarrow B$ og $AE \rightarrow F$ gir $\mathcal{R}_{AE}(ABEF)$

3. Kandidatnøklene er BC og CE (se side 6). Ingen av komponentene i punkt 2 inneholder noen kandidatnøkkel. Legg til $\mathcal{R}_0(BC)$ eller $\mathcal{R}_0(CE)$.

$\{ABD, AC, BDE, ABEF, CE\}$ er en EKNF-dekomposisjon.

(I dette tilfellet er alle komponentene ikke bare på EKNF, men på BCNF også.)

Algoritme for å finne minimale overdekninger

1. Initialiser $\mathcal{G} := \mathcal{F}$.
2. Lag atomære høyresider ved hjelp av splitting.
3. Gjør venstresidene minimale:
For hver $X \rightarrow A$ i \mathcal{G} og hver B i X :
 1. Beregn $(X-B)^+$ med hensyn på \mathcal{G} .
 2. Hvis A er et attributt i $(X-B)^+$, erstatt $X \rightarrow A$ med $(X-B) \rightarrow A$ i \mathcal{G} .
4. Fjern overflødige FDer:
For hver $X \rightarrow A$ i \mathcal{G} :
 1. Beregn X^+ med hensyn på \mathcal{G} uten å bruke $X \rightarrow A$.
 2. Hvis A er et attributt i X^+ , fjern $X \rightarrow A$ fra \mathcal{G} .

Eksempel 1

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn en minimal overdekning for \mathcal{F} .

1. $\mathcal{G} := \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

2. $\mathcal{G} := \{AB \rightarrow D, AB \rightarrow E, C \rightarrow A, BD \rightarrow E, AE \rightarrow B, AE \rightarrow F\}$

3. $AB \rightarrow D$: 1. $A^+ = A, B^+ = B$.

2. D er ikke med i A^+ eller B^+ , så $AB \rightarrow D$ har allerede minimal venstreside.

$AB \rightarrow E$: Det samme gjelder denne.

$BD \rightarrow E$: 1. $B^+ = B, D^+ = D$.

2. E er ikke med i B^+ eller D^+ , så $BD \rightarrow E$ har allerede minimal venstreside.

$AE \rightarrow B$: 1. $A^+ = A, E^+ = E$.

2. B er ikke med i A^+ eller E^+ , så $AE \rightarrow B$ har allerede minimal venstreside.

$AB \rightarrow D$: Det samme gjelder denne.

Resultat: \mathcal{G} er uendret etter dette trinnet.

4. $AB \rightarrow D$: 1. $AB^+ = ABEF$ når $AB \rightarrow D$ ikke brukes.

2. D er ikke med i $ABEF$, så $AB \rightarrow D$ er ikke overflødig.

$AB \rightarrow E$: 1. $AB^+ = ABDEF$ når $AB \rightarrow E$ ikke brukes.

2. E er med i $ABDEF$, så $AB \rightarrow E$ er overflødig. Fjern $AB \rightarrow E$ fra \mathcal{G} .

Resten av FDene undersøkes på tilsvarende måte: Bare $AB \rightarrow E$ er overflødig.

Resultat: $\mathcal{G} = \{AB \rightarrow D, C \rightarrow A, BD \rightarrow E, AE \rightarrow B, AE \rightarrow F\}$

Eksempel 2

$\mathcal{R}(A,B,C,D,E,F,G,H)$

$\mathcal{F} = \{AB \rightarrow CH, DE \rightarrow FG, D \rightarrow H, E \rightarrow B, H \rightarrow AG\}$

Finn en minimal overdekning for \mathcal{F} .

- $\mathcal{G} := \{AB \rightarrow CH, DE \rightarrow FG, D \rightarrow H, E \rightarrow B, H \rightarrow AG\}$
- $\mathcal{G} := \{AB \rightarrow C, AB \rightarrow H, DE \rightarrow F, DE \rightarrow G, D \rightarrow H, E \rightarrow B, H \rightarrow A, H \rightarrow G\}$
- $AB \rightarrow C$:
 - $A^+ = A, B^+ = B$.
 - C er ikke med i A^+ eller B^+ , så $AB \rightarrow C$ har allerede minimal venstreside. $AB \rightarrow H$: Det samme gjelder denne.
- $DE \rightarrow F$:
 - $D^+ = ADGH, E^+ = BE$.
 - F er ikke med i D^+ eller E^+ , så $DE \rightarrow F$ har allerede minimal venstreside. $DE \rightarrow G$:
 - $D^+ = ADGH, E^+ = BE$.
 - G er med i D^+ . Erstatt $DE \rightarrow G$ med $D \rightarrow G$:
 $\mathcal{G} := \{AB \rightarrow C, AB \rightarrow H, DE \rightarrow F, D \rightarrow G, D \rightarrow H, E \rightarrow B, H \rightarrow A, H \rightarrow G\}$
- $AB \rightarrow C$:
 - $AB^+ = ABGH$ når $AB \rightarrow C$ ikke brukes.
 - C er ikke med i $ABGH$, så $AB \rightarrow C$ er ikke overflødig. $AB \rightarrow H$:
 - $AB^+ = ABC$ når $AB \rightarrow H$ ikke brukes.
 - H er ikke med i ABC , så $AB \rightarrow H$ er ikke overflødig. $DE \rightarrow F$:
 - $DE^+ = ABCDEGH$ når $DE \rightarrow F$ ikke brukes.
 - F er ikke med i $ABCDEGH$, så $DE \rightarrow F$ er ikke overflødig. $D \rightarrow G$:
 - $D^+ = DGH$ når $D \rightarrow G$ ikke brukes.
 - G er med i DGH , så $D \rightarrow G$ er overflødig. Fjern den:
 $\mathcal{G} := \{AB \rightarrow C, AB \rightarrow H, DE \rightarrow F, D \rightarrow H, E \rightarrow B, H \rightarrow A, H \rightarrow G\}$ $D \rightarrow H$:
 - $D^+ = D$ når $D \rightarrow H$ ikke brukes (ser bare på FDene i den nyeste versjonen av \mathcal{G}).
 - H er ikke med i D , så $D \rightarrow H$ er ikke overflødig. $E \rightarrow B, H \rightarrow A, H \rightarrow G$: Gjør tilsvarende for disse. Ingen av dem er overflødige.

Resultat: $\mathcal{G} = \{AB \rightarrow C, AB \rightarrow H, DE \rightarrow F, D \rightarrow H, E \rightarrow B, H \rightarrow A, H \rightarrow G\}$

Hvordan avgjøre om en FD følger fra en gitt mengde integritetsregler

- A. Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} . Vi ønsker å teste om en FD $X \rightarrow Y$ er en konsekvens av integritetsreglene \mathcal{F} . Det er to måter å gjøre dette på:
1. Med tillukningsalgoritmen
 2. Med chasealgoritmen
- B. Gitt en relasjon \mathcal{R} , en mengde FDer \mathcal{F} og en mengde MVDer \mathcal{M} . Vi ønsker å teste om en FD $X \rightarrow Y$ er en konsekvens av integritetsreglene \mathcal{F} og \mathcal{M} . Når noen av reglene er MVDer, må vi bruke chasealgoritmen for å teste dette.

Bruk av tillukningsalgoritmen for å teste om $X \rightarrow Y$ følger fra en mengde FDer \mathcal{F}

1. Beregn X^+ mhp. \mathcal{F}
2. Hvis alle attributtene i Y er med i X^+ , følger $X \rightarrow Y$ av \mathcal{F}

Hvis ikke, følger ikke $X \rightarrow Y$ av \mathcal{F}

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn ut om $AB \rightarrow F$ og om $AB \rightarrow C$.

$AB^+ = ABDEF$

F er med i ABDEF. C er ikke med i ABDEF.

Resultat: $AB \rightarrow F$ følger fra \mathcal{F} , men det gjør ikke $AB \rightarrow C$.

Bruk av chasealgoritmen for å teste om $X \rightarrow Y$ følger fra en mengde FDer \mathcal{F}

1. Lag en tabell med en kolonne for hvert attributt i \mathcal{R} .
Lag to rader i tabellen:
Rad 1: I kolonnen for attributt A, skriv a
Rad 2: I kolonnen for attributt A, skriv a_1 hvis A er et attributt i X og skriv a_2 hvis A ikke er et attributt i X
2. Så lenge det skjer forandringer i tabellen:
For hver FD $V \rightarrow W$ i \mathcal{F} , hvis de to radene har lik V-verdi, gjør W-verdiene like. (Velg verdier uten subskript.)
3. Hvis de to radene har like verdier for alle attributtene i Y, følger $X \rightarrow Y$ fra \mathcal{F} . Hvis minst ett av attributtene i Y har forskjellige verdier, følger ikke $X \rightarrow Y$ fra \mathcal{F} .

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Finn ut om $AB \rightarrow F$ og om $AB \rightarrow C$.

Start med tabellen

A	B	C	D	E	F
a	b	c	d	e	f
a	b	c ₂	d ₂	e ₂	f ₂

Etter at chasealgoritmen er ferdig, ser tabellen slik ut:

A	B	C	D	E	F
a	b	c	d	e	f
a	b	c ₂	d ₂ d	e ₂ e	f ₂ f

De to radene er like i F, men ulike i C.

Resultat: $AB \rightarrow F$ følger fra \mathcal{F} , men det gjør ikke $AB \rightarrow C$.

Bruk av chasealgoritmen for å teste om $X \rightarrow Y$ følger fra en mengde FDer \mathcal{F} og en mengde MVDer \mathcal{M}

1. Lag en tabell med en kolonne for hvert attributt i \mathcal{R} .
Lag to rader i tabellen.
Rad 1: I kolonnen for attributt A, skriv a_1
Rad 2: I kolonnen for attributt A, skriv a_1 hvis A er et attributt i X og skriv a_2 hvis A ikke er et attributt i X
2. Så lenge det skjer forandringer i tabellen:
 - For hver FD $V \rightarrow W$ i \mathcal{F} , for alle par av rader med lik V-verdi, gjør W-verdiene like. (Velg verdier uten subskript.)
 - For hver MVD $V \twoheadrightarrow W$ i \mathcal{M} , for alle par av rader med lik V-verdi, lag to nye rader der W-verdiene er byttet om.
3. Hvis de to opprinnelige radene har like verdier for alle attributtene i Y, følger $X \rightarrow Y$ fra \mathcal{F} og \mathcal{M} . Hvis minst ett av attributtene i Y har forskjellige verdier, følger ikke $X \rightarrow Y$ fra \mathcal{F} og \mathcal{M} .

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

$\mathcal{M} = \{BD \rightarrow CF\}$

Finn ut om $BD \rightarrow F$ og om $BD \rightarrow C$.

Start med tabellen

A	B	C	D	E	F
a	b	c	d	e	f
a ₂	b	c ₂	d	e ₂	f ₂

Etter at chasealgoritmen er ferdig, ser tabellen slik ut:

A	B	C	D	E	F
a	b	c	d	e	f
a ₂ a	b	c ₂	d	e ₂ e	f ₂ f
a	b	c ₂	d	e	f ₂ f
a ₂ a	b	c	d	e	f

De to første radene er like i F, men ulike i C.

Resultat: $BD \rightarrow F$ følger fra \mathcal{F} og \mathcal{M} , men det gjør ikke $BD \rightarrow C$.

Hvordan avgjøre om en MVD følger fra en gitt mengde integritetsregler

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} og en mengde MVDer \mathcal{M} . Vi ønsker å teste om en MVD $X \twoheadrightarrow Y$ er en konsekvens av integritetsreglene \mathcal{F} og \mathcal{M} . Dette testes ved hjelp av chasealgoritmen.

Bruk av chasealgoritmen for å teste om $X \rightarrow Y$ følger fra en mengde FDer \mathcal{F} og en mengde MVDer \mathcal{M}

La Z være de attributtene som ikke forekommer i X eller Y .

1. Lag en tabell med en kolonne for hvert attributt i \mathcal{R} . Lag to rader i tabellen.
Rad 1: I kolonnen for attributt A , skriv a hvis A er et attributt i X eller Y og skriv a_1 hvis A ikke er et attributt i X eller Y
Rad 2: I kolonnen for attributt A , skriv a hvis A er et attributt i X eller Z og skriv a_2 hvis A ikke er et attributt i X eller Z
2. Så lenge det skjer forandringer i tabellen:
 - For hver FD $V \rightarrow W$ i \mathcal{F} , for alle par av rader med lik V -verdi, gjør W -verdiene like. (Velg verdier uten subskript.)
 - For hver MVD $V \twoheadrightarrow W$ i \mathcal{M} , for alle par av rader med lik V -verdi, lag to nye rader der W -verdiene er byttet om.
3. Hvis en av de resulterende radene er uten subskriptverdier, følger $X \rightarrow Y$ fra \mathcal{F} og \mathcal{M} . Hvis ikke, følger ikke $X \rightarrow Y$ fra \mathcal{F} og \mathcal{M} .

Eksempel

$$\mathcal{R}(A, B, C, D)$$
$$\mathcal{F} = \{A \rightarrow B\}$$
$$\mathcal{M} = \{B \rightarrow C\}$$

Finn ut om $A \rightarrow D$.

Start med tabellen

A	B	C	D
a	b_1	c_1	d
a	b	c	d_2

Etter at chasealgoritmen er ferdig, ser tabellen slik ut:

A	B	C	D
a	b_1 b	c_1	d
a	b	c	d_2
a	b	c	d
a	b	c_1	d_2

Den nest siste raden er uten subskriptverdier.

Resultat: $A \rightarrow D$ følger fra \mathcal{F} og \mathcal{M} .

Hvordan finne alle kandidatnøkler i nærvær av MVDer

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} og en mengde MVDer \mathcal{M} . La R være samtlige attributter i \mathcal{R} .

1. Sett X lik mengden av attributter som ikke forekommer i noen høyreside i \mathcal{F} .
2. Utvid systematisk X på alle mulige måter med attributter som forekommer i minst én venstreside i \mathcal{F} eller \mathcal{M} . For hver slik X , undersøk med chasealgoritmen om $X \rightarrow R$. Stans utvidelsen av en X hvis $X \rightarrow R$ holder.

For hver X der $X \rightarrow R$ holder: Hvis det uansett valg av en A i X **ikke** er slik at $(X-A) \rightarrow R$ følger fra \mathcal{F} og \mathcal{M} , er X en kandidatnøkkel.

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

$\mathcal{M} = \{A \twoheadrightarrow BC\}$

Finn alle kandidatnøklene til \mathcal{R} .

Attributter som ikke forekommer i noen høyreside i \mathcal{F} : C

Attributter som bare forekommer i høyresider i \mathcal{F} og \mathcal{M} : F

1. $X := C$. Bruker chase på to rader som er satt opp som beskrevet s. 28.

A	B	C	D	E	F
a	b	c	d	e	f
$a_2 a$	$b_2 b$	c	$d_2 d$	$e_2 e$	$f_2 f$
a	$b_2 b$	c	d	e	f
a	b	c	$d_2 d$	$e_2 e$	$f_2 f$

Etter at chasealgoritmen er ferdig, er de to første radene like på alle attributtene, så $C \rightarrow ABCDEF$. C er en kandidatnøkkel.

C er eneste kandidatnøkkel.

Hvordan avgjøre om en dekomposisjon er tapsfri når noen av reglene er MVDer

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} og en mengde MVDer \mathcal{M} . Gitt en dekomposisjon av \mathcal{R} til relasjonene S_1, \dots, S_k .

Chasealgoritme for å avgjøre om dekomposisjonen er tapsfri:

1. Lag en tabell med én kolonne for hvert attributt i \mathcal{R} og én rad for hver S_i .
2. I kolonnen for attributt A, for hver rad i:
 - Skriv a hvis A er et attributt i S_i
 - Skriv a_i hvis A ikke er et attributt i S_i
3. Så lenge det skjer forandringer i tabellen:
 - For hver FD $X \rightarrow Y$ i \mathcal{F} , for alle rader i tabellen med lik X-verdi, gjør Y-verdiene like. Hvis en av Y-ene er en verdi uten subskript, skal denne velges.
 - For hver MVD $X \twoheadrightarrow Y$ i \mathcal{M} , for alle par av rader med lik X-verdi, lag to nye rader der Y-verdiene er byttet om.
4. Hvis en rad er uten subskriptverdier, er dekomposisjonen tapsfri, ellers ikke.

Eksempel

$\mathcal{R}(A,B,C,D,E,F)$

$\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

$\mathcal{M} = \{A \rightarrow BC\}$

Finn ut om $\{ABDE, AC, BDE, ABEF\}$ er en tapsfri dekomposisjon.

	A	B	C	D	E	F
ABDE	a	b	c ₁	d	e	f ₁ f
AC	a	b ₂ b	c	d ₂ d	e ₂ e	f ₂ f
BDE	a ₃	b	c ₃	d	e	f ₃
ABEF	a	b	c ₄	d ₄ d	e	f
	a	b ₂ b	c	d	e	f
	a	b	c ₁	d ₂ d	e ₂ e	f ₂ f

Femte rad er uten subskriptverdier. Dekomposisjonen er tapsfri.

(Her har vi bare brukt MVDen på rad 1+2, men vi kunne ha brukt den på rad 1+4 og 2+4 også og fått ytterligere rader. Men siden vi finner en rad uten subskriptverdier bare ved å innføre to nye rader basert på rad 1+2, er ikke dette nødvendig.)

Hvordan dekomponere tapsfritt til 4NF

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} og en mengde MVDer \mathcal{M} .

Dekomponer i henhold til enten punkt 1 eller punkt 2 inntil det ikke lenger er noen brudd på BCNF og ikke lenger noen ekte MVDer i noen av relasjonene:

1. Hvis $X \rightarrow A$ er et brudd på BCNF i \mathcal{R} :
 - La Y være størst mulig slik at $X \rightarrow Y$ og Y ikke overlapper med X .
La Z være de attributtene i \mathcal{R} som ikke er i XY .[†]
 - Dekomponer \mathcal{R} til $S(XY)$ og $T(XZ)$.
2. Hvis $X \twoheadrightarrow Y$ er en ekte MVD i \mathcal{R} :
 - La Z være de attributtene i \mathcal{R} som ikke er i XY .
 - Dekomponer \mathcal{R} til $S(XY)$ og $T(XZ)$.

[†] Bruk chasealgoritmen på en tabell med to rader der den første raden er fullstendig uten subskriptverdier, og den andre er uten subskriptverdier i X , men med subskriptverdier i resten av attributtene. De av attributtene i andre rad som er uten subskriptverdier når chasealgoritmen avslutter, og som ikke er i X , er med i Y . Resten er i Z .

Eksempel 1

$$\mathcal{R}(A,B,C,D,E)$$

$$\mathcal{F} = \{C \rightarrow A\}$$

$$\mathcal{M} = \{E \twoheadrightarrow BC\}$$

Dekomponer \mathcal{R} tapsfritt til 4NF.

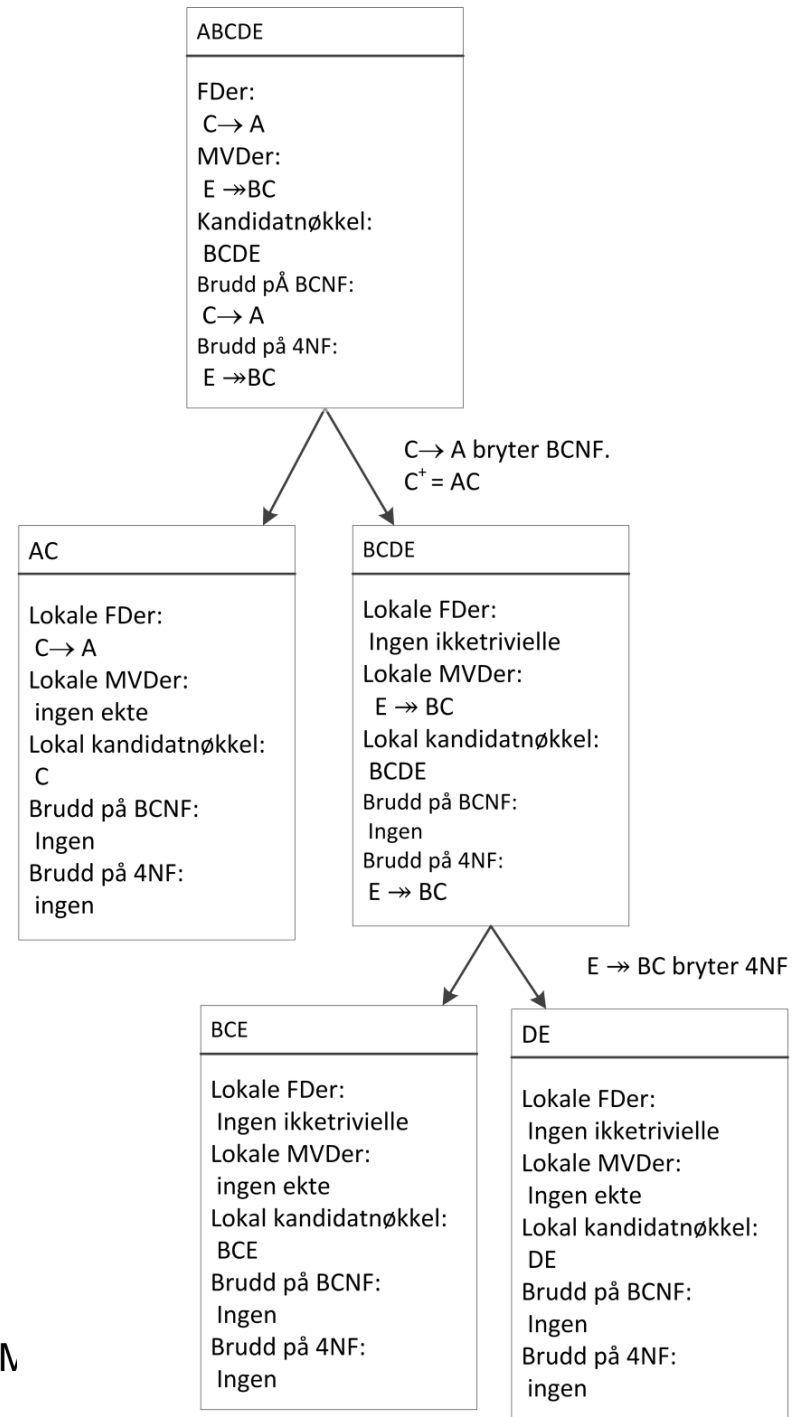
Kandidatnøkkel: BCDE.

$C \rightarrow A$ bryter BCNF.

MVDen $E \twoheadrightarrow BC$ er ekte.

1. Starter med FDer som bryter BCNF.
2. Deretter behandles ekte MVDer.

$\{AC, BCE, DE\}$ er en 4NF-dekomposisjon.



Eksempel 2

$\mathcal{R}(A,B,C,D,E)$

$\mathcal{F} = \{C \rightarrow A\}$

$\mathcal{M} = \{E \rightarrow BC\}$

(Samme eksempelrelasjon som forrige side.)

Dekomponer \mathcal{R} tapsfritt til 4NF.

Kandidatnøkkel: BCDE.

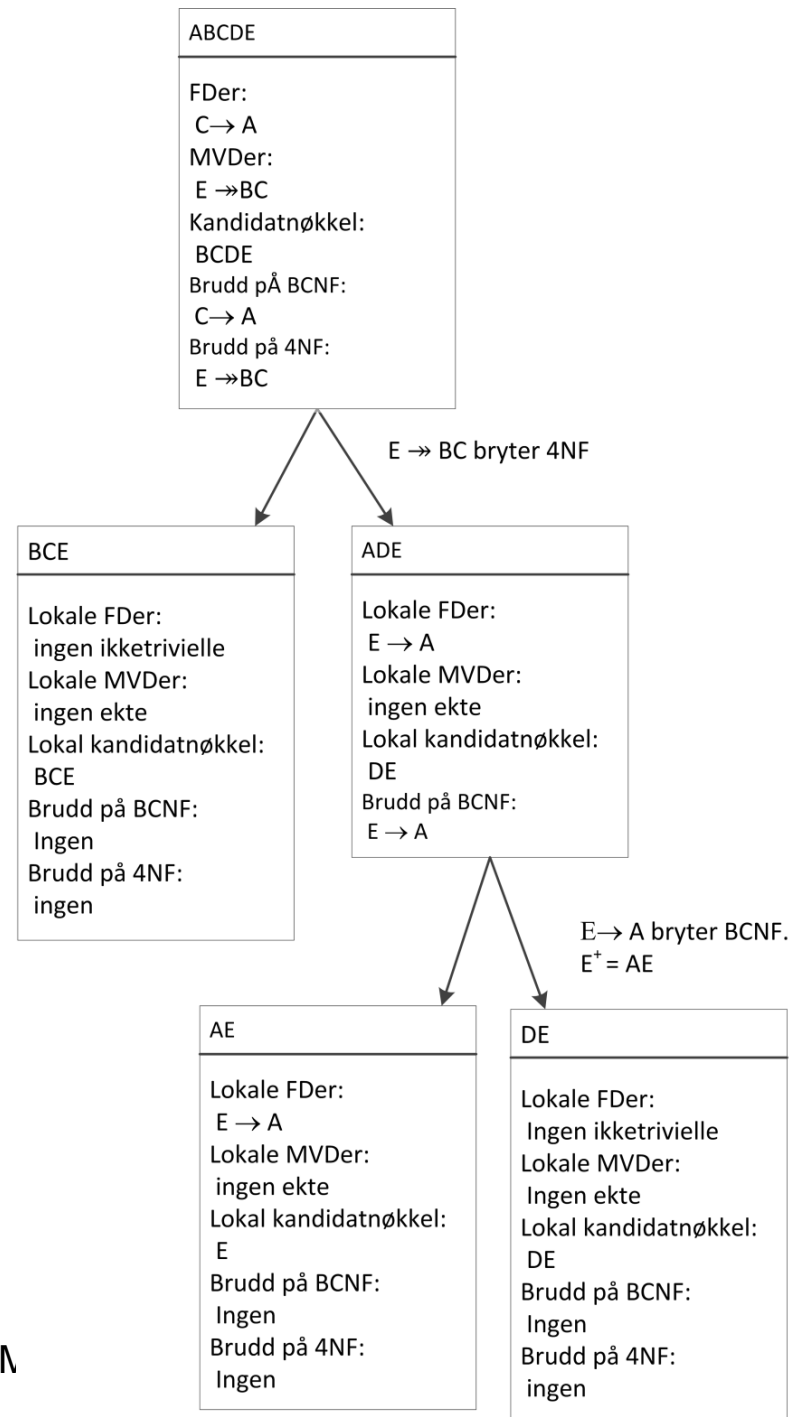
$C \rightarrow A$ bryter BCNF.

MVDen $E \rightarrow BC$ er ekte.

1. Starter med ekte MVDer.
2. Deretter tas FDer som bryter BCNF.

I tillegg til $C \rightarrow A$ har vi $E \rightarrow A$ (vises med chase); dette trenger vi underveis i dekomposisjonen.

$\{BCE, AE, DE\}$ er en 4NF-dekomposisjon.



Hvordan finne ut om en dekomposisjon kan ha støyinstanser

Gitt en relasjon \mathcal{R} med en mengde FDer \mathcal{F} . Gitt en tapsfri dekomposisjon av \mathcal{R} til relasjonene S_1, \dots, S_k .

1. Lag snittgrafene til dekomposisjonen:
 - a. Lag én node for hver S_i .
 - b. Lag en kant mellom to noder S_i og S_j hvis S_i og S_j overlapper, dvs. har felles attributter.
2. Hvis grafen ikke har sykler, har ikke dekomposisjonen støyinstanser. Hvis grafen har sykler, **kan det være** at dekomposisjonen har støyinstanser.

Om en dekomposisjon med syklisk snittgraf har støyinstanser eller ikke, avhenger til syvende og sist av hvordan de enkelte relasjonene i dekomposisjonen rent faktisk overlapper og hvilke FDer \mathcal{F} inneholder. Detaljene er ikke pensum.

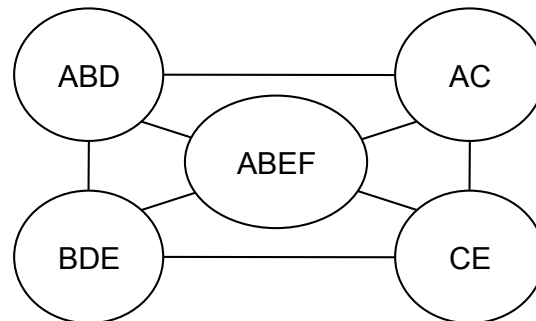
Eksempel

$\mathcal{R}(A,B,C,D,E,F)$, $\mathcal{F} = \{AB \rightarrow DE, C \rightarrow A, BD \rightarrow E, AE \rightarrow BF\}$

Tapsfri, FD-bevarende dekomposisjon (se side 19): $\{ABD, AC, BDE, ABEF, CE\}$

Finn ut om dekomposisjonen kan ha støyinstanser.

Lager snittgrafen:



Grafen har sykler, så det kan være at dekomposisjonen kan ha støyinstanser.

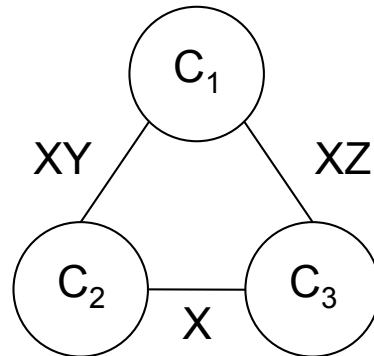
Se på følgende eksempel:

A	B	D	A	C	B	D	E	A	B	E	F	C	E
a_1	b_1	d_1	a_1	c_1	b_1	d_1	e_1	a_1	b_1	e_1	f_1	c_1	e_2
a_2	b_2	d_2	a_2	c_2	b_2	d_2	e_2	a_2	b_2	e_2	f_2	c_2	e_1

Instansene oppfyller alle FDene og «stemmer overens», dvs. $\pi_A(ABD) = \pi_A(AC) = \{a_1, a_2\}$, $\pi_{BD}(ABD) = \pi_{BD}(BDE) = \{b_1d_1, b_2d_2\}$ osv. Samtidig er $ABD \bowtie AC \bowtie BDE \bowtie ABEF \bowtie CE = \emptyset$, så dekomposisjonen har støyinstanser.

Sykliske snittgrafer uten støyinstanser I

- Hvis alle minimale sykler i snittgrafene er på formen under (der kantene er merket med de attributtene som nodene har felles), er dekomposisjonen uten støyinstanser.



- Dvs. alle minimale sykler har lengde 3, og i hver slik trekant finnes det en node C_1 som er slik at $C_2 \cap C_3 \subseteq C_1$ (der C_2 og C_3 er de andre nodene i trekanten).

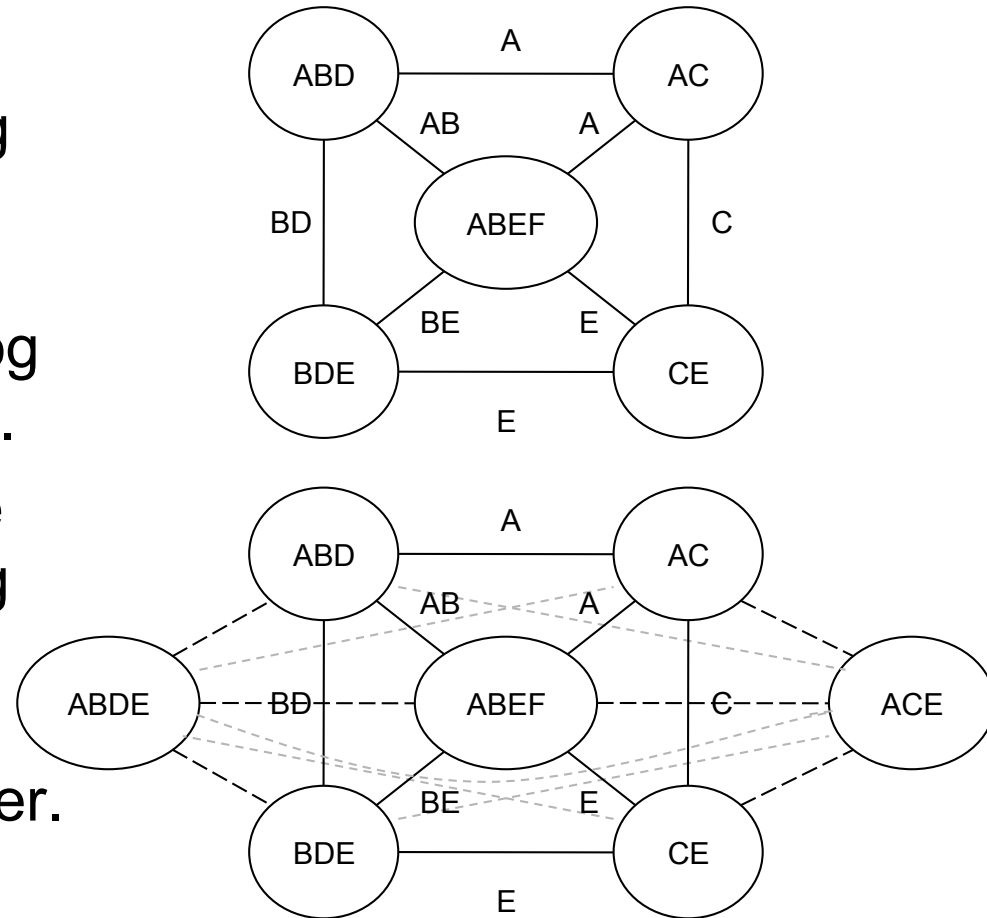
Sykliske snittgrafer uten støyinstanser II

- Hvis snittgrafene er syklisk og vi for hver minimal sykel C_1, \dots, C_n som ikke oppfyller kriteriet på forrige side, legger til en komponent C som inneholder attributtene i $C_i \cap C_{i+1}$ for $i=1, \dots, n$ (der vi med C_{n+1} mener C_1), er resultatet en tapsfri dekomposisjon uten støyinstanser.

Eksempel

Se på snittgrafene fra side 41.

- Syklene ABEF-ABD-AC og ABEF-CE-BDE oppfyller kriteriet på side 42, mens syklene ABEF-BDE-ABD og ABEF-AC-CE ikke gjør det.
- Vi legger til komponentene $ABDE = BE \cup BD \cup AB$ og $ACE = A \cup C \cup E$. Den resulterende dekomposisjonen er uten støyinstanser.



Relasjonsalgebra

- Notasjon (s. 46)
- Hvordan optimalisere relasjonsalgebrauttrykk (s. 47)
- Hvordan uttrykke integritetsregler i relasjonsalgebra (s. 48)

Notasjon

Bruk konsekvent **bag-tolkning** av alle operatorene under løsning av relasjonsalgebraoppgaver. Symboler:

- Bagunion: $R \cup S$
- Bagsnitt: $R \cap S$
- Bagdifferanse: $R - S$
- Kartesisk produkt: $R \times S$
- Seleksjon: $\sigma_C(R)$
- Projeksjon og utvidet projeksjon: $\pi_L(R)$
- Gruppering: $\gamma_L(R)$
- Naturlig join: $R \bowtie S$
- Theta-join: $R \bowtie_{\theta} S$
- Outerjoin: $R \bowtie^{\circ} S$, $R \bowtie_L S$, $R \bowtie_R S$
- Renavning: $\rho_S(R)$, $\rho_{S(A_1, A_2, \dots, A_n)}(R)$
- Duplikateliminasjon: $\delta(R)$
- Sortering: $\tau_L(R)$
- Divisjon: $R \text{ div } S$ eller R / S

Hvordan optimalisere relasjonsalgebrauttrykk

- De vanligste optimaliseringene er:
 - Dytt seleksjoner så langt ned som mulig.
Hvis seleksjonsbetingelsen består av flere deler, splitt i flere seleksjoner og dytt hver enkelt så langt ned i treet som mulig.
 - Dytt (legg til) projeksjoner så langt ned som mulig.
(Pass på at attributter som trengs lenger opp i treet, er inkludert i projeksjonsbetingelsene.)
 - Kombiner seleksjon og kartesisk produkt til en join av passende type
 - Duplikateliminasjoner kan fjernes hvis argumentet allerede er en mengde
- Men ikke ødelegg utnyttelse av indekser:
 - Dytting av projeksjon forbi en seleksjon kan ødelegge for bruk av indekser ved seleksjonen

Hvordan uttrykke integritetsregler i relasjonsalgebra

1. Hvis E er et relasjonsalgebrauttrykk, så er $E = \emptyset$ en integritetsregel som sier at E ikke har noen tupler.
2. Hvis E_1 og E_2 er to relasjonsalgebrauttrykk, så er $E_1 \subseteq E_2$ en integritetsregel som sier at ethvert tuppel i E_1 også skal være i E_2 .

Eksempler

$\mathcal{R}(A,B,C,D)$ med FDer $\{AB \rightarrow C, B \rightarrow AD\}$.

$\mathcal{S}(C,D,E,F)$ med FDer $\{CD \rightarrow E, E \rightarrow F\}$.

Det er en fremmednøkkel fra CD i \mathcal{R} til CD i \mathcal{S} .

1. Uttrykk fremmednøkkelen i relasjonsalgebra.

Idé: Mengden av tupler i CD i \mathcal{R} er en delmengde av tuplene i CD i \mathcal{S} .

Resultat (når vi bruker bagvarianten av relasjonsalgebraen):

$$\delta(\pi_{CD}(\mathcal{R})) \subseteq \pi_{CD}(\mathcal{S})$$

2. Uttrykk FDen $B \rightarrow AD$ i relasjonsalgebra.

Idé: Det skal ikke finnes to tupler i \mathcal{R} som er like i B, men ulike i A eller D.

Lag to kopier av \mathcal{R} og spleis alle tupler i den ene kopien med alle tupler i den andre. Ingen av de resulterende tuplene skal ha lik B og ulik A eller D.

Resultat: $\sigma_{T.B = U.B \text{ and } (T.A \neq U.A \text{ or } T.D \neq U.D)} (\rho_T(\mathcal{R}) \times \rho_U(\mathcal{R})) = \emptyset$