

# Plenum: Nøkler, normalformer og funksjonelle avhengigheter

Mathias Stang

14. november 2017

# Agenda

- Hva er god databasedesign?
- Atomære verdier
- Nøkler: Supernøkler, kandidatnøkler, primærnøkler, nøkkelattributter
- Funksjonelle avhengigheter (FD-er)
- Normalformer
  
- Mange eksempler og en eksamensoppgave

# God databasedesign: Hvorfor snakker vi om normalformer?

- Hver relasjon bør samle beslektet informasjon
- Vi ønsker så lite dobbeltlagring (redundans) som mulig
- Korrekt totalinformasjon kan gjenskapes nøyaktig ved join

# Atomære verdier: Adresse-eksempel

Gaustadalléen 23 B

0373 OSLO

Norge

# Normalisering: et eksempel



FulltNavn	Adresse	Fakturanr	Ordrenr	Fakturadato	Produkt1	Kost1	Antall1	Produkt2	Kost2	Antall2	Produkt3	Kost3
Ola Nordmann	Problemveien 11 0313 Oslo	3201	AB123	13/11/2017	Ting	249,-	2	Tang	1200,-	1	Greie	599,-
Kari Nilsen	Moltke Moes vei 35 0851 Oslo	3202	QZ93	13/11/2017	Ting	249,-	4					
Ole Olsen	Sem Sælands vei 7 0371 Oslo	3203	33AR	14/11/2017	Gjenstand	25,-	10	Ting	249,-	5		
Ola Nordmann	Problemveien 11 0313 Oslo	3204	AB130	14/11/2017	Tang	1200,-	1					
John Doe	Gaustadbekkalléen 23B 0372 Oslo	3205	656	14/11/2017	Ting	249,-	3					

<b>Fornavn</b>	<b>Etternavn</b>	<b>Adresse</b>	<b>Postnr</b>	<b>Poststed</b>	<b>Fakturanr</b>	<b>Ordrenr</b>	<b>Fakturadato</b>	<b>Produkt</b>	<b>Kost</b>	<b>Antall</b>
Ola	Nordmann	Problemveien 11	0313	Oslo	3201	AB123	13/11/2017	Ting	249,-	2
Ola	Nordmann	Problemveien 11	0313	Oslo	3201	AB123	13/11/2017	Tang	1200,-	1
Ola	Nordmann	Problemveien 11	0313	Oslo	3201	AB123	13/11/2017	Greie	599,-	1
Kari	Nilsen	Moltke Moes vei 35	0851	Oslo	3202	QZ93	13/11/2017	Ting	249,-	4
Ole	Olsen	Sem Sælands vei 7	0371	Oslo	3203	33AR	14/11/2017	Gjenstand	25,-	10
Ole	Olsen	Sem Sælands vei 7	0371	Oslo	3203	33AR	14/11/2017	Ting	249,-	5
Ola	Nordmann	Problemveien 11	0313	Oslo	3204	AB130	14/11/2017	Tang	1200,-	1
John	Doe	Gaustadbekalléen 23B	0372	Oslo	3205	656	14/11/2017	Ting	249,-	3

KundeID	Fornavn	Etternavn	Adresse	Postnr	Poststed	Fakturanr	Ordrenr	Fakturadato	ProduktID	Produkt	Kost	Antall
1	Ola	Nordmann	Problemveien 11	0313	Oslo	3201	AB123	13/11/2017	1	Ting	249,-	2
1	Ola	Nordmann	Problemveien 11	0313	Oslo	3201	AB123	13/11/2017	2	Tang	1200,-	1
1	Ola	Nordmann	Problemveien 11	0313	Oslo	3201	AB123	13/11/2017	3	Greie	599,-	1
2	Kari	Nilsen	Moltke Moes vei 35	0851	Oslo	3202	QZ93	13/11/2017	1	Ting	249,-	4
3	Ole	Olsen	Sem Sælands vei 7	0371	Oslo	3203	33AR	14/11/2017	4	Gjenstand	25,-	10
3	Ole	Olsen	Sem Sælands vei 7	0371	Oslo	3203	33AR	14/11/2017	1	Ting	249,-	5
1	Ola	Nordmann	Problemveien 11	0313	Oslo	3204	AB130	14/11/2017	2	Tang	1200,-	1
4	John	Doe	Gaustadbekkealléen 23B	0372	Oslo	3205	656	14/11/2017	1	Ting	249,-	3

KundeID	Fakturanr	Ordrenr	Fakturadato	ProduktID	Antall
1	3201	AB123	13/11/2017	1	2
1	3201	AB123	13/11/2017	2	1
1	3201	AB123	13/11/2017	3	1
2	3202	QZ93	13/11/2017	1	4
3	3203	33AR	14/11/2017	4	10
3	3203	33AR	14/11/2017	1	5
1	3204	AB130	14/11/2017	2	1
4	3205	656	14/11/2017	1	3



KundeID	Fakturanr	ProduktID	Antall
1	3201	1	2
1	3201	2	1
1	3201	3	1
2	3202	1	4
3	3203	4	10
3	3203	1	5
1	3204	2	1
4	3205	1	3

# Nøkler

# Nøkler

- X er en **supernøkkel** i R hvis  $X \subseteq R$ , og ingen instans av R får inneholde to forskjellige tupler  $t_1$  og  $t_2$  hvor  $t_1[X] = t_2[X]$
- X er en **kandidatnøkkel** i R hvis X er en supernøkkel i R, og for alle A i X er X-A ikke en supernøkkel i R (dvs. X er en minimal supernøkkel)
- En **primærnøkkel** X er en spesielt utpekt kandidatnøkkel i R

## Eksempel:

Filmgenre(filmid, genre, title)

filmid	genre	title
85908	Action	The Matrix
85908	Sci-Fi	The Matrix
85908	Thriller	The Matrix
26103	Action	Planet of the Apes
26103	Sci-Fi	Planet of the Apes
1320611	Action	Planet of the Apes
1320611	Sci-Fi	Planet of the Apes

**Supernøkler:** {filmid, genre},  
{filmid, genre, title}

**Kandidatnøkkel:** {filmid, genre}

**Primærnøkkel:** {filmid, genre}

# Nøkkelattributt

- Et **nøkkelattributt** er et attributt som er med i en kandidatnøkkel
- Et ikke-nøkkelattributt er et attributt som *ikke* er med i noen kandidatnøkkel

## Eksempel:

Filmgenre(filmid, genre, title)

filmid	genre	title
85908	Action	The Matrix
85908	Sci-Fi	The Matrix
85908	Thriller	The Matrix
26103	Action	Planet of the Apes
26103	Sci-Fi	Planet of the Apes
1320611	Action	Planet of the Apes
1320611	Sci-Fi	Planet of the Apes

**Kandidatnøkkel:** {filmid, genre}

**Nøkkelattributter:** filmid, genre

**Ikke-nøkkelattributt:** title

## Eksempel 2

Student(fnr, id, navn, adresse)

Primærnøkkel:

Kandidatnøkler:

Supernøkler:

Nøkkelattributter:

Ikke-nøkkelattributter:

# Funksjonelle avhengigheter

# Funksjonelle avhengigheter

- Y er **funksjonelt avhengig** av X hvis vi for enhver lovlig instans av R har at hvis instansen inneholder to tupler  $t_1$  og  $t_2$  hvor  $t_1[X] = t_2[X]$ , så må  $t_1[Y] = t_2[Y]$ 
  - Da skriver vi  $X \rightarrow Y$
- Omtales også som **FD-er** (functional dependencies)
- Vi sier at «Y følger av X» eller «X bestemmer Y»

## Eksempel:

Person(PID, Navn, Postnr, Poststed)

PID	Navn	Postnr	Poststed
1	Ola	0372	OSLO
2	Kari	5006	BERGEN
3	Per	0372	OSLO
4	Ola	1383	ASKER
5	Jo	1384	ASKER
6	Nils	0372	OSLO
7	Stein	0010	OSLO

Funksjonelle avhengigheter:

- Postnr  $\rightarrow$  Poststed

# Funksjonelle avhengigheter

- Merk: Hvis  $X$  er en supernøkkel, så holder  $X \rightarrow Y$  for enhver  $Y$ 
  - Altså: Hvis  $X$  er en primærnøkkel/kandidatnøkkel, holder  $X \rightarrow Y$  for alle  $Y$
- Og motsatt: Hvis  $X \rightarrow Y$  for enhver  $Y$ , så er  $X$  en kandidatnøkkel
- FD-en  $X \rightarrow Y_1, Y_2, \dots, Y_n$  kan også representeres som FD-ene  
 $X \rightarrow Y_1, X \rightarrow Y_2, \dots, X \rightarrow Y_n$   
(slik at høyresidene består bare av ett attributt)

## Eksempel:

Person(PID, Navn, Postnr, Poststed)

PID	Navn	Postnr	Poststed
1	Ola	0372	OSLO
2	Kari	5006	BERGEN
3	Per	0372	OSLO
4	Ola	1383	ASKER
5	Jo	1384	ASKER
6	Nils	0372	OSLO
7	Stein	0010	OSLO

Funksjonelle avhengigheter:

- Postnr  $\rightarrow$  Poststed
- PID  $\rightarrow$  Navn, Postnr, Poststed

*Alternativt skriver vi:*

- Postnr  $\rightarrow$  Poststed
- PID  $\rightarrow$  Navn
- PID  $\rightarrow$  Postnr
- PID  $\rightarrow$  Poststed



# Funksjonelle avhengigheter oppsummert

*Enkel definisjon:* Et attributt er «avhengig» av / bestemmes av et annet attributt

Eksempel: Student(id, brukernavn, navn, adresse, postnr, poststed)

Funksjonelle avhengigheter:

- id → brukernavn, navn, adresse, postnr, poststed
- brukernavn → id, navn, adresse, postnr, poststed

*Ny integritetsregel: Et postnr bestemmer et poststed*

- postnr → poststed

# FD-er: Enda et eksempel

Ordre(ordrenr, kundenr, kundenavn, antall, sum, mva)

I tillegg har vi følgende integritetsregler:

- Ordrenr er unikt

ordrenr  $\rightarrow$  . \ - -

- Kundenr bestemmer kundenavn

kundenr  $\leftrightarrow$  kundenavn

- Mva-verdi følger av sum

sum  $\rightarrow$  mva

# Normalformer

# Normalformer

Så langt har vi repetert:

- Atomære verdier
- Nøkler
  - Supernøkler
  - Kandidatnøkler (og primærnøkler)
  - Nøkkelattributt
- Funksjonelle avhengigheter

# Normalformer

Gitt en relasjon R, med et sett FD-er på formen  $X \rightarrow A$ , der X og A er et sett av attributter

## 1NF:

- R inneholder kun atomære verdier/attributter

## 2NF:

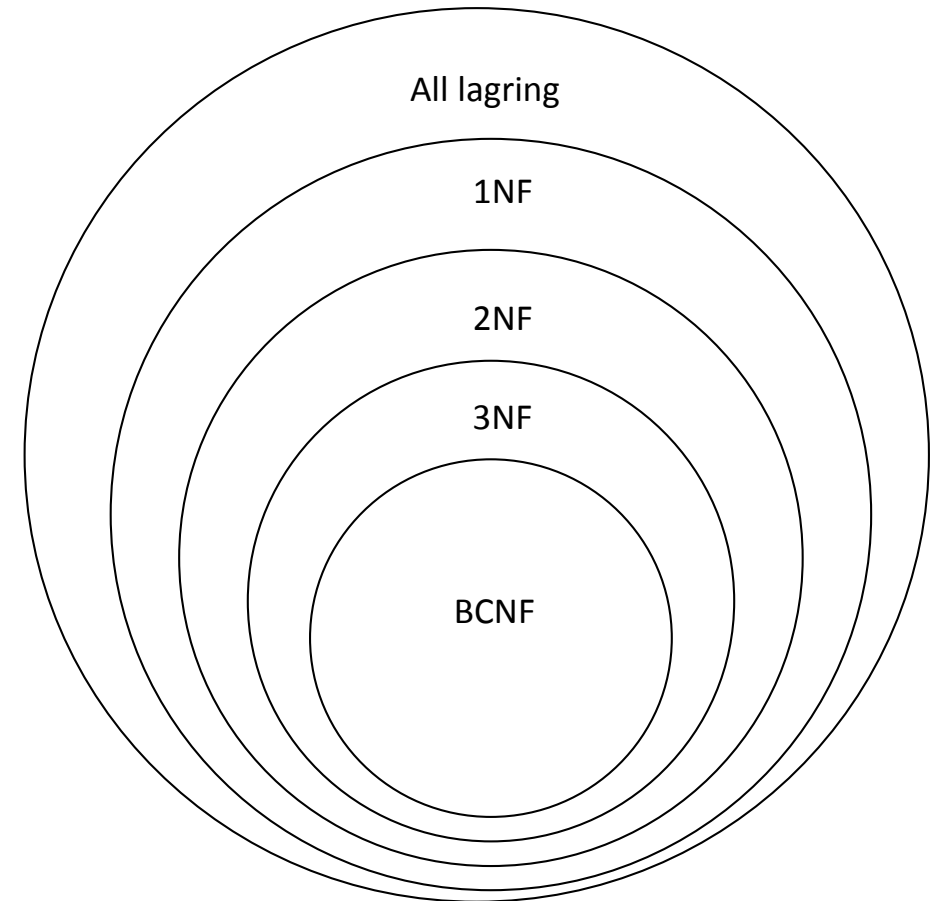
- X er en supernøkkel i R, eller
- A er et nøkkelattributt, eller
- X er *ikke* en delmengde av noen nøkler i R

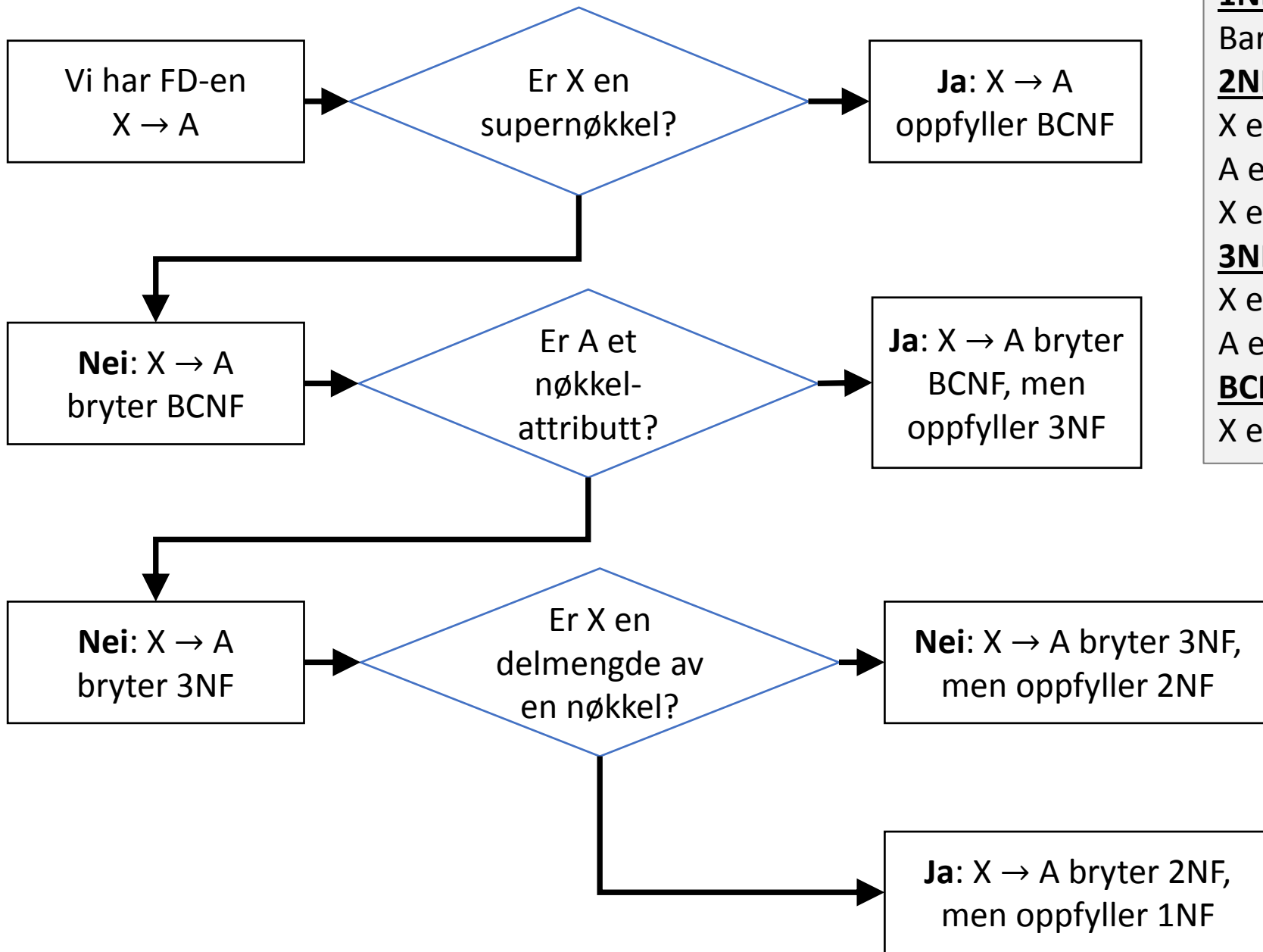
## 3NF:

- X er en supernøkkel i R, eller
- A er et nøkkelattributt

## BCNF:

- X er en supernøkkel i R





**1NF:**

Bare atomære verdier/attributter

**2NF:**

X er en supernøkkel i R, eller A er et nøkkelattributt, eller X er *ikke* en delmengde av noen nøkler i R

**3NF:**

X er en supernøkkel i R, eller A er et nøkkelattributt

**BCNF:**

X er en supernøkkel i R

# Eksempel 1

Timeliste(ansattnr, uke, år, navn, timer)

FD-er:

- ansattnr, uke, år  $\rightarrow$  navn, timer BCNF
- ansattnr  $\rightarrow$  navn 1NF

Hvilken normalform er relasjonen på? 1NF

## 1NF:

Bare atomære verdier/attributter

## 2NF:

X er en supernøkkel i R, eller

A er et nøkkelattributt, eller

X er *ikke* en delmengde av noen nøkler i R

## 3NF:

X er en supernøkkel i R, eller

A er et nøkkelattributt

## BCNF:

X er en supernøkkel i R

# Eksempel 2

Ordre(ordrenr, kundenr, kundenavn, antall, sum, mva)

FD-er:

- ordrenr → kundenr, kundenavn, antall, sum, mva    BCNF
- kundenr → navn    2NF
- sum → mva    2NF

Hvilken normalform er relasjonen på?    2NF

## 1NF:

Bare atomære verdier/attributter

## 2NF:

X er en supernøkkel i R, eller

A er et nøkkelattributt, eller

X er *ikke* en delmengde av noen nøkler i R

## 3NF:

X er en supernøkkel i R, eller

A er et nøkkelattributt

## BCNF:

X er en supernøkkel i R



# Eksempel 3

Student\_emne(id, emnekode, bnavn, karakter)

FD-er:

- id, emnekode → karakter BCNF
- bnavn, emnekode → karakter BCNF
- id → bnavn 3NF
- bnavn → id 3NF

Hvilken normalform er relasjonen på?

3NF

**1NF:**

Bare atomære verdier/attributter

**2NF:**

X er en supernøkkel i R, eller

A er et nøkkelattributt, eller

X er *ikke* en delmengde av noen nøkler i R

**3NF:**

X er en supernøkkel i R, eller

A er et nøkkelattributt

**BCNF:**

X er en supernøkkel i R

$X \rightarrow A$

# Eksempel 4

Student(fnr, id, navn, adresse)

FD-er:

- fnr  $\rightarrow$  id, navn, adresse      BCNF
- id  $\rightarrow$  fnr, navn, adresse      BCNF

Hvilken normalform er relasjonen på? BCNF

**1NF:**

Bare atomære verdier/attributter

**2NF:**

X er en supernøkkel i R, eller

A er et nøkkelattributt, eller

X er *ikke* en delmengde av noen nøkler i R

**3NF:**

X er en supernøkkel i R, eller

A er et nøkkelattributt

**BCNF:**

X er en supernøkkel i R

# Eksamensoppgave (2015)

I denne oppgaven skal vi bruke følgende relasjon:

Filmgenre (filmid, title, prodyear, genre)

Primærnøkkelen i tabellen er kombinasjonen av filmid og genre (filmid, genre). Videre vet vi også at filmid bestemmer tittel og produksjonsår for en film.

- Bestem alle supernøkklene i relasjonen Filmgenre. Skriv ned alle.
- Bestem alle FDene i relasjonen Filmgenre.
- Hvilken normalform er relasjonen Filmgenre på? Begrunn svaret ditt.

a :  $\{ \text{filmid}, \text{genre} \}, \{ f, g, \text{title} \}, \{ f, g, \text{prodyear} \}, \{ f, g, \text{title}, \text{prodyear} \}$

b :  $\text{filmid}, \text{genre} \rightarrow \text{title}$  BCNF

$\text{filmid}, \text{genre} \rightarrow \text{prodyear}$  BCNF

$\text{filmid} \rightarrow \text{prodyear}$  INF

$\text{filmid} \rightarrow \text{title}$  INF

c : INF

## 1NF:

Bare atomære verdier/attributter

## 2NF:

X er en supernøkkel i R, eller

A er et nøkkelattributt, eller

X er *ikke* en delmengde av noen nøkler i R

## 3NF:

X er en supernøkkel i R, eller

A er et nøkkelattributt

## BCNF:

X er en supernøkkel i R

# Hva gjør vi videre?

- Dekomposisjon til høyere normalform

Eksempel: Fra 3NF til BCNF

Student\_emne(id, emnekode, bnavn, karakter)

Student(bnavn, id)

StudentEmne(bnavn, emnekode, karakter)

Ikke pensum i INF1300, men om du vil lære mer om relasjonsdatabaseteori bør du ta **INF3100 – Databasesystemer** til våren!