



UNIVERSITETET  
I OSLO

# Relasjonsmodellen

## Relasjoner og funksjonelle avhengigheter

# Relasjonsdatabasemodellen

- **Datamodell**

Mengde av begreper for å beskrive strukturen til en database

- **Relasjonsmodellen**

Databasen kan betraktes som en samling av tabeller

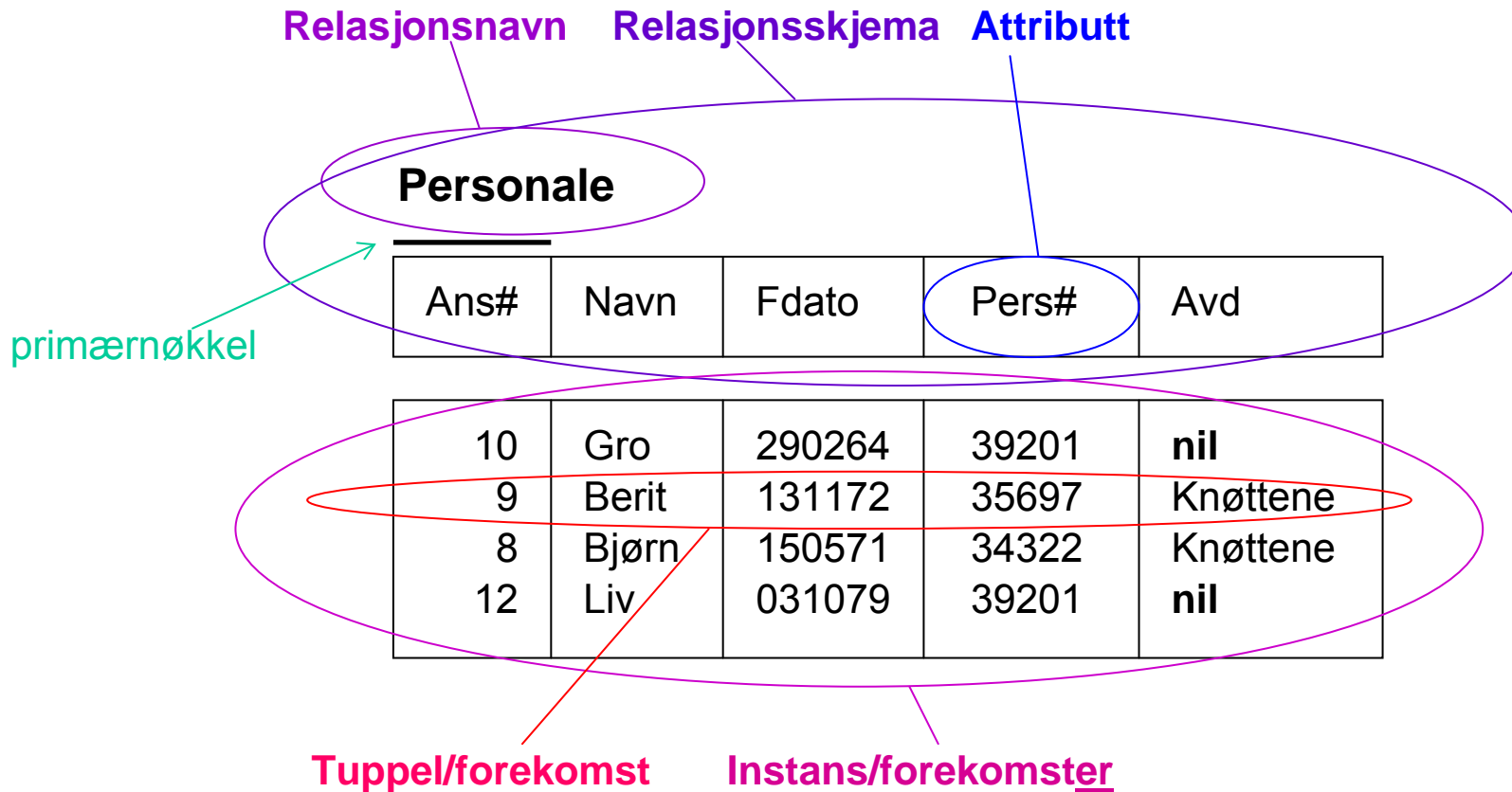
# Relasjoner og relasjonsdatabaser

## Personale

Ans#	Navn	Fdato	Pers#	Avd
10	Gro	290264	39201	nil
9	Berit	131172	35697	Knøttene
8	Bjørn	150571	34322	Knøttene
12	Liv	031079	39201	nil

- **Relasjon**: Et matematisk begrep som kan tolkes som en tabell med verdier. Presist: En mengde av **tupler**
- **Relasjonsdatabase**: En samling relasjoner

# Relasjoner - terminologi



# Formelle definisjoner

- **Domene:** En mengde *atomære* verdier
- **Attributt:** Et navn på en rolle spilt av et domene («*kolonnenavn*»)
- **Relasjonsskjema**  $R(A_1, A_2, \dots, A_n)$ : En navngitt mengde attributter  $R = \{A_1, A_2, \dots, A_n\}$  der  $R$  er **relasjonsnavnet**  $n$  kalles relasjonens *grad* eller *aritet*
- **Instans** av et relasjonsskjema  $R(A_1, A_2, \dots, A_n)$ : En mengde  $\{t_1, t_2, \dots, t_m\}$  der hver  $t_k$  er et  $n$ -tupple av verdier fra domene til  $A_1, A_2, \dots, A_n$  (noen kan være **nil**)
- **Relasjon:** Et relasjonsskjema med en tilhørende instans  
Relasjonsskjemaet kalles relasjonens **intensjon**  
Instansen kalles relasjonens **ekstensjon**

# Merk:

- Tuplens rekkefølge i en instans er vilkårlig
- Verdiens rekkefølge i et tuppel er vilkårlig
- I en instans kan det ikke finnes to like tupler
- Hver av verdiene i et tuppel er hentet fra et domene eller er **nil**
- Et domene kan være endelig eller uendelig
- To attributter i et relasjonsskjema kan ha samme domene, men ikke samme navn

# Relasjonsdatabaser - definisjoner

- **Relasjonsdatabaseskjema:**  
Samling av relasjonsskjemaer + integritetsregler
- **Relasjonsdatabaseinstans:**  
Samling av relasjonsinstanser
- **Relasjonsdatabase** =  
relasjonsdatabaseskjema +  
relasjonsdatabaseinstans

# Nøkler og nøkkelattributter



# Nøkler og nøkkelattributter

## Personale

Ans#	Navn	Fdato	Pers#	Avd
10	Gro	290264	39201	nil
9	Berit	131172	35697	Knøttene
8	Bjørn	150571	34322	Knøttene
12	Liv	031079	39201	nil

- Vi ønsker ikke at to ansatte skal kunne ha samme Ans#
- To personer kan aldri ha samme fødselsnummer = Fdato + Pers#

# Definisjon av nøkler

Gitt et skjema  $R(A_1, A_2, \dots, A_n)$  med tilhørende integritetsregler

La  $X$  være en delmengde av  $\{A_1, A_2, \dots, A_n\}$

Hvis  $t$  er et tuppel i en instans av  $R$ ,  
betegner  $t[X]$  verdiene i  $t$ 's  $X$ -attributter

- **Supernøkkel**: En delmengde  $X$  av  $\{A_1, A_2, \dots, A_n\}$  som er slik at hvis  $t$  og  $u$  er to tupler hvor  $t \neq u$ , så er  $t[X] \neq u[X]$
- **Kandidatnøkkel**: En minimal supernøkkel  
Dvs: Fjerning av hvilket som helst attributt fører til at de gjenværende attributtene ikke lenger utgjør en supernøkkel
- **Primærnøkkel**: En utvalgt blant kandidatnøkklene  
Alle relasjoner skal ha nøyaktig én primærnøkkel
- **Nøkkelattributt** (*prime attribute*): Attributt som er med i (minst) en kandidatnøkkel

Supernøkler benyttes til å uttrykke integritetsregler

# Fremmednøkler

## Barn

Løpe#	Navn	Fdato	Avd	TilknPers
2	Lisa	180502	Rosa Pantern	nil
5	Trym	030205	Knøttene	9
4	Anne	301102	Tommeliten	nil
7	Anne	151204	Knøttene	8

- Vi vil at TilknPers skal referere til forekomster i **Personale-**tabellen

# Fremmednøkler

- **Fremmednøkkel:**  
Ett eller flere attributter som peker ut/refererer primærnøkkelen i en annen relasjon

## Personale

Ans#	Navn	Fdato	Pers#	Avd
10	Gro	290264	39201	nil
9	Berit	131172	35697	Knøttene
8	Bjørn	150571	34322	Knøttene
12	Liv	031079	39201	nil

## Barn

Løpe#	Navn	Fdato	Avd	TilknPers
2	Lisa	180502	Rosa Pantern	nil
5	Trym	030205	Knøttene	9
4	Anne	301102	Tommeliten	nil
7	Anne	151204	Knøttene	8

# Fremmednøkler

- Fremmednøkkelen må ha samme antall attributter som primærnøkkelen i den relasjonen den peker ut, og attributtene må ha parvis samme domener
- Korresponderende attributter behøver ikke å ha samme navn
- Det er lov å ha fremmednøkler til «seg selv»
- Fremmednøkler benyttes til å uttrykke integritetsregler

# Påkrevde integritetsregler i relasjonsdatabaser

- **Entitetsintegritet:**  
Alle relasjonsskjemaer skal ha en og bare en primærnøkkel  
Ingen av attributtene i primærnøkkelene får være **nil**
- **Referanseintegritet:**  
Hvis fremmednøkkelene ikke er **nil**, så skal det finnes et tuppel i den refererte relasjonen hvor primærnøkkelene har samme verdi som fremmednøkkelene (dvs. at det refererte tuppelet skal eksistere)
- **Domeneintegritet:**  
Alle verdier skal være atomære og hentet fra vedkommende attributts domene  
(Dessuten kan **nil** være tillatt «verdi» for noen attributter)
- I tillegg kan databasen ha andre integritetsregler, for eksempel kandidatnøkler som ikke er primærnøkler

# Funksjonelle avhengigheter

# Formell definisjon av funksjonell avhengighet

- Gitt et relasjonsskjema  $R(A_1, A_2, \dots, A_n)$  og la  $X, Y$  være delmengder av  $\{A_1, A_2, \dots, A_n\}$
- $Y$  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  $t_1[Y] = t_2[Y]$
- I så fall skriver vi  $X \rightarrow Y$



# Funksjonelle avhengigheter (forts.)

- Ofte snakker vi for korthets skyld om «FDen  $X \rightarrow Y$ »  
(der **FD** står for Functional Dependency)
- Vi sier at «Y følger av X»,  
eller at «X bestemmer Y»
- Merk at hvis X er en supernøkkel,  
så holder  $X \rightarrow Y$  for enhver Y
- Omvendt: Hvis  $X \rightarrow Y$  for enhver Y,  
så er X en supernøkkel
- FDer er integritetsregler

# Ekvivalente mengder av FDer

La  $S$ ,  $T$  være to mengder av FDer

- **Definisjon:**

Vi sier at  **$S$  følger av  $T$**  hvis det er slik at enhver instans som oppfyller alle FDer i  $T$ , også oppfyller alle FDer i  $S$

- **Definisjon:**

Vi sier at  **$S$  og  $T$  er ekvivalente** hvis  $S$  følger av  $T$  og  $T$  følger av  $S$

# Armstrongs slutningsregler

1. **Refleksivitet:**  
Hvis  $Y$  er en delmengde av  $X$ , så  $X \rightarrow Y$
2. **Utvidelse:** Hvis  $X \rightarrow Y$ , så  $XZ \rightarrow YZ$
3. **Transitivitet:** Hvis  $X \rightarrow Y$  og  $Y \rightarrow Z$ , så  $X \rightarrow Z$

Regelsettet er

- **Sunt:** Vi kan ikke utlede selvmotsigelser
- **Komplett:** Alle FDer som kan vises å følge fra en mengde FDer ved bruk av definisjonen av FD, kan også vises ved å bare bruke slutningsreglene

# Trivielle FD-er

- En FD som følger av refleksivitetsregelen  
«Hvis  $Y$  er en delmengde av  $X$ , så  $X \rightarrow Y$ »  
kalles *triviell* fordi den er automatisk oppfylt
- En FD  $X \rightarrow Y$  hvor  $Y - X \neq \emptyset$ , kalles *ikke-triviell*

# Bevis for utvidelsesregelen:

Hvis  $X \rightarrow Y$ , så  $XZ \rightarrow YZ$

- Anta at  $X \rightarrow Y$ , og at det finnes to tupler  $t$  og  $u$  slik at  $t[XZ] = u[XZ]$
- Da har vi spesielt at  $t[X] = u[X]$  og at  $t[Z] = u[Z]$
- Siden  $X \rightarrow Y$ , og  $t[X] = u[X]$ , er  $t[Y] = u[Y]$
- Da har vi både at  $t[Y] = u[Y]$ , og at  $t[Z] = u[Z]$
- Sammen gir det at  $t[YZ] = u[YZ]$
- Dermed har vi bevist at  $XZ \rightarrow YZ$  holder

# Bevis for den transitive regelen:

Hvis  $X \rightarrow Y$  og  $Y \rightarrow Z$ , så  $X \rightarrow Z$

- Anta at  $X \rightarrow Y$ , at  $Y \rightarrow Z$ , og at det finnes to tupler  $t$  og  $u$  slik at  $t[X] = u[X]$
- Siden  $X \rightarrow Y$ , og  $t[X] = u[X]$ , er  $t[Y] = u[Y]$
- Siden  $Y \rightarrow Z$ , og  $t[Y] = u[Y]$ , er  $t[Z] = u[Z]$
- Dermed har vi bevist at  $X \rightarrow Z$  holder

# Tillukningen av $X$ mhp $F$

- **Definisjon** av **tillukningen av  $X$  mhp  $F$** :  
La  $X$  være en mengde attributter og  $F$  en mengde FDer  
Da er tillukningen av  $X$  med hensyn på  $F$  lik mengden av attributter som er bestemt av  $X$
- **Notasjon**:  $X^+$  (mhp  $F$ )

# Tillukningsalgoritmen

Algoritme som beregner  $X^+$  mhp. F:

1.  $T=X$
2. Sålenge T forandres:
  1. Søk etter  $Y \rightarrow Z$  i F hvor Y er en delmengde av T
  2. Hvis slik Y finnes: Legg Z inn i T  
(Dvs. Sett  $T = T \cup Z$ )
3.  $X^+=T$



# Test som avgjør om en FD $X \rightarrow Y$ følger fra en mengde FDer $F$

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

Hvis ikke, følger ikke  $X \rightarrow Y$  av  $F$

# Hvordan finne alle kandidatnøkler (minimale supernøkler)

Gitt et skjema  $R(A_1, A_2, \dots, A_n)$  og en mengde FDer  $F$

1. La  $S = \{A_1A_2 \dots A_n\}$  ( $A_1A_2 \dots A_n$  er en supernøkkel)  
La  $K = \{\}$
2. For hver supernøkkel  $X$  i  $S$ :
  - For hvert attributt  $A$  i  $X$ : Sett  $Y = X - \{A\}$  og beregn  $Y^+$  mhp.  $F$   
Hvis  $Y^+ = A_1A_2 \dots A_n$ , er  $X$  ikke en minimal supernøkkel;  
 $Y$  er også en supernøkkel, så legg  $Y$  i  $S$
  - Hvis  $Y^+ \neq A_1A_2 \dots A_n$  for alle slike  $Y$ , er  $X$  en minimal supernøkkel  
I så fall, legg  $X$  i  $K$
  - Fjern  $X$  fra  $S$
3. Nå er  $K$  mengden av alle kandidatnøkler mhp.  $F$

Algoritmen er ikke særlig effektiv, men den virker

# Hvordan effektivt finne alle kandidatnøkler

Eksempel:

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

- Siden C ikke er med i noen høyreside i F, må C være med i alle supernøkler (spesielt i alle kandidatnøkler)
- $C^+ = AC$ , så C er ikke en supernøkkel
- $BC^+ = ABCDE$ , så BC er en kandidatnøkkel
- $CD^+ = ACD$ , så CD er ikke en supernøkkel
- $CE^+ = ABCDE$ , så CE er en kandidatnøkkel
- For å utvide CD til en nøkkel, har vi to muligheter
  - BCD som inneholder kandidatnøkkelen BC
  - CDE som inneholder kandidatnøkkelen CE
- Altså er BC og CE de eneste kandidatnøkklene i R