

IN2090 – Databaser og datamodellering

08 – Design i praksis

Leif Harald Karlsen
leifhka@ifi.uio.no



Universitetet i Oslo

Hvordan designe til ca. BCNF?

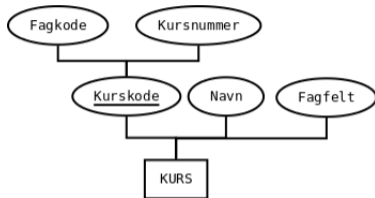
- ◆ Én tabell per entitetstype, altså ett tuppel = en entitet
- ◆ Relasjoner mellom entiteter representeres enten
 - ◆ via fremmenøkler fra en entitets-tabell til en annen (en-til-en og en-til-mange)
 - ◆ eller via egne tabeller (mange-til-mange)

ER-realisering og normalformer

- ◆ Prinsippene fra forrige slide minner veldig om realiseringsalgoritmen for ER
- ◆ Den algoritmen er laget slik at den (så langt det lar seg gjøre) gir et BCNF-skjema som resultat
- ◆ F.eks.:
 - ◆ Hver entitet blir én tabell hvor nøklene bestemmer alle andre attributter
 - ◆ Relasjoner blir kun del av en entitets tabell dersom kolonnen er bestemt av nøkkelen
 - ◆ Flerverdi-attributter blir egne tabeller
 - ◆ Utledbare attributter blir ikke del av realiseringen
- ◆ Men, merk at vi ikke er garantert BCNF, eller noe annet høyere enn 1NF!
- ◆ ER er ikke uttrykingskraftig nok til å uttrykke alle FDer

ER-realisering til 1NF – Eksempel

- ◆ ER-modell:



- ◆ FDer:

- ◆ $Fagkode, Kursnummer \rightarrow Navn$
- ◆ $Fagkode \rightarrow Fagfelt$

← Kan ikke uttrykkes i ER!

- ◆ Realiseres til:

$Kurs(\underline{Fagkode}, Kursnummer, Navn, Fagfelt)$

- ◆ $Fagkode \rightarrow Fagfelt$ bryter med 2NF fordi $Fagkode$ kun er en del av kandidatnøkkelen.

Når må man dekomponere?

- ◆ Etter ER-realisering i noen tilfeller
- ◆ Overtar dårlig designet database
- ◆ Databaser utvikler seg over tid
- ◆ Virkeligheten kan endre seg
- ◆ Migrere data fra f.eks. regneark til relasjonsskjema

sektor	art	år	statistikkvariabel	10721: Offentlig forvaltning, Inntekter og utgifter (mill. kr)
6100	Statsforvaltningen	A Totale inntekter	2000 Inntekter og utgifter	716095
6100	Statsforvaltningen	A Totale inntekter	2001 Inntekter og utgifter	728298
6100	Statsforvaltningen	A Totale inntekter	2002 Inntekter og utgifter	732725
6100	Statsforvaltningen	A Totale inntekter	2003 Inntekter og utgifter	745081
6100	Statsforvaltningen	A Totale inntekter	2004 Inntekter og utgifter	842956
6100	Statsforvaltningen	A Totale inntekter	2005 Inntekter og utgifter	965980
6100	Statsforvaltningen	A Totale inntekter	2006 Inntekter og utgifter	1125908
6100	Statsforvaltningen	A Totale inntekter	2007 Inntekter og utgifter	1182816
6100	Statsforvaltningen	A Totale inntekter	2008 Inntekter og utgifter	1336544
6100	Statsforvaltningen	A Totale inntekter	2009 Inntekter og utgifter	1157152
6100	Statsforvaltningen	A Totale inntekter	2010 Inntekter og utgifter	1223062
6100	Statsforvaltningen	A Totale inntekter	2011 Inntekter og utgifter	1367004
6100	Statsforvaltningen	A Totale inntekter	2012 Inntekter og utgifter	1444387
6100	Statsforvaltningen	A Totale inntekter	2013 Inntekter og utgifter	1430247
6100	Statsforvaltningen	A Totale inntekter	2014 Inntekter og utgifter	1450038
6100	Statsforvaltningen	A Totale inntekter	2015 Inntekter og utgifter	1420699
6100	Statsforvaltningen	A Totale inntekter	2016 Inntekter og utgifter	1407886
6100	Statsforvaltningen	A Totale inntekter	2017 Inntekter og utgifter	1500589
6100	Statsforvaltningen	A Totale inntekter	2018 Inntekter og utgifter	1666401
6100	Statsforvaltningen	A Totale inntekter	2019 Inntekter og utgifter	1695598
6100	Statsforvaltningen	A1 Skatter	2000 Inntekter og utgifter	208477
6100	Statsforvaltningen	A1 Skatter	2001 Inntekter og utgifter	209395

Data om offentlig forvaltning fra SSB

(<https://data.ssb.no/api/v0/dataset/928194?lang=no>)

BCNF = godt design?

$R(\text{Brukernavn}, \text{Navn}, \text{Etternavn}, \text{KursKode}, \text{KursNavn})$

FDer:

1. $\text{Brukernavn} \rightarrow \text{Navn}$
2. $\text{Brukernavn} \rightarrow \text{Etternavn}$
3. $\text{KursKode} \rightarrow \text{KursNavn}$

◆ Kan dekomponeres til:

- ◆ $S(\text{Brukernavn}, \text{Navn})$
- ◆ $T(\text{Brukernavn}, \text{Etternavn})$
- ◆ $U(\text{Brukernavn}, \text{KursKode})$
- ◆ $V(\text{KursKode}, \text{KursNavn})$

- ◆ Har her brukt algoritmen, men fjernet tillukningen av venstresiden
- ◆ Tapsfri dekomponering og BCNF, men vi har to tabeller som burde vært en
- ◆ Godt databasedesign er altså mer enn bare normalformer

Takk for nå!

Neste uke vil handle om aggregering og sortering i SQL.