

IN2090 – Databaser og datamodellering

07 – Introduksjon til databasedesign

Leif Harald Karlsen (Evgenij Thorstensen)

leifhka@ifi.uio.no



Universitetet i Oslo

Motiverende eksempel

La oss si vi skal lage en database for å lagre data om studenter, kurs, og karakterer.

- ◆ For studenter: Brukernavn, navn, etternavn, adresse...
- ◆ For kurs: Kurskode, tittel, beskrivelse, ant. studiepoeng...
- ◆ Hvem har fått hvilken karakter i hva.

Naiv løsning: Alt i en tabell.

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

- ◆ Hvorfor er dette et dårlig design, annet enn at det er rotete?
- ◆ Hvilke praktiske problemer har vi her?

Anomalier: Dataduplisering

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

I tabellen dupliseres data:

- ◆ Tar unødvendig med plass
- ◆ Gjør spørringer mindre effektive

Anomalier: Insetting og oppdatering

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
abcdef	Aber C.	Deflan	Addr4	IN9999	Quantum	Beskr3	10	

Dupliseringen gjør det vanskeligere å sette inn og oppdatere data:

- ◆ Må skrive inn all informasjon om en student, selvom vi bare ønsker å legge til et nytt resultat til en allerede eksisterende student
- ◆ Umulig å sette inn en student uten kurs eller et kurs uten en student (uten masse `NULLs`)
- ◆ Ved oppdateringer må vi være sikre på å oppdatere alle duplikatene

Anomalier: Sletting

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

- ◆ Sletting av et kurs kan føre til sletting av en student
- ◆ Sletting av en student kan føre til sletting av et kurs
- ◆ Vanskelig å komme unna

Et bra skjema

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

Merk: Samme attributtene og samme verdiene!

Godt og dårlig skjema: Hva er forskjellen?

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

- ◆ Godt skjema: Brnavn bestemmer alt om en student, Kurskode alt om et kurs!
- ◆ Dårlig skjema: Ingen (minste mengde) attributter bestemmer resten!

Prinsipper for databasedesign

Altså handler god database design om:

- ◆ Data om entiteter i hver sin tabell.
- ◆ Relasjoner mellom entiteter via referanser.

Kan dette gjøres mer systematisk?

Normalformer

- ◆ Det går an å formelt definere kriterier som gjør at anomalier ikke forekommer.
- ◆ Et skjema som oppfyller kriteriene, sies å være på en normalform.
- ◆ Vi skal se på normalformene 1NF, 2NF, 3NF, og BCNF.
- ◆ For å komme dit, trenger vi litt teori.

Takk for nå!

Neste video vil handle om funksjonelle avhengigheter.