IN2090 – Databaser og datamodellering

08 – Eksempler: Databasedesign og normalformer

Leif Harald Karlsen (Evgenij Thorstensen)
leifhka@ifi.uio.no

Universitetet i Oslo

Oppgave 1 – Løsning

Gitt følgende relasjon:

`Person(personnr, navn, initialer, fødselsdato, alder)`

med FDene:

- `personnr` → `navn, fødselsdato`
- `navn` → `initialer`
- `fødselsdato` → `alder`

Kandidatnøkler (fra forige uke): `personnr`

1. Finn ut hvilken normalform relasjonen er på
2. Dekomponer relasjonen til BCNF

Normalform:

- `personnr` → `navn, fødselsdato`: bryter ikke med BCNF
- `navn` → `initialer`:
  - Bryter med BCNF (`navn` ikke supernøkkel);
  - bryter med 3NF (`initialer` ikke nøkkelattributt);
  - bryter ikke med 2NF (`navn` ikke del av kandidatnøkkel).
- `fødselsdato` → `alder`:
  - Bryter med BCNF (`fødselsdato` ikke supernøkkel);
  - bryter med 3NF (`alder` ikke nøkkelattributt);
  - bryter ikke med 2NF (`fødselsdato` ikke del av kandidatnøkkel).

Altså er `Person` på 2NF.

Oppgave 1 – Løsning

Gitt følgende relasjon:

$\texttt{Person}(\texttt{personnr}, \texttt{navn}, \texttt{initialer}, \texttt{fødselsdato}, \texttt{alder})$

med FDene:

- $\texttt{personnr} \rightarrow \texttt{navn}, \texttt{fødselsdato}$
- $\texttt{navn} \rightarrow \texttt{initialer}$
- $\texttt{fødselsdato} \rightarrow \texttt{alder}$

Kandidatnøkler (fra forige uke): $\texttt{personnr}$

1. Finn ut hvilken normalform relasjonen er på
2. Dekomponer relasjonen til BCNF

Dekomponering:

- $\texttt{navn} \rightarrow \texttt{initialer}$: Bryter BCNF
  - $\texttt{navn}^+ = \{\texttt{navn}, \texttt{initialer}\}$
  - Dekomponerer til:
    $S_1(\texttt{navn}, \texttt{initialer})$ (BCNF)
    $S_2(\texttt{personnr}, \texttt{navn}, \texttt{fødselsdato}, \texttt{alder})$
- $\texttt{fødselsdato} \rightarrow \texttt{alder}$: Bryter BCNF for S_2
  - $\texttt{fødselsdato}^+ = \{\texttt{fødselsdato}, \texttt{alder}\}$
  - Dekomponerer til:
    $S_{21}(\texttt{fødselsdato}, \texttt{alder})$ (BCNF)
    $S_{22}(\texttt{personnr}, \texttt{navn}, \texttt{fødselsdato})$ (BCNF)

$\texttt{Person}$ dekomponeres altså tapsfritt til BCNF med:

- $S_1(\texttt{navn}, \texttt{initialer})$
- $S_{21}(\texttt{fødselsdato}, \texttt{alder})$
- $S_{22}(\texttt{personnr}, \texttt{navn}, \texttt{fødselsdato})$

Oppgave 2 – Løsning

$$\texttt{Produkt}(\texttt{produktID}, \texttt{navn}, \texttt{kategori}, \texttt{pris}, \texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr}, \texttt{poststed})$$

med FDene:

- $\texttt{produktID} \rightarrow \texttt{navn}$
- $\texttt{produktID} \rightarrow \texttt{kategori}$
- $\texttt{produktID} \rightarrow \texttt{pris}$
- $\texttt{navn}, \texttt{kategori} \rightarrow \texttt{produktID}$
- $\texttt{butikkID} \rightarrow \texttt{butikknavn}$
- $\texttt{butikkID} \rightarrow \texttt{butikktype}$
- $\texttt{butikkID} \rightarrow \texttt{adresse}$
- $\texttt{butikkID} \rightarrow \texttt{postnr}$
- $\texttt{postnr} \rightarrow \texttt{poststed}$

Kandiadtnøkler (fra forige uke): $\{\texttt{butikkID}, \texttt{produktID}\}$, $\{\texttt{butikkID}, \texttt{navn}, \texttt{kategori}\}$

1. Finn normalformen til relasjonen
2. Dekomponer relasjonen til BCNF

Normalform:

- $\texttt{produktID} \rightarrow \texttt{navn}$:
  - Bryter BCNF ($\texttt{produktID}$ ikke supernøkkel)
  - Bryter ikke 3NF ($\texttt{navn}$ er nøkkelattributt)
- $\texttt{produktID} \rightarrow \texttt{kategori}$: Samme som over
- $\texttt{produktID} \rightarrow \texttt{pris}$:
  - Bryter BCNF ($\texttt{produktID}$ ikke supernøkkel)
  - Bryter 3NF ($\texttt{pris}$ ikke nøkkelattributt)
  - Bryter 2NF ($\texttt{produktID}$ er del av kandidatnøkkel)

Altså er $\texttt{Produkt}$ på 1NF.

Oppgave 2 – Løsning

$\texttt{Produkt}(\texttt{produktID}, \texttt{navn}, \texttt{kategori}, \texttt{pris}, \texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr}, \texttt{poststed})$

med FDene:

- $\texttt{produktID} \rightarrow \texttt{navn}$
- $\texttt{produktID} \rightarrow \texttt{kategori}$
- $\texttt{produktID} \rightarrow \texttt{pris}$
- $\texttt{navn}, \texttt{kategori} \rightarrow \texttt{produktID}$
- $\texttt{butikkID} \rightarrow \texttt{butikknavn}$
- $\texttt{butikkID} \rightarrow \texttt{butikktype}$
- $\texttt{butikkID} \rightarrow \texttt{adresse}$
- $\texttt{butikkID} \rightarrow \texttt{postnr}$
- $\texttt{postnr} \rightarrow \texttt{poststed}$

Kandiadtnøkler (fra forige uke):
$\{\texttt{butikkID}, \texttt{produktID}\}$,
$\{\texttt{butikkID}, \texttt{navn}, \texttt{kategori}\}$

1. Finn normalformen til relasjonen
2. Dekomponer relasjonen til BCNF

Dekomponering av $\texttt{Produkt}$:

- $\texttt{produktID} \rightarrow \texttt{navn}$: Bryter BCNF
  - $\texttt{produktID}^+ = \{\texttt{produktID}, \texttt{navn}, \texttt{kategori}, \texttt{pris}\}$
  - Dekomponerer til:
    $S_1(\texttt{produktID}, \texttt{navn}, \texttt{kategori}, \texttt{pris})$ (på BCNF)
    $S_2(\texttt{produktID}, \texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr}, \texttt{poststed})$
- $\texttt{butikkID} \rightarrow \texttt{butikknavn}$: Bryter BCNF
  - $\texttt{butikkID}^+ = \{\texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr}, \texttt{poststed}\}$
  - Dekomponerer til:
    $S_{21}(\texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr}, \texttt{poststed})$
    $S_{22}(\texttt{produktID}, \texttt{butikkID})$ (på BCNF)

Oppgave 2 – Løsning

$\texttt{Produkt}(\texttt{produktID}, \texttt{navn}, \texttt{kategori}, \texttt{pris}, \texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr}, \texttt{poststed})$

med FDene:

- $\texttt{produktID} \rightarrow \texttt{navn}$
- $\texttt{produktID} \rightarrow \texttt{kategori}$
- $\texttt{produktID} \rightarrow \texttt{pris}$
- $\texttt{navn}, \texttt{kategori} \rightarrow \texttt{produktID}$
- $\texttt{butikkID} \rightarrow \texttt{butikknavn}$
- $\texttt{butikkID} \rightarrow \texttt{butikktype}$
- $\texttt{butikkID} \rightarrow \texttt{adresse}$
- $\texttt{butikkID} \rightarrow \texttt{postnr}$
- $\texttt{postnr} \rightarrow \texttt{poststed}$

Kandiadtnøkler (fra forige uke):
$\{\texttt{butikkID}, \texttt{produktID}\}$,
$\{\texttt{butikkID}, \texttt{navn}, \texttt{kategori}\}$

1. Finn normalformen til relasjonen
2. Dekomponer relasjonen til BCNF

Dekomponering av
$S_{21}(\texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr}, \texttt{poststed})$**:**

- $\texttt{postnr} \rightarrow \texttt{poststed}$: Bryter BCNF
  - $\texttt{postnr}^+ = \{\texttt{postnr}, \texttt{poststed}\}$ (på BCNF)
  - Dekomponerer til disse, begge på BCNF:
    $S_{211}(\texttt{postnr}, \texttt{poststed})$
    $S_{212}(\texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr})$

Oppgave 2 – Løsning

$$\texttt{Produkt}(\texttt{produktID}, \texttt{navn}, \texttt{kategori}, \texttt{pris}, \texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr}, \texttt{poststed})$$

med FDene:

- $\texttt{produktID} \rightarrow \texttt{navn}$
- $\texttt{produktID} \rightarrow \texttt{kategori}$
- $\texttt{produktID} \rightarrow \texttt{pris}$
- $\texttt{navn}, \texttt{kategori} \rightarrow \texttt{produktID}$
- $\texttt{butikkID} \rightarrow \texttt{butikknavn}$
- $\texttt{butikkID} \rightarrow \texttt{butikktype}$
- $\texttt{butikkID} \rightarrow \texttt{adresse}$
- $\texttt{butikkID} \rightarrow \texttt{postnr}$
- $\texttt{postnr} \rightarrow \texttt{poststed}$

Kandiadtnøkler (fra forige uke):
$\{\texttt{butikkID}, \texttt{produktID}\}$,
$\{\texttt{butikkID}, \texttt{navn}, \texttt{kategori}\}$

1. Finn normalformen til relasjonen
2. Dekomponer relasjonen til BCNF

Dekomponering blir alstå:

$S_1(\texttt{produktID}, \texttt{navn}, \texttt{kategori}, \texttt{pris})$

$S_{211}(\texttt{postnr}, \texttt{poststed})$

$S_{212}(\texttt{butikkID}, \texttt{butikknavn}, \texttt{butikktype}, \texttt{adresse}, \texttt{postnr})$

$S_{22}(\texttt{produktID}, \texttt{butikkID})$

Oppgave 3 (Vanskelig, ikke pensum!)

Lag et skjema på BCNF som inneholder dataene for 2019 i ”Fisketillatelser med fartøytilknytning og kvotestørrelser” fra Fiskeridirektoratet:

`https://fiskeridir.no/Tall-og-analyse/AApne-data/AApne-datasett/Fartoey-eier-og-fisketillatelser`

Oppgave 3 – Løsning: Lage tabell

- Starter med å laste ned filen og åpne i et regnearkprogram (e.g. Libreoffice) og åpner arket med navn ”2019”
- Lagrer filen som CSV med semicolon som ”delimiter” og all tekst omringet av enkle (`'`) fnutter (filnavn `kvoter.csv`)
- Flytter øverste linje fra CSV-filen inn i egen fil og skriver det om til en `CREATE TABLE`-kommando:

```
CREATE SCHEMA fiskeri;
CREATE TABLE fiskeri.kvoter_raw (
    Data_pr text,
    Fartøy_ID text,
    Registreringsmerke text,
    Tillatelse_kode text,
    Tillatelse text,
    Tillatelse_ID text,
    Tillatelse_fra_dato text,
    Tillatelse_til_dato text,
    Linjenummer int,
    Linjenummer_beskrivelse text,
    Kvotestørrelse float,
    Kvotestr_fra_dato text,
    Kvotestr_til_dato text
);
```

Oppgave 3 – Løsning: Datainnlasting

- Datoene er på feil (norsk) format, må oversettes til noe PostgreSQL-forstår
- Gjør dette med et view, `regexp_replace` og casting til riktig type:

```
CREATE VIEW fiskeri.kvoter AS
SELECT
    regexp_replace(Data_pr, '(\d\d)\.(\d\d)\.(\d\d\d\d)(.*)', '\3-\2-\1\4')::date AS Data_pr,
    Fartøy_ID,
    Registreringsmerke,
    Tillatelse_kode,
    Tillatelse,
    Tillatelse_ID,
    regexp_replace(Tillatelse_fra_dato, '(\d\d)\.(\d\d)\.(\d\d\d\d)(.*)', '\3-\2-\1\4')::timestamp
        AS Tillatelse_fra_dato,
    regexp_replace(Tillatelse_til_dato, '(\d\d)\.(\d\d)\.(\d\d\d\d)(.*)', '\3-\2-\1\4')::timestamp
        AS Tillatelse_til_dato,
    Linjenummer,
    Linjenummer_beskrivelse,
    Kvotestørrelse,
    regexp_replace(Kvotestr_fra_dato, '(\d\d)\.(\d\d)\.(\d\d\d\d)(.*)', '\3-\2-\1\4')::date
        AS Kvotestr_fra_dato,
    regexp_replace(Kvotestr_til_dato, '(\d\d)\.(\d\d)\.(\d\d\d\d)(.*)', '\3-\2-\1\4')::date
        AS Kvotestr_til_dato
FROM fiskeri.kvoter_raw;
```

Oppgave 3 – Løsning: Datainnlasting

Kjører følgende kommando for å laste dataene inn i tabellen vår:

```
cat kvoter.csv | psql <flagg> -c "COPY fiskeri.kvoter_raw FROM stdin CSV DELIMITER ';' NULL AS ''";
```

hvor `<flagg>` er de vanlige tilkobligsdetaljene til den personlige databasen

Oppgave 3 – Løsning: Bestemme FDer og dekomponering

FDer (fra dokumentasjonen om dataene):

1. `Fartøy_ID` → `Registreringsmerke`
2. `Tillatelse_kode` → `Tillatelse`
3. `Tillatelse_ID` → `Data_pr`
4. `Tillatelse_ID` → `Fartøy_ID`
5. `Tillatelse_ID` → `Tillatelse_kode`
6. `Tillatelse_ID` → `Tillatelse_Gjelder_Fra_Dato`
7. `Tillatelse_ID` → `Tillatelse_Gjelder_Til_Dato`
8. `Tillatelse_ID` → `Kvotestr_Gjelder_Fra_Dato`
9. `Tillatelse_ID` → `Kvotestr_Gjelder_Til_Dato`
10. `Tillatelse_ID, Linjenummer` → `Linjenummer_beskrivelse`
11. `Tillatelse_ID, Linjenummer` → `Kvotestørrelse`

Kandidatnøkkel: $\{\texttt{TillatelseID}, \texttt{Linjenummer}\}$

Oppgave 3 – Løsning: Dekomponering

```
BEGIN;

-- Fartøy_ID -> Registreringsmerke bryter med BCNF
-- Tillukningen til Fartøy_ID er {Fartøy_ID, Registreringsmerke},
-- altså får vi følgende (som ikke bryter med BCNF):
CREATE TABLE fiskeri.fartøy(
    Fartøy_ID text PRIMARY KEY,
    Registreringsmerke text
);
```

Oppgave 3 – Løsning: Dekomponering

```
-- Har nå en tabell med alle attributter bortsett fra Registreringsmerke
-- Tillatelse_kode -> Tillatelse bryter med BCNF
-- Tillukningen til Tillatelse_kode er {Tillatelse_kode, Tillatelse},
-- så får følgende (som ikke bryter BCNF):
CREATE TABLE fiskeri.tillatelsesInfo(
    Tillatelse_kode text PRIMARY KEY,
    Tillatelse text
);
```

Oppgave 3 – Løsning: Dekomponering

```
-- Har nå en tabell med alle attributter bortsett fra Registreringsmerke og Tillatelse
-- Tillatelse_ID -> Data_pr bryter med BCNF
-- Tillukningen til Tillatelse_ID er
--   {Tillatelse_ID, Data_pr, Fartøy_ID, Tillatelse_kode,
--    Tillatelse_Gjelder_Fra_Dato, Tillatelse_Gjelder_Til_Dato,
--    Kvotestr_Gjelder_Fra_Dato, Kvotestr_Gjelder_Til_Dato},
-- så får følgende (som ikke bryter BCNF):
CREATE TABLE fiskeri.tillatelse(
    Tillatelse_ID text PRIMARY KEY,
    Data_pr date,
    Fartøy_ID text REFERENCES fiskeri.fartøy(Fartøy_ID),
    Tillatelse_kode text REFERENCES fiskeri.tillatelsesInfo(Tillatelse_kode),
    Tillatelse_gjelder_fra_dato timestamp,
    Tillatelse_gjelder_til_dato timestamp,
    Kvotestr_gjelder_fra_dato timestamp,
    Kvotestr_gjelder_til_dato timestamp
);
```

Oppgave 3 – Løsning: Dekomponering

```
-- Står nå igjen med følgende tabell, som ikke bryter med BCNF:
CREATE TABLE fiskeri.tillatelsesLinje(
    Tillatelse_ID text REFERENCES fiskeri.tillatelse(Tillatelse_ID),
    Linjenummer int,
    Linjenummer_beskrivelse text,
    Kvotestørrelse float,
    PRIMARY KEY (Tillatelse_ID, Linjenummer)
);
```

Oppgave 3 – Løsning: Migrering

```sql
-- Setter så inn data fra fiskeri.kvoter i hver tabell:

INSERT INTO fiskeri.fartøy
SELECT DISTINCT Fartøy_ID, Registreringsmerke
FROM fiskeri.kvoter;

INSERT INTO fiskeri.tillatelsesInfo
SELECT DISTINCT Tillatelse_kode, Tillatelse
FROM fiskeri.kvoter;

INSERT INTO fiskeri.tillatelse
SELECT DISTINCT Tillatelse_ID, Data_pr, Fartøy_ID, Tillatelse_kode,
    Tillatelse_gjelder_fra_dato, Tillatelse_gjelder_til_dato
    Kvotestr_gjelder_fra_dato, Kvotestr_gjelder_til_dato
FROM fiskeri.kvoter;

INSERT INTO fiskeri.tillatelsesLinje
SELECT DISTINCT Tillatelse_ID, Linjenummer, Linjenummer_beskrivelse, Kvotestørrelse
FROM fiskeri.kvoter
WHERE Linjenummer IS NOT NULL; -- Finnes rader i kvoter som mangler linjenummer
```