

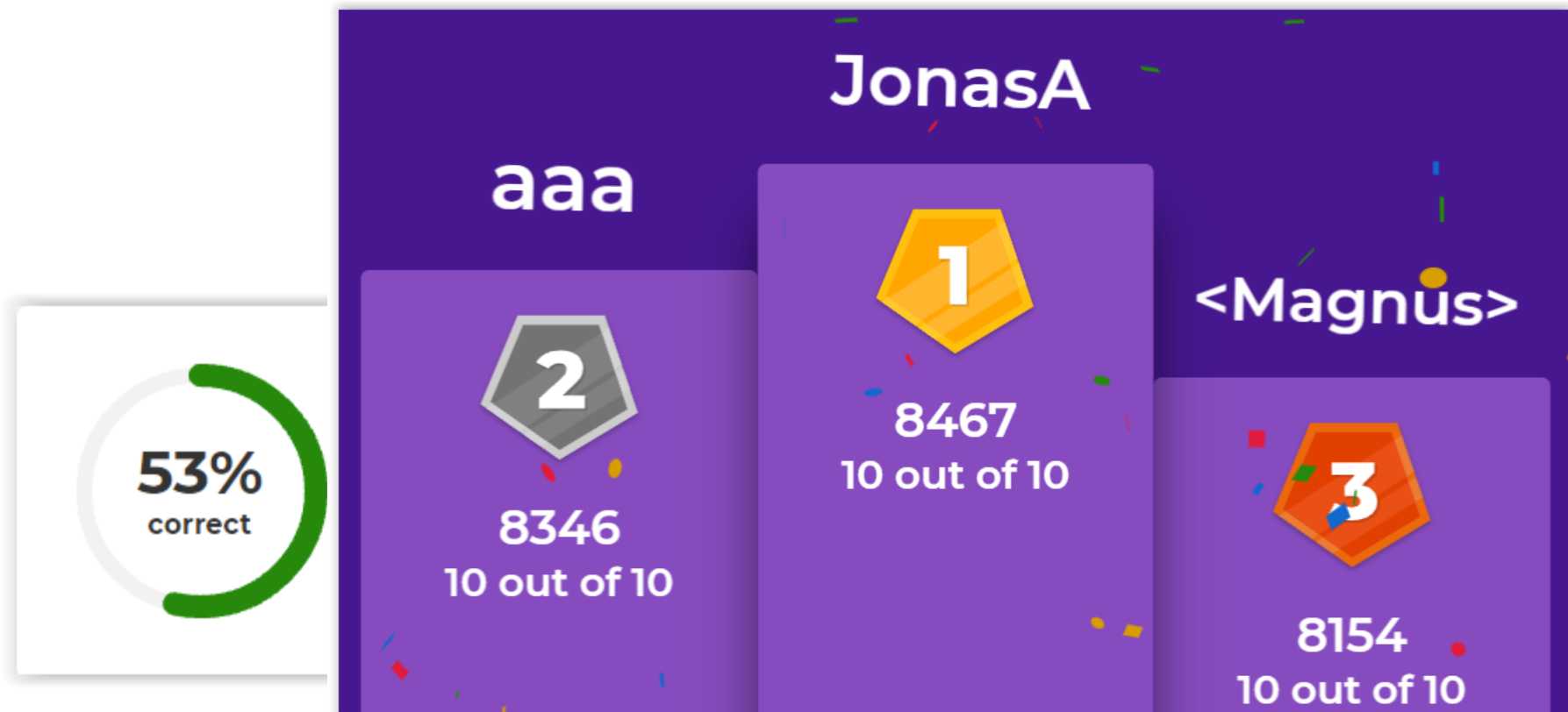


Økt 9 (av 12)

Per Lauvås / per.lauvas@kristiania.no

Yuan Lin / yuan.lin@kristiania.no

Gårdagens Kahoot - pallen



Dagens temaer

Dagens pensum: [Læreboka, kapittel 8.2.1 - 8.2.6](#)

- Dagens tema: [Normalisering](#)
 - Hensikt
 - Begreper
 - 1. til 3. normalform (1NF, 2NF og 3NF)

Classic WoW: DB modellering viktig!

- Når Blizzard skulle gi ut Classic WoW høsten 2019, var kunnskap om SQL og databaser, og spesielt ER modellering, veldig viktig!
 - Les gjerne her: [Dev Watercooler: World of Warcraft Classic](#)



Hensikten med normalisering

- Hensikten med **normalisering** er å finne de mest gunstige relasjoner for en gitt database.
- **Hovedkriteriet vi etterstreber** ved normalisering:
 - **Minimal dobbellagring** (redundans), så attributter kun lagres ett sted.
Unntak for Foreign keys: Disse trenger nødvendigvis å lagres flere steder, da de skal koble flere entiteter (tabeller) sammen.
- Fra læreboka:
 - «Normalisering går ut på å gjenta normaliseringssteg helt til all redundans er fjernet»

Problemer med dobbellagring

- Tabellen `ansatt_filial` (med primærnøkkel `anr`) inneholder **dobbellagring** (**redundans**) i flere kolonner.
 - FNR, FADRESSE og FPST: filialnummer, filialadresse og filialpostnummer.

ANR	NAVN	AADRESSE	APST	STILLING	LOENN	FNR	FADRESSE	FPST
3	Jon Hvit	Bruveien 7	4000	Leder	30000	1	Utleieveien 15	4000
4	Anne Strand	Strandgaten	2000	Megler	12000	1	Utleieveien 15	4000
20	Olav Gautesen	Galmannsveien 4	3000	Smegler	26000	1	Utleieveien 15	4000
5	David Opalsen	Gulerleveien 43	2000	Sekretær	18000	1	Utleieveien 15	4000
2	Marie Hovland	Strilegaten 8	5000	Leder	13000	2	Smuglerstien 67	5000
23	Ole Ås	Mor Åseveien 56	4000	Megler	17000	2	Smuglerstien 67	5000
21	Per Pollesen	Podlestadvveien 5	5000	Sekretær	15000	2	Smuglerstien 67	5000
7	Karl Hansen	Olavsgt 7	2000	Leder	25000	3	Snusveien 7	7000

- *Hvilke problemer gir denne tabellen oss, med tanke på innlegging, endring og sletting?*

INSERT-problemer

- Legge inn **ny ansatt**, problemer:
 - Må samtidig **legge inn allerede registrerte data på nytt** for filial. (Filialnummer, adresse og postnummer).
 - **Ved feiltasting får man problemer** med at adressen blir forskjellig for samme filial (får en inkonsistent database).

ANR	NAVN	AADRESSE	APST	STILLING	LOENN	FNR	FADRESSE	FPST
3	Jon Hvit	Bruveien 7	4000	Leder	30000	1	Utleieveien 15	4000
4	Anne Strand	Strandgaten	2000	Megler	12000	1	Utleieveien 15	4000

- Legge inn **ny filial uten ansatte**, problemer:
 - Må sette inn NULL på feltene som gjelder ansatte, for eksempel `anr`.
 - `anr` er primærnøkkel og NULL er ikke tillatt her.
 - Må dermed **legge inn en dummyansatt**.

DELETE-problemer

- Slette siste ansatte for en filial, problemer:
 - Mister samtidig alle opplysninger om filialen.

ANR	NAVN	AADRESSE	APST	STILLING	LOENN	FNR	FADRESSE	FPST
---	-----	-----	----	-----	-----	---	-----	----
3	Jon Hvit	Bruveien 7	4000	Leder	30000	1	Utleieveien 15	4000

UPDATE-problemer

- Endring av postnummer eller adresse til en filial, problemer:
 - Må gjøre de samme endringene i alle rader til alle ansatte for denne filialen.
 - Ved feiltasting får man problemer med at adressen blir forskjellig på samme filial (får en inkonsistent database).

ANR	NAVN	AADRESSE	APST	STILLING	LOENN	FNR	FADRESSE	FPST
3	Jon Hvit	Bruveien 7	4000	Leder	30000	1	Utleieveien 15	4000
4	Anne Strand	Strandgaten	2000	Megler	12000	1	Utleieveien 15	4000

Bedre tabellstruktur

- Løsning: Splitte ansatt_filial i to tabeller: ansatt og filial:

ANR	NAVN	ADRESSE	POST	STILLING	LOENN	FNR
---	-----	-----	----	-----	-----	---
3	Jon Hvit	Bruveien 7	4000	Leder	30000	1
4	Anne Strand	Strandgaten	2000	Megler	12000	1
20	Olav Gautesen	Galmannsveien 4	3000	Megler	26000	1
5	David Opalsen	Gulerleveien 43	2000	Sekretær	18000	1
2	Marie Hovland	Strilegaten 8	5000	Leder	13000	2
23	Ole Ås	Mor Åseveien 56	4000	Megler	17000	2
21	Per Pollesen	Podlestadveien 5	5000	Sekretær	15000	2
7	Karl Hansen	Olavsgt 7	2000	Leder	25000	3

FNR	ADRESSE	POST
---	-----	----
1	Utleieveien 15	4000
2	Smuglerstien 67	5000
3	Snusveien 7	7000

*Får vi nå problemer med
innlegging/endring/sletting?*

Bedre tabellstruktur, INSERT

- Legge inn ny ansatt:
 - Legger kun inn personopplysninger pluss korrekt filialnummer.
 - Filialadressen kan ikke bli forskjellig for en og samme filial, da filialopplysningene kun lagres i én rad i databasen. (Har en konsistent database).
- Legge inn ny filial uten ansatte:
 - Helt ok at det ikke finnes opplysninger om ansatte i filialtabellen.

Bedre struktur, DELETE/UPDATE

- Slette siste ansatte for en filial:
 - Filialen eksisterer fortsatt i filialtabellen.
- Endring av postnummer eller adresse til en filial:
 - Kun nødvendig å endre opplysningene for én rad i databasen.
(Har en konsistent database.)

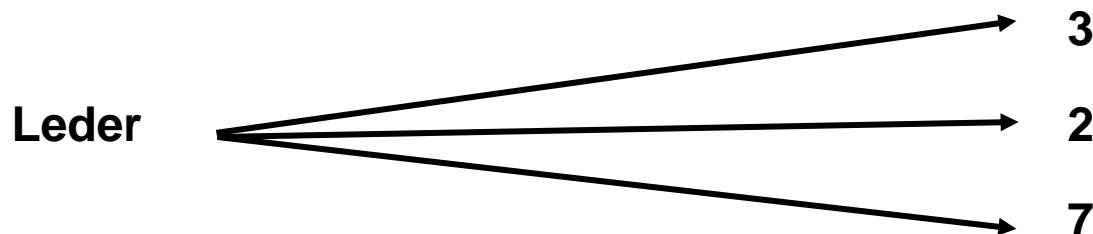
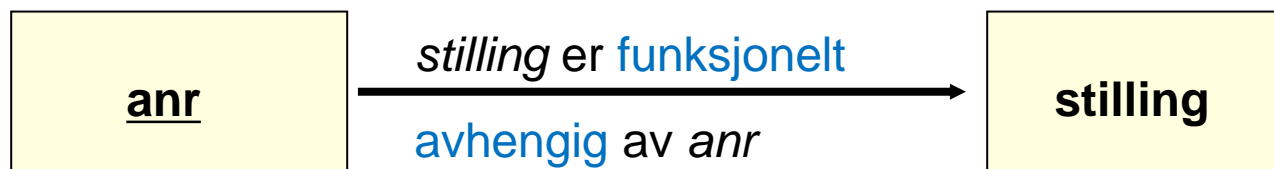
Normaliseringsprosessen

- Normalisering vil som regel si å **splitte databasen opp i flere tabeller**.
- Når en tabell skal splittes til flere, er det viktig at:
 - Vi får en tapsfri splitt: Alle rader i den opprinnelige tabellen må kunne dannes ved å koble sammen de nye tabellene. (Boka: Tapsfri dekomposisjon).
 - Vi bevarer avhengigheter: begrensninger i den originale tabellen bevares ved at vi lager tilsvarende begrensninger i de nye.
- Vi innfører begrepet "**funksjonell avhengighet**":
 - Merk at vi må forstå betydningen (semantikken) til attributtene for at dette skal ha mening.
 - Med andre ord: Vi må "forstå databasen" før vi kan normalisere den.

Definisjon, funksjonell avhengighet

- Hvis A og B er attributter i en relasjon R, så er B funksjonelt avhengig av A (kan skrives: $A \rightarrow B$) hvis det for hver verdi av A kan identifiseres kun én verdi av B, og hvis A gjentas får vi også samme verdi av B.
 - (A og B kan bestå av flere attributter).
- Altså: Har du A, så vet du B.
- Eksempel: Man kan si at 'stilling' er funksjonelt avhengig av 'ansattnummer' ('ansattnummer' \rightarrow 'stilling'). Da skal dette ansattnummeret alltid vise til samme stilling.
 - Det motsatte er ikke nødvendigvis tilfelle: Det er ikke sagt at det nødvendigvis er bare ett ansattnummer knyttet til en type stilling. Det kan være flere ledere (med hvert sitt ansattnummer, altså forskjellige personer) som leder hver sin avdeling, osv.
- *Litt vanskelige begreper, men kanskje lettere om vi ser et visuelt eksempel?*
 - *Se neste slide for eksempel for den tenkte databasen vi alt har sett på. :-)*

Funksjonell avhengighet



Determinant: viktig begrep

- Determinant:
 - En kolonne (eller fler) i en tabell som er bestemmende for minst en annen kolonne i tabellen.
- Generelt:
 - Ved $A \rightarrow B$ (B funksjonelt avhengig av A) er **A determinant for B**.
(Men B er ikke determinant for A.)
- Eksempelet på forrige side:
 - `anr` er determinant for `stilling`.
 - Men `stilling` er ikke determinant for `anr`.

Oppgave

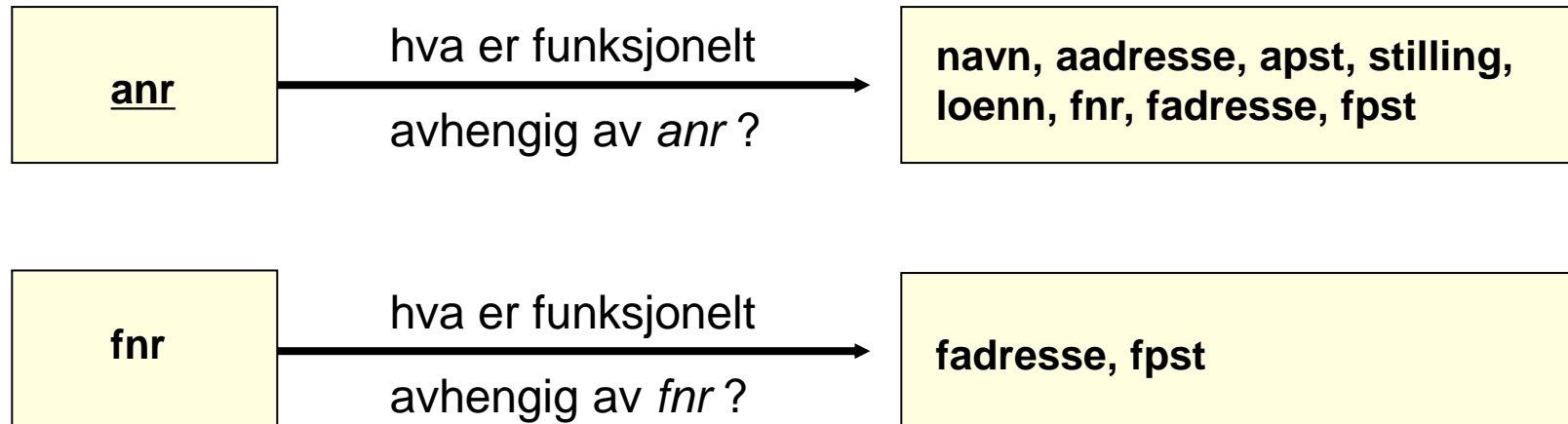
- Hvilke funksjonelle avhengigheter finner vi i tabellen nedenfor?

ANR	NAVN	AADRESSE	APST	STILLING	LOENN	FNR	FADRESSE	FPST
3	Jon Hvit	Bruveien 7	4000	Leder	30000	1	Utleieveien 15	4000
4	Anne Strand	Strandgaten	2000	Megler	12000	1	Utleieveien 15	4000
20	Olav Gautesen	Galmannsveien 4	3000	Smegler	26000	1	Utleieveien 15	4000
5	David Opalsen	Gulerleveien 43	2000	Sekretær	18000	1	Utleieveien 15	4000
2	Marie Hovland	Strilegaten 8	5000	Leder	13000	2	Smuglerstien 67	5000
23	Ole Ås	Mor Åseveien 56	4000	Megler	17000	2	Smuglerstien 67	5000
21	Per Pollesen	Podlestadveien 5	5000	Sekretær	15000	2	Smuglerstien 67	5000
7	Karl Hansen	Olavsgt 7	2000	Leder	25000	3	Snusveien 7	7000

– (Samme tabell ble brukt i eksempelet forrige video.)

Funksjonell avhengighet

ANSATT_FILIAL (anr, navn, aadresse, apst, stilling, loenn, fnr, fadresse, fpst)



Transitiv avhengighet

- Transitiv avhengighet er indirekte funksjonell avhengighet.
- A, B og C er attributter i en relasjon.
- $A \rightarrow B$ og $B \rightarrow C$.
- C er transitivt avhengig (indirekte funksjonelt avhengig) av A, via B.

Oppgave

- Hvilke transitive avhengigheter finner du i tabellen nedenfor?

ANR	NAVN	AADRESSE	APST	STILLING	LOENN	FNR	FADRESSE	FPST
3	Jon Hvit	Bruveien 7	4000	Leder	30000	1	Utleieveien 15	4000
4	Anne Strand	Strandgaten	2000	Megler	12000	1	Utleieveien 15	4000
20	Olav Gautesen	Galmannsveien 4	3000	Smegler	26000	1	Utleieveien 15	4000
5	David Opalsen	Gulerleveien 43	2000	Sekretær	18000	1	Utleieveien 15	4000
2	Marie Hovland	Strilegaten 8	5000	Leder	13000	2	Smuglerstien 67	5000
23	Ole Ås	Mor Åseveien 56	4000	Megler	17000	2	Smuglerstien 67	5000
21	Per Pollesen	Podlestadveien 5	5000	Sekretær	15000	2	Smuglerstien 67	5000
7	Karl Hansen	Olavsgt 7	2000	Leder	25000	3	Snusveien 7	7000

– (Fortsatt samme tabell.)

Transitiv anhengighet

ANR -----> FNR

FNR -> FADRESSE, FPST

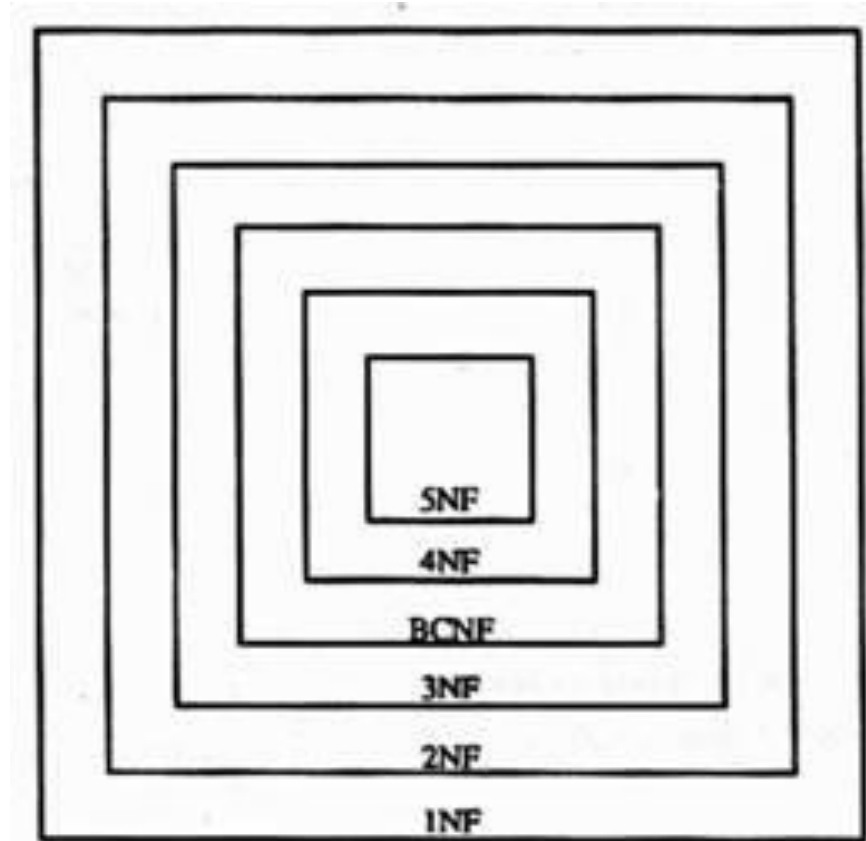
ANR -----> FADRESSE, FPST

ANR	NAVN	AADRESSE	APST	STILLING	LOENN	FNR	FADRESSE	FPST
3	Jon Hvit	Bruveien 7	4000	Leder	30000	1	Utleieveien 15	4000
4	Anne Strand	Strandgaten	2000	Megler	12000	1	Utleieveien 15	4000
20	Olav Gautesen	Galmannsveien 4	3000	Smegler	26000	1	Utleieveien 15	4000
5	David Opalsen	Gulerleveien 43	2000	Sekretær	18000	1	Utleieveien 15	4000
2	Marie Hovland	Strilegaten 8	5000	Leder	13000	2	Smuglerstien 67	5000
23	Ole Ås	Mor Åseveien 56	4000	Megler	17000	2	Smuglerstien 67	5000
21	Per Pollesen	Podlestadveien 5	5000	Sekretær	15000	2	Smuglerstien 67	5000
7	Karl Hansen	Olavsgt 7	2000	Leder	25000	3	Snusveien 7	7000

FADRESSE og FPST er **transitivt avhengig** av ANR, **via** FNR.

Normaliseringsprosessen

- I dag ser vi på 1NF til 3NF.
(Første til tredje normalform.)
- Neste gang ser vi på BCNF.
(Boyce-Codd normalform.)
- 4NF og 5NF (fjerde og femte normalform) er ikke relevant i dette emnet.



Unnormalisert form (UNF)

- Vi registrerer leieforhold som omfatter leietaker (lnr, lnavn), prospekt (pnr, padresse, fradato, tildato, leie) og eier (enr, enavn).
 - En tabell er unormalisert dersom den inneholder celler med mer enn ett dataelement. (Celler som ikke er atomære.)

<i>lnr</i>	<i>lnavn</i>	<i>pnr</i>	<i>padresse</i>	<i>fradato</i>	<i>tildato</i>	<i>leie</i>	<i>enr</i>	<i>enavn</i>	
5	Hansen	6	Aveien	1	01-JUL-94	01-SEP-96	3500	4	Olsen
?	?	1	Bveien	8	01-SEP-96	01-JAN-98	4400	12	Larsen
9	Persen	6	Aveien	1	01-SEP-96	01-SEP-97	3500	4	Olsen
?	?	2	Dveien	6	01-SEP-97	01-SEP-98	5500	6	Alfsen
?	?	5	Eveien	3	01-SEP-98	null	5000	6	Alfsen

Første normalform (1NF)

- En tabell er på **1NF** dersom:
 - Cellene inneholder bare ett dataelement.
 - (NB: "Ett dataelement" kan godt være flere ord, f.eks. i en varchar celle.)
- Vi kan fjerne cellene med flere dataelementer ved å lage én rad per element:
 - Da inneholder hvert felt kun en verdi.
 - Primærnøkkelen blir her sammensatt av (lnr, pnr).

Leietaker_Leiekontrakt_Prospekt_Eier

lnr	lnavn	pnr	padresse	fradato	tildato	leie	enr	enavn
---	-----	---	-----	-----	-----	----	---	-----
5	Hansen	6	Aveien 1	01-JUL-94	01-SEP-96	3500	4	Olsen
5	Hansen	1	Bveien 8	01-SEP-96	01-JAN-98	4400	12	Larsen
9	Persen	6	Aveien 1	01-SEP-96	01-SEP-97	3500	4	Olsen
9	Persen	2	Dveien 6	01-SEP-97	01-SEP-98	5500	6	Alfsen
9	Persen	5	Eveien 3	01-SEP-98	null	5000	6	Alfsen

Andre normalform (2NF)

- En tabell er på 2NF dersom:
 - Den er på 1NF.
 - Og ingen delmengde ("subset") av primærnøkkelen er determinant for en eller flere andre kolonner (ingen "delvis avhengighet" til PK).
 - Merk: Om en tabell er på 1NF og PK består av én kolonne, er tabellen også på 2NF!

Ok: avhengighet til hele PK.

lnr	lnavn	pnr	padresse	fradato	tildato	leie	enr	enavn
5	Hansen	6	Aveien 1	01-JUL-94	01-SEP-96	3500	4	Olsen
5	Hansen	1	Bveien 8	01-SEP-96	01-JAN-98	4400	12	Larsen
9	Persen	6	Aveien 1	01-SEP-96	01-SEP-97	3500	4	Olsen
9	Persen	2	Dveien 6	01-SEP-97	01-SEP-98	5500	6	Alfsen
9	Persen	5	Eveien 3	01-SEP-98	null	5000	6	Alfsen

Ikke ok: delvis avhengighet til PK.

Andre normalform (2NF) – forts.

- Primærnøkkelen er sammensatt.
 - Vi skiller ut kolonnene som er avhengige av deler av primærnøkkelen og plasserer disse i nye tabeller:

Leietaker

lnr	lnavn
---	-----
5	Hansen
9	Persen

Leiekontrakt

lnr	pnr	fradato	tildato
---	---	-----	-----
5	6	01-JUL-94	01-SEP-96
5	1	01-SEP-96	01-JAN-98
9	6	01-SEP-96	01-SEP-97
9	2	01-SEP-97	01-SEP-98
9	5	01-SEP-98	null

Prospekt_Eier

pnr	padresse	leie	enr	enavn
---	-----	-----	---	-----
6	Aveien	1 3500	4	Olsen
1	Bveien	8 4400	12	Larsen
2	Dveien	6 5500	6	Alfsen
5	Eveien	3 5000	6	Alfsen

Tredje normalform (3NF)

- En tabell er på 3NF dersom:
 - Den er på 2NF.
 - Og ingen ikke-primærnøkkel-attributt har en transitiv avhengighet til en primærnøkkel. (Ingen $A \rightarrow B$ og $B \rightarrow C \Rightarrow A \rightarrow C$.)
- Alternativt kan vi si det slik: Alle ikke-primærnøkkel-attributter er kun funksjonelt avhengige av primærnøkkel.

pnr	adresse	leie	enr	enavn
6	Aveien	1 3500	4	Olsen
1	Bveien	8 4400	12	Larsen
2	Dveien	6 5500	6	Alfsen
5	Eveien	3 5000	6	Alfsen

Merk: Tabellene
Leietaker og
Leiekontrakt er
allerede på 3NF.

Tredje normalform (3NF) – forts.

- Vi fjerner transitivt avhengige attributter fra tabellen og plasserer de i en ny tabell.
- **Determinanten** blir **nøkkelen**.

Prospekt

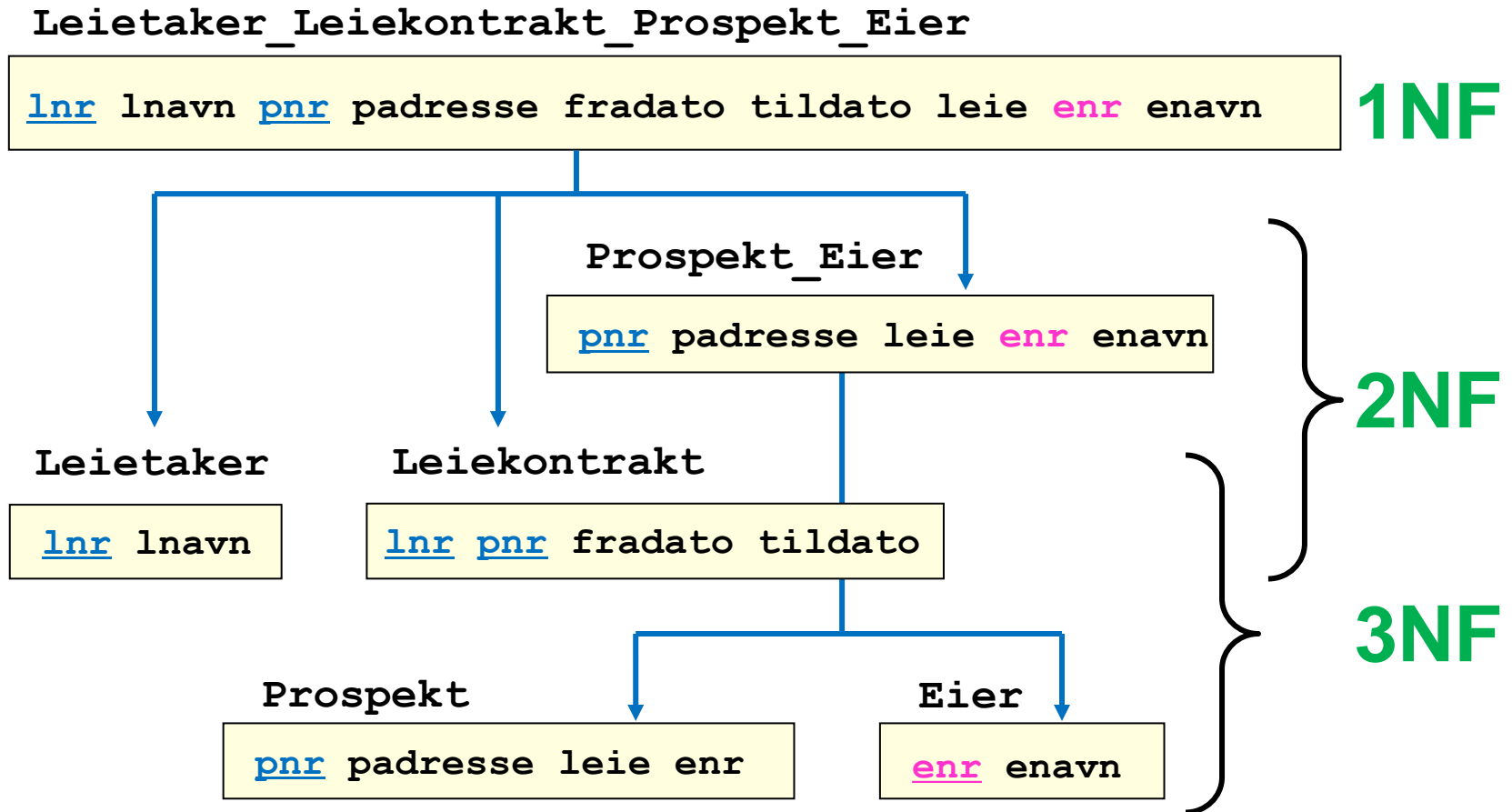
pnr	adresse	leie	enr
---	-----	----	---
6	Aveien 1	3500	4
1	Bveien 8	4400	12
2	Dveien 6	5500	6
5	Eveien 3	5000	6

Eier

enr	enavn
---	-----
4	Olsen
12	Larsen
6	Alfsen



Oppsummering: fra 1NF - 3NF



1NF til 3NF, oppsummert

- 1NF:
 - Cellene inneholder bare *ett* dataelement. (Er atomære.)
- 2NF, dersom den er på 1NF og:
 - Ingen avhengighet til *deler av primærnøkkelen*. (Ingen delmengde av PK er determinant for kolonner som ikke er del av PK.)
- 3NF, dersom den er på 2NF og:
 - Ingen kolonne utenfor primærnøkkelen har en *transitiv* (indirekte) avhengighet til *primærnøkkelen*. ($A \rightarrow B$ og $B \rightarrow C \Rightarrow A \rightarrow C$.)

Er NF, determinant, m.m. fortsatt litt hårete?

- Antakeligvis er normalisering fortsatt et tema med litt uklare punkter?
- I så fall – forslag:
 - Ta en titt på en (eller fler) av studentvideoene fra tidligere om normalisering. (Du finner de i lista med eksterne ressurser på Canvas.)
 - Og/eller google normaliseringsvideoer på nett.
 - Og/eller les de relevante avsnittene i læreboka en gang til.
 - Og/eller spør veiledere og foreleser i øvingstimene.
- Og, selvfølgelig: Gjør øvingsoppgavene!

Neste gang

- Boyce-Codd normalform (BCNF) og denormalisering.
- Pensum til neste gang: [Resten av kapittel 8](#).

Husk at arbeidskravet (del 1: modellen) skal leveres senest 27. oktober.