Tomas Sandnes
tomas.sandnes@kristiania.no

Week 10

DB1102 / PGR 111 – DATABASES

Today's topics

(Today's chapters: 8.2 in Norwegian book, 7.3 in English)

- Recap of 1NF to 3NF & task 1 + 4 from last week.

- Normalization, part 2:
  - Boyce-Codd Normal form (BCNF)
  - Denormalization

Regarding feedback on Coursework

- Regarding feedback on the coursework requirement.
  - Per Lauvås will host a Zoom session for this.
  - More info to come by Per (sometime after November 1st).

- Here's the info from Per: (in Norwegian)
  - *"Ja, jeg tenkte jeg kunne holde en Zoom-session for de som er interessert i en gjennomgang. Men det blir etter 1. nov da noen i DB1102 har fått utvidet frist til den datoen."*

1NF to 3NF

sum-up & exercises

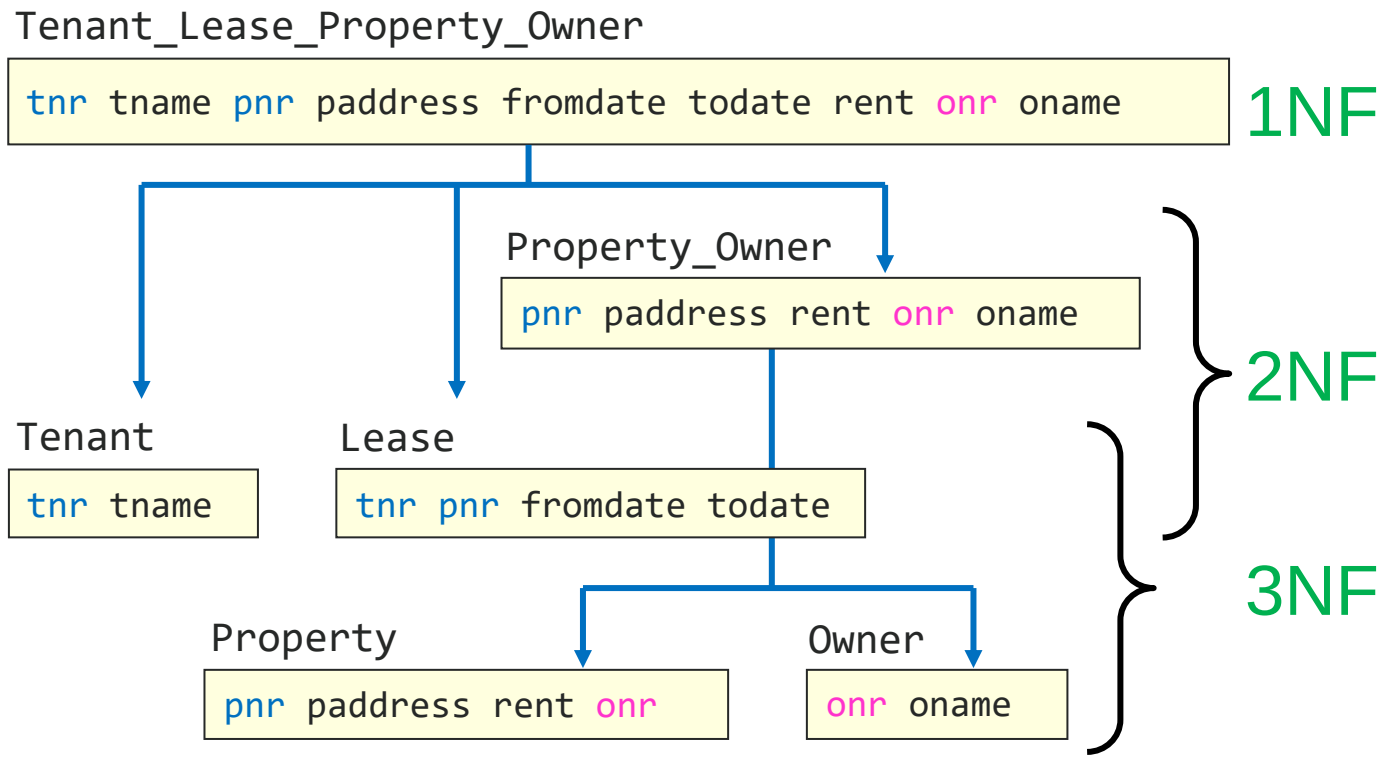
1NF to 3NF, short version *(repetition)*

- A table is 1NF if:
  - The cells contain only one data element. (Are atomic.)

- A table is 2NF if it is 1NF, and:
  - No subset of the PK is determinant for one or more other columns.
    ("For PKs of 2+ columns, no dependency to only some of those columns.")

- A table is 3NF if it is 2NF, and:
  - No non-PK attribute has a transitive dependency to a PK attribute.
    (*Transitive dependency:* A → B and B → C => A → C.)
    (*In plain English: "*No dependency to columns outside of the PK.")

Sum-up: 1NF to 3NF *(repetition)*



Format for showing normalization process

- A quite standard way of displaying normalization processes is with the following format:
  - (This is also used in the Norwegian textbook.)

- `Table1 (column1*, column2, column3, …, columnN)`
  `Table2 (column1, column2, column3, …, columnN*)`
  - PKs have an underline.
  - FKs end with an asterisk (*).

Exercises task 1

- From students, regarding task 1:
  - *"jeg lurte på hvordan det er best å gjennomføre task1 om normalisering fra uke 9 rent praktisk. Er det best å legge all dataen manuelt inn i excel og gjøre det slik du gjorde det i forelesningen, skal vi legge det inn i mysql, gjøre det med penn og papir eller hva? Lurte også da på hvordan det skal gjøres i en eventuell eksamen"*
  - *"sliter med å dra forhold mellom tabellene"*

- Will do a sum-up of task 1, focusing on the questions above.
  - For a table walkthrough, see the video (21 min.) shared on Canvas. :-)

Exercises task 1 + 4

- Task 1 solved on the textbook format: *(also in the solution pdf)*
  - `Grade (coursecode*, studentnumber*, examdate*, examresult)`
  - `Course (coursecode, coursetitle)`
  - `Student (studentnumber, studentname)`
  - `Exam (coursecode*, examdate, examregistrations)`

- From students, regarding task 4:
  - *"Kunne du forresten gått igjennom oppgave 4"*

- Will show task 4.
  - But not sure if we have time for all the details.

Normalization

part 2

Boyce-Codd normal form

- Boyce-Codd normal form (BCNF) is a further development of 3NF. (It is between 3NF and 4NF.)
  - *Note:* BCNF is normally met when 3NF is met!

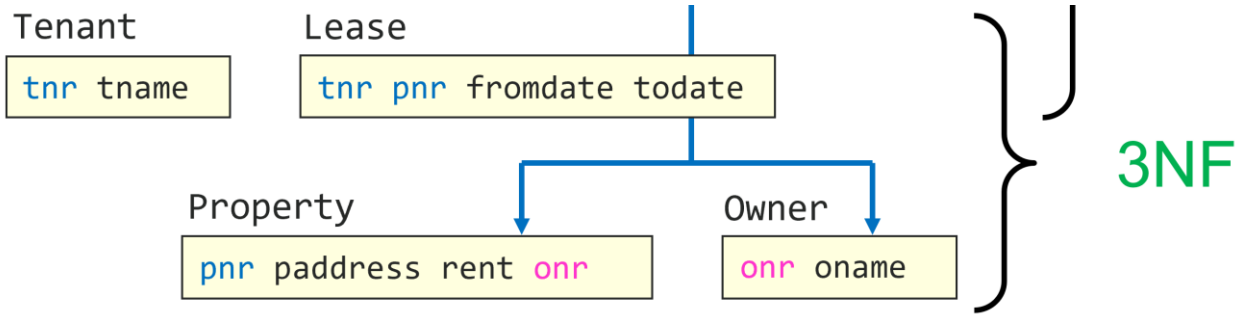
- A table is on BCNF if:
  - All minimal determinants are candidate keys.
  - (Can also be said as: All determinants are superkeys.)

- *In plain (plainer?) English:*
  - BCNF: *No determinant value can be repeated for 2+ rows in the table.*

Example 1, BCNF

- BCNF rule:
  - All minimal determinants are candidate keys.

- Lets look at the tables from last weeks example:



  - The tables Tenant, Property and Owner are BCNF: Each of them has their PK (a candidate key) as their only determinant.
  - But what about Lease? Let's find out. (See next slide.)

Example 1, BCNF – cont.

Lease

tnr	pnr	fromdate	todate
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

- Before we find the determinants, we need to know the business rules:
  - A tenant can only rent each property once.
  - A tenant can only rent one property at a time.
  - *Task:* How many (and which?) Determinants do we have here?

Example 1, BCNF – cont.

- 5 determinants in the table:
  - tnr, pnr → fromdate, todate
  - tnr, fromdate → pnr, todate
  - tnr, todate → pnr, fromdate
  - pnr, fromdate → tnr, todate
  - pnr, todate → tnr, fromdate

Lease

tnr	pnr	fromdate	todate
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

- Conclusion:
  - All determinants are candidate keys.
  - Ergo, last lessons example is not only on 3NF, but also on BCNF.

Example 2, BCNF

- We are *kind of* continuing from last lesson's example: Leasing properties.

- But now we'll look at the interview process. Prerequisites:
  - A tenant is never interviewed more than once per day. (But can be interviewed several times over several days.)
  - Each employee holds all their interviews for a given day in a single room. (The room can be used by other staff the same day as long as it is available.)

Interview_Location

tnr	date	time	enr	room
76	13-MAY-11	10:30	5	101
56	13-MAY-11	12:00	5	101
74	13-MAY-11	12:00	7	102
56	01-JUL-11	10:30	5	102

Example 2, BCNF – cont.

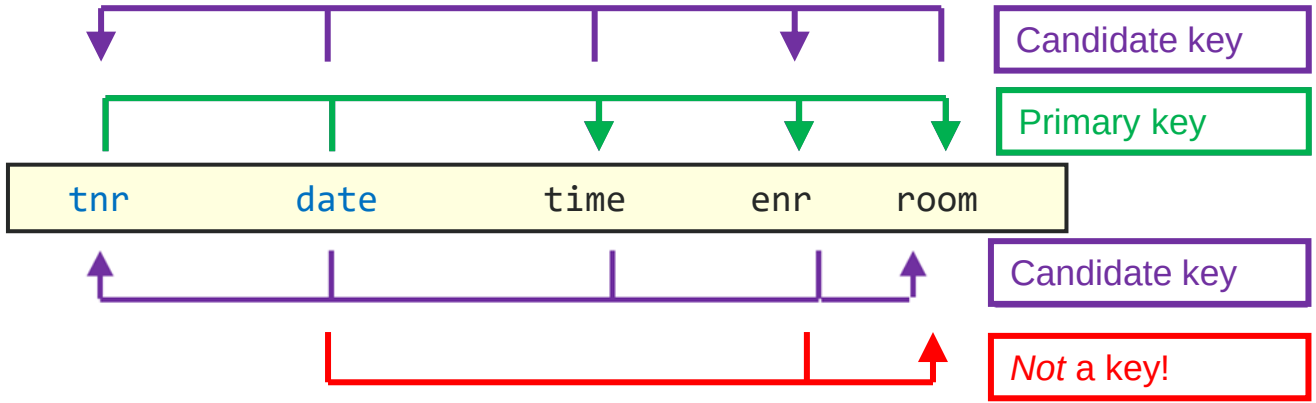
Interview_Location

tnr	date	time	enr	room
76	13-MAY-11	10:30	5	101
56	13-MAY-11	12:00	5	101
74	13-MAY-11	12:00	7	102
56	01-JUL-11	10:30	5	102

- Which 3 candidate keys ("minimal determinants") do we have here?
  - And can you also find a 4th *minimal* determinant?

Example 2, BCNF – cont.

- A table is on BCNF if:
  - All minimal determinants are candidate keys.



- We got: PK + 2 candidate keys,
  1 other determinant. (← Thus, *not* BCNF!)

Example 2, BCNF – cont.

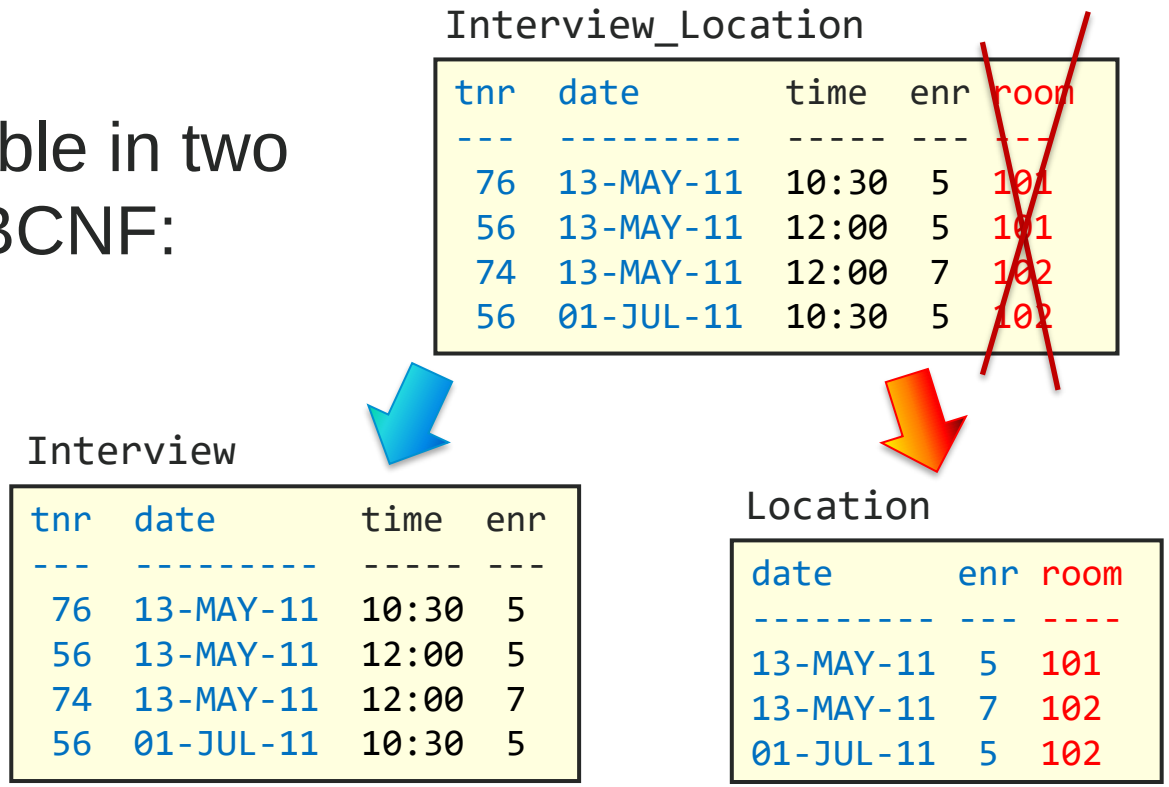
- PK and the 2 candidate keys are fine in BCNF.

- The 4th determinant, interview_date, employee → room, is not approved in BCNF: It is *not* a candidate key.
  - We can have several interviews done by the same employee during one day. This is a determinant, but it is not unique data, thus not a candidate key.

- We have to split the table in two to achieve BCNF.

Example 2, BCNF – cont.

Splits the table in two to achieve BCNF:



Denormalization

We are DONE with the theory on normalization! :-D

- But we'll do a really quick explanation (3 slides) on the opposite:
  - Denormalization
- Denormalization means:
  - Altering tables so that normalization is reduced by one degree or more.
  - *Example:* Going from BCNF / 3NF to 2NF.
- The reason for denormalization is usually that our JOINs take time:
  - We can make our SELECTs faster if they do not contain JOINs.

Denormalization – cont.

- Benefits of denormalization:
  - You avoid linking tables (fewer joins).
  - The speed of the database can therefore increase when looking up large amounts of data.

- Disadvantages:
  - Implementation becomes more difficult.
  - Double storage (redundancy).
  - Slower storage / updating.
  - Flexibility decreases.

Denormalization – cont.

- So when is it appropriate to denormalize? Very seldom!
  - When the system can not meet its query performance requirements with a normalized database.
  - (When the database is used for a lot more SELECTS than updates.)

- *Note:* Denormalization is the exception to the rule!
  - Both in this subject and in working life, plan for 3NF (or BCNF) unless you are specifically told otherwise!

Today's exercises & looking ahead

- Now: 2 hours of exercises.

- Exercises are on Canvas, as usual. Short summary:
  – Continue *where you left off* on the normalization exercises.

- Main contents for the next lesson:
  – Several smaller topics. (The last lesson with *new* content! :-D)

The