NORMAL FORMS 2

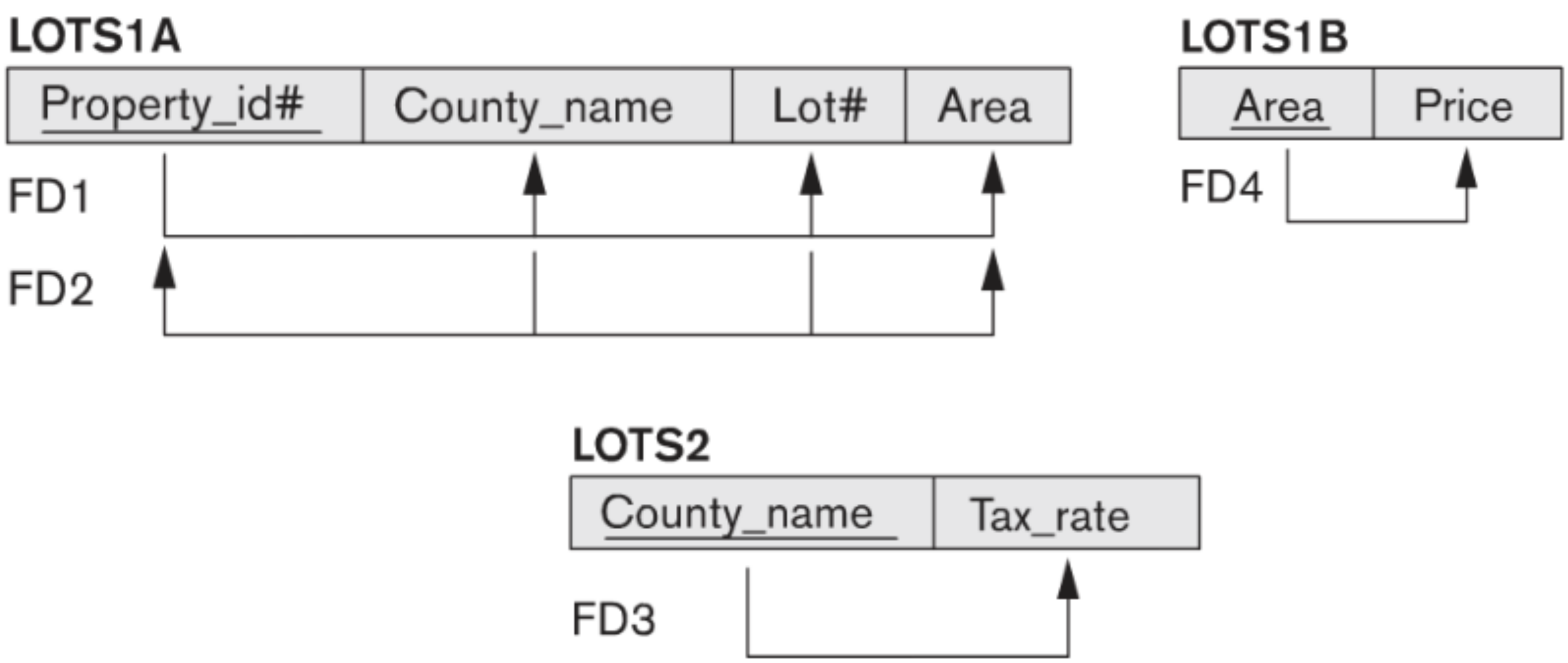
Corresponding Reading: Chapter 14.5-14.8

Third Normal Form (3NF)

- General Definition:
  - A relation schema R is in 3NF if, whenever a nontrivial functional dependency X --> A holds in R, either
    - (a) X is a superkey of R, **OR**
    - (b) A is a prime attribute of R.

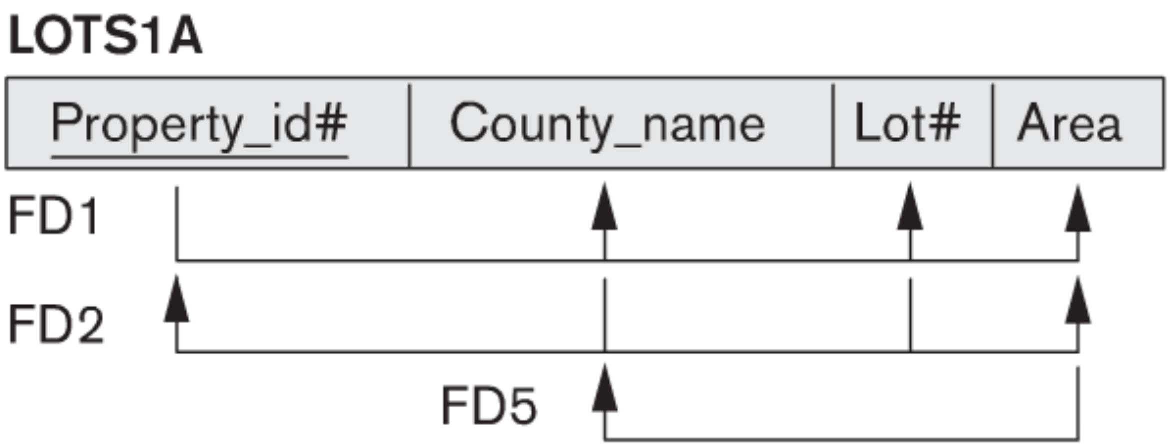
Boyce-Codd Normal Form (BCNF)

- BCNF was originally proposed to be a simpler form of 3NF, but it was found to be actually stricter than 3NF.
- Every relation in BCNF is also in 3NF.
- Let's re-examine LOTS Relation to understand BCNF:



LOTS Relation

- Suppose that we have 1000s of lots, but the lots are only from 2 counties: Providence and Portsmouth.
  - The lot sizes in Providence are only 0.5, 0.6, …1.0 acres.
  - The lot sizes in Portsmouth are only 1.1, 1.2, … 2.0 acres.
- In this situation, we would have an additional functional dependency FD5: Area --> County_name



  - LOTS1A would still be in 3NF, as County_name is a prime attr.

LOTS Relation

- Problem: As the number of lots grow, we will have redundant information in our table. There are only 16 possible lot sizes, so why don't we just store the lot sizes in a separate relation? R(Area, County_name)

LOTS1A

Property_id#	County_name	Lot#	Area

LOTS1AX

Property_id#	Area	Lot#

LOTS1AY

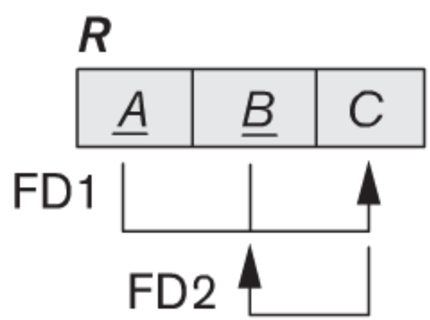
Area	County_name

BCNF

- A relation schema R is in BCNF if whenever a nontrivial functional dependency X-->A holds in R, then X is a superkey of R.
- This differs from 3NF which allows A to be prime.
  - FD5 satisfies 3NF in LOTS1A because County_name is a prime attribute.
  - FD5 violates BCNF in LOTS1A because Area is not a superkey of LOTS1A.
- In order to obtain BCNF, we create LOTS1AX and LOTS1AY. We lose FD2 because its attributes no longer coexist in the same relation after decomposition.

BCNF

- In practice, most relation schemas that are in 3NF are also in BCNF.
  - If the following condition exists, then the relation will not be in BCNF:
    - If X-->A holds in a relation schema R with X is **not** a superkey and A is a prime attribute of R.
- The following example illustrates this general case:



- Ideally, we should strive to achieve BCNF or 3NF.

Example: TEACH

- TEACH relation:

TEACH

Student	Course	Instructor
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Omiecinski
Zelaya	Database	Navathe
Narayan	Operating Systems	Ammar

Functional Dependencies:

FD1: {Student, Course} --> Instructor

FD2: Instructor --> Course (Each instructor teaches on course)

Candidate Key: {Student, Course}

Example: TEACH

- This example follows the general case:
  - Student as A, Course as B, and Instructor as C.
  - This relation is in 3NF, but not BCNF.
- Decomposition is not straightforward because it may be decomposed multiple ways:
  - 1) {Student, Instructor} and {Student, Course}
  - 2) {Course, Instructor} and {Course, Student}
  - 3) {Instructor, Course} and {Instructor, Student}
- All three lose FD1, but #3 is best because it won't generate spurious tuples after a JOIN.

Example: TEACH

- Nonadditive Decomposition is important during normalization.
  - You may have to forgo preservation of all FDs.
    - We lose FD1, but the relations are in BCNF.

- {Instructor, Course} and {Instructor, Student} is nonadditive

Multivalued Dependency

- Relations may have constraints that cannot be specified as functional dependencies.
- Multivalued Dependency (MVD)

EMP

Ename	Pname	Dname
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

  - A tuple in this relation represents the fact that an employee works on a project and has a dependent.
  - An employee may work on several projects and have several dependents, which are independent.

Multivalued Dependency (MVD)

- A separate tuple to represent every combination of an employee's dependent and an employee's project.
- This is a multivalued dependency.
  - Denoted as:

$$\text{Ename} \twoheadrightarrow \text{Pname}$$

$$\text{Ename} \twoheadrightarrow \text{Dname}$$

EMP

<u>Ename</u>	<u>Pname</u>	<u>Dname</u>
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John

- Whenever two independent 1:N relationships A:B and A:C are mixed in the same relation R(A,B,C), an MVD may arise.

$$A \twoheadrightarrow B|C$$

Multivalued Dependency

- Formal Definition:
  - A multivalued dependency $X \twoheadrightarrow Y$ specified on relation schema R, where X and Y are both subsets of R, specifies the following constraint on any relation state r of R:
    - If two tuples t_1 and t_2 exist in r such that $t_1[X] = t_2[X]$, then two tuples t_3 and t_4 should also exist in r with the following properties, where Z is $(R - (X \cup Y))$
      - $t_3[X] = t_4[X] = t_1[X] = t_2[X]$
      - $t_3[Y] = t_1[Y]$ and $t_4[Y] = t_2[Y]$
      - $t_3[Z] = t_2[Z]$ and $t_4[Z] = t_1[Z]$

Multivalued Dependency

- **Summary:**
  - Whenever $X \twoheadrightarrow Y$ holds, we say that X multidetermines Y.
  - Whenever $X \twoheadrightarrow Y$ holds in R, so does $X \twoheadrightarrow Z$.
  - Since $X \twoheadrightarrow Y$ and $X \twoheadrightarrow Z$, we can write: $X \twoheadrightarrow Y|Z$

- **Trivial MVD:** Z= $(R - (X \cup Y))$
  - An MVD $X \twoheadrightarrow Y$ in R is called a trivial MVD if:
    - (a) Y is a subset of X or
    - (b) The union of X and Y = R
  - Trivial MVDs do not specify any significant or meaningful constraints
- **Non-Trivial MVD:**
  - An MVD that satisfies neither (a) nor (b) is called nontrivial.

Multivalued Dependency

- Example: EMP_PROJECTS
  - Trivial MVD: Ename $\twoheadrightarrow$ Pname
- If there is a trivial MVD, then we will repeat values redundantly in tuples.
  - Undesired, however EMP is in BCNF because no functional dependencies hold in EMP
- We need a new normal form to prevent these situations!
  - Fourth Normal Form: Disallows relation schemas like EMP
- Notice: Relations that have nontrivial MVDs tend to be all-key-relations (their key is all their attributes together).

EMP_PROJECTS

Ename	Pname
Smith	X
Smith	Y

Fourth Normal Form (4NF)

- 4NF is violated when a relation has undesirable multivalued dependencies and can be used to identify and decompose the relation.

- Formal Definition:
  - A relation schema R is in 4NF with respect to a set of dependencies F (that includes functional dependencies and multivalued dependencies) if, for every nontrivial multivalued dependency $X \twoheadrightarrow Y$ in F^+, X is a superkey for R.

- The EMP relation is an all-key relation in BCNF, but not in 4NF since it contains the MVD

$$\text{Ename} \twoheadrightarrow \text{Pname} \mid \text{Dname}$$

Fourth Normal Form

- A relation that is not in 4NF due to a nontrivial MVD must be decomposed to convert it into a set of relation in 4NF.
- Decomposition removes the redundancy caused by the MVD.
- Normalization process for 4NF consists of decomposing the relation so that each nontrivial MVD is represented by a separate relation where it becomes a trivial MVD.

4NF - Example

- EMP is not in 4NF because of nontrivial MVDs

Ename $\twoheadrightarrow$ Pname Ename $\twoheadrightarrow$ Dname. and Ename is not a superkey

- Decompose EMP into EMP_PROJECTS and EMP_DEPENDENTS:

EMP_PROJECTS

Ename	Pname
Smith	X
Smith	Y

EMP_DEPENDENTS

Ename	Dname
Smith	John
Smith	Anna

- These relations are in 4NF as the MVDs are now trivial.
  - Also, there are no Functional Dependencies in these relations.

Fifth Normal Form (5NF)

- Normalization into 5NF is very rarely done in practice.
  - Only used for theoretical schema analysis.

- General Idea:
  - Achieving 4NF involves eliminating MVDs by repeated binary decompositions that are nonadditive.
  - In some cases, there may be no nonadditive decompositions into two relation schemas.
  - There may be nonadditive decompositions into more than 2 relation schemas.
  - We can then produce a multiway decomposition into 5NF.