Unit 4: Database Design and Development

Unit code : H/615/1622

Unit type : Core

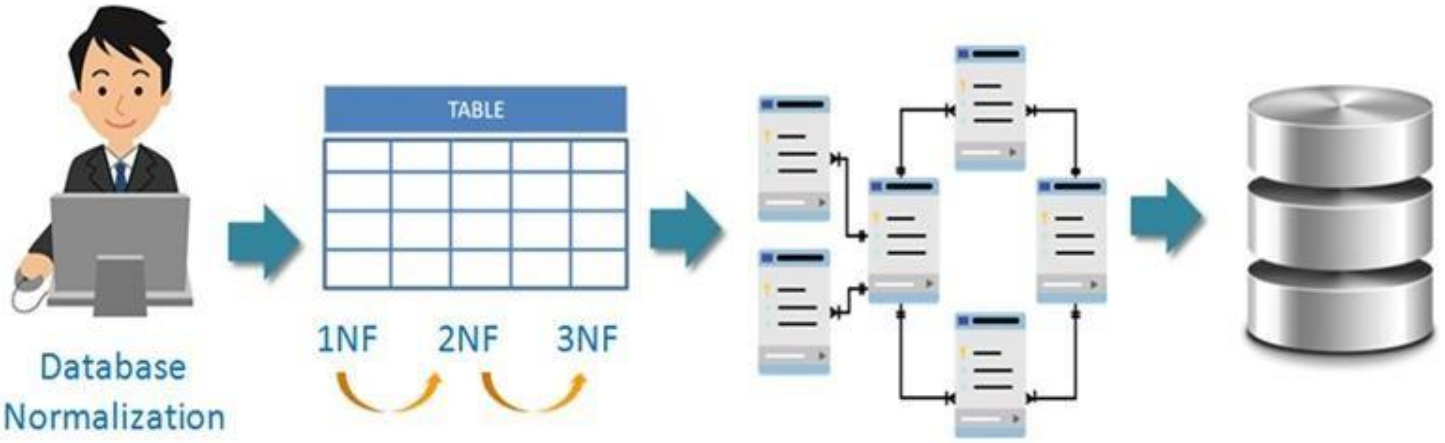
Credit value : 15

Lecture 04 (2 hrs)

Lecturer Name: Ms. Chathurangi D. Weerasinghe MSc(UCSC, Col), BSc(Ruh)

Approved Centre from Pearson

Normalization



Ms. Chathurangi D. Weerasinghe
MSc(UCSC, Col), BSc(Ruh)

Approved Centre from Pearson

Database Design process

1. Requirements Analysis

   What does the user want?

2. Conceptual Database Design

   Defining the entities and attributes, and the relationships between these --> The ER model

3. Logical Database Design (Map ER to Relational Schema)

4. **Schema Refinement**

5. Physical Database Design

   Implementation of the design using a Database Management System

Normalization

- Conceptual Modeling is a subjective process
- Therefore, the schema after the logical database design phase may not be very good (contain redundant data)
- However, there are formalisms to ensure that the schema is good.
- This process is called Normalization

Normalization

- Relational database schema = set of relations
- Relation = set of attributes
- How we group the attributes to relations is very important
- Normalization or Schema Refinement help determine "GOOD" relations

Purpose of Normalization

- To avoid redundancy by storing each ‘fact’ within the database only once.
- To put data into a form that conforms to relational principles (e.g., single valued attributes, each relation represents one entity) - no repeating groups.
- To put the data into a form that is more able to accurately accommodate change.
- To avoid certain updating ‘anomalies’.
- To facilitate the enforcement of integrity constraints.

Redundancy and Data Anomalies

Redundant data is where we have stored the same 'information' more than once. i.e., the redundant data could be removed without the loss of information.

Example: We have the following relation that contains staff and department details:

staffNo	job	dept	dname	city
SL10	Salesman	10	Sales	Stratford
SA51	Manager	20	Accounts	Barking
DS40	Clerk	20	Accounts	Barking
OS45	Clerk	30	Operations	Barking

ffNo	job	dept	dname	city
SL10	Salesman	10	Sales	Stratford
SA51	Manager	20	Accounts	Barking
DS40	Clerk	20	Accounts	Barking
OS45	Clerk	30	Operations	Barking

Such ‘redundancy’ could lead following ‘anomalies’

Insert Anomaly: We can’t add a new a dept without inserting a member of staff that works in that department

Update Anomaly: Change the name of the Accounts dept to Finance dept. We have to change all other records to avoid update anomaly.

Deletion Anomaly: Employee SL10 resigns. We remove the record. With that we lose all information pertaining to the Sales dept.

Repeating Groups

Is an attribute (or set of attributes) that can have more than one value

staffNo	job	dept	dname	city	contact number
SL10	Salesman	10	Sales	Stratford	018111777, 018111888, 079311122
SA51	Manager	20	Accounts	Barking	017111777
DS40	Clerk	20	Accounts	Barking	
OS45	Clerk	30	Operations	Barking	079311555

Repeating Groups are not allowed in a relational design, since all attributes have to be 'atomic' - i.e., there can only be one value per cell in a table!

Schema Refinement

A relation with redundancy can be refined by

Decomposing the relation into smaller relations…

- contain the same information
- with no redundancy

Formal Process

Formal process for good relational schema:

- To avoid the above mentioned issues in the relational schema, we can apply a formal process called Normalization
- Normalization is based on functional dependencies

Functional Dependencies

- FDs are used to specify *formal measures* of the "goodness" of relational designs
- FDs and keys are used to define **normal forms** for relations

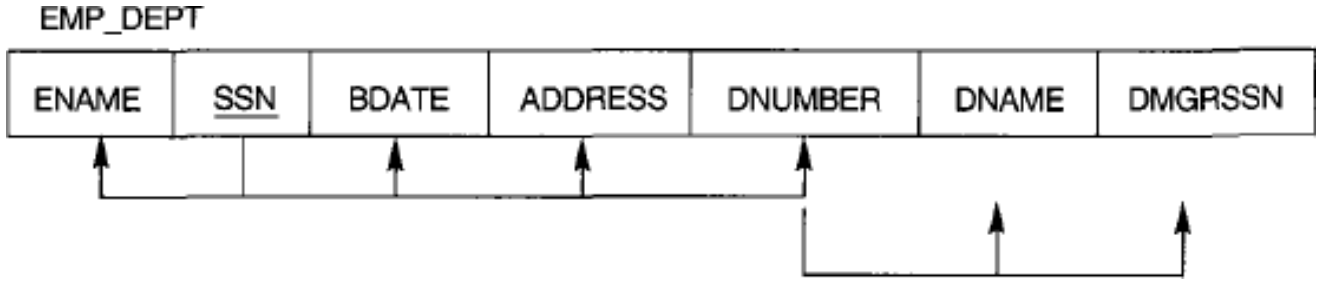
Functional dependency

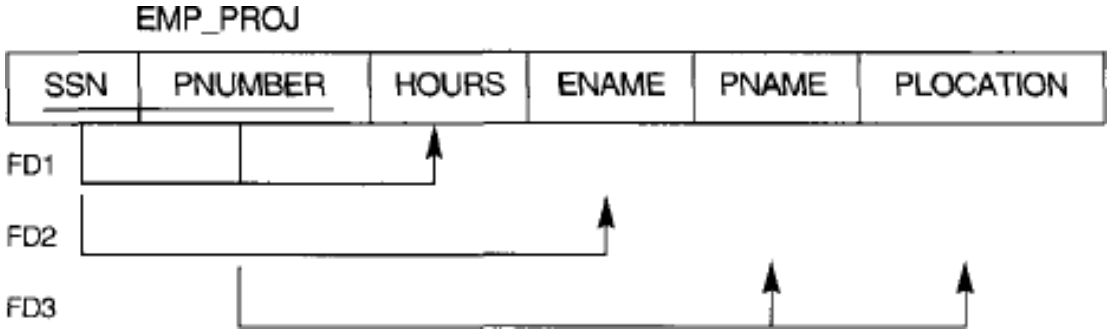
- A functional dependency, is a constraint between two sets of attributes
- denoted by X → Y,
  - X functionally determines Y
  - Y is functionally dependent on X

Functional dependency

- Describes the relationship between attributes in a relation.
- If A and B are attributes of relation R, B is functionally dependent on A (denoted A → B), if each value of A is associated with exactly one value of B.
- (A and B may each consist of one or more attributes)

Functional dependency





(Ssn,Pnumber) -> Hours (SSN & PNUMBER detrmines hrs emp work on a project)

Ssn -> Ename

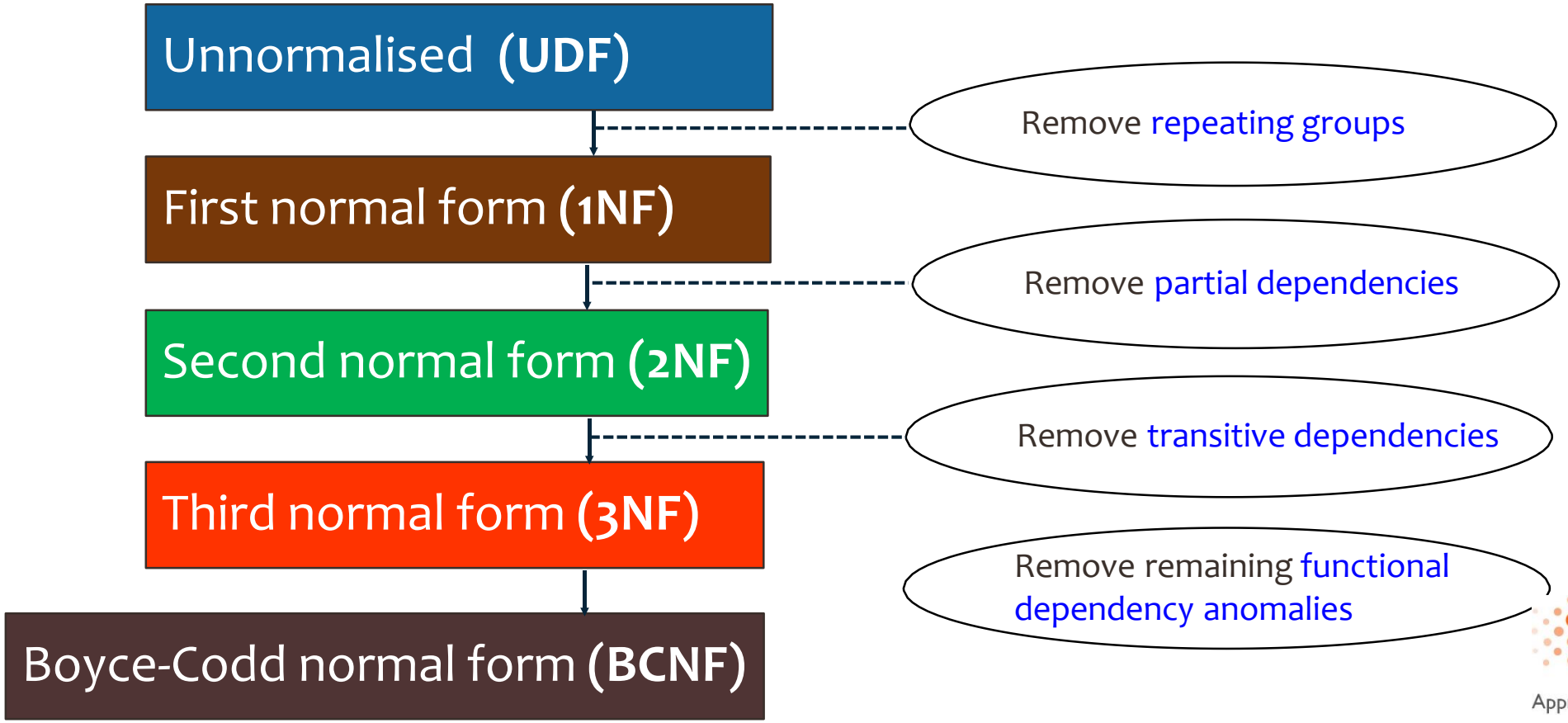
Pnumber -> (pname, plocation) (PNUMBER determines pname & location)

Database basics - review

- **Candidate Key:** Each key of a relation is called a candidate key
- **Primary Key:** A candidate key is chosen to be the primary key
- **Prime Attribute:** an attribute which is a member of a candidate key
- **Nonprime Attribute:** An attribute which is not prime

Stages of Normalization

- There are many Normal Forms proposed to reduce redundancies



Un-normalized Normal Form (UNF)

- A relation is un-normalized when it has not had any normalization rules applied to it, and it suffers from various anomalies

Normalization - 1st Normal Form

For example:

DEPARTMENT (Dname, Dnumber, DMGRSSN, (DLocation))

DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATIONS
Research	5	333445555	{Matara, Kandy, Colombo}
Administration	4	987654321	{Malabe}
Headquarters	1	888665555	{Colombo}

- Department is in UNF
- Department relation not in 1NF
- How to take into 1NF ?

Solution

: Create a separate DEPT_LOCATION relation with foreign key

DEPT_LOCATIONS

DNUMBER	DLOCATIONS
1	Colombo
4	Malabe
5	Matara
5	Kandy
5	Colombo

DEPARTMENT

DNAME	DNUMBER	DMGRSSN
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

- Remove the attribute DLOCTION and place it in a separate relation DEPT_LOCTIONS along with the primary key DNUMBER of DEPARTMENT.
- The PK is the combination {DNUMBER, DLOCTION}
- This decompose the non-INF relation into two INF relation.

Normalization – 2nd Normal Form

- A relation R is in second normal form (2NF) if every nonprime attribute A in R is **not partially dependent** on any key of R
- Remove partial functional dependencies into a new relation

Example: Not in 2NF

TEACHER	CAMPUS	COURSE	ADDRESS
Kapila	**Colombo**	**Database**	**BoC Merchant Tower**
Nuwan	**Malabe**	**Database**	**New Kandy Road**
Samantha	**Colombo**	**Operating Systems**	**BoC Merchant Tower**
Kapila	**Malabe**	**Operating Systems**	**New Kandy Road**

Partial dependency

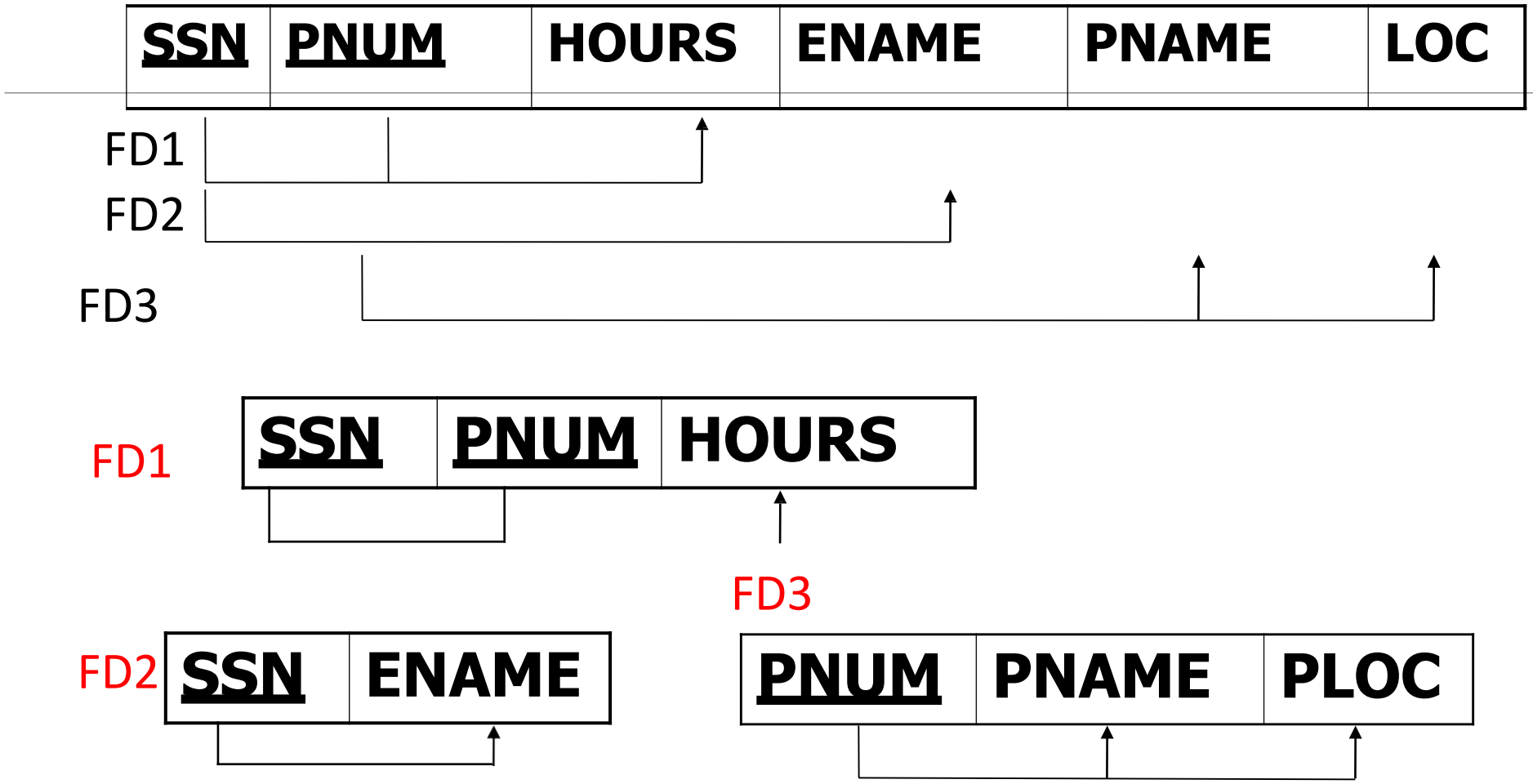
Normalization – 2nd Normal Form

Example: After normalized into 2NF

TEACHER	CAMPUS	COURSE
Kapila	**Metro**	**Database**
Nuwan	**Malabe**	**Database**
Samantha	**Metro**	**Operating Systems**
Kapila	**Malabe**	**Operating Systems**

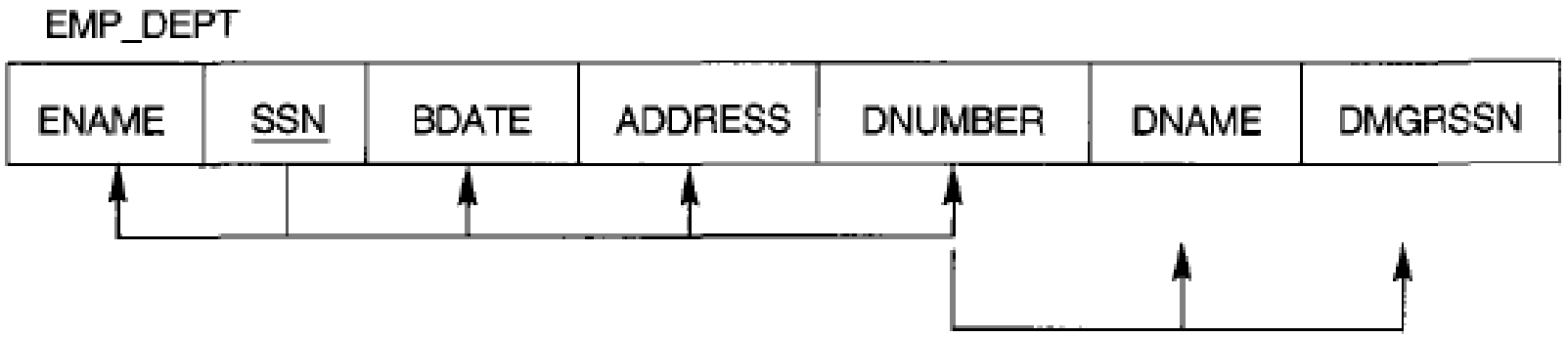
CAMPUS	ADDRESS
Metro	**BoC Merchant Tower**
Malabe	**Malabe Campus**

EMP_PROJ


SSN
PNUM
HOURS
ENAME
PNAME
LOC
FD1
FD2
FD3
FD1
SSN
PNUM
HOURS
FD3
FD2
SSN
ENAME
PNUM
PNAME
PLOC

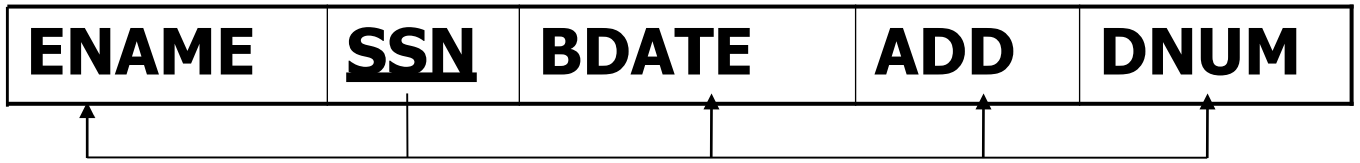
Normalization – 3rd Normal Form

- A relation R is in 3rd normal form (3NF) if every
  - R is in 2NF, and
  - No nonprime attribute is transitively dependent on any key
  - Remove transitive dependencies into a new relation

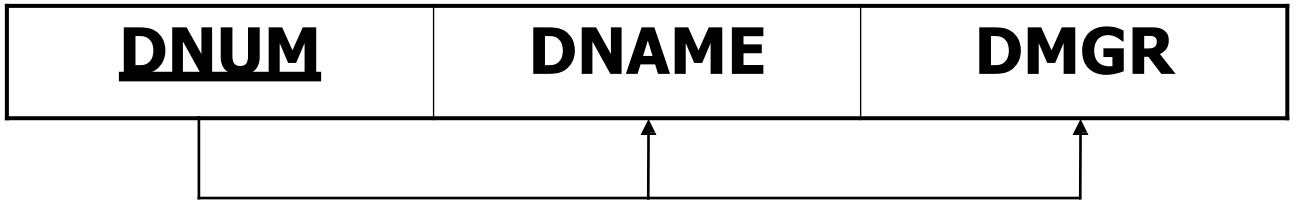


Normalization – 3rd Normal Form

FD1



FD2



Boyce - Codd Normal Form

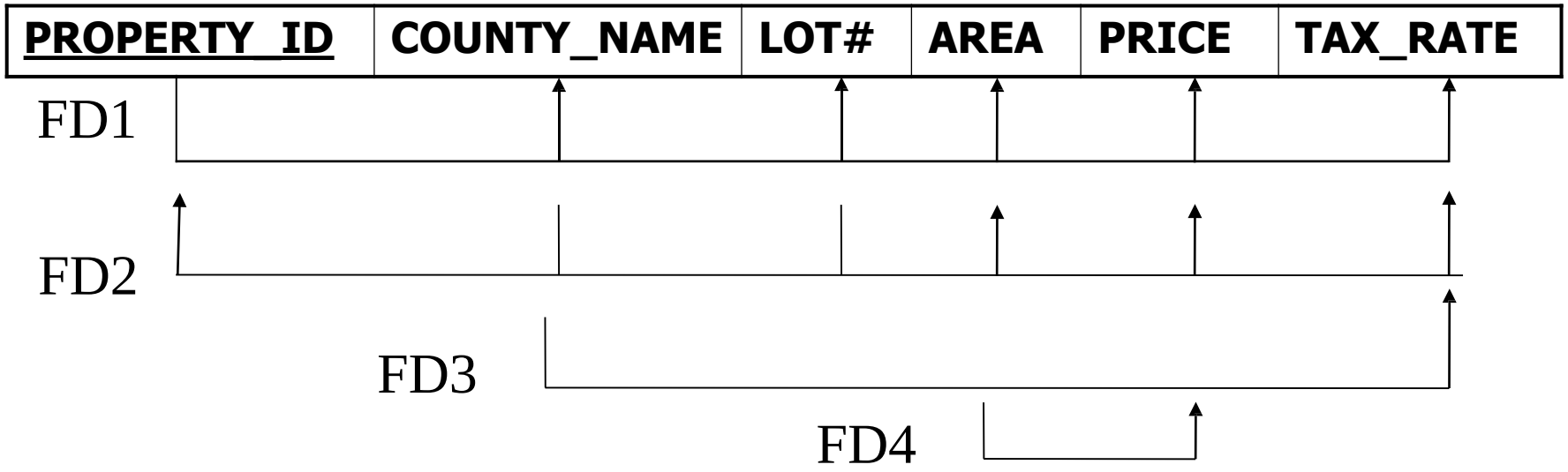
A relation schema is in Boyce- Codd Normal Form

- If every nontrivial functional dependency X→A hold in R, then X is a super key of R
- A relation is in BCNF if and only if, every determinant is a candidate key
- Every relation in BCNF is also in 3NF

- A relation is in BCNF, if and only if, every determinant is a candidate key

Approved Centre from Pearson

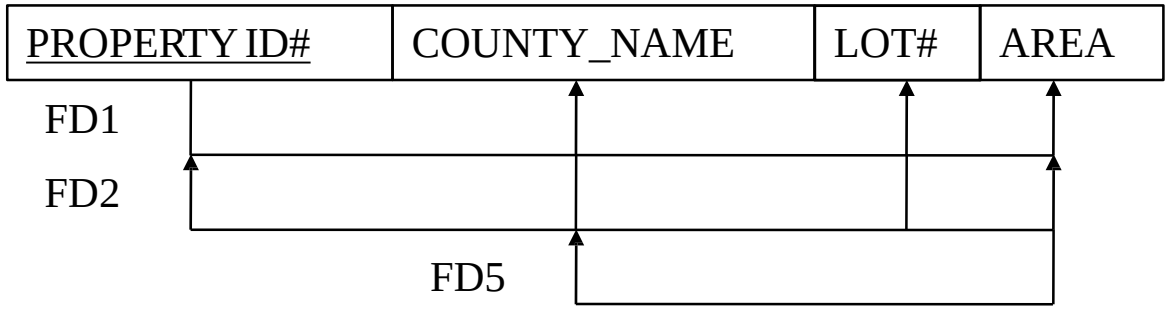
Normalization –(BCNF)



Keys: PropertyID, (County_Name, Lot#)

Normalization –BCNF

(a) LOTS1A



BCNF Normalization

LOTS1AX

PROPERTY ID#	AREA	LOT#

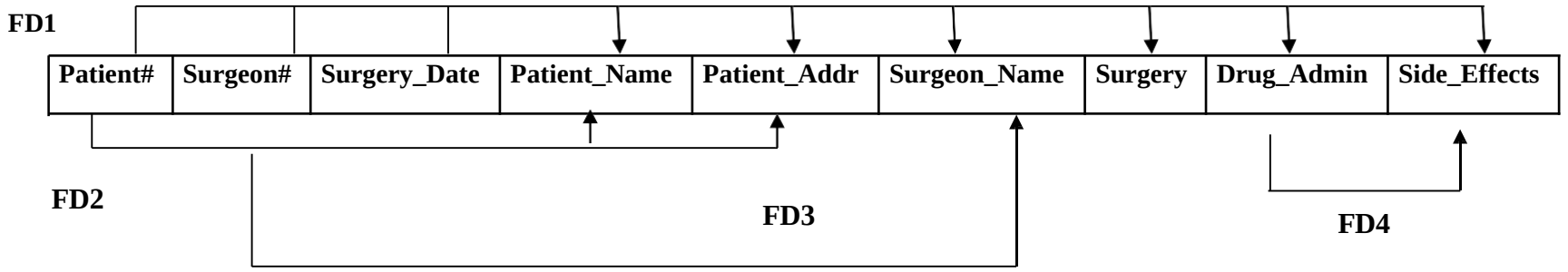
LOTS1AY

AREA	COUNTY_NAME

Approved Centre from Pearson

Exercise 1

Patient #	Surgeon #	Surgery Date	Patient Name	Patient Addr	Surgeon Name	Surgery	Drug admin	Side Effects
1111	311	12-Jun-95	John White	15 New St. New York, NY	Michael Diamond	Kidney stones removal	none	rash
1234	243	05-Apr-94	Mary Jones	10 Main St. Rye, NY	Charles Field	Eye Cataract removal	Tetracyclin e	Fever
1234	467	10-May-95	Mary Jones	10 Main St. Rye, NY	Patricia Gold	Thrombos is removal	none	none
2345	189	08-Jan-96	Charles Brown	Dogwood Lane Harrison, NY	David Rosen	Open Heart Surgery	Cephalosp orin	none



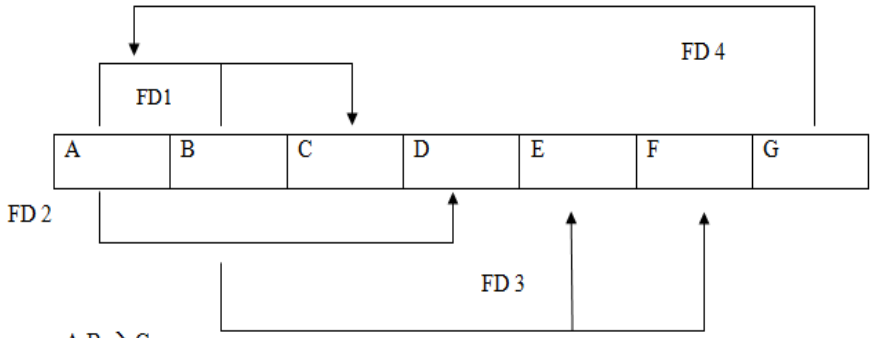
1. What normal form is the relation in?
2. If it is not in 3NF, convert it to 3NF. Explain your answer.

Exercise 2

Consider the following relational schema for R:

R(<u>A, B</u>,C,D,E,F,G)

AB is the primary key in the relation. Assume that the following dependencies exist:



A B → C

A → D

B → E F

G → A

1. What normal form is the relation in?
2. If R is not in BCNF, convert it to BCNF. Explain your answer.

Thank you!