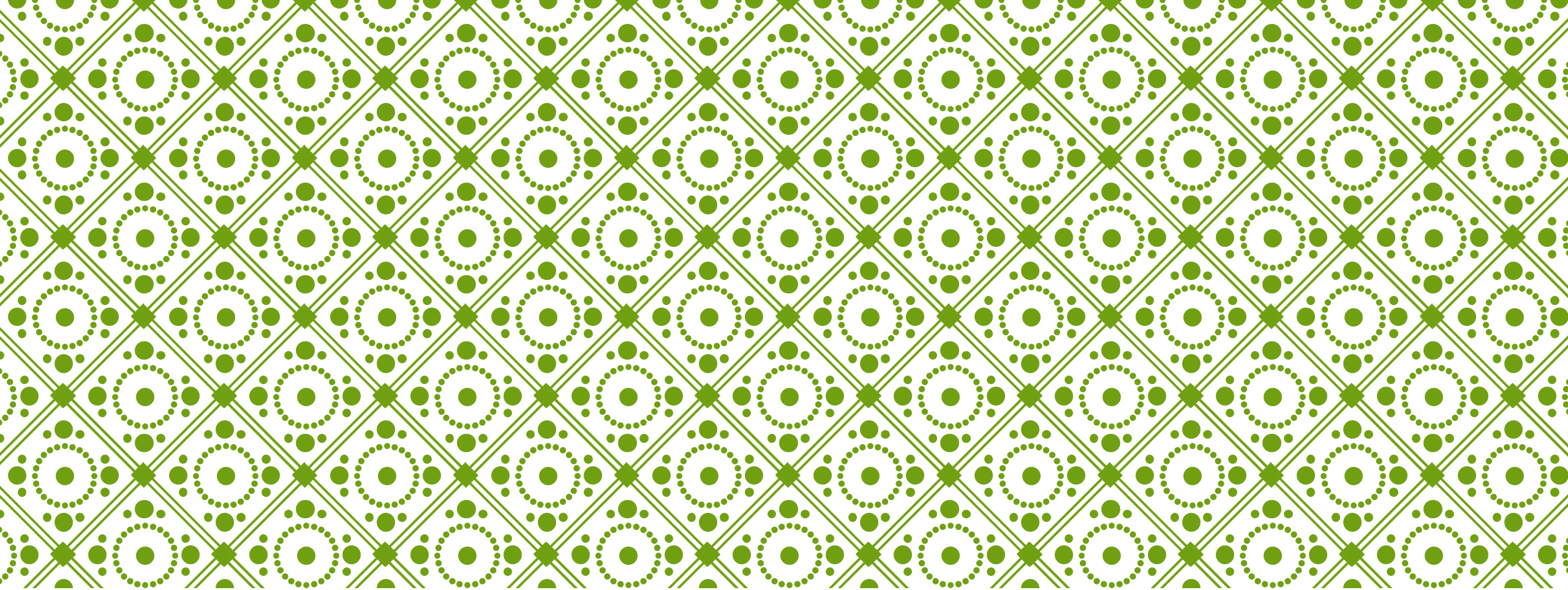


DATABASE NORMALIZATION

Vivek M R
Research Scholar
NIT Calicut

NORMALIZATION

Process of structuring relational schema design

Based on normal forms

Proposed by Edgar F. Codd (1972)

Advantages

- Reduces data redundancy and inconsistent dependency
- Improves data Integrity
- Minimized redesign while extending structure of database
- Minimize update anomalies

Contents & Examples based on
Elmasri & Navathe, Fundamentals of Database Systems, 6th Edition, Pearson. (2011). – Chapter 15

DESIGN GUIDELINES FOR RELATIONAL SCHEMAS

Do not combine attributes from multiple entities into single relation

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn

Contains attributes of an employee along with attributes of department(Dname, Dmgr_ssn)

Can be moved to another relation like DEPT_DETAILS with attributes (Dnumber, Dnane, Dmgr_ssn)

Avoid storing redundant information storage in tuples

Redundancy

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn

EMPLOYEE F.K.

Ename	Ssn	Bdate	Address	Dnumber

P.K.

DEPARTMENT F.K.

Dname	Dnumber	Dmgr_ssn

P.K.

UPDATE ANOMALIES

Caused due to storage of natural joins of base relations

EMP_DEPT

Ename	Ssn	Bdate	Address	Dnumber	Dname	Dmgr_ssn

Insertion anomalies

- e.g. Entering a new department with no employee - inserts NULL value for Ename, Ssn, Address
- e.g. Enter new employee not assigned to a dept. - insert department details as NULL

Deletion anomalies

- e.g. Delete last employee of a dept. - dept. information is lost

Modification anomalies

- e.g. Change manager of a particular dept - Need to change all employee tuples

So, design base relations without any update anomalies

Avoid using attributes having NULL value frequently

NULL occurs

- attribute does not apply
- value is unknown
- value is not entered

Can cause issues while using aggregate functions like COUNT, SUM.

- How to interpret NULL

Design relation schemas so that they can be joined with equality conditions on attributes that are appropriately related

- Between (primary key, foreign key) pairs
- Else duplicate tuples will be produced

FUNCTIONAL DEPENDENCIES

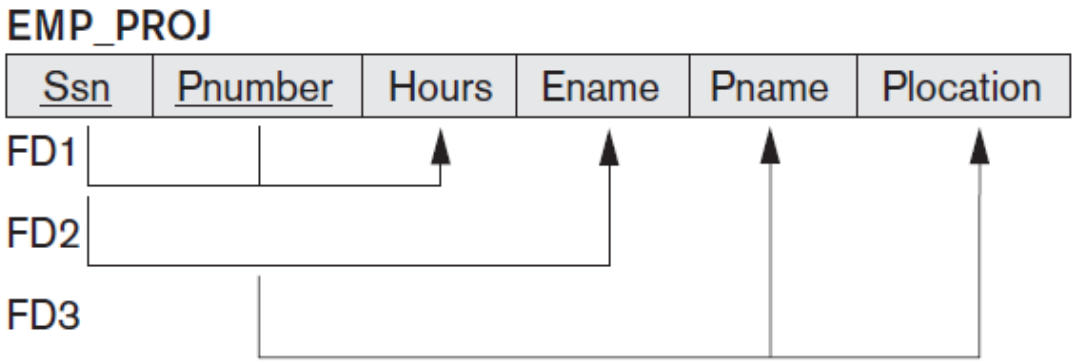
Denoted by **X → Y**

Defined between two sets of attributes X and Y that are subsets of R

Defines relationship/dependencies among attributes in a relation

X functionally determines Y OR Y is functionally dependent on X

e.g. Pnumber →{Pname, Plocation} - The value of a project's number (Pnumber) uniquely determines the project name (Pname) and location (Plocation)



NORMAL FORM

Normal form of a relation refers to the **highest** normal form condition that it meets, and hence indicates the **degree** to which it has been normalized.

e.g. 1NF, 2NF, 3NF, BCNF, 4NF, 5NF etc.

A relation in 3NF satisfies 1NF, 2NF and 3NF.

FIRST NORMAL FORM (1NF)

A relation is in INF if it does not contain composite, multi-valued attributes or their compositions.

Attribute values permitted by 1NF are single atomic (or indivisible) values.

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocations
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

DEPARTMENT

Dname	Dnumber	Dmgr_ssn	Dlocation
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Sol 1: *with redundancy

Sol 2: Decompose the relation into two 1NF relations, with violating attribute in another table along with primary key from original relation.

Sol 3: If a maximum number of values is known for the attribute e.g., if it is known that at most three locations can exist for a department - replace the Dlocations attribute by three atomic attributes: Dlocation1, Dlocation2, and Dlocation3.

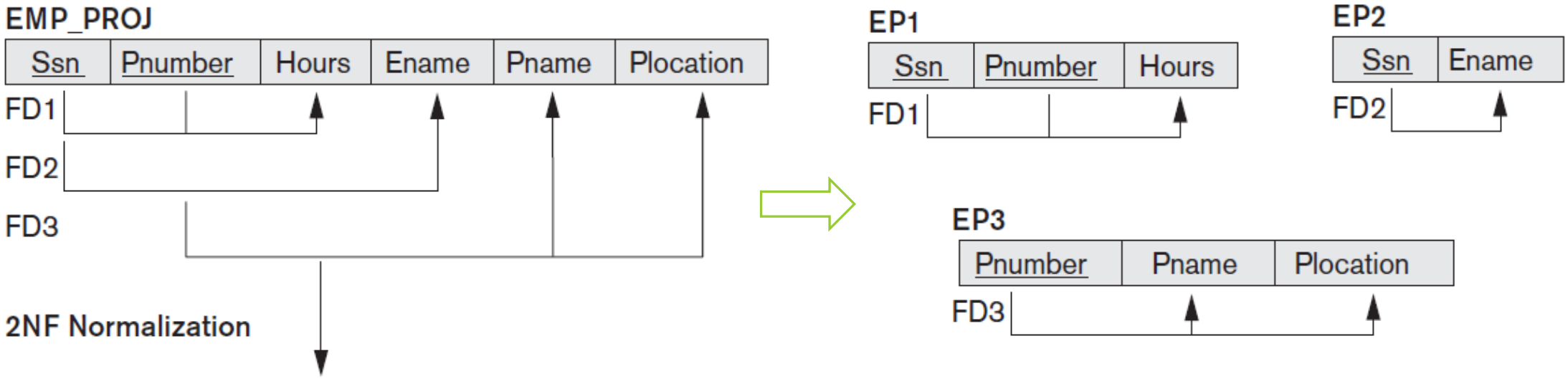
- NULL values in some places
- Search becomes issue (location can be under any 3 attributes)

SECOND NORMAL FORM (2NF)

A relation schema R is in 2NF if every nonprime attribute A in R is **fully functionally dependent** on the primary key of R.

No partial dependency allowed

Raises when primary contains **more than one** attribute.

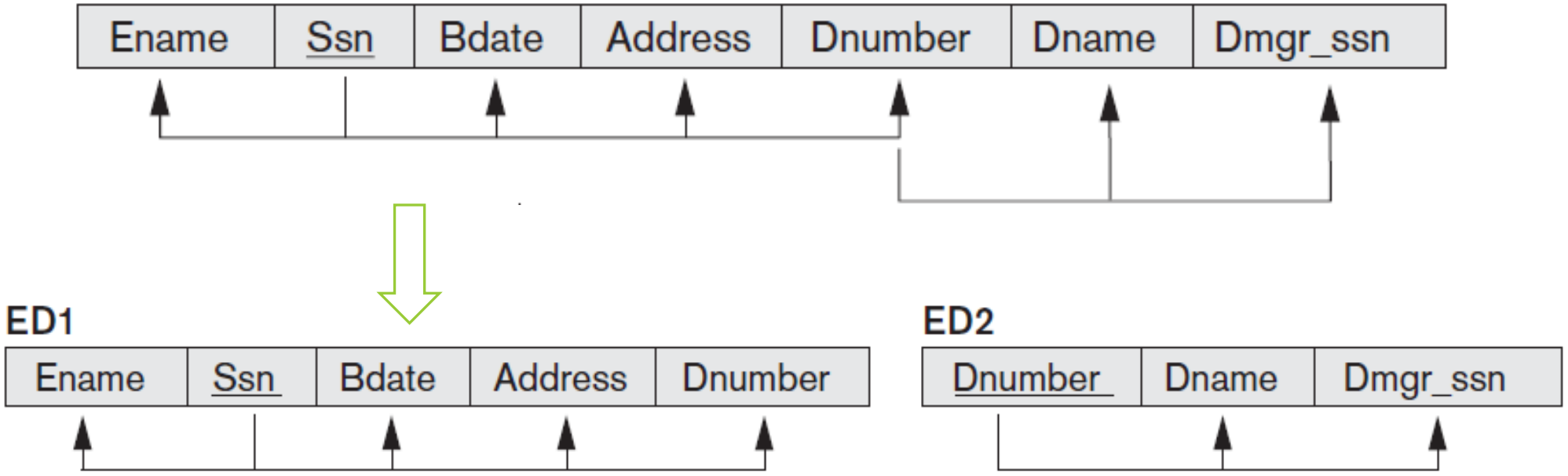


Prime attribute – An attribute, which is a part of the candidate-key

THIRD NORMAL FORM(3NF)

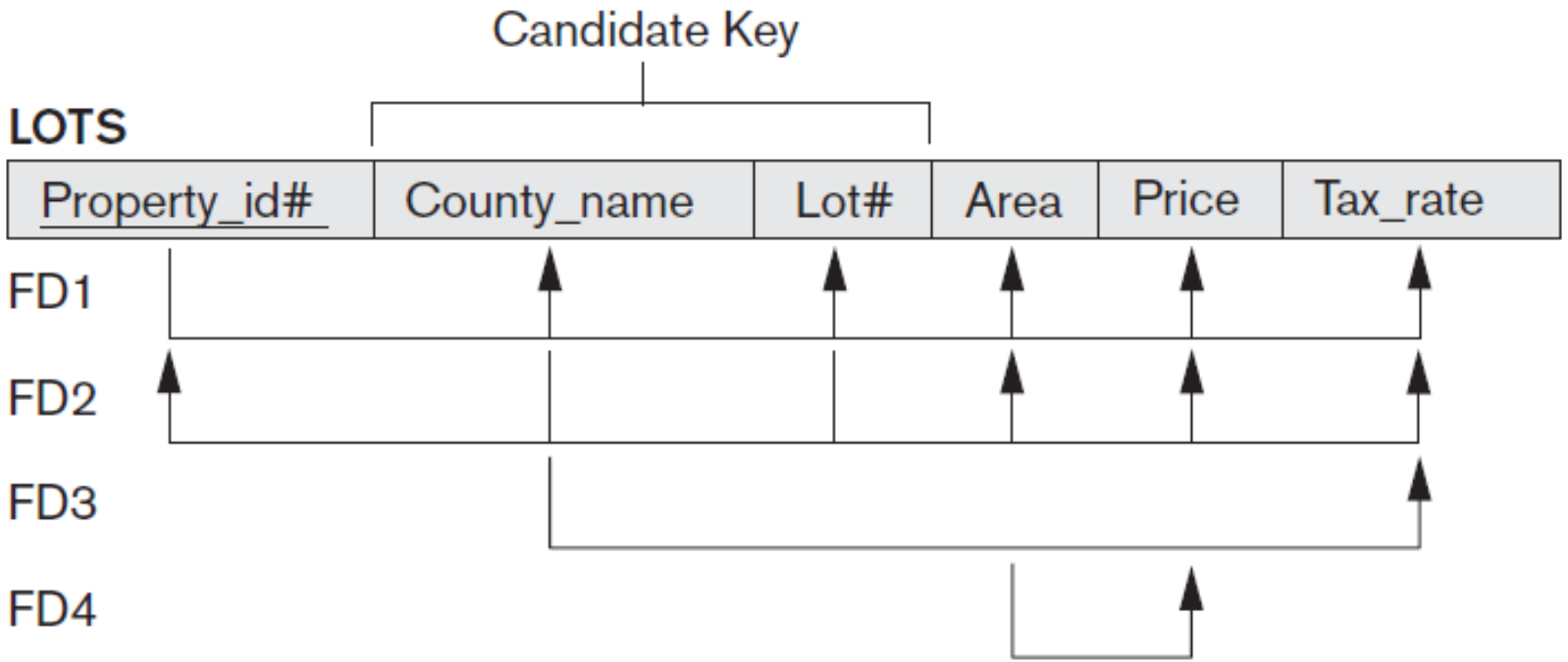
A relation schema R is in 3NF if it satisfies 2NF and no nonprime attribute of R is t**ransitively dependent** on the primary key.

A functional dependency X→Y in a relation schema R is a *transitive dependency* if there exists a set of attributes Z in R that is neither a candidate key nor a subset of any key of R, and both X→Z and Z→Y hold.



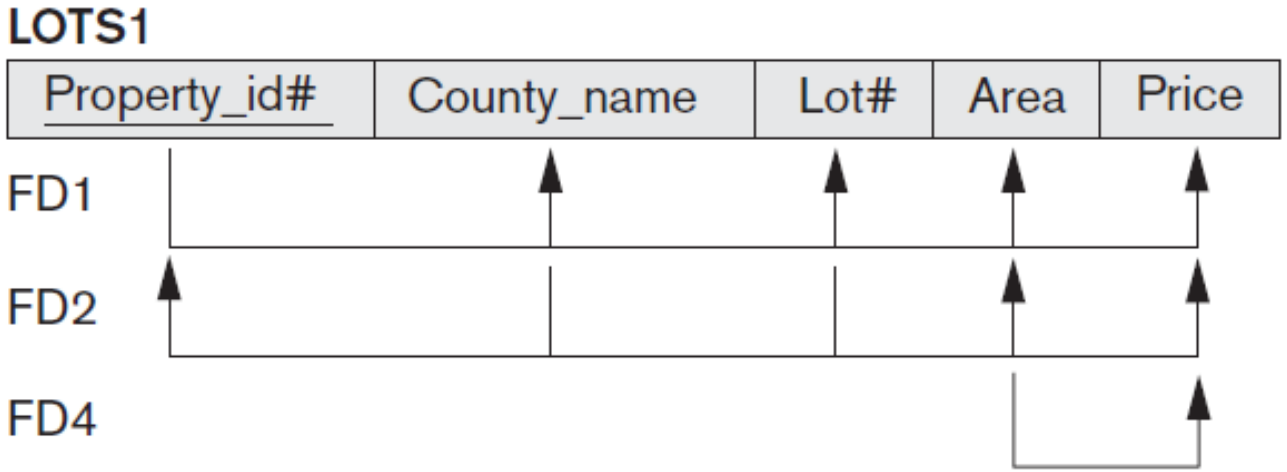
A relation schema R is in 2NF if every nonprime attribute A in R is not partially dependent on any key of R

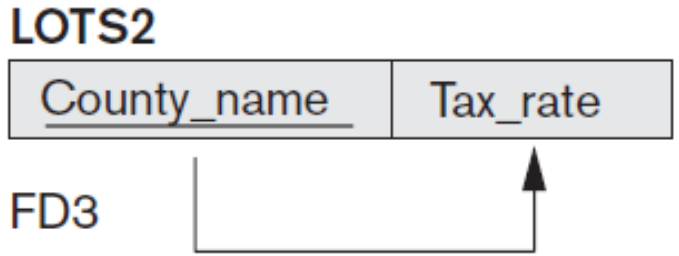
A relation schema R is in 3NF if, whenever a nontrivial functional dependency X→A holds in R, either (a) X is a superkey of R, or (b) A is a prime attribute of R.



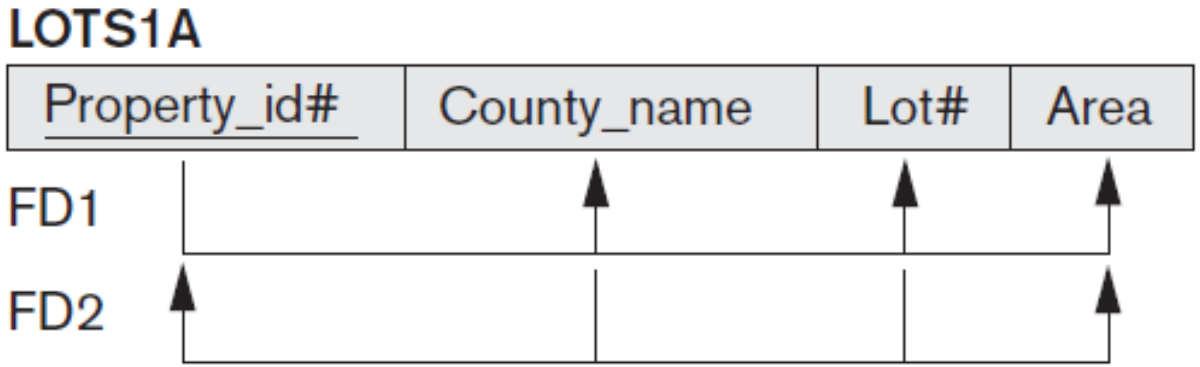
Trivial FD: X→Y and Y is a subset of X. They always hold.
e.g. {Country_name, Lot#} → Lot#

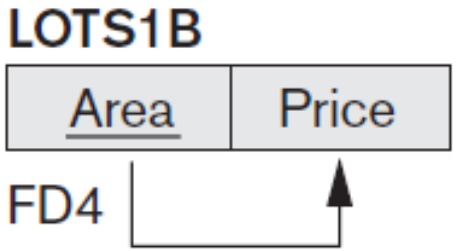
FD3 violates 2NF : Tax_rate is partially dependent on the candidate key {County_name, Lot#}. So Split LOTS into LOTS1 & LOTS2





FD4 in LOTS1 violates 3NF : Area is not a superkey and Price is not a prime attribute. So split LOTS1 into LOTS1A & LOTS1B

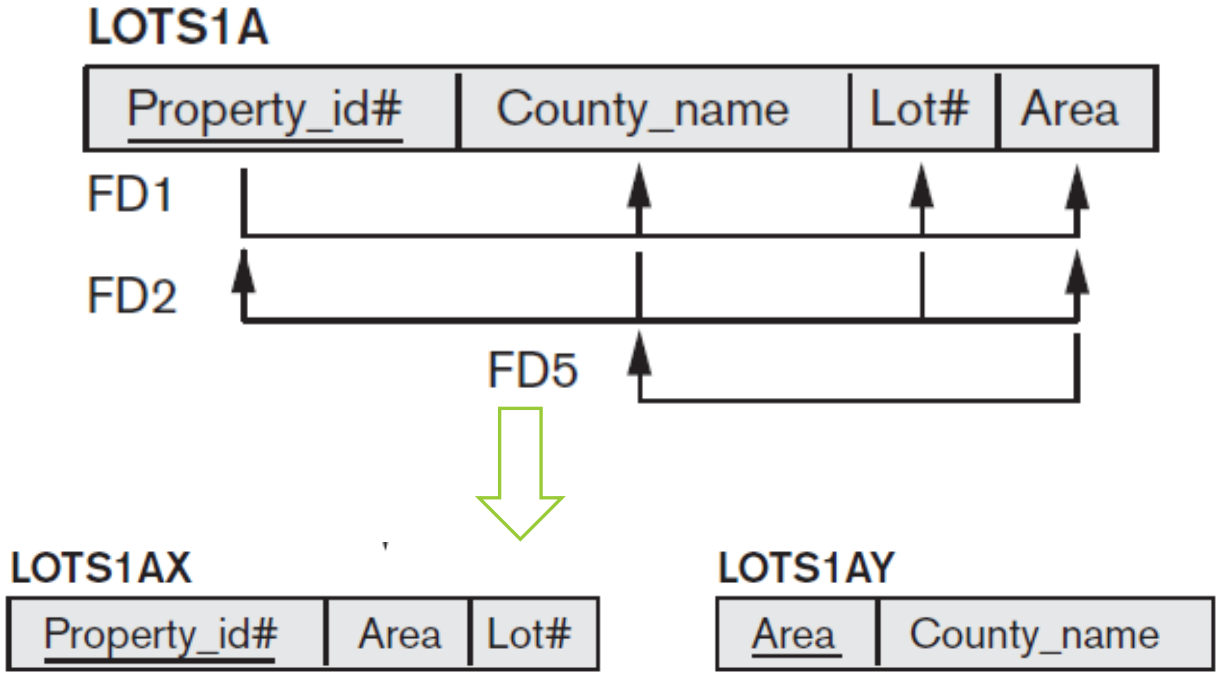


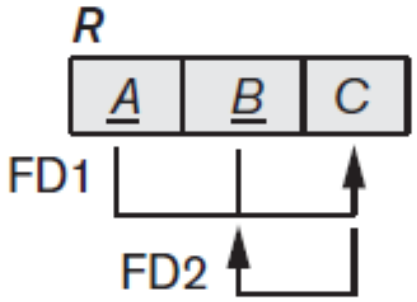


BOYCE-CODD NORMAL FORM (BCNF)

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.

Stricter than 3NF; every relation in BCNF is also in 3NF not vice versa.





in 3NF, but not in BCNF.

MULTIVALUED DEPENDENCY (MVD)

e.g. Employee can be assigned to multiple projects and can have multiple dependents

No functional dependencies

But lot of redundant data in EMP relation

Situation: There exist two independent 1:N relationship OR There are two or more multivalued independent attributes

So we say

- Ename multidetermines Pname &
- Ename multidetermines Dname

Denoted by X ↠ Y

EMP

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

Ename ↠ Pname
Ename ↠ Dname
Ename ↠ Pname | Dname

FOURTH NORMAL FORM (4NF)

A relation schema R is in 4NF with respect to a set of dependencies F if, for every nontrivial multivalued dependency $X \twoheadrightarrow Y$ in F^+, X is a superkey for R.

EMP

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

EMP_PROJECTS

Ename	Pname
Smith	X
Smith	Y

EMP_DEPENDENTS

Ename	Dname
Smith	John
Smith	Anna

Ename $\twoheadrightarrow$ Pname | Dname

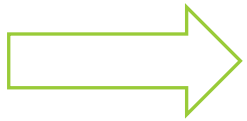
$X \twoheadrightarrow Y$ is called a trivial MVD if (a) Y is a subset of X, or (b) $X \cup Y = R$.

Update anomaly:

Brown starts working on a new additional project 'P'

EMP

Ename	Pname	Dname
Smith	X	John
Smith	Y	Anna
Smith	X	Anna
Smith	Y	John
Brown	W	Jim
Brown	X	Jim
Brown	Y	Jim
Brown	Z	Jim
Brown	W	Joan
Brown	X	Joan
Brown	Y	Joan
Brown	Z	Joan
Brown	W	Bob
Brown	X	Bob
Brown	Y	Bob
Brown	Z	Bob



EMP_PROJECTS

Ename	Pname
Smith	X
Smith	Y
Brown	W
Brown	X
Brown	Y
Brown	Z

EMP_DEPENDENTS

Ename	Dname
Smith	Anna
Smith	John
Brown	Jim
Brown	Joan
Brown	Bob

FIFTH NORMAL FORM (5NF) OR PROJECT-JOIN NORMAL FORM (PJNF)

A relation schema R is in 5NF with respect to a set F of functional, multivalued, and join dependencies if, for every nontrivial join dependency JD(R1, R2, ..., Rn) in F^+, every R_i is a superkey of R.

Till now, Normal forms were achieved using repeated binary decomposition.

All satisfied nonadditive join property, which ensures that no spurious tuples are generated when a NATURAL JOIN operation is applied to the relations resulting from the decomposition.

Whenever a supplier s supplies part p, and a project j uses part p, and the supplier s supplies at least one part to project j, then supplier s will also be supplying part p to project j.

SUPPLY

Sname	Part_name	Proj_name
Smith	Bolt	ProjX
Smith	Nut	ProjY
Adamsky	Bolt	ProjY
Walton	Nut	ProjZ
Adamsky	Nail	ProjX
Adamsky	Bolt	ProjX
Smith	Bolt	ProjY

JD(R1,R2, R3)
among the three projections R1(Sname, Part_name), R2(Sname, Proj_name), and R3(Part_name, Proj_name) of SUPPLY.

R_1

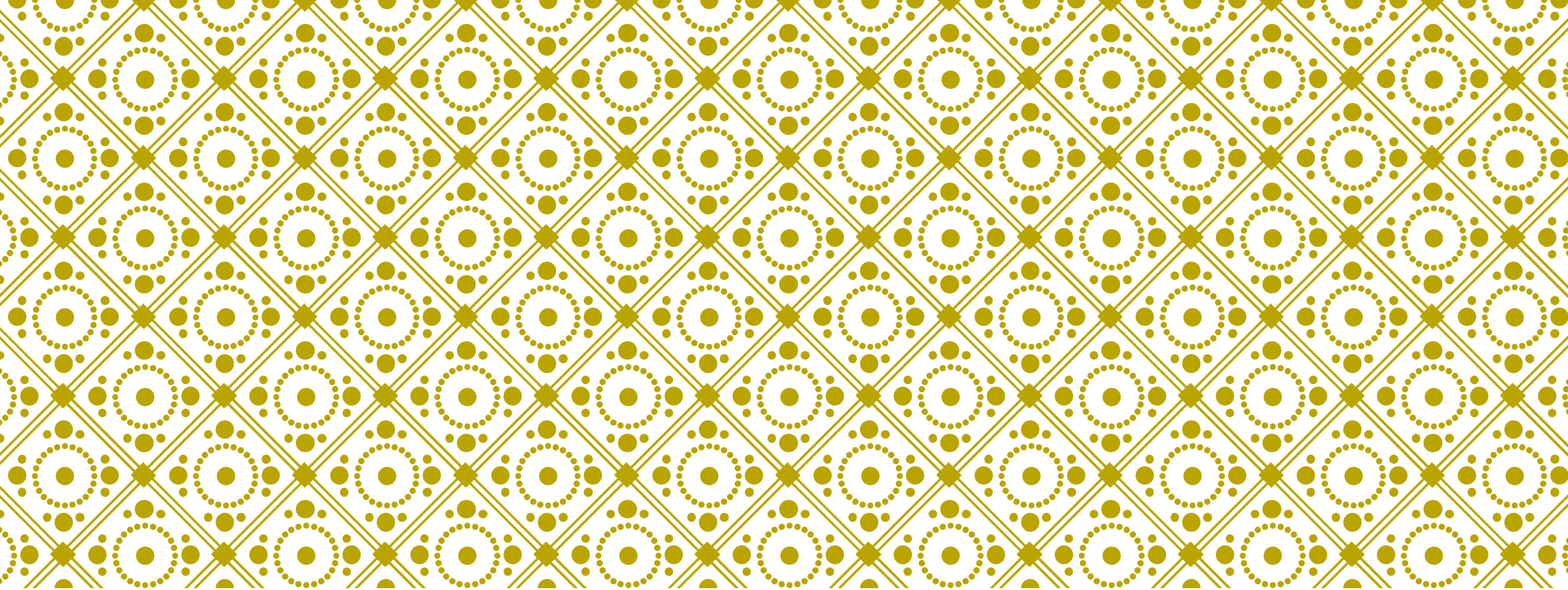
Sname	Part_name
Smith	Bolt
Smith	Nut
Adamsky	Bolt
Walton	Nut
Adamsky	Nail

R_2

Sname	Proj_name
Smith	ProjX
Smith	ProjY
Adamsky	ProjY
Walton	ProjZ
Adamsky	ProjX

R_3

Part_name	Proj_name
Bolt	ProjX
Nut	ProjY
Bolt	ProjY
Nut	ProjZ
Nail	ProjX



THANK YOU