Normalization for Relational Databases

政治大學

資訊科學系

沈錳坤

Outline

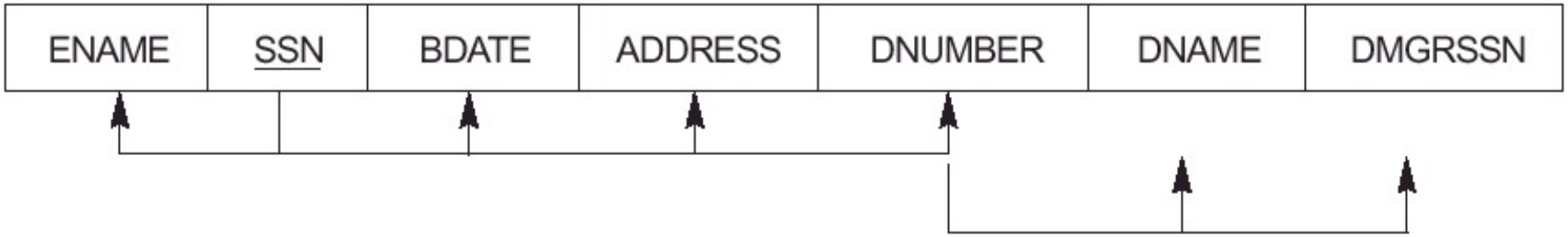
- Normal Forms based on Primary Keys
- General Definitions of Second & Third Normal Forms
- Boyce-Codd Normal Form

Normal Forms Based on Primary Keys

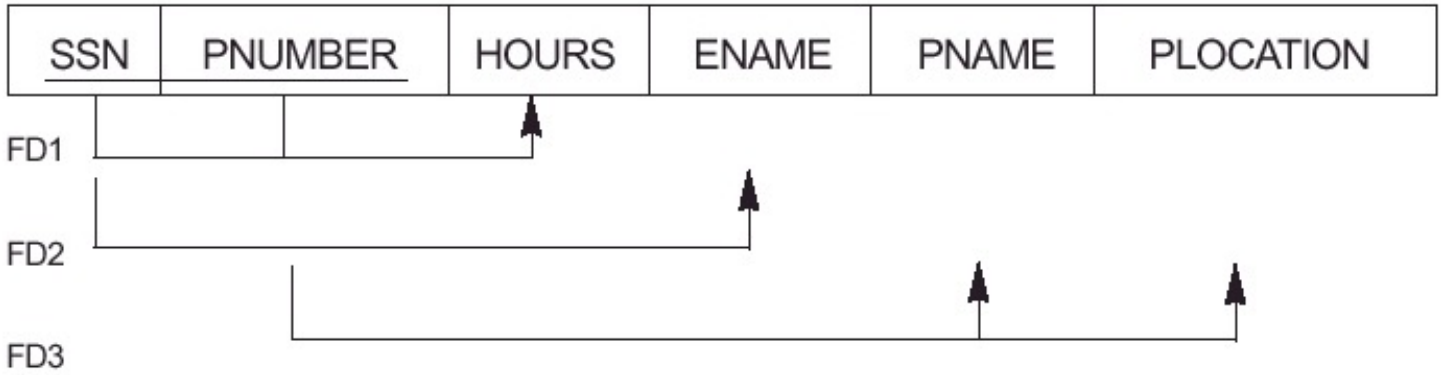
Informal Design Guidelines for Relation Schemas

- 4 Informal measures of quality for relation schema design
  - Semantics of the attributes
  - Reducing the redundant values in tuples
  - Reducing the null values in tuples
  - Disallowing the possibility of generating spurious tuples

(a)


ENAME
SSN
BDATE
ADDRESS
DNUMBER
DNAME
DMGRSSN

(b)


SSN
PNUMBER
HOURS
ENAME
PNAME
PLOCATION
FD1
FD2
FD3

EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
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

- To insert a new department that has no employee yet ?
- To delete an employee that happens to be the last employee working for a particular department ?
- To modify information of a department ?

Null Values in Tuples

- Guideline 3
  - Avoiding placing attributes in a base relation whose values may frequently be null
  - If nulls are unavoidable, make sure that they apply in exceptional cases only & do not apply to a majority of tuples in the relation
- e.g.: if only 10% employees have individual offices, there is little justification for including an attribute Office_Number in the Employee relation
- e.g.

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno

EMP_LOCS

ENAME	PLOCATION
Smith, John B.	Bellaire
Smith, John B.	Sugarland
Narayan, Ramesh K.	Houston
English, Joyce A.	Bellaire
English, Joyce A.	Sugarland
Wong, Franklin T.	Sugarland
Wong, Franklin T.	Houston
Wong, Franklin T.	Stafford
Zelaya, Alicia J.	Stafford
Jabbar, Ahmad V.	Stafford
Wallace, Jennifer S.	Stafford
Wallace, Jennifer S.	Houston
Borg,James E.	Houston

EMP_PROJ1

SSN	PNUMBER	HOURS	PNAME	PLOCATION
123456789	1	32.5	Product X	Bellaire
123456789	2	7.5	Product Y	Sugarland
666884444	3	40.0	Product Z	Houston
453453453	1	20.0	Product X	Bellaire
453453453	2	20.0	Product Y	Sugarland
333445555	2	10.0	Product Y	Sugarland
333445555	3	10.0	Product Z	Houston
333445555	10	10.0	Computerization	Stafford
333445555	20	10.0	Reorganization	Houston
999887777	30	30.0	Newbenefits	Stafford
999887777	10	10.0	Computerization	Stafford
987987987	10	35.0	Computerization	Stafford
987987987	30	5.0	Newbenefits	Stafford
987654321	30	20.0	Newbenefits	Stafford
987654321	20	15.0	Reorganization	Houston
888665555	20	null	Reorganization	Houston

	SSN	PNUMBER	HOURS	PNAME	PLOCATION	
	123456789	1	32.5	ProductX	Bellaire	Smith,John B.
*	123456789	1	32.5	ProductX	Bellaire	English,Joyce A.
	123456789	2	7.5	ProductY	Sugarland	Smith,John B.
*	123456789	2	7.5	ProductY	Sugarland	English,Joyce A.
*	123456789	2	7.5	ProductY	Sugarland	Wong,Franklin T.
	666884444	3	40.0	ProductZ	Houston	Narayan,Ramesh K.
*	666884444	3	40.0	ProductZ	Houston	Wong,Franklin T.
*	453453453	1	20.0	ProductX	Bellaire	Smith,John B.
	453453453	1	20.0	ProductX	Bellaire	English,Joyce A.
*	453453453	2	20.0	ProductY	Sugarland	Smith,John B.
	453453453	2	20.0	ProductY	Sugarland	English,Joyce A.
*	453453453	2	20.0	ProductY	Sugarland	Wong,Franklin T.
*	333445555	2	10.0	ProductY	Sugarland	Smith,John B.
*	333445555	2	10.0	ProductY	Sugarland	English,Joyce A.
	333445555	2	10.0	ProductY	Sugarland	Wong,Franklin T.
*	333445555	3	10.0	ProductZ	Houston	Narayan,Ramesh K.
	333445555	3	10.0	ProductZ	Houston	Wong,Franklin T.
	333445555	10	10.0	Computerization	Stafford	Wong,Franklin T.
*	333445555	20	10.0	Reorganization	Houston	Narayan,Ramesh K.
	333445555	20	10.0	Reorganization	Houston	Wong,Franklin T.

Approaches of Database Design

- Approaches
  - Bottom up design
    - Consider the basic relationships among individual attributes as the starting point
    - Not very popular in practice
    - Suffers from the problem of collecting a large number of binary attribute relationships
    - Design by synthesis
  - Top-down design
    - Start with a number of groupings of attributes into relations obtained from conceptual design
    - Further decomposition until all desirable properties are met
    - Design by analysis

Normalization

- Normalization
  - Proposed by Codd
  - Proceeds in a top-down fashion by evaluating each relation against the criteria for normal forms & decomposing relations as necessary
  - Relation design by analysis
- Normal forms
  - 1NF (First normal form)
  - 2NF (Second normal form)
  - 3NF (Third normal form)
  - BCNF (Boyce-Codd normal form)

  } functional dependencies among attributes

  - 4NF (Fourth normal form): multi-value dependency
  - 5NF (Fifth normal form): join-dependency

Normalization (cont.)

- Normalization procedure
  - A formal framework for analyzing relation schemas based on their keys & on the functional dependencies among their attributes
  - To achieve
    - Minimizing redundancy
    - Minimizing the insertion, deletion & update anomalies
  - Unsatisfactory relation schemas are decomposed into smaller relation schemas that meet the normal form test

Normalization (cont.)

- Superkey
- Key: minimal superkey

  (removal of any attribute will not be a superkey)
- Candidate key
- Primary key

  choose one (primary key is also a candidate key)
- Prime attribute of R
  - a member of some candidate key of R

EMPLOYEE

ENAME	SSN	BDATE	ADDRESS	DNUMBER
	p.k.			f.k.

DEPARTMENT

DNAME	DNUMBER	DMGRSSN
	p.k.	f.k.

DEPT_LOCATIONS

DNUMBER	DLOCATION
f.k.	p.k.

PROJECT

PNAME	PNUMBER	PLOCATION	DNUM
	p.k.		f.k.

WORKS_ON

SSN	PNUMBER	HOURS
f.k.	f.k.	
p.k.		

First Normal Form

- First normal form
  - (1) domain of an attribute must include only atomic values
    (2) value of any attribute in a tuple must be a single value from the domain of that attribute
  - Disallow multi-value attribute, composite attributes & their combinations
  - Be considered to be part of the formal definition of a relation in the basic relational model

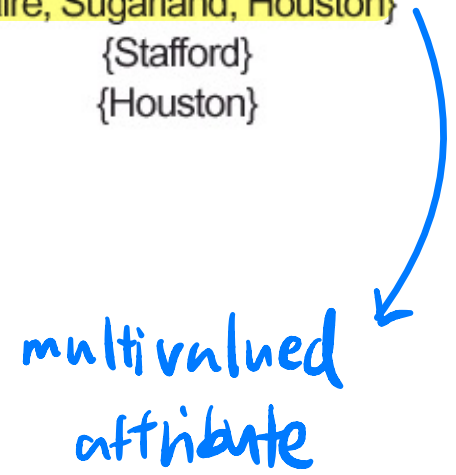
First Normal Form (cont.)

DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATIONS

DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATIONS
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}



Approach 1

DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATIONS
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

DEPARTMENT

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

DEPT_LOCATIONS

DNUMBER	DLOCATION
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

Approach 2

DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATIONS
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATION
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Approach 3

DEPARTMENT

DNAME	DNUMBER	DMGRSSN	DLOCATIONS
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

DName	DNumber	DMGRSSN	DLocation1	DLocation2	Dlocation3
Research	5	333445555	Bellaire	Sugarland	Houston
Administration	4	987654321	Stafford		
Headquarters	1	888665555	Houston		

EMP_PROJ

composite attribute

SSN	ENAME	PROJS	
		PNUMBER	HOURS

EMP_PROJ

SSN	ENAME	PNUMBER	HOURS
123456789	Smith,John B.	1	32.5
		2	7.5
666884444	Narayan,Ramesh K.	3	40.0
453453453	English,Joyce A.	1	20.0
		2	20.0
333445555	Wong,Franklin T.	2	10.0
		3	10.0
		10	10.0
		20	10.0
999887777	Zelaya,Alicia J.	30	30.0
		10	10.0
987987987	Jabbar,Ahmad V.	10	35.0
		30	5.0
987654321	Wallace,Jennifer S.	30	20.0
		20	15.0
888665555	Borg,James E.	20	null

EMP_PROJ

SSN	ENAME	PROJS	
		PNUMBER	HOURS

EMP_PROJ1

SSN	ENAME

EMP_PROJ2

SSN	PNUMBER	HOURS

Second Normal Form

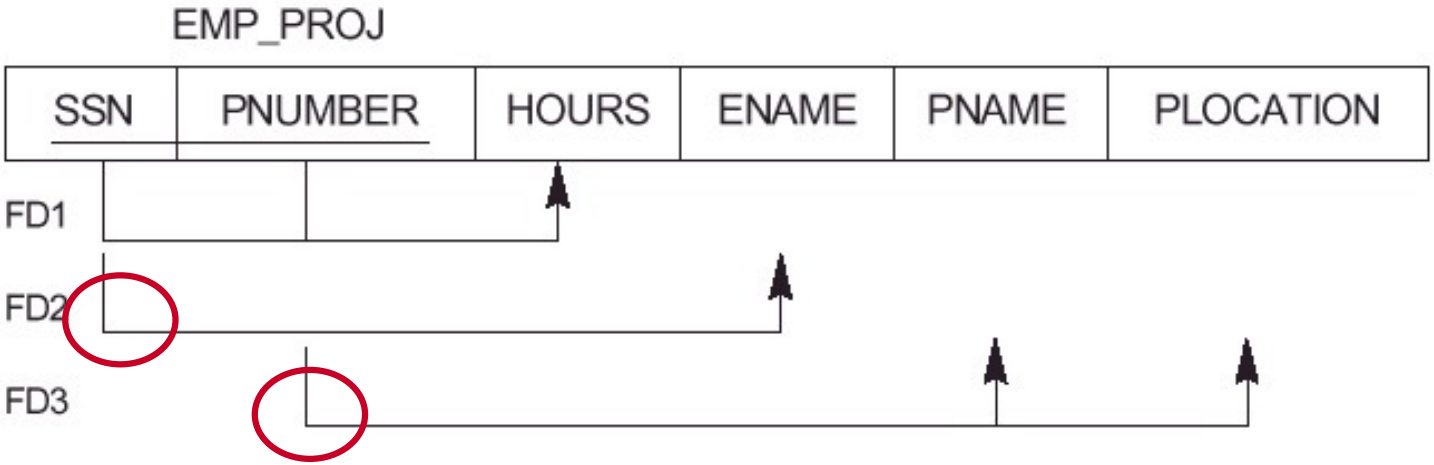
- Full functional dependency
  - X→Y is a full functional dependency

    if removal of any attribute A from X means that

    the dependency does not hold any more
  - X→Y is a Partial functional dependency

    if removal of any attribute A from X means that

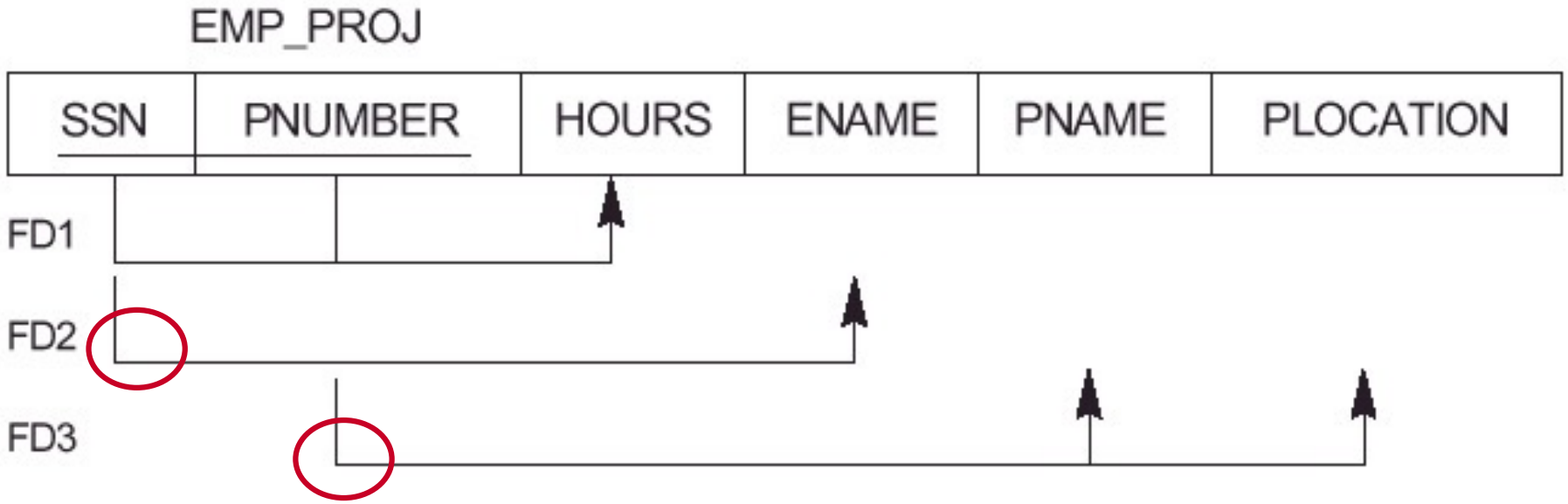
    the dependency still holds
  - e.g. {SSN, PNumber}→ {Hours} is a full FD
  - e.g. {SSN, PNumber} → {Ename} is a partial FD
  - e.g. {SSN, Pnumber} → {Pname, PLocation} is a partial FD



HOURS is full FD on {SSN, PNUMBER}

ENAME is partial FD on {SSN, PNUMBER}

PNAME, PLOCATION is partial FD on {SSN, PNUMBER}



EMP_PROJ1

ENAME	SSN	PNUMBER	HOURS	PNAME	PLOCATION
Smith,John B.	123456789	1	32.5	Product X	Bellaire
Smith,John B.	123456789	2	7.5	Product Y	Sugarland
Narayan,Ramesh K.	666884444	3	40.0	Product Z	Houston
English,Joyce A.	453453453	1	20.0	Product X	Bellaire
English,Joyce A.	453453453	2	20.0	Product Y	Sugarland
Wong,Franklin T.	333445555	2	10.0	Product Y	Sugarland
Wong,Franklin T.	333445555	3	10.0	Product Z	Houston
Wong,Franklin T.	333445555	10	10.0	Computerization	Stafford
Wong,Franklin T.	333445555	20	10.0	Reorganization	Houston
Zelaya, Alicia J.	999887777	30	30.0	Newbenefits	Stafford
Zelaya, Alicia J.	999887777	10	10.0	Computerization	Stafford
Jabbar,Ahmad V.	987987987	10	35.0	Computerization	Stafford
Jabbar,Ahmad V.	987987987	30	5.0	Newbenefits	Stafford
Wallace,Jennifer S.	987654321	30	20.0	Newbenefits	Stafford
Wallace,Jennifer S.	987654321	20	15.0	Reorganization	Houston
Borg,James E.	888665555	20	null	Reorganization	Houston

Redundancy appear in ENAME, PNAME, PLOCATION

Second Normal Form (cont.)

- 2NF
  - A relation schema R is in 2NF

    if every nonprime attribute A in R

    is fully functional dependent on the primary key of R
  - If R is not in 2NF,

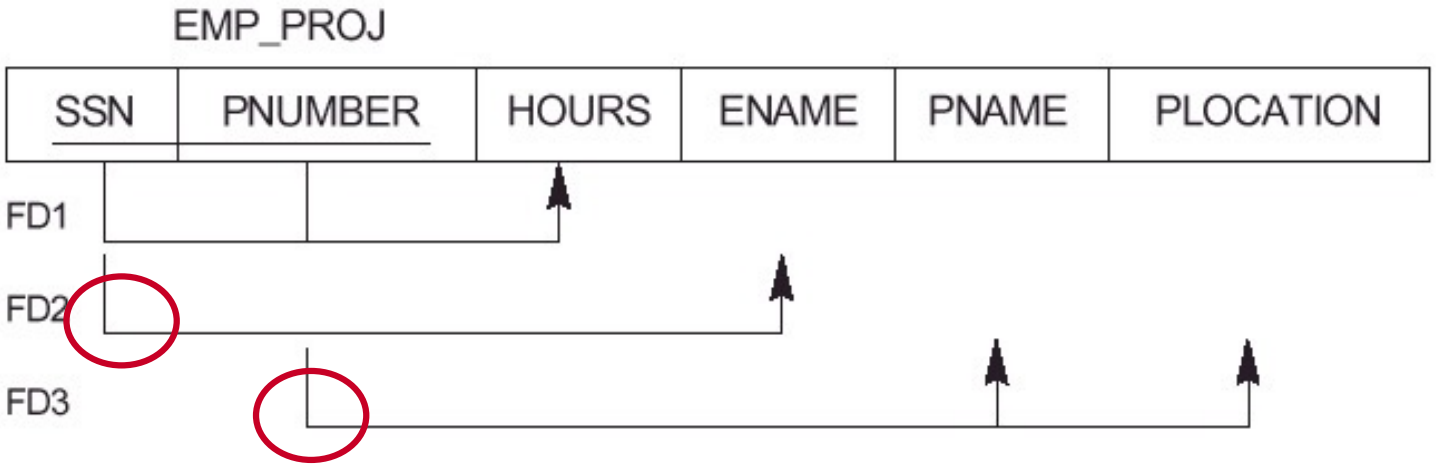
    it can be decomposed into a number of 2NF relations

    in which

    nonprime attributes are associated only with

    the part of the primary key

    on which they are fully functional dependent



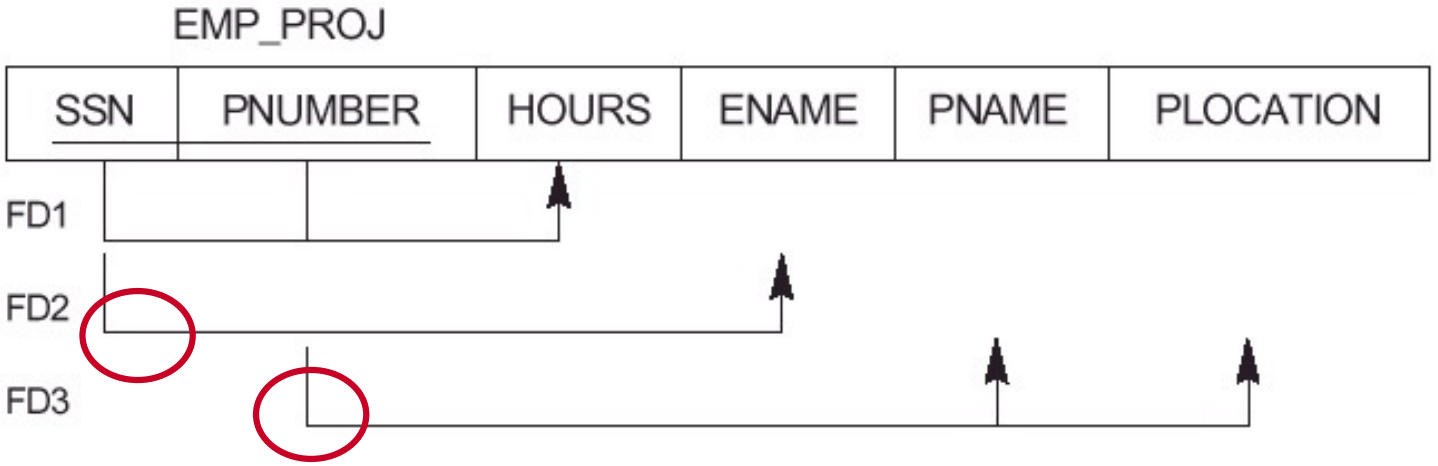
{SSN, PNUMBER} is primary key

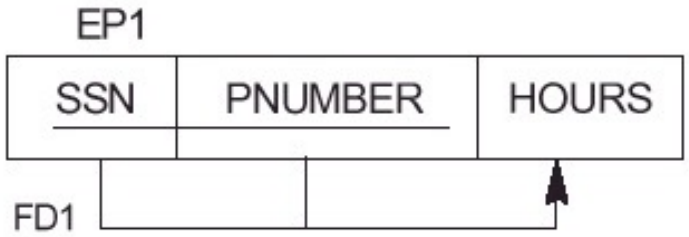
HOURS is full FD on {SSN, PNUMBER}

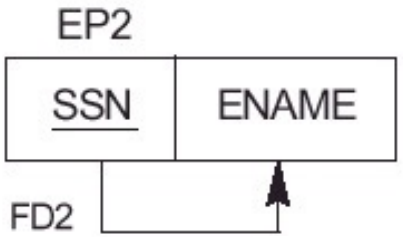
ENAME is partial FD on {SSN, PNUMBER}

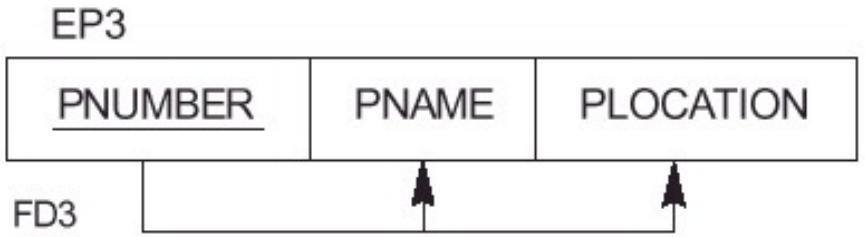
PNAME, PLOCATION is partial FD on {SSN, PNUMBER}

EMP_PROJ is not in 2NF


EMP_PROJ
SSN
PNUMBER
HOURS
ENAME
PNAME
PLOCATION
FD1
FD2
FD3


EP1
SSN
PNUMBER
HOURS
FD1


EP2
SSN
ENAME
FD2


EP3
PNUMBER
PNAME
PLOCATION
FD3

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

WORKS_ON

Essn	Pno	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	NULL

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Third Normal Form

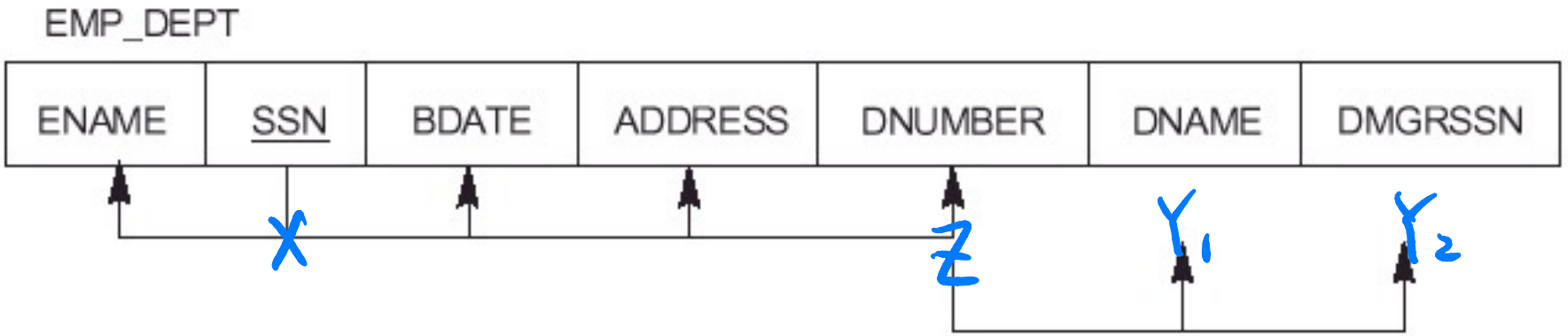
- Transitive dependency
  - $X \rightarrow Y$ is a transitive dependency

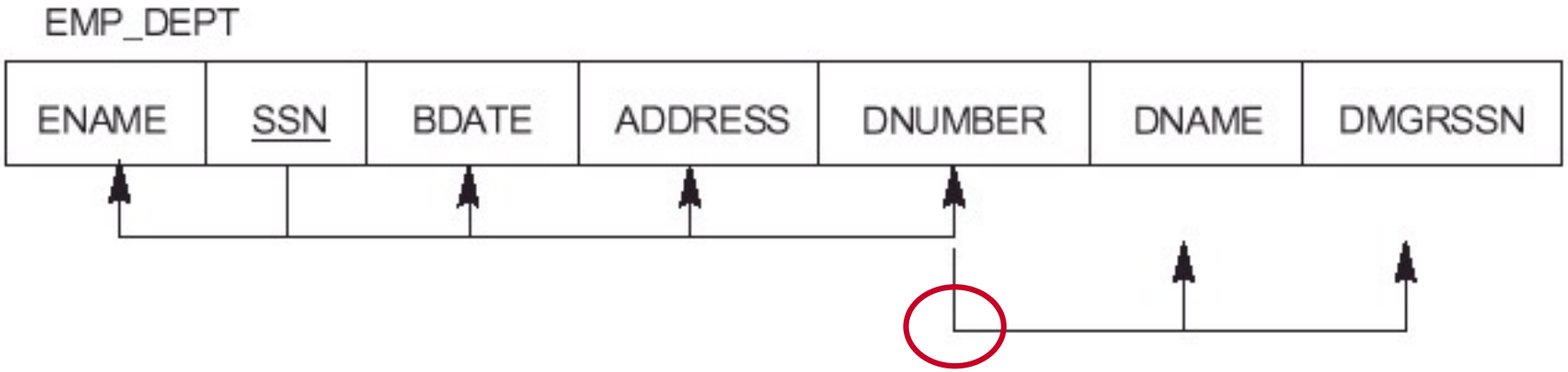
    if there is a set of attributes Z that is

    neither a candidate key

    nor a prime attribute of R &

    $X \rightarrow Z$ and $Z \rightarrow Y$
  - e.g. {SSN}→{DNumber} → {Dname, DMgrSSN}





{Dname, DMgrSSN} is Transitively Dependent on {SSN}

EMP_DEPT isn't in 3NF

* {SSN}→{DNumber} → {Dname, DMgrSSN}

X Z Y

EMP_DEPT

EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
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

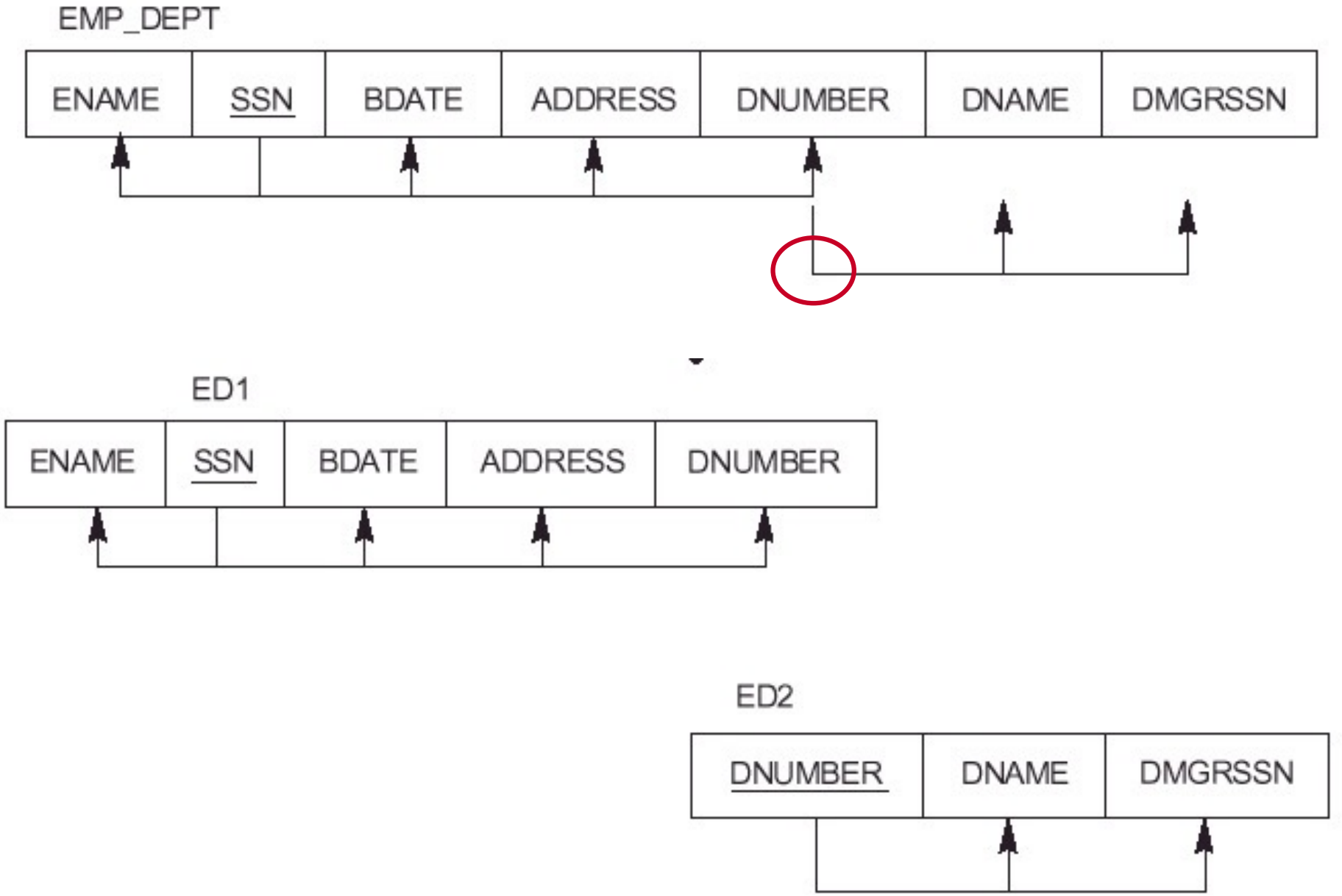
Third Normal Form (cont.)

- 3NF
  - A relation schema R is in 3NF

    if it satisfies 2NF &

    no nonprime attribute of R

    is transitively dependent on the primary key


EMP_DEPT
ENAME
SSN
BDATE
ADDRESS
DNUMBER
DNAME
DMGRSSN
ED1
ENAME
SSN
BDATE
ADDRESS
DNUMBER
ED2
DNUMBER
DNAME
DMGRSSN

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

WORKS_ON

Essn	Pno	Hours
123456789	1	32.5
123456789	2	7.5
666884444	3	40.0
453453453	1	20.0
453453453	2	20.0
333445555	2	10.0
333445555	3	10.0
333445555	10	10.0
333445555	20	10.0
999887777	30	30.0
999887777	10	10.0
987987987	10	35.0
987987987	30	5.0
987654321	30	20.0
987654321	20	15.0
888665555	20	NULL

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

PROJECT

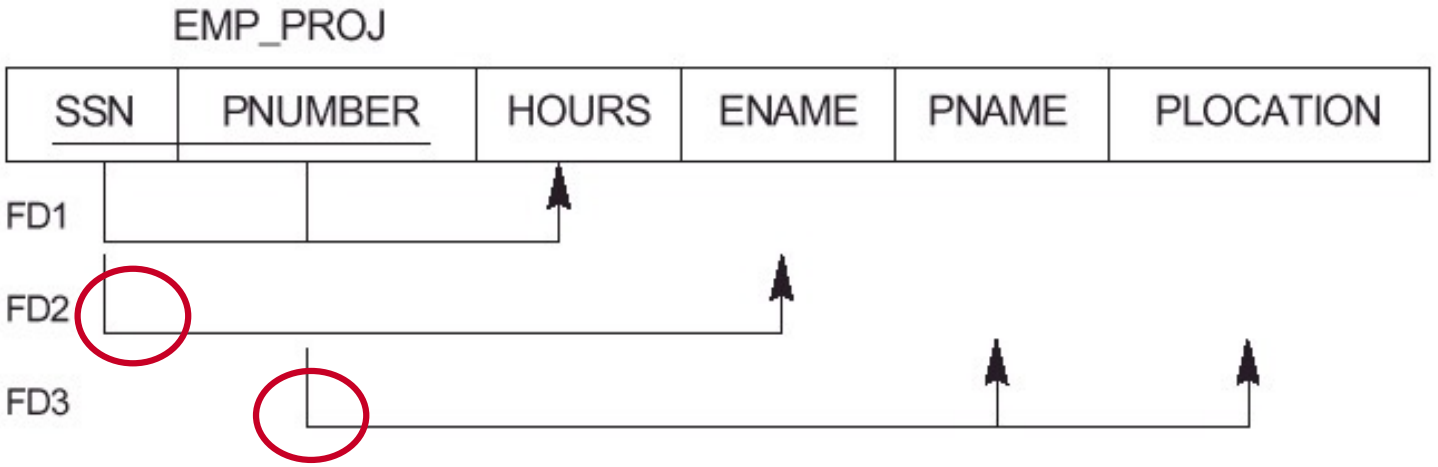
Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

Intuitively about Normalization

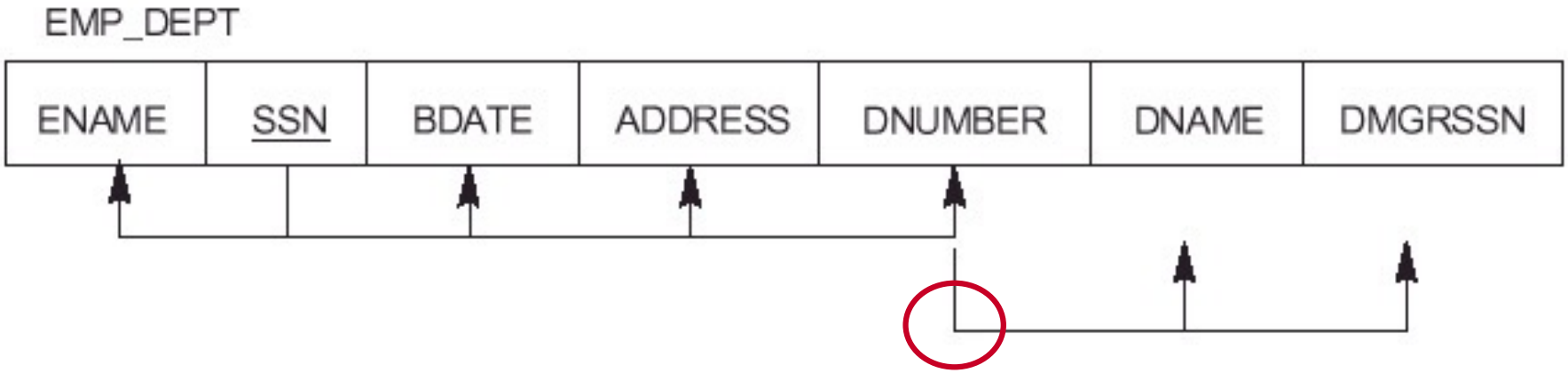
- Any functional dependency in which left-hand side is
  - Proper subset of the primary key (breaks 2nd NF)
  - Nonkey attribute (breaks 3rd NF)

  is a problematic functional dependency.

  * Proper subset of primary key: partial FD
  * Nonkey attribute: transitive dependency



Left hand side of FD is proper subset of the primary key



Left hand side of FD is nonkey attribute

General Definitions of Second & Third Normal Forms

2nd & 3rd Normal Form

建立在與 Primary Key 的 Functional Dependency 上，

那如果是 Candidate Key 的 Functional Dependency 呢 ？

General Definition of 2NF

- 2NF
  - A relation schema R is in 2NF

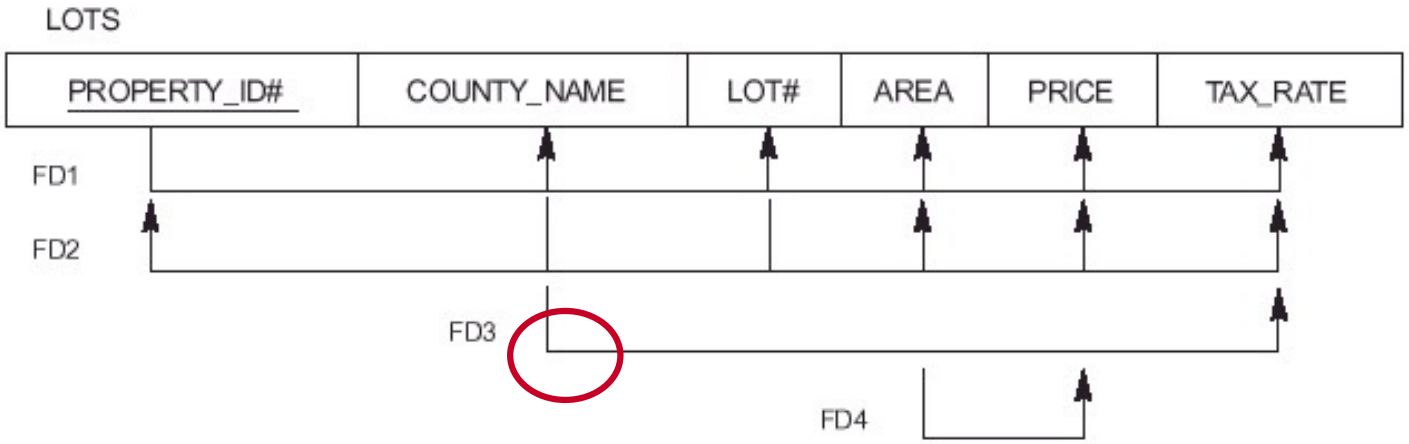
    if every nonprime attribute A in R is

    fully functional dependent on any key of R

    * Original definition
      - A relation schema R is in 2NF

        if every nonprime attribute A in R is

        fully functional dependent on the primary key of R

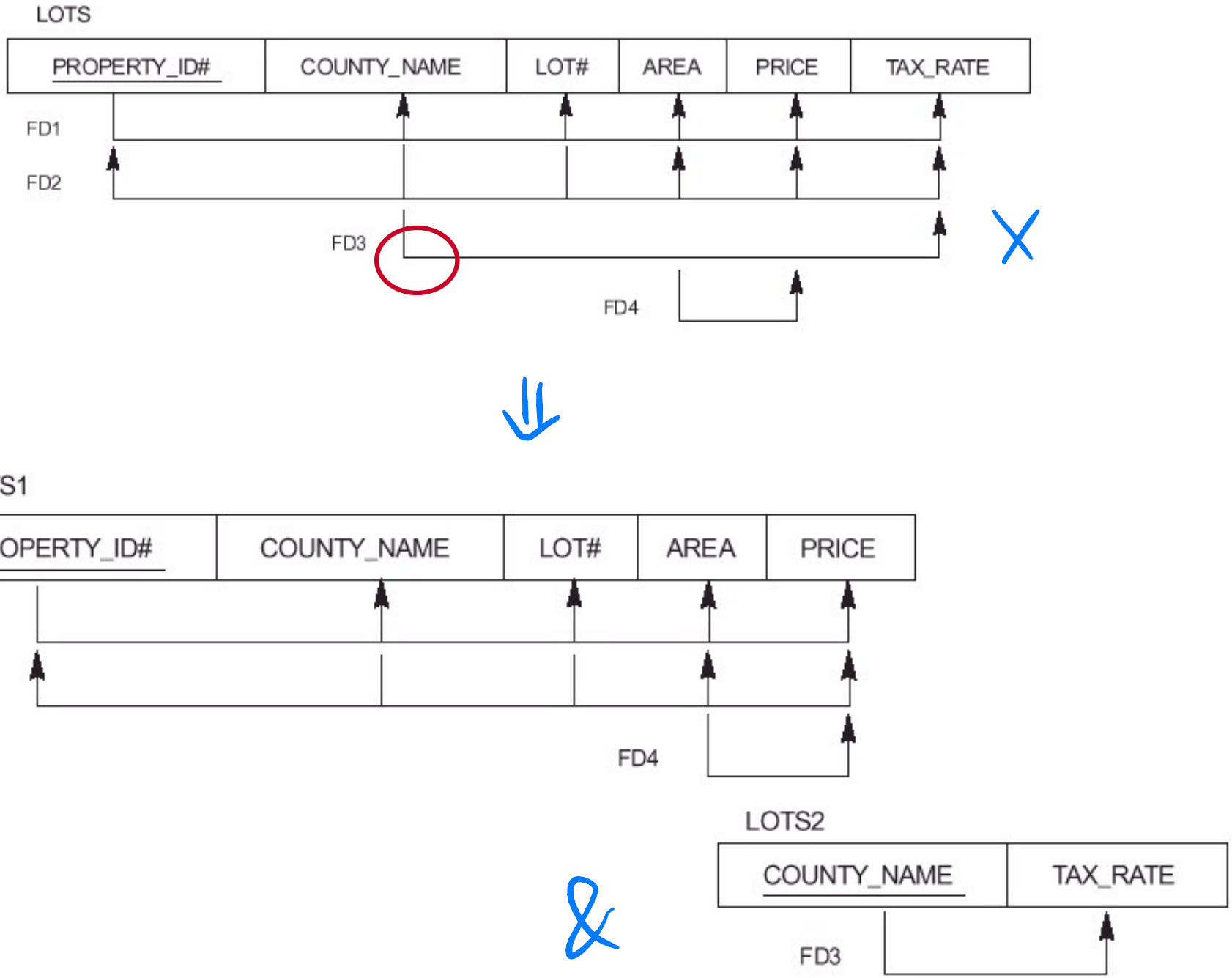


{RPROPERTY_ID#} is primary key

{COUNTY_NAME, LOT#} is a candidate key

TAX_RATE is partial FD on {COUNTY_NAME, LOT#}

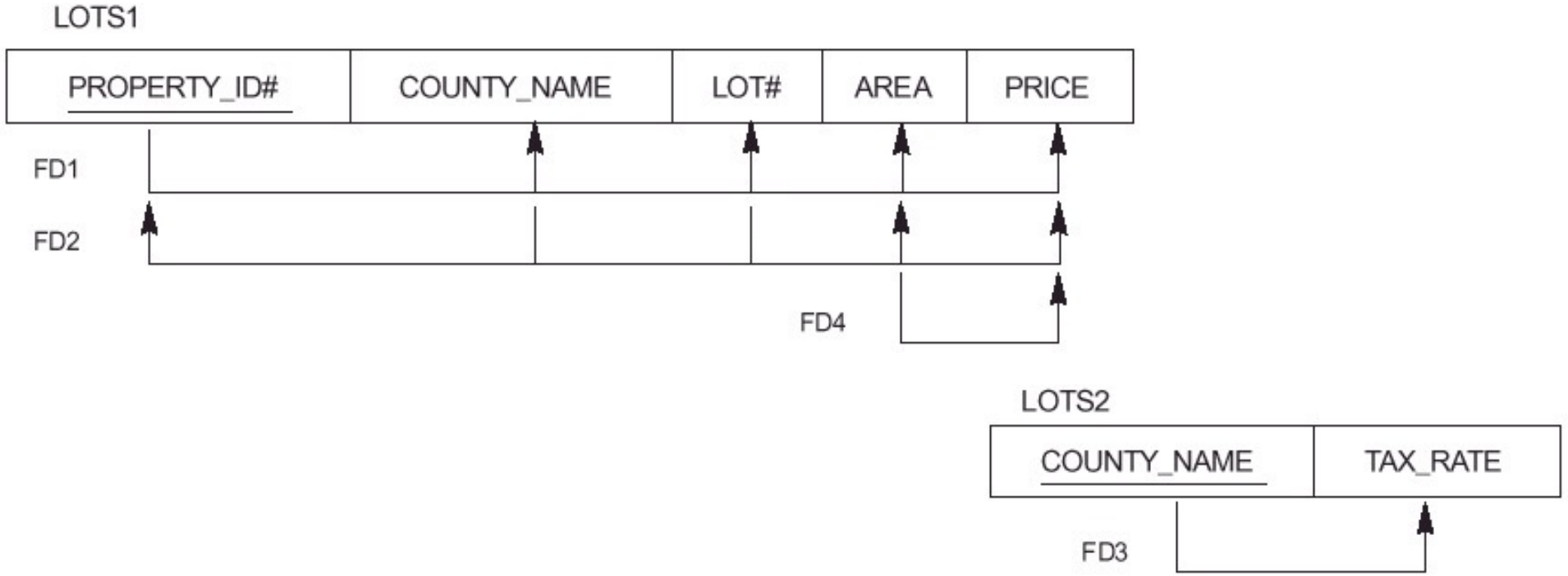
LOTS is not in 2NF


LOTS
PROPERTY_ID#
COUNTY_NAME
LOT#
AREA
PRICE
TAX_RATE
FD1
FD2
FD3
FD4
LOTS1
PROPERTY_ID#
COUNTY_NAME
LOT#
AREA
PRICE
FD1
FD2
FD4
LOTS2
COUNTY_NAME
TAX_RATE
FD3

General Definition of 3NF

- 3NF
  - A relation schema R is in 3NF
    if it satisfies 2NF &
      no nonprime attribute of R
        is transitively dependent on any key of R

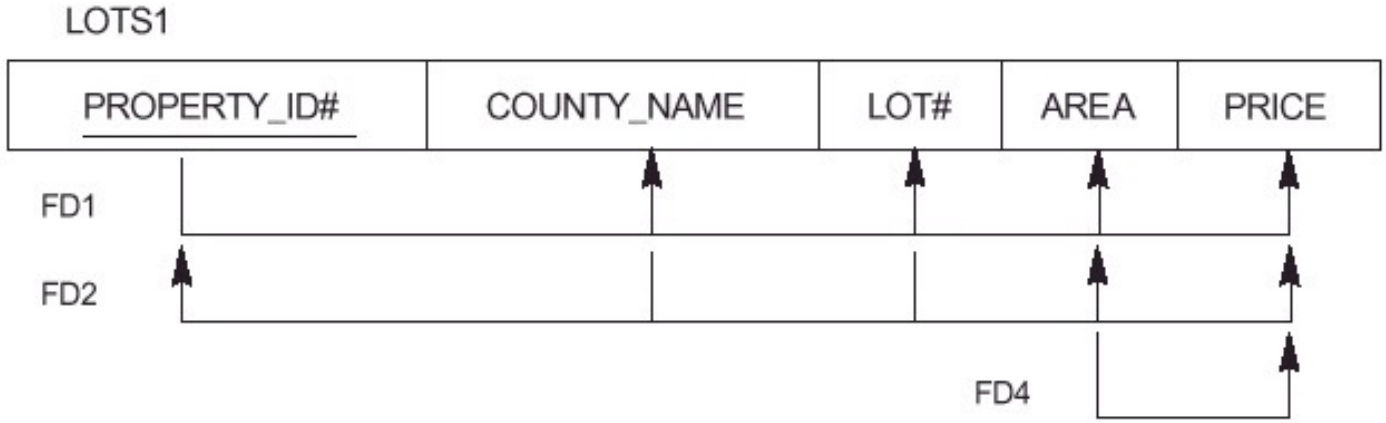
    * Original definition
      - A relation schema R is in 3NF
        if it satisfies 2NF &
          no nonprime attribute of R
            is transitively dependent on primary key of R

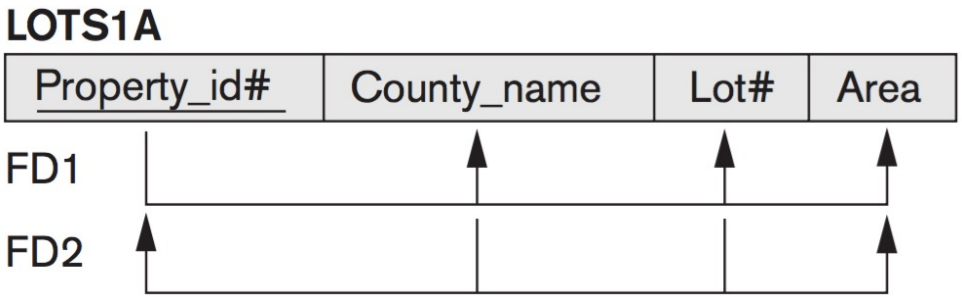


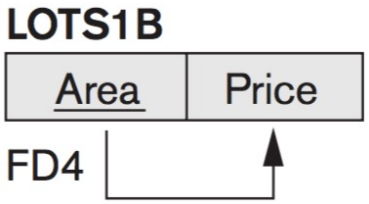
{COUNTY_NAME, LOT#} is a candidate key

PRICE is transitive depend on {COUNTY_NAME, LOT#}

LOTS1 isn't in 3 NF


LOTS1
PROPERTY_ID#
COUNTY_NAME
LOT#
AREA
PRICE
FD1
FD2
FD4


LOTS1A
Property_id#
County_name
Lot#
Area
FD1
FD2


LOTS1B
Area
Price
FD4

Alternative Definition of 3NF

- Alternative definition of 3NF
  - A relation schema R is in 3NF

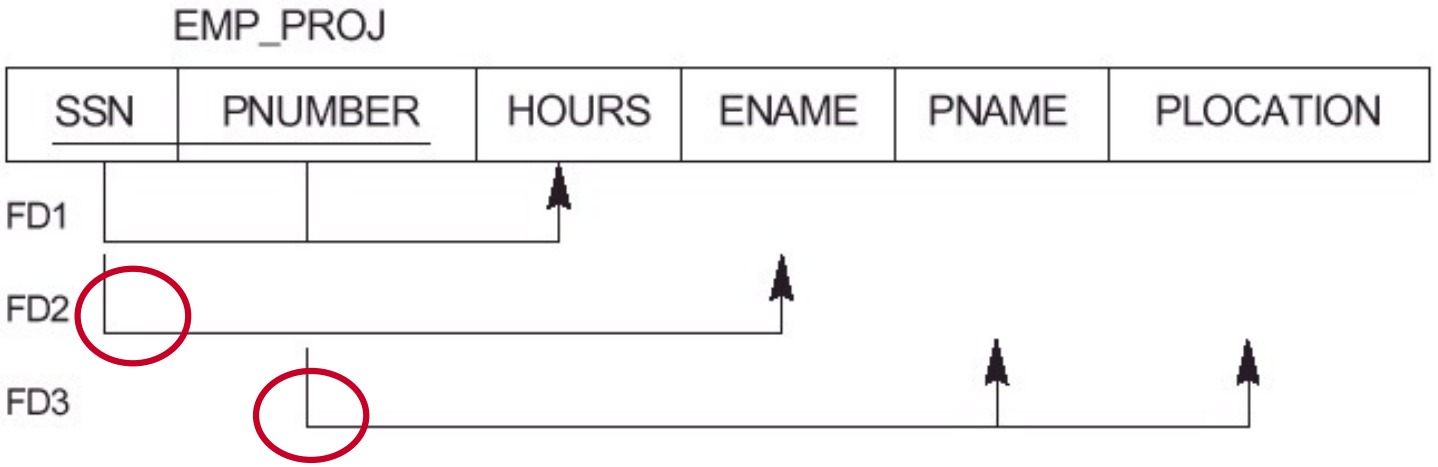
    if every nonprime attribute of R meets both of the following conditions
    - It is fully functionally dependent on every key of R (2nd NF)
    - It is non-transitively dependent on every key of R (3rd NF)

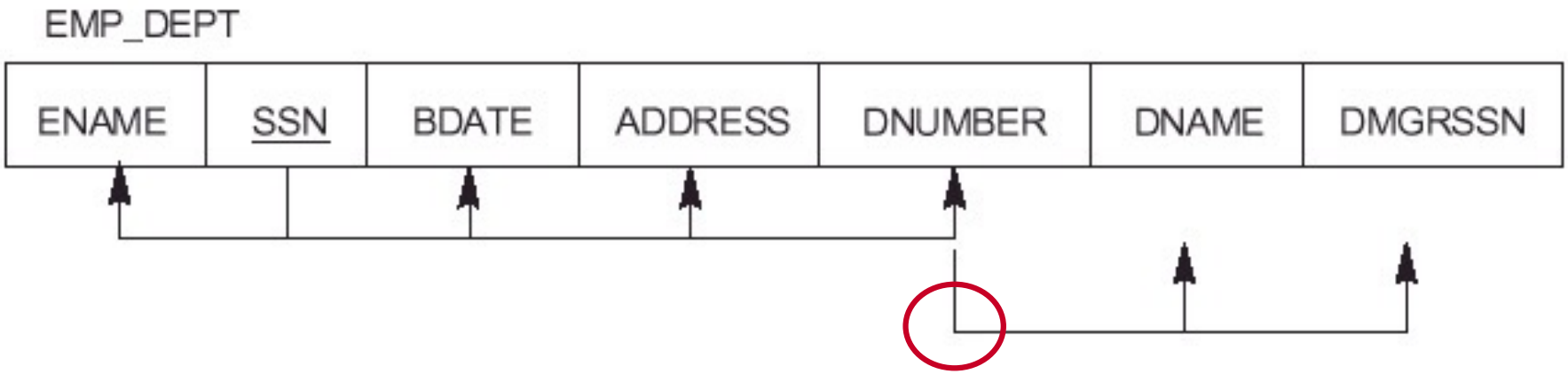
Intuitively about Normalization

- Any functional dependency in which left-hand side is
  - Proper subset of the key
  - Nonkey attribute

  is a problematic functional dependency.

  * Proper subset of key: partial FD
  * Nonkey attribute: transitive dependency


EMP_PROJ
SSN
PNUMBER
HOURS
ENAME
PNAME
PLOCATION
FD1
FD2
FD3


EMP_DEPT
ENAME
SSN
BDATE
ADDRESS
DNUMBER
DNAME
DMGRSSN

Boyce-Codd Normal Form

Boyce-Codd Normal Form

- BCNF
  - Simpler form of 3NF
  - Stricter than 3NF
  - Every relation in BCNF is also in 3NF
  - A relation in 3NF is not necessarily in BCNF
  - A relation schema R is in BCNF

    if whenever a non-trivial functional dependency $X \rightarrow A$ holds in R, then

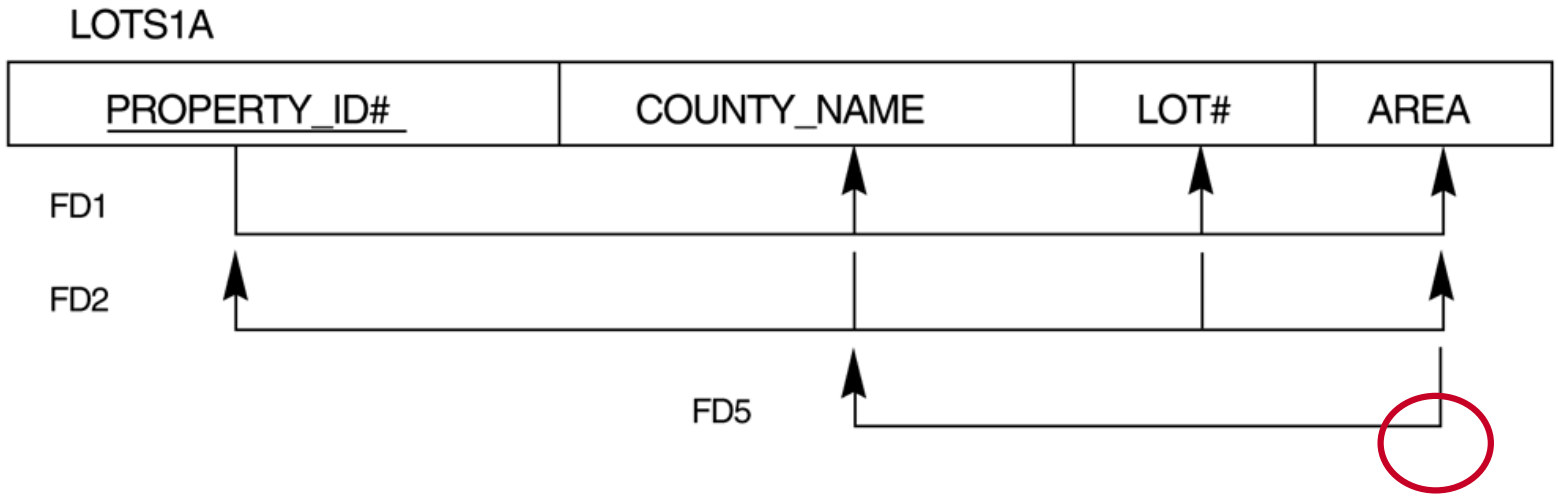
    X is a superkey of R

Boyce-Codd Normal Form (cont.)

- BCNF
  - A relation schema R is in BCNF

    if whenever a non-trivial functional dependency $X \rightarrow A$ holds in R, then

    X is a superkey of R

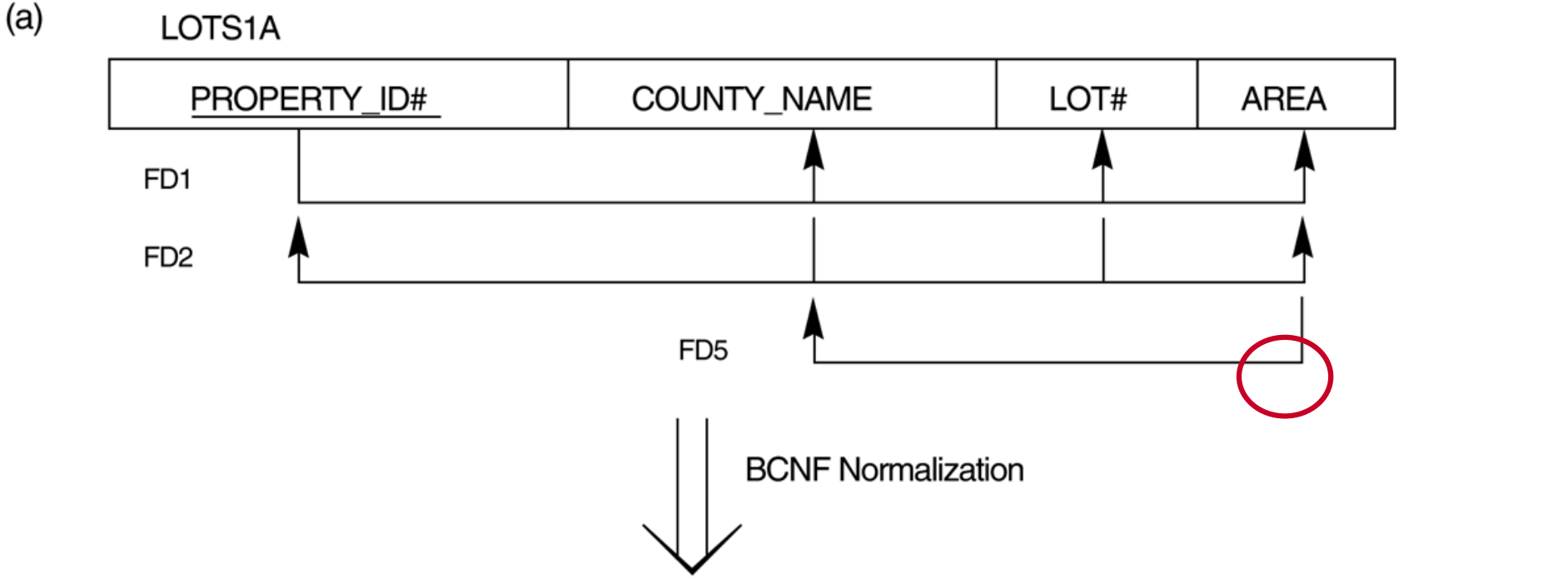


Assume that {AREA} → {COUNTY_NAME}
Area {1, 3, 5}: King county
Area {2, 4, 6}: Queen county

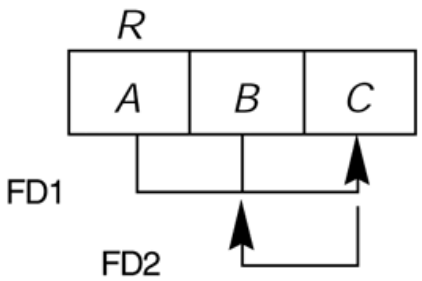
LOTS1A is in 3 NF
(FD1, FD2, FD5 do not violate general definition of 3NF)
but the redundancy appear in COUNTY_NAME

Property_ID#	County_Name	Lot#	Area
123456	Queen	123	6
123457	Queen	125	6
123458	Queen	136	2
123459	Queen	236	6
123460	King	125	1
123461	King	288	3
123462	King	369	1

(a)



(b)



Property_ID#	Lot#	Area
123456	123	6
123457	125	6
123458	136	2
123459	236	6
123460	125	1
123461	288	3
123462	369	1

Area	County_Name
6	Queen
4	Queen
2	Queen
1	King
3	King
5	King

Relationship between BCNF & 3NF

- BCNF
  - A relation schema R is in BCNF

    if whenever a non-trivial functional dependency $X \rightarrow A$ holds in R, then

    X is a superkey of R

* 3NF
  - A relation schema is in 3NF

    if whenever a nontrivial functional dependency $X \rightarrow A$ holds in R, either

    (a) X is a superkey of R **or**

    ~~(b) A is a prime attribute of R~~

General Definition of 3NF

- 3NF
  - A relation schema is in 3NF if whenever a nontrivial functional dependency $X \rightarrow A$ holds in R, either

    (a) X is a superkey of R

    (b) A is a prime attribute of R

    * Prime attribute: an attribute that is part of any candidate key

  * General Definition

  - A relation schema R is in 3NF

    if every nonprime attribute of R meets both of the following conditions

    - It is fully functionally dependent on every key of R
    - It is non-transitively dependent on every key of R

General Definition of 3NF (cont.)

- 3NF
  - A relation schema 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
  - A relation schema R **violates** 3NF

    if a functional dependency X→A holds in R

    that **violates** both

    (a) X is not superset of any key of R **and**

    (b) A is a nonprime attribute

General Definition of 3NF (cont.)

– A relation schema R **violates** 3NF

if a functional dependency X→A holds in R that both

(a) X is not superkey of R and

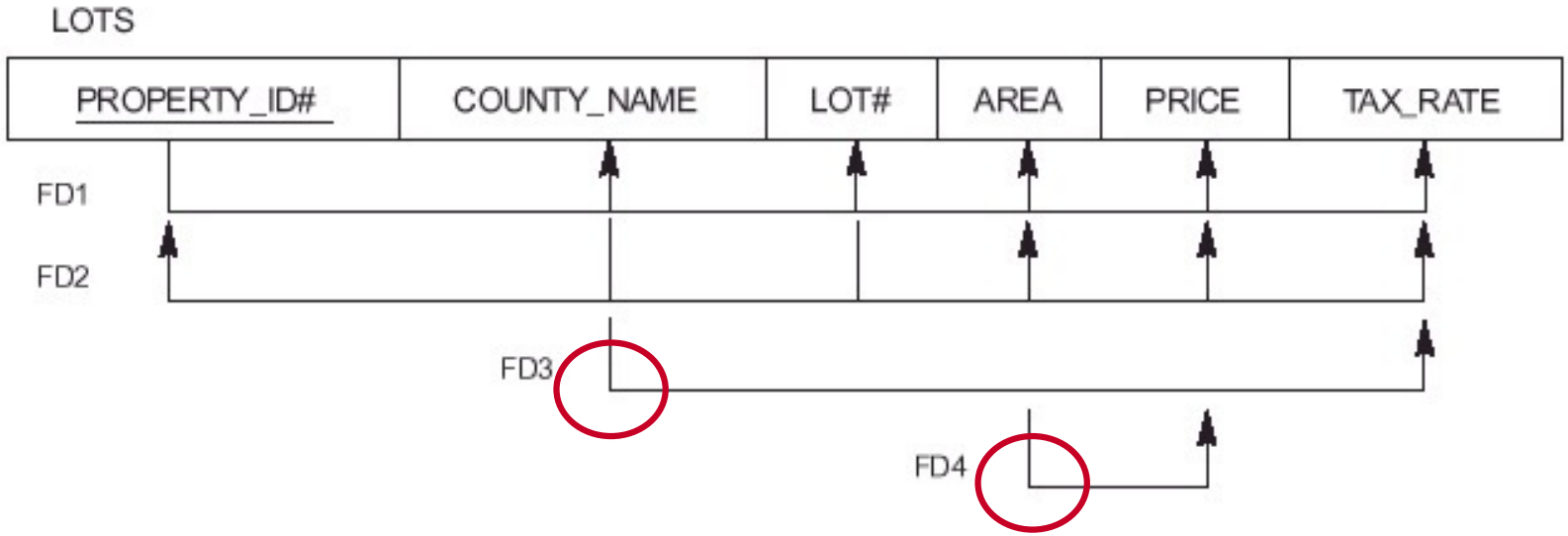
(b) A is a nonprime attribute

* X is not a superkey of R

=> X could be

– a proper subset of a key of R (partial dependency) or

– Nonprime (transitive dependency)



{COUNTY_NAME} → {TAX_RATE}

COUNTY_NAME is not superkey of LOTS,

TAX_RATE is a nonprime attribute of LOTS

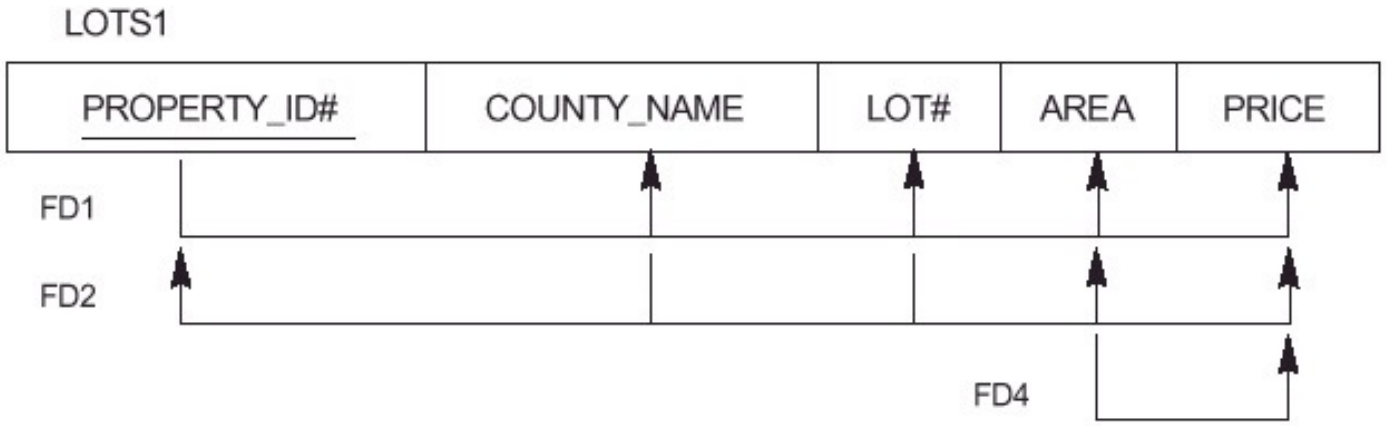
LOTS is not in 3NF

{AREA} → {PRICE}

AREA is not superkey of LOTS,

PRICE is a nonprime attribute of LOTS

LOTS is not in 3NF

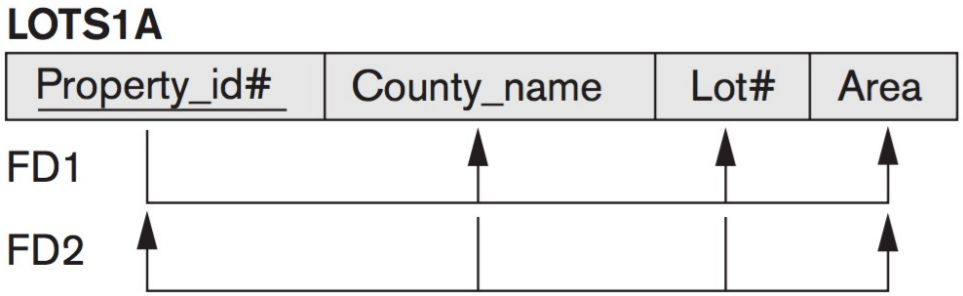


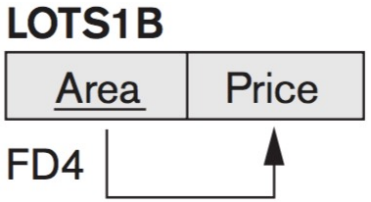
$\{Area\} \rightarrow \{Price\}$

Area is not superkey of LOTS1

Price is a nonprime attribute of LOTS1

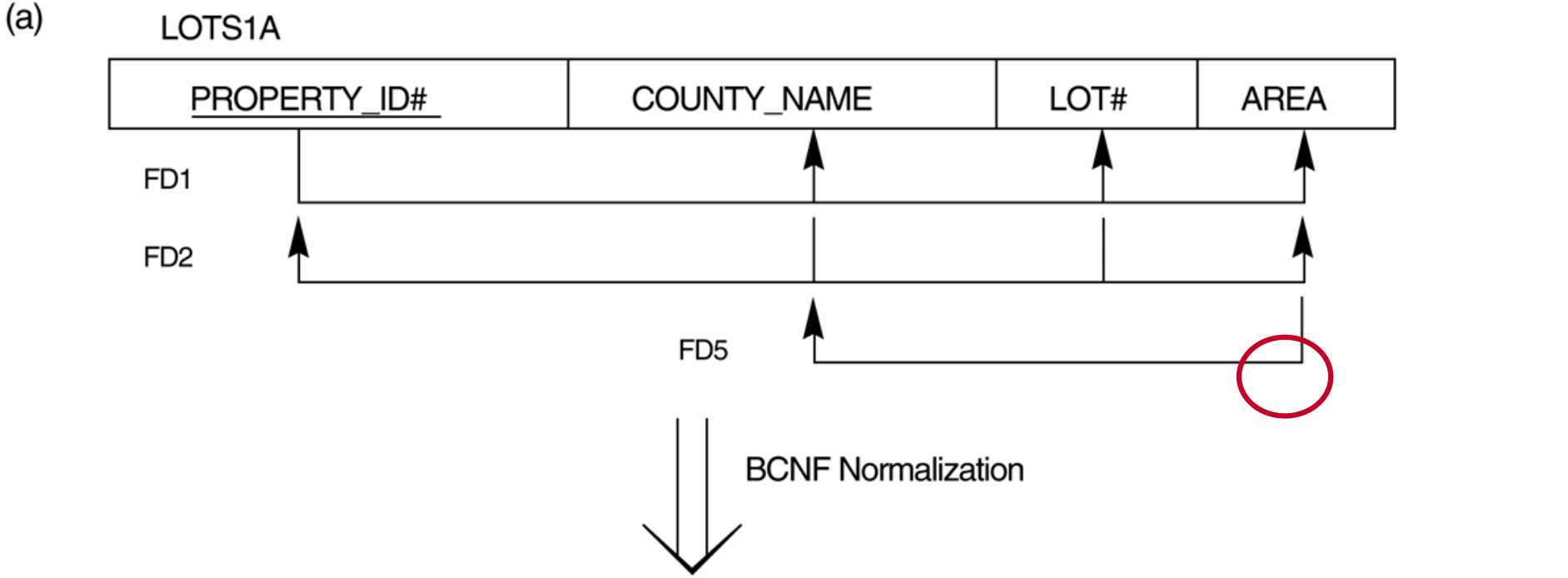
LOTS1 isn't in 3 NF



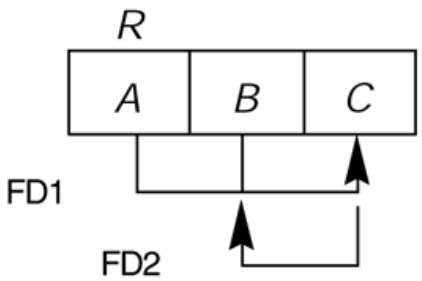


LOTS1A, LOTS1B are in 3 NF

(a)



(b)

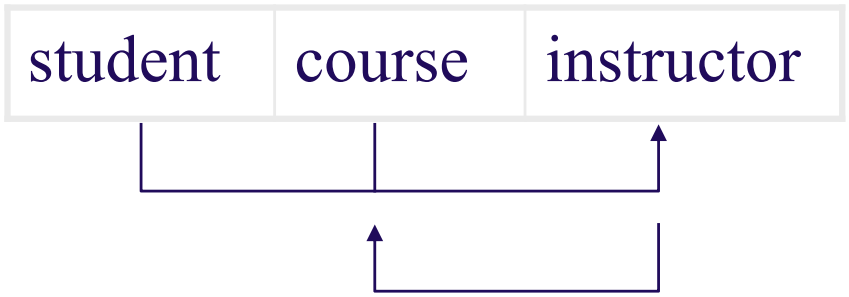


Property_ID#	Lot#	Area
123456	123	6
123457	125	6
123458	136	2
123459	236	6
123460	125	1
123461	288	3
123462	369	1

Area	County_Name
6	Queen
4	Queen
2	Queen
1	King
3	King
5	King

Another Example

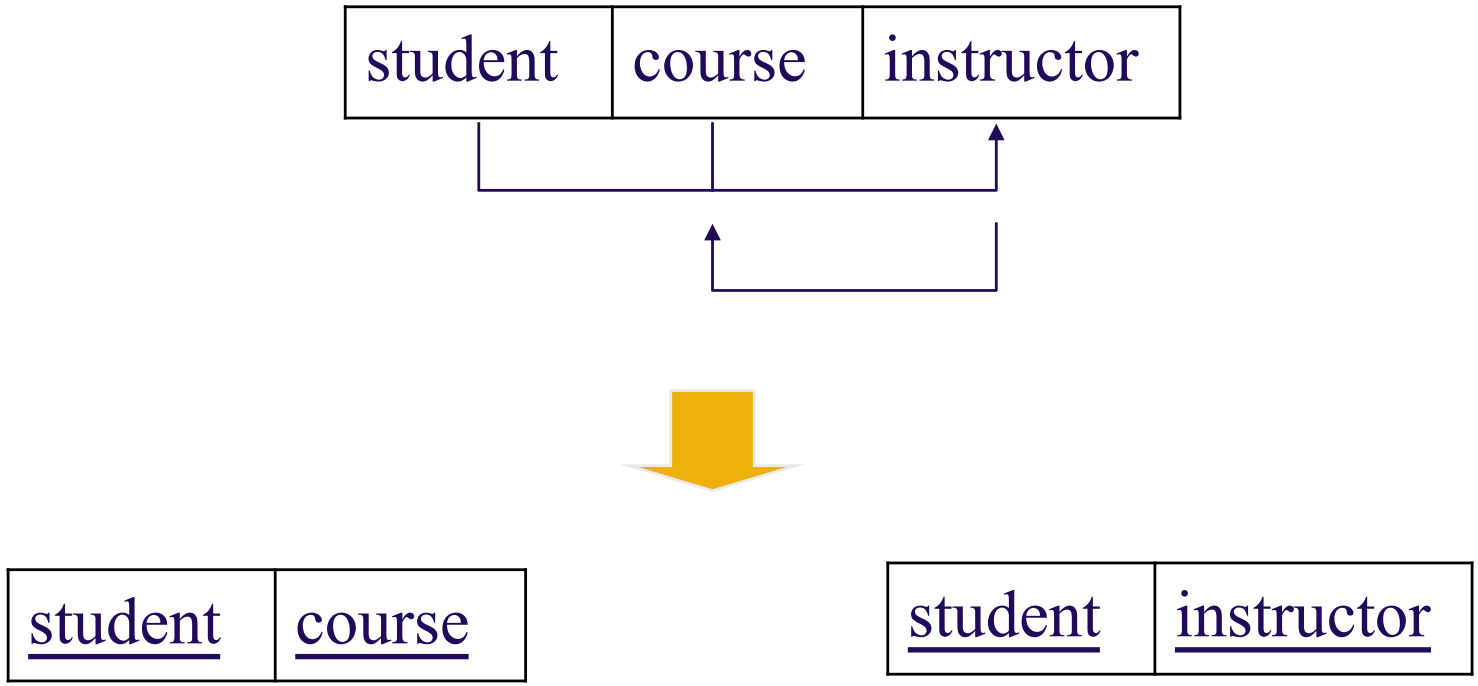
- Given
  - FD1: {student, course} -> {instructor}
  - FD2: {instructor} -> {course}



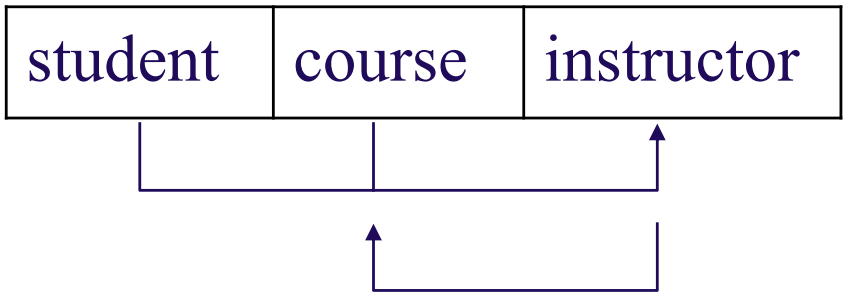
TEACH

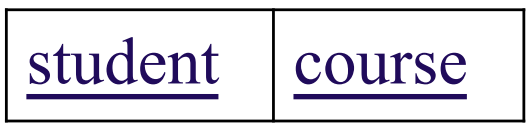
STUDENT	COURSE	INSTRUCTOR
Narayan	Database	Mark
Smith	Database	Navathe
Smith	Operating Systems	Ammar
Smith	Theory	Schulman
Wallace	Database	Mark
Wallace	Operating Systems	Ahamad
Wong	Database	Omiecinski
Zelaya	Database	Navathe

Decomposition Approach 1

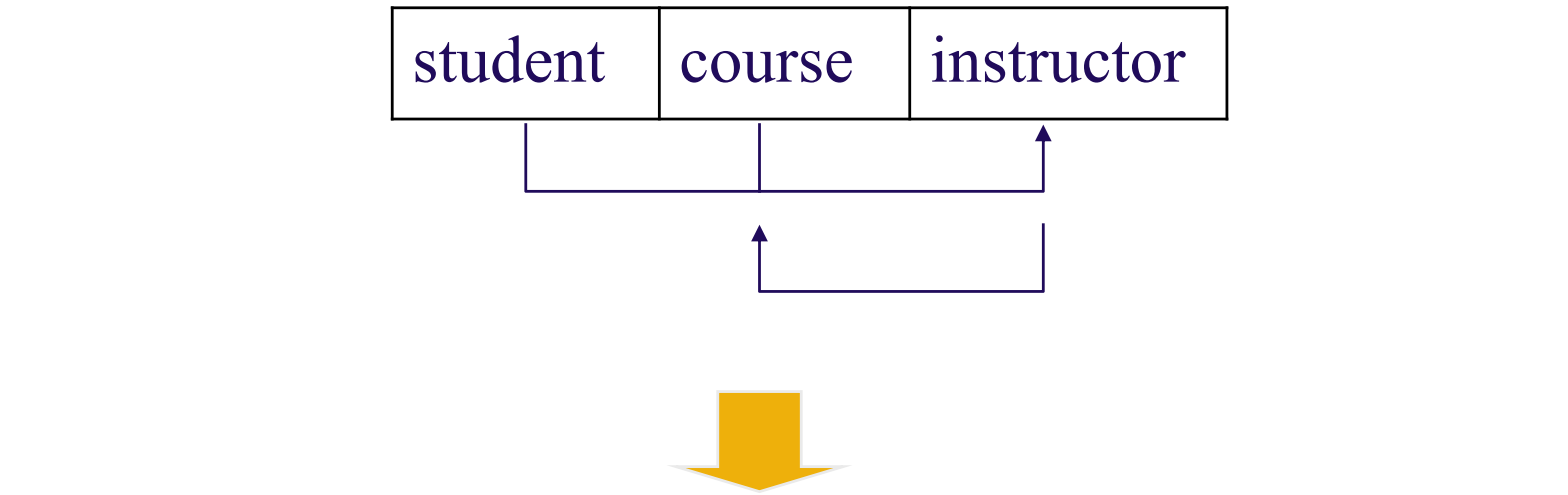


Decomposition Approach 2





Decomposition Approach 3



Conclusions

- Redundancy
  - Insertion anomaly
  - Deletion anomaly
  - Modification anomaly
- Normalization: decomposition based on functional dependency
  - First Normal Form: Relational Schema
  - Second Normal Form: full functional dependency
  - Third Normal Form: non-transitive functional dependency
  - Boyce-Codd Normal Form

Normalization (cont.)

- Relations may be left in a lower normalization status for performance reasons
- Denormalization: process of storing the join of higher normal form relations as a base relation

  (which is in a lower normal form)
- In general, it is advisable to
  - use normalized base relations
  - specify views that include the JOINs for placing together the attributes frequently referenced in important queries