COSC 3380
Design of Database Systems

Functional Dependencies and Normalization for Relational Databases

April 1, 2024

First Normal Form

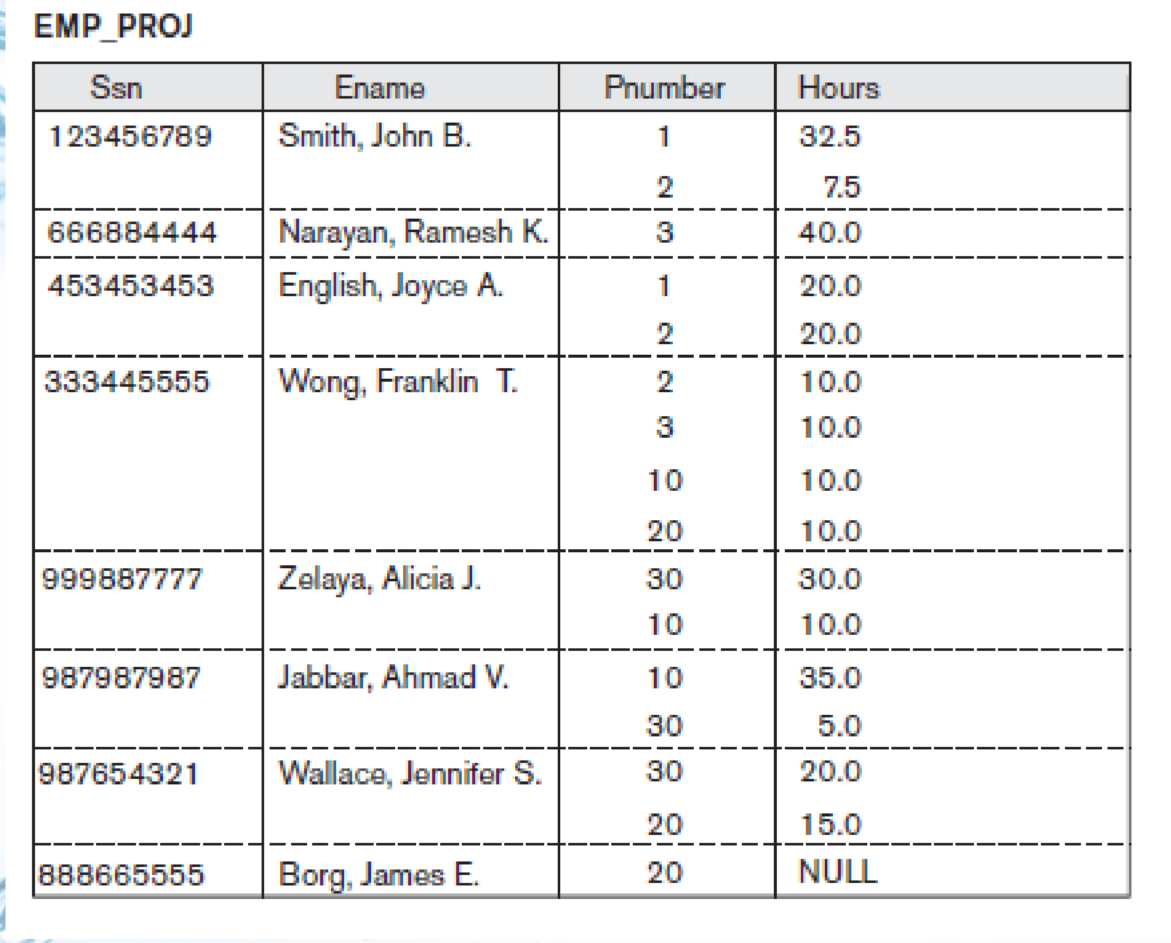
- Domain of an attribute must include only **atomic** (**simple, indivisible**) values
- Value of any attribute in a tuple must be a single value
- Disallows a set of values as an attribute value in a single tuple
  - Disallows composite attributes
  - Disallows multivalued attributes
  - Disallows relations within relations
- Most RDBMSs allow only those relations to be defined that are in 1NF

3 Techniques to achieve 1NF

- Remove attribute that violates 1NF and place in separate relation
- Expand the key
  - Introduces redundancy
- Use several atomic attributes
  - NULL Values
  - Querying and ordering problems
  - Scalability?

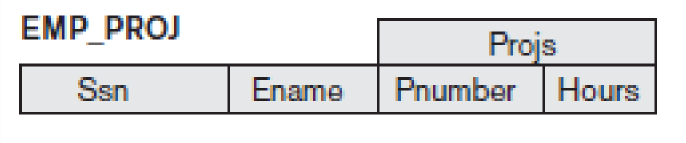
First Normal Form

EMP_PROJ(Ssn, Ename, {PROJS(Pnumber, Hours)})



EMP_PROJ

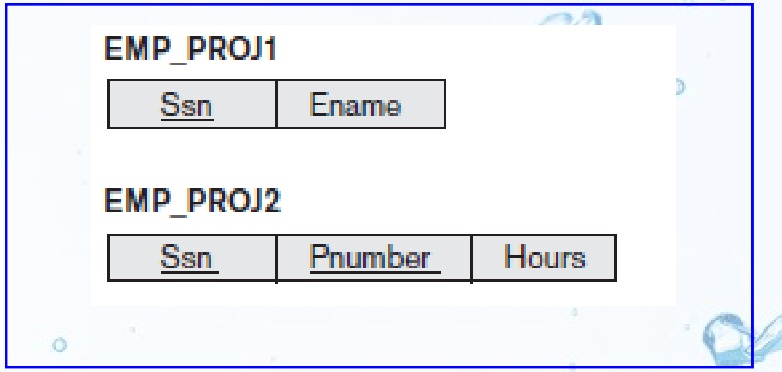
Ssn	Ename	Pnumber	Hours
123456789	Smith, John B.	1 2	32.5 7.5
666884444	Narayan, Ramesh K.	3	40.0
453453453	English, Joyce A.	1 2	20.0 20.0
333445555	Wong, Franklin T.	2 3 10 20	10.0 10.0 10.0 10.0
999887777	Zelaya, Alicia J.	30 10	30.0 10.0
987987987	Jabbar, Ahmad V.	10 30	35.0 5.0
987654321	Wallace,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



EMP_PROJ

		Projs	
Ssn	Ename	Pnumber	Hours

Decompose by propagating the primary key



EMP_PROJ1

Ssn	Ename

EMP_PROJ2

Ssn	Pnumber	Hours

First Normal Form

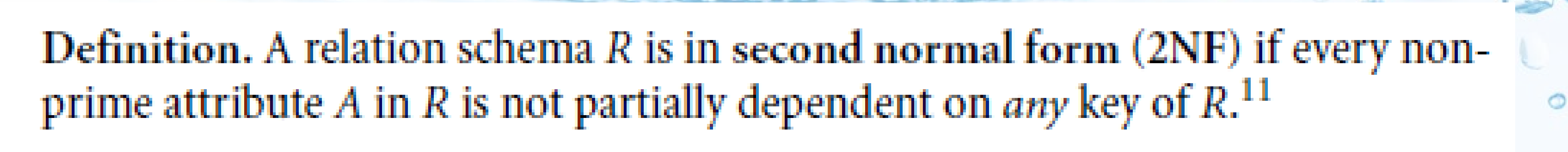
- Does not allow nested relations
- To change a relation to 1NF:
  - Move nested relation attributes into a new relation
  - Propagate the primary key into it
  - **Unnest** relation into a set of 1NF relations

Second Normal Form

- Uses the concepts of **FDs, primary key**
  - **Full functional dependency:**
  - For a FD X → Y where removal of any attribute A from X makes the FD not hold any more
  - (SSN, PNUMBER) →HOURS
  - (SSN, PNUMBER) → ENAME
    - SSN → ENAME holds
    - **Partial dependency**

Second Normal Form

- **Prime attribute**
  - Part of any candidate key will be considered as prime

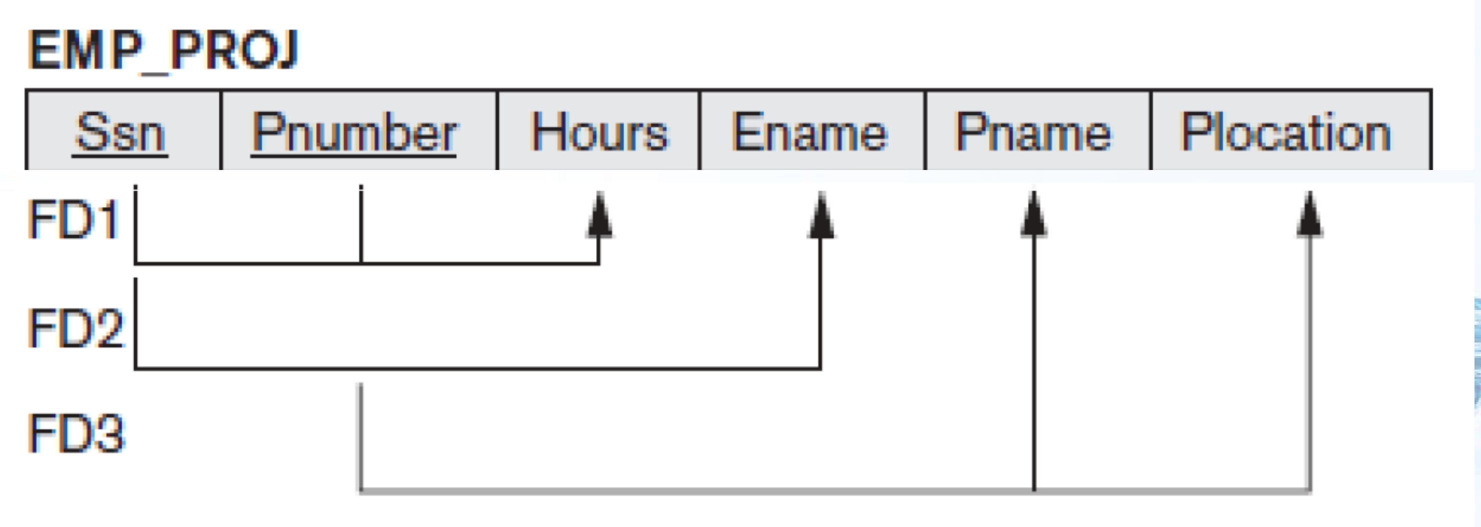


Definition. A relation schema *R* is in **second normal form (2NF)** if every nonprime attribute *A* in *R* is not partially dependent on *any* key of *R*.[11]

Second Normal Form

- A relation schema R is in **second normal form (2NF)** if every non-prime attribute A in R is fully functionally dependent on any key of R.
- Test for 2NF: Check FDs whose left-hand side attributes are part of the primary key
  - If the primary key contains a single attribute, no test needed.

Second Normal Form



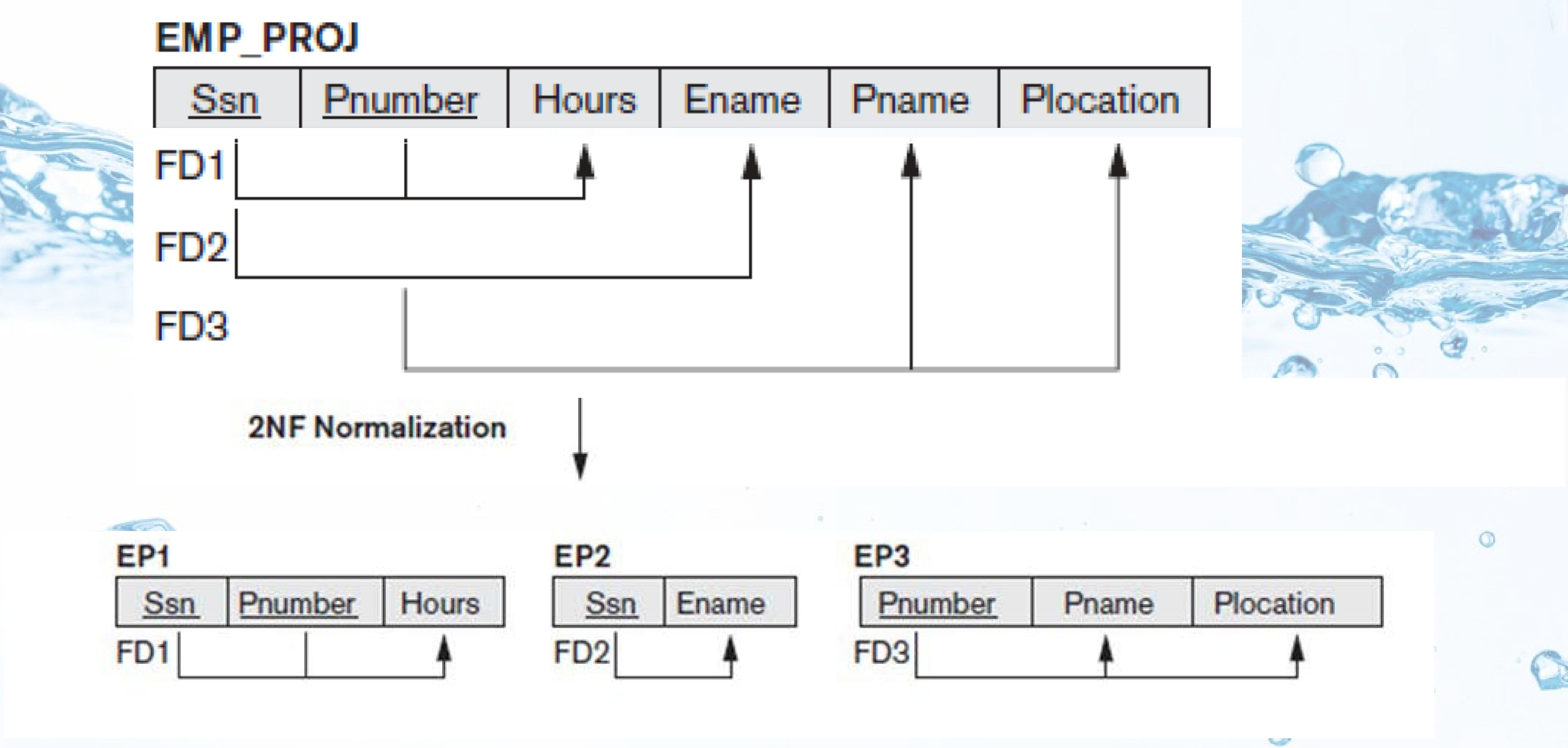
a. Ssn → Ename

b. Pnumber →{Pname, Plocation}

c. {Ssn, Pnumber} → Hours

Ssn and Pnumber are a part of the primary key {Ssn,Pnumber} violating the 2NF test

Second Normal Form



EMP_PROJ

Ssn	Pnumber	Hours	Ename	Pname	Plocation

FD1

FD2

FD3

2NF Normalization

EP1

Ssn	Pnumber	Hours

FD1

EP2

Ssn	Ename

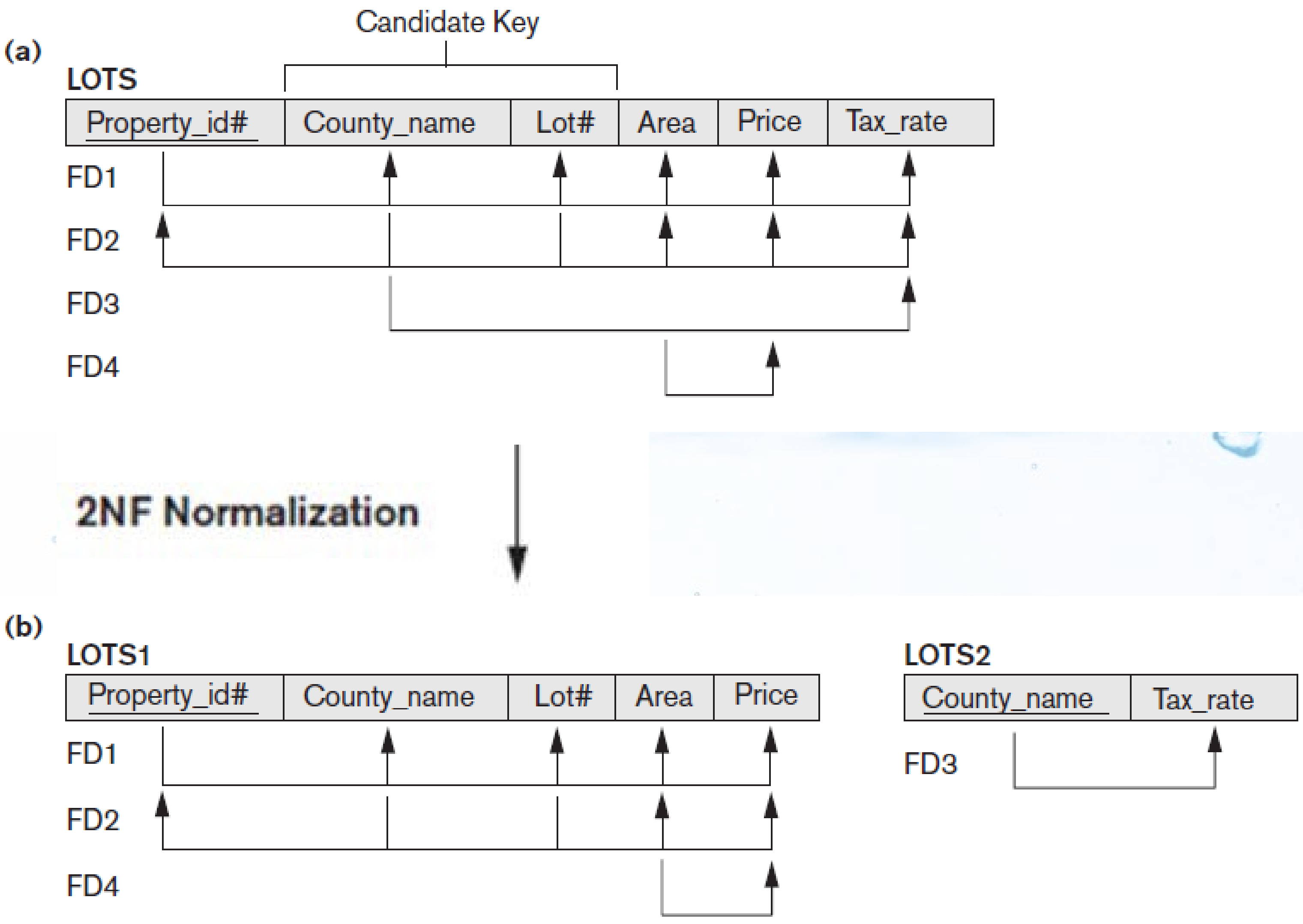
FD2

EP3

Pnumber	Pname	Plocation

FD3

Second Normal Form



(a)

Candidate Key

LOTS

Property_id# | County_name | Lot# | Area | Price | Tax_rate

FD1

FD2

FD3

FD4

2NF Normalization

(b)

LOTS1

Property_id# | County_name | Lot# | Area | Price

FD1

FD2

FD4

LOTS2

County_name | Tax_rate

FD3