Third Normal Form

- ***No partial dependency or transitive dependency from any key to any non-key attribute***
- Let X, Y be two <u>*sets*</u> of attributes, Z be a <u>*single*</u> attribute
- Partial dependency from some key to some non-key attribute
  - There exists some (non-trivial) FD: $X \longrightarrow Z$
  - X is a proper subset of some key, Z is NOT in any key
- Transitive dependency from some key to some non-key attribute
  - There exists some (non-trivial) FD: $Y \longrightarrow Z$
  - Y is NOT a superkey or a proper subset of any key
  - Z is NOT in any key

Third Normal Form

- *If there exists* partial dependency or transitive dependency from some key to some non-key attribute
- *Then there exists* some (non-trivial) FD: W ⟶ Z such that W is NOT a superkey *and* Z is NOT in any key
- A relation schema R is in third normal form (3NF) if
  for each (non-trivial) FD: W ⟶ Z
  either W is a superkey or Z is in some key

Superkeys

- Suppose R(A, B, C, D, E) has two candidate keys {A, B} and {C, D}
- What is a superkey:
  - {A, B}, {C, D}
  - {A, B, C}, {A, B, D}, {A, B, E}, {A, C, D}, {B, C, D}, {C, D, E}
  - {A, B, C, D}, {A, B, C, E}, {A, B, D, E}, {A, C, D, E}, {B, C, D, E}
  - {A, B, C, D, E}
- What is NOT a superkey:
  - {A}, {B}, {C}, {D}, {E}
  - {A, C}, {A, D}, {A, E}, {B, C}, {B, D}, {B, E}, {C, E}, {D, E}
  - {A, C, E}, {A, D, E}, {B, C, E}, {B, D, E}

Third Normal Form

- A relation schema R is in third normal form (3NF) if
  for each (non-trivial) FD: W ⟶ Z
  either W is a superkey or Z is in some key

- A (non-trivial) FD: W ⟶ Z where
  W is not a superkey but Z is in some key
  can <u>*cause redundancy*</u>

- {Student, Course} ⟶ {Instructor}
- {Instructor} ⟶ {Course}
- {Student, Course} is the only key

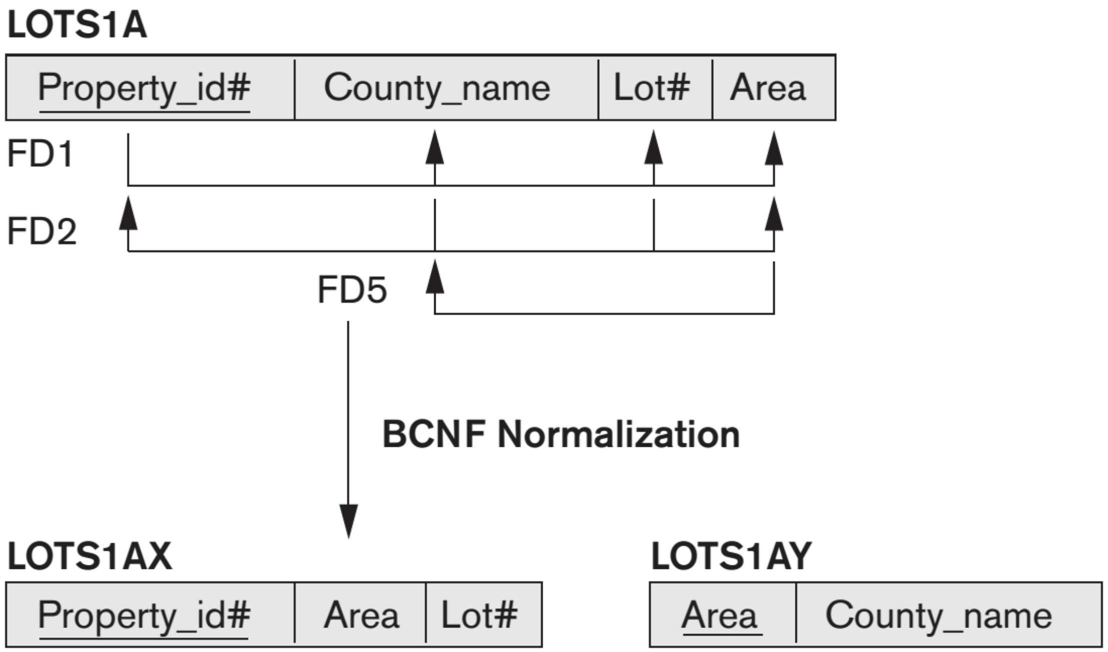
TEACH

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

Boyce-Codd Normal Form

- A relation schema R is in Boyce-Codd normal form (BCNF) if ***for each (non-trivial) FD: W ⟶ Z, W is a superkey***

- Example
  - R(A, B, C, D)
  - {A} ⟶ {B, C}
  - {B, C} ⟶ {A, D}
  - {D} ⟶ {B}
- Candidate keys
  - {A}, {B, C}



Boyce-Codd Normal Form

- **<u>Fact:</u>** Any two-attribute relation R(A, B) is in BCNF

- {A, B} is the key
  - no non-trivial FD's
- {A} is the only key (or symmetrically, {B} is the only key)
  - $A \longrightarrow B$
- Both {A} and {B} are keys
  - $A \longrightarrow B$
  - $B \longrightarrow A$

Decomposition into BCNF

- Find BCNF violations
- Decompose into two smaller relations
- Recurse on the two smaller relations
- Until there is no BCNF violation

- How to find BCNF violations in R?
  - Find a set of attributes A such that $A^+ \neq A$ and $A^+ \neq$ {all attributes}
  - A^+ is computed with respect to the entire set of FD's, and then discard attributes that are not in R

Suppose $A_1,\ldots,A_n \longrightarrow B_1$ violates BCNF
(i.e., $\{A_1,\ldots,A_n\}$ is not a superkey)
Find all attributes $B_1,\ldots,B_m$ such that
$A_1,\ldots,A_n \longrightarrow B_1,\ldots,B_m$

$R(A_1,\ldots,A_n,B_1,\ldots,B_m,C_1,\ldots,C_k)$

projection

$R_1(A_1,\ldots,A_n,B_1,\ldots,B_m)$

$R_2(A_1,\ldots,A_n,C_1,\ldots,C_k)$

$\{A_1,\ldots,A_n\}^+ = \{A_1,\ldots,A_n,B_1,\ldots,B_m\}$

Closure Algorithm

R(A, B, C, D, E, F)

$\{A, B\} \rightarrow \{C\}$
$\{A, D\} \rightarrow \{E\}$
$\{B\} \rightarrow \{D\}$
$\{A, F\} \rightarrow \{B\}$

Compute $\{A, B\}^+ = \{A, B, \qquad \}$

Compute $\{A, F\}^+ = \{A, F, \qquad \}$

Closure Algorithm

R(A, B, C, D, E, F)

$\{A, B\} \rightarrow \{C\}$
$\{A, D\} \rightarrow \{E\}$
$\{B\} \rightarrow \{D\}$
$\{A, F\} \rightarrow \{B\}$

Compute $\{A, B\}^+ = \{A, B, C, \quad \}$

Compute $\{A, F\}^+ = \{A, F, \quad \}$

Closure Algorithm

R(A, B, C, D, E, F)

{A, B} → {C}
{A, D} → {E}
{B} → {D}
{A, F} → {B}

Compute $\{A, B\}^+ = \{A, B, C, D, \quad \}$

Compute $\{A, F\}^+ = \{A, F, \qquad\qquad \}$

Closure Algorithm

R(A, B, C, D, E, F)

$\{A, B\} \rightarrow \{C\}$
$\{A, D\} \rightarrow \{E\}$
$\{B\} \rightarrow \{D\}$
$\{A, F\} \rightarrow \{B\}$

Compute $\{A, B\}^+ = \{A, B, C, D, E\}$

Compute $\{A, F\}^+ = \{A, F, \qquad\qquad \}$

Closure Algorithm

R(A, B, C, D, E, F)

$\{A, B\} \rightarrow \{C\}$
$\{A, D\} \rightarrow \{E\}$
$\{B\} \rightarrow \{D\}$
$\underline{\{A, F\} \rightarrow \{B\}}$

Compute $\{A, B\}^+ = \{A, B, C, D, E\}$

Compute $\{A, F\}^+ = \{A, F, B, \qquad \}$

Closure Algorithm

R(A, B, C, D, E, F)

$\{A, B\} \rightarrow \{C\}$
$\{A, D\} \rightarrow \{E\}$
$\{B\} \rightarrow \{D\}$
$\{A, F\} \rightarrow \{B\}$

Compute $\{A, B\}^+ = \{A, B, C, D, E\}$

Compute $\{A, F\}^+ = \{A, F, B, C, D, E\}$

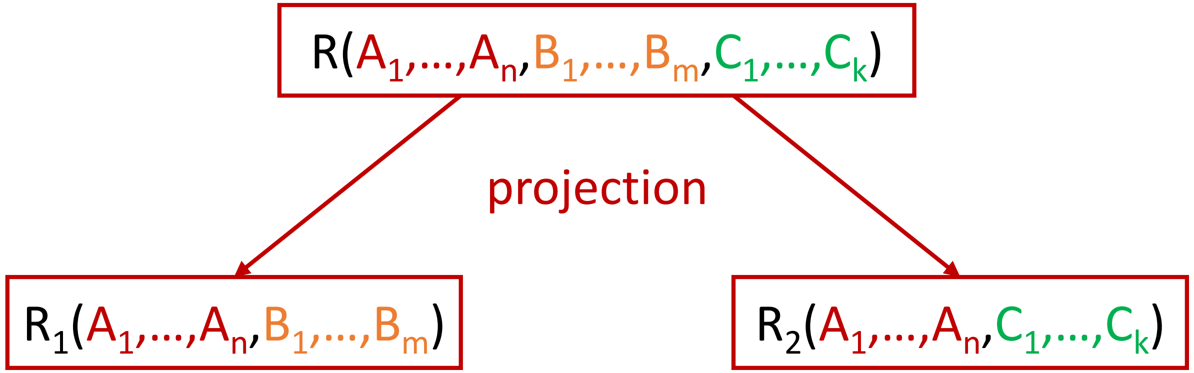
Decomposition into BCNF

R(A,B,C,D,E)

{A} → {B,C}
{C} → {D}

- Find BCNF violations
- Decompose into two smaller relations
- Recurse on the two smaller relations
- Until there is no BCNF violation

- How to find BCNF violations in R?
  - Find a set of attributes A such that $A^+ \neq A$ and $A^+ \neq$ {all attributes}
  - A^+ is computed with respect to the entire set of FD's, and then discard attributes that are not in R

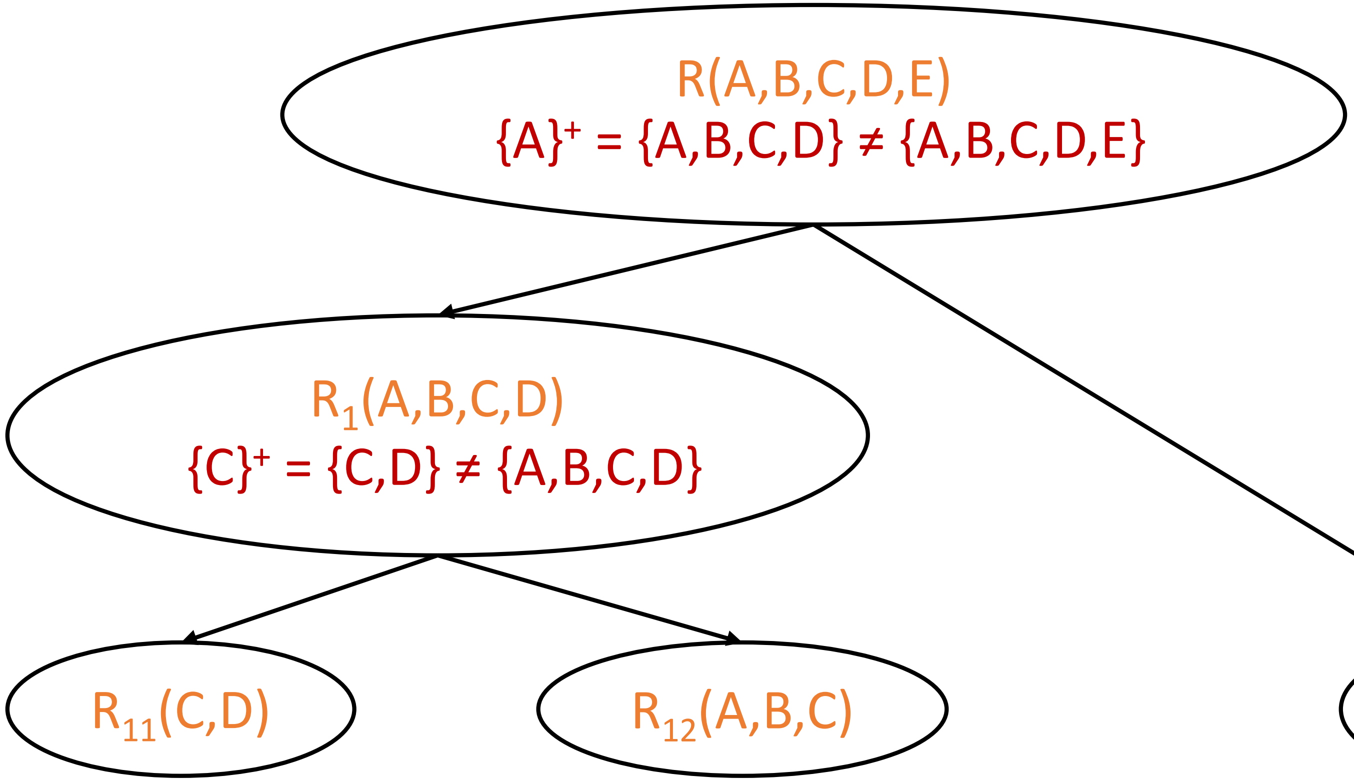


$\{A_1,\ldots,A_n\}^+ = \{A_1,\ldots,A_n,B_1,\ldots,B_m\}$

Decomposition into BCNF

R(A,B,C,D,E)

{A} → {B,C}
{C} → {D}



Lossy Decomposition

- Since any two-attribute relation is in BCNF
  why not randomly decompose into two-attribute relations?

Name	Price	Category
Gizmo	19.99	Gadget
OneClick	24.99	Camera
Gizmo	19.99	Camera

Name	Category
Gizmo	Gadget
OneClick	Camera
Gizmo	Camera

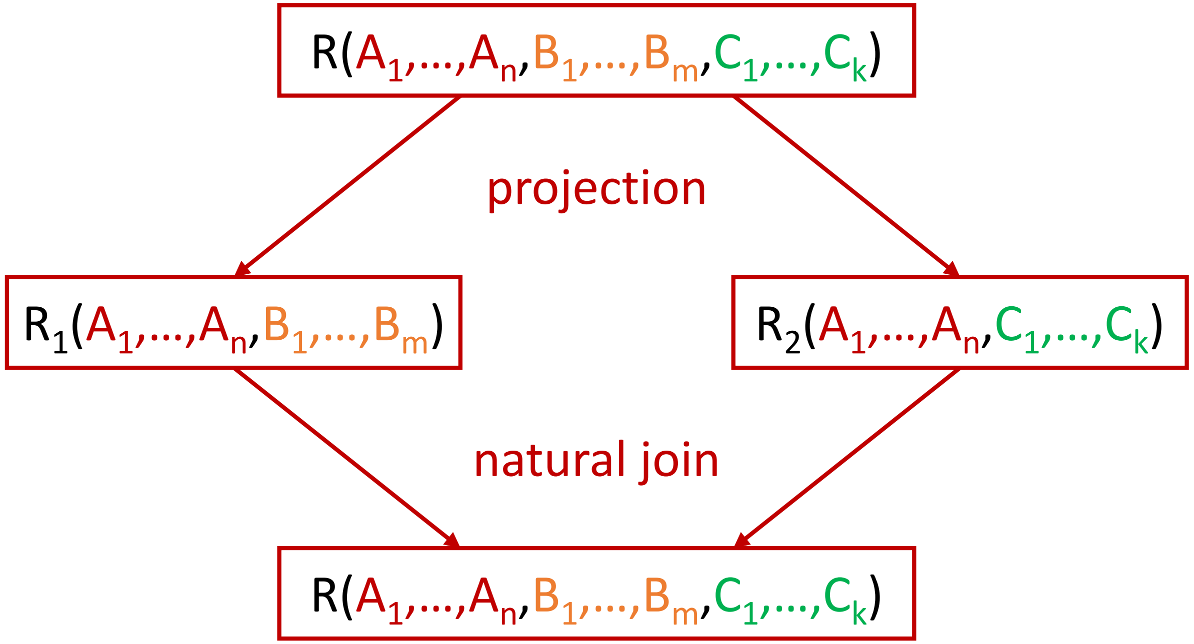
$\bowtie$

Price	Category
19.99	Gadget
24.99	Camera
19.99	Camera

Name	Price	Category
Gizmo	19.99	Gadget
OneClick	24.99	Camera
Gizmo	19.99	Camera
OneClick	**19.99**	**Camera**
Gizmo	**24.99**	**Camera**

Lossless Decomposition

- A decomposition of R into R_1 and R_2 is ***lossless*** if $R_1 \bowtie R_2 = R$



If A ⟶ B or A ⟶ C, then
the decomposition is lossless

BCNF decomposition is lossless!