

# PART 2: NORMAL FORMS

# Motivation for normal forms

- Identify a “good” schema
  - For some definition of “good”
  - Avoid anomalies, redundancy, etc.
- Many normal forms
  - 1<sup>st</sup>
  - 2<sup>nd</sup>
  - 3<sup>rd</sup>
  - Boyce-Codd
  - ... and several more we won't discuss...

*$BCNF \subseteq 3NF \subseteq 2NF \subseteq 1NF$  (focus on 3NF/BCNF)*

# First Normal Form: 1-NF

- Column names should be unique
- No duplicate rows
- Ordering between rows doesn't matter
- Each column should have atomic domain
- No multi-valued attributes
- No composite/complex attributes

Name	Address	Phone	Age
Sam	123, Main St, Allendale, MI-49401	999-888-7777, 111-222-3333	"23"
John	123, pacific St, GR, MI-49525	444-555-6666	34
John	123, pacific St, GR, MI-49525	123-456-7890	34

# Second Normal Form : 2-NF

- Must be in 1-NF
- No partial dependency
  - All non-prime attributes should be fully dependent on the candidate key

## 2-NF

Student ID	Course ID	Course Fee
Composite Key		
1	1	500
1	2	1000
2	4	200
2	3	750
3	5	1000
3	3	750

## 2-NF

- $R(A, B, C, D)$
- $FD = \{AB \rightarrow C, C \rightarrow D, B \rightarrow C\}$

## 2-NF

- $R(A, B, C, D)$
- $FD = \{AB \rightarrow C, C \rightarrow D, B \rightarrow C\}$

$CK = (AB)$  ,  $PA = \{A, B\}$  ,  $NPA = \{C, D\}$

Dependencies	PD?
$AB \rightarrow C$	No
$C \rightarrow D$	No
$B \rightarrow C$	Yes

## 2-NF

- $R(A, B, C, D, E, F)$  ,
- $FD = \{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow E\}$



## 2-NF

- $R(A, B, C, D, E, F)$  ,
- $FD = \{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow E\}$

$CK = (AF)$ ,  $PA = \{A, F\}$ ,  $NPA = \{B, C, D, E\}$

Dependencies	PD?
$A \rightarrow B$	Yes
$B \rightarrow C$	No
$C \rightarrow D$	No
$D \rightarrow E$	No

## 2-NF

- $R(A, B, C, D)$
- $FD = \{AB \rightarrow CD, C \rightarrow A, D \rightarrow B\}$

## 2-NF

- $R(A, B, C, D)$
- $FD = \{AB \rightarrow CD, C \rightarrow A, D \rightarrow B\}$

$CK = \{AB, BC, CD\}$  ,  $PA = \{A, B, C, D\}$ ,  $NPA = \{ \}$

Dependencies	PD?
$AB \rightarrow CD$	No
$C \rightarrow A$	No
$D \rightarrow B$	No

# Third Normal Form : 3-NF

- Must be in 1-NF and 2-NF
- No transitive dependency
  - All fields must be determinable by the primary key only, not by any other keys
  - No such incidents of FDs like:
    - Non-prime attribute  $\rightarrow$  Non-prime attribute

# 3-NF

## Composite Key

Tournament Name	Year	Winner	Winner's DOB
Indiana Invitational	1998	Al Fredrickson	21 July 1975
Cleveland Open	1999	Bob Albertson	28 September 1968
Des Moines Masters	1999	Al Fredrickson	21 July 1975
Indiana Invitational	1999	Chip Masterson	14 March 1977

# 3-NF

- $R(A,B,C,D)$
- $FD = \{ A \rightarrow B, B \rightarrow C, C \rightarrow D \}$

# 3-NF

- $R(A,B,C,D)$
- $FD = \{ A \rightarrow B, B \rightarrow C, C \rightarrow D \}$

$CK = \{A\}$ ,  $PA = \{A\}$ ,  $NPA \{B, C, D\}$

Dependencies	PD?	TD?
$A \rightarrow B$	No	No
$B \rightarrow C$	No	Yes
$C \rightarrow D$	No	Yes

# 3-NF

- $R(A,B,C,D,E,F)$
- $FD = \{ AB \rightarrow CDEF, BD \rightarrow F \}$



# 3-NF

- $R(A,B,C,D,E,F)$
- $FD = \{ AB \rightarrow CDEF, BD \rightarrow F \}$

$CK = \{AB\}$ ,  $PA = \{AB\}$ ,  $NPA \{C, D, E, F\}$

## 3-NF

- $R(\text{instructor}, \text{salary}, \text{dept}, \text{building})$
- $\text{FD} = \{ \text{instructor} \rightarrow R, \text{dept} \rightarrow \text{building} \}$

# 3-NF

- $R(\text{instructor}, \text{salary}, \text{dept}, \text{building})$
- $\text{FD} = \{ \text{instructor} \rightarrow R, \text{dept} \rightarrow \text{building} \}$

$\text{CK} = \{ (\text{instructor}) \},$

$\text{PA} = \{\text{instructor}\}, \text{NPA} = \{\text{salary}, \text{dept}, \text{building}\}$

Dependencies	PD?	TD?
$\text{instructor} \rightarrow R$	No	No
$\text{dept} \rightarrow \text{building}$	No	Yes

## 3-NF

- $R(A, B, C, D, E, F, G, H, I)$
- $FD = \{ AB \rightarrow C, AD \rightarrow GH, BD \rightarrow EF, A \rightarrow I, H \rightarrow J \}$

# Boyce-Codd Normal Form : BCNF

- Must be in 3-NF
- For all the dependencies, the LHS should be a super (candidate) key

# BCNF

- $R(A, B, C)$
- $FD = \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$

# BCNF

- $R(A, B, C)$
- $FD = \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$

$CK = \{A, B, C\}$  ,  $PA = \{A, B, C\}$  ,  $NPA = \{ \}$

Depend encies	PD?	TD?	BCNF Violation?
$A \rightarrow B$	No	No	No
$B \rightarrow C$	No	No	No
$C \rightarrow A$	No	No	No

# BCNF

- $R(J, K, L)$
- $F = \{ JK \rightarrow L, L \rightarrow K \}$



# BCNF

- $R(J, K, L)$
- $F = \{ JK \rightarrow L, L \rightarrow K \}$

$CK = \{ (JK), (JL) \}$ ,  $PA = \{ J, K, L \}$ ,  $NPA = \{ \}$

Dependencies	PD?	TD?	BCNF Violation?
$JK \rightarrow L$	No	No	No
$L \rightarrow K$	No	No	YES