

Midterm Help Session

FD and Normalization

CS127 TAs

Oct 28th, 2009

Normalization

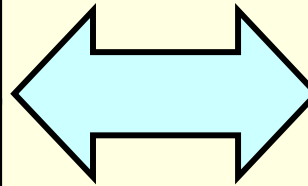
- Eliminate redundant data
 - Save disk space
 - Reduce update overhead/anomalies
- Decompose tables based on FD
 - $1NF \rightarrow 2NF \rightarrow 3NF \rightarrow BCNF \rightarrow 4NF \rightarrow 5NF$

1NF 1st Normal Form

- *All attributes are Atomic*

1NF

<u>Name</u>	<u>Age & Sex</u>
Hulk	56M

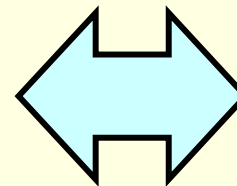


<u>Name</u>	<u>Age</u>	<u>Sex</u>
Hulk	56	M

- *No Repeating Attributes*

1NF

<u>Name</u>	<u>Class1</u>	<u>Class2</u>	<u>Class3</u>
Hulk	CS12	CS23	CS57



<u>Name</u>	<u>Class</u>
Hulk	CS12
Hulk	CS23
Hulk	CS57

Okay, Which is better?

<u>Name</u>	<u>A & S</u>	<u>Class1</u>	<u>Class2</u>
Hulk	56M	CS12	CS23
Linda	54F	CS46	CS57

Both are 1NF.
BUT....

<u>Name</u>	<u>A</u>	<u>S</u>	<u>Class</u>
Hulk	56	M	CS12
Hulk	56	M	CS23
Linda	54	F	CS46
Linda	54	F	CS57

<u>Name</u>	<u>A</u>	<u>S</u>
Hulk	56	M
Linda	54	F

+

<u>Name</u>	<u>Class</u>
Hulk	CS12
Hulk	CS23
Linda	CS46
Linda	CS57

FD: Functional Dependency

<u>Name</u>	<u>A</u>	<u>S</u>	<u>Class</u>
Hulk	56	M	CS12
Hulk	56	M	CS23
Linda	54	F	CS46
Linda	54	F	CS57

Name determines A, S (not class)

FD: $\text{Name} \rightarrow \{\text{A}, \text{S}\}$

Trivial FD: $\{\text{Name}, \text{A}\} \rightarrow \text{Name}$

super key: $\{\text{Name}, \text{A}, \text{S}, \text{Class}\}$, $\{\text{Name}, \text{A}, \text{Class}\}$,
 $\{\text{Name}, \text{S}, \text{Class}\}$, $\{\text{Name}, \text{Class}\}$

candidate key: $\{\text{Name}, \text{Class}\}$

key attributes: $\{\text{Name}, \text{Class}\}$

non-key attributes: $\{\text{A}, \text{S}\}$

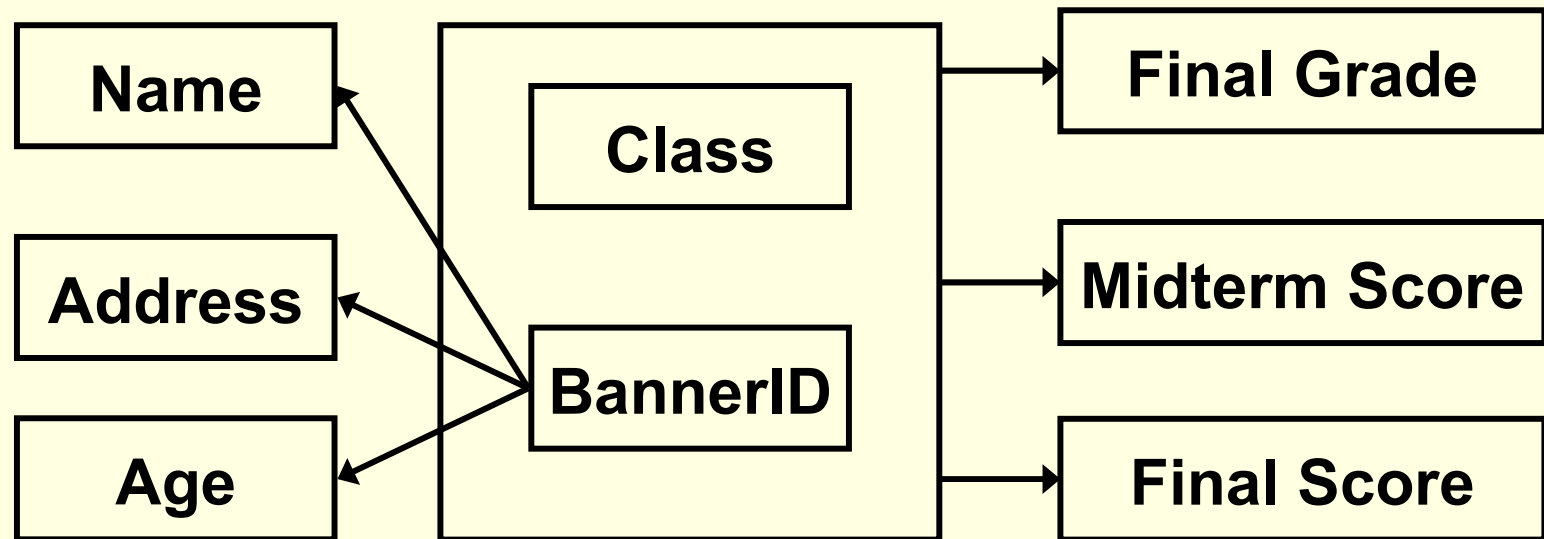
Type of FD

- $\{A,B\} \rightarrow C$, but neither $A \rightarrow C$ nor $B \rightarrow C$
Full FD: C is fully dependent on $\{A,B\}$
- $\{A,B\} \rightarrow C$, but $A \rightarrow C$ or $B \rightarrow C$
Partial FD: C is partially dep. on $\{A,B\}$
- $A \rightarrow B \rightarrow C$
Transitive FD: C is transitively dep. on A
(except when $B \rightarrow A$)

Analyze FDs to normalize!

Use your hand! - FD Diagram

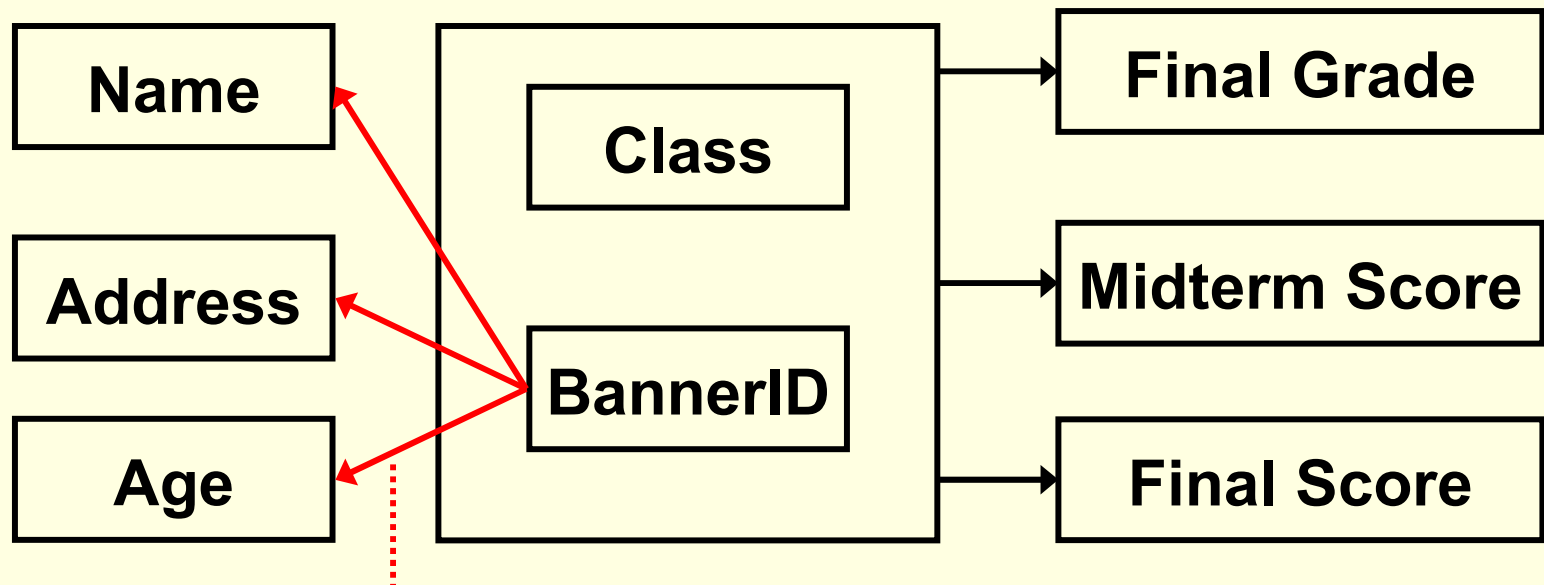
Box: Attribute, Arrow: FD



***BannerID* → {Name, Address, Age}**
{Class, BannerID} → {Grade, MScore, FScore}

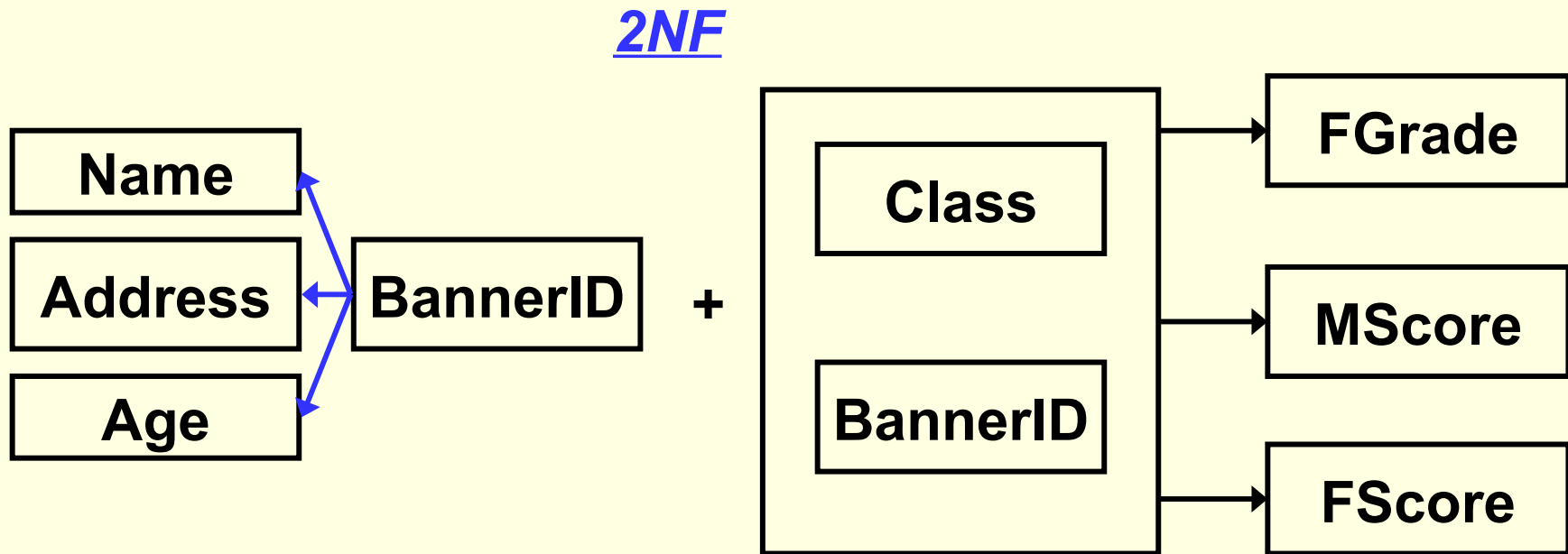
2NF

- 1NF
- All non-key attributes **are fully dependant on** candidate key



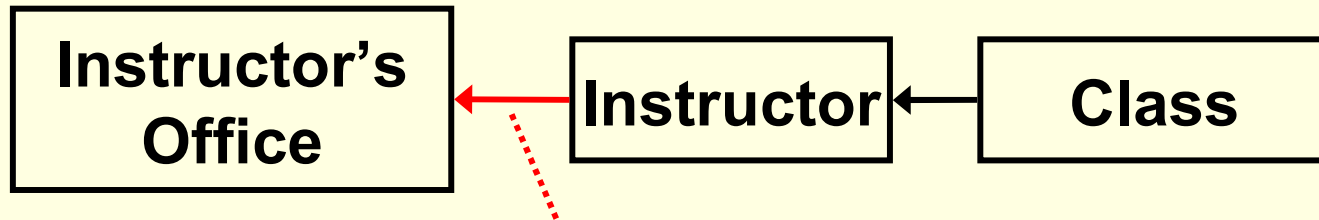
Eliminate these Partial FDs!

Normalization to 2NF



3NF

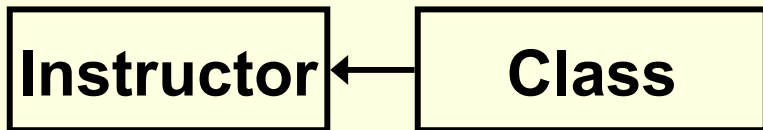
- 2NF
- All non-key attributes are non-transitively dependant on candidate key



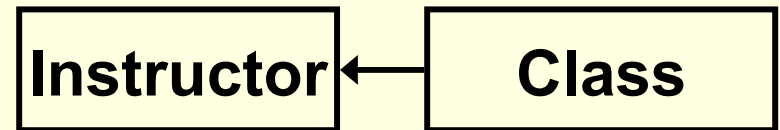
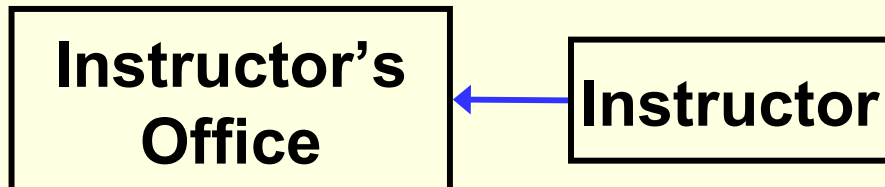
3NF

Eliminate it!

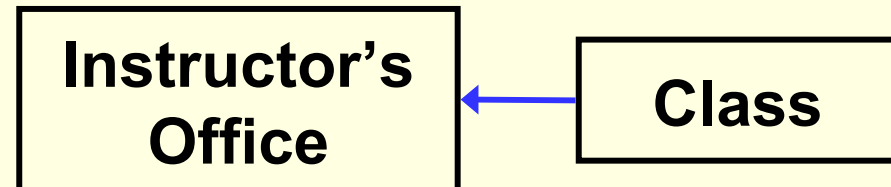
3NF, BUT...?



+



+



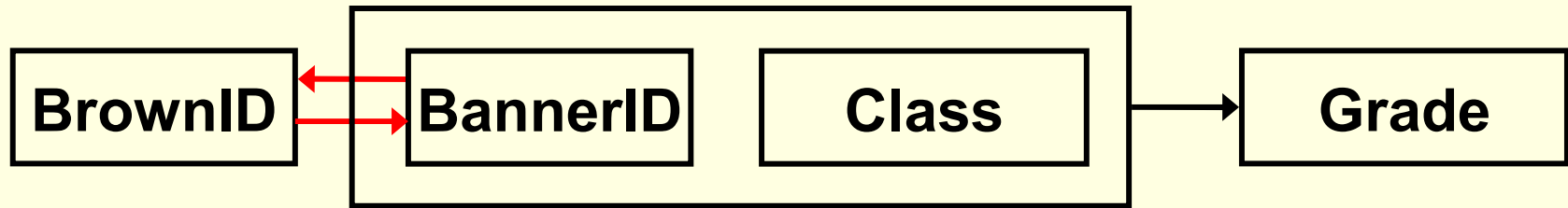
Good/Bad Decomposition

- Can Recover Original Table by JOIN
Lossless Join
- Preserve All FDs
Dependency Preserving
- A good decomposition is **always possible** up to 3NF (sometimes not for BCNF)
(1NF → 2NF → 3NF → BCNF → 4NF → 5NF)

BCNF

- All determinants are candidate key

= Remove **Partial FDs of key attributes**



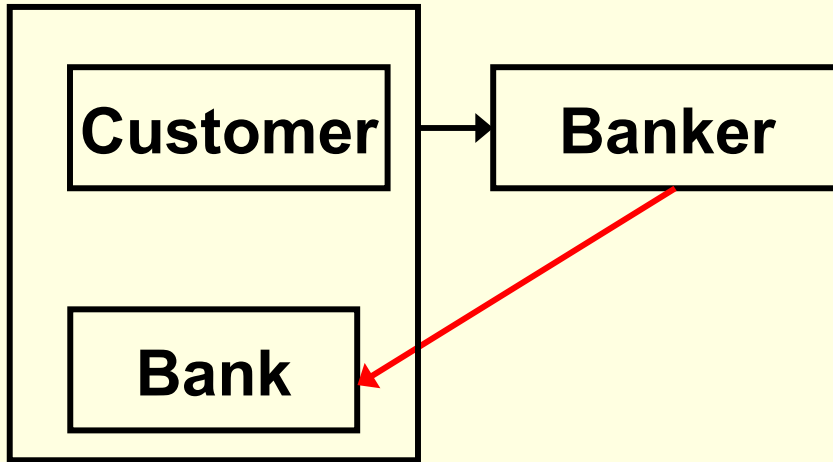
BannerID ↔ BrownID

{BannerID, Class} → {Grade}

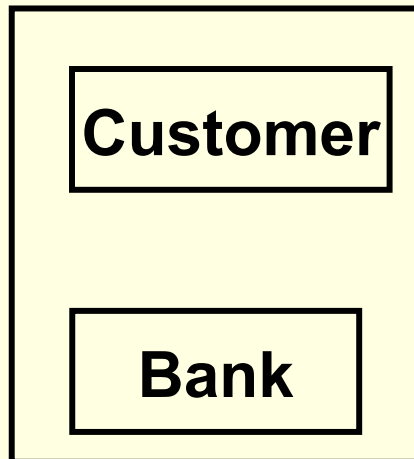
candidate key: {BannerID, Class}, {BrownID, Class}



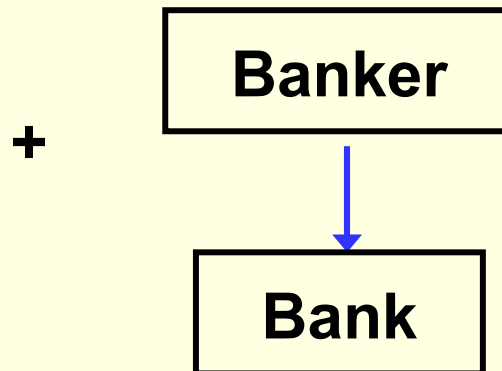
Is BCNF always good?



1. Banker \rightarrow Bank
2. {Customer, Bank} \rightarrow Banker



FD 2 was **LOST**



Denormalization

- *Pre-join: Alleviate Overheads of JOIN*
- *Computed Columns: 3NF → 2NF*
- *Duplicated Table*
 - *Different Clustered Index to speed up different queries*
 - *Not a denormalization, but against Chris Date's "The Principle of Orthogonal Design"*
- etc..