# Functional Dependencies



Imperial College London

# What is wrong with this schema?



| bank_data | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| no | sortcode | bname | cash | type | cname | rate? | mid | amount | tdate |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1000 | 2300.00 | 1999-01-05 |
| 101 | 67 | Strand | 34005.00 | deposit | McBrien, P. | 5.25 | 1001 | 4000.00 | 1999-01-05 |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1002 | -223.45 | 1999-01-08 |
| 107 | | | | | | | | 00.00 | 1999-01-11 |
| 103 | | | | | | | | 45.50 | 1999-01-12 |
| 100 | | | | | | | | 10.23 | 1999-01-15 |
| 107 | | | | | | | | 45.56 | 1999-01-15 |
| 101 | | | | | | | | 30.00 | 1999-01-15 |
| 119 | 56 | Wimbledon | 84340.45 | deposit | Poulov | | | | 1999-01-18 |



```
SELECT cash
FROM   bank_data
WHERE  sortcode=67
```

| cash |
|---|
| 34005.00 |
| 34005.00 |
| 34005.00 |
| 34005.00 |
| 34005.00 |

## What is wrong with this schema?



| bank_data | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| no | sortcode | bname | cash | type | cname | rate? | mid | amount | tdate |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1000 | 2300.00 | 1999-01-05 |
| 101 | 67 | Strand | 34005.00 | deposit | McBrien, P. | 5.25 | 1001 | 4000.00 | 1999-01-05 |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1002 | -223.45 | 1999-01-08 |
| 107 | | | | | | | | 00.00 | 1999-01-11 |
| 103 | | | | | | | | 45.50 | 1999-01-12 |
| 100 | | | | | | | | 10.23 | 1999-01-15 |
| 107 | | | | | | | | 45.56 | 1999-01-15 |
| 101 | | | | | | | | 30.00 | 1999-01-15 |
| 119 | 56 | Wimbledon | 84340.45 | deposit | Poulov | | | | 1999-01-18 |



```
SELECT DISTINCT cash
FROM   bank_data
WHERE  sortcode=67
```

| cash |
|---|
| 34005.0 |

# What is wrong with this schema?



bank_data

| no | sortcode | bname | cash | type | cname | rate? | mid | amount | tdate |
|---|---|---|---|---|---|---|---|---|---|
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1000 | 2300.00 | 1999-01-05 |
| 101 | 67 | Strand | 34005.00 | deposit | McBrien, P. | 5.25 | 1001 | 4000.00 | 1999-01-05 |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1002 | -223.45 | 1999-01-08 |
| 107 | | | | | | | | 00.00 | 1999-01-11 |
| 103 | | | | | | | | 45.50 | 1999-01-12 |
| 100 | | | | | | | | 10.23 | 1999-01-15 |
| 107 | | | | | | | | 45.56 | 1999-01-15 |
| 101 | | | | | | | | 30.00 | 1999-01-15 |
| 119 | 56 | Wimbledon | 84340.45 | deposit | Poulov | | | | 1999-01-18 |



```
SELECT DISTINCT rate
FROM   bank_data
WHERE  account=107
```

| rate |
|---|
| null |

# Problems with Updates on Redundant Data

```sql
INSERT INTO bank_data
VALUES (100,67,'Strand',33005.00,'deposit','McBrien, P.',null,
        1017,-1000.00,'1999-01-21')
```

```sql
UPDATE bank_data
SET    rate=1.00
WHERE  mid=1007
```





| no | sortco[illegible] | | | | | | | [illegible]ount | tdate |
|---|---|---|---|---|---|---|---|---|---|
| 100 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible]00.00 | 1999-01-05 |
| 101 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible]00.00 | 1999-01-05 |
| 100 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible]23.45 | 1999-01-08 |
| 107 | 56 | Wimbledon | 84340.45 | current | Poulova[illegible] | [illegible] | [illegible] | [illegible] | 1999-01-11 |
| 103 | 34 | Goodge St | 6900.67 | current | Boyd, M. | [illegible] | [illegible] | [illegible] | 1999-01-12 |
| 100 | 67 | Strand | 34005.00 | current | McBrie[illegible] | [illegible] | [illegible] | [illegible] | 1999-01-15 |
| 107 | 56 | Wimbledon | 84340.45 | current | Poulova[illegible] | [illegible] | [illegible] | [illegible] | 1999-01-15 |
| 101 | 67 | Strand | 34005.00 | deposit | McBrie[illegible] | [illegible] | [illegible] | [illegible] | 1999-01-15 |
| 119 | 56 | Wimbledon | 84340.45 | deposit | Poulovassilis, A. | 5.50 | 1009 | 5600.00 | 1999-01-18 |
| 100 | 67 | Strand | 33005.00 | deposit | McBrien, P. | null | 1017 | -1000.00 | 1999-01-21 |



```sql
SELECT DISTINCT cash
FROM   bank_data
WHERE  sortcode=67
```

⇒

| cash |
|---|
| 34005.00 |
| 33005.00 |

# Problems with Updates on Redundant Data

```sql
INSERT INTO bank_data
VALUES (100,67,'Strand',33005.00,'deposit','McBrien, P.',null,
        1017,-1000.00,'1999-01-21')
```

```sql
UPDATE bank_data
SET    rate=1.00
WHERE  mid=1007
```





| no | sortco[illegible] | | | | | | | [illegible]ount | tdate |
|---|---|---|---|---|---|---|---|---|---|
| 100 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible]00.00 | 1999-01-05 |
| 101 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible]00.00 | 1999-01-05 |
| 100 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible]23.45 | 1999-01-08 |
| 107 | 56 | Wimbledon | 84340.45 | current | Poulova | | | | 1999-01-11 |
| 103 | 34 | Goodge St | 6900.67 | current | Boyd, M. | | | | 1999-01-12 |
| 100 | 67 | Strand | 34005.00 | current | McBrie | | | | 1999-01-15 |
| 107 | 56 | Wimbledon | 84340.45 | current | Poulova | | | | 1999-01-15 |
| 101 | 67 | Strand | 34005.00 | deposit | McBrie | | | | 1999-01-15 |
| 119 | 56 | Wimbledon | 84340.45 | deposit | Poulovassilis, A. | 5.50 | 1009 | 5600.00 | 1999-01-18 |
| 100 | 67 | Strand | 33005.00 | deposit | McBrien, P. | null | 1017 | -1000.00 | 1999-01-21 |



```sql
SELECT DISTINCT rate
FROM   bank_data
WHERE  account=107
```

⇒

| rate |
|---|
| null |
| 1.00 |

# How do you know what is redundant?

## Functional Dependency



A **functional dependency (fd)** $X \rightarrow Y$ states that if the values of attributes $X$ agree in two tuples, then so must the values in $Y$.

## Using an F



If the FD

but $y$ and



| bank_data | | |
|---|---|---|
| no | mid | rate |
| 101 | 1001 | 5.25 |
| 101 | 1008 | $x$ |
| 119 | 1009 | $y$ |
| $z$ | 1010 | 5.25 |

# Quiz 1: FDs that hold in bank_data



| bank_data | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| no | sortcode | bname | cash | type | cname | rate? | mid | amount | tdate |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1000 | 2300.00 | 1999-01-05 |
| 101 | 67 | Strand | 34005.00 | deposit | McBrien, P. | 5.25 | 1001 | 4000.00 | 1999-01-05 |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1002 | -223.45 | 1999-01-08 |
| 107 | | | | | | | | 00.00 | 1999-01-11 |
| 103 | | | | | | | | 45.50 | 1999-01-12 |
| 100 | | | | | | | | 10.23 | 1999-01-15 |
| 107 | | | | | | | | 45.50 | 1999-01-15 |
| 101 | | | | | | | | 30.00 | 1999-01-15 |
| 119 | 56 | Wimbledon | 84340.45 | deposit | Poulovassilis, A. | 5.50 | 1009 | 5600.00 | 1999-01-18 |



Which set of FDs below do not hold for the data?



**A**

no → rate
no → bname

**B**

no → type
bname → no

**C**

no → type
mid → bname

**D**

amount → rate
bname → sortcode

## Quiz 2: Deriving FDs from other FDs



sortcode → bname
no → sortcode
no → cname
no → rate
mid → no

Given the F



**A**

no → bname

**B**

no,so



**C**

amount,tdate → amount

**D**

amount,tdate → mid

# Armstrong's Axioms

$X, Y$ and $Z$ are sets of attributes, and $XY$ is a shorthand for $X \cup Y$



## Reflexivit

$Y \subseteq X \models$

- Such a



## Applying reflexivity

If amount,tdate are attributes
By reflexivity



$\text{amount} \subseteq \text{amount}, \text{tdate} \models \text{amount}, \text{tdate} \rightarrow \text{amo}$
$\text{tdate} \subseteq \text{amount}, \text{tdate} \models \text{amount}, \text{tdate} \rightarrow \text{tdate}$

# Armstrong's Axioms



$X$, $Y$ and $Z$ are sets of attributes, and $XY$ is a shorthand for $X \cup Y$

### Augment

$X \to Y \models$



### Applying a

If no,cname,sortcode are attributes and no $\to$ cn

By augmentation

no $\to$ cname $\models$ no, sortcode $\to$ cname, sortcode

## Armstrong's Axioms

$X$, $Y$ and $Z$ are sets of attributes, and $XY$ is a shorthand for $X \cup Y$



**Transitiv**

$X \to Y, Y$



**Applying t**

If no $\to$ sortcode and sortcode $\to$ bname
By transitivity
no $\to$ sortcode, sortcode $\to$ bname $\models$ no $\to$ bname

## Union Rule

### Armstrong's Axioms

Reflexivity: $Y \subseteq X \models X \to Y$

Augmentation: $X \to Y \models XZ \to YZ$

Transitivity: $X \to Y, Y \to Z \models X \to Z$



### Union Rul

If $X \to Y,$



By augmen

$X \to Y \models XZ \to YZ$

$X \to Z \models X \to XZ$

By transitivity

$X \to XZ, XZ \to YZ \models X \to YZ$

$YZ \models YZ \to Y, YZ \to Z$

By tr

$X$

$X$



$$\therefore X \to Y, X \to Z \equiv X \to YZ$$

- Note that the union rules means that we can restrict ourselves to FD sets containing just one attribute on the RHS of each FD without loosing expressiveness

## Quiz 3: Deriving FDs from other FDs

Given a set $S = \{A \to BC, CD \to E, C \to F, E \to F\}$ of FDs

**Which set of FDs below follows from $S$?**



**A**

$A \to BF, A$



**B**

$A \to BD, A \to CF, A \to ABCF$



**C**

$A \to BD, A \to BF, A \to ABCF$

**D**

$A \to BD, A \to BF, A \to CF$

# Pseudotransitivity Rule

**Armstrong's Axioms**



Reflexivity: $Y \subseteq X \models X \rightarrow Y$

Augmentation: $X \rightarrow Y \models XZ \rightarrow YZ$

Transitivi

**Pseudotr**



If $X \rightarrow Y$,

By augmentation

$X \rightarrow Y \models WX \rightarrow WY$



By transitivity

$WX \rightarrow WY, WY \rightarrow Z \models WX \rightarrow Z$

$$\therefore X \rightarrow Y, WY \rightarrow Z \models WX \rightarrow Z$$

## Decomposition Rule

### Armstrong's Axioms



Reflexivity: $Y \subseteq X \models X \to Y$

Augmentation: $X \to Y \models XZ \to YZ$

Transitivi

### Decompo



If $X \to Y$,

By reflexivity

$Z \subseteq Y \models Y \to Z$



By transitivity

$X \to Y, Y \to Z \models X \to Z$

$$\therefore X \to Y, Z \subseteq Y \models X \to Z$$

# FDs and Keys

## Super-keys and minimal keys



- If a set of attributes $X$ in relation $R$ functionally determines all the other attributes of $R$, then $X$ must be a **super-key** of $R$
- If it is not possible to remove any attribute from $X$ to form $X'$, and $X'$ functi $R$



## Determin

Suppose branch(sortcode, bname, cash) has the FD set
{sortcode $\rightarrow$ bname, bname $\rightarrow$ sortcode, bname $\rightarrow$



1. {sortcode, bname} is a super-key since {sort
2. However, {sortcode, bname} is not a minimal key, s me, cash} and bname $\rightarrow$ {sortcode, cash}
3. sortcode and bname are both minimal keys of branch

## Quiz 4: Deriving minimal keys from FDs



Suppose the relation $R(A, B, C, D, E)$ has functional dependencies
$S = \{A \rightarrow$

Which of th



| A | B | C | |
|---|---|---|---|
| $A$ | $AB$ | $BC$ | |

## Quiz 5: Keys and FDs



Suppose the

Which FD d



| A | B | C | D |
|---|---|---|---|
| $ABC \rightarrow DE$ | $AC \rightarrow BDE$ | $AB$ | |

## Closure of a set of attributes with a set of FDs

### Closure $X^+$ of a set of attributes $X$ with FDs $S$

1. Set $X^+ := X$
2. Starting with $X^+$ apply each FD $X_s \to Y$ in $S$ where $X_s \subseteq X^+$ but $Y$ is not already in $X^+$, to find determined attributes $Y$
3. $X^+ :$
4. If $Y$
5. Retur



### Closure of attributes

Relation $R(A, B, C, D, E, F)$ has FD set $S = \{A$ … $\to F\}$
To compute $A^+$

- Start with $A^+ = A$, just $A \to BC$ matches, so $Y = BC$
- $A^+ = ABC$, just $C \to F$ matches, so $Y = F$
- $A^+ = ABCF$, no FDs apply, so we have the result

# Closure of a set of attributes with a set of FDs

## Closure $X^+$ of a set of attributes $X$ with FDs $S$

1. Set $X^+ := X$
2. Starting with $X^+$ apply each FD $X_s \to Y$ in $S$ where $X_s \subseteq X^+$ but $Y$ is not already in $X^+$, to find determined attributes $Y$
3. $X^+ :$
4. If $Y$
5. Retur



## Closure of a set of attributes

Relation $R(A, B, C, D, E, F)$ has FD set $S = \{A \to F\}$
To compute $AD^+$

- Start with $AD^+ = AD$, just $A \to BC$ matches, so $Y = BC$
- $AD^+ = ABCD$, $CD \to E, C \to F$ matches, so $Y = EF$
- $AD^+ = ABCDEF$, no FDs apply, so we have the result

## Quiz 6: Closure of Attribute Sets



Given a relation $R(A, B, C, D, E, F)$ and FD set
$S = \{A \rightarrow$

Which clo

| A | B | C | |
|---|---|---|---|
| $A^+$ | $BC^+$ | $BE^+$ | |

# Closure of a set of Functional Dependencies

## Closure of the FD Set



- The closure $S^+$ of a set of FDs $S$ is the set of all FDs that can be inferred from $S$
- Two sets of FDs $S, T$ are equivalent if $S^+ = T^+$
- For speed, we can ignore
  - t
  - L
  - fl $CD$ as $A \to C$
    a
- Apart from calculating equivalence, do not normally n sure





## Equivalent FDs

$S = \{A \to B, A \to C, B \to A, B \to D\}$
$T = \{A \to B, A \to C, A \to D, B \to A\}$
$S^+ = T^+ = \{A \to B, A \to C, A \to D, B \to A, B \to C, B \to D\}$
$\therefore S \equiv T$

# Minimal cover of a set of FDs

## Minimal cover $S_c$ of $S$

A minimal cover $S_c$ of FD set $S$ has the properties that:

- All the FDs in $S$ can be derived from $S_c$ (*i.e.* $S^+ = S_c^+$)
- It is not possible to form a new set $S_c'$ by deleting an FD from $S_c$ or deleting an attribute from an FD in $S_c$, and $S_c'$ can still derive all the FDs in $S$



In general, a s

## Deriving a



Suppose $S = \{A \to B, BC \to A, A \to C, B \to C\}$

1. Since $B \to C$
   $BC \to A \Rightarrow B \to A$
   Leaves $S' = \{A \to B, B \to A, A \to C, B \to$



$2_a$ Since $A \to B, B \to C \models A \to C$
   $A \to C \Rightarrow \emptyset$
   Leaves $S_c = \{A \to B, B \to A, B \to C\}$

$2_b$ Since $B \to A, A \to C \models B \to C$
   $B \to C \Rightarrow \emptyset$
   Leaves $S_c = \{A \to B, B \to A, A \to C\}$

# Quiz 7: Minimal Cover of a Set of FDs

Given an FD set $S = \{A \rightarrow BC, C \rightarrow D, BA \rightarrow E, BD \rightarrow F, EF \rightarrow B, BE \rightarrow ABC\}$

Which is a minimal cover of $S$?



**A**

$A \rightarrow BC, C$



**B**

$A \rightarrow BC, C \rightarrow D, BA \rightarrow E, BD \rightarrow F, EF \rightarrow B, B$



**C**

$A \rightarrow BCE, C \rightarrow D, BD \rightarrow F, EF \rightarrow B, BE \rightarrow$

**D**

$A \rightarrow BC, C \rightarrow D, B \rightarrow E, B \rightarrow F, EF \rightarrow B, BE \rightarrow A$

# Worksheet: Minimal Cover



$R(A, B, C, D, E, F, G, H)$

$S = \{AB \to DEH, BEF \to A, FGH \to C, D \to EG, EG \to BF, F \to BH\}$



1. Rewr on the RHS of
2. Consider each FD $X \to A$, and for each $B \in X$, consider if $X \to B$ from the other FDs. If so, replace $X \to A$ by $(X - B$
3. Consider each FD $X \to A$, and compute $X$ $A \in X^+$, delete $X \to A$ since it is rundundant. This will give a m of $S$.
4. Justify what are the minimal candidate keys of $R$ constrained by $S_c$

# Worksheet: Minimal Cover (Step 3)

1. $AB^+ = ABDEHGFC$
   Try removing $AB \rightarrow D$: find $AB^+ = ABEH$, so can't remove.
   Try removing $AB \rightarrow E$: find $AB^+ = ABDHEGFC$, so remove it from $S''$ to get $S'''$
   Try removing $AB \rightarrow H$: find $AB^+ = ABDEGFHC$, so remove it from $S'''$ to get $S'''' = \{AB \rightarrow D, EF \rightarrow A, FG \rightarrow C, D \rightarrow E, D \rightarrow G, EG \rightarrow B, EG \rightarrow F, F \rightarrow B, F \rightarrow H\}$
2. $EF^+ = EFABHDGC$
   Try re
3. $FG^+$
   Try re
4. $D^+ =$
   Try removing $D \rightarrow E$: find $D^+ = DG$, so can't rem
   Try removing $D \rightarrow G$: find $D^+ = DE$, so can't rem
5. $EG^+ = EGBFHADC$
   Try removing $EG \rightarrow B$: find $EG^+ = EGFBHA$ o get $S'''''$
   Try removing $EG \rightarrow F$: find $EG^+ = EG$, so can't remove.
6. $F^+ = FBH$
   Try removing $F \rightarrow B$: find $F^+ = FH$, so can't remove.
   Try removing $F \rightarrow H$: find $F^+ = FB$, so can't remove.
   Thus $S'''''$ is a minimal cover

$$S_c = \{AB \rightarrow D, EF \rightarrow A, FG \rightarrow C, D \rightarrow EG, EG \rightarrow F, F \rightarrow BH\}$$

# Normalisation



Imperial College London

# Using FDs to Formalise Problems in Schemas

bank_data

| no | sortcode | bname | cash | type | cname | rate? | mid | amount | tdate |
|---|---|---|---|---|---|---|---|---|---|
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1000 | 2300.00 | 1999-01-05 |
| 101 | 67 | Strand | 34005.00 | deposit | McBrien, P. | 5.25 | 1001 | 4000.00 | 1999-01-05 |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1002 | -223.45 | 1999-01-08 |
| 107 | 56 | Wimbledon | 84340.45 | current | Poulovassilis, A. | null | 1004 | -100.00 | 1999-01-11 |
| 103 | 34 | Goodge St | 6900.67 | current | Boyd, M. | null | 1005 | 145.50 | 1999-01-12 |
| 100 | | | | | | | | 10.23 | 1999-01-15 |
| 107 | | | | | | | | 45.56 | 1999-01-15 |
| 101 | | | | | | | | 30.00 | 1999-01-15 |
| 119 | | | | | | | | [illegible] | 1999-01-18 |



Formalise the intuition of redundancy by the statements of F

$\mathsf{mid} \rightarrow \{\mathsf{tdate}, \mathsf{amount}, \mathsf{no}\}$,
$\mathsf{no} \rightarrow \{\mathsf{type}, \mathsf{cname}, \mathsf{rate}, \mathsf{sortcode}\}$,
$\{\mathsf{cname}, \mathsf{type}\} \rightarrow \mathsf{no}$,
$\mathsf{sortcode} \rightarrow \{\mathsf{bname}, \mathsf{cash}\}$
$\mathsf{bname} \rightarrow \mathsf{sortcode}$

## 1st Normal Form (1NF)

Every attribute depends on the key

# Quiz 8: 1st Normal Form

| bank_data | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| no | sortcode | bname | cash | type | cname | rate? | mid | amount | tdate |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1000 | 2300.00 | 1999-01-05 |
| 101 | 67 | Strand | 34005.00 | deposit | McBrien, P. | 5.25 | 1001 | 4000.00 | 1999-01-05 |
| 100 | 67 | Strand | 34005.00 | current | McBrien, P. | null | 1002 | -223.45 | 1999-01-08 |
| 107 | 56 | Wimbledon | 84340.45 | current | Poulovassilis, A. | null | 1004 | -100.00 | 1999-01-11 |
| 103 | 34 | Goodge St | 6900.67 | current | Boyd, M. | null | 1005 | 145.50 | 1999-01-12 |
| 100 | | | | | | | | 10.23 | 1999-01-15 |
| 107 | | | | | | | | 45.56 | 1999-01-15 |
| 101 | | | | | | | | 30.00 | 1999-01-15 |
| 119 | | | | | | | | [illegible] | 1999-01-18 |



mid $\rightarrow$ {tdate, amount, no},
no $\rightarrow$ {type, cname, rate, sortcode},
{cname, type} $\rightarrow$ no,
sortcode $\rightarrow$ {bname, cash}
bname $\rightarrow$ sortcode

Is bank_data in 1st Normal form?

True

False

# Prime and Non-Prime Attributes

### Prime Attribute



An attribute $A$ of relation $R$ is **prime** if there is some minimal candidate key $X$ of $R$ such that $A \subseteq X$

Any other attribute $B \in Attrs(R)$ is **non-prime**

### Prime and n



bank_data(

Has FDs mid $\rightarrow$ {tdate, amount, no}, no $\rightarrow$ {type, cname, rate, sortcode}, {cname, type} $\rightarrow$ no, sortcode $\rightarrow$ {bname, cash}, b

Then



1. the only minimal candidate key is mid
2. the only prime attribute is mid
3. non-prime attributes are no,sortcode,bname,cash,type,cname,rate,amount,tdate

# 3rd Normal Form (3NF)

## 3rd Normal Form (3NF)



For every non-trivial FD $X \to A$ on $R$, either

1. $X$ is a super-key
2. $A$ is pr

*Every non-* *ut the key*



## Failure of

bank_data(no,sortcode,bname,cash,type,cname,rat



- Has the following FDs where the LHS is not a super-key:
  no $\to$ {type, cname, rate, sortcode}, {cname, t
  sortcode $\to$ {bname, cash}, bname $\to$ sortcode
- Each of the above FD causes the relation not to meet 3NF since the RHS contains non-prime attributes

## Quiz 9: Prime and nonprime attributes

Given a relation $R(A, B, C, D, E, F)$ and an FD set
$A \to BCE, C \to D, BD \to F, EF \to B, BE \to A$



What are the nonprime attributes?

**A**
$DEF$



**B**
$BC$



**C**
$CDF$

**D**
$CD$

## Quiz 10: 3rd Normal Form

Given a relation $R(A, B, C, D, E, F)$ and an FD set
$A \rightarrow BCE, C \rightarrow D, BD \rightarrow F, EF \rightarrow B, BE \rightarrow A$



**Which decomposition is not in 3NF?**

**A**

$R_1(B, D, F$



**B**

$R_1(A, B, C, E, F), R_2(C, D)$



**C**

$R_1(A, B, C, E, F), R_2(C, D), R_3(B, D, F)$

**D**

$R_1(B, E, F), R_2(A, C, E), R_3(C, D)$

# Lossless-join decomposition of relations

## Lossless-join decomposition of a Relation

A **lossless-join** decomposition of a relation $R$ with respect to FDs $S$ into relations $R_1, \ldots, R_n$ has the properties that:

- $Attrs(R_1) \cup \ldots \cup Attrs(R_n) = Attrs(R)$
- For all p $_{rs(R_n)} R = R$



## Lossless-j

bank_data(no,sortcode,bname,cash,type,cname,rate,mid,amount,tdate)

- Has FDs mid $\rightarrow$ {tdate, amount, no}, no $\rightarrow$ {
  {cname, type} $\rightarrow$ no, sortcode $\rightarrow$ {bname, cas
- Decomposing bank_data into
  branch $= \pi_{\text{sortcode,bname,cash}}$ bank_data
  account $= \pi_{\text{no,type,cname,rate,sortcode}}$ bank_data
  movement $= \pi_{\text{mid,amount,no,tdate}}$ bank_data
  satisfies the lossless-join decomposition property

# Problems if not a lossless-join decomposition



If a decomposition of $R$ into $R_1, \ldots, R_n$ is not lossless, then some tuples spread over $R_1, \ldots, R_n$ can result in phantom tuples appearing

$R(A, B, C,$



$R$

| $A$ | $B$ | $C$ | $D$ |
|---|---|---|---|
| 1 | 1 | 2 | 6 |
| 2 | 2 | 3 | 4 |
| 3 | 3 | 3 | 5 |

| $A$ | $B$ | $C$ |
|---|---|---|
| 1 | 1 | 2 |
| 2 | 2 | 3 |
| 3 | 3 | 3 |

| $C$ | $D$ |
|---|---|
| 2 | 6 |
| 3 | 4 |
| 3 | 5 |



| $A$ | $B$ | $C$ | $D$ |
|---|---|---|---|
| 3 | 3 | 3 | 4 |

# Quiz 11: Lossless join decomposition

Given a relation $R(A, B, C, D, E, F)$ and an FD set
$A \to BCE, C \to D, BD \to F, EF \to B, BE \to A$



Which is not a lossless-join decomposition of $R$?

**A**

$R_1(B, D, F$



**B**

$R_1(A, B, C, E, F), R_2(C, D)$



**C**

$R_1(A, B, C, E, F), R_2(C, D), R_3(B, D, F)$

**D**

$R_1(B, E, F), R_2(A, C, E), R_3(C, D)$

## Worksheet: Lossless Join Decomposition



1. $R(A, B, C, D, E)$ has the FDs $S = \{AB \rightarrow C, C \rightarrow DE, E \rightarrow A\}$.
   Whic
   1.
   2.



2. Deriv $, D, E, F)$
   with FDs $S = \{AB \rightarrow CD, C \rightarrow E, A \rightarrow F\}$
3. Derive a lossless join decomposition into three relation )
   with FDs $S = \{AB \rightarrow CD, C \rightarrow E, F \rightarrow A\}$

# Generating 3NF

## Generating 3NF

1. Given $R$ and a set of FDs $S$, find an FD $X \rightarrow A$ that causes $R$ to violate 3NF (*i.e.* for which $A$ is not a prime attribute and $X$ is not a superkey).
2. Decompose $R$ into $R_a(Attr(R) - A)$ and $R_b(XA)$ (Note because the two relati
3. Proje



Note that ste $R_a$ with $R_b$ will share $X$, and $X \rightarrow A$

## Canonical Example of 3NF Decomposition

Suppose $R(A, B, C)$ has FD set $S = \{A \rightarrow B, B$

- The only key is $A$, and so $B \rightarrow C$ violates 3NF (since $B$ is not a superkey and $C$ is nonprime).
- Decomposing $R$ into $R_1(A, B)$ and $R_2(B, C)$ results in two 3NF relations.

# Example: Decomposing bank_data into 3NF

## Bank Database as a Single Relation

bank_data(no,sortcode,bname,cash,type,cname,rate,mid,amount,tdate)
$S = \{$mid $\rightarrow \{$tdate, amount, no$\}$, no $\rightarrow \{$type, cname, rate, sortcode$\}$,
$\{$cname, type$\} \rightarrow$ no, sortcode $\rightarrow \{$bname, cash$\}$, bname $\rightarrow$ sortcode$\}$



Since sort e, cash
nonprime,

1. branc
   bnam



2. bank_data′(no, sortcode, type, cname, rate, mid
   mid $\rightarrow \{$tdate, amount, no$\}$, no $\rightarrow \{$type, cna
   $\{$cname, type$\} \rightarrow$ no



branch is in 3NF, but no $\rightarrow \{$type, cname, rate, so late
3NF, so we should decompose bank_data′ into:

3. account(no, type, cname, rate, sortcode) with FDs
   no $\rightarrow \{$type, cname, rate, sortcode$\}$, $\{$cname, type$\} \rightarrow$ no
4. movement(mid.amount, no, tdate) with FD mid $\rightarrow \{$tdate, amount, no$\}$

The relations branch, account, and movement are all in 3NF

# Preserving FDs during decomposition

## FD preserving decomposition

A lossless decomposition of $R$ with FDs $S$ into $R_a$ and $R_b$ preserves functional dependencies $S$ if the projection of $S^+$ onto $R_a$ and $R_b$ is equivalent to $S$



## FD preserving decomposition

Suppose $R(ABC)$ with $S = \{A \to B, B \to C, C \to A\}$ is decomposed into $R_a(AB)$ and $R_b(B$

- $S^+ =$

- The pr
- The projection of $S^+$ onto $R_b$ gives $S_b^+ = \{$
- Note that the union $S_u$ of the two subsets of he property that $S_u^+ = S^+$, and hence the decomposi dependencies.



## 3NF

There is always possible to decompose a relation into 3NF in a manner that preserves functional dependencies. Thus any *good* 3NF decomposition of a relation must also preserve functional dependencies.

# Quiz 12: Preserving FDs during Decomposition

Given a relation $R(A, B, C, D, E, F)$ and an FD set
$A \to BCE, C \to D, BD \to F, EF \to B, BE \to A$



Which decomposition preserves FDs?

A

$R_1(B, D, F$



B

$R_1(A, B, C, E, F), R_2(C, D)$



C

$R_1(A, B, C, E, F), R_2(C, D), R_3(B, D, F)$

D

$R_1(B, E, F), R_2(A, C, E), R_3(C, D)$

# Preserving FDs, lossless join, and 3NF

Given a relation $R(A, B, C, D, E, F)$ and an FD set
$A \to BCE, C \to D, BD \to F, EF \to B, BE \to A$



| Decom | eserves FDs |
|---|---|
| $R_1(B, D$ | |
| $R_1(A, B$ | |
| $R_1(A, B$ | |
| $R_1(B, E, F), R_2(A, C, E), R_3(C, D)$ | ✗ |





**Decomposing to 3NF**

Since it is always possible to decompose a relation into a 3NF form that is both a lossless join decomposition, and preserves FDs, you should always do so.

# Quiz 13: Preserving FDs during Decomposition to 3NF



Suppose the relation $R(A, B, C, D, E)$ has functional dependencies $S = \{AC \to DBE, BC \to DE, B \to A, E \to D\}$ (and hence has minimal keys $AC$ and $BC$)

Which is a los



**A**

$R_a(B, C, E), R_b(A, B, C), R_c(D, E)$

**B**

$R_a($



**C**

$R_a(A, C, D), R_b(A, C, E), R_c(A, B)$

**D**

$R_a(A, C, E), R_b(B, D, E)$

# Boyce-Codd Normal Form (BCNF)

## Boyce-Codd Normal Form (BCNF)

For every non-trivial FD $X \to A$ on $R$, $X$ is a super-key.
*Every attribute depends on the key, the whole key and nothing but the key*



## BCNF sch



branch(sor ... e $\to$ sortcode
is in BCNF since sortcode and bname are both candidate keys

account(no, type, cname, rate, sortcode) with FDs ... code},
{cname, type} $\to$ no is in BCNF since no and cn



movement(mid.amount, no, tdate) with FD mid $\to$ {tdate, amount, no} is in BCNF since mid is key

# Decomposition of Relations into BCNF

## Generating BCNF

1. Given $R$ and a set of FDs $S$, find an FD $X \to A$ that causes $R$ to violate BCNF (*i.e.* for which $X$ is not a superkey).
2. Decompose $R$ into $R_a(Attr(R) - A)$ and $R_b(X A)$ (Note because the two relations share $X$ and $X \to A$ this is lossless)
3. Proje



## Differenc



Suppose the relation address(no, street, town, county, postcode) has FDs $\{$no, street, town, county$\} \to$ postcode, postcode

- The relation is in 3NF (alternative keys no, ostcode).
- The relation is not in BCNF since postcode non-superkey as the determinant
  - Decompose the relation address on postcode $\to \{$street, town, county$\}$ to:
    postcode(postcode, street, town, county)
    streetnumber(no, postcode)
  - Note FD $\{$no, street, town, county$\} \to$ postcode cannot be projected over the relations.

## Worksheet: Normal Forms



$S_c = \{AB \to D, EF \to A, FG \to C, D \to EG, EG \to F, F \to BH\}$

1. Deco
2. Deco



3. Determine if your decompositions in (1) and (2) preserv not, suggest how to amend you schema to preserve FDs.