Functional Dependencies

P.J. McBrien

Imperial College London

What is wrong with this schema?

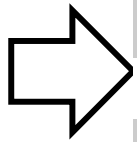
bank_data									
no	sortcode	bname	cash	type	cname	rate?	mid	amount	tdate
100	67	Strand	34005.00	current	McBrien, P.	null	1000	2300.00	1999-01-05
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1001	4000.00	1999-01-05
100	67	Strand	34005.00	current	McBrien, P.	null	1002	-223.45	1999-01-08
107	56	Wimbledon	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	67	Strand	34005.00	current	McBrien, P.	null	1006	10.23	1999-01-15
107	56	Wimbledon	84340.45	current	Poulovassilis, A.	null	1007	345.56	1999-01-15
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1008	1230.00	1999-01-15
119	56	Wimbledon	84340.45	deposit	Poulovassilis, A.	5.50	1009	5600.00	1999-01-18

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	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	67	Strand	34005.00	current	McBrien, P.	null	1006	10.23	1999-01-15
107	56	Wimbledon	84340.45	current	Poulovassilis, A.	null	1007	345.56	1999-01-15
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1008	1230.00	1999-01-15
119	56	Wimbledon	84340.45	deposit	Poulovassilis, A.	5.50	1009	5600.00	1999-01-18

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



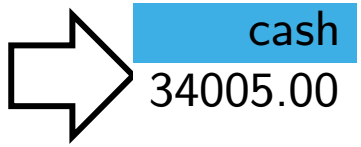
cash
34005.00
34005.00
34005.00
34005.00
34005.00

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	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	67	Strand	34005.00	current	McBrien, P.	null	1006	10.23	1999-01-15
107	56	Wimbledon	84340.45	current	Poulovassilis, A.	null	1007	345.56	1999-01-15
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1008	1230.00	1999-01-15
119	56	Wimbledon	84340.45	deposit	Poulovassilis, A.	5.50	1009	5600.00	1999-01-18

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

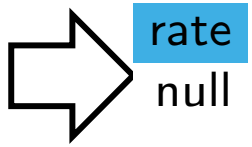


cash
34005.00

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	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	67	Strand	34005.00	current	McBrien, P.	null	1006	10.23	1999-01-15
107	56	Wimbledon	84340.45	current	Poulovassilis, A.	null	1007	345.56	1999-01-15
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1008	1230.00	1999-01-15
119	56	Wimbledon	84340.45	deposit	Poulovassilis, A.	5.50	1009	5600.00	1999-01-18

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



Problems with Updates on Redundant Data

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

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

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	67	Strand	34005.00	current	McBrien, P.	null	1006	10.23	1999-01-15
107	56	Wimbledon	84340.45	current	Poulovassilis, A.	1.00	1007	345.56	1999-01-15
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1008	1230.00	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

Problems with Updates on Redundant Data

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

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

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	67	Strand	34005.00	current	McBrien, P.	null	1006	10.23	1999-01-15
107	56	Wimbledon	84340.45	current	Poulovassilis, A.	1.00	1007	345.56	1999-01-15
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1008	1230.00	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

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

cash
34005.00
33005.00

Problems with Updates on Redundant Data

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

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

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	67	Strand	34005.00	current	McBrien, P.	null	1006	10.23	1999-01-15
107	56	Wimbledon	84340.45	current	Poulovassilis, A.	1.00	1007	345.56	1999-01-15
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1008	1230.00	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

```
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 FD to find a value

If the FD no $\rightarrow$ rate holds then x in the table below must always take the value 5.25, but y may take any value.

bank_data		
no	mid	rate
101	1001	5.25
101	1008	x
119	1009	y

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	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	67	Strand	34005.00	current	McBrien, P.	null	1006	10.23	1999-01-15
107	56	Wimbledon	84340.45	current	Poulovassilis, A.	null	1007	345.56	1999-01-15
101	67	Strand	34005.00	deposit	McBrien, P.	5.25	1008	1230.00	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 $\longrightarrow$ rate
no $\longrightarrow$ bname

B

no $\longrightarrow$ type
bname $\longrightarrow$ no

C

no $\longrightarrow$ type
mid $\longrightarrow$ bname

D

amount $\longrightarrow$ rate
bname $\longrightarrow$ sortcode

Quiz 2: Deriving FDs from other FDs

sortcode $\rightarrow$ bname
no $\rightarrow$ sortcode
no $\rightarrow$ cname
no $\rightarrow$ rate
mid $\rightarrow$ no

Given the FDs above, which FD below might not hold?

A

no $\rightarrow$ bname

B

no,sortcode $\rightarrow$ cname,sortcode

C

amount,tdate $\rightarrow$ amount

D

amount,tdate $\rightarrow$ mid

Armstrong's Axioms

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

Reflexivity

$Y \subseteq X \models X \rightarrow Y$

- Such an FD is called a **trivial** FD

Applying reflexivity

If amount,tdate are attributes
By reflexivity
$\mathsf{amount} \subseteq \mathsf{amount}, \mathsf{tdate} \models \mathsf{amount}, \mathsf{tdate} \rightarrow \mathsf{amount}$
$\mathsf{tdate} \subseteq \mathsf{amount}, \mathsf{tdate} \models \mathsf{amount}, \mathsf{tdate} \rightarrow \mathsf{tdate}$

Armstrong's Axioms

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

Augmentation

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

Applying augmentation

If no,cname,sortcode are attributes and no $\rightarrow$ cname
By augumentation
no $\rightarrow$ cname $\models$ no, sortcode $\rightarrow$ cname, sortcode

Armstrong's Axioms

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

Transitivity

$X \rightarrow Y, Y \rightarrow Z \models X \rightarrow Z$

Applying transitivity

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

Union Rule

Armstrong's Axioms

Reflexivity: $Y \subseteq X \models X \rightarrow Y$
Augmentation: $X \rightarrow Y \models XZ \rightarrow YZ$
Transitivity: $X \rightarrow Y, Y \rightarrow Z \models X \rightarrow Z$

Union Rule

If $X \rightarrow Y, X \rightarrow Z$
By augmentation
$X \rightarrow Y \models XZ \rightarrow YZ$
$X \rightarrow Z \models X \rightarrow XZ$
By transitivity
$X \rightarrow XZ, XZ \rightarrow YZ \models X \rightarrow YZ$

If $X \rightarrow YZ$
By reflexivity
$YZ \models YZ \rightarrow Y, YZ \rightarrow Z$
By transitivity
$X \rightarrow YZ, YZ \rightarrow Y \models X \rightarrow Y$
$X \rightarrow YZ, YZ \rightarrow Z \models X \rightarrow Z$

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

- 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 $S = \{A \rightarrow BC, CD \rightarrow E, C \rightarrow F, E \rightarrow F\}$ of FDs

Which set of FDs below follows from S?

A

$A \rightarrow BF, A \rightarrow CF, A \rightarrow ABCF$

B

$A \rightarrow BD, A \rightarrow CF, A \rightarrow ABCF$

C

$A \rightarrow BD, A \rightarrow BF, A \rightarrow ABCF$

D

$A \rightarrow BD, A \rightarrow BF, A \rightarrow CF$

Pseudotransitivity Rule

Armstrong's Axioms

Reflexivity: $Y \subseteq X \models X \rightarrow Y$
Augmentation: $X \rightarrow Y \models XZ \rightarrow YZ$
Transitivity: $X \rightarrow Y, Y \rightarrow Z \models X \rightarrow Z$

Pseudotransitivity Rule

If $X \rightarrow Y, WY \rightarrow Z$
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: $Y \subseteq X \models X \rightarrow Y$
Augmentation: $X \rightarrow Y \models XZ \rightarrow YZ$
Transitivity: $X \rightarrow Y, Y \rightarrow Z \models X \rightarrow Z$

Decomposition Rule

If $X \rightarrow Y, Z \subseteq Y$
By reflexivity
$Z \subseteq Y \models Y \rightarrow Z$
By transitivity
$X \rightarrow Y, Y \rightarrow Z \models X \rightarrow Z$

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

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 possibile to remove any attribute from X to form X', and X' functionally determine all attributes, then X is a **minimal key** of R

Determining keys of a relation

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

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

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 in S where the RHS is not already in X^+, to find determined attributes Y
3. If Y not empty, $X^+ := X^+ \cup Y$, goto (2)
4. Return X^+

Closure of attributes

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

- Start with $A^+ = A$, just $A \rightarrow BC$ matches, so $Y = BC$
- $A^+ = ABC$, just $C \rightarrow F$ 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 in S where the RHS is not already in X^+, to find determined attributes Y
3. If Y not empty, $X^+ := X^+ \cup Y$, goto (2)
4. Return X^+

Closure of a set of attributes

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

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

Quiz 4: Closure of Attribute Sets

Given a relation $R(A, B, C, D, E, F)$ and FD set
$S = \{A \rightarrow BC, C \rightarrow D, BA \rightarrow E, BD \rightarrow F, EF \rightarrow B, BE \rightarrow ABC\}$

Which closure of attributes of S does not cover R?

A	B	C	D
A^+	BC^+	BE^+	EF^+

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 infered from S
- Two sets of FDs S, T are equivalent if $S^+ = T^+$
- For speed, we can ignore trivial FDs, and flatten all FDs to have just one attribute in RHS
- Apart from calculating equivalence, do not normally need to compute closure

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 set of FDs may have more than one minimal cover

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 set of FDs may have more than one minimal cover

Deriving a minimal cover

Suppose $S = \{A \rightarrow B, BC \rightarrow A, A \rightarrow C, B \rightarrow C\}$
Since $B \rightarrow C$
$BC \rightarrow A \Rightarrow B \rightarrow A$
Leaves $S' = \{A \rightarrow B, B \rightarrow A, A \rightarrow C, B \rightarrow C\}$

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 set of FDs may have more than one minimal cover

Deriving a minimal cover

Suppose $S = \{A \rightarrow B, BC \rightarrow A, A \rightarrow C, B \rightarrow C\}$
Since $B \rightarrow C$
$BC \rightarrow A \Rightarrow B \rightarrow A$
Leaves $S' = \{A \rightarrow B, B \rightarrow A, A \rightarrow C, B \rightarrow C\}$
Since $A \rightarrow B, B \rightarrow C \models A \rightarrow C$
$A \rightarrow C \Rightarrow \emptyset$
Leaves $S_c = \{A \rightarrow B, B \rightarrow A, B \rightarrow C\}$

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 set of FDs may have more than one minimal cover

Deriving a minimal cover

Suppose $S = \{A \rightarrow B, BC \rightarrow A, A \rightarrow C, B \rightarrow C\}$
Since $B \rightarrow C$
$BC \rightarrow A \Rightarrow B \rightarrow A$
Leaves $S' = \{A \rightarrow B, B \rightarrow A, A \rightarrow C, B \rightarrow C\}$
Since $B \rightarrow A, A \rightarrow C \models B \rightarrow C$
$B \rightarrow C \Rightarrow \emptyset$
Leaves $S_c = \{A \rightarrow B, B \rightarrow A, A \rightarrow C\}$

Quiz 5: 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

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

B

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

C

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

D

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