Introduction to Database Systems

Normalization

Lee Mong Li

National University of Singapore

Recap: Functional Dependencies

Definition: X → Y

- If two tuples agree on the values of the attributes in X, then they agree on the values of the attributes in Y.
- **Example:** {position} → {salary}

If two employees have the same position, they also have the same salary.

Types of FDs:

- Trivial $\mathbf{Y \subset X}$
- Non-trivial $\mathbf{Y \not\subset X}$
- Completely non-trivial $\mathbf{Y \cap X = \varnothing}$
- **Example:** Given a relation R(A,B,C,D,E) with

$$F = \{AB \rightarrow CE, CD \rightarrow D, ACE \rightarrow BC, D \rightarrow E\}$$

How many FDs of each type are there?

Armstrong's Axioms

Let X, Y, Z be subsets of the scheme of relation R

- **Reflexivity:** If $Y \subset X$, then $X \rightarrow Y$
- **Augmentation:** If $X \rightarrow Y$, then $X \cup Z \rightarrow Y \cup Z$
- **Transitivity:** If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$

Example: Given a relation R(A,B,C,D,E) with

F = { AB → ABCE, C → BC }

Show that ACD is a **superkey** using the axioms.

In other words, show that ACD → ABCDE

Attribute Closure

For a set X of attributes, we call the **closure of X (*w.r.t. a set of FDs*)**, denoted by X^+, the maximum set of attributes such that $X \rightarrow X^+$.

Example: Given a relation R(A, B, C, D, E) with

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

Is A in A+? Is B in A+? Is C in A+? Is D in A+? Is E in A+?

Algorithm:

- Start with the given set X.
- Find and add new attributes which can be determined by any subset of X, until no more attributes can be added.

Useful for (a) finding ***candidate keys*** *and (b) proving* ***equivalence of sets of FDs***

Candidate Keys

Consider the relation R(A,B,C,D,E) with the FDs:

$AB \rightarrow CDE$

$AC \rightarrow BDE$

$B \rightarrow C$

$C \rightarrow B$

$C \rightarrow D$

$B \rightarrow E$

Find all **candidate keys**

Equivalence of 2 Sets of FDs

Algorithm based on attribute closure

Intuition:

- Given two sets of FDs F and G
- If we have in F, $A \rightarrow B$ and $A^+ = ABC$, then we know that $A \rightarrow A$, $A \rightarrow B$ and $A \rightarrow C$ is in F^+
- And if we have in G, $A^+ = AB$, then we know that there is some FD (i.e., $A \rightarrow C$) in F^+ but not in G^+
- Hence, F and G are not equivalent.

Minimal Cover

A set of FDs F is a minimal cover for a set of FDs G if and only if

- Every FD in F is of the form X $\rightarrow$ A where X is a set of attributes, A is a **single attribute** and X has **no redundant attributes**
- There are **no redundant FDs** in F
- F is equivalent to G, that is, **$F^+ = G^+$**

Minimal Cover

Consider the relation R(A,B,C,D,E) with FDs:

$AB \rightarrow CDE$

$AC \rightarrow BDE$

$B \rightarrow C$

$C \rightarrow B$

$C \rightarrow D$

$B \rightarrow E$

Find a **minimal cover** for the FDs

Normalization

Process of Decomposition

- Break down a relation into smaller relations such that all attributes are still present in at least one of the smaller relations.

Lossless / Lossy

- Whether the original relation can be recovered by combining the smaller relations.

Dependency preserving

- Whether all the dependencies still exist in the smaller relations.

Decomposition

Given a relation R(A,B,C) with {A} → {B,C}, which of the following are decompositions?

R1(A), R2(B)	No
R1(A,C), R2(B)	Yes
R1(A,B), R2(B,C)	Yes
R1(A,B), R2(A,C)	Yes
R1(A,C), R2(B,C)	Yes

Decomposition

Given a relation R(A,B,C) with {A} → {B,C}, which are lossless decompositions?

R1(A), R2(B)	**No**
R1(A,C), R2(B)	**No**
R1(A,B), R2(B,C)	**No**
R1(A,B), R2(A,C)	
R1(A,C), R2(B,C)	

A	B	C
1	Bill	Gates
2	Bill	Clinton

A	B
1	Bill
2	Bill

B	C
Bill	Gates
Bill	Clinton

A	B	C
1	Bill	Gates
1	Bill	Clinton
2	Bill	Gates
2	Bill	Clinton

Decomposition

Given a relation R(A,B,C) with {A} → {B,C}, which are lossless decompositions?

R1(A), R2(B)	**No**
R1(A,C), R2(B)	**No**
R1(A,B), R2(B,C)	**No**
R1(A,B), R2(A,C)	**Yes**
R1(A,C), R2(B,C)	

A	B	C
1	Bill	Gates
2	Bill	Clinton

A	B
1	Bill
2	Bill

A	C
1	Gates
2	Clinton

A	B	C
1	Bill	Gates
2	Bill	Clinton

Decomposition

Given a relation R(A,B,C) with {A} → {B,C}, which are lossless decompositions?

R1(A), R2(B)	No
R1(A,C), R2(B)	No
R1(A,B), R2(B,C)	No
R1(A,B), R2(A,C)	Yes
R1(A,C), R2(B,C)	No

A	B	C
1	Steve	Jobs
2	New	Jobs

A	C
1	Jobs
2	Jobs

B	C
Steve	Jobs
New	Jobs

A	B	C
1	Steve	Jobs
1	New	Jobs
2	Steve	Jobs
2	New	Jobs

Decomposition

Given a relation R(A,B,C) with {A} → {B,C}, which are dependency preserving decompositions?

R1(A), R2(B)	**No**		
R1(A,C), R2(B)	**No**		
R1(A,B), R2(B,C)	**No**		
R1(A,B), R2(A,C)	**Yes**	R1(A,B) R1(A,C)	**A → B** **A → C**
R1(A,C), R2(B,C)	**No**		

Normal Forms

Boyce-Codd Normal Form (BCNF)

- For any FD X → Y on a relation R, either the FD is trivial (i.e., $Y \subset X$) or X is a superkey of R

Third Normal Form (3NF)

- For any FD X → Y on a relation R, either the FD is trivial (i.e., $Y \subset X$), or X is a superkey, or Y is part of some candidate key

Example:

- Given a relation R(A, B, C, D, E, G) with FDs

  $$F = \{ AB \rightarrow CDEG, \quad E \rightarrow G \}$$

- Candidate key: AB
- $E \rightarrow G$ violates both BCNF and 3NF

Decomposition into BCNF

Given a set of relations and a set of FDs

1. **Find FD $X \rightarrow Y$ on a relation R that violates BCNF property**
2. **Decompose R into $(R - X^+ + X)$ and X^+ (and project the FDs)**
3. **Repeat Steps 1 & 2 until all relations are in BCNF**

Note: It is sometimes useful to find the minimal cover first.

Example: R(A, B, C, D, E, G) with F = { $AB \rightarrow CDEG$, $E \rightarrow G$ }

- **$E \rightarrow G$ violates BCNF**
- **BCNF decomposition into:**

  R1(E, G) with F1 = { $E \rightarrow G$ }

  R2(A, B, C, D, E) with F2 = { $AB \rightarrow CDE$ }

Decomposition into 3NF

Given a set of relations and a set of FDs

1. **Find the minimal cover G.**
2. **For every FD X → Y in G,**
   If $X \cup Y$ is not in the existing relations, then add a relation R with $X \cup Y$ (and project the FDs)
3. **If there is no relation containing the key, add a relation with the key.**

Note 1: The given relations are ignored.
Note 2: The extended minimal cover can be used, too.

Example: R(A, B, C, D, E, G) with F = { AB → CDEG, E → G }

- **E → G violates 3NF**
- **G = { AB → C, AB → D, AB →E, E → G}**
- **Extended minimal cover G' = { AB → CDE, E → G }**
- **3NF synthesis: { R1(<u>A,B</u>,C,D,E), R2(<u>E</u>,G) }**

BCNF Decomposition

Consider the relation R(A,B,C,D,E) with FDs:

$AB \rightarrow CDE$

$AC \rightarrow BDE$

$B \rightarrow C$

$C \rightarrow B$

$C \rightarrow D$

$B \rightarrow E$

Give a BCNF decomposition of R.

3NF Decomposition

Consider the relation R(A,B,C,D,E) with FDs:

$AB \rightarrow CDE$

$AC \rightarrow BDE$

$B \rightarrow C$

$C \rightarrow B$

$C \rightarrow D$

$B \rightarrow E$

Give a 3NF decomposition of R.