

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

Normal Form Review

Key

FDs F on schema R , $K \subseteq R$

1. $K^+ = R$ $K \rightarrow R$
2. No proper subset of K has this property

Can be multiple keys

An attribute A is prime if it is in some key K

Superkey: Just Part 1. above

Unit 2, Part 5

David Maier

1

Principles of Databases

Example

Quarter Q *dept* D *course Name* C *number* N *room* R *prof* P *time* T

$\{DC \rightarrow N, DN \rightarrow C, QDCT \rightarrow RP, P \rightarrow D, QRT \rightarrow DC\}$

$QRT^+ = QRTDCNP$

$QR^+ = QR$

$QT^+ = QT$

$RT^+ = RT$

$Q^+ = Q$

$R^+ = R$

$T^+ = T$

if none of these is a superkey then none of these are

Unit 2, Part 5

David Maier

2

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

More Examples

$\{DC \rightarrow N, DN \rightarrow C, QDCT \rightarrow RP, P \rightarrow D, QRT \rightarrow DC\}$

$QCPT^+ = QCPTDNR$] is a key

$CPT^+ = CPTDN$

$QPT^+ = QPTD$

$QCT^+ = QCT$

$QCP^+ = QCPDN$

What about $QRPT$?

$QRPT^+ = QRPTDCN$ Superkey, not a key
but $QRT \subseteq QRPT$
key

Unit 2, Part 5

David Maier

3

Principles of Databases

Different Example

Flight **D**ate **P**ilot **G**ate

$\{LD \rightarrow PG, L \rightarrow G\}$

LD is a key, so L, D are prime

P, G are not prime - not part of any key

Unit 2, Part 5

David Maier

4

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

Third Normal Form (3NF)

Schema R with FDs F is in third normal form if for any nontrivial FD $X \rightarrow A$,

then X is a superkey of R or A is prime

Last example was not in 3NF

Consider $L \rightarrow G$ ← not prime
/
not a superkey
L⁺ = LG

Unit 2, Part 5

David Maier

5

Principles of Databases

Another Example

not 3NF

FLight Date PilotID PilotName

{LD → IN, I → N, N → I}

non trivial
I, N not superkeys
I, N not prime

Original Def.
transitive dependency
not allowed
 $X \rightarrow Y, Y \twoheadrightarrow X, Y \rightarrow A, A \not\rightarrow XY$
 $LD \rightarrow I, I \twoheadrightarrow LD, I \rightarrow N, N \not\rightarrow LDI$

Unit 2, Part 5

David Maier

6

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

Aside: Boyce-Codd Normal Form (BCNF)

Schema R with FDs F is in Boyce-Codd normal form if for any nontrivial FD $X \rightarrow A$, X must be a super key for R

Unit 2, Part 5

David Maier

7

Principles of Databases

BCNF Example

Airport Company BillTo

$\{A C \rightarrow B, B \rightarrow C\}$

This schema is in 3NF

OK - C is prime

This schema is not in BCNF

B \rightarrow C

B not a superkey

AB BC

Unit 2, Part 5

David Maier

8

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

Normalization through Decomposition

If R not in 3NF for FDs F , find an FD $Y \rightarrow Z$ in F^+ where Y is not a key; split into

$R-Z$ and YZ

FLight Date PilotID PilotName

{LD \rightarrow IN, I \rightarrow N, N \rightarrow I}

I \rightarrow N

N \rightarrow I

LDI

IN

LDN

IN

Unit 2, Part 5

David Maier

9

Principles of Databases

Problems with Decomposition

1. Figuring out non-prime attributes can be hard
2. Might end up with more relations than you need
3. Might not be able to enforce FDs

Unit 2, Part 5

David Maier

10

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

Decomposition Example

$R = A B C D E$

$\{AB \rightarrow CDE, AC \rightarrow BDE, B \rightarrow C, C \rightarrow B, C \rightarrow D, B \rightarrow E\}$

AB, AC are keys, so $C \rightarrow D$ is a problem

Decompose on $C \rightarrow D$

$ABCE \quad CD$

$B \rightarrow E$ is a problem, decompose again

$ABC \quad BE$

But ... ABC, BDE in 3NF, only 2 relations

Unit 2, Part 5

David Maier

11

Principles of Databases

Enforcing FDs Example

$R = A B C D E$

$\{A \rightarrow BCDE, CD \rightarrow E, CE \rightarrow B\}$

A is the only key, so $CD \rightarrow E$ is a problem

Decompose on $CD \rightarrow E$

$ABCD, CDE$

How do you enforce $CE \rightarrow B$?

Unit 2, Part 5

David Maier

12

Normalization through Synthesis

Start with schema R and FDs F

Generate database schema

R_1, R_2, \dots, R_n such that

- Every relation schema is in 3NF
- Can enforce F
- Avoid extra relation schemas
- Have a lossless decomposition from R onto R_1, R_2, \dots, R_n .

$$F = F \cap [R_1, R_2, \dots, R_n]$$

Depends on a cover for F of a certain form

What's a Cover?

A set of FDs G is a cover for a set of FDs F if $F^+ = G^+$

G is a non-redundant cover if you cannot eliminate FDs and still be a cover

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

$$G = \{A \rightarrow B, B \rightarrow C\} \leftarrow \text{non-redundant cover for } F.$$

Can test FDs one by one to see if any can be left out

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

Minimum Cover

A minimum cover is one with fewest possible FDs

Non-redundant, but not minimum:

$\{A \rightarrow BC, B \rightarrow A, AD \rightarrow E, BD \rightarrow I\}$

minimum
 $\{A \rightarrow BC, B \rightarrow A, AD \rightarrow EI\}$

Unit 2, Part 5

David Maier

15

Principles of Databases

Finding a Minimum Cover

Replace every FD $X \rightarrow Y$ in F with $X \rightarrow X^+$, then find a non-redundant cover

$\{A \rightarrow BC, B \rightarrow A, AD \rightarrow E, BD \rightarrow I\}$

$A \rightarrow ABC, B \rightarrow ABC, AD \rightarrow ABCDEI, BD \rightarrow \cancel{ABCDEI}$

minimum
 $\{A \rightarrow ABC, B \rightarrow ABC, AD \rightarrow ABCDEI\}$

Unit 2, Part 5

David Maier

16

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

Reduced Covers

Are there extraneous attributes?
Can appear on the right side or the left

$$\{\cancel{AB} \rightarrow C, A \rightarrow B\}$$

$$\{A \rightarrow \cancel{B}, B \rightarrow C\}$$

How to remove?

For $XA \rightarrow Y$, see if $(F - \{XA \rightarrow Y\} \cup \{X \rightarrow Y\}) \stackrel{?}{=} F$

For $X \rightarrow YB$, see if $(F - \{X \rightarrow YB\} \cup \{X \rightarrow Y\}) \stackrel{?}{=} F$

F'

$F' \stackrel{?}{=} X \rightarrow B$

$F \stackrel{?}{=} X \rightarrow Y$

Unit 2, Part 5

David Maier

17

Principles of Databases

Reduce Left Sides First

$$G = \{A \rightarrow C, AB \rightarrow DE, AB \rightarrow CDI, \underline{AC} \rightarrow J\}$$

Reduce left sides

$$G' = \{A \rightarrow C, \underline{AB} \rightarrow \underline{DE}, \underline{AB} \rightarrow \underline{CDI}, A \rightarrow J\}$$

Reduce right sides

$$G'' = \{A \rightarrow C, AB \rightarrow E, AB \rightarrow DI, A \rightarrow J\}$$

Unit 2, Part 5

David Maier

18

CS 589 Principles of Databases, Unit 2, Part 5

Principles of Databases

Synthesis Algorithm

Start with FDs F

Find G = minimum cover for F

Find H = reduced cover for G

For each $X \rightarrow Y$ in H , add a schema XY

Unit 2, Part 5

David Maier

19

Principles of Databases

Synthesis Example

$$F = \{A \rightarrow BC, B \rightarrow A, AD \rightarrow E, BD \rightarrow I\}$$

$$F^+ = \{A \rightarrow ABC, B \rightarrow ABC, AD \rightarrow ABCDEI, BD \rightarrow ABCDEI\}$$

$$G = \{A \rightarrow \cancel{ABC}, B \rightarrow \cancel{ABC}, AD \rightarrow \cancel{ABC}EI\}$$

$$H = \{A \rightarrow BC, B \rightarrow A, AD \rightarrow EI\}$$

$$R1 = (\underline{A}BC) \quad R2 = (A\underline{B}) \quad R3 = (A\underline{DEI})$$

$$R12 (A B C) \\ \text{Keys} = \{A, B\}$$

Unit 2, Part 5

David Maier

20

Lossless Decomposition

If none of the FDs represents a key for all the attributes, need to add one more relation to get a lossless decomposition

$R = ABCDE$

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

"
H

$R_1 = (AB) \quad R_2 = (CD)$
add $R_3 = (ACE)$