Database design: E/R diagrams and BCNF

CSE 444 section

October 14, 2010
Today

• Database design with E/R diagrams
• Functional dependencies
• Boyce-Codd normal form (BCNF)
From English to an E/R diagram

- **Professors** have SSN, age, rank, and specialty
- **Projects** have IDs, sponsors, budgets, start and end dates
From English to an E/R diagram

- Each project is **managed by** one professor (principal investigator)
- A professor can **manage** multiple projects

From English to an E/R diagram

- Each project is **worked on** by one or more professors
- Professors can **work on** multiple projects
From E/R diagram to relations

- Professor (ssn, age, rank, specialty)
- Project (pid, sponsor, start_date, end_date, budget)
- Work_on (ssn, pid)
- Manages (ssn, pid)

Integrating the many-one relation

- Professor \((ssn, age, rank, specialty)\)
- Project \((pid, sponsor, start\_date, end\_date, budget, ssn)\)
- Work\_on \((ssn, pid)\)
SQL code for this database

CREATE TABLE Professor (  
    ssn  INT PRIMARY KEY,  
    age  INT,  
    urank  VARCHAR(30),  
    specialty  VARCHAR(30)  
);  

CREATE TABLE Project (  
    pid  INT PRIMARY KEY,  
    sponser  INT,  
    start_date  DATE,  
    end_date  DATE,  
    budget  FLOAT,  
    ssn  INT REFERENCES Professor(ssn)  
);  

CREATE TABLE Work_on (  
    ssn  INT REFERENCES Professor(ssn),  
    pid  INT REFERENCES Project(pid),  
    PRIMARY KEY (ssn, pid)  
);  

Today

• Database design with E/R diagrams
• Functional dependencies
• Boyce-Codd normal form (BCNF)
Garcia-Molina, problem 3.3.2 (i)

Consider a relation $S(A,B,C,D)$ with FDs
$A \rightarrow B$, $B \rightarrow C$, and $B \rightarrow D$.

a. Give the nontrivial FDs that follow from the given FDs. Restrict to 1 attr on right side.
b. What are all the keys of $S$?
c. What are the superkeys that aren’t keys?
Consider a relation $T(A,B,C,D)$ with FDs $AB \rightarrow C$, $BC \rightarrow D$, $CD \rightarrow A$, and $AD \rightarrow B$.

a. Give the nontrivial FDs that follow from the given FDs. Restrict to 1 attr on right side.

b. What are all the keys of $S$?

c. What are the superkeys that aren’t keys?
Today

• Database design with E/R diagrams
• Functional dependencies
• Boyce-Codd normal form (BCNF)
What is BCNF?

A relation R is in BCNF iff:

If $A_1, ..., A_n \rightarrow B$ is a non-trivial dependency in R, then $\{A_1, ..., A_n\}$ is a superkey for R
Why do BCNF decompositions?
BCNF decomposition algorithm

BCNF_Decompose(R)

find X s.t.: X ≠ X⁺ ≠ [all attributes]

if (not found) then “R is in BCNF”

let Y = X⁺ - X
let Z = [all attributes] - X⁺

decompose R into R1(X ∪ Y) and R2(X ∪ Z)
continue to decompose recursively R1 and R2
Consider the following FDs:

- $CD \rightarrow E$  
  BAD
- $D \rightarrow B$  
  BAD
- $A \rightarrow CD$

Which ones are the bad dependencies?

- $CD^+ = BCDE$
- $D^+ = BD$
- $A^+ = ABCDE$
- CD is not a superkey
- D is not a superkey
- A is a superkey
Consider the following FDs:

- CD → E  BAD
- D → B  BAD
- A → CD

BCNF example: table R(A, B, C, D, E)

R(A,B,C,D,E)  [CD+ = BCDE ≠ ABCDE]

R2(B,C,D,E)  [D+ = BD ≠ BCDE]

R3(A,C,D)  [BCNF]

R4(B,D)  [BCNF]

R5(C,D,E)  [BCNF]
2 more BCNF decompositions

\[ S(A, B, C, D) \]
\[ C \rightarrow D, C \rightarrow A, B \rightarrow C \]

\[ T(A, B, C, D, E) \]
\[ AB \rightarrow C, DE \rightarrow C, B \rightarrow D \]
Consider the following FDs:

- C → D, C+ = ACD  BAD
- C → A, C+ = ACD  BAD
- B → C, B+ = ABCD

S(A,B,C,D) solution

S(A,B,C,D) [C+ = ACD ≠ ABCD]
Consider the following FDs:

- $AB \rightarrow C$, $AB^+ = ABCD$  BAD
- $DE \rightarrow C$, $DE^+ = CDE$  BAD
- $B \rightarrow D$, $B^+ = BD$  BAD

T(A,B,C,D,E) 1st solution
Consider the following FDs:

- $AB \rightarrow C$, $AB^+ = ABCD$  BAD
- $DE \rightarrow C$, $DE^+ = CDE$  BAD
- $B \rightarrow D$, $B^+ = BD$  BAD

$T(A,B,C,D,E)$

$T(A,B,C,D,E)$  
[DE$^+$ = CDE ≠ ABCDE]

$T2(A,B,D,E)$
[B$^+$ = BD ≠ ABDE]

$T3(C,D,E)$  
[BCNF]

$T4(B,D)$  
[BCNF]

$T5(A,B,E)$  
[BCNF]
T(A,B,C,D,E) 3rd solution

Consider the following FDs:

- AB → C, AB+ = ABCD  BAD
- DE → C, DE+ = CDE  BAD
- B → D, B+ = BD  BAD