BCNF

Friday, May 24, 2019 11:41 AM

- Funtional Dependency (FD) for relation R: $A \to B$.
  - Trivial FD: $RHS \subseteq LHS$.

    e.g. $\{a, b\} \to \{a\}$

    - Trivial FDs do not provide real restrictions on the data in the relation, and they are usually ignored.
  - Buzzwords:
    - FD holds or does not hold on an instance.
    - If we can be sure that every instance of R will be one in which a given FD is true, then we say that R satisfies the FD.
    - If we say that R satisfies an FD, we are stating a constraint on R.
- Keys
  - Super key: attributes that determines all other attributes in a schema.
  - Minimal key / key: Set of attributes with smallest cardinality that does the work of super key. ⇒ A super key for which no subset is a superkey.
  - We can have more than 1 key, but in SQL we can only have 1 primary key.

- Closure

  - def: Given a set of attributes $\{A_1, \cdots, A_n\}$, the closure is the set of attributes B, noted $\{A_1, \cdots, A_n\}^+$ s.t. $A_1, \cdots, A_n \to B$.

  - Closure algorithm

$X = \{A_1, \cdots, A_n\}$.
repeat until X doesn't change do:
if $B_1, \cdots, B_n \to C$ and $B_1, \cdots, B_n \in X$
then add C to X.

    - Thus we can check that if $A \to Q$
      - Compute A^+, then check if $Q \subseteq A^+$.

    - Check if set P is the superkey of R.

      if $P^+ = R$, then superkey.

- Eliminating Anomalies

  $X \to A$ is OK to be a key if X is a (super)key.
  Not OK otherwise.

- Boyce-Codd Normal Form

  - def1: A relation R is in BCNF if whenever $X \to B$ is a non-trivial dependency, then X is a superkey.

    $\Updownarrow$

    def2. $\forall X$, either $\underset{X \text{ is not in any FD.}}{\underline{X^+ = X}}$ or $X^+ = R$.

- BCNF decomposition algorithm
  - → Normalize R.

find X s.t. $X \neq X^+$ and $X^+ \neq R$
if (not found) then R is in BCNF
let $Y = X^+ - X$; $Z = R - X^+$
decompose R into $R1(X \cup Y)$, $R2(X \cup Z)$
normalize($R1$); normalize($R2$);

- **Lossless Decomposition.**
  - decomposition is lossless if $R = S_1 \bowtie S_2 \dots \bowtie S_n$
    - every BCNF decomposition is lossless
    - $R(A_1, \dots, A_n, B_1, \dots, B_m, C_1, \dots, C_p)$

$$S_1(A_1, \dots, A_n, B_1, \dots, B_m) \qquad S_2(A_1, \dots, A_n, C_1, \dots, C_p)$$

S_1 = projection of R on $A_1, \dots, A_n, B_1, \dots, B_m$.

S_2 = projection of R on $A_1, \dots, A_n, C_1, \dots, C_p$.

The decomposition is called lossless if $R = S_1 \bowtie S_2$

  - The Chase Test for lossless join

e.g., We want to know that $R(A, B, C, D) = S1(A, D) \bowtie S2(A, C) \bowtie S_3(B, C, D)$

R satisfies $A \to B, B \to C, CD \to A$;

$\subseteq$) It is easy to know that $R \subseteq S_1 \bowtie S_2 \bowtie S_3$.

$\supseteq$) To prove $S_1 \bowtie S_2 \bowtie S_3 \subseteq R$, we use the chase test.

Suppose $(a, b, c, d) \in S_1 \bowtie S_2 \bowtie S_3$

A	B	C	D	
a	b1	c2 c	d	$(a, d) \in S_1$
a	b2 b1	c	d1	$(a, c) \in S_2$
a1 a	b	c	d	$(b, c, d) \in S_3$

thus we have (a, b, c, d), which implies that $R_1 \bowtie R_2 \bowtie R_3 \subseteq R$

$\because A \to B \quad \therefore b_2 = b_1$

$\because B \to C \quad \therefore c_2 = c$

$\because CD \to A \quad \therefore a_1 = a$

- Schema Refinements = Normal Forms
  - 1st Normal Form = all tables are flat
  - 2nd Normal Form = obsolete
  - Boyce Codd Normal Form = no bad FDs
  - 3rd Normal Form
    - BCNF removes anomalies, but may lose some FDs.
    - 3NF preserves all FD's, but may still have some anomalies.