CS 541

Course No.

5

Item No.

NORMALIZATION: WHY

Title

Bhargava CS 541 Lib. Reserve

NORMALIZATION

Why?

Three types of misbehavior

- UPDATE
- Insertion
- Delete

Each Relation should describe a single concept.

A relation is in third Normal Form if every determinant is key

$$A \xrightarrow{\text{determines}} B$$

determinant

NORMALIZATION

ITEM	PRICE	Date	QTY ordered
Toaster	20.00	1/10/78	2
Toaster	20.00	2/15/78	5
Mixer	28.00	4/6/75	3

Insertion Problem] Lack of inf.
Deletion Problem] loss

item	Price

item	Date	Qty

IF

$A \rightarrow B$

$B \rightarrow C$

THEN $A \rightarrow C$

② IF

$A \rightarrow B$

THEN

$AB \rightarrow B$

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

$A \rightarrow D$

$A \rightarrow E$

$AB \rightarrow C$

$R_1(\underline{A}, \underline{B}, C)$ $R_2(\underline{A}, D, E)$

Let X, Y, Z be subset of all attributes(U) of a relation R.

AXIOMS

① IF $Y \subseteq X \subseteq U$
Then $X \rightarrow Y$ Reflexivity

② IF $X \rightarrow Y$ and $Z \subseteq U$
Then $XZ \rightarrow YZ$ Augmentation

③ IF $X \rightarrow Y$ and $Y \rightarrow Z$
Then $X \rightarrow Z$ Transitivity

THEOREM

ARMSTRONG'S AXIOMS ARE SOUND and COMPLETE.

SOUND:

IF $X \to Y$ is deduced from F and Axio[ms]

THEN $X \to Y$ is true in any relation in which F is true

REFLEXIVITY: IF two tuples of r agree on X, they must agree on a subset of X.

AUGMENTATION:

IF two tuples agree on XZ but not on YZ then they must agree on X but not on Y

contradiction ($X \to Y$)

TRANSITIVITY:

IF two tuples agree on X, they agree on Y. If they agree on Y, they agree on Z. So $X \to Z$

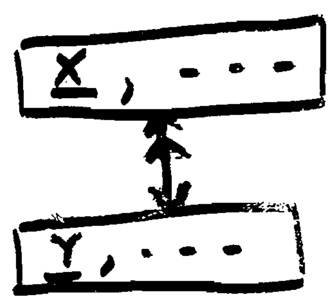
Functional Dependency (Revisited)

$$X \rightarrow Y$$

means
X functionally determines Y
or Y is functionally dependent on X
if it is not possible that relation R
has two tuples that
agree on value of X
and disagree on value of Y

Many - one mapping

$$X \rightarrow Y$$



one to one mapping

$$X \rightarrow Y$$

$$Y \rightarrow X$$

Some more rules

Union IF $X \to Y$, $X \to Z$

THEN $X \to YZ$

Pseudotransitivity

IF $X \to Y$

and $WY \to Z$

Then $XW \to Z$

Decomposition

IF

$X \to Y$ and $Z \subseteq Y$

Then $X \to Z$

IF F is a set of functional dependencies

F^+ is the set of all functional dependencies derivable from F

F^+ is called closure of F.

Proof

Union Rule

$X \to Y \Rightarrow X \to XY$

$X \to Z \Rightarrow XY \to ZY$ (or YZ)

So $X \to XY \to YZ$ QED

Pseudotransitivity Rule

$X \to Y \Rightarrow XW \to YW$

But $YW \to Z$

so $XW \to Z$ QED

Decomposition Rule

$X \to Y$

Tuples that agree on X do agree on Y and so they do agree on Subset of Y

But $Z \subseteq Y$ So they do agree on Z

So $X \to Z$ QED.

EXAMPLE

$R = (A, B, C, D, E, G)$

F :

$AB \to C$ $\quad$ $D \to EG$

$C \to A$ $\quad$ $BE \to C$

$BC \to D$ $\quad$ $CG \to BD$

$ACD \to B$ $\quad$ $CE \to AG$

$(BD)^+$ = Set of attributes that are dependent on attributes B, D

= ABCDEG = R

Thus BD determines R

We call BD as the key of R

A relation by definition is in FIRST NORMAL FORM

- A relation is in 2^{nd} NF if any one of the following is true

1. The key consists of a single attribute
2. There are no non-key attributes
3. Every non-key attribute depends on all of the key

Example

$R(\underline{A}, B)$ Case 1

$R(\underline{A}, \underline{B}, \underline{C})$ Case 2

$R(\underline{A}, \underline{B}, C, D)$ Case 3.

and $AB \rightarrow C$

$AB \rightarrow D$

- A relation is in 3NF, if it is in 2^{nd} NF and has no transitive dependencies

Mathematically

R is in 3^{NF}

if $\nexists$ key X for R and $Y \subseteq R$

and a non-key attribute A not in X or Y

Such that

1. $X \rightarrow Y$		$\overbrace{X}$ $\overbrace{A}$
2. $Y \rightarrow A$	$Y \nsubseteq X$	$\overbrace{X}$ $\overbrace{A}$ $\overbrace{Y}$
3 $Y \not\rightarrow X$	$Y \subseteq X$	$\overbrace{X}$ $\overbrace{Y}$ $\overbrace{A}$

IF Y is a subset of X

Then R has partial dependency

IF Y is not a subset of X

Then R has a transitive dependency

A set of F is minimal if

a) Every right side is a single attribute

b) For no $X \to A$, $F - \{X - A\} \equiv F$

c) For no $X \to A$, $Z \subset X$

$$[F - \{X \to A\}] \cup \{Z \to A\} \equiv F$$

Lemma:

F is covered by G in which no right side has more than one attribute

IF $X \to A$ in G $X \to Y$ in F and $A \in Y$

By decomposition $X \to A$ in F^+

So $G \subseteq F^+$

Let $Y = A_1 \cdot A_2 \cdot \ldots A_n$

IF $X \to Y$ in F Then

$X \to A_1$ in G
$X \to A_2$
$\ldots$
$X \to A_n$

So $F \subseteq G^+$

$F^+ \equiv G^+$

Lossless join Decomposition

Let r be a relation for scheme R satisfying dependencies D

Let $\rho = \{R_1, \ldots R_n)$ be a decomposition satisfying D.

Then the decomposition is lossless if

$$r = \pi_{R_1}(r) \bowtie \pi_{R_2}(r) \bowtie \cdots \bowtie \pi_{R_n}(r)$$

= Natural Join of its projections on the R_i's.

Let $m_\rho(r) = \bowtie_{i=1}^{n} \pi_{R_i}(r)$

Decomp is lossless if

$$r = m_\rho(r)$$

Let $r_i = \pi_{R_i}(r)$

Lemma

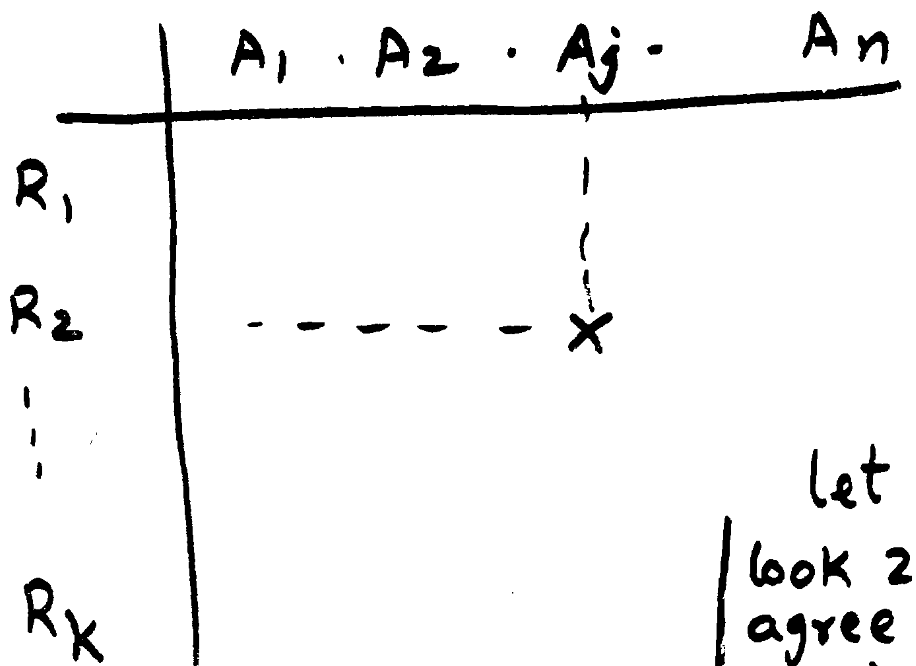
a) $r \subseteq m_\rho(r)$

b) If $s = m_\rho(r)$, then $\pi_{R_i}(s) = r_i$

c) $m_\rho(m_\rho(r)) = m_\rho(r)$

Testing Lossless Joins

$R = A_1 \cdots A_n \qquad \rho = (R_1, R_2, \ldots R_k)$



let $X \to Y$
look 2 rows that agree on X
Equate elements of Y ∋ if one is a both are a
if both are b, leave them as b.

$x = a_j$ if $A_j \in R_2$

$x = b_{2j}$ if $A_j \notin R_2$

Do Recursively:
If ONE ROW is all a's. HALT

$R = (A, B, C, D, E)$

$R_1 = AD$

$R_2 = AB$

$R_3 = BE$

$R_4 = CDE$

$R_5 = AE$

$A \rightarrow C$ ✓

$B \rightarrow C$ ✓

$C \rightarrow D$

$DE \rightarrow C$

$CE \rightarrow A$ ✓

	A	B	C	D	E
AD	a_1	b_{12}	b_{13}	a_4	b_{15}
AB	a_1	a_2	$b_{\cancel{2}3}$	a_4 $\cancel{b_{24}}$	b_{25}
BE	a_1 $\cancel{b_{31}}$	a_2	a_3 $\cancel{b_{33}}$	a_4 $\cancel{b_{34}}$	a_5
CDE	b_{41}	b_{42}	a_3	a_4	a_5
AE	a_1	b_{52}	$b_{\cancel{5}3}$	a_4 $\cancel{b_{54}}$	a_5

Proof: Algorithm correctly determines if a decomposition has a loss less join

One way $\longrightarrow$

Suppose the final table does not have a row with all a's. Let this be a relation r for R.

Then we must prove $r \neq m_\rho(r)$

Now for each R_i $\exists t_i \in r$ with all a's in ith row.

$m_\rho(r) = \overset{K}{\underset{r=1}{\bowtie}} \pi_{R_i}(r)$ contains a row with all a's

So r with no rows of a $\neq m_\rho(r)$ with a row of all a's

Reverse $\longleftarrow$

Please read yourself

Superkey — superset of a key

Candidate key — minimal set of attribute

key — one designated Candidate key

$R(\text{city}, \text{st}, \text{zip})$

$\text{city}, \text{st} \rightarrow \text{zip}$

$\text{zip} \rightarrow \text{city}$

$(\text{city}\ \text{st})$ (st, zip) are keys

An Att is prime att of R if it is a member of any key of R

Non prime att $\Rightarrow$ not a member of any key of R

3NF

X is a superkey of R

or A is a prime att of R

---

3 A rel is 3NF
if every non prime att of R

is 1) Fully functionally dep
on every key of R

& 2) Non-transitively dep
on every key of R

A rel is 2NF
if every nonprime att A
in R is not partially dep
on any key of R

Boyce Codd Normal Form

A rel R with dep F is in BCNF if

$X \longrightarrow A$ holds in R

and $A \notin X$

X is a superkey of R

X is or contains the key.

Diff between 3NF & BCNF

3NF allows A to be prime if X is not a superkey

NP Complete to determine if a R is in BC

Theorem

Every set of dependencies F is equivalent to a set of dependencies F' that is minimal.

Proof by construction

<u>Step 1</u> Change F to get single attribute on right side

<u>Step 2</u> If a dependency $X \to Y$ can be eliminated without changing F$^+$ do it.
(you may have several choices)

<u>Step 3</u> Eliminate attributes from the left side.

$$\left.\begin{array}{r} XY \to Z \\ Y \to Z \end{array}\right\} \text{Eliminate } X$$

Theorem

If a relation R is in BCNF

Then it is in 3^{NF}.

Proof

Let R in BCNF & not in 3^{NF}

Then $X \to Y \to A$ is in F (Partial or Transitive)

X is a key for R

$A \nsubseteq X$ or $A \subsetneq Y$ and $Y \not\to X$

If $Y \not\to X$

Then Y does not include the key for R

But $Y \to A$ violates that R in BCNF.

4^{th} NF, Multivalued Dependencies ?

Theorem:

If $\rho = (R_1, R_2)$

Then ρ has a lossless join w.r.t F

iff

$$\left.\begin{aligned} &R_1 \cap R_2 \rightarrow R_1 - R_2 \\ \text{or } &R_1 \cap R_2 \rightarrow R_2 - R_1 \end{aligned}\right\} \in F^+$$

Example

$R = (A, B, C)$

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

$R_1(A, B)$

$R_2(B, C)$

$R_1 \cap R_2 = B$

$R_1 - R_2 = A$

$R_2 - R_1 = C$

$B \not\rightarrow A$, $B \not\rightarrow C$

So Decomposition is lossy

$R_1(A, B)$

$R_2(A, C)$

$R_1 \cap R_2 = A$

$R_1 - R_2 = B$

$A \rightarrow B$

$R_1 \cap R_2 \rightarrow R_1 - R_2$

Decomposition has a lossless join

Algorithm for Lossless Join Decomposition into B-C NF.

Initially $\rho = R$.

For $S \in \rho$ if S not in BCNF

then $X \to A$ holds in S

$\ni$ X does not include a key for S
and $A \notin X$

let attribute $A_k \in S$

$\notin A$

$\notin X$

Then $S = (S_1, S_2)$

$\ni$ $S_1 = (X, A)$

$S_2 = (S - A) \neq \phi$ { $\because$ contains A_k }

Decomposition of S is (S_1, S_2)
Keep Iterating till all S_i in BCN

Lemma

$$\rho = (R_1, \ldots R_i, \ldots\ldots R_k)$$

Decomposition (lossless join) of R_i:

$$S_1, S_2 \ldots S_m$$

If ρ has a lossless join w.r.t F

$$\rho_1 = (R_1, \ldots R_{i-1}, S_1, S_2, \ldots S_m, R_{i+1}, \ldots R_k)$$

$\longrightarrow$ ρ_1 has a lossless join w.r.t F

$$\rho_2 = (R_1, \ldots, R_i, \ldots R_k, R_{k+1}, \ldots R_n)$$

ρ_2 include ρ and some more

then ρ_2 also has a lossless join w.r.t F

Decompositions that Preserve Dependencies

Projection of F onto Z ($\pi_Z(F)$)
is set of dependencies $X \to Y$ in F^+
such that $XY \subseteq Z$.

A decomposition ρ preserves a
set of dependencies F

if $\bigcup_{i=1}^{k} \pi_{R_i}(F) \subseteq F$

$$\left(\bigcup_{i=1}^{k} \pi_{R_i}(F) \right)^+ = F^+$$

Relation Schema

Lossless Join Decomposition

B-C NF

Lossless Join
and
Dependency preserving
Decomposition

3NF

$R(C, S, Z)$

$R_1(S, Z)$

$R_2(C, Z)$

$\Downarrow$

F:

$C, S \rightarrow Z$

$Z \rightarrow C$

Is it a lossless join Decomp?
Is it a FD preserving Decomp?

$R(A, B, C, D)$

$\rho = \{ AB \quad BC \quad CD \}$

$F: \; A \to B$

$B \to C$

$C \to D$

$D \to A$

$R_1(A, B)$

F_1

$A \to B \qquad \therefore \; \Pi_{R_1}(F^+) = F_1.$

and $B \to A$

$R(ABCD) \qquad \rho(AB, CD)$

$F: \; A \to B$

$C \to D.$

Preserves Functional Dep.

but yet lossy join.

Example

$$R_1 = (C, T, H, R, S, G)$$

$C \rightarrow T$

$HR \rightarrow C$

$HT \rightarrow R$

$CS \rightarrow G$

$HS \rightarrow R$

$(HS)^+ =$

Now $CS \rightarrow G$ violates that R_1 in BCNF

R_1

CSG   CSTHR

To break CSTHR, project F^+ on C, S, T, H, R

$C \rightarrow T$ $HT \rightarrow R$

$HR \rightarrow C$ $HS \rightarrow R$

What is the Key for CSTHR, HS

$C \to T$ violates that CSTHR in BCNF

CSTHR

CT   CSHR

Dependencies

$CH \to R$?

$HS \to R$

$HR \to C$

Key HS

$CH \to R$ violates CSHR in BCN

CSHR

CHR   CHS

$R_1 = \{ CGS, CT, CHR, CHS \}$ is in BC

Exponential Complexity

To test if R in BCNF is NP-COMPLETE.

✓ $CE \to A$ eliminated because $C \to A$

✓ $ACD \to B$ reduced to $CD \to B$

✓ $CG \to D$, $C \to A$, $ACD \to B$

$$\begin{array}{ccc} CG \to D & & ACD \to B \\ \Downarrow & & \Downarrow \\ CG \to CD & & CD \to B \\ & \Downarrow & \\ & CG \to B & \end{array}$$

DECOMPOSITION OF RELATION SCHEME

$$R = \{A_1, A_2, \ldots A_n\}$$

$$\Big\downarrow \text{Decompose}$$

$$\rho = \{R_1, R_2, \ldots R_K\} \ni R_1 \cup R_2 \cdots \cup R_K = [illegible]$$

Example $R = \{S, A, I, P\}$ $\quad R_1 = \{S, A\}$

$R_2 = \{S, I, P\}$

R

| lossless join

$\rho = (R_1, R_2, \cdots R_K)$

Each R_i in BCNF

& so in 3NF

Preserve set of dependencies

$\sigma = (S_1, S_2 \cdots\cdots S_m)$

Let $\tau = \sigma \cup \{X\}$

where X is the key for R

All relation in τ are in 3NF

Decomposition preserves Depend.

& has a lossless Join.

Att / DECOMP	Y Ai R - X	X
	row of a's	
X	a a a a a	a a a a

$Y_i \subseteq X \cup \{R - X\}$

$Y_i \rightarrow A_i$

Alg 5.4 Dependency Preserving Decomp. into 3NF

- If any attribute not in F it is one decomposition
- If any dependies contains all attributes of R, R is one decomp
- Otherwize Decomposition is X A where X → A in F

R = {C, T, H, R, S, G}

Minimal cover

C → T

HR → C

HT → R

CS → G

HS → R

T = {CT, HRC, HTR, CSG, HSR}