Normalization

Jatin

The normal form of relation refers to the highest normal form condition that it meets, and hence indicates the degree to which it has been normalized.

Normalization of data can be considered a process of analysing the given relation schemas based on their FDs and primary keys to achieve the desirable properties of (1) minimizing redundancy and minimizing the insertion, deletion and update anomalies.

Two propert must hold

→ (1) lossless Join
(2) Dependency preservation.

Denormalization is the process of storing the join of higher normal form relations as a base rel^n.

Attributes is considered as prime attribute of R if it is member of some candidate key

Attribute not member of candidate key are Non-Prime Attribute.

W

Guidelines for Database Design

(1) Design a relation schema so that it is ea
to explain its meaning. Do not combine
attributes from multiple entity types and
relationship types into a single relation.

(2) Design the base relation schemas so that
no insertion, deletion, or modification anomalies
are present in the relations

(3) Avoid placing attributes in a base relation
whose values may frequently be NULL. If NULLs
are unavoidable, make sure that they apply in
exceptional cases only and do not apply to
majority of tuples in the relation.

(4) Design relation schemas so that they can be
joined with equality condition on attributes
that are appropriately related pairs in a way
that guarantees that no spurious tuples are generat

Anomalies

Update anomalies

→ Insertion anomalies
→ Deletion anomalies
→ Modification anomalies

It is combine and

Functional Dependencies

A functional dependency is a constraint between two sets of attributes from the database. let $R = \{A_1, A_2 \ldots A_n\}$

A functional dependency, denoted by $X \rightarrow Y$ between two sets of attributes X and Y that are subsets of R specifies a constraint on the possible tuples that can form a relation state r of R. For any two tuples t_1 and t_2 in r that have $t_1[A] = t_2[A]$ they must also have $t_1[B] = t_2[B]$.

r(A, B, C, G, H, I)

$A \rightarrow B$
$A \rightarrow C$
$CG \rightarrow H$
$CG \rightarrow I$
$B \rightarrow H$

X		Y
a	$\rightarrow$	1
a	$\rightarrow$	2 X
d	$\rightarrow$	4
c	$\rightarrow$	3

X	Y
a	1
a	2
c	3
d	4

Type of FD

Trivial	Non Trivial
$AB \rightarrow A$	$Y \not\subseteq X$
$Y \subseteq X$	
Y is a proper subset of X	
→ valid.	
→ No new info found	

A → BC.
DE → C
C → DE
BC → A

R

A	B	C	D	E
a	2	3	4	5
2	a	3	4	5
a	2	3	6	5
a	2	3	6	6

F.D doesnot depend on data rather Data depends on F.D.

Check whether two value of A are giving same value or not:

for eg a → 2, 3

∴ A → BC is valid

DE → C ✓

4, 5 → 3
6, 5 → 3

Now all value on RHS are same ∴ rule is valid

C → DE

3 → 4, 5
3 → 6, 5

two different value for same value of C ∴ it is invalid

BC → A

2, 3 → a
a, 3 → 2

A	B	C	D	E
a	2	3	4	5
b	a	3	4	5
c	2	3	6	5
d	2	3	6	6

$A \to B$ ✓

$A \to CD$ ✓

$A \to BCDE$ ✓

if all the value on L.H.S are different then all the F.D are valid $\alpha \to \beta$

$A \to C$ ✓

$BD \to C$ ✓

$ABDE \to C$ ✓

if all the value on R.H.S are same that is shown in above example the all the rules are valid

CLOSURE SET of F.D

Given a relational schema r(R), a FD, f on R is logically implied by a set of functional dependencies, F on r if any instance of r(R) that satisfies F also satisfies f.

$$F = \left\{ \begin{array}{l} A \to B \\ A \to C \\ CG \to H \\ CG \to I \\ B \to H \end{array} \right\}$$

$$F^+ \left\{ \begin{array}{l} A \to H \\ A \to A \\ B \to B \\ CG \to C \\ CG \to G \end{array} \right\}$$

Computing F^+ is very lengthy and difficult process. We can only check whether a FD belongs to F closure or not by applying certain Axiom or rules of inference.

By using these rules, we can find F^+ given F.

This collection of rules is called Armstrong's axiom. It is sound because it doesn't generate incorrect F.D. They are complete because they allow to compute F^+ if F is given.

(1) Reflexivity :- If $\beta \subseteq \alpha$ then
$$\alpha \rightarrow \beta$$

(2) Augmentation :- If $\alpha \rightarrow \beta$ then
$$\gamma\alpha \rightarrow \gamma\beta$$

(3) Transitivity :- If $\alpha \rightarrow \beta$, $\beta \rightarrow \gamma$ then
$$\alpha \rightarrow \gamma$$

Some additional rules to simplify the computation of F^+.

(1) Union rule: If $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$ then
$$\alpha \rightarrow \beta\gamma$$

(2) Decomposition :- If $\alpha \rightarrow \beta\gamma$ then
$$\alpha \rightarrow \beta, \quad \alpha \rightarrow \gamma$$

(3) Pseudotransitivity :- If $\alpha \rightarrow \beta$, $\gamma\beta \rightarrow \delta$ then
$$\alpha\gamma \rightarrow \delta$$

Composition :- if $\alpha \rightarrow \beta$, $\gamma \rightarrow \delta$ then $\alpha\gamma \rightarrow \beta\delta$

Scanned by CamScanner

Let us apply these rules (3)

$R = (A, B, C, G, H, I)$

$$F = \begin{bmatrix} A \to B \\ A \to C \\ CG \to H \\ CG \to I \\ B \to H \end{bmatrix}$$

$F^+ = \{ A \to H$: Transitivity holds

$CG \to HI$: Union rule holds

$AG \to I \}$: pseudotransitivity

Closure of Attribute Sets

Size of R, n (attribute in rel^n)

No of subset $= 2^n$

$F^+ = \{2^n \times 2^n\}$ possible FD

2^{2n}

Compute $(AG)^+$

Initially result $= AG$ (A, G, AG)

$A \to B$ ∴ result = result ∪ B

$A \to C$ ∴ AGBC

$CG \to I$ ∴ AGBCI

$B \to H$ ∴ AGBCIH

There are several uses of attribute closure

→ Superkey: we compute α^+ and check if contains all attribute in R.

→ we can check if a FD $\alpha \to \beta$ holds by checking if $\beta \subseteq \alpha^+$ that is we compute α^+ by using attribute closure, and then check if it contain β.

→ It gives alternate way to calculate F^+. For each $Z \subseteq R$ we find closure Z^+ and for each $S \subseteq Z^+$ we compute FD $Z \to S$.

Q1

$R(A, B, C)$

$A^+ = \{A, B, C\}$

$B^+ = \{B, C\}$

$C^+ = \{C\}$

Q2 $R(A\ B\ C\ D\ E\ F\ G)$

$A \to B$

$BC \to DE$

$AEG \to G$

find AC^+

$= \{A, C\}$

$\{A, B, C\} \because A \to B$

$\{A, B, C, D, E\} \because BC \to DE$

if we use normalization the AC can determine only ABC, DE

$A \to BC$

$CD \to E$

$B \to D$

$E \to A$

$B^+ = \{B\}$

$\{B, D\}$

$AB \to C$

$BC \to AD$

$D \to E$

$CF \to B$

$(AB)^+$

$= \{A, B\}$

$= \{A, \underline{B, C}\} \because AB \to C$

$\{A, B, C, D\} \because BC \to D$

$\{ABCDE\} \because D \to E$

$A \to BC$

$CD \to E$

$E \to C$

$D \to AEH$

$ABH \to BD$

$DH \to BC$

Find if $BCD \to H$?

Find $(BCD)^+$

BCDE

$\boxed{ABCDEH}$

Equivalen of FD

R(A B C D E H)

F	G
F : A → C	G: A → CD
AC → D	E → AH
E → AD	
E → H	
F ⊆ G	G ⊆ F
$(A)^+$ ACD	$(A)^+$ {ACD}
$(AC)^+$ ACD	$(E)^+$ {EADHC}
$(E)^+$ EADCH	

It mean F = G

There are two set of FD but they are not at same start point. Take both set of FD to same start point by find closure of each left hand side of F.D by using FD set of opposite set. Similarly for set G.

Now check whether all the ~~At~~ FD in set F are valid already to the attribute closure computed. If all the FD are valid then Both sets are equal.

<u>Canonical Cover</u>: Whenever user update a relation, the DB must ensure that update does not violate any FD. i.e all FD are satisfied in a new DB state. So verification of simplified FD set is easier because both sets would give same closure.

Minimal Set of FD [Canonical Cov

Irreducible Set of FD

$\alpha \rightarrow \beta$

R(WXYZ)

$X \rightarrow W$

$WZ \rightarrow XY$

$Y \rightarrow WXZ$

Extraneous attribute: An attribute is said be extraneous, if we can remove it w/o changing the closure set of FD

if $AB \rightarrow C$, $A \rightarrow C$

the B is extraneous

if $AB \rightarrow CD$, $A \rightarrow C$

the C is extraneous in RHS

if $AB \rightarrow CD$, $A \rightarrow E$, $E \rightarrow C$,

compute $(AB)^+ = \{ABCDE\}$

f_1 $X \rightarrow W$ $\quad X^+ = XW$ $\quad$ Now compute X^+ w/o $X \rightarrow W$ which is $X^+ = X$

Since both X^+ and $X^{+'}$ are not same $\therefore$ it is essential

$\times f_2$ $WZ \rightarrow X$ $\quad (WZ)^+ = WZXY\}$

f_3 $WZ \rightarrow Y$ $\quad (WZ)^{+'} = \{WZYX\}$

$\therefore WZ \rightarrow X$ is not essential

$\times f_4$ $Y \rightarrow W$

f_5 $Y \rightarrow X$ $\quad Y^+ = \{YW, X, Z\}$

f_6 $Y \rightarrow Z$ $\quad Y^{+'} = (W, X, Y, Z)$ $\therefore$ it is not essential

f_5: $Y^+ = XYW, Z$ $\quad Y^{+'} = \{Y, Z$ $\quad$ is essential

f_6 is essent

$(WZ)^+ = \{WZYX\}$

$(W^+) = W$

$(Z^+) = Z$

} LHS

(5)

Canonical cover F_c for F is a set of dependencies such that F logically implies all dependence in F_c, and F_c logically implies all dependencies in F.

→ No functional dependency in F_c contains extraneous Attribute

→ Each LHS of F.D in F_c is Unique.

$A \to BC$
$B \to C$
$A \to B$
$AB \to C$

Find F_c

~~Combine~~ ~~$A \to BC$, $A \to B$~~

Write all the FD in decompose form

$A \to BC \Rightarrow$
$A \to B$
$A \to C$
$B \to C$
$A \to B$
$AB \to C$

Repeat

Rewrite

$A \to B$ ✓
$A \to C$ ✗
$B \to C$ ✓
$AB \to C$ ✗

Now Check one by one each FD

$A \to B$ find Closure of, $(A^+) = \{A B C\}$] Not equal

Now excluding this FD find $A^+ = \{A C\}$

∴ it is essential.

$A \to C$ find $A^+ = \{A B C\}$, $(A^+)' = \{AB C\}$ equal

∴ it is not essential

$B \to C$ find $(B^+) = \{B C\}$ $(B^+)' = \{B\}$ ∴ Not equal essential

$AB \to C$ find $(AB)^+ = (A, B C\}$] both are equal
$(AB^+)' = \{A B C\}$

∴ not essential.

$A \to B$
$B \to C$ $\} = F_c$

For removing extraneous Attribute on LHS of FD then find the closure of LHS of

eg $AB \rightarrow C$ $(AB)^+ = ABC$

Now find the closure of individual attribute on L.H.S

$(A^+) = \{ABC\}$

$(B^+\} = \{AB\}$

So in above case, if closure of (AB) w/o B i.e (A^+) is equivalent closure $(AB)^+$ then B is extraneous.

So in above case if closure of AB w/o A i.e $(B^+) = (A, B)$ which is not equivalent to (AB^+) ∴ A is essential Attribute on LHS.

Keys

Super key

Keys

Super Key : It is a key which can uniquely identify a row in a table. It may consist of any no. of attributes.

Candidate Key (minimal superkey)

A
B
C
AB
BC
CD
DA
AC

R(A, B, C, D)

	SK	Candidate key	PK
A → BCD	✓	✓	
AB → CD	✓	×	
ABC → D	✓	×	
BD → AB	✓	✓	
C → AD	×	All the attribute are not included in closure so it is not super key.	

If proper subset of Superkey is not super key then you are candidate key.

If there is not proper subset of super key exist then that is candidate key

Primary Key : Anyone among the set of candidate key are may be used as Primary key.

R(A B C D E F G H)

$AB \to C$

$A \to DE$

$B \to F$

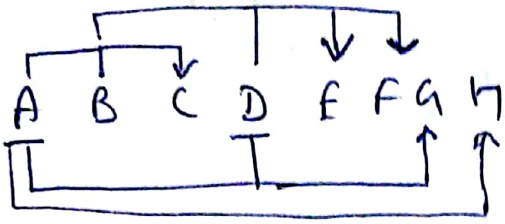
$F \to GH$

$(AB)^+ = (ABCDEFGH)$ ✓

$(B^+) = (BFGH)$ × $(A^+) = \{ADE\}$ ×

(2) R(A B C D E F G H)

AB → C
BD → EF
AD → G
A → H



Find the attributes which have no incoming edge. which mean are essential attributes and whatever be the Candidate key must contain ABD.

Step 1. ABD

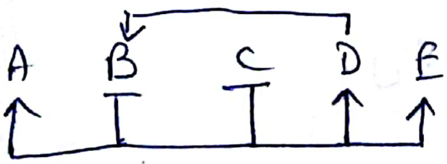
Step 2. $(ABD)^+ = \{A, B, C, D, E, F, G, H\}$

∴ ABD is candidate

It is the only candidate key because whatever be the candidate must be minimal and if an attribute is added to it then it becomes superkey not candidate key

(3) R(A B C D E)

BC → A DE
D → B



Step 1 :: C

Step 2 : $(C)^+ = \{C\}$ C alone can't find all the attributes

∴ C must be present in any of the candidate key found

$(AC)^+ = \{AC\}$ ✗
$(BC)^+ = (BCADE)$ ✓
$(CD)^+ = \{CD, B, A E\}$ ✓
$(CE)^+ = \{C, E\}$ ✗

for eg AC, BC, CD, CE

Any further combination $(CAE)^+ = \{ACE\}$ ✗

7

R(A B C D E)

AB → CD

D → A

BC → DE

In this case only 'B' is the attribute with no incoming edge

Step1 $(B)^+ =$ B   B itself can't find all the attribute ∴ check it with combination

AB, BC, BD, BE

$(AB)^+ =$ (A B C D E) ✓

$(BC)^+ =$ (A B C D E) ✓

$(BD)^+ =$ (A.B C D E) ✓

$(BE)^+ =$ (B E) X

ABE cannot be candidate key because AB is already candidate key

5

R(W X Y Z)

Z → W

Y → XZ

WX → Y

Attribute w/o any incoming edge. In this case no attribute is found. So check every possible combination

$(W)^+ =$ {W} X

$(X)^+ =$ {X} X

$(Y)^+ =$ {XYZW} ✓

$(Z)^+ =$ {ZW} X

$(WX)^+ \rightarrow$ {WXYZ} ✓

XZ = {XWZY} ✓

WZ = {WZ} X

6

R(A B C D E F G H)

CH → G

A → BC

B → CFH

E → A

F → EG

(a) $(DF)^+ =$ DF

2
3
(4) Candidate key
5

R (A B C D E F G H)

CH → G
A → BC
B → CFH
E → A
A → EG

(D^+) → D This is essential attribute

$(AD)^+$ = {ABCDEF GH} ✓
$(BD)^+$ = {B D C F H G} x
$(CD)^+$ = {C D} x
$(ED)^+$ = {A E D B C, C F G} ✓
$(FD)^+$ = {F D} x
$(GD)^+$ = {G D} x
$(HD)^+$ = {H D} x

$(BCD)^+$ = { B C D, F H G } x

$(BFD)^+$ = { B, F, D C H G } x
$(BGD)^+$ = { B, G, D, C F H } x
$(BHD)^+$ = { B H D C F G } x
$(CFD)^+$ = {C F D} x
$(GCD)^+$ = {G C D} x
$(HCD)^+$ = { H C D G } ✓
$(GFD)^+$ = { G F D } x
$(HFD)^+$ = { H F D } x
$(GHD)^+$ {G H D} x

BCFD B
BCGD
BCHD
BCFD