Module-5
Relational Database Design

Prof. Suja Jayachandran
Assistant Professor
Department of Computer Engineering
Vidyalankar Institute of Technology,Mumbai

Functional Dependency - For relation R, if there is set of attributes α and β $\alpha \subseteq R$, $\beta \subseteq R$ such that there is FDependency

$\alpha \rightarrow \beta$ if $t_1[\alpha] = t_2[\alpha]$

Then $t_1[\beta] = t_2[\beta]$

$\alpha \longrightarrow \beta$

determinant dependent

R

	α	β
$t_1 \rightarrow$	a	b
$t_2 \rightarrow$	a	b

If value of α is Distinct Then $\alpha \rightarrow \beta$ FD is always true

1

α	β
a	1
b	2
c	3
d	4

$\alpha \rightarrow \beta$?

If we have value of α, then we can find β

$\alpha = a$ $\beta = 1$

So $\alpha \rightarrow \beta$ is True

2

α	β
a	1
a	2
c	3
d	4

$\alpha \rightarrow \beta$?

$\because$ multiple value of β for same α

$\therefore$ $\alpha \rightarrow \beta$ is not possible

3

α	β
a	1
b	1
c	3
d	4

1) $\alpha \rightarrow \beta$?

As multiple value of β is possible for distinct α

$a \rightarrow 1$

$b \rightarrow 1$

$\alpha \rightarrow \beta$ is True

2) $\beta \rightarrow \alpha$?

This is not possible

Functional Dependency $\boxed{\alpha \to \beta}$

Trivial FD	Non Trivial FD
$AB \to A$	$AB \to ABC$
1. If $\beta \subseteq \alpha$	1. $\beta \not\subseteq \alpha$
2. Its valid, but no new information is gained	2. Its valid, new information is gained
3. Not so important	3. Important

1

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

FD is correct →

✓ 1 $A \to BC$

✓ 2 $DE \to C$

✗ 3 $C \to DE$

✓ 4 $BC \to A$

2

X	Y	Z
1	4	2
1	5	3
1	6	3
3	2	2

✗ 1 $XY \to Z$ && $Z \to Y$

✓ 2 $YZ \to X$ && $Y \to Z$

✗ 3 $YZ \to X$ && $X \to Z$

✗ 4 $XZ \to Y$ && $Y \to Z$

3

A	B	C
1	2	4
3	5	4
3	7	2
1	4	2

✗ 1 $A \to B$ && $BC \to A$

✗ 2 $C \to B$ && $CA \to B$

✓ 3 $B \to C$ && $AB \to C$

✗ 4 $A \to C$ && $BC \to A$

Rough work

$C \to DE$ ✗
$3 \to (4,5)$
$3 \to (4,5)$
$3 \to (6,5)$ ✗

$A \to BC$
$a \to 2,3$
$2 \to a,3$
$a \to 2,3$
$a \to 2,3$

$BC \to A$ ✓
$(2,3) \to a$ ✓
$(a,3) \to 2$
$(2,3) \to a$ ✓
$(2,3) \to a$ ✓

$XY \to Z$ ✓
$(1,4) \to 2$
$(1,5) \to 3$
$(1,6) \to 3$
$(3,2) \to 2$

$Z \to Y$ ✗
$2 \to 4$
$3 \to 5$
$3 \to 6$

$DE \to C$ ✓
$(4,5) \to 3$
$(4,5) \to 3$
$(6,5) \to 3$
$(6,6) \to 3$

Armstrong Axioms of FD

Axioms is a stmt that is taken to be True & serve as a strong point for functional arguements. It is basically used to generate closure set (F^+)

Primary Rules

1. Reflexivity

If $Y \subseteq X$ Then $X \rightarrow Y$ ($ABC \rightarrow AB$)

2. Augmentation

If $X \rightarrow Y$
Then $XZ \rightarrow YZ$

3. Transitivity

If $X \rightarrow Y$ & $Y \rightarrow Z$
Then $X \rightarrow Z$

Secondary Rules

1. Union : If $X \rightarrow Y$ & $X \rightarrow Z$
Then $X \rightarrow YZ$

2. Decomposition : If $X \rightarrow YZ$
Then $X \rightarrow Y$ & $X \rightarrow Z$

3. Pseudo transitivity

If $X \rightarrow Y$ & $WY \rightarrow Z$
Then $WX \rightarrow Z$

4. Composition If $X \rightarrow Y$ & $Z \rightarrow W$
Then $XZ \rightarrow YW$

Closure Set of Attributes

$F = F_1 + F_2$

$A \to B$ (direct dependency)

$A \to C$ (Transitive dependency) indirect

Let $R(A, B, C)$

If $A \to B$

$B \to C$

Then $\underline{\underline{A \to C}}$

$F = A \to BC$

$F^+ = (A\ B\ C)$

① If $R(A, B, C)$

$A \to B$

$B \to C$

Then $A^+ = (A, B, C)$

$B^+ = (B, C)$

$C^+ = (C)$

$A \to B$
$B \to C$
$A \to C$

② $R(A\ B\ C\ D\ E)$

$A \to BC$

$CD \to E$

$B \to D$

$E \to A$

$B^+ = (B, D)$

③ $R(A\ B\ C\ D\ E\ F\ G)$

$A \to B$

$BC \to DE$

$AEG \to G$

$(AC)^+ = (A, C, B, D, E)$

4. $R(ABCDEF)$

$AB \rightarrow C$

$BC \rightarrow AD$

$D \rightarrow E$

$CF \rightarrow B$

$(AB)^+ = (ABCDE)$

5. $R(ABCDEFGH)$

$A \rightarrow BC$

$CD \rightarrow E$

$E \rightarrow C$

$D \rightarrow AEH$

$ABH \rightarrow BD$

$DH \rightarrow BC$

$BCD \rightarrow H$ is true?

$\rightarrow$ True

$(AB)^+ = (ABC, \cancel{A}D, E)$

$AB \rightarrow C$

$BC \rightarrow AD$

$D \rightarrow E$

$(BCD)^+ = (B, C, D, E, A, H)$

$CD \rightarrow E$

$D \rightarrow AEH$

$(BCD) \rightarrow B$

$(BCD) \rightarrow C$

$(BCD) \rightarrow D$

$(BCD) \rightarrow E$

$(BCD) \rightarrow A$

$(BCD) \rightarrow H$ ✓

Equivalence of FD

GATE

① $R(A\ C\ D\ E\ H)$

$F:\ A \to C$
$AC \to D$
$E \to AD$
$E \to H$

$G:\ A \to CD$
$E \to AH$

a) $F \subseteq G$
b) $G \subseteq F$
✓c) $F = G$
d) $F \neq G$

For G:

$(A)^+ = (ACD)$
$(AC)^+ = (ACD)$
$(E)^+ = (EAHCD)$

$\therefore\ F \subseteq G$

For F

$(A)^+ = (ACD)$
$(E)^+ = (EADHC)$

$\therefore\ G \subseteq F$

So $\boxed{F = G}$

Compute all left side of FD's of G using F & vice versa.

② $R(PQRS)$

$X: P \to Q$
$Q \to R$
$R \to S$

$Y: P \to QR$
$R \to S$

a) $X \subseteq Y$
✓ b) $Y \subseteq X$
c) $X = Y$
d) $X \neq Y$

④ $R(ABC)$

$F: A \to B$
$B \to C$
$C \to A$

$G: A \to BC$
$B \to A$
$C \to A$

$F = G$ ✓

Soln

$X: P^+ = (PQRS)$
$Q^+ = (Q)$
$R^+ = (R, S)$

$Y: P^+ = (P, Q, R, S)$
$R^+ = (R, S)$

Soln F:
$A^+ = (ABC)$
$B^+ = (BAC)$
$C^+ = (ABC)$

$G: A^+ = (ABC)$
$B^+ = (BCA)$
$C^+ = (CAB)$

③ $R(VWXYZ)$

$F: W \to X$
$WX \to Y$
$Z \to WY$
$Z \to V$

$G: W \to XY$
$Z \to WX$

$F: (W)^+ = (WXY)$
$(WX)^+ = (WXY)$
$(Z)^+ = (ZWXY)$
$(Z)^+ = (ZWXY)$

$G: (W)^+ = (WXY)$
$(Z)^+ = (ZWYVX)$

✓ $G \neq F$ ✓

$G \subseteq F$ ✗

Irreducibility Set of FD (Cannonical form)

R(W X Y Z)

$X \to W$
$WZ \to XY$
$Y \to WXZ$

Decompose it $\longrightarrow$ (only RHS)

$X \to W$ — (1)
$WZ \to X$ — (2)
$WZ \to Y$ — (3)
$Y \to W$ — (4)
$Y \to X$ — (5)
$Y \to Z$ — (6)

$$\boxed{\begin{array}{c} X \to W \\ WZ \to Y \\ Y \to Z \end{array}}$$

Cannonical/Minimal form of R(W X Y Z)

(1) Now ignore $X \to W$ & compute X^+
$\therefore X^+ = (X)$ $\therefore$ $X \to W$ is important

(2) $(WZ)^+ = (WZYX)$ $\therefore$ $WZ \to X$ is not important

(3) $(WZ)^+ = (WZX)$ $\therefore$ $WZ \to Y$ is important

(4) $(Y)^+ = (YXZW)$ $\therefore$ $Y \to W$ is not important

(5) $(Y)^+ = (YWZX)$ $\therefore$ $Y \to X$ is not important

(6) $(Y)^+ = (YWX)$ $\therefore$ $Y \to Z$ is important

Keys Used in DBMS (Primary, Foreign, Candidate & super key)

R:

A	B	C
1	a	p
2	b	q
3	c	q
4	c	r

Key $\boxed{A \to BC}$ (given)

$A^+ = (A, B, C)$

A should be unique

∴ Key is A

Key is set of attribute which uniquely identify all tuples of R

① R(A B C D)

$ABC \to D$

$AB \to CD$

$A \to BCD$

Sol^n $(ABC)^+ = (ABCD)$

∴ $(ABC)^+$ is SK

$(AB)^+ = (ABCD)$

$(AB)^+$ is also SK

$(A)^+ = (ABCD)$

$(A)^+$ is SK

∴ $(A)^+$ is CK

Superkey - Set of attributes using which you can find value of all other attributes

$(\text{superkey})^+ = R$

Candidate Key - Minimal Superkey

Primary Key - Any one candidate Key can be primary key

(2) $R(ABCD)$

$B \rightarrow ACD$

$ACD \rightarrow B$

Sol^n

$(B)^+ = (BACD)$

$(ACD)^+ = (ACDB)$

Both $(B)^+$ & $(ACD)^+$ are SK

$(B)^+$ can be CK

$(B)^+$ can be PK

(3) $R(ABCD)$

$AB \rightarrow C$

$C \rightarrow BD$

$D \rightarrow A$

Sol^n

$(AB)^+ = (ABCD)$

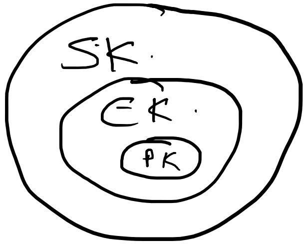
$(C)^+ = (CBDA)$

$(D)^+ = (DA)$

SK = $(AB)^+, (C)^+$

CK = $(C)^+$

PK = $(C)^+$



(4) $R(ABCD)$

$A \rightarrow BC$

Sol^n $(A)^+ = (ABC)$

As D is not reachable from A $\therefore$ A cannot be key.

How to identify a CK?

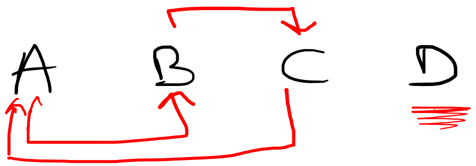
① $R(ABCD)$

$A \to B$

$B \to C$

$C \to A$

Sol^n



$(AD)^+ = (ADBC)$

$(BD)^+ = (BDCA)$

$(CD)^+ = (CDAB)$

$\therefore (AD)^+, (BD)^+, (CD)^+$ are SK & CK

① Find attribute who doesn't have an incoming edge

② So D is essential attribute as it cannot be reached from any other attribute

$(D)^+$ = No attribute

② R(A B C D)

AB → CD

D → A

Soln

A B C D

∴ B is essential attribute

$(AB)^+ = (ABCD)$

$(CB)^+ = (BC)$

$(DB)^+ = (DBAC)$

$SK = (AB)^+, (BD)^+$

$CK = (AB)^+, (BD)^+$

③ R(A B C D E)

AB → CD

D → A

BC → DE

Soln

A B C D E

B is essential att.

$(AB)^+ = (ABCDE)$

$(CB)^+ = (BCDEA)$

$(DB)^+ = (BDACE)$

$(EB)^+ = (EB)$

$SK = (AB)^+, (BC)^+, (BD)^+$

$CK = (AB)^+, (BC)^+,$

$(BD)^+$

④ R(A B C D E F)

AB → C

C → D

B → AE

Soln

A B C D E F

BF are essential

$(BF)^+ = (BFAECD)$

$(ABF)^+ = (ABFECD)$

$(CBF)^+ = (ABCDEF)$

$(DBF)^+ = (ABCDEF)$

$(EBF)^+ = (ABCDEF)$

$CK = (BF)^+$

$(ABF)^+, (BF)^+, (CBF)^+,$

$(DBF)^+, (EBF)^+ = SK$

⑤ R(AB CD)

AB → CD

C → A

D → B

Soln

A B C D

All attribute has incoming arrow

$(A)^+ = A$

$(B)^+ = B$

$(C)^+ = (C, A)$

$(D)^+ = (DB)$

$(AB)^+ = (ABCD)$ ✓

$(AC)^+ = (AC)$

$(AD)^+ = (ADBC)$ ✓

$(BC)^+ = (BCAD)$ ✓

$(BD)^+ = (BD)$

$(CD)^+ = (CDAB)$ ✓

$(ABC)^+$

$(ABD)^+$

$(BCD)^+$

$(ACD)^+$

S.K.

$CK = (AB)^+, (AD)^+, (BC)^+, (CD)^+$

SK = —, —, S.K

(Normalization) – FD, CK, Lossless & Lossy Decomposition, Dependency Preserving Decomposition

Lossless Decomposition

1. If R is decomposed into R1 & R2 Then union of attributes of R1 & R2 must be equal to attribute of R

$$\boxed{\text{Attribute}(R1) \cup \text{Attribute}(R2) = \text{Attribute}(R)}$$

2. Intersection of attribute of R1, R2 must not be null

$$\boxed{\text{Attribute}(R1) \cap \text{Attribute}(R2) \neq \phi}$$

3. Common Attribute must be a Key for atleast R1 or R2

$$\text{Attribute}(R1) \cap \text{Attribute}(R2) \rightarrow \text{Att}(R1)$$

or

$$\text{Att}(R1) \cap \text{Attribute}(R2) \rightarrow \text{Att}(R2)$$

Eg

Student Detail

Sid	SName	SAdd	SPhoneNo	Marks
1	Ajay	M		
2	Vijay	P		
3	Ram	D		

R1

Sid	SName	Marks
1		
2		
3		

R2

Sid	SAdd	SPhoneNo
1		
2	M	
3		

1. Marks, Sid, SName, SAdd, SPhone No.
2. Sid
3. Sid = Key

So this decomposition is Lossless Decomposition

NOTE – If any rule of lossless is not followed then we term the decomposition as Lossy.

Dependency Preserving Decomposition

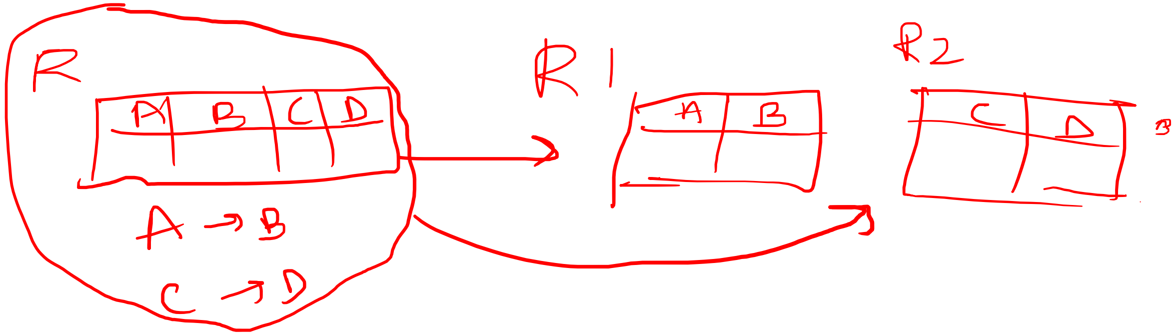
If we decompose a relation R into R1 & R2, All F.dependencies of R must be a part of R1 & R2 or must be derivable from combination of FDs of R1 & R2

Eg: R(ABCD)

$A \to BC$, $A \to D$

Sol^n R(ABCD)

R1(ABC) R2(CAD)



Eg: GATE 2005

R(ABCD)

$A \to B$ and $C \to D$

R1(AB) & R2(CD) is

a) Dependency preserving & lossless

b) Lossless but not dependency preserving

c) dependency preserving & lossy

d) not dependency preserving & lossy

Sol^n Here $A \to B$: R1(AB)

$C \to D$: R2(CD)

∴ It is Dependency Preserving Decomposition

(1) Attr(R1) ∪ Attr(R2) = Attr(R) ✓

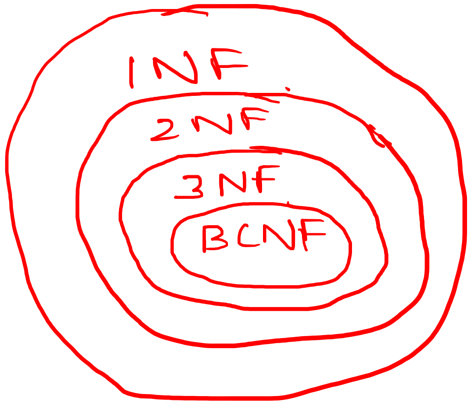
(AB) ∪ (C,D) = ABCD

(2) Att(R1) ∩ Attr(R2) ≠ Φ ✗

(A,B) ∩ (C,D) = Φ

∴ It is lossy

Normalization - To reduce data redundancy. Avoids insert anomalies delete anomalies & update anomalies. Brings the database to a consistent state.



Characteristics

1. Minimal use of Null value
2. Absence of Redundancy
3. Minimal loss of information.

2NF, 3NF, BCNF based on keys & FD of Relation

4NF based on key & multivalued dependency.

* 1NF, 2NF, 3NF, BCNF

① 1NF → All attributes in a relation (table) must have atomic value
→ No multivalued attributes
→ every column shd have unique name

Eg

Roll No.	Name	Course
101	Ajay	DBMS, WDL
102	Vijay	SPCC

$\xrightarrow{\text{1NF}}$

Roll No.	Name	Course
101	Ajay	DBMS
101	Ajay	WDL
102	Vijay	SPCC

* Prime Attribute — An attribute which is part of Primary Key/CK)
* Non Prime Attribute - An attribute which is not part of PK./CK

2NF

→ The Relation must be in 1NF
→ All non key attribute (Non Prime Attribute) must be functionally dependent on Primary key/Candidate Key

Eg

R (A B C D)

$AB \to D$

$B \to C$

ABCD

A	B	C	D

$\because AB \to D$ ✓

$\because B \to C$ = Partial Dependency

(when non-prime attribute instead of depending on CK, it depends on part of CK)

Then $(AB)^+ = (ABDC)$

$\therefore$ (AB) is candidate key

Prime Attribute = A, B

Non Prime Attribute = C, D

$\therefore$ It is not in 2NF

R(ABCD)

R1(ABD)

$\because$ AB is CK

$AB \to D$

(2NF)

R2(BC)

$\because B \to C$

$\therefore$ B is Prime Attribute (CK)

(2NF)

(2)

Sid	S_Name	P_id	P_Name	Grade
1	Ajay	P3	Ramesh	A
2	Vijay	P2	Mahesh	B
3	Ram	P1	Suresh	C

Sid → SName

Pid → PName

Sid, Pid → Grade

Sol^n

Sid	Pid	Grade
1	P3	A
2	P2	B
3	P1	C

Sid	SName
1	Ajay
2	Vijay
3	Ram

2NF

Pid	PName
P3	Ramesh
P2	Mahesh
P1	Suresh

(3) R(city, Street, HNo., Hcolor, Population)

Key : {city, street, HNo.}

{city, street, HNo.} → {Hcolor} ✓

{city} → {population}

Sol^n P → NP

P = {city, Street, HNo.}

NP = {Hcolor, Population}

New Schema is in 2NF

(1) {city, street, HNo., Hcolor}

(2) {city, population}

NP → NP ✗

P → NP ✓

3NF → Relation must be 2NF
→ There should be no transitive Dependency
→ For each FD $X \to Y$ atleast one of condition sld hold true
[X is superkey of table
Y is prime attribute of table

Eg R(empid, ename, ezipcode, estate, ecity, edistrict)
S.K. = {empid}, {empid, ename}
CK = {empid}

ezipcode → estate, ecity, edistrict
empid → ezipcode

2NF R1(emp-id, ename, ezipcode)
R2(ezipcode, estate, ecity, edistrict)
3NF

2NF: $\underline{X} \to \underline{Y}$
CK NP

T.D.
If $X \to Y$
$Y \to Z$
Then
$X \to Z$

$X \to Y$
SK or Prime attribute

2) Book Detail

Book id	Gener id	GenerType	Price
1	1	Science	300
2	2	Maths	350
3	1	Science	250
4	3	Physics	200
5	2	Maths	300

Book id → Gener id ✓

Generid → Genertype ×

CK : Book_id

↓ 2NF

Key

Book id	Gener id	Price
1	1	300
2	2	350
3	1	250
4	3	200
5	2	300

Will act as key attribute

Generid	GenerType
1	Science
2	Maths
3	Physics

Bookid → Generid

Bookid → Price

eliminating redundant data

Generid → GenerType

Now it is 3NF

BCNF : Boyce Codd Normal form or 3.5 NF

- Relation must be in 3NF
- For every FDs X → Y , X should be SUPER KEY

Eg :

S_Name	Course	P_Name
Priya	DBMS	Ajay
Raja	DBMS	Vijay
Amit	SPCC	Hari
Sumit	SPCC	Hari
Rohan	DBMS	Vijay

→

S_Name	Course
Priya	DBMS
Raja	DBMS
Amit	SPCC
Sumit	SPCC
Rohan	DBMS

Course	PName
DBMS	Ajay
DBMS	Vijay
SPCC	Hari

Key (SK) = {sname, course}

(sname, course) → PName

How to identify Normal form

1) First check for BCNF ($X \to Y$: X must be super key)
2) Next check for 3NF ($X \to Y$: either X shd be super key or Y shd be Prime Attribute
3) Next check for 2NF ($X \to Y$: Y shd be Non Prime & Y shd be fully functional dependent on candidate key (X))
4) Next check for 1NF (atomic value)

Eg : R(ABCDE)

$CE \to D$
$D \to B$
$C \to A$

Ans
It may be in 1NF

$A\ B\ C\ D\ E$

$(CE)^+ = (CEDBA)$

$\therefore$ CE is CK

Step 1

$\because D \to B, C \to A$ is not true for BCNF $\therefore$ given R is not BCNF

$\because$ B & A are not prime attribute $\therefore D \to B, C \to A$ is not true for 3NF, $\therefore$ given R is not in 3NF

$\because D \to B$ is not true for 2NF & $C \to A$ is partial dependent on CK
$\therefore$ R is not in 2NF

Example for Practice

1) $R(ABCDEF)$

$AB \rightarrow C$

$DC \rightarrow AE$

$E \rightarrow F$

2) $R(ABCDE)$

$AB \rightarrow CD$

$D \rightarrow A$

$BC \rightarrow DE$

3) $R(ABCDE)$

$BC \rightarrow ADE$

$D \rightarrow B$