#### **UNIT-IV**

### RELATIONAL DATABASE DESIGN

## **Features of Good Database Design:**

- > A Good database contains consistent data and there is no chance for inconsistency.
- > It also reduce the data redundency and there is no chance for duplication of data.
- > There is no wastage of space.
- > Uniform of data.

## **NORMALIZATION:**

> It is a process of organising the data in a database and also avoid the data redundency, insertion, deletion, updation anomaly.

#### **ANOMOLIES IN DATABASE:**

- → Definition: It is a possibly of certain data getting inconsistent.
- → There are 3 types of anomolies in database:
  - (1)Update Anomoly.
  - (2) Deletion Anomoly.
  - (3)Insert Anomoly.

## **Employee table**

eid	name	address	department
1	abc	guntur	cse
1	abc	guntur	ece
2	def	vizag	mec
3	xyz	vijayawada	eee
3	xyz	vijayawada	ce

## (1)Update Anomoly:

It is a data inconsistency and there is a partial update. **Example:** 

In above table,2 rows for employee "abc" and he belongs to 2 department. If we want to update the address of "abc", data will become inconsistent and there is a partial update and correct address can't be updated.

## (2)Deletion Anomoly:

It means loss of data due to deletion of other data. **Example:** 

Suppose an institute close the department "mec", the information of employee lost.

## (3)Insert Anomoly:

> It means inability to add the database due to absence of other data.

## **Example:**

Suppose new employee's are joining in an institute, those who are under training and currently not assigning to any department then we couldn't able to insert the data into table and it doesn't allow null values.

> To overcome these anomolies we need to use "**Normalization**" forms.

## **Functional Dependency(FD):**

- ➤ Definition: Association among attributes is known as "Functional Dependency" and it is represented as ( — )
- Let us consider a relation 'R' and A,B are two non-empty attributes in a relation 'R' then FD is represented as A → B. Where 'A' functionally determines B.So, 'A' is called "Determinant".
- > The Lefthandside(LHS) attributes determines the values of attributes on the rightside.
- > A primary key is a special case of FD.
- ➤ A Functional Dependency "A B" says that if two tuples agree on the values in the attribute 'A' they also agree on the values in attributes 'B'.

$$t_1.A = t_2.A$$
 then  $t_1.B = t_2.B$ 

here, A,B are two attributes and  $t_1$ ,  $t_2$  are tuples.

A FD is single to single(A → B), single to many(A → BC), many to single(BC → A), many to many(AB → CD).

**Example:** 

A	В	C	D
a1	b1	c1	d1
a1	b1	c1	d2
a1	b2	c2	d1
a2	b1	<b>c</b> 3	d1

#### Find the following FD's Satisfied are Not Satisfied ???

- (a)A -> B : Not satisfied..
- (b)B -> C : Not satisfied..
- (c)C -> D : Not satisfied..
- $(d)D \rightarrow A : Not satisfied..$
- (e)A -> BC : Not satisfied..
- (f)AB -> C: Satisfied..
- (g)AC -> D : Not satisfied..

## **Types Of Functional Dependencies:**

There are 3 types:

# (1)Trivial Functional Dependencies(FD's):

If the right hand side attributes are completely subset of left hand side attributes.

## **Symbolically:**

if A -> B then B is subset to A.

The following are also trivial FD's that are,

$$A -> A, B -> B.$$

**Example:** Let us consider a student relation it contains attributes student\_id,student\_name. {student\_id,student\_id} is a FD's. Also,

s name -> s name

## (a)Non-Trivial Functional Dependencies:

If one of the right hand side attribute is not subset of left hand side attributes.

## **Example:**

A student relation with 3 attributes student\_id,student\_name,student\_address. The following are non-trivial.

Sid -> name sid -> address

## (i)Completely Non-Trivial FD's:

If a Functional Dependency A->B holds true where A intersection B is NULL.

A intersection B =/ 0 Here,RHS attributes are independent of LHS attributes.

## (2)TRANSITIVE FD's:

A FD is said to be transitive if A->B,B->C then A->C. Transitive Dependency is occured with 3 or more than 3 attributes and also form with 2 or more FD's.

## **Example:**

sid -> name, name -> age, sid -> age.

## (3)MULTIVALUED FD's:

If more than one independent attributes depends on particular attribute is known as "Multivalued FD's". for **example**, A,B,C are 3 attributes, the attributes B and C are independent to each other. but depends on attribute A.

$$A ->-> B$$

$$A \rightarrow -> -> C$$

# \*\*<u>Armstrong Axioms/Inference Rules/Properties of</u> FD's: \*\*

- > Armstrong axioms is a set of rules.
- > It was developed by william w.Armstrong in 1974.
- > These rules reduces the set of functional dependencies.

## (1)Reflexivity Rule:

It is same as Trivial FD's.If A,B are two nonempty set of attributes and B is subset of A.

# (2)Transitivity Rule:

 $A \rightarrow B$ 

B->C

 $A \rightarrow C$ 

## (3) Augumentation Rule:

If attribute A holds B and C is a particular attribute then AC holds BC i.e., AC->BC. It means that attribute C is doesn't change the basic dependencies(A-> B).

## (4)Union Rule:

If attribute A holds B and attribute A holds C then A holds BC.

A -> B, A -> C therefore, A->BC.

## (5)Decomposition Rule:

If attribute A holds BC then A holds C.

 $A \rightarrow BC$ 

A->B, A->C

# (6)Pseudo Transitivity Rule:

If attribute A holds B and BC holds D then AC holds D.

A->B

BC->D then

AC->D.

## **Example:**

#### **Closure set:**

→ A functional dependency is said to be closure if it cover all attributes or not based on given condition.

# **Algorithm or procedure:**

**step1:** Equivate an attribute or attributes for which closure need to be identified( $x^+=x$ )

**step2:**Take FD one by one and verify whether LHS is available in X,if so add RHS attributes to x,if it is not available.

**Step3:** Repeat step2 as many times as possible to cover attributes.

**Step4:**Stop procedure after no more attribute added to X and declare X as closure set of attributes.

```
EX: suppose a relation R with schema R(ABCDEGHK) with set of FD's F:A->B B->DE
```

find

$$a)A^+$$
  $b)B^+$   $c)E^+$ 

```
step1: A^{\dagger} = A
    step2:A^+=A
             =AB(A->B)
             =ABDE(B->DE)
             =ABDEGH(E->GH)
             =ABDEGHK(B->K)
             =ABDEGHK
therefore, A+ is closure set
  it is super key because it covered all attributes
b)B<sup>+</sup>
    step1: B<sup>+</sup>=B
    step2: B<sup>+</sup>=B
                  =BDE(B->DE)
                  =BDEGH(E->GH)
                  =BDEGHK(B->K)
                  = BDEGHK
B<sup>†</sup> is not superkey
because it does not covered all attributes(i.e, A)
c)E<sup>+</sup>
    step1: E^{+}=E
    step2: E^+ = E
                  =EGH(E->GH)
                  =EGH
```

E<sup>+</sup> is not superkey because it does not covered all attributes(i.e, A,B,C,D,K)

```
Ex2: consider the following FD's
        AB->CD
        AF->D
        DE->F
        C->G
        F->F
        G \rightarrow A
which following is incorrect?
                                   [C and D]
a)(CF)^{+}=\{A,C,D,E,F,G\}
b)(BG)^{+}=\{A,B,C,D,G\}
c)(AF)^{+} = \{A,C,D,E,F,G\}
d)(AB)^{+} = \{A,C,D,F,G\}
a) (CF)^{+} = \{A,C,D,E,F,G\}
    (CF)^+ = CFE(F -> E)
        =CFEG(C->G)
        =CFEGA(G->A)
        =CFEGAD(AF->D)
                                   (correct)
        =CFEGAD
b)(BG)<sup>+</sup>={A,B,C,D,G}
    (BG)+=BG
        =BGA(G->A)
        =BGACD(AB->CD)
                                    (correct)
c)(AF)<sup>+</sup>=AF
        =AFE(F->E)
```

**DBMS** 

```
=AFED(AF->D) (incorrect)
\frac{d}{(AB)}^{+} = \{A,C,D,F,G\}
        AB^{+}=AB
            =ABCD(AB->CD)
            =ABCDG(C->G) (incorrect)
Ans: c and d
Ex:3 In a shema with attributes A,B,C,D,E and set of FD's
        A \rightarrow B
are
        A->C
        CD->E
        B->D
        E -> A
Which of the following FD's not implied by the above set
a)CD->AC b)BD->CD c)BC->CD d)AC->BC
                                             [B]
a)CD->AC
    (CD)^{\dagger} = CD
          =CDE (CD->E)
         =CDEA (E->A)
        =CDEAB (A->B)
b)BD->CD
    (BD)<sup>+</sup>=BD (it not contain CD)
c)BC->CD
    (BC)^{\dagger} = BC
        =BCD (B->D)
```

```
d)AC->BC
    (AC)^{+}=AC
         =ACB (A->B)
EX-4: Consider the relational schema R(EFGHIJKLMN)
and set of
FD's are
        EF->G
        F->IJ
        EH->KL
        K->M
        L > N
what is key for relation R
                                      \mathbf{B}
a){E,F}
b)\{E,F,H\}
c){E,F,H,K,L}
d)E
a)(EF)^+=EF
      =EFG(EF->G)
       =EFGIJ(F->IJ)
b)(EFH)<sup>+</sup>=EFH
       =EFHG
        =EFHGIJ
        =EFHGIJKL
        =EFHGIJKLMN
c)(EFHKL)<sup>+</sup>=EFHKL
            =EFHKLMN
            =EFHKLMNG
```

**DBMS** 

$$d)E^+=E$$

#### ans: b

EX:5 R(ABCDEH) what are the candidate keys for [ d ]

A->B

BC->D

E->C

D->A

### a)AE,BE

### b)AE,BE,DE

#### c)AEH,BEH,BCH

#### d)AEH,BEH,DEH

## **Equivalence of FD's:**

A two sets of functional dependencies F1 and F2 are equivalent if  $(F1^+=F2^+)$ 

**EX:** Find equivalence of FD's

# EX-2: R(ABCDEH) F1: $A \rightarrow C$ AC->D B->D

E->AH
$$A^{+}=ACD$$

$$E^{+}=EAH$$

=EAHCD

 $A \rightarrow CD$ 

**F2:** 

A->CD E->AH

F2:

$$A^+=A$$
 $=ACD$ 

#### .. F1!=F2

#### <u>Irreducible set/canonical cover/Minimal cover of FD's:</u>

It is set of FD's with a simplified version of FD's with a simplified version of FD's that is closure of particular attributes, if it cover all attributes that can be redundant otherwise it is irreducible or canonical.

#### **Procedure:**

**Step1:**If we have multiple attributes on RHS,if necessary FD use decomposition rule.

**Step2:**Removal of redundant FD that means to check whether Fd is redundant or not ,first hode that Fd and find closure of left in that Fd.If closure contains all attributes then Fd is redundant and remove that Fd from the set.

**Step3:**Look for Fd having more than one attribute on the LHS remove an attribute from the FD and take the closure of remaining attribute other FD.If removed attribute exists

 $E^{+}=EAH$ 

=FAHCD

in closure then remove it from FD.Repeat this process for remaining attributes.

## Step4:apply union rule

```
Ex:R(ABCD)
    A \rightarrow B
    C->B
    D->ABC
    AC->D
Ans:
Step1: using decomposition rule
    A \rightarrow B
    C->B
    D->A
    D->B
    D->C
    AC->D
step2: Finding redundant FD and remove it
    1.A->B
               hide 1, closure of A = > A^+ = A
    2.C->B
3.D->A
               hide 2,closure of C=>C+=C
                hide 3,closure of D=>D+=DBC
                hide 4,closure of D=>D+=DACB
    4.D->B
                hide 5,closure of D=>D+=DAB
    5.D->C
                   hide 6,closure of AC=>AC+=ACB
    6.AC->D
remove 4 as it covered all attributes
step3:
     AC->D
Remove A, closure C=CB
```

```
remove C,closure A=AB
as A and C doesnot exist we can't remove that FD
step4: By applying union,the resultant minimal set is
A->B
C->B
D->AC
AC->D
```

#### **ASSIGNMENT:**

```
1)R(WXYZ)
X->W
WZ->XY
Y->WXZ
2)R(ABCDE)
A->B
AB->C
D->AC
D->E
3)R(ABCDE)
A->BC
CD->E
B->D
E->A
```

4)R(ABC)

A->BC

B->C

 $A \rightarrow B$ 

AB->C

## **Finding Candidate Keys:**

If any attribute that cannot appears on the righthandside(RHS) from the given schema that is called "candidate key".

```
Ex-1: R(ABCDE)
A -> C
E -> D
B -> C
```

 $(ABE)^+ = ABCDE = R,$ 

... ABE is a candidate key because it covered all attributes.

```
Ex-2: R(ABCDE)
A -> BE
C -> BE
B -> D
(AC)+ = ACBED = R,
```

... AC is a candidate key because it covered all attributes.

```
Ex-3: R(ABCD)
AB->C
BC->D
CD->A
Ans: (AB)+=ABCD
```

(BC)<sup>+</sup>=BCDA (CD)<sup>+</sup>=CDA, CD not covered all attributes. ∴ AB and BC are candidate keys..

## **Types Of Decomposition:**

- **→ Decomposition:** The process of divide a single relation into two or more subrelations is called "Decomposition".
- → No information is lost from original relation during decomposition.
- → When a subrelaton are joined back the same relation is obtained i.e., Decomposition
- → There are two types:
  - (1)Loss less join decomposition
  - (2) Dependency preserving decomposition.

## (1)Loss Less Join Decomposition:

**Step-1:** Union of R1,R2,..,Rn must be equal to R.Each attribute of R must be in R1,R2,...Rn i.e.,

Attributes(R1)**U**Attributes(R2)**U**....=Attributes(R) **Step-2:**Intersection of attributes of R1,R2,...Rn must not be NULL i.e.,

Attributes(R1) **n** Attributes(R2) **n** ....not NULL.

**Step-3:**Common attributes must be a key for atleast one of the relation of R1 (or) R2 (or)....Rn.

Attr(R1)  $\mathbf{n}$  Attr(R2)  $\mathbf{n}$ ..... $\mathbf{n}$  Attr(Rn) -> Attr(R1)

Attr(R1) **n** Attr(R2) **n...n**  $Attr(Rn) \rightarrow Attr(R1)$ 

**Ex-1:** A relation R(ABCD) is decomposed into R1(ABC) and R2(AD) with FD set is {A->BC}. **Sol:** 

step-1: ABC  $\mathbf{U}$  AD =ABCD

step-2: ABC  $\mathbf{n}$  AD = A

step-3: A is key for R1(ABC) because A->BC.

**Ex-2:** A relation R(ABCD) is decomposed into R1(AB),R2(AD) and R3(BD) with FD set is {A->BC,B->C,C->D,D->B}.

step-1: AB  $\mathbf{U}$  BC  $\mathbf{U}$  BD =ABCD

step-2: AB **n** BC **n** BD = B

step-3: B is key for R2(BC) because B->C.

(2) Dependency Preserving Decomposition:

If we decompose a relation R into R1,R2,..,Rn then the dependencies of R must be part of R1 (or) R2 (or) Rn (or) must be derived from combination of FD's R1 and R2, R2 and R3,.....

Ex-1: R(ABCD) and FD is A->BC that can be decomposed into R1(ABC) and R2(AD).

Sol: Here Functional Dependency A->BC is part of R1.So it is Dependency Preserving.

Ex-2:Consider a schema R(ABCD) and FD's are A->B, C->D then decomposition of R into R1(AB) and R2(CD). [C]
(A)Dependency preserving and Loss less join (B)Loss less join but not dependency preserving (C)Dependency preserving but not loss less join (D)Neither nor

Ex-3:R(ABCD) that can be decomposed into R1(AB),R2(BC),R3(BD) and set of FD's are {A->B, B->C,C->D,D->B} is [B]

(A)Dependency preserving and Loss less join (B)Loss less join but not dependency preserving (C)Dependency preserving but not loss less join (D)Neither nor

## **Key and Non-Key Attributes:**

Key attributes are part of primary key. Non-Key attributes are not part of primary key.

## **Partial and Trasitive Dependency:**

All Non-key attributes in a table should be totally depend on part of primary key is called as "partial dependency(PD)".

## **Ex-1:** R(ABCD) AB->C B->D

Sol: Key: (AB)<sup>+</sup> =ABCD AB-> C (not PD) B-> D (PD because D is a non-key and B is a part of primary key AB).

```
Ex-2: R(ABCDEFGHIJ)
AB->C (not PD)
A->DE (PD)
B->F (PD)
F->GH (not PD)
D->IJ (not PD)
```

**Sol: Key:** (AB)<sup>+</sup> = ABCDEFGHIJ

A->DE,B->F are partial dependencies.

Transitive Dependency:

If there is a relationship between Non-key attributes(both LHS & RHS) is known as "Transitive Dependency(TD)".

## **Ex-1:** R(ABCD)

AB->C

C->D

## Sol: Key:(AB)+=ABCD

AB->C (not TD)

C->D (TD because C,D are Non-keys)

### Ex-2: R(ABCDEFGHIJ)

AB->C (not TD)

A->DE (not TD)

B->F (not TD)

F->GH (**TD**)

D->IJ (TD because Both LHS & RHS are

non-keys)

## **TYPES OF NORMAL FORMS:**

- (1)1NF(First-Normal Form)
- (2)2NF(Second-Normal Form)
- (3)3NF(Third-Normal Form)
- (4)BCNF(Boyce- codd Normal Form)

## (1)1NF(First-Normal Form):

A table is said to be in first normal form then it satisfies following conditions:

- (i) Column name should be unique.
- (ii) Rows must be Unique.
- (iii) Each column suppose to have same domain value.
- (iv) No restriction on No.of rows.

(v) Intersection of rows and columns must be a single value.

Ex:

Student table

sid	name	course		
1	a	DBMS,OS		
2	b	JAVA,DAA		
3	С	IDA,DBMS		
4	d	DS		

Sol:

The above given table not in first normal form because different courses in a course column.

1NF student table

sid	name	course		
1	a	DBMS		
1	a	OS		
2	b	JAVA		
2	b	DAA		
3	С	IDA		
3	С	DBMS		
4	d	DS		

Now, the table in a First Normal Form.

## (2)2NF(Second-Normal Form):

**Example:** R(ABCDEFGHIJ)

A table is said to be in 2NF if it is already in 1NF and also free from PD.

#### Note:

If there is a PD in a table,remove the partial dependent attributes from the original table and place it in a seperate table along with the copy of determinant.

```
AB \rightarrow C
    A -> DE (PD)
    B \rightarrow F (PD)
    F -> GH
    D -> IJ
Sol:
            Key:AB
    R<sub>1</sub>: ADEIJ
    R2: BFGH
    R3: AB
Example-2: R(ABCDEFGHIJ)
    AB \rightarrow C (PD)
    BD -> EF (PD)
    AD ->GH (PD)
    A \rightarrow I (PD)
    H \rightarrow J
```

Sol: Key: ABD

R<sub>1</sub>: ABCI — R11: ABC

**R12: AI** 

R<sub>2</sub>: BDEF

R<sub>3</sub>: ADGHJ

R<sub>4</sub>: ABD

ABC AI BDEF ADGHJ ABD

## (3)3NF(Second-Normal Form):

A table is said to be in 3NF if it is already in 1NF and also free from Transitive Dependency.

#### Note:

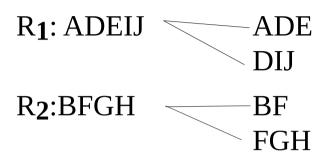
If there is a Transitive Dependency in a original table, remove those transitive dependency attributes from 2NF table and place it in a seperate table along with copy of determinant.

**Example:** R(ABCDEFGHIJ)

AB -> C

 $A \rightarrow DE (PD)$ 

Sol: Key: AB



**R3**: AB

Now the table in 3NF,

ADE DIJ BF FGH AB

# (4)BCNF(Boyce-Codd Normal Form):

A table is said to be in BCNF if it is already in 3NF and all determinants are keys.

**Example:** R(ABCDEFGHIJ)

$$\overrightarrow{AB} -> C$$

$$A \rightarrow DE (PD)$$

$$B \rightarrow F$$
 (PD)

$$D \rightarrow IJ$$
 (TD)

Sol: Key: AB

 $R_1$ : ADEIJ — ADE

DIJ

R<sub>2</sub>: BFGH — BF FGH

R<sub>3</sub>: AB

Now the table in 3NF,

ADE DIJ BF FGH AB

From the attributes, AB, A, B, F, D are keys.

**Note:** Every LHS attributes are Keys.