Normalization

Functional Dependency and Schema Refinement

How to find key?

- Conditions to find key
  - The attribute is a **part of key**, if it **does not occur on any side of FD**
  - The attribute is a **part of key**, if it **occurs on the left-hand side of an FD**, but **never occurs on the right-hand side**
  - The attribute is **not a part of key**, if it **occurs on the right-hand side of an FD**, but **never occurs on the left-hand side**
  - The attribute **may be a part of key or not**, if it **occurs on the both side of an FD**

How to find key? [Example]

THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY (Deemed to be University)

- Let a relation R with attributes ABCD with FDs C → A, B → C. Find keys for relation R.
  - attribute **not occur on any side** of FDs **(D) √**
  - attribute **occurs on only left-hand side** of an FDs **(B) √**
  - attribute **occurs on only right-hand side** of an FDs **(A)** X
  - attribute **occurs on both the sides** of an FDs **(C)** ?
- The **core is BD**.
- **B determines C** and **C determines A**, So using **transitivity rule B determines A** also.
- So **BD is a key**.

How to find key? [Exercise]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Let a relation R with attributes ABCD with FDs C → D, C → A and B → C. Find keys for relation R.
  - The core is B. B determines C which determines A and D, so **B is a key**. Therefore B is the key.

- Let a relation R with attributes ABCD with FDs B → C, D → A. Find keys for relation R.
  - The core is BD. B determines C and D determines A, so **BD is a key**. Therefore BD is the key.

- Let a relation R with attributes ABCD with FDs A → B, BC → D and A → C. Find keys for relation R.
  - The core is A. A determines B and C which determine D, so **A is a key**. Therefore A is the key.

Find (candidate) key [Example]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Suppose you are given a relation R with four attributes ABCD. For each of the following sets of FDs, do the following: **F = (B → C, D → A)**
  - Identify the candidate key(s) for R.

Candidate Key is BD

Find (candidate) key [Example]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Suppose you are given a relation R with four attributes ABCD. For each of the following sets of FDs, do the following: **F = (C → D, C → A, B → C)**
  - Identify the candidate key(s) for R.

Candidate Key is B

Find (candidate) key [Example]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Suppose you are given a relation R with four attributes ABCD. For each of the following sets of FDs, do the following: **F = (A → B, BC → D, A → C)**
  - Identify the candidate key(s) for R.

Candidate Key is A

Find (candidate) key [Example]

- Suppose you are given a relation R with four attributes ABCD. For each of the following sets of FDs, do the following: **F = (ABC → D, D → A)**
  - Identify the candidate key(s) for R.

Candidate Key are ABC & BCD

Functional Dependency (FD) and its types

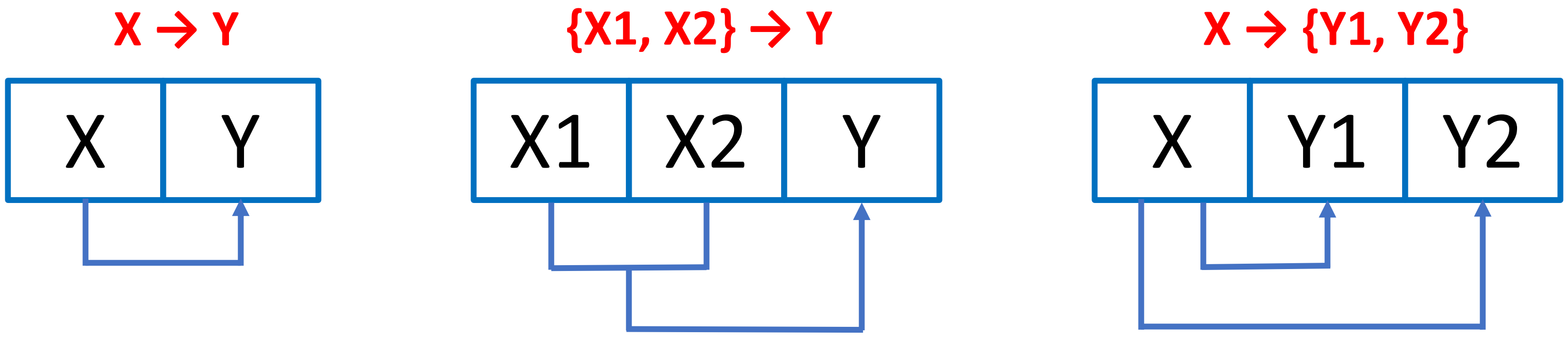
What is Functional Dependency (FD)?

- Let R be a relation schema having n attributes A1, A2, A3,..., An.
- Let attributes X and Y are two subsets of attributes of relation R.
- If the **values of the X component of a tuple uniquely** (or functionally) **determine the values of the Y component**, then there is a **functional dependency from X to Y**. This is denoted by **X → Y**.
- (i.e RollNo → Name, SPI, BL).
- It is referred as: **Y is functionally dependent on the X** or **X functionally determines Y**.

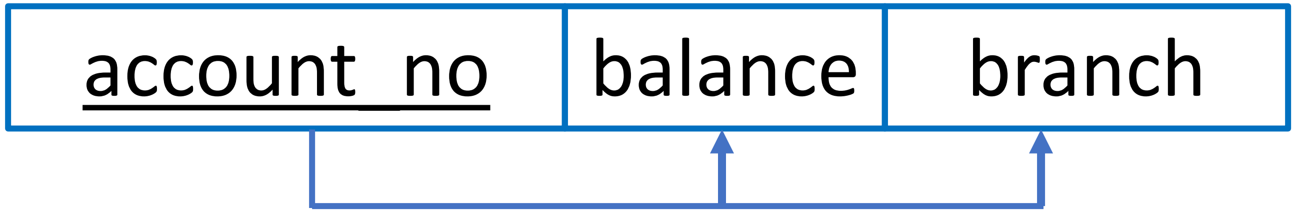
Student

RollNo	Name	SPI	BL
101	Raju	8	0
102	Mitesh	7	1
103	Jay	7	0

Diagrammatic representation of (FD)



- Example
- Consider the relation Account(account_no, balance, branch).
- account_no can determine balance and branch.
- So, there is a functional dependency from account_no to balance and branch.
- This can be denoted by account_no → {balance, branch}.



Functional Dependency

Types of Functional Dependency (FD)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Full Functional Dependency
  - In a relation, the attribute B is fully functional dependent on A if **B is functionally dependent on A, but not on any proper subset of A**.
  - Eg. {Roll_No, Semester, Department_Name} → SPI
  - We **need all three {Roll_No, Semester, Department_Name} to find SPI**.

Types of Functional Dependency (FD)

- Partial Functional Dependency
  - In a relation, the attribute B is partial functional dependent on A if **B is functionally dependent on A as well as on any proper subset of A**.
  - If there is some attribute that can be removed from A and the still dependency holds then it is partial functional dependancy.
  - Eg. {Enrollment_No, Department_Name} → SPI
  - **Enrollment_No is sufficient to find SPI**, Department_Name is not required to find SPI.

Types of Functional Dependency (FD)

- Transitive Functional Dependency
  - In a relation, if attribute(s) **A → B and B → C, then A → C (means C is transitively depends on A** via B).

Sub_Fac

Subject	**Faculty**	**Age**
DS	Shah	35
DBMS	Patel	32
DF	Shah	35

  - Eg. Subject → Faculty & Faculty → Age then Subject → Age
  - Therefore as per the rule of transitive dependency: Subject → Age should hold, that makes sense because if we know the subject name we can know the faculty's age.

Types of Functional Dependency (FD)

- Trivial Functional Dependency
  - X → Y is trivial FD if **Y is a subset of X**
  - Eg. {Roll_No, Department_Name, Semester} → Roll_No

- Nontrivial Functional Dependency
  - X → Y is nontrivial FD if **Y is not a subset of X**
  - Eg. {Roll_No, Department_Name, Semester} → Student_Name

Armstrong's axioms OR Inference rules

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Armstrong's axioms are a set of rules used to infer (derive) all the functional dependencies on a relational database.

Reflexivity
- If B is a subset of A
  - then A → B

Augmentation
- If A → B
  - then AC → BC

Self-determination
- If A → A

Transitivity
- If A → B and B → C
  - then A → C

Pseudo Transitivity
- If A → B and BD → C
  - then AD → C

Decomposition
- If A → BC
  - then A → B and A → C

Union
- If A → B and A → C
  - then A → BC

Composition
- If A → B and C → D
  - then AC → BD

Closure of a set of FDs

What is closure of a set of FDs?

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given a set F set of functional dependencies, there are certain other **functional dependencies that are logically implied by F**.
- E.g.: F = {A → B and B → C}, then we can infer that A → C (by transitivity rule)
- The set of **functional dependencies (FDs) that is logically implied by F** is called the closure of F.
- It is denoted by **F^+**.

Closure of a set of FDs [Example]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Suppose we are given a relation schema **R(A,B,C,G,H,I)** and the set of functional dependencies are:
  - **F = (A → B, A → C, CG → H, CG → I, B → H)**
- The functional dependency **A → H** is logical implied.

We have

A → B

B → H

Transitivity rule

A → H

Closure of a set of FDs [Example]

- Suppose we are given a relation schema **R(A,B,C,G,H,I)** and the set of functional dependencies are:
  - **F = (A → B, A → C, <u>CG → H</u>, <u>CG → I</u>, B → H)**
- The functional dependency **CG → HI** is logical implied.

We have

CG → H
CG → I

→ Union rule → CG → HI

Closure of a set of FDs [Example]

- Suppose we are given a relation schema **R(A,B,C,G,H,I)** and the set of functional dependencies are:
  - **F = (A → B, A → C, CG → H, CG → I, B → H)**
- The functional dependency **AG → I** is logical implied.

We have

A → C

CG → I

→ Pseudo-transitivity rule → **AG → I**

Closure of a set of FDs [Example]

- Suppose we are given a relation schema **R(A,B,C,G,H,I)** and the set of functional dependencies are:
  - **F = (A → B, A → C, CG → H, CG → I, B → H)**
- The functional dependency **AG → I** is logical implied.

We have

A → C → Augmentation rule → **AG → CG**

AG → CG
CG → I → Transitivity rule → **AG → I**

Closure of a set of FDs [Example]

- Suppose we are given a relation schema **R(A,B,C,G,H,I)** and the set of functional dependencies are:
  - **F = (A → B, A → C, CG → H, CG → I, B → H)**
- Find out the closure of F.

Several members of **F^+** are

F^+ = (A → H, CG → HI, AG → I)

Closure of a set of FDs [Example]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Compute the closure of the following set F of functional dependencies FDs for relational schema **R = (A,B,C,D,E,F):**
  - **F = (A → B, A → C, CD → E, CD → F, B → E)**
- Find out the closure of F.

A → B & A → C	Union Rule	**A → BC**
CD → E & CD → F	Union Rule	**CD → EF**
A → B & B → E	Transitivity Rule	**A → E**
A → C & CD → E	Pseudo-transitivity Rule	**AD → E**
A → C & CD → F	Pseudo-transitivity Rule	**AD → F**

F^+ = (A → BC, CD → EF, A → E, AD → E, AD → F)

Closure of a set of FDs [Example]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Compute the closure of the following set F of functional dependencies FDs for relational schema **R = (A,B,C,D,E):**
  - **F = (AB → C, D → AC, D → E)**
- Find out the closure of F.

D → AC	Decomposition Rule	**D → A & D → C**
D → AC & D → E	Union Rule	**D → ACE**

F^+ = (D → A, D → C, D → ACE)

What is an anomaly in database design?

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Anomalies are **problems that can occur in poorly planned, un-normalized database** where all the data are stored in one table.
- There are three types of anomalies that can arise in the database because of redundancy are
  - Insert anomaly
  - Delete anomaly
  - Update / Modification anomaly

Insert anomaly

- Consider a relation Emp_Dept(EID, Ename, City, DID, Dname, Manager) EID as a primary key

Emp_Dept

EID	Ename	City	DID	Dname	Manager
1	Raj	Rajkot	1	CE	Shah
2	Meet	Surat	1	CE	Shah
~~NULL~~	NULL	NULL	2	IT	NULL

An insert anomaly occurs when certain attributes cannot be inserted into the database without the presence of another attribute.

Want to insert new department detail (IT)

- Suppose a **new department (IT) has been started** by the organization but **initially there is no employee appointed** for that department.
- We **want to insert that department detail** in Emp_Dept table.
- But the **tuple for this department cannot be inserted** into this table as the **EID will have NULL value**, **which is not allowed because EID is primary key**.
- This kind of problem in the relation where some tuple cannot be inserted is known as insert anomaly.

Delete anomaly

- Consider a relation Emp_Dept(<u>EID</u>, Ename, City, DID, Dname, Manager) EID as a primary key

Emp_Dept

<u>EID</u>	**Ename**	**City**	**DID**	**Dname**	**Manager**
1	Raj	Rajkot	1	CE	Shah
2	Meet	Surat	1	CE	Shah
3	Jay	Baroda	2	IT	Dave

A delete anomaly exists when **certain attributes are lost because of the deletion of another attribute**.

Want to delete (Jay) employee's detail

- Now consider **there is only one employee in some department (IT)** and that **employee leaves the organization**.
- So we **need to delete tuple of that employee (Jay).**
- But in addition to that **information about the department also deleted**.
- This kind of problem in the relation where deletion of some tuples can lead to loss of some other data not intended to be removed is known as delete anomaly.

Update anomaly

- Consider a relation Emp_Dept(EID, Ename, City, Dname, Manager) EID as a primary key

Emp_Dept

EID	**Ename**	**City**	**Dname**	**Manager**
1	Raj	Rajkot	CE	Sah
2	Meet	Surat	C.E	Shah
3	Jay	Baroda	Computer	Shaah
4	Hari	Rajkot	IT	Dave

An update anomaly exists **when one or more records (instance) of duplicated data is updated, but not all**.

Want to update manager of CE department

- Suppose the **manager of a (CE) department has changed**, this requires that the **Manager in all the tuples corresponding to that department must be changed** to reflect the new status.
- If we **fail to update all the tuples of given department**, then **two different records of employee working in the same department might show different Manager lead to inconsistency** in the database.

How to deal with insert, delete and update anomaly

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Emp_Dept

EID	**Ename**	**City**	**DID**	**Dname**	**Manager**
1	Raj	Rajkot	1	CE	Shah
2	Meet	Surat	1	C.E	Shah
3	Jay	Baroda	2	IT	Dave
NULL	NULL	NULL	3	EC	NULL

Emp

EID	**Ename**	**City**	**DID**
1	Raj	Rajkot	1
2	Meet	Surat	1
3	Jay	Baroda	2

Dept

DID	**Dname**	**Manager**
1	CE	Shah
2	IT	Dave
3	EC	NULL

Such type of anomalies in the database design can be solved by using **normalization.**

closure of attribute sets

What is a closure of attribute sets?

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given a set of attributes α, the closure of α under F is the **set of attributes that are functionally determined by α under F**.
- It is denoted by α^+.

What is a closure of attribute sets?

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given a set of attributes α, the closure of α under F is the **set of attributes that are functionally determined by α under F**.
- It is denoted by **α^+**.

Algorithm

- Algorithm to compute α^+, the closure of α under F
  - Steps
    1. result = α
    2. *while* (changes to result) *do*
       - for each β → γ in F do
         - begin
           - if β ⊆ result then result = result U γ
           - else result = result
         - end

Closure of attribute sets [Example]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Consider the relation schema R = (A, B, C, G, H, I).
- For this relation, a set of functional dependencies F can be given as

$$F = \{A \rightarrow B,\ A \rightarrow C,\ CG \rightarrow H,\ CG \rightarrow I,\ B \rightarrow H\}$$

- Find out the closure of $(AG)^+$.

Algorithm

- Algorithm to compute α^+, the closure of α under F
  - Steps
    1. result = α
    2. *while* (changes to result) *do*
       - for each $\beta \rightarrow \gamma$ in F do
         - begin
           - if $\beta \subseteq$ result then result = result U γ
           - else result = result
         - end

- Step 1.

  result = α => result = AG

$A \rightarrow B$	$A \subseteq AG$	result = ABG
$A \rightarrow C$	$A \subseteq ABG$	result = ABCG
$CG \rightarrow H$	$CG \subseteq ABCG$	result = ABCGH
$CG \rightarrow I$	$CG \subseteq ABCGH$	result = ABCGHI
$B \rightarrow H$	$B \subseteq ABCGHI$	result = ABCGHI

$$AG^+ = ABCGHI$$

Closure of attribute sets [Exercise]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given functional dependencies (FDs) for relational schema R = (A,B,C,D,E):
- F = {A → BC, CD → E, B → D, E → A}
  - Find Closure for A
  - Find Closure for CD
  - Find Closure for B
  - Find Closure for BC
  - Find Closure for E

Answer

A^+ = ABCDE

CD^+ = ABCDE

B^+ = BD

BC^+ = ABCDE

E^+ = ABCDE

Closure of attribute sets [Exercise]

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given functional dependencies (FDs) for relational schema R = (A,B,C,D,E):

F = {A → B, B → D, C → DE, CD → AB}

  - Find Closure for A
  - Find Closure for B
  - Find Closure for C
  - Find Closure for D
  - Find Closure for E

Answer

A^+ = BDA

B^+ = BD

C^+ = ABCDE

D^+ = D

E^+ =E

From C, we can find all attributes.

X	Y	Z
1	1	1
2	1	2
2	1	3
3	1	1

Which of the following is Functional Dependency?

1. $X \rightarrow Y$
2. $XY \rightarrow Z$
3. $YZ \rightarrow X$

R(A,B,C,D,E,G)
{
$A \rightarrow B$
$B \rightarrow C$
$D \rightarrow E$
$E \rightarrow G$
}

The following functional dependencies are given:

$$AB \to CD, AF \to D, DE \to F, C \to G, F \to E, G \to A.$$

Which one of the following options is false?

(A) $\{CF\}^+ = \{ACDEFG\}$ (C) $\{AF\}^+ = \{ACDEFG\}$

(B) $\{BG\}^+ = \{ABCDG\}$ (D) $\{AB\}^+ = \{ABCDFG\}$

Determining the Candidate Key

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given functional dependencies (FDs) for relational schema R(A, B, C, D)

FD = {A → B, B → C, C → D, D → A}

Depending on the closure set find out the candidate keys of table R.

How many candidate keys should we test?

How many candidate keys are possible?

Determining the Candidate Key

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given a relation, how many candidate keys are possible?

with 2 attributes (A, B):

possible combinations (A, B, AB)

with 3 attributes (A, B, C):

possible combinations (A, B, C
AB, AC, BC, ABC)

with n attributes (A1, A2, ..., An):

\# possible combinations = 2^n-1

Determining the Candidate Key

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given functional dependencies (FDs) for relational schema R(A, B, C, D)

FD = {$A \to B$, $B \to C$, $C \to D$, $D \to A$}

Depending on the closure set find out the candidate keys of table R.

Possible CKs:

2^4-1 = 15

Question

From these 15 possible combinations, which are actually CKS?

CK of length 4:

(ABCD)

CK of length 3:

(ABC), (ABD), (BCD), (ACD)

CK of length2:

(AB), (AC), (AD), (BC), (BD), (CD)

CK of length1:

(A), (B), (C), (D)

NOTE

Start to check with the smallest length

Determining the Candidate Key

- Given functional dependencies (FDs) for relational schema R(A, B, C, D)

FD = {A → B, B → C, C → D, D → A}

Depending on the closure set find out the candidate keys of table R.

With length 1:
A^+ = {ABCD}
B^+ = {ABCD}
C^+ = {ABCD}
D^+ = {ABCD}

Answer

CKs = A, B, C, and D

Determining the Candidate Key

- Given functional dependencies (FDs) for relational schema R(E, F, G, H, I, J, K, L, M, N,)

  FD = { {E, F} → {G}, {F} → {I, J}, {E, H} → {K, L }, {K} → {M}, {L} → {M} }

What are the keys for table R?

a) {E, F}

b) {E, F, H}

c) {E, F, H, K, L}

d) {E}

Determining the Candidate Key

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given functional dependencies (FDs) for relational schema R(A, B, C, D, E, H)

$$FD = \{ A \rightarrow B, \quad BC \rightarrow D, \quad E \rightarrow C, \quad D \rightarrow A \}$$

What are the candidate keys of R?

a) AE, BE

b) AE, BE, DE

c) AEH, BEH, BCH

d) AEH, BEH, DEH

Determining the Candidate Key

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given functional dependencies (FDs) for relational schema R(A, B, C, D, E, F)

  FD = { $AB \rightarrow C$, $C \rightarrow D$, $D \rightarrow EB$, $E \rightarrow F$, $F \rightarrow A$ }

How many candidate keys for R?

Determining the FDs

- Given a relational schema R(A, B) with 2 attributes

How many FDs are possible?

Determining the FDs

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given a relational schema R(A, B, C) with 3 attributes and FDs = {A → B, B → C }

Set of all FDs that can be derived?

Equivalence

Equivalence of FDs

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

- Given a two FDs F and G, they will be equivalent iff
  - G is the subset of F, i.e., F is covering G, and
  - F is the subset of G, i.e., G is covering F,
- Example:

F: {A → B, AB → C, D → AC, D → E}

G: {A → BC, D → AB}

Not equivalent

- Example:

F: {A → B, B → C, C → A}

G: {A → BC, B → A, C → A}

Equivalent