DBMS Week 6 TA Session

Normalization and Normal Forms

- Normalization or Schema Refinement is a technique of organizing the data in the database
- Normalization is used for mainly two purpose:
  - Eliminating redundant (useless) data
  - Ensuring data dependencies
- A normal form specifies a set of conditions that the relational schema must satisfy in terms of its constraints

1NF (First Normal Form)

- All Domains should only have Atomic Values
- No Multivalued attributes

Example

SID	Sname	Cname
S1	A	C, C++
S2	B	C++, DB
S3	C	DB

- *Cname* is a multivalued attribute. So, the above table is not in *1NF*

2NF (Second Normal Form)

Relation R is in Second Normal Form (2NF) only if

- R is in 1NF and
- R contains no Partial Dependency

Partial Dependency

$A \rightarrow B$ is a Partial dependency only if

- A - Proper subset of Candidate Key
- B - Non Prime Attribute

Example

CHECK WHETHER IT IS IN 2NF OR NOT

Consider a relation $\mathbf{R}(A,B,C,D,E,F,G,H)$, where each attribute is atomic and the following functional dependencies hold:

$\mathcal{F}=\{AB \rightarrow CDE, D \rightarrow F, F \rightarrow GH, E \rightarrow AB\}$

3NF (Third Normal Form)

A relational schema R is in 3NF if for every FD $A \rightarrow B$ associated with R either

- R is in 2NF and
- $B \subseteq A$ (Trivial FD) (or)
- A is a superkey of R (or)
- B is a prime attribute

Note:

- 3NF preserves both dependency preservation and loseless join decomposition

Example

Consider a relational schema **Contacts**$(aadhaarNo, name, mobileNo, address)$.Assume that all the attributes have atomic values. Which of the following functional dependencies is/are example(s) of the third normal form?

$$\mathcal{F} = \{aadhaarNo \rightarrow (name, address),$$
$$mobileNo \rightarrow aadhaarNo,$$
$$(name, address) \rightarrow mobileNo\}$$

$$\mathcal{F} = \{aadhaarNo \rightarrow name,$$
$$mobileNo \rightarrow aadhaarNo,$$
$$(name, address) \rightarrow aadhaarNo\}$$

$$\mathcal{F} = \{(aadhaarNo, name) \rightarrow address,$$
$$mobileNo \rightarrow name,$$
$$(name, aadhaarNo) \rightarrow mobileNo\}$$

$$\mathcal{F} = \{(aadhaarNo, name) \rightarrow (address, mobileNo),$$
$$mobileNo \rightarrow aadhaarNo,$$
$$name \rightarrow address\}$$

BCNF (Boyce – Codd Normal Form)

A relational schema R is in BCNF if for every FD $A \rightarrow B$ associated with R either

- R is in 3NF and
- A must be a superkey (or)
- $B \subseteq A$ (Trivial FD)

Note:

- BCNF decomposition is loseless
- BCNF decomposition is may or may not be dependency preserving

Example

Consider the relational schema $\mathrm{R}(U, V, W, X, Y, Z)$ where the domain of every attribute consists of atomic values. The set of functional dependencies for the relation R is given as follows:

$$\mathcal{F} = \{UV \rightarrow W, W \rightarrow X, X \rightarrow VY, Y \rightarrow Z, Z \rightarrow U\}$$

What is the highest normal form of the given relation R?

MVD (Multivalued Dependency)

Let R be a relation schema and $\alpha \subseteq R$ and $\beta \subseteq R$. The multivalued dependency

$$\alpha \twoheadrightarrow \beta$$

holds on R if in any legal relation $r(R)$, for all pairs for tuples $t1$ and $t2$ in r such that $t1[\alpha] = t2[\alpha]$, there exist tuples t3 and t4 in r such that:

- $t1[\alpha] = t2[\alpha] = t3[\alpha] = t4[\alpha]$
- $t1[\beta] = t3[\beta] \quad \text{and} \quad t2[\beta] = t4[\beta]$
- $t2[R–\beta] = t3[R–\beta] \quad \text{and} \quad t1[R–\beta] = t4[R–\beta]$

For MVD,

- Total number of attributes should be more than two
- If there exist 3 attributes, then 2 attributes must be independent of each other

MVD Theory

Name	Rule
Complementation	If $X \twoheadrightarrow Y$, then $X \twoheadrightarrow (R - (X \cup Y))$
Augmentation	If $X \twoheadrightarrow Y$ and $W \supseteq Z$, then $WX \twoheadrightarrow YZ$
Transitivity	If $X \twoheadrightarrow Y$ and $Y \twoheadrightarrow Z$, then $X \twoheadrightarrow (Z - Y)$
Replication	If $X \rightarrow Y$, then $X \twoheadrightarrow Y$ but the reverse is not true
Coalescence	If $X \twoheadrightarrow Y$ and there is a W such that $W \cap Y$ is empty, $W \rightarrow Z$ and $Z \subseteq Y$, then $X \rightarrow Z$

Example

SId	Sname	Course	Instructor	Inst_Room
ME1001	David	Python	MK Singh	503
ME1001	David	Java	SN Joseph	505
ME1001	David	Python	SN Joseph	505
ME1001	David	Java	MK Singh	503

- {SId, Sname} $\twoheadrightarrow$ Course$
- SId $\twoheadrightarrow$ {Instructor, Inst_Room}

Example

course_name	instructor	book	edition
DBMS	Geeta	DBMS-Beginner	3
DBMS	Arjun	DBMS-Beginner	3
DBMS	Geeta	DBMS-Expert	2
DBMS	Arjun	DBMS-Expert	2
Java	Rahul	Java-Beginner	5
Java	Rahul	Java-Intermediate	3
Java	Rahul	Java-Expert	4
Java	Armaan	Java-Beginner	5
Java	Armaan	Java-Intermediate	3
Java	Armaan	Java-Expert	4

- course_name $\twoheadrightarrow$ instructor
- course_name $\twoheadrightarrow$ {book, edition}

Trivial MVD

A MVD $X \twoheadrightarrow Y$ in R is called a trivial MVD is

- Y is a subset of X $(Y \subseteq X)$ (or)
- $X \cup Y = R$

Example

- $AB \twoheadrightarrow B$ (trivial MVD)

4NF (Fourth Normal Form)

A relation schema R is in 4NF if and only if the following conditions are satisfied

- R is in BCNF and
- Should not have any multi-valued dependency