UNIT 2

Lecture 34

Normalization

4NF & 5NF (PJNF)

Multivalued Dependency

- A multivalued dependency X →→Y specified on relation schema R, where X and Y are both subsets of R, specifies the following constraints on any relation state r of R : If two tuples t1 and t2 exist in r such that t1[X] = t2[X], then two tuples t3 and t4 should also exist in r with the following properties, where we use Z to denote (R – (X ∪ Y)):

  t1[X] = t2[X] = t3[X] = t4[X].

  t1[Y] = t3[Y] and t2[Y] = t4[Y].

  t1[Z] = t4[Z] and t2[Z] = t3[Z].

- Whenever X →→Y holds, we say that X multidetermines Y.
- Because of the symmetry in the definition, whenever X →→Y holds in R, so does X→→Z. Hence, X→→Y implies X→→Z, and therefore it is sometimes written as X→→Y|Z.

Multivalued Dependency

For e.g., consider the relation EMP

EMP

ENAME	PNAME	DNAME
Kumar	X	Rakesh
Kumar	Y	Rao
Kumar	X	Rao
Kumar	Y	Rakesh

- From the above fig we have ENAME →→ PNAME and ENAME →→DNAME (or ENAME→→PNAME|DNAME) holds in the EMP relation.
- The employee with ENAME 'Kumar' works on projects with PNAME 'X' and 'Y' and has two dependents with DNAME 'Rakesh' and 'Rao'.

Trivial Multivalued Dependency

A MVD X→→Y in R is called a trivial MVD if

a) Y is a subset of X (i.e. X ⊇ Y) or

b) X ∪ Y = R.

For e.g., consider the relation EMP_PROJECTS

ENAME	PNAME
Kumar	X
Kumar	Y

From the above fig we have a trivial MVD

ENAME →→PNAME.

Inference Rules for Functional and Multivalued Dependencies

For the relation schema R(A1, A2, ….,An) and that X, Y, Z and W are subsets of R.

1. IR1 (Reflexive Rule for FDs) : If $X \supseteq Y$, then $X \rightarrow Y$.
2. IR2 (Augmentation Rule for FDs) : $\{X \rightarrow Y\} \models XZ \rightarrow YZ$.
3. IR3 (Transitive Rule for FDs) : $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$.
4. IR4 (Complementation Rule for MVDs) : $\{X \rightarrow\rightarrow Y\} \models \{X \rightarrow\rightarrow (R - (X \cup Y))\}$.
5. IR5 (Augmentation Rule for MVDs) : If $X \rightarrow\rightarrow Y$ and $W \supseteq Z$, then $WX \rightarrow\rightarrow YZ$.
6. IR6 (Transitive Rule for MVDs) : $\{X \rightarrow\rightarrow Y, Y \rightarrow\rightarrow Z\} \models X \rightarrow\rightarrow (Z - Y)$.
7. IR7 (Replication Rule for FD and MVD) : $\{X \rightarrow Y\} \models X \rightarrow\rightarrow Y$.
8. IR8 (Coalescence Rule for FDs and MVDs) : If $X \rightarrow\rightarrow Y$ and there exists W with the properties that
   a) $W \cap Y$ is empty,
   b) $W \rightarrow Z$, and
   c) $Y \supseteq Z$, then $X \rightarrow Z$.

Fourth Normal Form

- A relation schema R is in 4NF with respect to a set of dependencies F (that includes functional dependencies and multivalued dependencies) if, for every nontrivial multivalued dependency $X \rightarrow\rightarrow Y$ in F^+, X is a super key for R.
- For e.g., consider the EMP relation

ENAME	PNAME	DNAME
Kumar	X	Rakesh
Kumar	Y	Rao
Kumar	X	Rao
Kumar	Y	Rakesh

This relation is not in 4NF because in the nontrivial MVDs ENAME→→PNAME and ENAME→→DNAME, ENAME is not a super key of EMP.

Fourth Normal Form

We decompose EMP into EMP_PROJECTS and EMP_DEPENDENTS shown in fig below.

EMP_PROJECTS

ENAME	PNAME
Kumar	X
Kumar	Y

EMP_DEPENDENTS

ENAME	DNAME
Kumar	Rakesh
Kumar	Rao

- Both EMP_PROJECTS and EMP_DEPENDENTS are in 4NF, because the MVDs ENAME→→PNAME in EMP_PROJECTS and ENAME→→DNAME in EMP_DEPENDENTS are trivial MVDs.
- No other nontrivial MVDs hold in either EMP_PROJECTS or EMP_DEPENDENTS. No FDs hold in these relation schemas either.

Join Dependency (JD)

- A join dependency (JD), denoted by JD(R1, R2, ….,Rn), specified on relation schema R, specifies a constraint on the states r of R. The constraint states that every legal state r of R should have a non-additive join decomposition into R1, R2,...,Rn; that is, for every such r we have

  $$*(\prod_{R1}(r), \prod_{R2}(r), \ldots, \prod_{Rn}(r)) = r$$

- An MVD is a special case of a JD where n = 2. That is, a JD denoted as JD(R1, R2) implies a MVD $(R1 \cap R2) \rightarrow\rightarrow (R1 - R2)$ (or, by symmetry, $(R1 \cap R2 \rightarrow\rightarrow (R2 - R1))$.
- A join dependency JD(R1, R2,…,Rn), specified on relation schema R, is a trivial JD if one of the relation schemas Ri in JD(R1, R2,…,Rn) is equal to R.
- Such a dependency is called trivial because it has the nonadditive join property for any relation state r of R and hence does not specify any constraint on R.

Join Dependency

For e.g., consider the relation SUPPLY with JD(R1, R2, R3) shown below.

From the relations we have

*(R1, R2, R3) = SUPPLY

(R1 $\bowtie$ R2 $\bowtie$ R3) = SUPPLY.

SUPPLY

SNAME	PARTNAME	PROJNAME
Kumar	Bolt	ProjX
Kumar	Nut	ProjY
Prakash	Bolt	ProjY
Ramesh	Nut	ProjZ
Prakash	Nail	ProjX
Prakash	Bolt	ProjX
Kumar	Bolt	ProjY

Join Dependency

R1

SNAME	PARTNAME
Kumar	Bolt
Kumar	Nut
Prakash	Bolt
Ramesh	Nut
Prakash	Nail

R2

SNAME	PROJNAME
Kumar	ProjX
Kumar	ProjY
Prakash	ProjY
Ramesh	ProjZ
Prakash	ProjX

R3

PARTNAME	PROJNAME
Bolt	ProjX
Nut	ProjY
Bolt	ProjY
Nut	ProjZ
Nail	ProjX

Fifth Normal Form (PJNF)

- A relation schema R is in fifth normal form (5NF) (or project join normal form [PJNF]) with respect to a set of F of functional, multivalued, and join dependencies if, for every nontrivial join dependency JD(R1, R2,…,Rn) in F+ (that is implied by F), every Ri is a super key of R.
- For e.g., consider the SUPPLY relation shown above. The super key of this relation is {SNAME, PARTNAME, PROJNAME}. So, the SUPPLY relation is not in 5NF because we have a JD(R1, R2, R3) over R and every Ri is a not super key of R.
- So, The decomposition of R1(SNAME, PARTNAME), R2(SNAME, PROJNAME) and R3(PARTNAME, PROJNAME) is in 5NF.

Question

- Prove that a relation with 2 attributes is always in BCNF.

For Video lecture on this topic please subscribe to my youtube channel.

The link for my youtube channel is

https://www.youtube.com/channel/UCRWGtE76JlTp1iim6aOTRuw?sub_confirmation=1