

Normalization 4 NF, 5 NF

Dependencies

- Functional Dependencies FDs: Role in decomposition and schema refinement
- Multivalued Dependencies
- Join Dependencies

4 NF

- BCNF removes any anomalies due to FDs
- Further research has led to the identification of another type of dependency called Multi-valued Dependency (MVD)
- Proposed by R Fagin* in 1977
- MVDs can also cause data redundancy
- MVDs are a generalization of FDs

* R Fagin *Multi-valued Dependencies & a new normal form for relational databases*, ACM TODS2, No. 3 (Sept. 1977)

Multivalued Dependency

- Relation R (Course,Teacher,textbook) CTX
- Teacher T can teach course C and text X is recommended for this course
- No FDs exist
- Key is CTX
- Recommended texts for a course are independent of instructor
- CTX is in BCNF
- There is redundancy:
- Try inserting the fact that there is a new teacher for Physics101
- The text optics is a text for Physics 101 course is recorded once per potential teacher

MVD

- Consider the following relation CTX:

Course	Teacher	Texts
DBMS	M Bhise NDJ	RG Korth
MDA	M Bhise	G Booch McRobb

- In relational databases, repeating groups are not allowed

MVD

- 1 NF Version

CTX	COURSE	TEACHER	TEXTS
	DBMS	M Bhise	RG
	DBMS	M Bhise	Korth
	DBMS	NDJ	RG
	DBMS	NDJ	Korth
	MDA	M Bhise	G Booch
	MDA	M Bhise	McRobb

NO FDs in this relation

Anomalies

- New Teacher for DBMS
- New Text for DBMS
- Teacher teaching MDA leaves

Contd..

- MVD: Texts for a course are independent of the instructors
- CTX has no FDs at all (redundancies !!!)
- Use 2 binary relationship sets : Instructor (C,T) and Text (C, X) as these 2 are independent relationships
- If there is a tuple showing that C is taught by teacher T
- And there is a tuple showing that C has a book X as text
- Then there is tuple showing that C is taught by T and has text X
- If tuples $(c,t1,x1)$, $(c,t2,x2)$ appear then $(c,t1,x2)$ and $(c,t2,x1)$ also appear

MVD

- The relation CTX is not in 4NF as C->->T is a nontrivial MVD and C is not a key
- But each of CT and CX are in BCNF
- If a relation is in BCNF, and at least one of its keys consist of single attribute then it is also in 4NF

4 NF

- Relation in BCNF and non-trivial MVDs absent

4 NF

- Decompose CTX into CT & CX

CT

<u>COURSE</u>	<u>TEACHER</u>
DBMS	M Bhise
DBMS	NDJ
MDA	M Bhise

CX

<u>COURSE</u>	<u>TEXT</u>
DBMS	RG
DBMS	Korth
MDA	G Booch
MDA	McRobb

4 NF

- Decompose CTX into CT and CX
- Decomposition of CTX into CT & TX is not done on the basis of FDs (as there are no FDs)
- Decompose CTX into CT & TX is done on the basis of MVDs
- MVDs

Represents a dependency between attributes of a relation, such that for every value of A, there is a set of values of B & a set of values of C, The set of values for B & C are independent of each other

$\text{course} \rightarrow\!\!\! \rightarrow \text{teacher}$ (course multi-determines teacher)
 $\text{course} \rightarrow\!\!\! \rightarrow \text{text}$ (text multi-dependent on course)

MVD contd..

- Decompose CTX into CT and CX
- Phy taught by a new teacher
- Nonlossy decomposition
- CTX is in BCNF (all key)
- CT and CX are also in BCNF (all key)
- MVDs are generalization of FDs
- Every FD is an MVD (but converse is not true)
- CTX has 2 MVDs course->->teacher , course ->->Text (teacher multidependent on course)
- Each course has well defined set of teachers and well defined set of texts

Contd..

- MVD Definition : Let A,B,C be subsets of relation R, we can write $A->->B$ if and only if, in every possible legal value of R, the set of B values matching a given (A value, C value) pair depends only on the A value and is independent of C value
- Fagin theorem tells you that MVDs go in pairs
- $A->->B$ holds only if MVD $A->->C$ also holds $A->-> B| C$
- Every FD is an MVD in which the set of dependent (RHS) values matching a given determinant (LHS) value is always a singleton set

MVDs

- Some MVDs are not FDs
- The existence of such MVDs in CTX requires: to insert 2 tuples to add a new physics teacher
- These 2 tuples are needed to maintain an integrity constraint that is presented by $C \rightarrow\!-\!> X$
- CT and CX don't include any such MVDs
- A multi-valued dependency occurs when a determinant determines more than one dependent, and the dependents are independent of each other

Fagin Theorem

- Stronger version of Heath's Theorem
- Let $R\{A,B,C\}$ be a relation, where A,B,C are sets of attributes. Then R is equal to the join of its projections on $\{A,B\}$ and $\{A,C\}$ if and only if R satisfies the MVDs $A \rightarrow\rightarrow B|C$

4 NF

- An MVDs $A \rightarrow\!\!\!\rightarrow B$ is trivial if
 - (a) $B \subseteq A$ or
 - (b) $A \cup B = R$
- **A relation that is in BCNF & contains no non-trivial MVDs is said to be in 4NF**
- CTX is not in 4NF because *course* $\rightarrow\!\!\!\rightarrow$ *teacher* is a non trivial MVD
- What about CT and CX?

Multi-Valued Dependencies

- Most common source of redundancy in BCNF schemas is to put 2 or more M:M relationships in a single relation

Fourth Normal Form 4NF

- R is in 4NF if and only if, whenever there exist subsets A and B of attributes of R such that the nontrivial MVD $A \rightarrow\!-\!\rightarrow B$ is satisfied, then all attributes of R are also functionally dependent on A
- The only nontrivial dependency (FDs or MVDs) in R are of the form $K \rightarrow X$
- R is in 4NF if it is in BCNF and all (nontrivial) MVDs in R are in fact FDs out of keys
- CTX is not in 4NF since it involves MVD that is not an FD at all, let alone an FD out of a key
- MVD $A \rightarrow\!-\!\rightarrow B$ is trivial if either A is a superset of B or the union of A and B is the entire R

Suppliers-Parts-Projects Database SPJ

S#	P#	J#
S1	P1	J2
S1	P2	J1
S2	P1	J1
S1	P1	J1

Join Dependencies

- N- decomposable relation
- (but not any m where $m < n$)
- SPJ (supplier#, part#, project#) is the join of SP, PJ and JS
 - If pair (s1,p1) appears in SP
 - And pair (p1,j1) appears in PJ
 - And the pair (j1,s1) appears in JS
 - Then the triple (s1,p1,j1) appears in SPJ

Join

- Join (SP, PJ)
- Join (SP,PJ,JS)

Contd...

- If s_1 is linked to p_1 , p_1 is linked to j_1 and j_1 is linked to s_1 , then s_1, p_1 and j_1 coexist in the same tuple
- A relation is n decomposable for some $n > 2$ if and only if it satisfies some such n -way cyclic constraint
- If (s_2, p_1, j_1) is inserted then (s_1, p_1, j_1) must also be inserted to validate JD integrity constraint

Join Dependency

- JD is a constraint on the set of legal relations over a database scheme. An instance of relation R is subject to a join dependency if it can always be recreated by joining multiple tables each having a subset of the attributes of R
- If one of the tables in the join has all the attributes of the table T, the join dependency is called trivial

Join Dependency

Let R has subsets of attributes A,B,...,Z then we say that R satisfies JD *{A,B,...,Z} if and only if every possible legal value of R is equal to the join of its projections on A,B,...,Z

JD *{SP,PJ,JS}

Fagin's Theorem (modified)

$R\{A,B,C\}$ satisfies $JD^*\{AB,AC\}$ if and only if it satisfies the MVDs
 $A \rightarrow\!-\!\rightarrow B|C$

MVD is a special case of JD (Like FD is a special case of MVD)

Nontrivial JD

- $\text{JD}^* \{A, B, \dots, Z\}$ is trivial if and only if one of the projections A, B, \dots, Z is the identity projection of R (ie projection over all attributes of R)

JD and MVD

- 2-ary join dependencies are called multivalued dependency. More specifically if U is a set of attributes and R a relation over it, then R satisfies $X \twoheadrightarrow Y$

if R satisfies

$$*(X \cup Y, X \cup (U - Y))$$