Higher Normal Forms

Swarup Kr Ghosh

swarupg1@gmail.com

Introduction: Multivalued Dependency

In a relation R (A, B, C), consider the following assumption:

- There are at least three attribute (column) in table
- For each value of A there are number of values B.
- For each value of A there are number of values C.
- B and C are independent to each other.
- B and C are multivalued dependent to A

Definition: In a relation R, three different attributes A, B, C such that A, B and C are subset of R. If for every value of A there exists a set of values of B and C but B and C are independent to each other then B and C are multivalued dependent on A.

EX

EMP_Id	Project_Id	Hobby
01	01	Reading
01	02	Swimming
02	02	Swimming
02	04	Singing

EMP_Id →→ Hobby
EMP_Id →→ Project_Id

EMP_Id	Project_Id	Hobby
E01	01	Reading
E01	02	Swimming
E01	02	Reading
E01	01	Swimming
E02	02	Swimming
E02	04	Singing
E02	04	Swimming
E02	02	Singing

Introduction

- Consider a relation HEPD

HEPD

ename	pname	dname
Smith	X Y	John Anna
William	Y Z	Jenny

Multi-valued Attributes
Ename→→danme

pname and dname
are independent

Introduction

EPD

ename	pname	dname
Smith	X	John
Smith	X	Anna
Smith	Y	John
Smith	Y	Anna
William	Y	Jenny
William	Z	Jenny
---	---	---

❖EPD is an all key table

❖Constraint:

If tuples (e,p1,d1) and (e,p2,d2) both appear then tuples (e,p1,d2), (e,p2,d1) both appear also

❖**REDUNDANCY**

❖**UPDATE Anomalies**

Problem arises as pname and dname are independent

@SKG 11/23/2020 5

Introduction

EPD

ename	pname	dname
Smith	X	John
Smith	X	Anna
Smith	Y	John
Smith	Y	Anna
William	Y	Jenny
William	Z	Jenny
--	--	--

❖Decompose the table →

Decomposition must be non-loss

EP

ename	pname
Smith	X
Smith	Y
William	Y
William	Z
--	--

ED

ename	dname
Smith	John
Smith	Anna
William	Jenny
---	---

@SKG 11/23/2020 6

Few Important Issues

- ❖ EPD was not well designed
- ❖The decomposition into EP and ED was better not formally
- ❖EPD satisfies no non-trivial Functional Dependency
- ❖ ename→ename ……. were trivial
- ❖EPD is in BCNF ....
- ❖'Any "all-key" relation must necessarily be in BCNF'
- ❖EP and ED are all key………… Hence in BCNF

Few Important Issues

- ❖ EPD satisfies no non-trivial Functional Dependency
- ❖Attributes of EPD satisfies some different types of dependency
- ❖Consider ename and pname …… Every ename does not have a single corresponding pname….,i.e., FD ename→pname does not hold
- ❖Each ename does have a *well defined* set of corresponding pname
- ❖By *well defined* we mean…. For a given *ename e* and *dname d* the set of *pname p* matching the *(e,d)* pair in EPD depends on the value *e* alone……..it makes no difference which value of *d* we choose

Multi-valued Dependency

- Formally Multi-Valued Dependency (MVD) exists
- Multi-valued Dependency : Let R be a relation and let A,B,C be the subsets of the set of attributes of R. Then we say B is **multi-dependent** on A …….. in symbol

$$A \twoheadrightarrow B$$

- ….."A multi determines B" or "A double arrow B"…… if and only if, in every possible legal value of R, the set B vales matching a given (A value, C value pair) depends only on the A value and is independent of the C value

Multi-valued Dependency

- Multi-valued Dependency : Let R be a relation and let A,B,C be the subsets of the set of attributes of R. Then we say B is **multi-dependent** on A …….. in symbol

$$A \twoheadrightarrow B \mid C$$

- ….."A multi determines B" or "A double arrow B"…… if and only if, in every possible legal value of R, given a particular value of A, the set of values of B determined by this value of A, is completely determined by A alone, and does not depend on the values of the remaining attributes C of R

Multi-valued Dependency: Example

ename →→ pname ename →→ dname

- "Every FD is an MVD but vice versa is not true"
- EPD involves MVDs that are not FDs
- **Theorem (Fagin)**: Let R(A,B,C) be a relation, where A, B, and C are sets of attributes. Then R is equal to the projections on {A,B} and {A,C}, if and only if R satisfies the MVDs **A →→ B | C**

@SKG 11/23/2020 11

Multi-valued Dependency: Example

ename →→ pname ename →→ dname

- According to Fagin's theorem we can formally decompose the EPD relation into two relations: EP and ED
- Fourth Normal Form: Relation R is in 4NF if and only if, whenever there exists subsets A and B of the attributes of R such that the **nontrivial MVD** **A →→ B** is satisfied, then all attributes of R are also functionally dependent on A, i.e., A is a superkey of R.

An MVD A →→ B is **trivial** if either *A is a superset of B* or *the union of A and B is the entire heading*

@SKG 11/23/2020 12

Multi-valued Dependency: Example

❖ *Relation EPD is in 4NF or not……* Justify

❖ EP and ED are in 4NF or not……….. Justify

EPD

ename	pname	dname
Smith	X	John
Smith	X	Anna
Smith	Y	John
Smith	Y	Anna
William	Y	Jenny
William	Z	Jenny
--	--	--

EP

ename	pname
Smith	X
Smith	Y
William	Y
William	Z
--	--

ED

ename	dname
Smith	John
Smith	Anna
William	Jenny
---	---

Fourth Normal Form (4 NF)

- A relation in **Fourth normal form (4NF)** contains no non-trivial multi-valued dependency.
- In order to achieve 4NF, all non-trivial multi-valued dependencies are to be converted in trivial multi-valued dependencies in the given relation and it must be in BCNF.

- Multivalued Dependencies (MVDs):

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

$$\alpha \text{ ì ì } \beta$$

holds on R if in any legal relation $r(R)$, for all pairs for tuples t_1 and t_2 in r such that $t_1[\alpha] = t_2[\alpha]$, there exist tuples t_3 and t_4 in r such that:

$$t_1[\alpha] = t_2[\alpha] = t_3[\alpha] = t_4[\alpha]$$

$$t_3[\beta] = t_1[\beta]$$

Fourth Normal Form (4 NF)

EMP Table

Emp ID	Project ID	Hobby
101	P01	Singing
101	P01	Reading
101	P02	Singing
101	P02	Reading
102	P03	Singing
102	P04	Reading
102	P04	Singing
102	P03	Reading

EmpID ì ì ProjectID
EmpID ì ì Hobby

This is not in 4NF

E1 Table

Emp ID	Project ID
101	P01
101	P02
101	P03
102	P03
102	P04

E2 Table

Emp ID	Hobby
101	Singing
101	Reading
102	Singing
102	Reading

This is in 4NF

Join Dependency

In a relation R (A, B, C), consider the following assumption:

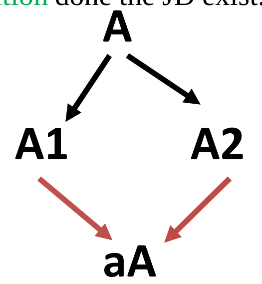
- B, C → A
- A, C → B
- A, B → C
- Then join dependency (JD) exists
- Sometimes it is called Project Join Normal Form

Definition: A join dependency (JD), denoted by JD(R1,R2, … Rn), specified on relation schema R, specifies a constraint on the states r of R.
Natural join (R1(r),R2(r), … Rn(r)) = r
Join dependency, multiway decomposition, results the fifth normal form (5NF)

Join Dependency

How does you ensure?

- Suppose there is a table A and decompose into two parts A1 and A2
- Now apply joining between A1 and A2
- If lossless decomposition done the JD exist.



JD does not exist if data is lost or new entries are created

Join Dependency and Fifth Normal Form

SPJ

S#	P#	J#
s1	p1	J2
s1	p2	J1
s2	p1	J1
s1	p1	J1

- ❖ SPJ is "all-key" relation
- ❖ Involves no non-trivial FD and MVDs
- ❖ "SPJ is in 4NF"….. Justify

2-decomposable……….

Join Dependency and Fifth Normal Form

SPJ

S#	P#	J#
s1	p1	J2
s1	p2	J1
s2	p1	J1
s1	p1	J1

SP

S#	P#
s1	p1
s1	p2
s2	p1

$\bowtie$

PJ

P#	J#
p1	J2
p2	J1
p1	J1

SJ

S#	J#
s1	J2
s1	J1
s2	J1

S#	P#	J#
s1	p1	J2
s1	p2	J1
s2	p1	J1
s2	p1	J2
s1	p1	J1

Spurious

$\bowtie$

Original SPJ

NON-LOSS DECOMPOSITION

Join Dependency

❖Join Dependency: Let R be a relation, and let A,B,…….,Z be subsets of the attributes of R. Then, we say that R satisfies the JD

* {A, B, C, …………,Z}

(read "star 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

❖Fagin's theorem can be restated as: R(A,B,C) satisfies the JD *{ AB,AC} if and only if it satisfies the MVDs $A \twoheadrightarrow$ **B | C**

Formally we have, $A \twoheadrightarrow$ **B | C** ≡ ***{AB, AC}**

Fifth Normal Form

❖ **Fifth Normal Form:** A relation is in 5NF….. also called projection-join normal form (PJ/NF)…… if and only if every non-trivial join dependency that holds for R is implied by the candidate key of R

❖Trivial Join Dependency: The JD *{A,B,……,Z} is trivial if and only if one of the projections A,B,………,Z is the identity projection of R,i.e., the projection over all attributes of R

Fifth Normal Form

❖ JD implied by the candidate key of R

Example-I: Let us have a relation R (s#, sname, status,city)
candidate key: s# and sname

R1 (s#, sname, status)   R2 (s#,city)

JD implied by the candidate key s#

Fifth Normal Form

- JD implied by the candidate key of R

Example-II: Let us have a relation R (s#, sname, status,city)

candidate key: s# and sname

R1 (s#, sname)   R2 (s#,status)   R3 (sname,city)

JD implied by both the candidate keys s# and sname

Join Dependency & 5NF

- A MVD is a special case of a JD with n=2.
- JD(R1,R2) ➔ MVD(R1 ∩ R2) ->> R1 – R2
- MVD(R1 ∩ R2) ->> R2 – R1
- A JD is trivial if any of Ri is R.
- The 5NF is also called project-join normal form (PJNF).

Definition: A relation schema is in 5NF or project-join normal form(PJNF) w.r.t a set of F of functional, multivalued and join dependencies if, for every join dependency JD(R1,R2, … , Rn) in closure of F, every Ri is a super key of R.

Fifth Normal Form (5 NF)

A Relation R is in 5NF iff every join dependency in R is implied by the candidate keys of R.

- Just breaking down the table

Company Table

Agent	Company	Product
A1	PQR	Nut
A1	PQR	Bolt
A1	XYZ	Nut
A1	XYZ	Bolt
A2	PQR	Nut

R1

Agent	Company
A1	PQR
A1	XYZ
A2	PQR

R2

Company	Product
PQR	Nut
PQR	Bolt
XYZ	Nut
XYZ	Bolt

R3

Agent	Product
A1	Nut
A1	Bolt
A2	Nut

This is not in 5 NF

R1, R2 and R3 are in 5NF

Go to: https://www.youtube.com/watch?v=mbj3HSK28Kk

Note: Join decomposition is a further generalization of Multivalued dependencies. If the join of R1 and R2 over C is equal to relation R, then we can say that a join dependency (JD) exists. Where R1 and R2 are the decompositions R1(A, B, C) and R2(C, D) of a given relations R (A, B, C, D).

Thank You

?