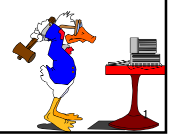
Multivalued Dependencies (MVDs)



1

Limitations of FD's

- Some redundancies cannot be detected using just functional dependencies
- Example:
  - suppose a teacher can teach several courses, several books can be recommended for a course, and same book can be recommended for different courses.

2

Limitations of FD's

- Suppose all this information is smushed together into one relation *CTB*
  - There are no FD's, so key is *CTB* and relation is in BCNF
  - But there is redundancy since course implies book
  - Eliminate redundancy by decomposing into *CT* and *CB*

3

Example of FD Limitations

Course (C)	*Teacher (T)*	*Book (B)*
Phys101	Green	Mechanics
Phys101	Green	Optics
Phys101	Brown	Mechanics
Phys101	Brown	Optics
Math301	Green	Mechanics
Math301	Green	Vectors
Math301	Green	Geometry

4

Multivalued dependencies (MVD's)

- MVD's express a condition among tuples of a relation that exists when the relation is trying to represent more than one many-many relationship.
- Then certain attributes become independent of one another, and their values must appear in all combinations.

5

Example

- Drinkers(name, addr, phones, beersLiked)
- A drinker's phones are independent of the beers they like.
- Thus, each of a drinker's phones appears with each of the beers they like in all combinations.
- If a drinker has 3 phones and likes 10 beers, then the drinker has 30 tuples

6

Example

- where each phone is repeated 10 times and each beer 3 times
- This repetition is unlike redundancy due to FD's, of which name->addr is the only one.

7

MVD's –

- A multivalued *dependency* of some attribute X on an attribute-set D is like a functional dependency except that X sometimes has several values for a given value of ***D***.
- The crucial point is that once the ***D*** value is specified, the X values are independent of other attributes.

8

MVD's –

- So, we can think of X as a multivalued attribute implemented by putting different values in different rows,
- where the set of X values is fully determined by just the value of ***D***.

E.g.: imagine multivalued car-colour being determined by just the make and year of the car.

9

Multivalued Dependencies (MVDs)

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

$$\alpha \rightarrow\rightarrow \beta$$

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:

$$t_1[\alpha] = t_2[\alpha] = t_3[\alpha] = t_4[\alpha]$$
$$t_3[\beta] = t_1[\beta]$$
$$t_3[R - \beta] = t_2[R - \beta]$$
$$t_4[\beta] = t_2[\beta]$$
$$t_4[R - \beta] = t_1[R - \beta]$$

10

Multi-valued Dependency

Definition: Multivalued dependency (MVD):

$A_1A_2\ldots A_n \rightarrow\rightarrow B_1B_2\ldots B_m$ holds for relation R if:

For all tuples *t, u* in R

If $t[A_1A_2\ldots A_n] = u[A_1A_2\ldots A_n]$, then there exists a *v* in R such that:

(1) $v[A_1A_2\ldots A_n] = t[A_1A_2\ldots A_n] = u[A_1A_2\ldots A_n]$

(2) $v[B_1B_2\ldots B_m] = t[B_1B_2\ldots B_m]$

(3) $v[C_1C_2\ldots C_k] = u[C_1C_2\ldots C_k]$, where $C_1C_2\ldots C_k$ is all attributes in R except $(A_1A_2\ldots A_n \cup B_1B_2\ldots B_m)$

11

Alt 1 Definition of MVD

- In relation R, we say MVD X->->Y holds
- If whenever there are tuples s and t in R
- such that $\pi X(s) = \pi X(t)$,
- then there is a tuple r in r such that:
  - 1. $\Pi_{XY}(r) = \Pi_{XY}(s)$.
  - 2. $\Pi_{(R - Y) \cup X}(r) = \Pi_{(R - Y) \cup X}(t)$.
- I.e., r agrees with s on the attributes mentioned, and with t on X and all the attributes not mentioned.

12

Alt 2 Definition of MVD

- Notion of MVD captures redundancy that FD's can't
- A *multivalued dependency* (MVD) on *R*, *X* ->-> *Y*, says that if two tuples of *R* agree on all the attributes of *X*, then their components in *Y* may be swapped, and the result will be two tuples that are also in the relation.
- i.e., for each value of *X*, the values of *Y* are independent of the values of *R-X-Y*.

13

MVD Example

R(x, y, z)

X	Y	Z
A	B1	C1
A	B2	C2

X ->-> Y

X	Y	Z
A	B1	C1
A	B2	C2
A	B2	C1
A	B1	C2

14

Example: C ->-> T

Course (C)	*Teacher (T)*	*Book (B)*
Phys101	Green	Mechanics
Phys101	Green	Optics
Phys101	Brown	Mechanics
Phys101	Brown	Optics
Math301	Green	Mechanics
Math301	Green	Vectors
Math301	Green	Geometry

15

Example: **name →→ street city**

Stars

	name	*street*	*city*	*title*	*year*
t	C. Fisher	123 Maple Str.	Hollywood	Star Wars	1977
	C. Fisher	5 Locust Ln.	Malibu	Star Wars	1977
v	C. Fisher	123 Maple Str.	Hollywood	Empire Strikes Back	1980
u	C. Fisher	5 Locust Ln.	Malibu	Empire Strikes Back	1980
	C. Fisher	123 Maple Str.	Hollywood	Return of the Jedi	1983
	C. Fisher	5 Locust Ln.	Malibu	Return of the Jedi	1983

16

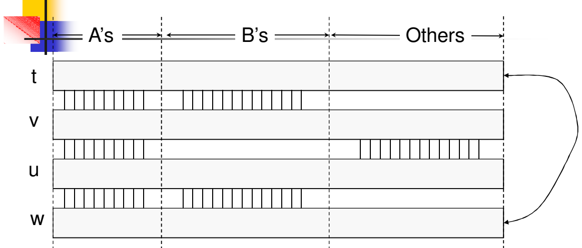
Example: **name →→ street city**

Stars

	name	*street*	*city*	*title*	*year*
u	C. Fisher	123 Maple Str.	Hollywood	Star Wars	1977
w	C. Fisher	5 Locust Ln.	Malibu	Star Wars	1977
v	C. Fisher	123 Maple Str.	Hollywood	Empire Strikes Back	1980
t	C. Fisher	5 Locust Ln.	Malibu	Empire Strikes Back	1980
	C. Fisher	123 Maple Str.	Hollywood	Return of the Jedi	1983
	C. Fisher	5 Locust Ln.	Malibu	Return of the Jedi	1983

17

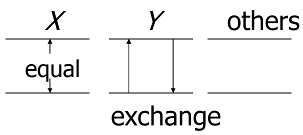
Pictorially Speaking...



- An MVD guarantees *v* exists
  - The existence of a fourth tuple *w* is implied by interchanging *t* and *u*

18

Picture of MVD *X* ->-> *Y*



MVD Rules

- Every FD is an MVD (*promotion*).
  - If *X* -> *Y*, then swapping *Y* 's between two tuples that agree on *X* doesn't change the tuples.
  - Therefore, the "new" tuples are surely in the relation, and we know *X* ->-> *Y*.
- In general If $A_1A_2...A_n \rightarrow B_1B_2...B_m$ then
  $A_1A_2...A_n \rightarrow\rightarrow B_1B_2...B_m$ holds

MVD Rules

- COMPLEMENTATION Rule
  - : If *X* ->-> *Y*, and *Z* is all the other attributes, then *X* ->-> *Z*.
  - If $A_1A_2...A_n \rightarrow\rightarrow B_1B_2...B_m$ then $A_1A_2...A_n \rightarrow\rightarrow C_1C_2...C_k$ where $C_1C_2...C_k$ is all attributes in R except ($A_1A_2...A_n \cup B_1B_2...B_m$)
- AUGMENTATION Rule
  - If $X \rightarrow\rightarrow Y$ and $W \supseteq Z$ then $WX \rightarrow\rightarrow YZ$

- TRANSITIVITY Rule

  If $X \rightarrow\rightarrow Y$ and $Y \rightarrow\rightarrow Z$ then $X \rightarrow\rightarrow (Z-Y)$

Fourth Normal Form

- The redundancy that comes from MVD's is not removable by putting the database schema in BCNF.
- There is a stronger normal form, called 4NF, that (intuitively) treats MVD's as FD's when it comes to decomposition, but not when determining keys of the relation.

Definition 4NF

- Given: relation R and set of MVD's for R
- Definition: R is in 4NF with respect to its MVD's if for every non-trivial MVD $A_1A_2...A_n \rightarrow\rightarrow B_1B_2...B_m$, $A_1A_2...A_n$ is a superkey

4NF Definition cont:

- A relation *R* is in 4NF if: whenever
  - *X*->->*Y* is a nontrivial MVD, then *X* is a superkey.
  - *Nontrivial MVD* means that:
    1. *Y* is not a subset of *X*, and
    2. *X* and *Y* are not, together, all the attributes.
  - Note that the definition of "superkey" still depends on FD's only.

25

Trivial MVDs

- $A_1A_2 \ldots A_n \rightarrow\rightarrow B_1B_2 \ldots B_m$ where $B_1B_2 \ldots B_m$ is a subset of $A_1A_2 \ldots A_n$ or ($A_1A_2 \ldots A_n \cup B_1B_2 \ldots B_m$) contains all attributes of R

26

4th Normal Form

- A Boyce Codd normal form relation is in fourth normal form if
  - (a) there is no multi value dependency in the relation or
  - (b) there are multi value dependency but the attributes, which are multi value dependent on a specific attribute, are dependent between themselves.

27

BCNF Versus 4NF

- Remember that every FD *X*->*Y* is also an MVD, *X*->->*Y*.
- Thus, if *R* is in 4NF, it is certainly in BCNF.
  - Because any BCNF violation is a 4NF violation.
- But *R* could be in BCNF and not 4NF, because MVD's are "invisible" to BCNF.

28

Decomposition and 4NF

- If *X*->->*Y* is a 4NF violation for relation *R*, we can decompose *R* using the same technique as for BCNF.
  1. *XY* is one of the decomposed relations.
  2. All but *Y* – *X* is the other.

29

Example

Consumers(<u>name</u>, addr, <u>phones</u>, <u>candiesLiked</u>)

FD: name -> addr

MVD's: name ->-> phones

name ->-> candiesLiked

- Key is {name, phones, candiesLiked}.
- All dependencies violate 4NF.

30

Example, Continued

- Decompose using name -> addr:

Consumers1(<u>name</u>, addr)
  - In 4NF; only dependency is name -> addr.

Consumers2(<u>name</u>, <u>phones</u>, <u>candiesLiked</u>)
  - Not in 4NF. MVD's name ->-> phones and name ->-> candiesLiked apply. No FD's, so all three attributes form the key.

(Sadly, no simple rule for projecting MVD's onto decomposed relations – use heuristics and knowledge of application)

31

Example: Decompose Consumers2

- Either MVD name ->-> phones or name ->-> candiesLiked tells us to decompose to:
  - Consumers3(<u>name</u>, <u>phones</u>)
  - Consumers4(<u>name</u>, <u>candiesLiked</u>)

32

Example

- $R = (A, B, C, G, H, I)$

  $F = \{ A \rightarrow\rightarrow B$

  $B \rightarrow\rightarrow HI$

  $CG \rightarrow\rightarrow H \}$
- R is not in 4NF since $A \rightarrow\rightarrow B$ and A is not a superkey for R

33

Example

- Decomposition

  a) $R_1 = (A, B)$ (R_1 is in 4NF)

  b) $R_2 = (A, C, G, H, I)$ (R_2 is not in 4NF)

  c) $R_3 = (C, G, H)$ (R_3 is in 4NF)

  d) $R_4 = (A, C, G, I)$ (R_4 is not in 4NF)

34

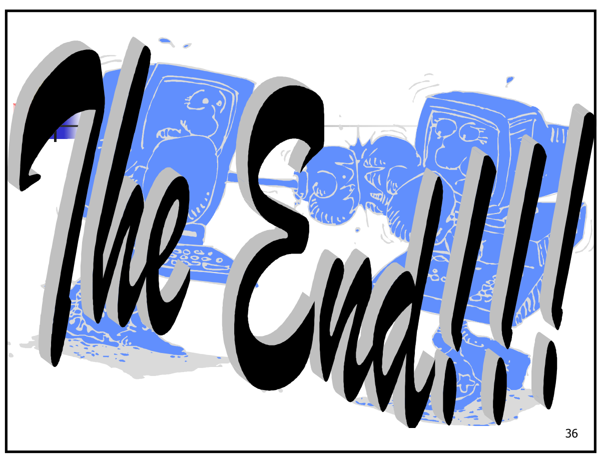
Example

- Since $A \rightarrow\rightarrow B$ and $B \rightarrow\rightarrow HI$, $A \rightarrow\rightarrow HI$, $A \rightarrow\rightarrow I$

  e) $R_5 = (A, I)$ (R_5 is in 4NF)

  f) $R_6 = (A, C, G)$ (R_6 is in 4NF)

35



36