Beyond BCNF

Data explosion problem

Employee

emp#	planguage	opsystem
100	C++	Win 2K
100	COBOL	Win 2K
100	Java	Win 2K
100	C++	OS 10
100	COBOL	OS 10
100	Java	OS 10
200	FORTRAN	Linux
200	C	Linux
200	C	Win 2K
200	FORTRAN	Win 2K

Nontrivial functional dependencies:

None

The highest normal form:

BCNF

Redundancies:

YES !

Multivalued dependency

Let $R = (A_1, \ldots, A_n)$ be a relational schema (a header of relational table) and let X, Y, Z be nonempty subsets of R

We say that multivalued dependency $X \rightarrow\rightarrow Y|Z$ is valid in relational schema R if ...

... for any relational table r with relational schema R, if for any two rows v and w in r such that $v[X] = w[X]$ there exist a row t in r such that ...

... $v[XY] = t[XY]$ and $w[XZ] = t[XZ]$

Multivalued dependency

Notations:

$X \rightarrow\rightarrow Y|Z$

X Y
X Z
X Y Z

```
CREATE VIEW R' AS (SELECT X,Y FROM R);
CREATE VIEW R" AS (SELECT X,Z FROM R);
SELECT R'.X, R'.Y, R".Z
FROM R' JOIN R"
ON R'.X = R".X
```

is always equal to R

4NF

A relational schema R is in 4NF if for every nontrivial multivalued dependency

$X \rightarrow\rightarrow Y|Z$

set of attributes X is a superkey in R

Alternative definition:
A relational schema R is in 4NF if no nontrivial multivalued dependencies are valid in schema R

4NF

Employee

emp#	planguage	opsystem
100	C++	Win 2K
100	COBOL	Win 2K
100	Java	Win 2K
100	C++	OS 10
100	COBOL	OS 10
100	Java	OS 10
200	FORTRAN	Linux
200	C	Linux
200	C	Win 2K
200	FORTRAN	Win 2K

Nontrivial functional dependencies:

None

Nontrivial multivalued dependencies:

$emp\# \rightarrow\rightarrow planguage | opsystem$

The highest normal form:

BCNF

Redundancies:

YES !

4NF

Languages

emp#	planguage
100	C++
100	COBOL
100	Java
200	FORTRAN
200	C

Systems

emp#	opsystem
100	OS 10
100	Win 2K
200	Linux
200	Win 2K

4NF

A relational schema R is in 4NF if for every nontrivial multivalued dependency

$$X \rightarrow\rightarrow Y \mid Z$$

set of attributes X is a superkey in R

Alternative definition:
A relational schema R is in 4NF if no nontrivial multivalued dependencies are valid in schema R

Join dependency

Let $R = (A_1, \ldots, A_n)$ be a relational schema (a header of relational table) and let $X, Y_1, \ldots, Y_n$ be nonempty subsets of R

We say that join dependency $\bowtie (X, Y_1, \ldots, Y_n)$ is valid in relational schema R if ...

... for any relational table r with relational schema R, if for any n rows $v_1, \ldots v_n$ in r such that $v_1[X] = \ldots = v_n[X]$ there exist a row t in r such that ...

$v_1[XY_1] = t[XY_1]$ and ... and $v_n[XY_n] = t[XY_n]$

Join dependency

Notations:

$\bowtie (X, Y_1, \ldots, Y_n)$

$X\ Y_1$
$X\ \ Y_2$
$\ldots\ \ \ldots$
$X\ \ \ \ \ \ Y_n$
$X\ Y_1\ Y_2\ \ldots\ Y_n$

5NF

A relational schema R is in 5NF if for every nontrivial join dependency

$$\bowtie (X, Y_1, \ldots, Y_n)$$

set of attributes X is a superkey in R

Alternative definition:
A relational schema R is in 5NF if no nontrivial join dependencies are valid in schema R

5NF example

Warehouse-part

Warehouse	part
Warehouse1	bolt
Warehouse2	nut
Warehouse2	bolt
Warehouse3	nut
Warehouse1	nail

Warehouse-supplier

Warehouse	Supplier
Warehouse1	Albert
Warehouse2	Albert
Warehouse2	Bob
Warehouse3	Carl
Warehouse1	Bob

Supplier-part

Supplier	part
Albert	bolt
Albert	nut
Bob	bolt
Carl	nut
Bob	nail

Higher normal forms

- 6NF!
- 7NF?!

References

Elmasri R., Navathe S. B., *Fundamentals of Database Systems*, chapters 11.3, 11.4