

CSCI235 – Database Systems

Beyond BCNF

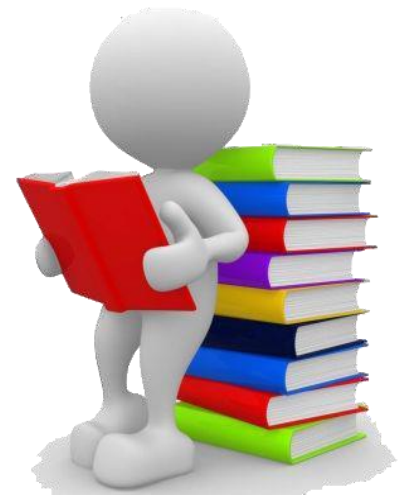
sjapit@uow.edu.au

29 September 2019



Beyond BCNF

Data Explosion Problem



Data Explosion Problem

Considering a relation schema

EMP(employee-number, programming-language, operating-system)

has no valid functional dependencies.

A relational table **EMPLOYEE** over a relational schema **EMP** contains information about the programming languages and operating systems known by employees.

Data Explosion Problem

Employee-number	Programming-language	Operating-system
200	Python	Linux
200	Java	Linux
200	Scala	Linux

Employee-number	Programming-language	Operating-system
200	Python	Windows 10
200	Java	Windows 10
200	Scala	Windows 10

Data Explosion Problem

Employee-number	Programming-language	Operating-system
200	Python	Linux
200	Java	Linux
200	Scala	Linux
200	Python	Windows 10
200	Java	Windows 10
200	Scala	Windows 10

Data Explosion Problem

Employee-number	Programming-language	Operating-system
200	Fortran	Linux
200	Fortran	Windows 10

Employee-number	Programming-language	Operating-system
200	Python	Unix
200	Java	Unix
200	Scala	Unix
200	Fortran	Unix

Data Explosion Problem

Employee-number	Programming-language	Operating-system
200	Python	Linux
200	Java	Linux
200	Scala	Linux
200	Python	Windows 10
200	Java	Windows 10
200	Scala	Windows 10
200	Fortran	Linux
200	Fortran	Windows 10
200	Python	Unix
200	Java	Unix
200	Scala	Unit
200	Fortran	Unix

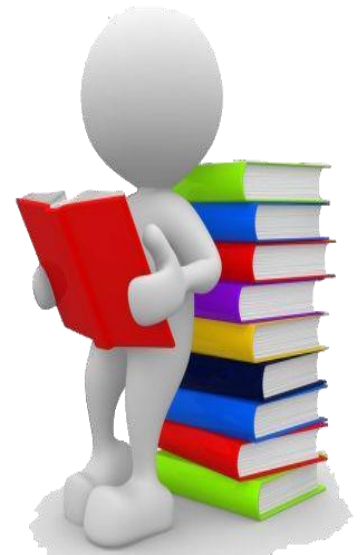
Data Explosion Problem

Which normal form?

- No valid functional dependencies means that does not exist a functional dependence such that its left hand side is not a superkey.
- The entire row will now be composited to form unique rows in the relation, and these composited attributes form a minimal superkey.
- This means there is no functional dependencies violate BCNF.
- Hence, the table is in BCNF, but the relation still contains a lot of redundancies.

Beyond BCNF

Multivalued Dependencies



Multivalued dependency

- Let $R = A_1, \dots, A_n$ be a relational schema and let X, Y, Z be nonempty subsets of R .
- We say that a multivalued functional dependency $X \twoheadrightarrow Y|Z$ is valid in a relational schema R if ...
 - for any relational table r created over a 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 $w[XY] = t[XZ]$

Multivalued dependency

Other notation

$$\frac{X \quad Y \quad \quad Z}{X \quad Y \quad Z}$$

It means that if a row $X Y \square$ is in a relational table and a row $X \square Z$ is in the same table, then a row $X Y Z$ must be in the same relational table.

Multivalued dependency

Examples

An employee knows many programming language and many operating systems.

empeNum \twoheadrightarrow *progLang* | *OS*

empeNum	progLang	
empeNum		OS
empeNum	progLang	OS

Multivalued dependency

Examples

A person owns many cars and has many skills

fName, lName \twoheadrightarrow regNumber | skill

fName, lName	regNumber	
fName, lName		skill
fName, lName	regNumber	skill

Multivalued dependency

Examples

A student has many friends and many hobbies

stdNum \twoheadrightarrow fName, lName | hobby

stdNum	fName, lName	
stdNum		hobby
stdNum	fName, lName	hobby

Multivalued dependency

X	Y	
X		Z
<hr/>		
X	Y	Z

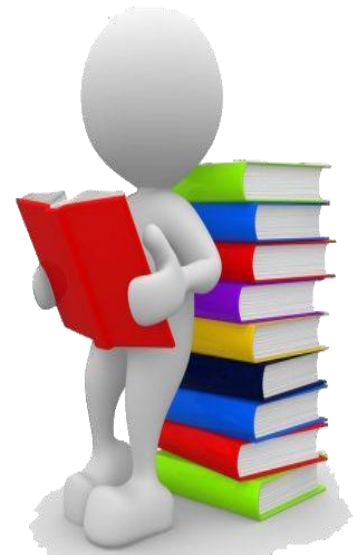
```
CREATE VIEW XY AS (SELECT X, Y FROM R);  
CREATE VIEW XZ AS (SELECT X, Z FROM R);
```

```
SELECT XY.X, XY.Y, XZ.Z  
FROM   XY JOIN XZ  
      ON XY.X=XZ.X
```

The result of SELECT is always
equal to R

Beyond BCNF

4NF (4th Normal Form)



4NF

A relational schema R is in the **Fourth Normal Form (4NF)** if for every nontrivial multivalued dependency $X \twoheadrightarrow Y|Z$ a set of attributes X is a superkey in a relational schema R

Alternative definition:

A relational schema R is in **4NF** if no nontrivial **multivalued dependencies** are valid in a relational schema R

4NF

For example, a multivalued dependency

empeNum \twoheadrightarrow *progLang* | *OS*

is valid in a relational schema

EMP(*empeNum*, *progLang*, *OS*)

Hence, the relational schema *EMP* is **NOT** in **4NF** because a nontrivial multivalued dependency is valid in *EMP*.

4NF

Decomposition into 4NF?

EMP(empNum, progLang, OS)

EPL(empNum, progLang)

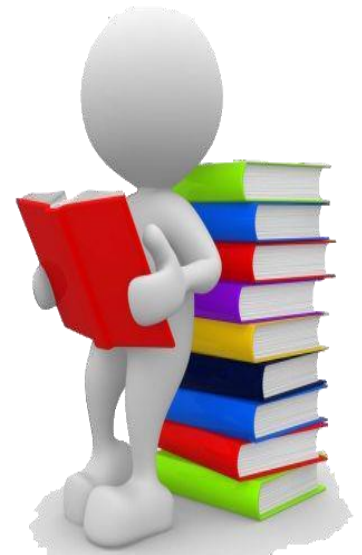
PK: (empNum, progLang)

EOS(empNum, OS)

PK: (empNum, OS)

Beyond BCNF

Join Dependency



Join dependency

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

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

... for any relational table r with relational schema R , if for any n rows v_1, \dots, v_n in r such that $v_1[X] = \dots = v_n[X]$ 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

Other notation

X	Y1				
X		Y2			
X			Y3		
...	
X	Yn
X	Y1	Y2	Y3	...	Yn

For example:

An employee knows many programming language, many OS, and has many hobbies.

⋈ (*empeNum, progLang, OS, hobby*)

empeNum	progLang		
empeNum		OS	
empeNum			hobby
empeNum	progLang	OS	hobby

Join dependency

Example:

A person owns many cars and has many skills and has many employers

⋈ $((fName, lName), regNum, skill, empr)$

fName	lName	regnum		
fName	lName		skill	
fName	lName			empr
fName	lName	regnum	skill	empr

Join dependency

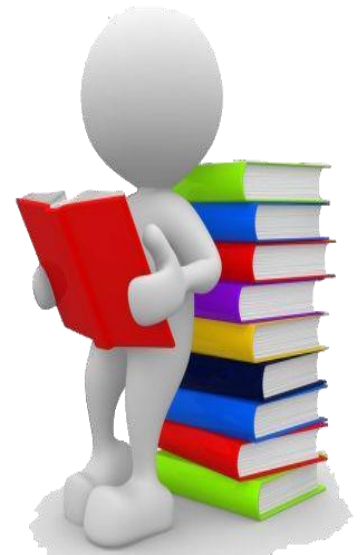
```
CREATE VIEW XY1 AS (SELECT X, Y1 FROM R);  
CREATE VIEW XY2 AS (SELECT X, Y2 FROM R);  
... ..  
CREATE VIEW Xyn AS (SELECT X, Yn FROM R);
```

```
SELECT    XY1.X, XY1.Y1, XY2.Y2, ... Xyn.Yn  
FROM      XY1  
          JOIN XY2 ON XY1.X=XY2.X  
          JOIN ...  
          JOIN Xyn ON XY1.X=Xyn.X
```

The result of **SELECT** is always equal to **R**.

Beyond BCNF

5NF (5th Normal Form)



5NF

A relational schema R is in the *Fifth Normal Form (5NF)* if for every nontrivial join dependency $\bowtie (X, Y_1, \dots, Y_n)$ a 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

For example, a join dependency $\bowtie(\text{empeNum}, \text{progLang}, \text{OS}, \text{hobby})$ is valid in a relational schema

EMP(empeNum, progLang, OS, hobby)

Hence, the relational schema *EMP* is **NOT** in **5NF** because a nontrivial join dependency is valid in *EMP*.

5NF

Decomposition into 5NF?

EMP(empeNum, progLang, OS, hobby)

EPL(empeNum, progLang)
PK: (empeNum, progLang)

EOS(empeNum, OS)
PK: (empeNum, OS)

EHB(empeNum, hobby)
PK: (empeNum, hobby)

References

- T. Connolly, C. Begg, Database Systems, A Practical Approach to Design, Implementation, and Management, Chapter 15.4 Fourth Normal Form (4NF), Chapter 15.5 Fifth Normal Form (5NF), Pearson Education Ltd, 2015