

# CSCI235 Database Systems

## Beyond BCNF

Dr Janusz R. Getta

School of Computing and Information Technology -  
University of Wollongong

# Beyond BCNF

## Outline

Data explosion problem

Multivalued dependency

4NF

Join dependency

5NF

# Data explosion problem

A relational 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

employee-number	programming-language	operating-system	EMPLOYEE
200	Python	Linux	
200	Java	Linux	
200	Scala	Linux	
200	Python	Windows 10	
200	Java	Windows 10	
200	Scala	Windows 10	

In HTML view press 'p' to see the lecture notes

[TOP](#)

Created by Janusz R. Getta, CSCI235 Database Systems, Spring 2022

3/18

# Data explosion problem

employee-number	programming-language	operating-system	EMPLOYEE
200	Python	Linux	
200	Java	Linux	
200	Scala	Linux	
200	Python	Windows	
200	Java	Windows	
200	Scala	Windows	
200	Fortran	Linux	
200	Fortran	Windows	
200	Python	Unix	
200	Java	Unix	
200	Scala	Unix	
200	Fortran	Unix	

In HTML view press 'p' to see the lecture notes

[TOP](#)

Created by Janusz R. Getta, CSCI235 Database Systems, Spring 2022

4/18

# Data explosion problem

employee-number	programming-language	operating-system	EMPLOYEE
200	Python	Linux	
200	Java	Linux	
200	Scala	Linux	
200	Python	Windows	
200	Java	Windows	
200	Scala	Windows	
200	Fortran	Linux	
200	Fortran	Windows	
200	Python	Unix	
200	Java	Unix	
200	Scala	Unix	
200	Fortran	Unix	

Normal form ?

No valid functional dependencies means that does not exists a functional dependence such that its left hand side is not a superkey

It means that no functional dependencies violate **BCNF**

**BCNF** but ... still a lot of redundancies !

[In HTML view](#) press 'p' to see the lecture notes

[TOP](#)

Created by Janusz R. Getta, CSCI235 Database Systems, Spring 2022

5/18

# Beyond BCNF

## Outline

[Data explosion problem](#)

[Multivalued dependency](#)

[4NF](#)

[Join dependency](#)

[5NF](#)

# 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 dependency  $X \twoheadrightarrow Y | Z$  is valid in 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  $w[XZ] = t[XZ]$

Other notation

X	Y	
X		Z
-----		
X	Y	Z

Multivalued dependency

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

employee-number  $\twoheadrightarrow$  programming-language | operating-system

employee-number	programming-language		Multivalued dependency
employee-number		operating-system	
<hr/>			
employee-number	programming-language	operating-system	

A person owns many cars and has many skills

first-name, last-name  $\twoheadrightarrow$  registration-number | skill

A students has many friends and many hobbies

student-number  $\twoheadrightarrow$  first-name, last-name | hobby

<pre>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</pre>	Multivalued dependency
---	------------------------

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



# Beyond BCNF

## Outline

[Data explosion problem](#)

[Multivalued dependency](#)

[4NF](#)

[Join dependency](#)

[5NF](#)

## 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  $R$

A multivalued dependency

$\text{employee-number} \twoheadrightarrow \text{programming-language} | \text{operating-system}$

is valid in a relational schema

$\text{EMP}(\text{employee-number}, \text{programming-language}, \text{operating-system})$

A relational schema  $\text{EMP}$  is **NOT** in **4NF** because a nontrivial multivalued dependency is valid in  $\text{EMP}$

# 4NF

Decomposition into 4NF ?

EPGM(employee-number, programming-language),  
EOPS(employee-number, operating-system)

# Beyond BCNF

## Outline

[Data explosion problem](#)

[Multivalued dependency](#)

[4NF](#)

[Join dependency](#)

[5NF](#)

# 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  $v_n[XY_n] = t[XY_n]$

Other notation

X	Y1				
X		Y2			
X			Y3		
...	...	...	...		
X	...	...	...	...	YN
<hr/>					
X	Y1	Y2	Y3	...	YN

Multivalued dependency

# Join dependency

## Examples

⌋(employee-number, programming-language, operating-system, hobby)

employee-number	programming-language			Join dependency
employee-number		operating-system		
employee-number			hobby	
<hr/>				
employee-number	programming-language	operating-system	hobby	

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

⌋((first-name, last-name), registration-number, skill, employer)

```
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

## Outline

[Data explosion problem](#)

[Multivalued dependency](#)

[4NF](#)

[Join dependency](#)

[5NF](#)

## 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  $R$

A join dependency  $\bowtie(\text{employee-number}, \text{programming-language}, \text{operating-system}, \text{hobby})$  is valid in a relational schema  $\text{EMP}(\text{employee-number}, \text{programming-language}, \text{operating-system}, \text{hobby})$

A relational schema  $\text{EMP}$  is **NOT** in **5NF** because a nontrivial join dependency is valid in  $\text{EMP}$



# 5NF

Decomposition into 5NF ?

EPGM(employee-number, programming-language),  
EOPS(employee-number, operating-system),  
EHOB(employee-number, 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