

# Data Management Systems

## Introduction to Design Theory

Matteo Devigili

June 4<sup>th</sup>, 2020

# Agenda

- Functional Dependencies
- Data Anomalies
- Normal Forms:
  - i 1NF
  - ii BCNF
  - iii 3NF

# Functional Dependencies

"If two tuples of  $R$  agree on all of the attributes  $A_1, A_2, \dots, A_n$  then they must also agree on all of another list of attributes  $B_1, B_2, \dots, B_m$ . We write this FD formally as  $A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_m$  and say that:

- ▷  $A_1, A_2, \dots, A_n$  functionally determine  $B_1, B_2, \dots, B_m$ "

Garcia-Molina, Ullman, Widom 2008

# Example

## Courses

Table: Courses

Name	Year	Weeks	Degree
DMS	2019/2020	6	Business Analytics
DMS	2019/2020	6	Actuarial Science
DMS	2019/2020	6	Actuarial Management
D-Viz	2019/2020	10	Business Analytics
DMS	2018/2019	2	Business Analytics
D-Viz	2018/2019	4	Business Analytics
D-Viz	2018/2019	4	Actuarial Management

What is the **FD**?

# Example

## Courses

Table: Courses

Name	Year	Weeks	Degree
DMS	2019/2020	6	Business Analytics
DMS	2019/2020	6	Actuarial Science
DMS	2019/2020	6	Actuarial Management
D-Viz	2019/2020	10	Business Analytics
DMS	2018/2019	2	Business Analytics
D-Viz	2018/2019	4	Business Analytics
D-Viz	2018/2019	4	Actuarial Management

*name year* → *weeks*

# Example

## Courses

Table: Courses

Name	Year	Weeks	Degree
DMS	2019/2020	6	Business Analytics
DMS	2019/2020	6	Actuarial Science
DMS	2019/2020	6	Actuarial Management
D-Viz	2019/2020	10	Business Analytics
DMS	2018/2019	2	Business Analytics
D-Viz	2018/2019	4	Business Analytics
D-Viz	2018/2019	4	Actuarial Management

What about:

*name year*  $\rightarrow$  *degree*

# Keys & Superkeys

- 🔑 "A superkey of a relation schema  $R = \{ A_1, A_2, \dots, A_n \}$  is a set of attributes  $S \subseteq R$  with the property that no two tuples  $t_1$  and  $t_2$  in any legal relation state  $r$  of  $R$  will have  $t_1[S] = t_2[S]$ "
- 🔑 "A **key**  $K$  is a superkey with the additional property that removal of any attribute from  $K$  will cause  $K$  not to be a superkey anymore" (a key has to be minimal)

Elmasri, Ramez, and Shamkant B. Navathe 2016

# Example

## Keys & Superkeys

Table: Courses

Name	Year	Weeks	Degree	Count
DMS	2019/2020	6	Business Analytics	75
DMS	2019/2020	6	Actuarial Science	6
DMS	2019/2020	6	Actuarial Management	16
D-Viz	2019/2020	10	Business Analytics	75
DMS	2018/2019	2	Business Analytics	37
D-Viz	2018/2019	4	Business Analytics	75
D-Viz	2018/2019	4	Actuarial Management	10

a possible superkey: {name, year, weeks, degree}

the key: {name, year, degree}



# Functional Dependencies

So what?

- i Look for FDs;
- ii Use FDs to design better relation schemas.

# Data Anomalies

- *Redundancy*: unnecessary repetition of information;
- *Update Anomalies*: we may replace information of a tuple, but forget about others;
- *Deletion Anomalies*: after deleting, we may accidentally lose some other information.

# Example

## Redundancy

Table: Courses

Name	Year	Term	Weeks	Degree
DMS	2019/2020	T3	6	Business Analytics
DMS	2019/2020	T3	6	Actuarial Science
DMS	2019/2020	T3	6	Actuarial Management
D-Viz	2019/2020	T1	10	Business Analytics
DMS	2018/2019	T3	2	Business Analytics
D-Viz	2018/2019	T2	4	Business Analytics
D-Viz	2018/2019	T2	4	Actuarial Management

# Example

## Update Anomalies

Table: Courses

Name	Year	Term	Weeks	Degree
DMS	2019/2020	T3	5	Business Analytics
DMS	2019/2020	T3	6	Actuarial Science
DMS	2019/2020	T3	6	Actuarial Management
D-Viz	2019/2020	T1	10	Business Analytics
DMS	2018/2019	T3	2	Business Analytics
D-Viz	2018/2019	T2	4	Business Analytics
D-Viz	2018/2019	T2	4	Actuarial Management

# Example

## Deletion Anomalies

Table: Courses

Name	Year	Term	Weeks	Degree
DMS	2019/2020	T3	6	<del>Business Analytics</del>
DMS	2019/2020	T3	6	Actuarial Science
DMS	2019/2020	T3	6	Actuarial Management
<del>D-Viz</del>	2019/2020	<del>T1</del>	10	<del>Business Analytics</del>
DMS	2018/2019	T3	2	<del>Business Analytics</del>
<del>D-Viz</del>	2018/2019	T2	4	<del>Business Analytics</del>
D-Viz	2018/2019	T2	4	Actuarial Management

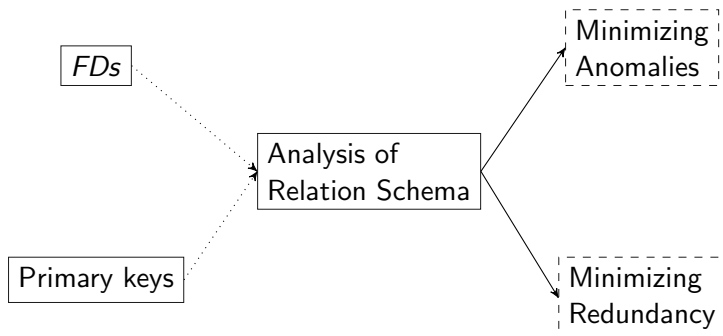
# Decomposition

A possible decomposition:

Name	Year	Term	Weeks
DMS	2019/2020	T3	6
D-Viz	2019/2020	T1	10
DMS	2018/2019	T3	2
D-viz	2018/2019	T2	4

Name	Year	Degree
DMS	2019/2020	Business Analytics
DMS	2019/2020	Actuarial Science
DMS	2019/2020	Actuarial Management
D-Viz	2019/2020	Business Analytics
DMS	2018/2019	Business Analytics
D-Viz	2018/2019	Business Analytics
D-Viz	2018/2019	Actuarial Management

# Normalization of data



# 1NF

## Atomic values

For the 1NF:

*"the domain of an attribute must include only **atomic** (simple, indivisible) values and that the value of any attribute in a tuple must be a single value from the domain of that attribute"*

Elmasri, Ramez, and Shamkant B. Navathe 2016



# 1NF

## Example

Name	Year	Weeks	Degree
DMS	2019/2020	6	{Business Analytics, Actuarial Science, Actuarial Management}
D-Viz	2019/2020	10	Business Analytics
DMS	2018/2019	2	Business Analytics
D-Viz	2018/2019	4	{Business Analytics, Actuarial Management}

This violates **1NF**

# 1NF

## PostgreSQL - Arrays

*"PostgreSQL allows columns of a table to be defined as variable-length multidimensional arrays"*

PostgreSQL 8.15

# Boyce-Codd Normal Form

*"A relation  $R$  is in BCNF if and only if: whenever there is a nontrivial FD  $A_1, A_2 \dots A_n \rightarrow B_1, B_2 \dots, B_m$  for  $R$ , it is the case that  $A_1, A_2 \dots A_n$  is a superkey for  $R$ "*

Garcia-Molina, Ullman, Widom 2008

# BCNF

## Example

Name	Year	Term	Weeks	Degree
DMS	2019/2020	T3	6	Business Analytics
DMS	2019/2020	T3	6	Actuarial Science
DMS	2019/2020	T3	6	Actuarial Management
D-Viz	2019/2020	T1	10	Business Analytics
DMS	2018/2019	T3	2	Business Analytics
D-Viz	2018/2019	T2	4	Business Analytics
D-Viz	2018/2019	T2	4	Actuarial Management

The (super) key is {name year degree}

So, the existence of {name year}  $\rightarrow$  {term weeks} violates BCNF.

# BCNF

## Example

When do we stop decomposing?

Name	Year	Term	Weeks
DMS	2019/2020	T3	6
D-Viz	2019/2020	T1	10
DMS	2018/2019	T3	2
D-viz	2018/2019	T2	4

Name	Year	Degree
DMS	2019/2020	Business Analytics
DMS	2019/2020	Actuarial Science
DMS	2019/2020	Actuarial Management
D-Viz	2019/2020	Business Analytics
DMS	2018/2019	Business Analytics
D-Viz	2018/2019	Business Analytics
D-Viz	2018/2019	Actuarial Management

# Decomposition

## Trade-off

- 1 Elimination of Anomalies
- 2 Recoverability of Information
- 3 Preservation of Dependencies

Garcia-Molina, Ullman, Widom 2008

# 3NF

*"Whenever  $A_1, A_2 \dots A_n \rightarrow B_1, B_2 \dots, B_m$  is a non-trivial FD, either  $A_1, A_2 \dots A_n$  is a superkey, or those of  $B_1, B_2, \dots B_m$  that are not among the A's, are each a member of some key (not necessarily the same key)."*

Garcia-Molina, Ullman, Widom 2008

# Normal Forms

1<sup>st</sup> Normal Form

2<sup>nd</sup> Normal Form

3<sup>rd</sup> Normal Form

Boyce-Codd Normal Form

4<sup>th</sup> Normal Form

5<sup>th</sup> Normal Form

...



# References

- 1 Hector Garcia-Molina, Jeff Ullman, and Jennifer Widom. Database Systems: The Complete Book, Pearson, 2008.
- 2 Elmasri, Ramez, and Shamkant B. Navathe. Fundamentals of Database Systems, Global Edition, Pearson Education Limited, 2016.