

Data Management Systems

Introduction to Design Theory

Matteo Devigili

June 2nd, 2023

Agenda

- ▶ Functional Dependencies
- ▶ Data Anomalies
- ▶ Normal Forms:
 1. 1NF
 2. BCNF
 3. 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

Name	Year	Weeks	Degree
NLP	2021/2022	7	Business Analytics
DMS	2021/2022	6	Business Analytics
DMS	2021/2022	6	Actuarial Science
DMS	2021/2022	6	Actuarial Management
D-Viz	2021/2022	6	Business Analytics
D-Viz	2021/2022	6	Actuarial Management
DMS	2020/2021	2	Business Analytics
D-Viz	2020/2021	4	Business Analytics

What is the **FD**?

Example

Name	Year	Weeks	Degree
NLP	2021/2022	7	Business Analytics
DMS	2021/2022	6	Business Analytics
DMS	2021/2022	6	Actuarial Science
DMS	2021/2022	6	Actuarial Management
D-Viz	2021/2022	6	Business Analytics
D-Viz	2021/2022	6	Actuarial Management
DMS	2020/2021	2	Business Analytics
D-Viz	2020/2021	4	Business Analytics

name year \rightarrow *weeks*

Example

Name	Year	Weeks	Degree
NLP	2021/2022	7	Business Analytics
DMS	2021/2022	6	Business Analytics
DMS	2021/2022	6	Actuarial Science
DMS	2021/2022	6	Actuarial Management
D-Viz	2021/2022	6	Business Analytics
D-Viz	2021/2022	6	Actuarial Management
DMS	2020/2021	2	Business Analytics
D-Viz	2020/2021	4	Business Analytics

What about:
name year \rightarrow *degree*

Key

*“We say a set of one or more attributes $\{A_1, A_2, \dots, A_n\}$ is a **key** for a relation R if:*

- 1. Those attributes functionally determine **all other attributes** of the relation. That is, it is impossible for two distinct tuples of R to agree on all of A_1, A_2, \dots, A_n .*
- 2. No proper subset of $\{A_1, A_2, \dots, A_n\}$ functionally determines all other attributes of R ; i.e., a key must be **minimal**.”*

Garcia-Molina, Ullman, Widom 2008

Superkey

A **Superkey** satisfies the first condition:

1. *Those attributes functionally determine all other attributes of the relation. That is, it is impossible for two distinct tuples of R to agree on all of A_1, A_2, \dots, A_n .*

Garcia-Molina, Ullman, Widom 2008

Example

Name	Year	Weeks	Degree	Count
NLP	2021/2022	7	Business Analytics	57
DMS	2021/2022	6	Business Analytics	45
DMS	2021/2022	6	Actuarial Science	15
DMS	2021/2022	6	Actuarial Management	9
D-Viz	2021/2022	6	Business Analytics	58
D-Viz	2021/2022	6	Actuarial Management	19
DMS	2020/2021	2	Business Analytics	10
D-Viz	2020/2021	4	Business Analytics	80

the key: {name, year, degree}

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

Functional Dependencies

So what?

- ▶ Look for FDs;
- ▶ Use FDs to design better relation schemas;
- ▶ Pay attention to local FDs!

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

Name	Year	Term	Weeks	Degree
NLP	2021/2022	T3	7	Business Analytics
DMS	2021/2022	T3	6	Business Analytics
DMS	2021/2022	T3	6	Actuarial Science
DMS	2021/2022	T3	6	Actuarial Management
D-Viz	2021/2022	T1	6	Business Analytics
D-Viz	2021/2022	T1	6	Actuarial Management
DMS	2020/2021	T2	2	Business Analytics
D-Viz	2020/2021	T2	4	Business Analytics

Example

Update Anomalies

Name	Year	Term	Weeks	Degree
NLP	2021/2022	T3	7	Business Analytics
DMS	2021/2022	T3	5	Business Analytics
DMS	2021/2022	T3	6	Actuarial Science
DMS	2021/2022	T3	6	Actuarial Management
D-Viz	2021/2022	T1	6	Business Analytics
D-Viz	2021/2022	T1	6	Actuarial Management
DMS	2020/2021	T2	2	Business Analytics
D-Viz	2020/2021	T2	4	Business Analytics

Example

Deletion Anomalies

Name	Year	Term	Weeks	Degree
NLP	2021///2022	T3	7	Business/Analytics
DMS	2021///2022	T3	6	Business/Analytics
DMS	2021/2022	T3	6	Actuarial Science
DMS	2021/2022	T3	6	Actuarial Management
D-Viz	2021///2022	T1	6	Business/Analytics
D-Viz	2021/2022	T1	6	Actuarial Management
DMS	2020///2021	T2	2	Business/Analytics
D-Viz	2020///2021	T2	4	Business/Analytics

Decomposition

A possible decomposition:

Name	Year	Term	Weeks
NLP	2021/2022	T3	7
DMS	2021/2022	T3	6
D-Viz	2021/2022	T1	6
DMS	2020/2021	T2	2
D-Viz	2020/2021	T2	4

Name	Year	Degree
NLP	2021/2022	Business Analytics
DMS	2021/2022	Business Analytics
DMS	2021/2022	Actuarial Science
DMS	2021/2022	Actuarial Management
D-Viz	2021/2022	Business Analytics
D-Viz	2021/2022	Actuarial Management
DMS	2020/2021	Business Analytics
D-Viz	2020/2021	Business Analytics

References

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