



Managing Data & Databases

Session 17

Data Modelling and Normalization

Building a database from the scratch...

- Database design is good when it is good for its specific use case
- That's why different types of databases exist
 - Relational
 - Object Relational
 - Document Databases (No-SQL)
 - Graph Databases
 - Dimensional Data for Analytics and Business Intelligence
- That's why you must take normalization as a guide, and not an inviolable set of rules (nonetheless, you need to master it to know when not to apply it)


What is normalization?

- A rather intuitive set of rules that help
 - Moving from single file-single list to relational data
 - Reducing redundancy
 - Improving data consistency
 - Sustaining data integrity over time and across rounds of data entry and updates
 - Making it possible to access structured data using SQL queries
 - Reducing complexity (I mean it!)
- But
 - Not always the most straightforward manner to save data!
 - Most importantly: May sacrifice speed and sometimes ease of use

Normalization...

What do we put in a single field?





Student Name
Abdelkader Hassan, Omar
Ahamed, Saharea
Amilhastre, Liam
Bajwa, Shehryar
Barber, Emily
Easton, Laura
Frank, Joshua
Fryml, Matthew
Guérette, Xavier

What not to do

Violation of the 1st Normal Form



B Com Major	Marketing *	UG Level Year 3	(438) 9:
B Com Major	Finance	UG Level Year 1 New	(0092)
B Com Major	Information Systems *	UG Level Year 3	(705) 8:
B Com Major	Information Systems	UG Level Year 3	(514) 7:
B Com Major	Information Systems *	UG Level Year 1Returning	(914) 5:
B Com Major	Information Systems	UG Level Year 3	(450) 6:
B Com General Management	General Management	UG Level Year 3	(450) 6:
B Com Major	Marketing	UG Level Year 2	(613) 3:
B Com Major	Information Systems *	UG Level Year 2	(418) 9:
B Com Major	Information Systems	UG Level Year 2	N/A
B Com Major	Finance	UG Level Year 2	N/A
B Com General Management	General Management	UG Level Year 3	(514) 4:
B Com Major	Information Systems	UG Level Year 3	(514) 4:
Exchange/Special/Visiting Mgmt	Exchange	Undergraduate Non-Degree	Intl:642

What not to do

Violation of the 1st Normal Form

Normalization...

What do we put in a single table?



course_enrollment


<i>student_id</i>	<i>course_id</i>	<i>course_title</i>	<i>stu_name</i>	<i>enrolled_on</i>
2423423	322	Flora Management	Mina	01/01/2015
5435432	544	Fauna Management	Pina	01/01/2015
5435432	322	Flora Management	Pina	01/01/2015
7465532	654	Jungle Strategy	Zina	03/01/2015

What not to do

Violation of the 2nd Normal Form

students

<i>student_id</i>	<i>first_name</i>	<i>last_name</i>	<i>group</i>	<i>exam_grade</i>	<i>project_grade</i>
2423423	Mina	Molson	1	A	B-
5435432	Pina	Paulson	2	A++	C
5465433	Sina	Samson	2	A+++	C
7465532	Zina	Zubrowka	1	Z	B-



What not to do

Violation of the 3rd Normal Form

student_interests

<i>student_id</i>	<i>subject</i>	<i>art</i>
2423423	<i>Flora Management</i>	<i>Lavatory Design</i>
5435432	<i>Fauna Management</i>	<i>Music</i>
5435432	<i>Flora Management</i>	<i>Dance</i>
7465532	<i>Jungle Strategy</i>	<i>Steampunk Produktion</i>

What not to do

Violation of the 4th Normal Form

Normalization...

No redundancy...



<i>student_id</i>	<i>first_name</i>	<i>last_name</i>	<i>group</i>	<i>courses_taken</i>
2423423	Mina	Molson	1	1
5435432	Pina	Paulson	2	2
5465433	Sina	Samson	2	0
7465532	Zina	Zubrowka	1	1

This data can be generated using other data in the database, so it's unnecessary

What not to do

Violation of the 5th Normal Form

<i>student_id</i>	<i>course_id</i>	<i>enrolment_date</i>
2423423	435	01/02/2015
5435432	534	02/02/2015
5435432	435	02/02/2015
7465532	345	27/01/2015

And Denormalization?

It happens...