

Explanation of the Database Normalization Process

Let's break down a messy student data table and tidy it up into a well-organised format using the normalisation process. The goal is to reduce repetition and make sure everything is structured cleanly for future analysis or database use. We'll walk through 1NF, 2NF, and 3NF, showing you what the table looks like at each stage and what changes were made.

Un-normalised Form (UNF) Here's the original table. As you can see, it contains repeating values in the course-related columns:

Student Number	Student Name	Exam Score	Support	Date of Birth	Course Name	Exam Boards	Teacher Name
1001	Bob Baker	78	No	25/08/2001	Computer Science, Maths	BCS, EdExcel	Mr Jones, Ms Parker
1002	Sally Davies	55	Yes	02/10/1999	Physics, Maths, Biology, Music	OCR, AQA, WJEC, AQA	Mr Peters, Ms Parker, Mrs Patel, Ms Daniels
1003	Mark Hanmill	90	No	05/06/1995	Computer Science, Maths, Physics	BCS, EdExcel, OCR	Mr Jones, Ms Parker, Mr Peters
1004	Anas Ali	70	No	03/08/1980	Maths, Physics, Biology	AQA, OCR, WJEC	Ms Parker, Mr Peters, Mrs Patel
1005	Cheuk Yin	45	Yes	01/05/2002	Computer Science, Maths, Music	BCS, EdExcel, AQA	Mr Jones, Ms Parker, Ms Daniels

First Normal Form (1NF) To get into 1NF, we need to remove the repeating groups so each row has a single course per student.

Student Number	Student Name	Exam Score	Support	Date of Birth	Course Name	Exam Board	Teacher Name
1001	Bob Baker	78	No	25/08/2001	Computer Science	BCS	Mr Jones
1001	Bob Baker	78	No	25/08/2001	Maths	EdExcel	Ms Parker
1002	Sally Davies	55	Yes	02/10/1999	Physics	OCR	Mr Peters
1002	Sally Davies	55	Yes	02/10/1999	Maths	AQA	Ms Parker
1002	Sally Davies	55	Yes	02/10/1999	Biology	WJEC	Mrs Patel
1002	Sally Davies	55	Yes	02/10/1999	Music	AQA	Ms Daniels
1003	Mark Hanmill	90	No	05/06/1995	Computer Science	BCS	Mr Jones
1003	Mark Hanmill	90	No	05/06/1995	Maths	EdExcel	Ms Parker
1003	Mark Hanmill	90	No	05/06/1995	Physics	OCR	Mr Peters
1004	Anas Ali	70	No	03/08/1980	Maths	AQA	Ms Parker
1004	Anas Ali	70	No	03/08/1980	Physics	OCR	Mr Peters
1004	Anas Ali	70	No	03/08/1980	Biology	WJEC	Mrs Patel
1005	Cheuk Yin	45	Yes	01/05/2002	Computer Science	BCS	Mr Jones
1005	Cheuk Yin	45	Yes	01/05/2002	Maths	EdExcel	Ms Parker

What Changed:

- Split rows to remove multiple values.
- Ensured each cell holds one piece of data only.

Second Normal Form (2NF) Now we remove partial dependencies—any column that depends only on part of the key is moved to a new table.

Table: Students

Student Number (Primary Key)	Student Name	Exam Score	Support	Date of Birth
1001	Bob Baker	78	No	25/08/2001
1002	Sally Davies	55	Yes	02/10/1999
1003	Mark Hanmill	90	No	05/06/1995
1004	Anas Ali	70	No	03/08/1980
1005	Cheuk Yin	45	Yes	01/05/2002

Table: Student_Courses

Student Number (Primary & Foreign Key)	Course Name (Foreign Key)
1001	Computer Science
1001	Maths
1002	Physics
1002	Maths
1002	Biology
1002	Music
1003	Computer Science
1003	Maths
1003	Physics
1004	Maths
1004	Physics
1004	Biology
1005	Computer Science
1005	Maths
1005	Music

Table: Course_Details

Course Name (Primary Key)	Exam Board	Teacher Name
Computer Science	BCS	Mr Jones
Maths	EdExcel	Ms Parker
Physics	OCR	Mr Peters
Biology	WJEC	Mrs Patel
Music	AQA	Ms Daniels

What Changed:

- Moved course-related data to its own table.
 - Linked students to their courses in a mapping table.
-

Third Normal Form (3NF) Let's go one step further and remove any indirect (transitive) dependencies.

Table: Students (same as before)

Student Number (Primary Key)	Student Name	Exam Score	Support	Date of Birth
1001	Bob Baker	78	No	25/08/2001
1002	Sally Davies	55	Yes	02/10/1999
1003	Mark Hanmill	90	No	05/06/1995
1004	Anas Ali	70	No	03/08/1980
1005	Cheuk Yin	45	Yes	01/05/2002

Table: Courses

Course ID (Primary Key)	Course Name
1	Computer Science
2	Maths
3	Physics
4	Biology
5	Music

Table: Exam_Boards

Board ID (Primary Key)	Exam Board
1	BCS
2	EdExcel
3	OCR
4	WJEC
5	AQA

Table: Teachers

Teacher ID (Primary Key)	Teacher Name
1	Mr Jones
2	Ms Parker
3	Mr Peters
4	Mrs Patel
5	Ms Daniels

Table: Course_Details

Course ID (Foreign Key)	Board ID (Foreign Key)	Teacher ID (Foreign Key)
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5

Table: Student_Courses

Student Number (Foreign Key)	Course ID (Foreign Key)
1001	1
1001	2
1002	3
1002	2
1002	4
1002	5
1003	1
1003	2
1003	3
1004	2
1004	3
1004	4
1005	1
1005	2
1005	5

What Changed:

- Broke out courses, boards, and teachers into separate entities with IDs.
 - Final schema is tidy, modular, and avoids unnecessary duplication.
-

Conclusion We've taken a messy, multi-value student table and step-by-step turned it into a clean set of relational tables that are much easier to manage. It's now ready for querying, analysis, or loading into a database.

Reflections on Learning Outcomes:

- Understood how to clean and organise raw data for structure.
- Applied each stage of database normalisation effectively.
- Created a professional, scalable data design suitable for real-world applications.