Normalisation Task

Original Dataset

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

1NF

We go about creating the first normal form where each row has to be unique and every value is atomic, meaning that only one value should be stored.

We need to expand the table to satisfy first normal form, a row needs to be available for each course name, exam board and teacher name. We can see that we end up with three entries per student.

Student Number	First Name	Last Name	Exam Score	Support	Date of Birth	Course Name	Exam Boards	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
1001	Bob	Baker	78	No	25/08/2001	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
1005	Cheuk	Yin	45	Yes	01/05/2002	Music	AQA	Ms Daniels

2NF

Everything in the 1NF table is related to the primary key and every value is atomic. The entirety of this table does not conform to second form if we take the course name it would not relate to the student number module if we were to make it the primary key.

To conform to second normal form the data will need to be restructured into two tables, The first table are details related to the students and the second table is related to teachers they are made relational by the course name

Student Number	First Name	Last Name	Exam Score	Support	Date of Birth	Course Name
1001	Bob	Baker	78	No	25/08/2001	Computer Science
1001	Bob	Baker	78	No	25/08/2001	Maths
1001	Bob	Baker	78	No	25/08/2001	Physics
1002	Sally	Davies	55	Yes	02/10/1999	Maths
1002	Sally	Davies	55	Yes	02/10/1999	Biology
1002	Sally	Davies	55	Yes	02/10/1999	Music
1003	Mark	Hanmill	90	No	05/06/1995	Computer Science
1003	Mark	Hanmill	90	No	05/06/1995	Maths
1003	Mark	Hanmill	90	No	05/06/1995	Physics
1004	Anas	Ali	70	No	03/08/1980	Maths
1004	Anas	Ali	70	No	03/08/1980	Physics
1004	Anas	Ali	70	No	03/08/1980	Biology
1005	Cheuk	Yin	45	Yes	01/05/2002	Computer Science
1005	Cheuk	Yin	45	Yes	01/05/2002	Maths
1005	Cheuk	Yin	45	Yes	01/05/2002	Music

Course Name	Exam Boards	Teacher Name
Computer Science	BCS	Mr Jones
Maths	EdExcel	Ms Parker
Physics	OCR	Mr Peters
Maths	AQA	Ms Parker
Biology	WJEC	Mrs Patel
Music	AQA	Ms Daniels

3NF

For the third normal form we need to ensure that each column is dependent on only the primary key. To conform to this, we need to separate data into three tables so that each table is dependent on only the primary key. The three tables are a student table, a course table with exam scores and a courses taught by which teacher table.

Students Tables

Student Number (PK)	First Name	Last Name	Support	Date of Birth
1001	Bob	Baker	No	25/08/2001
1002	Sally	Davies	Yes	02/10/1999
1003	Mark	Hanmill	No	05/06/1995
1004	Anas	Ali	No	03/08/1980
1005	Cheuk	Yin	Yes	01/05/2002

Student Scores

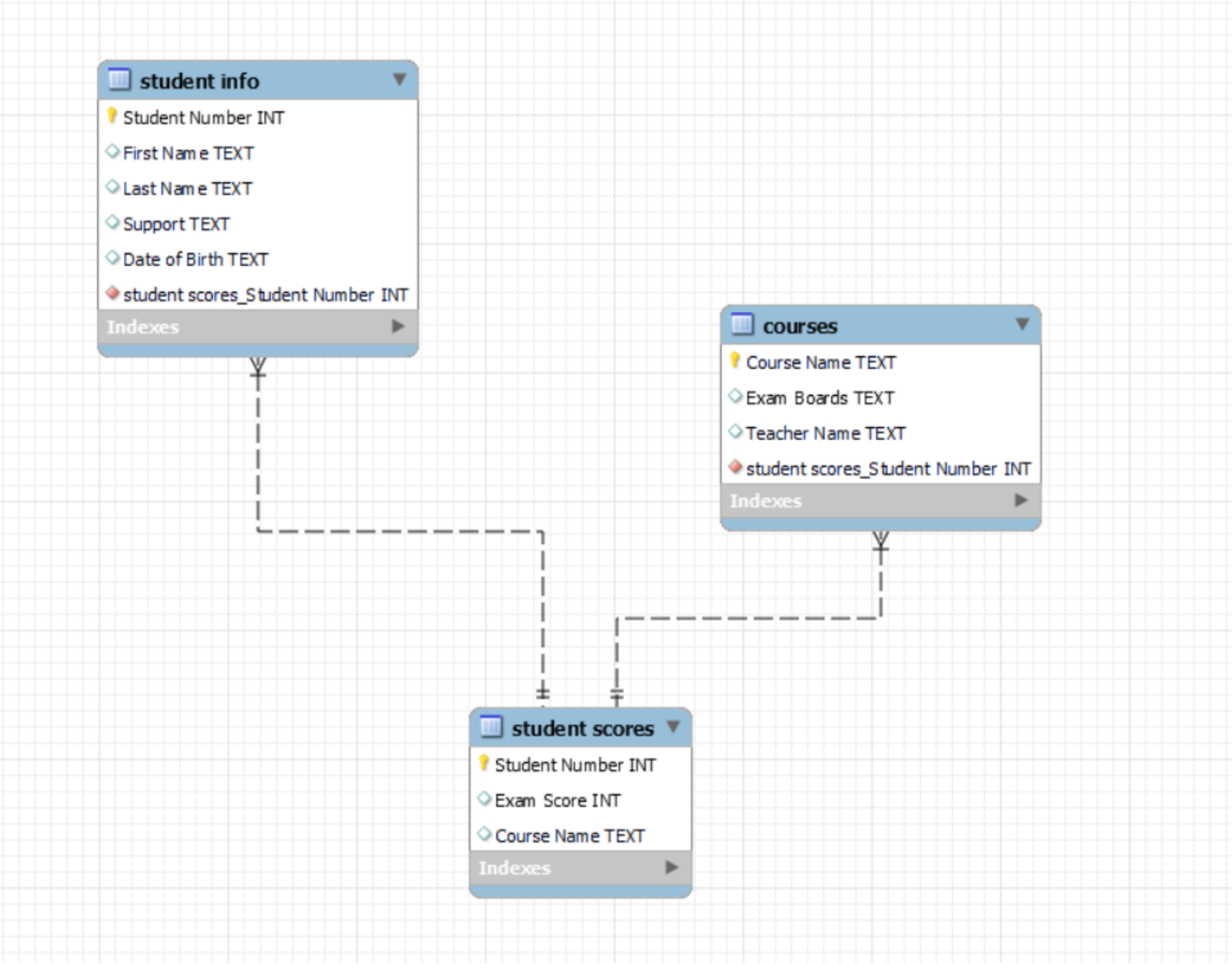
Student Number (PK)	Exam Score	Course Name
1001	78	Computer Science
1001	78	Maths
1001	78	Physics
1002	55	Maths
1002	55	Biology
1002	55	Music
1003	90	Computer Science
1003	90	Maths
1003	90	Physics
1004	70	Maths
1004	70	Physics
1004	70	Biology
1005	45	Computer Science
1005	45	Maths
1005	45	Music

Courses Taught

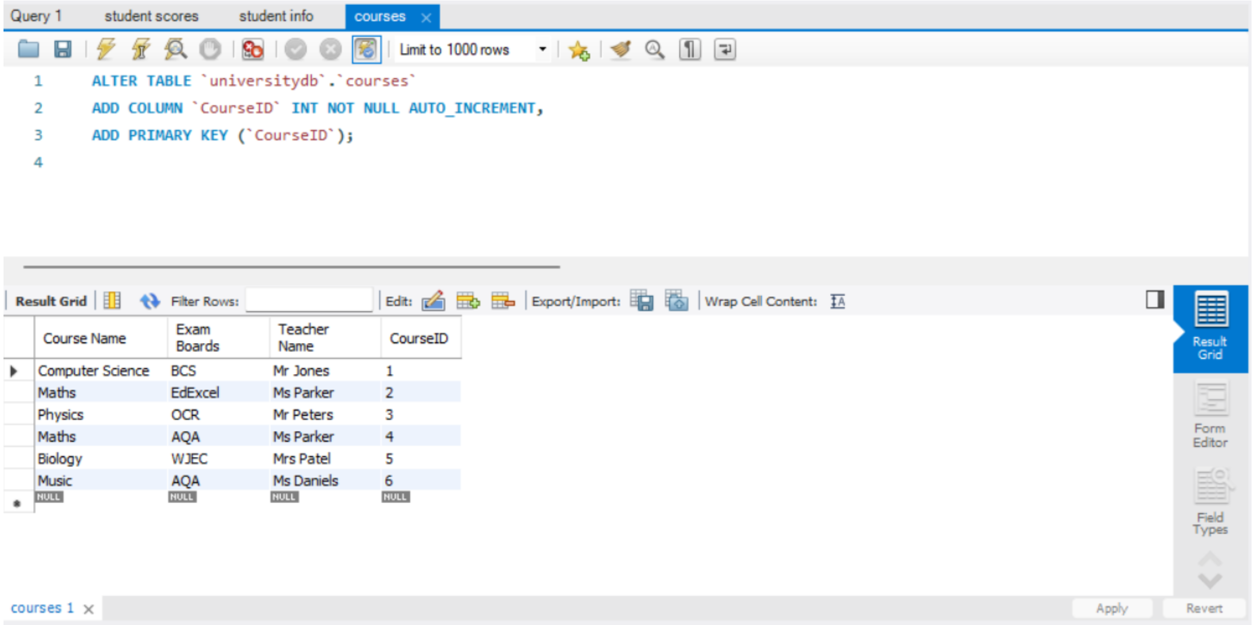
Course Name (PK)	Exam Boards	Teacher Name
Computer Science	BCS	Mr Jones
Maths	EdExcel	Ms Parker
Physics	OCR	Mr Peters
Maths	AQA	Ms Parker
Biology	WJEC	Mrs Patel
Music	AQA	Ms Daniels

Data Build Task

Entity Relation Diagram

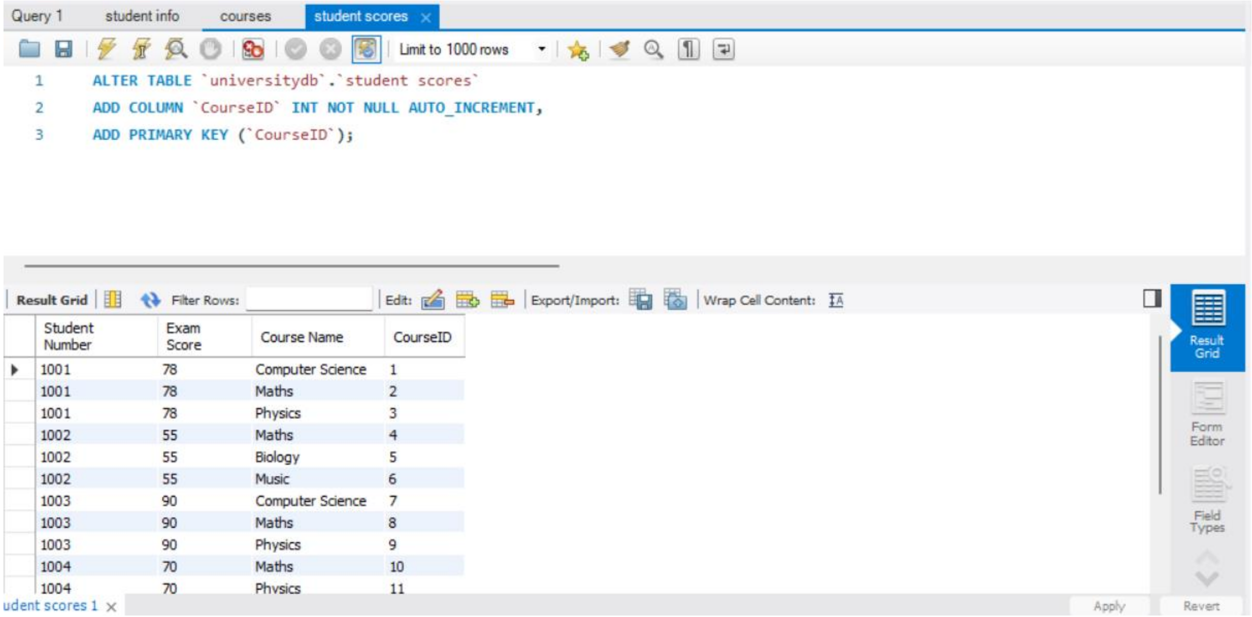


Surrogate primary keys created as duplicate values within columns let to an error message within MySQL



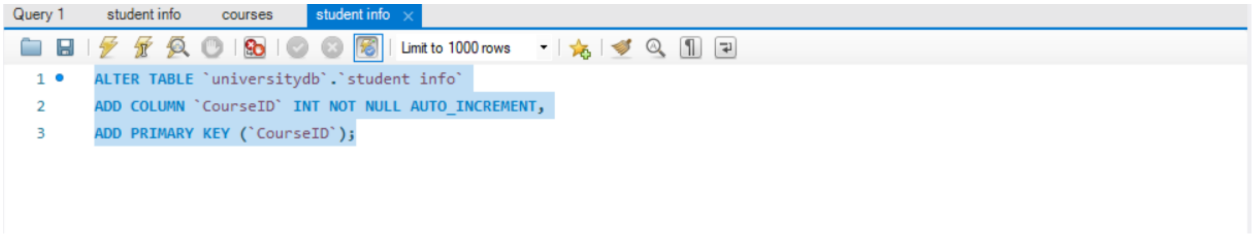
```sql
ALTER TABLE `universitydb`.`courses`
ADD COLUMN `CourseID` INT NOT NULL AUTO_INCREMENT,
ADD PRIMARY KEY (`CourseID`);
```

Course Name	Exam Boards	Teacher Name	CourseID
Computer Science	BCS	Mr Jones	1
Maths	EdExcel	Ms Parker	2
Physics	OCR	Mr Peters	3
Maths	AQA	Ms Parker	4
Biology	WJEC	Mrs Patel	5
Music	AQA	Ms Daniels	6
NULL	NULL	NULL	NULL



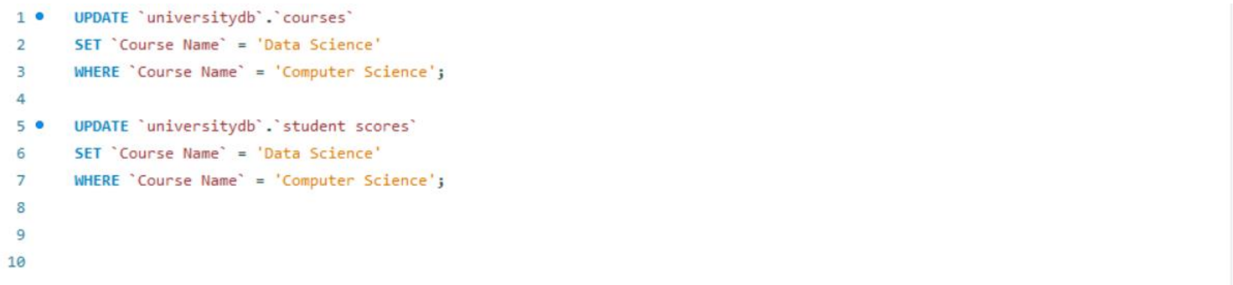
```sql
ALTER TABLE `universitydb`.`student scores`
ADD COLUMN `CourseID` INT NOT NULL AUTO_INCREMENT,
ADD PRIMARY KEY (`CourseID`);
```

Student Number	Exam Score	Course Name	CourseID
1001	78	Computer Science	1
1001	78	Maths	2
1001	78	Physics	3
1002	55	Maths	4
1002	55	Biology	5
1002	55	Music	6
1003	90	Computer Science	7
1003	90	Maths	8
1003	90	Physics	9
1004	70	Maths	10
1004	70	Physics	11



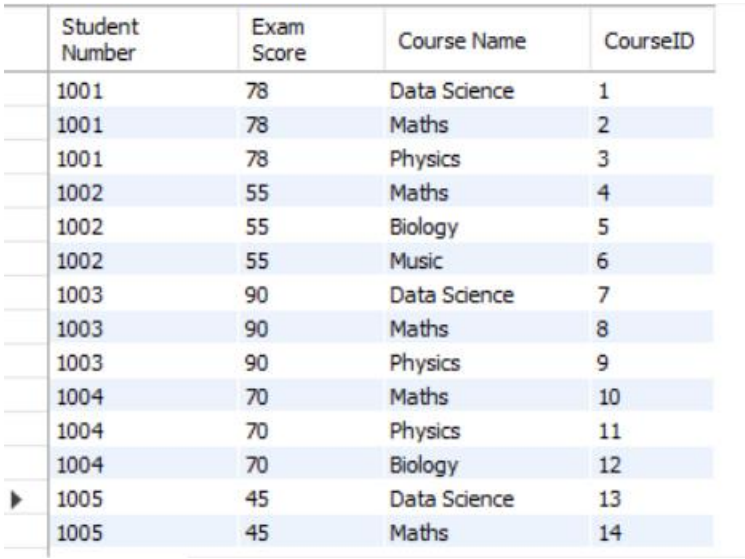
```sql
ALTER TABLE `universitydb`.`student info`
ADD COLUMN `CourseID` INT NOT NULL AUTO_INCREMENT,
ADD PRIMARY KEY (`CourseID`);
```

For example, if I wanted to change the name of the computer science course to Data science as the course may be changing then because the database is relational it will update all tables with data science.

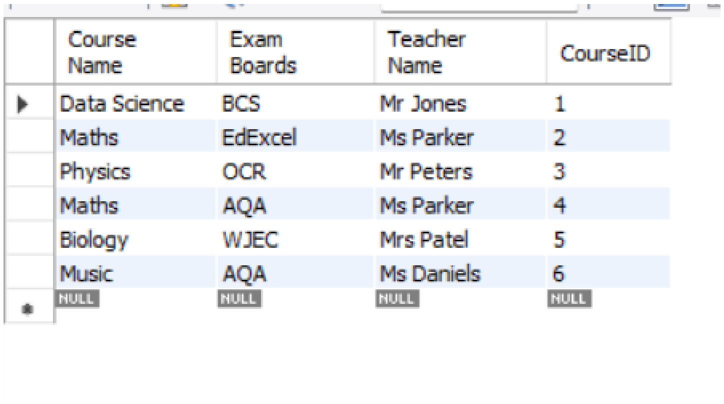


```
UPDATE `universitydb`.`courses`
SET `Course Name` = 'Data Science'
WHERE `Course Name` = 'Computer Science';

UPDATE `universitydb`.`student scores`
SET `Course Name` = 'Data Science'
WHERE `Course Name` = 'Computer Science';
```



Student Number	Exam Score	Course Name	CourseID
1001	78	Data Science	1
1001	78	Maths	2
1001	78	Physics	3
1002	55	Maths	4
1002	55	Biology	5
1002	55	Music	6
1003	90	Data Science	7
1003	90	Maths	8
1003	90	Physics	9
1004	70	Maths	10
1004	70	Physics	11
1004	70	Biology	12
1005	45	Data Science	13
1005	45	Maths	14



Course Name	Exam Boards	Teacher Name	CourseID
Data Science	BCS	Mr Jones	1
Maths	EdExcel	Ms Parker	2
Physics	OCR	Mr Peters	3
Maths	AQA	Ms Parker	4
Biology	WJEC	Mrs Patel	5
Music	AQA	Ms Daniels	6
NULL	NULL	NULL	NULL