

Solution: Normalization in SQL

Beginning Dataset

| studentID | studentName | campusAddress | studentPhone | courseNum | classes | instructor | major | scores1 |
|-----------|-------------|---------------|--------------|------------------------|-----------------------|------------------------|-------|------------|
| 1001 | Alice | Laferre Hall | 573-884-1234 | IT1040, IT3330, IT4005 | Python, Java, Android | Prof A, Prof B, Prof C | IT | 90, 86, 92 |
| 1002 | Bob | Laferre Hall | 573-884-2235 | IT1040, IT4005 | Python, Android | Prof A, Prof C | IT | 83, 77 |
| 1003 | Carla | Naka Hall | 573-884-3236 | CS1050, IT3330 | C, Java | Prof D, Prof B | CS | 84, 89 |
| 1004 | Devon | Naka Hall | 573-884-4237 | IT1040 | Python | Prof A | CS | 93 |
| 1005 | Emilio | Laferre Hall | 573-884-5238 | IT3330, IT4425 | Java, iOS | Prof B, Prof E | IT | 95, 79 |

1NF

| All values must be Atomic Values | | | | | | | | |
|----------------------------------|-------------|-------|--------------|---------------|-----------|------------|------------------|-------|
| studentID | studentName | major | studentPhone | campusAddress | courseNum | instructor | courseTitle | score |
| 1001 | Alice | IT | 573-884-1234 | Laferre Hall | IT1040 | Prof A | Intro to Python | 90 |
| 1001 | Alice | IT | 573-884-1234 | Laferre Hall | IT3380 | Prof B | Intro to Java | 86 |
| 1001 | Alice | IT | 573-884-1234 | Laferre Hall | IT4005 | Prof C | Intro to Android | 92 |
| 1002 | Bob | IT | 573-884-2235 | Laferre Hall | IT1040 | Prof A | Intro to Python | 83 |
| 1002 | Bob | IT | 573-884-2235 | Laferre Hall | IT4005 | Prof C | Intro to Android | 77 |
| 1003 | Carla | CS | 573-884-3236 | Naka Hall | CS1050 | Prof D | Intro to C | 84 |
| 1003 | Carla | CS | 573-884-3236 | Naka Hall | IT3380 | Prof B | Intro to Java | 69 |
| 1004 | Devon | CS | 573-884-4237 | Naka Hall | IT1040 | Prof A | Intro to Python | 93 |
| 1005 | Emilio | IT | 573-884-5238 | Laferre Hall | IT3380 | Prof B | Intro to Java | 95 |
| 1005 | Emilio | IT | 573-884-5238 | Laferre Hall | IT4425 | Prof E | Intro to iOS | 79 |

| Remove repeating groups | | | | |
|-------------------------|-------------|-------|--------------|---------------|
| studentID | studentName | major | studentPhone | campusAddress |
| 1001 | Alice | IT | 573-884-1234 | Laferre Hall |
| 1002 | Bob | IT | 573-884-2235 | Laferre Hall |
| 1003 | Carla | CS | 573-884-3236 | Naka Hall |
| 1004 | Devon | CS | 573-884-4237 | Naka Hall |
| 1005 | Emilio | IT | 573-884-5238 | Laferre Hall |

| studentID | courseNum | instructor | courseTitle | score |
|-----------|-----------|------------|------------------|-------|
| 1001 | IT1040 | Prof A | Intro to Python | 90 |
| 1001 | IT3380 | Prof B | Intro to Java | 86 |
| 1001 | IT4005 | Prof C | Intro to Android | 92 |
| 1002 | IT1040 | Prof A | Intro to Python | 83 |
| 1002 | IT4005 | Prof C | Intro to Android | 77 |
| 1003 | CS1050 | Prof D | Intro to C | 84 |
| 1003 | IT3380 | Prof B | Intro to Java | 69 |
| 1004 | IT1040 | Prof A | Intro to Python | 93 |
| 1005 | IT3380 | Prof B | Intro to Java | 95 |
| 1005 | IT4425 | Prof E | Intro to iOS | 79 |

Remove Repeating groups
and place in a separate table

The studentID column
becomes a foreign key in the
resulting table to reference
the students table.
studentID is also part of the
primary key in this table

2NF

Eliminate partial dependencies

| studentID | studentName | major | studentPhone | campusAddress |
|-----------|-------------|-------|--------------|---------------|
| 1001 | Alice | IT | 573-884-1234 | Laferre Hall |
| 1002 | Bob | IT | 573-884-2235 | Laferre Hall |
| 1003 | Carla | CS | 573-884-3236 | Naka Hall |
| 1004 | Devon | CS | 573-884-4237 | Naka Hall |
| 1005 | Emilio | IT | 573-884-5238 | Laferre Hall |

This table is already in 2NF

| studentID | courseNum | instructor | courseTitle | score |
|-----------|-----------|------------|------------------|-------|
| 1001 | IT1040 | Prof A | Intro to Python | 90 |
| 1001 | IT3380 | Prof B | Intro to Java | 86 |
| 1001 | IT4005 | Prof C | Intro to Android | 92 |
| 1002 | IT1040 | Prof A | Intro to Python | 83 |
| 1002 | IT4005 | Prof C | Intro to Android | 77 |
| 1003 | CS1050 | Prof D | Intro to C | 84 |
| 1003 | IT3380 | Prof B | Intro to Java | 69 |
| 1004 | IT1040 | Prof A | Intro to Python | 93 |
| 1005 | IT3380 | Prof B | Intro to Java | 95 |
| 1005 | IT4425 | Prof E | Intro to iOS | 79 |

Split the Table

A partial dependency exists because instructor and courseTitle depend on courseNum and not studentID. We break those columns into their own table.

| studentID | courseNum | score |
|-----------|-----------|-------|
| 1001 | IT1040 | 90 |
| 1001 | IT3380 | 86 |
| 1001 | IT4005 | 92 |
| 1002 | IT1040 | 83 |
| 1002 | IT4005 | 77 |
| 1003 | CS1050 | 84 |
| 1003 | IT3380 | 69 |
| 1004 | IT1040 | 93 |
| 1005 | IT3380 | 95 |
| 1005 | IT4425 | 79 |

| courseNum | instructor | courseTitle |
|-----------|------------|------------------|
| IT1040 | Prof A | Intro to Python |
| IT3380 | Prof B | Intro to Java |
| IT4005 | Prof C | Intro to Android |
| CS1050 | Prof D | Intro to C |
| IT4425 | Prof E | Intro to iOS |

3NF

Eliminate transitive dependencies

| students | | | | |
|-----------|-------------|-------|--------------|---------------|
| studentID | studentName | major | studentPhone | campusAddress |
| 1001 | Alice | IT | 573-884-1234 | Laferre Hall |
| 1002 | Bob | IT | 573-884-2235 | Laferre Hall |
| 1003 | Carla | CS | 573-884-3236 | Naka Hall |
| 1004 | Devon | CS | 573-884-4237 | Naka Hall |
| 1005 | Emilio | IT | 573-884-5238 | Laferre Hall |

A transitive dependency exists because campusAddress depends more on the student's major/department than it does their studentID. We split the table

| studentID | studentName | major | studentPhone |
|-----------|-------------|-------|--------------|
| 1001 | Alice | IT | 573-884-1234 |
| 1002 | Bob | IT | 573-884-2235 |
| 1003 | Carla | CS | 573-884-3236 |
| 1004 | Devon | CS | 573-884-4237 |
| 1005 | Emilio | IT | 573-884-5238 |

| department | buildingName |
|------------|--------------|
| IT | Laferre Hall |
| CS | Naka Hall |

| scores | | | |
|-----------|-----------|-------|--|
| studentID | courseNum | score | |
| 1001 | IT1040 | 90 | |
| 1001 | IT3380 | 86 | |
| 1001 | IT4005 | 92 | |
| 1002 | IT1040 | 83 | |
| 1002 | IT4005 | 77 | |
| 1003 | CS1050 | 84 | |
| 1003 | IT3380 | 69 | |
| 1004 | IT1040 | 93 | |
| 1005 | IT3380 | 95 | |
| 1005 | IT4425 | 79 | |

This table is already in 3NF

| courses | | |
|-----------|------------|------------------|
| courseNum | instructor | courseTitle |
| IT1040 | Prof A | Intro to Python |
| IT3380 | Prof B | Intro to Java |
| IT4005 | Prof C | Intro to Android |
| CS1050 | Prof D | Intro to C |
| IT4425 | Prof E | Intro to iOS |

This table is already in 3NF

Final Tables

The following solutions are all acceptable.

Solution 1

The following solution uses only the column information provided in the beginning dataset.

| students | | | | scores | | |
|-----------|-------------|------------------|----------------|--------------|-----------|-------|
| studentID | studentName | major | studentPhone | studentID | courseNum | score |
| 1001 | Alice | IT | 573-884-1234 | 1001 | IT1040 | 90 |
| 1002 | Bob | IT | 573-884-2235 | 1001 | IT3380 | 86 |
| 1003 | Carla | CS | 573-884-3236 | 1001 | IT4005 | 92 |
| 1004 | Devon | CS | 573-884-4237 | 1002 | IT1040 | 83 |
| 1005 | Emilio | IT | 573-884-5238 | 1002 | IT4005 | 77 |
| | | | | 1003 | CS1050 | 84 |
| | | | | 1003 | IT3380 | 69 |
| | | | | 1004 | IT1040 | 93 |
| | | | | 1005 | IT3380 | 95 |
| | | | | 1005 | IT4425 | 79 |
| courses | | | major_location | | | |
| courseNum | instructor | courseTitle | department | buildingName | | |
| IT1040 | Prof A | Intro to Python | IT | Laferre Hall | | |
| IT3380 | Prof B | Intro to Java | CS | Naka Hall | | |
| IT4005 | Prof C | Intro to Android | | | | |
| CS1050 | Prof D | Intro to C | | | | |
| IT4425 | Prof E | Intro to iOS | | | | |

Solution 2

There was not an instructorID column in the original dataset and if you look at the courses table from solution 1 you would recognize that a deletion anomaly is possible. For example, if you deleted the IT4425 iOS course we would delete any record of Prof E. The implication from solution 1 is that there would be a table for instructors similar to the students table where all instructor information is stored. We could clarify that assumption and address the possibility of the deletion anomaly by creating an instructors table as shown in solution 2.

| students | | | | scores | | |
|--------------|----------------|------------------|--------------|----------------|--------------|-------|
| studentID | studentName | major | studentPhone | studentID | courseNum | score |
| 1001 | Alice | IT | 573-884-1234 | 1001 | IT1040 | 90 |
| 1002 | Bob | IT | 573-884-2235 | 1001 | IT3380 | 86 |
| 1003 | Carla | CS | 573-884-3236 | 1001 | IT4005 | 92 |
| 1004 | Devon | CS | 573-884-4237 | 1002 | IT1040 | 83 |
| 1005 | Emilio | IT | 573-884-5238 | 1002 | IT4005 | 77 |
| | | | | 1003 | CS1050 | 84 |
| | | | | 1003 | IT3380 | 69 |
| | | | | 1004 | IT1040 | 93 |
| | | | | 1005 | IT3380 | 95 |
| | | | | 1005 | IT4425 | 79 |
| courses | | | | major_location | | |
| courseNum | instructor_ID | courseTitle | | department | buildingName | |
| IT1040 | 1 | Intro to Python | | IT | Laferre Hall | |
| IT3380 | 2 | Intro to Java | | CS | Naka Hall | |
| IT4005 | 3 | Intro to Android | | | | |
| CS1050 | 4 | Intro to C | | | | |
| IT4425 | 5 | Intro to iOS | | | | |
| instructors | | | | | | |
| instructorID | instructorName | | | | | |
| 1 | Prof A | | | | | |
| 2 | Prof B | | | | | |
| 3 | Prof C | | | | | |
| 4 | Prof D | | | | | |
| 5 | Prof E | | | | | |

Solution 3

Similar to the explanation for solution 2, there was not a buildingID column in the original dataset and we did not include one in solution 2. If you look at the major_location table from solution 2 you would recognize that a deletion anomaly and/or update anomaly is possible. For example, if we deleted the second row and there was no CS department we lose any record of Naka Hall. Similarly, if the CS department were to move locations to another building we would lose any record of Naka Hall, or if Naka Hall's name was changed we would have to update all instances of Naka Hall in the database. We can address these possible anomalies the same way we did in solution 2 and create a buildings table as shown below.

| students | | | | scores | | |
|--------------|----------------|------------------|--------------|----------------|--------------|-------|
| studentID | studentName | major | studentPhone | studentID | courseNum | score |
| 1001 | Alice | IT | 573-884-1234 | 1001 | IT1040 | 90 |
| 1002 | Bob | IT | 573-884-2235 | 1001 | IT3380 | 86 |
| 1003 | Carla | CS | 573-884-3236 | 1001 | IT4005 | 92 |
| 1004 | Devon | CS | 573-884-4237 | 1002 | IT1040 | 83 |
| 1005 | Emilio | IT | 573-884-5238 | 1002 | IT4005 | 77 |
| | | | | 1003 | CS1050 | 84 |
| | | | | 1003 | IT3380 | 69 |
| | | | | 1004 | IT1040 | 93 |
| | | | | 1005 | IT3380 | 95 |
| | | | | 1005 | IT4425 | 79 |
| courses | | | | major_location | | |
| courseNum | instructor_ID | courseTitle | | department | buildingID | |
| IT1040 | 1 | Intro to Python | | IT | 1 | |
| IT3380 | 2 | Intro to Java | | CS | 2 | |
| IT4005 | 3 | Intro to Android | | | | |
| CS1050 | 4 | Intro to C | | | | |
| IT4425 | 5 | Intro to iOS | | | | |
| instructors | | | | buildings | | |
| instructorID | instructorName | | | buildingID | buildingName | |
| 1 | Prof A | | | 1 | Laferre Hall | |
| 2 | Prof B | | | 2 | Naka Hall | |
| 3 | Prof C | | | | | |
| 4 | Prof D | | | | | |
| 5 | Prof E | | | | | |