

# Coursework commentary

## 2018–2019

### CO2209 Database systems

### Coursework assignment 2

#### General remarks

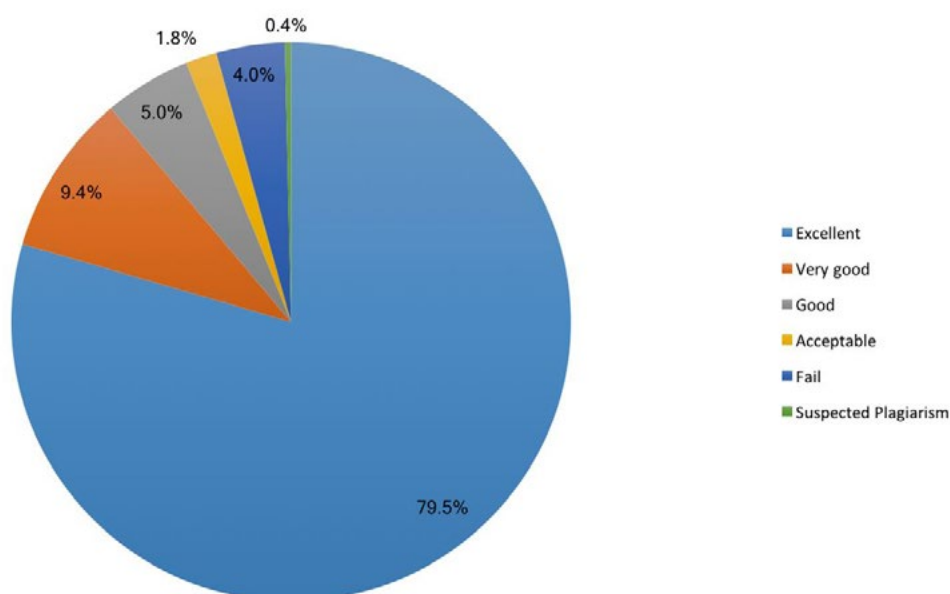
Coursework assignment 2 extended the problems in coursework assignment 1 by giving students a real, as opposed to a 'toy', database to work with, the well-known 'Mondial' database with thousands of records and several dozen relations.

A few students submitted only part of the coursework assignment and therefore lost marks. This may have been the result of poor planning, leaving too much to the last minute. It's highly advisable to begin your coursework as soon as you can, even if this means only writing down the main headings of the work you will submit. From then on, the coursework should be approached piecemeal. That is, instead of thinking, 'I have to work on coursework 2', think 'I have to do a little bit on question 1 of coursework 2.' Then complete the coursework assignment a bit at a time, working on it half an hour one night, a full hour another, perhaps only fifteen minutes on a third night. Don't think it has to be done in order: do the easy questions first and come back to the harder ones. And remember that the online forum is always available to extended explanations for anything that is not clear.

As a general rule, these coursework assignments, as opposed to written examinations, are designed so that anyone who works at them conscientiously, availing themselves of help from the web, and from the course discussion board where questions may be asked and clarifications sought, will be able to excel by getting most of the marks. On this particular coursework, 4 per cent of students failed, and nearly 80 per cent achieved an *Excellent* mark, which is in line with previous years.

See cohort mark distribution for 2018–2019 below:

**CO2209 CW2 Cohort mark distribution 2018-19**



## Comments on specific questions

### Part A

A student's first task in this coursework assignment, questions 1 and 2 of Part A, was to look at the accompanying descriptive diagrams of the data which they had downloaded: an E/R diagram, a referential dependency diagram, and a relational schema. The coursework questions should have led you to see that the diagrams and description, on the one hand, and the actual database, on the other, did not quite correspond. Furthermore, there was some ambiguous vocabulary naming the attributes of the relations, with the same name being used for somewhat different domains. The lesson here was, in learning about a database, don't take anything for granted. Whether this was made entirely clear is not certain, however. Most students noted the discrepancies, but did not comment on them, but then nor were they invited to. In any case, almost everyone who attempted this part of the coursework gained most of the available marks.

The next part of the coursework assignment, questions 3 and 4, simply required students to carry out the commands which would reveal the structure and size of the relations making up the database, and then answer some questions about it. Although this course is mainly concerned with the logical side of database design, it's also important for students to understand that the actual size of a database in bytes is a critical metric. Two databases can have exactly the same structure, but if the size of one is measured in kilobytes and the other, in gigabytes, the problems facing the Database Administrator will be quite different, because the 'performance' of the database system will vary enormously. Almost everyone carried out this task, but a few students, despite the evidence to the contrary being in front of their eyes, asserted that the size of a database is purely a function of its degree and cardinality. However, this is not true. Two relations may have exactly the same number of tuples and attributes, but if one is recording integer data, and the other, strings of text, the amount of disc space they occupy will necessarily be very different.

The last section of Part A, question 5 consisted of 10 questions about the database that could be answered by SQL queries. A few students lost marks because they ignored the instructions to list the natural language query, and the SQL to answer it, and the first few tuples generated by the query.

Questions that were answered incorrectly or only partially correctly by some students included Question 5(c) – when finding countries that had at least one religion in common with Japan, some students forgot to include the SQL clause that would exclude Japan itself as one of these countries. Although the query did not say, 'countries other than Japan' which had a religion in common with Japan, surely by the time students have reached university-attendance age they can figure out for themselves that including a country in a list of countries that have something in common with it is not exercising proper judgement.

Another question where some students did not gain marks was Question 5(e). It's easy to formulate an SQL query that will take the first country in a list, and glue it to the lowest inflation rate of those countries. This sort of error could be easily avoided if students would actually look at the data their queries generate and compare their results to the data in the tables instead of just uncritically accepting whatever the omniscient computer generates.

Despite the explicit hint that Question 5(g) required a set difference, with a subject guide page given for reference, quite a few students stumbled on this one. This sort of question – find all instances of an entity which do NOT

fulfil some criterion – is almost guaranteed to be on the examination paper. There is a “sucker-bait” answer involving the “not equals” operator, but it is not applicable in a one-to-many relationship, where only a set difference will find the answer.

Question 5(h) required some investigation and discussion. Two very similar queries were presented, which produced very different results. It generated some very good answers and some very poor ones. A few answers just pointed out that the queries were different because the second one contained an additional clause. The best answers resulted from an inspection of the actual data generated by the two queries – one giving a nonsensical answer and one giving a plausible answer.

## **Part B**

This part of the coursework assignment required students to think about the consequences of different choices for the Primary Key. This subject really reaches the heart of relational database design, and, while the concept of the primary key is simple, it's not always easy to grasp its implications. Most students grasped it and those who did were rewarded later when the same basic concept appeared in the examination. The lesson here is: the coursework helps you to do well in the actual examination.

## **Part C**

This part addressed the problem of functional dependency and normalization. It was a standard question of the sort that appears year after year in both coursework assignments and examinations. Most students got most of it right, but there were some errors which turned up again and again. Perhaps one kind of error is the following: in the table presented in the coursework assignment, sheep were weighed periodically and their weight on that date recorded. So there was a functional dependency from SheepID + Date to WeighingDate, because on a given date, a particular sheep has only one weight. However, some students saw a functional dependency from SheepID to Weight, without the Date. Probably they reasoned thus: any time I weigh a sheep it has only one weight. Therefore, SheepID determines Weight. But the functional dependency relates to the data in the table, not to the actual entities represented by the data in the table. A given sheep has only one physical manifestation in the here-and-now, and only one weight at that point. But in the table, a given sheep is represented over and over, once for each date on which it is weighed.

## **Part D**

The final part of the coursework assignment was bookwork – submitting definitions for some key database terminology. Almost everyone did well here except the very few who neglected to submit anything at all for this part, clearly a result of starting too late since this section could probably have been completed in under half an hour. Please follow the advice in ‘General remarks’.