Examiners' commentaries 2017–2018

CO2209 Database Systems – Zone B

General remarks

This examination should have held no surprises to anyone who had revised, especially if they had revised using past examination papers. A pleasing number of candidates did this and posted their proposed solutions to some past examination questions on the Student Discussion Forum, where they were answered at length, both by a lecturer and other students, hopefully clearing up some confusion. Probably the most important thing that those preparing for the examination can do is to follow the discussions on this board, even if they make no contribution themselves. One of the great advantages of an 'online' course such as this one is that 'class discussion' can actually be facilitated because it takes place in durable written form.

In any case, results for the examination were in line with previous results: a few abject failures, most of which seemed to be caused by a complete lack of any revision, and a small number of outstanding papers, with the great majority of marks distributed between the (increased) pass mark and the higher realms of excellence. Which is to say, the great majority of candidates received a passing mark. It should be said that some of the typical errors made on this exam provide food for thought for coursework setters in future.

Comments on specific questions

Question 1

This question allowed candidates to demonstrate their proficiency in the current standard database query language, SQL, and most rose to the challenge. As always, the difference between GROUP BY and ORDER BY eluded some, and the unique properties of the 'NULL' non-value caught others out, as did the tricky problem of 'negative queries' where the seduction of framing a query with the 'not equals' operator seduced a few. Some candidates had unfortunately not revised how to create a VIEW.

After the queries, an example was given of a poorly-performing database system, with the description of one relation practically demanding a horizontal partition to speed up operations on it. Gratifyingly, most candidates saw this. Another section asked about replication in distributed databases, and its justification. Again, this was dealt with satisfactorily by most. One section asked for a brief description of 'NoSQL' database systems – great detail was not expected here, since practical engagement with these increasingly-important but still minority systems was not part of the coursework. Almost all candidates were able to describe the circumstances that might suggest the employment of these systems.

Question 2

This question probed candidates' understanding of functional determinacy and primary key choice. Although this is the heart of database design, it has always proved a difficult topic for every student of the subject to grasp fully, and so it was on this examination.

The usual confusions arose with respect to describing 'anomalies' in the relation given as an example. Too many candidates thought that an 'anomaly' was just any possible error. But in the database world, the term is used for three different possible errors that only arise with an unnormalized relation.

Another area where some candidates were caught out was in identifying 'partial' and 'transitive' dependencies. This was probably more due to the vocabulary used to describe them than to inability to see them when the occurred. Thus, some candidates normalized the unnormalized relation without any problem, which meant they picked out and corrected its partial and transitive dependencies, but did not have the vocabulary to describe what they were doing.

In any case, most candidates were able to normalize the original relation, although some could not properly identify the Primary Keys. A note for anyone preparing for future exams: there is no such thing as a relation without a Primary Key. If none is designated by the designer, then the whole tuple is the Primary Key.

Most candidates were able to split the relation in the last part of this question – itself technically BCNF-compliant – into two relations that avoided its difficulties.

Question 3

This proved to be a fairly easy question for almost all candidates, who were able to take the description of a business and create a valid Entity/Relationship diagram from it, and then go on to derive an appropriate relational schema from either the diagram or the original description. Of course, there were some scripts that revealed confusion about just what a Primary Key is, and these were particularly obvious in answers to the third and last section of this question. However, only a small number candidates failed to get the majority of available marks.

Question 4

This question offered a relation and invited candidates to comment upon the bad consequences flowing from several proposed (wrong) Primary Keys. Candidates revising for future exams should remember: the Primary Key keeps out impossible data and lets in possible data. A poor choice of Primary Key fails in one, or the other, of these tasks: it either allows tuples which cannot exist in real life to be added, or it prevents valid tuples from being added. Some candidates did not seem to realize this. Those who didn't then could not give a valid criticism of the proposal to get around the key choice issue by making every tuple unique with a 'surrogate key', that is, a unique system-generated number as an additional attribute.

Other parts of this question asked if a relation's size, that is, the amount of storage space it occupies, is only a function of its degree and cardinality. This issue had been covered in the coursework and so it was pleasing to see that most candidates understood that it was not. Other parts asked about 'foreign keys', 'data dictionaries', levels of schemas, and physical data independence. Anyone not able to provide at least book-definitions for these concepts had simply not revised. Fortunately, there were few examples of this, as most candidates gained almost all the available marks here.

Question 5

The last question asked about the concurrent access problem. Most candidates were able to give satisfactory definitions and explanations where these were required, but an uncomfortable number were not able to see

where these definitions actually applied to a concrete example. This was given as a series of database operations, of which some sub-sequences were 'transactions' in the technical database sense and others were not. Candidates performed better on the remaining part of the question, which involved discussion of certain problems and solutions to them that arise in context of concurrent access.