Examiners' commentary 2018–2019

CO2209 Database systems – Zone A

General remarks

This year's examination did not present any surprises for those who completed the coursework and revised from previous years' examinations. A few candidates actively, and hopefully many more passively, took advantage of the course discussion board on the VLE to present proposed solutions to past examinations – both the VLE tutor and fellow students commented on these solutions at length. This was a particularly pertinent method for revision, because the same kinds of problems that proved challenging in past papers, also proved challenging in this year's paper, although perhaps not so challenging for those who followed the discussions on the VLE.

A small number of candidates presumed on the goodwill, and the eyesight, of the examiners, by submitting a script covered with barely-legible writing. One paper's answers were written in letters no larger, by measurement, than three millimetres.

For those reading this report in the hope of doing well in the next examination, it cannot be emphasized enough that close attention to past papers and past examination commentaries, and at least an acquaintance with the course discussion board, plus completion of the coursework assignments, is the key to success.

Comments on specific questions

Question 1

This question was all about primary keys, an understanding of which is at the heart of relational database design. It was generally answered in a satisfactory manner, although the wording of its first part was evidently difficult for some candidates to interpret. A tendency to fall back on the phrase 'would violate the uniqueness property' revealed itself. As an answer to a question this generality should usually be accompanied by a concrete example.

Question 2

Question 2 required candidates to review the case where a transaction (in the database sense) is interrupted by a system failure. It's a standard problem with a full explanation in the subject guide.

Some rather strange answers were received in response to the first part of section A, which simply asked for the definition of a 'checkpoint'. It may be the case that candidates did not adequately understand that in a typical database system, query results are first stored in fast but volatile storage, and only saved to permanent (compared to RAM) storage from time to time, due to the far longer time that disc access requires.

It might help to simply attach the concept of 'dump-to-disc' to the word 'checkpoint'.

Most candidates attempted the second part of this question, but many appeared not to have adequately revised all the cases. Quite a few answers offered the same case twice with only superficial differences.

The next part of the question was also standard and answered well although once again, some candidates resorted to guessing for some of the meaning of the words of this acronym.

The final part of the question was also standard. Where there was weakness in answers here, it was mainly in descriptions of the two-phase locking protocol.

The problem with learning about this aspect of database processing – dealing with system failure, guaranteeing consistency of transactions – is that it is difficult to present candidates with more than verbal descriptions of the problems. We cannot actually crash a running computer, or watch actions being taken at checkpoints. There is a market niche here for someone who will develop a good onscreen animation of a database system encountering all of these problems. Until we have one, candidates must study carefully the description of actions taken by the system, and come to the course discussion board with their questions where written explanations are not clear.

Question 3

Previous examinations have presented candidates with descriptions of a set of items which could be modelled with an entity-relationship diagram and asked them to draw one. This year's examination was somewhat different. Nonetheless, most candidates coped with the different format. The first part of this question required an illustration of the concept of cardinality, as the term is used to show a binary relationship between two entity types. Some candidates got the concept around backwards, forgetting that cardinality is 'look across'.

The second part of the question required a 'fan trap' to be illustrated and then resolved. Unfortunately, some candidates came adrift here. Some of them confused the 'fan trap' with the 'chasm trap', while others submitted solutions that made no sense. But the majority showed that they understood this idea.

The third part of the question simply required definitions for some basic ideas in relational database theory and was done well by almost everyone. Where some of the definitions were not answered well, it was clear that the candidate had not spent sufficient – if any – time going over the basic concepts. Answering that 'Query Optimization' means 'Optimizing Queries' or 'a determinant is an attribute that determines other attributes' will not attract any marks. Another weakness was the use of the word 'Key' instead of 'Attribute', as in 'a composite key is one which is made up of two or more different keys'.

The fourth part of the question dealt with distributed databases and replicated data. Most candidates saw that replicating (duplicating) the same data at different physical locations could speed up queries; a few missed out that this also makes our database more robust in the face of crashes or obstruction of the network.

The fifth and final part of the question raised the issue of databases that do not use the relational mode. Since the course does not involve use of a non-relational model database system, only a very general answer was expected here. Most candidates offered a 'document' database, where the structure is not known in advance, and received credit for doing so.

Question 4

Candidates who opted for this question, as almost all did, had to answer a series of questions using SQL. The only queries that gave a substantial number of candidates' difficulty were those involving negations, which required a

set operation approach – either directly through set operators or via a sub-query.

A few candidates evidently did not know the difference between **ORDER BY** and **GROUP BY**. Almost all candidates knew that an optimiser speeds up queries, rather fewer could present two plausible examples of how its purpose could be achieved.

Most candidates appeared to understand what a **VIEW** was, although some seemed to believe that it necessarily involved a new physical relation separate from the original relations, which it does not. Most saw that one of its advantages was in convenience for the user; fewer saw that it is also a security measure because it prevents unauthorized users from seeing confidential data while allowing them to view the data they need.

Question 5

Question 5 was a 'normalization' question. Candidates were required first to identify the functional dependencies in an unnormalized relation, then to find the insertion, deletion and modification anomalies in it, define Boyce-Codd Normal Form (BCNF) and then split the table up into that form.

This was generally done well, except that some redundant dependencies were depicted. As always, a few candidates did not understand what an 'anomaly' is in this context – it is not just a mistake, but a certain kind of mistake. If I mistakenly delete a tuple whose information I wanted to save, that's a mistake, but it is not an anomaly. If I delete a tuple because I wanted to delete its primary key, but by deleting the tuple I delete other information that I did not want to delete, that is an anomaly. The transformation of the single unnormalized table presented few problems to most candidates, which was gratifying. If there were problems here, they involved trying to use simple (one-attribute) Primary Keys where a compound Primary Key was required.

Conclusion

This year's examination results did not vary radically from previous years. The few outright failures seemed to be due to total lack of revision. Several candidates scored very well indeed.

It is hoped that the next cohort of students will not take the results this year – which show that, effectively, anyone who conscientiously revises will succeed in passing the course – as grounds for complacency, but will strive for even better results next year.