Please see separate pdf.
The paper was sat by 20 students who attained an overall average of 63%. Q1 was answered by 19 candidates gaining an average mark of 11.8 (out of 20). Q2 was answered by 16 (average 12.8), Q3 by 1 (average 18.0) and Q4 by 4 (average 13.3).

Only one candidate obtained more than 14 marks for Q1, so it was clearly difficult to work completely through the parts without making some error, despite the topic being popular.

Q2 was evidently easier for candidates: 5 students got first class marks (greater than or equal to 14). Possibly students were helped by having met adaptive quadrature in the lab.

Hardly surprisingly, only one candidate attempted Q3 (on cyclic reduction), but it was a good choice for this student.

Q4 was predominantly a coding problem and there was one very good and complete answer.
John Gurd/Graham Riley:

The paper was sat by 7 students who attained an overall average of 78%. Q1 is was answered by 5 candidates gaining an average mark of 16.6 (out of 20). Q2 was answered by 5 (average 14.6), Q3 by 1 (average 15.0) and Q4 by 3 (average 15.7).

Evidently candidates were well on top of the material. There was only one mark below 12 (out of 20) for the 14 questions that were answered; 5 answers got marks of 18 (90%) or above. The material is not particularly easy, so candidates must have prepared themselves well.

Mikel Lujan:

The exam contains two research papers not read by the students during the course. Each student has to select one of the papers and provide a critique of the paper following the answers provided.

Most of the students selected the research paper (Cruise) rather than the paper (DPJ).

In general the questions were answered quite well. A few answers concentrated too much on the precise and too little on the analysis. Most students followed the guidelines, established during the course unit, for the evaluation of research papers and thus the marks have been at the high end of the spectrum.

Most students did not go beyond the obvious limitations for the DeNovo paper.
Section A

The questions in Section A required essay-like answers involving a number of topics discussed during the lectures, and represented a balanced blend of book knowledge, application of technique with discussion and original thought questions.

A large number of students attempted to answer original thought questions by copying fractions of text from books, slides and articles, and therefore were not able to get full marks, since the combination of fractions of text taken from various sources would not answer the questions.

The students that showed deep understanding of the concepts involved in the questions and were able to describe and apply them in the context of the questions were better rewarded in terms of marks. However, those represented a small minority of the students. A significant number of students chose to rush into a quick and easy answer, without giving it any thought.

Section B

Question 3

a.
   i. Generally very well answered.
   ii. Some good answers and most attempts showed some reasonable idea of what was needed but a few seemed to misunderstand the OLAP operations.

B.
   i. Most answers made reasonable efforts showing understanding of the algorithm but only a few were excellent.
   ii. Generally very well answered.

C.
   i. Generally very well answered.
   ii. Some good, but some did not seem to understand the correlation aspect.

D. Some good answers but many didn’t really attempt to answer the question but just gave definitions.

E. Some very good answers, but a number didn’t really attempt to answer the question.

Question 4

a. Generally well answered.

B. Most answers made reasonable efforts showing some understanding of the algorithm but only a few were excellent (parts I and ii closely interlinked).

C. Mostly very good/excellent answers.

D. Some good answers but many didn’t really attempt to answer the question but just gave definitions.

E. Generally very well answered.
Students appear too often to have read the question and then have switched into ‘autopilot’ spouting generic truths which fail to show that they have understood the principles and only have loose fit to the answer sought. For example, when a question has been posed specifying the roles involved in a system, the students too often just make disconnected statements about ‘stakeholders’ or ‘owners’. They fail to show that they have grasped the first principle of IT Governance in the assignment of responsibilities for (say) data ownership, infrastructure management, and system development. Most of the answers of this type seem to hope that quantity will win over quality and perhaps the examiner might ignore the unfortunate fact that the question has not been answered.

This year, more than ever, suggests that a significant proportion of the students are struggling with the English language. For many papers it was extremely difficult to divine any understanding from the written answers. Perhaps the most alarming example which was a discussion on the use of leading and lagging metrics for forecasting. The student gives ‘weather forecasting’ as an example of the IT governance tool. Other substitutions would give a Perry and Croft script a run for its money.

Students are still memorising lists from the model answers from exams in previous years. Despite being told repeatedly that in the model answers, lists were there to show what to include, not as a rounded answer, students still left the regurgitated list as their entire solution. Some examples of the lists suffered from the understanding of English and were written down after presumably passing in and out of a mother tongue and not coming out well. I intend to remove any model questions from Moodle next year.
Q1.a

The main difficulty with this question was the interpretation of the term 'routes'. What is meant by that is the set of classical channels by which users and applications can interact with a DBMS server. The answer to the question is, therefore, that the routes are: JDBC-style drivers, user interfaces (e.g., web-based (or not) forms, and command-line interfaces) and embedded queries in general programming languages. Some interpreted 'routes' as denoting architectures (e.g., centralized, parallel, peer-to-peer) or languages (e.g., SQL, relational algebra, etc.).

Q1.b

No major problems here, though some forgot to emphasize that plan selection is based on cost estimates.

Q1.c.i

No major problems. Most students got it spot on. The commonest (but still rare) mistake was to forget the join conditions in the where clause.

Q1.c.ii

The great majority had no problems. Those that made mistakes often forgot that what is meant by a direct translation is one that generates a tree with a projection node for the SELECT clause, a selection node for the WHERE clause (if any) as a whole (i.e., not broken up into a cascading sequence of selections) and then a cascade of Cartesian products (and not joins!) with one leaf node for each relation mentioned in the FROM clause. The common mistake was to pre-optimize, or to replace products with joins, or both.

Q1.c.iii

This proved, somewhat surprisingly, a difficult question for many. The answer is not difficult to construct. One simply starts from the direct translation (see above), converts Cartesian products into joins by picking the join conditions from the selection node. Then takes the remaining predicates in the selection node and moves them down, in the direction of the leaves, as close as possible to the pertinent leaves. Finally, for each edge in the resulting graph we check which attributes from the child are needed downstream from it (all the way yo the root). If we ascertain that some attribute is not needed we split the edge by inserting a projection node that only keeps the needed attributes.

Q1.c.iv

The question proved quite hard for most students. Many students misread the question as being about distributed execution (i.e., about movements of data and queries) but the mention of "partitioned parallelism" should have been sufficient indication that the question was about parallel execution. Other common flaw was not to draw node boundaries and not to consider that when joins are parallelized, since the correctness of the operation depends on attribute values, one must be careful to use the attribute value as an inout to decision as to which copy of the join it is sent to. The simplest way to do this is to see the inputs of "each copy of the join" as hash-partitioned. So, above each child of the join there is a split that uses hash partitioning (thus, there are cross flows from one node to another!) and below each join there is a merge to receive the tuples sent by the splits.

Overall, Q1 turned out to be rather hard, with a mean of close to 60%. Parts c.iii and c.iv were worth quite a lot of marks and proved harder than expected.

Q2.a.i

This question proved hard. Although the lecture notes have an almost identical example,
lots of students failed to reason properly that we move the smallest, join, then move the result to the next smallest, then join again, and so on.

Q2.a.ii
This question was not so hard, but, still, many students failed to give the simple answer sought, again almost straight from the lecture notes.

Q2.a.iii
This question was better answered than Parts a.i and a.ii.

Q2.b, Q2.c, Q2.d and Q2.e
These were mostly recall and weren't badly answered with the exception of c. Many students related the intermediate key generated by the mapper function as the means to group by and that the reducer is inherently an aggregation function, but, still, too many students simply didn't remember this point, which was particularly commented on in lectures and is the subject of a worked out example in the lecture notes.

Q2.f
This (along with 1.c.iv) was the hardest question, but, once more, it's just an application of a techniques that is exemplified in the lecture notes. It is very surprising that so few students answered well.

Q2.g
This was meant to be and indeed was an easy question.

-----
One general comment (particularly with respect to Q2) is that the students seemed to have revised poorly, since so many mark were lost on questions of pure recall or of replicating an example in the lecture notes.

This 2012-2013 cohort had very high coursework marks and this may have instilled a sense of excessive confidence. Perhaps there was some complacency at play, e.g., in revision. Students were explicitly and emphatically warned not to take a high coursework mark to be taken as a predictor of a high exam mark. Indeed, over the years this course unit has been taught, one can see that the exam performance is lower than coursework performance. The difference between exam and coursework performance for this cohort is perhaps exacerbated: the coursework marks were higher than the trend, the exam lower than the trend (perhaps the exam was harder than the most recent ones, see comments below).

For future exams, we will take special care to state in which context the answer is to be given, since so many students lose marks for missing the cogent context for the question.

-----
Note
Part c.iv had a typo/mistake in that where the word 'parts' occurs the word 'catalogue' should have appeared instead. Although few answers seem to have been significantly made worse by the error (and some students used assumptions to generate a good answer), it was regrettable that the error slipped through the vetting/moderation process. The marking was adjusted in all cases to take into account the specific way in which the student interpreted the question in the presence of the mistake, i.e., all interpretations were accepted.
The average mark on the MCQ was 73%, and the maximum was 98%. On the whole this was very well done by most students.

In the short answer questions of the exam, the Naive Bayes question was the main problem. Many students forgot that the assumption is class conditional independence, not just independence. The rest of the exam turned out very well, with the mean as 71%. A few people mixed up the concept of filters versus wrappers, and a few also could not state the dynamics of an SVM clearly.

I will address these things next year by giving more practice on Bayes theorem questions during class time.
All students attempted questions in three sections. Their marks are roughly distributed in a normal distribution. The highest mark is 80% and the lowest one is merely 2% while most of marks are in the interval between 50% and 70%.

For short questions in Sect. A, nearly all students have a good understanding of basic concepts and book knowledge in general. As a result, a vast of majority received at least eight out of 10 marks from this section. A few failed to give correct answers to A.1 and A.2.

For essential questions in Sect. B, almost all students generally performed well in B.1 and B.2. However, a common problem appeared in B.1 where a few failed to describe how a given stress works via optimisation in MDS learning. For B.3, nearly all students chose a proper algorithm for such an application. However, a number of students did not describe their answers properly; they simply described a chosen in general without considering the context while the question is asking them how to apply an appropriate algorithm to a specific real-world problem. In addition, a number of students failed to address issues with critical analysis regarding this application.

For questions related to formal analysis in Sect. C, most of students made an attempt to C.1 instead of C.2. A number of students clearly show their understanding of the question in C.1(a), which asks students to make a derivation of the dual PCA algorithm presented in the PCA lecture with re-formulation by considering the data centralisation, and therefore gave a proper answer. However, others did not spot it and hence gave irrelevant details in their answers. Unfortunately, most students did not give correct answers to questions in C1.(b), and also none of those who chose the question in C.2 performed well.

In summary, the overall performance looks quite reasonable and accurately reflects what students actually achieved from this course unit. In particular, the distribution of examination marks is quite consistent with that of individual coursework assessment.
Q1. attempts: 67; mean: 62.5%; SD: 20.1%
Q2. attempts: 53; mean: 45.7%; SD: 20.3%
Q3. attempts: 62; mean: 54.1%; SD: 20.3%
Q4. attempts: 19; mean: 52.7%; SD: 32.1%
Question 1:
This question asks for how in SSL, client and server authenticate. Some students describe web authentication methods in general, rather than in SSL, and some only mention how server authenticates to client, but failed to mention how client authenticates to the server.

Question 2:
Students largely did well in this question. The most common mistakes are in (c) and (d).

Question 3:
This question appears to be too easy, as most students did really well.

Question 4:
This question appears to be the hardest one. The common mistakes are (1) public key certification was not mentioned, (2) using a public key to encrypt a large message (rather than using a symmetric key to encrypt the large message, but using public key to transport the symmetric key), and (3) in signature generation, the hash and sign paradigm was not used (signature should be signed on the hash value of the message, rather than on the message).

Some useful statistics:

In total, 66 students sit in the exam.

Question 1: taken by 58 students, and the average mark is 13.39/20 = 67%;
Question 2: taken by 63 students, and the average mark is 12.5/20 = 62.5%;
Question 3: taken by 42 students, and the average mark is 13.5/20 = 67.5%;
Question 4: taken by 35 students, and the average mark is 8.7/20 = 43.6%;

The overall average mark = 60%.
36 (out of 66) students get results >= 60%.
Please see the separate file containing feedback.
Question 1
The question examined the fundamentals of the course’s curriculum and most students managed to answer it without major problems.

Question 2
Some students struggled with the details of the W model, and its focus on component-based development with validation & verification. Another common oversight was that part c) has asked to discuss categories of component models, not only the models experienced in the lab exercises.

Question 3
Most problems with this question were associated with part b). The components of the modelled system should have been described in detail – to the level of service signatures. Many students failed to realise the interface of the system comprises system services, not the top-level connector itself (let alone the physical interface of the system). Another common source of errors was the specification of data channels within the system.

Question 4
Again, the practical part b) was more problematic than part a). Some students struggled to specify components of the environment control system to the required level of detail; surprisingly often the system architecture contradicted the description of the components or the control flow model.

Question 5
An insufficient level of details in the specification of the system’s components was the most common problem. Also, some students were confused about the meaning of required and provided services (e.g. modelling data inputs and data outputs instead). Often, the control flow specification was not accurate (e.g. omitting branching conditions).