
**Computing Science Examination
Autumn Semester 2016**

CSCU9YQ: NoSQL Databases

Saturday 10 December 2016

14:00 – 15:30 hours

Attempt **BOTH** questions.

The distribution of marks among the parts of each question is indicated.

IMPORTANT NOTE

Read the instructions on the front of each answer book carefully.

It is essential that you write your student number on the front of each answer book.

Also, when you have completed the examination, the number of answer books that you have used must be prominently written on the front of one book.

1.

- (a) NoSQL databases such as MongoDB are said to be schemaless, unlike relational databases. Explain what a database schema is in this context and then describe what it means for a database to be schemaless. Illustrate your answer with an example that applies to MongoDB. [6]
- (b) Replication and sharding are methods for distributing a database across a cluster.
- (i) Describe the difference between replication and sharding in terms of how the data is spread and what is achieved by doing so. [2]
- (ii) Describe the use of a shard key in MongoDB and explain two of the qualities you should ensure such a key possesses. [3]
- (iii) Describe the difference between a shard key based on a numeric field and one based on a string. How does each produce a key and what is the difference between them in terms of the relative location they choose for two entries whose values in the key field are close together. [4]
- (c) Describe some things you would need to consider when choosing suitable indexes for a MongoDB database. Why wouldn't you just index everything? [3]
- (d) Describe in principle how a Map-Reduce query works in MongoDB. Describe the job performed by the map, reduce and finalize functions and explain what qualities the reduce function must have, and why. [3]
- (e) A graph database looks superficially like a relational database in that it has relationships between entities. Describe the difference between these two types of database, commenting on the following:
- (i) The nature of the relationships between entities and how those relationships are represented in the database [2]
- (ii) The limitations on which entities may be connected in a given database. [2]

Continued/

2.

- (a) Imagine you need to design a data store for an online product review website where users will be encouraged to describe the products they are reviewing using a set of (descriptor: description) pairs. For example, Processor speed: Fast.
- (i) Explain why the use of a relational model would be unsuitable for this level of user flexibility. [2]
 - (ii) Using JSON notation, show how you would represent a review for a laptop made by Lenovo where the reviewer thought the battery life was too short, the screen quality very good and the list of things they liked about it were the design, the colour and the weight (express these as an array). [3]
 - (iii) Describe one disadvantage of the design decision to allow users to choose arbitrary field names like this. [2]
- (b) There are a number of different ways in which you might choose to organise the data using document embedding or references to other documents. Assuming you have a document for each user, discuss the relative merits of embedding or linking to things the user has reviewed in the past. What would be the best choice to make, and why? [4]
- (c) Write MongoDB commands to run the following queries on the database described above. Assume sensible collection and field names.
- (i) Add a user with user name bill1 and full name Bill Smith. [1]
 - (ii) Edit bill1's document to add the fact that his age is 25. [1]
 - (iii) Edit bill1's document to list two email addresses: bill1@bill.com and bill1@gmail.com. [2]
 - (iv) Add a third email address, bill1@work.com to bill1's account. [2]
 - (v) Find all the reviews for Apple products. [2]
 - (vi) Assume every item reviewed has a rating from 1 to 5 and a category that is either Electronics, Clothes or Cars. Write code that uses the `group()` method to calculate the average rating for each category. [6]

END OF EXAMINATION