

ITNPBD3 Assignment 2019

Design a Database for a Classified Adverts Website

Introduction

Imagine a website where people can sell things they no longer need. There are lots of examples you can look at: eBay, Gumtree, Craigslist etc.

Such sites have some easily identifiable entities about which data must be stored: Buyers, sellers, transactions (person A buys item B from person C, for example), items for sale, and messages (for example, asking the seller a question).

Classified adverts are organised into classifications, for example motor cars, rooms for rent, baby clothes etc. and each classification uses different features to describe items. Cars have make and model, clothes have sizes, etc. Note that each user can only have one account – there is no need for a buyer's account and a separate seller's account for the same person.

Your assignment

1. Imagine you choose a relational database to store the data from the classified adverts website.
 - a. Items for sale in a classified advert system have different features depending on their classification. Explain why this is a problem for a relational database and suggest two possible solutions to the problem. Compare the two solutions, giving advantages and disadvantages of each. Say which you would choose.
10 Marks
 - b. Design a set of normalised relations for storing data about people, objects for sale, transactions and messages. For each relation choose a primary key and explain the reasons behind the choice. Highlight the foreign keys in your design. Explain whether your design is in second or third normal form and say why you chose that level of normalisation. You needn't include every possible field (don't worry about product features, for example), just show some example fields for each table.
8 Marks
 - c. Draw an entity relation (ER) diagram for the normalised relations you have designed. There is no need to show non-key fields, but show primary and foreign keys and the cardinality of their relationship. Clearly state any assumptions you have made.
5 Marks
 - d. Give an example of the SQL code needed to select the names of the buyer and the seller of an example item if the item ID is known. The code should implement a join.
5 Marks
2. Now consider a NoSQL solution to the same problem.
 - a. Referring to your answer to part a. in question 1, explain how a document database such as MongoDB offers a solution to the classified advert problem.
5 Marks

- b. Describe two use cases for your database in terms of the task a user will want to perform, the data that will be accessed, and how often the task will be performed.

5 Marks

- c. People, objects and classifications are three potential aggregates that could be used for the database. Explain what this means and compare the advantages and disadvantages of using each of the suggested aggregates in a classified advert database. Refer to the use cases you gave in part b. above.

10 Marks

- d. MongoDB is a schemaless database. Describe what that means and then discuss what, if any, restrictions you might choose to impose on documents describing objects for sale. Say how these restrictions would be imposed and discuss the advantages and disadvantages of doing so.

5 Marks

- e. Thinking about the entries for objects for sale, discuss which fields you would index on. What would indexing gain for your website and why wouldn't you simply index every field?

5 Marks

- f. Imagine your database grows very large and is stored over a cluster. You will need to employ sharding to split the data up. Explain what this means and discuss which field(s) it would be sensible to use as a shard key. If you now realise that you need a field that you haven't mentioned yet, that is fine. Introduce it into the design now and explain why it is needed.

5 Marks

- g. Messages can be sent between two people (usually a buyer and a seller). Discuss different options for storing messages and linking each message to the people who sent and received it. Propose three different options and say which of the three you would choose, and why. In your discussion, refer to particular use cases such as listing all the messages a person has sent and consider the efficiencies of the different design choices you propose.

10 Marks

- 3. Now that you have two designs, one for SQL and one for NoSQL, discuss the relative merits of each and say which approach you would choose for the database behind a classified advert website.

7 Marks

- 4. Load the file provided, called AssgCars.json into a MongoDB collection and give the MongoDB code you would use to make the following queries. Also show the result of running each query. **4 Marks for each question**

- a. List the prices (and no other variables, including the ID) of all cars for sale where the make is Skoda and Model is Octavia
 - b. Calculate the average price of all Skoda cars for sale
 - c. Calculate the average price of each different model from the Skoda cars for sale
 - d. Find the model of Skoda car with the highest average price
- (part e. on next page)

- e. All cars have a field called Extras which holds an array of strings (e.g. {"Extras":["ABS","PAS","Aircon"]}). Write code to list all of the extras ever mentioned for each manufacturer in turn (i.e. the union of the Extras arrays grouped by manufacturer).

Submission

Please write your answers, save them to PDF and upload them on the course page on canvas. Include your student number on the front page of the report, but do not put your name anywhere on the report. The deadline for submission is Friday 22nd March at midnight. Late penalties of three marks per calendar day you are late will be applied, up to a maximum of seven calendar days, after which you will receive no grade for the assignment.

Plagiarism

Plagiarism is presenting somebody else's work as your own. Plagiarism is a form of academic misconduct and is taken very seriously by the University. Students found to have plagiarised work can have marks deducted and, in serious cases, even be expelled from the University. Do not submit any work that is not entirely your own.

The University's full guidance on academic misconduct can be found here:

<http://stir.ac.uk/1x0>