COMP207 Tutorial Exercises
Week 7 (16^{th}/18^{th} November)

The exercises below provide the opportunity to practice the concepts and methods discussed during previous week's videos/reading material. If you haven't done so, it is worthwhile to spend some time on making yourself familiar with these concepts and methods. Don't worry if you cannot solve all the exercises during the tutorial session, but try to tackle at least one or two of them. If at some point you do not know how to proceed, you could review the relevant material from the videos/reading material and return to the exercise later.

1 Distributed databases

The first question is not meant to be technical, but is more meant for understanding why we have these forms of partitioning.

Exercise 1 (Adapted from exercise 23.28 in [1])**.** Consider the following database for a chain of book stores in the USA: Books(Book_ no,Primary_ author,Topic,Title,Total_ stock,Price) Bookstore(Store_ no,City,State,Zip,Inventory_ value) Stock(Store_ no,Book_ no,Qty)

Total_ stock is the total number of books in stock and Inventory_ value is the total value of all books in that store.

(a) Give an example of two simple predicates that would be meaningful for the Bookstore relation for horizontal partitioning (consider perhaps warehouses that sends to a number of nearby stores)

(b) Give an example of a meaningful vertical partitioning for the Books relation (consider perhaps a sales register compared to headquarters compared to customers searching for books)

Exercise 2 (Exercise 20.4.1 in [2])**.** Suppose we want to take the natural join of $R(A,B)$ and $S(B,C)$, where R and S are at different sites, and the size of the data communicated is the dominant cost of the join. Suppose the sizes of R and S are s_R and s_S, respectively. Suppose that the size of $\pi_B(R)$ is fraction p_R of the size of R and $\pi_B(S)$ is fraction p_S of the size of S. Finally, suppose that fractions d_R and d_S of relations R and S, respectively, are dangling (i.e. $(1-d_R)|R| = |R \ltimes S|$ and $(1-d_S)|S| = |S \ltimes R|$). Write expressions, in terms of these six parameters, for the costs of the four strategies for evaluating $R \bowtie S$, and determine the conditions under which each is the best strategy. The four strategies are:

(a) Ship R to the site of S.

(b) Ship 5 to the site of R.

(c) Ship $\pi_B(S)$ to the site of R, and then $R \ltimes S$ to the site of S.

(d) Ship $\pi_B(R)$ to the site of S, and then $S \ltimes R$ to the site of R.

Exercise 3 (Exercise 20.5.2 in [2]). In this exercise, we need a notation for describing sequences of messages that can take place during a two-phase commit. Let (i, j, M) mean that site i sends the message M to site j, where the value of M and its meaning can be P (prepare), R (ready), D (don't commit), C (commit), or A (abort). We shall discuss a simple situation in which site 0 is the coordinator, but not otherwise part of the transaction, and sites 1 and 2 are the components. For instance, the following is one possible sequence of messages that could take place during a successful commit of the transaction:

$$(0,1,P),(0,2,P),(2,0,R),(1,0,R),(0,2,C),(0,1,C)$$

(a) Give an example of a sequence of messages that could occur if site 1 wants to commit and site 2 wants to abort.

(b) How many possible sequences of messages such as the above are there, if the transaction successfully commits?

(c) If site 1 wants to commit, but site 2 does not, how many sequences of messages are there, assuming no failures occur?

(d) If site 1 wants to commit, but site 2 is down and does not respond to messages, how many sequences are there? Some of the sequences are more reasonable than others in this scenario, try to distinguish which is which.

References

[1] Ramez Elmasri and Shamkant B. Navathe. *Fundamentals of Database Systems*. Pearson Education, 7th edition, 2016.

[2] Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems - The Complete Book*. Pearson Education, 2nd edition, 2009.