CS157B Midterm Spring 2019 Version 1

Name:
StudID:

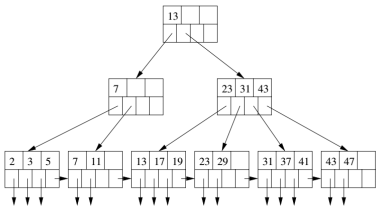
Instructions:

1. This midterm is due 11:59pm, March 12.
2. To complete the midterm, print it out, fill in your answers on the midterm, and scan it back into a file midterm.pdf where the total size is less than 10MB.
3. If you don't have a printer, copy and paste each problem into a word processor document. Then after each problem write your solution. Make a less than 10MB midterm.pdf file of the result and submit that.
4. Use the same submit mechanism as for the homeworks to submit your completed midterm.
5. Each problem on this midterm is worth the same amount (4pts).
6. If you have a question on the interpretation of a problem on the midterm, you can email me at chris@pollett.org.
7. Due to the coronavirus this is an open book, open internet midterm.
    a. What that means is that you can consult any static (on the order of static for weeks) source of information related to the midterm material.
    b. You cannot directly or indirectly ask another person how to do any problem off the midterm.
    c. To receive credit on problems that make use of your personal information, you need to have correctly filled in that personal information.
    d. When you submit your completed midterm, you are asserting all of the work in the midterm is your own.

1. Describe the ISAM file format as presented in class. (1pts). Give an example of performing a lookup according to this format where the number of cylinders and blocks of the file each depend on at least two digits from your Student ID (make obvious how they depend on your ID) (1pt). Explain the Oracle record parameters PCTFREE and PCTUSED (1pt). Give an example where they each could come in to play that depends on your Student ID (1pt).

2. Suppose the odds that a disk drive fails in a given year is 1/7. Suppose we are using Raid 4 with three drives. What is the mean time till data is lost if it takes 5 hours to replace a broken drive? (4pts)

3. Let *a* be the integer corresponding to the smallest two digits in your Student ID written left-to-right from smallest to largest digit. I.e., if your id was 314999768, then this would be 13. Let *b* be the integer corresponding to the next two smallest digits from your Student ID written left-to-right from smallest to largest digit. I.e., if your id was 314999768, then this would be 46. For the following B-tree, show:
    a. Step-by-step the blocks accessed in looking up the record which has key *a*. (2pts).
    b. Step-by-step inserting a new entry for a record with key *b*. (2pts).



4. View your Student ID as an integer *M*. Suppose we had a table PERSON with *M* rows. In these rows, the HAIR_COLOR attribute can take one of five values. The PERSON table has both an ID attribute and a SUPERVISOR_ID attribute. Solve the following problems (to receive credit please show all work):
    a. Find the records of people in PERSON that have your hair-color (use your specific hair color in answering this, for example, my hair color is chestnut brown). Write this as a relational algebra query (0.5pt), estimate the number of tuples returned (0.5pt).
    b. Estimate the number of blocks needed to be read to perform this query assuming 100 records fit in a block and the table is not ordered by hair-color (0.5pts). Suppose we have a clustered index on hair-color and 1000 index records fit in a block. How many blocks would be needed to be read in this situation? (0.5pts).
    c. Write the relational algebra need to return all PERSON rows whose supervisor shares their HAIR_COLOR (0.5pts). Assume we are using the clustered index above to perform this query. How many blocks would be read? (1pt).

5. Give the algorithms from class for each of the following (1pt each algorithm), then give an example stepping through the algorithm in a concrete situation dependent on your Student ID. (1pt each example):
    a. (Two Pass) Hash-based Join.
    b. (Two Pass) Sort based Duplicate Elimination.