CSc 460 — Database Design
Spring 2023 (McCann)

http://u.arizona.edu/~mccann/classes/460

Homework #1

(100 points)

Due Date: February 16th, 2023, at the beginning of class

Directions

1. **This is an INDIVIDUAL assignment; do your own work! Submitting answers created by other people is NOT doing your own work.**
2. Start early! Getting help is much easier n days before the due date/time than it will be n hours before.
3. Write complete answers to each of the following questions, in accordance with the given directions. Create your solutions as a PDF document such that each question is on a separate page; all parts of a multi–part question may be on the same page. Show your work, when appropriate, for possible partial credit.
4. Questions of the form "$\alpha.\beta$" are found in the Connolly/Begg text, available via D2L, as question β at the end of chapter α. For questions with sub–parts, note that a notation of "(w,z)" is asking you to complete parts w and z only, not parts x and y.
5. If you have questions about any aspect of this assignment, help is available from the class staff via piazza.com and our office hours.
6. When your answers are ready to be turned in, do so on gradescope.com. Be sure to assign pages to problems after you upload your PDF. Need help? Visit https://help.gradescope.com/ and search for "Submitting an Assignment."
7. Remember that you can use at most one late day on a homework assignment, because we will be distributing solutions after that time.

---

1. (5 points) 2.4 (just a paragraph or two)
2. (5 points) 3.2
3. (5 points) 3.18(a), but use only the Postgres DBMS (`www.postgresql.org`) as your subject
4. (5 points) Schemata.
   (a) What is the difference between a database file and that file's schema?
   (b) What is the schema of the hash bucket file of a dynamic hashing index?
5. (5 points) 4.8
6. (5 points) 12.8 (find examples not in the text and not used in class)
7. (10 points) 12.11(all parts). For the E–R diagram, use your choice of Crow's Feet or UML notations (but don't mix them — use one or the other!).
8. (15 points) Probabilities of disk failure. For all parts, assume an AFR probability of $p_f = 0.025$.
   (a) What is the probability of failure of a striped 5–disk system?
   (b) When your board of directors hears this, they are appalled. "That's too high! Build another system, just like this one, put it in our branch office in Frostbite Falls, and use it to mirror this system. That'll reduce the probability of the whole system failing!" Will it or won't it? Explain your answer, showing the math so that even the board members will understand.

(Continued . . .)

9. (10 points) We recently covered just four of the standard RAID levels in class. One of the levels we skipped is RAID 2. Using reliable sources, learn the details of RAID 2, and write, in your own words, a description of no more than three paragraphs that both describes the characteristics of RAID 2, and explains and why RAID 2 is rarely used.

10. (10 points) In Program #2, you dynamically–hashed using decimal digits. For this exercise, assume that our keys are in Base 3 instead of Base 10. Also assume that buckets are disk blocks that can hold at most three keys each. Build an **extendible** hashing (not dynamic hashing!) index structure using the keys listed below, reading the digits left to right, and draw the final structure. Be sure to include all global and local depths.

    1112, 2210, 0001, 0111, 1221, 0120, 1210, 2100, 1102, 0122, 0012, 2112

11. (15 points) Assuming a B^+–tree (note: plus!) of Order 2 (with Order defined as in our in-class (Comer's) B–Tree definition), insert the values K, B, G, D, S, T, and M, in the order presented. Show the tree after the completion of each insertion that causes the tree to grow by a level, and show the final tree.

12. (10 points) From your final tree from the last problem, delete the following keys, and show the final tree: M and G (in that order). You may show intermediate trees if you so desire.