Warm up
1. (5 points) Write your name and student id on your blue book.

Conceptual Data Model

2. (15 points) Consider people at USC who are specialized into students and faculty members. Multiple students have a faculty advisor. All individuals have a social security number (SS#), a name, and a start date at USC. Each student has a grade point average (gpa) and a unique student id (sid). Each faculty member has a salary and a faculty id (fid). Assume that the ISA relationship is non-overlapping, i.e., a USC person may NOT be both a student and a faculty member. Reduce this ER diagram to a collection of tables. Ensure to apply all possible optimizations that you are aware of.

SQL
Recall the employee relation that maintains the employees of a company, their salary, and the manager that an employee reports to:

   Emp (SS#, name, age, salary, dno, mgrSS#)

3) (10 points) Write the SQL command to delete (fire) all those employees whose salary is higher than their manager.

Normal Forms
4. (10 points) Both the 2nd and 3rd normal forms state that a non-key attribute must be a fact about the key, the whole key, and nothing but the key. They differ in how they are violated. Describe this difference and provide an example.

Relational Algebra
5. During lectures, we discussed the following simple program to give a 10% raise to those employees who work for the toy department using the tables Emp(SS#, name, age, salary, dno) and Dept(dno, dname, floor, mgrSS#):
   a. ToyEmp  Emp  name='toy' (Dept)
   b. Emp  Emp – ToyEmp
   c. ToyEmp  δ salary  1.1*salary (ToyEmp)
   d. Emp  Emp U ToyEmp
(15 points) If instruction b is swapped with instruction c then would the above program still give a 10% raise to the toy employees? Explain your answer.
Extendible hash index

6. (20 points) Consider a 4 bit machine from 1960s. (Instead of 32 bits, it has a 4 bit address space.) It implements extendible hashing where the size of a bucket is limited to 3 records. The index structure after the insertion of three keys (1110, RID0), (1010, RID1), and (1111, RID2) is shown to the right. In this picture, the depth of both the directory and the bucket is 0 because no overflow has taken place as yet. Show the structure of the directory once the following key (1000, RID3) is inserted. You may NOT assume the use of overflow chains. It is okay to show intermediate directory structures, however, make sure to identify the final directory structure and its buckets for full credit.

File System
7. (5 points) Provide the definition of a Record Identifier.

8. (5 points) State the difference between the seek time and rotational latency for a magnetic disk drive.

Simple Query Processing
9. Assume that 10 records of the Employee relation fit per disk page. Moreover, assume that this relation consists of 100,000 records.
9a) (5 points) Given an exact-match selection predicate that retrieves all employees whose salary equals $10,000 (Salary = 10,000), how many I/Os are required to process this operation with a heap file organization?

9b) (10 points) Assume a non-clustered hash index on the salary attribute of the Employee relation with a 50% loading factor and uniform distribution of index records across the bucket. How many I/Os are required to process the range selection predicate "Salary > 100,000"? Explain your answer.