Name: _____ **Student ID:** _____

MIDTERM EXAM, Fall 2022

CS 564
Department of Computer Science, University of Wisconsin, Madison

Exam Rules:

1) Close book and notes, cheat sheet allowed, 75 minutes

2) Please write down your name and student ID number NOW.

3) Please wait until being told to start reading and working on the exam.

4) If you think a problem is ambiguous, write down your assumptions, argue that they are reasonable, then work on the problem using those assumptions.

1. (25 points) Suppose you are asked to write a database application that captures the following information:

Students take courses. Each student has a student ID number and a name. Each course has a course ID number, name, and room number. Each student is either an undergraduate or graduate student. For each undergraduate student, we want to record also his/her year, GPA, and (possibly multiple) email addresses. For each graduate student, we want to record his/her advisor name, and also his/her multiple email addresses. In addition, each course is assigned to exactly one graduate student, who acts as the head TA of the course. For this head TA role, we want to record the start and end times (e.g., starts in 01/05 and ends in 12/05).

a) [10 points] Draw an ER diagram for this application. Identify and underline the attributes of the primary key of each entity set.

b) [10 points] Convert the above ER diagram into a relational schema. Use the OO approach to convert is-a hierarchies (that is, subclass entities). Specify the primary key of each relation in your schema.

c) [5 points] Why should we start modeling an application with ER diagram, why not starting the modeling process with relational schemas (that is, why not just directly create the tables)?

2. (15 points, 5 points each)
Consider a database schema with the following relations:

Person(ssn, name, phone, city)
Purchase(buyer-ssn, seller-ssn, store, pid)
Product(pid, name, price, category, cid)
Company(cid, name, stock-price, country)

where Purchase.pid and Product.pid refer to product id, and Product.cid
and Company.cid refer to company id.

a) Write a SQL query that finds the names and stock prices of companies that sold products of category "TV" to people living in Madison. (Returning duplicate names/stock prices is ok.)

b) Write a SQL query that lists the names of all companies that have sold at least two different products (that is, products with different pids).

c) Write a SQL query that lists the names of all people who (a) are living in Madison, and (b) buy at least five different products. For each person name, list also the total number of different products that that person has bought.

3. (10 points, 5 points each)

a) Is NULL value used to represent only missing values? If not, what else can it represent?

b) What would the SQL query

SELECT * FROM Sailors WHERE (age < 20) OR (age >= 20)

return from the table Sailors?

4. (15 points)

a) [10 points] Briefly discuss the suitability of heap file, hashed file, sorted file, and B+ tree file for equality and range searches.

b) [5 points] Can you describe a real-world situation where heap file organization would be the most appropriate?

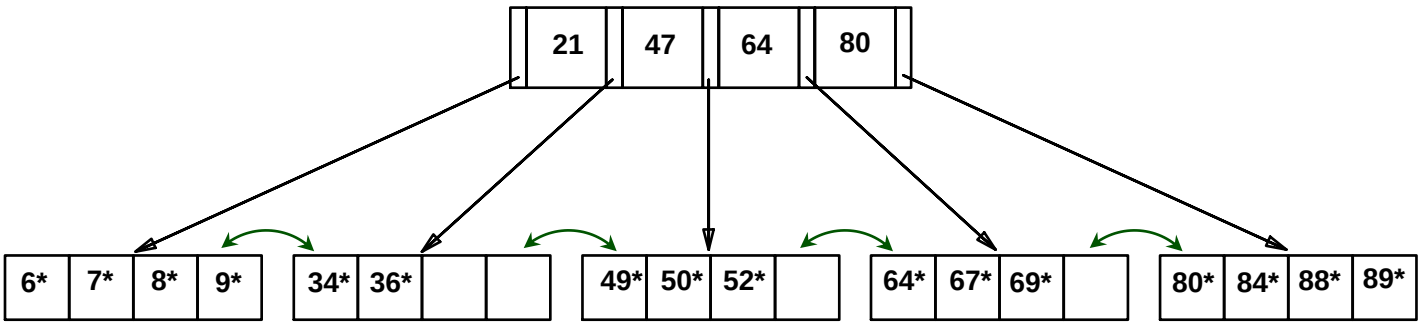
5. (15 points)

Assume a newly created slotted page P to store records of variable length. Consider the following sequence of actions:

Insert record A. Insert record B. Delete record A. Insert record C. Insert record D. Delete record B. Delete record D.

Draw the page AFTER the above sequence of action has been executed.

6. (20 points, 10 points each)
Consider the B+ tree below, where each node (except the root) must contain between 2 and 4 entries.



a) Draw the tree AFTER we have inserted 37, then inserted 95.

[See the next page for Part b of this question]

b) Continuing with the tree you have drawn in Part a, now we will delete 84, then delete 80. Draw the resulting tree. (Note that you must delete 84 then 80 from the tree you have drawn in Part a, NOT from the original tree.)