CSE 530 – Midterm Exam

Name:

(Print **_CLEARLY_**)

Question	Points Possible	Points Earned
1	25	
2	10	
3	20	
4	20	
5	15	
Total	90	

Question 1 – Heap Files

Suppose we want to create a heap file with a page size of 512 bytes. The tuples we wish to store in this heap page are 16 bytes long.

a.) (3 points) How many tuples will fit on a single heap page?

b.) (3 points) What will the size of the heap page header be, in bytes?

c.) (5 points) Explain why heap pages are a necessary part of the heap file structure – why do heap pages exist?

The heap file implementation we discussed in class and used for lab 1 is not the only way that heap files are implemented. One variation of a heap file does not have a header at all. Instead, it simply inserts each new tuple at the end of the heap file, on the very last page. To delete a record, the heap file is searched until the desired record is found and then removed.

d.) (7 points) While removing the header makes it seem as though it will lead to a more efficient use of space, that is not always true. Describe a situation when using a heap file without a header would lead to a lead to a less efficient use of space than a heap file with a header.

e.) (7 points)Which implementation of a heap file will be more efficient when searching for tuples (i.e. a relational select operation)? Explain how you know.

Question 2 – SQL Queries

You are given the following table definitions:

Airport		
Airport_ID	City	State

Flight		
Flight_ID	Departure_Airport_ID	Arrival_Airport_ID

Note that in this database, both the **Departure_Airport_ID** and **Arrival_Airport_ID** columns from the Flight table are foreign keys that are linked to the **Airport_ID** column of the Airport table. The underlined columns indicate the primary keys of each table. While no data is shown, you should assume that each table is filled with a large amount of flight data.

a.) (5 points) Write a query that will show the total number of flights that are departing from each city. Your results should include the name of each city in the database, as well as the number of flights departing from that city.

b.) (5 points) Write a query that lists the flight IDs of all of the flights that have arrived in the state of Missouri.

Question 3 – Query Trees and Optimization

(10 points) Construct a query tree for the following query. Your tree does not need to be optimized.

```
SELECT Title, Developer, Platform

FROM VideoGame JOIN Developer ON

      VideoGame.Developer_ID = Developer.DeveloperID

WHERE Year = 1989
```

(5 points) One of the steps of the optimization heuristic presented in class states that we should perform more restrictive select operations as early as possible. Explain what effect this rule has on the performance of joins, especially for queries that contain multiple joins.

(5 points) Describe one way that metadata can be used to optimize query performance.
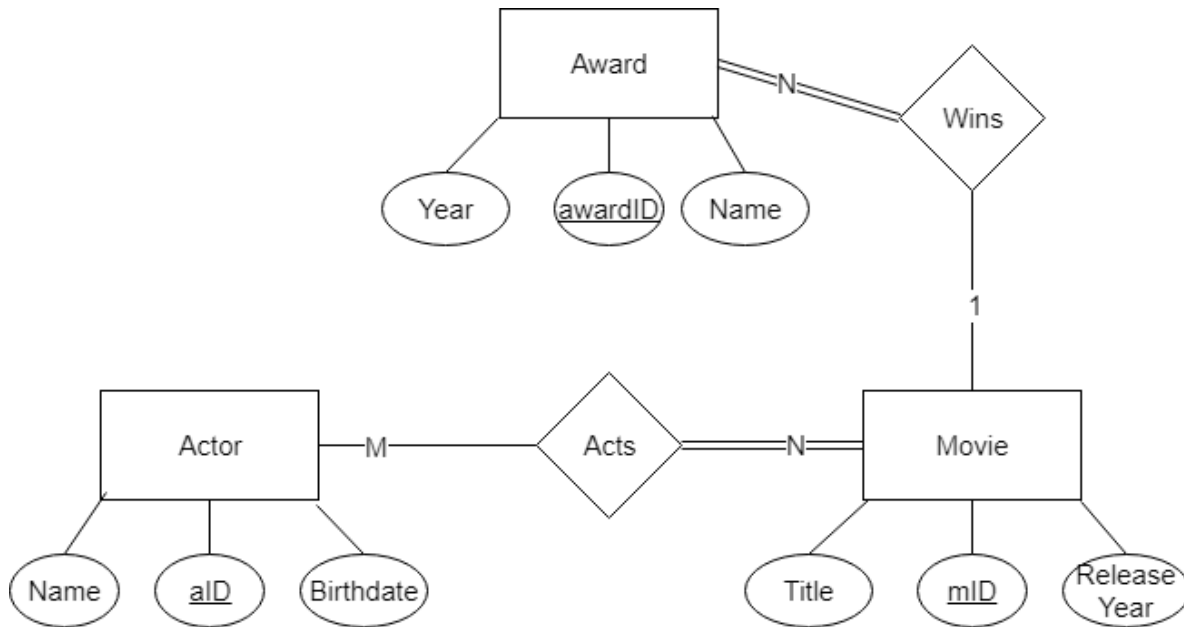
Question 4 – B+ Trees

Construct a B+ tree with $p_{inner} = 4$ and $p_{leaf} = 2$ that inserts the following values in order:

8, 6, 7, 5, 3, 0, 9, 10, 4, 2

For maximum partial credit, please show your work. (20 points)

Question 5 – ER Diagrams

You are given the following ER diagram:

a.) (4 points) Once this database is implemented, how many tables will it contain? Explain how you know.

b.) (6 points) Write a SQL statement that will create the Award table. This statement should create the table as well as create the relationship between the Award and Movie table (you can assume that the Movie table already exists).

c.) (5 points) Actors sometimes win awards, too, however this ER diagram is not currently tracking this information. Explain how we could change the design of this database to include awards that have been won by actors and actresses. You do not need to modify the above ER diagram to receive credit for this problem, but you may do so if it will help you answer this question.