Problem 1 (10)

State if the following statements are **TRUE** or **FALSE**. If a statement is false, please explain why. 2 points for **correct** answer, -1 point for **wrong** answer.

1. Any schedule produced by a lock scheduler using shared and exclusive locks is conflict serializable.
2. The LRU algorithm should not be used for the buffer management of B+-tree indexes.
3. A B+-tree is typically more efficient than extensible hashing for all types of queries.
4. Sequence pointers are used in leaf nodes of B+ trees to speed up range queries.
5. A hard disk with 5 double-sided platters, 500 tracks per surface, 1000 sectors per track, where each sector is 2048 bytes can store an 1 TBytes file.
Problem 2 (25)

Consider the following database that stores information about the current NBA season:

**Player** (playerID: integer, name: varchar(50), position: varchar(10), height: integer, weight: integer, teamName: varchar(30))

Each Player is assigned a unique playerID. The position of a player can either be Guard, Center or Forward. The height of a player is in inches while the weight is in pounds. Each player plays for only one team. The teamName field is a foreign key to Team.

**Team** (name: varchar(30), city: varchar(20))

Each Team has a unique name associated with it. There can be multiple teams from the same city.

**Game** (gameID: integer, homeTeam: varchar(30), awayTeam: varchar(30), homeScore: integer, awayScore: integer)

Each Game has a unique gameID. The fields homeTeam and awayTeam are foreign keys to Team. Two teams may play each other multiple times each season. There is an integrity check to ensure homeTeam and awayTeam are different.

**GameStats** (playerID: integer, gameID: integer, points: integer, assists: integer, rebounds: integer)

GameStats records the performance statistics of a player within a game. A player may not play in every game, in which case it will not have its statistics recorded for that game. gameID is a foreign key to Game. playerID is a foreign key to Player.

Write the SQL statements for the following queries:

1. List the name and the position of the players that play in a team in Los Angeles.  [5 points]

2. For each city that has at least one NBA team, list the city name and the height of the tallest player playing the "Center" position in that city.  [10 points]

3. For each win of the "Celtics", list the game ID and the name(s) of player(s) who scored the maximum number of points for the Celtics in that win, along with that maximum number of points.  [10 points]

You can use sub-queries and/or intermediate relations.
Problem 3 (20)

In the following schedules, Ri(A) stands for a Read(A) operation by transaction i; Wi(A) stands for a Write(A) operation by transaction i.

1. For each of the following schedules show if it is conflict-serializable and give a conflict-equivalent serial schedule. Show all the conflict operations.

   1. R1(A) W1(A) R2(A) R2(B) W3(B) W2(C) R4(A) R4(B) R4(C) R2(D) R3(E)
   2. R1(A) R4(A) W1(A) W3(B) R2(A) R2(B) W2(C) R4(B) R4(C) R2(D) R3(E)

2. Consider the following schedule example where an item is missing:

   W3(B) R2(A) W1(A) R3(?) R2(B) W1(B)

   and let:

   I. Producible using 2 Phase Locking
   II. Conflict Serializable

   Choose one answer for the following cases:

   a) If  ? = A, this schedule is which of the following:
      a. I & II b. I only c. II only d. neither I nor II

   b) If  ? = B, this schedule is which of the following:
      a. I & II b. I only c. II only d. neither I nor II

   c) If  ? = C, this schedule is which of the following:
      a. I & II b. I only c. II only d. neither I nor II

Problem 4 (20)

Consider the following schema:

Product (id(4), name(16), manufacturer(20), category(10), color(10), webpage(40))
Sales (pid(4), quantity(4), shippingaddress(20), date(12), shippingmethod(10))

Each attribute has a fixed length, with the size (in bytes) indicated by the number in parentheses. The number of tuples in each relation is: T(Product) = 1,000,000 and T(Sales)=2,000,000.

The size of one disk block is 1000 bytes, and there are 101 buffer blocks available in main memory.

1. Compute the number of blocks taken by each table B(Product) and B(Sales).
2. Consider the logical plan below:

\[ \Pi_{\text{name, date}} \]
\[ \sigma_{\text{category="Toy" AND quantity>100}} \]
\[ \text{id=pid} \]

\text{Product} \quad \text{Sales}

1. Assume the following execution plan: the join is implemented as a hash join, its result is pipelined into the selection operator, and, from there, pipelined into the projection operators. Compute the total cost of this physical plan.

2. Derive a new logical plan by pushing selections and projections down as far as possible (you have to draw a plan).

3. Consider a physical plan for your new logical plan in which all selections and projects are pipelined and the join is a block-nested loop join. Further assume that 1% of all Products are in category “Toy” and that 20% of all Sales have a quantity over 100. Compute the cost of your plan.

**Problem 5 (10 points)**

Consider the following log (next page) corresponding to a particular schedule at the time of a system crash for four transactions T1, T2, T3, and T4. Suppose that we use the immediate modification protocol with checkpointing. Describe the recovery process from the system crash. Show all the Undo and Redo operations and the values of the data items A, B, C, and D after the recovery.
Problem 6 (15 points)

Answer the following questions about Extendible Hash indexes. Assume that each bucket can store up to 4 entries (bucket capacity is 4.)

(1) Insert the following keys into an initially empty Extendible Hashing index:

3; 5; 7; 9; 10; 15; 25; 31; 44; 64

(2) Show the index after inserting a single key whose insertion causes a bucket split into the index that you created in (1).

(3) What is the maximum number of data entries that can be inserted into the index that you created in (1) before you have to split a bucket? Explain briefly.

(4) What is the minimum number of record insertions into the index that you created in (1) that will cause a split of all buckets? Explain briefly.