Homework 3

Spring 2020: Introduction to Database

Due Friday **May 1, 11:59 PM**

Overview

This assignment requires written answers to questions. You may have to copy and paste some SQL statements or scripts into the answer document you submit. You may also have to insert diagrams. We will only accept **typed** answers.

There are 12 questions worth 5 points each. The topics covered in this homework assignment touch on Chapters 11 to 16 in the textbook and the last few weeks worth of lectures.

Your submission should be a single PDF named [uni]_hw3.pdf.

Questions

1

1.1 An OS attempts to make sure that consecutive disk blocks store consecutive blocks of a particular file. Why is this practice especially important for magnetic disks?

1.2 A power failure that occurs while a disk block is being written could result in the block being only partially written. Assume that partially written blocks can be detected. An atomic block write is one where either the disk block is fully written or nothing is written (i.e., there are no partial writes). Suggest schemes for getting the effect of atomic block writes with the following RAID schemes.

Your schemes should involve work on recovery from failure.
 a. RAID level 1 (mirroring)
 b. RAID level 5 (block interleaved, distributed parity)

1.3 Suppose you have data that should not be lost on disk failure, and the application is write-intensive. How would you store the data? (Note: Your answer should consider

differences between primarily sequential writes and primarily random writes with regard to RAID).

2

2.1 Give benefits and disadvantages of variable length record management versus fixed length record management.

2.2 In the sequential file organization, why is an overflow block used even if there is, at the moment, only one overflow record?

2.3 Build and draw a B+ tree after inserting the following values. Assume the maximum degree of the B+ tree is 3. **This cannot be hand-drawn**

      Values: 3, 11, 12, 9, 4, 6, 21, 9, 15, 2

2.4 What would the occupancy of each leaf node of a B+-tree be if index entries were inserted in sorted order? Explain why.

3

3.1 Briefly explain the following normalization concepts:
    a. Functional dependency
    b. Data redundancy and update anomaly
    c. Decomposition
    d. BCNF
    e. First Normal Form and atomic domain

3.2 Assume (for simplicity in this exercise) that only one tuple fits in a block and memory holds at most three blocks. Show the runs created on each pass of the sort-merge algorithm when applied to sort the following tuples on the first attribute: (kangaroo, 17), (wallaby, 21), (emu, 1), (wombat, 13), (platypus, 3), (lion, 8), (warthog, 4), (zebra, 11), (meerkat, 6), (hyena, 9), (hornbill, 2), (baboon, 12).

4

4.1 Rewrite/transform the following query into an equivalent query that would be significantly more efficient.

```
SELECT
        people.playerid, people.nameLast, people.throws,
        batting.teamid, batting.yearid, ab, h, rbi
FROM
        (people JOIN batting USING(playerid))
WHERE teamid='BOS' AND yearID='1960';
```

4.2 Provide a two or three sentence definition of each of the following concepts, and provide an example:

   a. Structured data
   b. Semi-structured data
   c. Unstructured data
   d. Metadata
   e. BLOB