CS4432: Database Systems II
Spring D-Term

Release Date: March 21, 2014

Due Date: March 27, 2014 (11:59PM)

Total Points: 65

Problem 1 (Disks) [35 Points (5 each)]

Given a disk with the following configurations:

- There are 5 (double-sided) platters, each surface has 8,000 tracks, 200 sectors per track
- A sector holds 512 Bytes
- Disk rotates at 5400 rpm
- Seek time: A warm-up time is 1ms, and then the arm movement covers 500 tracks per ms
- Reading one sector takes 0.05 ms
- A disk block is of size 8K Bytes

Q1: Compute the disk capacity in GBs.

Q2: Compute the disk capacity in terms of the number of blocks it can hold.

Q3: What is the minimum, maximum, and average time needed to read one block from the disk? You need to divide the time into its three components (seek, rotation latency, transfer).

Q4: Given a file that consists of 100,000 records, and each record is 128 bytes. No record is allowed to span multiple blocks.

- How many records fit in one block?
- How many blocks are needed to store the file?
- How many sectors are needed to store the file?

Q5: Assume Blocks B1, B2, B3, …., B10, are stored on the same track (track number 100) and on adjacent sectors. If the disk head is positioned at track 0, what is the disk delay (time) needed to get read these 10 blocks? (Assume a half rotation is needed to reach the first block of these 10 blocks)

Q6: To optimize the reading of the file, we may store adjacent blocks, e.g., B1, B2, B3, …, on the same cylinder. Given the disk configuration above, how many blocks can be aligned on the cylinder position to be read by the disk arms at the same time?

Q7: Assume the file will be read sequentially, i.e., B1, B2, B3, ….until the last block B_{last}. Describe the best way to store these blocks on disk to speedup the sequential read. Then measure out the average time needed to read the file given this organization.
Hint: Make use of cylinders. Note that if Blocks B1 and B2 are aligned under each other in the same cylinder, then the transfer time to read B1, B2, or both is the same.

Problem 2 (Record Organization) [15 Points (5 each)]

Assume a database tables with the following fields:

ID (4 bytes), Name (25 bytes), age (4 bytes), DoB (10 bytes), gender (1 Byte), Address (60 bytes), state (2 bytes).

Each record on disk has a header part of size 8 bytes.

Q1: What will be the record size if each field has to start at 4-byte boundaries?

Q2: What will be the record size if each field has to start at 8-byte boundaries?

Q3: Assuming a disk block of size 4K bytes, the block uses 64 bytes of its own header (block header). How many records can fit in one block? Report the number under the 4- and 8-bytes boundaries.

Problem 3 (External Sorting) [15 Points (5 each)]

Given a relation R consisting of 10,000,000 tuples with each tuple 100 bytes long. Assume the disk is much larger than R, and each block is 4096 bytes. R is laid out on disk such that no record (tuple) spans more than one page. Assume that the main memory capacity is 256 MB, of which 60 MB is available for database system for sorting purposes. Answer the following questions.

Q1: How many disk I/Os are needed for sorting the above relation?

Q2: Now assume that the data is randomly laid out on disk, and the average seek time is 10 ms, the average rotational latency is 5 ms, and the average transfer rate is 0.1 ms per block of data. How much time will it take to sort the relation?

Q3: What is the size of the largest relation that can be sorted in 2 passes (pass 0 andpass 1) with the above buffer capacity (60 MB) in terms of number of tuples in the relation (tuple size is 100 bytes)?

What to Submit

- Include your answers in one file (.doc, .docx, or .pdf). This is the only file to submit.
- Include your name inside the file.

Where to Submit

- In WPI blackboard system

Late Submission Policy

- Follows the policy posted on the course website.