Database Systems: 1st Mid

Date: 8th February 2019 **Duration: 1.5 hrs**

1. **No clarifications during the exam.**
2. **Make *reasonable assumptions* and *clearly state* them to answer *ambiguous* questions.**
3. **Show your steps. Be concise and organized.**
4. **Calculators allowed. Sharing of calculators *not* allowed.**

1) The Megatron 777 disk has the following characteristics:

1. There are 15 surfaces, with 10,000 tracks each.
2. Tracks hold an average of 2000 sectors of 256 bytes each.
3. 20% of each track is used for gaps.
4. The disk rotates at 20,000 rpm.
5. The time it takes the head to move n tracks is $1+0.002n$ milliseconds.

(a) What is the capacity of the disk?
(b) If all tracks hold the same number of sectors, what is the density of bits in the sectors of a track?
(c) What is the maximum seek time?
(d) What is the maximum rotational latency?
(e) If a block is 16,384 bytes (i.e. 32 sectors), what is the transfer time of a block?

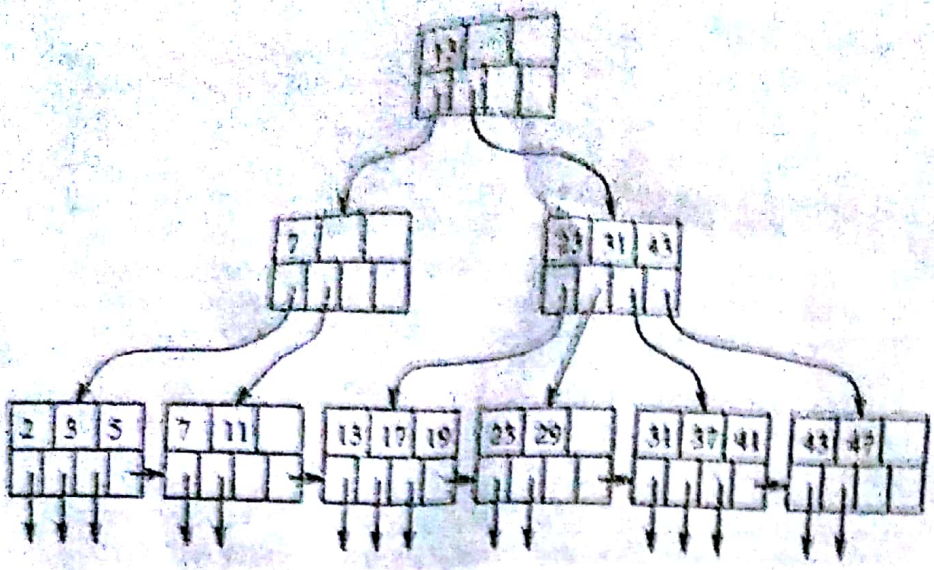
[10]

2) We have a 1 GB sized relation R of 10,000,000 tuples. Each tuple of 100 bytes has several fields, one of which is the *sort key* field, which may not be a primary key. The machine on which sorting occurs has one Megatron 777 disk (described above in Q1) and 50 MB of main memory available. Disk blocks are 4096 bytes. How long would it take to sort R using 2-phase, multiway merge sort. [10]

3) Suppose blocks hold either 3 records, or 10 key-pointer pairs. As a function of n, the number of records, how many blocks do we need to hold a data file and:
(a) A dense index? (b) A sparse index? [10]

4) Describe steps, using diagrams if necessary, to execute the following operations on the shown B+ tree:

a. Lookup all records in range 19 to 30
b. Lookup all records less than 20
c. Insert a record with key 6
d. Insert records with keys 14 through 16
e. Delete record with key 29



[10]

5) Show the state of a linear hashing file after each insert of keys in the sequence: 1, 2, 3, 6, 8, 13, 21, 34. Assume bucket capacity = 1 and family of hash functions = mod 2, mod 4, mod 8, mod 16, ... The file has no directory but may have an overflow area. For e.g. the following diagram shows the state after inserting keys 1, 2, 3. [10]

Next bucket to split	0	1					
H(i, k)	mod(k,2)	mod(k,2)					
H(i+1, k)		mod(k,4)					

0	2
1	1

0			
1	1	3	
2	2		

6) Suppose we store a relation $R(x,y)$ in a grid file. Both attributes have a range of values from 0 to 1000. The partitions of this grid file happen to be uniformly spaced; for x there are partitions every 20 units, at 20, 40, 60, and so on, while for y the partitions are every 50 units.

(a) How many buckets do we have to examine to answer the range query:
SELECT *
FROM R
WHERE $310 < x$ AND $x < 400$ AND $520 < y$ AND $y < 730$;

(b) How many disk accesses are needed to answer the above query using the grid file?

[10]