Assignment 4: Indexing and Query Processing (Spring 2019)

Instructor: Zhaonian Zou (znzou@hit.edu.cn)

Name: _____ Student ID: _____ Grade: _____

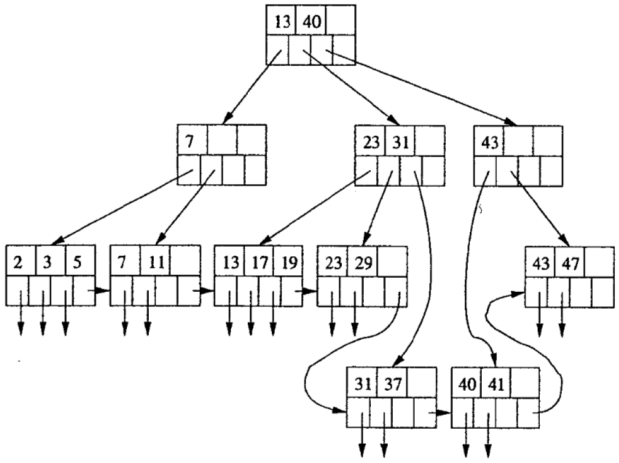
Question	1	2	3	4	5	6	7	Total
Score								

Notes

- **Print the assignment on A4 paper and answer the questions.**
- **Assignment due date: May 18/19, 2019 (the 2nd lab).**

Questions

1. (20 Points) We have a B+-tree depicted as follows.



   Complete the following operations on the tree.

   (a) (10 Points) Insert a tuple with search key 12. Describe the insertion process and draw the B+-tree obtained after the insertion.

   (b) (10 Points) Delete the tuple with search key 40. Describe the deletion process and draw the B+-tree obtained after the deletion.

2. (10 Points) Design a one-pass algorithm to implement the group-by operation $\gamma_{A;sum(B)}(R)$ and analyze the I/O cost and memory requirement of the algorithm.

3. (10 Points) Design a hash-based algorithm to implement the group-by operation $\gamma_{A;sum(B)}(R)$ and analyze the I/O cost and memory requirement of the algorithm.

4. (10 Points) Design a sort-based algorithm to implement the group-by operation $\gamma_{A;sum(B)}(R)$ and analyze the I/O cost and memory requirement of the algorithm.

5. (10 Points) Suppose there is a covering index on attributes A and B for relation R. Design an algorithm to implement the group-by operation $\gamma_{A;sum(B)}(R)$, which utilizes the covering index. Analyze the I/O cost and memory requirement of the algorithm.

6. (20 Points) We are given the statistics of 4 relations $W(a,b)$, $X(b,c)$, $Y(c,d)$, and $Z(d,e)$.

$W(a,b)$	$X(b,c)$	$Y(c,d)$	$Z(d,e)$
$T(W) = 100$	$T(X) = 200$	$T(Y) = 300$	$T(Z) = 400$
$V(W,a) = 20$	$V(X,b) = 50$	$V(Y,c) = 50$	$V(Z,d) = 40$
$V(W,b) = 60$	$V(X,c) = 100$	$V(Y,d) = 50$	$V(Z,e) = 100$

   (a) (5 Points) Estimate the cost of the following relational-algebra expression

$$\Pi_{b,c,d,e}(\sigma_{a=10 \wedge e>0}(W \bowtie X \bowtie Y \bowtie Z)).$$

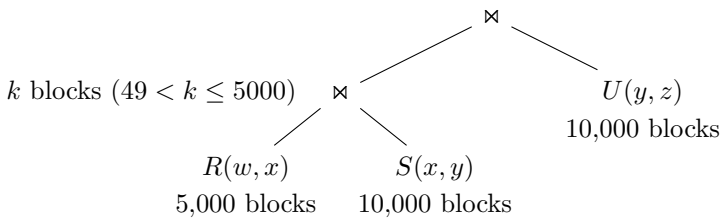
   (b) (5 Points) Transform the expression to an equivalent one that has lower estimated cost and give the estimated cost.

   (c) (10 Points) Determine the best order for evaluating $W \bowtie X \bowtie Y$, using left-deep join trees only.

7. (20 Points) Consider the following relational-algebra experession. The input relations $R(w,x)$, $S(x,y)$, and $U(y,z)$ are stored on disk in 5,000, 10,000, and 10,000 blocks, respectively. We are going to execute this expression using the following execution plan:

   - The operation $R \bowtie S$ is executed using the hash-join algorithm.
   - The join operation on $(R \bowtie S)$ and U is executed using the nested-loop join algorithm.
   - The tuples in $R \bowtie S$ are piplined to the join operation on $(R \bowtie S)$ and U.

   Suppose there are $M = 101$ blocks in the buffer pool available for executing the experession, and the tuples in $R \bowtie S$ occupy k blocks, where $49 < k \leq 5000$.



   Answer the following questions.

   (a) (10 Points) How many times has relation U been scanned during the execution of $(R \bowtie S) \bowtie U$?

   (b) (10 Points) Analyze the I/O cost for executing the expression according to the given plan.

Answers