CSCE-608 Database Systems

Spring 2024

Instructor: Dr. Jianer Chen
Office: PETR 428
Phone: 845-4259
Email: chen@cse.tamu.edu
Office Hours: MW 11:30 am–1:00 pm

Teaching Assistant: Ya-Ru Yang
Office: PETR 445
Phone: (979) 969-0095
Email: yaruyang@tamu.edu
Office Hours: TR 10:00 am–11:30 am

Assignment # 3

(Due April 24)

1. Below are the statistics for four relations:

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

Give the dynamic programming table entries that evaluates all join orders allowing: a) All trees, and b) left-deep trees only. What is the best choice in each case?

2. Suppose that we have the following key values:

$$29, 5, 7, 17, 19, 31, 2, 23, 11, 3.$$

Construct B^+-trees for the keys where the order of the B^+-tree is: (a) 3; (b) 7.

3. Discuss how to execute a deletion operation in an extensible hash table. What are the advantages and disadvantages of restructing the table if its smaller size after deletion allows for compression of certain blocks?

4. Discuss the necessary changes to insertion, deletion, and lookup algorithms for linear hash tables if the search keys are not unique

5. Suppose that we want to compute the set union of two relations S and R, where S and R are both sets and each takes $1,000$ blocks. Suppose that the main memory has $M = 200$ blocks.

(1) How many disk I/O's do you need if you use the two-pass sort-based algorithm discussed in class (and in the textbook)?

(2) Can you modify the two-pass sort-based algorithm to save some disk I/O's? (hint: use the idea of the hybrid algorithm as discussed in the hash-based algorithm).