DATABASE SYSTEMS (CSE 441)

Assignment 5

Deadline: Monday, 14th March 5 P.M.

Instruction:

1. You are not allowed to use any external library/jar files in this assignment.

2. Plagiarism will not be tolerated. **Copy from any source and in any form (friends/seniors/internet) will fetch you straight 0 marks in all the assignments and quiz.**

3. Be careful about your submissions. You must strictly follow the upload format. Failure to do so will fetch you a zero.

4. Languages allowed to code are C/C++/Java.

Given M memory blocks and two large relations R(X,Y) and S(Y,Z). Develop iterator for the following operations.

- **Sort-Merge join**
  - open() - Create sorted sublists for R and S, each of size M buffers.
  - getnext() - Use 1 block for each sublist and get min. of R & S. Join this minimum Y value with other table and return. Check for $B(R)+B(S)<M^2$
  - close() - close all files
- **Hash-Join**
  - open() - Create M-1 hashed sublists for R and S
  - getnext() - For each Ri and Si thus created, load the smaller of the two in the main memory and create a search structure over it. You can use M-1 buffers to achieve this. Then recursively load the other file in the remaining buffer and for each record of this file, search corresponding records (with same join attribute value) from the other file.
  - close() - close all files

Join condition (R.Y==S.Y).

Use 1 buffer for output which is filled by row returned by getnext() and when it gets full, append it to output file and continue.

You will be given as an input the files containing relations R and S and the value of M blocks.

Note that all attributes, X, Y and Z are strings and Y may be a non-key attribute.

Assume that each block can store 100 tuples for both relations, R and S.

Output: Vary M from 50 to 100 blocks in steps of 10 blocks and calculate the execution time. Plot the graph of M versus execution time separately for sort-merge join and hash-join.

Input Format: ./a.out <R file> <S file> <sort/hash> <M>
Output File: <R file>_<S file>_join
Languages allowed: C, C++ or Java.

Deadline: 14th March 5 PM

Submission Format:
Create a folder named your roll number and put all your files, README describing how to run your code and graph image in it. Compress it and upload **Assignment5.tar.gz**.