CSE441 DATABASE SYSTEMS
ASSIGNMENT-4

Question – To perform Union to two tables.

Function Prototype : union(Table1,Table2,n,M,type_of_index)

Table1 – Name of file containing the relation 1
Table2 – Name of file containing the relation 2
n – Number of attributes
M – Number of blocks.Note that B(table1) + B(table2) >M and M>2.
type_of_index – hash or Btree

Write a program to perform Union of two given relation.The union of the two relations should not contain any duplicates. To search whether a record is duplicate or not, use **B+Tree or Hashing** main memory structures for inserting and checking. The space required for B+Tree and Hash is not part of M.

Instructions :

1. Out of the M buffers, M-1 Buffers will be used as input buffers (which will hold the records from the input file). 1 buffer will be used as output buffer (holds the distinct records). If the output buffer gets filled, it should be flushed to the output file. If the input buffers get empty, next chunk of records should be read from the input file.
2. You need to construct a graph (time required for union vs size of relation2) where relation1 is of size 5 MB and the size of relation 2 changes as follows : 5 MB, 10 MB, 20 MB, 50 MB, 100 MB, 250 MB, 500 MB, 1 GB, 2 GB
3. Write a report containing the graph and your inferences drawn from the graph.
4. You have to write code for generation of the two tables. The generation can be done in following way : Generate R of 1 GB size .Generate R using random number function with r % duplication (r is an integer). After generating every 100 tuples, copy of any r tuples generated so far by selecting the same in the random manner.
5. Also, prepare a Readme file containing the details of how to execute your code and a file Rollno.sh to run your code.
6. **The languages allowed are C,C++,Java.**
7. **Any sought of copying from internet or from friends will lead to straight zero in all the assignments.**

Upload Format :

1. Create a folder with your rollnumber.
2. Put all the code files,ReadMe.txt,RollNo.sh and Report in to the folder created in 1.
3. Tar.gz the folder and name the archive as Assignment3.tar.gz

Deadline : 5th March 2016 5:00 pm