Concordia University
Dept. of Computer Science and Software Engineering
COMP 6521: Advanced Database Technology and Applications
Winter 2018

Mini-Project 2
==========
>> Total Points: 8
>> Project Reports due: **Tuesday April 10th at 17:00**
>> Project Demos due on: **Wednesday April 11th**

Project Description: Given two relations T1 and T2 defined below, we want to join them on the common attribute "Student ID." For this, we should compare the performance of (1) nested-loop join technique and (2) a sort based sort-based method. For the latter, adapt and use your implementation of the TPMMS method in MP1 to sort T1 and T2 first and then produce the join result in the last pass of phase 2. In addition to the number of tuples in the output and the output size (in MB), your report should also include total number of disk I/O's performed and the total processing time including the time to write the output on the disk. Also compute and store the GPA of every student in relation GPA(Student ID, ST-GPA).

Relation T1 has the same schema and parameters defined in MP1:

Student ID:	int (08)
First Name:	char(10)
Last Name:	char(10)
Department:	int (03)
Program:	int (03)
SIN Number:	int (09)
Address:	char(57)

Assume that every 40 tuples in T1 are stored in one block of 4K bytes. Suppose these relations are stored on the disk, and that each relation can have duplicated tuples. You may consider that each tuple appears on a separate line of the input file, stored on consecutive disk blocks. Note that there will be no separator (character) between the attribute values in a tuple and hence the length of the attributes is used to identify and extract the values. Examples of possible two records (lines) are as follows:

12345678John Smiths 111222999999999 1455 Maisonneuve West, Montreal, QC, H3G 1M8
88888888Roselyn Shirin 222444888888888 1515 Saint-Catherine West, Montreal, QC, H3G 1M7

Relation T2 is defined as follows with record size 27 and every 130 records stored in each block, leaving the rest of the block empty.

Enrollment(Student ID, Course ID, Semester, Year, Credits, Grade), where attribute types are int(8), char(8), int(1), int(4), int(2), and char(4), resp.

Evaluate the performance of your implementations on large instances of relations T1 and T2. To be more precise, in order to evaluate your work, create and use instances of T1 with about 50,000 tuples for relation T1 and 5,000,000 tuples for T2. Limit the size of the available main memory to 5MB. Report the results produced and the performance experienced using these instances and the 5MB main memory.

What tools you should use?

Use VM argument Xmx5m in Eclipse to restrict the main memory usage of Java Virtual Machine to 5MB.

What to submit on the due date (April 10th):

Submit your project report and the artifacts through MOODLE. Also include instruction to compile and run your code. Make sure your program compiles and runs on the lab computers.

Book a time slot with the lab assistants for the demo of your project on April 11th. Both lab instructors will be present for evaluating your implementations. Every member of your team MUST be present during your project demo.