Concordia University
Dept. of Computer Science and Software Engineering
COMP 6521: Advanced Database Technology and Applications
Winter 2019

Lab Assignment 1
==========
>> Points: 7 with at most 2 additional points for best work
>> Project Report: Wednesday February 20, 2019
>> Project Demo: Wednesday March 13, 2019

Description: An insurance company wants to identify for each client, the total amount paid to each client to compensate the damage to for each insured item. The claim records are recorded and maintained in the relation/table CLAIM since the company started 10 years ago. The company wants to find out the total amount paid to each client in the last 10 years. A client may insure multiple items, such as cars, homes, businesses, boats, etc, however the contract is different for each item insured. Each contract has a Contract ID which is unique. Table CLAIMS has a record for each item insured. The record layout is as follows:

The Schema of CLAIM table/relation:
==========
CNumber int(8) Primary Key
CDate Date,
CID int(9)
CName char(25)
CAddress char(150)
CEmail Address char(28)
Insured-Item Integer (2)
Amount-Damage Dec(9,2)
Amount-Paid Dec(9,2)

The claim number CNumber is the key attribute of the CLAIM table. Claim date Cdate is of type date"YYYY-MM-DD". The client denoted by CID is a 9 digit number (something like the social insurance number). CName is the client name, CAddress is his/her address, and CEmail is his/her email. Attribute Insured-Item is a 2 digit number which indicates the item insured. For example, 01 is for car, 02 for home, 03 for boat, 04 for motorcycle, 05 for jewelry, etc. Amount-Damage shows the assessment of the company of the damage to the item insured, and the Amount-Paid shows what the client is being paid to compensate the damage. This amount is to the focus of this project to process the SQL query given below. the focus of this project. The data type for the amount is 9 digits with two precisions, like $12.33. The length of each record/tuple is 250 bytes. Assume that each data block holds 15 tuples, and the data blocks are originally stored in consecutive disk cylinders.

The query to process is as follows:

```
SELECT CID, SUM(Amount-Paid)
FROM CLAIM
GROUP BY CID;
```

Your team is asked to develop a sort-based method to process this query and provide the desired output. Create and use different instances of the CLAIM table and use it to evaluate the performance of your implementation. The lab instructors will also create an instance of the table for you to use in your performance tests. They may create and use other instances of the CLAIM table for your demo

To evaluate your work, consider the following two cases:
(a) The number of CLAIM tuples is 100,000.
(b) The number of CLAIM tuples is 1000,000.

Also consider the following two cases for the amount of main memory available at run-time. (c) 5 MB or (d) 10 MB. Run your program in these two cases and report the following.

(1) Report the number of DISK I/Os for the 4 combination of CLAIM and Main Memory pairs: (a,c), (a,d), (b,c), and (b,d).
(2) Report the execution time (in minutes) in these 4 combinations.
(3) Include the top 10 costly clients and amount paid to them. The most costly client is the one(s) who received the highest total compensation.

Note: You should use VM argument Xmx5m in Eclipse to restrict the main memory usage of Java Virtual Machine. Ask PODs for help with Xmx5m.

What to submit on due date?

Each group should submit, through Moodle, a SINGLE zip file that contains: (1) your project report in PDF and (2) the source codes.

Please make sure to include the instructions to compile and run your code. It is important to ensure that your code compiles and runs on the computers in the labs. It is not acceptable to have project demos on your laptops.

Your project report and demo will be evaluated by at least two PODs. Every member of your team should be present for your project demo.

Bonus: The PODs may recommend your project for additional 1 or 2 points is it has the best performance.