**Assignment**

|  |  |
| --- | --- |
| **Title** | Assignment 1: Handling Large Data |
| **Due** | Jan 30, 2012 11:55 pm |
| **Grade Scale** | Points (max 100.0) |
| **Modified by instructor** | Jan 12, 2012 9:36 am |

**Instructions**

**Task**  
  
Write a java program to read in a table of data in the form of a CSV file and run queries on the table ***as efficiently as possible***. A template of the program is provided and your code should be added to the Assignment1.java file. A driver program, Driver.java is provided so that you can test your program. Your java program must support the following three operations:

* **load filename** : loads the csv file with filename into your program
* **searchEq columnNum value**: prints the rows of the table where the value in column number columnNum is equal to the given value.
* **searchGtr columnNum value**: prints the rows of the table where the value in column number columnNum is greater than the given value.

You must assume that ***the data may not fit in memory***, i.e., you will not be able to read all the data into an in-memory java data structure. You must NOT use any form of DBMS (eg. embedded DBMS) for this assignment, however, you may use any other third party libraries. You are free to design your program in any way you want while obeying the previous two constraints.  
  
This is an individual assignment and you may NOT work in groups. General discussions with peers are permitted and encouraged, but you must design and write your own code.  
  
A small sample data file is provided with the java code template. A larger 16 MB data file is also provided for further testing.  
  
One test case has been provided to ensure that your output conforms to the requirement. Type "runtest" and hit enter after logging onto the submission server.  
  
**Grading Criteria**  
  
To encourage students to get started early, 10 points will be awarded to students who submit a functionally complete program by the first submission dateline. The rest of the 90 points will be awarded based on the performance of the program as of the final submission.  
  
Assuming that the basic requirements described above are met, your grade will be contingent upon how fast your program runs on an unreleased data and workload. This assignment will be administered like a competition. Once your code is uploaded to the submission server, it will be benchmarked weekly and the results will be released in a leader board.  
  
**Submission Procedure**  
  
An account will be created for you on the submission server at:  
  
ec2-75-101-231-95.compute-1.amazonaws.com  
  
You should be able to login to your account remotely via SSH and be able to upload your java source file via SFTP or SCP.  
  
Your username will be your UH email id without the "@hawaii.edu". For example if you email is john@hawaii.edu, then you can login to the submission server using:  
  
ssh john@ec2-75-101-231-95.compute-1.amazonaws.com  
  
Your default password is "ics321". Please change it using the "passwd" command after first login.  
  
A directory "ics321a1" has been created with the default code template. Please upload your code to overwrite the default code template.  
  
**Running The Given Testcase**  
Once you have uploaded your code to the submission server and made sure that it compiles and executes, you can run the given testcase by typing "runtest" and <enter>.  
  
The entire testcase is located in /home/lipyeow/ics321a1/test01/.  
  
If there are differences in output, you might want to look at /home/lipyeow/ics321a1/test01/test01.exp for the correct output.