Test Data Sets for ALG 2011 Assignment

Data Set

There are 6 data sets. Each set has a pair of file, P1 and S1, P2 and S2, etc. P1, P2, etc. contain data about products, and S1, S2, etc. contain data about sales. File S6 is split into two because its size exceeds the size allowed by LMS.

All data sets are correct data sets, except for P1E.txt (which contains an error)

Set	Number of Products	Number of Sales	Maximum Number of Item Per Sale	Remarks
P1, S1	5	10	4	Tiny data set. You can inspect the data to verify your results
P2, S2	10	10	8	Small data set. Some of the top five lists have less than 5 products, some have more.
P3, S3	1000	10, 000	10	Should take a few seconds to run.
P4, S4	10,000	100,000	10	Likely to require more memory (than the default) to run. Request more memory with java –Xmx300m SaleInfoMiner …
P5, S5	10,000	500,000	10	Likely to require about 800MB to run (depending on your program)
P6, S6	15,000	1000,000	10	Likely to require about 1500MB to run. S6 is too big for LMS (> 100KB), so it is split into Part1 and Part2

How to use the test data sets

1. **Do NOT run your programs on latcs6.**
   Run them on one of the simula servers (simula1, simulal2, etc)

   If you run your program with the larger data sets on latcs6, you may slow down the server so much that it prevents other students on latcs6 from doing their work

2. You would log in and run your programs on simula the way you do it on latcs6.

3. To request more memory for Java execution, issue, for example:

   ```
   java –Xmx300m SaleInfoMiner … … … …
   ```

4. The data set will be put on the server in alg/assignment subdirectory very shortly (hopefully on Monday 02/05/2011)

5. To save disk space, do not copy the larger data sets to your own area.

6. You can copy the data sets to use on your own PC's.

∎