Fall 2006: CS 333 – UNIX Operating Systems Fundamentals

Homework – 2

100 points. Individual work only. **Due: October 13, 2006**.

1. Using the ***vi*** editor, type the following data and save it to the ***nr_timings_100_1*** in the ***hw2*** directory (under ***~/cs333/fall2006***). The data represents timings for executing a set of BLAST queries where the first column is the length of a BLAST query, the second one is the note this is the real system time taken for the run to complete, while the third column gives the actual time taken for the run. Please note that each column is separated by a colon as well that the way timings are reported vary between minutes and seconds. [**10 points**]

```
431:real:22.976
689:real:33.953
913:real:47.169
1943:real:1m27.655
48:real:6.459
584:real:30.445
672:real:34.512
1029:real:52.034
1959:real:1m29.088
51:real:6.292
93:real:8.615
662:real:34.390
978:real:46.155
2134:real:1m33.885
54:real:6.192
100:real:9.242
209:real:14.726
274:real:18.392
```

2. Using the ***vi*** editor, type the following data and save it to the ***nr_timings_100_2*** in the ***hw2*** directory (under ***~/cs333/fall2006***). The data represents timings for executing a set of BLAST queries again, but this time the single number on a line represents the query length while the timings are the formatted as direct output from UNIX ***time*** command. Note that tabs are used as delimiters within ***time*** command output. [**10 points**]

```
389
real  0m15.803s
user  0m30.140s
sys   0m1.000s

536
real  0m27.015s
user  0m52.880s
sys   0m0.640s

657
real  0m30.631s
user  0m59.690s
sys   0m0.930s

867
real  0m43.813s
user  1m26.140s
sys   0m0.970s

1984
real  1m24.468s
user  2m46.520s
```

```
sys    0m1.350s

41
real   0m6.460s
user   0m11.890s
sys    0m0.680s
```

3. Edit the file ***nr_timings_100_1*** using ***vi*** and add a new row at the top of the document that provides descriptions for the columns found in the file (*i.e.*, QuerySize:RealTime:Time). [**5 points**]

4. There is a separate file called ***num_threads_100_1*** where the number of threads used for this run is saved as the only value in the file (create this file and put number ***2*** as the only thing in the file). Given number of threads is shared between all values for the corresponding run. Use UNIX utilities to add this value to the end of each row of the ***nr_timings_100_1*** file (*e.g.*, `431:real:22.976:2`). Note that the added value should be separated from the last element in each row by a colon. Write down the command(s) used to make the addition and include the contents of the file with the homework submission (name the file ***nr_timings_100_1_t***) after the change. [**10 points**]

5. Write a shell script (using UNIX utilities) to combine the two files (***nr_timings_100_1 and nr_timings_100_2***) and create a new file ***nr_timings_100*** that has the same format as file ***nr_timings_100_1***, but tabs are used as separators instead of colons (*e.g.*, `274real 18.392)`. Please note that this involves parsing of the ***nr_timings_100_2*** file to extract value for the size of the query as well as real time stamp. With respect to the time stamp, make sure the format is the same as is in the original ***nr_timings_100_1*** file (*i.e.*, if time taken is less than one minute, `0m` should NOT displayed). The script should print the following status lines as it executes: [**35 points]**

   No. of entries in file ***nr_timings_100_1*** = #
   No. of entries in file ***nr_timings_100_2*** = #
   New run time file ***nr_timings_100***:
   <Query length real time>

6. For the following operations write down the command used along with the options and the output produced: [**20 points (10 points each)**]
   a. Convert all the timings in ***nr_timings_100*** file to use only numerical values for the run times and save those to the same file ***nr_timings_100.*** You must eliminate any characters and convert times that were using minutes into seconds (*e.g.,* `1m24.468s = 84.468)`
   b. Display the total min, max, and average time taken by all the runs saved in the ***nr_timings_100*** along with the corresponding query length (***e.g.***, `Min: 54 6.192)`

7. For the following commands explain what action is performed and write down the output produced: [**12 points (4 points each)**]

```
a. cut -f -3,5,7-9 -d ' ' infile1 > outfile1
b. sort -n -o report +1 -2 month[1-3]
c. head -4 nr_timings_100 | sed 's!real!real system time!;s/1m/60+/g'
```

NOTES:

1. Submit typed answers for questions 4, 5, 6 and 7 in class on the due date.
2. Create a compressed file with the directory ~/cs333/fall2006/hw2. Name the compressed file **<blazerID>_cs333f06_hw2.tar.gz** and email me the file (to **afgane@uab.edu**). The subject of the e-mail should be ***CS333-HW2***. Note that you should perform this step after you have finished the homework. [10 points penalty for each 24 hour period after due date (weekends count as 24 hours)].