CS245 Winter 2012
Assignment 1
Due: Thursday January 19 in class

Warning: This is an open-ended problem. Your numeric answers are not as important as the arguments you use, and the issues you uncover in your analysis. You will be given credit as long as you make a reasonable effort to answer the questions.

You want to purchase hardware to run a database system, with a database of approximately 1000 gigabytes. You are considering two types of storage:
** System M uses RAM memory to hold the data
** System D uses rotating hard disks for storage.
** (If you are ambitious, also consider a system S that uses solid state disks for storage.)

(a) Estimate the cost of each of these storage systems (ignoring processor costs). You can do this by going to the web site of your favorite vendor (Dell, IBM, HP, ...) and seeing what they offer, and what CURRENT prices are. If you cannot find the right product, it is Ok to guess. (Example, if you cannot find a memory sub-system of the right kind, see what a memory card costs, and then estimate what a "box" with the right number of cards would cost.)

Do not forget that for the RAM subsystem you need to purchase an un-interruptible power supply (so you do not lose your database during a power outage). Add the cost of this component to your budget.

Explain how you obtain your estimates, and state what assumptions or educated guesses you are making. Do you think it would it be realistic and/or cost effective to build such a storage system? Why or why not?

(b) Read about database benchmarks at
http://www.tpc.org/information/benchmarks.asp
Explore the site so you learn about the TPC Council and what it does.

(c) You have done some testing with the two systems, and your experiments indicate that system M is 10 times faster than system D, i.e., tpm(M) = 10*tpm(D). The non-storage related system cost is 5,000 dollars, and is the same for both systems. How do you think the two systems compare using the $/tpm metric? Use your data on storage costs (part a) for this estimate. Also, comment on whether you think the tpm(M) = 10*tpm(D) assumption is realistic.