Prepare your answers to the following questions either in a plain text file or in a file that can be opened with Microsoft Word. Submit your file to the Curator system (www.cs.vt.edu/curator) under the heading HW2, by the posted deadline for this assignment. No late submissions will be accepted.

1. Assume a system uses a hard drive with the following physical characteristics:

total capacity	256 GB
# of platters	4
# of tracks per surface	16384
avg # of sectors per track	8192
cluster size	4 KiB
spindle speed	10000 RPM
head start time	0.1 ms
track to track seek time	0.01 ms

The sector size is 512 bytes. In answering the following questions, express all final time values to the nearest hundredth of a millisecond (8.33 ms).

a) [15 points] What is the average random head seek time for this drive?

From the notes, the average random seek time is:

(1/3)*(#tracks per surface)*(track to track seek time) + (head start time)

Using the data above, that works out to be 54.7133 ms (which is not exactly great performance).

b) [15 points] What is the average rotational latency for this drive?

From the notes, the average rotational latency is the time required for ½ of a full rotation: $\frac{1}{2}\times\frac{60000}{R}$

where R is the number of rotations per minute. Since the spindle speed is 10000 RPM, the rotational latency is

$$\frac{1}{2}\times\frac{60000}{10000}=3.0000\text{ ms}$$

c) [10 points] What is the average total time required to read one randomly-chosen sector from this drive?

The time is the average seek time plus the average rotational latency plus the time for a sector to rotate past the read head (the transfer time). The first two were determined above. From the notes, the transfer time is given by

$$\text{Transfer}(n)=\left(\frac{n}{S_T}\right)\times\left(\frac{60000}{R}\right)$$

where n is the number of sectors to be read and S_T is the number of sectors per track and R is as in part b. Using the given values, the transfer time is $\left(\frac{1}{8192}\right)\times\left(\frac{60000}{10000}\right)=\frac{6}{8192}\approx 0.0007\text{ ms}$

So, the average time to read a random sector is about 54.7133 + 3.00 + .0007 ms or 57.7140 ms.

d) [10 points] What is the average total time required to read a file of 10 MiB from this drive if the clusters are randomly scattered on the drive?

The answer would be (# of clusters file occupies) * (average time to read one cluster). Since the cluster size is 4 KB and the file is 10 MB, the file will occupy exactly 2650 clusters.

The time to read one cluster would be the average seek time plus the average rotational latency plus the transfer time for a single cluster. We already know the first two values, and since the cluster size is 4 KiB and a sector is 512 bytes, there are 8 sectors in a cluster.

Thus, the transfer time for a cluster is given by:

$$\left(\frac{8}{8192}\right)\times\left(\frac{60000}{10000}\right)=\frac{48}{8192}\approx 0.005859 \text{ ms} \approx 0.0059 \text{ ms}$$

The average time to read one random cluster is 54.7133 + 3.0000 + .0059 = 57.7192 ms.

Since the file contains 2560 clusters, the total read time is about 2560 * 57.7192 or 147761.152 ms (which is nearly 2.5 minutes).

2. Assume each of the changes described below were made to the design of the drive described in the preceding question, and that no other changes were made. Indicate the effect of the change on the average random head seek time and the average rotational latency time for the drive. The parts are independent. Justify your conclusions if you want credit.

   a) [10 points] doubling the rotational speed, and modifying the read/write heads so that they can keep up with the increased transfer rate

   Doubling the rotational speed has no effect on the seek time.

   But it will cut the average rotational latency in half since that depends solely on the rotational speed.

   b) [10 points] reducing the track spacing by half and doubling the number of tracks per surface (which would alter the total capacity of the device); the track-to-track seek time remains the same

   Changing the track layout has no effect on the rotational latency.

   As for the seek time, doubling the number of tracks per surface while keeping the track-to-track seek time the same would increase the average seek time (almost doubling it since the head start time is small).

3. Let T be a PR quadtree storing data points that lie in a square with side *S*. According to the survey paper by Samet, if *D* is the minimum distance between any two of the data points, the maximum number of levels in a PR quadtree is:

$$\left\lceil \log\left(S\sqrt{2}/D\right)\right\rceil$$

   Suppose that the points in a given data set lie within a square whose side is 2^{20} units, and that the closest pair of data points are a distance of 2^3 units apart.

   a) [10 points] What is the maximum height of a PR quadtree that stores the entire data set?

   We are given that S is 2^{20} and that D is 2^3, so we can simply apply the formula to find the maximum height:

$$\left\lceil \log\left(2^{20}\sqrt{2}/2^3\right)\right\rceil = \left\lceil \log\left(2^{17.5}\right)\right\rceil = \lceil 17.5 \rceil = 18$$

   b) [10 points] Explain why the PR quadtree might actually be much shorter than the bound you gave in the previous part.

   The worst case bound is determined by the number of region splittings that must occur in order to separate the closest pair of data points.

   But, if the closest pair of data points are close to, but separated by, an earlier division then we will not get the worst case. For example, if the closest pair of data points are just on either side of the vertical line that separates the east and west halves of the world rectangle then the very first splitting would separate them from each other.

   c) [10 points] Explain, in one clear sentence, precisely why the number $\sqrt{2}$ gets to be involved in the formula for the bound.

   The bound depends on how far apart two points can be and still lie in the same square, and that depends on the length of the diagonal of the square, and that's given by $S\sqrt{2}$ if the length of the side is S.