1. You are to sort a file of 8,800,000 pages with a buffer of 220 pages and a disk whose seek+latency time is 14ms and whose page transfer time is 1ms. The initial pass will create $8,800,000/220=40,000$ runs (of 220 pages each). Determine the time for the subsequent passes using the following configurations:

   a) 200 input buffers of 1 page each, and 1 output buffer of 20 pages.
   b) 10 input and 1 output buffer, each of 20 pages.

2. Using the company schema, we want to determine the amount of disk I/O to compute the join of the employee and department tables. Suppose employee has 6000 pages and department has 1000 pages. What is the disk I/O for the following approaches?

   a) Nested loop join with 102 buffer pages.
   b) Sort-merge join with 101 buffer pages.

3. For the company database, write RA expressions for the following two queries. Also, write a DRC expression for one of the queries and a TRC expression for the other.

   a) List the names of all departments with an employee making more than $40,000.
   b) Give the names and birthdates of all dependents of any employee whose department is located in Sugarland.

4. See the other side
4. Consider the recovery process in the event of a crash given that the following logical log is found on disk at the time of restarting.

```
00   hello
10   update: t1 writes p1
20   update: t1 writes p2
30   update: t2 writes p1
40   update: t3 writes p3
50   checkpoint (*)
60   update: t3 writes p3
70   update: t3 writes p4
80   update: t4 writes p2
90   commit: t2
100  update: t1 writes p5
110  abort: t1
120  end: t2
130  CLR: t1 undo LSN 100
140  update: t4 writes p6
crash
```

* Suppose the checkpoint contains the DPT={ (p1,30), (p3,40)} and Xact table {t1,20,active), (t2,30,active), (t3,40,active)}. Note that this means that pages p1 and p2 were somehow cleaned prior to LSN 30.

a) Outline the Analysis phase.
b) From where does the Redo phase begin?
c) Show the log records that are added to the log during the Undo phase.