1. You are to sort a file of 4,000,000 pages with a buffer of 64 pages and a disk whose seek and latency time together is 12ms and whose page transfer time is 1ms. The initial pass will create 4,000,000/64 = 62,500 runs (of 64 pages each). Do not count the time of that initial pass. Determine the time for the subsequent passes using the following configurations:
   a) 3 input buffers and 1 output buffer, each of 16 pages.
   b) 5 input buffers of 8 pages each, and an output buffer of 24 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 5000 pages and department has 250 pages. What is the disk I/O for the following approaches?
   a) Block nested loop join with 52 buffer pages.
   b) Sort-merge join with 51 buffer pages.
   c) Hash join with 100 buffer pages. (NB 100 > √5000)

3. On the company database, write an SQL query for the following:
   Find the first and last names of all women who make over $35,000 and who work on a project located in Houston (a project’s location is given by plocation).
   a) Convert the SQL to a relational algebra tree.
   b) Optimize the tree, and justify your choices.

4. Suppose that, on the company database, the employee table has 1000 tuples, the works_on table 2500 tuples, and the project table 400 tuples. How many tuples will the natural join of these three tables have?

5. See the other side
5. 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   checkpoint (*)
10   update: t1 writes p2
20   update: t1 writes p7
30   update: t2 writes p2
40   update: t3 writes p1
50   update t4 writes p3
60   update: t2 writes p2
70   update: t4 writes p1
80   abort: t4
90   checkpoint (**)
100  commit: t1
110  end: t1
120  CLR: t4 uno LSN 70
130  update: t3 writes p4
```

* This checkpoint record contains an empty DPT and Xact table.
** This checkpoint contains the DPT={ (p1,70), (p2, 60), (p3, 50), (p7, 20)} and Xact table {(t1,20,active), (t2,60,active), (t3,40,active), (t4, 80, aborting)}. Note that this means that pages p1 and p2 were cleaned at some point before being used again.

a) From where does the Redo phase begin? Which LSNs need to be redone?
b) Show the log records that are added to the log during the Undo phase. Include prevLSNs and undonextLSNs for the entire log.