CSC 370 – Fall 2017 – Assignment 4
Due date: Dec 1, 11:55pm

Ex. 1 (7 pts)
Consider a disk with average seek time of 10 ms, average rotational latency of 5 ms, and a transfer time of 1 ms for a 4KB block. The cost of reading/writing a block is the sum of these values (i.e. 16 ms). We are asked to sort a large relation consisting of 10,000,000 blocks of 4KB each. For this, we use a computer on which the main memory available for buffering is 320 blocks (*somewhat small memory*). We begin as usual by creating sorted runs of 320 blocks each in phase 1. Then, we do 319-way merges. Determine the number of phases needed, and evaluate the cost of the Multi Phase Multiway Merge Sort.
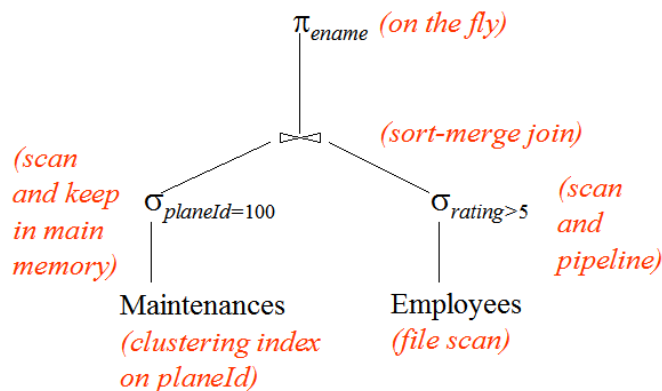
Ex. 2 (7 pts)
Build a B+ tree index with n=3 using the following sequence of keys:
        2, 15, 31, 32, 6, 18, 19, 20, 3, 4, 5, 40, 41, 42
Redraw the tree each time an insertion is done.

Ex. 3 (5 pts)
Consider the following query plan.

What is the cost in terms of number of I/Os for this plan?
Notes. The result of the left selection, being small, is kept in main memory, where it is sorted. The result of the right selection is pipelined to the join operator, i.e. the generation of the sorted sublists for the first phase of sort is done on the fly.
Do not count the I/Os for writing the final results (after projection).
Consult *queryeval.pdf* for the table statistics.

Ex. 4. (6 pts)
For each of the following relation schemas and sets of FD's:

a)   R(A,B,C,D) with FD's AB→C, B→D, CD→A, AD→B.
b)   R(A,B,C,D) with FD's A→B, B→C, C→D, D→A.

Indicate the BCNF violations (if any). Decompose the relations, as necessary, into relations that are in BCNF.