INF 551 – Fall 2017 (Afternoon section)
Quiz 11: Query execution (10 points), 15 minutes

Note: in all questions, we ignore the cost of outputting the result of join.

1. [5 points] Suppose M = **101** blocks, B(R) = 2,000 blocks, B(S) = 6,000 blocks. Suppose we use 100 blocks in sorting. Describe the steps (what each step does, what does it output, what is the size of the output) of computing R ⋈ S using the **simple sort-merge** join algorithm. What is the cost of the algorithm?

   Step 1: Sort R: load 100 blocks of R at a time, sort them and write back on disk. This generates 20 runs, each of size 100 blocks.
   Cost=2B(R)=4000

   Step 2: Load 20 runs of R at a time, and merge them. It generates 1 run and the run size is 2000 pages
   Cost=2B(R)=4000

   Step 3: Sort S: load 100 blocks of S at a time, sort them and write back on disk. This generates 60 runs, each of size 100 blocks

   Cost=2B(S)=12000

   Step 4: Load 60 runs of S at a time, and sort them. It generates 1 run and run size is 6000 pages

   Cost=2B(S)=12000

   Step 5: Joining R and S using two-way merging, since only one run from R and one from S.
   Cost=B(R)+B(S)=8000

   Total cost=40000.

2. [5 points] Suppose M = 101 blocks (i.e., pages), B(R) = 20,000 blocks, B(S) = 50,000 blocks. Describe the steps of computing R ⋈ S using the **partitioned-hash** join algorithm. What is the cost of the algorithm?

   Step 1: Hash R into 100 buckets using hash function h, and send all buckets to disk. 200 blocks/bucket.
   Step 2: Hash S into 100 buckets using hash function h, and send all buckets to disk. 500 blocks/bucket.

   The memory is not enough to hold an entire bucket of R (200 blocks) or an entire bucket of S (500 blocks).

   Step 3: Hash each bucket of R into 100 buckets using hash function h', and send all buckets to disk. 2 blocks/bucket.
   Step 4: Hash each bucket of S into 100 buckets using hash function h', and send all buckets to disk. 5 blocks/bucket.
   Step 5: Join corresponding buckets from R and S (i.e., with the same bucket number after both hash).

   Total cost: 5B(R)+5B(S)=350,000