Question 5 (8 marks)

Duplicate removal in projection can be implemented using either sorting or hashing.

Consider a projection scenario where

- the original tuples are R_o = 40 bytes long
- the projected tuples are R_p = 20 bytes long
- the original file has b_o = 200 pages
- the projected file (with duplicates) has b_p = 100 pages
- the result file (no duplicates) has b_r = 80 pages

Assume that we have a buffer pool with B = 31 buffers available for exclusive use of the projection operation. Assume also that the hash function distributes tuples evenly, and that the total number of pages in the hash partitions is the same as b_p.

a. calculate the total cost of projection using sorting for duplicate removal

b. calculate the total cost of projection using hashing for duplicate removal

Costs are in terms of total page read/writes. Include the cost of reading the original file and the cost of writing the result file.

State all assumptions. Show all working.

Instructions:

- Type your answer to this question into the file called `q5.txt`
- Submit via: **`give cs9315 exam_q5 q5.txt`**
  or via: Webcms3 > exams > Final Exam > Q5 submission > Make Submission

End of Question