ITWS-4250/6250 Database Applications and Systems

Lab 11

**Instructions:**

* In this lab, you will apply the concepts of index structures, query execution and query compilers to answer **6 questions**.
* You will have roughly 90 minutes during the scheduled lecture period to complete it, but it’s not due on LMS until **EOD tonight**.
* Before getting started, you should have a repo setup and have both Sindhu ([mandas@rpi.edu](mailto:mandas@rpi.edu)) and me rplotka ([rplotka@rpi.edu](mailto:rplotka@rpi.edu)) as collaborators. The repos must be private.
* Create folders for each HW, Lab, and other assignments as they come up. Remember to push your repo before submitting your folder zip to LMS for credit.
* You may consult with your groupmates, but you must do and submit your own work.

**Questions:**

1. Suppose blocks hold either three records, or ten key-pointer pairs. As a function of n, the number of records, how many blocks do we need to hold a data file and:
   1. A dense index
   2. A sparse index

For the data file, it requires n/3 blocks. For a dense index, it requires n/10 blocks and 13n/30 blocks in total. For sparse index, it requires n/30 blocks and 11n/30 blocks in total.

1. Repeat Question 1 if blocks can hold up to 30 records or 200 key-pointer pairs, but neither data- nor index-blocks are allowed to be more than 80% full.

For dense index, it requires n/30 blocks to store the data and n/200 to store the index. Because the blocks is restricted to 80% full, so the required block should time 1.25. The total number of block is 23n/600 \* 1.25 = 23n/400.

For sparse index, it requires 7n/200 \* 1.25= 7n/160.

1. Suppose that we use the hash function on integer keys *i* defined by h(i) = i2 mod B, where B is the number of buckets.
   1. What is wrong with this hash function if B = 10?
   2. How good is this hash function if B = 16?
   3. Are there values of B for which this hash function is useful?

The problem with value of 10 is that the bucket number can only be in a small range (0 to 9), which will generate much collision. Therefore, it is not idea to set B as 10. As for value of 16, it has the same problem with value of 10 for too small value. A useful value of B should be bigger the better.

1. Suppose we have a file of 1,000,000 records that we want to hash into a table with 1000 buckets. 100 records will fit in a block, and we wish to keep blocks as full as possible, but not allow two buckets to share a block. What are the minimum and maximum number of blocks that we could need to store this hash table?

For the best case, the records can perfectly fit in every block, so the minimum number of blocks will be 1,000,000/100 =10,000. However, if the records can not fit well and each block has only one record with the remaining record in the same block. Therefore, the remaining blocks has 999,001 records with 9991 blocks. So, the total number of blocks is 10,990.

1. For each of the three operators: Projection, Distinct, Grouping, tell whether the operator is blocking, by which we mean that the first output cannot be produced until all the input has been read. Put another way, a blocking operator is one whose only possible iterators have all the important work done by Open.

Projection and Distinct is not blocking. However, Grouping is blocking because it is a aggregation operation which means it require to consume all rows first and then produce the result.

1. Using the simple SQL grammar exhibited in this section, give parse trees for the following queries about relations R(a,b) and S(b,c):
   1. SELECT a, c FROM R, S WHERE R.b = S.b;
   2. SELECT a FROM R WHERE b IN (SELECT a FROM R, S WHERE R.b = S.b);



