

TASK 1: Evaluation of Performance

The SimpleDB buffer manager traverses the buffer pool sequentially when searching for pages, which is not an efficient technique. We have implemented a new traversal technique using data hash tables to improve search times.

Strategy: we have tested the improvements by incrementing the blocks & buffer size in 1000. As seen from the code below initially we set the buffer size to 1000 and we programmatically created 1000 blocks and pinned. Subsequently we unpinned these 1000 blocks. After the end of this operation we have SimpleDB buffer holding all unpinned blocks. Then we pin one valid block and calculate the time taken for the operation. Below snippet shows the performance testing strategy for 1000 blocks of size 400 bytes.

```
Block[] blocks = new Block[1000];
Buffer[] buffer = new Buffer[blocks.length];
SimpleDB.initFileLogAndBufferMgr("simpleDB");
BufferMgr bm = SimpleDB.bufferMgr();

for (int i = 0; i < blocks.length; i++) {
    blocks[i] = new Block("file.txt", i);
    buffer[i] = bm.pin(blocks[i]);
}

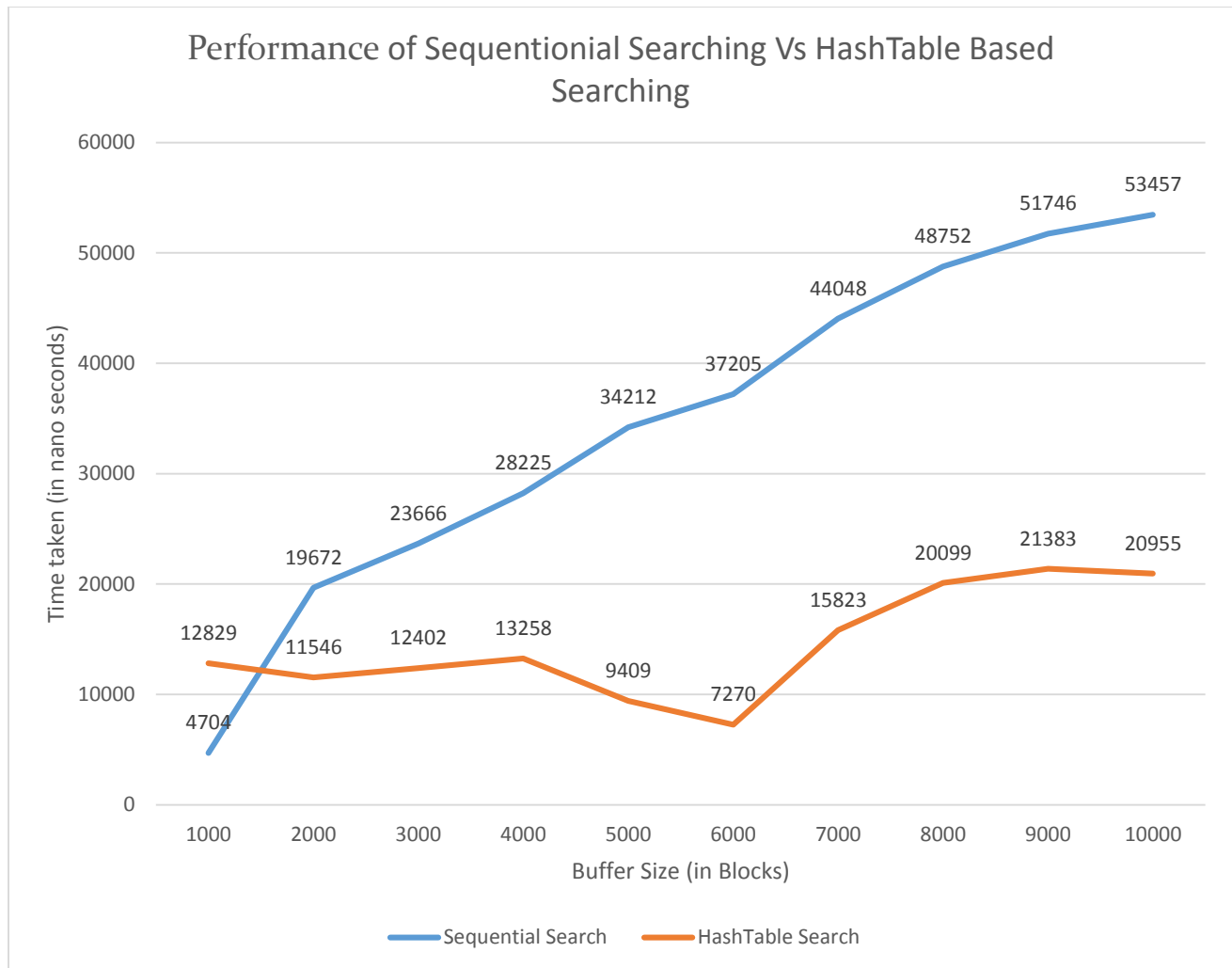
for (int i = 0; i < blocks.length; i++) {
    bm.unpin(buffer[i]);
}

Block blk = new Block("file.txt", blocks.length-1);
long starttime = System.nanoTime();
Buffer b = bm.pin(blk);
long endtime = System.nanoTime();
System.out.println("Time for operation is:" + (endtime - starttime));
```

Results:

Please find the experiment results collected.

Buffer Size (in Blocks)	Sequential Search	HashTable Search
1000	4704	12829
2000	19672	11546
3000	23666	12402
4000	28225	13258
5000	34212	9409
6000	37205	7270
7000	44048	15823
8000	48752	20099
9000	51746	21383
10000	53457	20955



Observations:

We see that the performance of searching has increased significantly in the case of Hash Table. For smaller values the sequential search is seen to be having better performance but as the number of blocks increases we see a clear distinction in the results in favor of HashTable. Hash Table based searching searches almost three times faster compared to sequential search. Also it is seen that sequential search time grows linearly as number of blocks increase where as HashTable based search has a very slow growth rate. Since in real-world enterprise level database, it is more likely that buffer size is huge in number HashTable based searching clearly has performance gain.