Bill Yang	Due: 3/9/19
Homework 5b	
CS386D Database Systems	Instructor: Daniel Miranker

1 Part A

15.3.2
$B(S) + (B(S)B(R))/(M-1) = 10000 + \lceil(10000/(1000-1)\rceil \times 10000) =$
120000 disk I/O.

15.3.3 a

$$\begin{aligned} 100000 &= 10000 + \lceil 10000/(M-1)\rceil(10000) \\ M &= 1112.11 \end{aligned}$$

$M \geq 1113$

16.2.6 a,b
a)
$\pi_{b+c\rightarrow x,y}(\pi_{b,c}(R) \bowtie \pi_{b,c,c+d\rightarrow y}(S))$
b)
$\pi_{a,b,a+d\rightarrow z}(\pi_{a,b,c}(R) \bowtie \pi_{b,c,d}(S))$

2 Part B

1.
Let c be the average seek time.
Then the total cost would be $1000c$, needing to seek and read from all $B(R)$ blocks.

2.
Let c be the average seek time and c' be the average weighted rotational latency and track to track seek time.
Then the cost would be $c + 999c'$, to seek the first block and then read the remaining blocks.

3.
a)
Let c be the average seek time.
Then the cost would be c, since $W.c$ is the primary key, it is unique, and we only need to read one record.

b)
Since it is the primary key, I expect there to be

1 record.

4.
a)
Let c be the average seek time.
Then the cost would be c,
since (c, b, a) is a compound primary key, it is unique, and we only need to read one record, ignoring the index (B-tree) lookups.

b)
1 record since (c, b, a) is a compound primary key.

5.
a)
Let c be the average seek time. There are 100 values for attribute a in the relation R, thus we expect there to be 100 values. A block contains 1000 values. Thus if sorted on primary key, all 100 values should fall into the same block.
Thus the total cost is c.

b)
Since there are 100 values for attribute a, then we expect there to be
100 records for the query.

6.
a)
Since there are 20 values of b, and 10 values of c, if we assume that all combinations exist in the table, there are 200 unique rows of W. Since there are 10000 total rows, we can assume for a specific tuple (b, c), there are 50 rows that match.
Thus we expect 50 rows for this query.

b)
For 20 values of b, we expect 500 rows for a specific b. For a specific c, we can expect 1000 rows. Thus the union, accounting for double-counting, would be $1500 - 50 =$
1450 rows in this query.

7.
a)
We assume 1/3 of the rows satisfy the inequality. Then for a specific c wold span 1/10 of those rows, getting us 1/30 of all rows, so
333 rows expected.

b)
Similar to $7a$, we have 3333 rows that meet the condition on b, and 1000 rows that meet the condition on c. Thus accounting for double counting, we get
4000 expected rows returned.

8.

0. No common attributes so no rows matched on a natural join.

9.
There are only 10 values of c in W, thus only 10/100 of the rows of V would get matched to a row in W. For a specific c value in W, there are 1000 rows of that value. Thus we have $100/10 = 10$ values of c that are matched to 1000 rows each in W.
10000 rows returned.

10.
From 6a, we saw that for a specific value of (b, c), there are 50 rows, and there are also only 200 total values of (b, c). In R, there are 10000 total values of (b, c). Thus for a only 200/10000 of the total rows of R would get matched in the join. 20000 rows of R joined to 50 rows each of W means 1000000 total rows returned.

11.
1/3 of the rows in R are matched by the inequality selection. $V(R, b) = 200$ and $V(U, b) = 400$, so all values of $R.b$ are contained in the values of $U.b$. For a single b value in U, there are 25 matching rows. Then we get
8333333 rows returned.

12.
Let us first consider the c join condition. There are 10 values of c in W so for a given c, there are 1000 rows. In R there are 50 values of c, so only 1/5 of the rows in R will be matched to something in W. Thus with 200000 rows matched to 1000 rows each, we get 200000000 rows returned.
Next we condition on $R.b > W.b$. Let us simplify further and consider $R.b > c$ for some constant c. Then 1/3 of the rows would satisfy this constraint based on our simplifying assumptions. Then for some value of $W.b$, we consider the rows that are created after joining on c, 1/3 of these rows should have $R.b > W.b$ based on our assumption.
Thus we have 66666667 rows returned.