Roy Chan
301202770

CMPT 354 Assignment 4: System Implementation, XML.

Storage and Indexing. 40 points total.

I. (20) Explain what the data entries in each of the following indexes contain. If the order of entries is significant, say so and explain why. For definition of terms, refer to text.

1. Index is not clustered or unclustered, index will remain as is like in the table.

    1.8, 2.0, 3.2, 3.4, 3.8

2. A clustered index is defined as the order of data entries being the same as the order of data records. By definition, order of entries is significant since it is a clustered index.

    $< 1.8, (1,1)>, < 3.4, (2,1)>$

3. An unclustered index is an index that is not clustered. From definition supplied in *1*, order of entries is not significant since it is an unclustered index.

    $< \text{Madayan}, (1,1)>, < \text{Guldu}, (1,2)>, < \text{Smith}, (1,3), (2.1)>, < \text{Jones}, (2,2)>$

II. (20) Consider a delete operation specified using an equality condition on a key. Assuming that no record qualifies, what is the cost for the three organizations: heap file, sorted file, unclustered hash index? Present your analysis using the same parameters as in the lectures.

a. Heap File: B(D+RC)
    Search the entire file anyway.

    For each page (B)
            Read page (D)
                    For each record (R), process record (C)

b. Sorted file: $D\log_2 B + C\log_2 R$

    Equality search to check that no qualifying record exists, despite no record qualifying. Then the cost is the same as an equality search.
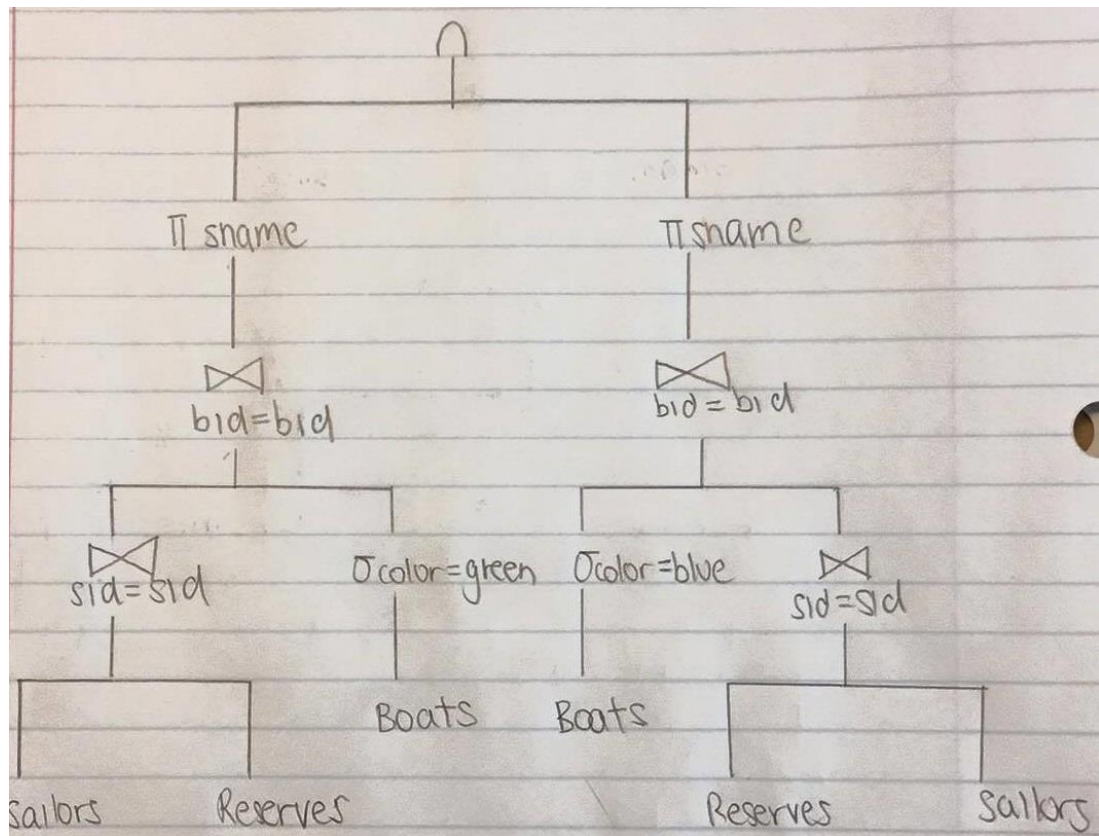
c. Unclustered hash index: $H + D + RC$

    For bucket containing hash value (H)

            Read Page (D)

                    For each record (R), process record (C)

Query Evaluation. 28 Points total.

I. (10) Write a query evaluation plan (relational algebra tree) for evaluating the given query.

II. (18) For each of the following indexes, list whether the index matches the given selection conditions. If there is a match, list the primary conjuncts.

a. Hash index on search key <Sailors.sid, Sailors.rating>
   - σ rating = 10 AND sid=500 (Sailors)

   Match.
   Primary Conjuncts: Sailors.sid=500 and Sailors.sid=500 ^ Sailors.rating=10

   - σ rating > 10 AND sid=500 (Sailors)

   Match.
   Primary Conjuncts: Sailors.sid=500

b. B+ tree on search key <Sailors.sid, Sailors.rating>
- σ sid< 500 AND rating=10 (Sailors)

   Match.
   Primary Conjuncts: Sailors.sid<500 and Sailors.sid<500 ^ Sailors.rating = 10

- σ sid= 500 AND rating > 10 (Sailors)

   Match.
   Primary Conjuncts: Sailors.sid<500 and Sailors.sid<500 ^ Sailors.rating>10

- σ sid< 500 (Sailors)

   Match.
   Primary Conjuncts: Sailors.sid<500

- σ rating > 10 (Sailors)

   No Match.
   Index is sorted on Sailors.sid. The entire relation would have to be searched for σ rating > 10 without Sailors.sid

Final Exam Question. 10 points total.

For the following table, explain what is included in a clustered index on *Departments* using Alternative (3). Explain why or why not this is a valid indices.

sID	Name	Department	Salary
10101	Srinivasan	Comp Sci.	65000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
33456	Gold	Physics	87000
45565	Katz	Comp Sci.	75000
98345	Kim	Elec. Eng.	80000