CS315 , IIT KANPUR

Assignment 5

Deadline: 5 April 2016, in class

Problem 1. Extendible Hashing

1. Suppose that we are using extendible hashing on a file that contains records with the following search-key values:
   (2369, 3760, 4692, 4871, 5659, 1821, 1074, 7115, 1620, 2428, 3943, 4750, 6975, 4981, 9208)
   Load these records into a file in the given order using extendible hashing (using least significant “global depth” bits). Assume that every block (bucket) can store up to four (4) values. Show the structure of the directory every 3 insertions. Use the hash function: $h(K) = K \bmod 128$.

2. Argue whether extendible hashing can handle duplicate values in the file or not.

Query Processing
The following problems use the schema and database statistics given below.

Sailors (sid: integer, sname: string, rating: integer, age: real)
Reserves (sid: integer, bid: integer, day: dates, rname: string)
Boats(bid: integer, bname: string, size: integer)

Reserves.sid is a foreign key to Sailors and Reserves.bid is a foreign key to Boats.bid. We are given the following information about the database:

1. Reserves contains 10000 records with 40 records per page.

2. Sailors contains 50,000 records with 20 records per page.

3. Boats contains 100 records with 10 records per page.

4. There are 50 values for Reserves.bid.

5. There are 10 values for Sailors.rating(1..10).

6. There are 10 values for Boat.size (1..10).

7. There are 2000 values for Reserves.day.

Problem 3. Consider the following query:

SELECT *
FROM Reserves R
WHERE R.bid = 10 AND R.sid = 1000

Suppose Reserves has a hash index on attribute sid and a B$^+$-tree clustered index on attribute bid. Evaluate the cost of the following plans in terms of the number of I/O operations. Also estimate the number of seek operations.

1. Scan the relation Reserves and retrieve matching tuples.
2. Use IndexScan on Reserves.sid and retrieve matching tuples.
3. Use IndexScan on Reserves.bid and retrieve matching tuples.
4. Use IndexScan on Reserves.bid and IndexScan on Reserves.sid, perform an record id intersection, and then retrieve the tuples.

Problem 4. Consider the following query.

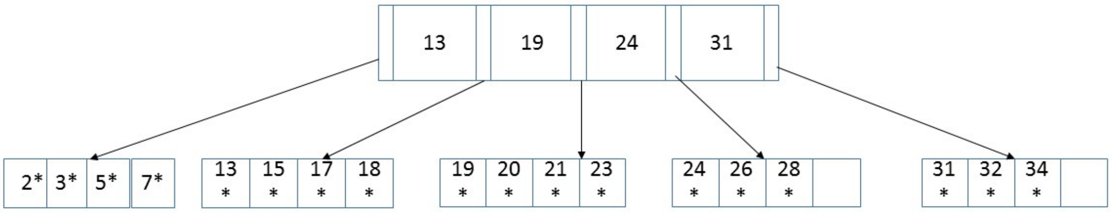
SELECT S.sid, S.sname, R.day
FROM Sailors S, Reserves R
WHERE S.sid=R.sid AND R.day = '31.3.2016'

For the following questions, assume uniform distribution of values and column independence. Suppose S has a clustered index on $S.Sid$ and R has an unclustered index on $R.day$ and an unclustered index on $R.sid$.

1. Estimate the number of tuples returned by the query.
2. Consider a block nested-loop join algorithm with R as the outer relation and S as the inner relation. Assume that there are 5 blocks of buffer allocated for R and one block for S. What is the cost of the solving the query under the following assumptions.
   a) A scan of relation R is used.
   b) An IndexScan of R.day is used.
3. Suppose a sort-merge join is used to solve the join predicate. Suppose there are 20 buffer blocks available in memory. Estimate the cost of the following plan.
   MergeJoin(IndexScan(S,S.sid),Sort(IndexScan(R, R.day = '31.3.2016'))

Problem 5. [B$^+$-tree] Consider a B^+-tree with $d = 4$.
Consider insertions and deletions as specified below using redistribution of keys.

1. Insert 29* into the tree.
2. In the tree resulting from (1), insert 30* into the tree.
3. In the tree resulting from (2), insert 16* into the tree.
4. In the original tree as given in the figure, delete 24*, 30*, and 32* and show the resulting tree.


13 19 24 31
2* 3* 5* 7*
13 * 15 * 17 * 18 *
19 * 20 * 21 * 23 *
24 * 26 * 28 *
31 * 32 * 34 *