CS3020 Database and Management Systems
Dr. Koninika Pal
Aniket Kumar,N Navya,Haripriya Chandra
IIT Palakkad
Exercise Sheet 6 : Hand out 20.04.23, Due 26.04.23 - 23 :59

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Assignment 1: Indexing

(1) Consider a database with a file containing 30,000 employee records of fixed length. Each record has two fields : Name (30 bytes) and city (9 bytes). A total of 1000 distinct cities are associated with employees. The file is ordered by the non-key field 'City' and you need to create an index on the 'City' using block anchors. The disk has a block size of 512 bytes, with a block pointer of 6 bytes and a record pointer of 7 bytes.

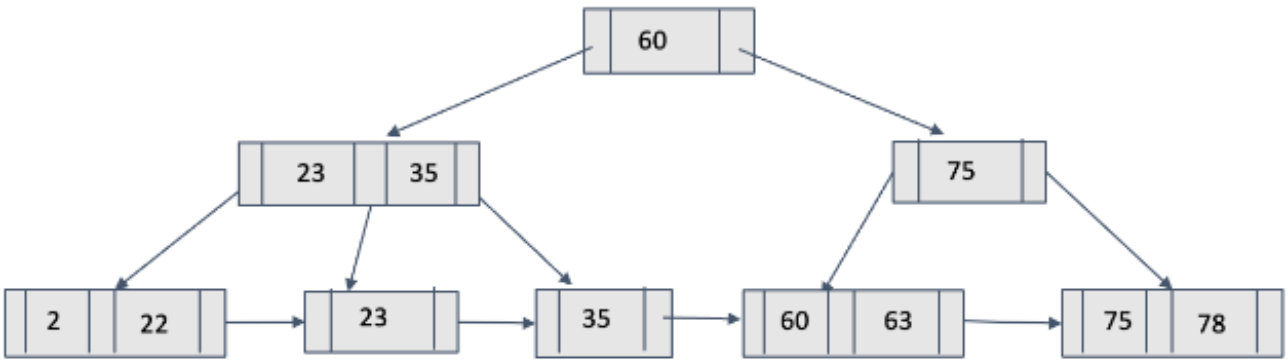
(i)What is the index blocking factor ?
(ii) How many index entries are there ?
(iii)How many levels are needed to create a multi-level index, such that the highest level can be fit into one block ?
(iv)What is the total number of blocks required by the multi-level index ?

Assignment 2: B+ Tree

(2)(i) Suppose we are building a B+ Tree on the EMP_ID attribute of the Employee relation. Assume that all employee names are of length 10 bytes. Disk block size is 512 bytes. Record pointer is 7 bytes. Block pointer is 6 bytes.

(a) What is the order of internal nodes of the B+ Tree ?
(b) What is the required number of elements in the leaf node of this Tree ?

(ii) Consider the following B+ Tree.



(a) What is the minimum number of nodes needed to be accessed (including the Root node) to retrieve all the records with a search key greater than or equal to 25 and less than 37".
(b) What is the height of the tree after the sequence of insertions 26, 27, 28, 30. Show intermediate results.
(c) On the final tree Delete 28, 63, 75. Show intermediate results.

Assignment 3: Query Optimization

(3) Consider the following Relations –
employee (employee_id , name , department_id)
department (department_id , department_name)
product (product_id, name, price, category)
sales(employee_id, product_id, quantity)

(i)Which type of join (hash or merge) would be more efficient for optimizing the following query and why ? Explain the procedure for the chosen join operation with illustration if any.

SELECT employee.employee name, department.department name
FROM employee JOIN department
ON employee.department id = department.department id ;

(ii) In the following query, which type of join (hash or merge) would be more efficient for optimizing the above query and why ? Explain the procedure for the chosen join operation with illustration if any.

SELECT employee.employee name, department.department name
FROM employee JOIN department
ON employee.department id = department.department id
ORDER BY dept id ;

(iii)Which index is better to improve the performance of the following query and why ?

SELECT name
FROM product
WHERE price <= 10000 AND price >= 5000

(iv)What is the equivalent relational algebra for the following query and what would be the optimized query plan ?

SELECT sales.quantity, product.name
FROM product NATURAL JOIN sales NATURAL JOIN employee
WHERE employee.name = Jerry and product.category = music and sales.quantity > 10

Assignment 4: Query Cost Estimation

(4) Consider a relation S (A, B) with the following characteristics :
- a total of 7,000 tuples appear in S and each block/page can hold maximum 70 tuples
- a hash index is created on attribute A
- The values of attribute A are integers and are uniformly distributed within [1, 200].

(i)Assuming that the aforesaid index on A is a non-cluster index, estimate the number of disk accesses needed to compute the query $\sigma_{A=18}$ (S).
(ii)What would be the cost estimate if the index were clustered ? Explain your reasoning.