National University of Computer and Emerging Sciences

Lahore Campus

Advance Database Concepts (CS4064)

Date: Sat, 12 Apr 2025

Course Instructor(s)

Muhammad Ishaq Raza, Muhammad Naveed

Sessional-2 Exam

Total Time (Hrs.): 1

Total Marks: 30

Total Questions: 4

_____ _____ _____

Roll No Section Student Signature

SOLUTION

Note: Please make sure to attempt all questions and their respective parts in the specified order. Failure to do so may result in a deduction of one mark for each incorrect part of a question.

CLO # 2 *Apply the models and approaches in order to become enabled to select and apply appropriate methods for a particular case.*

Q. No 1: Consider a file of book data that consists of *1,000,000* records, spread over *10,000* blocks. Assume there is 4 levels of secondary index exist on *BookID* key attribute, and 3 levels of cluster index exist on *Author* non-key attribute. For each of the following selection queries, estimate the I/O cost of the best possible solution, making use of the access paths available. Justify your answer. Take any valid assumption where needed. [6]

a. *SELECT * FROM book WHERE Author LIKE 'A%'; (Assume 5% of books return by this query)*

b. *SELECT * FROM book WHERE BookID= 7 or BookID= 9;*

Ans:

a. x + s/bfr = 3 + 50000/100 = 503 [s=50,000 i.e. 5% and bfr= 1,000,000/10,000= 100]

b. 2(x+1) = 2(4+1) = 10

CLO # 2 *Apply the models and approaches in order to become enabled to select and apply appropriate methods for a particular case.*

Q. No 2: Assume that the frequency of access to the following queries is very high. Identify the attributes that are more appropriate to create indexes to improve the performance of these queries and mention the type of each index (B-tree, Hash, or Bitmap). [6]

Query#1: *SELECT DISTINCT c.custName, c.gender, c.city*
FROM customer c JOIN order o ON c.custID=o.custID WHERE c.maritalStatus='Single';

Query#2: *SELECT custName, contact, highestDegree, rating*
FROM customer WHERE birthdate>= '01-Jan-2000';

Ans:

Index1 (for Query1): customer.maritalStatus – filter column (Bitmap Index)

Index2 (for Query1): order.custID – joining column (Hash Index)

Index3 (for Query2): customer.birthDate – filter column (B-tree Index)

CLO # 2 *Apply the models and approaches in order to become enabled to select and apply appropriate methods for a particular case.*

Q. No 3: Suppose you have the following search key values: *4, 15, 10, 16, 20, 12, 18, 5, 31*
Load the records with the above search key values into an expandable hash file based on extendible hashing. Show the structure of the directory at each step along with the global and local depths. Use the hash function *h(k)= k* and each bucket cannot hold more than *3* records. *Use lower bits (i.e., Right to Left) of hash value to determine a directory entry*. [8]

Ans:

H(Key)=key	4	15	10	16	20	12	18	5	31
Binary	00100	01111	01010	10000	10100	01100	10010	00101	11111

d=0

4	15	10

d=1

0: d'=1,	4	10	16	0	20 overflow
1:	15			d'=1	1

d=2

00:	4	16	20	d'=2	00	12 overflow
10:	10			d'=2	01	
01, 11:	15			d'=1	1	

d=3

000:	16			d'=3	000	
100:	4	20	12	d'=3	100	
010, 110:	10	18		d'=2	10	
001,011,101,111:	15	5	31	d'=1	1	

CLO # 2 *Apply the models and approaches in order to become enabled to select and apply appropriate methods for a particular case.*

Q. No 4: Assume: A block size is *B= 512* bytes, file has *r= 10,000,000* records, each record is *100* bytes long, a block pointer is *P= 6* bytes, a record pointer is P_R*= 7* bytes, and a key field for the index is *10 bytes* long. A database system uses a B^+-tree index on key field. A leaf node and non-leaf node are one block in size and contain as many keys (and appropriate pointers) as will fit in a block. How many blocks will this index use? Also estimate the number of block accesses needed to search for and retrieve a record from the file given its key value using the B^+-tree index. Show your working. [10]

Ans:

***order of p*: (p * 6) + (p-1) * 10) < 1024, which gives us order p = (512 + 10)/16 = 32**

***order of p_{leaf}*: (p * (10+7) + 6) < 1024, which gives us order pleaf = (512 - 6)/17 = 29**

This means the leaves (b1) will require [10,000,000/29] = 344,828 blocks.

Thus, our second level (b2) above the leaves will require ⌈344828/32⌉ = 10,776 blocks.
The third level (b3) above that will require ⌈10776/32⌉ = 337 blocks.
The fourth level (b4) above that will require ⌈337/32⌉ = 11 block.
The fifth level (b5) above that will require 1 block.
The total blocks this index uses are b1+b2+b3+b4+b5= 355,993 blocks.
Block access cost to search a key value= x+1= 5+1= 6