

National University of Computer and Emerging Sciences, Lahore Campus



Course:	Advance Database Concepts	Course Code:	CS4064
Program:	BS (Computer Science)	Semester:	Spring 2025
Out Date:	06-Mar-2025	Total Marks:	
Due Date:	Thu 13-Mar-2025 (Start of class)	Weight:	
		Page(s):	2
Assignment:	2 (File Structures & Hashing)		

Instructions:

- Use any valid assumption where needed.
- You are required to submit the hard copy of your assignment at the start of your class.
- For any queries, please contact your TA.

Question 1.

Consider a file system on a disk with block size $B=4096$ bytes. A file has $r=10,000,000$ STUDENT records of fixed length (un-spanned). Each record length is 150 bytes. Assume there are 50 departments and 75,000 students per department. Primary key of student file is RollNo.

Estimate the number of block fetches needed to compute the following queries:

- SELECT * FROM STUDENT WHERE rollno=9137; (Assume file is not ordered)
- SELECT * FROM STUDENT WHERE rollno=1371; (Assume file is ordered on rollno)
- SELECT * FROM STUDENT WHERE deptno= 34; (Assume file is not ordered)
- SELECT * FROM STUDENT WHERE deptno= 45; (Assume file is ordered on deptno)

Question 2.

A customer file has the following customer ID values:

(11, 24, 8, 2, 41, 6, 37, 27, 16)

- Consider these customer IDs as hash key values. The file uses 5 buckets named 0 to 4. One bucket cannot hold more than 2 records, meaning at max a bucket can hold 2 records. Load these records in file using hash function $h(k) = k \bmod 5$, in the given order.
- Calculate the average number of block accesses for random retrieval on customer ID.
- Load the given values in expandable hash files based on extendible hashing, show structure on each step, use hash function $h(k) = K \bmod 6$, max 2 records can be kept in one bucket. (Use most significant bits first i.e., left to right)
- Load the given values in expandable hash files based on dynamic hashing, show structure on each step, use hash function $h(k) = K \bmod 6$, max 2 records can be kept in one bucket. (Use most significant bits first i.e., left to right)
- Load the given values in expandable hash files based on linear hashing, show structure on each step, use hash function $h(k) = K \bmod 6$, max 2 records can be kept in one bucket.
- Show the structure after inserting an entry with customer ID value 13 into the final structure of part (c).
- Show the structure after inserting an entry with customer ID value 18 into the structure of part (f).

Question 3.

Suppose you are building an extensible hash index on a table of *100,000 rows*. Key values are *16 bytes*, a pointer (block/record) to a row is *16 bytes*, and a disk block is *4096 bytes*. Assume all keys are distinct.

- a. What is the (lowest possible) global depth? Provide valid reasons.
- b. What is the average occupancy of a bucket, assuming all buckets have a local depth equal to the global depth from part (a)? Justify your answer.