M.Sc. 2nd Sem 2022 (Mid-Sem) Examination

Full Marks-30 **Subject- Advance Database Management System (CSMC201)** **Time: 1 Hour**

1. Answer any five questions **(5 x 2 = 10)**

a. State the existing Functional Dependencies based on following instance.

X	Y	Z
1	4	2
1	5	3
1	6	3
3	2	2

b. R(A,B,C,D) is a relation. Which of the following does not have a lossless join, dependency preserving BCNF decomposition?

i. A →B, B →CD ii. A →B, B →C, C →D iii. AB →C, C→ AD iv. A →BCD

c. The following functional dependencies are given.

AB → CD, AF → D, DE → F, C→ G, F → E, G → A. What will be the Key?

d. The following functional dependencies hold true for the relational schema R (V,W,X,Y,Z).

V → W, VW → X, Y → VX, Y → Z.

Find the irreducible equivalent FD set for the given relation R.

e. Consider the set of FDs. F = { A→BC, CD → E, E → C, D → AEH, ABH → BD, DH → BC}.

Find out the canonical cover of F.

f. If an index contains data records as 'data entries', is it clustered or unclustered? Dense or sparse?

2. Answer any five from the following. **(5 x 4 = 20)**

a. Consider a relation stored as a randomly ordered file for which the only index is an unclustered index on a field called salary. If you want to retrieve all records with salary > 20000, is using the index always the best alternative? Explain.

b. Consider the following figure 1.

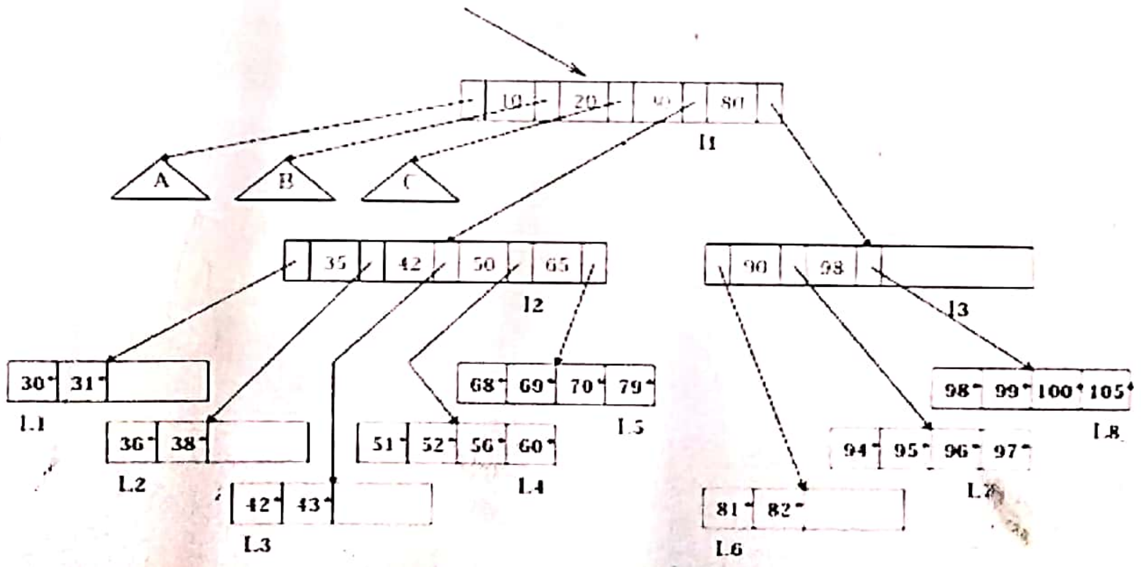


Figure 1

Insert a record with search key 109 into the tree.

Delete the record with search key 81 from the (original) tree.

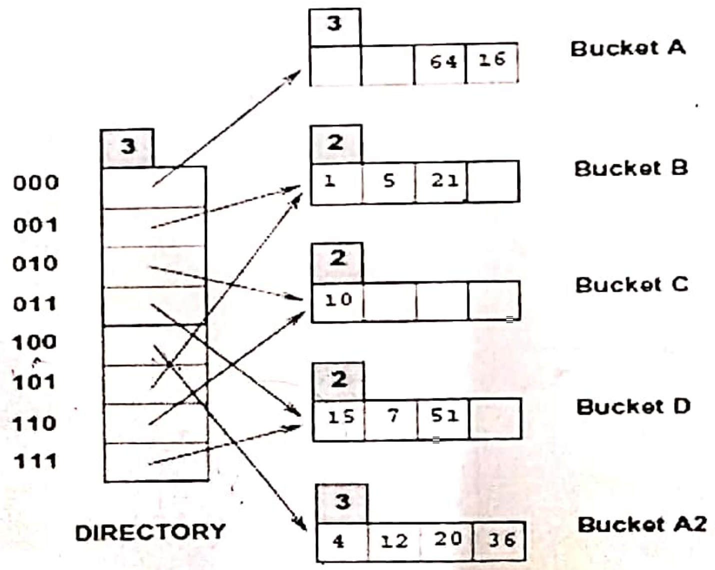
c. Deduce the expression for computing the cost of external sort algorithm that contains N records and four buffer pages.

d. Consider the following relational schema and the given query.
Employees (eno,ename, add, salary), Projects (pno, pname, pjob), Workin (eno,pno)
Select ename
Where eno IN
(Select eno
From Workin
Where Pno =
(Select Pno
From Projects
Where Pname = "DBMS"))
Order by ename

Give a detailed account of steps involved in optimizing the following query.

e. Consider a relation R (*a,b,c,d,e*) containing 5,000,000 records, where each data page of the relation holds 10 records. R is organized as a sorted file with dense secondary indexes. Assume that *R:a* is a candidate key for R, with values lying in the range 0 to 4,999,999, and that R is stored in *R:a* order.
For each of the following relational algebra queries, state which of the following three approaches is most likely to be the cheapest:
Access the sorted file for R directly.
Use a (clustered) B+ tree index on attribute *R:a*.
Use a linear hashed index on attribute *R:a*.
1. $a>50000$ ^ $a<50010(R)$
4. $a <> 50000(R)$

f. Consider the following figure 2.



Show the index after inserting an entry with hash value 68. Suppose you have told that there have been no deletions from the index so far. What can you say about the last entry whose insertion into the index causes a split?