

# Database Systems: 2<sup>nd</sup> Mid

Need to revise today

The simple question

Solve Panakrishna as well

Date: 17<sup>th</sup> March 2017

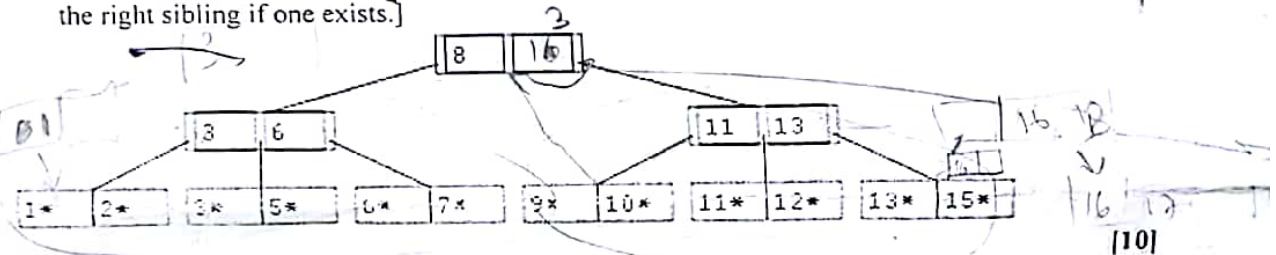
Duration: 1.5 hrs

1. No clarifications during the exam.
2. Make reasonable assumptions and clearly state them to answer ambiguous questions.
3. Show your steps. Be concise and organized.
4. Calculators allowed. Sharing of calculators not allowed.

1) Describe steps, using diagrams if necessary, to execute the following operations on the shown B+ tree:

- (a) Lookup all records in range 5 to 10
- (b) Lookup all records less than 14
- (c) Insert elements (16, 17, 18)
- (d) Insert record with key 0
- (e) Delete elements (11, 12, 16)

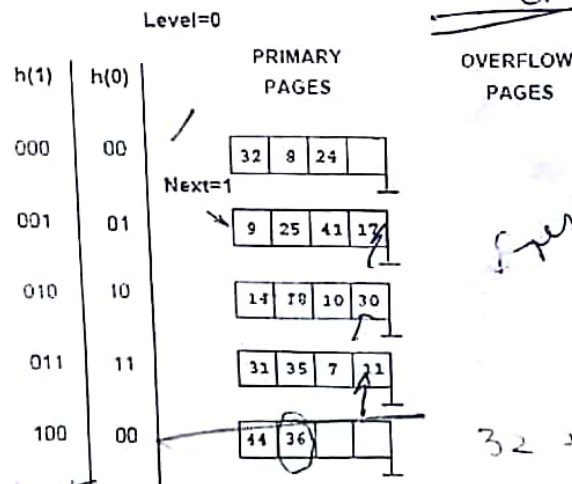
Assume the deletion algorithm tries to merge/redistribute with the right sibling if one exists.



2) Consider the Linear Hashing index shown in below Figure. Assume that we split whenever an overflow page is created. Answer the following questions about this index:

11.2

R. Panakrishna 11.2



(a) What can you say about the last entry that was inserted into the index if you know that there have been no deletions from this index so far?

(b) Suppose you know that there have been no deletions from this index so far. What

should have written assumption

00111

32 + 8 + 1  
41  
2 25 18

can you say about the last entry whose insertion into the index caused a split?

(c) Show the index after inserting an entry with hash value 4.

(d) Show the index after inserting an entry with hash value 15.

(e) Show the index after *fully* deleting the entries with hash values 36 and 44.

[10]

3) Suppose we store a relation  $R(x,y)$  in a grid file. Both attributes have a range of values from 0 to 1000. The partitions of this grid file happen to be uniformly spaced; for  $x$  there are partitions every 20 units, at 20, 40, 60, and so on, while for  $y$  the partitions are every 50 units.

(a) How many buckets do we have to examine to answer the range query:

SELECT \*

FROM R

WHERE  $310 < x$  AND  $x < 400$  AND  $520 < y$  AND  $y < 730$ ;

310 320 340 360 380 400 420 440 460 480 500 520 540 560 580 600 620 640 660 680 700 720 740

(b) How many disk accesses are needed to answer the above query using the grid file?

4) Compute the cost of  $r \bowtie_{A=B} s$  using the following methods:

[10]

(a) Nested loops

(b) Block-nested loops

(c) Index-nested loops with a hash index on B in  $s$ . (Do the computation for both clustered and unclustered index.)

tuple 0 49 part

144

8

where  $r$  occupies 2,000 pages, 20 tuples per page,  $s$  occupies 5,000 pages, 5 tuples per page, and the amount of main memory available for block-nested loops join is 402 pages. Assume that at most 5 tuples in  $s$  match each tuple in  $r$ .

[15]

5) Suppose a database has the following schema:

TRIP(fromAddrId: INTEGER, toAddrId: INTEGER, date: DATE)

ADDRESS(id: INTEGER, street: STRING, townState: STRING)

(a) Write an SQL query that returns the street of all addresses in 'Stony Brook NY' that are destination of a trip on '5/14/02'.

(b) Translate the SQL query in (a) into the corresponding "naive" relational algebra expression.

(c) Draw a query tree for the expression in (b).

(d) Translate the relational algebra expression in (b) into an equivalent expression using pushing of selections and projections.

(e) Translate the relational algebra expression in (c) into a most directly corresponding SQL query.

[10]