

# COL 362/632

## Introduction to Database Management Systems

Prof. Kaustubh Beedkar

Midterm Exam

Indian Institute of Technology Delhi  
Computer Science & Engineering

First Name:	KUSHAGRA	Entry Number:	2021CS50592
Last Name:	Gupta	Degree Program:	DUAL (B.TECH+M.TECH)

Do not write here

Question	Points	Score
1	3	
2	5	
3	2	
4	10	
5	10	
6	2	
7	2	
8	12	

Question	Points	Score
9	2	
10	10	
11	10	
12	12	
13	12	
14	8	
Total:	100	

### Exam Information

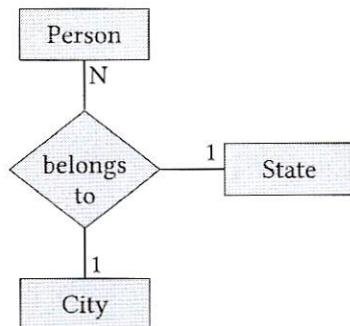
- Do not turn over the pages until told so.
- Make sure your exam is complete (22 pages). Note that the exam sheets are printed on both sides.
- Place your Student ID card on your desk.
- No auxiliaries (e.g., books, notes, laptop, tablet, phone, calculator, etc.) are allowed.
- You have **120 minutes** for the exam.
- There are 100 points in total. The number of achievable points per question can be found in the table as well as on the respective question sheet.
- Start with the questions that are easier for you.
- Answer the questions only in the spaces provided on the question sheets. Please be precise and succinct and avoid verbose answers. You may use the scratch space and empty sheets at the end.
- Please write in a clear and legible handwriting.
- Write your entry number on top of each page

I Wish you success!

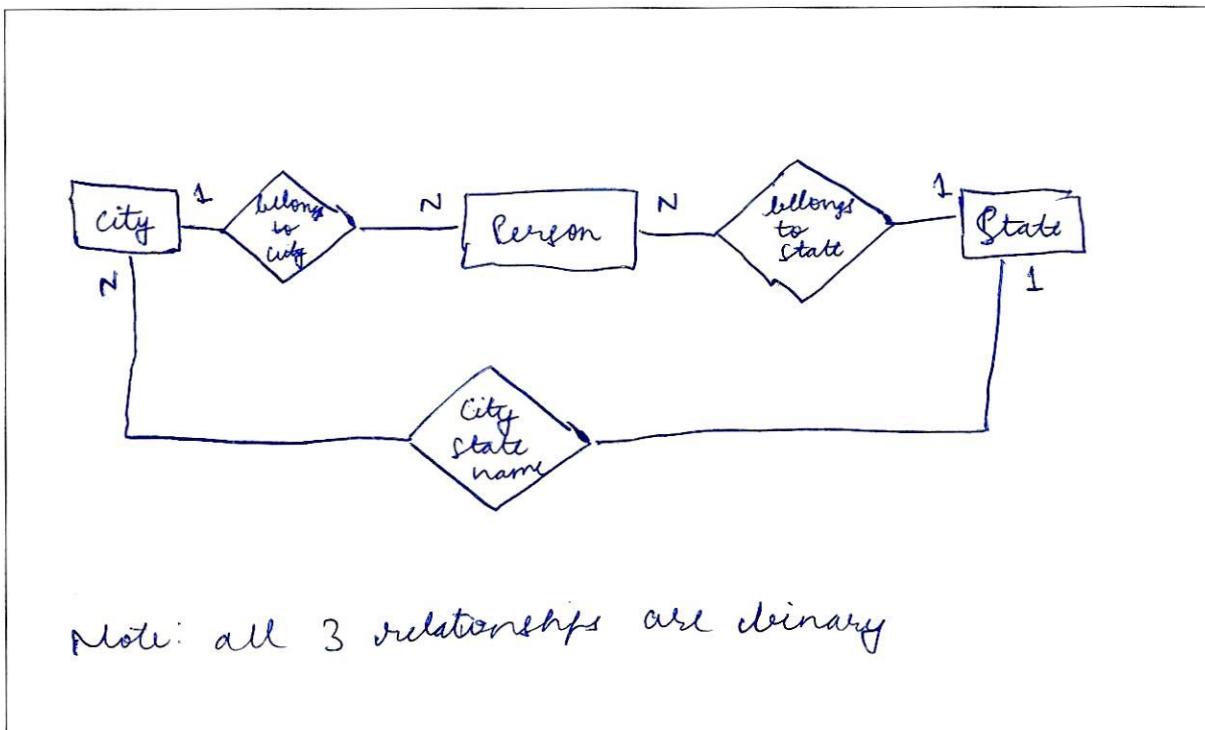
**Part 1. Database Design****Question 1 [E/R Modeling]**

3 points

Consider the following multiway relationship.



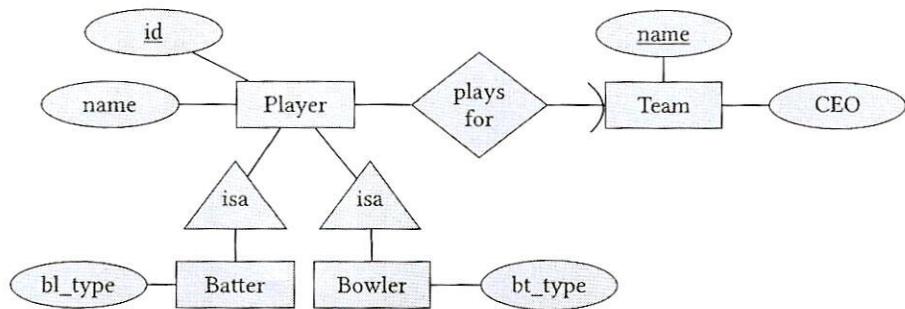
Convert the above multiway relationship into binary. In the space provided below, draw the simplified E/R diagram.



**Question 2 [E/R to Relations]**

5 points

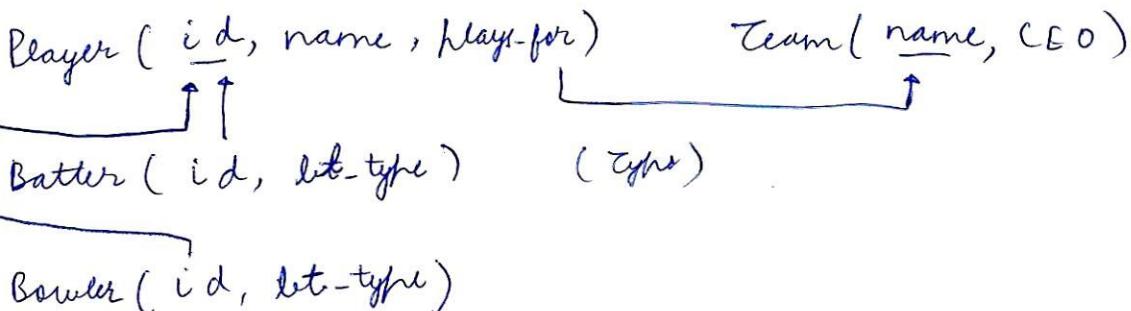
Consider the following E/R diagram below.



Convert the E/R diagram to relations in BCNF form. You may assume that there are no null values. Depict all keys and foreign keys respectively by underlying the key(s) and by drawing arrow(s) from foreign key(s) as depicted below for example relations  $R(a, b, c)$  and  $S(d, e, f)$ .

$\overline{R(a, b, c)}$

$S(d, e, f)$



This is BCNF form as Batter, Bowler have no FD's except  $\text{Player.id} = \text{Batter.id} \rightarrow \text{bl-type}$   
name is primary key, so ~~name~~  
 So F.O has superkey on the left  
 Player & Team are clearly in BCNF.  
 also are Batter & Bowler.

**Question 3 [E/R to Relations]**

2 points

Consider the following SQL schema.

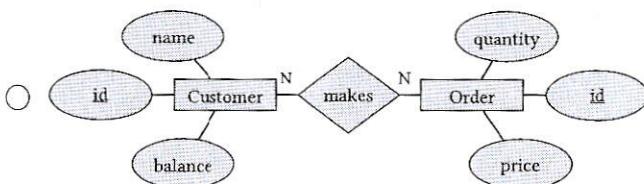
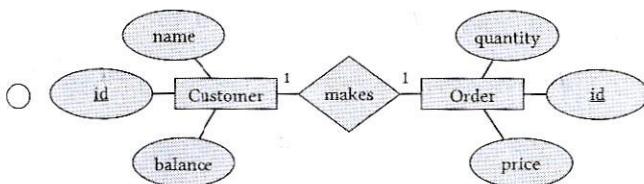
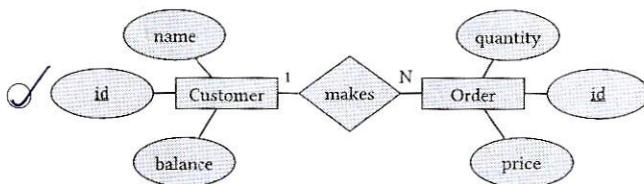
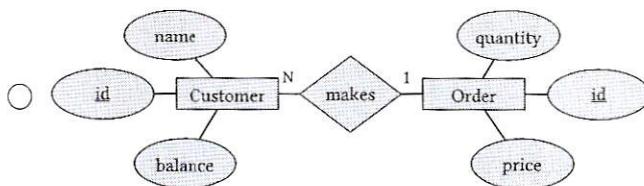
```

create table customer (
    cust_id int primary key,
    name varchar(50),
    wallet_balance float);

create table order(
    order_id int primary key,
    order_by int foreign key references customer,
    price float,
    quantity int not null);

```

Which of the following E/R diagrams could produce above SQL schema? Choose multiple options if you think that more than answer is correct. [+2 for correct answer; -1 for wrong answer; 0 if unanswered; no partial grading]

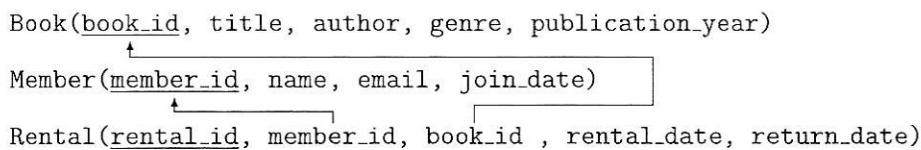


## Part 2. Relational Algebra &amp; SQL

## Question 4 [Relational Algebra]

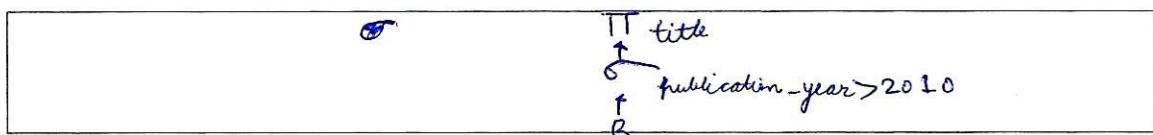
10 points

Consider the following three relations from a schema for a database representing a library system. primary keys are underlined, and foreign keys are shown via arrows.

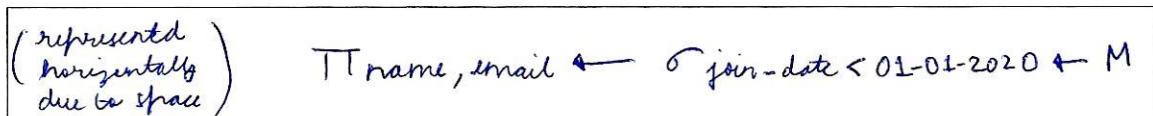


Write a relational algebra expression for the following queries using only basic operations including selection ( $\sigma$ ), projection ( $\pi$ ), natural join ( $\bowtie$ ), and aggregation ( $\gamma$ ). For succinctness, use B, M, and R for the above relations.

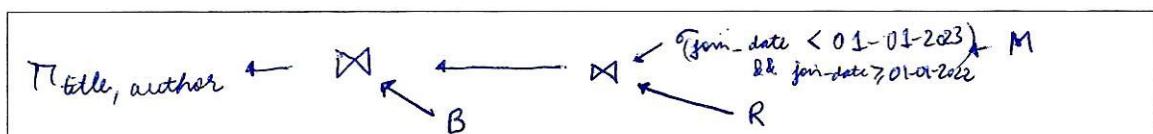
- (a) (1 point) Retrieve the titles of all books published after 2010.



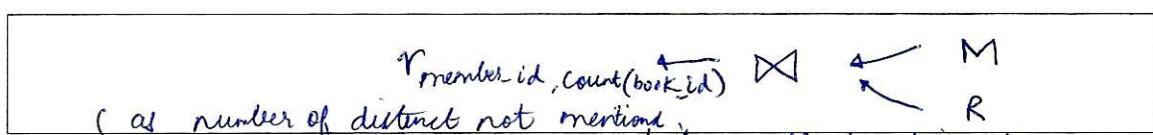
- (b) (1 point) List the names and email addresses of members who joined before 2020.



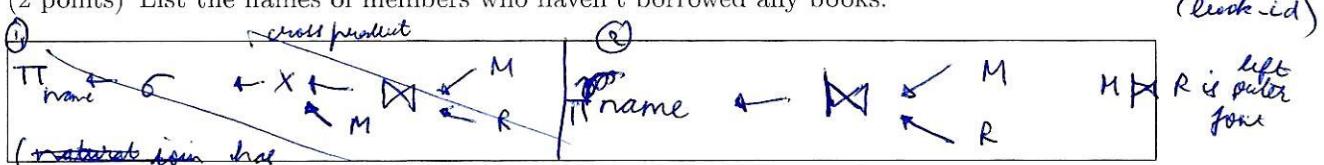
- (c) (2 points) Find titles and authors of books borrowed by members who joined in 2022.



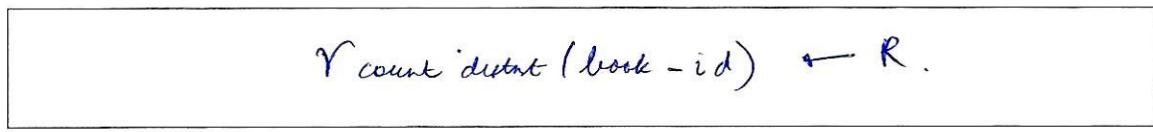
- (d) (2 points) Count the number of books borrowed by each member.



- (e) (2 points) List the names of members who haven't borrowed any books.



- (f) (2 points) Find the total number of books borrowed from the library.



**Question 5 [SQL]****10 points**

Consider the following database schema for an employee-manager database. Note that the employee-manager relationship can be hierarchical, meaning that managers themselves can have managers above them, and that a person can be managed by more than one manager.

```
create table person (
    id int primary key,
    name varchar(255),
    age int
);

create table manager (
    manager_id int,
    managed_id int,
    foreign key (manager_id) references person(id),
    foreign key (managed_id) references person(id),
    primary key (manager_id, managed_id)
);
```

(a) (5 points) Write a SQL query that returns total number of people managed by a manager.

In manager table , manager\_id refers to id of the manager (after the manager) & managed\_id refers to the id of the people managed by the manager .

```
MANAGER.manager_id,
SELECT COUNT(DISTINCT (MANAGER.managed_id))
FROM MANAGER
GROUP BY MANAGER.manager_id
AS COUNT_MANAGED
```

(b) (5 points) Write a SQL query that retrieves the name of the oldest manager.

We will use a recursive query for the same, with each level of recursion corresponding to one level of managers, and join with person after we obtain the top level.

WITH RECURSIVE FIND\_MANAGER AS

```

    (
        SELECT (*) FROM MANAGER
        JOIN FIND_MANAGER .manager_id ON
        FIND_MANAGER.manager_id = MANAGER.managed_id
    )
);

```

```

SELECT PERSON.NAME
FROM FIND_MANAGER
JOIN PERSON ON FIND_MANAGER.manager_id =
PERSON.id

```

Recursive CTE  
gives desired result,

Scratch

## Part 3. Functional Dependencies &amp; Normalization

## Question 6 [Keys]

2 points

Consider a relation  $R(A, B, C, D)$  and functional dependencies  $\{A \rightarrow B, B \rightarrow C\}$ . Which of the following statements is correct? [+2 for correct answer; -1 for wrong answer; 0 if unanswered; no partial grading]

- A is not a candidate key
- ABD is a candidate key
- AD is a candidate key
- ABCD is a candidate key

## Question 7 [Functional Dependency]

2 points

Consider two relations  $R(A, B, C)$  and  $S(C, D)$ . If the dependency  $B \rightarrow C$  holds on  $R$ , then which of the following statements is correct? [+2 for correct answer; -1 for wrong answer; 0 if unanswered; no partial grading]

- $B \rightarrow C$  must exists in  $R \times S$
- $B \rightarrow C$  must exists in  $R \bowtie S$
- $B \rightarrow C$  may not exists in  $R \times S$
- $B \rightarrow C$  may not exists in  $R \bowtie S$

$c_1 \quad c_2$   
 $R(A, B, C) \times S(C, D) \rightarrow$   
 (Because we have multiple C values,  $c_1, c_2$ )  
 Although  $B \rightarrow c_1$ , but we can't say  $B \rightarrow c_2$

## Question 8 [Normalization]

12 points

Consider a relation  $R(A, B, C, D, E)$  and functional dependencies  $F = \{D \rightarrow E, BD \rightarrow A, A \rightarrow C\}$ .

- (a) (4 points) List all minimal candidate keys in of  $R$  and a justification for your answer. Write none if you think this is no minimal candidate key. Hint: You don't have to compute closure  $F^+$  for this.

Candidate key =  $\{B, D\}$

Note that it is necessary to have B and D in the candidate key as they don't appear on RHS of any functional dependencies. We can also show that  $\{B, D\}$  is minimal because on adding even one attribute, we can safely remove it without disturbing the closure

$BD^+ = \{B, D, A, C, E\}$ . This is the reason why  $BD$  is a candidate key, or a minimal super key.

- (b) (2 points) Is  $R$  in 3NF? Briefly justify your answer.

Note that  $B$  and  $D$  are the prime attributes as only  $\{B, D\}$  is the candidate key.

Note: the F.D  $(X \rightarrow Y)$  does not satisfy any of the three conditions of 3NF:

- i) it is not trivial.
- ii)  $A$  is not a superkey.
- iii)  $C \setminus A = C$  is not a prime attribute.

$\therefore R$  is not in 3NF

(as at least one F.D violates the condition.)

- (c) (6 points) Decompose  $R$  into BCNF such that the decomposition is dependency preserving. Detail each step clearly and indicate the keys in the relation after decomposition by underlining them.

To decompose  $R$  into BCNF :- [neither trivial, nor  $A$  is superkey]  
 $A \rightarrow C$  and  $D \rightarrow E$  violate the conditions, choose  $A \rightarrow C$   
 Split into  $AC$  ( $X = A$ ,  $X^+ = AC$ ) &  $ABDE$  ( $BDE \cup A$ )

Note that  $R_1 = AC$  is BCNF

because it has  $F_1 = A \rightarrow C$  &  
 $A \rightarrow C$  has  $A$  as the superkey.

A	C
---	---

However,  $ABDE$  is not BCNF as  
 $D \rightarrow E$  violate the BCNF conditions  
 (as  $D \rightarrow E$  is neither trivial nor  $D$  is superkey)

A	B	P	E
---	---	---	---

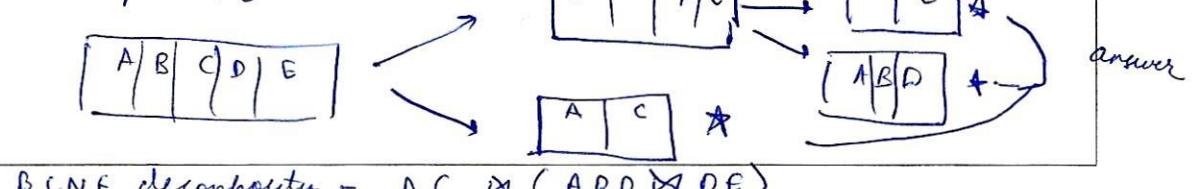
So, choose  $X = D$  ( $D \rightarrow E$  &  $X \rightarrow Y$ ) & compute  $X^+ = DE$ .

$\therefore R_2' = DE$ ,  $F_2' = D \rightarrow E$  (BCNF as  $D$  is superkey)

$R_3 = ABD$ ,  $F_3 = BD \rightarrow A$  (BCNF as  $BD$  is superkey)

We clearly observe that  $R_2'$  &  $R_3$  are in BCNF.

So we stop as we have obtained BCNF ( $AC, DE, ABD$ ) decomposition.



BCNF decomposition =  $AC \bowtie (ABD \bowtie DE)$

## Part 4. Database Storage

### Question 9 [File Organization]

2 points

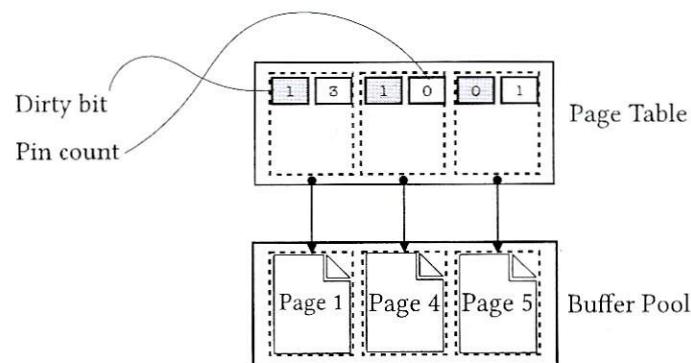
Briefly state the difference between a sequential file organization and a B+Tree file organization.

B+ tree file organization or index-based file organization where all files need not be contiguous unlike sequential file organization. Sequential helps in reducing the disk access time but indexing based approach has other methods (due to contiguous allocation).

### Question 10 [Buffer Manager]

10 points

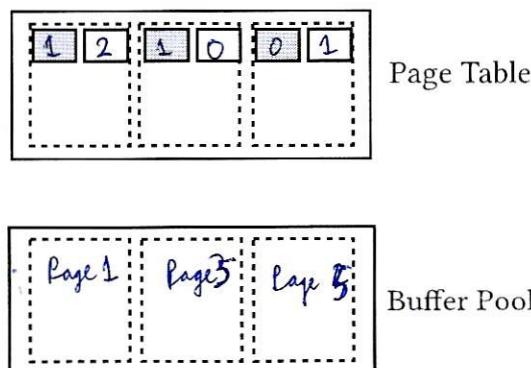
The figure below illustrates the current state of a buffer pool, which can hold 3 pages. In addition, the illustration shows the page table maintained by the buffer manager with dirty bit and pin count for pages 1, 4, and 5 that are currently in the buffer pool.



Now consider the following sequence of page requests that the buffer manager has to serve:  
 request page 4, release page 5, modify page 4, release page 4, request page 2,  
 release page 2, release page 1, request page 3, modify page 3, release page 3,  
 request page 5

Answer the following questions in the space provided. Use the space on the next page as scratch.

- (a) (5 points) Assuming that the buffer manager follows a LRU page replacement policy, show the state of the buffer pool and the page table by filling in the figure below after the above requests have been served.



- (b) (5 points) Fill the table below by specifying how many I/O(s) each request will incur. Compare LRU and MRU replacements policies and report the total.

Request	LRU	MRU
request page 4	0	0
release page 5	0	0
modify page 4	0	0
release page 4	0	0
request page 2	1	2
release page 2	0	0
release page 1	0	0
request page 3	2	1
modify page 3	0	0
release page 3	0	0
request page 5	1	0
Total I/O	4	3

p 5 evicted ←  
 but not  
 flushed  
 (LRU)

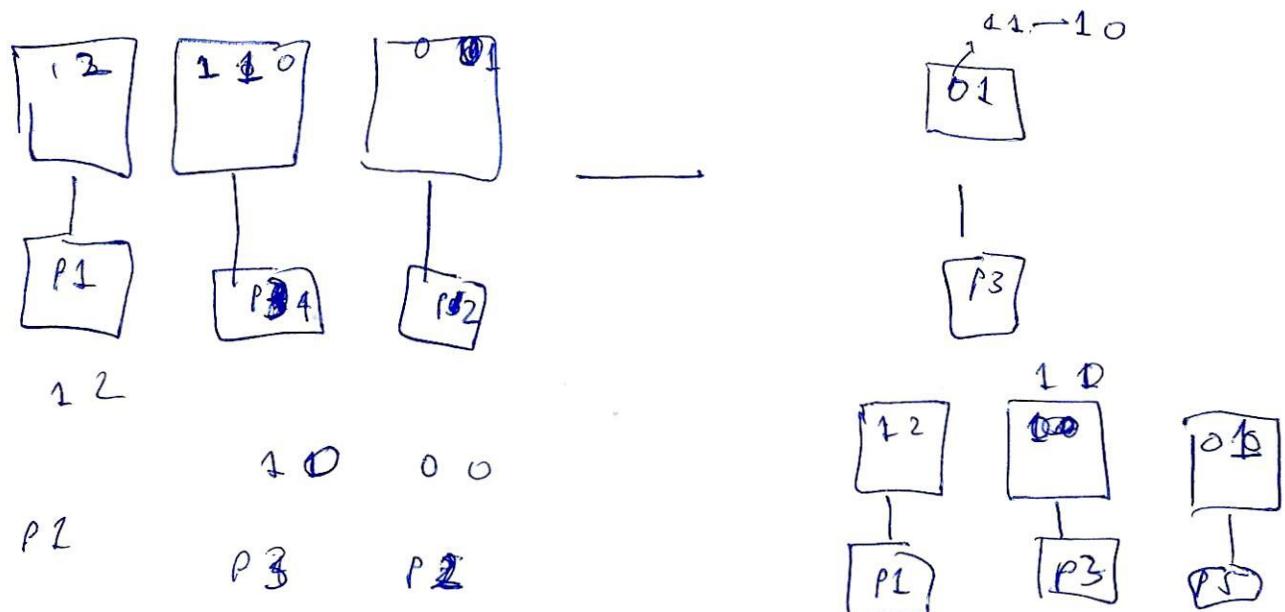
p 4 flushed ←

evicted ←  
 not  
 flushed

→ p 4 flushed (MRU)

→ p 2 evicted (MRU)

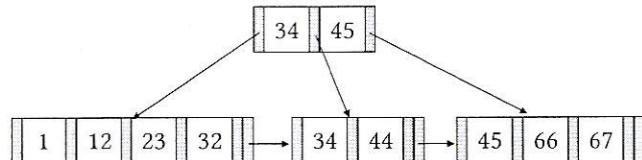
Scratch (for Question 10)



## Part 5. Indexing

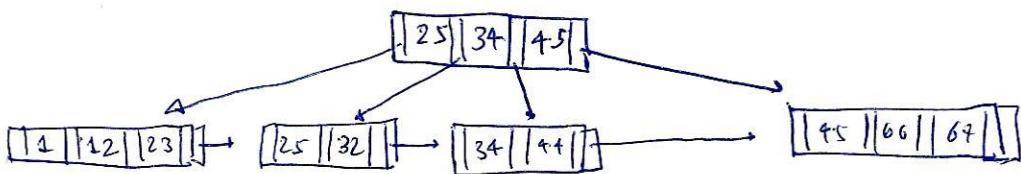
Question 11 [B<sup>+</sup>Tree]

10 points

Consider the following initial B<sup>+</sup>Tree, with  $m = 5$ . Use the next page as a scratch for this question.

- (a) (4 points) Draw the B<sup>+</sup>Tree after inserting 25.

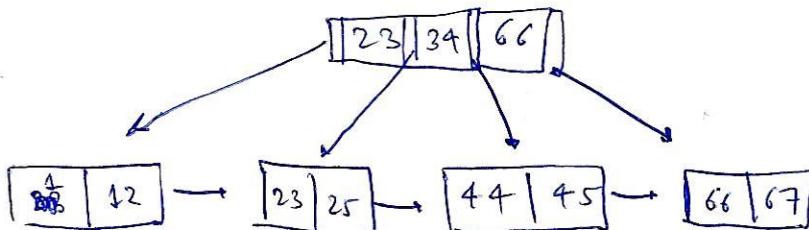
as maximum  $m - 1 = 5 - 1 = 4$  keys, we have to split the node when we insert 25.



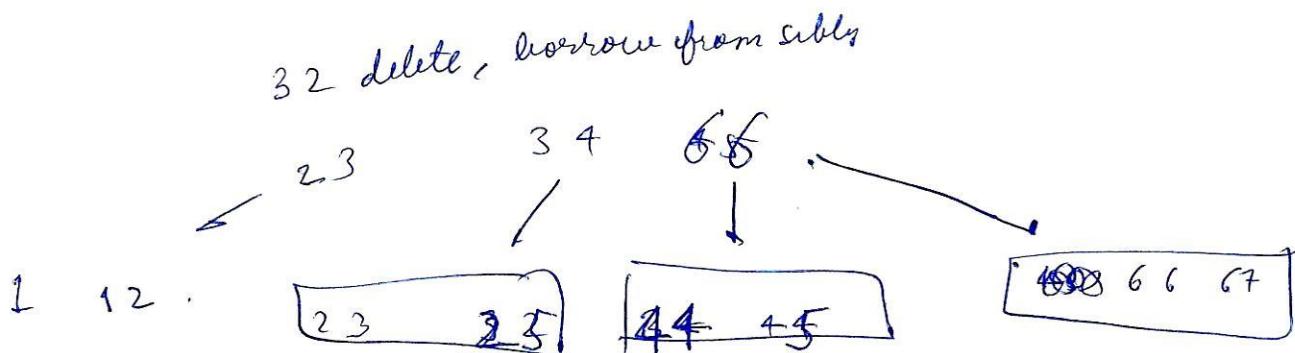
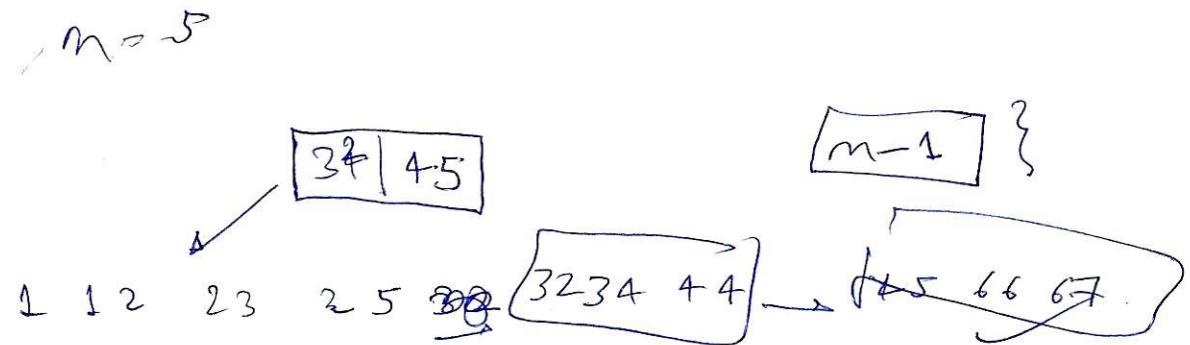
Satisfies all properties of B<sup>+</sup> tree

- (b) (6 points) After inserting 25 in part (a), draw the B<sup>+</sup>Tree after deleting keys 32 and 34.

$m - 1/2 = 2 \quad \lceil m/2 \rceil = 2$ , so B<sup>+</sup> tree cannot have more than 4 & less than 2 children.

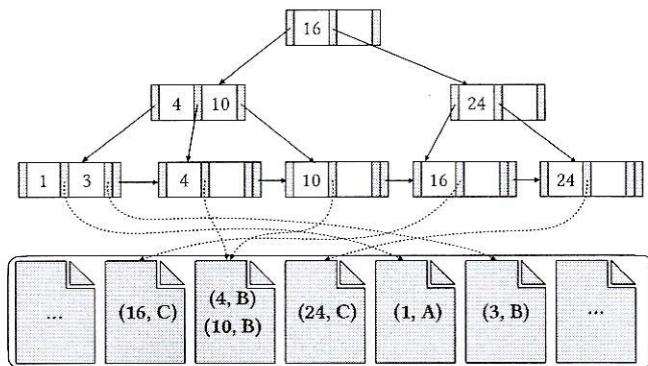


## Scratch (for Question 11)



**Question 12 [B+Trees]****12 points**

Consider the non-clustered  $B^+$ -Tree index shown below, where a node can hold at most 2 keys. Assume that each node is an index file. Answer the questions below along with brief explanation in the space provided.



- (a) (4 points) What is the maximum number entries we can add to this  $B^+$ -Tree index without increasing the height?

Since each node could hold at most 2 keys, we can add a total of 12 more entries to this  $B^+$ -tree without increasing the height [redistribution will occur] (as the first node can have 3 children, each of its children can have 3 so  $3 \times 2 = 18$  leaf node keys (9 leaf nodes)  $18 - 6 = 12$  entries)

- (b) (4 points) How many disk I/Os will the deletion of the record with  $key = 10$  incur?

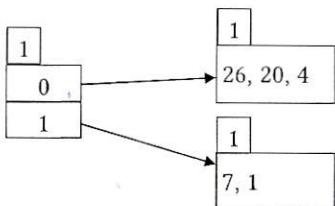
1 disk I/O on deletion of record with single key = 10. This is because the key 10 points to a block in to a page in the buffer pool, so it will be <sup>modified</sup> flushed to the disk. If we keep in the buffer pool page

- (c) (4 points) How many disk I/Os will reading of all records with  $key \geq 4$  incur?

3 disk I/O's to read all records keys 4, 10, 16, 24 because each record except (4, B), (10, B) are in different disk blocks so would need to read  $4 - 1 = 3$  disk I/O's. [one I/O will read (4, B) (10, B) simultaneously]

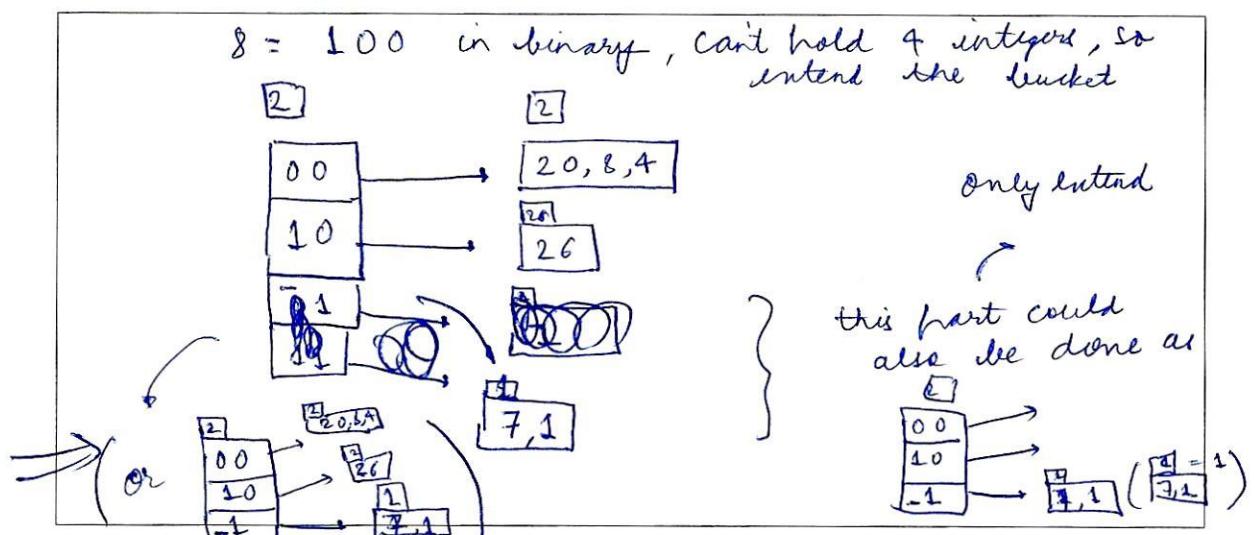
**Question 13 [Extendible Hashing]****12 points**

Consider the following hashing frame that illustrates the current state of the bucket address table and the buckets. Here, each bucket can at most hold three integers. The hash function used here takes as input an integer and returns a 5-bit binary integer that expresses the input key in its binary form. For example  $\text{hash}(4) = 00100$  and  $\text{hash}(7) = 00111$ . Further, note that the hashing scheme used here uses the *i lower order bits* of the hash(key) to determine the bucket

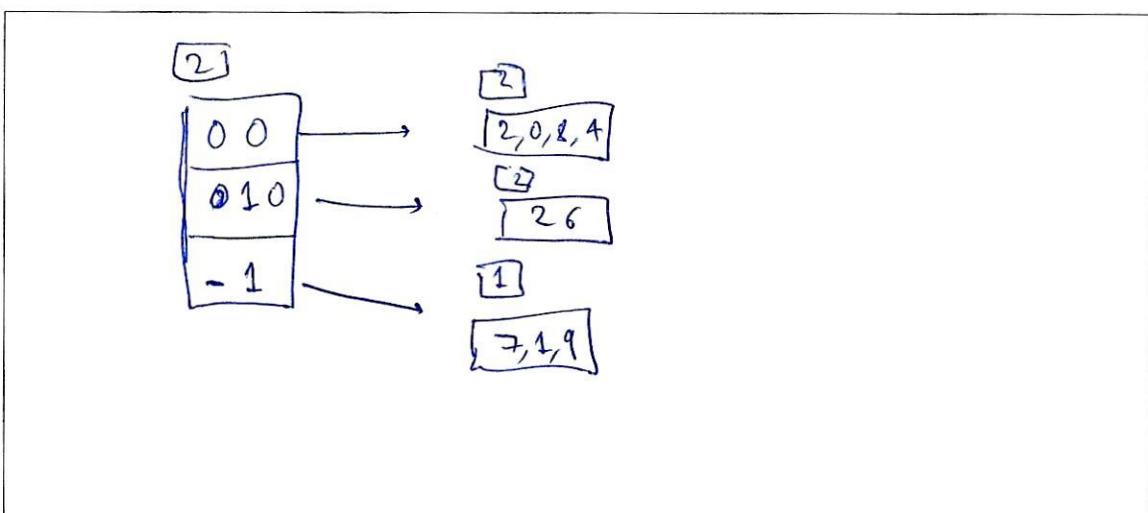


Illustrate the hashing frame in the space provided after inserting keys 8, 9, and 3. *Show all work!*

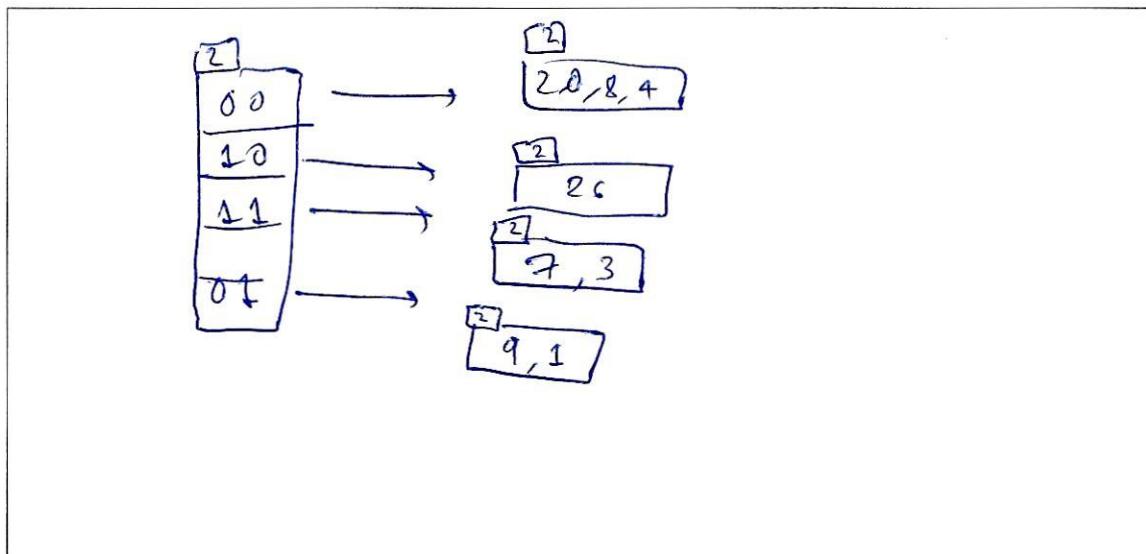
(a) (4 points) Hashing frame after inserting 8



(b) (4 points) Hashing frame after inserting 9



(c) (4 points) Hashing frame after inserting 3



Scratch (for Question 13)

**Question 14 [Bitmap Index]**

8 points

Consider the following Product table, Date table, and Sales table:

	RowId	ProdKey	ProdName	Qty	UnitPrice	Discontinued	CategoryKey
Product	1	p1	prod1	25	60	No	c1
	2	p2	prod2	45	60	Yes	c1
	3	p3	prod3	50	75	No	c2
	4	p4	prod4	50	100	Yes	c2
	5	p5	prod5	50	120	No	c3
	6	p6	prod6	70	110	Yes	c4

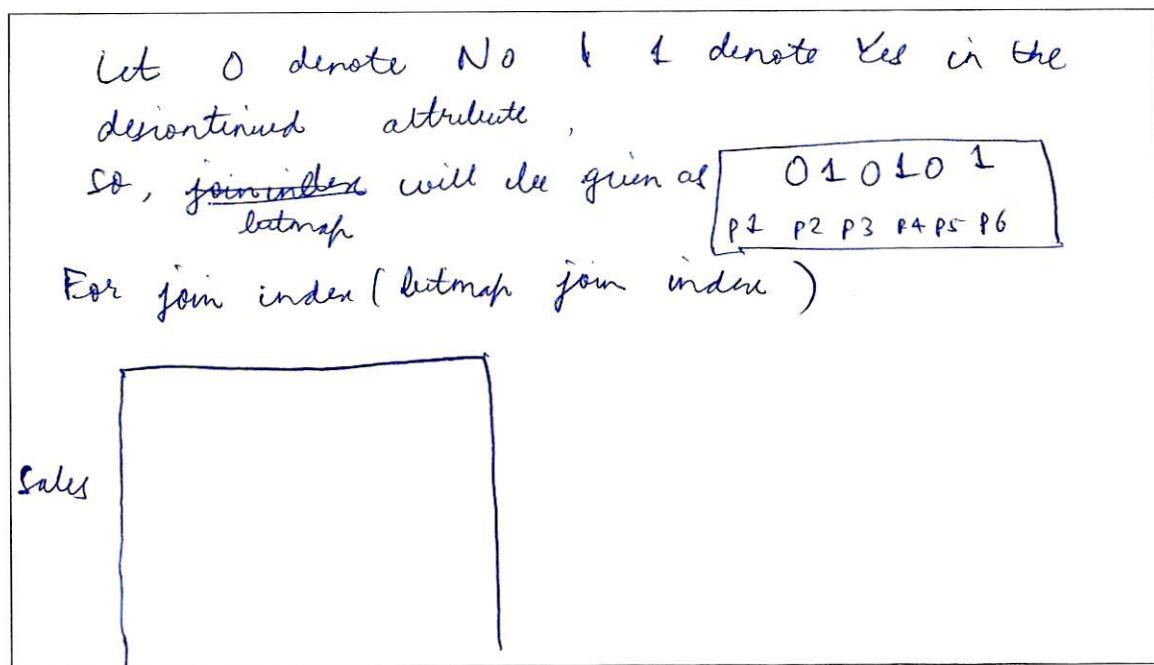
  

	RowId	DateKey	Day	Month	Year
Date	1	12	5	January	2021
	2	34	12	March	2021
	3	56	20	April	2021
	4	78	27	October	2021

	RowId	ProdKey	CustKey	DateKey	SalesAmt
Sales	1	p1	c1	12	100
	2	p1	c2	12	120
	3	p2	c2	78	150
	4	p2	c2	34	200
	5	p3	c3	12	105
	6	p4	c3	34	110
	7	p5	c4	56	125

(a) (2 points) Construct a bitmap join index for attribute Product:Discontinued and Sales.



- (b) (2 points) Construct a bitmap join index for attribute Date:Day and Sales.

<p>Take the days</p> <table border="0" style="margin-left: 20px;"> <tr><td>5 as 00</td></tr> <tr><td>12 as 01</td></tr> <tr><td>20 as 10</td></tr> <tr><td>27 as 11</td></tr> </table>	5 as 00	12 as 01	20 as 10	27 as 11	<p>} now construct the bitmap for each day</p> <p>12, 34, 56, 78</p> <table border="1" style="margin-left: 20px; border-collapse: collapse;"> <tr> <td style="width: 25px; height: 25px;">0</td><td style="width: 25px; height: 25px;">1</td><td style="width: 25px; height: 25px;">0</td><td style="width: 25px; height: 25px;">1</td><td style="width: 25px; height: 25px;">4</td><td style="width: 25px; height: 25px;">1</td><td style="width: 25px; height: 25px;">0</td><td style="width: 25px; height: 25px;">1</td><td style="width: 25px; height: 25px;">1</td></tr> <tr> <td style="text-align: center;">5</td><td style="text-align: center;">12</td><td style="text-align: center;">20</td><td style="text-align: center;">27</td><td></td><td></td><td></td><td></td><td></td></tr> </table>	0	1	0	1	4	1	0	1	1	5	12	20	27					
5 as 00																							
12 as 01																							
20 as 10																							
27 as 11																							
0	1	0	1	4	1	0	1	1															
5	12	20	27																				

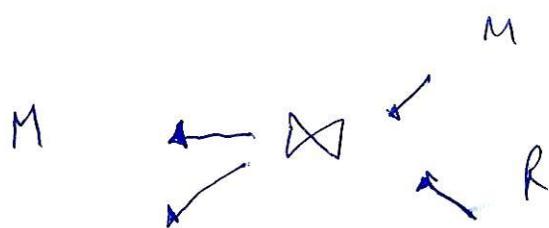
- (c) (4 points) Discuss how database can evaluate the following SQL query using bitmap indexes constructed in 14(a) and 14(b). *Show all bit operations.*

```
select prodkey, sum(salesamt)
from sales, product, date
where sales.prodkey = product.prodkey and sales.datekey = date.datekey
and product.discontinued = 'yes' and date.day < 25
group by prodkey
```

We can take the & (bitwise and of the bitmap)  
to evaluate the query containing "and"  
of the conditions "yes" & date.day < 25  
We can do it as < 25 has  
00, 01 and 10 as the possible bit values  
& yes has the possible bit value of 1.  
So when we and those records which have  
1, and 00, 01, 10, from this  
bitmap union we get our value.

Scratch

SELECT \* FROM members  
WHERE  
member\_id =



Entry No.: \_\_\_\_\_

Scratch

Entry No.: \_\_\_\_\_

Scratch

Entry No.: \_\_\_\_\_

Scratch